



UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE BIOLOGIA

DOUGLAS ADAMÓSKI MEIRA

REGULAÇÃO PÓS-TRANSCRICIONAL DA ENZIMA  
GLUTAMINASE POR HUR E SUA RELAÇÃO COM OS ALTOS  
NÍVEIS GLUTAMINOLÍTICOS TUMORAIS

POST-TRANSCRIPTIONAL REGULATION OF GLUTAMINASE  
ENZYME BY HUR AND ITS RELATIONSHIP WITH HIGH  
GLUTAMINOLYTIC LEVELS IN TUMORS

CAMPINAS

2018

**DOUGLAS ADAMÓSKI MEIRA**

**REGULAÇÃO PÓS-TRANSCRICIONAL DA ENZIMA GLUTAMINASE  
POR HUR E SUA RELAÇÃO COM OS ALTOS NÍVEIS  
GLUTAMINOLÍTICOS TUMORAIS**

**POST-TRANSCRIPTIONAL REGULATION OF GLUTAMINASE  
ENZYME BY HUR AND ITS RELATIONSHIP WITH HIGH  
GLUTAMINOLYTIC LEVELS IN TUMORS**

*Tese apresentada ao Instituto de Biologia da  
Universidade Estadual de Campinas como parte  
dos requisitos exigidos para a obtenção do  
Título de Doutor em Genética e Biologia  
Molecular, na Área de Genética Animal e  
Evolução*

*Thesis presented to the Institute of Biology of the  
University of Campinas in partial fulfillment of  
the requirements for the degree of Doctor in  
Genetics and Molecular Biology, in the area of  
Animal Genetics and Evolution*

ESTE ARQUIVO DIGITAL CORRESPONDE À  
VERSÃO FINAL DA TESE DEFENDIDA PELO  
ALUNO DOUGLAS ADAMÓSKI MEIRA E  
ORIENTADO PELA DRA. SANDRA MARTHA  
GOMES DIAS.

*Orientadora:* SANDRA MARTHA GOMES DIAS

**CAMPINAS**

**2018**

**Agência(s) de fomento e nº(s) de processo(s):** FAPESP, 2014/17820-3

**ORCID:** <https://orcid.org/0000-0001-5062-2586>

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Biologia  
Mara Janaina de Oliveira - CRB 8/6972

Ad16r Adamóski, Douglas, 1990-  
Regulação pós-transcricional da enzima glutaminase por HuR e sua  
relação com os altos níveis glutaminolíticos tumorais / Douglas Adamóski  
Meira. – Campinas, SP : [s.n.], 2018.

Orientador: Sandra Martha Gomes Dias.  
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de  
Biologia.

1. Glutaminase. 2. Expressão gênica. 3. Câncer. 4. Processamento de  
RNA. I. Dias, Sandra Martha Gomes. II. Universidade Estadual de Campinas.  
Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Post-transcriptional regulation of glutaminase enzyme by HuR and its relationship with high glutaminolytic levels in tumors

**Palavras-chave em inglês:**

Glutaminase

Gene expression

Cancer

RNA splicing

**Área de concentração:** Genética Animal e Evolução

**Titulação:** Doutor em Genética e Biologia Molecular

**Banca examinadora:**

Sandra Martha Gomes Dias [Orientador]

Fabio Papes

Paulo Sergio Lopes de Oliveira

Marcio Chaim Bajgelman

Maria Cristina Rodrigues Rangel

**Data de defesa:** 26-02-2018

**Programa de Pós-Graduação:** Genética e Biologia Molecular

Campinas, 26 de fevereiro de 2018

**COMISSÃO EXAMINADORA**

Dra. Sandra Martha Gomes Dias

Prof. Dr. Fabio Papes

Dr. Paulo Sergio Lopes de Oliveira

Dr. Marcio Chaim Bajgelman

Dra. Maria Cristina Rodrigues Rangel

*Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.*

## AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela bolsa de estudos fornecida sob o processo nº 2014/17820-3.

Ao Laboratório Nacional de Biociências (LNBio), pela infraestrutura fornecida para a execução deste projeto.

À minha orientadora, Dra. Sandra Martha Gomes Dias, que além de orientar a execução deste trabalho, contribuiu para a minha formação como um pesquisador por um todo.

Aos atuais e pretéritos alunos e funcionários do grupo de Pesquisa em Metabolismo Tumoral. Especialmente, Dra. Alessandra Girasole, Dr. Antônio Carlos Borges, Dra. Camila Cristina Pasquali, Grazielle Izalina Vasconcelos Bento, José Edwin Neciosup Quesñay, Larissa Damaceno, Dra. Larissa Menezes dos Reis, MSc. Luciana Paradela, Dr. Raghavendra Sashi Krishna Nagampalli e Renna Karoline Eloi Costa.

Ao professor Dr. Sílvio Roberto Consonni, pelas lições acerca de microscopia, acesso a técnicas histológicas utilizadas em colaborações com outros trabalhos do grupo e, principalmente, pela amizade e disposição.

Ao Dr. Andre Luis Berteli Ambrosio, não só pela colaboração na minha formação como pesquisador, mas especialmente por transformar minha concepção de representação de dados científicos. Além de, claro, compartilhar do entusiasmo com tecnologia e inovação.

Ao Dr. Daniel Maragno Trindade, pelo apoio (científico e pessoal) de todos os dias – especialmente finais de semana.

Ao Dr. Marcelo Falsarella Carazzolle, por apresentar-me o universo infinito de possibilidades do RNA-Seq.

Aos atuais e pretéritos alunos e funcionários do LNBio como um todo. Especialmente, Dra. Andrea Manrique Rincón, MSc. Anna Carolina Pereira Vieira de Carvalho, Caio Haddad Franco, Dr. Gustavo Fernando Mercaldi, Jhenifer Grunow Passos, Juliano Pereira Dias, Laís de Oliveira Ternero, Dra. Sami Yokoo, Valber Florêncio e Wanderley Pedroso da Graça – este último com o qual, entre chaves *allen* e *torx* (preferencialmente em uma combinação métrica e imperial), salvamos e *re-significamos* muitos equipamentos que imprescindíveis à conclusão deste trabalho.

Ao professor Dr. Pedro Manoel Mendes de Moraes Vieira pelo acesso ao oxígrafo *SeaHorse* e ao Felipe Corrêa da Silva pelo auxílio com o uso do mesmo.

À toda comunidade de software aberto, em especial *R core team* e todos os seus membros e colaboradores, e aos ativos membros dos diversos fóruns de discussão (em especial

do *StackOverflow*), que prontamente já haviam respondido minhas dúvidas antes mesmo delas serem concebidas.

Os resultados mostrados aqui são em parte baseados em dados gerados pelo *TCGA Research Network* (<http://cancergenome.nih.gov/>), estendendo o agradecimento ao envolvidos na produção dos resultados e, especialmente, às famílias e pacientes que, em meio ao momento de dor e sensibilidade, aceitaram contribuir com amostras.

Por fim, aos meus pais, Salomão e Wanda, que compreenderam a ausência em momentos difíceis durante estes cinco anos.

## RESUMO

A anaplerose do ciclo do TCA por glutamina, uma importante fonte energética e biossintética para células em elevada divisão celular, é prejudicada pela inibição da glutaminase. A inibição da glutaminase por CB-839 está em teste clínico como tratamento de múltiplos tumores sólidos e hematológicos. HuR (*Human antigen R*) é uma proteína codificada pelo gene *ELAVL1*. HuR apresenta três domínios de ligação à RNA e liga a elementos ricos em AU, afetando a estabilidade e o *splicing* do RNA. Achados prévios relacionam HuR à estabilização do mRNA da isoforma de glutaminase chamada de glutaminase C (GAC) em acidose renal e seu ortólogo, HuD, ao *splicing* alternativo de *GLS* em células neuronais. Com base nisto, nós hipotetizamos que HuR seria um importante regulador de glutaminase em câncer. Inicialmente nós avaliamos dados de transcriptômica pan-câncer do TCGA e observamos que pacientes com alta expressão de *ELAVL1* apresentam aumentado risco de recidiva tumoral, sendo que em câncer de mama observamos um *Cox hazard ratio* (HR) de 1,82. Além disto, utilizando dados públicos de RNA-Seq, RIP-Seq e PAR-CLIP-Seq nós mostramos que HuR liga ao intron 14 de *GLS* e controla seu *splicing*, levando à escolha da isoforma *kidney-type glutaminase* (KGA) frente à GAC. Em adição, empregando células de câncer de mama e próstata e HEK293 (com um *knock-in* produzido por CRISPR/Cas9 onde foi inserido gene para proteína fluorescente no C-terminal de KGA) nós confirmamos que HuR aumenta os níveis de KGA por afetar o *splicing* e a estabilidade do mRNA. O silenciamento lentiviral de *ELAVL1* prejudicou a proliferação, migração, invasão e aumentou o vício em glutamina de células de câncer de mama *in vitro*. O tratamento de células de câncer de mama silenciadas para *ELAVL1* com CB-839 teve efeito sinérgico no sentido de afetar proliferação e invasão celulares *in vitro*. Nós propomos HuR como um fator de prognóstico em câncer de mama, com impacto especialmente no metabolismo de glutamina através da regulação de *splicing* do gene *GLS*. Nesta tese, nós também apresentamos em anexo sete artigos publicados que participei como co-autor, sendo quatro em co-primeira autoria.

## ABSTRACT

TCA cycle anaplerosis by glutamine, an important energetic and biosynthetic source for rapidly dividing cells, is impaired by glutaminase inhibition. Glutaminase inhibition by CB-839 is currently under clinical trials for treating multiple solid and haematological tumors. HuR (Human antigen R) is a protein encoded by the *ELAVL1* gene. HuR contains three RNA-binding domains and binds cis-acting AU-rich elements, affecting mRNA stability and splicing. Previous findings have linked HuR to *GLS* mRNA stabilization in renal acidosis, and its ortholog, HuD, to *GLS* splicing in neuronal cells. We then hypothesized that HuR was an important glutaminase regulator in cancer. We first inquired a TCGA pan-cancer cohort and found that patients with higher *ELAVL1* expression had increased risk of tumor recidive (Cox HR of 1.82 for breast cancer). Moreover, by using publicly available RNA-Seq, RIP-Seq and PAR-CLIP-Seq data we showed that HuR binds to *GLS* intron 14 and controls *GLS* splicing, leading to kidney type glutaminase (KGA) choice over glutaminase C (GAC). Using breast and prostate cancer and HEK293 (with a CRISPR/Cas9 knock in) cell lines we further confirmed that HuR increases KGA by affecting splicing and mRNA stability. Lentiviral *ELAVL1* silencing impaired breast cancer cell's proliferation, migration, invasion and increased glutamine addiction *in vitro*. Treating silenced *ELAVL1* breast cancer cells with CB839 further impaired proliferation and invasion. We proposed HuR as a prognostic factor in breast cancer, with special impact on glutamine metabolism through splicing regulation. In this thesis, we also present seven published articles on which I participated as a co-author, being four as co-first author.

## LISTA DE FIGURAS

Figura 1 – Glicólise aeróbica e o ciclo do TCA .....	23
Figura 2 – Glutamina e a anaplerose alternativa do TCA .....	24
Figura 3 – Esquema dos domínios proteicos presentes nas isoformas de <i>GLS</i> e <i>GLS2</i> .....	26
Figura 4 – Interconexão entre o TCA e o ciclo da ureia no fígado .....	28
Figura 5 – Regulação do balanço ácido-base realizado pelo fígado e pelo rim .....	29
Figura 6 – Regulação da estabilidade do mRNA de GAC dentro do contexto fisiológico normal e de acidose metabólica .....	32
Figura 7 – Sítios de regulação na porção 3' não traduzida do mRNA das isoformas do gene humano <i>GLS</i> .....	33
Figura 8 – Esquema dos domínios proteicos presentes na proteína HuR e informação estrutural disponível .....	34
Figura 9 – Mecanismos de regulação do mRNA em que HuR atua e suas consequências no desenvolvimento tumoral. ....	36
Figura 10 – Representação simplificada dos vetores de expressão ou silenciamento utilizados .....	42
Figura 11 – Níveis de expressão do mRNA de <i>ELAVL1</i> em tecidos normais e tumorais humanos .....	72
Figura 12 – Razão entre os níveis de expressão de <i>ELAVL1</i> entre tecido tumoral e tecido normal adjacente .....	74
Figura 13 – Minimização de <i>p-value</i> para teste de razão de verossimilhança quanto ao nível de expressão de <i>ELAVL1</i> e sua relação com a redução do risco de recidiva tumoral .....	78
Figura 14 – Minimização de <i>p-value</i> para teste de razão de verossimilhança quanto ao nível de expressão de <i>ELAVL1</i> e sua relação com o aumento do risco de recidiva tumoral.....	80
Figura 15 – Avaliação dos elevados níveis de expressão de <i>ELAVL1</i> em pacientes de câncer de mama quanto a sobrevida, recidiva e metástase .....	82
Figura 16 – Observância da expressão de <i>ELAVL1</i> em pacientes com câncer de mama segundo a classificação molecular clinicamente utilizada.....	84
Figura 17 – Correlação entre a abundância do mRNA de <i>ELAVL1</i> e seu produto HuR em tumores de mama.....	85
Figura 18 – Níveis de expressão do mRNA de <i>GLS</i> em tecidos normais e tumorais humanos .....	86
Figura 19 – Razão entre os níveis de expressão de mRNA de <i>GLS</i> entre tecido tumoral e tecido normal adjacente.....	87
Figura 20 – Deconvolução do nível de expressão do mRNA das isoformas KGA e GAC em tecidos normais e tumorais humanos.....	90
Figura 21 – Expressão preferencial das isoformas de glutaminase em distintos tipos tumorais .....	92
Figura 22 – Correlação entre os níveis dos exons exclusivos de KGA e GAC e os níveis de <i>ELAVL1</i> em tumores de mama .....	94
Figura 23 – Busca de introns justapostos a exons envolvidos em <i>splicing</i> alternativo diferencialmente abundantes em experimento de imunoprecipitação de RNA seguida de sequenciamento (RIP-Seq) .....	96
Figura 24 – Densidade de introns justapostos a exons envolvidos em <i>splicing</i> alternativo e com ligação de HuR .....	98
Figura 25 – Análise das vias de <i>Gene Ontology</i> enriquecidas com a lista de genes em que HuR liga em ao menos um intron relacionado com <i>splicing</i> alternativo de acordo com os resultados de RIP-Seq.....	99

Figura 26 – Estratégia estatística de subamostragem e replicação aplicada para a busca de exons diferencialmente utilizados após o silenciamento de <i>ELAVLI</i> em célula HeLa .....	101
Figura 27 – Análise das vias de <i>Gene Ontology</i> enriquecidas com a lista de genes em que o silenciamento de <i>ELAVLI</i> impacta na abundância diferencial de algum de seus exons .....	101
Figura 28 – Exemplos resultantes da intersecção dos resultados de ligação à introns e escolha diferencial de exons .....	103
Figura 29 – Utilização diferencial de exons de <i>GLS</i> frente ao silenciamento de <i>ELAVLI</i> e a proposta ligação ao intron 14.....	105
Figura 30 – História evolutiva proposta para a glutaminase metazoária.....	108
Figura 31 – Expressão ectópica de <i>ELAVLI</i> e alterações observadas sobre o nível proteico e de mRNA das isoformas de glutaminase .....	110
Figura 32 – Consequências do silenciamento de <i>ELAVLI</i> sobre o nível proteico e de mRNA das isoformas de glutaminase .....	112
Figura 33 – Avaliação da dose de doxiciclina necessária para ativação do sistema de silenciamento induzível de HuR.....	113
Figura 34 – Consequências do silenciamento de <i>ELAVLI</i> em sistema induzível sobre o nível proteico das isoformas de glutaminase .....	113
Figura 35 – Silenciamento de <i>ELAVLI</i> em sistema induzível com sequência shHuR e o nível de mRNA das isoformas de glutaminase.....	114
Figura 36 – Silenciamento de <i>ELAVLI</i> em sistema induzível com sequência shHuR2 e o nível de mRNA das isoformas de glutaminase.....	115
Figura 37 – Avaliação por imunofluorescência do silenciamento de <i>ELAVLI</i> em sistema induzível e alteração em KGA .....	116
Figura 38 – Efeito do silenciamento de <i>ELAVLI</i> sobre o nível de mRNA e proteína das isoformas de glutaminase na linhagem de câncer de próstata DU-145.....	117
Figura 39 – Mensuração da estabilidade conferida por HuR via porções não traduzidas terminais do mRNA das isoformas de glutaminase .....	119
Figura 40 – Ensaio mensurando influência do intron 14 sobre o <i>splicing</i> alternativo de minigene repórter sob influência de HuR ectópica .....	122
Figura 41 – Ensaio mensurando influência do intron 14 sobre o <i>splicing</i> alternativo de minigene repórter sob influência do silenciamento de <i>ELAVLI</i> .....	123
Figura 42 – Ensaio mensurando influência do intron 14 sobre o <i>splicing</i> alternativo de minigene repórter sob influência do silenciamento de <i>ELAVLI</i> em sistema induzível .....	124
Figura 43 – Modificação do genoma de HEK293T utilizando CRISPR/Cas9 para expressão de KGA endógena com cauda fluorescente e avaliação dos efeitos do silenciamento e superexpressão de <i>ELAVLI</i> .....	127
Figura 44 – Ligação de HuR na região do intron 14 e 3'UTR das isoformas de <i>GLS</i> (KGA e GAC) .....	129
Figura 45 – Avaliação da interação de HuR e o intron 14 de <i>GLS in vitro</i> .....	130
Figura 46 – Níveis totais das proteínas HuR, KGA e GAC em células de câncer de mama .	132
Figura 47 – Nível proteico e localização celular de HuR em células de câncer de mama avaliadas por imunofluorescência .....	134
Figura 48 – Correlação entre níveis médios de HuR e localização citoplasmática mensurados por imunofluorescência com parâmetro <i>SER Valley (actin stress fibers)</i> .....	135
Figura 49 – Correlação entre a abundância de HuR e formação de <i>actin stress fibers</i> em células de câncer de mama em cultura .....	136
Figura 50 – Expressão e localização de HuR em células de câncer de mama desafiadas com a depleção de glutamina e avaliadas por imunofluorescência.....	137
Figura 51 – Padrão de presença de HuR e KGA em cultura tridimensional de células BT549 .....	139

Figura 52 – Proliferação de linhagens de câncer de mama silenciadas para <i>ELAVLI</i> .....	140
Figura 53 – Efeito do silenciamento de <i>ELAVLI</i> no crescimento de esferoides de BT549 ...	141
Figura 54 – Quantificação de <i>actin stress fibers</i> e da localização nuclear de $\beta$ -catenina frente ao silenciamento de <i>ELAVLI</i> .....	143
Figura 55 – Migração celular de células silenciadas para <i>ELAVLI</i> medida por <i>scratch-wound</i> .....	144
Figura 56 – Migração celular de células silenciadas para <i>ELAVLI</i> medida por <i>scratch-wound</i> embebido em matriz de colágeno do tipo I.....	145
Figura 57 – Invasão celular em experimento com câmara de <i>Boyden</i> com barreira de matrigel .....	146
Figura 58 – Efeitos do resgate da expressão de KGA em células com o silenciamento de <i>ELAVLI</i> .....	147
Figura 59 – Consumo de glutamina de linhagens celulares silenciadas para <i>ELAVLI</i> .....	149
Figura 60 – Dependência de glutamina para os fenótipos proliferativos e de respiração em linhagens silenciadas para <i>ELAVLI</i> .....	151
Figura 61 – Quantificação do <i>pool</i> geral de diversos metabólitos intracelulares em BT549 silenciada para <i>ELAVLI</i> .....	152
Figura 62 – Proliferação celular na combinação do silenciamento de <i>ELAVLI</i> com a inibição química de glutaminase .....	154
Figura 63 – Avaliação do IC50 do composto inibidor de glutaminase CB-839.....	155
Figura 64 – Avaliação do IC50 do composto inibidor de glutaminase BPTES .....	156
Figura 65 – Migração e invasão celulares de BT549 na combinação do silenciamento de <i>ELAVLI</i> com a inibição química de glutaminase.....	157
Figura 66 – Alterações nos níveis de expressão de <i>ELAVLI</i> e no splicing alternativo de <i>GLS</i> quando da patologia tumoral .....	159
Figura 67 – Relações observadas entre os pacientes, especificamente de câncer de mama, com alta expressão de <i>ELAVLI</i> .....	161
Figura 68 – Esquema da proposta importância fisiológica das glutaminases em uma perspectiva evolutiva .....	167
Figura 69 – Resumo da importância diferencial dos genes <i>GLS</i> e <i>GLS2</i> para a homeostase em uma perspectiva evolutiva .....	168
Figura 70 – Elementos regulatórios do (pré-)mRNA de <i>GLS</i> relacionados com HuR.....	170
Figura 71 – Consequências fenotípicas do silenciamento de <i>ELAVLI</i> em células de câncer de mama .....	176
Figura 72 – Estratégia terapêutica proposta de acordo com os resultados obtidos, com o simultâneo silenciamento de <i>ELAVLI</i> e inibição química de glutaminase. ....	178

## LISTA DE TABELAS

Tabela 1 – Sumário da literatura acerca do prognóstico clínico de pacientes de câncer de mama de acordo com o nível de HuR .....	37
Tabela 2 – Lista com os valores absolutos das razões das medianas de <i>ELAVL1</i> entre tecido tumoral e normal representadas na Figura 12A com o órgão afetado e o respectivo código empregado pelo TCGA.....	75
Tabela 3 – Minimização de <i>p-value</i> para teste de razão de verossimilhança quanto ao nível de expressão de <i>ELAVL1</i> e sua relação com a redução do risco de recidiva tumoral .....	76
Tabela 4 – Minimização de <i>p-value</i> para teste de razão de verossimilhança quanto ao nível de expressão de <i>ELAVL1</i> e sua relação com o aumento do risco de recidiva tumoral.....	77
Tabela 5 – Associações entre a expressão de <i>ELAVL1</i> e parâmetros clinicopatológicos nos tumores de mama do TCGA.....	83
Tabela 6 – Lista com os valores absolutos das razões das medianas de <i>GLS</i> entre tecido tumoral e normal representadas na Figura 19 com o órgão afetado e o respectivo código empregado pelo TCGA .....	88
Tabela 7 – Conversão entre as marcações exônicas do gene <i>GLS</i> entre as anotações RefSeq e UCSC.....	89
Tabela 8 – Lista com os valores absolutos das razões das medianas das isoformas KGA/GAC entre tecido tumoral e normal representadas na Figura 21 com o órgão afetado e o respectivo código empregado pelo TCGA.....	93
Tabela 9 – Nível de significância dos elementos relacionados com <i>splicing</i> afetados por HuR do gene <i>GLS</i> .....	106

## LISTA DE ABREVIATURAS E SIGLAS

4SU	<i>4-tiouridina</i>
6SG	<i>6-tioguanosina</i>
<i>ABCD3</i>	<i>ATP Binding Cassette Subfamily D Member 3</i>
ACC	<i>Adrenocortical carcinoma</i>
<i>ACTB</i>	<i>β-Actina</i>
ADP	<i>Adenosina Difosfato</i>
APC/C	<i>Anaphase-promoting complex/cyclosome</i>
API	<i>Application programming interface</i>
ARE	<i>AU-rich elements</i>
<i>ARG1</i>	<i>Arginase 1</i>
<i>ARMC8</i>	<i>Armadillo Repeat Containing 8</i>
<i>ASCT</i>	<i>Alanine-serine-cysteine-threonine 2</i>
<i>ASS1</i>	<i>Argininosuccinate lyase</i>
ATP	<i>Adenosina trifosfato</i>
AU	<i>Adenylate uridylate</i>
AUF1	<i>AU-rich element RNA-binding protein 1</i>
BAM	<i>Binary Alignment/Map</i>
BLCA	<i>Bladder Urothelial Carcinoma</i>
BleoR	<i>Bleomycin resistance</i>
BPTES	<i>Bis-2-(5-fenilacetamido-1,3,4-tiadiazol-2-il)ethyl sulfito</i>
BRCA	<i>Breast Cancer</i>
BSA	<i>Bovine serum albumin</i>
Cas9	<i>CRISPR associated protein 9</i>
<i>CCAT2</i>	<i>Colon cancer associated transcript 2</i>
CCDS	<i>Consensus coding sequences</i>
<i>CD47</i>	<i>Cluster of differentiation 42</i>
CDI	<i>Coefficient of drug interaction</i>
cDNA	<i>DNA complementar</i>
CESSC	<i>Cervical squamous cell carcinoma and endocervical adenocarcinoma</i>
ChIP-Seq	<i>Chromatin immunoprecipitation and sequencing</i>
CHOL	<i>Cholangiocarcinoma</i>
CI	<i>Confidence interval</i>
CMV	<i>Cytomegalovirus</i>
CO <sub>2</sub>	<i>Dióxido de carbono</i>
CoA	<i>Coenzima A</i>
COAD	<i>Colon adenocarcinoma</i>
CPM	<i>Contagem por milhão</i>
COX2	<i>Ciclooxygenase 2</i>
cPPT	<i>Central polypurine tract</i>
<i>CPS1</i>	<i>Carbamoyl-phosphate synthetase 1</i>
<i>CPS3</i>	<i>Carbamoyl-phosphate synthetase 3</i>
CRISPR	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
Ct	<i>Cycle threshold</i>
CTS	<i>Central terminal sequence</i>
DAPI	<i>4',6-diamidino-2-fenilindol</i>

DLBC	<i>Lymphoid Neoplasm Diffuse Large B-cell Lymphoma</i>
DMEM	<i>Dulbecco's Modified Eagle Medium</i>
DNA	<i>Deoxyribonucleic acid</i>
DON	<i>6-Diazo-5-oxo-L-norleucine</i>
Dox	<i>Doxiciclina</i>
DPSS	<i>Diode-pumped solid-state laser</i>
DTB	<i>Double Thymidine Block</i>
<i>ELAVL1</i>	<i>Embryonic lethal abnormal vision like 1</i>
<i>ELAVL2</i>	<i>Embryonic lethal abnormal vision like 2</i>
<i>ELAVL3</i>	<i>Embryonic lethal abnormal vision like 3</i>
<i>ELAVL4</i>	<i>Embryonic lethal abnormal vision like 4</i>
EMT	<i>Epithelial–mesenchymal transition</i>
ENA	<i>European nucleotide archive</i>
ESCA	<i>Esophageal carcinoma</i>
Faz	<i>First apoptosis signal receptor</i>
FDR	<i>False discovery ratio</i>
FPKM	<i>Fragments Per Kilobase Million</i>
FPLC	<i>Fast protein liquid chromatography</i>
FRAP	<i>Fluorescence recovery after photobleaching</i>
G2-M	<i>Gap 2 - mitosis phase</i>
GAB	<i>Glutaminase type B</i>
GABA	<i>γ-Aminobutyric Acid</i>
GAC	<i>Glutaminase type C</i>
<i>GART</i>	<i>Phosphoribosylglycinamide Formyltransferase, Phosphoribosylglycinamide Synthetase, Phosphoribosylaminoimidazole Synthetase</i>
GBM	<i>Glioblastoma multiforme</i>
GC	<i>Gas chromatography</i>
GDC	<i>Genomic data commons</i>
GDH	<i>Glutamate dehydrogenase</i>
gDNA	<i>DNA genômico</i>
GFF	<i>General Feature Format</i>
GFP	<i>Green fluorescent protein</i>
Gln	<i>Glutamina</i>
<i>GLS</i>	<i>Glutaminase</i>
<i>GLS2</i>	<i>Glutaminase 2</i>
<i>GOT1</i>	<i>Glutaminc-oxaloacetic transaminase</i>
GRC	<i>Genome reference consortium</i>
GTP	<i>Guanosine triphosphate</i>
HCl	<i>Ácido clorídrico</i>
HDR	<i>Homology directed repair</i>
HEPES	<i>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</i>
hg	<i>Human genome</i>
<i>HIF1A</i>	<i>Hypoxia inducible factor 1 alpha</i>
HIV-1	<i>Human immunodeficiency virus serotype 1</i>
HMG	<i>High mobility group</i>
HNS	<i>HuR nuclear-cytoplasmic shuttling sequence</i>
HNSC	<i>Carcinoma de células escamosas de cabeça e pescoço</i>
HR	<i>Hazard ratio</i>
HuB	<i>Human antigen B</i>

HuC	<i>Human antigen C</i>
HuD	<i>Human antigen D</i>
HuR/HuA	<i>Human antigen R</i>
IHC	<i>Immunohistochemistry</i>
IP	<i>Immunoprecipitation</i>
IRES	<i>Internal ribosome entry site</i>
IVT	<i>In vitro transcription</i>
JUN	<i>Jun proto-oncogene, AP-1 transcription factor subunit</i>
KEGG	<i>Kyoto encyclopedia of genes and genomes</i>
KEN	<i>Lysine-Glutamic Acid-Asparagine</i>
KGA	<i>Kidney-type glutaminase</i>
KICH	<i>Kidney Chromophobe</i>
KIRC	<i>Kidney renal clear cell carcinoma</i>
KIRP	<i>Kidney renal papillary cell carcinoma</i>
KSRP	<i>K-homology splicing regulatory protein</i>
LAML	<i>Acute Myeloid Leukemia</i>
LB	<i>Lysogeny broth</i>
LGA	<i>Liver-type Glutaminase</i>
LGG	<i>Brain Lower Grade Glioma</i>
LIHC	<i>Liver hepatocellular carcinoma</i>
LTR	<i>Long terminal repeat</i>
LUAD	<i>Adenocarcinoma de pulmão</i>
LUSC	<i>Carcinoma de células escamosas de pulmão</i>
MCS	<i>Multiple cloning site</i>
ME1	<i>Malic enzyme 1</i>
ME2	<i>Malic enzyme 2</i>
ME3	<i>Malic enzyme 3</i>
MESO	<i>Mesothelioma</i>
miR	<i>micro-RNA</i>
mKO2	<i>Monomeric Kusabira-orange 2</i>
MMLV	<i>Moloney murine leukemia virus</i>
MMP9	<i>Matrix metalloproteinase 9</i>
mRNA	<i>RNA mensageiro</i>
MS	<i>Mass spectrometry</i>
MYC	<i>MYC proto-oncogene, bHLH transcription factor</i>
NaCl	<i>Cloreto de sódio</i>
NAD	<i>Nicotinamide adenine dinucleotide</i>
NADH	<i>Nicotinamide adenine dinucleotide, reduced</i>
NAGS	<i>N-acetylglutamate synthase</i>
NCI	<i>National Cancer Institute</i>
NH <sub>3</sub>	<i>Amônia</i>
NH <sub>4</sub> <sup>+</sup>	<i>Amônio</i>
NHEJ	<i>Non-homologous end joining</i>
OD	<i>Optical density</i>
OTC	<i>Ornithine transcarbamylase</i>
OV	<i>Ovarian serous cystadenocarcinoma</i>
P2A	<i>Peptídeo 2A</i>
PAAD	<i>Pancreatic adenocarcinoma</i>

PAM	<i>Protospacer Adjacent Motif</i>
PAM50	<i>Prediction Analysis of Microarray 50</i>
PAR-CLIP-Seq	<i>Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation</i>
PARN	<i>Poly(A)-specific ribonuclease</i>
PBS	<i>Phosphate buffered saline</i>
PC	<i>Pyruvate carboxylase</i>
PCPG	<i>Pheochromocytoma and Paraganglioma</i>
PCR	<i>Polymerase chain reaction</i>
PDB	<i>Protein Data Bank</i>
PEG	<i>Polyethylene glycol</i>
PEI	<i>Polyethylenimine</i>
PGK	<i>Phosphoglycero kinase</i>
pH	<i>Potencial hidrogeniônico</i>
PHD	<i>Plant homeodomain</i>
PMSF	<i>Phenylmethanesulfonyl fluoride</i>
PRAD	<i>Prostate adenocarcinoma</i>
qPCR	<i>PCR quantitativa</i>
<i>RAC1</i>	<i>Ras-related C3 botulinum toxin substrate 1</i>
RBBH	<i>Reciprocal BLAST Best Hits</i>
RBP	<i>RNA-binding protein</i>
RDI	<i>Raw, description and inference</i>
READ	<i>Rectum adenocarcinoma</i>
RI	<i>Kovats retention index</i>
RIP-Seq	<i>RNA immunoprecipitation and sequencing</i>
RNA	<i>Ribonucleic acid</i>
RNAi	<i>RNA interference</i>
RNA-Seq	<i>Ribonucleic acid sequencing</i>
RPKM	<i>Reads Per Kilobase Million</i>
RPMI	<i>Roswell Park Memorial Institute medium</i>
RRE	<i>Rev responsive element</i>
RRM	<i>RNA recognition motif</i>
rRNA	<i>RNA ribossomal</i>
RT	<i>Reverse transcriptase</i>
rt	<i>Retention time</i>
S	<i>DNA Synthesis phase</i>
SARC	<i>Sarcoma</i>
SDS-PAGE	<i>Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis</i>
SER	<i>Saddles, Edges, Ridges</i>
SET	<i>Su(var)3-9, Enhancer-of-zeste and Trithorax</i>
sgRNA	<i>Single guide RNA</i>
shRNA	<i>short hairpin RNA</i>
<i>SIRT1</i>	<i>NAD-Dependent deacetylase Sirtuin 1</i>
SKCM	<i>Skin Cutaneous Melanoma</i>
SLC	<i>Solute carrier</i>
SN2	<i>System N 2</i>
<i>SNAI1</i>	<i>Snail Family Transcriptinal Repressor 1</i>
<i>SNAI2</i>	<i>Snail Family Transcriptinal Repressor 2</i>
snRNP	<i>small nuclear ribonucleoprotein</i>

STAD	<i>Stomach adenocarcinoma</i>
TBS	<i>Tris buffered saline</i>
TBS-T	<i>TBS - Tween 20</i>
TCA	<i>Tricarboxylic acid cycle</i>
TCGA	<i>The Cancer Genome Atlas</i>
tet	<i>Tetracyclina</i>
TetR	<i>Tetracycline Repressor</i>
TGCT	<i>Testicular Germ Cell Tumors</i>
TGF	<i>Transforming growth factor</i>
THCA	<i>Thyroid carcinoma</i>
THYM	<i>Thymoma</i>
TK	<i>Thymidine kinase</i>
TRA2	<i>Transformer-2 sex determining protein</i>
<i>TWIST1</i>	<i>Twist Family BHLH Transcription Factor 1</i>
UCEC	<i>Uterine Corpus Endometrial Carcinoma</i>
UCS	<i>Uterine Carcinosarcoma</i>
UCSC	<i>University of California, Santa Cruz</i>
UQ	<i>Upper-quantile normalization</i>
UTR	<i>Untranslated region</i>
UVM	<i>Uveal Melanoma</i>
<i>VEGF</i>	<i>Vascular endothelial growth factor</i>
<i>WHSC1</i>	<i>Wolf-Hirschhorn Syndrome Candidate 1</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>21</b>
1.1	O CICLO DO TCA E A DEMANDA POR CARBONOS .....	22
1.2	AS DIVERSAS GLUTAMINASES E IMPORTÂNCIA PARA O CÂNCER .....	25
1.3	A COORDENAÇÃO SISTÊMICA DAS ATIVIDADES DE <i>GLS</i> E <i>GLS2</i> EM SITUAÇÃO FISIOLÓGICA NORMAL .....	28
1.4	MECANISMOS GERAIS REGULATORIOS DE <i>GLS</i> .....	30
1.5	MECANISMO REGULATÓRIO DE <i>GLS</i> POR ESTABILIZAÇÃO DE 3' UTR EM ACIDOSE .....	31
1.6	HUR, REGULAÇÃO DE MRNA E CÂNCER .....	33
<b>2</b>	<b>OBJETIVOS</b> .....	<b>39</b>
2.1	OBJETIVOS ESPECÍFICOS .....	39
<b>3</b>	<b>MATERIAIS E MÉTODOS</b> .....	<b>40</b>
3.1	CLONAGEM E MODIFICAÇÃO DE VETORES .....	40
3.2	ANÁLISE DO CONJUNTO DE DADOS DISPONÍVEL NO <i>THE CANCER GENOME ATLAS</i> 45 .....	45
3.2.1	<i>Download</i> , organização, normalização e representação dos dados .....	45
3.2.2	Separação dos pacientes por níveis de expressão de <i>ELAVL1</i> , <i>GLS</i> e os níveis dos exons diferenciais do mRNA de GAC e KGA em tumores .....	46
3.3	ANÁLISE DE SEQUENCIAMENTOS DE RNA PREVIAMENTE DISPONÍVEIS .....	47
3.3.1	Busca e obtenção de sequências .....	47
3.3.2	Controle de qualidade dos fragmentos sequenciados e remoção das bases de baixa qualidade .....	49
3.3.3	Alinhamento contra o genoma humano .....	49
3.3.4	Produção de anotação dos introns envolvidos em <i>splicing</i> alternativo .....	50
3.3.5	Contagem de fragmentos .....	51
3.3.6	Análise de enriquecimento de vias e proposta de consequência funcional da troca dos exons nos genes exemplo .....	52
3.3.7	Visualização dos dados de PAR-CLIP-Seq alinhados e análise evolutiva .....	52
3.4	EXPERIMENTO <i>IN VITRO</i> DE INTERAÇÃO ENTRE HUR.WT E O INTRON 14 POR FRAP ( <i>FLUORESCENCE RECOVERY AFTER PHOTBLEACHING</i> ) .....	53
3.5	CULTURA DE CÉLULAS .....	55
3.6	WESTERN BLOT .....	55
3.7	TRANSFECCÃO CELULAR .....	56
3.8	VETORES VIRAIS .....	57
3.8.1	Empacotamento (sistema pLKO) .....	57
3.8.2	Empacotamento (sistemas tet-pLKO e pQC) .....	57
3.8.3	Transdução .....	58
3.9	EXTRAÇÃO DE RNA E PCR QUANTITATIVA .....	58
3.10	IMUNOPRECIPITAÇÃO DE RNA .....	59
3.11	SILENCIAMENTO INDUZÍVEL POR DOXICICLINA .....	59
3.12	IMUNOMARCAÇÃO E MICROSCOPIA DE FLUORESCÊNCIA .....	60
3.13	PROLIFERAÇÃO CELULAR EM MONOCAMADA OU TRIDIMENSIONAL ....	62
3.14	KNOCK IN DE CAUDA FLUORESCENTE NA ISOFORMA KGA DA GLUTAMINASE HUMANA .....	63
3.15	AVALIAÇÃO DO CICLO CELULAR UTILIZANDO MICROSCOPIA .....	63
3.16	ENSAIOS REPORTER .....	64
3.17	QUANTIFICAÇÃO DO CONSUMO DE GLUTAMINA .....	64

3.18	ENSAIO DE ATIVIDADE DE GLUTAMINASE COM LISADO CELULAR.....	65
3.19	MENSURAÇÃO DA DEPENDÊNCIA DE GLUTAMINA PARA RESPIRAÇÃO CELULAR UTILIZANDO O OXÍGRAFO <i>SEAHORSE XF24E</i> .....	65
3.20	MENSURAÇÃO METABÓLICA POR GAS <i>CHROMATOGRAPHY-MASS SPECTOMETRY</i> (GC-MS) .....	66
3.21	ENSAIOS DE MIGRAÇÃO E INVASÃO CELULAR .....	67
3.21.1	Extração de colágeno do tipo I .....	67
3.21.2	Ensaio de migração celular.....	68
3.21.3	Ensaio de invasão celular utilizando <i>scratch-and-wound</i> embebido em matriz .....	68
3.21.4	Ensaio de invasão celular utilizando câmara de Boyden.....	68
3.22	DESENHO EXPERIMENTAL CELULAR E AVALIAÇÃO ESTATÍSTICA.....	69
<b>4</b>	<b>RESULTADOS .....</b>	<b>70</b>
4.1	RELEVÂNCIA CLÍNICA DE HUR COMO ELEMENTO DE PROGNÓSTICO E SUA RELAÇÃO COM AS ISOFORMAS DE <i>GLS</i> .....	70
4.1.1	Correlação entre os níveis de mRNA de <i>ELAVL1</i> com parâmetros clínicos.....	71
4.1.2	Expressão de <i>GLS</i> e suas isoformas em tecidos tumorais .....	85
4.2	AVALIAÇÃO GENÔMICA DA RELAÇÃO DE HUR E <i>SPLICING</i> ALTERNATIVO 95	
4.2.1	Inferência da ligação à introns relacionados com <i>splicing</i> alternativo utilizando dados disponíveis publicamente de imunoprecipitação de RNA seguida de sequenciamento (RIP-Seq) 95	
4.2.2	Inferência de exons diferencialmente abundantes quando do silenciamento de HuR utilizando dados previamente disponíveis de sequenciamento de RNA e o contexto evolutivo dos hipotéticos elementos regulatórios.....	99
4.3	HISTÓRIA EVOLUTIVA DOS GENES <i>GLS</i> E <i>GLS2</i> .....	107
4.4	AVALIAÇÃO EXPERIMENTAL DA RELAÇÃO ENTRE HUR E A EXPRESSÃO DAS ISOFORMAS DE GLUTAMINASE .....	109
4.4.1	Consequências da modulação da expressão de <i>ELAVL1</i> nos níveis de GAC e KGA em linhagens celulares de câncer de mama .....	109
4.4.2	Consequências da modulação da expressão de <i>ELAVL1</i> nos níveis de GAC e KGA em linhagens celulares de câncer de próstata .....	117
4.4.3	Efeitos da modificação dos níveis de HuR sobre a estabilidade do mRNA das isoformas de glutaminase .....	118
4.4.4	Efeito regulatório de HuR através do intron 14 de <i>GLS</i> sobre <i>splicing</i> alternativo em um <i>minigene</i> repórter.....	121
4.4.5	Expressão de KGA endógena com fusão de mKO2 através da técnica de CRISPR/Cas9 e utilização do modelo para confirmação dos efeitos de HuR sobre a isoforma.....	125
4.4.6	Imunoprecipitação de RNA seguida de detecção por RT-PCR.....	128
4.4.7	Avaliação da interação <i>in vitro</i> de HuR e o RNA do intron 14 por <i>Fluorescence recovery after photobleaching</i> (FRAP) .....	129
4.5	FENÓTIPOS PROLIFERATIVOS, MIGRATÓRIOS E INVASIVOS DE CÉLULAS DE CÂNCER DE MAMA ALTERADOS PELOS NÍVEIS DE EXPRESSÃO DE <i>ELAVL1</i> 131	
4.5.1	Caracterização da expressão endógena de HuR em um painel de células de câncer de mama 132	
4.5.2	Silenciamento de HuR e o fenótipo obtido em células de câncer de mama.....	139
4.5.3	Fenótipo proliferativo .....	139
4.5.4	Fenótipo migratório e invasivo.....	141
4.5.5	Resgate fenotípico com a re-expressão ectópica da isoforma KGA em células com <i>ELAVL1</i> silenciada .....	147

4.5.6	Fenótipo glutaminolítico .....	148
4.5.7	Combinação do silenciamento de HuR a inibição química da glutaminólise celular .....	153
<b>5</b>	<b>DISCUSSÃO .....</b>	<b>158</b>
5.1	ANÁLISE PAN-CANCER DA IMPORTÂNCIA DE HUR PARA O PROGNÓSTICO DE PACIENTES .....	158
5.2	IMPORTÂNCIA DE HUR NA REGULAÇÃO DA GLUTAMINASE.....	162
5.3	ANÁLISE PAN-CANCER DA IMPORTÂNCIA DE <i>GLS</i> .....	163
5.3.1	A importância das isoformas KGA e GAC para o câncer .....	163
5.3.2	Relação entre expressão de <i>ELAVL1</i> e níveis de mRNA de GAC e KGA em diferentes tumores .....	164
5.3.3	Análise evolutiva e funcional da evolução das isoformas e aparecimento de GAC ..	165
5.4	IMPORTÂNCIA DE HUR PARA A ESCOLHA DAS ISOFORMAS DE <i>GLS</i> EM MODELOS CELULARES DE CÂNCER .....	168
5.4.1	Regulação via estabilização de mRNA – interação com 3’UTR.....	169
5.4.2	Regulação via <i>splicing</i> alternativo.....	170
5.4.3	Atuação de HuR na escolha das isoformas de <i>GLS</i> e fenótipo de linhagens tumorais de mama .....	173
<b>6</b>	<b>CONCLUSÃO.....</b>	<b>180</b>
	<b>REFERÊNCIAS .....</b>	<b>183</b>
	<b>APÊNDICES .....</b>	<b>203</b>
	<b>ANEXOS .....</b>	<b>206</b>

## 1 INTRODUÇÃO

O funcionamento da nossa sociedade depende de um conjunto de normas e hierarquias que todos devem aderir, sendo que, quando uma das pequenas partes constituintes não o faz, toda a estrutura pode ser abalada. Analogamente, nosso organismo procede da mesma maneira. Assim, estando constituído de células ordenadas em tecidos, órgãos e sistemas, os mesmos atuam em concordância com as funções fisiológicas e contribuem para o funcionamento do todo. Quando do acúmulo de desvios e inadequações além do limiar de tolerância do organismo, instala-se a patologia. Células com comportamento aberrante que se colocam à parte do funcionamento especializado tecidual e proliferam a taxas anormais impactam em um aumento local de massa celular, formando o que se conhece como tumores. Tumores benignos (não invasivos), usualmente não impactam na sobrevivência do paciente, salvo quando do acometimento de órgãos vitais. Tumores malignos (ou câncer) apresentam células agressivas que saem de seu contexto local e invadem outros tecidos, próximos ou distantes, podendo impactar negativamente com a função deste tecido e, ultimamente, do órgão ao qual o mesmo pertence. A invasividade de tumores malignos é a principal causa da mortalidade associada ao câncer<sup>1</sup>.

Assim, o tumor é formado por células que defletiram da ontogenia programada. Um grande esforço foi feito no decorrer do século XX, assim como ainda é feito, no sentido de se estabelecer quais características mínimas a célula deve adquirir para tornar-se insensível ao seu dever como constituinte de um ente multicelular. Propõem-se que células tumorais adquirem diversas características no decorrer de seu caminho de transformação – tais como a instabilidade do genoma, aumento de mutações e a inflamação regional<sup>2</sup> – que permitem o desenvolvimento de habilidades como a insensibilidade à sinais antiproliferativos (descontrolando a proliferação tecidual), auto-suficiência em sinalização de crescimento (desencadeando a proliferação contínua autônoma), capacidade replicativa indefinida (ignorando a restrição telomérica), evasão da apoptose (impedindo a morte celular programada), promoção da angiogênese (para suprir com nutrientes o crescente número de células), a capacidade de invadir tecido adjacente e distante (impactando na atividade de tecidos e órgãos)<sup>3</sup>, escape à imunovigilância (impedindo a resposta do organismo frente à agressão) e o desequilíbrio energético celular<sup>2</sup>. Em particular, a última característica está envolvida com a manutenção de três aspectos básicos: a geração de energia para suprir reações bioquímicas diversas, produção de blocos biossintéticos e a manutenção do balanço óxido-redutivo celular. Na base destas alterações energéticas está o intenso consumo de glicose, usado como base de

metodologias de diagnóstico tumoral, e glutamina, ambos importantes e praticamente ubíquas fontes de energia e blocos biossintéticos<sup>4</sup>.

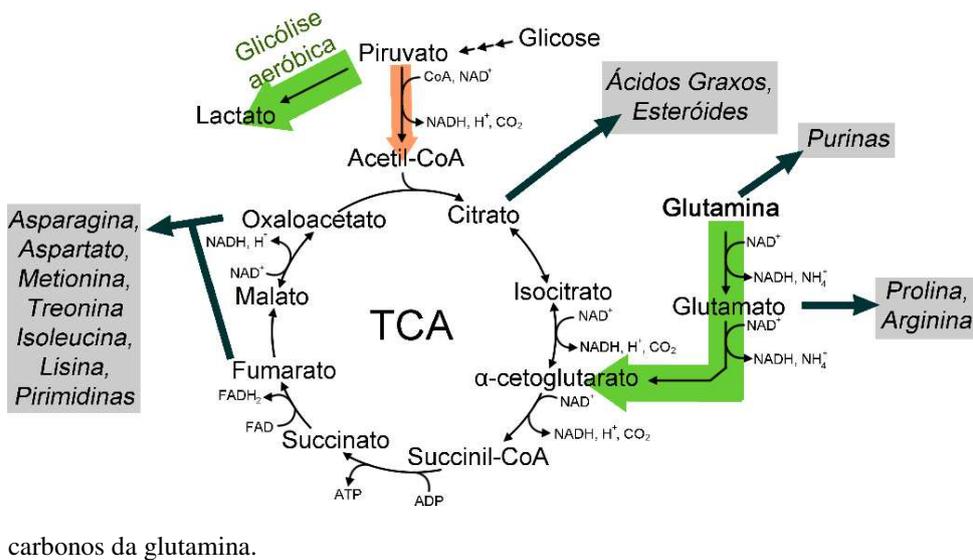
Há longa data, o laureado com um prêmio Nobel em 1931, Otto Heinrich Warburg, observou o elevado processo fermentativo ativo em células tumorais, gerando quantidades energéticas equiparáveis ao processo respiratório – observações designadas como glicólise aeróbica<sup>5</sup>. A utilização de seu sobrenome como descrição para este fenômeno veio muito posteriormente, em 1972, sendo apenas no atual século que a real efervescência de publicações na área tornou-se realidade<sup>6</sup>. Tal fenótipo é consequência da desregulação da via glicolítica, e se conecta intimamente com um deslocamento do comportamento do ciclo dos ácidos tricarboxílicos (TCA) de uma função primariamente oxidativa energética para um papel biossintético.

### 1.1 O CICLO DO TCA E A DEMANDA POR CARBONOS

A glicólise, processo pelo qual a célula transforma a glicose internalizada em duas moléculas de piruvato, provê energia e blocos biossintéticos envolvidos com a síntese de ácidos nucleicos – imprescindíveis para a proliferação celular. Sua molécula final, entretanto, pode ser transformada em lactato (processo da fisiologia normal em condições anaeróbicas, principalmente do tecido muscular levado à esforço extremo e sem o aporte necessário de oxigênio), ou condensada à coenzima A, formando o acetil-CoA, capaz de adentrar o ciclo do TCA<sup>7</sup>. O primeiro cenário, *a priori* energeticamente desfavorável, é amplamente utilizado por diversos tipos tumorais, possivelmente por distintos fatores, dentre eles: o descompromisso com o balanço energético do organismo (ou seja, a utilização indiscriminada dos nutrientes constantemente disponíveis na corrente sanguínea)<sup>8</sup>, a necessidade do encaminhamento dos carbonos da glicose para processos biossintéticos e não unicamente energéticos<sup>8</sup>, e as vantagens conferidas pelo estabelecimento de um ambiente ácido no desenvolvimento de mutações e manutenção das características tumorais<sup>9,10</sup>.

Entretanto, para a manutenção da biossíntese de diversos blocos celulares, as células proliferativas tumorais também dependem do TCA, dada a plêiade de vias que demandam de seus intermediários para a execução (Figura 1). Assim, sua importância majoritariamente energética é substituída pela função de fornecimento de moléculas para a síntese de aminoácidos diversos, de pirimidinas e purinas, de ácidos graxos e esteroides, entre outros<sup>7,11</sup>. Com a cataplerose (evasão) de moléculas do ciclo do TCA para estes processos biossintéticos, o mesmo precisa ser reabastecido por novas fontes de carbono. Piruvato pode

também ser convertido a outras moléculas, gerando vias anapleróticas alternativas. Assim, através da ação da piruvato carboxilase (*PC*), pode formar oxaloacetato e pela ação da enzima málica (*ME1*, *ME2* ou *ME3*), formar malato. Outras vias anapleróticas são exploradas, principalmente envolvendo aminoácidos, sendo um dos principais a glutamina<sup>11</sup>. Enquanto que a anaplerose do TCA via piruvato nas vias metabólicas descritas acima é importante para alguns tipos tumorais (especialmente de *PC* para pulmão<sup>12</sup>) a suplementação por glutamina tem se mostrado importante para diversos tipos tumorais tais como pulmão, pâncreas e mama<sup>13-17</sup>.



**Figura 1 – Glicólise aeróbica e o ciclo do TCA**

A intensificação da via glutaminolítica (seta em verde na direita), associada a cataplerose do TCA (para alimentar a síntese de moléculas nas caixas cinzas) definem a necessidade de suplementação do mesmo (anaplerose, seta em verde na direita) pelos

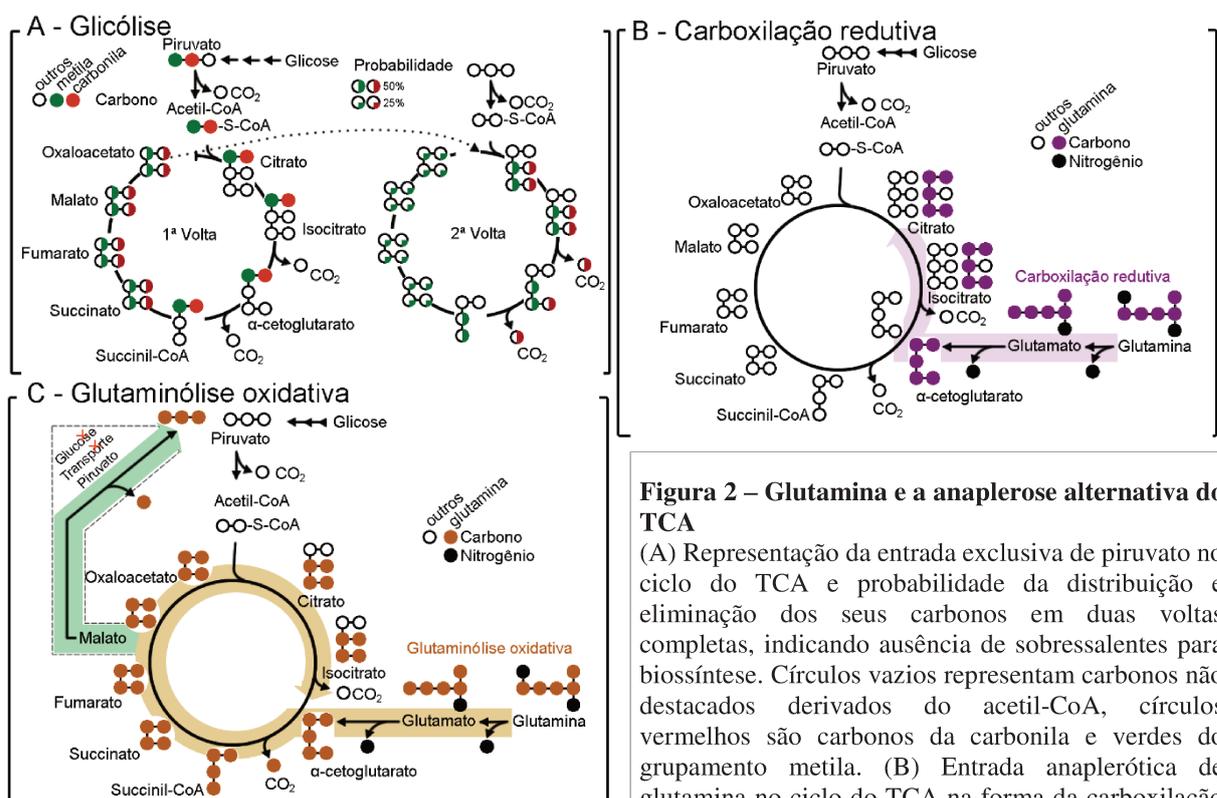
carbonos da glutamina.

A glutamina é o aminoácido mais abundante do plasma sanguíneo, sendo também utilizada para a síntese de peptídeos, glutatona, ácido fólico e neurotransmissores<sup>18</sup>. Para servir de substrato energético e biossintético, a glutamina é oxidada de forma análoga à oxidação da glicose, processo conhecido como glutaminólise. Neste processo a glutamina é convertida em glutamato, e este é transformado em α-cetoglutarato, que pode então entrar no ciclo do TCA (Figura 1).

Quando o ciclo do TCA não é utilizado como fonte de intermediários metabólicos para outras vias biossintéticas, ou seja, é empregado apenas para a produção de energia, o carbono do grupo carbonila doado pelo acetil-CoA (círculos vermelhos no esquema superior da Figura 2A) é inteiramente liberado como dióxido de carbono ( $\text{CO}_2$ ) em sua segunda volta no ciclo; já o carbono do grupamento metila (círculos verdes no esquema) é incorporado ao *pool* e depende de mais voltas para ser liberado como  $\text{CO}_2$ . Tal fato na produção do succinato, a molécula se torna simétrica e só é possível definir que um dos carbonos centrais é originado da metila e um dos carbonos externos é originado da carbonila (denotada pela probabilidade de 50% no esquema). Em situação de ausência de demanda biossintética celular, a oxidação do

piruvato glicolítico no TCA mantém o balanço de carbonos no ciclo, dado que toda entrada é compensada pela saída na forma de  $\text{CO}_2$ .

Criando-se uma necessidade de utilização do TCA como fonte biossintética, dá-se a evasão de carbonos do mesmo; isto, associado a intensificação da perda de carbonos glicolíticos na forma de lactato, faz-se com seja imprescindível a suplementação do ciclo. O mesmo é, como comentado acima, frequentemente suplementado com glutamina através da degradação da mesma em glutamato pela enzima glutaminase e posterior conversão em  $\alpha$ -cetogluturato. O metabolismo de glutamina no ciclo pode ser, em situações específicas, tais como hipóxia<sup>19,20</sup> ou em mutações em enzimas do ciclo do TCA ou da cadeia de transporte de elétrons<sup>21</sup>, realizado no sentido redutor (carboxilação redutiva); neste sentido, cinco dos seis carbonos do citrato passam a ser oriundos da glutamina e esta se torna imprescindível para a produção do mesmo (Figura 2B). Em situação mais geral acontece a metabolização da glutamina no sentido oxidativo (Figura 2C), com seus carbonos se mesclando aos carbonos do acetil-CoA derivado de glicose, gerando citrato com quatro dos seis carbonos oriundos da glutamina e os outros dois da glicose. Em outra situação específica, tais como limitação de glicose ou inibição do transporte de piruvato<sup>22,23</sup>, pode haver a conversão de malato (oriundo de glutamina) a piruvato e, então, em acetil-CoA, formando, assim, citrato com os 6 carbonos derivados da glutamina (Figura 2C).



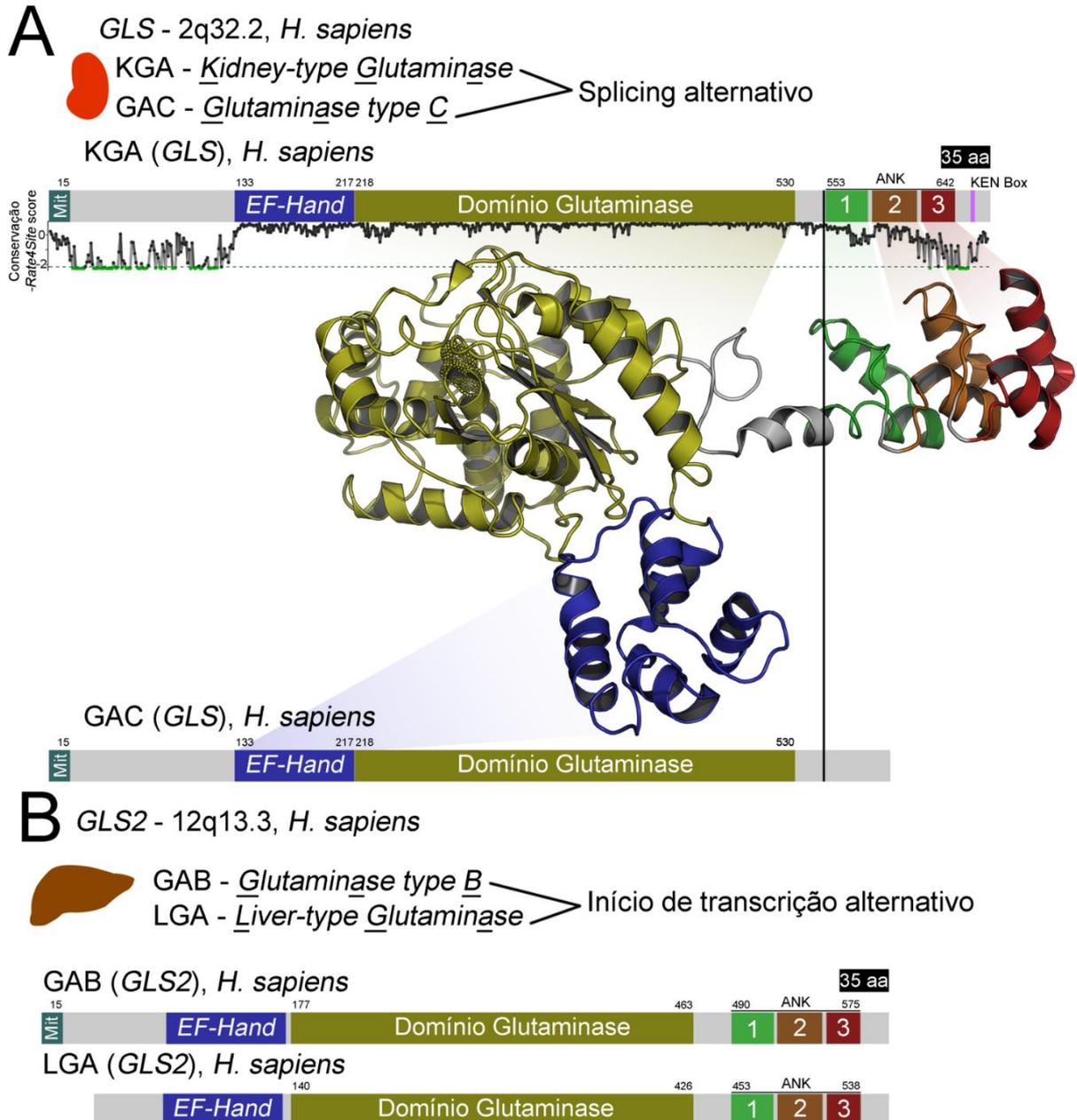
**Figura 2 – Glutamina e a anaplerose alternativa do TCA**

(A) Representação da entrada exclusiva de piruvato no ciclo do TCA e probabilidade da distribuição e eliminação dos seus carbonos em duas voltas completas, indicando ausência de sobressalentes para biossíntese. Círculos vazios representam carbonos não destacados derivados do acetil-CoA, círculos vermelhos são carbonos da carbonila e verdes do grupamento metila. (B) Entrada anaplerótica de glutamina no ciclo do TCA na forma da carboxilação

reduzida, com o provimento de carbonos especialmente para a geração de citrato e posterior síntese de lipídeos. Círculos vazios representam carbonos não-oriundos da glutamina, círculos roxos representam carbonos da glutamina e círculos pretos representam nitrogênios. (C) Entrada anaplerótica de glutamina no ciclo do TCA, tanto na forma da oxidação completa (com geração de piruvato e acetil-CoA), quanto na oxidação parcial com complexação de acetil-CoA da glicose. Caminho destacado em verde representa a utilização da enzima málica para a produção de piruvato, especialmente na ausência de glicose ou de transporte de piruvato para a mitocôndria, que pode sustentar o ciclo apenas a partir de glutamina. Círculos vazios representam carbonos não-oriundos da glutamina, círculos marrons representam carbonos da glutamina e círculos pretos representam nitrogênios. Glicólise (A) adaptada de Garret & Grisham<sup>7</sup> e glutaminólise (B e C) de Murai et al<sup>23</sup>.

## 1.2 AS DIVERSAS GLUTAMINASES E IMPORTÂNCIA PARA O CÂNCER

São observados, no genoma humano, dois *loci* responsáveis pela codificação de glutaminase: um está no cromossomo 2, e é chamado de gene *GLS*, e o outro está no cromossomo 12, sendo chamado de gene *GLS2* (Figura 3)<sup>24</sup>. O primeiro gene codifica, ainda, duas isoformas distintas geradas por *splicing* alternativo: a mais longa, KGA (*Kidney-type GlutaminAse*), e a mais curta, GAC (*GlutaminAse C*). Para gerar GAC há a troca dos 4 exons terminais da KGA por um distinto, contendo códon de parada e região terminal não traduzida (3'UTR) própria<sup>25</sup>. Por compartilharem o mesmo N-terminal, ambas apresentam a mesma sequência de localização mitocondrial (sendo, entretanto, a isoforma KGA eventualmente observada fora da mitocôndria<sup>26</sup>), e o motivo *EF-hand like*, de ligação a cálcio ainda não experimentalmente evidenciada. Além disto, ambas apresentam o mesmo domínio catalítico. A isoforma KGA se distingue no C-terminal pela presença de domínios do tipo anquirina, normalmente envolvido em interação proteína-proteína, enquanto GAC dispõe de um C-terminal desestruturado<sup>26,27</sup>. Apesar de KGA apresentar estas repetições anquirina ainda não foi descrita nenhuma proteína com a qual ela interagisse via este domínio. Funcionalmente, ambas isoformas apresentam atividade aumentada quando como tetrâmeros (oligomerização esta dependente da concentração proteica e da presença de fosfato<sup>26</sup>), sendo que GAC apresenta maior eficiência catalítica por formar mais facilmente oligômeros maiores do que tetrâmeros<sup>27-29</sup>. Quanto à expressão no contexto fisiológico normal, KGA foi descrita presente em todos os tecidos, com exceção do fígado, enquanto a presença da GAC já foi descrita em coração, pulmão, rim e pâncreas, tecidos mamários e células nervosas<sup>30</sup>. Observando os dados do projeto *Human BodyMap 2.0* disponibilizados como pistas do visualizador ENSEMBL, dados de RNA-Seq suportam a expressão de KGA e GAC em tecido adiposo, adrenal, mama, cólon, coração, rim, fígado, pulmão, linfonodos, ovário, próstata, testículos, osso e tireoide.



**Figura 3 – Esquema dos domínios proteicos presentes nas isoformas de *GLS* e *GLS2***

(A) Isoformas distintas do gene *GLS*. Gráfico abaixo da representação esquemática de KGA (superior) indica a conservação da sequência proteica em vertebrados. Pontos verdes definem os aminoácidos menos conservados em vertebrados acima do percentil de 90% do escore *Rate4Site*<sup>31</sup>. A estrutura proteica cristalográfica representada é da isoforma KGA (PDB:5UQE, cadeia B<sup>32</sup>) representada como *cartoon*; domínio N-terminal *EF-Hand-like* (com similaridade ao domínio de ligação a cálcio *EF-Hand*<sup>33</sup>) em azul, domínio de glutaminase em amarelo e C-terminal com as 3 repetições de anquirina (*Ank* 1, 2 e 3, *Ankyrin repeats*) denotadas em verde, marrom e vermelho (hélice de ligação em cinza). Linha vertical indica a posição do *splicing* alternativo que altera a porção c-terminal da isoforma GAC (inferior). (B) Representação esquemática das duas isoformas distintas de *GLS2*, GAB e LGA. Mit: sequência de direcionamento mitocondrial.

O gene *GLS2*, por sua vez, expressa suas duas isoformas em função da escolha distinta de promotores: o uso da região mais distal leva à expressão da isoforma GAB (*G*lutaminase *B*), enquanto que promotor localizado dentro do primeiro intron propicia a codificação de LGA (*L*iver-type *G*lutaminase)<sup>34</sup>. Construção contendo região do domínio N-

terminal comum a ambas isoformas, região catalítica e domínio C-terminal de *GLS2* apresenta menor atividade do que KGA e GAC<sup>26</sup>. Dados de atividade enzimática comparativos de LGA com GAB são pouco abundantes, mas indicam uma possível atividade superior da primeira, mais curta, em relação à segunda, mais longa<sup>34</sup>. GAB, com mais informação na literatura, apresenta localização mitocondrial menos evidente – sendo até mesmo nuclear em determinados tipos celulares<sup>35</sup> –, incapacidade de formar oligômeros maiores que tetrameros<sup>27,36</sup> e ativação por amônio em mecanismo de *feedback* positivo<sup>37–39</sup>. Na fisiologia normal, GAB é expressa no fígado, pâncreas e cérebro<sup>40</sup>, enquanto LGA foi detectada em fígado e cérebro<sup>34</sup>.

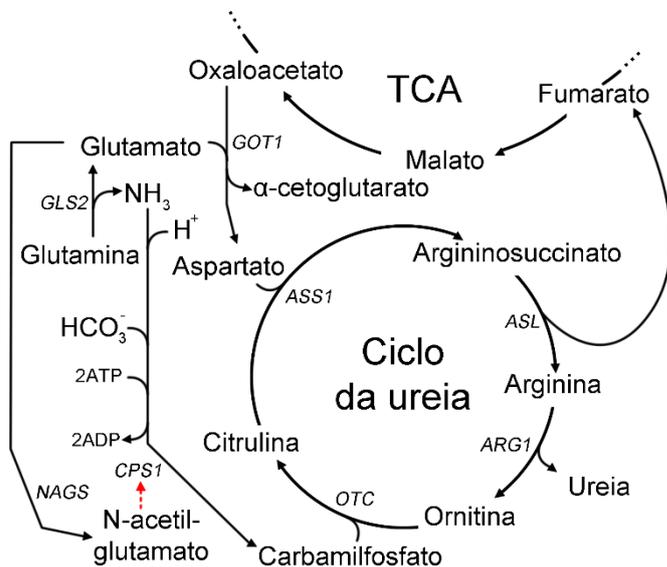
A inibição da via glutaminolítica é uma estratégia terapêutica para câncer que foi inicialmente proposta na década de 60 com a depleção da glutamina sérica (via administração da enzima glutaminase-asparaginase purificada de *Pseudomonas* sp). Tal abordagem reduziu a proliferação tumoral *in vivo* mas não trouxe benefício na sobrevivência aos ratos avaliados dada a toxicidade causada pela ausência de glutamina circulante<sup>41</sup>. Mais recentemente, as abundantes observações da correlação positiva entre os níveis glutaminolíticos e a capacidade proliferativa tumoral<sup>42</sup> têm justificado a contínua busca da compreensão da relação entre o metabolismo tumoral e este aminoácido. Como colocado acima, além de contribuir com a anaplerose do ciclo do TCA em condições de ativação de vias biossintéticas importantes para a proliferação celular, a glutamina tem mostrado atuar em outros mecanismos de malignidade. Especificamente, a glutamina, via seu produto metabólico,  $\alpha$ -cetoglutarato<sup>43</sup>, promove a transição epitélio-mesênquima (EMT), na qual as células se tornam indiferenciadas e podem migrar pelo organismo colonizando outros tecidos e órgãos<sup>44</sup>.

Neste sentido, estratégias têm sido traçadas para a inibição dos produtos de *GLS* pelo uso de inibidores competitivos (DON<sup>45</sup>) ou pelo uso de inibidores alostéricos. Nesta última alternativa, o inibidor CB-839 está em estudos clínicos fase I e II tanto em monoterapia ou em associação com quimioterápicos conhecidos, tendo mostrado resultados promissores para tumores de pâncreas<sup>46</sup>, leucemias<sup>47</sup>, pulmão<sup>48</sup>, mama<sup>17</sup>, entre outros.

Já *GLS2*, por ser alvo transcricional de p53<sup>49,50</sup>, é relacionado na literatura com a supressão tumoral. Em específico, foi mostrado que sua interação com a pequena GTPase Rac1 impede a ação promotora de metástase e invasão da mesma<sup>51</sup>. Todavia, existem também casos em que inibidores específicos à *GLS2* reduzem a proliferação celular e a formação de colônias independente de ancoragem<sup>52</sup>, implicando que a ação pró- ou anti-tumoral da mesma é contexto dependente.

### 1.3 A COORDENAÇÃO SISTÊMICA DAS ATIVIDADES DE *GLS* E *GLS2* EM SITUAÇÃO FISIOLÓGICA NORMAL

O gene *GLS2*, cuja expressão majoritariamente se dá nos tecidos hepáticos, apresenta importante papel no contexto do ciclo da ureia (Figura 4, com todos os passos enzimáticos, Figura 5 em um contexto sistêmico), em possíveis múltiplos níveis: provimento de amônio como substrato da enzima carbamoil fosfato sintase (*CPS1*)<sup>53</sup> que é o ponto de entrada do nitrogênio no ciclo; provimento de glutamato para a transaminase glutâmico-oxaloacética mitocondrial (*GOT1*) que fornece aspartato ao ciclo<sup>36</sup>; provimento de glutamato para a enzima n-acetilglutamato sintase (*NAGS*), que produz o n-acetilglutamato, regulador positivo de *CPS1*, passo lento e regulatório do ciclo<sup>54</sup>. Quando do não funcionamento adequado do fluxo do ciclo, mesmo na abundante presença de amônio, o fígado pode despejar este excesso na corrente sanguínea e levar à um quadro de encefalopatia hepática<sup>55</sup>. Esta síntese contínua de amônio, mesmo na presença de seus elevados níveis é possível, inclusive, por consequência do *feedback* positivo para amônio em *GLS2*<sup>37,38,54</sup>. Diferente de *GLS2*, *GLS* é inibida por amônio<sup>56</sup>, e potencialmente com pouca atividade potencial no fígado.



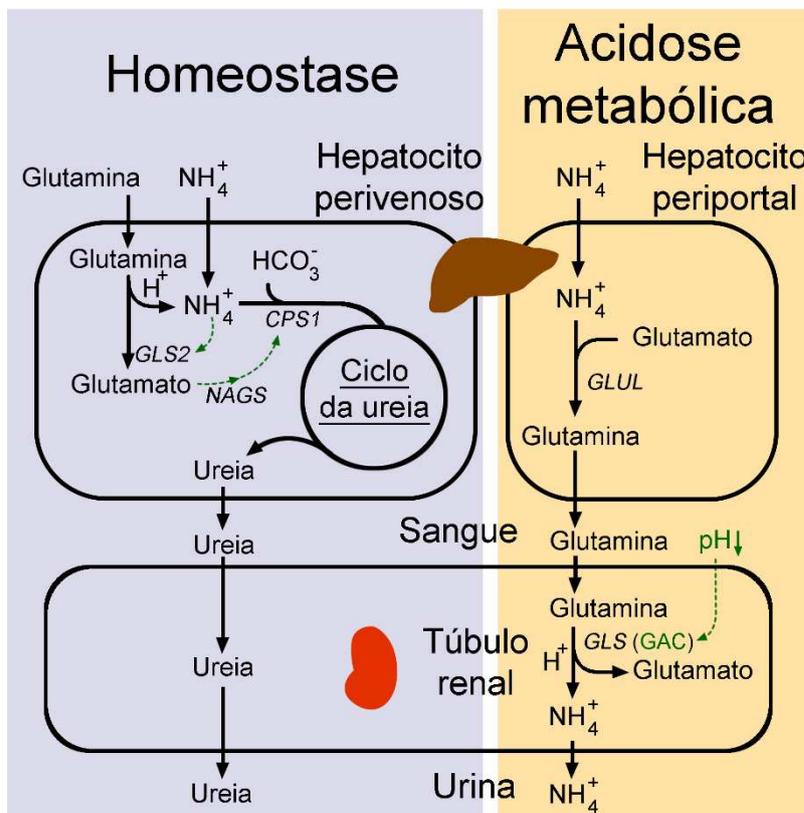
**Figura 4 – Interconexão entre o TCA e o ciclo da ureia no fígado**

*GOT1*: Glutamic-oxaloacetic transaminase 1. *GLS2*: Glutaminase 2. *ASS1*: Argininosuccinate synthase 1. *ASL*: Argininosuccinate lyase. *ARG1*: Arginase 1. *OTC*: ornithine transcarbamylase. *CPS1*: Carbamoyl-phosphate synthase 1. *NAGS*: N-acetylglutamate synthase.

Fisiologicamente, o produto do gene *GLS* se mostra importante especialmente no cérebro e rim. No primeiro caso, KGA é a isoforma que mais contribui para a síntese de ácido  $\gamma$ -aminobutírico (GABA), como molécula neurotransmissora<sup>30</sup>. Nos rins de mamíferos, por sua vez, em situação de acidose sanguínea, as glutaminases KGA e especialmente a GAC estão

envolvidas com o mecanismo de secreção de amônio ( $\text{NH}_4^+$ ), produto da catálise da glutamina<sup>57</sup>.

As atividades de *GLS* e *GLS2* se comunicam através do eixo hepático-renal de excreção nitrogenada<sup>58</sup> (Figura 5, esquerda). A glutamina, via *GLS2*, é degradada em glutamato e amônia, que pode se combinar a um próton e formar amônio (situação privilegiada dado o pKa elevado, de 9,25). Este (oriundo da degradação de glutamina e também coletado da circulação sanguínea<sup>54,58</sup>), pela ação da carbamoilfosfato sintase (*CPS1*), se une a bicarbonato e adentra ao ciclo da ureia, que é produzida e despejada na circulação sanguínea para posterior secreção pelo rim<sup>58</sup>.



**Figura 5 – Regulação do balanço ácido-base realizado pelo fígado e pelo rim**

Quadro esquerdo indica a situação de homeostase e na direita cenário de acidose metabólica. Adaptado de Häussinger<sup>58</sup>.

Em acidose metabólica (Figura 5, direita), há a redução do pH sanguíneo, o que, por sua vez, promove a regulação negativa do ciclo da ureia no fígado (devido a reduzida expressão da aquaporina mitocondrial 8, envolvida no transporte do amônio citoplasmático para a utilização pela enzima carbamoilfosfato sintase, *CPS1*), com a finalidade de economia de íons bicarbonato. Com a parada do ciclo da ureia, há acúmulo de  $\text{NH}_4^+$ , o qual precisa ser secretado do órgão. Para tanto, células de hepatócitos perivenosos (presentes na saída do sangue do fígado), complexam, pela ação da glutamina sintase, amônia ( $\text{NH}_3$ ) com glutamato, produzindo glutamina. Sendo uma carregadora não tóxica de nitrogênio na corrente sanguínea, a glutamina

é recaptada pelos rins onde é transformada em glutamato pelas glutaminases renais. Nesta reação, a amônia é formada novamente, complexada com prótons e liberada na urina na forma do ácido fraco amônio, eliminando, desta maneira o excesso de prótons da corrente sanguínea. Isto, aliado a economia de base bicarbonato no fígado, leva ao aumento e regularização do pH do sangue<sup>58</sup>. O processamento de glutamina nos rins quando da acidose metabólica se dá pela GAC, isoforma mais cataliticamente ativa, cujo mRNA é estabilizado por intermédio da sua 3'UTR, ao menos em *H. sapiens* e *S. scrofa*<sup>25</sup>, em mecanismo que será explicado abaixo.

#### 1.4 MECANISMOS GERAIS REGULATORIOS DE *GLS*

A avaliação de amplas coortes de pacientes mostram que a expressão aumentada de *GLS*, em relação ao tecido normal adjacente, só é observada em poucos tipos tumorais, tais como cólon, esôfago, fígado, estômago, tireoide e cabeça-e-pescoço<sup>59</sup>. O único fator de transcrição até então descrito como regulador da expressão de *GLS* é c-Jun, o qual se liga diretamente ao promotor do gene e aumenta sua expressão<sup>60</sup>.

O oncogene *MYC* (também conhecido como c-Myc) foi o primeiro a ser descrito como importante para o aumento da atividade de glutaminase em situação tumoral. *MYC* regula diversos passos do metabolismo de glutamina, seja pelo aumento de transportadores citoplasmáticos de glutamina, tais como os transportadores de aminoácidos neutros SLC (*solute carriers*, como *ASCT* e *SN2*)<sup>61</sup>, como também pela repressão transcricional da expressão de dois microRNAs, miR-23a/b, que almejam a 3'UTR da isoforma KGA de *GLS*<sup>14</sup>. Esta regulação impacta na abundância proteica da isoforma GAC por mecanismo não descrito<sup>14</sup>. Tumores com amplificação de *MYC* tornam-se “viciados” em glutamina<sup>62</sup>. Além de *MYC*, mostrou-se que fibroblastos murinos transformados com Dbl oncogênico (uma *Guanine Exchange Factor* necessária a ativação de Rho GTPases), tiveram o fenótipo tumoral inibido pelo composto 968, cujo alvo identificado foi a isoforma GAC<sup>63</sup>.

Nosso grupo demonstrou, em colaboração com o pesquisador George Callin do MD Anderson (Texas), que o *long non-coding RNA* (lncRNA) CCAT2, localizado na banda cromossômica 8q24 e detentor de SNP associado a risco de câncer de cólon rs6983267, regula o metabolismo tumoral *in vitro* e *in vivo*, de maneira alelo-específica<sup>64</sup>, artigo no qual colaborei diretamente (Anexo 1). O alelo G do SNP mencionado é associado com predisposição ao desenvolvimento de câncer de cólon, enquanto que o alelo T não o é. Os diferentes alelos (G e T) de CCAT2 se ligam ao complexo *Cleavage Factor I* (CFIm) com afinidades distintas, tendo G maior afinidade. O complexo CFIm-CCAT2 direciona o *splicing* alternativo de *GLS*, através

da interação com um sítio poli(A) no intron 14 do pre-mRNA de *GLS*, escolhendo a isoforma GAC, cataliticamente mais ativa<sup>64</sup>.

Em nível pós-traducional, a isoforma KGA, em específico, apresenta uma sequência Lisina - Ácido Glutâmico - Asparagina (KEN Box) e motivo de destruição (D-Box) no seu c-terminal. O complexo promotor da anáfase-ciclossomo-Cdh1 (APC/C-Cdh1) modula a degradação de KGA através desta região quando a célula avança para a fase G2/M do ciclo celular, impactando na síntese de bases nitrogenadas<sup>65</sup>. Este mecanismo afeta o consumo de glutamina e secreção de amônio durante estas fases do ciclo, com provável relação à demandas biossintéticas celulares específicas<sup>66</sup>. Por fim, diversas outras modificações pós-traducionais do tipo fosforilação, acetilação e succinilação já foram descritas para *GLS*, com impacto sobre a atividade e glutaminase e o fenótipo tumoral ainda desconhecido<sup>27,67</sup>.

## 1.5 MECANISMO REGULATÓRIO DE *GLS* POR ESTABILIZAÇÃO DE 3' UTR EM ACIDOSE

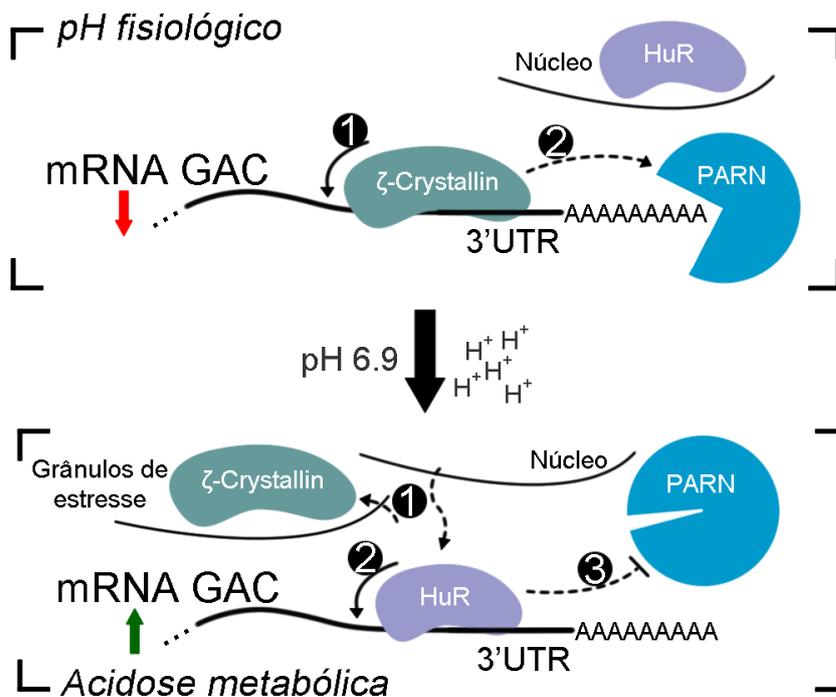
A regulação da estabilidade de um mRNA é um importante mecanismo de controle da expressão gênica. Tal estabilização pode ser modulada pela presença de elementos regulatórios em cis nos mRNAs e sua interação com o elementos regulatórios em trans<sup>68</sup>. Em células de mamíferos, os elementos ricos em adenosina e uridina presentes no 3'UTR mRNA, chamados de *AU-rich elements* (AREs), foram identificados pela sua habilidade em levar tais RNAs à degradação rápida, em um mecanismo dependente de de-adenilação (encurtamento da cauda poli-A)<sup>69</sup>.

O evento fisiológico já mencionado de acidose metabólica impacta o metabolismo renal de glutamina por elevar em até cinco vezes a atividade glutaminolítica em ratos expostos por sete dias à água acidificada com cloreto de amônio. O aumento da atividade glutaminolítica renal foi acompanhado de um aumento da abundância proteica de mesma ordem; tal efeito não foi observado na atividade glutaminolítica cerebral ou intestinal<sup>70</sup>. Este aumento no nível proteico foi explicado pelo aumento nos níveis de mRNA<sup>71</sup>. A restauração de água em pH normal para o consumo levava ao retorno dos níveis iniciais de mRNA, proteína e atividade<sup>72</sup>.

Em um modelo *in vitro* de células epiteliais do túbulo renal suíno LLC-PK-F+, foi mostrado um aumento dos níveis de mRNA de glutaminase em até oito vezes após a exposição a um pH 6.9, frente ao usual 7.4<sup>73</sup>. Tal aumento se mostrou relacionado com sequências localizadas na 3'UTR do mRNA, que impactaram no aumento da estabilidade em pH 6.9 de transcrito repórter e do *shift* em gel de RNA quando da incubação com extratos de células

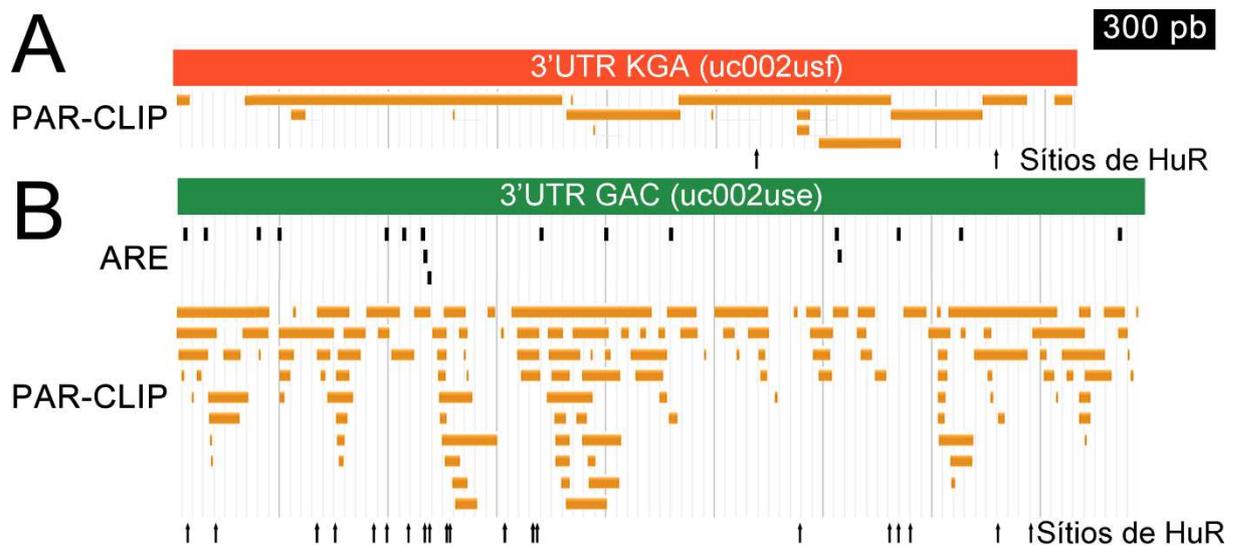
mantidas nesse pH<sup>73-75</sup>. Eventualmente, observou-se que tais elementos responsivos a pH estavam presentes no mRNA da isoforma GAC<sup>25</sup>.

Em adição, mostrou-se que a degradação do mRNA em pH fisiológico mediada pela 3'UTR (Figura 6) seria, no contexto renal, comandada pela proteína  $\zeta$ -Crystallin.  $\zeta$ -Crystallin liga-se a 3'UTR (passo 1 em pH fisiológico, Figura 6) e recruta a ribonuclease específica de poli(A) (PARN, passo 2 em pH fisiológico, Figura 6) que remove cauda protetora poli(A). A estabilização em situação de acidose, por outro lado, é dirigida pela proteína HuR, que transloca do núcleo para o citoplasma enquanto  $\zeta$ -Crystallin migra para grânulos de estresse (passo 1 em acidose metabólica, Figura 6) e interage com a 3'UTR do mRNA de GAC (passo 2 em acidose metabólica, Figura 6), o que impede o processo de degradação<sup>57</sup> (passo 3 em acidose metabólica, Figura 6).



**Figura 6 – Regulação da estabilidade do mRNA de GAC dentro do contexto fisiológico normal e de acidose metabólica**  
Adaptado de Ibrahim *et al*<sup>57</sup>.

Na Figura 7 mostram-se as sequências presentes no 3'UTR das isoformas de *GLS* já experimentalmente validadas como passíveis de ligação à diversas *RNA binding proteins* (sequências de ligação identificadas como caixas laranjas na linha identificada como “PAR-CLIP” no esquema). Observa-se que no mRNA de GAC estas sequências são mais abundantes; além disto, somente o mRNA de GAC contém *AU-rich elements* (ARE) preditos por *AURA2*. As setas verticais indicam a ligação especificamente da proteína HuR, observada também em KGA, mesmo na ausência de sítios preditos pela ferramenta.



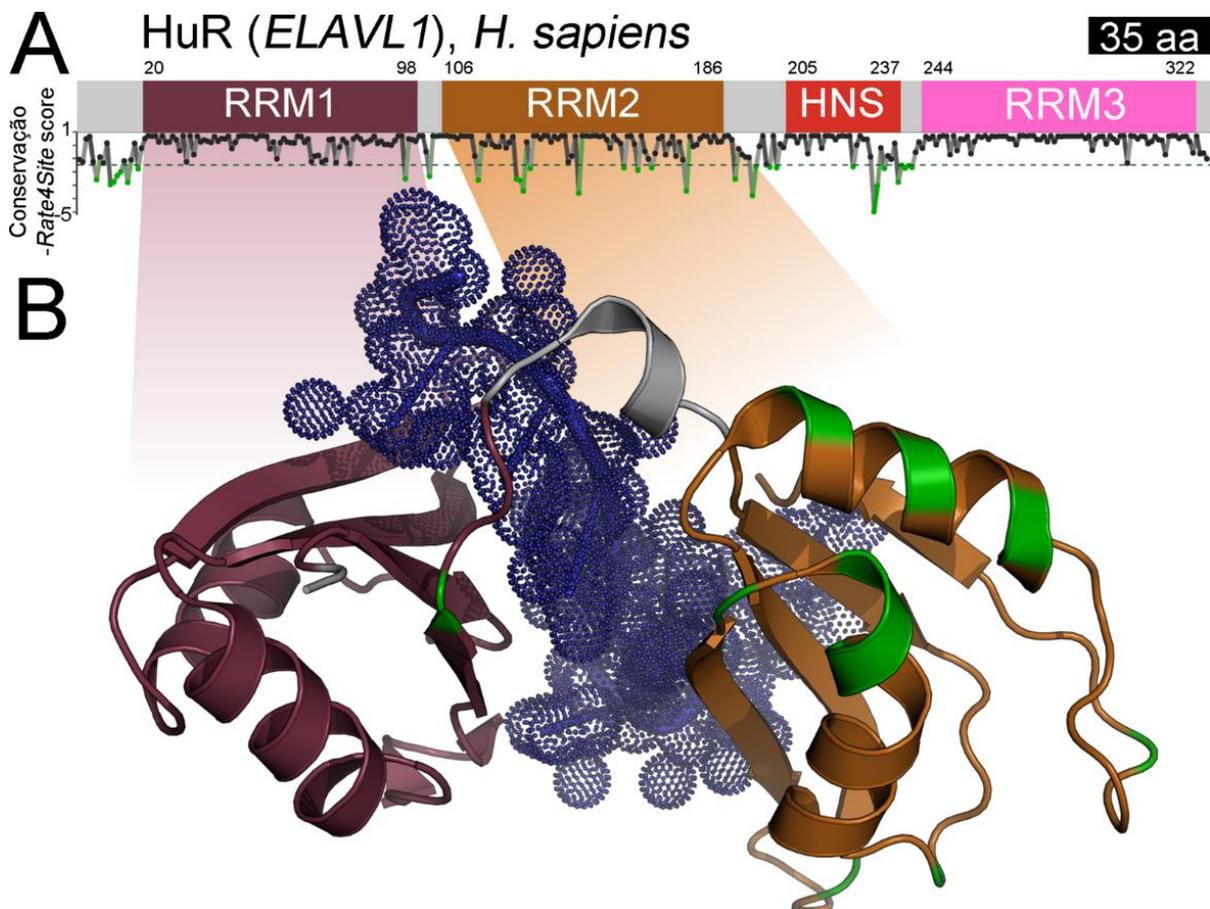
**Figura 7 – Sítios de regulação na porção 3' não traduzida do mRNA das isoformas do gene humano *GLS***  
 Região 3' UTR da isoforma KGA (A) ou GAC (B) com os respectivos sítios de ligação teóricos ARE (quando presentes, barras pretas) e sítios com evidência experimental (PAR-CLIP, barras amarelas, linhas suficientes para evitar empilhamento<sup>76</sup>) de ligação por alguma proteína de ligação a RNA. Setas pretas verticais indicam posições onde tal ligação é da proteína HuR. Busca realizada utilizando a plataforma *AURA2*<sup>76</sup>.

## 1.6 HUR, REGULAÇÃO DE MRNA E CÂNCER

A família das proteínas Hu (*Human antigen*), codificada pelos genes *ELAVn* (*Embryonic lethal abnormal vision-like*) é composta de quatro membros em vertebrados: HuR (eventualmente referenciada com HuA, *ELAVL1*), HuB (Hel-N1, *ELAVL2*), HuC (PLE21, *ELAVL3*) e HuD (*ELAVL4*), sendo a primeira de expressão ubíqua e as demais enriquecidas em neurônios – salvo HuB, também expressa nas gônadas<sup>77</sup>. A homóloga em *Drosophila* (*Embryonic Lethal Abnormal Vision*, *ELAV*) é restrita aos neurônios, sendo de grande importância no desenvolvimento do inseto, de maneira que seu fenótipo de letalidade nomeia o gene<sup>77</sup>. Estruturalmente (Figura 8), as proteínas da família Hu apresentam três distintos domínios de ligação a RNA, chamados de RRM (*RNA recognition motifs*). Não existe informação cristalográfica do terceiro domínio para HuR, o qual apresenta similaridade superior a 90% entre as variantes.

As sequências ARE, tais como o motivo canônico de ligação de HuR, AUUUA<sup>78</sup>, são usualmente localizadas nas regiões 5' ou 3'UTR de mRNAs. Estas sequências propiciam a degradação pelo complexo designado de *exossomo*, via interação das mesmas com elementos como *AUF1* e *KSRP*<sup>79</sup>. HuR, entretanto, impede esta ação, atividade que foi deduzida de sua incapacidade de co-imunoprecipitar com os elementos do exossomo ou ser detectada em espectrometria de massas de seus constituintes, disputando a ligação à regiões ricas em AU com

os elementos *AUF1* e *KSRP*<sup>79</sup>. Outra maneira de HuR definir o destino do mRNA é localizando o mesmo em polissomos, complexos de ribossomos com mRNA que impactam em elevados níveis de expressão dos mRNAs com eles associados<sup>77</sup>. De maneira antagônica, a proteína pode se ligar à região 5'UTR dos mRNAs e inibir a associação de ribossomos e o início de tradução, reduzindo os níveis proteicos gerados pelo mRNA<sup>80</sup>. Ainda no citoplasma, HuR pode definir a localização do mRNA e onde o mesmo será traduzido, e, assim, definir a localização celular do produto proteico. CD47, por exemplo, dispõe de um peptídeo sinal e hélices transmembrana, entretanto é solúvel no citoplasma; quem define a inserção da mesma na membrana é a localização do mRNA próxima ao retículo endoplasmático, que a envia para a membrana externa celular, mecanismo este regulado por HuR<sup>81</sup>.



**Figura 8 – Esquema dos domínios proteicos presentes na proteína HuR e informação estrutural disponível**  
 (A) Representação esquemática de HuR, indicando os domínios de ligação à RNA (*RNA Recognition Motif*, RRM) e o segmento de controle de localização citoplasma-núcleo (*HuR Nuclear-cytoplasmic Shuttling sequence*, HNS). Abaixo do esquema, gráfico indicando a conservação da sequência proteica em vertebrados (*OrthoDB* EOG090B07MJ<sup>82</sup>), onde os vales (pontos mais baixos) indicam regiões menos conservadas, com a linha tracejada indicando o percentil de 90% do escore Rate4Site<sup>31</sup> para a sequência e os pontos verdes para resíduos que ultrapassam este corte. (B) Estrutura cristalográfica dos primeiros domínios RRM da proteína (PDB:5UQE, cadeia A<sup>83</sup>) ligada a um RNA. A representação em *cartoon* contém o RRM1 (em bordô) e o RRM2 (em laranja), com os resíduos mais variáveis indicados na estrutura secundária em verde. Os resíduos mais variáveis não participam da interação com o RNA (em azul, representados o raio de van der Waals para os átomos, sobrepostos ao *cartoon* da fita).

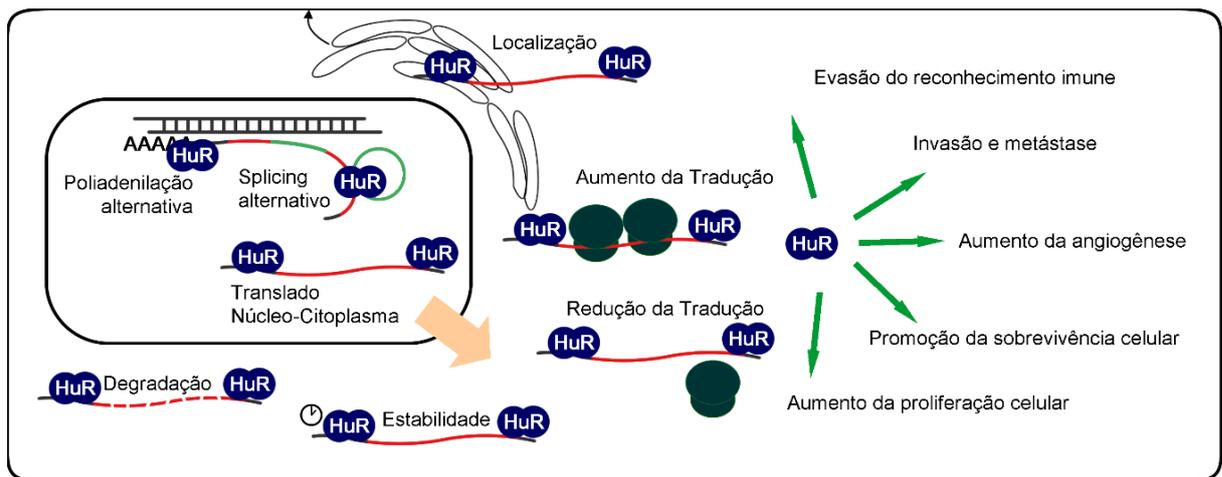
A região de alça entre o segundo e terceiros domínios RRM de HuR é responsável pelo transporte da proteína entre citoplasma e núcleo, função a qual é mediada pela Transportina 2<sup>84</sup>. Tal região, designada de HNS (*HuR nucleocytoplasmic shuttling sequence*), contém, ao mesmo tempo, um sinal de localização nuclear e de exportação ao citoplasma<sup>85</sup>, o que permite sua entrada e saída do núcleo, podendo, assim, mediar o transporte de alguns mRNAs para o citoplasma<sup>86</sup>. A localização celular de HuR é predominantemente nuclear, se tornando mais concentrada no citoplasma de tecidos tumorais, situação esta que favorece a transformação celular<sup>87,88</sup>.

No citoplasma, HuR realiza o processo de estabilização ou controle da tradução do mRNA descritos acima<sup>89</sup>. Quando nuclear, HuR pode influenciar o *splicing* alternativo de distintos alvos<sup>90-93</sup>, ocluindo suas sinalizações para a maquinaria de *splicing* ou mesmo por sua interação direta com o spliceossomo<sup>94,95</sup>. Sua atuação ocorre modulando a inserção ou remoção de exons em diversos transcritos<sup>96,97</sup>. A proteína *First apoptosis signal receptor* (Fas-CD95/APO-1, mediadora de resposta apoptótica), tem o *splicing* alternativo de seu gene (*FAS*) regulado pela ligação de HuR no intron 5, inibindo, assim, a inserção do exon 6<sup>90</sup>. Em outro contexto, o gene *Transformer-2 sex-determining protein* (*TRA2*, responsável por regular o *splicing* alternativo de diversos outros genes) tem seu *splicing* regulado pela ligação de HuR em seu segundo exon<sup>98</sup>.

Reforçando sua relação com *splicing* alternativo, análises ômicas de preferência de ligação em introns indicam uma preferência de ligação de HuR no começo ou o final dos mesmos<sup>99</sup>, ainda no mRNA nascente e impedindo o acesso da maquinaria de poliadenilação. Por este mecanismo, que acontece no próprio pre-mRNA de *ELAVL1*, leva a escolha de sítio mais distal<sup>100</sup> e a inclusão de diversas sequências de desestabilização, reduzindo, em um *feedback* negativo, os níveis de seu mRNA<sup>101</sup>. A avaliação da abundância de sítios condizentes com motivos de ligação de diversas RBPs<sup>102</sup> em regiões próximas a exons que são alvo de *splicing* alternativo revelou que, entre os motivos mais enriquecidos no genoma humano, estão não só os de HuR, assim como os de todas as proteínas de sua família<sup>103</sup>.

HuR pode ser regulada pós-traducionalmente por modificações como fosforilações, ubiquitinações, clivagens, metilações ou neddylações. Os resíduos S88, S100, T118, S158, S221 e S318 estão relacionados com fosforilações que modulam a ligação a distintos mRNAs de forma diversa, sendo que - salvo S88 e S100 - também impactam na localização nuclear ou citoplasmática da proteína<sup>104</sup>. Como esperado, ubiquitinações levam à degradação da proteína, efeito contrário de neddylações, que mantém a proteína reclusa no núcleo sem disponibilidade para degradação<sup>105</sup>.

Com diversas camadas regulatórias e possibilidades de atuação, HuR participa desde a geração do mRNA com a modulação do *splicing* alternativo até o seu destino no citoplasma de uma inúmera quantidade de RNAs relacionados com diversos eventos celulares<sup>97</sup> (Figura 9). Seu *knock-down* implicou na alteração de expressão de aproximadamente 13% do Transcriptoma *RefSeq* em células HeLa<sup>89</sup>.



**Figura 9 – Mecanismos de regulação do mRNA em que HuR atua e suas consequências no desenvolvimento tumoral.**

Adaptado de Silanes et al<sup>87</sup> e Wang et al<sup>106</sup>.

Dada a tamanha variedade de efeitos e atuações de HuR não é de se espantar que a mesma tenha efeitos pró- e anti-tumorais. HuR atua aumentando a proliferação (estabilizando diversas ciclinas e fatores de crescimento, reprimindo a tradução de, por exemplo, p27), levando a evasão da apoptose (com a promoção de fatores de sobrevivência), intensificando a angiogênese (especialmente pela estabilização ou aumento de tradução de *Hypoxia inducible factor 1 alpha*, *HIF1A*, e *Vascular endothelial growth factor*, *VEGF*), reduzindo a imunovigilância (com o aumento de TGF- $\beta$ ) ou agindo sobre alvos relacionados com o processo de EMT (como *Snail Family Transcriptional Repressor 1*, *SNAI1*, e *Matrix metalloproteinase 9*, *MMP9*)<sup>107</sup>.

Como anti-tumoral, a expressão induzível de *ELAVL1* em células mielóides causou proteção quanto ao surgimento de tumores colorretais em camundongos, sem a definição de um alvo específico mas balanceando situações inflamatórias<sup>108</sup>. De maneira geral, HuR já foi associada com alterações do fenótipo bioenergético do câncer de mama, com a sua alta expressão associada a grandes desvios (em relação ao tecido normal) da expressão de um painel de genes associados a vias metabólicas<sup>109</sup>.

Em diversos tipos tumorais, a abundância de HuR – total ou citoplasmática – se relaciona tanto com o prognóstico positivo quanto negativo de pacientes, eventualmente com informações contrastantes entre diversos trabalhos publicados<sup>110</sup>, literatura que resumimos na

Tabela 1 para câncer de mama. Nesta origem tumoral (mama) níveis totais elevados de expressão de *ELAVL1* são correlacionados com um prognóstico positivo em dois estudos (sendo um deles sem significância estatística)<sup>109,111</sup>, mas negativo em um terceiro<sup>112</sup>. A localização citoplasmática da proteína por sua vez, foi relacionada a um prognóstico negativo por quatro trabalhos de três grupos distintos<sup>113-116</sup> e se mostrou indiferente por apenas um<sup>117</sup>. Os dois trabalhos que caracterizaram sua abundância nuclear não observaram correlação preditiva com o prognóstico dos pacientes<sup>115,117</sup>. Em suma, tais resultados evidenciam uma inconsistência quanto ao resultado preditivo da abundância diferencial de HuR, de maneira que pacientes com presença citoplasmática elevada nos tumores de mama podem apresentar risco proporcional de morte aumentado em 6,69 vezes, enquanto análise de seus níveis totais proteicos indicam que altos níveis se relacionam ao um risco proporcional reduzido em 3,23 vezes.

**Tabela 1 – Sumário da literatura acerca do prognóstico clínico de pacientes de câncer de mama de acordo com o nível de HuR**

Fonte	Total			Prognóstico (alto)	Comentários
	Pacientes (baixo/alto)	Cox HR [CI]	<i>p-value</i>		
Ortega et al <sup>109</sup>	89(37/52)	0.31 [0.13-0.71]	0.006	Positivo	Recidiva, proteína - <i>western blot</i>
Yuan et al <sup>111</sup>	109 (??)	< 1 [?/?]	0.28	Positivo†	Recidiva, RNA - qPCR
Giaginis et al <sup>112</sup>	191 (67/54)	3.714 [2.443–5.140]	0.0004	Negativo	Recidiva, proteína - IHC
Fonte	Citoplasma			Prognóstico (alto)	Comentários
	Pacientes (baixo/alto)	Cox HR [CI]	<i>p-value</i>		
Denkert et al <sup>117</sup>	208 (145/63)	--	0.25	Indiferente	Sobrevida, proteína - IHC
Heinonen et al <sup>113</sup>	133 (95/38)	2.07 [1.05-4.07]	0.035	Negativo	Metástase, proteína - IHC
Heinonen et al <sup>114</sup>	525 (207/318)	1.74 [1.05-2.88]	0.0311	Negativo	Sobrevida, Não-BRCA, proteína - IHC
Zhu et al <sup>115</sup>	82 (44/38)	3.48 [1.53–4.40]	0.003	Negativo	Recidiva, proteína - IHC
Wang et al <sup>118</sup>	139 (99/40)	6.69 [2.47–18.13]	<0.0001	Negativo	Recidiva, proteína - IHC
Fonte	Núcleo			Prognóstico (alto)	Comentários
	Pacientes (baixo/alto)	Cox HR [CI]	<i>p-value</i>		
Denkert et al <sup>117</sup>	208 (80/128)	--	0.83	Indiferente	Sobrevida, proteína - IHC
Zhu et al <sup>115</sup>	82 (19/63)	1.74 [0.69–4.41]	0.240	Indiferente	Recidiva, proteína - IHC

NOTA: Cox HR: Proporção de risco de Cox; CI: intervalo de confiança; Recidiva: avaliou a sobrevida livre de recidiva; Sobrevida: avaliou a sobrevida global (irrespective à causa de morte); Não-BRCA: pacientes sem mutação em BRCA1/2; IHC: Imunohistoquímica. †Os autores propõem um prognóstico positivo, entretanto sem um *p-value* significativo associado.

Assim, tendo em vista a regulação da estabilidade do mRNA de GAC por HuR na situação de acidose renal e a importância de HuR dentro do contexto tumoral, hipotetizou-se que HuR pudesse ser um importante regulador da glutaminase em câncer. Sua múltipla atuação regulatória – e os eventos já descritos para seu parálogo, HuD – abrem a possibilidade não só de uma atuação sobre a 3'UTR dos transcritos de *GLS* como também uma possível implicação na regulação do *splicing* alternativo do gene. Por fim, o estudo do eixo HuR-GLS e seu impacto nos fenótipos de proliferação, invasão e o metabolismo de glutamina de células tumorais pode trazer subsídios para a proposição de terapias voltadas a inibição destas proteínas no combate a doença.

## 2 OBJETIVOS

O objetivo geral proposto foi compreender a ação regulatória da proteína de ligação à RNA HuR sobre os produtos do gene *GLS*, dentro do contexto tumoral, particularmente no *splicing* do gene *GLS* e a estabilidade de seus transcritos mediada por 3'UTR e os consequentes impactos fenotípicos desta regulação sobre o metabolismo, proliferação, migração e invasão.

### 2.1 OBJETIVOS ESPECÍFICOS

- Mensurar utilizando dados publicamente disponíveis de transcriptômica e genômica a abundância da expressão de *ELAVL1* em tecidos normal e tumoral e o efeito de seus níveis de expressão sobre o prognóstico dos pacientes. Identificar tumores onde HuR possa ter um papel importante para prognóstico da doença.
- Avaliar a abundância e localização núcleo/citoplasma de HuR em um painel de células tumorais.
- Modular os níveis de HuR em células de mamífero em cultura e avaliar seu impacto no nível de mRNA e nível proteico das isoformas GAC e KGA.
- Avaliar o impacto de HuR sobre o *splicing* alternativo de *GLS* e estabilidade do mRNA de GAC e KGA.
- Utilizando dados públicos ômicos de interação de HuR com regiões intrônicas, assim como informações de conservação evolutiva de sequência, identificar potenciais sítios de interação de HuR no intron 14 (envolvido no *splicing* que define as isoformas) de *GLS*.
- Avaliar o impacto do *knock down* de *ELAVL1* na atividade de glutaminase celular, metabolismo em geral, assim como na proliferação, migração e invasão de células tumorais.

### 3 MATERIAIS E MÉTODOS

#### 3.1 CLONAGEM E MODIFICAÇÃO DE VETORES

Todas as *Polymerase Chain Reaction* (PCR) realizadas empregaram a enzima *Phusion® High-Fidelity DNA Polymerase* (New England Biolabs), seguindo instruções do fabricante. Igualmente, a mesma foi utilizada para remoção de extremidades coesivas<sup>119</sup>. Para a adenilação de produtos de PCR, foi realizada a incubação com *Taq DNA Polimerase* (Invitrogen) nas condições indicadas pelo fabricante. A restrição de plasmídeos ou produtos de PCR foi realizada com enzimas de restrição comercialmente disponíveis (New England Biolabs) seguindo instruções do fabricante. A purificação dos produtos amplificados ou vetores restringidos com enzimas foi realizada com o *Wizard® SV Gel and PCR Clean-Up System* (Promega), seguindo instruções do fabricante. A ligação direta de oligonucleotídeos anelados foi realizada como descrito previamente<sup>120</sup>. Os oligonucleotídeos utilizados foram todos adquiridos com o fornecedor Exxtend, dessalinizados ou purificados por interação do grupo dimetoxitritil antes de sua remoção (*Oligonucleotide Purification Cartridge*). As modificações em vetores de expressão podem ser observadas no resumo da Figura 10 e no Quadro 1 e os oligonucleotídeos utilizados no Quadro 2.

**Quadro 1 – Resumo das construções para expressão ou silenciamento gênico utilizadas neste trabalho**

Vetor base	cDNAs/shRNA nas construções
pcDNA3.1-V5-His*	HuR.WT* KGA.WT*
pQC mKOrange IX*	pQC MCS IRES Puro (vetor vazio) pQC V5-MCS IRES Puro (vetor vazio) pQC V5-mKO2-MCS IRES Puro
pQC V5-MCS IRES Puro	HuR.WT HuR.ΔHNS
pQC V5-mKO2-MCS IRES Puro	HuR.WT HuR.ΔHNS
pQCXI Neo DsRed-LC3-GFP*	pQC MCS IRES G418 (vetor vazio) pQC V5-MCS IRES G418 (vetor vazio)
pLKO.1-TRC-Puro	shGFP shLuc shHuR shHuRII
tet-pLKO.1-puro	shHuR shHuRII
pET30b(+), mod*	mKO2-HuR.WT

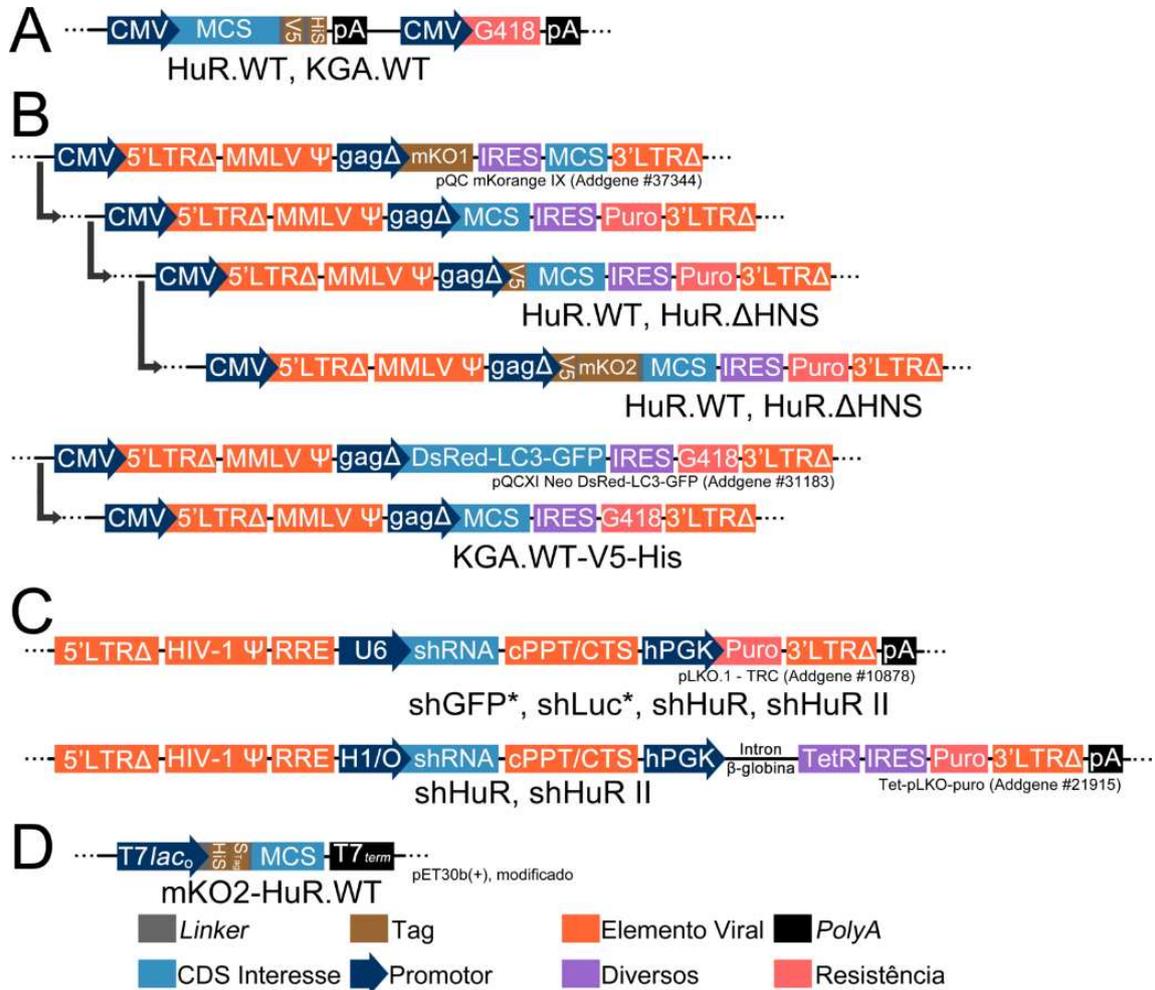
NOTA: \*Construções previamente disponíveis

Para a criação dos vetores pQC-MCS-IRES-PURO, o vetor pQC mKorange IX, gentilmente cedido pelo Dr. Connie Cepko (Addgene # 37344), foi modificado com a inserção do fragmento de resistência à puromicina a jusante de sua *Internal Ribosomal Entry Site* (IRES). A mesma foi aberta com EcoRV e fechada com fragmentos contendo a sequência codificante da puromicina N-acetiltransferase, removida com BamHI e KpnI e com remoção das extremidades protudentes a partir do vetor pLKO.1-puro. A região múltipla de clonagem (MCS) foi estabelecida removendo o inserto mKO2 a montante do IRES com EcoRI e AgeI e adicionando sequência de ligação sintética produzida a partir de oligos anelados - criando sítios únicos para BamHI, XbaI, AgeI, PmeI, NotI e EcoRI, além da possibilidade de clonagem TA com restrição dupla por XcmI. À essa região, na porção N-terminal, foi adicionada uma tag V5 com a restrição do primeiro sítio, BamHI, definindo um início de transcrição de Kozak e o epítopo viral, além de manter o sítio de BamHI em sua região 3'. Tal sítio, quando foi o caso, foi utilizado para a inserção de sequência codificante para mKO2, amplificada por PCR de vetor pGEM-T-mKO2 (Gentilmente cedido pelo Dr. Sílvio Roberto Consonni) e clivada com BglII e BamHI. HuR foi amplificada a partir de vetor pCDNA3.1-HuR.wt-V5-His previamente disponível no grupo. Para a remoção de sua região HNS, foram amplificados os fragmentos a jusante e a montante da mesma, unidos por sítio sintético de HindIII e inseridos no vetor da mesma maneira. A sequência removida compreendeu os aminoácidos 205-237 (RRFGGPVHHQAQRFRFSPMGVDHMSGVNSVNP) de HuR e a inclusão do sítio de HindIII substituí a mesma por uma lisina e uma leucina (KL).

O vetor pQC-MCS-G418 foi produzido substituindo no vetor pQCXI Neo DsRed-LC3-GFP, gentilmente cedido pelo Dr. David Sabatini (Addgene #31183), a região compreendida entre o início de seu promotor CMV e o início da entrada interna de ribossomo (IRES) pela mesma região do vetor, aqui produzido, pQC-MCS-IRES-Puro, utilizando as enzimas BglII e EcoRI em ambos os vetores. KGA-V5-6xHis foi extraído do vetor pCDNA3.1-KGA.wt-V5-His utilizando as enzimas de restrição BamHI e PmeI, e inserido com as mesmas no vetor pQC-MCS-G418 construído.

Os vetores repórteres para estudo da estabilidade da 3'UTR do mRNA das isoformas KGA e GAC foram produzidos utilizando o vetor psiCHECK2 modificado, gentilmente cedido pelo Dr. Marcio Chaim Bajgelman. Para a clonagem da 3'UTR de GAC, o vetor e foi clivado com XhoI e NotI e a região 3'UTR amplificada de DNA genômico extraído da linhagem de câncer de mama SKBR3 (utilizando *Wizard SV Genomic DNA Purification System*, seguindo instruções do fabricante, Promega), sendo clivada com as mesmas enzimas adicionadas aos oligonucleotídeos. Para a clonagem da 3'UTR de KGA, o vetor foi clivado

com SacI, com a ponta protundente removida, e NotI, sendo o inserto amplificado do mesmo DNA genômico com oligonucleotídeos contendo sítios de PmeI e NotI.



**Figura 10 – Representação simplificada dos vetores de expressão ou silenciamento utilizados**

(A) Vetores pcDNA3.1-V5-His-Topo (Invitrogen) previamente disponíveis no laboratório com a inserção das seqüências codificantes de HuR, KGA ou GAC. (B) Vetores de expressão retroviral (Gammaretrovirus), modificados para a expressão de *ELAVL1* selvagem ou deletada para a região HNS, que conferem resistência a puromicina, ou modificados para a expressão de KGA selvagem, conferindo resistência a geneticina. (C) Vetores de silenciamento retroviral (Lentivirus) para *short hairpin* controle ou almejando *ELAVL1*. (D) Vetor de expressão procariótica pET30b utilizado para a produção de proteína recombinante. CMV: Promotor do citomegalovírus humano. MCS: sítio múltiplo de clonagem. V5: pequeno epítipo do vírus símio Paramyxovirus. His: Sequência de 6 histinas. G418: Resistência ao agente de seleção geneticina. pA: região para finalização do mRNA e poliadenilação. 5'LTRΔ: região repetitiva 5' truncada. MMLV Ψ: Sequência *psi* de empacotamento do vírus da leucemia murina de Moloney. gagΔ: poliproteína gag truncada.mKO1/2: proteína fluorescente monomérica kusabira-orange. IRES: sítio de entrada interna de ribossomo. 3'LTRΔ: região repetitiva 3' truncada. HIV-1 Ψ: Sequência *psi* de empacotamento do vírus da síndrome de imunodeficiência humana do tipo 1. RRE: elemento responsivo à *Rev*. U6: promotor da polimerase III humana do tipo U6. shRNA: Sequência de *short hairpin* RNA. cPPT: Trato central de polipurina. CTS: Sequência central de terminação. hPGK: Promotor da fosfogliceroquinase humana. H1/O: promotor da RNA polimerase III humana com seqüência de ligação para o elemento responsivo a tetraciclina. TetR: repressor da tetraciclina.

**Quadro 2 – Oligonucleotídeos utilizados para modificação dos vetores de expressão**

Finalidade	Oligo	Sequência 5' – 3'
Inserção de MCS em pQC	MCS_Fow	CCGGGGATCCGGTTCTAGAGCCACCGGTTGTATG GAGTTTAAACTCCAAAGCATCGATGGTAGCGGCC GCGGTGAATTCTGATT
	MCS_Rev	AATTAATCAGAATTCACCGCGCCGCTACCATCG ATGCTTTGGAGTTTAAACTCCATACAACCGGTGG CTCTAGAACCGGATCC
Inserção de V5 em pQC-MCS	V5_Fow	GATCACCATGGGTAAGCCTATCCCTAACCTCTCC TCGGTCTCGATTCTACGG
	V5_Rev	GATCCCGTAGAATCGAGACCGAGGAGAGGGTTAG GGATAGGCTTACCCATGGT
Inserção de mKO2	mKO2.BglII-Kozak.Forward	AAGAGATCTACCATGGTGAGCGTGATCAAG
	mKO2.BamHI.Reverse	TCGGATCCCAGCCACCTCCGCCCG
Amplificação de <i>ELAVL1</i> para vetor pQC	HuR_AgeI_Fow	AGTACCGGTATGTCTAATGGTTATGAAGACCACA TGGC
	HuR_EcoRI_Rev	AGTGAATCTTATTGTGGACTGTTGGTTTTGA AGGA
Amplificação da porção pré-HNS de <i>ELAVL1</i>	HuR.KpnI.Forward	CCGGTACCATGTCTAATGGTTATGAAGACCAC
	HuR.PreHNS_HindIII_Rev	GAAAAGCTTCGCTGGCGAGTGGTACAG
Amplificação da porção pós-HNS de <i>ELAVL1</i>	HuR.XhoI.Reverse	TAGACTCGAGTTATTGTGGACTGTTGGTTTTG AAG
	HuR.PosHNS_HindIII_Fow	ATAAAGCTTGAAACGCCTCCTCCGGC
Amplificação do Intron 14 de <i>GLS</i> à partir do genoma	GAC.Exon14End_Fow	ACTATGATAATTTGAGACACTTT
	GAC.Exon15Start_Rev	TCCATTCTATATACTACAGTTGT
Amplificação do Intron 14 para restrição	Intron14.XbaI.Fow	TTCTAGAGGTAAGCAAAATCTTATTTAGATAAGT ATATA
	GLS.It14Sub.BamHI.Rev	TATGGATCCTAGTTGTTCAGCAACATACATAAT
Amplificação da região 3'UTR do mRNA de GAC	GAC3UTR.PmeI.Fow	TCGCCTCGAGAAATGGGTTCTAGTTTCAGAATG
	GAC3UTR.NotI.Rev	AGAATGCGGCCGCCAGCTTAAATAAATTTATTG TGC
Amplificação da região 3'UTR do mRNA de KGA	KGA3UTR.PmeI.Fow	TGTTTAAACGGTCTCAAATCCCAAGATTTAAATCA C
	KGA3UTR.NotI.Rev	GGCGGCCGCACATATATTTATTAATGCTGTAAAA AGCAAC
sgRNA para KGA	sgRNA_KGA_Stop_Fow	CACCGATCTTGATGGATTGTTGTAA
	sgRNA_KGA_Stop_Rev	AAACTTACAACAATCCATCAAGATC
Braço de homologia 5'	KGA_Homology5_SacI_Fow	ATCGAGCTCCAGGCTTCACAGTTTGTAAAGTGTA AGG
	KGA_Homology5_BamHI_Rev	ATCGGATCCCAACAATCCATCAAGATTCTTATGG ACGGT
Braço de homologia 3'	KGA_Homology3_BamHI_Fow	ATCGGATCCTAATGGTCTCAAATCCCAAGATTTA AATCACTTACCT
	KGA_Homology3_SalI_Rev	ATCGTCGACTTTCAGAATAACTTTGGTAGCTTAGA GCATTTCTG
P2A Porcine teschovirus-1	P2A_BamHI_Fow	GATCAGGAAGCGGAGCTACTAAGTTCAGCCTGCT GAAGCAGGCTGGAGACGTGGAGGAGAACCCTGG ACCTG
	P2A_BamHI_Rev	GATCCAGGTCCAGGGTTCCTCCACGTCTCCAGC CTGCTTCAGCAGGCTGAAGTTAGTAGCTCCGCTTC CT
Amplificação de BleoR	BleoR_BglII_Fow	ACAAGATCTATGGATGCCAAGTTGACCAGTGC
	BleoR_BamHI_Rev	GGCGGATCCGTCCTGCTCCTCGGCCA

Já a produção do vetor repórter para *splicing* alternativo RG6-Intron14 foi realizada utilizando o vetor homônimo, gentilmente cedido pelo Dr. Thomas A. Cooper<sup>121</sup>, com a clivagem do sítio DamI-DcmI bloqueado de XbaI e BamHI (a partir de minipreparação plasmidial de linhagem de *E. coli* defectiva para este sistema de metilação, JM110), que remove o primeiro intron do sistema. O intron 14 humano foi obtido de DNA genômico extraído da

linhagem de câncer de mama SKBR3. Fragmento amplificado inicialmente com oligonucleotídeos ancorando nas porções exônicas foi ligado em vetor pGEM-T easy (Promega), seguido posteriormente da amplificação do fragmento a ser inserido propriamente dito e clivagem com as mesmas enzimas de restrição (XbaI e BamHI).

Para a estratégia baseada em CRISPR/Cas9 de introdução de cauda fluorescente no genoma, o plasmídeo pX330<sup>122</sup> (gentilmente cedido pelo Dr. José Xavier Neto, originalmente pX330-U6-Chimeric\_BB-CBh-hSpCas9, Addgene#42230, gentilmente cedido pelo Dr. Feng Zhang) foi modificado com o intuito de adicionar resistência ao agente de seleção puromicina. Para tanto, o cassete pSV40-PuroR-PolyA foi removido do vetor pBABE.puro (gentilmente cedido pelo Dr. Marcio Chaim Bajgelman) utilizando os sítios de XbaI e BamHI, com protundências removidas e inserto fosforilado, e inserido no sítio de NotI de px330 logo após a cauda PolyA da hSpCas9, após a remoção das protundências. Ao mesmo vetor, cortado com BbsI, foi inserido o *single guide* RNA (sgRNA, Quadro 2) para KGA, anelado a partir de oligonucleotídeos sintéticos fosforilados.

A produção do fragmento doador foi realizada sobre o vetor pUC19, sem elementos eucarióticos de expressão. Clonou-se, sequencialmente, o braço de homologia 5' (SacI-BamHI) seguido do braço de homologia 3' (BamHI-SalI), ambos obtidos de DNA genômico extraído da célula SKBR3. Na cicatriz criada com BamHI, foi inserida a proteína fluorescente mKO2 (amplificada e cortada com BglII-BamHI), a região codificante para o peptídeo de auto-clivagem P2A do Teschovírus-1 suíno (P2A<sup>123,124</sup>), anelada e fosforilada – mantendo sítio de BamHI em sua 3' –, e a sequência codificante do gene BleoR de *Streptoalloteichus hindustanus* (conferindo resistência ao agente de seleção Zeocina), amplificada do vetor pFRT-lacZeo (Invitrogen) e cortada com BglII-BamHI.

Os vetores de silenciamento foram construídos utilizando os sistemas pLKO.puro-TRC (gentilmente cedido pelo Dr. David Root, Addgene #10878)<sup>125</sup> ou tet-pLKO-puro<sup>126</sup> (gentilmente cedido pelo Dr. Dmitri Wiederschain, Addgene #21915). Para ambos, foi realizado o corte com AgeI e EcoRI, seguido de defosforilação, e os oligonucleotídeos anelados (Quadro 3) e fosforilados foram ligados.

O vetor de expressão pET30b+-TEV (gentilmente cedido pelo Dr. Daniel Maragno Trindade), foi clivado com HindIII e XhoI e o cassete de expressão mKO2-HuR.WT amplificado por PCR e no mesmo clonado.

**Quadro 3 – Oligonucleotídeos utilizados para o silenciamento utilizando sistemas pLKO/tet-pLKO**

Transcrito alvo	Oligo	Sequência 5' – 3'
HuR (TRCN0000276129)	shHuR_CDS_Fow	CCGGCGAGCTCAGAGGTGATCAAAGCTCGAGCTTTGATCACCTCT GAGCTCGTTTTTG
	shHuR_CDS_Rev	AATTCAAAAACGAGCTCAGAGGTGATCAAAGCTCGAGCTTTGATC ACCTCTGAGCTCG
HuR II (TRCN0000276186)	shHuR_3UTR_Fow	CCGGTTGTTAGTGTACAACCTCATTCTCGAGAAATGAGTTGTACAC TAACAATTTTTG
	shHuR_3UTR_Rev	AATTCAAAAATTGTTAGTGTACAACCTCATTCTCGAGAAATGAGTT GTACACTAACAA
GFP (Controle)	GFP_Fow	CCGGCAAGCTGACCCCTGAAGTTCATCTCGAGATGAACTTCAGGGT CAGCTGTTTTTG
	GFP_Rev	AATTCAAAAACAAGCTGACCCCTGAAGTTCATCTCGAGATGAACTT CAGGGTCAGCTTG
GFP (Luc)	Luc_Fow	CCGGCTTACGCTGAGTACTTCGACCTGACCCATCGAAGTACTCAGC GTAAGTTTTG
	Luc_Rev	AATTCAAAAACTTACGCTGAGTACTTCGATGGGTCAGGTCGAAGT ACTCAGCGTAAG

NOTA: TRCN# Código de acesso do *The RNAi Consortium shRNA Library*

### 3.2 ANÁLISE DO CONJUNTO DE DADOS DISPONÍVEL NO *THE CANCER GENOME ATLAS*

#### 3.2.1 *Download*, organização, normalização e representação dos dados

Utilizando o portal de dados do *Genomic Data Commons* do *National Cancer Institute* (<https://portal.gdc.cancer.gov>) foram gerados os manifestos contendo a identificação única e o *hash* MD5 dos dados clínicos (*clinical xml*), contagem bruta de fragmentos em cada gene (*HTSeq counts*), fragmentos por milhar de transcrito por milhão de biblioteca (FPKM, e sua normalização por quartil superior, FPKM\_UQ) e a quantificação de leituras por milhar de pares de base do exon por milhão de biblioteca (RPKM). A comunicação foi feita através de API (*Application Programming Interface*) disponibilizada, utilizando o software R<sup>127</sup> versão 3.4.2, e as bibliotecas *httr*<sup>128</sup> e *jsonlite*<sup>129</sup>. A leitura das tabelas clínicas disponibilizadas em XML foi realizada utilizando o pacote *XML*<sup>130</sup>, para a busca dos parâmetros clinicopatológicos, da última atualização acerca do paciente (quanto a sobrevida) e a todas as atualizações quanto a primeira ocorrência de recidiva. Todas as entradas foram renomeadas em função do código único do paciente do TCGA e os valores de expressão de exon também foram normalizados pelo método do quartil superior<sup>131</sup>. Os dados aqui apresentados foram atualizados no dia 27 de novembro de 2017. Todos os arquivos foram verificados quanto a sua integridade e baixados novamente quando observadas exceções. Os tipos e origens tumorais avaliadas estão disponíveis no Apêndice 1.

Inicialmente, ilustrações com preto-e-branco foram produzidas, ilustrando o corpo humano e seus órgãos. Tais esquemas foram convertidos com a função *raster-to-vector* do programa de *Geographic Information System* QGIS<sup>132</sup> versão 2.18. A conversão para o

ambiente R do *ESRI Shapefile* gerado foi realizada com os pacotes *maptools*<sup>133</sup> e a sua simplificação em polígonos pelo pacote *gpclib*<sup>134</sup>. As medianas dos parâmetros avaliados (ou suas razões) foram então transformadas em escala de cor utilizada para colorir os respectivos órgãos, utilizando o pacote *ggplot2*<sup>135</sup>. Uma relação tecido-tipo tumoral-órgão está disponível no Apêndice 1. As representações RDI (*Raw (data), Description and Inference*) utilizadas nesta porção e em todo o trabalho foram produzidas com o pacote *yarr*<sup>136</sup>.

Para a comparação entre as anotações distintas, foi realizada uma harmonização dos resultados oriundos de cada uma das análises: os dados do TCGA com resolução de gene utilizam a versão GRC38 anotada pelo GENCODE versão 22, já quando em resolução de isoforma utilizam a sequência GRC37 (hg19) anotada pelo UCSC e, por fim, os dados experimentais aqui avaliados compartilham a primeira versão do genoma (GRC38) mas com uma anotação brevemente mais recente (GENCODE versão 25). A harmonização *in house* com a utilização de uma mesma anotação para todos as análises não foi possível devido a indisponibilidade pública dos arquivos com os fragmentos efetivamente lidos do TCGA e a indisponibilidade técnica do reprocessamento dos cerca de 96 terabytes de sequências curtas obtidas das 11096 amostras de RNA sequenciadas pelo projeto. Assim, o esforço de compatibilização das informações realizado era imprescindível para a obtenção de quaisquer informações relevantes.

### 3.2.2 Separação dos pacientes por níveis de expressão de *ELAVL1*, *GLS* e os níveis dos exons diferenciais do mRNA de GAC e KGA em tumores

Foi implementada uma rotina de minimização de *p-value*, como previamente proposto<sup>137,138</sup>, utilizando modelos de Cox como implementados nos pacotes *survival*<sup>139</sup> e *survminer*<sup>140</sup>. Descritivamente, os dados de expressão das amostras de tecido de um determinado tipo tumoral foram transformados em uma escala de *z-score* (cada valor foi subtraído da média de expressão daquele gene naquele tumor específico e o valor foi dividido pelo desvio padrão nas mesmas condições), que foi utilizada para ordenar numericamente os os pacientes da menor para a maior variação em relação à média. Considerando grupos formados com, ao menos, 4% dos dados ou 10 pacientes (o maior valor entre as duas opções), os mesmos foram avaliados com iterações do seguinte algoritmo:

1. Classificar os pacientes  $n$  primeiros pacientes como baixa expressão (inclusive)
2. Comparar o risco proporcional de Cox quanto a recidiva tumoral (por metástase distante ou loco regional), calcular o  $p$ -value para teste de *logrank* e armazenar tais informações
3. Adicionar 1 ao valor de  $n$ . Caso não infringida a regra dos 4% ou 10 pacientes remanescentes no grupo de alta expressão, retornar ao passo 1.

Com posse de todas as separações possíveis dos pacientes e risco relativo em cada situação, utilizou-se como critério para a distinção em dois grupos o ponto em que o  $p$ -value do teste *logrank* assumiu o menor valor. Estas separações, então, foram utilizadas para o cálculo da associação entre parâmetros clinicopatológicos e a categorização, utilizando testes de contingência  $\chi^2$ . A organização dos dados, quando necessária, se fez com o pacote *plyr*<sup>141</sup>. O controle de falso-positivo foi realizado com *false discovery ratio*<sup>142</sup> aplicado nos melhores  $p$ -value obtidos para todos os tumores. A título de simplificação das representações, os pacientes não considerados segundo o critério de tamanho mínimo de grupos foram ocultos no eixo x. Quando mencionada, foi aplicada a correção de  $p$ -value, para buscas de minimização, proposta por Altman *et al*<sup>143</sup>:

$$p_{\text{corrigido}} = -3,13 * p_{\text{mínimo}} * (1 + 1,65 * \log_e p_{\text{mínimo}}),$$

para 5% de erro do tipo I.

Pacientes com informação duvidosa ou incongruências no preenchimento da informação de sobrevida, recidiva ou metástase distante foram, em cada situação, desconsiderados. Pacientes com falecimento por causa não-relacionada e saudáveis desde o primeiro aparecimento da patologia foram considerados como casos censurados no dia de falecimento para as avaliações de recidiva e metástase distante. A produção da curva ROC foi realizada utilizando o pacote *pROC*<sup>144</sup>.

### 3.3 ANÁLISE DE SEQUENCIAMENTOS DE RNA PREVIAMENTE DISPONÍVEIS

#### 3.3.1 Busca e obtenção de sequências

Os dados da literatura foram obtidos no portal *Gene Expression Omnibus*<sup>145</sup>, após busca de experimentos em células humanas com imunoprecipitação de HuR ou silenciamento de seu gene. Os códigos de acesso se encontram no Quadro 4. O *download* dos arquivos foi realizado utilizando a ferramenta *AsperaConnect* a partir do servidor *European Nucleotide*

*Archive* (ENA)<sup>146</sup>. Os dados de RIP-Seq derivam de um experimento onde ocorreu imunoprecipitação utilizando anti-T7Tag em células expressando construção ectópica de HuR ou apenas vetor controle, seguido de extração de RNA e sequenciamento. Para obtenção dos dados de PAR-CLIP-Seq (*Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation*), a células foram suplementadas com um análogo de base (4SU ou 6SG) que realiza *crosslink* reversível com a proteína interatora após exposição à luz ultravioleta; após a imunoprecipitação de HuR, a ligação foi desfeita, deixando uma mutação na base que mediou a interação, o que, desta maneira, revela o sítio de interação.

**Quadro 4 – Conjuntos de dados externos utilizados e suas características**

SRR	Type	Fonte	Linhagem	Condição
SRR504455	RIP-Seq	ENCODE <sup>147</sup>	GM12878	IP:T7Tag (Controle)
SRR504456			GM12878	IP:T7Tag (Controle)
SRR504447			GM12878	IP:HuR
SRR504448			GM12878	IP:HuR
SRR504459			K562	IP:T7Tag (Controle)
SRR504460			K562	IP:T7Tag (Controle)
SRR504453			K562	IP:HuR
SRR504454			K562	IP:HuR
SRR248532			PAR-CLIP-Seq	Mukherjee <i>et al</i> <sup>97</sup>
SRR189779	Kishore <i>et al</i> <sup>148</sup>	HEK293T		MNase, 4SU; IP:HuR
SRR189780		HEK293T		MNase, 4SU; IP:HuR
SRR189781		HEK293T		T1 RNase, 4SU; IP:HuR
SRR309285		Lebedeva <i>et al</i> <sup>96</sup>		HeLa
SRR309286	HeLa			T1 RNase, 4SU - SILAC; IP:HuR
SRR309287	HeLa			T1 RNase, 6SG - SILAC; IP:HuR
SRR309284	HeLa			siRNA HuR 5 dias
SRR309282	HeLa			Mock 5 dias
	RNA-Seq			

NOTA: SRR#: Número de acesso do *Sequence Read Archive*; RIP-Seq: RNA immunoprecipitation followed by sequencing; GM12878: Linhagem celular linfoblástica transformada com vírus *Epstein-Barr*; K562: Linhagem celular de leucemia mieloide aguda; HEK293T: Linhagem celular de Rim embrionário humano transformada com o antígeno T do vírus SV40; HeLa: Linhagem celular de adenocarcinoma de cérvix uterino.

Os dados obtidos pelo ENCODE foram gerados com o objetivo inicial de, em uma abordagem ampla e focada em diversos elementos regulatórios, compreender a regulação da expressão gênica de forma genômica. Assim, seus resultados transformaram-se, basicamente, em pistas distintas para o ENSEMBL<sup>149</sup>. Mukherjee *et al*<sup>97</sup> se propuseram a avaliar a ligação de HuR sobre todo o transcriptoma, através de PAR-CLIP-Seq e o impacto gerado pelo silenciamento do gene na expressão gênica, utilizando *microarrays*. Com tais resultados, propuseram sítios preferenciais de ligação de HuR em introns e exons, elementos regulados e descreveram a característica observada dos sítios de forma contextual. Lebedeva *et al*<sup>96</sup> tinham

os mesmos objetivos, publicando, inclusive, na mesma edição da revista *Molecular Cell*. Metodologicamente, variaram com o uso de RNA-Seq para a quantificação da expressão após o silenciamento e quantificaram proteínas diferencialmente expressas utilizando proteômica. Kishore *et al*<sup>148</sup> estabeleceram uma busca técnica entre a utilização de T1 RNase ou MNase como ribonuclease para a digestão das regiões que sofreram *crosslink*, sendo um trabalho de caráter técnico e sem conclusões biológicas acerca.

### 3.3.2 Controle de qualidade dos fragmentos sequenciados e remoção das bases de baixa qualidade

Todos os conjuntos passaram por um processo homogeneizado de controle de qualidade. Inicialmente, a distribuição da qualidade dos fragmentos foi avaliada com o *software* *FastQC*<sup>150</sup>. Posteriormente, últimos nucleotídeos com qualidade inferior a 20 – na escala *PHRED*<sup>151,152</sup> – foram removidos sequencialmente até o primeiro nucleotídeo com qualidade superior a 20 ou até a sequência reduzir-se a um tamanho de 26 pares de base (o que implicava em seu descarte). Caso a sequência, após este processo, apresentasse qualidade média inferior a 30, também era descartada. Tais procedimentos eram aplicados nos dois pares para conjuntos *paired end* e, caso um dos mesmos não atendesse as demandas, os dois eram removidos. Tais procedimentos foram realizados com o auxílio do *software* *Skewer*<sup>153</sup> e seguem as sugestões previamente publicadas<sup>154,155</sup>. Após o processo, o conjunto era novamente avaliado com o *software* *FastQC*<sup>150</sup> para a confirmação da melhoria em termos de qualidade.

### 3.3.3 Alinhamento contra o genoma humano

Para alinhar os fragmentos contra a sequência do genoma humano, o *software* *STAR*<sup>156</sup> foi utilizado. O mesmo foi executado uma primeira vez utilizando o genoma humano com a montagem primária do GRCh38 (contendo todos os cromossomos e *scaffolds*, mas sem os *patches* e haplótipos para evitar alinhamento múltiplo de fragmentos únicos) utilizando a anotação completa de transcritos do GENCODE versão 25<sup>157,158</sup>. Junções exon-exon hipotéticas foram anotadas para cada sequenciamento. O arquivo foi alinhado uma segunda vez, incluindo agora as anotações adicionadas na primeira etapa. Assim, obteve-se uma precisão maior no posicionamento de fragmentos que sobrepujam a região de junção com poucas bases. Para o processamento do conjunto de RNA-Seq com objetivo de quantificação de expressão diferencial, as anotações de novas junções exon-exon dos dois arquivos foram reunidas em uma

única para o segundo alinhamento. Os arquivos gerados no formato BAM foram indexados utilizando o software samtools<sup>159</sup>.

### 3.3.4 Produção de anotação dos introns envolvidos em *splicing* alternativo

A anotação original do GENCODE (versão 25) foi carregada no software R<sup>127</sup> e processada com um código produzido de acordo com o algoritmo a seguir. Para cada *feature* identificada como “gene”, eram buscados todos as *features* relacionadas identificadas como transcritos e que apresentassem entrada no banco de dados CCDS. Caso restasse apenas um transcrito, o gene era descartado da anotação. No caso de dois transcritos, mas que variavam apenas quanto ao início de transcrição alternativo (ou seja, apenas o primeiro exon variava), o gene também era descartado. Aos remanescentes, eram selecionados os exons que não estavam presentes em absolutamente todos os transcritos, denominados então como envolvidos com *splicing* alternativo. Os introns imediatamente anterior e posterior a este exon eram então selecionados.

Quando do contexto da busca de introns relacionados com o *splicing* alternativo em que HuR apresentava ligação (como revelado por RIP-Seq), em um primeiro momento, a opção por uma intersecção de duas anotações (GENCODE filtrada por RefSeq/CCDS) poderia ser extremamente conservadora e impactar no alinhamento das leituras e mensuração da abundância de todos os exons de maneira negativa – vide a ausência de diversas junções exon-exon. Contudo, não houve impacto negativo na mensuração porque a anotação de junções utilizada para o alinhamento (GENCODE) é mais abrangente e o alinhador utilizado, STAR<sup>156</sup>, permite a adição de novas junções exon-exon na anotação *on-the-fly*<sup>160</sup>, ou seja, durante a execução do alinhamento em si utilizando a informação emanada dos próprios fragmentos observados, não demandando, então, de uma anotação extremamente abrangente para acurácia técnica.

Não se optou pela simples reutilização dos dados previamente processados na execução dos trabalhos que disponibilizaram os experimentos de RIP-Seq, PAR-CLIP-Seq e RNA-Seq<sup>96,147,148,161</sup>, mas sim seu novo processamento, devido a não disponibilidade total das contagens obtidas, ao fato dos resultados tabulares só incluírem uma fração dos achados ou porque as informação sobre regiões intrônicas estavam reduzidas ou ausente.

### 3.3.5 Contagem de fragmentos

O número de fragmentos que se sobrepunham em cada região anotada foi obtido utilizando o software HTSeq<sup>162</sup>. Para os dados de RIP-Seq, foi utilizada a anotação preparada no item 3.3.4. Para os dados de RNA-Seq, a anotação foi transformada para o formato GFF utilizando o script em Python *dexseq\_prepare\_annotation.py*<sup>163</sup>. Com o intuito de aplicar a abordagem de subamostragem modificada<sup>164</sup>, os arquivos de alinhamento do experimento de RNA-Seq sem replicatas foram aleatoriamente amostrados com o software samtools<sup>159</sup> por quatro vezes em porcentagens de 10 a 99% da quantidade total de fragmentos presentes, em incrementos de 1%, com o intuito da geração da estimativa do coeficiente de variação e do número de exons amostrados com o aumento de fragmentos avaliados. Para as reamostragens realizadas para a quantificação propriamente dita, foram selecionados 25% dos fragmentos por 128 vezes. Em ambas as situações, a semente geradora de números randômicos foi o número de segundos desde 1 de janeiro de 1970 até o início da execução de cada subamostragem, com pausas exponenciais entre cada início. Assim, a contagem com o HTSeq era realizada para cada alinhamento gerado, em cada uma das condições (controle ou siHuR).

Considerando cada intron como uma unidade individual, os dados de contagem do software HTSeq foram diretamente carregados no pacote DESeq2<sup>165</sup> e processados para a busca dos introns diferencialmente imunoprecipitados entre IP:HuR e IP:Controle. Para a representação gráfica global foi utilizada a ferramenta Circos<sup>166</sup>, com a produção de histogramas de contagem por milhão (CPM) IP:HuR subtraídas de IP:Controle para cada 7 milhões de pares de base (valor da janela do histograma).

Para a avaliação da expressão diferencial por exon foi utilizado o software DEXSeq<sup>163</sup>. As 128 subamostragens de cada amostra foram aleatoriamente sorteados três-a-três por 51 vezes, em cada uma realizando-se uma comparação siHuR contra Controle, similar a um *bootstrap* sobre as subamostragens. As diversas frações dos conjuntos originais foram reunidas três-a-três (sendo utilizado 75% de cada corrida por avaliação) e processadas controle contra siHuR em 51 pareamentos ao acaso, retornando à abordagem proposta por ISODE. A definição do nível de significância em termos de repetições do *bootstrap* obedece uma função de Bernoulli, atingindo uma probabilidade cumulativa de 95% com 31 das 51 repetições. Assim sendo, utilizando este critério, se um exon se mostrou diferencialmente expresso (FDR menor que 0,05) em ao menos 31 das 51 repetições, ele é considerado diferencialmente expresso. Foram considerados como exons diferencialmente expressos os que, acima do limiar estabelecido de *bootstraps*, apresentaram um FDR inferior a 5%. O limiar do *bootstrap*

consistiu no número de vezes esperados para 0,05 probabilidade de uma distribuição de Bernoulli, o mesmo era aceito como diferencialmente expresso. Tal abordagem é baseada em estratégia previamente publicada de IsoEM<sup>164</sup> utilizando para a mensuração de exons de DEXSeq<sup>163</sup>, o que relaxa a demanda de acurácia de anotação no que se refere a junções exon-exon e deixa a cargo do pesquisador a atribuição de cada exon à sua isoforma de relevância<sup>167</sup>. Caso a quantificação das isoformas fosse utilizando ferramentas centradas no transcrito – como a Tuxedo Suite<sup>168</sup>, RNA-Skim<sup>169</sup>, RSEM<sup>170</sup> ou IsoEM<sup>164</sup> nativa, por exemplo –, seria a acurácia do modelo de transcritos impactaria mais nas observações.

### 3.3.6 Análise de enriquecimento de vias e proposta de consequência funcional da troca dos exons nos genes exemplo

Para a avaliação de vias enriquecidas, os nomes dos genes foram padronizados utilizando o software *AnnotationDbi*<sup>171</sup> contra a base humana do Bioconductor *org.Hs.eg.db*<sup>172</sup>. A definição de qual via de Gene Ontology<sup>173</sup> cada gene pertencia foi realizada de acordo com a anotação *GO.db*<sup>174</sup> utilizando o pacote *goseq*<sup>175</sup> utilizando a aproximação de valores de *p* de Wallenius e o ajuste dos mesmos para FDR de Benjamini-Hochberg<sup>142</sup>. Para a base de dados REACTOME<sup>176</sup> a anotação dos genes foi realizada com a base *reactome.db*<sup>177</sup> e o enriquecimento propriamente dito foi realizado com o software *ReactomePA*<sup>178</sup>.

Para a definição de posições transmembrana foi utilizada a ferramenta TOPCONS<sup>179</sup>. Os domínios proteicos envolvidos foram buscados utilizando a base pfam<sup>180,181</sup> e a ferramenta SMART<sup>182,183</sup>. A via metabólica apresentada foi obtida na *Kyoto Encyclopedia of Genes and Genomes* (KEGG)<sup>184</sup>.

### 3.3.7 Visualização dos dados de PAR-CLIP-Seq alinhados e análise evolutiva

A região do intron 14 no genoma humano foi avaliada quanto aos fragmentos alinhados nos experimentos de PAR-CLIP-SEQ utilizando o *Integrative Genomics Viewer*<sup>185,186</sup>. A comparação evolutiva inicial foi realizada com a plataforma VISTA<sup>187,188</sup>. A busca de elementos transponíveis foi realizada com o *RepeatMasker*<sup>189</sup>. A identificação das sequências homólogas em outros organismos foi realizada inicialmente com um BLAST<sup>190</sup> contra o genoma do organismo, alinhamento local par-a-par para identificação dos exons e introns utilizando o LALIGN<sup>191</sup>, e confirmação pelo método do melhor hit de BLAST recíproco<sup>192</sup>. A visualização dos alinhamentos foi realizada com o software MEGA<sup>193</sup>. Os

organismos-modelo foram selecionados pela presença de genoma completo sequenciado e distância evolutiva, sendo *Ciona intestinalis*<sup>194</sup> (Urochordata), *Branchiostoma floridae*<sup>195</sup> (Cephalochordata), *Petromyzon marinus*<sup>196</sup> (Hyperoartia), *Callorhinchus milii*<sup>197</sup> (Chondrichthyes), *Danio rerio*<sup>198</sup> (Actinopterygii), *Xenopus tropicalis*<sup>199</sup> (Amphibia), e *Homo sapiens*<sup>200,201</sup> (Mammalia). A distribuição filogenética dos mesmos foi realizada reproduzindo a posição dos grupos aceita<sup>202</sup>.

#### 3.4 EXPERIMENTO *IN VITRO* DE INTERAÇÃO ENTRE HUR.WT E O INTRON 14 POR FRAP (*FLUORESCENCE RECOVERY AFTER PHOTOBLEACHING*)

Para a avaliação bioquímica da interação, são necessários dois constituintes: a proteína purificada e o RNA transcrito *in vitro* (*in vitro transcription*, IVT). Este segundo passo foi realizado utilizando o kit *Riboprobe In Vitro Transcription System* (Promega) seguindo instruções do fabricante. O vetor utilizado na transcrição foi cortado com BamHI (sítio localizado ao final RNA de interesse), para permitir a transcrição do tipo *run-off*, sem a existência de região terminadora. Como objeto de interesse, foi utilizada a sequência do intron 14 de *GLS* inserida no vetor pGEMT easy (com orientação para expressão via promotor T7) e, como controle, a sequência não específica do vetor pGEM *Express Positive Control* (Promega).

A expressão ectópica da sequência codificante do gene *ELAVL1* fusionada a *tag* de histidinas (para purificação por afinidade a metal) e proteína fluorescente mKO2 foi realizada em células Rosetta™ 2(DE3), utilizando o pET30b. Do pré-inóculo, crescido durante 16 horas a 37°C sob agitação e agentes de seleção canamicina e cloranfenicol, foi removido volume equivalente a 15 unidades de OD600nm e inoculados em 1 erlenmeyer de 2 litros contendo metade deste volume de meio LB com os respectivos antibióticos adicionados. Os frascos foram mantidos sob rotação de 200 rpm a 37°C até atingirem uma OD600 de 1, quando 1 mmol/L de IPTG foi adicionado e a cultura mantida por 16 horas sob agitação. Este cultivo foi centrifugado por 10 minutos a 8000 x g a 4°C, sendo os *pellets* obtidos reunidos em um tubo de cônico de 50 mL e prontamente congelados em nitrogênio líquido. Para ressuspensão do *pellet* único, utilizou-se 40 mL de Tampão de Ressuspensão (50 mmol/L de NaCl, 50 mmol/L de Tris-HCl pH 7,8, 5% de Glicerol, 1 mmol/L de PMSF e 2 mmol/L β-Mercaptoetanol, sendo os dois últimos adicionados logo antes do uso). Ao ressuspendido foi adicionado 120 mg de lisozima (Sigma-Aldrich) e 60 mg de desoxicolato de sódio (Sigma-Aldrich), seguido de uma incubação em gelo, por 1 hora. Após a incubação, o produto foi sonificado (Ponteira Vibra-Cell 630-0420-P) em pulsos de 1 segundos com amplitude de 30% e intervalos de 1 segundos entre os mesmos,

até a suficiente liquefação do lisado. Centrifugou-se o lisado por 45 minutos a 18.000 rpm (Rotor Sorvall SS45) a 4°C.

A cromatografia de afinidade a metais foi realizada por gravidade com 2 mL de resina de afinidade por metais (cobalto) TALON (Clontech, EUA), montada em coluna de vidro, sendo todo o procedimento realizado a 4°C. A resina foi equilibrada com 5 ml/min de resina de Solução A (10 mmol/L de NaCl, 50 mmol/L de Tris-HCl pH 7,8 e 2 mmol/L de  $\beta$ -Mercaptoetanol, adicionado prontamente antes do uso) antes de receber o lisado clarificado. Após passagem do lisado, a resina foi lavada com 5 mL de Solução A/mL resina e 2,5 mL de Solução A/mL resina acrescida de 450 mmol/L de NaCl (para remoção de ácidos nucléicos). Ao fim, a proteína foi eluída com 2,5 mL/mL resina de Solução A acrescida de 500 mmol/L de Imidazol.

A cromatografia de troca iônica foi realizada na plataforma ÄKTA FPLC (GE Healthcare, EUA) com uma coluna HiTrapQ FF (GE Healthcare, EUA), utilizando Solução A descrita acima e Solução B (1 mol/L NaCl, 50 mmol/L Tris-HCl pH 8,5 e 2 mmol/L de ditiotretitol, adicionado prontamente antes do uso) para a eluição da proteína. A amostra foi centrifugada por 10 minutos a 4.000 rpm a 4°C antes de ser aplicada na coluna. A eluição se deu por gradiente contínuo de 0-100% de Solução B por 40 minutos a taxa de 1 mL/minuto e a coleta em frações de 1 mL.

A cromatografia de filtração em gel foi realizada na mesma plataforma ÄKTA FPLC, com uma coluna Superdex 200 10/30 prep grade (GE Healthcare, EUA), utilizando um Solução GF (50 de mmol/L NaCl, 30 mmol/L de Tris-HCl pH 7,8 e 0,5 mmol/L de TCEP). Quinhentos microlitros da proteína previamente centrifugada por 10 minutos, 10.000 g a 4°C foram injetados na coluna e eluída de forma isocrática, em um fluxo de 0,5 mL/minuto. A proteína foi concentrada para o uso.

Para realização do experimento de FRAP propriamente dito, foram reunidos mKO2-HuR.wt e RNA (controle ou Intron 14), a uma concentração final de cada um a 5  $\mu$ M em 50 mmol/L Tris-HCl pH 7,4, 200 mmol/L de NaCl, 50% glicerol 0,55 U/ $\mu$ L de RNaseOUT em água DEPC. Cinquenta microlitros da solução foram depositados em uma microplaca de 96 poços CellCarrier (Perkin Elmer) com fundo de 0,17 mm de espessura e selada com filme plástico. A mesma foi incubada por 30 minutos no estágio do microscópio Leica DMi6000 mantido a 25°C e acoplado a um detector confocal SP8. A objetiva utilizada foi PL APO CS2 63X/NA1.2, laser DPSS (561nm) a 3% de potência na leitura e 100% para *bleach*, detector PMT e pinhole em 3 unidades de Airy. A área de escaneamento foi definida em 41  $\mu$ m x 41  $\mu$ m com um *bleach* central de 35  $\mu$ m x 35  $\mu$ m, a uma velocidade de 1400 Hz. As imagens foram

quantificadas com o módulo de análise FRAP Wizard do Leica LAS AF e os dados plotados e o ajuste de curva de regressão não linear realizado utilizando o software R.

### 3.5 CULTURA DE CÉLULAS

As linhagens utilizadas neste trabalho e respectivos meios de cultivo estão indicadas no Quadro 5. Os meios de cultivo foram adquiridos da Sigma-Aldrich e o soro fetal bovino, que em todas as situações suplementou em 10% (v/v) os meios, da Vitrocell. A atmosfera de incubação foi mantida com 5% de gás carbônico e 100% de umidade, a 37°C.

**Quadro 5 – Células utilizadas neste trabalho e condições de cultivo**

Nome utilizado	Catálogo ATCC	Organismo	Orig. Tumoral	Sítio	Meio	Tamponamento
BT549	HTB-122	<i>Homo sapiens</i>	Mama	Primário	RPMI 1640	2 g/L NaHCO <sub>3</sub>
HCC1143	CRL-2321					
HCC1806	CRL-2335					
HCC1937	CRL-2336					
HCC38	CRL-2314					
HCC70	CRL-2315					
Hs578t	HTB-126					
MDA-MB-157	HTB-24			Efusão pleural		
MDA-MB-231	HTB-26					
MDA-MB-436	HTB-130					
MDA-MB-453	HTB-131					
MDA-MB-468	HTB-132					
DU-145	HTB-81			Próstata		
PC-3	CRL-1435	Próstata	Osso	RPMI 1640		
HEK293T	CRL-11268	Imortalizada SV40	--	DMEM Alta Glicose	1,5 g/L NaHCO <sub>3</sub>	
NIH3T3	CRL-1658	<i>Mus musculus</i>	3T3	DMEM Alta Glicose		

NOTA: ATCC American Type Culture Collection; RPMI: Roswell Park Memorial Institute; DMEM: Dulbecco's Modified Eagle's Medium.

### 3.6 WESTERN BLOT

Os lisados celulares foram homogeneizados em Tampão de Extração (1% Triton-X 100, 100 mmol/L de Tris, pH 7.4, 100 mmol/L de pirofosfato de sódio, 100 mmol/L de fluoreto de sódio, 10 mmol/L EDTA, 10 mmol/L de ortovanadato de sódio, 2 mmol/L de PMSF, e 0.1 mg/mL de aprotinina) e processados como previamente descrito<sup>203</sup>. A estimativa da abundância proteica foi realizada pelo método de Bradford modificado<sup>204</sup>, com o uso de *Bovine Serum*

*Albumin* (BSA) como referência e o reagente comercialmente disponível (BioRad), seguindo instruções do fabricante. Entre 10 e 50 µg de conteúdo total de proteína foram utilizados na separação por gel de poliacrilamida gradiente de 3-15% em SDS-PAGE<sup>205</sup>, produzido com o PAGE Multi-Gel Caster e Gradient Maker 465 (BioRad). A transferência para PVDF de 0.22 µm (BioRad) foi realizada em sistema *Trans-Blot Turbo* (BioRad) com tampão de transferência Timmons<sup>206</sup> (250 mmol/L de Tris, 192 mmol/L de Glicina), seis filtros WypAll X60 (Kimberly-clark) em cada face, a 1.3A por 7 minutos. O bloqueio de interações inespecíficas foi realizado em solução de 3% de leite desnatado (Nestlé) em *Tris-Buffered Saline* (20 mmol/L de Tris pH 7,4, 150 mmol/L de cloreto de sódio) suplementada com 0,1% de *Tween* 20 (Sigma) (TBS-T), sob agitação constante durante 1 hora a temperatura ambiente. Após lavagens em TBS-T, as membranas foram incubadas *overnight* com os anticorpos primários indicados em cada caso, sob agitação constante, a 4°C. A incubação com os anticorpos secundários (Sigma) foi realizada por 3 horas a temperatura ambiente sob agitação constante. A detecção das proteínas foi realizada com o *SuperSignal West Pico Chemiluminescent Substrate* (Thermo Scientific) seguindo instruções do fabricante, e exposição por tempo necessário com o filme autorradiográfico (IBF). Os anticorpos primários foram diluídos 1:1000 em TBS-T e os secundários 1:20000, igualmente em TBS-T. Os primários utilizados foram anti-HuR (Thermo-Fisher, A-21277), anti-GLS (Abcam, ab156876, que reconhece as duas isoformas, KGA e GAC), anti-GAC (Rhea, IM-0323, específico para a isoforma GAC), anti-KGA (Aviva, ARP55098\_P050, específico para a isoforma KGA), Anti-V5 (Thermo-Fisher, R96025), anti-Vinculin (Abcam, ab18058) e anti-β-catenin (Santa Cruz, sc-7199).

### 3.7 TRANSFECCÃO CELULAR

Para a transfecção de células HEK293T foi utilizado o reagente polietilenimina linear em sua forma ácida (25 kDa PEI, Polysciences, 23966-2) ou salina (40 kDa PEI, Polysciences, 24765-2). Para o primeiro caso, o reagente foi solubilizado em água ultrapura a uma concentração de 0,5 mg/mL e a água aquecida até a dissolução completa do mesmo, com o ajuste do pH para 7,0 e esterilização por filtração. O segundo tipo seguiu o mesmo protocolo, mas sem o aquecimento. As células a serem transfectadas foram plaqueadas no dia imediatamente anterior para atingirem uma confluência de 70% a 80% no momento da transfecção. O DNA foi diluído em solução salina estéril (0,9% NaCl) e o PEI na mesma solução, na proporção de 1 µL para cada µg de DNA a ser transfectado. Tanto DNA quanto PEI foram incubados separadamente por 5 minutos, e então misturados por 20 minutos, antes da

adição por gotejamento sobre as células em cultura, em meio normal de cultivo - todavia trocado uma hora antes dessa adição. As células foram mantidas em cultivo por 12 a 16 horas, quando o meio foi trocado removendo os complexos sobrenadantes. As células foram então mantidas até a realização dos experimentos propriamente ditos. Para as demais células, foi utilizado o reagente de transfecção Lipofectamina 2000 (Thermo-Fisher), seguindo instruções do fabricante.

### 3.8 VETORES VIRAIS

#### 3.8.1 Empacotamento (sistema pLKO)

Para a produção de vetores virais – quando necessário – preparações plasmidiais em larga escala dos vetores de transferência (utilizando *PureYield™* Plasmid Maxiprep System, Promega ou NucleoBond Xtra Midiprep, Macherey-Nagel; eventualmente regenerando as colunas utilizadas<sup>207</sup>) foram submetidas ao VVL/LNBio (*Viral Vector Lab*) para empacotamento (segundo protocolo do laboratório) e, posteriormente, titulados utilizando diluição seriada dos mesmos em placas de 24 poços com  $1.10^4$  células NIH3T3 e agente de seleção por uma semana.

#### 3.8.2 Empacotamento (sistemas tet-pLKO e pQC)

Dois milhões de células HEK293T foram plaqueadas em placas de 60 mm de diâmetro no dia imediatamente anterior a transfecção. A transfecção ocorreu como já descrito (Item 3.7), utilizando um total de 7,4 µg de DNA total, distribuído, para vetores lentivirais, em 3,68 µg do vetor de transferência (contendo o gene de interesse), 2,8 µg do vetor psPAX2 (gentilmente cedido pelo Dr. Didier Trono, Addgene #12260) e 0,9 µg de pMD2.G (gentilmente cedido pelo Dr. Didier Trono, Addgene #12259). Para vetores retrovirais, transfectou-se 3,68 µg do vetor de transferência (contendo o gene de interesse), 2,8 µg de pUMVC (gentilmente cedido pelo Dr. Robert A. Weinberg, Addgene #8449) e 0,9 µg de pCMV-VSV-G (gentilmente cedido pelo Dr. Robert A. Weinberg, Addgene #8454).

As coletas foram realizadas a cada 12 horas, com a remoção dos 3 mL e adição de meio de cultura fresco, sendo o coletado centrifugado a  $1000 \times g$  por 5 minutos e o sobrenadante armazenado a 4°C até a realização de um total de seis coletas. Especialmente para os vetores tet-pLKO, as preparações foram concentradas<sup>208</sup>. Brevemente, ao sobrenadante foi adicionado

0,46 volume de uma solução 26,6% de PEG6000, 0,9 mol/L de NaCl e 0,25X de PBS pH7,4. O material foi incubado a 4°C por 24 horas e centrifugado a 7000 x g por 30 minutos a 4°C, sendo o *pellet* remanescente ressuspensionado em um volume dez vezes menor que o inicial com 50 mM de Tris-HCl pH 7,4. Em todos os cenários, os vírus foram titulados utilizando diluição seriada dos mesmos em placas de 24 poços com  $1.10^4$  células NIH3T3 e agente de seleção por uma semana.

### 3.8.3 Transdução

As coletas virais, após o título definido, foram transduzidas com um fator MOI (*multiplicity of infection*) máximo de 0,3 vírus/células plaqueadas – com o intuito de evitar a integração múltipla de vírus em uma única célula. Foram plaqueadas  $2.10^5$  células em placa de 6 poços e, após 24 horas, adicionado 1 mL total de meio contendo  $0,6 \times 10^5$  partículas virais suplementado com 1 µg/mL de polibreno (*hexadimethrine bromide*, Sigma). No dia seguinte, o meio foi substituído ainda sem o agente de seleção e, após outras 24 horas, o meio foi substituído e adicionado o agente para a seleção das células com integração positiva. Para puromicina (Thermo-Fisher, A1113803) empregou-se 1 µg/mL para seleção e manutenção, enquanto que para Geneticina (Sigma, A1720, diluída a 100 mg/mL em HEPES 100 mmol/L pH 7,4) empregou-se 1000 µg/mL para seleção e 200 µg/mL para manutenção posterior.

## 3.9 EXTRAÇÃO DE RNA E PCR QUANTITATIVA

O RNA total das células foi extraído utilizando o *TRI Reagent*® (Sigma) e a produção da primeira fita de cDNA feita através do uso de mix de hexâmetros randômicos (Life Technologies) com Oligo dT (NEB), em 7 partes para 5, e *GoScript*™ *Reverse Transcriptase* (Promega), seguindo instruções dos respectivos fabricantes. A quantificação dos transcritos foi realizada em triplicata técnica com o kit *SYBR*® *Green PCR Master Mix* (Life Technologies) seguindo instruções do fabricante. Os dados foram normalizados pela expressão do *housekeeping* 18S rRNA. Para a quantificação da expressão gênica foi utilizado o método comparativo  $2^{-\Delta\Delta Ct^{209}}$ . Os oligonucleotídeos utilizados estão sumarizados no Quadro 6.

### 3.10 IMUNOPRECIPITAÇÃO DE RNA

Para a imunoprecipitação de RNA foi utilizado anticorpo Anti-HuR (Thermo-Fisher, A-21277) e célula PC-3 selvagem, seguindo protocolo previamente estabelecido, com modificações<sup>210-212</sup>. Brevemente, as células foram lisadas em 20 mmol/L de Tris-HCl pH 7,4, 150 mmol/L de NaCl, 1% Triton X-100, 0,5 U/ $\mu$ L RNaseOUT, 1X *Protease Inhibitor Cocktail* (Roche) em água DEPC no gelo por 30 minutos. O lisado foi sonificado em 10 pulsos de 10 segundos com 30% de amplitude e centrifugadas a 4°C a 12.000 x g. O lisado foi dividido para as condições necessárias e incubados com Magnabeads (Invitrogen) bloqueados com esperma de salmão e BSA e os anticorpos incluídos, sendo mantidos sob rotação a 4°C por 16 horas. As *beads* foram lavadas com tampão NT2 (50 mmol/L de Tris-HCl pH 7,5, 150 mmol/L de NaCl, 1 mmol/L de MgCl<sub>2</sub> e 0.05% de NP-40) por 4 vezes e o RNA das *beads* extraído com TRI Reagent (Sigma). Os procedimentos de transcrição reversa e amplificação foram realizados como indicado para PCR quantitativo (Item 3.9).

**Quadro 6 – Oligonucleotídeos utilizados para a quantificação de transcritos por PCR quantitativa**

Transcrito alvo	Oligo	Sequência 5' – 3'
GLS	GLSFow_qPCR	AAAGCAGTCTGGAGGAAAGG
	GLSRev_qPCR	AGTAGAATGCCTCTGTCCATCTA
ELAVL1	HuRFow_qPCR	CATTAAGGTGTCGTATGCTC
	HuRRev_qPCR	CTGGACAAACCTGTAGTCTG
KGA	KGAFow_qPCR	TGGTGATCAAAGGGTAAAGTC
	KGARev_qPCR	TGCTGTTCTAGAATCATAGTCC
GAC	GACFow_qPCR	GATCAAAGGCATTCCTTTGG
	GACRev_qPCR	TACTACAGTTGTAGAGATGTCC
rRNA 18S	18SFow_qPCR	ATTCCGATAACGAACGAGAC
	18SRev_qPCR	TCACAGACCTGTTATTGCTC

### 3.11 SILENCIAMENTO INDUZÍVEL POR DOXICICLINA

Como forma de validação dupla dos achados, decidiu-se não só empregar o silenciamento utilizando uma única sequência de alvo para HuR, mas, também, estratégias distintas para o mesmo. O silenciamento constitutivo, utilizando o sistema pLKO<sup>125</sup>, estabelece a redução dos níveis da proteína alvo constantemente e de forma não-regulável. Especialmente quando o fenótipo causado apresenta reduz o *fitness* da célula, pode ocorrer o silenciamento epigenético do sistema<sup>213</sup> ou variação dos níveis de eficácia do silenciamento<sup>214</sup>. Com o intuito de controlar o tempo de silenciamento do gene de interesse, utilizou-se concomitantemente o sistema tet-pLKO<sup>126</sup>, o qual transduz as células, permite a seleção pela

expressão constitutiva de gene conferindo resistência e só confere o silenciamento frente a adição de tetraciclina ou algum de seus derivados, como a doxíciclina, mais estável<sup>215</sup>. Células transduzidas pelo sistema tet-pLKO foram distribuídas simultaneamente em duas placas de cultura que, 24 horas depois, tiveram os meios substituídos por meio de cultura adicionado ou não de 50 µg/mL de hiclato de doxíciclina (salvo quando da avaliação de múltiplas doses, sendo utilizada a indicada). As células passaram pela troca de meio diária até o sétimo dia, quando os experimentos propriamente ditos foram iniciados.

### 3.12 IMUNOMARCAÇÃO E MICROSCOPIA DE FLUORESCÊNCIA

As células sob experimentação foram plaqueadas em poços com espessura de fundo inferior a 0,2 mm. Especificamente quando na utilização de células HEK293T, as placas foram tratadas previamente com solução 0,3 mg/mL de colágeno do tipo I em ácido acético 20 mmol/L por 1 hora, seguido de lavagem por três vezes com PBS. A fixação, 24 a 48 horas posteriormente, foi realizada com a adição de agente de fixação por 30 minutos a temperatura ambiente. O mesmo foi preparado com a dissolução de paraformaldeído a 8% em água ultrapura sob aquecimento, e solução paralela de 0,2 mol/L de tampão fosfato pH 7,4 (53 mmol/L de NaH<sub>2</sub>PO<sub>4</sub> e 154 mmol/L de Na<sub>2</sub>HPO<sub>4</sub>), misturadas em proporção equivolúmica, aliquoteado e congelado. As células foram, então, lavadas por três vezes com PBS e permeabilizadas por 5 minutos com 0,2% de Triton X-100 em PBS 1X, por 5 minutos. Foram realizados dois bloqueios, inicialmente quanto às interações eletrostáticas inespecíficas, com 30 minutos de incubação com 5 mg/mL de Heparina Sódica, 5 mg/mL de Dextran (extraído de *Leuconostoc* spp.), 0,1% de Tween 20, 0,05% de azida de sódio em PBS pH 7,4. Após três lavagens com PBS, foi realizado o bloqueio de interações inespecíficas do anticorpo, com 10% de soro normal de cabra, 50 mmol/L de glicina, 0,1% de triton X-100 em PBS pH 7,4 por 1 hora. Tal solução, diluída cinco vezes em PBS pH 7,4, foi utilizada com o tampão para lavagens e diluições subsequentes. Após três lavagens, os anticorpos primários diluídos 1:300 foram adicionados e a placa incubada overnight a 4°C sob leve agitação, em câmara úmida. Outras três lavagens removeram os anticorpos primários para a adição de anticorpos secundários conjugados a fluoróforos (DyLight, Thermo-Scientific) com ou sem a presença de 0,75 µmol/L Faloidina-FITC (P5282, Sigma) por 3 horas a temperatura ambiente em câmara úmida. Após mais três lavagens, o material foi incubado com DAPI (2,5 µg/mL) por 20 minutos, lavado três vezes com PBS e mantido em PBS a 4°C até a leitura. O processamento foi realizado utilizando o software Columbus versão 2.4 (Perkin-Elmer), sendo exportados os dados de quantificação dos

diversos parâmetros acerca de cada célula independente. Brevemente, o software Columbus é uma interface gráfica para o motor de análise de imagens Acapella (Perkin-Elmer). O processamento inicia com a definição de cada elemento através da busca por círculos preenchidos na imagem respectiva ao DAPI (marcando DNA e apresentando formato arredondado). A partir de cada um dos núcleos detectados, a região citoplasmática – definida pela marcação de faloidina, por exemplo – é atribuída a cada um dos núcleos, designando agora células (com citoplasma e núcleo compartimentalizados). Objetos que toquem a borda da imagem são, então, descartados por não estarem completamente representados na coleta, evitando também duplicidade com campos adjacentes. Os valores de intensidade são quantificados utilizando a intensidade observada e, para a avaliação de texturas, máscaras matemáticas são aplicadas sobre a marcação, *binarizando* a imagem e quantificando a área recoberta pelos mesmos. O processamento e exploração gráfica foi realizado com o software R. Os primários utilizados foram anti-HuR (Thermo-Fisher, A-21277), anti-KGA (Aviva, ARP55098\_P050, específico para a isoforma KGA) e anti- $\beta$ -catenin (Santa Cruz, sc-7199).

Para os ensaios de imunomarcacão em esferoides tumorais foi utilizado o mesmo protocolo, salvo para as especificidades da técnica. Os esferoides foram fixados na mesma solução, entretanto por um período de 1 hora, sob agitaçao. Após as lavagens, os mesmos foram incluídos em 4% de agarose baixo ponto de fusão, *overnight* a 4°C. Tal material foi, então, colado na base de um vibratomo Leica VT1000S, com velocidade e vibraçao definidas na velocidade 6, sendo realizados cortes de 30  $\mu$ m de espessura, acondicionados em um poço de uma placa de 24. Com este material assim acondicionado, foi realizado o protocolo similar de imunomarcacão até o último passo, quando os cortes foram transferidos para uma lamínula e montados com ProLong Gold (Thermo-Fisher, P36934). A cura ocorreu por 48 horas, quando as lâminas foram capturadas utilizando o microscópio leitor de placas *Operetta* (Perkin Elmer) no modo confocal *spining disk* fotografando-se vários planos a fim de contemplar o critério de Nyquist de amostragem e o limite de resoluçao de Abbe<sup>216</sup>. As imagens geradas foram empilhadas e obtida uma projeçao máxima, a qual foi mensurada a abundância de marcaçao utilizando o software Fiji<sup>217</sup>/ImageJ<sup>218</sup> e a quantificaçao representada graficamente com regressão polinomial da intensidade em funçao da distância sobre o eixo maior do esferóide, utilizando o software R.

### 3.13 PROLIFERAÇÃO CELULAR EM MONOCAMADA OU TRIDIMENSIONAL

Para avaliar a proliferação celular,  $3 \times 10^3$  células com a modificação específica foram semeadas em placas de 96 wells em quadruplicatas, sendo incubadas em estufa de 5% de CO<sub>2</sub>, a 37 graus por 24 horas. Após este tempo, os meios de cultura foram substituídos, quando necessário, com as depleções ou adições indicadas. Quando na presença de inibição química, 1 µmol/L de CB-839 foi adicionado no meio de cultivo. Neste momento, uma placa fora fixada para a realização da normalização da quantidade inicial de células. A fixação se realizou com solução de formaldeído livre de metanol a 4% (paraformaldeído 4%) em tampão fosfato, por 20 minutos a temperatura ambiente. As células foram então lavadas duas vezes com PBS 1X, incubadas por 5 minutos com DAPI (2,5 µg/mL) a 4°C e, em seguida, novamente lavadas 3 vezes com PBS 1X para armazenamento a 4°C até a leitura. O mesmo procedimento foi realizado para as placas dos demais pontos de leitura. As imagens foram coletadas pelo microscópio leitor de placas *Operetta* (Perkin Elmer) e contadas pelo software *Columbus* versão 2.4.0 (Perkin Elmer). Para o cálculo do *Coefficient of drug interaction* (CDI), foi utilizada a seguinte fórmula<sup>219</sup>:

$$CDI = \frac{\text{Efeito (CB - 839 + Silenciamento ELAVL1)}}{\text{Efeito (CB - 839) x Efeito (Silenciamento ELAVL1)}}$$

Da qual podemos inferir sinergia (CDI < 1), aditividade (CDI = 1) ou antagonismo (CDI > 1) entre dois distintos tratamentos. Assim, quando observamos antagonismo concluímos que os dois tratamentos não apresentam seu desempenho máximo quando na combinação. Aditividade representa exatamente o desempenho máximo na combinação. Por fim, a sinergia é quando os dois tratamentos apresentam desempenho superior ao máximo individualmente obtido.

Para a mensuração da proliferação tridimensional das células BT549,  $4 \times 10^3$  células com o silenciamento de *ELAVL1* ou o respectivo controle foram semeadas em placas *Spheroid Microplate Ultra Low Attachment* (Corning, 4515), e mantidas em estufa de 5% de CO<sub>2</sub>, a 37°C com substituição de 50% do meio de cultivo a cada 48 horas, sendo avaliadas pelo microscópio leitor de placas *Operetta* (Perkin Elmer) em campo claro e analisadas pelo software Fiji<sup>217</sup>/ImageJ<sup>218</sup>. A mensuração do fenótipo proliferativo tridimensional se mostrou um desafio técnico, especialmente quanto a opção entre um parâmetro volumétrico ou de área. Como na ausência de marcação com fluoróforos que permitissem a visualização tridimensional e

consequente reconstrução de volume<sup>220</sup>, optou-se pela amplamente utilizada mensuração de parâmetros bidimensionais (de área ou diâmetro)<sup>221-224</sup>

### 3.14 KNOCK IN DE CAUDA FLUORESCENTE NA ISOFORMA KGA DA GLUTAMINASE HUMANA

Para a utilização da estratégia de *knock in* fluorescente com o sistema CRISPR/Cas9, células HEK293T foram transfectadas, como já descrito, utilizando o plasmídeo px330.puro adicionado do sgRNA para a região do códon de parada da isoforma KGA. Três dias após, o DNA genômico foi extraído (*Wizard SV Genomic DNA Purification System*, seguindo instruções do fabricante, Promega) e avaliada a presença de mutação de ponto com o ensaio da T7 Endonuclease I (segundo instruções do fabricante, NEB). Assim, o *knock in* propriamente dito foi realizado com nova transfecção, do mesmo plasmídeo utilizado, agora, na presença do plasmídeo doador de região homóloga de recombinação. Após três dias foi adicionado agente de seleção zeocina e as células mantidas até a morte celular de controle não-transfectado. Esse *pool* celular selecionado foi, então, utilizado nos experimentos subsequentes.

### 3.15 AVALIAÇÃO DO CICLO CELULAR UTILIZANDO MICROSCOPIA

A avaliação da posição no ciclo celular das células utilizando microscopia foi realizada conforme previamente descrito<sup>225</sup>, sendo as placas preparadas e capturadas como se para uma proliferação normal, mudando apenas o passo de quantificação. A mesma envolveu a mensuração da intensidade de fluorescência de cada célula individualmente em sua região nuclear, integrando-se a mesma pela área do núcleo. De posse do valor individual para cada célula, foram estimadas distribuições suavizadas para o logaritmo na base 10 da soma de intensidade de DAPI, proporcional à quantidade de DNA, e as posições de separação entre as fases estabelecidas do ciclo utilizando a intensidade de fluorescência para a célula controle do experimento.

O bloqueio duplo de timidina foi realizado adicionando 100 µg/mL de timidina (Sigma, T1895), seguido de 16 horas de incubação, lavagem e troca de meio, 12 horas de incubação e uma nova adição de 100 µg/mL de timidina por mais 12 horas. Ao final do último intervalo as células foram coletadas ou fixadas, dependendo do tipo de experimento a ser realizado e processadas como já indicado.

### 3.16 ENSAIOS REPORTER

Para o ensaio repórter com o sistema RG6, as células foram transfectadas como já indicado (Item 3.7) e avaliadas 48 horas após. A fixação se realizou com solução de formaldeído livre de metanol a 4% (paraformaldeído 4%) em tampão fosfato, por 20 minutos a temperatura ambiente. As células foram então lavadas duas vezes com PBS, marcadas com DAPI e armazenadas a 4°C até a leitura. Quando as células não expressavam as modificações de forma induzível ou constitutiva, durante a transfecção foi realizada a co-transfecção dos plasmídeos para a indução da alteração molecular e o prazo de incubação estendido para 3 dias.

Para os sistemas de mensuração da estabilidade utilizando a 3'UTR do mRNA das isoformas de *GLS*, os vetores psiCHECK respectivos foram transfectados na presença dos plasmídeos para as modificações de expressão de interesse e as células mantidas por 3 dias até a lise com o *Passive Lysis Buffer* por 20 minutos, sob agitação, a temperatura ambiente. O material lisado foi homogeneizado e avaliado quanto a luminescência com o *Dual-Luciferase® Reporter Assay System* (Promega), seguindo instruções do fabricante. A leitura foi realizada com o leitor de placas *EnSpire 2600* (Perkin-Elmer) em placas de fundo branco. Brevemente, foi adicionado 10 µL de lisado a todos os poços, sendo – posteriormente e de forma individual a cada poço – adicionados 25 µL do reagente *LARII*, mensurada a luminescência de *Firefly* por integração de 10 segundos de emissão, adicionados 25 µL do reagente *Stop&Glo* e mensurada a luminescência de *Renilla* por integração de 10 segundos de emissão. Os valores reportados são as razões de cada poço, *Renilla* sobre *Firefly*.

### 3.17 QUANTIFICAÇÃO DO CONSUMO DE GLUTAMINA

Uma versão modificada de protocolo pré-existente foi estabelecida<sup>226</sup>. Brevemente,  $3 \times 10^4$  (BT549) ou  $4 \times 10^4$  (MDA-MB-231) células foram semeadas em um volume de 100 µL numa placa de 96 poços e incubadas em estufa a 37°C e 5% de CO<sub>2</sub> por 16 horas. O meio foi removido e adicionados 50 µL de meio RPMI 1640 com FBS 10% fresco, do qual uma alíquota foi removida. Paralelamente, o mesmo foi adicionado a poços vazios da placa. Após o tempo de incubação, o meio foi removido e congelado a -80°C até a avaliação. A placa foi fixada e processada como em um experimento de proliferação regular para normalizar inconsistências na contagem realizada.

O ensaio consistiu em adicionar 10 µL deste meio a 50 mmol/L de Tris-acetato pH 8,6, 2 mmol/L NAD (Sigma) e 50 mmol/L de fosfato dipotássico (JT Baker) em um volume de

180  $\mu\text{L}$  e realizar uma leitura a 340 nm. Posteriormente, 0,5 unidades de L-GDH foram adicionadas em um volume de 10  $\mu\text{L}$  e a placa incubada por 1 hora a temperatura ambiente, selada. A absorbância a 340 nm foi novamente lida e 50 nmol/L final de glutaminase recombinante foi adicionada em um volume de 10  $\mu\text{L}$  para uma nova incubação por 1 hora a temperatura ambiente e leitura a 340 nm. A quantificação de glutamina é a subtração da terceira leitura pela segunda e a de glutamato da segunda pela primeira. Os valores são transpostos para quantidades através de curva padrão produzida em meio sem glutamina, sendo o consumo de glutamina definido pela subtração da quantidade de glutamina mensurada no meio incubado concomitantemente, mas sem a presença de células, pelos valores observados nos meios experimentais, normalizando-se pelo número de células contadas. Todas as leituras foram realizadas com o leitor de placas *EnSpire 2600* (Perkin-Elmer)

### 3.18 ENSAIO DE ATIVIDADE DE GLUTAMINASE COM LISADO CELULAR

As células foram lisadas em 25 mmol/L de HEPES pH 8,0 (Sigma-Aldrich), 150 mmol/L de NaCl (Merck), 1 mmol/L de EDTA (Merck), 0,01% de Triton X-100 (USB), 10 mmol/L de pirofosfato de sódio (Sigma), 10 mmol/L de fluoreto de Sódio (Sigma), 10 de mmol/L ortovanadato de sódio (Sigma), 2 de mmol/L PMSF (Sigma), 10  $\mu\text{mol/L}$  de leupeptina (Sigma), 1  $\mu\text{mol/L}$  de pepstatina (Sigma), 2  $\mu\text{g/mL}$  de aprotinina, seguido de 20 passagens em agulha de 26 gauges. O ensaio de atividade foi realizado conforme previamente publicado <sup>26</sup>. Brevemente, 10  $\mu\text{g}$  de lisado celular (quantificado pelo método de Bradford), foi adicionado a uma placa de 96 wells em uma solução de 50 mM Tris-Acetato pH 8,6, 0,5 unidades de L-GDH Bovina (Sigma), 2 mM NAD (Sigma), 20 mM fosfato dipotássico (Sigma) e 7,5 mmol/L de L-Glutamina (Sigma). A absorção de NADH foi avaliada no tempo no comprimento de onda de 340nm com o leitor de placas *EnSpire 2600* (Perkin-Elmer). A inclinação da reta que construída com ao menos metade dos pontos avaliados que apresentasse o melhor  $r^2$  foi utilizada para a comparação entre as amostras.

### 3.19 MENSURAÇÃO DA DEPENDÊNCIA DE GLUTAMINA PARA RESPIRAÇÃO CELULAR UTILIZANDO O OXÍGRAFO *SEAHORSE XF24E*

Uma versão modificada do *Seahorse XFp Mito Fuel Flex Test Kit* (Agilent) foi utilizada. Brevemente,  $1,6 \times 10^5$  células BT549 foram semeadas em um volume de 100  $\mu\text{L}$  numa placa XF24 de 24 poços e incubadas em estufa a 37°C e 5% de CO<sub>2</sub> por 1 hora. Foram

adicionados mais 150  $\mu\text{L}$  de meio e as células incubadas por mais 15 horas nas mesmas condições. Simultaneamente, cada poço do XFe24 *sensor cartridge* foi equilibrado com 1000  $\mu\text{L}$  de *Seahorse XF Calibrant Solution* à 37°C em atmosfera ambiente. As placas foram transportadas em atmosfera hermética de mesma constituição gasosa até o equipamento, lavadas com 500  $\mu\text{L}$  de *SeaHorse XF Assay Medium* adicionado de glutamina, glutamato e piruvato de sódio e pH ajustado para 7,4 e mantidas em 500  $\mu\text{L}$  do mesmo meio por 1 hora a 37°C em atmosfera ambiente. Após a calibração do XFe24 *sensor cartridge* previamente equilibrado, o experimento foi realizado com 3 ciclos de leitura da respiração basal, adição de 3  $\mu\text{mol/L}$  de CB-839, para a inibição da enzima glutaminase, seguido de 3 ciclos de leitura da respiração após inibição. Os valores percentuais representados indicam a fração da respiração que foi inibida após a adição do composto.

### 3.20 MENSURAÇÃO METABÓLICA POR GAS CHROMATOGRAPHY-MASS SPECTROMETRY (GC-MS)

Três milhões de células, com o silenciamento de *ELAVL1* induzido ou não, foram semeadas em placas de 60 mm e incubadas por 24 horas. Quando do prazo, o meio foi removido e as células lavadas com PBS por três vezes e uma última com água ultrapura, sendo removido todo líquido remanescente, para a adição de *quench* de 400  $\mu\text{L}$  de acetonitrila gelada e 300  $\mu\text{L}$  de água ultrapura gelada contendo 5 nmoles de norleucina. As células foram raspadas, transferidas para um microtubo que fora reservado. Outra quantidade igual de acetonitrila e água foi adicionada à placa, na ausência de norleucina, para a remoção dos resíduos remanescentes e recolhida no mesmo microtubo. Após congelamento em nitrogênio líquido e armazenamento a -80°C até o processamento, foram adicionados 400  $\mu\text{L}$  de clorofórmio gelado a cada tubo, que foi agitado vigorosamente e centrifugado a 10.000 x g por 10 minutos a 4°C. A fase superior foi coletada, congelada em nitrogênio líquido e liofilizada. As amostras foram então enviadas à *facility* Laboratório de Metabolômica (CTBE/CNPEM), onde foram derivatizadas por MTBSTFS e avaliadas no cromatógrafo gasoso Agilent acoplado a um detector de massa por tempo de voo Pegasus (Leco).

Os resultados *splitless* da injeção foram convertidos pelo software ChromaTOF (Leco) no padrão aberto NetCDF, o qual pode ser carregado no software R para a utilização do pacote *metaMS*<sup>227</sup> e proposta de pipeline inspirada no Workflow4Metabolomics<sup>228,229</sup>. Foi utilizado o banco de dados espectral *GMD\_20111121\_MDN35\_FAME*, produzido pela *GOLM Metabolome Database*<sup>230</sup>. Inicialmente, as cromatografias foram avaliadas grosseiramente para

a correção do tempo de retenção (*rt*) estimado para a escada de metil-ésteres de ácido graxo inserida com o intuito de corrigir oscilações no domínio cromatográfico. Com os valores corretos de média e desvio das corridas, foi ajustado o cromatograma utilizando o índice de retenção de Kovats (*RI*) que permite a comparação de dados cromatográficos obtidos em diferentes condições experimentais. Então, foram produzidos pseudo-espectros (apenas a partir de um pico cromatográfico os fragmentos de massa eram integrados) e comparados inicialmente entre as amostras, com uma tolerância de *drift* cromatográfico de 5000 unidades de *RI* (cerca de 5 segundos). Esses objetos (*features*) metabólicas identificadas foram, então, comparadas com a biblioteca, em uma tolerância de 60% do índice de similaridade, para identificação. *Features* mais abundantes nas lavagens foram excluídas, tal qual as que variaram mais do que 25% entre os *pools* injetados entre as cromatografias. A mensuração dos metabólitos foi feita a partir da abundância de contagens dos picos do pseudospectro e a comparação das mesmas realizadas com o pacote *limma*<sup>231</sup> e a análise de PCA com o *ggbiplot*<sup>232</sup>.

### 3.21 ENSAIOS DE MIGRAÇÃO E INVASÃO CELULAR

#### 3.21.1 Extração de colágeno do tipo I

Foi utilizada metodologia previamente disponível com modificações<sup>233</sup>. Brevemente, 12 caudas de rato armazenadas a -80°C foram descongeladas por 24 horas a 4°C, lavadas em água ultrapura e secas. As vertebras foram rompidas e puxadas, removendo os tendões de colágeno que foram armazenados em PBS 1X até o processamento. Os tendões foram então sequencialmente lavados em PBS (5 minutos), acetona (5 minutos) e isopropanol 70% (5 minutos) e depois mantidos em agitação a 4°C em 750 mL 20 mmol/L de ácido acético em água ultrapura por 48 horas. Após, os tendões foram homogeneizados com liquidificador utilizando raspas de gelo produzido com água ultrapura. O material foi então centrifugado por 1 hora a 10.000 x g a 4°C e o sobrenadante congelado a -20°C. Após o completo congelamento e fracionamento do material, o mesmo foi liofilizado e armazenado a -80°C. A solubilização foi realizada mensurando massa suficiente para 8 mg por mL e homogeneizada em liquidificador em 20 mL de 20 mmol/L de ácido acético em água e dialisado contra ácido acético de mesma concentração. A esterilização foi realizada por vapor de clorofórmio por uma semana a 4°C.

### 3.21.2 Ensaio de migração celular

Para ambos ensaios de *scratch-and-wound* (de migração e invasão) as placas foram previamente tratadas com uma solução 0,3 mg/mL de colágeno do tipo I em ácido acético 20 mmol/L por 1 hora, seguido de lavagem por três vezes com PBS. Então,  $3 \times 10^4$  (BT549) ou  $4 \times 10^4$  (MDA-MB-231) células foram semeadas em um volume de 100  $\mu$ L nesta placa tratada de 96 poços e incubadas em estufa a 37°C e 5% de CO<sub>2</sub> por 16 horas. Após o período de incubação, o meio foi removido, as células lavadas com PBS por uma vez, e meio sem soro fetal bovino (mas com 0,1% de BSA inativado termicamente) foi adicionado, para uma segunda incubação nas mesmas condições por 24 horas. Quando este passo foi concluído, o meio foi novamente removido e, com o auxílio de um aspirador multicanal, riscos foram feitos no fundo de todos os poços. Para a remoção das células raspadas, a placa foi lavada com PBS e adicionado meio DMEM alta glicose tamponado com 25 mmol/L de HEPES pH 7,4. A placa foi então incubada no microscópio *Operetta* (Perkin Elmer) com estágio aquecido a 37°C durante todo o período das capturas de campo claro. As imagens obtidas foram organizadas e cada um dos poços teve a cobertura celular mensurada nas contínuas medidas comparadas com a primeira, utilizando o software Fiji<sup>217</sup>/ImageJ<sup>218</sup> com macro produzida para este fim.

### 3.21.3 Ensaio de invasão celular utilizando *scratch-and-wound* embebido em matriz

Quando da avaliação da invasão celular em matriz tridimensional, após a realização dos riscos, a placa foi mantida em gelo por 5 minutos e sobreposta com 50  $\mu$ L de solução de 2 mg/mL de colágeno do tipo I, em DMEM alta glicose tamponado 25 mmol/L de HEPES pH7,4, 20% de soro fetal bovino e hidróxido de sódio suficiente para o ajuste do pH entre 7 e 8. A matriz foi polimerizada a 37°C por 30 minutos e sobreposta com 200  $\mu$ L de meio de cultura, de mesmas condições de tamponamento, para ser lida e quantificada de mesma maneira. Quando na presença de inibição química, 1  $\mu$ mol/L de CB-839 foi adicionado no meio sobreposto e na matriz.

### 3.21.4 Ensaio de invasão celular utilizando câmara de Boyden

Foi utilizada variação de protocolos previamente disponíveis<sup>234-239</sup>. As células foram preliminarmente semeadas em placas de 60 mm e incubadas por 16 horas, quando foram então lavadas com meio sem soro e mantidas em meio na presença de 0,1% de BSA por 24

horas. Nesse instante, foram tripsinizadas (utilizando metade da concentração usual, 0,025% m/v) e a mesma inativada com meio contendo 0,1% de BSA na ausência de soro fetal bovino. Após duas centrifugações a 200 x g por 5 minutos seguidas de ressuspensão em meio com 0,1% de BSA, as mesmas foram contadas e adicionadas às câmaras de Boyden previamente preparadas com solução quimioatrativa de 5 ng/mL de fator de crescimento epitelial. As células invadiram por 8 horas, quando o inserto contendo as membranas foi fixado em formaldeído por 30 minutos e lavado em PBS. As células não invadidas foram removidas com um cotonete e as invadidas marcadas com DAPI em solução contendo 1% de Triton X-100. A leitura foi realizada utilizando o microscópio *Operetta* (Perkin Elmer), em seu modo confocal, e a contagem de núcleos foi realizada com a projeção máxima da pilha de imagens obtidas para cada câmara, pilha essa centrada na membrana que serviu como barreira, sob a qual as células invadidas estavam aderidas. A preparação preliminar das placas envolveu a adição de 6 µg totais de matrigel (em concentração de 120 µg/mL) diluído em água ultrapura em cada parte superior da câmara de Boyden, permitindo a secagem 24 horas em fluxo laminar. No dia da execução propriamente dita do experimento, a placa foi reconstituída na porção superior com meio de cultura sem soro e na porção inferior com colágeno do tipo I a 60 µg/mL, para permitir a adesão celular. Antes da adição das células, a porção superior teve o meio removido e a inferior foi lavada com PBS.

### 3.22 DESENHO EXPERIMENTAL CELULAR E AVALIAÇÃO ESTATÍSTICA

A quantidade de replicatas experimentais técnicas são demonstradas sobre os gráficos, na forma de pontos. Dada a utilização do método do  $2^{-\Delta\Delta C_t}$  para mensuração das diferenças dos experimentos de qPCR, não é possível a representação individualizada dos pontos, sendo representados nos gráficos duplicatas ou triplicatas técnicas. Para os gráficos de migração e proliferação, as médias sumarizadas são originadas de, ao menos, quadruplicatas técnicas. A normalidade dos resultados foi acessada com o teste de Shapiro-Wilk. Quando em comparações binárias, foram realizados testes-t bicaudais de médias, empregando a correção de Welch para variâncias distintas. Para mais de um par de comparações, foram realizadas análises de variância seguidas de teste posterior de diferença significativa honesta de Tukey.

## 4 RESULTADOS

### 4.1 RELEVÂNCIA CLÍNICA DE HUR COMO ELEMENTO DE PROGNÓSTICO E SUA RELAÇÃO COM AS ISOFORMAS DE *GLS*

Com o intuito de conhecer os efeitos de HuR em diversos tecidos tumorais de acordo com sua origem topográfica, foram utilizados os dados do *The Cancer Genome Atlas* (TCGA). O projeto contempla dados de transcriptômica, epigenômica, proteômica, clínicos e de imagem para 33 tumores distribuídos, dos quais 10 considerados raros, em 27 tecidos de origem (Quadro 7).

**Quadro 7 – Legenda da representação topográfica dos órgãos cujos dados tumorais (e de tecido normal) estão disponíveis no TCGA**

Sexo	#	Tecido	Código TCGA	Tipo
Indiferente	1	Cérebro	GBM	Glioblastoma multiforme
			LGG	Glioma de baixo grau
	2	Olho	UVM	Melanoma uveal
	3	Cabeça e pescoço	HNSC	Carcinoma de células escamosas de cabeça e pescoço
	4	Tireóide	THCA	Carcinoma papilar de tireóide
	5	Esôfago	ESCA	Câncer de esôfago
	6	Pleura	MESO	Mesotelioma
	7	Pulmão	LUAD	Adenocarcinoma de pulmão
			LUSC	Carcinoma de células escamosas de pulmão
	8	Linfonodos	DLBC	Linfoma difuso de grandes células B
	9	Timo	THYM	Timoma
	10	Fígado	LIHC	Carcinoma hepatocelular
	11	Ducto biliar	CHOL	Colangiocarcinoma
	12	Pâncreas	PAAD	Adenocarcinoma pancreático ductal
	13	Estômago	STAD	Adenocarcinoma de estômago
	14	Pele	SKCM	Melanoma cutâneo
	15	Adrenal	ACC	Carcinoma Adrenocortical
			PCPG	Feocromocitoma & Paraganglioma
	16	Rim	KIRC	Carcinoma renal de células claras
			KICH	Carcinoma renal cromóforo
			KIRP	Carcinoma papilar renal
17	Intestino grosso	COAD	Adenocarcinoma de cólon	
18	Bexiga	BLCA	Cancer de bexiga	
19	Intestino delgado	READ	Adenocarcinoma de reto	
20	Tecidos macios	SARC	Sarcoma	
21	Medula óssea	LAML	Leucemia mielóide aguda	
Feminino	22	Mama*	BRCA	Câncer de mama
	23	Útero	UCEC	Câncer de endométrio
			UCS	Carcinossarcoma uterino
	24	Ovário	OV	Cistadenocarcinomas serosos do ovário
25	Cérvice uterino	CECSC	Câncer cervical	
Masculino	26	Prostata	PRAD	Adenocarcinoma de próstata
	27	Testículos	TGCT	Câncer de células germinativas de testículo

NOTA: † Tumores considerados raros pelo TCGA. \* Indicado como tumor *tipicamente* feminino, também acometendo homens. # Numeração encontrada nas imagens.

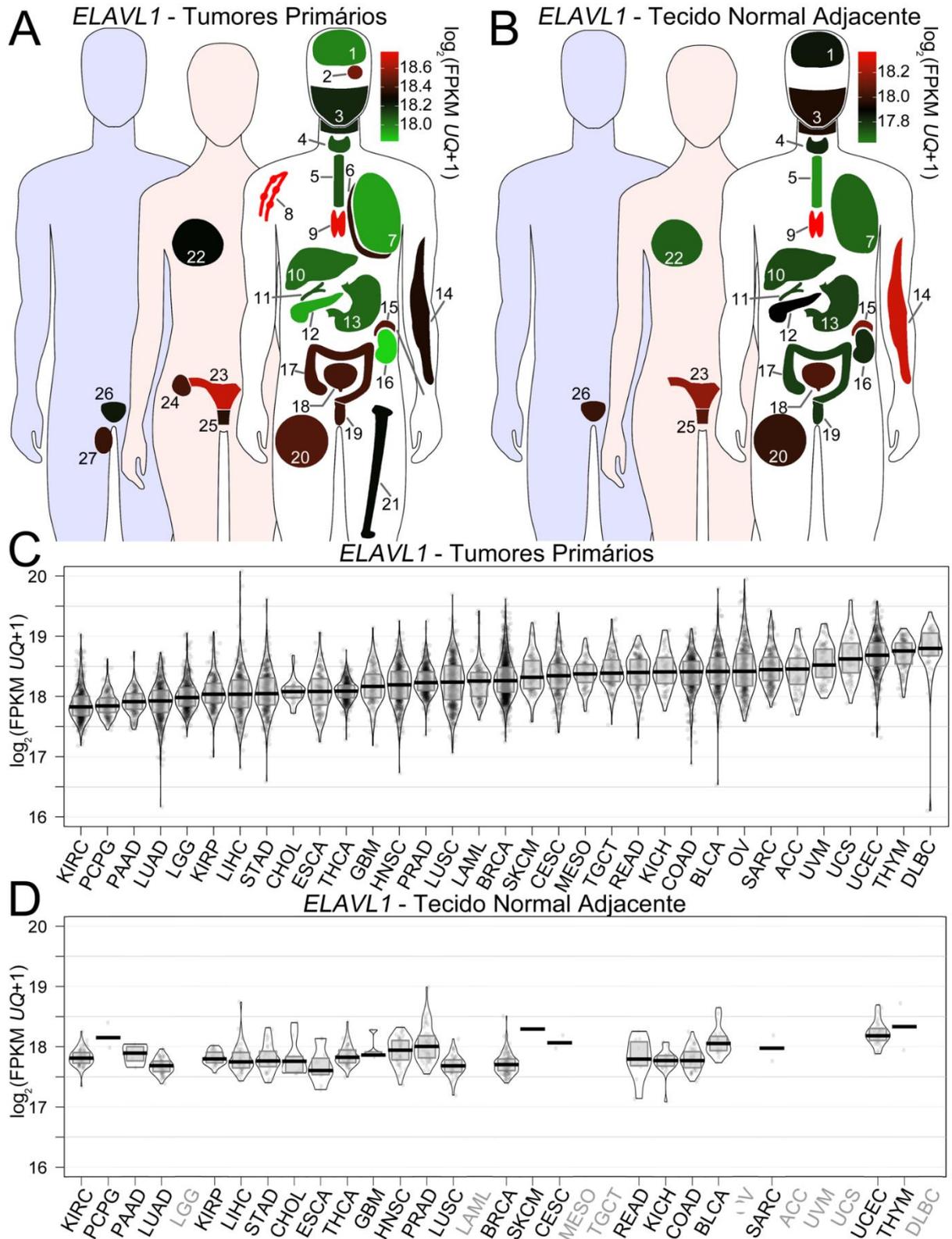
#### 4.1.1 Correlação entre os níveis de mRNA de *ELAVLI* com parâmetros clínicos

Para a avaliação da expressão em um contexto topograficamente identificado, decidiu-se criar uma representação gráfica que consiste em um mapa de calor representado sobre um conjunto de corpos humanos estilizados (descrevendo um homem e seus tumores específicos, uma mulher e seus tumores específicos e um corpo sem identificação de gênero para os tumores comuns a ambos, com numeração e associação descritas no Quadro 7 e no Apêndice 1). Tal representação permite a observação direta de parâmetros numéricos simples em um contexto fisiológico, graças à associação com escala de cor. Utilizando tal representação topográfica, então, podemos observar a expressão de *ELAVLI*, gene que codifica para a HuR, mais elevada em tecidos de tumores primários de linfoma difuso de grandes células B, timoma e câncer de endométrio (representados pelos números 8, 9 e 23, respectivamente, Figura 11A), sendo que timoma e câncer de endométrio também representam os tecidos normais com maior expressão (Figura 11B). É preciso observar que não há amostras de células normais de linfócitos B (assim como de testículo, medula, ovário, mesotélio e olho) dentro do TCGA e o tamanho amostral dos tecidos normais de timo é pequeno.

Para a melhor visualização dos valores de expressão em FPKM, tais valores, normalizados pelo método do quartil superior (*upper quartile*, UQ), são também apresentados na forma de gráfico (Figura 11C para tecido tumoral e Figura 11D para tecido normal adjacente); é possível ver que, tanto para tecido tumorais como normais, as medianas estão acima de 17. Para se ter uma noção do nível de expressão deste gene em relação aos demais genes, em mama, a mediana da expressão, mensurada em  $\log_2$  (FPKM UQ +1) de todos os genes é  $7,4 \pm 3,2$ , tanto entre as amostras normais e tumorais, quando consideramos os transcritos sem expressão; removendo estes transcritos zerados, a mediana é,  $13,5 \pm 0,40$ . Desta maneira, fica implícito que a expressão de *ELAVLI* varia de forma intensa e está acima da média de expressão dos genes em tecido mamário, sugerindo grande abundância, e possível importância deste gene<sup>240</sup>.

Na comparação intertumor, a oscilação das medianas é de uma unidade de  $\log_2$ , implicando em uma quase duplicação do nível de expressão (em escala linear) entre o tumor de maior (DLBC, 17,5) para os de menor mediana (KIRC, 18,5) (Figura 11C). Com relação aos valores medianos dos tecidos normais com tamanho amostral superior a 4 (que apresentam definições de primeiro e terceiro quartis, além da distribuição suavizada) encontra-se uma variação em uma escala próxima a 0,4 unidades de  $\log_2$  (entre THYM e ESCA), muito inferior a observada nos tipos tumorais (Figura 11D). A grande magnitude de variação entre tecidos

tumorais pode indicar que este gene pode ter uma função biológica importante para determinados tipos tumorais.



**Figura 11 – Níveis de expressão do mRNA de *ELAVL1* em tecidos normais e tumorais humanos**  
 Representação topográfica dos níveis de expressão de *ELAVL1* em tumores primários (A) e em tecidos normais adjacentes a tumores primários (B). Os valores representados nas escalas de cor representam a mediana dos níveis

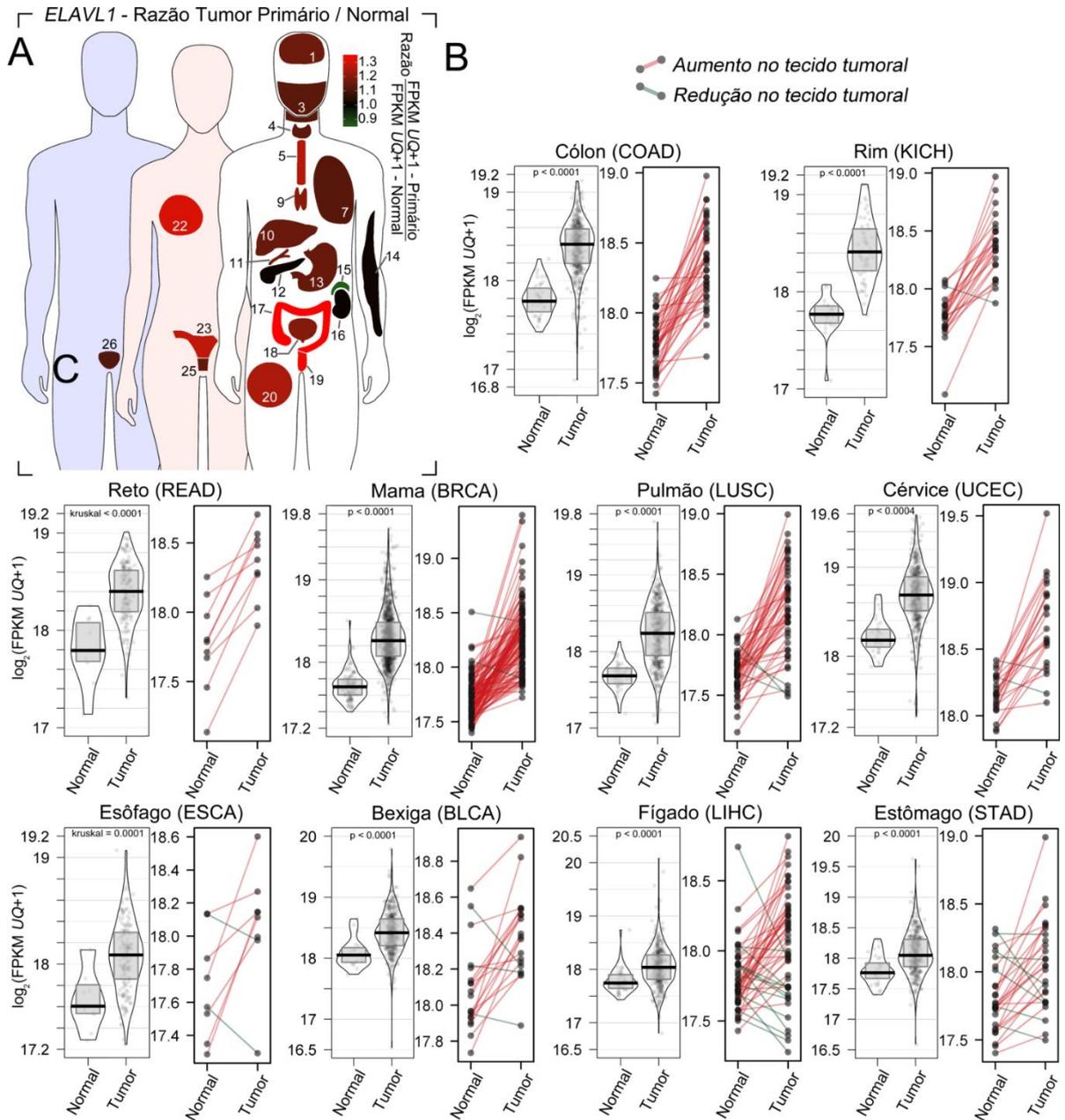
de abundância de mRNA mensurados em FPKM normalizado pelo quartil superior em escala logarítmica e estão dimensionados de forma independente entre os gráficos. Gráfico RDI (*Raw (data), Description and Inference*) dos níveis de expressão de *ELAVLI* em tumores primários, ordenados crescentemente de acordo com a mediana (C) e em tecidos normais adjacentes a tumores primários, obedecendo a ordenação de C (D). Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados como círculos. A numeração dos órgãos está presente no Quadro 7.

Entre as amostras de tecido de um mesmo tipo tumoral, o intervalo interquartil aproxima-se a 0,4 unidades de  $\log_2$  em vários tipos tumorais (Figura 11C, mais facilmente observado nos exemplos da Figura 12C), o que implica variações lineares de 30% entre primeiro e terceiro quartil. Comparativamente, os tecidos normais apresentam variações na metade desta escala, o que indica uma heterogeneidade superior no contexto tumoral quanto à expressão de *ELAVLI*.

Para facilitar a visualização da expressão relativa entre os tecidos tumorais e seus pares normais, a razão de expressão entre tecido tumoral e normal foi calculada (Figura 12A) e as diferenças estatisticamente significativas (*p-value* <0,05) foram definidas (Tabela 2). Observa-se que todos os tumores, com exceção de quatro tipos (melanoma – SKCM –, adenocarcinoma de pâncreas – PAAD –, carcinoma renal de células claras – KIRC – e feocromocitoma & paraganglioma – PCPG), a razão é superior a 1,1; vale ressaltar, entretanto, que para três deles, PAAD, PCPG e SKCM o limitado tamanho amostral para tecido normal impede inferências amplas. Assim, conclui-se que *ELAVLI* tem expressão aumentada na grande maioria dos tumores em relação ao respectivo tecido normal, quando observados os dados disponíveis no TCGA (Figura 12C).

De posse das informações disponibilizadas pelo TCGA, decidimos investigar o efeito da expressão de *ELAVLI* no prognóstico dos pacientes. Uma maneira de fazê-lo seria a separação dos pacientes em alta e baixa expressão seguida da análise dos parâmetros de sobrevivência. Caso a distribuição de expressão apresentasse um comportamento bimodal (com clara distinção entre grupos com diferentes níveis de expressão), modelos matemáticos de mistura gaussiana<sup>241,242</sup>, já aplicados em alguns exemplos clássicos, poderiam ser empregados<sup>243,244</sup>. Entretanto, em tecidos tumorais, observa-se que a expressão de *ELAVLI* segue perfil normal com apenas uma moda (após a transformação logarítmica, necessária devido a extrema assimetria dos dados na escala linear) ou duas modas muito próximas, não havendo mais de uma distribuição normal clara com médias e desvios independentes. Tal fato pode ser observado nas distribuições suavizadas nos gráficos mostrados e não por testes de normalidade, dada a redução do poder frente a grande tamanhos amostrais<sup>245</sup>. Outra possível abordagem<sup>246</sup>, consiste na utilização dos valores de *z-score* (quantos desvios-padrão um

determinado paciente se desloca em relação à média da população global<sup>247</sup>) com um limiar arbitrário a partir do qual os grupos de alta e baixa expressão são determinados.



**Figura 12 – Razão entre os níveis de expressão de *ELAVL1* entre tecido tumoral e tecido normal adjacente** (A) Representação topográfica da razão entre a mediana de expressão de *ELAVL1* em tecido tumoral sobre o tecido normal adjacente denotada pela fração dos valores de FPKM normalizados pelo quartil superior em escala linear. (B) À esquerda, Gráficos RDI dos níveis de expressão de *ELAVL1* em tecidos normais adjacentes (Normal) comparado aos tumorais primários (Tumor) dos 10 tipos tumorais mais sobreexpressos. À direita, gráficos pareados entre tecidos normais adjacentes e seus respectivos tecidos tumorais, ligados por linhas vermelhas quando do aumento da expressão e verdes quando da redução. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados sobre os dados como círculos. Os valores são mensurados em FPKM normalizado pelo quartil superior em escala logarítmica e estão dimensionados de forma independente entre os gráficos. *P-values* originados de teste t não-pareado; para tamanhos amostrais normais pequenos ( $\leq 10$ ) é representado o *p-value* para o teste de não-paramétrico de Kruskal-Wallis. A numeração dos órgãos está presente no Quadro 7.

**Tabela 2 – Lista com os valores absolutos das razões das medianas de *ELAVLI* entre tecido tumoral e normal representadas na Figura 12A com o órgão afetado e o respectivo código empregado pelo TCGA.**

Tecido	Código TCGA	Tipo	Razão	<i>p-value</i>	FDR	<i>ks</i>
Intestino grosso	COAD	Adenocarcinoma de cólon	1,56	0,001443 †	0,0022	<0,0001 †
Rim	KICH	Carcinoma renal cromóforo	1,56	<0,0001 †	<0,0001	<0,0001 †
Intestino delgado	READ	Adenocarcinoma de reto	1,52	<0,0001 †*	<0,0001	<0,0001 †*
Mama	BRCA	Câncer de mama	1,47	<0,0001 †	<0,0001	<0,0001 †
Pulmão	LUSC	Carcinoma de células escamosas de pulmão	1,47	<0,0001 †	<0,0001	<0,0001 †
Útero	UCEC	Câncer de endométrio	1,42	<0,0001 †	<0,0001	<0,0001 †
Esôfago	ESCA	Câncer de esôfago	1,40	0,0005 †*	0,0008	0,0014 †*
Tecidos macios	SARC	Sarcoma	1,37	0,1015 *	0,1282	0,0637 *
Timo	THYM	Timoma	1,30	0,05 *	0,0689	0,3796 *
Bexiga	BLCA	Cancer de bexiga	1,29	<0,0001 †	<0,0001	<0,0001 †
Ducto biliar	CHOL	Colangiocarcinoma	1,26	0,0014 †*	0,0022	0,0009 †*
Cérebro	GBM	Glioblastoma multiforme	1,24	0,2174 *	0,2609	0,0217 †*
Fígado	LIHC	Carcinoma hepatocelular	1,22	<0,0001 †	<0,0001	<0,0001 †
Estômago	STAD	Adenocarcinoma de estômago	1,22	<0,0001 †	<0,0001	<0,0001 †
Cérvice uterino	CESC	Câncer cervical	1,22	0,3548 *	0,3871	0,0492 †*
Tireóide	THCA	Carcinoma papilar de tireóide	1,20	<0,0001 †	<0,0001	<0,0001 †
Cabeça e pescoço	HNSC	Carcinoma de células escamosas de cabeça e pescoço	1,19	<0,0001 †	<0,0001	<0,0001 †
Pulmão	LUAD	Adenocarcinoma de pulmão	1,18	<0,0001 †	<0,0001	<0,0001 †
Rim	KIRP	Carcinoma papilar renal	1,18	<0,0001 †	<0,0001	<0,0001 †
Prostata	PRAD	Adenocarcinoma de próstata	1,17	<0,0001 †	<0,0001	<0,0001 †
Pele	SKCM	Melanoma cutâneo	1,02	0,9643 *	0,9643	1 *
Pâncreas	PAAD	Adenocarcinoma pancreático ductal	1,01	0,8955 *	0,9344	0,8742 *
Rim	KIRC	Carcinoma renal de células claras	1,01	0,2646	0,3024	0,0141 †
Adrenal	PCPG	Feocromocitoma & Paraganglioma	0,81	0,0517 *	0,0689	0,0273 *

NOTA: Indicações de \* (asterisco) denotam tumores com pequeno tamanho amostral ( $\leq 10$ ) para o tecido normal adjacente e de † (cruz) *p-value* significativo ( $< 0,01$ ). Correção dos valores de *p* em FDR (*false discovery ratio*) pelo método de Benjamini-Hochberg<sup>142</sup>. *Ks*: *p-value* para o teste de ranqueamento Kolmogorov-Smirnov.

Para eliminar-se a questão da arbitrariedade na definição de *z-score* e separação entre grupos, escolheu-se aquele que gerava uma separação com o menor *p-value* (segundo teste *log-rank*) para risco de recidiva da doença. Distintas implementações desta estratégia implicam na separação em dois ou mais grupos de pacientes<sup>137,138</sup>, optando-se aqui na separação em apenas dois. Paralelamente, foi calculada a proporção de risco de Cox (Cox Hazard Ratio ou apenas Cox HR) entre os grupos definidos, para observar se os pacientes classificados com alta expressão de *ELAVLI* apresentavam piora ou melhora do prognóstico, quanto a recidiva. Desta forma, identificamos tumores onde a alta expressão de *ELAVLI* (para o *z-score* definido

pelo menor *p-value*) traria um melhor (Tabela 3) ou pior (Tabela 4) prognóstico para o paciente para todos os tipos tumorais disponíveis no TCGA.

**Tabela 3 – Minimização de *p-value* para teste de razão de verossimilhança quanto ao nível de expressão de *ELAVL1* e sua relação com a redução do risco de recidiva tumoral**

Tipo	Código	Corte <i>z-score</i>	Cox HR (para alta expressão)					Pacientes baixa	
			HR	CI95 Min	CI95 Max	<i>p-value</i>	FDR		
Adenocarcinoma pancreático ductal	PAAD	0,96	0,17	0,04	0,72	<b>0,0015</b>	0,0197	105	83%
Carcinoma renal cromóforo	KICH	-0,42	0,11	0,02	0,55	<b>0,0031</b>	0,0197	19	32%
Melanoma uveal	UVM	0,98	0,13	0,02	0,93	<b>0,0045</b>	0,0197	60	82%
Adenocarcinoma de cólon	COAD	-0,79	0,19	0,05	0,72	<b>0,0169</b>	0,0366	76	19%
Câncer de endométrio	UCEC	-1,03	0,43	0,22	0,81	<b>0,0171</b>	0,0366	63	13%
Carc. células esc. de pulmão	LUSC	-0,92	0,39	0,18	0,84	<b>0,0237</b>	0,0474	75	19%
Carcinoma papilar de tireóide	THCA	-1,46	0,31	0,11	0,88	0,0572	0,0903	19	4%
Adenocarcinoma de estômago	STAD	1,46	0,36	0,05	2,65	0,2370	0,2539	257	92%
Câncer cervical	CESC	-1,53	0,62	0,15	2,58	0,5369	0,5369	9	9%

NOTA: *p-value* significativos a 0,05 estão em negrito. Tumores com *p* significativo < 0,01 estão com fundo cinza. Correção dos valores de *p* em FDR (*false discovery ratio*) pelo método de Benjamini-Hochberg<sup>142</sup>

Foram observados 9 tipos tumorais em que o prognóstico dos pacientes com alta expressão de *ELAVL1* é positivo (Tabela 3). Observamos três cenários em que há redução abaixo do limiar de *p-value* de 0,01 (ou empatados no menor valor de FDR, ainda significativo a 0,05), de seis cenários abaixo do limiar de 0,05 (que também retém FDR significativo). Para os tipos THCA, STAD e CESC, o menor *p-value* obtido não é significativo.

Dentre os três significativos a *p-value* < 0,01, em câncer de pâncreas (PAAD, Figura 13A, superior), observamos que até o valor de *z-score* (eixo x) 0,5, o Cox HR (linha azul cheia) se mantém próximo a 1 (separações sem impacto prognóstico). No menor *p-value* obtido (representado pela linha vertical verde), atingimos um Cox HR de 0,17 (5,9 vezes menor, 1/0,17) para os pacientes com alta expressão de *ELAVL1* (Figura 13B-C, superior). Cabe ressaltar que pâncreas não apresenta uma alteração significativa dos níveis de *ELAVL1* no tecido tumoral em relação ao normal adjacente (Figura 12A-B), dispondo de poucas amostras normais.

Em carcinoma renal cromóforo (KICH), um dos dez tumores raros do TCGA e de menor disponibilidade de dados na literatura, a minimização de *p-value* (Figura 13A, centro) gera um valor de Cox HR abaixo de 1, em praticamente toda a extensão dos cortes avaliados em função de *z-score*, sendo observado no *z-score* de -0,42 o *p-value* mínimo (0,0031), estando

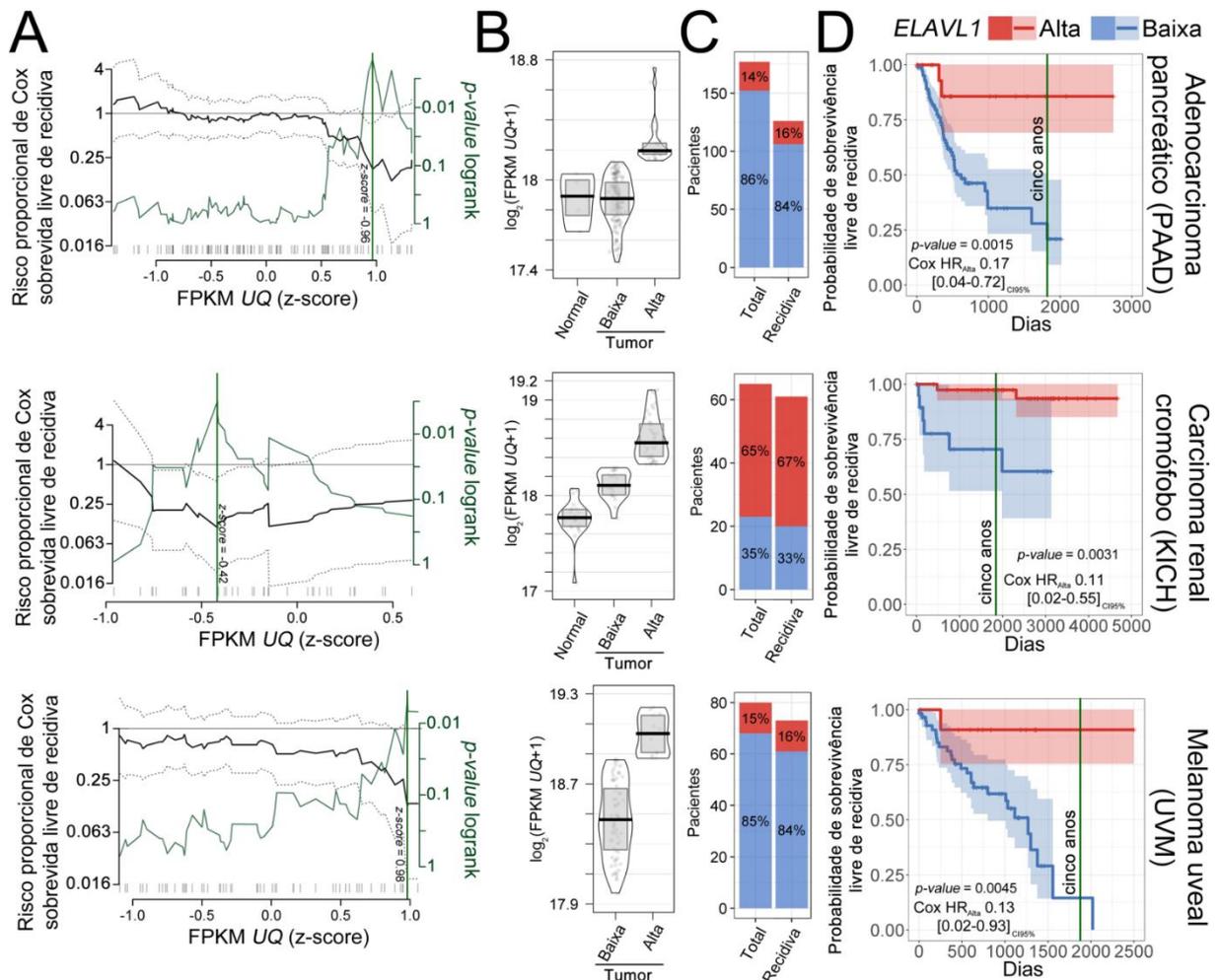
associado a este valor um Cox HR de 0,11. Assim, o grupo de pacientes com alta expressão de *ELAVL1*, 67% dos pacientes com informação de recidiva disponível (Figura 13B-C, centro), tem 9,1 (1/0,11) vezes menos probabilidade de apresentarem retorno da doença.

**Tabela 4 – Minimização de *p-value* para teste de razão de verossimilhança quanto ao nível de expressão de *ELAVL1* e sua relação com o aumento do risco de recidiva tumoral**

Tipo	Código	Corte z-score	Cox HR (para alta expressão)				<i>p-value</i>	FDR	Pacientes baixa	
			HR	CI95 Min	CI95 Max					
Câncer de mama	BRCA	-0,23	1,82	1,24	2,68	<b>0,0017</b>	0,0197	490	48%	
Carc. células esc. de cabeça e pescoço	HNSC	-0,78	4,32	1,34	13,92	<b>0,0024</b>	0,0197	73	22%	
Carcinoma hepatocelular	LIHC	0,26	1,64	1,19	2,26	<b>0,0034</b>	0,0197	237	74%	
Feocromocitoma & Paraganglioma	PCPG	0,14	5,20	1,45	18,66	<b>0,0046</b>	0,0197	107	62%	
Adenocarcinoma de pulmão	LUAD	-0,86	6,52	0,88	48,35	<b>0,0130</b>	0,0366	82	21%	
Adenocarcinoma de próstata	PRAD	-0,10	1,85	1,13	3,04	<b>0,0131</b>	0,0366	260	53%	
Sarcoma	SARC	0,20	1,60	1,11	2,31	<b>0,0141</b>	0,0366	157	69%	
Carcinoma papilar renal	KIRP	0,82	3,90	1,41	10,73	<b>0,0155</b>	0,0366	218	85%	
Melanoma cutâneo	SKCM	-0,81	6,49	0,86	49,08	<b>0,0164</b>	0,0366	17	20%	
Câncer de células germinativas de testículo	TGCT	-0,60	2,76	0,97	7,83	<b>0,0311</b>	0,0583	32	24%	
Glioma de baixo grau	LGG	0,62	1,92	1,08	3,43	<b>0,0379</b>	0,0669	354	85%	
Câncer de esôfago	ESCA	-0,45	1,73	1,01	2,98	<b>0,0411</b>	0,0685	42	32%	
Colangiocarcinoma	CHOL	-0,46	2,61	0,86	7,89	0,0663	0,0995	11	33%	
Cancer de bexiga	BLCA	1,81	1,98	1,00	3,89	0,0731	0,1044	313	96%	
Carcinoma Adrenocortical	ACC	-0,54	1,78	0,84	3,78	0,1162	0,1585	22	29%	
Timoma	THYM	0,80	2,09	0,82	5,35	0,1389	0,1812	84	75%	
Mesotelioma	MESO	-0,78	1,56	0,82	2,95	0,1658	0,2073	15	29%	
Carcinoma renal de células claras	KIRC	0,77	6,53	0,41	104,60	0,2133	0,2509	328	87%	
Linfoma difuso de grandes células B	DLBC	-0,28	3,47	0,39	31,21	0,2188	0,2509	16	36%	
Carcinossarcoma uterino	UCS	-0,87	1,75	0,67	4,60	0,2258	0,2509	9	20%	
Cistadenocarcinomas serosos do ovário	OV	1,13	4,73	0,43	52,23	0,2586	0,2675	283	90%	

NOTA: *p-value* significativos a 0,05 estão em negrito. Tumores com *p* significativo a 0,01 estão com fundo cinza. Correção dos valores de *p* em FDR (*false discovery ratio*) pelo método de Benjamini-Hochberg<sup>142</sup>

O terceiro tumor com prognóstico positivo para a elevada expressão de *ELAVL1* e *p-value* significativo (0,0045) foi o melanoma uveal, tal qual o anterior, raro. De maneira análoga ao carcinoma renal cromóforo, ao longo dos diferentes *z-score* testados, verificou-se um Cox HR inferior a 1 (Figura 13A, inferior), sendo em 0,98 obtida a maior significância – neste ponto 16% dos pacientes com dados de recidiva disponíveis se enquadraram no grupo de alta expressão de *ELAVL1* (Figura 13B-C, inferior), com Cox HR que indica 7,7 (1/0,13) vezes menos chance de recidiva para os mesmos (Figura 13D, inferior).



**Figura 13 – Minimização de *p-value* para teste de razão de verossimilhança quanto ao nível de expressão de *ELAVL1* e sua relação com a redução do risco de recidiva tumoral**

(A) Gráfico de linha indicando os valores de *z-score* (eixo x) utilizados na separação entre alta e baixa expressão de *ELAVL1* e sua consequência no risco proporcional de Cox (eixo y da esquerda, preto) e no *p-value* do teste *log-rank* (eixo y da direita, verde). Ambas ordenadas foram produzidas em escala logarítmica e os valores transcritos em seus correspondentes em escala linear para facilitar a compreensão numérica do eixo. Linha pontilhada horizontal azul indica o valor de Cox HR 1, onde não há distinção de prognóstico. Os *p-values* estão plotados em ordem inversa (de menores para maiores valores). Linhas tracejadas pretas representam os limites do intervalo de confiança de 95% da estimativa do risco proporcional de Cox. Linhas verticais cinzas na porção inferior do gráfico indicam a posição do paciente dado o valor de expressão do gene. Linha vertical verde indica o valor de separação que gerou o menor *p-value*. Os limites do eixo x foram definidos entre os grupos formados, omitindo da representação apenas os pacientes abaixo ou acima dos valores limítrofes de corte definido (cada grupo com ao menos 4% ou 10 pacientes, o maior valor).

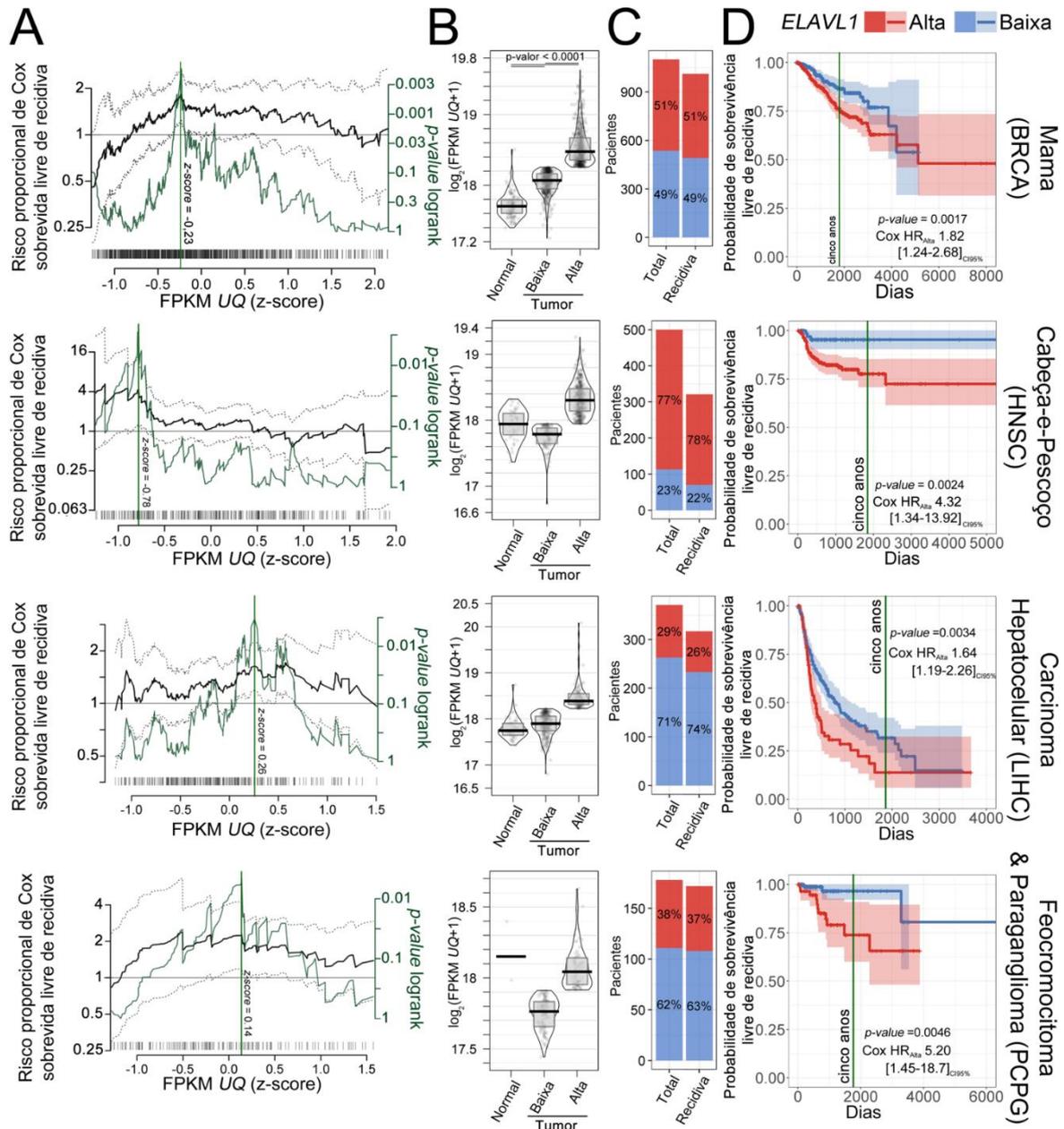
(B) Gráficos RDI dos níveis de expressão de *ELAVL1* em tecidos normais (Normal), e tumores primários (Tumor) separados entre pacientes classificados com baixa e alta expressão de *ELAVL1*. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados como círculos.

(C) Quantidade dos pacientes classificados em cada categoria utilizando o menor *p-value* obtido, do grupo completo e daqueles sobre os quais os dados de recidiva estavam disponíveis.

(D) Curva de sobrevivência para a separação pelo *z-score* definido, sendo a linha vermelha os pacientes com alta e a linha azul com baixa expressão. Linha vertical verde indica o limite de 5 anos de sobrevida. Valores indicam o risco proporcional de Cox (Cox HR) e o valor do teste *log-rank*. Área sombreada representa, em cada respectiva cor, o intervalo de confiança da curva de sobrevivência. Cada linha da figura representa um tumor (PAAD, KICH ou UVM) indicado na margem direita da mesma.

Foram encontrados 21 tumores nos quais o *z-score* com o menor *p-value* implica em prognóstico negativo para pacientes de alta expressão. Destes, 4 tumores (câncer de mama, carcinoma de células escamosas de cabeça e pescoço, carcinoma hepatocelular e feocromocitoma & paraganglioma adrenal) tiveram *p-value* significativo (Tabela 4). Assim, dos 30 tumores analisados (após remoção de leucemias – LAML –, câncer retal – READ – e gliomas de alto grau – GBM por inconsistências na anotação de recidivas), 21 apresentaram prognóstico negativo (Cox HR > 1) contra 9 de prognóstico positivo (Cox HR < 1) para os pacientes de alta expressão de *ELAVL1*, relativo a parâmetro de risco de recidiva tumoral. Observando a significância de *p-value* de 0,05, ainda temos 12 tumores com Cox HR superior a 1 contra apenas 6 tumores com Cox HR inferiores a este valor. Neste sentido, para 21 dos 30 tumores analisados, conclui-se que a alta expressão de *ELAVL1* está relacionada a pior prognóstico – mensurado pela probabilidade de recidiva após tratamento.

O tumor com *p-value* mais significativo entre os *p-values* mínimos obtidos na separação de grupos de alta e baixa expressão de *ELAVL1* por um dado *z-score*, e com Cox HR positivo, foi o de mama. Quanto à análise de minimização de *p-value*, cortes baixos impactaram em riscos relativos inferiores a 1 (melhor prognóstico, mas de forma não-significativa), que prontamente (após valores de -0,75) inverteram a predição e se mantiveram razoavelmente constantes pelo restante do espaço amostral (Figura 14). Como forma de validação estatística da observação para este tumor em específico, 51,87% de dez mil reamostragens (análogo a dez mil escolhas de subgrupos de validação) realizadas utilizando 40% do conjunto de pacientes apresentaram significância (considerando *p-value* < 0,05). A título de referência, para uma reamostragem de bernoulli (com as características indicadas) ser considerada significativa, precisa extrapolar o limite de 50,82%. Cabe mencionar que em tais reamostragens o valor de risco proporcional de Cox esteve acima de 1 em 99,51% dos eventos. Caso consideremos apenas intervalos de acompanhamento de cinco e dez anos, obtivemos 58,37% e 58,58% de valores de *p* significativos, respectivamente. Com outra estratégia de correção do *p-value* obtido especificamente para a metodologia de minimização de *p-value*<sup>143</sup>, ainda obtém-se 0,0387. Tais validações se unem, evidentemente, à correção de *p-value* pelo método do *false discovery ratio*, já mostrada (Tabela 4).



**Figura 14 – Minimização de  $p\text{-value}$  para teste de razão de verossimilhança quanto ao nível de expressão de *ELAVL1* e sua relação com o aumento do risco de recidiva tumoral**

(A) Gráfico de linha indicando os valores de  $z\text{-score}$  (eixo x) utilizados na separação entre alta e baixa expressão de *ELAVL1* e sua consequência no risco proporcional de Cox (eixo y da esquerda, preto) e no  $p\text{-value}$  do teste log-rank (eixo y da direita, verde). Ambas ordenadas foram produzidas em escala logarítmica e os valores transcritos em seus correspondentes em escala linear para facilitar a compreensão numérica do eixo. Linha pontilhada horizontal azul indica o valor de Cox HR 1, onde não há distinção de prognóstico. Os  $p\text{-values}$  estão plotados em ordem inversa (de menores para maiores valores). Linhas tracejadas pretas representam os limites do intervalo de confiança de 95% da estimativa do risco proporcional de Cox. Linhas cinzas na porção inferior do gráfico indicam a posição do paciente dado o valor de expressão do gene. Linha vertical verde indica o valor de separação que gerou o menor  $p\text{-value}$ . Os limites do eixo x foram definidos entre os grupos formados, omitindo da representação apenas os pacientes abaixo ou acima dos valores limítrofes de corte definido (cada grupo com ao menos 4% ou 10 pacientes, o maior valor). (B) Gráficos RDI dos níveis de expressão de *ELAVL1* em tecidos normais (Normal), e tumores primários (Tumor) separados entre pacientes classificados com baixa e alta expressão de *ELAVL1*. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados como círculos. (C) Quantidade dos pacientes classificados em cada categoria utilizando o menor  $p\text{-value}$  obtido, do grupo completo e daqueles sobre os quais os dados de recidiva estavam disponíveis. (D) Curva de sobrevivência para a separação pelo  $z\text{-score}$  definido, sendo a linha vermelha os

pacientes com alta e a linha azul com baixa expressão. Linha vertical verde indica o limite de 5 anos de sobrevida. Valores indicam o risco proporcional de Cox (Cox HR) e o valor do teste *log-rank*. Área sombreada representa, em cada respectiva cor, o intervalo de confiança da curva de sobrevivência. Cada linha da figura representa um tumor (BRCA, HNSC, LIHC ou PCPG) indicado na margem direita da mesma.

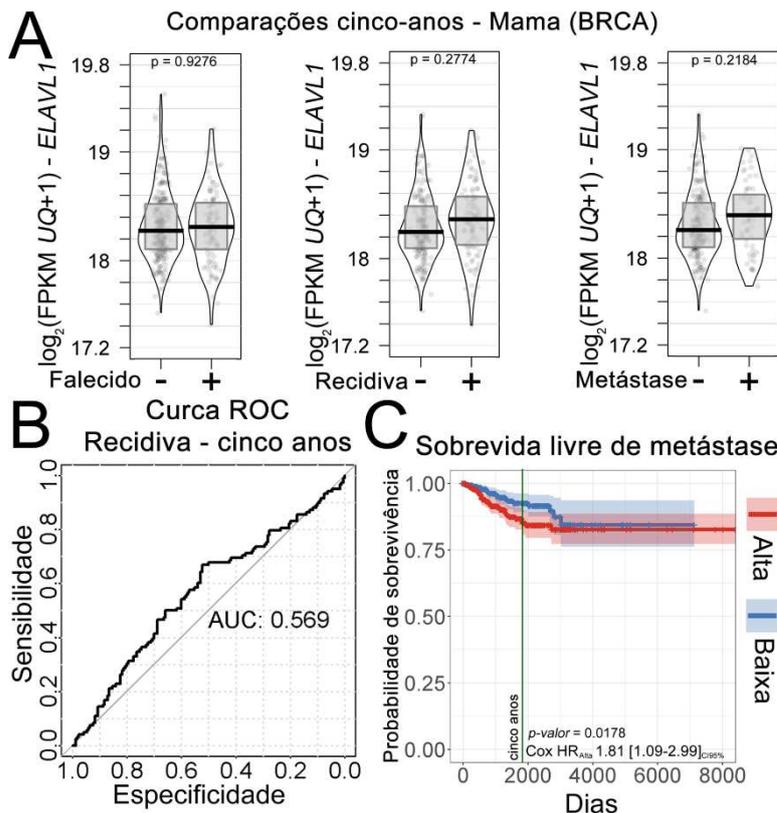
Para os tumores de mama, a separação proposta implicou em um aumento da mediana de expressão em mais de 70% (em escala linear), frente aos tecidos normais adjacentes disponíveis (Figura 14B), levando a divisão das amostras tumorais em duas porções iguais (Figura 14C). Para a análise de sobrevida livre de recidiva, as bandas do intervalo de confiança das curvas de sobrevivência (Figura 14D, área sombreada) fundem-se em apenas após o período de 5 anos de acompanhamento, o que – somado ao grande tamanho amostral – justifica a presença de um *p-value* significativo. Avaliando-se a recidiva em uma janela de cinco anos, o risco proporcional de Cox eleva-se para 2,00 (*logrank p-value* 0,0008) – frente a 1,82 utilizando todo o intervalo de acompanhamento disponível.

Os tumores de cabeça-e-pescoço aqui designados compreendem origens primárias na cavidade oral, faringe, laringe, sinos paranasais, cavidade nasal e glândulas salivares<sup>248</sup>. O menor *p-value* obtido (0,0024) foi relativo a *z-score* de -0,78, ponto no qual observou-se risco proporcional HR Cox de 4,32 (Figura 14A). Em suma, a alta expressão de *ELAVLI* neste contexto aumenta o risco de recidiva em 4,32 vezes (Figura 14D).

O terceiro tumor em ordem crescente de *p-value* é o carcinoma hepatocelular. Em todos os cortes de *z-score* avaliados, foram verificados valores de Cox HR superiores a 1, com mínimo *p-value* (0,0034) em *z-score* de 0,23 (Figura 14A). Tal corte produziu uma fração de 26% dos pacientes no grupo de alta expressão (Figura 14B-C) com risco proporcional de recidiva 1,19 vezes superior para o grupo de alta em relação ao grupo com baixa expressão de *ELAVLI*. O gráfico de Kaplan-Meier mostra que, até o período de 5 anos pós-incidência, há a clara separação entre as curvas de sobrevida (e até de fração de suas bandas de confiança), as quais, após este período, terminam por se unir (Figura 14D).

Por fim, o último tumor com *p-value* inferior a 0,01 para aumento do risco proporcional de recidiva relacionado a alta expressão de *ELAVLI* é o raro tumor de glândulas adrenais, feocromocitomas e paragangliomas<sup>249</sup>. O Cox HR se mantém acima do valor de 1 em todos os cortes de *z-score* avaliados, com *p-value* mínimo (0,0046) em *z-score* de 0,14. Tal corte impacta em 37% dos pacientes alocados no grupo de alta expressão e *ELAVLI* (Figura 14B-C), que apresentam risco superior de recidiva em 1,45 vezes em relação aos de baixa expressão, com a intersecção entre as bandas de confiança das curvas de sobrevivência ocorrendo apenas após 5 anos (Figura 14D).

Sendo os tumores de mama detentores do mais significativo *p-value* para pior prognóstico segundo a análise univariada *pan-câncer* realizada (Tabela 4), os parâmetros clínicos foram mais profundamente explorados (Figura 15 e Tabela 5). Realizando comparações binárias após 5 anos da detecção do tumor dentro de parâmetros preditivos de sobrevida (falecimento ou não, recidiva tumoral ou não e presença de metástase ou não), não observamos nenhuma diferença significativa no nível de expressão de *ELAVL1* (Figura 15A). A não significativa diferença de expressão entre os grupos mostra-se na curva ROC (*Receiver operating characteristic*) para recidiva, com área de 0,569, brevemente superior ao esperado para um classificador linear randômico, 0,5 (Figura 15B). Quando observada a classificação binária para a sobrevida livre de metástase distante (Figura 15C), entretanto, não só foi obtida a significância, quanto o mínimo do intervalo de confiança da estimativa de Cox HR também permaneceu acima de 1, confirmando um prognóstico negativo para o paciente relacionado a alta expressão de *ELAVL1*.



**Figura 15 – Avaliação dos elevados níveis de expressão de *ELAVL1* em pacientes de câncer de mama quanto a sobrevida, recidiva e metástase**

(A) Gráficos RDI dos níveis de expressão de *ELAVL1* em tumores primários de pacientes de acordo com o falecimento (esquerda), recidiva (centro) ou metástase distante em 5 anos após diagnóstico inicial. *P-values* oriundos de testes-t. (B) Curva de *receiver operating characteristic* (ROC) para o nível de expressão de *ELAVL1* e a ocorrência de recidiva em uma janela de 5 anos. A linha cinza indica preditor completamente randômico e não-relacionado com o evento. AUC: área sob a curva. (C) Curva de sobrevivência para o melhor *p-value* obtido em recidiva quanto a sobrevida livre de metástase, sendo a linha vermelha os pacientes com alta e a linha azul com baixa expressão. Linha vertical verde indica o limite de

5 anos de sobrevida. Valores indicam o risco proporcional de Cox (Cox HR) e o valor do teste *log-rank*. Área sombreada representa, em cada respectiva cor, o intervalo de confiança da curva de sobrevivência. Para os gráficos RDI, barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados sobre os dados como círculos.

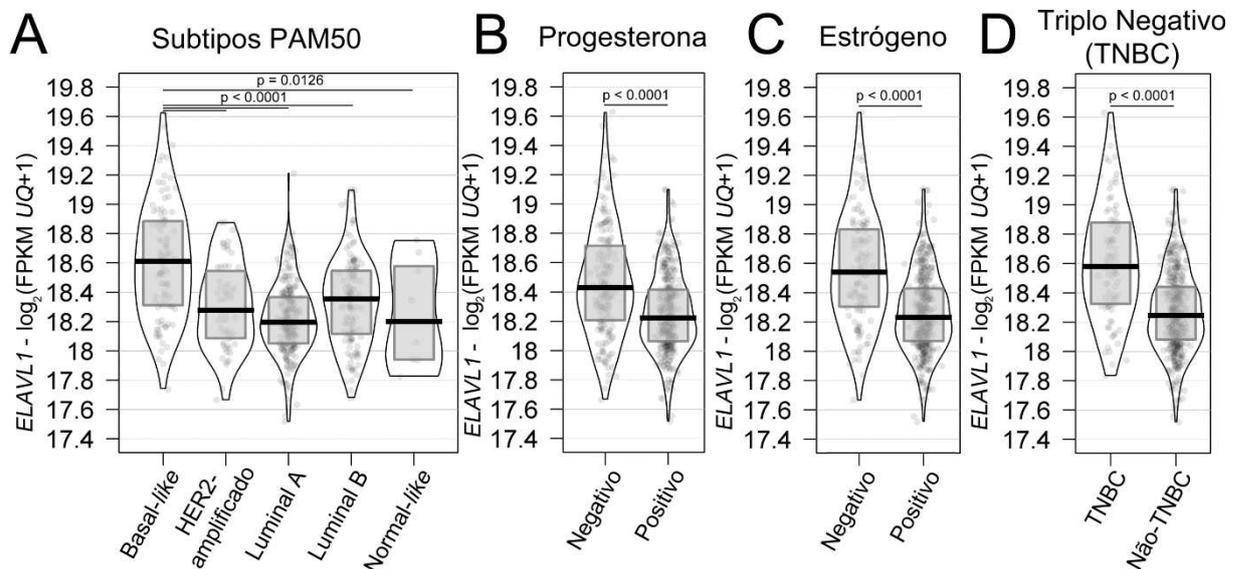
Avaliou-se, também, a relação entre parâmetros clinicopatológicos (coletados da época do diagnóstico) em câncer de mama e os grupos de expressão de *ELAVLI* (Tabela 5). O único parâmetro com significância quando avaliada a ausência de homogeneidade é o estadiamento T, com um aumento proporcional da categoria T2 – de tumores com mais de 20 mm e menos de 50 mm em sua maior dimensão. Aproveita-se para mencionar, que o câncer de mama é uma patologia predominantemente feminina, mas a glândula mamária masculina – mesmo que atrofiada – também pode ser acometida por tumores<sup>250</sup> os quais representam 12 pacientes do TCGA, sendo 75% dos mesmos associados a alta expressão de *ELAVLI* (mas de maneira não significativa pelo baixo tamanho amostral). Por fim, vale comentar que, apesar de em câncer de mama haver um consenso de que elevados níveis citoplasmáticos de HuR impactam negativamente no prognóstico de parâmetros como sobrevida e recidiva da doença (Tabela 1), nós concluímos que neste tumor, altos níveis totais de expressão de *ELAVLI* podem também potencialmente serem usados como fator prognóstico para sobrevida livre de recidiva e livre de metástase.

**Tabela 5 – Associações entre a expressão de *ELAVLI* e parâmetros clinicopatológicos nos tumores de mama do TCGA**

Característica	Expressão de <i>ELAVLI</i>		<i>p-value</i>
	Alta (%)	Baixa (%)	
Estadiamento histológico	553 (51,3)	525 (48,7)	0,0898
I	81 (7,5)	101 (9,4)	
II	341 (31,6)	286 (26,5)	
III	121 (11,2)	128 (11,9)	
IV	10 (0,9)	10 (0,9)	
Estadiamento T	562 (51,2)	536 (48,8)	<b>0,0005</b>
T1	116 (10,6)	164 (14,9)	
T2	359 (32,7)	280 (25,5)	
T3	68 (6,2)	71 (6,5)	
T4	19 (1,7)	21 (1,9)	
Estadiamento N	556 (51,4)	525 (48,6)	0,3065
N0	268 (24,8)	253 (23,4)	
N1	189 (17,5)	175 (16,2)	
N2	67 (6,2)	53 (4,9)	
N3	32 (3)	44 (4,1)	
Estadiamento M	491 (52,5)	445 (47,5)	1
M0	479 (51,2)	435 (46,5)	
M1	12 (1,3)	10 (1,1)	
Sexo	565 (51,3)	536 (48,7)	0,1738
Feminino	556 (50,5)	533 (48,4)	
Masculino	9 (0,8)	3 (0,3)	

NOTA: Valores de *p*, derivados de  $\chi^2$ , significativos estão em negrito. Para características com subníveis (por exemplo, T1a, T1b e T1c, os valores foram colapsados).

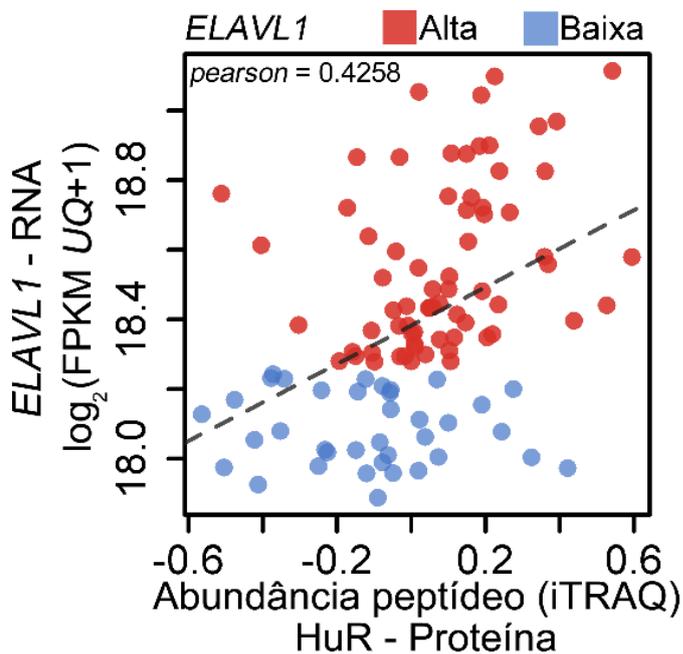
Para a compreensão da distribuição da expressão de *ELAVL1* dentro dos subtipos moleculares clinicamente avaliados, confrontamos a abundância do gene com uma classificação que utiliza um painel gênico capaz de prever tipos moleculares de câncer de mama associados com o prognóstico metastático (*Prediction Analysis of Microarray 50*, PAM50)<sup>251</sup> (Figura 16A). Dentro das categorias, os subtipos Luminal B, HER-amplificado e Basal-like apresentam os piores prognósticos<sup>251</sup>, sendo a expressão de *ELAVL1* aumentada significativamente no último subgrupo. Também são elevados em pacientes com reduzidos níveis do receptor de progesterona (Figura 16B), estrógeno (Figura 16C) e negativos para ambos simultaneamente com a ausência de amplificação de *ERBB2* (Figura 16D), tumores estes designados triplo-negativos e de conhecido pior prognóstico para os pacientes acometidos<sup>252</sup>.



**Figura 16 – Observância da expressão de *ELAVL1* em pacientes com câncer de mama segundo a classificação molecular clinicamente utilizada**

Gráficos RDI da separação dos pacientes de câncer de mama do TCGA segundo Koboldt *et al*<sup>253</sup> para o painel gênico PAM50 (A), presença do receptor de progesterona (B), de estrógeno (C), ou na ausência de ambos e não amplificação de *ERBB2*, tumores designados triplo-negativos (TNBC). Para os gráficos RDI, barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados sobre os dados como círculos. Os *p-values* indicados foram gerados após uma análise de variância seguida de um teste de Tukey para as médias.

Avaliou-se, por fim, se a abundância de mRNA implicava em alteração nos níveis proteicos de HuR utilizando a fração de amostras de tumores de mama com dados de mRNA e massas disponíveis (Figura 17). Os resultados mostraram um coeficiente de correlação de Pearson positivo de 0,4258 entre os níveis normalizados do mRNA de *ELAVL1* e a abundância dos peptídeos de HuR detectados, implicando que os níveis de mRNA medidos refletem a abundância proteica.



**Figura 17 – Correlação entre a abundância do mRNA de *ELAVL1* e seu produto HuR em tumores de mama**

Gráfico de dispersão indicando entre a abundância de peptídeos não-únicos de HuR nas amostras disponíveis no TCGA e seus respectivos níveis de mRNA. Cores dos pontos simbolizam a classificação dos pacientes de acordo com o nível de mRNA.

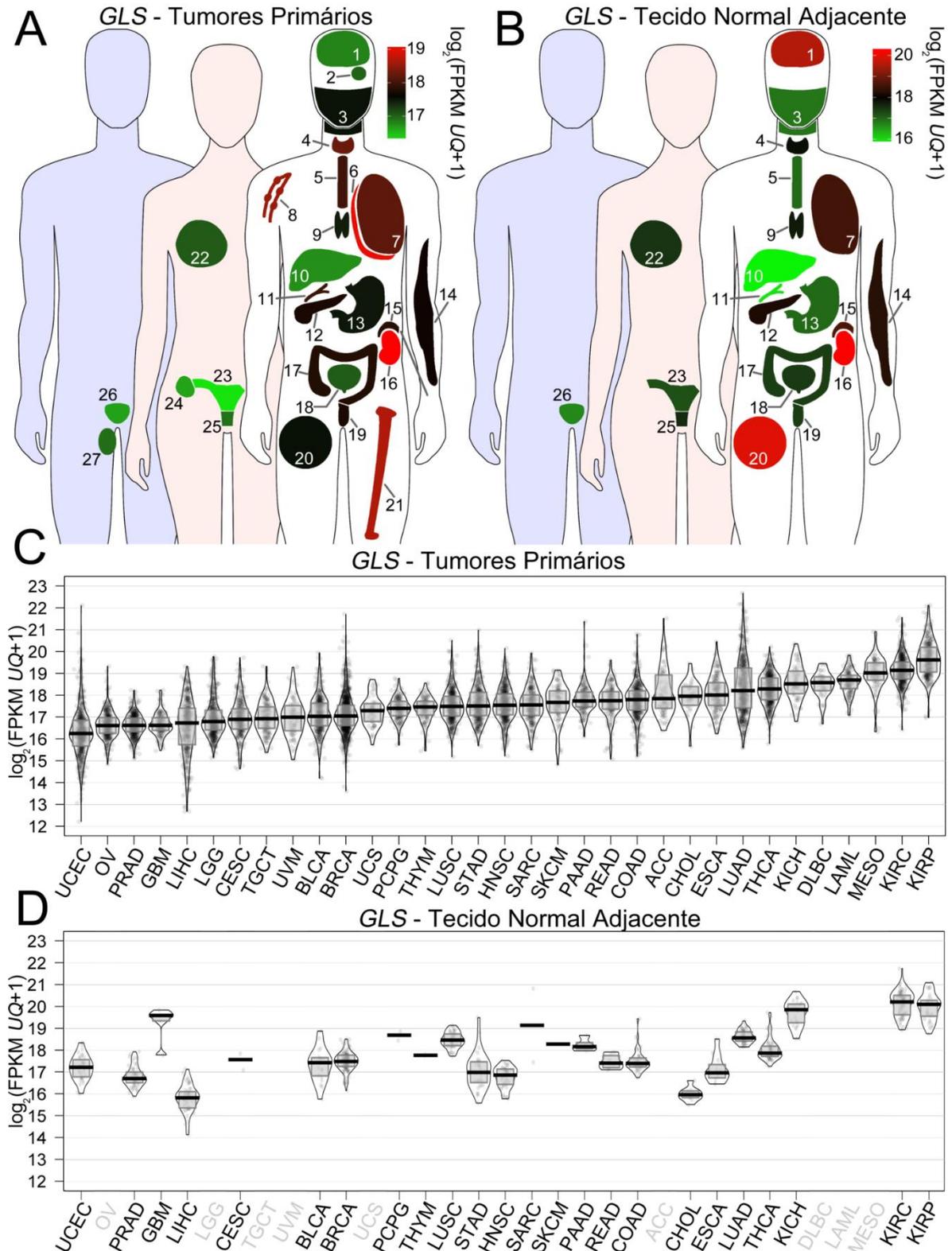
#### 4.1.2 Expressão de *GLS* e suas isoformas em tecidos tumorais

Os níveis totais de expressão de *GLS* e do mRNA de suas isoformas, KGA e GAC, foi avaliado na mesma coorte de pacientes, para a avaliação da alteração entre o tecido normal e tumoral e correlação com a expressão de *ELAVL1*. Inicialmente, utilizando o nível de expressão total do gene, uma nova análise exploratória topográfica permitiu avaliar o padrão de expressão do gene *GLS* em diversos tecidos (Figura 18).

O rim, órgão que inclusive nomeia uma das isoformas (KGA – *kidney-type glutaminase*), aparece como um dos órgãos de expressão mais abundante de *GLS*, seja na situação patológica (Figura 18A) ou normal (Figura 18B). No lado oposto, observamos órgãos relacionados ao sistema genito-urinário e o fígado – órgão que nomeia isoforma do gene parálogo *GLS2* (LGA - *liver-type glutaminase*). A amplitude de variação da mediana de expressão de *GLS* em tumores primários vai de 16,5 a 19,5, cerca de 3 unidades de log<sub>2</sub> ou 8 vezes em escala linear (Figura 18C).

Tal qual *ELAVL1*, a expressão de *GLS* também está acima da mediana de expressão do genoma e apresenta uma variação inter-tumoral ainda superior. Intra-tumoralmente também existe uma variação de grande magnitude. Mama (BRCA), por exemplo, que tem um valor intermediário de expressão, contempla uma variação de até 128 vezes (em escala linear) entre seus pacientes limítrofes. Para as amostras normais, a variação das medianas ultrapassa 4 unidades de log<sub>2</sub>, implicando em uma variação de até 16 vezes entre a menor (fígado, LIHC) e

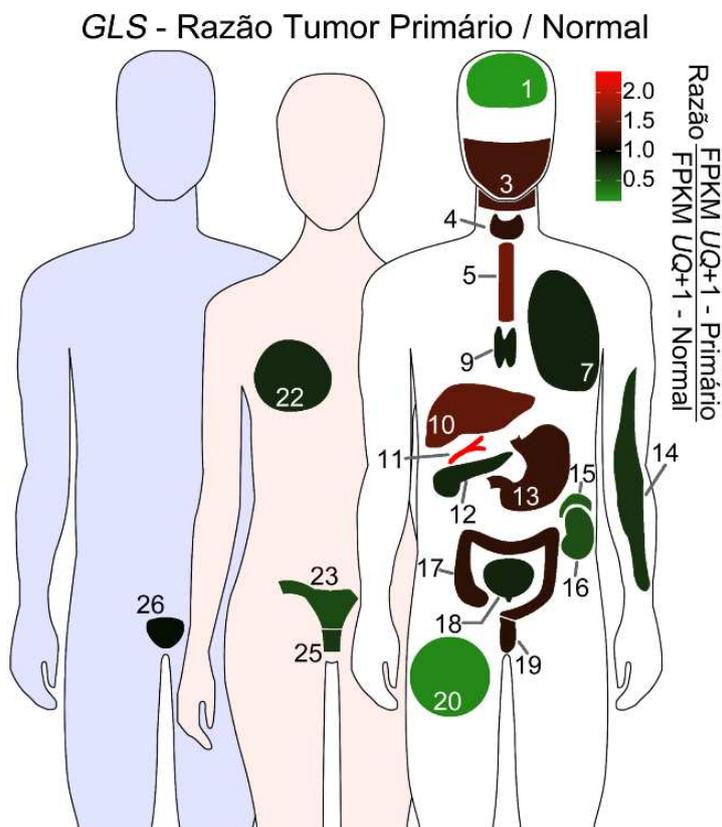
a maior (rim, KIRP+KIRC+KICH) expressão do gene, compatível com a utilização dos genes de forma tecido-específica.



**Figura 18 – Níveis de expressão do mRNA de *GLS* em tecidos normais e tumorais humanos** (A) Representação topográfica dos níveis de expressão de *GLS* em tumores primários. (B) Representação topográfica dos níveis de expressão de glutaminase (*GLS*) em tecidos normais adjacentes a tumores primários, sendo mostrado somente aqueles para os quais dados eram disponíveis. Os valores representados nas escalas de

cor representam a mediana dos níveis de abundância de mRNA mensurados em FPKM normalizado pelo quartil superior em escala logarítmica e estão dimensionados de forma independente entre os gráficos. (C) Gráfico RDI dos níveis de expressão de *GLS* em tumores primários, ordenados crescentemente de acordo com a mediana. (D) Gráfico RDI dos níveis de expressão de *GLS* em tecidos normais adjacentes a tumores primários, obedecendo a ordenação de A. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados sobre os dados como círculos. A numeração dos órgãos está presente no Quadro 7.

Comparando a razão da expressão entre tecidos tumorais e normais adjacentes, em 16 dos 24 tumores para os quais a comparação é possível, verifica-se uma redução na expressão global do gene em tumores (Figura 19). Por exemplo, os três tipos de tumores renais (KIRP, KIRC e KICH) apresentam intensa redução na expressão do mRNA total do gene quando comparado com tecido normal, tendo sido observado os valores 0,71, 0,47 e 0,4, respectivamente (Tabela 6). A expressão global de *GLS* aumentou somente em tumores de ducto biliar, esôfago, fígado, cabeça-e-pescoço e estômago (tireóideo, cólon e reto de forma não significativa), quando comparado aos tecidos normais.



**Figura 19 – Razão entre os níveis de expressão de mRNA de *GLS* entre tecido tumoral e tecido normal adjacente**

Representação topográfica da razão entre a mediana de expressão de *ELAVL1* em tecido tumoral sobre o tecido normal adjacente denotada pela fração dos valores de FPKM normalizados pelo quartil superior em escala linear. A numeração dos órgãos está presente no Quadro 7.

**Tabela 6 – Lista com os valores absolutos das razões das medianas de *GLS* entre tecido tumoral e normal representadas na Figura 19 com o órgão afetado e o respectivo código empregado pelo TCGA**

Tecido	Código TCGA	Tipo	Razão	<i>p-value</i>	FDR	<i>ks</i>
Ducto biliar	CHOL	Colangiocarcinoma	3,99	< 0,0001 *†	< 0,0001	<0,0001 *†
Esôfago	ESCA	Câncer de esôfago	2,06	0,0007 *†	0,0015	0,0004 *†
Fígado	LIHC	Carcinoma hepatocelular	1,87	< 0,0001 †	< 0,0001	<0,0001 †
Cabeça e pescoço	HNSC	Carcinoma de células escamosas de cabeça e pescoço	1,60	< 0,0001 †	< 0,0001	<0,0001 †
Estômago	STAD	Adenocarcinoma de estômago	1,44	0,0002 †	0,0005	0,0056 †
Tireóide	THCA	Carcinoma papilar de tireóide	1,33	0,0243	0,0486	0,0001
Intestino grosso	COAD	Adenocarcinoma de cólon	1,32	0,1169	0,1754	0,0001
Intestino delgado	READ	Adenocarcinoma de reto	1,27	0,5365 *	0,5853	0,0493 *
Prostata	PRAD	Adenocarcinoma de próstata	0,94	0,0723	0,1239	0,1663
Timo	THYM	Timoma	0,81	0,3231 *	0,4081	0,1366 *
Pulmão	LUAD	Adenocarcinoma de pulmão	0,78	0,2850	0,3800	<0,0001
Bexiga	BLCA	Cancer de bexiga	0,76	0,2734	0,3800	0,3060
Pâncreas	PAAD	Adenocarcinoma pancreático ductal	0,74	0,3515 *	0,4218	0,0365 *
Mama	BRCA	Câncer de mama	0,73	< 0,0001 †	< 0,0001	<0,0001 †
Rim	KIRP	Carcinoma papilar renal	0,71	0,1117	0,1754	0,0219
Pele	SKCM	Melanoma cutâneo	0,66	0,6916 *	0,7217	1,0000 *
Cérvice uterino	CESC	Câncer cervical	0,62	0,4865 *	0,5560	0,1369 *
Útero	UCEC	Câncer de endométrio	0,51	< 0,0001 †	< 0,0001	<0,0001 †
Pulmão	LUSC	Carcinoma de células escamosas de pulmão	0,51	< 0,0001 †	< 0,0001	<0,0001 †
Rim	KIRC	Carcinoma renal de células claras	0,47	0,8738 †	0,8738	<0,0001 †
Adrenal	PCPG	Feocromocitoma & Paraganglioma	0,41	0,0002 *	0,0005	<0,0001 *
Rim	KICH	Carcinoma renal cromóforo	0,40	< 0,0001 †	< 0,0001	<0,0001 †
Tecidos macios	SARC	Sarcoma	0,19	0,0301 *	0,0556	0,5095 *
Cérebro	GBM	Glioblastoma multiforme	0,13	< 0,0001 *†	< 0,0001	<0,0001 *†

NOTA: Indicações de \* (asterisco) denotam tumores com pequeno tamanho amostral ( $\leq 10$ ) para o tecido normal adjacente e de † (cruz) *p-value* significativo ( $< 0,01$ ). Correção dos valores de *p* em FDR (*false discovery ratio*) pelo método de Benjamini-Hochberg<sup>142</sup>. *Ks*: *p-value* para o teste de ranqueamento Kolmogorov-Smirnov.

Uma vez que já foi mostrada a diferença na expressão e atividade entre as isoformas de *GLS*<sup>26,27</sup>, decidiu-se avaliar a expressão das mesmas individualmente. Enquanto que para a abundância global dos genes utilizou-se a anotação do GENCODE<sup>157,158,254</sup>, para analisarmos as isoformas foi necessário realizar a conversão das posições genômicas e compreender a anotação utilizada pelo Genome Browser UCSC<sup>255</sup> (utilizada pelo TCGA para prover dados de expressão com resolução de isoformas, exons e junções exon-exon;) e a anotação RefSeq<sup>256</sup>. Uma vez que a anotação RefSeq possui somente as isoformas KGA e GAC, amplamente validadas experimentalmente e estudadas, enquanto que o UCSC apresenta sete, nós preterimos a segunda em relação a primeira. Essa abundância de isoformas sem validação experimental leva a modificação de exons com a inclusão de introns retidos e novos exons não validados.

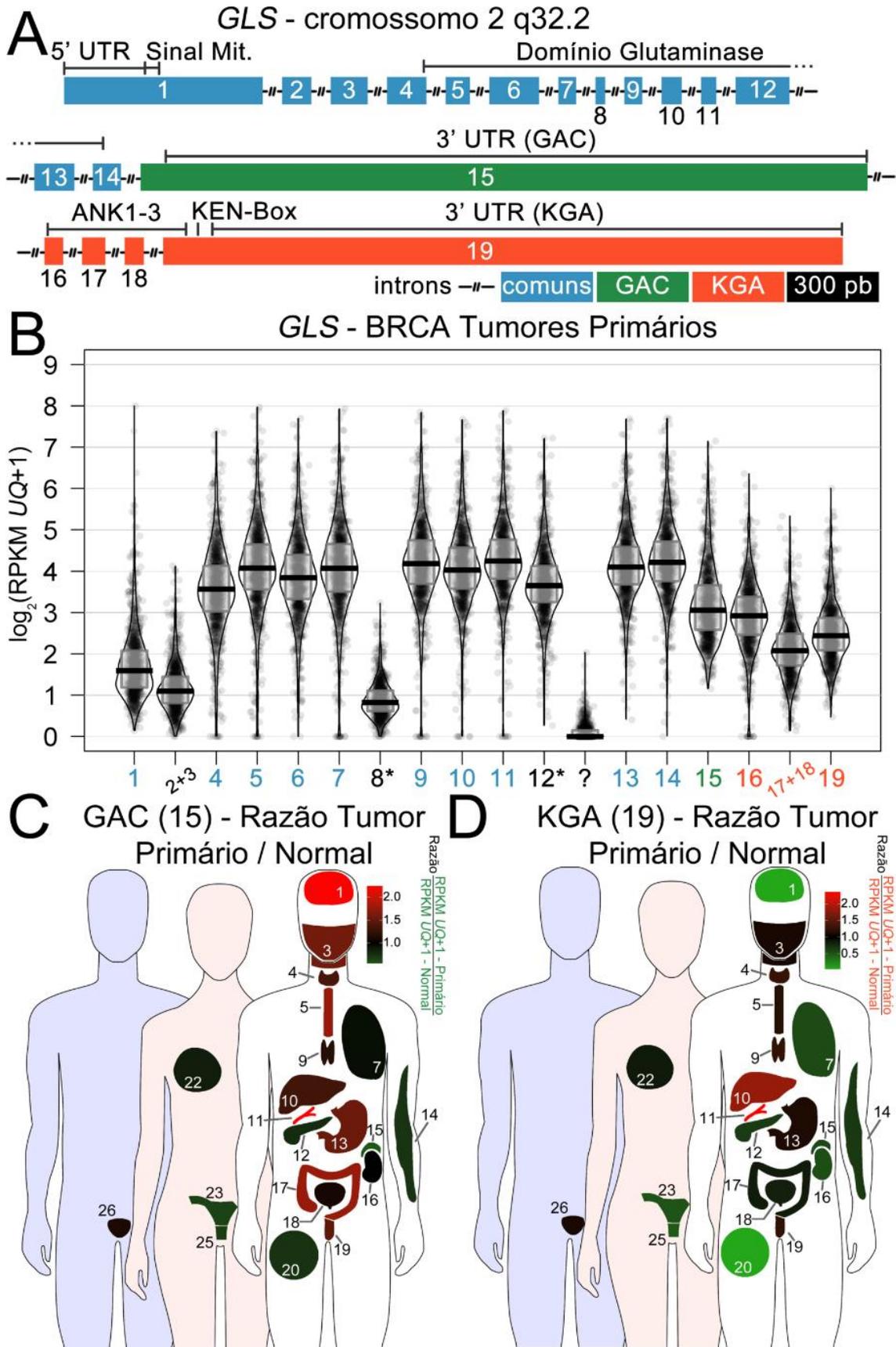
Pode-se observar na Tabela 7 a consequência de algumas destas diferenças: por exemplo, o exon 2 e 3 de RefSeq estão fusionados em um único exon com retenção de intron na anotação UCSC; em outro exemplo, vê-se que os exons 17 e 18 e os exons 8 e 12 de RefSeq estão apresentados com sítios doadores de *splicing* distintos em UCSC.

**Tabela 7 – Conversão entre as marcações exônicas do gene *GLS* entre as anotações RefSeq e UCSC**

RefSeq versão 85				UCSC Genes				Observação
Exon	Início	Fim	pb	Exon	Início	Fim	pb	
1	191745547	191746196	649	1	191745547	191746196	649	
2	191759878	191759974	96	2	191759878	191761227	1349	<i>UCSC une os exons</i>
3	191760330	191760451	121					
4	191765290	191765419	129	3	191765290	191765419	129	
5	191766673	191766752	79	4	191766673	191766752	79	
6	191769730	191769893	163	5	191769730	191769893	163	
7	191774989	191775047	58	6	191774989	191775047	58	
8	191785750	191785782	32	7	191783989	191785782	1793	<i>UCSC é maior</i>
9	191785871	191785929	58	8	191785871	191785929	58	
10	191788643	191788709	66	9	191788643	191788709	66	
11	191789269	191789319	50	10	191789269	191789319	50	
12	191792032	191792208	176	11	191791996	191792208	212	<i>UCSC é maior</i>
	Sem equivalência			12	191792416	191792456	40	
13	191795163	191795294	131	13	191795163	191795294	131	
14	191796271	191796363	92	14	191796271	191796363	92	
15	191797442	191799817	2375	15	191797442	191799817	2375	<i>Exon exclusivo GAC</i>
16	191818291	191818352	61	16	191818291	191818352	61	<i>Exon exclusivo KGA</i>
17	191819310	191819386	76	17	191819310	191819544	234	<i>UCSC une os exons</i>
18	191819481	191819544	63					
19	191827556	191829776	2220	18	191827556	191829776	2220	<i>Exon exclusivo KGA</i>

Nota: Posições referem-se à montagem GRCh37/hg19 do genoma humano.

Como inicialmente descrito (Figura 3), as duas isoformas de *GLS* variam exclusivamente em seus C-terminais, com a KGA mais longa e incluindo domínios anquirina e de regulação KEN-Box e a GAC mais curta e sem os respectivos domínios. Observando a estrutura genômica que gera esta alteração, temos 14 primeiros exons comuns às duas isoformas, o exon 15 exclusivo à GAC (que contém um códon de parada, porção 3'UTR única e gera a terminação do mRNA através da adição de uma cauda poli-A ao mesmo), e os exons 16, 17, 18 e 19 exclusivos à KGA (sendo o último também com seu códon de parada, porção 3'UTR e terminação de mRNA exclusivos), como representado no esquema da Figura 20A. Considerando as peculiaridades e possíveis inadequações da anotação do UCSC, observamos a variação dos níveis de abundância individuais dos exons de glutaminase em amostras de tumores de mama (Figura 20B).



**Figura 20 – Deconvolução do nível de expressão do mRNA das isoformas KGA e GAC em tecidos normais e tumorais humanos**

(A) Representação esquemática dos exons do gene *GLS*, com indicação das principais regiões. Os exons comuns a ambas isoformas estão indicados em azul, exclusivo a GAC em verde e exclusivos a KGA em laranja. Os introns

encontram-se fora de escala. Com exceção às regiões terminais não traduzidas (UTRs), as demais indicações referem-se a domínios proteicos. (B) Gráfico RDI dos níveis de expressão de cada exon individual de *GLS* em amostras derivadas de tecido tumoral de mama. A nomenclatura dos exons refere-se à tradução para a anotação RefSeq (Consultar Tabela 7). Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados sobre os dados como círculos. Asteriscos (\*) indicam exons que não são fusões, mas diferem entre as anotações. (C) Representação topográfica da razão da expressão do exon 15 de *GLS* (exclusivo à GAC) em tumores primários sobre tecido normal adjacente. (D) Representação topográfica da razão da expressão do exon 19 de *GLS* (exclusivo à KGA) em tumores primários sobre tecido normal adjacente. Os valores representados nas escalas de cor representam a razão linear das mediana dos níveis de abundância de mRNA mensurados em RPKM normalizados pelo quartil superior e estão dimensionados de forma independente entre os gráficos. A numeração dos órgãos está presente no Quadro 7.

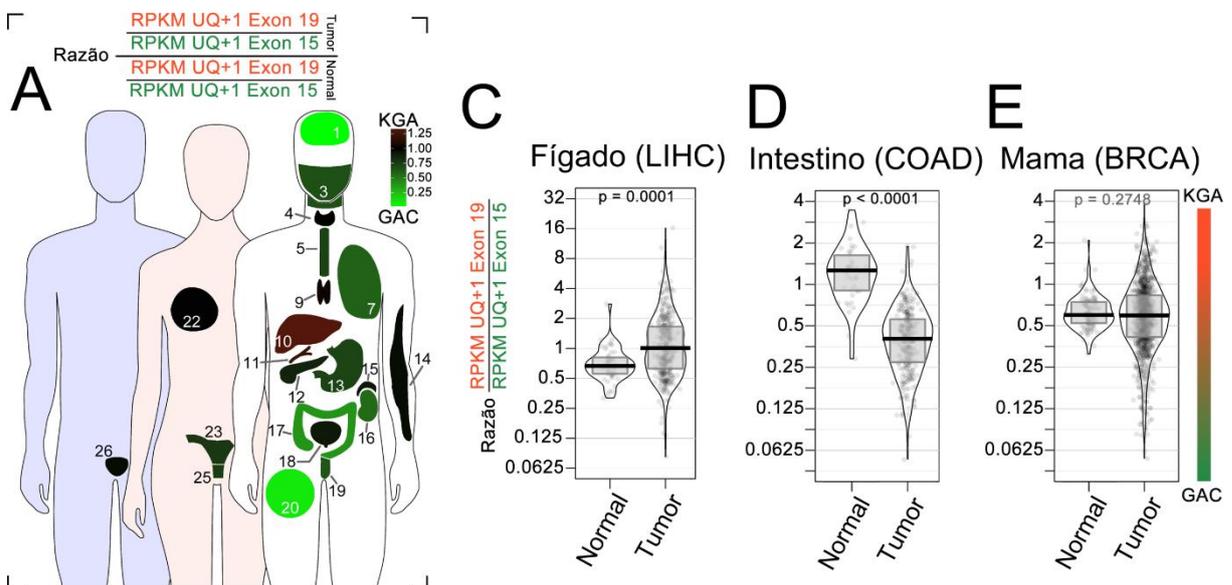
A menor abundância observada para o exon 1 pode ser justificada como uma inadequação da identificação do início de transcrição exato para o gene, o que aumenta o tamanho teórico do fragmento – impactando na normalização da unidade utilizada (RPKM, *Reads Per Kilobase Million*, leituras por mil pares de base de transcrito por milhão de leituras sequenciadas). Ainda, na região comum à ambas isoformas, observamos reduções também nos exons com incompatibilidades entre as anotações do RefSeq e UCSC: o exon 2+3, 8, 12 e o sem equivalente (indicado como "?" na figura). Para os exons isoforma-específicos, GAC detém o 15 sem nenhuma incongruência evidente (seus níveis são inferiores aos dos exons comuns, o que é esperado dada a divisão entre mRNAs produzidos por *splicing* alternativo). Já KGA detém o exon 16 com nível de expressão próximo ao do exon exclusivo de GAC, o exon 17+18 reduzido e o exon 19 reduzido frente à GAC e ao restante dos exons comuns. Dado que o exon 16 apresenta um tamanho diminuto e, conseqüentemente, implica em maior alteração numérica gerada pela normalização por tamanho do mesmo, decidiu-se utilizar os exons 15 e 19 como “leitores” dos níveis de GAC e KGA, respectivamente. Tais exons são terminais e portadores das 3'UTR, são de tamanho razoavelmente similar entre as duas isoformas e tem anotação compatível entre RefSeq e UCSC.

Com a capacidade de discernir adequadamente as duas isoformas de glutaminase nos dados disponíveis no TCGA, avaliou-se a razão de expressão entre os tecidos tumorais e normais para GAC (Figura 20C) e KGA (Figura 20D), utilizando a abundância dos exons isoforma-específicos. Observamos frequente aumento de expressão GAC nos tecidos tumorais (Figura 20C), diferente de *GLS* (Figura 19) e KGA (Figura 20D).

Para confirmar a observação de que GAC estaria mais expressa em diversos tipos tumorais, avaliamos a razão entre tumor e tecido normal das razões entre os exons 19 e 15. Nesta operação, tanto numerador quando denominador se deslocam para valores maiores quando na maior abundância de KGA e valores menores quando da maior abundância de GAC. Assim, o resultado final representa se a proporção entre KGA e GAC se manteve estável

(próxima a 1), mudou a proporção na direção de KGA (valores superiores a 1) ou caminhou na direção de aumentar a expressão de GAC (valores inferiores a 1), como demonstrado gráfica (Figura 21A) e numericamente (Figura 21B).

Fica evidente agora que grande parte dos tumores ocorre um movimento de modificação na escolha das isoformas, em direção a preferir GAC preterindo KGA. Apenas fígado, ducto biliar, rim e timo movem sua escolha para enriquecimento em KGA (sendo de forma significativa apenas para o primeiro). Por exemplo, para o caso de fígado (Figura 21C), o tecido normal apresenta uma razão inferior a 1, indicando uma escolha por GAC, que se move para um valor próximo a 1, indicando um equilíbrio entre as isoformas, o que implica numa mudança de priorização das isoformas, sendo esta escolha deslocada para KGA no contexto tumoral em relação ao normal. Cabe ressaltar que este é o único tipo tumoral em que, significativamente ( $p$ -value 0,0002) ocorre essa priorização. Para intestino (Figura 21D), temos o tecido normal com maior abundância da isoforma KGA (valores superiores a 1), sendo que o tecido tumoral reduz a proporção (para valores inferiores a 1), com uma grande predileção por GAC. Por fim, em mama (Figura 21E), os valores se mantêm muito próximos, tanto o tecido normal quanto o tecido tumoral apresentam uma preferência por GAC – a qual aumenta muito brevemente no contexto tumoral, de forma não significativa.



**Figura 21 – Expressão preferencial das isoformas de glutaminase em distintos tipos tumorais**

(A) Representação topográfica da razão das razões de expressão do exon 19 de *GLS* (exclusivo à KGA) em tumores primários sobre tecido normal adjacente sobre o exon 15 de *GLS* (exclusivo à GAC) em tumores primários sobre tecido normal adjacente, indicando o movimento da expressão preferencial entre as isoformas. (B) Gráfico RDI representativo para tumores de fígado, indicando uma preferência por GAC nos tecidos normais (razão menor que 1), movendo-se para uma proporção de igualdade entre as isoformas (próxima a 1), representando um movimento pró-KGA na patologia. (C) Gráfico RDI representativo para tumores de cólon intestinal, indicando uma preferência por KGA nos tecidos normais (razão maior que 1), movendo-se para uma proporção de seleção de GAC (inferior a 1), representando um movimento pró-GAC na patologia. (D) Gráfico RDI representativo para tumores de mama,

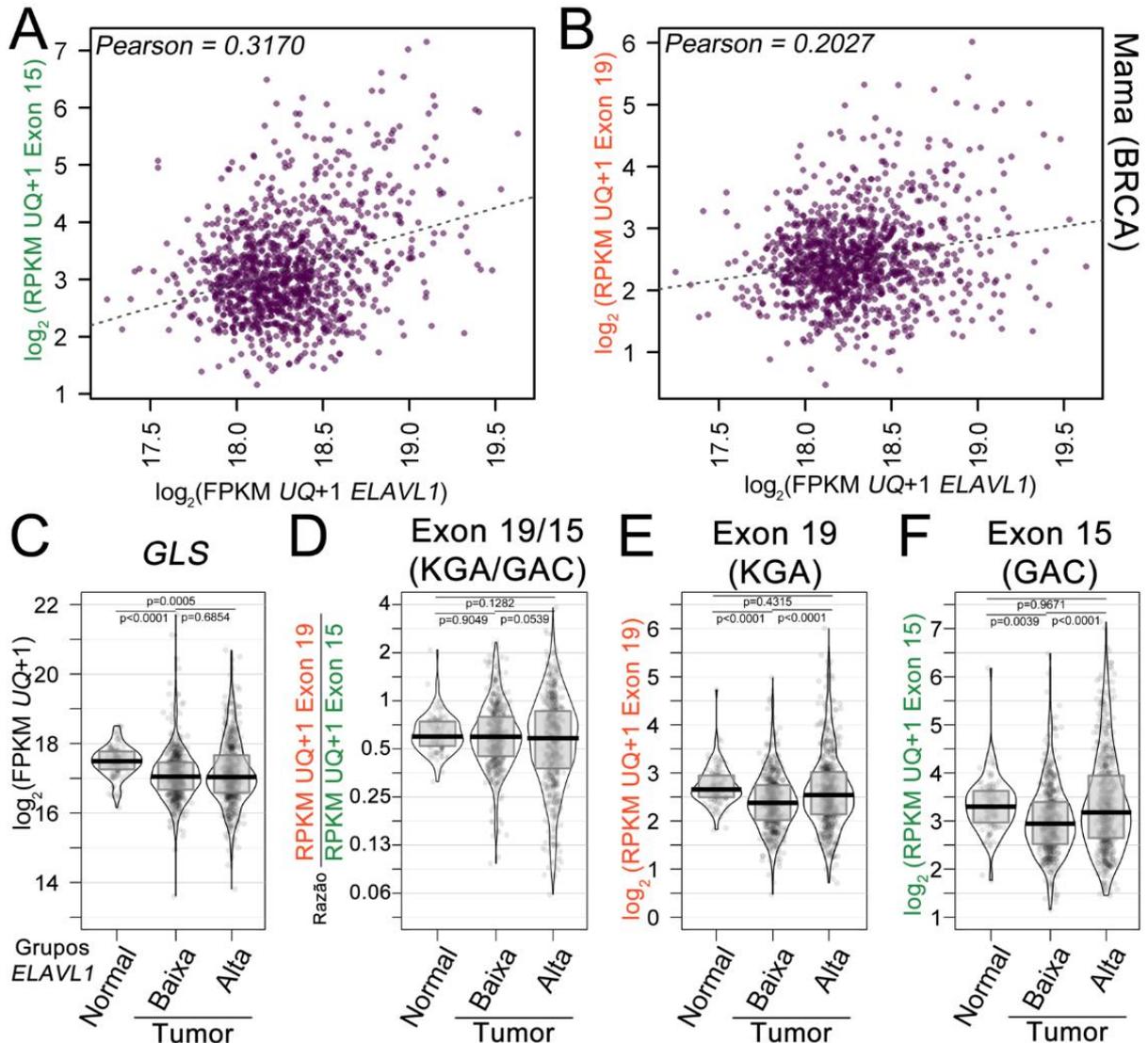
indicando uma breve preferência por GAC similar tanto no tecido normal quanto na patologia, sem uma alteração da escolha. As ordenadas foram produzidas em escala logarítmica e os valores transcritos em seus correspondentes na escala linear para facilitar a compreensão numérica do eixo. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados sobre os dados como círculos. *P-values* originados de testes-t não-pareados. A numeração dos órgãos está presente no Quadro 7.

**Tabela 8 – Lista com os valores absolutos das razões das medianas das isoformas KGA/GAC entre tecido tumoral e normal representadas na Figura 21 com o órgão afetado e o respectivo código empregado pelo TCGA**

Tecido	Código TCGA	Tipo	Razão	<i>p-value</i>	FDR
Fígado	LIHC	Carcinoma hepatocelular	1,51	0,0002 †	0,0005
Ducto biliar	CHOL	Colangiocarcinoma	1,32	0,2983 *	0,4475
Rim	KICH	Carcinoma renal cromóforo	1,14	0,7223	0,8641
Timo	THYM	Timoma	1,04	0,7561 *	0,8641
Mama	BRCA	Câncer de mama	0,99	0,2748	0,4397
Tireóide	THCA	Carcinoma papilar de tireóide	0,99	0,5118	0,6465
Prostata	PRAD	Adenocarcinoma de próstata	0,96	0,0229	0,0423
Pele	SKCM	Melanoma cutâneo	0,95	0,9971 *	1,0000
Adrenal	PCPG	Feocromocitoma & Paraganglioma	0,90	1,0000 *	1,0000
Bexiga	BLCA	Cancer de bexiga	0,89	0,1233	0,2114
Pâncreas	PAAD	Adenocarcinoma pancreático ductal	0,83	0,8651 *	0,9437
Útero	UCEC	Câncer de endométrio	0,71	0,4592	0,6123
Cérvix uterino	CESC	Câncer cervical	0,68	0,3218 *	0,4543
Esôfago	ESCA	Câncer de esôfago	0,65	0,0108 *	0,0236
Estômago	STAD	Adenocarcinoma de estômago	0,65	<0,0001 †	<0,0001
Cabeça e pescoço	HNSC	Carcinoma de células escamosas de cabeça e pescoço	0,59	<0,0001 †	<0,0001
Intestino delgado	READ	Adenocarcinoma de reto	0,57	0,0154 *	0,0308
Pulmão	LUAD	Adenocarcinoma de pulmão	0,52	<0,0001 †	<0,0001
Rim	KIRC	Carcinoma renal de células claras	0,48	<0,0001 †	<0,0001
Rim	KIRP	Carcinoma papilar renal	0,46	<0,0001 †	<0,0001
Pulmão	LUSC	Carcinoma de células escamosas de pulmão	0,37	<0,0001 †	<0,0001
Intestino grosso	COAD	Adenocarcinoma de cólon	0,32	<0,0001 †	<0,0001
Tecidos macios	SARC	Sarcoma	0,07	0,0004 *†	0,0010
Cérebro	GBM	Glioblastoma multiforme	0,02	<0,0001 *†	<0,0001

NOTA: Indicações de \* (asterisco) denotam tumores com pequeno tamanho amostral ( $\leq 10$ ) para o tecido normal adjacente e de † (cruz) *p-value* significativo ( $< 0,01$ ). Correção dos valores de p em FDR (*false discovery ratio*) pelo método de Benjamini-Hochberg<sup>142</sup>. Valores superiores a 1 indicam priorização da escolha de KGA e inferiores a 1, GAC.

Focando no contexto de câncer de mama, origem na qual a expressão de *ELAVL1* apresentou a maior significância de relação com prognóstico negativo para o paciente (Figura 15), decidimos avaliar a correlação da sua expressão com as isoformas KGA e GAC (Figura 22).



**Figura 22 – Correlação entre os níveis dos exons exclusivos de KGA e GAC e os níveis de *ELAVL1* em tumores de mama**

Gráfico de dispersão entre os níveis do mRNA de *ELAVL1* e do exon 15 exclusivo de GAC (A) ou do exon 19 exclusivo de KGA (B) em tecido tumoral de mama. Correlação de Pearson indicada no canto superior esquerdo dos gráficos. Gráficos RDI com a separação dos pacientes por nível de expressão de *ELAVL1* na separação escolhida pelo método da minimização do *p-value* de recidiva frente *GLS* (C), razão da abundância do exon 19/exon 15 de *GLS* (D), abundância do exon 19 (E) e abundância do exon 15 (F). Os *p-values* indicados foram gerados após uma análise de variância seguida de um teste de Tukey para as médias. As ordenadas foram produzidas em escala logarítmica e os valores transcritos em seus correspondentes na escala linear para facilitar a compreensão numérica do eixo quando na representação de razões. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; todos os pacientes estão individualmente representados sobre os dados como círculos.

Ao compararmos a expressão do exon 15 (GAC) com os níveis de *ELAVL1*, encontramos uma correlação positiva (pearson = 0,3170), superior a observada com o exon de KGA (Figura 22B, Pearson = 0,2027). Em ambos os casos as correlações não foram numericamente elevadas, mas justificam uma tendência de co-ocorrência. Tal tendência fica mais evidente quando separamos os pacientes de acordo com a categorização proposta para o

nível de expressão de *ELAVLI*. Os níveis absolutos de expressão de *GLS* não são impactados com a separação (Figura 22C), nem a razão entre KGA/GAC, mas a abundância tanto de KGA (Figura 22E) quando de GAC (Figura 22F) são superiores. Assim concluímos que, em tumores de mama, a expressão de *ELAVLI* correlaciona-se positivamente com níveis de mRNA de GAC e KGA.

## 4.2 AVALIAÇÃO GENÔMICA DA RELAÇÃO DE HUR E *SPLICING* ALTERNATIVO

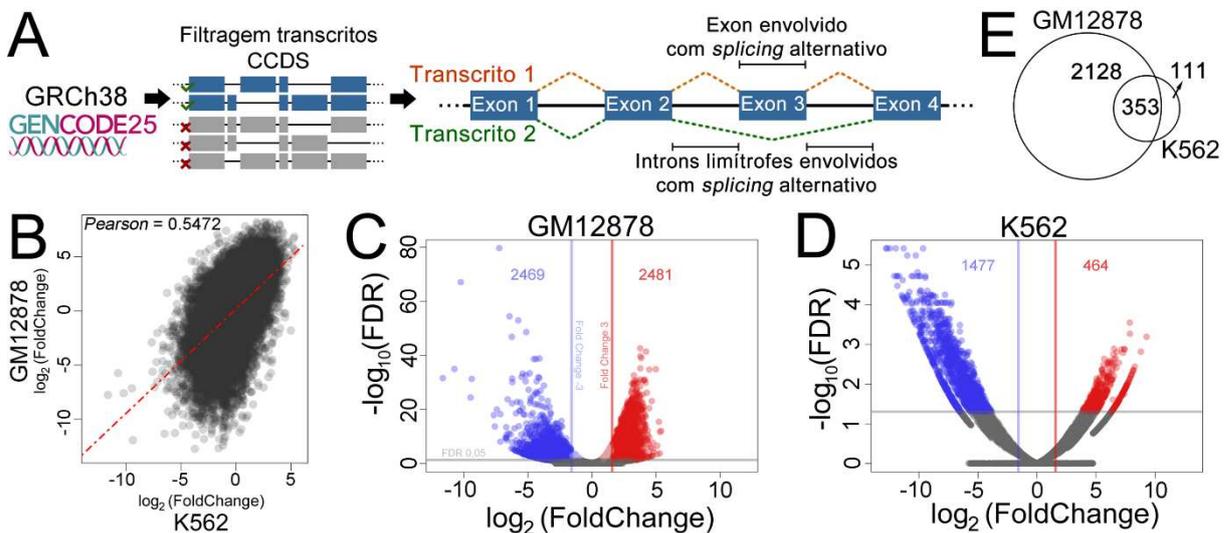
Como colocado na introdução, HuR pode impactar em diversos eventos como na estabilização de mRNA pela interação com elementos regulatórios presentes no 3'UTR assim como no *splicing* alternativo de genes. Assim, utilizamos dados disponíveis em bancos de dados para avaliar a atuação deste regulador no *splicing* do genoma como um todo e especialmente de *GLS*.

### 4.2.1 Inferência da ligação à introns relacionados com *splicing* alternativo utilizando dados disponíveis publicamente de imunoprecipitação de RNA seguida de sequenciamento (RIP-Seq)

O experimento de RIP-Seq pode ser descrito como a imunoprecipitação total dos RNAs que estejam de alguma forma associados a proteína de interesse, seguido da avaliação por sequenciamento massivo de RNA<sup>257</sup>. Como sua especificidade recai na interação específica do anticorpo com a proteína de interesse, usualmente utiliza-se como controle a imunoprecipitação paralela de um controle de isotipo, por exemplo. A metodologia de análise dos dados usualmente passa pelo tratamento dos mesmos como experimentos convencionais de expressão diferencial (observando quantos fragmentos alinham em determinada região e comparando a imunoprecipitação com o seu controle) ou como experimentos de ChIP-Seq (comparando a presença de “picos” de fragmentos alinhando em determinadas regiões, independente do contexto regulatório genômico)<sup>258</sup>. Entretanto, dado que 15-35% dos fragmentos obtidos em um experimento de RNA-Seq (com depleção apenas de rRNA) alinham em regiões intrônicas e não codificantes<sup>259</sup> (presentes no pre-mRNA e de tamanho muito superior às regiões codificantes), a maioria restante (65%-85%) representa uma região menor que 1% do genoma, composta por exons e regiões codificantes. Assim sendo, existe uma exposição diferenciada das regiões intrônicas à técnica. Como diversos algoritmos de contagem e avaliação levam em consideração a frequência da ligação em todas as regiões do genoma de

forma não diferenciada para a estimativa da variância global e consequente cálculo de *p-values*<sup>163,165,260,261</sup>, o sinal menos intenso em tais regiões acaba se assemelhando a ruído experimental de regiões exônicas, o que as levam a desconsideração na análise. Os dados de RIP-Seq utilizados foram produzidos na iniciativa ENCODE<sup>147</sup> de compreensão dos elementos regulatórios do genoma, incluindo proteínas da ligação à mRNA, sendo disponíveis como *tracks* para avaliação de ligação de proteínas regulatórias em regiões 3'UTR de diversos genes.

Para avaliarmos preliminarmente se HuR estaria interagindo com o pré-mRNA de glutaminase de alguma forma, e dado que tais regiões estariam menos abundantes, produziu-se uma anotação diferenciada do genoma humano para a contagem dos fragmentos alinhados apenas em introns possivelmente relacionados com *splicing* alternativo (imediatamente a jusante ou a montante de exons que sofrem *splicing* alternativo)<sup>262</sup>. Recentemente, inclusive, um estudo utilizou o mesmo conjunto de dados com abordagem similar – de avaliação apenas intrônica – para mensurar o impacto de HuR em *splicing* alternativo<sup>263</sup>. Utilizando-se a versão 25 da anotação do projeto GENCODE<sup>157,158</sup> sobre a versão 38 do *Human Genome Reference Consortium* (GRCh38)<sup>264</sup>, separou-se os genes e transcritos que compusessem a *Consensus Coding Sequence Database* (CCDS)<sup>265</sup> (Figura 23A).



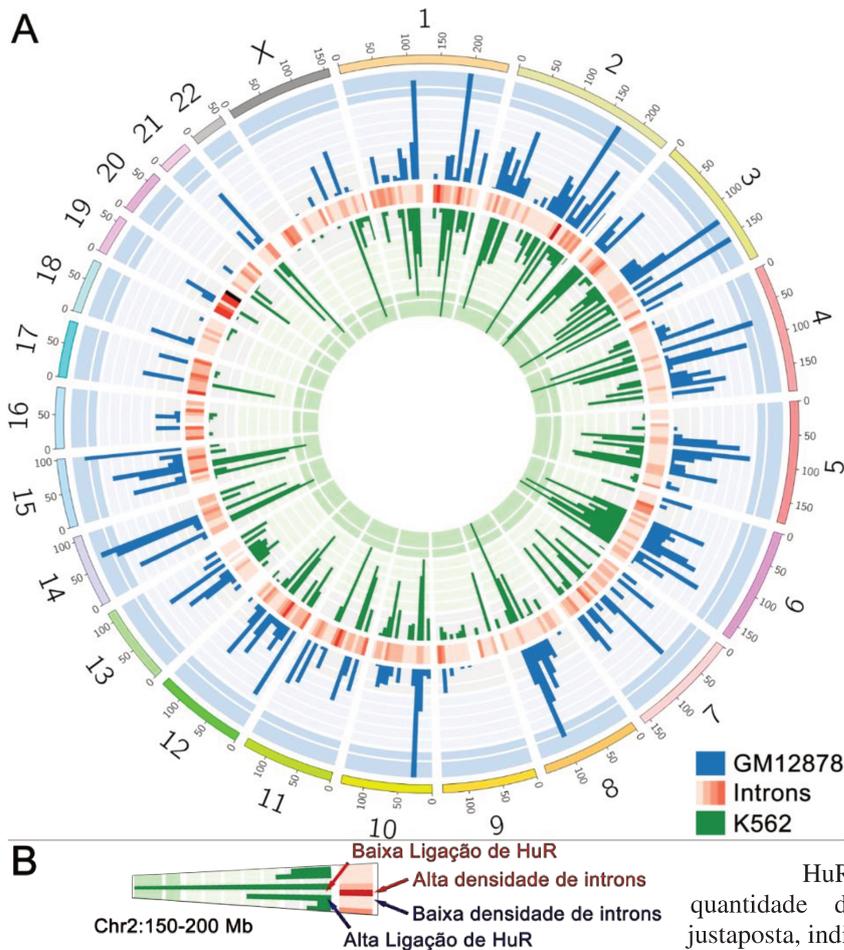
**Figura 23 – Busca de introns justapostos a exons envolvidos em *splicing* alternativo diferencialmente abundantes em experimento de imunoprecipitação de RNA seguida de sequenciamento (RIP-Seq)**

(A) Esquema da estratégia utilizada para a definição da anotação a ser utilizada para a contagem de fragmentos imunoprecipitados. (B) Correlação entre os valores de *fold change* entre a imunoprecipitação de HuR e o controle inespecífico em dois tipos celulares distintos. *Volcano plots* indicando os introns mais abundantes no controle (IP: Controle, em azul) e os introns mais abundantes na imunoprecipitação de HuR (IP:HuR, em vermelho) para as células GM12878 (C) e K562 (D). Linhas verticais indicam o enriquecimento de três vezes e linha horizontal o limite de significância do *false discovery ratio* de 0,05. (E) Diagrama de Venn da intersecção dos introns imunoprecipitados em ambas condições de acordo dos critérios de significância estabelecidos.

Lançando mão de tal anotação foi possível, então, avaliar o alinhamento contando fragmentos observados na precipitação de HuR ou de T7Tag, o controle experimental, para as duas réplicas disponíveis de experimentos realizados em 2 linhagens celulares, a GM12878 (linfoblastos não malignos) e K562 (linhagem de leucemia mielóide crônica). Comparando as quantidades normalizadas de fragmentos alinhados em um mesmo intron na amostra IP:HuR e em IP:Controle entre as duas linhagens celulares, observamos a existência de uma correlação positiva, mostrando que HuR tende a se ligar aos mesmo introns nas duas linhagens (Figura 23B). GM12878 (Figura 23C), ao contrário de K562 (Figura 23D), apresentou quantidade menor de ruído, exemplificada pela menor fração de introns com significativa ligação na IP:Controle e maiores FDRs de maneira geral. Quando comparados qualitativamente, os introns selecionados em K562 estão em grande parte (76%) contidos na lista de introns selecionados em GM12878 (Figura 23E).

Com a identificação de uma abundante quantidade de regiões intrônicas em que HuR interage, levantou-se a hipótese de uma ligação ubíqua no genoma, sem necessariamente uma seleção específica de alvos – ocasionada por problemas com as condições experimentais da imunoprecipitação. A visualização Circos (Figura 24A) permite observar se há correlação entre regiões com elevada densidade de introns e a ligação de HuR, se a distribuição é homogênea entre todas as regiões cromossômicas e a concordância entre os tipos celulares. Observa-se a presença de várias regiões de alta densidade de introns possivelmente regulados, mas com baixa ligação de HuR, assim como o exato oposto (Figura 24B). Tal cenário indica que HuR não está ligando ubiquamente aos RNAs produtos do genoma e, sim, em transcritos produzidos em determinadas regiões. Também, podemos confirmar os dados de correlação entre as duas linhagens celulares (Figura 23B) de forma espacialmente localizada e observar a heterogeneidade de regiões com *splicing* alternativo ao longo do genoma, denotado na pista do mapa de calor. Em conjunto, tais dados sugerem que a ligação de HuR em introns potencialmente relacionados com a regulação do *splicing* alternativo tem relação com uma reprogramação celular para um objetivo em específico, e não se dá ao acaso<sup>89,96,97</sup>.

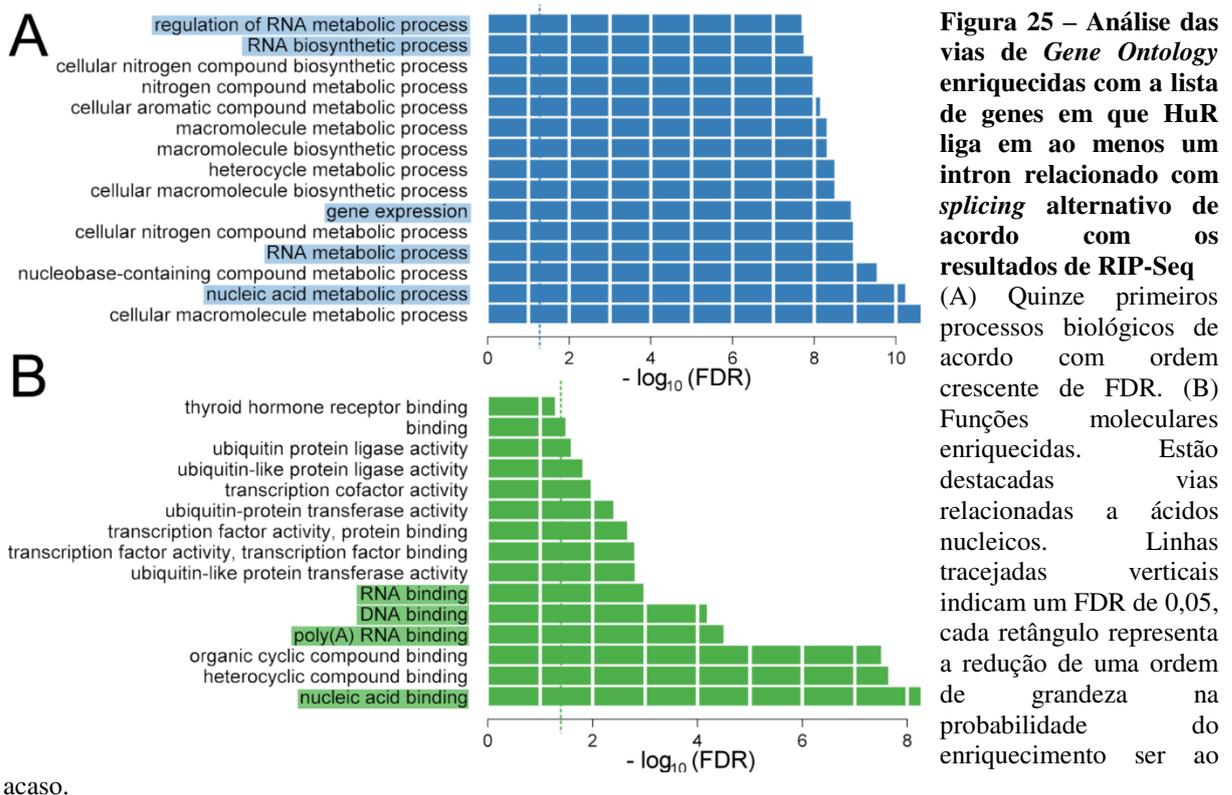
GM12878 foi selecionada para as análises posteriores por apresentar 2481 introns imunoprecipitados de forma significativa (Figura 23E), sinal reduzido em IP:Controle, FDRs superiores (Figura 23C) e padrão de regulação genômico correlacionado numérica (Figura 23B) e espacialmente (Figura 24A) com K562.



**Figura 24 – Densidade de introns justapostos a exons envolvidos em *splicing* alternativo e com ligação de HuR**

(A) Gráfico Circos<sup>166</sup> indicando na pista externa os cromossomos humanos. Na pista 2 imediatamente interna, verifica-se histograma em azul com o somatório de contagens por milhão (CPM) de IP:HuR subtraídas de IP:Controle, sobre regiões seccionadas de 7 Mpb, para a célula GM12878. Na pista 3 verifica-se *heatmap* da densidade de introns passíveis de envolvimento com *splicing* alternativo (em tons de vermelho). Na pista 4 subsequente, verifica-se histograma em verde similar ao primeiro, para a célula K562. (B) Demonstrativo da região do cromossomo 2 indicando uma posição com baixa ligação de HuR mesmo na presença de elevada quantidade de introns reguláveis e outra, justaposta, indicando o evento oposto.

Buscando compreender o conjunto de eventos biológicos regulados pelos genes cujo *splicing* pode ser influenciado por HuR, realizou-se uma análise de enriquecimento de vias com os 1459 genes que apresentaram a ligação de HuR a ao menos um intron envolvido com *splicing* alternativo (Figura 25). As vias enriquecidas, de maneira geral, apontam para eventos relacionados com a regulação da expressão gênica geral através da modulação do metabolismo de RNA e DNA e remodelação de cromatina. Assim, HuR estaria regulando processos nos quais ela própria participa, orquestrando a regulação da expressão gênica no nível pós-transcricional de forma direta e através da modificação e modulação de efetores diversos. Curiosamente, outros processos metabólicos gerais e especializados encontram-se nas listas, ressaltando uma possível relação de HuR com metabolismo de nutrientes.



**Figura 25 – Análise das vias de Gene Ontology enriquecidas com a lista de genes em que HuR liga em ao menos um intron relacionado com *splicing* alternativo de acordo com os resultados de RIP-Seq** (A) Quinze primeiros processos biológicos de acordo com ordem crescente de FDR. (B) Funções moleculares enriquecidas. Estão destacadas vias relacionadas a ácidos nucleicos. Linhas tracejadas verticais indicam um FDR de 0,05, cada retângulo representa a redução de uma ordem de grandeza na probabilidade do enriquecimento ser ao

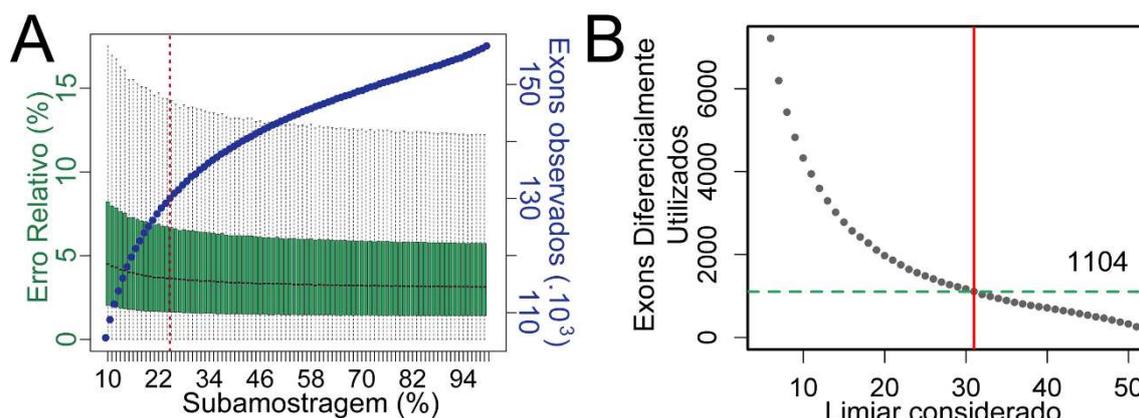
#### 4.2.2 Inferência de exons diferencialmente abundantes quando do silenciamento de HuR utilizando dados previamente disponíveis de sequenciamento de RNA e o contexto evolutivo dos hipotéticos elementos regulatórios

Para transformar a informação de ligação a um intron em modulação da expressão diferencial do gene de forma experimental, buscou-se na literatura um conjunto de dados de silenciamento de HuR seguido de RNA-Seq. Lebedeva *et al*<sup>96</sup> realizaram transcriptômica e proteômica de células silenciando *ELAVL1* e buscaram novos alvos regulados, além de confirmar a interação com experimentos de PAR-CLIP-Seq. Mesmo sendo o foco deste trabalho o mapeamento da ligação de HuR em regiões regulatórias 5' e 3'UTR, os autores também observaram alterações na ligação à regiões intrônicas e o efeito de HuR no *splicing* alternativo. Entretanto, a abordagem desenvolvida para a busca de exons diferencialmente expressos envolvia a comparação de três exons consecutivos de um mesmo gene, realizando-se a média de fragmentos alinhados no terceiro e no primeiro e comparando com a do central. Caso o mesmo variasse mais do que 50% em relação à média seria colocado como diferencialmente expresso, parâmetro que pode ser demasiado estrigente e sofrer redução de performance em situações como a do gene *GLS*, em que os quatro últimos exons em sequência são diferencialmente expressos em relação aos demais iniciais.

Diversas abordagens para a definição de transcritos estão disponíveis na literatura, como a Tuxedo Suite<sup>168</sup>, RSEM<sup>170</sup>/EBSeq<sup>260</sup> e ISODE<sup>164</sup>, as quais consideram a unidade mínima como isoformas, ou MISO<sup>266</sup> e HTSeq<sup>162</sup>/DEXSeq<sup>163</sup>, que consideram a unidade mínima como o exon. Quando na ausência de réplicas (técnicas ou biológicas, caso do conjunto de dados de *Lebedeva et al*<sup>96</sup>), apenas ISODE se propõe a estabelecer valores de confiança, utilizando uma reamostragem por *bootstrap* sobre o arquivo de alinhamento. Cabe mencionar que as estratégias aplicadas por transcrito tem redução de performance tanto na ausência quanto no excesso de anotação de isoformas, o que faz os mecanismos baseados em exon demonstrarem uma performance melhor<sup>167</sup>, especialmente o conjunto HTSeq<sup>162</sup>/DEXSeq<sup>163</sup>.

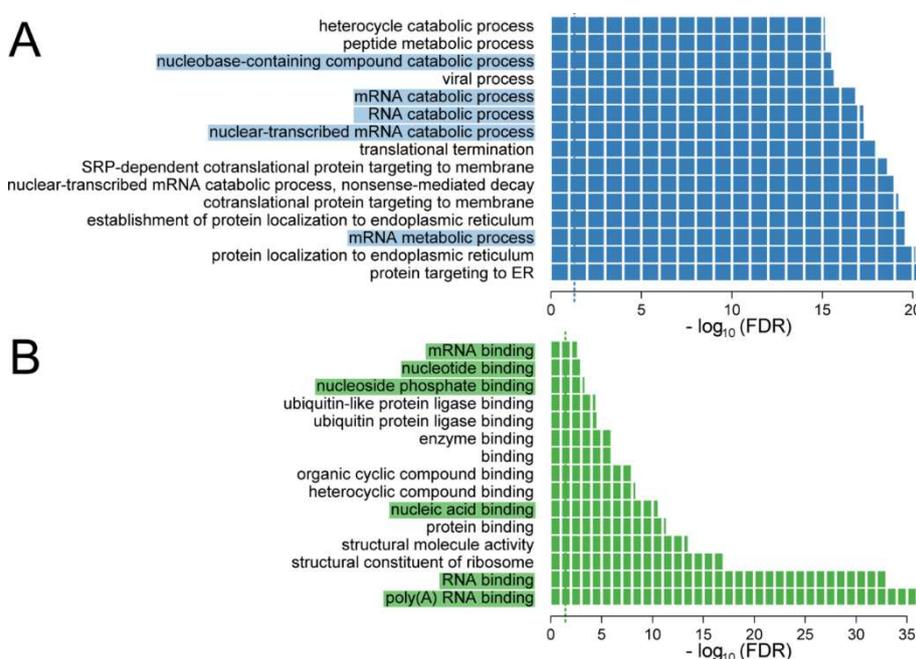
Assim, com o intuito de reprocessar os dados disponíveis com a unidade mínima sendo o exon, a metodologia de ISODE<sup>164</sup> foi modificada. Ao invés da reamostragem direta sobre o alinhamento baseada em *bootstrap* com reposição, aplicou-se inicialmente uma subamostragem (onde a reamostragem se dá sem reposição de fragmentos e gerando um subconjunto menor do original, dada a massiva quantidade de fragmentos sequenciados, superior a 54 milhões de fragmentos alinhados por condição, análogo a um *jackknife<sub>n</sub>*). Comparando-se as diferentes porcentagens de subamostragem, percebe-se que, mesmo em frações pequenas, o erro relativo em relação ao conjunto completo é baixo e pouco se altera com o aumento da fração de reamostragem (Figura 26A). Já quanto a curva de adição de novos exons observados (exons com ao menos 5 fragmentos observados na quadruplicada da subamostragem), não há um platô evidente, indicando que quanto mais fragmentos forem sequenciados, maior será o número de exons detectados – incluindo cada vez mais exons de expressão extremamente baixa e não relevantes para análise em questão. Com base nestes resultados, utilizou-se uma reamostragem de 25% dos fragmentos originais. Foi possível observar, utilizando esta estratégia, 1104 exons diferencialmente utilizados (Figura 26B) a um nível de significância de 5%.

Por fim, utilizando a lista dos genes com exons diferencialmente expressos, realizou-se um novo enriquecimento de vias (Figura 27), tal qual realizada para ligação de HuR aos introns relacionados a *splicing* alternativo. Coincidentemente ao primeiro resultado, vias de metabolismo de RNA e DNA e ligação a ambos se mostraram enriquecidas. Assim, avaliando os dois resultados em conjunto, corroboramos o evento de HuR regular processos nos quais ela própria participa, orquestrando a regulação da expressão gênica no nível pós-transcricional de forma direta e através da modificação e modulação de efetores diversos.



**Figura 26 – Estratégia estatística de subamostragem e replicação aplicada para a busca de exons diferencialmente utilizados após o silenciamento de *ELAVL1* em célula HeLa**

(A) Aumento da fração de subamostragem (quadruplicata para cada porcentagem) dos fragmentos alinhados ao genoma e sua consequência na redução do erro relativo em relação ao conjunto de fragmentos total (ordenada esquerda, verde, representada nos diagramas de caixa, com barras pretas indicando a mediana, *whiskers* estendendo-se até o percentil de 90% e limites da caixa no primeiro e terceiro quartis) e na coleta de novos exons (ordenada direita, em azul, representada nos pontos). (B) Quantidade de exons diferencialmente abundantes, de acordo com o modelo de DEXSeq<sup>163</sup>, para distintos limiares de confirmação em número de repetições de reamostragem. Linha vertical vermelha indica número mínimo de reamostragens necessárias para atingir significância de 5 % (de acordo com distribuição Bernoulli da probabilidade cumulativa do evento binário, em 51 replicações). Linha horizontal tracejada verde indica o número de exons observados.



**Figura 27 – Análise das vias de *Gene Ontology* enriquecidas com a lista de genes em que o silenciamento de *ELAVL1* impacta na abundância diferencial de algum de seus exons**

(A) Quinze primeiros processos biológicos de acordo com ordem crescente de FDR. (B) Funções moleculares enriquecidas. Estão destacadas vias relacionadas a ácidos nucleicos. Linhas tracejadas verticais indicam uma redução de uma ordem de grandeza na probabilidade

do enriquecimento ser ao acaso.

Com a disponibilidade das duas listas, uma contendo os genes em que HuR apresenta ligação em introns relacionados com o *splicing* alternativo e a outra de genes cujo *splicing* é alterado na depleção de HuR, foi possível extrair da intersecção das mesmas, que resumiria os genes candidatos a terem o *splicing* alternativo regulado diretamente por HuR.

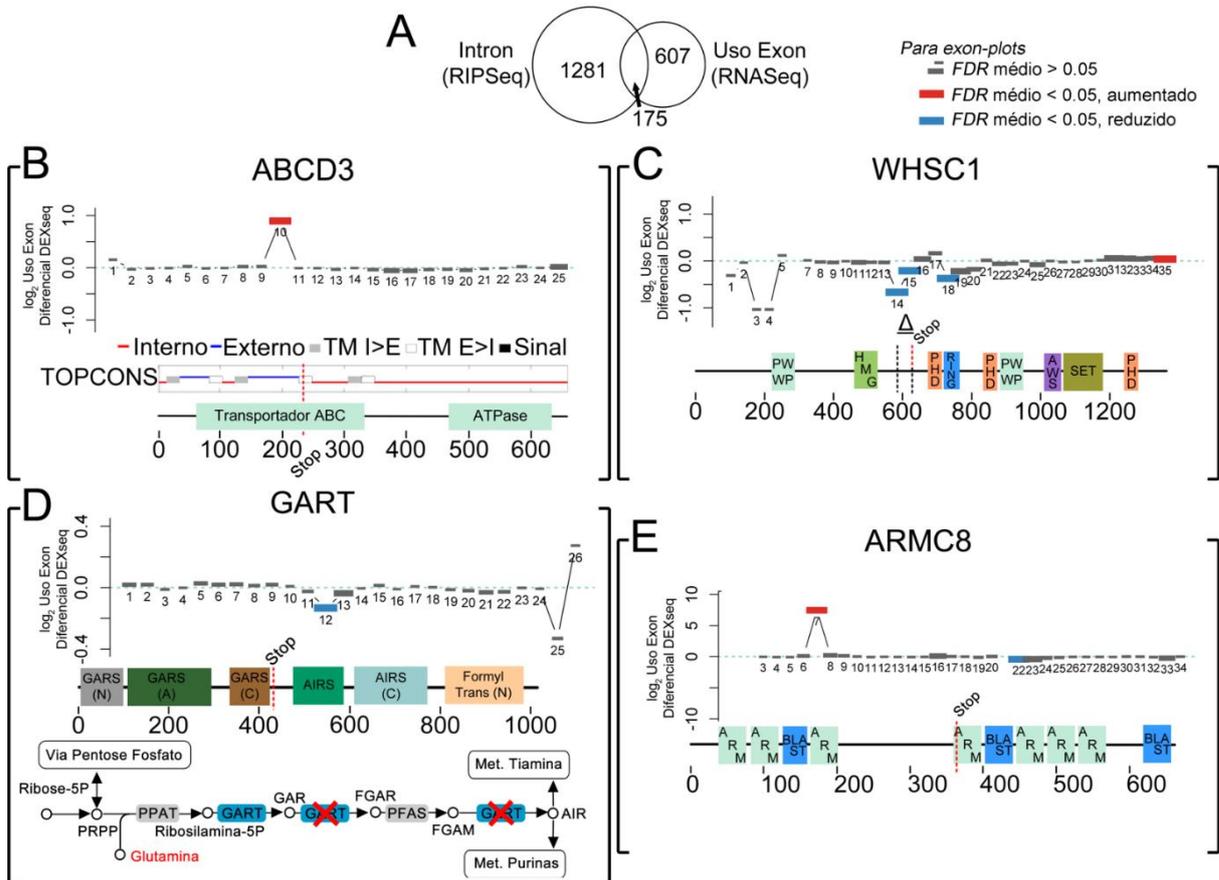
Observaram-se 175 genes (Figura 28A, lista no Apêndice 2) candidatos, diversos deles envolvendo apenas substituição de regiões não-codificantes do mRNA (5'UTR e 3'UTR), assim como de regiões codificantes sem informação acerca da importância funcional (regiões sem domínios caracterizados e sem informação estrutural). Buscaram-se alguns exemplos que envolvessem a substituição ou remoção de domínios completos, os quais são elencados abaixo.

O transportador *ABCD3* (*ATP-binding cassette sub-family D member 3*, Figura 28B) é envolvido com o transporte de moléculas através da membrana do peroxissomo com a utilização de ATP. Estruturalmente, é composto por um domínio transmembrana (responsável pelo transporte propriamente dito) e um domínio ATPase (responsável pela obtenção de energia para o transporte)<sup>267</sup>. Especificamente, *ABCD3* está implicado na realização de  $\beta$ -oxidação como fonte de energia para a célula e seu silenciamento implica na redução dessa via energética<sup>268</sup>. O silenciamento de *ELAVL1* implicou no aumento da isoforma truncada, sem o domínio ATPase, logo incapaz de realizar transporte – ao menos do tipo ativo. A troca de isoformas gera uma abrupta alteração funcional da proteína, podendo *splicing* impactar na manutenção da via de beta-oxidação.

A metiltransferase de histona *WHSC1* (*Wolf-Hirschhorn Syndrome Candidate 1*, Figura 28C) contém um domínio SET envolvido com o processo de metilação de histona H3 na lisina 27<sup>269,270</sup> (levando a repressão da atividade transcricional vizinha) ou na lisina 36 (levando a um aumento de expressão de alvos<sup>271</sup>). Sua superexpressão em tumores de próstata leva a aumento da invasividade através da execução de EMT, dentro outros fatores, por ativar a expressão de *TWIST1*<sup>271</sup>. Além de próstata, *WHSC1* também está superexpressa em outros tipos tumorais<sup>269</sup>. O silenciamento de *ELAVL1* reprime exons envolvidos com a expressão de uma isoforma menor, truncada e sem o domínio SET, logo sem atividade metiltransferase. O silenciamento também diminui a expressão de outra isoforma sem uma fração entre o domínio HMG e PHD, cujo impacto funcional é desconhecido. Assim, a presença de HuR reduziria a atividade metiltransferase de *WHSC1*, podendo influenciar características associadas a agressividade tumoral através de mecanismos epigenéticos.

A enzima trifuncional GARS-AIRS-GART (*glycinamide ribonucleotide synthetase, aminoimidazole ribonucleotide synthetase, glycinamide ribonucleotide formyltransferase*, Figura 28D), produto do gene *GART*, é responsável pela catálise de três passos na síntese do metabolismo de purinas a partir da Ribose-5P (Figura 28D, inferior)<sup>272,273</sup> e sua superexpressão é associada com pior prognóstico de gliomas<sup>274</sup> e hepatocarcinomas<sup>275,276</sup> – o que é justificável pela dependência de bases nitrogenadas para a elevada proliferação tumoral. O silenciamento de *ELAVL1* diminui a expressão do exon 12, que trunca a proteína

logo após a porção GARS, permitindo a expressão do polipeptídeo completo e a execução da síntese de purinas de forma total (Figura 28D, superior). Logo, a presença de HuR introduz a isoforma incompleta, apenas com a porção GARS, com potencial consequência no metabolismo de purinas.



**Figura 28 – Exemplos resultantes da intersecção dos resultados de ligação à introns e escolha diferencial de exons**

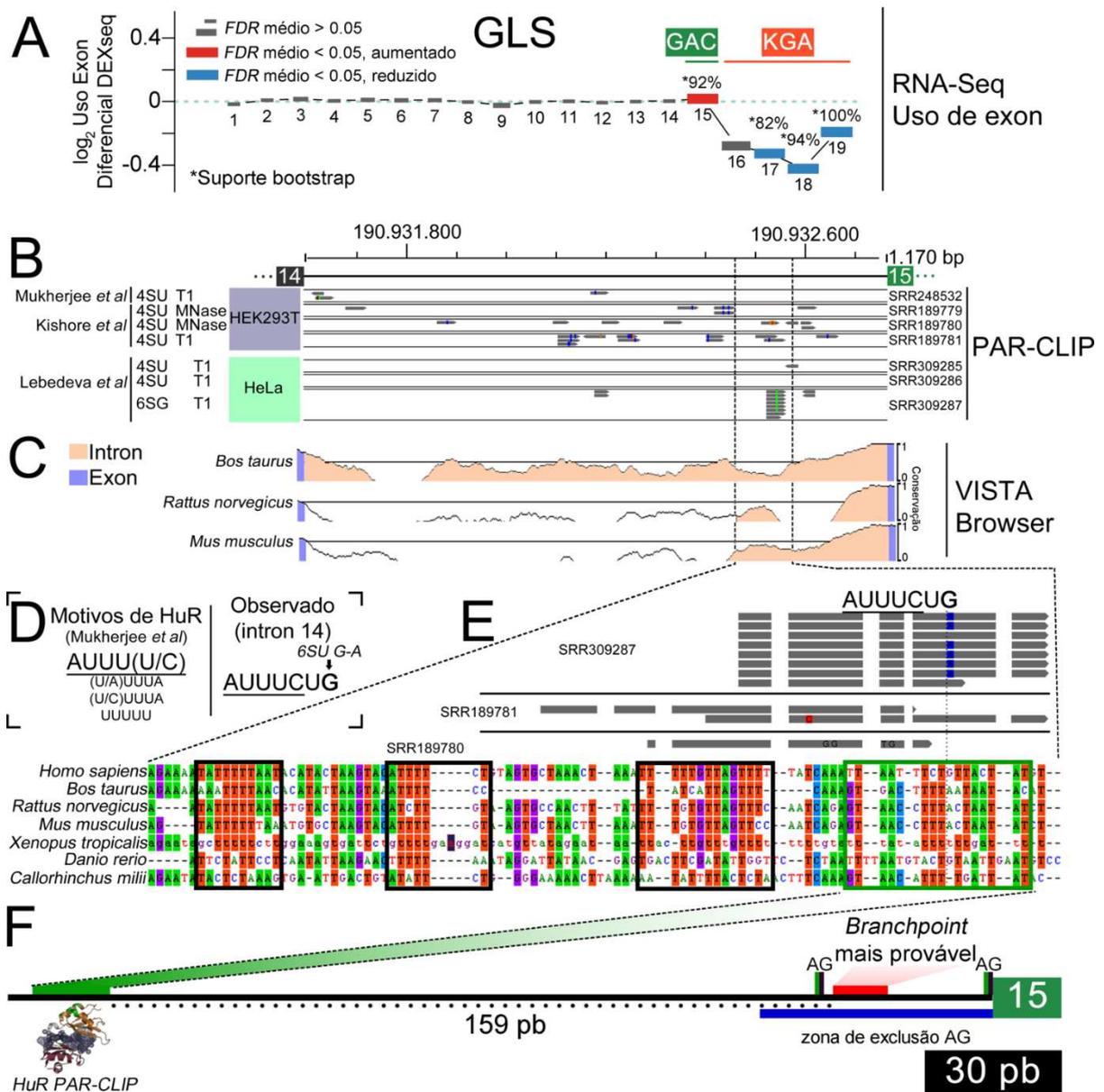
(A) Intersecção dos genes em que HuR liga-se a ao menos um intron relacionado com *splicing* alternativo (RIP-Seq) e genes com a abundância diferencial de ao menos um exon relacionado a *splicing* alternativo (RNA-Seq). Seta indica a intersecção. (B-E) Exemplos da lista de genes candidatos a regulação direta por HuR através da utilização de gráficos de exon. Nos mesmos, o tamanho da barra representativa do exon está proporcional a redução do FDR médio entre todas as reamostragens. Valores do eixo y representam a abundância diferencial pela métrica de DEXSeq. (B, superior) Regulação de *ABCD3*, indicando a inserção do exon 10 após o silenciamento de *ELAVL1*, o que leva a um códon de parada prematuro. (B, inferior) Predição de região transmembrana de *ABCD3* e de domínios proteicos, indicando a consequência funcional do truncamento utilizando a ferramenta TOPCONS<sup>179</sup> e a predição de domínios transmembrana orientação interno-externo (TM I>E) ou externo-interno (TM E>I). (C, superior) Regulação de *WHSC1*, indicando a repressão dos exons 14, 15 e 18 e a superexpressão do exon 35. (C, inferior) Esquema com a região deletada e a posição do truncamento que é reprimida no silenciamento de *ELAVL1*. (D, superior) Regulação de *GART*, mostrando a repressão do exon 12. (D, inferior) Esquema indicando a isoforma truncada reprimida no silenciamento de *ELAVL1* e o bloqueio das reações catalisadas pelos domínios removidos. (E, superior) Regulação de *ARMC8* indicando a superexpressão do exon 7. (E, inferior) Esquema de isoforma truncada superexpressa na ausência de HuR. Para os gráficos de exon, o tamanho da barra representativa é proporcional a estimativa por bootstrap do FDR de o mesmo ser diferencialmente expresso. Exons com cores ultrapassaram o limiar de 31 repetições de *bootstrap*, sendo azuis quando reprimidos ou vermelhos quando superexpressos.

A proteína ARMC8 (*Armadillo Repeat Containing 8*, Figura 28E) faz parte de um complexo proteico relacionado com regulação da migração celular<sup>277</sup> e, conseqüentemente, invasão e agressividade em tumores de mama<sup>278</sup> e cólon<sup>278</sup>. O silenciamento de *ELAVL1* aumenta a expressão da isoforma menor através da inserção de um exon com um stop códon prematuro, possivelmente modulando a interação com o complexo CTLH e propriedades invasivas tumorais.

Por fim, observamos as alterações relacionadas a *GLS* (Figura 29). Confirmamos a interação de HuR com o pré-mRNA de *GLS* nos experimentos de RIP-Seq (Tabela 9), nos dois introns justapostos ao exon exclusivo de GAC. Quando do silenciamento de *ELAVL1*, ocorreu o aumento do exon exclusivo de GAC (15) e redução dos exons de KGA (17,18 e 19) (Figura 29A). O *fold-change* direto em escala linear é de 1,3 para o exon exclusivo de GAC (15) e 0,3 para o último exon de KGA (19).

Com o intuito de confirmar o sítio de interação de HuR com o pré-mRNA de glutaminase, foi empregada a técnica de PAR-CLIP-Seq (*Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation*) com dados disponíveis publicamente da imunoprecipitação de HuR. A técnica permite a detecção de sítios específicos de interação proteína-RNA ao se basear na suplementação do meio de cultivo celular com análogos de nucleotídeos fotoreativos. A irradiação de células com UV permite o *crosslink* da proteína ao RNA contendo o resíduo modificado na região do contato. Após a imunoprecipitação da proteína, isola-se o RNA crossligado, realiza-se a síntese de cDNA e o sequenciamento do mesmo. Regiões de contato que incorporaram o nucleotídeo fotoreativo podem ser detectados pela inserção de uma mutação na posição<sup>257,279</sup>.

Três conjuntos de dados foram obtidos da literatura. Lebedeva *et al*<sup>96</sup>, previamente mencionado, realizou o PAR-CLIP-Seq com dois diferentes nucleotídeos análogos (6SG e 4SU) e em dois tipos de meio de cultivo para o 4SU, análogo mais usual. Mukherjee *et al*<sup>97</sup> realizou análises de PAR-CLIP-Seq e RIP-Chip, além de transcriptômica, para avaliar sítios de ligação de HuR de maneira global. Neste processo, os autores identificaram sítios de ligação de HuR em introns e confirmaram a sua atividade frente *splicing* alternativo, sendo glutaminase identificada como alvo em sua análise. Kishore *et al*<sup>148</sup>, em uma publicação de objetivo mais técnico, comparam a utilização de distintas enzimas para a degradação do RNA fora do sítio de interação, como MNase e T1 RNase, em alguns modos de digestão. Dado que seu objetivo não era exatamente elucidar a biologia de HuR, limitou-se a avaliação de alvos exônicos, sem relação com *splicing* alternativo.



**Figura 29 – Utilização diferencial de exons de *GLS* frente ao silenciamento de *ELAVL1* e a proposta ligação ao intron 14**

(A, superior) Gráfico de exon de *GLS*, mostrando o aumento da expressão de 15 e a repressão de 16 (não-significativa), 17, 18 e 19. O tamanho da barra representativa do exon está proporcional a redução do FDR (média entre todas as reamostragens). Valores do eixo y representam a abundância diferencial pela métrica de DEXSeq. Números sobre as barras significativas indicam a fração das reamostragens que suportaram a diferença de expressão. Exons com cores ultrapassaram o limiar de 31 repetições de *bootstrap*, sendo azuis quando reprimidos ou vermelhos quando superexpressos. (B) Avaliação da presença de sítios de ligação de HuR no intron 14 de glutaminase (*GLS*) através de PAR-CLIP-Seq. Barras cinzas representam o posicionamento dos fragmentos observados em relação ao intron 14, linhas coloridas nos fragmentos denotam mutações do PAR-CLIP-Seq (verdes sendo as esperadas para 6SG e azuis para 4SU). Posições referem-se à montagem GRCh38 do genoma humano. (C) Gráfico de conservação evolutiva das regiões, coloração rosa claro denota conservação acima de 50% em relação a sequência humana, em exons ou introns. Linhas tracejadas verticais isolam a região de interesse. (D) Motivos de HuR propostos por Mukherjee *et al*<sup>97</sup> e o compatível observado na região com homologia entre as três espécies e com maior acúmulo de fragmentos suportando a interação. (E) Alinhamento da região selecionada através do PAR-CLIP com sequências homólogas de diversos vertebrados e guiada pela análise VISTA. Cores indicam conservação do nucleotídeo nas sequências avaliadas. Blocos acima representam os fragmentos do PAR-CLIP e as mutações (em azul). Destacados ao redor das sequências indicam possíveis sítios de ligação; em verde, a região rica em AU mais provável de ser o sítio de interação de HuR no intron 14 de *GLS*. (F) Região de

*branchpoint* mais provável<sup>280</sup> frente a zona de exclusão de AG<sup>281</sup> para mediação do *splicing* alternativo e a distância da região de ligação de HuR para o mesmo.

Após a análise destes conjuntos de dados, observamos nos fragmentos sequenciados alinhados ao intron 14 as mutações diagnósticas do crosslink em experimentos de PAR-CLIP-Seq, de forma mais evidente nos dados de Lebedeva *et al*<sup>96</sup> com a utilização de 6SG. (Figura 29B). A região de ligação se mostra conservada entre os genomas de mamíferos disponíveis na plataforma VISTA, mais um elemento indicativo de possível importância regulatória (Figura 29C). No conjunto de fragmentos (especialmente suportado pelos dados de Lebedeva *et al*<sup>96</sup>), encontramos um dos motivos de ligação propostos para HuR (Figura 29D), especialmente com a observação da troca de G para A esperada na presença de 6SU (ocorrendo justaposta ao motivo). A informação estrutural disponível aponta 11 nucleotídeos participando diretamente da interação com os dois primeiros domínios de ligação à RNA de HuR<sup>83</sup>, o que deixa claro o motivo de ligação como um *core* da sequência, sendo flanqueada por sequências menos conservadas mas que participam da interação.

Quando comparada às demais sequências de vertebrados homólogas ao intron 14 de glutaminase (Figura 29E), o sítio hipotético de ligação demonstra conservação especialmente das bases timina (que são transcritos em uracila no RNA e estão diretamente relacionados com a ligação de HuR, segundo os motivos estimados na literatura<sup>282</sup>). Localizando espacialmente este sítio em relação às estruturas de *splicing* alternativo, o sítio de ligação de HuR proposto está 159 pares de bases distante da sequência predita<sup>280</sup> de *branchpoint* – na qual a porção 5' clivada do intron estabelece uma ligação covalente, prévia à clivagem da porção 3' e ligação entre os dois exons<sup>283</sup>. Desta maneira, concluímos que HuR interage com o intron 14 de *GLS*, mas não em região diretamente ligada ao *splicing*, com mecanismo não relacionado a uma simples oclusão física do sítio de *branching*.

**Tabela 9 – Nível de significância dos elementos relacionados com *splicing* afetados por HuR do gene *GLS***

Dado	Elemento	Bootstrap (%)	<i>p-value</i>	FDR
RIP-Seq	Intron 14	NA	0,011408	0,038609
	Intron 15	NA	1,18E-22	1,41E-20
RNA-Seq	Exon 15	92,2		0,010431
	Exon 16	45,1		0,100961
	Exon 17	82,4		0,030563
	Exon 18	94,1		0,013694
	Exon 19	100		1,92E-14

NOTA: Dados de RIP-Seq referentes à linhagem GM12878. NA: Não-aplicável. Para RNA-Seq, FDR médio entre todas as pseudoreplicatas.

### 4.3 HISTÓRIA EVOLUTIVA DOS GENES *GLS* E *GLS2*

Com o interesse de avaliar o surgimento evolutivo do exon 15 e, conseqüentemente, da isoforma GAC originada por *splicing* alternativo, buscamos a possível origem do evento de translocação deste exon para a região onde ele é encontrado em vertebrados. Reconstruímos a estrutura exônica da glutaminase metazoária (Figura 30A) utilizando como estratégia de busca dos genes no filo Chordata utilizando a sequência das proteínas codificadas pelos genes *GLS* e *GLS2*. Em um segundo momento, apenas a região correspondente ao exon 14 (último comum a ambas isoformas) e do exon 15 (exclusivo de GAC) foi utilizada para a busca de elementos homólogos nos genes identificados.

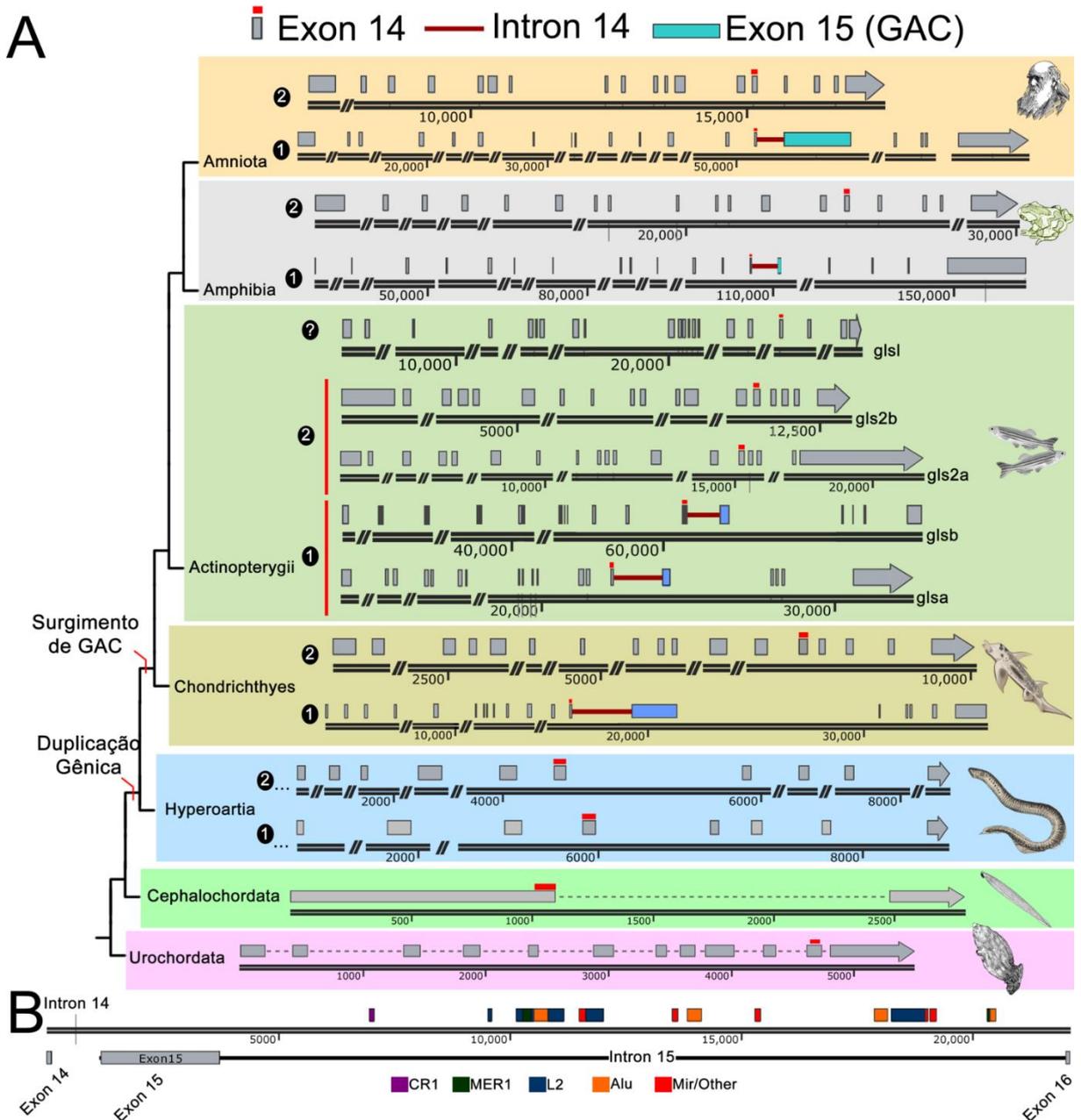
Na base evolutiva dos cordados, *Ciona intestinalis*<sup>194</sup> apresenta apenas uma sequência similar a *GLS/GLS2* contendo 12 exons, distribuição completamente distinta da atual, e o exon 14 (último comum a KGA e GAC) presente na região interna de outro exon. Desta maneira, esta sequência não aparenta gerar troca de isoformas por *splicing* alternativo.

O cefalocordado anfioxo (*Branchiostoma floridae*<sup>195</sup>), por sua vez, apresenta uma re-organização estrutural gênica abrupta, compactando o gene para apenas dois exons. Tal evento não envolveu apenas exon *shuffling*, comum na evolução dos eucariotos<sup>284-286</sup>, mas também uma reestruturação intrônica completa. Entretanto, a partir de então, já começamos a detectar características homólogas e permanentes, como a junção de *splicing* da região homóloga ao exon 14 – grosseiramente – ao final do domínio catalítico, evento comum evolutivamente<sup>285</sup>. Na região a jusante da mesma não fora observada homologia ao exon exclusivo de GAC.

O animal modelo de Hyperoartia, a lampreia-Marinha (*Petromyzon marinus*<sup>196</sup>), tem o genoma extremamente incompleto, a ponto de não incluir o gene *GLS* completo em nenhum dos *contigs* ou *scaffolds* disponíveis. Entretanto, buscando individualmente os exons através de *Reciprocal BLAST Best Hits* (RBBH<sup>192</sup>) foi possível identificar dois fragmentos com homologia a glutaminase, nenhum dos mesmos contendo homologia ao exon 15, mas mantendo o padrão de exons-introns de cefalocordados e sem homologia completa entre os mesmos, denotando a duplicação gênica de *GLS* em o que atualmente designamos *GLS* e *GLS2*.

Os peixes cartilagosos, representados pelo tubarão-elefante (*Callorhynchus milii*<sup>197</sup>), são os primeiros a possuir o exon diferencial a GAC (mesmo que não identificado na sua anotação genômica original) e, provavelmente, a regulação por *splicing* alternativo do gene. Para os ósseos, representados pelo peixe-zebra (*Danio rerio*<sup>198</sup>), diversos eventos de duplicação ocorreram, mas se mostram como uma autapomorfia do grupo, dado que não foram mantidos

no clado irmão. As 5 cópias observadas sugerem a duplicação de *GLS* e *GLS2*, além de uma cópia não completamente homóloga, mas similar a *GLS2*, dado a ausência do exon exclusivo de GAC.



**Figura 30 – História evolutiva proposta para a glutaminase metazoária.**

(A) Representação filogenética da história evolutiva do gene *GLS*, com os eventos de *exon shuffling* e o surgimento o exon 15 (retângulos azuis) e, conseqüentemente, do intron 14 (linhas bordô) e 15, usando a numeração humana como referência. Barras verticais vermelhas denotam a duplicação autapomórfica de cada um dos genes de glutaminase em *Danio rerio* (Actinopterygii). Pontilhado em *Petromyzon marinus* (Hyperoartia) indica a incompleta seqüência do gene disponível em contigs gerados pelo sequenciamento do genoma do organismo. (B) Busca por elementos retrotransponíveis na região genômica que compreende o exon 15 e os introns 14 e 15 da glutaminase humana usando a ferramenta *Transposgene*<sup>287</sup>. As imagens utilizadas, créditos e respectivos direitos de uso são: Darwin, por Chandler B. Bleach, em domínio público; sapo, por VxD, em CC-BY-SA-2.5; peixe-zebra, por DataBase Center for Life Science, em CC-BY-4.0; peixe-elefante, por George Raper, em domínio

público; lampreia, por F. Barthélemy, em CC-BY-SA-3.0; anfioxo, por V. Gewin, em CC-BY-SA-3.0 e urocordado, por Wellcome Library, CC-BY-4.0.

Para o representante anfíbio, *Xenopus tropicalis*<sup>199</sup>, observamos uma estrutura gênica similar a humana, com uma única cópia de *GLS* e *GLS2*, sendo o segundo compactado em uma região genômica menor e o primeiro espalhado em um locus com mais quilobases. Nos humanos<sup>200,201</sup>, a estrutura é a mesma observada em anfíbios, apenas com um tamanho de sequência final reduzido.

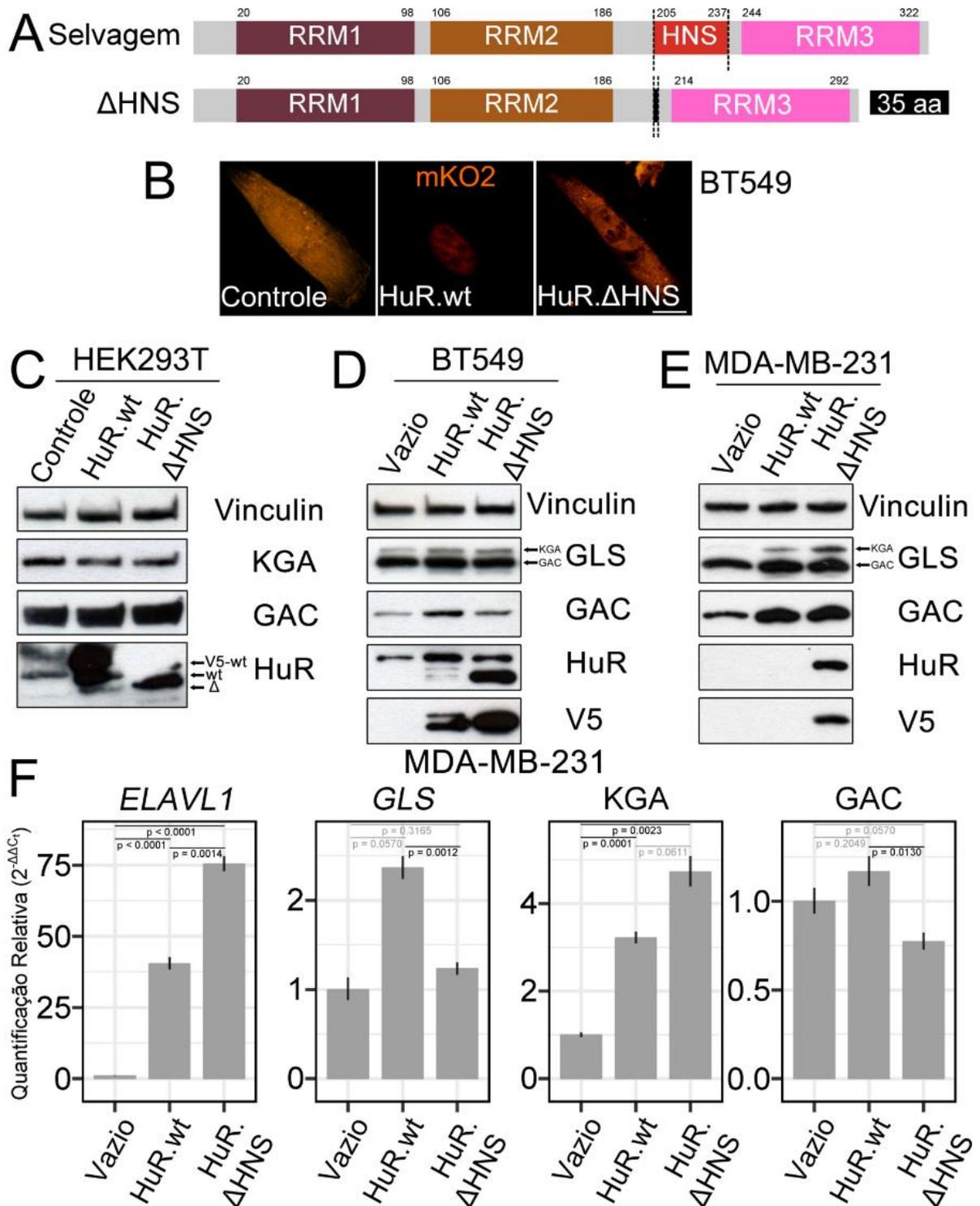
Diversos elementos transponíveis, como Alu e L2 (Figura 30B) foram observados na região do intron 15 da *GLS* humana, sugerindo que a inserção do exon de GAC foi consequência de um evento de retrotransposição que gerou uma “exonização” de sequência não codificante<sup>288</sup>. Entretanto, buscas por blastn<sup>190</sup> dentro do genoma humano foram incapazes de apontar uma possível origem do exon 15. Cabe ressaltar que as observações evolutivas aqui observadas foram publicadas no trabalho do **Anexo 2**<sup>32</sup>.

Por fim, avaliando a região do intron 14 no projeto 1000genomes<sup>289</sup>, apenas três variantes são detectadas nesta região: rs534382779 (A para G), rs753576052 (A para G) e rs545013471 (C para A). As variantes são encontradas em frequência alélica diminuta, sempre em apenas uma única observação, não evidenciando nenhuma clara relação com algum fenótipo e frustrando análises posteriores de correlação.

#### 4.4 AVALIAÇÃO EXPERIMENTAL DA RELAÇÃO ENTRE HUR E A EXPRESSÃO DAS ISOFORMAS DE GLUTAMINASE

##### 4.4.1 Consequências da modulação da expressão de *ELAVL1* nos níveis de GAC e KGA em linhagens celulares de câncer de mama

Duas construções foram geradas para a expressão de *ELAVL1*: selvagem (denominada HuR.WT) ou com a deleção da sequência de translocação núcleo-citoplasma (denominada HuR.ΔHNS), conhecida por não estar envolvida diretamente na interação com o RNA<sup>69</sup> (Figura 31A). A construção com deleção da região HNS, reguladora da localização núcleo-citoplasma<sup>85</sup>, foi empregada de maneira a distribuir a proteína entre os dois compartimentos e melhor observar os efeitos citoplasmáticos de sua ação como estabilizadora de mRNA.



**Figura 31** – Expressão ectópica de *ELAVL1* e alterações observadas sobre o nível proteico e de mRNA das isoformas de glutaminase

(A) Esquema com a região de controle núcleo-citoplasma deletada. (B) Microscopia confocal de fluorescência de células transduzidas e selecionadas com a expressão de *ELAVL1* (WT ou  $\Delta$ HNS) fusionada à proteína fluorescente monomérica *kusabira-orange2*. Barra de escala de 10  $\mu$ m. (C) Western blot após expressão transitente de HuR *ELAVL1* (WT ou  $\Delta$ HNS) em células HEK293T. Setas em HuR indicam o tamanho esperado para cada construção. Western blot após transdução viral e seleção de pool celular expressando HuR (WT ou  $\Delta$ HNS) em célula BT549 (D) ou MDA-MB-231 (E). Setas em *GLS* distinguem as isoformas pelo peso molecular. (F) PCR quantitativa da célula MDA-MB-231 após a expressão de HuR (WT ou  $\Delta$ HNS). Barras representam erro padrão da média de triplicatas técnicas, *p-values* originados de teste-t de comparação de médias entre os valores de  $\Delta$ Ct.

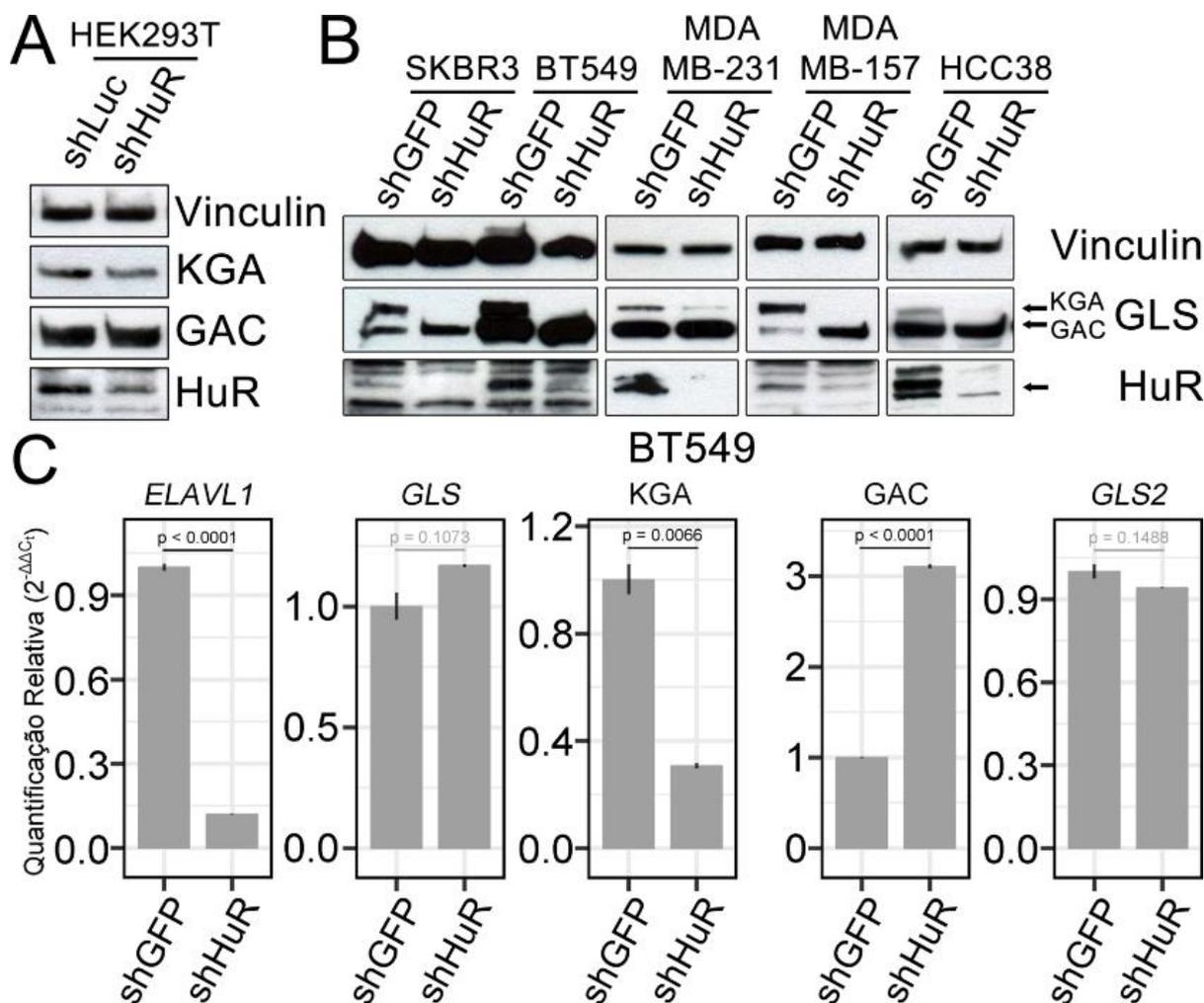
Primeiramente, confirmamos em BT549, com a expressão de fusões fluorescentes de HuR, que as proteínas estavam se localizando nos compartimentos esperados, com HuR.WT majoritariamente no núcleo e HuR. $\Delta$ HNS distribuído pelo núcleo e pelo citoplasma (Figura 31B). Expressamos então *ELAVLI* (HuR.WT ou HuR. $\Delta$ HNS) ectopicamente e de maneira transiente em células HEK293T. Apesar do sucesso na expressão das proteínas (Figura 31C), não observamos notada alteração nos níveis proteicos de GAC e KGA (uma ligeira diminuição em KGA). Curiosamente, observamos redução da expressão da isoforma endógena apenas quando da expressão ectópica desta construção deletada para a região de translado, o que era esperado em ambas dado o *feedback* regulatório negativo de HuR sobre ela própria<sup>290</sup>.

Testamos então um sistema estável de expressão nas linhagens de câncer de mama BT549 e MDA-MB-231, obtido através de transdução viral. Na célula BT549, a expressão de HuR.WT e HuR. $\Delta$ HNS causou um ligeiro aumento em ambas as isoformas (Figura 31D), mais saliente para GAC na presença de HuR.WT ectópica. Em linhas gerais, os mesmos efeitos foram observados em MDA-MB-231 (Figura 31E).

Apesar de HuR.WT não ter sido observada por *western blot* em MDA-MB-231, observamos um aumento de 30-75 vezes nos níveis do transcrito de *ELAVLI* em células expressando a construção selvagem, confirmando a expressão ectópica da mesma (Figura 31F). Os resultados de qPCR mostraram também que, quando na presença de HuR.WT e HuR. $\Delta$ HNS, houve aumento do mRNA de KGA e não alteração a diminuição do nível de GAC (Figura 31F). A não alteração no nível de mRNA de GAC, apesar de verificarmos aumento no nível da proteína, pode implicar um papel de HuR em algum mecanismo de regulação pós-transcricional relacionado com modulação da tradução do mRNA<sup>211</sup>.

Uma vez observado que a expressão ectópica de *ELAVLI* levava a alteração nos níveis das isoformas, fizemos seu silenciamento nas células. Em HEK293T, o silenciamento de *ELAVLI* levou à redução dos níveis da isoforma KGA (Figura 32A). Seis linhagens de câncer de mama (SKBR3, BT549, MDA-MB-231, MDA-MB-157 e HCC38) foram então avaliadas após o silenciamento estável gerado pela transdução com o sistema viral pLKO. Inicialmente, verificamos que houve diminuição no nível de HuR (Figura 32B), como esperado. Nas seis linhagens avaliadas, o silenciamento de *ELAVLI* foi acompanhado da diminuição de KGA, e em SKBR3 e MDA-MB-157 foi observado concomitante aumento de GAC (Figura 32B). Para os níveis de mRNA de BT549, confirmamos uma redução de 7 vezes em *ELAVLI* acompanhada de uma redução de KGA e aumento de GAC (Figura 32C) – sem impacto nos níveis globais de

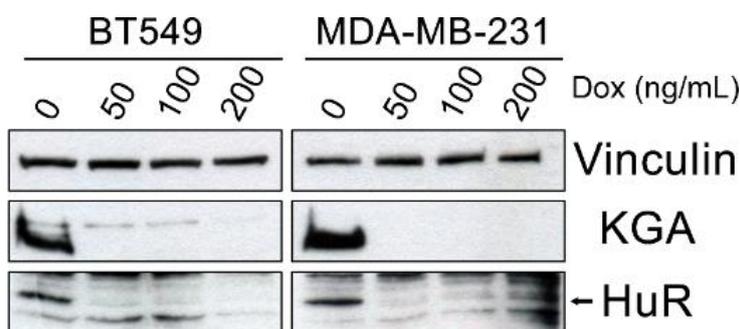
*GLS* ou de seu parálogo, *GLS2*. Tais resultados corroboram o papel de HuR no *splicing* alternativo e troca de isoformas de *GLS*.



**Figura 32** – Consequências do silenciamento de *ELAVL1* sobre o nível proteico e de mRNA das isoformas de glutaminase

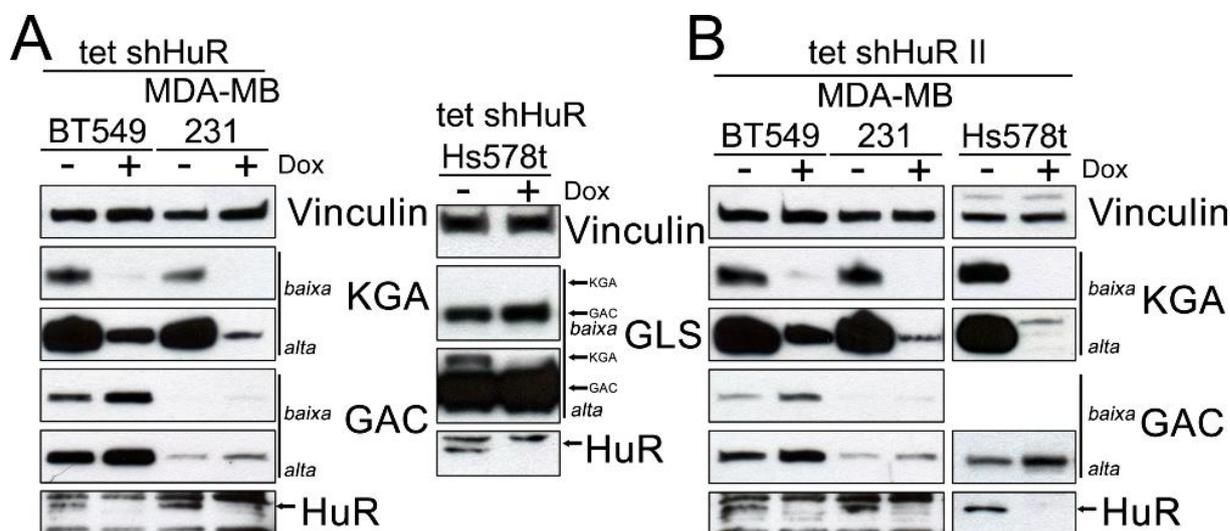
(A) *Western blot* após silenciamento transitente de *ELAVL1* utilizando células HEK293T. (B) *Western blot* após silenciamento constitutivo de *ELAVL1* utilizando *pools* celulares transduzidos com o sistema pLKO-shHuR ou controle. Setas em *GLS* distinguem as isoformas pelo peso molecular e a banda com o peso molecular esperado para HuR. (C) PCR quantitativa da célula BT549 após o silenciamento constitutivo de *ELAVL1*. Barras representam erro padrão da média de triplicatas técnicas, *p-values* originados de teste-t de comparação de médias entre os valores de  $\Delta C_t$ .

Por fim, um sistema de silenciamento induzível (tet-pLKO) foi utilizado para avaliar os efeitos de HuR sobre GAC e KGA. Doses de 50, 100 e 200 ng/mL de hiclato de doxiciclina (mais estável que a molécula original de uso com o sistema, tetraciclina) foram avaliadas. Verificamos que 50 ng/mL já foi suficiente não só para silenciar a expressão de HuR, mas também para impactar nos níveis de KGA (Figura 33).



**Figura 33 – Avaliação da dose de doxiciclina necessária para ativação do sistema de silenciamento induzível de HuR**  
Western blot de células transduzidas com o sistema tet-pLKO expressando shRNA para *ELAVL1* tratadas com concentrações crescentes de doxiciclina e mantidas na concentração indicada por 6 dias. Seta indica a banda com o peso molecular esperado.

De posse deste sistema, avaliamos o efeito do silenciamento de *ELAVL1* utilizando uma segunda sequência de shRNA almejando a região 3'UTR do transcrito de *ELAVL1*. A eficácia do silenciamento de *ELAVL1* com a sequência original (shHuR) foi observada nas células BT549, MDA-MB-231 e uma linhagem não testada anteriormente, Hs578t, resultando na redução dos níveis de KGA em todas e aumento dos níveis de GAC nas duas primeiras (Figura 34). Quando da utilização da sequência alternativa, shHuR II, o silenciamento de HuR foi observado nas três linhagens após a adição de doxiciclina, concomitante a redução dos níveis de KGA e aumento dos níveis de GAC.

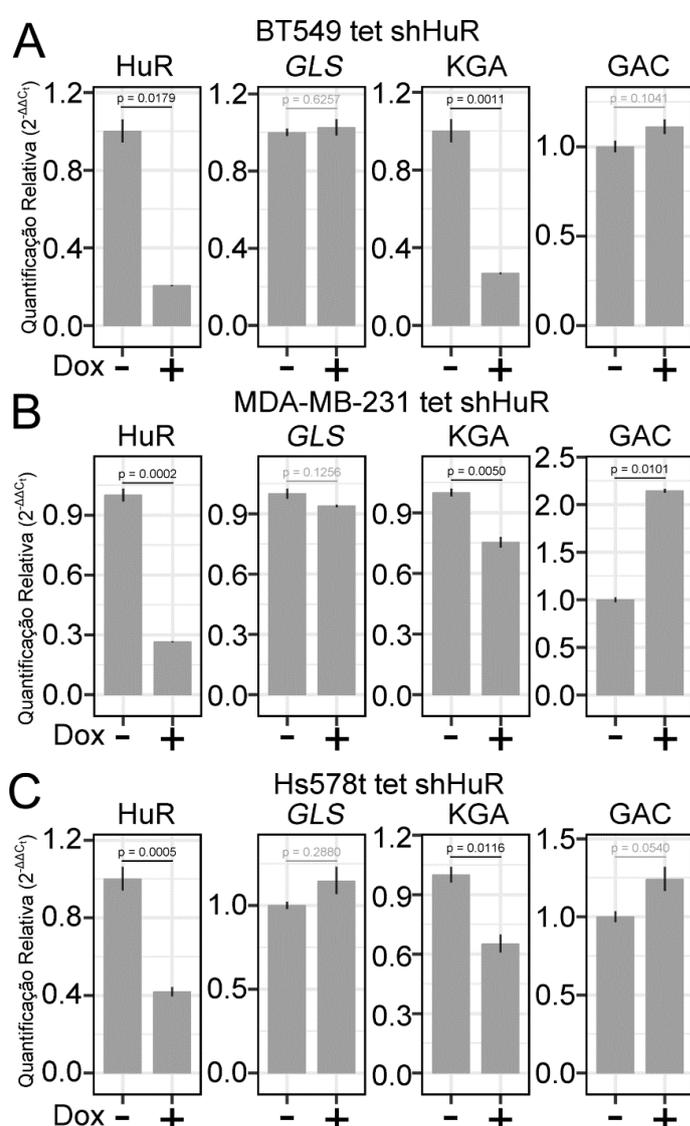


**Figura 34 – Consequências do silenciamento de *ELAVL1* em sistema induzível sobre o nível proteico das isoformas de glutaminase**

(A) Western blot de células BT549, MDA-MB-231 e Hs578t transduzidas com o sistema tet-pLKO-shHuR após a indução do silenciamento de *ELAVL1* com doxiciclina. (B) Western blot das células BT549, MDA-MB-231 e Hs578t após a indução do silenciamento de *ELAVL1* com doxiciclina com sequência de shRNA alternativa (shHuR II). Setas em *GLS* distinguem as isoformas pelo peso molecular. Seta em HuR indica a banda com o peso molecular esperado. Alta e baixa indicam exposições distintas.

Os níveis de mRNA foram avaliados nas linhagens BT549 (Figura 35A), MDA-MB-231 (Figura 35B) e Hs578t (Figura 35C) no sistema tet-pLKO. O silenciamento de HuR

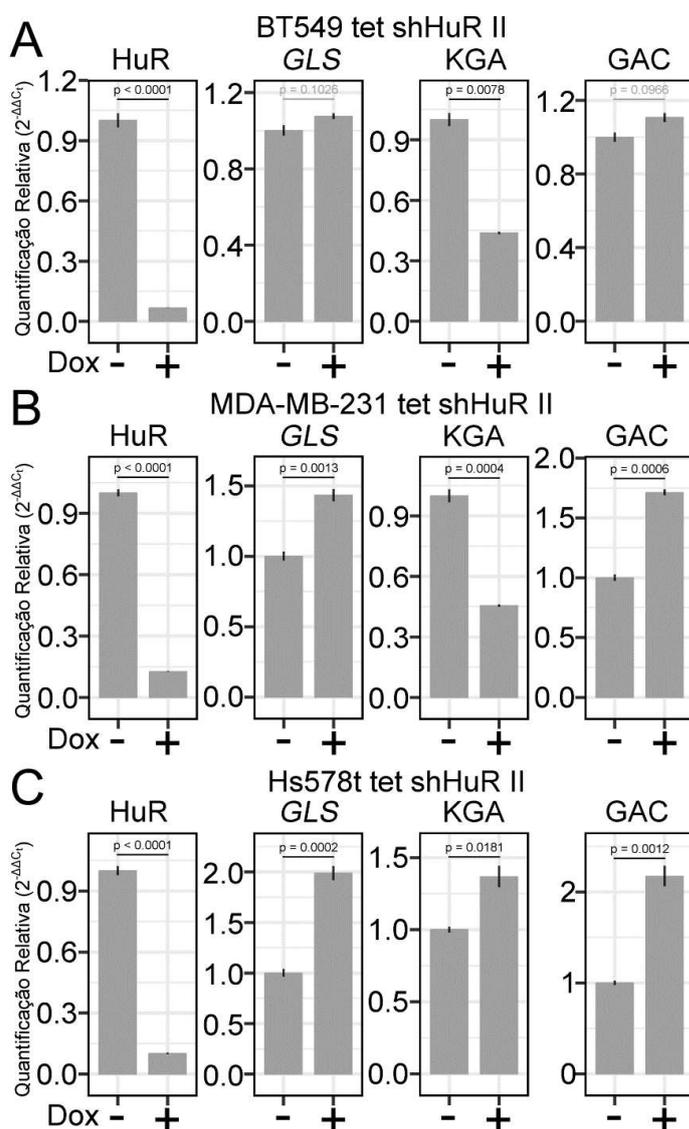
foi observado nas três linhagens celulares, com reduções de 5 (BT549), 3,5 (MBA-MB-231) e 2,5 (Hs578t) vezes em relação ao nível basal de expressão das células, sendo significativo em todos os cenários. Os níveis globais de *GLS* não sofreram alteração de maneira significativa, tal qual no sistema constitutivo. Quando observadas as isoformas de maneira independente, KGA apresentou reduções significativas nas três linhagens avaliadas, mas de maneira mais intensa na célula BT549 (3,3 vezes, contra cerca de 1,3 em MDA-MB-231 e Hs578t). Já GAC teve seu nível aumentado de forma significativa em MDA-MB-231, mas não alterada em BT549 e Hs578t.



**Figura 35 – Silenciamento de *ELAVL1* em sistema induzível com sequência shHuR e o nível de mRNA das isoformas de glutaminase** PCR quantitativa utilizando células BT549 (A), MDA-MB-231 (B) e Hs578t (C) transduzidas com o sistema tet-pLKO-shHuR após a indução do silenciamento de *ELAVL1* com doxiciclina. Barras representam erro padrão da média de triplicatas técnicas, valores de *p* originados de teste-t de comparação de médias entre os valores de  $\Delta Ct$ .

Utilizando a sequência alternativa de shRNA nas linhagens de câncer de mama BT549 (Figura 36A), MDA-MB-231 (Figura 36B) e Hs578t (Figura 36C), verificamos uma eficiência de silenciamento superior, com redução do nível basal de expressão de cada célula

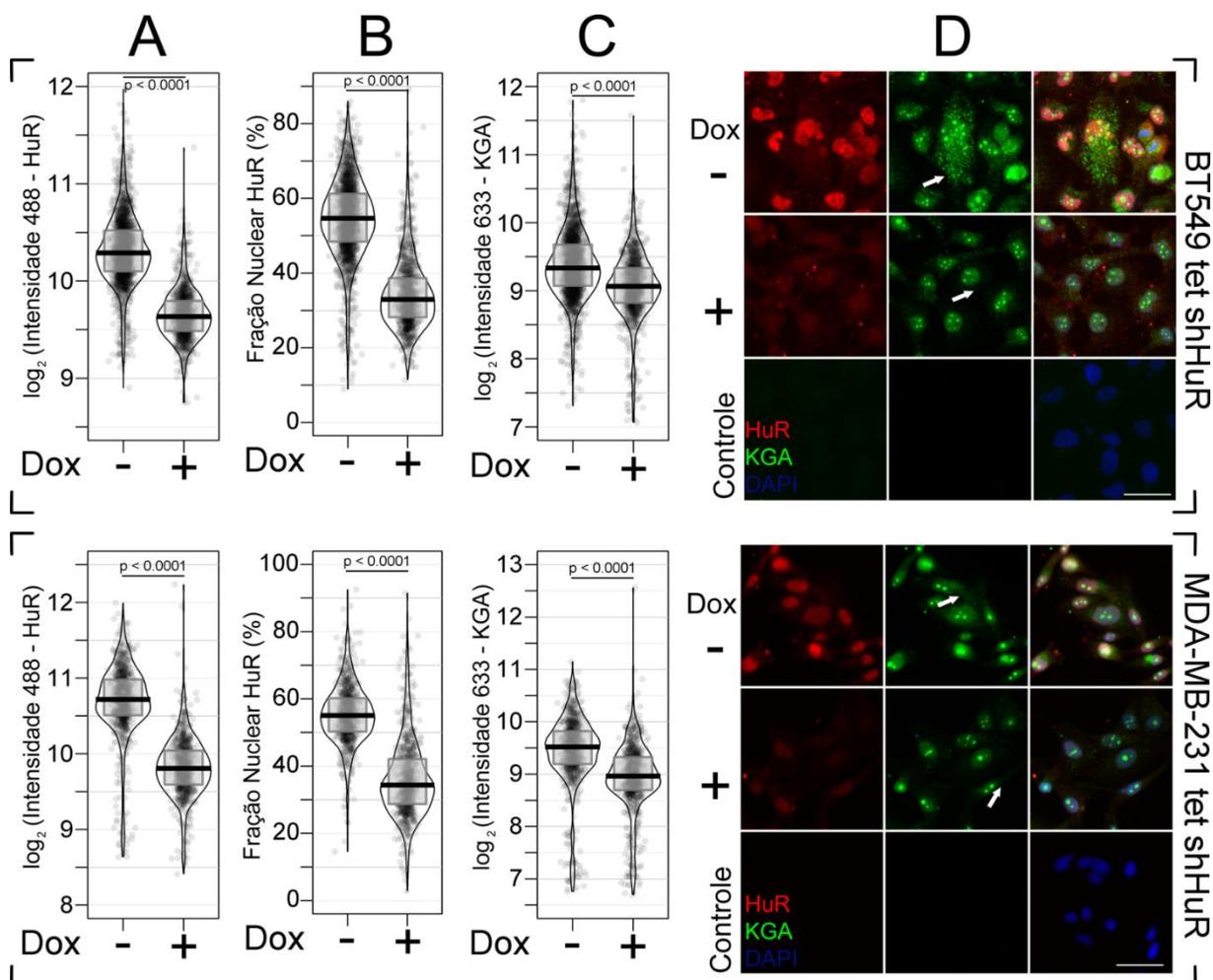
em 6,5 vezes, tendo sido esta sequência, de maneira geral, mais eficiente que shHuR. Por razões desconhecidas, alterações foram vistas no nível global do mRNA de *GLS* em MDA-MB-231 e Hs578t. Enquanto que em BT549 e MDA-MB-231, shHuRII levou a diminuição do mRNA para KGA (cerca de duas vezes), em Hs578t, curiosamente, houve elevação dos níveis de KGA (em 1,3 vezes). O nível do mRNA de *GAC*, por outro lado, como esperado, aumentou nas 3 linhagens, tendo sido este aumento mais pronunciado em MDA-MB-231 (1,7 vezes) e Hs578t (2,1 vezes) do que em BT549 (1,1 vez, não significativo).



**Figura 36 – Silenciamento de *ELAVL1* em sistema induzível com sequência shHuRII e o nível de mRNA das isoformas de glutaminase PCR quantitativa utilizando células BT549 (A), MDA-MB-231 (B) e Hs578t (C) transduzidas com o sistema tet-pLKO-shHuRII após a indução do silenciamento de HuR com doxíciclina. Barras representam erro padrão da média de triplicatas técnicas, valores de  $p$  originados de teste-t de comparação de médias entre os valores de  $\Delta Ct$ .**

Por fim, avaliamos a variação dos níveis proteicos de HuR e KGA por imunomarcacão dupla nas linhagens BT549 e MDA-MB-231 transduzidas com o sistema induzível e expostas a doxíciclina (Figura 37). Como esperado, observou-se uma reduço significativa na intensidade media da marcaço de HuR geral e nuclear em BT549 (Figura 37A,

superior) e MDA-MB-231 (Figura 37, inferior) quando expostas a doxiciclina. De acordo, o silenciamento de *ELAVL1* levou a diminuição no sinal de imunofluorescência medido para KGA (Figura 37C). Especialmente, reduziu-se a marcação citoplasmática de KGA (Figura 37D, setas brancas), de forma não tão evidente na marcação nuclear.



**Figura 37 – Avaliação por imunofluorescência do silenciamento de *ELAVL1* em sistema induzível e alteração em KGA**

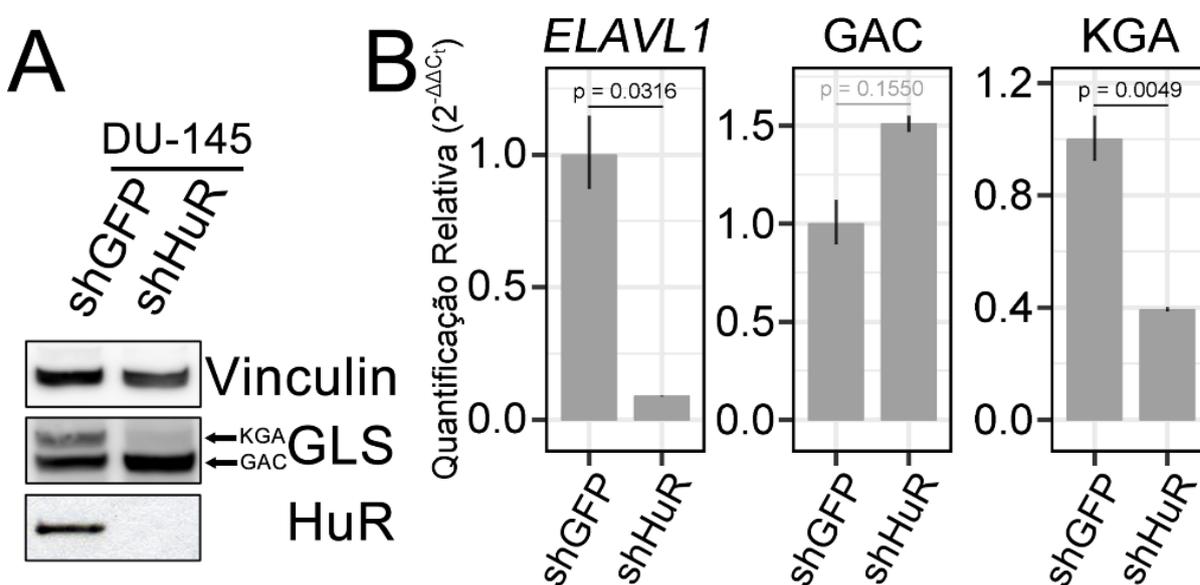
(A) Gráfico RDI da intensidade de imunomarcção de HuR por fluorescência com ou sem adição de doxiciclina. (B) Gráfico RDI da fração de intensidade de fluorescência identificada na região determinada como núcleo em relação a região total da célula com ou sem adição de doxiciclina. (C) Gráfico RDI da intensidade de imunomarcção de KGA por fluorescência observada com ou sem adição de doxiciclina. (D) Imagens representativas das microscopias. Controle indica a não adição de anticorpo primário. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os valores das células quantificadas estão individualmente representados sobre os dados como círculos. Valores de  $p$  originados de teste-t de comparação de médias. Barras de escala representam  $50\mu\text{m}$ . Setas brancas indicam regiões citoplasmáticas com menor marcação.

Em conclusão, verificamos que a expressão ectópica de HuR eleva os níveis de KGA e, de maneira concordante, o silenciamento de HuR reduz a abundância da isoforma; a alteração em GAC entretanto é variável, tendo sido observado, nos modelos de silenciamento,

o aumento desta isoforma acompanhado à diminuição de KGA. A manutenção de HuR no citoplasma pelo uso da construção HuR. $\Delta$ HNS, enquanto não conclusiva para GAC, aumentou os níveis de KGA.

#### 4.4.2 Consequências da modulação da expressão de *ELAVL1* nos níveis de GAC e KGA em linhagens celulares de câncer de próstata

Em próstata, verificamos que o Cox HR para alta expressão de *ELAVL1* foi 1,85 (FDR de 0,0366, Tabela 4), mostrando um prognóstico negativo para o paciente. Sendo este um tipo tumoral de grande incidência em homens (representando 28,6% dos casos de tumores masculinos, o mais incidente para este gênero no Brasil<sup>291</sup>) e para o qual possuíamos linhagens celulares no laboratório, avaliamos a importância de HuR para a escolha das isoformas de *GLS* no mesmo. Assim, a linhagem celular de câncer de próstata DU-145 foi transduzida com o sistema de silenciamento constitutivo pLKO shHuR (Figura 38). No caso, foi utilizada a sequência-alvo presente a porção de sua região codificante (CDS, designada shHuR).



**Figura 38 – Efeito do silenciamento de *ELAVL1* sobre o nível de mRNA e proteína das isoformas de glutaminase na linhagem de câncer de próstata DU-145**

(A) *Western blot* após silenciamento constitutivo de *ELAVL1* utilizando *pools* celulares transduzidos com o sistema pLKO-shHuR ou controle. Setas em *GLS* distinguem as isoformas pelo peso molecular. (B) PCR quantitativa de DU-145 após o silenciamento constitutivo de HuR. Barras representam erro padrão da média de duplicatas ou triplicatas técnicas, *p-values* originados de teste-t de comparação de médias entre os valores de  $\Delta C_t$ .

Verificamos que, em relação às células controle (expressando shGFP), houve diminuição da banda proteica de HuR (Figura 38A), confirmado pelo qPCR (Figura 38B,

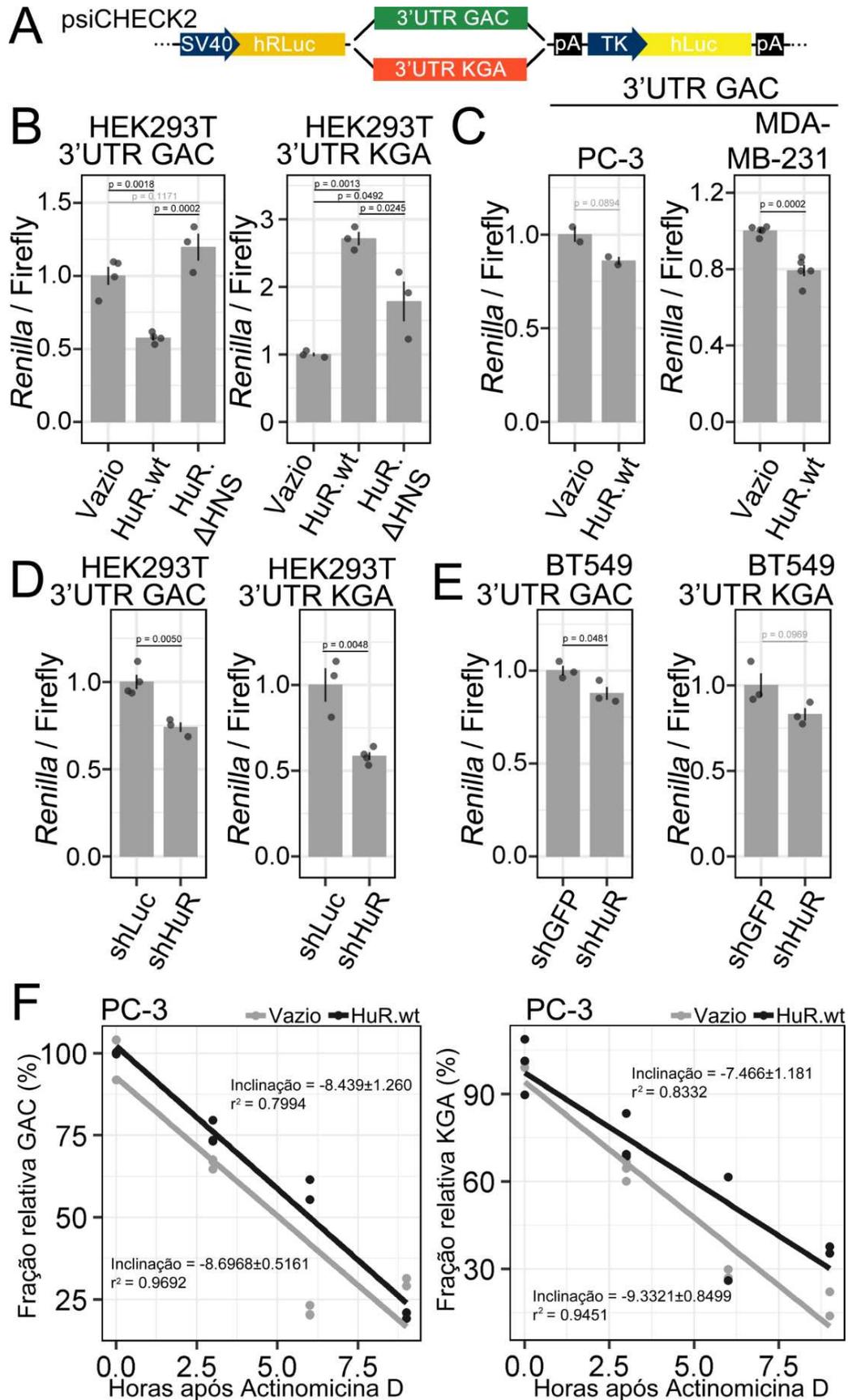
diminuição de 11,2 vezes). O silenciamento de *ELAVL1* reduziu os níveis proteicos de KGA, com aumento dos níveis de GAC (Figura 38A), o que foi corroborado por qPCR (Figura 38B).

Desta maneira, confirma-se em um segundo tipo tumoral o papel de HuR no controle das isoformas de *GLS*, reforçando sua ação em outros tecidos e potencial importância global neste processo.

#### 4.4.3 Efeitos da modificação dos níveis de HuR sobre a estabilidade do mRNA das isoformas de glutaminase

Conforme colocado na Introdução, HuR age no metabolismo de RNA tanto no processo de *splicing* sobre pré-mRNA, quanto na estabilidade de mRNA, pela sua interação com a região 3'UTR dos mesmos. Com o intuito de observar a regulação das isoformas de glutaminase mediada pela interação de HuR com suas respectivas 3'UTR, foram construídos sistemas repórteres baseados na família de vetores psiCHECK2. Basicamente, como ilustra o esquema (Figura 39A), o sistema consiste na expressão, sob o controle de um promotor do vírus vacuolante símio 40 (SV40), da sequência codificante humanizada da luciferase derivada de rim-do-mar (*Renilla reniformis*), seguida da 3'UTR das isoformas KGA ou GAC. Para normalizar distintas eficiências de transfecção entre as células avaliadas o mesmo vetor expressa, sob o controle do promotor da timidina cinase do vírus do herpes simples (TK), sequência codificante humanizada da luciferase derivada de vagalume (do inglês, *firefly*, sendo derivada de *Photinus pyralis*)<sup>292</sup>. As duas enzimas filogeneticamente distantes utilizam substratos distintos, luciferina e coelenterazina, o que permite a detecção da atividade das duas enzimas em um mesmo experimento<sup>292</sup>.

Em células HEK293T co-transfectadas com o vetor repórter para a 3'UTR de GAC e vetores para expressão (de HuR.WT ou o mutante deletional HuR.ΔHNS) verificamos que a expressão de HuR.WT (mas não de HuR.ΔHNS) levou a redução da proporção de sinal de *Renilla/Firefly* (Figura 39B, esquerda). Para o vetor repórter com a 3'UTR de KGA, HuR.WT e HuR.ΔHNS levaram a um aumento na razão *Renilla/Firefly* (Figura 39B). A transfecção do vetor repórter em MDA-MB-231 e PC-3 estavelmente expressando HuR.WT também levou a diminuição ( $p = 0,0018$  e  $0,0894$ , respectivamente, significativa apenas para MDA-MB-231) na razão *Renilla/Firefly* gerada pelo mRNA contendo a 3'UTR de GAC (Figura 39C), confirmando os achados em HEK293T (Figura 39B, esquerda). O mesmo não foi avaliado para a 3'UTR de KGA em MDA-MB-231 e PC-3.



**Figura 39 – Mensuração da estabilidade conferida por HuR via porções não traduzidas terminais do mRNA das isoformas de glutaminase**

(A) Esquema representativo dos plasmídeos gerados para a avaliação da estabilidade conferida pelo 3'UTR do mRNA de KGA ou GAC. Promotor SV40 coordena a expressão de mRNA codificando luciferase derivada de

Rim-do-mar (*Renilla*) contendo um 3'UTR curto (quadrado pequeno cinza) (controle), ou 3'UTR de GAC (retângulo verde) ou KGA (retângulo vermelho). Concomitantemente, promotor mínimo expressa luciferase derivada de vagalume (*Firefly*) com o intuito de normalização de transfecção. Aumentos na razão da expressão *Renilla / Firefly* implicam em aumento da estabilidade ou da tradução do mRNA. (B) Avaliação dos efeitos da expressão ectópica de *ELAVL1* na estabilidade/tradução mediada pelo 3'UTR de GAC e KGA na célula HEK293T. (C) Avaliação dos efeitos da expressão ectópica de *ELAVL1* na estabilidade mediada pelo 3'UTR de GAC nas células PC-3 e MDA-MB-231. Avaliação dos efeitos do silenciamento de *ELAVL1* na estabilidade mediada pelo 3'UTR de GAC e KGA na célula HEK293T (D), e BT549 (E). (F) Experimento de decaimento do mRNA de GAC (esquerda) ou KGA (direita) com inibição da tradução por actinomicina D em células PC-3. Valores próximos às curvas indicam inclinação das regressões lineares e o ajuste de  $r^2$ . Para gráficos de barras, as mesmas representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico como pontos individuais. Valores de  $p$  originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias.

Avaliamos também o efeito do *knock down* de *ELAVL1* sobre o vetor repórter. Em HEK293T, a supressão de HuR diminuiu a razão *Renilla/Firefly* gerada pelo mRNA contendo a 3'UTR tanto de GAC quanto de KGA (Figura 39D). A coerência com a superexpressão se deu apenas em KGA (Figura 39B), mas não GAC. Tal observação aponta para potenciais outros fatores regulatórios não avaliados nestes experimentos. Em BT549 os resultados se confirmaram: *knock down* de *ELAVL1* diminuiu a razão *Renilla/Firefly* do mRNA contendo a 3'UTR de GAC e KGA, não significativo para a segunda (Figura 39E).

Para confirmarmos os achados, realizamos, em PC-3, ensaio de decaimento do mRNA endógeno de GAC, após a adição de actinomicina, um bloqueador da transcrição de RNA. Verificamos que a presença ectópica de HuR.WT não afetou a estabilidade do mRNA de GAC (Figura 39E, esquerda), enquanto que o mesmo aumentou a estabilidade do mRNA de KGA (Figura 39E, direita). Em células PC-3 expressando *ELAVL1*, o decaimento do mRNA endógeno de KGA não só apresentou redução na velocidade (indicada pela redução da inclinação da reta, linha escura, Figura 39E, direita), mas como também apresentou um aumento do  $-\Delta Ct$ , o que reforça o impacto positivo de HuR sobre este mRNA.

Com o conjunto dos dados, concluímos que, enquanto HuR apresenta um papel claro no aumento da estabilidade do mRNA de KGA, seu papel sobre o mRNA de GAC mostrou-se de diminuição de estabilidade na expressão ou no silenciamento de *ELAVL1*, sendo nossos resultados pouco conclusivos. Sendo este processo predominantemente citoplasmático e considerando que a forma selvagem de HuR é mais pronunciadamente nuclear (Figura 31B), é curioso que a expressão de HuR. $\Delta$ HNS (com maior presença citoplasmática) não tenha aumentado ainda mais estabilização do mRNA repórter contendo a 3'UTR de KGA.

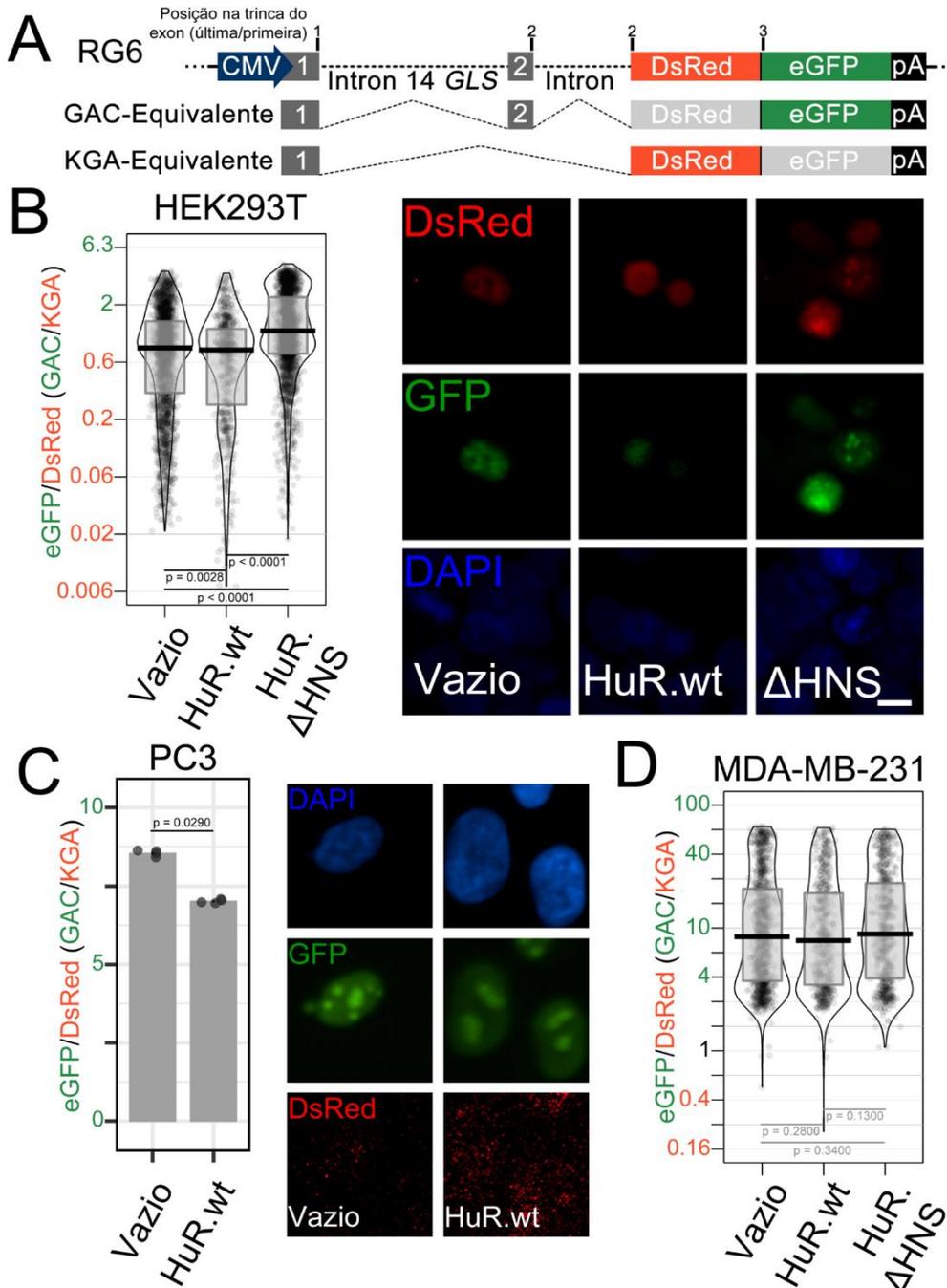
#### 4.4.4 Efeito regulatório de HuR através do intron 14 de *GLS* sobre *splicing* alternativo em um *minigene* repórter

Nossos estudos do potencial papel de HuR no *splicing* alternativo de *GLS* utilizando dados de transcriptômica e suas variantes disponíveis em bancos públicos nos levou a avaliar o papel de HuR na regulação do *splicing* deste gene pela interação com o intron 14. Orengo *et al*<sup>121</sup> desenvolveram um sistema com *minigene* repórter baseado na alteração do quadro de leitura após realização de *splicing* alternativo (Figura 40A). Basicamente, o sistema original consistia em um promotor do citomegalovírus (CMV) dirigindo a expressão de gene formado por três exons e dois introns. O primeiro exon conta com um início de transcrição e também o início de tradução da proteína, sendo que seu sítio de doação de *splicing* “entrega” o quadro de leitura com o primeiro nucleotídeo da trinca preenchido. Ocorrem os eventos de *splicing* que podem gerar dois transcritos: com ou sem a presença do exon 2. A inserção do exon 2 implica em modificação da “entrega” de posição do quadro de leitura para o próximo exon do primeiro para o segundo nucleotídeo. No terceiro exon, mais longo, temos a presença da região codificante para duas proteínas fluorescentes, DsRed e eGFP, mas fora de quadro de leitura uma em relação a outra: o quadro de DsRed inicia com uma trinca incompleta começando pela segunda posição e o de eGFP pela terceira.

Assim, quando o exon 2 é inserido, o peptídeo final gera fluorescência verde (emissão a 509nm quando excitado a 489nm), de eGFP; por outro lado, quando o exon 2 está ausente é observada fluorescência laranja-avermelhada (emissão a 592nm quando excitada a 557nm) de DsRed. Tal ferramenta pode ser modificada pela alteração de seus elementos regulatórios e se aplicar a outros eventos de regulação de *splicing* alternativo<sup>293</sup>. Assim sendo, a região do intron 1 do repórter foi substituída pelo intron 14 de *GLS*, avaliando seu efeito sobre a escolha de inserção ou não do exon 2, analogamente a posição no contexto genômico original, quando de escolha similar referente ao exon 15, exclusivo de GAC. Assim sendo, maiores valores da razão da intensidade de fluorescência eGFP/DsRed simbolizam maior inclusão do exon 2 (e conseqüentemente uma resposta pró-GAC) enquanto que menores valores implicam em maior pulo (*skip*) do mesmo (e uma resposta pró-KGA) (Figura 40A).

Verificamos que, em HEK293T, a expressão de HuR.WT leva a uma ligeira redução da mediana da razão eGFP/DsRed (equivalente a expressar a isoforma KGA, Figura 40B e imagem representativa à direita); como esperado, a expressão de HuR. $\Delta$ HNS (que desloca a proteína para maior presença citoplasmática), leva a um aumento da mediana, indicando maior escolha de GAC. A avaliação em PC-3 (feita somente para HuR.WT, Figura

40D, e imagem representativa à direita) evidenciou o mesmo resultado; em MDA-MB-231 vimos a mesma tendência, mas não de maneira significativa, tanto para a expressão de HuR.WT quanto de HuR. $\Delta$ HNS (Figura 40F).



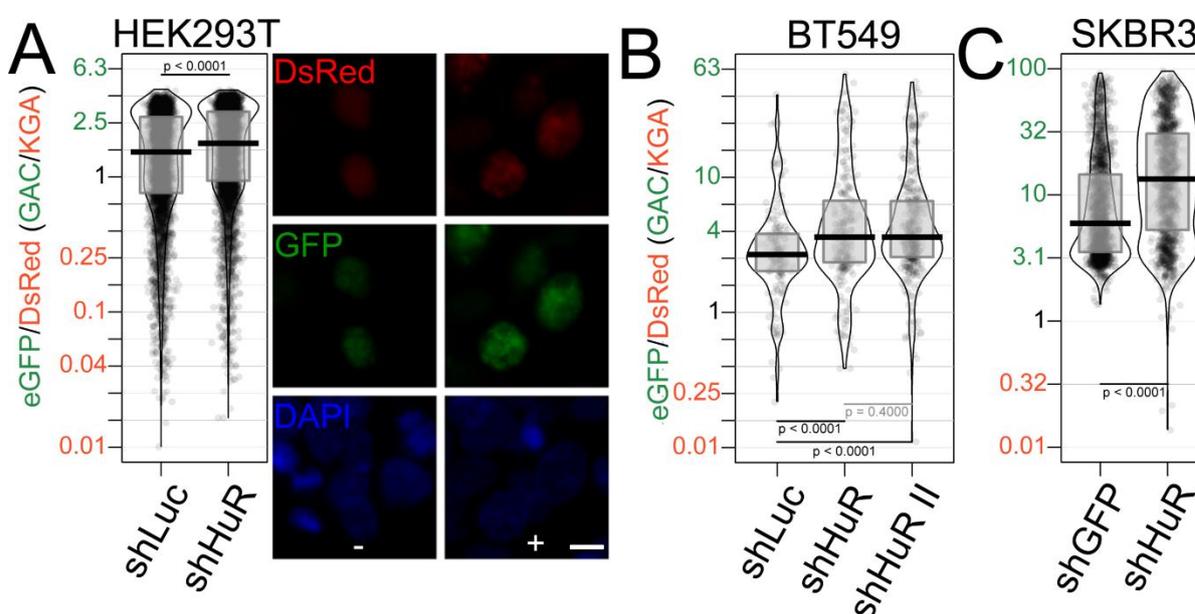
**Figura 40 – Ensaio mensurando influência do intron 14 sobre o *splicing* alternativo de minigene repórter sob influência de HuR ectópica**

(A) Esquema representativo do sistema repórter modificado<sup>121</sup>. O minigene é constituído de 3 exons, sendo o primeiro separado do segundo pela sequência do intron 14 de *GLS*. Na inserção do exon 2 (equivalente a inserção do exon 15 exclusivo à GAC), o quadro de leitura é deixado na segunda posição da trinca o que coloca, no terceiro exon, GFP em quadro de leitura, a qual, quando excitada a 489nm emite em 509nm (verde). Quando da omissão do exon 2, o terceiro exon terá a DsRed em quadro de leitura e expressão de proteína que quando excitada em 557nm emite a 592nm (vermelho). Assim, na razão

GFP/DsRed, valores superiores indicam maior expressão equivalente de GAC e valores inferiores maior expressão equivalente de KGA. (B) Gráfico RDI após expressão ectópica transitente de HuR (selvagem, WT, ou deletada para a sua região de controle de translado núcleo-citoplasmático,  $\Delta$ HNS) em células HEK293T e imagens representativas. (D) Gráfico de barras após expressão ectópica de HuR em células PC-3 e imagens representativas. (F) Gráfico RDI após expressão ectópica de HuR (selvagem, WT, ou deletada para a sua região de translado núcleo-citoplasmático,  $\Delta$ HNS) em células MDA-MB-231. Para gráficos de barras, as mesmas representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico. Para gráficos RDI barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os

valores das células quantificadas estão individualmente representados sobre os dados como círculos. Os eixos foram produzidos em escala logarítmica de base 10 e os valores convertidos para escala linear. *P-values* originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias. Barras de escala representam 10  $\mu\text{m}$ .

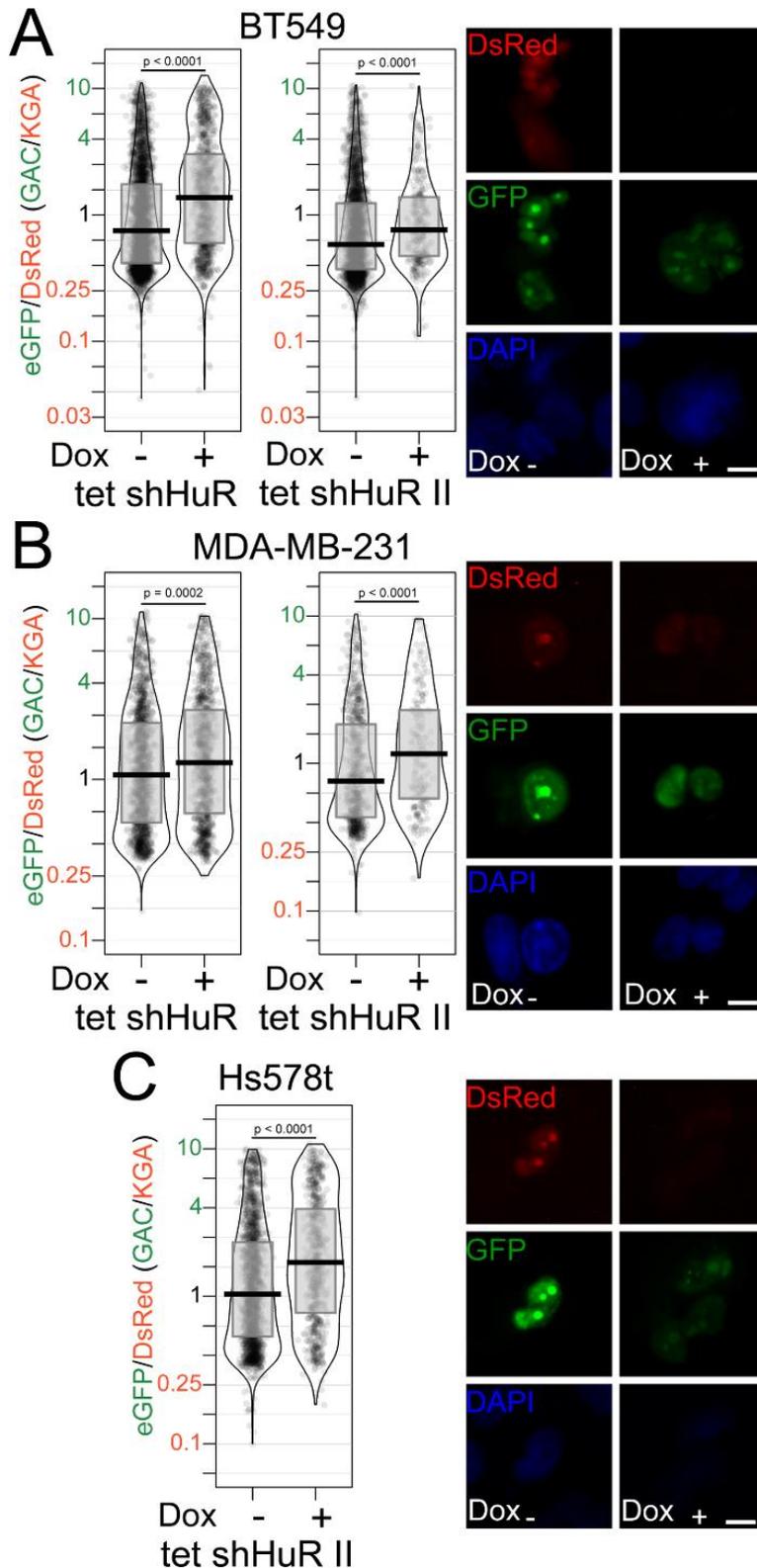
Avaliamos também o silenciamento transitente do gene de HuR nas células HEK293T. Verificamos que, como esperado, o silenciamento de *ELAVL1* levou a um aumento da razão eGFP/DsRed, indicando a escolha de GAC por *splicing* alternativo (Figura 41A). O mesmo foi observado em linhagens expressando estavelmente o shRNA, SKBR3 e BT549 (neste último as sequências shHuR e shHuR II foram testadas) (Figura 41B e C, respectivamente).



**Figura 41 – Ensaio mensurando influência do intron 14 sobre o *splicing* alternativo de minigene repórter sob influência do silenciamento de *ELAVL1***

Gráfico RDI da mensuração da razão GFP/DsRed para o silenciamento transitente de *ELAVL1* em HEK293T (A), e estável em BT549 (B) e SKBR3 (C). Imagem representativa das células HEK293T à direita. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os valores das células quantificadas estão individualmente representados sobre os dados como círculos. Os eixos foram produzidos em escala logarítmica de base 10 e os valores convertidos para escala linear. *P-values* originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias. Barras de escala representam 10  $\mu\text{m}$ .

Avaliamos o papel de HuR no *splicing* do intron 14 utilizando o sistema induzível de expressão de shRNA almejando o gene de HuR. Confirmamos neste sistema que o *knock down* de *ELAVL1* (empregando ambas as sequências) levou a aumento da razão eGFP/DsRed (GAC/KGA) tanto em BT549 quanto em MDA-MB-231 e Hs578t (Figura 42A, B e C, respectivamente, imagens representativas a direita).



**Figura 42 – Ensaio mensurando influência do intron 14 sobre o *splicing* alternativo de minigene repórter sob influência do silenciamento de *ELAVL1* em sistema induzível**

Gráfico RDI da mensuração da razão GFP/DsRed para o silenciamento induzível de *ELAVL1* em célula BT549 (A), MDA-MB-231 (B) e Hs578T (C) transduzidas com o sistema tet-pLKO-shHuR (esquerda) ou tet-pLKO-shHuRII (direita). Imagens representativas das células à direita. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os valores das células quantificadas estão individualmente representados sobre os dados como círculos. Os eixos foram produzidos em escala logarítmica de base 10 e os valores convertidos para escala linear. *P-values* originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias. Barras de escala representam 10  $\mu\text{m}$ .

É curioso observar que, em BT549 (Figura 42A), claramente não observamos uma distribuição normal dos valores, mesmo após conversão logarítmica dos mesmos. Observa-se, por exemplo para o shHuR (Figura 42A), pelo menos três pontos de aumento da distribuição

(abaixo de 1, entre 1 e 4 e outro entre 4 e 10); assim, por mais que vejamos um deslocamento da média e mediana para valores superiores (pró-GAC), não é possível a aplicação de testes paramétricos. A justificativa da utilização dos mesmos foi o do tamanho amostral de milhares de células, o que enquadra o problema em um caso do teorema do limite central<sup>294</sup>. À parte das opções estatísticas, a não-parametricidade das distribuições refletem a heterogeneidade celular observada quanto ao *splicing* alternativo mediado pelo intron 14 indicando o papel de importância de HuR, mas não excluindo outros elementos regulatórios.

#### 4.4.5 Expressão de KGA endógena com fusão de mKO2 através da técnica de CRISPR/Cas9 e utilização do modelo para confirmação dos efeitos de HuR sobre a isoforma

Utilizamos a ferramenta de modificação do genoma CRISPR/Cas9, conforme esquema representado na Figura 43A, para inserir a sequência codificante da proteína fluorescente mKO2 no C-terminal da isoforma KGA. Uma região de 100 nucleotídeos nas imediações do códon de parada existente no exon 19 de *GLS*, referente à KGA, foi avaliada pela ferramenta CRISPRDesign<sup>295</sup>. Encontramos 7 possíveis RNA-guias (sgRNA), sendo que um deles, coincidentemente, indicou o corte exatamente antes do códon de parada de KGA. Isto provavelmente aconteceu porque o códon *Stop* (de uma sequência TAA) é imediatamente justaposto por uma sequência TGG, passível de ser um “motivo adjacente ao protoespaçador” (*Protospacer Adjacent Motif*, PAM, que, para a Cas9 de *Streptococcus pyogenes*, se define como NGG<sup>296</sup>). Quando avaliamos os *offtargets* em sequências codificantes, a predição mais provável envolvia 3 mutações no sgRNA e uma que descaracterizava o motivo PAM. Para sequências não-codificantes, a predição mais provável envolvia, também, a descaracterização do motivo PAM com duas mutações.

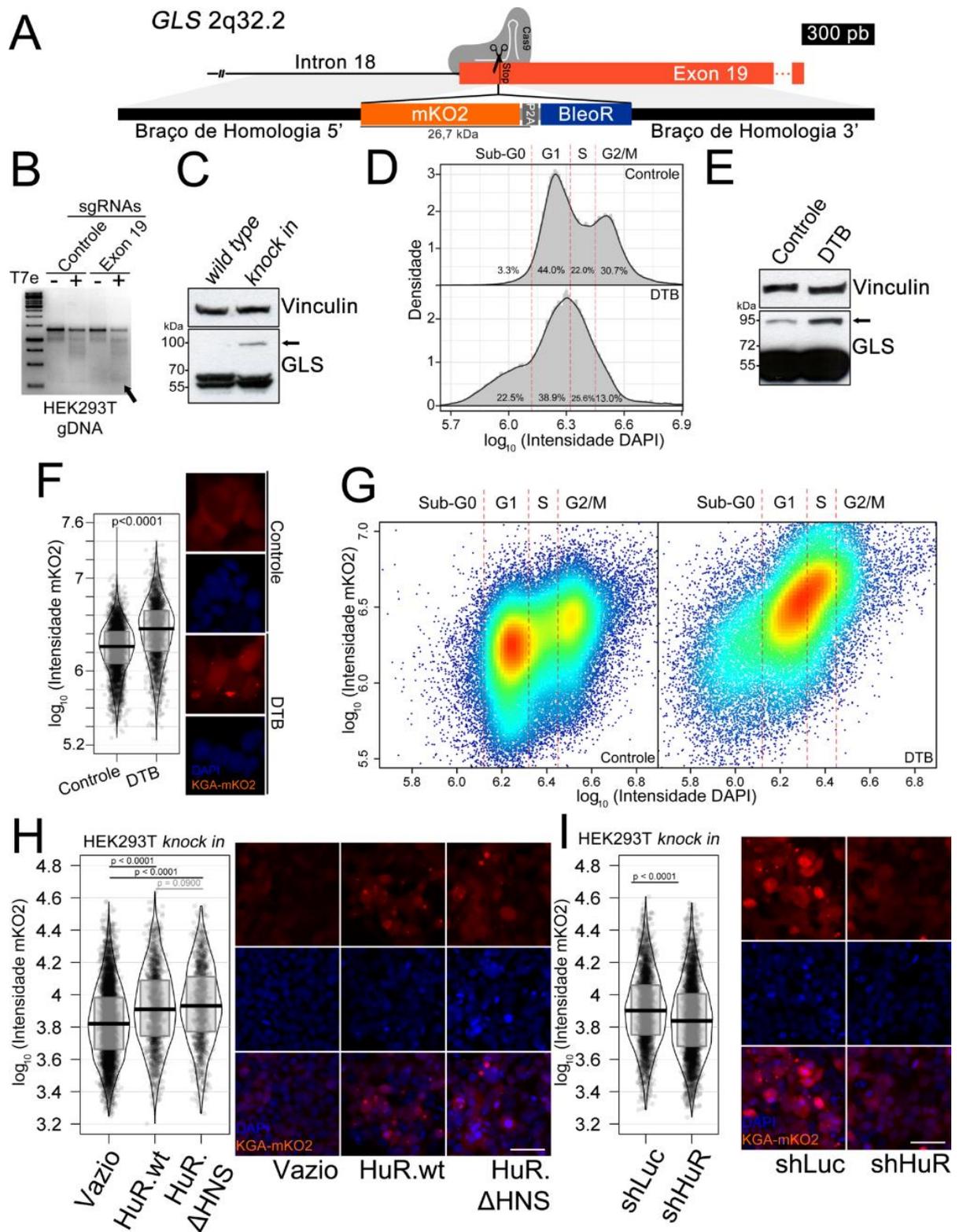
Dado que haveria uma quebra de dupla fita na porção final de KGA, para evitar o reparo por junção de pontas não-homólogas (NHEJ, *Non-Homologous End Joining*) e que a sequência-base utilizada para o reparo dirigido por homologia (HDR, *Homology Directed Repair*) não fosse o outro cromossomo, foi desenhado um plasmídeo doador de sequência homóloga. O cassete criado consiste em mil bases de homologia genômica a montante do corte (compreendendo uma porção do intron 18 e o início do exon 19), sequência codificante para a proteína fluorescente monomérica kusabira-orange2 (mKO2), sequência do peptídeo de auto-clivagem 2A do Teschovírus-1 suíno (P2A<sup>123,124</sup>), sequência codificante do gene BleoR de *Streptoalloteichus hindustanus* (conferindo resistência ao agente de seleção Zeocina) e, por fim, mil bases de homologia genômica a justante do corte (compreendendo o códon de parada e

uma porção da 3'UTR de KGA). Assim, células com o genoma positivamente alterado irão expressar fluorescência laranja fusionada à enzima KGA e proteína que confere resistência à quebra de DNA promovida pela Zeocina<sup>297, 298</sup>. A proteína que confere resistência, vale ressaltar, é liberada da KGA pela auto-clivagem do P2A, evitando, assim, problemas estruturais em KGA ou o carregamento da resistência para a mitocôndria.

Inicialmente, o plasmídeo expressão Cas9 pX330<sup>295</sup>, modificado para conter o sgRNA e resistência a puromicina (removendo a necessidade de co-entrega com plasmídeo carregando agente de seleção), foi transfectado em células HEK293T. As células tiveram seu DNA genômico extraído e a região do códon de parada de KGA amplificada por PCR para ensaio com T7 endonuclease I (Figura 43B). Basicamente, quando da mistura de fragmentos selvagens e com mutações é anelado, sequências com pareamentos inadequados serão cortadas pela T7 endonuclease I, como indicado pela seta. Com o indício de atividade da sequência sgRNA escolhida, o mesmo vetor foi novamente transfectado, agora com o braço de homologia construído linearizado, e observada uma banda de maior peso molecular, compatível com a adição da cauda construída (Figura 43C).

Para avaliar se modificação gerada estava representando a abundância de KGA intracelularmente, decidiu-se explorar a relação entre o acúmulo de KGA com o ciclo celular, como previamente observada<sup>65,66</sup>. A região designada como KEN-Box (sequência de lisina - ácido glutâmico – asparagina, situada 11 aminoácidos antes do final de KGA) é marcada pelo complexo promotor da anáfase APC/C-Cdh1, direcionamento a proteína para degradação com o avanço para a fase G2/M do ciclo celular. Colombo<sup>65</sup> *et al* mostraram um acúmulo dos níveis de KGA quando as células foram sincronizadas em G1/S. Desta maneira, realizando a sincronização das células nesta fase, esperaríamos que nosso sistema repórter mimetizasse tal evento, quando acompanhado pelo nível proteico total ou por fluorescência.

A sincronização das células foi realizada pelo método do bloqueio duplo de timidina<sup>299</sup>, que atua inibindo a síntese de DNA (com o excesso de timidina bloqueando a via de síntese de desoxirribonucleotídeos) e realizando uma sincronia preliminar em diversos estágios da fase de síntese, seguido por uma liberação (permitindo a conclusão da síntese pelas células em diversas etapas) e um novo bloqueio, sincronizando agora as células em fase G1 ou S-inicial. Com o intuito de avaliar o a sincronização do ciclo, a intensidade de fluorescência de DAPI foi utilizada para mensurar a quantidade de DNA<sup>225</sup>. O bloqueio (*Double thymidine block*, DTB) mostrou uma redução de 30,7% para 13% das células em fase G2/M (Figura 43D).



**Figura 43 – Modificação do genoma de HEK293T utilizando CRISPR/Cas9 para expressão de KGA endógena com cauda fluorescente e avaliação dos efeitos do silenciamento e superexpressão de *ELAVL1***

(A) Esquema da estratégia de modificação empregada. Parte superior indica a região do genoma editada, exon 19 de *GLS*. Na porção inferior, o cassete introduzido entre os braços de homologia no plasmídeo doador: proteína fluorescente monomérica Kusabira-Orange2, peptídeo 2A de auto clivagem e gene de resistência ao agente de seleção Zeocina. (B) Ensaio com T7 endonuclease em DNA genômico de células controle ou transfetadas com o plasmídeo contendo o RNA guia e Cas9. Seta indica região do fragmento gerado pelo corte. (C) *Western blot* de células selvagens ou do *pool* celular selecionado com a introdução do cassete fluorescente. Seta indica banda de maior peso molecular observada em anticorpo que reconhece as duas isoformas de glutaminase. (D) Avaliação da eficácia na sincronização do ciclo celular com utilização da mensuração da

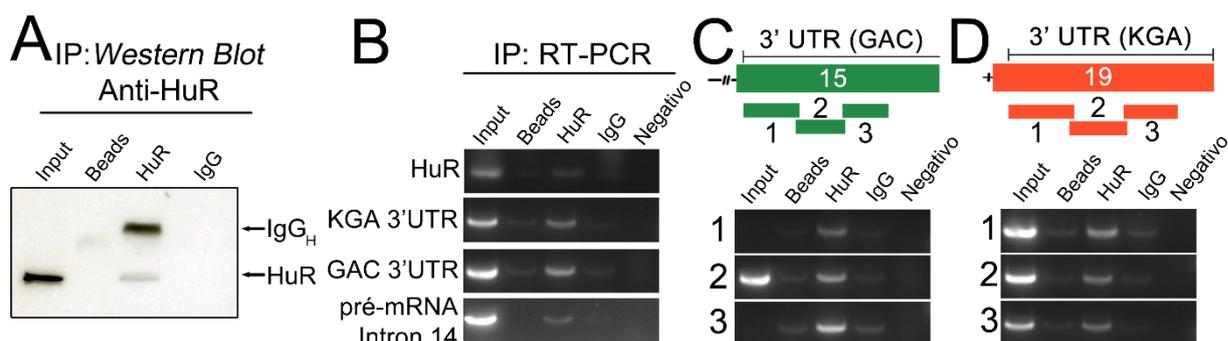
fluorescência de DAPI<sup>225</sup>. Linhas verticais indicam as definições de fases do ciclo celular e porcentagens a fração das células na respectiva etapa. (E) *Western blot* após bloqueio duplo de timidina (*Double thymidine block*, DTB) para sincronização de ciclo celular em G1/S. Seta indica a banda gerada no *knock in*. (F) Gráfico RDI comparando a fluorescência da célula após o DTB. Imagens representativas do aumento de intensidade de fluorescência à direita. (G) Gráfico de dispersão correlacionando a quantidade de DNA e a intensidade de fluorescência de KGA-mKO2 da célula. Cores representam a densidade kernel bidimensional suavizada<sup>300</sup>, onde azul claro são menores acúmulos de pontos e vermelho escuro as maiores concentrações. (H) Gráfico RDI da fluorescência quando na expressão ectópica de *ELAVLI* (selvagem, WT, ou deletada para a sua região de translado núcleo-citoplasma,  $\Delta$ HNS); a direita, imagens representativas. (I) Gráfico RDI da fluorescência quando no silenciamento transiente de *ELAVLI*; à direita, imagens representativas. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os valores das células quantificadas estão individualmente representados sobre os dados como círculos. *P-values* originados de teste-t de comparação de médias.

Tal qual esperado, o processo DTB gerou um aumento da banda de altura elevada gerada com o *knock in* frente às células não bloqueadas (Figura 43E), acompanhado de um aumento na intensidade de fluorescência de mKO2 (Figura 43F, imagem representativa ao lado). Observando em gráfico de dispersão de correlação entre os níveis de KGA-mKO2 e DNA, verificamos, no controle, dois centros de massa na distribuição inicial, em G1 e G2/M; o procedimento de DTB levou a formação de um único pico em G1/S, havendo uma concordância entre o aumento de KGA e acúmulo de células nestas fases. Em escala linear, a diferença na intensidade de KGA-mKO2 nas células-controle presentes em G0-1 para as células na divisa entre G1/S (obtido com DTB) é de cerca de 80% (Figura 43G). O procedimento levou a um grande aumento em sub-G0, indicando toxicidade.

Validado o sistema, avaliamos o impacto da expressão transiente de HuR.WT e HuR. $\Delta$ HNS na intensidade de fluorescência de mKO2. A expressão de HuR.WT e de HuR. $\Delta$ HNS causou aumento de KGA (Figura 43H, imagem representativa à direita) enquanto que o *knock down* de *ELAVLI* levou a diminuição deste sinal (Figura 43I), coerente com a presença de HuR relacionando-se positivamente com o nível proteico de KGA.

#### 4.4.6 Imunoprecipitação de RNA seguida de detecção por RT-PCR

Para confirmar a interação de HuR com a região 3'UTR do mRNA de GAC e KGA, e no intron 14 do gene *GLS* foi realizada uma imunoprecipitação de HuR seguida da extração dos RNAs ligados.



**Figura 44 – Ligação de HuR na região do intron 14 e 3'UTR das isoformas de *GLS* (KGA e GAC)**

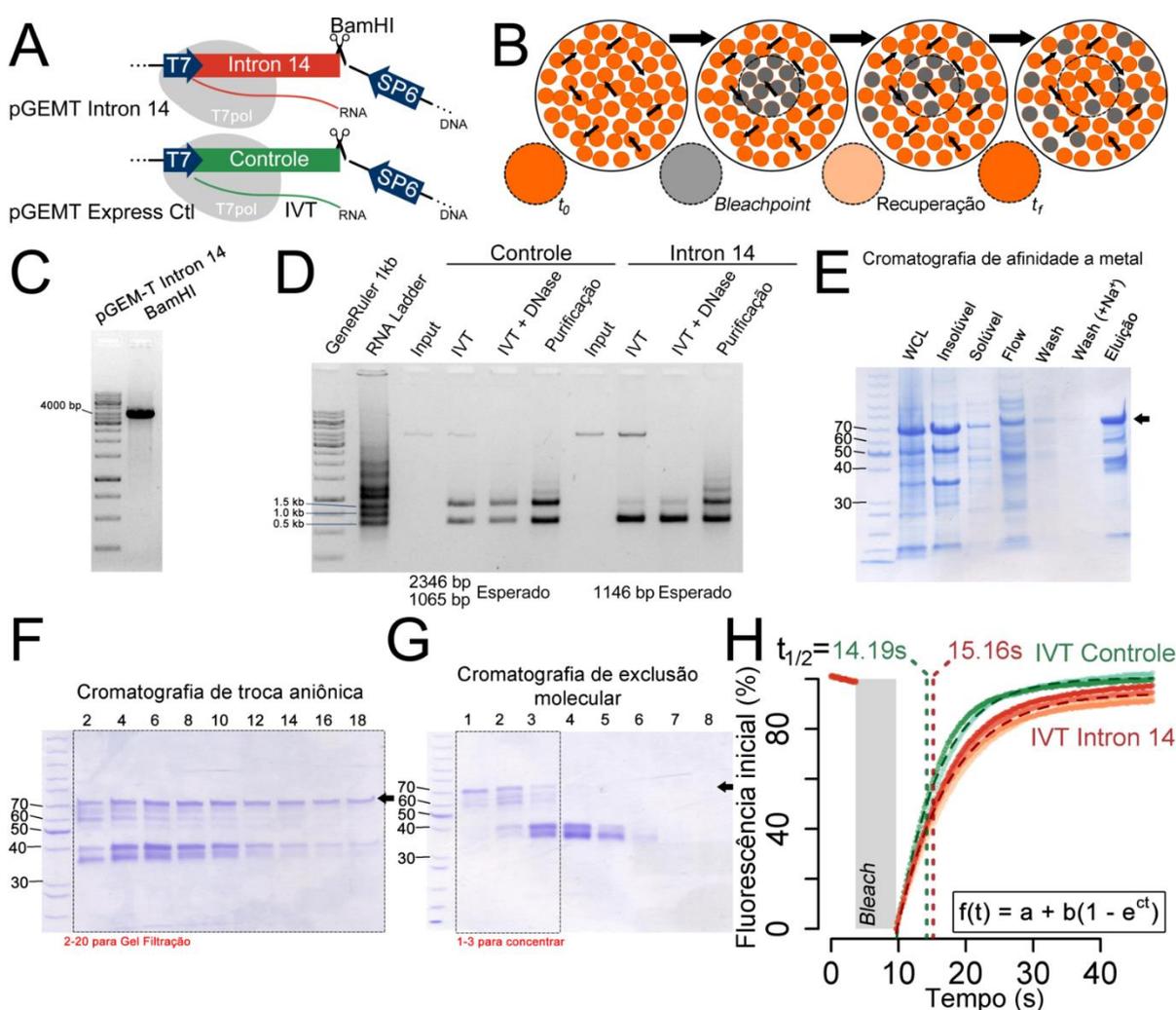
(A) *Western blot* mostrando a imunoprecipitação de HuR na presença do anticorpo de anti-HuR e ausência de ligação inespecífica detectável. (B) Transcrição Reversa seguida de amplificação por PCR do RNA extraído dos resultados da imunoprecipitação, com amplificação superior no material inicial (*Input*) e na presença de anti-HuR (HuR) utilizando oligonucleotídeos para transcrito de *ELAVL1*, 3'UTR de KGA, 3'UTR de GAC e pré-mRNA de *GLS* - na região do intron 14. (C) Utilização de outros oligonucleotídeos para regiões definidas do 3'UTR de GAC (C) ou KGA (D). Esquema superior indica a posição de cada fragmento numerado e seu respectivo resultado de amplificação. Experimento realizado utilizando células de câncer de próstata PC-3.

A proteína HuR endógena foi imunoprecipitada de PC-3 adequadamente (Figura 44A) e o RNA ligado extraído. Após transcrição reversa seguida de PCR, confirmamos a ligação de HuR a sua própria 3'UTR (como já descrito na literatura<sup>101</sup>), assim como à região 3'UTR de GAC e KGA (Figura 44B). Ao pré-mRNA de *GLS*, especificamente na região do intron 14, também foi observada interação (Figura 44B). Oligonucleotídeos para regiões distintas da 3'UTR de GAC (Figura 44C) ou KGA (Figura 44D) também foram empregados, apresentando amplificação em todas as imunoprecipitações de HuR. Conforme já apresentado, já foram mostrados sítios de ligação teóricos e validados por PAR-CLIP-Seq na 3'UTR de GAC<sup>76</sup> (Figura 7) e, como proposto no Item 4.2, no intron 14 de glutaminase (Figura 29).

#### 4.4.7 Avaliação da interação *in vitro* de HuR e o RNA do intron 14 por *Fluorescence recovery after photobleaching* (FRAP)

Para a confirmação *in vitro* da interação direta entre a proteína HuR com a sequência do intron 14 de glutaminase realizamos a expressão heteróloga em sistema procaríoto e purificação da proteína HuR, bem como a transcrição *in vitro* do RNA correspondente ao intron 14. O sistema de transcrição *in vitro* utilizado foi do tipo *run-off* (Figura 45A): o DNA molde, originário de um plasmídeo previamente clivado exatamente no final da região desejada, tem a transcrição iniciada pela T7 RNA Polimerase que irá “cair” da fita ao chegar a região cortada e liberar o RNA. Aproveitamos a fusão fluorescente utilizada para acompanhar a expressão da proteína em um experimento de FRAP (*Fluorescence recovery after photobleaching*, Figura 45B). Basicamente, o mesmo consiste na avaliação da capacidade de

difusão de uma molécula em uma solução: após quantificar a fluorescência inicial em um volume de líquido, todos os fluoróforos da região são “queimados”, em nosso caso, com um laser bombeado a diodo de estado sólido (DPSS) de 561 nm até a extinção da fluorescência das partículas. Durante o tempo de recuperação (com a fluorescência acompanhada), as moléculas “queimadas” – por movimento de difusão – irão sair do volume focal, sendo substituídas por moléculas com fluorescência. A velocidade de difusão, diminuída para moléculas maiores (como complexos RNA:Proteína), pode ser averiguada pelo tempo de recomposição de metade da fluorescência ( $t_{1/2}$ )<sup>301,302</sup>.



**Figura 45 – Avaliação da interação de HuR e o intron 14 de GLS *in vitro***

(A) Representação esquemática do preparo das construções para a transcrição *in vitro*, evidenciando a clivagem com BamHI ao final dos fragmentos de interesse. (B) Representação esquemática do experimento de FRAP: A situação inicial ( $t_0$ ) indica a livre mobilidade do fluoróforo avaliado, que é queimado com laser DPSS 561nm, gerando uma zona sem fluorescência (*Bleachpoint*); após um tempo, há um processo de redistribuição das moléculas (Recuperação) o qual culmina no novo platô de fluorescência final ( $t_f$ ). (C) Clivagem do molde de Intron 14 para transcrição *run-off*. (D) Transcrição *in vitro* (IVT) utilizando T7 RNA Polimerase para o controle (pGEM Express) e para Intron 14 indicando o DNA inicial (*Input*), o resultado total da transcrição (IVT), o resultado da degradação do molde através do tratamento com DNase (IVT + DNase) e o proveniente da purificação final (Purificação). Seta indica a banda do tamanho esperado para a construção 6xHis-mKO2-HuR.WT. (E) Avaliação

por SDS-PAGE da cromatografia de afinidade realizada, indicando a presença de banda de cerca de 70 kDa na fração da eluição (seta preta). *WCL*: *whole cell lysate*; Insolúvel: ressuspensão do *pellet* após ultracentrifugação em tampão de corrida; Solúvel: quantidade de proteína remanescente diluída no sobrenadante da ultracentrifugação; *Flow*: remanescente não-ligado na coluna de purificação por afinidade à cobalto; *Wash*: lavagens da coluna; *Wash (+Na<sup>+</sup>)*: lavagem com adição de 500 mmol/L de NaCl; Eluição: Proteína obtida após a adição de imidazol. (F) Avaliação por SDS-PAGE da cromatografia de troca iônica realizada com o eluído da afinidade em suas diversas frações. (G) Avaliação por SDS-PAGE da cromatografia de gel filtração com as frações concentradas do passo anterior. Em ambos os casos (F e G), números representam eluições de 1 mL e caixas tracejadas as frações coletadas e utilizadas no próximo passo cromatográfico. (H) Avaliação por FRAP do percentual de fluorescência após o *bleach* em função do tempo, o ajuste da curva e a meia-vida de recuperação ( $t_{1/2}$ , linhas verticais) do RNA Controle (curvas verdes) e do RNA do Intron 14 (curvas vermelhas). Região sem valores marcada em cinza representa o ponto de bleach no qual não é avaliada a fluorescência e a função dentro da caixa o formato de exponencial escolhido para o ajuste com o pacote *nls* do R.

O RNA do intron 14 foi produzido através de transcrição *run-off* utilizando como molde o vetor pGEM-T Intron 14 clivado com BamHI (Figura 45C). Tanto a produção do controle de ligação inespecífica quanto do Intron 14 apresentaram sucesso (Figura 45D), tendo o processo de degradação do DNA molde e purificação ocorrendo adequadamente. Acredita-se que as bandas duplas observadas (não reproduzidas em outras preparações), sejam ocasionadas por estrutura secundária não-desnaturada do RNA. A construção 6xHis-mKO2-HuR.WT foi expressa e purificada por afinidade a cobalto (Figura 45E) devido a cauda de histidinas, realizada uma troca iônica para remoção de ácidos nucléicos contaminantes (Figura 45F) e filtrada em gel para um polimento final da purificação (Figura 45G), e remoção de fragmentos de menor peso molecular em relação ao esperado (~70kDa) – resultantes de contaminação remanescente ou degradação da proteína.

De posse da proteína purificada e do RNA relativo ao controle e ao intron 14, o experimento de FRAP foi realizado (Figura 45H). Observando-se as curvas de recuperação, ocorreu um deslocamento do tempo de meia vida de recuperação do controle (curvas verdes) para o intron 14 (linhas vermelhas) em 970 milisegundos ao utilizarmos uma proporção equimolar de intron e proteína. Tal resultado indica que o intron 14 interage diretamente com a proteína HuR, dada a reduzida mobilidade da mesma em solução quando na presença do RNA transcrito pelo intron.

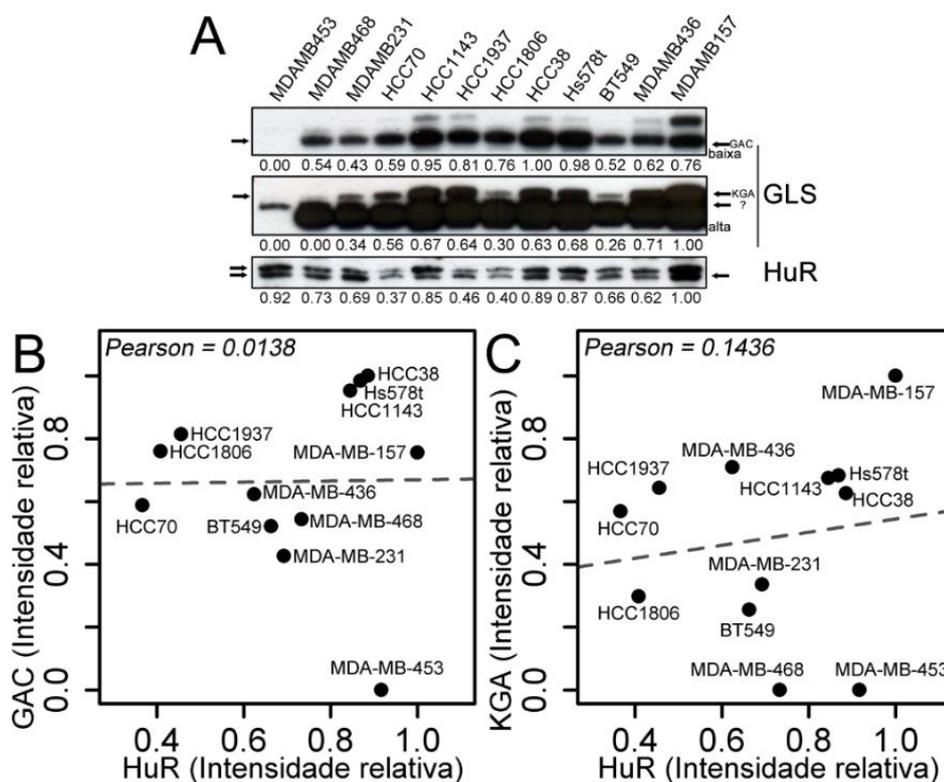
#### 4.5 FENÓTIPOS PROLIFERATIVOS, MIGRATÓRIOS E INVASIVOS DE CÉLULAS DE CÂNCER DE MAMA ALTERADOS PELOS NÍVEIS DE EXPRESSÃO DE *ELAVL1*

Tanto HuR quanto *GLS* já foram relacionados aos processos de crescimento e invasão celular. Em câncer de mama, por exemplo, o silenciamento direto de *ELAVL1* reduziu

a capacidade invasiva<sup>303</sup> e a sua inibição química a capacidade migratória<sup>304</sup>, em células MDA-MB-231. O silenciamento de *GLS*, por sua vez, já foi observado reduzindo a capacidade proliferativa celular de SKBR3 e MDA-MB-231<sup>305</sup> e sensibilizando a última a tratamentos quimioterápicos, como o taxol<sup>306</sup>. Assim, resolvemos avaliar o impacto HuR sobre a escolha das isoformas nestes processos tumorais.

#### 4.5.1 Caracterização da expressão endógena de HuR em um painel de células de câncer de mama

Utilizando *western blot*, detectamos a presença da proteína HuR em um painel de 12 linhagens celulares de câncer de mama (Figura 46A). Para as isoformas de *GLS*, GAC foi praticamente ubíqua em todas as linhagens (salvo em MDA-MB-453) e KGA não foi detectada em 2 das 12 linhagens (MDA-MB-453 e MDA-MB-468). Quando comparamos a correlação entre a intensidade da marcação de HuR com a intensidade da banda referente à isoforma GAC (Figura 46B) e KGA (Figura 46C), observamos uma correlação positiva apenas com a segunda isoforma, coerente com os resultados *in vitro* até então apresentados.



**Figura 46 – Níveis totais das proteínas HuR, KGA e GAC em células de câncer de mama**

(A) *Western blot* de linhagens celulares de câncer mama para anti-GLS (KGA e GAC) e anti-HuR. Setas na esquerda indicam a(s) banda(s) quantificadas segundo a numeração abaixo do quadro. Baixa e alta referem-se ao tempo de exposição dos filmes. Gráficos de correlação entre as intensidades relativas a banda de maior intensidade (denotada por 1.00) de GAC (B) ou KGA (C) e HuR. Linha de correlação baseada em regressão linear, correlação de Pearson indicada no topo dos gráficos.

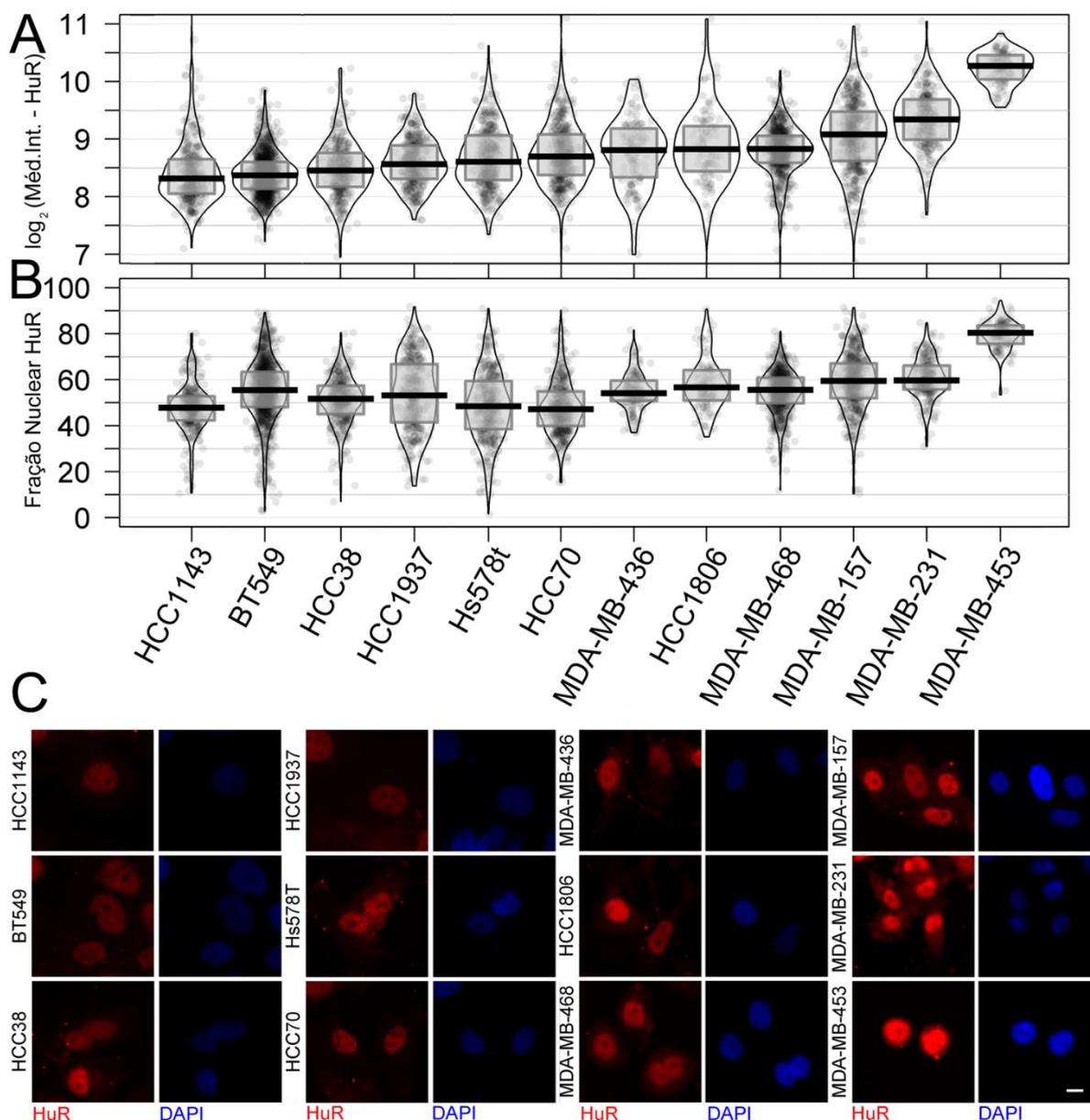
A habilidade de invasão é necessária à uma célula tumoral para a concretização do processo metastático<sup>307</sup>. Para invadir, entretanto, a célula tumoral precisa desenvolver habilidades de movimento direcional, o que decorre de processos de remodelamento do citoesqueleto<sup>308,309</sup> e de degradação de matriz extracelular, todos parte do processo de transição epitélio-mesênquima<sup>308</sup>. Para mensurar tal evento, diversos marcadores são propostos, dentre eles a formação de *actin stress fibers* (fibras de estresse de actina)<sup>310</sup>, que geram a força contrátil necessária para a movimentação celular<sup>311</sup>. Uma maneira prática de quantificar este fenótipo se dá pela avaliação do citoesqueleto celular por microscopia de alto conteúdo com a mensuração do parâmetro *SER Valley* (*Spots Edges Ridges – Valley*) em marcações de faloidina<sup>312</sup>, o qual é superior na presença destas formações. Já foi observada uma relação entre os níveis totais de HuR e a formação de *actin stress fibers* em células MDA-MB-231<sup>303</sup> e seu efeito estabilizador sobre o mRNA de actina<sup>313</sup>. Desta maneira, decidimos avaliar nestas doze linhagens celulares de câncer de mama, por imunofluorescência, os níveis proteicos totais de HuR e o parâmetro *SER Valley*.

As intensidades médias de fluorescência variaram cerca de 1,5 unidade de  $\log_2$ , com MDA-MB-453 apresentando (em escala linear) cerca de 3 vezes mais de marcação média do que de HCC1143 (Figura 47A). Quanto a localização nuclear de HuR, todas as células apresentam, ao menos, cerca de metade da proteína localizada em seu núcleo, com o extremo sendo de MDA-MB-453 alcançando 80% de concentração nuclear (20% de fração citoplasmática, Figura 47B), justificado pelo diminuto citoplasma da célula que, por mais que contenha elevada marcação de HuR, representa pouco do volume celular.

Relacionamos então a marcação média de HuR e *actin stress fibers*, observando uma correlação positiva (Figura 48A), mais evidenciada na remoção da célula de pequeno citoplasma MDA-MB-453 (Figura 48B). Apesar de haver esta correlação positiva, fica claro dos dados que em algumas linhagens a alta marcação de HuR não implica em maior marcação para *actin stress fibers*. Para outras, entretanto, esta relação fica claro, como na comparação (Figura 48C) de HCC1143 e BT549 (com menor marcação de HuR) e MDA-MB-468 e HCC1806 (com maior marcação de HuR), onde as duas primeiras tem nitidamente menor formação de *actin stress fibers* do que as duas últimas.

Avaliamos então a correlação entre estes dois parâmetros célula a célula para cada linhagem com o interesse de encontrar modelos em que estes parâmetros se correlacionassem positivamente. Das células avaliadas, oito apresentaram correlações acima de 0,10 entre *SER Valley* e a intensidade média de HuR (MDA-MB-453, MDA-MB-436, MDA-MB-468, HCC1806, HCC38, MDA-MB-231, HCC1143 e BT549), três apresentaram correlação entre -

0,10 e 0,10 (MDA-MB-157, HCC70 e HCC1937) e apenas uma apresentou correlação negativa (Hs578t) (Figura 49).

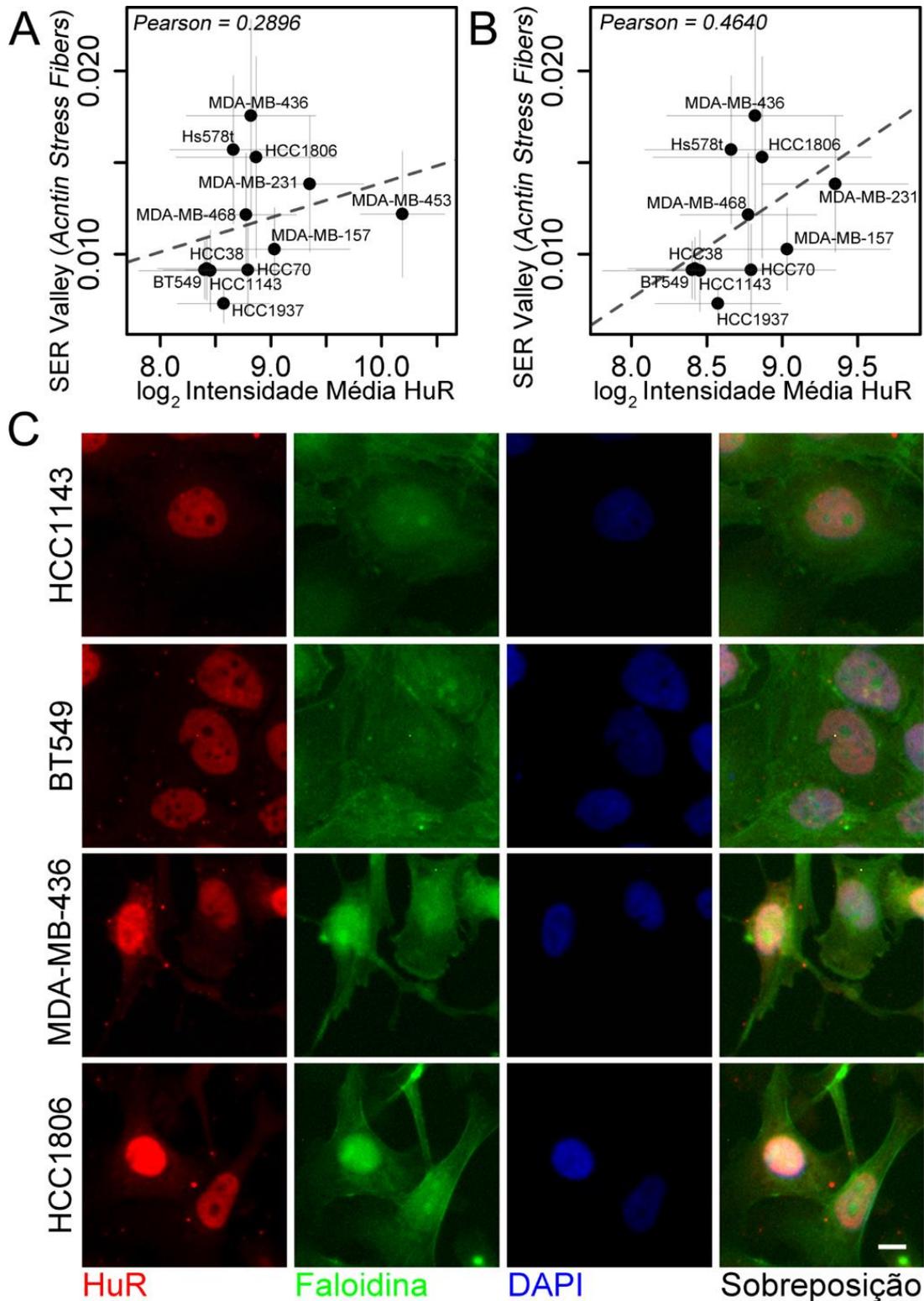


**Figura 47 – Nível proteico e localização celular de HuR em células de câncer de mama avaliadas por imunofluorescência**

Gráfico RDI com intensidade média (A) e da fração nuclear (B) de HuR nas células de câncer de mama avaliadas. (C) Imagens representativas das células avaliadas. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os valores das células quantificadas estão individualmente representados sobre os dados como círculos. Barra de escala representa 10  $\mu$ m.

Para verificar se os níveis proteicos de HuR e sua localização celular poderiam ser impactados pelo metabolismo de glutamina, avaliou-se os efeitos da remoção de glutamina do

meio de cultivo por um período de 24 horas sobre a intensidade e localização celular de HuR (Figura 50).

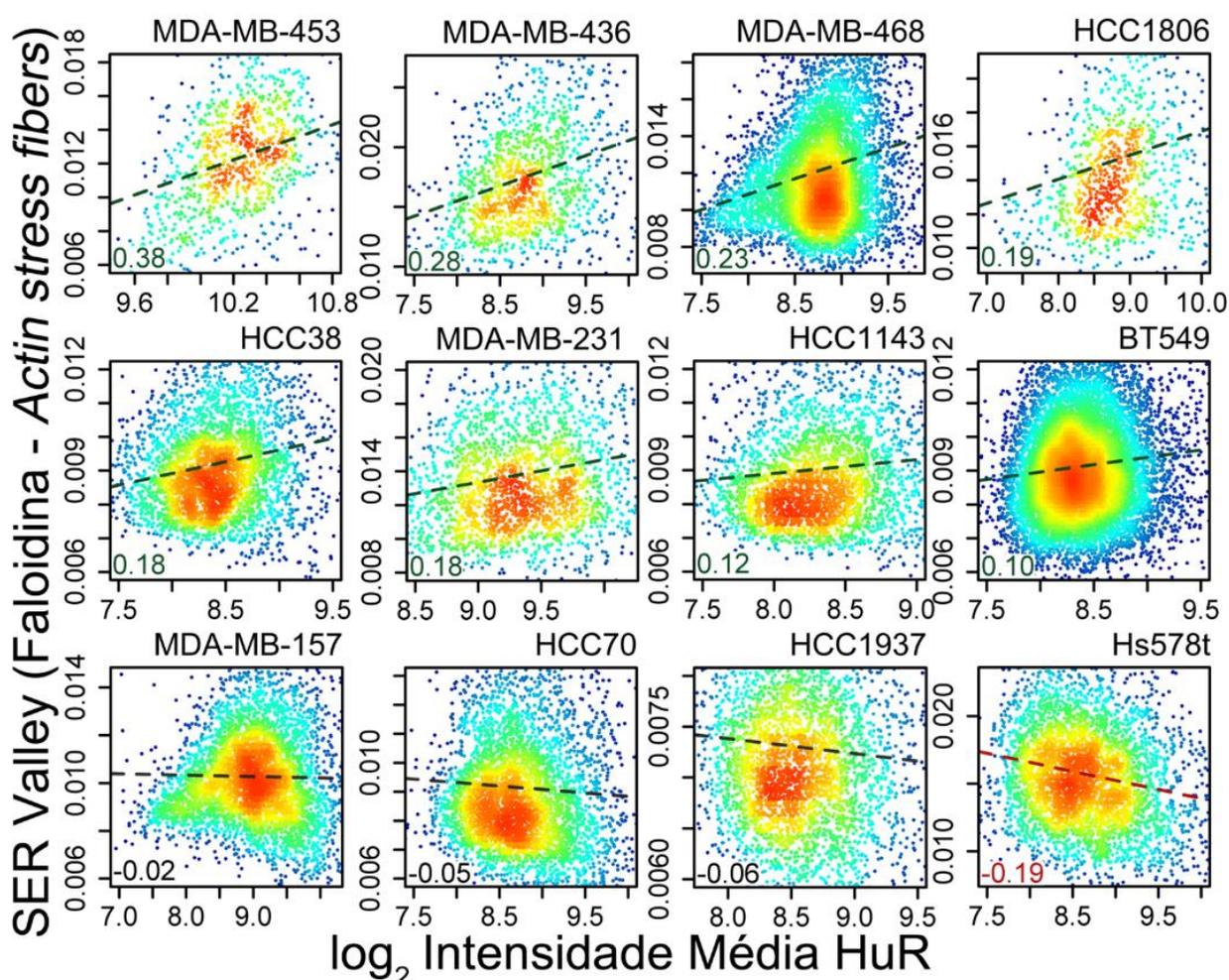


**Figura 48 – Correlação entre níveis médios de HuR e localização citoplasmática mensurados por imunofluorescência com parâmetro *SER Valley* (*actin stress fibers*)**

Correlação entre os níveis de intensidade média de marcação fluorescente para HuR e *SER Valley*, com todas as 12 linhagens (A) e na ausência de MDA-MB-453 (B). Barras representam desvio padrão da média, linha de correlação baseada em regressão linear, correlação de Pearson indicada no topo dos gráficos. (C) Imagens

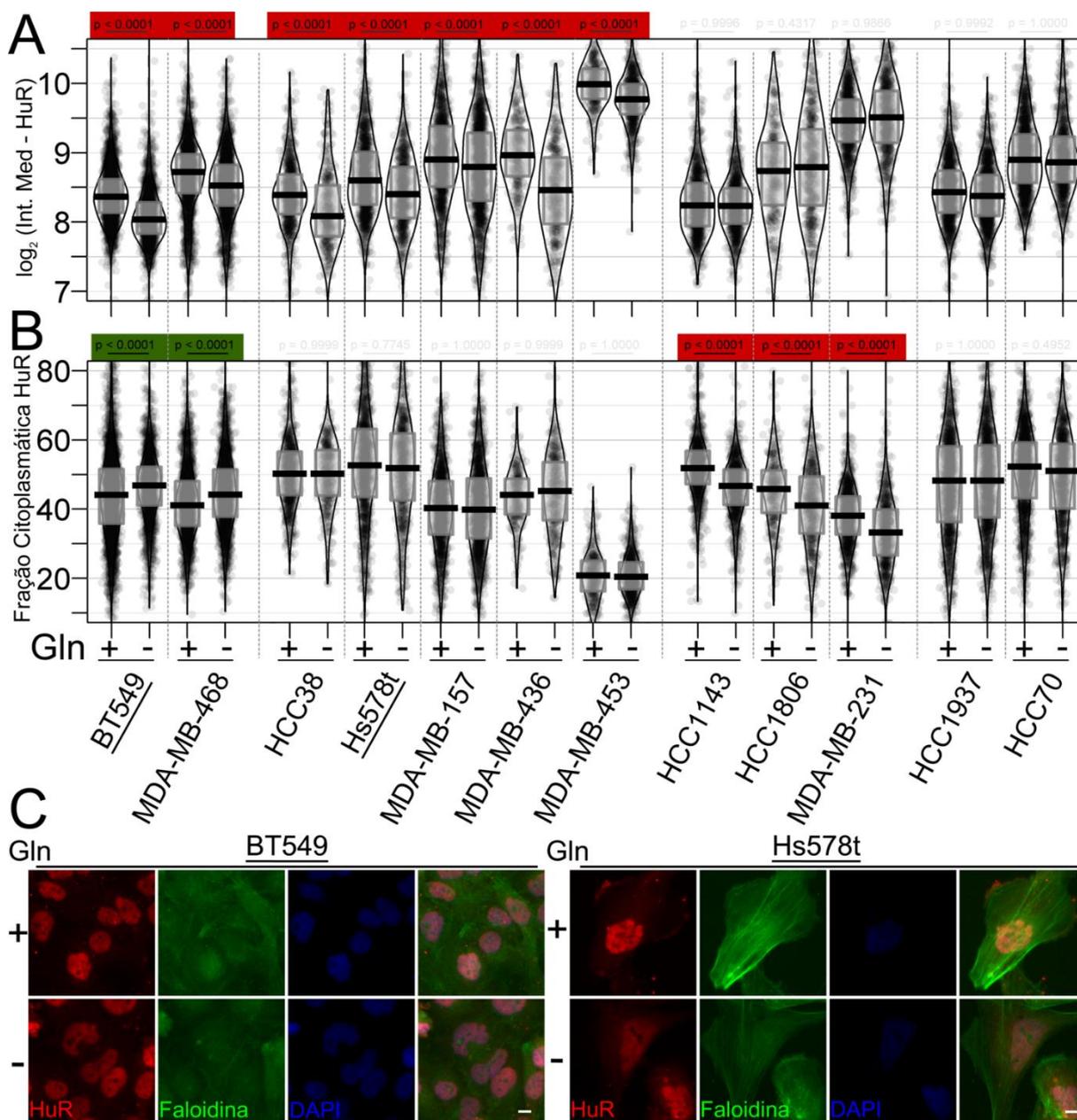
representativas de células com menor (HCC1143 e BT549) e células com maior (MDA-MB-468 e Hs578t) marcação de HuR e o impacto na formação de *actin stress fibers*. Barra de escala representa 10  $\mu$ m.

Das doze linhagens avaliadas, sete apresentaram redução dos níveis de HuR quando da supressão de glutamina (BT549, MDA-MB-468, HCC38, Hs578t, MDA-MB-157, MDA-MB-436 e MDA-MB-453) (Figura 50A). Quanto a redistribuição núcleo-citoplasma (Figura 50B), das cinco linhagens em que ocorreu alguma alteração, HCC1143, HCC1806 e MDA-MB-231 apresentaram uma menor localização citoplasmática e BT549 e MDA-MB-468 apresentaram um deslocamento para o citoplasma – sendo também as únicas a terem os dois parâmetros (abundância e localização) afetados pela supressão de glutamina.



**Figura 49 – Correlação entre a abundância de HuR e formação de *actin stress fibers* em células de câncer de mama em cultura**

Gráficos de dispersão correlacionando as intensidades médias da marcação de imunofluorescência de HuR com o parâmetro *SER Valley* de textura de imagem, que representa a formação de *actin stress fibers*, comumente relacionadas à invasividade e malignidade. Linhas tracejadas representam regressão linear dos pontos. Valores no canto inferior direito de cada gráfico o coeficiente de correlação de Pearson. Valores acima de 0.10 estão em verde e abaixo de -0.10 em vermelho. Cores dos pontos representam a densidade kernel bidimensional suavizada<sup>300</sup>, onde azul claro são menores acúmulos de pontos e vermelho escuro as maiores concentrações.



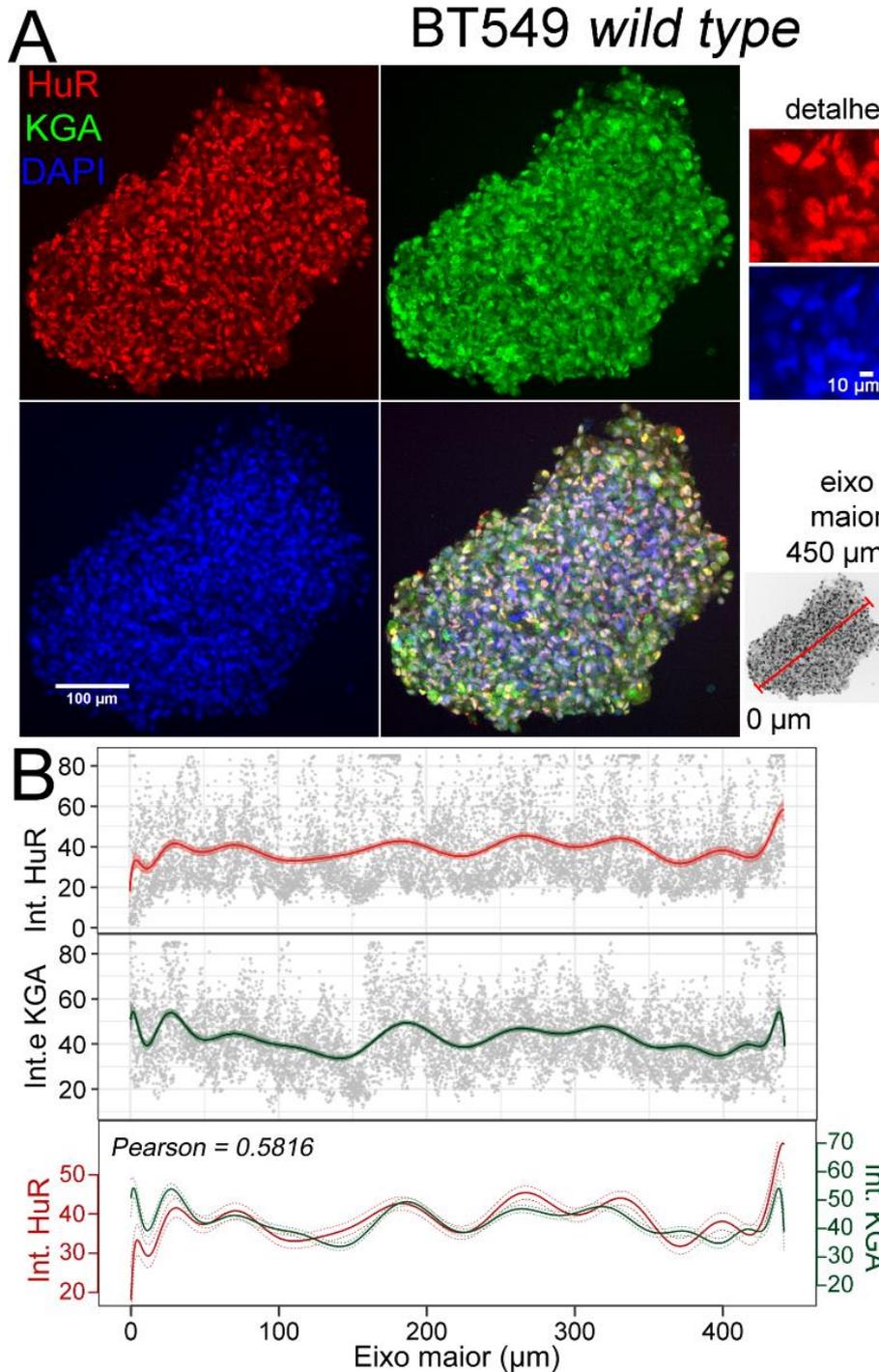
**Figura 50 – Expressão e localização de HuR em células de câncer de mama desafiadas com a depleção de glutamina e avaliadas por imunofluorescência**

(A) Gráfico RDI com intensidade média de HuR em células de câncer de mama cultivadas utilizando soro dialisado e com a suplementação ou não de glutamina. (B) Gráfico RDI com a localização celular de HuR nas mesmas células. (C) Imagens representativas das células avaliadas. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os valores das células quantificadas estão individualmente representados sobre os dados como círculos. Valores de  $p$  originados de análise de variância seguida de teste de Tukey, quando indicados em vermelho representam uma diminuição das médias observadas e em verde um aumento.

Como a supressão de glutamina impactou na localização celular de HuR em BT549, avaliou-se a distribuição da proteína em um ambiente no qual a limitação de nutrientes se dá de forma natural, tais como os esferóides celulares. Sabe-se que esferóides celulares com 134  $\mu\text{m}$  de diâmetro já apresentam limitações de permeação ao oxigênio<sup>314</sup> e os de 200  $\mu\text{m}$  apresentam

dificuldade de acesso para diversas moléculas importantes para o metabolismo celular<sup>315</sup>. Obtivemos esferoides de 400 µm de diâmetro no eixo maior e analisamos cortes dos mesmos quanto a imunomarcção de HuR e KGA. Não observamos uma maior abundância ou localização diferenciais de acordo com a proximidade da borda dos esferoides (Figura 51A), sendo apenas confirmada uma correlação positiva ( $Pearson = 0,5816$ ) entre a expressão de HuR e KGA nas diferentes posições do esferoide (Figura 51B, inferior). Também é possível observar que HuR se localiza predominantemente no núcleo, entretanto, com presença citoplasmática também observável (Figura 51A, detalhe).

Dadas as características de empacotamento das células, inerentes a formação das massas celulares, e falta de marcador citoplasmático nas análises não foi possível quantificar a porcentagem de localização citoplasmática de HuR em função da localização da célula. Concluímos que a abundância de HuR correlaciona-se positivamente com a formação de *actin stress fibers* e com a presença da isoforma KGA – para este último fato, tanto em cultura celular bidimensional quanto tridimensional. Além disto, a disponibilidade de glutamina no meio de cultura impacta de maneira variada nos níveis e localização núcleo-citoplasma de diferentes linhagens celulares de câncer de mama.

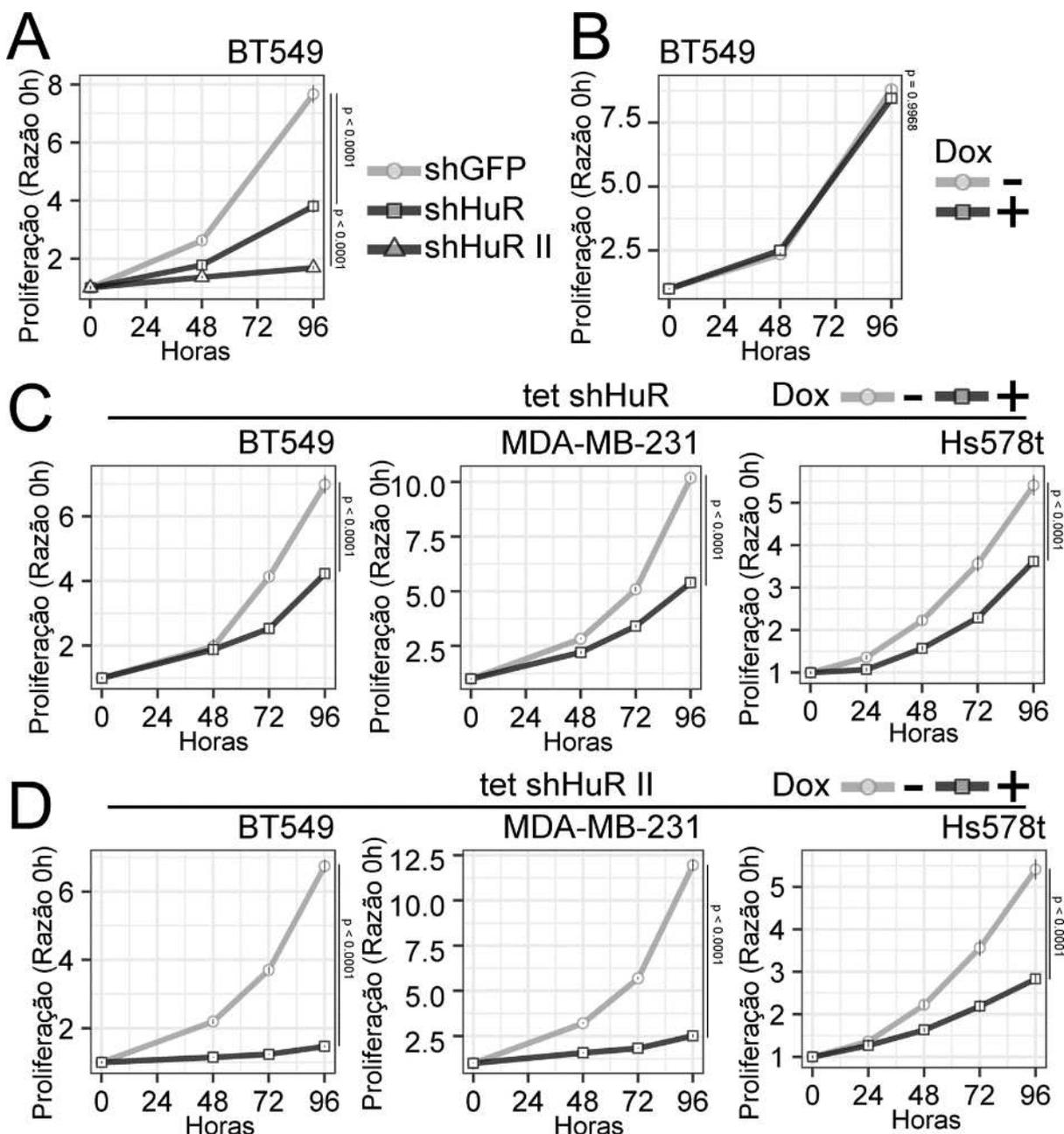


#### 4.5.2 Silenciamento de HuR e o fenótipo obtido em células de câncer de mama

#### 4.5.3 Fenótipo proliferativo

O efeito do silenciamento de *ELAVL1* sobre a proliferação das linhagens BT549, MDA-MB-231 e Hs578t foi avaliado com os sistemas constitutivo e induzível. O silenciamento constitutivo de *ELAVL1* na célula BT549 levou a uma redução do crescimento de 50% (shHuR)

ou 75% (shHuRII) frente ao controle expressando shRNA para GFP, em uma janela de 96 horas (Figura 52A).



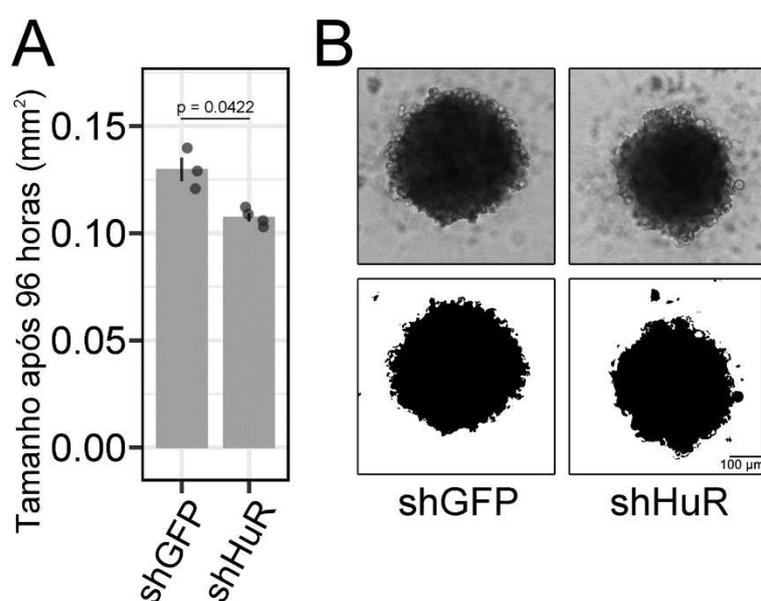
**Figura 52 – Proliferação de linhagens de câncer de mama silenciadas para *ELAVL1***

(A) Curva de proliferação da célula BT549 transduzida com o sistema de silenciamento constitutivo pLKO, para as duas sequências de shRNA. (B) Curva de proliferação da célula BT549 selvagem exposta a doxiciclina. Curvas de proliferação para as células BT549 (esquerda), MDA-MB-231 (centro) e Hs578t (direita) para o sistema induzível tet-pLKO shHuR (C) ou com a sequência de shRNA alternativa (D). Pontos representam a média dos valores obtidos e barras de erro representam o erro padrão da média das quadruplicatas. Valores de  $p$  originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias.

Avaliou-se também o impacto do silenciamento induzido por doxiciclina utilizando sistema induzível tet-pLKO. Inicialmente, incubou-se a linhagem BT549 selvagem com 50

ng/mL de doxiciclina, o que mostrou que doxiciclina não afeta a proliferação da linhagem (Figura 52B). Assim, as células BT549, MDA-MB-231 e Hs578t foram incubadas com doxiciclina durante 7 dias para indução do *knock down*, seguido da avaliação de proliferação por 96 horas. Confirmaram-se os resultados anteriores para a sequência shHuR (Figura 52C) e shHuRII (Figura 52D).

Com intuito de avaliar o efeito do silenciamento constitutivo de *ELAVLI* na linhagem BT549 sobre a proliferação de esferóides, produzimos os mesmos utilizando sistema de *ultra-baixa aderência*. Foi observada uma ligeira, mas significativa, redução na área média dos esferóides produzidos quando no silenciamento de *ELAVLI* (Figura 53A), observada nas imagens representativas de campo claro (Figura 53B, superior). Observa-se uma boa correlação observada entre as máscaras propostas (Figura 53B, inferior) e as imagens reais (Figura 53B, superior), que assim foram utilizadas para a mensuração das áreas.



**Figura 53 – Efeito do silenciamento de *ELAVLI* no crescimento de esferóides de BT549**

(A) Tamanho dos esferóides de células BT549 quando do silenciamento de HuR através do sistema constitutivo pLKO após 96 horas de avaliação. As barras representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico como pontos individuais. *P-values* originados de teste-t de comparação de médias. (B, superior) Imagem representativa de campo claro para os esferóides e (B, inferior) respectiva máscara de quantificação.

#### 4.5.4 Fenótipo migratório e invasivo

Como já comentado, a formação de *actin stress fibers*, mensurada pela coloração com faloidina tem relação com a transição epitélio-mesênquima<sup>309</sup>. Outro elemento importante para o processo de ganho de traços mesenquimais, intimamente relacionado a ganho de capacidade invasiva, é a translocação de beta-catenina para o núcleo; o fator de transcrição beta-catenina, quando translocado para o núcleo, participa da ativação de diversos genes relacionados a tumorigênese<sup>316,317</sup>. A via de sinalização Wnt/beta-catenin apresenta funções

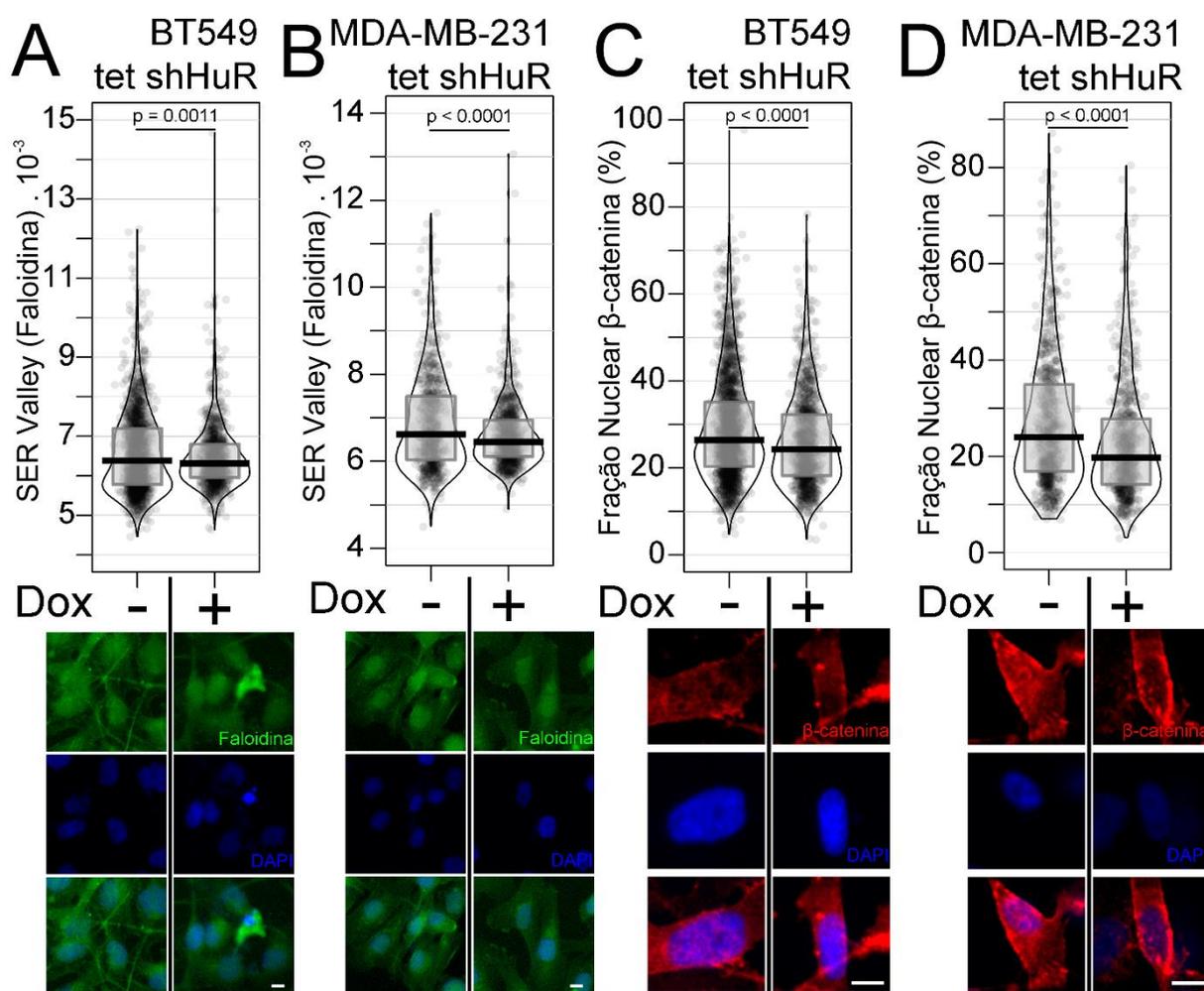
proliferativas e de diferenciação imprescindíveis no desenvolvimento de todos os cordados, apresentando ortólogos de todos os seus constituintes gênicos desde cnidária<sup>318</sup>. Em seu estado de sinalização inativo, a via mantém a proteína  $\beta$ -catenina fosforilada no citoplasma, eventualmente sendo direcionada para a degradação pelo proteassomo; quando ativada pela sinalização extracelular de Wnt na proteína transmembrana Frizzled, ocorre uma liberação de  $\beta$ -catenina (tanto de sua fosforilação quanto da degradação), que permite sua translocação nuclear e atuação como fator de transcrição<sup>318</sup>. Em tumores, por exemplo, a ativação da via Wnt/ $\beta$ -catenin leva ao aumento da *Zinc finger protein SNAI2*, responsável pela troca das caderinas envolvidas nas junções celulares e aumento do fenótipo mesenquimal<sup>319</sup>.

O silenciamento de HuR levou a diminuição da formação do parâmetro de textura *SER Valley*, relacionado com *actin stress fibers*, medido sobre a coloração de faloidina em BT549 (Figura 54A, imagem representativa abaixo) e MDA-MB-231 (Figura 54B, imagem representativa abaixo). Como observado na curva de distribuição das densidades, a região de máximo de densidade na presença de doxiciclina é superior em ambos tipos celulares, mesmo com a redução de média significativa para teste-t. Mesmo com o teorema do limite central assegurando a utilização de abordagem paramétrica, teste de Kruskal-Wallis confirmou (*p-value* inferior a 5%) apenas para MDA-MB-231 a redução, no caso, da mediana. Já foi observado para MDA-MB-231 que a inibição de HuR (através do eixo miR-29a-TTP-HuR) leva a redução de *actin stress fibers*<sup>303</sup>. Em ambas linhagens celulares também observamos que no silenciamento de *ELAVL1* ocorreu a diminuição da localização nuclear de  $\beta$ -catenina (Figura 54C-D, respectivamente, imagens abaixo), mais intensamente observada nas células MDA-MB-231.

O processo migratório celular consiste na capacidade da célula modificar sua posição no espaço, enquanto que o processo invasivo refere-se à capacidade de realizar este mesmo processo com a imposição de uma barreira tridimensional que mimetize a matriz extracelular - envolvendo o remodelamento deste ambiente de forma ativa<sup>236</sup>. Para medir migração celular, realizamos experimentos *in vitro* de *scratch-wound* ou *wound-healing*. Assim, uma monocamada celular confluenta, deprivada de fatores proliferativos (no caso, soro fetal bovino), sofre uma ruptura em formato de linha, a qual tem seu preenchimento por células adjacentes avaliado no tempo<sup>236,320</sup> (Figura 55A).

Mensurando a fração da área aberta no centro do poço que foi recoberta durante o tempo, observamos a ausência de efeito da doxiciclina sobre a capacidade migratória da célula BT549 selvagem (Figura 55B). Quando silenciemos *ELAVL1*, ocorreu significativa da velocidade migratória de BT549 (Figura 55C) e MDA-MB-231 (Figura 55D), com ambas

sequências de shRNA. De forma análoga aos resultados de proliferação, o impacto do shHuR2 foi mais pronunciado.

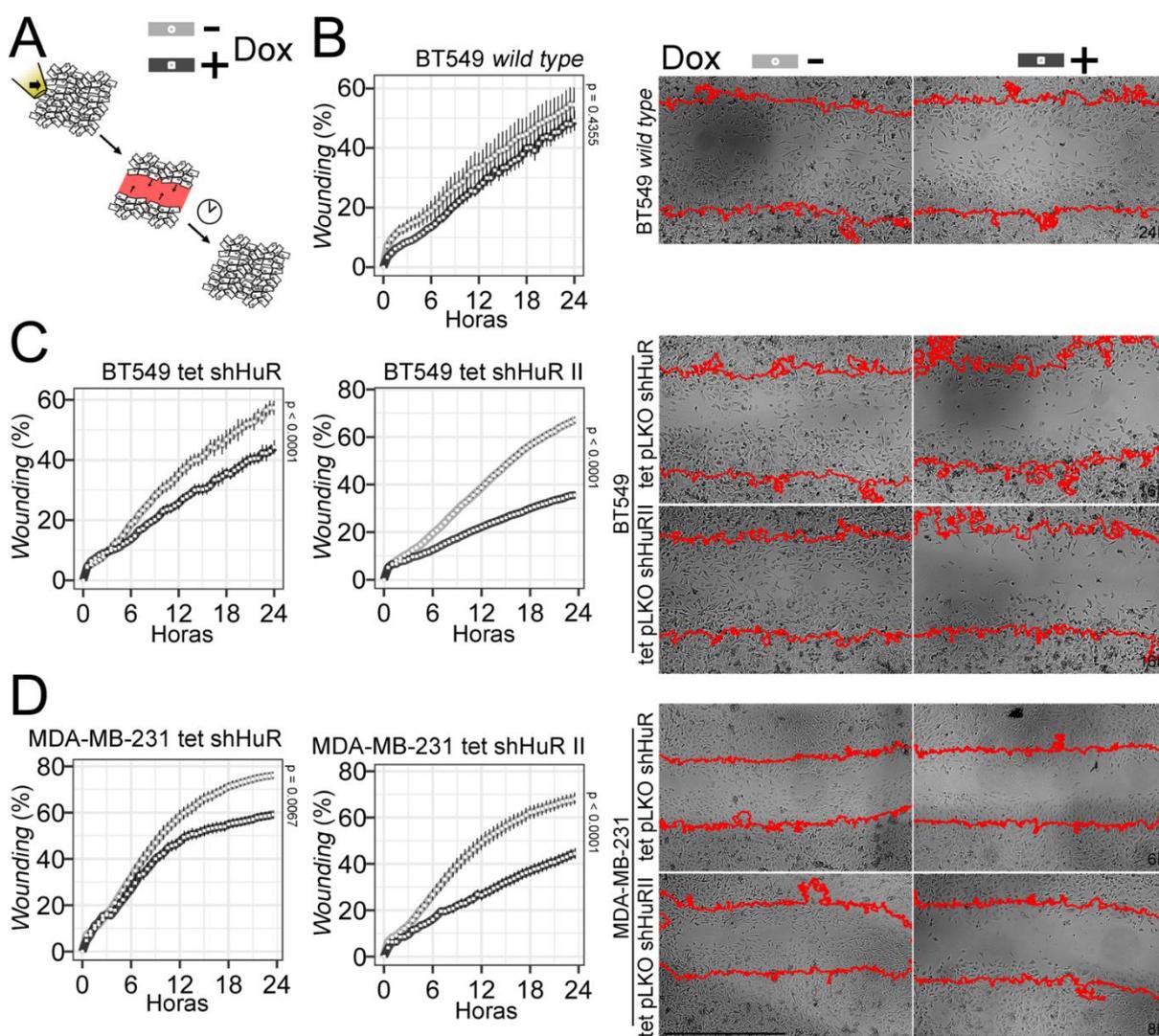


**Figura 54 –** Quantificação de *actin stress fibers* e da localização nuclear de β-catenina frente ao silenciamento de *ELAVL1*

Gráfico RDI para a quantificação do parâmetro *SER Valley* de textura de imagem, que representa a formação de *actin stress fibers* para BT549 (A) e MDA-MB-231 (B). Imagens representativas para marcação de faloidina. Gráfico RDI indicando a fração nuclear de β-catenina medida em BT549 (C) e MDA-MB-231 (D). Imagens representativas para marcação de β-catenina. Barras centrais indicam a mediana; limites inferior e superior da caixa indicam primeiro e terceiro quartis, respectivamente; linhas periféricas representam a densidade suavizada dos dados; os valores das células quantificadas são individualmente representados sobre os dados como círculos. *P-values* originados de teste-t de comparação de médias.

Para avaliar a capacidade invasiva celular, duas abordagens distintas foram utilizadas: ensaio de *scratch-and-wound* embebido em matriz extracelular tridimensional de colágeno do tipo I e ensaio de invasão através de barreira porosa (também conhecida como câmara de Boyden ou *Transwell*<sup>®</sup>) recoberta com matriz extracelular do sarcoma murino Engelbreth-Holm-Swarm (Matrigel<sup>®</sup>)<sup>236</sup>. O primeiro apresenta como vantagem a composição mais densa da matriz, já o segundo é capaz de confrontar a célula não só com a capacidade de

interagir e remodelar a matriz extracelular ao entorno como também demanda a reorganização estrutural e passagem em pequenos poros de diâmetro fixo, os quais não podem ser degradados de maneira enzimática. Uma desvantagem associada ao sistema de *scratch-wound* embebido é a não direcionalidade da quimioatração oferecida em direção ao risco criado na monocamada (o que pode fazer as células migrarem verticalmente e não só horizontalmente).

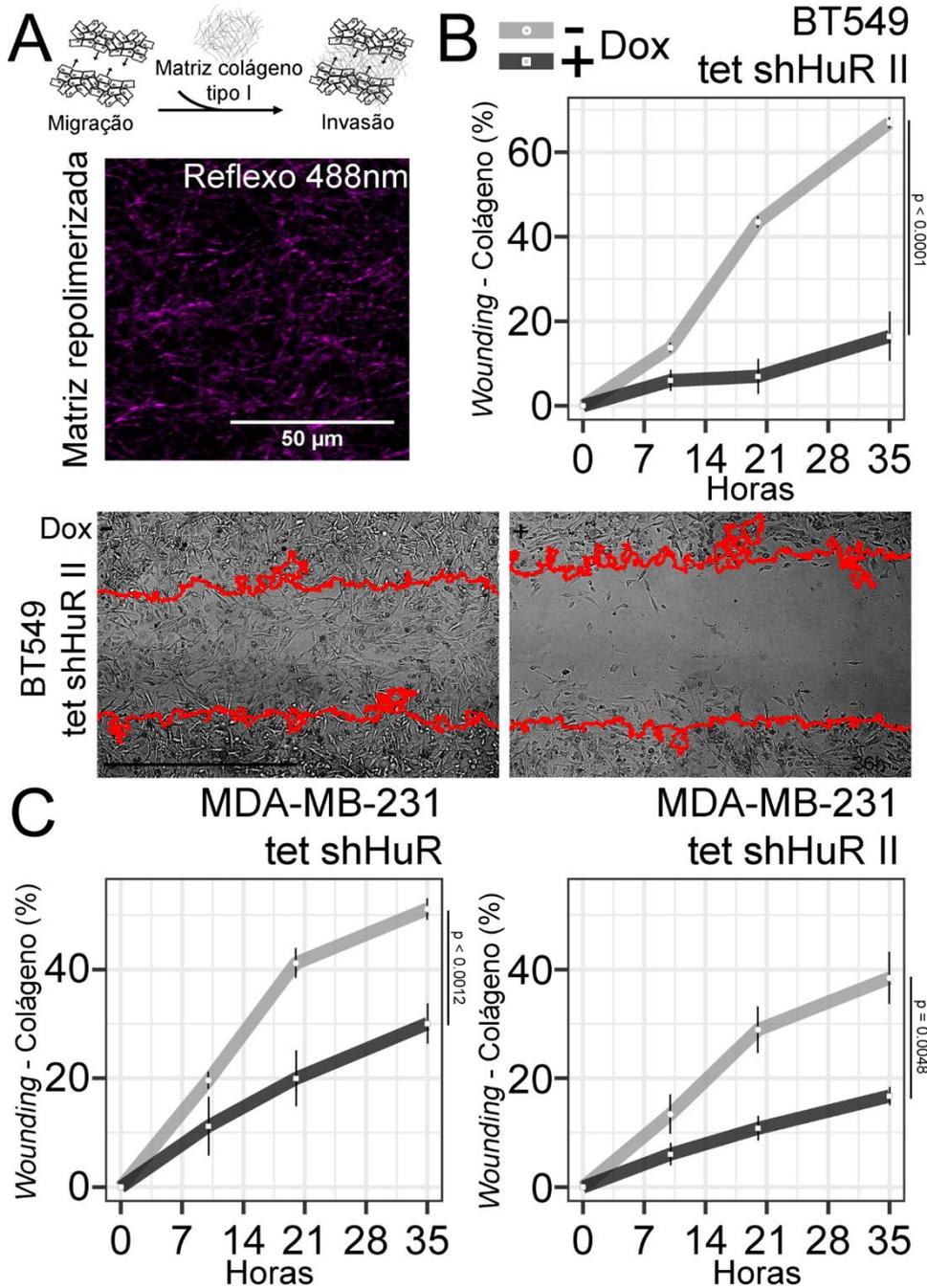


**Figura 55 – Migração celular de células silenciadas para *ELAVL1* medida por *scratch-wound***

(A) Representação esquemática do experimento de *scratch-wound* realizado. (B) Curva de migração da célula BT549 selvagem incubada com doxiciclina. Curva de migração após indução do *knock down* de *ELAVL1* (shHuR e shHuR II) com doxiciclina (e controle) das linhagens BT549 (C) e MDA-MB-231 (D); imagens representativas a direita. Pontos representam a média dos valores obtidos e barras de erro representam o erro padrão da média das replicatas (de ao menos 4 em cada condição). Valores de *p* originados de teste-t de comparação de médias das últimas observações. Barra de escala representa 500  $\mu$ m. Horários indicam a posição da imagem representativa no gráfico de acompanhamento.

Foi realizada inicialmente a extração e purificação do colágeno dos tendões da cauda de ratos Wistar, o qual precisa polimerizar sobre a camada de células, formando a barreira para a movimentação das células (Figura 56A, superior). Para confirmar a polimerização do

colágeno extraído, utilizamos microscopia confocal de reflexão<sup>321</sup> do material, lançando mão da propriedade natural de fibras de colágeno de refletir luz em uma ampla faixa de comprimentos de onda (aqui sendo utilizados 488 nm)<sup>322</sup>. Como esperado, obtivemos fibras de colágeno reconstituídas, como avaliado pela reflexão de luz a 488 nm (Figura 56A, inferior).



**Figura 56** – Migração celular de células silenciadas para *ELAVL1* medida por *scratch-wound* embebido em matriz de colágeno do tipo I

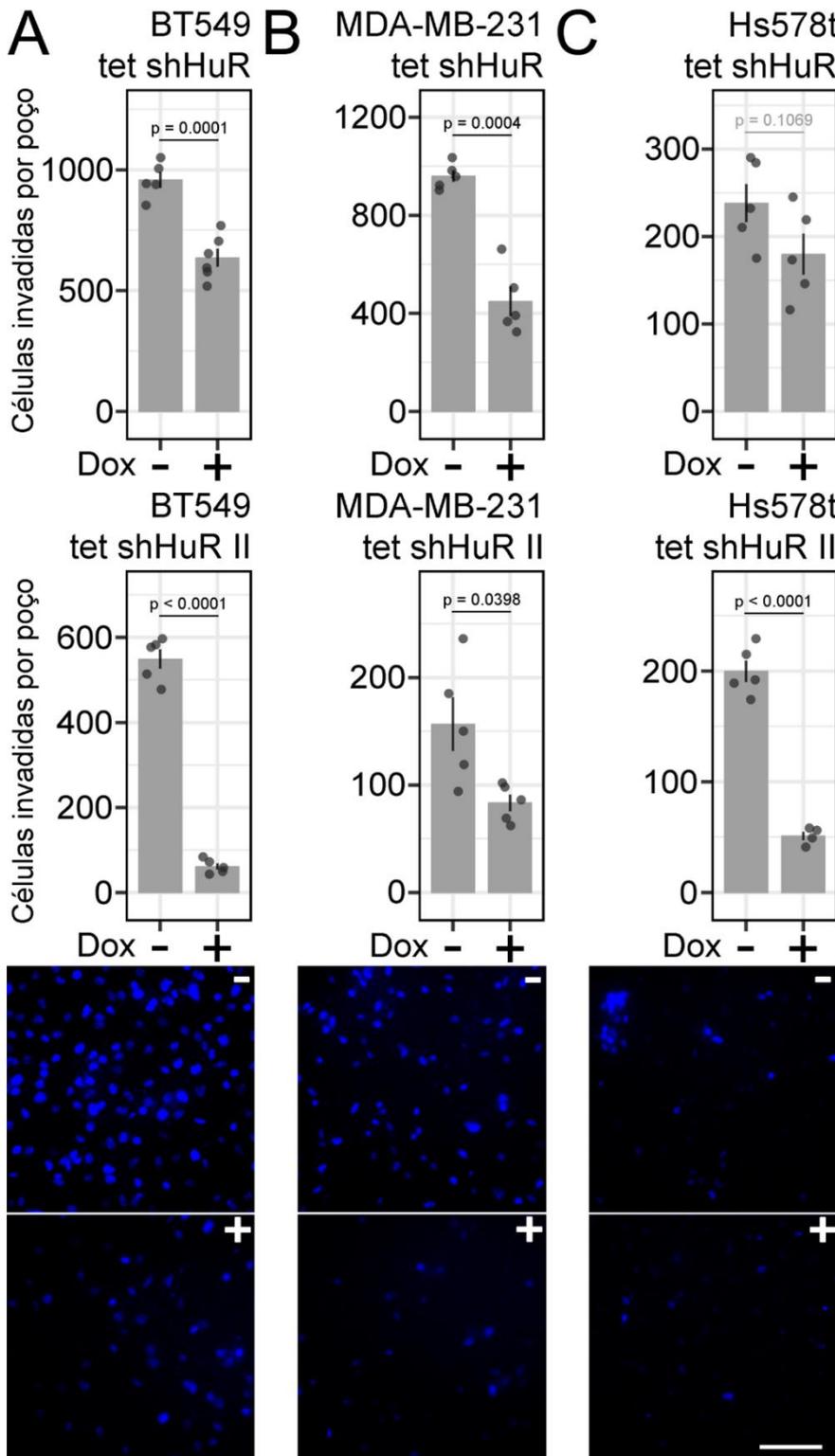
(A) Esquema do experimento realizado, mostrando que células para avançar sobre uma área livre precisam invadir uma matriz de Colágeno tipo I (imagem superior).

Microscopia confocal de reflexão *label-free* para a confirmação da polimerização em fibras do colágeno do tipo I extraído de rabo de *Rattus norvegicus* (imagem inferior). Curva de invasão após indução do *knock down* de *ELAVL1* com doxicilina (e controle) das linhagens BT549 (B, shHuR II, imagens representativas abaixo) e MDA-MB-231 (C, shHuR e shHuR II). Pontos representam a média dos valores obtidos e barras de erro representam o erro padrão da média de, ao menos,

quadruplicatas. *P-values* originados de teste-t de comparação de médias das últimas observações. Barra de escala representa 500 μm.

A segunda metodologia utilizada para mensurar a invasão foi a câmara de Boyden recoberta com Matrigel<sup>®</sup> utilizando como quimioatrativo o fator de crescimento epitelial

(epidermal growth factor, EGF). O silenciamento de *ELAVL1* empregando as sequências shHuR e shHuRII levou a redução na capacidade invasiva das linhagens BT549 (Figura 57A, imagens representativas abaixo), MDA-MB-231 (Figura 57B, imagens representativas abaixo) e Hs578t (Figura 57C, imagens representativas abaixo), não significativo somente para Hs578t na utilização da sequência shHuR.



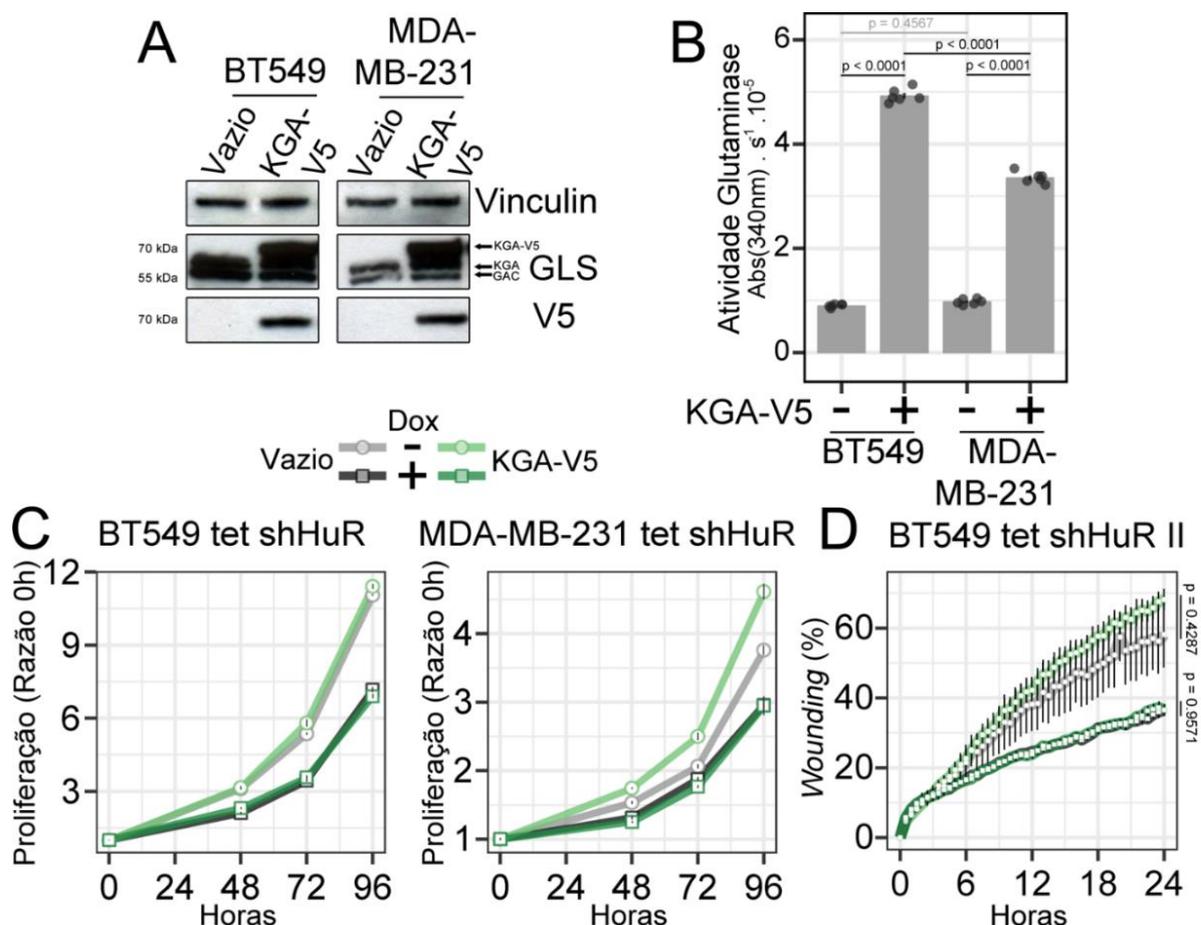
**Figura 57 – Invasão celular em experimento com câmara de Boyden com barreira de matrigel**

Invasão celular através de barreira de matrigel e remodelamento de citoesqueleto por poro de diâmetro médio de 8  $\mu\text{m}$  das célula BT549 (A), MDA-MB-231 (B) e Hs578t (C), com shHuR e shHuRII e imagens representativas da sequência shHuR (abaixo). Barras representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico como pontos individuais. *P-values* originados de teste-t de comparação de médias. Barra de escala representa 100  $\mu\text{m}$ .

O silenciamento de *ELAVL1* reduziu significativamente a capacidade das linhagens BT549 (Figura 56B para shHuRII; imagem representativa abaixo) e MDA-MB-231 (Figura 56C para shHuR e shHuRII, respectivamente) moverem-se na matriz de colágeno, em relação aos respectivos controles sem doxiciclina.

#### 4.5.5 Resgate fenotípico com a re-expressão ectópica da isoforma KGA em células com *ELAVL1* silenciada

Dado que verificamos que o *knock down* de *ELAVL1* leva a diminuição de KGA, assim como diminuição da proliferação e migração celulares, avaliamos a importância desta modulação em glutaminase sobre estes fenótipos. Assim, linhagens celulares estabelecidas com o sistema tet-pLKO-shHuR foram transduzidas com vetores retrovirais, vazio (controle) ou contendo sequência de cDNA para a isoforma KGA (fusionado à cauda imunogênica V5, KGA-V5), e avaliadas quanto aos fenótipos proliferativo e migratório (Figura 58).



**Figura 58 – Efeitos do resgate da expressão de KGA em células com o silenciamento de *ELAVL1***

(A) *Western blot* confirmatório da expressão ectópica de KGA nas células BT549 e MDA-MB-231. Setas indicam as isoformas de *GLS*, incluindo KGA-V5 ectópica. (B) Confirmação da atividade glutaminolítica elevada em lisado celular total quando na expressão ectópica de KGA-V5. Curva de proliferação de BT549 (C) e MDA-MB-231 (D)

na presença ou ausência de doxiciclina e KGA-V5. (E) Curva de migração de BT549 na presença ou ausência de doxiciclina e KGA-V5. Para gráficos de barras, as mesmas representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico como pontos individuais. Para gráficos de linha, pontos representam a média dos valores obtidos e barras de erro representam o erro padrão da média das quadruplicatas. *P-values* originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias.

A expressão de KGA-V5 foi confirmada por *western blot* nas linhagens BT549 e MDA-MB-231 (Figura 58A); como esperado, células expressando KGA-V5 tiveram maior atividade de glutaminase medida de lisado total (Figura 58B). Entretanto, KGA-V5 não foi capaz de recuperar a redução de proliferação causada pelo *knock down* de *ELAVL1* em BT549 (Figura 58C) e MDA-MB-231 (Figura 58D). A mesma ausência de resgate fenotípico foi observada quanto a capacidade migratória (Figura 58E).

Em suma, a diminuição da expressão de KGA resultante do silenciamento de *ELAVL1* não é o fator crucial na redução da proliferação e migração celulares causadas por este último. Sendo HuR um regulador pleiotrópico de diversos passos do metabolismo de RNA na célula, não é de surpreender que o mesmo possa impactar estes fenótipos por outras vias, assim como já publicado<sup>323</sup>.

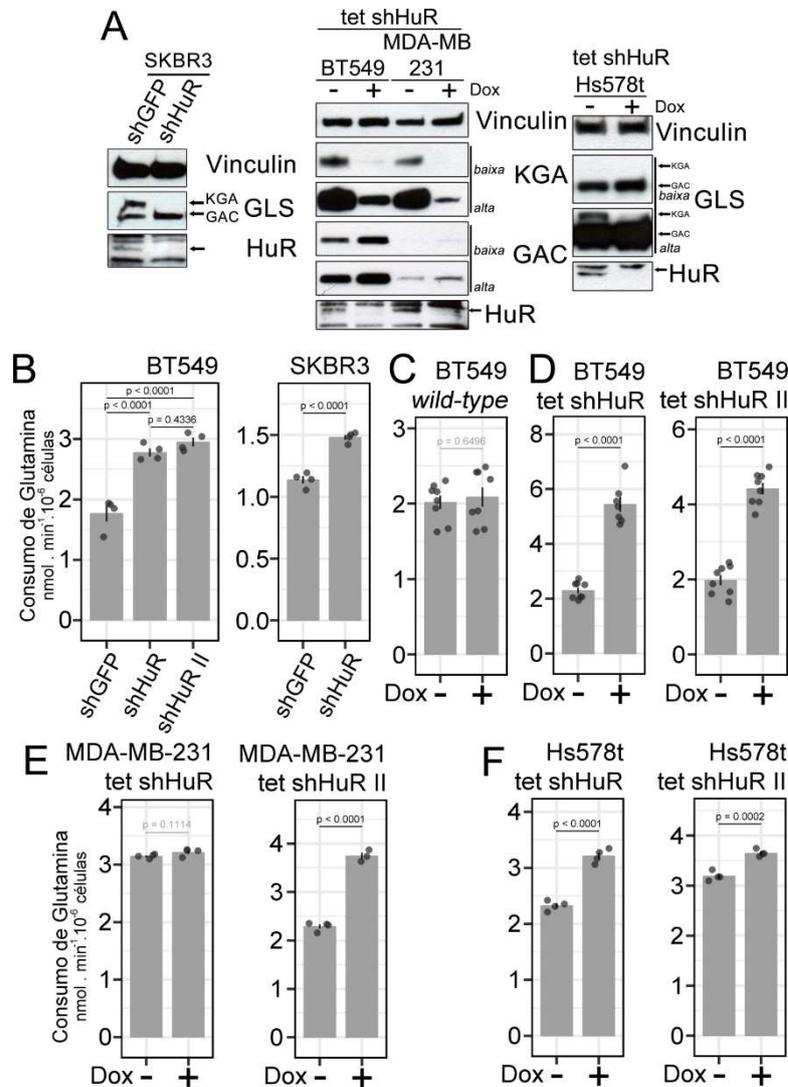
#### 4.5.6 Fenótipo glutaminolítico

Dado que HuR altera a escolha entre as isoformas, favorecendo KGA, especulamos que o mesmo pudesse impactar o metabolismo celular de glutamina. O silenciamento constitutivo de *ELAVL1* (com as sequências shHuR e shHuR2), que causa a troca de KGA por GAC (Figura 59A), isoforma cataliticamente mais ativa<sup>27</sup>, causou aumento no consumo de glutamina das linhagens BT549 (Figura 59B) e SKBR3 (Figura 59C). Estes resultados são coerentes com o fato de que o silenciamento de *ELAVL1* leva a diminuição de KGA e aumento de GAC, isoforma mais. Avaliamos o impacto do silenciamento induzível sobre este fenótipo também. Primeiramente verificamos que doxiciclina não impactava no consumo de glutamina celular de BT549 selvagem (Figura 59C). O *knock down* induzido de *ELAVL1* (com shHuR e shHuR2) levou ao aumento do consumo de glutamina em BT549 (Figura 59D), MDA-MB-231 (Figura 59E, somente para a sequência shHuR2) e Hs578t (Figura 59F). Quando observados em conjunto, os resultados apontam que o silenciamento de *ELAVL1* aumentou o consumo de glutamina das células avaliadas.

Avaliamos também o impacto do silenciamento de *ELAVL1* sobre a atividade glutaminolítica das linhagens BT549 e SKBR3. O silenciamento constitutivo de *ELAVL1* levou

o aumento ligeiro, mas significativo da atividade glutaminolítica de BT549 e SKBR3 (Figura 60A); estes resultados foram confirmados em BT549 com o uso do sistema induzível (sequência shHuR) (Figura 60B). Estes achados vão em acordo com a observação de que células com supressão de HuR consomem mais glutamina.

Especulamos que este aumento da capacidade glutaminolítica das células seguido do silenciamento de *ELAVL1* pudesse impactar na dependência da célula a este aminoácido para proliferar. BT549 demonstrou ser dependente de glutamina para proliferar, uma vez que a remoção de glutamina diminuiu a proliferação de BT549 em 46% após 96 horas (Figura 60C, ausência de doxiciclina). O silenciamento de *ELAVL1* deixou BT549 ainda mais sensível a remoção de glutamina, com diminuição de 65% em relação ao controle com glutamina em 96 horas (Figura 60C).



**Figura 59 – Consumo de glutamina de linhagens celulares silenciadas para *ELAVL1***

(A) *Western blots* previamente mostrados indicando a troca das isoformas frente ao silenciamento de *ELAVL1*. (B) Consumo de glutamina em BT549 (esquerda) e SKBR3 (direita) frente ao silenciamento de *ELAVL1* com as sequências shHuR e shHuR II. (C) Consumo de glutamina de BT549 selvagem na adição de doxiciclina. Impacto do *knock down* induzido de *ELAVL1* no consumo de glutamina de BT549 (D), MDA-MB-231 (E) e Hs578t (F). Gráfico da esquerda, sequência shHuR e da direita, sequência shHuR II. Barras representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico como pontos individuais. *P-values* originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias.

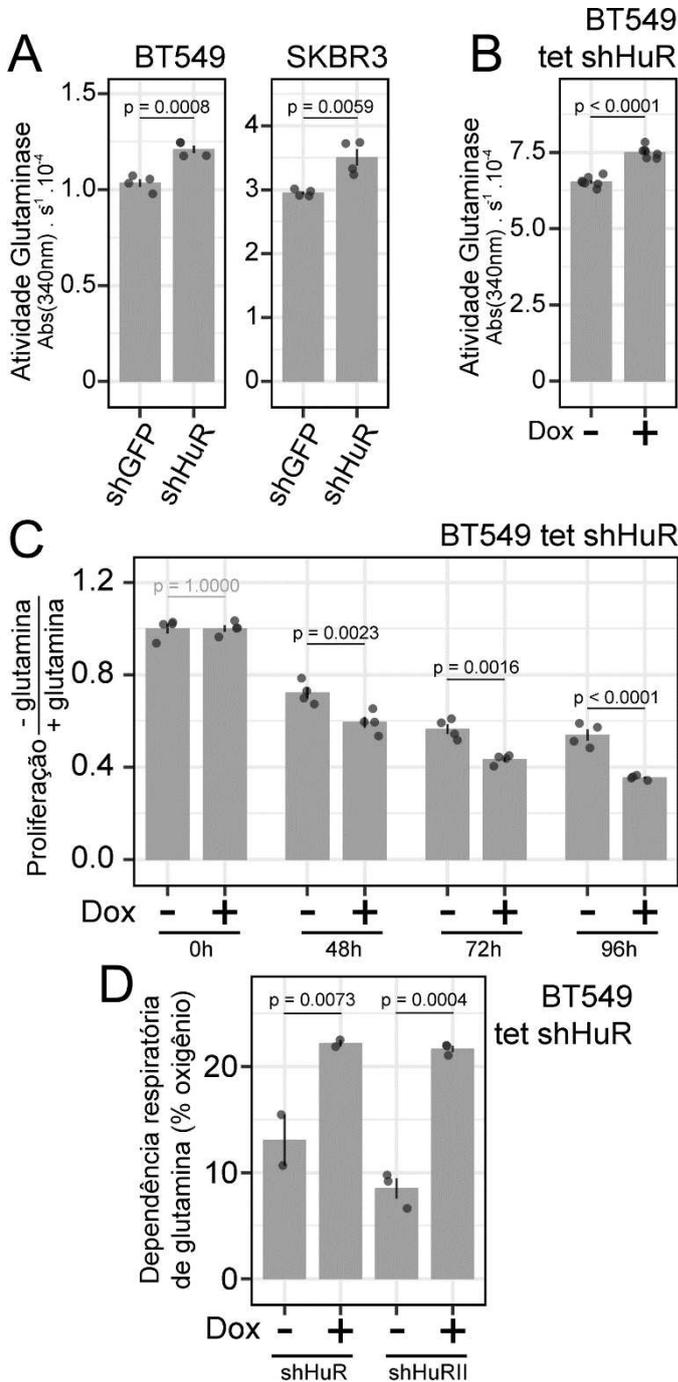
A dependência de glutamina para o consumo de oxigênio foi mensurada utilizando o equipamento Seahorse XF. Neste ensaio, avaliamos em oxígrafo multipoços a respiração celular antes (definida como basal) e depois da adição de inibidor da atividade glutaminolítica (pela molécula CB-839), cessando o fluxo de carbono originado da glutamina para o ciclo do TCA. A razão do consumo de oxigênio depois e antes da inibição mensura a fração da respiração que depende exclusivamente de glutamina, ou seja, que a plasticidade celular não é capaz de imediatamente direcionar para outra fonte de carbono. Verificamos que, em acordo com o aumento da dependência de glutamina ditado pelo silenciamento de *ELAVL1* em BT549, este silenciamento também levou a um aumento na dependência intrínseca de glutamina para ambas as sequências de shRNA avaliadas (Figura 60D). Desta maneira, concluímos que o silenciamento de *ELAVL1* aumenta a dependência celular ao aminoácido glutamina para proliferação e respiração – relacionadas com demandas biossintéticas e energéticas.

As consequências do silenciamento de *ELAVL1* sobre o nível total de diversos metabólitos intracelulares de BT549 foi avaliado por cromatografia gasosa seguida de espectrometria de massas (Figura 61). Metabólitos do ciclo do TCA (citrato e malato), além de glutamina, glutamato, ureia e aspartato foram avaliados. Observamos que o silenciamento de *ELAVL1* levou ao aumento significativo de aspartato, não tendo impactado nos demais (Figura 61A). Adicionalmente, diversos outros metabólitos foram identificados, alguns deles mostrados. Destes, verificamos que o *knock down* de *ELAVL1* levou a aumento significativo de metionina (Figura 61B).

A estratégia de análise utilizada implicou na detecção de *features* metabólicas compartilhadas que, posteriormente, são comparadas com uma base de dados de fragmentação de massas. Neste sentido, algumas *features* identificadas podem não ser relacionadas a moléculas conhecidas da base empregada. Análise de componentes principais dos metabólitos avaliados mostrou que o silenciamento de *ELAVL1* impactou no metabolismo celular gerando uma grande separação espacial das réplicas biológicas analisadas (Figura 61C). Observamos, então, que a diminuição dos níveis de HuR altera a abundância de alguns metabólitos intracelulares, especialmente aspartato e de metionina.

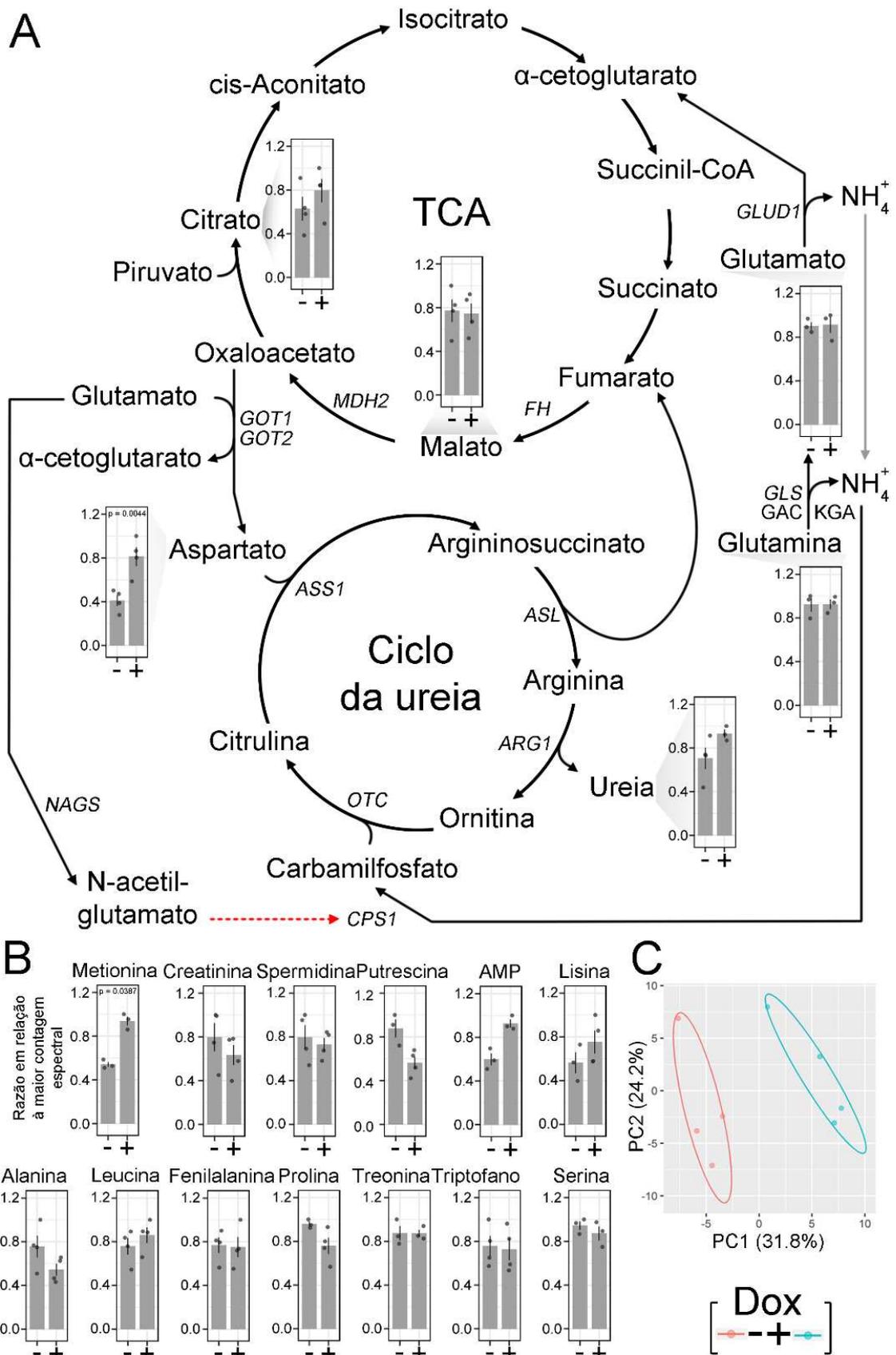
Dado que o silenciamento de *ELAVL1* levou ao aumento de GAC, seguido de maior atividade glutaminolítica e dependência das células a glutamina, verificamos o impacto da combinação do *knock down* de *ELAVL1* com a inibição de glutaminase por CB-839, atualmente em fase clínica de estudos para diversos tipos tumorais, incluindo mama. CB-839 é um inibidor de ambas isoformas de glutaminase, KGA e GAC, não afetando a atividade do produto do parálogo *GLS2*<sup>17</sup>. Nossa hipótese seria que células parcialmente responsivas ao tratamento com

CB-839 poderiam ser direcionadas para um fenótipo mais glutaminolítico com a supressão de HuR e, nesta situação, responderem mais ao fármaco.



**Figura 60 – Dependência de glutamina para os fenótipos proliferativos e de respiração em linhagens silenciadas para *ELAVL1***

(A) Atividade glutaminolítica de lisado celular de BT549 (esquerda) e SKBR3 (direita) com *knock down* constitutivo de *ELAVL1* e controle. (B) Atividade glutaminolítica de lisado celular total de BT549 com *knock down* induzido de *ELAVL1* e controle. (C) Impacto da remoção de glutamina do meio de cultivo e silenciamento induzido de *ELAVL1* na proliferação de BT549. (D) Dependência respiratória de glutamina mensurada pela razão entre o consumo de oxigênio no basal celular e após a adição de CB-839, inibidor de glutaminase, feito no oxígrafo SeaHorse. Para gráficos de barras, as mesmas representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico como pontos individuais. Para gráficos de linha, pontos representam a média dos valores obtidos e barras de erro representam o erro padrão da média das quadruplicatas. *P-values* originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias.



**Figura 61 –** Quantificação do *pool* geral de diversos metabólitos intracelulares em BT549 silenciada para *ELAVL1*  
 (A) Mapa simplificado do TCA, de sua anaplerose por glutamina e a relação com o ciclo da ureia, com a indicação dos intermediários identificados e seus níveis antes e após adição de doxiciclina. (B) Metabólitos diversos

identificados, com ênfase em aminoácidos. (C) Análise de componentes principais denotando a separação entre expressão e silenciamento de *ELAVLI* quando utilizando todos os metabólitos identificados e desconhecidos. *P-value* oriundos da avaliação de *limma*. Para gráficos de barras, as mesmas representam a média dos valores obtidos, barras de erro representam o erro padrão da média e as réplicas experimentais estão indicadas sobre o gráfico como pontos individuais.

#### 4.5.7 Combinação do silenciamento de HuR a inibição química da glutaminólise celular

*ELAVLI* foi silenciada constitutivamente em BT549 e tratada ou não com CB-839 por até 96 horas. Tanto o uso do shHuR quando shHuR2 levou a redução na proliferação de BT549, como esperado, em relação ao controle shGFP (Figura 62A). Quando adicionada a inibição química, enquanto que em células controle o tratamento com CB-839 levou apenas a uma redução de 5 vezes da proliferação celular, em células shHuR e shHuR2, o tratamento levou a diminuição do número de células em relação ao plaqueado, indicando aumento de morte celular (Figura 62A).

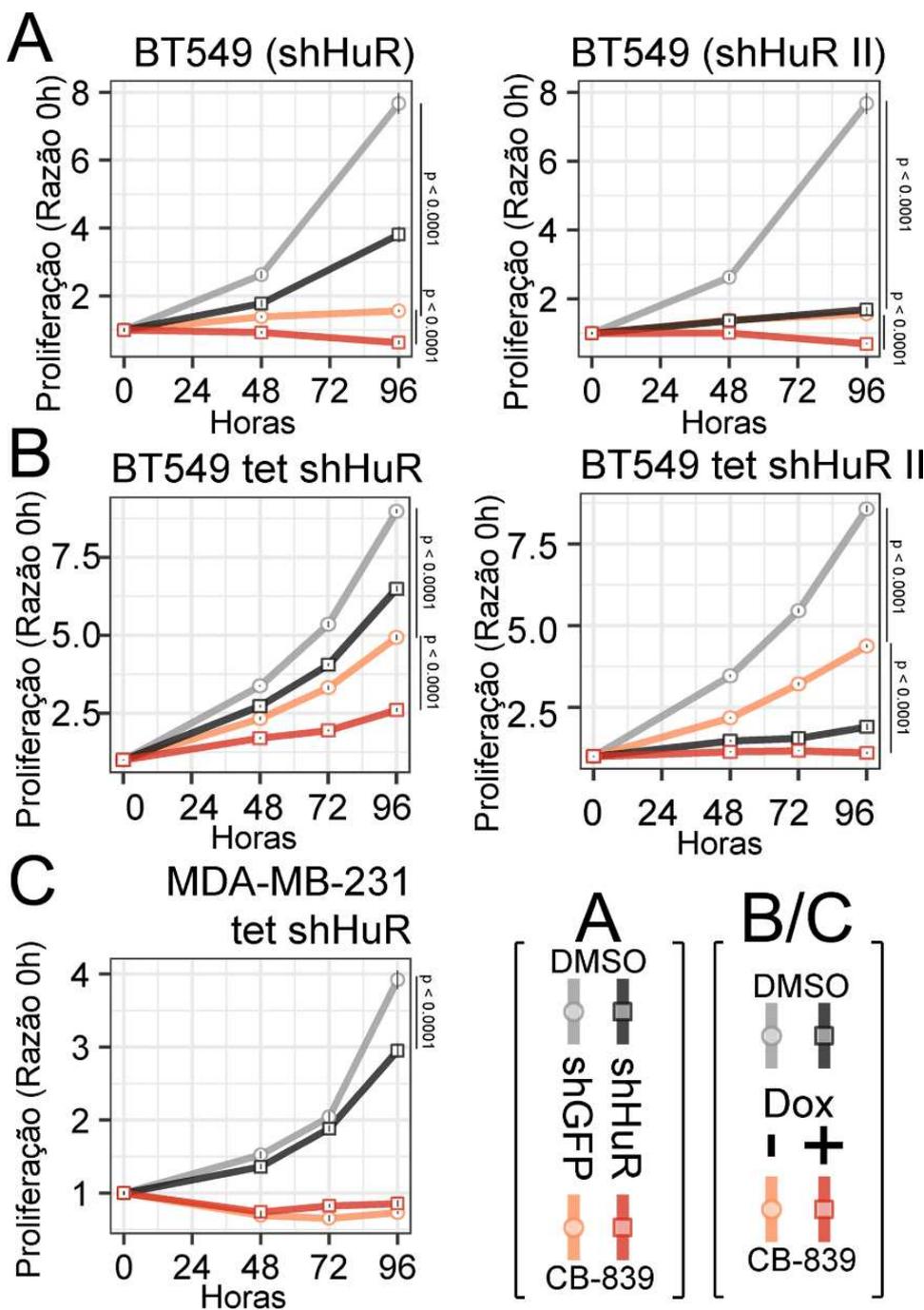
O silenciamento induzido de *ELAVLI* levou a uma redução de 1,4 e 4,5 vezes (shHuR e shHuR2, respectivamente) da proliferação, como esperado (Figura 62B). Enquanto que em células controle o tratamento com CB-839 levou a uma redução 4,9 vezes da proliferação celular, o tratamento combinado ao silenciamento de *ELAVLI* levou a redução da proliferação em 3,5 vezes na combinação com shHuR e diminuição do número de células em relação ao plaqueado na combinação com shHuR2 (morte celular) (Figura 62B). A linhagem MDA-MB-231, extremamente sensível ao composto CB-839 na dose utilizada, como descrito na literatura<sup>17</sup>, não apresentou ganho de inibição de proliferação/morte com o *knock down* de *ELAVLI* (Figura 62C).

Realizando o cálculo do coeficiente de interação de drogas<sup>219</sup> (CDI, *Coefficient of drug interaction*) em 96 horas para a célula BT549, a combinação entre a sequência shHuR e 1 µmol/L de CB-839 indicou sinergia entre os efeitos (CDIs de 0,81 e 0,73, para silenciamento constitutivo e induzível, respectivamente). Dado o efeito pronunciado do silenciamento mediado pelo shHuR2, a sua combinação com CB-839 indicou antagonismo entre os efeitos (CDIs de 2,00 e 1,14, para expressão de shRNA constitutivo e induzível, respectivamente), visto que a combinação dos efeitos anti-proliferativos dos eventos em separado supera o efeito anti-proliferativo da combinação experimental.

Cabe ressaltar que a observação de antagonismo não implica em que as duas abordagens – quando utilizadas em conjunto – não apresentam maior efeito no fenótipo celular em avaliação. Este resultado pode ser interpretado da seguinte maneira: como o efeito do shHuR2 (que pode ter efeitos inespecíficos *off target* levando a uma redução mais intensa da

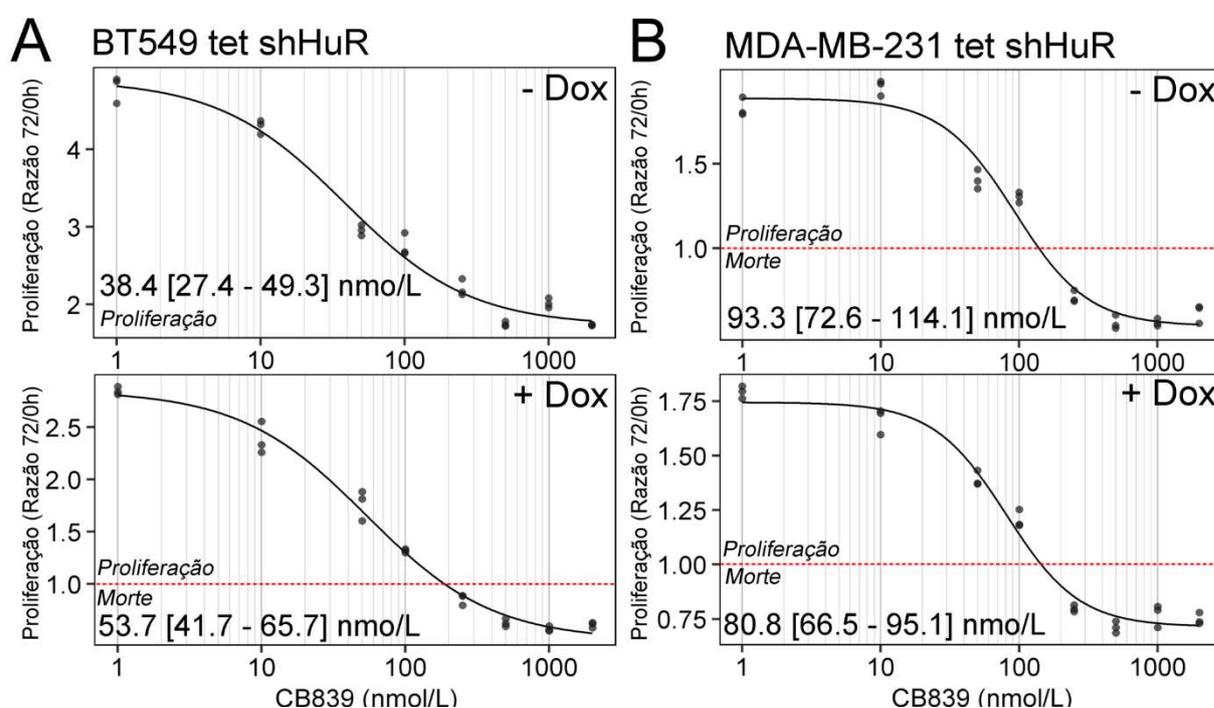
proliferação em relação a shHuR) e tratamento com CB-839, separadamente, levam a completa parada de crescimento em 96 horas, e a combinação leva a morte em 96 horas (não sendo possível mensurar maior janela de efeito da combinação dos tratamentos) obtém-se da matemática empregada esta “impressão” de não haver sinergia. Assim, o efeito exacerbado de shHuRII pode estar mascarando a sinergia com CB-839 em BT549.

Diferente de BT549, MDA-MB-231 com shHuR+CB-839 apresentou CDI de 1,56, indicando algum antagonismo. Assim sendo, a sinergia ou o antagonismo de CB-839 e o silenciamento de *ELAVL1* se mostrou dependente da linhagem avaliada.



**Figura 62** – Proliferação celular na combinação do silenciamento de *ELAVL1* com a inibição química de glutaminase. Curva de proliferação de BT549 expressado constitutivamente (A) ou de maneira induzida (B) shHuR (esquerda) ou shHuRII (direita). (C) Curva de proliferação em células MDA-MB-231 expressado de maneira induzida shHuR. Pontos representam a média dos valores obtidos e barras de erro representam o erro padrão da média das quadruplicatas. *P-values* originados de análise de variância seguida de um teste de Tukey para as médias.

Avaliamos o efeito do *knock down* de *ELAVL1* sobre o IC50 de CB-839 e inibidor de glutaminase BPTES (análogo menos potente do CB-839). Não foi observada uma variação do IC50 de CB-839 (Figura 63) e BPTES (Figura 64) que não residisse no intervalo de confiança dos valores estimados. Como esperado, em BT549, maiores doses tanto de CB-839 quanto BPTES leva a diminuição celular abaixo do nível plaqueado (sugerindo morte celular) quando o silenciamento de *ELAVL1* é induzido; em MDA-MB-231 o tratamento celular com os inibidores já evidencia este potencial efeito de morte nas mais altas doses tanto com expressão ou silenciamento de *ELAVL1*. Assim, a supressão de HuR leva a aumento de GAC o que confere maior sensibilidade ao inibidor CB-839 (potencial aumento de morte celular) sem impactar no IC50 do inibidor.

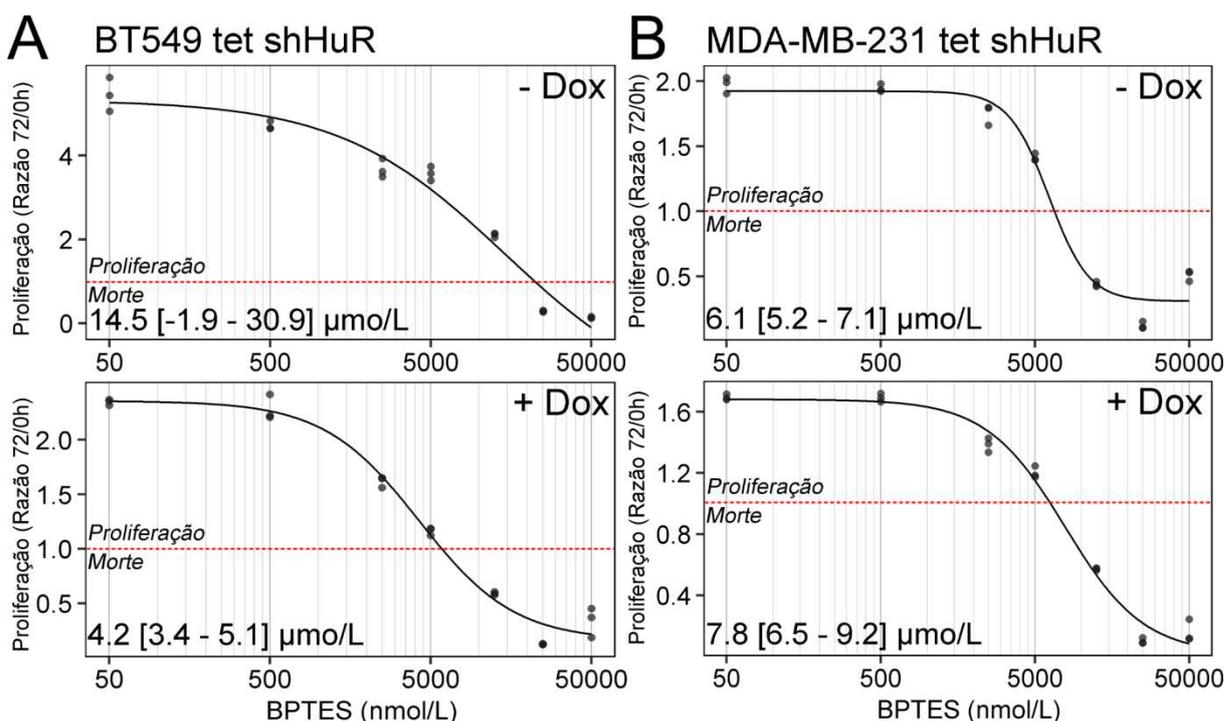


**Figura 63 – Avaliação do IC50 do composto inibidor de glutaminase CB-839**

Curvas de IC50 para BT549 (A) e MDA-MB-231 (B) com silenciamento de *ELAVL1* doxíciclina para o composto CB-839. Pontos são os valores obtidos experimentalmente, eixo das abscissas em escala logarítmica com os valores traduzidos em escala linear para melhor compreensão. Linha fechada representa a regressão log-logística com 4 parâmetros<sup>324</sup> que deriva os valores de IC50 e intervalo de confiança do mesmo. Linha tracejada vermelha indica ponto acima do qual ocorreu proliferação celular e abaixo morte celular em relação ao plaqueado.

Avaliamos também o efeito combinado do silenciamento de *ELAVL1* com o tratamento com CB-839 sobre a migração e invasão celular de BT549. Como esperado, o silenciamento de *ELAVL1* levou a diminuição da migração celular (Figura 65A, linha cinza escuro) enquanto que o tratamento com CB-839 não diminuiu este parâmetro (Figura 65A, linha vermelha clara); a combinação do silenciamento de *ELAVL1* com o tratamento com CB-839

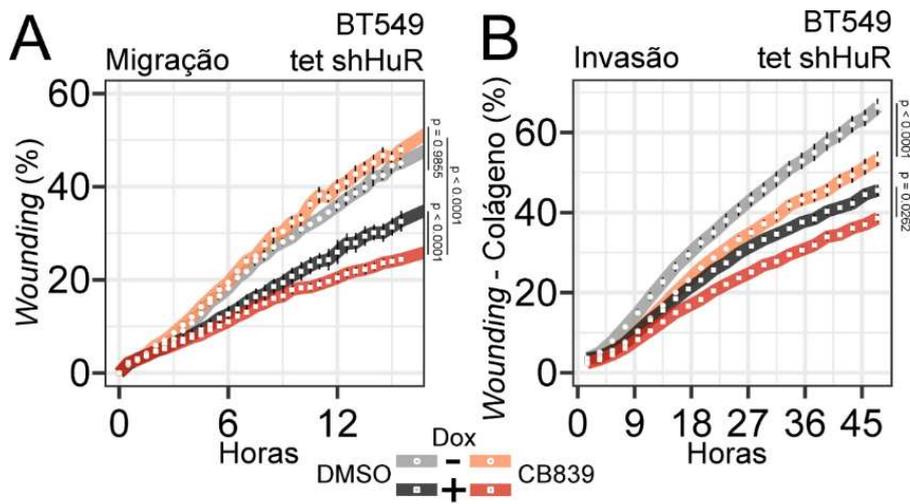
leveu a uma maior diminuição da migração celular que o observado só com o *knock down* de *ELAVL1* (linha vermelha escura). Enquanto que tanto o tratamento com CB-839 quando o silenciamento de *ELAVL1*, levaram, por si, a diminuição da invasão celular (Figura 65, linhas escuras), a combinação dos dois aumentou a inibição da invasão. Observando a interação entre os tratamentos, ocorre sinergia (CDI = 0,70) para migração e aditismo/breve antagonismo para invasão (CDI = 1,06).



**Figura 64 – Avaliação do IC50 do composto inibidor de glutaminase BPTES**

Curvas de IC50 para BT549 (A) e MDA-MB-231 (B) com silenciamento de *ELAVL1* doxíciclina para o composto BPTES. Pontos são os valores obtidos experimentalmente, eixo das abscissas em escala logarítmica com os valores traduzidos em escala linear para melhor compreensão. Linha fechada representa a regressão log-logística com 4 parâmetros<sup>324</sup> que deriva os valores de IC50 e intervalo de confiança do mesmo. Linha tracejada vermelha indica ponto acima do qual ocorreu proliferação celular e abaixo morte celular em relação ao plaqueado.

Concluimos que a supressão de HuR aumenta a dependência de células BT549 a atividade glutaminolítica. O silenciamento de *ELAVL1* combinado com a inibição de glutaminase diminui a proliferação, migração e invasão celular *in vitro*, sugerindo potencial efeito terapêutico combinatório positivo para o tratamento de câncer de mama.



**Figura 65 – Migração e invasão celulares de BT549 na combinação do silenciamento de *ELAVL1* com a inibição química de glutaminase** (A) Curva de migração da linhagem BT549 expressando de maneira induzida shHuR na presença ou ausência do inibidor de glutaminase CB-839 e de doxiciclina. (B) Curva de invasão com matriz tridimensional de colágeno da linhagem BT549 expressando de

maneira induzida shHuR na presença ou ausência do inibidor de glutaminase CB-839 e de doxiciclina. Pontos representam a média dos valores obtidos e barras de erro representam o erro padrão da média de, ao menos, quadruplicatas. Valores de  $p$  originados de análise de variância seguida de um teste de Tukey para as médias em cenários com mais de dois grupos ou teste-t para comparações binárias.

## 5 DISCUSSÃO

### 5.1 ANÁLISE PAN-CANCER DA IMPORTÂNCIA DE HUR PARA O PROGNÓSTICO DE PACIENTES

HuR (codificada pelo gene *ELAVL1*) é uma proteína de ligação à RNA apresentando-se atuando em diversas fases do processamento do mRNA, como no *splicing* alternativo<sup>92</sup>, poliadenilação alternativa<sup>101</sup>, transporte para o citoplasma<sup>325</sup>, promoção<sup>326</sup> ou redução<sup>327</sup> da estabilidade, promoção<sup>328</sup> ou redução<sup>329</sup> da tradução ou mesmo localização do mRNA e as proteínas traduzidas<sup>81</sup>.

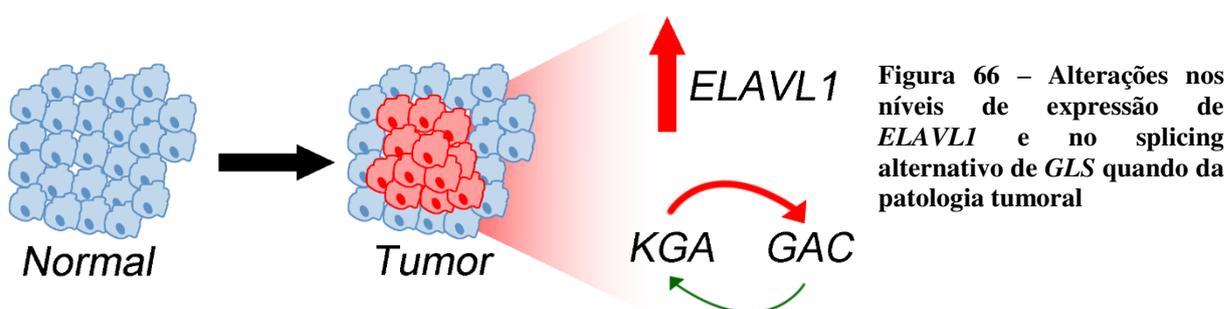
Dadas contribuições para a regulação da expressão gênica como um todo, sua relação com o câncer foi proposta devido ao possível impacto pleiotrópico de sua regulação. Em tumores de distintas origens topográficas, como bexiga<sup>330</sup>, pulmão<sup>116</sup>, esôfago<sup>331</sup>, laringe<sup>332</sup>, mama<sup>117</sup> entre outras<sup>110</sup>, a elevada abundância citoplasmática da proteína se correlaciona com um prognóstico pior para o paciente. Tal mecanismo é justificado pela importante relação com a estabilização de mRNAs de proteínas como ciclooxigenase-2 (COX-2) e o *Vascular endothelial growth factor* (VEGF)<sup>88</sup>, a primeira relacionada com a síntese de prostaglandinas e a modulação da resposta inflamatória do organismo<sup>333</sup> e a segunda com o processo de angiogênese<sup>334</sup>.

Em outro conjunto de evidências da literatura, a localização nuclear de HuR em si se relaciona com um prognóstico negativo para o paciente, tais como em tumores de vesícula biliar<sup>335</sup> ou ovário<sup>336</sup>. Alguns autores levantam, então, que o provável efeito de HuR em ambos os compartimentos (núcleo e citoplasma) levam à progressão tumoral e a métrica mais adequada estaria na razão de localização da proteína entre os dois espaços<sup>337</sup>.

Independente das métricas, existe um crescente questionamento acerca do impacto das funções nucleares de HuR na homeostase e na patologia<sup>338</sup>. Por exemplo, HuR participa da regulação do *splicing* do fator de *splicing* TRA2B promovendo a inclusão do exon 2 do gene e impedindo, assim, a tradução correta da proteína<sup>98</sup>. O gene do receptor promotor de apoptose Fas, por sua vez, tem o exon 6 removido por HuR durante o *splicing*, o que leva a expressão de uma forma solúvel capaz de prevenir algumas vias de apoptose<sup>90</sup>. Outro exemplo é o gene da histona deacetilase III (*SIRT1*), onde HuR promove a exclusão do exon que codifica a porção catalítica (8), modulando sua função regulatória de maneira ainda não completamente compreendida<sup>91</sup>. Tais evidências, em conjunto, suportam um crescente interesse nas funções

nucleares de HuR nos contextos fisiológico e tumoral, em especial na regulação do *splicing* alternativo.

Avaliamos o nível de expressão global de *ELAVL1* e a relação do mesmo com a de recidiva de pacientes em diferentes tipos tumorais empregando dados disponíveis publicamente pelo *The Cancer Genome Atlas*. Encontramos que *ELAVL1* é superexpresso em tumores de diversas origens (Figura 12, Figura 66 e Tabela 2), como estômago, bexiga, pulmão, cabeça-e-pescoço, rim, cérvix uterino, fígado, mama e ovário, em acordo com dados da literatura<sup>104,109,110,116,330–332,339–345</sup>. Também é conhecido que em tumores de pele há elevados níveis de expressão de *ELAVL1* em relação a tecido normal<sup>346</sup>; nos dados avaliados por nós, provavelmente devido ao baixo número amostral de tecido normal, não encontramos a mesma relação. Com relação aos raros tumores adrenais (feocromocitomas e paragangliomas, PCPG) encontramos menor expressão de *ELAVL1* em tecidos tumorais, o que vai de encontro ao previamente publicado<sup>347</sup>, o que pode, acreditamos, ser também devido ao baixo tamanho amostral dos tecidos normais. Assim, concluímos, a expressão de *ELAVL1* é elevada em diversos tipos tumorais, indicando uma potencial importância global deste regulador de RNA para o câncer.



Buscamos compreender em mais detalhes o potencial efeito de HuR para a doença, buscando correlações entre padrões de expressão gênica e o prognóstico ou características clinicopatológicas dos pacientes acometidos<sup>348–352</sup>. Dos trinta tumores analisados, em três (com *p-value* inferior a 1%, sendo seis controlando a descoberta falsa em 5%) a alta expressão de *ELAVL1* relacionou-se com maior probabilidade de recidiva, tendo sido eles adenocarcinoma pancreático ductal (PRAD), carcinoma renal cromóforo (KICH) e melanoma uveal (UVM) (Figura 13, e Tabela 3).

Os pacientes com câncer de pâncreas com alta expressão de *ELAVL1* apresentaram um prognóstico positivo, com uma proporção de risco 5,9 vezes menor para a ocorrência de recidiva (Tabela 3). Todavia, HuR é apontada como proteína imprescindível para o

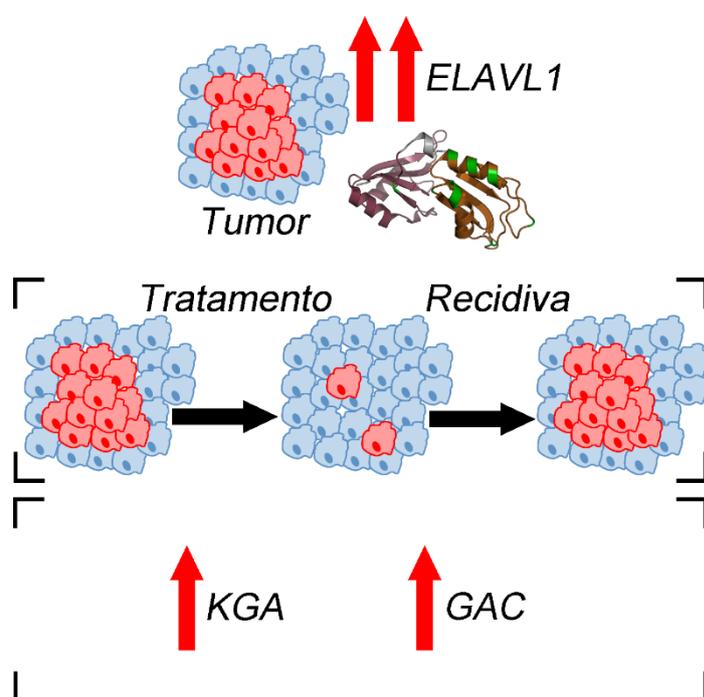
desenvolvimento do tumor de pâncreas, dado que seu silenciamento implica em redução de parâmetros de malignidade (como migração, invasão e desenvolvimento de tumores *in vivo*)<sup>353-355</sup>. Neste sentido, inclusive, já foi proposto uma oportunidade terapêutica para inibição de HuR concomitante a inibição da Poli (ADP-Ribose) Polimerase<sup>353,355</sup>. Gemcitabina, medicamento indicado como primeira-linha de tratamento para câncer de pâncreas<sup>356</sup>, correlaciona-se com a translocação de HuR do núcleo para o citoplasma em células desta origem, onde a mesma promove a tradução da enzima deoxicitina cinase<sup>357</sup>. Tal enzima é imprescindível para a atuação de gemcitabina como análogo de pirimidina, inibindo a síntese de DNA e, conseqüentemente, a proliferação celular. Costantino e colaboradores<sup>357</sup> reportaram um risco relativo aumentado em 4,48 [1,49–13,5] (*p*-valor 0,0036) para pacientes com reduzidos níveis de HuR em regime terapêutico de gemcitabina frente aos pacientes com elevados níveis de HuR sob o mesmo tratamento. Existe, inclusive, estudo de caso na literatura em que o mecanismo de resistência e escape à terapia com o composto se deu pela redução dos níveis de HuR<sup>358</sup>. Coerentemente, do conjunto de dados do TCGA, para os quais a informação era disponível, ao menos 36% dos pacientes com alta expressão de *ELAVL1* realizaram abordagem terapêutica com gemcitabina, o que poderia justificar, pelo menos em parte, o melhor prognóstico verificado neste tumor nos dados do TCGA. Uma outra explicação poderia ser uma incongruência entre nível de mRNA e proteína, o que precisaria ser avaliado com futuros dados de proteômica adicionados à plataforma.

Carcinoma renal cromóforo e melanoma uveal são tumores raros para os quais também observamos significativa relação com prognóstico positivo (com um valor de *p* inferior a 1%). Não há estudos na literatura correlacionando a presença de HuR e um prognóstico positivo ou negativo para o paciente para estes tipos tumorais.

Para quatro outros tumores (câncer de mama - BRCA -, carcinoma de células escamosas de cabeça e pescoço - HNSC -, carcinoma hepatocelular - LIHC - e feocromocitoma & paraganglioma adrenal - PCPG) verificamos aumento do risco de recidiva na alta expressão de *ELAVL1* (Figura 14, Figura 15 e Tabela 4). Dentre estes, nossas análises apontou o câncer de mama como o detentor do Cox HR de prognóstico negativo mais significativo (Figura 67).

Dado que, por exemplo, em câncer de mama, estudos utilizando grandes painéis gênicos randômicos (superiores a 100 genes), 90% das permutações levam à descoberta de assinaturas gênicas que se correlacionam com o prognóstico dos pacientes<sup>359</sup>, a metodologia de validação estatística destes eventos é amplamente discutida<sup>137,138,143,360-363</sup>. Com a abordagem aqui utilizada, que compara diversos valores de cortes da expressão a uma variável clínica<sup>137,138,360</sup>, o risco de erros do tipo I (observação de tumores com *p-value* significativo nos

quais não há real impacto na sobrevida) é elevado<sup>361</sup>. Para confirmar nosso achado em câncer de mama, buscamos validar a significância da alta expressão de *ELAVL1* sobre a recidiva tumoral através de dois métodos correção do *p-value*, tais como o método genérico de *false discovery ratio*<sup>142</sup> ou específico para busca de cortes<sup>143</sup>, e pela utilização de pseudoreplicação<sup>364</sup>. Em todas as estratégias utilizadas, observamos significância estatística na correlação entre alta expressão de *ELAVL1* e aumento do risco de recidiva em pacientes com câncer de mama, indicando a importância de se estudar o impacto de HuR sobre este tipo tumoral.



**Figura 67 – Relações observadas entre os pacientes, especificamente de câncer de mama, com alta expressão de *ELAVL1***

Outros trabalhos também avaliaram o nível total de expressão de *ELAVL1* frente o prognóstico do câncer de mama. Um destes estudos concordam com nossos achados<sup>112</sup>, enquanto que outros dois são opostos<sup>109,111</sup>. É importante ressaltar que a maior parte dos trabalhos onde a importância de HuR foi avaliada em câncer de mama, se ateve a informação sobre a localização núcleo/citoplasma da proteína, tendo sido frequentemente encontrado uma correlação entre localização citoplasmática e piora na sobrevida<sup>113–115,118</sup>. Assim sendo, demonstrar uma importância prognóstica irrespectivamente à localização da proteína, através da abundância do seu mRNA, aponta a importância da proteína correlacionando-se independentemente da localização – seja nuclear ou citoplasmática – com o prognóstico dos pacientes em determinadas origens tumorais – o que corrobora com funções nucleares ainda negligenciadas de HuR no contexto patológico.

No estudo da expressão global alta de *ELAVL1* e sua relação com a sobrevida dos pacientes, vale ressaltar que esta relação foi significativa sobre a melhora da sobrevida livre de recidiva e livre de metástase, mas não sobrevida global (dados não mostrados). Importante lembrar, entretanto, que a sobrevida global (usualmente mencionada na literatura como *overall survival*, ou ainda *all-cause survival*) tem a vantagem de apresentar um *ponto final* extremamente bem definido, sendo o falecimento uma definição absoluta<sup>365</sup>; entretanto, nesta análise, também são inclusos falecimentos não (*a priori*) relacionados câncer, tais como fatalidades acidentais diversas. Neste sentido, o estimador de recidiva tem a vantagem de mostrar a relação direta com a recorrência local ou distante (metástase) da doença; as desvantagens desta medida, entretanto, são a dificuldade de definição da não-ocorrência de recidiva (relacionada com a sensibilidade dos exames realizados) e a muitas vezes impossível distinção entre novos eventos tumorais independentes e recidivas de células da mesma origem tumoral<sup>366</sup>. Por fim, o parâmetro de sobrevida livre de metástase, por mais que capaz de indicar os eventos de mais severa agressividade dos tumores, ignora recidivas locais e recai na possível indistinção entre novos eventos tumorais e metástases originárias do primeiro tecido tumoral. Faz sentido, então, observar os três eventos simultaneamente e ponderar o prognóstico final em função dos mesmos. Assim sendo, no conjunto, e em luz do que está na literatura, propomos que pacientes de câncer de mama com elevada expressão de *ELAVL1* apresentam risco superior para recidivas (Figura 67).

## 5.2 IMPORTÂNCIA DE HUR NA REGULAÇÃO DA GLUTAMINASE

A atuação regulatória de HuR sobre *GLS* no contexto fisiológico está relacionada com a ligação da proteína em elementos responsivos à acidose<sup>57</sup>, localizados na 3'UTR da isoforma GAC humana e, curiosamente, na 3'UTR da isoforma KGA de rato<sup>25</sup>. Quando em situação de acidose metabólica, na qual o pH sistêmico se reduz para 6,9, HuR transloca-se do núcleo para o citoplasma e contrabalancearia os efeitos desestabilizadores dos elementos responsivos a acidose, aumentando a estabilidade do mRNA de GAC (em humanos) e seu consequente nível proteico<sup>57</sup> (Figura 6). Também regulando o destino do mRNA de *GLS*, o *knock out* do gene parálogo *ELAVL4* (HuD) em camundongos reduz os níveis de mRNA da isoforma KGA e aumenta os níveis de GAC no tecido cerebral do animal, com mecanismo provavelmente relacionado a *splicing* devido a ligação observada, dentre outros locais, no intron 14 do gene *GLS* murino<sup>367</sup>.

A glutamina, junto com a glicose, é uma importante fonte biossintética e energética em câncer<sup>14</sup>. O primeiro passo de processamento da glutamina é feito pela enzima glutaminase, gerada por dois genes distintos: *GLS* e *GLS2*. *GLS* codifica por *splicing* alternativo duas isoformas, KGA e GAC, e, de maneira genérica, tem sido associado ao processo tumorigênico, enquanto que as isoformas de *GLS2*, LGA e GAB, geradas por uso alternativo de promotor, têm sido associadas à supressão tumoral<sup>34</sup>. O metabolismo de glutamina, e a consequente dependência do gene *GLS*, são observadas em diversos tecidos tumorais<sup>59</sup>, tendo isto sido demonstrado tanto *in vitro* quanto *in vivo*<sup>368</sup>. De fato, inibição desta enzima pela molécula CB-839 está em estudos clínicos fase 1 e 2 para diversos tipos tumorais tais como hematológicos, renais, pulmonares, colorretais e tumores de mama triplo negativo<sup>17,369</sup>.

Apesar da associação genérica de *GLS* com a condição tumoral, há trabalhos que apontam a maior importância da isoforma GAC<sup>370</sup>, seja pela sua maior abundância observada (por exemplo, em tumores de mama<sup>26</sup>, em câncer colorretal<sup>371</sup>, do sistema nervoso central<sup>371</sup> e de outras origens topográficas<sup>372</sup>) ou por sua maior atividade glutaminolítica (quando comparada a isoforma KGA<sup>27</sup>). Em tumores de cabeça-e-pescoço, entretanto, a isoforma KGA se mostrou importante para um fenótipo agressivo celular, uma vez que seu silenciamento levou a uma redução de agressividade e proliferação *in vivo*<sup>373</sup>.

Os exemplos regulatórios já conhecidos de HuR sobre *GLS* se encaixam no contexto fisiológico normal de sua função e não há descrição na literatura da atuação de HuR sobre as isoformas de glutaminase em câncer e, principalmente, a importância da regulação das mesmas para o fenótipo tumoral, questões estas que motivaram este trabalho.

### 5.3 ANÁLISE PAN-CANCER DA IMPORTÂNCIA DE *GLS*

#### 5.3.1 A importância das isoformas KGA e GAC para o câncer

Dentro deste contexto da importância de HuR e de *GLS* para o câncer, começamos por revisar a importância do gene *GLS* e de suas isoformas para diversos tipos tumorais usando dados disponíveis do TCGA.

Em uma análise global, verificamos que, curiosamente, em apenas cinco dos tumores avaliados (ducto biliar, esôfago, fígado, cabeça e pescoço e estômago) a expressão de *GLS* está significativamente (*p-value* a 5% de significância) aumentada em relação a tecido normal (Figura 18 e Tabela 6). Comparando nossos achados com avaliações similares da literatura, Cluntun *et al*<sup>59</sup> também concluíram que o aumento de expressão em tecidos tumorais

frente aos seus pares normais se dá apenas nos tumores de cólon, esôfago, fígado, estômago, tireoide e cabeça-e-pescoço, sendo que em nossa avaliação aponta a não-significância de estômago, cólon e tireoide, além de incluir ducto biliar. Outros trabalhos também suportam o aumento em tumores de origens coloretal<sup>374</sup> e hepática<sup>375</sup>.

Entretanto, um ponto importante a ser levado em consideração com relação a este gene é a importância relativa das isoformas KGA e GAC. Como já mencionado, a isoforma GAC é a cataliticamente mais ativa e mais abundante em diversos tipos tumorais<sup>26,27,370-372,376</sup>, sendo identificada como alvo do composto 968, identificado pela inibição de fibroblastos transformados por Rho GTPase da família Dbl<sup>63</sup>. Xia *et al*<sup>372</sup> também evidenciaram a opção tumoral pela isoforma GAC em sete tumores do TCGA (pulmão, útero, bexiga, mama, cabeça-e-pescoço e rim), utilizando como justificativa o fato de seu 3'UTR ser mais curto e desprovido de sequências regulatórias encontradas no mRNA de KGA.

Nossa avaliação pan-câncer evidenciou que a expressão do mRNA de GAC é priorizada frente a de KGA em diversos tipos tumorais em relação ao tecido normal, confirmando uma tendência de escolha desta isoforma (Figura 21, Figura 66 e Tabela 8). A escolha preferencial de GAC não só condiz com o observado na literatura, como reforça a existência de mecanismos ainda não elucidados para a seleção – através de *splicing* alternativo ou estabilização diferencial – de seu mRNA. Quanto a expressão da isoforma KGA em relação ao tecido normal, a mesma se demonstra particularmente aumentada apenas em tumores de fígado, e reduzida ou com níveis mantidos na maior parte das demais origens topográficas.

### 5.3.2 Relação entre expressão de *ELAVL1* e níveis de mRNA de GAC e KGA em diferentes tumores

Avaliamos o nível de mRNA de GAC frente à KGA (via avaliação dos níveis de expressão do exon 15 e 19, respectivamente) em tumores de mama com alta expressão de *ELAVL1*. Nossa avaliação mostrou que em tumores com alta expressão de *ELAVL1* (em nível que impacta a sobrevida do paciente) houve maior expressão das duas isoformas, comparado a tumores com baixa expressão de *ELAVL1* (Figura 22 e Figura 67). Desta maneira, especulamos que HuR deve influir simultaneamente na escolha das isoformas por *splicing* alternativo e também na estabilidade de seus mRNAs.

Os dados aqui observados para câncer de mama indicam que altos níveis de expressão de *ELAVL1* correlacionam-se com a abundância tanto de KGA quanto de GAC, todavia não do gene *GLS* como um todo. Tal incoerência está relacionada com a mensuração

de outros transcritos presentes no GENCODE. Tais transcritos incluem introns retidos e inícios de transcrição alternativos (como na anotação da UCSC), que são menos abundantes no contexto tumoral e, assim, reduzem a abundância do mRNA total de *GLS*.

### 5.3.3 Análise evolutiva e funcional da evolução das isoformas e aparecimento de GAC

Nossos estudos evolutivos apontam para a possibilidade de que GAC tenha sido gerada por retrotransposição e mantida pela seleção natural durante a evolução dos metazoários, o que implica que a mesma tenha alguma peculiaridade (regulatória, bioquímica ou de interação) que justificasse a sua manutenção no genoma como unidade funcional e também dos processos de regulação de sua expressão, especialmente em região intrônica altamente variável<sup>286</sup>. Similarmente, a duplicação gênica de *GLS2* conferiu algum tipo de vantagem frente aos desafios ambientais que justificaram sua manutenção no genoma. Compreender a história evolutiva das glutaminases e, em específico, do exon de GAC (15), pode nos trazer sugestões das suas funções na homeostase e na patologia.

Em mamíferos a excreção de nitrogênio está intimamente relacionada com a regulação do pH sistêmico, mencionada na introdução e retomada aqui. Em situação de pH equilibrado, a glutamina, via *GLS2*, é degradada em glutamato e amônia, que pode se combinar a um próton e formar amônio no fígado, somando-se aos íons amônio ( $\text{NH}_4^+$ ) captados do sangue. O amônio, pela ação da carbamoilfosfato sintase (*CPSI*), se une a bicarbonato e adentra ao ciclo da ureia, que é produzida e despejada na circulação sanguínea para posterior excreção pelo rim<sup>58</sup>. Quando da acidose sistêmica, o ciclo da ureia se reduz, poupando bicarbonato, e os hepatócitos periportais se encarregam de sintetizar glutamina a partir do excesso de amônio circulante, secretando o aminoácido que será capturado nos rins. Nestes, ocorre a estabilização do mRNA da isoforma GAC (e aumento de seu nível proteico), levando a liberação de amônia ( $\text{NH}_3$ ) e sua excreção na urina com um próton, na forma de amônio<sup>377</sup>, promovendo o reestabelecimento do pH sanguíneo. Assim, a isoforma GAC se apresenta como a isoforma de escolha para a rápida eliminação destes prótons em excesso e reestabelecimento do equilíbrio do pH do sangue, levando a acidificação da urina. Desta maneira, não é de se espantar que trabalho recente tenha mostrado que, em pH 6.0, GAC é a isoforma mais ativa<sup>376</sup>.

Este particular papel de GAC na acidificação da urina mostra-se potencialmente relevante na observação de que GAC surgiu no período aproximado de 445 milhões de anos atrás<sup>378</sup>, no ancestral comum aos vertebrados (Figura 30 e Figura 68A). Correlacionando a evolução anatômica concomitante, é o passo em que o pronefro ancestral está especializando-

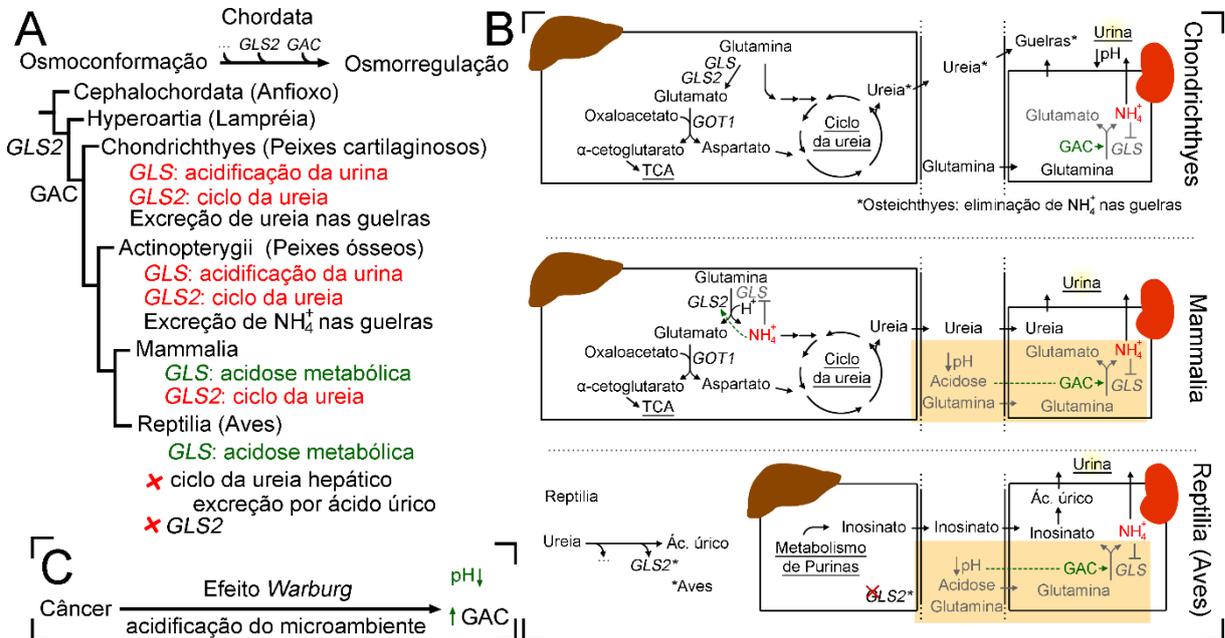
se com a mudança de animais ainda com mecanismos osmoconformes (como as lampreias) – que não regulam a pressão osmótica de forma distinta ao meio de forma completa – para animais osmorreguladores<sup>379</sup> – que mantém osmolaridade completamente distinta do meio. Para manter esta osmolaridade distinta, estes animais precisam constantemente eliminar sais na urina e reter água.

Observando a evolução dos distintos sistemas de excreção (Figura 68B), com o avanço para a osmorregulação, o aparecimento de GAC em peixes cartilagosos potencialmente se relaciona com a necessidade da manutenção do pH da urina reduzido, importante para evitar a precipitação dos sais excretados<sup>380</sup>. Portanto, a necessidade de efluxo de prótons superior nos rins pode ter sido a pressão de seleção necessária para manutenção de uma isoforma retrotransposta capaz de degradar glutamina de forma mais eficiente, gerando mais amônio e amenizando a precipitação de sais na urina (Figura 68B, superior). Na contraparte hepática, nestes organismos, *GLS2* ainda não se tornou imprescindível para o ciclo da ureia, dada a presença da carbamoilfosfato sintase III, capaz de utilizar diretamente a glutamina para a suplementação do ciclo.

Com a invasão do ambiente terrestre, a necessidade de acidificação da urina, para a não precipitação de sais, restringiu-se. O órgão renal assumiu parte da tarefa de regular o pH sistêmico, com a ativação de GAC na acidose metabólica, já mencionada na introdução, com a redução do ciclo da ureia frente ao pH reduzido no sangue<sup>381</sup>. Assim, mesmo com a mudança da pressão seletiva em si, permaneceu a demanda de secreção de amônio no rim, agora, com uma fina camada regulatória através da estabilização de mRNA – como a observada em humanos<sup>57</sup>. Do ponto de vista do parálogo duplicado, *GLS2*, sua liberação da pressão seletiva renal (quando da duplicação gênica) permitiu sua modificação da ativação por fosfato para a ativação por amônio<sup>37,39,53,54</sup> – que inibe as enzimas de *GLS*<sup>56</sup> –, em um *feedback* positivo, possivelmente relacionado com a necessidade do fornecimento ininterrupto de amônio<sup>53</sup> e aspartato para o ciclo da ureia ou a regulação do mesmo ciclo pelo provimento de glutamato para a síntese do regulador positivo N-acetilglutamato<sup>54</sup> (Figura 68B, centro).

Tal hipótese evolutiva é corroborada, ainda, pelo fato do gene *GLS2* não ser mais observado no genoma de aves<sup>32</sup> – que não realizam o ciclo da ureia<sup>36</sup> e excretam nitrogênio basicamente utilizando o ácido úrico (tal como em todo o grupo monofilético Reptilia), não diretamente relacionado com o metabolismo de glutamina<sup>382</sup> (Figura 68B, inferior). Em aves, a síntese de purina ocorre, no fígado, até a formação do inosinato, o qual é transportado para o rim, onde a enzima xantina oxidase converte o mesmo para ácido úrico e realiza a excreção propriamente dita<sup>383</sup>. Assim, com o afrouxamento da pressão seletiva sobre *GLS2*, seu gene

acabou sendo perdido em algum processo aleatório de remodelação do genoma – mantendo-se a importância de *GLS* para a regulação da acidose metabólica<sup>384</sup>.



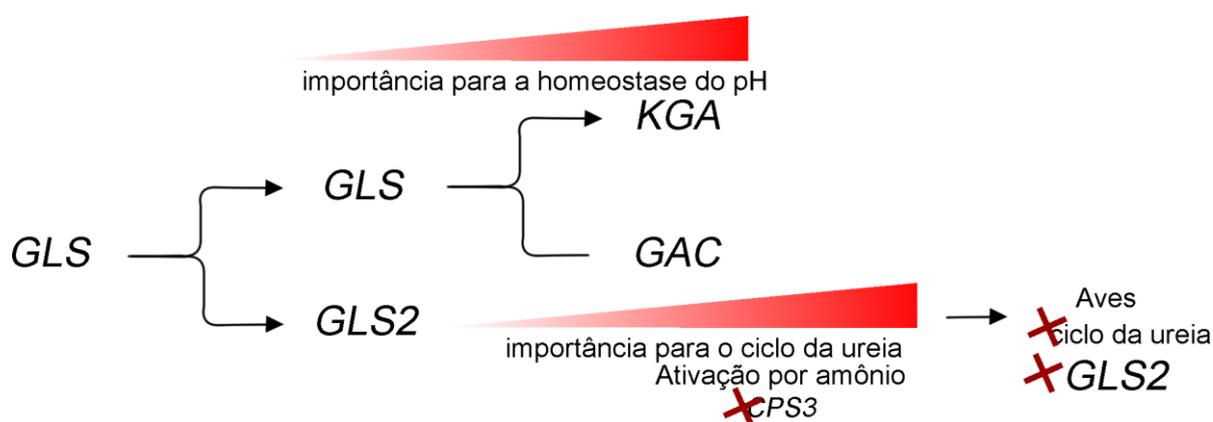
**Figura 68 – Esquema da proposta importância fisiológica das glutaminases em uma perspectiva evolutiva** (A) Eventos evolutivos relacionados com a função fisiológica das glutaminases e a estrita relação com o balanço de pH da urina ou sistêmico, (B) comparação da excreção em peixes e aves frente ao ciclo de mamíferos, este adaptado de Häussinger<sup>58</sup> (C) Relação entre a isoforma GAC e câncer, além da anaplerose do TCA, adaptada de Huang et al<sup>376</sup>

Por fim, a especialização funcional proposta a partir da análise filogenética feita acima pode auxiliar na compreensão dos resultados de expressão tecido tumoral das isoformas contra tecido normal. Tumores de fígado apresentam elevada expressão de *GLS* frente ao tecido normal; entretanto, em tecidos normais, a expressão do gene adaptado à necessidade de suprir o ciclo da ureia, *GLS2*, é aumentado em relação a tecidos tumorais. Logo, com a perda de funcionalidade e des-diferenciação que ocorre durante o processo tumorigênico, provavelmente não há mais a necessidade de alta expressão de *GLS2* para a manutenção do ciclo da ureia; neste contexto, *GLS*, que é ativada por fosfato, se tornaria mais eficaz para a proliferação do tumor. Curiosamente, em fígado, a isoforma priorizada é a KGA e não a GAC, por razões que podem envolver o domínio extra glutaminolítico C-terminal e alguma potencial função pró-tumoral ainda não elucidada.

Em rins, em contrapartida, tecidos normais expressam elevados níveis de *GLS* (priorizando a isoforma KGA) enquanto que, nos tecidos tumorais, os níveis de GAC não se alteram, mas ocorre uma grande redução de KGA – condizente com uma des-diferenciação celular imposta pela condição tumoral. O porquê de alguns tipos tumorais optarem por uma

isoforma ou outra precisaria ser validado em nível proteico mas coloca uma questão interessante sobre a importância relativa das mesmas. Como já colocado acima, Huang *et al*<sup>376</sup> propuseram que a escolha da isoforma GAC em situação de acidose sanguínea se daria pela maior atividade da mesma em situação de pH ácido em relação à KGA ou mesmo produtos de *GLS2*<sup>376</sup>. Desta maneira, uma hipótese interessante é a de que, dado que se sabe que o microambiente tumoral é frequentemente ácido devido a secreção de lactato de células em efeito Warburg<sup>376,385,386</sup>, considerando-se que esta acidez precisa ser combatida intracelularmente, GAC teria desempenho superior que as outras isoformas para a produção de amônia, o qual, na captura de prótons para a formação de amônio (e secreção do mesmo), poderia ajudar a balancear o pH celular (Figura 68C).

Em suma (Figura 69), com a duplicação gênica, *GLS* foi selecionada no contexto de regulação do pH (seja acidificando a urina de peixes para evitar a precipitação de sais ou atuando na economia de bicarbonato e secreção de prótons na acidose metabólica de vertebrados). *GLS2*, por sua vez, adquiriu a regulação positiva por amônio e estreitou relação com o ciclo da ureia, evidenciada com a ausência do gene em aves, que não realizam mais o tal ciclo.



**Figura 69 – Resumo da importância diferencial dos genes *GLS* e *GLS2* para a homeostase em uma perspectiva evolutiva**

#### 5.4 IMPORTÂNCIA DE HUR PARA A ESCOLHA DAS ISOFORMAS DE *GLS* EM MODELOS CELULARES DE CÂNCER

Dado que HuR pode impactar no metabolismo de RNA, não só em nível de RNA maduro (pela interação com 3'UTR e modulação da meia-vida do mRNA e/ou tradução), mas também sobre o *splicing* alternativo do pre-mRNA, avaliamos o impacto desta proteína nestes dois processamentos do transcrito do gene *GLS*.

#### 5.4.1 Regulação via estabilização de mRNA – interação com 3'UTR

HuR já foi mostrada estabilizar o mRNA de GAC pela interação com elemento *AU-rich* do 3' UTR de GAC humana (Figura 7) em situação de acidose metabólica, como já discutido<sup>25,57,74,387,388</sup>. Nossos estudos de imunoprecipitação de HuR mostraram que há interação de HuR com o 3'UTR do mRNA das isoformas humanas KGA e GAC (Figura 44). Em nossos estudos nas linhagens HEK293T, MDA-MB-231 e PC-3, não observamos um aumento da atividade de luciferase traduzida de mRNA contendo o 3' UTR de GAC quando HuR.WT é expressa; em contrário, observamos uma redução desta atividade nesta condição (Figura 39 e Figura 70). Tais ensaios na literatura são utilizados para mensurar a estabilidade<sup>389</sup> e o potencial de tradução<sup>390</sup> conferido por uma determinada sequência de RNA a um transcrito, dado que o experimento por si é incapaz de discernir os dois eventos<sup>391</sup>. Uma vez que o estudo de decaimento (após tratamento com actinomicina) a meia vida do mRNA endógeno de GAC não foi impactado pela expressão ectópica de *ELAVL1*, é possível que HuR atue reduzindo o processo de tradução do mRNA contendo de GAC, o que deveria ser confirmado com experimento de associação a poli-ribossomos, que mensura a atividade de síntese proteica para um determinado transcrito<sup>392</sup>.

Como também já discutido acima, os rins são conhecidos como responsável pelo balanço do pH, mantendo o tamponamento do sangue. A ativação desse sistema se dá quando células do túbulo renal são expostas cronicamente à um pH externo ( $pH_e$ ) próximo de 6,9<sup>393</sup>. Neste  $pH_e$  foi publicado que ocorre a estabilização do mRNA de GAC especificamente, e não de KGA<sup>25</sup>. Curiosamente, o microambiente tumoral, em decorrência do efeito Warburg apresenta um  $pH_e$  entre 6,9 e 7,0<sup>386</sup> e um  $pH_i$  (interno) próximo a 7,2<sup>385</sup>. Se HuR estabiliza o mRNA de GAC em condições de baixo pH, o mesmo não foi mimetizado em nossos estudos, os quais envolveram a manutenção das células em sistema tamponado a  $pH_e$  7,4. A realização de tais experimentos em pH mais ácidos poderia levar a resultados diferentes.

Além da questão do pH vale lembrar que os resultados obtidos com o mRNA de GAC, onde a expressão de HuR.WT diminuiu a estabilidade/tradução segundo avaliado com vetor repórter e não afetou estabilidade no estudo de decaimento, refletem a presença de HuR primariamente nuclear. Assim, é curioso a observação de que a expressão de HuR. $\Delta$ HNS, que apresenta maior localização citoplasmática, onde os processos de estabilidade acontecem, tenha tido um impacto ligeiramente positivo (mas não significativo), na atividade de luciferase cujo mRNA estava associado a 3'UTR de GAC. A expressão desta construção não foi avaliada para

a estabilidade do mRNA endógeno de GAC. Tal resultado indica que, para o destino do mRNA de GAC, a localização celular citoplasmática de HuR pode ser determinante. Além disto, é possível que a associação dos estudos de HuR. $\Delta$ HNS com pH mais ácido levem a uma melhor resposta no sentido de se detectar a estabilidade do mRNA de GAC dirigida por HuR. Assim, estudos da estabilidade do mRNA de GAC por HuR não estão completos e demandariam mais ensaios. De uma maneira geral, estes estudos poderiam elucidar porque, em cortes tumorais, há uma relação positiva entre a expressão de *ELAVL1* e os níveis de mRNA de GAC, situação que não recapitulamos em nossos estudos *in vitro* e em pH levemente básico.

Com relação ao mRNA da KGA humana, para o qual não existiria regulação por HuR em acidose renal<sup>25,57</sup>, o cenário observado foi coerente entre os experimentos de silenciamento e expressão de *ELAVL1*: verificamos que o primeiro levou a redução da atividade de luciferase dirigida por 3'UTR do mRNA de KGA enquanto que a expressão ectópica (da forma selvagem) levou ao aumento da mesma atividade, assim como da estabilidade do mRNA endógeno de KGA (Figura 39). Cabe ressaltar que, mesmo na ausência de predição de sítios *AU-Rich* na 3'UTR de KGA, os dados previamente existentes na literatura já indicavam que a ligação também ocorria (Figura 7), o que concorda com o observado aqui nos estudos de imunoprecipitação de HuR (Figura 44 e Figura 70).

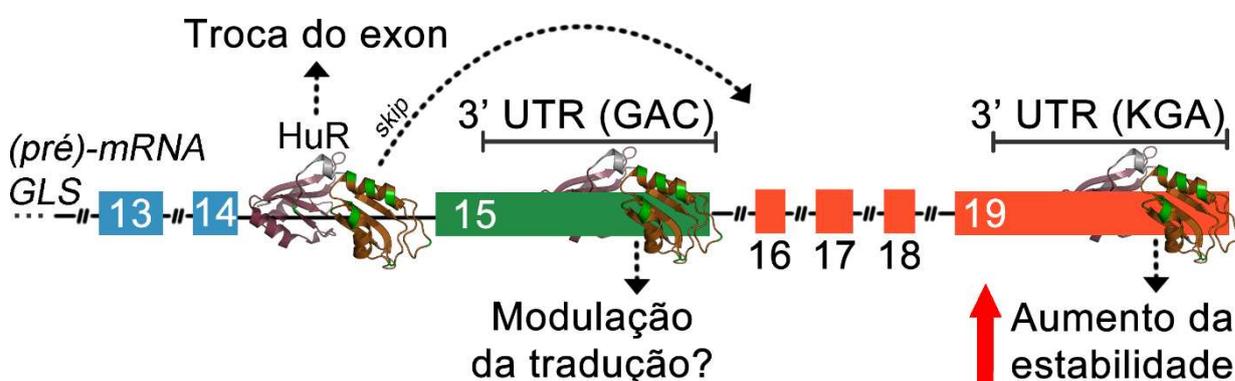


Figura 70 – Elementos regulatórios do (pré)-mRNA de *GLS* relacionados com HuR

#### 5.4.2 Regulação via *splicing* alternativo

A participação de HuR sobre o *splicing* alternativo de diversos genes tem sido cada vez mais conhecida<sup>394</sup>, tendo sido a proteína descrita como presente no complexo supra-spliceossomo<sup>95,395,396</sup>. Como já colocado, HuD, presente majoritariamente no tecido cerebral, regula o *splicing* e a tradução do mRNA de *GLS*<sup>367</sup> em camundongos, sendo que o seu *knock out* reduz a expressão de KGA. Em dois estudos<sup>96,97</sup> com avaliação ômica da ligação de HuR à

RNA, dos quais empregamos os dados analisados neste trabalho, mostrou-se uma relação de HuR com o *splicing* alternativo de diversos genes, não tendo sido *GLS* identificado nos mesmos.

Buscamos uma abordagem para avaliar a relação de HuR e o *splicing* alternativo, na qual cruzamos as informações obtidas da identificação de sua ligação a introns justapostos a exons passíveis de *splicing* alternativo (obtidos de dados de RIP-Seq), com informações de abundância de exons (obtidos de dados de RNA-Seq) diferencialmente expressos em situação de presença versus *knock down* de *ELAVL1* (Figura 23 e Figura 24). Tal estratégia levou a proposta de 175 alvos onde HuR potencialmente age sobre o *splicing* alternativo (Figura 28); entre estes genes, encontramos *GLS* (Figura 29 e Figura 70).

Cabe menção à publicação recente que mensurou a importância de HuR sobre *splicing* alternativo de maneira análoga, utilizando conjuntos de dados similares (utilizando a análise de PAR-CLIP-Seq de Lebedeva *et al*<sup>96</sup>, avaliando sua intersecção com a predição ab initio de sítios *AU-Rich* intrônicos), e obteve resultados compatíveis com os aqui apresentados<sup>263</sup>. Sua abordagem também observou a ligação e a presença de sítios *AU-rich* no intron 14 de glutaminase, mas não traz inferência específica da regulação do *splicing* do mesmo – apenas observação global da atuação de HuR sobre o *splicing* de seus possíveis alvos baseada em experimento de microarranjo (disponibilizado por Mukherjee *et al*<sup>97</sup>), validando experimentalmente alguns selecionados.

Avaliando apenas a presença de ligação a introns (irrespective se justapostos à exons relacionados com *splicing* alternativo) e utilizando o experimento de PAR-CLIP-Seq de Lebedeva *et al*<sup>96</sup> como fonte dos dados, este trabalho concluiu que 7071 genes são regulados por *splicing* alternativo guiado por HuR<sup>263</sup>, não realizando estudos de enriquecimento de vias para avaliar os processos alterados nas células. Nosso trabalho, em contrapartida, levou a identificação de 2418 genes (apenas na linhagem GM12878) em que HuR mostrou ligação a introns. Este menor número de alvos está provavelmente relacionado ao nosso modo de análise, onde aplicamos uma maior restringência na escolha dos introns, e priorizamos somente aqueles justapostos a exons trocados em ao menos um transcrito e, conseqüentemente, apresentam maior chance de relação com o *splicing* alternativo<sup>262,397</sup>.

Genes detectados em nossos estudos também foram observados neste trabalho, tais como *ABCD3*, *WHSC1*, *GART*, *ARMC8* (Figura 24) e *GLS*<sup>263</sup> (Figura 29 e Figura 70). Da lista de 175 genes aqui propostos como efetivamente regulados por HuR depois do cruzamento com os dados de silenciamento do gene, 157 foram também propostos por Bakheet *et al*<sup>263</sup>. Cabe mencionar, também, que nenhum dos 8 alvos confirmados experimentalmente neste trabalho (*POT1*, *PBLD*, *FKBP14*, *NR6A1*, *QARS*, *RPE*, *ELF2S2* e *PRAF2*) foram observados em nossa

lista de 175 genes, o que nos leva a crer que alguns dos filtros aplicados por nós em nossa metodologia tenha sido muito restritivo, levando a falsos-negativos. Entretanto, é plausível de se esperar que, sem o filtro de expressão gênica, os resultados de Bakheet *et al*<sup>263</sup> contemplem uma grande fração de falsos-positivos.

Encontramos que HuR potencialmente impacta na regulação do *splicing* alternativo de diversos alvos relacionados com o metabolismo de DNA e RNA. Não de maneira direta e evidente, mas a literatura já dá exemplos desta “regulação da regulação”, como a modulação de fatores de transcrição<sup>211</sup>, metilação de DNA<sup>398</sup> e de si própria, uma reguladora de mRNA<sup>101,399</sup>. A regulação de HuR sobre outras proteínas relacionadas ao processamento de mRNA, aqui observada, também é corroborada pelo previamente disponível na literatura<sup>96,103</sup>. Os exemplos utilizados para demonstrar a atuação tanto em inserção como excisão de exons (*ABCD3*, *WHSC1*, *GART*, *ARMC8* e, também, *GLS*), entretanto, não são alvos de nenhum estudo em específico.

Dentre os diversos alvos conhecidos como regulados por HuR por *splicing* alternativo<sup>90,98,400,401</sup>, identificamos o parálogo de eIF4E, EIF4E2<sup>92</sup>, e NF1<sup>402</sup>. Quanto ao gene *GLS*, diferentemente de nossa análise, a avaliação de Mukherjee *et al*<sup>97</sup> encontrou sítios de ligação de HuR em introns de *GLS*, entretanto seus dados de expressão (baseados em *microarray*) não mostraram uso diferencial dos exons exclusivos à KGA ou GAC. Já no caso de Lebedeva *et al*<sup>96</sup>, do qual oriundam os dados de RNA-Seq aqui utilizados, foi observado a ligação de HuR à introns não especificados de *GLS* no experimento de PAR-CLIP-Seq, mas não a abundância diferencial de expressão. Neste estudo, é importante mencionar que foi usada uma estratégia de mensuração distinta, baseada em quantificação de isoforma, tal como implementado por Tuxedo Suite<sup>168,403,404</sup>.

Por fim, a utilização dos dados de PAR-CLIP Seq para identificar os sítios de interação de HuR no intron 14 do gene *GLS*, quando cruzada com a informação acerca da mudança evolutiva da região, abordagem inédita utilizada por nós no estudo do *splicing* de *GLS* por HuR, permite localizar o possível elemento responsável a HuR (Figura 29). Tal identificação permite, por exemplo, estudos de mutação sítio-dirigida e confirmação da sequência regulatória proposta. Além disto, seria importante realizar o bloqueio da região com o uso de sequências ASO (*antisense oligonucleotides*) para provar sua importância para o processo de *splicing*, como já reportado anteriormente<sup>64</sup>.

Para validar experimentalmente o mecanismo de atuação de HuR sobre o *splicing* alternativo de *GLS* (Figura 70), empregamos um sistema repórter<sup>121</sup> adaptados por nós (Figura 40, Figura 41 e Figura 42). Os resultados confirmaram o envolvimento de HuR com *splicing*

de *GLS* mediado pelo intron 14: enquanto que o *knock down* de *ELAVL1* levou à inclusão do exon 15 e formação de proteína repórter equivalente a escolha de GAC, a super-expressão deste gene levou ao *skipping* do exon 15 e formação de proteína repórter equivalente a escolha de KGA. Concomitantemente, através de experimentos *in vitro*, mostramos a interação direta da proteína HuR expressa em sistema heterólogo procariótico e purificada com o intron 14 de *GLS* transcrito *in vitro* (Figura 45).

Observando o proposto sítio de ligação de HuR (Figura 29), localizado a 159 pares de base a montante da região do *branchpoint* (onde ocorre a ligação covalente entre o 5' do intron excisado, cujo reconhecimento por snRNP U2 é imprescindível para a concretização da excisão<sup>405</sup>) e a sua capacidade de interagir (com os dois primeiros domínios) com apenas 11 pares de base, a hipótese simples do impedimento físico de acesso aos elementos regulatórios chave do *splicing* não parece provável. A identificação de HuR como parte do *supraspliceossomo*, realizada por grupos independentes utilizando técnicas de espectrometria de massas<sup>94,395,396,406</sup>, não indica um parceiro específico de interação, mas a participação no conjunto completo de proteínas que compõe a estrutura responsável por realizar o *splicing* alternativo dos transcritos.

Uma proposta existente na literatura justifica a relação de HuR com o *skipping* de exons através da ligação a regiões distais dos mesmos e impedindo o reconhecimento pela snRNP U1 (*small nuclear ribonucleo protein U1*) da região 5' do intron subsequente, o que implica no não reconhecimento daquele exon e sua remoção do transcrito final<sup>90</sup>. Tal mecanismo não se aplica em nosso cenário dado o evento de regulação ser mediado pelo intron que antecede o exon e não por sequência dentro do próprio exon, como no modelo levantado. Também, o modelo proposto demanda de um bloqueio físico da ligação na região, o que não poderia ser transposto para o snRNP U2 (*small nuclear ribonucleo protein U2*, responsável pela ligação ao *branchpoint* intrônico), salvo ligação de HuR em regiões mais proximais ao sítio de reconhecimento de *splicing* não observadas nos experimentos avaliados. Assim, a caracterização mecânica de HuR sobre as diversas formas de atuação sobre o *splicing* alternativo de genes ainda permanece em aberto.

#### 5.4.3 Atuação de HuR na escolha das isoformas de *GLS* e fenótipo de linhagens tumorais de mama

Como comentado acima, detectamos experimentalmente a ligação de HuR tanto na 3'UTR de GAC quanto de KGA, assim como no intron 14 do gene *GLS*, indicando que HuR

regula o metabolismo do RNA desta enzima em diversos níveis. *In vitro*, verificamos um efeito negativo de HuR (pelo menos em sua versão completa, WT) sobre estabilidade/tradução/escolha em *splicing* ou qualquer outro evento relacionado ao mRNA de GAC, assim como o efeito positivo de HuR sobre KGA é corroborado pela observação de que o silenciamento de *ELAVLI* em diversas linhagens celulares de câncer de mama gera aumento dos níveis proteicos (e de mRNA) de GAC e diminuição dos mesmos níveis de KGA (Figura 31 e Figura 32). Com o objetivo de confirmar a alteração dos níveis de KGA na expressão e silenciamento de *ELAVLI*, utilizamos a técnica de CRISPR/Cas9<sup>120,122,296</sup> para inserir uma proteína fluorescente no C-terminal de KGA (Figura 43). Com esta ferramenta, confirmamos os achados de que HuR impacta no aumento dos níveis de KGA, sendo os mesmos reduzidos em seu silenciamento e aumentados em sua expressão ectópica.

HuR apresentou variados níveis de expressão (detectado por imunofluorescência e *western blot*) em linhagens de câncer de mama. Sua abundância proteica no estado basal destas linhagens se correlacionou fraca e positivamente com os níveis de KGA, mas não de GAC, confirmando que HuR impacta de maneira mais ubíqua em KGA, sendo a sua ação sobre GAC contexto dependente.

Confirmamos o observado em outros estudos acerca da localização nuclear de HuR em cultura celular<sup>85</sup>; cabendo ressaltar que HuR transita entre núcleo e citoplasma, sendo sua localização citoplasmática dependente de regulação através da fosforilação de sua sequência HNS<sup>407,408</sup>. HuR foi mostrada regular a formação de *actin stress fibers* pela estabilização do mRNA de *ACTB* (gene que codifica a  $\beta$ -actina)<sup>313</sup>. Quando avaliados simultaneamente a abundância de HuR e o fenótipo das *actin stress fibers* (propostas como importantes para a força mecânica de migração celular<sup>308</sup>), observamos que algumas linhagens com um maior conteúdo total HuR apresentam mais destas estruturas (Figura 48), não tendo sido obtido uma correlação positiva alta entretanto quando diversas linhagens foram avaliadas (Figura 49). Tal observação reforça que, apesar de HuR estar relacionada a este fenômeno, como já publicado<sup>313</sup>, não é o regulador máster do mesmo pelo menos neste tipo tumoral.

Em MDA-MB-231 e BT549 observamos uma correlação positiva célula a célula entre os níveis de HuR e a formação de *actin stress fibers*; em acordo, o silenciamento de *ELAVLI* reduziu os fenótipos invasivos e a própria formação destas estruturas nestas linhagens (Figura 49). Em Hs578t, que apresentou relação negativa entre HuR e a formação de *actin stress fibers*, o silenciamento de *ELAVLI* foi capaz de reduzir capacidade invasiva em experimento de *transwell*, o que sugere que HuR afeta este parâmetro de maneira independente a formação

das fibras (Figura 57). O efeito do silenciamento de *ELAVLI* sobre a formação de *actin stress fibers* não foi medido para esta linhagem celular.

Também observamos que algumas linhagens de câncer de mama respondem à remoção de glutamina do meio de cultura com a redução dos níveis totais de HuR e, em ao menos duas, BT549 e MDA-MB-468, foi observada uma translocação do núcleo para o citoplasma quando desta supressão (Figura 50). O efeito da depleção de aminoácidos sobre a localização de HuR já foi mostrado de forma similar, mas não especificamente relacionado apenas com glutamina<sup>409</sup>. Na tentativa de verificar se esta regulação por acesso a glutamina é limitada por questões espaciais de difusão e impacta na localização de HuR, estabelecemos um modelo tridimensional de esferóides de BT549. Desta maneira, observando se havia alteração da localização de HuR em função da distância da célula da borda da massa celular, esperaríamos mimetizar situação análoga a um tumor real que apresenta limitação de acesso à nutrientes por ausência de vascularização. Neste sentido, também esperaríamos mimetizar uma situação onde células em massa tumoral tem variação na localização núcleo/citoplasma de HuR (como já observamos preliminarmente de cortes de estudos de enxertos xenográficos – dados não mostrados) e podem assim, impactar diferentemente nas isoformas.

Nossos resultados mostraram que o modelo de esferóide não mimetizou uma massa tumoral autêntica, já que o posicionamento celular de HuR não variou em função da localização celular na esfera (Figura 51). Desta maneira, não pudemos testar a hipótese de que maior localização citoplasmática de HuR (combinado a depleção de nutrientes/oxigênio) poderia ser a situação ideal para se observar o aumento de GAC. Entretanto, ao longo do diâmetro analisado, evidenciamos uma correlação positiva entre os níveis de HuR e KGA.

Nas linhagens de câncer de mama aqui avaliadas, verificamos que o silenciamento de *ELAVLI* levou à significativa redução da proliferação tanto em um sistema bidimensional (Figura 52) como tridimensional (Figura 53). Este último modelo agrega diversas similaridades ao crescimento tumoral *in vivo*, tais como interação celular<sup>410</sup> e a já mencionada limitação da distribuição de nutrientes e gases<sup>411</sup>. Importante, o silenciamento também levou à menor localização nuclear de  $\beta$ -catenina, o que pode estar associada a redução da migração e invasão medidas *in vitro* (Figura 54). Tal observação acerca da via Wnt não se articula com alguns achados da literatura, nos quais HuR foi mostrada promover a localização citoplasmática de  $\beta$ -catenina<sup>412</sup>. Tal incongruência pode estar relacionada com a regulação da estabilidade de alguns transcritos por HuR/ $\beta$ -catenina em câncer de mama especificamente no contexto de hipóxia<sup>413</sup>. De forma coerente aos nossos achados em MDA-MB-231, outros trabalhos já mostram a inibição de HuR levando à diminuição da migração desta linhagem em experimentos de

*scratch-wound*<sup>304</sup> (Figura 55 e Figura 56) e invasão em câmara de boyden<sup>303,414</sup> (Figura 57). No intuito de entender se a diminuição nos níveis de KGA pudesse estar ligada a perda na capacidade proliferativa e invasiva das células decorrente do silenciamento de *ELAVL1*, expressamos esta proteína ectopicamente nas células (Figura 58). A reintrodução de KGA nas células não resgatou tais fenótipos celulares e acredita-se que os mesmos sejam ligados a regulação que HuR exerce sobre ciclinas do ciclo celular<sup>323</sup> ou alvos relacionados a migração e invasão<sup>104,106</sup>. Além disto, reforça a ideia de que a isoforma KGA não contribui para a agressividade de tumores de mama. O panorâma geral dos achados está resumido na Figura 71.

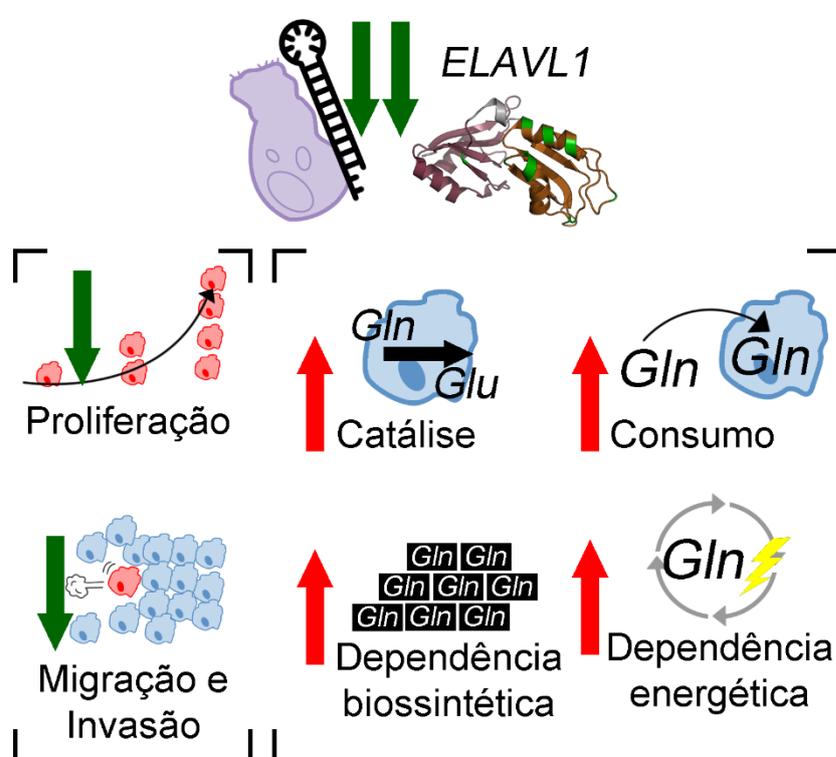


Figura 71 – Consequências fenotípicas do silenciamento de *ELAVL1* em células de câncer de mama

Como a resposta da manipulação de HuR sobre KGA se dá claramente no nível proteico em diversos modelos estudados, acreditamos que HuR seja um regulador primordial desta isoforma, o que se dá por *splicing* e aumento de estabilidade de mRNA. Entretanto, HuR tem o potencial estabilizar o mRNA de GAC no citoplasma em situações específicas de acidose, aspecto este que não foi completamente estudado neste trabalho. Assim, na avaliação de uma massa tumoral como um todo, em algumas células onde HuR teria maior presença citoplasmática, HuR poderia impactar positivamente no mRNA e proteína de GAC. Tal fenômeno poderia explicar porque observamos que em tumores de mama do TCGA maior expressão de *ELAVL1* está relacionado com aumento do mRNA de GAC e KGA.

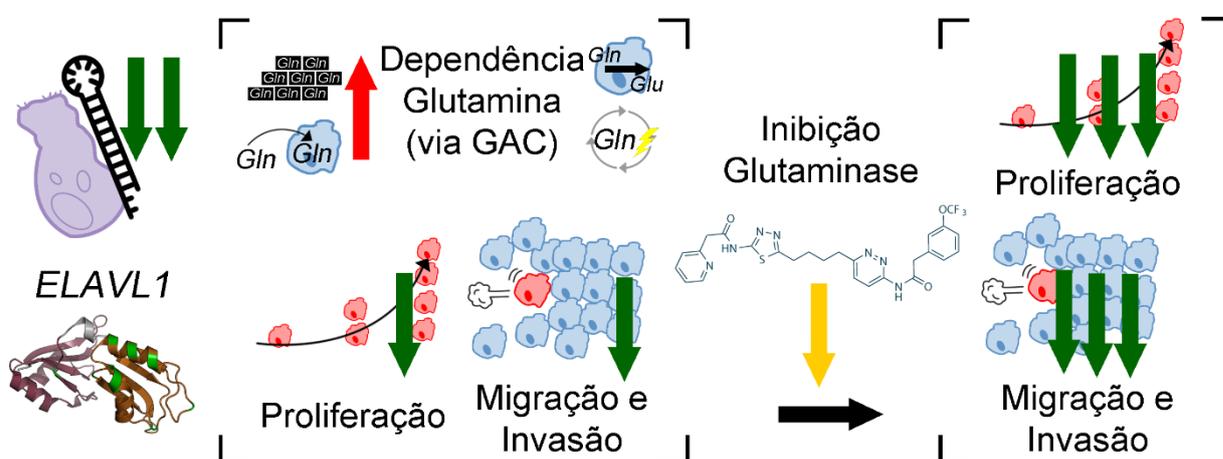
Enquanto que a diminuição de KGA não foi ligada as alterações na capacidade migratória das células, acreditamos que o aumento de GAC tenha sido o elemento envolvido com o ganho de capacidade glutaminolítica (medido pelo aumento de consumo de glutamina e atividade de glutaminase) e dependência de glutamina para biossíntese (medido por proliferação) e energética celular (inferido das medições de respiração) (Figura 59, Figura 60 e Figura 71). Já foi aqui discutido que a isoforma GAC é mais ativa que a KGA<sup>26</sup>, o que, acredita-se, estar relacionado a um C-terminal mais curto e desestruturado propiciando a formação de fibras supra-tetraméricas mais ativas<sup>27</sup>. Neste sentido, a supressão de HuR levou a um aumento da dependência metabólica celular à glutamina, dado este inédito na literatura. Desta maneira, a inibição química de GAC utilizando a molécula CB-839 afetou mais profundamente a respiração celular, proliferação, migração e invasão de células BT549 silenciadas para *ELAVL1*. Uma contra-prova importante não-executada para nossa argumentação seria o *knock down* de GAC combinado com o *knock down* de *ELAVL1* e verificação do retorno desta sensibilidade para o nível do controle.

Nossas análises de GC-MS de células com *knock down* de *ELAVL1* (Figura 61) mostraram que os níveis de glutamina e glutamato intracelular não se alteraram mesmo com a intensificação do consumo do primeiro. Tal resultado pode refletir que HuR comanda de maneira pleiotrópica o metabolismo deste aminoácido na célula, e, além de definir maior consumo do meio externo (observado experimentalmente) também comanda maior vazão metabólica do mesmo (coerente com a expressão da isoforma mais ativa) – mantendo os estoques intracelulares. Além disto, observamos um acúmulo de aspartato e de metionina nas células com *knock down* de *ELAVL1*, acúmulo para os quais desconhecemos o mecanismo envolvido assim como potenciais implicações. Apesar disto, é curioso que, no conjunto dos metabólitos identificados assim como de *features* para as quais não foi possível a identificação molecular, células com níveis normais de HuR ou com supressão do mesmo apresentam perfil metabólico distinto, implicando que HuR impacta no metabolismo celular de maneira significativa, o que precisaria ser melhor descrito. A identificação de tais *features* demanda da utilização de bibliotecas de fragmentação maiores e mais adaptadas aos subprodutos do metabolismo humano, e está em andamento.

Diversos inibidores para as glutaminases já foram e estão sendo propostos<sup>17,63,368,415</sup> e HuR tem se mostrado um crescente alvo para a inibição química igualmente<sup>304,416</sup>, já sendo proposta a inibição conjunta de HuR com outros alvos sensibilizados após sua inativação<sup>355</sup>. Uma vez que células de câncer de mama com supressão de HuR se tornaram mais dependentes a glutamina, resolvemos explorar esta dependência combinando o *knock down* de *ELAVL1* com

a inibição química de glutaminase (Figura 72). Como já colocado acima, nossos resultados mostraram que a combinação do *knock down* de *ELAVL1* com a inibição de glutaminase por CB-839 foram mais efetivos na redução da proliferação (levando a morte celular), migração e invasão das células comparado ao uso destes elementos isoladamente (Figura 62 e Figura 65). Estudos pré-clínicos *in vivo* combinando o *knock down* de *ELAVL1* com inibição de glutaminase estão em andamento em colaboração com a Dra. Maria Cristina R Rangel, no laboratório do Dr. Roger Chammas do ICESP/SP.

Dezenas de estratégias químicas já foram propostas para a inibição de HuR no tratamento de tumores<sup>338</sup>, desde moléculas amplamente inespecíficas, como quercetina<sup>417</sup>, até moléculas avaliadas como inibidores específicos através de ensaios de *high throughput screening*<sup>416,418,419</sup>, contemplando inibidores com ação anti-proliferativa em tumores de mama<sup>304</sup>. Apesar disto, até a data não há estudo clínico para nenhuma das moléculas registrado em *ClinicalTrials.gov*. Com relação a glutaminase, diversos estudos clínicos estão em andamento para aplicações do composto CB-839, em monoterapia ou associação, para diversos origens tumorais, dentre elas mama, pulmão, rim, entre outras<sup>369</sup>. Especificamente, o crescente número de estudos clínicos com associações de CB-839 aponta que sua aplicação como monoterapia não se sustenta na maior parte dos tipos tumorais.



**Figura 72 – Estratégia terapêutica proposta de acordo com os resultados obtidos, com o simultâneo silenciamento de *ELAVL1* e inibição química de glutaminase.**

Para dar força à nossa proposta terapêutica, os achados *in vitro* precisam ser confirmados *in vivo*. A próxima lacuna a ser respondida seria de quais subtipos tumorais de câncer de mama melhor responderiam ao tratamento proposto. Dada a elevada expressão de *ELAVL1* ocorrer no subtipo basal-like e triplo-negativo para os marcadores clínicos (ausentes da expressão dos receptores de estrógeno e progesterona e sem amplificação do receptor do tipo

2 de fator de crescimento epidérmico<sup>244</sup>), tal fração dos pacientes se mostra como uma sugestão inicial de tumores que responderiam ao tratamento (Figura 16). Cabe salientar que a célula utilizada para a avaliação da hipótese de inibição combinada, BT549, pertence aos subtipos triplo-negativo e basal-like. Nosso laboratório já tem interesse neste modelo tumoral por serem tumores agressivos e de pior prognóstico e nos quais já foi demonstrada a dependência por glutamina e sensibilidade a inibição da glutaminase<sup>17</sup>. Entretanto, dados do nosso grupo (não publicados) e de outros grupos<sup>17</sup> têm mostrado uma grande variação da sensibilidade de linhagens deste subtipo tumoral à inibição química de *GLS*, o que nos sugere heterogeneidade à resposta nos pacientes tratados com CB-839. Assim, uma proposta seria que a terapia combinada de CB-839 com inibição de HuR poderia ser mais eficaz em tumores com menor dependência de glutamina. Com o silenciamento de *ELAVL1*, tais tumores se tornariam mais dependentes ao aminoácido e, conseqüentemente, mais sensíveis à inibição química de *GLS*. Se esta abordagem seria mais efetiva em tumores triplo-negativos de sensibilidade reduzida a glutaminase ou poderia mesmo ser efetiva em tumores não-triplo negativos (tornando-os dependentes deste aminoácido) é uma questão em aberto.

## 6 CONCLUSÃO

- Utilizando dados de transcriptômica disponíveis publicamente pelo TCGA, estabelecemos um panorama de expressão de *ELAVL1* e das isoformas KGA e GAC em diferentes tipos tumorais. Verificamos que a expressão de *ELAVL1* está aumentada em diferentes tumores, incluindo mama; para muitos tumores a alta expressão de *ELAVL1* se correlaciona com pior sobrevida livre de recidiva (e livre de metástase, no caso de mama), enquanto que para outros se correlaciona com melhor sobrevida livre de recidiva. De maneira inédita, nosso estudo é o primeiro que realiza uma análise pan-câncer da relação do nível total de HuR sobre o impacto na sobrevida dos pacientes, expandindo as observações existentes, até então focada em um ou outro tumor e priorizando a informação de localização da proteína.
- Com os estudos de transcriptômica, observamos também uma maior escolha de GAC sobre KGA (medida pela razão de expressão dos seus exons exclusivos) em diferentes tipos tumorais, incluindo mama. Apesar da observação da preferência de diversos tumores por GAC estar exemplificada em alguns trabalhos da literatura, este é o primeiro onde esta relação é apresentada de maneira geral em diversos tumores e confirma a predileção da grande maioria deles pela GAC, pelo menos em nível transcricional.
- Em tumores de mama do TCGA, a expressão alta de *ELAVL1* se correlacionou com maiores níveis de GAC e KGA. Também, tumores do subtipo triplo-negativo e basal-like mostram maior abundância de *ELAVL1*.
- Acrescentamos evidências à atuação de HuR como proteína com importante papel na regulação do *splicing* alternativo, propomos possíveis 175 alvos em que HuR liga-se a algum intron relacionado ao *splicing* alternativo e ocorre troca das isoformas expressas quando do silenciamento de *ELAVL1*.
- Observamos também uma regulação específica de introns observada espacialmente, sendo a ligação de HuR não correlacionada apenas com a abundância de introns em uma determinada região genômica, todavia, sim, aos elementos regulatórios presentes.
- Os propostos genes cujo *splicing* é regulado por HuR enriquecem vias relacionadas com o metabolismo geral da célula e, mais especificamente, o metabolismo de RNA e DNA, indicando que HuR tem um papel preponderante em regular outras proteínas que, como ela, atuam no controle fino da expressão gênica.
- Propomos que o mecanismo regulatório do *splicing* de *GLS* por HuR não consiste na simples oclusão física do sítio de *branchpoint* utilizado pelo spliceossomo. A região

identificada, utilizando conservação de sequência e experimentos ômicos de imunoprecipitação de RNA, não é sobreposta aos elementos mínimos reconhecidos pelo spliceossomo.

- Conseguimos estabelecer ferramentas moleculares capazes de suprimir ou aumentar a expressão de *ELAVL1* em linhagens celulares, utilizando sistemas transientes, constitutivos e induzíveis.
- Estabelecemos também sistemas repórter para a estabilidade mediada pela 3'UTR das isoformas KGA e GAC de *GLS*, além de minigene capaz de indicar por fluorescência a escolha entre as duas isoformas durante o *splicing* alternativo. Em conjunto, tais ferramentas são capazes de discernir efeitos de *splicing* ou de estabilidade/tradução sobre o destino do mRNA.
- Observamos a expressão proteica de HuR em um painel de células de câncer de mama, e verificamos uma correlação positiva da abundância da mesma com os níveis proteicos da isoforma KGA, mas não de GAC. Os níveis totais de HuR se correlacionaram com aumento de formação de *actin stress fibers*.
- O gene *ELAVL1* foi introduzido ectopicamente ou silenciado em linhagens de tumor de mama, com a presença de HuR impactando sempre positivamente pela escolha de KGA e, em muitos casos, sua ausência levando ao aumento de GAC. O mesmo foi observado em linhagens de tumor de próstata.
- Mostramos por RIP-RT-PCR que HuR se associa a 3'UTR do mRNA de GAC, KGA e ao intron 14 do gene *GLS*, diretamente ligado à escolha entre as isoformas durante o processo de *splicing*. Para o intron 14 de *GLS*, também validamos a interação *in vitro* utilizando técnica de FRAP.
- Utilizamos ensaios celulares com genes repórteres que mostraram que HuR.WT aumenta a estabilidade do mRNA de KGA (e potencialmente diminui tradução do mRNA de GAC) e propicia a escolha do mRNA de KGA sobre GAC durante o processo de *splicing*. Não obtivemos conclusões assertivas acerca do efeito do 3'UTR de GAC, sugerindo apenas que HuR atua de alguma forma nesta regulação, possivelmente envolvendo modulação de taxas de tradução.
- Estabelecemos, utilizando a tecnologia CRISPR/Cas9, célula com introdução de cauda fluorescente no C-terminal da proteína KGA endógena, capaz de prontamente reportar os níveis de abundância da mesma quando de maneira independente de imunomarcção. Tal linhagem celular respondeu como esperado segundo a literatura acerca da regulação de KGA durante o ciclo celular.
- Empregando a linhagem construída como sensora dos níveis de KGA, confirmamos que a expressão de HuR aumenta os níveis de KGA e a sua supressão diminui os níveis da isoforma.

- De maneira geral, HuR se localiza preferencialmente no núcleo (o que é mais frequentemente observado em cultura de células) onde influencia a escolha do mRNA de KGA sobre o de GAC.
- Ressaltamos, também, que em linhagens onde a presença de KGA esta nivelada com a de GAC, observamos claramente a troca das isoformas com a manipulação de HuR e ação de HuR sobre *GLS* deve-se dar primordialmente em nível de *splicing* alternativo do gene. Em outras onde KGA já se mostra escasso frente a GAC, a manipulação de HuR surtiu efeitos diversos sobre o nível proteico e de mRNA de GAC. Nestas, as quais já foram mostradas ser mais dependentes de glutamina (triplo-negativas de alto conteúdo de GAC), acreditamos que outros mecanismos são acionados para garantir a máxima presença de GAC, havendo pouca janela para a ação de HuR.
- Realizamos a reconstrução da história evolutiva da isoforma GAC, tendo concluído que GAC surgiu em evento de retrotransposição. Observando a evolução do sistema excretor dos animais e o aparecimento das isoformas de *GLS*, assim como do seu parálogo *GLS2*, concluímos que os dois genes apresentam funções fisiológicas distintas relacionadas ao metabolismo sistêmico do nitrogênio. Em específico, com especialização permitida pela duplicação de *GLS*, *GLS2*, que é mais expressa em fígado, tem papel mais importante para o ciclo da ureia. Concomitantemente ao processo de seleção funcional, ocorreu uma adaptação do mecanismo de regulação das isoformas de *GLS*, permitindo a expressão da cataliticamente mais ativa (GAC) em um contexto metabólico adequado, mediante a utilização dos elementos regulatórios retrotranspostos e modificados. O elevado grau de especialização tecidual gerado por esta adaptação evolutiva fica claro com a redução dos níveis de *GLS2* em tumores hepáticos e de *GLS* em tumores renais, dada a desdiferenciação celular intimamente ligada ao processo tumoral.
- O *knock down* de *ELAVL1* levou a diminuição da proliferação (bi- e tridimensional), migração e invasão de linhagens de tumor de mama, assim como aumento de dependência metabólica de glutamina (em célula menos dependente relacionada a célula mais dependente).
- O resgate da expressão de KGA não restaurou a capacidade proliferativa e migratória das células, indicando que a troca das isoformas de *GLS* não é o principal efetor deste fenótipo mediado por HuR.
- Mostramos que o *knock down* de *ELAVL1* pode apresentar sinergia com a inibição química de glutaminase no sentido de diminuir a proliferação, invasão e migração da linhagem de tumor de mama BT549 *in vitro*. Propomos que a terapia combinada de inibição destas duas

proteínas tem o potencial de ter efeito terapêutico aumentado frente a inibição isolada das mesmas no tratamento de tumores de mama.

## REFERÊNCIAS

1. Weinberg, R. A. *The biology of cancer*. (Garland Sciences, 2014).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
3. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
4. Cairns, R. A. & Mak, T. W. The current state of cancer metabolism. *Nature Reviews Cancer* **16**, 613–614 (2016).
5. Warburg, O. On the Origin of Cancer Cells. *Science (80-. )*. **123**, 309–314 (1956).
6. Otto, A. M. Warburg effect(s)—a biographical sketch of Otto Warburg and his impacts on tumor metabolism. *Cancer Metab.* **4**, 5 (2016).
7. Garret, R. G. & Grisham, C. M. *Biochemistry*. (Cengage Learning, 2017).
8. Heiden, M. G. Vander, Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science (80-. )*. **324**, 1029–1033 (2009).
9. Vizán, P., Mazurek, S. & Cascante, M. Robust metabolic adaptation underlying tumor progression. *Metabolomics* **4**, 1–12 (2008).
10. Tannock, I. F. & Rotin, D. Acid pH in tumors and its potential for therapeutic exploitation. *Cancer Res.* **49**, 4373–84 (1989).
11. Altman, B. J., Stine, Z. E. & Dang, C. V. From Krebs to clinic: Glutamine metabolism to cancer therapy. *Nature Reviews Cancer* **16**, 619–634 (2016).
12. Sellers, K. *et al.* Pyruvate carboxylase is critical for non – small-cell lung cancer proliferation. *J. Clin. Invest.* **125**, 687–698 (2015).
13. Liu, W. *et al.* Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. *Proc. Natl. Acad. Sci.* **109**, 8983–8988 (2012).
14. Gao, P. *et al.* c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458**, 762–5 (2009).
15. Cheng, T. *et al.* Pyruvate carboxylase is required for glutamine-independent growth of tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8674–9 (2011).
16. Jiang, L. *et al.* Reductive carboxylation supports redox homeostasis during anchorage-independent growth. *Nature* **532**, 255–258 (2016).
17. Gross, M. I. *et al.* Antitumor Activity of the Glutaminase Inhibitor CB-839 in Triple-Negative Breast Cancer. *Mol. Cancer Ther.* **13**, 890–901 (2014).
18. Williamson, D. H. & Brosnan, J. T. *Methods of Enzymatic Analysis*. (Academic Press, 1974).
19. Gameiro, P. A. *et al.* In vivo HIF-mediated reductive carboxylation is regulated by citrate levels and sensitizes VHL-deficient cells to glutamine deprivation. *Cell Metab.* **17**, 372–385 (2013).
20. Golias, T. *et al.* Hypoxic repression of pyruvate dehydrogenase activity is necessary for metabolic reprogramming and growth of model tumours. *Sci. Rep.* **6**, 1–11 (2016).
21. Mullen, A. R. *et al.* Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature* **481**, 385–388 (2012).
22. Yang, C. *et al.* Glutamine Oxidation Maintains the TCA Cycle and Cell Survival during Impaired Mitochondrial Pyruvate Transport. *Mol. Cell* 1–11 (2014).

- doi:10.1016/j.molcel.2014.09.025
23. Murai, S. *et al.* Inhibition of malic enzyme 1 disrupts cellular metabolism and leads to vulnerability in cancer cells in glucose-restricted conditions. *Oncogenesis* **6**, 1–9 (2017).
  24. Pérez-Gómez, C. *et al.* Genomic organization and transcriptional analysis of the human l-glutaminase gene. *Biochem. J.* **370**, 771–784 (2003).
  25. Porter, L. D., Ibrahim, H., Taylor, L. & Curthoys, N. P. Complexity and species variation of the kidney-type glutaminase gene. *Physiol. Genomics* **9**, 157–66 (2002).
  26. Cassago, A. *et al.* Mitochondrial localization and structure-based phosphate activation mechanism of Glutaminase C with implications for cancer metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1092–7 (2012).
  27. Ferreira, A. P. S. A. P. S. *et al.* Active glutaminase C self-assembles into a supratetrameric oligomer that can be disrupted by an allosteric inhibitor. *J. Biol. Chem.* **288**, 28009–20 (2013).
  28. Olsen, B. R., Torgner, I. A., Christensen, T. B. & Kvamme, E. Ultrastructure of pig renal glutaminase. Evidence for conformational changes during polymer formation. *J. Mol. Biol.* **74**, 239–51 (1973).
  29. Olsen, B. R., Svenneby, G., Kvamme, E., Tveit, B. & Eskeland, T. Formation and ultrastructure of enzymically active polymers of pig renal glutaminase. *J. Mol. Biol.* **52**, 239–45 (1970).
  30. Elgadi, K. M., Meguid, R. a, Qian, M., Souba, W. W. & Abcouwer, S. F. Cloning and analysis of unique human glutaminase isoforms generated by tissue-specific alternative splicing. *Physiol. Genomics* **1**, 51–62 (1999).
  31. Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. & Ben-Tal, N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* **18 Suppl 1**, S71-7 (2002).
  32. Pasquali, C. C. *et al.* The origin and evolution of human glutaminases and their atypical C-terminal ankyrin repeats. *J. Biol. Chem.* **292**, (2017).
  33. Lewit-Bentley, A. & Réty, S. EF-hand calcium-binding proteins. *Curr. Opin. Struct. Biol.* **10**, 637–643 (2000).
  34. Martín-Rufián, M. *et al.* Mammalian glutaminase Gls2 gene encodes two functional alternative transcripts by a surrogate promoter usage mechanism. *PLoS One* **7**, (2012).
  35. Olalla, L. *et al.* Nuclear localization of L-type glutaminase in mammalian brain. *J. Biol. Chem.* **277**, 38939–38944 (2002).
  36. Watford, M. Hepatic glutaminase expression: relationship to kidney-type glutaminase and to the urea cycle. *FASEB J.* **7**, 1468–74 (1993).
  37. Verhoeven, A. J., van Iwaarden, J. F., Joseph, S. K. & Meijer, A. J. Control of rat-liver glutaminase by ammonia and pH. *Eur. J. Biochem.* **133**, 241–244 (1983).
  38. McGivan, J. D. & Bradford, N. M. Characteristics of the activation of glutaminase by ammonia in sonicated rat liver mitochondria. *Biochim. Biophys. Acta* **759**, 296–302 (1983).
  39. Campos-Sandoval, J. a. *et al.* Expression of functional human glutaminase in baculovirus system: Affinity purification, kinetic and molecular characterization. *Int. J. Biochem. Cell Biol.* **39**, 765–773 (2007).
  40. Gómez-Fabre, P. M. *et al.* Molecular cloning, sequencing and expression studies of the human breast cancer cell glutaminase. *Biochem. J.* **345 Pt 2**, 365–75 (2000).
  41. Greenberg, D. M., Blumenthal, G. & Ramadan, M.-E. A. Effect of administration of the enzyme glutaminase on the growth of cancer cells. *Cancer Res.* **24**, 957–63 (1964).
  42. Knox, W. E., Horowitz, M. L. & Friedell, G. H. The Proportionality of Glutaminase Content to Growth Rate and Morphology of Rat Neoplasms. *Cancer Res.* **29**, 669–680

- (1969).
43. Lee, S. Y. *et al.* Dlx-2 and glutaminase upregulate epithelial-mesenchymal transition and glycolytic switch. *Oncotarget* **7**, 7925–39 (2016).
  44. Alderton, G. K. Epithelial to mesenchymal and back again. *Nat. Rev. Cancer* **13**, (2013).
  45. Dion, H. W., Fusari, S. A., Jakubowski, Z. L., Zora, J. G. & Bartz, Q. R. 6-Diazo-5-oxo-L-norleucine, A New Tumor-inhibitory Substance. II.1 Isolation and Characterization. *J. Am. Chem. Soc.* **78**, 3075–3077 (1956).
  46. Elgogary, A. *et al.* Combination therapy with BPTES nanoparticles and metformin targets the metabolic heterogeneity of pancreatic cancer. *Proc. Natl. Acad. Sci.* **113**, E5328–E5336 (2016).
  47. Emadi, A. *et al.* Inhibition of glutaminase selectively suppresses the growth of primary acute myeloid leukemia cells with IDH mutations. *Exp. Hematol.* **42**, 247–251 (2014).
  48. Lee, J. S. *et al.* Glutaminase 1 inhibition reduces thymidine synthesis in NSCLC. *Biochem. Biophys. Res. Commun.* **477**, 374–382 (2016).
  49. Suzuki, S. *et al.* Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. *Proc. Natl. Acad. Sci.* **107**, 7461–7466 (2010).
  50. Hu, W. *et al.* Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7455–7460 (2010).
  51. Zhang, C. *et al.* Glutaminase 2 is a novel negative regulator of small GTPase Rac1 and mediates p53 function in suppressing metastasis. *Elife* **5**, 1–20 (2016).
  52. Lee, Y.-Z. *et al.* Discovery of selective inhibitors of Glutaminase-2, which inhibit mTORC1, activate autophagy and inhibit proliferation in cancer cells. *Oncotarget* **5**, 6087–101 (2014).
  53. Meijer, A. J. Channeling of ammonia from glutaminase to carbamoyl-phosphate synthetase in liver mitochondria. *FEBS Lett.* **191**, 249–51 (1985).
  54. Brosnan, J. T. & Brosnan, M. E. Hepatic glutaminase--a special role in urea synthesis? *Nutrition* **18**, 455–7 (2002).
  55. Meijer, A. J., Lamers, W. H. & Chamuleau, R. A. Nitrogen metabolism and ornithine cycle function. *Physiol. Rev.* **70**, 701–48 (1990).
  56. Kvamme, E. & Lenda, K. Regulation of glutaminase by exogenous glutamate, ammonia and 2-oxoglutarate in synaptosomal enriched preparation from rat brain. *Neurochem. Res.* **7**, 667–78 (1982).
  57. Ibrahim, H., Lee, Y. J. & Curthoys, N. P. Renal response to metabolic acidosis: role of mRNA stabilization. *Kidney Int.* **73**, 11–8 (2008).
  58. Häussinger, D. Ammonia, urea production and pH regulation. in *Textbook of Hepatology* (eds. Rodés, J., Benhamou, J.-P., Blei, A., Reichen, J. & Rizzetto, M.) **75**, 181–192 (GastroHep, 2004).
  59. Cluntun, A. A., Lukey, M. J., Cerione, R. A. & Locasale, J. W. Glutamine Metabolism in Cancer: Understanding the Heterogeneity. *Trends in Cancer* **3**, 169–180 (2017).
  60. Lukey, M. J., Greene, K. S., Erickson, J. W., Wilson, K. F. & Cerione, R. A. The oncogenic transcription factor c-Jun regulates glutaminase expression and sensitizes cells to glutaminase-targeted therapy. *Nat. Commun.* **7**, 11321 (2016).
  61. Wise, D. R. *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18782–7 (2008).
  62. Javeshghani, S. *et al.* Carbon source and myc expression influence the antiproliferative actions of metformin. *Cancer Res.* **72**, 6257–6267 (2012).
  63. Wang, J.-B. *et al.* Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. *Cancer Cell* **18**, 207–19 (2010).

64. Redis, R. S. *et al.* Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2. *Mol. Cell* **61**, 520–534 (2016).
65. Colombo, S. L. *et al.* Molecular basis for the differential use of glucose and glutamine in cell proliferation as revealed by synchronized HeLa cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 21069–74 (2011).
66. Colombo, S. L. *et al.* Anaphase-promoting complex/cyclosome-Cdh1 coordinates glycolysis and glutaminolysis with transition to S phase in human T lymphocytes. *Proc. Natl. Acad. Sci.* **107**, 18868–18873 (2010).
67. McDonald, C. J., Acheff, E., Kennedy, R., Taylor, L. & Curthoys, N. P. Effect of lysine to alanine mutations on the phosphate activation and BPTES inhibition of glutaminase. *Neurochem. Int.* (2014). doi:10.1016/j.neuint.2014.12.003
68. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biol.* **3**, 1–10 (2002).
69. Chen, C. A., Xu, N. & Shyu, A. Highly selective actions of HuR in antagonizing AU-rich element-mediated mRNA destabilization. *Mol. Cell. Biol.* **22**, 7268–78 (2002).
70. Tong, J., Harrison, G. & Curthoys, N. P. The effect of metabolic acidosis on the synthesis and turnover of rat renal phosphate-dependent glutaminase. *Biochem. J.* **233**, 139–44 (1986).
71. Tong, J., Shapiro, R. A. & Curthoys, N. P. Changes in the levels of translatable glutaminase mRNA during onset and recovery from metabolic acidosis. *Biochemistry* **26**, 2773–7 (1987).
72. Hwang, J. J., Perera, S., Shapiro, R. A. & Curthoys, N. P. Mechanism of altered renal glutaminase gene expression in response to chronic acidosis. *Biochemistry* **30**, 7522–6 (1991).
73. Kaiser, S. & Curthoys, N. P. Effect of pH and bicarbonate on phosphoenolpyruvate carboxykinase and glutaminase mRNA levels in cultured renal epithelial cells. *J. Biol. Chem.* **266**, 9397–402 (1991).
74. Gstraunthaler, G. *et al.* Differential expression and acid-base regulation of glutaminase mRNAs in gluconeogenic LLC-PK(1)-FBPase(+) cells. *Am. J. Physiol. Renal Physiol.* **278**, F227–37 (2000).
75. Hwang, J. J. & Curthoys, N. P. Effect of acute alterations in acid-base balance on rat renal glutaminase and phosphoenolpyruvate carboxykinase gene expression. *J. Biol. Chem.* **266**, 9392–6 (1991).
76. Dassi, E. *et al.* AURA 2: Empowering discovery of post-transcriptional networks. *Translation* **2**, e27738 (2014).
77. Hinman, M. N. & Lou, H. Diverse molecular functions of Hu proteins. *Cell. Mol. Life Sci.* **65**, 3168–3181 (2008).
78. Myer, V. E., Fan, X. C. & Steitz, J. A. Identification of HuR as a protein implicated in AUUUA-mediated mRNA decay. *EMBO J.* **16**, 2130–2139 (1997).
79. Chen, C. Y. *et al.* AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**, 451–464 (2001).
80. Kullmann, M., Göpfert, U., Siewe, B. & Hengst, L. ELAV/Hu proteins inhibit p27 translation via an IRES element in the p27 5'UTR. *Genes Dev.* **16**, 3087–3099 (2002).
81. Berkovits, B. D. & Mayr, C. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363–7 (2015).
82. Kriventseva, E. V. *et al.* OrthoDB v8: Update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* **43**, D250–D256 (2015).
83. Wang, H. *et al.* The structure of the ARE-binding domains of Hu antigen R (HuR) undergoes conformational changes during RNA binding. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 373–80 (2013).

84. Güttinger, S., Mühlhäusser, P., Koller-Eichhorn, R., Brennecke, J. & Kutay, U. Transportin2 functions as importin and mediates nuclear import of HuR. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2918–23 (2004).
85. Fan, X. C. & Steitz, J. a. HNS, a nuclear-cytoplasmic shuttling sequence in HuR. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 15293–8 (1998).
86. Gallouzi, I. E. & Steitz, J. A. Delineation of mRNA export pathways by the use of cell-permeable peptides. *Science (80-. )*. **294**, 1895–1901 (2001).
87. Silanes, I. L. de, Lal, A. & Gorospe, M. HuR: Post-Transcriptional Paths to Malignancy. *RNA Biol.* **2**, 11–13 (2005).
88. Young, L. E. *et al.* The mRNA Binding Proteins HuR and Tristetraprolin Regulate Cyclooxygenase 2 Expression During Colon Carcinogenesis. *Gastroenterology* **136**, 1669–1679 (2009).
89. Uren, P. J. *et al.* Genomic analyses of the RNA-binding protein Hu antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. *J. Biol. Chem.* **286**, 37063–6 (2011).
90. Izquierdo, J. M. Hu Antigen R (HuR) Functions as an Alternative Pre-mRNA Splicing Regulator of Fas Apoptosis-promoting Receptor on Exon Definition. *J. Biol. Chem.* **283**, 19077–19084 (2008).
91. Zhao, W. *et al.* HuR and TIA1/TIAL1 are involved in regulation of alternative splicing of SIRT1 pre-mRNA. *Int. J. Mol. Sci.* **15**, 2946–2958 (2014).
92. Chang, S.-H. *et al.* ELAVL1 regulates alternative splicing of eIF4E transporter to promote postnatal angiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18309–14 (2014).
93. Barnhart, M. D., Moon, S. L., Emch, A. W., Wilusz, C. J. & Wilusz, J. Changes in cellular mRNA stability, splicing, and polyadenylation through HuR protein sequestration by a cytoplasmic RNA virus. *Cell Rep.* **5**, 909–17 (2013).
94. Kotzer-Nevo, H., de Lima Alves, F., Rappsilber, J., Sperling, J. & Sperling, R. Supraspliceosomes at Defined Functional States Portray the Pre-Assembled Nature of the Pre-mRNA Processing Machine in the Cell Nucleus. *Int. J. Mol. Sci.* **15**, 11637–64 (2014).
95. Chen, Y.-I. I. G. *et al.* Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.* **35**, 3928–3944 (2007).
96. Lebedeva, S. *et al.* Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. *Mol. Cell* **43**, 340–352 (2011).
97. Mukherjee, N. *et al.* Integrative Regulatory Mapping Indicates that the RNA-Binding Protein HuR Couples Pre-mRNA Processing and mRNA Stability. *Mol. Cell* **43**, 327–339 (2011).
98. Akaike, Y. *et al.* HuR Regulates Alternative Splicing of the TRA2 Gene in Human Colon Cancer Cells under Oxidative Stress. *Mol. Cell. Biol.* **34**, 2857–2873 (2014).
99. Jens, M. *Dissecting Regulatory Interactions of RNA and Protein*. (Springer International Publishing, 2014). doi:10.1007/978-3-319-07082-7
100. Elkon, R., Ugalde, A. P. & Agami, R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* **14**, 496–506 (2013).
101. Dai, W., Zhang, G. & Makeyev, E. V. RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. *Nucleic Acids Res.* **40**, 787–800 (2012).
102. Giudice, G., Sánchez-Cabo, F., Torroja, C. & Lara-Pezzi, E. ATtRACT-a database of RNA-binding proteins and associated motifs. *Database* **2016**, 1–9 (2016).
103. Carazo, F., Romero, J. P. & Rubio, A. Upstream analysis of alternative splicing: a review of computational approaches to predict context-dependent splicing factors. *Brief.*

- Bioinform.* 1–18 (2018). doi:10.1093/bib/bby005
104. Abdelmohsen, K. & Gorospe, M. Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdiscip. Rev. RNA* **1**, 214–29 (2010).
  105. Embade, N. *et al.* Murine double minute 2 regulates Hu antigen R stability in human liver and colon cancer through NEDDylation. *Hepatology* **55**, 1237–48 (2012).
  106. Wang, J. *et al.* Multiple Functions of the RNA-Binding Protein HuR in Cancer Progression, Treatment Responses and Prognosis. *Int. J. Mol. Sci.* **14**, 10015–10041 (2013).
  107. Srikantan, S. & Gorospe, M. HuR function in disease. *Front. Biosci.* **17**, 189–205 (2012).
  108. Yiakouvaki, A. *et al.* Myeloid cell expression of the RNA-binding protein HuR protects mice from pathologic inflammation and colorectal carcinogenesis. *J. Clin. Invest.* **122**, 48–61 (2012).
  109. Ortega, Á. D., Sala, S., Espinosa, E., González-Barón, M. & Cuezva, J. M. HuR and the bioenergetic signature of breast cancer: A low tumor expression of the RNA-binding protein predicts a higher risk of disease recurrence. *Carcinogenesis* **29**, 2053–2061 (2008).
  110. Kotta-Loizou, I., Giaginis, C. & Theocharis, S. Clinical significance of HuR expression in human malignancy. *Med. Oncol.* **31**, 161 (2014).
  111. Yuan, Z., Sanders, A. J., Ye, L., Wang, Y. & Jiang, W. G. Prognostic value of the human antigen R (HuR) in human breast cancer: high level predicts a favourable prognosis. *Anticancer Res.* **31**, 303–10 (2011).
  112. Giaginis, C. *et al.* Elevated Hu-Antigen Receptor (HuR) Expression is Associated with Tumor Aggressiveness and Poor Prognosis but not with COX-2 Expression in Invasive Breast Carcinoma Patients. *Pathol. Oncol. Res.* (2017). doi:10.1007/s12253-017-0288-1
  113. Heinonen, M. *et al.* Cytoplasmic HuR expression is a prognostic factor in invasive ductal breast carcinoma. *Cancer Res.* **65**, 2157–2161 (2005).
  114. Heinonen, M. *et al.* Prognostic role of HuR in hereditary breast cancer. *Clin. Cancer Res.* **13**, 6959–6963 (2007).
  115. Zhu, Z. *et al.* Cytoplasmic HuR expression correlates with P-gp, HER-2 positivity, and poor outcome in breast cancer. *Tumor Biol.* **34**, 2299–2308 (2013).
  116. Wang, J., Wang, B., Bi, J. & Zhang, C. Cytoplasmic HuR expression correlates with angiogenesis, lymphangiogenesis, and poor outcome in lung cancer. *Med. Oncol.* **28** **Suppl 1**, S577-85 (2011).
  117. Denkert, C. *et al.* Expression of the ELAV-like protein HuR is associated with higher tumor grade and increased cyclooxygenase-2 expression in human breast carcinoma. *Clin. Cancer Res.* **10**, 5580–6 (2004).
  118. Wang, J., Li, D., Wang, B. & Wu, Y. Predictive and prognostic significance of cytoplasmic expression of ELAV-like protein HuR in invasive breast cancer treated with neoadjuvant chemotherapy. *Breast Cancer Res. Treat.* **141**, 213–224 (2013).
  119. Costa, G. L. & Weiner, M. P. Polishing with T4 or Pfu polymerase increases the efficiency of cloning of PCR fragments. *Nucleic Acids Res.* **22**, 2423 (1994).
  120. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–308 (2013).
  121. Orengo, J. P., Bundman, D. & Cooper, T. a. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res.* **34**, e148 (2006).
  122. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–23 (2013).
  123. Szymczak-Workman, A. L., Vignali, K. M. & Vignali, D. A. A. Design and Construction of 2A Peptide-Linked Multicistronic Vectors. *Cold Spring Harb. Protoc.* **2012**,

- pdb.ip067876-ip067876 (2012).
124. Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6**, 1–8 (2011).
  125. Moffat, J. *et al.* A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell* **124**, 1283–1298 (2006).
  126. Wiederschain, D. *et al.* Single-vector inducible lentiviral RNAi system for oncology target validation. *Cell Cycle* **8**, 498–504 (2009).
  127. R Development Core Team. R: A Language and Environment for Statistical Computing. (2011).
  128. Wickham, H. httr: Tools for Working with URLs and HTTP. R package version 1.3.1. (2017).
  129. Ooms, J. The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. (2014). doi:1403.2805
  130. Lang, D. T. & CRAN Team. XML: Tools for Parsing and Generating XML Within R and S-Plus. (2017).
  131. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).
  132. QGIS Development Team. QGIS Geographic Information System. (2009).
  133. Bivand, R. & Lewin-Koh, N. mapproj: Tools for Reading and Handling Spatial Objects. (2017).
  134. Peng, R. D., Murdoch, D., Rowlingson, B. & Murta, A. gpcplib: General Polygon Clipping Library for R. (2017).
  135. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2009).
  136. Phillips, N. yarr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R. (2017).
  137. Camp, R. L., Dolled-Filhart, M. & Rimm, D. L. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin. cancer Res.* **10**, 7252–9 (2004).
  138. Budczies, J. *et al.* Cutoff Finder: A Comprehensive and Straightforward Web Application Enabling Rapid Biomarker Cutoff Optimization. *PLoS One* **7**, 1–7 (2012).
  139. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model*. (Springer, 2000).
  140. Kassambara, A. & Kosinski, M. survminer: Drawing Survival Curves using 'ggplot2'. (2017).
  141. Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* **40**, 1–29 (2011).
  142. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statistical Soc. Ser. B* **57**, 289–300 (1995).
  143. Altman, D. G., Lausen, B., Sauerbrei, W. & Schumacher, M. Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. *J. Natl. Cancer Inst.* **86**, 829–35 (1994).
  144. Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, (2011).
  145. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
  146. Leinonen, R. *et al.* The European Nucleotide Archive. *Nucleic Acids Res.* 1–4 (2010). doi:10.1093/nar/gkq967
  147. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
  148. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* **8**, 559–564 (2011).

149. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, 754–761 (2017).
150. Andrews, S. FastQC: A quality control tool for high throughput sequence data. (2014).
151. Ewing, B. & Green, P. Base-calling of automated sequencer traces using Phred. II. Error Probabilities. *Genome Res.* **8**, 186–194 (1998).
152. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res.* **8**, 175–185 (1998).
153. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
154. MacManes, M. D. On the optimal trimming of high-throughput mRNAseq data. *bioRxiv* (2013). doi:10.1101/000422
155. Del Fabbro, C., Scalabrin, S., Morgante, M. & Giorgi, F. M. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One* **8**, e85024 (2013).
156. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
157. Harrow, J. *et al.* GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
158. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7 Suppl 1**, S4.1-9 (2006).
159. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
160. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinforma.* **51**, 11.14.1-11.14.19 (2015).
161. Mukherjee, N., Lager, P. J., Friedersdorf, M. B., Thompson, M. A. & Keene, J. D. Coordinated posttranscriptional mRNA population dynamics during T-cell activation. *Mol Syst Biol* **5**, 288 (2009).
162. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–9 (2015).
163. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–17 (2012).
164. Al Seesi, S., Tiagueu, Y. T., Zelikovsky, A. & Măndoiu, I. I. Bootstrap-based differential gene expression analysis for RNA-Seq data with and without replicates. *BMC Genomics* **15 Suppl 8**, S2 (2014).
165. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
166. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–45 (2009).
167. Sonesson, C., Matthes, K. L., Nowicka, M., Law, C. W. & Robinson, M. D. Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage. *Genome Biol.* **17**, 12 (2016).
168. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–78 (2012).
169. Zhang, Z. & Wang, W. RNA-Skim: A rapid method for RNA-Seq quantification at transcript level. *Bioinformatics* **30**, 283–292 (2014).
170. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
171. Pages, H., Carlson, M., Falcon, S. & Li, N. AnnotationDbi: Annotation Database Interface. (2016).
172. Carlson, M. org.Hs.eg.db: Genome wide annotation for Human.
173. Blake, J. A. *et al.* Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).

174. Carlson, M. GO.db: A set of annotation maps describing the entire Gene Ontology.
175. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
176. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, 472–477 (2014).
177. Ligtenberg, W. reactome.db: A set of annotation maps for reactome.
178. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. BioSyst.* **12**, 477–479 (2016).
179. Tsirigos, K. D., Peters, C., Shu, N., Lukas, K. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, 401–407 (2015).
180. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–22 (2010).
181. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, 222–230 (2014).
182. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci.* **95**, 5857–5864 (1998).
183. Letunic, I., Doerks, T. & Bork, P. SMART: recent updates , new developments and status in 2015. *Nucleic Acids Res.* **43**, 257–260 (2015).
184. Kanehisa, M., Goto, S., Sato, Y. & Kawashima, M. Data, information, knowledge and principle: back to metabolism in KEGG. **42**, 199–205 (2014).
185. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
186. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV ): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2012).
187. Mayor, C. *et al.* VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046–1047 (2000).
188. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, 273–279 (2004).
189. Tarailo-graovac, M. & Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinforma.* **Unit4.10**, 1–14 (2009).
190. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
191. Huang, X. & Miller, W. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12**, 337–357 (1991).
192. Ward, N. & Moreno-Hagelsieb, G. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLoS One* **9**, e101850 (2014).
193. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
194. Dehal, P. *et al.* The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–67 (2002).
195. Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).
196. Smith, J. J. *et al.* Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* **45**, 415–21, 421–2 (2013).
197. Venkatesh, B. *et al.* Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–9 (2014).
198. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).

199. Hellsten, U. *et al.* The Genome of the Western Clawed Frog *Xenopus tropicalis*. *Science* (80-. ). **328**, 633–636 (2010).
200. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* (80-. ). **291**, 1304–1351 (2001).
201. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
202. Holland, P. W. H. & Garcia-Fernández, J. Hox genes and chordate evolution. *Dev. Biol.* **173**, 382–95 (1996).
203. Flores, M. B. S. *et al.* Obesity-induced increase in tumor necrosis factor- $\alpha$  leads to development of colon cancer in mice. *Gastroenterology* **143**, 741-53–4 (2012).
204. Zor, T. & Selinger, Z. Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies. *Anal. Biochem.* **236**, 302–8 (1996).
205. Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680–685 (1970).
206. Timmons, T. M. & Dunbar, B. S. Protein blotting and immunodetection. in *Methods in Enzymology* **182**, 679–688 (1990).
207. Siddappa, N. B., Avinash, A., Venkatramanan, M. & Ranga, U. Regeneration of commercial nucleic acid extraction columns without the risk of carryover contamination. *Biotechniques* **42**, 186–192 (2007).
208. Kutner, R. H., Zhang, X.-Y. & Reiser, J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat. Protoc.* **4**, 495–505 (2009).
209. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
210. Tenenbaum, S. a, Lager, P. J., Carson, C. C. & Keene, J. D. Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods* **26**, 191–8 (2002).
211. Mazan-Mamczarz, K. *et al.* RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8354–9 (2003).
212. Chu, P., Kulp, S. K. & Chen, C. Insulin-like growth factor-I receptor is suppressed through transcriptional repression and mRNA destabilization by a novel energy restriction-mimetic agent. *Carcinogenesis* **34**, 2694–705 (2013).
213. Hofmann, A. *et al.* Epigenetic regulation of lentiviral transgene vectors in a large animal model. *Mol. Ther.* **13**, 59–66 (2006).
214. Ellis, J. Silencing and Variegation of Gammaretrovirus and Lentivirus Vectors. *Hum. Gene Ther.* **16**, 1241–1246 (2005).
215. Hassani, M., Lázaro, R., Pérez, C., Condón, S. & Pagán, R. Thermostability of oxytetracycline, tetracycline, and doxycycline at ultrahigh temperatures. *J. Agric. Food Chem.* **56**, 2676–80 (2008).
216. Pawley, J. Fundamental limits in confocal microscopy. *Handb. Biol. Confocal Microsc.* 19–37 (1995). doi:10.1007/978-0-387-45524-2\_2
217. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
218. Schneider, C. a, Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–5 (2012).
219. Zhao, Y. *et al.* Cytotoxicity enhancement in MDA-MB-231 cells by the combination treatment of tetrahydropalmatine and berberine derived from *Corydalis yanhusuo* W. T. Wang. *J. Intercult. Ethnopharmacol.* **3**, 68–72 (2014).
220. Pampaloni, F., Reynaud, E. G. & Stelzer, E. H. K. The third dimension bridges the gap

- between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol.* **8**, 839–845 (2007).
221. Cisneros Castillo, L. R., Oancea, A.-D., Stüllein, C. & Régnier-Vigouroux, A. Evaluation of Consistency in Spheroid Invasion Assays. *Sci. Rep.* **6**, 28375 (2016).
  222. Vinci, M., Box, C. & Eccles, S. A. Three-dimensional (3D) tumor spheroid invasion assay. *J. Vis. Exp.* e52686 (2015). doi:10.3791/52686
  223. Vinci, M. *et al.* Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. *BMC Biol.* **10**, 29 (2012).
  224. Friedrich, J., Seidel, C., Ebner, R. & Kunz-Schughart, L. a. Spheroid-based drug screen: considerations and practical approach. *Nat. Protoc.* **4**, 309–324 (2009).
  225. Roukos, V., Pegoraro, G., Voss, T. C. & Misteli, T. Cell cycle staging of individual cells by fluorescence microscopy. *Nat. Protoc.* **10**, 334–348 (2015).
  226. Lund, P. L-Glutamine and L-Glutamate: UV-Method with Glutaminase and Glutamate Dehydrogenase. in *Methods of Enzymatic Analysis* (ed. Bergmeyer, H. U.) 357–363 (VCH Verlagsgesellschaft, 1986).
  227. Wehrens, R., Weingart, G. & Mattivi, F. MetaMS: An open-source pipeline for GC-MS-based untargeted metabolomics. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **966**, 109–116 (2014).
  228. Giacomoni, F. *et al.* Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics* **31**, 1493–1495 (2015).
  229. Guitton, Y. *et al.* Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics. *Int. J. Biochem. Cell Biol.* **93**, 89–101 (2017).
  230. Kopka, J. *et al.* GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **21**, 1635–1638 (2005).
  231. Smyth, G. K. Limma: linear models for microarray data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. R, G., V, C., S, D. & WH, I.) 397–420 (Springer, 2005).
  232. Vu, V. Q. ggbiplot: A ggplot2 based biplot. (2017).
  233. Rajan, N., Habermehl, J., Coté, M.-F., Doillon, C. J. & Mantovani, D. Preparation of ready-to-use, storable and reconstituted type I collagen from rat tail tendon for tissue engineering applications. *Nat. Protoc.* **1**, 2753–2758 (2006).
  234. Albini, A. *et al.* A rapid in vitro assay for quantitating the invasive potential of tumor cells. *Cancer Res.* **47**, 3239–45 (1987).
  235. Islam, S. & Flaherty, P. *Assay Methods. Protocol: Cell Invasion Assay (CLS-DL-CC-031REV1)*. Corning Life Sciences (2012).
  236. Kramer, N. *et al.* In vitro cell migration and invasion assays. *Mutat. Res. Mutat. Res.* **752**, 10–24 (2013).
  237. Price, J. T. & Thompson, E. W. Models for studying cellular invasion of basement membranes. *Methods Mol. Biol.* **129**, 231–49 (1999).
  238. Hall, D. M. S. & Brooks, S. A. In Vitro Invasion Assay Using Matrigel™: A Reconstituted Basement Membrane Preparation. in *Methods in Molecular Biology* **1070**, 1–11 (2014).
  239. Chen, H.-C. Boyden chamber assay. *Methods Mol. Biol.* **294**, 15–22 (2005).
  240. Huang, H. *et al.* Identifying reproducible cancer-associated highly expressed genes with important functional significances using multiple datasets. *Sci. Rep.* **6**, 1–9 (2016).
  241. Grün, B. & Leisch, F. FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. *J. Stat. Softw.* **28**, 1–35 (2008).
  242. Fraley, C., Raftery, A. E., Murphy, T. B. & Scrucca, L. *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*.

- Technical Report 597* (2012).
243. Lehmann, B. D. B. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).
  244. Quintero, M. *et al.* Guanylate-binding protein-1 is a potential new therapeutic target for triple-negative breast cancer. *BMC Cancer* **17**, 727 (2017).
  245. Yazici, B. & Yolacan, S. A comparison of various tests of normality. *J. Stat. Comput. Simul.* **77**, 175–183 (2007).
  246. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
  247. Curtis, A. E., Smith, T. A., Ziganshin, B. A. & Elefteriades, J. A. The Mystery of the Z-Score. *AORTA* **4**, 124–130 (2016).
  248. National Cancer Institute. Head and Neck Cancers. (2017). Available at: <https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet>. (Accessed: 17th January 2018)
  249. Façal, S. & Shiota, D. Feocromocitoma: atualização diagnóstica e terapêutica. *Rev. Assoc. Med. Bras.* **43**, 237–244 (1997).
  250. Fentiman, I. S., Fourquet, A. & Hortobagyi, G. N. Male breast cancer. *Lancet* **367**, 595–604 (2006).
  251. Parker, J. S. *et al.* Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
  252. Lehmann, B. D. & Pietenpol, J. A. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. *J. Pathol.* **232**, 142–50 (2014).
  253. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
  254. Frankish, A. *et al.* Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. *BMC Genomics* **16 Suppl 8**, S2 (2015).
  255. Casper, J. *et al.* The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.* **46**, D762–D769 (2018).
  256. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
  257. Ascano, M., Hafner, M., Cekan, P., Gerstberger, S. & Tuschl, T. Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip. Rev. RNA* **3**, 159–77 (2011).
  258. Zambelli, F. & Pavesi, G. RIP-Seq data analysis to determine RNA-protein associations. *Methods Mol. Biol.* **1269**, 293–303 (2015).
  259. Sultan, M. *et al.* Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics* **15**, 675 (2014).
  260. Leng, N. *et al.* EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**, 1035–1043 (2013).
  261. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
  262. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–13 (2008).
  263. Bakheet, T., Hitti, E., Al-Saif, M., Moghrabi, W. N. & Khabar, K. S. A. The AU-rich element landscape across human transcriptome reveals a large proportion in introns and regulation by ELAVL1/HuR. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1861**, 167–177 (2018).
  264. Church, D. M. *et al.* Modernizing reference genome assemblies. *PLoS Biol.* **9**, e1001091

- (2011).
265. Farrell, C. M. *et al.* Current status and new features of the Consensus Coding Sequence database. *Nucleic Acids Res.* **42**, D865-72 (2014).
  266. Katz, Y., Wang, E. T., Airoidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
  267. Vasiliou, V., Vasiliou, K. & Nebert, D. W. Human ATP-binding cassette (ABC) transporter family. *Hum. Genomics* **3**, 281–90 (2009).
  268. Di Benedetto, R. *et al.* RNAi-mediated silencing of ABCD3 gene expression in rat C6 glial cells: a model system to study PMP70 function. *Neurochem. Int.* **52**, 1106–13 (2008).
  269. Kim, J.-Y. *et al.* Multiple-myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity. *Mol. Cell. Biol.* **28**, 2023–34 (2008).
  270. Stec, I. *et al.* WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmophy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum. Mol. Genet.* **7**, 1071–82 (1998).
  271. Ezponda, T. *et al.* The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial–mesenchymal transition and invasive properties of prostate cancer. *Oncogene* **32**, 2882–2890 (2013).
  272. Welin, M. *et al.* Structural studies of tri-functional human GART. *Nucleic Acids Res.* **38**, 7308–19 (2010).
  273. Brodsky, G. *et al.* The human GARS-AIRS-GART gene encodes two proteins which are differentially expressed during human brain development and temporally overexpressed in cerebellum of individuals with Down syndrome. *Hum. Mol. Genet.* **6**, 2043–50 (1997).
  274. Liu, X. *et al.* Glycinamide ribonucleotide formyl transferase is frequently overexpressed in glioma and critically regulates the proliferation of glioma cells. *Pathol. Res. Pract.* **210**, 256–263 (2014).
  275. Cong, X. *et al.* Increased expression of glycinamide ribonucleotide transformylase is associated with a poor prognosis in hepatocellular carcinoma, and it promotes liver cancer cell proliferation. *Hum. Pathol.* **45**, 1370–8 (2014).
  276. Zhang, H. *et al.* Genetic variations in the one-carbon metabolism pathway genes and susceptibility to hepatocellular carcinoma risk: a case–control study. *Tumor Biol.* **36**, 997–1002 (2015).
  277. Kobayashi, N. *et al.* RanBPM, Muskelin, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8alpha and ARMC8beta are components of the CTLH complex. *Gene* **396**, 236–47 (2007).
  278. Fan, C. *et al.* Armc8 expression was elevated during atypia-to-carcinoma progression and associated with cancer development of breast carcinoma. *Tumour Biol.* **35**, 11337–43 (2014).
  279. Hafner, M. *et al.* Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
  280. Corvelo, A., Hallegger, M., Smith, C. W. J. & Eyras, E. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput. Biol.* **6**, e1001016 (2010).
  281. Smith, C. W., Chu, T. T. & Nadal-Ginard, B. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell. Biol.* **13**, 4939–52 (1993).
  282. López de Silanes, I., Zhan, M., Lal, A., Yang, X. & Gorospe, M. Identification of a target

- RNA motif for RNA-binding protein HuR. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2987–92 (2004).
283. Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
284. Patthy, L. Genome evolution and the evolution of exon-shuffling--a review. *Gene* **238**, 103–14 (1999).
285. Liu, M. & Grigoriev, A. Protein domains correlate strongly with exons in multiple eukaryotic genomes--evidence of exon shuffling? *Trends Genet.* **20**, 399–403 (2004).
286. Long, M., de Souza, S. J. & Gilbert, W. Evolution of the intron-exon structure of eukaryotic genes. *Curr. Opin. Genet. Dev.* **5**, 774–8 (1995).
287. Levy, A., Sela, N. & Ast, G. TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. *Nucleic Acids Res.* **36**, D47-52 (2008).
288. Shapiro, J. Nothing in Evolution Makes Sense Except in the Light of Genomics: Read–Write Genome Evolution as an Active Biological Process. *Biology (Basel)*. **5**, 27 (2016).
289. Clarke, L. *et al.* The 1000 Genomes Project: data management and community access. *Nat. Methods* **9**, 1–4 (2012).
290. Eberhardt, W., Doller, A. & Pfeilschifter, J. Regulation of the mRNA-binding protein HuR by posttranslational modification: spotlight on phosphorylation. *Curr. Protein Pept. Sci.* **13**, 380–90 (2012).
291. INCA - Instituto Nacional de Cancer José Alencar Gomes da Silva. *Estimativa 2016: incidência de câncer no Brasil. Ministério da Saúde Instituto Nacional de Cancer José Alencar Gomes da Silva* (2016). doi:978-85-7318-283-5
292. Sherf, B. B. A., Navarro, S. L., Hannah, R. R. & Wood, K. V. Dual-Luciferase Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays. *Promega Notes Mag.* **2**, (1996).
293. Takeuchi, A., Hosokawa, M., Nojima, T. & Hagiwara, M. Splicing reporter mice revealed the evolutionally conserved switching mechanism of tissue-specific alternative exon selection. *PLoS One* **5**, e10946 (2010).
294. Coma, I., Herranz, J. & Martin, J. Statistics and Decision Making in High-Throughput Screening. in *High Throughput Screening. Methods in Molecular Biology (Methods and Protocols)* (eds. Janzen, W. & Bernasconi, P.) **565**, 69–105 (Humana Press, 2009).
295. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
296. He, X. *et al.* Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* **44**, e85 (2016).
297. Ehrenfeld, G. M. *et al.* Copper-Dependent Cleavage of DNA by Bleomycin. *Biochemistry* **26**, 931–942 (1987).
298. Drocourt, D., Calmels, T., Reynes, J. P., Baron, M. & Tiraby, G. Cassettes of the *Streptoalloteichus hindustanus* ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Res.* **18**, 4009 (1990).
299. Ma, H. T. & Poon, R. Y. C. Synchronization of HeLa Cells. in *Cell Cycle Synchronization. Methods in Molecular Biology (Methods and Protocols)* (ed. Banfalvi, G.) 151–161 (Humana Press, 2011). doi:10.1007/978-1-61779-182-6\_10
300. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S.* (Springer, 2002).
301. Ishikawa-Ankerhold, H. C., Ankerhold, R. & Drummen, G. P. C. Advanced fluorescence microscopy techniques--FRAP, FLIP, FLAP, FRET and FLIM. *Molecules* **17**, 4047–132 (2012).
302. Pucadyil, T. J. & Chattopadhyay, A. Confocal fluorescence recovery after photobleaching of green fluorescent protein in solution. *J. Fluoresc.* **16**, 87–94 (2006).

303. Al-Ahmadi, W., Al-Ghamdi, M., Al-Souhibani, N. & Khabar, K. S. A. miR-29a inhibition normalizes HuR over-expression and aberrant AU-rich mRNA stability in invasive cancer. *J. Pathol.* **230**, 28–38 (2013).
304. Manzoni, L. *et al.* Interfering with HuR-RNA Interaction: Design, Synthesis and Biological Characterization of Tanshinone Mimics as Novel, Effective HuR Inhibitors. *J. Med. Chem.* [acs.jmedchem.7b01176](https://doi.org/10.1021/acs.jmedchem.7b01176) (2018). doi:10.1021/acs.jmedchem.7b01176
305. Qie, S., Chu, C., Li, W., Wang, C. & Sang, N. ErbB2 activation upregulates glutaminase 1 expression which promotes breast cancer cell proliferation. *J. Cell. Biochem.* 1–29 (2013). doi:10.1002/jcb.24684
306. Fu, A., Yu, Z., Song, Y. & Zhang, E. Silencing of glutaminase 1 resensitizes Taxol-resistant breast cancer cells to Taxol. *Mol. Med. Rep.* **11**, 4727–4733 (2015).
307. Stetler-Stevenson, W. G., Aznavoorian, S. & Liotta, L. A. Tumor cell interactions with the extracellular matrix during invasion and metastasis. *Annu. Rev. Cell Biol.* **9**, 541–73 (1993).
308. Sun, B., Fang, Y., Li, Z., Chen, Z. & Xiang, J. Role of cellular cytoskeleton in epithelial-mesenchymal transition process during cancer progression (Review). *Biomed. Reports* **3**, 603–610 (2015).
309. Shapiro, I. M. *et al.* An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet.* **7**, e1002218 (2011).
310. Mori, M. *et al.* Zyxin mediates actin fiber reorganization in epithelial-mesenchymal transition and contributes to endocardial morphogenesis. *Mol. Biol. Cell* **20**, 3115–24 (2009).
311. Monzani, E., Shtil, A. a & La Porta, C. a M. The water channels, new druggable targets to combat cancer cell survival, invasiveness and metastasis. *Curr. Drug Targets* **8**, 1132–7 (2007).
312. Fassler, M. & Schick Tanz, S. *Analysis of Stem Cell Differentiation using the Operetta High Content Imaging System. PerkinElmer Application Note* (2011).
313. Dormoy-Raclet, V. *et al.* The RNA-Binding Protein HuR Promotes Cell Migration and Cell Invasion by Stabilizing the beta-actin mRNA in a U-Rich-Element-Dependent Manner. *Mol. Cell. Biol.* **27**, 5365–5380 (2007).
314. Langan, L. M. *et al.* Direct measurements of oxygen gradients in spheroid culture system using electron paramagnetic resonance oximetry. *PLoS One* **11**, 1–13 (2016).
315. Lin, R. Z. & Chang, H. Y. Recent advances in three-dimensional multicellular spheroid culture for biomedical research. *Biotechnol. J.* **3**, 1172–1184 (2008).
316. Arend, R. C., Londoño-Joshi, A. I., Straughn, J. M. & Buchsbaum, D. J. The Wnt/ $\beta$ -catenin pathway in ovarian cancer: A review. *Gynecol. Oncol.* **131**, 772–779 (2013).
317. Howe, L. R. & Brown, A. M. C. Wnt signaling and breast cancer. *Cancer Biol. Ther.* **3**, 36–41 (2004).
318. Nusse, R. & Clevers, H. Wnt/ $\beta$ -Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985–999 (2017).
319. Zhan, T., Rindtorff, N. & Boutros, M. Wnt signaling in cancer. *Oncogene* **36**, 1461–1473 (2017).
320. Essen BioScience. *Detailed Scratch Wound Cell Invasion Protocol for Collagen-1 + HT1080 cells Report 8000-0341-B00.* (2015).
321. Paddock, S. Confocal reflection microscopy: The ‘other’ confocal mode. *Biotechniques* **32**, 274–278 (2002).
322. Brightman, A. *et al.* Time-Lapse Confocal Reflection Microcopy of Collagen Fibrillogenesis and Extracellular Matrix Assembly In Vitro. *Biopolymers* **54**, 222–234 (2000).
323. Zhang, Z., Huang, A., Zhang, A. & Zhou, C. HuR promotes breast cancer cell

- proliferation and survival via binding to CDK3 mRNA. *Biomed. Pharmacother.* **91**, 788–795 (2017).
324. Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-response analysis using R. *PLoS One* **10**, 1–13 (2015).
  325. Gantt, K., Cherry, J., Tenney, R., Karschner, V. & Pekala, P. H. An early event in adipogenesis, the nuclear selection of the CCAAT enhancer-binding protein  $\beta$  (C/EBP $\beta$ ) mRNA by HuR and its translocation to the cytosol. *J. Biol. Chem.* **280**, 24768–24774 (2005).
  326. Sengupta, S. *et al.* The RNA-binding protein HuR regulates the expression of cyclooxygenase-2. *J. Biol. Chem.* **278**, 25227–33 (2003).
  327. Liu, L. *et al.* RNA-binding protein HuR promotes growth of small intestinal mucosa by activating the Wnt signaling pathway. *Mol. Biol. Cell* **25**, 3308–18 (2014).
  328. Mazan-Mamczarz, K. *et al.* Post-transcriptional gene regulation by HuR promotes a more tumorigenic phenotype. *Oncogene* **27**, 6151–63 (2008).
  329. Meng, Z. *et al.* The ELAV RNA-stability factor HuR binds the 5'-untranslated region of the human IGF-IR transcript and differentially represses cap-dependent and IRES-mediated translation. *Nucleic Acids Res.* **33**, 2962–79 (2005).
  330. Miyata, Y. *et al.* High Expression of HuR in Cytoplasm, but Not Nuclei, Is Associated with Malignant Aggressiveness and Prognosis in Bladder Cancer. *PLoS One* **8**, 1–8 (2013).
  331. Zhang, C. *et al.* Cytoplasmic expression of the ELAV-like protein HuR as a potential prognostic marker in esophageal squamous cell carcinoma. *Tumor Biol.* **35**, 73–80 (2014).
  332. Cho, N.-P., Han, H.-S., Soh, Y., Lee, K.-Y. & Son, H.-J. Cytoplasmic HuR over-expression is associated with increased cyclooxygenase-2 expression in laryngeal squamous cell carcinomas. *Pathology* **39**, 545–50 (2007).
  333. Greenhough, A. *et al.* The COX-2/PGE2 pathway: Key roles in the hallmarks of cancer and adaptation to the tumour microenvironment. *Carcinogenesis* **30**, 377–386 (2009).
  334. Dixon, D. a *et al.* Altered expression of the mRNA stability factor HuR promotes cyclooxygenase-2 expression in colon cancer cells. *J. Clin. Invest.* **108**, 1657–65 (2001).
  335. Sun, D. P. *et al.* Clinicopathological significance of HuR expression in gallbladder carcinoma: With special emphasis on the implications of its nuclear and cytoplasmic expression. *Tumor Biol.* **34**, 3059–3069 (2013).
  336. Yi, X., Zhou, Y., Zheng, W. & Chambers, S. K. HuR expression in the nucleus correlates with high histological grade and poor disease-free survival in ovarian cancer. *Aust. New Zeal. J. Obstet. Gynaecol.* **49**, 93–98 (2009).
  337. Lauriola, L., Granone, P., Ramella, S., Lanza, P. & Ranelletti, F. O. Expression of the RNA-binding protein HuR and its clinical significance in human stage I and II lung adenocarcinoma. *Histol Histopathol* **27**, 617–626 (2012).
  338. Zucal, C. *et al.* Targeting the Multifaceted HuR Protein, Benefits and Caveats. *Curr. Drug Targets* **16**, 1–17 (2015).
  339. Kang, M.-J. *et al.* NF-kappaB activates transcription of the RNA-binding factor HuR, via PI3K-AKT signaling, to promote gastric tumorigenesis. *Gastroenterology* **135**, 2030–2042 (2008).
  340. Liang, P. I. *et al.* HuR cytoplasmic expression is associated with increased cyclin A expression and poor outcome with upper urinary tract urothelial carcinoma. *BMC Cancer* **12**, 1–11 (2012).
  341. Abdelmohsen, K. *et al.* miR-519 suppresses tumor growth by reducing HuR levels. *Cell Cycle* **9**, 1354–9 (2010).
  342. Cho, N. H. *et al.* Elevation of cyclin B1, active cdc2, and HuR in cervical neoplasia with

- human papillomavirus type 18 infection. *Cancer Lett.* **232**, 170–8 (2006).
343. Danilin, S. *et al.* Role of the RNA-binding protein HuR in human renal cell carcinoma. *Carcinogenesis* **31**, 1018–26 (2010).
  344. Xu, F. *et al.* Loss of repression of HuR translation by miR-16 may be responsible for the elevation of HuR in human breast carcinoma. *J. Cell. Biochem.* **111**, 727–34 (2010).
  345. Raspaglio, G. *et al.* HuR regulates beta-tubulin isotype expression in ovarian cancer. *Cancer Res.* **70**, 5891–900 (2010).
  346. Koljonen, V., Böhling, T., Haglund, C. & Ristimäki, A. Expression of HuR in Merkel cell carcinoma and in normal skin. *J. Cutan. Pathol.* **35**, 10–4 (2008).
  347. Leijon, H. *et al.* HuR in pheochromocytomas and paragangliomas - overexpression in verified malignant tumors. *APMIS* **124**, 757–63 (2016).
  348. West, M. *et al.* Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11462–11467 (2001).
  349. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–6 (2002).
  350. Sabatier, R., Finetti, P., Cervera, N., Birnbaum, D. & Bertucci, F. Gene expression profiling and prediction of clinical outcome in ovarian cancer. *Crit. Rev. Oncol. Hematol.* **72**, 98–109 (2009).
  351. Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113**, 913–23 (2004).
  352. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8418–23 (2003).
  353. Jimbo, M. *et al.* Targeting the mRNA-binding protein HuR impairs malignant characteristics of pancreatic ductal adenocarcinoma cells. *Oncotarget* **6**, 27312–31 (2015).
  354. Burkhart, R. A. *et al.* HuR is a post-transcriptional regulator of core metabolic enzymes in pancreatic cancer. *RNA Biol.* **10**, 1312–23 (2013).
  355. Chand, S. N. *et al.* Post-transcriptional regulation of PARG mRNA by HuR facilitates DNA repair and resistance to PARP inhibitors. *Cancer Res.* **1**, canres.2704.2016 (2017).
  356. Heinemann, V. Gemcitabine: progress in the treatment of pancreatic cancer. *Oncology* **60**, 8–18 (2001).
  357. Costantino, C. L. *et al.* The role of HuR in gemcitabine efficacy in pancreatic cancer: HuR Up-regulates the expression of the gemcitabine metabolizing enzyme deoxycytidine kinase. *Cancer Res.* **69**, 4567–72 (2009).
  358. Oba, A. *et al.* Correlation between the Acquisition of Resistance to Gemcitabine Therapy and the Expression of HuR in Pancreatic Ductal Adenocarcinoma: a case report. *Int. Surg.* INTSURG-D-15-00278.1 (2016). doi:10.9738/INTSURG-D-15-00278.1
  359. Venet, D., Dumont, J. E. & Detours, V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput. Biol.* **7**, (2011).
  360. Mazumdar, M. & Glassman, J. R. Prognostic Variables: Categorizing a Prognostic Variable: Review of Methods, Code for Easy Implementation and Applications to Decision-Making about Cancer Treatments. *Stat. Med.* **1**, 187–208 (2005).
  361. Hilsenbeck, S. G., Clark, G. M. & McGuire, W. L. Why do so many prognostic factors fail to pan out? *Breast Cancer Res. Treat.* **22**, 197–206 (1992).
  362. Cantor, A. B. & Shuster, J. J. Re: Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. *J. Natl. Cancer Inst.* **86**, 1798–9 (1994).
  363. Altman, D. Suboptimal analysis using 'optimal' cutpoints. *Br. J. Cancer* **78**, 550–557 (1998).
  364. Iasonos, A., Schrag, D., Raj, G. V & Panageas, K. S. How to build and interpret a

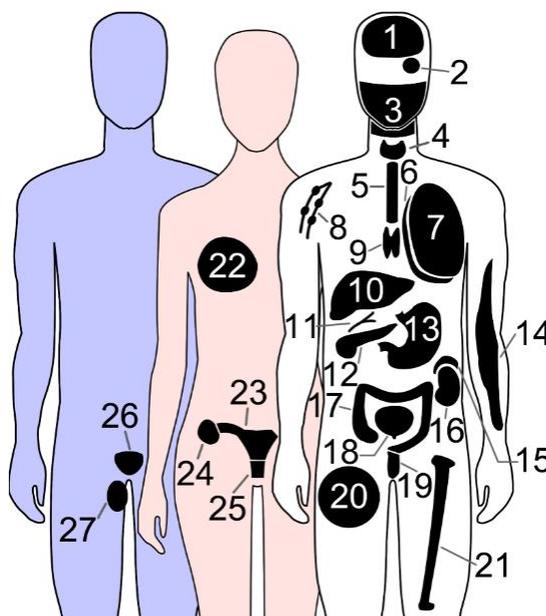
- nomogram for cancer prognosis. *J. Clin. Oncol.* **26**, 1364–70 (2008).
365. Clark, T. G., Bradburn, M. J., Love, S. B. & Altman, D. G. Survival Analysis Part I: Basic concepts and first analyses. *Br. J. Cancer* **89**, 232–238 (2003).
  366. Andersen, P. K., Geskus, R. B., de Witte, T. & Putter, H. Competing risks in epidemiology: possibilities and pitfalls. *Int. J. Epidemiol.* **41**, 861–70 (2012).
  367. Ince-Dunn, G. *et al.* Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. *Neuron* **75**, 1067–80 (2012).
  368. Xiang, Y. *et al.* Targeted inhibition of tumor-specific glutaminase diminishes cell-autonomous tumorigenesis. *J. Clin. Invest.* **125**, 2293–2306 (2015).
  369. National Library of Medicine. Glutaminase + Cancer Clinical Trials. (2018). Available at: <https://clinicaltrials.gov/ct2/results?cond=Cancer&term=glutaminase&cntry=&state=&city=&dist=>. (Accessed: 24th January 2018)
  370. Van Den Heuvel, A. P. J., Jing, J., Wooster, R. F. & Bachman, K. E. Analysis of glutamine dependency in non-small cell lung cancer: GLS1 splice variant GAC is essential for cancer cell growth. *Cancer Biol. Ther.* **13**, 1185–1194 (2012).
  371. Szeliga, M. *et al.* Relative expression of mRNAs coding for glutaminase isoforms in CNS tissues and CNS tumors. *Neurochem. Res.* **33**, 808–813 (2008).
  372. Xia, Z. *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* **5**, 5274 (2014).
  373. Kamarajan, P. *et al.* Head and Neck Squamous Cell Carcinoma Metabolism Draws on Glutaminolysis, and Stemness Is Specifically Regulated by Glutaminolysis via Aldehyde Dehydrogenase. *J. Proteome Res.* **16**, 1315–1326 (2017).
  374. Huang, F., Zhang, Q., Ma, H., Lv, Q. & Zhang, T. Expression of glutaminase is upregulated in colorectal cancer and of clinical significance. *Int. J. Clin. Exp. Pathol.* **7**, 1093–1100 (2014).
  375. Szeliga, M. & Obara-Michlewska, M. Glutamine in neoplastic cells: Focus on the expression and roles of glutaminases. *Neurochem. Int.* **55**, 71–75 (2009).
  376. Huang, W. *et al.* A proposed role for glutamine in cancer cell growth through acid resistance. *Cell Res.* **23**, 724–7 (2013).
  377. Weiner, I. D. & Verlander, J. W. Renal ammonia metabolism and transport. *Compr. Physiol.* **3**, 201–220 (2013).
  378. Robinson-Rechavi, M., Boussau, B. & Laudet, V. Phylogenetic Dating and Characterization of Gene Duplications in Vertebrates: The Cartilaginous Fish Reference. *Mol. Biol. Evol.* **21**, 580–586 (2004).
  379. Ditrich, H. The origin of vertebrates: a hypothesis based on kidney development. *Zool. J. Linn. Soc.* **150**, 435–441 (2007).
  380. Carrier, J., Musick, J. & Heithaus, M. *Biology of Sharks and their relatives*. (CRC Press, 2004).
  381. Bean, E. S. & Atkinson, D. E. Regulation of the rate of urea synthesis in liver by extracellular pH: a major factor in pH homeostasis in mammals. *J. Biol. Chem.* **259**, 1552–1559 (1984).
  382. Keilin, J. The biological significance of uric acid and guanine excretion. *Biol. Rev.* **34**, 265–294 (1958).
  383. Krakoff, I. H. & Karnofsky, D. A. Inhibition of uric acid biosynthesis in birds by o-diazoacetyl-L-serine (azaserine) and 6-diazo-5-oxo-L-norleucine (DON). *Am. J. Physiol.* **195**, 244–250 (1958).
  384. Craan, A. G., Lemieux, G., Vinay, P. & Gougoux, A. The kidney of chicken adapts to chronic metabolic acidosis: in vivo and in vitro studies. *Kidney Int* **22**, 103–11 (1982).

385. Zhang, X., Lin, Y. & Gillies, R. J. Tumor pH and its measurement. *J. Nucl. Med.* **51**, 1167–70 (2010).
386. Swietach, P., Vaughan-Jones, R. D. & Harris, A. L. Regulation of tumor pH and the role of carbonic anhydrase 9. *Cancer Metastasis Rev.* **26**, 299–310 (2007).
387. Laterza, O. F. & Curthoys, N. P. Specificity and functional analysis of the pH-responsive element within renal glutaminase mRNA. *Am. J. Physiol. Renal Physiol.* **278**, F970-7 (2000).
388. Curthoys, N. P. Glutamine Metabolism : Nutritional and Clinical Significance Role of Mitochondrial Glutaminase in Rat Renal Glutamine Metabolism. *J. Nutr.* 2491–2495 (2001).
389. Paukku, K. *et al.* Regulation of AT1R expression through HuR by insulin. *Nucleic Acids Res.* **40**, 5250–61 (2012).
390. Humphreys, D. T., Westman, B. J., Martin, D. I. K. & Preiss, T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16961–6 (2005).
391. Cok, S. J. & Morrison, A. R. The 3'-Untranslated Region of Murine Cyclooxygenase-2 Contains Multiple Regulatory Elements that Alter Message Stability and Translational Efficiency. *J. Biol. Chem.* **276**, 23179–23185 (2001).
392. Harding, H. P. *et al.* Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* **6**, 1099–1108 (2000).
393. Laterza, O. F., Hansen, W. R., Taylor, L. & Curthoys, N. P. Identification of an mRNA-binding protein and the specific elements that may mediate the pH-responsive induction of renal glutaminase mRNA. *J. Biol. Chem.* **272**, 22481–8 (1997).
394. Zhu, H., Hasman, R. A., Barron, V. A., Luo, G. & Lou, H. A nuclear function of Hu proteins as neuron-specific alternative RNA processing regulators. *Mol. Biol. Cell* **17**, 5105–14 (2006).
395. Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419** VN-, 182–185 (2002).
396. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–45 (2002).
397. Zhang, X. H. F., Leslie, C. S. & Chasin, L. A. Dichotomous splicing signals in exon flanks. *Genome Res.* **15**, 768–779 (2005).
398. López De Silanes, I. *et al.* The RNA-binding protein HuR regulates DNA methylation through stabilization of DNMT3b mRNA. *Nucleic Acids Res.* **37**, 2658–2671 (2009).
399. Al-Ahmadi, W., Al-Ghamdi, M., Al-Haj, L., Al-Saif, M. & Khabar, K. S. a. Alternative polyadenylation variants of the RNA binding protein, HuR: abundance, role of AU-rich elements and auto-Regulation. *Nucleic Acids Res.* **37**, 3612–24 (2009).
400. Lema, I. *et al.* HuR-Dependent Editing of a New Mineralocorticoid Receptor Splice Variant Reveals an Osmoregulatory Loop for Sodium Homeostasis. *Sci. Rep.* **7**, 1–13 (2017).
401. Li, H. *et al.* HuR antagonizes the effect of an intronic pyrimidine-rich sequence in regulating WT1 +/-KTS isoforms. *RNA Biol.* **12**, 1364–1371 (2015).
402. Zhu, H., Hinman, M. N., Hasman, R. A., Mehta, P. & Lou, H. Regulation of Neuron-Specific Alternative Splicing of Neurofibromatosis Type 1 Pre-mRNA. *Mol. Cell. Biol.* **28**, 1240–1251 (2008).
403. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
404. Trapnell, C. *et al.* transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat. Biotechnol.* **28**, 511–515 (2010).

405. Jeanteur, P. *Regulation of Alternative Splicing. Bioelectrochemistry and Bioenergetics* **31**, (Springer Berlin Heidelberg, 2003).
406. Neubauer, G. *et al.* Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* **20**, 46–50 (1998).
407. Doller, A., Schlepckow, K., Schwalbe, H., Pfeilschifter, J. & Eberhardt, W. Tandem phosphorylation of serines 221 and 318 by protein kinase Cdelta coordinates mRNA binding and nucleocytoplasmic shuttling of HuR. *Mol. Cell. Biol.* **30**, 1397–410 (2010).
408. Doller, A. *et al.* Posttranslational modification of the AU-rich element binding protein HuR by protein kinase Cdelta elicits angiotensin II-induced stabilization and nuclear export of cyclooxygenase 2 mRNA. *Mol. Cell. Biol.* **28**, 2608–25 (2008).
409. Yaman, I. *et al.* Nutritional control of mRNA stability is mediated by a conserved AU-rich element that binds the cytoplasmic shuttling protein HuR. *J. Biol. Chem.* **277**, 41539–46 (2002).
410. Kim, J. Bin, Stein, R. & O’Hare, M. J. Three-dimensional in vitro tissue culture models of breast cancer-- a review. *Breast Cancer Res. Treat.* **85**, 281–91 (2004).
411. Gatenby, R. A. *et al.* Cellular adaptations to hypoxia and acidosis during somatic evolution of breast cancer. *Br. J. Cancer* **97**, 646–653 (2007).
412. Kim, I., Hur, J. & Jeong, S. HuR represses Wnt/ $\beta$ -catenin-mediated transcriptional activity by promoting cytoplasmic localization of  $\beta$ -catenin. *Biochem. Biophys. Res. Commun.* **293**, (2014).
413. D’Uva, G. *et al.* Beta-catenin/HuR post-transcriptional machinery governs cancer stem cell features in response to hypoxia. *PLoS One* **8**, 1–17 (2013).
414. Woo, H.-H. *et al.* Posttranscriptional Suppression of Proto-Oncogene c-fms Expression by Vigilin in Breast Cancer. *Mol. Cell. Biol.* **31**, 215–225 (2011).
415. Erickson, J. W. & Cerione, R. a. Glutaminase: a hot spot for regulation of cancer cell metabolism? *Oncotarget* **1**, 734–740 (2010).
416. Meisner, N. C. *et al.* Identification and mechanistic characterization of low-molecular-weight inhibitors for HuR. *Nat. Chem. Biol.* **3**, 508–515 (2007).
417. Chae, M. J. *et al.* Chemical inhibitors destabilize HuR binding to the AU-rich element of TNF- $\alpha$  mRNA. *Exp. Mol. Med.* **41**, 824–831 (2009).
418. Wu, X. *et al.* Identification and Validation of Novel Small Molecule Disruptors of HuR-mRNA Interaction. *ACS Chem. Biol.* **10**, 1476–1484 (2015).
419. D’Agostino, V. G., Adami, V. & Provenzani, A. A Novel High Throughput Biochemical Assay to Evaluate the HuR Protein-RNA Complex Formation. *PLoS One* **8**, 1–9 (2013).

## APÊNDICES

## Apêndice 1: Legenda da representação topográfica dos órgãos dados disponíveis no TCGA



Sexo	#	Tecido	Código TCGA	Tipo
Indiferente	1	Cérebro	GBM	Glioblastoma multiforme
			LGG	Glioma de baixo grau
	2	Olho	UVM†	Melanoma uveal
	3	Cabeça e pescoço	HNSC	Carcinoma de células escamosas de cabeça e pescoço
	4	Tireóide	THCA	Carcinoma papilar de tireóide
	5	Esôfago	ESCA	Câncer de esôfago
	6	Pleura	MESO†	Mesotelioma
	7	Pulmão	LUAD	Adenocarcinoma de pulmão
			LUSC	Carcinoma de células escamosas de pulmão
	8	Linfonodos	DLBC	Linfoma difuso de grandes células B
	9	Timo	THYM†	Timoma
	10	Fígado	LIHC	Carcinoma hepatocelular
	11	Ducto biliar	CHOL†	Colangiocarcinoma
	12	Pâncreas	PAAD	Adenocarcinoma pancreático ductal
	13	Estômago	STAD	Adenocarcinoma de estômago
	14	Pele	SKCM	Melanoma cutâneo
	15	Adrenal	ACC†	Carcinoma Adrenocortical
			PCPG†	Feocromocitoma & Paraganglioma
	16	Rim	KIRC	Carcinoma renal de células claras
			KICH†	Carcinoma renal cromóforo
			KIRP	Carcinoma papilar renal
	17	Intestino grosso	COAD	Adenocarcinoma de cólon
	18	Bexiga	BLCA	Cancer de bexiga
	19	Intestino delgado	READ	Adenocarcinoma de reto
	20	Tecidos macios	SARC†	Sarcoma
	21	Medula óssea	LAML	Leucemia mielóide aguda
	Feminino	22	Mama*	BRCA
UCEC				Câncer de endométrio
23		Útero	UCS†	Carcinossarcoma uterino
	24	Ovário	OV	Cistadenocarcinomas serosos do ovário
	25	Cérvice uterino	CESC	Câncer cervical
Masculino	26	Prostata	PRAD	Adenocarcinoma de próstata
	27	Testículos	TGCT†	Câncer de células germinativas de testículo

† Tumores considerados raros pelo TCGA. \* Indicado como tumor *tipicamente* feminino, também acometendo homens.

**Apêndice 2: Lista de genes em que HuR liga em introns justapostos a exons envolvidos com *splicing* alternativo e apresentam ao menos um exon diferencialmente abundante quando do silenciamento de HuR (continua)**

Ensembl ID	Gene ID	HUGO	Ensembl ID	Gene ID	HUGO
ENSG00000126070	192669	AGO3	ENSG00000151914	667	DST
ENSG00000117528	5825	ABCD3	ENSG00000104047	56986	DTWD1
ENSG00000138443	10152	ABI2	ENSG00000135930	9470	EIF4E2
ENSG00000123983	2181	ACSL3	ENSG00000134759	55250	ELP2
ENSG00000008277	53616	ADAM22	ENSG00000132199	55556	ENOSF1
ENSG00000101126	23394	ADNP	ENSG00000112851	55914	ERBIN
ENSG00000173744	3267	AGFG1	ENSG00000120705	2107	ETF1
ENSG00000004455	204	AK2	ENSG00000166801	63901	FAM111A
ENSG00000131503	54882	ANKHD1	ENSG00000113391	83989	FAM172A
ENSG00000084234	334	APLP2	ENSG00000172775	80011	FAM192A
ENSG00000100852	394	ARHGAP5	ENSG00000158169	2176	FANCC
ENSG00000162980	26225	ARL5A	ENSG00000145907	10146	G3BP1
ENSG00000113966	84100	ARL6	ENSG00000159131	2618	GART
ENSG00000114098	25852	ARMC8	ENSG00000090863	2734	GLG1
ENSG00000112249	10973	ASCC3	ENSG00000115419	2744	GLS
ENSG00000066279	259266	ASPM	ENSG00000141404	2774	GNAL
ENSG00000017260	27032	ATP2C1	ENSG00000087460	2778	GNAS
ENSG00000033627	535	ATP6V0A1	ENSG00000089916	55668	GPATCH2L
ENSG00000085224	546	ATRX	ENSG00000151233	283464	GXYLT1
ENSG00000029363	9774	BCLAF1	ENSG00000105968	94239	H2AFV
ENSG00000148120	84909	C9orf3	ENSG00000048052	9734	HDAC9
ENSG00000128595	813	CALU	ENSG00000100644	3091	HIF1A
ENSG00000153113	831	CAST	ENSG00000092199	3183	HNRNPC
ENSG00000175455	64770	CCDC14	ENSG00000138668	3184	HNRNPD
ENSG00000152492	152137	CCDC50	ENSG00000125944	10236	HNRNPR
ENSG00000149231	79780	CCDC82	ENSG00000166130	121457	IKBIP
ENSG00000196776	961	CD47	ENSG00000115604	8809	IL18R1
ENSG00000188419	1121	CHM	ENSG00000071243	54556	ING3
ENSG00000141367	1213	CLTC	ENSG00000068745	51447	IP6K2
ENSG00000143771	29097	CNIH4	ENSG00000090376	11213	IRAK3
ENSG00000125107	23019	CNOT1	ENSG00000102781	84056	KATNAL1
ENSG00000214078	8904	CPNE1	ENSG00000121774	10657	KHDRBS1
ENSG00000009307	7812	CSDE1	ENSG00000115919	8942	KYNU
ENSG00000113712	1452	CSNK1A1	ENSG00000073803	9175	MAP3K13
ENSG00000044115	1495	CTNNA1	ENSG00000152601	4154	MBNL1
ENSG00000109861	1075	CTSC	ENSG00000175471	79772	MCTP1
ENSG00000158290	8450	CUL4B	ENSG00000185787	10933	MORF4L1
ENSG00000257923	1523	CUX1	ENSG00000198015	28977	MRPL42
ENSG00000055163	26999	CYFIP2	ENSG00000057935	57504	MTA3
ENSG00000100523	80821	DDHD1	ENSG00000100330	8897	MTMR3
ENSG00000198563	7919	DDX39B	ENSG00000197535	4644	MYO5A
ENSG00000198947	1756	DMD	ENSG00000244754	10443	N4BP2L2
ENSG00000119661	83544	DNAL1	ENSG00000132780	4678	NASP
ENSG00000188641	1806	DPYD	ENSG00000158092	4690	NCK1

**Apêndice 2: Lista de genes em que HuR liga em introns justapostos a exons envolvidos com *splicing* alternativo e apresentam ao menos um exon diferencialmente abundante quando do silenciamento de HuR (continuação)**

Ensembl ID	Gene ID	HUGO	Ensembl ID	Gene ID	HUGO
ENSG00000196712	4763	NF1	ENSG00000165416	10910	SUGT1
ENSG00000181163	4869	NPM1	ENSG00000102710	55578	SUPT20H
ENSG00000110713	4928	NUP98	ENSG00000135316	10492	SYNCRIP
ENSG00000109180	54940	OCIAD1	ENSG00000179152	285343	TCAIM
ENSG00000115947	5000	ORC4	ENSG00000108064	7019	TFAM
ENSG00000197111	5094	PCBP2	ENSG00000114126	7029	TFDP2
ENSG00000100731	22990	PCNX1	ENSG00000151923	7073	TIAL1
ENSG00000170248	10015	PDCD6IP	ENSG00000143183	54499	TMCO1
ENSG00000113448	5144	PDE4D	ENSG00000117500	50999	TMED5
ENSG00000204138	65979	PHACTR4	ENSG00000106460	54664	TMEM106B
ENSG00000118482	23469	PHF3	ENSG00000109133	55161	TMEM33
ENSG00000115762	55041	PLEKHB2	ENSG00000103978	25963	TMEM87A
ENSG00000135241	50640	PNPLA8	ENSG00000120802	7112	TMPO
ENSG00000145725	23262	PPIP5K2	ENSG00000083312	3842	TNPO1
ENSG00000100614	5494	PPM1A	ENSG00000076554	7163	TPD52
ENSG00000138814	5530	PPP3CA	ENSG00000153827	9320	TRIP12
ENSG00000080815	5663	PSEN1	ENSG00000135451	10024	TROAP
ENSG00000100902	5687	PSMA6	ENSG00000163728	151613	TTC14
ENSG00000197170	5718	PSMD12	ENSG00000182670	7267	TTC3
ENSG00000117569	58155	PTBP2	ENSG00000101019	55245	UQCC1
ENSG00000110958	10728	PTGES3	ENSG00000101558	9218	VAPA
ENSG00000184007	8073	PTP4A2	ENSG00000196715	154807	VKORC1L1
ENSG00000112531	9444	QKI	ENSG00000197969	23230	VPS13A
ENSG00000161800	29127	RACGAP1	ENSG00000118965	57539	WDR35
ENSG00000116473	5906	RAP1A	ENSG00000133316	54663	WDR74
ENSG00000162521	5928	RBBP4	ENSG00000109685	7468	WHSC1
ENSG00000117625	55758	RCOR3	ENSG00000060237	65125	WNK1
ENSG00000137075	152006	RNF38	ENSG00000198373	11060	WWP2
ENSG00000008988	6224	RPS20	ENSG00000143324	9213	XPR1
ENSG00000143252	6391	SDHC	ENSG00000079246	7520	XRCC5
ENSG00000138674	22872	SEC31A	ENSG00000015153	10138	YAF2
ENSG00000183291	9403	SEP15	ENSG00000058673	9877	ZC3H11A
ENSG00000168385	4735	SEPT2	ENSG00000121741	7750	ZMYM2
ENSG00000097033	51100	SH3GLB1	ENSG00000010244	7756	ZNF207
ENSG00000136603	6498	SKIL	ENSG00000167380	7769	ZNF226
ENSG00000111371	81539	SLC38A1	ENSG00000112200	26036	ZNF451
ENSG00000113658	4090	SMAD5	ENSG00000171466	54811	ZNF562
ENSG00000159140	6651	SON	ENSG00000179195	144348	ZNF664
ENSG00000153914	140890	SREK1	ENSG00000197472	57116	ZNF695
ENSG00000188529	10772	SRSF10	ENSG00000120963	51123	ZNF706
ENSG00000141380	6760	SS18	ENSG00000160352	148206	ZNF714
ENSG00000114850	6747	SSR3	ENSG00000198040	7637	ZNF84
ENSG00000165209	55342	STRBP	ENSG00000036549	26009	ZZZ3
ENSG00000113387	10923	SUB1			

## ANEXOS

## ANEXO 1

Redis, R. S., Vela, L. E., Lu, W., Ferreira de Oliveira, J., Ivan, C., Rodriguez-Aguayo, C., ..., Adamoski, D., ... Calin, G. A. (2016). Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2. *Molecular Cell*, 61(4), 520–534. <https://doi.org/10.1016/j.molcel.2016.01.015>

A política editorial do grupo *Cell Press* permite a reprodução do texto em teses e dissertações (<https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>).

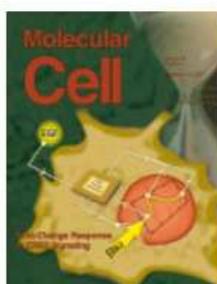


RightsLink®

Home

Create Account

Help



**Title:** Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2

**Author:** Roxana S. Redis, Luz E. Vela, Weiqin Lu, Juliana Ferreira de Oliveira, Cristina Ivan, Cristian Rodriguez-Aguayo, Douglas Adamoski, Barbara Pasculli, Ayumu Taguchi, Yunyun Chen, Agustin F. Fernandez, Luis Valledor, Katrien Van Roosbroeck, Samuel Chang et al.

**Publication:** *Molecular Cell*

**Publisher:** Elsevier

**Date:** 18 February 2016

Copyright © 2016 Elsevier Inc. All rights reserved.

LOGIN

If you're a **copyright.com user**, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

# Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2

Roxana S. Redis,<sup>1</sup> Luz E. Vela,<sup>2</sup> Weiqin Lu,<sup>3,4</sup> Juliana Ferreira de Oliveira,<sup>5</sup> Cristina Ivan,<sup>6</sup> Cristian Rodriguez-Aguayo,<sup>1</sup> Douglas Adamoski,<sup>5</sup> Barbara Pasculli,<sup>1</sup> Ayumu Taguchi,<sup>3</sup> Yunyun Chen,<sup>8</sup> Agustin F. Fernandez,<sup>9</sup> Luis Valledor,<sup>10</sup> Katrien Van Roosbroeck,<sup>1</sup> Samuel Chang,<sup>1</sup> Maitri Shah,<sup>1</sup> Garrett Kinnebrew,<sup>11</sup> Leng Han,<sup>12</sup> Yaser Atlasi,<sup>13,22</sup> Lawrence H. Cheung,<sup>1</sup> Gilbert Y. Huang,<sup>1</sup> Paloma Monroig,<sup>1</sup> Marc S. Ramirez,<sup>14</sup> Tina Catela Ivkovic,<sup>1,15</sup> Long Van,<sup>2</sup> Hui Ling,<sup>1</sup> Roberta Gafà,<sup>16</sup> Sanja Kapitanovic,<sup>15</sup> Giovanni Lanza,<sup>16</sup> James A. Bankson,<sup>14</sup> Peng Huang,<sup>3</sup> Stephen Y. Lai,<sup>8</sup> Robert C. Bast,<sup>1</sup> Michael G. Rosenblum,<sup>1</sup> Milan Radovich,<sup>11</sup> Mircea Ivan,<sup>17</sup> Geoffrey Bartholomeusz,<sup>1</sup> Han Liang,<sup>18</sup> Mario F. Fraga,<sup>19</sup> William R. Widger,<sup>2</sup> Samir Hanash,<sup>7</sup> Ioana Berindan-Neagoe,<sup>1,20,21</sup> Gabriel Lopez-Berestein,<sup>1,6</sup> Andre L.B. Ambrosio,<sup>5</sup> Sandra M. Gomes Dias,<sup>5</sup> and George A. Calin<sup>1,6,\*</sup>

<sup>1</sup>Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>2</sup>Department of Biology & Biochemistry, University of Houston, Houston, TX 77204, USA

<sup>3</sup>Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>4</sup>Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>5</sup>Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas 13083-100, Brazil

<sup>6</sup>Center for RNA Interference and Non-coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>7</sup>Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>8</sup>Department of Head & Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>9</sup>Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA), HUCA, Universidad de Oviedo, Oviedo 33006, Spain

<sup>10</sup>Department of Organisms and Systems Biology, University of Oviedo, Oviedo 33006, Spain

<sup>11</sup>Department of Surgery, Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>12</sup>Department of Biochemistry and Molecular Biology, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX 77030, USA

<sup>13</sup>Department of Pathology, Josephine Nefkens Institute, Erasmus Medical Center, Rotterdam 3015, the Netherlands

<sup>14</sup>Department of Imaging Physics, Division of Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>15</sup>Laboratory for Personalized Medicine, Division of Molecular Medicine, Ruder Boskovic Institute, Zagreb 10000, Croatia

<sup>16</sup>Department of Morphology, Surgery and Experimental Medicine, University of Ferrara, Ferrara 44121, Italy

<sup>17</sup>Department of Medicine, Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

<sup>18</sup>Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>19</sup>Nanomaterials and Nanotechnology Research Center (CINN-CSIC), Asturias 33424, Spain

<sup>20</sup>Research Center for Functional Genomics, Biomedicine and Translational Medicine, University of Medicine and Pharmacy Iuliu Hatieganu, Cluj-Napoca 400012, Romania

<sup>21</sup>Department of Functional Genomics, The Oncology Institute, Cluj-Napoca 400015, Romania

<sup>22</sup>Present address: Radboud Institute for Molecular Life Sciences (RIMLS) Department of Molecular Biology, Nijmegen 6525, the Netherlands

\*Correspondence: [gcalin@mdanderson.org](mailto:gcalin@mdanderson.org)

<http://dx.doi.org/10.1016/j.molcel.2016.01.015>

## SUMMARY

Altered energy metabolism is a cancer hallmark as malignant cells tailor their metabolic pathways to meet their energy requirements. Glucose and glutamine are the major nutrients that fuel cellular metabolism, and the pathways utilizing these nutrients are often altered in cancer. Here, we show that the long ncRNA CCAT2, located at the 8q24 amplicon on cancer risk-associated rs6983267 SNP, regulates cancer metabolism in vitro and in vivo in an allele-specific manner by binding the Cleavage Factor I (CFIm) complex with distinct affinities for the two subunits (CFIm25 and CFIm68). The CCAT2 interaction with the CFIm complex fine-tunes the alternative splicing of *Glutaminase* (*GLS*) by selecting the poly(A) site in intron 14 of the precursor mRNA. These

findings uncover a complex, allele-specific regulatory mechanism of cancer metabolism orchestrated by the two alleles of a long ncRNA.

## INTRODUCTION

Long non-coding RNAs (lncRNAs) form the largest part of the mammalian non-coding transcriptome (Mercer et al., 2009) and are generally expressed in a developmental-, tissue-, or disease-stage-specific manner, which makes them attractive therapeutic targets (Ling et al., 2013a). Although the underlying molecular mechanisms are not yet entirely understood, lncRNAs control gene expression at various levels including chromatin modification and transcriptional and post-transcriptional processing (Wilusz et al., 2009).

The revival of Warburg's theory of cancer (Warburg et al., 1924), complemented with novel discoveries in the field, has

promoted cellular metabolism as an essential molecular mechanism for driving malignant transformation and progression (Boroughs and DeBerardinis, 2015). Various studies have exposed the fine interplay between metabolic pathways orchestrated by protein-coding oncogenes and tumor suppressor genes (Chen and Russo, 2012) and, more recently, by ncRNAs (miRNAs and lncRNAs) (Gao et al., 2009; Yang et al., 2014). Glutamine, one of the essential nutrients, is deaminated by *Glutaminase* (*GLS*) to produce glutamate, which further serves as substrate for a variety of metabolic pathways (e.g., tricarboxylic cycle, TCA). Glutamine metabolism is modulated by MYC via miR-23a/b in prostate cancer and B cell lymphoma (Gao et al., 2009) and by NF- $\kappa$ B p65 subunit also through miR-23a downregulation in leukemic cells (Rathore et al., 2012; Wang et al., 2010b).

Colon Cancer-Associated Transcript 2 (*CCAT2*), a lncRNA that spans the highly conserved 8q24 region harboring the rs6983267 SNP (Ling et al., 2013b; Redis et al., 2013), is associated with increased risk for various types of cancer (Tomlinson et al., 2007; Tuupanen et al., 2009) and is specifically overexpressed in the microsatellite stable colorectal cancer (CRC MSS). The two alleles of the rs6983267 SNP present in the general population have been shown to render distinct risks of CRC; namely, the G allele was associated with greater predisposition to CRC than the T allele (Tomlinson et al., 2007). *CCAT2* induces chromosomal instability and metastases (Ling et al., 2013b) and regulates the expression levels of MYC oncogene, known to coordinate multiple molecular pathways supporting cell proliferation, metastases and cancer metabolism (Carroll et al., 2015; Stine et al., 2015). However, it is not clear how the two alleles are specifically involved in the malignant process. In this study, we demonstrate that the lncRNA, *CCAT2*, modulates cellular energy metabolism in an allele-specific manner by interacting with the Cleavage Factor I (CFIm) complex to regulate the alternative splicing of *GLS*.

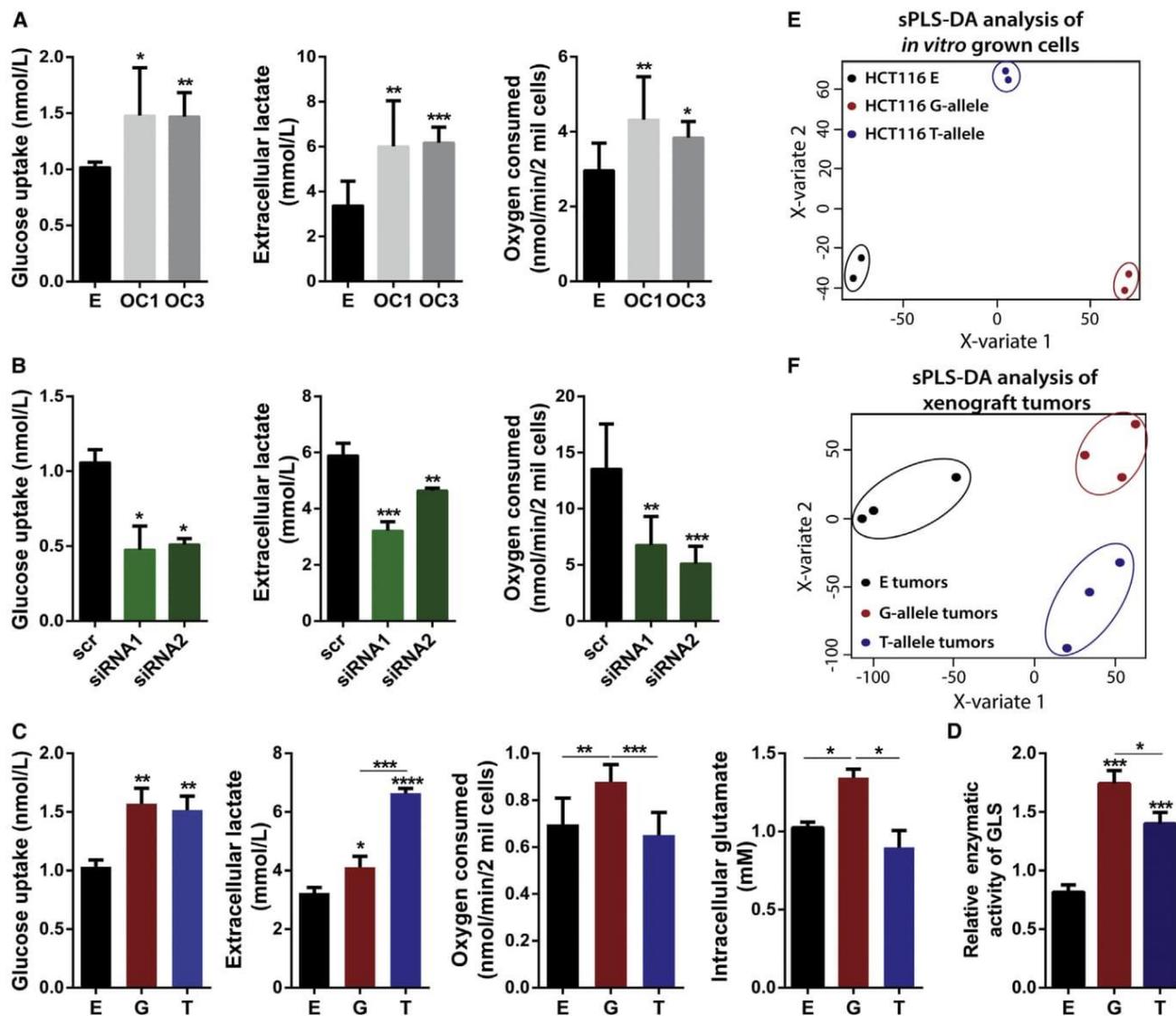
## RESULTS

### **CCAT2 Modulates Energy Metabolism in an Allele-Specific Manner In Vitro and In Vivo**

We observed an unexpected change in the color of the media of in vitro grown cells when modulating the expression of *CCAT2* that suggested a possible shift in the energy metabolism consequent to *CCAT2* expression. We tested this hypothesis by measuring metabolic parameters in HCT116 colon cancer cells that stably overexpress *CCAT2* (OC1 and OC3) (Ling et al., 2013b) versus control cells and observed a significant and reproducible increase in glucose uptake, lactate secretion and oxygen consumption in the *CCAT2*-overexpressing clones (Figure 1A). These results were further confirmed in KM12SM cells with *CCAT2* downregulated expression (Figure 1B). Moreover, we explored whether these metabolic changes were occurring in vivo as well by injecting HCT116 *CCAT2*-overexpressing cells and control cells subcutaneously into nude mice and subjecting them to hyperpolarized magnetic resonance imaging (HP-MRI). We detected a significant increase in the flux of hyperpolarized [ $1\text{-}^{13}\text{C}$ ]pyruvate to [ $1\text{-}^{13}\text{C}$ ]lactate for the xenograft tumors derived from the *CCAT2*-overexpressing cells compared to the tumors derived from control cells (Figures S1A and S1B), consistent

with our in vitro findings. These findings confirm that *CCAT2* alters metabolism, boosting glycolysis and cellular respiration. The coexistence of increased glycolysis with increased respiration, in highly proliferative cells, translates into enhanced anaplerotic reactions that replenish the TCA cycle intermediates (Ward and Thompson, 2012). Since glutamine is the main source for replenishing the intermediates of the TCA cycle, we measured the intra- and extracellular glutamate concentration, as well as the glutamine uptake in HCT116 cells with *CCAT2* overexpression and control cells. We found higher levels of both intra- and extracellular glutamate corresponding to higher levels of *CCAT2* (Figures S1C and S1D), suggesting *CCAT2* is boosting glutamine metabolism (glutaminolysis). Surprisingly, the glutamine uptake was not significantly different between the three clones (Figure S1D), implying the higher glutamate is not due to increased glutamine consumption. Therefore, we measured the enzymatic activity of *GLS*, the rate-limiting enzyme of glutaminolysis, in the whole lysate of the same cells and detected significantly higher activity in the cells with increased *CCAT2* expression (Figure S1F). In addition, both metabolic pathways (glycolysis and glutaminolysis) have been shown to be regulated by many factors, including the MYC oncogene (Carroll et al., 2015; Stine et al., 2015), a target of *CCAT2* by our previous report (Ling et al., 2013b).

We next explored whether the rs6983267 SNP influences these metabolic changes and assessed the glucose and glutamine uptake, oxygen consumption, lactate secretion, and intra- and extracellular glutamate concentration in HCT116 stably overexpressing either the G or the T allele of *CCAT2* and control HCT116 cells. Interestingly, we found on one hand higher glucose uptake and secreted glutamate in both G and T allele cells compared to control cells, while on the other hand we observed significant differences in lactate secretion, oxygen consumption, and intracellular glutamate production between the alleles (Figures 1C and S1E). Moreover, the glutamine consumption was not significantly different between the clones, similar to our previous results (Figure S1E). Consequently, we measured *GLS* enzymatic activity in these cells and observed that both *CCAT2* alleles induced a remarkable increase in activity compared to control, but the cells overexpressing the G allele displayed a significantly higher enzymatic activity compared to the T allele-overexpressing cells (Figure 1D). We also analyzed by mass spectroscopy the metabolites obtained from in vitro culturing of the HCT116 *CCAT2* G or T allele and control cells and from in vivo xenografted tumors derived from subcutaneous injection of the same cells. We observed contrasting distribution patterns when performing partial least squares discriminant analysis (PLS-DA) for both the in vitro (Figure 1E) and in vivo analyses (Figure 1F), and similarly for the principal component analysis (PCA) (Figures S1G and S1H). We detected 85.04% (in vitro) and 59.55% (in vivo) of metabolic pathways upregulated by *CCAT2* G allele compared to the T allele (Figure S1I, Table S1B). We then compared the pathway analysis for both datasets and identified 40 common pathways for the G allele and five common pathways for the T allele (Figure S1J). For these pathways, metabolic cluster distribution of differentially accumulated compounds revealed a significant overall enhancement of metabolic pathways related to glucose metabolism, TCA cycle, and glutamine metabolism for the G allele cells compared to the T allele cells (Tables S1A and S1B). We



**Figure 1. CCAAT2 Regulates Cancer Metabolism In Vitro and In Vivo**

(A) Glucose uptake, lactate production, and oxygen consumption assays in HCT116 stable clones (E: empty control vector; OC1 and OC3: CCAAT2-overexpressing, GG genotype).

(B) Glucose uptake, lactate production, and oxygen consumption assays in KM12SM cells with CCAAT2 downregulation (GT genotype).

(C) Whole-cell lysate *Glutaminase* activity measured in HCT116 CCAAT2-overexpressing G or T allele and control cells.

(D) Partial least-squares discriminant analysis (sPLS-DA) of HCT116 cells stably overexpressing CCAAT2 with either G or T alleles, and control cells (E, empty vector) *in vitro* allowed an adequate classification of the different cell lines according to its metabolome. sPLS-DA algorithm allows the classification of the samples based on the different abundances of each metabolite trying to find the maximum covariance between treatments and metabolome, in this way finding the most important metabolites for explaining the different effects of the treatments.

(E) Xenograft tumors derived from the same cell lines were also correctly classified by sPLS-DA analysis.

Results are presented as normalized mean values  $\pm$  SD. See also Figure S1 and Table S1.

evaluated the expression of MYC between the HCT116 CCAAT2 G and T allele cells but did not find impressive differences, consistent with our previous findings (Ling et al., 2013b) (Figure S2A). These results imply that in cells with high CCAAT2 expression, MYC activates energy metabolism in a general fashion; however, the fine-tuning of distinct metabolic pathways may occur through MYC-independent, but SNP-dependent, mechanisms.

We therefore decided to direct our efforts toward exploring the CCAAT2 MYC-independent mechanism of regulation.

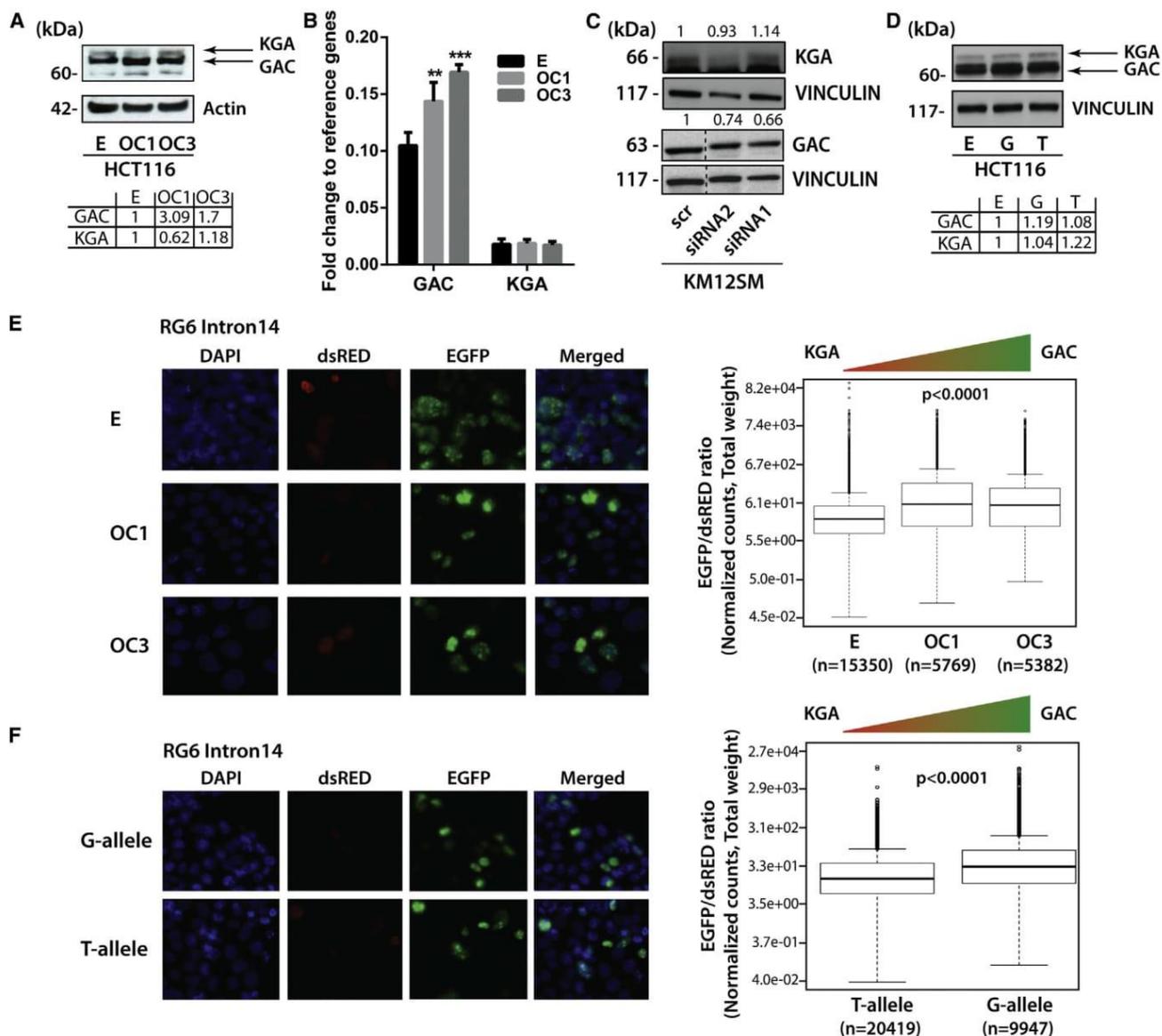
#### CCAAT2 Regulates the Expression of GLS Isoforms

Our group has previously shown that CCAAT2 induces chromosomal instability, a process highly reliant on the supply of nucleotides (Bester et al., 2011) and intimately linked to glutamine

metabolism (Jeong et al., 2013). Further supported by the metabolic data, we directed our focus toward pathways metabolizing glutamine. We first assessed the protein expression of the two alternative splicing isoforms of *GLS*, KGA (glutaminase kidney isoform) and GAC (glutaminase isoform C) as a function of *CCAT2*. We used specific antibodies for each isoform (recognizing the distinct C termini) (Cassago et al., 2012) and/or a common antibody, recognizing the N terminus shared by the two isoforms, depending on the cell line. Although the two isoforms share the same active site, GAC has a higher catalytic activity than KGA and therefore may be more relevant for replenishing intermediates of the TCA cycle (Cassago et al., 2012; Le et al., 2012). While for GAC we observed an increase in protein expression in the HCT116 *CCAT2*-overexpressing cells with both antibodies, for KGA the protein expression presented an inconsistent variation with *CCAT2* upregulation (Figures 2A, S2B, and S2C). Analogously, the mRNA expression pattern for GAC and KGA reflected the protein expression (Figure 2B). Moreover, downregulation of *CCAT2* in KM12SM cells reduced GAC protein expression with 34% and 26%, respectively (Figures 2C, S2D, and S2E), whereas KGA protein expression was either unaltered or slightly increased (Figures 2C, S2D, and S2E). Similar results were obtained when measuring the mRNA expression of the two isoforms in the same cellular model (Figure S2F). Interestingly, when we evaluated the expression of the two isoforms in the HCT116 cells overexpressing the *CCAT2* G or T allele and control cells at both the mRNA and protein levels, we found unanimously higher expression of GAC when *CCAT2* G allele is upregulated (Figures 2D and S2G–S2I). However, changes in the KGA mRNA expression pattern did not concur with the protein expression (Figures 2D and S2G–S2I). The discordance observed for the KGA isoform is probably due to the regulation by the MYC-miR-23 axis (Gao et al., 2009; additional data available from authors upon request). These results alluded to the idea that *CCAT2* may preferentially induce the splicing of the GAC isoform. To determine this, we cloned the intron 14 of *GLS* precursor mRNA, known to encompass the alternative splicing site, in the RG6 bichromatic fluorescent reporter (Orengo et al., 2006). If the splicing machinery binds to the intron 14, it induces the splicing of a EGFP-tagged protein, which is the equivalent of GAC; otherwise, a dsRED-tagged protein will be produced, which is the equivalent of KGA (Figure S2J). We transfected the HCT116 control and *CCAT2*-overexpressing cells with the fluorescent reporter and determined the ratio between the expression levels of EGFP and dsRED fluorescence using the VECTRA automated imaging system. Although the EGFP-tagged protein, GAC-equivalent, was predominant in all models, we observed a 50% significantly higher EGFP/dsRED ratio in the *CCAT2*-overexpressing cells, corresponding to 50% more alternative splicing events (Figure 2E). Similar results were obtained when cells were analyzed by flow cytometry (Table S2). Moreover, when we compared the alternative splicing events occurring in the G and T alleles, using the RG6 reporter, we found significantly higher EGFP/dsRED ratio for the *CCAT2* G allele compared to the *CCAT2* T allele (Figure 2F and Table S2). Altogether, these data demonstrate that the *CCAT2* G allele is more efficient in boosting the alternative splicing of GAC isoform.

### CFIm25 Governs the Switch between *GLS* Splicing Isoforms

To elucidate the underlying mechanism of *CCAT2*-induced regulation of the *GLS* alternative splicing, we introduced the MS2 tag (24 repeats) into vectors containing either the *CCAT2* G allele or T allele, pulled down the proteins that bind *CCAT2*, and analyzed them by mass spectrometry (Yoon et al., 2012). Pathway analysis on the QIAGEN platform identified Cleavage and Polyadenylation of Pre-mRNA among the top pathways associated with the G allele (Figure S3A), with CFIm25, the small (25 kDa) subunit of CFIm encoded by the *NUDT21* gene (Elkon et al., 2013; Yang et al., 2011), as the main protein correlated with the pathway. Notably, in the T allele pull down, both CFIm25 and the larger subunit (68 kDa) of the CFIm complex, CFIm68 (encoded by the *CPSF6* gene), were detected (CFIm68 was not detected in the G allele pull down); however, for CFIm25, the area under the peak had a 1.48 higher fold change in the G allele compared to the T allele (Figure S3B). We screened the *GLS* intron 14 for potential splicing and/or alternative polyadenylation (APA) sites using ASTRA (Alternative Splicing and Transcription Archives database; <http://dbarchive.biosciencedbc.jp/en/astra/desc.html>) (Nagasaki et al., 2006) and identified a type 2 (skipping exon) poly(A) site (Figure S3C) (Lutz and Moreira, 2011) and multiple conserved binding motifs (UGUA) for CFIm25, consistent with a previous report mapping two poly(A) motifs within the same intron of the *GLS* pre-mRNA (Tian et al., 2007). Considering that CFIm25 and the CFIm heterotetramer complex have been previously linked to alternative splicing (Millevoi et al., 2006; Zhou et al., 2002), we first downregulated *NUDT21* and assessed the protein levels of GAC and KGA. We noticed a significant decrease in GAC protein expression in both cellular models and a clear increase in the KGA protein expression (Figure 3A). We also measured mRNA expression for the two isoforms and observed a significant reduction of the GAC/KGA mRNA ratio with *NUDT21* knockdown (Figure 3B). In order to evaluate whether the switch in isoform expression is a consequence of CFIm25 binding to the UGUA sequences within intron 14 of *GLS* pre-mRNA, we designed antisense synthetic oligonucleotides (ASOs) to block the binding of the 25 kDa subunit to these motifs (Figure 3C). Out of the four tested ASOs, we could identify two that were able to reverse the GAC to KGA protein expression ratio similar to the specific downregulation of *NUDT21* in HCT116 OC1 cells (Figure 3C). This result suggested that binding of CFIm25 to intron 14 is responsible for inducing the preferential expression of GAC. To further confirm the direct interaction of CFIm25 with *GLS* pre-mRNA and *CCAT2*, we immunoprecipitated the RNA bound to constituting proteins of the CFIm complex, CFIm25 and CFIm68, in HCT116 cells overexpressing either the G or T allele of *CCAT2* and control cells, and measured the differences in RNA enrichment between the distinct pull-down lysates by real-time quantitative PCR (qRT-PCR) (Figures 3D–3F and S3D–S3F). We included two lncRNAs as controls: *NEAT1* was chosen to be a positive control as it has been previously shown to interact with the CFIm complex (Naganuma et al., 2012), and *GAS5* was chosen as negative control due to minimal sequence similarity with *CCAT2* (Figure S3G). When assessing the fold enrichment of *GLS* and *CCAT2*, we detected, respectively, 5.77 and 13.6 times more RNA bound to CFIm25 in the



**Figure 2. CCAT2 Induces the Preferential Splicing of GAC**

(A) Western blot analysis of GAC, KGA in HCT116 CCAT2-overexpressing cells (OC1 and OC3) and control cells.

(B) qRT-PCR assessing the mRNA expression of GAC and KGA in HCT116 CCAT2-overexpressing cells (OC1 and OC3) and control cells.

(C) Western blot analysis of GAC, KGA in KM12SM cells with CCAT2 downregulation.

(D) Western blot analysis of GAC, KGA in HCT116 stably overexpressing CCAT2 G or T alleles and control cells.

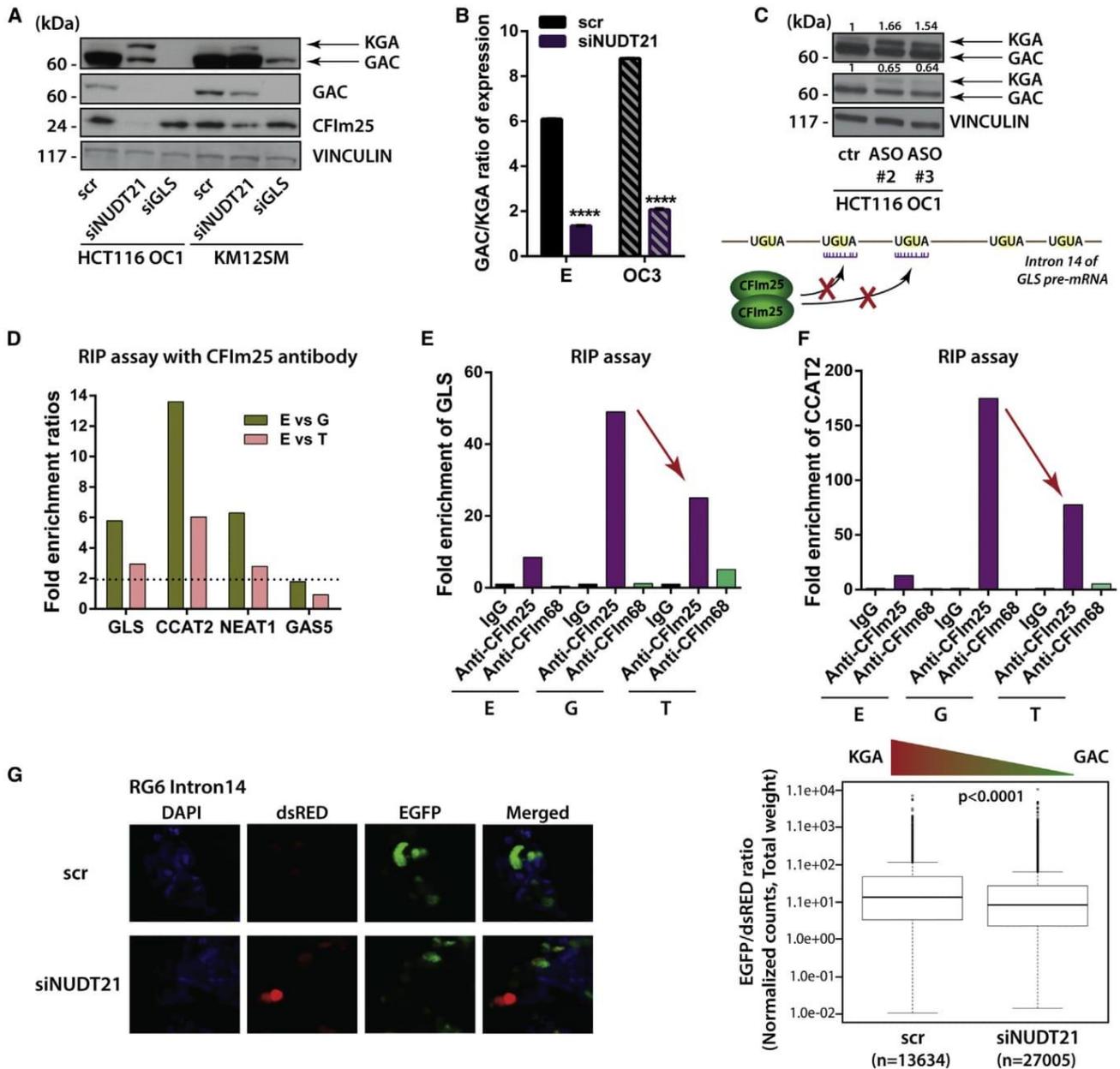
(E) Fluorescence microscopy images of HCT116 stable clones (E, empty control vector; OC1 and OC3, CCAT2 overexpressing) transfected with the RG6 intron 14 vector and the analysis of the EGFP/dsRED ratio ( $n$  = number of analyzed cells).

(F) Fluorescence microscopy images of HCT116 CCAT2 G allele and T allele transfected with the RG6 intron 14 vector and the analysis of the EGFP/dsRED ratio ( $n$  = number of analyzed cells).

Results are presented as normalized mean values  $\pm$  SD. See also Figure S2 and Table S2.

cells overexpressing CCAT2 G allele compared to control cells, while in the cells overexpressing CCAT2 T allele compared to control cells, the fold enrichment ratios were only about half (2.95 and 6.02, respectively) (Figure 3D). Moreover, comparing the fold enrichment in the G-overexpressing cells to the T-overexpressing cells, we observed roughly double fold enrichment in

the G-overexpressing cells for both GLS and CCAT2 (1.95 and 2.26) (Figures 3E and 3F). The positive control (*NEAT1*) presented 6.3-fold more RNA bound to CFIm25 in the cells overexpressing CCAT2 G allele compared to control cells and 2.79-fold more RNA enrichment in the cells overexpressing CCAT2 T allele compared to control cells (Figures 3D and S3D). The negative



### Figure 3. CFIm Protein Complex Binds *GLS* Pre-mRNA

(A) Western blot analysis of CFIm25, GAC, and KGA in HCT116 OC1 and KM12SM cells transiently transfected with siRNA for NUDT21, GLS (targeting the coding sequence shared by the two isoforms), and siRNA control.

(B) qRT-PCR assessing the GAC/KGA mRNA ratio in HCT116 CCAT2-overexpressing cells (OC1 and OC3) and control cells (E) with modulated CFIm25 expression.

(C) Western blot analysis of GAC and KGA in HCT116 OC1 cells with transient blockage of CFIm25 binding motifs (UGUA) by antisense oligonucleotides (ASOs). Schematic representation of the mechanism is presented below.

(D) qRT-PCR assessing the fold enrichment of *GLS*, *CCAT2*, *NEAT1*, and *GAS5* RNA bound to CFIm25 protein (RNA immunoprecipitation). Data are presented as fold enrichment ratios between control HCT116 cells (E) and *CCAT2*-overexpressing G or T allele.

(E and F) qRT-PCR assessing the fold enrichment of *GLS* mRNA (E) and *CCAT2* (F) bound to CFIm25 and CFIm68 in HCT116 cells, *CCAT2*-overexpressing G or T alleles, and control cells (E).

(G) Fluorescence microscopy images of KM12SM cells transfected with siNUDT21 and scr, followed by transfection with the RG6 intron 14 vector, and the analysis of the EGFP/dsRED ratio (n = number of analyzed cells).

Results are presented as normalized mean values  $\pm$  SD. See also Figure S3 and Table S2.

control (*GAS5*) revealed only a 1.79-fold increase in the RNA bound to CFIm25 in the cells overexpressing *CCAT2* G allele compared to control cells and no difference in RNA enrichment between the overexpressing *CCAT2* T allele and control cells (Figures 3D and S3E). Thus, we concluded that, overall, in the cells overexpressing *CCAT2*, there is an augmentation of interaction between CFIm25 and *GLS* and *CCAT2*, with the highest levels in the G-overexpressing cells. The low binding of the RNAs to the CFIm68 was not surprising, as the main function of the protein is merely to enhance RNA binding and facilitate RNA looping, while the 25 kDa subunit has the leading role in binding the RNA via the UGUA elements (Yang et al., 2011) (Figures 3E, 3F, S3D, and S3E). Nonetheless, for our RNAs of interest, we observed an increase in enrichment in the T allele overexpressing cells compared to the G allele overexpressing and control cells, suggesting the secondary structure of *CCAT2* T allele may ease the interaction between the 68 kDa subunit and RNA molecules. In addition, we confirmed these results by end-point PCR in both the same cellular model and the KM12SM cell line, with a GT-heterozygous genotype (Figures S4A and S4B). As a further validation, we expressed the RG6 bichromatic fluorescent reporter with the intron 14 in KM12SM cells with knockdown of *NUDT21*. We assessed, by both VECTRA and FACS, the ratio of EGFP to dsRED and found a significantly lower (28%) EGFP/dsRED ratio in the cells with *NUDT21* downregulation, corresponding to fewer splicing events and consequently lower expression of the GAC equivalent (Figure 3G and Table S2).

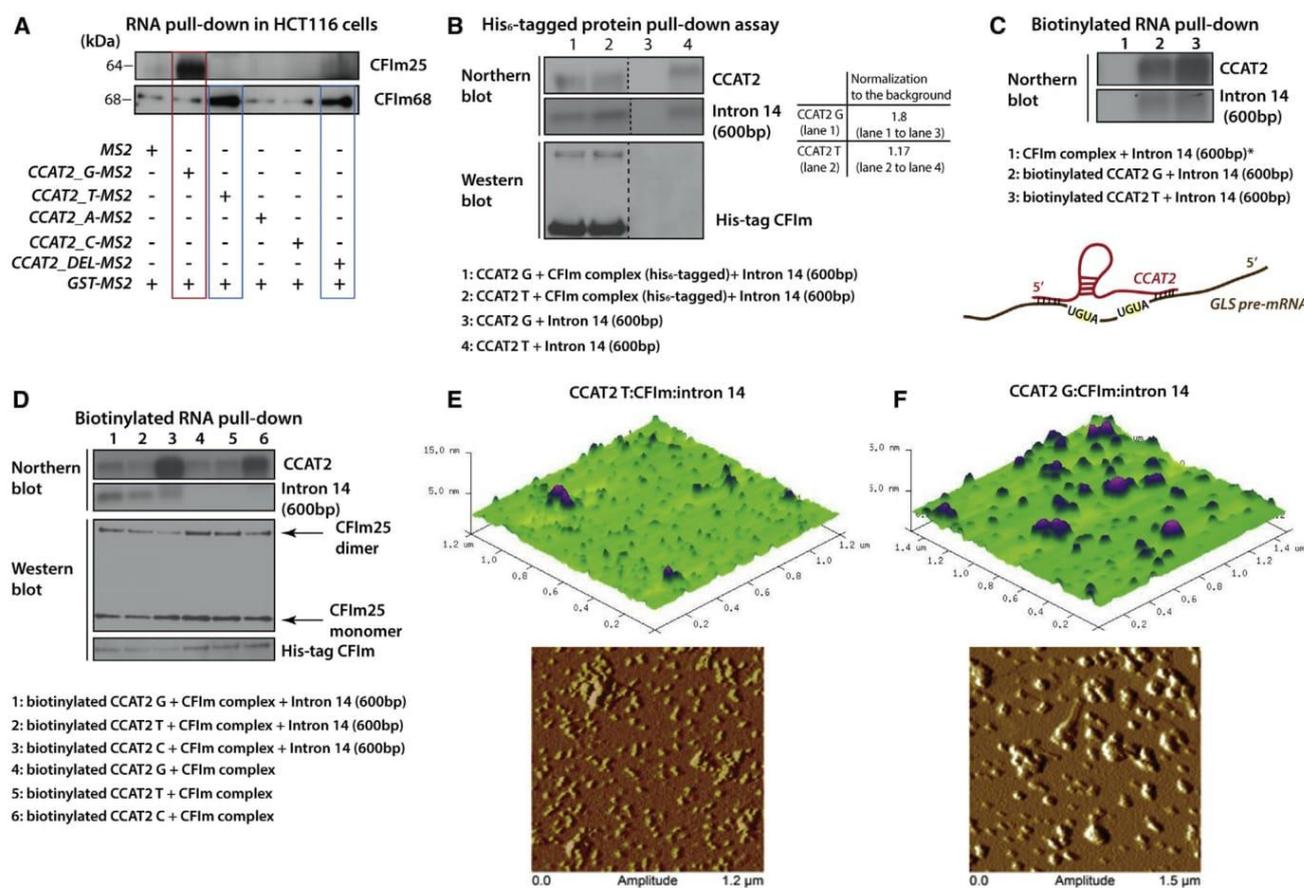
#### Alternative Splicing of *GLS* Is Associated with the Interaction between *CCAT2* and the CFIm Complex

We next aimed to investigate in more detail the mechanism leading to the difference in binding affinity of CFIm25 and CFIm68 to the distinct *CCAT2* alleles. We first scanned the *CCAT2* RNA sequence and identified two CFIm25 binding motifs surrounding the rs6983267, one upstream and the other downstream of the SNP. We performed secondary structure predictions using the RNAfold Webserver (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) and noticed major local structural changes induced near the putative upstream CFIm25 binding sequence, by the single nucleotide variation, especially among the G and T alleles (Figure S4C). Such changes may plausibly translate into distinct tertiary folds that could in principle explain the binding of the G and T alleles with different affinities. To further validate our results and evaluate the contribution of a single nucleotide to the structural changes and binding affinity, we mutated the SNP into an A or C (nucleotides never detected in the human population) and also deleted the 7 nucleotide region encompassing the SNP, which has been reported to have enhancer activity (Tuupainen et al., 2009). We repeated the RNA pull-down assay using different vectors containing the MS2 tags (*CCAT2*-G, -T, -A, -C, and DEL) and analyzed by western blot the proteins bound to RNAs. When CFIm25 antibody was hybridized on the blot, we detected a strong band of approximately 64 kDa unique to the G allele (band corresponds to the CFIm25 dimer; Figure 4A). When CFIm68 antibody was hybridized on the blot, we distinguished bands of approximately 68 kDa for both the T allele and the allele with the deleted region

(Figure 4A). This implied that the T allele preferentially binds the 68 kDa subunit, but not in the region of the SNP. Notably, neither of the two mutated *CCAT2* alleles (A and C) interacted with the CFIm complex, and interestingly, both present secondary structures different from the G and T alleles (Figure S4C).

Additional evidence for the direct interaction between the *CCAT2* G and T alleles with the CFIm complex was provided by a His<sub>6</sub>-tag pull-down assay using heterologously expressed CFIm68:CFIm25 complex (His<sub>6</sub>-tagged CFIm68 subunit; Figure 4B) incubated with in vitro synthesized RNAs. We detected strong affinity of the CFIm complex for the *CCAT2* G allele, followed by moderate binding of the *CCAT2* T allele. We also identified in the pull down the 600 nt long region of the *GLS* pre-mRNA intron 14, containing one type 2 poly(A) site, with affinities for the protein complex corresponding to the two alleles (Figure 4B). This suggested that the intron 14 may also interact with *CCAT2*. To test this, we added a biotin tag to in vitro transcribed *CCAT2* RNAs, combined them with the intron 14 fragment and/or CFIm complex, and pulled down the complex with Streptavidin beads. We confirmed not only that the *CCAT2* G allele preferentially binds CFIm25, but also that *CCAT2* interacts with the intron 14 fragment in a SNP-independent fashion (Figures 4C and S4D). To ensure the specificity of the interaction, we repeated the biotin RNA pull-down assay to include the biotinylated *CCAT2* C allele as a negative control. We also added the whole intron 14, previously used for the RG6 splicing assay, to determine if *CCAT2* can interact with the entire region. We prepared mixes of *CCAT2* (G, T, and C) and the CFIm complex with and without the intron 14 fragment to evaluate how it impacts the interaction between *CCAT2* and CFIm complex. We discovered that in the presence of the intron 14, *CCAT2* G and T alleles displayed increased binding affinity to the CFIm complex, compared to the C allele (Figures 4D and S4E and Table S3). The G and T alleles presented a remarkable specificity for the CFIm25 dimer in the presence of the intron 14 (both the smaller fragment and the whole intron) with enhanced binding to the G allele (Figure S4E and Table S3), supporting the hypothesis that the secondary structure of *CCAT2* influences the interaction with the protein complex. In the absence of the intron, although the specificity of the interaction with the CFIm25 dimer is partly retained, it appeared to be revoked in the case of CFIm complex (Figures 4D and S4E and Table S3). We confirmed that *CCAT2* G and T alleles can bind the whole intron 14 as well and aligned the *CCAT2* genomic sequence with *GLS* genomic sequence to determine the extent of the interaction (Figure S4F). We observed multiple short fragments (13–18 nt) of sequence complementarity spanning the entire *GLS* sequence, present in both introns and exons (Figure S5A). For additional validation, we performed the His<sub>6</sub>-tag pull-down assay using the G, T, and C alleles in the presence and absence of the whole intron 14. As expected, when the intron was included in the mix, we could detect only *CCAT2* G and T alleles in the pull-down lysates (Figure S4G).

Furthermore, the same mixes of *CCAT2* RNA (G/T), CFIm protein complex, and intron 14 RNA were prepared, as well as solutions of individual components, and subjected to atomic force microscopy (AFM) for visualizing the formation of the RNA:protein:RNA quaternary complex (Lyubchenko et al., 2011) (Figures



**Figure 4. The rs6983267 SNP Affects the Interaction of CCAT2 with the CFIm Protein Complex**

(A) Western blot analysis of the proteins pulled down with the MS2-CCAT2 vectors (G, T, A, C, and DEL) showing the presence of CFIm25 for the G allele and CFIm68 for the T allele.

(B) Northern blot analysis showing the presence of CCAT2 and intron 14 (600 bp fragment) in the lysate pulled down with TALON resin (upper panel). Western blot analysis showing the presence of the His<sub>6</sub>-tagged CFIm complex in the lysate pulled down with the TALON resin (lower panel).

(C) Northern blot analysis showing the presence of CCAT2 and intron 14 (600 bp fragment) in the lysate pulled down with Streptavidin beads. Lane 1 marked with the star symbol is identical to lane 3 in Figure S4D. Schematic illustration of the interaction of CCAT2 with GLS pre-mRNA (representation is not at scale).

(D) Northern blot analysis showing the presence of CCAT2 and intron 14 (600 bp fragment) in the lysate pulled down with Streptavidin beads (upper panel). Western blot analysis showing the presence CFIm25, monomer (26 kDa) and dimer (64 kDa), and His<sub>6</sub>-tagged CFIm68 (38 kDa) (lower panel).

(E and F) AFM images of CCAT2:CFIm:intron 14 quaternary complex including either CCAT2 T allele (E) or CCAT2 G allele (F).

See also Figures S4 and S5 and Table S3.

4E and 4F). Particles of substantial size were detected when scanning the mica surface, suggesting the formation of potential CCAT2:CFIm:intron14 complexes, with higher frequency for the G allele compared to the T allele (Figures 4E and 4F), while the individual components (CFIm protein complex, CCAT2 RNA, and intron 14 RNA) appeared to be evidently smaller (Figures S5B–S5D). Moreover, when measuring the diameter of the complexes (G allele complex, 226.706 nm; T allele complex, 182.844 nm), we found that it corresponds approximately to the sum of the diameter of individual components (CCAT2, 81.865 nm; CFIm, 67.999 nm; intron 14, 48.360 nm).

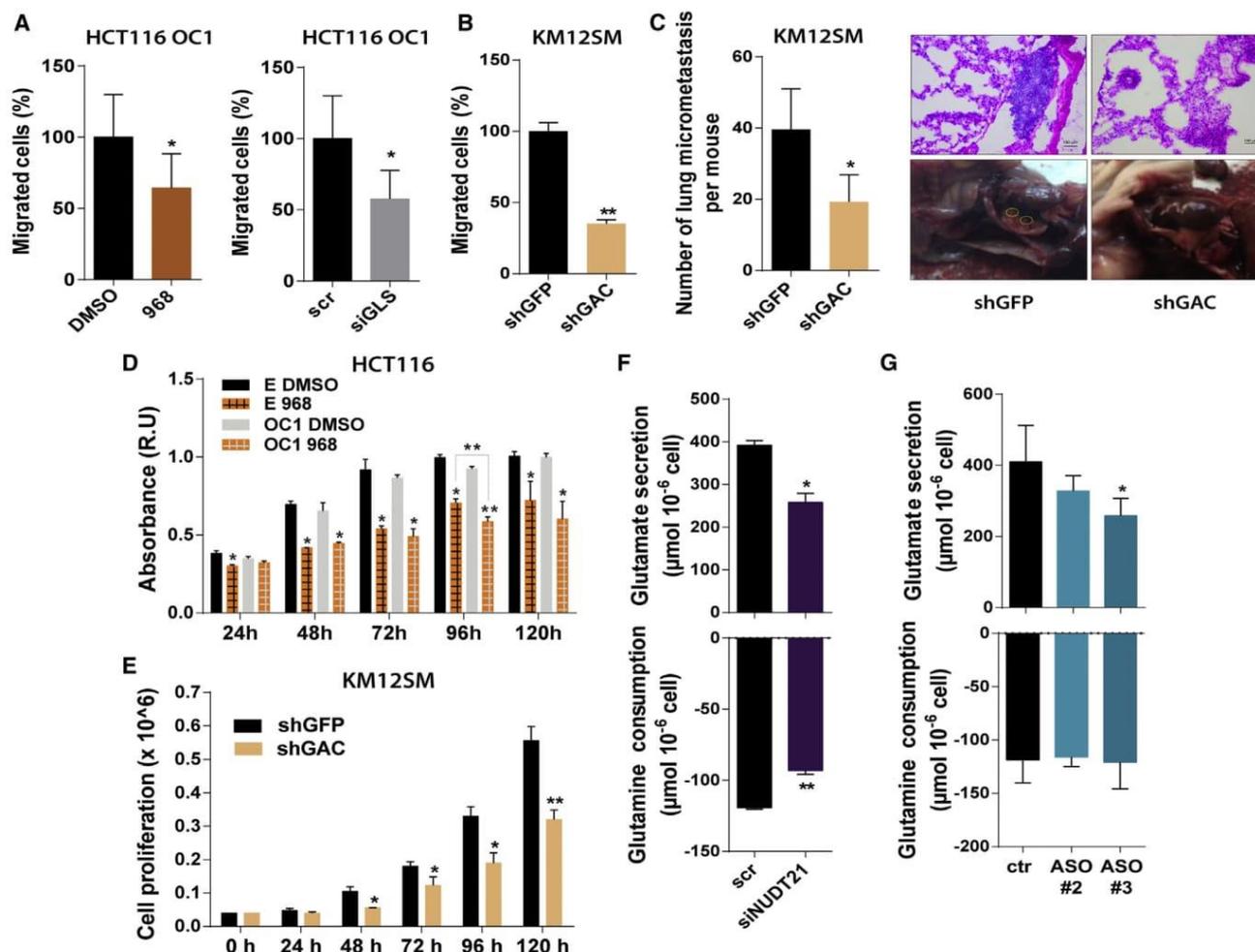
Collectively, these findings suggested that: (1) GLS pre-mRNA impacts the interaction of CCAT2 with CFIm complex; (2) CCAT2 may be acting as a scaffold or assembly platform, promoting the selection of the poly(A) site in intron 14 of GLS pre-mRNA by

directly binding both the pre-mRNA and the CFIm complex; and (3) the formation of the RNA:RNA:protein complex is dependent on the rs6983267 SNP and the secondary structure of CCAT2.

As an additional layer of regulation, we discovered that MYC is a transcription factor for *NUDT21* (data available from authors upon request).

### GAC Promotes Metastases and Cell Proliferation

We proceeded with evaluating the contribution of GAC to aggressiveness of CRC. Considering that we have previously demonstrated that CCAT2 promotes metastases (Ling et al., 2013b), we first assessed the in vitro migration potential of HCT116 CCAT2-overexpressing cells treated with either the allosteric inhibitor 968 or siRNA for GLS (Katt et al., 2012). We



**Figure 5. GLS Promotes In Vivo Metastases and In Vitro Cell Proliferation and Migration**

(A) Migration of HCT116 OC1 cells (GG genotype) treated with the inhibitor 968 and DMSO (left panel) and with siGLS and scrambled siRNA (right panel). (B) Migration of KM12SM cells (GT genotype) with stable downregulation of GAC. KM12SM shGFP cells represent the control cells. (C) Number of the lung micrometastases for two groups (shGFP, 4 mice; shGAC, 4 mice) assessed by IHC (left panel). IHC images showing micrometastases in the three groups and images showing the presence or absence of lung metastases for mice injected in the tail vein with KM12SM shGFP and shGAC cells, respectively (right panel). (D and E) Growth curves for HCT116 OC1 and control cells (GG genotype) treated with DMSO (control) or the GLS allosteric inhibitor 968 (10 µM) (D) and KM12SM cells (GT genotype) with stable downregulation of GAC (E). (F and G) Glutamine and glutamate concentration in the media relative to the empty well 24 hr after seeding HCT116 CCAT2-overexpressing cells (OC1, GG genotype) transfected with siRNA against *NUDT21* and scrambled (F) and ASOs for inhibiting the binding sites of CFIm25 (G). Results are presented as normalized mean values ± SD. See also Figure S6 and Table S4.

observed a reduction by half of the migration in the cells where *GLS* was either inhibited or downregulated (Figure 5A). We next assessed the individual contribution of GAC to the migration, in KM12SM cells with stable downregulation of the isoform (Figure S6A), and found that 60% fewer cells migrated when GAC expression was reduced (Figure 5B). We then injected the KM12SM stable clones in the tail vein of nude mice, sacrificed the mice 8 weeks after injections, and evaluated the in-lung macro- and micrometastases. Supporting our in vitro results, the incidence of metastases to the lung was 50% higher in the mice injected with the control cells (shGFP) compared to the shGAC group (Figure 5C). Moreover, when

we assessed the proliferation of HCT116 control and *CCAT2*-overexpressing cells treated with the *GLS* inhibitor 968, we found that cells overexpressing *CCAT2* were more sensitive to *GLS* (GAC in this case) inhibition, implying that cells with high *CCAT2* expression are dependent on GAC for survival (Figure 5D) (Katt et al., 2012; Wang et al., 2010a). Similar results were obtained when using KM12SM cells with GAC downregulation compared to control cells in vitro (Figure 5E). Additional confirmation of the higher dependency of the *CCAT2* G allele on GAC was provided by the colony formation assay for NIH 3T3 cells transfected with the *CCAT2*-overexpressing vectors (Figures S6B and S6C).

We next sought to determine if the shift in *GLS* isoform expression and CFIm25 is responsible for the marked metabolic changes observed in cells overexpressing *CCAT2*. We therefore modulated the expression of the isoforms using the ASOs and downregulated *NUDT21* in HCT116 *CCAT2*-overexpressing cell line (OC1) and measured the extracellular lactate concentration and glutamine metabolism. We observed that downregulation of *NUDT21* significantly decreased lactate and glutamate secretion and glutamine consumption similarly to the effects of *CCAT2* overexpression in the same cell line (Figures S6D and 5F). However, these metabolic changes appeared to only partly mirror the metabolic effects of *CCAT2* overexpression, suggesting there are additional layers of regulations independent of *NUDT21*, most likely through the bona fide cancer metabolism modulator and *CCAT2* target, *MYC*. On the other hand, the switch in *GLS* isoforms reflected in lower secretion of glutamate with the glutamine consumption remaining relatively constant but had a modest effect on the extracellular levels of lactate (Figures 5G and S6E). These findings suggested that *GLS* is not the only contributor to *CCAT2*-induced metabolic profile. We hypothesized that other metabolic targets might be regulated by *CCAT2* via the same mechanism; thus, we performed Affymetrix HTA 2.0 array to compare the whole transcriptome splicing pattern in HCT116 *CCAT2*-overexpressing G and T allele cells. Pathway analysis revealed that several genes associated with two major metabolic pathways, “Metabolism of Carbohydrates” and “Fructose and Mannose Metabolism,” are significantly spliced between the *CCAT2* G and T alleles (Table S4). Moreover, validating our previous results, we identified a negative splicing index (SI = -1.19) for *GLS* for the exclusion junction connecting exon 14 and exon 15, suggesting higher signal for the exclusion junction in the T allele-expressing cells. This translates into less GAC being spliced in the T allele cells compared to the G allele cells, supportive of our findings (Figure S6F).

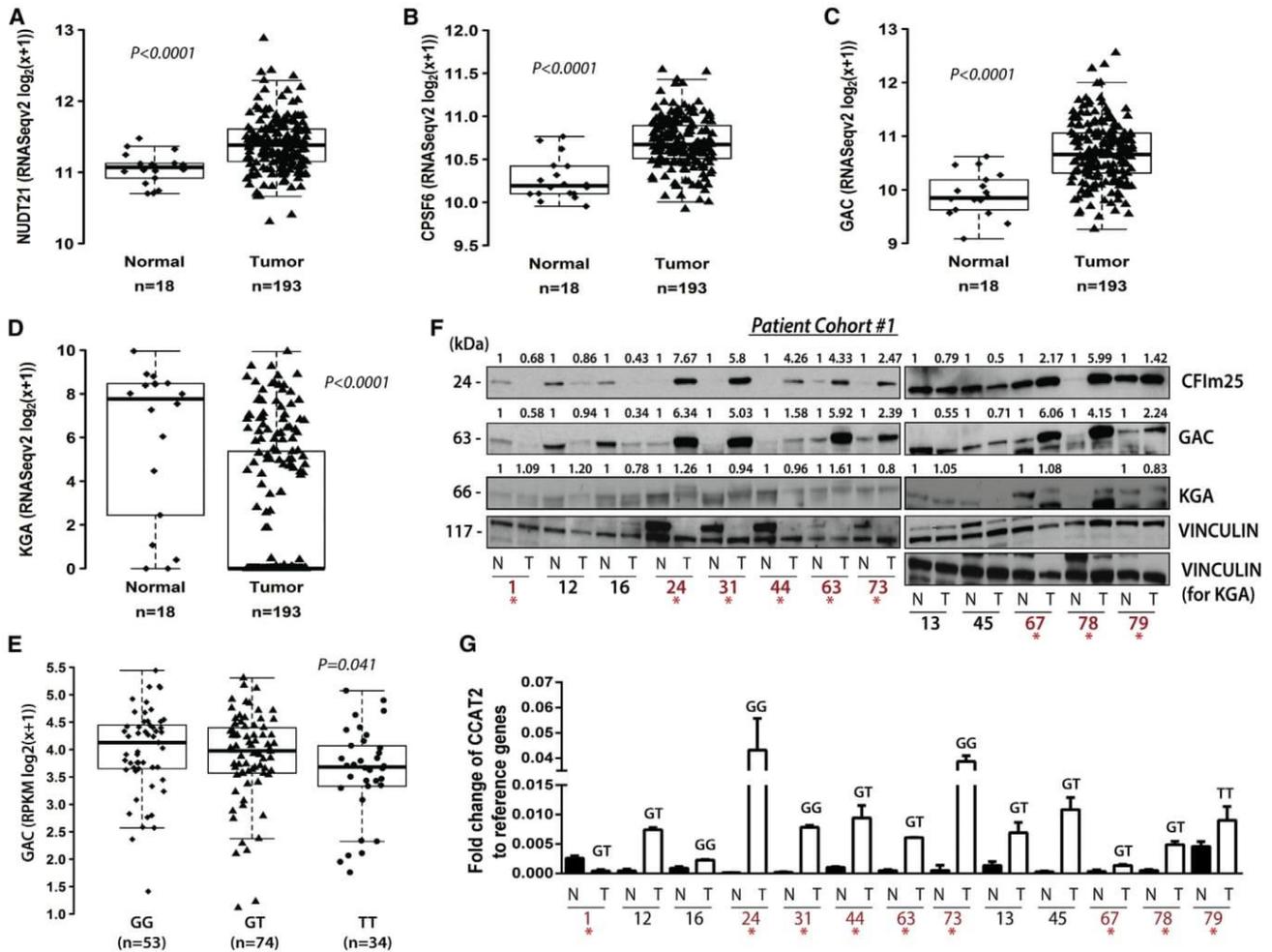
Collectively, these data demonstrate that GAC adds to the CRC phenotype; however, it is not solely responsible for the metabolic phenotype observed in cells overexpressing *CCAT2*.

### CCAT2-CFIm-GLS Regulation Axis in CRC Tumors

We continued with evaluating the expression pattern of *CCAT2*, *GLS*, *NUDT21*, and *CPSF6* in colon tumors by analyzing the publicly available TCGA database of colon cancer (<http://cancergenome.nih.gov/>). We first compared 18 normal samples to 193 tumor samples and identified an enrichment of the GAC isoform, as well as *NUDT21* and *CPSF6* in tumor tissue, whereas *KGA* showed the opposite pattern (Figures 6A–6D). We further analyzed the associations between *CCAT2*, *NUDT21*, *CPSF6*, *GAC*, and *KGA* in the TCGA dataset of colon cancer samples (Cancer Genome Atlas Network, 2012) and detected direct correlations of *CCAT2*, *CPSF6*, and *NUDT21* with *GAC* ( $r_s = 0.26$ ,  $p = 0.0006$ ;  $r_s = 0.68$ ,  $p < 0.0001$ ; and  $r_s = 0.72$ ,  $p < 0.0001$ , respectively; Figure S7A) and inverse correlations between *CCAT2*, *CPSF6*, *NUDT21*, and *KGA* ( $r_s = -0.17$ ,  $p = 0.0271$ ;  $r_s = -0.47$ ,  $p < 0.0001$ ; and  $r_s = -0.60$ ,  $p < 0.0001$ ; Figure S7A), as well as between *KGA* and *GAC* ( $r_s = -0.590$ ,  $p < 0.0001$ ) (Figure S7A). We also obtained significant direct associations of *CCAT2* with *CPSF6* and *NUDT21* ( $r_s = 0.20$ ,  $p = 0.0082$  and  $r_s = 0.26$ ,  $p =$

0.0007; Figure S7A). We did not find any significant correlations between the expression of *MYC* and *NUDT21*, *GAC*, or *KGA*. Since our in vitro findings advanced the concept that *CCAT2* G allele is promoting the expression of *GAC*, we compared the levels of *GAC* and *KGA* between patients having GG, GT, and TT genotypes. We observed a significant association with the genotypes for *GAC*, having the highest expression in patients with GG genotype, but not for *KGA* (Figures 6E and S7B). Similarly, we did not find any association of *NUDT21* or *MYC* with the genotypes (Figures S7C and S7D). These results suggested the molecular mechanism uncovered in vitro is present in CRC patients. Moreover, we analyzed the correlation between the expression of the two isoforms and the overall survival of patients from the TCGA dataset and observed a significant association of high *GAC* expression combined with low *KGA* expression with shorter overall survival (Figure S7E). This suggested that the *GAC* isoform may accelerate the progression of cancer. In addition, we interrogated the TCGA colon cancer dataset for genes that significantly correlate with the lncRNA and performed gene set enrichment analysis (GSEA) and ingenuity pathway analysis (IPA) (QIAGEN) to identify the *CCAT2* gene signatures. Various metabolically relevant pathways were found significantly associated with *CCAT2* expression by both analyses (Figures 7A and 7B). We further inquired as to whether the genes that were found significantly correlated with *CCAT2* expression were also associated with the genotypes of rs6983267 SNP. We repeated the GSEA and screened for pathways that correlated with either of the genotypes (FDR  $q$  value  $< 0.25$  and nominal  $p$  value  $< 0.05$ ). We identified various metabolic and RNA processing pathways that were associated with certain genotypes (Figure 7C). Interestingly, several of the pathways were related to the ones identified by the Affymetrix HTA 2.0 array (highlighted in green in Figure 7C), suggesting that *CCAT2* may modulate cellular metabolism in CRC patients through a network of metabolic genes regulated most likely via the same mechanism of alternative splicing (see Table S4).

To assess the protein expression of CFIm68, CFIm25, *GAC*, and *KGA*, we performed western blot analysis on paired normal mucosae and CRC samples (patient cohort #1) and identified the same high protein levels of CFIm25 and *GAC* in tumor tissue compared to normal mucosae for 61.5% (8/13) of the pairs (Figure 6F). As for the *KGA* isoform, protein levels were mostly lower in tumor samples or comparable to the levels in normal samples (Figure 6F). The 68 kDa subunit of the CFIm complex was either very low or not expressed in approximately half of the paired samples (7/13), while in the rest of the pairs it was clearly overexpressed in tumors (Figure S7F). We also measured *CCAT2* expression by qRT-PCR in the same set of samples and confirmed a positive correlation between *CCAT2*, CFIm25, and *GAC* for 69.2% of the samples (Figures 6F, 6G, S7G, and S7H). Moreover, the samples that had elevated levels of *CCAT2* displayed a high *GAC/KGA* ratio ( $GAC/KGA = 2-5.7$ ) (Figure S7I). We repeated the measurements in a second set of paired normal mucosae and CRC samples (patient cohort #2) and obtained similar results (Figures S7J and S7K). Both CFIm25 and *GAC* proteins were overexpressed in 60% (3/5) of tumor samples. In all samples, elevated protein levels of CFIm25 and *GAC* matched the increased RNA levels of



**Figure 6. *CCAT2*, *NUDT21*, *CPSF6*, and *GLS* Expression Pattern in TCGA Dataset and CRC Patient Samples**

(A–D) Analysis of *NUDT21* (A), *CPSF6* (B), *GAC* (C), and *KGA* (D) mRNA expression in TCGA RNA-seq colon cancer sample set.

(E) Association of *GAC* mRNA expression with the genotypes (GG, GT, and TT) of the rs6983267 SNP for CRC patients (TCGA RNA-seq dataset).

(F) Western blot analysis of CFIm25, *GAC*, and *KGA* expression in paired CRC samples (patient cohort #1).

(G) qRT-PCR analysis for *CCAT2* in the same paired CRC samples (patient cohort #1).

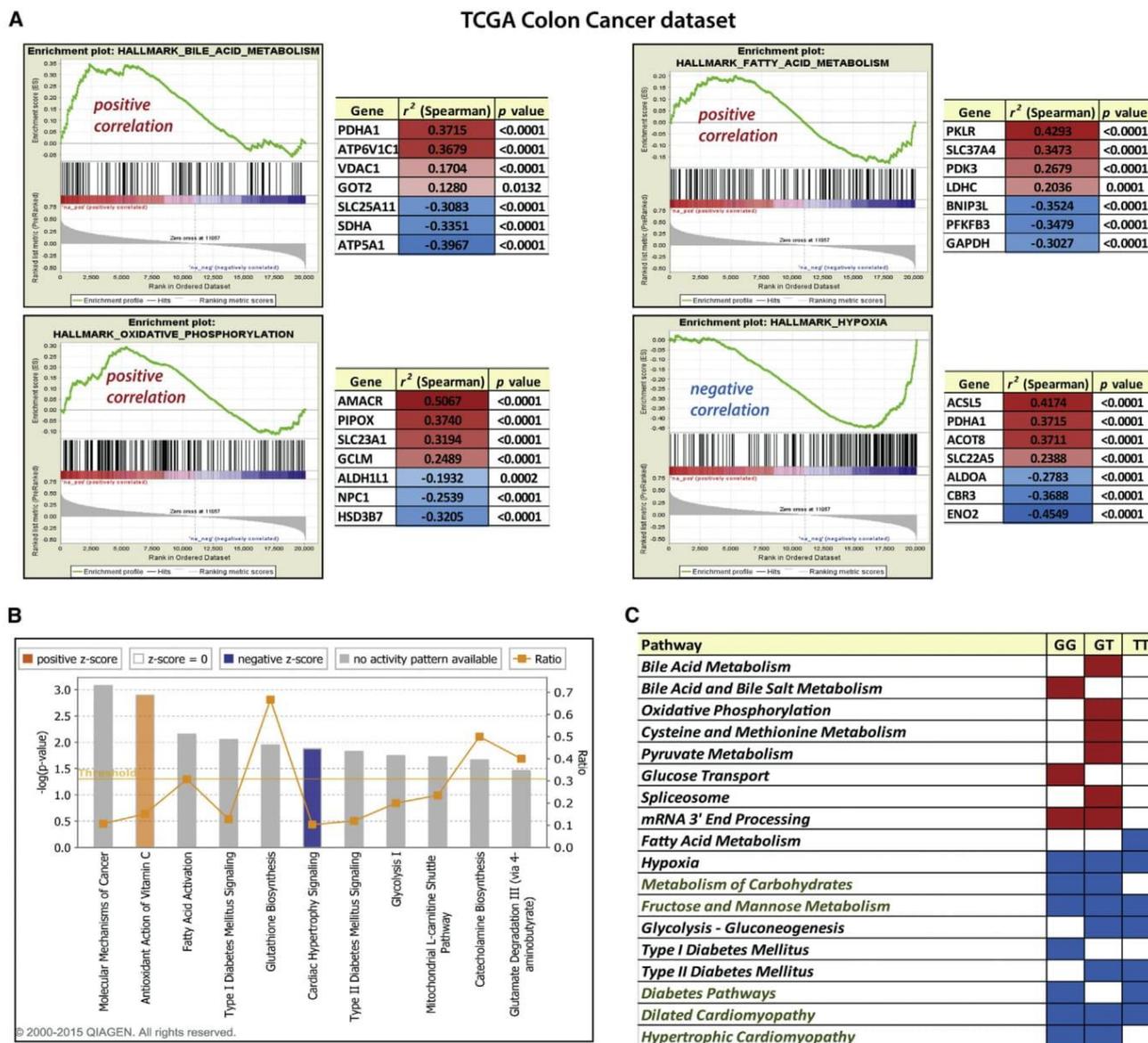
Results are presented as normalized mean values  $\pm$  SD. See also Figure S7.

*CCAT2*. Additionally, we have genotyped the tumors from patient cohort #1 and found that the association of GG genotype with higher *GAC* and CFIm25 protein expression (samples marked in red) was consistent for 75% of patients (3/4), while for the GT genotype the association was present in only 62.5% of patients (5/8) (Figure 6G). No conclusion can be drawn for the TT genotype due to the limited representation of the genotype in this cohort (one patient). Altogether, this mechanism of *GLS* regulation was detected in the majority of the analyzed CRC cases (61%, 11/18).

## DISCUSSION

Our study demonstrates that the rs6983267 SNP (G/T) induces changes in the secondary structure of the lncRNA, *CCAT2*, initiating a domino effect mechanism, which leads to allele-specific

reprogramming of cellular energy metabolism. The consequence of the allele-specific interaction between *CCAT2*, CFIm, and *GLS* pre-mRNA appears to be the selection of the poly(A) site within intron 14 of *GLS*, resulting in the preferential splicing to the *GAC* isoform, the more catalytically active of the two *GLS* isoforms (Cassago et al., 2012). Although a recent study has described the negative regulation of *GLS* by CFIm25 in glioblastoma via 3' UTR processing mechanisms (Masamha et al., 2014), suggesting a tumor-suppressive role for CFIm25, in our model, neither of the *GLS* isoforms is subjected to 3' UTR shortening (data not shown). In the context of these findings and considering that *CCAT2* is not expressed in glioblastoma (data not shown), our results reveal an intriguing aspect of lncRNA mechanism of action, namely the ability of an lncRNA to alter the function of the partner RNA-binding protein/complex. The enrichment on metabolites related to TCA cycle that we



**Figure 7. CCAT2 Gene Signature in Colon Cancer Patients—TCGA Dataset**

(A and B) Genes associated with CCAT2 were analyzed by gene set enrichment analysis (GSEA) (A) and ingenuity pathway analysis (IPA, QIAGEN) (B). Relevant examples for each analysis are presented in (A) and (B).

(C) Table containing the pathways significantly associated with CCAT2 expression and rs6983267 genotype (FDR q value < 0.25 and nominal p value < 0.05). Pathways that were positively correlated are marked with red, and the ones that are negatively correlated are marked with blue. Highlighted in green are the pathways found common between GSEA analysis and the Affymetrix HTA 2.0 pathways analysis.

See also Figure S7 and Table S4.

observed for tumors derived from HCT116 cells overexpressing CCAT2 the G allele are supported by Kaldma and colleagues' findings describing that CRC tumors are not purely glycolytic, but rather dependent on OXPHOS for ATP production (Kaldma et al., 2014).

The aberrant expression of GLS has been reported in many types of cancer, including CRC (Huang et al., 2014) Furthermore, various studies have described GAC as the more abundant isoform in lung adenocarcinoma, head and neck squamous

cell carcinoma, kidney renal clear cell carcinoma, and AML (Jacque et al., 2015; van den Heuvel et al., 2012; Xia et al., 2014). Glutamine metabolism has also been associated with genomic instability (Jeong et al., 2013), commonly encountered in CRC and previously shown to be promoted by CCAT2 (Ling et al., 2013b), and appears to be endorsed by the G risk allele. However, it must be stated that CCAT2 is modulating energy metabolism in a general fashion via MYC and in an allele-specific manner via GLS and other metabolic enzymes and/or

metabolites, whose expression is finely regulated by the interaction of *CCAT2* with CFIm. Although the differences in regulation between the G and T alleles may not be impressive, the variation in expression of multiple enzymes/metabolites may have an additive effect towards a clear phenotypical change. The complex mechanism presented in this manuscript encompassing lncRNA, protein complexes, oncogenes, and transcription factors opens several windows for targeted therapy. The metabolic enzyme *GLS* is already considered a therapeutic target for cancer (Vander Heiden, 2011); however, our work introduces the opportunity of targeting the cancer-specific GAC isoform in particular.

Lastly, our study reveals the complexity and refinement of the interaction networks among the alleles of a non-coding RNA, the components of a protein complex, and the splicing isoforms of a metabolic enzyme that contribute to the malignant transformation and progression of CRC.

## EXPERIMENTAL PROCEDURES

Full details of the [Experimental Procedures](#) are presented in the [Supplemental Experimental Procedures](#). Primer sequences and information regarding the antibodies used in this study can be found in [Tables S5](#) and [S6](#).

### Patient Samples

18 paired samples, normal colon mucosa and colon tumor, were used in this study. The samples were obtained from two different sources: The Ruder Boskovic Institute, Croatia (15 paired samples) and University of Ferrara, Italy (5 paired samples). Tissue samples were obtained from fresh surgical specimens frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . All the samples were obtained with the patients' informed consent and under the approval and supervision of the institutional review boards. The samples were histologically confirmed prior to use.

### In Vivo Models and Tissue Processing

70 male athymic nude mice were purchased from the National Cancer Institute, Frederick Cancer Research and Development Center (Frederick, MD) and were cared for according to guidelines set forth by the American Association for Accreditation of Laboratory Animal Care and the U.S. Public Health Service policy on Human Care and Use of Laboratory Animals. All mouse studies were approved and supervised by the MD Anderson Cancer Center Institutional Animal Care and Use Committee.

### Glucose Uptake Assay

Cells were plated in 96-well plates (25,000 cells/well) 16 hr before performing the assay. The medium was removed and cells were washed twice with PBS. To the wells containing the blanks, 50  $\mu\text{l}$  of PBS was added, while for the wells with the samples 50  $\mu\text{l}$  of 2-NBDG (100  $\mu\text{M}$ ) (Sigma) was added and the mixture was incubated for 10 min at  $37^{\circ}\text{C}$  and 5%  $\text{CO}_2$ . After incubation, cells were washed twice with ice-cold PBS to stop the reaction, and 200  $\mu\text{l}$  PBS was added to each well. Fluorescence measurements were performed at 485/520 nm with the PHERAstar FS (BMG Labtech).

### Lactate Production Assay

To measure lactate production, cells that were 80% confluent were replenished with fresh medium. Aliquots of the medium were removed at the indicated time points (24 or 48 hr) for measurement of lactate using an Accutrend lactate analyzer (Roche). At each time point, cell numbers were also counted for normalization of lactate generation.

### Intracellular Glutamate Assay

The glutamate concentration in cell lysates was measured using the Glutamate Colorimetric Assay Kit (Biovision) following the manufacturer's protocol.

Briefly,  $1 \times 10^6$  cells per tested sample were homogenized in 100  $\mu\text{l}$  of assay buffer and centrifuged to remove insoluble material. 100  $\mu\text{l}$  of reaction mix was added to the supernatant, standards, and background control samples, and after 30 min incubation at  $37^{\circ}\text{C}$ , absorbance was measured at 450 nm with a SpectraMax Plus384 MicroPlate Reader (Molecular Devices). The experiment was performed in quadruplicate.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2016.01.015>.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: R.S.R., L.E.V., W.L., J.F.O., A.F.F., M.S.R., M.I., G.B., M.F.F., S.H., W.R.W., A.L.B.A., S.M.G.D., and G.A.C. Performed the experiments: R.S.R., L.E.V., W.L., J.F.O., C.R.-A., D.A., B.P., A.T., Y.C., K.V.R., S.C., M.S., Y.A., L.H.C., G.Y.H., P.M., M.S.R., T.C.I., L.V., and H.L. Analyzed the data: R.S.R., L.E.V., W.L., J.F.O., C.I., C.R.-A., D.A., A.T., A.F.F., L.V., G.K., J.A.B., M.R., A.L.B.A., S.M.G.D., and G.A.C. Contributed reagents/materials/analysis tools: L.H., S.K., R.G., G.L., S.Y.L., R.C.B., P.H., M.G.R., I.B.-N., and G.L.-B. Wrote the first draft of the manuscript: R.S.R., S.M.G.D., A.L.B.A., and G.A.C. Contributed to the writing of the manuscript: R.S.R., S.M.G.D., A.L.B.A., and G.A.C. Agree with manuscript results and conclusions: all authors. Critical revision of the manuscript for important intellectual content: all authors. Statistical analysis: C.I., L.H., G.K., M.R., and H.L. Study supervision: S.M.G.D. and G.A.C.

## ACKNOWLEDGMENTS

G.A.C. is The Alan M. Gewirtz Leukemia & Lymphoma Society Scholar. Work in G.A.C.'s laboratory is supported in part by the NIH/NCI grants 1UH2TR00943-01 and 1 R01 CA182905-01, the UT MD Anderson Cancer Center SPORE in Melanoma grant from NCI (P50 CA093459), Aim at Melanoma Foundation and the Miriam and Jim Mulva research funds, the Brain SPORE (2P50CA127001), the Center for Radiation Oncology Research Project, the Center for Cancer Epigenetics Pilot project, a 2014 Knowledge GAP MDACC grant, a CLL Moonshot pilot project, the UT MD Anderson Cancer Center Duncan Family Institute for Cancer Prevention and Risk Assessment, a SINF grant in colon cancer, the Laura and John Arnold Foundation, the RGK Foundation, and the Estate of C.G. Johnson, Jr. I.B.-N. was financed by a grant entitled Non-Invasive Intelligent Systems for Colorectal Cancer Diagnosis and Prognosis Based on circulating miRNAs Integrated in the Clinical Workflow – INTELCCOR. S.M.G.D., A.L.B.A., and D.A. are supported by the São Paulo Research Foundation FAPESP under grants 2014/15968-3, 2014/20673-2, and 2014/17820-3, respectively. W.L. was partly supported by grants from The University of Texas MD Anderson Cancer Center Sheikh Ahmed Bin Zayed Al Nahyan Center for Pancreatic Cancer Research. J.A.B. was supported by the Cancer Center Support Grant (P30 CA016672), and the HP imaging program of the Small Animal Facility (SAIF) was supported by the Cancer Prevention and Research Institutes of Texas grant RP-101243P5. H.L. was supported by NIH/NCI grant R01CA175486, a grant (RP140462) from the Cancer Prevention and Research Institute of Texas, and the R. Lee Clark Fellow Award from The Jeanne F. Shelby Scholarship Fund. I.B.-N. was financed by a Fulbright fellowship and by a grant entitled Non-Invasive Intelligent Systems for Colorectal Cancer Diagnosis and Prognosis Based on circulating miRNAs Integrated in the Clinical Workflow – INTELCCOR. We would like to thank Dr. Riccardo Fodde (Erasmus Medical Center) for the scientific support and advice, Dr. Sylvie Doublet for the generous gift of CFIm25 and CFIm68 plasmids, Dr. Riccardo Spizzo for generating the HCT116 *CCAT2* stable clones, and Dr. Thomas A. Cooper for the generous gift of RG6 plasmid. We would also like to thank the IM Bioscope II-UT core facility and Dr. Ana Maria Zaske for performing the AFM imaging. Additionally, we thank the members of Flow Cytometry & Cellular Imaging Core Facility (Department of Leukemia) and Dr. Jared Burks for performing the VECTRA imaging and analysis and

Drs. Liuqing Yang (MDACC), Yibin Zhou (IBT), Clifford Stephan (IBT), Shawn Bratton (MDACC), Charles V. Kingsley (MDACC), Jorge Delacerda (MDACC), Iva Maestri (University of Ferrara), and Linda Ulazzi (University of Ferrara) for the technical support. We also would like to thank the Biological Imaging Facility and the Protein Purification and Bioassay Laboratories (LNBio/CNPem/Brazil) for access to the equipment.

Received: April 3, 2015

Revised: October 23, 2015

Accepted: January 8, 2016

Published: February 4, 2016

## REFERENCES

- Bester, A.C., Roniger, M., Oren, Y.S., Im, M.M., Sami, D., Chaoat, M., Bensimon, A., Zamir, G., Shewach, D.S., and Kerem, B. (2011). Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* **145**, 435–446.
- Boroughs, L.K., and DeBerardinis, R.J. (2015). Metabolic pathways promoting cancer cell survival and growth. *Nat. Cell Biol.* **17**, 351–359.
- Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337.
- Carroll, P.A., Diolaiti, D., McFerrin, L., Gu, H., Djukovic, D., Du, J., Cheng, P.F., Anderson, S., Ulrich, M., Hurley, J.B., et al. (2015). Deregulated Myc requires MondoA/Mlx for metabolic reprogramming and tumorigenesis. *Cancer Cell* **27**, 271–285.
- Cassago, A., Ferreira, A.P., Ferreira, I.M., Fornezari, C., Gomes, E.R., Greene, K.S., Pereira, H.M., Garratt, R.C., Dias, S.M., and Ambrosio, A.L. (2012). Mitochondrial localization and structure-based phosphate activation mechanism of Glutaminase C with implications for cancer metabolism. *Proc. Natl. Acad. Sci. USA* **109**, 1092–1097.
- Chen, J.Q., and Russo, J. (2012). Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. *Biochim. Biophys. Acta* **1826**, 370–384.
- Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* **14**, 496–506.
- Gao, P., Tchernyshyov, I., Chang, T.C., Lee, Y.S., Kita, K., Ochi, T., Zeller, K.I., De Marzo, A.M., Van Eyk, J.E., Mendell, J.T., and Dang, C.V. (2009). c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458**, 762–765.
- Huang, F., Zhang, Q., Ma, H., Lv, Q., and Zhang, T. (2014). Expression of glutaminase is upregulated in colorectal cancer and of clinical significance. *Int. J. Clin. Exp. Pathol.* **7**, 1093–1100.
- Jacque, N., Ronchetti, A.M., Larrue, C., Meunier, G., Birsens, R., Willems, L., Saland, E., Decroocq, J., Thiago, T.T., Lambert, M., et al. (2015). Targeting glutaminolysis has antileukemic activity in acute myeloid leukemia and synergizes with BCL-2 inhibition. *Blood* **126**, 1346–1356.
- Jeong, S.M., Xiao, C., Finley, L.W., Lahusen, T., Souza, A.L., Pierce, K., Li, Y.H., Wang, X., Laurent, G., German, N.J., et al. (2013). SIRT4 has tumor-suppressive activity and regulates the cellular metabolic response to DNA damage by inhibiting mitochondrial glutamine metabolism. *Cancer Cell* **23**, 450–463.
- Kaldma, A., Klepinin, A., Chekulayev, V., Mado, K., Shevchuk, I., Timohhina, N., Tepp, K., Kandashvili, M., Varikmaa, M., Koit, A., et al. (2014). An in situ study of bioenergetic properties of human colorectal cancer: the regulation of mitochondrial respiration and distribution of flux control among the components of ATP synthasome. *Int. J. Biochem. Cell Biol.* **55**, 171–186.
- Katt, W.P., Ramachandran, S., Erickson, J.W., and Cerione, R.A. (2012). Dibenzophenanthridines as inhibitors of glutaminase C and cancer cell proliferation. *Mol. Cancer Ther.* **11**, 1269–1278.
- Le, A., Lane, A.N., Hamaker, M., Bose, S., Gouw, A., Barbi, J., Tsukamoto, T., Rojas, C.J., Slusher, B.S., Zhang, H., et al. (2012). Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. *Cell Metab.* **15**, 110–121.
- Ling, H., Fabbri, M., and Calin, G.A. (2013a). MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat. Rev. Drug Discov.* **12**, 847–865.
- Ling, H., Spizzo, R., Atlasi, Y., Nicoloso, M., Shimizu, M., Redis, R.S., Nishida, N., Gafà, R., Song, J., Guo, Z., et al. (2013b). CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res.* **23**, 1446–1461.
- Lutz, C.S., and Moreira, A. (2011). Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip. Rev. RNA* **2**, 23–31.
- Lyubchenko, Y.L., Shlyakhtenko, L.S., and Ando, T. (2011). Imaging of nucleic acids with atomic force microscopy. *Methods* **54**, 274–283.
- Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.B., Li, W., and Wagner, E.J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* **510**, 412–416.
- Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159.
- Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. (2006). An interaction between U2AF 65 and CF Im links the splicing and 3' end processing machineries. *EMBO J.* **25**, 4854–4864.
- Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N., and Hirose, T. (2012). Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* **31**, 4020–4034.
- Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., and Gotoh, O. (2006). Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns. *Bioinformatics* **22**, 1211–1216.
- Orengo, J.P., Bundman, D., and Cooper, T.A. (2006). A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res.* **34**, e148.
- Rathore, M.G., Saumet, A., Rossi, J.F., de Bettignies, C., Tempé, D., Lecellier, C.H., and Villalba, M. (2012). The NF- $\kappa$ B member p65 controls glutamine metabolism through miR-23a. *Int. J. Biochem. Cell Biol.* **44**, 1448–1456.
- Redis, R.S., Sieuwerts, A.M., Look, M.P., Tudoran, O., Ivan, C., Spizzo, R., Zhang, X., de Weerd, V., Shimizu, M., Ling, H., et al. (2013). CCAT2, a novel long non-coding RNA in breast cancer: expression study and clinical correlations. *Oncotarget* **4**, 1748–1762.
- Stine, Z.E., Walton, Z.E., Altman, B.J., Hsieh, A.L., and Dang, C.V. (2015). MYC, Metabolism, and Cancer. *Cancer Discov.* **5**, 1024–1039.
- Tian, B., Pan, Z., and Lee, J.Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res.* **17**, 156–165.
- Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W., et al.; COGIC Consortium (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39**, 984–988.
- Tuupainen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., Kivioja, T., Björklund, M., Wei, G., Yan, J., Niittymäki, I., et al. (2009). The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.* **41**, 885–890.
- van den Heuvel, A.P., Jing, J., Wooster, R.F., and Bachman, K.E. (2012). Analysis of glutamine dependency in non-small cell lung cancer: GLS1 splice variant GAC is essential for cancer cell growth. *Cancer Biol. Ther.* **13**, 1185–1194.
- Vander Heiden, M.G. (2011). Targeting cancer metabolism: a therapeutic window opens. *Nat. Rev. Drug Discov.* **10**, 671–684.
- Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., Chen, N., Sun, F., and Fan, Q. (2010a). CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.* **38**, 5366–5383.
- Wang, J.B., Erickson, J.W., Fuji, R., Ramachandran, S., Gao, P., Dinavahi, R., Wilson, K.F., Ambrosio, A.L., Dias, S.M., Dang, C.V., and Cerione, R.A.

- (2010b). Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. *Cancer Cell* 18, 207–219.
- Warburg, O., Posener, K., and Negelein, E. (1924). Über den Stoffwechsel der Carcinomzelle. *Biochem. Zeitschr* 152, 309–344.
- Ward, P.S., and Thompson, C.B. (2012). Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell* 21, 297–308.
- Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23, 1494–1504.
- Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., and Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* 5, 5274.
- Yang, Q., Gilmartin, G.M., and Doublié, S. (2011). The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5' capping and splicing. *RNA Biol.* 8, 748–753.
- Yang, F., Zhang, H., Mei, Y., and Wu, M. (2014). Reciprocal regulation of HIF-1 $\alpha$  and lincRNA-p21 modulates the Warburg effect. *Mol. Cell* 53, 88–100.
- Yoon, J.H., Srikantan, S., and Gorospe, M. (2012). MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. *Methods* 58, 81–87.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182–185.

**ANEXO 2**

**Pasquali, C. C.\*, Islam, Z.\*, Adamoski, D.\***, Ferreira, I. M., Righeto, R. D., Bettini, J., ... Ambrosio, A. L. B. (2017). The origin and evolution of human glutaminases and their atypical C-terminal ankyrin repeats. *Journal of Biological Chemistry*, 292(27). <https://doi.org/10.1074/jbc.M117.787291>

**\*Contribuição como co-primeiro autor.**

A política editorial do *Journal of Biological Chemistry* permite a reprodução do texto em teses e dissertações (<http://www.jbc.org/site/misc/edpolicy.xhtml#copyright>).

**For noncommercial use:**

Other parties are welcome to copy, distribute, transmit and adapt the work at no cost and without permission for noncommercial use as long as they attribute the work to the original source using the citation above.

Examples of noncommercial use include:

- Reproducing a figure for educational purposes, such as schoolwork or lecture presentations.
- Appending a reprinted article to a Ph.D. dissertation.



# The origin and evolution of human glutaminases and their atypical C-terminal ankyrin repeats

Received for publication, March 21, 2017, and in revised form, May 11, 2017. Published, Papers in Press, May 19, 2017, DOI 10.1074/jbc.M117.787291

Camila Cristina Pasquali<sup>†1</sup>, Zeyaul Islam<sup>†1</sup>, Douglas Adamoski<sup>†1</sup>, Igor Monteze Ferreira<sup>†5</sup>, Ricardo Diogo Righeto<sup>¶||2</sup>, Jefferson Bettini<sup>¶</sup>, Rodrigo Villares Portugal<sup>¶</sup>, Wyatt Wai-yin Yue<sup>§</sup>, Ana Gonzalez<sup>\*\*</sup>, Sandra Martha Gomes Dias<sup>†3</sup>, and Andre Luis Berteli Ambrosio<sup>†4</sup>

From the <sup>†</sup>Laboratório Nacional de Biotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, São Paulo 13083-970, Brazil, the <sup>§</sup>Structural Genomics Consortium, University of Oxford, Oxford OX3 7DQ, United Kingdom, the <sup>¶</sup>Laboratório Nacional de Nanotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, São Paulo 13083-970, Brazil, the <sup>||</sup>School of Electrical and Computer Engineering, University of Campinas, São Paulo 13083-852, Brazil, and the <sup>\*\*</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025

Edited by Norma Allewell

On the basis of tissue-specific enzyme activity and inhibition by catalytic products, Hans Krebs first demonstrated the existence of multiple glutaminases in mammals. Currently, two human genes are known to encode at least four glutaminase isoforms. However, the phylogeny of these medically relevant enzymes remains unclear, prompting us to investigate their origin and evolution. Using prokaryotic and eukaryotic glutaminase sequences, we built a phylogenetic tree whose topology suggested that the multidomain architecture was inherited from bacterial ancestors, probably simultaneously with the hosting of the proto-mitochondrion endosymbiont. We propose an evolutionary model wherein the appearance of the most active enzyme isoform, glutaminase C (GAC), which is expressed in many cancers, was a late retrotransposition event that occurred in fishes from the Chondrichthyes class. The ankyrin (ANK) repeats in the glutaminases were acquired early in their evolution. To obtain information on ANK folding, we solved two high-resolution structures of the ANK repeat-containing C termini of both kidney-type glutaminase (KGA) and GLS2 isoforms (glutaminase B and liver-type glutaminase). We found that the glutaminase ANK repeats form unique intramolecular contacts through two highly conserved motifs; curiously, this arrangement occludes a region usually involved in ANK-mediated protein-protein interactions. We also solved the crystal structure of full-length KGA and present a small-angle X-ray scattering model for full-

length GLS2. These structures explain these proteins' compromised ability to assemble into catalytically active supra-tetrameric filaments, as previously shown for GAC. Collectively, these results provide information about glutaminases that may aid in the design of isoform-specific glutaminase inhibitors.

Glutamine, the most abundant amino acid in human plasma and muscles (1), is consumed by rapidly proliferating tumor cells to meet their increased energy and biosynthetic precursor demands (2, 3). The enzyme glutaminase converts glutamine into glutamate, which is further catabolized to produce ATP, nucleotides, amino acids, lipids, and glutathione. In this regard, glutaminase is a well established target for the inhibition of cell transformation (4). Accordingly, several oncogenes and tumor suppressors have already been described to be involved in glutaminase-dependent glutamine breakdown (4). Notably, c-Myc stimulates glutamine uptake and degradation by GLS (specifically the product of the *GLS* gene) to sustain cellular viability and Krebs cycle anaplerosis (5). Similarly, nuclear factor- $\kappa$ B (NF- $\kappa$ B) enhances glutamine metabolism by decreasing the expression of miR-23a, which targets the *GLS* mRNA (6); Rho GTPases, ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C)-Cdh1, and c-Jun also regulate glutamine metabolism by acting on *GLS* and consequently supporting cancer growth (7–9). However, a second gene encodes another glutaminase enzyme, named *GLS2*. In contrast to *GLS*, *GLS2* is regulated by the tumor suppressor p53 and has tumor suppressor features toward gliomas, glioblastomas, lung cancer, and hepatocarcinomas (10–14).

The existence of multiple glutaminases in mammals was first reported by Hans Krebs (15), on the basis of the detection of tissue-specific kinetic parameters and their susceptibility to inhibition by their catalytic products, when he probed the conversion of glutamine into glutamate in tissue extracts. In addition to being expressed by two different genes, at least four glutaminase isoforms have been described in mammals. The kidney-type isoforms, kidney-type glutaminase (KGA)<sup>5</sup> and

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo Grant 2014/20673-2 (to A. L. B. A.), Grant 2015/25832-4 (to S. M. G. D.), Scholarship 2014/19518-2 (to C. C. P.), Fellowship 2014/12663-7 (to Z. I.), Scholarship 2014/17820-3 (to D. A.), and Scholarship 2013/05668-0 (to I. M. F.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains supplemental Table S1 and Fig. S1.

The atomic coordinates and structure factors (codes 5u0i, 5u0j, 5u0k, and 5uqe) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Present address: Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Basel, CH-4058, Switzerland.

<sup>3</sup> To whom correspondence may be addressed. E-mail: sandra.dias@lnbio.cnpem.br.

<sup>4</sup> To whom correspondence may be addressed: Laboratório Nacional de Biotecnologia, LNbio, CNPEM, Rua Giuseppe Máximo Scolfaro, 10.000, Pólo II de Alta Tecnologia, Campinas, São Paulo 13083-970, Brazil. Tel.: 55-19-3512-1115; Fax: 55-19-3512-1004; E-mail: andre.ambrosio@lnbio.cnpem.br.

<sup>5</sup> The abbreviations used are: KGA, kidney-type glutaminase; LGA, liver-type glutaminase; GAB, glutaminase B; ANK, ankyrin; STAS, sulfate transporter and anti- $\sigma$  factor antagonist; EH, Eps15 homology; SAD, single-wavelength anomalous dispersion; r.m.s., root mean square; BPTES, (bis-2[5-phenylac-

## The multidomain architecture of human glutaminases

glutaminase C (GAC), are generated by alternative splicing of the *GLS* gene (2q32.2). Both isoforms are activated by inorganic phosphate and inhibited by glutamate (16–18). In contrast, the liver-type isozymes, liver-type glutaminase (LGA) and glutaminase B (GAB), originate from the *GLS2* gene (12q13.3) through the use of alternative transcription initiation sites (19). The recombinant construct that spans a common region between the *GLS2* isoforms responds poorly to phosphate (20) and is not inhibited by glutamate (19, 21).

Although both the kidney isozymes and GAB are expected to be localized in the mitochondrial matrix because of the transit peptide sequences at their N termini, KGA and LGA can also localize to the cytosol and nuclei (20, 22). However, GAC has been shown to be exclusively located in the mitochondrion (20). Moreover, GAC provides key growth advantages to cancer cells (20). KGA, LGA, and GAB share a glutaminase domain that is well conserved in sequence and structurally similar to bacterial glutaminases, and it is flanked by a long N-terminal domain folded in an EF-hand-like four-helix bundle and a C-terminal domain with three putative ANK repeats (20). In contrast, GAC has a shorter, 48-amino acid-long C terminus with no canonical motif or domain. GAC is also unique because it assembles into highly active, long double-stranded helical filaments in the presence of inorganic phosphate, whereas KGA forms shorter and less active structures under the same conditions (23).

Intrigued by the diversity in numbers, architecture, mechanism of activation, and enzymatic capabilities of the human glutaminases, we felt motivated to investigate the origin, evolution, and possible functions of such complex features. First, by identifying homologous prokaryotic and eukaryotic sequences and building a phylogenetic tree, we propose that the multidomain architecture of the mammalian glutaminases *GLS* and *GLS2* is a feature inherited from bacterial ancestors, probably simultaneously with the hosting of the proto-mitochondrion endosymbiont. From this phylogenetic tree, we also propose an evolutionary model wherein a *GLS*-like predecessor gene in tunicates gave rise to the currently known human isozymes through exon remodeling, gene duplication, and retrotransposition events. In addition, we solved high-resolution crystallographic structures of the C-terminal ANK repeats from both human KGA and GAB/LGA. These structures display a unique type of dimerization for ANK domains, governed by two short amino acid motifs that are highly conserved from bacteria to higher eukaryotes. Finally, we present the first C terminus-containing crystallographic structure of human KGA, which explains its inability to assemble into the catalytically active supra-tetrameric filaments, as previously shown for GAC. A similar model for human *GLS2* is also proposed on the basis of small-angle X-ray scattering data and cryo-electron microscopy. Collectively, these results contribute to the general knowledge on the many human glutaminase isoforms as well as explain the mechanistic diversity that poses challenges that still must be overcome to

increase the chances of successfully inhibiting the enzyme in a clinical setting.

## Results

### The multidomain origin of glutaminases

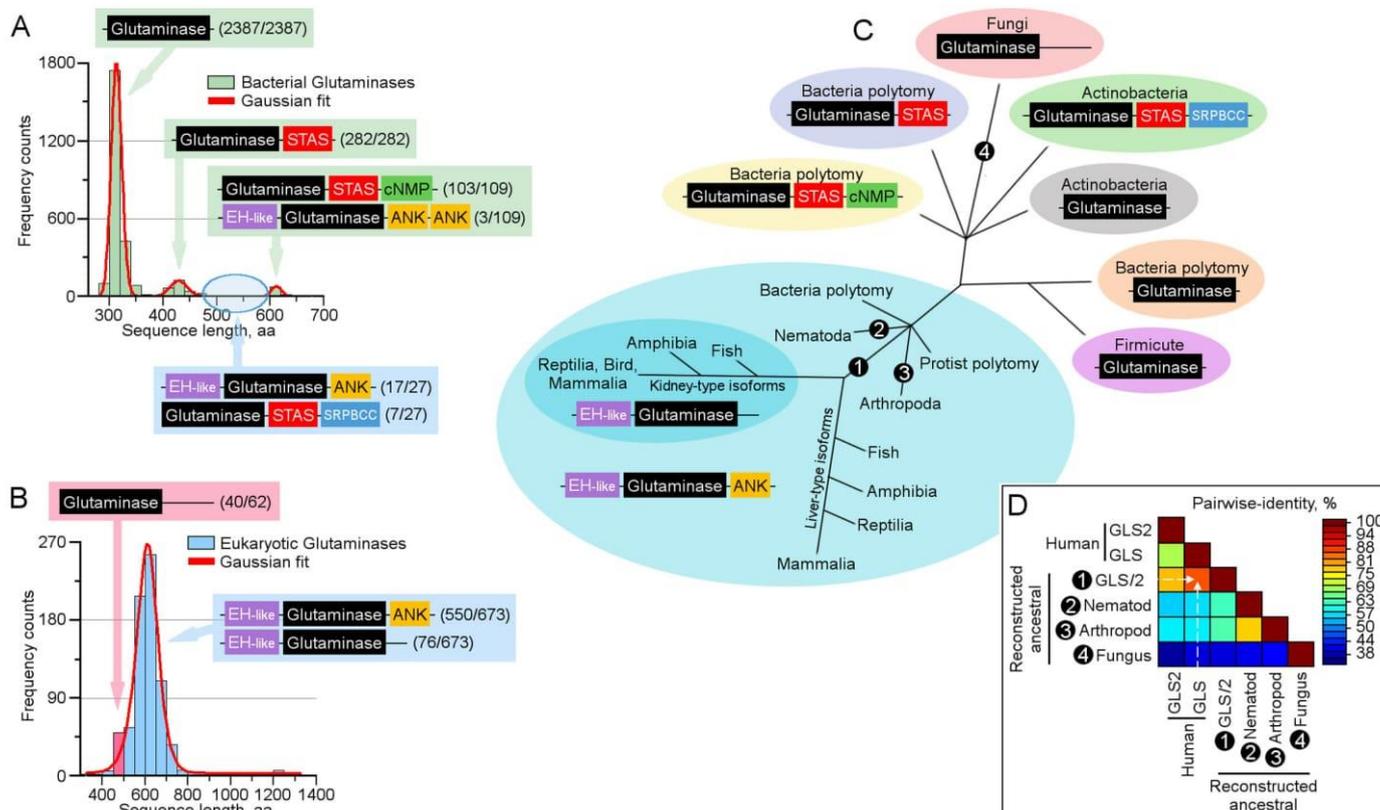
An evolutionary analysis of 2,796 bacterial glutaminase protein sequences and 789 eukaryotic sequences selected from a BLAST search was performed using the glutaminase domain from human KGA (Ile<sup>221</sup>-Arg<sup>544</sup>) as the query sequence (*E* value  $\leq 0.0001$ ). Particularly, no glutaminase homologs were identified in archaea. The bacterial sequence sizes presented three normal distributions. The first and major distribution (~85% of the sequences) was centered at  $313 \pm 10$  amino acids (Fig. 1A), thus suggesting that the proteins predominantly contained only the catalytic domain (24). The second distribution was smaller (10% of the sequences) and centered at  $429 \pm 16$  amino acids. Detailed analysis suggested that this group possesses a sulfate transporter and anti- $\sigma$  factor antagonist (STAS) domain located C-terminal to the glutaminase domain (Fig. 1A). Finally, 4% (109) of bacterial glutaminases clustered in a third distribution centered at 613 amino acids. These sequences, in addition to the STAS domain, have an extra cyclic nucleotide-binding domain located N-terminal to the glutaminase domain. Notably, three cyanobacteria sequences from this group possess a glutaminase domain flanked by an Eps15 homology-like domain (EH-like) at the N terminus and two clusters of ANK repeats at the C terminus (Fig. 1A). The Eps15 homology domain is a pair of EF-hand motifs that recognize and bind to proteins containing Asn-Pro-Phe (NPF) sequences (25). Aside from these three-domain cyanobacteria sequences, 17 proteo- and actinobacteria sequences situated between the second and third normal distributions (Fig. 1A, *blue oval*) also exhibit a glutaminase domain flanked N-terminal by an EH-like domain and C-terminal by only one cluster of ANK repeats.

Eukaryotic glutaminases, which primarily belong to the kingdoms Animalia and Fungi, are consistently longer and present a single Gaussian size distribution averaging 631 amino acids in length (mode at 615), with a skewed standard deviation of 45 amino acids toward shorter sequences and 26 amino acids toward longer sequences (Fig. 1B). The left-most end of the distribution contained the fungal glutaminases, which are 450 to 500 amino acids long and present a unique architecture, wherein the glutaminase domain is followed by a long C terminus ( $156 \pm 7$  amino acids) with an unknown fold (Fig. 1B, *pink bar*). The remaining eukaryotic glutaminases that are longer than 500 amino acids in length mostly belong to the Animalia kingdom and have extended N-terminal sequences of varying lengths (Fig. 1B). The N-terminal region immediately adjacent to the glutaminase domain is more than 40% similar to the EF-hand-like four-helix bundle, which is structurally similar to the EH-like domain found in bacteria. Indeed, the X-ray structures of human *GLS* containing this domain have an EH-like structure (20).

Putative ANK repeats are predicted for about 550 of the longer eukaryotic glutaminases and are located C-terminal to the glutaminase domain with no exception. Another 76 sequences,

etamido-1,2,4-thiadiazol-2-yl]ethyl sulfide; BIS-TRIS propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; ISAC, iterative stable alignment and clustering; PDB, Protein Data Bank; Rh, hydrodynamic radius.

## The multidomain architecture of human glutaminases



**Figure 1. Sequence and architecture analysis of bacterial and eukaryotic glutaminases.** The amino acid sequence corresponding to the glutaminase domain of the *GLS* gene (Ile<sup>221</sup>-Arg<sup>544</sup>) was used as input of blastp search against the non-redundant protein sequences (nr) database ( $E$  value  $\leq 0.0001$ ). The sequences obtained were analyzed by number of amino acids (aa); the normal distribution of the sequence length for bacterial and eukaryote glutaminases are, respectively, represented in A and B, with the respective architectures indicated in rectangles; in parentheses, the number of sequences containing the corresponding architecture/the total number of sequences contained in the referred region of the distribution plot. C, cladogram based on architectural organization obtained from a maximum likelihood phylogenetic reconstruction approach. D, the generated phylogenetic tree and respective alignment were used to reconstruct ancestral protein sequences for specific nodes (ancestral chordates (1); nematodes (2); arthropods (3); and fungi (4)) using maximum parsimony. The obtained ancestral sequences were aligned with the human *GLS* (KGA and GAC) and *GLS2* (GAB and LGA) and the sequence similarity displayed as a heatmap of pairwise distances constructed in SDT using MUSCLE alignment. The obtained result suggests that *GLS* is the most primitive gene, which has been further duplicated to generate *GLS2*. *cNMP*, cyclic nucleotide-binding domain; *SRPBCC*, START/RHO $\alpha$ C/PITP/Bet v1/CoxG/CalC.

all of mammalian origin, have a shorter C terminus that closely resembles human GAC in sequence (Fig. 1B). Finally, a few glutaminase sequences (11 in total) belonging to the major haptophyte and heterokont lines of eukaryotes were identified. These sequences are much longer, ranging from 800 to 1250 amino acids in length, and the predicted architectures of some of them consist of two complete tandemly associated ANK-containing glutaminases (data not shown).

To detect evolutionary relationships based on the architectural organization, we next reconstructed an unrooted cladogram using the analyzed sequences (Fig. 1C). The derived tree structure has a taxa organization similar to that found in rRNA-based phylogenetic trees (26). Many unresolved nodes (polytomy) were observed, particularly for the bacterial glutaminases. We believe that such observation stems from the fact that many bacterial species possibly have more than one glutaminase, often as a result of the combination of a single protein with a multidomain isozyme. Except for fungal glutaminases, which clustered closer to the bacterial homologs, the eukaryotic proteins had a better-resolved tree. Protists, nematodes, and arthropods form a polytomic group with chordates, as well as with ANK-containing proteo- and cyanobacteria; this grouping suggests that the appearance of this multidomain feature was

an early event, concomitant with the symbiotic association related to the appearance of eukaryotes, because archaea does not carry any glutaminase-like coding gene. A subsequent lineage split linked to the appearance of multiple variations around the same basic architecture (EH-like + glutaminase + ANK) suggests that the gene duplication event that produced the *GLS* and *GLS2* genes and isoform differentiation (with the appearance of GAC-like isoforms; Fig. 1B, dark blue oval) appeared with chordates. However, interestingly, birds have completely lost the *GLS2* gene.

Finally, based on a theoretical reconstruction of the ancestral glutaminases from each of the main eukaryotic branches (chordata, arthropods, nematodes, and fungi), the most primitive glutaminase acquired from a bacterial ancestor was most probably the kidney-type glutaminase (KGA-like), as shown in Fig. 1D. The general organization EH-like + glutaminase + ANK is preserved in all four reconstructed ancestral sequences. Comparison of the sequence identity of human genes with the reconstructed common ancestor of *GLS* and *GLS2* from Gnathostomata suggested that *GLS* was restrained during evolution to maintain the same function, whereas *GLS2* exhibited more changes, probably acquiring new functions and/or regulatory mechanisms.

## The multidomain architecture of human glutaminases

### The origin of glutaminase isoforms

Having established a plausible origin for the multidomain architecture of the glutaminases, we next asked how the different variants of the enzyme, currently observed in humans, might have evolved. We first performed a “top-down” search for glutaminase genes in phylum Chordata using the complete genomic sequence of the human *GLS* and *GLS2* genes as templates. Second, we identified regions homologous to exon 14 of *GLS* (after which the *GLS* splicing event occurs) and exon 15 (the “GAC exon”).

In the lower chordates, we verified that the genome of the tunicate sea squirt (27) (*Ciona intestinalis*) has a single *GLS*-like gene containing 12 exons. A region similar to human exon 14 (70% identity, red bar in the tunicate branch, Fig. 2A) is inserted in exon 11. The Cephalochordate amphioxus (28) (*Branchiostoma floridae*) also has a single glutaminase gene, which, interestingly, has only two exons. In this gene, the region homologous to the human exon 14 is appropriately located at the 3' boundary of exon 1; however, no homology to the GAC exon was identified in the downstream sequence. A gene duplication event was observed further upstream in lampreys (29) (*Pekinusus marinus*, Hyperortia), which generated distinct sequences related to the human *GLS* and *GLS2* genes.

Next, the Chondrichthyes *Callorhynchus milii* (elephant shark (30)) also has two glutaminases that are phylogenetically related to the *GLS* and *GLS2* genes without annotated splicing regulation. However, in the *GLS* gene, a region homologous to the human GAC exon is for the first time observed and located downstream of an exon 14-equivalent sequence, thereby suggesting an origin for the splicing variant GAC. Zebrafish (*Danio rerio* (31)) and other fish species possess 5 glutaminase genes, thus suggesting subsequent autapomorphic duplication events for both *GLS* and *GLS2*. Interestingly, the GAC isoform was maintained in both copies of the *GLS* gene, thus supporting the hypothesis that the GAC exon appeared before this duplication; a new *GLS*-like (*GLSL*) sequence was identified in species within this branch, such as zebrafish, elephant shark, and others.

Of note, the amphibian *Xenopus tropicalis* (Western clawed frog (32)), with only one copy of both *GLS* and *GLS2*, points to a compaction of the *GLS2* introns and exons against the spread of the *GLS* across a larger genomic region. Finally, the human glutaminase gene structures retained the same architecture observed in the amphibian genes, although the sequence lengths were smaller.

The identification of several transposable elements such as Alu and L2 (Fig. 2B), within the human *GLS* intron 15, strongly suggests that the insertion of the GAC exon was due to an early retrotransposition event; however, no consensus matching elements were found in the rapid evolving intron 14. Nevertheless, a thorough tblastn analysis failed to identify a region homologous to the GAC exon (exon 15) within the human genome, preventing us from proposing an original location for this sequence.

### Crystal structures of the human glutaminase ankyrin repeats

We next determined the crystal structures of the ANK-containing C-terminal regions of both human KGA (KGA.ANK: Val<sup>551</sup>-Leu<sup>669</sup>) and GLS2 (GLS2.ANK: Lys<sup>485</sup>-Val<sup>602</sup>). Both

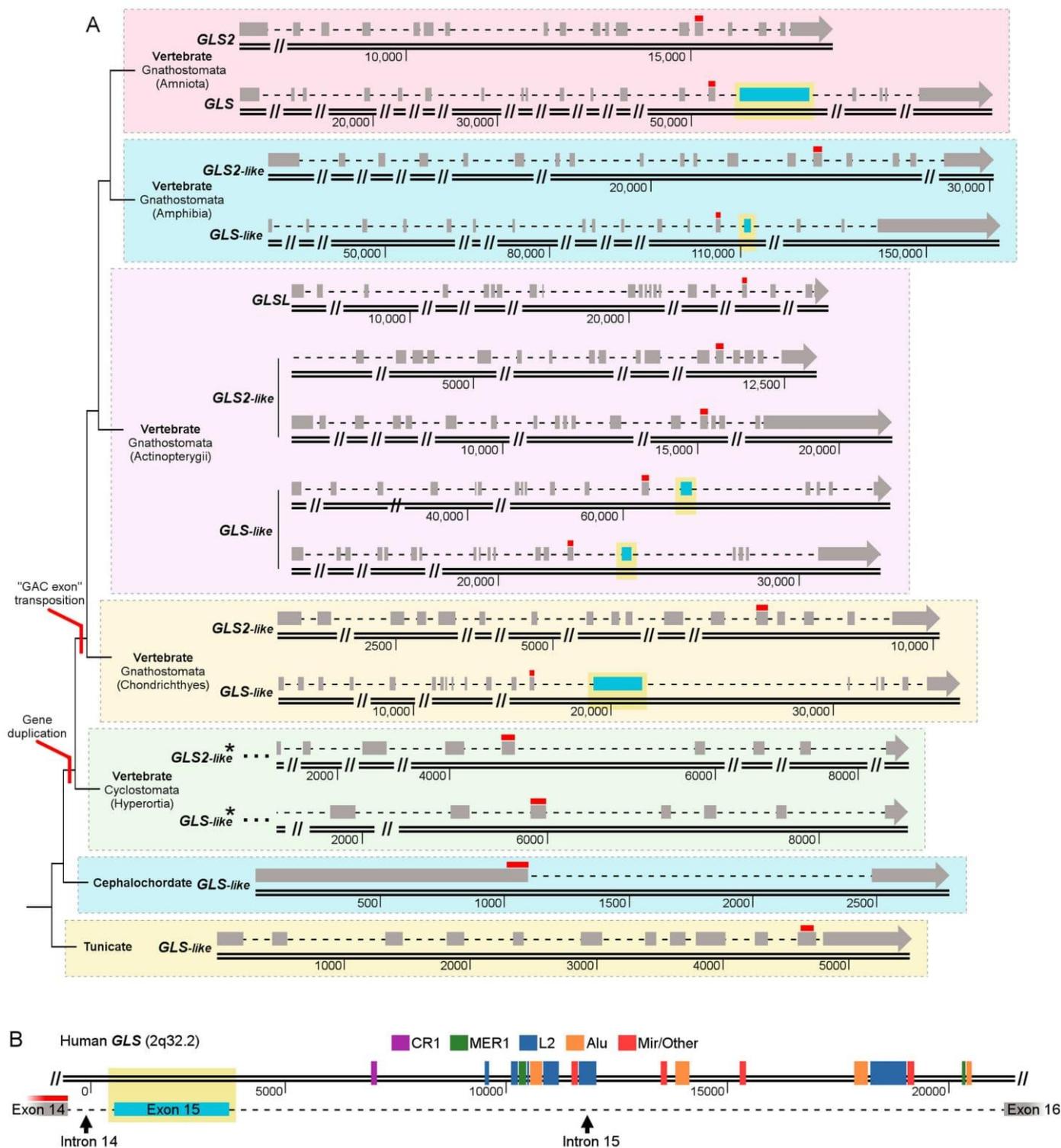
proteins were subjected to limited proteolysis with trypsin (for KGA.ANK) or chymotrypsin (for GLS2.ANK) during purification to remove the likely intrinsically flexible regions and obtain well diffracting crystals. X-ray diffraction datasets were obtained for KGA.ANK crystals belonging to two different space groups, namely the tetragonal P4<sub>3</sub>2<sub>1</sub>2 and a monoclinic P12<sub>1</sub>1. Sulfur single-wavelength anomalous dispersion (SAD) phasing was used to produce a higher resolution molecular model (at 1.41 Å maximum) from the tetragonal crystals. The obtained structure was then used as a search model for molecular replacement of the monoclinic crystals, which diffracted up to 1.74-Å maximum resolution. Finally, the structure, which was collected at 2.55-Å maximum resolution, of the GLS2.ANK hexagonal crystals (space group P6<sub>5</sub>) was solved by molecular replacement using the KGA.ANK structure. The statistics of the data collection, processing, phasing, and model refinement data for all the structures are shown in Table 1. Because of the high structural similarity observed for both KGA.ANK structures (backbone r.m.s. deviations of 0.4 Å), only the higher resolution model was used for the subsequent detailed analysis and comparison to the GLS2.ANK structure.

Overall, the C-terminal domains of KGA and GLS2 each contain three ANK repeats (labeled ANK1 to ANK3) and are very similar, both in sequence composition (80% identical) and overall structure (r.m.s. deviation of 0.5 Å over 89 equivalent C $\alpha$  positions; Fig. 3A). The largest structural variation (a shift between 1.3 and 1.8 Å of the C $\alpha$  atom positions) occurs in the  $\beta$ -hairpin that links ANK1 to ANK2 ( $\beta$ -hairpin 2), whereas the minimal deviations ( $\sim$ 0.1 Å) are observed in the inner and outer  $\alpha$ -helices of ANK2. The primary sequences of ANK1 and ANK3 repeats in both KGA and GLS2 diverge from the canonical ANK repeats at the TPLH tetrapeptide motif, which is located at the beginning of the inner  $\alpha$ -helices and is responsible for the stabilization of ANK repeats through reciprocal hydrogen bonding between the threonine hydroxyl and histidine imidazole groups (33, 34) (Fig. 3B). The ANK1 inner helix of both KGA and GLS2 is a half-turn longer and lacks the  $\beta$ 1 turn compared with the typical inner helix observed in ANK repeats (Fig. 3B), whereas in the ANK3 inner helix, the histidine is replaced with an aspartate, thus preventing the formation of the conserved hydroxyl-imidazole hydrogen bond. Moreover, in both structures, ANK2 contains an extra residue, a surface-exposed lysine (Lys<sup>611</sup> at KGA and Lys<sup>543</sup> at GLS2, indicated by a plus sign in Fig. 3B), at the C terminus of the outer helix. Finally, the GLS2 ANK3 outer helix is one turn longer than the canonical ANK repeats (Fig. 3B).

### Glutaminase ankyrin repeats assemble into atypical dimers

A search for the biologically relevant assembly of the KGA and GLS2 ANK domains was performed using the PISA (protein interfaces, surfaces and assemblies) server (35). The program identified a thermodynamically stable interface involving a crystallographic homodimer, with a solvation free energy gain of  $-5.4$  kcal/M upon formation (Fig. 3C). The identified interface occludes an area of 800 Å<sup>2</sup> and corresponds to a symmetric groove-to-groove interaction in both structures. Twelve hydrogen bonds and 12 salt bridges are formed between the short loops present in the  $\beta$ -hairpins 1 (linking ANK1 to ANK2) and

## The multidomain architecture of human glutaminases



**Figure 2. Evolution of glutaminase intron-exon structure.** *A*, protein sequence corresponding to human *GLS* exon 14 was used as a query of translated blast (tblastn) searches against available genomes from *C. intestinalis* (Tunicate), *B. floridae* (Cephalochordate), *P. marinus* (Cyclostomata), *C. milii* (Chondrichthyes), *D. rerio* (Actinopterygii), and *X. tropicalis* (Amphibia). After an initial Reciprocal Best Hit region was found, the correct position of the exon was determined by pairwise LALIGN between sequences. Region downstream genomic sequence homologous to exon 14 were also evaluated by LALIGN against human GAC-exclusive exon (exon 15). As both *P. marinus* glutaminases are incomplete, available exons were found in contig using LALIGN approach. The human genes are represented at the top branch (Amniota). The *thin red bars* represent the homologous region to the *GLS* exon 14 and the *light blue boxes* indicate homologous regions to the exon 15 of human *GLS*. The introns, represented as *dashed lines*, are not drawn to scale and the omitted sections are indicated by *two bars* (||). The *asterisk* (\*) indicates that the deposited sequences are incomplete. *B*, a search for retrotransposition evidences in the genomic region comprehending exon 15 and introns 14 and 15 of human *GLS* using TranspoGene resulted in the identification of various transposable elements in the intron 15; however, no consensus matching elements were found in the intron 14.

## The multidomain architecture of human glutaminases

**Table 1**  
X-ray crystallography data collection parameters and structure refinement statistics

Data for the outer shell are shown in parentheses.

Data collection and processing				
Beamline	12–2 at SSRL, USA	12–2 at SSRL, USA	12–2 at SSRL, USA	I03 at DLS, UK
Wavelength (Å)	0.9795	0.9795	0.9795	0.9762
Space group	P 43 21 2	P 1 21 1	P 65	P 42 21 2
Cell parameters (Å)	$a = b = 75.3, c = 88.3$	$a = 76.5, b = 47.2, c = 115.7$ ( $\beta = 90.1^\circ$ )	$a = b = 85.2, c = 336.8$	$a = b = 144.1, c = 615.2$
Resolution range (Å)	75.0 – 1.42 (1.45 – 1.42)	64.0 – 1.72 (1.75 – 1.72)	84.0 – 2.55 (2.64 – 2.55)	50.0 – 3.6 (3.67 – 3.60)
Unique reflections	48,262 (2,342)	82,659 (4,303)	45,005 (4,413)	76,934 (4,463)
Multiplicity	22.7 (23.2)	6.4 (6.3)	8.1 (8.1)	23.4 (20.8)
$R_{\text{pim}}$ (%)	2.6 (30.9)	2.4 (31.8)	3.5 (25.2)	7.4 (49.4)
CC (1/2)	0.994 (0.441)	0.999 (0.902)	0.999 (0.860)	0.998 (0.716)
Completeness (%)	100 (100)	94.1 (92.1)	99.9 (99.8)	100 (100)
$\langle I/\sigma(I) \rangle$	17.1 (2.9)	15.7 (2.1)	14.1 (3.4)	8.8 (1.9)
Average mosaicity ( $^\circ$ )	0.62	0.55	0.49	0.30
Wilson B-factor (Å <sup>2</sup> )	13.6	25.7	48.5	104.6
Monomers/AU	2	4*	10	5
Solvent content (%)	46.7	36.4*	53.0	77.1
Matthews coefficient (Å <sup>3</sup> Da <sup>-1</sup> )	2.32	1.93 <sup>a</sup>	2.62	5.37
Model refinement				
Resolution range (Å)	57.3 – 1.42	64.0 – 1.72	73.8 – 2.55	50.0 – 3.6
Test set reflections (%)	2500 (5.2)	4285 (5.2)	2366 (5.3)	7160 (5.0)
$R_{\text{factor}}/R_{\text{free}}$ (%)	14.4/17.9	18.3/22.2	21.8/25.7	27.4/31.2
Clashscore	2	2	5	3
Side chain outliers (%)	0.6	0.7	2.4	2.6
RSRZ outliers (%)	1.6	0.6	2.8	5.7
Average B, all atoms (Å <sup>2</sup> )	22.0	34.0	53.0	137.0
Number of protein atoms	1,506	2,943	6,860	18,905
Solvent molecules	285	640	110	
Inorganic atoms	1 Cl <sup>-</sup> , 3 Na <sup>+</sup>	6 Na <sup>+</sup>		
Ligand molecules				3 BPTES
R.m.s. deviations from standard geometry				
Bond length (Å)	0.015	0.019	0.002	0.003
Bond angles ( $^\circ$ )	1.313	1.454	0.479	0.566
Ramachandran plot (%)				
Most favorable	98.0	98.6	95.9	87.5
Outliers	0	0	0.2	2.1
<b>PDB ID</b>	5u0i	5u0j	5u0k	5uqe

<sup>a</sup> Translational non-crystallographic symmetry was detected for this dataset.

2 (linking ANK2 to ANK3, Fig. 3B) in chain A with their respective equivalents in chain B (Fig. 3C, *left panel*, for  $\beta$ -hairpin 1 loop interactions and *right panel* for  $\beta$ -hairpin 2 loop interactions). The short loops of  $\beta$ -hairpins 1 and 2 contain the motifs Asp-Tyr-Asp (DYD, residues 583–585 in KGA and 516–518 in GLS2) and Asp-Arg-Trp (DRW, residues 617–619 in KGA and 550–552 in GLS2), respectively (Fig. 3C). Despite the overall relatively low sequence identity of the eukaryotic and bacterial glutaminase ANK repeats, both the DYD and DRW motifs are highly conserved (Fig. 3, D and E, respectively). Finally, additional polar interactions are observed between the  $\beta$ -hairpin 1 loop and the ANK3 inner helix and between the  $\beta$ -hairpin 2 loop and the ANK2 inner helix, complementing the identified symmetrical interface (data not shown).

### Crystal structure of full-length human KGA

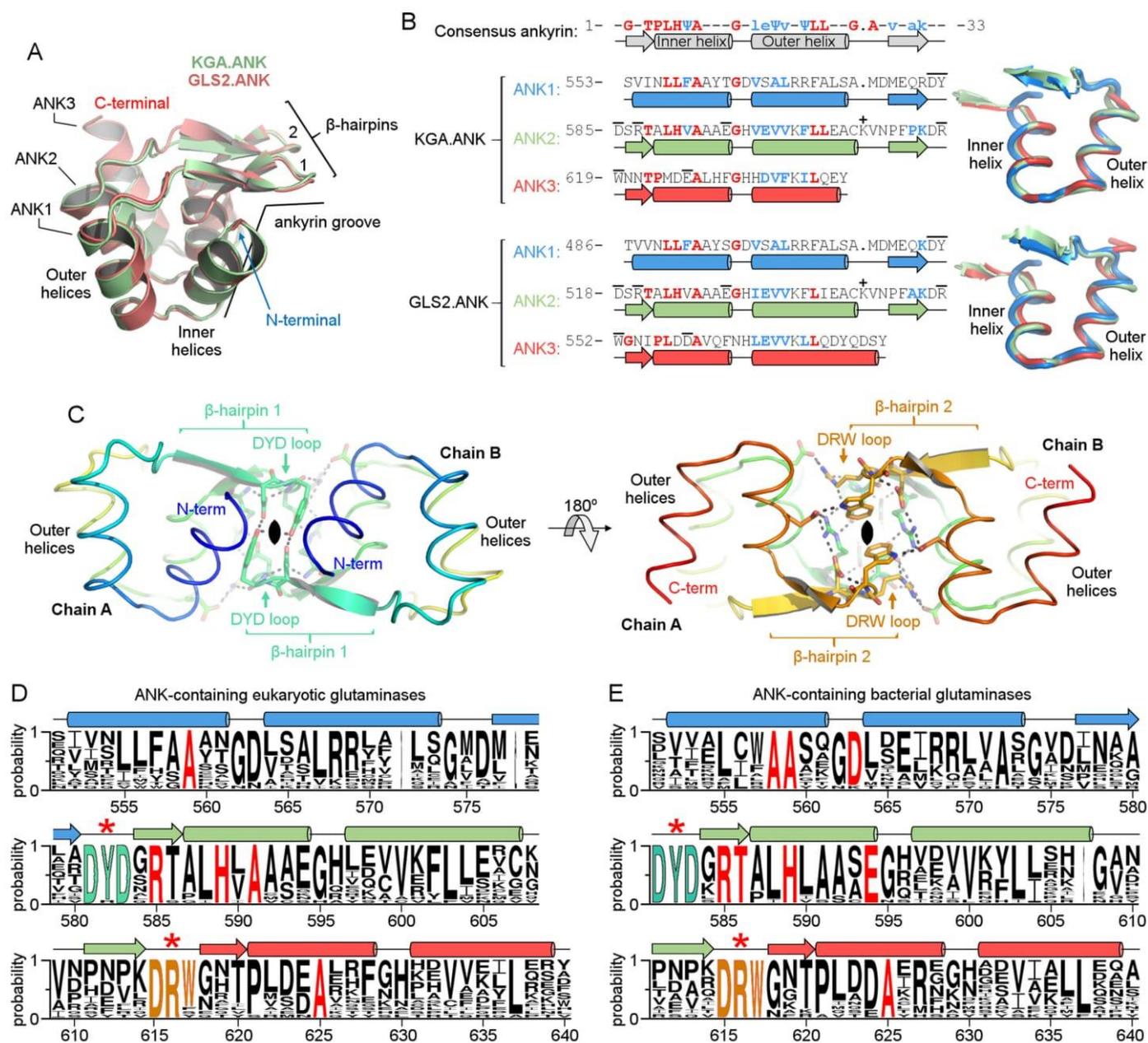
In addition to the determination of the crystal structures of the isolated C-terminal ANK dimers, we also solved the novel crystal structure of KGA containing the C terminus, which was bound to the inhibitor bis-2[5-phenylacetamido-1,2,4-thiadiazol-2-yl]ethyl sulfide (BPTES) (36). The diffraction dataset was collected at a 3.6-Å maximum resolution (space group P4<sub>2</sub>2<sub>1</sub>2) with lengthy cell parameters and a high solvent content of ~78% (Table 1). Phasing was achieved by molecular replacement using the coordinates of the previously solved N-terminal and glutaminase domains from GAC bound to BPTES (PDB code 4jkt) and the KGA.ANK dimer described here as

search models. The P4<sub>2</sub>2<sub>1</sub>2 asymmetric unit contains five KGA monomers; accordingly, one canonical tetramer (a dimer of dimers) is formed within the asymmetric unit by four monomers and, the fifth monomer is related to another tetramer by the crystal 2-fold screw axis. The final model obtained (Fig. 4A) was refined to  $R_{\text{factor}}$  and  $R_{\text{free}}$  of 27.3 and 31.8%, respectively.

The ANK dimers in the KGA structure (*green surface* in Fig. 4B) are spatially located between the N-terminal EF-hand-like domains (*blue surfaces* in Fig. 4B). Notably, both ANK dimers of a tetramer lie on the same side of a sectional plane defined by the two longest axes of the protein (Fig. 4B), breaking the previously established 2-fold dihedral symmetry of the glutaminase structures lacking the C terminus (20).

The low-resolution model did not allow us to provide a detailed description of the side chain interactions between the ANK dimers and the N-terminal domain. However, the close contact between the ANK1 outer helix of chain A with the EF-hand-like domain helix H1 of chain B is unmistakable (Fig. 4C). The interface contains large, exposed polar residues, such as glutamic acid and arginine, which are likely to be responsible for the contact. An equivalent arrangement was also observed for the ANK1 outer helix of chain C with the EF-hand-like domain helix H1 of chain D (data not shown). The majority of the C-terminal region of KGA adjacent to the ANK repeats (residues Thr<sup>647</sup>-Leu<sup>669</sup>) is likely heterogeneous in conforma-

## The multidomain architecture of human glutaminases



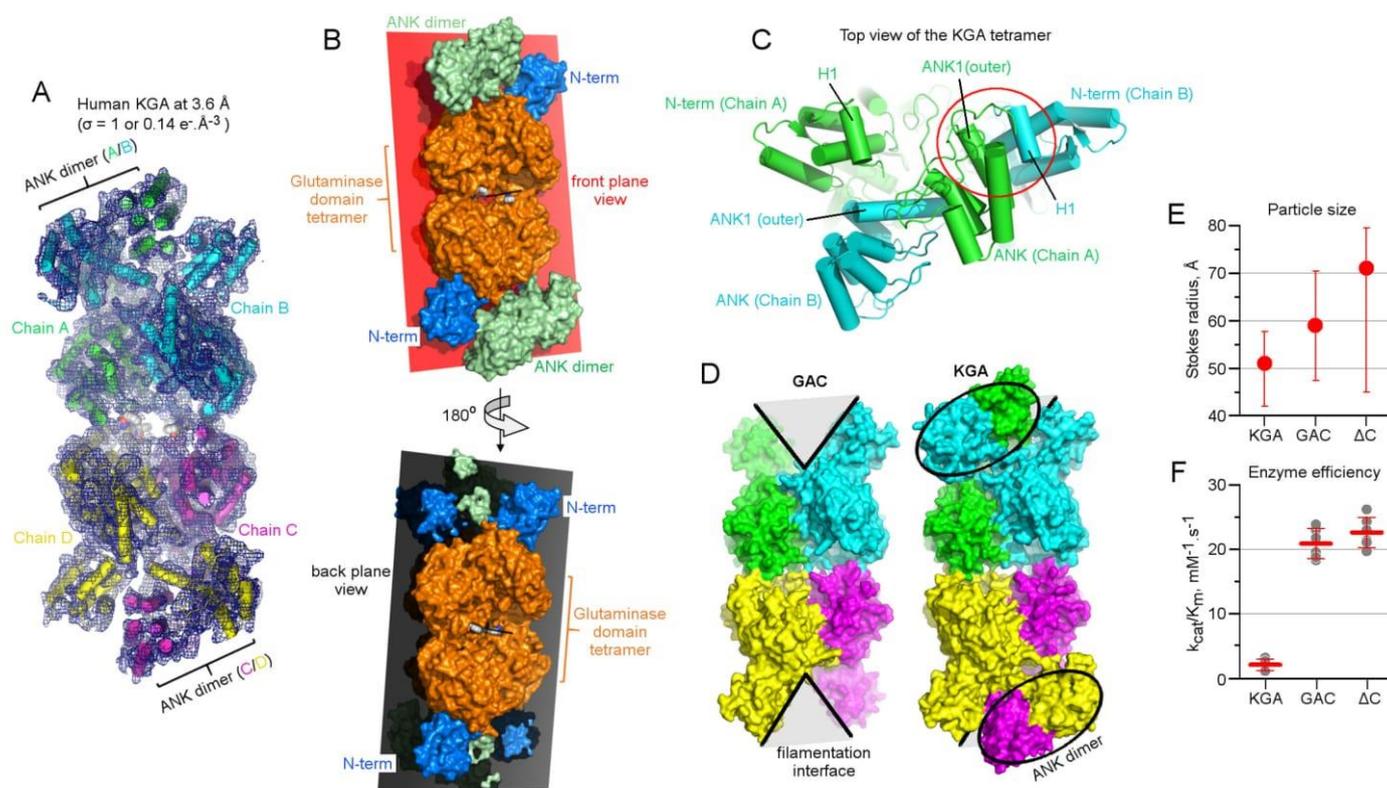
**Figure 3. Overall structural features of ANK domain of glutaminases.** *A*, schematic representation of three ANK repeats (ANK1 to ANK3) present as a C-terminal domain of KGA and GLS2. Both ANK repeats are very similar in sequence as well as in the overall structure. *B*, alignment of the glutaminase ANK repeats and the consensus sequence (shown on top), proposed in Ref. 39. Highly conserved residues are capitalized and in red, semiconserved residues are colored cyan and not capitalized. Residues involved in the dimer interface are indicated by a bar. ANK2 in both KGA and GLS2 contains an extra surface-exposed lysine, which is indicated by the + signal. On the right, the superposition of ANK1, ANK2, and ANK3 is represented. *C*, dimer interface and associated interaction of ANK repeats. The side chains of the motifs DYD (left) and DRW (right) are represented in sticks. *D*, representation of sequence conservation of glutaminases ANK from eukaryotes and *E*, bacteria. The size of the letters is proportional to the degree of conservation of residues. The motifs DYD and DRW are highlighted. The residues in red shows the 100% conservation throughout the alignment. Residues are numbered according to human KGA.

tion and therefore could not be modeled. According to secondary structure predictions by the JPred4 server (37), this region is mostly unstructured. This information, added to the fact that such region is present in a KEN box involved on KGA degradation (8), suggest that this segment is likely functionally independent from the ankyrin domain.

As shown in our previous study (23), the phosphate-dependent enzymatic activity of the GLS isoforms is directly related to their ability of self-assembling into supra-tetrameric helical filaments; GAC is the most active isoform and forms longer filaments. Moreover, we previously demonstrated that this

assembly was generated by end-to-end associations via the N-terminal domains; the KGA assembly was always shorter, although it formed the same structure. Based on the KGA crystal structure presented here, we hypothesized that the ANK repeats located between the N-terminal domains prevent KGA from forming longer higher active filaments due to the fact that it destabilizes the polymerization. More specifically, the ANK dimers limit the association of the KGA tetramers into long polymers (23) because they limit the formation and availability of the filamentation interface (Fig. 4D). This interface is the groove region by which the single strand filaments of the GAC

## The multidomain architecture of human glutaminases



**Figure 4. Multidomain structure of KGA and biophysical characterization of KGA, GAC, and  $\Delta C$ .** *A*, electron density and final model for the complete tetramer of KGA. *B*, planar representation, dividing the tetramer into front and back halves to show the position of two ANK dimers at the same side of the plane. *C*, upper view of the tetramer, indicating the close contact between the ANK1 outer helix and the helix H1 of the N-terminal EF-hand like domain. *D*, the filamentation interface between two N-terminal domains (region delimited by gray), and by which the single strand polymer of GAC (PDB code 4jkt) grows (23), is occluded in KGA because of the presence of the ankyrin dimers. *E*, the Stokes radii calculated from the size exclusion chromatographic profiles of KGA, GAC, and  $\Delta C$  (KGA lacking the ANK repeats), indicating that ANK repeats prevent the self-assembly of glutaminases into the higher supra-tetrameric filament forms. *F*, the comparative enzymatic efficiency of KGA, GAC, and  $\Delta C$  toward glutamine in the presence of 20 mM phosphate. The lack of ANK repeats enhances the catalytic efficiency of KGA. Red lines indicate mean  $\pm$  S.D.

isoform grows indefinitely, via end to end interaction between pairs of N-terminal domains (23).

To further confirm this hypothesis, we generated a C terminally truncated GLS construct containing the common sequence between KGA and GAC (N terminus and glutaminase domains, called  $\Delta C$ ), but lacking the ANK repeats. Accordingly, the construct had a larger Stokes radius in comparison to KGA and GAC and was more heterogeneous in size (Fig. 4E). Moreover, as expected, the  $\Delta C$  construct was as active as GAC, possibly due to its increased ability to form extended filaments (Fig. 4F).

#### Low resolution structure of GLS2

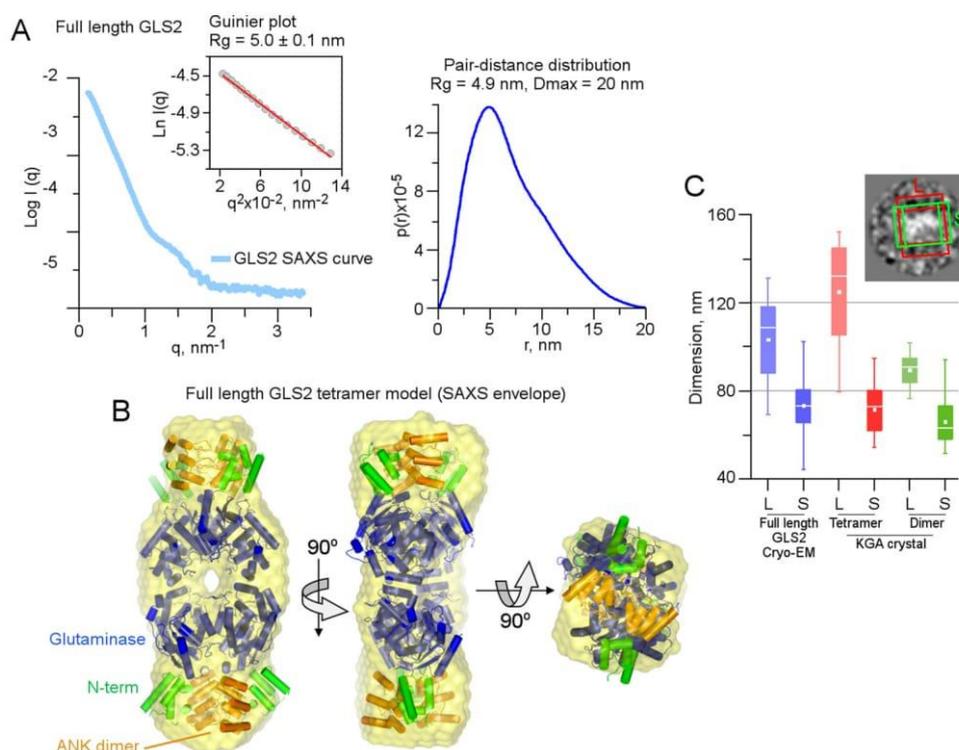
Last, the availability of the crystal structure of individual domains of GAB/LGA, such as the C-terminal ANK dimers reported here and the glutaminase domain tetramer (PDB code 4bqm), allowed for a proposition of the full-length structure based on small-angle X-ray scattering data (Fig. 5, A and B) for the multidomain portion common to both isoforms. The C-terminal ANK domains are expected to be spatially located between the N-terminal domains, in an organization similar to that observed for KGA. Consequently, this protein is also expected not to assemble in filament-like higher order oligomers, being stable at the tetramer form. This has been previously observed by our group (23) and is here corroborated by measurements on cryo-electron microscopy images (Fig. 5C).

#### Discussion

The human genome has two copies of glutaminase-coding genes, which are known to produce at least four isoforms through alternative splicing or transcription at alternative initiation sites (4). Genetic evolutionary models often predict an increase in the gene copy number prior to the specialization or gain of new functions, with a momentary intermediate state of specialization with an overlap of functions (38). In addition, this apparent redundancy in glutaminase isoforms can be explained by differential tissue expression patterns, as well as the cell proliferation state. In this regard, KGA is expressed in kidney tubule epithelial cells, in which ammonia production is key to renal acid-base regulation (16); LGA is expressed in periportal hepatocytes, where it participates in urea synthesis (17). In addition, although GAC meets the glutamine-dependent metabolic demands of several types of cancer (20), GAB has tumor-suppressing activity in hepatocarcinomas and gliomas (40).

Despite this functional diversity, the human glutaminase isoforms converge into a conserved multidomain structure. Based on a combination of computational predictions and experimental evidence mainly produced by X-ray crystallography studies, we found that all four isoforms contain an N-terminal EF-hand-like domain, followed by a glutaminase domain in the middle and three ANK repeats at the C terminus. The only exception is GAC, in which the C terminus is a short

## The multidomain architecture of human glutaminases



**Figure 5. Low-resolution solution model for full-length GLS2 (GAB/LGA).** Intensity (left panel) and pair-distance (right panel) distribution functions, obtained for GAB/LGA subject to SAXS experiments. *B*, orthogonal views of the superposition between the SAXS envelope and the collection of X-ray crystal structures from GLS2 glutaminase domain (PDB code 4BQM), N-terminal portion of human GLS (PDB code 3CZD) and GAB/LGA.ANK. *C*, box plot of the two perpendicular dimensions for GAB/LGA, large (*L*) and short (*S*) as taken from the cryo-EM micrographs (inset). KGA dimer and tetramer crystal structures projections were used as a reference for the dimensions of these two oligomeric states.

unstructured region, because of alternative splicing of the *GLS* gene.

In this work, the multidomain architecture was shown to be shared by all chordates, including fishes, reptiles, birds, and amphibians, as well as by arthropods and nematodes. By surveying the database-deposited glutaminase protein sequences, we observed similar organizational characteristics in free-living proteo- and cyanobacteria, but not fungi. From our data, we propose that the multidomain glutaminase structure has a very ancient origin. In addition, we verified that multiple *GLS* genes and isoforms are present in a wide range of vertebrates. By further analyzing the genomes of representative species of the phylum Chordata, we were able to identify early events of exon repositioning among tunicates and cephalochordates, followed by gene duplication in Hyperoartia and exon retrotransposition in Chondrichthyes, concomitant with the change from a simple kidney structure pronephros to a more complex mesonephros with tubules (41). Together, all these features are likely the basis of the isozyme diversity observed in vertebrates. We also identified a glutaminase ancestor that was more similar to the human kidney-type glutaminase and gave rise to the liver-type glutaminase, whereas the appearance of GAC-like glutaminases, with shorter C termini, was a late event in glutaminase evolution. GAC is the most active isoform compared with KGA and LGA, a feature linked to its capacity to assemble into long filament-like superstructures (23). In the present study, we found that this feature is possible only because GAC lacks the original C-terminal bulky ANK repeats of parental KGA, because the  $\Delta$ C-terminal construct assembles in longer fila-

ments and is as active as GAC. Therefore, we propose that filament formation is a gain-of-function characteristic of vertebrate glutaminases, which has been positively selected to create a more active enzyme due to the substrate channeling effect.

Our group and others have previously published the structure of the N-terminal EF-hand-like and glutaminase domains of GLS (20, 42), but the structure of the C-terminal ANK repeats remained unsolved. Here, we provide novel crystal structures for the ANK domains of KGA and GLS2, which surprisingly form an atypical dimer. ANK repeats usually mediate the interaction of a protein with a different partner. However, KGA.ANK and GLS2.ANK mediated the formation of homooligomers, a rare feature. A survey of structures in the Protein Data Bank revealed 68 unique crystal structures of ANK-containing proteins, containing between 2 to 24 ANK repeats (supplemental Table S1). Of these crystal structures, only 5 displayed an ANK-to-ANK association (supplemental Fig. S1). However, none contained an interface that resembles the glutaminase ANK dimer (contact between the DYD and DRW motifs). Therefore, we propose that the ANK dimer association described here is structurally unique to glutaminases and was selected as a conserved feature throughout glutaminase evolution. The explanation for this positive selection remains elusive.

The short regions located immediately after the ANK repeats in KGA and GLS2 (KEN and ESMV motif, respectively) are involved in E3 ubiquitin ligase-dependent degradation (8) and direct interactions with PDZ domain-containing proteins (43), respectively. An analysis of protein complexes including ANK-containing proteins and their interaction partners (44–47)

## The multidomain architecture of human glutaminases

showed that most of the contacts necessary for ANK-to-ANK interaction involves the ANK groove (the concave surface). Although most of the known cases involve the concave surface, some ANK repeats mediate protein-protein contacts through the convex face. One example is the human oncoprotein g-ankyrin, which was crystallized in complex with an antibody fragment. In this structure, the interaction occurred through the outer  $\alpha$ -helices of the ANK4-ANK6 repeats (48). The vaccinia virus K1 protein consists entirely of ANK repeats that are involved in interactions mediated by the convex surface (49). In addition, the VPS9-domain ANK repeat protein binds to Rab32 through the convex side of its ANK repeat (50).

In this regard, although the concave faces of the ANK repeats of KGA and GLS2 make contacts within the glutaminase, we predict that the ANK repeats may still mediate protein-protein interactions through the convex surface. Recently, GAB has been shown to bind and inhibit the small GTPase Rac1 by preventing its interaction with a guanine-nucleotide exchange factor, an interaction that involves the C terminus of GLS2 (residues 464 to 602) (40). Because the region contains the ANK repeats of GAB, we predict that the contact involves the ANK concave face. Considering the long list of partners that were shown to interact with glutaminases using mass spectrometry-based approaches (51), further investigations are required to confirm the mechanism by which the ANK repeats in glutaminases mediate protein-protein interactions.

Last, although at low resolution, the novel ANK-containing structure of human KGA confirms our previous hypothesis that the long-range propagation of filaments for this isoform was thwarted by the presence of the ankyrin repeats themselves (23). As observed in the crystallographic model, the ANK dimer occludes the formation of the filamentation groove, by which the single strand filament of the GAC isoform grows indefinitely, via end to end interaction between pairs of N-terminal domains. Because GAC possesses a shorter unstructured C-terminal sequence, this hypothesis is confirmed when the ANK domain is completely deleted in a mutant construct, which can form longer polymeric species, and therefore, resulting in a more active protein.

### Experimental procedures

#### Phylogenetic reconstruction

726 Glutaminase protein sequences obtained from GenBank<sup>TM</sup> (52), by performing a BLAST search with human glutaminase domain sequence from *GLS* as the query, were aligned using kalign (53). Maximum likelihood phylogenetic reconstruction was performed with RaxML (54) with 260 bootstrap pseudoreplicates and using automated model search. Replicates were summarized with sumtrees from DendroPy (55). The generated phylogenetic tree and respective alignment were used to reconstruct ancestral states for specific nodes using maximum parsimony, as implemented in phangorn (56). Heatmaps from pairwise distances were constructed in SDT (57) using MUSCLE (58) alignment.

#### Evolution of glutaminase exon structure

Protein sequence corresponding to human *GLS* exon 14 was used as input of translated blast (tblastn) searches against avail-

able genomes from *C. intestinalis* (27), *B. floridae* (28), *P. marinus* (29), *C. milii* (30), *D. rerio* (31), and *X. tropicalis* (32). After an initial Reciprocal Best Hit (59) region was found, the correct position of the exon was determined by pairwise LALIGN between sequences (60). Regions downstream from the genomic sequence homologous to exon 14 were also evaluated by LALIGN against human GAC-exclusive exon (exon 15). As both *P. marinus* glutaminases are incomplete, available exons were found in contig using the LALIGN approach. The *GLS* region comprising intron 14, exon 15, and intron 15 was used as input of TranspoGene (61) retrotransposon searching.

#### Protein expression and purification

The constructs of human KGA Val<sup>124</sup>-Leu<sup>669</sup> and Val<sup>551</sup>-Leu<sup>669</sup> were amplified from the pcDNA3.1/hKGA-V5 clone, kindly provided by Dr. Richard Cerione (Cornell University, Ithaca, NY), and subcloned into the pET28a plasmid (Novagen) using the NdeI and XhoI restriction sites with N-terminal His<sub>6</sub> tag. KGA Val<sup>124</sup>-Arg<sup>550</sup> was generated from the pET28a KGA Val<sup>124</sup>-Leu<sup>669</sup> construct by site-directed mutagenesis of the Val<sup>551</sup> residue into a stop codon (TAA) using the QuikChange II XL Site-directed Mutagenesis Kit (Agilent Technologies). The construct of human GAB Lys<sup>485</sup>-Val<sup>602</sup> was cloned into pNIC28-Bsa4 plasmid. The construct of GAC Met<sup>128</sup>-Ser<sup>603</sup> was amplified from a mouse fetal brain tissue cDNA library and cloned into the pET28a plasmid. The four constructs were transformed into *Escherichia coli* Rosetta-2 thermocompetent cells (Merck). Overnight cultures, grown in LB medium in the presence of 50  $\mu\text{g ml}^{-1}$  of kanamycin and 50  $\mu\text{g ml}^{-1}$  of chloramphenicol, were inoculated in a ratio of 1:100 in 3-liter cultures supplemented with the same antibiotics and left shaking at 200 rpm for 5 h at 37 °C. The cultures were then down-tempered to 18 °C for 1 h before induction with 200 nM isopropyl  $\beta$ -D-1-thiogalactopyranoside for 16 h at 18 °C. Cells were collected by rapid centrifugation and resuspended in 500 mM NaCl, 50 mM Tris HCl, pH 8.5 (or 50 mM HEPES, pH 7.5, for GAB construct), 10% glycerol, and 2 mM  $\beta$ -mercaptoethanol and 1 mM PMSF (phenylmethylsulfonyl fluoride). Cell lysis was performed chemically, by incubation with hen egg white lysozyme, DNase I, and deoxycholate (all three reagents from Sigma) for about 1 h, incubated on ice. The soluble fractions were separated from the debris by high speed centrifugation and subsequently loaded, by gravity and in a cold room, on immobilized metal affinity column, Co<sup>2+</sup>-charged TALON (Clontech) for the GLS constructs and nickel-nitrilotriacetic acid Superflow (Qiagen) for the GAB construct, previously equilibrated with the running buffer 10 mM NaCl and 50 mM Tris-HCl, pH 8.5, HEPES, pH 7.5. The constructs were eluted stepwise using running buffer to which up to 500 mM imidazole (v/v) had been added. For the GLS constructs, the tag was removed by overnight digestion with bovine thrombin (Sigma) and the samples loaded into a HiTrap Q HP anion exchange chromatography column (GE Healthcare). Elution was done by performing a linear gradient with a buffer containing 1 M NaCl, 50 mM Tris-HCl, pH 8.5, and 2 mM  $\beta$ -mercaptoethanol. The fractions containing the GLS constructs were loaded in a HiLoad 16/600 Superdex 200 (for KGA Val<sup>124</sup>-Leu<sup>669</sup>, KGA Val<sup>124</sup>-Arg<sup>550</sup>, and GAC Met<sup>128</sup>-Ser<sup>603</sup> constructs) or 75  $\mu\text{g}$  (GE

## The multidomain architecture of human glutaminases

Healthcare) (for KGA Val<sup>551</sup>-Leu<sup>669</sup>). The final buffer conditions were 150 mM NaCl, 30 mM HEPES, pH 8, and 0.5 mM tris(2-carboxyethyl)phosphine). For the GAB construct, the eluate from the affinity chromatography was directly loaded in a HiLoad 16/600 75 µg. Protein concentration was determined by UV 280 nm using calculated extinction coefficients. The hydrodynamic parameters (Stokes radii, Rh) of human KGA, GAC, and the deletion mutant ΔC were determined by gel filtration chromatography using a prepacked Superdex 200 HR 10/30 column (GE Healthcare) in 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine). For each purified protein, ~1 mg ml<sup>-1</sup> concentration in a 500-µl volume was injected into the column. To induce the formation of the higher-order oligomeric species, 20 mM K<sub>2</sub>HPO<sub>4</sub> as a final concentration was added in the protein solution before loading it onto the column. The gel filtration buffer was also supplemented with 20 mM K<sub>2</sub>HPO<sub>4</sub>. The flow rate was maintained at 0.5 ml min<sup>-1</sup>. To establish the hydrodynamic radius (Rh) and elution volume relationship of the protein, proteins of known Rh were run on the same column. The proteins used as standards were ferritin (440 kDa, Rh = 60.8 Å), aldolase (158 kDa, Rh = 48.1 Å), conalbumin (75 kDa, Rh = 36.4 Å), ovalbumin (44 kDa, Rh = 30.5 Å), and ribonuclease A (13.7 kDa, Rh = 15.9 Å).

### Crystallization

Following size-exclusion purification, KGA Val<sup>124</sup>-Leu<sup>669</sup> was concentrated using an Amicon 30-kDa cutoff concentrator (Millipore) to a final concentration of ~7.5 mg ml<sup>-1</sup>. Crystallization experiments were performed at 277 K using the conventional sitting drop vapor diffusion technique. Drops were made by mixing two parts of protein previously incubated with 1.2 mM BPTES to one part of the well solution, containing 1.8 M sodium formate, 0.5 M NaCl, and 0.1 M BIS-TRIS propane, pH 6.8. Before data collection at cryogenic temperature (100 K), harvested crystals were cryoprotected with 10% ethylene glycol added to the mother liquor.

For large-scale limited proteolysis, KGA Val<sup>551</sup>-Leu<sup>669</sup> after a size-exclusion purification step was incubated with 1:100 trypsin (Sigma) at 23 °C for 20 min. The proteolysis was stopped by the addition of 1.5 mM PMSF. For limited proteolysis of GAB Lys<sup>485</sup>-Val<sup>602</sup>, purified protein was incubated overnight with 1:1000 α-chymotrypsin at 23 °C. The digested fragments were immediately purified by size-exclusion chromatography. For crystallization trials, digested KGA Val<sup>551</sup>-Leu<sup>669</sup> and GAB Lys<sup>485</sup>-Val<sup>602</sup> were concentrated, respectively, to 25 and 50 mg ml<sup>-1</sup> using Amicon 10-kDa cutoff concentrators.

Both constructs were crystallized by sitting drop vapor diffusion, by mixing equal parts of protein solution and mother liquor. KGA Val<sup>551</sup>-Leu<sup>669</sup> crystals were grown in (a) 3.2 M NaCl, 0.1 M Tris, pH 8.5 and (b) 3.8 M NH<sub>4</sub>Cl, 0.1 M BIS-TRIS propane, pH 7. GAB Lys<sup>485</sup>-Val<sup>602</sup> crystals were grown in 1.1 M trisodium citrate, 0.1 M imidazole, pH 8, 20 mM L-glutamine.

### X-ray crystallography

Diffraction data were collected at beamlines I03 at the Diamond Light Source (UK) and 12-2 at the Stanford Synchrotron Radiation Lightsource, respectively, for KGA Val<sup>124</sup>-Leu<sup>669</sup> and

for KGA/GAB ANK crystals. Datasets were integrated using Mosflm (62) (for native datasets) and XDS (63) (for sulfur-SAD dataset) and scaled with Aimless (64). The first set of phases of KGA Val<sup>124</sup>-Leu<sup>669</sup> was obtained by the molecular replacement technique as implemented in the program Phaser (65), using the coordinates of the mouse GAC isoform (PDB code 3SS3). KGA ANK was solved by sulfur-SAD using SHELX (66) and the model was refined using higher resolution native datasets. The model obtained for KGA was employed as a search model for solving the GAB ANK structure by molecular replacement (Table 1). Positional and B-factor refinement cycles, as well as solvent modeling, were performed with Refmac (67) and Phenix (68), followed by visual inspection using COOT (69).

### Glutaminase activity assay

To obtain the kinetic parameters for KGA Val<sup>124</sup>-Leu<sup>669</sup>, KGA Val<sup>124</sup>-Arg<sup>550</sup>, and GAC Met<sup>128</sup>-Ser<sup>603</sup>, a mixture containing 10 nM glutaminase, 50 mM Tris acetate, pH 8.6, 3 units of bovine L-glutamate dehydrogenase (Sigma), 2 mM NAD (Sigma) was pipetted into 96-well plates previously filled with 6 or 12 serial dilutions of L-glutamine, to achieve a range of concentrations from 60 to 0.15 mM. K<sub>2</sub>HPO<sub>4</sub> (2 M stock, pH 9.4) was added to the mixture at a final concentration of 20 mM. The formation of NADH was tracked by absorbance readings at 340 nm, for up to five consecutive minutes, at room temperature. Measurements were done in triplicate. The initial velocities, in picomoles of NADH produced per second, were calculated using an extinction coefficient for NADH of 6,220 M<sup>-1</sup> cm<sup>-1</sup> at 340 nm and 0.5 cm of path length. The total volume per reaction was 200 µl. Plate-reader used was an EnSpire (PerkinElmer Life Sciences). Measurements were done in triplicates and analyzed using GraphPad Prism 5.00 (GraphPad Software, San Diego, CA).

### Small-angle X-ray scattering

Scattering data were collected at λ = 1.488 Å, for sample-detector distances of 1.1 m covering the momentum transfer ranges 0.015 < s < 0.442 Å<sup>-1</sup> (s = 4π sinθ/λ, where 2θ is the scattering angle). The data were normalized to the intensity of the incident beam and corrected for the detector response using an in-house program. Two frames of 250 s were collected and compared for radiation damage using the program PRIMUS (70). The same program was used to average the frames and subtract the buffer. The different protein concentrations were evaluated for aggregation by following increases in the measured R<sub>g</sub> (radius of gyration) as calculated by auto R<sub>g</sub>. The R<sub>g</sub> was confirmed by using the indirect Fourier transform program GNOM (71), which was also used to calculate the distribution function P(r) and D<sub>max</sub>. The data were analyzed and processed, including *ab initio* construction and model averaging, using the programs contained in the ATSAS package (72).

### Cryo-electron microscopy

For visualization of cryogrids, purified GAB samples were frozen onto a Gatan 626 sample holder, prepared with FEI Vitrobot Mark II (force of -5 for 2s). Images were acquired using a JEM 2100 (200 kV) electron microscope with a LaB6 filament Gatan 4k × 4k slow scan CCD camera (US4000). Exposure time

## The multidomain architecture of human glutaminases

was 1 s/frame, with a dose of  $20 e \text{ \AA}^{-2} \text{ s}^{-1}$ . The micrographs were processed using IMAGIC (73) and EMAN 2.1 (74). Iterative stable alignment and clustering (ISAC) (75) was used to generate reference-free class-averages from both the IMAGIC stack (25,000 particles) and the EMAN2.1 stack (33,608 particles). Using IMAGIC and UCSF Chimera (76), 10,000 projection images in random orientations of the GAC crystallographic model (PDB code 3SS3) were generated, in both dimer and tetramer configurations. The bounding rectangle dimensions of 200 particles randomly extracted from the IMAGIC, EMAN2.1, ISAC/IMAGIC, and ISAC/EMAN2.1 datasets (50 from each) were classified according to longer (L) and shorter (S) dimensions. Images were low-pass filtered to reduce noise. Only the particles contained in the 50 “best” class averages generated by each program were considered (total: 200 particles). Measurements were performed in Digital Micrograph based on the integration profile across each perpendicular direction, as shown in Fig. 5C. Measurements were taken as the distance, in pixels, between the valleys confining the particle signal. For comparison, we also measured 200 randomly selected projections of the GAC atomic model in dimer and also in tetramer configuration.

**Author contributions**—S. M. G. D. and A. L. B. A. conceived and initiated the research; C. C. P., Z. I., D. A., I. M. F., R. D. R., J. B., R. V. P., and A. G. conducted the research; all authors analyzed and discussed the data; C. C. P., Z. I., D. A., S. M. G. D., and A. L. B. A. wrote the manuscript. All authors revised and approved the final manuscript.

**Acknowledgments**—We thank LNBio for accessibility to core facilities as well as for financial support. We are very grateful to Dr. Alessandra Girasole for expert technical support. We thank the Biotechnology group at the SGC Oxford for cloning and expression testing of glutaminase constructs. We thank Dr. Artur T. Cordeiro and Américo T. Ranzani for help with crystallographic data collection. We thank Dr. Wah Chiu for access to the cryo-electron microscopy facility at Baylor College of Medicine. We acknowledge the staff of all the beamlines used in this work. The KGA.ANK and GLS2.ANK data were collected during the RapiData 2015 Course for Data Collection and Structure Solving. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and National Institutes of Health, National Institute of General Medical Sciences Grant P41GM103393.

### References

- Bergström, J., Fürst, P., Norée, L. O., and Vinnars, E. (1974) Intracellular free amino acid concentration in human muscle tissue. *J. Appl. Physiol.* **36**, 693–697
- DeBerardinis, R. J., and Cheng, T. (2010) Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene* **29**, 313–324
- Levine, A. J., and Puzio-Kuter, A. M. (2010) The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. *Science* **330**, 1340–1344
- Altman, B. J., Stine, Z. E., and Dang, C. V. (2016) From Krebs to clinic: glutamine metabolism to cancer therapy. *Nat. Rev. Cancer* **16**, 619–634
- Wise, D. R., DeBerardinis, R. J., Mancuso, A., Sayed, N., Zhang, X.-Y., Pfeiffer, H. K., Nissim, I., Daikhin, E., Yudkoff, M., McMahon, S. B., and Thompson, C. B. (2008) Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18782–18787
- Rathore, M. G., Saumet, A., Rossi, J.-F., de Bettignies, C., Tempé, D., Lecellier, C.-H., and Villalba, M. (2012) The NF- $\kappa$ B member p65 controls glutamine metabolism through miR-23a. *Int. J. Biochem. Cell Biol.* **44**, 1448–1456
- Wang J. B., Erickson, J. W., Fuji, R., Ramachandran, S., Gao, P., Dinavahi, R., Wilson, K. F., Ambrosio, A. L., Dias, S. M., Dang, C. V., and Cerione, R. A. (2010) Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. *Cancer Cell.* **18**, 207–219
- Colombo, S. L., Palacios-Callender, M., Frakich, N., De Leon, J., Schmitt, C. A., Boorn, L., Davis, N., and Moncada, S. (2010) Anaphase-promoting complex/cyclosome-Cdh1 coordinates glycolysis and glutaminolysis with transition to S phase in human T lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18868–18873
- Lukey, M. J., Greene, K. S., Erickson, J. W., Wilson, K. F., and Cerione, R. A. (2016) The oncogenic transcription factor c-Jun regulates glutaminase expression and sensitizes cells to glutaminase-targeted therapy. *Nat. Commun.* **7**, 11321
- Suzuki, S., Tanaka, T., Poyurovsky, M. V., Nagano, H., Mayama, T., Ohkubo, S., Lokshin, M., Hosokawa, H., Nakayama, T., Suzuki, Y., Sugano, S., Sato, E., Nagao, T., Yokote, K., Tatsuno, I., and Prives, C. (2010) Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7461–7466
- Hu, W., Zhang, C., Wu, R., Sun, Y., Levine, A., and Feng, Z. (2010) Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7455–7460
- Szeliga, M., and Albrecht, J. (2015) Opposing roles of glutaminase isoforms in determining glioblastoma cell phenotype. *Neurochem. Int.* **88**, 6–9
- Szeliga, M., Bogacińska-Karaš M., Różycka, A., Hilgier, W., Marquez, J., and Albrecht, J. (2014) Silencing of GLS and overexpression of GLS2 genes cooperate in decreasing the proliferation and viability of glioblastoma cells. *Tumour Biol.* **35**, 1855–1862
- Liu, J., Zhang, C., Lin, M., Zhu, W., Liang, Y., Hong, X., Zhao, Y., Young, K. H., Hu, W., and Feng, Z. (2014) Glutaminase 2 negatively regulates the PI3K/AKT signaling and shows tumor suppression activity in human hepatocellular carcinoma. *Oncotarget* **5**, 2635–2647
- Krebs, H. (1935) Metabolism of amino-acids: The synthesis of glutamine from glutamic acid and ammonia, and the enzymic hydrolysis of glutamine in animal tissues. *Biochem. J.* **29**, 1951–1969
- Curthoys, N. P., Kuhlenschmidt, T., Weiss, R. F., and Godfrey, S. S. (1976) Phosphate-dependent activity glutaminase from rat kidney and identity with glutaminase in response to acidosis from other tissues. *Arch. Biochem. Biophys.* **172**, 162–167
- Curthoys, N. P., and Watford, M. (1995) Regulation of glutaminase activity and glutamine metabolism. *Annu. Rev. Nutr.* **15**, 133–159
- Elgadi, K. M., Meguid, R. A., Qian, M., Souba, W. W., and Abcouwer, S. F. (1999) Cloning and analysis of unique human glutaminase isoforms generated by tissue-specific alternative splicing. *Physiol. Genomics* **1**, 51–62
- Martín-Rufián, M., Tosina, M., Campos-Sandoval, J. A., Manzanares, E., Lobo, C., Segura, J. A., Alonso, F. J., Matés, J. M., and Márquez, J. (2012) Mammalian glutaminase *Gls2* gene encodes two functional alternative transcripts by a surrogate promoter usage mechanism. *PLoS ONE* **7**, e38380
- Cassago, A., Ferreira, A. P., Ferreira, I. M., Fornezari, C., Gomes, E. R., Greene, K. S., Pereira, H. M., Garratt, R. C., Dias, S. M., and Ambrosio, A. L. (2012) Mitochondrial localization and structure-based phosphate activation mechanism of glutaminase C with implications for cancer metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1092–1097
- Campos-Sandoval, J. A., López de la Oliva, A. R., Lobo, C., Segura, J. A., Matés, J. M., Alonso, F. J., and Márquez, J. (2007) Expression of functional human glutaminase in baculovirus system: affinity purification, kinetic and molecular characterization. *Int. J. Biochem. Cell Biol.* **39**, 765–773

## The multidomain architecture of human glutaminases

22. Olalla, L., Gutiérrez, A., Campos, J. A., Khan, Z. U., Alonso, F. J., Segura, J. A., Márquez, J., and Aledo, J. C. (2002) Nuclear localization of L-type glutaminase in mammalian brain. *J. Biol. Chem.* **277**, 38939–38944
23. Ferreira, A. P., Cassago, A., Gonçalves Kde, A., Dias, M. M., Adamoski, D., Ascensão, C. F., Honorato, R. V., de Oliveira, J. F., Ferreira, I. M., Fornezari, C., Bettini, J., Oliveira, P. S., Paes Leme, A. F., Portugal, R. V., Ambrosio, A. L., and Dias, S. M. (2013) Active glutaminase C self-assembles into a supratetrameric oligomer that can be disrupted by an allosteric inhibitor. *J. Biol. Chem.* **288**, 28009–28020
24. Brown, G., Singer, A., Proudfoot, M., Skarina, T., Kim, Y., Chang, C., Dementieva, I., Kuznetsova, E., Gonzalez, C. F., Joachimiak, A., Savchenko, A., and Yakunin, A. F. (2008) Functional and structural characterization of four glutaminases from *Escherichia coli* and *Bacillus subtilis*. *Biochemistry* **47**, 5724–5735
25. de Beer, T., Carter, R. E., Lobel-Rice, K. E., Sorkin, A., and Overduin, M. (1998) Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. *Science* **281**, 1357–1360
26. Pace, N. R. (2009) Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* **73**, 565–576
27. Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M., Harafuji, N., Hastings, K. E., Ho, I., Hotta, K., Huang, W., et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–2167
28. Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutiérrez, E. L., Dubchak, I., Garcia-Fernández, J., Gibson-Brown, J. J., Grigoriev, I. V., et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071
29. Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., Yandell, M. D., Manousaki, T., Meyer, A., Bloom, O. E., Morgan, J. R., Buxbaum, J. D., Sachidanandam, R., Sims, C., Garruss, A. S., et al. (2013) Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* **45**, 415–421
30. Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., Ohta, Y., Flajnik, M. F., Sutoh, Y., Kasahara, M., Hoon, S., Gangu, V., Roy, S. W., Irimia, M., Korzh, V., et al. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179
31. Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., et al. (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503
32. Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., Blitz, I. L., Blumberg, B., Dichmann, D. S., Dubchak, I., Amaya, E., et al. (2010) The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636
33. Sedgwick, S. G. and Smerdon, S. J. (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316
34. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C., and Peng, Z.-Y. (2004) The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**, 1435–1448
35. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
36. Robinson, M. M., McBryant, S. J., Tsukamoto, T., Rojas, C., Ferraris, D. V., Hamilton, S. K., Hansen, J. C., and Curthoys, N. P. (2007) Novel mechanism of inhibition of rat kidney-type glutaminase by bis-2-(5-phenylacetamido-1,2,4-thiadiazol-2-yl)ethyl sulfide (BPTES). *Biochem. J.* **406**, 407–414
37. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394
38. Bergthorsson, U., Andersson, D. I., and Roth, J. R. (2007) Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci.* **104**, 17004–17009
39. Mosavi, L. K., Minor, D. L., Jr., and Peng, Z.-Y. (2002) Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci.* **99**, 16029–16034
40. Zhang, C., Liu, J., Zhao, Y., Yue, X., Zhu, Y., Wang, X., Wu, H., Blanco, F., Li, S., Haffty, B. G., Hu, W., and Feng, Z. (2016) Glutaminase 2 is a novel negative regulator of small GTPase Rac1 and mediates p53 function in suppressing metastasis. *Elife* **5**, e10727
41. Scott, A. P. (1987) Reproductive endocrinology of fish. In *Fundamentals of Comparative Vertebrate Endocrinology* (Chester-Jones, I., Ingleton, P. M., Philips, and J. G., eds), pp. 223–256, Springer US, New York, NY
42. DeLaBarre, B., Gross, S., Fang, C., Gao, Y., Jha, A., Jiang, F., Song, J. J., Wei, W., and Hurov, J. B. (2011) Full-length human glutaminase in complex with an allosteric inhibitor. *Biochemistry* **50**, 10764–10770
43. Olalla, L., Gutiérrez, A., Jiménez, A. J., López-Téllez, J. F., Khan, Z. U., Pérez, J., Alonso, F. J., de la Rosa, V., Campos-Sandoval, J. A., Segura, J. A., Aledo, J. C., and Márquez, J. (2008) Expression of the scaffolding PDZ protein glutaminase-interacting protein in mammalian brain. *J. Neurosci. Res.* **86**, 281–292
44. Russo, A. A., Tong, L., Lee, J.-O., Jeffrey, P. D., and Pavletich, N. P. (1998) Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* **395**, 237–243
45. Chiswell, B. P., Zhang, R., Murphy, J. W., Boggon, T. J., and Calderwood, D. A. (2008) The structural basis of integrin-linked kinase-PINCH interactions. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20677–20682
46. Takeda, S., Minakata, S., Koike, R., Kawahata, I., Narita, A., Kitazawa, M., Ota, M., Yamakuni, T., Maéda, Y., and Nitanai, Y. (2010) Two distinct mechanisms for actin capping protein regulation: steric and allosteric inhibition. *PLoS Biol.* **8**, e1000416
47. Holzer, S., Ban, N., and Klinge, S. (2013) Crystal structure of the yeast ribosomal protein rpS3 in complex with its chaperone Yar1. *J. Mol. Biol.* **425**, 4154–4160
48. Robin, G., Sato, Y., Desplancq, D., Rochel, N., Weiss, E., and Martineau, P. (2014) Restricted diversity of antigen binding residues of antibodies revealed by computational alanine scanning of 227 antibody–antigen complexes. *J. Mol. Biol.* **426**, 3729–3743
49. Li, Y., Meng, X., Xiang, Y., and Deng, J. (2010) Structure function studies of vaccinia virus host range protein k1 reveal a novel functional surface for ankyrin repeat proteins. *J. Virol.* **84**, 3331–3338
50. Hesketh, G. G., Pérez-Dorado, I., Jackson, L. P., Wartosch, L., Schäfer, I. B., Gray, S. R., McCoy, A. J., Zeldin, O. B., Garman, E. F., Harbour, M. E., Evans, P. R., Seaman, M. N., Luzio, J. P., and Owen, D. J. (2014) VARP is recruited on to endosomes by direct interaction with retromer, where together they function in export to the cell surface. *Dev. Cell.* **29**, 591–606
51. Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., et al. (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440
52. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2016) GenBank. *Nucleic Acids Res.* **44**, D67–D72
53. Lassmann, T., and Sonnhammer, E. L. (2005) Kalign: an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* **6**, 298
54. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313
55. Sukumaran, J., and Holder, M. T. (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571
56. Schliep, K. P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593
57. Muhire, B. M., Varsani, A., and Martin, D. P. (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE* **9**, e108277
58. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797
59. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families. *Science* **278**, 631–637
60. Huang, X., and Miller, W. (1991) A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12**, 337–357

## The multidomain architecture of human glutaminases

61. Levy, A., Sela, N., and Ast, G. (2008) TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. *Nucleic Acids Res.* **36**, D47–D52
62. Leslie, A. G. W., and Powell H. R. (2007) Processing diffraction data with mosflm. In *Evolving methods for macromolecular crystallography* (Read, R. J., and Sussman, J. L., eds), pp. 41–51, Springer, Dordrecht, Netherlands
63. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
64. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214
65. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
66. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485
67. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195
68. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
69. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
70. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282
71. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
72. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1: towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40**, s223–s228
73. van Heel, M., Portugal, R., Rohou, A., Linnemayr, C., Bebeacqua, C., Schmidt, R., Grant, T., and Schatz, M. (2012) Four-dimensional cryo-electron microscopy at quasi-atomic resolution: IMAGIC 4D. *Int. Tables. Acta Crystallogr.* **Vol. F**, ch. 19.9, 624–628
74. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46
75. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J., and Penczek, P. A. (2012) Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237–247
76. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612

### ANEXO 3

Ferreira, A. P. S. A. P. S., Cassago, A., Gonçalves, K. D. A., Dias, M. M., Adamoski, D., Ascensão, C. F. R. C. F. R., ... Dias, S. M. G. S. M. G. (2013). Active glutaminase C self-assembles into a supratetrameric oligomer that can be disrupted by an allosteric inhibitor. *The Journal of Biological Chemistry*, 288(39), 28009–20. <https://doi.org/10.1074/jbc.M113.501346>

A política editorial do *Journal of Biological Chemistry* permite a reprodução do texto em teses e dissertações (<http://www.jbc.org/site/misc/edpolicy.xhtml#copyright>).

#### **For noncommercial use:**

Other parties are welcome to copy, distribute, transmit and adapt the work at no cost and without permission for noncommercial use as long as they attribute the work to the original source using the citation above.

Examples of noncommercial use include:

- Reproducing a figure for educational purposes, such as schoolwork or lecture presentations.
- Appending a reprinted article to a Ph.D. dissertation.

# Active Glutaminase C Self-assembles into a Supratetrameric Oligomer That Can Be Disrupted by an Allosteric Inhibitor\*

Received for publication, July 14, 2013, and in revised form, August 6, 2013. Published, JBC Papers in Press, August 8, 2013, DOI 10.1074/jbc.M113.501346

Amanda Petrina Scotá Ferreira<sup>†1</sup>, Alexandre Cassago<sup>§1</sup>, Kaliandra de Almeida Gonçalves<sup>†1</sup>, Marília Meira Dias<sup>†</sup>, Douglas Adamoski<sup>†</sup>, Carolline Fernanda Rodrigues Ascensão<sup>†</sup>, Rodrigo Vargas Honorato<sup>†</sup>, Juliana Ferreira de Oliveira<sup>†</sup>, Igor Monteze Ferreira<sup>†</sup>, Camila Fornezari<sup>†</sup>, Jefferson Bettini<sup>§</sup>, Paulo Sérgio Lopes Oliveira<sup>†</sup>, Adriana Franco Paes Leme<sup>†</sup>, Rodrigo Villares Portugal<sup>§</sup>, Andre Luis Berteli Ambrosio<sup>†2,3</sup>, and Sandra Martha Gomes Dias<sup>†2,4</sup>

From the Laboratórios Nacionais de <sup>†</sup>Biociências e <sup>§</sup>Nanotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas-SP 13083-100, Brazil

**Background:** GAC supplies for increased metabolic needs of tumors because of exclusive localization and kinetic properties.

**Results:** Higher than tetramer oligomers are the active form in *in vitro* and in cellular assays. Bis-2-(5-phenylacetamido-1,3,4-thiadiazol-2-yl)ethyl sulfide disrupts oligomers.

**Conclusion:** A novel molecular mechanism for GAC activation is proposed.

**Significance:** The data affect the development of therapies targeting GAC in tumors, with emphasis on allosteric inhibitors.

The phosphate-dependent transition between enzymatically inert dimers into catalytically capable tetramers has long been the accepted mechanism for the glutaminase activation. Here, we demonstrate that activated glutaminase C (GAC) self-assembles into a helical, fiber-like double-stranded oligomer and propose a molecular model consisting of seven tetramer copies per turn per strand interacting via the N-terminal domains. The loop <sup>321</sup>LRFNKL<sup>326</sup> is projected as the major regulating element for self-assembly and enzyme activation. Furthermore, the previously identified *in vivo* lysine acetylation (Lys<sup>311</sup> in humans, Lys<sup>316</sup> in mouse) is here proposed as an important down-regulator of superoligomer assembly and protein activation. Bis-2-(5-phenylacetamido-1,3,4-thiadiazol-2-yl)ethyl sulfide, a known glutaminase inhibitor, completely disrupted the higher order oligomer, explaining its allosteric mechanism of inhibition via tetramer stabilization. A direct correlation between the tendency to self-assemble and the activity levels of the three mammalian glutaminase isozymes was established, with GAC being the most active enzyme while forming the longest structures. Lastly, the ectopic expression of a fiber-prone superactive GAC mutant in MDA-MB 231 cancer cells provided considerable

proliferative advantages to transformed cells. These findings yield unique implications for the development of GAC-oriented therapeutics targeting tumor metabolism.

Cancer cells have a well established dependence on glutamine metabolism to support their highly proliferative status. Apart from acting as a source for nitrogen as well as reductive power, its anabolic carbon skeletons can be siphoned from the TCA cycle to be used as building blocks for the growing and dividing cells (1, 2). Both Myc and Rho GTPases have been shown to stimulate glutaminase C (GAC),<sup>5</sup> an isozyme that possesses a distinct cellular localization as well as activation levels purported to provide the aforesaid proliferative advantage to cancer cells (3–5).

We have recently described the structural determinants of the phosphate-dependent activation mechanism of GAC, based on the tetramerization-induced lifting of a so-called gating loop (<sup>321</sup>LRFNKL<sup>326</sup>; NCBI sequence NP\_001106854.1), which controls substrate accessibility to the active site. We showed that phosphate binds inside the catalytic pocket, resulting in allosteric stabilization of tetramers and facilitating substrate entry by outcompeting with the product, glutamate, to guarantee enzyme cycling (5). Additionally, recent publications have provided structural insights into glutaminase inhibition by making use of the small molecules BPTES and 968 (6–9). Nevertheless, the precise descriptions of their modes of inhibition are still lacking.

Renewed interest in cancer metabolism has prompted an innovative warfront against metabolic enzymes, aiming at the development of alternative and efficient therapeutic opportunities. Glutaminase C is a key target in this sense (2, 4, 10, 11),

\* This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grants 2009/10875-9 (to S. M. G. D.) and 2010/05003-0 and 2012/14298-9 (to A. L. B. A.) and Fellowships 2010/05987-0 (to A. P. S. F.), 2010/05987-0 (to A. C.), 2011/06654-7 (to K. A. G.), and 2009/54067-3 (to A. F. P. L.). This work was also supported in part by the Laboratório Nacional de Biociências.

The atomic coordinates and structure factors (code 4jkt) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> These authors contributed equally to this work.

<sup>3</sup> To whom correspondence may be addressed: LNBio, CNPEM, Rua Giuseppe Máximo Solfaro, 10.000, Pólo II de Alta Tecnologia, Campinas, SP, 13083-100, Brazil. Tel.: 55-19-3512-1115; Fax: 55-19-3512-1004; E-mail: andre.ambrosio@lnbio.cnpem.br.

<sup>4</sup> To whom correspondence may be addressed: LNBio, CNPEM, Rua Giuseppe Máximo Solfaro, 10.000, Pólo II de Alta Tecnologia, Campinas, SP, 13083-100, Brazil. Tel.: 55-19-3512-1115; Fax: 55-19-3512-1004; E-mail: sandra.dias@lnbio.cnpem.br.

<sup>5</sup> The abbreviations used are: GAC, glutaminase C; BPTES, bis-2-(5-phenylacetamido-1,3,4-thiadiazol-2-yl)ethyl sulfide; TEM, transmission electron microscopy; KGA, kidney-type glutaminase; LGA, liver-type glutaminase; DSS, disuccinimidyl suberate.

## ANEXO 4

**de O Barsottini, M. R.\*, de Oliveira, J. F.\*, Adamoski, D.\***, Teixeira, P. J. P. L., do Prado, P. F. V, Tiezzi, H. O., ... Ambrosio, A. L. B. (2013). Functional diversification of cerato-platanins in *Moniliophthora perniciosa* as seen by differential expression and protein function specialization. *Molecular Plant-Microbe Interactions*, 26(11), 1281–1293. <https://doi.org/10.1094/MPMI-05-13-0148-R>

### **\*Contribuição como co-primeiro autor.**

A política editorial do grupo *APS Journals* permite a reprodução do texto em teses e dissertações ([https://apsjournals.apsnet.org/page/Copyright Permissions](https://apsjournals.apsnet.org/page/Copyright%20Permissions)).

### **Rights and Permissions**

All APS copyrighted materials must be used in accord with current United States and international copyright law, except as specifically stated or licensed otherwise by APS. These copyright restrictions apply to publications in all media including print, photocopy, photographic, digital, and online. See [APS Intellectual Property Rights Policies](#) for details.

Except as stated below, authors wishing to reproduce materials copyrighted by APS must obtain permission to do so in writing. Permissions can be obtained through the [Copyright Clearance Center](#).

1. Authors of APS-copyrighted material who wish to reprint figures or tables in their own other personal works and in other works of the authors' institutions may do so without restriction. The reprinted material should be credited as follows:

Reprinted from [authors], [year of publication], [title], [journal] [volume]:[page range].

Reprinted from **de O Barsottini, M. R.\*, de Oliveira, J. F.\*, Adamoski, D.\***, Teixeira, P. J. P. L., do Prado, P. F. V, Tiezzi, H. O., ... Ambrosio, A. L. B., 2013, Functional diversification of cerato-platanins in *Moniliophthora perniciosa* as seen by differential expression and protein function specialization, *Molecular Plant-Microbe Interactions*, 26(11), 1281-1293.

# Functional Diversification of Cerato-Platanins in *Moniliophthora perniciosa* as Seen by Differential Expression and Protein Function Specialization

Mario R. de O. Barsottini,<sup>1,2</sup> Juliana F. de Oliveira,<sup>2</sup> Douglas Adamoski,<sup>2</sup> Paulo J. P. L. Teixeira,<sup>1</sup> Paula F. V. do Prado,<sup>1,2</sup> Henrique O. Tiezzi,<sup>2</sup> Mauricio L. Sforça,<sup>2</sup> Alexandre Cassago,<sup>3</sup> Rodrigo V. Portugal,<sup>3</sup> Paulo S. L. de Oliveira,<sup>2</sup> Ana C. de M. Zeri,<sup>2</sup> Sandra M. G. Dias,<sup>2</sup> Gonçalo A. G. Pereira,<sup>1,2</sup> and Andre L. B. Ambrosio<sup>2</sup>

<sup>1</sup>Departamento de Genética e Evolução, UNICAMP, Campinas, SP 13083-970, Brazil; <sup>2</sup>Laboratório Nacional de Biotecnologia, CNPEM, Campinas, SP 13083-100, Brazil; <sup>3</sup>Laboratório Nacional de Nanotecnologia, CNPEM, Campinas, SP 13083-100, Brazil

Submitted 8 March 2013. Accepted 22 July 2013.

Cerato-platanins (CP) are small, cysteine-rich fungal-secreted proteins involved in the various stages of the host–fungus interaction process, acting as phytotoxins, elicitors, and allergens. We identified 12 CP genes (MpCP1 to MpCP12) in the genome of *Moniliophthora perniciosa*, the causal agent of witches' broom disease in cacao, and showed that they present distinct expression profiles throughout fungal development and infection. We determined the X-ray crystal structures of MpCP1, MpCP2, MpCP3, and MpCP5, representative of different branches of a phylogenetic tree and expressed at different stages of the disease. Structure-based biochemistry, in combination with nuclear magnetic resonance and mass spectrometry, allowed us to define specialized capabilities regarding self-assembling and the direct binding to chitin and N-acetyl-

glucosamine (NAG) tetramers, a fungal cell wall building block, and to map a previously unknown binding region in MpCP5. Moreover, fibers of MpCP2 were shown to act as expansin and facilitate basidiospore germination whereas soluble MpCP5 blocked NAG6-induced defense response. The correlation between these roles, the fungus life cycle, and its tug-of-war interaction with cacao plants is discussed.

The basidiomycete *Moniliophthora perniciosa* is the causal agent of the witches' broom disease (WBD) and considered one of the major fungal plagues of cacao crops (Aime and Phillips-Mora 2005). WBD severely affects the production of cocoa beans, the raw material of chocolate and other derivatives, resulting in great agro-economic losses throughout the cultivating fields of Central and South America (Evans 2007). In this regard, it is urgent to gain knowledge on important proteins for the fungus attack onto the plant.

Defense responses in plants upon a pathogen attack usually result in deposition of callose and production of reactive oxygen species, nitric oxide, and phytoalexins, along with the transcription of defense-related genes (Desender et al. 2007). These responses are promoted by elicitors, which are recognized by plant receptors. They initiate the microbe-associated molecular pattern (MAMP)-triggered immunity or induce a second layer of defense called the effector-triggered immunity (Jones and Dangl 2006).

One family of elicitors spread among the fungal kingdom is the cerato-platanin (CP) family. The founding member of this family and the most studied protein is CP from *Ceratocystis platani* (CpCP) (Pazzagli et al. 1999). CpCP, recently characterized as a MAMP, induces defense responses and tissue necrosis in plants (Fontana et al. 2008; Lombardi et al. 2013; Pazzagli et al. 1999) and may have a structural role in fungal cell wall formation (Bacelli et al. 2012; Boddi et al. 2004). The tridimensional structure of CpCP was determined by nuclear magnetic resonance (NMR) and is similar to proteins that bind to polysaccharides (de Oliveira et al. 2011). The CpCP interacts with chitin (an insoluble polymer of N-acetylglucosamine [NAG]) and soluble NAG oligomers (Bacelli et al. 2013, de Oliveira et al. 2011) and, although it does not bind to cellulose,

Nucleotide sequences were deposited at the GenBank under the accession numbers JX847578 (MpCP6), JX422024 (MpCP7), EU250343 (MpCP8), JX422025 (MpCP9), JX847579 (MpCP10), JX422026 (MpCP11), JX422027 (MpCP12), JX422037 (SMpCP1), JX422028 (SMpCP2), JX422029 (SMpCP3), JX422036 (SMpCP6), JX422030 (SMpCP7), JX422031 (SMpCP8), JX422032 (SMpCP9), JX422033 (SMpCP10), JX422034 (SMpCP11), JX422035 (SMpCP12), JX422041 (CcCP1), JX422043 (CcCP2), JX422039 (CcCP3), JX422042 (CcCP4), JX422040 (CcCP5), JX422044 (CcCP6), JX422038 (MrCP1), JX426104 (MrCP2), JX426107 (MrCP3), JX426108 (MrCP5), JX426106 (MrCP6), JX426113 (MrCP7), JX426110 (MrCP8), JX426105 (MrCP9), JX426109 (MrCP10), JX426112 (MrCP11), and JX426111 (MrCP12).

The atomic coordinates and structure factors under PDB ID 3SUJ (MpCP1), 3SUK (MpCP2), 3SUL (MpCP3), and 3SUM (MpCP5) are available at the Protein Data Bank.

M. R. de O. Barsottini, J. F. de Oliveira, and D. Adamoski contributed equally to this work. S. M. G. Dias and A. L. B. Ambrosio contributed equally to this work.

Corresponding authors: A. L. B. Ambrosio or S. M. G. Dias; E-mail: [andre.ambrosio@lnbio.cnpem.br](mailto:andre.ambrosio@lnbio.cnpem.br) or [sandra.dias@lnbio.cnpem.br](mailto:sandra.dias@lnbio.cnpem.br); Telephone: +(55) 19 3512 1115; Fax: +(55) 19 3512 1004.

\* The e-Xtra logo stands for “electronic extra” and indicates that nine supplementary figures, three supplementary tables, and supplementary information about experimental procedures are published online.

## ANEXO 5

Fala, A. M., Oliveira, J. F., Adamoski, D., Aricetti, J. A., Dias, M. M., Dias, M. V. B., ... Ambrosio, A. L. B. (2015). Unsaturated fatty acids as high-affinity ligands of the C-terminal Per-ARNT-Sim domain from the Hypoxia-inducible factor 3 $\alpha$ . *Scientific Reports*, 5(August), 12698. <https://doi.org/10.1038/srep12698>

A política editorial da *Scientific Reports* licencia o trabalho como *Creative Commons Attribution 4.0*, permitindo a reprodução do texto em teses e dissertações (<https://creativecommons.org/licenses/by/4.0/>).



The image is a banner for the Creative Commons Attribution 4.0 International (CC BY 4.0) license. It features a blue header with the CC and BY icons, followed by the text 'Attribution 4.0 International (CC BY 4.0)'. Below this, a grey section contains a disclaimer, the heading 'You are free to:', and two bullet points: 'Share' and 'Adapt'. A circular seal on the right side of the grey section reads 'Free Cultural Works APPROVED FOR'. At the bottom of the grey section, it states that the licensor cannot revoke these freedoms as long as the user follows the license terms.

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

### You are free to:

- Share** — copy and redistribute the material in any medium or format
- Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.



# SCIENTIFIC REPORTS

OPEN

## Unsaturated fatty acids as high-affinity ligands of the C-terminal Per-ARNT-Sim domain from the Hypoxia-inducible factor 3 $\alpha$

Received: 26 March 2015

Accepted: 07 July 2015

Published: 03 August 2015

Angela M. Fala<sup>1,\*</sup>, Juliana F. Oliveira<sup>1,\*</sup>, Douglas Adamoski<sup>1</sup>, Juliana A. Aricetti<sup>2</sup>, Marília M. Dias<sup>1</sup>, Marcio V. B. Dias<sup>1,†</sup>, Maurício L. Sforça<sup>1</sup>, Paulo S. Lopes-de-Oliveira<sup>1</sup>, Silvana A. Rocco<sup>1</sup>, Camila Caldana<sup>2,3</sup>, Sandra M. G. Dias<sup>1,†</sup> & Andre L. B. Ambrosio<sup>1,†</sup>

Hypoxia-inducible transcription factors (HIF) form heterodimeric complexes that mediate cell responses to hypoxia. The oxygen-dependent stability and activity of the HIF- $\alpha$  subunits is traditionally associated to post-translational modifications such as hydroxylation, acetylation, ubiquitination, and phosphorylation. Here we report novel evidence showing that unsaturated fatty acids are naturally occurring, non-covalent structural ligands of HIF-3 $\alpha$ , thus providing the initial framework for exploring its exceptional role as a lipid sensor under hypoxia.

HIF, closely linked to disorders of the circulatory system and cancer progression<sup>1</sup>, consists of two multidomain subunits including the constitutively expressed HIF-1 $\beta$  and an oxygen-labile HIF-1 $\alpha$  (or paralogs HIF-2 $\alpha$  and -3 $\alpha$ ). The gene encoding for HIF-3 $\alpha$  (*HIF3A*) is exclusively susceptible to post-transcriptional processing in humans, with six isoforms shown to exist at mRNA levels and three already confirmed as functional protein<sup>2,3</sup>. HIF-3 $\alpha$ 9, the longest (669 aa) and canonical isoform<sup>3</sup>, is a potent oxygen-labile transcription factor, tightly regulating the expression of a unique set of genes<sup>4</sup>. HIF-3 $\alpha$ 4 (363 aa), on the other hand, while insensitive to intracellular oxygen levels and incapable of transactivation, is a dominant-negative regulator of HIF-1 $\alpha$ <sup>5</sup>. Due to the use of different transcription initiation sites and a combination of splicing events in *HIF3A*, the C-terminal Per-Arnt-Sim sensor domain (PASb) is the only common architectural feature among the HIF-3 $\alpha$  isoforms<sup>3</sup>. The PASb domains in both HIF- $\alpha$  and - $\beta$  subunits have a crucial role in forming active HIF heterodimers and recruiting co-regulators<sup>6,7</sup>.

PAS members are versatile sensory modules of stimuli such as oxygen tension, redox state and light intensity through binding to cofactors such as heme, flavin adenine dinucleotide and even divalent metal ions<sup>8</sup>. In the case of the HIF proteins, PAS domains-mediated protein-protein interactions have been shown independently of a cofactor and simply based on the presence of the partner PAS<sup>9,10</sup>. However, recently, the identification of a 290 Å<sup>3</sup> polar cavity in HIF-2 $\alpha$  PASb domain led to the proposition that endogenous ligands may exist<sup>11</sup>; this cavity was later exploited for the selection of artificial small molecules that disrupt the full-length HIF-2 heterodimer<sup>10-12</sup>.

Hereby, we present the crystal structure of the recombinant human HIF-3 $\alpha$  PASb, or PASb-3 $\alpha$ , at 1.15 Å maximum resolution (Table 1 and Supplementary Fig. 1). Although with a backbone virtually identical to that of the PASb domains of both HIF-1 $\alpha$ <sup>9</sup> and HIF-2 $\alpha$ <sup>10</sup> (Fig. 1a), PASb-3 $\alpha$  has a unique

<sup>1</sup>Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil, 13083-100. <sup>2</sup>Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil, 13083-100. <sup>3</sup>Max Plack-partner group at the Laboratório Nacional de Ciência e Tecnologia do Bioetanol. <sup>4</sup>Current address: Institute of Biomedical Science, University of São Paulo, São Paulo, SP, 05508-000, Brazil. <sup>\*</sup>These authors contributed equally to this work. <sup>†</sup>These authors jointly supervised this work. Correspondence and requests for materials should be addressed to S.M.G.D. (email: sandra.dias@lnbio.cnpem.br) or A.L.B.A. (email: andre.ambrosio@lnbio.cnpem.br)

## ANEXO 6

Quintero, M.\*, Adamoski, D.\*, Reis, L. M. dos\*, Ascensão, C. F. R., Oliveira, K. R. S. de, Gonçalves, K. de A., ... Dias, S. M. G. (2017). Guanylate-binding protein-1 is a potential new therapeutic target for triple-negative breast cancer. *BMC Cancer*, 17(1), 727. <https://doi.org/10.1186/s12885-017-3726-2>

**\*Contribuição como co-primeiro autor.**

A política editorial da *BMC Cancer* licencia o trabalho como *Creative Commons Attribution 4.0*, permitindo a reprodução do texto em teses e dissertações (<https://creativecommons.org/licenses/by/4.0/>).



## Attribution 4.0 International (CC BY 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

### You are free to:

**Share** — copy and redistribute the material in any medium or format

**Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.



## RESEARCH ARTICLE

## Open Access



# Guanylate-binding protein-1 is a potential new therapeutic target for triple-negative breast cancer

Melissa Quintero<sup>1†</sup>, Douglas Adamoski<sup>1,3†</sup>, Larissa Menezes dos Reis<sup>1,3†</sup>, Carolline Fernanda Rodrigues Ascenção<sup>1,3</sup>, Krishina Ratna Sousa de Oliveira<sup>1,3</sup>, Kalandra de Almeida Gonçalves<sup>1</sup>, Marília Meira Dias<sup>1</sup>, Marcelo Falsarella Carazzolle<sup>2</sup> and Sandra Martha Gomes Dias<sup>1\*</sup>

## Abstract

**Background:** Triple-negative breast cancer (TNBC) is characterized by a lack of estrogen and progesterone receptor expression (*ESR* and *PGR*, respectively) and an absence of human epithelial growth factor receptor (*ERBB2*) amplification. Approximately 15–20% of breast malignancies are TNBC. Patients with TNBC often have an unfavorable prognosis. In addition, TNBC represents an important clinical challenge since it does not respond to hormone therapy.

**Methods:** In this work, we integrated high-throughput mRNA sequencing (RNA-Seq) data from normal and tumor tissues (obtained from The Cancer Genome Atlas, TCGA) and cell lines obtained through in-house sequencing or available from the Gene Expression Omnibus (GEO) to generate a unified list of differentially expressed (DE) genes. Methyome and proteomic data were integrated to our analysis to give further support to our findings. Genes that were overexpressed in TNBC were then curated to retain new potentially druggable targets based on in silico analysis. Knocking-down was used to assess gene importance for TNBC cell proliferation.

**Results:** Our pipeline analysis generated a list of 243 potential new targets for treating TNBC. We finally demonstrated that knock-down of Guanylate-Binding Protein 1 (*GBP1*), one of the candidate genes, selectively affected the growth of TNBC cell lines. Moreover, we showed that *GBP1* expression was controlled by epidermal growth factor receptor (EGFR) in breast cancer cell lines.

**Conclusions:** We propose that *GBP1* is a new potential druggable therapeutic target for treating TNBC with enhanced *EGFR* expression.

**Keywords:** Breast cancer, Triple-negative breast cancer, Gene expression, RNA-Seq, Transcriptomics, Therapeutic target

## Background

The emergence of next-generation sequencing (NGS) technology has provided a large amount of data, much of which is publicly available [1, 2]. Specifically, RNA-Seq has been used for the estimation of RNA abundance [3, 4], alternative splicing detection [5–7], and

the discovery of novel genes and transcripts. As such, RNA-Seq has become an important tool in cancer studies [6], contributing to reduced costs and less time being spent in benchtop experiments, thus speeding up the resolution of biological problems. However, a challenge remains in achieving intelligible data analysis and efficient laboratory validation.

Triple-negative breast cancer (TNBC) is characterized by a lack of estrogen and progesterone receptor expression (*ESR* and *PGR*, respectively) and an absence of human epithelial growth factor receptor (*ERBB2*) amplification. Approximately to 15–20% of breast malignancies are TNBC [8]. Patients with TNBC often

\* Correspondence: sandra.dias@lnbio.cnpem.br

†Equal contributors

<sup>1</sup>Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPem), Campinas, São Paulo 13083-970, Brazil  
Full list of author information is available at the end of the article



## ANEXO 7

**Nagampalli, R.S.K.\***, **Quesñay, J.E.N.\***, **Adamoski, D.\***, Islam, Z., Birch, J., Sebinelli, H. G., Girard, R.M.B.M., Ascensão, C.F.R., Fala, A.M., Pauletti, B.A., Consonni, S.R., de Oliveira, J.F., Silva, A.C.T., Franchini, K.G., Leme, A.F.P., Silver, A.M., Ciancaglini, P., Moraes, I., Dias, S.M.G., Ambrosio, A.L.B. (2018) Human mitochondrial pyruvate carrier 2 as an autonomous membrane transporter. *Scientific Reports*, 8:3510.

### \*Contribuição como co-primeiro autor.

A política editorial da *Scientific Reports* licencia o trabalho como *Creative Commons Attribution 4.0*, permitindo a reprodução do texto em teses e dissertações (<https://creativecommons.org/licenses/by/4.0/>).



The image shows a Creative Commons Attribution 4.0 International (CC BY 4.0) license banner. It features a blue header with the CC and BY icons, followed by the text "Attribution 4.0 International (CC BY 4.0)". Below this, there is a disclaimer: "This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#)." The main body of the banner is light gray and contains the heading "You are free to:" followed by two bullet points: "Share — copy and redistribute the material in any medium or format" and "Adapt — remix, transform, and build upon the material for any purpose, even commercially." At the bottom, it states: "The licensor cannot revoke these freedoms as long as you follow the license terms." On the right side, there is a circular seal that says "Free Cultural Works APPROVED FOR".



## Attribution 4.0 International (CC BY 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

### You are free to:

- Share** — copy and redistribute the material in any medium or format
- Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.



# SCIENTIFIC REPORTS

OPEN

## Human mitochondrial pyruvate carrier 2 as an autonomous membrane transporter

Raghavendra Sashi Krishna Nagampalli<sup>1</sup>, José Edwin Neciosup Quesñay<sup>1</sup>, Douglas Adamoski<sup>1</sup>, Zeyaul Islam<sup>1</sup>, James Birch<sup>2,3</sup>, Heitor Gobbi Sebinelli<sup>4</sup>, Richard Marcel Bruno Moreira Girard<sup>5</sup>, Carolline Fernanda Rodrigues Ascensão<sup>1</sup>, Angela Maria Fala<sup>1,6</sup>, Bianca Alves Pauletti<sup>1</sup>, Sílvio Roberto Consonni<sup>1,7</sup>, Juliana Ferreira de Oliveira<sup>1</sup>, Amanda Cristina Teixeira Silva<sup>1</sup>, Kleber Gomes Franchini<sup>1</sup>, Adriana Franco Paes Leme<sup>1</sup>, Ariel Mariano Silber<sup>5</sup>, Pietro Ciancaglini<sup>4</sup>, Isabel Moraes<sup>2,3,8</sup>, Sandra Martha Gomes Dias<sup>1</sup> & Andre Luis Berteli Ambrosio<sup>1</sup>

The active transport of glycolytic pyruvate across the inner mitochondrial membrane is thought to involve two mitochondrial pyruvate carrier subunits, MPC1 and MPC2, assembled as a 150 kDa heterotypic oligomer. Here, the recombinant production of human MPC through a co-expression strategy is first described; however, substantial complex formation was not observed, and predominantly individual subunits were purified. In contrast to MPC1, which co-purifies with a host chaperone, we demonstrated that MPC2 homo-oligomers promote efficient pyruvate transport into proteoliposomes. The derived functional requirements and kinetic features of MPC2 resemble those previously demonstrated for MPC in the literature. Distinctly, chemical inhibition of transport is observed only for a thiazolidinedione derivative. The autonomous transport role for MPC2 is validated in cells when the ectopic expression of human MPC2 in yeast lacking endogenous MPC stimulated growth and increased oxygen consumption. Multiple oligomeric species of MPC2 across mitochondrial isolates, purified protein and artificial lipid bilayers suggest functional high-order complexes. Significant changes in the secondary structure content of MPC2, as probed by synchrotron radiation circular dichroism, further supports the interaction between the protein and ligands. Our results provide the initial framework for the independent role of MPC2 in homeostasis and diseases related to dysregulated pyruvate metabolism.

Nearly four decades after the demonstration of the protein-mediated transport of pyruvate across the inner mitochondrial membrane (IMM)<sup>1</sup>, two concurrent studies identified the oligomeric complex formed by MPC1 and MPC2 as necessary and sufficient for this task<sup>2,3</sup>; MPC1 and MPC2 were proposed to function together via the formation of an oligomeric structure of approximately 150 kDa<sup>2</sup>. These original findings inspired many subsequent investigations that have further characterized MPC-dependent pyruvate transport in the cellular context<sup>4-6</sup>, with multiple groups showing that either the loss of MPC1 or MPC2 in mitochondria is sufficient to confer

<sup>1</sup>Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, 13083-970, Brazil. <sup>2</sup>Membrane Protein Laboratory, Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, England. <sup>3</sup>Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell, Didcot, Oxfordshire OX11 0FA, England. <sup>4</sup>Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, 14040-901, Brazil. <sup>5</sup>Laboratory of Biochemistry of Tryps – LaBTryps, Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, 05508-000, Brazil. <sup>6</sup>Present address: Structural Genomics Consortium (SGC), Universidade Estadual de Campinas, Campinas, SP, 13083-886, Brazil. <sup>7</sup>Present address: Departamento de Bioquímica e Biologia Tecidual, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, 13083-862, Brazil. <sup>8</sup>Present address: National Physical Laboratory, Teddington, Middlesex, TW11 0LW, England. Raghavendra Sashi Krishna Nagampalli, José Edwin Neciosup Quesñay and Douglas Adamoski contributed equally to this work. Correspondence and requests for materials should be addressed to S.M.G.D. (email: [sandra.dias@lnbio.cnpem.br](mailto:sandra.dias@lnbio.cnpem.br)) or A.L.B.A. (email: [andre.ambrosio@lnbio.cnpem.br](mailto:andre.ambrosio@lnbio.cnpem.br))

## ANEXO 8 – Parecer da Comissão Interna de Biossegurança (CIBio) do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM)

Use exclusivo da CIBio:

Número de projeto / processo: 2015-45

### Formulário de encaminhamento de projetos de pesquisa para análise pela CIBio - Comissão Interna de Biossegurança do CNPEM – Centro Nacional de Pesquisa em Energia e Materiais

**Título do projeto:** Regulação pós-transcricional da enzima glutaminase por HuR e sua relação com os altos níveis glutaminolíticos tumorais

**Pesquisador responsável:** Sandra Martha Gomes Dias

**Experimentador:** Douglas Adamoski Meira

**Classe de risco do OGM:**  Risco I  Risco II

Nível do treinamento do experimentador:  -Iniciação científica,  -mestrado,  -doutorado,  -doutorado direto,  -pós-doutorado,  -nível técnico,  -outro, especifique: \_\_\_\_\_

#### Resumo do projeto:

As diferenças genéticas e metabólicas das células tumorais permitem elevadas taxas de divisão, normalmente associadas a um grande consumo de glutamina para a produção de energia e formação de blocos biossintéticos celulares (lipídeos, proteínas e ácidos nucleicos). O primeiro passo do metabolismo deste aminoácido é a conversão em glutamato, catalisada pela enzima glutaminase. Glutaminase possui três isoformas oriundas de dois genes distintos: GLS que codifica para a enzima a kidney-type glutaminase (KGA) e a glutaminase C (GAC), produtos de splicing alternativo, e GLS2 que codifica para a liver-type glutaminase (LGA). Estudos preliminares do laboratório indicam que a KGA e a GAC, diferentes apenas na região C-terminal, são mais expressas em tecidos tumorais do que nos tecidos não-transformados adjacentes. Além disto, a isoforma GAC apresenta níveis de mRNA e proteicos mais elevados em relação à KGA nas linhagens tumorais de mama, quando comparado com cultura de células epiteliais mamárias normais. Estes mesmos estudos indicaram a proteína HuR como potencial reguladora da estabilidade do mRNA da GAC. HuR, altamente expressa em células de câncer, pode estar relacionada com os elevados níveis proteicos de GAC em tecidos tumorais. Avaliação preliminar de bancos de microarranjo indicou que há uma correlação heterogênea entre HuR, GAC e KGA quanto ao tipo de tecido avaliado. Serão expandidas as avaliações utilizando maior volume de dados de microarranjo e também e RNAseq para validar estes achados. Tal heterogeneidade está provavelmente envolvida com a sublocalização celular da proteína, dado que HuR pode atuar em várias instâncias regulatórias do mRNA, tais como splicing, estabilização e modulação da tradução. Silenciamento da proteína HuR na linhagem de tumor de próstata PC-3, através de sistema retroviral, levou ao aumento do nível proteico de GAC e diminuição de KGA, enquanto que superexpressão da construção HuR-V5-6xHis levou à redução de GAC e aumento de KGA. Ensaio de atividade de luciferase dirigido pelo 3'UTR destas isoformas e avaliação in silico de sequências regulatórias do intron 14 (decisivo para a definição entre as proteínas através de splicing alternativo) apontam para a possibilidade de que HuR possa regular a estabilidade/tradução do mRNA de GAC enquanto que também está potencialmente envolvido no splicing alternativo do pre-mRNA GLS e na escolha de KGA, ao menos na linhagem PC-3. A ação de HuR sobre o splicing alternativo de GLS será avaliada pelo (já construído) plasmídeo repórter bicromático contendo o minigene RG6 e o intron 14. A atividade enzimática da glutaminase, assim como os níveis extra e intra-celulares de nutrientes, especialmente glutamina, serão avaliados por RMN após realizar o knock down ou superexpressão de HuR. Todos os experimentos serão realizados no contexto das linhagens celulares do câncer mais relevante encontrado nas análises de bioinformática. Resultados preliminares de pull-down mostraram interação direta entre HuR recombinante produzida em sistema heterólogo e o 3'UTR de GAC. A confirmação de tais resultados e estudos de R-EMSA do elemento full length ou de seus subclones irão confirmar a interação, assim como definir as sub-regiões envolvidas. Por

fim, experimentos in vitro e em modelo animal serão utilizados para verificar a correlação entre obesidade / inflamação (Através NFkB / HuR) e a atividade glutaminolítica.

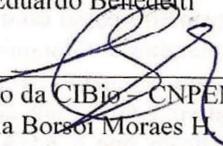
Os genes ELAVL1 (Proteína HuR, humano, *Homo sapiens*, com adição de mutações ou em sua forma selvagem), mKO2 (Coral, *Fungia concinna*, com adição de mutações), e as sequências regulatórias de GLS (humano, *Homo sapiens*) serão clonados em vetores retrovirais murinos para avaliar efeitos da superexpressão. O *knockdown* será realizado por meio de vetores lentivirais codificando sequências de RNAi. Os vetores retrovirais e lentivirais utilizados são defectivos, não se replicam, não são patogênicos e nem infecciosos. O experimentador recebeu treinamento de biossegurança e está sendo orientado pelo supervisor para o cumprimento de todos os requisitos necessários para manipulação, armazenamento e descarte de OGM risco II, conforme regulamentações da CTNBio e ANVISA

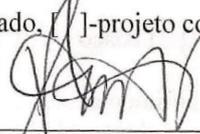
A CIBio analisou este projeto em reunião realizada no dia: 5/8/2015

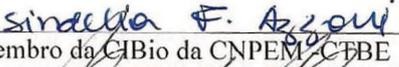
**Parecer final:** -projeto aprovado, [ ]-projeto recusado, [ ]-projeto com deficiências, comentários anexo.

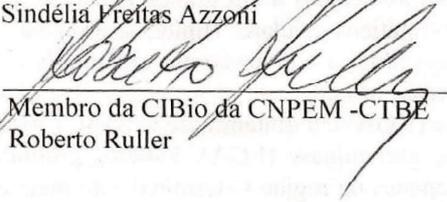
  
 Presidente da CIBio – CNPEM-LNBio  
 Marcio Chaim Bajgelman

  
 Membro da CIBio – CNPEM -LNBio  
 Celso Eduardo Benedetti

  
 Membro da CIBio – CNPEM -LNBio  
 Carolina Borsoi Moraes H. Freitas

  
 Membro da CIBio da CNPEM -CTBE  
 Fabio Márcio Squina

  
 Membro da CIBio da CNPEM -CTBE  
 Sindélia Freitas Azzoni

  
 Membro da CIBio da CNPEM -CTBE  
 Roberto Ruller

**ANEXO 9 – Declaração de Direitos Autorais****DECLARAÇÃO DE PERMISSÃO DE PUBLICAÇÃO DE ARTIGO**

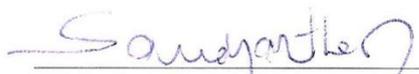
As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha tese de Doutorado, intitulada “Regulação pós-transcricional da enzima glutaminase por HuR e sua relação com os altos níveis glutaminolíticos tumorais” não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 05 de março de 2018.



Nome do Autor: Douglas Adamóski Meira

RG n.º 8.081.003 2 PR



Nome da Orientadora: Sandra Martha Gomes Dias

RG n.º M7854719