

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



RENATO MILANI

**“DESENVOLVIMENTO DE UM *PIPELINE* PARA ANÁLISE EM
LARGA ESCALA DE UM *CHIP* DE PROTEÍNAS QUINASES”**

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
RENATO MILANI
[assinatura]
e aprovada pela Comissão Julgadora.

Dissertação apresentada ao Instituto de
Biologia para obtenção do título de
Mestre em Biologia Funcional e
Molecular, na área de Bioquímica.

Orientador: Prof. Dr. Eduardo Galembeck
Co-Orientadora: Prof^a. Dr^a. Carmen Veríssima Ferreira

Campinas, 2010

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

M589d Milani, Renato
Desenvolvimento de um pipeline para análise em larga escala de um *chip* de proteínas quinases / Renato Milani. – Campinas, SP: [s.n.], 2010.

Orientadores: Eduardo Galembeck, Carmen Veríssima Ferreira.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Bioinformática. 2. Fosforilação. 3. Biotecnologia.
I. Milani, Renato. II. Ferreira, Carmen Veríssima. III. Universidade Estadual de Campinas. Instituto de Biologia.
IV. Título.

Título em inglês: Developing a pipeline for large scale analysis of a protein kinase chip.

Palavras-chave em inglês: Bioinformatics; Phosphorylation; Biotechnology.

Área de concentração: Bioquímica.

Titulação: Mestre em Biologia Funcional e Molecular.

Banca examinadora: Eduardo Galembeck, Hiroshi Aoyama, José Miguel Ortega.

Data da defesa: 26/02/2010.

Programa de Pós-Graduação: Biologia Funcional e Molecular.

Campinas, 26 de fevereiro de 2010

BANCA EXAMINADORA

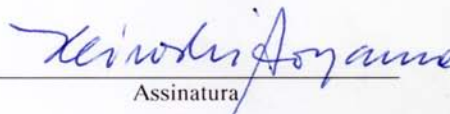
Prof. Dr. Eduardo Galembeck (Orientador)


Assinatura

Prof. Dr. José Miguel Ortega


Assinatura

Prof. Dr. Hiroshi Aoyama


Assinatura

Prof. Dr. José Mauro Granjeiro

Assinatura

Prof. Dr. Fábio Trindade Maranhão Costa

Assinatura

SUMÁRIO

AGRADECIMENTOS	xiii
RESUMO	xvii
ABSTRACT	xix
ABREVIATURAS	xxi
1 INTRODUÇÃO	01
1.1 Princípios gerais de sinalização celular	02
1.2 Modificações covalentes alteram a atividade enzimática	03
1.3 Quinases transferem um grupo fosforil do ATP para resíduos de aminoácidos	04
1.4 A fosforilação é um meio eficiente de regulação da atividade proteica	05
1.5 Quinases possuem diferentes especificidades determinadas pela sequência de aminoácidos do substrato	06
1.6 Tipos de quinases	07
1.7 O conceito de quinoma	08
1.8 Biologia de Sistemas: o passo após o reducionismo em sistemas complexos	11
1.9 O <i>chip</i>	13
1.10 Bioinformática	16
2 OBJETIVO	19
3 RESULTADOS	21

3.1 PepMatrix	21
3.1.1 As particularidades do <i>chip</i>	21
3.1.2 O algoritmo e a estatística	22
3.1.3 A normalização	25
3.1.4 Testes estatísticos	29
3.1.5 Núcleo de informações da matriz	29
3.1.6 O código	31
3.1.7 A interface com o usuário	35
3.2 Perfil Quinômico de Células TF1 Expostas Á Violaceína	37
3.3 Perfil Quinômico de Pré-Osteoblastos MC3T3-E1 em Adesão	51
4 DISCUSSÃO E CONCLUSÕES	63
5 BIBLIOGRAFIA	67
6 APÊNDICE	71
6.1 Artigos publicados	71
6.2 Artigos com parecer favorável em revisão	71
6.3 Artigos em processo de submissão	71
6.4 Apresentações em congressos	72

A vida é muito importante para ser levada a sério.

Oscar Wilde

Aos meus pais, Jair e Rosângela,
fontes de todo o apoio e compreensão
que sempre recebi.

Também dedico esta dissertação aos meus avós,
alguns dos quais não tive a felicidade de conhecer.

E a toda a minha família
pelo incentivo a continuar por esse caminho.

Dedico todo este trabalho àquela sem a qual ele não existiria.
Dani, você foi responsável por tantas mudanças fantásticas em minha vida que
seriam necessárias várias dissertações como essa para narrá-las todas.
Amo você!

AGRADECIMENTOS

A todos os meus amigos, que sempre fizeram parte dos momentos realmente importantes de minha vida. Caso não tenha mencionado especificamente alguém, não há dúvida de que essa pessoa se encaixa aqui.

Aos professores que contribuíram para construir o conhecimento que tenho hoje, com a paciência necessária a esse tipo de tarefa. Também agradeço àqueles que a perderam, já que me ensinaram não ser uma boa ideia bagunçar na sala de aula.

Especialmente ao meu orientador, Eduardo, por ter acreditado em mim desde o início, em 2005 e já me dar dicas nas aulas em 2004. Mas, principalmente, pela empolgação de realizar qualquer trabalho em conjunto numa relação em que amizade é a palavra mais adequada.

À professora Carmen, que surgiu em minha vida acadêmica por um acaso bastante feliz. Espero que eu sempre possa continuar aprendendo com ela a ter ótimas ideias.

Aos meus amigos de Americana, Everton, Fabio, Gustavo, Ricardo, Rodrigo e Adriano. O povão sempre se reunirá!

Aos meus amigos de Campinas, Mateus, Tiago, Victor e Melissa. Pelas discussões e pelo abrigo nas noites em que não voltei para Americana.

Ao meu primo Márcio, por toda a confiança depositada, além dos ensinamentos cujo valor é impossível mensurar. Mostre lixar chão e olho de tigre!

Ao Willian, que mostrou-se um grande e valiosíssimo amigo ao compartilhar de seu fascínio pela Ciência e me inspirar a realizar grandes feitos, mesmo que pareçam pequenos aos olhos dos outros.

Ao Professor Hiroshi, que com sua impressionante habilidade didática e experiência de laboratório foi capaz de fazer com que eu vivesse a Bioquímica ainda mais intensamente.

Ao professor do Laboratório de Proteômica Claudio Werneck, pelas discussões, oportunidades e colocações no exame de qualificação.

Ao Professor Miguel, por causar profundas alterações em minha vida computacional e por me ensinar que a Bioinformática é a Informática que existe dentro de nós.

À Nani, Quico e Millie, pelos ensinamentos sobre a vida.

Aos pesquisadores holandeses Sander e Maikel, por permitirem as análises com o PepChip e disponibilizarem seu conhecimento para que fossem realizadas.

Ao pessoal do Laboratório de Tecnologia Educacional, Thanuci e Cia., por tornar os dias mais alegres e menos gelados de coração. Agradeço ao Danilo por ter me ajudado nos momentos finais da impressão. Também ao Anderson e Itaraju, pelos ensinamentos que me permitiram desenvolver meu trabalho.

Ao pessoal do Laboratório de Transdução de Sinal e Bioensaios, que me deu as boas-vindas e sempre esteve à disposição para qualquer dificuldade que eu tive. Também por estar presente em minha primeira SBBq.

Agradeço também ao Antonio, Karla, Luciana, Willian e todos mais que acreditaram que uma análise mais elaborada do PepChip seria possível e cederam seus dados experimentais para validação pela nova técnica. Sem o apoio de vocês este trabalho jamais teria acontecido.

À secretária do programa de pós-graduação, Andréia, por todo o trabalho que dei.

Ao CNPq, pela oportunidade financiada de desenvolver o projeto.

E, finalmente, agradeço a todos os que se dispuseram a gerar artes e tecnologias que foram capazes de me fascinar e entreter durante todos os anos de minha vida e que, sem dúvida, continuarão trabalhando e conseguindo isso ao longo de toda ela.

AGRADECIMENTOS ESPECIAIS



RESUMO

A atividade de proteínas quinases é responsável pela regulação de muitos processos biológicos através de cascatas de sinalização que levam a diferentes efeitos celulares. No entanto, a análise da reação de fosforilação reversível catalisada por estas enzimas é prejudicada pela complexidade inerente às interações entre proteínas neste sistema de sinalização. Consequentemente, o foco em uma única proteína quinase pode não ser suficiente para revelar completamente os mecanismos por trás dos fenótipos observados. Nesse sentido, em conjunto com outras técnicas de análise em larga-escala como *chips* de expressão de mRNA, arranjos de peptídios contendo substratos de proteínas quinases vem sendo cada vez mais utilizados por pesquisadores. No entanto, a falta de uniformidade na análise estatística desses *chips* tem sido um grande empecilho à obtenção de dados relevantes com o uso dessa técnica. Por conta disso, o objetivo desse trabalho foi desenvolver uma metodologia, chamada de PepMatrix, capaz de aplicar estatística básica de forma automatizada visando a seleção de replicações com baixa variabilidade e a obtenção da anotação das proteínas envolvidas nos eventos de fosforilação ocorridos no *chip*. Esse novo método foi aplicado em vários conjuntos de dados de diferentes experimentos biológicos e seus resultados revelaram atividades quinásicas significativamente alteradas, muitas das quais tiveram confirmação por *Western blot*. Além disso, os resultados ressaltaram a importância da análise sistêmica dos eventos de sinalização celular em conjunto com uma análise crítica das replicações. O alto grau de uniformidade analítica obtido por esse método faz com que ele seja uma poderosa e confiável ferramenta na análise quinômica em larga-escala.

ABSTRACT

The activity of protein kinases governs many biological processes through signaling cascades that lead to distinct outputs, from homeostasis to disease. However, analysis of the reversible phosphorylation perpetrated by these enzymes is hindered by the inherent complexity of interactions in this signaling system. Consequently, focusing on a single kinase may not be enough to completely unfold the mechanisms behind observed phenomena. In this sense, together with other omics approaches like mRNA expression analysis, peptide arrays have shown increasing popularity, particularly ones containing kinase substrate sequences. The lack of uniformity in statistical analysis of these chips, though, has been a major issue for the field. In this paper, we propose PepMatrix, a fast and accurate method for selecting so called "reliable replicate spots" and automatically retrieving differential activity and annotation information about phosphorylation events identified in a peptide array of kinase substrates. Here, we present several cases where this new methodology was applied to biological datasets. We successfully identified putative up and down-regulated kinases, many of which were confirmed to have altered activity by Western blot. Moreover, the results emphasized the need for a true systems biology approach to the cellular signaling events alongside a critical replicate selection method. The high degree of analysis uniformity we achieved with this method provides a powerful and reliable addition for high-throughput kinome analysis.

ABREVIATURAS

A	Média
ADP	Adenosina difosfato
AKT	Proteína quinase B
AMPK	Proteína quinase ativada por 5'-AMP
aPK	Proteína quinase atípica
Arg	Arginina
ATM	Proteína serina quinase Ataxia Telangiectasia Mutated
ATP	Adenosina trifosfato
c-Mer	Proteína tirosina quinase Mer
CaMK	Proteína quinase dependente de Ca^{2+} /calmodulina
CaMKK	Proteína quinase quinase dependente de Ca^{2+} /calmodulina
CDC2	Homólogo da proteína de controle da divisão celular 2
CDK5	Quinase dependente de ciclina 5
cDNA	DNA complementar
CK	Caseína quinase
DAPK1	Quinase associada à morte 1
DNA	Ácido desoxirribonucleico
EDTA	Ácido tetra-acético etileno diamina
EGFR	Receptor do fator de crescimento epidérmico
EGTA	Ácido tetra-acético etileno glicol
ePK	Proteína quinase eucariótica

ERK2	Quinase regulada por sinal extracelular 2
FAK	Quinase de adesão focal
GCN2	Quinase 4 do fator eucariótico de iniciação da tradução 2-alfa
GSK3β	Glicogênio-sintase quinase beta
HCR	Quinase 1 do fator eucariótico de iniciação da tradução 2-alfa
InsR	Receptor de insulina
KEGG	Enciclopédia de Genes e Genomas de Kyoto
M	Mediana
MAP3K11	Proteína quinase quinase quinase ativada por mitógeno 11
MAPK	Proteína quinase ativada por mitógeno
MC3T3-E1	Linhagem celular osteoblástica de camundongo
mRNA	Ácido ribonucleico mensageiro
NCBI	Centro Nacional para Informação Biotecnológica (EUA)
NIH	Institutos Nacionais de Saúde (EUA)
PBS	Tampão fosfato-salino
pH	Potencial hidrogeniônico
PIM1	Proto-oncogene proteína quinase serina/treonina PIM1
PAK1	Proteína quinase serina/treonina ativada por p21
PDGFRβ	Receptor do fator de crescimento derivado de plaquetas do tipo beta
PKA	Proteína quinase dependente de cAMP
PKC	Proteína quinase C
PKG	Proteína quinase dependente de cGMP

PKR	Proteína quinase ativada por RNA dupla-fita induzida por interferon
PLK	Quinase do tipo polo
pp90rsk	Proteína quinase S6 ribossomal
Ser	Serina
RTK	Receptor tirosina quinase
SD	Desvio-padrão
SGK	Quinase serina/treonina regulada por soro/glicocorticoide
Src	Proteína tirosina quinase Src (sarcoma)
TF1	Linhagem celular eritroleucêmica
Thr	Treonina
VBScript	Microsoft Visual Basic Scripting Edition

1 INTRODUÇÃO

A Biologia Molecular tem apresentado, nos últimos anos, descobertas apoiadas em metodologias que anteriormente não estavam disponíveis ou que possuíam um custo demasiadamente elevado. A extensão dos avanços tecnológicos a um número cada vez maior de cientistas fez com que a quantidade de dados biológicos gerados aumentasse de forma exponencial, já que técnicas *high-throughput* como *microarrays* de DNA e de proteínas, *phage display* e duplo híbrido vem sendo associadas a ferramentas computacionais capazes de armazenar, manipular, combinar e recuperar tais dados. Nesse contexto, o presente trabalho tem por objetivo o desenvolvimento de um *pipeline* para análise em larga escala de uma técnica voltada para a obtenção de informações sobre eventos de fosforilação proteica. Esse mecanismo controla a maior parte da transdução de sinal em células eucarióticas através de proteína quinases, enzimas capazes de transferir o grupo fosfato terminal do ATP para resíduos de aminoácidos em uma cadeia polipeptídica. A técnica consiste de um *microarray* contendo peptídios que são substratos conhecidos de proteína quinases dispostos numa lâmina (*chip*). Quando submetidos ao lisado de uma célula, os substratos são fosforilados por suas quinases e posteriormente identificados através de um marcador radioativo. Como nas outras técnicas de *microarray*, existe um grande número de informações que precisam ser pré-analisadas para depois fornecer ao pesquisador um conjunto de dados com o qual trabalhar. No entanto, pelo fato de ser uma técnica nova, o *chip* não possui uma metodologia de análise estabelecida, levando a dificuldades para interpretar os dados brutos gerados e transformá-los em informações de significância biológica. Este trabalho descreve a construção de um sistema capaz de automatizar a análise dos dados do

chip e fornecer aos cientistas que o utilizam informações quantitativas baseadas em parâmetros estatísticos sobre eventos de fosforilação passíveis de ocorrência no ambiente celular.

1.1 PRINCÍPIOS GERAIS DE SINALIZAÇÃO CELULAR

A capacidade celular de receber e reagir a estímulos externos é fundamental para a vida. Células bacterianas recebem constantemente sinais de proteínas de membrana relacionados ao meio que as cerca, trazendo informações como as condições de pH, salinidade, presença de nutrientes, oxigênio e luz, além de substâncias tóxicas, predadores ou competidores. Esses sinais desencadeiam respostas apropriadas, como o movimento em direção ao alimento, a fuga de substâncias tóxicas ou a formação de esporos num meio sem nutrientes. Em organismos multicelulares, células com diferentes funções trocam uma grande quantidade de sinais. Células vegetais respondem a hormônios de crescimento e a variações luminosas. Células animais trocam informações sobre a concentração de íons e de glicose em fluidos extracelulares, sobre atividades metabólicas ocorrendo em diferentes tecidos e sobre o correto posicionamento de cada uma durante o desenvolvimento de um embrião. Em todos esses casos, o sinal representa informação, detectada por receptores e convertida numa resposta celular, sempre envolvendo um sinal químico. A conversão de informação em um sinal químico é chamada de transdução de sinal, uma propriedade universal das células (Figura 1).

O número de sinais biológicos a que as células respondem é grande, bem como a quantidade de respostas celulares a cada um deles. Contudo, existem mecanismos conservados para a detecção de sinais extracelulares e para a sua transdução em modificações intracelulares.

Frequentemente, esses mecanismos resultam na fosforilação de proteínas que fazem parte de uma cascata de sinalização. (NELSON & COX, 2004, pp. 421-422).

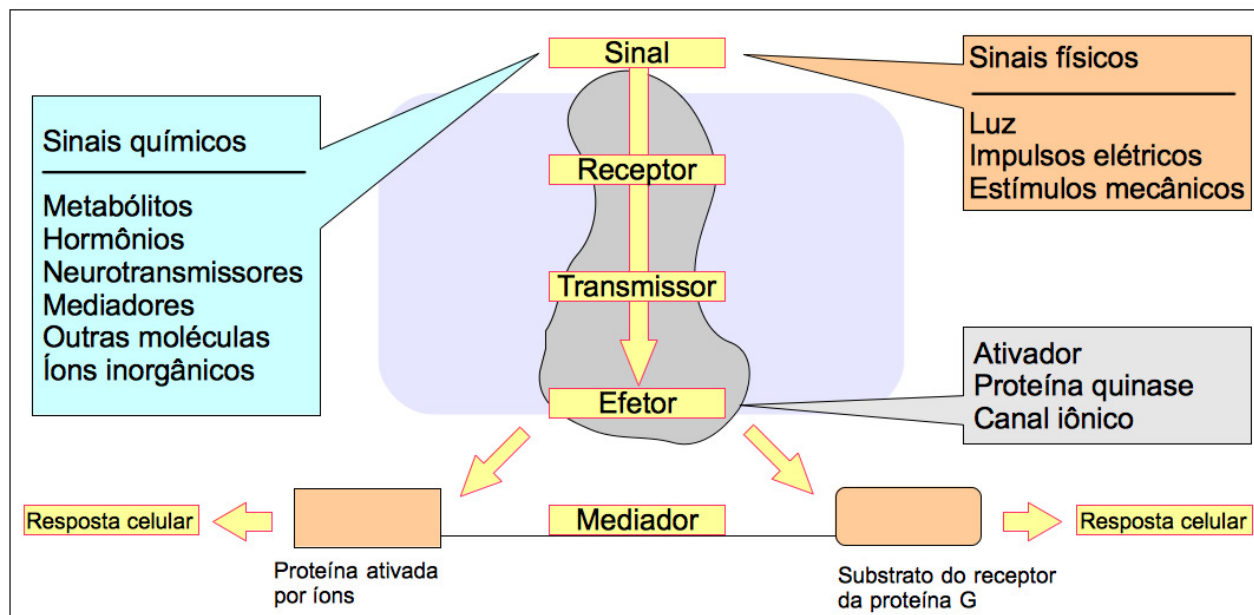


Figura 1. Mecanismo básico da atuação de um receptor de sinal celular. Ao interagir com um sinal específico que pode ser físico (luz, impulsos elétricos ou estímulos mecânicos) ou químico (metabólitos, hormônios, neurotransmissores etc.), o receptor é capaz de transmitir a informação recebida até um domínio efetor, que por sua vez pode ativar proteínas mediadoras (levando a uma cascata de sinalização) ou então passar a funcionar como um canal iônico (frequente em receptores de neurotransmissores). Adaptado de Koolman & Röhm, 2004.

1.2 MODIFICAÇÕES COVALENTES ALTERAM A ATIVIDADE ENZIMÁTICA

A atividade de uma enzima pode ser drasticamente alterada pela ligação covalente de certos grupos à cadeia polipeptídica, que causa modificações importantes em sua conformação. A alteração de atividade pode ser decorrente de mudanças na afinidade pelo substrato ou na sensibilidade a efetadores alostéricos. A modificação covalente constitui uma reação química e, portanto, catalisada por enzimas. A mudança de atividade não é definitiva, pois o sistema celular também inclui enzimas capazes de catalisar a retirada do grupo adicionado, retornando a enzima

modificada à sua conformação original. Várias são as modificações possíveis (metilação, adenilação, acetilação etc.), porém a mais prevalente consiste na fosforilação da cadeia polipeptídica. A fosforilação é catalisada por proteína quinases, enzimas que constituem uma das maiores famílias proteicas conhecidas, contando com mais de 500 integrantes em *Homo sapiens* (MARZZOCO & TORRES, 2007, p. 263). Também conhecidas como proteína cinases, possuem esse nome porque são capazes de alterar as propriedades cinéticas de seus substratos através de uma reação de fosforilação. Tal reação é capaz de alterar as propriedades conformacionais da proteína-alvo e, assim, alterar parâmetros cinéticos como atividade catalítica e afinidade com seu próprio substrato.

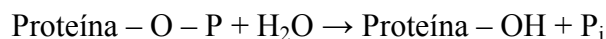
1.3 QUINASES TRANSFEREM UM GRUPO FOSFORIL DO ATP PARA RESÍDUOS DE AMINOÁCIDOS

As atividades de muitas enzimas, canais de membrana e outras proteínas são controladas por fosforilação catalisada por quinases, uma reação que envolve a transferência do grupo fosforil terminal (γ) do ATP para resíduos específicos de serina e treonina (muito embora outros resíduos também, raramente, possam sofrer fosforilação) por uma classe de quinases ou para resíduos de tirosina por outra classe, formando uma ligação éster fosfórico:



Os substratos de uma reação de fosforilação localizam-se no interior celular, onde o ATP doador de grupos fosforil é abundante. Proteínas exclusivamente extracelulares não são reguladas por fosforilação.

Um outro grupo de enzimas, as proteína fosfatases, é capaz de reverter os efeitos da fosforilação através da remoção por hidrólise do grupo fosforil ligado covalentemente a cadeias proteicas:



É importante notar que a reação catalisada por fosfatases não é a simples reversão da reação de fosforilação. Ambas são irreversíveis em condições fisiológicas. Além disso, ambas ocorrem a velocidades extremamente baixas na ausência de enzimas. Dessa forma, a fosforilação de uma proteína só ocorrerá pela intervenção direta de uma quinase, às custas de ATP, enquanto a desfosforilação depende da atuação de uma fosfatase. A predominância da forma fosforilada ou desfosforilada depende essencialmente das atividades relativas de quinases e fosfatases (MARZZOCO & TORRES, 2007, p. 263; STRYER *et al.*, 2002, pp. 275-278).

1.4 A FOSFORILAÇÃO É UM MEIO EFICIENTE DE REGULAÇÃO DA ATIVIDADE PROTEICA

Estruturalmente, um grupo fosforil adiciona duas cargas negativas à cadeia proteica. Interações eletrostáticas previamente estabelecidas podem ser rompidas e novas interações podem ser formadas, o que pode levar a alterações drásticas na capacidade de interação da proteína com seu substrato e em sua atividade catalítica. Além disso, um fosfato pode formar três ou mais ligações de hidrogênio, o que permite novas e específicas interações com outras moléculas.

Termodinamicamente, a energia livre da reação de fosforilação equivale a $-50\text{kJ}\cdot\text{mol}^{-1}$, sendo que a energia provém do ATP. Aproximadamente metade dessa energia é conservada na

ligação covalente. Levando em conta que uma alteração de energia livre de $5,69\text{kJ}\cdot\text{mol}^{-1}$ corresponde a um fator de 10 numa constante de equilíbrio, pode-se concluir que a reação de fosforilação é capaz de promover uma grande alteração do equilíbrio conformacional entre diferentes estados funcionais, correspondente a um fator de aproximadamente 10^4 .

As reações de fosforilação e de desfosforilação podem também ser reguladas cineticamente para ocorrerem de acordo com as necessidades de um processo fisiológico, ocorrendo em uma fração de segundo ou ao longo de várias horas. Além disso, esse tipo de reação é capaz de gerar efeitos em larga escala, uma vez que uma única quinase ativa pode fosforilar centenas de substratos num curto intervalo e, por sua vez, alguns desses substratos podem ser outras enzimas, levando a uma nova cascata de reações e efeitos celulares. Por fim, o uso do ATP como doador do grupo fosforil faz com que essas reações estejam intimamente relacionadas à regulação do metabolismo (STRYER *et al.*, 2002, pp. 275-278).

1.5 QUINASES POSSUEM DIFERENTES ESPECIFICIDADES

DETERMINADAS PELA SEQUÊNCIA DE AMINOÁCIDOS DO SUBSTRATO

Ao passo que existem quinases específicas, capazes de fosforilar apenas uma única proteína ou um grupo de proteínas similares, também existem quinases pouco específicas, capazes de fosforilar substratos dos mais variados, tendo um grande alcance e regulando uma miríade de processos. Independentemente do tipo, as quinases são capazes de reconhecer seu substrato por conta de uma *sequência consenso*, que corresponde a uma sequência determinada de aminoácidos presente na cadeia da proteína-substrato. Como exemplo, a sequência consenso para a proteína quinase A (PKA) é Arg-Arg-X-Ser-Z ou Arg-Arg-X-Thr-Z, sendo que X é um

resíduo qualquer contendo um radical pequeno e Z é um resíduo hidrofóbico. O sítio de fosforilação é o resíduo de serina ou de treonina. Essa sequência consenso, no entanto, não é estritamente necessária: um resíduo de lisina pode substituir um dos resíduos de arginina, levando a uma redução de afinidade com a PKA. Peptídeos curtos sintéticos (como os presentes no *chip*), contendo uma sequência consenso são quase sempre fosforilados por serina-treonina quinases. Dessa forma, o fator mais importante na determinação da especificidade é o conjunto de resíduos adjacentes ao sítio de fosforilação. No entanto, resíduos distantes do sítio de fosforilação também podem contribuir para a determinação da especificidade, na medida em que são capazes de alterar a conformação da proteína e alterar a acessibilidade do sítio de fosforilação às proteína quinases (STRYER *et al.*, 2002, pp. 275-278).

1.6 TIPOS DE QUINASES

O sequenciamento do genoma humano revelou que aproximadamente 20% dos genes humanos codificam proteínas envolvidas na transdução de sinal (MANNING *et al.*, 2002). Dentre elas, mais de 500 são proteína quinases, que podem ser subdivididas em duas classes, dependendo de sua especificidade catalítica em relação a tirosina ou a serina/treonina. Tirosina quinases, que incluem RTK (*receptor tyrosine kinases*) ativam numerosas vias de sinalização, levando a proliferação celular, diferenciação, migração e alterações metabólicas. Além disso, a atividade elevada de tirosina quinases é uma característica marcante de uma considerável fração de tumores e outras doenças proliferativas. Serina/treonina quinases também tem um papel fundamental na homeostase e na sinalização celular, já que são responsáveis pela fosforilação de fatores de transcrição, reguladores do ciclo celular e de um grande número de efetores

citoplasmáticos e nucleares. Também estão associadas a alguns tipos de câncer, principalmente através da perda de regulação de MAPK (*mitogen-activated protein kinases*), Aurora e PLK (*Polo-like kinase*), famílias que estão relacionadas ao crescimento de tumores, metástases e a um prognóstico clínico ruim (HUNTER & COOPER, 1985; EDELMAN *et al.*, 1987; CAPRA *et al.*, 2006).

1.7 O CONCEITO DE QUINOMA

O quinoma corresponde ao complemento quinásico do proteoma, ou seja, o conjunto de proteína quinases codificadas pelo genoma de um organismo (GOMASE & TAGORE, 2008). Manning *et al.* identificaram 518 genes no genoma humano que codificam proteínas contendo domínios catalíticos quinásicos. Esses genes foram classificados em duas famílias principais: ePK e aPK. A primeira agrupa todas as enzimas que possuem o domínio catalítico das quinases de eucariotos, correspondendo a 478 genes. A segunda família contém todas as quinases que não possuem esse domínio mas que apresentam atividade quinásica, sendo portanto chamadas de atípicas e totalizando 40 genes. Como cada uma das 518 proteína quinases catalisa a fosforilação de um conjunto distinto de substratos, é possível construir redes complexas interligando a atividade de uma quinase a toda uma cascata de reações e a uma resposta celular. Para isso, existem duas questões principais: quais substratos e sítios cada quinase é capaz de fosforilar, e qual quinase é responsável pela fosforilação de um sítio sabidamente fosforilado? Classicamente, as reações de fosforilação são estudadas por métodos bioquímicos bastante laboriosos, como mapeamento de peptídios, análise de fosfoaminoácidos e purificação de quinases. Apesar de essas técnicas terem permitido a obtenção de informações previamente indisponíveis e ainda hoje apresentarem utilidade, o desafio que se apresenta no mapeamento das redes de sinalização

protagonizadas pelas enzimas integrantes do quinoma possui uma escala que demanda o uso combinado de diferentes técnicas, conforme apresentado na figura 2 (JOHNSON & HUNTER, 2005). Uma delas, a de *screening* via arranjo de proteínas, é capaz de responder não só qual é o substrato que uma determinada quinase é capaz de fosforilar, mas quais são as quinases expressas e ativas numa determinada condição celular, o que permite associar as vias de sinalização ativas numa célula com os processos biológicos apresentados por ela. Essa abordagem busca integrar vários eventos celulares envolvendo quinases considerando que o conjunto deles é responsável pelo fenótipo celular, culminando numa abordagem sistêmica da sinalização celular.

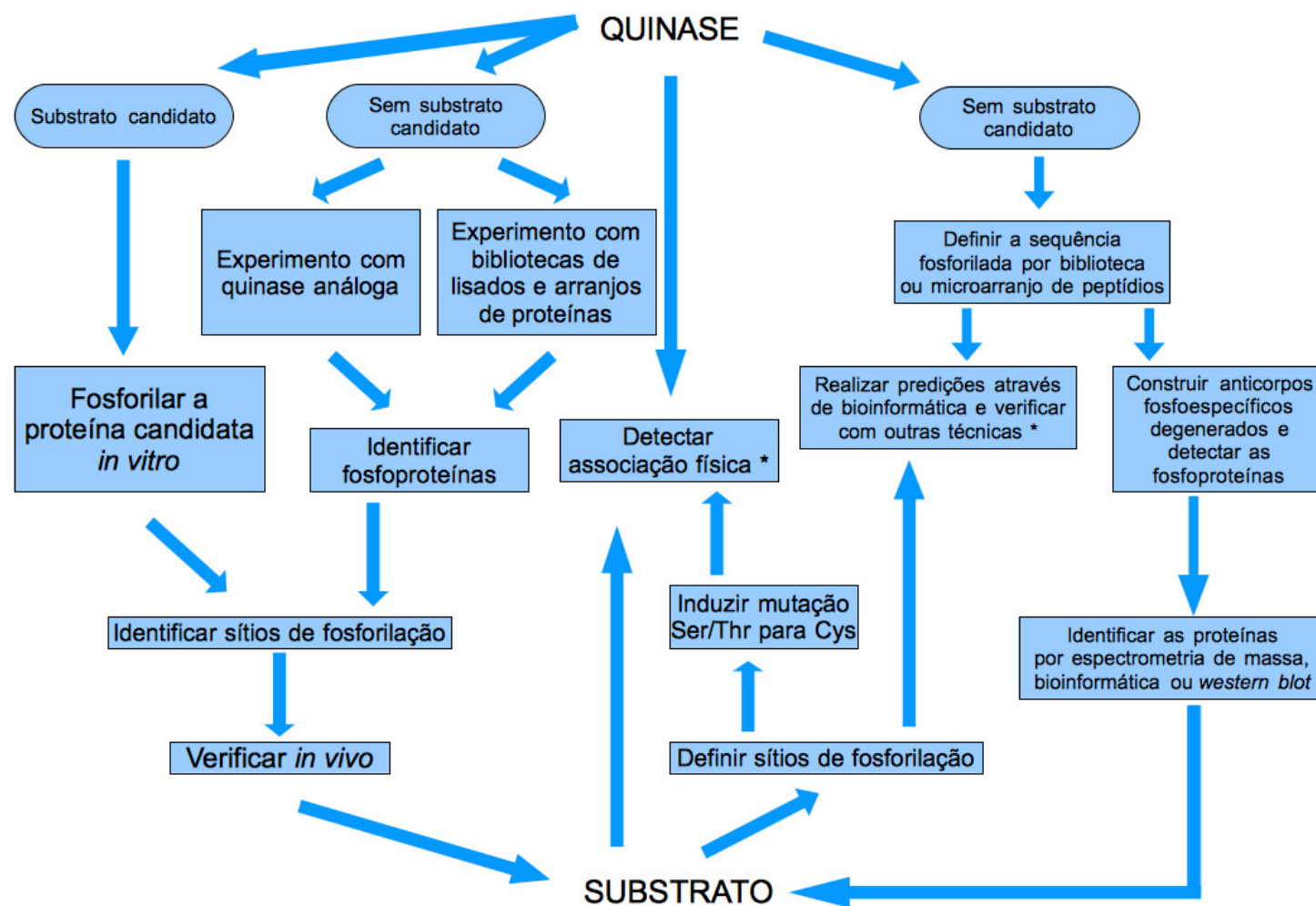


Figura 2. Fluxograma de estratégias disponíveis para a identificação de um substrato ou quinase. Técnicas marcadas com * oferecem resultados que precisam ser posteriormente validados usando metodologias experimentais. Adaptado de Johnson & Hunter, 2005.

1.8 BIOLOGIA SISTÊMICA: O PASSO APÓS O REDUCIONISMO EM SISTEMAS COMPLEXOS

A definição de Biologia Sistêmica (também conhecida como Biologia de Sistemas) envolve o estudo de redes gênicas, proteicas, metabólicas e assim por diante, de modo a compreender os princípios de funcionamento de sistemas biológicos complexos. Esses estudos podem focar em redes em larga escala (formadas por genomas, proteomas etc.) ou em pequenos subcircuitos da rede. A Biologia de Sistemas começa com um fenômeno biológico complexo visando um arcabouço explicativo simples e lógico evidenciando o mecanismo por trás da ocorrência dos eventos que compõem o fenômeno estudado, de modo a gerar um conjunto de dados capaz de descrever processos biológicos e identificar interações entre seus componentes (Figura 3).

A Biologia Sistêmica não é a antítese do reducionismo, que por sua vez é uma abordagem para o entendimento da complexidade através do estudo específico de suas partes. Muitas vezes os componentes individuais de um evento já foram descritos através de abordagens reducionistas e é sobre essa fundação que a abordagem sistêmica se apoia. As propriedades de sistemas complexos são diferentes das propriedades de proteínas individuais. Essas propriedades, chamadas emergentes, são o objeto de estudo da Biologia Sistêmica (FERRELL JR., 2009). No caso do *chip* de substratos de proteína quinases, o objetivo é exatamente partir dos eventos individuais de fosforilação e analisá-los em conjunto, visando compreender como a ocorrência deles é capaz de regular um processo biológico complexo.

No contexto dos microarranjos de DNA ou de proteínas, a importância dessas metodologias é justamente a de permitir uma análise ampla das vias metabólicas e de sinalização

ao verificar a atividade de milhares de enzimas, no caso dos *chips* de proteínas, ou os níveis de expressão de milhares de genes, no caso dos *chips* de DNA, ao mesmo tempo. No entanto, biologicamente, é mais importante o foco na atividade enzimática do que na concentração de metabólitos ou na expressão gênica (van BAAL *et al.*, 2006; DIKS *et al.*, 2004; IRISH *et al.*, 2004; ANAMIKA & SRINIVASAN, 2007), justamente por conta de que a presença de um metabólito ou do produto de um gene na célula não garante que a mesma esteja ativa, uma vez que isso depende de modificações pós-traducionais.

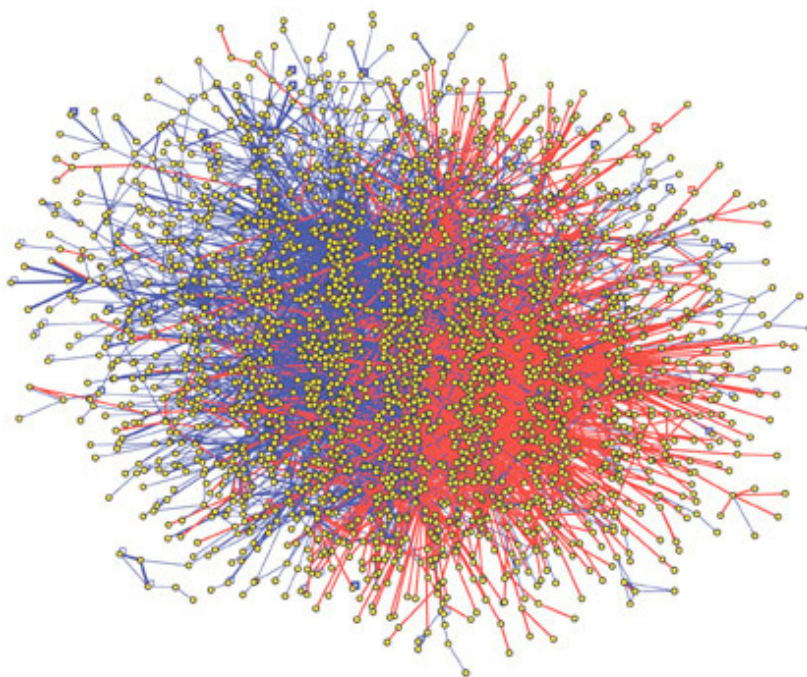


Figura 3. Exemplo extremo de rede de interações entre proteínas humanas, obtida a partir de um estudo na área de Biologia de Sistemas. Cada ponto corresponde a uma proteína, enquanto as retas correspondem a interações entre proteínas. Retas vermelhas representam interações obtidas a partir de um estudo de duplo-híbrido ao passo que retas azuis correspondem a interações obtidas na literatura a partir de um conjunto de 7200 genes que codificam proteínas. Adaptado de Rual, *et al.*, 2005.

1.9 O *CHIP*

Este trabalho visa a construção de um fluxo metodológico que permita a análise dos dados produzidos por uma técnica relativamente nova na área da Biologia Molecular, uma placa de microarranjos contendo substratos de proteínas quinases, conhecida como PepChip® e produzida pela empresa Pepscan Presto BV. Este *chip* foi desenvolvido pelo grupo do Dr. Maikel Peppelenbosch, na Universidade de Groningen, Holanda, e sua utilização no Laboratório de Transdução de Sinal e Bioensaios – IB – Unicamp é constante, como parte da intensa colaboração existente entre os dois grupos. Devido ao seu recente e contínuo desenvolvimento, esta técnica é ainda incipiente em relação a processos de análise que produzam informações relevantes aos pesquisadores da área biológica.

O *chip* é constituído por uma lâmina de vidro coberta com uma superfície hidrofílica de 50nm a 100nm de espessura contendo peptídios aderidos correspondentes a sequências consenso reconhecidas por proteína quinases. Essa superfície, segundo o fabricante, possui excelentes propriedades para o reconhecimento de peptídios por enzimas e com uma baixa taxa de ligação não-específica, resultando em altas razões sinal/ruído. Os peptídios são covalentemente ligados à superfície de forma estável e em alta densidade. Na superfície, são alocados em *spots* e agrupados blocos de 32x32 em triplicata, totalizando 1024 peptídios no caso do PepChip® Kinomics™ ou de 28x42 em duplicata, totalizando 1176 peptídios, no caso do PepChip® Kinase™. Nos dois casos há espaço para canaletas e bordas de 3mm. A distância de um *spot* para o outro (centro a centro) é de 560µm (Figura 4). O PepChip® Kinomics™ contém somente sequências oriundas de *Homo sapiens*, com foco especial em substratos de relevância fisiológica, como os relacionados a estresse, crescimento e diferenciação celular. O PepChip® Kinase™, por sua vez, contém

sequências de várias espécies, permitindo investigações evolutivas a respeito da especificidade de quinases e caracterização da atividade de novas quinases.

A preparação do *chip* consiste em obter uma cultura das células desejadas e promover sua lise com um tampão de lise comercial adicionado de Tris-HCl 20mM (pH 7,5), NaCl 150mM, EDTA 1mM, EGTA 1 mM, 1% Triton X-100, pirofosfato de sódio 2,5mM, h-glicerofosfato 1mM, vanadato de sódio 1mM, fluoreto de sódio 1mM, leupeptina 1µg/mL e aprotinina 1µg/mL. Os lisados são então centrifugados a 20000g por 10 minutos a 4°C e o precipitado é descartado. Na sequência, 50µL de lisado é adicionado a 12µL do mix de ativação contendo glicerol 50%, MgCl₂ 60mM, Brij-35 0,05% (v/v), albumina de soro bovino 0,25mg/mL e [γ -³³P] ATP 250µmol/L a 2000 Ci/mL. O *chip* é então incubado com os lisados por 90 minutos numa estufa a 37°C e depois lavado duas vezes em PBS contendo Triton X-100 0,1%, duas vezes em NaCl 2M contendo Tween 20 0,1% e duas vezes em água destilada. Depois, é seco ao ambiente e exposto à uma tela sensível a radiação por 72 horas. O objetivo até aqui é que o [γ -³³P] ATP tenha fornecido o grupo fosforil radioativo como substrato para a reação de fosforilação promovida pelas quinases presentes no lisado. Geralmente é realizado um controle para garantir a ligação específica do [γ -³³P] ATP, substituindo-o por [α -³³P] ATP: como na reação de fosforilação o γ -fosfato do ATP é o fosfato incorporado na proteína, espera-se que com o uso de [α -³³P] ATP não seja detectada radioatividade nas amostras.

O *chip*, então, após os procedimentos experimentais, tem suas imagens processadas pelo aplicativo gratuito EisenLab ScanAlyze, versão 2.50. Após alinhamento do *grid* específico correspondente a posição dos *spots* e otimização da localização dos *spots*, o programa cria arquivos DAT contendo informações sobre a intensidade de radiação de cada *spot*, gerada pela fosforilação dos substratos do *chip* a partir de γ -³³ATP.

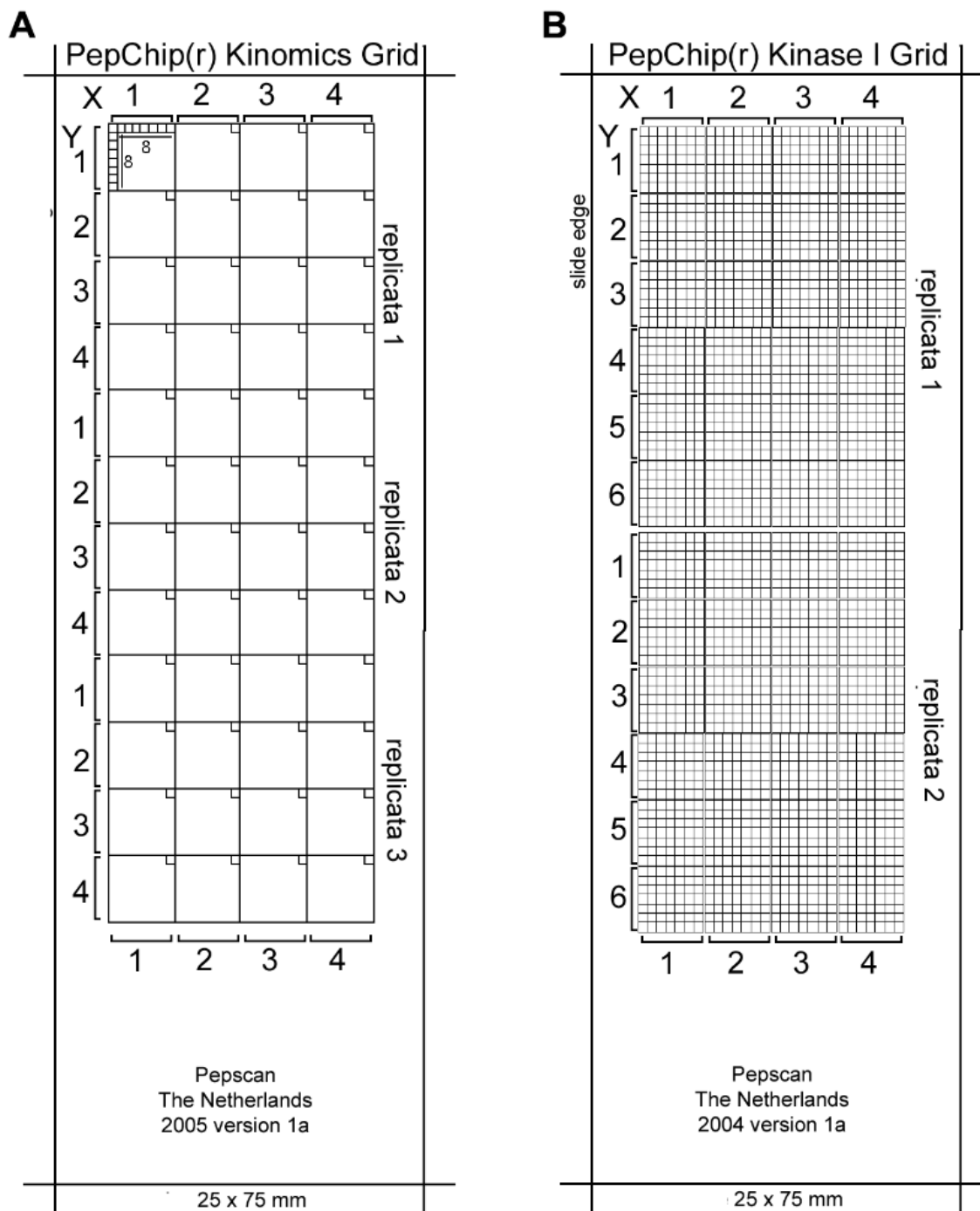


Figura 4. Estrutura de *spots* no PepChip® Kinomics™ (a) e Kinase™ (b). O primeiro apresenta 1024 *spots* (32x32) em triplicata, contendo apenas sequências humanas, enquanto o segundo contém 1176 *spots* (28x42) em duplicata, contendo sequências de diversas espécies.

1.10 BIOINFORMÁTICA

A Bioinformática pode ser definida como o estudo da informação existente nos seres vivos. Longe de ser apenas uma ferramenta operacional, também pode ser definida como a aplicação da Tecnologia da Informação no manejo de dados biológicos (GIBAS & JAMBECK, 2001). Isso envolve o desenvolvimento e o uso de ferramentas computacionais visando a expansão e facilitação do uso de dados biológicos através da capacidade de adquirir, armazenar, organizar, analisar e visualizar tais dados, de acordo com o NIH – NCBI (*National Institutes of Health – National Center for Biotechnology Information*). Longe de ser um fim em si mesma, a Bioinformática é um dentre os vários componentes das Ciências Biológicas.

Base da Tecnologia da Informação, os computadores são dispositivos poderosos para a compreensão de qualquer sistema que pode ser descrito matematicamente. Processos biológicos geralmente envolvem proteínas, que por sua vez são sintetizadas a partir do fluxo de informação codificada no DNA. O estudo desses processos ocorre, na maioria das vezes, uma proteína de cada vez: cientistas determinam a sequência ou a estrutura ou analisam os perfis de expressão de um produto gênico de cada vez, levando a uma compreensão detalhada dessa entidade. No entanto, o mecanismo pelo qual uma via bioquímica ou uma proteína interage com outros componentes biológicos é um problema de escala muito maior e mais difícil de ser abordado. São necessárias técnicas para realizar experimentos em larga escala e ferramentas para armazenar e analisar os dados gerados por esses experimentos. Usando os microarranjos de DNA ou de peptídios como exemplo: os *chips* permitem que sejam realizados, efetivamente, milhares de experimentos individuais de expressão gênica ou de atividade enzimática simultaneamente numa pequena lâmina de vidro. Os resultados desses experimentos precisam ser armazenados num

banco de dados relacional, que proporcione rápido acesso posterior e eficiente manipulação dos dados, tal como buscas cruzadas e exportação para outros formatos. No entanto, devido à natureza complexa e interligada da informação biológica, cada *spot* num microarranjo de DNA está conectado não só a informações sobre sua intensidade, mas a várias camadas de informação relacionadas a localização no genoma, sequência de DNA, estrutura, função etc. Os genes ou proteínas de interesse estão presentes em bases de dados *online* que contém informações geradas por inúmeros outros experimentos de pesquisadores espalhados pelo mundo. Esses genes ou proteínas possuem sequências biológicas também disponíveis na *World Wide Web*. As sequências podem apontar para genes ou proteínas similares no mesmo ou em outros organismos. Além disso, cada gene no genoma não é uma entidade independente. Múltiplos genes interagem para formar vias bioquímicas, que por sua vez alimentam outras vias e sofrem a influência do ambiente externo, como patógenos e outros estímulos. Existem bancos de dados catalogando interações entre proteínas, como o DIP (*Database of Interacting Proteins*) (XENARIOS, 2000) e o IntAct (ARANDA, 2009), o que permite a um pesquisador desenhar um mapa de interações que descreva um processo biológico observado a partir de algumas proteínas identificadas como relevantes em seus próprios experimentos. Ainda no campo das interações entre proteínas, existe um meta banco chamado STRING (JENSEN, 2009), que permite a busca por interações entre proteínas catalogadas em diferentes bancos de dados (incluindo os bancos citados acima) armazenando esse tipo de informação derivada desde interações verificadas experimentalmente até aparições conjuntas de proteínas em resumos de artigos científicos. Dessa forma, é possível obter a informação de diferentes bancos de dados em uma visualização concentrada, ideal para que o pesquisador seja capaz de integrar tal informação e inferir suas conclusões de modo mais completo e prático.

1.10.1 Conclusão

A Biologia Molecular atual gera tal quantidade de informação que seria impraticável realizar análises sistêmicas eficientes sem ferramentas computacionais adequadas, além da tendência de disponibilidade pública de dados permitir aos cientistas uma grande economia de tempo e recursos. Contudo, é de extrema importância ressaltar que a Bioinformática não é capaz de resolver por si só os problemas enfrentados pela Biologia Molecular. Ela é também uma ferramenta capaz de atribuir significados e gerar interpretações falsas a partir de resultados experimentais. A busca por padrões e a inferência de predições a partir de dados biológicos fatalmente resultará em erros sem uma compreensão profunda de onde provêm os dados biológicos e em que contexto estão inseridos. O potencial da Bioinformática está em exatamente ser uma ferramenta, como os métodos recentes da Biologia Molecular, expandindo os limites do que pode ser estudado.

2 OBJETIVO

Com base nos avanços que a Biologia Molecular vem apresentando nos últimos anos, diversas novas técnicas são desenvolvidas e disponibilizadas aos pesquisadores. Contudo, muitas vezes esse desenvolvimento não acompanha metodologias de análise apropriadas para a geração de dados relevantes e confiáveis no contexto da pesquisa em questão. Um exemplo de técnica que até hoje recebe críticas quanto aos resultados que produz é o microarranjo de DNA, apesar de já ser bem-aceita e amplamente utilizada pela comunidade científica. Outros tipos de microarranjos, como o de substratos peptídicos de proteínas quinases, encontram dificuldades em se estabelecer como técnica corrente nos laboratórios devido à falta de metodologias apropriadas de análise que facilitem a obtenção de informações relevantes pelo usuário final. Nesse contexto, o objetivo deste trabalho é o desenvolvimento de um *pipeline* automatizado que permita a análise dos resultados do PepChip®, um microarranjo contendo substratos de proteínas quinases voltado a elucidação da atividade quinásica e a integração dessa informação no contexto de vias metabólicas e de sinalização celular.

3 RESULTADOS

3.1 PepMatrix

O principal resultado deste trabalho é o desenvolvimento de um método para análise estatística qualitativa automatizado do PepChip®. Batizado de PepMatrix, ele permite a determinação da confiabilidade de cada *spot* do *chip*, uma vez que o número baixo de replicações disponíveis aos pesquisadores é uma característica inerente à técnica e essa prática é corrente em outros tipos de microarranjos. As particularidades do PepChip®, no entanto, exigiram uma abordagem peculiar, que contemplasse as necessidades de uso dos pesquisadores.

3.1.1 AS PARTICULARIDADES DO *CHIP*

Microarranjos, sejam de proteínas ou de ácidos nucleicos, apresentam como característica marcante um baixo número de replicações, tornando extremamente difícil avaliar os resultados obtidos e determinar diferenças estatisticamente significativas entre os níveis de fosforilação, no caso do PepChip®, em *chips* representando condições biológicas distintas. Com o aumento no uso desse tipo de tecnologia, especialmente a de microarranjos de DNA, a variabilidade dos dados e a avaliação da qualidade dos resultados tem sido bastante discutida (MURIE *et al.*, 2009; ROCKE *et al.*, 2009; KOOPERBERG *et al.*, 2002; JEFFERY *et al.*, 2006) e gerou diversos métodos de análise (SCHUCHHARDT *et al.*, 2000; KAUFFMANN *et al.*, 2009), bem como algumas ferramentas para avaliação da qualidade de tipos específicos de arranjos (DUNNING *et al.*, 2007; BUNESS *et al.*, 2005). No entanto, como é de se esperar, as metodologias desenvolvidas para microarranjos de DNA não são facilmente adaptáveis para microarranjos de proteínas, sendo a principal razão para isso a diferença fundamental no objetivo de cada

experimento: enquanto os primeiros procuram avaliar os níveis de expressão do mRNA associado a um determinado gene e seu produto proteico, os últimos identificam a atividade de uma enzima ao invés de sua presença na célula. Os parâmetros ideais para determinação desses distintos objetivos são também, naturalmente, diferentes. Como exemplo, as ferramentas de avaliação de microarranjos de DNA frequentemente trazem limites inferiores de intensidade de *spot* para afirmar que o mRNA de determinado gene hibridizou-se com o cDNA presente na placa. Quanto maior a intensidade, maior a expressão do gene em questão. Já no caso de arranjos de proteínas, tomando como exemplo o PepChip®, intensidades muito altas podem indicar não uma grande quantidade de quinases presentes na célula promovendo essas reações (ainda mais levando em consideração que as vias de sinalização não dependem de grandes quantidades de enzimas, uma vez que uma única é capaz de fosforilar centenas de substratos, como visto na Introdução). Ao contrário, intensidades altas podem indicar fosforilações inespecíficas, com várias quinases fosforilando um determinado substrato no *chip*. Assim, a metodologia de análise do PepChip® demanda uma abordagem específica e desenvolvida levando em conta tais particularidades.

3.1.2 O ALGORITMO E A ESTATÍSTICA

O primeiro passo da metodologia de análise envolve a importação dos dados presentes nos arquivos DAT gerados a partir das imagens do *chip* de forma que seja possível manipulá-los satisfatoriamente. Esses arquivos contêm informações como a intensidade bruta de cada *spot*, seus respectivos *backgrounds* (*pixels* ao redor do *spot*) médios e medianos, além de outras informações que permitem identificar cada *spot*. Os arquivos DAT contêm informações em formato texto, separadas por tabulações e facilmente importáveis para um editor de planilhas.

Levando em consideração o *feedback* de usuários do PepChip®, estabeleceu-se que o objetivo específico era determinar quais *spots* apresentavam condições de serem levados adiante em análises posteriores. Para isso, conceitos estatísticos foram calculados a partir das informações dos arquivos DAT, visando a construção de uma matriz contendo apenas *spots* que apresentaram replicações com variações aceitáveis, determinadas por dois índices de confiabilidade.

No caso do PepChip® Kinomics™, que possui 1024 substratos em três replicações por *chip*, três conceitos estatísticos principais agrupados em dois índices de confiabilidade são calculados: o desvio-padrão (SD), a média (A) e a mediana (M) das três replicações (cada uma definida como o valor bruto de intensidade de cada *spot* subtraído da intensidade da mediana dos *pixels* do *background*), sendo que um dos índices é composto pela razão SD/A e o outro pela razão A/M. O segundo índice é irrelevante para a versão Kinase™ do *chip*, uma vez que com apenas duas replicações, $M = A$. Os valores recomendados de cada índice são $SD/A < 20\%$ e $80\% < A/M < 120\%$, uma vez que nos testes realizados com dados reais foram os valores que apresentaram o melhor equilíbrio entre a qualidade e a quantidade dos *spots* retornados. No entanto, o usuário é livre para alterar esses valores conforme for apropriado.

Ajustando-se cada um desses índices de forma que seja possível selecionar apenas *spots* que satisfaçam as condições estabelecidas, tem-se uma maneira extremamente simples de determinar individualmente a robustez dos valores de intensidade apresentados por um *spot* com base em suas replicações. O conceito de avaliação individual é extremamente importante porque as metodologias de análise anteriores, derivadas de técnicas para análise de microarranjos de DNA, valiam-se da correlação entre os *spots* de diferentes replicatas, analisando o *chip* como um todo. No entanto, se apenas alguns poucos *spots* apresentarem replicações muito divergentes e, portanto, inadequados para futura inferência biológica, a correlação entre as replicações não será

afetada significativamente. Na prática, um pesquisador que se baseie apenas na correlação geral entre replicações pode utilizar em sua pesquisa *spots* inadequados, acreditando que a qualidade do *chip* como um todo é alta, conforme visto na figura 5.

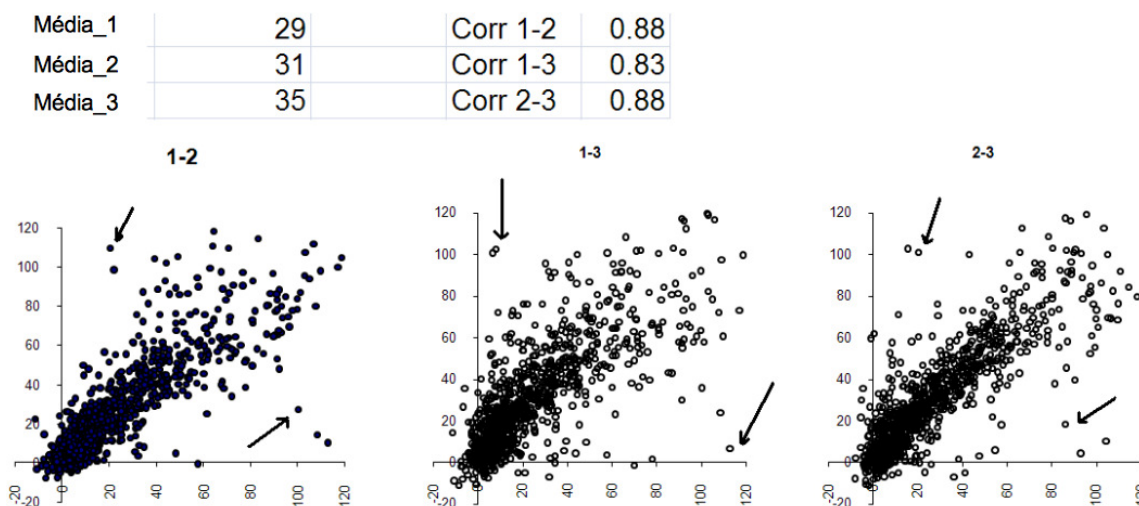


Figura 5. Índices de correlação e respectivos gráficos de um PepChip® Kinomics™. Os índices de correlação apresentam-se todos acima de 0,8, o que é tomado como adequado pelos pesquisadores usuários do *chip*, já que índices próximos de 1 são inatingíveis na prática. No entanto, observa-se claramente que muitos *spots* apresentam replicações inadequadas, muitas vezes chegando a variações de até 1000% (setas).

Existem diversas possibilidades de uso dos *chips*: a principal é a comparação entre *chips* de diferentes condições biológicas realizada de forma a determinar o perfil quinômico de cada uma. Uma outra possibilidade é verificar quais são os substratos mais fosforilados num determinado momento celular, o que envolveria apenas um *chip*. No primeiro caso, a comparação se dá *spot* a *spot*, observando as diferenças de fosforilação entre os dois *chips*. Da mesma forma, a seleção dos *spots* chamados confiáveis ocorre simultaneamente nos dois *chips*. De nada adiantaria um *spot* apresentar replicações dentro dos parâmetros estabelecidos se sua contraparte no outro *chip* também não o fizer.

3.1.3 A NORMALIZAÇÃO

Os *spots* selecionados são, então, submetidos a um processo de normalização para permitir a comparação entre *chips* diferentes. Um valor p é definido para rotular os valores extremos de intensidade, com base em um percentil. Seu valor é definido como v_p , correspondendo ao p -ésimo percentil para um conjunto de dados contendo Z elementos com valores v_1, v_2, \dots, v_N , como em

$$n = \frac{p}{100}(Z-1)+1 \quad (1)$$

onde n é um valor intermediário posteriormente separado em seus componentes inteiro, k e decimal, d , tal que $n = k + d$. v_p é então calculado como

$$v_p = v_k + d(v_{k+1} - v_k) \quad (2)$$

A normalização prossegue definindo os valores de percentis nos dois extremos dos dados (de modo a eliminar os $p\%$ *spots* menos intensos e os $p\%$ *spots* mais intensos do *chip*, visando evitar incluir nos dados tanto ausências completas de fosforilação como fosforilações inespecíficas, que não são o foco de análise do *chip*. De qualquer forma, esses parâmetros são ajustáveis e, caso o usuário deseje não excluir esses valores, isso é perfeitamente possível. Os valores normalizados são dados por

$$N = \frac{R - v_L}{v_H - v_L} \cdot 100 \quad (3)$$

onde N corresponde aos valores normalizados, v_L é o percentil inferior e v_H é o percentil superior.

A tabela 1 apresenta como os valores de intensidade são alocados para os cálculos que geram os valores de intensidade que serão submetidos à metodologia de seleção, além da normalização que

será utilizada na comparação entre *chips*. A tabela 2 mostra os cálculos dos parâmetros que compõem os índices e a seleção propriamente dita.

Tabela 1. Primeiros 43 *spots* de um PepChip® Kinomics™. A. As colunas Int_1, Int_2 e Int_3 correspondem aos valores brutos de intensidade de cada replicação, respectivamente. As colunas BG_1, BG_2 e BG_3 correspondem aos valores de intensidade de background relativos a cada *spot*. Os resultados da subtração das colunas BG das colunas Int resultam nos valores das colunas Val_1, Val_2 e Val_3. Esses são os valores utilizados para a seleção dos *spots* com base nos índices SD/A e A/M. B. Os valores normalizados estão apresentados nas colunas Norm_1, Norm_2 e Norm_3 e são utilizados na comparação entre dois *chips*, através dos *spots* selecionados previamente.

A

Image	controlmc3w								Min	-2692	-2838	-2929
Date	17/01/08								Range	27135	29689	26505
Spot_Nr	Row	Col	Int_1	BG_1	Int_2	BG_2	Int_3	BG_3	Val_1	Val_2	Val_3	
1	1	1	17897	17354	17445	17284	19926	19078	543	161	848	
2	1	2	13452	16726	18300	17668	19305	19388	-3274	632	-83	
3	1	3	17520	16224	18642	17844	17160	19664	1296	798	-2504	
4	1	4	19414	16180	18098	18220	19286	19894	3234	-122	-608	
5	1	5	14197	16162	17898	18394	17833	20180	-1965	-496	-2347	
6	1	6	16220	16454	17072	18376	17463	20586	-234	-1304	-3123	
7	1	7	17301	17056	17356	18302	22231	20878	245	-946	1353	
8	1	8	29306	16678	32082	18176	25907	20884	12628	13906	5023	
9	2	1	20100	17094	17040	17088	18647	18684	3006	-48	-37	
10	2	2	19849	16338	18341	17136	21604	18430	3511	1205	3174	
11	2	3	15381	16302	11236	17404	20353	18552	-921	-6168	1801	
12	2	4	16029	16604	18789	17852	20523	19126	-575	937	1397	
13	2	5	35889	16672	38405	18106	29410	19308	19217	20299	10102	
14	2	6	18178	17474	18343	18570	18629	20006	704	-227	-1377	
15	2	7	17478	17484	17774	18464	17886	20240	-6	-690	-2354	
16	2	8	14465	17282	19288	18216	23787	20092	-2817	1072	3695	
17	3	1	16261	17028	19232	16980	19195	18356	-767	2252	839	
18	3	2	17572	16604	19225	17000	21469	17884	968	2225	3585	
19	3	3	17324	17010	20803	17600	16584	18206	314	3203	-1622	
20	3	4	15877	17488	18429	18350	21905	19234	-1611	79	2671	
21	3	5	18207	17444	19118	18570	21595	19372	763	548	2223	
22	3	6	14882	18280	18245	19166	19239	20156	-3398	-921	-917	
23	3	7	19523	17782	20802	18460	20746	20030	1741	2342	716	
24	3	8	23526	17234	28924	18026	28138	19756	6292	10898	8382	
25	4	1	16733	17274	17511	17272	17012	18350	-541	239	-1338	
26	4	2	18212	16838	17196	17372	17157	17698	1374	-176	-541	
27	4	3	14988	17010	16381	17844	16558	18488	-2022	-1463	-1930	
28	4	4	20024	17282	15886	18280	17913	19212	2742	-2394	-1299	
29	4	5	54381	17274	59728	18390	41000	19292	37107	41338	21708	
30	4	6	17370	18288	21869	18698	20193	20006	-918	3171	187	
31	4	7	28956	17936	26426	18366	23949	19800	11020	8060	4149	
32	4	8	18147	17692	18234	18118	17564	19516	455	116	-1952	
33	5	1	20018	17702	19292	17474	17189	18484	2316	1818	-1295	
34	5	2	16118	17608	17022	17924	18549	18180	-1490	-902	369	
35	5	3	17603	17702	18247	18362	20423	19068	-99	-115	1355	
36	5	4	16784	17722	16960	18834	29635	19898	-938	-1874	9737	
37	5	5	16139	17572	17682	18584	19022	19502	-1433	-902	-480	
38	5	6	19809	17674	23559	18708	24516	19804	2135	4851	4712	
39	5	7	16154	17400	20414	18152	16335	19346	-1246	2262	-3011	
40	5	8	15438	17196	18099	18060	20166	19118	-1758	39	1048	
41	6	1	20039	18318	18279	17848	20460	19166	1721	431	1294	
42	6	2	38574	18596	33487	18688	32980	19406	19978	14799	13574	
43	6	3	18014	18438	16255	19008	20348	20434	-424	-2753	-86	

B

Spot_Nr	Norm_1	QC_1	Norm_2	QC_2	Norm_3	QC_3
1	12	0	10	0	14	0
2	-2	1	12	0	11	0
3	15	0	12	0	2	0
4	22	0	9	0	9	0
5	3	0	8	0	2	0
6	9	0	5	0	-1	1
7	11	0	6	0	16	0
8	56	0	56	0	30	0
9	21	0	9	0	11	0
10	23	0	14	0	23	0
11	7	0	-11	1	18	0
12	8	0	13	0	16	0
13	81	0	78	0	49	0
14	13	0	9	0	6	0
15	10	0	7	0	2	0
16	0	1	13	0	25	0
17	7	0	17	0	14	0
18	13	0	17	0	25	0
19	11	0	20	0	5	0
20	4	0	10	0	21	0
21	13	0	11	0	19	0
22	-3	1	6	0	8	0
23	16	0	17	0	14	0
24	33	0	46	0	43	0
25	8	0	10	0	6	0
26	15	0	9	0	9	0
27	2	0	5	0	4	0
28	20	0	1	0	6	0
29	147	2	149	2	93	0
30	7	0	20	0	12	0
31	51	0	37	0	27	0
32	12	0	10	0	4	0
33	18	0	16	0	6	0
34	4	0	7	0	12	0
35	10	0	9	0	16	0
36	6	0	3	0	48	0
37	5	0	7	0	9	0
38	18	0	26	0	29	0
39	5	0	17	0	0	1
40	3	0	10	0	15	0
41	16	0	11	0	16	0
42	84	0	59	0	62	0
43	8	0	0	0	11	0

Tabela 2. Os mesmos 43 *spots* mostrados na tabela anterior agora alocados no passo seguinte da metodologia PepMatrix, que consiste no cálculo dos índices SD/A e A/M. Isso permite a seleção dos *spots* com base na prévia determinação dos parâmetros relacionados a esses índices, como observado na última coluna da tabela: dentre todos os *spots* mostrados, apenas o *spot* presente na linha 29 apresentou replicações consistentes o bastante para sobreviver aos parâmetros selecionados ($SD/A < 20\%$ e $80\% < A/M < 120\%$). Como esse é um *chip* contendo uma condição controle e existe um outro *chip* contendo a condição tratada, o *spot* de número 29 ainda não está necessariamente aprovado: apenas o será se seu correspondente no outro *chip* também apresentar replicações consistentes.

1	SL1 - controlmc3w										
2	Val_1	Val_2	Val_3	Average	SD	Median	SD/A	Below 20%	A/M	Between 80% and 120%	✓
3	543	161	848	517.33	344.22	543	66.54%		95.27%	1	
4	-3274	632	-83	-908.33	2079.69	-83	228.96%		1094.38%		
5	1296	798	-2504	-136.67	2065.24	798	1511.15%		-17.13%		
6	3234	-122	-608	834.67	2092.04	-122	250.64%		-684.15%		
7	-1965	-496	-2347	-1602.67	977.25	-1965	60.98%		81.56%	1	
8	-234	-1304	-3123	-1553.67	1460.59	-1304	94.01%		119.15%	1	
9	245	-946	1353	217.33	1149.75	245	529.03%		88.71%	1	
10	12628	13906	5023	10519.00	4802.38	12628	45.65%		83.30%	1	
11	3006	-48	-37	973.67	1760.06	-37	180.77%		-2631.53%		
12	3511	1205	3174	2630.00	1245.54	3174	47.36%		82.86%	1	
13	-921	-6168	1801	-1762.67	4050.62	-921	229.80%		191.39%		
14	-575	937	1397	586.33	1031.71	937	175.96%		62.58%		
15	19217	20299	10102	16539.33	5601.08	19217	33.87%		86.07%	1	
16	704	-227	-1377	-300.00	1042.42	-227	347.47%		132.16%		
17	-6	-690	-2354	-1016.67	1207.60	-690	118.78%		147.34%		
18	-2817	1072	3695	650.00	3276.45	1072	504.07%		60.63%		
19	-767	2252	839	774.67	1510.53	839	194.99%		92.33%	1	
20	968	2225	3585	2259.33	1308.84	2225	57.93%		101.54%	1	
21	314	3203	-1622	631.67	2428.14	314	384.40%		201.17%		
22	-1611	79	2671	379.67	2156.78	79	568.07%		480.59%		
23	763	548	2223	1178.00	911.36	763	77.36%		154.39%		
24	-3398	-921	-917	-1745.33	1431.25	-921	82.00%		189.50%		
25	1741	2342	716	1599.67	822.16	1741	51.40%		91.88%	1	
26	6292	10898	8382	8524.00	2306.28	8382	27.06%		101.69%	1	
27	-541	239	-1338	-546.67	788.52	-541	144.24%		101.05%	1	
28	1374	-176	-541	219.00	1016.77	-176	464.28%		-124.43%		
29	-2022	-1463	-1930	-1805.00	299.73	-1930	16.61%	1	93.52%	1	1
30	2742	-2394	-1299	-317.00	2705.16	-1299	853.36%		24.40%		
31	37107	41338	21708	33384.33	10330.92	37107	30.95%		89.97%	1	
32	-918	3171	187	813.33	2115.23	187	260.07%		434.94%		
33	11020	8060	4149	7743.00	3446.45	8060	44.51%		96.07%	1	
34	455	116	-1952	-460.33	1302.89	116	283.03%		-396.84%		
35	2316	1818	-1295	946.33	1956.96	1818	206.79%		52.05%		
36	-1490	-902	369	-674.33	950.18	-902	140.91%		74.76%		
37	-99	-115	1355	380.33	844.12	-99	221.94%		-384.18%		
38	-938	-1874	9737	2308.33	6450.41	-938	279.44%		-246.09%		
39	-1433	-902	-480	-938.33	477.54	-902	50.89%		104.03%	1	
40	2135	4851	4712	3899.33	1529.54	4712	39.23%		82.75%	1	
41	-1246	2262	-3011	-665.00	2684.08	-1246	403.62%		53.37%		
42	-1758	39	1048	-223.67	1421.32	39	635.46%		-573.50%		
43	1721	431	1294	1148.67	657.17	1294	57.21%		88.77%	1	
44	19978	14799	13574	16117.00	3399.36	14799	21.09%		108.91%	1	
45	-424	-2753	-86	-1087.67	1452.09	-424	133.50%		256.53%		

3.1.4 TESTES ESTATÍSTICOS

Após os valores de intensidade terem sido normalizados, podem então ser comparados de modo a definir se são ou não significativamente diferentes entre dois *chips*. O teste estatístico utilizado para isso é o teste t bicaudal. A natureza paramétrica deste teste, que exige condições normais, não interfere devido ao baixo número de replicações, o que impede determinações precisas de normalidade. No entanto, um segundo requerimento do teste é o conhecimento sobre as variâncias das replicações. Para isso, é utilizado o teste F . Se as variâncias não forem estatisticamente diferentes, o teste t aplicado é o homoscedástico. Caso contrário, é utilizado o teste t heteroscedástico, garantindo maior precisão nos resultados finais.

3.1.5 NÚCLEO DE INFORMAÇÕES DA MATRIZ

Após todos os testes terem sido realizados e um destaque aplicado aos *spots* que apresentaram fosforilação diferencial, todos os *spots* que se mostraram confiáveis de acordo com os parâmetros definidos são associados a suas respectivas informações-chave, obtidas diretamente com o fabricante do *chip*: um identificador de localização, a sequência peptídica, o identificador UniProt ou RefSeq da proteína doadora da sequência, o nome da proteína e a provável quinase responsável pelo evento de fosforilação observado. O resultado final da análise é mostrado na tabela 3.

Tabela 3. A. *Spots* selecionados a partir de dois *chips* e com índices definidos como $SD/A < 20\%$ e $80\% < A/M < 120\%$. Dos 1024 *spots* presentes no PepChip® Kinomics™, 32 apresentaram replicações consistentes nos dois *chips* e foram selecionados para a análise diferencial. Destes, 23 apresentaram fosforilação diferencial ($p < .05$) na coluna R (Parte B) e 9 não apresentaram diferenças estatisticamente significativas na fosforilação. As colunas B, C, D e H, I, J contêm os valores de intensidade normalizados para os *spots* selecionados de cada *chip*, respectivamente. As colunas E, F e K, L apresentam os valores de média e mediana para cada *spot*. Já as colunas O e P contêm a razão de alteração observada (conhecida como FoldChange ou FC) entre as médias na coluna O e entre as medianas na coluna P. B. A coluna Q contém os valores do teste *F* para a determinação da existência de diferença significativa entre as variâncias das replicações entre os dois *chips*. Caso exista diferença ($p < .05$), o teste *t* realizado na coluna R é heteroscedástico. Caso contrário, o teste *t* realizado é homoscedástico. Os *spots* estão ordenados de forma ascendente pela coluna R. Nas colunas T, U, V, W e X estão dispostas as informações específicas sobre cada *spot*: respectivamente, o identificador no *chip*, a sequência presente no *spot*, o identificador RefSeq ou SwissProt dessa sequência, o nome da proteína de qual a sequência foi retirada e a provável quinase responsável pelo evento de fosforilação naquele *spot*.

A

	B	C	D	E	F	H	I	J	K	L	O	P
1	controlr	SL1		average	median	L0352-2	SL2		average	median	FC avg	FC median
2	15.383	13.754	17.152	15.43	15.383	43.179	44.836	42.095	43.37	43.179	2.810855	2.80701901
3	178.23	162.39	174.72	171.78	174.72	362.12	346.12	340.21	349.48	346.12	2.034493	1.98099469
4	47.33	36.247	42.702	42.093	42.702	398.76	369.27	347.93	371.99	369.27	8.837242	8.64760198
5	168.16	139.47	158.76	155.46	158.76	379.67	372.31	357.45	369.81	372.31	2.378766	2.34515584
6	184.33	161.36	185.02	176.9	184.33	350.59	373.38	368.11	364.03	368.11	2.057829	1.99704442
7	167.14	122.59	158.52	149.41	158.52	372.72	364.51	368.62	368.62	368.62	2.46707	2.32541373
8	19.565	21.403	24.204	21.724	21.403	50.78	50.869	58.144	53.265	50.869	2.451872	2.37673339
9	168.59	148.39	166.78	161.25	166.78	376.02	317.66	351.11	348.26	351.11	2.159701	2.10522571
10	67.12	80.751	85.936	77.936	80.751	166.73	158	141.83	155.52	158	1.995469	1.95661681
11	78.449	78.531	111.25	89.409	78.531	374.74	376.18	375.4	375.44	375.4	4.199099	4.78021361
12	23.859	23.148	28.607	25.204	23.859	71.727	55.653	58.653	62.011	58.653	2.46033	2.45834255
13	72.327	62.266	62.593	65.729	62.593	38.448	33.268	26.8	32.839	33.268	0.49961	0.53150117
14	138.43	137.22	148.01	141.22	138.43	291.22	314.14	245.36	283.58	291.22	2.008053	2.10376811
15	86.836	68.982	88.826	81.548	86.836	41.585	37.255	33.035	37.292	37.255	0.4573	0.42902721
16	116.27	141.25	134.07	130.53	134.07	82.378	71.718	93.886	82.661	82.378	0.633257	0.61442241
17	91.019	80.478	69.535	80.344	80.478	138.79	154.84	119.88	137.84	138.79	1.715623	1.72459767
18	180.14	160.08	177.82	172.68	177.82	257.94	348.16	286.43	297.51	286.43	1.722904	1.61076858
19	85.229	90.172	93.965	89.789	90.172	70.989	68.599	50.079	63.222	68.599	0.704124	0.76075886
20	88.495	114.84	123.1	108.81	114.84	163.88	143.17	145.27	150.77	145.27	1.385614	1.26491958
21	165.64	141.93	159.81	155.79	159.81	255.13	248.63	188.8	230.86	248.63	1.481824	1.55585709
22	165.45	148.82	154.21	156.16	154.21	138.44	110.79	98.945	116.06	110.79	0.743186	0.71842856
23	169.32	141.78	152.3	154.47	152.3	241.77	256.89	183.21	227.29	241.77	1.471468	1.58751273
24	175.46	162.14	175.78	171.13	175.46	139.49	158.77	125.08	141.11	139.49	0.824624	0.79499349
25	149.61	136.6	155.69	147.3	149.61	112.35	136.73	104.44	117.84	112.35	0.800011	0.75097251
26	35.994	32.818	30.904	33.239	32.818	30.146	18.024	20.332	22.834	20.332	0.686972	0.61953098
27	160.66	150.38	167.59	159.54	160.66	373.67	235.84	280.81	296.77	280.81	1.860134	1.74788455
28	73.58	58.544	72.678	68.267	72.678	85.435	86.434	73.17	81.68	85.435	1.196465	1.17552769
29	97.402	97.316	110.39	101.7	97.402	85.854	99.262	83.347	89.488	85.854	0.879891	0.88143843
30	102.62	83.654	87.751	91.34	87.751	76.521	86.258	69.801	77.527	76.521	0.848767	0.87203048
31	131.96	112.7	106.83	117.16	112.7	109.21	87.56	113.27	103.35	109.21	0.882061	0.96899181
32	104.1	102.17	89.547	98.606	102.17	100.08	89.289	69.35	86.24	89.289	0.874596	0.87390871
33	163.32	150.68	156.12	156.71	156.12	183.57	176.27	138.72	166.19	176.27	1.060499	1.12905443

B

	Q	R	T	U	V	W	X
1	F-test (norm)	T-test (norm)	Scanalyze	MOTIF	PROTEIN ID	PROTEIN	UP KIN
2	0.79487	0.00002	578	DVLKFYDSNTV	NP_002568	PAK2	Src
3	0.70077	0.00003	771	RKGHEYTNIKY	NP_002825	SHP2	PDGFRbeta
4	0.09083	0.00003	894	KKPKKKSCLLL	NP_002875	Ras related	PKA
5	0.74888	0.00004	627	RKKRISVKKKQ	NP_059431	Cytohesin 2	PKC
6	0.87939	0.00006	189	RKKKVSSTKRH	NP_004753	Cytohesin-1	PKC
7	0.05852	0.00009	383	APKAPSKKEKK	NP_001334	DAB2	PKC
8	0.46799	0.00035	711	SptCtrl	0	Used for pro na	
9	0.25420	0.00049	746	RSKKNLALSL	NP_000116	Estrogen rec	PAK1
10	0.74360	0.00108	573	PTPPLSPSRRS	NP_002458	c-Myc	CK2;GSK
11	0.00289	0.00143	250	TRVPPSRGPD	NP_000672	Alpha 2A adre	PKC
12	0.21521	0.00214	758	AEVLSPRGQR	NP_001058	DNA topoisom	Proline_directed
13	0.97948	0.00222	95	REWHKTTQMSA	NP_002410	MAP3K11	MAP3K11
14	0.05537	0.00226	605	RLRPRTKVKKS	NP_038202	Neutrophil c	PKC
15	0.26548	0.00284	700	GPRLVSNHSLH	NP_001447	Filamin A	CAMKII
16	0.85251	0.00814	756	MRRQRSAPDLK	NP_006603	Kinesin fami	CK2
17	0.54742	0.00834	399	KTNLSYEEYDK	AAC08966	Bone marrow	FAK
18	0.10723	0.01033	127	KFRTPSFLKKN	NP_058432	Adducin 3	PKCalpha
19	0.25551	0.01986	64	GFTEESGDDEY	NP_005988	TCFL1	CK2
20	0.56925	0.02726	689	RSNPPSRKSGS	NP_000157	Connexin 32	PKC;PKA
21	0.20498	0.02808	635	LKPGSSHRKTK	NP_000052	Bruton's tyr	PKCbeta
22	0.29846	0.03413	739	KKGARSRLFS	NP_002059	Guanine nucl	PKCalpha
23	0.22621	0.03794	634	KKKKPSRLKGD	NP_000029	APC	PKA
24	0.34999	0.04917	104	MRKKISNAQLQ	NP_003469	Cullin 5	PKA
25	0.50286	0.05859	699	LLSELSRRRIR	NP_004085	eIF2 alpha	PKR;GCN2;Eukaryo
26	0.27516	0.06005	489	PSGLWSPAYAS	NP_079199	NUP210	CyclinB;p34cdc2
27	0.02989	0.07530	613	RGKKKSGCLVL	NP_001655	RhoA	PKA
28	0.86832	0.10696	53	KALPLSPRKRL	NP_001245	CDC6	CDK2
29	0.87210	0.13692	422	TKALQSPKRPR	NP_127492	TFII-I	EphB2
30	0.81498	0.13867	537	IHMVYSKRSRK	NP_003080	Small nuclea	nd
31	0.95019	0.27791	670	APLLSTPKRER	NP_015556	Kinesin like	Cyclin_B2
32	0.40897	0.28768	548	I AVRKSRDKAK	NP_005185	CCAAT/Enhanc	ERK2
33	0.12995	0.54541	633	LSRRPSYRKIL	NP_604391	cAMP respons	PKA;MAPK;CaMK;RS

3.1.6 O CÓDIGO

A metodologia PepMatrix foi implementada visando a maior usabilidade e disponibilidade possível. Por conta disso, foi escrita em VBScript, linguagem universal em sistemas Microsoft® Windows™ e facilmente compilada para execução nesses sistemas. O

código, conforme implementado na versão 2DAT (ver adiante) do PepMatrix para o Microsoft® Excel™ 2003, segue abaixo.

```
=====
strTemplateFile = Replace(WScript.ScriptFullName, WScript.ScriptName, "") &
                  "PepMatrix.xls"

Set objExcel = CreateObject("Excel.Application")
Const xlUp = -4162
Const xlDown = -4121
Const xlAnd = 1
Const xlAscending = 1
Const xlDescending = 2
Const xlYes = 1
objExcel.Visible = True

objExcel.Workbooks.Open strTemplateFile
Set objTemplateWB = objExcel.ActiveWorkbook

Set objDialog = CreateObject("UserAccounts.CommonDialog")
objDialog.Filter = "DAT Files (*.dat)|*.dat;All Files (*.*)|*.*"
objDialog.FilterIndex = 1
objDialog.InitialDir = "."
intResult = objDialog.ShowOpen

If intResult = 0 Then
    MsgBox "No file was selected."
    objExcel.Quit
Else
    objExcel.Workbooks.Open objDialog.FileName
    Set objDATWB = objExcel.ActiveWorkbook
    objDATWB.Sheets(1).UsedRange.Copy objTemplateWB.Sheets(6).Cells(65536,
        "A").End(xlUp).Offset(0, 0)
    objDATWB.Activate
    objExcel.ActiveWorkbook.Close False

    objDialog.Filter = "DAT Files (*.dat)|*.dat;All Files (*.*)|*.*"
    objDialog.FilterIndex = 1
    objDialog.InitialDir = "."
    intResult = objDialog.ShowOpen

    If intResult = 0 Then
        Set objDATWB = objExcel.ActiveWorkbook
        objDATWB.Activate
        objExcel.ActiveWorkbook.Close False
        objTemplateWB.Activate
        objExcel.ActiveWorkbook.SaveAs objExcel.ActiveWorkbook.Path &
            Replace(objExcel.ActiveWorkbook.Name, ".xls", "") &
            FormatDateToString(Now) & ".xls"
        objExcel.ActiveWorkbook.Close False
        objExcel.Quit
    Else

```

```

objExcel.Workbooks.Open objDialog.FileName
Set objDATWB = objExcel.ActiveWorkbook
objDaTWB.Sheets(1).UsedRange.Copy objTemplateWB.Sheets(5).Cells(65536,
    "A").End(xlUp).Offset(0, 0)

objDATWB.Activate
objExcel.ActiveWorkbook.Close False
Set objTemplateWB = objExcel.ActiveWorkbook
objTemplateWB.Activate
Set objTemp2 = objTemplateWB.Sheets("FC_norm_B")
Set objRange = objTemp2.UsedRange
Set objRange2 = objTemplateWB.Sheets("FC_norm_B").Range("R1")
objRange.Sort objRange2, xlAscending, , , , , xlYes

objExcel.ActiveWorkbook.SaveAs objExcel.ActiveWorkbook.Path & "\" &
    Replace(objExcel.ActiveWorkbook.Name, ".xls", "") & "_" &
    FormatDateTime(Now) & ".xls"
objExcel.ActiveWorkbook.Close False
objExcel.Quit

End If
End If

Function FormatDateToString(dDateTime)
    FormatDateToString = Year(dDateTime) & "-" & _
        Right("00" & Month(dDateTime), 2) & "-" & _
        Right("00" & Day(dDateTime), 2) & "-" & _
        Right("00" & Hour(dDateTime), 2) & "-" & _
        Right("00" & Minute(dDateTime), 2) & "-" & _
        Right("00" & Second(dDateTime), 2)
End Function
=====

```

O código utiliza um arquivo base consistindo numa tabela na qual os dados de replicações serão inseridos, permitindo que os parâmetros utilizados para os índices sejam facilmente alteráveis pelo usuário e permitindo a comparação imediata entre diferentes parâmetros de confiabilidade. Como padrão, a tabela base traz os índices $SD/A < 20\%$ e $80\% < A/M < 120\%$, estabelecidos de forma a atingir o máximo equilíbrio entre o número de *spots* selecionados e a qualidade de suas replicações. O objetivo geral do método é selecionar e trazer para o usuário apenas os *spots* que apresentem variações entre as replicatas dentro desses limites.

Existem diversas possibilidades de uso dos *chips*, o que demanda a adequação da metodologia PepMatrix a cada uma delas. Como ela foi construída em VBScript e depende do aplicativo Microsoft® Excel™, foram criadas versões adequadas para o Microsoft® Excel™ 2007, bem como para Microsoft® Excel™ 2003 e anteriores. Além disso, os *chips* podem ser usados em conjunto para a comparação de uma condição controle com uma condição alterada, o que demanda uma análise comparativa entre *chips*. Para esse caso existe uma versão do PepMatrix chamada de 2DAT, na qual, após a seleção dos *spots* confiáveis, são calculados o *FoldChange* (definido como a razão entre as médias ou medianas de intensidade de cada *spot* do tratamento em relação aos *spots* controle); o teste *F*, de modo a determinar se as variâncias entre as amostras eram comparáveis e o teste *t* apropriado em relação às variâncias de cada amostra (homoscedástico para variâncias semelhantes e heteroscedástico para variâncias diferentes). A partir do teste *t* são selecionados os *spots* que apresentam diferença significativa no valor da média ($p < 0,05$), permitindo ao pesquisador proceder à análise dos substratos e quinases correspondentes.

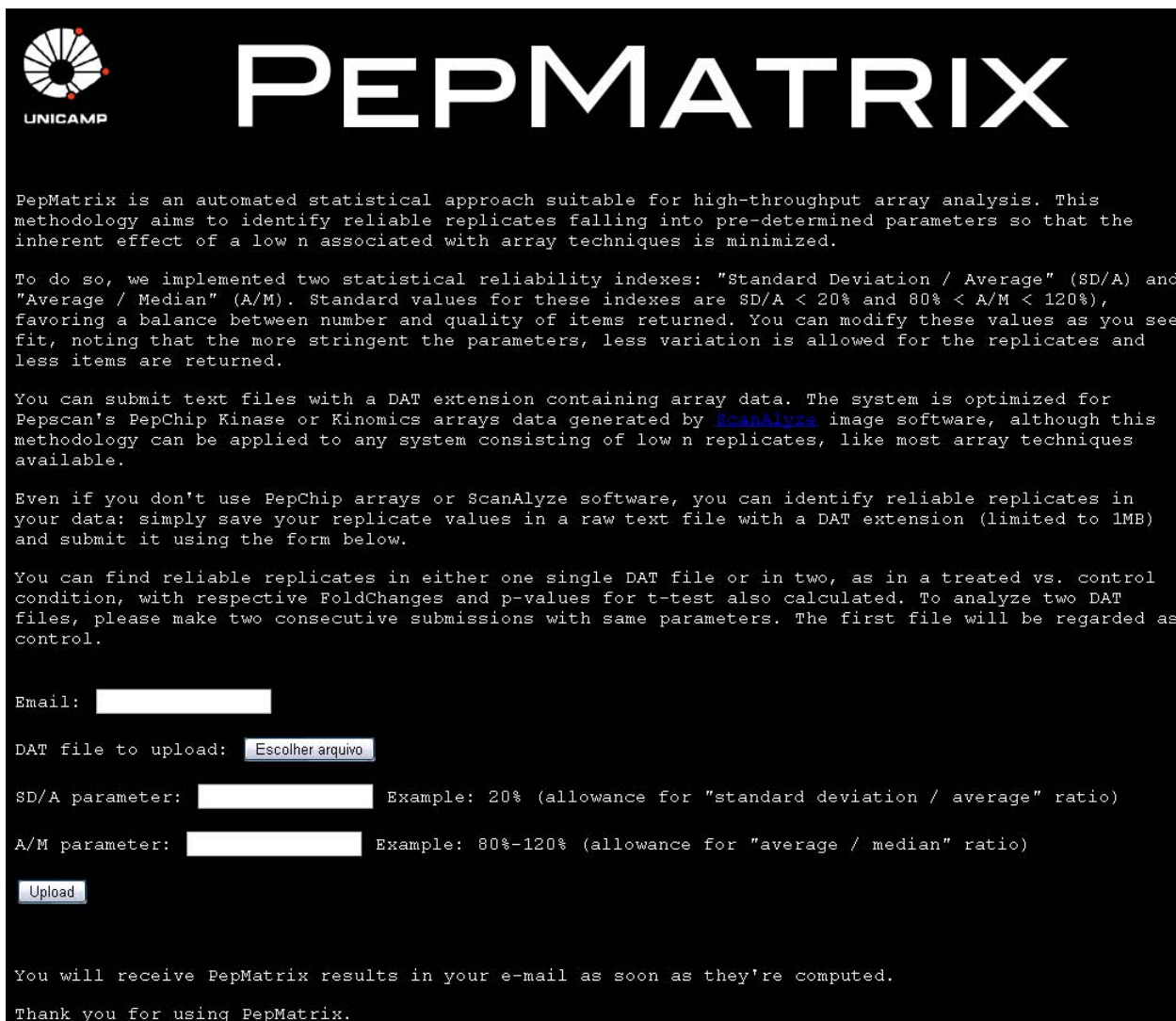
Uma outra possibilidade de uso do *chip* é verificar quais são os substratos mais fosforilados num determinado momento celular, o que envolveria apenas um *chip*. Para isso, também existe uma versão do PepMatrix chamada de SingleDAT, adaptada para realizar a seleção dos *spots* confiáveis e retorná-los por ordem de intensidade apresentada. Além disso, conforme já exposto, existe ainda um outro tipo de PepChip®, o Kinase™. Ao contrário da versão Kinomics™, que apresenta substratos quinásicos chave para o estudo de alterações da sinalização celular em doenças humanas, a versão Kinase™ possui substratos oriundos de diferentes espécies e é voltado para a caracterização de quinases e potenciais inibidores. O PepChip® Kinase™ possui 1176 substratos em apenas duas replicações, o que exige uma versão


do PepMatrix, a Kinase, adequada tanto para o número de replicações e substratos quanto para os índices estatísticos que são utilizados, já que com apenas duas replicações o uso da mediana torna-se irrelevante. É importante ressaltar o fato de que não é necessário utilizar apenas um *chip* para cada condição analisada, de modo que quanto mais replicações, tanto técnicas (utilizando a mesma amostra) quanto biológicas (utilizando amostras equivalentes), estiverem disponíveis, existe uma tendência ao enriquecimento da qualidade dos dados analisados. Contudo, o uso de mais de um *chip* para cada condição biológica é uma impossibilidade técnica ou financeira para muitos laboratórios.

3.1.7 A INTERFACE COM O USUÁRIO

Foi construído e disponibilizado um *website* onde pesquisadores usuários do PepChip® podem submeter os arquivos DAT relacionados a seus experimentos à metodologia PepMatrix. No próprio *site* o usuário define os parâmetros estatísticos desejados e depois recebe uma tabela com os *spots* selecionados no endereço de *e-mail* informado. Além disso, a metodologia, devido a sua natureza de seleção de replicações consistentes, pode ser aplicada para qualquer conjunto de dados com baixas replicações, como por exemplo outros tipos de microarranjos e fosfoproteomas determinados por espectrometria de massas. O *website* está também preparado para aceitar submissões desses tipos de dados. A figura 6 mostra a página inicial do *website*, de onde é possível também acessar os arquivos executáveis associados à metodologia. Através do *download* desses arquivos, um pesquisador pode implementar a metodologia PepMatrix localmente.

A página está disponível no endereço <http://pro-161-70.ib.unicamp.br/~renatomilani/PHP/>. Um endereço de atalho, conhecido como *quicklink* de fácil memorização está localizado em <http://tinyurl.com/pepmatrixonline>.



 **PEPMATRIX**

PepMatrix is an automated statistical approach suitable for high-throughput array analysis. This methodology aims to identify reliable replicates falling into pre-determined parameters so that the inherent effect of a low n associated with array techniques is minimized.

To do so, we implemented two statistical reliability indexes: "Standard Deviation / Average" (SD/A) and "Average / Median" (A/M). Standard values for these indexes are $SD/A < 20\%$ and $80\% < A/M < 120\%$, favoring a balance between number and quality of items returned. You can modify these values as you see fit, noting that the more stringent the parameters, less variation is allowed for the replicates and less items are returned.

You can submit text files with a DAT extension containing array data. The system is optimized for Pepscan's PepChip Kinase or Kinomics arrays data generated by [Scanalyze](#) image software, although this methodology can be applied to any system consisting of low n replicates, like most array techniques available.

Even if you don't use PepChip arrays or Scanalyze software, you can identify reliable replicates in your data: simply save your replicate values in a raw text file with a DAT extension (limited to 1MB) and submit it using the form below.

You can find reliable replicates in either one single DAT file or in two, as in a treated vs. control condition, with respective FoldChanges and p-values for t-test also calculated. To analyze two DAT files, please make two consecutive submissions with same parameters. The first file will be regarded as control.

Email:

DAT file to upload:

SD/A parameter: Example: 20% (allowance for "standard deviation / average" ratio)

A/M parameter: Example: 80%-120% (allowance for "average / median" ratio)

You will receive PepMatrix results in your e-mail as soon as they're computed.

Thank you for using PepMatrix.

Figura 6. *Website* da metodologia PepMatrix. Através dessa página, pode-se submeter arquivos DAT contendo replicações obtidas através do PepChip®.

3.2 Perfil quinômico de células TF1 expostas à violaceína

A violaceína é um composto antibiótico produzido pela bactéria *Chromobacterium violaceum*. Esta bactéria é gram-negativa e anaeróbica facultativa, sendo parte da água e solo de regiões tropicais e subtropicais. A violaceína mostra atividade contra amebas e tripanossomas, sendo que existem evidências de que pode auxiliar no tratamento de câncer colorretal e outros tipos de câncer através da modulação de vias de sinalização (KODACH *et al.*, 2006).

Como parte de um estudo amplo em colaboração sobre os efeitos da violaceína em sistemas celulares, o perfil quinômico de células TF1 (derivadas de um paciente com leucemia mieloide) foi obtido através do PepChip® Kinomics™ pelo Laboratório de Transdução de Sinal e Bioensaios – IB – Unicamp e os conjuntos de dados obtidos foram submetidos à metodologia PepMatrix. As células TF1 foram expostas à violaceína por 30 minutos, 16 horas e 24 horas, antes de terem seus lisados processados e submetidos ao *chip* conforme descrito anteriormente. Um controle sem exposição foi utilizado. A variabilidade encontrada nas replicações de todos os *chips* foi, conforme esperado, bastante alta, sendo que a metodologia PepMatrix provou-se útil em selecionar apenas os *spots* com replicações consistentes. A figura 7A mostra as três replicações para o *chip* de 16 horas e a distribuição das intensidades de fosforilação. Na figura 7B pode-se observar as razões SD/A para cada *spot*, em duas escalas diferentes.

No *chip* de 30 minutos, 142 *spots* foram selecionados como tendo replicações confiáveis de acordo com os parâmetros $SD/A < 20\%$ e $80\% < M/A < 120\%$, que foram selecionados buscando o melhor equilíbrio entre o número de *spots* selecionados e a qualidade de suas replicações. Nos tempos de 16h e 24h foram selecionados, respectivamente, 69 e 75 *spots*. De todos os *spots* selecionados, 84 apresentaram atividade quinásica diferencial em relação ao

controle ($p < .05$) (Tabelas 4, 5 e 6 e figuras 8, 9 e 10). Dos 84 *spots* com fosforilação diferencial, foram realizadas análises por *Western blot* para GSK3 β e Src devido à sua importância para vias de proliferação celular. Os resultados mostram um notável aumento da fosforilação do resíduo inibitório S9 da enzimas GSK3 β , confirmando a redução de 26% da atividade de todas as fosforilações atribuídas a GSK3 β no *chip* de 30 minutos (Figuras 8, 11 e tabela 4). Além disso, a atividade da quinase Src não mostrou diferenças significativas ($p > .05$) no mesmo tempo de exposição, confirmando a densidade constante no *blot* para os resíduos Y416 e Y527, respectivamente responsáveis pela ativação e inibição de Src (Figura 11).

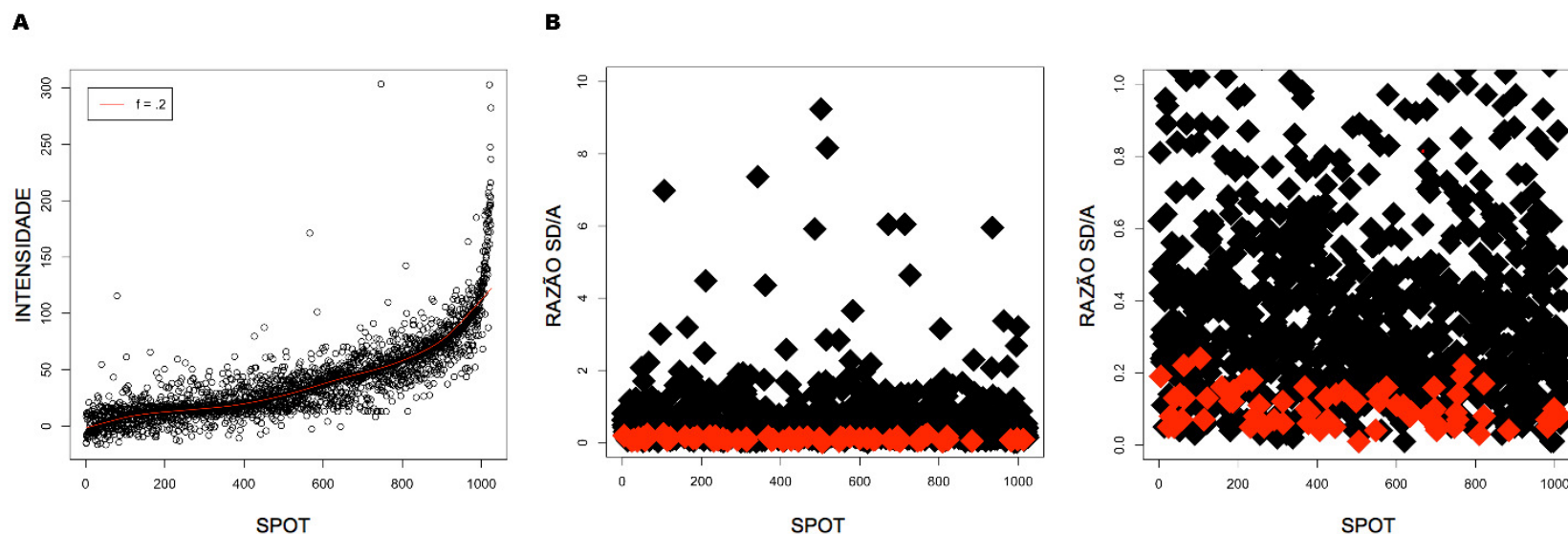


Figura 7. As replicatas no *chip* podem ser altamente variáveis, o que torna os *spots* associados a elas inadequados para posterior inferência biológica. A. Intensidade de cada *spot* em triplicata para as células TF1 expostas à violaceína por 16h. Os *spots* são ordenados pela mediana da intensidade de forma ascendente, formando uma trilha de pontos seguida quase perfeitamente pela linha de regressão mais clara. As replicações afastam-se da linha bastante frequentemente. B. Razões SD/A calculadas a partir de valores normalizados. Para um índice de SD/A < 20% em valores não-normalizados, a máxima razão SD/A em valores normalizados ficou em 24%. A primeira figura apresenta as razões SD/A numa escala de 0-10 e a segunda numa escala de 0-1 (sendo que SD/A = 1 significa que o desvio-padrão das replicações é igual à média delas). Os pontos claros indicam os *spots* selecionados pela metodologia PepMatrix (considerando também o *chip* controle).

Tabela 4. Primeiros 29 *spots* confiáveis selecionados a partir de células TF1 expostas à violaceína por 30 minutos *versus* um controle sem exposição. A. *Spots* selecionados com índices definidos como $SD/A < 20\%$ e $80\% < A/M < 120\%$. Ao todo, 142 *spots* foram selecionados e, destes, 11 apresentaram fosforilação diferencial ($p < .05$) na coluna R (Parte B). As colunas B, C, D e H, I, J contêm os valores de intensidade normalizados para os *spots* selecionados de cada *chip*, respectivamente. As colunas E, F e K, L apresentam os valores de média e mediana para cada *spot*. Já as colunas O e P contêm a razão de alteração observada (conhecida como FoldChange ou FC) entre as médias na coluna O e entre as medianas na coluna P. B. A coluna Q contém os valores do teste *F* para a determinação da existência de diferença significativa entre as variâncias das replicações entre os dois *chips*. Caso exista diferença ($p < .05$), o teste *t* realizado na coluna R é heteroscedástico. Caso contrário, o teste *t* realizado é homoscedástico. Os *spots* estão ordenados de forma ascendente pela coluna R. Nas colunas T, U, V, W e X estão dispostas as informações específicas sobre cada *spot*: respectivamente, o identificador no *chip*, a sequência presente no *spot*, o identificador RefSeq ou SwissProt dessa sequência, o nome da proteína de qual a sequência foi retirada e a provável quinase responsável pelo evento de fosforilação naquele *spot*.

A

	B	C	D	E	F	H	I	J	K	L	O	P
1	AUTO0	SL1		average	median	AUTO3	SL2		average	median	FC avg	FC median
2	53.45	56.26	52.94	54.22	53.45	42.61	44.69	43.85	43.72	43.85	0.806344856	0.820379552
3	65.04	72.91	69.96	69.3	69.96	96.47	97.83	89.95	94.75	96.47	1.367188659	1.379035207
4	84.44	82.16	78.35	81.65	82.16	119	103.3	105.3	109.2	105.3	1.337405279	1.281235043
5	58.18	65.29	62.26	61.91	62.26	75.18	81.03	73.25	76.49	75.18	1.235400261	1.207404412
6	-6.76	-4.588	-6.865	-6.071	-6.76	-1.734	1.798	4.736	1.6	1.798	-0.263516325	-0.265926857
7	55.46	68.99	60.05	61.5	60.05	39.93	47.75	49.58	45.76	47.75	0.743986738	0.795228018
8	59.95	61.48	63.08	61.51	61.48	45.2	55.25	53.94	51.46	53.94	0.836724121	0.877318728
9	77.46	84.02	81.27	80.91	81.27	86.86	93.81	100	93.57	93.81	1.156410482	1.154351108
10	45.76	50.71	54.8	50.42	50.71	39.82	43.21	43.35	42.13	43.21	0.83547681	0.852100491
11	84.44	88.59	103.7	92.23	88.59	79.31	68.99	71.76	73.35	71.76	0.795361925	0.810015446
12	112.4	119.6	123.5	118.5	119.6	73.52	88	107.1	89.56	88	0.755885909	0.735762758
13	62.55	62.68	56.79	60.67	62.55	44.47	44.16	56.68	48.44	44.47	0.798340394	0.710968192
14	57.23	61.59	71.94	63.59	61.59	33.73	49.86	53.44	45.68	49.86	0.718344298	0.80959218
15	83.97	77.92	80.45	80.78	80.45	45.09	57.79	76.12	59.67	57.79	0.738644931	0.718294435
16	212.9	216.2	205.3	211.5	212.9	208	196.8	198.1	201	198.1	0.950416128	0.930386168
17	53.57	50.27	59.59	54.47	53.57	57.92	62.65	67.77	62.78	62.65	1.152417883	1.16955927
18	78.76	65.73	71.13	71.87	71.13	55.75	66.24	63.28	61.76	63.28	0.859249733	0.889730057
19	259.6	231.4	261.4	250.8	259.6	220.8	180.5	234.7	212	220.8	0.845404594	0.850587818
20	83.37	74	92.23	83.2	83.37	67.53	69.83	69.26	68.87	69.26	0.827818405	0.830744429
21	43.99	57.34	59.24	53.52	57.34	61.84	62.22	72.38	65.48	62.22	1.223461956	1.085071696
22	264.2	234.4	229.7	242.8	234.4	205.4	220.7	227.7	217.9	220.7	0.897698147	0.941319557
23	68.35	76.72	87.68	77.59	76.72	55.85	65.92	70.01	63.93	65.92	0.823950899	0.859219765
24	92.25	77.05	86.98	85.43	86.98	60.4	76.59	78.11	71.7	76.59	0.839335638	0.880602881
25	66.81	62.9	79.75	69.82	66.81	56.16	63.7	57.68	59.18	57.68	0.847606428	0.863269529
26	90.94	96.42	113.6	100.3	96.42	62.67	77.44	95.3	78.47	77.44	0.782276102	0.803080084
27	25.53	21.1	25.78	24.14	25.53	25.14	35.92	45.72	35.59	35.92	1.474752958	1.406743471
28	64.1	71.06	58.3	64.49	64.1	66.8	76.8	75.24	72.95	75.24	1.131250591	1.173911138
29	66.1	58.65	57.25	60.67	58.65	52.75	58.85	52.57	54.72	52.75	0.901978181	0.899315172
30	59.24	59.74	61.8	60.26	59.74	60.29	65.61	63.91	63.27	63.91	1.049886309	1.069719566

B

	Q	R	T	U	V	W	X
1	F-test (norm)	T-test (norm)	Scanalyze	MOTIF	PROTEIN ID	PROTEIN	UP KIN
2	0,50872	0,00092	915	PNSSKTYGIKW	NP_005177	Calpain, large polypeptide	CKI;PKC (NetworkKIN pre
3	0,94224	0,00160	554	LKRSHSDSLDH	NP_001780	CDC 25A	CDC2;PIM1
4	0,22918	0,00629	753	MRRSVSEAAAL	NP_005348	Lipase hormone sensitive	PKA
5	0,87411	0,00949	42	VLAQPSTSRKR	NP_009125	CHK2	PKA;PKG
6	0,27202	0,01891	532	GSVLSTACGTP	NP_003647	CAM kinase 1	Calcium-calmodulin_dep
7	0,71380	0,03363	430	GSRGGSFDTAG	NP_004406	Desmoplakin	PKA
8	0,15157	0,03763	62	YQRRASDDGKL	NP_002871	RAF1	PKA
9	0,39929	0,04095	559	LKRSLSEMEIG	NP_003122	Serum response factor	Potassium_voltage_gate
10	0,32576	0,04399	247	PEKRASPSKPA	NP_002366	Microtubule associated p	CDC2_kinase
11	0,43693	0,04601	548	IAVRKSRDKAK	NP_005185	CCAAT/Enhancer binding p	ERK2
12	0,20131	0,04795	447	NRSFLSLKHTP	NP_077719	Notch2	GSK3beta
13	0,36362	0,05501	235	DDLMLSPDDIE	NP_000537	p53	Homeodomain_interactir
14	0,68194	0,07456	730	FSEITSPSKRS	NP_004144	ORC1	CDK2
15	0,07313	0,08283	242	LRRQHSYDTFV	NP_000825	NR2B	CaMKII;PKC
16	0,90385	0,09352	380	LEVSDSESDEA	Q01892	SPIB transcription factor	CK2
17	0,95750	0,10281	617	RAREASGAPTS	NP_000035	Androgen receptor	AKT1
18	0,81150	0,10814	251	FWSTLSPIAPR	NP_005220	ELK 1	MAPK14;p46SAPK;p54SAP
19	0,52707	0,10985	703	SSTDRSPYEKV	P15941	Mucin 1 transmembrane	GSK3beta
20	0,03397	0,11019	448	LVSPDSPRSID	NP_002301	Leukemia inhibitory facto	MAPkinase
21	0,68176	0,11309	917	IEQWFTEDFGP	NP_000537	p53	ERK2
22	0,54053	0,12127	243	MQPDNSSSDSY	NP_055022	CD5	PKA
23	0,72270	0,12283	667	TWRRGSTAGGC	NP_005177	Calpain, large polypeptide	nd
24	0,76364	0,12969	63	SRKGYSRKGFD	NP_000910	Peptidylglycine alpha am	PKC
25	0,33996	0,12969	492	SLLRLSLYNNC	NP_000237	MHC class II transactivat	Protein_kinase,_cAMP-c
26	0,68533	0,13376	573	PTPPLSPSRRS	NP_002458	c-Myc	CK2;GSK
27	0,12293	0,13508	152	ERGQEYLILEK	NP_003206	Tec tyrosine kinase	Tec_tyrosine_kinase
28	0,82981	0,15406	552	ASGYISSLEYF	NP_001724	Complement component 1, s	CK2
29	0,72107	0,15847	852	VLDDQYVSSVG	AAC08966	Bone marrow kinase BMX	c-Src
30	0,39834	0,16122	142	PSEEGYQDYEP	NP_000336	Synuclein alpha	SYK

Tabela 5. Primeiros 29 *spots* confiáveis selecionados a partir de células TF1 expostas à violaceína por 16 horas *versus* um controle sem exposição. A. *Spots* selecionados com índices definidos como $SD/A < 20\%$ e $80\% < A/M < 120\%$. Ao todo, 69 *spots* foram selecionados e, destes, 33 apresentaram fosforilação diferencial ($p < .05$) na coluna R (Parte B) (nem todos são exibidos). As colunas B, C, D e H, I, J contêm os valores de intensidade normalizados para os *spots* selecionados de cada *chip*, respectivamente. As colunas E, F e K, L apresentam os valores de média e mediana para cada *spot*. Já as colunas O e P contêm a razão de alteração observada (conhecida como FoldChange ou FC) entre as médias na coluna O e entre as medianas na coluna P. B. A coluna Q contém os valores do teste *F* para a determinação da existência de diferença significativa entre as variâncias das replicações entre os dois *chips*. Caso exista diferença ($p < .05$), o teste *t* realizado na coluna R é heteroscedástico. Caso contrário, o teste *t* realizado é homoscedástico. Os *spots* estão ordenados de forma ascendente pela coluna R. Nas colunas T, U, V, W e X estão dispostas as informações específicas sobre cada *spot*: respectivamente, o identificador no *chip*, a sequência presente no *spot*, o identificador RefSeq ou SwissProt dessa sequência, o nome da proteína de qual a sequência foi retirada e a provável quinase responsável pelo evento de fosforilação naquele *spot*.

A

	B	C	D	E	F	H	I	J	K	L	O	P
1	AUTO01	SL1		average	median	AUTO11	SL2		average	median	FC avg	FC median
2	105.02	103.06	100.62	102.9	103.06	52.523	55.049	56.581	54.718	55.049	0.531750866	0.534131063
3	61.135	56.585	58.186	58.635	58.186	117.19	131.13	128.42	125.58	128.42	2.141722532	2.207128396
4	37.836	42.323	42.448	40.869	42.323	70.863	77.28	74.572	74.239	74.572	1.816491049	1.761960088
5	36.295	40.364	48.273	41.644	40.364	90.178	87.079	96.415	91.224	90.178	2.190557099	2.23410581
6	67.522	63.988	69.378	66.963	67.522	104.26	116.9	112.63	111.26	112.63	1.661569521	1.668020041
7	86.452	83.688	88.493	86.211	86.452	160.41	144.52	142.02	148.98	144.52	1.728120007	1.671665204
8	55.341	63.335	69.028	62.568	63.335	189.87	159.29	156.17	168.45	159.29	2.692216265	2.515111148
9	99.698	110.03	87.214	98.982	99.698	192.66	177.52	204.39	191.52	192.66	1.934933704	1.932415482
10	106.8	106.98	113.21	109	106.98	69.221	60.984	78.281	69.495	69.221	0.637599586	0.647043149
11	54.038	61.483	52.705	56.075	54.038	38.12	34.338	32.813	35.09	34.338	0.625772478	0.635444564
12	50.256	58.11	47.344	51.904	50.256	119.97	98.405	97.104	105.16	98.405	2.026081822	1.958068321
13	79.233	85.214	93.51	85.986	85.214	118.99	114	124.85	119.28	118.99	1.387233016	1.396390871
14	48.837	44.283	51.421	48.18	48.837	67.748	63.055	72.653	67.819	67.748	1.40760271	1.38722831
15	61.49	63.549	72.525	65.855	63.549	100.65	103.93	87.76	97.446	100.65	1.479709113	1.583843533
16	43.514	39.058	36.852	39.808	39.058	52.362	54.913	49.433	52.236	52.362	1.312196695	1.340631538
17	75.806	67.034	73.575	72.138	73.575	59.399	53.528	50.257	54.395	53.528	0.754032045	0.727529782
18	51.082	63.442	61.913	58.813	61.913	88.052	100.06	81.72	89.945	88.052	1.529344739	1.422188775
19	112.36	119.61	123.47	118.48	119.61	196.26	161.37	211.81	189.81	196.26	1.602078117	1.640823389
20	71.426	59.304	66.93	65.887	66.93	101.31	86.115	90.374	92.6	90.374	1.405451902	1.350289213
21	79.826	64.529	70.542	71.632	70.542	138.96	114.42	106.17	119.85	114.42	1.673109663	1.621977425
22	89.524	93.485	93.624	92.211	93.485	83.141	83.215	76.498	80.951	83.141	0.877891497	0.889352149
23	63.148	70.407	70.307	67.954	70.307	91.651	79.766	87.213	86.21	87.213	1.268652787	1.240458616
24	83.968	77.918	80.449	80.778	80.449	93.616	100.06	108.78	100.82	100.06	1.248090157	1.243792869
25	125.25	100.45	108.43	111.38	108.43	247.49	198.5	172.11	206.04	198.5	1.849917986	1.830730988
26	259.62	231.4	261.38	250.8	259.62	203.46	196.02	215.79	205.09	203.46	0.817744155	0.783683478
27	84.439	88.587	103.65	92.226	88.587	112.44	122.43	117.71	117.53	117.71	1.274325964	1.328748711
28	72.613	54.192	76.952	67.919	72.613	42.049	34.338	44.352	40.246	42.049	0.592568607	0.579087491
29	59.006	54.192	56.203	56.467	56.203	65.131	72.31	62.762	66.734	65.131	1.181838113	1.158863595
30	39.256	42.436	50.257	43.983	42.436	60.059	52.563	63.31	58.644	60.059	1.333340307	1.415291865

B

	Q	R	T	U	V	W	X
1	F-test (norm)	T-test (norm)	Scanalyze	MOTIF	PROTEIN_ID	PROTEIN	UP_KIN
2	0.92643	0.00001	702	ELEGISPDELK	AAH05359	Glia maturation factor	PKA
3	0.17768	0.00012	314	KVRFNSVSSYS	NP_000611	Nitric oxide synthase	CAMKIIalpha
4	0.79896	0.00016	406	ALDFRTPRNAK	NP_000905	Opioid receptor mu 1	Beta-adren
5	0.75728	0.00037	982	LRGSATRALPP	NP_001885	Casein kinase 1, epsilon	CK1epsilon
6	0.30687	0.00039	700	GPRLVSNHSLH	NP_001447	Filamin A	CAMKII
7	0.11051	0.00045	426	AKIQASFRGHM	NP_006167	Neurogranin	PKCalpha;PK
8	0.23996	0.00076	254	GFSRKSHTFLP	NP_149124	CNPase	PKC
9	0.83682	0.00082	291	KKWKQSVRLIS	NP_004929	Death associated protein	Death_assoc
10	0.30228	0.00189	179	IRYIESLQELL	NP_005584	Myogenic factor 5	CK2
11	0.50037	0.00265	822	TTGGESCDELE	NP_002529	Occludin	CK2beta
12	0.31642	0.00275	557	KRRQTSMTDFY	NP_510867	Cyclin-dependent kinase	AKT1
13	0.72898	0.00304	444	SVRMLSGSKEK	NP_000905	Opioid receptor mu 1	CAMKII
14	0.72368	0.00480	981	RKRRPTSGLHP	NP_009225	BRCA1	AKT1
15	0.64075	0.00618	637	SQRRESFLYRS	NP_002591	cAMP-specific 3',5'-cy	PKA
16	0.79001	0.00785	265	HPFLRRNSGC	NP_001079	Aralkylamine N-acetyl	PKA
17	0.98408	0.00911	638	KARKKSSCQLL	NP_056461	Ras related protein Ra	PKA
18	0.68687	0.00938	697	PAARASKKILL	NP_001610	Beta-adrenergic recept	PKC
19	0.09103	0.00950	447	NRSFLSLKHTP	NP_077719	Notch2	nd
20	0.75854	0.00966	634	KKKKPSRLKGD	NP_000029	APC	PKA
21	0.33910	0.01114	756	MRRQRSAPDLK	NP_006603	Kinesin family member	CK2
22	0.53410	0.01236	699	LLSELSRRRIR	NP_004085	eIF2 alpha	PKR;GCN2;E
23	0.64897	0.01238	58	EPQRRSARLSA	NP_005508	HMG17	Ribosomal_j
24	0.27505	0.01330	242	LRRQHSYDTFV	NP_000825	NR2B	CaMKII;PKC
25	0.19742	0.01523	766	GLRRSSKFCLK	NP_000015	Beta-2-adrenergic rece	PKC
26	0.52106	0.01553	703	SSTDSPYEKV	P15941	Mucin 1 transmembrane	GSK3beta
27	0.39277	0.01776	548	I AVRKSRDKAK	NP_005185	CCAAT/Enhancer binding	ERK2
28	0.31703	0.02200	40	ARPPASPSPQR	NP_008881	Synapsin I	CDK5
29	0.38260	0.03237	755	CMRNFSRSDHL	NP_001955	Early growth response	CK2
30	0.97308	0.03246	751	IKRLRSQVQVS	NP_000448	Hepatocyte nuclear fac	AMPK

Tabela 6. Primeiros 29 *spots* confiáveis selecionados a partir de células TF1 expostas à violaceína por 24 horas *versus* um controle sem exposição. A. *Spots* selecionados com índices definidos como $SD/A < 20\%$ e $80\% < A/M < 120\%$. Ao todo, 75 *spots* foram selecionados e, destes, 40 apresentaram fosforilação diferencial ($p < .05$) na coluna R (Parte B) (nem todos são exibidos). As colunas B, C, D e H, I, J contêm os valores de intensidade normalizados para os *spots* selecionados de cada *chip*, respectivamente. As colunas E, F e K, L apresentam os valores de média e mediana para cada *spot*. Já as colunas O e P contêm a razão de alteração observada (conhecida como FoldChange ou FC) entre as médias na coluna O e entre as medianas na coluna P. B. A coluna Q contém os valores do teste *F* para a determinação da existência de diferença significativa entre as variâncias das replicações entre os dois *chips*. Caso exista diferença ($p < .05$), o teste *t* realizado na coluna R é heteroscedástico. Caso contrário, o teste *t* realizado é homoscedástico. Os *spots* estão ordenados de forma ascendente pela coluna R. Nas colunas T, U, V, W e X estão dispostas as informações específicas sobre cada *spot*: respectivamente, o identificador no *chip*, a sequência presente no *spot*, o identificador RefSeq ou SwissProt dessa sequência, o nome da proteína de qual a sequência foi retirada e a provável quinase responsável pelo evento de fosforilação naquele *spot*.

A

	B	C	D	E	F	H	I	J	K	L	O	P
1	AUTO0H	SL1		average	median	AUTO24	SL2		average	median	FC avg	FC median
2	105.02	103.06	100.62	102.9	103.06	47.947	41.617	44.273	44.613	44.273	0.43355037	0.429579547
3	216.21	234.56	236.43	229.07	234.56	44.947	48.829	55.018	49.598	48.829	0.216523783	0.208178546
4	106.8	106.98	113.21	109	106.98	66.385	61.303	66.258	64.649	66.258	0.593131513	0.619343867
5	125.72	112.1	121.02	119.61	121.02	285.54	274.51	248.15	269.4	274.51	2.252273513	2.268327248
6	116.02	141.38	132.33	129.91	132.33	37.302	33.652	42.598	37.851	37.302	0.291357558	0.281885886
7	50.256	58.11	47.344	51.904	50.256	115.1	106.53	102.16	107.93	106.53	2.079437013	2.119656237
8	55.341	63.335	69.028	62.568	63.335	128.14	154.45	145.8	142.8	145.8	2.282270746	2.301969803
9	73.555	74.652	81.968	76.725	74.652	106.41	115.39	108.04	109.95	108.04	1.433002115	1.447249396
10	44.34	44.068	47.924	45.444	44.34	95.616	77.678	90.256	87.85	90.256	1.93314081	2.035539008
11	61.49	63.549	72.525	65.855	63.549	115.26	104.42	101.66	107.11	104.42	1.626487495	1.643135401
12	86.452	83.688	88.493	86.211	86.452	136.99	137.93	160.06	144.99	137.93	1.681819719	1.595462174
13	71.31	78.571	84.417	78.099	78.571	120.05	117.94	133.21	123.73	120.05	1.584317172	1.527886449
14	99.698	110.03	87.214	98.982	99.698	185.56	173.84	210.23	189.87	185.56	1.918280609	1.861175263
15	63.148	70.407	70.307	67.954	70.307	89.77	87.742	95.29	90.934	89.77	1.338172832	1.276826472
16	58.18	65.294	62.263	61.912	62.263	78.829	82.939	87.236	83.001	82.939	1.34062583	1.332076055
17	125.25	100.45	108.43	111.38	108.43	208.79	208.4	171.13	196.11	208.4	1.760773246	1.921980764
18	68.709	67.468	83.597	73.258	68.709	38.504	43.272	43.44	41.739	43.272	0.569748117	0.629792985
19	112.36	119.61	123.47	118.48	119.61	202.34	162.42	185.23	183.33	185.23	1.547376012	1.548592645
20	84.439	82.163	78.351	81.651	82.163	53.793	46.125	63.238	54.385	53.793	0.666070271	0.654717832
21	63.619	68.994	82.433	71.682	68.994	44.652	40.413	41.929	42.331	41.929	0.590546391	0.607729427
22	57.232	61.59	71.941	63.587	61.59	42.101	41.469	37.399	40.323	41.469	0.634137298	0.673314251
23	66.935	70.407	78.935	72.092	70.407	53.491	55.288	55.183	54.654	55.183	0.758111925	0.783772919
24	78.878	98.383	89.313	88.858	89.313	64.137	60.246	63.238	62.54	63.238	0.703820469	0.708045066
25	73.2	87.388	84.531	81.706	84.531	61.438	52.583	60.72	58.247	60.72	0.712886979	0.718319895
26	48.482	50.273	58.301	52.352	50.273	40.155	37.413	35.047	37.538	37.413	0.717034129	0.744190383
27	66.103	67.907	82.433	72.148	67.907	93.515	111.48	104.68	103.23	104.68	1.430775142	1.541548677
28	128.09	124.83	111.81	121.58	124.83	67.882	66.409	95.455	76.582	67.882	0.629903421	0.543776328
29	48.837	39.384	54.104	47.442	48.837	73.728	89.249	105.19	89.387	89.249	1.884148103	1.827502645
30	82.549	71.713	74.504	76.255	74.504	64.439	59.95	60.72	61.703	60.72	0.80916792	0.814999985

B

	Q	R	T	U	V	W	X
1	F-test (norm)	T-test (norm)	Scanalyze	MOTIF	PROTEIN_ID	PROTEIN	UP_KIN
2	0.64996	0.00001	702	ELEGISPDELK	AAH05359	Glia maturation factor	PKA
3	0.34252	0.00001	701	SSDDSDSEEP	NP_002829	CD45	CK2alpha1;CK2alp
4	0.77318	0.00008	179	IRYIESLQELL	NP_005584	Myogenic factor 5	CK2
5	0.22964	0.00022	771	RKGHEYTNIKY	NP_002825	SHP2	PDGFRbeta
6	0.21829	0.00030	59	IRESESTAGSF	AAH13200	Lck	PKC
7	0.83448	0.00035	557	KRRQTSMTDFY	NP_510867	Cyclin-dependent kinase	AKT1
8	0.41635	0.00077	254	GFSRKSHFLP	NP_149124	CNPase	PKC
9	0.95459	0.00096	377	NSRRPSRAMWL	NP_000629	Vitronectin	PKA
10	0.10356	0.00148	772	IYSGDYRQGR	NP_006334	c-Mer	c-Mer
11	0.80015	0.00153	637	SQRRESFLYRS	NP_002591	cAMP-specific 3',5'-cycl	PKA
12	0.06599	0.00155	426	AKIQASFRGHM	NP_006167	Neurogranin	PKCalpha;PKM;PK
13	0.77281	0.00171	369	RKRKNSRVTF	NP_054829	NIPP1	PKA
14	0.54889	0.00195	291	KKWKQSVRLIS	NP_004929	Death associated prot	Death_associated
15	0.93655	0.00222	58	EPQRRSARLSA	NP_005508	HMG17	Ribosomal_protein
16	0.83815	0.00270	42	VLAQPSTSRKR	NP_009125	CHK2	PKA;PKG
17	0.51025	0.00424	766	GLRRSSKCLK	NP_000015	Beta-2-adrenergic rece	PKC
18	0.17771	0.00438	886	RLLGHSPVLRN	NP_068660	CDC 25B	CDC2
19	0.14697	0.00570	447	NRSFLSLKHTP	NP_077719	Notch2	nd
20	0.22823	0.00658	753	MRRSVSEAALA	NP_005348	Lipase hormone sensit	nd
21	0.09365	0.00688	741	DEQFVSLYGTE	NP_037386	TBK1	nd
22	0.20484	0.00722	730	FSEITSPSKRS	NP_004144	ORC1	CDK2
23	0.05194	0.00848	183	YGGLTSPGLSY	NP_002264	Keratin 8	p42_kinase
24	0.08346	0.01025	695	RPRNYSVGRP	NP_006203	6-phosphofructo-2-kin	AKT
25	0.60102	0.01059	574	QLRRPSDRELS	NP_068810	NFKB3	PKA
26	0.38574	0.01162	504	KTKFASDDEHD	NP_003133	Autoantigen La	nd
27	0.98671	0.01344	381	LRRRLSDSSFI	NP_598328	Synapsin II	PKA
28	0.43399	0.01353	563	DFEGFSYVNPQ	NP_002728	Protein kinase C alpha	PKCalpha
29	0.36714	0.01399	5	ALVHSYMTGRR	NP_005481	SH2D3A	nd
30	0.30795	0.01460	1013	SRVTFSEDEI	NP_054829	NIPP1	CK2

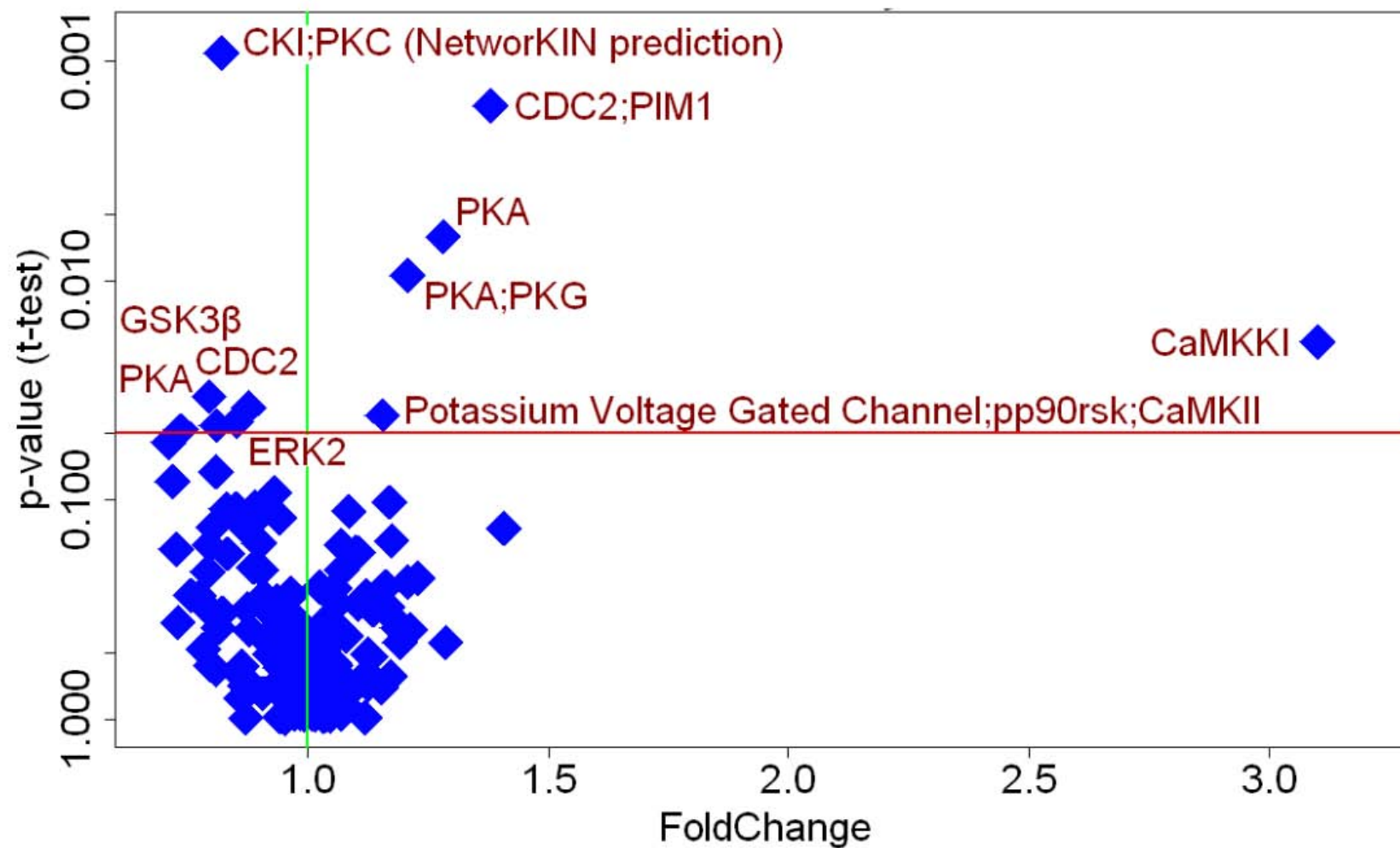


Figura 8. Quinases *up*-reguladas e *down*-reguladas com base na fosforilação dos *spots* confiáveis de acordo com os parâmetros estabelecidos para um *chip* de células TF1 expostas à violaceína por 30 minutos *versus* um controle sem exposição selecionados pela metodologia *PepMatrix*. Apenas os nomes das quinases potencialmente responsáveis pelas fosforilações e que apresentaram fosforilação diferencial são mostrados.

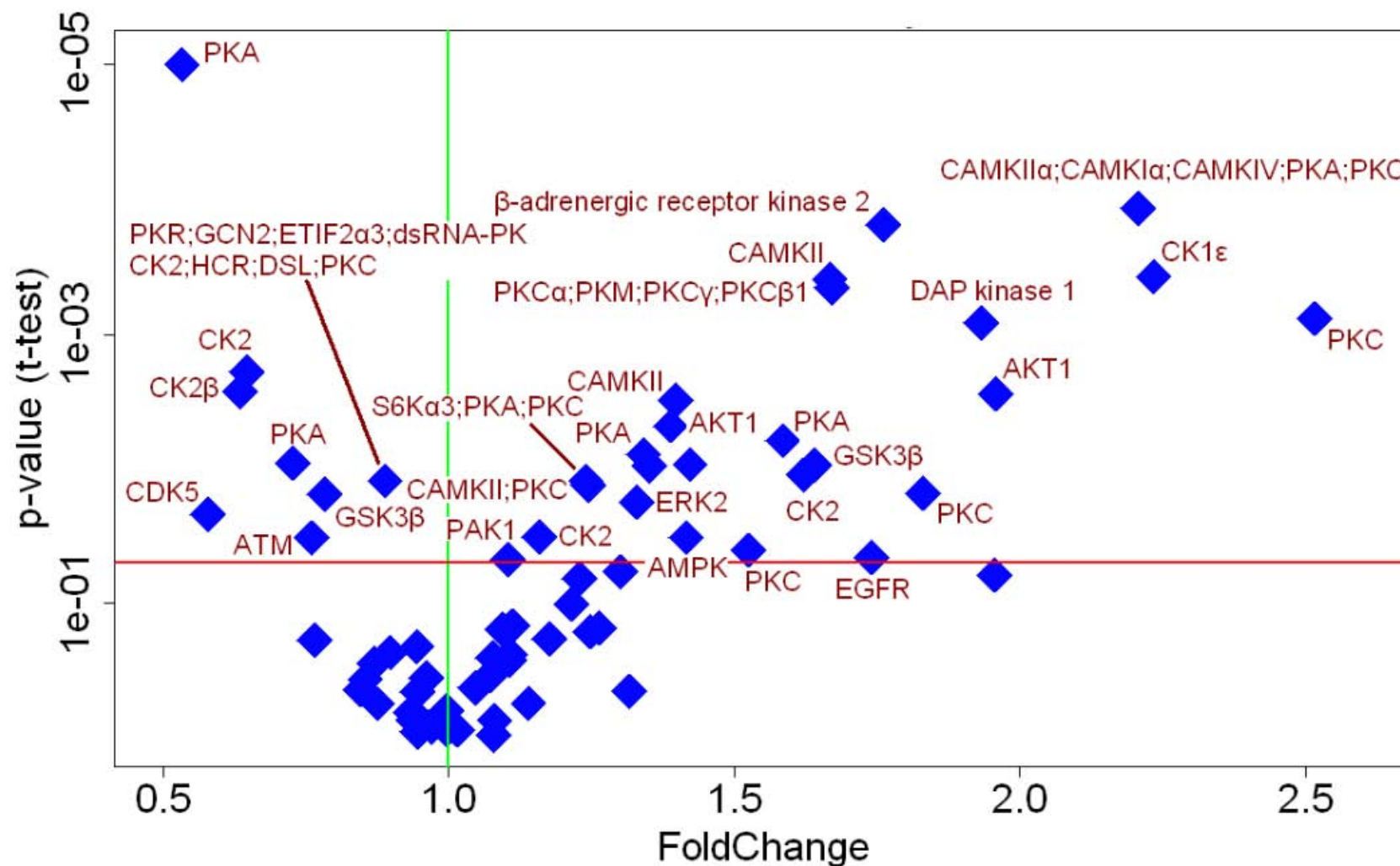


Figura 9. Quinases *up*-reguladas e *down*-reguladas com base na fosforilação dos *spots* confiáveis de acordo com os parâmetros estabelecidos para um *chip* de células TF1 expostas à violaceína por 16 horas *versus* um controle sem exposição selecionados pela metodologia *PepMatrix*. Apenas os nomes das quinases potencialmente responsáveis pelas fosforilações e que apresentaram fosforilação diferencial são mostrados.

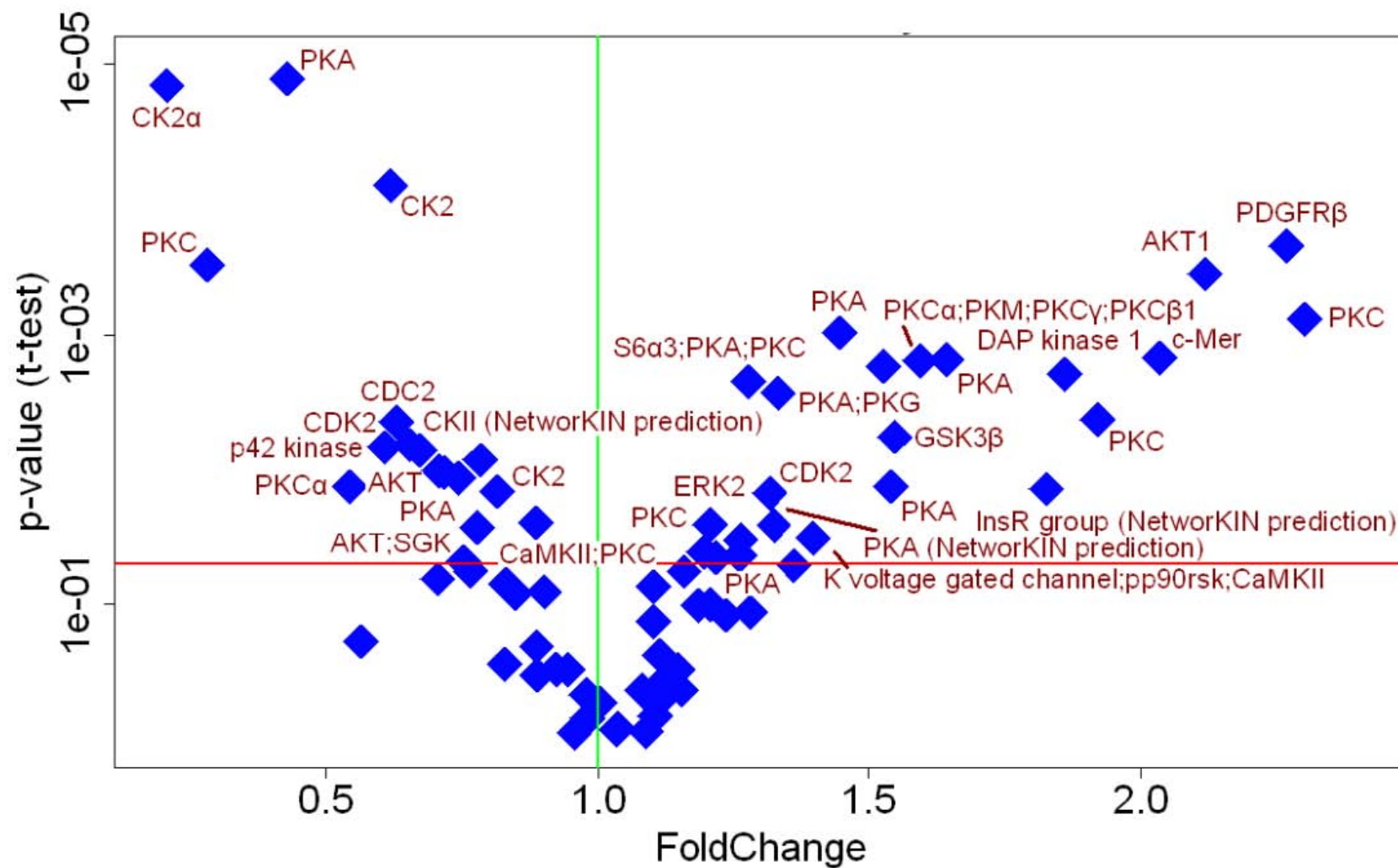


Figura 10. Quinases *up*-reguladas e *down*-reguladas com base na fosforilação dos *spots* confiáveis de acordo com os parâmetros estabelecidos para um *chip* de células TF1 expostas à violaceína por 24 horas *versus* um controle sem exposição selecionados pela metodologia *PepMatrix*. Apenas os nomes das quinases potencialmente responsáveis pelas fosforilações e que apresentaram fosforilação diferencial são mostrados.

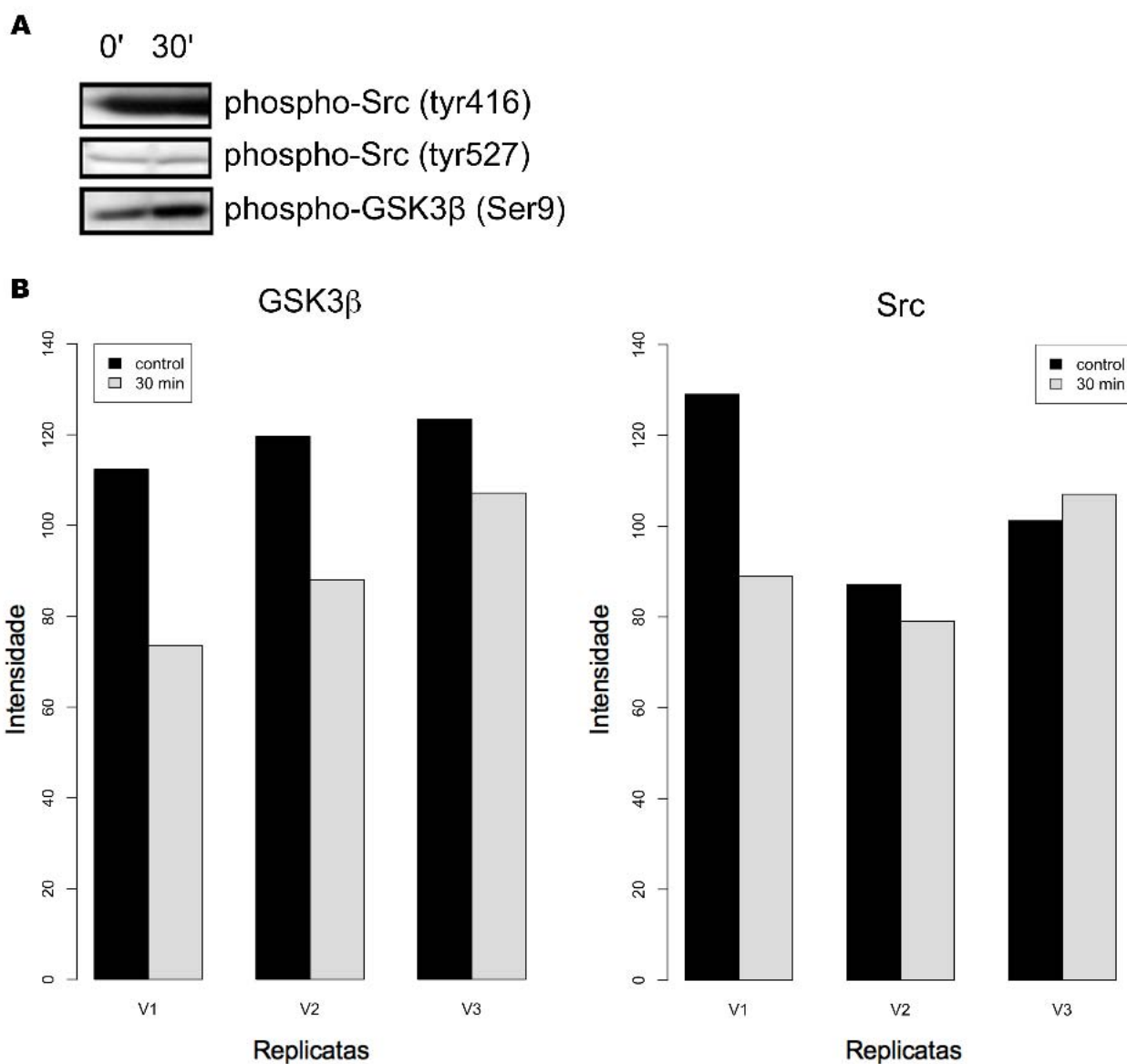


Figura 11. Avaliação dos *spots* confiáveis no *chip* originou resultados confirmados por outras técnicas. A. *Western blots* de células TF1 expostas a violaceína por 30 minutos para pGSK3 β (S9 - inibitório) e pSrc (Y416 - ativador e Y527 - inibitório). B. Intensidades das replicações de *spots* contendo substratos de GSK3 β e Src num *chip* contendo lisados de células TF1 expostas a violaceína por 30 minutos. É notável a redução da intensidade de fosforilação no substrato de GSK3 β ($p < .05$) enquanto o substrato de Src não apresentou diferença significativa ($p > .05$).

3.3 Perfil quinômico de pré-osteoblastos MC3T3-E1 em adesão

A adesão a uma matriz é um processo essencial na diferenciação de osteoblastos. O processo de remodelação óssea, pelo qual o tecido ósseo passa ao longo de sua existência visando a regulação da homeostase de cálcio, microrreparos na estrutura e definição do esqueleto durante o crescimento, envolve intensa relação entre a matrix extracelular e os osteoblastos, apresentando forte influência nos mecanismos de transdução de sinal. A expressão de proteínas da matriz extracelular, do citoesqueleto e de membrana, como integrinas e caderinas, é finamente regulada por mecanismos de sinalização celular (ANSELME, 2000). Durante a vida, algumas condições como envelhecimento e estresse mecânico podem alterar essa regulação. Nesse contexto, a busca por novas terapias envolvendo o tecido ósseo tem obtido destaque na área de biomateriais. A engenharia de tecidos tem oferecido subsídios pelos quais células osteoprogenitoras podem ser associadas a biomateriais com o objetivo de desenvolver osteossustitutos ideais capazes de restaurar e manter as características do tecido (ZAMBUZZI, 2008).

Nas últimas décadas, os osteoblastos vêm sendo amplamente utilizados *in vitro* para o estudo do metabolismo ósseo e interações células-biomateriais. Para este fim são utilizadas linhagens celulares provenientes de ratos (UMR-106), camundongos (MC3T3-E1), fetos humanos (hFOB 1.19) e osteosarcoma (SAOS-2). Particularmente, a linhagem celular de pré-osteoblastos MC3T3-E1, obtida por Sudo *et al.* em 1983, tem sido muito utilizada para o estudo do processo de diferenciação de osteoblastos e subsequente mineralização biológica.

Nesse contexto, como parte de um estudo em colaboração com o Laboratório de Transdução de Sinal e Bioensaios – IB - Unicamp, o perfil quinômico de células MC3T3-E1 foi estudado através do PepChip® Kinomics™ e os resultados foram submetidos à metodologia

PepMatrix. As células foram cultivadas sobre superfícies de poliestireno e as amostras foram coletadas 2 horas, usada como condição experimental, e 10 dias depois, usada como controle. Os resultados mostraram diversas quinases com atividade diferencial, conforme a tabela 7.

Tabela 7. Primeiros 29 *spots* confiáveis selecionados a partir de células MC3T3-E1 submetidas a adesão em uma placa de poliestireno por 2 horas versus um controle de células já aderidas em cultura por 10 horas. A. *Spots* selecionados com índices definidos como $SD/A < 20\%$ e $80\% < A/M < 120\%$. Ao todo, 32 *spots* foram selecionados e, destes, 23 apresentaram fosforilação diferencial ($p < .05$) na coluna Q (Parte B). As colunas A, B, C e G, H, I contêm os valores de intensidade normalizados para os *spots* selecionados de cada *chip*, respectivamente. As colunas D, E e J, K apresentam os valores de média e mediana para cada *spot*. Já as colunas N e O contêm a razão de alteração observada (conhecida como FoldChange ou FC) entre as médias na coluna N e entre as medianas na coluna O. B. A coluna P contém os valores do teste *F* para a determinação da existência de diferença significativa entre as variâncias das replicações entre os dois *chips*. Caso exista diferença ($p < .05$), o teste *t* realizado na coluna Q é heteroscedástico. Caso contrário, o teste *t* realizado é homoscedástico. Os *spots* estão ordenados de forma ascendente pela coluna Q. Nas colunas S, T, U, V e W estão dispostas as informações específicas sobre cada *spot*: respectivamente, o identificador no *chip*, a sequência presente no *spot*, o identificador RefSeq ou SwissProt dessa sequência, o nome da proteína de qual a sequência foi retirada e a provável quinase responsável pelo evento de fosforilação naquele *spot*.

A

	A	B	C	D	E	G	H	I	J	K	N	O
1	control	SL1		average	median	L0352-2	SL2		average	median	FC avg	FC median
2	15.383	13.754	17.152	15.43	15.383	43.179	44.836	42.095	43.37	43.179	2.81085522	2.80701901
3	178.23	162.39	174.72	171.78	174.72	362.12	346.12	340.21	349.48	346.12	2.03449341	1.98099469
4	47.33	36.247	42.702	42.093	42.702	398.76	369.27	347.93	371.99	369.27	8.83724159	8.64760198
5	168.16	139.47	158.76	155.46	158.76	379.67	372.31	357.45	369.81	372.31	2.37876552	2.34515584
6	184.33	161.36	185.02	176.9	184.33	350.59	373.38	368.11	364.03	368.11	2.05782854	1.99704442
7	167.14	122.59	158.52	149.41	158.52	372.72	364.51	368.62	368.62	368.62	2.46706995	2.32541373
8	19.565	21.403	24.204	21.724	21.403	50.78	50.869	58.144	53.265	50.869	2.45187227	2.37673339
9	168.59	148.39	166.78	161.25	166.78	376.02	317.66	351.11	348.26	351.11	2.15970066	2.10522571
10	67.12	80.751	85.936	77.936	80.751	166.73	158	141.83	155.52	158	1.99546865	1.95661681
11	78.449	78.531	111.25	89.409	78.531	374.74	376.18	375.4	375.44	375.4	4.19909873	4.78021361
12	23.859	23.148	28.607	25.204	23.859	71.727	55.653	58.653	62.011	58.653	2.4603295	2.45834255
13	72.327	62.266	62.593	65.729	62.593	38.448	33.268	26.8	32.839	33.268	0.49960968	0.53150117
14	138.43	137.22	148.01	141.22	138.43	291.22	314.14	245.36	283.58	291.22	2.00805321	2.10376811
15	86.836	68.982	88.826	81.548	86.836	41.585	37.255	33.035	37.292	37.255	0.45729964	0.42902721
16	116.27	141.25	134.07	130.53	134.07	82.378	71.718	93.886	82.661	82.378	0.63325685	0.61442241
17	91.019	80.478	69.535	80.344	80.478	138.79	154.84	119.88	137.84	138.79	1.71562276	1.72459767
18	180.14	160.08	177.82	172.68	177.82	257.94	348.16	286.43	297.51	286.43	1.72290389	1.61076858
19	85.229	90.172	93.965	89.789	90.172	70.989	68.599	50.079	63.222	68.599	0.70412379	0.76075886
20	88.495	114.84	123.1	108.81	114.84	163.88	143.17	145.27	150.77	145.27	1.38561415	1.26491958
21	165.64	141.93	159.81	155.79	159.81	255.13	248.63	188.8	230.86	248.63	1.48182445	1.55585709
22	165.45	148.82	154.21	156.16	154.21	138.44	110.79	98.945	116.06	110.79	0.74318637	0.71842856
23	169.32	141.78	152.3	154.47	152.3	241.77	256.89	183.21	227.29	241.77	1.47146817	1.58751273
24	175.46	162.14	175.78	171.13	175.46	139.49	158.77	125.08	141.11	139.49	0.82462374	0.79499349
25	149.61	136.6	155.69	147.3	149.61	112.35	136.73	104.44	117.84	112.35	0.80001131	0.75097251
26	35.994	32.818	30.904	33.239	32.818	30.146	18.024	20.332	22.834	20.332	0.68697227	0.61953098
27	160.66	150.38	167.59	159.54	160.66	373.67	235.84	280.81	296.77	280.81	1.86013437	1.74788455
28	73.58	58.544	72.678	68.267	72.678	85.435	86.434	73.17	81.68	85.435	1.19646458	1.17552769
29	97.402	97.316	110.39	101.7	97.402	85.854	99.262	83.347	89.488	85.854	0.87989081	0.88143843
30	102.62	83.654	87.751	91.34	87.751	76.521	86.258	69.801	77.527	76.521	0.84876742	0.87203048

B

	P	Q	S	T	U	V	W
1	F-test (norm)	T-test (norm)	Scanalyze	MOTIF	PROTEIN ID	PROTEIN	UP KIN
2	0.79487	0.00002	578	DVLKPYDSNTV	NP_002568	PAK2	Src
3	0.70077	0.00003	771	RKGHEYTNIKY	NP_002825	SHP2	PDGFRbeta
4	0.09083	0.00003	894	RKPKKKSCLLL	NP_002875	Ras related protein 1A	PKA
5	0.74888	0.00004	627	RKKRISVKKKQ	NP_059431	Cytohesin 2	PKC
6	0.87939	0.00006	189	RKKKVSSTKRH	NP_004753	Cytohesin-1	PKC
7	0.05852	0.00009	383	APKAPSKKEKK	NP_001334	DAB2	PKC
8	0.46799	0.00035	711	SptCtrl	0	Used for production/QC	na
9	0.25420	0.00049	746	RSKKNLSLALS	NP_000116	Estrogen receptor, alpha	PAK1
10	0.74360	0.00108	573	PTPPLSPSRRS	NP_002458	c-Myc	CK2;GSK
11	0.00289	0.00143	250	TRVPPSRRGPD	NP_000672	Alpha 2A adrenergic rec	PKC
12	0.21521	0.00214	758	AEVLSPSPRGQR	NP_001058	DNA topoisomerase II alpha	Proline_directed_kinase
13	0.97948	0.00222	95	REWHKTTQMSA	NP_002410	MAP3K11	MAP3K11
14	0.05537	0.00226	605	RLRPRTKRVKS	NP_038202	Neutrophil cytosolic factor	PKC
15	0.26548	0.00284	700	GPRLVSNHSLH	NP_001447	Filamin A	CAMKII
16	0.85251	0.00814	756	MRRQRSAPDLK	NP_006603	Kinesin family member 3	CK2
17	0.54742	0.00834	399	KTNLSYYEYDK	AAC08966	Bone marrow kinase BMX	FAK
18	0.10723	0.01033	127	KFRTPSFLKKN	NP_058432	Adducin 3	PKCalpha
19	0.25551	0.01986	64	GFTEESGDDEY	NP_005988	TCFL1	CK2
20	0.56925	0.02726	689	RSNPSPSRKSG	NP_000157	Connexin 32	PKC;PKA
21	0.20498	0.02808	635	LKPGSSHRKTK	NP_000052	Bruton's tyrosine kinase	PKCbeta
22	0.29846	0.03413	739	KKGARSRLFS	NP_002059	Guanine nucleotide binding	PKCalpha
23	0.22621	0.03794	634	KKKKPSRLKGD	NP_000029	APC	PKA
24	0.34999	0.04917	104	MRKKISNAQLQ	NP_003469	Cullin 5	PKA
25	0.50286	0.05859	699	LLSELSRRRIR	NP_004085	eIF2 alpha	PKR;GCN2;Eukaryotic_tra
26	0.27516	0.06005	489	PSGLWSPAYAS	NP_079199	NUP210	CyclinB;p34cdc2
27	0.02989	0.07530	613	RGKKKSGCLVL	NP_001655	RhoA	PKA
28	0.86832	0.10696	53	KALPLSPRKRL	NP_001245	CDC6	CDK2
29	0.87210	0.13692	422	TKALQSPKRPR	NP_127492	TFII-I	EphB2
30	0.81498	0.13867	537	IHMVYSKRSGL	NP_003080	Small nuclear ribonucleon	

As quinases Src, PKC e FAK apresentaram significativa atividade diferencial, sendo pelo menos duas vezes mais ativas durante a adesão (Figura 12). Essas proteínas estão entre os principais reguladores da adesão celular através do estabelecimento de adesões focais. O mecanismo envolve a ativação de integrinas, ao que se segue a interação direta entre Src, PKC e FAK, levando a uma cascata de sinalização através de Rho, ROCK, paxilina e talina e alcançando o elemento final da cascata, a cofilina. A cofilina é responsável pela determinação do comprimento dos filamentos de actina e consequentemente pelo rearranjo do citoesqueleto (YANG *et al.*, 1998) (Figura 14).

Além disso, a cascata também ativa GSK3 β , responsável por fosforilar e consequentemente inibir a proteína β -catenina, reduzindo a sinalização para a proliferação celular. As proteínas PAK e RAP1 também fazem parte da cascata de sinalização durante a adesão, sendo importantes para a sinalização de sobrevivência celular através de Bcl-2 (Figura 14).

A adesão celular também depende da sinalização de sobrevivência celular, bem como da inibição da proliferação. Os resultados do *chip* mostraram que a quinase GSK3 β apresentou atividade aumentada durante a adesão na fosforilação de um peptídeo derivado de c-Myc contendo o sítio T58, levando à sua inativação e subsequente inibição da proliferação celular (SMITH *et al.*, 2002). A GSK3 β também é responsável pela fosforilação e consequente inibição de β -catenina (um co-ativador transcricional de c-Myc), portanto inibindo a proliferação celular também pela via Wnt (Tabela 7 e figuras 13 e 14). Na tabela 7, observa-se que o pseudosubstrato derivado de c-Myc pode ser fosforilado por CK2 ou GSK (*Glycogen Synthase Kinase*), sendo que ambas atuam em conjunto num *scaffold* proteico que faz parte da via Wnt (Figura 13).

Um substrato presente no *chip* derivado de RAP1A (*Ras related protein 1A*) foi mais fosforilado consistentemente ao longo das replicações durante a adesão. A intensidade no *chip* foi mais de oito vezes maior durante a adesão do que durante a condição controle. A enzima responsável é provavelmente uma PKA (Tabela 7), estabelecendo uma possível rota para a manutenção da viabilidade celular através da sinalização por MAPK e BCL-2 desencadeadas por integrinas (Figura 14).

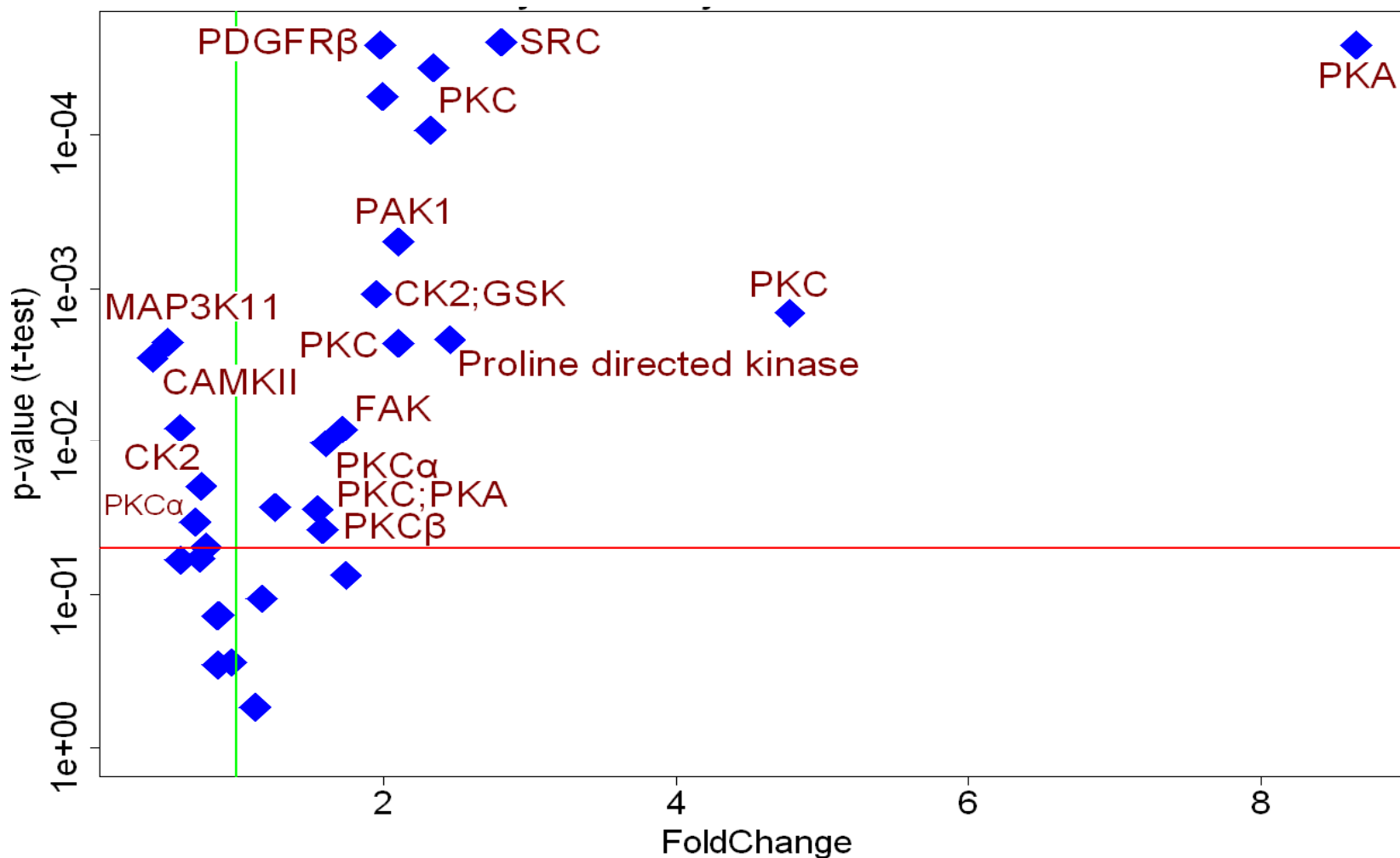


Figura 12. Quinases *up*-reguladas e *down*-reguladas com base na fosforilação dos *spots* confiáveis de acordo com os parâmetros estabelecidos para um *chip* de células MC3T3-E1 submetidas a adesão em uma placa de poliestireno por 2 horas *versus* um controle de células já aderidas em cultura por 10 horas selecionados pela metodologia *PepMatrix*. Apenas os nomes das quinases potencialmente responsáveis pelas fosforilações e que apresentaram fosforilação diferencial são mostrados. As quinases Src, PKC, PAK1 e FAK estão relacionadas a vias reguladoras do remodelamento do citoesqueleto de actina e ao estabelecimento de adesões focais, processos relacionados à adesão celular.

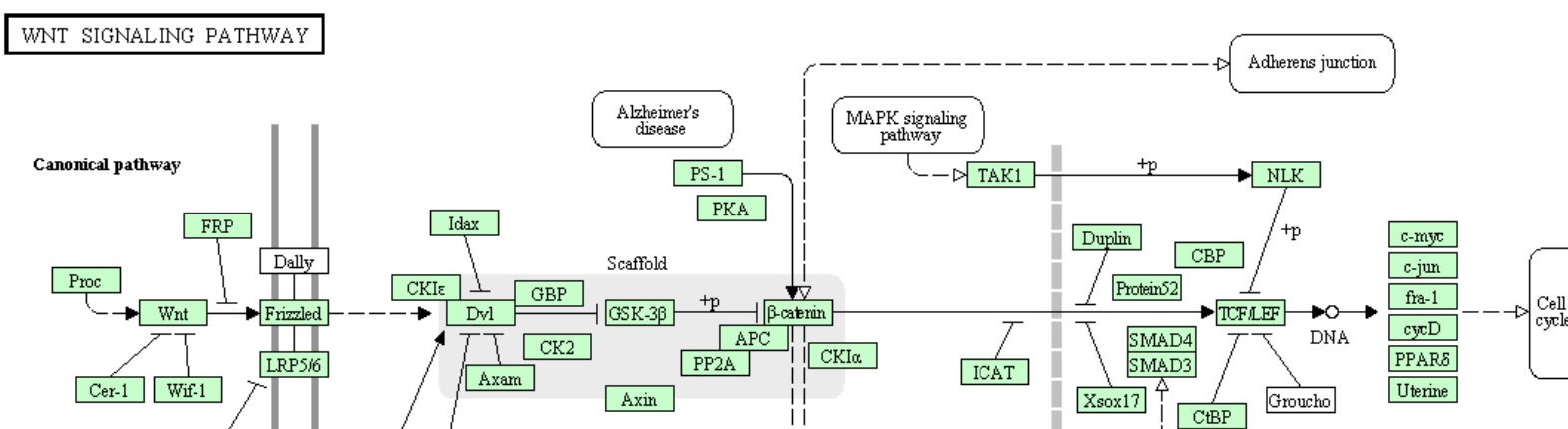
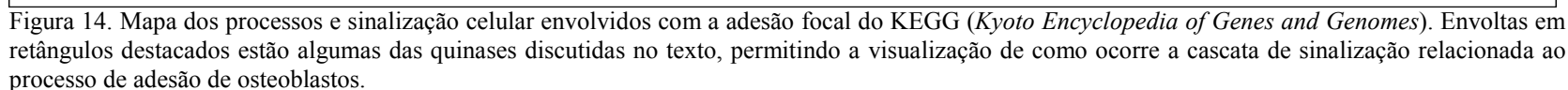


Figura 13. Via de sinalização Wnt canônica mostrando o *scaffold* proteico responsável pela fosforilação de β -catenina envolvendo CK2 e GSK3 β . À direita encontra-se o fator de transcrição c-Myc, responsável pela proliferação celular. Adaptado de KEGG (*Kyoto Encyclopedia of Genes and Genomes*).

Como mostrado acima, houve alterações significativas nos *spots* relacionados a replicações consistentes. No entanto, uma análise da atividade quinômica global também pode trazer resultados significativos e correlacionados aos resultados anteriores. O gráfico mostrado na figura 15 mostra a mediana das intensidades de todos os 1024 *spots*, em condições de adesão *versus* condições controle, normalizadas de 0 a 100 e com uma cauda de exclusão de 5%. No geral, observa-se que há quinases mais ativas durante a fase de adesão. A possibilidade de fosforilações inespecíficas nos *spots* acima da cauda de normalização e abaixo da linha horizontal em 100 em 2h de adesão é desconsiderada porque nas condições controle o limite de 100 não foi ultrapassado. Assim, Src, PKA, FAK, GSK3 β , PKC dentre outras, apresentam forte evidência de ativação específica durante a adesão. Isso foi confirmado através da técnica de *Western blot*, de acordo com a disponibilidade de anticorpos, que validou as atividades quinásicas observadas no *chip* (Figura 16).

Uma análise da composição dos pseudosubstratos presentes no *chip* que apresentaram replicações consistentes *versus* o *chip* como um todo revelou que substratos contendo o aminoácido serina como sítio conhecido de fosforilação estão presentes em maior proporção no conjunto dos *spots* selecionados como confiáveis (Figura 17). Isso provavelmente é um reflexo da presença de várias quinases do grupo AGC (ex.: PKC e PKA) sendo capazes de fosforilar múltiplos pseudosubstratos que apresentaram fosforilação diferencial dentre os selecionados, indicando um papel fundamental desta família de quinases no processo de adesão.



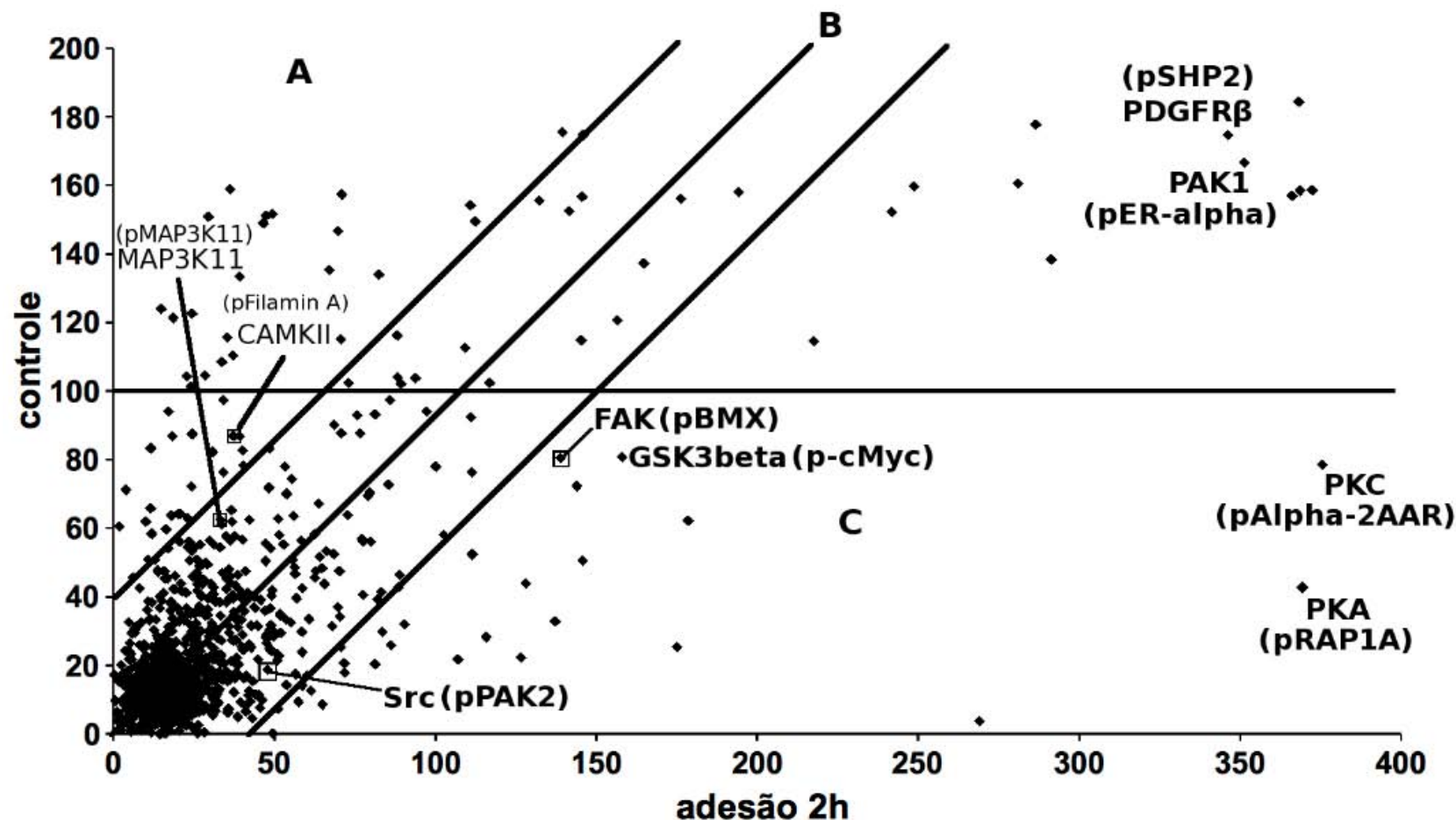


Figura 15. Análise quinômica global da adesão de células MC3T3-E1 por 2h sobre uma superfície de poliestireno *versus* um controle de células já aderidas em cultura por 10 horas. O gráfico classifica cada *spot* representando a condição de fosforilação do substrato presente no *chip*. Os substratos (indicados entre parênteses) associados às quinases Src, FAK, GSK3 β , PKC, PKA, PAK1, PDGFR β apresentaram aumento na intensidade de fosforilação enquanto os substratos associados às quinases MAP3K11 e CAMKII apresentaram redução na intensidade de fosforilação. As condições de fosforilação ao longo do *chip* podem variar dramaticamente durante o processo de adesão em relação à condição controle: a região A indica eventos em que a fosforilação foi reduzida durante a adesão, enquanto a região C indica eventos em que a fosforilação foi aumentada. A região B indica *spots* em que a variação foi menos pronunciada. *Spots* na região C abaixo da linha horizontal em 100 são os que apresentaram maior fosforilação diferencial, ao mesmo tempo em que provavelmente não indicam eventos de fosforilação inespecífica, já que a mediana correspondente no controle não apresentou intensidade acima da cauda de exclusão.

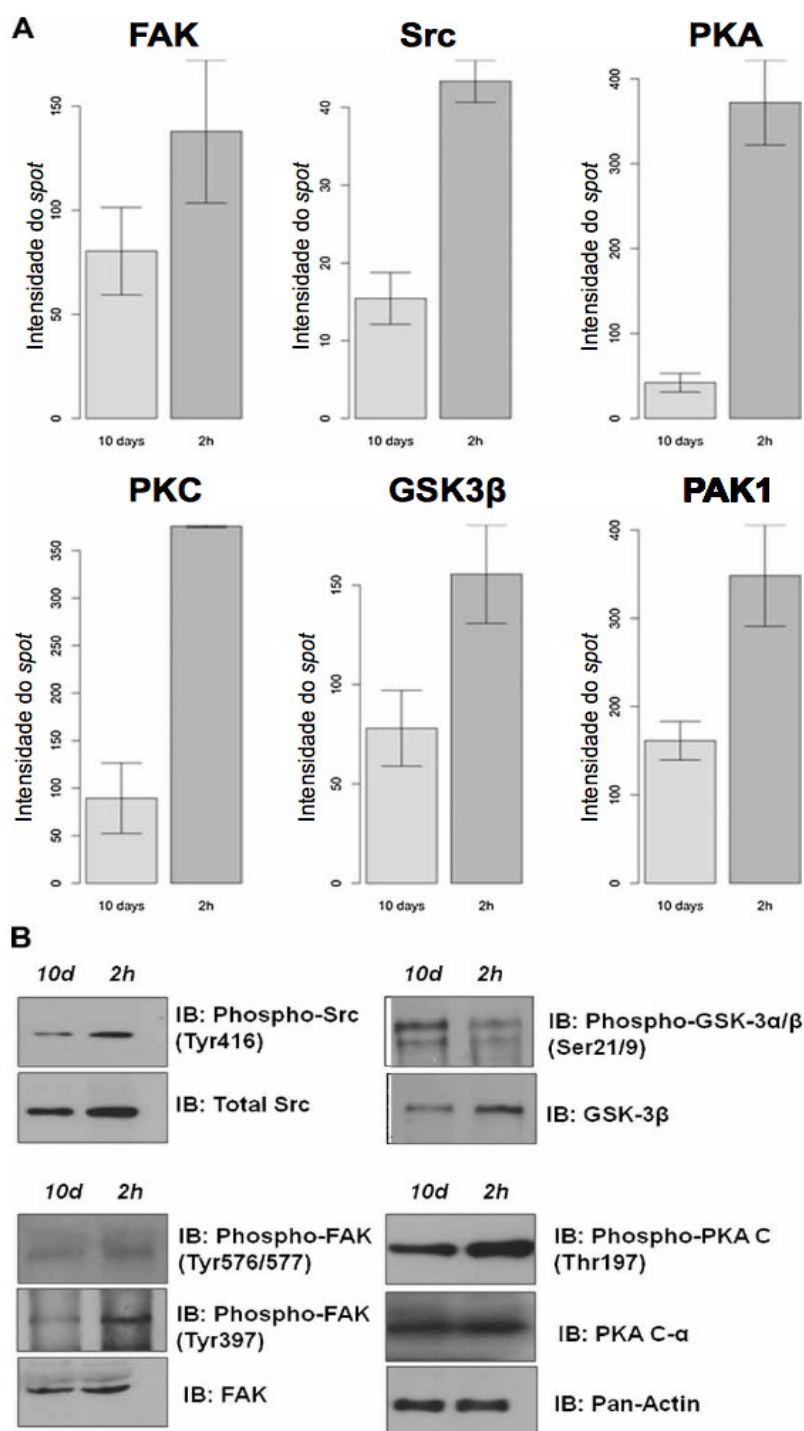


Figura 16. A. Replicações de *spots* no PepChip® para as quinases FAK, Src, PKA, PKC, GSK3β e PAK1, todas relevantes no processo de adesão osteoblástica. B. Os resultados do *chip* foram validados pela técnica de *Western blot*, específica para cada enzima, evidenciando que as quinases acima foram fosforiladas em resíduos específicos necessários à sua ativação. No caso de GSK3β, que apresenta-se menos fosforilada durante a adesão, o resíduo em questão é inibitório. Portanto, a enzima apresenta-se mais ativa, inibindo a proliferação celular.

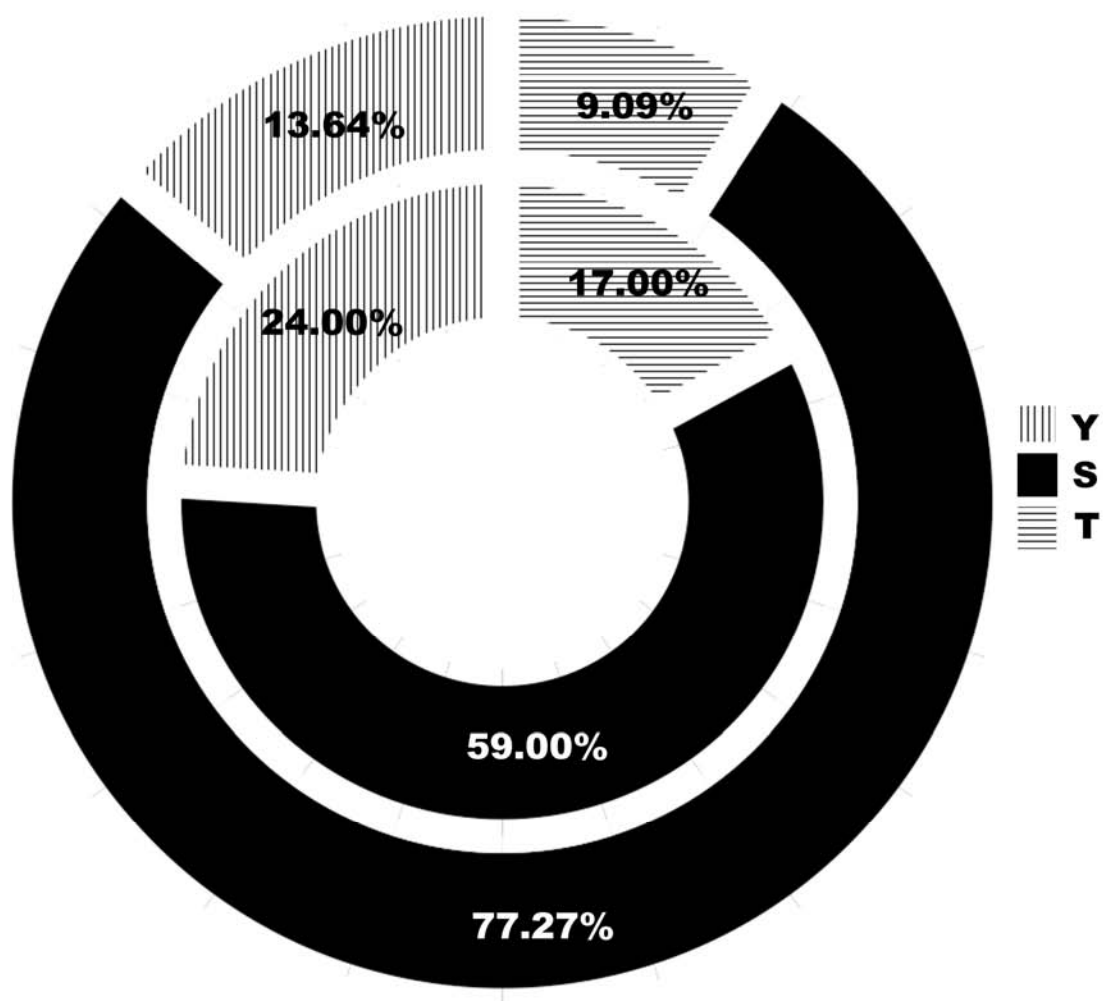


Figura 17. Os sítios de fosforilação dos *spots* selecionados durante a adesão possuem uma distribuição diferenciada em relação ao *chip* como um todo. Um significativo aumento dos peptídeos fosforilados em serina pode ser observado quando se compara apenas os *spots* diferencialmente fosforilados (círculo externo) em relação à distribuição observada em todo o *chip* (círculo interno). Y, S e T correspondem, respectivamente, à resíduos de tirosina, serina e treonina.

4 DISCUSSÃO E CONCLUSÕES

A análise dos resultados de técnicas de biologia molecular é bastante complexa, envolvendo questões que vão desde a seleção da metodologia até a validação dos resultados. Este desafio é ainda agravado pelos inúmeros *softwares* que existem hoje e são desenvolvidos a cada dia, aumentando a complexidade da tarefa.

As próprias metodologias laboratoriais estão sujeitas à dificuldades técnicas que podem resultar em dados imprecisos. No caso do PepChip®, essas dificuldades envolvem, entre outras: a) fosforilações inespecíficas devido ao pequeno tamanho do pseudosubstrato e b) acessibilidade reduzida da enzima ao substrato por conta do emparelhamento de substratos ao longo do *chip*. Essas dificuldades acabam gerando uma grande variabilidade entre as replicações.

O método de seleção de *spots* com base na consistência de replicações aqui proposto e aplicado em conjuntos de dados de laboratório reais expôs tais limitações técnicas verificando quais são as vantagens e desvantagens presentes na tecnologia do PepChip®. Nos 4 conjuntos de dados analisados pela metodologia PepMatrix com parâmetros $SD/A < 20\%$ e $80\% < A/M < 120\%$, houve um retorno médio de 80 *spots* selecionados para cada *chip*. Esse número representa aproximadamente 8% de todos os *spots* presentes no PepChip® Kinomics™, demonstrando que a porção útil do *chip* é relativamente pequena. Apesar disso, comprovou-se possível distinguir *spots* afetados pelas limitações do *chip* de outros com replicações consistentes. Estudos adicionais poderão comprovar se a tendência da reduzida área útil se mantém independentemente do tamanho do *chip* ou se seria mais eficiente fabricar *chips* contendo um número menor de *spots* selecionados de acordo com o interesse dos pesquisadores, ao contrário do *chip* produzido hoje,

que contém substratos de um grande número de enzimas pertencentes a processos biológicos não-relacionados.

Por outro lado, o PepChip® foi capaz de identificar com precisão eventos de fosforilação críticos para determinados processos biológicos, em meio a centenas de pseudosubstratos e quinases presentes num único *chip*. No caso de células TF1 expostas à violaceína por 30 minutos, a técnica de *Western blot* demonstrou que este composto promoveu considerável aumento da fosforilação no resíduo inibitório S9 da enzima GSK3 β , relacionada a diversas vias de sinalização regulando o crescimento celular, diferenciação e desenvolvimento (SIGFRIED *et al.*, 1990; BOUROUIS *et al.*, 1990). O PepChip®, por sua vez, apontou redução significativa e consistente da atividade da enzima. O PepChip® também evidenciou um aumento na atividade de GSK3 β após 24h de exposição à violaceína, o que pode indicar a redução da atividade do composto após esse período. Além disso, a enzima Src, associada a proliferação e diferenciação em uma grande variedade de tipos celulares, não apresentou diferença significativa em sua atividade no *chip* e em sua expressão no ensaio de *Western blot*.

Já no caso de células MC3T3-E1 pré-osteoblásticas durante a adesão a superfícies de poliestireno, o PepChip® foi capaz de apontar aumento de atividade em enzimas como Src, FAK, PKA e GSK3 β , cruciais no processo de adesão osteoblástica por seu envolvimento com a reorganização do citoesqueleto e a formação da adesão focal (ZAMBUZZI *et al.*, 2010), sendo que todos os eventos de fosforilação foram validados por um aumento na expressão das respectivas enzimas no *Western blot*.

No contexto da Biologia de Sistemas, a busca pela integração de resultados obtidos em estudos em larga escala é uma constante. No entanto, nem sempre é possível agrupar todos resultados num mesmo processo biológico. O PepChip®, por exemplo, pela sua própria natureza

variada em relação aos substratos presentes, muitas vezes retorna eventos de fosforilação relacionados a mais de um evento celular. No caso das células MC3T3-E1, a figura 12 e a tabela 7 apresentam um evento de fosforilação envolvendo o receptor PDGFR β , relacionado a doenças mieloproliferativas. Contudo, o foco da análise era a adesão osteoblástica e as enzimas relacionadas a ela, o que pode levar a estudos adicionais buscando elucidar se PDGFR β efetivamente participa do processo de adesão ou não.

Ferramentas voltadas a análise de microarranjos, como o PepMatrix, tem sido objeto de intenso debate na comunidade científica (MURIE *et al.*, 2009; ROCKE *et al.*, 2009; KOOPERBERG *et al.*, 2002; JEFFERY *et al.*, 2006). As particularidades de cada *chip* muitas vezes tornam um método adequado para um tipo de microarranjo completamente inadequado para outro. Nesse contexto, é importante definir um método simples e facilmente adaptável para a análise de vários tipos de microarranjos, de modo que os resultados sejam consistentes independentemente da técnica utilizada.

Mesmo ferramentas desenvolvidas visando uma abordagem no contexto da Biologia de Sistemas não trazem a facilidade de implementação de forma diversa daquela na qual é distribuída: o KegArray (WHEELLOCK *et al.*, 2009) é uma ferramenta desenvolvida para o mapeamento dos resultados de estudos em larga escala, incluindo dados de transcriptoma, proteoma e metaboloma, em mapas metabólicos do banco de dados KEGG (OGATA, 1999; KANEHISA, 2008). Enquanto apresenta um valor incalculável para estudos sistêmicos e a conexão entre as entidades biológicas de escala molecular e os fenômenos a elas associados, esta ferramenta não permite uma análise ideal dos dados de quinoma, uma vez que não contempla as particularidades de microarranjos de proteínas, que demandam análises voltadas à atividade enzimática e não à expressão gênica.

O uso da Estatística básica para análise de conjuntos de dados derivados de estudos de microarranjos provou-se adequado para lidar com as limitações técnicas do *chip*, permitindo a produção de conjuntos de dados adequados à análise posterior pelos pesquisadores. Apesar de não ser uma abordagem nova, as ferramentas disponíveis são focadas, na maioria das vezes, em microarranjos de nucleotídeos. Os parâmetros e procedimentos desses tipos de arranjos apresentam notáveis diferenças em relação a arranjos de proteínas. Por exemplo, arranjos de DNA consideram sinais bastante intensos como sinal de um alto nível de expressão, geralmente apresentando um limite inferior para considerar um *spot* válido. O PepChip®, por outro lado, avalia a atividade enzimática, o que exige valores de intensidade num intervalo que não seja considerado nem fraco demais nem tão forte que sugira uma fosforilação inespecífica.

Em vista do exposto, o PepMatrix apresenta-se como uma metodologia de análise capaz de integrar conceitos estatísticos simples a uma técnica promissora de análise em larga-escala de reações de fosforilação. De fácil uso e implementação, os resultados apresentados demonstram a robustez da metodologia de análise em vista da validação obtida por técnicas mais específicas. Sua adaptabilidade a outras técnicas não a deixa atrelada a apenas um modelo de *chip*, já que suas premissas também são facilmente alteráveis e sua usabilidade, com a possibilidade de uso através de um *website* ou através da execução local, tornam o PepMatrix uma ferramenta útil na obtenção de informações biologicamente relevantes a partir de microarranjos de proteínas.

5 BIBLIOGRAFIA

- ANAMIKA, K & SRINIVASAN, N (2007) Comparative kinomics of plasmodium organisms: unity in diversity. *Protein Pept. Lett.*, 14:509-517.
- ANSELME, K (2000) Osteoblast adhesion on biomaterials. *Biomaterials*, 21:667-681.
- ARANDA, B, ACHUTHAN, P, ALAM-FARUQUE, Y, ARMEAN, I, BRIDGE, A, DEROW, C, FEUERMANN, M, GHANBARIAN, AT, KERRIEN, S, KHADAKE, J, KERSSEMAKERS, J, LEROY, C, MENDEN, M, MICHAUT, M, MONTECCHI-PALAZZI, L, NEUHAUSER, N, ORCHARD, S, PERREAU, V, ROECHERT, B, van EIJK, K, HERMJAKOB, H (2009) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, Database Issue 38:525-531.
- BOUROIS, M, MOORE, P, RUEL, L, GRAU, Y, HEITZLER, P, SIMPSON, P (1990) An early embryonic product of the gene shaggy encodes a serine/threonine protein kinase related to the CDC28/cdc2+ subfamily. *EMBO J.*, 9:2877-2884.
- BUNESS, A, HUBER, W, STEINER, K, SÜLTMANN, H, POUTSKA, A (2005) arrayMagic: two-colour cDNA microarray quality control and preprocessing. *Bioinformatics*, 21:554-556.
- CAPRA, M, NUCIFORO, PG, CONFALONIERI, S, QUARTO, M, BIANCHI, M, NEBULONI, M, BOLDORINI, R, PALLOTTI, F, VIALE, G, GISHIZKY, ML, DRAETTA, GF, DI FIORE, PP (2006) Frequent alterations in the expression of serine/threonine kinases in human cancers. *Cancer Res.*, 66:8147-8154.
- DIKS, SH, KOK, K, O'TOOLE, T, HOMMES, DW, van DIJKEN, P, JOORE, J, PEPPELENBOSCH, MP (2004) Kinome profiling for studying lipopolysaccharide signal transduction in human peripheral blood mononuclear cells. *J. Biol. Chem.* 279:49206-49213.
- DUNNING, MJ, SMITH, ML, RITCHIE, ME, TAVARÉ, S (2007) beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics*, 23:2183-2184.
- EDELMAN, AM, BLUMENTHAL, DK, KREBS, EG (1987) Protein Serine/Threonine Kinases. *Ann. Rev. Biochem.* 56:567-613.
- FERRELL JR., JE (2009) Q&A: Systems biology. *Journal of Biology*, 8:2.
- GIBAS, C & JAMBECK, P (2001) Developing Bioinformatics Computer Skills. 1ed. Sebastopol: O'Reilly.
- GOMASE, VS & TAGORE S (2008) Kinomics. *Current Drug Metabolism*, 9:255-258.
- HUNTER, T & COOPER, JA (1985) Protein-Tyrosine Kinases. *Ann. Rev. Biochem.*, 54:897-930.
- HUNTER, T & JOHNSON, SA (2005) Kinomics: methods for deciphering the kinome. *Nature Methods*, 2:17-25.
- IRISH, JM, HOVLAND, R, KRUTZIK, PO, PEREZ, OD, BRUSERUD, Ø, GJERTSEN, BT, NOLAN, GP (2004) Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell*, 118:217-228.
- JEFFERY, IB, HIGGINS, DG, CULHANE, AC (2006) Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics*, 7:359.

- JENSEN, LJ, KUHN, M, STARK, M, CHAFFRON, S, CREEVEY, C, MULLER, J, DOERKS, T, JULIEN, P, ROTH, A, SIMONOVIC, M, BORK, P, von MERING, C (2009) STRING 8 – a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, Database Issue 37:412-416.
- KAUFFMANN, A, GENTLEMAN, R, HUBER, W (2009) arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25:415-416.
- KANEHISA, M, ARAKI, M, GOTO, S, HATTORI, M, HIRAKAWA, M, ITOH, M, KATAYAMA, T, KAWASHIMA, S, OKUDA, S, TOKIMATSU, T, YAMANISHI, Y (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, 36:480-484.
- KODACH, LL, BOS, CL, DURÁN, N, PEPPELENBOSCH, MP, FERREIRA, CV, HARDWICK, JC (2006) Violacein synergistically increases 5-fluorouracil cytotoxicity, induces apoptosis and inhibits Akt-mediated signal transduction in human colorectal cancer cells. *Carcinogenesis*, 27:508-516.
- KOOLMAN, J & RÖHM, K (2004) Color Atlas of Biochemistry, 2ed. Stuttgart-New York: Thieme.
- KOOPERBERG, C, SIPIONE, S, LEBLANC, M, STRAND, AD, CATTANEO, E, OLSON, JM (2002) Evaluating test statistics to select interesting genes in microarray experiments. *Hum. Mol. Genet.* 11:2223-2232.
- MARZZOCO, A & TORRES, BB (2007) Bioquímica Básica, 3ed. Rio de Janeiro: Guanabara Koogan.
- MANNING, G, WHYTE, DB, MARTINEZ, R, HUNTER, T, SUDARSANAM, S (2002) The protein kinase complement of the human genome. *Science*, 298:1912-1934.
- MURIE, C, WOODY, O, LEE, AY, NADON, R (2009) Comparison of small n statistical tests of differential expression applied to microarrays. *BMC Bioinformatics*, 10:45.
- NCBI (2004) Bioinformatics Factsheet. Disponível em:
<<http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>>. Acesso em: 13 out. 2009.
- NELSON, D & COX, M (2004) Lehninger Principles of Biochemistry, 4ed. New York: W.H.Freeman.
- OGATA, H, GOTO, S, SATO, K, FUJIBUCHI, W, BONO, H, KANEHISA, M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, 27:29-34.
- ROCKE, DM, IDEKER, T, TROYANSKAYA, O, QUACKENBUSH, J, DOPAZO, J (2009) Papers on normalization, variable selection, classification or clustering of microarray data. *Bioinformatics*, 25:701-702.
- RUAL, J-F, VENKATESAN, K, HAO, T, *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437:1173-1178.
- SCHUCHHARDT, J, BEULE, D, MALIK, A, WOLSKI, E, EICKHOFF, H, LEHRACH, H, HERZEL, H (2000) Normalization strategies for cDNA microarrays. *Nucleic Acids Res.*, 28:47.
- SIGFRIED, E, PERKINS, LA, CAPACI, TM, PERRIMON, N (1990) Biological activity of soluble *wingless* protein in cultured *Drosophila* imaginal disc cells. *Nature*, 345:825-829.
- SMITH, E, COETZEE, GA, FRENKEL, B (2002) Glucocorticoids inhibit cell cycle progression in differentiating osteoblasts through glycogen synthase kinase-3beta. *J. Biol. Chem.*, 277:18191-18197.
- STRYER, L, BERG, JM, TYMOCZKO, JL (2002) Biochemistry, 5ed. New York: W.H.Freeman.

- SUDO, H, KODAMA, HA, AMAGAI, Y, YAMAMOTO, S, KASAI, S (1983) *In vitro* differentiation and calcification in a new clonal osteogenic cell line derived from newborn mouse calvaria. *J. Cell Biol.*, 96:191-198.
- van BAAL, JW, DIKS, SH, WANDERS, RJ, RYGIEL, AM, MILANO, F, JOORE, J, BERGMAN, JJ, PEPPELENBOSCH, MP, KRISHNADATH, KK (2006) Comparison of kinome profiles of Barrett's esophagus with normal squamous esophagus and normal gastric cardia. *Cancer Res.*, 66:11605-11612.
- WHEELLOCK, CE, WHEELLOCK, AM, KAWASHIMA, S, DIEZ, D, KANEHISA, M, van ERK, M, KLEEMANN, R, HAEGGSTROM, JZ, GOTO, S (2009) Systems biology approaches and pathway tools for investigating cardiovascular disease. *Mol Biosyst.*, 5:588-602.
- XENARIOS, I, RICE, DW, SALWINSKI, L, BARON, MK, MARCOTTE, EM, EISENBERG, D (2000) DIP: the Database of Interacting Proteins. *Nucleic Acids Res.*, 28:289-291.
- YANG, N, HIGUCHI, O, OHASHI, K, NAGATA, K, WADA, A, KANGAWA, K, NISHIDA, E, MIZUNO, K (1998) Cofilin phosphorylation by LIM-kinase 1 and its role in Rac-mediated actin reorganization. *Nature*, 393:809-812.
- ZAMBUZZI, WF (2008) Mecanismos de Transdução de Sinal Envolvidos com a Diferenciação de Osteoblastos e Osteócitos. Tese de Doutorado, Universidade Estadual de Campinas.
- ZAMBUZZI, WF, MILANI, R, TETI, A (2010) Expanding the role of Src and protein-tyrosine phosphatases balance in modulating osteoblast metabolism: Lessons from mice. *Biochimie*, In Press. Disponível online em 18 de janeiro. DOI: 10.1016/j.biochi.2010.01.002.

6 APÊNDICE

6.1 ARTIGOS PUBLICADOS

MILANI, R. , FERREIRA, C. V. , GRANJEIRO, J. M. , PAREDES-GAMERO, E. J. , SILVA, R. A. , JUSTO, G. Z. , NADER, H. B. , PEPPELENBOSCH, M. P. , AOYAMA, H. , GALEMBECK, E. , ZAMBUZZI, W. F. . Phosphoproteome reveals an atlas of protein signaling networks during osteoblast adhesion. *Journal of Cellular Biochemistry* - JCB-09-048. Aceito em 26/11/2009 .

ZAMBUZZI, W. F. , MILANI, R. , TETI, A. . Expanding the role of Src and protein tyrosine phosphatases balance in modulating osteoblast metabolism: lessons from mice. *Biochimie – BIOCHI-D-09-00378*. Aceito em 06/01/2010.

6.2 ARTIGOS COM PARECER FAVORÁVEL EM REVISÃO

CHAVES-NETO, A. , QUEIROZ, K. C. , MILANI, R. , PAREDES-GAMERO, E. J. , JUSTO, G. Z. , PEPPELENBOSCH, M. P. , FERREIRA, C. V. . Metabolic overview of the initial-stage of ascorbic acid/ β -glycerophosphate-induced osteoblastic differentiation. Submetido para *Journal of Cellular Biochemistry* sob o número JCB-09-0401 em 31/07/2009. Aceito para revisão em 31/08/2009.

6.3 ARTIGOS EM PROCESSO DE SUBMISSÃO

MILANI, R. , QUEIROZ, K. C. , ZAMBUZZI, W. F. , SOUZA, T. A. , AZOUBEL, S. , DIKS, S. H. , PEPPELENBOSCH, M. P. , FERREIRA, C. V. , GALEMBECK, E. . PepMatrix: Strict Replicate Evaluation towards Differential Phosphorylation in Peptide Arrays.

6.4 APRESENTAÇÕES EM CONGRESSOS

MILANI, R. ; MARTINS, A. R. ; NETO, A. H. C. ; FERREIRA, C. V. ; PEPPELENBOSCH, M. P. ; GALEMBECK, E. . Developing a pipeline for large-scale analysis of a protein kinase *chip*. In: Bioinformatics to Systems Biology - BSB '08, 2008, Congresso Online.

MILANI, R. ; ZAMBUZZI, W. F. ; DIKS, SH. ; PEPPELENBOSCH, M. P. ; AOYAMA, H. ; FERREIRA, C. V. ; GALEMBECK, E. . Bioinformatizing Osteoblast Adhesion: A Critical View of Peptide Arrays. In: XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2009, Águas de Lindóia - SP.

PERES, L. L. ; ZAMBUZZI, W. F. ; MILANI, R. ; DIKS, SH. ; PEPPELENBOSCH, M. P. ; GALEMBECK, E. ; FERREIRA, C. V. . On the Road to Targeted Therapy of Resistant Human Breast Cancer. In: XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2009, Águas de Lindoia - SP.

SILVA, R. A. ; MACHADO, D. ; MILANI, R. ; PEPPELENBOSCH, M. P. ; SHISHIDO, S. M. ; SOUZA, A. C. S. ; FERREIRA, C. V. ; GALEMBECK, E. ; JUSTO, G. Z. . Phosphokinome Profiling of Human Keratinocyte Metabolic Reprogramming to Different Stressors. In: XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2009, Águas de Lindoia - SP.