Centro Nacional de Pesquisa em Energia e Materiais (CNPEM)

Laboratório Nacional de Biociências (LNBio)

Pipeline para modelagem dinâmica automatizada de redes biológicas associadas ao câncer

Orientadora: Dra. Gabriela Vaz Meirelles (LNBio/CNPEM)

Co-orientador: Prof. Dr. Jorg Kobarg (LNBio/CNPEM)

Colaboradores:

Prof. Dr. Paulo Sérgio Lopes de Oliveira (LNBio/CNPEM)

Prof. Dra. Adriana Franco Paes Leme (LNBio/CNPEM)

Prof. Dr. Guilherme P. Telles (UNICAMP)

MSc. Marcelo Falsarella Carazzolle (UNICAMP)

I. INTRODUÇÃO E JUSTIFICATIVA

1. Biologia de Sistemas: análise de redes de interação biológicas e descoberta de drogas

Na última década, a pesquisa na área biológica ultrapassou a capacidade humana de analisar o imenso volume de dados acumulados, gerados a partir do desenvolvimento de tecnologias em larga escala, sendo fundamental o uso de programas computacionais para o armazenamento, busca, análise e integração desses dados. A Biologia de Sistemas é uma área multidisciplinar que visa integrar os dados relativos a genes e proteínas individuais e investigar o comportamento e relações entre os diversos elementos de um sistema biológico para explicar o seu funcionamento¹.

A Biologia de Sistemas encontra-se centrada na construção de modelos que descrevam os sistemas biológicos. Esses modelos proporcionam uma plataforma para (i) a interrogação sistemática e verificação experimental de um dado processo biológico, (ii) o gerenciamento da complexidade de centenas de componentes celulares e interações, e (iii) a revelação de propriedades emergentes e predição de consequências a partir da perturbação de determinados componentes das diferentes vias.

Um importante objetivo da modelagem computacional de redes biológicas é a identificação de potenciais alvos de drogas. A aplicação prática da Biologia de Sistemas à descoberta de novas drogas reside em três abordagens principais: a integração computacional de dados oriundos de experimentos em larga escala, a modelagem computacional de vias metabólicas, regulatórias e/ou de sinalização de células saudáveis e doentes (câncer etc.) a partir de dados da literatura (para a seleção de alvos e desenhos de testes clínicos), e o uso dos próprios sistemas de células humanas complexos para interpretar e prever as atividades biológicas de drogas e alvos gênicos (abordagem experimental direta para classificar respostas biológicas relevantes e validar alvos/drogas)².

1.1. Modelagem dinâmica de redes biológicas: perspectivas na compreensão e tratamento do câncer

O câncer é entendido, atualmente, não apenas como uma patologia altamente heterogênea com relação ao tipo celular e tecido de origem, mas também como uma doença envolvendo a desregulação de múltiplas vias interconectadas, abrangendo processos celulares fundamentais, como proliferação, diferenciação, migração e morte³. Além disso, essas redes se conectam a componentes externos às próprias células tumorais, incluindo outras células e componentes da matriz extracelular^{4,5}. Assim, as atividades das redes moleculares que executam processos metabólicos ou de organização do citoesqueleto, ou regulam estes através de transdução de sinal, encontram-se alteradas de uma forma complexa por diversas mutações gênicas em conjunto com o contexto ambiental. Portanto, para uma efetiva compreensão da doença e propostas de novas terapias, torna-se necessária uma abordagem que avalie o problema nos seus diferentes níveis, a partir da (i) identificação das vias desreguladas, (ii) elucidação das consequências de mutações nas atividades da rede, (iii) integração da rede ao comportamento celular, (iv) integração do comportamento celular aos processos no nível dos tecidos, e (v) predição dos efeitos a partir de intervenções moleculares.

Esta perspectiva leva a, no mínimo, duas consequências: (i) a caracterização da desregulação patológica deverá ser multivariada e quantitativa, e (ii) a compreensão preditiva ou mecanicista da patologia dependerá de uma abordagem de Biologia de Sistemas, fortemente baseada na modelagem computacional de redes complexas⁶. Devido à multidisciplinaridade desta área, metodologias para análises computacionais podem variar consideravelmente, dependendo da pergunta biológica e dos dados experimentais disponíveis, abrangendo diferentes tipos de formalismos, desde abordagens preditivas qualitativas (discretas) a quantitativas (contínuas), incluindo: modelos baseados em lógica (redes Booleanas, lógica *fuzzy*, etc), redes Bayesianas, Petri nets, álgebra de processos, modelos baseados em restrições,

equações diferenciais (ODE/PDE), redes neurais artificiais, modelos baseados em regras, *interacting state machines* (ISM), autômatos celulares, e modelos baseados em agentes (ABM)⁶⁻⁸. Esta diversidade pode levar à fragmentação de esforços em modelagem, uma vez que ela dificulta a integração de modelos provenientes de diferentes fontes. Portanto, simulações no nível de uma célula inteira se beneficiariam do desenvolvimento de uma plataforma para modelagem, análise e simulação baseada em um único ou poucos formalismos, capazes de integrar os diferentes componentes e suas relações, englobando todos os tipos de redes biológicas.

Nesse contexto, o desenvolvimento de novos algoritmos híbridos, combinando os métodos mais eficientes para cada tipo de rede, proporcionaria uma modelagem automatizada das redes de interação em um pipeline único, sendo o objetivo central deste projeto.

2. Resultados prévios do grupo

2.1. NIMA-related kinases (Neks) e câncer

Como protótipo dessa família, a cinase NIMA de *Aspergillus nidulans* contribui para múltiplos aspectos da progressão da mitose⁹. Os mamíferos apresentam uma extensa família de 11 Neks estruturalmente homólogas à NIMA¹⁰, e, de modo geral, os genes que codificam as Neks humanas encontram-se super-expressos¹¹, ou são alvos de mutações em câncer¹². Sendo assim, é de fundamental importância o conhecimento das vias de sinalização em que as Neks atuam nas células, os mecanismos moleculares envolvidos e a identificação de potenciais inibidores que possam, eventualmente, ser utilizados em novas estratégias terapêuticas anti-câncer.

Nosso grupo realizou a identificação de parceiros de interação para a hNek1¹³, hNek6¹⁴, hNek7¹⁵, hNek3, hNek4, hNek5, hNek9, hNek10 e hNek11. Além disso, realizamos triagens com compostos comerciais utilizando a técnica de *Thermal Shift Assay* para a hNek1, hNek6 e hNek7, além de análises de docking e ensaios de fosforilação *in vitro* para a confirmação desses compostos como possíveis inibidores¹⁶. Novos inibidores também serão testados em bioensaios e utilizados nas análises *in silico* propostas neste projeto.

2.2. Plataforma IIS

A plataforma IIS (*Integrated Interactome System*), desenvolvida em colaboração com o MSc. Marcelo F. Carazzolle, em projeto coordenado pela Dra. Gabriela V. Meirelles¹⁷, apresenta interface web para anotação, análise e visualização de perfis de interação de proteínas/genes, metabólitos e drogas de interesse, a partir da criação de um projeto (protegido por senha) para cada usuário. IIS funciona em 4 módulos conectados: (i) Módulo “Submissão”, que recebe cromatogramas, derivados de sequenciamento Sanger a partir de diferentes métodos de screening (duplo-híbrido, transcrito, etc) e realiza o processamento dessas seqüências; (ii) Módulo “Busca”, que possibilita ao usuário fazer buscas pelos reads processados ou pelas listas de proteínas/genes, metabólitos e drogas de interesse, selecioná-los e adicioná-los ao seu projeto; (iii) Módulo “Anotação”, que realiza a montagem dos reads em contigs e singlets, ou utiliza listas de proteínas/genes carregadas pelo usuário, e os compara contra diversos bancos de dados, gerando tabelas com anotação automática que podem ser manualmente curadas; e (iv) Módulo “Interactoma”, que faz o blast dos contigs e singlets ou compara as listas de proteínas/genes, metabólitos e drogas contra o nosso banco GPMGDID (Global Protein-Metabolite-Gene-Drug Interaction Database), formado pelo agrupamento de diferentes bancos de interação, gerando as redes de interação que podem ser visualizadas via web ou analisadas através do software Cytoscape.

A plataforma IIS surgiu da necessidade crescente por ferramentas apropriadas a uma análise sistemática e padronizada de interações físicas, genéticas e químico-genéticas a partir de dados processados ou que necessitam de um processamento automatizado, e requer agora que novos módulos sejam implementados no sentido de ampliar sua funcionalidade, possibilitando a modelagem automatizada destas redes de interação, como parte dos objetivos propostos por este projeto.

II. OBJETIVOS

1. Desenvolvimento de novos algoritmos para a simulação de redes complexas aplicadas na compreensão e predição de mecanismos moleculares envolvidos com câncer e outras doenças.
2. Desenvolver um *pipeline* para a modelagem automatizada de redes de interação baseado nos algoritmos selecionados a partir do item II.1, e sua integração à plataforma IIS como o módulo “Modelagem Dinâmica de Redes Complexas”.

III. METODOLOGIA

1. Revisão bibliográfica e avaliação dos principais algoritmos

Será realizado um levantamento bibliográfico dos principais métodos/algoritmos utilizados para a simulação de redes biológicas, desde abordagens preditivas qualitativas (discretas) a quantitativas (contínuas). Esta revisão já encontra-se em andamento e, através dela, será realizada uma análise e seleção dos algoritmos descritos de melhor performance para cada um dos tipos de redes: (i) redes de interação proteína-proteína, (ii) redes de sinalização, (iii) redes regulatórias e (iv) redes metabólicas.

2. Desenvolvimento, implementação e teste de novos algoritmos promissores

Serão desenvolvidos novos algoritmos híbridos, buscando combinar os métodos mais eficientes para cada caso. Estes novos algoritmos serão implementados e sua eficácia avaliada através da comparação com os algoritmos selecionados no item III.1. Para isso será realizada a simulação de redes construídas a partir de experimentos *in house*, em particular o interactoma das Neks, em colaboração com o grupo do Prof. Dr. Jorg Kobarg, e o proteoma/secretoma de células de câncer, em colaboração com o grupo da Profa. Dra. Adriana F. Paes Leme, além de redes construídas a partir de dados da literatura, depositados em bancos de dados públicos. A validação experimental dos novos resultados obtidos pelas simulações das diferentes redes será realizada em colaboração com o grupo do Prof. Dr. Jorg Kobarg.

3. Desenvolvimento do módulo “Modelagem Dinâmica de Redes Complexas” e sua integração à plataforma IIS

O módulo “Modelagem Dinâmica de Redes Complexas” será desenvolvido como um *pipeline on-line* a ser posteriormente integrado à página do projeto de cada usuário na plataforma IIS, assim como os módulos “Anotação” e “Interactoma”, previamente desenvolvidos em nosso laboratório¹⁷. O módulo será desenvolvido a partir dos algoritmos mais eficazes, definidos por ranking no item III.2, e terá como objetivo a simulação automática e otimizada de redes biológicas de interesse, possibilitando ao usuário inserir perturbações na rede (mutações gênicas, inibidores, etc) e prever comportamentos celulares e mecanismos moleculares emergentes em cada condição/tempo de estudo. Isso facilitará novas descobertas biológicas, bem como poderá ser aplicado ao diagnóstico e tratamento de doenças.

Este módulo consistirá nas seguintes etapas: (i) *upload* dos dados ao projeto do usuário, (ii) construção das redes de interação nas condições definidas pelo usuário (incluindo possíveis perturbações), e (iii) envio por email do relatório da análise e do link para visualização da rede, gráficos e tabelas gerados pelas simulações. Os dados carregados como *input* em cada projeto poderão ser qualitativos (cromatogramas ou listas de genes/proteínas e metabólitos) ou quantitativos, em forma de tabelas, oriundos de experimentos de transcrito (microarranjo, RNAseq), proteoma (MS), e/ou metaboloma (RMN). As redes de interação condicionais serão construídas a partir do banco de dados de interações globais GPMGDID, construído em nosso laboratório, considerando-se a seleção dos seguintes parâmetros: (i) organismo, (ii) tipo celular ou tecido, (iii) condição normal ou patológica, (iv) compartimento celular, (v) processo biológico, (vi) integração com dados quantitativos públicos (transcrito, proteoma e/ou metaboloma), e (vii) perturbação (mutação gênica, inibidores, etc).

IV. REFERÊNCIAS BIBLIOGRÁFICAS

1. Kitano H. (2002) Systems Biology: a brief overview. *Science* 295: 1662-1664.
2. Butcher EC, Berg EL, Kunkel EJ. (2004) Systems biology in drug discovery. *Nat Biotechnol.* 22(10): 1253-9.
3. Pawson T. et al. (2008) Network medicine. *FEBS Lett.*, 582, 1266–1270.
4. Kenny PA. et al. (2007) Targeting the tumor microenvironment. *Front. Biosci.*, 12, 3468–3474.
5. Weinberg RA. (2008) Mechanisms of malignant progression. *Carcinogenesis*, 29, 1092–1095.
6. Kreeger PK, Lauffenburger DA. (2010) Cancer systems biology: a network modeling perspective. *Carcinogenesis*, 31(1):2-8.
7. Feiglin A, Hacohen A, Sarusi A, Fisher J, Unger R, Ofra Y. (2012) Static network structure can be used to model the phenotypic effects of perturbations in regulatory networks. *Bioinformatics*, 28(21):2811-8.
8. Machado D, Costa RS, Rocha M, Ferreira EC, Tidor B, Rocha I. (2011) Modeling formalisms in Systems Biology. *AMB Express*. 1:45.
9. O'Regan L, Blot J, Fry AM. (2007) Mitotic regulation by NIMA-related kinases. *Cell Div.* 2: 25.
10. Quarumby LM, Mahjoub MR. (2005) Caught Nek-ing: Cilia and centrioles. *J Cell Sci.* 118: 5161-5169.
11. Capra M, Nuciforo PG, Confalonieri S, Quarto M, Bianchi M, et al. (2006) Frequent alterations in the expression of serine/threonine kinases in human cancers. *Cancer Res*, 66 (16), 8147–8154.
12. Greenman CP, Stephens R, Smith GL, Dalgliesh C, Hunter G, et al. (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 446: 153-158.
13. Surpili, M.J.; Delben, T.M.; Kobarg, J. (2003) Identification of proteins that interact with the central coiled coil region of the human protein kinase NEK1 involved in the cell cycle regulation. *Biochemistry* 42: 15369-15376.
14. Meirelles GV, Lanza DCF, Silva JC, Bernachi JS, Leme AFP, Kobarg J. Characterization of hNek6 Interactome Reveals an Important Role for Its Short N-Terminal Domain and Colocalization with Proteins at the Centrosome. *J Proteome Res* 2010, 9(12): 6298-316.
15. Souza EE, Meirelles GV, Smetana J, Kobarg J. The interactome profile of Nek7 is distinct from that of Nek6 and the substrate recognition is mediated by the short N-terminal domains. (Manuscrito em preparação).
16. Moraes EC, Meirelles GV, Souza EE, Souza TACB, Murakami M, Kobarg J. Comparison of the kinase inhibitor profile for Nek1, 2, 6 and 7 and analysis of the structural basis for the inhibitor specificity. (Manuscrito em preparação)
17. Carazzolle MF, Carvalho LM, Vidal RO, Pereira GAG, Kobarg J, Meirelles GV. IIS – Integrated Interactome System: a web-based platform for the annotation, analysis and visualization of protein-metabolite-gene-drug interactions by integrating a variety of data sources and tools. (Manuscrito em preparação)