

Pipeline para modelagem dinâmica de redes biológicas no câncer

Pesquisadora responsável: Dra. Gabriela Vaz Meirelles (LNBio/CNPEM)

Colaboradores:

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

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

Dr. Marcelo Falsarella Carazzolle (UNICAMP)

I. INTRODUÇÃO E JUSTIFICATIVA

1. Biologia de Sistemas: análise de redes de interação biológicas

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.

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 morte². Além disso, essas redes se conectam a componentes externos às próprias células tumorais, incluindo outras células e componentes da matriz extracelular^{3,4}. 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 complexas⁵. 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, etc.), redes Bayesianas, Petri nets, álgebra de processos, equações diferenciais (ODE/PDE), redes neurais artificiais e autômatos celulares⁶. 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 na unificação destes formalismos, capazes de integrar os diferentes componentes e suas relações, englobando todos os tipos de redes biológicas.

Apesar de nenhum formalismo individualmente ser capaz de cobrir todo o espectro de características requeridas para a modelagem de todos os componentes biológicos, Petri nets está entre os formalismos que abrangem a maior parte

das características⁶. Nesse contexto, a implementação de um algoritmo combinando as características mais eficientes dos modelos descritos em Petri nets 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. Plataforma IIS

A plataforma IIS (*Integrated Interactome System*), desenvolvida em colaboração com o Dr. Marcelo F. Carazzolle, em projeto coordenado pela Dra. Gabriela V. Meirelles⁷, 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 sequências; (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. Seleção das três vias metabólicas ou de sinalização mais enriquecidas no interactoma de carcinoma e melanoma⁸ e obtenção dos seus mapas através do banco KEGG⁹.
2. Desenvolvimento de um algoritmo para conversão dos arquivos KGML (mapas de vias do banco KEGG) para SBML.
3. Avaliação do uso de Petri nets e modificação do algoritmo para posterior integração na plataforma IIS. Predição de comportamentos celulares a partir de perturbações na rede (mutações gênicas, inibidores, etc).

III. METODOLOGIA

1. Seleção de vias de sinalização/metabólicas de carcinoma e melanoma

Serão selecionadas as três vias (redes) de sinalização e/ou metabólicas mais enriquecidas ($p \leq 0.05$) a partir de experimentos *in house* do secretoma de carcinoma e melanoma⁸, em colaboração com o grupo da Profª. Dra. Adriana F. Paes Leme. As vias enriquecidas serão extraídas dos interactomas de carcinoma e melanoma (três vias de cada interactoma), previamente analisados pelo “Módulo 4: Interactoma” da plataforma IIS, e os mapas correspondentes às vias enriquecidas serão obtidos a partir do banco de dados KEGG (Kyoto Encyclopedia of Genes and Genomes)⁹.

2. Desenvolvimento de um algoritmo para conversão de arquivos KGML para SBML

Após a seleção dos mapas das vias de interesse, estas serão acessadas por meio de download do banco KEGG. Será desenvolvido um algoritmo para a conversão dos arquivos das vias no formato KGML para SBML (Systems Biology Markup Language), de modo similar ao programa KEGGtranslator¹⁰ com a diferença de que o algoritmo a ser desenvolvido será capaz de converter tanto vias metabólicas (com os campos de reação bem definidos), quanto vias de sinalização (que não apresentam informação de reação).

3. Avaliação do uso de Petri nets para modelagem de redes

Os arquivos convertidos para o formato SBML serão, por sua vez, utilizados como entrada no programa Snoopy¹¹ para realizar as simulações das redes em Petri nets, a fim de se observar o desempenho e eficácia do uso de Petri nets para os dados em questão. Serão inseridas perturbações nas redes, como mutações gênicas e inibidores direcionados para proteínas de interesse em cada via, visando prever comportamentos celulares e possíveis mecanismos moleculares característicos das células de carcinoma e melanoma, para posterior comparação entre elas e com células normais. O algoritmo utilizado para simular as redes em Petri nets será, posteriormente, modificado para integração à plataforma IIS.

IV. PERSPECTIVAS

Como perspectivas deste projeto, pretendemos otimizar os algoritmos desenvolvidos e modificados neste projeto para a implementação do módulo “Modelagem Dinâmica de Redes”, que será desenvolvido como um *pipeline on-line* integrado à plataforma IIS, assim como os módulos “Anotação” e “Interactoma”, previamente desenvolvidos em nosso laboratório⁷. O módulo 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.

V. REFERÊNCIAS BIBLIOGRÁFICAS

1. Kitano H. (2002) Systems Biology: a brief overview. *Science* 295: 1662-1664.
2. Pawson T. et al. (2008) Network medicine. *FEBS Lett.*, 582, 1266–1270.
3. Kenny PA. et al. (2007) Targeting the tumor microenvironment. *Front. Biosci.*, 12, 3468–3474.
4. Weinberg RA. (2008) Mechanisms of malignant progression. *Carcinogenesis*, 29, 1092–1095.
5. Kreeger PK, Lauffenburger DA. (2010) Cancer systems biology: a network modeling perspective. *Carcinogenesis*, 31(1):2-8.
6. Machado D, Costa RS, Rocha M, Ferreira EC, Tidor B, Rocha I. (2011) Modeling formalisms in Systems Biology. *AMB Express*. 1:45.
7. Carazzolle MF, Carvalho LM, Slepicka HH, Vidal RO, Pereira GAG, Kobarg J, Meirelles GV. (2014) 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. *PLoS ONE* 9(6): e100385.
8. Kawahara R, Meirelles GV, Heberle H, Domingues RR, Granato DC, Yokoo S, Canevarolo R, Winck FV, Ribeiro ACP, Brandão TB, Filgueiras PR, Cruz KSP, Barbuto JA, Poppi RJ, Minghim R, Telles GP, Fonseca FP, Fox JW, Santos-Silva AR, Coletta RD, Sherman NE, Leme AFP. Integrative Analysis to Select Cancer Candidate Biomarkers to Targeted Validation. In revision.
9. Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008) KEGG for linking genomes to life and the environment. *Nucleic acids research* 36, D480-484.
10. Wrzodek C, Dräger A, Zell A. (2011) KEGGtranslator: visualizing and converting the KEGG PATHWAY database to various formats. *Bioinformatics*. 27(16):2314-5.
11. Rohr C, Marwan W, Heiner M. (2010) Snoopy--a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics*. 26(7):974-5.