

## A integração das “ômicas” através de ferramentas de Bioinformática

Laércio Pol Fachin<sup>1,2</sup>

<sup>1</sup>Instituto Aggeu Magalhães, Fiocruz-PE, Recife, Brasil

<sup>2</sup>Departamento de Química Fundamental, UFPE, Recife, Brasil

Os avanços na área de biotecnologia possibilitaram que estudos em pesquisa biomédica respondessem, de forma eficiente, a importantes questões em diferentes escalas “ômicas”, incluindo genética, genômica, transcriptômica, epigenômica, proteômica, metabolômica, lipidômica e glicômica. Cada uma dessas áreas, com suas características e aplicações específicas, acumulam um volume de dados extremamente elevado, porém permitindo o melhor entendimento de processos celulares e metabólicos relacionados a diagnóstico, tratamento e prevenção de doenças. Ao mesmo tempo que o uso e disponibilidade de diferentes técnicas aplicáveis a essas “ômicas” vem aumentando, é crescente também o conceito de que nenhuma delas é capaz de lidar com a complexidade de sistemas biológicos em sua totalidade por si só. Como exemplos, o nível de expressão de um gene não reflete diretamente a quantidade de proteína sintetizada, sua localização, atividade biológica, perfil de glicosilação ou sua relação funcional com metabolomas. Ainda, vários níveis de regulação ocorrem após a transcrição gênica, ou seja, de forma pós-transcricional, traducional ou pós-traducional, bem como através de controles bioquímicos, tais como alosteria ou regulação por *feedback*. Dessa forma, para obter um panorama mais completo de processos biológicos de uma forma geral, tais dados experimentais devem ser integrados para serem analisados. Nesse contexto, embora o desenvolvimento de métodos que consigam realizar a análise integrativa desses conjuntos de dados seja um dos problemas mais relevantes no campo das ciências computacionais, o uso dessas ferramentas vem crescendo em publicações internacionais.

Os objetivos dessas análises integrativas podem ser resumidos, basicamente, em três tipos: (1) a descoberta de novos mecanismos moleculares; (2) o agrupamento de amostras, ou indivíduos; ou (3) a predição de um desfecho, como a sobrevivência ou eficácia de uma terapia. Ainda, tal integração pode ser feita de forma (1) sequencial, na qual um conjunto de dados é analisado inicialmente, e correlacionado com o segundo, de forma causal; ou (2) simultânea, na qual cada conjunto de dados é considerado um entidade separada, porém etapas preliminares de integração dos dados costumam ser necessárias. Diferentes metodologias computacionais vem sendo empregadas nesse sentido, incluindo aprendizagem de máquina, métodos de estatística multivariada, e métodos baseados em rede associados ou não a métodos estatísticos bayesianos, entre outros – empregados conforme sua melhor adequação a um dado conjunto de dados e “ômicas” a ser integradas. A implementação dessas metodologias atualmente se encontra dividida como ferramentas individuais em vários programas de uso comum na literatura, inclusive disponibilizados de forma gratuita e na forma de código aberto.

Do ponto de vista aplicação, enquanto a transcriptômica oferece uma análise global da expressão gênica (e, portanto, dos RNA mensageiros produzidos por uma célula ou por um conjunto de células), a proteômica possibilita a análise simultânea de todas as proteínas de fato expressas nessas células, em um

tecido ou organismo em condições fisiológicas ou pré-determinadas. A proximidade entre essas duas “ômicas” possibilita que a integração dos conjuntos de dados advindos dessas duas áreas pode ser realizada com diversos fins. A mais óbvia, de complementação dos dados, possibilita que algum viés de detecção de alguma das tecnologias possa ser evitado, ou que uma melhor cobertura das mudanças metabólicas possa ser observado. Da mesma forma, a integração dos dados pode ser realizada para validação ou para auxiliar na revelação de novos aspectos biológicos não acessíveis por apenas um dos conjuntos de dados. A partir do que já foi observado até então, embora se possa pensar que a correlação entre os níveis de expressão de RNA mensageiro e abundância proteica estejam fortemente baseados no dogma central da biologia molecular, dados de ferramentas integrativas não apontam necessariamente na mesma direção. A não-correlação, ou correlação fraca entre esses aspectos, vem sendo associado a diferentes fatores. Entre eles, destacam-se a regulação da síntese proteica através de modificações pós-transcricionais, diferenças na meia-vida de RNA mensageiro e proteínas, a regulação proteica realizada através de modificações pós-traducionais, bem como erros experimentais significativos nas técnicas.

Em paralelo a essas “ômicas”, que lidam com produtos primários de genes, outras áreas encontram-se menos maduras. A metabolômica, por exemplo, auxilia na identificação da diversidade e na abundância de pequenas moléculas em células ou grupos celulares. Sua integração com o transcriptoma, feito especialmente em vegetais, são uma importante ferramenta de correlação entre elementos de informação – genes e transcritos – e elementos funcionais – metabólitos – em células. A integração de dados em múltiplas escalas vem sendo empregada também, de forma crescente, em diversas esferas, avaliando a resposta a estresse em diferentes organismos (integrando dados de transcriptômica, proteômica e metabolômica), até, no âmbito das ciências biomédicas, o aprimoramento da capacidade das técnicas disponíveis atualmente na diferenciação de casos de câncer em pacientes dos controles (integrando dados de proteômica, metabolômica e glicômica).

Como fechamento, fica claro que, por possuírem técnicas mais antigas, de uso mais acessível, e por tratarem com produtos primários de genes, a genômica e a transcriptômica possuem métodos de integração mais bem estabelecidos. No entanto, para o futuro, a investigação científica integrando diferentes “ômicas” além dessas, e avaliando a interação dessas diferentes moléculas com as células (através de interactomas) poderão ampliar a visão atual sobre o metabolismo celular, conferindo uma descrição mais quantitativa a esses modelos.