

Ambiente Computacional para Modelagem de Redes de Regulação Gênica

Raquel Lopes Costa¹, Marcelo Ribeiro Alves², Fábio Porto¹

¹Laboratório Nacional de Computação Científica (LNCC)
Av. Getúlio Vargas 333 - Petrópolis - RJ - Brasil

²Laboratório de Pesquisas sobre o Timo. Instituto Osvaldo Cruz (Fiocruz)
Av. Brasil 4365 - Rio de Janeiro - RJ - Brasil

{quelopes, mribalves}@gmail.com, fporto@lncc.br

Resumo. O recente desenvolvimento de ensaios moleculares para estudo do transcriptoma, associados a métodos estatísticos, matemáticos e computacionais, trouxe grandes avanços na compreensão de sistemas biológicos e, mais especificamente, na modelagem de redes de regulação gênica, GRNs (Genetic Regulatory Networks). As GRNs podem auxiliar na identificação de elementos-chave do processo de regulação gênica, contribuindo, por exemplo, para a compreensão de doenças complexas, bem como na escolha de alvos para o desenvolvimento de fármacos e testes com medicamentos. Apesar dos sistemas biológicos estarem em constante mudança e adaptação, grande parte dos dados experimentais de transcriptoma assumem estacionalidade e invariância no tempo. Medidas temporais provêm uma importante dimensão extra aos estudos de causalidade baseados na intuição que os eventos passados têm efeitos de causa nos eventos atuais e futuros e não vice-versa. Tais relações podem sugerir dependências entre eventos, servindo de base para inferência de ações regulatórias. De um lado, o processo de modelagem de GRNs envolve diferentes etapas, tais como a aquisição dos dados de transcriptoma, integração de informações disponíveis em bancos de dados biológicos, reconstrução do modelo regulatório, validação do modelo, visualização e análises da GRN obtida. Por outro lado, durante o processo de modelagem, são utilizados diversos sistemas de software, com diferentes pressupostos de organização e forma dos dados de entrada, fazendo com que todo esse processo seja além de trabalhoso e fragmentado, passível de erros. Apresentamos uma proposta de ambiente computacional integrativo por meio de workflow científico, proveniência e banco de dados com o propósito de analisar séries temporais de transcriptoma de eucariotos. O modelo de workflow científico está baseado no uso do Taverna como sistemas de gerenciador de workflows científico (Workflow Management Systems), o PROV-O (Prov Ontology) como modelo de proveniência e o Graph Database como modelo de banco de dados usado no armazenamento das informações biológicas. As informações biológicas são constituídas de dados provenientes dos repositórios de dados de transcriptoma como GEO (Gene Express Omnibus) integrado a outras fontes de conhecimento como Jaspár, Kegg e Reactome.