Mapping Omics datasets on KEGG Metabolic Pathways

Duarte Velho1, Andreia Salvador2 and João Sequeira3

1 University of Minho, Braga, Portugal

**Abstract.** The abstract should summarize the contents of the paper in short terms, i.e. 150-250 words.

**Keywords:** Bioinformatic-tools, Metabolic pathways mapping, Metagenomics, Metabolic networks.

1. Introduction
   1. Importance of Metagenomic Approaches

As abordagens metagenómicas são cada vez mais reconhecidas como uma base para compreender a ecologia e a evolução dos ecossistemas microbianos, bem como para o estudo e perceção da evolução de certas doenças.

As comunidades microbianas são compostas por bactérias, archaea, fungos, leveduras, eucariotas e vírus, que frequentemente coabitam em conjunto num mesmo habitat. Com esta dimensão e complexidade das comunidades microbianas, os dados ómicos que delas resultam, tornam a análise desses mesmos dados numa tarefa difícil (Sequeira et al.), sendo necessária uma abordagem diferente para o tratamento desses dados, pois a diversidade total da biosfera não pode ser estudada se não for enquadrada num meta-contexto adequado.

Até há pouco tempo, não existiam ferramentas para estudar sistematicamente a função global da comunidade e o ambiente a nível molecular, porque as comunidades microbianas complexas não são geralmente passíveis de estudo em laboratório. As abordagens metagenómicas permitiram os primeiros conhecimentos em grande escala sobre a função de comunidades microbianas complexas e são cada vez mais reconhecidas como uma base para compreender a ecologia e a evolução dos ecossistemas microbianos como redes genéticas e metabólicas.

* 1. Development of Bioinformatics Tools

Durante a última década, a análise de vias surgiu como uma ajuda preciosa para compreender os dados gerados por várias tecnologias "ómicas". Como resultado, foram desenvolvidas várias ferramentas de software robustas para apoiar a análise de vias para estudos genómicos e proteómicos (Xia & Wishart). Estas ferramentas combinam uma poderosa análise estatística com capacidades de visualização para ajudar os investigadores a identificar vias significativas envolvidas nas condições em estudo. As primeiras ferramentas de análise de vias foram normalmente concebidas para obter uma lista de genes diferencialmente expressos (Draghici et al) e comparar o número de genes diferencialmente expressos detectados em cada via de interesse com o número de genes que se esperava encontrar na via em causa apenas por acaso - um procedimento conhecido como análise de sobre-representação (Khatri et al.).

* 1. Current Challenges in Omics Data Analysis

A chave para associar esses conjuntos de dados em grande escala a valores práticos reside nas tecnologias de bioinformática, não só em termos de métodos computacionais, mas também em termos de bases de conhecimento. Deste modo, o novo panorama que se enfrenta exige o desenvolvimento de novas ferramentas de software (Filipo et al.) para caraterizar, reconstruir e modelar vias bioquímicas e reguladoras, como resultado da integração de funções em interações de comunidades microbianas.

As redes metabólicas constituem um quadro intuitivo e poderoso para a compreensão dos sistemas celulares. Até à data, praticamente todas as ferramentas de análise de vias foram concebidas e desenvolvidas para a análise de dados genómicos ou proteómicos, mas não de dados metabolómicos. (Xia & Wishart)

Embora estejam disponíveis vários recursos bioinformáticos para análise de meta-ómicas, muitos deles exigem conhecimentos computacionais significativos. As interfaces Web são mais fáceis de utilizar, mas na maior parte dos casos, têm dificuldade em lidar com grandes ficheiros de dados, como os obtidos em experiências de metagenómica, metatranscriptómica ou metaproteómica (Sequeira J. et al.).

A interpretação biológica destes grandes conjuntos de dados beneficia de ferramentas fáceis de utilizar, que convertem automaticamente os dados brutos em informações completas, idealmente com um pequeno contributo dos utilizadores e um tempo de análise reduzido.

As ferramentas bioinformáticas para mapear vias em análises meta-ómicas são instrumentos essenciais que procuram descodificar as complexas redes de genes e proteínas envolvidas no metabolismo, criando mapas que elucidam as vias no interior de uma célula, proporcionando clareza sobre a forma como os genes e as proteínas interagem e funcionam em vários processos biológicos, como o metabolismo, a sinalização e a regulação.

As vantagens da utilização destas ferramentas são bastante significativas, pois possuem a capacidade de processar e analisar conjuntos de dados volumosos, que outrora constituía uma pedra angular na compreensão de mecanismos biológicos complexos. Além disso, ajudam a identificar potenciais alvos para intervenção terapêutica e biomarcadores para estados de doença. Por outro lado, também existem desvantagens associadas a estas ferramentas, pois exigem frequentemente uma grande capacidade computacional e podem ser complexas, o que constitui um desafio para os utilizadores, que podem ter de ultrapassar uma curva de aprendizagem acentuada.

O desenvolvimento destas ferramentas envolve frequentemente linguagens de programação adequadas à análise de dados e à ciência computacional, como Python, R ou Java, e normalmente, incorporam pacotes ou bibliotecas especializadas em bioinformática como o Bioconductor para R e Biopython para Python.

No que diz respeito ao tipo de identificadores que estas ferramentas utilizam, baseiam-se em identificadores biológicos estabelecidos, como por exemplo os IDs Ensembl, Gene Ontology IDs ou KEGG Orthology Ids.

Muitas ferramentas de mapeamento de vias vêm com uma funcionalidade de identificação interativa, que permite aos utilizadores interagir com elementos específicos de uma via para uma exploração mais aprofundada ou para obter dados adicionais sobre genes ou proteínas específicas. Do mesmo modo, algumas ferramentas representam várias amostras em simultâneo. Isto permite fazer análises comparativas, como o contraste de estados doentes com estados saudáveis, revelando assim expressões diferenciais dentro das vias. Esta caraterística é fundamental para os investigadores que pretendem compreender mais pormenorizadamente as implicações biológicas dos seus dados experimentais.

A integração da taxonomia nestas ferramentas é também de grande importância, uma vez que permite aos investigadores identificar as relações e distinções evolutivas entre espécies, proporcionando uma compreensão mais profunda da diversidade genética e metabólica em diferentes organismos. Este facto pode ser particularmente esclarecedor em estudos que abrangem várias espécies ou na comparação de funções conservadas em diversos contextos biológicos.

Estas ferramentas permitem ainda a análise da expressão diferencial, fornecendo informações sobre quais os genes ou proteínas que são regulados positivamente ou negativamente em condições específicas, como o stress ambiental, estados de doença ou fases de desenvolvimento.

* 1. Discussion on methodologies and tools

Com a disponibilidade de diagramas de vias de redes metabólicas com curadoria manual e o volume cada vez maior de dados ómicos a visualizar, foram desenvolvidas várias ferramentas para facilitar a interpretação dos resultados da anotação funcional e representar os genes ou proteínas identificadas nas vias metabólicas. Por exemplo, o KEGG Mapper é uma coleção de ferramentas de mapeamento de vias, BRITE e MÓDULOS, apoiando a integração e interpretação de grandes conjuntos de dados genômicos e moleculares. Fornece uma base de dados abrangente que integra informações funcionais sistêmicas com dados genômicos e químicos, permitindo a visualização e análise de redes de interação molecular (Kanehisa et al.). Apesar de necessitar de compreensão dos sistemas biológicos e da estrutura organizacional do KEGG para uso eficaz, o KEGG Mapper oferece funcionalidades interativas, possibilitando que os usuários explorem informações detalhadas sobre interações e funções moleculares. Abrangendo uma vasta gama de organismos, desde procariontes até eucariontes, a ferramenta se adapta à especificidade dos mapas de vias e informações genômicas de cada organismo. Suas ferramentas, como "Search&Color Pathway", permitem mapear dados de expressão gênica, indicando regulações positivas ou negativas (Kanehisa M. et al.). Aceita listas de identificadores de genes, proteínas e pequenas moléculas, com formatos dependentes da ferramenta específica usada dentro do KEGG Mapper, e oferece saídas visuais e gráficos para vias e hierarquias funcionais em formatos de imagem variados para apresentações e publicações (Kanehisa & Sato).

KEGGCharter, reCOGnizer e UPIMAPI são ferramentas computacionais que se complementam na análise e interpretação de dados ómicos (Sequeira J. et al.). KEGGCharter foca na representação de resultados ómicos em vias metabólicas do KEGG, destacando a expressão gênica diferencial e a classificação taxonômica de enzimas, o que é particularmente útil em estudos metagenômicos. ReCOGnizer oferece anotação baseada em homologia de domínios de proteínas, integrando múltiplas bases de dados funcionais e fornece dados sobre nomes de domínios e números EC. UPIMAPI realiza anotações baseadas em homologia de sequências, coletando dados do UniProtKB e outros bancos de dados externos (Sequeira J. et al.). Essas ferramentas exigem formatos específicos de entrada de dados e conhecimento detalhado das bases de dados utilizadas, mas são úteis para a caracterização funcional abrangente e a classificação taxonômica com base em atribuições de enzimas ou domínios, permitindo uma análise rica e integrativa de conjuntos de dados biológicos complexos.

O MetPA, é uma ferramenta de análise de vias metabólicas desenhada para a visualização de dados metabolómicos (Xia & Wishart). Utiliza linguagens R e Java e pacotes como KEGGgraph, Graphviz e ImageMagick para renderizar redes metabólicas detalhadas, disponíveis numa interface web que suporta zoom sem perda de qualidade e manipulação dinâmica. A ferramenta é projetada para lidar com dados metabolômicos em larga escala, cobrindo 11 modelos de organismos comuns, incluindo humanos e *Escherichia coli*. Oferece análises estatísticas como t-test, ANOVA e regressão linear para investigar diferenças de concentração de metabolitos individuais em relação aos fenótipos representados numa tabela de concentração de compostos. As informações são visualizadas como uma rede em uma interface web, proporcionando um ambiente de análise de vias metabólicas robusto e de fácil uso (Xia & Wishart).

A DAVID é um software acessível pela web que integra anotações funcionais genômicas com resumos gráficos intuitivos, simplificando a anotação e sumarização rápida de dados conforme categorias compartilhadas (Dennis Jr. et al), como Gene Ontology e domínios de proteínas. Embora seu desempenho possa ser limitado pela conectividade com a internet e pelo estado atual dos bancos de dados integrados, a DAVID oferece uma interface interativa que facilita a navegação, a exploração detalhada de vias e a visualização de dados genômicos. Com linguagens de programação como Perl, Java e Visual Basic para automação, e suportando vários tipos de identificadores de genes ou proteínas, o DAVID ajuda na interpretação de dados expressos diferencialmente, facilitando a transição da coleta de dados para o significado biológico (Dennis Jr. et al). Os resultados são fornecidos em tabelas HTML para revisão fácil e análise adicional, incluindo links para fontes de anotação originais para exploração aprofundada.

Outra ferramenta, o CellDesigner 3.5, é uma ferramenta de modelagem para redes bioquímicas e gene-regulatórias, utilizando notação gráfica padronizada e SBML (Systems Biology Markup Language) para facilitar a troca de modelos. Oferece representação gráfica abrangente e interoperabilidade SBML, integra-se com softwares de simulação e análise e possui uma interface intuitiva. Requer conhecimento de SBML e sua notação gráfica para uso efetivo e, como é um software de computador, pode ter limitações colaborativas em comparação com ferramentas baseadas na web. É programado em Java integra-se com pacotes habilitados para o Systems Biology Workbench (SBW) para análise e simulação (Funahashi et al.). O CellDesigner suporta vários identificadores biológicos e não se aplica diretamente à representação de amostras de dados experimentais. Pode incluir expressões diferenciais de genes nos seus modelos. Exporta modelos em formato SBML e formatos gráficos como PNG e SVG para inclusão em publicações (Funahashi et al.).

O GenMAPP versão 2, é uma ferramenta que permite aos usuários visualizar e analisar dados de escala genômica, como dados de microarray, em vias biológicas (Salomonis et al). Implementada em Visual Basic 6.0, oferece suporte a múltiplas anotações de genes e espécies e permite a criação personalizada de bancos de dados para um número potencialmente ilimitado de espécies. Esta ferramenta integra processos de ETL para extração de dados de recursos públicos como Ensembl, Entrez Gene e Affymetrix, suportando uma ampla gama de identificadores, como IDs de genes Ensembl, IDs UniProt, IDs Entrez Gene e IDs de conjunto de sondas Affymetrix. Uma característica única é a coloração dinâmica dos genes nas vias com base em critérios definidos pelo usuário, fornecendo acesso a anotações de genes e links externos de bancos de dados por meio de "backpages" HTML (Salomonis et al.). A ferramenta foi projetada para lidar com conjuntos de dados de escala de genoma e interpretar mudanças a nível de via numa ampla gama de organismos, incluindo, mas não se limitando, a humanos, ratos, leveduras, minhocas entre outros, e permite a visualização de dados de expressão gênica diferencial e outros tipos de dados biológicos (por exemplo, SNP, variações de splicing) no contexto das vias. Usa o formato Microsoft Jet para conjuntos de dados experimentais (.gex), bancos de dados de genes (.gdb) e vias (.mapp), e permite a exportação para HTML, e para formatos de exportação gráfica para visualização, como PNG e SVG.

O KGML-ED, permite a visualização dinâmica, navegação interativa e edição de diagramas de vias do KEGG. Esta ferramenta suporta a exploração interativa e dinâmica das vias, permitindo a edição e criação de novas vias, facilitando a integração de dados específicos do usuário (Klukas & Schreiber). Embora requeira familiaridade com o formato KGML e a estrutura subjacente das vias do KEGG, o KGML-ED oferece técnicas de visualização semi-estáticas e dinâmicas para análises de vias melhoradas e específicas ao usuário. Implementado em Java e utilizando o toolkit de visualização de grafos Gravisto, o KGML-ED suporta identificadores KGML para genes, permitindo mapeamento e anotação detalhados dentro do framework KEGG. Possui capacidades interativas para navegar e modificar diagramas de vias, com opções de personalização conduzida pelo usuário. O foco do KGML-ED é a manipulação e visualização de diagramas de vias em vez da representação de amostras de dados experimentais, e permite que os usuários colorem componentes das vias com base em resultados de análise externa (Klukas & Schreiber).

MetaCore™ é uma plataforma computacional baseada na web para biologia de sistemas, que analisa os dados moleculares de alto rendimento e integra-os para visualização e análise estatística no contexto de redes, vias e processos biológicos. Com acesso pago e complexidade para novos usuários, exige dados de alta qualidade e usa Perl, HTML/JavaScript e Flash Player Plug-in (Ekins S. et al.). Aceita identificadores como LocusLink, SwissProt, RefSeq e Unigene. A interatividade é possível e depende do conjunto de dados específicos utilizados pelo usuário, focando em *Homo sapiens* e mamíferos. Suporta análise de expressão gênica diferencial e aceita vários formatos de entrada e saída de dados, incluindo XML, JSON, CSV, TSV, PDF, PNG, JPEG, HTML, XLS e XLSX (Ekins S. et al.). Da mesma forma, o MetaDrug™ foi concebido para prever os principais metabolitos xenobióticos em humanos, juntamente com as suas interacções enzimáticas e propriedades ADME/Tox. Utiliza algoritmos para dar prioridade aos metabolitos e integra modelos de relação quantitativa estrutura-atividade (QSAR) e fornece a visualização de dados previstos e redes de assinatura de genes. Porém, requer a integração com o Pipeline Pilot para desenvolvimentos futuros (Ekins S. et al.). O MetaDrug™ suporta vários identificadores moleculares e bioquímicos e, oferece capacidades interactivas. Ambas as ferramentas suportam a integração com outras bases de dados e ferramentas computacionais, o que facilita a análise complexa de sistemas biológicos e permite aos usuários criar redes e vias que são relevantes para seus dados experimentais e objetivos de pesquisa (Ekins S. et al.).

Pathway Tools é um software de bioinformática projetado simulação e visualização de coleções integradas de dados genômicos, de vias metabólicas e regulatórias. Oferece suporte a uma vasta gama de tipos de dados em bioinformática e permite reconstruções metabólicas e previsões, além da visualização de interações regulatórias e análise de dados ômicos (Karp. Et al., 2015). Ao longo de suas versões, Pathway Tools aprimorou as suas capacidades e expandiu seu banco de dados, passando de 800 bases de dados de genoma/vias (PGDBs) (Karp. Et al., 2009) para mais de 20.000 (Karp. Et al., 2019), e abrangendo uma taxonomia diversa, incluindo *E. coli*, *B. subtilis* e *P. difficile*. Enquanto que a aprendizagem da ferramenta pode ser íngreme para novos usuários e algumas funcionalidades estão restritas aos modos software ou web, a ferramenta mantém a interatividade e a compatibilidade com diferentes formatos de entrada e saída de dados, como GenBank, SBML e BioPAX, facilitados em versões recentes pelo recurso SmartTables. A integração de dados inferidos por máquina e curados manualmente oferece uma representação fidedigna e rica dos dados para os usuários (Karp. Et al., 2015).

A ferramenta iPath é uma ferramenta web para visualização e análise de vias celulares que fornece mapas interativos para o metabolismo central, biossíntese de metabólitos secundários e vias regulatórias (Yamada T. et al.). Oferece uma interface interativa e amigável para a exploração de uma ampla gama de dados de vias metabólicas, com a capacidade de mapear uma variedade de identificadores, como KEGG, e customizar mapas com dados como expressão diferencial. Ao longo das atualizações, iPath expandiu seu conjunto de dados e funcionalidades, incluindo novos módulos de vias KEGG, reações e espécies (Darzi Y. et al.). Apesar da melhoria na interface e na experiência do usuário, a ferramenta continua dependente do acesso à internet, o que pode limitar a acessibilidade para usuários com largura de banda limitada. A versão mais recente suporta análise interativa com detalhes em nós e arestas, refletindo avanços no motor de exibição e na gestão de dados (Darzi Y. et al.). As opções de exportação foram aprimoradas, permitindo a inclusão de mapas personalizados em publicações ou outros documentos.

Da mesma forma, PathVisio é uma ferramenta de software para visualizar, editar e analisar vias biológicas, com integração a bases de dados biológicos para enriquecer as informações das vias (P van Iersel et al.). É multiplataforma devido à sua implementação em Java e integra-se com GenMAPP e MAPPFinder para análise estatística, utilizando GPML, um formato XML para armazenamento de vias que é extensível e compatível com versões anteriores. Embora exija entrada manual de dados, o que pode ser demorado, PathVisio é reforçado pela sua capacidade de interagir com outras ferramentas científicas, como Cytoscape e Eu.Gene, aumentando sua aplicabilidade em fluxos de trabalho de bioinformática. Suporta uma vasta gama de identificadores de entidades biológicas e permite a análise visual de dados de alto rendimento (P van Iersel et al.). A ferramenta oferece exportação para vários formatos gráficos, tornando-a útil em várias áreas de pesquisa biológica.

Pathview é um conjunto de ferramentas para integração e visualização de dados baseados em vias metabólicas, mapeando dados do usuário em gráficos de vias relevantes para facilitar a análise e interpretação. Como parte do projeto Bioconductor, ele integra-se com diversos pacotes R para análise de dados abrangente e suporta uma ampla gama de identificadores para genes/proteínas e compostos/metabólitos (Luo W. et al.). Embora dependa de bancos de dados externos, como o KEGG, para informações de vias e exija conhecimento em R para uso eficaz, Pathview compensa com funcionalidade completa, acessibilidade e facilidade de automação. É interativo, lidando com conjuntos de dados de diferentes escalas e complexidade, e compatível com mais de 2000 espécies. Oferece suporte à visualização de expressão gênica diferencial e outros tipos de dados quantitativos, aceita dados em formatos de matriz ou vetor e gera visualizações de vias nas visualizações nativas do KEGG e Graphviz, com opções de personalização para nós e atributos de arestas.

A ferramenta Pathway Tools é uma aplicação baseada em Java (Rahman S. et al.) e oferece uma representação visual da rede bioquímica de um organismo, gerada automaticamente a partir de um Pathway/Genome Database (PGDB). O layout é automatizado, suportando a análise de dados em larga escala e facilitando a interpretação biológica de dados ômicos (Paley S. & Karp P). A ferramenta é projetada para trabalhar com conjuntos de dados de organismos inteiros, integrando-se com o Omics Viewer para sobrepor dados de expressão gênica e outros dados quantitativos. No entanto, exige uma curva de aprendizado para aproveitar totalmente suas capacidades e depende da completude do PGDB subjacente. Os aprimoramentos ao longo das versões incluem visualizações interativas na web e desktop, suporte a zoom semântico, geração de pôsteres e visualizações coloridas e animadas para dados de séries temporais (Paley S. & Karp P). Aceita arquivos delimitados por tabulações fornecidos pelo usuário, aumentando sua interatividade e aplicabilidade na visualização de redes celulares e na análise de dados de alto rendimento.

A Reactome Knowledgebase, é uma base de dados online que detalha processos celulares como transdução de sinal e replicação de DNA em uma rede de transformações moleculares ordenada. É uma ferramenta tanto para arquivar processos biológicos quanto para descobrir relações funcionais em conjuntos de dados (Fabregat et al.). Oferece uma cobertura abrangente de processos biológicos humanos, com Diagramas de Alto Nível Aprimorados para visualização aprimorada e ferramentas de análise de dados em memória para acessibilidade melhorada. Requer uma curva de aprendizado devido à complexidade dos dados e foco em processos biológicos humanos. A Reactome, é uma base de dados curada manualmente que fornece uma interface web intuitiva para o conhecimento de vias e uma suíte de ferramentas de análise de dados. Permite visualização interativa de dados ômicos sobre diagramas de vias e oferece exportação de diagramas de vias para análise e personalização adicionais pelo usuário (Haw R. et al.). Ambas versões suportam uma variedade de identificadores e são capazes de lidar com grandes conjuntos de dados.

**Table 1.** The available analytical tools and web servers designed for mapping pathway data, including their functions, advantages, disadvantages, supported languages, packages, types of identifiers, interactivity, sample representation, taxonomy considerations, differential expression analysis capabilities, and input/output formats.

[Table]

A partir da tabela 1, é possível inferir que

* 1. Context and importance/relevance of the project

Na vanguarda da bioinformática, o projeto em análise insere-se no campo emergente das abordagens metagenómicas, uma área reconhecida por seu potencial em decifrar a complexidade e a dinâmica dos ecossistemas microbianos, bem como sua relevância no estudo da evolução de doenças. Neste contexto, a capacidade de mapear conjuntos de dados ómicos em vias metabólicas representa uma ferramenta indispensável na interpretação de interações complexas em redes genéticas e metabólicas. A metagenómica abre portas para uma compreensão mais abrangente da ecologia microbiana, permitindo avanços significativos na ciência e na medicina, particularmente na identificação de novos biomarcadores e alvos terapêuticos.

Definição dos Objetivos do Projeto:

Expansão das representações gráficas do KEGGCharter: Este projeto visa desenvolver representações interativas sobre as visualizações de vias metabólicas existentes. O objetivo é enriquecer os mapas metabólicos com informações multi-nível de expressão gênica, proporcionando uma análise mais detalhada e integrada das vias metabólicas.

Aperfeiçoamento de análises comparativas: A segunda meta do projeto é aprimorar a capacidade das ferramentas de bioinformática para realizar comparações entre diferentes estados, como condições saudáveis e patológicas, utilizando dados meta-ómicos. Isso permitirá revelar expressões diferenciais críticas e compreender melhor os mecanismos subjacentes das doenças.

Integração taxonômica e funcional: O projeto também busca integrar a taxonomia nas ferramentas existentes, melhorando a capacidade de diferenciar e compreender a diversidade genética e metabólica entre organismos. Tal integração será crucial para estudos envolvendo várias espécies ou para a comparação de funções conservadas em diferentes contextos biológicos.

Automatização do mapeamento de vias: Por fim, o projeto almeja a automatização do processo de mapeamento de vias, reduzindo o tempo de análise e a intervenção manual necessária. A implementação de algoritmos avançados e interfaces intuitivas facilitará a tradução de dados brutos em informações compreensíveis, acelerando o progresso das pesquisas em bioinformática.

References

1. Darzi, Y., Letunic, I., Bork, P., and Yamada, T.: iPath3.0: interactive pathways explorer v3. Nucleic Acids Research W510-W513 (2018). doi: 10.1093/nar/gky299. Published online 30 April 2018.
2. Draghici, S., Khatri, P., Martins, R.P., Ostermeier, G.C., and Krawetz, S.A.: Global functional profiling of gene expression. Genomics 81, 98–104 (2003). doi:10.1016/S0888-7543(02)00021-6.
3. De Filippo, C., Ramazzotti, M., Fontana, P., Cavalieri, D.: Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. Briefings in Bioinformatics 13(6), 696–710 (2012). doi:10.1093/bib/bbs070
4. Dennis Jr, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., & Lempicki, R.A.: DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology 4(9), R60 (2003).
5. Ekins S., Nikolsky Y., Bugrim A., Kirillov E., Nikolskaya T. Pathway mapping tools for analysis of high content data. In: Taylor D. L., Haskins J. R., Giuliano K. A. (eds.) High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery, vol. 356, pp. 319–350. Humana Press, Totowa, NJ (2007).
6. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Duenas Roca, C., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D’Eustachio, P.: The Reactome Pathway Knowledgebase. Nucleic Acids Research 46(D1), D649–D655 (2018). DOI: 10.1093/nar/gkx1132.
7. Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., Kitano, H.: CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. Proceedings of the IEEE 96(8), 1254–1265 (2008). Digital Object Identifier: 10.1109/JPROC.2008.925458
8. Haw, R., Hermjakob, H., D'Eustachio, P., & Stein, L. (2011). Reactome pathway analysis to enrich biological discovery in proteomics data sets. Proteomics, 11(18), 3598–3613.
9. Jin, L., Zuo, X.-Y., Su, W.-Y., Zhao, X.-L., Yuan, M.-Q., Han, L.-Z., Zhao, X., Chen, Y.-D., Rao, S.-Q.: Pathway-based Analysis Tools for Complex Diseases: A Review. Genomics Proteomics Bioinformatics 12(5), 210–220 (2014).
10. Kanehisa, M., Sato, Y.: KEGG Mapper for inferring cellular functions from protein sequences. Protein Science 29, 28–35 (2020). DOI: 10.1002/pro.3711
11. Kanehisa, M., Sato, Y., Kawashima, M.: KEGG mapping tools for uncovering hidden features in biological data. Protein Science 31(1), 47–53 (2022). DOI: 10.1002/pro.4172
12. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M.: KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Research 40(D1), D109–D114 (2012). DOI: 10.1093/nar/gkr988.
13. Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., Caspi, R.: Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Briefings in Bioinformatics 11(1), 40–79 (2009). Advance Access published on 2 December 2009, doi:10.1093/bib/bbp043.
14. Karp, P.D., Latendresse, M., Paley, S.M., Krummenacker, M., Ong, Q.D., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P., Spaulding, A., Fulcher, C., Keseler, I.M., Caspi, R.: Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. Briefings in Bioinformatics 17(5), 877-890 (2016). DOI: 10.1093/bib/bbv079. Advance Access Publication Date: 10 October 2015. Published by Oxford University Press.
15. Karp, P.D., Midford, P.E., Billington, R., Kothari, A., Krummenacker, M., Latendresse, M., Ong, W.K., Subhraveti, P., Caspi, R., Fulcher, C., Keseler, I.M., & Paley, S.M.: Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Briefings in Bioinformatics 22(1), 109–126 (2021). doi: 10.1093/bib/bbz104. Advance Access Publication Date: 8 December 2019. Review article.
16. Khatri, P., Draghici, S., Ostermeier, G.C., Krawetz, S.A.: Profiling Gene Expression Using Onto-Express. Genomics 79(2), 266–270 (2002). DOI:10.1006/geno.2002.6698. Available online at IDEAL
17. Klukas, C., Schreiber, F.: Dynamic exploration and editing of KEGG pathway diagrams. Bioinformatics 23(3), 344–350 (2007). doi:10.1093/bioinformatics/btl611.
18. Luo, W., Brouwer, C.: Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 29(14), 1830–1831 (2013). doi:10.1093/bioinformatics/btt285
19. van Iersel, M.P., Kelder, T., Pico, A.R., Hanspers, K., Coort, S., Conklin, B.R., & Evelo, C.: Presenting and exploring biological pathways with PathVisio. BMC Bioinformatics 9, 399 (2008). DOI:10.1186/1471-2105-9-399.
20. Paley, S.M., Karp, P.D.: The Pathway Tools cellular overview diagram and Omics Viewer. Nucleic Acids Research 34(13), 3771–3778 (2006). doi:10.1093/nar/gkl334.
21. Rahman, S.A., Advani, P., Schunk, R., Schrader, R., Schomburg, D.: Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). Bioinformatics 21(7), 1189–1193 (2005). doi:10.1093/bioinformatics/bti116.
22. Salomonis, N., Hanspers, K., Zambon, A.C., Vranizan, K., Lawlor, S.C., Dahlquist, K.D., Doniger, S.W., Stuart, J., Conklin, B.R., Pico, A.R.: GenMAPP 2: new features and resources for pathway analysis. BMC Bioinformatics 8, 217 (2007).
23. Sequeira, J.C., Rocha, M., Alves, M.M., Salvador, A.F.: UPIMAPI reCOGnizer and KEGGCharter: Bioinformatics tools for functional annotation and visualization of (meta)-omics datasets. Comput. Struct. Biotechnol. J. 20, 1798–1810 (2022).
24. Werner, T.: Bioinformatics applications for pathway analysis of microarray data. In: Current Opinion in Biotechnology 19, pp. 50–54 (2008). DOI 10.1016/j.copbio.2007.11.005
25. Xia, J., & Wishart, D.S.: MetPA: a web-based metabolomics tool for pathway analysis and visualization. In: Bioinformatics Applications Note, Vol. 26, No. 18, pp. 2342–2344. Oxford University Press (2010). DOI:10.1093/bioinformatics/btq418
26. Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., Bork, P.: iPath2.0: interactive pathway explorer. Nucleic Acids Research 39(suppl\_2), W412–W415 (2011). doi:10.1093/nar/gkr313. Published online 5 May 2011.