

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 23/07/2020.

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS BIOLÓGICAS - GENÉTICA

Giordano Bruno Sanches Seco

**Análise Comparativa de Métodos de Determinação de
Vias Biológicas Alteradas em Dados Toxicogenômicos
de estudos *in vitro* e *in vivo***

Botucatu, Junho de 2018

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
INSTITUTO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS BIOLÓGICAS - GENÉTICA

Giordano Bruno Sanches Seco

Análise Comparativa de Métodos de Determinação de Vias Biológicas Alteradas em Dados Toxicogenômicos de estudos *in vitro* e *in vivo*

Dissertação apresentada ao Instituto de Biociências, Campus de Botucatu, UNESP, em preenchimento dos requisitos para a obtenção do título de Mestre no Programa de Pós-Graduação em Ciências Biológicas - Genética.

Orientador: Prof. Dr. José Luiz Rybarczyk Filho

Botucatu, Junho de 2018.

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Seco, Giordano Bruno Sanches.

Análise comparativa de métodos de determinação de vias biológicas alteradas em dados toxicogenômicos de estudos in vitro e in vivo / Giordano Bruno Sanches Seco. - Botucatu, 2018

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: José Luiz Rybarczyk Filho

Capes: 20204000

1. Agentes antineoplásicos. 2. Bioinformática. 3. Microarranjos de DNA. 4. Análise de microarranjo. 5. Toxicogenética.

Palavras-chave: Antineoplásicos; Bioinformática; Microarranjo; Toxicogenômica.

Agradecimentos

Agradeço aos processos CNPq 134467/2016-7, CNPq 458810/2013-4 e CNPq 473789/2013-2 cujo auxílio financeiro possibilitou a execução desse trabalho.

Agradeço do Prof. Dr. José Luiz Rybarczyk Filho, cuja a orientação foi de fundamental importância para solucionar os diversos desafios que apareceram durante a realização deste trabalho.

Agradeço a Dr(a) Agnes Alessandra Sekijima Takeda, cujo suporte fornecido na interpretação da parte biológica dos dados e na montagem de figuras permitiu a confecção dos resultados e discussões.

Agradeço aos colegas de laboratório André Luiz Molan, Carlos Alberto Oliveira de Biagi Junior e José Rafael Pilan cujo suporte tanto em nível técnico quanto pessoal foi essencial para execução das tarefas necessárias para a confecção deste trabalho

Agradeço a todas as demais pessoas (professores, funcionários e amigos) que de alguma forma, direta ou não, tenham contribuído para a realização deste trabalho.

Agradeço imensamente a minha família, em especial ao meu pai Adalberto Alexandre Godoy Sanches Seco e mãe Márcia Regina Ogava Seco por todo suporte pessoal fornecido durante a realização deste trabalho.

Por fim agradeço a Deus pela vida.

Resumo

A toxicogenômica é uma área de estudo que se utiliza de dados provindos de tecnologias das ciências ômicas e de metodologias de análise de dados da bioinformática para caracterizar perfis biológicos de compostos a fim realizar um estudo sobre seu comportamento no sistema biológico. Um dos problemas na análise de dados de expressão gênica é que existem diversas metodologias disponíveis, cada uma aborda os dados brutos de maneira diferente e consequentemente fornecem resultados diferentes. Porém estes resultados não estão necessariamente errados, por essa razão abordar dados toxicogenômicos de diversas maneiras pode ajudar pesquisadores a tecer hipóteses sobre os mecanismos das drogas. Isso pode implicar em redução de custos na fase pré-clínica do desenvolvimento de medicamentos e numa melhor qualidade de tratamento para os pacientes. A abordagem mais comum na análise de dados de expressão gênica é o enriquecimento funcional dos genes que se mostrem diferencialmente expressos. Entretanto é possível que informações relevantes possam ser obtidas utilizando métodos com outras considerações. Para avaliar esta hipótese o presente trabalho tem por objetivo analisar dados de expressão gênica disponíveis em bases de dados on-line referentes a 3 drogas antineoplásicas testadas em fígado de *Rattus norvegicus* (estudos *in vivo* e *in vitro*) e *Homo sapiens* (estudo *in vivo* apenas) para avaliação de ontologias alteradas em cada droga. Duas metodologias de bioinformática foram utilizadas para análise desses dados. A tradicional análise de genes diferencialmente expressos (EFDEGs) e a análise de grupos de genes funcionalmente associados determinados pela entropia de shannon (AGFAGs). Para cada um dos três métodos foram testados três normalizações: RMA, MAS5, GCRMA. Os resultados sugerem que os 3 tipos de estudos tem pouca similaridade entre si e a pouca que existe se limita, aos principais mecanismos das drogas antineoplásicas como ciclo celular, apoptose, dano e reparo de DNA. Portanto não se pode afirmar ainda que um tipo de estudo possa substituir o outro. Também determinamos que resultados diferentes fornecidos pelas 2 metodologias para um mesmo caso podem não estar necessariamente errados e algumas vezes podem ser complementares. A meta-análise desse trabalho não só mostrou similaridade e complementariedade entre os métodos como também encontrou lncRNAs potencialmente responsáveis com a regulação pós-transcricional em hepatócitos humanos. Isso evidencia a importância da prática de se utilizar dados previamente obtidos em análises exploratórias para confecção de hipóteses. Assim o uso de ambas as metodologias pode ter valor como ferramenta para análises exploratórias em dados biológicos disponíveis.

Abstract

Toxicogenomics is an area that uses data generated by *omics technologies* and *bioinformatics* to characterize biological profiles of chemical compounds in order to analyze the biological system's behavior in response to said compounds. A common problem in the analysis of gene expression data is the large amount of methodologies available, each one approaches the raw data differently and consequently provide different results. But these results are not necessarily wrong, for that reason approaching toxicogenomic data in several ways may help researchers to hypothesize about a drug's mechanisms. This may imply cost reduction in the pre-clinical phase of drug development and in a better quality of treatment for patients. The most common approach in the analysis of gene expression data is the functional enrichment of differentially expressed genes (EFDEGs). However, it is possible that relevant information can be obtained using methods with other considerations. To evaluate this hypothesis this work aims to analyze gene expression data available in the OPEN TG-GATEs online database about 3 antineoplastic drugs tested in the liver of *Rattus norvergicus* (*in vivo* and *in vitro*) and *Homo sapiens* (*in vivo* only) for the evaluation of altered ontologies in each drug. Two bioinformatics methodologies were used to analyze these data. The traditional analysis of differentially expressed genes (EFDEGs) and the analysis of groups of functionally associated genes determined by shannon entropy (AGFAGs). For each of the two methods, three normalizations were tested: RMA, MAS5, GCRMA. Results suggest that the 3 types of studies have little similarity to each other and the few that exist are limited to the main mechanisms of antineoplastic drugs such as cell cycle, apoptosis and DNA damage and repair. Therefore, it still cannot be said that one type of study can replace the other. We also found out that the different results provided by the two methodologies for the same case may not necessarily be wrong and sometimes may be complementary. The meta-analysis done in this work not only showed similarity and complementarity between methods but also found differentially expressed lncRNAs potentially responsible for post-transcriptional regulation in human hepatocytes. This evidences the importance of using previously obtained data in exploratory analyzes to make hypotheses. Thus the use of both methodologies may have value as a tool for exploratory analyzes.

Lista de Figuras

1.1	Relação hierárquica das quatro frentes das ciências ômicas	p. 15
1.2	<i>Workflow</i> descrevendo um experimento geral de Toxicogenômica	p. 18
1.3	Ilustração das etapas de um experimento de microarranjo de dois canais (duas cores)	p. 29
1.4	<i>Workflow</i> de um experimento de toxicogenômica envolvendo o uso de dados de expressão gênica	p. 30
1.5	Cronologia dos experimentos <i>in vivo</i> no <i>Rattus norvegicus</i> . Adaptado de (TAKASHIMA et al., 2006)	p. 38
3.1	<i>Workflow</i> da pesquisa	p. 50
3.2	Exemplo de uma segmento da ontologia BP	p. 52
4.1	Boxplots das 3 normalizações para os chips da ciclofosfamida e estudo <i>Homo sapiens in vitro</i>	p. 62
4.2	Distribuição d GDEs das 3 drogas no método EFDEGs	p. 64
4.3	Gráficos de linhas para ilustrar os GDEs das 3 drogas no método EFDEGs com <i>cutoff</i> de 0,001	p. 65
4.4	Distribuição de BPs das 3 drogas no método EFDEGs	p. 68
4.5	Gráficos de linhas para ilustrar os GDEs das 3 drogas no método EFDEGs com <i>cutoff</i> de 0,001	p. 69
4.6	Distribuição de GDEs da ciclofosfamida no método EFDEGs	p. 71
4.7	Gráficos de linhas para ilustrar os GDEs da ciclofosfamida no método EF- DEGs com <i>cutoff</i> de 0,001	p. 72
4.8	Distribuição de BPs da ciclofosfamida no método EFDEGs	p. 74

4.9	Gráficos de linhas para ilustrar os BPs da ciclofosfamida no método EF-DEGs com <i>cutoff</i> de 0,001	p. 75
4.10	Expressão dos genes anotados na ontologia GO:0002933	p. 78
4.11	Expressão dos genes anotados na ontologia GO:0009822	p. 79
4.12	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Homo sapiens in vitro</i> na ciclofosfamida com uso do STRING	p. 82
4.13	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Rattus norvegicus in vitro</i> na ciclofosfamida com uso do STRING	p. 86
4.14	Rede PPI derivada do universo de GDEs do estudo <i>Rattus norvegicus in vitro</i> na ciclofosfamida com uso do STRING	p. 89
4.15	Gráficos de linhas para ilustrar os GDEs do etopósido no método EFDEGs com <i>cutoff</i> de 0,001	p. 90
4.16	Distribuição de GDEs do etopósido no método EFDEGs	p. 91
4.17	Distribuição de BPs do etopósido no método EFDEGs	p. 93
4.18	Gráficos de linhas para ilustrar os BPs do etopósido no método EFDEGs com <i>cutoff</i> de 0,001	p. 94
4.19	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Homo sapiens in vitro</i> no etopósido com uso do STRING	p. 98
4.20	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Homo sapiens in vitro</i> no etopósido com uso do STRING	p. 100
4.21	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Rattus norvegicus in vivo</i> no etopósido com uso do STRING	p. 102
4.22	Distribuição de GDEs da lomustina no método EFDEGs	p. 104
4.23	Gráficos de linhas para ilustrar os GDEs da lomustina no método EFDEGs com <i>cutoff</i> de 0,001	p. 105
4.24	Distribuição de BPs da lomustina no método EFDEGs	p. 106

4.25	Gráficos de linhas para ilustrar os BPs da lomustina no método EFDEGs com <i>cutoff</i> de 0,001	p. 107
4.26	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Homo sapiens in vitro</i> na lomustina com uso do STRING	p. 109
4.27	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Rattus norvegicus in vitro</i> na lomustina com uso do STRING	p. 111
4.28	Rede PPI derivada dos GDEs em comum aos casos não nulos do estudo <i>Rattus norvegicus in vivo</i> na lomustina com uso do STRING	p. 113
4.29	Distribuição de BPs das 3 drogas no método AGFAGs (atividade)	p. 115
4.30	Gráficos de linhas para ilustrar os BPs das 3 drogas no método AGFAGs (atividade) com <i>cutoff</i> de 0,001	p. 116
4.31	Distribuição de BPs das 3 drogas no método AGFAGs (diversidade)	p. 118
4.32	Gráficos de linhas para ilustrar os BPs das 3 drogas no método AGFAGs (diversidade) com <i>cutoff</i> de 0,001	p. 119
4.33	Distribuição de BPs da ciclofosfamida drogas no método AGFAGs (atividade)p.	121
4.34	Gráficos de linhas para ilustrar os BPs da ciclofosfamida no método AGFAGs (atividade) com <i>cutoff</i> de 0,001	p. 122
4.35	Distribuição de BPs da ciclofosfamida drogas no método AGFAGs (diversidade)	p. 123
4.36	Gráficos de linhas para ilustrar os BPs da ciclofosfamida no método AGFAGs (diversidade) com <i>cutoff</i> de 0,001	p. 124
4.37	Distribuição de BPs do etopósido no método AGFAGs (atividade)	p. 132
4.38	Gráficos de linhas para ilustrar os BPs do etopósido no método AGFAGs (atividade) com <i>cutoff</i> de 0,001	p. 133
4.39	Distribuição de BPs do etopósido no método AGFAGs (diversidade)	p. 134
4.40	Gráficos de linhas para ilustrar os BPs do etopósido no método AGFAGs (diversidade) com <i>cutoff</i> de 0,001	p. 135

4.41	Distribuição de BPs da lomustina no método AGFAGs (atividade)	p. 142
4.42	Gráficos de linhas para ilustrar os BPs da lomustina no método AGFAGs (atividade) com <i>cutoff</i> de 0,001	p. 143
4.43	Distribuição de BPs da lomustina no método AGFAGs (diversidade)	p. 144
4.44	Gráficos de linhas para ilustrar os BPs da lomustina no método AGFAGs (diversidade) com <i>cutoff</i> de 0,001	p. 145

Lista de Tabelas

1.1	Resumo do design experimental dos estudos <i>in vivo</i>	p. 40
1.2	Resumo do Design Experimental dos estudos <i>in vitro</i>	p. 41
1.3	Tabela de relação entre drogas antineoplásicas e estudos realizados no OPEN TG-GATEs	p. 44
1.4	Tabela com o protocolo utilizado nos estudos <i>in vivo</i> no OPEN TG-GATEs	p. 44
1.5	Tabela com o protocolo utilizado nos estudos <i>in vitro</i> no OPEN TG-GATEs	p. 45
3.1	Métodos de normalização de dados de microarranjos e suas características .	p. 56
4.1	Processos chave encontrados nos estudos/drogas no EFDEGs	p. 152
4.2	Processos chave encontrados nos estudos/drogas no AGFAGs (atividade) .	p. 152
4.3	Processos chave encontrados nos estudos/drogas no AGFAGs (diversidade)	p. 153

Sumário

Resumo	p. 2
Abstract	p. 3
1 Introdução	p. 12
1.1 Advento das Tecnologias de Alto Desempenho: As ciências Ômicas	p. 12
1.2 Toxicogenômica	p. 16
1.2.1 Definição	p. 16
1.2.2 O Potencial da Toxicogenômica: <i>Design</i> Experimental e Aplicações Comuns	p. 17
1.2.3 Redes Biológicas: A Toxicogenômica nos Dias Atuais	p. 24
1.3 Microarranjo (<i>microarray</i>)	p. 28
1.3.1 Princípios da Técnica	p. 28
1.3.2 <i>Design</i> do Experimento de Expressão Gênica e Análise de Dados: Aplicações, Exemplos, Limitações e Perspectivas	p. 30
1.4 O Projeto Toxicogenômico Japonês Open TG-GATEs e os Dados Seleccionados	p. 36
1.4.1 OPEN TG-GATEs: <i>Design</i> Experimental	p. 37
1.4.1.1 Estudos <i>In vivo</i>	p. 37
1.4.1.2 Estudos <i>In vitro</i>	p. 40
1.4.2 Sobre as Drogas	p. 41
1.4.3 Antineoplásicos e o OPEN TG-GATEs	p. 43

1.5	Justificativa	p. 45
2	Objetivos	p. 48
2.1	Objetivos Principais	p. 48
2.2	Objetivos Específicos	p. 48
3	Materiais e Metodos	p. 49
3.1	<i>Workflow</i>	p. 49
3.2	<i>Gene Ontology (GO)</i>	p. 51
3.3	STRING: Um Banco de Dados de Interações Proteína-Proteína	p. 53
3.4	Normalização dos dados	p. 54
3.4.1	<i>Robust Multi-array Average (RMA)</i>	p. 56
3.4.2	<i>GC Robust Multi-array (GCRMA)</i>	p. 57
3.4.3	<i>Microarray Suite 5 (MAS5)</i>	p. 57
3.5	Enriquecimento Funcional de Genes Diferencialmente Expressos (EFDEGs)	p. 58
3.6	Análise de Grupos de Genes Funcionalmente Associados (AGFAGs)	p. 59
4	Resultados Discussão	p. 61
4.1	Sobre as análises e normalizações	p. 61
4.2	Método EFDEGs	p. 63
4.2.1	Aspectos Gerais: Análise Numérica e Descritiva	p. 63
4.2.2	Ciclofosfamida	p. 70
4.2.3	Etopósido	p. 90
4.2.4	Lomustina	p. 103
4.3	Método AGFAGs	p. 114

4.3.1	Aspectos Gerais: Análise Numérica e Descritiva	p. 114
4.3.2	Ciclofosfamida	p. 120
4.3.3	Etopósido	p. 131
4.3.4	Lomustina	p. 141
4.4	Convergência dos Resultados	p. 150
5	Conclusão	p. 154
	Referências Bibliográficas	p. 156

1 *Introdução*

1.1 **Advento das Tecnologias de Alto Desempenho: As ciências Ômicas**

Hoje pesquisadores tem acesso a tecnologias que são capazes de gerar informações sobre diferentes aspectos do material genético com vazão, eficiência e precisão muito maiores do que as tecnologias disponíveis nas décadas de 80 e 90, utilizadas para realização do projeto de sequenciamento do genoma humano e com um custo consideravelmente menor (GALAS; MCCORMACK, 2003) (NRC et al., 2007). O conjunto dessas tecnologias capazes de gerar grandes quantidades de informação sobre os diferentes aspectos do material genético, será aqui referido como “**tecnologias ômicas**”. As **ciências ômicas** são, essencialmente, a aplicação das tecnologias ômicas, ferramentas de **bioinformática** e de métodos de integração e análise de dados de grande porte para realizar um estudo global de um determinado sistema biológico exposto a determinadas condições. Alguns exemplos de **ciências ômicas** são: **genômica**, **proteômica**, **transcriptômica** e **metabolômica**. Esses quatro exemplos são a principal forma de busca de informações que a **toxicogenômica** utiliza para compreender um sistema biológico.

A **genômica** e suas tecnologias relacionadas englobam principalmente o sequenciamento do genoma de organismos, a mapeamento dos genomas (genes) e análises genotípicas. As tecnologias genômicas se desenvolveram muito na década passada (2000-2010), seja na capacidade de sequenciamento de genomas inteiros ou seja na capacidade de realizar análises genotípicas (NRC et al., 2007). Isto possibilita a avaliação de múltiplas variáveis em todo o genoma de diversos tipos de organismos ou mesmo a avaliação de polimorfismos genéticos em grandes populações de indivíduos de uma determinada espécie (GRIFFITHS, 2005). A avaliação da variação do genoma entre indivíduos pode fornecer informações sobre as dife-

rentes reações que humanos tem à agentes químicos e outros fatores ambientais. A avaliação de diferenças entre genomas de indivíduos é composta principalmente pela busca de SNPs (*single nucleotide polymorphisms*), ou seja variações na própria sequência de nucleotídeos no genoma dos indivíduos e portanto na informação genética codificada. Essa avaliação também é realizada pela busca de fatores epigenéticos, que são mudanças reversíveis e herdáveis capazes de alterar a maneira como os genes se expressam sem causar mudança na informação codificada por eles (sequência de nucleotídeos). Alguns fatores epigenéticos comuns são: metilação do DNA, *Imprinting* e modificações das histonas (NRC et al., 2007; HANDE, 2017).

A **transcriptômica** é capaz de determinar o perfil de expressão gênica de um único gene ou de vários (isso depende da técnica a ser utilizada) a partir da análise do transcriptoma, este representa o conjunto de todos os transcritos celulares em um determinado momento. A técnica *Northern Blot* (ALWINE; KEMP; STARK, 1977), por exemplo, é capaz de mensurar a expressão relativa de um único gene entre um caso e um controle (SCHULZE; DOWNWARD, 2001). Hoje em dia existem técnicas capazes de mensurar a expressão relativa de milhares de genes simultaneamente como os *chips* de microarranjo e o RNA-seq (WANG; GERSTEIN; SNYDER, 2009). Todas essas técnicas usam como base a determinação de quantidades relativas de transcritos celulares para determinar quais genes se encontram subexpressos ou superexpressos. A vantagem dessas últimas tecnologias é que elas podem fornecer informações de diversos genes simultaneamente (SCHULZE; DOWNWARD, 2001). Dessa maneira é possível determinar perfis de expressão gênica, dando ao pesquisador a capacidade de determinar genes que estão super ou subexpressos em diversas situações. Assim é possível por meio do uso de bases de dados que relacionam os genes a sua função (proteína que codifica ou ação reguladora), determinar processos biológicos que estão super ou subexpressos. Cabe lembrar que, entretanto, a relação entre quantidade relativa de transcritos celulares e concentração das proteínas nem sempre é linear e isto constitui um desafio para análise do transcriptoma (NRC et al., 2007; HANDE, 2017).

A **Proteômica** estuda o conjunto de todas as proteínas expressas em um determinado momento na célula ou tecido o qual é chamado de **proteoma**. As proteínas são as efetoras diretas da informação codificada no genoma e por isso ter acesso a sua distribuição na célula/tecido pode fornecer informações que não podem ser obtidas por vias da genômica ou

transcriptômica. Ter acesso a essas informações permite o pesquisador criar bases de dados que associam sequências do genoma com sua respectiva proteína e suas concentrações, assim como associações entre duas ou mais proteínas. O proteoma não é imutável ou fixo como o genoma, as concentrações das proteínas estão sempre variando como resposta aos estímulos que o ambiente fornece ao organismo. Não existe nenhuma tecnologia análoga ao PCR (*polimerase chain reaction*), que permite amplificação do genoma e transcriptoma, para aplicar ao proteoma. As proteínas devem ser analisadas em suas concentrações nativas, que podem abranger até 6 ordens de magnitude (de 1 a 10^6) e isto constitui um desafio considerável para análise do proteoma (NRC et al., 2007; HANDE, 2017).

Outros desafios incluem modificações constantes do proteoma (danos), causados por agentes químicos intermediários formados por substâncias tóxicas e estresse oxidativo endógeno. Devido a essas razões as tecnologias proteômicas envolvem elaboradas análises seriais de poucos produtos. Isso é o oposto do que ocorre com tecnologias paralelas, como *chips* de microarranjo, que permitem avaliação de milhares de produtos. As técnicas mais utilizadas para determinação do proteoma, atualmente, constituem um conjunto de diversas variações e adaptações da espectrometria de massa (*Mass Spectrometry* (MS)) (AEBERSOLD; MANN, 2003).

A **metabolômica** estuda o conjunto de todos os metabólitos do organismo em um determinado momento o qual é chamado de **metaboloma**. Metabólitos intermediáveis refletem a interferências de proteínas em vias bioquímicas e portanto representam o estado biológico de uma célula ou tecido de uma maneira análoga ao proteoma. O metaboloma possui uma grande diversidade química, ao contrário do genoma, do transcriptoma e do proteoma. Alguns exemplos de tipos de metabólitos são: pequenos peptídeos, lipídios, precursores de ácidos nucleicos, diversos produtos de degradação de agentes químicos intermediários em processos de biosíntese e de catabolismo e metabólitos de compostos exógenos (derivados da dieta e intervenções médicas, por exemplo) (NRC et al., 2007; HANDE, 2017). Essa diversidade química do metaboloma constitui um desafio para sua análise pois dificulta sua avaliação a partir de uma única técnica.

A metabolômica também sofre de algumas dificuldades presentes na proteômica como o fato de o metaboloma não ser imutável ou fixo como o genoma e que as abundâncias nativas dos metabólitos podem ter muitas ordens de magnitude (NRC et al., 2007; HANDE, 2017). Algumas técnicas utilizadas para avaliação do metaboloma são diversas variações da

ressonância magnética nuclear (NMR) e da espectrometria de massa (ERNST et al., 1987; WU et al., 2008).

A figura 1.1 abaixo demonstra as quatro frentes das ciências ômicas e seus objetos de estudo em uma maneira hierárquica, considerando a direção do fluxo de informação genética.

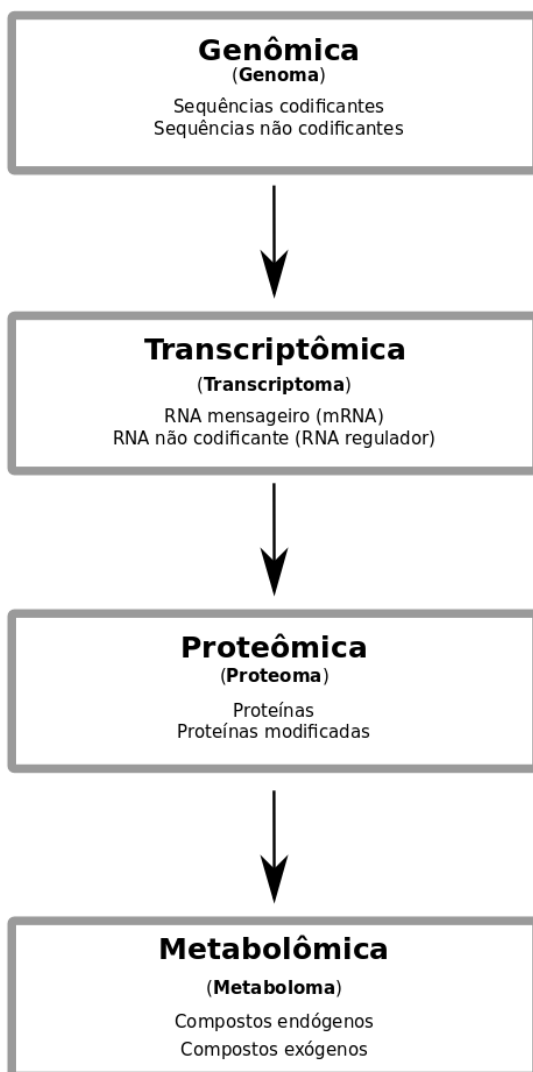


Figura 1.1: Relação hierárquica das quatro frentes das ciências ômicas. Adaptado de (NRC et al., 2007; HANDE, 2017).

5 Conclusão

Com base na análise numérico-descritiva concluímos que o método EFDEGs possui resultados (GDEs e BPs) que variam conforme a dose e o tempo. Em geral o número de resultados é maior conforme a dose da droga é maior, e o mesmo se aplica ao tempo. O método AGFAGs possui resultados (BPs apenas) cuja distribuição varia muito entre as drogas e estudos, não havendo um comportamento claro entre as distribuições como mostram os gráficos de linhas. O único comportamento observado é que quanto menor o número de GDEs em um estudo do EFDEGs menor, maior é o número de BPs encontrados nesse mesmo estudo no método AGFAGs. Os estudos *in vitro* no método EFDEGs tem, em geral, um comportamento semelhante entre si através do aumento da dose e do tempo. Sendo que a magnitude (quantidade) de resultados é o que os difere entre si. No método AGFAGs os estudos *in vitro* não possuem, entre si a mesma similaridade que é mostrada no método EFDEGs. Isso é consequência da maneira peculiar que o AGFAGs funciona.

Quanto a questão *in vivo vs in vitro*, o uso de ambas as metodologias deixou evidente que existe pouca similaridade entre os 3 estudos. As tabelas 4.1, 4.2 e 4.3 na sessão anterior deixam isso claro, a pouca convergência existente entre os 3 tipos de estudo se limita aos processos que normalmente são alterados pelas drogas antineoplásicas como apoptose, dano e reparo do DNA e controle do ciclo celular. A convergência entre os métodos também é escassa e se limita feita aos termos: ciclo celular, apoptose, reparo e dano ao DNA. Todos processos que normalmente são alterados pelas drogas antineoplásicas estudadas. Isso é interessante pois observamos processos cuja alteração era esperada em ambos os métodos, dando robustez a observação. É interessante notar que as diferenças entre os métodos muitas vezes são complementares, como no caso *Homo sapiens in vitro* da ciclofosfamida, no método EFDEGs encontramos GDEs/BPs associados aos seus principais mecanismos (ciclo celular), ao metabolismo e ainda potencial regulação pós-transcricional pelo lncRNA EGOT. O método AGFAGs para esse mesmo estudo encontrou também BPs associados também aos

principais mecanismos da droga (ciclo celular) e a potenciais efeitos adversos como fadiga e anemia (os quais não são observados no método EFDEGs). Isso não apenas deixa evidente a importância da prática de se utilizar dados biológicos para análises exploratórias na busca de hipóteses. E a meta-análise deste trabalho deixa claro como o uso de diferentes metodologias pode ampliar a gama de pontos interessantes em um mesmo conjunto de dados.

Referências Bibliográficas

- AEBERSOLD, R.; MANN, M. Mass spectrometry-based proteomics. *Nature*, Nature Publishing Group, v. 422, n. 6928, p. 198–207, 2003.
- AFFYMETRIX, I. Statistical algorithms description document. *Technical paper*, 2002.
- AKUTSU, T.; MIYANO, S.; KUHARA, S. Algorithms for identifying boolean networks and related biological networks based on matrix multiplication and fingerprint function. *Journal of computational biology*, Mary Ann Liebert, Inc., v. 7, n. 3-4, p. 331–343, 2000.
- ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Error and attack tolerance of complex networks. *nature*, Nature Publishing Group, v. 406, n. 6794, p. 378–382, 2000.
- ALEXA, A.; RAHNENFUHRER, J. topgo: enrichment analysis for gene ontology. *R package version*, v. 2, n. 0, 2010.
- ALEXANDER, J. et al. Biomarkers of exposure to heterocyclic amines: approaches to improve the exposure assessment. *Food and chemical toxicology*, Elsevier, v. 40, n. 8, p. 1131–1137, 2002.
- ALMEIDA, V. D. et al. Câncer e agentes antineoplásicos ciclo-celular específicos e ciclo-celular não específicos que interagem com o dna: uma introdução. *Quim. Nova*, SciELO Brasil, v. 28, n. 1, p. 118–129, 2005.
- ALWINE, J. C.; KEMP, D. J.; STARK, G. R. Method for detection of specific rnas in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with dna probes. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 74, n. 12, p. 5350–5354, 1977.
- AMIN, R. P. et al. Identification of putative gene based markers of renal toxicity. *Environmental health perspectives*, National Institute of Environmental Health Science, v. 112, n. 4, p. 465, 2004.
- ANTONIADIS, A.; LAMBERT-LACROIX, S.; LEBLANC, F. Effective dimension reduction methods for tumor classification using gene expression data. *Bioinformatics*, Oxford University Press, v. 19, n. 5, p. 563–570, 2003.
- ARMSTRONG, R. A. When to use the bonferroni correction. *Ophthalmic and Physiological Optics*, Wiley Online Library, v. 34, n. 5, p. 502–508, 2014.
- ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. *Nature genetics*, Nature Publishing Group, v. 25, n. 1, p. 25–29, 2000.

- BARABÁSI, A.-L.; GULBAHCE, N.; LOSCALZO, J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, Nature Publishing Group, v. 12, n. 1, p. 56–68, 2011.
- BARABASI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, v. 5, n. 2, p. 101–113, Feb 2004. Disponível em: <<http://dx.doi.org/10.1038/nrg1272>>.
- BEGLEY, T. J. et al. Damage recovery pathways in *saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Molecular cancer research*, American Association for Cancer Research, v. 1, p. 103–112, 2002.
- BEGLEY, T. J.; SAMSON, L. D. Network responses to dna damaging agents. *DNA repair*, Elsevier, v. 3, n. 8, p. 1123–1132, 2004.
- BERRA, E.; GINOUVÈS, A.; POUYSSÉGUR, J. The hypoxia-inducible-factor hydroxylases bring fresh air into hypoxia signalling. *EMBO reports*, EMBO Press, v. 7, n. 1, p. 41–45, 2006.
- BIAGI-JUNIOR, C. A. O. de. Meta-análise do projeto toxicogenômico japonês: diferenças entre modelos in vivo e in vitro.
- BLOOM, G. et al. Multi-platform, multi-site, microarray-based human tumor classification. *The American journal of pathology*, Elsevier, v. 164, n. 1, p. 9–16, 2004.
- BODDY, A. V.; YULE, S. M. Metabolism and pharmacokinetics of oxazaphosphorines. *Clinical pharmacokinetics*, Springer, v. 38, n. 4, p. 291–304, 2000.
- BOLSTAD, B. M. et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, Oxford University Press, v. 19, n. 2, p. 185–193, 2003.
- BOULESTEIX, A.-L.; TUTZ, G.; STRIMMER, K. A cart-based approach to discover emerging patterns in microarray data. *Bioinformatics*, Oxford University Press, v. 19, n. 18, p. 2465–2472, 2003.
- BROWN, M. P. et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 97, n. 1, p. 262–267, 2000.
- BUESEN, R. et al. Applying'omics technologies in chemicals risk assessment: Report of an ecetoc workshop. *Regulatory Toxicology and Pharmacology*, Elsevier, 2017.
- BUTTE, A. J.; KOHANE, I. S. Unsupervised knowledge discovery in medical databases using relevance networks. In: AMERICAN MEDICAL INFORMATICS ASSOCIATION. *Proceedings of the AMIA Symposium*. [S.l.], 1999. p. 711.
- BUTTE, A. J. et al. Discovering functional relationships between rna expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 97, n. 22, p. 12182–12186, 2000.

- CASTRO, M. A. A. et al. Viacomplex: software for landscape analysis of gene expression networks in genomic context. *Bioinformatics*, v. 25, n. 11, p. 1468–1469, Jun 2009. Disponível em: <<http://dx.doi.org/10.1093/bioinformatics/btp246>>.
- CHARITOU, T.; BRYAN, K.; LYNN, D. J. Using biological networks to integrate, visualize and analyze genomics data. *Genetics Selection Evolution*, BioMed Central, v. 48, n. 1, p. 27, 2016.
- CHELLAPPAN, S. P. Hog on the promoter: regulation of the osmotic stress response. *Science Signaling*, Science Signaling, v. 2001, n. 93, p. pe1–pe1, 2001.
- CONRADT, L. et al. Mdm2 inhibitors synergize with topoisomerase ii inhibitors to induce p53-independent pancreatic cancer cell death. *International journal of cancer*, Wiley Online Library, v. 132, n. 10, p. 2248–2257, 2013.
- COUNCIL, N. R. et al. *Health risks from dioxin and related compounds: evaluation of the EPA reassessment*. [S.l.]: National Academies Press, 2006.
- DALMA-WEISZHAUSZ, D. D. et al. [1] the affymetrix genechip® platform: An overview. *Methods in enzymology*, Elsevier, v. 410, p. 3–28, 2006.
- DATTA, A. et al. External control in markovian genetic regulatory networks: the imperfect information case. *Bioinformatics*, Oxford University Press, v. 20, n. 6, p. 924–930, 2004.
- DICKINSON, D. A. et al. Differentiation of dna reactive and non-reactive genotoxic mechanisms using gene expression profile analysis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, Elsevier, v. 549, n. 1, p. 29–41, 2004.
- DOBBIN, K. K. et al. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clinical Cancer Research*, AACR, v. 11, n. 2, p. 565–572, 2005.
- DONG, Z. et al. Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC cancer*, BioMed Central, v. 15, n. 1, p. 489, 2015.
- DONIGER, S. W. et al. Mappfinder: using gene ontology and genmapp to create a global gene-expression profile from microarray data. *Genome biology*, BioMed Central, v. 4, n. 1, p. R7, 2003.
- DRĂGHICI, S. *Statistics and data analysis for microarrays using R and bioconductor*. [S.l.]: Chapman and Hall/CRC, 2016.
- DUDOIT, S. et al. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica sinica*, JSTOR, p. 111–139, 2002.
- DYER, M. D.; MURALI, T.; SOBRAL, B. W. The landscape of human proteins interacting with viruses and other pathogens. *PLoS pathogens*, Public Library of Science, v. 4, n. 2, p. e32, 2008.

- EISEN, M. B. et al. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 95, n. 25, p. 14863–14868, 1998.
- EKHART, C. et al. Influence of polymorphisms of drug metabolizing enzymes (cyp2b6, cyp2c9, cyp2c19, cyp3a4, cyp3a5, gsta1, gstp1, aldh1a1 and aldh3a1) on the pharmacokinetics of cyclophosphamide and 4-hydroxycyclophosphamide. *Pharmacogenetics and genomics*, LWW, v. 18, n. 6, p. 515–523, 2008.
- ELLIS, M. et al. Development and validation of a method for using breast core needle biopsies for gene expression microarray analyses. *Clinical Cancer Research*, AACR, v. 8, n. 5, p. 1155–1166, 2002.
- ERNST, R. R. et al. Principles of nuclear magnetic resonance in one and two dimensions. Clarendon Press Oxford, 1987.
- FAMILI, I.; MAHADEVAN, R.; PALSSON, B. O. k-cone analysis: determining all candidate values for kinetic parameters on a network scale. *Biophysical journal*, Elsevier, v. 88, n. 3, p. 1616–1625, 2005.
- FINKEL, T.; HOLBROOK, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature*, Nature Publishing Group, v. 408, n. 6809, p. 239–247, 2000.
- FREEMAN, T. C. et al. A gene expression atlas of the domestic pig. *BMC biology*, Springer, v. 10, n. 1, p. 90, 2012.
- FRIEDMAN, N. et al. Using bayesian networks to analyze expression data. *Journal of computational biology*, Mary Ann Liebert, Inc., v. 7, n. 3-4, p. 601–620, 2000.
- GALAS, D. J.; MCCORMACK, S. J. An historical perspective on genomic technologies. 2003.
- GAUTIER, L. et al. affy-analysis of affymetrix genechip data at the probe level. *Bioinformatics*, Oxford Univ Press, v. 20, n. 3, p. 307–315, 2004.
- GENTLEMAN, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, BioMed Central, v. 5, n. 10, p. 1, 2004.
- GIAEVER, G. et al. Functional profiling of the saccharomyces cerevisiae genome. *nature*, Nature Publishing Group, v. 418, n. 6896, p. 387, 2002.
- GOLUB, T. R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, American Association for the Advancement of Science, v. 286, n. 5439, p. 531–537, 1999.
- GRIFFITHS, A. J. *An introduction to genetic analysis*. [S.l.]: Macmillan, 2005.
- HAMADEH, H. K. et al. Prediction of compound signature using high density gene expression profiling. *Toxicological Sciences*, Oxford University Press, v. 67, n. 2, p. 232–240, 2002.

- HANDE, M. P. Book review: Toxicogenomics in predictive carcinogenicity. *Frontiers in Genetics*, Frontiers, v. 8, p. 11, 2017.
- HARTWELL, L. H. et al. From molecular to modular cell biology. *Nature*, Nature Publishing Group, v. 402, p. C47–C52, 1999.
- HASHIMOTO, R. F. et al. Growing genetic regulatory networks from seed genes. *Bioinformatics*, Oxford University Press, v. 20, n. 8, p. 1241–1247, 2004.
- HERRERO, J.; VALENCIA, A.; DOPAZO, J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics*, Oxford University Press, v. 17, n. 2, p. 126–136, 2001.
- HOEN, P. A. t et al. Fluorescent labelling of crna for microarray applications. *Nucleic acids research*, Oxford University Press, v. 31, n. 5, p. e20–e20, 2003.
- HOSACK, D. A. et al. Identifying biological themes within lists of genes with ease. *Genome biology*, BioMed Central, v. 4, n. 10, p. R70, 2003.
- HUANG, Z.; ROY, P.; WAXMAN, D. J. Role of human liver microsomal cyp3a4 and cyp2b6 in catalyzing n-dechloroethylation of cyclophosphamide and ifosfamide. *Biochemical pharmacology*, Elsevier, v. 59, n. 8, p. 961–972, 2000.
- HUMANS, I. W. G. on the Evaluation of Carcinogenic Risks to et al. Pharmaceuticals. volume 100 a. a review of human carcinogens. *IARC monographs on the evaluation of carcinogenic risks to humans*, Various, v. 100, n. PT A, p. 1, 2012.
- IARC et al. *Overall evaluations of carcinogenicity: an updating of IARC monographs volumes 1 to 42*. [S.l.]: IARC Lyon, 1987.
- IMOTO, S. et al. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *Journal of bioinformatics and computational biology*, World Scientific, v. 1, n. 02, p. 231–252, 2003.
- IRIZARRY, R. A. et al. Multiple-laboratory comparison of microarray platforms. *Nature methods*, Nature Publishing Group, v. 2, n. 5, p. 345, 2005.
- JANSSEN, A. W. et al. The impact of ppar α activation on whole genome gene expression in human precision cut liver slices. *BMC genomics*, BioMed Central, v. 16, n. 1, p. 760, 2015.
- JELINSKY, S. A. et al. Regulatory networks revealed by transcriptional profiling of damaged saccharomyces cerevisiae cells: Rpn4 links base excision repair with proteasomes. *Molecular and cellular biology*, Am Soc Microbiol, v. 20, n. 21, p. 8157–8167, 2000.
- JELINSKY, S. A.; SAMSON, L. D. Global response of saccharomyces cerevisiae to an alkylating agent. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 96, n. 4, p. 1486–1491, 1999.

JENSEN, L. J. et al. String 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research*, Oxford Univ Press, v. 37, n. suppl 1, p. D412–D416, 2009.

JIANG, N. et al. Methods for evaluating gene expression from affymetrix microarray datasets. *BMC bioinformatics*, BioMed Central, v. 9, n. 1, p. 284, 2008.

JONSSON, P. F.; BATES, P. A. Global topological features of cancer proteins in the human interactome. *Bioinformatics*, Oxford University Press, v. 22, n. 18, p. 2291–2297, 2006.

KANEHISA, M.; GOTO, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, Oxford University Press, v. 28, n. 1, p. 27–30, 2000.

KILLICK, K. E. et al. Key hub and bottleneck genes differentiate the macrophage response to virulent and attenuated mycobacterium bovis. *Frontiers in immunology*, Frontiers Media SA, v. 5, 2014.

KIM, S. K. et al. A gene expression map for caenorhabditis elegans. *Science*, American Association for the Advancement of Science, v. 293, n. 5537, p. 2087–2092, 2001.

KING, C. et al. Udp-glucuronosyltransferases. *Current drug metabolism*, Bentham Science Publishers, v. 1, n. 2, p. 143–161, 2000.

LAGE, K. et al. Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 109, n. 35, p. 14035–14040, 2012.

LARKIN, J. E. et al. Independence and reproducibility across microarray platforms. *Nature methods*, Nature Publishing Group, v. 2, n. 5, p. 337, 2005.

LAWLESS, N. et al. Microrna regulation of bovine monocyte inflammatory and metabolic networks in an in vivo infection model. *G3: Genes, Genomes, Genetics*, G3: Genes, Genomes, Genetics, v. 4, n. 6, p. 957–971, 2014.

LE, Q.-T. et al. Identification of osteopontin as a prognostic plasma marker for head and neck squamous cell carcinomas. *Clinical Cancer Research*, AACR, v. 9, n. 1, p. 59–67, 2003.

LEAR, L.; NATION, R.; STUPANS, I. Effects of cyclophosphamide and adriamycin on rat hepatic microsomal glucuronidation and lipid peroxidation. *Biochemical pharmacology*, Elsevier, v. 44, n. 4, p. 747–753, 1992.

LEE, F. Y. et al. Clinical pharmacokinetics of oral ccnu (lomustine). *Cancer chemotherapy and pharmacology*, Springer, v. 14, n. 2, p. 125–131, 1985.

LEHMANN, J. M. et al. The human orphan nuclear receptor pxx1 is activated by compounds that regulate cyp3a4 gene expression and cause drug interactions. *The Journal of clinical investigation*, Am Soc Clin Investig, v. 102, n. 5, p. 1016–1023, 1998.

- MAANEN, J. V. et al. Mechanism of action of antitumor drug etoposide: a review. *JNCI: Journal of the National Cancer Institute*, Oxford University Press, v. 80, n. 19, p. 1526–1533, 1988.
- MALIK, A. et al. Network analysis for the identification of differentially expressed hub genes using myogenin knock-down muscle satellite cells. *PloS one*, Public Library of Science, v. 10, n. 7, p. e0133597, 2015.
- MANFREDI, J. J. The mdm2–p53 relationship evolves: Mdm2 swings both ways as an oncogene and a tumor suppressor. *Genes & development*, Cold Spring Harbor Lab, v. 24, n. 15, p. 1580–1589, 2010.
- MARSELOS, M.; VAINIO, H. Carcinogenic properties of pharmaceutical agents evaluated in the iarc monographs programme. *Carcinogenesis*, Oxford University Press, v. 12, n. 10, p. 1751–1766, 1991.
- MEDEIROS, A.; BIRD, M.; WITZ, G. Potential biomarkers of benzene exposure. *Journal of toxicology and environmental health*, Taylor & Francis, v. 51, n. 6, p. 519–539, 1997.
- MERING, C. V. et al. String 7—recent developments in the integration and prediction of protein interactions. *Nucleic acids research*, Oxford Univ Press, v. 35, n. suppl 1, p. D358–D362, 2007.
- MERING, C. V. et al. String: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic acids research*, Oxford Univ Press, v. 33, n. suppl 1, p. D433–D437, 2005.
- MONTESANO, R.; PEGG, A. E.; MARGISON, G. P. Alkylation of dna and carcinogenicity of n-nitroso compounds. *Journal of Toxicology and Environmental Health, Part A Current Issues*, Taylor & Francis, v. 6, n. 5-6, p. 1001–1008, 1980.
- NACU, Ş. et al. Gene expression network analysis and applications to immunology. *Bioinformatics*, Oxford University Press, v. 23, n. 7, p. 850–858, 2007.
- NETZER, N. et al. Hypoxia, oxidative stress and fat. *Biomolecules*, Multidisciplinary Digital Publishing Institute, v. 5, n. 2, p. 1143–1150, 2015.
- NGUYEN, D. V.; ROCKE, D. M. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*, Oxford University Press, v. 18, n. 1, p. 39–50, 2002.
- NIELSEN, H. B.; GAUTIER, L.; GAUTIER, M. L. Package affypdnn.
- NRC et al. *Applications of toxicogenomic technologies to predictive toxicology and risk assessment*. [S.l.]: National Academies Press (US), 2007.
- ORR, M.; SCHERF, U. Large-scale gene expression analysis in molecular target discovery. *Leukemia*, Nature Publishing Group, v. 16, n. 4, p. 473, 2002.

- PEGG, A. E. Repair of o 6-alkylguanine by alkyltransferases. *Mutation Research/Reviews in Mutation Research*, Elsevier, v. 462, n. 2, p. 83–100, 2000.
- PERERA, F.; SANTELLA, R.; POIRIER, M. Biomonitoring of workers exposed to carcinogens: immunoassays to benzo (a) pyrene-dna adducts as a prototype. *Journal of Occupational and Environmental Medicine*, LWW, v. 28, n. 10, p. 1117–1123, 1986.
- PETRICOIN, E. F. et al. Serum proteomic patterns for detection of prostate cancer. *Journal of the National Cancer Institute*, Oxford University Press, v. 94, n. 20, p. 1576–1578, 2002.
- PETROS, W. P. et al. Associations between drug metabolism genotype, chemotherapy pharmacokinetics, and overall survival in patients with breast cancer. *Journal of clinical oncology*, American Society of Clinical Oncology, v. 23, n. 25, p. 6117–6125, 2005.
- RAMASWAMY, S. et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 98, n. 26, p. 15149–15154, 2001.
- RAYCHAUDHURI, S.; STUART, J. M.; ALTMAN, R. B. Principal components analysis to summarize microarray experiments: application to sporulation time series. In: NIH PUBLIC ACCESS. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*. [S.l.], 2000. p. 455.
- RUDEN, D. M.; GURDZIEL, K.; ASCHNER, M. Frontiers in toxicogenomics in the twenty-first century, the grand challenge: To understand how the genome and epigenome interact with the toxic environment at the single-cell, whole-organism, and multi-generational level. *Frontiers in Genetics*, Frontiers, v. 8, p. 173, 2017.
- SAID, M. R. et al. Global network analysis of phenotypic effects: protein networks and toxicity modulation in *saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 101, n. 52, p. 18006–18011, 2004.
- SAVOIE, C. J. et al. Use of gene networks from full genome microarray libraries to identify functionally relevant drug-affected genes and gene regulation cascades. *Dna Research*, Oxford University Press, v. 10, n. 1, p. 19–25, 2003.
- SCHAEFER, M. H.; SERRANO, L.; ANDRADE-NAVARRO, M. A. Correcting for the study bias associated with protein–protein interaction measurements reveals differences between protein degree distributions from different cancer types. *Frontiers in genetics*, Frontiers Media SA, v. 6, 2015.
- SCHLEKER, S.; TRILLING, M. Data-warehousing of protein-protein interactions indicates that pathogens preferentially target hub and bottleneck proteins. *Frontiers in microbiology*, Frontiers Media SA, v. 4, 2013.
- SCHMÄHL, D.; HABS, M. Carcinogenic action of low-dose cyclophosphamide given orally to sprague-dawley rats in a lifetime experiment. *International journal of cancer*, Wiley Online Library, v. 23, n. 5, p. 706–712, 1979.

- SCHULZE, A.; DOWNWARD, J. Navigating gene expression using microarrays—a technology review. *Nature cell biology*, Nature Publishing Group, v. 3, n. 8, p. E190, 2001.
- SELVAKUMAR, E. et al. Mitigation of oxidative stress in cyclophosphamide-challenged hepatic tissue by dl- α -lipoic acid. *Molecular and cellular biochemistry*, Springer, v. 272, n. 1-2, p. 179–185, 2005.
- SHMULEVICH, I. et al. Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, Oxford University Press, v. 18, n. 2, p. 261–274, 2002.
- SHUGART, L.; HOLLAND, J.; RANN, R. Dosimetry of pah skin carcinogenesis: covalent binding of benzo [a] pyrene to mouse epidermal dna. *Carcinogenesis*, Oxford University Press, v. 4, n. 2, p. 195–198, 1983.
- SKIPPER, P. L. et al. Protein adducts as biomarkers of human carcinogen exposure. *Drug metabolism reviews*, Taylor & Francis, v. 26, n. 1-2, p. 111–124, 1994.
- SMYTH, G. K. Limma: linear models for microarray data. In: *Bioinformatics and computational biology solutions using R and Bioconductor*. [S.l.]: Springer, 2005. p. 397–420.
- SOINOV, L. *Supervised classification for gene network reconstruction*. [S.l.]: Portland Press Limited, 2003.
- SOUKAS, A. et al. Leptin-specific patterns of gene expression in white adipose tissue. *Genes & development*, Cold Spring Harbor Lab, v. 14, n. 8, p. 963–980, 2000.
- SOUZA, T. M.; KLEINJANS, J.; JENNEN, D. G. Dose and time dependencies in stress pathway responses during chemical exposure: Novel insights from gene regulatory networks. *Frontiers in genetics*, Frontiers, v. 8, p. 142, 2017.
- SUMIDA, K. et al. Effects of dms0 on gene expression in human and rat hepatocytes. *Human & experimental toxicology*, Sage Publications Sage UK: London, England, v. 30, n. 10, p. 1701–1709, 2011.
- SZKLARCZYK, D. et al. The string database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*, Oxford Univ Press, v. 39, n. suppl 1, p. D561–D568, 2011.
- SZKLARCZYK, D. et al. String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, Oxford University Press, v. 43, n. D1, p. D447–D452, 2014.
- TAKASHIMA, K. et al. Effect of the difference in vehicles on gene expression in the rat liver—analysis of the control data in the toxicogenomics project database. *Life sciences*, Elsevier, v. 78, n. 24, p. 2787–2796, 2006.

- TAMADA, Y. et al. Estimating gene networks from gene expression data by combining bayesian network model with promoter element detection. *Bioinformatics*, Oxford University Press, v. 19, n. suppl_2, p. ii227–ii236, 2003.
- TAMAYO, P. et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 96, n. 6, p. 2907–2912, 1999.
- THEILHABER, J. et al. Finding genes in the c2c12 osteogenic pathway by k-nearest-neighbor classification of expression data. *Genome research*, Cold Spring Harbor Lab, v. 12, n. 1, p. 165–176, 2002.
- TÖRÖNEN, P. et al. Analysis of gene expression data using self-organizing maps. *FEBS letters*, Wiley Online Library, v. 451, n. 2, p. 142–146, 1999.
- TUSHER, V. G.; TIBSHIRANI, R.; CHU, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 98, n. 9, p. 5116–5121, 2001.
- UEHARA, T. et al. Species-specific differences in coumarin-induced hepatotoxicity as an example toxicogenomics-based approach to assessing risk of toxicity to humans. *Human & experimental toxicology*, Sage Publications Sage UK: London, England, v. 27, n. 1, p. 23–35, 2008.
- UEHARA, T. et al. Toxicogenomic biomarkers for renal papillary injury in rats. *Toxicology*, Elsevier, v. 303, p. 1–8, 2013.
- UEHARA, T. et al. Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. *Toxicology and applied pharmacology*, Elsevier, v. 255, n. 3, p. 297–306, 2011.
- UEHARA, T. et al. The japanese toxicogenomics project: application of toxicogenomics. *Molecular nutrition & food research*, Wiley Online Library, v. 54, n. 2, p. 218–227, 2010.
- VEAL, G. J. et al. Cyclophosphamide pharmacokinetics and pharmacogenetics in children with b-cell non-hodgkin's lymphoma. *European Journal of Cancer*, Elsevier, v. 55, p. 56–64, 2016.
- VEZINA, C. M.; WALKER, N. J.; OLSON, J. R. Subchronic exposure to tcdd, pecdf, pcb126, and pcb153: effect on hepatic gene expression. *Environmental health perspectives*, National Institute of Environmental Health Science, v. 112, n. 16, p. 1636, 2004.
- WANG, J. et al. Clustering of the som easily reveals distinct gene expression patterns: results of a reanalysis of lymphoma study. *BMC bioinformatics*, BioMed Central, v. 3, n. 1, p. 36, 2002.
- WANG, Z.; GERSTEIN, M.; SNYDER, M. Rna-seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, Nature Publishing Group, v. 10, n. 1, p. 57–63, 2009.

- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of small-world networks. *nature*, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998.
- WEINSTEIN, J. N. et al. An information-intensive approach to the molecular pharmacology of cancer. *Science*, American Association for the Advancement of Science, v. 275, n. 5298, p. 343–349, 1997.
- WEIS, B. Standardizing global gene expression analysis between laboratories and across platforms. *Nature methods*, Nature Publishing Group, v. 2, n. 5, p. 351, 2005.
- WEIZMANN. *GeneCards: Human Gene Database*. jun. 2018. Disponível em: <[www-genecards.org](http://www.genecards.org)>.
- WEN, X. et al. Large-scale temporal gene expression mapping of central nervous system development. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 95, n. 1, p. 334–339, 1998.
- WIBACK, S. J.; MAHADEVAN, R.; PALSSON, B. Ø. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the escherichia coli spectrum. *Biotechnology and bioengineering*, Wiley Online Library, v. 86, n. 3, p. 317–331, 2004.
- WU, H. et al. High-throughput tissue extraction protocol for nmr-and ms-based metabolomics. *Analytical biochemistry*, Elsevier, v. 372, n. 2, p. 204–212, 2008.
- WU, Y. et al. Long noncoding rna eosinophil granule ontogeny transcript inhibits cell proliferation and migration and promotes cell apoptosis in human glioma. *Experimental and therapeutic medicine*, Spandidos Publications, v. 14, n. 4, p. 3817–3823, 2017.
- WU, Z. et al. A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American statistical Association*, Taylor & Francis, v. 99, n. 468, p. 909–917, 2004.
- XENARIOS, I. et al. Dip, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic acids research*, Oxford University Press, v. 30, n. 1, p. 303–305, 2002.
- XU, S.-p. et al. Downregulation of the long noncoding rna egot correlates with malignant status and poor prognosis in breast cancer. *Tumor Biology*, Springer, v. 36, n. 12, p. 9807–9812, 2015.
- YU, H. et al. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS computational biology*, Public Library of Science, v. 3, n. 4, p. e59, 2007.
- ZEEBERG, B. R. et al. Gominer: a resource for biological interpretation of genomic and proteomic data. *Genome biology*, BioMed Central, v. 4, n. 4, p. R28, 2003.

ZEMPLICKIS, D. et al. Teratogenicity and carcinogenicity in a twin exposed in utero to cyclophosphamide. *Teratogenesis, carcinogenesis, and mutagenesis*, Wiley Online Library, v. 13, n. 3, p. 139–143, 1993.

ZOU, M.; CONZEN, S. D. A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, Oxford University Press, v. 21, n. 1, p. 71–79, 2004.