Overview of Bioinformatic Sequence Analysis and Applications in Toxicology

Tamer A. Mansour, MBChB, MSc, PhD
University of California, Davis

Ahmed Abdelmoneim, BVMS, MSc, PhD
Louisiana State University

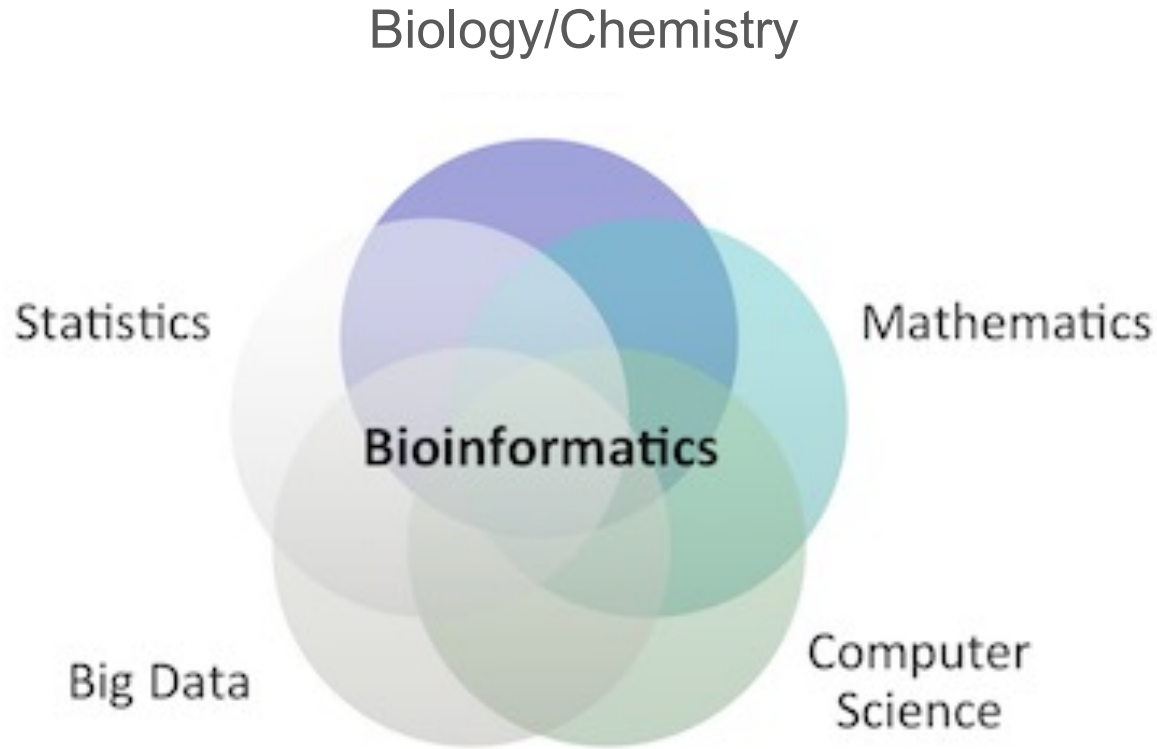
Bioinformatics

Bioinformatics is an interdisciplinary field that

develops/applies computational methods

for processing and understanding

of biological data



Need for Big Data Handling

- **Accumulation of data in huge databases** e.g., sequence databases or public health records
- **Advancement in high throughput data generation** e.g., omics including DNA and RNA seq, proteomics, metabolomics
- **Availability of complex data types** e.g., macromolecular 3D structure

- **Exponential increase of computational power** (1 trillion-fold increase in 60 years!)

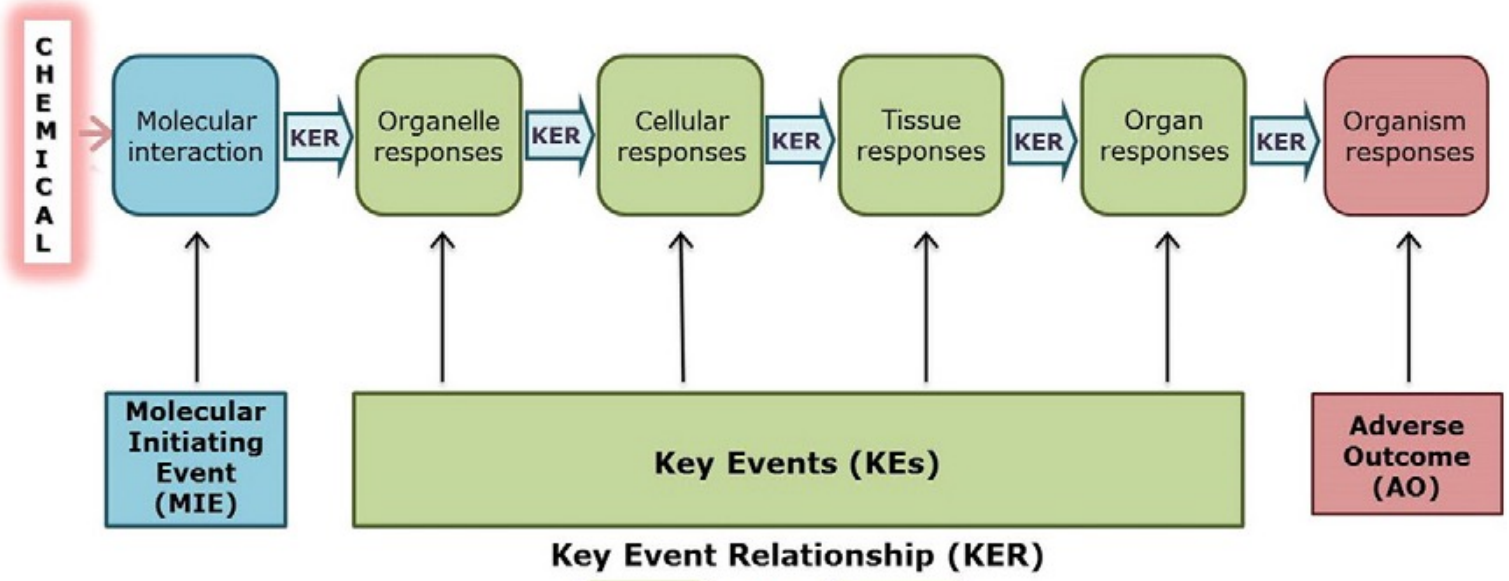
https://www.visualcapitalist.com/visualizing-trillion-fold-increase-computing-power/

Domains of Bioinformatic Applications

- Human and Veterinary Medicine
- Pharmacology and Drug Design
- Toxicology & Environmental Health
- Microbiology and Immunology
- Computational Anatomy and Imaging
- Ecology & Evolution Sciences
- Agriculture and Food Science
- Astronomy
- Oceanography

Toxicology and Impacts on the Molecular Machinery

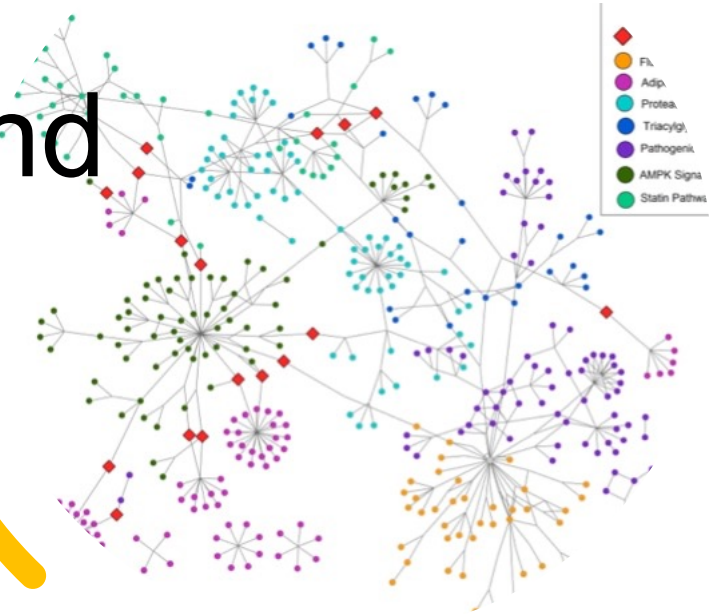
- Exposure and adverse effects
- Genetic code to function
- Integrating molecular biology to toxicology

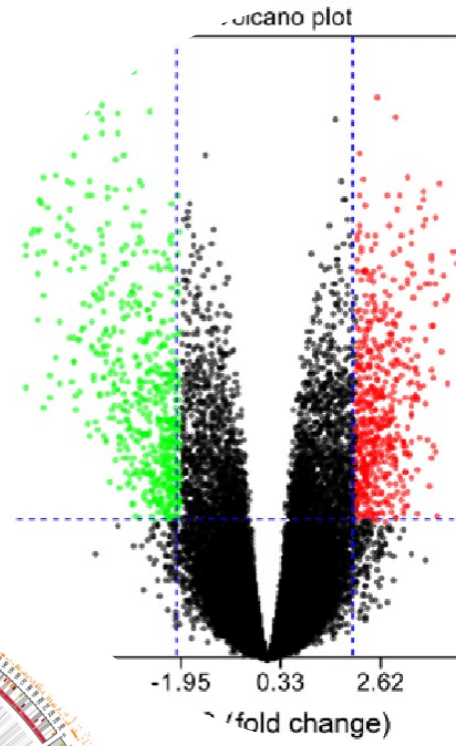


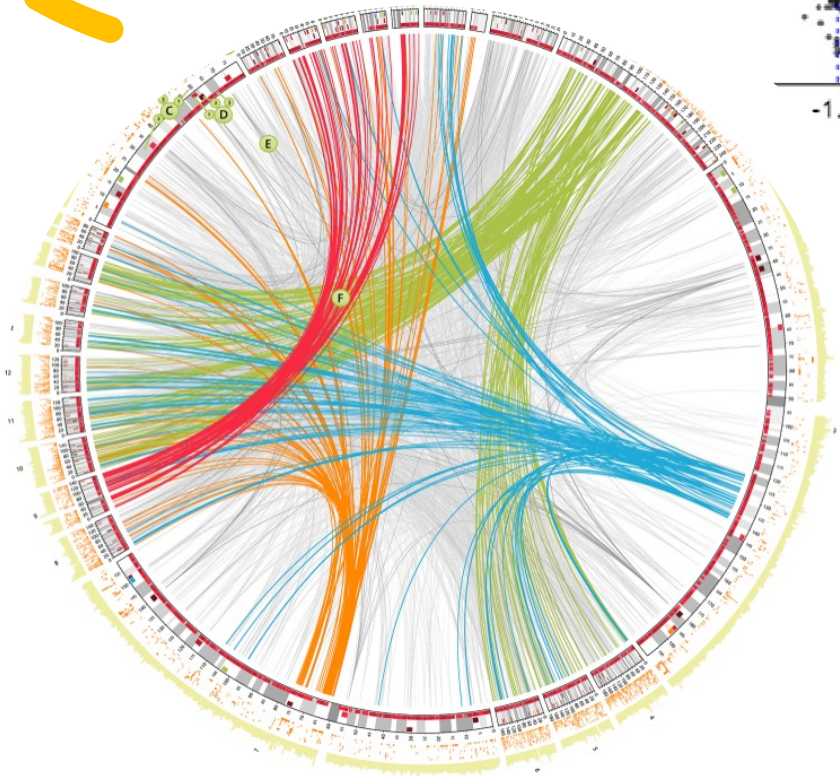
Bal-Price & Meek, 2017

Need for Databases and Analytical Tools

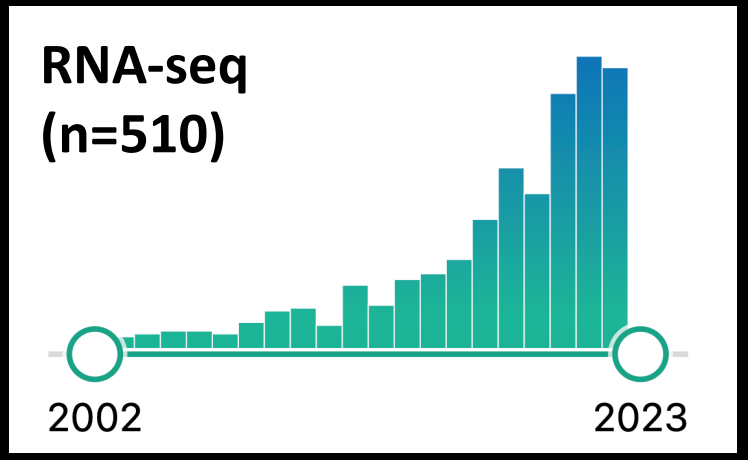
- Traditional techniques
- Working at the whole-genome level

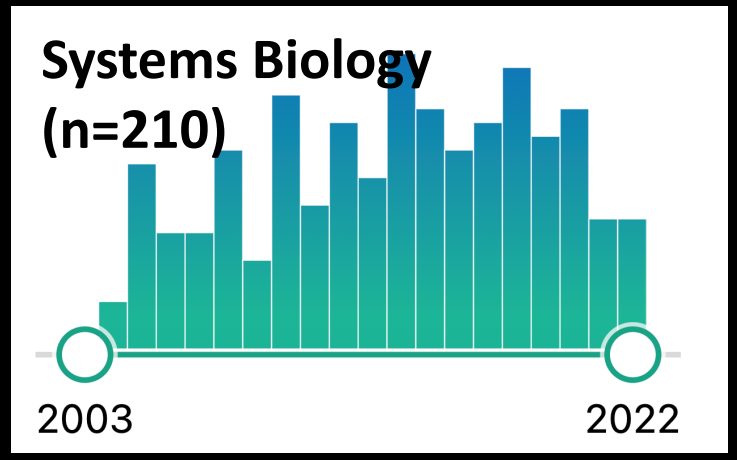


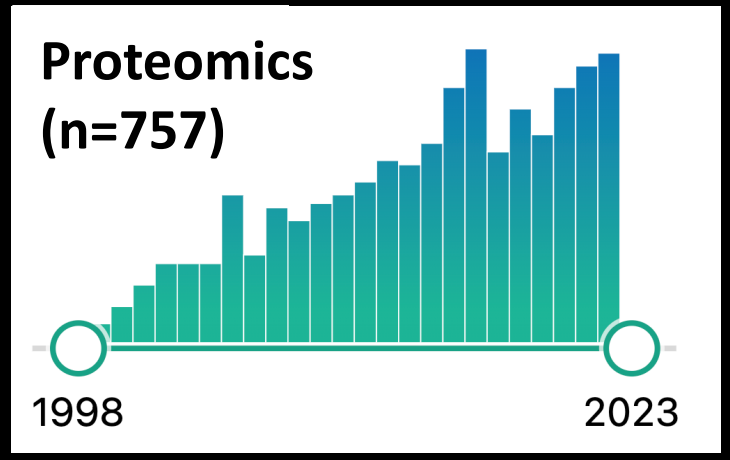


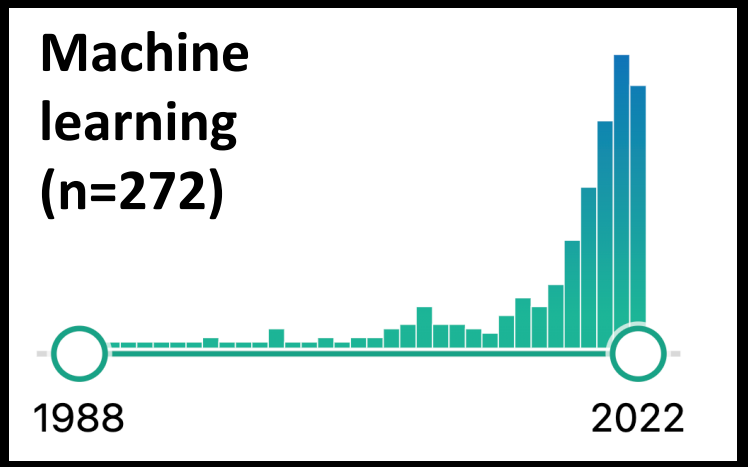


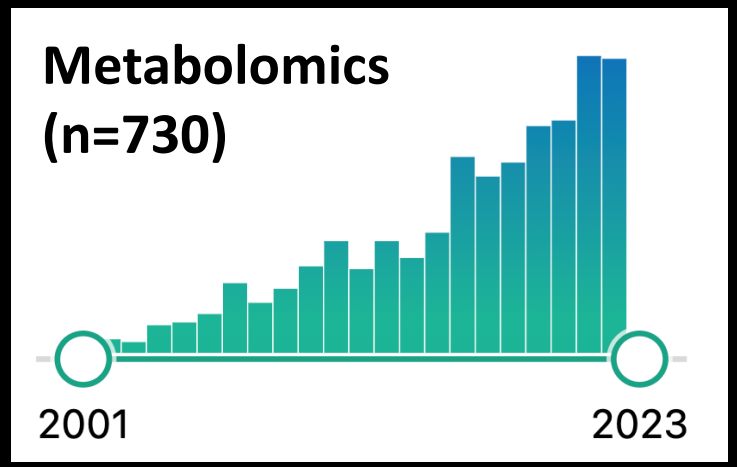
Bioinformatics in Toxicology

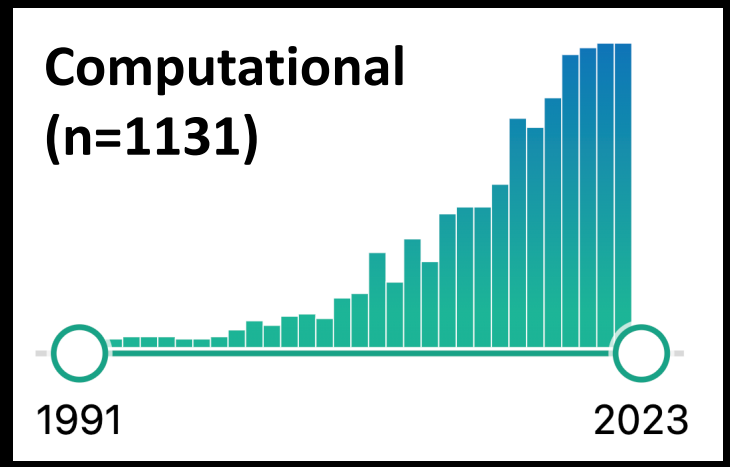










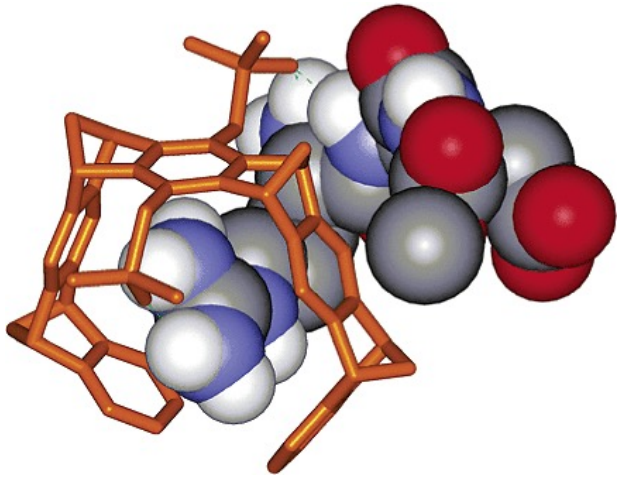


Fields & Techniques of Bioinformatics

- **Structural Bioinformatics**
- **Sequence Analysis (Genomics)**
- **Networks and Systems Biology**
- Signal and Image Processing
- Biomedical Informatics
- Database Curation

Fields & Techniques of Bioinformatics

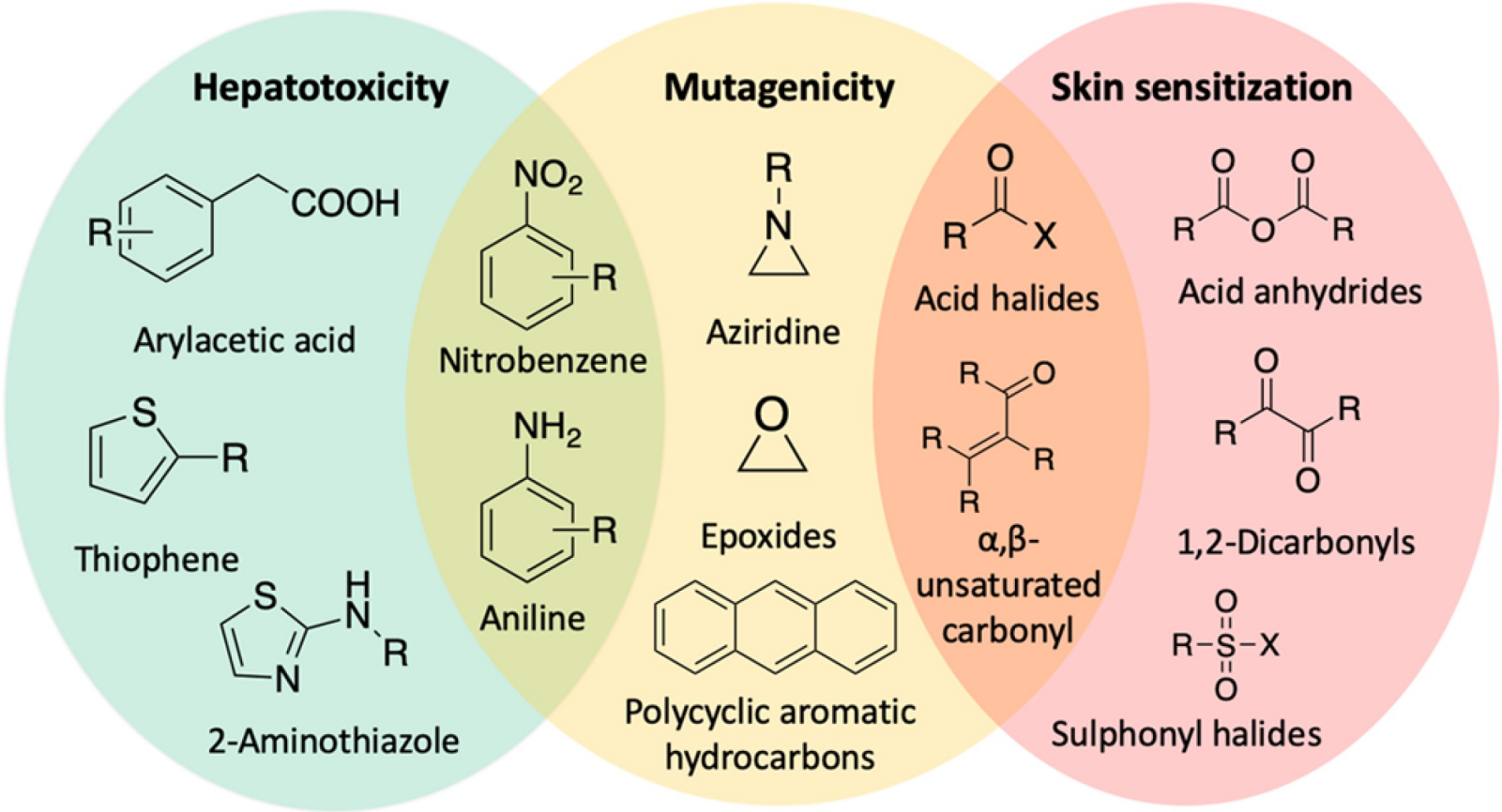
- **Structural Bioinformatics**
- **Sequence Analysis (Genomics)**
- **Networks and Systems Biology**
- Signal and Image Processing
- Biomedical Informatics
- Database Curation



- Molecular structure prediction
- Molecular interaction modeling
- Molecular docking & Drug design

Structural Bioinformatics

- Toxicity Prediction from structural alerts



Yang et al. 2020

Structural Bioinformatics

- **Expert systems**: Derek Nexus, LeadScope, AGDH, ChemoTyper, ToxAlerts
- **Frequency analysis**: Moss, Gaston, SARpy, CASE
- **Machine learning**: Bioalerts, FP, CNN

Yang et al. 2020

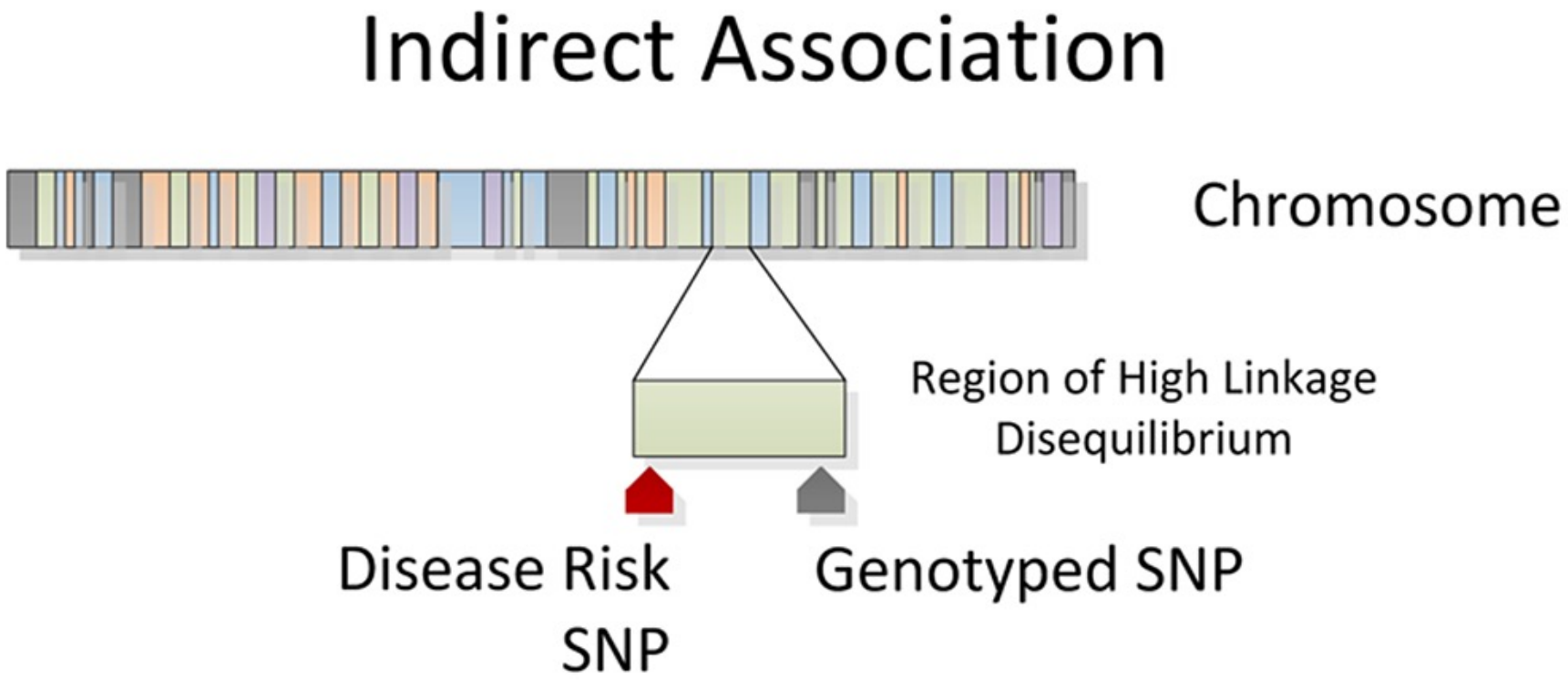
Fields & Techniques of Bioinformatics

- **Structural Bioinformatics**
- **Sequence Analysis (Genomics)**
- **Networks and Systems Biology**
- Signal and Image Processing
- Biomedical Informatics
- Database Curation

- Sequence search
- Genome mining
- Population genomics
- Comparative genomics
- Variant discovery
- Expression analysis
- Sequence binding analysis
- Epigenetic analysis

(Toxico)genomics

A) Variant Calling and Association Studies



(Toxico)genomics

A) Variant Calling and Association Studies

- Genetic Markers: genome-wide (GWAS), exome-wide or targeted. Typically genotyped by NGS or microarrays.
- Test the association of genetic markers with a phenotype to identify risk variants.
- Further analysis to identify the causative mutation

Journal of *Personalized Medicine*

MDPI

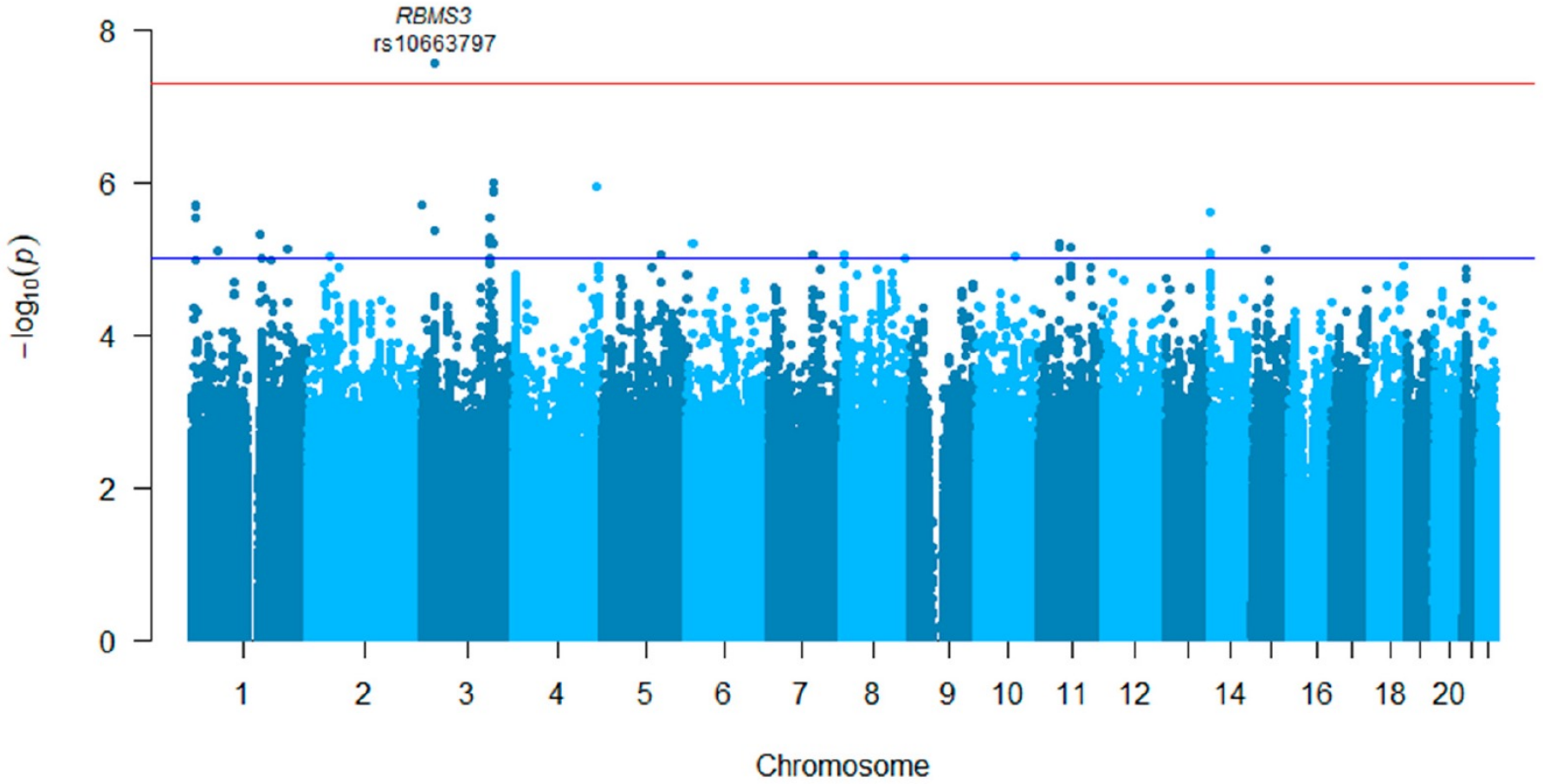
Article

Genome-Wide Analyses of Nephrotoxicity in Platinum-Treated Cancer Patients Identify Association with Genetic Variant in *RBMS3* and Acute Kidney Injury

Marije J. Klumpers [1], Ward De Witte [2], Giovanna Gattuso [3], Elisabetta Schiavello [3], Monica Terenziani [3], Maura Massimino [3], Corrie E. M. Gidding [4], Sita H. Vermeulen [5], Chantal M. Driessen [6], Carla M. van Herpen [6], Esther van Meerten [7], Henk-Jan Guchelaar [8], Marieke J. H. Coenen [2,†] and D. Maroeska W. M. te Loo [1,*,†]

- Platinum agents are chemotherapeutic agents.
- Nephrotoxicity is a common and dose-limiting side effect.
- Genetic risk loci for platinum-induced nephrotoxicity.

- 195 platinum-treated patients.
- 9,799,032 DNA variants tested.
- Association between RBMS3 rs10663797 and acute kidney injury.



Klumpers et al., 2022

(Toxico)genomics

B) Expression Studies

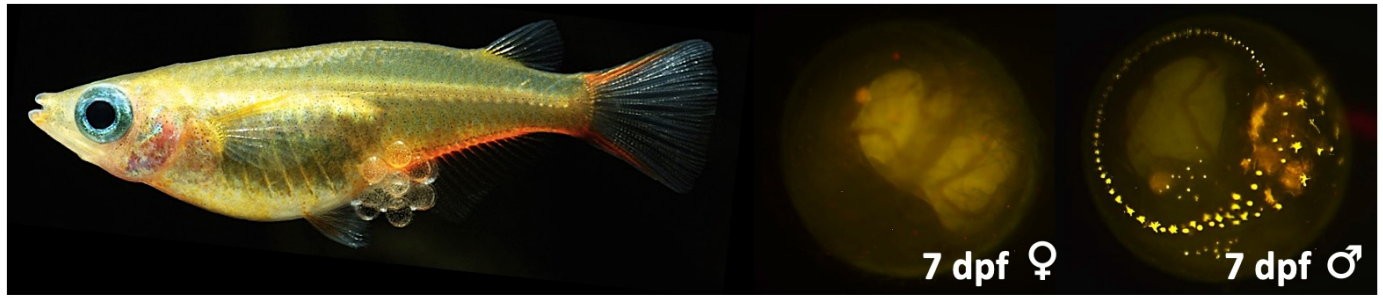
- Identify a list of differentially expressed genes or gene products.
- These lists are typically tested for enrichment in functional pathways, annotation terms, interactome modules, etc. to understand the underlying molecular change.
- RNA-seq, proteomics, and metabolomics are the commonly used techniques.

Aquatic Toxicology 208 (2019) 187–195

Molecular signaling pathways elicited by 17α-ethinylestradiol in Japanese medaka male larvae undergoing gonadal differentiation

Ahmed Abdelmoneim[a,b], Amira Abdu[a,c], Shuai Chen[a], Maria S. Sepúlveda[a,*]

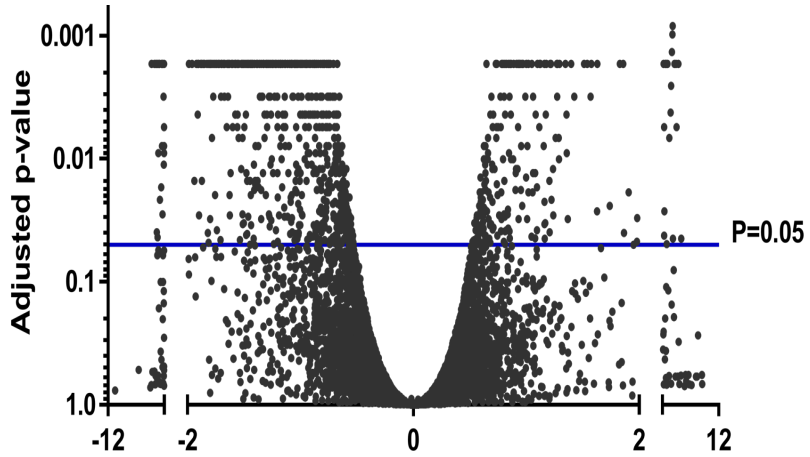
- Estrogenic contaminants alter gonadal development
- Induce gonadal intersex and sex reversal
- Transcriptomic approaches to identify molecular initiating events



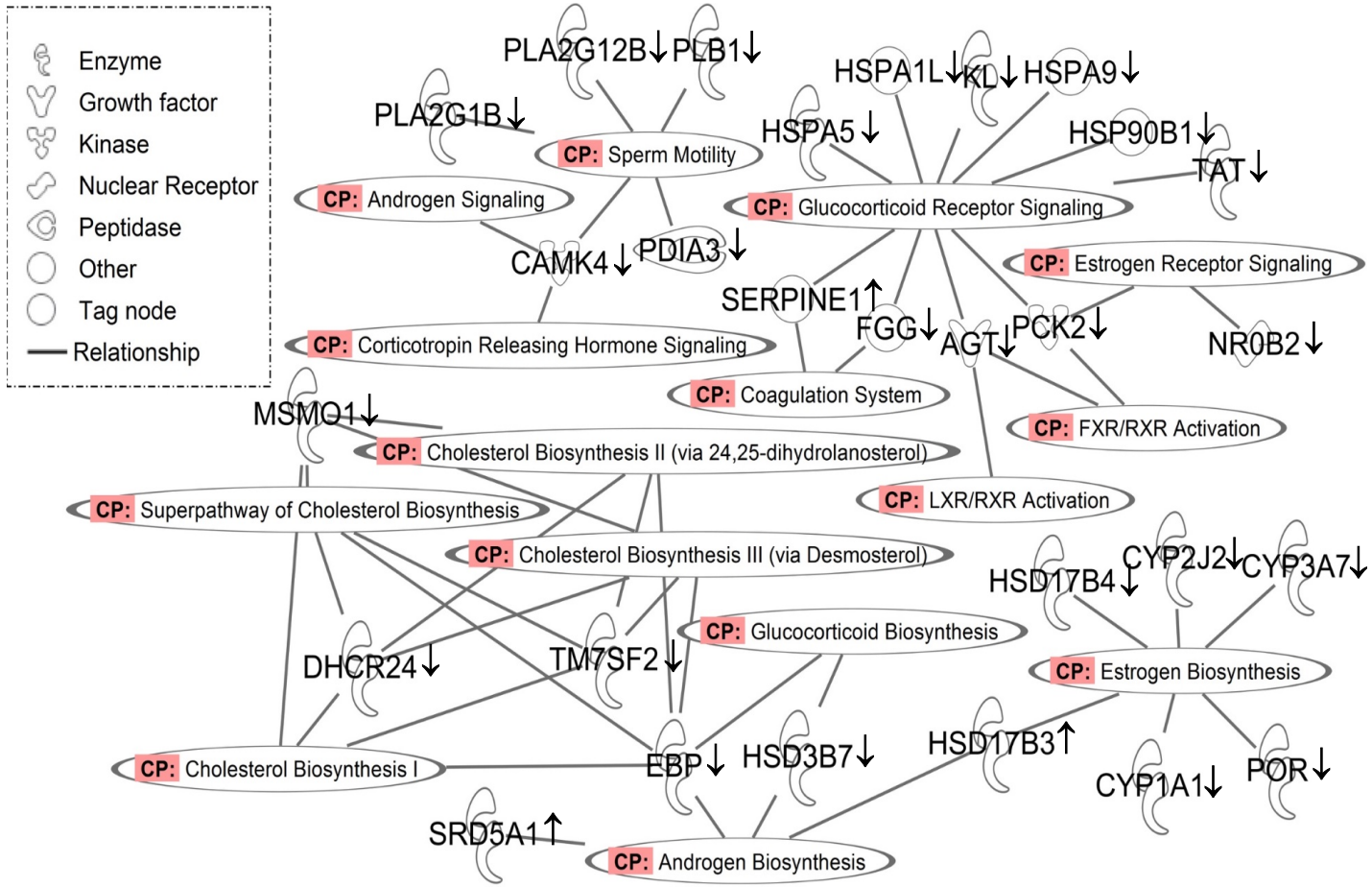
dpf = days post fertilization

- Hundreds of DEGs identified in relation to different exposures.
- Various molecular signaling pathways were impacted by these exposures, suggesting potential role in the development of female phenotypic changes.

Differentially expressed genes (DEGs) between control and intersex males



Genes and canonical pathways altered in intersex males



Abdelmoneim et al., 2019

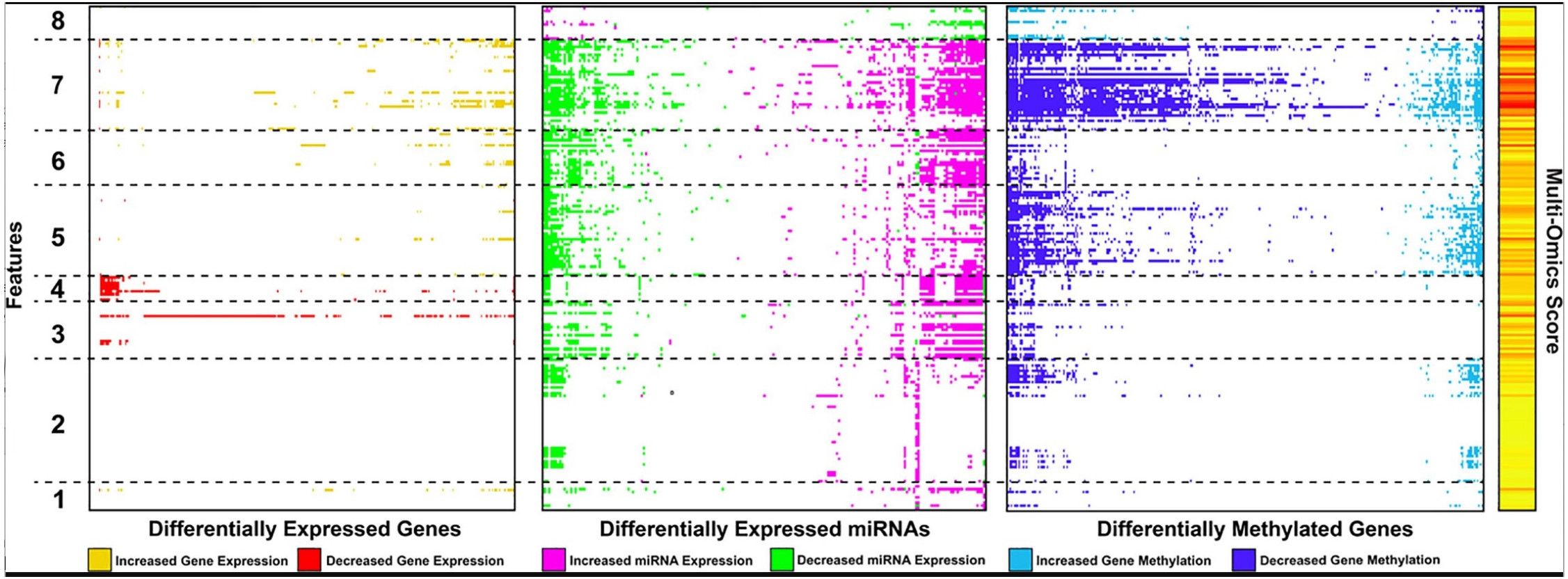
Fields & Techniques of Bioinformatics

- **Structural Bioinformatics**
- **Sequence Analysis (Genomics)**
- **Networks and Systems Biology**
- Signal and Image Processing
- Biomedical Informatics
- Database Curation

Systems Biology

- Integrate multiple omic techniques to minimize the high false positive rate of of individual techniques
- Commonly coupled with mathematical data modeling based on perturbation studies
- Typically, present the data as a network of interactions
- Networks can be easily annotated to add extra molecular and non-molecular data (e.g., ptn-ptn interactions and phenotype similarity).

Systems Biology

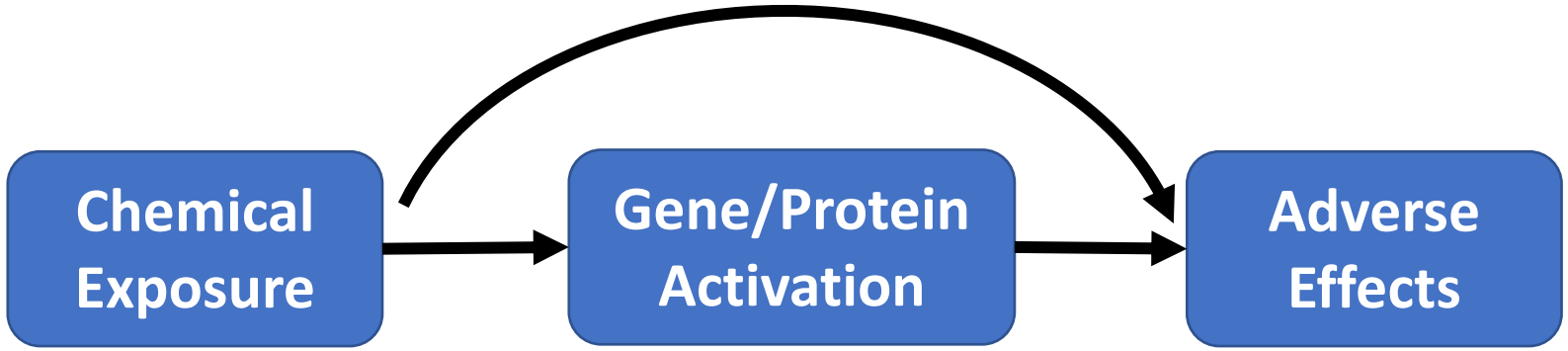


Chao et al. Integrative exposomic, transcriptomic, epigenomic analyses of human placental samples links understudied chemicals to preeclampsia. Environment International 2022.

Integration of Adverse Outcome Pathways, Causal Networks and 'Omics to Support Chemical Hazard Assessment

Edward J. Perkins[1], E. Alice Woolard[2] and Natàlia Garcia-Reyero[1]*

- Molecular events for regenerative proliferation.
- Three carcinogens and two non-carcinogens.

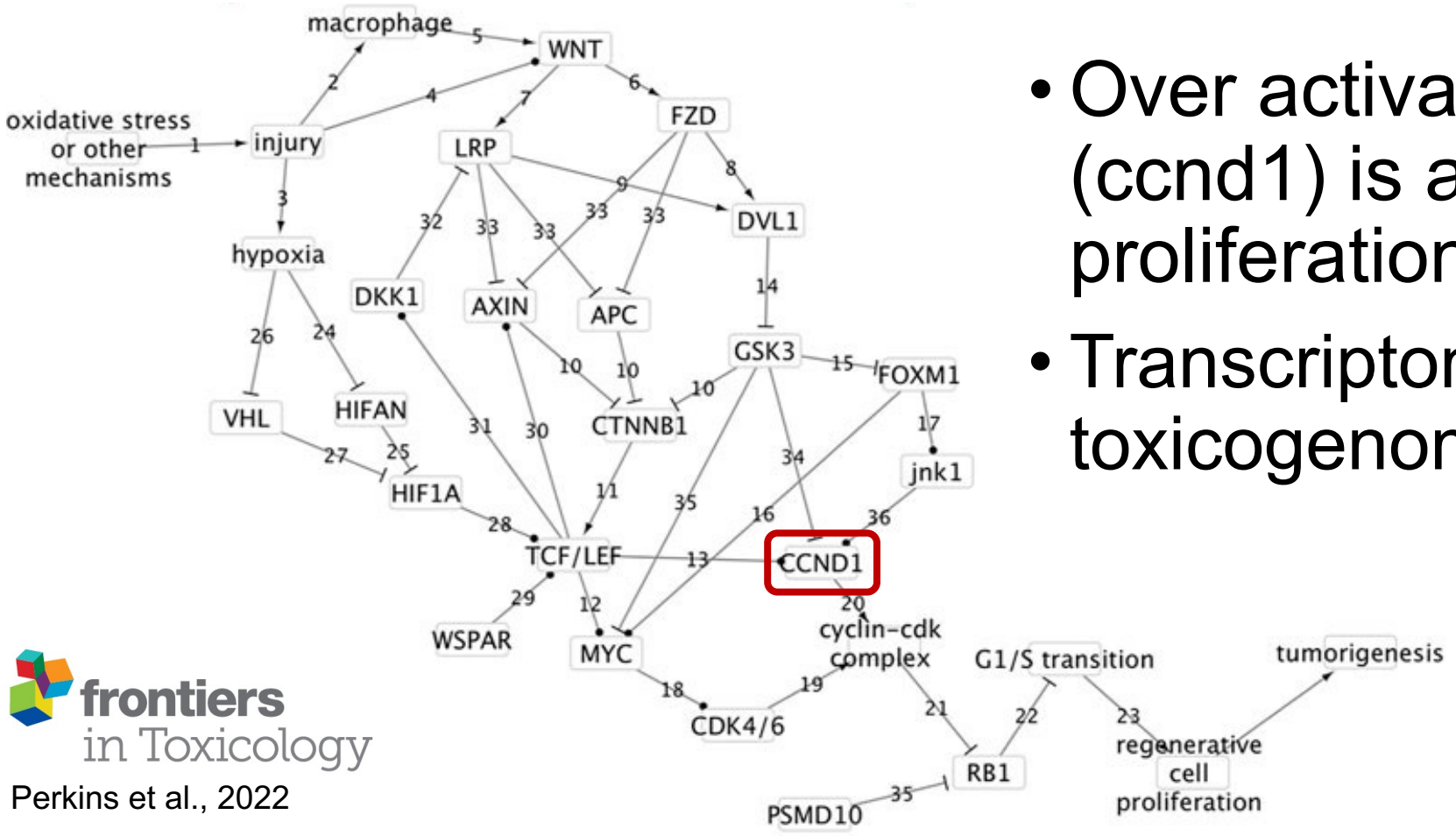


frontiers in Toxicology

Perkins et al., 2022

Integration of Adverse Outcome Pathways, Causal Networks and 'Omics to Support Chemical Hazard Assessment

Edward J. Perkins[1], E. Alice Woolard[2] and Natàlia Garcia-Reyero[1]*

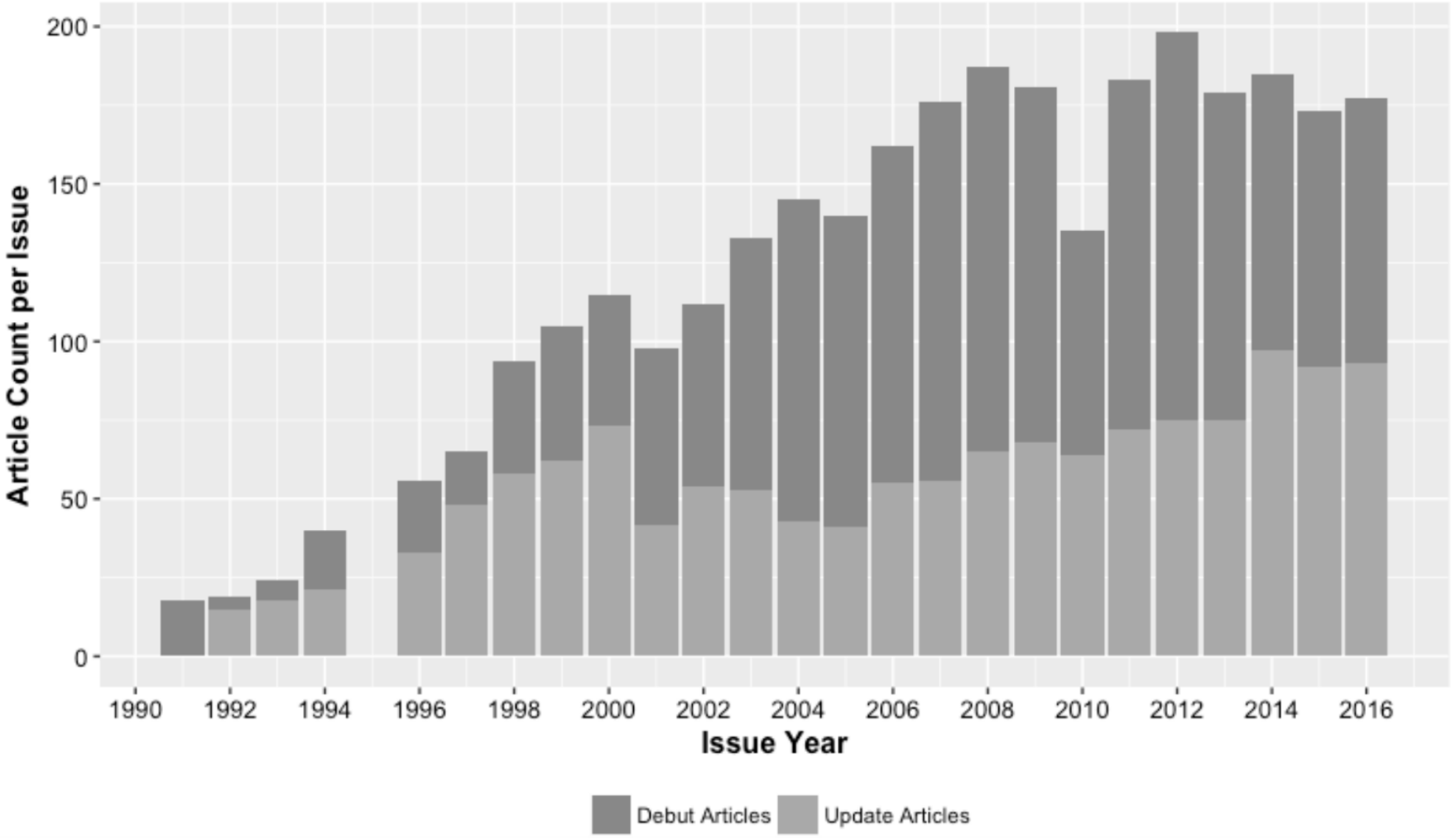


- Over activation of Cyclin D1 (ccnd1) is a MIE for regenerative proliferation.
- Transcriptomic data from a toxicogenomics database.

frontiers in Toxicology

Perkins et al., 2022

Bioinformatics Databases



Heidi J. Imker. 25 Years of Molecular Biology Databases: A Study of Proliferation, Impact, and Maintenance. BioRxiv preprint first posted online Mar. 9, 2018;

Bioinformatics Databases

Molecular databases

- Gene DB (at NCBI) and Ensembl DB (at the European Bioinformatics Institute)
- Uniprot database of proteins
- Human Metabolome Database (HMDB)

Sequence variant databases

- dbSNP: DB of Short Variants (e.g., SNP, and small indels).
- dbVar: DB of Structural Variants (e.g., large indels, translocations and inversions)
- ClinVar: For relationships between human variants and health status with supporting evidence.

Phenotype databases

- OMIM: NCBI database for human genes and genetic phenotypes
- Disgenet: Integrates data from expert curated repositories, GWAS catalogues, animal models and the scientific literature.

Bioinformatics Databases

- **Drug information portal**
- **SIDER**
- **DrugBank:** drugs and drug targets (genes and pathways)
- **SMPDB (Small Molecule Pathway Database)**
- **Comparative Toxicogenomics Database**

Drug Information Portal

Quick Access to Quality Drug Information

NIH National Library of Medicine

Home | News and Features | NLM Resources | NLM Research Resources | Resources by Audience / Class | Other Resources

Home ▸ Substance

Search DRUG INFORMATION PORTAL

◉ By Name ○ By Category

ACETAMINOFEN Go

1 result for Name/Synonym equals ACETAMINOFEN

Drug Name: Acetaminophen [USP:JAN] View Synonyms View Structure

Description: Analgesic antipyretic derivative of acetanilide. It has weak anti-inflammatory properties and is used as a common analgesic, but may cause liver, blood cell, and kidney damage.

Categories: Analgesics Show more categories

Summary

- **Summary of drug information (MedlinePlusDrug)**
- **Summary of consumer health information (MedlinePlusTopics)**
- **Summary of the effect on breastfeeding (LactMed)**
- **Summary of Drug-Induced Liver Injury (LiverTox)**
- **Manufacturers drug label (DailyMed)**
- **Clinical trials (ClinicalTrials.gov)**

SIDER 4.1 | Home | Drug list | Side Effects | Download | About

SIDER 4.1 : Side Effect Resource

SIDER contains information on marketed medicines and their recorded adverse drug reactions. The information is extracted from public documents and package inserts. The available information include side effect frequency, drug and side effect classifications as well as links to further information, for example drug–target relations.

Search for drugs or side effects :

type 3 or more characters...

Database statistics

Number of drugs and side effects

# of SE	# of drugs	# of drug-SE pairs	Pairs with frequency information
5868	1430	139756	39.9%

Number of drug–side effect pairs in different frequency ranges

	frequent (with exact data)	infrequent (with exact data)	rare (with exact data)	postmarketing	total
drug	24562 (23601)	16765 (11426)	11784 (6013)	19265	55730
placebo	7133 (7133)	3294 (3294)	2512 (2512)	0	10748

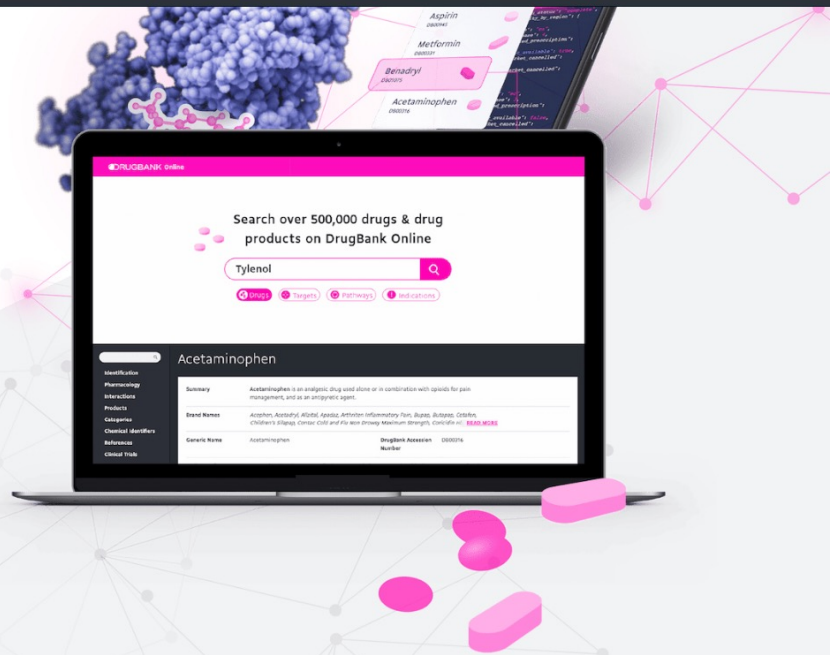
DRUGBANK Online | Browse | COVID-19 | Search | Interaction Checker | Downloads | Solutions | About

Learn to leverage the **Clinical API: Therapeutic Alternatives** Read Blog!

Building the foundation for better health outcomes

Access the right information at the right time, with our intelligent clinical drug data API and in-depth knowledge database.

Learn about our solutions >



Search over 500,000 drugs & drug products on DrugBank Online

Tylenol

Drugs | Targets | Pathways | Indications

Welcome to the

Small Molecule Pathway Database

Version 2.0

Brought to you by the creators of the Human Metabolome Database (HMDB) and DrugBank:

SMPDB (The Small Molecule Pathway Database) is an interactive, visual database containing more than 30 000 small molecule pathways found in humans only. The majority of these pathways are not found in any other pathway database. SMPDB is designed specifically to support pathway elucidation and pathway discovery in metabolomics, transcriptomics, proteomics and systems biology. It is able to do so, in part, by providing exquisitely detailed, fully searchable, hyperlinked diagrams of human metabolic pathways, metabolic disease pathways, metabolite signaling pathways and drug-action pathways. All SMPDB pathways include information on the relevant organs, subcellular compartments, protein_complex cofactors, protein_complex locations, metabolite locations, chemical structures and protein_complex quaternary structures. Each small molecule is hyperlinked to detailed descriptions contained in the HMDB or DrugBank and each protein_complex or enzyme complex is hyperlinked to UniProt. All SMPDB pathways are accompanied with detailed descriptions and references, providing an overview of the pathway, condition or processes depicted in each diagram. The database is easily browsed and supports full text, sequence and chemical structure searching. Users may query SMPDB with lists of metabolite names, drug names, genes/protein_complex names, SwissProt IDs, GenBank IDs, Affymetrix IDs or Agilent microarray IDs. These queries will produce lists of matching pathways and highlight the matching molecules on each of the pathway diagrams. Gene, metabolite and protein_complex concentration data can also be visualized through SMPDB's mapping interface. All of SMPDB's images, image maps, descriptions and tables are downloadable.

Get started now: ★ Browse Pathways ★

Comparative Toxicogenomics Database

ctd *Illuminating how chemicals affect human health.*

Comparative Toxicogenomics Database

Home | Search | Analyze | Download | Commercial Users | Help

Connect. Compare.

CTD is a robust, publicly available database that aims to advance understanding about how environmental exposures affect human health. *More...*

Discover.

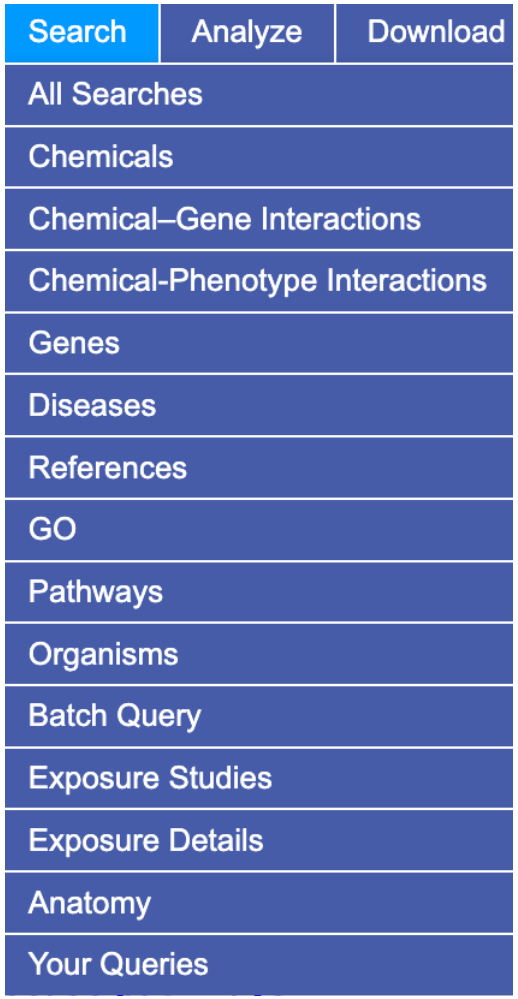
1. What human diseases are associated with a gene/protein? (Example)
2. What human diseases are associated with a chemical? (Example)
3. What genes/proteins interact with a chemical? (Example)
4. What chemicals interact with a gene/protein? (Example)
5. What references report a chemical–gene/protein interaction? (Example)
6. What cellular functions (GO terms) are affected by a chemical? (Example)

News

- *October 31, 2022*
  New data available!

http://ctdbase.org/

Comparative Toxicogenomics Database



Batch Query
Download custom data associated with a set of chemicals, diseases, genes, Gene Ontology terms, pathways, or references.

Set Analyzer
Perform analyses such as set-based enrichment for collections of chemicals or genes, and pathway generation for collections of genes.

MyGeneVenn
Compare your gene list to genes associated with up to two chemicals or diseases.

MyVenn
View relationships among your lists of CTD chemicals, diseases, genes, GO terms or pathways, or any other data.

VennViewer
Compare associated data sets for up to three chemicals, diseases, or genes.

CTD Tetramers
Download CTD Tetramers, which are computationally predicted information blocks, derived by integrating five direct evidence statements found in CTD.

http://ctdbase.org/

Thank you for your attention

Tamer A. Mansour (tahmed@ucdavis.edu)
Ahmed Abdelmoneim (aahmed10@lsu.edu)