Computational approach for enhanced RNA-sequencing analysis of hepatotoxic compounds

Benedetta Leoni[1*], Tamara Krsmanovic[1], Katica Jankovic[1], Jonathan Frampton[1], Gordana Apic[1], Robert B. Russell[2]

[1] Metisox, Cambridge, UK; [2] University of Heidelberg, Cell Networks, Heidelberg, Germany, * blm@metisox.com

Objectives

The recent years have seen increased interest in the field of toxicogenomics, as a way of predicting and understanding compound toxicity, with the aim to reduce the need for animal testing and improve the high clinical attrition rates of new drug candidates. In particular, gene expression data plays a big part in this field thanks to the rapid development of methods for its production and analysis[1].

The prediction of toxicity from RNA-sequencing or microarray data, however, remains quite difficult, as the output usually consists of just a list of dysregulated genes. While it is possible to extract Gene Ontology (GO) terms and dysregulated pathways from this list of genes through enrichment analysis, it remains challenging to interpret the results and link the observed gene expression signatures to precise and relevant toxicological endpoints.

In this study we have built a computational tool that allows us to easily find gene expression signatures from gene expression data and correlate them with real toxicities; we have then validated the system using gene expression data of primary human hepatocyte (PHH) spheroid cultures[2], with the dual aim to find a gene expression signature for hepatotoxicity following the administration of any of three hepatotoxic compounds, namely amiodarone, aflatoxin and chlorpromazine, and to correlate this gene expression signature to liver toxicological endpoints.

Methods

In order to make the output of RNA-sequencing easier to interpret in terms of toxicology,we created a computational model of biological pathways by manually annotating and processing molecular information from the literature from the public domain including PubMed articles, FDA reports, UniProt and RefSeq annotations, PubChem and IntAct information.

Our expert knowledge system utilizes a network-based approach that considers prior knowledge of genes, proteins and chemicals and their associations with each other and to our proprietary ontology that includes 2500 organ and tissue pathologies and pathways (Figure 1).

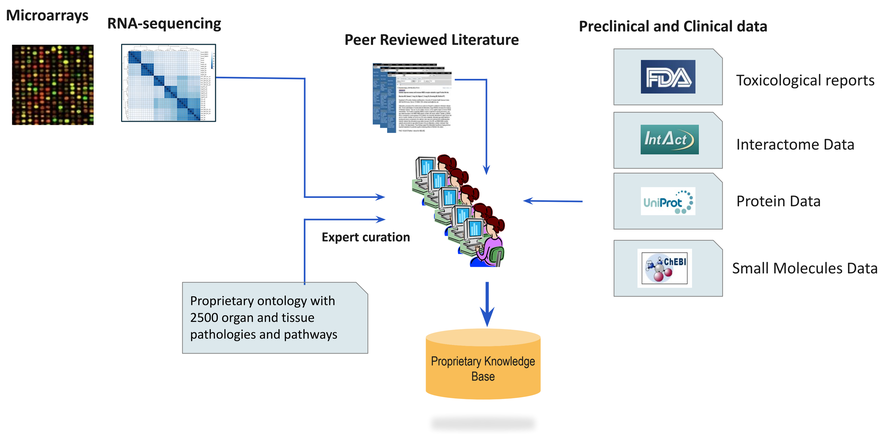

Figure 1. Manually curated database. Data was extracted from many diverse sources. Database contains complete human interactome, 2.500 organ and tissue pathologies, 500.000 synonyms/ontologies, 4 million linked articles (PubMed, FDA, clinical, patents), >10 million relationships, 50.000 biologically active chemicals.

Using this developed expert knowledge system we analyzed gene expression data of primary human hepatocyte (PHH) spheroid cultures[2].

Results

We curated and processed the knowledge on protein function, distribution and involvement in specific pathologies. The available pathological data for each target were classified into the hierarchical structure of an ontology of cell types, tissues and organs allowing for detailed mechanistic hypothesis testing.

We validated the system by exploring transcriptomic datasets from primary human hepatocyte spheroids treated with three hepatotoxic compounds: chlorpromazine, aflatoxin and amiodarone. We identified 21 genes that were dysregulated when any of these compounds was administered, making them good candidates as a signature of hepatotoxicity (Figure 2).

As it is often the case with a small gene set spread across several heterogeneous samples, the conventional tools for enrichment analysis returned scarce and very generic results. In particular, there were no results reported by Gene Ontology, while Reactome reported a list of quite generic dysregulated pathways from which no clear toxicities could be easily extracted. Conversely, our proprietary database and algorithm correlated these genes to several liver toxicity endpoints (Table 1).

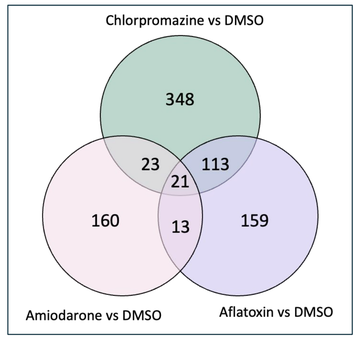

Figure 2. Identification of 21 genes dysregulated when any of the three hepatotoxic compounds are administered to PHH spheroid cultures.

Hepatotoxicity endpoint	Genes from signature
Liver steatosis	CSF1R, FGF21, G0S2, PDK4, SERPINE1
Liver inflammation	FGF21, G0S2, SERPINE1
Liver fibrosis	FGF21, G0S2, SERPINE1

Table 1. A sample of hepatotoxicities associated to the gene expression signature.

The system allows fast retrieval of literature references that report toxicologically relevant information about signature genes. For example, the FGF21 gene from our signature is consistently downregulated across the samples (Figure 3). The literature review confirmed downregulation in several liver conditions, including hepatic fibrosis, and treatment with FGF21 is a putative therapy for liver disease[3].

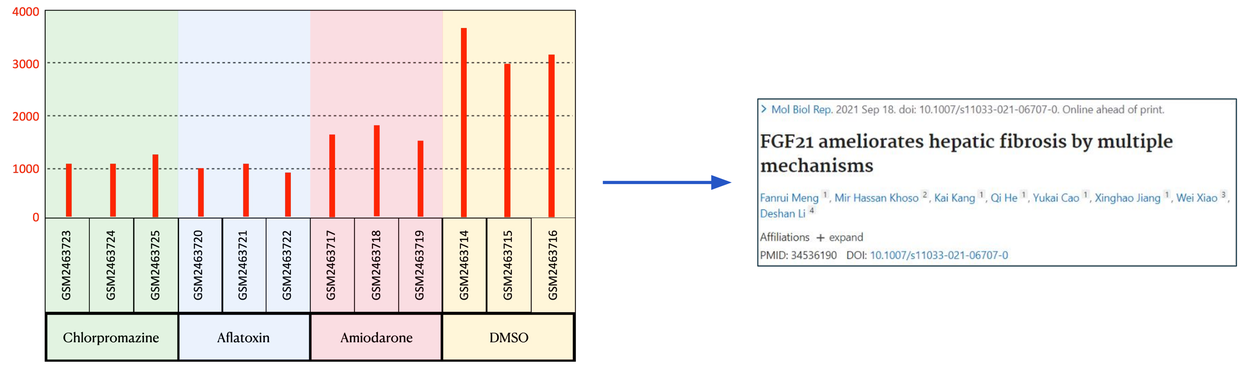

Figure 3. Downregulation of FGF21 in samples treated with hepatotoxic compounds compared to control samples (DMSO). Downregulation of FGF21 has been associated with amelioration of hepatic fibrosis[3].

Conclusion

We have developed a manually curated database of proteins biology, chemical entities and their involvement in specific pathologies. The developed system allows us to easily find and analyze gene expression signatures. By applying bioinformatics analysis tools to gene expression data and combining with toxicants data, we have connected the hepatotoxicity gene expression signature to several liver toxicity endpoints and have obtained more information on single genes of interest.

References

[1] Alexander-Dann *et al.*, 2018, Developments in Toxicogenomics: Understanding and Predicting Compound-Induced Toxicity from Gene Expression Data. *Molecular Omics* 14, n. 4 (2018): 218–36.

[2] Bell *et al.*, 2017, Transcriptional, Functional, and Mechanistic Comparisons of Stem Cell–Derived Hepatocytes, HepaRG Cells, and Three-Dimensional Human Hepatocyte Spheroids as Predictive In Vitro Systems for Drug-Induced Liver Injury. *Drug Metabolism and Disposition* 45, n. 4 (April 2017): 419–29.

[3] Meng *et al.*, 2021, FGF21 Ameliorates Hepatic Fibrosis by Multiple Mechanisms. *Molecular Biology Reports* 48, n. 11 (November 2021): 7153–63.