

Exploiting advances in transcriptomics to improve on human-relevant toxicology

T.A. Darde, F. Chalmel, T. Svingen

▶ **To cite this version:**

T.A. Darde, F. Chalmel, T. Svingen. Exploiting advances in transcriptomics to improve on human-relevant toxicology. Current Opinion in Toxicology, 2018, 11-12, pp.43-50. 10.1016/j.cotox.2019.02.001 . hal-02161231

HAL Id: hal-02161231

https://univ-rennes.hal.science/hal-02161231

Submitted on 8 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACCEPTED MANUSCRIPT

Exploiting advances in transcriptomics to improve on human-relevant toxicology

Thomas A. DARDE[1], Frédéric CHALMEL[1] and Terje SVINGEN[2,*]

1. Université de Rennes, Inserm, EHESP, Irset (Institute de recherche en santé, environnement et travail) – UMR_S1085, F-35000 Rennes, France

2. Division of Diet, Disease Prevention and Toxicology, National Food Institute, Technical University of Denmark, Kemitorvet Building 202, Kgs. Lyngby DK-2800, Denmark

* Corresponding author: tesv@food.dtu.dk

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Modern toxicology is challenged in its aspiration to test and risk-assess the myriad of chemicals humans potentially can be exposed to. We still lack sufficient data to perform proper human risk assessments for a large proportion of environmental chemicals and we still know very little about many of the molecular mechanisms that cause adverse effects. In turn, lack of mechanistic insight stalls the development of more cost-efficient and high-throughput alternative assays, which is a prerequisite if we are to deal with the large number of untested compounds. One way to help speed up the effort is to take advantage of the many advances in genomics technologies and apply them to toxicity testing strategies, or at the very least use them to better characterize the causative molecular mechanisms. For instance, single-cell Digital Gene Expression-(DGE-) and single-cell RNA (scRNA-) sequencing techniques hold great promise for not only enabling the analyses of large sample sizes at low cost, but also capturing toxico-molecular events and genomic susceptibilities at the cellular level. In this paper, we discuss some of the advances in transcriptomics and how they can be applied to toxicology to advance human-relevant risk assessment.

Key words: toxicogenomics; risk assessment; mechanisms; RNA-seq; toxicology

Introduction

Contemporary toxicology is grappling with increasing demands for faster, cheaper and more reliable test strategies. There is also an increasing demand to develop and incorporate more human-relevant models and assays when assessing the risk to human health; and ideally, real-life exposure scenarios and sub-group susceptibilities should also be accounted for [1]. These are all individually big challenges, but combined they represent an enormous challenge that cannot be solved using traditional toxicology approaches of assessing single chemicals one at a time by standard testing regimens. This is especially true if we considering the tens-of-thousands of chemicals that are continuously being manufactured and used across the world, chemicals for which we still lack sufficient data to perform proper risk assessments. In fact, global chemical pollution has been considered so immense and complex that its impact on the planet and its inhabitants is beyond quantifiable [2]. To help solve some of these challenges, modern genomics methods can be combined with toxicology approaches to assess toxicity in new ways – a modern field coined toxicogenomics.

The term toxicogenomics first appeared in the late 1990s. It heralded a new era for chemical toxicology by advancing our understanding of mechanisms of toxicity and thereby predictive capabilities [3]. In 2007, the American National Research Council defined toxicogenomics as "combin[ing] toxicology with information-dense genomic technologies to integrate toxicant-specific alterations in gene, protein and metabolite expression patterns with phenotypic responses of cells, tissues and organisms" [4]. The main hypothesis was that an organismal response to a toxicant leading to pathological or phenotypical changes in certain organs at a low dose could be detected as changes to gene, protein, or metabolite expression. The field

has yet to deliver on all its promises, but has come a long way. Currently gene expression is the most sensitive of the above-mentioned parameters, also with the highest cost-benefit ratio. Consequently, toxicogenomics has by some been defined more narrowly to only include ‘‘the study of altered gene expression after toxicant exposure’’ [5].

Irrespective of current state of the science, toxicogenomics has a broad application potential and can be used for mechanistic analyses, biomarker discovery, risk assessments and drug development, and so forth. And perhaps more importantly, toxicogenomics holds the potential to deliver more discriminate, sensitive and predictive information with regard to pharmaceuticals and environmental chemicals and their potential effects on human health. A looming challenge, however, comes from the successes achieved with genomics technologies that now are capable of retrieving mass amount of data all the way down to the scale of single cells. Unless handled appropriately, such terabytes of information can serve to obscure rather than illuminate the picture; or to quote Cervante’s immortal character Don Quixote, “facts [can become] the enemy of truth”.

Transcriptomics for human-relevant toxicity testing

Although genomics can be applied to toxicology in numerous ways, here we will focus on where transcriptomics can, and already do, make an impact on human-relevant chemical toxicity. Three important areas are mechanistic insight, biomarker identification, and mixture risk assessment. One: A recent comparative case study on human risk to benzo(a)pyrene in drinking water found that genomics-informed assessments yielded comparative points-of departure (POD) values as did traditional assessment approaches; however, toxicogenomics could also be used to delineate more detailed modes-of-action in

the mouse toxicity model and asserting human relevance by showing consistency in perturbed pathways between the mouse tissue and expression data from human cells [6]. Two: A recent report by Makarov and colleagues presented an efficient method for identifying reliable biomarkers using toxicogenomics datasets. Using two different algorithms ('search marker' and 'recognition algorithm') on the well-known toxicogenomics dataset DrugMatrix [7], they were able to identify genomic patterns specific to chemical compounds regardless of dose or time. Three: Several studies have attempted to implement mixture modelling in transcriptomics analysis to better understand the interactions of combined chemicals [8-10], a topic considered of utmost relevance if we are to appreciate the true impact environmental chemicals can have on human health [11, 12]. Although our understanding of the molecular and cellular responses to complex mixtures remains lacking – and the use of classical mixture toxicity models to predict mixture effects from toxicogenomics data is still in its infancy [13] – we predict great advances in this area within the near future.

Genomics in the twenty-first Century

Modern genomics has seen enormous advances within a relatively short space of time. What seemed like impossible less than two decades ago is now possible even for a small research laboratory. Today one can sequence a person's genome in two days for less than $4,000, as well as analyse the transcriptome of single cells from complex tissues. In comparison, the first human genome took 13 years to complete at a cost of upwards of $1 billion [14]. Hence, the main challenge seems no longer to be what data we can extract, but what to do with the information once it has been collected.

Advances in ultra-high-throughput sequencing technologies have greatly reduced the overall cost of running these assays, and thus also greatly boosted their use [15]. Today, RNA sequencing (RNA-seq) is the dominant technique in transcriptomics and has several advantages over other approaches such as the early Sanger sequencing or microarray-based methods. Indeed, RNA-seq has quickly proven to produce accurate quantification of gene expression and to correlate well with other technologies to eventually outperform the alternative methods in terms of accurately quantifying low-abundance transcripts [16-18]. Because of the many advantages coinciding with parallel development of more sophisticated software and algorithms to handle ever larger data sets, RNA-seq has opened up possibilities for comprehensive characterization of the genomic landscapes to the point of transcriptome data at single-cell resolution by Single-cell RNA-sequencing (scRNA-seq). This has made it possible to catalogue the expression of individual genes in individual cells from a given tissue with reasonable accuracy. This powerful tool can ultimately reveal the inherent variability between cells, both in normal and disease states [19]. These advances can give access to a vast amount of information that will allow us to address a myriad of scientific questions regarding cell differentiation and lineage specification, cellular heterogeneity, and development of disease from earliest of stages, all within one tissue from one individual. The potential applications to biology, toxicology and medicine are enormous.

Another recent development in genomics is Digital Gene Expression (DGE) profiling, a variant of the SAGE method using ultra-high-throughput sequencing techniques [20]. Simply put, DGE can be used to quantify levels of expression by directly counting reads from raw data obtained from sequencing the short fragments located at the 3' poly(A) of RNA

messenger, as opposed to RNA sequencing which covers the entire genome. This technique, originally developed to study scRNA-seq or small cell populations, provides a reliable estimate of the gene expression level at extremely low costs – €30-50 per sample – compared with conventional "bulk" RNA sequencing [21].

Together these emerging genomics technologies can be applied across many biomedical research fields, including diagnostics [22], disease profiling [23], gene function and annotation [24], as well as response to environment and toxicants [25]. Below, we will discuss more how transcriptomics can be used in toxicology to both advance our knowledge of modes of action and perform better risk assessments at lower costs than previously possible.

Toxicogenomics in the twenty-first Century

Depending on endpoint to be evaluated, many toxicity assessments use intact animals or cell cultures to measure several parameters, including morphometric and histopathological evaluations, cell viability, cytotoxicity, hepatotoxicity, or nephrotoxicity. These classical approaches can detect toxicity and be used to assess points-of-departure levels (POD) for specific chemicals. Data generated by omics technologies on the other hand, are envisioned to detect toxicities that may not be observable by conventional assessment strategies, thereby facilitating more accurate and predictive decision-making based on toxicity mechanisms [26]. It comes with certain challenges, however, for example lack of uniform study designs, multiplicity of normalization and analysis strategies [27], reproducibility of microarray data across platforms [28], absence of data quality control measures and standards, and lack of effective data sharing and reporting standards.

The ability to profile global transcriptional responses has become a valuable resource in toxicology. One challenge, however, is that *in vivo* toxicity studies that include transcriptomics rarely exceed 10-12 samples. It is close to impossible to detect robust toxicogenomic signatures in humans with such small samples sizes [29, 30]. To overcome this problem, a viable alternative is to use low-cost sequencing technologies such as DGE to allow for the study of larger collection of samples in order to detect stronger toxicogenomic signatures with greater statistical power. This would also help discriminate sub-groups of individuals based on their susceptibility to disease, making it possible to identify common constitutional factors.

Susceptibility to an adverse effect may also exist within cellular subpopulations within a defined tissue. Classical "bulk" transcriptomic technologies cannot detect this phenomenon at the organ level. Recent advances in scRNA-seq approaches make it possible to study the impact of chemical exposure at the level of cellular subpopulation, as well as to identify specific biomarkers [31] to enhance the toxicological characterization and underlying mechanisms at the cellular level. Single-cell technologies have already been used successfully in studies on tumor heterogeneity. In a landmark study by Dalerba and colleagues [32], scRNA-seq revealed that multi-lineage cell differentiation determines the transcriptional diversity of cancer tissues. With regard to toxicology, scRNA-seq has been highlighted as a potential powerful approach to discover new biomarkers that reflect the efficacy, metabolism, pharmacodynamics, absorption and toxicology of chemicals [31].

Using toxicogenomics to bridge the gaps between rodent and human data

For more than a century, rodents have been indispensable for experimental medicine, drug development and toxicology. Rats and mice continue to serve as model organisms in drug design, biomarker discovery and chemical risk assessment, but the use of animal models is also under increasing pressure, both from an ethical point of view and for not necessarily providing human-relevant data. Several studies have shown that rats and humans share a high degree of genomic and physiologic similarities [33, 34], but simultaneously, important differences have been reported, including functional differences within non-pharmacokinetic metabolism. Such differences could influence how and if a compound induce toxicity or elevate a biomarker [35, 36]. It is thus paramount that we characterize any differences between rodents and humans if we are to accurately interpret animal studies used for biomarker discovery, toxicity testing or comparative toxicogenomics analyses [37, 38]. To this, transcriptomic analyses can help reveal inter-species similarities and differences [39].

Efforts have been put towards setting up large, comprehensive collections of comparable rodent and human transcriptomic data. New computational methods for integrating gene expression measurements into genome-scale network reconstruction have also been developed. ‘Systems toxicology’ approaches rely on the integration of multiple toxicology data (e.g. transcriptomic, metabolomic, epigenomic) with quantitative analysis of large networks of molecular and functional changes occurring across multiple levels of biological organization. It can be used to study the full spectrum of toxicological responses from xenobiotic insult, as well as enable comparative predictions across individual patients, treatment conditions, tissue types and species [40, 41]. These efforts have been made possible by the implementation of services for toxicogenomics data management.

Potential impact of toxicogenomics on 'uncertainty factors' in chemical risk assessment

In order to account for uncertainties when using effect assessment data for human chemical risk assessment, so-called assessment factors (AF; also referred to as uncertainty factors) are used to take into account inter- and intra-species variabilities and uncertainties [42]. Although the 'default' AF of 100 is often referred to (10x to account for differences in experimental animal to humans; 10x to account for individual differences in human population), there are in fact many different AFs employed within regulatory frameworks depending on experimental design and substances that are tested [42, 43]. AF values must be adjusted in cases where information is available that allows for greater certainty. Such amended AFs can be contentious however, and a source of debate [43]. In any case, it is likely that toxicogenomics can impact how AFs are determined in the future. For instance, by mapping the genetic differences between populations associated with susceptibility to a given disease, concomitant with elucidation of the detailed regulatory pathways that are affected by specific chemicals, there will be much higher certainty in the use of effect data for calculating health-based guidance values. Hence, the AF for intra-species (i.e. human) differences could be set much lower than the default 10x. This line of though can also be extended to default AFs for inter-species variations. By gaining more knowledge about the genomic landscape and how different species respond to the exposome, it is likely that more precise AFs can be set for individual species (e.g. rat), or even strains of the same species, depending on how well they relate to humans for any given effect endpoint.

Toxicogenomics data management and online services

The sheer volume of available transcriptomic data has now grown so large that it starts to impose additional challenges to research communities, including toxicology; namely the complex tasks of processing, hosting and interpreting data [44]. Among the many practical points that should be considered when funding, planning and reporting toxicology studies in the future is improved data transparency [30, 45]. This should also include negative (read: no effect) data or contradictory results. The European Commission has started to address this issue by attempting to establish a model service for systems biology data management through the European Research Infrastructure Consortium (ERIC). Its objective is to make biological data FAIR: Findable, Accessible, Interoperable and Reusable. This demands that investigators make raw data publicly available. Although great efforts have been made to store raw data in public repositories [46, 47], the all-important processed data (e.g. list of regulated genes) remains inaccessible to the scientists (i.e. supplementary files provided by authors).

To meet a growing demand of accessing pre-processed omics data without having to download and re-process raw data, several on-line resources have been developed (Table 1). For example, the CTD [48], diXa [49], ToxDB [50], CEBS [51], Drug2Gene [52] and ToxCast [53], which have paved the way for improved storage, exchange, and analysis of toxicological data [54]. More recently, we launched TOXsIgN [55] as an innovative platform designed to share standardized processed data that also provide tools for predictive toxicology. The success of this and other repositories, however, ultimately depends on researchers' compliance to the FAIR principles. But conversely, the success of FAIR depends on the availability of suitable repositories such as TOXsIgN. In essence, we encourage the uploading of toxicogenomic and toxicological signatures in dedicated repositories such as

TOXsIgN before submission of manuscripts for publication. This would increase both their visibility and accessibility.

Table 1: List of toxicogenomics web resources

Name	**Ressource type**	**Address**	**Ref**
Open TG-GATEs	Database	http://toxico.nibio.go.jp/english/index.html	[56]
Drug Matrix	Database	https://ntp.niehs.nih.gov/drugmatrix/index.html	[57]
ChemDIS	Tool	https://cwtung.kmu.edu.tw/chemdis/	[58]
Toxygates	Tool	http://toxygates.nibiohn.go.jp/toxygates/	[59]
CEBS	Tool	https://www.niehs.nih.gov/research/resources/databases/cebs/index.cfm	[51]
ChemProt	Tool	http://potentia.cbs.dtu.dk/ChemProt/	[60]
LTMap	Tool	http://tcm.zju.edu.cn/ltmap	[61]
ToxDBScan	Tool	http://www.ra.cs.uni-tuebingen.de/software/ToxDBScan/welcome_e.html	[62]
Drug2Gene	Tool	http://www.drug2gene.com	[52]
Comparative Toxicogenomics Database (CTD)	Tool / Database	http://ctdbase.org	[41]
The Connectivity Map	Tool / Database	https://clue.io/cmap	[63]
LINCs	Tool / Database	http://www.lincsproject.org/	[64]
diXa	Tool / Database	http://www.dixa-fp7.eu	[49]
ToxDB	Tool / Database	http://toxdb.molgen.mpg.de	[50]
TOXsIgN	Tool / Database	http://toxsign.genouest.org	[55]
ToxCast	Tool / Database	https://www.epa.gov/chemical-research/toxicity-forecasting	[53]
NFFinder	Tool / Database	http://nffinder.cnb.csic.es/	[65]

Perspectives on modern toxicogenomics

Less than two decades ago, proposing to sequence the transcriptome at single cell-resolution in a complex organ and use this data to extract a personalized toxicogenomic profile to predict future effects from exposure to a real-life mixture of environmental chemicals, would likely have been considered quixotic. With recent advances in genomics and computer technologies, however, this proposition does no longer seem exceedingly idealistic or delusional. Massive amounts of gene expression data are now available in the public domain, enabling new biological questions to be addressed through data re-use without the need for further experimentation [56, 57, 63, 64]. And we are optimistic that the future of toxicogenomics will deliver on many of its promises and thereby also aid contemporary toxicology to deliver on increasing demands from governing bodies and the public to safeguard human health against xenobiotic insults at ever lower costs and better predictability. To end, we like to highlight five points where we believe toxicogenomics will have significant impact in the coming years.

- Enhance our knowledge about mechanism of action at the molecular and cellular levels induced by exposure to xenobiotics.
- Enable the progressive implementation of the 3R-principles (replace, reduce, refine) for animal toxicity studies, not least by helping to verify the suitability of emerging technologies such as human-based organoids and organs-on-chip. Genomics technologies will likely play a major role in future chemical risk assessment.
- Facilitate comparative studies between species, in particular with regard to human relevance and thereby enable more accurate extrapolations from model organisms to human risk assessment.

- Support the development of integrated approaches needed to advance the area of mixture risk assessment by offering more integrated knowledge about molecular and cellular responses to exogenous compounds.
- Allow for the identification of susceptible sub-populations or individuals with regard to effects caused by chemical exposures. This could also include refinements to the arbitrary ’10x uncertainty factor’ used in risk assessments.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Acknowledgements

We thank Dr Julie Boberg for helpful comments on the manuscript. This work was partially funded by the French agency for food and safety [ANSES n °EST-13-081 to F.C], the French National Research Agency [ANR n° 16-CE14-0017-02 to F.C], and the Fondation pour la recherche médicale [FRM n °DBI20131228558 to FC].

ACCEPTED MANUSCRIPT

References

1. F.S. Collins, G.M. Gray, and J.R. Bucher, Toxicology. Transforming environmental health protection. Science. 319 (2008) 906-907. DOI: 10.1126/science.1154619.

*2. J. Rockström, W. Steffen, K. Noone, A. Persson, F.S.3rd. Chapin, E.F. Lambin, T.M. Lenton, M. Scheffer, C. Folke, H.J. Schellnhuber, et al., A safe operating space for humanity. Nature. 461 (2009) 472-475. DOI: 10.1038/461472a.

A provokative review proposing finite planetary boundaries, including chemical pollution, that must not be transgressed by human activities if we are to avoid unacceptable changes to the environment.

3. E.F. Nuwaysir, M. Bittner, J. Trent, J.C. Barrett, and C.A. Afshari, Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog. 24 (1999) 153-159. DOI: 10.1002/(SICI)1098-2744(199903)24:3<153::AID-MC1>3.0.CO;2-P.

4. National Research Council (NRC), Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment (2007), National Academy Press: Washington DC. *Available from:* https://www.ncbi.nlm.nih.gov/books/NBK10219/pdf/Bookshelf_NBK10219.pdf

5. W.M. Haschek, C.G. Rosseaux, and M.A. Wallig, Haschek and Rousseaux's handbook of toxicological pathology. 3 ed (2013), London: Academic Press.

**6. I. Moffat, N. Chepelev, S. Labib, J. Bourdon-Lacombe, B. Kuo, J.K. Buick, F. Lemieux, A. Williams, S. Halappanavar, A. Malik, et al., Comparison of toxicogenomics and traditional approaches to inform mode of action and points of departure in human health risk assessment of benzo[a]pyrene in drinking water. Crit Rev Toxicol. 45 (2015) 1-43. DOI: 10.3109/10408444.2014.973934.

Seminal publication introducing transcriptomic as a starting point for benchmark dose analysis in human health risk assessment.

7. V. Makarov and A. Gorlin, Computational method for discovery of biomarker signatures from large, complex data sets. Comput Biol Chem. 76 (2018) 161-168. DOI: 10.1016/j.compbiolchem.2018.07.008.

8. F.A. Tilton, S.C. Tilton, T.K. Bammler, R.P. Beyer, P.L. Stapleton, N.L. Scholz, and E.P. Gallagher, Transcriptional impact of organophosphate and metal mixtures on olfaction: copper dominates the chlorpyrifos-induced response in adult zebrafish. Aquat Toxicol. 102 (2011) 205-215. DOI: 10.1016/j.aquatox.2011.01.012.

9. C.M. Hutchins, D.F. Simon, W. Zerges, and K.J. Wilkinson, Transcriptomic signatures in Chlamydomonas reinhardtii as Cd biomarkers in metal mixtures. Aquat Toxicol. 100 (2010) 120-127. DOI: 10.1016/j.aquatox.2010.07.017.

10. R. Altenburger, S. Scholz, M. Schmitt-Jansen, W. Busch, and B.I. Escher, Mixture toxicity revisited from a toxicogenomic perspective. Environ Sci Technol. 46 (2012) 2508-2522. DOI: 10.1021/es2038036.

11. T. Svingen and A.M. Vinggaard, The risk of chemical cocktail effects and how to deal with the issue. J Epidemiol Community Health. 70 (2016) 322-323. DOI: 10.1136/jech-2015-206268.

12. A. Kortenkamp and M. Faust, Regulate to reduce chemical mixture risk. Science. 361 (2018) 224-226. DOI: 10.1126/science.aat9219.

13. Y. Song, J. Asselman, K.A.C. De Schamphelaere, B. Salbu, and K.E. Tollefsen, Deciphering the Combined Effects of Environmental Stressors on Gene Transcription: A Conceptual Approach. Environ Sci Technol. 52 (2018) 5479-5489. DOI: 10.1021/acs.est.8b00749.
14. National Human Genome Research Institute, The cost of sequencing a human genome. (2016). *Available from:* https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/
15. D. Sims, I. Sudbery, N.E. Ilott, A. Heger, and C.P. Ponting, Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 15 (2014) 121-132. DOI: 10.1038/nrg3642.
16. F. Izadi, H.N. Zarrini, G. Kiani, and N.B. Jelodar, A comparative analytical assay of gene regulatory networks inferred using microarray and RNA-seq datasets. Bioinformation. 12 (2016) 340-346. DOI: 10.6026/97320630012340.
17. L. Chen, F. Sun, X. Yang, Y. Jin, M. Shi, L. Wang, Y. Shi, C. Zhan, and Q. Wang, Correlation between RNA-Seq and microarrays results using TCGA data. Gene. 628 (2017) 200-204. DOI: 10.1016/j.gene.2017.07.056.
18. C. Wang, B. Gong, P.R. Bushel, J. Thierry-Mieg, D. Thierry-Mieg, J. Xu, H. Fang, H. Hong, J. Shen, Z. Su, et al., The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol. 32 (2014) 926-932. DOI: 10.1038/nbt.3001.
19. S. Linnarsson and S.A. Teichmann, Single-cell genomics: coming of age. Genome Biol. 17 (2016) 97. DOI: 10.1186/s13059-016-0960-x.
20. S. Audic and J.M. Claverie, The significance of digital gene expression profiles. Genome Res. 7 (1997) 986-995. DOI: 10.1101/gr.7.10.986.
21. M. Soumillon, D. Cacchiarelli, S. Semrau, A. van Oudenaarden, and T.S. Mikkelsen, Characterization of directed differentiation by high-throughput single-cell RNA-Seq. bioRxiv, (2014). DOI: 10.1101/003236.
22. A.S. Nikitina, V.V. Babenko, K.A. Babalyan, A.O. Vasiliev, A.V. Govorov, E.A. Prilepskaya, S.A. Danilenko, O.V. Selezneva, and E.I. Sharova, [Primary candidate rna biomarker screening by RNA-seq for prostate cancer diagnostics]. Biomed Khim. 61 (2015) 781-784. DOI: 10.18097/PBMC20156106781.
23. V. Costa, M. Aprile, R. Esposito, and A. Ciccodicola, RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet. 21 (2013) 134-142. DOI: 10.1038/ejhg.2012.129.
24. T. Bischler, H.S. Tan, K. Nieselt, and C.M. Sharma, Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in Helicobacter pylori. Methods. 86 (2015) 89-101. DOI: 10.1016/j.ymeth.2015.06.012.
25. R. Yin, L. Gu, M. Li, C. Jiang, T. Cao, and X. Zhang, Gene expression profiling analysis of bisphenol A-induced perturbation in biological processes in ER-negative HEK293 cells. PLoS One. 9 (2014) e98635. DOI: 10.1371/journal.pone.0098635.
26. T. Urushidani, Prediction of hepatotoxicity based on the toxicogenomics database, in *Hepatoxicity: from genomics to in vitro and in vivo models*, S.C. Sahu, Editor (2008), John Wiley & Sons, Ltd.
27. J. Quackenbush, Microarray data normalization and transformation. Nat Genet. 32 Suppl. (2002) 496-501. DOI: 10.1038/ng1032.
28. N. Mah, A. Thelin, T. Lu, S. Nikolaus, T. Kühbacher, Y. Gurbuz, H. Eickhoff, G. Klöppel, H. Lehrach, B. Mellgård, et al., A comparison of oligonucleotide and cDNA-based

microarray systems. Physiol Genomics. 16 (2004) 361-370. DOI: 10.1152/physiolgenomics.00080.2003

**29. G.W. Miller, Improving reproducibility in toxicology. Toxicol Sci. 139 (2014) 1-3. DOI: 10.1093/toxsci/kfu050.

A thoughtful editorial arguing for data-driven rather than agenda-driven science in toxicology, with a keen focus on minimising biases in experimental design and data interpretations.

30. C.A. Poland, M.R. Miller, R. Duffin, and F. Cassee, The elephant in the room: reproducibility in toxicology. Part Fibre Toxicol. 11 (2014) 42. DOI: 10.1186/s12989-014-0042-8.

31. F. Niu, D.C. Wang, J. Lu, W. Wu, and X. Wang, Potentials of single-cell biology in identification and validation of disease biomarkers. J Cell Mol Med. 20 (2016) 1789-1795. DOI: 10.1111/jcmm.12868.

*32. P. Dalerba, T. Kalisky, D. Sahoo, P.S. Rajendran, M.E. Rothenberg, A.A. Leyrat, S. Sim, J. Okamoto, D.M. Johnston, D. Qian, et al., Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol. 29 (2011) 1120-1127. DOI: 10.1038/nbt.2038.

Landmark study using scRNA-seq to highlight that multi-lineage cell differentiation determines the transcriptional diversity in colon cancer.

33. R.A. Gibbs, G.M. Weinstock, M.L. Metzker, D.M. Muzny, E.J. Sodergren, S. Scherer, G. Scott, D. Steffen, K.C. Worley, P.E. Burch, et al., Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 428 (2004) 493-521. DOI: 10.1038/nature02426.

34. W. Mattes, K. Davis, E. Fabian, J. Greenhaw, M. Herold, R. Looser, W. Mellert, S. Groeters, H. Marxfeld, N. Moeller, et al., Detection of hepatotoxicity potential with metabolite profiling (metabolomics) of rat plasma. Toxicol Lett. 230 (2014) 467-478. DOI: 10.1016/j.toxlet.2014.07.021.

35. L. Luo, S. Schomaker, C. Houle, J. Aubrecht, and J.L. Colangelo, Evaluation of serum bile acid profiles as biomarkers of liver injury in rodents. Toxicol Sci. 137 (2014) 12-25. DOI: 10.1093/toxsci/kft221.

36. J.H. Shannahan, M.C. Schladweiler, J.H. Richards, A.D. Ledbetter, A.J. Ghio, and U.P. Kodavanti, Pulmonary oxidative stress, inflammation, and dysregulated iron homeostasis in rat models of cardiovascular disease. J Toxicol Environ Health A. 73 (2010) 641-656. DOI: 10.1080/15287390903578208.

37. T.A. Clayton, J.C. Lindon, O. Cloarec, H. Antti, C. Charuel, G. Hanton, J.P. Provost, J.L. Le Net, D. Baker, R.J. Walley, et al., Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature. 440 (2006) 1073-1077. DOI: 10.1038/nature04648.

38. T. Uehara, A. Ono, T. Maruyama, I. Kato, H. Yamada, Y. Ohno, and T. Urushidani, The Japanese toxicogenomics project: application of toxicogenomics. Mol Nutr Food Res. 54 (2010) 218-227. DOI: 10.1002/mnfr.200900169.

39. M.S. Rye, M.F. Bhutta, M.T. Cheeseman, D. Burgner, J.M. Blackwell, S.D. Brown, and S.E. Jamieson, Unraveling the genetics of otitis media: from mouse to human and back again. Mamm Genome. 22 (2011) 66-82. DOI: 10.1007/s00335-010-9295-1.

40. E.M. Blais, K.D. Rawls, B.V. Dougherty, Z.I. Li, G.L. Kolling, P. Ye, A. Wallqvist, and J.A. Papin, Reconciled rat and human metabolic networks for comparative

toxicogenomics and biomarker predictions. Nat Commun. 8 (2017) 14250. DOI: 10.1038/ncomms14250.

**41. A.P. Davis, C.J. Grondin, R.J. Johnson, D. Sciaky, B.L. King, R. McMorran, J. Wiegers, T.C. Wiegers, and C.J. Mattingly, The Comparative Toxicogenomics Database: update 2017. Nucleic Acids Res. 45 (2017) D972-D978. DOI: 10.1093/nar/gkw838.

Reference database in toxicogenomics providing links between genes, compounds, and pathologies.

42. ECHA, Guidance on information requirements and chemical safety assessment. Chapter R.8: Characterisation of dose [concentration]-response for human health. European Chemicals Agency (2012). Accessed online 6 December 2018: https://echa.europa.eu/documents/10162/13632/information_requirements_r8_en.pdf/e153243a-03f0-44c5-8808-88af66223258

43. K. de Raat, ECHA Science: Understanding the importance of assessment factors in finding safe human exposure levels. ECHA Newsletter February 2014, Issue 1 (2014) Accessed on-line 6 December 2018: https://newsletter.echa.europa.eu/home/-/newsletter/entry/1_14_understanding-the-importance-of-assessment-factors-in-finding-safe-human-exposure-levels

44. K. Raja, M. Patrick, Y. Gao, D. Madu, Y. Yang, and L.C. Tsoi, A review of recent advancement in integrating omics data with literature mining towards biomedical discoveries. Int J Genomics. 2017 (2017) 6213474. DOI: 10.1155/2017/6213474.

45. M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, et al., The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 3 (2016) 160018. DOI: 10.1038/sdata.2016.18.

46. T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, et al., NCBI GEO: archive for functional genomics data sets--update. Nucleaic Acids Res. 41 (2013) D991-995. DOI: 10.1093/nar/gks1193.

47. N. Kolesnikov, E. Hastings, M. Keays, O. Melnichuk, Y.A. Tang, E. Williams, M. Dylag, N. Kurbatova, M. Brandizi, T. Burdett, et al., ArrayExpress update--simplifying data submissions. Nucleaic Acids Res. 43 (2015) D1113-1116. DOI: 10.1093/nar/gku1057.

48. A.P. Davis, C.J. Grondin, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B.L. King, T.C. Wiegers, and C.J. Mattingly, The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. Nucleaic Acids Res. 43 (2015) D914-920. DOI: 10.1093/nar/gku935.

49. D.M. Hendrickx, H.J. Aerts, F. Caiment, D. Clark, T.M. Ebbels, C.T. Evelo, H. Gmuender, D.G. Hebels, R. Herwig, J. Hescheler, et al., diXa: a data infrastructure for chemical safety assessment. Bioinformatics. 3 (2015) 1505-1507. DOI: 10.1093/bioinformatics/btu827.

50. C. Hardt, M.E. Beber, A. Rasche, A. Kamburov, D.G. Hebels, J.C. Kleinjans, and R. Herwig, ToxDB: pathway-level interpretation of drug-treatment data. Database (Oxford). 2016 (2016) pii: baw052. DOI: 10.1093/database/baw052.

51. I.A. Lea, H. Gong, A. Paleja, A. Rashid, and J. Fostel, CEBS: a comprehensive annotated database of toxicological data. Nucleaic Acids Res. 45 (2017) D964-D971. DOI: 10.1093/nar/gkw1077.

52. H.G. Roider, N. Pavlova, I. Kirov, S. Slavov, Z. Uzunov, and B. Weiss, Drug2Gene: an exhaustive resource to explore effectively the drug-target relation network. BMC Bioinformatics. 15 (2014) 68. DOI: 10.1186/1471-2105-15-68.

53. A.M. Richard, R.S. Judson, K.A. Houck, C.M. Grulke, P. Volarath, I. Thillainadarajah, C. Yang, J. Rathman, M.T. Martin, J.F. Wambaugh, et al., ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chem Res Toxicol. 29 (2016) 1225-1251. DOI: 10.1021/acs.chemrestox.6b00135.

54. G.W. Miller, Data sharing in toxicology: beyond show and tell. Toxicol Sci. 143 (2015) 3-5. DOI: 10.1093/toxsci/kfu237.

55. T.A. Darde, P. Gaudriault, R. Beranger, C. Lancien, A. Caillarec-Joly, O. Sallou, N. Bonvallot, C. Chevrier, S. Mazaud-Guittot, B. Jégou, et al., TOXsIgN: a cross-species repository for toxicogenomic signatures. Bioinformatics. 34 (2018) 2116-2122. DOI: 10.1093/bioinformatics/bty040.

*56. Y. Igarashi, N. Nakatsu, T. Yamashita, A. Ono, Y. Ohno, T. Urushidani, and H. Yamada, Open TG-GATEs: a large-scale toxicogenomics database. Nucleaic Acids Res. 43 (2015) D921-927. DOI: 10.1093/nar/gku955.

This paper presents Open TG-GATES, a collaborative project between the National Institute of Biomedical Innovation (NIB), the NIES and around 15 pharmaceutical companies that sought to study 150 chemicals and their transcriptional responses in rat tissues.

*57. B. Ganter, R.D. Snyder, D.N. Halbert, and M.D. Lee, Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix database. Pharmacogenomics. 7 (2006) 1025-1044. DOI: 10.2217/14622416.7.7.1025

The second largest dataset in toxicogenomics. Set up by the National Institute of Environmental Sciences (NIES, USA), Drugmatrix aimed at studying transcriptional responses in rats of 376 compounds in five different tissues at several doses and multiple exposure times.

58. C.W. Tung, C.C. Wang, S.S. Wang, and P. Lin, ChemDIS-Mixture: an online tool for analyzing potential interaction effects of chemical mixtures. Sci Rep. 8 (2018) 10047. DOI: 10.1038/s41598-018-28361-6.

59. J. Nyström-Persson, Y. Natsume-Kitatani, Y. Igarashi, D. Satoh, and K. Mizuguchi, Interactive Toxicogenomics: Gene set discovery, clustering and analysis in Toxygates. Sci Rep. 7 (2017) 1390. DOI: 1038/s41598-017-01500-1.

60. J. Kringelum, S.K. Kjaerulff, S. Brunak, O. Lund, T.I. Oprea, and O. Taboureau, ChemProt-3.0: a global chemical biology diseases mapping. Database (Oxford). 2016 (2016). DOI: 10.1093/database/bav123.

61. L. Xing, L. Wu, Y. Liu, N. Ai, X. Lu, and X. Fan, LTMap: a web server for assessing the potential liver toxicity by genome-wide transcriptional expression data. J Appl Toxicol. 34 (2014) 805-809. DOI: 10.1002/jat.2923.

62. M. Römer, L. Backert, J. Eichner, and A. Zell, ToxDBScan: Large-scale similarity screening of toxicological databases for drug candidates. Int J Mol Sci. 15 (2014) 19037-19055. DOI: 10.3390/ijms151019037.

63. J. Lamb, E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.P. Brunet, A. Subramanian, K.N. Ross, et al., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 313 (2006) 1929-1935. DOI: 10.1126/science.1132939

64. C. Liu, J. Su, F. Yang, K. Wei, J. Ma, and X. Zhou, Compound signature detection on LINCS L1000 big data. Molecular bioSystems. 11 (2015) 714-722. DOI: 10.1039/c4mb00677a.

65. J. Setoain, M. Franch, M. Martinez, D. Tabas-Madrid, C.O. Sorzano, A. Bakker, E. Gonzales-Couto, J. Elvira, and A. Pascual-Montano, NFFinder: an online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning. Nucleaic Acids Res. 43 (2015) W193-199. DOI: 10.1093/nar/gkv445.

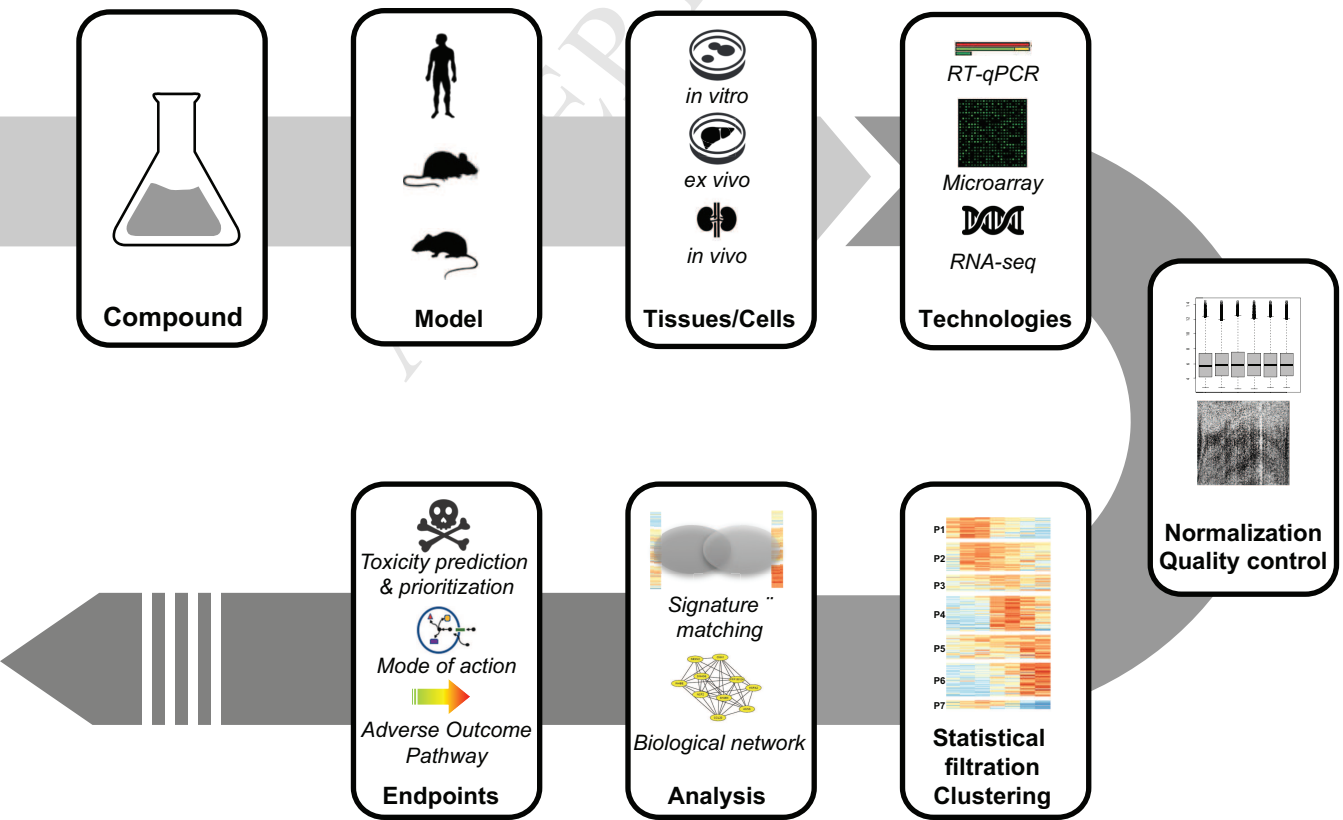
ACCEPTED MANUSCRIPT

Figure Legends

Figure 1 – Typical toxicogenomics analysis pipeline based on gene expression analysis. Several technologies such as RT-qPCR, microarray or RNA-Seq are used to evaluate the transcriptomic response following exposure to one chemical in one tissue or cell type, at a given time and dose. Various computational methods are used to measure differential gene expression and are considered significant based on pre-determined statistical cut-off values.

Using DGE, many analyses could be performed, for instance signature matching methods assuming that compounds which induce similar gene expression signatures will also have similar effects in a biological system; or the biological networking (i.e. co-expression network, protein-protein interaction), especially directed signalling networks, which allow scientists to follow the cellular response of a compound treatment from the compound's target to the differentially expressed genes. From these methods, and using a sufficient dataset, there is an obvious opportunity to estimate the toxicity of a compound, its mechanism of toxicity, and a related Adverse Outcome Pathway. With increasing data becoming available, and increasing sophistication of methods, the aim is that this will, over time, result in decreased animal testing and decreased the number of failures during drug development.

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT