Induced Pluripotent Stem Cells in chemical safety testing- data management and bioinformatics workflows

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

PRANIKA SINGH

Basel, 2022

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Dr. Thomas Exner
Prof. Dr. Martin F. Wilks
Prof. Dr. Alex Odermatt
Prof. Dr. Florian Caiment

Basel, 21.06.2022

Prof. Dr. Marcel Mayor
Dekan

Abstract

In recent years, there has been a push toward the adoption of alternatives to animal-based approaches for chemical safety assessment. This has led to the development of better human *in vitro* models. Within the in3 project, human-induced pluripotent stem cells (iPSCs) were used in combination with *in silico* tools for application in mechanistic toxicity testing. Project-wide data management and bioinformatics workflows were applied for the analysis of the transcriptomics data to uncover tissue-specific markers of toxicity and stress response pathways activated in iPSCs models against chemical exposure. In this thesis, first, the utility of iPSC-derived renal proximal tubule-like (PTL) cells in chemical testing was investigated. The cells were exposed to cadmium, arsenic, rotenone, tunicamycin, doxorubicin, amiodarone, and GW788388, and samples were collected at multiple time points. The time points were used to look at the temporal profile of the differentially expressed genes related to four important stress responses namely oxidative stress, metal response, p53 signaling pathway, and unfolded protein response. The results showed that iPSCs could unravel the activation of these stress responses. Oxidative stress response and metal response-related genes were found to be activated very early (around 1h-2h after exposure) in response to arsenic and cadmium but in later hours (around 12h-16h after exposure) in other compounds. Tunicamycin exposure led to the strongest expression levels of genes associated with unfolded protein response, which were activated very early (around 2h). Doxorubicin exposure led to the strongest expression levels of genes associated with the p53 signaling pathway showing activation from around 4h after exposure. Second, different iPSC models for the brain, blood-brain barrier, kidney, and liver were used to see their responses to paraquat (PQ) exposure. The pathways like oxidative stress and unfolded protein response known to be induced in response to PQ were found to be activated. Genes associated with these pathways showed a concentration-dependent increase for all the cell models. The brain model was found to be the most sensitive toward PQ based on the cytotoxicity curves. To investigate the reason for this, the expression level of transporters of PQ was checked as they are used by PQ to enter the cell. The amino acid transporters SLC3A2 and SLC7A11 showed high expression levels in the brain models leading to the high uptake of PQ in cells. In conclusion, these studies showed the potential of human-iPSCs in future studies for animal-free chemical safety assessment, determining the sensitivity of multiple models towards a chemical and signifying the applicability of temporal data in unraveling the evolution of mechanisms.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to Dr. Thomas Exner and Dr. Barry Hardy for giving me the opportunity to work on this project. I would particularly like to thank Dr. Thomas Exner for providing me with his great supervision and mentorship throughout my Ph.D. journey. He has been a constant source of support and motivation.

I would also like to thank Dr. Tatyana Doktorova and the rest of my colleagues from Edelweiss Connect GmbH for providing their assistance in helping me perform various tasks related to the project. My sincere thanks to Prof. Dr. Alex Odermatt and Prof. Dr. Martin F. Wilks for providing me with their constant support and constructive feedback in bringing this Ph.D. to completion. Thanks to Prof. Dr. Florian Caiment for agreeing to be my external supervisor and supporting me.

I am thankful to Prof. Dr. Paul Jennings who has been very encouraging of my decisions regarding the management and analysis of the project data and for providing his supervision and aid in finalizing the publications. It has been a great experience of being part of the in3 project and I am very thankful to all the in3 supervisors and ESRs for their collaboration and trust. Prof. Dr. Nynke Kramer and Dr. Anja Wilmes thank you for providing me with your expert knowledge and guidance throughout.

Lastly but most importantly I would like to thank my family and friends for being with me on this journey and giving me all their love and care. I am specifically grateful to my husband George who has been very patient, supportive, and motivating during challenging situations. His guidance has helped me achieve many personal and professional goals. My friends Meemansa, Menorca and Karmen thank you for being there to listen to me and motivate me. To my grandfather, parents, and siblings, thank you for always being there for me and cheering me up. You are all pillars of my strength and spirit to keep going in life.

Abbreviations

3Rs	Replacement, Reduction and Refinement
AO	Adverse outcome
AOPs	Adverse Outcome Pathways
BLECs	Brain - like endothelial cells
BS	Brain Spheres
DE	Differentially expression
DEGs	Differentially expressed genes
EC	Endothelial cells
ECHA	European Chemicals Agency
HLC	Hepatocyte - like cells
IATA	Integrated Approaches to Testing and Assessment
IPA	Ingenuity Pathway Analysis
iPSCs	Induced pluripotent stem cells
KE	Key events
KER	Key event relationship
MIE	Molecular initiating event
MT	Metallothionein
MTF1	Metal regulatory transcription factor 1
NAMs	New approach methodologies
NC	Neural cells
Nrf2	Nuclear factor erythroid - 2 - related factor 2
OECD	The Organisation for Economic Co-operation and Development
PODO	Podocytes
PQ	Paraquat
PTL	Proximal tubule - like cells
PTL	Proximal tubular like cells
qAOPS	Quantitative adverse outcome pathways
QSAR	Quantitative structure - activity relationships
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
UPR	Unfolded Protein Response

Table of Contents

1. Introduction

1.1 Ambitions and objectives

Chemical production has increased globally at a faster rate in the previous decade (UN 2015). This has also pushed the assessment of the safety of these chemicals for the protection of humans and the environment (Monosson 2005). The safety assessment still relies mostly on animal testing procedures. But these methods are costly, and time-consuming, raise ethical concerns, and cannot capture the inter-species differences sufficiently (Knight et al. 2021). Additionally, the pressure from the EU regulatory bodies to find alternatives to animal testing leading to the development of 3Rs (stands for Replacement, Reduction, and Refinement) principles has driven a paradigm shift in the field of chemical safety assessment (Fitzi-Rathgen 2019). Therefore, there is a need to develop alternatives to animal methods like *in vitro* and *in silico* tools which are time and cost-effective, more human-relevant, and useful to study the mechanisms of toxicity in more detail (Doke and Dhawale 2015). The recent years have seen significant growth in the development of these approaches (Krewski et al. 2020). The complete ban on animal tests in the cosmetic industry has also pushed other industries to look for alternatives to animal testing methods (Fentem et al. 2021). But the development of this animal-free safety assessment is hampered by the missing link between the quantitative and qualitative aspects of chemical exposures and mechanistic toxicology; the lack of standardized *in vitro* systems; and the lack of integration of *in silico* and *in vitro* tools/models (Pistollato et al. 2012; Price et al. 2020). Many past projects like SEURAT-1 (stands for Safety Evaluation Ultimately Replacing Animal Testing-1) (http://www.seurat-1.eu) (Kohonen et al. 2013), Tox21 (https://ntp.niehs.nih.gov/whatwestudy/tox21/index.html) (Richard et al. 2021) and EU-ToxRisk (stands for An Integrated European 'Flagship' Programme Driving Mechanism-based Toxicity Testing and Risk Assessment for the 21st century) (https://www.eu-toxrisk.eu) (Moné et al. 2020) and ongoing projects like RISK-HUNT3R (stands for Risk assessment of chemicals integrating the human-centric next-generation testing strategies promoting the 3Rs) (https://www.risk-hunt3r.eu), ONTOX project (https://ontox-project.eu), PrecisionTox project (https://precisiontox.org) and PARC project (stands for The European Partnership for the Assessment of Risks from Chemicals) (https://www.efsa.europa.eu/en/funding-calls/european-partnership-assessment-risks-chemicals-parc) have focused on different aspects of animal-free chemical safety assessment. in3 (stands for integrated *in vitro* and *in silico* tools) (https://estiv.org/in3/) project is one of the EU's Marie Sklodowska-Curie Action - Innovative Training Network (MSCA-ITN) funded projects built upon the knowledge and learning from these projects and train a new generation of researchers for the animal-free chemical safety assessment. It has utilized mechanism-based studies for hazard assessment, linking exposure to hazard, and integrating *in vitro* data with *in silico* tools. The main aim of the project was to apply human-induced pluripotent stem cells (iPSCs) *in vitro* cell models derived from the same genetic makeup to mechanistic toxicology assays and integrate them with *in silico* tools. The project consists of 15 sub-projects which were assigned to nine *in vitro* Ph.D. candidates and six *in silico* Ph.D. candidates appointed by eleven beneficiaries across Europe.

This thesis is part of the in3 project which focused on project-wide data management and the development of project-wide bioinformatics analysis workflows to uncover tissue-specific markers of toxicity and stress response pathways. The specific aim was to apply these approaches to a) investigate the potential of *in vitro* iPSCs models in predicting the known and unknown affected genes and associated mechanism of toxicity pathways due to exposure to different sets of chemicals b) resolve the temporal alteration of the toxicity pathways activated due to these chemicals c) see the response of different iPSC cell models against paraquat (PQ) d) see if such analysis can highlight differences w.r.t sensitivity of these different models towards PQ.

In the following sections of Chapter 1, the list of publications; an overview of the general concepts like 3Rs, new approach methodologies (NAMs), Integrated Approaches to Testing and Assessment (IATAs), and Adverse Outcome Pathway (AOP) to understand the rationale behind using specific methods within the in3 project; an overview of the specific concepts like Induced Pluripotent Stem Cells (iPSCs), transcriptomics application in mechanistic studies using TempO-Seq technology and biological pathways used within in3 project; description on the application and importance of data management and data analysis workflows developed for achieving the goals of the project; summary of the case studies demonstrating the application of bioinformatics workflows to analyze the response of iPSC-derived models on chemical exposure will be provided. Chapter 2 includes the publication based on the study related to temporal alterations in human-iPSC-derived renal proximal tubule-like (PTL) cells due to exposure to cadmium and also a shorter study to investigate the clustering of the related genes according to temporal profile. Chapter 3 focuses on the study related to temporal alterations in the stress responses activated in human-iPSC-derived renal PTL cells due to exposure to 6 organic compounds. Chapter 4 includes the publication based on the study of exposure of multiple human-iPSCs models to paraquat. Finally, Chapter 5 gives the cumulative conclusions highlighting the importance of the work done during the thesis and its future applicability.

1.2 List of publications

The following publications (with first authorship and shared first authorship) are based on the case studies which were performed to achieve the specific aims of this thesis as described in Section 1.1:

- **Singh P***, Chandrasekaran V*, Hardy B, Wilmes A, Jennings P, Exner T; Temporal transcriptomic alterations of cadmium exposed human iPSC-derived renal proximal tubule-like cells. Toxicology in Vitro. (2021) DOI: 10.1016/j.tiv.2021.105229.

- Nunes C*, **Singh P***, Mazidi Z*, Murphy C*, Bourguignon A, Wellens S, Chandrasekaran V, Ghosh S, Zana M, Pamies D, Thomas A, Verfaillie C, Culot M, Dinnyes A, Hardy B, Wilmes A, Jennings P, Grillari R, Grillari J, Zurich M, Exner T; An in vitro strategy using multiple human induced pluripotent stem cell-derived models

to assess the toxicity of chemicals: A case study on paraquat. Toxicology in Vitro. (2022) DOI: 10.1016/j.tiv.2022.105333.

- **Singh P,** Hardy B, Wilmes A, Jennings P, Exner T; Temporal alterations in the stress responses activated in human iPSC-derived renal proximal tubule-like cells due to exposure to 6 organic compounds (**in preparation**)

***shared first authorship**

The application of project-wide data management and developed bioinformatics analysis workflows were utilized by the first authors in the following publications to achieve their study goals. The support as co-author was provided by customizing the workflows and helping in the application of the workflows for the different study goals.

- Wellens S, Dehouck L, Chandrasekaran V, **Singh P**, Azevodo Loiola R, Sevin M, Exner T, Jennings P, Gosselet F, Culot M; Evaluation of a human iPSC-derived BBB model for repeated dose toxicity testing with cyclosporine A as a model compound. Toxicology in Vitro. (2021) DOI: 10.1016/j.tiv.2021.105112.

- Chandrasekaran V, Wellens S, Bourguignon A, Djidrovski I, Fransen L, Ghosh S, Mazidi Z, Murphy C, Nunes C, **Singh P**, Zana M, Armstrong L, Dinnyes A, Grillari J, Grillari R, Leonard M, Verfaillie C, Wilmes A, Zurich M, Exner T, Jennings P, Culot M; Evaluation of the impact of iPSC differentiation protocols on transcriptomic signatures (**in preparation**)

- Mazidi Z, Wieser M, Spinu N, Weidinger A, Kozlov A, Singh P, Vukovic K, Wellens S, Murphy C, Moralagares L, Reddy Bobbili M, Liendl L, Schosserer M, Diendorfer A, Bettelheim B, Eilenberg W, Exner T, Culot M, Jennings P, Grillari R, Grillari J; Cyclosporin A toxicity on endothelial cells differentiated from iPSC cells: assembling an AOP (**in preparation**)

1.3 The paradigm shift in toxicology towards non-animal test methods

This section is focused on the general concepts which are important for risk assessment starting from the role of the 3Rs principles in the use of alternatives to animal approaches, new approach methodologies (NAMs) which support the 3Rs principles, and the concept of Integrated Approaches to Testing and Assessment (IATA) and Adverse Outcome Pathways (AOPs).

1.3.1 3Rs and their contribution to non-animal test approaches

3Rs stands for Replacement, Reduction, and Refinement and were first proposed in 1960 by WMS Russell and RL Burch (Russell and Burch 1959). The goal of these principles is to replace animal methods with alternative non-animal methods, reduce the use of animal tests wherever possible and refine the existing methods to decrease animal suffering. The objective is to assure that optimum steps are taken to avoid the stress caused to animals as much as possible during the experimental tests and to utilize and build alternative non-animal methods (Hubrecht and Carter 2019). In many countries around the world, these principles serve as both legal and ethical principles and guide the use of animals in scientific procedures (Maestri 2021). The European Union (EU) has incorporated the use of the 3Rs as the key concept for the humane utilization of animals for scientific research in various legislations. These principles have been very well incorporated into the regulatory testing guidelines for conducting safety assessments of various types of chemicals used in different types of industries (Törnqvist et al. 2014).

The REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) is the regulation developed in 2007 in the EU. It is aimed at regulating the manufacturing and utilization of various chemical substances. This regulation requires industries to fulfill the conditions to ensure the safety of chemicals for both humans and the environment. The industries need to register the chemicals along with the safety information that they produce, use, or sell to the European Chemicals Agency (ECHA) (Taylor 2018). REACH also constantly promotes its objectives to encourage the use of non-animal testing methods as mentioned in the following statement in its article "This regulation should also promote the development of alternative methods for the assessment of hazards of substances" (REACH Legislation - ECHA,2020). To register the chemical under REACH regulations, the industries are not allowed to conduct any new experiments which involve vertebrate animals without prior approval from ECHA. In case the industry wants to use the animal approaches, they are obliged to show that alternative methods were considered for the testing and then ECHA can review their proposals for deciding if animal approaches are needed or not (ECHA 2020).

These regulations and legislations have put pressure on the industries to reduce the use of animals for scientific purposes and align with the 3Rs principles (GreenFacts 2022). This has led to a paradigm shift in the cosmetic industry to completely ban the usage of animal approaches in the toxicity testing (Rovida and Hartung 2009). EU Cosmetics Regulations 1223/2009 led to the complete ban on testing animals with the finished cosmetic products in 2004 and the raw cosmetic ingredients in 2009. In 2013, there was a complete ban on the marketing of cosmetic products including the ones manufactured before the regulations were made which in any form affect human health irrespective of the availability of alternatives for animal tests (Eixarch et al. 2019). The ban also signifies that the use of non-animal methods is not just a desire now but a necessity. This is also pushing the potential of the ban in other industry sectors. The industries are compelled to look for viable alternatives to meet the obligations led by different regulatory bodies.

1.3.2 New approach methodologies (NAMs)

The use of NAMs is gaining more importance in the modern day approach to toxicology. It focuses more on a mechanistic understanding of toxicity rather than histopathology which helps in better prediction and safer by design approaches. A topical Scientific Workshop held by the European Chemicals Agency (ECHA) in 2016 states new approach methodologies (NAMs) as methods that include approaches supporting 3Rs for alternatives for animal testing (e.g., *in silico*, *in chemico,* and *in vitro* methods) as well as in chemical associated hazard and risk information. Many of these approaches and tools are being developed and used for chemical testing like cell-based *in vitro* models (like primary cell culture, immortalized cell lines, organotypic culture, stem cells, iPSCs, and organoid culture) and computational modeling approaches. Many new advanced testing tools like high-throughput screening and high-content technologies (including genomics, proteomics, and metabolomics) are applied to these NAMs to test the chemicals these days (Cohen Hubal et al. 2010). NAMs have the potential to become promising alternatives to animal-based methods by reducing time and cost, offering more higher throughput, and being able to reduce animal usage in the chemical testing (Madden et al. 2020). Implementing standardization in the NAMs applicability and constant research demonstrating their usability will further help them in becoming alternatives to animal testing.

In vitro approaches are capable to gain more insights into the toxicity mechanisms and the cell types interactions (Ghallab 2013). The continuous advances made in *in vitro* cell models like the introduction of telomerase overexpression to immortalize cells leading to the creation of human cells without viral oncogenes and normal phenotypes; induced pluripotent stem cells (iPSC) opening the possibility for producing target tissues from any individual etc. have increased their utilization as alternatives approach. *In vitro* alternative approaches established for eye irritation, skin irritation, acute oral toxicity, and pyrogenicity have now completely replaced the animal testing methods (Perdomo-Morales et al. 2011; Alépée et al. 2014; Prinsen et al. 2017). *In vitro* approaches also help in understanding the mechanisms causing the toxicity and mode of action. The *in vitro* methods in combination with the use of advanced omics technologies offer a broad analysis of the effect of chemicals at the molecular level and point out to the induction of toxicity pathways that can lead to adverse outcomes (Anadón et al. 2014).

Many *in silico* approaches are also being developed and integrated with *in vitro* techniques as an alternative to animal methods. *In silico* methods work upon the computational assessment of toxicity of chemical substances (Madden et al. 2020)s. These methods can organize, analyze, model, visualize and predict the toxicity of a chemical and utilize the knowledge from the pre-existing data. The various *in silico* tools like quantitative structure-activity relationships (QSARs), read-across, biokinetic model, etc. are being used for toxicology-relevant studies (Madden et al. 2020). Based on the structure of the chemical, QSAR helps in predictions of chemical behavior inside a body. Read across approaches utilize the information from the endpoint of one chemical to predict the endpoint of another similar chemical like common functional groups in the absence of toxicological data (Punt et al. 2020).

However, NAMs also have some limitations for example the ability of *in vitro* models to mimic the conditions in the cells in a whole organism and to get the accurate extrapolation from *in vivo* to *in vitro* doses (Graudejus et al. 2018). But the efforts are being made in the creation of systems with interacting cells. For example, to study the communication among different organs like the liver, pancreas, gastrointestinal tract, brain, and hormone-producing organs multiple organoids have been joined to do that (Kim et al. 2020).

1.3.3 Integrated approaches to testing and assessment (IATAs) and adverse outcome pathway (AOP)

IATAs are structured approaches that integrate and weigh different types of pre-existing and new data extracted from different sources and methods for the hazard identification, hazard characterization, and/or safety assessment of chemicals to aid regulatory decision-making. It incorporates many NAMs for data generation, interpretation, and integration (OECD 2021). The different methods and sources used for this integration include methodological approaches (physicochemical properties, QSAR(s), read-across, *in vivo* and *in vitro*, *in chemico*) and/or omics technologies. The process of IATA is divided into three main steps: a) collection of pre-existing data generated by multiple approaches, b) evaluation of the weight of evidence (WoE) for this data usability in making a regulatory decision, and c) generation of new data if the collected data is not enough. IATAs consist of different testing approaches like integrated testing strategies (ITS), sequential testing strategies (STS), as well as the weight of evidence (WoE) considerations (OECD, 2014a) (Tollefsen et al. 2014).

Many IATAs use the adverse outcome pathways (AOPs) for assembling the pre-existing and new data in a structured format. AOPs provide a structured framework for the assembly and representation of the mechanism-based knowledge (Tollefsen et al. 2014). AOPs hold the potential in being used for regulatory and scientific applications like the selection of chemicals for the testing and development of NAMs, evaluation of novel biomarkers, providing the push for read-across methods, and thus contributing to 3Rs principles (Carusi et al. 2018; Coady et al. 2019). AOPs start from the chemical/stressor-induced molecular initiating event (MIE) and progress to the adverse outcome (AO) via a series of intermediate key events (KE) occurring at different biological levels like cellular, tissue, organ, organism, and population. The key events are connected using key event relationship (KER) representing the evidence to support the organization of the KEs in AOP qualitatively or quantitatively. AOPs provide relevance and reliability biologically for the results from structural data, “omics-based” data, *in silico, in chemico*, and *in vitro* data used in regulatory settings (Vinken 2013; Svingen et al. 2021; OECD 2021). Chemical-induced skin sensitization is one of the first AOP with an application in the risk assessment of cosmetics ingredients (Macmillan and Chilton 2019; Gautier et al. 2020; OECD 2014). The Organisation for Economic Co-operation and Development (OECD) has developed a handbook for AOP development (OECD 2021) in which it gives the standards for the building of high-quality and structurally similar AOPs, with comparable weight-of-evidence (WoE) evaluations (Villeneuve et al., 2014a). OECD monitors the utilization of the

AOP framework to ensure the building of hypothesis-driven hazard assessment tools across all continents. The integration of AOPs into IATAs has facilitated more drive toward the animal-free risk assessment (Krewski et al. 2020).

The AOPs are validated by exploring the following things as mentioned in (Meek et al. 2014): (1) biological concordance, (2) essentiality of Key Events, (3) concordance of empirical observation (encompasses dose-response and temporal concordance, and beyond), (4) consistency (among different biological contexts) and (5) analogy (consistency across chemicals). The results of these steps direct the use of different AOPs for different regulatory applications (Perkins et al. 2013; Patlewicz et al. 2014; Tollefsen et al. 2014).

The quantitative adverse outcome pathways (qAOPS) have been developed recently to capture the measurement of the perturbations caused by chemical exposure at different KE levels and how will that affect AO (Patlewicz et al. 2015; Conolly et al. 2017). The KEs can be measured to evaluate dose responses using experimental methods. The tipping points of change between KEs from MIE to AO are derived using the mathematical enabled KERs (Schultz and Watanabe 2018). Various NAMs like deep learning (DL) and machine learning (ML) models, *in vitro* tests, and batteries of *in vitro* assays, organs-on-a-chip models can contribute to populating the qAOPs by giving more information on chemical-induced toxicity (Mahony et al. 2020).

AOPwiki (https://aopwiki.org) coordinated by OECD is a collaboration of the European Commission-DG Joint Research Centre (JRC) and the U.S Environmental Protection Agency (EPA) transparently facilitates the development of AOPs globally. Different AOPs are connected automatically if they share any KE and various AOPs can be searched via keywords like AOP titles and key events. To date, around 400 AOPs and 5017 KEs have been developed in the AOPwiki (Ramšak et al. 2022).

1.4 Overview of specific concepts used within the in3 project

This section will provide an overview of the specific methods used within the in3 project starting from introduction to iPSCs utilization as the *in vitro* tool, transcriptomics importance in mechanistic studies, TempO-seq technology, and introduction to biological pathways. These concepts were used for achieving the specific goals of this thesis.

1.4.1 Induced pluripotent stem cells (iPSC):

Over the past years, various human cell line-based *in vitro* models have been developed for the chemical safety assessment (Suter-Dick et al. 2015). But many of these cell lines like primary cell lines have non-typical physiology as they manifest cancerous phenotypes and can sometimes also lose their differentiated functionalities (Rauch et al. 2014). Also, it is difficult to collect complex functioning tissues like the brain from such cell lines. iPSCs developed in 2006 by Takahashi and Yamanaka developed a method to reprogram the simple somatic cells

into stem cells in mice using 4 transcription factors namely Oct3/4, Klf4, Sox2, and c-Myc (Shi et al. 2017). These reprogrammed cells behaved like pluripotent stem cells and had the ability to self-renewal, differentiate into any cell type, and commit to a specific cell lineage (Takahashi and Yamanaka 2006; Suter-Dick et al. 2015; Omole and Fakoya 2018). The use of non-integrating techniques like self-replicating RNA has further improved the stem cell technology (Rauch et al. 2014; Hawkins and Kotton 2015; Suter-Dick et al. 2015).

As compared to embryonic stem cells, iPSCs are derived from adult living donors, have less ethical concerns, allow selection of donors, use a less invasive sample collection approach, are human-relevant, and can contribute to the patient-specific regenerative medicine (Scott et al. 2013; Hawkins and Kotton 2015). These iPSCs also make the study of diseases and their time-based evolution much easier by including gene expression relationships (Shi et al. 2017).

The utilization of iPSC cells in toxicity testing is still under development. The protocols used to differentiate the iPSCs into various target tissues need proper optimization and curation to help in showing their usage in toxicology studies (Jennings 2015). Often, the cells are derived in different laboratory setups, from different donors, and cultured in different media making it difficult to compare them and integrate information at the whole organism level. in3 project has utilized standardized and high-quality human iPSC cells from the same donors banked by StemBANCC(Morrison et al. 2015), EbiSC (Steeg et al. 2020), and in3 partners to generate various cell types. Different cell types namely liver (hepatocytes), kidney (podocytes and proximal tubule cells), brain (aggregates, neurons, and blood-brain-barrier), lung (conducting airways and alveolar cells), and endothelial cells (for vasculature and as co-culture partner cells) were differentiated from human-iPSCs by optimizing the developed differentiation protocols further. As the iPSCs are not still fully validated cell models representing the whole organ, in3 partners will also provide data for the further validation of iPSC models by comparing the responses of different *in vitro* models representing the same organ to different chemicals.

1.4.2 Transcriptomics application in mechanistic studies using TempO-Seq technology

The mechanism-based studies are important for understanding the pathobiology behind the adverse events caused due to exposure to chemicals. These studies are also useful in making the decisions in risk assessment (National Research Council (US) Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology 2007). The mechanisms facilitate better prediction, the elucidation of the routes of toxicity that points to the cause and progression of certain diseases, and safer by design methodologies (Houck and Kavlock 2008; Cui and Paules 2010). These mechanistic approaches utilize transcriptomics which offers non-biased experimentation and high content data (Jennings et al. 2014).

Transcriptomics is beneficial in such cases as it allows the study of all transcripts or mRNA in a cell or an organism simultaneously and can be applied in read-across studies, understanding the mechanism of toxicity as well as finding points of departure (Wilmes et al. 2013; Fröhlich

2017; Joseph 2017). Various stress response pathways can be studied on a cellular level activated on exposure to xenobiotics using the transcriptomics (Wilmes et al. 2013; Lowe et al. 2017). Specific transcriptional signatures (oxidative stress, DNA damage, ER stress, metal response, hypoxia, etc.) are exhibited by these stress responses (Jennings et al. 2013). It allows the understanding of alteration in gene expression at cellular levels due to stress caused by chemical exposure (Jennings et al. 2013). The OECD Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST) has also made the Transcriptomic Reporting Framework (TRF) for increasing the use of transcriptomics data in regulatory studies. Transcriptomics data has the potential to play an important role in the development of IATAs even further. They offer further knowledge into the mechanism of actions that can be used to enrich the AOPs at different biological levels (Garcia-Reyero et al. 2014a, b).

The transcriptomics data obtained from the toxicity studies can be analyzed using bioinformatics-based statistical approaches. It helps in understanding molecular mechanisms underlying the toxicity of the chemical by quantitatively assessing and analyzing the transcriptomics data (Joseph 2017). The biggest statistical challenges found during the analysis of the transcriptomics data are the limited number of samples and few biological replicates, which affects the normalization and mean due to sample variability. Various methods have been developed for differential expression analysis to estimate such variance of genes (Love et al. 2014a). In the in3 project, the DeSeq2 package was used as a statistical tool for the differential expression analysis of the transcriptomics data. It offers an advantage by removing the noise in the data by estimating standard errors for the log fold changes, flagging the genes/outliers which should not be considered in the downstream analysis and it can be applied to a limited number of replicates. It also offers a regularized logarithmic transformation for the visualization of the raw read count data for quality assessment and clustering and principal component analysis of over dispersed data (Love et al. 2014a).

The traditional transcriptomics techniques are low throughput and costly. The cost of running the whole genome is still relatively expensive. Besides that, the complexity increases for RNA sample preparation and post bioinformatics analysis with the number of samples. Due to this, the study of multiple compound concentrations and/or temporal effects is still in limited capacity (Wilmes et al. 2011a, 2013). However, the cheaper, high throughput, and targeted transcriptomic-based methods have brought the cost of sequencing lower, include more target-specific selection from whole transcriptome sequencing libraries, and help make the way for better hazard and risk assessment using more mechanistic-based studies (Yeakley et al. 2017; Limonciel et al. 2018).

One of these cheaper methods is TempO-Seq technology provided by BioSpyder (http://biospyder.com/), which was used in the in3 project to generate transcriptomics data for the case studies mentioned in the later chapters. TempO-Seq is an NGS library preparation method that is based on the hybridization and sequencing of highly specific detector oligos. The ligated oligos are hybridized on the RNA targets directly without the need for RNA extraction. It only requires small picograms of total RNA without the preamplification step and

thus increases the use of the samples (Limonciel et al. 2018; Harrill et al. 2021). TempO-Seq is advantageous as it doesn't need the isolation of RNA, is cost-effective due to its targeted behavior, quantifies specific mRNA targets, and allows sequencing from sample lysates which leads to the removal of the variation causing factors like library construction in the RNA-sequencing (Su et al. 2014). Additionally, it requires fewer data storage and allows simplified bioinformatics workflows application which gives faster and more manageable results. It has moderate variation, noise, and error in comparison to other methods and has been reported to be similar to microarray and RNA-sequencing when applied for SEQC/MAQC3 MOA toxicogenomics data analysis as described in (Bushel et al. 2018). The main limitation of this method is its targeted focus on the toxicological responses and therefore the data is less useful for other research not looking at toxicology.

in3 project has used TempO-Seq utilizing the EU-ToxRisk v2.1 panel 3257 genes (3565 probes) which contains the genes curated specifically relevant for the toxicology. This gene set was created with the help of bioinformatics approaches by using publicly available human transcriptomics data and scoring each gene based on its interaction with other genes, transcriptional diversity, and coverage of the biological space (Mav et al. 2018). These genes fully cover all known and well-annotated biological pathways which accurately predicts the pathway perturbations after chemical exposure.

1.4.3 Biological pathways

The National Human Genome Research Institute defines a biological pathway as "A series of interactions among molecules in a cell that leads to a certain product or a change in a cell. Such a pathway can trigger the assembly of new molecules, such as a fat or protein". The most known pathways are related to metabolism activities, regulation of genes, and transmission of signals.

The information for the development of the biological pathways is derived from the literature studies based on the experiments performed by scientists on various cells and species. The extracted information is aggregated into a pathway either manually or by using software tools using appropriate formats, information standards, statistical values implicating their significance, and pathway-building tools. These pathways are then topped up with additional information like species, disease, cell type, etc., and are regularly checked and curated by the experts in the domain (Viswanathan et al. 2008). They are stored, represented, and accessed by collecting in pathway libraries/knowledgebases using computational tools. In this thesis, three pathway libraries were used namely Reactome (https://reactome.org) (Jassal et al. 2020), Ingenuity Pathway Analysis (IPA) (https://digitalinsights.qiagen.com/products-overview/) (Krämer et al. 2014), literature derived pathway library of in3 (developed by Vidya Chandrasekaran, Vrije Universiteit Amsterdam, details in Chapter 2) and an integrated library called ConsensusPathDB (http://cpdb.molgen.mpg.de) (Kamburov et al. 2009) which integrates many important pathway libraries. These pathway libraries help in making the information available for genes, their regulation, and correlation easily accessible and usable.

The limitation of these libraries is the non-standardization of the pathway names which also leads to repetitions in integrated libraries like ConsensusPathDB.

1.5 Data management and analysis workflows application within in3 project

After describing the background in the previous sections needed for working on the specific aims of the thesis, this section will give an overview of the data management and analysis workflows, which will be then applied in the different studies summarized in the next section and then detailed in the attached publications in chapters 2-4.

The design of the interlinked studies within the in3 project led to the generation of transcriptomic data using TempO-Seq technology to characterize the newly developed iPSCs models both under basal and chemical exposure with comparable parameters (time, concentrations, cell types, etc.). This data was also used to achieve specific objectives like the evaluation of the iPSC tools for safety assessment including determination of model robustness, concentration-response, time series analysis, assessment sensitivity and specificity to chemical exposure, grouping and read-across approaches, and providing mechanistic validation for the QSARs. Central data management and common workflows for data processing and analysis were needed for the alignment of the study plans within in3 to allow comparative studies, collaborative work, and foster re-use of data within and outside the project. The output of this was a standardized data management and analysis workflow which also allowed customization according to the objectives of the individual study plans. It made both cell model comparison on a gene level and identification of similarities and differences in baseline and exposure scenarios possible.

The reuse of data is very essential for optimizing data-driven research and innovation. This is most often framed under the FAIR principles (findable, assessable, interoperable, and re-usable) and named FAIRification of data (Wilkinson et al. 2016). The focus of these principles is to facilitate the long-term availability and re-use of the data. However, the importance of rich metadata following domain-relevant community standards cannot be stressed enough, is extremely relevant also for the sharing inside a project and a starting point for the generation of integrated analysis workflows. It is very necessary to document the needs for the metadata in parallel to the data collection to avoid missing and later irretrievable information making the independent interpretation of the data impossible and the experiments irreproducible. There is consensus that metadata can be defined as information defining and describing other data and the ISO/IEC JTC1 SC32 Working Group as a central standardization body has adopted this definition (Anon 2014). Another more obvious advantage of early and complete metadata collection is that this information can directly be applied in the development of the integrated analysis pipeline as well as in hypothesis formulation and interpretation of the results.

A careful metadata collection process was executed following good (meta) data management standards to guarantee the completeness of information for the analysis of the data in each

study but also for cross-study analysis to compare cell models and differences in sensitivities. A metadata template was developed for in3 to provide details on the assays, compounds used, and exposure scenarios optimized to the specificities of iPSCs testing procedures. The metadata included information to characterize the cell model, differentiation stage, exposure scenario including co-stimuli, vehicle, and plate layout as well the detailed standard information like SMILES, InChI, etc. on the chemicals used for treating the cell systems. The raw read count data files for all samples were provided by BioClavis (https://www.bioclavis.co.uk) who performed the TempO-Seq experiments and the quality control. The metadata as well as the data files were uploaded to the EdelweissData™ data management system (SaferWorldbyDesign2020) after multiple cycles of metadata schema curation and validation distributed among different partners. From the EdelweissData™ data management system, the data and metadata can be directly inspected, queried, and without the need to first download the files integrated into custom-made workflows. These workflows could be used to select the samples of interest and the corresponding controls, filter the data according to quality measures, and then finally apply it in integrated processing and analysis pipelines.

Access to all the data was provided to all partners within in3 and a standardized workflow was created in form of the python script implementing the application programming interface (API) calls to obtain the data based on metadata fields selected by the user interactively. The script was provided to everyone in form of a Jupyter notebook (https://jupyter.org/) along with documentation of the steps performed including intermediate results to make the use of the script simple without the need to install Python or later R. As the first step to use this workflow, the user must select the study of interest. It is also possible to select more than one study if needed for cross-study comparisons e.g., gene profiles of the cells treated with the same chemical. The next step allows the user to filter the data based on the metadata fields such as compound, cell lines, organ concentration, and exposure time. The last step of the script offers an option to filter the genes based on statistics of raw read counts across all selected samples. This is used to remove genes from the raw data, that show extremely low read counts across all samples. Criteria of this selection are the sum of counts (rowsum), the maximum, mean, and median value over all samples. These low read count levels with the small number of replicates used (3 in most cases) can result in high fold changes with seemingly reasonable significance as expressed in the adjusted p-values even if they are a result of the variance in the measurement and not of a biological effect. The data can be written to the local disk on the computer before as well as after the optional filtering step.

An R-based bioinformatics script was created to perform integrated analysis and was provided as a Jupyter notebook to be able to use the DESeq2 library (Love et al. 2014a). This standardized script performs several steps, which can also be customized based on specific questions. More advanced functions for quality control, differential expression analysis, and interpretation are also added to this script. The script starts with a few methods for the quality check of the raw read counts for which the generated graphs can be saved as images. The boxplot functionality from R was used to visualize the raw count distribution for manual inspection of the sample's overall quality. Even if the normalization performed later is

compensation for this in some respect, this gave a chance to decide if it is better to remove the samples or at least flag them when interpreting the results if they keep them.

The clustering method available as part of the DESeq2 library (Love et al. 2014a) was also added to the workflow for the exploratory analysis. The r-log transformation was applied to the raw data for sample-based normalization steps before clustering to compensate for different overall number raw read count levels in the different samples (Love et al. 2014). The clustering here uses a hierarchical approach with default methods set as Euclidean distance with a complete linkage method. Another way of looking at groups of samples with similar properties is principal component analysis (PCA). This method looks at the variance in the data and generates linear combinations of the input variables perpendicular to each other, i.e., the principal components, to optimally characterize this variance.

After the steps of quality checks on the raw read count data, the next step was the normalization which uses the median of ratios method (Anders and Huber 2010) recommended for further processing with the DESeq2 library. Normalization brings the numeric values in the selected dataset onto one common scale which is important for doing comparative analysis. The median of ratios method generated normalization factors based on ratios for each gene between the sample and a pseudo-reference sample created as the geometric mean across all samples. It considers sequencing depth and RNA composition and is robust concerning avoiding imbalance caused by samples with extreme expression levels or a large number of differentially expressed genes (DEGs). It is also suited for gene count comparisons between and within the group of samples and for differential expression analysis between genes in one sample. The normalized values are saved in a separate file for additional analysis outside the standard workflow.

Differential expression (DE) analysis is performed as the final step of the workflow to determine the genes which are significantly influenced by differentiation protocols or exposure to various chemicals. The DESeq2 library was further used to perform differential expression analysis using the Wald statistical test (Agresti 2006) for significance evaluation. In this way, different experimental groups (like treated vs untreated) can be compared regarding changes in gene expression levels. The DESeq () function of the DESeq2 library calculates fold changes, i.e., relative up and down-regulation compared to the reference group, and p-values and adjusted p-values, i.e., measures for the statistical significance of the changes, for all groups defined by the user. A p-value of 0.05 means that there is a 5% percent chance that the gene is a false positive, i.e., 5% of the genes are identified as significantly differentially expressed even if the changes could also be explained by random fluctuations (the null hypothesis) (Burden et al. 2014). The adjusted p-values (also known as false discovery rates) take care of the fact that the number of false positives will increase if more genes are tested. A cutoff value of 0.05 means that 5% of the tests (genes) considered significant are accepted as false positives due to multiple testing (Benjamini and Hochberg 1995; Burden et al. 2014).

The metadata files created for each study include the “Group ID” column to automate the group definition in the workflow. Group ID was a unique identifier assigned to each group of

replicates representing the same conditions such as cell model, test compound, and exposure scenario. To make this grouping used in all DESeq2 functions, the "design" parameter is set in the DESeqDataSetFromMatrix () function to point to this column in the metadata file (design = ~ Group. ID). Depending on the study design, the most relevant of these group comparisons are then automatically written to files. The pairing control columns were added to the metadata to avoid manual selection of the comparisons. These specify the control group for a treatment group. The standard workflow uses the "Pairing Control.1" column to identify the control groups. Some studies also include additional pairing columns, e.g., "Pairing Control.2", which can be selected by simple modification of the code. Also, comparisons to the other cell model are possible besides the negative controls by mentioning the Group IDs one wants to compare in the "contrast" parameter of the "results" function. The genes can then be further filtered using log fold changes and adjusted p-values (padj) for choosing the statistically significant ones.

1.6 Application of bioinformatics workflows to analyze the response of iPSC-derived models on chemical exposure

The data management and analysis workflows were also utilized to achieve the specific aims of the thesis as mentioned before in Section 1.1. The main results will be summarized here, and details will be provided in Chapters 2-4 in the form of reproductions of the corresponding peer-reviewed publications.

1.6.1 Summary: Utilization of iPSC PTL cells to study cadmium-induced stress responses

1.6.1.1 Temporal resolution of the stress responses activated due to cadmium

The first study was aimed at a) investigating the utilization of a human-derived iPSC PTL cell model exposed to cadmium in chemical safety assessment and b) deriving the temporal profiles of the well-known stress response pathways like Nrf2-mediated oxidative stress and MTF1-mediated metal response known to be activated on exposure to xenobiotic cadmium. Proximal tubule cells are the most affected part of the kidney on being exposed to cadmium (Barnett and Cummings 2018). To study the effects of cadmium on proximal tubules, well-studied and established models like HK2 and RPTEC/TERT1 have been used (Wilmes et al. 2011b). But iPSCs differ from them by being more donor-specific (Jennings 2015; Wilmes and Jennings 2015; Pamies et al. 2018; Murphy et al. 2019).

The iPSCs were differentiated using the protocol described in (Chandrasekaran et al. 2021) and were exposed to cadmium chloride (non-cytotoxic concentration of 5 μM) and medium controls. The resulting samples were collected at 1h, 2h, 4h, 6h, 8h, 12h, 20h, 24h, 72h, and 168h. The raw read count data and metadata exposed to cadmium and the corresponding controls for all the time points in iPSC-derived PTL cells were analyzed for finding the differentially expressed genes using DESeq2 (Love et al. 2014b). These genes were then

associated with pathways using three pathway libraries namely Reactome (https://reactome.org) (Jassal et al. 2020), IPA (https://digitalinsights.qiagen.com/products-overview/) (Krämer et al. 2014), and literature studies. The pathways derived from all three sources were scored using the z-score (Kutmon et al. 2015; Wellens et al. 2021). The significant and overrepresented pathways were selected using the threshold of 1.96. A fitting procedure was developed in python to represent the temporal profiles of the genes related to the stress responses. The profiles were modeled using 10 pre-defined functions from the standard SciPy library (Virtanen et al. 2020).

In total 115 significantly differentially expressed genes were obtained, summing up the number of differentially expressed genes (passing the cut-offs) at different time points. The heatmap showed these genes forming three clusters of different behavior. The first cluster contained genes with a very high expression like metallothioneins (MT) MT1 and MT2, HSPA6, and HMOX-1. They attained the maxima between 4h-6h and then reached a plateau slighter lower than the first maxima with an increase in late hours except for HSPA6 which reached baseline levels in later hours. The MT genes were related to metal regulatory transcription factor 1 (MTF1) (Sabolic et al. 2010) and HMOX1 to nuclear factor erythroid-2-related factor 2 (Nrf2) (Okawa et al. 2006; Reichard et al. 2007; Yates et al. 2009) based on literature studies. The second cluster was formed by genes with early maxima between 4h and 8h, then a drop-down to baseline levels by 24h and no or little over-expression in later hours. Genes showing this behavior were GCLM, MAFF, MAFG, etc., and were seen to be associated with the Nrf2 transcription factor. The third cluster contained genes like PCNA and SESN1which showed no expression or a lot of noise until 24h but higher expression values in the later hours. These higher expressions were caused due to a decrease in the expression values in the untreated samples. Most of these genes were associated with the p53 transcription factor.

The pathway analysis performed using z-scores showed different significance values and results. The transcription factors associated with the differentially expressed genes based on literature evidence showed the relation of MTF1 to metal response and Nrf2 to oxidative stress. All these were also seen as the overrepresented pathways based on z-score calculation. Exploring the same knowledge using Reactome knowledgebase also showed metal response overrepresented. But there was no specific pathway in Reactome which represented oxidative stress response. IPA used as another knowledgebase showed pathways representing oxidative stress as significant but there was no specific pathway for a metal response. Instead, it showed "SPINK1 General Cancer Pathway" associated with all MTs as overrepresented. The combination of these pathway libraries along with the transcription factor knowledge derived from literature studies helped in the identification of important known stress response pathways.

The results showed that the iPSC PTL cell model can help in the identification of the known stress responses activated against cadmium exposure. The genes related to Nrf2-mediated oxidative stress response and MTF1-mediated metal response showed early activation. While Nrf2-mediated oxidative stress response-related genes came to baseline levels by 24h, the MTF1-mediated metal response-related genes sustained the response until 168h. The two

responses showing different behavior could be due to activation of both responses in reducing the oxidative stress response. Nrf2-mediated oxidative stress started coming down since the MTF1-mediated metal response is removing the free intracellular cadmium. This could increase oxidated glutathione and decrease the amount of intracellular free cadmium leading to genes related to oxidative stress response going down by 24h.

1.6.1.2 Clustering genes according to temporal profiles

In the previous study, the related genes were mapped to the pathways using the pathway libraries. This study aimed to see if a clustering method can group the related genes based on their temporal profiles and point to their involvement in a mechanism. First, two known clustering methods namely hierarchal clustering (Zolfaghari et al. 2019)and Short Time-series Expression Miner (STEM) clustering tool (Ernst et al. 2005) were used. Besides these, the third type of clustering was performed based on the two criteria developed to group the gene profiles modeled using the fitting procedure developed in (Singh et al. 2021). The criterion was: a) t_{max} is the time point when the first maximum or the plateau is reached or in cases where there is no clear maxima point, then the time point when 95% of the maxima height is reached b) residual which is the amount of the maxima retained at the end of 24h.

The results showed that hierarchal clustering clusters the genes in two groups: first where genes with similar temporal patterns were assigned to two different clusters and second where the genes with different temporal patterns were assigned to the same cluster. The reason for that is due to hierarchal clustering grouping the genes based on the absolute values of log fold changes and not considering the temporal relationships between genes. STEM even if could assign genes with similar profiles to the same cluster, also included many different profiles in the same cluster. The reason for that is due to STEM not able to use its model profiles to distinguish the noise levels from the significance levels of the gene expression values. This made STEM group many genes together to the same standard profile. The third method based on the two criteria applied to the profiles modeled using the fitting procedure showed distinct behavior for both Nrf2-mediated oxidative stress and MTF1-mediated metal response related genes. Both started activating early but only genes related to MTF1-mediated metal response sustained the response until the end of the measurement period. Nrf2-mediated oxidative stress went to baseline by 24h. The MTF1-mediated metal response related MT genes showed dense clusters with t_{max} achieved around 6h and a residual of 60-80% left for all the genes. Nrf2-mediated oxidative stress related genes were seen to be distributed over a wider area. Most of them had distinct t_{max} and different residual percentages at the end of the measurement. This method could help in defining a distinct pattern of MTF1-mediated metal response but fails in the case of Nrf2-mediated oxidative stress. The distribution of the Nrf2-mediated oxidative stress related genes showed that most of them have different behavior and they overlap with other genes from other mechanisms. The reason could be a complex pathway like Nrf2-mediated oxidative stress includes many other mechanisms which have different activation time points. The MTF1-mediated metal response has most of the genes belonging to one class called MT and is thus expected to behave similarly.

Therefore, clustering the related genes together without the use of a knowledgebase was not possible, especially in the case of a complex pathway like Nrf2-mediated oxidative stress. The mechanisms activated in response to chemical exposure usually are not independent and involve many other mechanisms and process happening at the same time or later. Also, many genes are not specific to one mechanism and can be shared by many other mechanisms which also influence their activation time and expression levels.

1.6.2 Summary: Utilization of iPSC PTL cells to study six different chemicals and the induced stress responses

The second study aimed to a) see if iPSC-derived PTL cells can be also utilized to uncover the well-known pathways affected by other compounds as they did in the previous study and b) resolve the temporal resolution of these pathways due to exposure to different compounds. The four stress response pathways p53 signaling pathway, oxidative stress response, metal response and unfolded protein response known to get activated in response to cellular perturbations due to chemical exposure (Jennings et al. 2013) were chosen for the study.

The iPSC-derived PTL cells were exposed to multiple compounds namely arsenic (10 µM), amiodarone (50 µM), doxorubicin (15 µM), GW788388 (1 µM), rotenone (5 nM), and tunicamycin (300 nM) and the samples were collected at 1, 2, 4, 6, 8, 12, 16, 20 and 24 hours. The data was analyzed using DESeq2 (Love et al. 2014a) and the fitting method (Singh 2020) was applied to model the temporal profiles of genes as done in the previous study. ConsensusPathDB V35 (http://cpdb.molgen.mpg.de) (Kamburov et al. 2009) was used as the pathway library to find the mechanisms associated with the DEGs. The z-scores and the corresponding number of DEGs for each time point per compound were also calculated for finding the overrepresented pathways.

The results for the stress response activation due to arsenic exposure were compared to the cadmium from the previous study to find the differences or similarities in the temporal resolution between metal compounds. The highest number of DEGs and their expression related to Nrf2 mediated oxidative stress was found due to exposure to arsenic and cadmium as compared to other compounds. Exposure to both the metals showed early activation of genes starting from around 1h-2h. The Nrf2-mediated oxidative stress was sustained until 24h in the case of arsenic but it started going down by 24h in the case of cadmium. The metal response was also activated in response to both metals in the early hours and was sustained until 24h. The reason for the sustained response for both metal response and oxidative stress responses in arsenic could be due to the high level of reactive oxygen species (ROS) produced due to remaining arsenic left even after full expression of MTs. This leads to oxidative stress activation until 24h to counterattack the ROS production.

Exposure to amiodarone, rotenone, and doxorubicin led to no or few DEGs activations related to Nrf2-mediated oxidative stress in the early time points. But a steady increase in the number of DEGs activation happened in the later hours around 12h. This means that many of the genes related to Nrf2-mediated oxidative stress for these compounds must have reached their highest

expression at 24h and therefore, more time points beyond 24h would be needed to show the full activation of Nrf2-mediated oxidative stress in these compounds. Tunicamycin and GW788388 showed only 2 and 1 associated DEGs, respectively. The literature studies also showed the activation of oxidative stress due to these compounds (Carvalho et al. 2009; Chakraborty et al. 2014; Afsar et al. 2020). The upregulation of the Nrf2-mediated oxidative stress related genes like MAFF, MAFG, and SQSTM1 and downregulation of SLC2A1 were found to be common in arsenic, cadmium, amiodarone, rotenone, and doxorubicin.

The MTF1-mediated metal response related genes were activated due to amiodarone exposure and showed a considerable increase in the expression in the later hours around 12h. Doxorubicin exposure also showed the related DEGs for metal response in later hours, but they were downregulated which may mean the involvement of these DEGs in another mechanism.

UPR pathway related genes showed the highest expression levels in the early hours and started coming to baseline levels by 24h in the case of cadmium and arsenic. Tunicamycin exposure led to the genes' early activation and the highest overexpression but went down by 24h. In the case of amiodarone exposure, the overexpression for the genes was highest around 12h-16h and went down by 24h. In rotenone, the activation started in the later hours around 16h which means it needs more time to reach its full activation. Doxorubicin showed the activation of only one gene ATF3 which was also found to be part of the p53 signaling pathway (Yan and Boyd 2006) and can therefore not be called a specific gene for this pathway. GW788388 didn't show any activation of the genes related to UPR. The upregulation of TRIB3, DDIT3, ASNS, and PPP1R15A could be seen as the potential biomarkers for the UPR as they are seen in arsenic, cadmium, amiodarone, rotenone, and tunicamycin. Their direct association with UPR could also be seen in the literature evidences (Gjymishka et al. 2009; Nicoletti-Carvalho et al. 2010; Rashid et al. 2015; Young et al. 2016).

Doxorubicin exposure led to the highest number of DEGs related to the p53 signaling pathway, and overexpression as compared to other compounds. The p53 signaling pathway was sustained in the case of doxorubicin until 24h. In the case of the metal compounds, the response sustained until 24h for arsenic exposure but went down to baseline levels by 24h for cadmium exposure like oxidative stress response. In the case of doxorubicin, the activation was early between 2h and 4h as compared to oxidative stress response (6h or 8h). Amiodarone and rotenone exposure showed the activation start in later hours around 12h-16h. Tunicamycin exposure led to very weak activation of a few genes related to p53 and GW788388 didn't lead to any activation of the genes related to the p53 signaling pathway. ATF3, GDF15, and RGCC being upregulated due to arsenic, cadmium, doxorubicin, amiodarone, and rotenone could be potential biomarkers for the p53 signaling pathway. Their association with the pathway was also found in the literature (Tsui et al. 2015; Counts and Mufson 2017).

GW788388 exposure showed no activation of genes related to any stress response pathway. It is a known transforming growth factor-β (TGF-β) receptor kinases inhibitor designed to halt renal fibrosis (Petersen et al. 2008) with no toxic effects reported so far. Only 4 out of the 16 genes activated due to exposure to GW788388 were associated with the pathways leading to

the renal fibrosis (Petersen et al. 2008). WT1 and TXNIP showed downregulation and CYP24A1 and SERPINE1 showed upregulation in renal fibrosis due to GW788388 which has also been reported in literature studies (Qi et al. 2007; Ma and Fogo 2009; Sakairi et al. 2011; Miller-Hodges and Hohenstein 2012; Hu and Zhang 2018).

The study also found that z-scores should not be used as a sole criterion for choosing the significantly overrepresented pathways and the gene expression profiles should be used as a primary source to look more into the derived results. Therefore, the results and discussion were based on gene expression profiles of the stress responses in this study. The z-scoring showed few issues in the study. The first issue was seen when the number of genes remained the same, and the fold changes increased for these genes, but the z-scores dropped down. This was seen for metal response in cadmium between 4h and 8h. The reason could be the total number of DEGs across all pathways which influenced the z-score making metal response less important. The second issue was when the z-score for the Nrf2 pathway was very high in arsenic at 1h even when there was just 1 DEG expressed for this pathway. This happened due to the low number of DEGs in total expressing at 1h and thereby increasing the z-score. It should be used with caution when comparing time points for the relevance of the pathway and specifically when comparing to other compounds or time points as this can mislead the interpretation. The third issue is that the z-scores are influenced by the number of total DEGs at a time point due to chemical exposure and the number of total DEGs belonging to the pathway of interest. Therefore, oxidative stress was not automatically identified as an overrepresented mechanism for amiodarone, doxorubicin, and rotenone even if the genes and literature showed the activation of this pathway.

The study showed that the iPSC PTL cells were able to show major known stress response activation due to exposure to amiodarone, arsenic, rotenone, tunicamycin and doxorubicin like it did for cadmium. The temporal resolution was very useful to see the difference in the timelines of the responses against exposure to different compounds. GW788388 showed no adversity and stress response until 24h. Both metals cadmium and arsenic showed similar strong expression of genes related to stress response mechanisms. Tunicamycin showed strong induction of UPR, and doxorubicin showed strong activation of the p53 signaling pathway.

1.6.3 Summary: Utilization of different iPSC cell models to compare their response to paraquat

This study was focused on comparing the response of different cell models to paraquat (PQ) and determining their sensitivity to PQ. PQ is a well-studied herbicide that produces ROS using redox cycling in mitochondria and induces an oxidative stress response in several organs (Dinis-Oliveira et al. 2008; Gawarammana and Buckley 2011; Blanco-Ayala et al. 2014).

iPSCs cells were differentiated into different cell types namely brain spheres (BS), neural cells (NC), brain-like endothelial cells (BLECs), podocytes (PODO), proximal tubular-like cells (PTL), the endothelial cells (EC), and the hepatocyte-like cells (HLC) using new and developed protocols by different partner laboratories. The different cells were exposed to different

concentrations of PQ for 24h with brain spheres exposed also at 48h. The viability assays were run for deriving the cytotoxicity concentration-dependent curves after the exposure of the different cells to a wide range of PQ concentrations. The raw read count data and metadata exposed to PQ and the corresponding controls for all the iPSC cell models were analyzed for finding the differentially expressed genes using DESeq2 (Love et al. 2014b). These genes were then associated with pathways using the ConsensusPathDB (http://cpdb.molgen.mpg.de) (Kamburov et al. 2009) knowledgebase. The default settings i.e., the minimum overlap of 2 genes between the input gene list and a P-value cutoff of 0.01 were used to filter the most overrepresented pathways. UpSetR package (Conway et al. 2017) in R was used to find the common pathways shared by different combination sets of models.

The cytotoxicity curves showed the difference in the sensitivity of each model to PQ. Out of the two brain models, the NC model was found to be the most sensitive whereas the BS model was less sensitive. EC showed the same range whereas BLECs showed even lesser sensitivity. For the kidney, PTL was less sensitive as compared to PODO. HLC representing liver was found to be the least sensitive.

EC model was not considered further for analysis as it had about 64% probes showing very low read count probes and at least one replicate in every treatment group had no correlation to other replicates. This high variance among different samples could have led to biased statistics. So, the rest of the analysis was done using only NC, BS, BLECs, HLC, PTL, and PODO.

The results showed the variance seen in PCA was seen mostly due to differences in cell models and not due to PQ concentrations. 341 probes were found to be commonly differentially expressed in NC, BS, BLECs, HLC, PTL, and PODO. No common pathways were found among all 6 cell models. Nrf2-mediated oxidative stress response which is known to be activated in response to PQ exposure (Dou et al. 2016) was found to be the common specific stress response pathway in all models except for BS. However, the associated genes showed a concentration-dependent increase in expression for all the cell models including BS. Even after showing the concentration effect, the Nrf2-mediated oxidative stress response was not found in BS as the expression values of the associated genes were not high enough to cross the thresholds. UPR showed a concentration-dependent increase in BLECs, NC, PTL, and HLC. The endoplasmic reticulum (ER) stress pathway which activates UPR has also been linked to PQ exposure in lung epithelial cells of lungs (Chinta et al. 2008; Omura et al. 2013). BS again didn't show this response, but the associated genes showed a concentration-dependent behavior after 24hrs, and therefore repeated exposure is needed in BS to see this response. ESR-mediated signaling was found in 4 cell models BS, NC, BLECs, and HLC. It has also been recently reported in the literature study that the toxicity of PQ in rat neurons leads to estrogen dysfunction (Moyano et al. 2020).

Like pathways, there were no common genes found in all 6 models. MAFF and PPP1R15A were found to be common in a combination of all models except for BS. VEGFA, ATF3, and GDF15 were found to be commonly expressed in a set of combinations of 4 models. Apart

from the shared genes and pathways, the different cell models also have a few specific genes and pathways expressed. BS and NC had the top upregulated genes related to the cell cycle and top downregulated genes related to cytoskeleton and mitosis, respectively. Few and weak stress response-related gene activation showed that the brain doesn't have a strong defense mechanism against PQ and therefore, shows its high sensitivity towards PQ (Kuter et al. 2010).

The genes related to PQ transport and enzymes involved in redox cycling and antioxidant defense were also looked upon to see if they can help in explaining the sensitivity of organs to PQ as seen in the cytotoxic curves. The amino acid transporters SLC3A2 and SLC7A11 showed the highest basal expression in BS, NC, and PODO. Antioxidant defense SOD2, SOD3, GSTA2, and GSTM3 showed the highest basal expression in HLC. The sensitivity to PQ in the brain was the most and could be explained due to high expression of amino acid transporters SLC3A2 and SLC7A11 in BS and NC which has also been reported in literature studies by Wang et al 2021 (under review) (Kuter et al. 2010). The highest resistance of HLC against PQ could be explained due to the antioxidant defense genes like SOD and GSTs.

The study, therefore, helped in the assessment of chemical toxicity in different iPSC models in parallel. The known stress responses activated against PQ were found for these cell models. Also, BS and NC were found to be the most sensitive, and HLC was found to be the most resistant based on the cytotoxic curves.

2. iPSC PTL model for studying cadmium

2.1 Temporal transcriptomic alterations of cadmium exposed human iPSC-derived renal proximal tubule-like cells

Pranika Singh[*1,2], Vidya Chandrasekaran[*3], Barry Hardy[1], Anja Wilmes[3], Paul Jennings[#3] and Thomas E. Exner[#1,4].

[1]Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, 4057 Basel, Switzerland

[2] Division of Molecular and Systems Toxicology, Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

[3]Division of Molecular and Computational Toxicology, Chemistry and Pharmaceutical Sciences, AIMMS, Vrije Universiteit Amsterdam, The Netherlands.

[4]current address: Seven Past Nine d.o.o., Hribljane 10, 1380 Cerknica, Slovenia

* equally contributing first authors
# Corresponding authors

Paul Jennings, VU Amsterdam p.jennings@vu.nl
Thomas Exner, Seven Past Nine, thomas.exner@sevenpastnine.com

Contribution:

- Planned and proposed the study the effects of cadmium on iPSC cells
- Quality controlled, analysed, visualized, and interpreted the data.
- Drafted the manuscript
- Worked in alignment with the other shared first author in the revision process

Toxicology in Vitro 76 (2021) 105229

Contents lists available at ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

Temporal transcriptomic alterations of cadmium exposed human iPSC-derived renal proximal tubule-like cells

Pranika Singh [a,b,1], Vidya Chandrasekaran [c,1], Barry Hardy [a], Anja Wilmes [c], Paul Jennings [c,*], Thomas E. Exner [d,*]

[a] Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, 4057 Basel, Switzerland
[b] Division of Molecular and Systems Toxicology, Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland
[c] Division of Molecular and Computational Toxicology, Chemistry and Pharmaceutical Sciences, AIMMS, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[d] Seven Past Nine d.o.o., Hribljane 10, 1380 Cerknica, Slovenia

ARTICLE INFO

Keywords:
HSPA6
Metallothionein
iPSC
In vitro
Cadmium
Metal responsive transcription factor-1
Temporal expression patterns
Proximal tubules
Nuclear factor erythroid-2-related factor 2

ABSTRACT

Cadmium is a well-studied environmental pollutant where the kidney and particularly the proximal tubule cells are especially sensitive as they are exposed to higher concentrations of cadmium than other tissues. Here we investigated the temporal transcriptomic alterations (TempO-Seq) of human induced pluripotent stem cell (iPSC)-derived renal proximal tubule-like (PTL) cells exposed to 5 μM cadmium chloride for 1, 2, 4, 8, 12, 16, 20, 24, 72 and 168 h. There was an early activation (within 4 h) of the metal and oxidative stress responses (metal-responsive transcription factor-1 (MTF1) and nuclear factor erythroid-2-related factor 2 (Nrf2) genes). The Nrf2 response returned to baseline within 24 h. The Activator Protein 1 (AP-1) regulated genes HSPA6 and FOSL-1 followed the Nrf2 time course. While the MTF1 genes also spiked at 4 h, they remained strongly elevated over the entire exposure period. The data and cell culture model utilised will be useful in further research aimed at the refinement of safe human exposure limits for cadmium, other metals and their mixtures.

1. Introduction

Nephrotoxicity, i.e., injury to one or more cell types among over 20 different cell types present in the nephron, is a potential manifestation of exposure to a wide variety of xenobiotics and can lead to acute and chronic kidney disease (Pfaller and Gstraunthaler, 1998). The most frequently reported affected part of the nephron to xenobiotic injury is the proximal tubule epithelium (Barnett and Cummings, 2018). This is primarily due to the abundant and diverse transport capabilities of these cells which can result in high cytosolic levels of certain xenobiotics (George et al., 2017). Cadmium is an example of a proximal tubule nephrotoxin (Prozialeck and Edwards, 2012) and is used here as a case study compound to test if induced human Pluripotent Stem Cell (iPSC)-derived proximal tubule (PTL) cells can be utilised in hazard characterisation studies, at their present stage of development. Existing renal cell lines including the RPTEC/TERT1 and HK2 cells (Wilmes et al., 2011) serve well as biological tools representing the proximal tubule to perform mechanism driven omic studies to understand nephrotoxicity (Aschauer et al., 2015). However, iPSCs which have the potential to differentiate into any cell type, are a promising emerging tool to overcome some of the limitations of traditional cell lines such as donor specificity (Jennings, 2015; Murphy et al., 2019; Pamies et al., 2018; Wilmes and Jennings, 2015). This study is part of the Marie Sklodowska-Curie Action - Innovative Training Network in3, an integrated interdisciplinary approach to animal-free nanomaterial and chemical safety assessment, (https://estiv.org/in3/, grant no. 721975), which had a focus on the utilisation of iPSC technology for chemical safety assessment. In the present study we utilised a recently described iPSC-PTL differentiation protocol, which produces proximal tubule like (PTL) cells, as characterised by epithelial polarisation, responsiveness to parathyroid hormone, p-glycoprotein activity and megalin (LRP2)-mediated uptake of albumin (Chandrasekaran et al., 2021).

Cadmium is a common environmental pollutant associated with several adverse health conditions including renal disease (Satarug and

Abbreviations: iPSC, induced pluripotent stem cell; Nrf2, Nuclear factor erythroid-2-related factor 2; MTF1, Metal-responsive transcription factor-1; AP-1, Activator Protein 1; PTL, Proximal tubule-like cells; LFC, Log fold change; ER, Endoplasmic Reticulum Stress; IPA, Ingenuity Pathway Analysis; RC, Read count.

* Corresponding authors.
E-mail addresses: p.jennings@vu.nl (P. Jennings), thomas.exner@sevenpastnine.com (T.E. Exner).

[1] Equally contributing first authors.

https://doi.org/10.1016/j.tiv.2021.105229
Received 15 March 2021; Received in revised form 25 June 2021; Accepted 26 July 2021
Available online 3 August 2021

Moore, 2004). Cadmium exposure is mainly through food, via plant uptake of cadmium in soil, and via smoking (Ismael et al., 2019). On a cellular and molecular level cadmium, which is not a Fenton metal, is thought to cause oxidative stress by binding intracellular thiols (Liu et al., 2009; Prozialeck and Edwards, 2012). Cells can defend themselves from damage induced by oxidative stress by activating the nuclear factor erythroid-2-related factor 2 (Nrf2) pathway (Limonciel and Jennings, 2013; Wilmes et al., 2011). In addition to oxidative stress, cadmium can displace biological metals such as zinc (Funk et al., 1987; Tang et al., 2014), which can cause aberrant function of a wide variety of proteins. Here too, cells can mount an adaptive defence by activating the metal stress response via the metal-responsive transcription factor 1 (MTF1, also termed MRE-binding transcription factor or metal regulatory transcription factor) (Jennings et al., 2013). This pathway involves the induction of several metallothionein isomers that bind free metals in order to eliminate them from the cell and body, ultimately via renal clearance. However, the metallothionein chelation of metals is potentially adverse at the proximal tubule, as they are taken up into the cytosol via megalin-mediated endocytosis (Sabolić et al., 2010). Thus, the proximal tubule is exposed twice, once via the original ion and second via MT-metal.

Some studies have explored the temporal effect of compounds involving organs like kidney and liver (Aguayo-Orozco et al., 2018; Fan et al., 2014) and also several studies have investigated the effect of cadmium on transcriptional networks but none have extensively addressed temporal patterns of cadmium exposure. Therefore, the aim of the current study was twofold: (1) to investigate the utility of iPSC differentiated into proximal tubules for metal exposure studies, (2) to resolve the temporal alteration of metal stress and oxidative stress pathways induced by cadmium.

2. Materials and methods

2.1. Cell culture

2.1.1. iPSC culture

The human iPSC line SBAD2 clone 1, generated using Cytotune 2.0 (Thermofisher) from dermal fibroblasts within the IMI funded Stem-BANCC project (grant agreement no 115439, http://stembancc.org), (Morrison et al., 2015) was utilised here. These cells were further genetically manipulated by Prof. A. Dinnyés laboratory, Biotalentum, Hungary to contain a Green Fluorescent Protein (GFP) tag downstream of one allele of the HMOX-1 gene and designated as SBAD2 HMOX1-eGFP (although, this fluorescent property was not utilised in this study). SBAD2 HMOX1-eGFP iPSCs were cultured on Geltrex coated plates (Life Technologies A1413302), in mTeSR1 medium (StemCell Technologies 05850) and were routinely passaged with EDTA (0.02% Versene, Lonza BE17-711E) as previously described (Wilmes et al., 2017).

2.1.2. PTL differentiation

iPSCs were differentiated using the protocol described by (Chandrasekaran et al., 2021). In brief, iPSCs were cultured until approximately 70–80% confluent, washed with PBS and detached by incubating with accutase for approximately 3–5 min. Cells were resuspended at the concentration of 2×10^4 / cm^2 in medium 1 (DMEM: F12 containing a final concentration of 5 mM glucose, 2 mM glutamax (Thermo fisher 35050038), 5 μg/mL insulin, 5 μg/mL transferrin and 5 ng/mL sodium selenite (Sigma Aldrich I1884)) supplemented with 3 μM CHIR992021 (abcam 120,190), 1 μM TTNPB and 10 μM Rock inhibitor (abcam 120129). The cells were seeded on 96-well plates (cell carrier ultra) coated with Geltrex (day 0). After 42 h (day 2), the medium was refreshed with medium 1, supplemented with 1 μM TTNPB. After 72 h (day 3), the medium was replaced with medium 2 (consisting of medium 1 plus 10 ng/mL epithelial growth factor (Sigma Aldrich E9644) and 36 ng/mL hydrocortisone (Sigma-Aldrich H0135)), supplemented with 10 ng/mL FGF9 (Thermo Fisher PHG0194). After 120 h (day 5), cells were fed with medium 2. From then on, medium 2 was refreshed every 2 to 3 days without additional growth factors. Differentiated cells (PTL) were used for treatment from day 14 onwards. To maintain long term stability of the cells, medium 2 was supplemented with 1 μM GW788388 hydrate (GW7) (catalogue number: SML0116, 98% purity) from day 14 onwards.

2.2. Sample preparation

Cadmium Chloride ($CdCl_2$) was purchased from Sigma Aldrich (catalogue number 202908, purity 99.99%). Stock solution was made in deionised water at 20 mM. At day 14 of differentiation, PTL cells were exposed to medium controls and a non-cytotoxic concentration of 5 μM cadmium chloride both containing 0.1% DMSO and samples were collected at 1, 2, 4, 6, 8, 12, 16, 20, 24, 72 and 168 h after exposure. The DMSO was added to allow for internal comparisons with other DMSO soluble compounds. For samples collected at 72 h and 168 h, PTL cells were obtained from separate differentiation and were repeatedly exposed every 24 h from day 14 of differentiation to 5 μM cadmium chloride in medium 2, supplemented with 1 μM GW788388 hydrate (Fig. 1A). For sample collection, supernatant was removed, and the cells were incubated with 100 μL TempO-Seq lysis buffers. The lysates were frozen at −80 °C.

The samples were sent to BioClavis technologies Ltd. Glasgow for performing the TempO-Seq experiment using the EU-ToxRisk v2.1 panel (3565 probes) (Limonciel et al., 2018; Mav et al., 2018) with standard attenuators including primary processing to derive gene-annotated raw read counts and quality control. Standard QA/QC tests were performed on read counts and included in the BioClavis service report (Supporting information 1). As metrics, the mapped reads in positive RNA controls, signal: noise ratio and percentage of mapped reads in positive controls, and average reads/probes are reported. In addition to these, the replicate analysis of positive RNA controls and untreated negative control samples was performed to check for reproducibility of different samples.

2.3. Data accessibility and sample selection

To guarantee interoperability and reusability across all 17 TempO-Seq based studies performed in the in3 project, central data management which adhered to the FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al., 2016) was adopted. The raw read count data file (data not shown) as provided by BioClavis (after passing the QA/QC tests) was used for further analysis. This was complemented by metadata to define the cell models, differentiation states and exposure scenarios in a customized format used for all studies. The metadata file is using a tabular layout providing information on sample IDs to unambiguously associate metadata to the read counts (RCs), compound names and their identifiers, cell models, exposure conditions, number of biological replicates and other experimental conditions. Group IDs were used to group biological replicates, which represent the same experimental conditions with respect to treatment and controls, to be used later to automate the differential expression analysis. Control groups for each treatment group were defined directly in the file using 'Pairing Control ID' columns listing the group ID of the corresponding control samples. To allow easy access and sharing of metadata and data within the in3 project already anticipated in the joint development of the study designs, all the raw data and related metadata files were uploaded to an internal instance of the EdelweissData™ management system (SaferWorldbyDesign, 2020). This system has the advantage to provide, beside advanced search and browse functionality, an application programming interface (API) to allow access directly using different scripting languages.

A Python workflow was developed to obtain the data and select the samples by metadata filtering per user-provided criteria. For the temporal analysis, samples exposed to cadmium chloride and the corresponding negative controls were selected with additional filters for iPSC-derived PTL cells, dose of 5 μM and time points 1, 2, 4, 6, 8, 12, 16, 20,

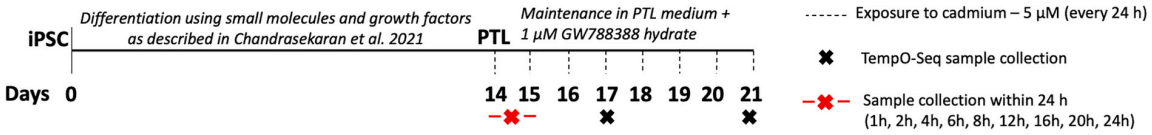


Fig. 1a. A) Treatment regime: Treatment regime showing the days of exposure to cadmium and the days at which sample was collected. Please note that between 14 and 15 days, the sample collection was done at 1, 2, 4, 6, 12, 16, 20 and 24 h.

24, 72 and 168 h. This resulted in samples from two different in3 study designs with slightly different differentiation and exposure protocols as described above, with the first group including samples up to 24 h and the second with the 72 h and 168 h samples. Probes showing low RCs across all conditions/samples were removed using the threshold of "rowsum" (which calculates the total sum of RCs in all samples per probe) of 1000 per probe across all samples (also available in the DeSeq2 package (Love et al., 2014)) directly in the Python script and specific files with the selected raw counts and metadata were produced as input for the next steps.

All data, pre-processed and processed can be found on the MolTox/VU synology data server at http://gofile.me/4bct7/GKRC36MW4.

2.4. Differential expression analysis

Normalisation and differential expression analysis were performed using the DESeq2 library (Love et al., 2014) as previously described in (Limonciel et al., 2018). Sample and treatment group-based quality control of the raw count data was performed to ensure the good quality at both sample and probe level before performing any statistical analysis. Box plots were used first for visualisation of distribution of raw count data to detect outliers in the samples. Further, clustering and Principal Component Analysis (PCA) were performed on the r-log transformed (HBC Training, 2019; Love et al., 2014) normalised RCs (for moderating variance across means) to see the variance within and between treatment groups. Standard approaches and parameters as proposed by the DESeq2 library were used for the clustering (hierarchical clustering method with Euclidean distance and complete linkage (Love et al., 2014)) and PCA (based on the top 500 genes which show the most amount of variance). The results were explored to validate that the replicates cluster together and that the exposure conditions constitute a major source of variation without any additional batch effects.

The normalisation was performed using median ratio method recommended for the further processing with the DESeq2 library (Love et al., 2014). It generates normalisation factors based on ratios for each probe between the sample and a pseudo-reference sample created as the geometric mean across all the samples. The normalised RCs were further used as an input for the differential expression analysis also using the DESeq2 library for calculating fold changes, *p*-values, base mean and adjusted *p*-values as measures for the statistical significance of the changes (Jafari and Ansari-Pour, 2019). It was further used to compare the expression of the treated group to its corresponding untreated 'control' group separately for individual time points. An R script for automating the group comparisons using Group IDs and Control pairing IDs was developed to output all treated vs untreated comparison statistics. The default Wald statistics test (Society, 2016) was used for significance evaluation. Outlier probes (low RCs and zero RCs, extreme count outlier) were detected by the automatic independent filtering method in DESeq2 (Love et al., 2014) and their corresponding adjusted *p*-values were set to NA. For initial selection of the most significant differentially expressed probes cut-offs were set to 0.05 for adjusted *p*-values and |2| for log2 fold changes (LFCs). Probes were removed if these criteria were not met at any time point.

2.5. Gene/probe–pathway association based on literature evidence

Toxicology/stress response focussed gene pathway association was made based on key transcription factors involved for the known stress responses such as Nrf2, ER stress, metal response, apoptosis, etc. Genes regulated by these transcription factors are identified through existing relevant stress response information available in the literature or from the ChIP-Seq data (data not shown) (Jennings et al., 2013; Limonciel et al., 2018, 2015; Zgheib et al., 2018). This list was then used to find if these pathways are activated at any given timepoint by using z-score calculation (Section 2.7) upon cadmium chloride exposure. Specific focus was put on the stress response pathways, which are activated in response to chemical exposure and especially on the MTF1 (metal response) and Nrf2 (oxidative stress) pathways linked to adverse effects of cadmium.

2.6. Transcription factor and gene mapping using Reactome and IPA

The literature derived probe-pathway associations were further explored and compared against the following well known existing pathway databases and software as follows:

a) Reactome (Jassal et al., 2020)(Reactome, 2020) (version 72, last accessed Sept, 2020) was used to find pathways associated with the most highly differentially expressed genes (corresponding to the differentially expressed probes) identified in the cadmium chloride samples at each time point. The "Analyse Data" and "Project to human" options were selected to translate the identifiers from the TempO-Seq differentially expressed gene list to the equivalent human identifiers used in the definition of the pathways. The pathways at different time points were selected based on their significance (as done by Reactome) as well as their relation or equivalence to the literature-derived pathways in Section 2.5. The full list of genes associated to these pathways were extracted and then further used to calculate the z-scores.
b) Ingenuity Pathway Analysis (IPA, Updated October 2020) (Krämer et al., 2014)(QIAGEN Ingenuity Pathway Analysis, 2020) was used as a second, additional knowledge source of gene-pathway associations. With the default IPA settings, the differentially expressed genes (corresponding to the differentially expressed probes) with their log fold change values at each time point were uploaded as an input to identify the associated pathways. Additionally, the pathway diagram for the pathways which were significant (based on the IPA ranking) and also related or equivalent to the literature-derived pathways in Section 2.5 were selected to get the full set of gene associations by selecting "Pathway and Tox list", "Group" and "Complex" node types to further calculate z-scores. IPA Grow tool was used to expand the group and complex nodes by one step.

2.7. Z-Score Calculation

Z-scores were calculated using the approach described in Kutmon et al. (2015), Wellens et al. (2021). The total number of probes 'N' was taken as the total number of TempO-Seq probes measured and total number of differentially expressed probes 'R' was taken as the total number of differentially expressed TempO-Seq probes meeting the criteria of adjusted *p*-value <0.05 and LFC $> |2|$ at the specified time point. For literature, the total number of differentially expressed probes belonging to the chosen pathway 'r' was taken as the total number of differentially expressed TempO-Seq probes (meeting the criteria

adjusted p-value <0.05 and LFC $> |2|$) associated to the chosen transcription factor at the specified time point and total number of probes 'n' was taken as the total number of TempO-Seq probes associated to the chosen transcription factor based on literature review. For Reactome, 'n' was taken as the total number of TempO-Seq probes found to be associated to the specific pathway in Reactome (the genes which were not part of TempO-Seq panel were excluded from this list) and 'r' was taken as the total number of differentially expressed TempO-Seq probes out of R found to be associated to the chosen pathway in Reactome. For IPA, 'n' was taken as the total number of the TempO-Seq probes found to be associated to the specific pathway in IPA (the genes which were not part of TempO-Seq panel were excluded from this list) and 'r' was taken as the total number of differentially expressed TempO-Seq probes out of R found to be associated to the pathway of interest in IPA. A z-score > 1.96 was applied as a threshold to find significant and overrepresented pathways. The following formula was used in the calculation of the z-score.

$$Z-score = \frac{\left(r - n\frac{R}{N}\right)}{\sqrt{n\left(\frac{R}{N}\right)\left(1-\frac{R}{N}\right)\left(1-\frac{n-1}{N-1}\right)}}$$

2.8. Fit temporal expression levels to profiles modelled by standard functions

The analysis based on well-established approaches described so far are targeting the first aim of this study of investigating the utility of iPSC differentiated into proximal tubular-like cells for metal exposure studies. Resolving the temporal alteration of pathways, being the second aim of the studies, is a less developed field and existing approaches like the STEM (Ernst and Bar-Joseph, 2006) method is mainly concentrating on clustering the temporal profiles. In contrast, we decided to develop a new method able to represent the general trend of individual probe expression profiles. These can then be used, on the one hand, to compare different probes and, on the other hand, to additionally evaluate the significance of the changes in the expression levels and reduce noise by not looking at the time points independently but comparing neighbouring time points.

Modelled profiles of probe expression were generated by fitting a set of predefined functions. A python script was developed to perform the fits to all functions in this set for all probes based on the standard scipy library (Virtanen et al., 2020) (scipy.optimize.curve_fit() and the functions from the scipy.stats package). The functions were selected to represent profiles of the strongest differentially expressed probes. Profile fitting is essential to reduce noise caused by the experiment actually not following one system over time but by the time evolution being approximated by the use of independent samples for each time-point. Visual inspection of the profiles, both of the mean normalised RCs and LFCs, seen in the cadmium chloride samples guided the selection of fit functions. Most overexpressed probes can be grouped into a small number of general behaviours:

1. Probes, which show a strong maximum and then a decrease to baseline level.
2. Probes, which show a maximum and then decrease to a plateau still significantly above baseline level.
3. Probes, which increase to reach a plateau. This plateau can also be reached by going through a minimum first.
4. Probes, which show a constant increase until the end of the measurement period (longer measurement times would probably move these genes into group 3).

Down-regulated probes show the same profiles with the difference that, for the fold changes, the maximum needs to be substituted for the minimum and the increase for the decrease and vice versa. Using this simple substitution schema, we will refer in the remaining description of the method to "overexpressed probes" meaning both over-expression and under-expression to avoid the complication to have to mention both variants all the time.

The selected functions used for fitting are exactly chosen to represent these general trends. Maxwell probability and cumulative probability distributions were used to describe probes, which show a significant change during one time period but then go back to baseline level and genes, which reach a plateau in the expression level and stay there until the maximum analysis time, respectively. However, Maxwell distributions are not well suited for profiles, where the increase in the values until the maximum is much faster than the following decrease back to baseline level. Therefore, we also added the gamma distribution to the list of fitting functions. Since many probes are not reaching baseline level after the first maximum, we also tried combinations of Maxwell or gamma distributions to model the maximum with Maxwell cumulative probability distributions to describe the increased levels at the end of the analysis time. Besides the normal parameters of these functions, a scaling factor in y-direction and a shift in the x-direction were optimised during the fitting procedure of the Maxwell distributions. For example, the function used for fitting the Maxwell distribution became:

$$y = c\cdot\sqrt{\frac{2}{\pi}}\ \frac{(x-d)^2 e^{-(x-d)^2/\left(2a^2\right)}}{a^3}$$

with a, c (scaling in y) and d (shift in x) being the fit parameters. For comparison, also a linear fit was used to evaluate if one of the more complex functions is able to describe the profile significantly better than this simple profile. This resulted in the following 10 functions:

1) Maxwell probability distribution without shift in x-direction;
2) Maxwell probability distribution with shift in x-direction;
3) Gamma distribution;
4) Maxwell cumulative distribution;
5) Maxwell probability distribution without shift in x-direction combined with Maxwell cumulative distribution;
6) Gamma distribution combined with Maxwell cumulative distribution;
7) Maxwell probability distribution with shift in x-direction combined with Maxwell cumulative distribution;
8) two Maxwell probability distributions;
9) one Maxwell and one Gamma distribution and
10) linear function.

Splines and high-degree polynomials were avoided in order to be less dependent on outliers and noise when comparing sequential time points. It has to be noted here that not always an optimal fit could be obtained with these functions and additional functions might have to be added when results from exposure to other chemicals are analysed using temporal profiles. However, these functions were able to describe the most relevant probes for the observed mechanisms and pathways seen after cadmium chloride exposure and were not influenced by noise in the data and were not subjected to overfitting as this would have been the case e.g., for polynomial functions with the same number of parameters.

3. Results

3.1. Quality control of TempO-Seq data and differentially expressed gene analysis

All standard quality control measures performed by BioClavis Ltd. as seen in supporting information 1 demonstrated acceptable quality of all samples well suited for further analysis. The additional quality control was performed on 2197 probes left after filtering with row sum based on RCs distributions, clustering and PCA (Supporting information 1)

confirmed the quality. All probes showed high and similar total RCs and a similar distribution across all samples with the treated samples showing some very high expressed probes. The clustering and PCA showed good similarity between different variables like replicates from one treatment scenario, different time points as well as separation between treated samples and negative control samples.

The normalised RCs and fold changes resulting from differential expression analysis showed that many probes were influenced as a

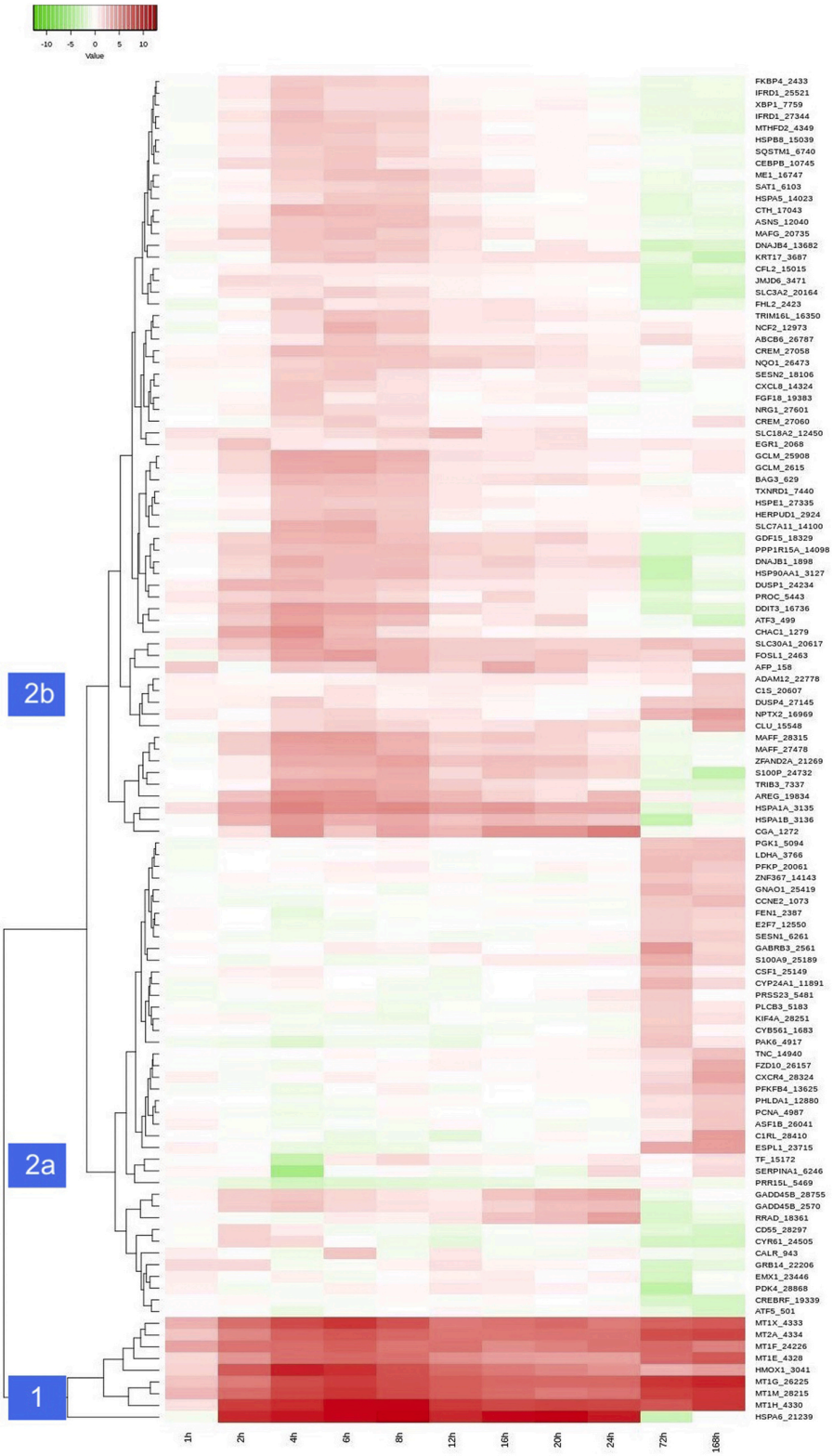


Fig. 1b. Hierarchical clustering of the 115 most differentially expressed probes. Heatmap presenting the relative expression (LFCs calculated treated vs time matched controls) of probes in the iPSC PTL cell model significantly affected on exposure to cadmium chloride (5 μM concentration) at time points 1, 2, 6, 8, 12, 16, 20, 24, 72 and 168 h. The shade of red colour represents the positive overexpression and shade of green colour represents the negative overexpression log2 fold changes. The clusters formed due to hierarchal clustering were divided into three clusters 1, 2a and 2b representing the different temporal patterns of the significantly expressed probes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

response to the treatment with cadmium chloride. To concentrate on the major changes, strong criteria were applied to select the set of probes used for further analysis. Based on the automatic independent filtering in DESeq2, adjusted *p*-values of the probes having low mean normalised counts were automatically set to NA (Love et al., 2014). As a result, of a total of 2197 probes 341 probes at 1 h (mean count per probe ≤32.54) and 43 probes at 168 h (mean count per probe ≤15.65) (data not shown) were flagged as outliers. For all other time points, no probes needed to be removed based on this criterion. Most of these filtered probes were not significantly differentially expressed at any time point except for probes ABCB6_26787 (significantly differentially expressed at 6 h and 8 h) and ATF3_499 (significantly differentially expressed from 2 h to 8 h and at 20 h). Applying cut-offs of adjusted *p*-values <0.05 and of LFCs) > |2| thus, resulted in a total of 115 significantly differentially expressed

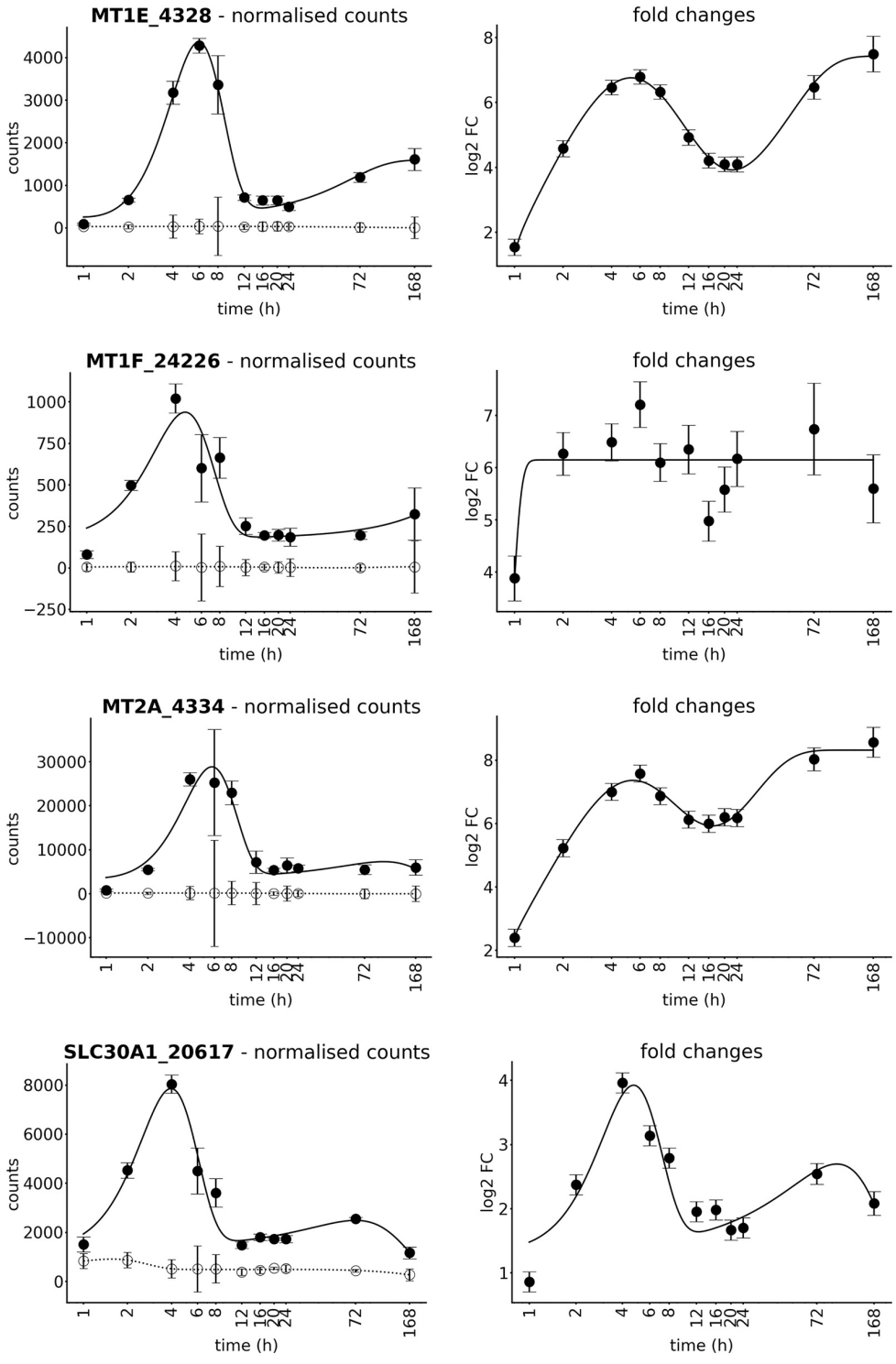


Fig. 2a. The temporal expression profiles of the differentially expressed (A) MTF1 induced metal response pathway related genes over all the time points. RCs of the treated (•) and control samples (○) are given in columns 1 (left) and the corresponding LFCs in columns 2 (right). Values of low significance (adjusted *p*-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard deviation and the standard error for the normalised counts and fold changes, respectively. The logarithmic scale was used in the x-axis so that the time points until 24 h can be better separated from each other.

probes, summing up the number of differentially expressed genes (passing the cut-offs) over all different time points. Fig. 1B presents these in the form of a heatmap where the expression levels of these probes across all the time points are shown to highlight and explore the relative expression of these probes between different time points.

The hierarchical clustering visualised in the form of the heatmap (using the heatmap.2 function of the gplots library) (Warnes et al., 2016) showed that these probes can be roughly separated into three clusters (clusters 1, 2a and 2b), based on both their expression levels and temporal behaviour. Cluster 1 combined all the genes with very high overexpression levels across all the timepoints. One of the highest expressed probes in cluster 1 is metallothionein MT1X_4333 with a LFC of 9.27 ± 0.38 at 6 h (resulting from RCs of 3237.73 ± 581.44 for treated samples and 5.18 ± 3.89 for negative control samples at that time point). Here and in the following, standard deviations are given for the mean values of RCs and standard errors for the LFCs. When it comes to absolute expression levels, MT1G_26225 (142,499.68 ± 50,186.02 for treated samples and 311.14 ± 28.47 for negative control samples at 6 h), MT1M_28215 (34,082.74 ± 4947.76 for treated samples and 53.18 ± 12.75 for negative control samples at 6 h) and MT2A_4334 (26,034.67 ± 1539.91 for treated samples and 203.21 ± 57.95 for negative control samples at 4 h) were among the highest expressed probes. All profiles from the cluster 1 showed a strong increase in the initial hours after exposure reaching a maximum around 4–6 h and then a decrease to a minimum at 24 h (Fig. 2A). After that, the profiles showed a plateau or slight increase at later time points (72 h and 168 h). It should be noted that since the RC expression levels of the metallothionein in the negative control samples decreased over time, the decrease after the first maxima was more pronounced in the RCs than in the LFCs (data not shown). This can result in higher fold change values at the late time points than at the first maximum even if the RC is lower.

The other two probes of cluster 1 were HMOX1_3041 and HSPA6_21239. The profiles of HSPA6_21239 are shown in Fig. 2C. The RCs showed a similar behaviour as the metallothionein in the first hours but did not stay at a plateau and returned to baseline level at the later time points. Even with the strong decrease in the RCs, the fold changes were calculated as being almost constant up to 24 h, which was caused by decrease of the RCs of the negative control samples paralleling the RCs in the treated sample. Also, HMOX1_3041 showed high maximum fold change similar to other metallothioneins which is the reason that it is grouped with them in Cluster1.

Based on the literature studies, we found that most of the probes in Cluster 1 correspond to metallothionein encoding genes, which are clearly related to transcription factor MTF1 and metal stress response pathway except for HMOX1_3041 and HSPA6_21239. The temporal profile of the RC and LFC of these representative metallothionein genes are plotted together along with the zinc transporter encoding gene SLC30A1_20617 (which belongs to cluster 2b due to its lower expression levels but is also related to transcription factor MTF1) in Fig. 2A. The other probes associated to MTF1 are depicted in the supporting information 2.

Cluster 2b consists of probes like GCLM_2615, MAFG_20735, MAFF_27478 etc. which showed the first maximum at a time between 4 h and 8 h and more variance at later time points. Most of the probes had reduced expression at later time points (72 h and 168 h) and then stayed at the base level except for few like NCF2_12973, ZFAND2A_21269 which showed fluctuating pattern both in the treated samples and in the negative control samples (Fig. 2B). Most of the probes in Cluster 2b are related to transcription factor Nrf2 and oxidative stress based on the literature studies and are shown in Fig. 2B. The additional probes associated to Nrf2 are depicted in supporting information 2.

Cluster 2a included probes like PCNA_4987, SESN1_6261 which showed very noisy signals in the early time points (until 24 h) and no significant change compared to the negative control samples. In contrast, they then showed high log fold changes ranging between 2 and 3 in the later time points (72 h and 168 h), which is the reason they were grouped into Cluster 2a (Fig. 2C). However, this was not caused by a large increase of the expression (RCs) in the treated samples (none of the proposed fitting functions gives a better agreement with the data than a linear, almost horizontal function) but instead due to sharp decrease in the expression in the negative control samples. These probes correspond to the genes that are related to p53 transcription factor and apoptotic pathway based on the current version of the gene pathway association from the literature review (see Fig. 2C). However, other p53 related and apoptotic pathway related probes like DUSP1_24234 and GDF15_18329 (which are also related to Nrf2 transcription factor and oxidative stress pathway and belong to cluster 2b) show similar first maxima in 24 h as seen in probes from cluster 2b but showed the opposite effect as compared to PCNA_4987, SESN1_6261 with a significant increase of the expression levels in the negative control samples at the later time points (72 h and 168 h) resulting in negative LFCs and the under expression already noticed in Fig. 1B. However, it should be noted, that these later time points showed very high standard deviations in the RCs of especially the negative control samples and thus, the fold changes have a high uncertainty even if this was not fully represented by the standard errors due to the logarithmic scale. Please note that in Fig. 2 as well as later in Figs. 3-5, logarithmic x-axes were used for better separating the early time points until 24 h.

3.2. Pathway enrichment analysis

Pathway enrichment analysis was performed using manually curated gene-pathway associations based on literature review (as mentioned in Section 2.5). In total, 11 pathways were curated (data not shown). The literature derived MTF1 based metal response, Nrf2 mediated oxidative stress and Activating Transcription Factor 4 (ATF4) mediated Unfolded Protein Response pathways showed the highest overrepresentation and significance with respect to z-scores across most of the time points and are depicted in Fig. 3. The highest overrepresentation (overrepresentation is defined as z-score over 1.96) is seen for MTF1 (contributed by 8 differentially expressed genes out of 18) in most of the time points. On the other hand, maximum numbers of differentially expressed genes were seen at 6 h for Nrf2 (contributed by 14 differentially expressed genes out of 139 genes) as well as for ATF4 (contributed by 15 differentially expressed genes out of 137 genes)(Wellens et al., 2021). Nrf2 and ATF4 pathways were majorly over-represented in all the early time points between 2 h and 16 h but not in the later time points.

Gene-pathway association based on literature review were compared to commonly used pathway enrichment analysis tool such as Reactome and IPA. The pathways identified using Reactome and IPA as most relevant for the DEGs in the cadmium chloride samples at 8 h and 168 h were selected for further analysis and are summarised below. Both these time points showed the most different and distinct pathways while the other time points showed some combinations of the pathways identified at 8 h and 168 h. Please note that the statistics and ranking performed for the pathways by Reactome and IPA is generated under the assumption that all genes in these knowledge bases are also measured by the user in their experiment and study. However, since the TempO-Seq panel used in this study only covered a targeted probe set (3565 probes), we performed a significance test using the z-score based only on the TempO-Seq gene set as already described above for the literature-curated gene-pathway association.

In Fig. 4, six important pathways identified by Reactome are presented with their z-score profile over time. “Metallothioneins bind metals” (R-HSA-5661231.2) clearly showed up as the most important pathway across all time points. The z-score is more than 1.96 in all the time points (contributed by 7 differential expressed genes out of 10 genes). However, this pathway only defined the binding process by the metallothioneins (as per the definition in Reactome) but did not show any relation to the associated transcription factor MTF1 as found in the literature. Another pathway which was found was “MTF1 activates gene

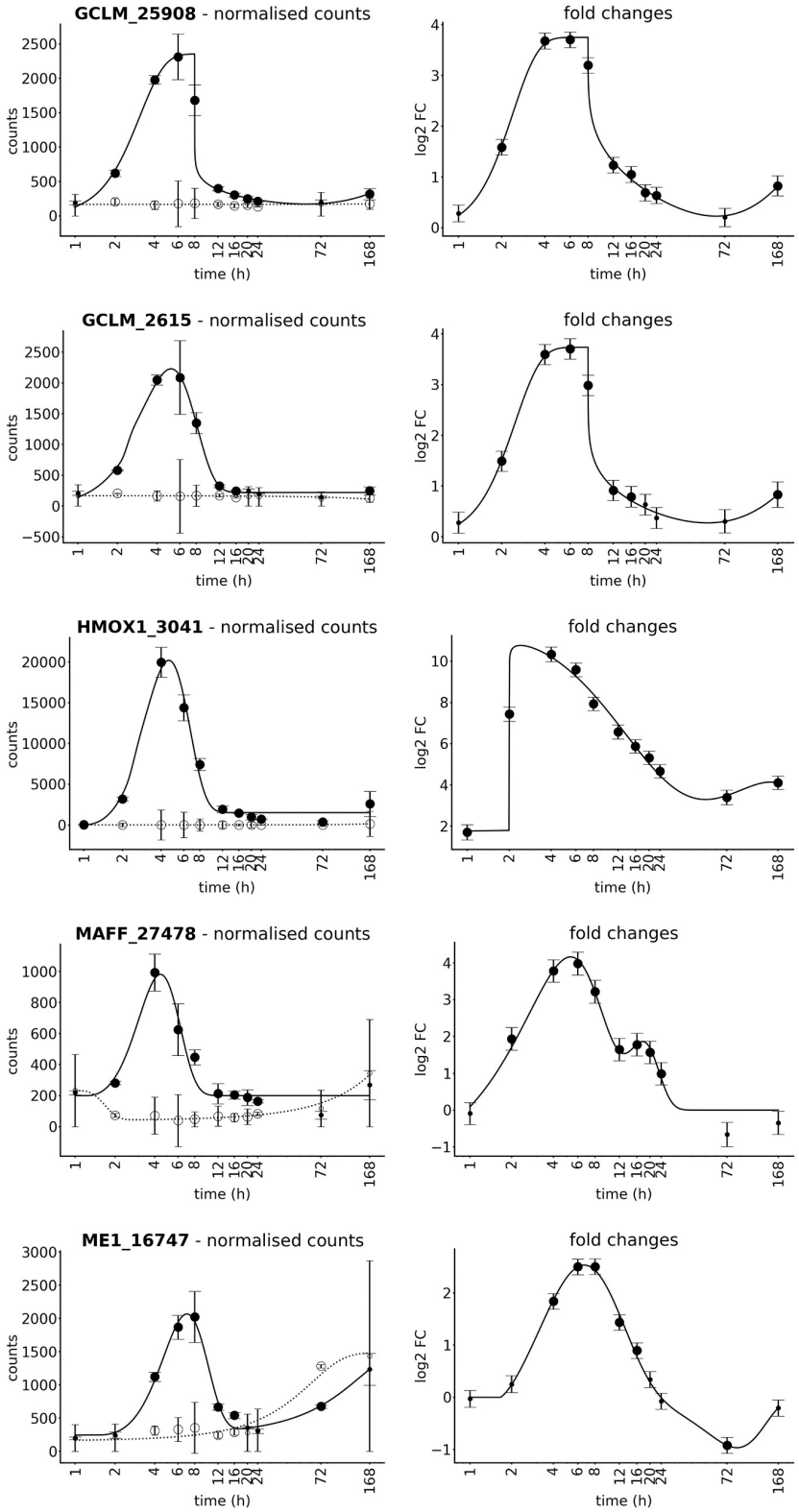


Fig. 2b. The temporal expression profiles of the differentially expressed (B) Nrf2 mediated oxidative stress pathway related genes over all the time points. RCs of the treated (•) and control samples (○) are given in columns 1 (left) and the corresponding LFCs in columns 2 (right). Values of low significance (adjusted *p*-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard deviation and the standard error for the normalised counts and fold changes, respectively. The logarithmic scale was used in the x-axis so that the time points until 24 h can be better separated from each other.

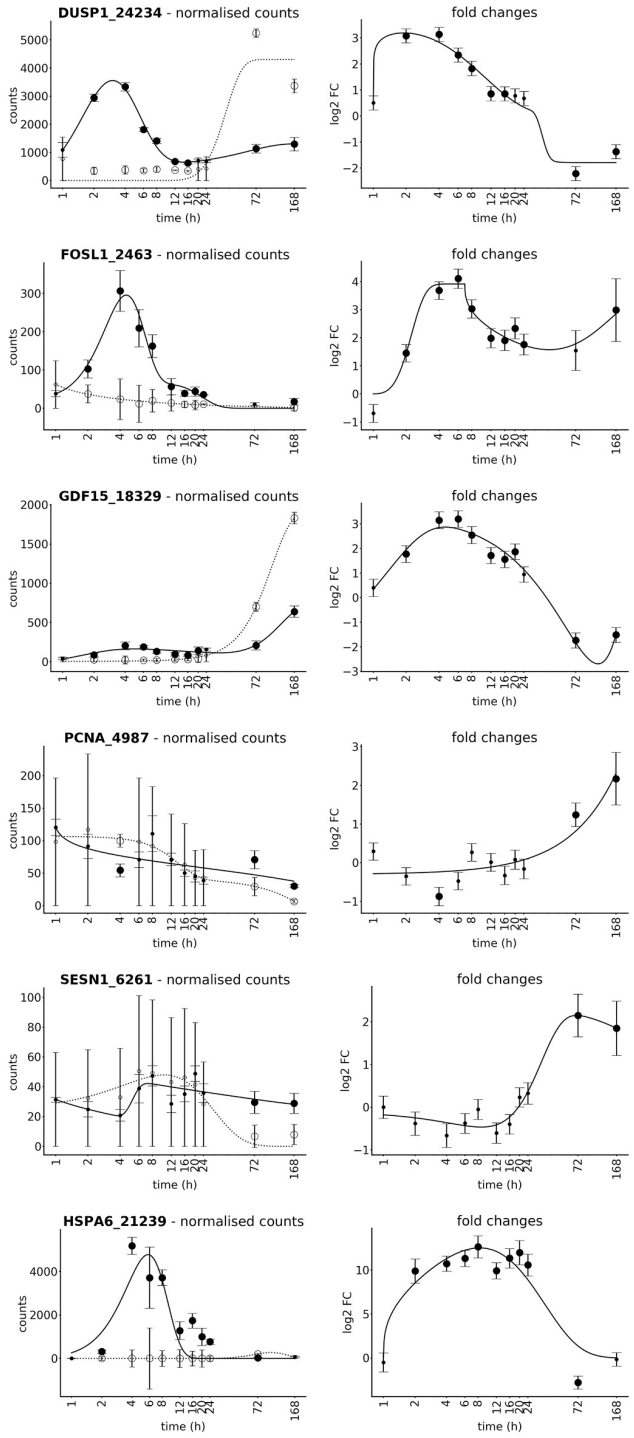


Fig. 2c. The temporal expression profiles of the differentially expressed (C) Miscellaneous genes with both normalised RCs and the LFCs over all the time points. RCs of the treated (•) and control samples (○) are given in columns 1 (left) and the corresponding LFCs in columns 2 (right). Values of low significance (adjusted *p*-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard deviation and the standard error for the normalised counts and fold changes, respectively. The logarithmic scale was used in the x-axis so that the time points until 24 h can be better separated from each other.

expression" (R-HSA-5660489.1)" which looked specifically at the transcription factor was then further examined. This pathway had only 2 genes associated, CSRP1 and SNCB, and only the first was available in the TempO-Seq probe set and had no metallothioneins associated to it. Using the hierarchy of the Reactome pathways, the two specific pathways are grouped together into the subcategory of the pathway "Response to metal ions" (R-HSA-5660526). R-HSA-5660526 was also identified using the pathway analysis to be significant. But because the number of differentially expressed genes does not change as compared to the metallothioneins specific pathway (R-HSA-5661231.2) and the number of genes in this combined pathway is larger, it was found to be on a lower rank and was therefore, not identified to be among the most important pathways.

Other pathways identified as significant by Reactome (among top 20 in the list) at 8 h were the "ATF4 activates genes in response to endoplasmic reticulum stress" (R-HSA-380994) well corresponding to the genes associated with ATF4 in the manual curated list, "Unfolded protein response" (R-HSA-381119), "Attenuation phase" (R-HSA-3371568) and "PPARA activates gene expression" (R-HSA-1989781) which are depicted with the z scores per time point in Fig. 3. The highest z score for R-HSA-380994 was seen at 6 h (contributed by 5 differentially expressed genes out of 16), for R-HSA-381119 at 6 h (contributed by 7 differentially expressed genes out of 43), for R-HSA-3371568 at 4 h (contributed by 6 differentially expressed genes out of 18) and for R-HSA-1989781 from 4 h to 8 h (contributed by 3 differentially expressed genes out of 61).

We also observed that some pathways identified by Reactome showed high overlap of the gene set between each other. We discuss here only one example of heat shock related pathways. These were identified with high z-scores with large overlap with the unfolded protein response. Therefore, they do not generate new information and are not shown here. Also, there were some very general pathways like "Cellular responses to external stimuli" (R-HSA-8953897) and "Cellular responses to stress" (R-HSA-2262752) that showed genes overlapping with the manually curated Nrf2/oxidative stress pathway. However, there were no more specific pathways assigned to these genes and especially no relation to the Nrf2 transcription factor was evident from the Reactome results.

The z-scores per time point of the three most significant pathways identified by IPA using the DEGs seen at 8 h, are shown in Fig. 5. In contrast to Reactome, IPA has a dedicated pathway for Nrf2, "NRF2-mediated Oxidative Stress Response", with a large overlap 1) with the genes associated with the literature-based gene-pathway association list as well as 2) with the "Aldosterone Signalling in Epithelial Cells" as the fourth most significant IPA pathway. Because of this overlap, the latter is not included in the figure. Also, "Unfolded protein response" was identified as highly significant. The highest z-score for "NRF2-mediated Oxidative Stress Response" was observed at 6 h (contributed by 15 differentially expressed genes out of 126). Finally, the "SPINK1 General Cancer Pathway" was identified as the most significant pathway at 8 h. The z-scores were observed to be highest at all time points and 7 differentially expressed genes out of 35. Even if this pathway seems less important for the effects of cadmium chloride on kidney cells, the gene list associated with this pathway (data not shown) covers all the metallothionein genes and thus, is the only pathway related to metal ion response in IPA. For "Unfolded protein response" the highest z-score was seen at 6 h (contributed by 8 differentially expressed genes out of 29).

All data related to gene-pathway association and z-score for literature annotated, Reactome and IPA pathways can be found on the MolTox/VU synology data server at http://gofile.me/4bct7/GKRC36MW4.

4. Discussion

This study investigated the temporal transcriptional alterations of exposure to cadmium in a recently developed iPSC derived PTL (Chandrasekaran et al., 2021). Human systemic exposure of cadmium (a non-

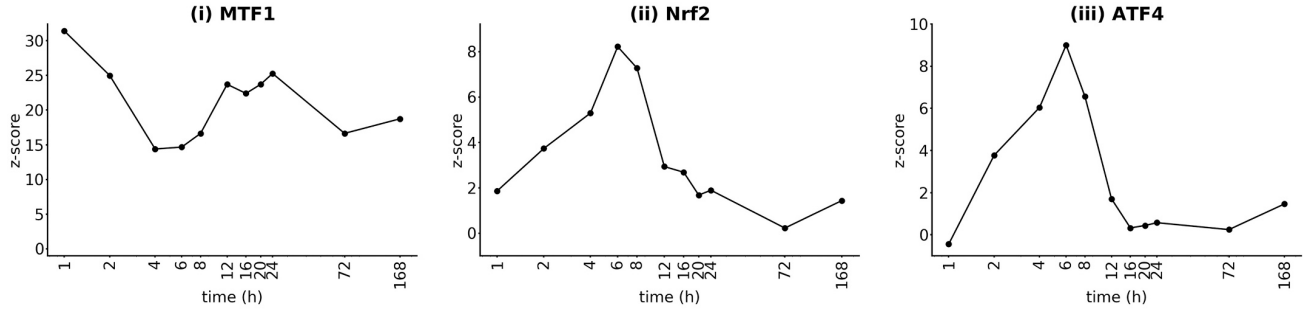


Fig. 3. *Z*-score profiles for 3 pathways based on literature evidence: The profiles of (i) MTF1 mediated metal response pathway (ii) Nrf2 mediated oxidative stress pathway (iii) ATF4 mediated Unfolded Protein Response pathway evaluating the significance of the pathways based on the expression profile across all the time points. The logarithmic scale was used in the x-axis so that the time points until 24 h can be better separated from each other.

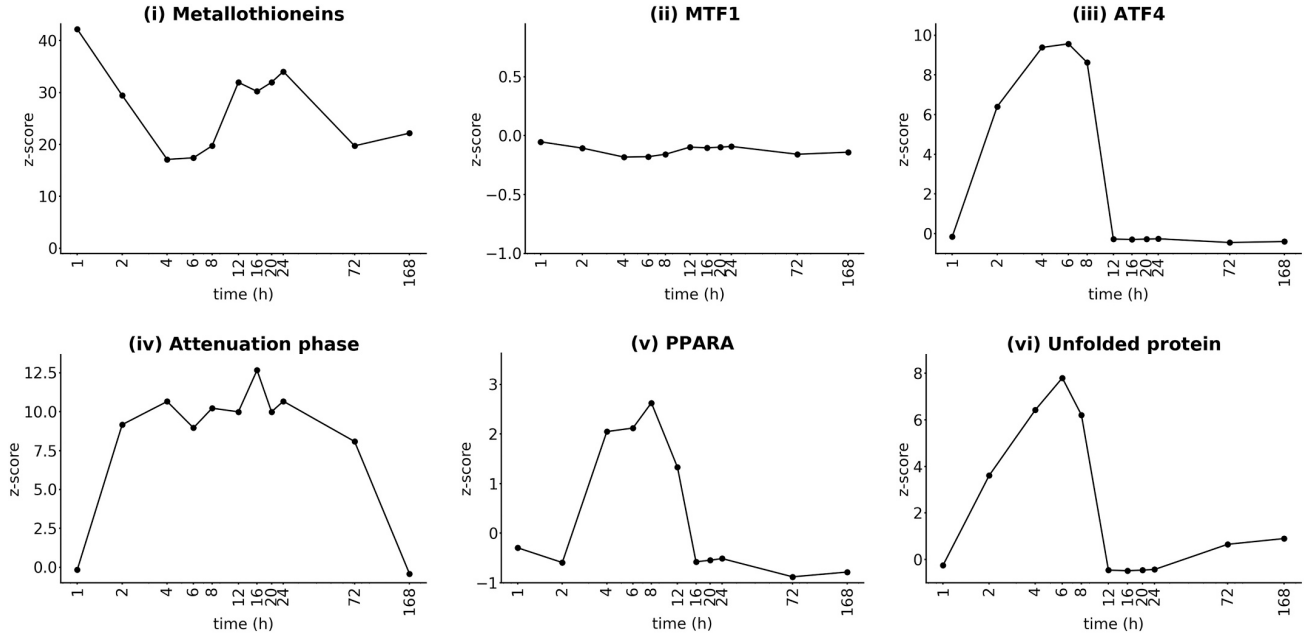


Fig. 4. Z-score profiles for 6 pathways based on Reactome library: The profiles of the (i) Metallothioneins bind metals (ii) MTF1 activates gene expression (iii) ATF4 activates genes in response to endoplasmic reticulum stress (iv) Attenuation phase (v) PPARA activates gene expression (vi) Unfolded Protein Response (UPR) evaluating the significance of the pathways based on the expression profile across all the time points. The logarithmic scale was used in the x-axis so that the time points until 24 h can be better separated from each other.

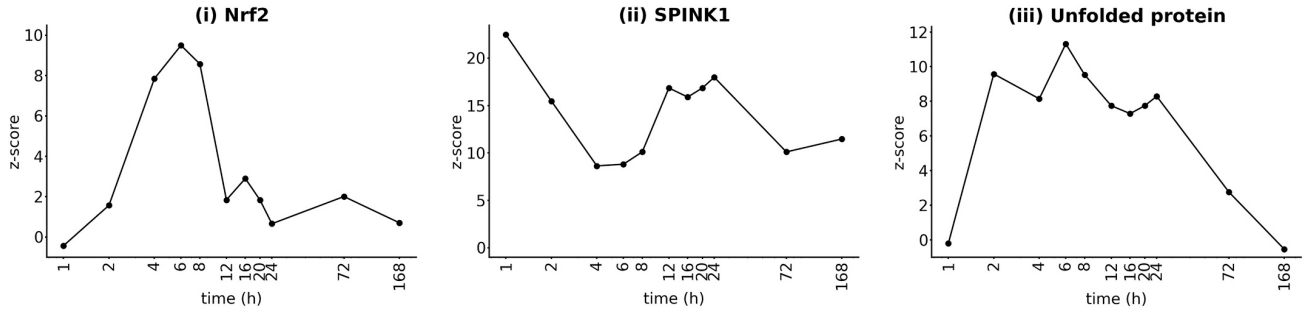


Fig. 5. Z-score profiles for 3 pathways based on IPA library: The profiles of the (i) NRF2-mediated Oxidative Stress Response (ii) SPINK1 General Cancer Pathway (iii) Unfolded protein response evaluating the significance of the pathways based on the expression profile across all the time points. The logarithmic scale was used in the x-axis so that the time points until 24 h can be better separated from each other.

biological metal) is primarily through food and water consumption but also via occupational exposure and exposure to cigarette smoke. Cadmium has and extremely long estimated half-life of between 14 and 24 years (Suwazono et al., 2009), reviewed in (EFSA, 2009; Satarug et al., 2010), accumulating in several tissues in the body particularly in the kidney and liver with concentrations reported in low to mid μg / g tissue (Uetani et al., 2006). The European Food and Safety Authority recommend safety limits of cadmium of 1 μg / g creatine corresponding to a weekly intake of 2.5 μg / kg body weight (EFSA, 2009). Roughly this would equate to an average plasma concentration safe limit of 5 μg / L. The concentration tested here is 5 μM cadmium chloride, equivalent to 562 μg cadmium / L medium, and approximately 0.5 μg cadmium load / 1 million cells. Without conducting actual cadmium measurements within the cells, it is not possible to provide the actual in vitro tissue concentration. The nominal medium concentration is in the order of 100 times the calculated safe plasma concentration.

The main aim and novelty here was to demonstrate the utility of the iPSC derived model and to investigate temporal alterations in order to better delineate the various adaptive responses to a metal response, of which the expected ones are the Nrf2 oxidative stress response and the MTF1 metal response.

Adverse effects of cadmium in human renal proximal tubular cells have been characterised at a cellular and organism level (Satarug and Moore, 2004; Wilmes et al., 2011) and arise due to oxidative injury and endogenous metal replacement. Nrf2 (aka NFE2L2), is a member of the cap "n" collar (CNC) subfamily of bZIP transcription factors activated in response to oxidative or electrophilic stress (Baird and Dinkova-Kostova, 2011). Nrf2 is regulated by KEAP1, a substrate adaptor for a CUL3 based E3 ubiquitin ligase complex, which targets Nrf2 for ubiquitination and 26S proteasomal degradation (Cullinan et al., 2004). Cytosolic KEAP1 senses oxidative or electrophilic stress through direct modification of cysteine thiol residues within the protein, leading to an inhibition of degradation and an accumulation of Nrf2 which translocate to the nucleus and heterodimerises with members of the small maf protein family and initiates transcription by binding to antioxidant / electrophilic response elements (ARE/EpRE) (Itoh et al., 1997). Genes with ARE elements in their promotor region include those involved in glutathione synthesis and reduction including glutamate-cysteine ligase, catalytic and modifier subunits (GCLC, GCLM) and glutathione reductase (GSR); genes involved in reduction processes including heme oxygenase 1 (HMOX1), NAD(*P*)H dehydrogenase, quinone 1 (NQO1); thioredoxin reductase 1 (TXNRD1), and sulfiredoxin (SRXN1) (Okawa et al., 2006; Reichard et al., 2007; Yates et al., 2009). In this study, the Nrf2 regulated genes, GCLM, HMOX1, NQO1, TXNRD1, MAFF and ME1 spiked within the first 12 h and most returned to baseline by 24 h (Fig. 1 Band 2B). *[It is noteworthy that this is more evident by observing at the RCs rather than the* log*2 normalised data. We have observed a similar data skewing by log 2 normalisation of RCs before* (Limonciel et al., 2018) *and thus, recommend always carefully interrogating the non-normalised RCs.*] We have previously shown that cadmium exposure induces Nrf2 regulated genes and proteins in human renal and pulmonary in vitro models (Aschauer et al., 2015; Forti et al., 2010; Wilmes et al., 2011).

MTF1 is primarily involved in protecting the cell from the oxidative and protein malfunctions induced by free metal ions (Gunther et al., 2012). MTF1 remains for the most part in the cytoplasm via continuous nuclear export though the nuclear MTF1 transporter XPO1. MTF1 is activated through zinc binding either directly in the case of zinc exposure or through secondary release of zinc from other proteins by metals such as cadmium (Gunther et al., 2012). Activated MTF1 homodimers accumulate in the nucleus and target genes with a metal-response element. The main target genes for MTF1 are the metallothionein families (MT1 and MT2), which act to bind and eliminate excess metal ions (Sabolic et al., 2010). It is worth noting that metallothionines were first discovered in the 1950s in the kidney of a horse poisoned with cadmium (Margoshes, 1957). Other direct targets for MTF1 include zinc transporters (SLC30A1, SLC30A2, SLC39A10), hepatitis A virus cellular receptor 1 (HAVCR1, aka KIM1), selenoproteins (SEPW1, SELH) and thioredoxin reductase 2 (TXNRD2) (Gunther et al., 2012). In the present study the metal transcription factor response was activated at 1 h and remained active throughout the experiment. Genes such as MT1E, MT1F, MT1G, MT1H, MT1M, MT1X, MT2A and SLC30A1 were robustly induced from 2 h, spiked within the first 12 h, but unlike the Nrf2 genes, never returned to baseline and remained elevated over the 168 h exposure, although at a substantially lower level than peak (Fig. 2A). We have previously identified MTF1 regulated genes, including MT1G, MT1H, MT1X and SLC30A1 as highly and persistently upregulated in the renal human cell line RPTEC/TERT1 cells, cultured on microporous supports and treated for 1, 3 and 14 days with 4 and 11 μM cadmium chloride (Aschauer et al., 2015).

The HSPA6 (aka heat shock protein 70 B) response is interesting; it followed a similar pattern to the Nrf2 genes spiking at 4 h and reducing to base levels by 24 h. HSPA6 has been observed as deregulated in several studies using cadmium (Forti et al., 2010; Han et al., 2007; Hofmann et al., 2014; Wada et al., 2007); however, it has also been studied in many other stress related contexts. Recent studies point to HSPA6 as a transcription regulator. Heat shock proteins themselves were originally characterised by their induction in thermal stress, but other stressors have also been identified, which may or may not be related to heat generated by intracellular events. While the exact transcriptional regulation of HSPA6 is not known, some studies have shown positive and negative regulators and one study identifies an AP-1 site (Ramirez et al., 2015). This is interesting as AP-1 was first discovered as a transcription factor that bound to a cis-regulatory element of the MT2A promoter (Lee et al., 1987). FOSL1 followed the same temporal pattern and is also thought to be regulated by AP-1 (Young and Colburn, 2006). Thus, it is possible that some of the metallothionein genes share the same early regulation as HSPA6 and FOSL1.

The results from this study can be interpreted as follows. The MTF1 pathway is activated within minutes as the intracellular free cadmium increases. The metal chelating genes and the metal exporting transporter genes take time to be operational. The Nrf2 genes, which respond to oxidative stress, spike simultaneously with the metal genes, peaking at approximately 4 h. Cadmium induced oxidative stress begins to reduce at this point due to the dual action of the Nrf2 and MTF proteins in reducing oxidative stress, increasing oxidated glutathione and decreasing the amount of intracellular free cadmium. Nrf2 activation thus begins to subside and returns to baseline levels by 24 h. The sampling frequency after the first 24 h is not sufficient to know if there was an Nrf2 spike on subsequent exposures. However, repeated exposure to cadmium chloride resulted in a sustained activation of the MTF pathway.

The ultimate goal in the development of new *in vitro* and *in silico* tools is to improve the health of human populations, eventually entirely without the use of animal studies. Biological pathway compilations are an excellent way to condense the knowledge from literature into an easily accessible resource. Many such tools are now available, including Reactome and IPA or even comparative tools combining multiple of these pathway libraries such as CONSENSUSPathDB. However, these are not necessarily developed with a toxicology or safety focus and are often suboptimal for finding mechanisms associated with adverse effects. Therefore, the identified areas of concern are highly dependent on the pathway library chosen as demonstrated here. Only the Nrf2 pathway and the associated stress response can be clearly identified when using IPA. The metallothioneins are associated with the "SPINK1 General Cancer Pathway" as the only pathway including metal response as one of its subparts. In contrast, Reactome has a specific metallothioneins pathway and using the Reactome hierarchy of pathways, this can also be linked to the transcription factor MTF1. Linking from here to other genes regulated by MTF1, like specific membrane metal exporters such as SLC30A1, is then possible but would need analysis beyond pathway enrichment.

More concerningly, important pathways may also be missing also for

other adverse / adaptive effects. For example, Nrf2 oxidative response is not included in Reactome. Tools combining multiple pathway libraries might reduce this problem, but these could result in outputting many unrelated and very specific pathways only because they show gene overlap with the general response pathways of toxicological relevance. More work has to be performed on multiple compounds to see how the consensus approach compares to the transcription-factor-based pathway enrichment used specifically in the in3 project. Looking at multiple time points or the full temporal profile as done in this study has the advantages, while clearly increasing the needed experimental effort, to 1) identify all relevant pathways, e.g., the Nrf2 pathway would have been identified as less relevant if only the 24 h time point would have been examined, and 2) give evidence on how the cells react to the perturbations in early responses to redress alterations in homeostasis, but often instead create a new allostatic state. This can further be used to develop and validate key events of adverse outcome pathways (AOPs) at the molecular and cellular level and even derive quantitative relationships between these events. Additionally, the temporal profiling and noise-reduction using profile fitting helps to increase the confidence in the observed effects, especially in cases of borderline significance (adjusted *p*-values around 0.05 or log2 fold changes below 1) and for genes with overall low RCs. Regardless, some additional literature curation is still needed for the transcription factor association.

In summary, the study shows the utility of conducting high temporal resolution transcriptomic studies. Unravelling the time scale of compound induced alterations is highly informative. We could clearly delineate an early response (within 6 h) from a sustained response (from 24 to 168 h). In the early response Nrf2 and MTF1 genes were present along with AP-1 regulated genes HSPA6 and FOSL1. MTF1 was the main constituent of the sustained response. The new renal model which is a human iPSC derived PTL performed well and is thus well suited to such toxicological investigations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was funded by the EU project in3 a Marie Sklodowska-Curie Action - Innovative Training Network under grant no. 721975.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tiv.2021.105229.

References

Aguayo-Orozco, A., Bois, F.Y., Brunak, S., Taboureau, O., 2018. Analysis of time-series gene expression data to explore mechanisms of chemical-induced hepatic steatosis toxicity. Front. Genet. 9, 1–15. https://doi.org/10.3389/fgene.2018.00396.

Aschauer, L., Limonciel, A., Wilmes, A., Stanzel, S., Kopp-Schneider, A., Hewitt, P., Lukas, A., Leonard, M.O.M.O., Pfaller, W., Jennings, P., 2015. Application of RPTEC/TERT1 cells for investigation of repeat dose nephrotoxicity: a transcriptomic study. Toxicol. Vitr. 30, 106–116. https://doi.org/10.1016/j.tiv.2014.10.005.

Baird, L., Dinkova-Kostova, A.T., 2011. The cytoprotective role of the Keap1-Nrf2 pathway. Arch. Toxicol. 85, 241–272. https://doi.org/10.1007/s00204-011-0674-5.

Barnett, L.M.A., Cummings, B.S., 2018. Nephrotoxicity and renal pathophysiology: a contemporary perspective. Toxicol. Sci. https://doi.org/10.1093/toxsci/kfy159.

Chandrasekaran, V., Carta, G., da Costa Pereira, D., Gupta, R., Murphy, C., Feifel, E., Kern, G., Lechner, J., Cavallo, A.L., Gupta, S., Caiment, F., Kleinjans, J.C.S., Gstraunthaler, G., Jennings, P., Wilmes, A., 2021. Generation and characterization of iPSC-derived renal proximal tubule-like cells with extended stability. Sci. Rep. 11, 11575. https://doi.org/10.1038/s41598-021-89550-4.

Cullinan, S.B., Gordan, J.D., Jin, J., Harper, J.W., Diehl, J.A., 2004. The Keap1-BTB protein is an adaptor that bridges Nrf2 to a Cul3-based E3 ligase: oxidative stress sensing by a Cul3-Keap1 ligase. Mol. Cell. Biol. 24, 8477–8486. https://doi.org/10.1128/MCB.24.19.8477-8486.2004.

EFSA, 2009. Cadmium in food - scientific opinion of the panel on contaminants in the food chain. EFSA J. Wiley-Blackwell Publishing Ltd. https://doi.org/10.2903/j.efsa.2009.980.

Ernst, J., Bar-Joseph, Z., 2006. STEM: a tool for the analysis of short time series gene expression data. BMC Bioinform. 7, 1–11. https://doi.org/10.1186/1471-2105-7-191.

Fan, Y., Wei, C., Xiao, W., Zhang, W., Wang, N., Chuang, P.Y., He, J.C., 2014. Temporal profile of the renal transcriptome of HIV-1 transgenic mice during disease progression. PLoS One 9. https://doi.org/10.1371/journal.pone.0093019.

Forti, E., Bulgheroni, A., Cetin, Y., Hartung, T., Jennings, P., Pfaller, W., Prieto, P., 2010. Characterisation of cadmium chloride induced molecular and functional alterations in airway epithelial cells. Cell. Physiol. Biochem. 25, 159–168. https://doi.org/10.1159/000272060.

Funk, A.E., Day, F.A., Brady, F.O., 1987. Displacement of zinc and copper from copper-induced metallothionein by cadmium and by mercury: in vivo and ex vivo studies. Comp. Biochem. Physiol. Part C. Comp. 86, 1–6. https://doi.org/10.1016/0742-8413(87)90133-2.

George, B., You, D., Joy, M.S., Aleksunes, L.M., 2017. Xenobiotic transporters and kidney injury. Adv. Drug Deliv. Rev. https://doi.org/10.1016/j.addr.2017.01.005.

Gunther, V., Lindert, U., Schaffner, W., 2012. The taste of heavy metals: gene regulation by MTF-1. Biochim. Biophys. Acta. https://doi.org/10.1016/j.bbamcr.2012.01.005.

Han, S.G., Castranova, V., Vallyathan, V., 2007. Comparative cytotoxicity of cadmium and mercury in a human bronchial epithelial cell line (BEAS-2B) and its role in oxidative stress and induction of heat shock protein 70*. J. Toxicol. Environ. Heal. Part A 70, 852–860. https://doi.org/10.1080/15287390701212695.

HBC Training, 2019. accessed November 2020, https://github.com/hbctraining/DGE_workshop>.

Hofmann, U., Priem, M., Bartzsch, C., Winckler, T., Feller, K.-H., 2014. A sensitive sensor cell line for the detection of oxidative stress responses in cultured human keratinocytes. Sensors 14, 11293–11307. https://doi.org/10.3390/s140711293.

Ismael, M.A., Elyamine, A.M., Moussa, M.G., Cai, M., Zhao, X., Hu, C., 2019. Cadmium in plants: uptake, toxicity, and its interactions with selenium fertilizers. Metallomics. https://doi.org/10.1039/c8mt00247a.

Itoh, K., Chiba, T., Takahashi, S., Ishii, T., Igarashi, K., Katoh, Y., Oyake, T., Hayashi, N., Satoh, K., Hatayama, I., Yamamoto, M., Nabeshima, Y., 1997. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem. Biophys. Res. Commun. 236, 313–322.

Jafari, M., Ansari-Pour, N., 2019. Why, when and how to adjust your P values? Cell J. 20 https://doi.org/10.22074/cellj.2019.5992.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2020. The reactome pathway knowledgebase. Nucleic Acids Res. 48 https://doi.org/10.1093/nar/gkz1031.

Jennings, P., 2015. The future of in vitro toxicology. Toxicol. Vitr. 29, 1217–1221. https://doi.org/10.1016/j.tiv.2014.08.011.

Jennings, P., Limonciel, A., Felice, L., Leonard, M.O., 2013. An overview of transcriptional regulation in response to toxicological insult. Arch. Toxicol. 87, 49–72. https://doi.org/10.1007/s00204-012-0919-y.

Krämer, A., Green, J., Pollard, J., Tugendreich, S., 2014. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 30, 523–530. https://doi.org/10.1093/bioinformatics/btt703.

Kutmon, M., van Iersel, M.P., Bohler, A., Kelder, T., Nunes, N., Pico, A.R., Evelo, C.T., 2015. PathVisio 3: an extendable pathway analysis toolbox. PLoS Comput. Biol. 11, 1–13. https://doi.org/10.1371/journal.pcbi.1004085.

Lee, W., Haslinger, A., Karin, M., Tjian, R., 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature 325, 368–372. https://doi.org/10.1038/325368a0.

Limonciel, A., Jennings, P., 2013. A review of the evidence that ochratoxin A is an Nrf2 inhibitor: implications for nephrotoxicity and renal carcinogenicity. Toxins (Basel). 6, 371–379. https://doi.org/10.3390/toxins6010371.

Limonciel, A., Moenks, K., Stanzel, S., Truisi, G.L., Parmentier, C., Aschauer, L., Wilmes, A., Richert, L., Hewitt, P., Mueller, S.O., Lukas, A., Kopp-Schneider, A., Leonard, M.O., Jennings, P., 2015. Transcriptomics hit the target: monitoring of ligand-activated and stress response pathways for chemical testing. Toxicol. Vitr. 30, 7–18. https://doi.org/10.1016/j.tiv.2014.12.011.

Limonciel, A., Ates, G., Carta, G., Wilmes, A., Watzele, M., Shepard, P.J., VanSteenhouse, H.C., Seligmann, B., Yeakley, J.M., van de Water, B., Vinken, M., Jennings, P., 2018. Comparison of base-line and chemical-induced transcriptomic responses in HepaRG and RPTEC/TERT1 cells using TempO-Seq. Arch. Toxicol. 92, 2517–2531. https://doi.org/10.1007/s00204-018-2256-2.

Liu, J., Qu, W., Kadiiska, M.B., 2009. Role of oxidative stress in cadmium toxicity and carcinogenesis. Toxicol. Appl. Pharmacol. https://doi.org/10.1016/j.taap.2009.01.029.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

Margoshes, M., 1957. A cadmium protein from equine kidney cortex. J. Am. Chem. Soc. 79, 4813.

Mav, D., Shah, R.R., Howard, B.E., Auerbach, S.S., Bushel, P.R., Collins, J.B., Gerhold, D. L., Judson, R.S., Karmaus, A.L., Maull, E.A., Mendrick, D.L., Merrick, B.A., Sipes, N. S., Svoboda, D., Paules, R.S., 2018. A hybrid gene selection approach to create the S1500+ targeted gene sets for use in high-throughput transcriptomics. PLoS One 13, 1–19. https://doi.org/10.1371/journal.pone.0191105.

Morrison, M., Klein, C., Clemann, N., Collier, D.A., Hardy, J., Heisserer, B., Cader, M.Z., Graf, M., Kaye, J., 2015. StemBANCC: governing access to material and data in a large stem cell research consortium. Stem Cell Rev. 5, 681–687. https://doi.org/10.1007/s12015-015-9599-3.

Murphy, C., Feifel, E., Jennings, P., Gstraunthaler, G., Wilmes, A., 2019. A protocol for one-step differentiation of human induced pluripotent stem cells into mature podocytes. Methods Mol. Biol. 1994, 93–99. https://doi.org/10.1007/978-1-4939-9477-9_8.

Okawa, H., Motohashi, H., Kobayashi, A., Aburatani, H., Kensler, T.W., Yamamoto, M., 2006. Hepatocyte-specific deletion of the keap1 gene activates Nrf2 and confers potent resistance against acute drug toxicity. Biochem. Biophys. Res. Commun. 339, 79–88. https://doi.org/10.1016/j.bbrc.2005.10.185.

Pamies, D., Bal-Price, A., Chesné, C., Coecke, S., Dinnyes, A., Eskes, C., Grillari, R., Gstraunthaler, G., Hartung, T., Jennings, P., Leist, M., Martin, U., Passier, R., Schwamborn, J.C., Stacey, G.N., Ellinger-Ziegelbauer, H., Daneshian, M., 2018. Advanced good cell culture practice for human primary, stem cell-derived and organoid models as well as microphysiological systems. ALTEX 35, 353–378. https://doi.org/10.14573/altex.1710081.

Pfaller, W., Gstraunthaler, G., 1998. Nephrotoxicity Testing In Vitro - What we Know and What we Need to Know, in: Environmental Health Perspectives. Public Health Services, US Dept of Health and Human Services, pp. 559–569. https://doi.org/10.1289/ehp.98106559.

Prozialeck, W.C., Edwards, J.R., 2012. Mechanisms of cadmium-induced proximal tubule injury: new insights with implications for biomonitoring and therapeutic interventions. J. Pharmacol. Exp. Ther. https://doi.org/10.1124/jpet.110.166769.

QIAGEN Ingenuity Pathway Analysis, 2020, QIAGEN, accessed November 2020. https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis.

Ramirez, V.P., Stamatis, M., Shmukler, A., Aneskievich, B.J., 2015. Basal and stress-inducible expression of HSPA6 in human keratinocytes is regulated by negative and positive promoter regions. Cell Stress Chaperones 20, 95–107. https://doi.org/10.1007/s12192-014-0529-0.

Reactome, 2020, Lincoln Stein, Peter D'Eustachio, Henning Hermjakob, Guanming Wu, accessed November 2020, https://reactome.org.

Reichard, J.F., Motz, G.T., Puga, A., 2007. Heme oxygenase-1 induction by NRF2 requires inactivation of the transcriptional repressor BACH1. Nucleic Acids Res. 35, 7074–7086. https://doi.org/10.1093/nar/gkm638.

Sabolic, I., Breljak, D., Skarica, M., Herak-Kramberger, C.M., 2010. Role of metallothionein in cadmium traffic and toxicity in kidneys and other mammalian organs. Biometals 23, 897–926. https://doi.org/10.1007/s10534-010-9351-z.

Sabolić, I., Breljak, D., Škarica, M., Herak-Kramberger, C.M., 2010. Role of metallothionein in cadmium traffic and toxicity in kidneys and other mammalian organs. BioMetals. https://doi.org/10.1007/s10534-010-9351-z.

SaferWorldbyDesign, 2020. accessed November 2020. https://saferworldbydesign.com.

Satarug, S., Moore, M.R., 2004. Adverse health effects of chronic exposure to low-level cadmium in foodstuffs and cigarette smoke. Environ. Health Perspect. 112 https://doi.org/10.1289/ehp.6751.

Satarug, S., Garrett, S.H., Sens, M.A., Sens, D.A., 2010. Cadmium, environmental exposure, and health outcomes. Environ. Health Perspect. https://doi.org/10.1289/ehp.0901234.

Society, T.E., 2016. On the formulation of wald tests of nonlinear restrictions author (s): P.C.B. Phillips and Joon Y. Park. In: Published by: The Econometric Society Stable URL. http://www.jstor.org/stable/1911359. REFERENCES Linked references are available on JSTOR for 56, 1065–1083.

Suwazono, Y., Kido, T., Nakagawa, H., Nishijo, M., Honda, R., Kobayashi, E., Dochi, M., Nogawa, K., 2009. Biological half-life of cadmium in the urine of inhabitants after cessation of cadmium exposure. Biomarkers 14, 77–81. https://doi.org/10.1080/13547500902730698.

Tang, L., Qiu, R., Tang, Y., Wang, S., 2014. Cadmium-zinc exchange and their binary relationship in the structure of Zn-related proteins: a mini review. Metallomics. https://doi.org/10.1039/c4mt00080c.

Uetani, M., Kobayashi, E., Suwazono, Y., Honda, R., Nishijo, M., Nakagawa, H., Kido, T., Nogawa, K., 2006. Tissue cadmium (Cd) concentrations of people living in a Cd polluted area, Japan. BioMetals 19, 521–525. https://doi.org/10.1007/s10534-005-5619-0.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. Pietro, Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Wada, K.I., Taniguchi, A., Okano, T., 2007. Highly sensitive detection of cytotoxicity using a modified HSP70B′ promoter. Biotechnol. Bioeng. 97, 871–876. https://doi.org/10.1002/bit.21293.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B., 2016. Package "gplots": various R programming tools for plotting data. In: R Packag. version 2.17.0.

Wellens, S., Dehouck, L., Chandrasekaran, V., Singh, P., Loiola, R.A., Sevin, E., Exner, T., Paul, J., Gosselet, F., Culot, M., 2021. Evaluation of a human iPSC-derived BBB model for repeated dose toxicity testing with cyclosporine A as model compound. Toxicol. Vitr. 105112 https://doi.org/10.1016/j.tiv.2021.105112.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Van Der Lei, J., Van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. Comment: the FAIR guiding principles for scientific data management and stewardship. Sci. Data. https://doi.org/10.1038/sdata.2016.18.

Wilmes, A., Jennings, P., 2015. The Use of Renal Cell Culture for Nephrotoxicity Investigations, Predictive Toxicology: From Vision to Reality. Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/10.1002/9783527674183.ch10.

Wilmes, A., Crean, D., Aydin, S., Pfaller, W., Jennings, P., Leonard, M.O., 2011. Identification and dissection of the Nrf2 mediated oxidative stress pathway in human renal proximal tubule toxicity. Toxicol. Vitr. 25, 613–622. https://doi.org/10.1016/j.tiv.2010.12.009.

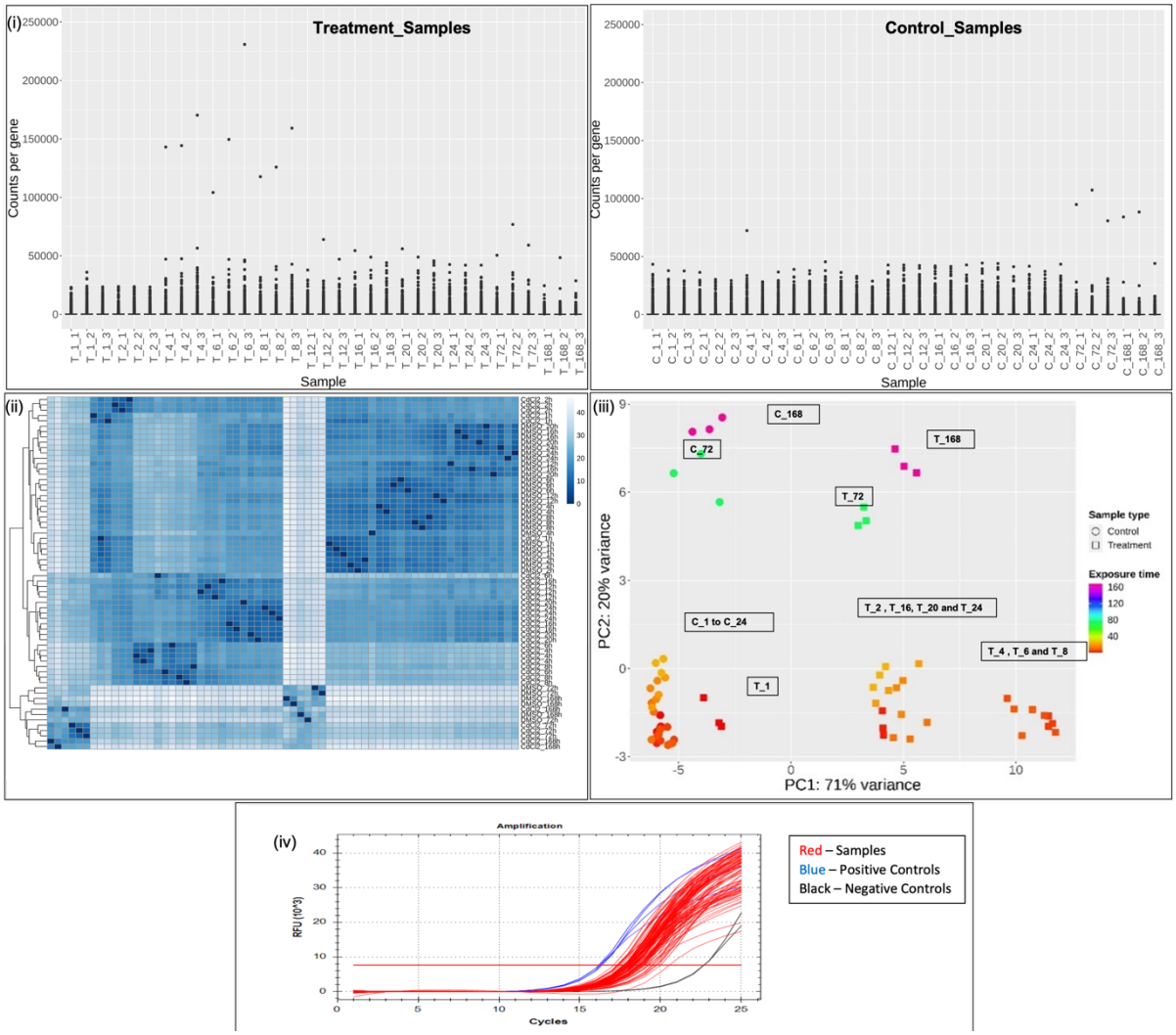
Wilmes, A., Rauch, C., Carta, G., Kern, G., Meier, F., Posch, W., Wilflingseder, D., Armstrong, L., Lako, M., Beilmann, M., Gstraunthaler, G., Jennings, P., 2017. Towards optimisation of induced pluripotent cell culture: extracellular acidification results in growth arrest of iPSC prior to nutrient exhaustion. Toxicol. Vitr. 45, 445–454. https://doi.org/10.1016/j.tiv.2017.07.023.

Yates, M.S., Tran, Q.T., Dolan, P.M., Osburn, W.O., Shin, S., McCulloch, C.C., Silkworth, J.B., Taguchi, K., Yamamoto, M., Williams, C.R., Liby, K.T., Sporn, M.B., Sutter, T.R., Kensler, T.W., 2009. Genetic versus chemoprotective activation of Nrf2 signaling: overlapping yet distinct gene expression profiles between Keap1 knockout and triterpenoid-treated mice. Carcinogenesis 30, 1024–1031. https://doi.org/10.1093/carcin/bgp100.

Young, M.R., Colburn, N.H., 2006. Fra-1 a target for cancer prevention or intervention. Gene. https://doi.org/10.1016/j.gene.2006.05.001.

Zgheib, E., Limonciel, A., Jiang, X., Wilmes, A., Wink, S., Van De Water, B., Kopp-Schneider, A., Bois, F.Y., Jennings, P., 2018. Investigation of Nrf2, AhR and ATF4 activation in toxicogenomic databases. Front. Genet. 9, 1–17. https://doi.org/10.3389/fgene.2018.00429.

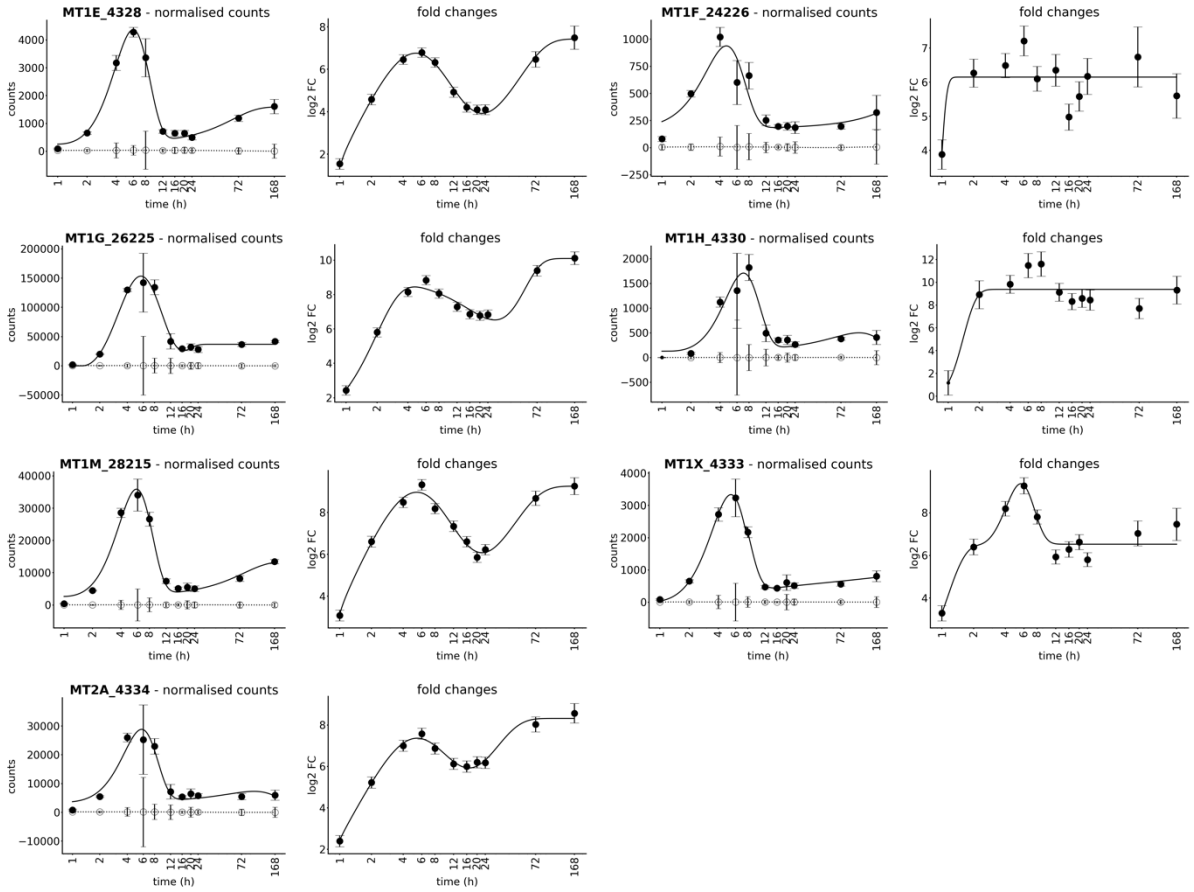
Supporting Information



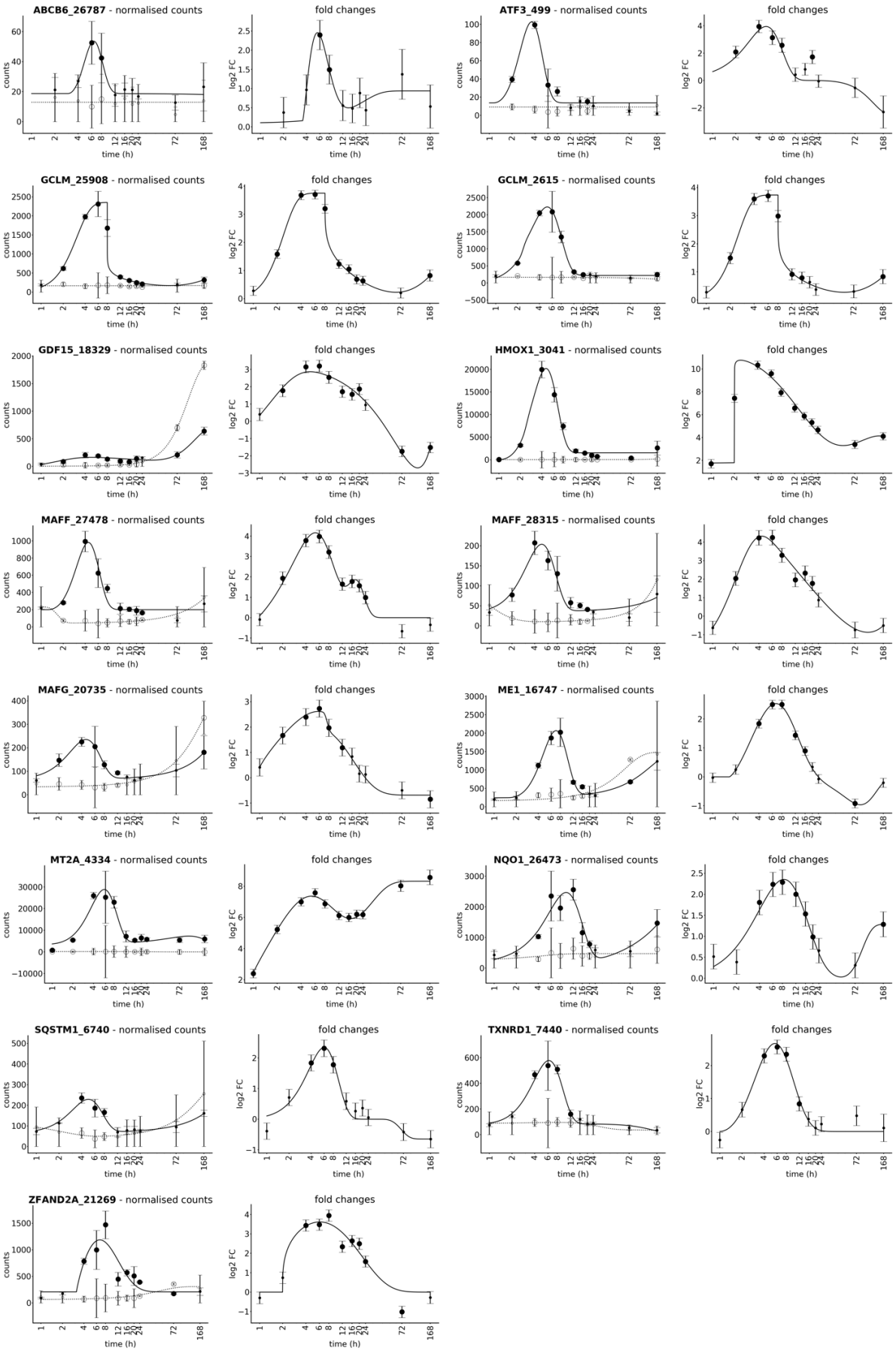
Supporting Information 1: (i) Bar plots for Treatment groups and Control groups showing raw distribution of samples and the replicates where T represents the treated samples and C represents the Control samples. It shows that the sum of total raw read counts per sample and many genes for both the controls and treatments is more than 10,000 counts, which is specified as minimum values for good quality samples. Also, the distribution of raw counts is similar between treatment and control, with some genes with high expression levels in the treatment sample as expected. Combined, this shows that all samples show a quality to be used in the further analysis and no specific approaches to handle batch effects have to be applied. (ii)The heatmap graph represents the similarities and dissimilarities between different samples. Standard hierarchal clustering was used with default Euclidean distance which calculates the distance between the samples (using the DESeq2 library (Love et al. 2014) with standard parameters). The replicate samples for individual exposure scenarios (control or treatment) and time points group together signifying changes happening in the cells due to exposure. Interesting to note is that the 4h 6h, and 8h time points are clustered together showing a significant difference to the 2h, 12h and later time points, which are forming another cluster whereas 168h and 72h forms another cluster. (iii) The Principal Component analysis (using the DESeq2 library (Love et al. 2014) with standard parameters) shows similar results as the clustering above. T represents the treated samples and the C represents the control samples and different color represents different time points the samples belong to. The first PC indicates the presence of three clusters in the sample: 1 h time point and all negative controls from 1h to 24h as first cluster, 4h, 6h and 8h time points as second cluster and 2h, 20h 24h and 16h time points as seconcluster. The second PC indicates a possible correlation between groups of negative controls at 168h and 72h to the 168h and 72h time point. The formation of clusters clearly shows the cell activity and sensitivity to $CdCl_2$. Also, no clear outliers are visible,

which was used as the final criterion of quality control. (iv) QC report by BioClavis compared the samples to both positive and negative controls before they were processed for sequencing step. The samples passed the cut-off after or alongside the positive controls showing that the samples quality is enough to proceed with sequencing. Please note that the figure is a cumulative figure for all the samples from cadmium chloride and other compounds used in another study (Arsenic, Amiodarone, Doxorubicin, Cyclosporine A, Rotenone, TTNPB, Tunicamycin and GW788388) as the samples were sent together for the sequencing.

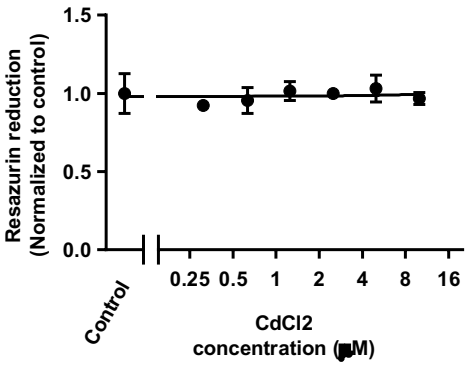
(A)



(B)


ABCB6_26787 - normalised counts
fold changes
ATF3_499 - normalised counts
fold changes
GCLM_25908 - normalised counts
fold changes
GCLM_2615 - normalised counts
fold changes
GDF15_18329 - normalised counts
fold changes
HMOX1_3041 - normalised counts
fold changes
MAFF_27478 - normalised counts
fold changes
MAFF_28315 - normalised counts
fold changes
MAFG_20735 - normalised counts
fold changes
ME1_16747 - normalised counts
fold changes
MT2A_4334 - normalised counts
fold changes
NQO1_26473 - normalised counts
fold changes
SQSTM1_6740 - normalised counts
fold changes
TXNRD1_7440 - normalised counts
fold changes
ZFAND2A_21269 - normalised counts
fold changes
counts
log2 FC
time (h)

Supporting Information 2: A) Expression profiles of probes with the normalized read counts (left) and fold changes (right) belonging to the MTF1 mediated metal response pathway. **B)** Expression profiles of probes with the normalized read counts (left) and fold changes (right) belonging to the Nrf2 mediated oxidative stress response pathway.



Supporting Information 3: Viability curve showing response to cadmium chloride. SBAD2 HMOX1-eGFP iPSC were differentiated into PTL and treated with cadmium chloride in a 96-well format from 10 µM to 0.31 µM with 2-fold dilution steps for 168 h with every 24 h repeated dose exposure. The differentiation and treatment regime were performed as mentioned in the figure 1 schematic representation. At 168 h, the resazurin assay was performed by incubating resazurin for 1 h. The conversion of resazurin to fluorescent resorufin which measures the redox potential of the cells was measured at excitation at 540 nm, emission at 590 nm. Values represent the mean ± SD (n=3) of the relative fluorescence units (RFU) for CdCl2 treatments normalized to RFU of controls (n=3) over different concentrations.

2.2 Case study on clustering genes according to temporal profiles

Pranika Singh[1,2], and Thomas E. Exner[3]

[1]Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, 4057 Basel, Switzerland

[2] Division of Molecular and Systems Toxicology, Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

[3]current address: Seven Past Nine d.o.o., Hribljane 10, 1380 Cerknica, Slovenia

Thomas Exner, Seven Past Nine, thomas.exner@sevenpastnine.com

Abstract

The iPSC-PTL cells can show the well-known pathways induced on exposure to cadmium: oxidative stress (mediated via the nuclear factor erythroid-2-related factor 2 (Nrf2)) and the metal stress response (mediated via the metal-responsive transcription factor 1 (MTF1, also termed MRE-binding transcription factor or metal regulatory transcription factor)) and could also derive the temporal evolution of these pathways by using the temporal patterns of the associated genes. The association between the genes and the pathways was made using the pathway libraries like Reactome, Ingenuity Pathway Analysis (IPA), and literature studies. Here we investigated if the clustering methods can group the related genes based on their temporal profiles and point to their association to a common pathway. Hierarchal clustering, STEM clustering, and fitting procedure were applied to the differentially expressed genes. Even though the genes showed a similar profile pattern for the metal response but it was not the same case with oxidative stress. It might be because complex pathways like oxidative stress include several mechanisms which are depending on each other and are activated at different time intervals. Therefore, without the use of pathway libraries, the grouping of the profiles of related genes is difficult and not possible with any clustering method.

Keywords: Hierarchal clustering, STEM clustering, fitting procedure, Pathway libraries.

Abbreviations: iPSC, induced pluripotent stem cell; PTL, proximal tubule like cells; Nrf2, nuclear factor erythroid-2-related factor 2; MTF1, the metal-responsive transcription factor 1; IPA, Ingenuity Pathway Analysis; STEM, Short Time-series Expression Miner; MT, metallothioneins.

1. Introduction

In the previous study (Singh et al. 2021), it was investigated that the human induced pluripotent Stem Cell (iPSC)-derived proximal tubule (PTL) cells can be utilized in the chemical study. The two most important stress responses are oxidative stress (mediated via the nuclear factor erythroid-2-related factor 2 (Nrf2)) and the metal stress response (mediated via the metal-responsive transcription factor 1 (MTF1, also termed MRE-binding transcription factor or metal regulatory transcription factor)) were seen to be activated in iPSC-PTL cells in response to cadmium chloride exposure. For modeling the temporal profiles of the genes, a fitting procedure was developed to remove any noise as described in the study before. The temporal profiles of the genes were grouped based on their association to these stress response pathways using existing pathway libraries like Reactome, Ingenuity Pathway Analysis (IPA), and literature evidence studies. Additionally, using these pathway libraries for grouping the temporal profiles of genes, we were also able to find the time points at which these pathways were activated and if they sustain or not until the end of the measurement period. We found that both oxidative stress and metal response related genes activated in the early hours. The metal response related genes showed a sustained response until 168h as compared to oxidative stress related genes which came down to baseline levels by 24h. The association between genes and pathways was done using pathway libraries therefore, the goal of this study here is to find if the clustering method that can help a) identify related temporal genes and point the involvement of these related genes toward a common pathway or mechanism without using existing pathway libraries.

2. Material and Methods

As described in more detail in chapter (Singh et al. 2021), the 115 differentially expressed (based on the adjusted p-values and log fold change) were obtained. These genes along with their log fold changes in different time points were used as an input for the clustering techniques described in the following sections.

2.1 Hierarchal clustering

Clustering is an unsupervised machine learning method that is used to group unlabeled data into clusters. The algorithm for such a method assigns the most similar entities to a cluster that are dissimilar to the entities in another cluster (Datacamp 2021). In this study, we started with the hierarchical clustering method (Zolfaghari et al. 2019) using the agglomerative approach to cluster the genes. The hierarchical clustering method has been successful in clustering the genes based on their dose-response behavior (Black et al. 2012) and as well in grouping the compounds with a similar mode of action (Wink et al. 2018). It is a data-driven approach that we used to see if the temporal profiles can be compared to find clusters of simultaneously activated genes. As, the first step, the data frame containing the log fold change values of significant genes from 1h to 168h was provided as an input to the function dist(). The method

was set to ‘euclidean’ in the dist() function. This function outputted a distance matrix containing distances calculated between the different combinations of input genes. Next, the hclust() function was applied to this matrix to perform the clustering. The method in this function was set to “complete linkage method” (4.0.0. 2020). To generate clusters with maximal dissimilarity, we chose 5 as the cluster number in the cutree () function after the visual inspection of the dendrogram tree of the hierarchical clustering.

2.2 STEM clustering

The next method used was Short Time-series Expression Miner (STEM) Clustering tool (Ernst and Bar-Joseph 2006) to see if the time dependency among successive time points could be captured which couldn’t be done by hierarchical clustering. STEM is specifically designed for clustering short time-series data gene expression data (Ernst and Bar-Joseph 2006). It has the advantage of being able to differentiate between real and random patterns by capturing the profiles from the dataset, which fit the pre-defined and pre-validated model profiles. STEM clustering algorithm first selects distinct and representative pre-defined temporal profiles (known as model profiles too). These profiles correspond to probable profiles of a temporal expression pattern of input genes. The algorithm maps the input gene’s expression profile to the most similar model profile based on the correlation coefficient (calculated inside the tool). The statistical significance of each profile is calculated by computing the enrichment score for the input genes belonging to this profile (Ernst et al. 2005). The parameter “maximum number of model profiles” represents the number of pre-defined temporal profiles which could best represent the input genes’ temporal profiles. Different values (32, 40, and 50) were tried to select the most representative and distinct profiles (Ernst and Bar-Joseph 2006). After manual inspection, 32 profiles were chosen for representing the distinct patterns of the input gene list. The results showed that the input list gene profiles were only assigned a maximum of 8 distinct profiles even if the model profiles were more than 32. Therefore, increasing the model profiles didn’t change the number of model profiles assigned to the input genes profiles. Other parameters were set to standard/default values.

2.3 Characterization of temporal profiles

A fitting procedure as described in (Singh et al. 2021) was developed to resolve the temporal patterns of the pathways which couldn’t be done by the other two methods. It concentrates on comparing the genes as well as reducing noise by comparing neighboring time points instead of looking at them independently. The modeled profiles used the set of 10 pre-defined functions (Gamma distribution, Maxwell probability distribution, Maxwell cumulative distribution, etc) which have been described in more detail in (Singh et al. 2021). This gave 10 fitted profiles for each gene. They were then manually checked for every gene to choose the one that describes the temporal profile of the gene best.

The next step was to find the criterion to characterize the resulting gene profiles (obtained from this fitting procedure) based on their time of activation and their relation to biological processes/pathways. Four general behaviors could best describe the gene profiles as described in (Singh et al. 2021): a) Genes showing a strong maximum expression and then dropping to baseline levels. b) Genes showing a strong maximum expression, dropping to plateau but still

above baseline levels. c) Genes increase in expression and attain a plateau. This plateau could also be attained by dropping to a minima first. d) Genes showing a constant increase in the expression values until the last time point. Two criteria were developed to characterize these profiles based on these gene behavioral pattern.

The first was the time point t_{max} when the first maximum or the plateau (h_1) was reached. For this, we either chose the maximum (left-hand side of **Fig 1**first maximum in cases, where the final plateau is higher) or when 95% of the maximum height of the cumulative function was reached (right-hand side of **Fig 1,** being a good approximation for the starting time point of the plateau. For the second criterion, we normalized the profile on the height of this first maximum and calculated the percentage of this value (h_2) at the end of the measurement period t_{end} (24h for the detailed short-term analysis and 168h for the long-term analysis), called *residual* in the following. The latter described the percentage of the first maximum heights remaining at the end of the analysis period with a value of 0 describing the full return to baseline, between 0 and 100 a remaining but lowered activation, and values above 100 a continuous increase over the analysis time after a short decrease separating the first maximum from the later time points.

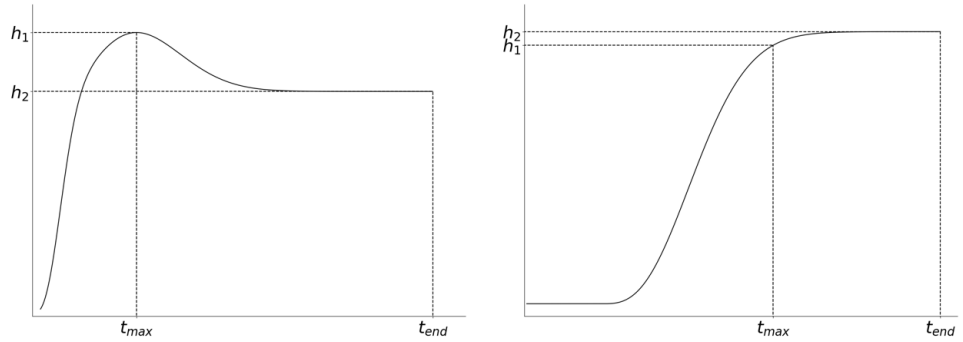


Fig 1: Definition of the criteria to characterize temporal profiles: tmax is the time point where the first maximum is reached. If the profile doesn't show a clear maximum, t_{max} gives the time when 95 % of the maximum is reached to be able to differentiate plateau from constantly increasing profiles. $residual = \frac{h_1}{h_2} \cdot 100$ describes how much of the signal of the maximum is conserved at the end of the measuring period tend.

The four behaviors described above can then easily be differentiated when plotting the two criteria in an x-y plot as follows:

- Behavior 4: genes showing their first maximum at the last time point
- Behavior 1: genes showing a residual of approximately 0%
- Behavior 3: genes showing their first maximum before the last time point and a residual of approximately 100%
- Behavior 2: genes not belonging to any of the other behaviors

Examples of these behaviors and their location in the plot of the residual vs. time point of maximum are shown in **Fig 2** useful for finding similarities in profiles.

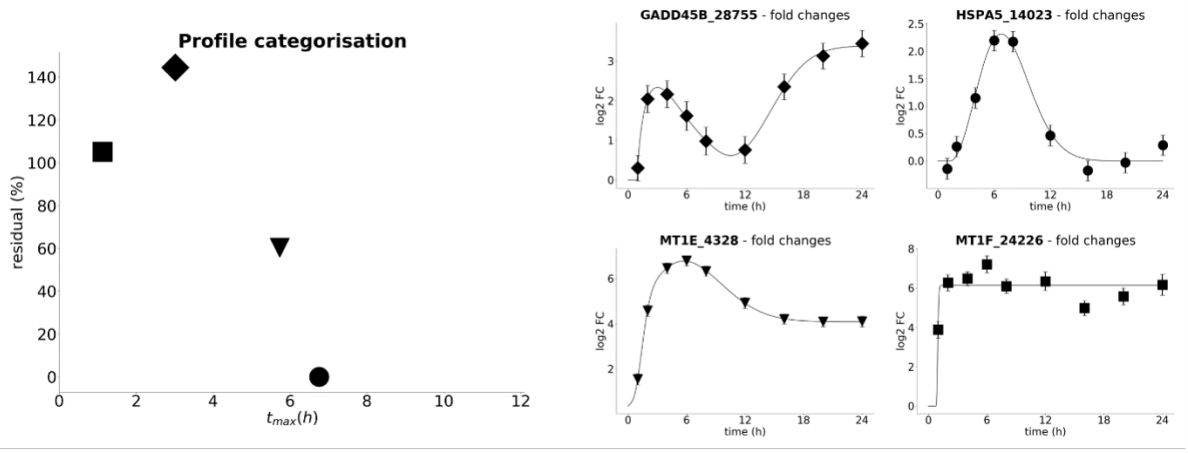


Fig 2: Plot residual vs. time of first maximum t_{max} using four profiles of different behaviors: ●behavior 1 (profile for probe HSPA5_14023) showing a maximum at around 7h and decreases to ground level (residual = 0), ● behavior 2 (profile for MT1E_4328 and profile for GADD45B_28775) showing a plateau after the first maximum being lower and higher than the first maxima respectively (residual below and above 100%) and ●behavior 3 (profile for MT1F_24226) showing a stable plateau from very early on. For behavior 4, no examples are available in the analysis of the early 24h and therefore cannot be shown in this figure.

3. Results

We used hierarchical clustering to divide 115 differentially expressed genes into five clusters. The number of genes assigned to these five clusters is shown as a histogram in **Fig 3**.

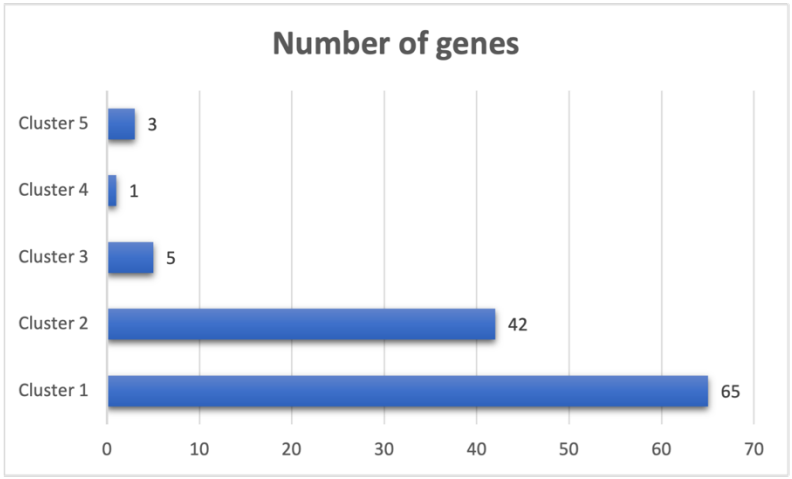


Fig 3: The histogram shows the number of genes contained in each cluster using hierarchal clustering

The results showed two types of scenarios where a) genes with a similar temporal pattern were assigned to different clusters and b)genes with different temporal patterns were assigned to the same cluster.

Cluster 1 showed for example genes SLC30A1 and EGR1 combined into one cluster even if we would argue that they are quite different in their temporal behavior as shown in **Fig 4**. SLC30A1 showed varying expression levels throughout the full-time period while EGR1 only showed one time point with a more strongly increased expression level (2h) and all others

being almost constant around a log2 fold change of 1, i.e. a two-fold increase compared to the ground levels.

Another interesting finding was for metallothioneins (MTs) genes which got divided into two different clusters. All genes showed very similar profiles with the maximum attained around 4h-6h and then an increased expression until the end time point in **Fig 5**. Metallothioneins MT1E, MT1X, MT1F and MT2A, and HMOX1 (with similar expression patterns) were assigned to cluster 3 and the other three metallothioneins MT1G, MT1H, and MT1M were assigned to cluster 5 in **Fig 4**.

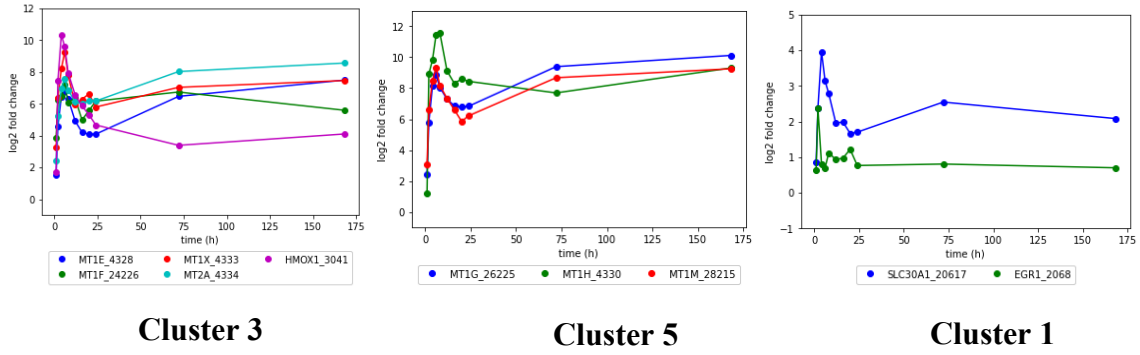


Fig 4. Temporal Profiles of the genes based on hierarchal clustering method.

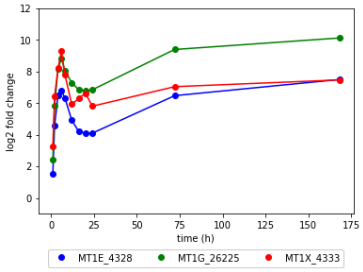


Fig 5: Temporal Profiles of the metallothionines genes based on different clusters.

STEM clustering was the next method that was used to cluster the genes. As described more in detail in the discussion in the study before (Singh et al. 2021), we saw very different behavior of the cell response in the first 24h compared to the later time point, therefore we only applied this tool to group the genes based only on the early events (until 24h). Also, STEM is designed to apply to short time-series data only (up to 8 time points).

Only 71 of the 115 genes showed high expression levels in this early time interval and therefore, only these were used as input. Out of the 32 model profiles, only 8 of them had genes assigned to them. The two profiles with the largest numbers of assigned genes (37 and 26, respectively) as shown in **Fig 6** covered almost 90% of the 71 genes on the input list. The other 6 profiles were only assigned to one or two genes.

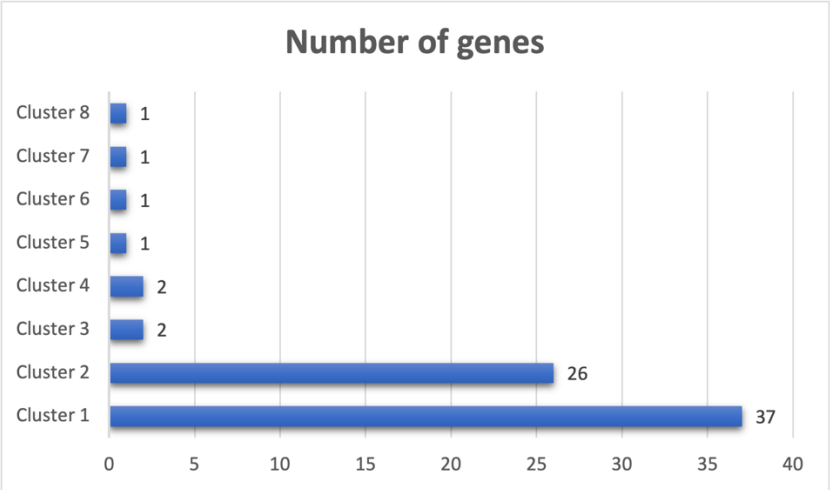


Fig 6: The histogram shows the number of genes contained in each cluster using the STEM clustering tool

The results showed two types of scenarios with STEM clustering where a) genes with similar temporal profiles were assigned to the same cluster and b) genes with different profiles were assigned to the same cluster. Model Profile 1 (consists of the maximum number of genes assigned) had all the MTs and HMOX1 assigned to it. They were all extremely highly overexpressed and sustained high expression levels until the last measurement (left of **Fig 7**). However, there were also many other genes that were assigned to the model profiles defining this cluster. Two such genes were SLC7A11 and PROC. In the figure shown on the right of **Fig 7**, these two genes had been placed together along with one of the metallothioneins (MT1E) for comparison. They were different concerning their temporal expression, where the MT1E sustained the high expression level until 24h but the other two genes dropped down to the baseline levels.

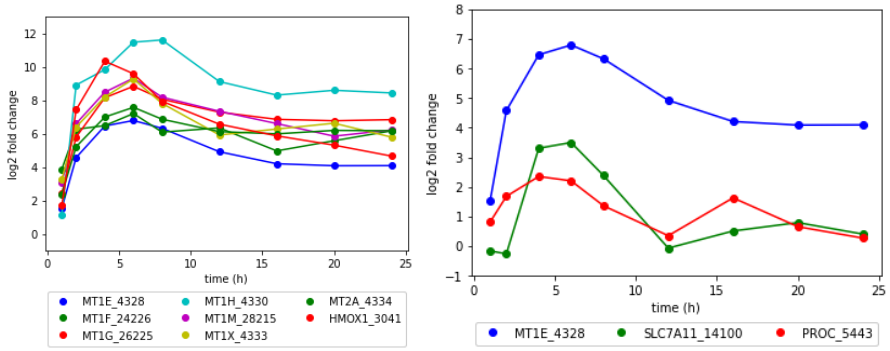


Fig 7: Temporal Profiles of the genes based on STEM clustering method.

The fitting procedure was used to overcome the limitations of the other two clustering methods which couldn't capture the temporal alteration of stress responses. These results have been taken from the publication (Singh et al. 2021) and shown in **Fig 8 and Fig 9** and have been duplicated here for characterizing the profiles.

The first set of genes with a similar expression pattern were metallothioneins (belonging to the MTF1-mediated metal response as described (Singh et al. 2021)) showed a strong increase in

the first few hours after the exposure. They were seen to be associated with MTF1-mediated metal response. The maximum was reached around 4-6h with a slight decrease in the expression yet high levels until 24 hours (**Fig 8**). At the time points 72h and 168h, the profiles showed a plateau or slight increase.

The other set of genes showing similar expression patterns were GCLM and MAFF. They were seen to be associated with Nrf2-mediated oxidative stress response as described in (Singh et al. 2021). They showed the first maxima between 4 hours and 8 hours but showed a lot of variance in the later hours 72h and 168h. But by the end of 24h, their expression levels dropped down to ground levels (Figure 9). In contrast to this behavior, HMOX1 another important oxidative stress related gene showed similar temporal profiles as MTs. It activated very early and maintained high expression levels until the last hour (Figure 9). Another interesting gene was ME1 which had maxima in the early time points like GCLM but started downregulating by 24h.

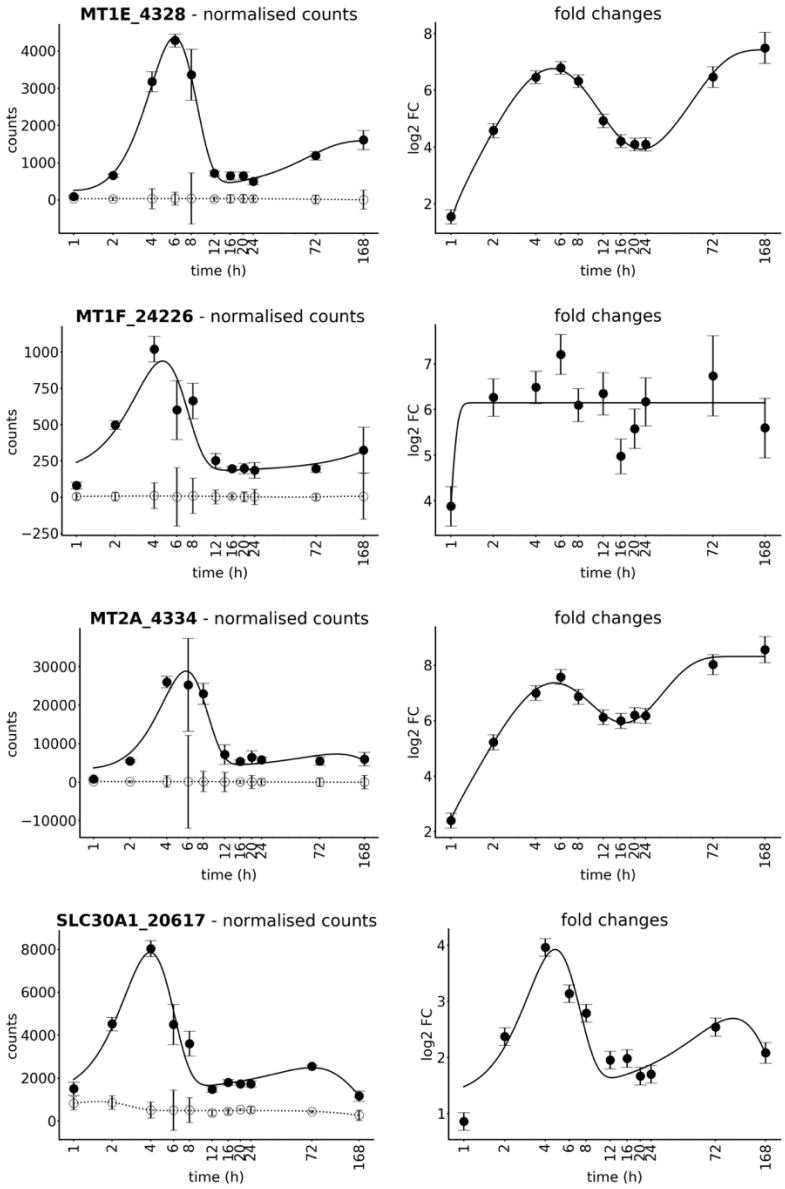
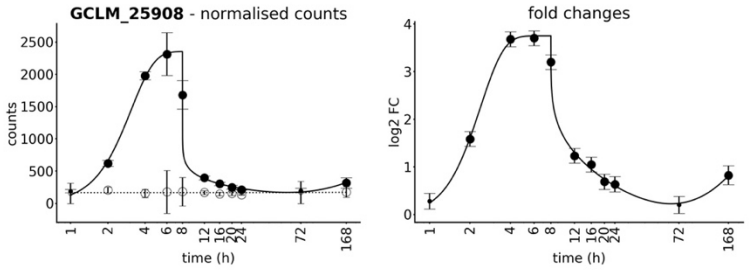


Fig 8: Temporal profiles of MTF1 genes with both normalized RCs and the LFCs over all the time points. RCs of the treated (•) and control samples (○) are given in columns 1 (left) and the corresponding LFCs in columns 2 (right). Values of low significance (adjusted p-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard deviation and the standard error for the normalized counts and fold changes, respectively. The logarithmic scale was used on the x-axis so that the time points until 24h can be better separated from each other.



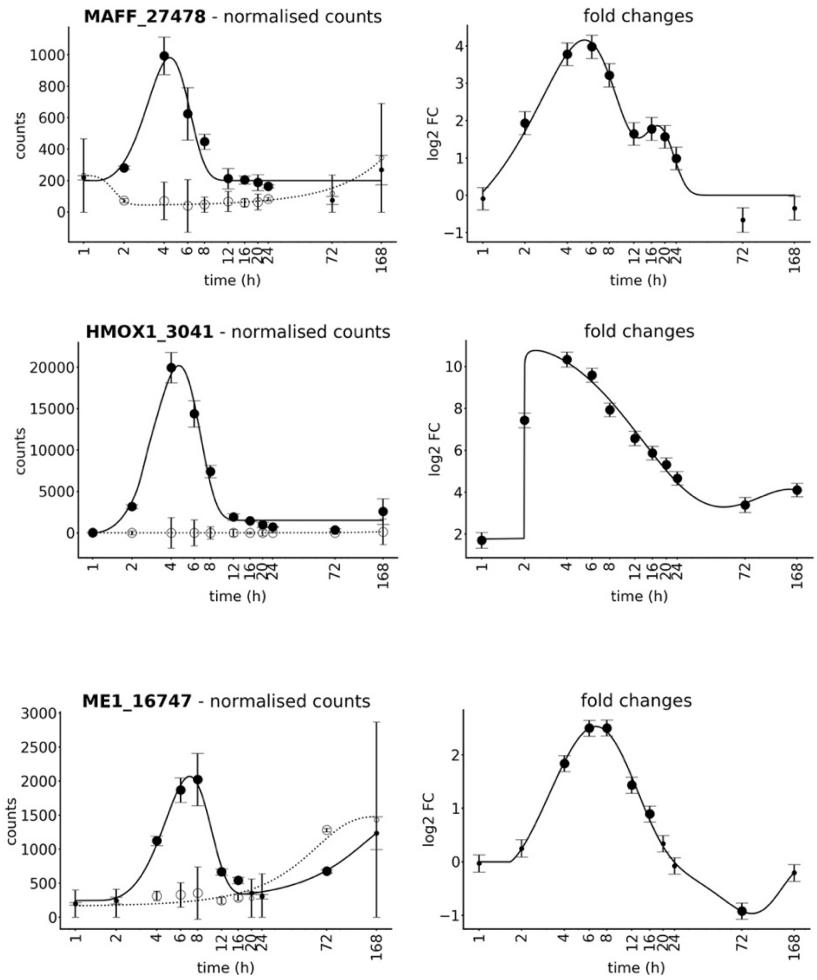


Fig 9: Temporal profiles of Nrf2 genes with both normalized RCs and the LFCs over all the time points. RCs of the treated (•) and control samples (○) are given in columns 1 (left) and the corresponding LFCs in columns 2 (right). Values of low significance (adjusted p-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard deviation and the standard error for the normalized counts and fold changes, respectively. The logarithmic scale was used on the x-axis so that the time points until 24h can be better separated from each other.

These similar and related temporal profiles were further characterized as mentioned in (Section 2.4). The characterization represented by the residual vs. time of the first maximum plot as introduced in Section 2.3 has been shown here for both Nrf2-mediated oxidative stress and MTF1-mediated metal response. These responses were associated with the many genes which were seen to be differentially expressed between 1h and 168h (Singh et al 2021).

To highlight the two responses, the corresponding probes are marked. All metallothioneins (metal response) built a dense cluster with t_{max} around 6h and a residual between 60 and 80 % (left-hand side **Fig 10)**. This represented the clear first maximum and the high levels of overexpression throughout the measurement period. The only exception is MT1F with a slightly earlier t_{max} and higher residual. This correlated with the less pronounced maximum for this probe seen in **Fig 9** compared to the other MTs. The transporter gene SLC30A1 (also assigned to this response) was separated from the rest representing the narrower maximum

reaching the maximum earlier and then the more significant decrease to around 50 % at the end of the considered time interval, which was chosen here to be the first 24h.

In contrast, the oxidative stress related genes were distributed in a larger area and overlapped with almost all other genes except the most distinct profiles with maxima before 2h and/or residuals of more than 100 % or negative. The diverse behavior of the profiles of the oxidative stress related genes can be seen on (the right side of **Fig 10)**. They all had one thing in common to have a distinct maximum in the earlier time points but varying residual ranges. HMOX1 showed a residual of more than 40% whereas genes like TXNRD1 and SQSTM1 showed a residual of 0%.

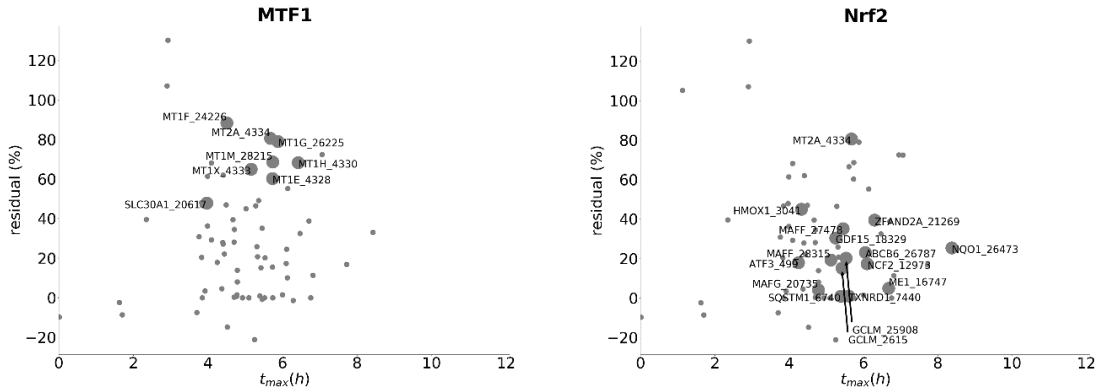


Fig 10: Residual (percentage of the first maximum heights remaining at the end of the analysis period) vs. time of first maximum t_{max} plots highlighting the genes associated with transcription factor MTF1 (left) and Nrf2 (right). To show similarity between these genes, all other of the 115 differentially expressed genes are represented as smaller spheres.

4. Discussion

This study tried to investigate a clustering method that can help identify the relationship between similar temporal gene profiles. Clustering is an unsupervised machine learning algorithm that is used to find meaningful information and patterns from a set of data. It divides the dataset into a small number of clusters and keeps the similar data points in the same group and the most dissimilar data points in the different groups (GeeksForGeeks 2021). The clustering techniques when applied to gene expression data can not only help in identifying the most interesting expression pattern in the overall data but also removes unnecessary noise from the data (Oyelade et al. 2016). We wanted to see if the clustering could group similar profiles with the idea that such similarity would imply a relationship between the genes and that could point to some mechanism related to these genes.

Hierarchical clustering used in the study clustered genes in a predetermined order called hierarchy (R-Bloggers 2021). It is one of the most known methods used for clustering. The agglomerative hierarchical clustering technique used in this study started by considering each data point as a single cluster. Then, at every step, it clustered the most similar clusters until the defined number of clusters is formed (Towards data science 2021). Our results showed that the

hierarchical clustering approach used only the absolute values of the fold changes and cluster-based on height. But it didn't take into account their temporal relationship when clustering the genes.

The results obtained from this clustering showed two types of behavior a) similar temporal profile genes assigned to different clusters and b) different temporal pattern genes assigned to the same cluster. It grouped all the MTs into two different clusters even though they all showed similar expression patterns over time. Whereas SLC30A1 and EGR1 were combined into one cluster even if their temporal profiles were quite different with varying expression levels throughout the full-time period.

The behavior can be explained by the fact that the hierarchical clustering is only concentrating on the absolute values of the changes. It calculates the Euclidean distance between two genes as the distance between the log2 fold changes at each time point. This results in a separation of highly differentially expressed genes from the ones which are only slightly influenced. The division into more clusters is then splitting each of the main groups into smaller ones with more limited ranges of fold change values. For the pair of MT1G and MT1X, the difference in the last time points was enough to separate them into two clusters even if the profile for the first 24h is very similar. This could be fixed by looking at the early and late time points. But, this would then probably lead MT1E to be grouped into another cluster since the height of the first maximum is much smaller for this gene. Also, since MT1E had values at 72h and 168h very similar to MT1X, these two genes were clustered together when the full-time range is considered. Overall, the large variation of the maximum fold changes for the MTs resulted in very high Euclidean distances. The fact that they were with HMOX1 the gene showing such high overexpression leads to the seemingly reasonable cluster. For the genes with the much smaller overexpression levels, SLC30A1 and EGR1, again the fact that they both had small expression levels is the reason why they were in the same cluster. The absolute differences at each time point were on average smaller than for the two groups of MTs despite their different profiles.

To overcome the limitation of hierarchical clustering which doesn't use the time dependency factor among successive time points, we used the STEM clustering method as our second approach. STEM clustering method clusters genes by mapping them to the pre-defined set of model profiles (Ernst and Bar-Joseph 2006). It has been known to cluster the short time-series data.

As compared to the hierarchal clustering, STEM put all the metallothioneins together along with HMOX1 in one cluster. They were all highly overexpressed and sustained extreme expression levels throughout the complete period. But genes like SLC7A11 and PROC which had different temporal patterns were also part of the same cluster. As compared to MTs, they went down to levels not significantly above ground level and partly even became underexpressed in the late time point. These were very different biological behaviors representing a continuous activation in contrast to a very short-lived effect, which will also differ in their importance for the adversity of the test compound. Therefore, we saw that the definition of the model profiles is not able to differentiate profiles based on the significance of

the deviation compared to ground/noise level and, thus, too many genes are grouped to the same standard profile, extremely reducing the information gain from the clustering.

Therefore, both hierarchical clustering and STEM clustering couldn't cluster the related genes. This could be due to the reason that the first one only look at the time points individually and the second one overstated the importance of noise. The third method called the fitting procedure method as described in (Singh et al. 2021) was made so that it considers the most important aspect that the other methods didn't take into account. This aspect was evaluating the significance of the changes in the expression levels of genes by taking into account the changes in the neighboring time points. It described the individual general profile of the genes using the pre-defined mathematical distribution functions which also removed noise to an extent.

The results in the residual vs t_{max} plot showed almost all metal response related genes showing an early activation for all these genes and sustaining the response until the last measurement time point. They clustered around each other with t_{max} around 6h and a residual between 60 and 80 %. They didn't overlap with other differentially expressed genes. Whereas we couldn't derive any distinct temporal pattern for the oxidative stress related genes. They were distributed in a larger area and overlapped with almost all other genes which belonged to other mechanisms and pathways. They showed common maxima appearing between 4h to 8h but then went back to baseline or kept some overexpression or underexpression as seen in the results. Also, the huge number of genes were seen to be clustered closer even if they belonged to different pathways.

Therefore, the fitting procedure helped to concentrate on the most important patterns represented by the genes profiles by considering their time dependency. It also helped in noise-reduction and helped increase the confidence in the observed effects, especially in cases of borderline significance (adjusted p-values around 0.05 or log2 fold changes below 1) and for genes with overall low RCs (Singh et al 2021). Further, profile fitting and optimization of the criteria used for defining that two profiles are similar (time of max and % of max height are starting points but need improvement) removed some basic issues of the clustering methods. But this clustering still could not cluster the temporal profiles belonging to the same pathway. Even though the genes showed a similar profile pattern for the metal response but it was not the same case with oxidative stress. It might be because complex pathways like oxidative stress include several mechanisms which are depending on each other and are activated at different time intervals. Additionally, many pathways share common genes which makes it even more difficult to differentiate pathways based on these temporal profiles. The metal response is grouping only very similar genes with the same function and thus, could also be expected to behave similarly. Therefore, finding a clustering method to group the profiles of related genes is difficult and not possible without the use of pathway libraries as done in the previous study.

5. Conflict of interest statement

The authors declare no competing interests.

6. Acknowledgements

This work was funded by the EU project in3 a Marie Sklodowska-Curie Action - Innovative Training Network under grant no. 721975.

7. References

4.0.0. RDCT (2020) A Language and Environment for Statistical Computing. R Found Stat Comput

Black MB, Budinsky RA, Dombkowski A, et al (2012) Cross-species comparisons of transcriptomic alterations in human and rat primary hepatocytes exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. Toxicol Sci 127:. https://doi.org/10.1093/toxsci/kfs069

Ernst J, Bar-Joseph Z (2006) STEM: A tool for the analysis of short time series gene expression data. BMC Bioinformatics 7:1–11. https://doi.org/10.1186/1471-2105-7-191

Ernst J, Nau GJ, Bar-Joseph Z (2005) Clustering short time series gene expression data. Bioinformatics 21:. https://doi.org/10.1093/bioinformatics/bti1022

Oyelade J, Isewon I, Oladipupo F, et al (2016) Clustering algorithms: Their application to gene expression data. Bioinform Biol Insights 10:. https://doi.org/10.4137/BBI.S38316

Singh P, Chandrasekaran V, Hardy B, et al (2021) Temporal transcriptomic alterations of cadmium exposed human iPSC-derived renal proximal tubule-like cells. Toxicol Vitr 76:. https://doi.org/10.1016/j.tiv.2021.105229

Wink S, Hiemstra SW, Huppelschoten S, et al (2018) Dynamic imaging of adaptive stress response pathway activation for prediction of drug induced liver injury. Arch Toxicol 92:. https://doi.org/10.1007/s00204-018-2178-z

Zolfaghari F, Khosravi H, Shahriyari A, et al (2019) Hierarchical cluster analysis to identify the homogeneous desertification management units. PLoS One 14:1–21. https://doi.org/10.1371/journal.pone.0226355

3. Temporal alterations in the stress responses activated in human iPSC-derived renal proximal tubule-like cells due to exposure to 6 organic compounds

Pranika Singh[1,2], Barry Hardy[1], Anja Wilmes[3], Paul Jennings[3] and Thomas E. Exner[4]

[1]Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, 4057 Basel, Switzerland

[2]Division of Molecular and Systems Toxicology, Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

[3]Division of Molecular and Computational Toxicology, Chemistry and Pharmaceutical Sciences, AIMMS, Vrije Universiteit Amsterdam, The Netherlands.

[4]Seven Past Nine d.o.o., Hribljane 10, 1380 Cerknica, Slovenia

Thomas Exner, Seven Past Nine, thomas.exner@sevenpastnine.com

Abstract

Stress response pathways are activated in response to perturbations or disturbances in the cells. These responses prevent the cells from being damaged and help restore or maintain homeostasis. Cellular homeostasis is very significant for the proper functioning of cellular mechanisms. In this case study, we have studied the temporal alterations in four stress responses relevant to toxicology namely oxidative stress-mediated by Nrf2, metal response mediated by MTF-1, p53 signaling pathway, and unfolded protein response (UPR) in human-derived iPSC cells. These iPSC cells were exposed to a set of a wide variety of chemicals namely GW788388 (1 μM), arsenic (10 μM), amiodarone (50 μM), rotenone (5 nM), tunicamycin (300 nM), and doxorubicin (15 μM) for 1, 2, 6, 8, 12, 16, 20 and 24 hours. All four stress response pathways showed different temporal patterns and severity in response to exposure to different compounds. GW788388 showed no activation of any stress response pathways until 24h. The oxidative stress response pathway was seen to be activated in all other compounds and was strongly activated in arsenic and remained sustained until 24h. The metal response was seen to be strongly activated in arsenic and was sustained until 24h like the oxidative stress pathway. Arsenic also showed the early activation of the p53 signaling pathway and UPR pathway. p53 pathway was seen to be activated in all compounds except for tunicamycin with doxorubicin showing the strongest expression. UPR pathway was seen to be activated in all compounds except for doxorubicin with tunicamycin showing the strongest expression. In comparison to other compounds, rotenone showed late activation of these pathways from around 16h. The data shows the utilization of iPSC cell models in elucidating the stress responses and in deriving their severity due to exposure to different compounds.

Keywords: human-induced pluripotent stem cells, stress response pathways, temporal resolution, proximal tubule cells, chemical substances

Highlights

- A human iPSC-derived PTL model can be used for toxicity testing
- Oxidative stress mediated by Nrf2, metal response mediated by MTF-1, p53 signaling pathway, and Unfolded Protein Response pathway show different severity and temporal patterns in different compounds.
- Metal compounds show the strongest activation of oxidative stress and metal response pathways.
- Doxorubicin showed the strongest activation of the p53 pathway and tunicamycin showed the strongest activation of the UPR pathway.

Abbreviations: EMT, Epithelial-to-mesenchymal transition; ER, Endoplasmic reticulum; LFC, Log2 fold change; MT, Metallothionine; MTF1, Metal regulatory transcription factor 1; Nrf2, Nuclear factor erythroid-2-related factor 2; ROS, Reactive oxygen species; TF, Transcription Factor; UPR, Unfolded Protein Response.

1. Introduction

Multicellular organisms react to changes in their environment first at the cellular level. Cells are considered the fundamental units of life and are part of every biological organization in the body (Cooper and Hausman 2007). Cells have different methods to react to different perturbations happening around them (Wigner et al. 2021). For example, they react to metabolic signaling via endogenous growth factors which are necessary for the development of an organism. Similarly, the cells also respond to different stressors like xenobiotics, chemical substances, etc. via adaptive stress response pathways (Simmons et al. 2009). The role of these pathways is to help the cells to repair the damages and maintain or reestablish the homeostasis (Jennings 2013). The adaptive stress response pathways are therefore defined as the signal transduction pathways which are activated due to cellular stress and lead to the activation of cytoprotective genes at the transcriptional level. These genes are the first ones to be activated in response to the cellular stress (Kültz 2005). Under the stress, the enzymatic transducers such as kinases and hydroxylases are activated, and they covalently modify both the cell sensors and transcription factors (TF). As a result, the sensor's activity is blocked which paves the way to the activation and nuclear translocation of the TF. This translocation then leads to the upregulation of many target genes. This helps the cells to cope with the stress (Simmons et al. 2009; Vihervaara et al. 2018).

Many stress response pathways like DNA damage response pathway, oxidative stress pathway, hypoxia, osmotic stress, heat shock response pathway, Endoplasmic reticulum (ER) stress including unfolded protein response (UPR), inflammation, metal response pathway, etc are activated in response to the different types of chemicals (Jennings 2013). In this study, we will focus on only four of these important stress response pathways: the p53 signaling pathway, oxidative stress response, metal response, and UPR.

p53 signaling pathway is one of the major pathways studied in carcinogenicity, aging, and toxicity and is activated in response to the DNA damaging agents. It safeguards the DNA integrity and removal of cells that are beyond repair. P53 family along with its family members is the major transcription factor that mediates this response. On activation in response to stress, it leads to the transcription of genes which leads to cycle arrest or apoptosis or senescence(Horn and Vousden 2007; Jennings et al. 2013). Nuclear factor erythroid-2-related factor 2 (Nrf2) mediated oxidative stress response (Baird and Dinkova-Kostova 2011) is of high importance in toxicology studies. Oxidative stress responds to a wide variety of toxicants and cellular reactions which involve the production of reactive oxygen species (ROS) (Simmons et al. 2009). The activation of Nrf2 is mediated via cytosolic KEAP1 which leads to its nuclear translocation and ultimately the transcript of genes related to glutathione synthesis and reduction process (Itoh et al. 1997). The metal response is mediated by metal regulatory transcription factor 1 (MTF1) and is activated in response to certain metals (Lichtlen and Schaffner 2001). The oxidative and protein malfunction caused due to free metal ions can harm the cell and the MTF1 is, therefore, activated to protect these cells (Gunther et al. 2012). The

activated MTF1 leads to activation of the metallothionine (MT) genes which remove the excess metal ions (Saydam et al. 2002). It has also been seen that in absence of heavy metals too, MTs can get activated by MTF-1 against oxidative stress (Andrews 2000). UPR is activated against the disturbances caused in the ER due to the deposition of denatured or misfolded proteins. It is activated to restore homeostasis and create a balance between the disturbed ER and stress of the protein folding (Foufelle and Fromenty 2016; Guha et al. 2017). This is also collectively known as ER stress which can be caused by factors like glucose lack, amino acid scarcity, ROS production, etc. (Lin and LaVail 2010). Three transcription regulators ATF4, ATF6, and XBP1 are activated which leads to the transcription of genes responsible for protecting ER against stress (Jennings 2013).

In this case study, we have used a diverse set of chemicals namely, GW788388 (TGF- ß receptor inhibitor), arsenic (metal toxin), amiodarone (antiarrhythmic drug), rotenone (an organic pesticide), tunicamycin (antibiotic drug) and doxorubicin (anthracycline drug). These chemicals are tested on human-induced pluripotent stem cells (iPSC) derived proximal tubule-like cells (PTL) which were differentiated as described in (Chandrasekaran et al. 2021). This study was aimed at further testing the iPSC cells in chemical safety assessment using other types of chemicals as done in the previous case study with cadmium. Additionally, it was also used to see if different stress response pathways show different temporal patterns and severity of adversity in different compounds or not.

2. Material and Methods

2.1 iPSC cell Culture

iPSCs were maintained in mTeSR1 medium (StemCell Technologies 05850) on 6-well plates coated with Geltrex (Life Technologies A1413302) in a humified incubator at 36.7°C containing 5% CO_2 and sub-cultured with EDTA (0.02% Versene, Lonza BE17-711E) twice per week. In this study, we employed an iPSC-line (SBAD2 HMOX1-eGFP) that had previously been tagged with a GFP-reporter for the HMOX1 expression (Snijders et al. 2021).

PTL differentiation

SBAD2 HMOX1-eGFP were differentiated into renal proximal tubular-like cells as previously described (Chandrasekaran et al. 2021). Briefly, cells were detached with accutase and centrifuged at 300 g for 5 min. After resuspension in differentiation medium (1:1 mixture of DMEM/F12 (Thermo Fisher Scientific), 2 mM Glutamax, and ITS (5 µg/mL, 5 µg/mL, 5 ng/mL, Sigma-Aldrich), supplemented with 3 µM CHIR99021 (Sigma-Aldrich), 1 µM TTNPB (Sigma-Aldrich) and 10 µM Rock inhibitor (abcam) cells were seeded on Geltrex coated 96-well plates at 35,000 cells per cm^2. After 42h, the medium was replaced with a differentiation medium supplemented with 1 µM TTNPB. After an additional 30h, the medium was replaced with a proximal tubular medium (1:1 mixture of DMEM/F12 (Thermo Fisher Scientific), 2 mM Glutamax, ITS (5 µg/mL, 5 µg/mL, 5 ng/mL, Sigma-Aldrich), 10 ng/ml EGF (Sigma-Aldrich) and 36 ng/ml hydrocortisone (Sigma-Aldrich) supplemented with 10 ng/ml FGF9

(Thermo Fisher Scientific). FGF9 was removed after 60h and cells were fed with the proximal tubular medium. Cells were then fed every 2 to 3 days with the proximal tubular medium until day 14.

2.2 Compound exposure and sample collection

On day 14 of differentiation, iPSC-derived PTL cells were treated with either 0.01% DMSO (controls) or with the following diverse set of chosen compounds: arsenic (10 µM), amiodarone (50 µM), doxorubicin (15 µM), GW788388 (1 µM), rotenone (5 nM) and tunica--mycin (300nM) over a temporal course (1, 2, 4, 6, 8, 12, 16, 20 and 24 h). Three biological replicates were used. During compound exposure, cells were imaged for GFP expression (excitation at 488nm and emission at 510nm) in a High Content Imager (HCI) Operetta (Perkin Elmer) and the intensity of the GFP signal was determined with the software Harmony 4.8 (Perkin Elmer). In addition, cells were lysed in TempO-Seq buffer and collected for TempO-analysis as described before in Singh et al (Singh et al. 2021).

2.3 TempO-Seq experiments

The samples obtained for different compounds for all the time points (1, 2, 4, 6, 8, 12, 16, 20, and 24 h) were collected and sent to execute TempO-Seq experiments using EU-ToxRisk v2.1 panel (3565 probes) (Mav et al. 2018; Limonciel et al. 2018) with standard attenuators at the BioClavis technologies Ltd. Glasgow. These samples were measured for the derivation of the raw read counts and were also checked for standard quality control.

2.4 Data accessibility and sample/gene selection

The raw data and metadata corresponding to the compounds arsenic, amiodarone, doxorubicin, GW788388, rotenone, and tunicamycin were accessed by making use of the EdelweissData™ management system (SaferWorldbyDesign2021) using the python script as described in Singh et al (Singh et al. 2021). The samples for both treatment and corresponding controls were selected for each of these compounds with additional filters for the iPSC-derived PTL cells and the time points 1h,2h,4h,6h,8h,12h,16h, 20h, and 24h. This resulted in 54 samples for each compound with three biological replicates each for treatment and control samples. The low read count probes were removed separately using the row sum threshold as 100 per probe across all samples per compound (including both treatment and control) before performing differential expression analysis (Love et al. 2014).

2.5 Differential expression analysis

The analysis was performed as described in and available in form of an R script from (Singh et al. 2021). The data obtained after filtering the low read count probes were visualized, after r-log transformation (HBC Training 2019, (Love et al. 2014)), using PCA and hierarchical clustering to see the variance between and among the different treatment and control groups (Love et al. 2014). The data were then normalized using the standard median-ratio method and on this normalized data, differential expression analysis was performed for the

treated group vs the corresponding control group for each compound (Love et al. 2014). The statistical values base mean log2 Fold Change (LFC), adjusted p-value, p-value, and log fold change error (lfcSE) were generated for each probe for each time point per compound. Additional filters were set to identify the most significantly expressed genes: the adjusted p-value was set to 0.01 and |LFC| > 2.

2.6 Temporal profiles of genes

Since the different time points are not continuous measures of the same system but from independent exposure experiments, the fitting method as described in (Singh et al. 2021) was used in the visualizations to reduce noise and model continuous temporal profile for normalized read counts of the treated and control groups as well as the resulting fold changes.

2.7 Pathway enrichment analysis

The full list of pathways and their associated genes were downloaded from the ConsensusPathDB V35 (http://cpdb.molgen.mpg.de) (Kamburov et al. 2009). This list was then modified to only include the genes represented by one or multiple probes of the used TempO-Seq panel.

Z-scores were calculated for the pathways to see their overrepresentation over time since the scores from ConsensusPathDB can't be used as they are based on the full list of genes which is not covered by the TempO-Seq panel used in our experiments. A python script was used to automate the calculation of the z-score per time point per compound for all pathways seen in the ConsensusPathDB list. Z-scores were then calculated using the formula as described in (Kutmon et al. 2015; Wellens et al. 2021) for each compound. Time profiles of the z-scores and the number of DEGs are generated automatically by the script for all pathways and were selected based on the thresholds for the z-score > 1.96 reached at any time point and the number of corresponding differentially expressed genes >= 2.

3. Results

3.1 Differential expression analysis and pathway enrichment analysis

Filtering out low expressed probes per each chemical based on the row sum of read counts (which was set to 100 across all samples per probe as the threshold as described in Section 2.4) left 2789 probes for arsenic, 2774 probes for amiodarone, 2853 probes for doxorubicin, 2794 probes for GW788388, 2787 probes for rotenone and 2775 probes for tunicamycin out of the 3656 probes available in the TempO-Seq panel. These filtered raw data were first analyzed using PCA. The PCA plots (see **Supporting Information 1**) showed good separation of individual groups for all the compounds except for GW788388. Controls and different exposure times showed separated clusters with, in most cases, early time points clustering closer to the control samples than later ones showing the temporal progression of the changes. Differentially expressed genes (DEGs) were selected based on thresholds for adjusted p-value < 0.01 and |LFC| > 2. Amiodarone, arsenic, doxorubicin, and rotenone showed the highest

number of DEGs (represented by 81, 247, 399, and 93 probes respectively) at the 24h time point. Tunicamycin showed the highest DEGs (58 probes) at 20h. GW788388 showed very few DEGs compared to the other compounds with 7 probes being the highest count (at both 4h and 20h) and a sum of 16 probes differentially expressed at any time point. As a comparison, cadmium chloride discussed in detail in the previous paper (Singh et al. 2021) showed the maximum of DEGs at 6h (85 probes) and 126 probes in total over all time points.

Pathway enrichment analysis was performed using ConsensusPathDB (Kamburov et al. 2009) to avoid bias from the pathway coverage of a specific pathway library. The pathways were ranked based on the z-score values calculated using the python script as described in Section 2.7 taking into account only the target gene set in the TempO-Seq panel. To identify significantly overrepresented pathways, thresholds for the maximum z-score > 1.96, i.e., the highest z-score seen across all the time points, and the associated number of differentially expressed genes >= 2 were applied. The top 5 pathways are listed in **Table 1.** The full lists of pathways that cross the threshold of z-score and the minimum number of DEGs for each compound are provided in the supplementary data (available on request). Since this study is focusing on similarities and differences seen during exposure to the set of chemicals in important toxicity-related biological mechanisms, four pathways representing these biological mechanisms are also listed in **Table 1**. The differentially expressed gene list from the cadmium chloride exposure presented in the previous study (Singh et al. 2021) was also reanalyzed with ConsensusPathDB to be able to base comparisons on the same set of knowledgebase. Even if the selected mechanisms are clearly relevant for the adverse effects of the compounds and sometimes are seen in the highest ranks, many other pathways showed high significance and are ranked much higher. For example, the first of the selected pathways Nrf2-mediated oxidative stress (named as the Nrf2 pathway in the pathway library) is listed as rank 242 in the case of amiodarone exposure. The reason for this is the high number of DEGs resulting in a high number of selected pathways. Much important information can be extracted from this list of pathways and the corresponding genes, which can be guiding a full risk assessment of the compounds. However, doing this full analysis including an evaluation of the relevance of the suggested pathways for toxicology is out of the scope of this study concentrating on known stress response pathways.

Arsenic		
Rank	**Pathway**	**Max. Z-score**
1	heme degradation (PWY-5874)	32,9
2	**Metallothioneins bind metals (R-HSA-5661231)**	30,5
3	Response to metal ions (R-HSA-5660526)	27,8
4	Zinc homeostasis (WP3529)	25,7
5	Mineral absorption - Homo sapiens (human) (path: hsa04978)	21,9
43	**Photodynamic therapy-induced unfolded protein response (WP3613)**	7,7
48	**Direct p53 effectors (p53downstreampathway)**	7,2
58	**NRF2 pathway (WP2884)**	6,6
Cadmium		
1	**Metallothioneins bind metals (R-HSA-5661231)**	40,3
2	Response to metal ions (R-HSA-5660526)	36,8
3	Zinc homeostasis (WP3529)	34

4	Mineral absorption - Homo sapiens (human) (path: hsa04978)	25,4
5	Copper homeostasis (WP3286)	18,2
11	**Photodynamic therapy-induced unfolded protein response (WP3613)**	10,3
16	**NRF2 pathway (WP2884)**	9
71	**Direct p53 effectors (p53downstreampathway)**	5,2
Amiodarone		
1	**Metallothioneins bind metals (R-HSA-5661231)**	24
2	Celecoxib Pathway, Pharmacokinetics (PA165816736)	22,1
3	Response to metal ions (R-HSA-5660526)	21,9
4	Celecoxib Metabolism Pathway (SMP00644)	21,9
5	Zinc homeostasis (WP3529)	20,2
65	**Photodynamic therapy-induced unfolded protein response (WP3613)**	8,4
242	**NRF2 pathway (WP2884)**	3,6
405	**Direct p53 effectors (p53downstreampathway)**	2,2
Doxorubicin		
1	**Direct p53 effectors (p53downstreampathway)**	10,8
2	Validated transcriptional targets of TAp63 isoforms (tap63pathway)	10,7
3	TP53 Regulates Transcription of Cell Cycle Genes (WP3804)	10,1
4	Interleukin-18 signaling (R-HSA-9012546)	9,9
5	p73 transcription factor network (p73pathway)	8,5
346	**Photodynamic therapy-induced unfolded protein response (WP3613)**	2,6
411	**Metallothioneins bind metals (R-HSA-5661231)**	2,3
NA	**NRF2 pathway (WP2884)**	NA
Tunicamycin		
1	ATF4 activates genes (WP2753)	19,4
2	ATF6 (ATF6-alpha) activates chaperone genes (WP2655)	18
3	**Photodynamic therapy-induced unfolded protein response (WP3613)**	16,5
4	Transcriptional cascade regulating adipogenesis (WP4211)	14,2
5	asparagine biosynthesis (ASPARAGINE-BIOSYNTHESIS)	12,1
NA	**NRF2 pathway (WP2884)**	NA
NA	**Direct p53 effectors (p53downstreampathway)**	NA
NA	**Metallothioneins bind metals (R-HSA-5661231)**	NA
Rotenone		
1	Vitamin C metabolism (Vitamin C metabolism)	17,1
2	Interleukin-18 signaling (R-HSA-9012546)	16,4
3	Cellular hexose transport (R-HSA-189200)	14,8
4	Vitamin C (ascorbate) metabolism (R-HSA-196836)	14
5	Fructose and mannose metabolism (Fructose and mannose metabolism)	13,6
35	**Photodynamic therapy-induced unfolded protein response (WP3613)**	6,8
135	**Direct p53 effectors (p53downstreampathway)**	4,5
148	**Metallothioneins bind metals (R-HSA-5661231)**	4,3
239	**NRF2 pathway (WP2884)**	3,5
GW788388		
1	Nevirapine Metabolism Pathway (SMP00642)	26,3
2	Anti-diabetic Drug Nateglinide Pathway, Pharmacokinetics (PA154423659)	24,9
3	Celecoxib Metabolism Pathway (SMP00644)	22,8
4	Losartan Pathway, Pharmacokinetics (PA164713428)	21,5
5	Fluoxetine Pathway, Pharmacokinetics (PA161749012)	19,2
357	**Direct p53 effectors (p53downstreampathway)**	2,3
360	**NRF2 pathway (WP2884)**	2,1
2155	**Metallothioneins bind metals (R-HSA-5661231)**	-0,1
3090	**Photodynamic therapy-induced unfolded protein response (WP3613)**	-0,1

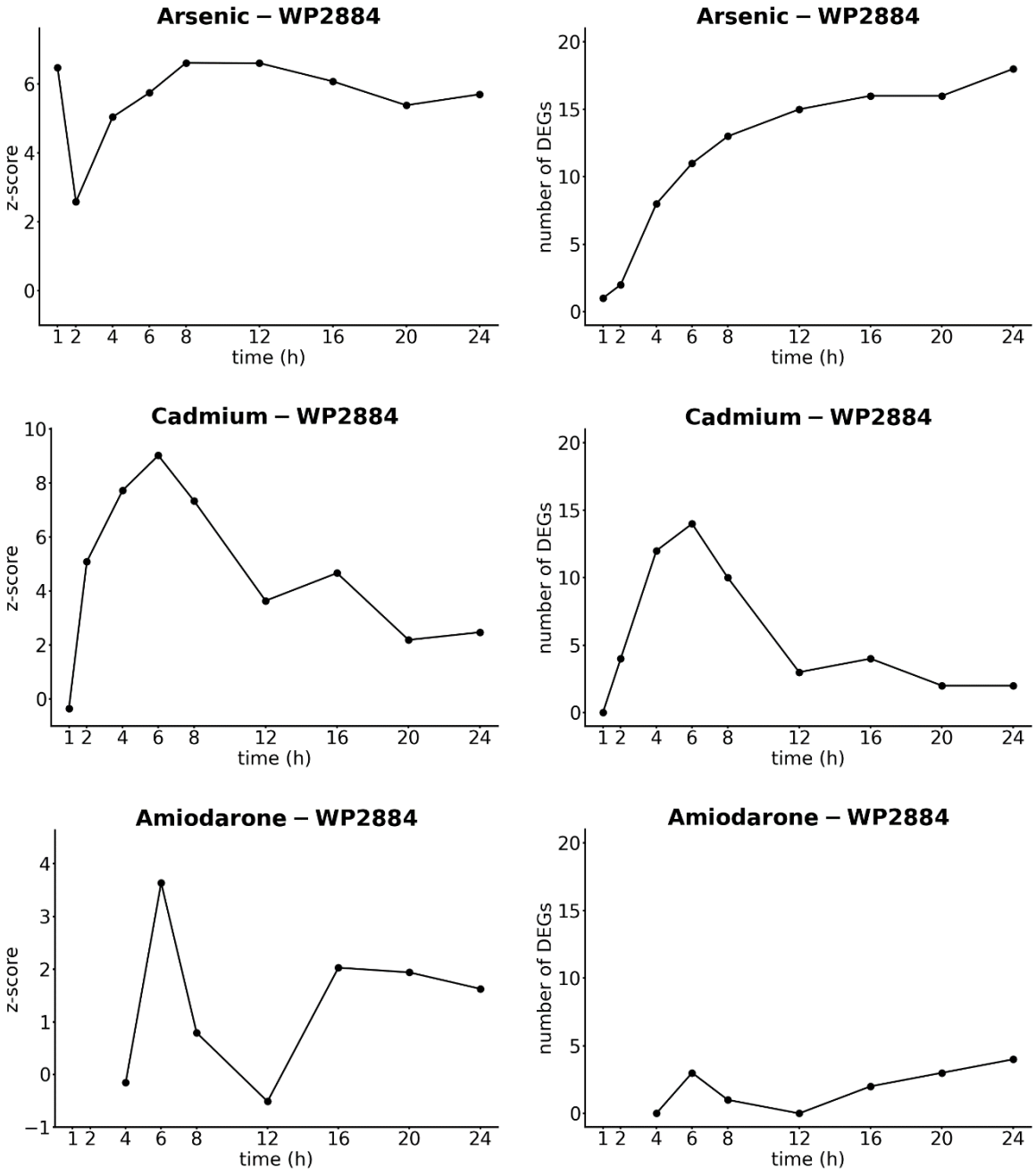
Table 1: The top 5 ranked pathways based on the maximum z-score for arsenic, cadmium, amiodarone, doxorubicin, rotenone, tunicamycin and GW788388 including the list of selected pathways (highlighted in bold).

The four stress response mechanisms, 1) oxidative stress mediated by the Nrf2, 2) metal response mediated by MTF1, 3) UPR, and 4) the tumor suppressor p53 signaling pathway are represented by many entries in CONSENSUSPathDB collected from different pathway libraries**.** For example, in the case of doxorubicin, 10 of the 15 first ranks of the z-score sorted pathway list are related to p53, genotoxicity, or cell death/apoptosis with even more related pathways in lower ranks. This demonstrates the general importance of the pathway being covered in all underlying libraries, e.g., "TP53 Regulates Transcription of Cell Cycle Genes" (WP3804) from WikiPathways (Martens et al. 2018; Slenter et al. 2018) on rank 3 and "p53 signaling pathway - Homo sapiens (human)" (path: hsa04115) from KEGG (Kanehisa et al. 2016) on rank 6. These different representations of the same biological response share a core of common genes. But many additional genes are added in some pathways representing the focus of the research group generating these entries, how general or specific a pathway was meant to be, the evidence used to define gene-pathway relationships, the knowledge at the point of time, the pathway was created or updated, and how contradicting evidence is treated. Therefore, conclusions about similarities between compounds extracted from the results will depend on the chosen pathway to represent the response. To be able to base our further results on only one pathway per mechanism, we chose the one showing up consistently on high ranks of multiple components: 1) "Nrf2 pathway" (WikiPathway ID: WP2884, 78 total genes), 2) "MTF1-mediated metal response" (Reactome ID: R-HAS-5661231, 10 total genes), 3) "Photodynamic therapy-induced unfolded protein response" (WikiPathway ID: WP3613, 18 total genes) and 4) Direct "p53 effectors" (Pathway Interaction Database: p53downstreampathway, 80 total genes). We already showed in the previous publication (Singh et al. 2021) with the drastic example of "SPINK1 General Cancer Pathway" being a good representation of metal response that the relevance of pathways should not be judged by their name only. A milder case is seen here with the UPR pathway being quite specifically defined but being a better representation of the effects caused by chemical exposure than the one used in the previous publication.

3.2 Oxidative stress response represented by the Nrf2 pathway

The Nrf2 pathway showed the maximum number of genes differentially expressed for the two metal compounds arsenic and cadmium. For both, the number of DEGs started increasing already at 1h. A sustained response until 24h was seen for arsenic exposure whereas the number of DEGs started decreasing from 8h in the case of cadmium exposure (see **Fig 2**). Please note that since the z-score depends on both the DEGs part of the pathway and the total number of DEGs at the specific time point, we will present both values in the following. The highest number of DEGs against arsenic exposure were seen at 24h (18 out of 231 DEGs, z-score 5.70), and for cadmium exposure at 6h (14 out of 81 DEGs, highest z-score 9.02). But the highest z-score for arsenic 6.61344 was seen at 8h resulting from 13 out of 112 DEGs. Amiodarone, rotenone, and doxorubicin exposure showed no or few DEGs in the early time points. The numbers were then steadily increasing in the later time points finally reaching 4, 5, and 8 DEGs

at 24h, respectively (see also **Fig 2**). Since the total number of DEGs also increased with time, the maximum z-scores were found at earlier timepoint (amiodarone: 3.6382 at 6h, 3 out of 21 DEGS; rotenone: 3.52245 at 12h, 3 out of 22 DEGs; tunicamycin: 1.48651 at 6h, 1 out of 11 DEGs; doxorubicin: 0.918559 at 6h, 3 out of 77 DEGs). Tunicamycin and GW788388 exposure showed a maximum of two (HMOX1_3041 and SLC7A11_14100) and one DEG (SERPINA1_6246), respectively associated with the Nrf2 pathway.



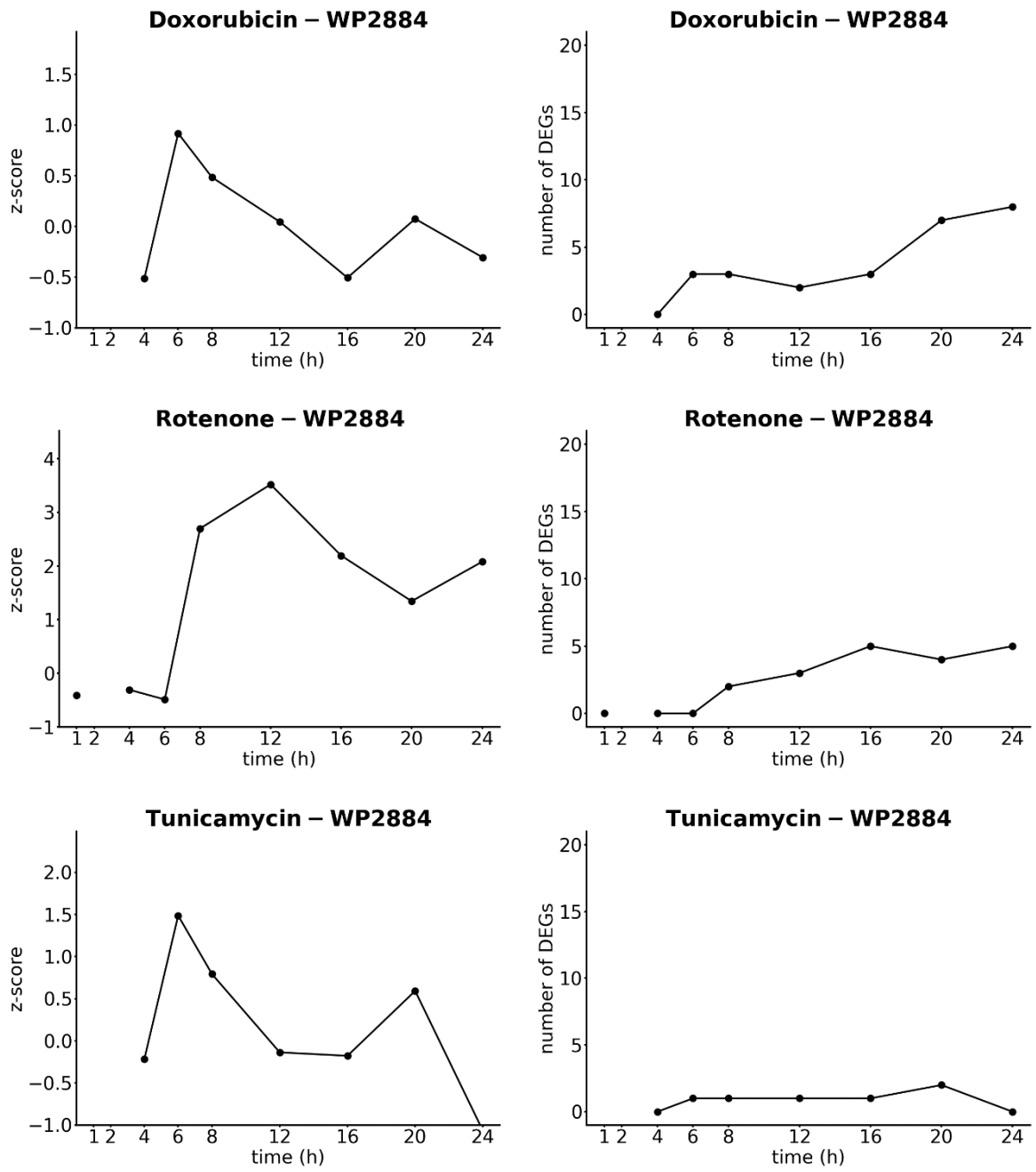


Fig 2 Overrepresentation profiles for Nrf2 pathway based on ConsensusPathDB for arsenic, cadmium, amiodarone, doxorubicin, rotenone, and tunicamycin: z-scores are given in the left column, and the corresponding number of differentially expressed genes in the right column.

Looking at the gene level, MAFF, HMOX1, HSP90AA1, MAFG, NRG1, NOQ1, DNAJB1, EGR1, HSPA1A, SLC2A1, SLC7A11, and SQSTM1 were the most commonly occurring genes overexpressed in response to at least 3 compounds. MAFF was differentially expressed in response to all compounds except tunicamycin and GW788388. Both probes associated with this gene, MAFF has the strongest overrepresentation in arsenic (highest LFC of 3.2 and 3.3 at 6h, respectively) and cadmium (highest LFC of 4.3 and 2.7 at 6h, respectively). Most of the other upregulated genes HSPA1A, DNAJB1, NQO1, SLC7A11, SQSTM1, HMOX1,

HSP90AA1, and MAFG seen in three or more compounds also had the highest overexpression levels for arsenic and cadmium (see **Table 2**).

An exception is EGR1_2068 showing the highest expression value of 2.4 at 8h in amiodarone. The NRG1 results need to be considered with care since the two available probes showed very different activity. NRG1_27598 showed the highest overexpression in doxorubicin exposure with LFC of 5.06 at 24h. But the other probe NRG1_27600 showed no overexpression at all and the temporal profile looks noisy as compared to the other probe (see **Fig 3**). Finally, SLC2A1_14159 showed downregulation for the Nrf2 pathway in all other compounds except for tunicamycin and cadmium. It showed the highest downregulation of -3.6 at 24h for doxorubicin exposure. The changes in the controls were almost neglectable compared to the overexpression levels seen in the metal compounds but made a considerable difference in the underexpressed genes.

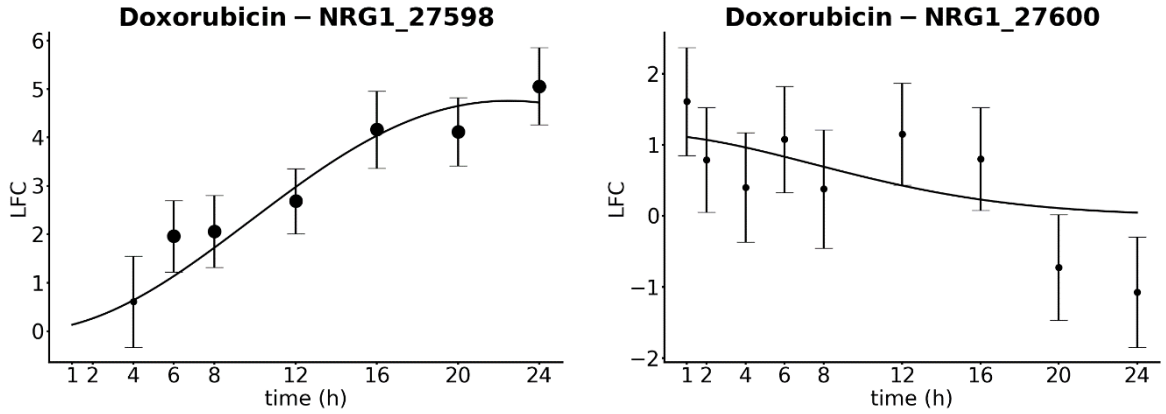


Fig 3 Temporal profiles of LFCs for the two probes of NRG1 included in the TempO-Seq panel after exposure to doxorubicin. Values of low significance (adjusted p-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard error for the fold changes.

Probes	**Log Fold Changes (LFCs)**	
	Arsenic	**Cadmium**
HSPA1A_3135	4.4 (4h)	5.4 (4h)
DNAJB1_1898	2.0 (4h)	3.4 (4h)
NQO1_26473	3.8 (16h)	2.3 (8h)
NQO1_4668	3.9 (16h)	2.0 (8h)
SLC7A11_14100	5.0 (6h)	3.5 (6h)
SQSTM1_6740	3.0 (20h)	2.3 (6h)
HMOX1_3041	10.6 (8h)	10.4 (8h)
HSP90AA1_3127	2.7 (8h)	3.0 (8h)
MAFG_20735	3.3 (6h)	2.8 (4h)

Table 2: The highest expression values (LFCs) seen for probes associated to Nrf2 pathway in compounds arsenic and cadmium.

The temporal behavior of the number of DEGs can also be seen in many individual probes. As an example, we will use the arsenic-cadmium comparison again. While the response of all the associated genes was sustained or only slightly decreases until 24h in arsenic, the profiles

of genes in cadmium are going down significantly after the first maximum and in some cases even to baseline (see **Fig 4**) For example, HMOX1 in the case of cadmium exposure, the normalized read counts drop from 20,000 to less than 10 % of this maximum (890 counts) by 24h with read counts for controls remaining constant around 20-30.

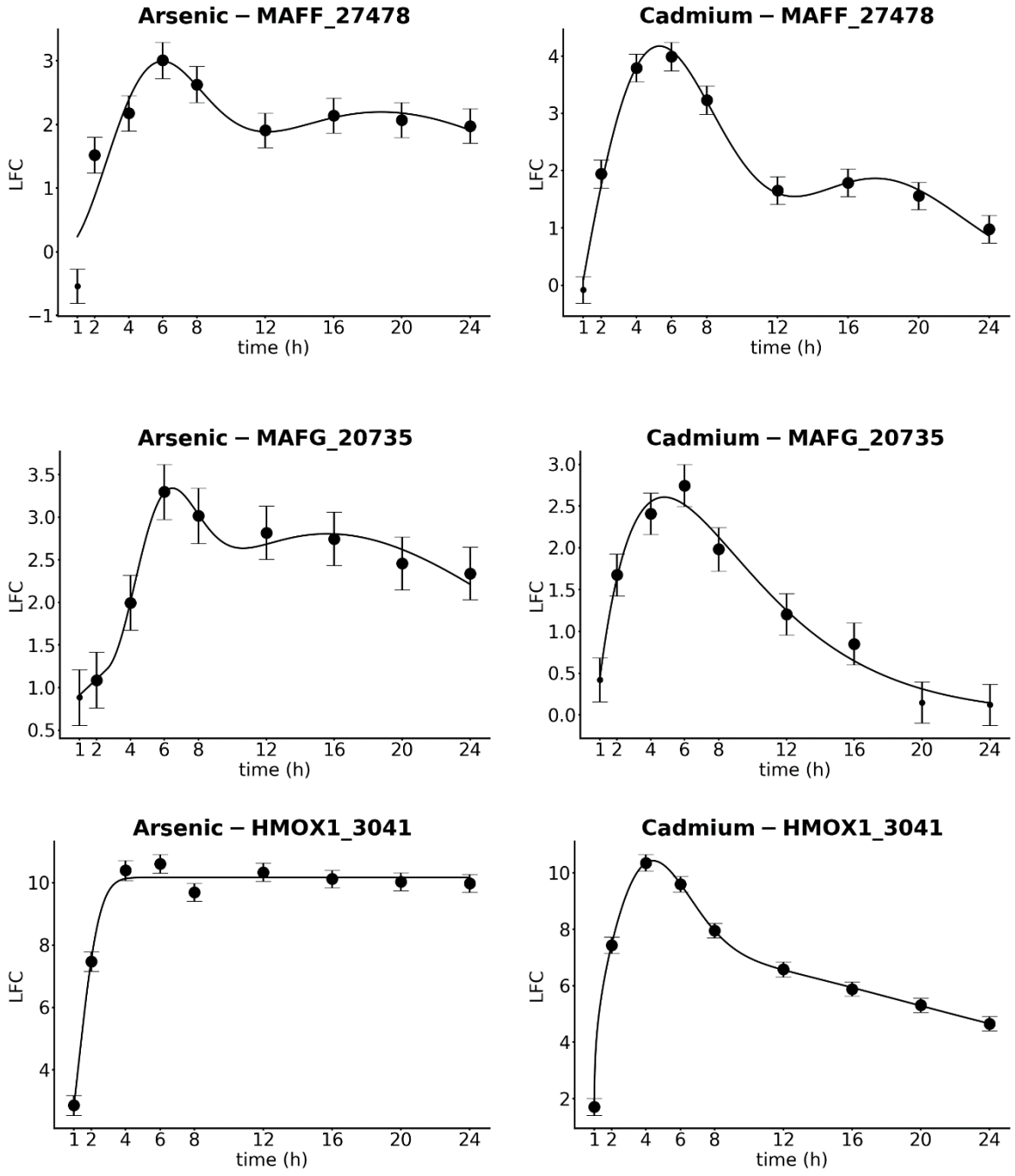
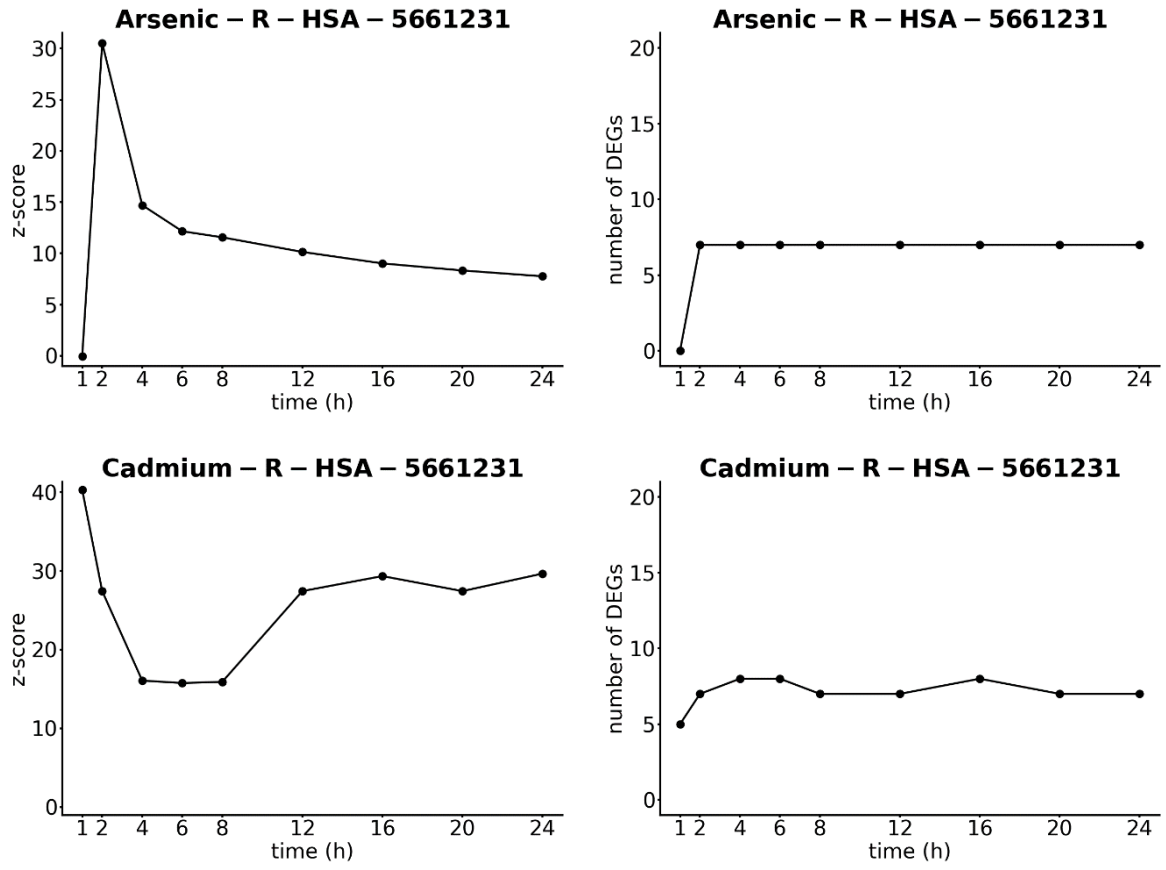


Fig 4 Temporal Profiles of Nrf2 pathway related genes: LFCs of the genes expression pattern over time for arsenic and cadmium are given in the left and right columns, respectively. Values of low significance (adjusted p-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard error for the fold changes.

3.3 Metal response

Metal response and specifically the "MTF1-mediated metal response" pathway from Reactome (ID: R-HAS-5661231) was identified as the most important pathway according to z-scores for arsenic, cadmium, and amiodarone. As for oxidative stress, it showed the maximum number of genes differentially expressed for the metal compounds arsenic and cadmium (see **Fig 5**). For both, the number of DEGs started increasing already at 1h and maintained constant high values from 2h onwards (arsenic: 7 DEGs at 2h to 24 h, highest z-score 30.5283 at 2h; cadmium: up to 8 DEGs from 4h to 24h, highest z-score 29.3437 at 16h). The profiles of individual genes showed, as for the Nrf2 pathway, many similarities with the number of DEGs. In arsenic, the fold changes directly jumped to very high values and then stayed constant while they slightly decreased after an early first maximum following cadmium exposure (see **Fig 6**).



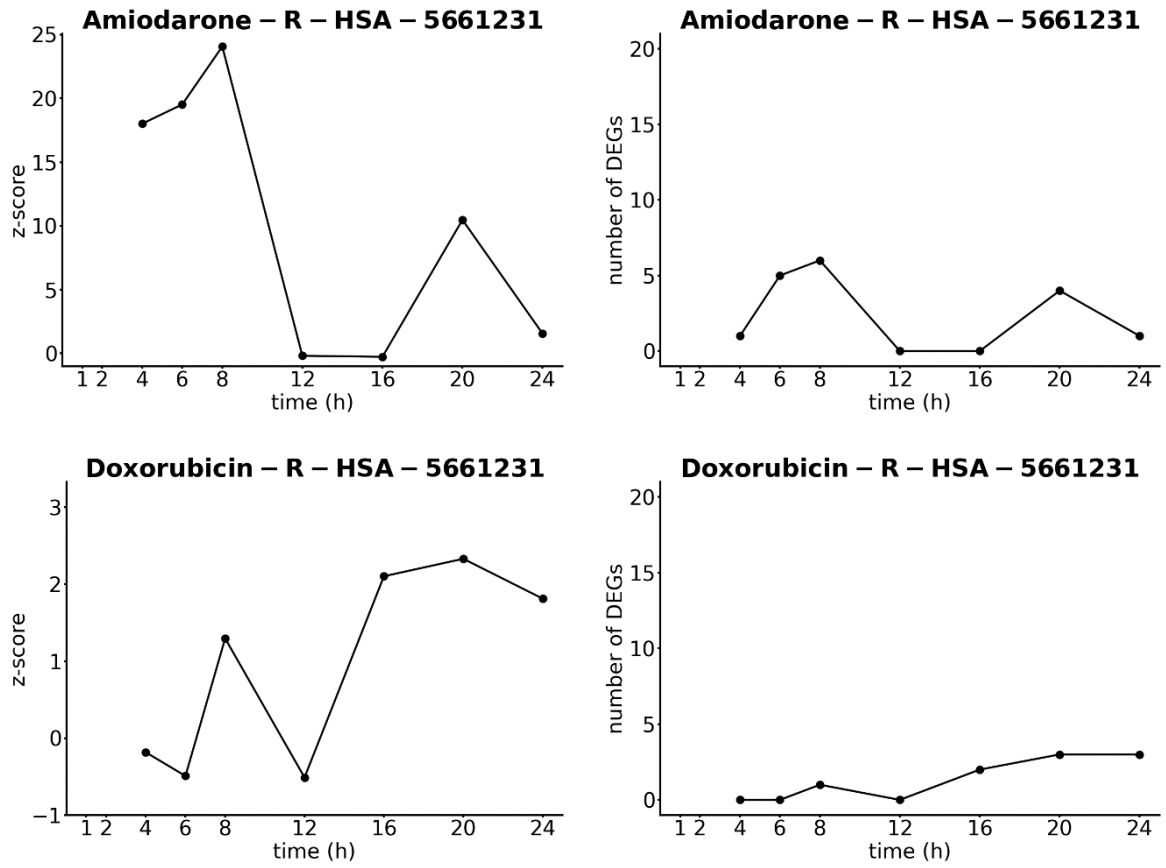
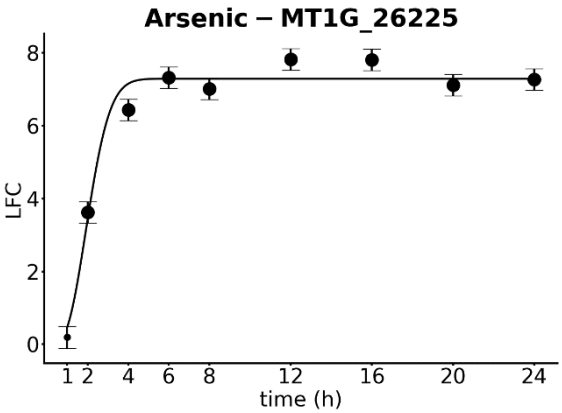
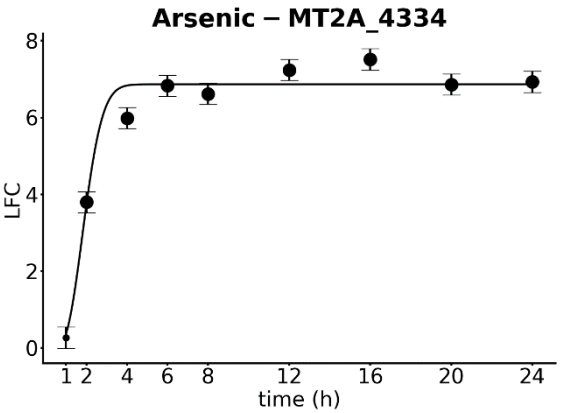


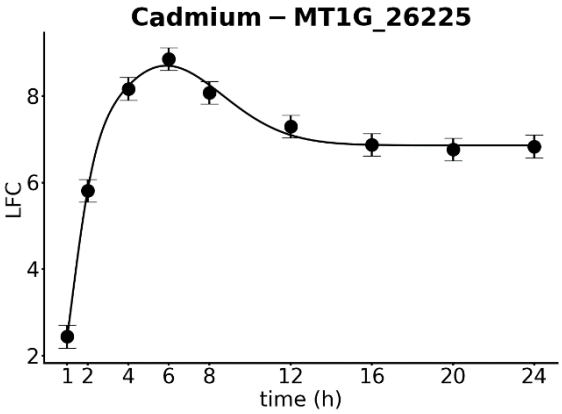
Fig 5 Overrepresentation profiles for MTF1 pathway based on ConsensusPathDB: z-scores are given in the left column and the corresponding number of differentially expressed genes in the right column.

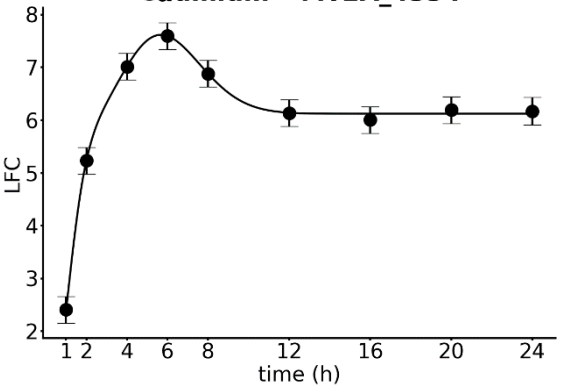
Amiodarone and doxorubicin showed an increase in the number of DEGs at slightly and much later hours compared to the metal compounds, respectively (see **Fig 5**). This difference in the temporal behavior also becomes obvious when looking at the time point when the highest z-score is reached with 24.0685 at 8h due to 6 out of 20 DEGs and 2.33096 at 20h due to 3 out of 292 DEGs for the amiodarone and doxorubicin, respectively. The profile of amiodarone with the two maxima was caused by the fact that the LFCs of the genes were very close to 2 and fell all below this cut-off at 12 and 16h (see **Fig 6**). However, the profiles especially of amiodarone and cadmium looked very similar not considering the absolute values of the overexpression levels. In contrast, the increase in the number of DEGs in doxorubicin was not caused by an increased expression but by many genes showing underexpression (see also **Fig 6**) and thus, had to be attributed to a different mode of action. Doxorubicin was therefore seen to be very different as compared to other compounds.

Rotenone exposure results in only 1 DEG at 8h and 2 at 20h but because of the low overall number of DEGs, this still results in a maximum z-score of 4.30596. GW788388 and tunicamycin do not show any DEG at any time point related to this pathway.


Arsenic − MT1G_26225
LFC
time (h)
0
2
4
6
8
1 2 4 6 8 12 16 20 24


Arsenic − MT2A_4334
LFC
time (h)
0
2
4
6
8
1 2 4 6 8 12 16 20 24


Cadmium − MT1G_26225
LFC
time (h)
2
4
6
8
1 2 4 6 8 12 16 20 24


Cadmium − MT2A_4334
LFC
time (h)
2
3
4
5
6
7
8
1 2 4 6 8 12 16 20 24

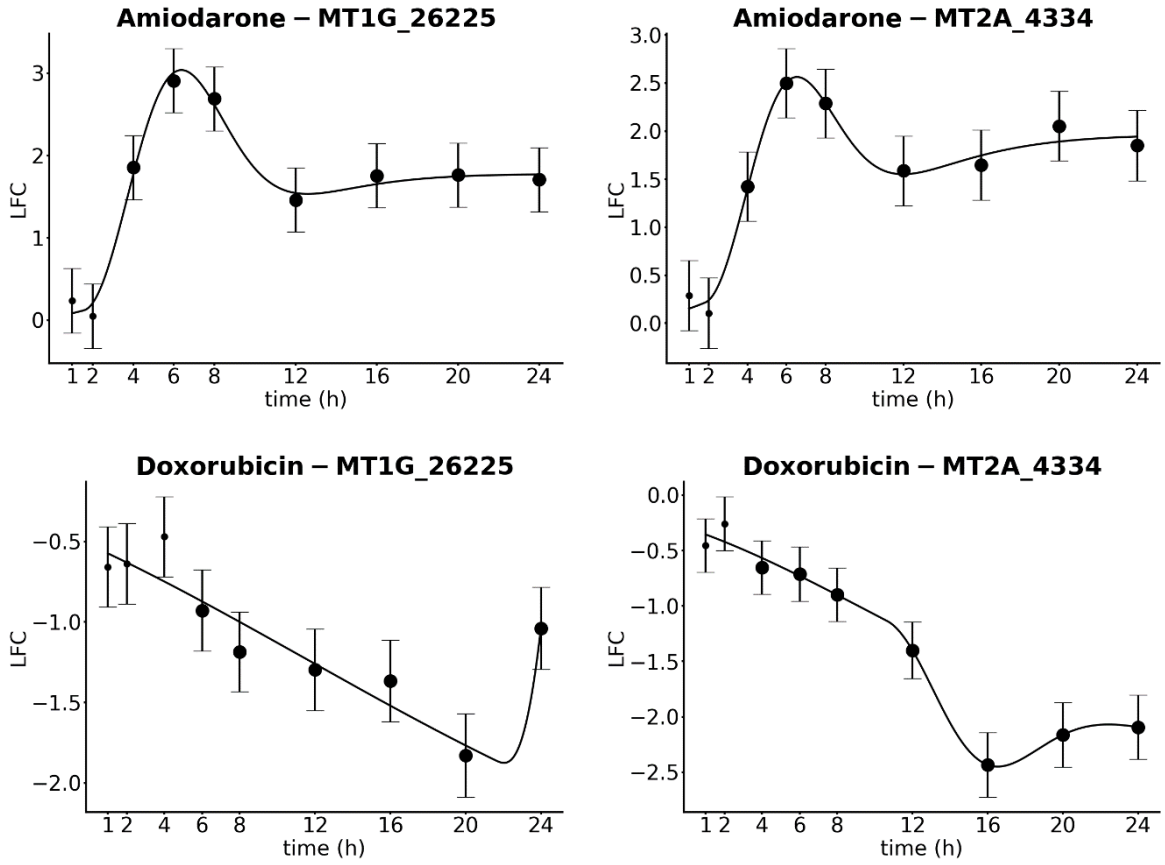


Fig 6 Temporal Profiles of MTF1 pathway related genes: Values of low significance (adjusted p-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard error for the fold changes.

3.4 Unfolded protein response

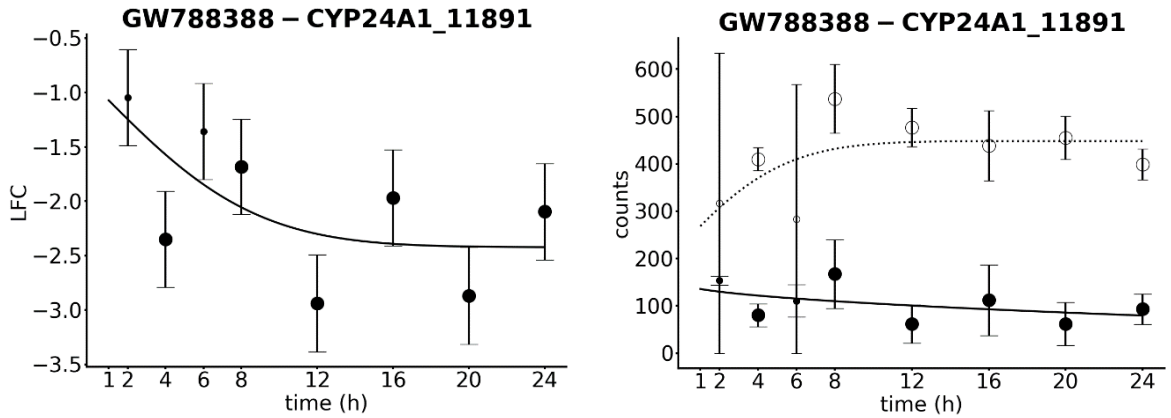
For the UPR pathway, ASNS_12040, ATF3_409, CALR_943, DDIT3_16736, PPP1R15A_14098, and TRIB3_7337 were the most differentially expressed genes activated in response to our set of chemicals (seen in at least 3 compounds). TRIB3_7337 and DDIT3_16736 showed the strongest overexpression in arsenic (highest LFC 5.5 at 6h, highest LFC 4.46 at 6h) and tunicamycin (highest LFC 4.8 at 12h, highest LFC 4.5 at 12h). PPP1R15A_14098 showed overexpression in all compounds except for tunicamycin, doxorubicin, and GW788388. The highest expression value was observed in rotenone (highest LFC of 3.1 at 24h). ASNS_12040 showed the highest expression in arsenic (highest LFC of 4.16 at 6h) and tunicamycin (highest LFC of 3.5 at 12h) and was also seen in amiodarone. ATF3_409 overexpression was observed in all compounds other than tunicamycin and GW78388. It showed the highest expression value in doxorubicin (highest LFC of 4.01 at 20h). Finally, CALR_943 showed the highest expression value in rotenone (highest LFC of 2.9 at 6h) and was also seen in doxorubicin and tunicamycin. The largest number of DEGs at a specific time point were 7 for cadmium (6h), 8 for tunicamycin (16h), 6 for arsenic (6h and 8h), 4 for amiodarone (24h), 5 for rotenone (24h) and 3 for doxorubicin (20h and 24h). GW788388 didn't show any DEG at any time point related to this pathway.

3.5 P53 signaling pathway

HSPA1A_3135, ATF3_499, CASP1_26966, RGCC_16894, GDF15_18329, PMAIP1_5221 and GADD45A_2569 were the p53 signalling pathway related genes most commonly influenced by the set of chemicals (at least 3 compounds). GDF15_18329 was differentially expressed in all compounds other than GW788388. It showed the strongest overexpression in doxorubicin (highest LFC of 6.4 at 16h) and arsenic (highest LFC 5.5 at 6h). GADD45A_2569 was differentially expressed in all compounds except for cadmium, tunicamycin, and GW788388. GADD45A_2569 and PMAIP1_5221 showed the highest expression in arsenic (highest LFC 2.8 at 24h, the highest LFC 3.5 at 8h, respectively) and doxorubicin (highest LFC 3.3 at 20h, the highest LFC 4.3 at 20h, respectively). RGCC_16894 showed the highest expression in doxorubicin (the highest LFC 4.5 at 24h) and cadmium (the highest LFC 3.5 at 4h). The maximum number of differentially expressed genes for this pathway resulted from doxorubicin exposure with 21 DEGs at 16h. Arsenic showed the highest number of DEGs at 24h (12 out of 231 total DEGs; z-score 2.84) and maintained a high number of DEGs until 24h. Cadmium showed the highest number of DEGs at 4h (9 out of 78 total DEGs; z-score 5.29), which is followed by a strong decrease. Amiodarone and rotenone showed an increase in the number of DEGs at the later hours with the highest number of DEGs (4 out of 79 total DEGs; 6 out of 89 total DEGs) at 24h. Amiodarone and rotenone had not more than 2 DEGs at any time point and GW788388 and tunicamycin only showed 1 DEG.

3.6 Gene profile of GW788388

GW788388 only showed a maximum of 1 differentially expressed probe associated with any of the pathways discussed above. CONSENSUSPathDB also didn't point to any other toxicity related pathway since all showed only a very low number of DEGs associated with them. Taking into account what is known about the compound (Petersen et al. 2008), the differential expression of WT1_17358, TXNIP_28030, CYP24A1_11891, and SERPINE1_6253 (LFC and normalized count profiles are given in **Fig 7**) can be associated to renal fibrosis. WT1_17358 and TXNIP_28030 showed upregulation (highest LFC of 2.1, 3.6 respectively) while CYP24A1_11891 and SERPINE1_6253, showed downregulation (LFC of -2.9 and -2.1 respectively).



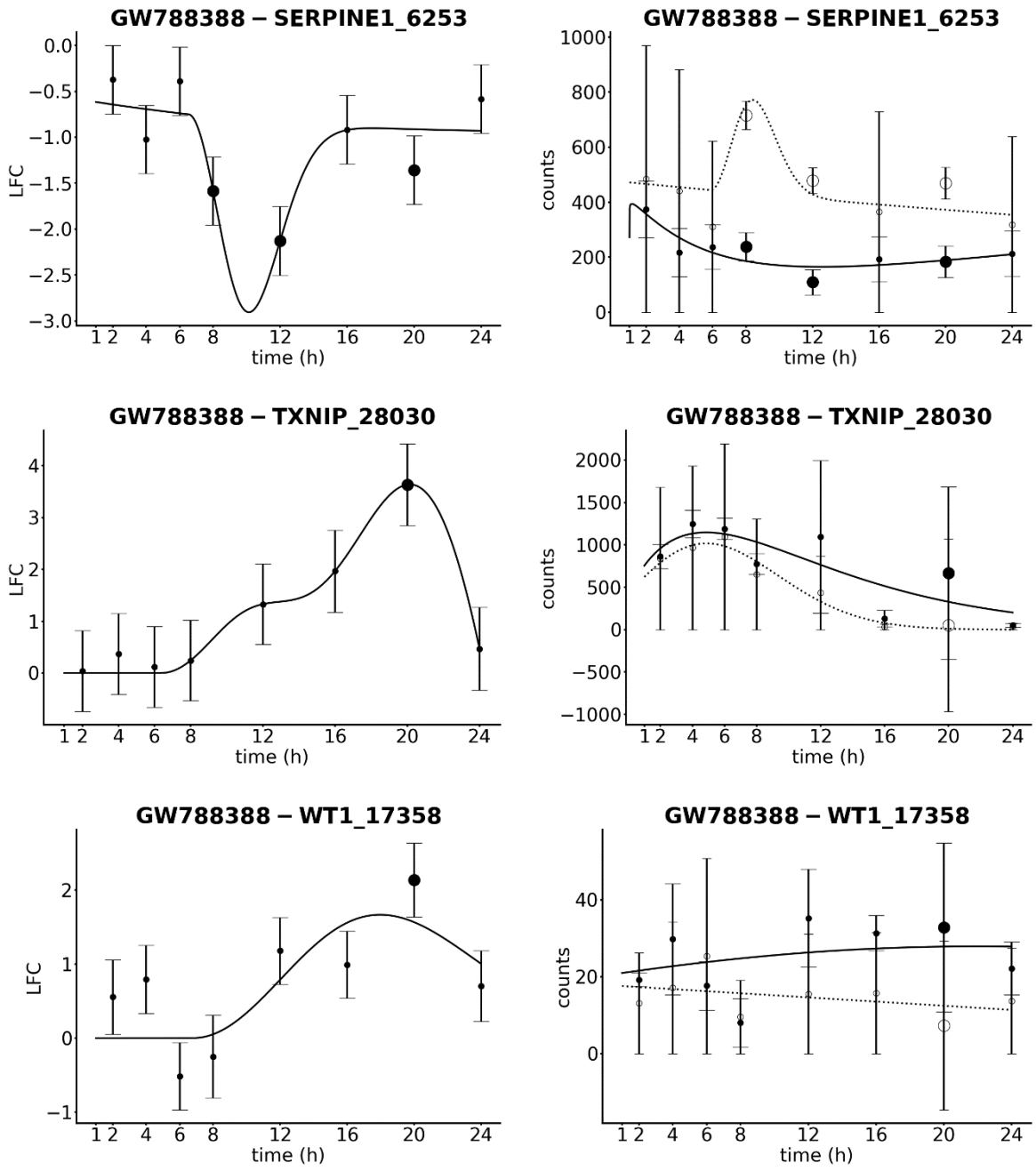


Fig 7 Temporal profiles of the TGF-β signaling pathway related DEGs on exposure to GW788388: The LFCs in column 1 (left) and the corresponding RCs of the treated (•) and control samples (○) are given in column 2 (right). Values of low significance (adjusted p-value above 0.05) are represented as smaller symbols. The error bars correspond to the standard deviation and the standard error for the normalized counts and fold changes, respectively. Interpolation lines were generated via the curve fitting procedure described in the material and methods

4. Discussion

4.1 Expression profile of GW788388

GW788388 was designed to be a potent and selective transforming growth factor-β (TGF-β) receptor kinases inhibitor to prevent kidney fibrosis and has shown no toxicity and adverse events so far (Petersen et al. 2008). Our data also showed that exposure to this compound didn't trigger any of the major stress responses discussed in detail below for the other compounds. Therefore, we will discuss this compound first by looking at the small number of genes, for which differential expression was seen, but then neglect it in the remaining part of the discussion since the focus of this paper is on the similarity of responses for toxicology relevant pathways.

Despite, having important functions like tumor suppression, immune-modulatory, etc., TGF-β which is an anti-inflammatory cytokine has been linked to causing fibrosis in the injured kidney. It promotes fibrosis by mediating the epithelial-to-mesenchymal transition (EMT) (Petersen et al. 2008; Nieto et al. 2016). EMT leads the epithelial cells to transit into the mesenchymal cells like myofibroblasts thereby disturbing the tubular basement membranes and enabling the cells to possess a greater ability for migration and invasion and, even if required during embryogenesis and wound healing, also paves the way to renal fibrogenesis (Park et al. 2007; Nieto et al. 2016).

Based on the chosen adjusted p-values and LFC thresholds, GW788388 exposure showed only 16 differentially expressed genes. Out of these, 4 genes, WT1, TNXIP, CYP24A1, and SERPINE1, are related to the target pathway TGF-β (Qi et al. 2007; Ma and Fogo 2009; Sakairi et al. 2011; Miller-Hodges and Hohenstein 2012; Hu and Zhang 2018; Xu et al. 2018). Our data showed overexpression of the genes WT1 and TXNIP after longer exposure to GW788388, which was the result of decreasing read counts of these genes in the control samples (see **Fig 7)** and a slight increase in the treated samples. This agreed with what is known from the literature. Studies have reported that a decrease in the levels of WT1 mRNA and protein levels in podocytes is due to elevated levels of TGF-β and lead to glomerulosclerosis, which is a characteristic of renal fibrosis following an increased activation of the EMT mechanism (Sakairi et al. 2011; Miller-Hodges and Hohenstein 2012). TGF-β has also been shown to negatively regulate the expression of TXNIP (Qi et al. 2007). However, it has to be noted that the fold changes need to be considered with caution since they can only be considered significant at the 20h time point according to the adjusted p-values. CYP24A1 and SERPINE1 showed under expression. Literature studies suggest that SERPINE1 aka PAI-1 and CYP24A1 both are expressed under the influence of the TGF-β signaling pathway and lead to an increase in EMT-related gene markers (Ma and Fogo 2009; Hu and Zhang 2018; Xu et al. 2018). Again, the fold changes were a result of changing levels of the genes in the control samples (see also **Fig 7**) with the increase speaking for some kind of activation of the TGF-β signaling pathway in the undisturbed cells, which is inhibited by GW788388.

Blocking of the TGF-β signaling pathway has not shown any short-term effects and even lifetime blocking did not cause adverse side effects in mice (Isaka 2018). This explains why our single-exposure, 24h experiments showed also mainly the wanted effect on genes associated with EMT following blocking of the TGF-β pathway. However, this does not exclude the unacceptable toxicity of continuous suppression of TGF-β with its highly pleiotropic function as cautioned in the review by Isaka (Isaka 2018). The other 12 differentially expressed genes not related to the TGF-β pathway were mostly found to be associated with metabolism pathways. The only exceptions were FSTL3, TF, SERPINA1, and ALB, which were associated with the pathways "Post-translational protein phosphorylation" and "Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)" based on CONSENSUSPathDB. None of the genes have been directly linked to any adverse effect and, thus, we saw no evidence for toxicity of GW788388 from the results of our study.

4.2 Oxidative stress and metal response

Oxidative stress is one of the most important mechanisms leading to adverse effects. This becomes easily obvious, besides the immense amount of scientific literature linking oxidative stress to the adversity of chemicals, by looking at the AOP-Wiki (https://aopwiki.org) as the primary source of cumulative information on curated adverse outcome pathways. Using the search term "oxidative stress" already resulted in 12 key events (KE) having the concept in their title and being part of 24 adverse outcome pathways (AOPs) (access date 5 August 2021). Full text search in the KE descriptions resulted even in 30 KE and searching for other synonyms will give even more results. Oxidative stress response, which is activated to remove excessive ROS, is mediated by the Nrf2 transcription factor. Nrf2 is responsible for maintaining cell homeostasis and aids in the expression of antioxidant enzymes encoding genes as well as pro-oxidant genes (Cullinan et al. 2004; Baird and Dinkova-Kostova 2011).

"Nrf2 pathway" pathway as defined in WikiPathways under entry ID WP2884 (Martens et al. 2021) was identified as an important pathway based on expression levels and the number of DEGs associated with this pathway in 5 of the studied compounds (arsenic, amiodarone, rotenone, doxorubicin, and tunicamycin) as well as cadmium chloride reported in the previous publication (Singh et al. 2021) and used here for comparison. The highest expression levels of the associated genes were seen in arsenic and cadmium. Similar to cadmium, the temporal profiles of the differential expressed genes (DEGs) on arsenic exposure showed very early and extremely high activation levels. Arsenic induces oxidative stress pathways by forming ROS like superoxide anion ($O_2^{\bullet-}$), hydroxyl radical (•OH), hydrogen peroxide (H_2O_2), singlet oxygen ($^1O_2$), and peroxyl radicals (Flora 2011; Jomova et al. 2011). Superoxide anions are formed in the arsenic-triggered deformed mitochondria and are converted to other reactive species. These species are then responsible for the interaction with various other molecules causing oxidative stress, DNA damage, and signaling pathways (Flora 2011; Jomova et al. 2011). Arsenic activated the p53 pathway to cause DNA damage, perturbations in the cell cycle, and apoptosis (Yin and Yu 2018).

However, in arsenic the response was sustained until 24h but all profiles start going to baseline level by 24h in the case of cadmium exposure (**Fig 4**). The only exception to this difference in behavior was HSPA1A, which was going down after the first maximum in both exposure scenarios. This gene belonged to the heat shock proteins like HSP6A discussed in detail in the earlier publication (Singh et al. 2021), which were originally characterized by their induction in thermal stress, but other stressors have also been identified, which may or may not be related to heat generated by intracellular events. Therefore, this continuous upregulation could be caused by its involvement in other stress pathways, like UPR (Terrab and Wipf 2020) and DNA repair (Boysen et al. 2019), and only partly related to oxidative stress.

The similarities and differences in the behavior of the two compounds could also be seen in the number of DEGs associated with the Nrf2 pathway, which kept increasing with time of exposure to arsenic while it went up to a maximum but then started going down after 8h in the case of cadmium (see **Fig 3**). This is translated to z-scores, which showed high significant overrepresentation of the pathway (significance cutoff for overrepresentation defined by a z-score >= 1.96) over the complete period for arsenic but started going down from around 12h in cadmium (see also **Fig 3**).

As discussed in detail in the previous paper (Singh et al. 2021) "MTF1-mediated metal response" pathway (R-HAS-5661231) ("MTF1-mediated metal response ". Reactome, release 79, https://reactome.org/content/detail/R-HSA-5661231 with StableID: R-HAS-5661231 (13 August 2021)) became the most important pathway at later hours during cadmium exposure as a long-term defense against metal ion. The same holds for arsenic. The response was sustained at high levels until 24h in response to both compounds as could be seen either by looking at individual gene profiles in **Fig 6** or the z-scores and corresponding number of DEGs in **Fig 5**. Interesting to note is that also the three organic compounds inducing oxidative stress, amiodarone, doxorubicin, and rotenone, discussed below also showed the metal response pathways significantly.

In the previous paper (Singh et al. 2021), we argued that the sustained response of MTF1-mediated stress response pathway and the increase in metallothionines and metal transporters in cells reduces the free cadmium levels to a degree low enough to not cause oxidative stress and, thus, the Nrf2 pathway can return to normal levels. In contrast, the remaining arsenic levels even after full activation of the metallothionein production are still producing high levels of oxidative species, to which the cells react by keeping the Nrf2 pathway up during the complete time. Further investigations are needed to see if higher cadmium levels also lead to such an overload of the MTF1-mediated stress response pathway and the need for support from other stress response pathways.

For the two compounds, arsenic and cadmium, and the two pathways Nrf2 and MTF1 discussed so far, the gene profiles and the profiles of the z-scores resulted in the same conclusions regarding the similarities and differences of the time evolution of the most important mechanism of actions. However, issues with the z-score could also already be seen, which became even more severe for the organic compounds. In the previous paper (Singh et al. 2021),

the seemingly decrease in the significance of the MTF1-mediated stress response pathway between 4h and 8h was already discussed. This was exactly the time when the corresponding genes were reaching their maximum level and also the number of DEGs was not different from the later time points showing higher z-score. Therefore, the reduction in z-score was not a result of changes in this pathway but was caused by the other pathways becoming important around this time, like Nrf2, but was going down to baseline later. Since the total number of DEGs across all pathways had a strong influence on the z-score, the MTF1-mediated stress response pathway became less important. While this was desired when comparing different pathways for one exposure scenario, i.e., one compound at one specific time point, since here the relative importance was relevant, it was less optimal to follow the behavior of one pathway in multiple exposure scenarios, either changing the compound, the time point, or the concentration. This became even more obvious when looking at the z-scores for arsenic exposure. MTF1-mediated stress response pathway seemed to lose significance after its peak at 2h. However, the number of DEGs stayed the same and the expression levels of the DEGs even doubled from 2h to 4h. Therefore, we concluded that this pathway stayed equally important over the complete period for both arsenic and cadmium in contrast to the Nrf2 pathway in cadmium, which became less important at later times even if the z-score profile looked very similar to the one of the MTF1 mediated stress response pathway in arsenic.

A second possible issue with z-scores was the high value at 1h in the Nrf2 pathway profile of arsenic even if the number of DEGs associated with this pathway was only 1. Even if this fitted into the picture from the analysis of the time series that this pathway was extremely relevant, judging the relevance of a pathway just from this one-time point and especially when comparing to other compounds or time points could be misleading. The reason is the very low number of DEGs in corresponding samples, which led to the high z-scores even with very few DEGs for a specific pathway. Similar to the conclusion given in the previous paper, that LFCs are mainly useful to identify interesting genes but that conclusions should be based on or at least backed up by normalized read counts, the z-scores should be used to identify pathways, but conclusions need to be based on the number of DEGs or even better analysis of the corresponding genes since the latter is less influenced by cutoffs used for the fold changes and p-values. Therefore, we will use gene expression profiles as the primary information to guide the discussion and only reference z-scores from time to time mainly to demonstrate these issues even more.

Nrf2 pathway related genes were also seen to be activated in the organic compounds namely amiodarone, doxorubicin, rotenone, and tunicamycin as presented in **Fig 8** in the form of a heatmap. The biggest difference however was that HMOX1, strongly overexpressed in the metal compounds, was seen to be weakly expressed in the organic compounds or not at all. Even if more data is needed to confirm this, the results gave some evidence that the details of the oxidative stress response are different in metal-containing and organic compounds or that HMOX1, besides related to oxidative stress, is also relevant for the metal ion response.

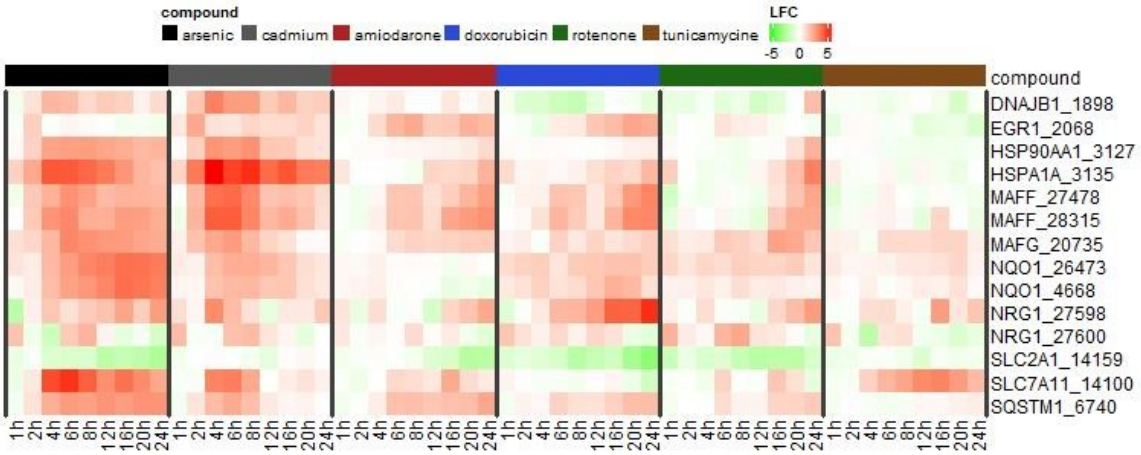


Fig 8 Heatmap representing LFCs of Nrf2 pathway related genes: The shade of red colour represents overexpression and the shade of green color underexpression.

Amiodarone and doxorubicin showed the highest overrepresentation of the Nrf2 pathway with a very similar expression pattern of the related genes as the metal compounds. Based on the study performed by Chakraborty et al.(Chakraborty et al. 2014), the kidney is affected by amiodarone, which releases iodine triggering the production of ROS and ultimately leading to oxidative stress. Doxorubicin is also known to target the proximal tubule cells leading to acute renal failure. It leads to activation of oxidative stress pathway, ROS production, and apoptosis as a result of inflammation (Carvalho et al. 2009; Abdelmeguid et al. 2010; Afsar et al. 2020). Especially the high expression levels of MAFF, MAFG, and partly SQSTM1 and some downregulation of SLC2A1 seemed to be characteristic of the oxidative stress causing compounds. Larger differences were seen, besides HMOX1 discussed above, for DNAJB1 and HSP90AA1 (only differentially expressed in the metal compounds), NQO1 (not differentially expressed in amiodarone), and SLC7A11 (not differentially expressed in doxorubicin).

One interesting case is NRG1. One of the probes, NRG1_27598, showed the highest overexpression seen for Nrf2 pathway related genes for doxorubicin, especially at late hours. The temporal profile also showed a steady buildup of the signal over time, which increased the confidence that this signal was associated with a real cellular mechanism. However, the other probe, NRG1_27600, didn't show this increase in the signal at the end of the period and had a quite noisy time profile across all compounds with absolutely no correlation to the other probe of the same gene. This different behavior of the two probes will need further investigation in the future.

Tunicamycin showed almost no change in the expression levels of Nrf2 pathway related genes (see **Fig 8**). Especially MAFF as a possible biomarker across all the other oxidative stress-causing compounds discussed above was not influenced and MAFG showed only a slight increase over baseline. The most overexpressed Nrf2 pathway related genes are NRG1, with the issues described above, and SLC7A11, which was only seen in the metal compounds but not the other organic compounds. Therefore, our results suggest that oxidative stress is not a major response to the adversity of tunicamycin.

Rotenone being used widely as a pesticide has been banned currently in many countries due to its toxic character (Heinz et al. 2017). It causes severe oxidative stress due to excessive production of ROS, glutathione, and malonaldehyde, and by blocking the superoxide dismutase (SOD) and glutathione peroxidase (GSH-Px) activity in the kidney (Jiang et al. 2017). Despite this strong literature evidence, rotenone showed relatively low deregulation of the Nrf2 pathway genes in our experiments. Relevant genes were coming up only close to the maximum exposure times. MAFF, MAFG, HSP90AA1, HSPA1A, NRG1_27598, and SQSTM1 showed significant upregulation and SLC2A1 downregulation in the last 1 to 3 time points (16h, 20h, and 24h) with most of them steadily increasing/decreasing in this period. Therefore, by looking at the time profiles, it can be concluded that rotenone also caused oxidative stress in iPSC-derived kidney cells but that the effect was starting slower than in the other compounds and the stress response was probably not reaching its full activation levels in the first 24h. When compared to the metal compounds, which already showed quite high overexpression or even the highest values for most genes at 2h or 4h, the strong increase in the signals of the organic molecules is, in general, shifted to later time points. EGR1 in amiodarone started to increase at 4h as one of the earliest signals in any organic compound, while the other genes showed increased overexpression for this compound from 6h on. 6h was also the time point when the changes start to occur in doxorubicin. Finally, an observation in all the organic compounds, already discussed for rotenone above, was that the highest differential expression levels for many of the Nrf2 pathway related genes were seen at 24h which is the maximum exposure time, which suggests that the maximum activation of the pathway might be even later, and longer exposure times would be needed to show the full effect.

As the last point of the discussion of the Nrf2 pathway, we wanted to check if it would have been possible to automatically identify this stress response as a major mechanism for amiodarone, rotenone, and doxorubicin just using the z-score measure without any additional knowledge from literature about the compounds. The z-score temporal profiles are shown in **Fig 2**. Compared to the z-scores of the metal compounds also shown in the same figure, values of the organic compounds were very low with maximum values below 4 for amiodarone and rotenone and even below 1 for doxorubicin. This would suggest that oxidative stress was of low importance in doxorubicin since the corresponding response pathway was not significantly overrepresented (significance limit = 1.96). At 24h, the z-score was negative for doxorubicin, representing an underrepresentation in our samples, even when the number of DEGs was higher (doxorubicin: z-score = -0.31 resulting from 8 DEGs) as compared to the other compounds namely amiodarone and rotenone (z-score=1.63 resulting from 4 DEGs, z-score=2.08 resulting from 5 DEGs, respectively) at this time point and around 50% and 70 % of the maximum seen in arsenic and cadmium, respectively. Tunicamycin, where oxidative stress was less relevant than in doxorubicin, showed very similar low z-scores but even a higher maximum of around 1.49 at 6h. This could be reasoned by the extreme influence of the calculated z-scores on the total number of differentially expressed genes compared to the overall number of measured genes. Doxorubicin showed 399 DEGs at 24h, which was around 14 % of the total genes covered by the TempO-Seq panel. For pathways like the

Nrf2 pathway, with which many genes were associated (76 genes of the TempO-Seq panel) but only a small number was influenced by toxins, this resulted in extremely low z-scores even if a reasonable number of genes of the pathway were differentially expressed. Therefore, this pathway was at a low rank in the list of relevant pathways for this compound. This first seemed to contradict the literature evidence and also our discussion above. However, the large number of DEGs was a sign that many processes in the cell were adversely affected and multiple defense mechanisms were activated with the p53 pathway and related mechanisms like DNA damage, genotoxicity, and apoptosis being on the highest ranks according to z-score. As we discuss below in the section on the p53 pathway, the experiments can thus correctly identify the adversity of the compound and its mode of action, in which oxidative stress plays only one of the many roles. However, it also showed that the z-score might only be useful to judge importance in one specific exposure scenario, but different scenarios should not be compared based on absolute values. In any case, detailed analysis should be performed on all values used as input for the z-score calculation and especially the number of totals and pathway-specific DEGs. Additionally, a detailed analysis is also needed for pathways of most importance, also the underlying LFC, and raw data to evaluate what genes, pathways, and mechanisms are relevant for assessing the risk of chemical compounds.

4.3 Unfolded protein response

Protein misfolding or (partial) unfolding in the ER is an often-seen adverse side effect of drugs and chemicals. To reduce such ER stress, restore its homeostasis, and also maintain a balance between the disrupted ER and stress of protein folding, the UPR cell signaling system is activated (Li et al. 2011; Foufelle and Fromenty 2016; Guha et al. 2017). Many recent studies have also revealed the importance of the UPR pathway in full-body physiology and in different disease mechanisms including inflammation, neurodegeneration, and organ toxicity (Cornejo et al. 2013; Foufelle and Fromenty 2016; Lindholm et al. 2017). Two pathways are included in CONSENSUSPathDB describing UPR, “Unfolded Protein Response (UPR)” (R-HSA-381119) as defined by the Reactome Pathway library (Jassal et al. 2020) and “Photodynamic therapy-induced unfolded protein response” (WP3613) from WikiPathways (Slenter et al. 2018; Martens et al. 2021). It is interesting to note that only four genes (ATF6, EIF2AK3, ERN1, and HSPA5) from the TempO-Seq panel were shared between the pathways. This little overlap resulted in very different results for the two pathways with WP3613 identified as one of the most important pathways in 4 out of 6 studied compounds while R-HSA-381119 was classified as of low significance for all the tested molecules (a maximum of two genes associated with R-HAS-381119 were found in rotenone). This showed, as already highlighted in the earlier publication (Singh et al. 2021), how important the choice of the pathway library is and the advantage of using the CONSENSUSPathDB which combines multiple of these libraries. However, we also need the mention the influence of the gene selection in the targeted TempO-Seq panel. Only 6 from the 22 genes associated with R-HSA-381119 are in the panel while there are 17 from the 28 genes associated with WP3613.

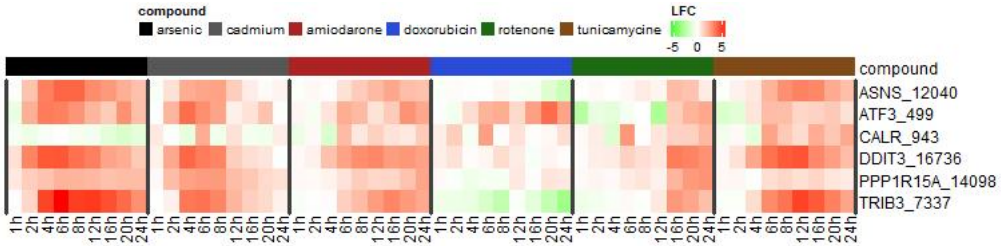


Fig 9 Heatmap representing LFCs of UPR pathway related genes: The shade of red color represents the overexpression and the shade of green color under expression.

Based on R-HSA-381119 identified as the relevant UPR pathway for our study, related genes were overexpressed in the organic compounds amiodarone, rotenone, and tunicamycin as well as the metal compounds arsenic and cadmium chloride (see **Fig 9**). Arsenic and cadmium showed again very high overexpression levels and the many DEGs using our selection criteria. The list of DEGs also shared 6 genes demonstrating a common mode of action (see **Fig 10**). Similar to the Nrf2 pathway in cadmium, the signals were going down with longer exposure for both metal compounds providing evidence that the activation of the MTs is reducing oxidative as well as ER/unfolded protein stress. This agrees with existing knowledge that arsenic induces UPR triggering eukaryotic translation initiation factor 2 subunit α (eIF2α) mediated induction of ATF4 and processing of ATF6 (Weng et al. 2014). ATF4 induction could be directly seen in our data even if the levels only reached the LFC cutoff criteria at one time point (6h). ER stress can either be directly induced by arsenic as proven by many studies (Ramadan et al. 2009; Jacobson et al. 2012) or UPR can be induced by the overproduction of ROS in mitochondria by making the respiratory chain process ill-functioned (Paul et al. 2008; Vineetha et al. 2015; Foufelle and Fromenty 2016), which then can lead to misfolding of proteins in the ER (Inagi 2009). The latter process was verified in our data by the correlation in the activation of the Nrf2 and UPR pathways.

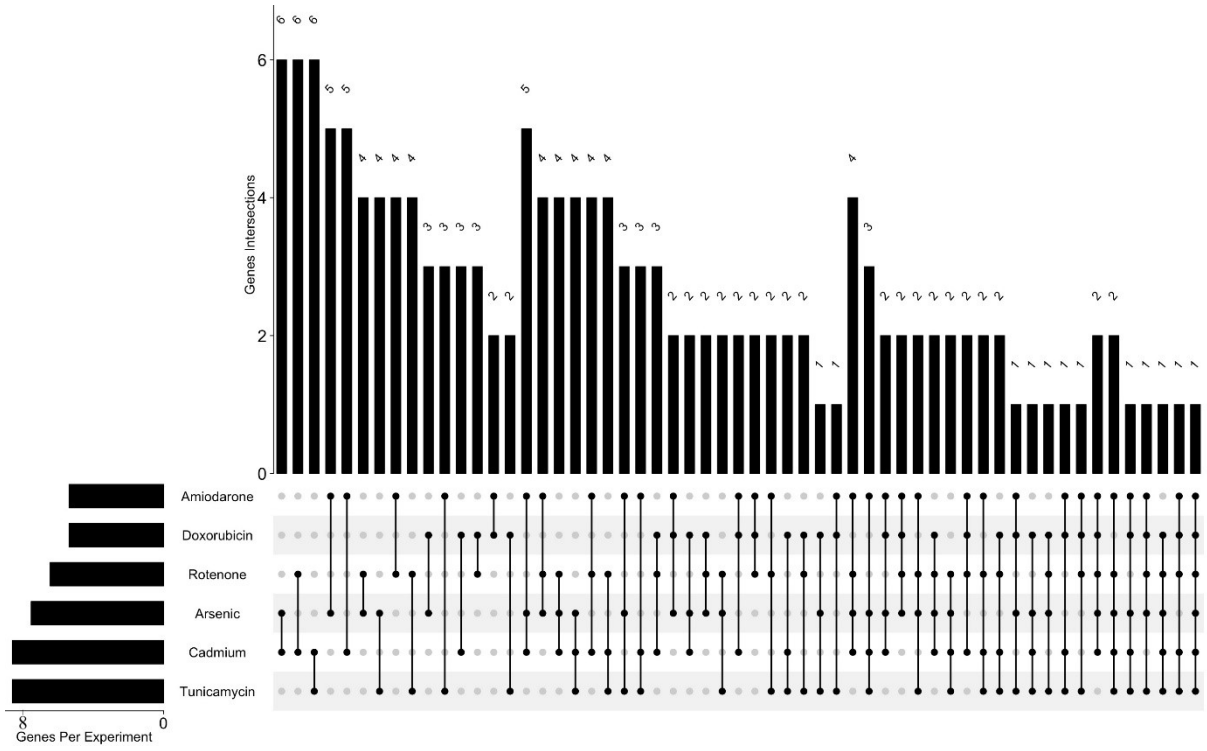


Fig 10 Intersection plot of differentially expressed genes associated with the UPR pathway using the UpSetR package (Conway et al. 2017) in R: The horizontal bars on left represent the total number of UPR related DEGs associated with compounds amiodarone, doxorubicin, rotenone, arsenic, cadmium, and tunicamycin. The vertical bars on the top represent the number of DEGs found common between different sets of compounds. The dots joined with the black line below connect the compounds in the set that represents the intersection number above.

Tunicamycin showed an even larger number of DEGs associated with this pathway (8 DEGs at 16h compared to 6 genes for arsenic and 7 genes for cadmium both at the 6h time point) and similar expression levels. Thus, tunicamycin is the strongest organic introducer of UPR (at least in the chosen concentrations). Tunicamycin is a drug well-known to induce ER stress leading to acute renal tubular necrosis by directly attacking the protein glocalization leading to many unfolded proteins (Guha et al. 2017; Mohammed-Ali et al. 2017). Overlap of DEGs with the metal compounds as well as the two other organic compounds triggering UPR, amiodarone, and rotenone, was quite high with at least 3 common genes. This was quite remarkable since amiodarone had only 4 DEGs and rotenone 5 DEGs associated with this pathway and therefore showed overlaps of 75% and 60%. Amiodarone is well-known to induce ER stress in thyroid cells triggering the unfolded protein response (UPR) (Lombardi et al. 2015) and has also been linked to renal failure due to hypothyroidism by increasing the glomerular filtration and renal plasma flow levels (Luciani et al. 2009). Exposure to rotenone was shown to cause ER stress due to ROS formation in the ER. UPR activation can reduce this stress but prolonged exposure leads to the apoptotic death of neurons to avoid further damage to other nearby cells (Goswami et al. 2016).

In contrast, doxorubicin only showed overexpression of a single gene, ATF3, of the UPR pathway over a longer period. ATF3 is also known to play a part in stabilizing the p53 signaling pathway in the event of DNA damage (Yan and Boyd 2006). Therefore, we would

argue that doxorubicin is not causing ER stress and that the overexpression of this gene is not related to UPR. This is strengthened by the fact that ATF3 is also of lesser relevance for tunicamycin, which showed strong UPR but low p53 pathway activation (see below). The exclusion of doxorubicin from the UPR-causing compounds became even more evident when looking at the other UPR related gene, TRIB3 was seen to be downregulated due to doxorubicin exposure while it was upregulated in the rest of the compounds. Taking all discussions so far and the additional information from the DEG overlap into account, overexpression of TRIB3 and DDIT3 can be looked upon as good biomarkers for UPR in the iPSC-derived PTL model since they are seen in all compounds with clear links to UPR as reported in the literature but not in doxorubicin as the only UPR-inactive compound. TRIB3 has been reported in the literature to be upregulated in response to UPR and ER stress and is involved in managing many events like the sustainability of the UPR, promotion of homeostasis, etc. (Nicoletti-Carvalho et al. 2010; Mondal et al. 2016) In the event of stress, one of the ER stress sensor PERK leads to the translocation of transcription factor ATF4 which in turn activates DDIT3. The activation of DDIT3 has also been linked to the upregulation of the TRIB3 (Rashid et al. 2015).

ASNS and PPP1R15A are other candidates even if the expression levels didn't reach the LFC threshold in all compounds. PERK was reported to lead, besides DDIT3 activation described above, to the activation of ASNS, which along with other ER stress response elements form a responsive sequence to ER stress (Gjymishka et al. 2009). PERK-induced translocation of ATF4 also leads to the direct activation of PPP1R15A, which in turn acts in the proper reestablishment of the protein synthesis disrupted during ER stress (Young et al. 2016). In contrast, ATF3 can not be used as discussed above and the relationship between CALR and UPR is not clear from our results since it is seen in 4 UPR-active compounds but show no differential expression in arsenic.

Besides the similarities described so far, the main difference between the expression profiles was seen in the activation time of the genes. Similar to the Nrf2 pathway, the genes reached high activation levels already very early after exposure to the two metal compounds (4h to 6h time point). As already discussed above, the levels were then decreased over time. Buildup in tunicamycin and amiodarone was slightly delayed starting at around 6-8h and reaching maximum overexpression at 12-16h. After that, the levels were decreasing probably due to metabolic deactivation. Rotenone showed even longer delayed UPR expression levels starting to rise at 16h and some of them probably did not reach their full activation levels in the first 24h.

4.4 P53 signaling pathway

The p53 transcriptional network is one of the most important responses to stresses that can disrupt the fidelity of DNA replication and the cell division process (Harris and Levine 2005). Cellular homeostatic mechanisms that regulate DNA replication, chromosome segregation, and cell division are affected by the intrinsic and extrinsic stress signals and lead to the activation of the p53 transcriptional network (Vogelstein et al. 2000). The ability to trigger cell cycle arrest and ultimately cell death by apoptosis makes the p53 signaling pathway the most

important preventive cancer factor and tumor suppressor and activation of this pathway speaks for the adversity of a chemical compound (Chen 2016). As already mentioned above, "Direct p53 effectors" (p53downstreampathway) defined in Pathway Interaction Database (Schaefer et al. 2009) was identified as the most relevant pathway for doxorubicin based on z-scores and also included most genes relevant for the evaluation of p53 activation caused by chemical exposure, we will base all following results also for the other compounds on this pathway definition.

When analyzing the temporal profiles of the p53 pathway-related genes during doxorubicin exposure, it is interesting to note that the increase started earlier (as early as at the 2h or 4h time point) than the oxidative stress response (6h or 8h time points or even later) (see **Fig 11** in comparison to **Fig 8** above). This is in line with a study reporting the activation of p53 as early as 6h after exposing H9c2 cardiomyoblasts to the doxorubicin (Sardão et al. 2009).

Doxorubicin increases the levels of inflammatory factors TNF-α and interleukin increasing the renal matrix and glomerular sclerosis in the kidney. However, it has also been shown that it damages the renal cells indirectly by lipid peroxidation (Liu et al. 2016), which is mediated by the Nrf2 pathway (Dodson et al. 2019), in glomerular epithelial cells leading to functional damage in protein and glucose metabolism (Liu et al. 2016). The increase in the levels of p53 protein happens due to an increase in the inflammatory factor TNF-α in response to damage to DNA (Ladiwala et al. 1999). Activation of the p53 gene in response to DNA damage by doxorubicin has also been seen as one of the early events in cardiotoxicity and may regulate the oxidative stress (Sardão et al. 2009).

Arsenic, cadmium, amiodarone, and rotenone all also showed the p53 pathway as an important response to the exposure however in lower ranks according to z-scores with arsenic and amiodarone with the highest (12) and the lowest (4) number of DEGs associated to the pathway at any specific time point. The temporal profiles were quite comparable to the ones of Nrf2 pathway related genes providing some evidence of a direct link between oxidative stress and DNA damage in form of a correlation between the two responses (see **Fig 11**). Oxidative stress was, *e.g.*, already reported to be mediating the arsenic-induced, p53-mediated apoptosis (Liu et al. 2003). It should also be noted that the evidence for cadmium activating the p53 pathways is much stronger when the conclusions are based on the CONSENSUSPathDB compared to the transcription-factor-based association used in the previous publication (Singh et al. 2021) with now 9 related DEGs compared to the 4 reported before. Finally, tunicamycin showed a weaker but overall consistent activation of p53 genes. This is consistent with reports that tunicamycin induces ER stress, which triggers UPR to cause the elevation of p53 gene levels by activating another stress response pathway, NF-κB, and finally resulting in cell death (Lin et al. 2012). The lower levels seen in our experiment for all but 1 gene not reaching the LFC threshold for our DEG definition can be seen as a sign that UPR probably with some help from p53 can counter the toxic effects of tunicamycin under the applied concentrations and the cell has not relied on more severe defense mechanisms including its death to protect the organ or organism.

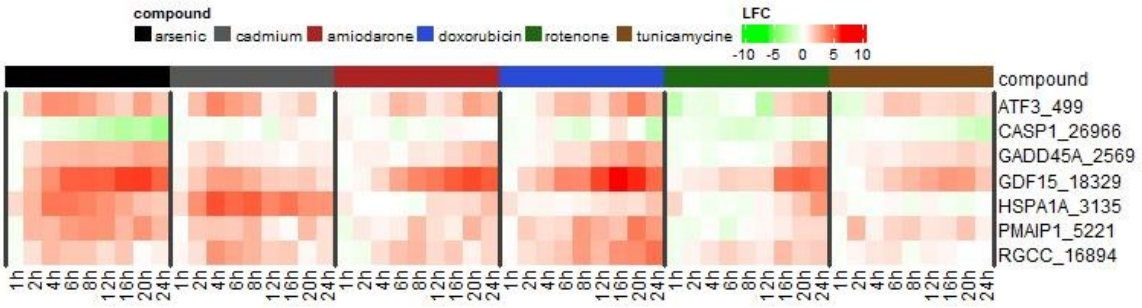


Fig 11 Heatmap representing LFCs of p53 pathway related genes: The shade of red colour represents the overexpression and the shade of green color under expression.

Looking at the gene overlap in the expression profiles related to p53 (see **Fig 12**) reveals that there was a clear subset of genes relevant for the p53 response to adverse chemicals, a phenomenon already seen for UPR. From the 79 genes associated with the pathway, a maximum of 21 were differentially expressed (doxorubicin exposure for 16h, supplementary data available on request). As expected, more genes were activated by more active chemicals like doxorubicin and arsenic but the overlap with the DEGs of the less potent chemicals was almost perfect. Doxorubicin with the largest number of DEGs had an overlap of 11 genes with arsenic and 5 with cadmium, amiodarone, and rotenone. All these five compounds had 3 genes in common, ATF3, GDF15, and RGCC, also showing the highest overexpression levels. The first two also showed an increased expression level in the tunicamycin exposure even if the levels didn't reach the LFC threshold as already discussed above. Links to these genes are already evident in the literature. GDF15 has been seen to be upregulated both at gene and protein levels by the activated p53 in response to doxorubicin in bladder carcinoma cells (Tsui et al. 2015). The p53 triggers high expression RGCC in many cancerous cells and has been involved in the cell cycle process specifically G1/S and G2/M phase transitions (Counts and Mufson 2017).

However, if these genes can be used as biomarkers for stresses that can disrupt the fidelity of DNA replication and cell division in the IPSC-derived kidney model cannot be judged since there is not a negative control compound available for which these genes should not show any change in the expression.

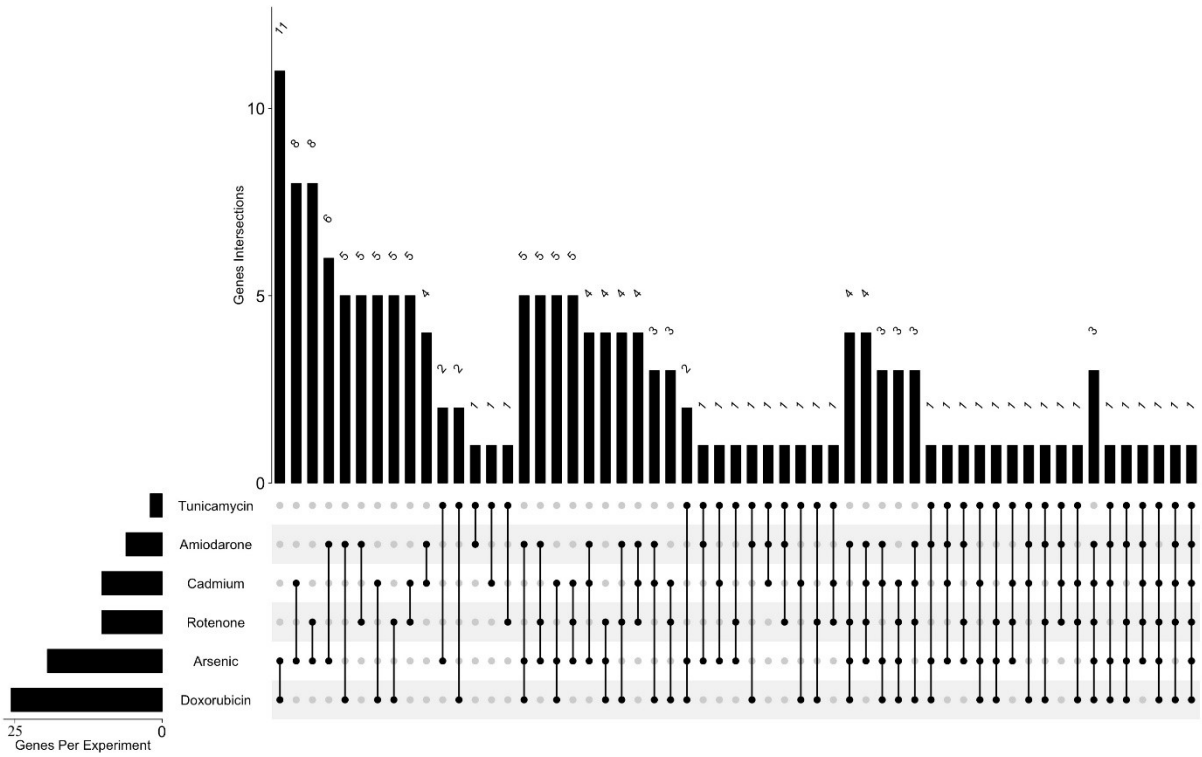


Fig 12 Intersection plot of DEGs associated with the p53 pathway using the UpSetR (Conway et al. 2017) package in R: The horizontal bars on left represents the total number of p53 related DEGs associated with compounds amiodarone, doxorubicin, rotenone, arsenic, cadmium, and tunicamycin. The vertical bars on the top represent the number of DEGs found common between different sets of compounds. The dots joined with the black line below connect the compounds in the set that represents the intersection number above.

5. Conclusions

This study validates the usefulness of the combination of the IPSC-derived kidney model with targeted transcriptomics (TempO-Seq) analysis at multiple time points as a tool for identifying mechanistic causes for adverse effects of chemicals as proposed in (Singh et al. 2021). Looking at the temporal profiles of genes related to major stress response pathways demonstrated that adversity not only differs concerning its severity but also in the timeframe this adversity is seen. Cadmium, for example, showed a very strong oxidative stress response at early time points but this went down quickly again. In contrast, expression levels of Nrf2 pathway related genes only reached high levels at the longest exposure times for doxorubicin, known to cause high levels of ROS, and these levels might even go higher if the experiment would be extended to include longer exposure times. Therefore, without taking the time parameter into account, *i.e.,* just using one defined exposure time, the full effect would not be seen for one or both compounds, or activation of the Nrf2 pathway might be completely missed. For example, when only using the 24h time point as often done in transcriptomics studies, oxidative stress would be not be identified as a major stress response due to cadmium exposure.

Regarding the specific chemicals used as reference compounds in this study to compare the results to existing knowledge, the major outcomes can be summarised as follows:

1. The gene profiles after exposure to GW788388 did not show any evidence of at least the short-term toxic effects of this compound. The only signals seen were related to the TGF-β inhibition being the therapeutic effect and metabolism.
2. Most of the compounds showed oxidative stress leading either to the UPR pathway or p53 pathway activation or both.
3. The metal compounds showed very similar behavior with the MTF1-mediated stress response pathway as the major pathway and strong activation of the Nrf2-mediated oxidative stress response, UPR, and the p53 signaling pathway. However, there are differences in how successful the MTF1-mediated stress response pathway counters the adverse effect alone and how long it needs to be supported by the other pathways.
4. Tunicamycin showed a different mode of action compared to the other compounds for activating UPR without causing oxidative stress response and only weak activation of p53.
5. Doxorubicin was a very potent p53 activator that masked the molecular initiating event and other effects like oxidative stress.

5. Conflict of interest statement

The authors declare no competing interests.

6. Acknowledgements

This work was funded by the EU project in3 a Marie Sklodowska-Curie Action - Innovative Training Network under grant no. 721975.

7. References

Abdelmeguid NE, Chmaisse HN, Zeinab NSA (2010) Protective effect of silymarin on Cisplatin-induced nephrotoxicity in rats. Pakistan J Nutr 9:. https://doi.org/10.3923/pjn.2010.624.636

Afsar T, Razak S, Almajwal A, Al-Disi D (2020) Doxorubicin-induced alterations in kidney functioning, oxidative stress, DNA damage, and renal tissue morphology; Improvement by Acacia hydaspica tannin-rich ethyl acetate fraction. Saudi J Biol Sci 27:. https://doi.org/10.1016/j.sjbs.2020.07.011

Andrews GK (2000) Regulation of metallothionein gene expression by oxidative stress and metal ions. In: Biochemical Pharmacology

Baird L, Dinkova-Kostova AT (2011) The cytoprotective role of the Keap1-Nrf2 pathway.

Arch Toxicol 85:241–272. https://doi.org/10.1007/s00204-011-0674-5

Boysen M, Kityk R, Mayer MP (2019) Hsp70- and Hsp90-Mediated Regulation of the Conformation of p53 DNA Binding Domain and p53 Cancer Variants. Mol Cell 74:. https://doi.org/10.1016/j.molcel.2019.03.032

Carvalho C, Santos R, Cardoso S, et al (2009) Doxorubicin: The Good, the Bad and the Ugly Effect. Curr Med Chem 16:. https://doi.org/10.2174/092986709788803312

Chakraborty A, Mondal C, Sinha S, et al (2014) Amiodarone induced oxidative stress in stress - Vulnerable organs of adult male rats. Asian J Pharm Clin Res 7:

Chandrasekaran V, Carta G, da Costa Pereira D, et al (2021) Generation and characterization of iPSC-derived renal proximal tubule-like cells with extended stability. Sci Rep 11:11575. https://doi.org/10.1038/s41598-021-89550-4

Chen J (2016) The cell-cycle arrest and apoptotic functions of p53 in tumor initiation and progression. Cold Spring Harb Perspect Med 6:. https://doi.org/10.1101/cshperspect.a026104

Cooper GM, Hausman RE (2007) The Cell: A Molecular Approach 2nd Edition

Cornejo VH, Pihán P, Vidal RL, Hetz C (2013) Role of the unfolded protein response in organ physiology: Lessons from mouse models. IUBMB Life 65

Counts SE, Mufson EJ (2017) Regulator of cell cycle (RGCC) expression during the progression of Alzheimer's disease. Cell Transplant 26:. https://doi.org/10.3727/096368916X694184

Cullinan SB, Gordan JD, Jin J, et al (2004) The Keap1-BTB protein is an adaptor that bridges Nrf2 to a Cul3-based E3 ligase: oxidative stress sensing by a Cul3-Keap1 ligase. Mol Cell Biol 24:8477–8486. https://doi.org/10.1128/MCB.24.19.8477-8486.2004

Dodson M, Castro-Portuguez R, Zhang DD (2019) NRF2 plays a critical role in mitigating lipid peroxidation and ferroptosis. Redox Biol. 23

Flora SJS (2011) Arsenic-induced oxidative stress and its reversibility. Free Radic. Biol. Med. 51

Foufelle F, Fromenty B (2016) Role of endoplasmic reticulum stress in drug-induced toxicity. Pharmacol. Res. Perspect. 4

Gjymishka A, Su N, Kilberg MS (2009) Transcriptional induction of the human asparagine synthetase gene during the unfolded protein response does not require the ATF6 and IRE1/XBP1 arms of the pathway. Biochem J 417:. https://doi.org/10.1042/BJ20081706

Goswami P, Gupta S, Biswas J, et al (2016) Endoplasmic Reticulum Stress Plays a Key Role

in Rotenone-Induced Apoptotic Death of Neurons. Mol Neurobiol 53:. https://doi.org/10.1007/s12035-014-9001-5

Guha P, Kaptan E, Gade P, et al (2017) Tunicamycin induced endoplasmic reticulum stress promotes apoptosis of prostate cancer cells by activating mTORC1. Oncotarget 8:. https://doi.org/10.18632/oncotarget.19277

Gunther V, Lindert U, Schaffner W (2012) The taste of heavy metals: Gene regulation by MTF-1. Biochim Biophys Acta. https://doi.org/10.1016/j.bbamcr.2012.01.005

Harris SL, Levine AJ (2005) The p53 pathway: Positive and negative feedback loops. Oncogene 24

Heinz S, Freyberger A, Lawrenz B, et al (2017) Mechanistic Investigations of the Mitochondrial Complex i Inhibitor Rotenone in the Context of Pharmacological and Safety Evaluation. Sci Rep 7:. https://doi.org/10.1038/srep45465

HBC Training, accessed November 2020, <https://github.com/hbctraining/DGE_workshop>

Horn HF, Vousden KH (2007) Coping with stress: Multiple ways to activate p53. Oncogene 26

Hu N, Zhang H (2018) CYP24A1 depletion facilitates the antitumor effect of vitamin D3 on thyroid cancer cells. Exp Ther Med 16:. https://doi.org/10.3892/etm.2018.6536

Inagi R (2009) Endoplasmic Reticulum Stress in the Kidney as a Novel Mediator of Kidney Injury. Nephron Exp Nephrol 112:. https://doi.org/10.1159/000210573

Isaka Y (2018) Targeting TGF-β signaling in kidney fibrosis. Int. J. Mol. Sci. 19

Itoh K, Chiba T, Takahashi S, et al (1997) An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem Biophys Res Commun 236:313–322

Jacobson T, Navarrete C, Sharma SK, et al (2012) Arsenite interferes with protein folding and triggers formation of protein aggregates in yeast. J Cell Sci 125:. https://doi.org/10.1242/jcs.107029

Jennings P (2013) Stress response pathways, toxicity pathways and adverse outcome pathways. Arch. Toxicol. 87

Jennings P, Limonciel A, Felice L, Leonard MO (2013) An overview of transcriptional regulation in response to toxicological insult. Arch. Toxicol. 87

Jiang XW, Qiao L, Feng XX, et al (2017) Rotenone induces nephrotoxicity in rats: oxidative damage and apoptosis. Toxicol Mech Methods 27:. https://doi.org/10.1080/15376516.2017.1333553

Jomova K, Jenisova Z, Feszterova M, et al (2011) Arsenic: Toxicity, oxidative stress and human disease. J. Appl. Toxicol. 31

Kamburov A, Wierling C, Lehrach H, Herwig R (2009) ConsensusPathDB - A database for integrating human functional interaction networks. Nucleic Acids Res 37:. https://doi.org/10.1093/nar/gkn698

Kanehisa M, Sato Y, Kawashima M, et al (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:. https://doi.org/10.1093/nar/gkv1070

Kültz D (2005) Molecular and evolutionary basis of the cellular stress response. Annu. Rev. Physiol. 67

Kutmon M, van Iersel MP, Bohler A, et al (2015) PathVisio 3: An Extendable Pathway Analysis Toolbox. PLoS Comput Biol 11:1–13. https://doi.org/10.1371/journal.pcbi.1004085

Ladiwala U, Li H, Antel JP, Nalbantoglu J (1999) p53 Induction by tumor necrosis factor-α and involvement of p53 in cell death of human oligodendrocytes. J Neurochem 73:. https://doi.org/10.1046/j.1471-4159.1999.0730605.x

Li C, Xu J, Li F, et al (2011) Unfolded protein response signaling and MAP kinase pathways underlie pathogenesis of arsenic-induced cutaneous inflammation. Cancer Prev Res 4:. https://doi.org/10.1158/1940-6207.CAPR-11-0343

Lichtlen P, Schaffner W (2001) The “metal transcription factor” MTF-1: Biological facts and medical implications. Swiss Med. Wkly. 131

Lin JH, LaVail MM (2010) Misfolded proteins and retinal dystrophies. In: Advances in Experimental Medicine and Biology

Lin WC, Chuang YC, Chang YS, et al (2012) Endoplasmic reticulum stress stimulates p53 expression through NF-κB activation. PLoS One 7:. https://doi.org/10.1371/journal.pone.0039120

Lindholm D, Korhonen L, Eriksson O, Kõks S (2017) Recent insights into the role of unfolded protein response in ER stress in health and disease. Front. Cell Dev. Biol. 5

Liu L, Trimarchi JR, Navarro P, et al (2003) Oxidative stress contributes to arsenic-induced telomere attrition, chromosome instability, and apoptosis. J Biol Chem 278:. https://doi.org/10.1074/jbc.M303553200

Liu L, Zhao Y fu, Han W hao, et al (2016) Protective effect of antioxidant on renal damage caused by doxorubicin chemotherapy in mice with hepatic cancer. Asian Pac J Trop Med 9:. https://doi.org/10.1016/j.apjtm.2016.08.003

Lombardi A, Inabnet WB, Owen R, et al (2015) Endoplasmic reticulum stress as a novel

mechanism in amiodaroneinduced destructive thyroiditis. J Clin Endocrinol Metab 100:. https://doi.org/10.1210/jc.2014-2745

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Luciani R, Falcone C, Principe F, et al (2009) Acute renal failure due to amiodarone-induced hypothyroidism. Clin Nephrol 72:. https://doi.org/10.5414/cnp72079

Ma LJ, Fogo AB (2009) PAI-1 and kidney fibrosis. Front Biosci 14:. https://doi.org/10.2741/3361

Martens M, Ammar A, Riutta A, et al (2021) WikiPathways: Connecting communities. Nucleic Acids Res 49:. https://doi.org/10.1093/nar/gkaa1024

Martens M, Verbruggen T, Nymark P, et al (2018) Introducing WikiPathways as a Data-Source to Support Adverse Outcome Pathways for Regulatory Risk Assessment of Chemicals and Nanomaterials. Front Genet 9:1–9. https://doi.org/10.3389/fgene.2018.00661

Miller-Hodges E, Hohenstein P (2012) WT1 in disease: Shifting the epithelial-mesenchymal balance. J. Pathol. 226

Mohammed-Ali Z, Carlisle RE, Nademi S, Dickhout JG (2017) Animal Models of Kidney Disease. In: Animal Models for the Study of Human Disease: Second Edition

Mondal D, Mathur A, Chandra PK (2016) Tripping on TRIB3 at the junction of health, metabolic dysfunction and cancer. Biochimie 124

Nicoletti-Carvalho JE, Nogueira TCA, Gorjão R, et al (2010) UPR-mediated TRIB3 expression correlates with reduced AKT phosphorylation and inability of interleukin 6 to overcome palmitate-induced apoptosis in RINm5F cells. J Endocrinol 206:. https://doi.org/10.1677/JOE-09-0356

Nieto MA, Huang RYYJ, Jackson RAA, Thiery JPP (2016) EMT: 2016. Cell 166

Park SH, Choi MJ, Song IK, et al (2007) Erythropoietin decreases renal fibrosis in mice with ureteral obstruction: Role of inhibiting TGF-β-induced epithelial-to-mesenchymal transition. J Am Soc Nephrol 18:. https://doi.org/10.1681/ASN.2005080866

Paul MK, Kumar R, Mukhopadhyay AK (2008) Dithiothreitol abrogates the effect of arsenic trioxide on normal rat liver mitochondria and human hepatocellular carcinoma cells. Toxicol Appl Pharmacol 226:. https://doi.org/10.1016/j.taap.2007.09.020

Petersen M, Thorikay M, Deckers M, et al (2008) Oral administration of GW788388, an inhibitor of TGF-β type I and II receptor kinases, decreases renal fibrosis. Kidney Int

73:. https://doi.org/10.1038/sj.ki.5002717

Qi W, Chen X, Gilbert RE, et al (2007) High glucose-induced thioredoxin-interacting protein in renal proximal tubule cells is independent of transforming growth factor-β1. Am J Pathol 171:. https://doi.org/10.2353/ajpath.2007.060813

Ramadan D, Rancy PC, Nagarkar RP, et al (2009) Arsenic(III) species inhibit oxidative protein folding in Vitro. Biochemistry 48:. https://doi.org/10.1021/bi801988x

Rashid HO, Yadav RK, Kim HR, Chae HJ (2015) ER stress: Autophagy induction, inhibition and selection. Autophagy 11:. https://doi.org/10.1080/15548627.2015.1091141

SaferWorldbyDesign accessed November 2020, < https://saferworldbydesign.com>

Sakairi T, Abe Y, Kopp JB (2011) TGF-beta1 reduces Wilms' tumor suppressor gene expression in podocytes. Nephrol Dial Transplant 26:. https://doi.org/10.1093/ndt/gfr061

Sardão VA, Oliveira PJ, Holy J, et al (2009) Doxorubicin-induced mitochondrial dysfunction is secondary to nuclear p53 activation in H9c2 cardiomyoblasts. Cancer Chemother Pharmacol 64:. https://doi.org/10.1007/s00280-009-0932-x

Saydam N, Adams TK, Steiner F, et al (2002) Regulation of metallothionein transcription by the metal-responsive transcription factor MTF-1: Identification of signal transduction cascades that control metal-inducible transcription. J Biol Chem 277:. https://doi.org/10.1074/jbc.M110631200

Simmons SO, Fan CY, Ramabhadran R (2009) Cellular stress response pathway system as a sentinel ensemble in toxicological screening. Toxicol. Sci. 111

Singh P, Chandrasekaran V, Hardy B, et al (2021) Temporal transcriptomic alterations of cadmium exposed human iPSC-derived renal proximal tubule-like cells. Toxicol Vitr 76:. https://doi.org/10.1016/j.tiv.2021.105229

Slenter DN, Kutmon M, Hanspers K, et al (2018) WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res 46:. https://doi.org/10.1093/nar/gkx1064

Snijders KE, Fehér A, Táncos Z, et al (2021) Fluorescent tagging of endogenous Heme oxygenase-1 in human induced pluripotent stem cells for high content imaging of oxidative stress in various differentiated lineages. Arch Toxicol 95:3285–3302. https://doi.org/10.1007/s00204-021-03127-8

Terrab L, Wipf P (2020) Hsp70 and the Unfolded Protein Response as a Challenging Drug Target and an Inspiration for Probe Molecule Development. ACS Med. Chem. Lett. 11

Tsui KH, Hsu SY, Chung LC, et al (2015) Growth differentiation factor-15: A p53- and demethylation-upregulating gene represses cell proliferation, invasion, and

tumorigenesis in bladder carcinoma cells. Sci Rep 5:. https://doi.org/10.1038/srep12870

Vihervaara A, Duarte FM, Lis JT (2018) Molecular mechanisms driving transcriptional stress responses. Nat. Rev. Genet. 19

Vineetha VP, Soumya RS, Raghu KG (2015) Phloretin ameliorates arsenic trioxide induced mitochondrial dysfunction in H9c2 cardiomyoblasts mediated via alterations in membrane permeability and ETC complexes. Eur J Pharmacol 754:. https://doi.org/10.1016/j.ejphar.2015.02.036

Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. Nature 408:. https://doi.org/10.1038/35042675

Wellens S, Dehouck L, Chandrasekaran V, et al (2021) Evaluation of a human iPSC-derived BBB model for repeated dose toxicity testing with cyclosporine A as model compound. Toxicol Vitr 105112. https://doi.org/https://doi.org/10.1016/j.tiv.2021.105112

Weng CY, Chiou SY, Wang L, et al (2014) Arsenic trioxide induces unfolded protein response in vascular endothelial cells. Arch Toxicol 88:. https://doi.org/10.1007/s00204-013-1101-x

Wigner P, Zielinski K, Michlewska S, et al (2021) Disturbance of cellular homeostasis as a molecular risk evaluation of human endothelial cells exposed to nanoparticles. Sci Rep 11:. https://doi.org/10.1038/s41598-021-83291-0

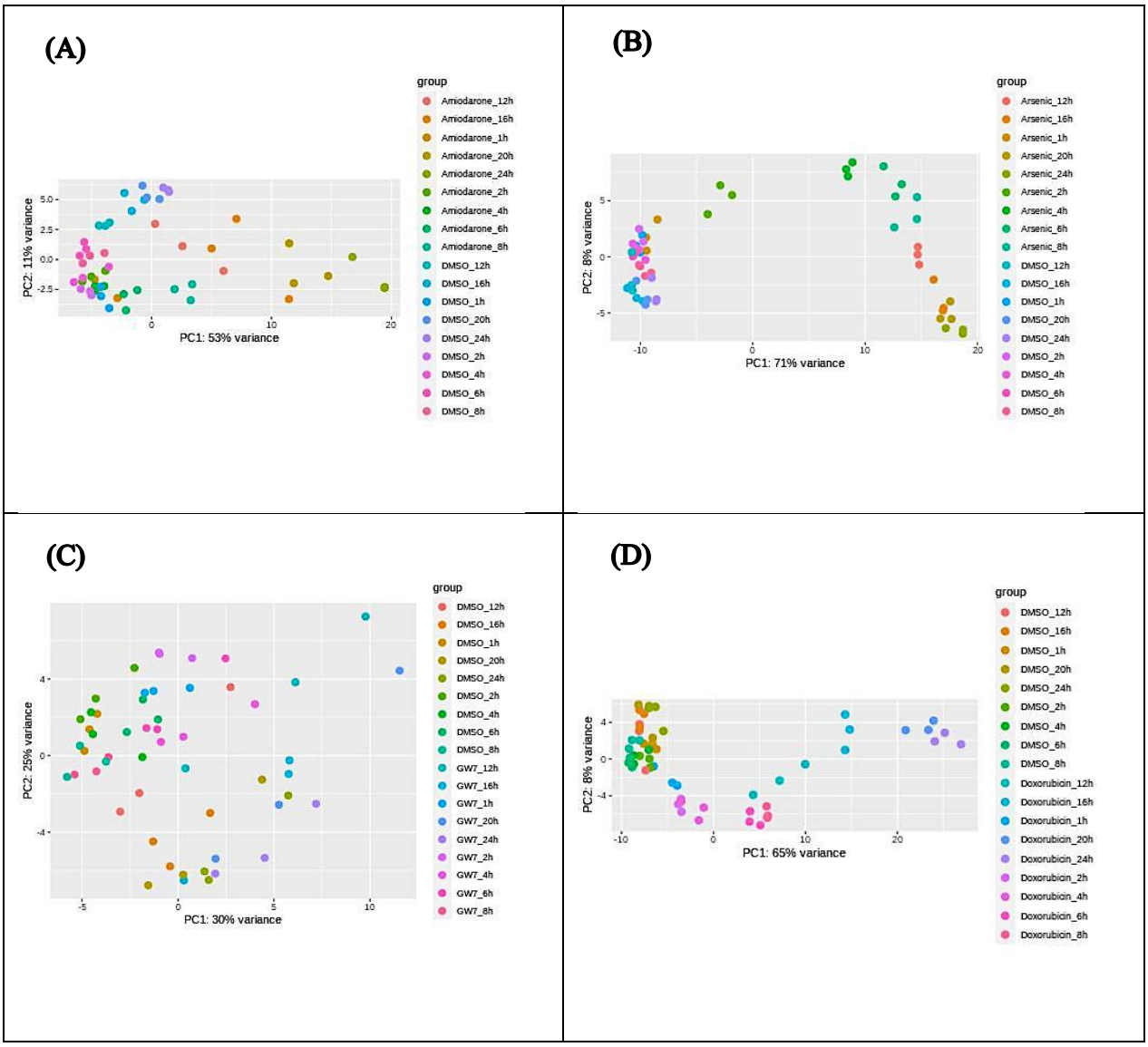
Xu J, Zhang W, Tang L, et al (2018) Epithelial-mesenchymal transition induced PAI-1 is associated with prognosis of triple-negative breast cancer patients. Gene 670:. https://doi.org/10.1016/j.gene.2018.05.089

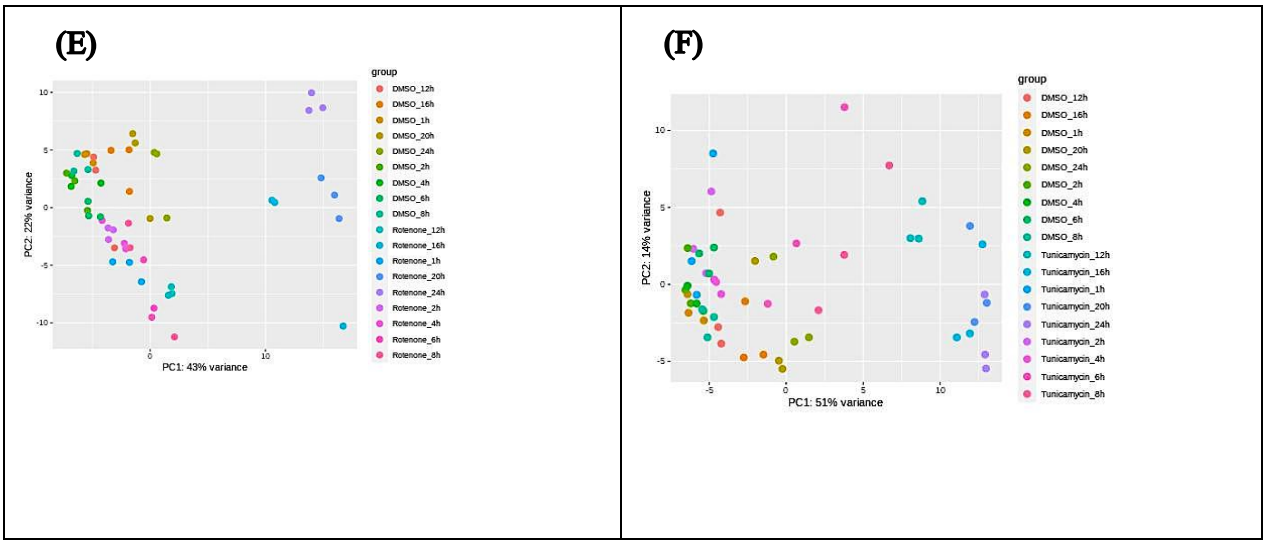
Yan C, Boyd DD (2006) ATF3 regulates the stability of p53: A link to cancer. Cell Cycle 5

Yin L, Yu X (2018) Arsenic-induced apoptosis in the p53-proficient and p53-deficient cells through differential modulation of NFkB pathway. Food Chem Toxicol 118:. https://doi.org/10.1016/j.fct.2018.06.053

Young SK, Shao Y, Bidwell JP, Wek RC (2016) Nuclear matrix protein 4 is a novel regulator of ribosome biogenesis and controls the unfolded protein response via repression of Gadd34 expression. J Biol Chem 291:. https://doi.org/10.1074/jbc.M116.729830

Supporting Material:





Supporting Information 1: PCA plots for A) Amiodarone B) Arsenic C) GW788388 D) Doxorubicin E) Rotenone F) Tunicamycin

4. An in vitro strategy using multiple human induced pluripotent stem cell-derived models to assess the toxicity of chemicals: a case study on paraquat

Carolina Nunes[1,2*], Pranika Singh[3,4*], Zahra Mazidi[5,6*], Cormac Murphy[7*], Aurore Bourguignon[8,9], Sara Wellens[10], Vidya Chandrasekaran[7], Sreya Ghosh[11], Melinda Zana[8], David Pamies[1,2], Aurélien Thomas[12,13], Catherine Verfaillie[11], Maxime Culot[10], Andras Dinnyes[8,9,14], Barry Hardy[3], Anja Wilmes[7], Paul Jennings[7], Regina Grillari[5], Johannes Grillari[5,6,15], Marie-Gabrielle Zurich [1,2,$,#] and Thomas Exner[3,16,$,#]

1. Department ofBiomedical Sciences, University of Lausanne, Rue du Bugnon 7, 1005 Lausanne, Switzerland
2. Swiss Centre for Applied Human Toxicology (SCAHT), University of Basel Missionsstrasse 64, 4055 Basel, Switzerland
3. Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, 4057 Basel, Switzerland
4. Division of Molecular and Systems Toxicology, Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland
5. Evercyte GmbH, Vienna, Austria
6. Institute of Molecular Biotechnology, Department of Biotechnology, BOKU - University of Natural Resource and Life science (BOKU), Vienna, Austria
7. Division of Molecular and Computational Toxicology, Department of Chemistry and Pharmaceutical Sciences, AIMMS, Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, The Netherlands.
8. BioTalentum Ltd, Gödöllő, Hungary
9. Department of Physiology and Animal Health, Institute of Physiology and Animal Health, Hungarian University of Agriculture and Life Sciences, Gödöllõ, Hungary
10. University of Artois, UR 2465, Laboratoire de la Barrière Hémato-Encéphalique (LBHE), Faculté des sciences Jean Perrin, Rue Jean Souvraz SP18, F-62300 Lens, France
11. Department of Development and Regeneration, Stem Cell Institute, KU Leuven, Leuven, Belgium
12. Unit of Forensic Toxicology and Chemistry, CURML, Lausanne and Geneva University Hospitals, Geneva, Switzerland.
13. Faculty Unit of Toxicology, CURML, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland.
14. Department of Cell Biology and Molecular Medicine, University of Szeged, Szeged, Hungary
15. Ludwig Boltzmann Institute for Traumatology Research Center in cooperation with AUVA Vienna, Austria
16. Seven Past Nine d.o.o., Hribljane 10, 1380 Cerknica, Slovenia

* shared first authorship
$ shared senior authorship # corresponding authors

Corresponding authors:

Thomas Exner
+49 171 3807485 thomas.exner@sevenpastnine.com

Marie-Gabrielle Zurich
+41 21 692 55 42 mzurich@unil.ch

Contribution:

- Planned and proposed the study the effects of paraquat on different iPSC cells
- Quality controlled, analysed, visualized, and interpreted the data.
- Drafted the manuscript
- Worked in alignment with the other shared first author in the revision process

Toxicology in Vitro 81 (2022) 105333

Contents lists available at ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

An *in vitro* strategy using multiple human induced pluripotent stem cell-derived models to assess the toxicity of chemicals: A case study on paraquat

Carolina Nunes [a,b,1], Pranika Singh [c,d,1], Zahra Mazidi [e,f,1], Cormac Murphy [g,1], Aurore Bourguignon [h,i], Sara Wellens [j], Vidya Chandrasekaran [g], Sreya Ghosh [k], Melinda Zana [h], David Pamies [a,b], Aurélien Thomas [l,m], Catherine Verfaillie [k], Maxime Culot [j], Andras Dinnyes [h,i,n], Barry Hardy [c], Anja Wilmes [g], Paul Jennings [g], Regina Grillari [e], Johannes Grillari [e,f,o], Marie-Gabrielle Zurich [a,b,*,2], Thomas Exner [c,p,*,2]

[a] Department of Biomedical Sciences, University of Lausanne, Rue du Bugnon 7, 1005 Lausanne, Switzerland
[b] Swiss Centre for Applied Human Toxicology (SCAHT), University of Basel, Missionsstrasse 64, 4055 Basel, Switzerland
[c] Edelweiss Connect GmbH, Technology Park Basel, Hochbergerstrasse 60C, 4057 Basel, Switzerland
[d] Division of Molecular and Systems Toxicology, Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland
[e] Evercyte GmbH, Vienna, Austria
[f] Institute of Molecular Biotechnology, Department of Biotechnology, BOKU - University of Natural Resource and Life science (BOKU), Vienna, Austria
[g] Division of Molecular and Computational Toxicology, Department of Chemistry and Pharmaceutical Sciences, AIMMS, Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, the Netherlands
[h] BioTalentum Ltd, Gödöllő, Hungary
[i] Department of Physiology and Animal Health, Institute of Physiology and Animal Nutrition, Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary
[j] University of Artois, UR 2465, Laboratoire de la Barrière Hémato-Encéphalique (LBHE), Faculté des sciences Jean Perrin, Rue Jean Souvraz SP18, F-62300 Lens, France
[k] Department of Development and Regeneration, Stem Cell Institute, KU Leuven, Leuven, Belgium
[l] Unit of Forensic Toxicology and Chemistry, CURML, Lausanne and Geneva University Hospitals, Geneva, Switzerland
[m] Faculty Unit of Toxicology, CURML, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland
[n] Department of Cell Biology and Molecular Medicine, University of Szeged, Szeged, Hungary
[o] Ludwig Boltzmann Institute for Traumatology Research Center in cooperation with AUVA, Vienna, Austria
[p] Seven Past Nine d.o.o., Hribljane 10, 1380 Cerknica, Slovenia

ARTICLE INFO

Editor: Dr. P. Jennings

Keywords:
Human induced pluripotent stem cells
Paraquat
New approach methodology
Toxicology
Acute toxicity

ABSTRACT

Most OECD guidelines for chemical risk assessment include tests performed on animals, raising financial, ethical and scientific concerns. Thus, the development of human-based models for toxicity testing is highly encouraged. Here, we propose an *in vitro* multi-organ strategy to assess the toxicity of chemicals. Human induced pluripotent stem cells (hiPSCs)-derived models of the brain, blood-brain barrier, kidney, liver and vasculature were generated and exposed to paraquat (PQ), a widely employed herbicide with known toxic effects in kidneys and brain. The models showed differential cytotoxic sensitivity to PQ after acute exposure. TempO-Seq analysis with a set of 3565 probes revealed the deregulation of oxidative stress, unfolded protein response and estrogen receptor-mediated signaling pathways, in line with the existing knowledge on PQ mechanisms of action. The main advantages of this strategy are to assess chemical toxicity on multiple tissues/organs in parallel, exclusively in human cells, eliminating the interspecies bias, allowing a better evaluation of the differential sensitivity of the

Abbreviations: 3Rs, Replacement, Reduction and Refinement of Animal Experiments; BBB, Blood-brain barrier; BLECs, Brain-like endothelial cells; BS, BrainSphere; DEG, Differentially expressed gene; EC, Endothelial cell; ER, Endoplasmic Reticulum; ESC, Embryonic stem cell; ESCs, Embryonic stem cells; ESR, Estrogen Receptor; hiPSC, Human induced pluripotent stem cell; HLC, Hepatocyte-like cell; LFC, log2 fold change; NAM, New approach methodology; NC, Neural cell; PODO, Podocyte; PTL, Proximal Tubular like cell; ROS, Reactive oxygen species; UPR, Unfolded protein response.

* Corresponding authors.
E-mail addresses: mzurich@unil.ch (M.-G. Zurich), thomas.exner@sevenpastnine.com (T. Exner).

[1] Shared first authorship.
[2] Shared senior authorship.

https://doi.org/10.1016/j.tiv.2022.105333
Received 27 August 2021; Received in revised form 4 December 2021; Accepted 11 February 2022
Available online 16 February 2022

models representing the diverse organs, and increasing the chance to identify toxic compounds. Furthermore, although we focused on the mechanisms of action of PQ shared by the different models, this strategy would also allow for organ-specific toxicity testing, by including more cell type-specific probes for TempO-Seq analyses. In conclusion, we believe this strategy will participate in the further improvement of chemical risk assessment for human health.

1. Introduction

The need to develop new approach methodologies (NAMs) providing mechanistic understanding for toxicology testing is now well recognized, and governmental agencies worldwide have been pushing towards the development of human-based assays for human safety assessment (EUR-Lex, 2010; National Research Council, 2007; Oredsson et al., 2019; Sewell et al., 2017). For instance, the US Environmental Protection Agency plans to eliminate animal testing from regulatory requirements for pesticides and industrial chemicals by 2035 (EPA, 2019). However, the move from animal experimentation to NAMs highly depends on the ability of the NAMs to mirror human rather than animal physiology (Herrmann et al., 2019). Since their development, a little over a decade ago (Takahashi et al., 2007), human induced pluripotent stem cells (hiPSCs) have represented a potential paradigm shift across the biological sciences from developmental biology to regenerative medicine. For the field of toxicology, they have been seen as a valuable new tool to enable NAMs to achieve the ideals of the 3Rs through the production of *in vitro* models with human relevant phenotypes that can replace animal testing (Steimberg et al., 2020). While they share capacity for self-renewal and pluripotent potential of embryonic stem cells (ESCs), their generation from adult somatic cells come with a range of advantages while avoiding some of the controversies associated to ESCs (Liu et al., 2017). The relative ease by which hiPSCs can be generated from cells obtained with minimal discomfort on the part of the donor, from sources such as blood, skin and even urine (Zhou et al., 2011), has allowed for the establishment of large cell banks. StemBANCC is one such bank that generated hiPSCs lines from 500 healthy and diseased individuals, and most of these well described lines are now widely available for drug and chemical screening (Morrison et al., 2015) through the European Bank For Induced Pluripotent Stem Cells (EBiSC).

While traditionally, cancer or immortalised cell lines, and to a lesser extent, primary cells, have been the main cell sources for *in vitro* chemical safety assessment, each have their own drawbacks that hiPSCs can potentially overcome. Human primary cells, that would be the most adequate cells to use, can be difficult to obtain and are only available in a limited quantity for a limited time, while undifferentiated hiPSCs can, in theory, proliferate indefinitely. Immortalised cell lines, once made, can be continuously available; however, these models often lack some important biological aspects of their unaltered counterparts leading to concerns that they may be poor models for healthy human cells and lack the sensitivity for toxicity testing (Liu et al., 2017). While hiPSCs have their own challenges, chief among them being the establishment of robust and repeatable differentiation protocols that produce the desired cell type, once this is done, they are a renewable source of cells that can express many important intrinsic phenotypic features often lost in traditional cell lines. For example, a recently published hiPSC differentiation protocol produces kidney proximal tubular-like cells that, alongside other characteristic markers, expressed megalin (LRP2), an apical transporter that plays a key role in reabsorption and is absent in most available human immortalised cell lines (Chandrasekaran et al., 2021).

With the establishment of reliable differentiation protocols for more cell types, and with the emergence of complex culture systems, the possibility has emerged to use hiPSC-derived cells to model the impact of toxins on the different cell types/organs from the same human donor. In this study, we hypothesized that an *in vitro* multi-organ human iPSC-derived strategy will allow the reliable assessment of the toxicity of chemicals. To test this, hiPSCs-derived models of the brain (2D and 3D mixed-cell cultures), blood-brain barrier (brain capillary endothelial cells), kidney (podocytes and proximal tubular cells), liver (hepatocyte) and vasculature (endothelial cells) were generated from two donors, using published or newly developed protocols, and were exposed to paraquat (PQ), a widely employed herbicide used as a reference compound chosen for its known toxic effects on a range of tissues across the body (Cicchetti et al., 2009; Dinis-Oliveira et al., 2008; Gao et al., 2020). TempO-Seq analysis showed the disruption of "Oxidative stress induced gene expression via Nrf2 markers", as expected since oxidative stress is the well-known PQ mechanism of action, but also of two other pathways previously demonstrated to be involved in PQ toxicity, "Unfolded protein response" and "ESR-mediated signaling", suggesting this strategy as a future tool for prioritization, screening of chemicals and risk assessment.

2. Materials and method

2.1. Cell cultures

Human iPSCs used in this study (SBAD3 clone 1 and SBAD2 clone 1) were generated in the IMI-funded StemBANCC project (http://stembancc.org) (Morrison et al., 2015) from purchased primary fibroblasts of non-diseased donors (Lonza). They were cultured on Matrigel® Growth Factor Reduced (GFR) (Corning) or Geltrex® (Gibco) in mTeSR™1 (StemCell Technology) feeder-free medium at 37 °C in a humidified 5% CO_2 incubator. Medium was replaced every day and cells were passaged every 4–5 days using Versene® according to the supplier's protocol (Thermo Fisher Scientific).

Human iPSCs were differentiated into different cell types to obtain *in vitro* models for several tissues/organs. Details on methodologies and characterization are found in the references given for each model, in Chandrasekaran et al. (in preparation – for this special issue), as well as in Table 1. BrainSpheres (BS), a 3D model containing neurons, astrocytes and oligodendrocytes, were prepared from neuroprogenitor cells (NPCs) derived from the SBAD3 clone 1 using a Gibco protocol (Nunes and Zurich, 2020; Pamies et al., 2017). Neural cells (NC), a 2D model containing neurons and astrocytes were generated from NPCs derived from the SBAD2 clone 1 using dual SMAD inhibition protocol (Chambers et al., 2009; Chandrasekaran et al., 2017; Lo Giudice et al., 2019; Ochalek et al., 2017). The brain-like endothelial cells (BLECs) transwell model was prepared from the SBAD3 clone 1 (Sevin et al., 2019; Wellens et al., 2021). The kidney cells, podocytes (PODO) (Murphy et al., 2019; Rauch et al., 2018) and proximal tubular like cells (PTL) (Chandrasekaran et al., 2021) were derived from the SBAD2 clone 1 containing a Green Fluorescent Protein (GFP) tag downstream of one allele of the HMOX-1 gene (SBAD2 clone 1 HMOX1-eGFP, prepared by Prof. Dinnyes's laboratory, Biotalentum, Hungary), although this fluorescent property was not utilised in this study. The endothelial cells (EC) (Gholami et al., 2021) representing the vascular system and the hepatocyte-like cells (HLC) (Boon et al., 2020) representing the liver were derived from SBAD2 clone 1.

2.2. Paraquat exposure

PQ (Sigma–Aldrich, catalogue No 36541, lot #BCBW5264, CAS #75365-73-0) was dissolved in ultrapure water. The different cell cultures were exposed to PQ for 24 h and BS was also exposed for 48 h, after

Table 1
Characterization overview of the different models used.

Name	Organ	References	Time in culture before exposure	Characterization							Functional Assay	
				Cell Type	Marker	Technique (Ref at protocol DOI)					Assay	Observations
						RT-qPCR	Immunocytochemistry	FACS	Western-blot	TempO-Seq		
BS	Brain	[1]. Pamies et al., 2017	42d	Neurons	ACHE					x	Multi-Electrode Array MEA (Pamies et al., 2017)	Electrical activity
					B3T					x		
					GABRA1	x [1]				x		
					GAD1	x [1]				x		
					GRIN1	x [1]				x		
					MAP2		x [1]			x		
					NFH		x [1]			x		
					SYN	x [1]	x [1]			x		
					SYP					ND		
					TH	x [1]	x [1]			x		
					TUJ1		x [1]			NP		
					VGLUT1		x [1]			x		
				Astrocytes	GFAP	x [1]	x [1]			x		
					S100	x [1]				ND		
					VIM					x		
				Oligodendrocytes	CNPase		x [1]			NP		
					MBP	x [1]	x [1]			x		
					NOGOA		x [1]			NP		
					O1	–	x [1]			NP		
					O2	–	x [1]			NP		
					O4	–				NP		
					OLIG2	x[1]				x		
NC	Brain	[1]. Ochalek et al., 2017, [2]. Chandrasekaran et al., 2017	21d	Neurons	TUBB3	x[2]	x [1, 2]			x	Calcium current measurements (Lo Giudice et al., 2019) and patch clamp (Chandrasekaran et al., 2017)	Electrical activity at later time-points
					MAP2	x[2]	x [1, 2]			x		
					NF200		x [1]			NP		
				NPC	Ki67	x[2]	x[2]			NP		
					SOX1	x[2]	x[2]	x[2]		x		
					NESTIN	x[2]	x [1, 2]	x [1, 2]		NP		
					PAX6	x[2]	x [1, 2]	x [1, 2]		NP		
					SOX9	x[2]		x[2]		NP		
				Astrocytes	GFAP	x[2]	x[2]			x		
					AQP4	x[2]	x[2]			NP		
BLECs	BBB	[1]. Wellens et al., 2021	8d	Brain-like endothelial cells	ABCG2					x 1	Pump-out assay (Wellens et al., 2021 and Sevin et al., 2019)	Reduced Rhodamine123 extrusion under influence from elacridar and verapamil efflux pump inhibitors.
					ABCB1					x [1]		
					CDH5		x [1]			x [1]		
					CLDN5		x [1]			ND		
					vWF					x [1]		
					CD31		ND			ND		
					ABCC4					x [1]		
					ABCC5					x [1]		
					SLC25A46					x [1]		
					SLC26A2					x [1]		
					SLC27A2					x [1]		
					SLC27A3					x [1]		
					SLC2A1					x [1]		
					SLC35A3					x [1]		
					SLC7A5					x [1]		
EC	Vasculature	[1]. Gholami et al., 2021	8d	Endothelial cells	TBXT	x [1]				NP	Matrigel assay	Endothelial cells could form tubes at one of the beginning stages of angiogenesis
					CD31	x [1]	x [1]	X [1]		x		
					CDH5	x [1]	x [1]	x [1]		NP		

(*continued on next page*)

Table 1 (*continued*)

Name	Organ	References	Time in culture before exposure	Characterization		Technique (Ref at protocol DOI)					Functional Assay	
				Cell Type	Marker	RT-qPCR	Immunocytochemistry	FACS	Western-blot	TempO-Seq	Assay	Observations
					VEGFR2			x [1]		NP		
					vWF		x [1]			ND		
PTL	Kidney	[1]. Chandrasekaran et al., 2021	14d	Proximal tubule-like cells	LRP2		x [1]		x [1]	NP	Parathyroid hormone (PTH) response assay; Calcein-AM efflux assay; Albumin uptake assay for megalin facilitated endocytosis	Increase levels of cAMP in response to PTH; Functional P-gp efflux; uptake of labelled albumin
					TJP3		x [1]			NP		
					PTH1R					NP		
					FBP1					NP		
PODO	Kidney	[1]. Rauch et al., 2018, [2]. Murphy et al., 2019	12-14d	Podocytes	WT1		x [1,2]		x [1]	ND	Albumin uptake assay	Uptake of albumin by podocyte-like cells
					SYNPO	x [1]	x [1,2]		x [1]	NP		
					NPHS1		x [1]		x [1]	NP		
					NPHS2		x [1]		x [1]	NP		
HLC	Liver	[1]. Ghosh et al. (under revision for this special issue)	42d	Hepatocyte-like cells	ALB	x [1]				x	Albumin secretion assay; CYP3A4 activity assay	Albumin secretion was observed at about 40% of that of primary hepatocytes plated for 12 h. CYP3A4 activity was observed (about one-third that of primary hepatocytes).
					AFP	x [1]	x [1]			x		
					HNF4A	x [1]	x [1]			x		
					CYP3A4	x [1]	x [1]			x		
					CYP2C9	x [1]				x		
					G6PC	x [1]				x		
					PEPCK	x [1]				NP		
					SLC10A1	x [1]				x		
					SERPINA1					x		

NT: Non tested; x: detected; ND: not detected; NP: not present in the TempO-Seq panel; Superscript number: refers to publication listed under references column.

various duration in culture, corresponding to a maturation state defined by each lab. BS was exposed to PQ at day *in vitro* (DIV)42, NC at DIV21, BLECs at DIV8, PODO at DIV12-14, PTL at DIV14, EC at DIV8, and HLC at DIV40. Concentrations of paraquat are given in figure or figure captions (Figs. 2, 3 and Table 2). For cytotoxicity testing, 2–9 samples were run per group, and for TempO-Seq analysis 2–16 samples were run per group. Exposure of the cell models to PQ was conducted in 6 different labs, NC at BioTalentum Ltd. (Hungary), BS at University of Lausanne (Switzerland), PODO and PTL at Vrije Universiteit Amsterdam (Netherlands), EC at Evercyte GmbH (Austria), HLC at Katholieke Universiteit Leuven (Belgium) and BLECs at Université d'Artois (France).

2.3. Cytotoxicity assays

After exposure of the different iPSC-derived cell models to a large range of PQ concentrations, viability assays were run in order to establish cytotoxicity concentration-dependent curves.

2.3.1. Resazurin assay

After treatment with PQ, the medium was removed and the cells were incubated with 44 μM resazurin diluted in cell culture medium for a duration indicated in Table 2. The fluorescent product resorufin was measured at 530–560 nm excitation and 590 nm emission using a plate reader. After subtraction of the background (only resazurin solution), results of the treated cultures were expressed as percentage of the mean of the control cultures.

2.3.2. ATP assay

ATP viability assay was performed using the CellTiter-Glo® 3D Cell Viability Assay, according to the manufacturer's protocol (Promega). NC were lysed with 100 μL CellTiter-Glo® 3D Reagent for 90 min at RT. Luminescence signal was recorded with a Thermo VarioScan Flash plate reader (Thermo Fisher Scientific). Results of the treated cultures were expressed as percentage of the mean of the control cultures.

2.4. TempO-Seq sample collection

Samples of the different cell systems were treated with PQ at concentrations chosen on the basis of the previously established cytotoxicity curves (final concentrations and time points in Table 2 and indicated in figure captions). DMSO was added to all samples (final concentration 0.01%) due to the simultaneous testing of water-insoluble chemicals (results not presented in this manuscript). At the end of exposure, cells were collected and lysed with 1× TempO-Seq Lysis Buffer in a ratio of 0.25 to 2 million cells/mL. Lysates were frozen at −80 °C and shipped to BioClavis Technologies Ltd. (Glasgow) on dry ice where the TempO-Seq assay using the EU-ToxRisk v2.1 panel (3565 probes representing 3257 genes) was conducted with standard attenuators. The service also included primary processing to derive gene-annotated raw read counts and quality control, following a previously described procedure (Limonciel et al., 2018; Mav et al., 2018). Each sample FASTQ file was aligned against the TempO-Seq transcriptome using the Bowtie aligner (Li and Durbin, 2009). The output of this analysis generated a table of counts per gene per sample.

2.5. TempO-Seq data analysis and visualization

2.5.1. Sample and probe filtering

The raw read count data of the samples provided by BioClavis was used for differential expression analysis. The raw data and corresponding metadata files were uploaded and maintained on an internal instance of the EdelweissData™ management system (SaferWorldbyDesign2020) for easier accessibility. All samples exposed to PQ and the respective controls from the different hiPSC-derived cell models were extracted using a python workflow as previously described (Singh et al., 2021).

Table 2
Description of all models used in the study and respective IC50/30/20/10 and PQ concentrations used for TempOSeq. BrainSpheres (BS), Neural Cells (NC), Blood-Brain Barrier (BBB), Endothelial cells (EC), renal proximal tubular epithelial-like cells (PTL) podocytes (PODO), hepatocytes-like cells (HLC).

Model	Organ	PQ exposure (h)	Cytotoxicity						TempOSeq
			Assay for assessment	Incubation time (h)	Concentration (μM)				Concentrations (μM)
					IC50	IC30	IC20	IC10	
BS	Brain	24	Resazurin	3	200	161	132	97	-
		48			60	51	45	38	1–2.5–5
NC	Brain	24	ATP	1 h30	49	31	24	16	0.05–0.1
BLECs	BBB	24	Resazurin	2-2 h30	749	654	132	97	50–100–250
EC	Vasculature	24	Resazurin	1 h30	114	55	35	17	1.25–2.5–5
PTL	Kidney	24	Resazurin	1	561	405	327	237	25–100–300
PODO	Kidney	24	Resazurin	2	373	164	97	45	5–50–100
HLC	Liver	24	Resazurin	2	–	–	–	–	12.5–25–50–100
		48			1093	797	590	376	-

Before downstream analyses (Fig. 1), samples and probes were quality controlled. The samples were checked as described elsewhere (Singh et al., 2021) by investigating a) the total number of probe read counts per sample with a threshold of 50,0000 under which samples were removed and b) the Pearson's correlation coefficient among replicates using Morpheus online tool with a threshold of 0.80 under which samples were removed. Concerning the probes, only the ones with median of raw read counts higher than 5 across all conditions for each model were used.

2.5.2. Differential expression analysis

The filtered raw data was analysed for differential expression using the R based library DESeq2 (Love et al., 2014). Samples underwent an r-log transformation and posterior unsupervised clustering using standard DeSeq2 suggested settings to visualize the variance within and between different cell models, as well as between treatment and control groups. Differential expression analysis was performed using DeSeq2 library after normalization of read counts using its standard mean-median ratio method. An automatic script (Singh et al., 2021) was used to generate the statistical values log2fold changes, adjusted *p*-values and base means. To select the most significantly expressed probes in each cell model for further analyses, thresholds were set at: log2 fold changes $> |0.58|$ and adjusted *p*-values <0.05.

2.5.3. Gene-pathway association

ConsensusPathDB (v.34) (http://cpdb.molgen.mpg.de), a web-based meta-database integrating information from around 30 different publicly available databases including WikiPathway, Reactome and KEGG (Kamburov et al., 2013), was used to identify enriched pathways in each cell culture system. The analysis was done using the ConsensusPathDB's

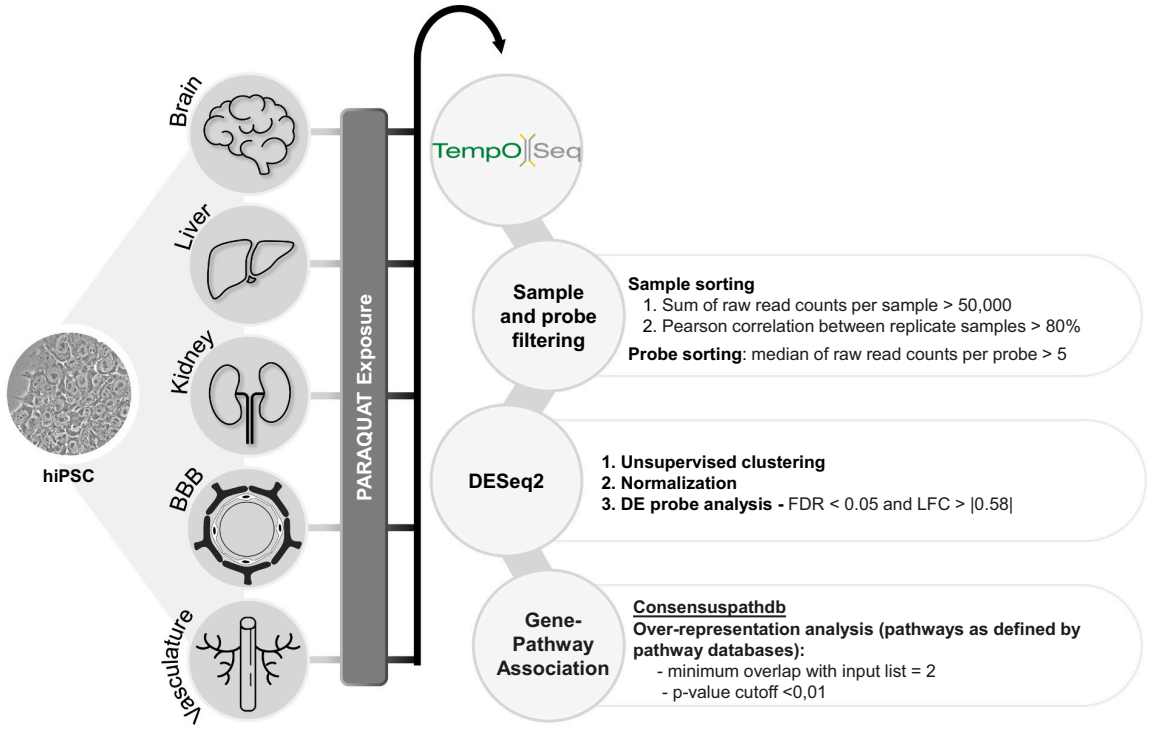


Fig. 1. *In vitro* strategy using multiple hiPSC-derived models to assess the toxicity of chemicals. Workflow used to compare the response of different hiPSC-derived cell models to PQ acute exposure based on TempOSeq data. The 7 models were differentiated from either SBAD2 or SBAD3 hiPSC lines to represent brain, liver, kidney, BBB and vasculature and exposed to different concentrations of PQ chosen on the basis of viability curves. Data from TempOSeq was analysed using R package DSeq2, after filtering and normalization of the raw data. The differentially expressed genes per model were analysed for pathway over-representation with ConsensusPathDB.

over-representation analysis (ORA) and the list of the most significantly differentially expressed genes of each system (taking all concentrations into account) as an input. HUGO gene nomenclature Committee (HGNC) symbols were employed as identifiers. ORA compares input gene list to all the genes associated to each pathway available in the databases. The default Consensus ORA settings require a minimum overlap of 2 genes between the input gene list and a given pathway associated gene list along with a *p*-value cutoff of 0.01 to define a pathway as over-represented. ORA *p*-value calculations are based on the hypergeometric distribution between the input list and existing pathway list (Kamburov et al., 2009).

2.5.4. Gene and pathway intersection analysis

To find deregulated genes and pathways shared by the different cell systems, the lists of differentially expressed genes and enriched pathways from each cell model were compared using UpSetR package (Conway et al., 2017). UpSetR package helps in both visualization and extraction of the different intersections in a matrix format by assembling different set size combinations. Mode was set to "intersect" in the function "make_comb_mat()" and the function "extract_comb()" was then used to obtain the intersections. Finally, intersection plots were created using UpSet() function in the UpSetR package (Conway et al., 2017).

3. Results

3.1. Human iPSC-derived organ-specific models exhibit different sensitivity to paraquat

In order to assess the differential sensitivity of the iPSC-derived cell models to PQ, and to choose sub-cytotoxic concentrations for subsequent experiments to elucidate early pathways of toxicity, the different iPSC-derived cell culture models were exposed to a wide range of PQ concentrations for 24 h and/or 48 h, then concentration-dependent curves for cytotoxicity were established.

The most sensitive model was found to be 2D neural cell cultures (NC), with an IC50 at 49 μM (Fig. 2 and Table 2). BS, the 3D brain cell model, was less sensitive to PQ (IC50: 200 μM after 24 h of exposure; 60 μM after 48 h). The toxicity of PQ to endothelial cells (EC) was found in the same range (IC50: 114 μM), whereas BLECs, used as a model for blood-brain barrier and later referred to as BBB, exhibited a much higher IC50 (749 μM). Among the two cell culture models representing kidney, PTL was less sensitive to PQ than PODO (IC50: 561 μM and 373 μM, respectively). Finally, the hepatocyte cell-like model (HLC) seemed to be the least sensitive model out of the seven models tested; however, the IC50 was determined only at 48 h (1093 μM, cytotoxicity data for HLC also used in Ghosh et al., submitted to Toxics) (Fig. 2 and Table 2).

The results reveal a differential response of the models to the

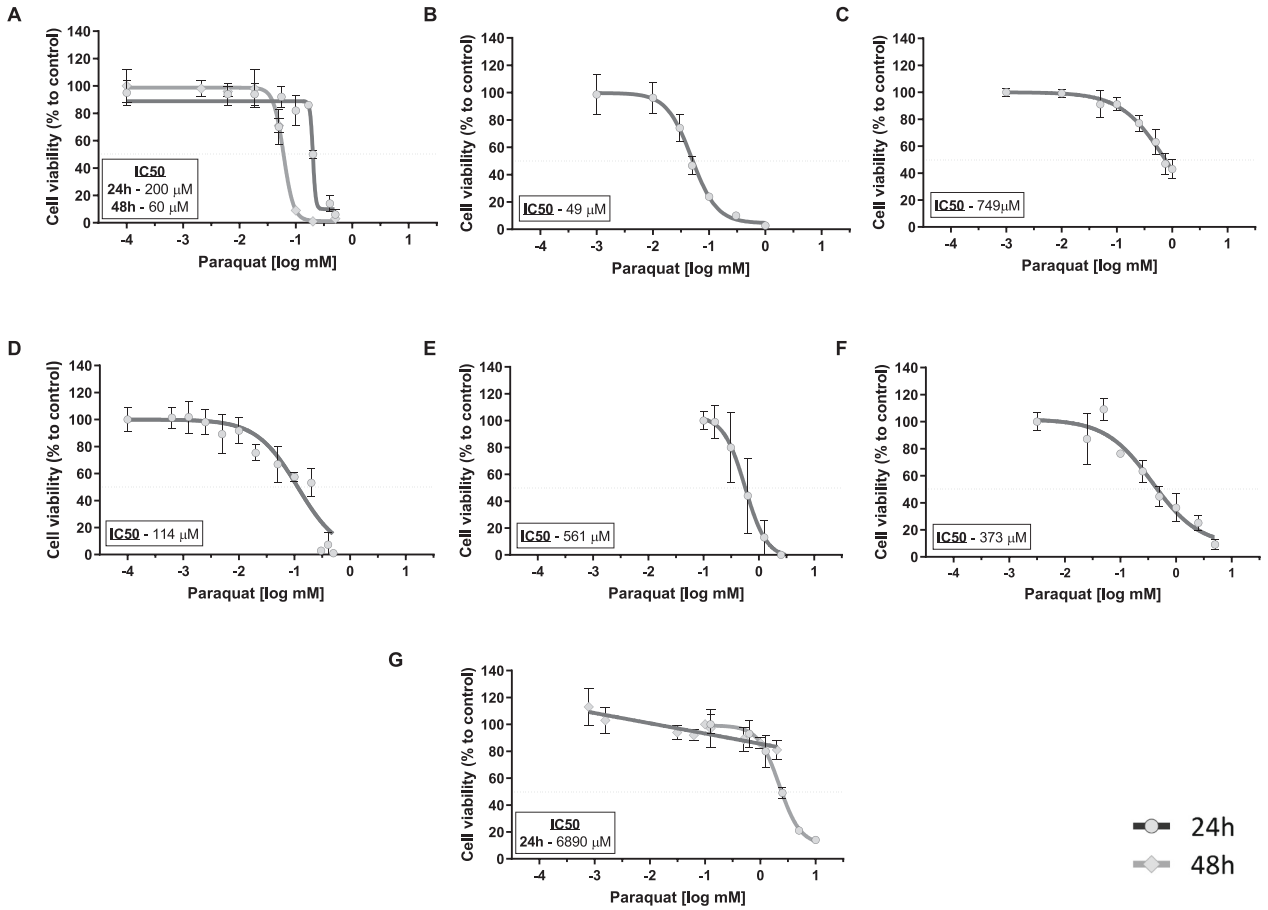


Fig. 2. Models present different sensitivity to PQ exposure. Concentration-dependent curves after 24 h and/or 48 h exposure to PQ: A) BS (0–0,5 mM), B) NC (0,001–1,5 mM), C) BLECs (0–1 mM), D) EC (0–0,5 mM), E) PTL (0–10 mM), (F) PODO (0–5 mM), G) HLC (0–10 mM; data for 48 h also used in Ghosh et al., submitted to Toxics). Results are expressed as % of control cultures a non-linear regression (log (inhibitor) *vs* response - variable slope (four parameters)) was performed in order to calculate the inhibitory concentration (IC) using GraphPad Prism®. Means ±SD of 2–9 samples obtained in 1–3 independent experiments are shown. Black line: 24 h, grey line: 48 h.

cytotoxic effects of PQ, and show that the sensitivity to PQ depend not only on the organ but also on the specific cell type considered (NC *vs* BS, modelling the brain, and PTL *vs* PODO, modelling the kidney). To identify toxicity related pathways induced by PQ, transcriptomic analysis was performed using TempO-Seq as a cost-efficient tool. Each model was exposed to at least one concentration below cytotoxicity IC10 (Table 2).

3.2. *hiPSC-derived models drive sample clustering of TempO-Seq data after paraquat exposure*

All human iPSC-derived models, except EC, exhibited a sum of read counts per sample above 50,000 counts, and therefore passed our first sample quality filter. The second filter, namely the Pearson correlation, led us to the elimination of one replicate of the control group from NC; two control replicates and one replicate of the 25 μM group from PTL; one replicate of treated 250 μM group from BBB and one replicate sample from the 5 μM and from the 100 μM groups from PODO. For HLC (raw data for HLC also used in Ghosh et al., submitted to Toxics) and BS, the samples from all groups showed Pearson correlation ≥0.80 and therefore were all kept for further analyses. For EC model, very low sums of probe read counts (5000–12,000) were found for one replicate from each treatment group and therefore these samples were also extremely poorly correlated to their corresponding replicates (Pearson correlation coefficients between 0.09 and 0.22). After removing the low correlated samples from each cell model, probes with median read counts across all samples of the model considered <5 were filtered out. This caused the removal of 1456 probes from NC, 1208 from PTL, 1445 probes from HLC, 1175 probes from BBB, 1319 probes from PODO and 1517 probes from BS. For the EC model, 64% of the probes (2097) had to be eliminated. Furthermore, although the controls of the EC model passed the 50′000 counts threshold, for undetermined reasons their read counts were much lower than in all the other models (Suppl. Fig. 1), causing a high variance potentially leading to biased normalization and distortion of the differential expression analysis. Based on these results, the data form the EC cell model were considered as unreliable, and the EC model was removed from further analyses.

After these steps of sample and probe filtering, unsupervised PCA of controls and treated samples for the 6 remaining models (Fig. 3A) showed four very distinct clusters. One cluster was formed by PODO and another cluster by HLC samples. A third cluster comprised samples from both brain models (NC and BS), whereas PTL and BBB samples were grouped together. On this PCA plot, BS was the only cell model to show clear separation between treated and control samples; however, the separation is visible when PCA is performed for each model individually, with the exception of PODO (Suppl. Fig. 2). It is remarkable that most of the variance was observed between cell models and less was due to paraquat treatment, even if the list of probes used did not comprise a lot of cell type-specific genes.

3.3. *Paraquat deregulated several identical pathways in the different hiPSC-derived models*

In total, we found 341 probes differentially expressed across the cell models and treatment groups upon PQ exposure. A heatmap of the 100 highest log2 FC probes (Fig. 3B), each one crossing the statistical thresholds in at least one treatment group, with values scaled for each model between the range − 1 to 1 for better relative comparison among models, clearly shows concentration-dependent effects of paraquat in the different models.

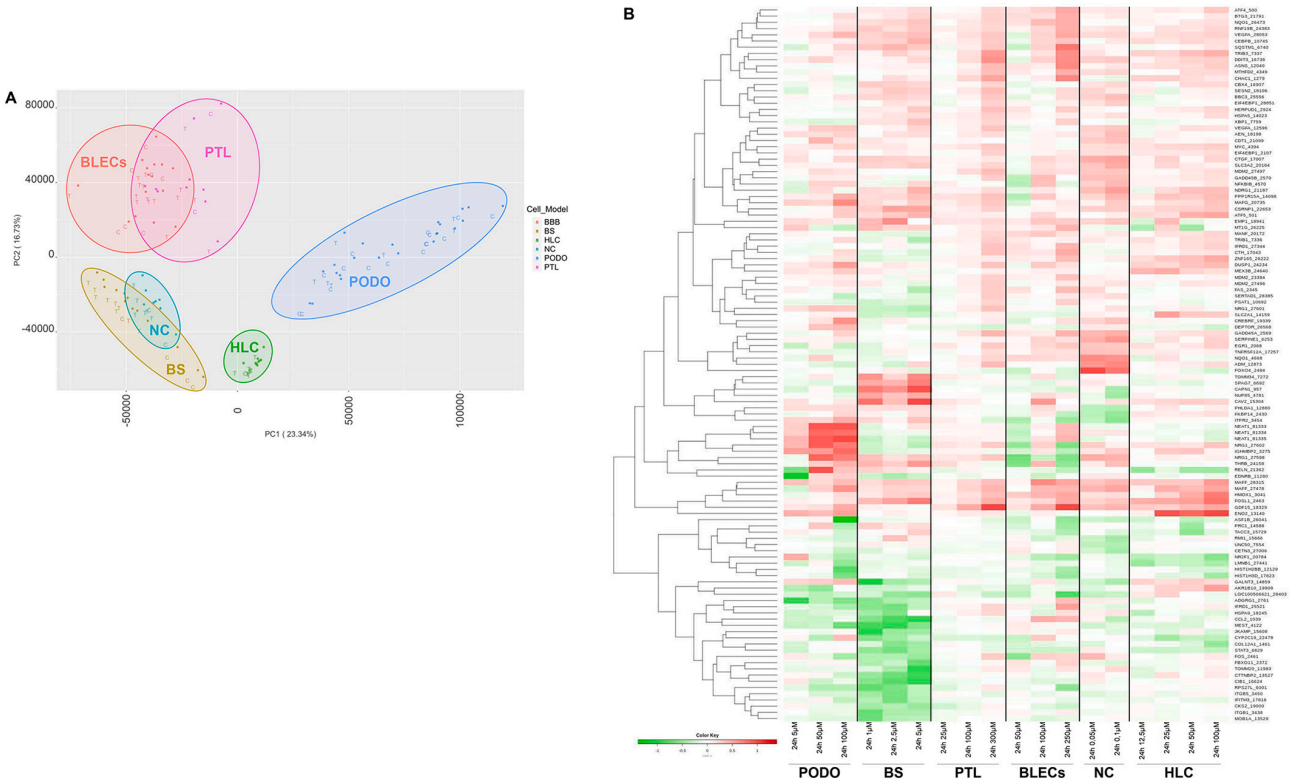


Fig. 3. Clustering of samples and the most deregulated genes after PQ exposure. A) PCA plot of all control and PQ treated samples Each dot represents a sample and each colour represents a cell model. Neuronal Cells (NC, turquoise, 2 samples), Brainsphere (BS, yellow), Podocytes (PODO, blue), Proximal tubule like cells (PTL, purple), Hepatocytes like cells (HLC, green) and Blood Brain Barrier (BLECs, pink). B) A heatmap of the 100 highest log2 FC probes (gene symbol_probe ID), each one crossing the statistical thresholds in at least one treatment group. The relative expression values scaled between −1 (green) and 1 (red).

To investigate the mechanisms of PQ toxicity shared by the different models, pathway enrichment analysis using ConsensusPathDB was performed based on the list of differentially expressed genes (DEG) derived from each model. In total, 36 pathways were identified in HLC, 110 in PODO, 136 in NC, 134 in BBB, 164 in BS and 278 pathways in PTL. The number of pathways given here includes all similar pathways described by different databases aggregated by ConsensusPathDB. Then, UpSetR package was used to find the intersections of these affected pathways in different combinations of 2, 3, 4 and 5 cell models. The pathway intersection graph (Fig. 4) shows that each cell model shared affected pathways with up to four other cell models (Fig. 4, intersections between 5 models). Furthermore, PODO and PTL that are both kidney related cell models shared 18 pathways, whereas BS and NC which are both brain-related cell models shared 17 pathways. No pathway shared by all 6 cell models was found with the criteria used in this study.

In the combinations of 5 cell models, "Oxidative stress induced gene expression via Nrf2" pathway was not found in BS (Table 3). However, read counts for Heme Oxygenase 1 (HMOX1), NAD(*P*)H Quinone Dehydrogenease (NQO1), MAF BZIP Transcription Factor F (MAFF) and G (MAFG) genes, all belonging to the oxidative stress pathway, show a concentration-dependent increase for the six models (Fig. 5), including BS. The other stress pathways found in the combination of 5 models are "Cellular responses to external stimuli" and "Cellular responses to stress" (Table 3). Among combinations of 4 cell models, many other stress pathways were identified, such as "Quercetin and Nf-kB- AP-1 Induced Cell Apoptosis", "Photodynamic therapy-induced NFE2L2 (Nrf2) survival signalling" and "Photodynamic therapy-induced unfolded protein response". Concentration-dependent increases were observed for Protein Phosphatase 1 Regulatory Subunit 15A (PPP1R15A), Activating Transcription Factor 4 (ATF4), DNA Damage Inducible Transcript 3 (DDIT3) and Asparagine Synthetase (ASNS) genes belonging to the latter pathway (UPR), even in BS, although this pathway was not identified in this model (Fig. 5). Only PODO did not show any regulation of these genes. Finally, more generic pathways such as "Cell cycle", "Cellular responses to external stimuli" and "Generic Transcription Pathway" (Table 3) were found in several combinations of four models, and "ESR-mediated signalling" pathway appeared in one of these combinations.

UpSetR package was also used to identify the genes differentially expressed in the combinations of 2, 3, 4 and 5 cell models in an unbiased way. The gene intersection graph shows that each cell model shared affected genes with up to four other cell models (Fig. 6, intersections between 5 models). No gene shared by all 6 cell models was found with the criteria used in this study. MAFF and PPP1R15A were found deregulated in 5 models (Table 4), but not in the same combination thereof, whereas Vascular Endothelial Growth Factor A (VEGFA), ATF4 and Growth Differentiation Factor 15 (GDF15) were found deregulated in 4 models. NC and BS, both CNS models, shared 6 deregulated genes (Table 4): Chromobox 4 (CBX4), Eukaryotic Translation Initiation Factor 1 (EIF1), PPP1R15A, Solute Carrier Family 3 member 2 (SLC3A2), and VEGFA, that were all upregulated in both models, as well as one downregulated gene, the tubulin alpha-1B chain (TUBA1B). Finally, PODO and PTL, models of two different kidney cell types, shared 9 upregulated genes: CD55 Molecule (CD55), Choline Kinase Alpha (CHKA), Dual Specificity Phosphatase 1 (DUSP1), Enolase 2 (ENO2), Heat Shock Protein Family A Member 1B (HSPA1B), MAFF, MAFG, PPP1R15A and VEGFA (Table 4), that were deregulated in the same direction, except for HSPA1B, downregulated in PODO but upregulated in PTL (Suppl. Fig. 3).

3.4. *Specificity of hiPSC-derived models response to Paraquat*

In addition to the cellular responses shared by the different models, we also interrogated the data for the genes that are specifically deregulated in each model. The number of deregulated probes (22–190) per model is given in Table 5. The number of down-regulated genes was very similar to the number of up-regulated genes in BS, NC and BBB models, whereas the proportion of up-regulated genes was much higher for PTL (73% of the deregulated genes), PODO (70%) and HLC (95%)

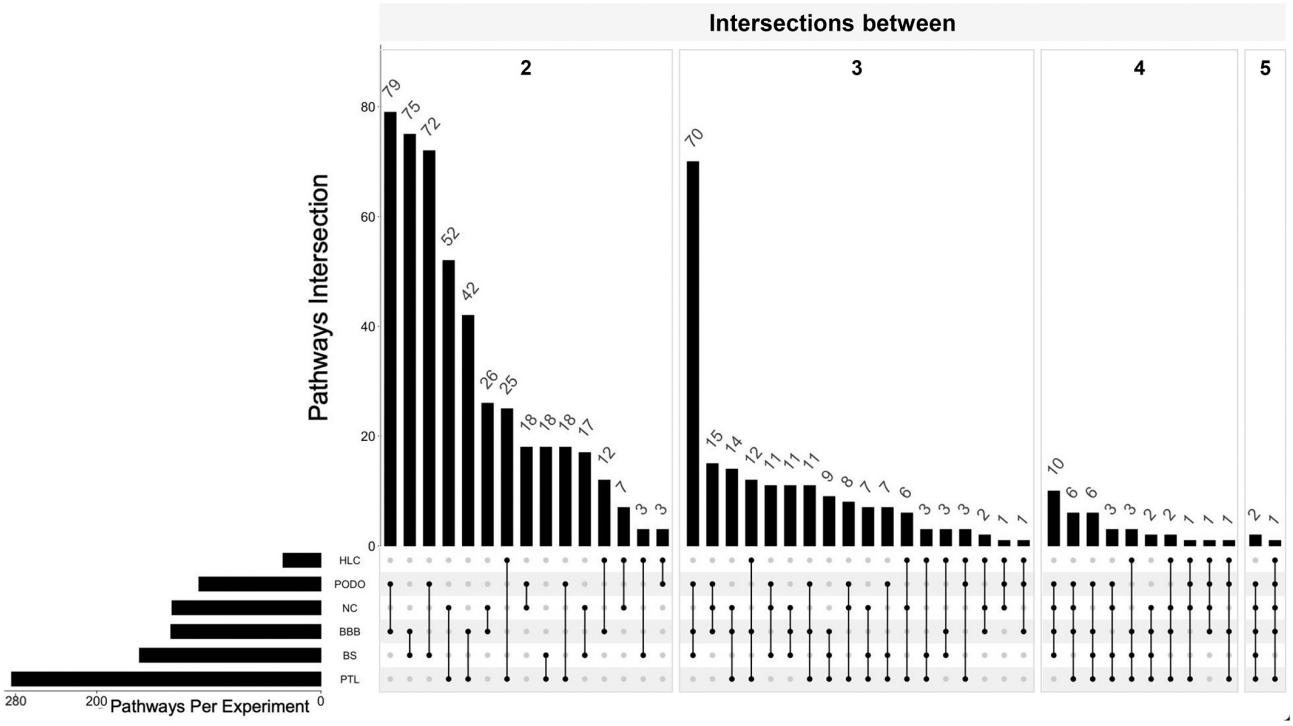


Fig. 4. All the models share perturbated pathways with at least 4 other systems. Intersection plot of the potentially affected pathways by PQ exposure shared by 2–5 models. For each model, the pathways were derived using the differentially expressed genes in the ConsensusPathDB for over-representation analysis (ORA).

Table 3
List of pathways potentially disturbed by PQ exposure in at least 4 different models.

Models	Example
NC + PODO+PTL + HLC + BBB	Oxidative stress induced gene expression *via* Nrf2
NC + BS + PODO+PTL + BBB	Cellular responses to external stimuli; Cellular responses to stress
NC + PODO+PTL + BBB	Oxidative stress induced gene expression *via* Nrf2; Cell cycle; Cellular responses to external stimuli; Cellular responses to stress; Direct p53 effectors; Generic transcription pathway
NC + BS + PODO+PTL	Cellular responses to external stimuli; Cellular responses to stress; Quercetin and Nf-kB- AP-1 induced cell apoptosis
NC + PTL + HLC + BBB	Oxidative stress induced gene expression *via* Nrf2 Photodynamic therapy-induced unfolded protein response;
NC + BS + PODO+BBB	ESR-mediated signaling; B-WICH complex positively regulates rRNA expression; Cell cycle checkpoints; Cellular responses to external stimuli; Cellular responses to stress; Epigenetic regulation of gene expression; Formation of the beta-catenin: TCF transactivating complex; Mitotic prophase; Signaling by nuclear receptors; TCF dependent signaling in response to WNT
NC + BS + PTL + BBB	Cellular responses to external stimuli; Cellular responses to stress
BS + PTL + HLC + BBB	NRF2-ARE regulation; Photodynamic therapy-induced NFE2L2 (Nrf2) survival signaling; Phytochemical activity on Nrf2 transcriptional activation
PODO+PTL + HLC + BBB	Oxidative stress induced gene expression *via* Nrf2

models (Table 5).

Among the top upregulated genes in BS (Table 5) genes involved in essential cellular functions were found, such as caveolin 2 (CAV2) and calpain 1 (CAPN1), but also genes involved in development such as SRY-BOX Transcription Factor 1 (SOX1), INSM transcriptional repressor 1 (INSM1) and Thyroid hormone receptor beta (THRB). On the other hand, the most down-regulated genes in BS were associated with cytoskeleton, such as Tubulin Alpha 1B chain (TUBA1B) and Cortactin Binding Protein 2 (CTTNBP2) and cell adhesion (Cadherin (CDH8)). In the other iPSC-derived model of brain cells, NC, the top 10 up-regulated genes included 4 genes involved in cellular stress responses, such as PPP1R15A, GDF15, Growth Arrest And DNA Damage Inducible Alpha (GADD45A) and MAFF, but also genes implicated in cell adhesion, such as Cellular Communication Network Factor 1 (CCN1 also called CYR61), Cellular Communication Network Factor 2 (CCN2 also called Connective Tissue Growth Factor (CTGF)) and Serpin Family E Member 1 (SERPINE1). In the down-regulated genes for this model, we found genes involved in mitosis (RecQ Mediated Genome Instability 1 (RMI1) and Centrin 3 (CETN3)), another one potentially mediating apoptosis during neuronal development (Pleckstrin Homology Like Domain Family A Member 1 (PHLDA1)) and an essential component of the nuclear pore complex, nucleoporin 85 (NUP85). In BBB model, most of the top ten up-regulated genes were associated with stress responses: Tribbles Pseudokinase 3 (TRIB3), DDIT3, CD55 Molecule (CD55), PPARGC1A, Sequestosome (SQSTM1), and GDF15, whereas among the top ten downregulated genes, five were histones, and one was a histone chaperone (Anti-Silencing Function 1B Histone Chaperone (ASF1B)).

For PTL, the up-regulated genes were all stress genes involved in oxidative stress and/or inflammation, except FOSL1 that is implicated in proliferation, differentiation and transformation; and among the downregulated genes are 2 genes related to detoxification (Epoxide Hydrolase 2 (EPHX2) and Aldehyde Dehydrogenase 7 Family member A1 (ALDH7A1)), as well as pyruvate dehydrogenase kinase isoenzyme 4 (PDK4), that plays a crucial role in the control of metabolic flexibility under various physiological. For the other kidney cell type, PODO, 5 of the top upregulated genes were involved in stress pathways: Matrix Metallopeptidase 2 (MMP2), Immunoglobulin Mu DNA Binding Protein 2 (IGHMBP2), CD55, MAFG and CREB3 regulatory Factor (CREBRF). Among the other upregulated genes we found Myelin Protein Zero Like 1 (MPZL1) and Ankyrin Repeat Domain 36 (ANKRD36B) involved in cell adhesion and cell membrane integrity. The downregulated genes in PODO include five histones and one histone chaperone, ASF1B, just as observed for BBB.

The model for hepatocytes, HLC, showed 6 of its top ten upregulated genes associated with oxidative stress and/or inflammation, namely the rate limiting enzyme of glutathione synthesis, Glutamate-Cysteine Ligase Modifier Subunit (GCLM), HMOX1, C-X-C Motif Chemokine Ligand 8 (CXCL8), GDF15, TRIB3 and MAFF. Besides, Ras Association Domain Family Member 1 (RASSF) involved in cell cycle and Aldo-Keto Reductase Family 1 member 1 (AKR1B10) were also upregulated. The only downregulated gene found with the criteria applied was a type of collagen (COL12A1).

Finally, basal expression of selected genes related to PQ transport, and enzymes involved in redox cycling and anti-oxidant defence were extracted from the TempO-Seq data in order to investigate their potential contribution to the sensitivity of the various models to PQ. The basal expression of Solute Carrier Family 3 Member 2 (SLC3A2) and Solute Carrier Family 7 Member 11 (SLC7A11), two amino acid transporters, was higher in BS, NC and PODO than in the other models (Fig. 7). The expression of the NAD(*P*)H-cytochrome P450 oxidoreductase (POR), involved in the redox cycling of the PQ ion, was highest in HLC than in the other models, but also robustly expressed in BS, NC and PTL. Superoxide dismutases (SOD2 and SOD3) and glutathione S-transferases (GSTA2 and GSTM3), involved in anti-oxidant defence, showed a much higher basal expression in HLC than in all the other models, even being absent from some of them, whereas SOD 1 was more expressed in BS, NC and PTL than in HLC.

4. Discussion

The development of new human-based assays for human safety assessment, highly encouraged by various authorities (EUR-Lex, 2010;

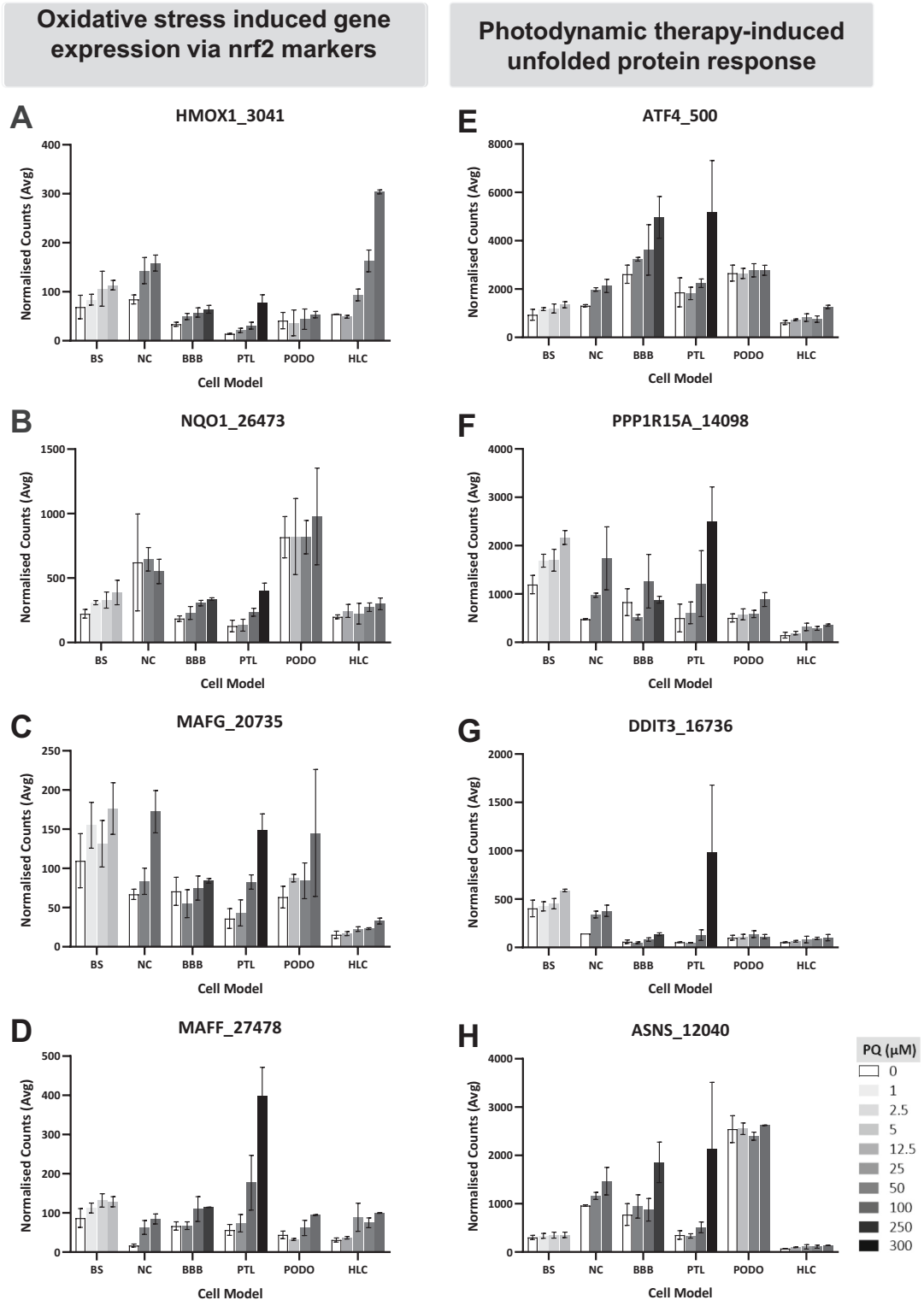


Fig. 5. PQ exposure induces upregulation of genes involved in the oxidative stress pathways Nrf2 and UPR (A-D) normalized read counts for HMOX1, NQO1, MAFG and MAFF probes (gene symbol_probe ID) involved in the "oxidative stress induced gene expression via Nrf2" pathway from BioCarta database; (*E*-H) normalized read counts for ATF4, PPP1R15A, DDIT3 and ASNS probes (gene symbol_probe ID) involved in "photodynamic therapy-induced unfolded protein response" pathway described in the Wikipathways database. Average normalized read counts ±SD are shown.

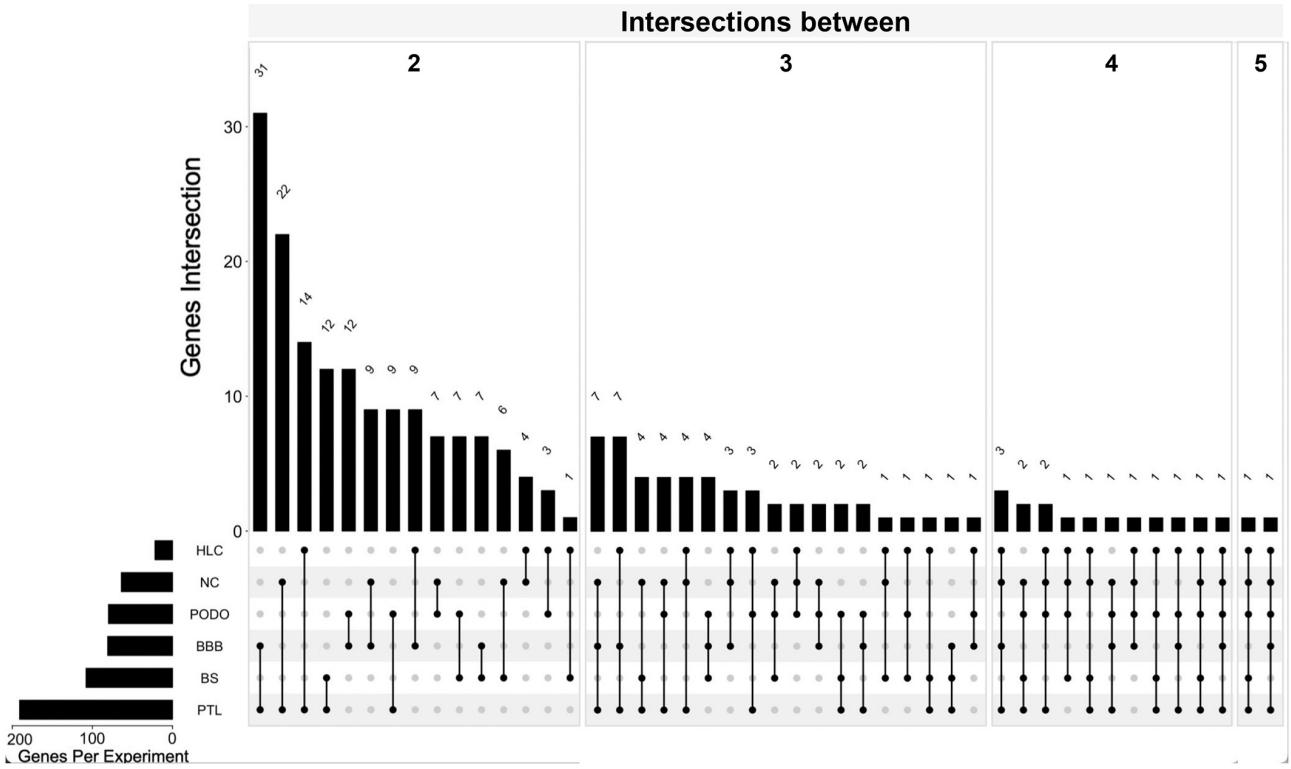


Fig. 6. All the models share differentially expressed genes with at least 4 other systems. Intersection plot of the PQ-induced DEG shared by 2–5 models using UpSetR pack. For each model, the DEG list was obtain with the DSeq2 Package.

Table 4
List of the DEGs shared by at least 4 models or by modeled organ (brain and kidney).

Organ	Intersections	Gene symbol_probe ID
	NC + PODO+PTL + HLC + BBB	MAFF_27478
	NC + BS + PODO+PTL + HLC	PPP1R15A_14098
	NC + BS + PODO+PTL	PPP1R15A_14098; VEGFA_28053
	NC + PODO+PTL + HLC	PPP1R15A_14098; MAFF_27478;
	NC + PTL + HLC + BBB	ATF4_500; GDF15_18329; MAFF_27478
Brain	NC + BS	CBX4_16907; EIF1_2074; PPP1R15A_14098; SLC3A2_20164; TUBA1B_7399; VEGFA_28053
Kidney	PODO+PTL	CD55_28297; CHKA_14315; DUSP1_24234; ENO2_13140; HSPA1B_3136; MAFF_27478; MAFG_20735; PPP1R15A_14098; VEGFA_28053

National Research Council, 2007; Oredsson et al., 2019; Sewell et al., 2017), has been made easier by the development of standardized ways to derive hiPSC from somatic cells and by the description of new methods for their differentiation into various cell types (Chandrasekaran et al., 2021; Murphy et al., 2019; Pamies et al., 2017). This raises the possibility to test for specific organ toxicity within the same genetic background to more fully assess the potential hazard of a chemical. Here we propose an *in vitro* multi-organ strategy with hiPSCs-derived models to assess the toxicity of chemicals though TempO-Seq analysis. We believe that a multi-organ approach associated with high throughput omics technology would improve the efficiency and reliability of chemical toxicity screening.

To test this approach, PQ, a herbicide widely employed with recognized toxic effects in many different organs, such as lung, kidney, central nervous system and liver (Delirrad et al., 2015; Gawarammana and Buckley, 2011; Saravu et al., 2013; Sittipunt, 2005), was used as a reference compound. The cytotoxic sensitivity to PQ we observed in this study was not the same for the different hiPSCs-derived models, with liver and kidney models being less sensitive than brain models. This is in line with a report showing that in rats, repeated, systemic low doses of PQ are much better tolerated by peripheral organs than by brain, and suggesting this to be due to easier induction of adaptive mechanisms (Kuter et al., 2010). Furthermore, this observation suggests that the application of the diverse differentiation protocols was successful. The differential sensitivity observed might be explained by the higher basal expression of amino acid transporters, such as Solute Carrier Family 3 Member 2 (SLC3A2) and Solute Carrier Family 7 Member 11 (SLC7A11), shown by TempO-Seq analysis in the most sensitive models. Indeed, amino acid transporters have been suggested to be used by PQ to enter cells (for review, see Wang et al., 2021), their stronger expression might thus potentially lead to a higher cellular uptake of PQ. The highest level of genes coding for enzymes involved in anti-oxidant defence, such as SODs and GSTs, observed in HLC, might also explain the relatively highest resistance of this model to PQ, even so NAD(*P*)H-cytochrome P450 oxidoreductase (POR), engaged in ion redox cycling of PQ was a bit more expressed in HLC. However, other genes related to transport, redox cycling and anti-oxidant defence important for PQ mode of action and cellular response were unfortunately not available in the chosen TempO-Seq panel and more experiments are needed to fully characterize the causes of the differential sensitivity to PQ observed among the

Table 5
List of the 10 most upregulated and downregulated genes after exposure to the highest concentration of PQ for each model.

Model	Downregulated			Upregulated			Total DEG
	Gene symbol_probe ID	LFC	Number (% of total)	Gene symbol_probe ID	LFC	Number (% of total)	
BS	SPC25_6704	−3.0	55 (56)	CAV2_15304	2.5	43 (44)	98
	TOMM20_11983	−2.3		CAPN1_957	2.4		
	CIB1_16624	−2.3		SPAG7_6692	1.8		
	CTTNBP2_13527	−2.2		TOMM34_7272	1.7		
	FBXO11_2372	−1.8		THRB_24158	1.4		
	TRIM2_7343	−1.4		SOX1_27894	1.4		
	TUBA1B_7399	−1.4		CSRNP1_22653	1.3		
	NDUFA4_14125	−1.4		INSM1_21254	1.2		
	CDH8_19001	−1.3		TFAP2A_7039	1.2		
	HIST1H2BG_2953	−1.3		HIF1A_2945	1.2		
NC	SGK2_10628	−6.1	30 (47)	CYR61_24505	6.3	34 (53)	64
	EDNRA_25237	−5.1		PLK1_5203	3.6		
	IFI44L_19848	−3.0		EXOSC5_27191	3.4		
	TCF4_15698	−2.6		SERPINE1_6253	2.7		
	FKBP14_2430	−2.2		CTGF_17007	2.3		
	NUP85_4781	−2.1		MAFF_27478	2.3		
	MYBL1_22509	−2.0		SLC3A2_20164	2.2		
	RMI1_15666	−2.0		GADD45A_2569	2.1		
	CETN3_27006	−1.8		GDF15_18329	1.9		
	PHLDA1_12880	−1.7		PPP1R15A_14098	1.9		
BBB	NT5DC2_4737	−2.6	43 (54)	GDF15_18329	1.9	36 (46)	79
	NCAM1_23834	−1.3		SPP1_6720	1.5		
	HMMR_3040	−1.2		SQSTM1_6740	1.5		
	HIST1H3H_2957	−1.0		PPARGC1A_12225	1.4		
	HIST1H2BG_2953	−1.0		CD55_1141	1.3		
	PSRC1_5548	−1.0		ASNS_12040	1.3		
	HIST2H2AA4_2963	−1.0		DDIT3_16736	1.2		
	HIST1H2BM_2956	−1.0		TRIB3_7337	1.2		
	HIST1H1E_14544	−1.0		CXCL2_28232	1.2		
	ASF1B_26041	−1.0		HSD17B11_3115	1.1		
PTL	CXCL14_22017	−2.2	51 (27)	GDF15_18329	6.8	139 (73)	190
	CSF3R_25969	−1.8		DDIT3_16736	4.2		
	PRR15L_5469	−1.7		NMRAL2P_81336	3.9		
	EPHX2_18456	−1.7		TRIB3_7337	3.8		
	CASP1_26966	−1.6		CHAC1_1279	3.4		
	GBP1_19263	−1.6		FOSL1_2463	3.2		
	DEPTOR_26568	−1.5		ATF3_499	3.1		
	ALDH7A1_227	−1.5		GADD45A_2569	2.9		
	CAPS_26136	−1.5		MAFF_27478	2.8		
	PDK4_28868	−1.4		IL1A_22714	2.8		
PODO	ASF1B_26041	−2.5	24 (30)	USP32P2_28186	2.2	55 (70)	79
	MYBL2_4391	−2.1		MPZL1_4249	2.1		
	HIST1H2BH_2954	−1.6		ANKRD36B_290	1.8		
	HIST1H2BB_12129	−1.4		MMP2_4213	1.8		
	HIST1H2BN_11971	−1.4		DMRTA1_14819	1.7		
	NFYB_16601	−1.3		NEAT1_81333	1.7		
	MEST_4122	−1.2		IGHMBP2_3275	1.5		
	HIST1H2BI_11969	−1.2		CD55_28297	1.5		
	IGFBP7_26667	−1.2		MAFG_20735	1.2		
	HIST2H2AA4_2963	−1.1		CREBRF_19339	1.1		
HLC	COL12A1_1461	−1.7	1 (5)	FOSL1_2463	2.7	21 (95)	22
				HMOX1_3041	2.5		
				AKR1B10_19908	1.8		
				MAFF_27478	1.7		
				GDF15_18329	1.7		
				CXCL8_14324	1.6		
				ZNF165_26222	1.6		
				TRIB3_7337	1.5		
				GCLM_25908	1.5		
				RASSF1_28369	1.5		

DEG: differentially expressed genes; LFC: log2 fold change.

various cell models. Both brain models also differ in their sensitivity to PQ. This may be explained by the cell composition and the type of cultures. The most sensitive model, NC, is a 2D model comprising neurons and a few astrocyte progenitors, whereas BS exhibit a higher astrocyte to neuron ratio, comprise oligodendrocytes and form 3D structures, possibly explaining its slightly highest resistance to paraquat.

PQ is known to produce high levels of reactive oxygen species (ROS) through mitochondrial redox cycling, and the induction of oxidative stress is one of its better described mechanisms of action in different organs (Blanco-Ayala et al., 2014; Dinis-Oliveira et al., 2008; Gawarammana and Buckley, 2011). In our study, one of the three pathways deregulated by PQ in five cell models, identified by enrichment analysis using ConsensusPathDB, was “Oxidative stress induced gene expression via Nrf2 markers”. Keap1/Nrf2 activation functions mainly as an antioxidant mechanism helping cells to cope with oxidative stress and has been previously associated with PQ-induced injury (Dou et al., 2016; Kheiripour et al., 2021). Nrf2 transactivation induces the transcription of genes with upstream antioxidant response elements (ARE) to

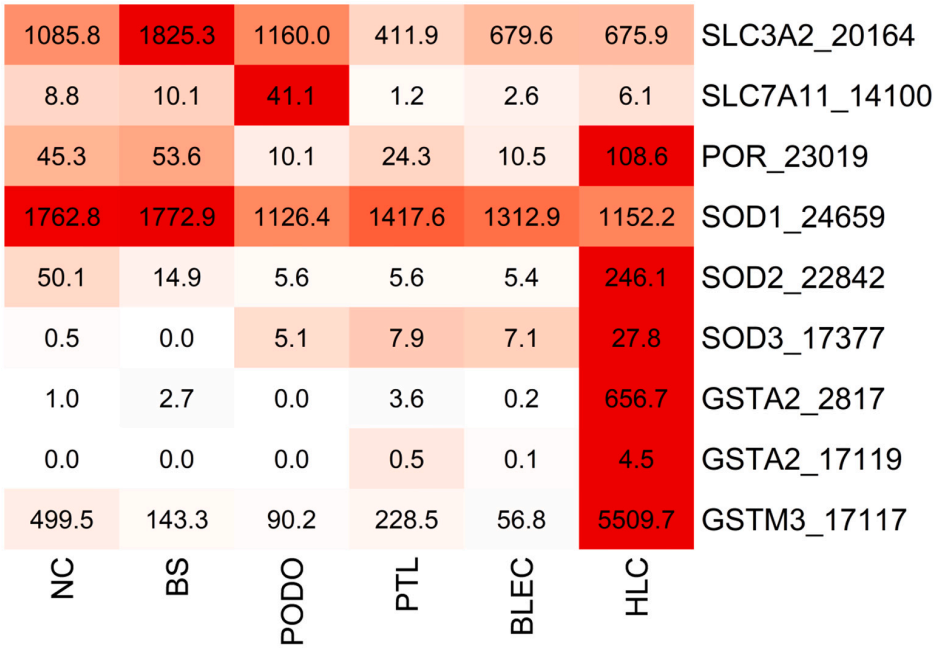


Fig. 7. Heatmap of TempO-Seq normalized read counts data for selected genes involved in PQ transport, redox cycling and anti-oxidant defence. White to red intensity represents 0 (white) to highest normalized read counts (red) of the TempO-Seq probe in the displayed data set. Annotations represent the gene symbol and Tempo-Seq probe ID. Non-normalized values are printed to allow for a quantitative comparison between genes.

counteract oxidative stress, including glutathione- reduction, −synthesis, and -recycling along with other reducing enzymes (Bugno et al., 2015; Wang et al., 2020). HMOX and NQO1 are such prototypical Nrf2 response genes (Hichor et al., 2017; Xu et al., 2017), and were induced in all the hiPSCs-derived cell systems exposed to PQ in this study. Indeed, although the thresholds were not reached to significantly activate this pathway for BS after acute exposure, we observed a concentration-dependent increase in the expression of the associated genes, and the pathway itself was impacted after 1 week of repeated exposure to PQ (Nunes et al., 2021, unpublished observation).

Endoplasmic reticulum (ER), the site of synthesis and folding of proteins, maintains a strict quality control targeting incorrectly folded proteins for proteasomal degradation (Smith et al., 2011). Upon accumulation of protein-folding alterations occurring during ER stress, cells activate a series of complementary adaptive mechanisms, the unfolded protein response (UPR), primarily serving to return normal ER function, and failure to adapt to ER stress may result in apoptosis triggered by UPR (Hetz, 2012; Lin and LaVail, 2010). Perturbations in ER calcium, high ROS production, depletion of oxidised glutathione and proteotoxicity all have the potential to activate the UPR. Indeed, disturbance of ER and UPR have been described in human lung epithelial cells after PQ exposure (Chinta et al., 2008; Omura et al., 2013). Furthermore, recently UPR was reported after PQ treatment in primary cultures of human CD34+ derived dendritic cells, but not in macrophages (Fransen and Leonard, 2021). In the present study, the UPR pathway was highlighted in four out of the six models. As for the oxidative stress pathway, a repeated exposure to PQ was necessary to detect the UPR using ConsensusPathBD in BrainSpheres (Nunes et al., 2021, unpublished observation), although a clear concentration-dependent increase was observed in the expression of several genes involved in this pathway, such as ATF4, PPP1R15A and DDIT3, already after 24 h of exposure. The second model, for which ConsensusPathDB did not highlight UPR, the podocyte model (PODO), showed only a slight modification of PPP1R15A, suggesting no or only very weak activation of UPR-PERK branch, contrary to the observation made in the other kidney cell type, PTL. Further studies are needed to determine whether PQ activates UPR through the induction of ER stress, as described in lung epithelial cells, or directly through the induction of ROS, since an interaction between UPR and oxidative stress has been described (Kupsco and Schlenk, 2015), involving the phosphorylated form of eIF2α, ATF4, CHOP (transcription factor C/EBP homologous protein, also called DDIT3) and PPP1R15A (Harding et al., 2009; Hetz, 2012; Pavitt and Ron, 2012; Tabara et al., 2018). Altogether, our results show that even though we used a minimum overlap of two genes between the input gene list and a given pathway associated gene list to define a pathway as overrepresented, our strategy using several hiPSC-derived models, with various cell type or organ specificities, allowed to detect the main mechanisms of action of PQ.

Another pathway revealed after PQ exposure in four cell culture systems, including both brain models (NC and BS), BLECs and HLC, was "ESR-mediated signaling". This is interesting given that estrogenic dysfunction has recently been identified to be involved in toxicity induced by PQ on rat primary hippocampal neurons *in vitro* (Moyano et al., 2020a, 2020b) and that estradiol has been shown to protect a neuronal line from PQ toxicity (Gélinas et al., 2004). It is therefore remarkable to note that our strategy not only retrieved the main, known mechanisms of action of PQ, but also a recently proposed one that clearly deserves further studies.

Most of the top ten upregulated genes per model were involved in stress responses, in particular oxidative stress and inflammation for four hiPSC-derived models out of the six tested. The two brain models, NC and BS, however behaved differently, exhibiting genes involved in development, such as OX1 and INSM1 that promote generation and expansion of neuronal progenitor cells, and the thyroid receptor that is also important in the development and maintenance of brain structures. The lower proportion of genes involved in stress responses among the top upregulated genes in brain models suggest that brain cells have the weakest self-defence capacity against toxins or can less easily induce their adaptative mechanisms than peripheral organs (Kuter et al., 2010). These results show that although stress pathways shared by most of the models have been identified, each model seemed to also react in a specific way to paraquat. This is remarkable since in this study we used a

restricted number of probes, that were not specifically designed for cell type-specific discrimination.

Finally, VEGFA was induced by PQ in four of the models. VEGFA, a prototypical HIF1 alpha gene, primarily induced as a response to hypoxia, is mainly known for its activities related to angiogenesis, vascular permeability and vascular survival (Ferrara, 2004). There is a growing body of evidence, however, that VEGFA fulfils additional less 'traditional' functions in multiple organs during development, as well as homeostatic functions in fully developed organs (for review, see (Licht and Keshet, 2013)). Our results showing the upregulation of VEGFA shared by both brain models and both kidney models are perfectly in line with the current knowledge on the importance of VEGFA for these two organs. VEGFA, mainly expressed in astrocytes and neurons is upregulated in all brain cell types upon severe hypoxia (Jin et al., 2000). It is also expressed in glomerular podocytes, where it is critical for the establishment and maintenance of glomerular filtration (Eremina et al., 2003), and ample evidence supports important roles for this growth factor in the maintenance of tubulointerstitial integrity and the response to acute kidney injury (Doi et al., 2010; Schrijvers et al., 2004). The upregulation of VEGFA in brain and kidney models after exposure to PQ, again shows that the hiPSCs-derived models used in this study are reproducing the organ patho-physiology, and furthermore suggests this growth factor as marker of chemical injury.

Some limitations of the study should be stated. The cell models utilised here although from the same donors were cultured, differentiated and treated separately. Our systems and the majority of *in vitro* models in general represent a reductionist approach and does not fully reflect the physiology of a whole organism due to the lack of connections and communications between them through blood flow, endocrine and nervous systems. This could possibly be better recapitulated in the future by using organ-on-a-chip models and/or by the application of *in silico* modelling. Furthermore, the iPSCs-derived models used in this study, beside their great advantages, come also with some limitations. For example, although the BS model is multicellular and highly complex it does not (at this point in development) possess microglia, precluding the development of the full neuroinflammatory cascade. The HLC are not completely mature hepatocytes and lack cooperation with Kupfer cells, and the kidney model PODO lacks fenestrated endothelial cells and glomerular basement membrane to form a full glomerular model. These limitations can be overcome in the near future by co-culturing cells with their missing partners. Finally, in order to overcome the multiple differences linked to cell culture conditions that could impact the true comparisons of the effects of PQ in the different models, the distribution kinetics of PQ has to be established for each model, to correlate the observed toxic effects to the effective concentrations reaching the cells and not to the nominal concentrations used. This will allow a better comparison of the effects of chemicals on the various models.

5. Conclusions

Our strategy, taking advantage of the iPSC technology, allowed to determine the known and less known mechanisms of PQ toxicity with TempO-Seq analysis, using a cost-effective panel of probes specific for toxicity testing. The main advantages of this strategy are to assess chemical toxicity on cell types from multiple organs in parallel, exclusively in human cells, eliminating interspecies bias and allowing a better evaluation of the differential sensitivity of the models representing the diverse organs. To apply this approach to the general assessment of chemical toxicity, the strategy presented herein could be further improved by adding hiPSC-derived models for other organs, such as the heart and the lungs, and by standardization of procedures to reach a higher level of harmonization between the different laboratories. Furthermore, organ-specific toxicity testing could be reached by adding organ-specific genes to our set of TempO-Seq probes.

Given the present results showing the differential sensitivity of the various human models to PQ, we believe this strategy will contribute to the further improvement of chemical risk assessment for human health, providing the above cited improvements combined with the evaluation of the distribution kinetics of the chemicals, as a key step for the *in vitro* to *in vivo* extrapolation (IVIVE) (Punt et al., 2020), absolutely required for regulatory decision-making.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tiv.2022.105333.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

The work was funded by the European Union's Horizon 2020 Research and Innovation Programme MarieSklodowska-Curie Action-Innovative Training Network project in3, under grant agreement no. 721975 (fellowships to Carolina Nunes, Pranika Singh, Zahra Mazidi, Cormac Murphy, Aurore Bourguignon, Sara Wellens, Vidya Chandrasekaran and Sreya Ghosh).

References

Blanco-Ayala, T., Andérica-Romero, A.C., Pedraza-Chaverri, J., 2014. New insights into antioxidant strategies against paraquat toxicity. Free Radic. Res. 48, 623–640. https://doi.org/10.3109/10715762.2014.899694.

Boon, R., Kumar, M., Tricot, T., Elia, I., Ordovas, L., Jacobs, F., One, J., De Smedt, J., Eelen, G., Bird, M., Roelandt, P., Doglioni, G., Vriens, K., Rossi, M., Vazquez, M.A., Vanwelden, T., Chesnais, F., El Taghdouini, A., Najimi, M., Sokal, E., Cassiman, D., Snoeys, J., Monshouwer, M., Hu, W.S., Lange, C., Carmeliet, P., Fendt, S.M., Verfaillie, C.M., 2020. Amino acid levels determine metabolism and CYP450 function of hepatocytes and hepatoma cell lines. Nat. Commun. 11, 1–16. https://doi.org/10.1038/s41467-020-15058-6.

Bugno, M., Daniel, M., Chepelev, N.L., Willmore, W.G., 2015. Changing gears in Nrf1 research, from mechanisms of regulation to its role in disease and prevention. Biochim. Biophys. Acta - Gene Regul. Mech. 1849, 1260–1276. https://doi.org/10.1016/j.bbagrm.2015.08.001.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., Studer, L., Author, N.B., 2009. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling HHS Public Access Author manuscript. Nat. Biotechnol. 27, 275–280. https://doi.org/10.1038/nbt.1529.

Chandrasekaran, A., Avci, H.X., Ochalek, A., Rösingh, L.N., Molnár, K., László, L., Bellák, T., Téglási, A., Pesti, K., Mike, A., Phanthong, P., Bíró, O., Hall, V., Kitiyanant, N., Krause, K.H., Kobolák, J., Dinnyés, A., 2017. Comparison of 2D and 3D neural induction methods for the generation of neural progenitor cells from human induced pluripotent stem cells. Stem Cell Res. 25, 139–151. https://doi.org/10.1016/j.scr.2017.10.010.

Chandrasekaran, V., Carta, G., Pereira, C., Gupta, R., Murphy, C., Feifel, E., Kern, G., Lechner, J., Cavallo, A.L., Gupta, S., Caiment, F., Kleinjans, J.C.S., Gstraunthaler, G., 2021. Generation and characterization of iPSC – derived renal proximal tubule – like cells with extended stability. Sci. Rep. 1–17 https://doi.org/10.1038/s41598-021-89550-4.

Chinta, S.J., Rane, A., Poksay, K.S., Bredesen, D.E., Andersen, J.K., Rao, R.V., 2008. Coupling endoplasmic reticulum stress to the cell death program in dopaminergic cells: effect of paraquat. Neuro Mol. Med. 10, 333–342. https://doi.org/10.1007/s12017-008-8047-9.

Cicchetti, F., Drouin-Ouellet, J., Gross, R.E., 2009. Environmental toxins and Parkinson's disease: what have we learned from pesticide-induced animal models? Trends Pharmacol. Sci. 30, 475–483. https://doi.org/10.1016/j.tips.2009.06.005.

Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.

Delirrad, M., Majidi, M., Boushehri, B., 2015. Clinical features and prognosis of paraquat poisoning: a review of 41 cases. J. Clin. Exp. Med 8 (5), 8122–8128.

Dinis-Oliveira, R.J., Duarte, J.A., Sánchez-Navarro, A., Remião, F., Bastos, M.L., Carvalho, F., 2008. Paraquat poisonings: mechanisms of lung toxicity, clinical features, and treatment. Crit. Rev. Toxicol. 38, 13–71. https://doi.org/10.1080/10408440701669959.

Doi, K., Noiri, E., Fujita, T., 2010. Role of vascular endothelial growth factor in kidney disease. Curr. Vasc. Pharmacol. 8, 122–128. https://doi.org/10.2174/157016110790226606.

Dou, T., Yan, M., Wang, X., Lu, W., Zhao, L., Lou, D., Wu, C., Chang, X., Zhou, Z., 2016. Nrf2/ARE pathway involved in oxidative stress induced by paraquat in human neural progenitor cells. Oxidative Med. Cell. Longev. 2016 https://doi.org/10.1155/2016/8923860.

EPA, 2019. Administrator Memo Prioritizing Efforts to Reduce Animal Testing, September 10, 2019. United States Environ. Prot. Agency.

Eremina, V., Sood, M., Haigh, J., Nagy, A., Lajoie, G., Ferrara, N., Gerber, H.P., Kikkawa, Y., Miner, J.H., Quaggin, S.E., 2003. Glomerular-specific alterations of

VEGF-A expression lead to distinct congenital and acquired renal diseases. J. Clin. Invest. 111, 707–716. https://doi.org/10.1172/JCI17423.
EUR-Lex, 2010. Directive 2010/63/EU of the European parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Off. J. Eur. Union 1–61.
Ferrara, N., 2004. Vascular endothelial growth factor: basic science and clinical progress. Endocr. Rev. https://doi.org/10.1210/er.2003-0027.
Fransen, L.F.H., Leonard, M.O., 2021. CD34+ derived macrophage and dendritic cells display differential responses to paraquat. Toxicol. in Vitro 75, 105198. https://doi.org/10.1016/j.tiv.2021.105198.
Gao, L., Yuan, H., Xu, E., Liu, J., 2020. Toxicology of paraquat and pharmacology of the protective effect of 5-hydroxy-1-methylhydantoin on lung injury caused by paraquat based on metabolomics. Sci. Rep. 10, 4–8. https://doi.org/10.1038/s41598-020-58599-y.
Gawarammana, I.B., Buckley, N.A., 2011. Medical management of paraquat ingestion. Br. J. Clin. Pharmacol. 72, 745–757. https://doi.org/10.1111/j.1365-2125.2011.04026.x.
Gélinas, S., Bureau, G., Valastro, B., Massicotte, G., Cicchetti, F., Chiasson, K., Gagne, B., Blanchet, J., Martinoli, M.G., 2004. Alpha and beta estradiol protect neuronal but not native PC12 cells from paraquat-induced oxidative stress. Neurotox. Res. 6, 141–148. https://doi.org/10.1007/BF03033216.
Gholami, S., Mazidi, Z., Pahlavan, S., Moslem, F., Hosseini, M., Taei, A., Hesaraki, M., Barekat, M., Aghdami, N., Baharvand, H., 2021. A novel insight into endothelial and cardiac cells phenotype in systemic sclerosis using patient-specific induced pluripotent stem cell. Cell J. 23 https://doi.org/10.22074/cellj.2021.7244.This.
Harding, H.P., Zhang, Y., Scheuner, D., Chen, J.J., Kaufman, R.J., Ron, D., 2009. Ppp1r15 gene knockout reveals an essential role for translation initiation factor 2 alpha (eIF2α) dephosphorylation in mammalian development. Proc. Natl. Acad. Sci. U. S. A. 106, 1832–1837. https://doi.org/10.1073/pnas.0809632106.
Herrmann, K., Pistollato, F., Stephens, M.L., 2019. Food for thought … beyond the 3Rs: expanding the use of human-relevant replacement methods in biomedical research. ALTEX 36, 343–352. https://doi.org/10.14573/altex.1907031.
Hetz, C., 2012. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. Nat. Rev. Mol. Cell Biol. 13, 89–102. https://doi.org/10.1038/nrm3270.
Hichor, M., Sampathkumar, N.K., Montanaro, J., Borderie, D., Petit, P.X., Gorgievski, V., Tzavara, E.T., Eid, A.A., Charbonnier, F., Grenier, J., Massaad, C., 2017. Paraquat induces peripheral myelin disruption and locomotor defects: crosstalk with LXR and Wnt pathways. Antioxid. Redox Signal. 27, 168–183. https://doi.org/10.1089/ars.2016.6711.
Jin, K.L., Mao, X.O., Nagayama, T., Goldsmith, P.C., Greenberg, D.A., 2000. Induction of vascular endothelial growth factor and hypoxia-inducible factor-1α by global ischemia in rat brain. Neuroscience 99, 577–585. https://doi.org/10.1016/S0306-4522(00)00207-4.
Kamburov, A., Wierling, C., Lehrach, H., Herwig, R., 2009. ConsensusPathDB - a database for integrating human functional interaction networks. Nucleic Acids Res. 37, 623–628. https://doi.org/10.1093/nar/gkn698.
Kamburov, A., Stelzl, U., Lehrach, H., Herwig, R., 2013. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 41, 793–800. https://doi.org/10.1093/nar/gks1055.
Kheiripour, N., Plarak, A., Heshmati, A., Asl, S.S., Mehri, F., Ebadollahi-Natanzi, A., Ranjbar, A., Hosseini, A., 2021. Evaluation of the hepatoprotective effects of curcumin and nanocurcumin against paraquat-induced liver injury in rats: modulation of oxidative stress and Nrf2 pathway. J. Biochem. Mol. Toxicol. 1–9 https://doi.org/10.1002/jbt.22739.
Kupsco, A., Schlenk, D., 2015. Oxidative Stress, Unfolded Protein Response, and Apoptosis in Developmental Toxicity, International Review of Cell and Molecular Biology. Elsevier Ltd. https://doi.org/10.1016/bs.ircmb.2015.02.002
Kuter, K., Nowak, P., Gołembiowska, K., Ossowska, K., 2010. Increased reactive oxygen species production in the brain after repeated low-dose pesticide paraquat exposure in rats. A comparison with peripheral tissues. Neurochem. Res. 35, 1121–1130. https://doi.org/10.1007/s11064-010-0163-x.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
Licht, T., Keshet, E., 2013. Delineating multiple functions of VEGF-A in the adult brain. Cell. Mol. Life Sci. https://doi.org/10.1007/s00018-013-1280-x.
Limonciel, A., Ates, G., Carta, G., Wilmes, A., Watzele, M., Shepard, P.J., VanSteenhouse, H.C., Seligmann, B., Yeakley, J.M., van de Water, B., Vinken, M., Jennings, P., 2018. Comparison of base-line and chemical-induced transcriptomic responses in HepaRG and RPTEC/TERT1 cells using TempO-Seq. Arch. Toxicol. 92, 2517–2531. https://doi.org/10.1007/s00204-018-2256-2.
Lin, J.H., LaVail, M.M., 2010. Misfolded Proteins and Retinal Dystrophies, in: Advances in Experimental Medicine and Biology. Springer, New York, NY, pp. 115–121. https://doi.org/10.1007/978-1-4419-1399-9_14.
Liu, S., Yin, N., Faiola, F., 2017. Prospects and Frontiers of stem cell toxicology. Stem Cells Dev. 26, 1528–1539. https://doi.org/10.1089/scd.2017.0150.
Lo Giudice, M., Mihalik, B., Turi, Z., Dinnyés, A., Kobolák, J., 2019. Calcilytic NPS 2143 reduces amyloid secretion and increases sAβPPα release from PSEN1 mutant iPSC-derived neurons. J. Alzheimers Dis. 72, 885–899. https://doi.org/10.3233/JAD-190602.
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.
Mav, D., Shah, R.R., Howard, B.E., Auerbach, S.S., Bushel, P.R., Collins, J.B., Gerhold, D. L., Judson, R.S., Karmaus, A.L., Maull, E.A., Mendrick, D.L., Merrick, B.A., Sipes, N. S., Svoboda, D., Paules, R.S., 2018. A hybrid gene selection approach to create the S1500+ targeted gene sets for use in high-throughput transcriptomics. PLoS One 13, e0191105. https://doi.org/10.1371/JOURNAL.PONE.0191105.
Morrison, M., Klein, C., Clemann, N., Collier, D.A., Hardy, J., Heiβerer, B., Cader, M.Z., Graf, M., Kaye, J., 2015. StemBANCC: governing access to material and data in a large stem cell research consortium. Stem Cell Rev. Rep. 11, 681–687. https://doi.org/10.1007/s12015-015-9599-3.
Moyano, P., Sanjuan, J., García, J.M., Anadon, M.J., Lobo, M., Pelayo, A., García, J., Frejo, M.T., del Pino, J., 2020a. Primary hippocampal estrogenic dysfunction induces synaptic proteins alteration and neuronal cell death after single and repeated paraquat exposure. Food Chem. Toxicol. 136, 110961 https://doi.org/10.1016/j.fct.2019.110961.
Moyano, P., Sanjuan, J., García, J.M., Anadon, M.J., Naval, M.V., Sola, E., García, J., Frejo, M.T., Pino, J. del, 2020b. Dysregulation of prostaglandine E2 and BDNF signaling mediated by estrogenic dysfunction induces primary hippocampal neuronal cell death after single and repeated paraquat treatment. Food Chem. Toxicol. 144 https://doi.org/10.1016/j.fct.2020.111611.
Murphy, C., Feifel, E., Jennings, P., Gstraunthaler, G., Wilmes, A., 2019. A protocol for One-step differentiation of human induced pluripotent stem cells into mature podocytes. In: Mandenius, C.-F., Ross, J.A. (Eds.), Cell-Based Assays Using iPSCs for Drug Development and Testing. Springer, New York, New York, NY, pp. 93–99. https://doi.org/10.1007/978-1-4939-9477-9_8.
National Research Council, 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy [WWW Document]. Natl. Acad. Press, Washington, DC. URL. https://books.google.ch/books?hl=pt-PT&lr=&id=3AWfAwAAQBAJ&oi=fnd&pg=PT21&dq=Toxicity+Testing+in+the+21st+Century:+A+Vision+and+a+Strategy&ots=UwnqSFwazc&sig=KCxUqgElnoh_RuE2YHTzTTuNRMU&redir_esc=y#v=onepage&q=ToxicityTestinginthe21stCentury%3AAVis (accessed 4.30.21).
Nunes, C., Zurich, M.-G., 2020. Neurotoxicology and Disease Modelling. Springer, Cham, pp. 229–246. https://doi.org/10.1007/978-3-030-43939-2_12.
Ochalek, A., Mihalik, B., Avci, H.X., Chandrasekaran, A., Téglási, A., Bock, I., Giudice, M. Lo, Táncos, Z., Molnár, K., László, L., Nielsen, J.E., Holst, B., Freude, K., Hyttel, P., Kobolák, J., Dinnyés, A., 2017. Neurons derived from sporadic Alzheimer's disease iPSCs reveal elevated TAU hyperphosphorylation, increased amyloid levels, and GSK3B activation. Alzheimers Res. Ther. 9, 1–19. https://doi.org/10.1186/s13195-017-0317-z.
Omura, T., Asari, M., Yamamoto, J., Oka, K., Hoshina, C., Maseda, C., Awaya, T., Tasaki, Y., Shiono, H., Yonezawa, A., Masuda, S., Matsubara, K., Shimizu, K., 2013. Sodium tauroursodeoxycholate prevents paraquat-induced cell death by suppressing endoplasmic reticulum stress responses in human lung epithelial A549 cells. Biochem. Biophys. Res. Commun. 432, 689–694. https://doi.org/10.1016/j.bbrc.2013.01.131.
Oredsson, S., Coecke, S., van der Valk, J., Vinken, M., 2019. What is understood by "animal-free research"? Toxicol. in Vitro 57, 143–144. https://doi.org/10.1016/j.tiv.2019.03.001.
Pamies, D., Barreras, P., Block, K., Makri, G., Kumar, A., Wiersma, D., Smirnova, L., Zhang, C., Bressler, J., Christian, K.M., Harris, G., Ming, G.L., Berlinicke, C.J., Kyro, K., Song, H., Pardo, C.A., Hartung, T., Hogberg, H.T., 2017. A human brain microphysiological system derived from induced pluripotent stem cells to study neurological diseases and toxicity. ALTEX 34, 362–376. https://doi.org/10.14573/altex.1609122.
Pavitt, G.D., Ron, D., 2012. New insights into translational regulation in the endoplasmic reticulum unfolded protein response. Cold Spring Harb. Perspect. Biol. 4, 1–13. https://doi.org/10.1101/cshperspect.a012278.
Punt, A., Bouwmeester, H., Blaauboer, B.J., Coecke, S., Hakkert, B., Hendriks, D.F.G., Jennings, P., Kramer, N.I., Neuhoff, S., Masereeuw, R., Paini, A., Peijnenburg, A.A.C. M., Rooseboom, M., Shuler, M.L., Sorrell, I., Spee, B., Strikwold, M., Van der Meer, A. D., Van der Zande, M., Vinken, M., Yang, H., Bos, P.M.J., Heringa, M.B., 2020. New approach methodologies (NAMs) for human-relevant biokinetics predictions. Meeting the paradigm shift in toxicology towards an animal-free chemical risk assessment. ALTEX 37, 607–622. https://doi.org/10.14573/altex.2003242.
Rauch, C., Feifel, E., Kern, G., Murphy, C., Meier, F., Parson, W., Beilmann, M., Jennings, P., Gstraunthaler, G., Wilmes, A., 2018. Differentiation of human iPSCs into functional podocytes. PLoS One 13. https://doi.org/10.1371/journal.pone.0203869.
Saravu, K., Sekhar, S., Pai, A., Barkur, A.S., Rajesh, V., Earla, J.R., 2013. Paraquat – a deadly poison: report of a case and review. Indian J. Crit. Care Med. 17, 182–184. https://doi.org/10.4103/0972-5229.117074.
Schrijvers, B.F., Flyvbjerg, A., De Vriese, A.S., 2004. The role of vascular endothelial growth factor (VEGF) in renal pathophysiology. Kidney Int. 65, 2003–2017. https://doi.org/10.1111/j.1523-1755.2004.00621.x.
Sevin, E., Dehouck, L., Versele, R., Culot, M., Gosselet, F., 2019. A miniaturized pump out method for characterizing molecule interaction with ABC transporters. Int. J. Mol. Sci. 20 https://doi.org/10.3390/ijms20225529.
Sewell, F., Aggarwal, M., Bachler, G., Broadmeadow, A., Gellatly, N., Moore, E., Robinson, S., Rooseboom, M., Stevens, A., Terry, C., Burden, N., 2017. The current status of exposure-driven approaches for chemical safety assessment: a cross-sector perspective. Toxicology 389, 109–117. https://doi.org/10.1016/j.tox.2017.07.018.
Singh, P., Chandrasekaran, V., Hardy, B., Wilmes, A., Jennings, P., Exner, T.E., 2021. Temporal transcriptomic alterations of cadmium exposed human iPSC-derived renal proximal tubule-like cells. Toxicol. in Vitro 73.
Sittipunt, C., 2005. Paraquat Poisoning.
Smith, M.H., Ploegh, H.L., Weissman, J.S., 2011. Road to ruin: targeting proteins for degradation in the endoplasmic reticulum. Science (80-.). https://doi.org/10.1126/science.1209235.

Steimberg, N., Bertero, A., Chiono, V., Dell'Era, P., Di Angelantonio, S., Hartung, T., Perego, S., Raimondi, M., Xinaris, C., Caloni, F., de Angelis, I., Alloisio, S., Baderna, D., 2020. iPS, organoids and 3d models as advanced tools for in vitro toxicology. ALTEX 37, 136–140. https://doi.org/10.14573/altex.1911071.

Tabara, K., Iwata, Y., Koizumi, N., 2018. The unfolded protein response. Methods Mol. Biol. 1691, 223–230. https://doi.org/10.1007/978-1-4939-7389-7_17.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors, pp. 861–872. https://doi.org/10.1016/j.cell.2007.11.019.

Wang, X., Zhang, Y., Wan, X., Guo, C., Cui, J., Sun, J., Li, L., 2020. Responsive expression of MafF to β -amyloid-induced oxidative stress. Dis. Markers 2020. https://doi.org/10.1155/2020/8861358.

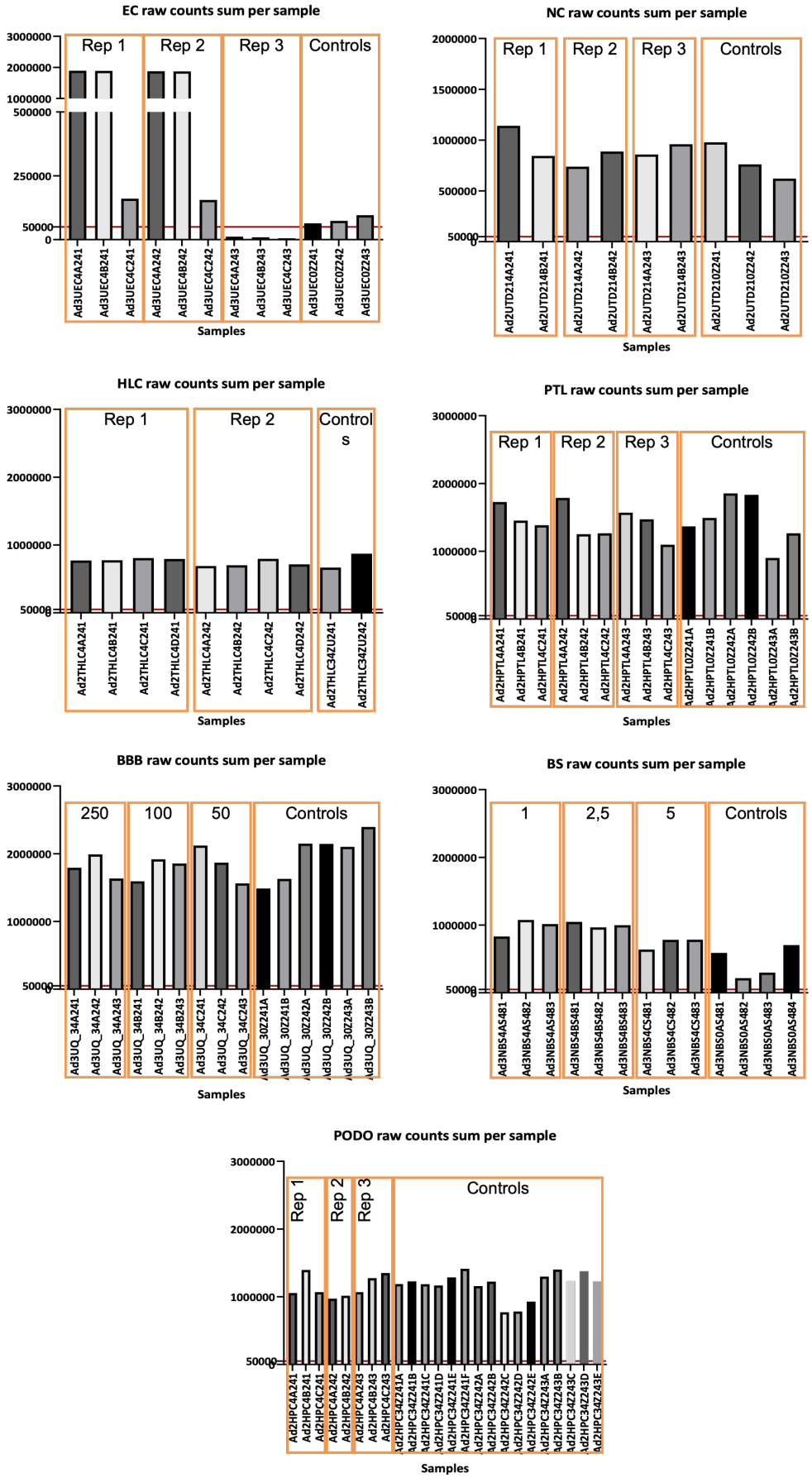
Wang, X., Wang, X., Zhu, Y., Chen, X., 2021. ADME/T-based strategies for paraquat detoxification: transporters and enzymes. Environ. Pollut. 291, 118137 https://doi.org/10.1016/j.envpol.2021.118137.

Wellens, S., Dehouck, L., Chandrasekaran, V., Singh, P., Loiola, R.A., Sevin, E., Exner, T., Jennings, P., Gosselet, F., Culot, M., 2021. Evaluation of a human iPSC-derived BBB model for repeated dose toxicity testing with cyclosporine A as model compound. Toxicol. in Vitro 73, 105112. https://doi.org/10.1016/j.tiv.2021.105112.

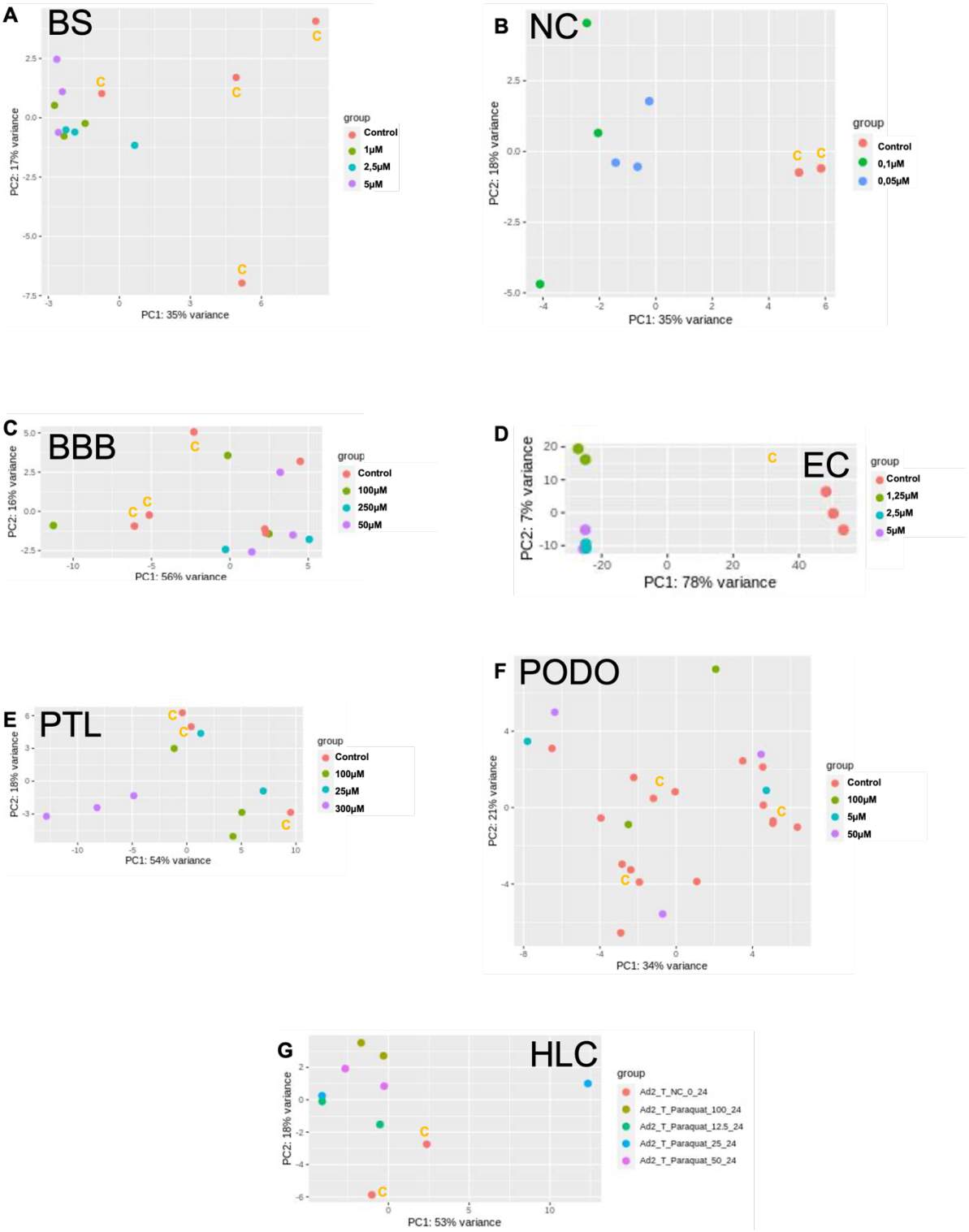
Xu, W., Li, F., Xu, Z., Sun, B., Cao, J., Liu, Y., 2017. Tert-butylhydroquinone protects PC12 cells against ferrous sulfate-induced oxidative and inflammatory injury via the Nrf2/ARE pathway. Chem. Biol. Interact. 273, 28–36. https://doi.org/10.1016/j.cbi.2017.05.021.

Zhou, T., Benda, C., Duzinger, S., Huang, Y., Li, X., Li, Y., Guo, X., Cao, G., Chen, S., Hao, L., Chan, Y.C., Ng, K.M., Ho, J.C., Wieser, M., Wu, J., Redl, H., Tse, H.F., Grillari, J., Grillari-Voglauer, R., Pei, D., Esteban, M.A., 2011. Generation of induced pluripotent stem cells from urine. J. Am. Soc. Nephrol. 22, 1221–1228. https://doi.org/10.1681/ASN.2011010106.

Supporting Information



Supplementary Fig. 1. TempO-Seq raw read counts per sample for each hiPSC-derived model



Supplementary Fig. 2. PCA of all samples per model after filtering. (A) BS (B) NC) (C) BBB (D) EC (E) PTL (F) PODO (G) HLC

Gene symbol_probe ID	LFC per model per condition (µM)																	
	BS			NC		BBB			PTL			PODO			HLC			
	1	2,5	5	0,05	0,1	50	100	250	25	100	300	5	50	100	12,5	25	50	100
MAFF_27478		x		1,88	2,31	0,02	0,72	0,78	0,39	1,64	2,82	-0,43	0,51	1,11	0,24	1,52	1,26	1,68
PPP1R15A_14098	0,49	0,50	0,85	1,03	1,86		x		0,28	1,27	2,32	0,19	0,22	0,81	0,38	1,12	0,98	1,28
VEGFA_28053	0,60	0,79	0,90	0,93	1,06		x		0,13	1,21	2,14	0,05	0,37	0,74			x	
ATF4_500		x		0,60	0,70	0,31	0,47	0,93	-0,04	0,27	1,48		x		0,25	0,42	0,30	1,04
GDF15_18329		x		1,58	1,93	0,29	0,91	1,89	1,09	3,18	6,75		x		0,97	0,84	1,27	1,67

Gene symbol_probe ID	LFC per model per condition (µM)					
	PODO			PTL		
	5	50	100	25	100	300
CD55_28297	0,49	1,01	1,51	0,51	1,09	1,27
CHKA_14315	0,26	0,44	0,69	0,06	0,43	0,75
DUSP1_24234	0,20	0,36	0,99	0,29	0,70	2,33
ENO2_13140	0,82	0,67	0,96	0,18	0,80	0,95
HSPA1B_3136	-0,35	-0,12	-0,82	0,59	0,42	1,28
MAFF_27478	-0,43	0,51	1,11	0,39	1,64	2,82
MAFG_20735	0,47	0,42	1,20	0,28	1,20	2,06
PPP1R15A_14098	0,19	0,22	0,81	0,28	1,27	2,32
VEGFA_28053	0,05	0,37	0,74	0,13	1,21	2,14

Gene symbol_probe ID	LFC per model per condition (µM)				
	BS			NC	
	1	2,5	5	0,05	0,1
CBX4_16907	0,72	0,67	0,56	0,66	1,05
EIF1_2074	0,28	0,28	0,63	0,53	0,88
PPP1R15A_14098	0,49	0,50	0,85	1,03	1,86
SLC3A2_20164	0,60	0,56	0,74	1,48	2,17
TUBA1B_7399	-1,34	-0,97	-1,40	-0,40	-0,84
VEGFA_28053	0,60	0,79	0,90	0,93	1,06

Supplementary Fig. 3. List of the DEGs shared by at least 4 models or by modeled organ (brain and kidney). Log2 fold changes (LFC) of genes presented in Table 4.

4. Conclusions and future work

Recent years have seen the development of many human cell-based tests and other non-animal methods in the chemical safety assessment (Burden et al. 2015; Singh et al. 2018). While their effective implementation is on its way, it is also clear that no single method or approach is or even should be a one-to-one replacement for animal-based standardized testing guidelines (Madden et al. 2014; Sakuratani et al. 2018). The researchers have therefore started developing approaches that integrate various non-animal tools to not only pace up the transition to more innovative strategies but also become less dependent on animals, cheaper and more human relevant (Meigs et al. 2018; 2019).

The in3 project as described earlier has focused on integrating *in vitro* methodologies with *in silico* tools using iPSC technology for mechanistic toxicity testing. All the study designs and plans for the *in vitro-based* experiments were discussed within the complete consortium of in3 to be able to align them so that the data can be reused within the project. This also led to many collaborative research studies like cell model comparisons based on the controls and PQ as mentioned in Section 1.2. For achieving the maximum benefit out of the harmonized study designs and make the reuse of data as easy as possible, this thesis focused on coordinating the partners' *in vitro* data collection, standardization, accessibility, and hosting on an optimized database and on the development of project-wide bioinformatics analysis workflows to uncover tissue specific markers of toxicity and stress response pathways. The central data management was applied and common bioinformatics based workflows for data processing and analysis were developed for proper evaluation within the project to foster the re-use of data within and outside the project. A careful metadata collection process was executed to guarantee the completeness of information for the analysis of the data in each study and for cross-study analysis. The metadata as well as the corresponding raw data files were uploaded to the EdelweissData™ data management system (SaferWorldbyDesign2020) as described in Section 1.5. A standard workflow accessing the application programming interface (API) was developed to provide and share data hosted on the data management system with everyone in the project. A bioinformatics workflow was also integrated for further processing and analysis of the data in a standardized way as described in Section 1.5. These workflows were provided to all the partners to help them in the evaluation of the iPSC cell models and their application to safety assessment, including determination of model robustness and specificity and elucidating the mechanisms related to toxicity. This common strategy led to the analysis in a collaborative and standardized manner. The standardized strategy also helped partners in utilizing this data in the grouping, read-across approaches, and providing mechanistic validation for the QSARs. Overall, this led to around 10 publications so far (both published and in preparation) within the project using the data sharing infrastructure and the bioinformatics workflows and/or extending the latter.

The above discussed workflows created as the first part of this thesis were also utilized to achieve the specific aims as mentioned in Section 1.1 which focused on the utilization of iPSCs in toxicity testing, usage of temporal data to resolve mechanisms patterns, and comparing the

cell models' responses to PQ. The work done in this thesis is an important step towards being able to use iPSCs in standardized testing for risk assessment and regulations.

The first aim was to utilize an iPSC-derived PTL cell model in chemical toxicity testing. The targeted transcriptomics (TempO-Seq) data and pathway libraries allowed the identification of stress responses caused in PTL cells on being exposed to cadmium, arsenic, amiodarone, doxorubicin, rotenone, tunicamycin, and GW788388. Different stress responses were identified in a time-dependent way, showing how they progress differently over time depending on the exposure to chemicals. The pathway libraries helped translate the differential expressed genes (DEGs) pattern (derived from the transcriptomics data analysis) to the risk-relevant stress response mechanisms. The combination of the two pathway libraries (Reactome (https://reactome.org) (Jassal et al. 2020) and Ingenuity Pathway Analysis (IPA) (https://digitalinsights.qiagen.com/products-overview/) (Krämer et al. 2014) and the transcription factor associations curated from literature (by Vidya Chandrasekaran, Vrije Universiteit Amsterdam) showed the association of the DEGs activated due to cadmium exposure to two known major stress responses: Nrf2-mediated oxidative stress response and MTF1-mediated metal stress response (Andrews 2000; Chen and Shaikh 2009). The integrated pathway library ConsensusPathDB V35 (http://cpdb.molgen.mpg.de) (Kamburov et al. 2009) showed known stress responses like UPR, Nrf2- mediated oxidative stress response, MTF1-mediated metal response, and p53 signaling pathway related to the DEGs activated in different degrees and time points in response to tunicamycin, arsenic, rotenone, doxorubicin, and amiodarone(Carvalho et al. 2009; Gjymishka et al. 2009; Nicoletti-Carvalho et al. 2010; Chakraborty et al. 2014; Rashid et al. 2015; Tsui et al. 2015; Young et al. 2016; Counts and Mufson 2017; Afsar et al. 2020). GW788388 was the only compound that showed no DEGs activated related to stress response based on the cut-offs set for both on LFCs and adjusted p-value. But even without the use of the cut-offs, no concentration-based responses were seen for the stress response related genes in GW788388.

The second aim was to evaluate the advantages of using temporal gene expression data. The temporal sequence of the mechanisms was analyzed based on the associated gene profiles due to exposure to different chemicals. The MTF1-mediated stress response related genes for both the arsenic and cadmium showed an early activation of the genes and a sustained response until the last measured time point. There was also an early activation of Nrf2-mediated oxidative stress related genes for both the compounds. But, in cadmium, the expression started going to the baseline levels by 24h and in arsenic, the response was sustained until 24h. This difference in the exposure to metal compounds would not have been possible to see if only one single time point would have been used. Additionally, a single time point is not always enough to see all responses. In the case of exposure to cadmium, the Nrf2- mediated oxidative stress response would have never been identified if only the 24h time point would have been chosen as often done in transcriptomics studies. The use of different time points before 24h showed that the genes related to this pathway reached the highest expression very early but didn't sustain the expression until 24h. In contrast, the Nrf2- mediated oxidative stress response started reaching the highest level only at the longest exposure time points in the case of doxorubicin. Thus,

whatever single time point would have been chosen, this important pathway would have been missed in the one or other chemical exposure scenario.

The profile fitting method as described in Chapter 2.1 was developed exclusively to apply in the case studies from Chapters 2 and 3 to reduce the influence of noise and, in this way, better identify similar profiles. The functions used in the fitting procedure producing the best profiles to describe the expression pattern were chosen manually for each gene. However, automating the step of choosing the best temporal profile for the gene will be useful for the easy application of this method on the broader set of time resolution data by future researchers.

The conclusion of the results addressing the first two aims which were achieved for the case studies from Chapter 2 and Chapter 3 is that the temporal-based studies utilizing iPSC models and targeted transcriptomics technology help unravel the time-based evolution of the responses induced due to chemical exposure. The targeted transcriptomics data used in TempO-Seq was useful to find the expression pattern of the most important genes and their associated responses known to be activated against the chemical exposure of all the compounds used. In future applications of this approach, the addition of the full genome information would benefit in exploring the expression pattern of the genes which were not present in this targeted gene list. The targeted gene set can miss novel transcripts which might be important in different activities happening in the associated mechanisms. Additionally, these novel transcripts may play an important role in the induction of other connected responses in cells as well.

The third aim was to analyze the activation of stress responses in different IPSC-derived models namely proximal tubule-like cells (PTL), podocytes (PODO), brain-like endothelial cells (BLECs), brain spheres (BS), neuronal cells (NC), and hepatocyte-like cell (HLC) after being exposed to PQ. TempO-Seq analysis using selection criteria (cut-offs) based on the maximum fold changes showed the induction of Nrf2-mediated oxidative stress response, which is known to be affected by PQ (Dinis-Oliveira et al. 2008; Gawarammana and Buckley 2011; Blanco-Ayala et al. 2014), in all the models except for BS. But the associated genes like HMOX1 and NQO1 showed a concentration-dependent increase for all the six models, even for BS, but with different sensitivities. PTL showed the highest expression levels of the genes associated with the oxidative stress response. UPR also known to be induced on PQ exposure (Chinta et al. 2008; Omura et al. 2013) was activated in all models except for BS and PODO. Like oxidative stress, BS showed a concentration-dependent increase, but the genes and the associated pathway didn't cross the cut-offs. In contrast, the UPR-related genes like PPP1R15A did not show the concentration-dependent increase in PODO and very weak expression in general suggesting that this mechanism is not or less important in PODO.

This case study was based on the re-use of the data generated as part of studies performed by other in3 partners, which means that the experiments done for each cell model were performed based on different study goals and not exclusively for this case study. Therefore, even if the same cut-offs were used to choose the DEGs and associated responses across all cell models, the comparison of absolute values of expression levels should be done with care as the concentrations chosen for these models are not harmonized. For the future application of this

approach, the full concentration-response curves would be needed requiring more than 3 concentrations and a broader concentration range that can show the response until saturation levels are reached. Also, the distribution kinetics of PQ must be established for each model, to harmonize the concentrations better, and correlate the observed toxic effects to the effective concentrations reaching the cells and not to the nominal concentrations used here. Additionally, temporal information along with the different concentrations can be also used to analyze responses of different cell models based on both the time and concentration.

The fourth aim was to explore the high sensitivity of the brain towards PQ which was seen while analyzing the results addressing the third aim. To evaluate the sensitivity of these models towards PQ, a wide range of PQ concentrations for 24h and/or 48h were used to expose these different cell models and establish the cytotoxicity curves. The brain was found to be the most sensitive model to PQ based on these curves. To investigate this finding, the expression level of PQ transporters was looked upon. SLC3A2 and SLC7A11, are two amino acid transporters that are used by PQ to enter cells for review, see Wang et al., 2021) (Kuter et al. 2010). showed high expression levels in both the models BS and NC representing the brain which might be the reason for the higher uptake of PQ by these cells.

The conclusion of the results addressing aims three and four for the case study from Chapter 4 is that this strategy can be applied to evaluate the toxicity of a chemical in different human-derived cell types simultaneously. This is also helpful in the assessment of differential sensitivity of cell models representing different organs. The future application of this approach in chemical toxicity assessment can also be supplemented with iPSC models for other organs like the heart and lungs.

One limiting factor in the studies was the usage of a very limited number of replicates. It is very necessary to have more than 3 replicates for a group (treatment or control) in the experimental setup to avoid losing data. In the PQ case study, the EC cell model was removed due to dissimilarity among replicates from the same treatment group as mentioned in detail in Chapter 4. The results were not interpretable as the statistics with 3 replicates were biased due to huge variance and impossibile with just 2 replicates. The only solution would have been to repeat the experiments for the EC model which was not possible within the timeframe of the project.

The work done in this thesis can be integrated into future applications with additional further developments or modifications as mentioned below:

1. The cell models can be further improved to represent the corresponding organs more closely. The BS model is multicellular and highly complex and does not (at this point in development) possess microglia, precluding the development of the full neuroinflammatory cascade. The HLC is not completely mature hepatocytes and lacks cooperation with Kupfer cells, and the kidney model PODO lacks fenestrated endothelial cells and a glomerular basement membrane to form a full glomerular model. Co-culturing cells with their missing partners in the future can be one possible solution

to this issue. Also, these cell models which are *in vitro* represent a reductionist approach, and the physiology of a whole organism is not fully replicated due to the lack of connections and communications between them through blood flow, endocrine and nervous systems. In the future, the use of organ-on-a-chip models and/or the application and integration of *in silico* modeling can recapitulate it better.

2. The time points used for the first 24h in the cadmium study helped look at the first exposure and the first 24h. However, for looking at long(er)-term effects and multi-dosing, a similar resolution as in the first 24h would be needed. In the cadmium study, there were only two time points after 24h used in the study presented here to interpolate the profile to include the long-term effects of cadmium exposure to PTL cells. The cells were repeatedly exposed after every 24h. However, the behavior of genes in the first 24h was very different as compared to 72h and 168h. For example, Nrf2 related genes came to baseline levels by 24h and did not show any activation at 72h and 168h. Genes like PCNA and SESN1 showed no signal until 24h but showed higher expression levels at 72h and 168h. This higher expression was not caused due to the increase in the raw counts of treated samples but due to the decrease in the read counts of the control samples. The course of the mechanism over time between 24h and 72h as well as between 72h and 168h was not identifiable due to no time points in between these two ranges. Also, for genes like MAFF and MAFG which came to baseline levels by 24h and didn't show any activation at 72h and 168h, there is a chance to miss important information if the addition of cadmium after 24h causes another maximum as seen before 24h. Therefore, this would need more investigations including more time points between 24h and 168h to better understand the effect of multi-dosing and analyze long-term effects.

3. a) The pathway libraries like ConsensusPathDB integrates information from all pathway libraries into one large library. This leads to repetitions of many pathways for example in the case of doxorubicin, 10 pathways showed relation to the p53 signaling pathway, genotoxicity, or cell death/apoptosis. Some common sets of genes were shared by these different representations. But many additional genes were added to some pathways highlighting the area focused by the research group generating these entries, the generality or specificness of a pathway, and the knowledge at the point of time. But this can further be improved by subgrouping the related pathways coming from different pathway libraries in a nested manner under one common name representing the pathway of interest.

   b) The pathway libraries need to be made more toxicology relevant for their broader application in risk assessment. The transcriptomic information related to exposed cells can help to do that. For example, genes, which never show up in any exposure scenario, could be removed from a tox-optimized pathway. Also, using the temporal resolution information related to risk assessment can be advantageous. For example, the temporal resolution might help to better structure biological pathways to represent the temporal relationships of the involved genes, proteins, etc. Adding more data might make it

possible to group or divide the genes related to a pathway based on the time points of their occurrence when comparing responses due to different chemicals.

5. References

Afsar T, Razak S, Almajwal A, Al-Disi D (2020) Doxorubicin-induced alterations in kidney functioning, oxidative stress, DNA damage, and renal tissue morphology; Improvement by Acacia hydaspica tannin-rich ethyl acetate fraction. Saudi J Biol Sci 27:. https://doi.org/10.1016/j.sjbs.2020.07.011

Agresti A (2006) An Introduction to Categorical Data Analysis: Second Edition

Alépée N, Grandidier MH, Cotovio J (2014) Sub-categorisation of skin corrosive chemicals by the EpiSkinTM reconstructed human epidermis skin corrosion test method according to UN GHS: Revision of OECD Test Guideline 431. Toxicol Vitr 28:. https://doi.org/10.1016/j.tiv.2013.10.016

Anadón A, Martínez MA, Castellano V, Martínez-Larrañaga MR (2014) The role of in vitro methods as alternatives to animals in toxicity testing. Expert Opin. Drug Metab. Toxicol. 10

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:. https://doi.org/10.1186/gb-2010-11-10-r106

Andrews GK (2000) Regulation of metallothionein gene expression by oxidative stress and metal ions. In: Biochemical Pharmacology

Anon (2014) Metadata Standards, International Organization for Standardization. In: Work. Gr. SO/IEC JTC1 SC32. http://metadata-standards.org/. Accessed 17 Mar 2020

Barnett LMA, Cummings BS (2018) Nephrotoxicity and renal pathophysiology: A contemporary perspective. Toxicol. Sci. 164:379–390

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B 57:. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blanco-Ayala T, Andérica-Romero AC, Pedraza-Chaverri J (2014) New insights into antioxidant strategies against paraquat toxicity. Free Radic Res 48:623–640. https://doi.org/10.3109/10715762.2014.899694

Burden CJ, Qureshi SE, Wilson SR (2014) Error estimates for the analysis of differential expression from RNA-seq count data. PeerJ 2:e576. https://doi.org/10.7717/peerj.576

Burden N, Sewell F, Chapman K (2015) Testing Chemical Safety: What Is Needed to Ensure the Widespread Application of Non-animal Approaches? PLoS Biol 13:. https://doi.org/10.1371/journal.pbio.1002156

Bushel PR, Paules RS, Auerbach SS (2018) A Comparison of the TempO-Seq S1500+ Platform to RNA-Seq and Microarray Using Rat Liver Mode of Action Samples. Front Genet 9:. https://doi.org/10.3389/fgene.2018.00485

Carusi A, Davies MR, De Grandis G, et al (2018) Harvesting the promise of AOPs: An assessment and recommendations. Sci. Total Environ. 628–629

Carvalho C, Santos R, Cardoso S, et al (2009) Doxorubicin: The Good, the Bad and the Ugly Effect. Curr Med Chem 16:. https://doi.org/10.2174/092986709788803312

Chakraborty A, Mondal C, Sinha S, et al (2014) Amiodarone induced oxidative stress in stress - Vulnerable organs of adult male rats. Asian J Pharm Clin Res 7:

Chandrasekaran V, Carta G, da Costa Pereira D, et al (2021) Generation and characterization of iPSC-derived renal proximal tubule-like cells with extended stability. Sci Rep 11:11575. https://doi.org/10.1038/s41598-021-89550-4

Chen J, Shaikh ZA (2009) Activation of Nrf2 by cadmium and its role in protection against cadmium-induced apoptosis in rat kidney cells. Toxicol Appl Pharmacol 241:. https://doi.org/10.1016/j.taap.2009.07.038

Chinta SJ, Rane A, Poksay KS, et al (2008) Coupling endoplasmic reticulum stress to the cell death program in dopaminergic cells: Effect of paraquat. NeuroMolecular Med 10:. https://doi.org/10.1007/s12017-008-8047-9

Coady K, Browne P, Embry M, et al (2019) When Are Adverse Outcome Pathways and Associated Assays “Fit for Purpose” for Regulatory Decision-Making and Management of Chemicals? Integr Environ Assess Manag 15:. https://doi.org/10.1002/ieam.4153

Conolly RB, Ankley GT, Cheng W, et al (2017) Quantitative Adverse Outcome Pathways and Their Application to Predictive Toxicology. Environ Sci Technol 51:. https://doi.org/10.1021/acs.est.6b06230

Conway JR, Lex A, Gehlenborg N (2017) UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics 33:2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Counts SE, Mufson EJ (2017) Regulator of cell cycle (RGCC) expression during the progression of Alzheimer’s disease. Cell Transplant 26:. https://doi.org/10.3727/096368916X694184

Cui Y, Paules RS (2010) Use of transcriptomics in understanding mechanisms of drug-induced toxicity. Pharmacogenomics 11

Dinis-Oliveira RJ, Duarte JA, Sánchez-Navarro A, et al (2008) Paraquat poisonings: Mechanisms of lung toxicity, clinical features, and treatment. Crit. Rev. Toxicol. 38

Doke SK, Dhawale SC (2015) Alternatives to animal testing: A review. Saudi Pharm J SPJ Off Publ Saudi Pharm Soc 23:223–229. https://doi.org/10.1016/j.jsps.2013.11.002

Dou T, Yan M, Wang X, et al (2016) Nrf2/ARE Pathway Involved in Oxidative Stress Induced by Paraquat in Human Neural Progenitor Cells. Oxid Med Cell Longev 2016:. https://doi.org/10.1155/2016/8923860

ECHA (2020) The Use of Alternatives to Testing on Animals for the REACH Regulation

Eixarch H, Wyness L, Sibanda M (2019) The regulation of personalized cosmetics in the EU. Cosmetics 6:. https://doi.org/10.3390/COSMETICS6020029

Ernst J, Nau GJ, Bar-Joseph Z (2005) Clustering short time series gene expression data. Bioinformatics 21:. https://doi.org/10.1093/bioinformatics/bti1022

Fentem J, Malcomber I, Maxwell G, Westmoreland C (2021) Upholding the EU's Commitment to "Animal Testing as a Last Resort" Under REACH Requires a Paradigm Shift in How We Assess Chemical Safety to Close the Gap Between Regulatory Testing and Modern Safety Science. Altern Lab Anim 49:. https://doi.org/10.1177/02611929211040824

Fitzi-Rathgen J (2019) The 3Rs and replacement methods - Better research, less animal harm. ALTEX 36

Fröhlich E (2017) Role of omics techniques in the toxicity testing of nanoparticles. J. Nanobiotechnology 15

Garcia-Reyero N, Ekman DR, Habib T, et al (2014a) Integrated approach to explore the mechanisms of aromatase inhibition and recovery in fathead minnows (Pimephales promelas). Gen Comp Endocrinol 203:. https://doi.org/10.1016/j.ygcen.2014.03.022

Garcia-Reyero N, Tingaud-Sequeira A, Cao M, et al (2014b) Endocrinology: Advances through omics and related technologies. Gen Comp Endocrinol 203:. https://doi.org/10.1016/j.ygcen.2014.03.042

Gautier F, Tourneix F, Assaf Vandecasteele H, et al (2020) Read-across can increase confidence in the Next Generation Risk Assessment for skin sensitisation: A case study with resorcinol. Regul Toxicol Pharmacol 117:. https://doi.org/10.1016/j.yrtph.2020.104755

Gawarammana IB, Buckley NA (2011) Medical management of paraquat ingestion. Br J Clin Pharmacol 72:. https://doi.org/10.1111/j.1365-2125.2011.04026.x

Gjymishka A, Su N, Kilberg MS (2009) Transcriptional induction of the human asparagine synthetase gene during the unfolded protein response does not require the ATF6 and IRE1/XBP1 arms of the pathway. Biochem J 417:. https://doi.org/10.1042/BJ20081706

GreenFacts (2022), Alternatives to animal testing for safety assessment of chemicals in Europe. https://www.greenfacts.org/en/reach-animal-testing-alternatives/l-2/index.htm. Accessed April 2022

Harrill JA, Everett LJ, Haggard DE, et al (2021) High-Throughput Transcriptomics Platform for Screening Environmental Chemicals. Toxicol Sci 181:. https://doi.org/10.1093/toxsci/kfab009

Hawkins F, Kotton DN (2015) Embryonic and induced pluripotent stem cells for lung regeneration. In: Annals of the American Thoracic Society

Houck KA, Kavlock RJ (2008) Understanding mechanisms of toxicity: Insights from drug discovery research. Toxicol. Appl. Pharmacol. 227

Hu N, Zhang H (2018) CYP24A1 depletion facilitates the antitumor effect of vitamin D3 on thyroid cancer cells. Exp Ther Med 16:. https://doi.org/10.3892/etm.2018.6536

Hubrecht RC, Carter E (2019) The 3Rs and humane experimental technique: Implementing change. Animals 9:. https://doi.org/10.3390/ani9100754

Jassal B, Matthews L, Viteri G, et al (2020) The reactome pathway knowledgebase. Nucleic Acids Res 48:. https://doi.org/10.1093/nar/gkz1031

Jennings P (2015) “The future of in vitro toxicology.” Toxicol Vitr 29:1217–1221. https://doi.org/10.1016/j.tiv.2014.08.011

Jennings P, Limonciel A, Felice L, Leonard MO (2013) An overview of transcriptional regulation in response to toxicological insult. Arch. Toxicol. 87

Jennings P, Schwarz M, Landesmann B, et al (2014) SEURAT-1 liver gold reference compounds: a mechanism-based review. Arch. Toxicol. 88

Joseph P (2017) Transcriptomics in toxicology. Food Chem Toxicol 109:650–662. https://doi.org/10.1016/j.fct.2017.07.031

Kamburov A, Wierling C, Lehrach H, Herwig R (2009) ConsensusPathDB - A database for integrating human functional interaction networks. Nucleic Acids Res 37:. https://doi.org/10.1093/nar/gkn698

Kim J, Koo BK, Knoblich JA (2020) Human organoids: model systems for human biology and medicine. Nat. Rev. Mol. Cell Biol. 21

Knight DJ, Deluyker H, Chaudhry Q, et al (2021) A call for action on the development and implementation of new methodologies for safety assessment of chemical-based products in the EU – A short communication. Regul Toxicol Pharmacol 119:. https://doi.org/10.1016/j.yrtph.2020.104837

Kohonen P, Benfenati E, Bower D, et al (2013) The ToxBank data warehouse: Supporting the replacement of in vivo repeated dose systemic toxicity testing. Mol Inform 32:. https://doi.org/10.1002/minf.201200114

Krämer A, Green J, Pollard J, Tugendreich S (2014) Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 30:. https://doi.org/10.1093/bioinformatics/btt703

Krewski D, Andersen ME, Tyshenko MG, et al (2020) Toxicity testing in the 21st century: progress in the past decade and future perspectives. Arch. Toxicol. 94

Kuter K, Nowak P, Gołembiowska K, Ossowska K (2010) Increased reactive oxygen species production in the brain after repeated low-dose pesticide paraquat exposure in rats. A comparison with peripheral tissues. Neurochem Res 35:. https://doi.org/10.1007/s11064-010-0163-x

Kutmon M, van Iersel MP, Bohler A, et al (2015) PathVisio 3: An Extendable Pathway Analysis Toolbox. PLoS Comput Biol 11:1–13. https://doi.org/10.1371/journal.pcbi.1004085

Limonciel A, Ates G, Carta G, et al (2018) Comparison of base-line and chemical-induced transcriptomic responses in HepaRG and RPTEC/TERT1 cells using TempO-Seq. Arch Toxicol 92:2517–2531. https://doi.org/10.1007/s00204-018-2256-2

Love MI, Huber W, Anders S (2014a) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Love MI, Huber W, Anders S (2014b) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Lowe R, Shirley N, Bleackley M, et al (2017) Transcriptomics technologies. PLoS Comput Biol 13:. https://doi.org/10.1371/journal.pcbi.1005457

Ma LJ, Fogo AB (2009) PAI-1 and kidney fibrosis. Front Biosci 14:. https://doi.org/10.2741/3361

Macmillan DS, Chilton ML (2019) A defined approach for predicting skin sensitisation hazard and potency based on the guided integration of in silico, in chemico and in vitro data using exclusion criteria. Regul Toxicol Pharmacol 101:. https://doi.org/10.1016/j.yrtph.2018.11.001

Madden JC, Enoch SJ, Paini A, Cronin MTD (2020) A Review of In Silico Tools as Alternatives to Animal Testing: Principles, Resources and Applications. Altern. Lab. Anim. 48

Madden JC, Rogiers V, Vinken M (2014) Application of in silico and in vitro methods in the

development of adverse outcome pathway constructs in wildlife. Philos. Trans. R. Soc. B Biol. Sci. 369

Maestri E (2021) The 3rs principle in animal experimentation: A legal review of the state of the art in europe and the case in Italy. BioTech 10:. https://doi.org/10.3390/biotech10020009

Mahony C, Ashton RS, Birk B, et al (2020) New ideas for non-animal approaches to predict repeated-dose systemic toxicity: Report from an EPAA Blue Sky Workshop. In: Regulatory Toxicology and Pharmacology

Mav D, Shah RR, Howard BE, et al (2018) A hybrid gene selection approach to create the S1500+ targeted gene sets for use in high-throughput transcriptomics. PLoS One 13:. https://doi.org/10.1371/journal.pone.0191105

Meek ME, Boobis A, Cote I, et al (2014) New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol 34:. https://doi.org/10.1002/jat.2949

Meigs L, Smirnova L, Rovida C, et al (2018) Animal testing and its alternatives - the most important omics is economics. ALTEX 35:. https://doi.org/10.14573/altex.1807041

Miller-Hodges E, Hohenstein P (2012) WT1 in disease: Shifting the epithelial-mesenchymal balance. J. Pathol. 226

Moné MJ, Pallocca G, Escher SE, et al (2020) Setting the stage for next-generation risk assessment with non-animal approaches: the EU-ToxRisk project experience. Arch. Toxicol. 94

Monosson E (2005) Chemical mixtures: Considering the evolution of toxicology and chemical assessment. Environ. Health Perspect. 113

Morrison M, Klein C, Clemann N, et al (2015) StemBANCC: Governing Access to Material and Data in a Large Stem Cell Research Consortium. Stem Cell Rev Reports 11:. https://doi.org/10.1007/s12015-015-9599-3

Moyano P, Sanjuan J, García JM, et al (2020) Primary hippocampal estrogenic dysfunction induces synaptic proteins alteration and neuronal cell death after single and repeated paraquat exposure. Food Chem Toxicol 136:. https://doi.org/10.1016/j.fct.2019.110961

Murphy C, Feifel E, Jennings P, et al (2019) A Protocol for One-Step Differentiation of Human Induced Pluripotent Stem Cells into Mature Podocytes. Methods Mol Biol 1994:93–99. https://doi.org/10.1007/978-1-4939-9477-9_8

National Research Council (US) Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology (2007) Applications of toxicogenomic technologies to predictive toxicology and risk assessment

Nicoletti-Carvalho JE, Nogueira TCA, Gorjão R, et al (2010) UPR-mediated TRIB3 expression correlates with reduced AKT phosphorylation and inability of interleukin 6 to overcome palmitate-induced apoptosis in RINm5F cells. J Endocrinol 206:. https://doi.org/10.1677/JOE-09-0356

Okawa H, Motohashi H, Kobayashi A, et al (2006) Hepatocyte-specific deletion of the keap1 gene activates Nrf2 and confers potent resistance against acute drug toxicity. Biochem Biophys Res Commun 339:79–88. https://doi.org/10.1016/j.bbrc.2005.10.185

Omole AE, Fakoya AOJ (2018) Ten years of progress and promise of induced pluripotent stem cells: Historical origins, characteristics, mechanisms, limitations, and potential applications. PeerJ 2018:. https://doi.org/10.7717/peerj.4370

Omura T, Asari M, Yamamoto J, et al (2013) Sodium tauroursodeoxycholate prevents paraquat-induced cell death by suppressing endoplasmic reticulum stress responses in human lung epithelial A549 cells. Biochem Biophys Res Commun 432:. https://doi.org/10.1016/j.bbrc.2013.01.131

Pamies D, Bal-Price A, Chesné C, et al (2018) Advanced Good Cell Culture Practice for human primary, stem cell-derived and organoid models as well as microphysiological systems. ALTEX 35:353–378. https://doi.org/10.14573/altex.1710081

Patlewicz G, Ball N, Boogaard PJ, et al (2015) Building scientific confidence in the development and evaluation of read-across. Regul Toxicol Pharmacol 72:. https://doi.org/10.1016/j.yrtph.2015.03.015

Patlewicz G, Kuseva C, Kesova A, et al (2014) Towards AOP application - Implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. Regul Toxicol Pharmacol 69:. https://doi.org/10.1016/j.yrtph.2014.06.001

Perdomo-Morales R, Pardo-Ruiz Z, Spreitzer I, et al (2011) Monocyte activation test (MAT) reliably detects pyrogens in parenteral formulations of human serum albumin. ALTEX 28:. https://doi.org/10.14573/altex.2011.3.227

Perkins EJ, Ankley GT, Crofton KM, et al (2013) Current perspectives on the use of alternative species in human health and ecological hazard assessments. Environ. Health Perspect. 121

Petersen M, Thorikay M, Deckers M, et al (2008) Oral administration of GW788388, an inhibitor of TGF-β type I and II receptor kinases, decreases renal fibrosis. Kidney Int 73:. https://doi.org/10.1038/sj.ki.5002717

Pistollato F, Bremer-Hoffmann S, Healy L, et al (2012) Standardization of pluripotent stem cell cultures for toxicity testing. Expert Opin. Drug Metab. Toxicol. 8

Price PS, Jarabek AM, Burgoon LD (2020) Organizing mechanism-related information on

chemical interactions using a framework based on the aggregate exposure and adverse outcome pathways. Environ Int 138:. https://doi.org/10.1016/j.envint.2020.105673

Prinsen MK, Hendriksen CFM, Krul CAM, Woutersen RA (2017) The Isolated Chicken Eye test to replace the Draize test in rabbits. Regul Toxicol Pharmacol 85:. https://doi.org/10.1016/j.yrtph.2017.01.009

Punt A, Bouwmeester H, Blaauboer BJ, et al (2020) New Approach Methodologies (NAMs) for Human-Relevant Biokinetics Predictions: Meeting the Paradigm Shift in Toxicology Towards an Animal-Free Chemical Risk Assessment. ALTEX 37:. https://doi.org/10.14573/altex.2003242

Qi W, Chen X, Gilbert RE, et al (2007) High glucose-induced thioredoxin-interacting protein in renal proximal tubule cells is independent of transforming growth factor-β1. Am J Pathol 171:. https://doi.org/10.2353/ajpath.2007.060813

Ramšak Ž, Modic V, Li RA, et al (2022) From Causal Networks to Adverse Outcome Pathways: A Developmental Neurotoxicity Case Study. Front Toxicol 4:. https://doi.org/10.3389/ftox.2022.815754

Rashid HO, Yadav RK, Kim HR, Chae HJ (2015) ER stress: Autophagy induction, inhibition and selection. Autophagy 11:. https://doi.org/10.1080/15548627.2015.1091141

Rauch C, Jennings P, Wilmes A (2014) Use of Induced Pluripotent Stem Cells in Drug Toxicity Screening

REACH Legislation - ECHA (no date). Available at: https://echa.europa.eu/regulations/reach/legislation (Accessed: 20 April 2022)

Reichard JF, Motz GT, Puga A (2007) Heme oxygenase-1 induction by NRF2 requires inactivation of the transcriptional repressor BACH1. Nucleic Acids Res 35:7074–7086. https://doi.org/10.1093/nar/gkm638

Richard AM, Huang R, Waidyanatha S, et al (2021) The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. Chem. Res. Toxicol. 34

Rovida C, Hartung T (2009) Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals - A report by the transatlantic think tank for toxicology (t4). ALTEX 26:. https://doi.org/10.14573/altex.2009.3.187

Sabolic I, Breljak D, Skarica M, Herak-Kramberger CM (2010) Role of metallothionein in cadmium traffic and toxicity in kidneys and other mammalian organs. Biometals 23:897–926. https://doi.org/10.1007/s10534-010-9351-z

SaferWorldbyDesign (2022), https://saferworldbydesign.com. Accessed November 2020

Sakairi T, Abe Y, Kopp JB (2011) TGF-beta1 reduces Wilms' tumor suppressor gene expression in podocytes. Nephrol Dial Transplant 26:. https://doi.org/10.1093/ndt/gfr061

Sakuratani Y, Horie M, Leinala E (2018) Integrated Approaches to Testing and Assessment: OECD Activities on the Development and Use of Adverse Outcome Pathways and Case Studies. Basic Clin. Pharmacol. Toxicol. 123

Schultz IR, Watanabe KH (2018) The development of quantitative AOPs. In: A Systems Biology Approach to Advancing Adverse Outcome Pathways for Risk Assessment

Scott CW, Peters MF, Dragan YP (2013) Human induced pluripotent stem cells and their use in drug discovery for toxicity testing. Toxicol. Lett. 219

Shi Y, Inoue H, Wu JC, Yamanaka S (2017) Induced pluripotent stem cell technology: A decade of progress. Nat. Rev. Drug Discov. 16

Singh P (2020) Total erworbene Kreditpunkte : 10

Singh P, Chandrasekaran V, Hardy B, et al (2021) Temporal transcriptomic alterations of cadmium exposed human iPSC-derived renal proximal tubule-like cells. Toxicol Vitr 76:. https://doi.org/10.1016/j.tiv.2021.105229

Singh S, Khanna VK, Pant AB (2018) Development of In Vitro Toxicology: A Historic Story. In: In Vitro Toxicology

Steeg R, Neubauer JC, Müller SC, et al (2020) The EBiSC iPSC bank for disease studies. Stem Cell Res 49:. https://doi.org/10.1016/j.scr.2020.102034

Su Z, Łabaj PP, Li S, et al (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol 32:. https://doi.org/10.1038/nbt.2957

Suter-Dick L, Alves PM, Blaauboer BJ, et al (2015) Stem cell-derived systems in toxicology assessment. Stem Cells Dev. 24

Svingen T, Villeneuve DL, Knapen D, et al (2021) A Pragmatic Approach to Adverse Outcome Pathway Development and Evaluation. Toxicol. Sci. 184

Takahashi K, Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126:. https://doi.org/10.1016/j.cell.2006.07.024

Taylor K (2018) Ten years of reach - an animal protection perspective. ATLA Altern. to Lab. Anim. 46

Tollefsen KE, Scholz S, Cronin MT, et al (2014) Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). Regul

Toxicol Pharmacol 70:. https://doi.org/10.1016/j.yrtph.2014.09.009

Törnqvist E, Annas A, Granath B, et al (2014) Strategic focus on 3R principles reveals major reductions in the use of animals in pharmaceutical toxicity testing. PLoS One 9:. https://doi.org/10.1371/journal.pone.0101638

Tsui KH, Hsu SY, Chung LC, et al (2015) Growth differentiation factor-15: A p53- and demethylation-upregulating gene represses cell proliferation, invasion, and tumorigenesis in bladder carcinoma cells. Sci Rep 5:. https://doi.org/10.1038/srep12870

UN (2015) Transforming our world: The 2030 Agenda for sustainable development. A/RES/70/1. In. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E. Accessed April 2022

Vinken M (2013) The adverse outcome pathway concept: A pragmatic tool in toxicology. Toxicology 312

Virtanen P, Gommers R, Oliphant TE, et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

Viswanathan GA, Seto J, Patil S, et al (2008) Getting Started in Biological Pathway Construction and Analysis. PLoS Comput Biol 4:. https://doi.org/10.1371/journal.pcbi.0040016

Wellens S, Dehouck L, Chandrasekaran V, et al (2021) Evaluation of a human iPSC-derived BBB model for repeated dose toxicity testing with cyclosporine A as model compound. Toxicol Vitr 105112. https://doi.org/https://doi.org/10.1016/j.tiv.2021.105112

Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al (2016) Comment: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:. https://doi.org/10.1038/sdata.2016.18

Wilmes A, Crean D, Aydin S, et al (2011a) Identification and dissection of the Nrf2 mediated oxidative stress pathway in human renal proximal tubule toxicity. Toxicol Vitr 25:613–622. https://doi.org/10.1016/j.tiv.2010.12.009

Wilmes A, Crean D, Aydin S, et al (2011b) Identification and dissection of the Nrf2 mediated oxidative stress pathway in human renal proximal tubule toxicity. Toxicol Vitr 25:613–622. https://doi.org/10.1016/j.tiv.2010.12.009

Wilmes A, Jennings P (2015) The Use of Renal Cell Culture for Nephrotoxicity Investigations. Wiley-VCH Verlag GmbH & Co. KGaA

Wilmes A, Limonciel A, Aschauer L, et al (2013) Application of integrated transcriptomic, proteomic and metabolomic profiling for the delineation of mechanisms of drug induced

cell stress. J Proteomics 79:180–194. https://doi.org/10.1016/j.jprot.2012.11.022

Yan C, Boyd DD (2006) ATF3 regulates the stability of p53: A link to cancer. Cell Cycle 5

Yates MS, Tran QT, Dolan PM, et al (2009) Genetic versus chemoprotective activation of Nrf2 signaling: overlapping yet distinct gene expression profiles between Keap1 knockout and triterpenoid-treated mice. Carcinogenesis 30:1024–1031. https://doi.org/10.1093/carcin/bgp100

Yeakley JM, Shepard PJ, Goyena DE, et al (2017) A Trichostatin a expression signature identified by TempO-Seq targeted whole transcriptome profiling. PLoS One 12:. https://doi.org/10.1371/journal.pone.0178302

Young SK, Shao Y, Bidwell JP, Wek RC (2016) Nuclear matrix protein 4 is a novel regulator of ribosome biogenesis and controls the unfolded protein response via repression of Gadd34 expression. J Biol Chem 291:. https://doi.org/10.1074/jbc.M116.729830

Zolfaghari F, Khosravi H, Shahriyari A, et al (2019) Hierarchical cluster analysis to identify the homogeneous desertification management units. PLoS One 14:1–21. https://doi.org/10.1371/journal.pone.0226355

(2019) Research and Testing Without Animals: Where Are We Now and Where Are We Heading? In: Animal Experimentation: Working Towards a Paradigm Change