USING TRANSCRIPTOMICS TO INFORM THE NEXT GENERATION RISK ASSESSMENT OF DOXORUBICIN AND NIACINAMIDE

A thesis submitted to the University of Nottingham for the degree of MRes in Bioinformatics

Elizabeth Ann Tulum

September 2023
Supervisors: Dr Nathan Archer and Dr Adam Blanchard

ABSTRACT

Human health risk assessments without generating new animal data have resulted in extreme efforts over the past few decades from industry, academia, and regulatory bodies to develop and apply new method approaches. Various non-animal approaches for chemical hazard characterisation has appeared known as new approach methodologies (NAMs) which form the basis of integrated testing and assessment strategies designed to prevent harm to human health.

A bioinformatics workflow was developed to aid the exploration of Next Generation Risk Assessment (NGRA) to determine whether NAMs can be used for safety decisions using various consumer products. The analysis within this thesis continues the analysis conducted in the previous case studies and explores the key concepts relating to HepaRG, HepG2 and MCF-7 cells dosed with doxorubicin and niacinamide. The aim of this thesis has been achieved by addressing a gap of knowledge within Unilever by conducting biological interpretation of exposure to chemicals at different concentrations and the general use of *in vitro* methods for non-animal risk assessments using high throughput transcriptomics (HTTr) data. By looking at doxorubicin and niacinamide, this has enabled the interpretation of these chemicals using developed methodologies for future analysis of potential chemicals for NGRA. The bioinformatics workflow consisted of using R Studio to align, quantify and conduct differential expression analysis using DESeq2 for all chemical concentrations and cell lines. Furthermore, BMDExpress has been successfully conducted to derive Points of Departure (PoDs) for doxorubicin and niacinamide as well as pathway investigation alongside Ingenuity Pathway Analysis (IPA) to determine mechanisms of action (MoA) for the chemicals of interest.

At least 20 pathways were detected to apply the pathway-level tests. Using these selections, the observed pathway-level PoD ranged from 0.0219µM to 0.4854µM for doxorubicin and 2552.94µM to 24977.10µM for niacinamide across cell lines.
The selected gene or pathway PoDs were derived using BMDExpress2 for HepaRG, HepG2 and MCF-7 cells dosed with doxorubicin were 0.1557µM, 0.0313µM & 0.0219µM respectively. The selected PoDs for HepaRG, HepG2 and MCF-7 cells dosed with niacinamide were 2552.94µM, 5046.89µM & 5444.65µM respectively.

Doxorubicin is a medication that belongs to the anthracycline class of medications. It works by slowing or stopping the growth of cancer cells in your

body interfering with the function of DNA. Niacinamide plays a significant role in DNA repair, maintenance of genomic stability and cellular responses to injury including inflammation and apoptosis (cell death). For both chemicals these modes of action have been identified in the BMDExpress and IPA analysis.

Both BMDExpress and IPA are very different software packages with BMDExpress being used for gene and pathway level analysis, deriving PoDs and pathway related analysis while IPA is used solely for pathway interpretation and determining Mechanism of Action (MoA). The combination of using BMDExpress and IPA has demonstrated more robust data can be generated for gene and pathways interpretation. More investigations are required using both software to depict correlations between them as well as considering individual results. This thesis has explained how developing a bioinformatics workflow using BMDExpress and IPA has expanded knowledge in the NGRA space using NAMs for safety decision making as a non-animal alternative approach.

ACKNOWLEDGMENTS

I'd like to express my love and gratitude to my husband Predrag for supporting me through my MRes degree some 20 years after university, he was there during my BSc degree and has helped me through emotional and exciting times. Big thanks to my children Sofija and Emilija for their patience and understanding while being locked away in my office drafting this thesis and during my first year. A big thanks to my employer (Unilever) who supported me during this apprenticeship, allowing me to focus on my studies and to my colleagues who have supported and encouraged me to upskill myself as a bioinformatician. Thanks to current and recent line managers Christelle Billon and Steve Gutsell for all their help and support throughout the MRes course. Many thanks to Nottingham University Professors Nathan Archer and Adam Blanchard for taking the time to supervise me during this project, they have given me advice and being a novice in this area I have appreciated all their input and patience during the course. Thanks to my course mates on the apprenticeship programme and for the online coffees and emotional and practical support. Finally, my huge thanks to Safety and Environmental Assurance Centre (SEAC) Bioinformatics team for introducing me to the wonderful world of Bioinformatics and especially to Mark Liddell and Jade Houghton for their help with R studio and BMDExpress. Predrag Kukic for his amazing mentoring and support; and reviewing of this thesis, Predrag has helped me throughout this MRes course with his enthusiasm and experience in the bioinformatics area in addition to keeping me focused when stress kicked in during the last few months! Many thanks to Carl Westmoreland for reviewing this thesis on a Unilever programme leader basis.

PUBLICATIONS

TULUM, L., DEAG, Z., BROWN, M., FURNISS, A., MEECH, L., LALLJIE, A. & COCHRANE, S. 2018. Airborne protein concentration: a key metric for type 1 allergy risk assessment-in home measurement challenges and considerations. *Clin Transl Allergy,* 8**,** 10.

ZHANG, Z., ZHANG, Y., LI, Y., JIANG, S., XU, F., LI, K., CHANG, L., GAO, H., KUKIC, P., CARMICHAEL, P. L., LIDDELL, M., LI, J., ZHANG, Q., LYU, Z., PENG, S., ZUO, T., TULUM, L. & XU, P. 2022. Quantitative phosphoproteomics reveal cellular responses from caffeine, coumarin and quercetin in treated HepG2 cells. *Toxicol Appl Pharmacol,* 449**,** 116110.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Abbreviation	Definition
AIC	Akaike Information Criterion
ATP	Adenosine Triphosphate
BER	Bioactivity Exposure Ratio
BMD	Benchmark Dose
BMDL	Benchmark Dose Lower
BMDS	Benchmark Dose Software
BMDU	Benchmark Dose Upper
BMR	Benchmark response
DEG	Differentially Expressed Gene
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DO's	Detector Oligo's
ECACC	Public Health England European Collection of Cell Cultures
FC	Fold Change
GO	Gene Ontology
ICCR	International Cooperation on Cosmetic Regulation
LDH	Lactate dehydrogenase
MoA	Mechanism of Action
NAM's	New Approach Methodologies
NGRA	Next Generation Risk Assessment
NOPEL	No Observed Protein Effect Level

NOTEL	No Observed Transcriptomics Effect Level
PBK	Physiologically based kinetic
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PoD	Point of Departure
QC	Quality Control
SEAC	Safety and Environmental Assurance Centre
TempO-Seq	Templated Oligo assay
URA	Upstream Regulator Analysis
USEPA	United States Environmental Protection Agency

CHAPTER 1

1. INTRODUCTION

1.1 Introduction to non-animal new approach methodologies (NAMs)

Human health risk assessments without generating new animal data have resulted in extreme efforts over the past few decades from industry, academia, and regulatory bodies to develop and apply new method approaches. Various non-animal approaches for chemical hazard characterisation have appeared known as NAMs which form the basis of integrated testing and assessment strategies designed to prevent harm to human health (Carmichael et al., 2009; Council, 2007; Desprez et al., 2018; Thomas et al., 2019; Westmoreland et al., 2010). Any technology, methodology, approach or combination of both that can be used to provide information on chemical hazard and risk that avoids the use of intact animals can be defined as a NAM (U.S. Environmental Protection Agency (USEPA 2018)). This can include many different types of *in vitro* bioactivity studies, *in silico* modelling of bioactivities and exposure predictions and cheminformatics. The USEPA has been evaluating high-throughput screening (HTS) through the ToxCast program and computational toxicology tools for over 10 years (Judson et al. 2010; Richard et al. 2016). Some of the outcomes that came out those efforts are NAM-based screening of endocrine disrupting chemicals for use in a regulatory setting (USEPA 2016) and high-throughput transcriptomics (HTTr) using targeted RNA-Seq as a broad coverage screening assay (Harrill et al. 2021).

New, non-animal approaches have been developed for conducting toxicological safety assessments being motivated by several factors including ethical considerations, regulatory action with certain types of ingredients being banned for animal testing, plus the need to assure the safety of chemicals using efficient, cost-effective, and robust methods (Dent et al., 2018, 2021; Thomas et al., 2019). NAMs also have the potential to improve safety assessments by using more human-relevant tools through coverage of key biological pathways or targets. However, as animal testing is illegal for the cosmetic industry, the safety assessment of new chemicals continues to rely on *in vivo* testing in animals, especially for higher tier hazard endpoints, such as systemic toxicity, but there have been advances in biotechnology and computational modelling. As a consequence, we see increasing use of different cell-based assays including high-content screening, omics, and reporter cell lines as well as a variety of computational models. In order to address these higher tied endpoints, a weight of evidence approach to combine NAMs to ensure the robustness and transparency of future risk assessments is required. Recently, the International Cooperation on Cosmetics Regulation (ICCR), a voluntary international group of cosmetic regulatory authorities, has defined the major principles for incorporating NAMs into an integrated strategy for Next-Generation Risk Assessment (NGRA) (Dent et al., 2018a). In line with the ICCR principles, NGRA provides a way to integrate NAM data from various sources into the decision-

making process, allowing for safety assessments to be conducted without the use of animal data (Dent et al., 2018, 2021; Thomas et al., 2019).

NGRA of cosmetic ingredients looks at the risk by comparing the exposure to an ingredient against the bioactivity of the ingredient itself. The bioactivity of the ingredient is probed using *in vitro* NAMs, such as a primary cell culture, co-culture, or micro tissue. In those assays, the biological model is exposed to a chemical, typically over a range of concentrations and timepoints, where differences in biomarker(s) of interest are measured. This results in a concentration–response dataset which can be analysed using statistical methods to obtain concentrations of chemicals which lead to a response of a prechosen amount. For example, concentrations of chemical resulting in differences no larger than background variation may serve as points of departure (PoDs) for comparison against an estimate of exposure within an NGRA (Baltazar et al, 2020).

One of the first examples of the use of NGRA in decision making has emerged recently where risk assessment of coumarin in a cosmetic product was assessed. The workflow for the coumarin case study, Figure 1, was based on the principles underpinning the use of NAMs in the safety assessment of cosmetic ingredients (Berggren et al., 2017; Dent et al., 2018a; Baltazar et al, 2020). The workflow uses a hypothesis-driven decision-making process to move the risk assessment from problem formulation to safety decision. The first tier starts with the estimation of exposure levels based on the use scenario and consumer habits ("Exposure Estimation" step), incorporated with problem formulation including molecular structure, *in silico* predictions and information from literature ("Collate existing information" step). Relevant internal exposures were estimated using a physiologically based kinetic (PBK) model for coumarin for exposure scenarios based on the habits and practices of the European demographic (Hall et al., 2007; SCCS, 2018). The second tier involves *in vitro* biological activity charactersation. The overall strategy for the second tier involves collecting and generating a wide range of bioactivity data to provide a comprehensive set of biomarkers which can be used to measure the bioactivity of the ingredient, and, corresponding PoDs, at consumer-relevant concentrations to identify or develop mechanistic hypotheses, or to obtain initial PoD. The third tier consisted of metabolism refinement which increased the certainty in the PoD using metabolite identification and 3D models. Once all these tiers were conducted, the margin of safety (MoS) was determined along with sufficient data and high certainty the risk assessment could be concluded with a low-risk conclusion based on the margin of safety calculations. All PoDs were compared with exposure estimates (plasma C_{max}) to calculate the margin of safety (also known as bioactivity exposure ratio (BER)) distribution which is used in the risk assessment decision (Middleton et al, 2020). It is envisaged that such an approach could be used to decide, depending on the BER, whether a given chemical-exposure scenario is low risk, or whether to use higher tier approaches to refine the risk assessment further.

The PoD estimation module consists of 3 of the *in vitro* bioactivity platforms used in (Baltazar et al., 2020) to obtain a BER estimate: high-throughput transcriptomics (Harrill et al., 2021), a cell stress panel (Hatherell et al., 2020) and *in vitro* pharmacological profiling (Bowes et al., 2012). The latter 2 platforms were selected to cover cellular stress and targeted biological effects, respectively, whereas the transcriptomics platform (generated using multiple cell models—HepG2, HepaRG, and MCF-7) was included to provide a non-targeted approach to capture biological effects potentially not detected using the other tools as detailed in Middleton et al, 2020.

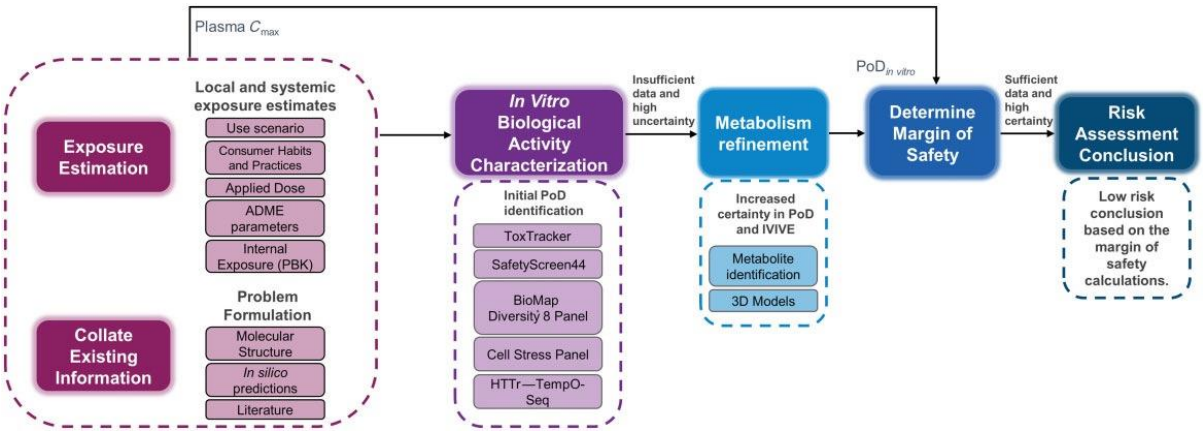


Figure 1. An example of workflow used for systemic toxicity end point in the Baltazar et al (2020) paper. Workflow based on a tired approach with the first tier involving local & systemic exposure estimates and problem formulations, second tier involving in vitro biological activity characterisation and the third tier involving metabolism refinement.

For a given workflow as detailed in Figure 1 above, the BER is defined as the ratio between the minimum PoD from bioactivity assays and the relevant plasma Cmax estimate (Figure 2). If the exposure level of a chemical in humans is far below the concentration needed for it to have any biological effect, then it is unlikely to trigger any toxicity. In contrast, if the exposure level is above the minimum concentration of the biological effect, the distributions of Cmax and bioactivities, and ultimately of the BER, can be further used to inform the decision and associated uncertainties.

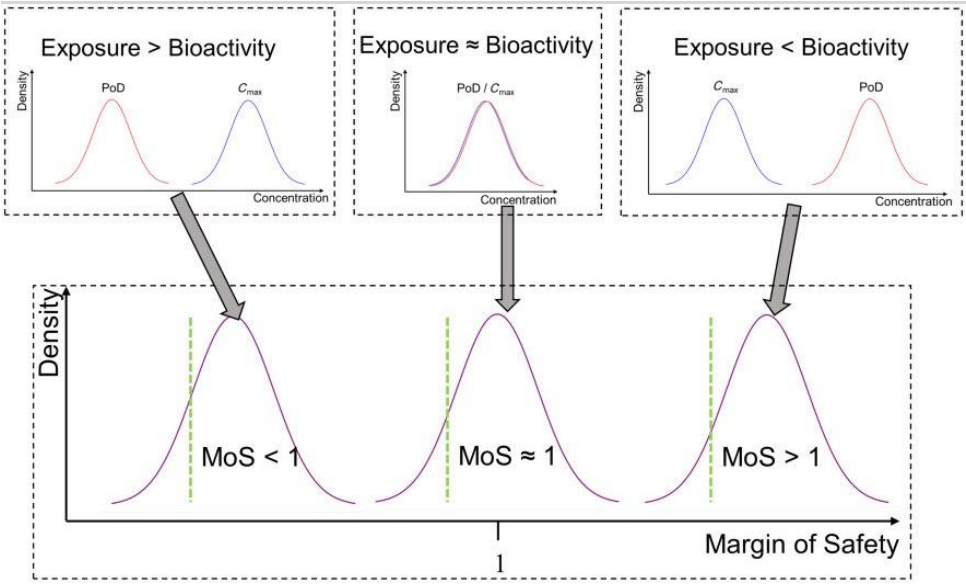


Figure 2. BER distribution for decision-making. The BER was defined as the PoD/C_{max} ratio. The uncertainty in the C_{max} and PoD estimates is represented as a distribution, and hence the BER estimate is also a distribution. When the distribution for the PoD is predominantly lower than the distribution for the C_{max} (exposure > bioactivity), this produces a distribution for the BER<1 and the safety decision for the exposure scenario is deemed uncertain (Middleton et al, 2020). Alternatively, if the distribution for the PoD is predominantly greater than the C_{max} distribution (exposure < bioactivity), this produces a distribution for the BER>1 and the safety decision for the exposure scenario is deemed safe (Middleton et al, 2020). When the distributions for the C_{max} and PoD strongly overlap (exposure≈ bioactivity), this results in an MoS distribution centred around 1 and uncertain decision.

In order to understand the meaning of BER in the context of chemicals with a known history of use, BER distributions are obtained for the benchmark chemical-exposure scenarios, using the distributions from the C_{max} error distribution model and the minimum POD. As a first step in benchmarking, the BER distributions were compared with the risk classifications assigned at stage 1 to each of the benchmark chemical-exposure scenario (Figure 3). Here, exposure scenarios are ranked by the median estimated BER, from smallest to largest along the y-axis, and color-coded according to their assigned risk-categories for doxorubicin and niacinamide (see Table 1). A BER<1 indicates the plasma C_{max} is above the minimum PoD measured across the bioactivity platforms. Based on this ranking, the first 6 exposure scenarios (Figure 3) were all high-risk benchmark chemical-exposure scenarios, and all have a median BER less than 1 (uncertain risk). The last 13 exposure scenarios were all safe (Middleton et al, 2020).

1.2 Case study chemicals

In this thesis, we will use two benchmark chemicals from (Middleton et al 2022): doxorubicin and niacinamide, these chemicals are used for Unilever business purposes as reference materials.

Doxorubicin is a chemotherapy drug and is a treatment for many different types of cancer. It slows or stops the growth of cancer cells by blocking enzymes called topoisomerase I and II. Cancer cells need these enzymes to divide and grow. Topoisomerase I inhibitors, such as irinotecan and topotecan, and topoisomerase II inhibitors, such as etoposide, teniposide, and anthracyclines, induce DNA strand breaks and hinder the action of topoisomerases that are involved in the DNA replication and process of transcription. Doxorubicin stabilizes the topoisomerase II complex after it has broken the DNA chain for replication, preventing the DNA double helix from being released and thereby stopping the process of replication. It may also increase quinone type free radical production, hence contributing to its cytotoxicity (Tacar et al. 2013). doxorubicin has been demonstrated to have significant therapeutic potential and is recognized as one of the most efficient chemotherapy medications that have been approved by the Food and Drug Administration (FDA) for the treatment of various cancers (Kciuk et al. 2023)

The planar aromatic chromophore portion of the molecule intercalates between two base pairs of the DNA, while the six-membered daunosamine sugar sits in the minor groove and interacts with flanking base pairs immediately adjacent to the intercalation site, as evidenced by several crystal structures (Frederick et al. 1990).Doxorubicin acts by inhibiting topoisomerase II (TopoII) resulting in DNA double-strand breaks. Cells then activate the DNA damage response (DDR) signalling cascade to guide recruitment of the repair machinery to these breaks. If this fails, the DNA repair programme initiates apoptosis. Rapidly replicating cells such as tumour cells are presumed to exhibit greater sensitivity to the resulting DNA damage than normal cells, thus constituting a chemotherapeutic window (Pang et al. 2013)

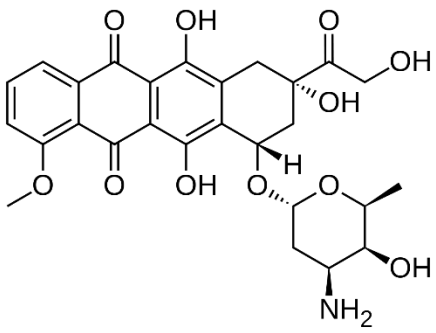


Figure 3. Chemical structure of doxorubicin

Niacinamide, an amide of vitamin B3 (niacin), is a hydrophilic endogenous substance. Given a sufficient bioavailability, niacinamide has antipruritic, antimicrobial, vasoactive, photo-protective and lightening effects depending on its concentration. There is tentative evidence for a potential role of niacinamide in treating acne, rosacea, autoimmune blistering disorders, ageing skin, and atopic dermatitis (National Cancer Institute. 2011). niacinamide also inhibits poly(ADP-ribose) polymerases (PARP-1), enzymes involved in the rejoining of DNA strand breaks induced by radiation or chemotherapy. (Chen et al. 2014).niacinamide is a well-tolerated and safe substance often used in cosmetics (Knip et al. 2000).

The structure of nicotinamide consists of a pyridine ring to which a primary amide group is attached in the *meta* position. It is an amide of nicotinic acid. As an aromatic compound, it undergoes electrophilic substitution reactions and transformations of its two functional groups(Knip et al. 2000).

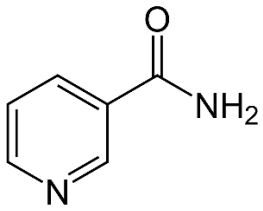


Figure 4. Chemical structure of niacinamide

Doxorubicin falls in the high-risk benchmark chemical exposure scenario with a BER of approximately 10^{-4} and niacinamide falls into the low-risk benchmark chemical exposure scenario with BER at approximately 1000 (Figure 3).

Risk classifications of “high” or “low” were assigned for the chosen chemicals investigated in this thesis, doxorubicin and niacinamide, to each benchmark scenario, for the purpose of safety decision-making in the context of a consumer product (e.g. personal care products). Therefore, if the documented safety profile of the benchmark chemical-exposure was used as a decision for inclusion in a consumer product, it would be considered high or low risk accordingly (Table 1).

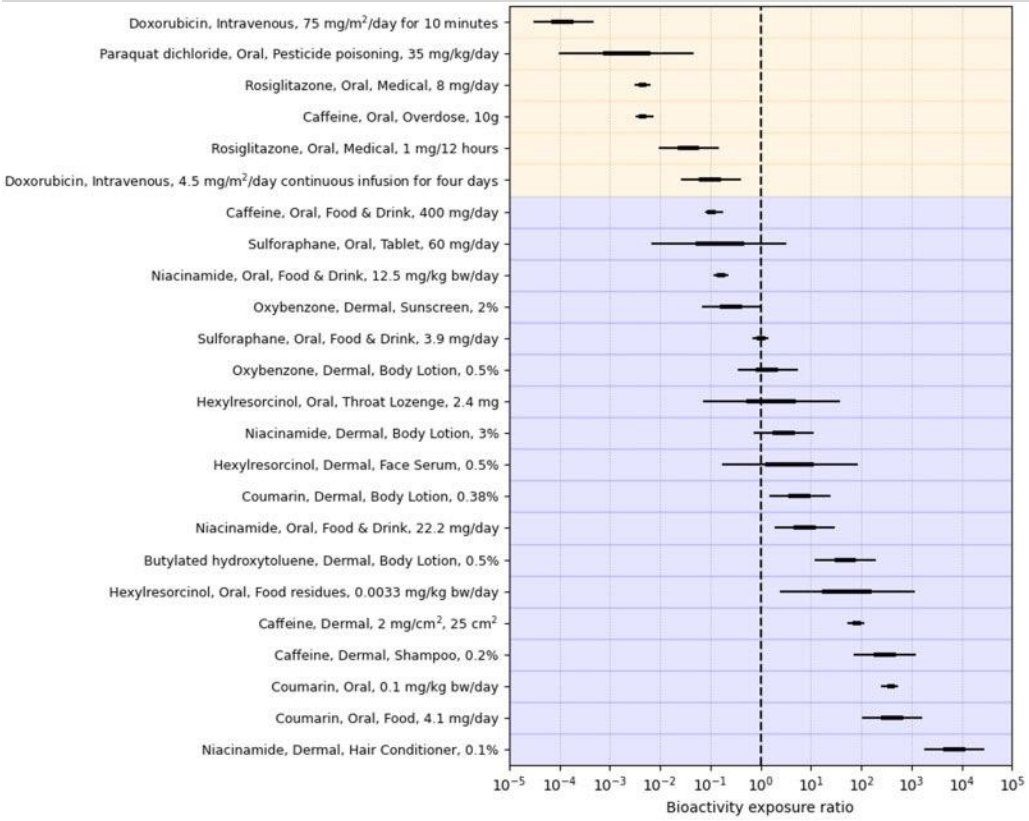


Figure 5. The distribution of the bioactivity exposure ratio (BER) when using all available predicted Cmax estimates. Background colours indicate the assigned risk category for each benchmark chemical-exposure scenario assigned at stage 1 (blue—low, yellow—high). The vertical dashed line indicates a BER equal to 1. For doxorubicin the risk category is in the yellow defined as high with BER<1 and for niacinamide the risk category is in the blue area defined as low were BER>1.

Compound	**Use Scenario**	**Risk Classification***	**Risk Classification Reasoning**	**Reference**
Doxorubicin Hydrochloride	IV bolus 75 mg/m2, 10 min	High risk	The incidence of symptomatic chronic heart failure is estimated to be 3–4% after a cumulative dose of 450 mg/m^2 if doxorubicin is administered as a bolus of 45–	Injac et al, 2008 Biganzoli et al, 2003 Rahman et al, 2007

			75 mg/m^2 every 3–4 weeks.	
Niacinamide	Tolerable daily intake (TDI) 12.5 mg/bw/day	Low risk	History of safe use. No evidence for concern with respect to systemic toxicity from the available toxicological data, as concluded by the Scientific Committee on Food and Scientific Panel on Dietetic Products, Nutrition and Allergies. niacinamide is a form of vitamin B3 with a recommended intake of 10-15 mg/day of niacin equivalent.	Cosmetic Ingredient Review Expert Panel, 2005; EFSA NDA Panel, 2014; EFSA Panel on Nutrition, Novel Foods and Food Allergens, 2022
	Norwegian dietary intake 22.2 mg/day	Low risk		
	0.1% in a hair conditioner	Low risk		
	3% in a body lotion	Low risk		

*from a consumer goods perspective with respect to systemic exposure

Table 1 – Summary of risk classification and rationale for each chemical exposure scenario with doxorubicin defined as high risk for various in use scenarios and niacinamide defined as low risk for various in use scenarios.

1.3 Introduction to Transcriptomics

HTTr is the study of mRNA molecules in a cell. mRNA is copied from pieces of DNA and contains information to make proteins and perform other important functions in the cell. Transcriptomics is used to learn more about how genes are turned on in different types of cells and assesses changes in gene expression (Black et al. 2022). HTTr is a type of NAM that uses gene expression profiling as an endpoint for rapidly evaluating the effects of large numbers of chemicals on *in vitro* cell culture systems. As compared to targeted high-throughput screening (HTS) approaches that measure the effect

of chemical *X* on target *Y*, HTTr is a non-targeted approach that allows researchers to more broadly characterize the integrated response of an intact biological system to chemicals that may affect a specific biological target or many biological targets under a defined set of treatment conditions (time, concentration, etc.). HTTr screening performed in concentration-response mode can provide potency estimates for the concentrations of chemicals that produce perturbations in cellular response pathways (Harrill et al. 2021).

Gene expression profiling has long been considered an informative method for evaluating the biological activity and/or toxicity of chemicals. Previous research focused on using gene expression data from *in vivo* animal studies to characterise the toxicity of environmental chemicals, using concentration-response modelling of gene expression measurements to identify molecular mechanisms-of-action, and to define transcriptional PoDs (Blomme et al. 2009; Cui and Paules 2010; Farmahin et al. 2017; Harrill et al. 2021; Thomas et al. 2013). Such studies were necessarily low-throughput given the use of laboratory animals and due to the changes in technology to high throughput transcriptomics. (Harrill et al. 2021)

Over the years, advances in transcriptomics research have included technological improvements in transcriptomics assay platforms, the establishment of large-scale, open-access transcriptome profiling datasets housing both *in vivo* and *in vitro* chemical bioactivity data (Igarashi et al. 2015; Lamb et al. 2006; Svoboda et al. 2019), and development of many computational strategies for analysing such data. However, the latter two topics have primarily focused on mechanism-of-action characterization and chemical clustering/read-across in past research (De Abrew et al. 2016). Fortunately, increasing efficiency and declining costs associated with generating whole transcriptome profiles have made *in vitro* HTTr screening in concentration-response mode a feasible option for NAMs-based hazard characterisation of environmental chemicals (Harrill et al. 2021).

In vitro biological activity characterisation can be conducted with a high throughput Transcriptomics method using TempO-Seq (Figure 4), a novel ligation-based targeted whole transcriptome expression profiling assay, to determine whether previously unreported compound-responsive genes could be identified and incorporated into a broad but specific compound signature (Yeakley et al, 2017)

TempO-Seq (Templated Oligo assay with Sequencing readout) uses a different approach to targeted sequencing (Figure 4) for mRNA only. The whole transcriptome TempO-Seq assay targets and measures a specific sequence within each gene, while measuring every gene and isoform in the transcriptome, doing so by directly targeting the RNA contained in crude cellular lysates in a homogenous progressive addition assay. TempO-Seq relies on the hybridization of two novel “detector” oligos (DOs) to adjacent target sequences so that when properly hybridized they can be ligated. Excess

unhybridized DOs are removed enzymatically in a process that is readily scaled and automated, then the ligated DO pairs are amplified to add a sample-specific sequence and the adaptors required for sequencing. As there is no poly-(A)+ selection, there is no positional bias in target location, so DOs are designed to maximize hybridisation specificity. The ligation step also provides specificity for single base differences, making even highly homologous genes distinguishable. The unique biochemistry of TempO-Seq also eliminates mis-ligation and assay background, making the whole transcriptome content possible, and allowing precise dose-response and single-cell level measurements while maximizing sequencing flow cell productivity. Importantly, the sequencing process delivers only already known ligated DO sequences, so there is no complex bioinformatic analysis; the output of the assay is a simple table with expression levels of each gene in each sample (Yeakley et al, 2017).

TempO-Seq exhibits 99.6% specificity, single cell sensitivity, and excellent correlation with fold differences measured by RNA-Seq (R^2 = 0.9) for 20,629 targets. Unlike many expression assays, TempO-Seq does not require RNA purification, cDNA synthesis, or capture of targeted RNA, and lacks a 3′ end bias (Yeakley et al, 2017).

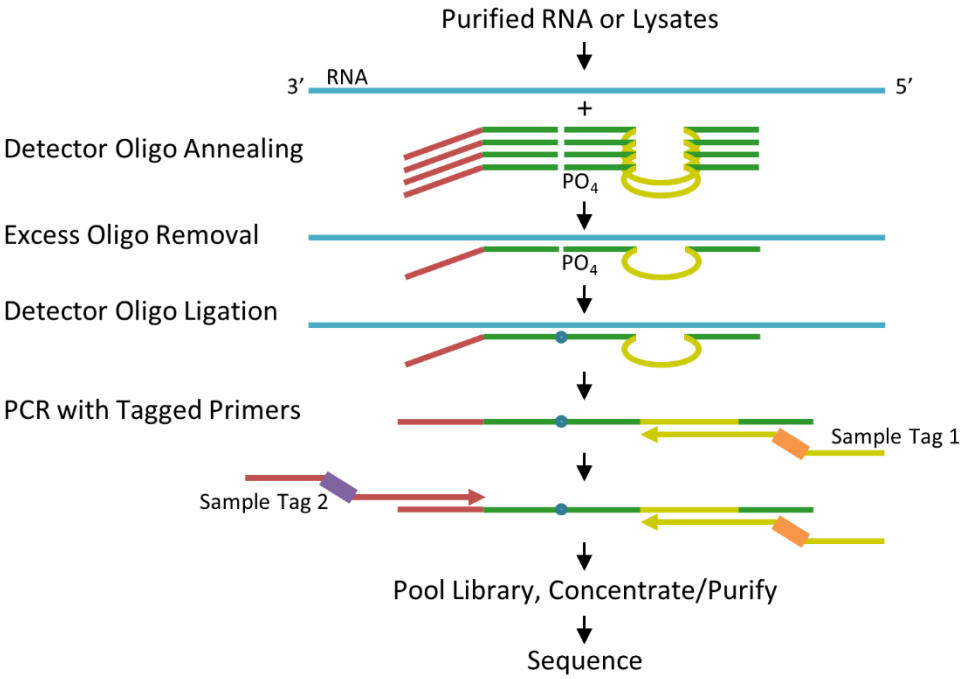


Figure 6. TempO-Seq biochemical scheme. RNAs are targeted by annealing to detector oligo (Dos) that contain target- specific sequences (green) as well as primer landing sites (red and yellow) that are shared across all DOs. Excess oligos are removed by a 3^0 exonuclease, then the hybridized oligos are ligated and amplified using primers that contain sample tag (index) sequences

(orange and purple bars), and adaptors required for sequencing. The amplified assay products are pooled for a library, purified/concentrated and sequenced (Yeakley et al, 2017).

The main focus for this thesis will be HTTr data generated at a third-party contract organisation and three cell lines of interest were HepG2, HepaRG and MCF-7 cells. HepG2 cells are used as an *in vitro* model of the human liver and to study toxic effects of drugs *in vitro*, HepaRG cells exhibit many characteristics of primary human hepatocytes including key metabolic enzymes, expression of nuclear receptors and drug transporters while MCF-7 cells are human breast cancer cell lines used for assessing the disruption in the endocrine system due to the expression of endocrine system related receptors (Middleton et al, 2020). .

1.4 Aims and Objectives

The main aim of this project is to address a gap of knowledge within Unilever by conducting biological interpretation of exposure to chemicals at different concentrations and the general use of *in vitro* methods for non-animal risk assessments.The project will interpret dose response HTTr data produced from TempO-Seq for HepG2, HepaRG and MCF7 cells dosed with doxorubicin and niacinamide at seven different doses and comparisons with the DMSO control. The first objective is to use dose response modelling software BMDExpress to potentially derive transcriptional PoDs and compare them to Cmax value determined for the given exposure scenarios.
Doxorubicin and niacinamide were selected as chemicals of interest for this project as they exemplify high and low risk benchmark chemical exposures, respectively, as determined using BER activities (Middleton et al 2022) (Figure 3). The null hypothesis states that there are no significant transcriptional differences in cell lines between control samples and samples treated with chemicals at different doses. The alternative hypothesis is that there are significant transcriptional differences in the various cell lines between control samples and samples treated with doxorubicin and niacinamide at seven different concentrations.

The second objective of this project is to go beyond the PoD determination from HTTr and further carry out biological interpretation of the detected differentially expressed genes from the dose response HTTr experiment. The outcome of the biological interpretation is used to check the hypothesis whether the already known mechanism of action (MoA) of the chemicals can be observed from these pathway analyses. For that reason, pathway analysis will be carried using Ingenuity Pathway Analysis (IPA) software and the outcomes will be compared with the known MoA of the chemicals.

Technical questions were explored as follows:

1) Are there relevant transcriptional differences between control and treated samples in Hep G2, HepaRG and MCF7 cells at different doses?
2) Is it possible to derive a PoD from dose response software BMDExpress?
3) Identify pathways associated with the differential expressed gene profiles and determine whether the MoA of the chemicals in each cell line can be observed from these pathway analyses using BMDExpress and IPA

The following objectives were investigated for this project:

Objective 1:
a) Align and quantify transcriptomics data (from fastq files) from HepG2, HepaRG and MCF7 cell lines treated with doxorubicin and niacinamide at seven different doses.

b) Conduct differential expression analysis using DESeq2 between each treatment chemical and DMSO controls.

Objective 2:
a) Using dose response software BMDExpress to potentially derive PoD.

Objective 3:
a) Investigate the enriched pathways from the significant differentially expressed genes and determine whether the MoA of the chemical can be inferred using BMDExpress and IPA.

CHAPTER 2

2. MATERIALS AND METHODS

2.1 Introduction

The work described in this report contains the analysis output of high-throughput transcriptomics data (HTTr) using the TempO-Seq assay for targeted sequencing-based RNA expression analysis of HepaRG, HepG2 and MCF7 cell lysates, generated at a third party contract research organisation (CRO) Cyprotex, where cells were treated with doxorubicin and niacinamide for 24 hours prior to cell lysate generation and subsequent shipping to another third party contract research organisation, Bioclavis for QC analysis and results generated from the TempO-Seq assays were performed. The data generated was analysed by SEAC, Unilever using BMDExpress to generate Points of Departure (PoD) to be used in NGRA as well as pathway analysis using BMDExpress and Ingenuity Pathway Analysis (IPA).
This report relates solely to the cell lysate samples that were exposed to doxorubicin and niacinamide, although these samples were part of a much larger study.

2.2 Cell Culture

HepG2 cells (human hepatoblastoma) were obtained from the Public Health England European Collection of Cell Cultures (ECACC, Salisbury, UK). Cells were cultured in complete MEM (Gibco) supplemented with 10% fetal bovine serum (FBS) (Sigma), 2 mM GlutaMAX (Gibco), 1% nonessential amino acids (Gibco), 53 U/mL penicillin (Sigma), and 53 µg/mL streptomycin (Sigma). HepG2 cells were maintained in 75 cm^2 cell culture flasks in a humidified atmosphere incubator with 5% CO_2 at 37 °C; the cells were kept at a confluence below 85% and were not maintained in culture for more than 4 weeks (8 passages). Cells were seeded into 384-well, clear-bottom black-walled tissue culture plates at a density of 6000 cells/well and were left overnight to attach.

MCF-7 cells (human Caucasian breast adenocarcinoma) were obtained from ECACC (Salisbury, UK). Cells were cultured in complete RPMI 1640 medium (Gibco) supplemented with 10% FBS (Sigma), 2 mM GlutaMAX (Gibco), 53 U/mL penicillin (Gibco), and 53 µg/mL streptomycin (Gibco). MCF-7 cells were seeded into 384-well, clear-bottom black-walled tissue culture plates at a density of 6000 cells/well and were left overnight for attachment.

HepaRG cells (proliferative human hepatoma-derived cell line) were obtained from Life Technologies and cultured in Williams' E medium supplemented (Gibco) with 2 mM L-glutamine (Gibco) and HPRG670 supplement (Lonza, UK), in collagen-coated, 384-well, clear-bottom, black-walled, tissue culture plates,

at a density of 20,000 cells/well. HepaRG cells were then transferred to serum-free medium following the initial 24 h seeding procedure (Williams E medium supplemented with 2 mM GlutaMAX, 100 units/mL penicillin, 100 µg/mL streptomycin, and HPRG640 supplement), for 6 days prior to dosing, with media replenishment every second day.

2.3 Cytotoxicity testing

Cytotoxicity measurement was used to inform the dose range of doxorubicin and niacinamide. The experiments were done and reported in Middleton et al, (2022) by measuring the level of induced Lactate dehydrogenase (LDH) in all three cell lines. From the LDH measurements, doxorubicin became cytotoxic at approximately 0.5µM for all cell lines while niacinamide became cytotoxic at approximately 3000µM for Hep G2, >8000µM for HepaRG and 10,000µM for MCF-7 cells.

2.4 Preparation of dose solutions

Test compounds, doxorubicin (LGC Standards) and niacinamide (Sigma-Aldrich) were prepared as stock solutions in 200× higher concentrations than the highest concentration to be tested. Dimethyl sulfoxide (DMSO) was used as the solvent, and its final concentration in treatment media was maintained at 0.5% v/v. Serial dilutions were performed using the custom dilution series for each compound. Doxorubicin was prepared at doses of 0.000064 - 1 µM for all cell line and niacinamide were prepared at doses of 0.512 - 8000 µM (HepaRG) and 3.84 - 60000 µM (HepG2 and MCF-7). Cells were treated at seven concentrations of each test compound, and five biological replicates were generated. Compound treatment was performed for 24 h in a humidified atmosphere with 5% CO_2 at 37 °C. Cells were washed in calcium- and magnesium-free phosphate buffered saline (PBS) (Sigma). With all residual PBS removed, the 2X TempO-Seq lysis buffer (BioSpyder Technologies, proprietary kit) was diluted to 1× with PBS and added at a volume of 1 µL per 1000 cells with a minimum of 10 µL per well and incubated for 10 min at room temperature. Following lysis, the samples were frozen at −80 °C prior to sequencing.

Doxorubicin and niacinamide were produced on separate exposure plates. 21 DMSO (Sigma) solvent controls are produced per exposure plate, distributed across various columns and rows of the 384 well plate, resulting in 21 positive control samples. HTTr sequencing was performed using TempO-Seq (BioClavis) version two of the Human whole transcriptome panel.

Probes were filtered to include only those which had a median count, across all samples, of 5 or above. Samples were filtered to only include those with more than a sum of 2,500,000 counts within the remaining probes and those with a mapped read percentage over 55%.

2.5 RNA Sequencing and Data Analysis

TempO-Seq analysis was performed as described in the introduction (Section 1.2) (Yeakley et al. 2017), with a targeted sequence depth of 200 mapped read counts per transcript including the use of the general attenuation panel. Raw count data was produced using a STAR algorithm (Dobin et al., 2013) and TempO-Seq R software package. Raw counts were processed using the R package DESeq2 (Love, Huber, and Anders 2014) as detailed in Appendix 1 Supplementary Script 1 and GitHub link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/DESeq2%20MRes%20script.R).

2.6 TempO-Seq assay protocol summary

In TempO-Seq, each Detector Oligo (DO) consists of a sequence complementary to an mRNA target plus a universal primer binding site (i.e. same for every targeted gene). They anneal in immediate juxtaposition to each other on the targeted RNA template such that they can be ligated together. Ligated detector oligos are PCR-amplified using a primer set (single-plex 25-cycle PCR reaction, with a single primer pair for each sample) that introduces both the adaptors required for sequencing and a sample-specific barcode. The barcode sequences flank the target sequence and are inserted appropriately into the standard Illumina adaptors to permit standard dual-index sequencing of the barcodes and deconvolution of sample-specific reads from the sequencing data using the standard Illumina software. Up to 384 PCR-amplified and barcoded samples are pooled into a single library for sequencing. Sequencing reads are demultiplexed using the standard sequencing instrument software for each sample using the barcodes to give a FASTQ file for each.

2.7 Data analysis protocol

TempO-Seq sequence files were analysed using the Tempo-SeqR software package. The input for TempO-Seq data analysis is a folder of zipped FASTQ files. Each FASTQ file contains the reads and quality scores for one sample. Each FASTQ file is aligned using the STAR algorithm to a pseudo-transcriptome corresponding to the gene panel used in the assay. The primary output of the Tempo-SeqR software was a table of counts with each column representing a sample and each row representing a gene.

2.8 Data pre-filtering, differential expression analysis and visualisation

Transcriptomics data was analysed using R Studio (2022.07.2 Build 576) (R script detailed in Appendix 1 Supplementary Script 1 and GitHub link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/DESeq2%20MRes%20script.R) to produce visualisation

plots including principal component analysis (PCA), MA plots, volcano plots and ggplots to identify any outliers in the data.

For transcriptomics, a typical pre-filtering criterion of median counts over all samples < 5 was carried out. The normalization and analysis of differentially expressed genes (DEGs) were conducted by using DESeq2 (version 1.38.3 (Love et al. 2014), which is regarded as one of the leading tools for pairwise differential expression analysis when fewer than 12 replicates are used (Schurch et al., 2016).

Raw counts were processed using the R package DESeq2 (version 1.38.3) (Love et al., 2014) separately per chemical/cell-line dataset. Probes were filtered to include only those which had a median count, across all samples, of 5 or above and samples were filtered to only include those with more than a sum of 2.5 million counts within the remaining probes and with a mapped read percentage over 55%. Outliers were removed where biological replicates had a correlation of <85% and could identified using principal component analysis.

Data were normalized using the negative binomial distribution in DESeq2 with model "_ VESSEL_ID + CONCENTRATION" where "VESSEL_ID" is given per treatment 384 well plate and is identified as a strong source of variation between biological replicates, and therefore set as a confounding factor. Rlog-transformed normalized counts were used as input into benchmark response (BMR) modelling software BMDExpress 2.3 (Version BUILD released on July 15, 2020) (Phillips et al., 2019) where data were modelled to calculate PoDs per chemical/cell-line dataset. Figure 5 shows the analysis flowchart from raw counts to gene/pathway level PoD's.

Figure 7: Overview of analysis flowchart from raw counts to gene/pathway level PoD's.

All pre-processing steps were performed in R Studio (2022.07.2 Build 576) and visualizations were generated in the ggplot2 R package (2022.07.2 Build 576). RNA sequencing data was pre-processed in R using the raw RNA sequencing read counts. PCA was performed for the transcriptome data based on raw counts after variance stabilizing transformation using the 'vst' and 'plotPCA' functions of the R package DESeq2. MA and volcano plots were performed using plotMA and plot_EnhancedVolcano functions for each cell line and dosed chemical.

For both doxorubicin and niacinamide, differentially expressed genes (DEGs) were identified by using a pre-filtering procedure with Benjamin-Hochberg

adjusted p-value < 0.05 with a fold change (FC) threshold of 1.5.

DESeq2 normalised and rlog transformed data was used to investigate biological replicate quality by plotting PCA biplots coloured by replicate group (Concentration) and calculating correlation values for the top 20 differentially expressed genes and coloured by concentration groups (ggplots).
MA plots are commonly used to represent log fold-change versus mean expression between 2 treatments. This is visually displayed as a scatter plot with base-2 log fold change along the y axis and normalised mean expression along the x axis.
A volcano plot is a type of scatterplot that shows statistical significance (P value) versus magnitude of change (log 2 fold Change). It enables a visual identification of genes with large fold changes that are also statistically significant which may be the most biologically significant genes (add Volcano plots).
Upset plots are a data visualisation method for showing set data with more than three intersecting sets, showing intersections in a matrix with the rows of the matrix corresponding to the sets and the columns to the intersections between these sets. The size of the sets and intersections are shown as bar charts. (R script details in Appendix 1 Supplementary Script 2 and GitHub link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/IPA_gene_intersect%20upset.R).

2.9 Dose response based on Benchmark dose (BMD) method and pathway enrichment analysis

Chemical risk assessment aims to establish acceptable levels of exposures based on toxicological dose–response studies. Traditional methods that apply the lowest-observed-adverse-effect- level or no-observed-adverse-effects-level, may be limited by the selection of doses, sample sizes required to detect subtle effects and by technical and biological variability that limits ability to detect significant changes (Crump, 1984). In contrast, benchmark dose (BMD) modelling fits experimental dose–response data with a statistical model to identify a defined level of response relative to a control group. BMD was developed to overcome the limitations of the lowest-observed-adverse-effect-level/no-observed-ad- verse-effects-level approach (Crump, 1984). Regulatory agencies have increasingly adopted BMD modelling for human health risk assessment (Budtz-Jorgensen et al., 2013; Health Canada, 2013).

For a comprehensive statistical analysis of the dose response data, BMDExpress 2 software (version 2.3) was used for multivariate dose-response analyses, which provides functionality for benchmark dose (BMD) computations using the same curve-fitting methods as implemented in the U.S. Environmental Protection Agency's Benchmark Dose Software (BMDS) (U.S. Environmental Protection Agency)). The input for BMDExpress 2 was prepared from the pre-processed intensity and read count data after quality filtering.

BMDExpress computations are presented in tabular format viewable in the BMDExpress software. However, because of the limited capability to perform additional analyses and data visualization in the BMDExpress application, the results are typically exported to separate software (e.g., a spreadsheet) for further exploration. There are two major types of outputs that can be exported, i.e., (1) "BMD Analysis," and (2) "Functional Classifications." A "BMD Analysis" output file contains gene (or microarray probe), BMD and BMD lower confidence (BMDL) values for each statistical model, as well as the information required for model selection. A "Functional Classifications" output file was exported as "Gene Ontology Analyses" (Figure 6).

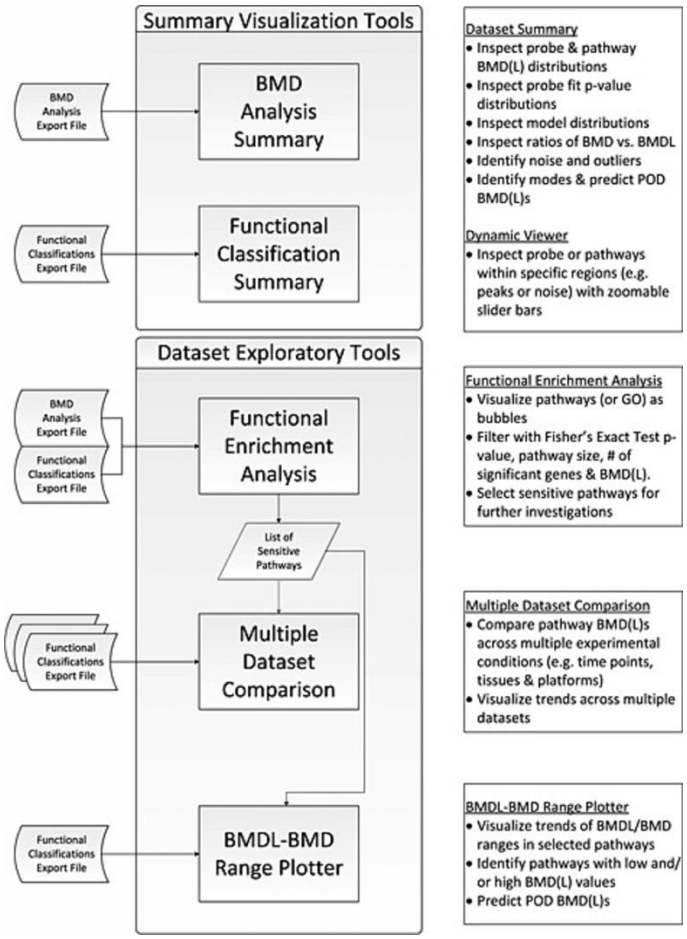


Figure 8 - Workflow demonstrating the features and functionality of BMDExpress Data Viewer. BMD, benchmark dose; BMDL, benchmark dose lower confidence (values); GO, gene ontology; POD, point of departure.

2.9.1 Preparation of input data for BMDExpress 2

For all datasets, the pre-processed and filtered intensity/expression data was stored in tab-separated plain text files. These files were formatted according to the requirements of the BMDExpress 2 software (Phillips et al. 2019) i.e. with one column per sample and one row per feature, where the first column

contains the feature identifier and the first two rows contain the sample identifier (row 1) and the applied compound concentration (row 2).

2.9.2 Benchmark dose analysis with BMDExpress 2

The actual dose response analysis was performed using the BMDExpress 2 desktop application (version 2.30, build 0439) (Phillips et al. 2019). To limit the number of models that need to be fitted in the actual benchmark dose (BMD) analysis, the data was pre-filtered within BMDExpress 2. Pre-filtering was performed using the "Williams Trend Test" with "P-Value Cutoff" set to 0.05, "number of Permutations" set to 100 and without checking the "Multiple Testing Correction" and checking the "Filter Out Control Genes" options. Together with the Williams Trend Test, a Fold Change Filter was applied with the "Fold Change Value" set to 1.5 for all datasets. Execution parameters were set at "Number of Threads" to 20.

Subsequently, the features passing pre-filtering in each dataset were used as input for the actual BMD analysis, i.e. the fitting of the dose response models. For each feature the following six types of model equations were fitted to the measurement data: four exponential models called Exponential 2 (Exp 2), Exponential 3 (Exp 3), Exponential 4 (Exp 4) and Exponential 5 (Exp 5), a second-degree Polynomial model (Poly 2), a Hill model (Hill), and a Power model (Power). After fitting all models, for each feature the best model was selected using the Akaike information criterion (AIC) (Cavanaughet al. 2019) as quality measure. Further parameters for curve fitting and model selection were set as follows:

- P-Value Cutoff: 0.05
- Best Poly Model Test: Lowest AIC
- Flag Hill Model with 'k' Parameter < 1/3 of Lowest Positive Dose
- Maximum Iterations: 250
- Confidence Level: 0.95
- Constant Variance: true
- Restrict Power: No Restriction
- BMR Factor: 1.349 (10%)
- Multiple Threads: 20

Finally, the BMD analysis output was filtered to keep only features for which the BMD of the best model was at most as high as the maximum tested concentration (Best BMD < Highest concentration of each sample) and the best model fitted the data sufficiently well (Best fitPValue > 0.1). Data was filtered further to keep only features with sufficiently small confidence intervals for the best BMD (BMDU/BMDL < 40). These parameters were conducted for all samples in this analysis.

Downstream examination of the BMD results was performed directly within BMDExpress 2. Overlap plots were generated using the R package UpSetR (R script detailed in Appendix 1 Supplementary Script 2 and GitHub link:

https://github.com/liztulum/MRes-thesis-scripts/blob/main/IPA_gene_intersect%20upset.R) and further visualizations were generated using ggplot2 (R script detailed in Appendix 1 Supplementary Script 1 and GitHub link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/DESeq2%20MRes%20script.R).

2.9.3 Functional classification analysis

To identify biological functions or pathways affected by the applied compound, we performed a functional classification analysis (Category Analysis) within BMDExpress 2. In this analysis, the features in the BMD analysis output are matched to different functional classifications based on their associated Entrez ID and summary values for the BMD values (from the best model) and the corresponding benchmark dose upper and lower confidence limits (BMDU, BMDL) are computed for each functional category. Functional classification analysis was performed for Gene Ontology (GO) terms (THE GENE ONTOLOGY, C. 2017*;* Mi et al. 2019*)* and REACTOME pathways (B. Jassal *et al;* Gillespie et al. 2022*)*. The parameters for the functional classification analysis were set as follows:

- Remove Promiscuous Probes: true
- Remove BMD > Highest Dose from Category Descriptive Statistics: true
- Remove BMD with p-Value < Cutoff: 0.1
- Remove Genes with BMD/BMDL >: null
- Remove Genes with BMD/BMD >: null
- Remove Genes with BMDU/BMDL >: 40
- Remove Genes with BMD Values > N Fold Below the Lowest Positive Dose: null
- Identify Conflicting Probe Sets: true
- Correlation cutoff for conflicting probes sets: 0.5

All further parameters were left at the default values.

For downstream examinations, the output of the functional classification analysis was filtered to keep only categories with a total size of more than two genes (All Genes (Platform) > 2), a sufficiently small p-value in the enrichment test (Fisher's Exact Two Tail < 0.1), and for which more than one feature in the dataset was passing all input filters (Genes that Passed All Filters > 1).

In addition to these five approaches suggested by Farmahin et al., we applied a sixth approach and computed a gene level global BMD /NOPEL/NOTEL as mean BMD of all genes associated to adverse reactions, cellular stress responses or AOPs.

The BMDL mean value of each pathway for transcriptomics was calculated based on the average of BMDL values of the genes enriched in the particular pathway. Global POD values were determined at pathway and gene level

based on the previously published approaches (Farmahin et al. 2017):

The mean BMDL was calculated by taking the mean of all significant probe level BMDLs in the given Reactome pathway. PoDs presented in this report are as follows with (1) and (2) being at the gene level and (3), (4) & (5) being at the pathway level

1) the average of BMDLs of the 20 probes with highest fold change, gene level
2) the average of probe BMDLs within the 25th to 75thpercentile, gene level
3) the average of the lowest 20 mean pathway BMDLs,
4) the average of the 20 mean pathway BMDLs with highest significance (lowest p value)
5) the lowest mean pathway BMDL. PoDs presented are based on the nominal concentrations tested.

2.10 Pathway and upstream regulator enrichment analysis using Ingenuity Pathway Analysis

Pathway and upstream regulator enrichment analysis was carried out using QIAGEN IPA software that is built on the manually curated content of the QIAGEN Knowledge Base (Figure 7).

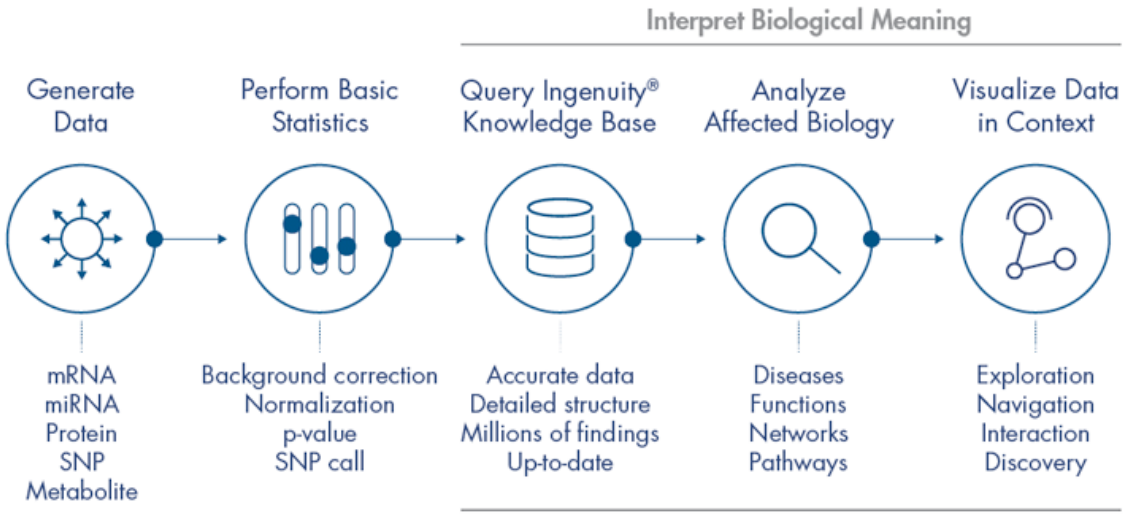


Figure 9 – workflow of IPA software (image copied from https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/insightful/)

IPA; QIAGEN, Redwood City, California (version 94302991, Build ing-lapis) was used to identify perturbed upstream regulators and canonical pathways. For each chemical tested, Excel files were imported into IPA containing gene IDs (Ensembl and Gene Symbol), p value, adjusted p value and the log2 fold-changes of the gene relative to solvent controls (exported data from R

Studio). Using Visual Studio (VR) Code (version 1.77.0), (Python script detailed in Appendix 1 Supplementary Script 3 and GitHub link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/gene%20name%20change.py), the unique protein/site labels in the one JSON object were extracted and inserted into the corresponding slots of the other JSON object (with annotation information) to replace the 'gene' column in the excel file as "Gene IDs". The resulting modified JSON objects with unique protein/site labels ("Gene IDs") were imported as "txt files" into IPA for each cell line and dosed chemical. IPA is only able to process pathways with a minimum of 100 DEG's therefore not all samples concentration were imported into IPA. The samples imported into IPA for pathway analysis were doxorubicin MCF-7 1, 0.2 and 0.04µM, HepG2 1, 0.2, 0.04 and 0.00032µM, HepaRG 1, 0.2µM; niacinamide MCF-7 60000 and 12000µM, HepG2 60000 and 12000µM and HepaRG 8000µM.

This approach allowed us to identify enrichment of genes showing robust concentration-responses to the exposures. IPA Core Analysis with a gene expression threshold of log2 fold change 1.5 and FDR-adjusted p value 0.05 was used with the direct and indirect relationship settings based on experimental and highly predicted data (focusing on human sources from breast cancer cell lines). Statistical significance of the overlap (FDR-adjusted p value .05) between the data set and known targets of upstream regulators in IPA were calculated using Fisher's exact tests. The z-score was calculated using Fisher's exact test based on the expected relationship for directions between upstream regulators and target genes and those observed in the data set. A z-score of >2 (activated) or <2 (inhibited) was considered statistically significant.

The following parameters in IPA were set as default as follows:

Core Analysis – Expression analysis

- Measurement type – Expr Log Ratio
- Population of genes to consider for p value calculations – Ingenuity Knowledge Base (Genes only)
- Direct and indirect relationships
- Interaction networks – Endogenous chemicals included with 35 molecules per network and 24 networks per analysis
- Causal networks – score using causal networks only
- Data sources – Third party information including MicroRNA-mRNA interactions, Protein-protein interactions, additional sources i.e. Gene Ontology (GO); DrugBank; BioGrid etc
- Species – Mammal (Human, Mouse and Rat)
- Various Tissues, Primary cells and cell lines

Chapter 3

3. RESULTS USING BMDEXPRESS

To identify changes in the number of transcripts induced by doxorubicin and niacinamide exposure, we compared the treatment time profiles with control profiles derived from time-matched DMSO-treated cell lines. The choice of doses was informed by changes in cytotoxicity after 24 hours, as measured in the previous work (Middleton et al. 2022). Initial cytotoxicity pre-screens – cellular adenosine triphosphate (ATP) and lactate dehydrogenase (LDH) release measurements were used to set the maximum concentration to be tested for the transcriptomics platforms for each compound and cell line. For the transcriptomics platforms, the concentrations were 36 identified by using a standardised setting procedure which involved specifying the maximum concentration based on the minimum of either a chemical's solubility limit or the concentration at which cytotoxicity is observed. The dilution series used for doxorubicin and niacinamide was the minimum concentration at approximately 4 orders of magnitude smaller than the maximum concentration. Cell toxicity for niacinamide was estimated to be ~3000µM for HepG2 cells; <8000µM for HepaRG and 11000µM for MCF-7 cells and for doxorubicin was estimated to be ~0.5µM for all cell lines.

For each time point, each experiment was carried out using 6 doses and the vehicle control in 3 biological replicates. In order to maintain consistency with the previous risk assessment studies of coumarin and caffeine (Baltazar et al. 2020; Hatherell et al. 2020; Rajagopal et al. 2022), doses for doxorubicin treatment were preselected, ranging between 0.000064 – 1 µM for all cell lines. Similarly, doses for the niacinamide treatment were preselected, ranging between 0.512 - 8000 µM (HepaRG) and 3.84 - 60000 µM (HepG2 and MCF-7).

Raw counts data was inputted into R studio and analysed by using DESeq2 to produce differentially expressed genes (DEGs) for each chemical and cell line. The number of differentially expressed genes was calculated for each chemical concentration and cell line giving an indication as to how many genes had been differentially expressed using this methodology. These DEGs were then used for BMDExpress 2.3 analysis to identify genes and pathways relevant to doxorubicin and niacinamide for each cell line. IPA was also conducted using the DESeq2 data with only the number of DEGs above 100 meeting the criteria, meaning not all chemical concentrations and cell lines were analysed using IPA.

As expected, for all cell lines, the number of DEGs at higher doses of doxorubicin and niacinamide were higher compared to the lower doses as detailed in Table 2. The top 20 DEG's for each cell line, doses and chemicals are illustrated in Appendix 2, Tables 1-28, this data was used in DESeq2 and

inputted into BMDExpress for data analysis. The data detailed for each cell line dosed with doxorubicin and niacinamide at various concentrations the top 20 genes identifed with up and down regulated DEG's as calculated by the log2fold change results. Some of the lower concentrations contained less than 20 DEG's and some didn't contain any DEG's which is further illustrated in Table 2 below.

Chemical/cell line names	Concentrations (μM)	No of DEG's
HepaRG Doxorubicin	1	1157
	0.2	101
	0.04	6
	0.08	18
	0.0016	0
	0.00032	0
	6.4e.05	0
HepG2 Doxorubicin	1	5021
	0.2	1994
	0.04	183
	0.08	0
	0.0016	0
	0.00032	61
	6.4e.05	0
MCF-7 Doxorubicin	1	5388
	0.2	727
	0.04	9
	0.08	0
	0.0016	0
	0.00032	0
	6.4e.05	0
HepaRG Niacinamide	8000	167
	1600	2
	320	0
	64	0
	12.8	0
	2.56	0
	0.512	7
HepG2 Niacinamide	60000	3813
	12000	151
	2400	236
	480	2
	96	2
	19.2	57
	3.84	1
MCF-7 Niacinamide	60000	7689
	12000	1622

	2400	76
	480	0
	96	0
	19.2	0
	3.84	0

Table 2. The number of DEGs in all cell lines dosed with doxorubicin and niacinamide.

In general, across the cell lines, treatment with doxorubicin resulted in limited gene expression changes at concentrations below 0.04µM suggesting limited cellular effects at lower concentrations. Specifically, in HepaRG cells there were no significant gene changes (p-adj < 0.05) at concentrations under 0.2µM. By 0.2µM, only 101 genes were found to be differentially expressed, which increased to 1157 genes at 1µM. In the HepG2 cells the overall gene expression response to doxorubicin was stronger with 5021 DEGs identified at the highest concentration of 1µM, decreasing to 1994 DEGs at 0.2µM and 183 at 0.04µM with no significant gene changes (p-adj < 0.05) at concentrations under 0.08µM. DEGs were detected in MCF7 cells at 1µM and 0.2µM only with 5388 and 727 genes, respectively. No significant differential gene expression was observed below 0.2µM for MCF7 cells.

Treatment with niacinamide resulted in limited gene expression changes at concentrations below 2400µM for HepG2 and MCF7 cells suggesting limited cellular effects at lower concentrations, whereas HepaRG cells shows limited gene expression changes at concentrations below 8000µM. Specifically, in HepG2 cells there were no significant gene changes (p-adj < 0.05) at concentrations under 2400µM. By 12000µM, only 151 genes were found to be differentially expressed, which increased to 3813 genes at 60000µM. In the HepaRG cells, the overall gene expression response to niacinamide was significantly lower with 167 differentially expressed genes identified at the highest concentration of 8000µM. The lower number of genes in HepaRG cells could be due to the smaller top dose of 8000µM compared to 60000µM in HepG2 and MCF-7 cells. DEGs were detected in MCF7 cells at 60000µM, 12000µM and 2400µM only with 7689, 1622 and 76 genes, respectively. No DEGs were observed below 2400µM for MCF7 cells.

Estimation of PoDs from high-throughput transcriptomics data is an active area of research and there is considerable debate about the selection of which method or PoD definition is most appropriate for NGRA (Baltazar et al., 2020; Farmahin et al., 2017; Harrill et al., 2019, 2021; Reynolds et al., 2020). To begin to explore the potential impact of selecting one approach over another, the transcriptomics data were analysed using 2 different methods, BMDExpress2 and IPA, however, IPA is used for pathway identification and functional analysis for individual doses only, whereas the BMDExpress2 BMDL PoDs represent the lowest concentration at which mechanistic changes occur, inferred by Reactome pathways, and an estimate of apical endpoints

(Farmahin et al., 2017). A comparison of the two approaches will not be conducted in this thesis, rather an interpretation of the results for each individual approach and if any pathway similarities can be identifed these will be discussed.

3.1 Doxorubicin

3.1.1 Quality Control

Firstly, PCA analysis was conducted to make sure there is a separation between the samples and there were no outliers. For doxorubicin we observed clear separation for MCF7 and HepG2 cells while for HepaRGs we observed no clear separation between treated and untreated samples, but there was a clear separation between samples treated with different concentrations (see Figure 8). The ggplots (Figure 9) showed a significant separation between the different concentrations for the top 20 genes in all cell lines dosed with doxorubicin, good correlation was observed showing the normalised counts increasing with the increased concentrations. The higher the concentration, the larger the number of normalised counts for the top 20 significantly differentiated genes.

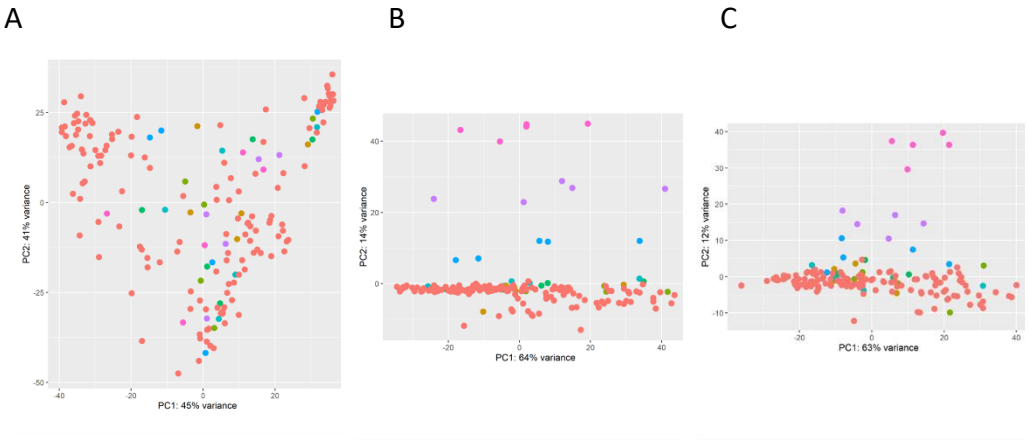


Figure 10: PCA biplots that show the distribution of the samples using top 2 principal components PC1 and PC2 – (A) HepaRG doxorubicin; (B) HepG2 doxorubicin; (C) MCF-7 doxorubicin

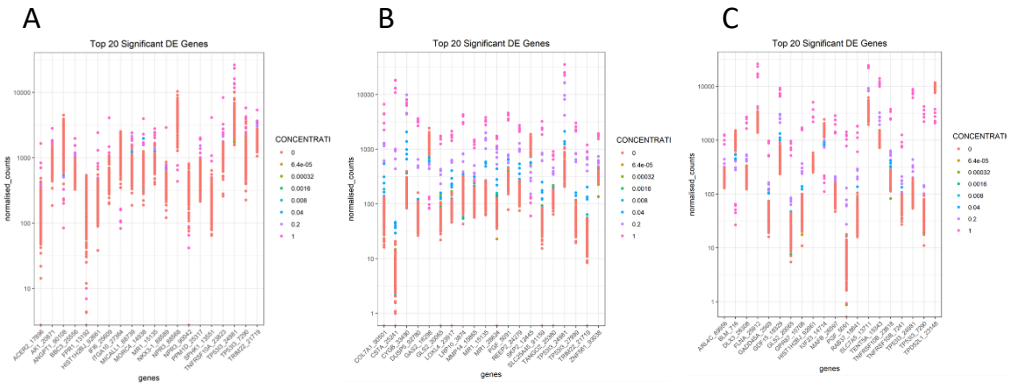


Figure 11: ggplot plots that depict the normalised count of top 20 DEGs coloured by concentration – (A) HepaRG doxorubicin; (B) HepG2 doxorubicin; (C) MCF-7 doxorubicin.

MA plots were conducted for doxorubicin samples and we observed a large number of DEGs with HepaRG showing the largest number of DEGs, followed by HepG2 and then MCF7 cells (Figures 10). A p-value cut off of <0.1 was implemented which illustrated a MA plot with the most significantly expressed genes. The blue dots depict the differentially expressed genes with a p value of <0.1 while the grey dots represent genes with a p value >0.1. The line in the middle of the plot represents the threshold, with genes above this line being upregulated and genes below this line being downregulated. From these plots, HepaRG cells appear to have the most DEG's, followed by HepG2 cells with the least amount in MCF-7 cells. The increased number of blue dots means the more DEG's are present in these cells, with similar proportion of up and down regulated genes for HepaRG and MCF-7; slightly more down regulated genes compared to up regulated for HepG2 cells.

A B C

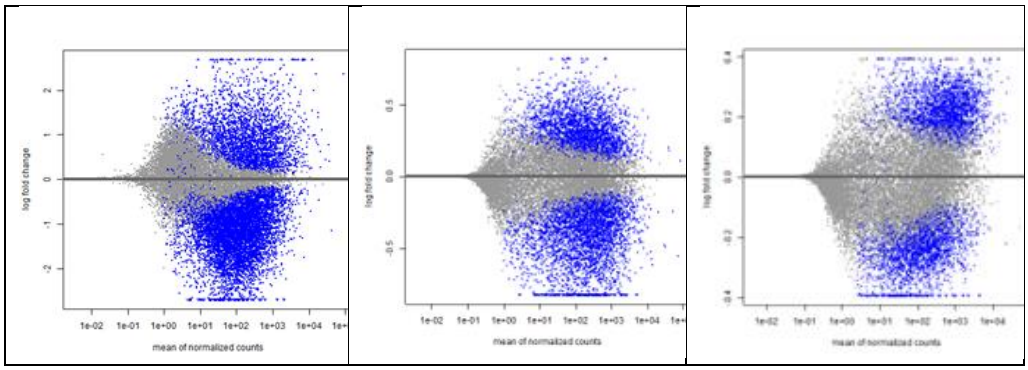


Figure 12: MA plots that depict the distribution of significantly expressed genes in doxorubicin-treated samples with p-value cut off of <0.1 in (A) HepaRGs; (B) HepG2s; (C) MCF-7s. The blue dots depict the DEGs with a p-value of <0.1 while the grey dots represent genes with a p-value >0.1. The line

in the middle of the plot represents the threshold, with genes above this line being upregulated and genes below this line being downregulated.

For doxorubicin samples, volcano plots were constructed as illustrated in Figures 11. showing HepaRG cells had a similar number of up regulated and down regulated genes (red spots), while HepG2 and MCF7 cells showed to have slightly increased numbers of up regulated genes compared to down regulated. The amount of significantly expressed genes are similar for all 3 cell lines. These volcano plots correlate to the MA plots illustrated above.

A B C

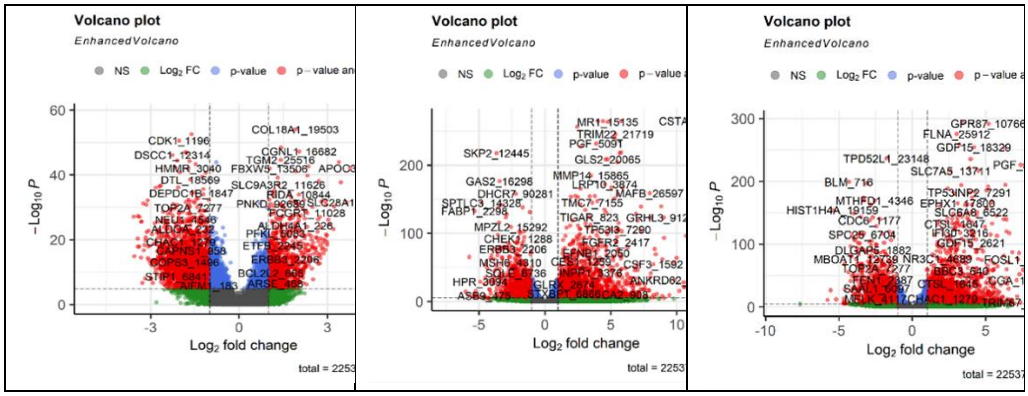


Figure 13: Volcano plots that depict the distribution of p-value and fold change for DEGs (A) HepaRG doxorubicin; (B) HepG2 doxorubicin; (C) MCF-7 doxorubicin

3.1.2 Benchmark dose analysis

For a comprehensive statistical analysis of the dose response data, we used the BMDExpress 2 software (See Materials and Methods). The input for BMDExpress 2 was prepared from the pre-processed intensity and read count data after quality filtering as described in the methods section for BMDExpress2. Within BMDExpress2, probes were first filtered for a significant concentration response using a Williams Trend Test with threshold p-value <0.05 and minimum fold change of 1.5 across concentrations tested. The data were then modelled using the following seven parametric models: Linear, Poly 2, Hill, Power, Exponential 3, 4, and 5, with recommended default configurations. For more details see Materials and Methods.

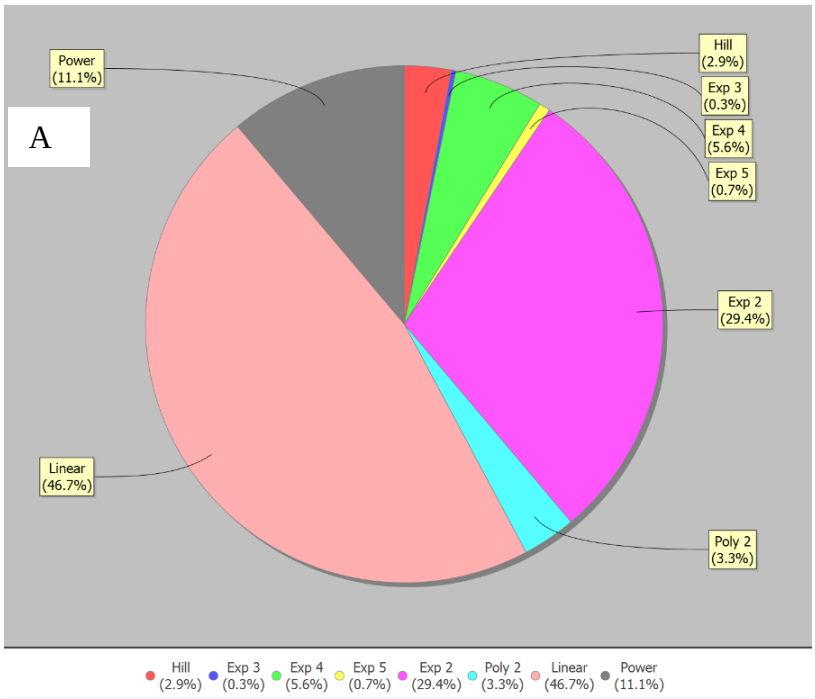
Table 3 summarises the number of features passing the further pre-filtering with Williams trend test and Benchmark dose analysis within BMDExpress 2. During pre-filtering, the data is restricted to features that show some indication of a dose response trend, to limit the amount of data for the subsequent model fitting. Thus, the number/fraction of features passing pre-filtering can give a first rough estimate how many transcripts are affected by the compound treatment in a dose-dependent manner.

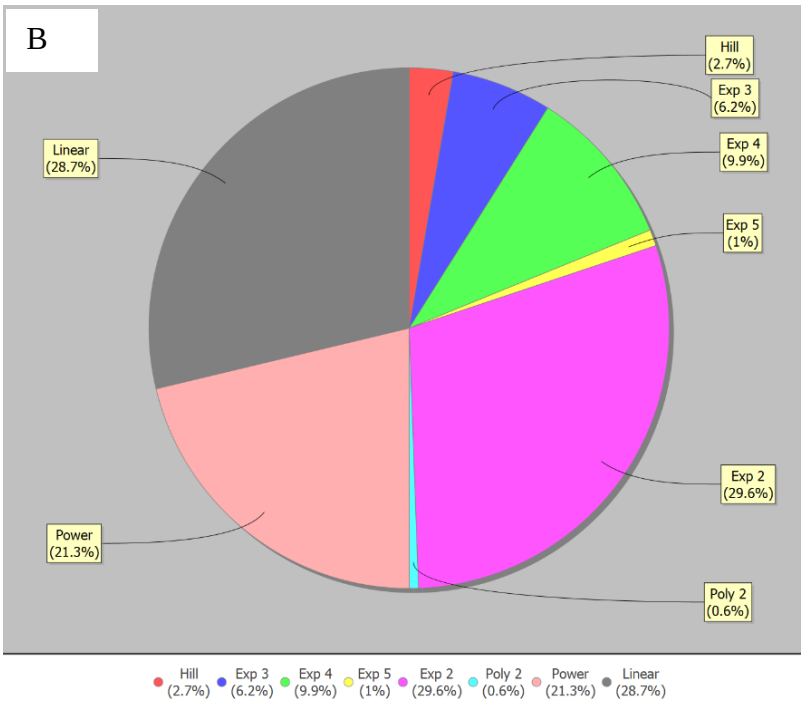
Dataset name	Number of features	Passing Williams Trend test prefiltering	Passing BMD analysis
HepaRG doxorubicin	22537	718	498
MCF-7 doxorubicin	22537	7793	5967
HepG2 doxorubicin	22537	7983	5557

Table 3. Number of features passing the pre-filtering steps in BMDExpress for HepaRG, HepG2 and MCF-7 cells dosed with doxorubicin

The features passing pre-filtering were subsequently used as input for the actual BMD analysis, where seven different model equations were fitted to the dose response data and the best model was selected using the Akaike information criterion (AIC) (Bevans, R, 2023) as described in the methods. From the fitted best model the so-called benchmark dose (BMD) can be obtained, which is the lowest concentration in the curve at which a critical effect size (in the present study set to 10% difference) is observed. Additionally, also a 95% confidence interval for the BMD value is computed and the lower (BMDL) and upper (BMDU) borders of this interval are returned. For the downstream analyses, we wanted to focus on features with reliable curve fits and meaningful BMD values and therefore filtered the BMD analysis output as described in the methods section.

The overall distributions of model types among the best models vary between the datasets indicating that the actual shapes of the dose response curves might differ. Nevertheless, we also observe some common patterns. For example in most datasets a large fraction of the best models followed either Linear or the exponential (Exp2) equations followed by Power for HepaRG and MCF7 samples with HepG2 showing either Linear, Power followed by exponential (Exp4) as illustrated in Figures 12.


A
Power (11.1%)
Hill (2.9%)
Exp 3 (0.3%)
Exp 4 (5.6%)
Exp 5 (0.7%)
Exp 2 (29.4%)
Linear (46.7%)
Poly 2 (3.3%)
Hill (2.9%)
Exp 3 (0.3%)
Exp 4 (5.6%)
Exp 5 (0.7%)
Exp 2 (29.4%)
Poly 2 (3.3%)
Linear (46.7%)
Power (11.1%)


B
Hill (2.7%)
Exp 3 (6.2%)
Exp 4 (9.9%)
Exp 5 (1%)
Exp 2 (29.6%)
Poly 2 (0.6%)
Power (21.3%)
Linear (28.7%)
Hill (2.7%)
Exp 3 (6.2%)
Exp 4 (9.9%)
Exp 5 (1%)
Exp 2 (29.6%)
Poly 2 (0.6%)
Power (21.3%)
Linear (28.7%)

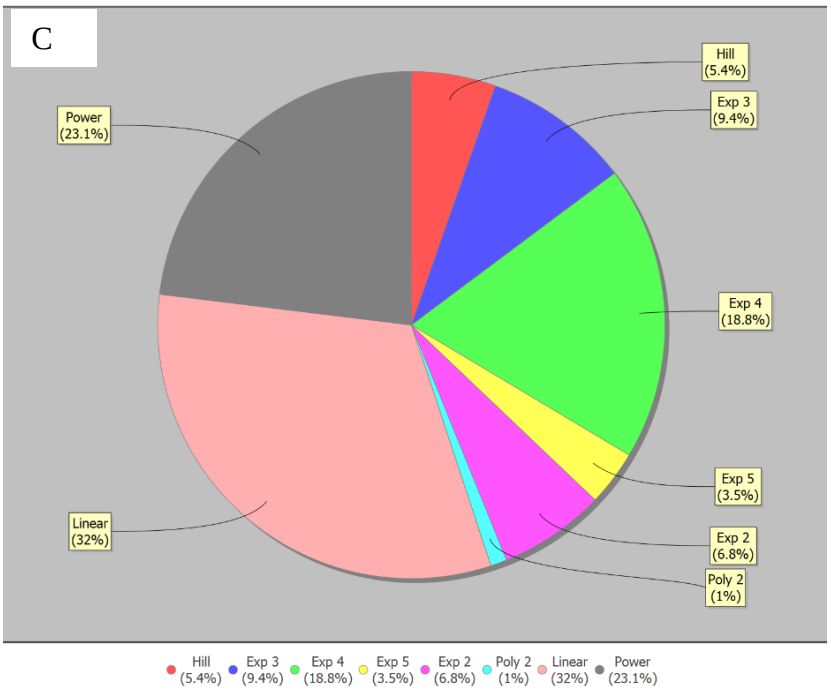


Figure 14: Distribution of equation types among best models ("BMDS Model Counts"). Pie charts show for each dataset the fraction of best models per equation type that pass the output filter criteria. (A) HepaRG doxorubicin (B) MCF-7 doxorubicin (C) HepG2 doxorubicin

Within BMDExpress2, probes which passed the Williams Trends Test are mapped to Reactome pathways with pathway level BMDLs are calculated as detailed in the materials and method section.

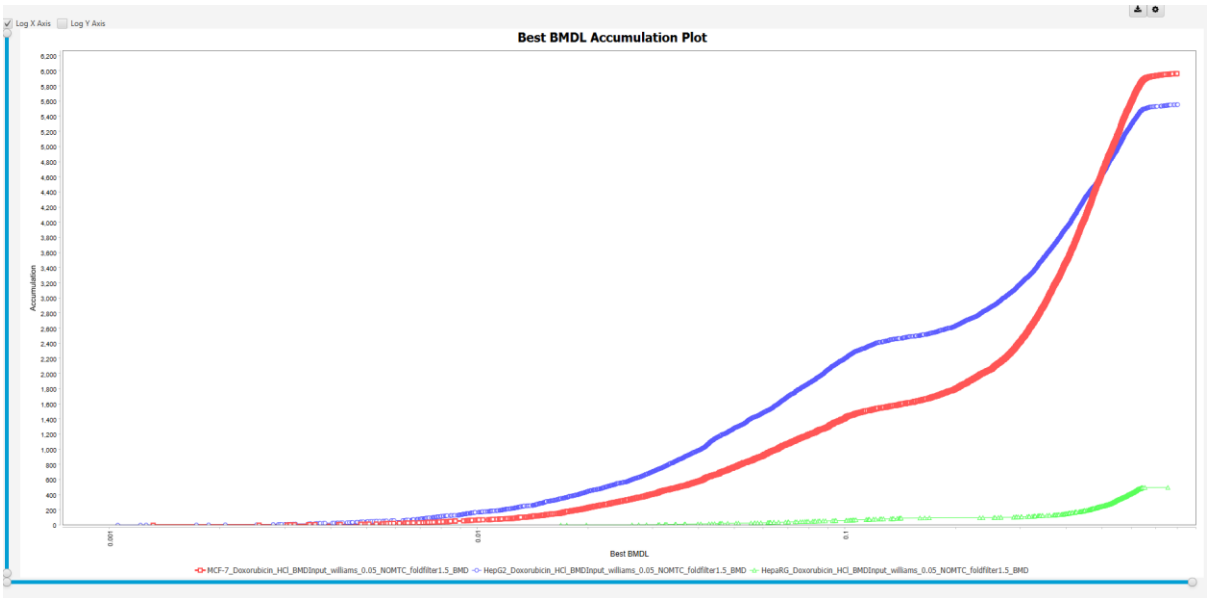


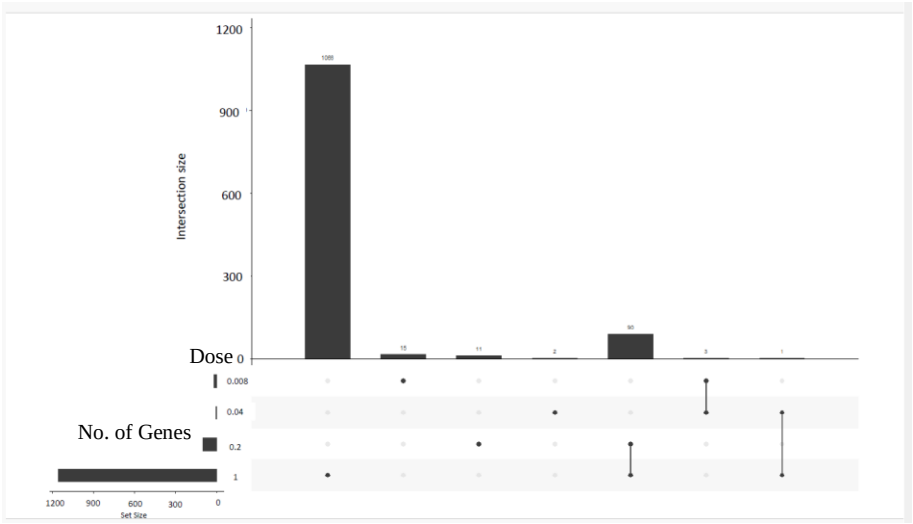
Figure 15: Overlayed accumulation plots of the best model BMDL values for doxorubicin in HepaRG (green), HepG2 (blue) and MCF-7 (red) cells. Accumulation plots include all features passing output filter criteria.

BMDExpress2 also allows generation of accumulation plots, which plot for each dataset the ranks of all features (that pass output filtering) sorted by increasing BMD/BMDL value, against the BMD/BMDL values. Thus, these accumulation plots allow global assessment of the sensitivity against the compound within a dataset and an overall comparison between datasets. Figure 13 illustrates the comparison accumulation plots for doxorubicin in HepaRG, HepG2 and MCF7 cells at the gene level. As per Figure 13, both HepG2s and MCF-7 show higher sensitivity to doxorubicin in comparison to HepaRG cells. This trend is similar to the results for the number of DEGs for the three cell lines before the pre-filtering procedure (Table 1).

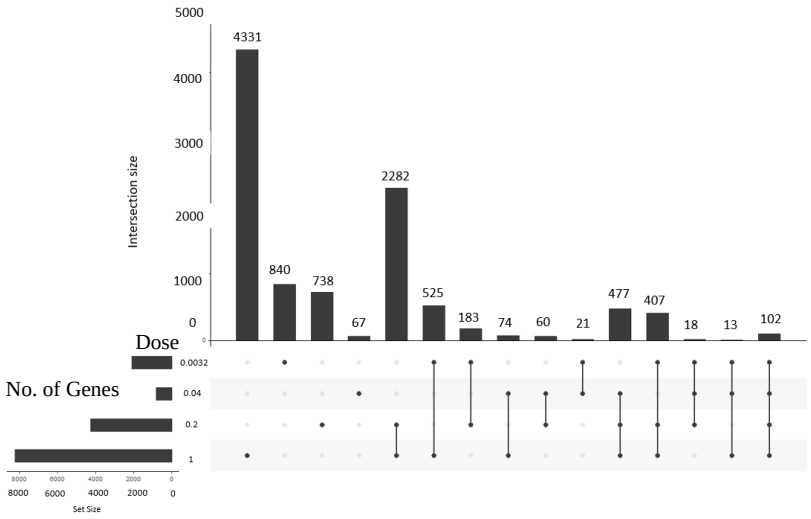
Appendix 3 Supplementary Tables 1-3 illustrate the six lowest probe BMDL individual curve fits for HepaRG, HepG2 and MCF-7 cells dosed with doxorubicin respectively. Each graph shows the BMDL, BDM and BMDU values along with the best curve fit for each gene.

While the accumulation plots provide an overview of the overall sensitivity per dataset, it is also important to examine whether the affected features are the same between doses for each cell line dosed with increasing concentration of doxorubicin. To this end, we examined the overlap of (dose-responsive) features between the different datasets as illustrated in Figure 14,

A


1200
900
600
300
0
Intersection size
Dose
0
0.008
0.04
0.2
1
No. of Genes
1200
900
600
300
0
Set Size

B


5000
4000
3000
2000
1000
0
Intersection size
4331
840
738
67
2282
525
183
74
60
21
477
407
18
13
102
Dose
0.0032
0.04
0.2
1
No. of Genes
8000
6000
4000
2000
0
Set Size

C

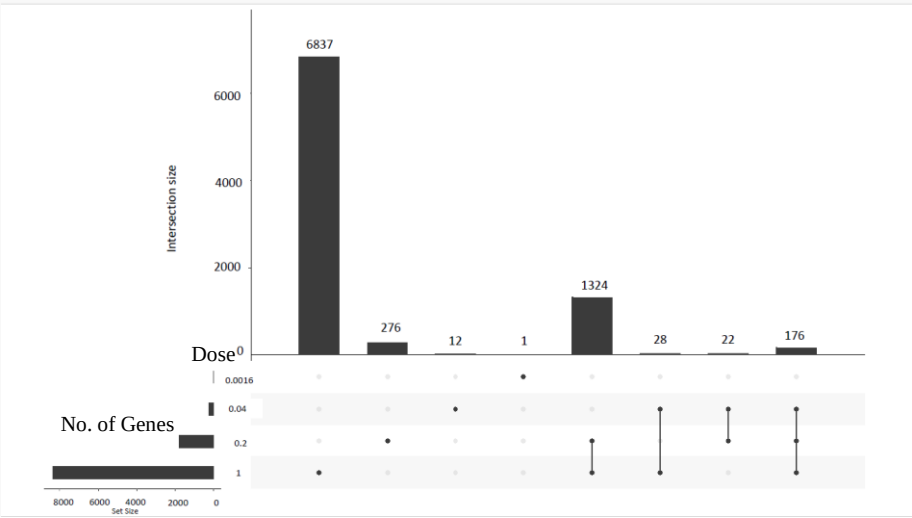


Figure 14: Feature overlap between doxorubicin doses. UpSet plots visualize the feature overlap between the different concentrations for HepaRG, HepG2 and MCF-7 datasets dosed with doxorubicin, considering all features imported into BMDExpress (A), HepaRG at 1, 0.2, 0.04 & 0.008µM (B) HepG2 at 1, 0.2, 0.04 & 0.00032µM (C) MCF-7 at 1, 0.2, 0.04 & 0.08µM

The upset plots were constructed using R Studio with R script detailed in Appendix 1, supplementary script2 and GitHub link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/IPA_gene_intersect%20upset.R

For HepaRG and MCF-7 cells there was a small number of overlaps between the doses, with 90 and 30 features identified between 1 and 0.2µM for HepaRG and MCF-7 datasets respectively, with ≤ 3 features identified between the other doses.
Hep G2 datasets identifed many overlaps between doses, with the largest overlap being 2282 features identified between 1 and 0.2µM, and 525 features identified between 1 and 0.00032µM, showing that HepG2 datasets have larger feature overlaps than the other two cell lines.

Appendix 4 Supplementary Tables 1-3 show the summaries of the top twenty most significantly enriched genes based on the Highest Fold Change Absolute for HepaRG, HepG2 and MCF-7 cells respectively dosed with doxorubicin at the gene level.

3.1.3 Functional classification analysis - BMDExpress

To examine a potential mechanism of action of doxorubicin we next identified biological functions or pathways that were affected by the compound. For this purpose, we performed a functional classification analysis in BMDExpress 2 as described in the methods section Functional classification analysis, searching

for over-/under-represented Gene Ontology (GO) terms and REACTOME pathways among the identified compound-responsive features. To focus on reliably identified functional categories, we filtered the category analysis output based on the criteria as described in the methods section Functional classification analysis. First, we looked at the number of identified pathways across the 3 cell lines as represented by the accumulation plot (Figure 17). Not surprisingly, the accumulation plots correlate to the BMDExpress data at the gene level showing HepaRG identified a lower number of pathways compared to HepG2 and MCF-7 dosed with doxorubicin.

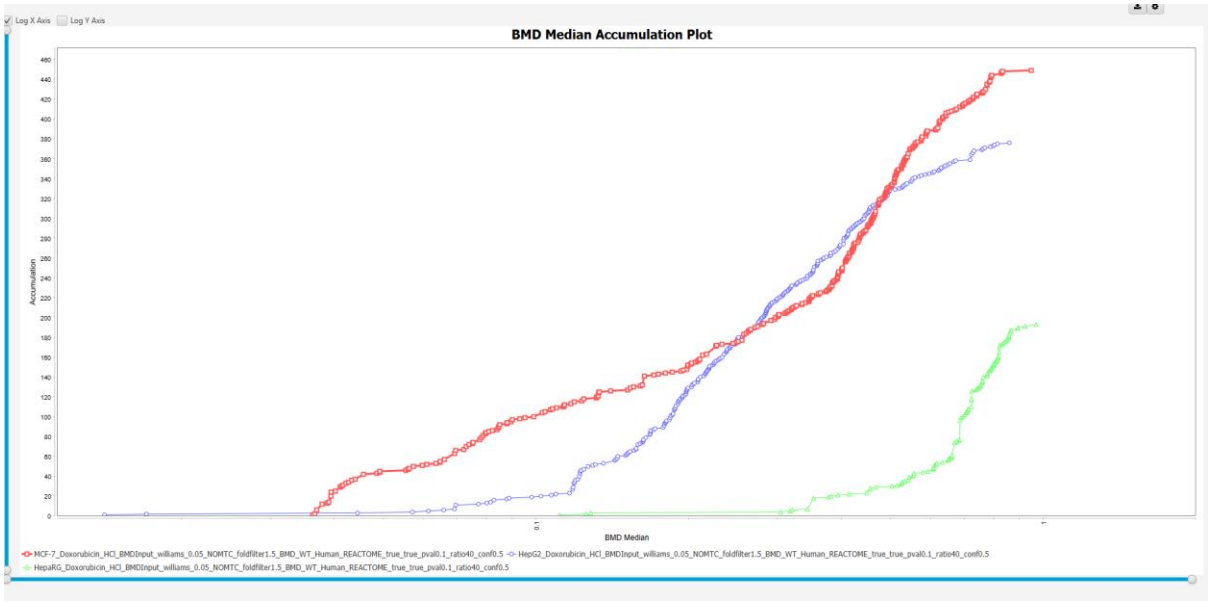


Figure 17: Overlaid accumulation plots of median BMDL values per REACTOME pathway for doxorubicin in HepaRG (green), HepG2 (blue) & MCF7(red) cells. Accumulation plots show median BMDL values across all features associated to a REACTOME pathway that were obtained by category analyses on all features passing the output filter criteria.

Then we looked at the top twenty most significantly enriched pathways based on the lowest BMDL values with P value <0.05 for HepaRG, HepG2 and MCF-7 cells respectively dosed with doxorubicin at the pathway level (see Appendix 4 Supplementary Tables 4-6). This approach revealed the top pathway that was perturbed at the lowest dose was TP53 Regulates Transcription of cell death genes pathway for HepaRG; G2/M DNA replication checkpoint for MCF-7 cells and APC-Ccd20 mediated degradation of Nek2A for Hep G2 cells.

All the cell lines, HepaRG, HepG2 and MCF-7 met the recommendation (Farmahin et al., 2017) that at least 20 pathways were detected to apply the pathway-level tests (detailed in Materials and Methods section). Using this

selection, the observed lowest mean pathway-level BMDL ranged from 0.0219µM to 0.1557µM for doxorubicin across cell lines as detailed in Table 10.

A

Cell line/Chemical	**Average of 20 lowest pathway BMDLs (PoD) µM**	**Average BMDL with the lowest p value (PoD) µM**	Number of features
HepaRG Doxorubicin	0.3065	0.4854	20
Hep G2 Doxorubicin	0.0954	0.2308	20
MCF7 Doxorubicin	0.0610	0.2313	20

B

Cell line/Chemical	**Lowest mean pathway BMDL**	**Number of features**	**Lowest mean pathway at the lowest dose**
HepaRG Doxorubicin	0.1557	20	TP53 Regulates Transcription of Cell Death Genes
Hep G2 Doxorubicin	0.0392	20	APC-Cdc20 mediated degradation of Nek2A
MCF7 Doxorubicin	0.0219	20	G2/M DNA replication checkpoint

Table 4. Summary of BMDL computation results for HepaRG, HepG2 & MCF-7 cell lines dosed with doxorubicin. Tables summarize the values obtained with two different approaches for all features passing the previously described output filter criteria. A –Mean of top 20 lowest pathway BMDL and–average BMDL with the lowest Fischer Exact Two Tail (P Value); B – Lowest mean pathway BMDL with the most sensitive pathways.

3.2 NIACINAMIDE

3.2.1 Quality Control

As described above for doxorubicin, PCR analysis was conducted to make sure there is a separation between the samples and there were no outliers. For niacinamide we observed no clear separation between treated and untreated samples, but there was a clear separation between different treated sample

concentrations (Figure 16) No outlying replicates could be identified in the replicate correlation analysis where it is normally expected replicate correlation should be over 85%. The ggplots (Figure 17) all showed a varied separation between the different concentrations for the top 20 genes in all cell lines dosed with niacinamide, with some genes showing good correlation with normalised counts increasing with the increased concentrations, however there was still a clear separation between the different treated samples. For some genes, the higher the concentration, the larger the number of normalised counts for the top 20 significantly differentiated genes, but for others it was the opposite. The ggplots correlate to the PCA biplots showing no clear separation between untreated and treated samples, which could be due to errors in the experimental analysis of these samples and beyond the scope of this thesis.

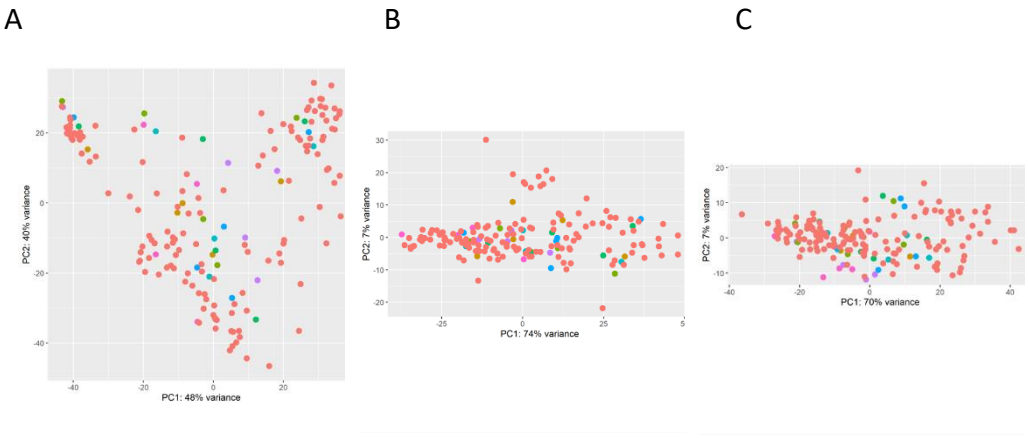


Figure 18: PCA biplots that show the distribution of the samples using top 2 principal components PC1 and PC2 – (A) HepaRG niacinamide; (B) HepG2 niacinamide; (C) MCF-7 niacinamide

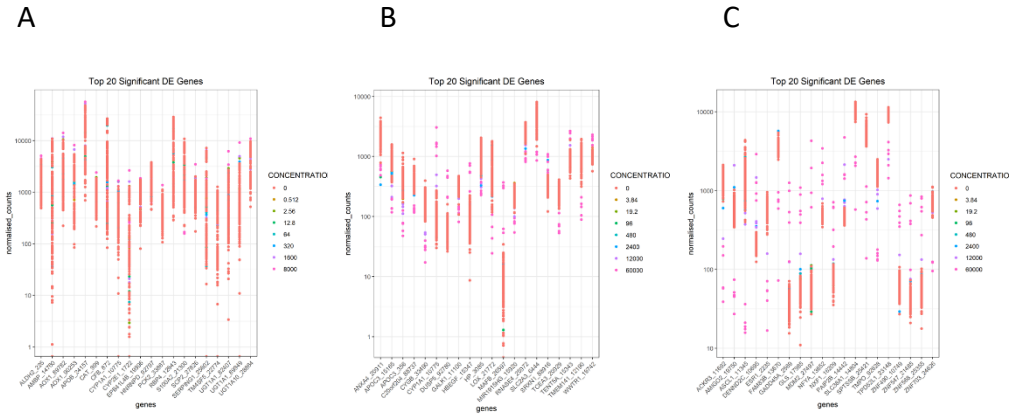


Figure 19: ggplots that depict the normalised count of top 20 DEGs – (A) HepaRG doxorubicin; (B) HepG2 doxorubicin; (C) MCF-7 doxorubicin.

MA plots were conducted for niacinamide samples and we observed a large number of differentially expressed genes with HepaRG showing the largest number of differentially expressed genes, followed by HepG2 and then MCF7 cells (Figures 18). p value cut off of <0.1 was implemented which illustrated a MA plot with the most significantly expressed genes. The blue dots depict the differentially expressed genes with a p value of <0.1 while the grey dots represent genes with a p value >0.1. The line in the middle of the plot represents the threshold, with genes above this line being upregulated and genes below this line being downregulated. From these plots, they show that HepaRG cells appear to have the most DEG's, followed by HepG2 cells with the least amount in MCF-7 cells. The increased number of blue dots means the more DEG's are present in these cells, with similar proportion of up and down regulated genes for all cell lines.

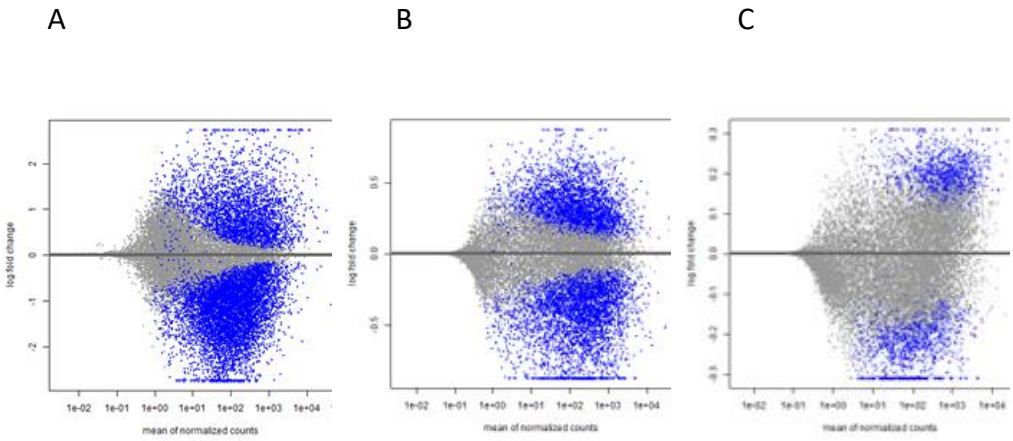


Figure 20: MA plots that depict the distribution of significantly expressed genes in niacinamide-treated samples with p-value cut off of <0.1 in (A) HepaRGs; (B) HepG2s; (C) MCF-7s. The blue dots depict the DEGs with a p-value of <0.1 while the grey dots represent genes with a p-value >0.1. The line in the middle of the plot represents the threshold, with genes above this line being upregulated and genes below this line being downregulated.

For niacinamide samples, volcano plots were constructed as illustrated in Figures 19. showing HepG2 cells had a smaller number of up regulated and down regulated genes (red spots), while MCF7 and HepaRG cells showed to have slightly increased numbers of up regulated genes. There were more significantly expressed genes in HepaRG and MCF7 cells compared to HepG2 cells.

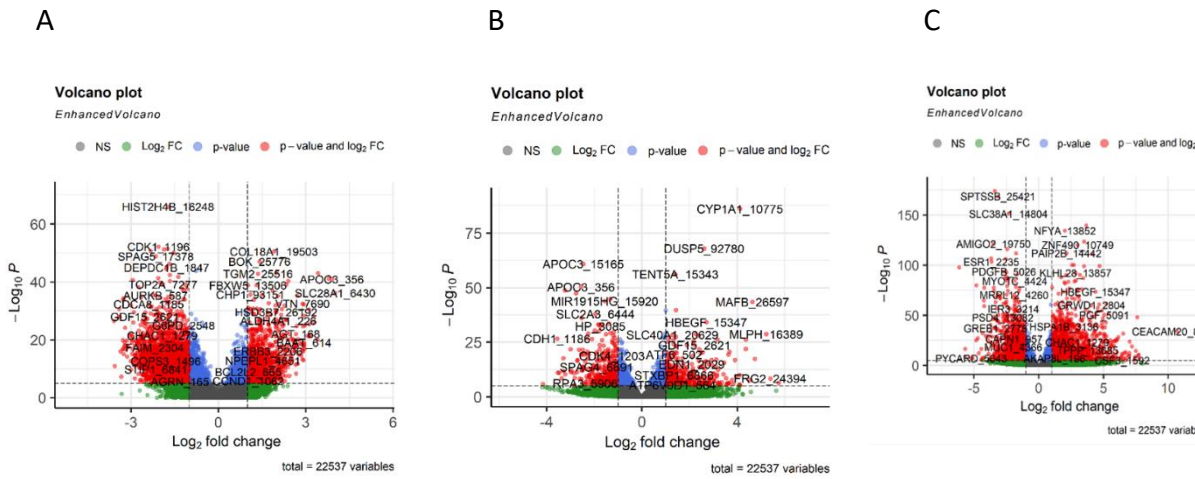


Figure 21: Volcano plots that depict the distribution of p-value and fold change for DEGs – (A) HepaRG niacinamide; (B) HepG2 niacinamide; (C) MCF-7 niacinamide

3.2.2 Benchmark dose analysis

For a comprehensive statistical analysis of the dose response data, we used the BMDExpress 2 software.

The input for BMDExpress 2 was prepared from the pre-processed intensity and read count data after quality filtering as described in the methods section for BMDExpress 2. Within BMDExpress2, probes were first filtered for a significant concentration response using a Williams Trend Test with threshold p value <0.05 and minimum fold change of 1.5 across concentrations tested. The data were then modelled using the following seven parametric models (Linear, Poly 2, Hill, Power, Exponential 3, 4, and 5, with recommended default configurations).

Table 5 summarizes the number of features passing the further pre-filtering with Williams trend test and Benchmark dose analysis within BMDExpress 2.

During pre-filtering, the data is restricted to features that show some indication of a dose response trend, to limit the amount of data for the subsequent model fitting. Thus, the number/fraction of features passing pre-filtering can give a first rough estimate how many transcripts are affected by the compound treatment in a dose-dependent manner.

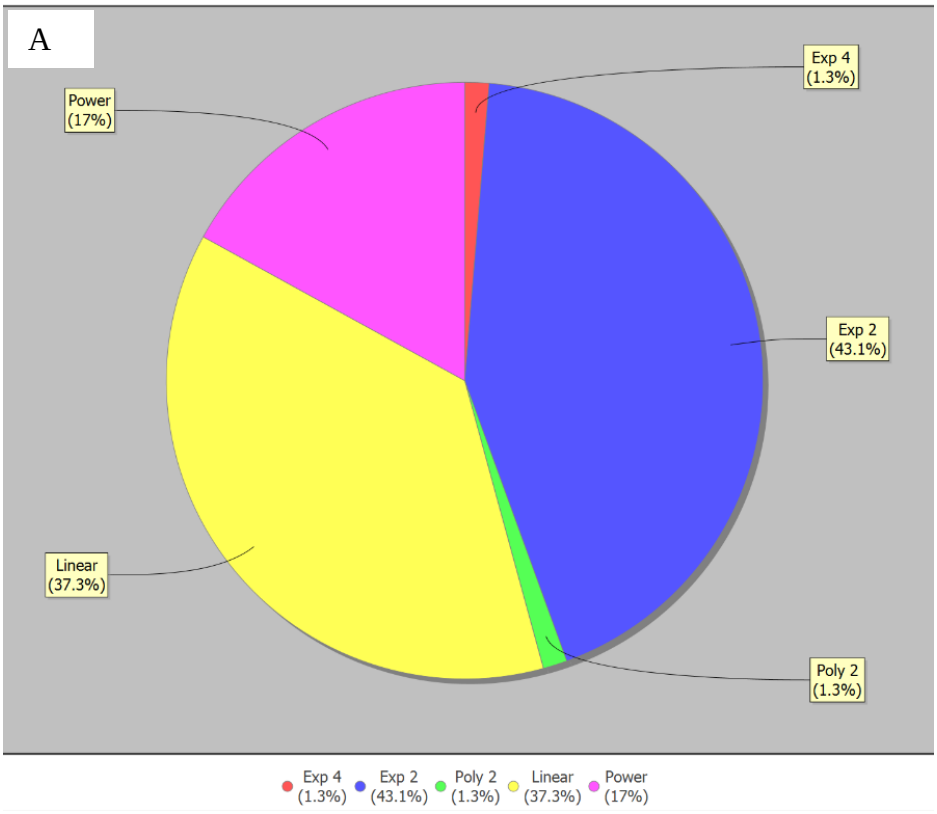
Dataset name	**Number of features**	**Passing Williams Trend test prefiltering**	**Passing BMD analysis**
HepaRG Niacinamide	22537	153	18
MCF-7 Niacinamide	22537	6831	5671

HepG2 Niacinamide	22537	3082	1965

Table 5. NµMber of features passing the pre-filtering steps in BMDExpress for HepaRG, HepG2 and MCF-7 cells dosed with niacinamide

The features passing pre-filtering were subsequently used as input for the actual BMD analysis, as described in the doxorubicin results section.

The overall distributions of model types among the best models vary between the datasets indicating that the actual shapes of the dose response curves might differ. Nevertheless, we also observe some common patterns. For example in most datasets a large fraction of the best models followed either Linear or the exponential (Exp2) equations followed by Power for HepaRG, HepG2 and MCF7 samples as illustrated in Figures 20.



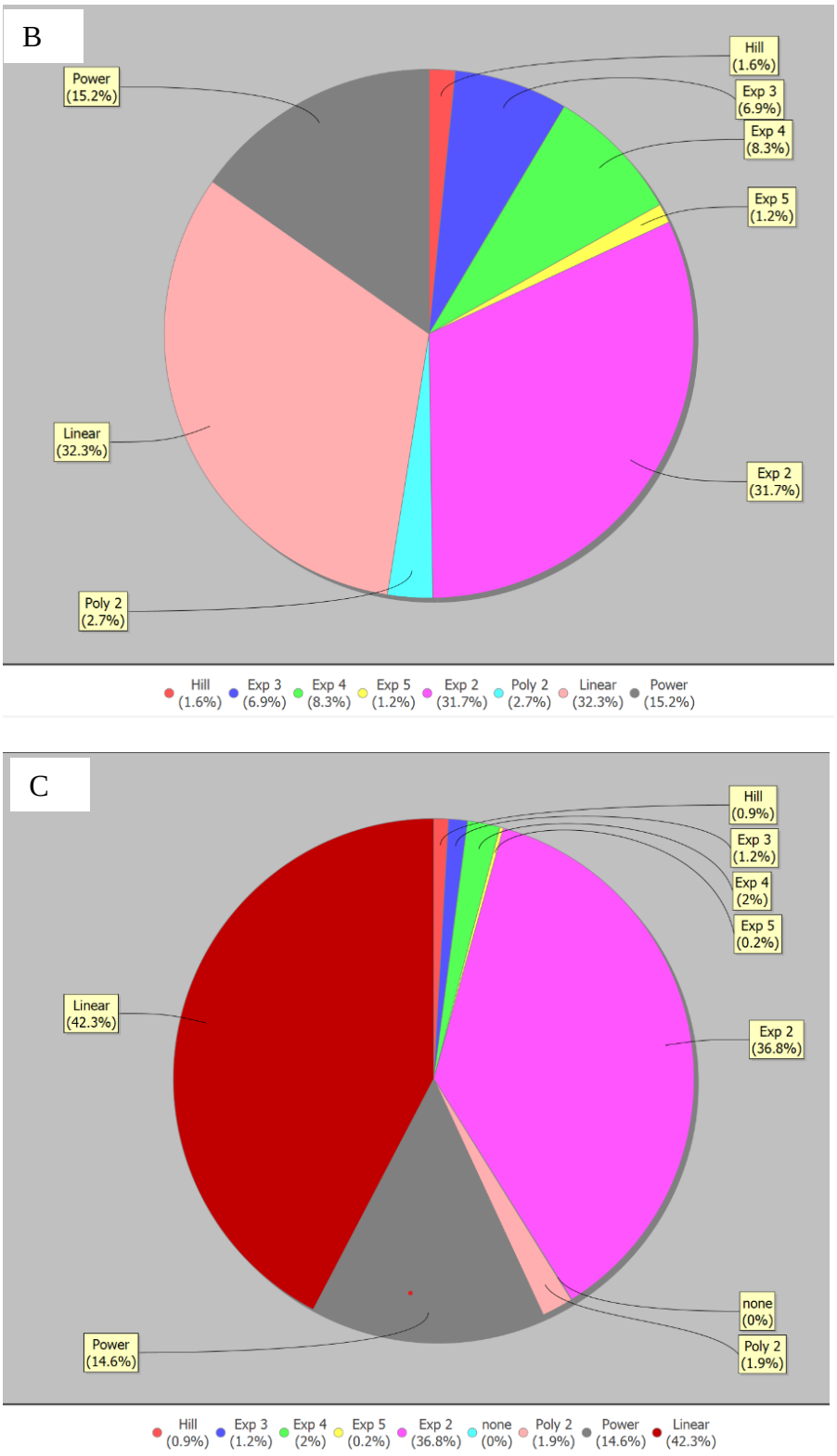


Figure 22: Distribution of equation types among best models ("BMDS Model Counts"). Pie charts show for each dataset the fraction of best models per equation type that the output filter criteria. (A) HepaRG niacinamide (B) MCF-7 niacinamide (C) HepG2 niacinamide

Within BMDExpress2, probes which passed the Williams Trends Test are mapped to Reactome pathways with pathway level BMDLs are calculated as detailed in the materials and method section.

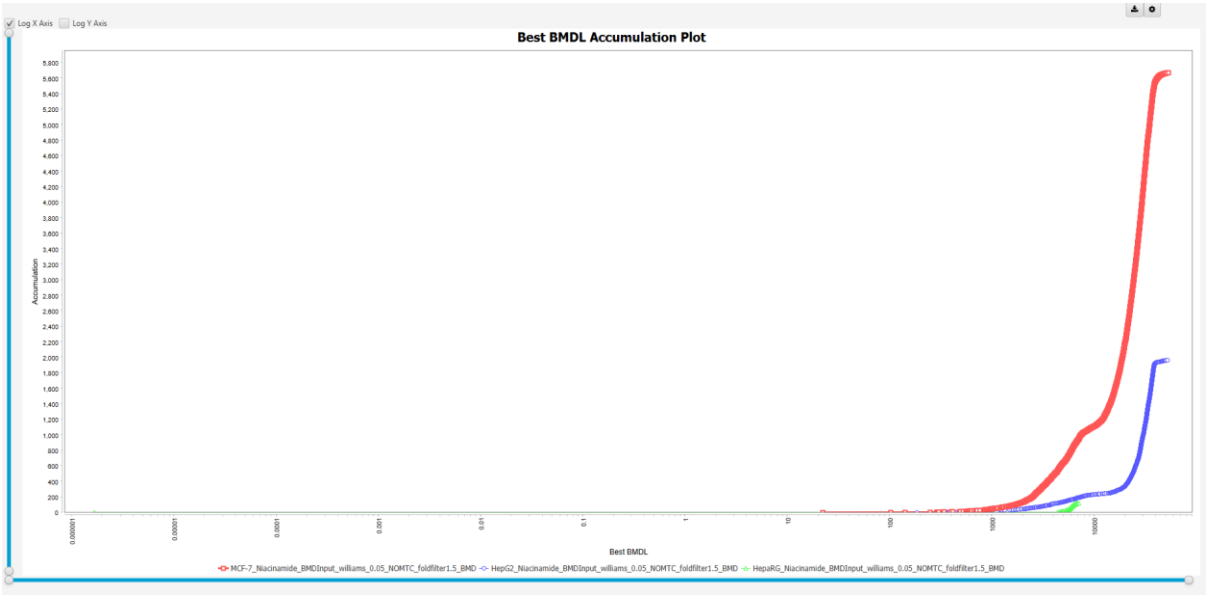


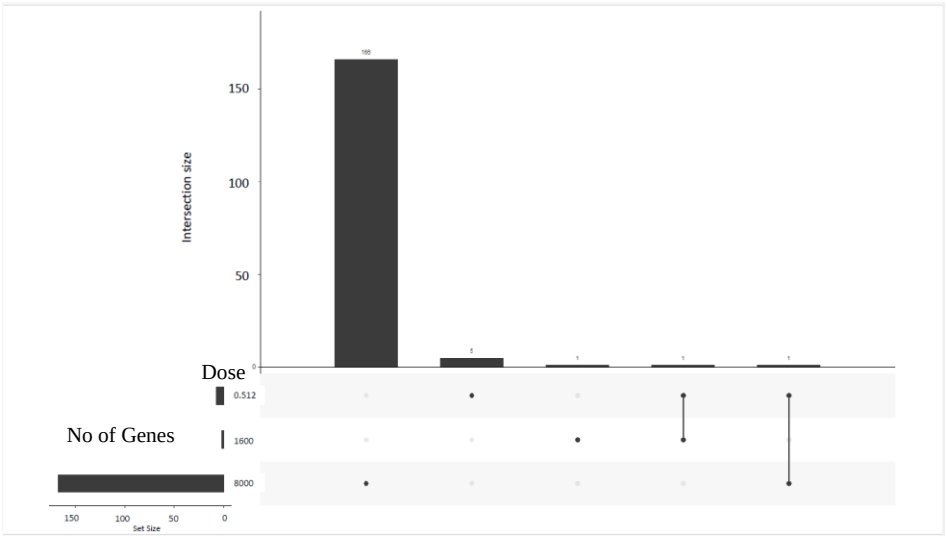
Figure 23: Overlaid accumulation plots of the best model BMDL values for niacinamide in HepaRG (green), HepG2 (blue) and MCF-7 (red) cells. Accumulation plots include all features passing output filter criteria.

BMDExpress2 also allows generation of accumulation plots, which plot for each dataset the ranks of all features (that pass output filtering) sorted by increasing BMD/BMDL value, against the BMD/BMDL values. Thus, these accumulation plots allow global assessment of the sensitivity against the compound within a dataset and an overall comparison between datasets. Figure 21. illustrates the comparison accumulation plots for niacinamide in HepaRG, HepG2 and MCF7 cells at the gene level. As per Figure 21, both HepG2s and MCF-7 show higher sensitivity to niacinamide in comparison to HepaRG cells. This trend is similar to the results for the number of DEGs for the three cell lines before the pre-filtering procedure (Table 5).

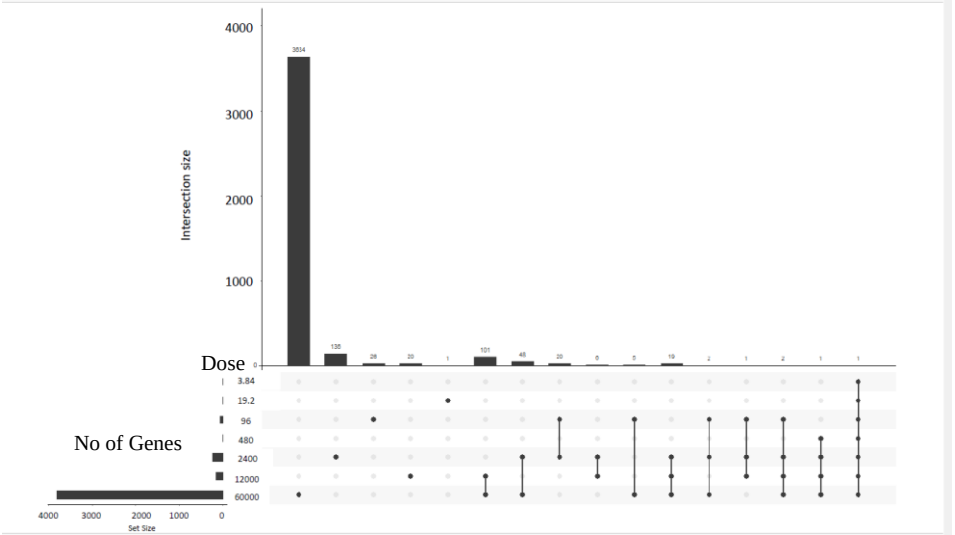
Appendix 4 Supplementary Figures 4-6 illustrate the six lowest probe BMDL individual curve fits for HepaRG, HepG2 and MCF-7 cells dosed with niacinamide respectively. Each graph shows the BMDL, BMD and BMDU values along with the best curve fit for each gene.

While the accumulation plots provide an overview of the overall sensitivity per dataset, it is also important to examine whether the affected features are the same between doses for each cell line dosed with increasing concentration of niacinamide. To this end, we examined the overlap of (dose-responsive) features between the different datasets as illustrated in Figure 22.

A


Intersection size
150
100
50
0
Dose
0.512
1600
8000
No of Genes
150
100
50
0
Set Size

B


Intersection size
4000
3000
2000
1000
0
Dose
3.84
19.2
96
480
2400
12000
60000
No of Genes
4000
3000
2000
1000
0
Set Size

C

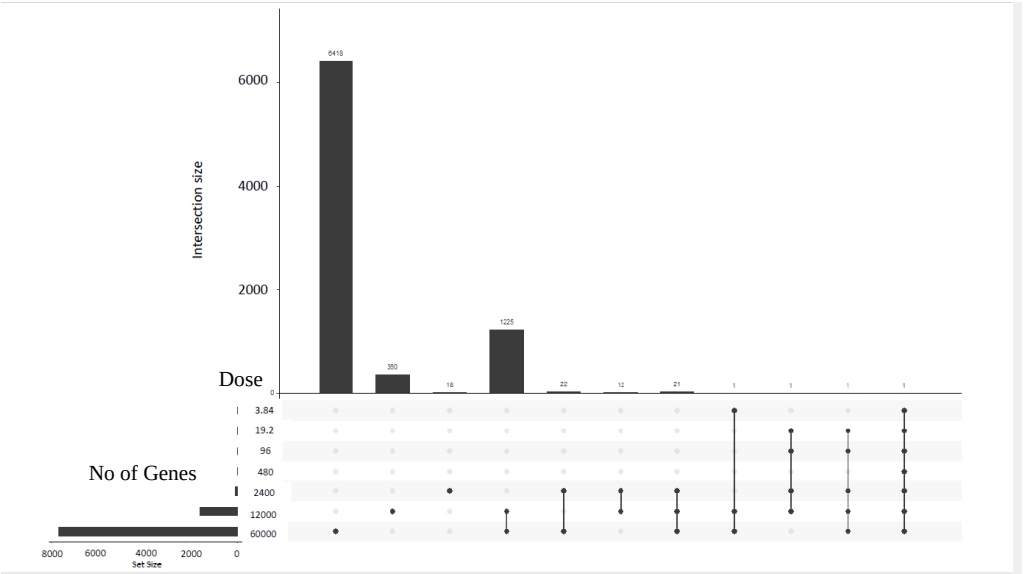


Figure 24: Feature overlap between niacinamide doses. UpSet plots visualize the feature overlap between the different concentrations for HepaRG, HepG2 and MCF-7 datasets dosed with niacinamide, considering all features imported into BMDExpress (A), HepaRG at 8000, 1600, 0.512µM (B) HepG2 at 60000, 12000, 2400, 480, 96, 19.2, 3.84µM (C) MCF-7 at 60000, 12000, 2400, 480, 96, 19.2, 3.84µM

For HepaRG and MCF-7 cells there were a small number of overlaps between the doses, with 1 feature identified between 8000, 1600 and 0.512µM. MCF7 datasets identified several more overlaps between doses compared to HepaRG with the largest overlap being 1225 features identified between 60000 and 12000µM and ≤22 features for the other dose overlaps.
Hep G2 datasets identifed several more overlaps between doses compared to HepaRG, with the largest overlap being 101 features identified between 60000 and 12000µM, 48 features identified between 60000 and 2400µM and ≤20 features for the other dose overlaps.

Appendix 5 Supplementary Tables 7-9 show the summaries of the top twenty most significantly enriched genes based on the Highest Fold Change Absolute for HepaRG, HepG2 and MCF-7 cells respectively dosed with niacinamide at the gene level.

3.2.3 Functional classification analysis - BMDExpress

To examine a potential mechanism of action of niacinamide, we next identified biological functions or pathways that were affected by the

compound. For this purpose, we performed a functional classification analysis in BMDExpress 2 as described in the methods section Functional classification analysis, searching for over-/under-represented Gene Ontology (GO) terms and REACTOME pathways among the identified compound-responsive features. To focus on reliably identified functional categories, we filtered the category analysis output based on the criteria as described in the methods section Functional classification analysis. First, we looked at the number of identified pathways across the 3 cell lines as represented by the accumulation plot (Figure 23). Not surprisingly, the accumulation plots correlate to the BMDExpress data at the gene level showing HepaRG identified a lower number of pathways compared to HepG2 and MCF-7 dosed with doxorubicin.

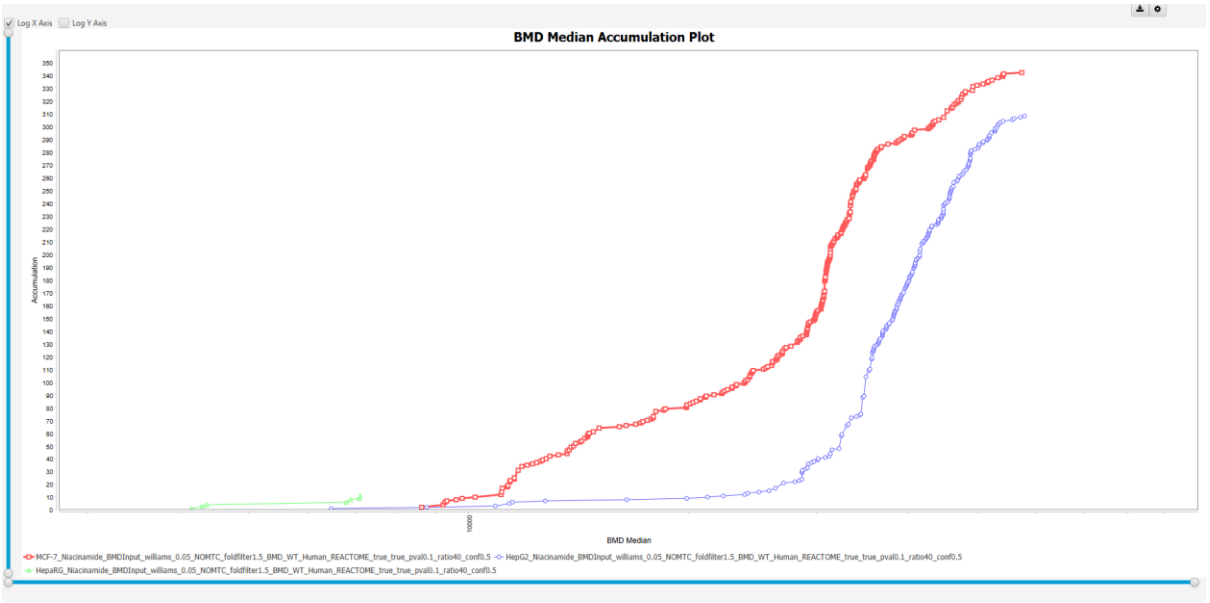


Figure 25: Overlaid accumulation plots of median BMD values per REACTOME pathway for niacinamide in HepG2 (blue), HepaRG (green) & MCF7 (red) cells. Accumulation plots show median BMD values across all features associated to a REACTOME pathway that were obtained by category analyses on all features passing the output filter criteria.

Then we looked at the top twenty most significantly enriched pathways based on the lowest BMDL values with P value <0.05 for HepaRG, HepG2 and MCF-7 cells respectively dosed with niacinamide at the pathway level (see Appendix 5 Supplementary Tables 10-12). This approach revealed the pathway that was perturbed at the lowest dose was Metabolism of lipids pathway for HepaRG; Condensation of Prophase Chromosomes for MCF-7 cells and The AIM2 inflammasome for Hep G2 cells.

All the cell lines, HepaRG, HepG2 and MCF-7 cell lines met the recommendation (Farmahin et al., 2017) that at least 20 pathways were detected to apply the pathway-level tests (detailed in Materials and Methods section). Using this selection, the observed lowest mean pathway-level BMDL ranged from 2552.94µM to 5444.65µM for niacinamide across cell lines as detailed in Table 6

A

Cell line/Chemical	**Average of 20 lowest pathway BMDLs (PoD) µM**	**Average BMDL with the lowest p value (PoD) µM**	Number of features
HepaRG Niacinamide	3492.19	3492.19	20
Hep G2 Niacinamide	14857.81	24977.10	20
MCF7 Niacinamide	8169.58	24395.43	20

B

Cell line/Chemical	**Lowest mean pathway BMDL**	**Number of features**	**Lowest mean pathway at the lowest dose**
HepaRG Niacinamide	2552.94	20	Metabolism of lipids
Hep G2 Niacinamide	5046.89	20	The AIM2 inflammasome
MCF7 Niacinamide	5444.65	20	Condensation of Prophase Chromosomes

Table 6. Summary of BMDL computation results for HepaRG, HepG2 & MCF-7 cell lines dosed with niacinamide. Tables summarize the values obtained with two different approaches for all features passing the previously described output filter criteria. A –Mean of top 20 lowest pathway BMDL and–average BMDL with the lowest Fischer Exact Two Tail (P Value); B – Lowest mean pathway BMDL with the most sensitive pathways.

3.2.4 Points of Departure calculations

PoDs were calculated for both chemicals using an R scripts generated by Unilever, SEAC (R script detailed in Appendix 4 Supplementary Script 4 and GitHub links: https://github.com/liztulum/MRes-thesis-scripts/blob/main/calculate_pods.R and Appendix 4 Supplementary Script 5 and GitHub links: https://github.com/liztulum/MRes-thesis-scripts/blob/main/Calculate_PoDs_from%20_BMDExpress2.R) – doxorubicin

and niacinamide based on probe BMDLs. The script applies the methodology published in Farmahin et al. (2017) where BMDExpress2 *in vitro* PoDs were calculated in a way to correlate with the reference/benchmark PoD data set derived from *in vivo* studiesµMAll the cell lines, HepaRG, HepG2 and MCF-7 cell lines met the recommendation (Farmahin et al., 2017) that at least 20 genes were detected to apply the gene-level tests. Using this selection, the lowest observed gene-level PoD ranged from 0.0855µM to 0.2569 M for doxorubicin and 4132.79µM to 11150.48 µM for niacinamide across cell lines using the lowest average BMDL values with the highest fold change values.

The points of departures for each cell line are defined as the point on a toxicological dose-response curve generally corresponding to an estimated low effect level or no effect level at which a biological response is first observed. The niacinamide results in Table 7 identify that the PoD's for HepG2 and MCF-7 are higher compared to HepaRG, meaning the no effect level for these latter cells is lower compared to the others. This would indicate that HepaRG cells response to niacinamide is more potent at lower concentrations compared to HepG2 and MCF7 cells. For doxorubicin, this is the opposite with HepaRG cells having a higher PoD compared to HepG2 and MCF-7 meaning the latter two cell lines indicate the response to doxorubicin is more potent at lower concentrations.

Cell line/ Chemical	**Number of features**	**Average BMDL with the highest fold change (PoD) µM**	**Average BMDL within 25th to 75th percentile (PoD) µM**	**Average of 20 lowest pathway BMDLs (PoD) µM**	**Average pathway BMDL with the lowest p value (PoD) µM**	**Lowest mean pathway BMDL**
HepaRG niacinamide	20	4132.79	5068.61	3492.19	3492.19	**2552.94**
Hep G2 niacinamide	20	8210.79	37242.64	14857.81	24977.10	**5046.89**
MCF7 niacinamide	20	11150.48	34934.10	8169.58	24395.43	**5444.65**
HepaRG doxorubicin	20	0.2569		0.3065		**0.1557**

			0.6086		0.4854	
Hep G2 doxorubicin	20	**0.0313**	0.5365	0.0954	0.2308	0.0392
MCF7 doxorubicin	20	0.0856	0.5625	0.0610	0.2313	**0.0219**

Table 7 – Summary of the average BMDL values (PoD) derived using the approach published in Farmahin et al. (2017) for HepaRG, HepG2 and MCF-7 cells dosed with doxorubicin and niacinamide including the lowest PoD for each cell line selected in bold. The average BMDL values at the gene level were calculated by averaging the lowest BMDL values with the highest fold change and the average of BMDLs within the 25th to 75th percentile.

The selected PoDs were derived using BMDExpress2 for HepaRG, HepG2 and MCF-7 cells dosed with doxorubicin were 0.1557µM, 0.0313µM & 0.0219µM respectively. The selected PoD's for HepaRG, HepG2 and MCF-7 cells dosed with niacinamide were 2552.94µM, 5046.89µM & 5444.65µM respectively.

CHAPTER 4

4. RESULTS FOR IPA

4.1 INGENUITY PATHWAY ANALYSIS (IPA) - PATHWAY AND UPSTREAM REGULATOR ENRICHMENT ANALYSIS

The lists of genes with BMDs were uploaded, along with the maximum fold change and Williams trend test p-values, into IPA to determine if there was enrichment of genes fitting models to specific canonical pathways and associated with regulation by specific upstream molecules. Visual Studio code (version 1.77.1) was conducted to convert the DESeq results to compatible results for IPA as detailed in Appendix 1, supplementary script 3 and GitHub link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/gene%20name%20change.py. This approach allowed the comparison of chemicals across all concentrations tested and perturbed gene sets associated with predicted upstream regulators and canonical pathways. For IPA analysis, it is recommended that datasets with DEGs between 100-3000 only would be analysed, therefore not all concentrations for doxorubicin and niacinamide were analysed using IPA. HepaRG niacinamide dataset gave less than 100 DEGs and were not processed using IPA, various lower concentrations for both chemicals were also not processed using IPA. Table 8 shows the cell line concentrations used in IPA.

Cell line/Chemical	Concentrations used in IPA (µM)
HepaRG doxorubicin	1, 0.2
Hep G2 doxorubicin	1, 0.2, 0.04, 0.00032
MCF-7 doxorubicin	1, 0.2, 0.04
HepaRG niacinamide	-
Hep G2 niacinamide	60000, 12000, 2400
MCF-7 niacinamide	60000, 12000

Table 8. Cell line concentrations used in IPA for doxorubicin and niacinamide

4.1.1 Doxorubicin

Canonical pathways for each cell line and concentration for doxorubicin are illustrated in Figure 26. The orange bars showed a positive z-score which means the pathways was activated and gave upregulated responses while the blue bars showed negative z-scores which means the pathways are inhibited and gave down regulated responses. The grey bars detailed no activity patterns were available and the white bars showed a z-score of 0 therefore no up or down regulation of pathways.

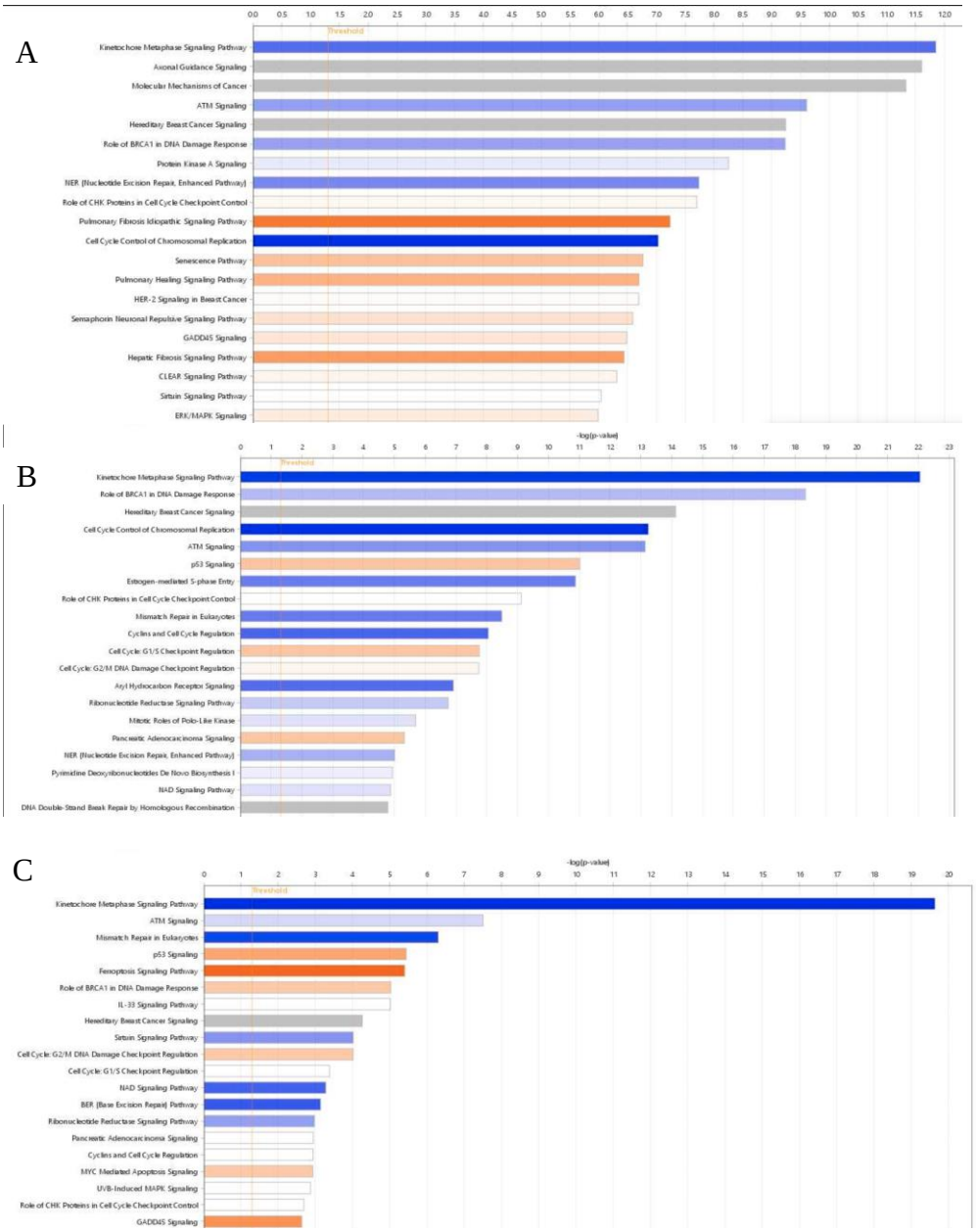
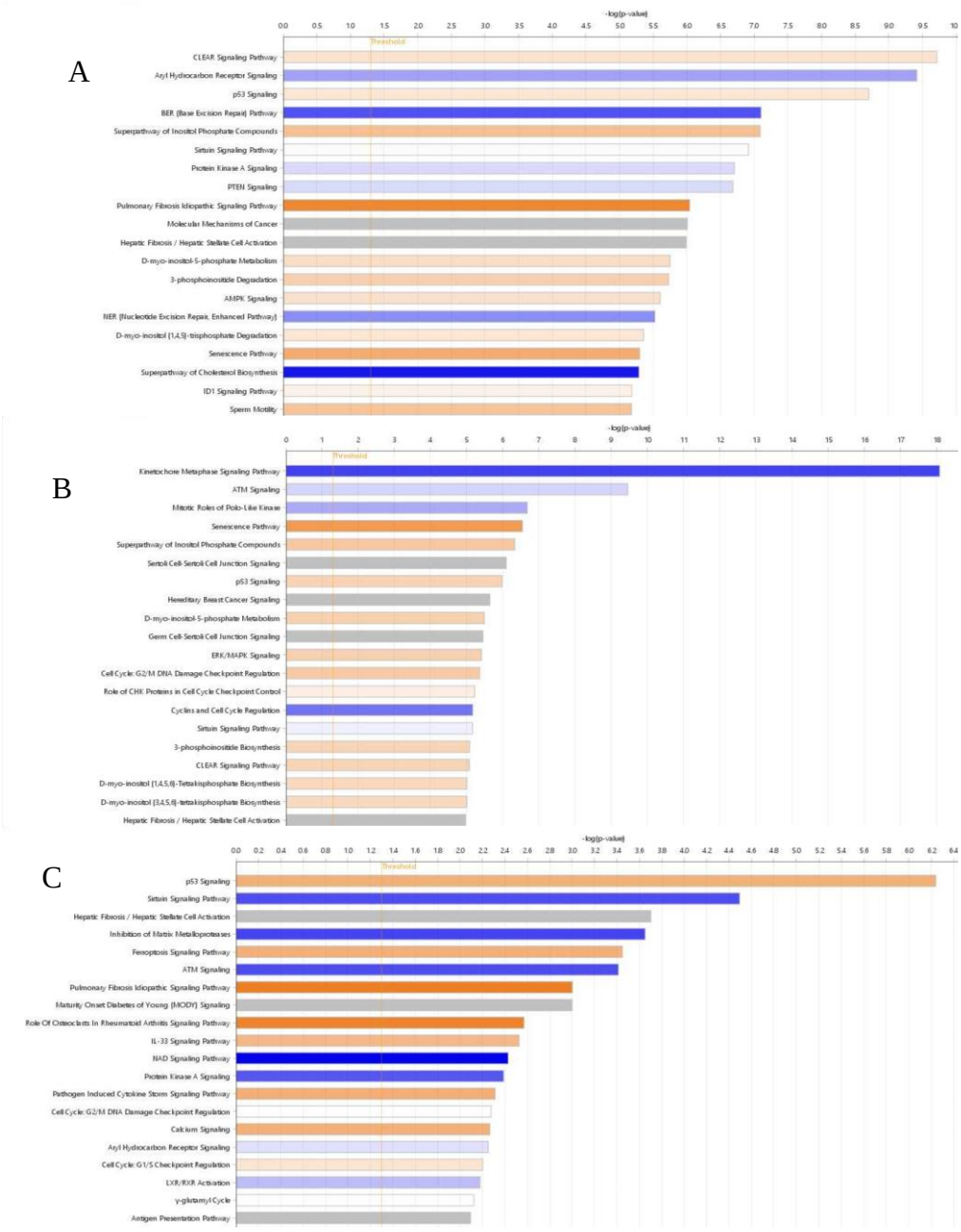


Figure 26: Top 20 canonical pathways identifed using IPA for doxorubicin for each concentration analysed – (A) MCF7 doxorubicin 1µM; (B) MCF7 doxorubicin 0.2µM; (C) MCF7 doxorubicin 0.04µM

For MCF-7 cells, the canonical pathways illustrate most pathways are inhibited rather than activated (Figure 226). For all concentrations, the Kinetochore Metaphase Signalling pathway, p53 signalling and the ATM signalling pathway were identifed to be inhibited the most in the MCF-7 cell line. The Kinetochore Metaphase Signalling pathway is an important checkpoint in the middle of mitosis during which the cell ensures that it is ready to divide, but as it is inhibited this pathway doesn't occur due to the reaction with doxorubicin in the cell process (Navarro et al,. 2021). ATM & P53 Signalling relates to DNA damage repair and apoptosis, therefore cell death. Most of the inhibited

pathways relate to cell cycle, DNA damage, apoptosis, cell death, while the activated pathways relate to immune response pathways as the cells try to recover from chemical treatment i.e. p53 signalling (Abuetabh et al,. 2022). This adheres to the nature of doxorubicin being a highly toxic chemical slowing or stopping the growth of cells during the cell cycle.



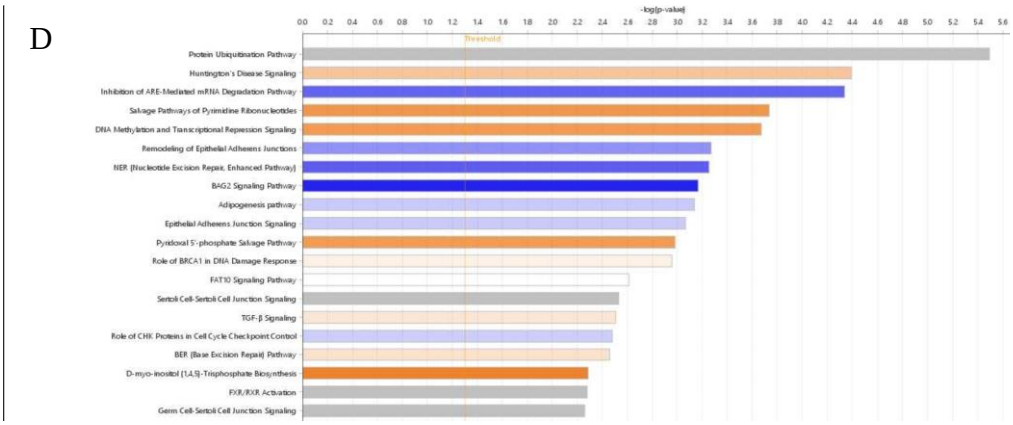
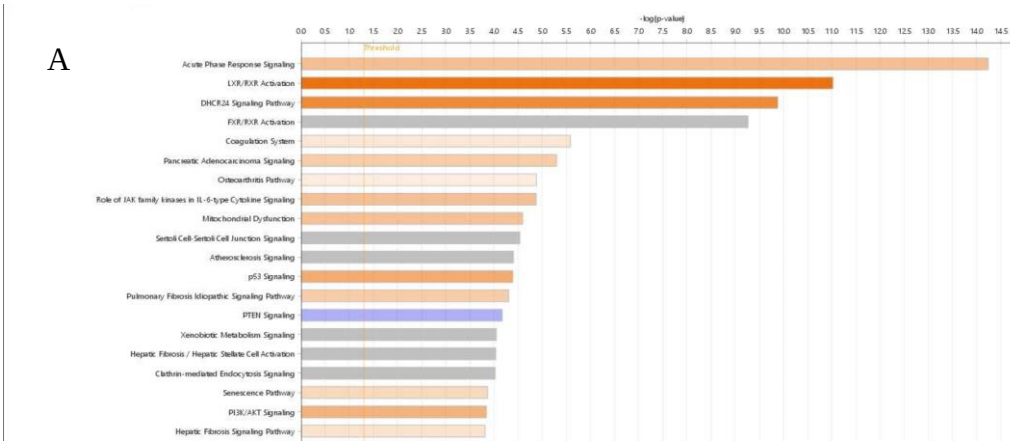


Figure 27: Top 20 canonical pathways identifed using IPA for doxorubicin for each concentration analysed – (A) HepG2 doxorubicin 1µM; (B) HepG2 doxorubicin 0.2µM; (C) HepG2 doxorubicin 0.04µM; (D) HepG2 doxorubicin 0.00032µM

For HepG2 cells, the canonical pathways illustrate most pathways are a mixture of inhibited and activated pathways (Figure 27). For all concentrations, the BER Signalling pathway, PTEN signalling and the ATM signalling pathway were identifed to be inhibited the most in HepG2cell line. ATM & PTEN Signalling relates to DNA damage repair and apoptosis, therefore cell death. The majority of the inhibited pathways relate to DNA damage, apoptosis, cell death, while the activated pathways i.e. p53 signalling relate plays an important role in the co-ordination of the cellular response to different types of stress such as DNA damage and hypoxia with the downstream signals leading to apoptosis and cell cycle arrest (Abuetabh et al,. 2022).



B

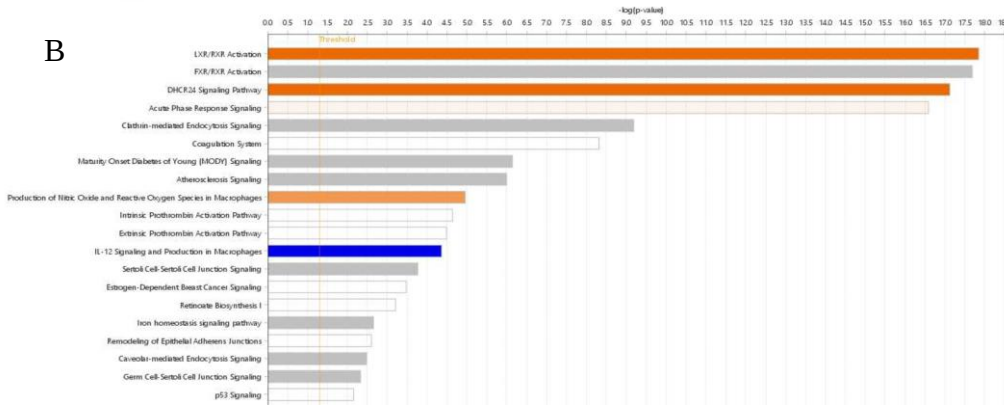


Figure 28: Top 20 canonical pathways identifed using IPA for doxorubicin for each concentration analysed – (A) HepaRG doxorubicin 1µM; (B) HepaRG doxorubicin 0.2µM

For HepaRG cells, the canonical pathways illustrate most pathways are activated (Figure 28). For all concentrations, LXR/RXR activation, p53 signaling and FXR/RXR activation were identifed to be activated the most for the HepaRG cell line which are metabolism related and in agreement with the metabolising capacity of HepaRG cells (Jiang et al,. 2022). PTEN Signaling was identified as inhibited which relates to DNA damage repair and apoptosis, therefore cell death. The majority of the inhibited pathways relate to DNA damage, apoptosis, cell death, while the activated pathways i.e. p53 signaling plays an important role in the co-ordination of the cellular response to different types of stress such as DNA damage and hypoxia with the downstream signals leading to apoptosis and cell cycle arrest (Abuetabh et al,. 2022).

Comparison heatmaps for doxorubicin at the various concentrations are illustrated in Figure 29. These heatmaps show whether the various cell lines have been activated (orange) or inhibited (blue) by the treatment of doxorubicin using the z-score.

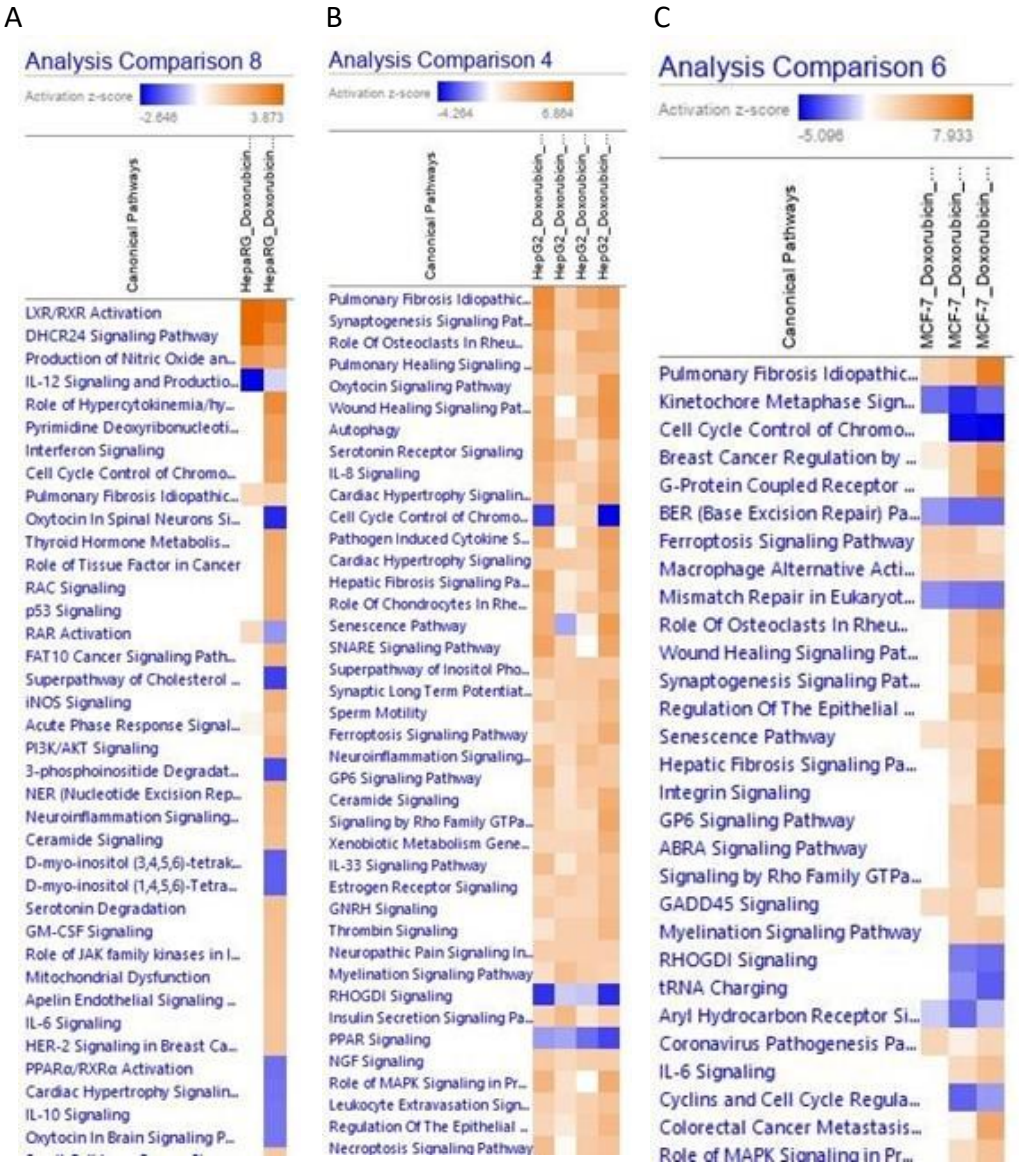


Figure 29: Comparison Heatmaps of the top 20 canonical pathways comparing doxorubicin for each concentration analysed – (A) HepaRG doxorubicin (first colμMn = 0.2μM; second column 1μM); (B) Hep G2 doxorubicin (first column = 1μM; second column = 0.00032μM; third coumn = 0.04μM; fourth coumn = 0.2μM); (C) MCF7 doxorubicin (first column = 0.04μM; second column 0.2μM; third column = 1μM)

This approach allowed us to compare chemicals across all concentrations tested and revealed that the HepaRG, HepG2 and MCF-7 cell lines doses with doxorubicin perturbed gene sets associated with predicted upstream regulators and canonical pathways. For HepaRG, the top 3 pathways showed activated pathways with LXR/RXR activation, DHCR24 Signaling Pathway and production of Nitric Oxide, with IL-12 Signaling and Production showing inhibition. The darker the colour the higher the z-score which means for Hep

G2 1µM sample gives more pathway activation/inhibition compared to 0.2µM.
HepG2 and MCF-7 samples show more activation pathways compared to inhibition, with the higher concentrations having a higher z-score which is to be expected, with Pulmonary Fibrosis Idiopathic pathway in both being activated and Kinetochore Metaphase Signaling being inhibited. Some of these pathways are very similar to what was seen in the BMDExpress results with HepaRG showing cellular response to different types of stress such as DNA damage and hypoxia with the downstream signals leading to apoptosis and cell cycle arrest i.e. p53 signaling (Navarro et al,. 2021). The Kinetochore Metaphase Signaling pathway is an important checkpoint in the middle of mitosis during which the cell ensures that it is ready to divide, but as it is inhibited in HepG2 and MCF7 samples this pathway doesn't occur due to the reaction with doxorubicin in the cell process. The gene set enrichment analyses were remarkably similar across the concentrations for each cell line dosed with doxorubicin, which indicates a high degree of concordance in the transcriptional alterations that they induce.

The upstream regulator analysis (URA) tool is a novel function in IPA which can, by analysing linkage to DEGs through coordinated expression, identify potential upstream regulators including transcription factors (TFs) and any gene or small molecule that has been observed experimentally to affect gene expression.

A B C

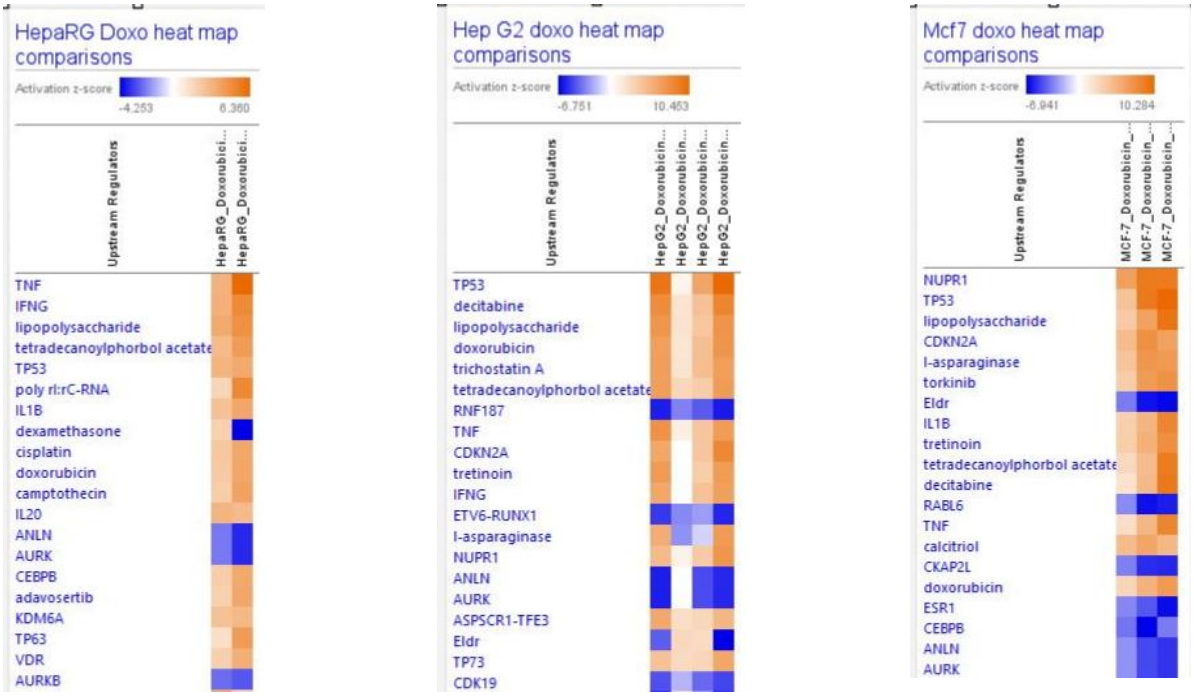


Figure 30: Comparison Heatmaps of the top 20 Upstream regulators comparing doxorubicin for each concentration for each chemical analysed – (A) HepaRG doxorubicin (first column = 0.2µM; second column 1µM); (B) Hep

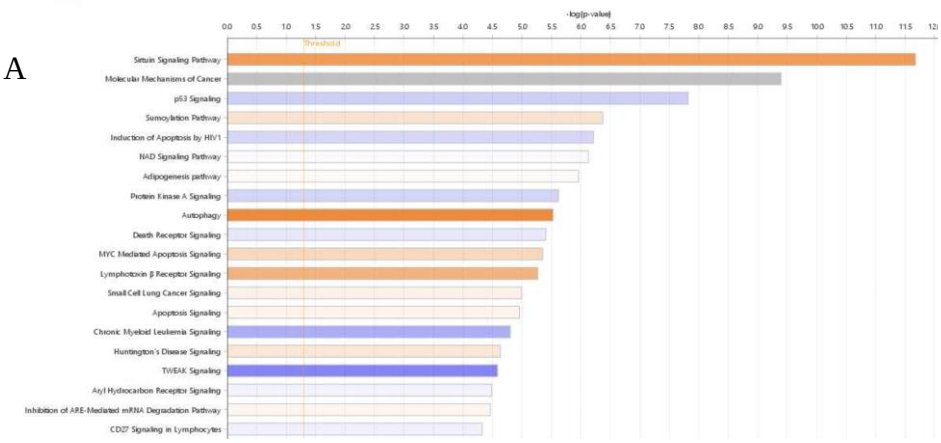
G2 doxorubicin (first column = 1µM; second column = 0.00032µM; third column = 0.04µM; fourth column = 0.2µM); (C) MCF7 doxorubicin (first column = 0.04µM; second column 0.2µM; third column = 1µM)

Interestingly, doxorubicin is actually one of the reference chemicals used for the upstream analysis in the IPA. In all the heatmaps illustrated in Figure 30, doxorubicin was indeed identifed as an activated upstream regulator which gives the results a weight of evidence that this chemical has an activation effect on all the cell lines. TP53 upstream regulator appears in all the cell lines, this might be because doxorubicin is an anti-cancer drug or simply because these are cancer cell lines. The TP53 gene encodes the p53 protein, which has been recognised as a cell cycle and apoptosis regulator. TP53 is the most frequently mutated gene in human cancers rendering the gene inactive and resulting in chemo-resistance to chemotherapies that both halt cell cycle progression and trigger apoptosis via the p53 pathways (McSweeney et al, 2019).

4.1.2 Niacinamide

Canonical pathways for each cell line and concentration for niacinamide are illustrated in Figure 31. The orange bars showed a positive z-score which means the pathways was activated and gave upregulated responses while the blue bars showed negative z-scores which means the pathways are inhibited and gave down regulated responses. The grey bars detailed no activity patterns were available and the white bars showed a z-score of 0 therefore no up or down regulation of pathways.

A



B

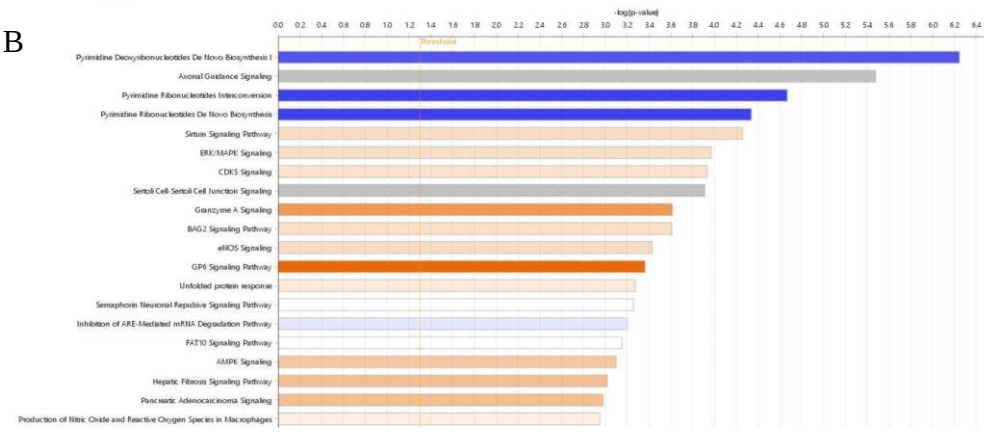


Figure 31: Top 20 canonical pathways identifed using IPA for niacinamide for each concentration analysed – (A) MCF7 Niac 60000µM; (B) MCF7 Niac 12000µM

For MCF-7 cells, the canonical pathways illustrate most pathways are activated rather than inhibited meaning they are upstream regulated pathways. For the highest concentration 60000µM, p53 signalling, induction of apoptosis by HIV1, protein kinase A signalling and the death receptor signalling were identified to be inhibited whereas sirtuin signalling pathway, sumoylation pathway and lymphotoxin B receptor were identified to be activated the most for the MCF-7 cell line. The induction of apoptosis by HIV1 relates to apoptosis of the cells resulting in cell death, P53 Signalling relates to DNA damage repair and apoptosis, therefore cell death. Sumoylation pathway affects normal cells in various ways, and usually plays a negative role in regulating transcription factor activity by changing the interaction with DNA and chromatin to repress gene expression, this pathway has been activated and will affect DNA damage of the cells. Most of the inhibited pathways relate to DNA damage, apoptosis, cell death, while the activated pathways relate to immune response pathways as the cells try to recover from chemical treatment. For the 12000µM concentration, different pathways were identified as activated or inhibited including pyrimidine deoxyribonucleotide de novo biosynthesis as well as other DNA/RNA related pathways. Even though niacinamide isn't a potent or highly toxic chemical compared to doxorubicin, the treatment of cells with any chemical can cause the slowing or stopping of cell growth during the cell cycle resulting in changes to the cells.

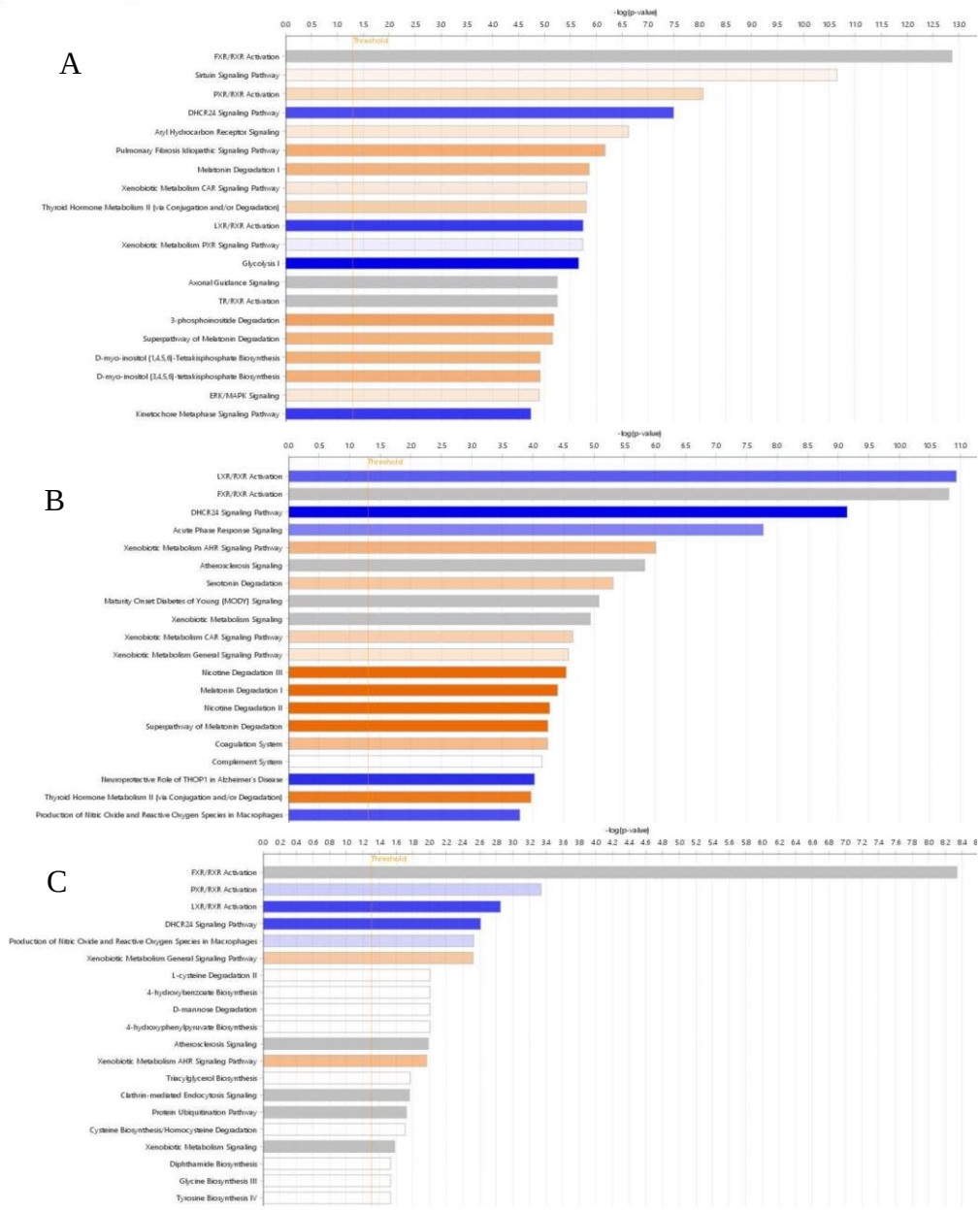


Figure 32: Top 20 canonical pathways identifed using IPA for niacinamide for each concentration analysed – (A) HepG2 Niac 60000µM; (B) HepG2 Niac 12000µM; (C) HepG2 Niac 2400µM

For HepG2 cells, the canonical pathways illustrate the majority of pathways are a mixture of inhibited and activated meaning some pathways are downstream regulated pathways and some are upstream regulated pathways (Figure 32). For the highest concentration 60000µM, PXR/RXR activation, melatonin degradation I and pulmonary fibrosis idiopathic signalling pathway were identified to be activated, the LXR/RXR Signaling pathway, DHCR24 Signalling pathway and the kinetochore metaphase signalling pathway were identified to be inhibited the most for the HepG2 cell line. LXR/RXR relates to the regulation of lipid metabolism and inflammation, while DHCR24 signalling

pathway is involved in cholesterol biosynthesis and kinetochore metaphase signalling pathway is an important checkpoint in the middle of mitosis during which the cells divide. As all these pathways are inhibited, it means they are not able to occur resulting in issues with metabolism and the cell cycle. The majority of the activated pathways relate to DNA damage, apoptosis, cell death, while the activated pathways i.e. PXP/RXR signaling plays an important role in drug metabolism and excretion, it can also regulate several endogenous transport, lipid metabolism and cholesterol homeostasis.

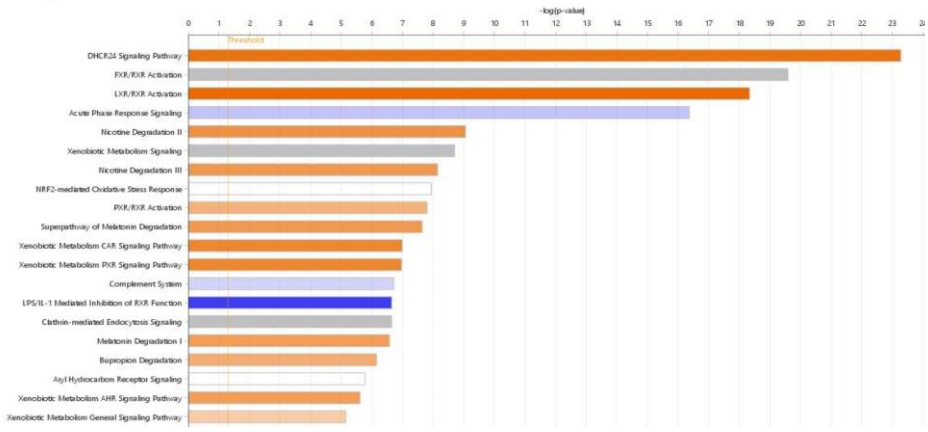


Figure 33: Top 20 canonical pathways identifed using IPA for niacinamide for HepaRG Niac 8000µM;

For HepaRG cells, the canonical pathways illustrate most pathways are activated meaning these pathways are upstream regulated pathways (Figure 33). For the one concentration, LXR/RXR activation, DHCR24 signaling pathway and FXR/RXR activation were identified to be activated the most for the HepaRG cell line. LPS/IL-1 mediated inhibition of RXR function was identified as inhibited which relates to lipid metabolism. Most of the activated pathways relate to metabolism processes in the cells.

Comparison heatmaps for niacinamide at the various concentrations and cell lines are illustrated in Figures 34 and 35, these heatmaps show whether the various cell lines have been activated (orange) or inhibited (blue) by the treatment of niacinamide using the z-score.

A

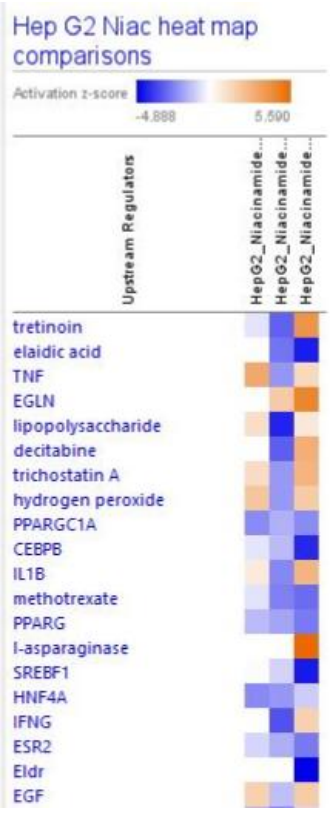
B

Figure 34: Comparison Heatmaps of the top 20 canonical pathways comparing niacinamide for each concentration analysed – (A) Hep G2 niacinamide (first colµMn = 2400µM; second column = 12000µM; third column = 60000µM); (B) MCF7 niacinamide (first column = 12000µM; second column = 60000µM

This approach allowed us to compare chemicals across all concentrations tested and revealed that the, HepG2 and MCF-7 cell lines doses with niacinamide perturbed gene sets associated with predicted upstream regulators and canonical pathways. For HepG2, the top 2 pathways showed inhibited pathways with LXR/RXR activation and DHCR24 Signaling Pathway

showing inhibition. The darker the colour the higher the z-score which means for Hep G2 1µM sample gives more pathway activation/inhibition compared to 0.2µM.
MCF-7 samples show a mixture of activation and inhibition pathways, with the higher concentrations having a higher z-score which is to be expected, with Pulmonary Fibrosis Idiopathic pathway and GP6 signalling pathway both being activated and ATM signalling, pyrimidine deoxyribonucleotide de novo biosynthesis and pyrimidine ribonucleotide interconversion being inhibited. Some of these pathways are very similar to what was seen in the BMDExpress results with HepaRG relating to lipid metabolism. HepG2 and MCF7 samples, the Kinetochore Metaphase Signaling pathway is an important checkpoint in the middle of mitosis during which the cell ensures that it is ready to divide, but as it is inhibited this pathway doesn't occur due to the reaction with niacinamide in the cell process. The gene set enrichment analyses were remarkably similar across the concentrations for each cell line dosed with niacinamide, indication activation and inhibitions of metabolism processes and DNA/RNA processes in the cells.

A



B

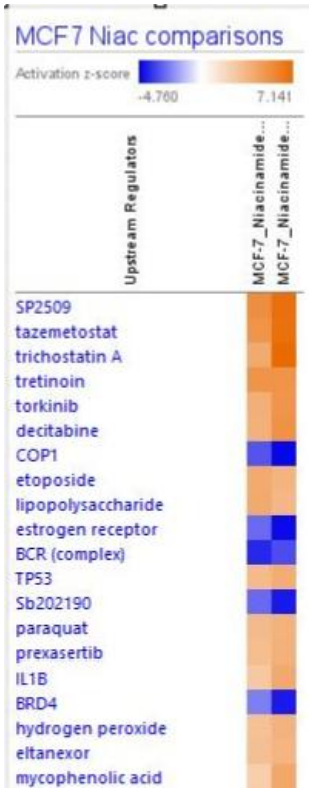


Figure 35: Comparison Heatmaps of the top 20 Upstream regulators comparing niacinamide for each concentration analysed – (A) Hep G2 niacinamide (first column = 2400µM; second column = 12000µM; third column = 60000µM); (B) MCF7 niacinamide (first column = 12000µM; second column = 60000µM)

For doxorubicin most upstream regulators related to cell cycle death and apoptosis, while niacinamide being a less toxic chemical showed less significant cell death related pathways or regulators, instead relating to immune and inflammatory responses. IPA has been a valuable tool for determination of relevant pathways and demonstrating mechanism of action of both chemicals.

CHAPTER 5

5. DISCUSSION

For NGRA, there is a continuing need to determine whether NAMs can be used to make robust safety decisions that are protective of human health. How this could be done has been demonstrated using a hypothetical case study in which coumarin was used as an ingredient in various consumer products (Baltazar et al., 2020) as well as within this thesis for doxorubicin and niacinamide using similar techniques. A key aspect of the coumarin case study was that a BER estimate (or margin of safety) obtained using NAMs (*in vitro* assays, PBK models, etc.) was combined with other toxicity data (eg, *in silico* predictions) to make safety decisions. It remains uncertain the degree the tools and approaches used in the coumarin case study could be utilised more generally to guarantee systemic safety for a wider range of chemical-exposure scenarios. Furthermore, in line with the principles of NGRA (Dent et al., 2018), a key purpose is to provide human health protection, rather than to be necessarily predictive of various adverse effects in animals. This is particularly important in the area of systemic toxicity, where a wide range of potential adverse outcomes must be covered, with many often not being fully characterised in relation to mechanism of action or adverse outcome pathways.

Several hypothetical frameworks describing a tiered approach for NGRA have been published over the past few years (Andersenet al., 2019; Berggren et al., 2017; Thomas et al., 2019), but NGRA examples of how to analyse, integrate and interpret all the data obtained from NAMs to inform a safety decision are still not common. Consequently, a milestone has been reached in the development and application of non-animal approaches to assess human safety, demonstrating for the first time that in chemico, *in silico*, and *in vitro* approaches can be combined to reach a consumer safety decision for systemic effects. However, this work demonstrates several key principles of NGRA (Dent et al., 2018a). The overall goal was to perform an exposure-led human safety assessment designed to prevent harm by applying robust and relevant methods in a hypothesis-driven way without animal testing. The philosophy behind this type of risk assessment aimed at preventing harm is based on the premise of “Protection not Prediction” (Kavlock et al., 2018; Thomas et al., 2019). Such a safety assessment approach is possible because it does not attempt to replicate the results of the animal tests historically used in safety assessment. Instead, the hypothesis underpinning this type of NGRA is that if there is no bioactivity observed at consumer-relevant concentrations, there can be no adverse health effects.

The analysis within this thesis continues the analysis conducted in the coumarin case study and further explores the bioactivity of doxorubicin and niacinamide.

The doses used for all the cell lines are reflective of what has been used in the literature (Middleton et al., 2022) by looking at the Cmax values associated with the various in use exposure scenarios of doxorubicin (Injac et al., 2008; Biganzoli et al., 2003; Rahman et al., 2007) and niacinamide (Cosmetic Ingredient Review Expert Panel., 2005; EFSA NDA Panel., 2014; EFSA Panel on Nutrition, Novel Foods and Food Allergens., 2022) in the treatment of patients as detailed in Table 1 of this thesis. The Cmax values for both chemicals are in the range of doses used for the experimental analysis conducted which covers the dose ranges seen for doxorubicin and niacinamide.

In particular, we have applied the PoD analysis to the dose response HTTr and carried out biological interpretation of the identified DEGs. By looking at doxorubicin and niacinamide, this has enabled the interpretation of these chemicals using developed methodologies for future analysis of potential chemicals for NGRA. The objectives outlined in the introduction section have been completed by using R Studio to align, quantify and conduct differential expression analysis using DESeq2 for all chemical concentrations and cell lines. Furthermore, BMDExpress has been successfully conducted to derive PoD's for doxorubicin and niacinamide as well as pathway investigation alongside IPA to determine a putative mechanism of action for the chemicals of interest.

HTTr is usually used in NGRA as a non-targeted approach for characterising biological responses potentially not covered by the other tools (Baltazar et al., 2020). In this thesis, we have applied different approaches for aggregating gene and pathway-level BMDs from HTTr based on previous work by Farmahin and colleagues (Baltazar et al., 2020). The results are combined in a weight of evidence to provide an overall understanding of transcriptional responses. Farmahin et al. (2017) indicated that summarizing BMD modelling in different ways has a comparatively small impact on the PoD, a similar result was seen in this study across all cell lines.

There is still significant discussion of what approaches to use for deriving a PoD at both gene and pathway level using BMDExpress, and therefore multiple PoDs were derived using several published methods (Farmahin et al., 2017) that have been shown to correlate closely to BMDL derived from equivalent treated samples using standard pathology studies. These included the mean of the 20 pathways with the lowest p value, or the 20 pathways with the lowest transcriptional BMDs and finally the lowest pathway BMDL that meets the significant enrichment criteria. At the gene level this included both the mean BMDL of 20 genes with largest fold change and the mean BMD of genes between 25th and 75th percentile. All the cell lines, HepaRG, HepG2 and MCF-7 cell lines met the recommendation (Farmahin et al., 2017) that at least 20 genes were detected to apply the gene-level tests. We have shown in this thesis that no matter whether we used gene level-derived PoDs or

pathway level-derived PoDs, the relative potency of doxorubicin was higher than that of niacinamide.

Niacinamide, an amide of vitamin B_3, is a hydrophilic endogenous substance. Given a sufficient bioavailability, niacinamide has antipruritic, antimicrobial, vasoactive, photo-protective, sebostatic and lightening effects depending on its concentration. Within a complex metabolic system niacinamide controls the NFκB-mediated transcription of signalling molecules by inhibiting the nuclear poly (ADP-ribose) polymerase-1 (PARP-1) (Wohlrab et al). Niacinamide is a well-tolerated and safe substance often used in cosmetics (Wohlrab et al).
Niacinamide plays a significant role in DNA repair, maintenance of genomic stability and cellular responses to injury including inflammation and apoptosis (cell death) (Boo et al,. 2021). Our PoD analysis shows that niacinamide is toxic to the cell lines in this thesis only at relatively high concentrations that are well above the use scenarios of niacinamide. At high concentrations niacinamide elicits a broad cellular stress response and affects primarily the metabolism processes and DNA/RNA related processes. All these related genes gave a PoD closer to the highest concentration for HepaRG cells and closer to the second highest concentration for HepG2 and MCF-7 cells, this means niacinamide affects the cells at the higher concentrations depicting this chemical is less toxic to cells compared to doxorubicin. Further work is required to understand the advantages of this analysis for human health protection due to the added uncertainty in modelling the compound's concentration *in vitro* following repeat dosing. For all the top pathways for HepaRG cells relates to metabolism processes i.e. Regulation of lipid metabolism by PPARalpha and Metabolism. The top pathways for MCF-7 cells relates to DNA/RNA related processes i.e. cleavage of the damaged pyrimidine and DNA methylation which could be due to the nature of the cell line itself as it is a cancer cell line (Comşa et al, 2015). For HepG2 cells, the top pathways identifed related to more generic cellular processes rather than any specific processes.

For all the top 20 pathways for HepaRG cells, it was identified that these pathways also related to predominately metabolism processes i.e. metabolism of lipids, metabolism; while the top 20 pathways for MCF-7 cell identifed as predominately DNA/RNA related processes i.e. viral mRNA Translation, RNA Polymerise II transcription.
Recent advances have provided physiological mechanisms of action of niacinamide on lipid metabolism and atherosclerosis (Kamanna et al) as related to the top pathway results in BMDExpress for HepaRG cells. Inflammatory benefits using niacinamide have been documented in the literature (Fivenson et al., Gehring et al) which relates to the top pathway for HepG2 cells. DNA related processes including the top pathway for MCF-7 cells – condensation of phosphate chromosomes were identified.

Doxorubicin is an antibiotic derived from the *Streptomyces peucetius* bacterium. It has had wide use as a chemotherapeutic agent since the 1960s. Doxorubicin is part of the anthracycline group of chemotherapeutic agents; other anthracyclines include daunorubicin, idarubicin, and epirubicin. Commonly, doxorubicin is an agent used in the treatment of solid tumors in adult and pediatric patients. Doxorubicin may be used to treat soft tissue and bone sarcomas and cancers of the breast, ovary, bladder, and thyroid. It is also used to treat acute lymphoblastic leukemia, acute myeloblastic leukemia, Hodgkin lymphoma, and small cell lung cancer (Yu et al,. 2019). The primary mechanism of action of doxorubicin involves the drug's ability to intercalate within DNA base pairs, causing breakage of DNA strands and inhibition of both DNA and RNA synthesis. Doxorubicin inhibits the enzyme topoisomerase II, causing DNA damage and induction of apoptosis (Sritharan et al,. 2021). Doxorubicin induces DNA strand breaks and triggers the activation of p53 to initiate DNA damage responses. Activation of p53 induces increased transcription of genes involved in the extrinsic and intrinsic apoptotic pathway as illustrated in this thesis.

Both PoD and functional analyses carried out in this thesis indicate that doxorubicin is a strongly toxic chemical. It is found that doxorubicin elicits a broad cellular stress response in HepG2 cells, and affects primarily the metabolism processes for HepaRG cells, and DNA/RNA related processes for MCF-7 cells. All these related genes gave a PoD closer to the second highest concentration for HepaRG cells and closer to the lower concentrations for HepG2 and MCF-7 cells, meaning doxorubicin affects the cells at the lower concentrations depicting this chemical is extremely toxic to cells compared to niacinamide.

The top pathway for HepaRG cells relates to cell death processes which could be due to the potency of doxorubicin on cells i.e. programmed cell death and apoptosis (McSweeney et all, 2019). The top pathway for MCF-7 cell relates to cell cycle and DNA replication in particular i.e. unwinding of DNA, DNA strand elongation, DNA methylation and this could be due to the nature of the cell line itself as it is a cancer cell line (Comşa et al., 2015). Similarly, for HepG2 cells, the top pathway refers to the cell cycle and in particular to the cell cycle-regulated protein kinase Nek2A (Faragher et al., 2003).

Furthermore, we have identified biological processes related to metabolism and immune response as more frequent among the top 20 pathways for HepaRG cells e.g. interferon signalling, metabolism of lipids, cytokine signaling in immune system (see Table 4). For MCF-7 and HepG2 cells we identified DNA/RNA related processes as the most common among the top 20 pathways e.g. metabolism of RNA, GTP hydrolysis and joining of the 60s ribosomal subunit and diseases of DNA repair. This difference between HepaRGs and MCF-7/HepG2s is not surprising, considering that HepaRG cells are more metabolically competent in comparison to MCF-7 and HepG2 cells (Duivenvoorde et al., 2021).

For HepG2 cell lines stabilization of endogenous p53 by doxorubicin, decreases NEK2 expression (Fang et al) and treatment with doxorubicin reduces HepG2 cell viability through initiating cell apoptosis and strong G2/M phase cell cycle arrest (Fan C et al). For HepaRG cells a similar top pathway was identifed relating to P53 and cell death as TP53 regulates Transcription of cell death genes with TP53 being the gene encoding p53 (McSweeney et al) and for MCF-7 cells modulation of G2/M phase arrest induced by doxorubicin is depend on the characteristic of breast cancer cells especially the p53 status (Junedi et al), which is supported by the pathway results generated from BMDExpress and IPA. Literature evidence states that the functionally of the different cell lines dosed with doxorubicin and niacinamide relate to the results produced in this thesis.

The significant up- and down-regulated gene lists for the two compounds also recovered enriched ontology pathway elements that were associated with aspects of the toxic response and MOA seen *in vivo* with these compounds. However, pathway elements associated with cell-cycle processes, mitotic processes, DNA damage, and some indirect indications of cell stress response and apoptosis signaling dominated.

Full interpretation using IPA of altered canonical pathways is beyond the scope of this study but serves to group the cell lines dosed with doxorubicin and niacinamide based on possible mode of action. In summary, each cell line dosed with doxorubicin and niacinamide perturbed similar canonical pathways and increased the z-score based on concentration from high concentrations illustrating higher z-scores for activation (orange) or lower z-scores for inhibition (blue) (Figures 26-28 and Figures 31-33). As with BMDExpress, various pathways were identified mainly relating to DNA/RNA process, metabolism, DNA damage for both cell lines and for doxorubicin more dominant cell death (apoptosis) related pathways. Upstream regulator analysis provides additional support to the weight of evidence and refers to any molecule that can affect the expression, transcription or phosphorylation of another molecule. For doxorubicin most upstream regulators related to cell cycle death and apoptosis, while niacinamide being a less toxic chemical showed less significant cell death related pathways or regulators, instead relating to immune and inflammatory responses (Figures 30 and 35).

Both BMDExpress and IPA are very different software packages with BMDExpress being used for gene and pathway level analysis, deriving PoDs and pathway related analysis, while IPA is used solely for pathway interpretation and determining MoA. The combination of using BMDExpress and IPA has demonstrated more robust data can be generated for gene and pathways interpretation. More investigations are required using both software to depict correlations between them as well as considering individual results. This thesis has explained how BMDExpress and IPA are conducted separately and the difference/correlations are not the aim of this thesis.

Finally, this thesis shows that the majority of biological processes that were upregulated or downregulated by the chemicals are very much related to the underlying biology of the cancer cell lines used here. This leaves open the question of choice when selecting an immortalized cancer cell line and what constitutes a meaningful surrogate for *in vivo* response of compounds whose cellular affects related to potential adverse outcomes are known to target specific organs or where metabolic competency of the *in vitro* system is insufficient to serve as a surrogate for *in vivo* response. Further efforts will be needed to better inform choice of cell lines, or design assays with greater physiological relevance when designing high throughput chemical screening assays.

In conclusion, this thesis has demonstrated that NAMs can provide robust insights to address a gap of knowledge within Unilever by conducting biological interpretation of exposure to chemicals at different concentrations and the general use of *in vitro* methods for non-animal risk assessments. The continuous development and application of NAMs in a decision-making context will participate in fulfilling the drive to assure safety of novel ingredients without the need for any animal testing, but confidence in NAMs will only occur with extensive learning by conducting and sharing more case studies.

6. APPENDICES

Appendix 1

Supplementary Script 1 - DESeq2 script using R Studio (Github reference link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/DESeq2%20MRes%20script.R).

```
#R script - DESeq2 MRes script - Sample data sets for HepG2, HepaRG and MCF7 cell line dosed with Doxorubicin and Niacinamide

#Datasets analysed using DESeq2.

#load libraries relevant to this analysis

library(DESeq2)

library(tidyverse)

library(dplyr)

library(EnhancedVolcano)

library(org.Hs.eg.db)

library(ggplot2)

library(plyr)

#read the csv files containing the sample data for each cell line and chemical

counts <- read.csv("HepG2_Doxorubicin_HCl_counts.csv", row.name=1)

head(counts)

counts2 <- read.csv("HepG2_Niacinamide_counts.csv", row.name=1)

head(counts2)

counts3 <- read.csv("HepaRG_Niacinamide_counts.csv", row.name=1)

head(counts3)

counts4 <- read.csv("HepaRG_Doxorubicin_HCl_counts.csv", row.name=1)

head(counts4)

counts5 <- read.csv("MCF-7_Niacinamide_counts.csv", row.name=1)

head(counts5)



counts6 <- read.csv("MCF-7_Doxorubicin_HCl_counts.csv", row.name=1)
```

```
head(counts6)

colSums(counts)

#create list of all files to input into a loop

files <- list.files(path = ".", pattern = "_counts.csv", full.names = T)

files

#loop created to analyse all the cell lines and chemicals individually

for(i in (files)){

 print(i)

 counts_file <- i

 meta_files <- gsub("_counts.csv", "_metadata.csv", i) # create metadata csv files from the counts data files

 counts <- read.csv(counts_file, row.names = 1)

 SampleTable <- read.csv(meta_files, header = T)

 SampleTable$CONCENTRATION <- as.factor(SampleTable$CONCENTRATION)

 dds <- DESeqDataSetFromMatrix(countData = counts,

                               colData = SampleTable,

                               design = ~ VESSEL_ID + CONCENTRATION)

 dds$condition <- relevel(dds$CONCENTRATION, ref = "0")

 dds <- estimateSizeFactors(dds)

 sizeFactors(dds)

  dds <- DESeq(dds)

 res <- results(dds)

 comparisons <- resultsNames(dds)

 normalisedCounts <- counts(dds, normalized = TRUE)

  for (j in comparisons[6:12]) {

   res <- results(dds, name = j)

   padj.cutoff <- 0.05

   lfc.cutoff <- 0.58

   res_table <- res %>%
```

```
    data.frame() %>%
    rownames_to_column(var="gene") %>%
    as_tibble() #convert the results table into a tibble
  sigOE <- res_table %>%
    filter(padj < padj.cutoff & abs(log2FoldChange) > lfc.cutoff)
  #write.table(sigOE, file = paste0(j, ".txt", i)) #need to add cell line & chemical?
  write.table(sigOE, file = paste0(gsub("_counts.csv", "_DESeq", i), "_", j, ".txt"), col.names = T, row.names = T, quote = F, sep = "\t")
 }
res_norm <- lfcShrink(dds=dds, coef=2, type="normal")
save(res_norm, file = "res_norm.RData")
png(gsub("_counts.csv", "_MA.png", i))
ma <- plotMA(res_norm)
print(ma)
dev.off()
 png(gsub("_counts.csv", "_MA2.png", i))
ma2 <- plotMA(res_norm, alpha = 0.05)
print(ma)
dev.off()
 plot_EnhancedVolcano <- EnhancedVolcano(res_norm,
                         lab = rownames(res),
                         x = "log2FoldChange",
                         y = "pvalue")
ggsave(plot_EnhancedVolcano, filename = gsub("_counts.csv", "_EV.png", i))
 normalised_counts <- normalisedCounts %>%
data.frame() %>%
 rownames_to_column(var="gene") %>%
 as_tibble()
 top20_sigOE_genes <- res_table %>%
```

```
  arrange(padj) %>%
  pull(gene) %>%
  head(n=20)
  top20_sigOE_norm <- normalised_counts %>%
  filter(gene %in% top20_sigOE_genes)
  pivot_top20_sigOE <- top20_sigOE_norm %>%
  pivot_longer(!gene, names_to = "samplename", values_to = "normalised")
 pivot_top20_sigOE_join <- pivot_top20_sigOE %>%
  left_join(dplyr::select(SampleTable, X, CONCENTRATION), by = c("samplename" = "X"))
 ggplot <- ggplot(pivot_top20_sigOE_join,aes(x = gene, y = normalised, color = CONCENTRATION)) +
  geom_point() +
  scale_y_log10() +
  xlab("genes") +
  ylab("normalised_counts") +
  ggtitle("Top 20 Significant DE Genes") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5))
  ggsave(ggplot, filename = gsub("_counts.csv", "_ggplot.png", i))
 rld <- vst(dds, blind=TRUE)
 pca <- plotPCA(rld, intgroup="CONCENTRATION")
 ggsave(pca, filename = gsub("_counts.csv", "_pca.png", i))
 vst_write <- assay(vst(dds, blind=FALSE))
 write.csv(as.data.frame(vst_write), file=gsub("_counts.csv", "_results.csv", i))
 vst_write_copy <- vst_write # COPY SO WE DONT RUIN DATA
 colnames(vst_write_copy) <- NULL # Getting rid of colnames so we can append to data
 cols <- colnames(vst_write) # pulling sample ids
 vst_dose <- SampleTable[SampleTable$X %in% colnames(vst_write),"CONCENTRATION" ] # ordering
the metadata in the order our sampleid in the normalised data then pulling concentration out
```

```
  vst_dose_dt <- data.frame(matrix(ncol=175, nrow=1)) # making the dose dataframe ready to rbind

  vst_dose_dt[1, ] <- vst_dose # setting the first row to doses

  bmd_input <- rbind(vst_dose_dt, data.frame(vst_write_copy)) # combining dose row with gene normalised data

  rownames(bmd_input)[1] <- "Dose" # getting dose in the rownames

  toprow <- data.frame(matrix(ncol = 175, nrow=1))  # new data_frame to get the 'SampleID' bit in

  toprow[1, ] <- cols # set first row to sample ID. This will be our actual colnames name but we dont set it here

  rownames(toprow) <- "SampleID" # corner value

  final_bmd_input <- rbind(toprow, bmd_input)

  write.table(final_bmd_input, file = gsub("_counts.csv", "_BMDInput.txt", i), col.names = F, row.names =T, quote = F, sep = "\t") # write withour colnames as these are current X1. X2

  print(paste(i, "finished"))

}
```

Supplementary Script 2: Generation of Upset Plots using R Studio (Github reference link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/IPA_gene_intersect%20upset.Rlibrary(glue)

```
library(purrr)

library(dplyr)

library(UpSetR)

upset_gene_intersect <- function(cell_line,chemical, conc1, conc2, conc3, conc4, conc5, conc6, conc7){

  # check if there are 6 or 7 concentrations available and make list of all the concentrations

   tsv_files <- c(file.path(
glue::glue("{cell_line}_{chemical}_DESeq_CONCENTRATION_{conc1}_vs_0.txt")),

            file.path( glue::glue("{cell_line}_{chemical}_DESeq_CONCENTRATION_{conc2}_vs_0.txt")),

            file.path( glue::glue("{cell_line}_{chemical}_DESeq_CONCENTRATION_{conc3}_vs_0.txt")),

            file.path( glue::glue("{cell_line}_{chemical}_DESeq_CONCENTRATION_{conc4}_vs_0.txt")),

            file.path( glue::glue("{cell_line}_{chemical}_DESeq_CONCENTRATION_{conc5}_vs_0.txt")),

            file.path( glue::glue("{cell_line}_{chemical}_DESeq_CONCENTRATION_{conc6}_vs_0.txt")),

            file.path( glue::glue("{cell_line}_{chemical}_DESeq_CONCENTRATION_{conc7}_vs_0.txt")))
```

```
  conc_list <- c(conc1,conc2, conc3, conc4, conc5, conc6, conc7)
 # make a list of dataframes containing the data from the concentration with 1 or more degs
 dfs <- list()
 for (i in 1:length(tsv_files)){
  num_rows <- nrow(read.csv(tsv_files[i], sep = "\t"))
  conc <- toString(conc_list[i])
  if (num_rows > 0) {
  dfs[[conc]] <- read.csv(tsv_files[i], sep = "\t")
  }
 }
 # check how many of the concentrations have degs, based on the number of data frames in the list
 if (length(dfs) == 7){
  #combine data frames, given all the concentrations have 1 or more degs
  combined <- purrr::reduce(list(data.frame(gene = dfs[[1]]$gene, conc1 = 1),
                                 data.frame(gene = dfs[[2]]$gene, conc2 = 1),
                                 data.frame(gene = dfs[[3]]$gene, conc3 = 1),
                                 data.frame(gene = dfs[[4]]$gene, conc4 = 1),
                                 data.frame(gene = dfs[[5]]$gene, conc5 = 1),
                                 data.frame(gene = dfs[[6]]$gene, conc6 = 1),
                                 data.frame(gene = dfs[[7]]$gene, conc7 = 1)), full_join)
  combined[is.na(combined)] <- 0
  # give columns more meaningful names
  names(combined) <- c("genes",
                       names(dfs)[1],
                       names(dfs)[2],
                       names(dfs)[3],
                       names(dfs)[4],
                       names(dfs)[5],
```

```
                names(dfs)[6],
                names(dfs)[7])
  } else if (length(dfs) == 6){
    #combine data frames, given 6 of the concentrations have 1 or more degs
    combined <- purrr::reduce(list(data.frame(gene = dfs[[1]]$gene, conc1 = 1),
                                   data.frame(gene = dfs[[2]]$gene, conc2 = 1),
                                   data.frame(gene = dfs[[3]]$gene, conc3 = 1),
                                   data.frame(gene = dfs[[4]]$gene, conc4 = 1),
                                   data.frame(gene = dfs[[5]]$gene, conc5 = 1),
                                   data.frame(gene = dfs[[6]]$gene, conc6 = 1)), full_join)
    combined[is.na(combined)] <- 0
    # give columns more meaningful names
    names(combined) <- c("genes",
                names(dfs)[1],
                names(dfs)[2],
                names(dfs)[3],
                names(dfs)[4],
                names(dfs)[5],
                names(dfs)[6])
  } else if (length(dfs) == 5){
    #combine data frames, given 5 of the concentrations have 1 or more degs
    combined <- purrr::reduce(list(data.frame(gene = dfs[[1]]$Probe, conc1 = 1),
                                   data.frame(gene = dfs[[2]]$Probe, conc2 = 1),
                                   data.frame(gene = dfs[[3]]$Probe, conc3 = 1),
                                   data.frame(gene = dfs[[4]]$Probe, conc4 = 1),
                                   data.frame(gene = dfs[[5]]$Probe, conc5 = 1)), full_join)
    combined[is.na(combined)] <- 0
```

```
  # give columns more meaningful names
  names(combined) <- c("genes",
                       names(dfs)[1],
                       names(dfs)[2],
                       names(dfs)[3],
                       names(dfs)[4],
                       names(dfs)[5])
 } else if (length(dfs) == 4){
  #combine data frames, given 4 of the concentrations have 1 or more degs
  combined <- purrr::reduce(list(data.frame(gene = dfs[[1]]$gene, conc1 = 1),
                                 data.frame(gene = dfs[[2]]$gene, conc2 = 1),
                                 data.frame(gene = dfs[[3]]$gene, conc3 = 1),
                                 data.frame(gene = dfs[[4]]$gene, conc4 = 1)), full_join)
  combined[is.na(combined)] <- 0
  # give columns more meaningful names
  names(combined) <- c("genes",
                       names(dfs)[1],
                       names(dfs)[2],
                       names(dfs)[3],
                       names(dfs)[4])
 } else if (length(dfs) == 3){
  #combine data frames, given 3 of the concentrations have 1 or more degs
  combined <- purrr::reduce(list(data.frame(gene = dfs[[1]]$gene, conc1 = 1),
                                 data.frame(gene = dfs[[2]]$gene, conc2 = 1),
                                 data.frame(gene = dfs[[3]]$gene, conc3 = 1)), full_join)
  combined[is.na(combined)] <- 0
  # give columns more meaningful names
  names(combined) <- c("genes",
```

```
                names(dfs)[1],
                names(dfs)[2],
                names(dfs)[3])
  } else if (length(dfs) == 2){
    #combine data frames, given 2 of the concentrations have 1 or more degs
    combined <- purrr::reduce(list(data.frame(gene = dfs[[1]]$gene, conc1 = 1),
                                   data.frame(gene = dfs[[2]]$gene, conc2 = 1)), full_join)
    combined[is.na(combined)] <- 0
    # give columns more meaningful names
    names(combined) <- c("genes",
                names(dfs)[1],
                names(dfs)[2])
  } else {
    # the chemical either has only one or no concentration with degs
    # an upset plot is not needed
    print("There are not enough data sets containing data!")
  }
  # check the chemical has at least 2 or more concentrations with degs
  if (length(dfs) > 1) {
    # get  vector names of concentrations being compared from the combined data frame column
headers
    # needed to set concentrations displayed order in the upset plot
    upset_sets <- names(combined)[-1]
    # create upset plot
    plot <- UpSetR::upset(combined, nsets = length(names(combined)),keep.order = T, sets = upset_sets)
    # export upset plot
   png(file.path(glue::glue("{cell_line}_{chemical}_shared_genes_upset.png")), width = 1300, height =
800, res = 100)
   print(plot)
```

```
  dev.off()

 }

}

upset_gene_intersect("HepaRG", "Niacinamide", 8000, 1600, 320, 64, 12.8, 2.56, 0.512)

upset_gene_intersect("HepG2", "Niacinamide", 60000, 12000, 2400, 480, 96, 19.2, 3.84)

upset_gene_intersect("MCF-7", "Niacinamide", 60000, 12000, 2400, 480, 96, 19.2, 3.84)

upset_gene_intersect("HepaRG", "Doxorubicin_HCl", 1, 0.2, 0.04, 0.008, 0.0016, 0.00032, "6.4e.05")

upset_gene_intersect("HepG2", "Doxorubicin_HCl", 1, 0.2, 0.04, 0.008, 0.0016, 0.00032, "6.4e.05")

upset_gene_intersect("MCF-7", "Doxorubicin_HCl", 1, 0.2, 0.04, 0.008, 0.0016, 0.00032, "6.4e.05")
```

Supplementary Script 3 – IPA gene name change using Visual Studio Code
(Github reference link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/gene%20name%20change.py)

```python
import json
import pandas as pd

# add your file path to the json file
gene_file = open('C:/Users/liz.tulum/OneDrive - Unilever/MRes course/190620_Human_Whole_Transcriptome_2.0_Manifest_probe_to_gene (3).json')

# load in json file of gene names as data frame
gene_names = json.load(gene_file)

# the concentrations of the file your looking add
niac_conc= ["0.512", "2.56", "12.8", "64", "320", "1600", "8000"]
niac_conc2= ["3.84", "19.2", "96", "480", "2400", "12000", "60000"]
doxo_conc= ["6.4e.05", "0.00032", "0.0016", "0.008", "0.04", "0.2", "1"]

# for each concentration it reads in the file as a data frame and replaces the probe name with the gene name
# HepaRG
for conc in niac_conc:
    df = pd.read_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes course/HepaRG_Niacinamide_DESeq_CONCENTRATION_{conc}_vs_0.txt', sep='\t')
    print(df)
    # change 'Unnamed:0' to 'gene' or whatever the first column of the file is caaled
    df['gene'].replace(gene_names,inplace = True)

    # export df with the gene names as a tsv
```

```
    df.to_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/HepaRG_Niacinamide_DESeq_CONCENTRATION_{conc}_vs_0.csv', index=False)

for conc in doxo_conc:
    df = pd.read_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/HepaRG_Doxorubicin_HCl_DESeq_CONCENTRATION_{conc}_vs_0.txt', sep='\t')
    # change 'Unnamed:0' to 'gene' or whatever the first column of the file
is caaled
    df['gene'].replace(gene_names,inplace = True)
    # export df with the gene names as a tsv
    df.to_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/HepaRG_Doxorubicin_HCl_DESeq_CONCENTRATION_{conc}_vs_0.csv',
index=False)

#HepG2
for conc in niac_conc2:
    df = pd.read_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/HepG2_Niacinamide_DESeq_CONCENTRATION_{conc}_vs_0.txt', sep='\t')
    # change 'Unnamed:0' to 'gene' or whatever the first column of the file
is caaled
    df['gene'].replace(gene_names,inplace = True)

    # export df with the gene names as a tsv
    df.to_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/HepG2_Niacinamide_DESeq_CONCENTRATION_{conc}_vs_0.csv', index=False)

for conc in doxo_conc:
    df = pd.read_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/HepG2_Doxorubicin_HCl_DESeq_CONCENTRATION_{conc}_vs_0.txt', sep='\t')
    # change 'Unnamed:0' to 'gene' or whatever the first column of the file
is caaled
    df['gene'].replace(gene_names,inplace = True)
    # export df with the gene names as a tsv
    df.to_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/HepG2_Doxorubicin_HCl_DESeq_CONCENTRATION_{conc}_vs_0.csv',
index=False)

#MCF7
for conc in niac_conc2:
    df = pd.read_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes
course/MCF-7_Niacinamide_DESeq_CONCENTRATION_{conc}_vs_0.txt', sep='\t')
    # change 'Unnamed:0' to 'gene' or whatever the first column of the file
is caaled
    df['gene'].replace(gene_names,inplace = True)

    # export df with the gene names as a tsv
    df.to_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes course/MCF-
7_Niacinamide_DESeq_CONCENTRATION_{conc}_vs_0.csv', index=False)
```

```
for conc in doxo_conc:
    df = pd.read_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes course/MCF-7_Doxorubicin_HCl_DESeq_CONCENTRATION_{conc}_vs_0.txt', sep='\t')
    # change 'Unnamed:0' to 'gene' or whatever the first column of the file is caaled
    df['gene'].replace(gene_names,inplace = True)
    # export df with the gene names as a tsv
    df.to_csv(f'C:/Users/liz.tulum/OneDrive - Unilever/MRes course/MCF-7_Doxorubicin_HCl_DESeq_CONCENTRATION_{conc}_vs_0.csv', index=False)
```

Supplementary Script 4 – Calculating PoD using R Studio (Github reference link: https://github.com/liztulum/MRes-thesis-scripts/blob/main/calculate_pods.R)

```
library(logger)

library(glue)

library(tibble)

#' Create output directory if it doesn't exist. Warn if existing directory is not empty

#' @param path directory path to create (string)

#' @importFrom logger log_debug log_warn

#'

#' @export

create_dir <- function(path) {

 if (!dir.exists(path)) {

  logger::log_debug("Creating output directory {dQuote(path)}")

  dir.create(path, recursive = T)

 } else {

  logger::log_debug("Output directory {dQuote(path)} is existing")

  if (length(dir(path = path, all.files = FALSE)) > 0) {

   logger::log_warn("Output directory {dQuote(path)} is not empty, files may be overwritten")

  }

 }

 return(path)
```

```
}


#' BMD and pathway data checks and error handling

#'

#'

#' @param data variable to check intended to be either bmd data or pathway data from BMDExpress2

#' @param function_name name of the function running this check function. This is used in error messages

#

#' @importFrom logger log_fatal

#' @importFrom glue glue

check_data <- function(data, function_name) {

  list(

    type = function() {

      if (!class(data) == "data.frame") {

        logger::log_fatal("data passed to function '{function_name}' is not a data.frame")

        stop(glue::glue("Please supply a data.frame object to function '{function_name}' with columns from BMDExpress2 results BMD export."))

      }

    },

    cols = function(keep_cols) {

      is_missing <- !(keep_cols %in% colnames(data))

      if (any(is_missing)) {

        missing_cols <- keep_cols[is_missing]

        logger::log_fatal("data passed to function '{function_name}' is missing column(s) {missing_cols}")

        stop(glue::glue("Please supply a data.frame object to function '{function_name}' with columns from BMDExpress2 results BMD export."))

      }

    },

    empty = function() {
```

```
    if (nrow(data) == 0) {
      e <- "data passed to function '{function_name}' has no data."
      logger::log_fatal(e)
      stop(e)
    }
  },
  analysis_length = function() {
    if (length(unique(data$Analysis)) != 1) {
      logger::log_fatal("data passed to function '{function_name}' has more than one analysis group.")
      stop("pathway_filtered passed to function '{function_name}' has more than one analysis group.
Please pass data with one unique value of column 'Analysis' only")
    }
  }
 )
}

#' Write filtered data set
#'
#' Write a data frame of data to csv with a table of filter parameters used writtten above data.
#'
#' @param filtered_data Dataframe of data to be written
#' @param filter_info Dataframe of filter parameters to be written above the filtered_data
#' @param output_file_name the file name (included '.csv') to be written
#' @param output_dir the directory where the file should be written.
#' @importFrom logger log_fatal log_info
#' @importFrom glue glue
write_filtered_data <- function(filtered_data, output_file_name, filter_info = NULL, output_dir = ".") {
 # filter parameters
 filter_by <- tibble::lst(filtered_data,
                          filter_info)
```

```
  for (f in names(filter_by)) {
    filter_value <- filter_by[[f]]
    if(!is.null(filter_value)) {
      if(!is.data.frame(filter_value)) {
        logger::log_fatal("{f} '{filter_value}' passed to function 'write_filtered_data' is not a dataframe")
        stop(glue::glue("{filter_value} passed to function 'write_filtered_data' is not a dataframe. Please supply a dataframe"))
      }
    }
  }
  create_dir(output_dir)
  # Write filter parameters table and the filtered data to a csv file by appending tables on top another (with blank dataframe for spacing)
  file_path <- file.path(output_dir, output_file_name)
  write.table(filter_info, file_path, col.names = T, sep = ", ", append = F, row.names = F)
  suppressWarnings(write.table(data.frame(), file_path, col.names = T, sep = ",", append = T))
  suppressWarnings(write.table(filtered_data, file_path, col.names = T, sep = ",", append = T, row.names = F))
  logger::log_info("Filtered data written to {file_path}.")
}

#' Filter Significant Gene BMDs
#'
#' Filtering for significant can include filtering out bmds which exceed the highest tested concentration, filtering based on the fit-p value and on the BMDU and BMDL ratio.
#' This function can be run on different analysis groups (cell line - chemical groups) but it must be noted that different analysis groups likely have different highest concentrations and therefore require a different highest_conc_filter value. If this this the case the function should be run on a subset of the input data for one analysis.
#' This function will write the returned data.frame of filtered gene level BMDs to csv file at {output_dir} with file name {output_prefix}_gene_bmd_filtered.csv
#' @param bmd_data Dataframe subset of BMDExpress gene bmd data. Dataframe should only include one analysis (cell line - chemical) group.
```

#' Must include the following columns as a minimum

#' - Analysis

#' - Probe ID

#' - Entrez Gene IDs

#' - Genes Symbols

#' - Best Model

#' - Best BMD

#' - Best BMDL

#' - Best BMDU

#' - Best fitPValue

#' - Best fitLogLikelihood

#' - Best AIC

#' - Best BMD/BMDL

#' - Best BMDU/BMDL

#' - Best BMDU/BMD

#' - Max Fold Change

#' - Max Fold Change Absolute Value

#' @param highest_conc_filter Numeric parameter indicating the value to filter 'Best BMD'. Data are filtered 'Best BMD' <= highest_conc_filer. To not use the filter set highest_conc_filter as NULL.

#' @param bmdl_bmdu_filter Numeric parameter indicating the value to filter 'Best BMDU/BMDL'. Data are filtered 'Best BMDU/BMDL' < bmdl_bmdu_filter. To not use the filter set bmdl_bmdu_filter as NULL. Default = 40

#' @param fitP_filter Numeric parameter indicating the value to filter on 'Best fitPValue'. Data are filtered 'Best fitPValue' > fitP_filter. To not use the filter set fitP_filter as NULL. Default = 0.1

#' @param gene_csv_output_prefix Character string to use as the prefix the the outputed csv of filtered data. Recommended is the analysis group from BMDExpress. If no csv is required, use NULL (default).

#' @param gene_csv_output_dir The directory path to where the output csv of filtered data should be written to. This directory should already exist.

#' @importFrom logger log_fatal log_info

#' @importFrom glue glue

#' @return dataframe of filtered gene BMD data

```
#' @export

filter_gene_bmds <- function(bmd_data, highest_conc_filter = NULL, bmdl_bmdu_filter = 40, fitP_filter = 0.1, gene_csv_output_prefix = NULL, gene_csv_output_dir = ".") {

  keep_cols <- c(

    "Analysis", # chosen as the most relavent data columns to save in csvs

    "Probe ID",

    "Entrez Gene IDs",

    "Genes Symbols",

    "Best Model",

    "Best BMD",

    "Best BMDL",

    "Best BMDU",

    "Best fitPValue",

    "Best fitLogLikelihood",

    "Best AIC",

    "Best BMD/BMDL",

    "Best BMDU/BMDL",

    "Best BMDU/BMD",

    "Max Fold Change",

    "Max Fold Change Absolute Value"

  )

  ## Parameter Error Handling

  # bmd_data

  check <- check_data(bmd_data, "filter_gene_bmds")

  check$type()

  check$cols(keep_cols)

  check$empty()

    # filter parameters

  filter_by <- tibble::lst(fitP_filter,
```

```
                    bmdl_bmdu_filter,
                    highest_conc_filter)
  for (f in names(filter_by)) {
  filter_value <- filter_by[[f]]
  if(!is.null(filter_value)) {
   if(!is.numeric(filter_value)) {
    logger::log_fatal("{f} '{filter_value}' passed to function 'filter_gene_bmds' is not numeric")
    stop(glue::glue("{filter_value} passed to function 'filter_gene_bmds' is not numeric. Please supply numeric value or NULL."))
   }
  }
 }
 FC_cols <- colnames(bmd_data)[grep("FC Dose Level", colnames(bmd_data))]
 keep_cols <- c(keep_cols, FC_cols)
 bmd_filtered <- bmd_data[bmd_data$`Best BMD` != "none" & bmd_data$`Best BMD` != "NaN", keep_cols] # get rid of nones and remove unneeded columns
 for (i in 1:ncol(bmd_filtered)) {
  if (!any(is.na(suppressWarnings(as.numeric(bmd_filtered[, i]))))) { # check whether columns can be converted to numeric, if so convert. (NAs appear when non numeric characters are attempted to be converted so we check for NAs to determine suitability)
   bmd_filtered[, i] <- as.numeric(bmd_filtered[, i])
  }
 }
 if (!is.null(highest_conc_filter)) {
  bmd_filtered <- bmd_filtered[bmd_filtered[, "Best BMD"] <= highest_conc_filter, ]
 } else {
  highest_conc_filter <- "None"
 }
 if (!is.null(bmdl_bmdu_filter)) {
  bmd_filtered <- bmd_filtered[bmd_filtered$`Best BMDU/BMDL` < bmdl_bmdu_filter, ]
```

```
  } else {
    bmdl_bmdu_filter <- "None"
  }
  if (!is.null(fitP_filter)) {
    bmd_filtered <- bmd_filtered[bmd_filtered$`Best fitPValue` > fitP_filter, ]
  } else {
    fitP_filter <- "None"
  }
  if (!is.null(gene_csv_output_prefix)) {
    # Get table of filter parameters used
    filters_used <- data.frame(c(
      paste0("Best BMD <= ", highest_conc_filter),
      paste0("Best BMDU/BMDL < ", bmdl_bmdu_filter),
      paste0("Best fitPValue > ", fitP_filter)
    ))
    colnames(filters_used) <- "These data are filtered using the following column filters:"
    write_filtered_data(filtered_data = bmd_filtered,
                        filter_info = filters_used,
                        output_file_name = paste0(gene_csv_output_prefix, "_gene_bmd_filtered.csv"),
                        output_dir = gene_csv_output_dir)
    }
  return(bmd_filtered)
}
#' Calculate gene level PoDs from BMD data
#'
#' Using data from the filter_gene_bmds function, PoDs are calculated via 2 method:
#' 1. Average of BMD/BMDLs of 20 genes with the largest fold change.
#' 2. Average of the 25th and 75th percentile of BMD/BMDLs
```

```
#' This function assumed all genes inputted to the function have significant BMD/BMDLs and that only one analysis group is passed (eg one cell line- chemical)

#' @param bmd_filtered output of the filter_gene_bmds. Data needs columns "Probe ID", "Max Fold Change Absolute Value" and "Best BMD" or "Best BMDL"

#' @param bmd_param "BMD" or "BMDL" indicating which value to base calculations around.

#' @param bmdExpress_input_dir directory path where BMD input files for BMDExpress are located

#' @importFrom logger log_info

#' @importFrom glue glue

#' @importFrom stats quantile

#'

#' @return dataframe of gene level PoDs for one analysis group

#' @export

calculate_gene_PoD <- function(bmd_filtered, bmd_param, bmdExpress_input_dir) {

 # Parameter Error Handling

 # bmd_param

 check_cols <- c("BMD", "BMDL")

 if (!toupper(bmd_param) %in% check_cols) {

  e <- glue::glue("bmd_param '{bmd_param}' passed to function 'filter_gene_bmds' is not {paste(check_cols, collapse = ' or ')}")

  logger::log_fatal(e)

  stop(e)

 }

 col_names <- c(

  "Analysis", # minimum columns needed for filtering

  "Best BMDU/BMDL",

  "Max Fold Change Absolute Value",

  paste0("Best ", bmd_param)

 )

 check <- check_data(bmd_filtered, "calculate_gene_PoD")

 check$type()
```

```
  check$cols(col_names)

  check$empty()

  check$analysis_length()

  bmd_filtered[, paste0("Best ", bmd_param)] <- as.numeric(bmd_filtered[, paste0("Best ",
bmd_param)])

  if(nrow(bmd_filtered) >= 20) {

    top20FC <- bmd_filtered[order(bmd_filtered$`Max Fold Change Absolute Value`, decreasing =
T)[1:20], paste0("Best ", bmd_param)] # get BMD/BMDL values for the genes with 20 highest FC

    PoD_top20FC <- mean(top20FC)

    analysis_name <- unique(bmd_filtered$Analysis)

  } else {

    PoD_top20FC <- mean(bmd_filtered[, paste0("Best ", bmd_param)])

    analysis_name <- paste0(unique(bmd_filtered$Analysis), " *(", nrow(bmd_filtered), " bmds)")

  }

  PoD_percentile_25 <- quantile(bmd_filtered[, paste0("Best ", bmd_param)], c(.25))

  PoD_percentile_75 <- quantile(bmd_filtered[, paste0("Best ", bmd_param)], c(.75))

  PoD_percentile <- mean(bmd_filtered[bmd_filtered[, paste0("Best ", bmd_param)] >=
PoD_percentile_25 &

                           bmd_filtered[, paste0("Best ", bmd_param)] >= PoD_percentile_75,

                         paste0("Best ", bmd_param)])

  lowest_probe <-  bmd_filtered[order(bmd_filtered[, paste0("Best ", bmd_param)], decreasing = F)[1],]

  lowest_probe_BMD_param <-  lowest_probe[, paste0("Best ", bmd_param)] # get lowest BMD/BMDL
Conc

  lowest_probe <-  lowest_probe[, "Probe ID"] # get probe name with lowest bmd param

  lowest_conc <- get_highest_lowest_conc_value(analysis = analysis_name, bmdExpress_input_dir =
bmdExpress_input_dir, get_value = "Lowest" )

  bmd_filtered_above_1st_conc <- bmd_filtered[bmd_filtered[, paste0("Best ", bmd_param)] >=
lowest_conc,]

  lowest_bmd_filtered_above_1st_conc <-
bmd_filtered_above_1st_conc[order(bmd_filtered_above_1st_conc[, paste0("Best ", bmd_param)],
decreasing = F)[1],]

  lowest_bmd_filtered_above_1st_conc_BMD_param <- lowest_bmd_filtered_above_1st_conc[,
paste0("Best ", bmd_param)] # get lowest BMD/BMDL Conc above lowest tested concentration
```

```
  lowest_bmd_filtered_above_1st_conc_probe <- lowest_bmd_filtered_above_1st_conc[, "Probe ID"] # get lowest BMD/BMDL Probe above lowest tested concentration



  PoD_percentiles <- t(as.data.frame(quantile(bmd_filtered[, paste0("Best ", bmd_param)], c(.05, 0.10, .2, .3, .4, .5, .6 ,.7, .8, .9, .95))))

  gene_PoDs <- data.frame(analysis_name,

                          PoD_top20FC,

                          PoD_percentile,

                          "",

                          lowest_probe_BMD_param,

                          lowest_probe,

                          lowest_bmd_filtered_above_1st_conc_BMD_param,

                          lowest_bmd_filtered_above_1st_conc_probe,

                          stringsAsFactors = F)

  colnames(gene_PoDs) <- c("Analysis",

                           paste0("Avg of 20 ", bmd_param, "s with highest FC"),

                           paste0("Avg of 25th-75th percentile of ", bmd_param, "s"),

                           "-",

                           paste0("Lowest probe ", bmd_param),

                           paste0("Lowest probe ID"),

                           paste0("Lowest probe ", bmd_param, "after lowest dose"),

                           paste0("Lowest probe ID after lowest dose")

  )

  gene_PoDs <- cbind(gene_PoDs, PoD_percentiles)

  row.names(gene_PoDs) <- analysis_name

  return(gene_PoDs)

}

#' Filter Significant Pathway BMDs

#'
```

#' BMDExpress2 Pathway level BMD results can be extracted using bmdexpress2-cmd export --bm2-file {bm2_file_name.bm2} --analysis-group categorical --output-file-name {filename.txt}. These data should be filtered using this function for significance.

#' Filtering for significant pathway level BMDs include:

#' 1. Filtering by number of total genes (with a dose response) in the pathway.

#' 2. Filtering by the number of significant dose responsive genes found in the pathway. NOTE: significant genes to filter pathway BMDs are defined in the BMDExpress2 configuration file not by this R package (NOT filter_gene_bmds()).

#' 3. Filtering by Fishers 2-tail p value.

#' This function can be run on multiple analysis groups if all analysis groups should be filtered the same.

#' This function will write the returned data.frame of filtered pathway level BMDs to csv file at {path_csv_output_dir} with file name {path_csv_output_prefix}_path_bmd_filtered.csv

#'

#' @param path_data Dataframe of BMDExpress pathway bmd data. Minimum required columns:

#' - Input Genes

#' - Genes That Passed All Filters

#' - Fisher's Exact Two Tail

#' @param min_total_genes Numeric parameter indicating the value to filter 'Input Genes'. Data are filtered 'Input Genes' >= min_total_genes To not use the filter set min_total_genes as NULL. Default = 3

#' @param min_sig_genes Numeric parameter indicating the value to filter 'Genes That Passed All Filters'. Data are filtered 'Genes That Passed All Filters' <= min_sig_genes. To not use the filter set min_sig_genes as NULL. Default = 2

#' @param fishers_p_val Numeric parameter indicating the value to filter on 'Fisher's Exact Two Tail'. Data are filtered 'Fisher's Exact Two Tail' < fishers_p_val To not use the filter set fishers_p_val as NULL. Default = 0.1

#' @param path_csv_output_prefix Character string to use as the prefix the the outputed csv of filtered data. Recommended is the analysis group from BMDExpress. If no csv is required, use NULL (default).

#' @param path_csv_output_dir The directory path to where the output csv of filtered data should be written to. This directory should already exist.

#'

#' @importFrom glue glue

#' @importFrom logger log_fatal log_info

#' @return dataframe of filtered pathway BMD data

#' @export

```
filter_pathway_bmds <- function(path_data, min_total_genes = 3, min_sig_genes = 2, fishers_p_val = 0.1, path_csv_output_prefix = NULL, path_csv_output_dir = ".") {

  # Parameter Error Handling

  # path_data

  col_names <- c("Analysis", "Input Genes", "Genes That Passed All Filters", "Fisher's Exact Two Tail") # minimum columns for filtering

  check <- check_data(path_data, "filter_pathway_bmds")

  check$type()

  check$cols(col_names)

  check$empty()

    # filter parameters

  filter_by <- tibble::lst(min_total_genes,

                           min_sig_genes,

                           fishers_p_val)

  for (f in names(filter_by)) {

    filter_value <- filter_by[[f]]

    if(!is.null(filter_value)) {

      if(!is.numeric(filter_value)) {

        logger::log_fatal("{f} '{filter_value}' passed to function 'filter_pathway_bmds' is not numeric")

        stop(glue::glue("'{filter_value}' passed to function 'filter_pathway_bmds' is not numeric. Please supply numeric value or NULL."))

      }

    }

  }

  # path_csv_output_dir

  create_dir(path_csv_output_dir)

  path_data <- path_data[path_data$`Genes That Passed All Filters` != 0, ] # remove blank pathway enrichment rows (as no genes present)

  for (i in 1:ncol(path_data)) {

    if (!any(is.na(suppressWarnings(as.numeric(path_data[, i]))))) {
```

```
    path_data[, i] <- as.numeric(path_data[, i])
  }
 }
 filtered_path_data <- path_data
 if (!is.null(min_total_genes)) {
  filtered_path_data <- filtered_path_data[filtered_path_data$`Input Genes` >= min_total_genes, ]
 } else {
  min_total_genes <- "None"
 }
 if (!is.null(min_sig_genes)) {
  filtered_path_data <- filtered_path_data[filtered_path_data$`Genes That Passed All Filters` >=
min_sig_genes, ]
 } else {
  min_sig_genes <- "None"
 }
 if (!is.null(fishers_p_val)) {
  filtered_path_data <- filtered_path_data[filtered_path_data$`Fisher's Exact Two Tail` < fishers_p_val,
]
 } else {
  fishers_p_val <- "None"
 }
 if (!is.null(path_csv_output_prefix)) {
  # Get a table of filter values
  filters_used <- data.frame(c(
   paste0("Input Genes >= ", min_total_genes),
   paste0("Genes That Passed All Filters >= ", min_sig_genes),
   paste0("Fisher's Exact Two Tail < ", fishers_p_val)
  ))
  colnames(filters_used) <- "These data are filtered using the following column filters:"
```

```
    write_filtered_data(filtered_data = filtered_path_data,
                        filter_info = filters_used,
                        output_file_name = paste0(path_csv_output_prefix, "_pathway_bmd_filtered.csv"),
                        output_dir = path_csv_output_dir)
  }
  return(filtered_path_data)
}

#' Calculate pathway level PoDs from BMD data
#'
#' Using data from the filter_pathway_bmds function, PoDs are calculated via 3 methods.
#' 1. Average of BMD/BMDLs of 20 pathways with the lowest Fishers 2-tail p values.
#' 2. Average of BMD/BMDLs of 20 pathways with the lowest BMD/BMDLs mean values.
#' 3. Taking the lowest pathways BMD/BMDLs mean value.
#'
#' This function assumed all pathways inputted to the function have significant BMD/BMDLs and that only one analysis group is passed (eg one cell line- chemical)
#'
#' @param pathway_filtered output of the filter_pathway_bmds. Minimum columns required:
#' - Analysis
#' - GO/Pathway/Gene Set Name
#' - GO/Pathway/Gene Set ID
#' - Fisher's Exact Two Tail
#' - Mean BMD or Mean BMDL (dependent on the bmd_param parameter)
#'
#' @param bmd_param "BMD" or "BMDL" indicating which value to base calculations around.
#' @param bmdExpress_input_dir directory path where BMD input files for BMDExpress are located
#' @importFrom logger log_info
#' @importFrom stats quantile
#'
```

```
#' @return dataframe of pathway level PoDs for one analysis group

#' @export

calculate_pathway_PoD <- function(pathway_filtered, bmd_param, bmdExpress_input_dir) {

 # Parameter Error Handling

 # bmd_param

 if (toupper(bmd_param) != "BMD" & toupper(bmd_param) != "BMDL") {

  logger::log_fatal("bmd_param '{bmd_param}' passed to function 'calculate_gene_PoD' is not 'BMD' or 'BMDL'.")

  stop("bmd_param passed to function 'calculate_gene_PoD' is not 'BMD' or 'BMDL'.")

 } else {

  bmd_param <- toupper(bmd_param)

 }

 # pathway_filtered

 #replace version 2.3 columns anems with version 2.0 column names

 colnames(pathway_filtered) <- gsub("GO/Pathway/Gene Set/Gene Name", "GO/Pathway/Gene Set Name", colnames(pathway_filtered))

 colnames(pathway_filtered) <- gsub("GO/Pathway/Gene Set/Gene ID", "GO/Pathway/Gene Set ID", colnames(pathway_filtered))

 col_names <- c("Analysis", "GO/Pathway/Gene Set Name", "GO/Pathway/Gene Set ID", "Fisher's Exact Two Tail", paste0(bmd_param, " Mean"))

 check <- check_data(pathway_filtered, "calculate_pathway_PoD")

 check$type()

 check$cols(col_names)

 check$empty()

 check$analysis_length()

 if(nrow(pathway_filtered) >= 20) {

  min20bmd <- pathway_filtered[order(pathway_filtered[, paste0(bmd_param, " Mean")], decreasing = F)[1:20], paste0(bmd_param, " Mean")] # Get 20 lowest BMD/BMDL values

  minbmd <- min(min20bmd) # get the lowest BMD/BMDL value (complete.cases in line above retains orignal order of data so min has to be used in this line)

  min20p <- pathway_filtered[order(pathway_filtered$`Fisher's Exact Two Tail`, decreasing = F)[1:20], paste0(bmd_param, " Mean")] # Get BMD/BMDL values for the pathways with the 20 lowest fishers 2 tail p value
```

```
  analysis_name <- unique(pathway_filtered$Analysis)

  } else {

  min20bmd <- pathway_filtered[, paste0(bmd_param, " Mean")] # Get 20 lowest BMD/BMDL values

  minbmd <- min(min20bmd) # get the lowest BMD/BMDL value (complete.cases in line above retains orignal order of data so min has to be used in this line)

  min20p <- pathway_filtered[, paste0(bmd_param, " Mean")]

  analysis_name <- paste0(unique(pathway_filtered$Analysis), " *(", length(min20bmd), " bmds)")

 }

 PoD_min20bmd <- mean(min20bmd)

 PoD_minbmd <- minbmd

 PoD_min20p <- mean(min20p)

 lowest_conc <- get_highest_lowest_conc_value(analysis = analysis_name, bmdExpress_input_dir = bmdExpress_input_dir, get_value = "Lowest")

 path_filtered_above_1st_conc <- pathway_filtered[pathway_filtered[, paste0(bmd_param, " Mean")] >= lowest_conc,]

 lowest_path_filtered_above_1st_conc <-
path_filtered_above_1st_conc[order(path_filtered_above_1st_conc[, paste0(bmd_param, " Mean")], decreasing = F)[1],]

 lowest_path_filtered_above_1st_conc_BMD_param <- lowest_path_filtered_above_1st_conc[, paste0(bmd_param, " Mean")] # get lowest BMD/BMDL Conc above lowest tested concentration

 lowest_path_filtered_above_1st_conc_path <-
lowest_path_filtered_above_1st_conc[,"GO/Pathway/Gene Set Name"] # get lowest BMD/BMDL pathway above lowest tested concentration

 PoD_percentiles <- t(as.data.frame(quantile(pathway_filtered[, paste0(bmd_param, " Mean")], c(.05, 0.10, .2, .3, .4, .5, .6 ,.7, .8, .9, .95))))

 pathway_PoDs <- data.frame(analysis_name,

                            PoD_min20bmd,

                            PoD_minbmd,

                            PoD_min20p,

                            "",

                            lowest_path_filtered_above_1st_conc_BMD_param,

                            lowest_path_filtered_above_1st_conc_path,

                            stringsAsFactors = F)
```

```
  colnames(pathway_PoDs) <- c("Analysis",
            paste0("Avg of 20 lowest pathway ", bmd_param, "s"),
            paste0("The lowest pathway ", bmd_param),
            paste0("Avg of 20 pathway ", bmd_param, "s with lowest 2-tail fisher P values"),
            "-",
            paste0("Lowest pathway ", bmd_param, "after lowest dose"),
            paste0("Lowest pathway after lowest dose")
  )
  pathway_PoDs <- cbind(pathway_PoDs, PoD_percentiles)
  rownames(pathway_PoDs) <- analysis_name
  return(pathway_PoDs)
}

#' Determine the highest concentration filter to be used based on a combination of highest_conc_filter and bmdExpress_input_dir parameters
#'
#' This function's set the rules of high_conc_filter is determined for use in function calculate_PoDs_from_BMDExpress2.
#' - Rule 1: When a numeric value is given to highest_conc_filter, then this value will always be used.
#' - Rule 2: When both highest_conc_filter and bmdExpress_input_dir are NULL then the final highest concentration wont be filtered and therefore final value set to NULL.
#' - Rule 3: When highest_conc_filter is NULL but bmdExpress_input_dir is a valid file path to text files of the input files for BMDExpress2, the highest concentration value is set to the highest value found in the files which matches the BMDExpress2 analysis group.
#'
#' @param analysis BMDExpress2 analysis group name. This shall be used to match bmdExpress_input_dir files if present.
#' @param bmdExpress_input_dir The directory path for the BMDExpress2 input text file of which highest concentration shall be extracted. Can also be NULL in cases of rule 1 and 2.
#' @param highest_conc_filter The parameter value passed to calculate_PoDs_from_BMDExpress2. Can be numeric value for rule 1 or NULL for rules 2 and 3.
#' @param get_value "Highest" or "Lowest" for concentration to be used
#'
#' @importFrom logger log_info
```

```
#' @importFrom utils read.delim

#' @return The final determined highest_conc_filter value

get_highest_lowest_conc_value <- function(analysis, bmdExpress_input_dir, highest_conc_filter = NULL, get_value = "Highest") {

 get_value_accepted <- c("Highest", "Lowest")

 if(!get_value %in% get_value_accepted){

  logger::log_fatal("get_value passes to get_highest_lowest_conc_value is not either of {get_value_accepted}. Value passed: {get_value}.")

  stop("Incorrect get_value parameter.")

 }

 if (is.null(highest_conc_filter) | get_value == "Lowest") {

  if (!is.null(bmdExpress_input_dir)) {

   input_list <- list.files(path = bmdExpress_input_dir)

   analysis_input <- input_list[unlist(lapply(

    X = gsub(".txt", "", input_list),

    FUN = grepl,

    x = analysis

   ))] # Get BMDExpress2 input file from bmdExpress_input_dir which has the part of the analysis string in - The original data

   if(length(analysis_input) == 0) {

    logger::log_fatal("BMDExpress input file for {analysis} cannot be found at {bmdExpress_input_dir}. This means a concentration to filter cannot be determined and analysis is terminated.")

    stop("Analysis group not found in provide BMDExpress input files.")

   }

   input_file_path <- file.path(bmdExpress_input_dir, analysis_input)

   logger::log_debug("{get_value} concentration for {analysis} is pulled from {input_file_path}")

   input_data <- read.delim(input_file_path,

                            header = F, skip = 1, nrows = 1)[1, -1]

   if(get_value == "Highest") {

    chosen_filter <- max(input_data) # Get max concentration for the BMD input file from the second row of the text file.
```

```
    } else {

      input_data_not0 = input_data[input_data!=0]

      chosen_filter <- min(input_data_not0) # Get min concentration for the BMD input file from the second row of the text file.

    }

  } else {

    chosen_filter <- NULL

  }

 } else {

  chosen_filter <- highest_conc_filter

 }

 if(get_value == "Highest") {

  logger::log_info("For {analysis} the chosen highest_conc_filter is {chosen_filter}.")

 } else {

  logger::log_info("For {analysis} the chosen lowest concentration is {chosen_filter}.")

 }

 return(chosen_filter)

}

#' Write PoD file

#'

#' PoD's shall be written to a csv file at {pod_csv_output_prefix} with file name {pod_csv_output_dir}_PoDs.csv. File will include filters applied to the data and gene and pathway level PoDs

#

#' @param defined_conc concentration to be used (TODO Jade to add more detail here)

#' @param bmd_param "BMD" or "BMDL" indicating which value to base calculations around

#' @param collated_gene_PoDs Calculated gene level PoDs (data to be included in csv)

#' @param collated_pathway_PoDs Calculated pathway level PoDs (data to be included in csv)

#' @param pod_csv_output_prefix Character string to use as the prefix the the outputed csv of PoDs calculated at both the gene and pathway level. Recommended is the analysis group from BMDExpress. If no csv is required, use NULL (default).
```

```
#' @param pod_csv_output_dir The directory path to where the output csv of PoDs calculated at both the gene and pathway level should be written to. This directory should already exist. If NULL no file will be written.


#' @inheritParams filter_gene_bmds

#' @inheritParams filter_pathway_bmds

#'

#' @importFrom logger log_info

write_PoDs_to_file <- function(pod_csv_output_prefix,

                               defined_conc,

                               highest_conc_filter,

                               bmd_param,

                               bmdl_bmdu_filter,

                               fitP_filter,

                               min_total_genes,

                               min_sig_genes,

                               fishers_p_val,

                               pod_csv_output_dir,

                               collated_gene_PoDs,

                               collated_pathway_PoDs) {


  filters_used_gene <- data.frame(c(

    paste0("PoDs based on ", bmd_param),

    paste0("Best ", bmd_param, " <= ", defined_conc),

    paste0("Best BMDU/BMDL < ", bmdl_bmdu_filter),

    paste0("Best fitPValue > ", fitP_filter)

  ))

  colnames(filters_used_gene) <- "Gene BMDExpress2 results data are filtered using the following column filters:"

  if (is.null(min_total_genes)) min_total_genes <- "None"
```

```
  if (is.null(min_sig_genes)) min_sig_genes <- "None"

  if (is.null(fishers_p_val)) fishers_p_val <- "None"



  filters_used_path <- data.frame(c(

    paste0("Input Genes >= ", min_total_genes),

    paste0("Genes That Passed All Filters >= ", min_sig_genes),

    paste0("Fisher's Exact Two Tail < ", fishers_p_val)))

  colnames(filters_used_path) <- "Pathway BMDExpress2 results are filtered using the following column filters:"

  create_dir(pod_csv_output_dir)

  file_path <- file.path(pod_csv_output_dir, paste0(pod_csv_output_prefix," ", bmd_param, "_PoDs.csv"))

  write.table(filters_used_gene, file_path, col.names = T, sep = ",", row.names = F)

  suppressWarnings(write.table(data.frame(), file_path, col.names = T, sep = ",", append = T))

  suppressWarnings(write.table(filters_used_path, file_path, col.names = T, sep = ",", append = T, row.names =F))

  suppressWarnings(write.table(data.frame(), file_path, col.names = T, sep = ",", append = T))

  suppressWarnings(write.table(collated_gene_PoDs, file_path, col.names = T, sep = ",", append = T, row.names = F))

  suppressWarnings(write.table(data.frame(), file_path, col.names = T, sep = ",", append = T))

  suppressWarnings(write.table(collated_pathway_PoDs, file_path, col.names = T, sep = ",", append = T, row.names = F))

  logger::log_info("PoD file successfully written to {file_path}.")

}

#' Calculate gene level PoDs from BMDExpress2 output for a group of analysis

#'

#' Loop through each analysis group and call get_highest_lowest_conc_value, filter_gene_bmds and calculate_gene_PoD and then collate and output all gene PoDs.

#'

#' @param gene_bmd_file_path File path for the gene level bmd text file {filename.txt}, directly from bmdexpress2-cmd export --bm2-file {bm2_file_name.bm2} --analysis-group bmd --output-file-name {filename.txt}.
```

```
#' @param bmdExpress_input_dir Directory path for where the BMDExpress2 input files are. These files are read in for each analysis to get the highest concentration used to filter gene level BMDs by. This will override any value used set by highest_conc_filter.

#' @inheritParams filter_gene_bmds

#' @inheritParams calculate_gene_PoD

#'

#' @importFrom logger log_info

#' @importFrom utils read.delim

#' @return dataframe gene level PoDs for all analysis groups in the BMD files.

run_gene_level_BMD_analysis <- function(gene_bmd_file_path,

                        gene_csv_output_dir,

                        bmdExpress_input_dir = NULL,

                        highest_conc_filter = NULL,

                        bmd_param = "BMDL",

                        bmdl_bmdu_filter = 40,

                        fitP_filter = 0.1

                        ) {

 # gene_bmd_file_path

 if(!file.exists(gene_bmd_file_path)) {

  logger::log_fatal("gene_bmd_file_path '{gene_bmd_file_path}' passed to function 'calculate_PoDs_from_BMDExpress2' has no such file or directory")

  stop("gene_bmd_file_path passed to function 'calculate_PoDs_from_BMDExpress2' does not exist")

 }

 # bmdExpress_input_dir

 if (!is.null(bmdExpress_input_dir)) {

  if (!dir.exists(bmdExpress_input_dir)) {

   logger::log_fatal("bmdExpress_input_dir '{bmdExpress_input_dir}' passed to function 'calculate_PoDs_from_BMDExpress2' does not exist")

   stop("bmdExpress_input_dir passed to function 'calculate_PoDs_from_BMDExpress2' does not exist")

  }
```

```
  }

  gene_data <- as.data.frame(read.delim(gene_bmd_file_path, sep = "\t", header = F, stringsAsFactors =
F, skip = 1)) # header= T throws errors due to NA columns

  # set column names

  gene_data_cols <- as.character(gene_data[1, ])

  colnames(gene_data) <- gene_data_cols

  gene_data <- gene_data[-1, ]

  gene_data <- gene_data[, !apply(is.na(gene_data), 2, all)] # remove NA columns

  collated_gene_PoDs <- data.frame(matrix(nrow = 0, ncol = 19))

  for (analysis in unique(gene_data$Analysis)) {

    analysis_short_name <- gsub("_williams_0.05_NOMTC_foldfilter1.5_BMD", "", analysis)

    analysis_bmd <- gene_data[gene_data$Analysis == analysis, ]

    defined_conc <- get_highest_lowest_conc_value(analysis, bmdExpress_input_dir, highest_conc_filter)

    if (nrow(analysis_bmd) > 0) {

      bmd_filtered <- filter_gene_bmds(

        bmd_data = analysis_bmd,

        highest_conc_filter = defined_conc,

        bmdl_bmdu_filter = bmdl_bmdu_filter,

        fitP_filter = fitP_filter,

        gene_csv_output_prefix = analysis_short_name,

        gene_csv_output_dir = gene_csv_output_dir

      )

      if (nrow(bmd_filtered) > 0) {

        gene_PoDs <- calculate_gene_PoD(bmd_filtered = bmd_filtered, bmd_param = bmd_param,
bmdExpress_input_dir = bmdExpress_input_dir)

        collated_gene_PoDs <- rbind(collated_gene_PoDs, gene_PoDs)

      } else {

        collated_gene_PoDs[paste0(analysis, "_*(0 bmds)"),] <- c(paste0(analysis, "_*(0 bmds)"), rep(NA, 2),
"", rep(NA, 15)) # NA for when no pathway data is returned so no PoD can be calculated.

      }
```

```
  }

 }

 for (i in 1:ncol(collated_gene_PoDs)) {

  if (!any(is.na(suppressWarnings(as.numeric(collated_gene_PoDs[!is.na(collated_gene_PoDs[,i]), i]))))) { # check whether columns can be converted to numeric, if so convert. (NAs appear when non numeric characters are attempted to be converted so we check for NAs to determine suitability)

   collated_gene_PoDs[, i] <- as.numeric(collated_gene_PoDs[, i])

  }

 }

 return(collated_gene_PoDs)

}

#' Calculate pathway level PoDs from BMDExpress2 output for a group of analysis

#'

#' Loop through each analysis group and call filter_pathway_bmds and calculate_pathway_PoD and then collate and output all pathway PoDs.

#'

#' @param path_bmd_file_path File path for the pathway level bmd text file {filename.txt}, directly from bmdexpress2-cmd export --bm2-file {bm2_file_name.bm2} --analysis-group categorical --output-file-name {filename.txt}.

#' @param highest_conc_filter The parameter value passed to calculate_PoDs_from_BMDExpress2. Can be numeric value for rule 1 or NULL for rules 2 and 3.

#' @inheritParams filter_pathway_bmds

#' @inheritParams calculate_pathway_PoD

#' @importFrom logger log_fatal

#' @importFrom utils read.delim

#' @return dataframe pathway level PoDs for all analysis groups in the BMD files.

#' @export

run_pathway_level_BMD_analysis <- function(path_bmd_file_path,

                                           highest_conc_filter,

                                           bmd_param = "BMDL",

                                           min_total_genes = 3,
```

```
                                min_sig_genes = 2,

                                fishers_p_val = 0.1,

                                path_csv_output_dir,

                                bmdExpress_input_dir) {

 # path_bmd_file_path

 if(!file.exists(path_bmd_file_path)) {

  logger::log_fatal("path_bmd_file_path '{path_bmd_file_path}' passed to function
'calculate_PoDs_from_BMDExpress2' has no such file or directory")

  stop("path_bmd_file_path passed to function 'calculate_PoDs_from_BMDExpress2' does not exist")

 }

  path_data <- as.data.frame(read.delim(path_bmd_file_path, sep = "\t", header = F, stringsAsFactors =
F, skip = 1)) # header= T throws errors due to NA columns

 # set column names

 path_data_cols <- as.character(path_data[1, ])

 colnames(path_data) <- path_data_cols

 path_data <- path_data[-1, ]

 path_data <- path_data[, !apply(is.na(path_data), 2, all)] # remove na columns

  collated_pathway_PoDs <- data.frame(matrix(nrow = 0, ncol = 18), stringsAsFactors = F)

 colnames(collated_pathway_PoDs) <- c("Analysis",

                                "Avg of 20 lowest pathway BMDLs",

                                "The lowest pathway BMDL",

                                "Avg of 20 pathway BMDLs with lowest 2-tail fisher P values",

                                "-",

                                "Lowest pathway BMDLafter lowest dose",

                                "Lowest pathway after lowest dose",

                                "5%",

                                "10%",

                                "20%",

                                "30%",
```

```
                    "40%",
                    "50%",
                    "60%",
                    "70%",
                    "80%",
                    "90%",
                    "95%")
  for (analysis in unique(path_data$Analysis)) {
    analysis_short_name <-
gsub("_williams_0.05_NOMTC_foldfilter1.5_BMD_WT_Human_REACTOME_true_true_pval0.1_ratio40"
, "", analysis)
    analysis_bmd <- path_data[path_data$Analysis == analysis, ]
    analysis_pathway_filtered <- filter_pathway_bmds(
      path_data = analysis_bmd,
      min_total_genes = min_total_genes,
      min_sig_genes = min_sig_genes,
      fishers_p_val = fishers_p_val,
      path_csv_output_prefix = analysis_short_name,
      path_csv_output_dir = path_csv_output_dir
    )
    if (nrow(analysis_pathway_filtered) > 0) {
      path_PoDs <- calculate_pathway_PoD(pathway_filtered = analysis_pathway_filtered, bmd_param =
bmd_param, bmdExpress_input_dir = bmdExpress_input_dir)
      collated_pathway_PoDs <- rbind(collated_pathway_PoDs, path_PoDs)
    } else {
      collated_pathway_PoDs[paste0(analysis, " *(0 bmds)"),] <- c(paste0(analysis, " *(0 bmds)"), rep(NA,
3), "", rep(NA,13)) # NA for when no pathway data is returned so no PoD can be calculated.
    }
  }
  for (i in 1:ncol(collated_pathway_PoDs)) {
```

```
  if (!any(is.na(suppressWarnings(as.numeric(collated_pathway_PoDs[!is.na(collated_pathway_PoDs[,i]), i]))))) { # check whether columns can be converted to numeric, if so convert. (NAs appear when non numeric characters are attempted to be converted so we check for NAs to determine suitability)

    collated_pathway_PoDs[, i] <- as.numeric(collated_pathway_PoDs[, i])

  }

 }

 return(collated_pathway_PoDs)

}

#' Calculate gene and pathway level PoDs from BMDExpress2 output

#'

#' Point of departures are calculated from BMDExpress2's gene results for each analysis group (BMDExpress2 input file). Gene level BMDs can be extracted with bmdexpress2-cmd export --bm2-file {bm2_file_name.bm2} --analysis-group bmd --output-file-name {filename.txt}.

#' Pathway level BMDs can be extracted with bmdexpress2-cmd export --bm2-file {bm2_file_name.bm2} --analysis-group categorical --output-file-name {filename.txt}.

#' PoDs are calculated from a subset of the genes/pathways which are deemed as significant based on given filter parameters and the chemical PoD are defined using different methods:

#'

#' Gene Level PoDs

#' 1. Average of BMD/BMDLs of 20 genes with the largest fold change.

#' 2. Average of the 25th and 75th percentile of BMD/BMDLs

#'

#' Pathway Level PoDs

#' 3. Average of BMD/BMDLs of 20 pathways with the lowest Fishers 2-tail p values.

#' 4. Average of BMD/BMDLs of 20 pathways with the lowest BMD/BMDLs mean values.

#' 5. Taking the lowest pathways BMD/BMDLs mean value.

#'

#' This function loops through each analysis group (cell line- chemical) in the BMDExpress2 text files and run functions 'filter_gene_bmds', 'calculate_gene_PoD', 'filter_pathway_bmds' and 'calculate_pathway_PoD' and collate results.

#' Each BMDExpress analysis will have filtered bmd and pathway csv exported. Output files will have BMDExpress2 analysis name with appended with "_bmd_filtered" and "_pathway_bmd_filtered".Where BMDExpress2 bmd analysis names have "_williams_0.05_NOMTC_foldfilter1.5_BMD" and pathway analysis names have
```

"_williams_0.05_NOMTC_foldfilter1.5_BMD_WT_Human_REACTOME_true_true_pval0.1_ratio40" in them, this text will be removed from the file output name to reduce characters, otherwise the whole BMDExpress analysis name will be seen.

#'

#' PoD's shall be written to a csv file at {pod_csv_output_prefix} with file name {pod_csv_output_dir}_PoDs.csv

#'

#' @param gene_bmd_file_path File path for the gene level bmd text file {filename.txt}, directly from bmdexpress2-cmd export --bm2-file {bm2_file_name.bm2} --analysis-group bmd --output-file-name {filename.txt}.

#' @param path_bmd_file_path File path for the pathway level bmd text file {filename.txt}, directly from bmdexpress2-cmd export --bm2-file {bm2_file_name.bm2} --analysis-group categorical --output-file-name {filename.txt}.

#' @param bmdExpress_input_dir Directory path for where the BMDExpress2 input files are. These files are read in for each analysis to get the highest concentration used to filter gene level BMDs by. This will override any value used set by highest_conc_filter.

#' @param bmd_param "BMD" or "BMDL" indicating which value to base calculations around.

#' @inheritParams filter_gene_bmds

#' @inheritParams filter_pathway_bmds

#' @inheritParams write_PoDs_to_file

#'

#' @return list of 2 dataframes, gene_PoDs = gene level PoDs and pathway_PoDs = pathway level PoDs for all analysis groups in the BMD files.

#' @export

calculate_PoDs_from_BMDExpress2 <- function(gene_bmd_file_path,

path_bmd_file_path,

bmdExpress_input_dir = NULL,

highest_conc_filter = NULL,

bmd_param = "BMDL",

bmdl_bmdu_filter = 40,

fitP_filter = 0.1,

min_total_genes = 3,

min_sig_genes = 2,

fishers_p_val = 0.1,

```
                          gene_csv_output_dir = ".",
                          path_csv_output_dir = ".",
                          pod_csv_output_prefix = "",
                          pod_csv_output_dir = ".") {
 collated_gene_PoDs <- run_gene_level_BMD_analysis(gene_bmd_file_path = gene_bmd_file_path,
                                  bmdExpress_input_dir = bmdExpress_input_dir,
                                  highest_conc_filter = highest_conc_filter,
                                  bmd_param = bmd_param,
                                  bmdl_bmdu_filter = bmdl_bmdu_filter,
                                  fitP_filter = fitP_filter,
                                  gene_csv_output_dir = gene_csv_output_dir)
 collated_pathway_PoDs <- run_pathway_level_BMD_analysis(path_bmd_file_path =
path_bmd_file_path,
                                    bmd_param = bmd_param,
                                    highest_conc_filter = highest_conc_filter,
                                    min_total_genes = min_total_genes,
                                    min_sig_genes = min_sig_genes,
                                    fishers_p_val = fishers_p_val,
                                    path_csv_output_dir = path_csv_output_dir,
                                    bmdExpress_input_dir = bmdExpress_input_dir)
 if (!is.null(pod_csv_output_dir)) {
   if (is.null(bmdExpress_input_dir) & is.null(highest_conc_filter)) {
     conc <- "None"
   } else if (!is.null(bmdExpress_input_dir) & is.null(highest_conc_filter)) {
     conc <- "Highest tested concentration"
   } else {
     conc <- highest_conc_filter
   }
 write_PoDs_to_file(pod_csv_output_prefix,
```

```
            conc,
            highest_conc_filter,
            bmd_param,
            bmdl_bmdu_filter,
            fitP_filter,
            min_total_genes,
            min_sig_genes,
            fishers_p_val,
            pod_csv_output_dir,
            collated_gene_PoDs,
            collated_pathway_PoDs)

  }
  return(list(gene_PoDs = collated_gene_PoDs,
         pathway_PoDs = collated_pathway_PoDs))
}
```

Supplementary Script 5 – Calculating PoD from BMDExpress2 using R Studio (Github reference link:https://github.com/liztulum/MRes-thesis-scripts/blob/main/Calculate_PoDs_from%20_BMDExpress2.R)

```
calculate_PoDs_from_BMDExpress2(gene_bmd_file_path = "bmd_gene.txt",
                    path_bmd_file_path = "bmd_pathway.txt",
                    bmdExpress_input_dir = "BMD_input",
                    gene_csv_output_dir = "BMD_output_filtered/gene_bmds_filtered",
                    path_csv_output_dir = "BMD_output_filtered/pathway_bmds_filtered",
                    pod_csv_output_prefix = "pods",
                    pod_csv_output_dir = ".")
```

APPENDIX 2

The top 20 DEG's for HepaRG, HepG2 and MCF-7 cell lines dosed with doxorubicin and niacinamide at various doses

Supplementary Table 1 - HepaRG Doxorubicin 1uM

Gene	baseMean	log2Fold Change	lfcSE	stat	pvalue	padj
TSC22D3_7366	3066.3732	2.67985490	0.420136	6.3785469	1.79E-10	3.44E-08
MUC1_4366	42.950120	2.51909863	0.52762	4.7744541	1.80E-06	0.00011754
COX4I1_1515	545.26190	0.82955680	0.263074	3.1533201	0.00161425	0.02669541
IER3_3214	175.68949	1.34579300	0.382493	3.5184734	0.00043404	0.00991082
C3_886	1449.9124	1.1931151	0.354167	3.3687953	0.00075497	0.01494612
CCNB1IP1_1054	490.40493	0.88674597	0.219813	4.0340974	5.48E-05	0.00187987
ARPC5L_463	771.04092	0.79558759	0.25373	3.1355704	0.00171523	0.0276663
CDC26_1162	256.96022	0.80127001	0.225663	3.5507364	0.00038415	0.00903647
CALU_944	37.345272	2.00997350	0.536098	3.7492684	0.00017735	0.00480727
CCND1_1062	98.508150	1.55313804	0.421007	3.6890996	0.00022505	0.00580237
ALDH7A1_227	36.896090	1.64870435	0.511464	3.2235007	0.00126634	0.02212648
ARL6IP4_442	135.96376	1.26947646	0.381754	3.3253796	0.00088298	0.01690393
COL5A1_1480	513.18098	1.28552925	0.396414	3.2428963	0.00118321	0.02091961
CTSD_1642	2756.8044	0.67965930	0.233531	2.9103556	0.00361018	0.0484474
DDB2_1798	595.53087	0.98200603	0.160495	6.1186088	9.44E-10	1.55E-07
F2_2286	1069.7516	2.75284869	0.704666	3.9065987	9.36E-05	0.00286983
FHL2_2423	79.755736	1.79872609	0.470573	3.8224167	0.00013215	0.00380909
DRAP1_1957	287.65094	1.12596204	0.248754	4.5264000	6.00E-06	0.00031221
FEN1_2387	106.39603	1.66498619	0.348837	4.7729618	1.82E-06	0.00011794
HSPG2_3153	172.88822	0.82013851	0.246472	3.3275055	0.00087627	0.01683027

Supplementary Table 2 - HepaRG Doxorubicin 0.2uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
F2_2286	1069.75	2.713072	0.7045	3.8506	0.000118	0.01903
ORM1_4833	12107.7	2.177266	0.5221	4.1697	3.05E-05	0.008714
APOC3_356	4130.58	3.281813	0.7505	4.3724	1.23E-05	0.004563
FABP1_2298	81.4128	3.293434	0.8234	3.9996	6.34E-05	0.01219
TWIST1_7416	1165.4	-0.86234	0.2293	3.7605	0.00017	0.024858
SYTL1_6923	145.509	1.141579	0.3138	3.6377	0.000275	0.034867
VTN_7690	948.910	2.218009	0.5505	4.0284	5.61E-05	0.01219
TP53I3_7290	1006.11	0.854612	0.2315	3.6913	0.000223	0.030315
TUBA3D_7401	30.9310	-2.24717	0.5280	4.2553	2.09E-05	0.006494
ALB_217	17799.1	2.943891	0.7269	4.0493	5.14E-05	0.011775
CFB_872	5630.3	1.998505	0.4040	4.9463	7.56E-07	0.000702
HP_3085	88948.2	2.407425	0.6229	3.8646	0.000111	0.018232
ORM2_4836	2856.77	1.834237	0.4960	3.6980	0.000217	0.029899
CP_10542	1898.80	1.869051	0.5220	3.5803	0.000343	0.041128
GPR87_10766	617.617	0.923777	0.2571	3.5927	0.000327	0.039631
SERPINC1_11069	501.634	3.61364	0.9050	3.9928	6.53E-05	0.01233
FGL1_11631	900.767	3.471283	0.8198	4.2341	2.29E-05	0.006728
ARMC9_12765	557.600	-0.65645	0.1799	3.6489	0.000263	0.034311
DRAXIN_12848	61.0077	2.876833	0.638214	4.50763	6.56E-06	0.00281
GPX8_13313	153.058	-1.14464	0.323939	3.5335	0.00041	0.045253

Supplementary Table 3 - HepaRG Doxorubicin 0.04uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
CYP11B1_12735	1.53971	-17.1058	3.684528	4.64261	3.44E-06	0.012884
EI24_15156	106.181	1.560374	0.325608	4.79E+00	1.65E-06	0.007414
SERPINB12_15476	2.24729	-17.8041	3.003874	5.93E+00	3.08E-09	1.73E-05

OR51S1_19220	5.50985	-19.6494	2.767616	7.10E+00	1.25E-12	1.40E-08
AC011525.2_20202	3.64461	-18.4095	3.002714	6.13094	8.74E-10	6.54E-06
KRTAP10-3_27410	4.59055	-18.9629	2.385924	7.94781	1.90E-15	4.27E-11

Supplementary Table 4 - HepaRG Doxorubicin 0.008uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
DEFB103A_11955	11.948	-20.7024	2.300407	9.00E+00	2.27E-19	1.27E-15
CHST10_12175	3.9388	-18.749	3.046171	6.15E+00	7.51E-10	1.13E-06
INSYN2B_12783	3.8112	-19.5885	2.96E+00	6.61274	3.77E-11	7.06E-08
CYP11B1_12735	1.5397	-17.5226	3.684528	4.75572	1.98E-06	0.002614
DHRS2_15289	61.309	-3.7339	8.62E-01	4.32956	1.49E-05	0.018651
OR51S1_19220	5.5098	-20.257	2.77E+00	7.3193	2.49E-13	6.22E-10
AC011525.2_20202	3.6446	-19.0866	3.00E+00	6.35644	2.06E-10	3.57E-07
OR6K3_23328	4.4097	-19.0838	2.832289	6.73796	1.61E-11	3.28E-08
TRIM64_23528	8.3123	-20.3978	2.453131	8.31502	9.17E-17	3.44E-13
PRAMEF13_26610	14.130	-21.4023	2.604811	8.21643	2.10E-16	6.73E-13
SSX7_28872	6.2241	-19.813	2.160241	9.17168	4.66E-20	3.49E-16
TRBV6-2_87654	6.0211	-20.0597	2.30E+00	-8.72481	2.67E-18	1.20E-14
TRBV27_88690	7.3410	-19.6724	2.51E+00	-7.83116	4.83E-15	1.36E-11
IGHV3OR16-12_88869	5.8875	-18.4634	2.937635	-6.28514	3.28E-10	5.26E-07
PRAMEF13_90473	23.032	-20.8498	2.24E+00	-9.30585	1.33E-20	1.49E-16
KRTAP4-12_91049	7.2800	-19.2624	2.724716	-7.06951	1.55E-12	3.49E-09
GLIPR1L1_91181	3.8458	-18.7472	3.35E+00	-5.58963	2.28E-08	3.20E-05
FABP5_92946	19.376	-21.9984	1.91E+00	-11.5432	7.99E-31	1.80E-26

Supplementary Table 5 - HepaRG Niacinamide 8000uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
TOP2A_7277	478.998	-1.17198	2.90E-01	4.05E+00	5.17E-05	0.009957
CYP2E1_1722	218.219	3.570769	7.39E-01	4.83E+00	1.37E-06	0.00082
MAPRE1_4017	273.026	-7.78E-01	0.224648	3.46E+00	5.37E-04	0.030212
ORM1_4833	11786.6	1.960275	0.543065	3.61E+00	0.000307	0.021491
PLP2_5208	369.057	-1.05E+00	0.316313	3.33E+00	0.000871	0.038919
RNASEH2A_5875	260.757	-1.16E+00	0.308225	3.75E+00	1.74E-04	0.016038
TMSB10_7224	4437.71	-1.24E+00	0.338329	3.68E+00	2.35E-04	0.020143
VTN_7690	924.518	1.830714	0.553661	3.31E+00	9.44E-04	0.039224
ALDH2_225	1919.88	1.191075	0.223903	5.32E+00	1.04E-07	0.00033
ALB_217	17959.5	2.933615	0.734278	4.00E+00	6.46E-05	0.010753
CFB_872	5624.49	1.907168	0.414635	4.60E+00	4.23E-06	0.001793
TMEM263_820	456.149	-6.42E-01	0.187034	3.43E+00	6.01E-04	0.031833
CALR_943	19931.3	-1.22E+00	0.340933	3.58E+00	3.43E-04	0.022699
CAT_999	930.547	1.055826	0.213152	4.95E+00	7.29E-07	0.000579
C1R_871	2008.82	8.07E-01	0.228502	3.53E+00	4.10E-04	0.026053
CDK1_1196	924.811	-0.98139	0.267513	3.67E+00	2.44E-04	0.020393
DDIT4_1803	242.428	-9.52E-01	0.290869	3.27E+00	1.06E-03	0.041215
ERO1A_2227	3423.65	-6.71E-01	0.162195	4.13E+00	3.55E-05	0.007524
FGB_2397	3454.24	2.35E+00	0.68296	3.44434	0.000572	0.031357
NUSAP1_4789	1507.60	-7.16E-01	0.181237	3.95336	7.71E-05	0.011387

Supplementary Table 6 - HepaRG Niacinamide 1600uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CETN1_23919	3.851851	-1.85E+01	2.78E+00	6.64E+00	3.04E-11	6.83E-07
GABPA_92886	120.2599	-1.53E+00	3.30E-01	4.64E+00	3.48E-06	0.039133

Supplementary Table 7 - HepG2 Doxorubicin 1uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
GLRX_2674	1.01E+03	1.46E+00	1.39E-01	1.05E+01	9.99E-26	2.01E-24
STXBP1_6866	3.15E+01	1.45E+00	2.06E-01	7.04E+00	1.92E-12	1.55E-11
AKAP8L_196	61.6881	1.38E+00	2.52E-01	5.468016	4.55E-08	2.33E-07
CTSL_1645	562.283	7.21E-01	0.168528	4.27581	1.90E-05	6.85E-05
IGFBP4_3271	49.5608	1.75E+00	0.213574	8.185989	2.70E-16	2.95E-15
RAP1GAP_5729	1.83996	1.83E+00	5.22E-01	3.512543	0.000444	0.001256
COQ8A_114	353.700	8.80E-01	1.36E-01	6.484345	8.91E-11	6.16E-10
FDFT1_2382	125.636	-1.19E+00	1.66E-01	7.17488	7.24E-13	6.10E-12
AGRN_165	1015.58	1.80E+00	3.27E-01	5.502022	3.75E-08	1.95E-07
TSC22D3_7366	4.03577	3.16E+00	4.94E-01	6.396451	1.59E-10	1.07E-09
ERBB3_2206	858.769	-2.41E+00	1.29E-01	18.6581	1.09E-77	1.40E-75
SQLE_6736	232.902	-2.19E+00	1.60E-01	13.653	1.94E-42	8.60E-41
AFMID_157	1228.74	-1.08E+00	1.58E-01	6.82459	8.82E-12	6.70E-11
TOP2A_7277	1261.73	-6.39E-01	1.64E-01	3.8943	9.85E-05	0.000315
ALDH18A1_218	273.351	-1.90E+00	2.69E-01	7.04125	1.91E-12	1.54E-11
AARS_3	970.041	-8.64E-01	2.12E-01	4.07233	4.65E-05	0.000157
CHD1L_1283	938.872	-1.00E+00	1.17E-01	8.57202	1.02E-17	1.24E-16
GNS_2725	338.696	7.49E-01	1.49E-01	5.021646	5.12E-07	2.30E-06
AKT1_210	300.364	-1.77346	0.227763	7.78645	6.89E-15	6.83E-14
RFC5_5812	100.114	-1.28706	2.70E-01	4.77167	1.83E-06	7.57E-06

Supplementary Table 8 – HepG2 Doxorubicin 0.2uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
STXBP1_6866	3.15E+01	1.03E+00	2.09E-01	4.94E+00	7.65E-07	8.24E-06
IGFBP4_3271	4.96E+01	1.35E+00	2.15E-01	6.28E+00	3.40E-10	6.49E-09
FDFT1_2382	1.26E+02	-5.95E-01	0.15697	3.79E+00	1.50E-04	0.000924
AGRN_165	1.02E+03	0.861009	0.32721	2.63E+00	8.50E-03	0.029119
COPS3_1496	6.93E+02	-0.75535	0.263291	2.87E+00	4.12E-03	0.015984
SQLE_6736	2.33E+02	-1.15E+00	0.145954	-7.89707	2.86E-15	1.04E-13

TOP2A_7277	1.26E+03	-1.06E+00	0.164377	6.48E+00	9.34E-11	1.96E-09
GNS_2725	3.39E+02	6.17E-01	0.14915	4.14E+00	3.48E-05	0.000254
APOE_358	4.21E+02	8.53E-01	0.201148	4.24E+00	2.24E-05	0.000172
RFC5_5812	1.00E+02	-9.01E-01	0.264401	3.41E+00	6.56E-04	0.003347
CDK4_1203	4.18E+03	-7.61E-01	0.139307	5.46E+00	4.69E-08	6.30E-07
STOML2_6858	1.14E+02	-9.19E-01	0.22657	4.06E+00	5.01E-05	0.000352
IER3_3214	3.21E+02	8.26E-01	0.101898	8.11E+00	5.15E-16	2.01E-14
PI4KA_5128	6.43E+02	5.81E-01	0.172422	3.37E+00	0.000745	0.003731
C3_886	2.10E+03	6.83E-01	0.128866	5.30E+00	1.15E-07	1.43E-06
ATF3_499	7.48E+00	1.50E+00	0.367699	4.09E+00	4.34E-05	0.000309
RNPS1_5895	3.24E+03	-6.47E-01	0.14981	4.32E+00	1.56E-05	0.000124
CDC25A_1158	1.63E+02	-1.31E+00	0.189285	6.91E+00	4.76E-12	1.18E-10
AZIN1_593	95.94268	-1.0135	0.250859	4.04E+00	5.34E-05	0.000372
CCNB1IP1_1054	993.1013	-6.47E-01	0.121392	5.33E+00	9.83E-08	1.24E-06

Supplementary Table 9 - HepG2 Doxorubicin 0.04uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
ADSL_153	1.37E+01	-1.08E+00	3.42E-01	3.15E+00	1.64E-03	0.02417
APOO_361	1.74E+02	-6.20E-01	1.81E-01	3.43E+00	5.94E-04	0.011609
DDB2_1798	2.71E+02	1.224207	1.94E-01	6.32E+00	2.61E-10	3.78E-08
CHERP_1290	209.650	-0.66184	1.72E-01	3.86E+00	0.000114	0.003482
GABPB1_2555	641.222	-0.81465	2.23E-01	3.65E+00	0.000259	0.006313
HSPG2_3153	6.95E+0	0.753843	0.165381	4.56E+00	5.16E-06	0.000279
HIST1H4J_2959	3.67E+02	-2.49449	5.18E-01	4.82E+00	1.45E-06	9.53E-05
HIST1H3H_2957	2.19E+03	-2.91894	2.32E-01	1.26E+01	3.27E-36	6.84E-33
HIST1H2BH_2954	1.89E+01	-1.87523	4.78E-01	3.93E+00	8.66E-05	0.002776
HIST1H1C_2950	1.78E+03	-2.27769	2.03E-01	1.12E+01	2.60E-29	2.92E-26
HES1_2925	2.24E+02	0.730899	1.72E-01	4.25E+00	2.10E-05	0.000881
HIST1H4E_2958	1.52E+03	-1.80561	2.39E-01	7.54E+00	4.54E-14	1.09E-11
HIST1H2BM_2956	3.76E+02	-3.51783	4.49E-01	7.83E+00	4.87E-15	1.29E-12
HS2ST1_3104	2.41E+01	-0.83022	2.89E-01	2.86927	0.004114	0.046036
HIST1H2BG_2953	4.12E+01	-2.24454	4.72E-01	4.75E+00	2.01E-06	0.000126
LRPAP1_3876	5.73E+01	0.724031	2.49E-01	2.91E+00	0.003583	0.041583

LPIN3_3865	2.23E+01	1.00587	2.67E-01	3.76E+00	0.000168	0.004653
LAMA3_3717	1.59E+01	1.549756	2.91E-01	5.33E+00	9.79E-08	8.70E-06
LPAR2_3854	11.5139	0.948153	3.04E-01	3.12E+00	0.001802	0.025912
NFKB2_4567	2.86E+01	0.728765	2.15E-01	3.39E+00	6.97E-04	0.012916

Supplementary Table 10 - HepG2 Doxorubicin 0.00032uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
IGFBP4_3271	4.96E+01	9.95E-01	2.19E-01	4.55E+00	5.34E-06	0.00034
BCL2L2_665	2.07E+02	8.05E-01	2.32E-01	3.47E+00	0.000516	0.007183
ALDH18A1_218	273.3511	8.98E-01	2.60E-01	3.46E+00	5.44E-04	0.007389
ECH1_2022	41.36794	1.39E+00	4.06E-01	3.42775	0.000609	0.007861
GNS_2725	338.6962	-7.14E-01	1.54E-01	-4.64579	3.39E-06	0.000259
APOE_358	420.6231	1.033298	2.01E-01	5.14E+00	2.77E-07	5.15E-05
RFC5_5812	100.1146	-7.52E-01	2.64E-01	-2.85E+00	4.43E-03	0.028518
AGPAT2_163	21.81578	-7.69E-01	2.87E-01	-2.68E+00	7.34E-03	0.039641
ELAC2_2114	32.2519	8.48E-01	2.25E-01	3.76E+00	0.000168	0.003545
CHIC2_1295	126.4555	-7.47E-01	1.86E-01	-4.02E+00	5.73E-05	0.001786
TMEM230_7197	1472.313	-5.92E-01	1.54E-01	-3.85E+00	0.00012	0.002844
AZIN1_593	95.94268	-6.37E-01	2.48E-01	-2.57E+00	1.03E-02	0.049342
BZW2_805	202.2757	-1.00E+00	3.75E-01	-2.67E+00	7.56E-03	0.040405
BAG6_630	1250.878	6.61E-01	0.167598	3.941557	8.10E-05	0.002243
BMI1_732	7.46865	-1.85E+00	5.82E-01	-3.18E+00	0.001489	0.014024
AMD1_247	743.0572	-8.15E-01	2.20E-01	-3.70023	0.000215	0.004159
ADGRE5_1154	18.76131	-1.61E+00	3.29E-01	-4.89737	9.71E-07	0.000113
CD40_1129	7.830089	1.54E+00	5.23E-01	2.95E+00	3.16E-03	0.022649
CCDC130_1019	55.97117	5.91E-01	2.31E-01	2.562345	0.010397	0.049754
CIC_1347	5.822303	2.49E+00	5.15E-01	4.83E+00	1.40E-06	0.000138

Supplementary Table 11 - HepG2 Niacinamide 60000uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
IGFBP4_3271	4.96E+01	9.95E-01	2.19E-01	4.55E+00	5.34E-06	0.00034
BCL2L2_665	2.07E+02	8.05E-01	2.32E-01	3.47E+00	0.000516	0.007183
ALDH18A1_218	273.3511	8.98E-01	2.60E-01	3.46E+00	5.44E-04	0.007389
ECH1_2022	41.36794	1.39E+00	4.06E-01	3.42775	0.000609	0.007861
GNS_2725	338.6962	-7.14E-01	1.54E-01	-4.64579	3.39E-06	0.000259
APOE_358	420.6231	1.033298	2.01E-01	5.14E+00	2.77E-07	5.15E-05
RFC5_5812	100.1146	-7.52E-01	2.64E-01	-2.85E+00	4.43E-03	0.028518

AGPAT2_163	21.81578	-7.69E-01	2.87E-01	-2.68E+00	7.34E-03	0.039641
ELAC2_2114	32.2519	8.48E-01	2.25E-01	3.76E+00	0.000168	0.003545
CHIC2_1295	126.4555	-7.47E-01	1.86E-01	-4.02E+00	5.73E-05	0.001786
TMEM230_7197	1472.313	-5.92E-01	1.54E-01	-3.85E+00	0.00012	0.002844
AZIN1_593	95.94268	-6.37E-01	2.48E-01	-2.57E+00	1.03E-02	0.049342
BZW2_805	202.2757	-1.00E+00	3.75E-01	-2.67E+00	7.56E-03	0.040405
BAG6_630	1250.878	6.61E-01	0.167598	3.941557	8.10E-05	0.002243
BMI1_732	7.46865	-1.85E+00	5.82E-01	-3.18E+00	0.001489	0.014024
AMD1_247	743.0572	-8.15E-01	2.20E-01	-3.70023	0.000215	0.004159
ADGRE5_1154	18.76131	-1.61E+00	3.29E-01	-4.89737	9.71E-07	0.000113
CD40_1129	7.830089	1.54E+00	5.23E-01	2.95E+00	3.16E-03	0.022649
CCDC130_1019	55.97117	5.91E-01	2.31E-01	2.562345	0.010397	0.049754
CIC_1347	5.822303	2.49E+00	5.15E-01	4.83E+00	1.40E-06	0.000138

Supplementary Table 12 - HepG2 Niacinamide 12000uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
APOE_358	3.90E+02	-8.56E-01	2.10E-01	4.08E+00	4.48E-05	0.006161
AGPAT2_163	2.05E+01	-1.38E+00	3.11E-01	4.44598	8.75E-06	0.001721
CDH1_1186	7.61E+01	-1.64E+00	2.55E-01	6.45E+00	1.14E-10	2.32E-07
HMOX1_3041	3.86E+02	-8.21E-01	0.180815	4.54002	5.62E-06	0.001285
PDLIM1_5041	7.54E+02	-7.80E-01	0.171542	4.54E+00	5.50E-06	0.001279
ORM1_4833	1252.23	-9.81E-01	2.54E-01	3.86E+00	1.14E-04	0.012539
UBE2L6_7483	2.95E+02	-5.93E-01	1.51E-01	3.92E+00	9.03E-05	0.010659
APOC3_356	4.22E+02	-1.40E+00	1.61E-01	8.70E+00	3.22E-18	1.52E-14
PDGFB_5026	3.90E+01	-7.90E-01	1.99E-01	3.96959	7.20E-05	0.008867
TP53INP2_7291	2.91E+02	-5.86E-01	1.58E-01	3.71E+00	0.000211	0.018674
VTN_7690	1.95E+02	-1.12E+00	1.51E-01	7.43066	1.08E-13	3.83E-10
HP_3085	8.96E+02	-1.28E+00	1.44E-01	8.93E+00	4.20E-19	2.97E-15
SERPINE1_6253	1.18E+03	-7.71E-01	2.01E-01	3.84E+00	0.000123	0.013066
MYOF_10516	1.75E+01	-1.37731	0.352972	3.90E+00	9.54E-05	0.010984
CYP1A1_10775	1.53E+02	1.35E+00	2.12E-01	6.368191	1.91E-10	3.39E-07

SERPINC1_11069	6.72E+02	-5.97E-01	0.111527	5.35216	8.69E-08	5.13E-05
ISG20_11135	4.11E+01	-9.03E-01	0.253235	3.56E+00	0.000365	0.02797
LGALS7B_11147	1.06E+01	-2.26E+00	6.31E-01	3.59E+00	0.000329	0.026142
COL16A1_11319	4.87E+02	-1.01E+00	0.237194	4.24164	2.22E-05	0.003698
IDNK_11284	2.64E+01	-1.20E+00	2.92E-01	4.13E+00	3.62E-05	0.005229

Supplementary Table 13 - HepG2 Niacinamide 2400uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
AGPAT2_163	2.05E+01	-1.34E+00	3.11E-01	4.32E+00	1.58E-05	0.00448
ELAC2_2114	3.33E+01	8.93E-01	0.218807	4.08E+00	4.52E-05	0.007218
ADGRE5_1154	1.86E+01	-9.14E-01	2.82E-01	3.24E+00	1.18E-03	0.043553
CDH1_1186	7.61E+01	-1.34004	0.249507	5.37E+00	7.84E-08	0.000271
DPM2_1947	3.92E+01	1.136696	3.27E-01	3.48E+00	0.000503	0.029628
DYNLL1_1997	9.48E+02	-1.24E+00	3.48E-01	3.55E+00	0.000389	0.02554
MYLK_4414	2.71E+01	-1.07E+00	3.12E-01	3.44E+00	0.00059	0.031862
KLHDC3_3631	2.10E+02	8.35E-01	2.03E-01	4.11E+00	4.00E-05	0.00688
MCEE_4067	4.36E+01	-9.15E-01	0.215375	4.25E+00	2.16E-05	0.005114
MT2A_4334	3.61E+02	1.71E+00	5.22E-01	3.283308	0.001026	0.041364
NT5DC2_4737	6.39E+00	1.17E+00	0.358154	3.28E+00	1.06E-03	0.041753
PDLIM1_5041	7.54E+02	-6.44E-01	1.71E-01	3.76E+00	1.70E-04	0.016976
PAIP1_4909	8.27E+01	-7.41E-01	2.10E-01	3.5225	0.000427	0.026677
TRIB3_7337	2713.657	-0.67079	2.02E-01	3.33E+00	0.000875	0.038648
ELMO3_2123	1.05E+01	1.24E+00	0.321613	3.86E+00	1.16E-04	0.013862
PPARA_5300	4.35E+01	-0.63815	0.195618	3.26E+00	1.11E-03	0.042393
TMEM184B_7192	2.02E+01	0.710841	2.14E-01	3.321323	0.000896	0.03872
ERCC8_2220	1.94E+02	-8.55E-01	2.35E-01	3.64498	0.000267	0.020536
APOC3_356	4.22E+02	-0.61615	0.158243	3.89E+00	9.87E-05	0.012788
TMC7_7154	1.62E+01	-1.32E+00	3.62E-01	3.64E+00	0.000277	0.020858

Supplementary Table 14 – HepG2 Niacinamide 480uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
APOC3_356	4.22E+02	-7.42E-01	1.59E-01	-4.68E+00	2.91E-06	0.032645
APOL1_20931	8.32E+02	1.450616	2.68E-01	5.42E+00	5.96E-08	0.001336

Supplementary Table 15 - HepG2 Niacinamide 96uM

gene	base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
CPNE1_1529	1.46E+03	-6.03E-01	1.53E-01	3.94E+00	8.06E-05	0.0276
TMEM203_7196	50.09829	-1.27E+00	2.59E-01	4.91E+00	9.29E-07	0.0038
TP53INP2_7291	2.91E+02	-6.93E-01	1.59E-01	4.3677	1.26E-05	0.0167
ABCE1_32	211.5218	-6.86E-01	1.65E-01	4.1476	3.36E-05	0.0212
ARHGEF2_423	8.92E+01	5.95E-01	0.165	3.5951	0.000324	0.0499
CRYZ_1586	2.18E+02	-8.01E-01	0.193	4.1449	3.40E-05	0.0212
KPNA3_3672	7.94E+01	-9.04E-01	0.243	3.7148	0.000203	0.0416
MSMO1_4315	5.56E+03	-6.83E-01	1.88E-01	3.6266	0.000287	0.0499
PCBP2_4957	921.751	6.80E-01	1.75E-01	3.8949	9.82E-05	0.0287
POT1_5287	6.77E+0	0.980593	0.230	4.2458	2.18E-05	0.0197
SF3B1_6268	893.622	-7.18E-01	1.80E-01	-3.993	6.52E-05	0.0256
TATDN2_6963	3.10E+01	9.12E-01	2.31E-01	3.9426	8.06E-05	0.0276
USO1_7576	2.62E+02	-0.82798	0.2112	3.9196	8.87E-05	0.0287
RALGAPA2_10829	6.02E+01	-9.35E-01	0.2607	3.5878	0.000333	0.0499
LRCH4_11541	20.8846	1.14E+00	3.02E-01	3.7805	0.000156	0.0375
SLC7A6OS_11998	350.579	-8.87E-01	2.19E-01	4.0472	5.18E-05	0.0230
TMEM169_12325	2.31E+01	0.974671	0.2552	3.8181	0.000134	0.0350
ACADSB_13280	5.26E+02	-0.68797	0.1711	4.0201	5.82E-05	0.0240
RAB11FIP5_13349	116.6166	6.24E-01	1.53E-01	4.0868	4.37E-05	0.0212
TMEM131_13546	5.18E+01	1.61E+00	0.389	4.1242	3.72E-05	0.0212

Supplementary Table 16 - HepG2 Niacinamide 19.2uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
COQ9_11694	2.93E+02	-1.68E+00	3.43E-01	4.90E+00	9.39E-07	0.01053
APOL1_20931	8.32E+02	1.54E+00	2.68E-01	5.77E+00	8.11E-09	0.000182

Supplementary Table 17 - HepG2 Niacinamide 3.84uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
APOL1_20931	8.32E+02	1.48E+00	2.68E-01	5.51E+00	3.50E-08	0.000784

Supplementary Table 18 – MCF-7 Doxorubicin 1uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CHAC1_1279	3.19E+01	2.06E+00	2.77E-01	7.44E+00	1.03E-13	8.88E-13
GLRX_2674	5.17E+01	8.97E-01	1.81E-01	4.944021	7.65E-07	3.32E-06
MEFV_4116	5.431577	2.74E+00	0.427015	6.428306	1.29E-10	8.53E-10
AKAP8L_196	35.53614	1.73E+00	0.218079	7.952083	1.83E-15	1.81E-14
CTSL_1645	752.1987	2.431977	0.188362	12.9112	3.89E-38	1.25E-36
RAP1GAP_5729	1.784273	1.77E+00	0.658025	2.685288	0.007247	0.016241
COQ8A_114	109.9445	1.287839	0.114887	11.20963	3.66E-29	7.84E-28
AGRN_165	2463.815	1.029802	0.358162	2.875241	0.004037	0.009624
ABCG1_37	4.357605	1.56E+00	0.477774	3.265766	0.001092	0.00294
COPS3_1496	1485.223	-7.88E-01	0.263966	-2.98634	0.002823	0.006944
TSC22D3_7366	523.9508	2.09E+00	0.149985	13.91652	5.03E-44	2.04E-42
ERBB3_2206	1051.834	-8.27E-01	0.119058	-6.94814	3.70E-12	2.81E-11
SQLE_6736	372.5088	-1.73E+00	0.141966	-12.1809	3.93E-34	1.05E-32
MUC1_4366	90.27375	-1.66656	0.202262	-8.2396	1.73E-16	1.83E-15
TOP2A_7277	571.08	-2.45312	0.147268	-16.6576	2.67E-62	1.90E-60
ALDH18A1_218	70.60763	-9.50E-01	0.270824	-3.50737	0.000453	0.0013
AARS_3	1245.21	-1.44637	0.174272	-8.29951	1.05E-16	1.13E-15
RPS5_6011	1640.987	-6.81E-01	0.287237	-2.36953	0.017811	0.036328
CHD1L_1283	568.6987	-1.11E+00	0.110605	-10.0412	1.00E-23	1.66E-22
GNS_2725	580.2582	1.52E+00	0.193846	7.8352	4.68E-15	4.48E-14

Supplementary Table 19 - MCF-7 Doxorubicin 0.2uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
AKAP8L_196	3.55E+01	7.92E-01	2.24E-01	3.53E+00	4.12E-04	0.004254
CTSL_1645	7.52E+02	5.94E-01	0.1890	3.14E+00	1.67E-03	0.01383
COPS3_1496	1.49E+03	-0.91464	0.2639	3.47E+00	5.30E-04	0.005251
TSC22D3_7366	5.24E+02	1.185531	0.1504	7.88E+00	3.28E-15	2.72E-13
MUC1_4366	90.27375	-1.41522	0.1921	7.37E+00	1.74E-13	1.17E-11

AARS_3	1.25E+03	-0.58011	0.1729	3.35467	0.000795	0.007409
GNS_2725	580.258	0.739432	0.1942	3.81E+00	1.40E-04	0.001678
RFC5_5812	155.511	-1.15692	0.2042	5.66292	1.49E-08	4.80E-07
CDK4_1203	7.41E+03	-0.82674	0.1395	5.92413	3.14E-09	1.15E-07
ACD_60	6.42E+01	-0.61353	0.1630	3.76192	0.000169	0.001976
CDC25A_1158	7.81E+01	-1.37636	0.2578	5.34E+00	9.42E-08	2.65E-06
CD320_1116	9.24E+01	-0.85037	0.1727	4.92E+00	8.52E-07	1.93E-05
CBR3_1007	1.29E+02	-0.75863	0.1274	5.95E+00	2.64E-09	9.81E-08
CDCA5_1184	159.497	-1.65034	0.2906	5.68E+00	1.36E-08	4.44E-07
CCND1_1062	197.579	0.823447	0.2679	3.07E+00	2.12E-03	0.016891
AURKB_587	4.35E+02	-0.90396	0.2450	3.68836	0.000226	0.00254
APOO_361	296.546	-0.73083	0.1329	5.50E+00	3.84E-08	1.14E-06
CGA_1272	1.19E+01	1.615526	0.564	2.864409	0.004178	0.028827
CDK2_1202	4.32E+02	-0.8603	0.0875	9.83E+00	8.24E-23	1.38E-20
DEPDC1B_1847	6.38E+02	-1.18129	0.1507	7.84E+00	4.64E-15	3.81E-13

Supplementary Table 20 - MCF-7 Doxorubicin 0.04uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
RFC5_5812	1.56E+02	-7.00E-01	2.01E-01	-3.48E+00	0.0005	0.02586
AURKB_587	4.35E+02	-0.82439	2.45E-01	-3.37E+00	0.000765	0.034851
CKLF_1359	127.2494	-0.71994	2.20E-01	-3.27E+00	0.001061	0.044405
FEN1_2387	217.5066	-0.60437	1.26E-01	-4.80E+00	1.57E-06	0.000324
HIST1H4J_2959	150.9058	-1.81067	5.12E-01	-3.54E+00	4.05E-04	0.022325
HIST1H3H_2957	1255.967	-1.0643	0.229977	-4.62786	3.69E-06	0.000633
HIST1H2BH_2954	140.7993	-1.15347	2.92E-01	-3.95E+00	7.68E-05	0.006389

HIST1H1C_2950	1135.832	-0.99287	0.182294	-5.45E+00	5.14E-08	2.21E-05
HIST1H4E_2958	974.6977	-1.02962	0.216175	-4.76E+00	1.91E-06	0.000384
HIST1H2BM_2956	824.9539	-1.73839	0.38178	-4.55339	5.28E-06	0.00082
HIST1H2BG_2953	38.73987	-1.42413	4.31E-01	-3.30E+00	9.57E-04	0.040949
SPC25_6704	410.7606	-0.62989	1.15E-01	-5.48E+00	4.34E-08	2.04E-05
SOX2_6683	80.52757	-0.76009	1.96E-01	-3.87E+00	1.09E-04	0.008442
GDF15_2621	253.8098	0.960226	1.89E-01	5.07E+00	4.02E-07	0.000135
GREB1_2778	488.13	0.761046	2.30E-01	3.31E+00	0.000929	0.040421
BBC3_640	51.01453	0.728915	0.220063	3.312301	0.000925	0.040406
H2AFX_2861	4662.36	-0.76677	1.70E-01	-4.51E+00	6.38E-06	0.000907
GADD45G_2571	34.98628	0.652624	0.187621	3.478416	0.000504	0.025862
SYTL1_6923	100.6078	0.744213	1.62E-01	4.61E+00	4.10E-06	0.000687
HIST1H4K_2960	1739.456	-1.51227	3.24E-01	-4.66E+00	3.14E-06	0.000556

Supplementary Table 21 - MCF-7 Doxorubicin 0.0016uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
HRAS_3101	1.40E+03	-8.20E-01	1.63E-01	-5.024	5.06E-07	0.002269

Supplementary Table 22 - MCF-7 Niacinamide 60000uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CHAC1_1279	3.31E+01	2.96E+00	0.29401	1.01E+01	7.19E-24	1.68E-22
MEFV_4116	6.42E+00	3.81E+00	0.448476	8.496452	1.95E-17	2.93E-16
STXBP1_6866	1.27E+02	1.34E+00	0.133069	10.07775	6.93E-24	1.63E-22
AKAP8L_196	3.31E+01	1.21E+00	2.25E-01	5.403116	6.55E-08	3.59E-07
IGFBP4_3271	1.74E+02	-1.84E+00	2.21E-01	-8.32443	8.47E-17	1.20E-15
ERBB3_2206	1022.544	-0.91039	1.38E-01	-6.59124	4.36E-11	3.50E-10
CLN3_1392	7.82E+00	-1.34E+00	5.05E-01	-2.65203	0.008001	0.018283

MUC1_4366	88.33541	-1.96E+00	2.19E-01	-8.95E+00	3.66E-19	6.39E-18
ALDH18A1_218	67.71459	-1.55E+00	2.79E-01	-5.5563	2.76E-08	1.59E-07
ECH1_2022	29.64415	-1.97453	4.60E-01	-4.29241	1.77E-05	6.73E-05
GRK2_152	2.98E+02	-1.13E+00	1.34E-01	-8.40174	4.40E-17	6.40E-16
CHD1L_1283	5.63E+02	-1.00E+00	1.09E-01	-9.22E+00	2.97E-20	5.58E-19
GNS_2725	5.67E+02	9.99E-01	1.98E-01	5.048617	4.45E-07	2.16E-06
MFAP1_4142	6.09E+01	1.18E+00	2.27E-01	5.198993	2.00E-07	1.02E-06
AKT1_210	1.41E+03	-9.11E-01	0.21773	-4.18301	2.88E-05	0.000106
APOE_358	1.254733	2.676592	0.992311	2.697331	0.00699	0.016202
RFC5_5812	1.53E+02	-8.19E-01	2.02E-01	-4.0637	4.83E-05	0.000172
CDK4_1203	7122.304	-1.50375	0.140164	-10.7285	7.48E-27	2.23E-25
FOS_2461	1.31E+01	2.65E+00	3.66E-01	7.245988	4.29E-13	4.25E-12
STOML2_6858	1.60E+02	-8.68E-01	2.43E-01	-3.5736	0.000352	0.001076

Supplementary Table 23 - MCF-7 Niacinamide 12000uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
GLRX_2674	5.13E+01	1.349971	2.12E-01	6.36E+00	2.01E-10	4.44E-08
MEFV_4116	6.42E+00	1.359591	0.485396	2.80E+00	5.09E-03	0.040439
STXBP1_6866	1.27E+02	0.810678	0.135659	5.98E+00	2.29E-09	3.13E-07
IGFBP4_3271	1.74E+02	-8.08E-01	0.209837	-3.85E+00	1.18E-04	0.002431
GNS_2725	5.67E+02	8.98E-01	0.197978	4.54E+00	5.72E-06	0.000216
RFC5_5812	153.3336	-6.96E-01	0.200563	-3.47E+00	5.21E-04	0.007592
MRPS2_4292	1.69E+02	-9.43E-01	0.307065	-3.07109	0.002133	0.02159
CD320_1116	8.87E+01	-7.79E-01	1.66E-01	-4.69E+00	2.76E-06	0.000123
CBR3_1007	1.23E+02	-7.09E-01	0.14479	-4.90E+00	9.70E-07	5.11E-05
CCND1_1062	180.8657	-9.49E-01	0.280169	-3.39E+00	7.04E-04	0.00944
ALDOC_233	1.79E+01	9.57E-01	0.341151	2.81E+00	5.01E-03	0.039889
CAPN1_957	1.87E+02	-5.87E-01	1.60E-01	-3.66E+00	2.52E-04	0.004353
CYP46A1_1732	3.16E+00	1.78E+00	0.53248	3.34E+00	8.26E-04	0.010591
CGA_1272	1.51E+01	4.20E+00	0.695414	6.04E+00	1.59E-09	2.49E-07
CFLAR_1269	8.18E+01	-0.60008	0.158082	-3.80E+00	0.000147	0.002873
COL5A1_1480	44.57019	0.695042	0.190336	3.651663	0.000261	0.004456
CTSD_1642	3.03E+03	-1.13E+00	0.22872	-4.96E+00	7.20E-07	4.02E-05
CYP2E1_1722	4.703489	1.325203	0.48436	2.74E+00	6.22E-03	0.046387
CEBPD_1242	1.77E+01	-1.50E+00	0.362078	-4.14E+00	3.43E-05	0.000893
GDPD5_2628	2.20E+00	1.87E+00	0.640642	2.922542	0.003472	0.030441

Supplementary Table 24 - MCF-7 Niacinamide 2400uM

gene	Base Mean	log2Fold Change	lfcSE	stat	pvalue	padj
PSME2_5535	414.2825	-6.90E-01	1.90E-01	-3.63E+00	2.87E-04	0.04162
RASL10B_5744	20.04721	-1.0838	2.59E-01	-4.19E+00	2.79E-05	0.009287
MATK_4040	163.6744	-0.67588	1.51E-01	-4.49E+00	7.11E-06	0.003732
GJB2_2653	9.48E+01	-0.62839	1.67E-01	-3.76E+00	0.000167	0.031715
MGST3_10518	1.76E+02	0.627795	1.75E-01	3.59E+00	0.00033	0.045909
QPCT_10614	4.70E+01	0.679338	1.89E-01	3.59E+00	0.000335	0.045909
LRFN4_11101	6.30E+01	-0.63001	0.155703	-4.04622	5.21E-05	0.01531
NAALADL2_11468	3.00E+01	9.20E-01	1.92E-01	4.80E+00	1.55E-06	0.001783
TSPAN1_11441	4.00E+01	-1.05663	2.13E-01	-4.96E+00	6.99E-07	0.001331
AASS_11850	1.86E+01	1.381702	3.13E-01	4.42E+00	1.00E-05	0.004767
LIMA1_12066	1.29E+02	0.642139	1.67E-01	3.85E+00	0.000118	0.0251
DSCC1_12314	2.33E+03	-6.29E-01	1.61E-01	-3.90E+00	9.44E-05	0.020961
CDC14A_12788	3.11E+01	-0.99826	2.32E-01	-4.30E+00	1.70E-05	0.006306
KLHDC1_13148	2.81E+01	0.812348	2.12E-01	3.84E+00	1.23E-04	0.025569
MAOB_13383	13.29947	-1.5759	4.23E-01	-3.72446	0.000196	0.035233
SPHK1_13551	439.2118	-0.64929	0.126727	-5.12354	3.00E-07	0.001145
BDH2_14883	1.24E+02	0.787317	2.06E-01	3.83E+00	1.28E-04	0.02578
HMGN5_15632	96.17149	0.710803	1.78E-01	3.98E+00	6.78E-05	0.017358
HMCN1_15657	3.11E+01	0.742609	1.99E-01	3.73E+00	0.000194	0.035233
ZNF189_15740	1.83E+01	0.88826	0.248686	3.57181	0.000355	0.045909

Supplementary Table 25 - MCF-7 Niacinamide 480uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CCDC9_20830	3.11E+02	1.10E+00	1.90E-01	5.79E+00	6.92E-09	0.000156

Supplementary Table 26 - MCF-7 Niacinamide 96uM

gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CCDC9_20830	3.11E+02	9.11E-01	1.90E-01	4.79E+00	1.67E-06	0.018786
ACACA_91463	7.00E+02	-6.43E-01	1.42E-01	-4.53958	5.64E-06	0.04221
GABPA_92886	2.07E+02	-1.25E+00	2.54E-01	-4.9338	8.06E-07	0.018117

Supplementary Table 27 - MCF-7 Niacinamide 19.2uM

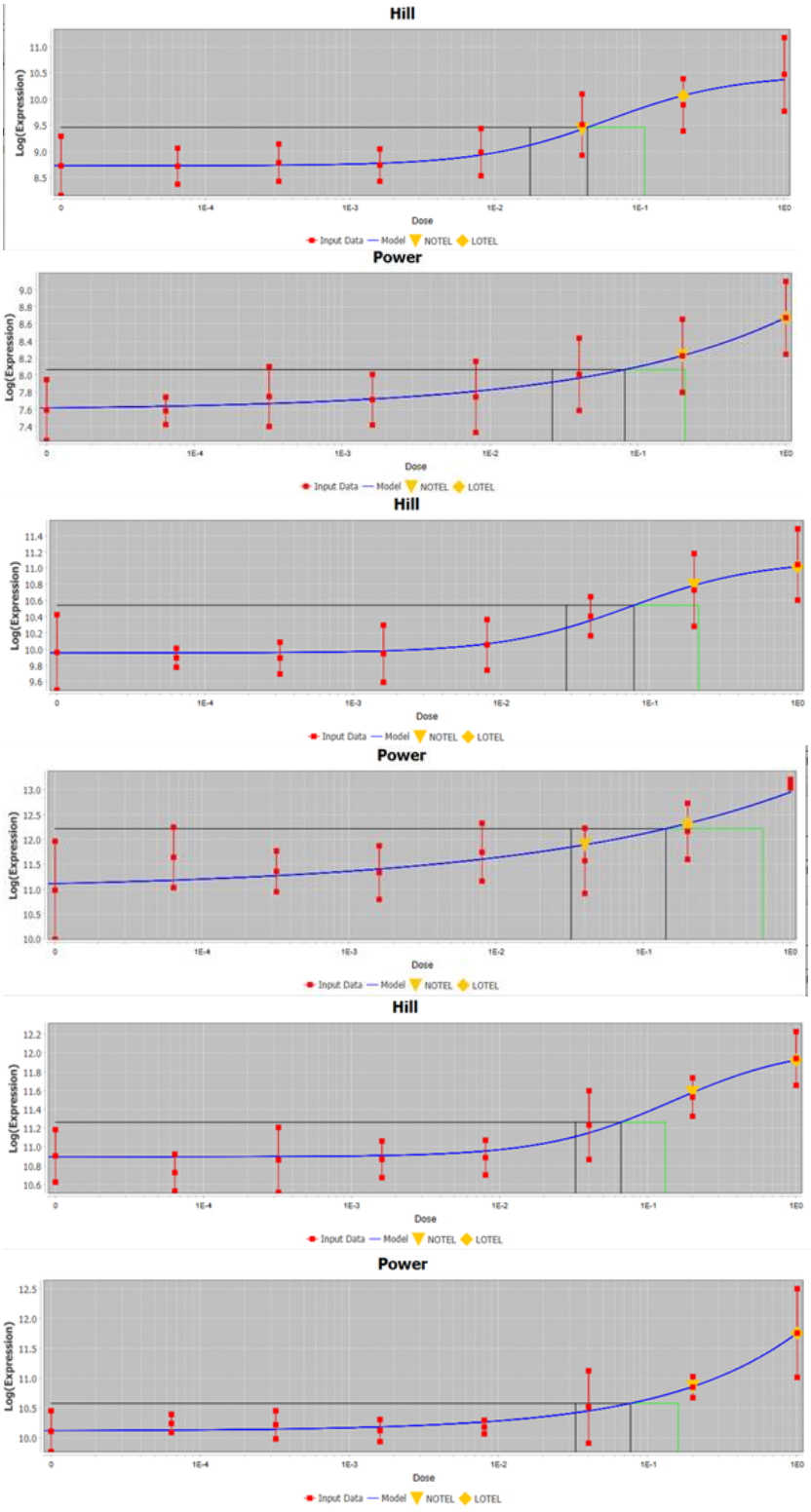
gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CCDC9_20830	3.11E+02	1.02E+00	1.90E-01	5.37E+00	8.04E-08	0.000602
ACACA_91463	7.00E+02	-0.82938	1.42E-01	-5.83E+00	5.49E-09	0.000123
GABPA_92886	2.07E+02	-1.40318	2.56E-01	-5.48564	4.12E-08	0.000463

Supplementary Table 28 - MCF-7 Niacinamide 3.84uM

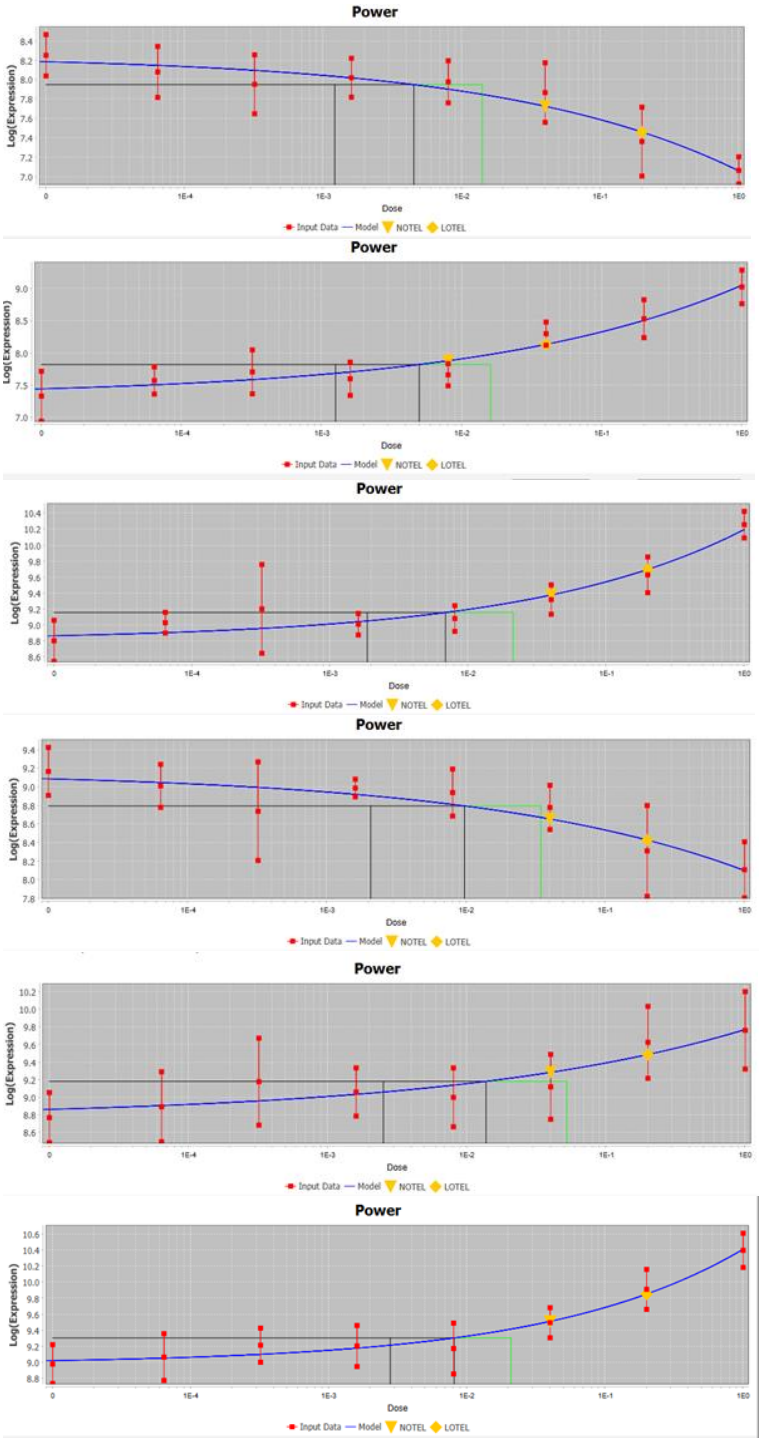
gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
CCDC9_20830	3.11E+02	9.73E-01	1.90E-01	5.11E+00	3.15E-07	0.003542
MMP10_28510	5.30E+00	-1.58E+01	1.779038	-8.89E+00	6.12E-19	1.37E-14

APPENDIX 3

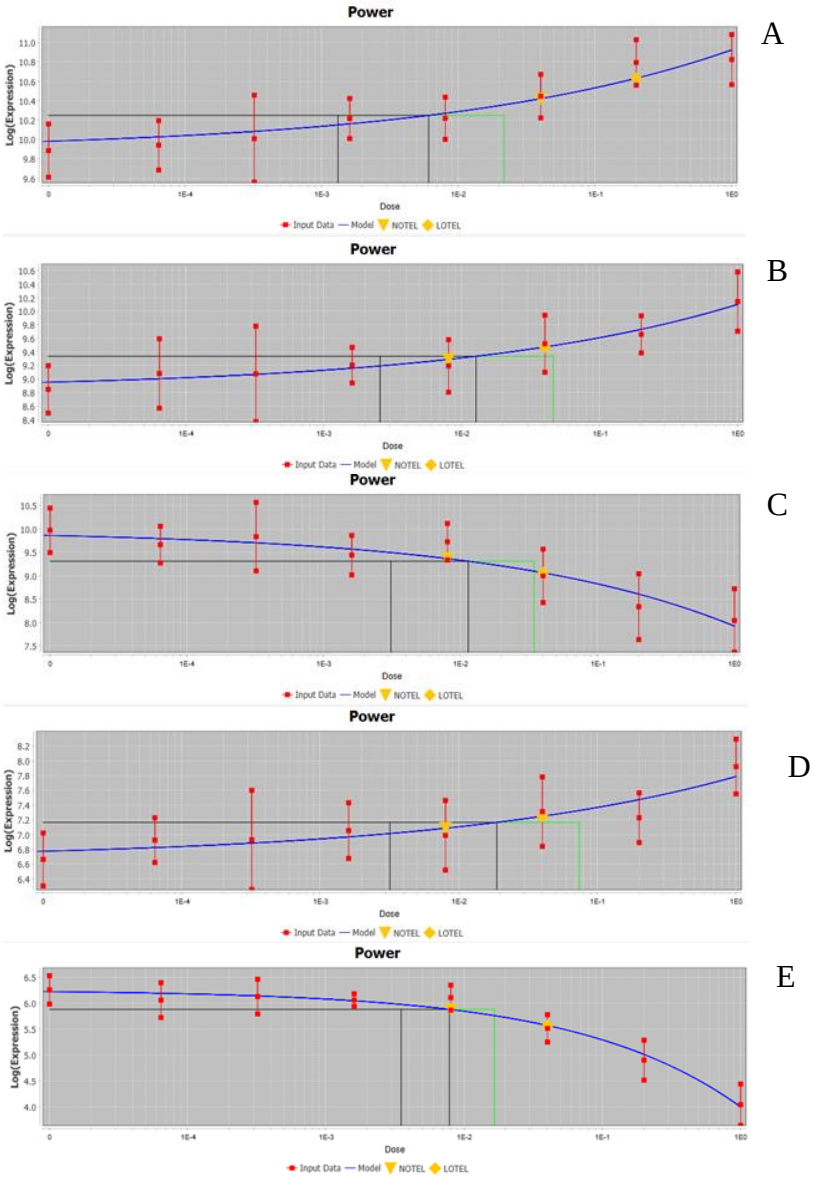
Supplementary Figures 1-6

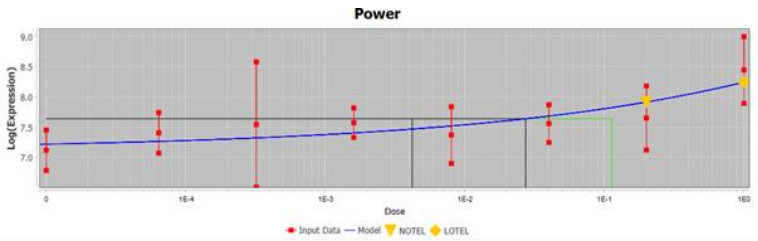


Supplementary Figure 1: The six lowest probe BMDL individual curve fits for HepaRG doxorubicin. Red points = mean and standard deviation data points, first black vertical line = BMDL, secondblack vertical line = BMD, green vertical line = BMDU, blue curve = model fit to the data. A - GDNF_24860; response – Hill; BMDL = 0.0174935; B - MAST4_13899; RESPONSE – Power; BMDL= 0.0263351; C - DSE_15124; response – Hill; BMDL = 0.0274406 ;D - AKR1B10_19908; response – Power; BMDL = 0.0323136; E - TRIM22_21719; response – Hill; BMDL = 0.0324324 ; F - TNFRSF10A; response – Power; BMDL = 0.0327428



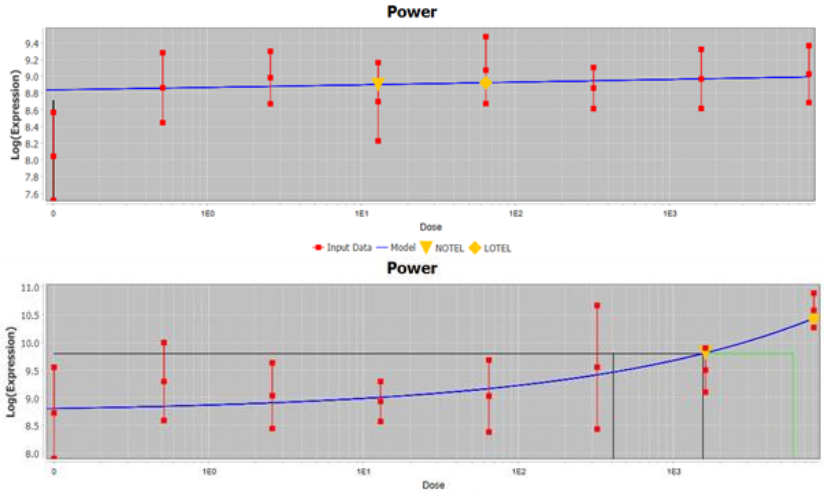
Supplementary Figure 2: The six lowest probe BMDL individual curve fits for Hep G2 doxorubicin. Red points = mean and standard deviation data points, first black vertical line = BMDL, second black vertical line = BMD, green vertical line = BMDU, blue curve = model fit to the data. A - KCTD6_27391; response – Power; BMDL = 0.00121298;; B - GOLGA8A_19056; response – Power; BMDL = 000125679; C - INPPL1_14655; response – Power; BMDL = 0.00185812;D - TOR1AIP1_10953; response – Power; BMDL = 0.0020632; E - ARF3_16792; response – Power; BMDL = 0.00251049; F - NAGK_23047; response – Power; BMDL = 0.00278434





F

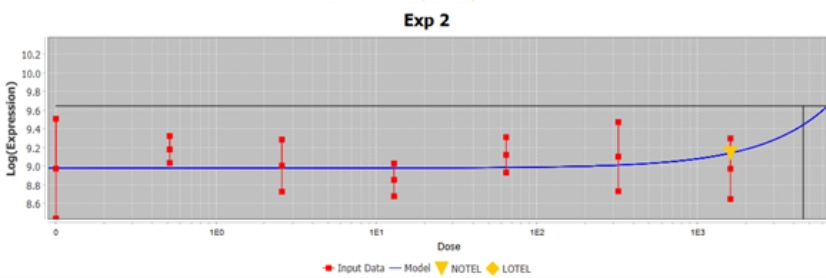
Supplementary Figure 3: The six lowest probe BMDL individual curve fits for MCF7 doxorubicin. Red points = mean and standard deviation data points, first black vertical line = BMDL, second black vertical line = BMD, green vertical line = BMDU, blue curve = model fit to the data. A - ULK1_26707; response – Power; BMDL = 0.0013132; B - SMAD3_6562; response – Power; BMDL = 0.00254904; C - HIST1H4E_2958; response – Power; BMDL = 0.00308809;D - SYTL1_6923; response – Power; BMDL = 0.00318123; E - RIBC2_10735; response – Power; BMDL = 0.00349724; F - HELZ2_10675; response – Power; BMDL = 0.00419383



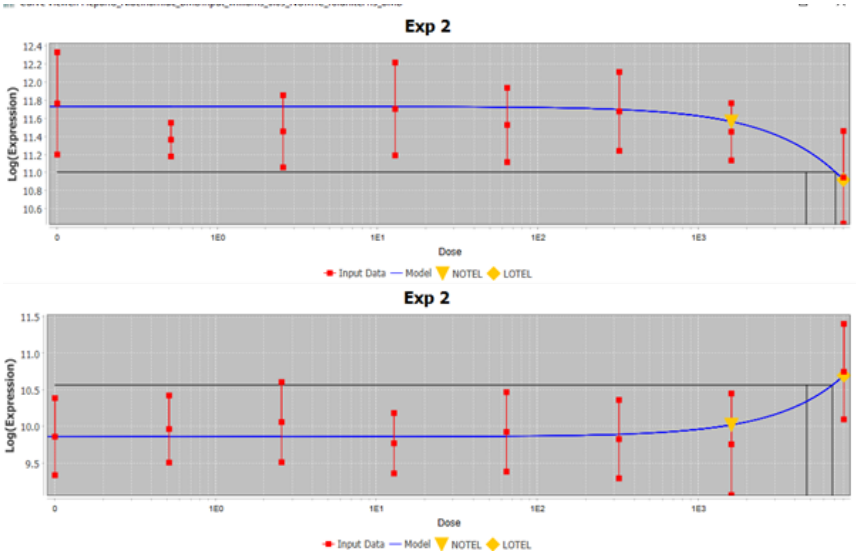
A

B

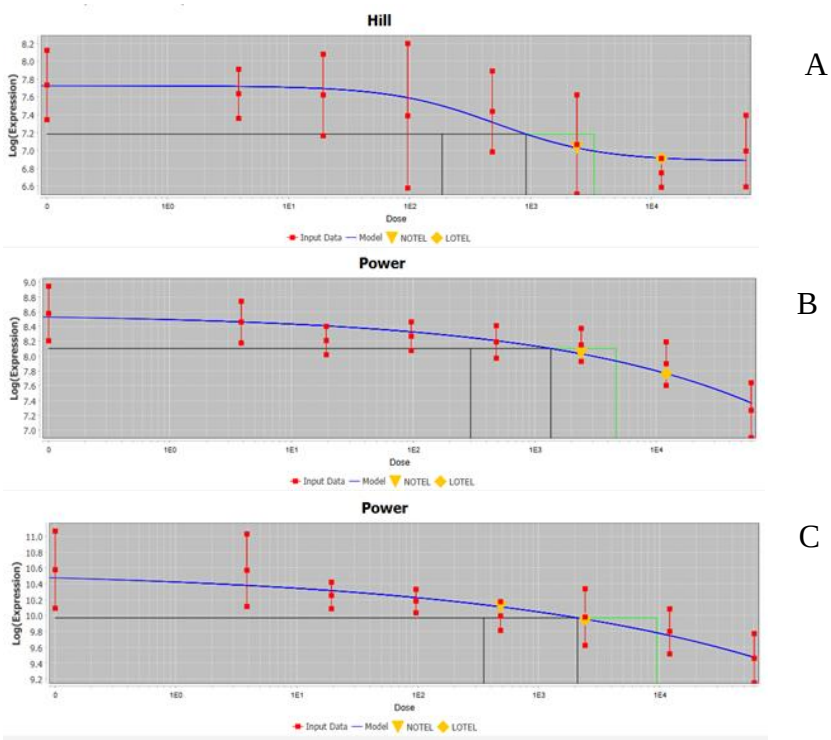
C

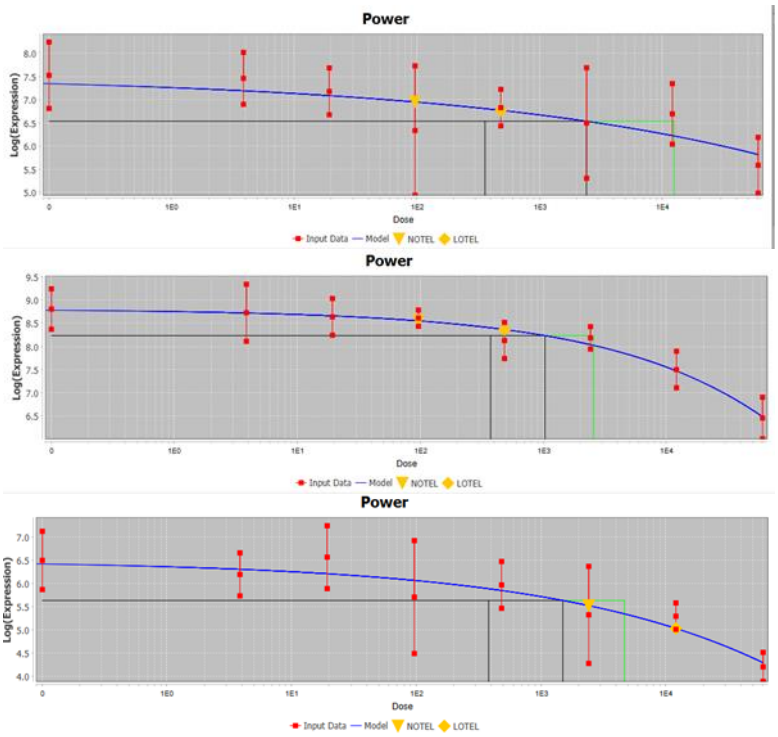


D

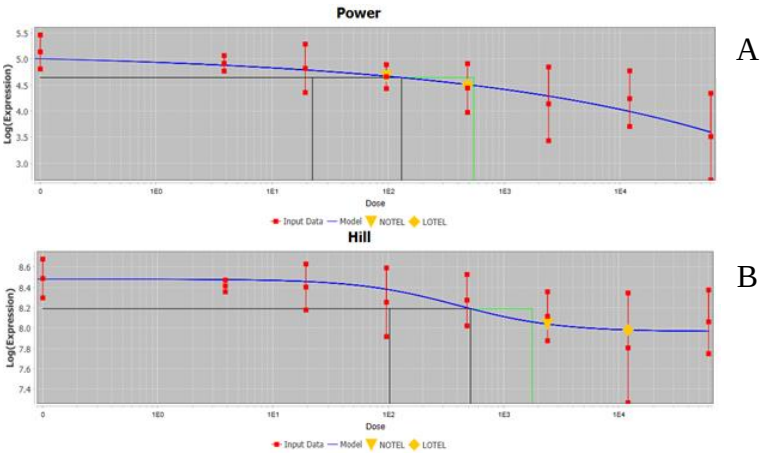


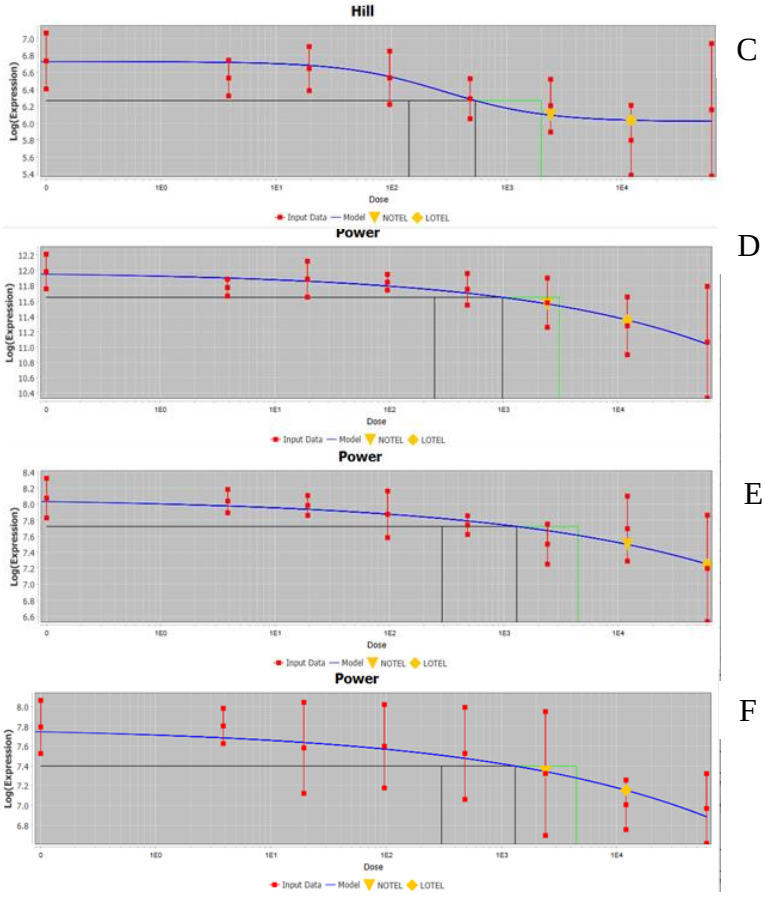
Supplementary Figure 4: The six lowest probe BMDL individual curve fits for HepaRG niacinamide. Red points = mean and standard deviation of data points, first black vertical line = BMDL, second black vertical line = BMD, green vertical line = BMDU, blue curve = model fit to the data. A - CCDC_20830; response – Power; BMDL = 1.66319e-6; B - CYP1A1_10775; response – Power; BMDL = 406.466; C - SDS_25069; response – Power; BMDL = 1044.44; D - PCK2_33867; response – Exp2; BMDL = 4542.77; E - TP5313_24981; response – Exp2; BMDL = 4680.81; F - SCP2_27826; response – Exp2; BMDL = 4699.41





Supplementary Figure 5: The six lowest probe BMDL individual curve fits for HepG2 niacinamide. Red points = mean and standard deviation data points, first black vertical line = BMDL, second black vertical line = BMD, green vertical line = BMDU, blue curve = model fit to the data. A - VTN_7690; response – Hill; BMDL = 184.541; B - FAM3B_11565; response – Power; BMDL = 295.221; C - APOC1_19115; response – Power; BMDL = 348.788; D ADH4_25100; response – Power; BMDL = 356.957; E - APOC3_356; response – Power; BMDL = 370.622; F - CDH1_1186; response – Power; BMDL = 378.298





Supplementary Figure 6: The six lowest probe BMDL individual curve fits for MCF-7 niacinamide. Red points = mean and standard deviation data points, first black vertical line = BMDL, second black vertical line = BMD, green vertical line = BMDU, blue curve = model fit to the data. A - ACOX2_19283; response – Power; BMDL = 22.1636; B - PDSS1_22903; response – Hill; BMDL = 102.175; C - TNFRSF18_20262; response – Hill; BMDL = 141.107; D - NME2_28197; response – Power; BMDL = 243.988; E - MGLL_17894; response – Power; BMDL = 287.703; F - ADRAZC_22752; response – Power; BMDL = 300.555.

APPENDIX 4

Supplementary Tables 1-6

Supplementary Table 1: Summary of top twenty enriched GO terms for HepaRG dosed with doxorubicin. Shown are for HepaRG dosed with doxorubicin the top twenty most significantly enriched genes based on the Highest Fold Change Absolute

Probe ID	Gene symbols	Best Model	Best BMD	Best BMDL	Best BMDU	Max Fold Change Value (Absolute)
SERPINA3_17487	serpina3	Linear	0.8289	0.5673	1.5379	7.8689
ITGA10_27364	itga10	Linear	0.3010	0.2537	0.3699	5.5664
S100A2_21300	s100a2	Linear	0.6615	0.4822	1.0528	4.9837
TP53I3_24981	tp53i3	Exp 4	0.1188	0.0738	0.2114	4.9545
ANGPT1_90108		Exp 2	0.3630	0.2873	0.4845	4.8920
TSC22D3_7366	tsc22d3	Exp 2	0.6848	0.4676	1.1886	4.7476
NPR3_88568		Exp 2	0.2867	0.2368	0.3603	4.7139
GDF15_18329	gdf15	Poly 2	0.1936	0.1181	0.4296	4.6885
AKR1B10_19908	akr1b10	Power	0.1422	0.0323	0.6500	4.4137
TP53I3_7290	tp53i3	Poly 2	0.1612	0.1051	0.3103	4.1233
PADI4_23775	padi4	Poly 2	0.1167	0.0839	0.1854	4.1126
TUBA4A_7404	tuba4a	Linear	0.3528	0.2911	0.4475	3.8804
CDC42BPG_23788	cdc42bpg	Exp 4	0.1597	0.0944	0.3112	3.7222
GPR87_10766	gpr87	Exp 4	0.2091	0.1049	0.4801	3.6455
ACTBL2_24968	actbl2	Linear	0.3544	0.2922	0.4500	3.6442
ZSWIM4_87363		Linear	0.6816	0.4930	1.1036	3.6242
ACER2_17896	acer2	Exp 4	0.1302	0.0755	0.2486	3.5735
SPHK1_13551	sphk1	Exp 2	0.3772	0.3160	0.4706	3.5684
HIST1H2BJ_92661		Exp 2	0.3088	0.2649	0.3711	3.5200
EGLN3_90114		Linear	0.7297	0.5182	1.2328	3.4683

Supplementary Table 2: Summary of top twenty enriched GO terms for Hep G2 dosed with doxorubicin. Shown are for HepG2 dosed with doxorubicin the top twenty most significantly enriched genes based on the Highest Fold Change Absolute

Probe ID	Gene symbols	Best Model	Best BMD	Best BMDL	Best BMDU	Max Fold Change Value (Absolute)
CSTA_25241	csta	Exp 4	0.01382	0.01230	0.01570	716.79669
PADI4_23775	padi4	Hill	0.04418	0.03201	0.06589	121.29038
WFIKKN2_20142	wfikkn2	Exp 5	0.02422	0.01777	0.03324	100.32687
MAFB_26597	mafb	Linear	0.08905	0.08085	0.09889	91.22452
GAST_87534		Hill	0.09018	0.05326	0.18205	82.69598
TEX37_22098	tex37	Hill	0.02153	0.01756	0.03102	82.44215
ABCA12_18812	abca12	Exp 4	0.02365	0.02065	0.02750	64.00106

UBE2QL1_23376	ube2ql1	Hill	0.09083	0.05440	0.18423	58.30764
TREM2_20915	trem2	Hill	0.01516	0.01162	0.01950	57.19973
CGB5_21001	cgb1;cgb2;cgb3;cgb5;cgb8	Exp 4	0.02218	0.01939	0.02575	54.72734
TP53I3_24981	tp53i3	Exp 4	0.02017	0.01665	0.02470	54.51785
PLTP_18836	pltp	Exp 4	0.02866	0.02257	0.03677	53.78918
GRHL3_91237		Hill	0.01296	0.01081	0.01848	49.28883
PXDN_23717	pxdn	Power	0.16826	0.13209	0.22267	48.37337
COL7A1_90501		Hill	0.01741	0.01292	0.02387	47.83280
ACHE_13991	ache	Hill	0.04930	0.03177	0.08065	44.44639
TP53I3_7290	tp53i3	Exp 4	0.03145	0.02440	0.04102	44.27894
CGB8_28964	cgb5;cgb8	Exp 4	0.03327	0.02818	0.04012	43.56483
DRAXIN_12848	draxin	Exp 4	0.01111	0.00957	0.01301	43.38172
GDNF_24860	gdnf	Hill	0.02353	0.01766	0.03797	42.54195

Supplementary Table 3: Summary of top twenty enriched GO terms for MCF-7 cells dosed with doxorubicin. Shown are for MCF-7 cells dosed with doxorubicin the top twenty most significantly enriched genes based on the Highest Fold Change Absolute

Probe ID	Gene symbols	Best Model	Best BMD	Best BMDL	Best BMDU	Max Fold Change Value (Absolute)
TP53I3_24981	tp53i3	Exp 4	0.06320	0.05095	0.08152	71.06427
PGF_5091	pgf	Poly 2	0.04023	0.03463	0.04789	62.29139
TP53I3_7290	tp53i3	Power	0.06019	0.04146	0.08717	61.82729
NPTX1_18716	nptx1	Power	0.31073	0.20407	0.88476	46.81840
FOSL1_2463	fosl1	Linear	0.11691	0.10537	0.13103	43.18584
TP53I3_27999	tp53i3	Exp 4	0.07519	0.05891	0.10042	40.45102
HIST1H4A_19159	hist1h4a	Exp 4	0.04107	0.03269	0.05298	39.31838
SLC18A2_12450	slc18a2	Linear	0.11248	0.10151	0.12587	37.07502
FSCN1_22817	fscn1	Power	0.32532	0.20573	0.90762	36.56473
RASD1_17573	rasd1	Exp 4	0.07911	0.06113	0.10808	34.38891
INSYN2A_23631		Exp 2	0.15998	0.14637	0.17610	33.72083
PSG4_25648	psg4	Linear	0.12505	0.11245	0.14059	31.58548
ACHE_13991	ache	Power	0.07719	0.05045	0.11708	30.85504
ADAMTS7_24182	adamts7	Linear	0.11045	0.09973	0.12351	30.22548
GRIN2C_25162	grin2c	Power	0.01350	0.00732	0.02426	29.21803
CGA_1272	cga	Exp 2	0.18484	0.16895	0.20379	28.93337
JSRP1_89320		Exp 4	0.03724	0.03085	0.04606	28.64470
GPR87_10766	gpr87	Power	0.00956	0.00570	0.01589	28.10369
HAPLN3_21676	hapln3	Exp 2	0.16973	0.15465	0.18785	27.96263
GDNF_24860	gdnf	Hill	0.04941	0.03904	0.08021	26.97343

Supplementary Table 4: Summary of the top 20 significantly enriched REACTOME pathways. Shown are for HepaRG dosed with doxorubicin the top twenty most significantly enriched REACTOME pathways (Fisher's Exact Right P-Value < 0.05) based on the lowest mean BMDL values.

GO/Pathway ID	GO/Pathway Name	All Genes (Platform)	Genes that passed all filters	Fischers Exact two tail P value	BMD Mean	BMDL Mean	BMDU Mean
R-HSA-5633008	TP53 Regulates Transcription of Cell Death Genes	44	3	0.0074967	0.1099	0.0606	0.2212
R-HSA-1660661	Sphingolipid de novo biosynthesis	44	2	0.061151	0.253708	0.19574515	0.359575
R-HSA-109581	Apoptosis	179	5	0.024052	0.31143158	0.20561814	0.5569162
R-HSA-5357801	Programmed Cell Death	209	5	0.042607	0.31143158	0.20561814	0.5569162
R-HSA-75153	Apoptotic execution phase	52	2	0.081828	0.2912725	0.2066906	0.455861
R-HSA-8964539	Glutamate and glutamine metabolism	14	2	0.0070179	0.3025785	0.224515	0.4692075
R-HSA-6791312	TP53 Regulates Transcription of Cell Cycle Genes	48	2	0.071241	0.3667115	0.24565945	0.6201215
R-HSA-422475	Axon guidance	549	9	0.099314	0.375552967	0.261183967	0.658030222
R-HSA-1482801	Acyl chain remodelling of PS	23	2	0.018498	0.414665	0.261754	0.7726665
R-HSA-1793185	Chondroitin sulfate/dermatan sulfate metabolism	50	2	0.076475	0.40335015	0.2724123	0.7212225
R-HSA-6806667	Metabolism of fat-soluble vitamins	48	2	0.071241	0.435148	0.2748488	0.9390765
R-HSA-975634	Retinoid metabolism and transport	44	2	0.061151	0.435148	0.2748488	0.9390765
R-HSA-1483206	Glycerophospholipid biosynthesis	129	4	0.030458	0.423044075	0.27528625	0.7936905
R-HSA-428542	Regulation of commissural axon pathfinding by SLIT and ROBO	10	2	0.003554	0.455771	0.280632	0.9393225
R-HSA-1640170	Cell Cycle	687	2	0.093512	0.548989	0.289044	3.3592975

R-HSA-109606	Intrinsic Pathway for Apoptosis	55	2	0.090065	0.4491 2225	0.2909 8335	0.8567 005
R-HSA-114452	Activation of BH3-only proteins	30	2	0.030519	0.4491 2225	0.2909 8335	0.8567 005
R-HSA-111453	BH3-only proteins associate with and inactivate anti-apoptotic BCL-2 members	9	2	0.0028602	0.4491 2225	0.2909 8335	0.8567 005
R-HSA-499943	Interconversion of nucleotide di- and triphosphates	29	2	0.028652	0.5715 33	0.3046 42	3.3953 585
R-HSA-3371497	HSP90 chaperone cycle for steroid hormone receptors (SHR) in the presence of ligand	55	2	0.090065	0.3812 685	0.3082 135	0.4984 35

Supplementary Table 5: Summary of the top 20 significantly enriched REACTOME pathways. Shown are for MCF-7 dosed with doxorubicin the top twenty most significantly enriched REACTOME pathways (Fisher's Exact Right P-Value < 0.05) based on the lowest mean BMDL values.

GO/Pathway ID	GO/Pathway Name	All genes (Platform)	Genes that passed all filters	Fischers Exact two tail P value	BMD Mean	BMDL Mean	BMDU Mean
R-HSA-69478	G2/M DNA replication checkpoint	5	4	0.012526	0.0532 9732	0.0218 73	0.1218 17775
R-HSA-2514853	Condensation of Prometaphase Chromosomes	11	7	0.0053221	0.0552 482	0.0243 31371	0.1195 663
R-HSA-176974	Unwinding of DNA	12	6	4.18E-02	0.0545 1738	0.0330 9355	0.1191 856
R-HSA-69091	Polymerase switching	14	8	0.007064	0.0842 7262	0.0436 54087	0.1764 18225
R-HSA-69109	Leading Strand Synthesis	14	8	7.06E-03	0.0842 7262	0.0436 54087	0.1764 18225
R-HSA-69190	DNA strand elongation	32	16	0.0010128	0.0873 2631	0.0441 83894	0.2127 25619
R-HSA-176187	Activation of ATR in response	37	18	7.55E-04	0.0702 7603	0.0506 62178	0.1067 68417

	to replication stress						
R-HSA-69186	Lagging Strand Synthesis	20	10	9.02E-03	0.10701166	0.0508381	0.26884963
R-HSA-69166	Removal of the Flap Intermediate	14	7	2.82E-02	0.1319536	0.06079086	0.34477 4643
R-HSA-69183	Processive synthesis on the lagging strand	15	7	4.23E-02	0.1319536	0.06079086	0.34477 4643
R-HSA-73728	RNA Polymerase I Promoter Opening	62	23	0.011544	0.10244335	0.06915 8907	0.17729 9765
R-HSA-68962	Activation of the pre-replicative complex	33	19	2.79E-05	0.10829501	0.07012 7341	0.19112 0805
R-HSA-2299718	Condensation of Prophase Chromosomes	73	30	6.51E-04	0.11947681	0.07544 8686	0.23184 3523
R-HSA-9710421	Defective pyroptosis	72	28	0.0026091	0.12109338	0.07648 4797	0.24818 2182
R-HSA-212300	PRC2 methylates histones and DNA	72	26	1.11E-02	0.11764773	0.07667 4697	0.23678 1235
R-HSA-5334118	DNA methylation	64	26	1.77E-03	0.11758751	0.07720 5535	0.23515 8446
R-HSA-5625886	Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3	66	24	0.013105	0.11483446	0.07976 062	0.19169 5358
R-HSA-912446	Meiotic recombination	85	37	3.73E-05	0.11673672	0.08116 8125	0.19125 6216
R-HSA-1362300	Transcription of E2F targets	16	7	0.060342	0.13601206	0.08744 7314	0.22899 1314

	under negative control by p107 (RBL1) and p130 (RBL2) in complex with HDAC1						
R-HSA-5685939	HDR through MMEJ (alt-NHEJ)	10	6	0.014714	0.13890176	0.09096335	0.23632875

Supplementary Table 6: Summary of the top 20 significantly enriched REACTOME pathways. Shown are for Hep G2 dosed with doxorubicin the top twenty most significantly enriched REACTOME pathways (Fisher's Exact Right P-Value < 0.05) based on the lowest mean BMDL values.

GO/Pathway ID	GO/Pathway Name	All genes (Platform)	Genes that passed all filters	Fischers Exact two tail P value	BMD Mean	BMDL Mean	BMDU Mean
R-HSA-179409	APC-Cdc20 mediated degradation of Nek2A	26	2	0.096225	0.081404	0.03922515	0.179067
R-HSA-419408	Lysosphingolipid and LPA receptors	14	6	0.094678	0.07453945	0.0417311	0.14256993
R-HSA-936440	Negative regulators of DDX58/IFIH1 signaling	35	3	0.064156	0.09263 44	0.04372 95	0.201517
R-HSA-3656237	Defective EXT2 causes exostoses 2	14	6	0.094678	0.10687965	0.06785 2742	0.16527822
R-HSA-3656253	Defective EXT1 causes exostoses 1, TRPS2 and CHDS	14	6	0.094678	0.10687965	0.06785 2742	0.16527822
R-HSA-8849470	PTK6 Regulates Cell Cycle	5	3	0.071574	0.11968 3667	0.07630 59	0.217426
R-HSA-8939236	RUNX1 regulates transcription of genes involved in differentiation of HSCs	128	18	0.039623	0.14421 2906	0.09732 4463	0.24224724
R-HSA-69615	G1/S DNA Damage Checkpoints	67	8	0.053241	0.15891 4683	0.10532544	0.26674531
R-HSA-9616222	Transcriptional regulation of granulopoiesis	89	29	0.01914	0.15248 9442	0.10679 7463	0.24415795
R-HSA-9659787	Aberrant regulation of mitotic G1/S transition in cancer due to RB1 defects	17	7	0.07127	0.16540 1486	0.10734 7343	0.30008519

R-HSA-9661069	Defective binding of RB1 mutants to E2F1,(E2F2, E2F3)	17	7	0.07127	0.165401486	0.107347343	0.30008519
R-HSA-8856825	Cargo recognition for clathrin-mediated endocytosis	106	14	0.032766	0.166177182	0.108570561	0.32504439
R-HSA-174143	APC/C-mediated degradation of cell cycle proteins	87	7	9.76E-04	0.178645571	0.114180914	0.31294379
R-HSA-453276	Regulation of mitotic cell cycle	87	7	9.76E-04	0.178645571	0.114180914	0.31294379
R-HSA-187577	SCF(Skp2)-mediated degradation of p27/p21	58	7	0.079439	0.180236129	0.116857	0.31201983
R-HSA-69563	p53-Dependent G1 DNA Damage Response	65	6	0.014625	0.175076211	0.117156636	0.29013191
R-HSA-69580	p53-Dependent G1/S DNA damage checkpoint	65	6	0.014625	0.175076211	0.117156636	0.29013191
R-HSA-2214320	Anchoring fibril formation	15	8	0.0070881	0.172883087	0.118144038	0.22065089
R-HSA-2024096	HS-GAG degradation	22	10	0.015569	0.18200851	0.118455615	0.31554513
R-HSA-8857538	PTK6 promotes HIF1A stabilization	5	3	0.071574	0.157304667	0.122872867	0.226486

Appendix 5 Supplementary Tables 7-12

Supplementary Table 7. Summary of top twenty enriched GO terms for HepaRG dosed with niacinamide. Shown are for HepaRG dosed with niacinamide the top twenty most significantly enriched genes based on the Highest Fold Change Absolute

Probe ID	Gene symbols	Best Model	Best BMD	Best BMDL	Best BMDU	Max Fold Change Value (Absolute)
UGT1A1_89949		Linear	6611.47	4528.86	12238.00	4.22
CYP1A1_10775	cyp1a1	Power	1540.53	406.47	5911.82	3.59
S100A2_21300	s100a2	Exp 2	7617.74	4747.93	16929.00	3.46
CYP1A2_82402		Exp 2	5931.31	4367.74	9609.49	2.77
ALDH2_225	aldh2	Exp 2	7196.13	4913.14	13991.20	2.25
SULT2A1_24513	sult2a1	Linear	7077.59	4746.51	13905.80	2.09
CCDC9_20830	ccdc9	Power	0.000	0.000	0.000	2.04

Supplementary Table 8. Summary of top twenty enriched GO terms for HepG2 dosed with niacinamide. Shown are for Hep G2 dosed with niacinamide the top twenty most significantly enriched genes based on the Highest Fold Change Absolute

Probe ID	Gene symbols	Best Model	Best BMD	Best BMDL	Best BMDU	Max Fold Change Value (Absolute)
UGT1A1_89949		Power	1514.98	591.057	3517.19	16.24256516
CYP1A1_10775	cyp1a1	Poly 2	6649.46	4856.92	10201.4	15.49775791
UGT1A1_82407		Hill	3697.02	2043.28	6596.71	13.36897564
MAFB_26597	mafb	Power	55018.1	20144.6	55499.9	9.473223686
MLPH_16389	mlph	Exp 3	31261.6	16960.2	55861.8	9.453998566
UGT1A10_28864	ugt1a6;ugt1a10;ugt1a9;ugt1a1;ugt1a5;ugt1a7;ugt1a3;ugt1a8;ugt1a4	Power	1412.28	576.39	3179.21	8.860595703
CYP1A1_1698	cyp1a1	Linear	16862.7	14324.5	20478.4	7.990476131
LOX_21772	lox	Linear	24340.1	19652.5	31949.5	7.487246037
UGT1A8_28230	ugt1a8;ugt1a4;ugt1a6;ugt1a10;ugt1a9;ugt1a1;ugt1a5;ugt1a7;ugt1a3	Hill	3081.42	1609.34	5758.38	6.84641695
TMEM141_12190	tmem141	Exp 2	11787.3	10078.3	14106.1	6.16596508
HBEGF_15347	hbegf	Linear	17348.7	14689.4	21167.9	5.580536842
TENT5C_25075		Exp 2	17366	15388.5	19998.9	5.38710022
APOC3_356	apoc3	Power	1023.84	370.622	2536.45	5.107273579
APOC3_15165	apoc3	Power	1048.84	382.068	2588.86	5.07951498
DUSP5_92780		Exp 2	14264.8	12626	16412.9	5.029593945
SESN3_14278	sesn3	Linear	19614.9	16355.9	24480.5	4.939711094
SAA2-SAA4_28476		Power	1508.34	455.312	4194.88	4.929204464
CDH1_1186	cdh1	Power	1487.54	378.298	4644.85	4.902386665

OR2J3_22186	or2j3	Poly 2	10804.4	6812.94	21634.9	4.68742466
UNC93A_88810		Power	14281.7	5919.69	29715.9	4.681125164

Supplementary Table 9. Summary of top twenty enriched GO terms for MCF-7 dosed with niacinamide. Shown are for MCF-7 cells dosed with niacinamide the top twenty most significantly enriched genes based on the Highest Fold Change Absolute

Probe ID	Gene symbols	Best Model	Best BMD	Best BMDL	Best BMDU	Max Fold Change Value (Absolute)
ASCL1_11345	ascl1	Linear	9768.09	8677.59	11155.80	41.33
TMPRSS9_20419	tmprss9	Hill	10784.30	5327.84	11635.50	20.06
HBEGF_15347	hbegf	Linear	10598.80	9370.33	12181.60	16.08
DHRS2_15289	dhrs2	Linear	12829.60	11188.90	15017.70	16.07
CCDC83_88508		Power	22248.30	13248.50	54818.80	14.98
IQCN_90496		Linear	10879.80	9602.72	12532.10	14.02
ACKR3_11692	ackr3	Linear	8214.05	7359.28	9277.78	12.74
GOLGA6D_18150	golga6a;golga6d;golga6b;golga6c	Exp 4	7228.80	4919.95	11767.00	12.50
DHRS2_1857	dhrs2	Exp 4	8972.11	5664.87	16220.40	12.46
DKK1_25251	dkk1	Exp 3	49312.10	19298.00	53821.60	12.31
VHLL_88591		Linear	12064.10	10571.80	14030.70	12.13
RGS2_28241	rgs2	Linear	11917.70	10452.90	13843.50	11.79
SNAI1_24734	snai1	Exp 2	13533.20	12195.40	15210.10	11.63
LSMEM1_27464	lsmem1	Linear	14389.30	12424.50	17076.10	11.52
C5orf47_22862	c5orf47	Linear	12380.40	10827.60	14436.70	11.50
ESR1_12234	esr1	Power	54499.40	21263.40	54926.40	11.22
PYGM_24473	pygm	Linear	10461.20	9256.19	12010.60	11.00
HIST1H3F_93274		Exp 4	8364.82	4918.62	16007.60	10.98
SLC17A8_19336	slc17a8	Exp 2	13190.50	11958.30	14707.70	10.89
MDM2_27497	mdm2	Exp 3	52806.80	24482.90	55669.60	10.82

Supplementary Table 10. Summary of the top 20 significantly enriched REACTOME pathways. Shown are for HepaRG dosed with niacinamide the top twenty most significantly enriched REACTOME pathways (Fisher's Exact Right P-Value < 0.05) based on the lowest mean BMDL values.

GO/Pathway ID	GO/Pathway Name	All gene (platform)	Genes that passed all filters	Fischers Exact two tail P value	BMD Mean	BMDL Mean	BMDU Mean
R-HSA-556833	Metabolism of lipids	739	2	0.0133	4309.0600	2576.4880	9908.8100
R-HSA-1989781	PPARA activates gene expression	117	2	0.0003	4309.0600	2576.4880	9908.8100

R-HSA-400206	Regulation of lipid metabolism by PPARalpha	119	2	0.0004	4309.0600	2576.4880	9908.8100
R-HSA-211945	Phase I - Functionalization of compounds	106	2	0.0003	4368.3300	2659.8030	9951.5100
R-HSA-1430728	Metabolism	2102	3	0.0073	5271.4167	3355.3720	11269.6067
R-HSA-211859	Biological oxidations	222	3	0.0000	5271.4167	3355.3720	11269.6067

Supplementary Table 11. Summary of the top 20 significantly enriched REACTOME pathways. Shown are for MCF-7 cells dosed with niacinamide the top twenty most significantly enriched REACTOME pathways (Fisher's Exact Right P-Value < 0.05) based on the lowest mean BMDL values.

GO/Pathway ID	GO/Pathway Name	All Genes (Platform)	Genes that passed all filters	Fischers Exact two tail P value	BMD Mean	BMDL Mean	BMDU Mean
R-HSA-2299718	Condensation of Prophase Chromosomes	73	24	0.03335	10266.94833	5444.654583	27606.56417
R-HSA-5625886	Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3	66	24	0.010548	11563.69833	5764.829583	28621.68083
R-HSA-73728	RNA Polymerase I Promoter Opening	62	25	0.0011469	12423.5904	6752.9464	30732.1516
R-HSA-5334118	DNA methylation	64	27	3.79E-04	13006.36667	7340.405556	31954.8663
R-HSA-427389	ERCC6 (CSB) and EHMT2 (G9a) positively regulate rRNA expression	75	26	0.011828	12761.43308	7366.265769	29989.85154
R-HSA-110329	Cleavage of the damaged pyrimidine	60	20	0.042657	14167.08	7711.7755	32483.6575
R-HSA-73928	Depyrimidination	60	20	0.042657	14167.08	7711.7755	32483.6575
R-HSA-110328	Recognition and association of DNA glycosylase with site containing an affected pyrimidine	60	20	0.042657	14167.08	7711.7755	32483.6575
R-HSA-73929	Base-Excision Repair, AP Site Formation	62	21	0.03153	14247.8381	7990.505238	31845.48333
R-HSA-9710421	Defective pyroptosis	72	28	0.0014579	14451.455	8221.722857	34900.88893
R-HSA-427359	SIRT1 negatively regulates rRNA expression	67	27	9.36E-04	14465.58741	8518.863333	31995.63111
R-HSA-3214842	HDMs demethylate histones	50	22	5.21E-04	13911.12136	8561.191818	25647.32909

R-HSA-212300	PRC2 methylates histones and DNA	72	30	2.75E-04	14324.18467	8634.954667	32387.51633
R-HSA-977225	Amyloid fiber formation	110	35	0.020325	15028.88229	8675.159743	35065.13857
R-HSA-9616222	Transcriptional regulation of granulopoiesis	89	36	9.51E-05	15259.9525	8991.78963	30213.06204
R-HSA-201722	Formation of the beta-catenin:TCF transactivating complex	90	31	0.0071234	15022.4929	9169.015161	31781.72065
R-HSA-9031525	NR1H2 & NR1H3 regulate gene expression to limit cholesterol uptake	5	3	0.075878	24159.1	9279.474667	36636.73333
R-HSA-3214815	HDACs deacetylate histones	94	32	0.0083247	15780.82	9711.959688	33561.81063
R-HSA-5625740	RHO GTPases activate PKNs	94	32	0.0083247	16645.55187	9849.819375	37612.15075
R-HSA-912446	Meiotic recombination	85	31	0.0024084	16262.33419	9982.735806	34839.07226

Supplementary Table 12. Summary of the top 20 significantly enriched REACTOME pathways. Shown are for HepG2 cells dosed with niacinamide the top twenty most significantly enriched REACTOME pathways (Fisher's Exact Right P-Value < 0.05) based on the lowest mean BMDL values.

GO/Pathway ID	GO/Pathway Name	All Genes (Platform)	Genes that passed all filters	Fischers Exact two tail P value	BMD Mean	BMDL Mean	BMDU Mean
R-HSA-844615	The AIM2 inflammasome	3	2	0.0210	11468.3350	3729.7675	30213.4250
R-HSA-9660826	Purinergic signaling in leishmaniasis infection	26	5	0.0681	14432.6740	5046.8850	51766.3300
R-HSA-9664424	Cell recruitment (pro-inflammatory response)	26	5	0.0681	14432.6740	5046.8850	51766.3300
R-HSA-844456	The NLRP3 inflammasome	16	4	0.0434	15201.1675	5371.4163	57314.9375
R-HSA-8963898	Plasma lipoprotein assembly	19	4	0.0753	16475.0725	10169.9632	45091.3488
R-HSA-8963901	Chylomicron remodeling	10	4	0.0076	17613.7575	10672.2088	46656.2638
R-HSA-425410	Metal ion SLC transporters	26	5	0.0681	21371.1400	13010.7780	41216.9000
R-HSA-8963888	Chylomicron assembly	10	3	0.0485	21274.7133	13443.6883	56997.9517

R-HSA-8963684	Tyrosine catabolism	5	2	0.0624	21233.5150	14342.7765	36981.9350
R-HSA-3000480	Scavenging by Class A Receptors	19	4	0.0753	23906.7500	16533.3300	72871.6000
R-HSA-611105	Respiratory electron transport	93	3	0.0629	26591.0333	16973.3967	83182.3667
R-HSA-8963691	Phenylalanine and tyrosine metabolism	11	3	0.0626	23862.3100	17148.8843	38119.5567
R-HSA-381753	Olfactory Signaling Pathway	392	6	0.0000	25881.8400	17478.6150	48504.5500
R-HSA-70370	Galactose catabolism	5	3	0.0056	22430.9333	18253.8667	29226.6667
R-HSA-1799339	SRP-dependent cotranslational protein targeting to membrane	110	2	0.0056	32404.9500	18500.7800	94108.3500
R-HSA-72689	Formation of a pool of free 40S subunits	99	2	0.0114	32404.9500	18500.7800	94108.3500
R-HSA-927802	Nonsense-Mediated Decay (NMD)	113	2	0.0039	32404.9500	18500.7800	94108.3500
R-HSA-975956	Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	93	2	0.0236	32404.9500	18500.7800	94108.3500
R-HSA-975957	Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	113	2	0.0039	32404.9500	18500.7800	94108.3500
R-HSA-72764	Eukaryotic Translation Termination	91	2	0.0230	32404.9500	18500.7800	94108.3500

7. References

ABUETABH, Y., WU, H. H., CHAI, C., AL YOUSEF, H., PERSAD, S., SERGI, C. M. & LENG, R. 2022. DNA damage response revisited: the p53 family and its regulators provide endless cancer therapy opportunities. *Exp Mol Med,* 54**,** 1658-1669.

AGENCY, U. E. P. Benchmark Dose Software (BMDS).

ANDERSEN, M. E., MCMULLEN, P. D., PHILLIPS, M. B., YOON, M., PENDSE, S. N., CLEWELL, H. J., HARTMAN, J. K., MOREAU, M., BECKER, R. A. & CLEWELL, R. A. 2019. Developing context appropriate toxicity testing approaches using new alternative methods (NAMs). *ALTEX,* 36**,** 523-534.

BAINS, P., KAUR, M., KAUR, J. & SHARMA, S. 2018. Nicotinamide: Mechanism of action and indications in dermatology. *Indian J Dermatol Venereol Leprol,* 84**,** 234-237.

BALTAZAR, M. T., CABLE, S., CARMICHAEL, P. L., CUBBERLEY, R., CULL, T., DELAGRANGE, M., DENT, M. P., HATHERELL, S., HOUGHTON, J., KUKIC, P., LI, H., LEE, M. Y., MALCOMBER, S., MIDDLETON, A. M., MOXON, T. E., NATHANAIL, A. V., NICOL, B., PENDLINGTON, R., REYNOLDS, G., REYNOLDS, J., WHITE, A. & WESTMORELAND, C. 2020. A Next-Generation Risk Assessment Case Study for Coumarin in Cosmetic Products. *Toxicol Sci,* 176**,** 236-252.

BERGGREN, E., WHITE, A., OUEDRAOGO, G., PAINI, A., RICHARZ, A. N., BOIS, F. Y., EXNER, T., LEITE, S., GRUNSVEN, L. A. V., WORTH, A. & MAHONY, C. 2017. Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods. *Comput Toxicol,* 4**,** 31-44.

BEVAN, R. 2023. Akaike Information Criterion: When & How to use it (Example). *Scribbr.*

BIGANZOLI, L., CUFER, T., BRUNING, P., COLEMAN, R. E., DUCHATEAU, L., RAPOPORT, B., NOOIJ, M., DELHAYE, F., MILES, D., SULKES, A., HAMILTON, A. & PICCART, M. 2003. Doxorubicin-paclitaxel: a safe regimen in terms of cardiac toxicity in metastatic breast carcinoma patients. Results from a European Organization for Research and Treatment of Cancer multicenter trial. *Cancer,* 97**,** 40-5.

BLACK, M. B., STERN, A., EFREMENKO, A., MALLICK, P., MOREAU, M., HARTMAN, J. K. & MCMULLEN, P. D. 2022. Biological system considerations for application of toxicogenomics in next-generation risk assessment and predictive toxicology. *Toxicol In Vitro,* 80**,** 105311.

BLOMME, E. A., YANG, Y. & WARING, J. F. 2009. Use of toxicogenomics to understand mechanisms of drug-induced hepatotoxicity during drug discovery and development. *Toxicol Lett,* 186**,** 22-31.

BOO, Y. C. 2021. Mechanistic Basis and Clinical Evidence for the Applications of Nicotinamide (Niacinamide) to Control Skin Aging and Pigmentation. *Antioxidants (Basel),* 10.

BORAH, N. A. & REDDY, M. M. 2021. Aurora Kinase B Inhibition: A Potential Therapeutic Strategy for Cancer. *Molecules,* 26.

BUDTZ-JORGENSEN, E., BELLINGER, D., LANPHEAR, B., GRANDJEAN, P. & INTERNATIONAL POOLED LEAD STUDY, I. 2013. An international

pooled analysis for obtaining a benchmark dose for environmental lead exposure in children. *Risk Anal,* 33**,** 450-61.

CARMICHAEL, P., DAVIES, M., DENT, M., FENTEM, J., FLETCHER, S., GILMOUR, N., MACKAY, C., MAXWELL, G., MEROLLA, L., PEASE, C., REYNOLDS, F. & WESTMORELAND, C. 2009. Non-animal approaches for consumer safety risk assessments: Unilever's scientific research programme. *Altern Lab Anim,* 37**,** 595-610.

CAVANAUGH, J. & NEATH, A. 2019. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics,* 11.

CHEN, A. C. & DAMIAN, D. L. 2014. Nicotinamide and the skin. *Australas J Dermatol,* 55**,** 169-75.

COMSA, S., CIMPEAN, A. M. & RAICA, M. 2015. The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. *Anticancer Res,* 35**,** 3147-54.

CUI, Y. & PAULES, R. S. 2010. Use of transcriptomics in understanding mechanisms of drug-induced toxicity. *Pharmacogenomics,* 11**,** 573-85.

DE ABREW, K. N., SHAN, Y. K., WANG, X., KRAILLER, J. M., KAINKARYAM, R. M., LESTER, C. C., SETTIVARI, R. S., LEBARON, M. J., NACIFF, J. M. & DASTON, G. P. 2019. Use of connectivity mapping to support read across: A deeper dive using data from 186 chemicals, 19 cell lines and 2 case studies. *Toxicology,* 423**,** 84-94.

DENT, M., AMARAL, R. T., DA SILVA, P. A., ANSELL, J., BOISLEVE, F., HATAO, M., HIROSE, A., KASAI, Y., KERN, P., KREILING, R., MILSTEIN, S., MONTEMAYOR, B., OLIVEIRA, J., RICHARZ, A., TAALMAN, R., VAILLANCOURT, E., VERMA, R., POSADA, N. V. O. R. C., WEISS, C. & KOJIMA, H. 2018. Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Computational Toxicology,* 7**,** 20-26.

DENT, M. P., VAILLANCOURT, E., THOMAS, R. S., CARMICHAEL, P. L., OUEDRAOGO, G., KOJIMA, H., BARROSO, J., ANSELL, J., BARTON-MACLAREN, T. S., BENNEKOU, S. H., BOEKELHEIDE, K., EZENDAM, J., FIELD, J., FITZPATRICK, S., HATAO, M., KREILING, R., LORENCINI, M., MAHONY, C., MONTEMAYOR, B., MAZARO-COSTA, R., OLIVEIRA, J., ROGIERS, V., SMEGAL, D., TAALMAN, R., TOKURA, Y., VERMA, R., WILLETT, C. & YANG, C. 2021. Paving the way for application of next generation risk assessment to safety decision-making for cosmetic ingredients. *Regulatory Toxicology and Pharmacology,* 125**,** 105026.

DESPREZ, B., DENT, M., KELLER, D., KLARIC, M., OUEDRAOGO, G., CUBBERLEY, R., DUPLAN, H., EILSTEIN, J., ELLISON, C., GREGOIRE, S., HEWITT, N. J., JACQUES-JAMIN, C., LANGE, D., ROE, A., ROTHE, H., BLAAUBOER, B. J., SCHEPKY, A. & MAHONY, C. 2018. A strategy for systemic toxicity assessment based on non-animal approaches: The Cosmetics Europe Long Range Science Strategy programme. *Toxicol In Vitro,* 50**,** 137-146.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

DUIVENVOORDE, L. P. M., LOUISSE, J., PINCKAERS, N. E. T., NGUYEN, T. & VAN DER ZANDE, M. 2021. Comparison of gene expression and biotransformation activity of HepaRG cells under static and dynamic culture conditions. *Sci Rep,* 11**,** 10327.

EPA, U. 2014. Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational and systems Biology (Final Report). *US Environmental Protection Agency, Washington, DC*.

FAN, C., ZHENG, W., FU, X., LI, X., WONG, Y. S. & CHEN, T. 2014. Strategy to enhance the therapeutic effect of doxorubicin in human hepatocellular carcinoma by selenocystine, a synergistic agent that regulates the ROS-mediated signaling. *Oncotarget,* 5**,** 2853-63.

FANG, Y. & ZHANG, X. 2016. Targeting NEK2 as a promising therapeutic approach for cancer treatment. *Cell Cycle,* 15**,** 895-907.

FARAGHER, A. J. & FRY, A. M. 2003. Nek2A kinase stimulates centrosome disjunction and is required for formation of bipolar mitotic spindles. *Mol Biol Cell,* 14**,** 2876-89.

FARMAHIN, R., WILLIAMS, A., KUO, B., CHEPELEV, N. L., THOMAS, R. S., BARTON-MACLAREN, T. S., CURRAN, I. H., NONG, A., WADE, M. G. & YAUK, C. L. 2017. Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Archives of Toxicology,* 91**,** 2045-2065.

FIVENSON, D. P. 2006. The mechanisms of action of nicotinamide and zinc in inflammatory skin disease. *Cutis,* 77**,** 5-10.

FREDERICK, C. A., WILLIAMS, L. D., UGHETTO, G., VAN DER MAREL, G. A., VAN BOOM, J. H., RICH, A. & WANG, A. H. 1990. Structural comparison of anticancer drug-DNA complexes: adriamycin and daunomycin. *Biochemistry,* 29**,** 2538-49.

GEHRING, W. 2004. Nicotinic acid/niacinamide and the skin. *J Cosmet Dermatol,* 3**,** 88-93.

GILLESPIE, C., WASALATHANTHRI, D. P., RITZ, D. B., ZHOU, G., DAVIS, K. A., WUCHERPFENNIG, T. & HAZELWOOD, N. 2022. Systematic assessment of process analytical technologies for biologics. *Biotechnol Bioeng,* 119**,** 423-434.

HARRILL, J. A., EVERETT, L. J., HAGGARD, D. E., SHEFFIELD, T., BUNDY, J. L., WILLIS, C. M., THOMAS, R. S., SHAH, I. & JUDSON, R. S. 2021. High-Throughput Transcriptomics Platform for Screening Environmental Chemicals. *Toxicol Sci,* 181**,** 68-89.

HATHERELL, S., BALTAZAR, M. T., REYNOLDS, J., CARMICHAEL, P. L., DENT, M., LI, H., RYDER, S., WHITE, A., WALKER, P. & MIDDLETON, A. M. 2020. Identifying and Characterizing Stress Pathways of Concern for Consumer Safety in Next-Generation Risk Assessment. *Toxicol Sci,* 176**,** 11-33.

HU, D., JABLONOWSKI, C., CHENG, P. H., ALTAHAN, A., LI, C., WANG, Y., PALMER, L., LAN, C., SUN, B., ABU-ZAID, A., FAN, Y., BRIMBLE, M., GAMBOA, N. T., KUMBHAR, R. C., YANISHEVSKI, D., MILLER, K. M., KANG, G., ZAMBETTI, G. P., CHEN, T., YAN, Q., DAVIDOFF, A. M. & YANG, J. 2018. KDM5A Regulates a Translational Program that Controls p53 Protein Expression. *iScience,* 9**,** 84-100.

INJAC, R. & STRUKELJ, B. 2008. Recent advances in protection against doxorubicin-induced toxicity. *Technol Cancer Res Treat,* 7**,** 497-516.

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**,** D498-D503.

JIANG, L., ZHANG, H., XIAO, D., WEI, H. & CHEN, Y. 2022. Corrigendum to: "Farnesoid X receptor (FXR): Structures and ligands" [Comput. Struct. Biotechnol. J. 19 (2021) 2148-2159]. *Comput Struct Biotechnol J,* 20**,** 1227-1228.

JUNEDI, S., HERMAWAN, A., FITRIASARI, A., SETIAWATI, A., SUSIDARTI, R. A. & MEIYANTO, E. 2021. The Doxorubicin-Induced G2/M Arrest in Breast Cancer Cells Modulated by Natural Compounds Naringenin and Hesperidin. *Indonesian Journal of Cancer Chemoprevention,* 12**,** 83-89.

KAMANNA, V. S., GANJI, S. H. & KASHYAP, M. L. 2013. Recent advances in niacin and lipid metabolism. *Curr Opin Lipidol,* 24**,** 239-45.

KAVLOCK, R. J., BAHADORI, T., BARTON-MACLAREN, T. S., GWINN, M. R., RASENBERG, M. & THOMAS, R. S. 2018. Accelerating the Pace of Chemical Risk Assessment. *Chem Res Toxicol,* 31**,** 287-290.

KCIUK, M., GIELECINSKA, A., MUJWAR, S., KOLAT, D., KALUZINSKA-KOLAT, Z., CELIK, I. & KONTEK, R. 2023. Doxorubicin-An Agent with Multiple Mechanisms of Anticancer Activity. *Cells,* 12.

KNIP, M., DOUEK, I. F., MOORE, W. P., GILLMOR, H. A., MCLEAN, A. E., BINGLEY, P. J., GALE, E. A. & EUROPEAN NICOTINAMIDE DIABETES INTERVENTION TRIAL, G. 2000. Safety of high-dose nicotinamide: a review. *Diabetologia,* 43**,** 1337-45.

KREWSKI, D., ACOSTA, D., JR., ANDERSEN, M., ANDERSON, H., BAILAR, J. C., 3RD, BOEKELHEIDE, K., BRENT, R., CHARNLEY, G., CHEUNG, V. G., GREEN, S., JR., KELSEY, K. T., KERKVLIET, N. I., LI, A. A., MCCRAY, L., MEYER, O., PATTERSON, R. D., PENNIE, W., SCALA, R. A., SOLOMON, G. M., STEPHENS, M., YAGER, J. & ZEISE, L. 2010. Toxicity testing in the 21st century: a vision and a strategy. *J Toxicol Environ Health B Crit Rev,* 13**,** 51-138.

KREWSKI, D., ANDERSEN, M. E., TYSHENKO, M. G., KRISHNAN, K., HARTUNG, T., BOEKELHEIDE, K., WAMBAUGH, J. F., JONES, D., WHELAN, M., THOMAS, R., YAUK, C., BARTON-MACLAREN, T. & COTE, I. 2020. Toxicity testing in the 21st century: progress in the past decade and future perspectives. *Arch Toxicol,* 94**,** 1-58.

KUO, B., FRANCINA WEBSTER, A., THOMAS, R. S. & YAUK, C. L. 2016. BMDExpress Data Viewer - a visualization tool to analyze BMDExpress datasets. *J Appl Toxicol,* 36**,** 1048-59.

KUO, B., WEBSTER, A. F., THOMAS, R. S. & YAUK, C. L. 2016. BMDExpress Data Viewer - a visualization tool to analyze BMDExpress datasets. *Journal of Applied Toxicology,* 36**,** 1048-1059.

LAMB, J., CRAWFORD, E. D., PECK, D., MODELL, J. W., BLAT, I. C., WROBEL, M. J., LERNER, J., BRUNET, J. P., SUBRAMANIAN, A., ROSS, K. N., REICH, M., HIERONYMUS, H., WEI, G., ARMSTRONG, S. A., HAGGARTY, S. J., CLEMONS, P. A., WEI, R., CARR, S. A., LANDER, E. S. & GOLUB, T. R. 2006. The Connectivity Map: using

gene-expression signatures to connect small molecules, genes, and disease. *Science,* 313**,** 1929-35.
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.
MCSWEENEY, K. M., BOZZA, W. P., ALTEROVITZ, W. L. & ZHANG, B. 2019. Transcriptomic profiling reveals p53 as a key regulator of doxorubicin-induced cardiotoxicity. *Cell Death Discov,* 5**,** 102.
MI, X., TANG, M., LIAO, H., SHEN, W. & LEV, B. 2019. The state-of-the-art survey on integrations and applications of the best worst method in decision making: Why, what, what for and what's next? *Omega,* 87**,** 205-225.
MIDDLETON, A. M., REYNOLDS, J., CABLE, S., BALTAZAR, M. T., LI, H., BEVAN, S., CARMICHAEL, P. L., DENT, M. P., HATHERELL, S., HOUGHTON, J., KUKIC, P., LIDDELL, M., MALCOMBER, S., NICOL, B., PARK, B., PATEL, H., SCOTT, S., SPARHAM, C., WALKER, P. & WHITE, A. 2022. Are Non-animal Systemic Safety Assessments Protective? A Toolbox and Workflow. *Toxicol Sci,* 189**,** 124-147.
NAVARRO, A. P. & CHEESEMAN, I. M. 2021. Kinetochore assembly throughout the cell cycle. *Semin Cell Dev Biol,* 117**,** 62-74.
PANEL, C. 2005. Final report of the safety assessment of niacinamide and niacin. 24**,** 1-31.
PANEL, E. N. 2014. Scientific opinion on dietary reference values for niacin. *EFSA J.,* 12**,** 3759.
PANG, B., QIAO, X., JANSSEN, L., VELDS, A., GROOTHUIS, T., KERKHOVEN, R., NIEUWLAND, M., OVAA, H., ROTTENBERG, S., VAN TELLINGEN, O., JANSSEN, J., HUIJGENS, P., ZWART, W. & NEEFJES, J. 2013. Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat Commun,* 4**,** 1908.
PHILLIPS, J. R., SVOBODA, D. L., TANDON, A., PATEL, S., SEDYKH, A., MAV, D., KUO, B., YAUK, C. L., YANG, L., THOMAS, R. S., GIFT, J. S., DAVIS, J. A., OLSZYK, L., MERRICK, B. A., PAULES, R. S., PARHAM, F., SADDLER, T., SHAH, R. R. & AUERBACH, S. S. 2019. BMDExpress 2: enhanced transcriptomic dose-response analysis workflow. *Bioinformatics,* 35**,** 1780-1782.
POURQUIER, P., MONTAUDON, D., HUET, S., LARRUE, A., CLARY, A. & ROBERT, J. 1998. Doxorubicin-induced alterations of c-myc and c-jun gene expression in rat glioblastoma cells: role of c-jun in drug resistance and cell death. *Biochem Pharmacol,* 55**,** 1963-71.
PRODUCTS, E. P. F. D. 2015. Nutrition and Allergies. Scientific opinion on safety of caffeine. *EFSA J.,* 13**,** 4102.
R., C. N. 2007. Toxicity Testing in the 21st Century: A vision and a strategy. *The National Academies Press, Washington DC*.
RAHMAN, A. M., YUSUF, S. W. & EWER, M. S. 2007. Anthracycline-induced cardiotoxicity and the cardiac-sparing effect of liposomal formulation. *Int J Nanomedicine,* 2**,** 567-83.
RAJAGOPAL, R., BALTAZAR, M. T., CARMICHAEL, P. L., DENT, M. P., HEAD, J., LI, H., MULLER, I., REYNOLDS, J., SADH, K., SIMPSON, W., SPRIGGS, S., WHITE, A. & KUKIC, P. 2022. Beyond AOPs: A

Mechanistic Evaluation of NAMs in DART Testing. *Front Toxicol,* 4**,** 838466.
SCHURCH, N. J., SCHOFIELD, P., GIERLIŃSKI, M., COLE, C., SHERSTNEV, A., SINGH, V., WROBEL, N., GHARBI, K., SIMPSON, G. G., OWEN-HUGHES, T., BLAXTER, M. & BARTON, G. J. 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *Rna,* 22**,** 839-51.
SRITHARAN, S. & SIVALINGAM, N. 2021. A comprehensive review on time-tested anticancer drug doxorubicin. *Life Sci,* 278**,** 119527.
SVOBODA, M., KONVALINKA, J., TREMPE, J. F. & GRANTZ SASKOVA, K. 2019. The yeast proteases Ddi1 and Wss1 are both involved in the DNA replication stress response. *DNA Repair (Amst),* 80**,** 45-51.
TACAR, O., SRIAMORNSAK, P. & DASS, C. R. 2013. Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *J Pharm Pharmacol,* 65**,** 157-70.
THE GENE ONTOLOGY, C. 2017. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res,* 45**,** D331-D338.
THOMAS, A. F., KELLY, G. L. & STRASSER, A. 2022. Of the many cellular responses activated by TP53, which ones are critical for tumour suppression? *Cell Death Differ,* 29**,** 961-971.
THOMAS, R. S., BAHADORI, T., BUCKLEY, T. J., COWDEN, J., DEISENROTH, C., DIONISIO, K. L., FRITHSEN, J. B., GRULKE, C. M., GWINN, M. R., HARRILL, J. A., HIGUCHI, M., HOUCK, K. A., HUGHES, M. F., HUNTER, E. S., ISAACS, K. K., JUDSON, R. S., KNUDSEN, T. B., LAMBERT, J. C., LINNENBRINK, M., MARTIN, T. M., NEWTON, S. R., PADILLA, S., PATLEWICZ, G., PAUL-FRIEDMAN, K., PHILLIPS, K. A., RICHARD, A. M., SAMS, R., SHAFER, T. J., SETZER, R. W., SHAH, I., SIMMONS, J. E., SIMMONS, S. O., SINGH, A., SOBUS, J. R., STRYNAR, M., SWANK, A., TORNERO-VALEZ, R., ULRICH, E. M., VILLENEUVE, D. L., WAMBAUGH, J. F., WETMORE, B. A. & WILLIAMS, A. J. 2019. The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. *Toxicol Sci,* 169**,** 317-332.
WESTMORELAND, C., CARMICHAEL, P., DENT, M., FENTEM, J., MACKAY, C., MAXWELL, G., PEASE, C. & REYNOLDS, F. 2010. Assuring safety without animal testing: Unilever's ongoing research programme to deliver novel ways to assure consumer safety. *ALTEX,* 27**,** 61-5.
WOHLRAB, J. & KREFT, D. 2014. Niacinamide - mechanisms of action and its topical use in dermatology. *Skin Pharmacol Physiol,* 27**,** 311-5.
YEAKLEY, J. M., SHEPARD, P. J., GOYENA, D. E., VANSTEENHOUSE, H. C., MCCOMB, J. D. & SELIGMANN, B. E. 2017. A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. *PLoS One,* 12**,** e0178302.
YU, A. F., CHAN, A. T. & STEINGART, R. M. 2019. Cardiac Magnetic Resonance and Cardio-Oncology: Does T(2) Signal the End of Anthracycline Cardiotoxicity? *J Am Coll Cardiol,* 73**,** 792-794.