

The way forward in multi-omics data analyses

Citation for published version (APA):

Nguyen, N. (2022). *The way forward in multi-omics data analyses: from the methylome to the transcriptome and the proteome in drug-induced cardiotoxicity and hepatotoxicity*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20221122tn>

Document status and date:

Published: 01/01/2022

DOI:

[10.26481/dis.20221122tn](https://doi.org/10.26481/dis.20221122tn)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

The way forward in multi-omics data analyses

From the methylome to the transcriptome and the proteome in
drug-induced cardiotoxicity and hepatotoxicity

Nhan Nguyen



*The research described in this thesis was conducted at School for
Oncology and Reproduction (GROW), Maastricht University.*

© Nhan Nguyen, Maastricht, 2022

ISBN: 978-94-6469-056-9

Layout: Nhan Nguyen

Cover design: Nhan Nguyen & The Creative Hub, Maastricht University

(the cover contains a drawing made by Nhan Nguyen for this thesis)

Printed by the Canon Shop, Maastricht University

The way forward in multi-omics data analyses

From the methylome to the transcriptome and the proteome in
drug-induced cardiotoxicity and hepatotoxicity

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof. dr. Pamela Habibović
in accordance with the decision of the Board of Deans,
to be defended in public
on Tuesday, 22nd of November, 2022 at 10:00 hours

by

Nguyễn Thị Nhân (Nhan Nguyen)

born on 14th of December 1993, in Hanoi, Vietnam

Supervisor:

Prof. dr. J.C.S. Kleinjans

Co-supervisor:

Dr. D.G.J. Jennen

Assessment committee:

Prof. dr. R.L.M. Peeters (chair of the committee, Department of Data Science and Knowledge Engineering, Maastricht University, Netherlands)

Prof. dr. L.F.A. Wessels (Netherlands Cancer Institute, Amsterdam, Netherlands)

Dr. M. Summer-Kutmon (Department of Bioinformatics and Maastricht Centre for Systems Biology, Maastricht University, Netherlands)

Prof. dr. R. Schneider (Luxembourg Centre for Systems Biomedicine, Luxembourg)

Table of Contents

Preface.....	7
Chapter 1 General introduction.....	9
<i>Drug side effects</i>	10
<i>High throughput technologies</i>	12
<i>Aims and outline of the thesis</i>	21
Chapter 2 DNA methylation modifications analysis in Epirubicin-induced cardiotoxicity investigation	31
<i>Introduction</i>	33
<i>Experimental design and dataset</i>	35
<i>Chapter 2.1: Establishing a MeDIP-seq analysis workflow</i>	36
<i>Chapter 2.2: Epirubicin alters DNA methylation profiles related to cardiotoxicity</i>	42
<i>General remarks</i>	52
Chapter 3 Transcriptomics and proteomics analysis in Anthracycline-induced cardiotoxicity investigation	63
<i>Introduction</i>	64
<i>Experimental design and dataset</i>	66
<i>Chapter 3.1: Transcriptome analysis of long noncoding RNAs reveals their potential roles in anthracycline-induced cardiotoxicity</i>	67
<i>Chapter 3.2: Proteomics analysis of anthracycline-induced cardiotoxicity from cardiac microtissues to human heart biopsies</i>	80
<i>General remarks</i>	100
Chapter 4 Transcription factor and target relations in drug-induced hepatotoxicity.....	117
<i>Introduction</i>	118
<i>Chapter 4.1: RegOmics tool – Regulatory omics information</i>	119
<i>Chapter 4.2: TF-target relations related to Rifampicin-induced hepatotoxicity from DNA methylation to transcriptome and proteome</i>	124
<i>General remarks</i>	139
Chapter 5 General discussion	149
Impact paragraph.....	155
Addendum	159
<i>About the author</i>	161
<i>List of publications</i>	163
<i>Acknowledgments</i>	165

Preface

Drug side effects play an important role in drug safety. Research on drug adverse effects helps to unveil cellular disturbances under treatment as well as support drug discovery and development. High throughput technologies have been well established and considered as affordable approaches in biomedical research, especially for drug side effect investigations. These modern technologies can systematically capture the cells' epigenetic, transcriptomic, proteomic, and metabolomics alterations. Thus, they provide quantitative and multi-dimensional omics data and offer hypothesis-free observations. They are powerful approaches to recognize new drug-related targets and mechanisms on the molecular levels.

High throughput or omics technologies have been known for their extensive data generation and, consequently, expanded the possibility of drug side effects research. However, it is a genuine challenge to distil meaningful information from the massive omics dataset. Different omics data types also require particular data processing, filtering, and analysis that are not standardized. These aspects are the hurdles of omics data analyses and interpretation. In this thesis, we explore and demonstrate different approaches to inspect various omics data types derived from human microtissues exposed to different drugs. The purpose of this thesis is to not only contribute to drug side effect research but also partially tackle the omics analysis obstacle.

Keywords

Drug side effects, high throughput technologies, omics, toxicogenomics, DNA methylation, MeDIP-seq, transcriptomic, RNA sequencing, long non-coding RNA (lncRNA), proteomics, anthracycline, doxorubicin, epirubicin, idarubicin, rifampicin, cardiotoxicity, hepatotoxicity.

Abbreviations

ANT	anthracycline
circRNA	circular RNAs
DE	differentially expressed
DMR	differentially methylated region
DOX	doxorubicin
EPI	epirubicin
GC-MS	gas-chromatography mass spectrometry
GO	Gene Ontology
HPLC	high-performance liquid chromatography
HTPs	high throughput technologies
IDA	idarubicin
lncRNA	long non-coding RNA
Log ₂ FC	log ₂ fold change
log ₂ FC_avg	average log ₂ fold change
MBD-seq	methylated-CpG domain-binding proteins
MeDIP-seq	methylated DNA immunoprecipitation-sequencing
miRNA	microRNA
MS	mass spectrometry
ncRNA	noncoding RNA
NGS	next-generation sequencing
NMR	nuclear magnetic resonance
RNA-seq	RNA sequencing
RIF	rifampicin
TF	transcription factor
UPLC	ultra-performance liquid chromatography

Chapter 1 | General introduction

Adapted from

Nhan Nguyen, Danyel Jennen, Jos Kleinjans. “*Omics technologies to understand drug toxicity mechanisms.*” *Drug Discovery Today* (2022): 103348.

Drug side effects

Drug safety plays an important role in medical care as well as in drug development. The main focus of drug safety relates to the occurrence of adverse drug events, which are defined as injuries resulting from any drug-related intervention. In the United States, annual adverse drug events cause 2 million hospital stays in inpatient settings and over 3.5 million physician office visits in outpatient settings [1]. Parts of these events are preventable such as medication errors, overdoses, and allergic reactions [2]. Nevertheless, other adverse drug events, namely adverse drug reactions or drug side effects, are caused by underlying drug-induced mechanisms. Preventing those drug side effects in health care is essential not only for the clinical outcome but also for the economic impact [3]. Understanding drug adverse effects is also beneficial to drug screening and drug development in the pharmaceutical industry. In particular, drug toxicity has been responsible for the attrition of around 30% of drug candidates and contributed to the high cost and time-consuming process of drug discovery [4,5]. Thus, fundamental research on drug side effects is needed for disclosing undesired events and exploring their mechanisms.

With the exponential proliferation of the drug market, researchers demand feasible and effective methods to inquire about drug toxicity. Next to a wide range of animal-based models, toxicologists increasingly use *in vitro* cell systems exposed to different drug doses and measure targeted endpoints such as cytotoxicity, mitochondrial damage, oxidative stress, and apoptosis [4]. While this targeted approach could directly link the drug toxicity to individual elements of phenotype and suggest useful biomarkers for follow-up clinical investigations, it is time-consuming and poses difficulty in explaining causality at molecular levels. Additionally, a single biomarker or endpoint is unlikely to portray the complete adverse effects of a drug [6]. Therefore, new and more global methods are desired to overcome these obstacles.

Over the last decades, multiple high throughput technologies (HTPs) have emerged and shown their capability of providing a broader molecular detection in a single observation at acceptable prices and time (Figure 1-1). These HTPs dissolve the aforementioned obstacles in drug toxicity studies by recognizing multiple drug-induced alterations inside cells without prior knowledge. In this chapter, we discuss the availability of these cutting-edge technologies, and how they are utilized in drug side effect research. We have also foreseen the HTPs' prospects

and their obstructions and briefly explain the aim of this thesis related to this topic.

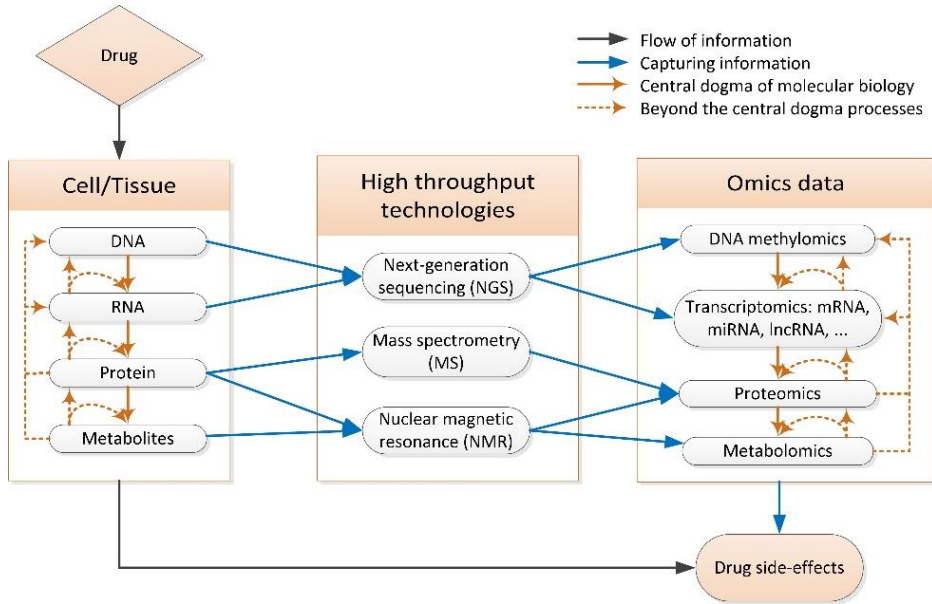


Figure 1-1: High throughput technologies (HTPs) in drug side effects studies

High throughput technologies

Although there are various technologies to detect biological molecules, only technologies that can generate an extensive amount of data in a short time are considered as HTPs. Current HTPs can measure different biological molecules, from DNA, RNA, and protein to metabolites (Figure 1-1). They can be categorized into three branches, comprising next-generation sequencing (NGS), mass spectrometry (MS), and nuclear magnetic resonance (NMR). NGS, also called high throughput sequencing, can be applied for DNA and RNA sequencing as well as NGS-based epigenetic profiling. MS is commonly used in protein identification, while both MS and NMR can be used for metabolites depending on the particular research purpose. Since HTPs can capture multiple biomolecules in one observation, they provide a high-dimensional output. The output data on each molecular level is named with the suffix “-omics”, such as transcriptomics and proteomics (Figure 1-1). This suffix also refers to a field of study that analyzes the large-scale data of a particular molecule [7]. Generally, HTPs or omics approaches illustrate the cellular molecular behaviours as a whole, guarantee a better understanding of the biological complexity, and serve as a system-wide approach to research on the adverse effects of drugs.

Epigenetic modifications: DNA methylation

Epigenetic modifications imply the heritable phenotype without altering DNA sequences that can affect gene expression. Investigations on this subject have become a general trend in pharmacology [8]. Although there is a wide variety of epigenetic alterations such as modifications of histone and other tudor domain proteins, the majority of drug research has focused on DNA methylation processes, in which methyl groups are added to or subtracted from cytosines in DNA (5-methylcytosine, i.e. 5mC). Changing 5mC levels can influence the functional state of genome regions by monitoring the gene expression or alternative splicing [9]. These epigenetic alterations can be caused by drugs and may persist even after the treatment. For instance, differences in DNA methylation sites can contribute to liver injury caused by rifampicin, a well-known antibacterial drug [10]. Another study in rats treated with doxorubicin illustrated the association of epigenetic dynamics with cardiac mitochondrial functions that can last a long time and affect the transcriptome and metabolome [11]. A study in human hepatocytes exposed to valproic acid, a common worldwide antiepileptic drug, identified methylation biomarkers not only in nuclear DNA but also in mitochondrial DNA [12]. Together with 5mC, researchers have discovered 5-

Hydroxymethylcytosine (5hmC), an epigenetic modification, which also happens in cytosines and plays some roles in gene expression regulation [13].

The DNA methylation-detecting technologies are classified into 3 main method groups: affinity enrichment-based methods, restriction enzymes-based methods, and bisulfite conversion-based methods [9,13]. The affinity enrichment-based methods group uses either antibodies (methylated DNA immunoprecipitation, MeDIP-seq) or proteins (methylated-CpG domain-binding proteins, MBD-Seq) to compile the methylated genomic regions for sequencing [13]. By contrast, the restriction enzyme-based methods identify 5mC in selected sequences by using restriction enzymes to cleave the recognition sequence at the site of DNA methylation. In the bisulfite conversion-based methods, DNA is denatured and subjected to bisulfite treatment. Thus, the unmodified cytosine is converted to uracil, while methylated cytosine remains unchanged and is detectable [9]. Recently, novel DNA methylation-detected methods have also been developed, including bisulfite conversion-based and capture-based methods or third-generation sequencing [9,13]. However, most of these methods are capable of detecting 5mC but could not differentiate between 5mC and 5hmC cases due to their similar structures. To specifically detect 5hmC, either altered enzymes or another oxidative transformation is required in the DNA preparation procedure [13,14].

Overall, this wide range of methods offers diverse possibilities for DNA methylation investigation in drug side effects research. The DNA methylation method, per se, has particular application due to its DNA input requirement, sensitivity, and specificity. In a large-scale approach, the whole-genome methylation profiling using MeDIP-seq or MBD-Seq can provide a comprehensive and unbiased view of the epigenome and discover new epigenetic biomarkers. On the contrary, other approaches, such as targeted bisulfite conversion, can evaluate the DNA methylation within specific DNA regions of interest and confirm particular mechanisms [15]. Together with potential applications, epigenetic research also requires intensive infrastructure as well as deep bioinformatics knowledge and skills to analyze the data. Furthermore, epigenetics research needs to standardize both experimental and computational procedures for enabling reliable comparisons between experiments, especially among different laboratories [9].

Transcriptome

The term transcriptome refers to the total RNA from a single cell or a population of cells at a particular condition. The transcriptome studies highly focus on differentially expressed genes and alternative splicing under disease and treatment conditions [16]. The premier subject in transcriptomic research is mRNA, in which the expression of mRNA can serve as a proxy to the relevant protein [17]. Nevertheless, besides mRNA, the RNA population also includes noncoding RNAs (ncRNA), which are not translated into protein and have a variety of important biological functions. For instance, the long noncoding RNAs (> 200 nucleotides) promote chromatin remodelling, regulate transcription and mRNA turnover, also monitor translation and post-translational modification [18]. MicroRNAs (miRNAs) are other well-known ncRNAs (approximate 22 nucleotides) that can cause mRNA degradation, translational repression, and gene silencing [19]. Furthermore, circular RNAs (circRNAs), a new class of ncRNA with diverse molecular functions, have conjointly gained acceptance in the genomics research community [20]. The high-profile function of circRNAs is to form miRNA sponges for regulating miRNA activities; this has been employed in therapeutic treatments but has not received vital attention in drug side effects research [21]. Thus, aside from traditional mRNAs, all these ncRNAs are also interesting targets in studying drug adverse effects.

RNA sequencing (RNA-seq) leverages the applications of NGS and provides a broad dynamic range and sensitivity to detect and quantify RNA in biological samples. There are different RNA-seq techniques from whole transcriptome sequencing to targeted RNA-seq. While total RNA or whole transcriptome sequencing can examine both coding and non-coding RNAs simultaneously, the mRNA sequencing applies poly(A) selection to sequence all protein-coding genes expression. Instead of the regular RNA-seq, the small RNA-seq has been developed to quantify microRNA (miRNA) due to the transcript length concern. While these foregoing RNA-seq techniques cover the global transcriptome in a biological sample with hypothesis-free experimental designs, targeted RNA sequencing focuses on specific genes of interest at a lower cost [22]. All these technologies have been based on the short-read sequencing concept, in which fragments under 200 nucleotides are selected for sequencing. This helps to provide high-throughput and robust data across platforms but also limits the transcript isoforms' detecting ability. To tackle this limitation, researchers have proposed a long-read sequencing concept without fragment size selection to enhance the capacity of capturing individual transcripts [23], even though the

current long-read technologies, such as PacBio and ONT platforms, have lower sensitivity and specificity, as well as higher error rates compared to short-read platforms. Overall, all these RNA-seq innovations and technologies have undoubtedly contributed to transcriptome-wide analysis in adverse drug effect research.

RNA sequencing significantly elevates the capacity to understand drug toxic mechanisms and has a positive impact on drug discovery and development. This HTP yields quantitative information about gene expression features including the presence/absence and quantification of a transcript, as well as the alternative splicing related to protein isoforms based on different drug doses and time exposures. It has shown numerous advantages in discovering new biomarkers, recognizing drugs-target genes, and identifying pharmaceutical mechanisms compared to other technologies [16]. The application of RNA sequencing gained even more benefits from recent bioinformatics developments including algorithm analysis and high-performance computing [24] with suggestive standardized preprocessing pipelines [25]. However, the downstream analysis to detect differential expression genes is still controversial, which gives some flexibility in research but also can lead to transparency issues. To tackle this issue, the R-ODAF has recently been released and aims to suggest a more rigorous transcriptomic data analysis framework [26].

Proteome

Proteomics focuses on quantifying protein levels in the cell, which is the terminal result of the gene expression control and directly relates to functional biological processes. For instance, a whole-cell proteomic study can reveal the oxaliplatin mechanisms of action, in which oxaliplatin triggers the DNA damage responses and nucleolar and ribosomal stress [27]. A global proteomic analysis in mice under rifampicin treatments convincingly determined the change of the proteome and characterized the related hepatotoxicity mechanistic pathways [28]. Thus, proteomic profiling can capture the cellular response to drugs and help to predict the side effects of drugs [29].

Different protein identification techniques may be utilized in drug side effects research. Western blotting and ELISA are traditional techniques that can detect protein expression by using specific antibodies to bind to the targeted proteins. However, these techniques are confined by prior assumptions about the protein characteristics and the availability of specific antibodies, and can only detect a few proteins in a single experiment [30]. Mass spectrometry (MS) is an advanced

proteomic technique and has been elected in recent proteomic studies due to its protein-detecting ability [31]. MS methods can be divided into global and targeted proteomics. Global proteomics attempts to quantify and identify all the proteins in a given sample. Thus, researchers can measure the difference in the proteome between samples without hypotheses [32]. On the other hand, targeted proteomics limits the number of detected proteins and optimizes the methods to obtain the highest sensitivity [33]. Furthermore, the MS analysis procedures can also be divided into the peptide-based approach and protein-based approach. The peptide-based approach is the older method and a bottom-up strategy, which analyzes the MS of peptides, resulting from the enzymatic digestion of proteins, to identify proteins. The protein-based approach is a recently developed method based on a top-down strategy by analyzing the entire proteins [33]. Thus, different MS methods and MS analysis procedures can be selected in the interest of particular research purposes.

Although MS methods provide high-throughput proteomics data and become a potential tool for pharmacology, it has several limitations. To date, none of the currently available proteomics techniques is able to detect all proteins within the cells. This limits the usage of proteomics compared to transcriptomics, which indeed captures all transcripts inside the cell. Additionally, proteomics techniques require comparatively large-sized samples and have difficulties with detecting low abundant proteins. The proteomics pipeline analysis is also not user-friendly and considerably differs among laboratories, which can generate a transparency problem. All these constraints restrict the application of proteomics, although, from the traditional point of view, the knowledge about protein is easier transposable from *in vitro* to clinical applications compared to transcriptome and epigenomics.

Metabolome

In recent years, toxicologists have been interested in the metabolome, the comprehensive set of small-molecule metabolites inside the cell [34,35]. While drug-induced changes in DNA methylation, transcriptome, and proteome, can only depict the potential outcomes, metabolites are the end products of all these cellular processes. Understanding the metabolome provides a direct correlation between drug influences and cellular responses [36]. Cohort studies have described the changing endogenous metabolites under pulmonary tuberculosis treatments, such as isoniazid and rifampicin, and enlightened unknown toxicity mechanisms of these medicines [37]. Researchers have eventually utilized the metabolome to predict the mode of action of new antimicrobial compounds [38].

Thus, by recognizing changes in metabolome under drug exposure, researchers can discover potential drug adverse effects, identify toxicity-related biomarkers, and predict drug toxicity.

There are different technologies to detect metabolites in biological samples. The two well-known technologies are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. They can be used in combination with other hyphenated platforms such as gas-chromatography mass spectrometry (GC-MS), high-performance liquid chromatography (HPLC), ultra-performance liquid chromatography (UPLC), or the combination of these platforms [34]. MS is a technology with high sensitivity and resolution and is able to detect an extensive number of metabolites for both qualitative and quantitative analysis. Its limitation is requiring separation steps to reduce sample complexity and minimize ionization suppression effects. Another downside of MS is the limited coverage of MS libraries for peak observation, causing a sectional view of the metabolome. The current high-resolution MS analysis focuses mainly on identifying lipids, namely lipidomics [39], and bile acids [40]. Contrary to MS, NMR technology can be directly used to analyze several classes of metabolites simultaneously and does not require complex sample preparation. Thus, NMR can be applied to compounds that are difficult to ionize or would need to be derivatized in MS analysis. NMR can also be used to study metabolites from biological fluids, such as urine, saliva, as well solid tissue samples derived from biopsy or bulk materials; and the biological sample does not need to be destructed and can be reused for further analysis. Nevertheless, NMR has low sensitivity and does not allow for the investigation of a large proportion of metabolites at low abundance as is possible with MS [34]. Eventually, each metabolomics technology has its own advantages and constraints and is suitable for different study purposes.

The main advantage of metabolomics is its close association with physiological responses. This helps to expand the knowledge about drug toxicity mechanisms and thus supports precision medicine, biomarker discovery, and drug development studies. However, the application of metabolomics is still restricted due to the limitation of metabolomics technologies with respect to their speed, costs, and data quality [41]. Additionally, metabolomics requires some standard data analysis procedures and comprehensive metabolites databases for capably identifying detected metabolites.

Future developments: single-cell and multi-omics approaches

One aspect of the biological system that also has become an obstacle for drug-related studies is cellular heterogeneity. Cellular heterogeneity has been widely recognized, in which the expression of genes, proteins, metabolites, and individual cell activities can vary among cellular populations [42-44]. Researchers also noticed the heterogeneous methylation at tissue levels and the links between cells' phenotype and genotype [9]. Thus, measuring cell behavior individually can give more accurate information and clarify the link between individual cell behavior to functional tissue rather than measuring bulk tissue, which normalizes all cellular variations. Single-cell RNA measurement was applied to track the transcriptomic landscape of human heart development [42]. The current single-cell HTPs are not limited to the transcriptome but are also available at the methylome, proteome, and metabolome levels [9,43-45]. Single-cell HTPs have been used to optimize treatment strategies [46] and to improve the drug discovery and development process [47]. The application of single-cell technologies, therefore, can extend scientific knowledge and encourage new insight into drug-induced mechanisms.

The development of modern HTPs also supported integrative omics analyses. While each HTP can provide a comprehensive view of the state of particular molecules, biological mechanisms are not limited to particular molecular status but are extended by how these molecules interact with each other. The interdependent relationships of cellular molecules contribute to cellular processes such as transcriptional regulation, protein-protein interaction, and metabolic regulation. Because of this, analyses focusing on a single molecular type cannot elucidate the entire biological response to drug treatment [48]. Multi-omics approaches, which combine different omics layers, encourage an integrative and systems-based approach towards understanding drug-related mechanisms as compared to single omics data analyses. For example, researchers have incorporated gene expression and methylation patterns to provide a better understanding of the drug's long-term side effects [49,50]. The multi-omics approach can be named differently such as "multi-omics," "vertical omics," or "trans-omics" [48,51], but they all refer to the integration of several omics data to constitute a flow of information and help to reveal the complete drug-related mechanisms.

The advancement of multi-omics analyses considers drug-induced mechanisms as the integrated response of the whole biological system rather than the sum of each -omics regime. These approaches not only uncover biological

interactions but also initiate substantial hypotheses across multiple omics layers for dedicated empirical validation. In general, multi-omics analyses can be divided into two major approaches: using prior knowledge or data-driven approaches [51]. The prior knowledge approach relies on the known relationships between biological molecules (extrinsic information), such as transcription factor (TF) - target regulation, protein-protein interaction, or gene expression regulation by miRNAs and circRNA. This approach can draw reliable conclusions but also limits itself to established references on interactions. On the contrary, the data-driven approach, such as matrix factorization, classification, regression, or deep learning, only relies on the omics data themselves (intrinsic information) [52]. This approach can suggest potential causality and interactions based on the dependencies within and between omics layers data. However, predictions based on any approach certainly require experimental validation. To leverage the benefits of both approaches, researchers have attempted to establish hybrid integration such as composite network methods [52]. While researchers have been prompted to assimilate two omics types into their analysis; recent studies have endeavored to integrate more omics types to disclose the complex adverse drug responses [49,53]. With the accumulation of various omics datasets, the multi-omics analyses can elevate their power predictions for either refining drug treatments or accelerating the drug discovery process [52]. Thus, multi-omics approaches are on the cutting-edge and sequentially promoted in adverse drug effect investigations.

Notably, multi-omics analysis is the predominant bottleneck in the field of toxicogenomics. A gold standard for multi-omics study design as well as analysis methodologies is certainly not yet available. For the study design, the high costs of *in vitro* setting and the availability of precious biopsy samples restrain the sampling capacity to obtain all omics layers per replicate generated from one sample source. Furthermore, the difference in the response time and the lifespan of biomolecules within and between the omics layers also initiate issues: making sampling time points complicated as well as accumulating noise signals in the integrative process [54]. For the data analysis, multi-omics analysis inherits original challenges from single omics analyses plus new challenges of integrating multi-omics layers into an identical biological framework. Besides the limit of available “gold standard” datasets to serve as unbiased ground truth, the multi-omics analysis has to face sufficient benchmarks and reproducibility issues [55]. While a range of different approaches, as mentioned in the previous paragraph, is available for multi-omics analysis, particular computational methods have mainly been developed for a very specific case and can only operate on certain criteria.

No multi-omics approach has been able to provide informative molecular interpretation combined with time and dose-response relationship as well as support reproducible research [54]. Compared to these challenges in multi-omics analysis, the obstacle of single-cell omics has been partly resolved by the advanced technologies for measurements as well as adjusted the ongoing data analysis approaches of bulk cells for single-cell data interpretation. Researchers have attempted to combine single-cell omics with multi-omics analysis to gain advances in both single-cell resolutions and a systematic understanding of cellular events [56]. Thus, the development of multi-omics data analysis could leverage the single-cell multi-omics approach and augment the exploration of toxicology. Multi-omics approaches should be a focus in future toxicogenomics/toxicology analyses.

Aims and outline of the thesis

Understanding drug side effects research is essential for drug safety, as well as drug development and discovery. As explained in this introduction, HTPs have been well established in biomedical research, especially for drug adverse effect investigations; however, this demands compatible bioinformatics analyses. Thus, the flourishing of effective omics data analyses could support the progress in this research field.

This thesis thus focuses on omics data analyses and explores different data analysis approaches as well as research angles to interpret omics data. We did data analyses among different omics data layers from epigenomics (MeDIP-seq) and transcriptomics (RNA-seq) to proteomics (MS). To respond to the need of developing multi-omics analysis approaches, as argued above, we did not limit the analysis scale to a single omics layer, but also pursued cross-omics interpretation as a multi-omics approach. These bioinformatics analyses can serve as introductory and accessible workflows to interpret omics data to understand biological mechanisms and direct further research. Besides investigating new omics data analyses, in this thesis, we also demonstrated how we utilized these approaches in particular drug side effects (anthracycline, rifampicin) as case studies.

Dataset

As discussed in previous sections, there are not many available datasets consisting of multiple omics data types. In this thesis, we fortunately had access to a comprehensive dataset, which comprised different omics data types derived from the same samples. The dataset was acquired from the Hepatic and Cardiac Toxicity Systems (HeCaToS) modeling project that was funded by the European Union Seventh Framework Programme (FP7/2007-2013) with grant number ID: 602156.

The HeCaToS project (<https://cordis.europa.eu/project/id/602156>) aimed at developing integrative *in silico* models for predicting human heart and liver toxicity. This project used advanced *in vitro* 3D liver and heart human microtissues. The human cardiac microtissues (3D InSight™ Human Cardiac Microtissues from InSphero) contained 4000 iPSC-derived human cardiomyocytes from a female Caucasian donor and 1000 cardiac fibroblasts from a male Caucasian donor. The human cardiac microtissues were cultured in 3D InSight™ Human Cardiac Microtissues Maintenance Medium (InSphero) [57]. The human 3D

hepatic microtissue model comprised of 1000 primary human hepatocytes mixed from 5 males and 5 females donor (7-59 years old) and 1000 primary human Kupffer cells from a Caucasian 27-year-old with unreported gender. The microtissues were cultured in 3D Insight™ Human Liver Microtissues Maintenance Medium - AF (InSphero) [58].

According to the experimental design, the microtissues were separately exposed to 20 targeted drugs, at 2 dose administrations (therapeutic and toxic) in 14 days. The therapeutic dose reflected the clinical dose and the toxic dose was the IC₂₀ value based on the ATP production (cell viability) previously determined after 7 days of exposure [59]. The dose treatment regimes reflected the specific drug interstitial concentrations over time based on the dose administration that was calculated by reverse physiologically based pharmacokinetic (PBPK) modeling (Table S 1-1, Table S 1-2) [60]. For the functional evaluation, the ATP content in microtissues was measured during toxic treatment. Along with the drug exposure, the biomolecules of microtissues were measured by HTPs to generate epigenomics, transcriptomics, and proteomics data. In parallel, biopsies were obtained from patients with cardiomyopathies and liver injuries due to drug side effects. The biopsies' molecular status was also measured by similar HTPs. The HeCaToS dataset is deposited in the BioStudies platform with identified accession numbers (<http://www.ebi.ac.uk/biostudies>). This thesis focused on analyzing omics data derived from cardiac and hepatic microtissues exposed to anthracyclines (ANTs) and rifampicin (RIF), respectively.

Outline of the thesis

We have presented the background information and the general aims in this current chapter - **chapter 1**.

In **chapter 2**, we explored the epigenetic modifications related to drug-induced side effects, in this case, EPI-induced cardiotoxicity. We first established a workflow to analyze the methylation data in **chapter 2.1**, then in **chapter 2.2**, we applied this workflow to the MeDIP-seq data deriving from cardiac microtissues exposed to EPI.

In **chapter 3**, we investigated how drugs, in particular the ANT family that are well-known anti-tumor and cardiotoxic agents, can impact the transcriptome and proteome and lead to adverse side effects. In **chapter 3.1**, we analyzed the transcriptome of microtissues exposed to ANTs and specifically focused on lncRNAs, a newly emerged biomarker type that appeared in recent research. In **chapter 3.2**, we analyzed the proteome of the same microtissues exposed to ANT.

In this **chapter 3**, we not only detected some candidate lncRNAs and proteins as potential targets for further ANT-induced side effect investigation in the *in vitro* experiment but also checked these genes and protein expression in biopsies samples obtained from ANT-treated and control patients.

In **chapter 4**, we focused on biological interaction and cross-omics analysis. Omics data analysis can result in an extensive list of genes or proteins, which form difficulties in perceiving TF-target relations. In **chapter 4.1**, we developed Regomics, a tool for downstream omics data analysis, to retrieve potentially transcription factors (TFs) that regulated the given genes/proteins. In **chapter 4.2**, we analyzed the epigenomics, transcriptomics, and proteomics from hepatic samples exposed to RIF, a drug with hepatotoxicity side effects. We also deployed the Regomics tool to determine TF- target relations possibly related to RIF side effects.

Chapter 5 is the general discussion while we compiled and recapped remarks on all of the preceding analysis research. In this chapter, we also criticize the data used and established workflows, as well as discuss the further research direction. The impact of this thesis work is discussed in the **Impact paragraph**.

The thesis is concluded by the **Addendum**, in which all the thanks are given to people who support me during my PhD (**Acknowledgments**), as well as all my professional information (**Curriculum vitae** and **List of publications**)

Supplementary Materials

Supplementary tables

Table S 1-1: The anthracycline concentration profile in the microtissue testing

Sampling time point	Time range (hours)	Incubation time (hours)	Respective concentration [uM]						
			Control DMSO	Doxorubicin		Epirubicin		Idarubicin	
				Therapeutic	Toxic	Therapeutic	Toxic	Therapeutic	Toxic
T2	0 - 2	2	0.062	0.211	0.602	0.221	0.9	0.005	0.087
T8	2 - 8	6	0.006	0.016	0.046	0.013	0.051	0.001	0.023
T24	8 - 24	16	0.002	0.008	0.024	0.004	0.018	0.001	0.01
	24 - 26	2	0.063	0.217	0.62	0.224	0.912	0.006	0.106
T72	26 - 32	6	0.007	0.022	0.063	0.015	0.061	0.002	0.037
	32 - 48	16	0.004	0.013	0.038	0.006	0.025	0.002	0.016
	48 - 50	2	0.064	0.221	0.631	0.225	0.917	0.006	0.115
	50 - 56	6	0.008	0.026	0.073	0.016	0.066	0.003	0.046
	56 - 72	16	0.004	0.016	0.046	0.007	0.029	0.002	0.02
	72 - 144	72	0.013	0.039	0.111	0.028	0.116	0.003	0.038
T168	144 - 146	2	0.065	0.227	0.649	0.226	0.923	0.007	0.123
	146 - 152	6	0.009	0.031	0.09	0.018	0.072	0.003	0.053
	152 - 168	16	0.005	0.021	0.06	0.008	0.034	0.002	0.023
	168 - 170	2	0.065	0.228	0.65	0.227	0.924	0.007	0.124
	170 - 176	6	0.01	0.032	0.091	0.018	0.072	0.003	0.053
	176 - 192	16	0.006	0.021	0.061	0.008	0.034	0.002	0.023
	192 - 194	2	0.065	0.228	0.651	0.227	0.924	0.007	0.124
	194 - 200	6	0.01	0.032	0.092	0.018	0.072	0.003	0.053
	200 - 216	16	0.006	0.022	0.062	0.008	0.034	0.002	0.023
	216 - 218	2	0.065	0.228	0.652	0.227	0.924	0.007	0.124
T240	218 - 224	6	0.01	0.032	0.092	0.018	0.073	0.003	0.053
	224 - 240	16	0.006	0.022	0.062	0.008	0.034	0.002	0.023
	240 - 312	72	0.014	0.042	0.119	0.029	0.118	0.003	0.039
	312 - 314	2	0.065	0.229	0.653	0.227	0.925	0.007	0.124
	314 - 320	6	0.01	0.033	0.093	0.018	0.073	0.003	0.053
T336	320 - 336	16	0.006	0.022	0.063	0.008	0.034	0.002	0.024

Table S 1-2: The rifampicin concentration profile in the microtissue testing

Sampling time point	Time range (hours)	Incubation time (hours)	Control with 0.1 % DMSO	Respective concentration [uM]	
				RIF therapeutic	RIF toxic
To	0 h	0 h	0.1%	0.000	0.000
T2	0 - 2 h	2 h	0.1%	1.886	55.527
T8	2 - 8 h	6 h	0.1%	0.890	29.768
T24	8 - 24 h	16 h	0.1%	0.213	9.747
	24 - 26 h	2 h	0.1%	1.965	60.102
	26 - 32 h	6 h	0.1%	0.954	33.543
	32 - 48 h	16 h	0.1%	0.247	12.100
	48 - 50 h	2 h	0.1%	1.987	61.711
	50 - 56 h	6 h	0.1%	0.975	35.017
	T72	56 - 72 h	16 h	0.1%	0.265
72 - 144 h		72 h	0.1%	0.615	41.651
144 - 146 h		2 h	0.1%	2.044	65.094
146 - 152 h		6 h	0.1%	1.034	38.427
T168	152 - 168 h	16 h	0.1%	0.321	16.567
	168 - 170 h	2 h	0.1%	2.057	65.841
	170 - 176 h	6 h	0.1%	1.047	39.189
	176 - 192 h	16 h	0.1%	0.333	17.322
	192 - 194 h	2 h	0.1%	2.069	66.571
	194 - 200 h	6 h	0.1%	1.059	39.934
	200 - 216 h	16 h	0.1%	0.345	18.059
	216 - 218 h	2 h	0.1%	2.081	67.286
T240	218 - 224 h	6 h	0.1%	1.072	40.664
	224 - 240 h	16 h	0.1%	0.357	18.781
	240 - 312 h	72 h	0.1%	0.701	-
	312 - 314 h	2 h	0.1%	2.124	-
T336	314 - 320 h	6 h	0.1%	1.115	-
	320 - 336 h	16 h	0.1%	0.399	-

Notes: RIF, rifampicin; -, not applicable.

References

- [1] U.D.o. Health, H. Services, National action plan for adverse drug event prevention, Washington, DC (2014) 50-98.
- [2] I.J. Onakpoya, C.J. Heneghan, J.K. Aronson, Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature, *BMC Med* 14 (2016) 10. <https://doi.org/10.1186/s12916-016-0553-2>
- [3] D. Formica, J. Sultana, P.M. Cutroneo, S. Lucchesi, R. Angelica, S. Crisafulli, Y. Ingrassiotta, F. Salvo, E. Spina, G. Trifirò, The economic burden of preventable adverse drug reactions: a systematic review of observational studies, *Expert Opinion on Drug Safety* 17 (2018) 681-695. <https://doi.org/10.1080/14740338.2018.1491547>
- [4] F.P. Guengerich, Mechanisms of drug toxicity and relevance to pharmaceutical development, *Drug metabolism and pharmacokinetics* 26 (2011) 3-14. <https://doi.org/10.2133/dmpk.dmpk-10-rv-062>
- [5] R.K. Harrison, Phase II and phase III failures: 2013–2015, *Nature Reviews Drug Discovery* 15 (2016) 817-818. <https://doi.org/10.1038/nrd.2016.184>
- [6] H. Yun-Fu, K. June, H. Yiwu, From Traditional Biomarkers to Transcriptome Analysis in Drug Development, *Current Molecular Medicine* 5 (2005) 29-38. <http://dx.doi.org/10.2174/1566524053152915>
- [7] S.P. Yadav, The wholeness in suffix -omics, -omes, and the word om, *Journal of biomolecular techniques : JBT* 18 (2007) 277-277.
- [8] Z. Sessions, N. Sánchez-Cruz, F.D. Prieto-Martínez, V.M. Alves, H.P. Santos, E. Muratov, A. Tropsha, J.L. Medina-Franco, Recent progress on cheminformatics approaches to epigenetic drug discovery, *Drug Discovery Today* 25 (2020) 2268-2276. <https://doi.org/10.1016/j.drudis.2020.09.021>
- [9] D. Barros-Silva, C.J. Marques, R. Henrique, C. Jerónimo, Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications, 9 (2018) 429.
- [10] C. Huai, Y. Wei, M. Li, X. Zhang, H. Wu, X. Qiu, L. Shen, L. Chen, W. Zhou, N. Zhang, G. Zhu, Y. Zhang, Z. Zhang, L. He, S. Qin, Genome-Wide Analysis of DNA Methylation and Antituberculosis Drug-Induced Liver Injury in the Han Chinese Population, 106 (2019) 1389-1397. <https://doi.org/10.1002/cpt.1563>
- [11] M. Tantawy, F.G. Pamittan, S. Singh, Y. Gong, Epigenetic Changes Associated With Anthracycline-Induced Cardiotoxicity, 14 (2021) 36-46. <https://doi.org/10.1111/cts.12857>
- [12] J.E.J. Wolters, S.G.J. van Breda, F. Caiment, S.M. Claessen, T. de Kok, J.C.S. Kleinjans, Nuclear and Mitochondrial DNA Methylation Patterns Induced by Valproic Acid in Human Hepatocytes, *Chem Res Toxicol* 30 (2017) 1847-1854. <https://doi.org/10.1021/acs.chemrestox.7b00171>
- [13] S. Li, T.O. Tollefsbol, DNA methylation methods: Global DNA methylation and methylomic analyses, *Methods* 187 (2021) 28-43. <https://doi.org/10.1016/j.ymeth.2020.10.002>
- [14] J.J. Briedé, L. Deferme, J.E.J. Wolters, S.M.H. Claessen, T. van den Beucken, R.J. Wagner, S.G. van Breda, J.C.S. Kleinjans, A cross-omics approach to investigate temporal gene expression regulation by 5-hydroxymethylcytosine via TBH-derived oxidative stress

showed involvement of different regulatory kinases, *Toxicology in Vitro* 48 (2018) 318-328. <https://doi.org/10.1016/j.tiv.2018.02.006>

[15] S. Kurdyukov, M. Bullock, DNA Methylation Analysis: Choosing the Right Method, *Biology* 5 (2016) 3. <https://doi.org/10.3390/biology5010003>

[16] X. Yang, L. Kui, M. Tang, D. Li, K. Wei, W. Chen, J. Miao, Y. Dong, High-Throughput Transcriptome Profiling in Drug and Biomarker Discovery, *Frontiers in Genetics* 11 (2020). <https://doi.org/10.3389/fgene.2020.00019>

[17] B.T. Gufford, J.D. Robarge, M.T. Eadon, H. Gao, H. Lin, Y. Liu, Z. Desta, T.C. Skaar, Rifampin modulation of xeno- and endobiotic conjugating enzyme mRNA expression and associated microRNAs in human hepatocytes, *Pharmacology Research & Perspectives* 6 (2018) e00386. <https://doi.org/10.1002/prp2.386>

[18] R.-W. Yao, Y. Wang, L.-L. Chen, Cellular functions of long noncoding RNAs, *Nature Cell Biology* 21 (2019) 542-551. <https://doi.org/10.1038/s41556-019-0311-8>

[19] J. O'Brien, H. Hayder, Y. Zayed, C. Peng, Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation, *Frontiers in Endocrinology* 9 (2018). <https://doi.org/10.3389/fendo.2018.00402>

[20] K.-Y. Hsiao, H.S. Sun, S.-J. Tsai, Circular RNA – New member of noncoding RNA with novel functions, *Experimental Biology and Medicine* 242 (2017) 1136-1141. <https://doi.org/10.1177/1535370217708978>

[21] A. Lavenniah, T.D.A. Luu, Y.P. Li, T.B. Lim, J. Jiang, M. Ackers-Johnson, R.S.Y. Foo, Engineered Circular RNA Sponges Act as miRNA Inhibitors to Attenuate Pressure Overload-Induced Cardiac Hypertrophy, *Molecular Therapy* 28 (2020) 1506-1517. <https://doi.org/10.1016/j.ymthe.2020.04.006>

[22] R. Hrdlickova, M. Toloue, B. Tian, RNA-Seq methods for transcriptome analysis, *WIREs RNA* 8 (2017) e1364. <https://doi.org/10.1002/wrna.1364>

[23] R. Stark, M. Grzelak, J. Hadfield, RNA sequencing: the teenage years, *Nature Reviews Genetics* 20 (2019) 631-656. <https://doi.org/10.1038/s41576-019-0150-2>

[24] B. Schmidt, A. Hildebrandt, Next-generation sequencing: big data meets high performance computing, *Drug Discovery Today* 22 (2017) 712-717. <https://doi.org/10.1016/j.drudis.2017.01.014>

[25] A. Federico, A. Serra, M.K. Ha, P. Kohonen, J.-S. Choi, I. Liampa, P. Nymark, N. Sanabria, L. Cattalani, M. Fratello, P.A.S. Kinaret, K. Jagiello, T. Puzyn, G. Melagraki, M. Gulumian, A. Afantitis, H. Sarimveis, T.-H. Yoon, R. Grafström, D. Greco, Transcriptomics in Toxicogenomics, Part II: Preprocessing and Differential Expression Analysis for High Quality Data, *Nanomaterials (Basel, Switzerland)* 10 (2020) 903. <https://doi.org/10.3390/nano10050903>

[26] M. Verheijen, W. Tong, L. Shi, T.W. Gant, B. Seligman, F. Caiment, Towards the development of an omics data analysis framework, *Regulatory Toxicology and Pharmacology* 112 (2020) 104621. <https://doi.org/10.1016/j.yrtph.2020.104621>

[27] T. Ozdian, D. Holub, Z. Maceckova, L. Varanasi, G. Rylova, J. Rehulka, J. Vaclavkova, H. Slavik, P. Moudry, P. Znojek, J. Stankova, J.B. de Sanctis, M. Hajduch, P. Dzubak, Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main

responses to oxaliplatin treatment in cancer cells, *Journal of Proteomics* 162 (2017) 73-85. <https://doi.org/10.1016/j.jpropt.2017.05.005>

[28] J.-H. Kim, W.S. Nam, S.J. Kim, O.K. Kwon, E.J. Seung, J.J. Jo, R. Shresha, T.H. Lee, T.W. Jeon, S.H. Ki, H.S. Lee, S. Lee, Mechanism Investigation of Rifampicin-Induced Liver Injury Using Comparative Toxicoproteomics in Mice, *International journal of molecular sciences* 18 (2017) 1417. <https://doi.org/10.3390/ijms18071417>

[29] M. Zitnik, M. Agrawal, J. Leskovec, Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 34 (2018) 1457-1466. <https://doi.org/10.1093/bioinformatics/bty294>

[30] R.A. Alharbi, Proteomics approach and techniques in identification of reliable biomarkers for diseases, *Saudi Journal of Biological Sciences* 27 (2020) 968-974. <https://doi.org/10.1016/j.sjbs.2020.01.020>

[31] L. Monaci, E. De Angelis, N. Montemurro, R. Pilolli, Comprehensive overview and recent advances in proteomics MS based methods for food allergens analysis, *TrAC Trends in Analytical Chemistry* 106 (2018) 21-36. <https://doi.org/10.1016/j.trac.2018.06.016>

[32] N. Nguyen, T. Souza, M.C.T. Verheijen, H. Gmuender, N. Selevsek, R. Schlapbach, J. Kleinjans, D. Jennen, Translational Proteomics Analysis of Anthracycline-Induced Cardiotoxicity From Cardiac Microtissues to Human Heart Biopsies, *Front Genet* 12 (2021) 695625. <https://doi.org/10.3389/fgene.2021.695625>

[33] B. Deracinois, C. Flahaut, S. Duban-Deweer, Y. Karamanos, Comparative and Quantitative Global Proteomics Approaches: An Overview, 1 (2013) 180-218.

[34] K.A. Stuart, K. Welsh, M.C. Walker, R. Edrada-Ebel, Metabolomic tools used in marine natural product drug discovery, *Expert Opinion on Drug Discovery* 15 (2020) 499-522. <https://doi.org/10.1080/17460441.2020.1722636>

[35] K. Segers, S. Declerck, D. Mangelings, Y.V. Heyden, A.V. Eeckhaut, Analytical techniques for metabolomic studies: a review, 11 (2019) 2297-2318. <https://doi.org/10.4155/bio-2019-0014>

[36] G. Quintás, T. Martínez-Sena, I. Conde, E. Pareja Ibars, J. Kleinjans, J.V. Castell, Metabolomic analysis to discriminate drug-induced liver injury (DILI) phenotypes, *Archives of Toxicology* 95 (2021) 3049-3062. <https://doi.org/10.1007/s00204-021-03114-z>

[37] M. Combrink, D.T. Loots, I. du Preez, Metabolomics describes previously unknown toxicity mechanisms of isoniazid and rifampicin, *Toxicology Letters* 322 (2020) 104-110. <https://doi.org/10.1016/j.toxlet.2020.01.018>

[38] M. Zampieri, B. Szappanos, M.V. Buchieri, A. Trauner, I. Piazza, P. Picotti, S. Gagneux, S. Borrell, B. Gicquel, J. Lelievre, B. Papp, U. Sauer, High-throughput metabolomic analysis predicts mode of action of uncharacterized antimicrobial compounds, *Science Translational Medicine* 10 (2018) eaal3973. <https://doi.org/doi:10.1126/scitranslmed.aal3973>

[39] T. Züllig, H.C. Köfeler, High resolution mass spectrometry in lipidomics, *Mass Spectrometry Reviews* 40 (2021) 162-176. <https://doi.org/10.1002/mas.21627>

[40] D. Dewaele, P. Annaert, E. Hoeben, LC-MS/MS Analysis of Bile Acids in In Vitro Samples, in: M. Vinken (Ed.), *Experimental Cholestasis Research*, Springer New York, New York, NY, 2019, pp. 15-23.

- [41] X.-w. Zhang, Q.-h. Li, Z.-d. Xu, J.-j. Dou, Mass spectrometry-based metabolomics in health and medical science: a systematic review, *RSC Advances* 10 (2020) 3092-3104. <https://doi.org/10.1039/C9RA08985C>
- [42] Y. Cui, Y. Zheng, X. Liu, L. Yan, X. Fan, J. Yong, Y. Hu, J. Dong, Q. Li, X. Wu, S. Gao, J. Li, L. Wen, J. Qiao, F. Tang, Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart, *Cell Reports* 26 (2019) 1934-1950.e5. <https://doi.org/10.1016/j.celrep.2019.01.079>
- [43] L. Zhang, A. Vertes, Single-Cell Mass Spectrometry Approaches to Explore Cellular Heterogeneity, *Angewandte Chemie (International ed. in English)* 57 (2018) 4466-4477. <https://doi.org/10.1002/anie.201709719>
- [44] A. Gérard, A. Woolfe, G. Mottet, M. Reichen, C. Castrillon, V. Menrath, S. Ellouze, A. Poitou, R. Doineau, L. Briseno-Roa, P. Canales-Herrerias, P. Mary, G. Rose, C. Ortega, M. Delincé, S. Essono, B. Jia, B. Iannascoli, O. Richard-Le Goff, R. Kumar, S.N. Stewart, Y. Pousse, B. Shen, K. Grosselin, B. Saudemont, A. Sautel-Caillé, A. Godina, S. McNamara, K. Eyer, G.A. Millot, J. Baudry, P. England, C. Nizak, A. Jensen, A.D. Griffiths, P. Bruhns, C. Brenan, High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics, *Nature Biotechnology* 38 (2020) 715-721. <https://doi.org/10.1038/s41587-020-0466-7>
- [45] D. Wang, S. Bodovitz, Single cell analysis: the new frontier in 'omics', *Trends in biotechnology* 28 (2010) 281-290. <https://doi.org/10.1016/j.tibtech.2010.03.002>
- [46] S. Yamada, S. Nomura, Review of Single-Cell RNA Sequencing in the Heart, *International Journal of Molecular Sciences* 21 (2020) 8345. <https://doi.org/10.3390/ijms21218345>
- [47] J.R. Heath, A. Ribas, P.S. Mischel, Single-cell analysis tools for drug discovery and development, *Nature Reviews Drug Discovery* 15 (2016) 204-216. <https://doi.org/10.1038/nrd.2015.16>
- [48] K. Yugi, H. Kubota, A. Hatano, S. Kuroda, Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple 'Omics' Layers, *Trends in Biotechnology* 34 (2016) 276-290. <https://doi.org/10.1016/j.tibtech.2015.12.013>
- [49] N. Selevsek, F. Caiment, R. Nudischer, H. Gmuender, I. Agarkova, F.L. Atkinson, I. Bachmann, V. Baier, G. Barel, C. Bauer, S. Boerno, N. Bosc, O. Clayton, H. Cordes, S. Deeb, S. Gotta, P. Guye, A. Hersey, F.M.I. Hunter, L. Kunz, A. Lewalle, M. Lienhard, J. Merken, J. Minguet, B. Oliveira, C. Pluess, U. Sarkans, Y. Schrooders, J. Schuchhardt, I. Smit, C. Thiel, B. Timmermann, M. Verheijen, T. Wittenberger, W. Wolski, A. Zerck, S. Heymans, L. Kuepfer, A. Roth, R. Schlapbach, S. Niederer, R. Herwig, J. Kleinjans, Network integration and modelling of dynamic drug responses at multi-omics levels, *Communications biology* 3 (2020) 573-573. <https://doi.org/10.1038/s42003-020-01302-8>
- [50] C.-H. Hsu, H. Tomiyasu, C.-H. Liao, C.-S. Lin, Genome-wide DNA methylation and RNA-seq analyses identify genes and pathways associated with doxorubicin resistance in a canine diffuse large B-cell lymphoma cell line, *PLOS ONE* 16 (2021) e0250013. <https://doi.org/10.1371/journal.pone.0250013>
- [51] Y. Hasin, M. Seldin, A. Lusic, Multi-omics approaches to disease, *Genome Biology* 18 (2017) 83. <https://doi.org/10.1186/s13059-017-1215-1>

- [52] M.A. Wörheide, J. Krumsiek, G. Kastenmüller, M. Arnold, Multi-omics integration in biomedical research – A metabolomics-centric review, *Analytica Chimica Acta* 1141 (2021) 144-162. <https://doi.org/10.1016/j.aca.2020.10.038>
- [53] B. Lee, S. Zhang, A. Poleksic, L. Xie, Heterogeneous Multi-Layered Network Model for Omics Data Integration and Analysis, *Frontiers in Genetics* 10 (2020). <https://doi.org/10.3389/fgene.2019.01381>
- [54] S. Canzler, J. Schor, W. Busch, K. Schubert, U.E. Rolle-Kampczyk, H. Seitz, H. Kamp, M. von Bergen, R. Buesen, J. Hackermüller, Prospects and challenges of multi-omics data integration in toxicology, *Archives of Toxicology* 94 (2020) 371-388. <https://doi.org/10.1007/s00204-020-02656-y>
- [55] M. Krassowski, V. Das, S.K. Sahu, B.B. Misra, State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing, *Frontiers in Genetics* 11 (2020). <https://doi.org/10.3389/fgene.2020.610798>
- [56] J. Lee, D.Y. Hyeon, D. Hwang, Single-cell multiomics: technologies and data analysis methods, *Experimental & Molecular Medicine* 52 (2020) 1428-1442. <https://doi.org/10.1038/s12276-020-0420-2>
- [57] M. Verheijen, Transcriptomics close to my heart: advanced models & methods for toxicogenomics research illustrated by anthracycline-induced cardiotoxicity. Chapter 4., in: *Toxicogenomic, GROW school, Maastricht University, ProefschriftMaken Maastricht, 2019*, pp.^79-109.
- [58] M. Verheijen, Transcriptomics close to my heart: advanced models & methods for toxicogenomics research illustrated by anthracycline-induced cardiotoxicity. Chapter 2., in: *Toxicogenomic, GROW school, Maastricht University, ProefschriftMaken Maastricht, 2019*, pp.^24-51.
- [59] M. Verheijen, Y. Schrooders, H. Gmuender, R. Nudischer, O. Clayton, J. Hynes, S. Niederer, H. Cordes, L. Kuepfer, J. Kleinjans, F. Caiment, Bringing in vitro analysis closer to in vivo: Studying doxorubicin toxicity and associated mechanisms in 3D human microtissues with PBPK-based dose modelling, *Toxicology Letters* 294 (2018) 184-192. <https://doi.org/10.1016/j.toxlet.2018.05.029>
- [60] L. Kuepfer, O. Clayton, C. Thiel, H. Cordes, R. Nudischer, L.M. Blank, V. Baier, S. Heymans, F. Caiment, A. Roth, D.A. Fluri, J.M. Kelm, J. Castell, N. Selevsek, R. Schlapbach, H. Keun, J. Hynes, U. Sarkans, H. Gmuender, R. Herwig, S. Niederer, J. Schuchhardt, M. Segall, J. Kleinjans, A model-based assay design to reproduce in vivo patterns of acute drug-induced toxicity, *Archives of Toxicology* 92 (2018) 553-555. <https://doi.org/10.1007/s00204-017-2041-7>

Chapter 2 | DNA methylation modifications analysis in Epirubicin-induced cardiotoxicity investigation

Adapted from

Nhan Nguyen, Matthias Lienhard, Ralf Herwig, Jos Kleinjans, and Danyel Jennen. "A bioinformatics workflow to detect genes with DNA methylation alterations: a case study of analyzing MeDIP-seq data in cardiac microtissue exposed to epirubicin." In 2022 12th International Conference on Bioscience, Biochemistry and Bioinformatics (2022): 63-69.

Nhan Nguyen, Matthias Lienhard, Ralf Herwig, Jos Kleinjans, and Danyel Jennen. "Epirubicin alters DNA methylation profiles related to cardiotoxicity." *Frontiers in Bioscience-Landmark* 27.6 (2022): 173.

Introduction

Epigenetic modifications are heritable alterations that can influence the functional state of genome regions without changing the DNA sequence. One of the major epigenetic modifications is DNA methylation, in which a methyl group is added to the fifth carbon position of the cytosine base [1]. The DNA methylation status is regulated by both methylation and demethylation processes. This epigenetic modification plays an essential role in different biological activities based on where it is located in genomic regions. DNA methylation in intergenic regions can repress the expression of potentially harmful genetic elements, while DNA methylation in CpG islands can diminish transcription factor binding, and recruit repressive methyl-binding protein, resulting in gene silencing [2]. In gene regions, DNA methylation at the first exon can lead to gene silencing [2], whereas DNA methylations in other gene regions can also be signals for RNA splicing regulators [3]. Alterations in DNA methylation status can then lead to changes in gene and protein expressions. Studies have determined the relationship between human DNA methylation alterations and external factors such as environmental exposure, nutritional status, and disease [1,4]. Recent research also shows that DNA methylation can be a useful tool to understand a drug's mode of action and adverse outcome pathway [5]. Therefore, studying DNA methylation could provide an added value to cellular mechanisms, especially for drug-induced mechanisms investigation.

Among different DNA methylation detection approaches as mentioned in chapter 1 [6], methylated DNA immunoprecipitation-sequencing (MeDIP-seq) is a cost-effective technology, which requires low DNA input while providing adequate accuracy, genome coverage, and resolution [7]. MeDIP-seq uses a specific antibody to immunoprecipitate methylated DNA, and then evaluates the obtained fractions by high-throughput sequencing. Thus, MeDIP-seq can estimate the relative enrichment of methylated DNA across the majority of the genome (>95%). This genome coverage of the MeDIP-seq is higher compared to other methods such as the reduced representation bisulfite (RRBS) or the whole-genome bisulfite (WGBS) protocol which both have approximately a 75% coverage [8]. Thus, although MeDIP-seq can only determine the regional changes in DNA methylation, it is an effective tool to establish pioneer information about the genome-wide methylation status under specific conditions.

In this chapter, we focused on developing a bioinformatics workflow to analyze MeDIP-seq data (chapter 2-1) and then utilized this workflow to interpret

the MeDIP-seq data from cardiac microtissues exposed to epirubicin (EPI) in chapter 2-2.

Experimental design and dataset

The human cardiac microtissues (3D InSight™ Human Cardiac Microtissues from InSphero) were exposed to either a clinically therapeutic or a toxic (IC₂₀) dose. Every weekday, the sample medium was renewed 3 times corresponding to the cardiac drug interstitial concentration profile at 2, 8, and 24 hours calculated by the physiologically based pharmacokinetic (PBPK) modeling. EPI was dissolved in 0.1% DMSO before utilization, thus control samples were also exposed to similar DMSO concentrations over time (Table S 1-1,) [9].

The DNA and RNA were collected from microtissues in triplicates at 2, 8, 24, 72, 168, 240, and 336 hours of exposure. After DNA extraction, the methylated DNA fragments were isolated by anti-5-methylcytosine antibody and then paired-end sequenced (MeDIP-seq) with 50 bp read length [10]. After RNA extraction, the total RNA in each sample was isolated using Qiagen AllPrep DNA/RNA/miRNA Universal Kit (Cat #80224). Ribosomal RNAs were depleted by using the Illumina RiboZero Gold kit (Cat #MRZG12324), and then samples were prepared by the Lexogen SENSE total RNA library preparation kit (Cat #009.96). The RNA quality and quantity of the samples were checked by the Agilent 420 TapeStation and the Qubit™ before they were sequenced by an HiSeq2000 with 100bp paired-end reads [11].

Chapter 2.1: Establishing a MeDIP-seq analysis workflow

Objectives of the study

The outcome of the MeDIP-seq method is sequenced reads derived from methylated DNA regions. This data need to be processed by mapping the reads to the reference genome, calculating methylation levels, and identifying differentially methylated regions (DMRs) among input samples. To assist the MeDIP-seq data analysis, different computational methods have been developed such as Batman [12], MEDIPS [13], MeDUSA [14], MeQA [15], and recently QSEA [16]. Some of the tools were published quite a long time ago such as Batman was launched in 2008 [12], or are currently unavailable such as MeQA [15]. Furthermore, among those accessible tools, the most comprehensive pipelines end at determining and annotating DMRs between samples [17].

QSEA is a recent R package that is launched in 2017 as the successor of the MEDIPS package to analyze the MeDIP-seq data. QSEA provides a straight MeDIP analysis process from quantifying and normalizing MeDIP-seq data to detecting and annotating differentially methylated regions between samples [16]. This certainly supports the general MeDIP-seq data analysis. However, analyzing MeDIP-seq data at the regional level could result in thousands of DMRs, which is unmanageable for researchers to extract meaningful information. Some studies have upgraded the DNA methylation analysis at the DMR level to the gene level by identifying genes that had DMR located in the promoter region [18]. Thus, exploring different ways to elucidate the MeDIP-seq data at the gene level could elevate the application of DNA methylation analysis.

In this chapter, we aimed to build a bioinformatics workflow based on the recently developed MeDIP-seq analysis method (QSEA) to address the DNA methylation at the gene levels. This workflow could aid researchers to detect the DMRs from MeDIP-seq data and refine the extensive DMRs list into a shortlist of methylated candidate genes. The code is written in R and is publicly available (<https://github.com/NhanNguyenooo/MeDIP>).

Analysis procedure

The MeDIP-seq paired-end reads were aligned to human reference genome hg38 using Burrows-Wheeler Alignment tool (BWA) version 0.7.17 [19] and converted to .bam files using Samtools version 1.10 [20] in the Linux environment (Figure 2-1).

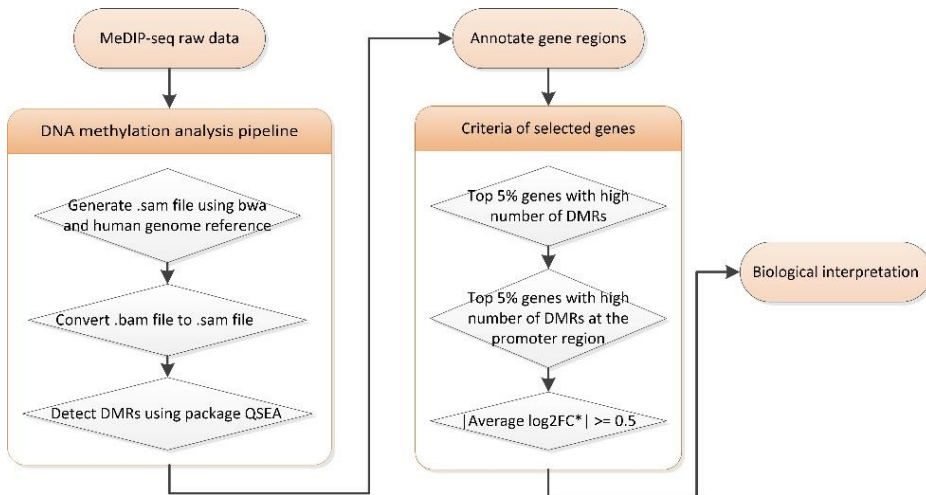


Figure 2-1: The bioinformatics workflow for MeDIP-seq data analysis. (*) The average \log_2FC of each gene was the average $\log_2 FC$ from all DMRs assigned to that gene.

Thereafter, the aligned MeDIP-seq data were processed in R version 3.6.3 [21] using the QSEA package and human genome build hg38 with default window size (250 bases) excluding sex chromosomes [16]. Copy Number Variations (CNVs) were calculated from input and MeDIP reads based on 1 megabase windows. The CpG enrichment was estimated using the ‘Blind calibration’ option. The quality of the mapped MeDIP-seq data was performed using `getOffset()` and `plotEPmatrix()` functions. Thereafter, the QSEA package detected DMRs (p -value < 0.01) using generalized linear models (GLMs) with pairwise comparisons between EPI-treated and control samples (Figure 2-1).

The DMRs were annotated to genomic regions using the “annotatr” package [22]. The “annotatr” package provides comprehensive annotations and genomic context of DMRs. The average p -value and \log_2 fold change (\log_2FC) per gene were calculated based on the average of the p -value and \log_2FC from all DMRs assigned to that gene. Thereafter, candidate genes were selected for further analysis based on criteria: (i) select the top 5% genes with the highest number of DMRs across their gene regions, (ii) select the top 5% genes with the highest number of DMRs in their promoter region, and (iii) select genes that had the absolute $\log_2FC \geq 0.5$ (Figure 2-1).

Results

The methylation enrichment efficiency was sufficient in all harvested samples (Figure S 2-1). In general, DNA methylation profiles differed between EPI-treated and control samples (Figure 2-2). This indicates that EPI treatment could significantly alter the DNA methylation in cardiac tissues and MeDIP-seq was able to capture these DNA methylation modifications.

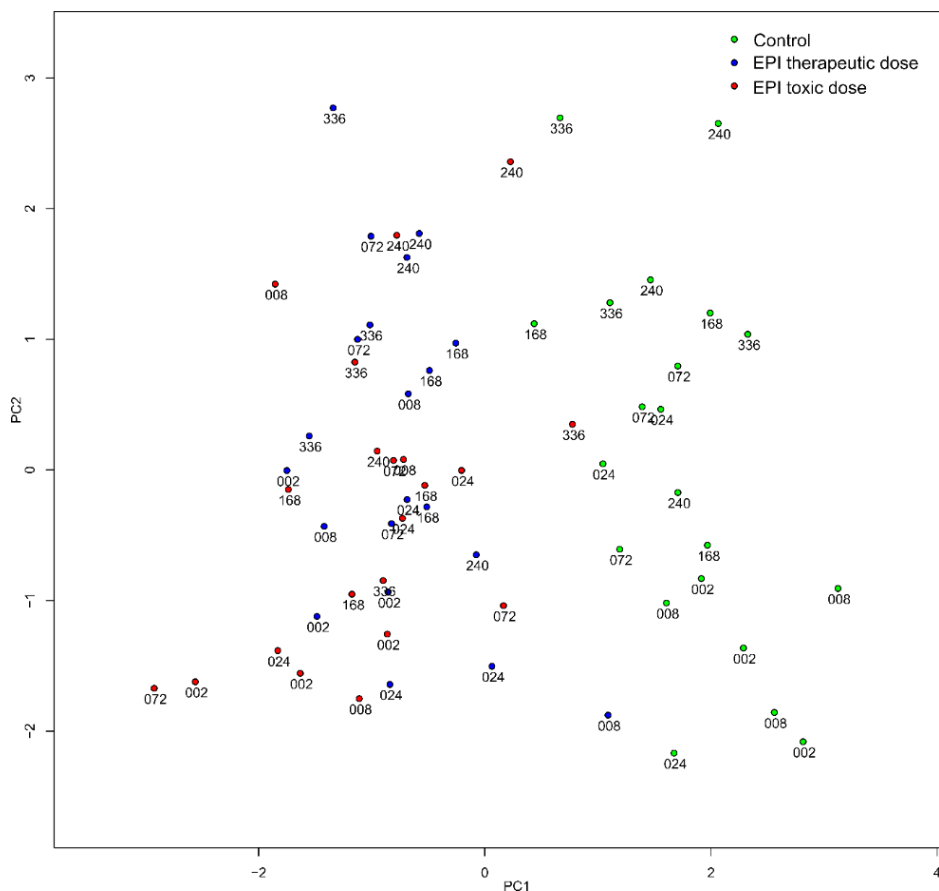


Figure 2-2: The PCA plots of all EPI-treated and control samples. The numbers are the exposure times in hours.

The DNA methylation analysis between all EPI-treated and control samples unveiled 161,356 unique DMRs corresponding to 19,825 genes. Per gene, the high amount of DMRs, especially in the promoter region, suggests a strong influence of the EPI treatment on the DNA methylation status of that gene compared to control samples. After filtering, 966 genes were in the top 5% of genes that had the highest number of DMR regions across gene regions (Figure 2-3A), while 47 out of these 966 genes had the highest number DMRs in the promoter region (Figure 2-3B). We also enhanced the filtering criteria by only selecting genes that had absolute average $\log_2FC \geq 0.5$. Based on all these selection criteria, the workflow derived 35 selected genes with strong methylated alterations from the enormous number of detected DMRs (Figure 2-4).

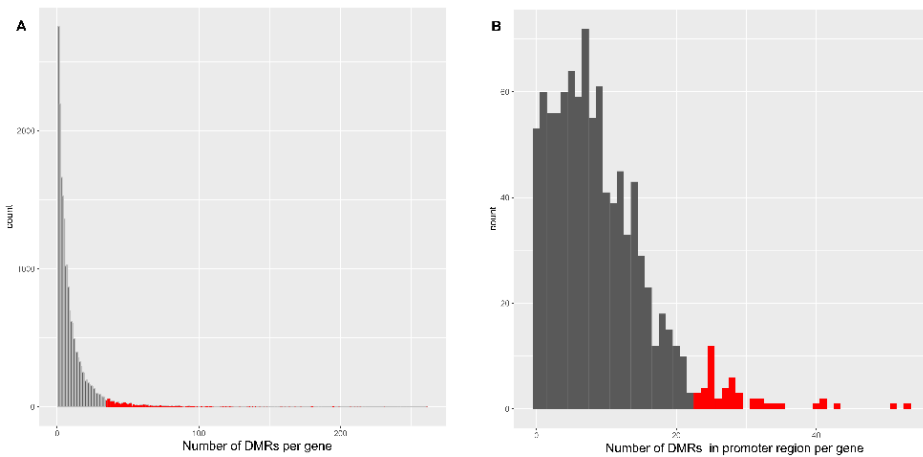


Figure 2-3: The distribution of genes according to the number of DMRs per gene (A) and the number of DMRs in the promoter region per gene (B) after the differentially methylated analysis between all EPI-treated samples and control samples. The red area indicates selected genes for the next analysis steps.

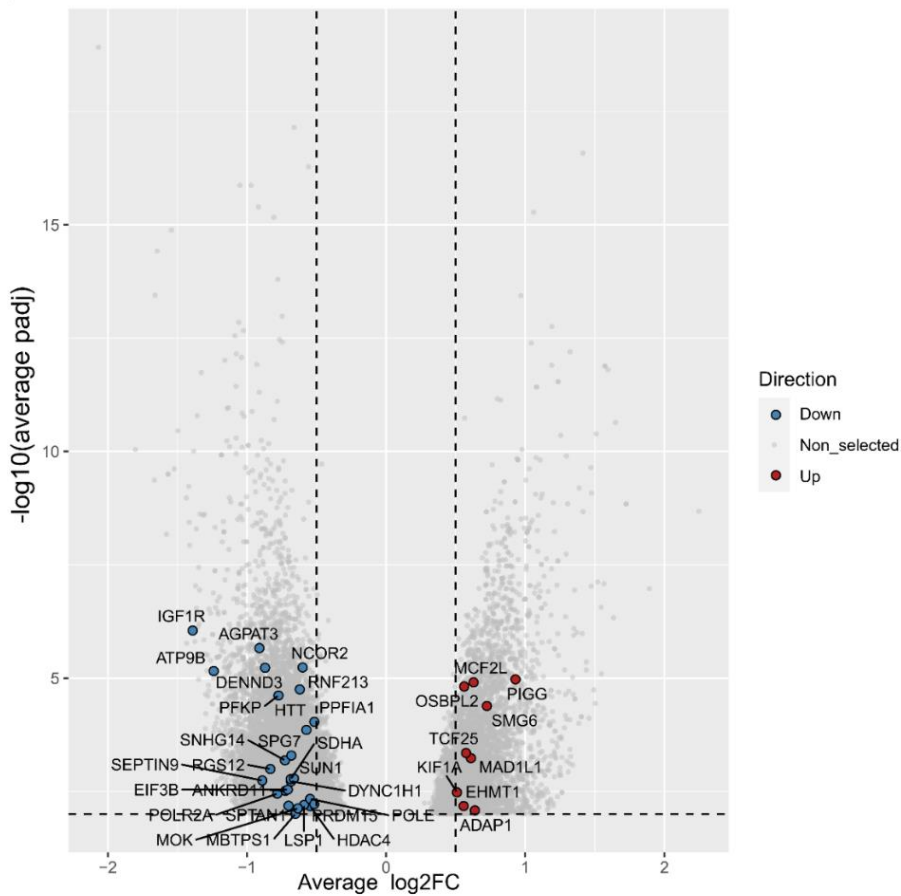


Figure 2-4: The volcano plot of the differentially methylated genes in EPI-treated samples compared to control. Genes that satisfied all the selection criteria were highlighted in blue and red for hypo-methylated and hyper-methylated status respectively.

Discussion

In this chapter 2-1, we developed a workflow to detect genes that had strong methylation alterations based on the traditional QSEA R package. While several tools have been developed to analyze MeDIP-seq data, QSEA is the most recent one that offers a straight procedure to inspect MeDIP-seq data [16]. This workflow not only obtains DMRs but also filters top candidate genes with strong methylation alterations (Figure 2-1). Thereby, it could suggest potential genes for further investigation and certainly elevate the application of DNA methylation analysis.

Furthermore, this chapter 2-1 has demonstrated initial results about the change of genome-wide DNA methylation in cardiac microtissues exposed to EPI. Hence, the DNA methylation status in cells can also be influenced by disease progression and drug treatment as mentioned in previous studies [23,24]. DNA methylation can, therefore, serve as a potential approach to understand drug mechanisms as well as to discover biomarkers for drug safety and early drug screening. The MeDIP-seq data analysis of cardiac microtissues under EPI exposure is given in detail in the coming chapter 2-2.

Chapter 2.2: Epirubicin alters DNA methylation profiles related to cardiotoxicity

Objectives of the study

Epirubicin (EPI) is an important anticancer drug that is widely used in multiple types of cancer treatments even though its utilization leads to a high risk of heart failure [25]. To improve the therapeutic application of EPI, clinicians restricted its dose usage, because a very high dose of EPI (around 900 mg/m²) can cause acute heart failure circumstances. However, long-term observational studies have shown that using EPI also at lower doses can still provoke substantial cardiotoxicity [26,27]. Even though researchers have suggested updated signal transduction models [28], and studied the side effect of EPI on cardiomyocytes on gene expression [29] and protein [30] levels, deeper insights into EPI toxic mechanisms are still in demand. Since epigenetic modification can influence the functional state of genome regions without changing the DNA sequence, studying epigenetic signals could be beneficial to understanding EPI-related toxic mechanisms. Research on EPI in gastric cancer demonstrates that the changes in DNA methylation can be beneficial to understanding the biological mechanism of drug resistance [31].

In this chapter 2-2, we intently focused on analyzing and interpreting the MeDIP-seq data deriving from cardiac microtissues exposed to EPI. While researchers usually analyze DNA methylation at the DMRs level [32,33], we would like to apply the recent MeDIP-seq analysis workflow as developed in the foregoing chapter 2-1 that targets candidate methylated genes. By utilizing this bioinformatics workflow, we can identify candidate genes that had strong DNA methylation alterations related to the EPI-induced cardiotoxicity mechanism. Furthermore, we also examine how changes in DNA methylation of those candidate genes affects their expressions on the transcriptome level. The outcome of this study could suggest potential genes with epigenetic regulations for EPI-induced cardiotoxicity research.

Analysis procedures

MeDIP-seq data analysis

We utilized the same MeDIP-seq data analysis workflow mentioned in previous chapter 2-1. In this chapter 2-2, we performed the DNA methylation analysis not only between all EPI-treated and control samples but also between EPI either

therapeutic or toxic-treated samples compared to controls. All the filtering steps were used with the default settings from the bioinformatics workflow (Figure 2-1). The overlapping genes within different DNA methylation analyses were identified using InteractiVenn tools [34]. The Gene Ontology (GO) enrichment analysis on differential methylated gene sets was performed by the PANTHER version 14 using the GO molecular function annotation dataset, no correction after Fisher's Exact test, and default reference for Homo sapiens [35].

RNA sequencing data analysis

After RNA sequencing, the adapter sequences in the paired-end sequenced reads were removed by using Trimmomatic version 0.36 [36]. The sample sequencing quality was examined by FastQC version 0.11.7 [37], and summarized by MultiQC [38] before and after trimming the reads. Then, the RNA sequencing data were mapped to the human genome version GRCh38.p12, Ensembl Archive Release 12 93 [39] using RSEM version 1.3.1 [40] and Bowtie2 version 2.3.4.1 [41] with the paired-end option. All samples had more than 5 million read counts and were used for further analysis. Due to the limited amount of microtissues after 336 hours exposed to EPI-related toxic dose, only RNA data were available from samples treated with toxic dose at 2, 8, 24, 72, 168, and 240 hours of exposure. Thereafter, the RNA read counts between EPI-treated and control samples were normalized using the “DEseq2” package [42].

Results

DNA methylation analysis

As mentioned in the previous chapter 2-1, the DNA methylation profiles were dissimilar among EPI-treated and control samples (Figure 2-2). The DNA methylation analysis unveiled 161,356 unique DMRs corresponding to 19,825 genes between all EPI-treated and control samples. After different filtering steps, the workflow derived 35 genes with strong methylated alterations from the enormous number of detected DMRs (Table 2-1, Figure 2-4).

Table 2-1: The numbers of DMRs and selected methylated genes in EPI-treated samples compared to control

Samples compared to controls	All EPI-treated samples	EPI therapeutic-treated samples	EPI toxic-treated samples
DMRs regions	161,356	169,214	42,226
Annotated genes	19,825	20,270	11,637
Top 5% genes with the highest number of DMRs across the gene regions	966	987	521
Top 5% genes with the highest number of DMRs in the promoter region	47	46	23
Genes with average log ₂ FC ≥ 0.5	35	37	19
Hypo-methylated genes	26	22	18
Hyper-methylated genes	9	15	1

Note: DMR, differentially methylated region; EPI, epirubicin.

We deployed the same DNA methylation analysis procedure to analyze the DMRs and corresponding differential methylated genes between EPI-treated and control samples per dose. The workflow again distilled the excessive number of detected DMRs into a shortlist of strong differentially methylated genes (Table 2-1, Figure 2-5). Intriguingly, EPI therapeutic-treated samples showed a higher number of DMRs regions and gradually a higher number of strong differentially methylated genes compared to EPI toxic-treated samples. The DNA methylation analysis between therapeutic-treated samples compared to controls indicated 37 candidates comprising 15 hyper-methylated and 22 hypo-methylated genes. This is quite compatible with the outcome of DNA methylation analysis between all EPI-treated samples and controls, which had 35 candidates including 9 hyper-methylated and 26 hypo-methylated genes. However, the DNA methylation analysis between EPI toxic-treated samples compared to controls demonstrated 19 candidates, and only one candidate gene, *SPG7*, was hyper-methylated (Table 2-1). The GO gene enrichment analysis demonstrated that the differential methylated genes are involved in different functional classes (Table 2-2). While a major of these genes were concentrated in the catalytic activity (GO:0003824) and binding (GO:0005488) groups, the rest engages in regulator and transporter activities.

Furthermore, we identified the overlapping differentially methylated genes within all foregoing DNA methylation analyses. This resulted in 8 common candidate genes in both EPI therapeutic and toxic-treated samples compared to

controls. However, one of these genes, *ATP11A*, was not recognized as a differentially methylated candidate gene when we compared the DNA methylation profiles between all EPI-treated samples and control (Figure 2-4, Table 2-3). Interestingly, while *SPG7* was hyper-methylated at the EPI toxic-treated condition (average $\log_2FC = 0.84$) (Table S 2-3), it was hypo-methylated at EPI therapeutic-treated condition (average $\log_2FC = -0.69$) (Table S 2-2) and at all EPI-treated samples compared to controls (average $\log_2FC = -0.68$) (Table S 2-1). By contrast, while the rest of the candidate genes at the EPI toxic-treated conditions were in hypo-methylated status, some of them were in hyper-methylated status at other conditions. For instance, *MAD1L1* was hyper-methylated if comparing EPI therapeutic-treated or all EPI-treated samples to control. On the other hand, *NCOR2* was hyper-methylated at the EPI therapeutic-treated condition but hypo-methylated if comparing all EPI-treated samples or EPI toxic-treated samples to control (Table S 2-3). Thus, specific doses and how the MeDIP-seq data were processed had influenced the outcome of the DNA methylation analysis.

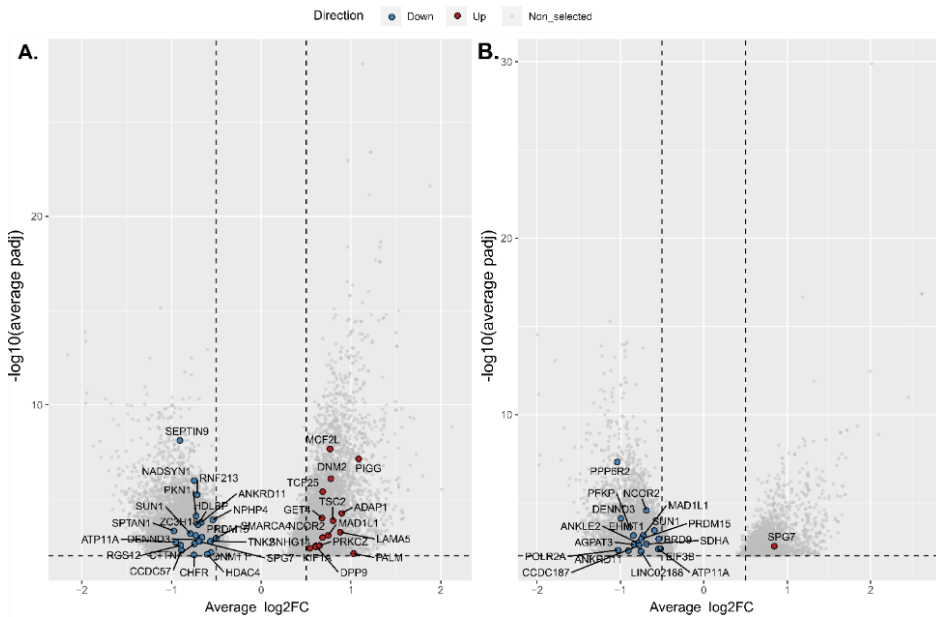


Figure 2-5: Differential methylated comparison between control and either EPI therapeutic-treated (A) or toxic-treated samples (B). EPI, epirubicin; \log_2FC , \log_2 fold change.

Table 2-2: GO enrichment analysis for differentially methylated gene set per treatment condition compared to control

Gene Ontology (GO) Enrichment	Number of differentially methylated genes		
	All EPI-treated samples	EPI therapeutic-treated samples	EPI toxic-treated samples
Transporter activity (GO:0005215)	3	1	1
Transcription regulator activity (GO:0140110)	2	2	2
Catalytic activity (GO:0003824)	13	12	5
Molecular function regulator (GO:0098772)	1	2	1
ATP-dependent activity (GO:0140657)	3	3	1
Molecular adaptor activity (GO:0060090)	2	2	1
Binding (GO:0005488)	14	10	6
Cytoskeletal motor activity (GO:0003774)	2	1	-
Translation regulator activity (GO:0045182)	1	-	1
Molecular transducer activity (GO:0060089)	1	-	-

Notes: EPI, epirubicin; -, not applicable.

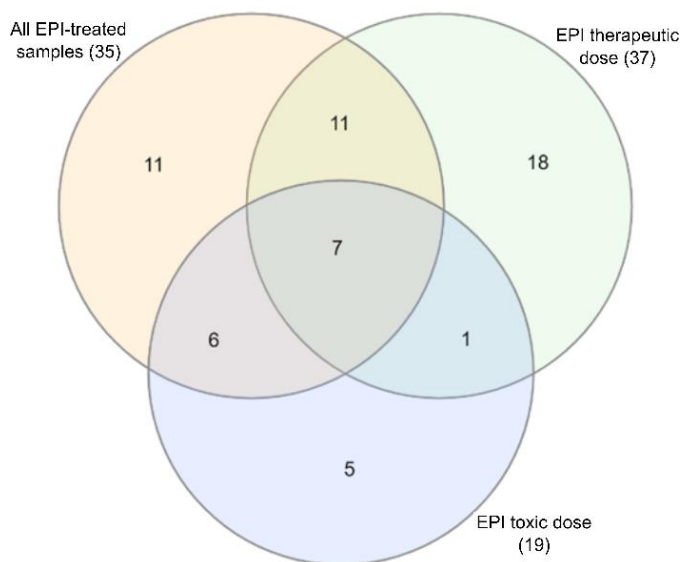


Figure 2-6: The Venn diagram of differentially methylated genes resulting from different DNA methylation comparisons. This includes all EPI-treated, EPI therapeutic-treated, and EPI toxic-treated samples compared to controls.

Table 2-3: The differentially methylated genes resulting from three DNA methylation analyses compared to control

Samples compared to controls	All EPI-treated samples	EPI therapeutic-treated samples	EPI toxic-treated samples
Overlapping differential methylated genes	7 <i>MADI1L1, PRDM15, NCOR2, SUN1, SPG7, ANKRD11, DENND3</i>		
Other differential methylated genes	28	30	12
	<i>PIGG, SMG6, ADAP1, MCF2L, TCF25, OSBPL2, EHMT1, KIF1A, PPF1A1, HDAC4, POLE, HTT, LSP1, RNF213, MOK, MBTPS1, DYNC1H1, SDHA, SPTAN1, EIF3B, SNHG14, PFKP, POLR2A, RGS12, SEPTIN9, AGPAT3, ATP9B, IGF1R</i>	<i>ADAP1, ATP11A, CCDC57, CHFR, CTTN, DNMT2, DNMT1, DPP9, GET4, HDAC4, HDLBP, KIF1A, LAMA5, MCF2L, NADSYN1, NPHP4, PALM, PIGG, PKN1, PRKCZ, RGS12, RNF213, SEPTIN9, SMARCA4, SNHG14, SPTAN1, TCF25, TNK2, TSC2, ZC3H18</i>	<i>AGPAT3, ANKLE2, ATP11A, BRD9, CCDC187, EHMT1, EIF3B, LINC02188, PFKP, POLR2A, PPP6R2, SDHA</i>

From DNA methylation to gene expression

The MeDIP data and the transcriptome data were harvested from the same microtissues exposed to EPI; therefore, we were able to evaluate the influence of changing DNA methylation status on the gene expression. The gene expression of 8 common differential methylated genes in EPI therapeutic and toxic-treated conditions is shown in Figure 2-7. The methylation status of some genes, such as *SUN1*, demonstrated a coherent relation with its gene expression on the transcriptome level. *SUN1* was hypo-methylated with $\log_2FC_{avg} = -0.66, -0.79,$ and -0.59 for the DNA methylation analysis between all EPI-treated, EPI therapeutic-treated, and EPI toxic-treated samples versus control, respectively. The expression of *SUN1* in almost all EPI-treated samples was higher than its expression in corresponding control samples (Figure 2-7). Nevertheless, in some genes, the regulation at the DNA methylation level could not entirely be related to the gene expression at the transcriptome level. For example, *SPG7* was hyper-methylated in the EPI toxic-treated condition (Table S 2-3) and hypo-methylated in EPI therapeutic-treated condition (Table S 2-2); however, on the transcriptome level, *SPG7* was overexpressed in samples treated with both EPI doses compared to its expression in controls (Figure 2-7). Similarly, *MADI1L1* and *NCOR2* were

hyper-methylated during EPI therapeutic treatments (Table S 2-2) and hypo-methylated at EPI toxic treatments (Table S 2-3) but their gene expressions did not clearly reflect this (Figure 2-7). Hence, the change in methylated status could explain the expression regulation of certain genes, but not for all affected genes in the human genome.

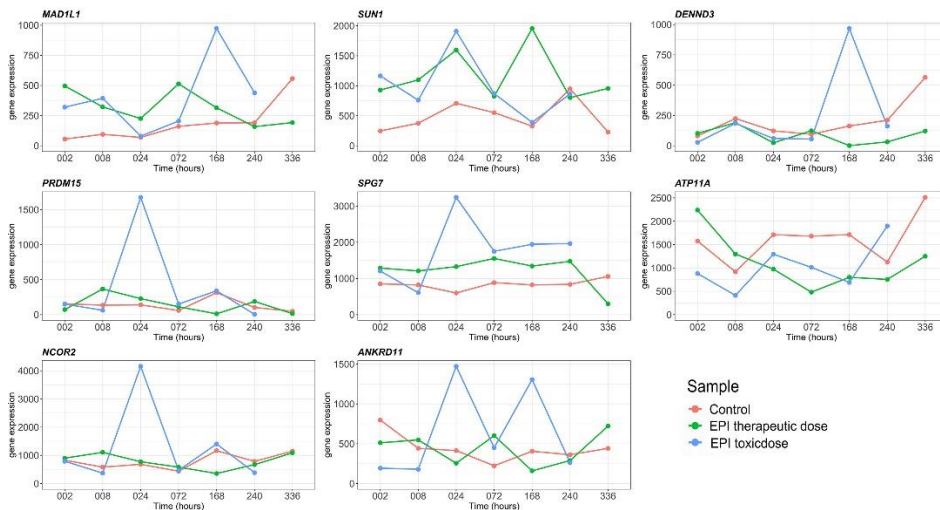


Figure 2-7: The gene expression of overlapped differentially methylated genes between EPI therapeutic and toxic-treated conditions.

We also demonstrated the expression of some differential methylated genes that demonstrated distinct expressions on the transcriptome in EPI-treated samples from control. For instance, *DPP9* was hyper-methylated at the EPI therapeutic-treated condition (average $\log_2FC = 0.61$, Table S 2-2), and its gene expression in EPI-treated samples was mostly lower than that in control samples (Figure 2-8A). While *SMARCA4*, *HDAC4*, *PKN1*, and *RGS12* were hypo-methylated at the EPI therapeutic-treated condition (average $\log_2FC = -0.50$, -0.60 , -0.72 , and -0.90 respectively, Table S 2-2), only *SMARCA4* and *PKN1* were up-regulated on the transcriptome level in EPI therapeutic-treated samples compared to control across roughly all time of exposure (Figure 2-8A). At the EPI toxic-treat condition, *SDHA* and *POLR2A* (average $\log_2FC = -0.69$ and -1.03 respectively, Table S 2-3) were hypo-methylated at the DNA methylation level and consequently showed noticeable up-regulated on the transcriptome level (Figure 2-8B). By contrast, although *AGPAT3* was hypo-methylated at the EPI toxic-treated condition compared to control (average $\log_2FC = -0.84$, Table S3), *AGPAT3* had lower RNA expression levels after 24 hours of EPI exposure compared to corresponding control samples (Figure 2-8B).

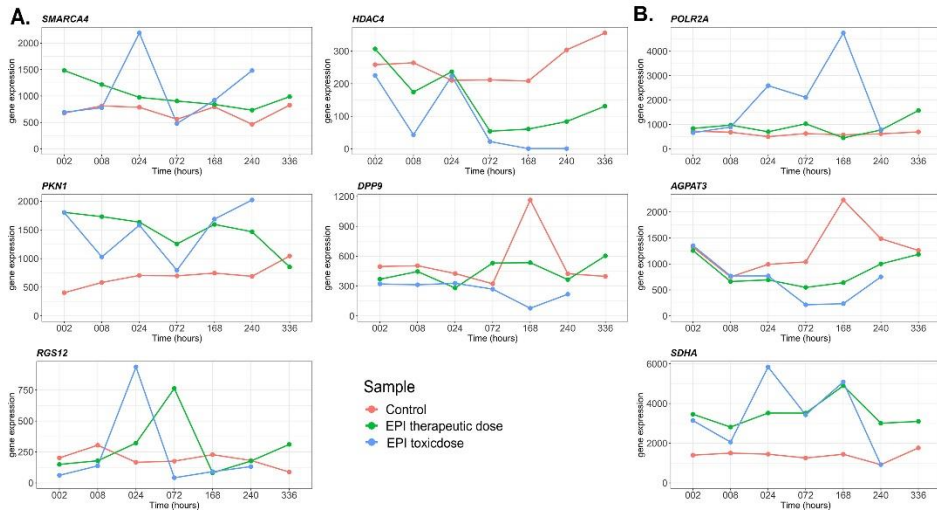


Figure 2-8: The gene expression of differentially methylated genes in EPI therapeutic (A) and toxic-treated (B) conditions.

Discussion

This chapter 2-2 demonstrated the undeniable impact of EPI on the DNA methylation profiles as well as the changing of gene expression as the consequence of DNA methylation alterations under EPI exposure in *in vitro* human cardiac microtissues. By using the MeDIP-seq analysis workflow developed in chapter 2-1, we were able to refine the extensive amount of detected DMRs into a shortlist of strongly differentially methylated genes. While the same data analysis procedure was applied, the outcome in each analysis step was different among particular EPI dose conditions (Table 2-1) and resulted in slightly different lists of candidate genes (Table 2-3). For example, we detected 35 candidate genes in all EPI-treated samples compared to control, while we encountered 37 and 15 candidate genes between EPI therapeutic and toxic dose-treated conditions compared to controls, respectively. Although there were still overlapped genes among these DNA methylation analyses (Table 2-3, Figure 2-6), it is clear that various sample grouping approaches can generate different outcomes. It also shows that dose-dependence can lead to different numbers of DMRs and corresponding genes. Besides, the massive change in DNA methylation could also be the signal of genomic imprinting interruption.

The differential methylated genes are involved in different cellular functions such as molecular interactions, regulations, and transportations (Table 2-2). In

particular, some genes changed their methylation status under EPI treatment, which is potentially associated with EPI cardiotoxic adverse effects. For instance, *SMARCA4* (also known as *BRG1*) was hypo-methylated and provoked up-regulation on the transcriptome level in EPI therapeutic-treated samples (Figure 2-7). This gene has a critical role in regulating heart muscle development and disease via the myosin heavy chain switch. *SMARCA4* is generally turned off in cardiomyocytes; however, it is re-activated under stress and its level is correlated with hypertrophic cardiomyopathy severity [43]. Thus, EPI could afflict the expression of *SMARCA4* via DNA methylation alterations and stimulate cardiac dysfunctions. The gene expression of *PKN1* was also up-regulated by hypo-methylation in EPI therapeutic-treated conditions (Figure 2-7). The *PKN1* activation can initiate cardiac hypertrophy and fibrosis-associated gene expression and can be involved in heart failure development [44]. *RGS12* was hypo-methylated and up-regulated on the transcriptome level in a part of the samples exposed to EPI due to different times of exposure (Figure 2-7). A rodent study has demonstrated that *RGS12* contributes to angiotensin II-induced hypertrophy, and its over-expression has been observed in cardiac hypertrophy and heart failure pathology [45]. Another hypo-methylated gene, *HDAC4*, is known for rapid histone methylation regulation in response to elevated cardiac load [46]. On the other hand, *DPP9* was hyper-methylated and led to a lower gene expression in EPI therapeutic-treated samples compared to control (Figure 2-7); the drug-induced *DPP9* inhibition can impair the CaMKII-PLB and PKC signaling and cause cardiac dysfunction [47]. All these genes were differentially methylated and potentially related to cardiotoxicity even under EPI therapeutic-treated conditions; this is consistent with the observation in cancer survivors who underwent EPI treatment have a higher risk of late-onset cardiac disease [26,27].

Eight genes were consistently differently methylated across EPI-treated samples as well as some genes specifically showed strong methylated alterations at the EPI toxic-treated condition (Table 2-3, Figure 2-6, Figure 2-7, Figure 2-8). *NCOR2* was hyper-methylated in the EPI therapeutic-treated condition but hypo-methylated in the EPI toxic-treated conditions, respectively. As a nuclear receptor, *NCOR2* can regulate the expression of other genes and influence the metabolic oxidative balance in cardiomyocytes [48]. A study has suggested that the differential methylation signature of *NCOR2* in CD4+ T cells could be a non-invasive biomarker to identify pulmonary arterial hypertension patients [49]. Furthermore, *SDHA*, *POLR2A*, and *AGPAT3* genes were hypo-methylated at the EPI toxic-treated condition and also play important roles in cardiac dysfunctions. While *POLR2A* has been considered a stable heart failure reference gene across

rodents and humans [50], *SDHA* participates in the tricarboxylic acid cycle and mitochondrial respiratory chain. The change of *SDHA* expression, due to the methylation modification at the DNA level (Table S 2-3, Figure 2-7), can impact mitochondrial acetyl-CoA homeostasis and energy metabolism which contribute to heart failure [51]. *AGPAT3* is also an enzyme involved in mitochondrial oxidation; thereupon, the change in its expression can consequently affect ATP production [52]. Thus, *SDHA* and *AGPAT3* can be potential drivers in the EPI-induced energy metabolic dysregulation and contribute to heart failure development.

In conclusion, measuring genome-wide DNA methylation profiles could provide further insights into the EPI-induced cardiotoxicity. Differential DNA methylation alterations could offer a supportive explanation for understanding EPI cardiotoxic mechanisms along with transcriptome and proteome study. In this chapter 2-2, a handful of genes that had strong EPI-related DNA methylation alterations were named as candidates for further investigation. A part of them, such as *SMARCA4*, *PKN1*, *RGS12*, *DPP9*, *NCOR2*, *SDHA*, *POLR2A*, and *AGPAT3*, has disclosed their roles in cardiac dysfunctions as well as potential biomarkers for heart failure in different contexts. This is coherent with the well-known EPI cardiotoxicity adverse effects. Those genes, together with other detected candidate genes, can be candidates for further investigations of EPI-related toxic mechanisms.

General remarks

In this chapter, we first established a DNA methylation analysis workflow and then employ this workflow to analyze the DNA methylation profile in cardiac microtissues exposed to EPI compared to controls. We were able to inspect the DNA methylation alterations of several genes, and how changes in DNA methylation could affect their gene expression at the transcriptome level in relation to EPI-induced cardiotoxicity.

From the bioinformatics perspective, we established a workflow built on the QSEA package to detect genes that had strong methylation alterations. While several tools have been developed to analyze MeDIP-seq data, there is still a demand for interpreting the differentially methylated status among samples from genome regions (DMRs) toward the gene levels. Our bioinformatics workflow built on the recent QSEA R package can analyze the DNA methylation profiles not only on the DMRs level but also on the gene levels as well as suggest candidate genes that have strong DNA methylation alteration between samples. The workflow is written in R and publicly available via Github; thereby, this could aid researchers with the DNA methylation analysis procedure.

For the drug side effect, we explored the genome-wide DNA methylation alteration under EPI treatment. EPI is a popular chemotherapeutic agent with cardiotoxic effects [26]. Although different studies have investigated the impact of EPI on cellular mechanisms on the transcriptome and protein levels [29,30], there is not much research on EPI-induced epigenetic modifications. By analyzing the genome-wide methylation status of human cardiac tissues exposed to EPI, we identified candidate genes that had strong DNA methylation alterations related to the EPI-induced cardiotoxicity mechanism. According to transcriptome data from the same samples, we also examined how changes in DNA methylation of those genes affects their gene expressions on the transcriptome level. In some genes, their gene expressions were up or down-regulated, which compatibly reflected the DNA methylation regulation. Together with other differentially methylated genes, these genes can be interesting targets for further investigation in EPI-induced toxic mechanisms. Thus, studying epigenetic modification such as DNA methylation can help to understand the EPI-related toxic mechanisms in cardiac tissue.

Therefore, DNA methylation can also be a useful tool to reveal drug-induced adverse side effects. Studying DNA methylations, as epigenetic signals, can be allied with gene and protein expression research in order to produce a multi-

layered and extensive view of biological mechanisms, in this case, EPI-induced mechanisms. This chapter 2 not only establishes a new DNA methylation analysis workflow but also provides new insight into the gene expression and regulation at DNA methylated level related to EPI-induced cardiotoxicity.

Supplementary Materials

Data & code accessibility

The data is deposited in BioStudies (<http://www.ebi.ac.uk/biostudies>) under accession numbers S-HECA433 and S-HECA434 for the MeDIP-seq data and S-HECA11 for the RNA-seq data. The R code is available on Github (<https://github.com/NhanNguyenooo/MeDIP>).

Supplementary figures

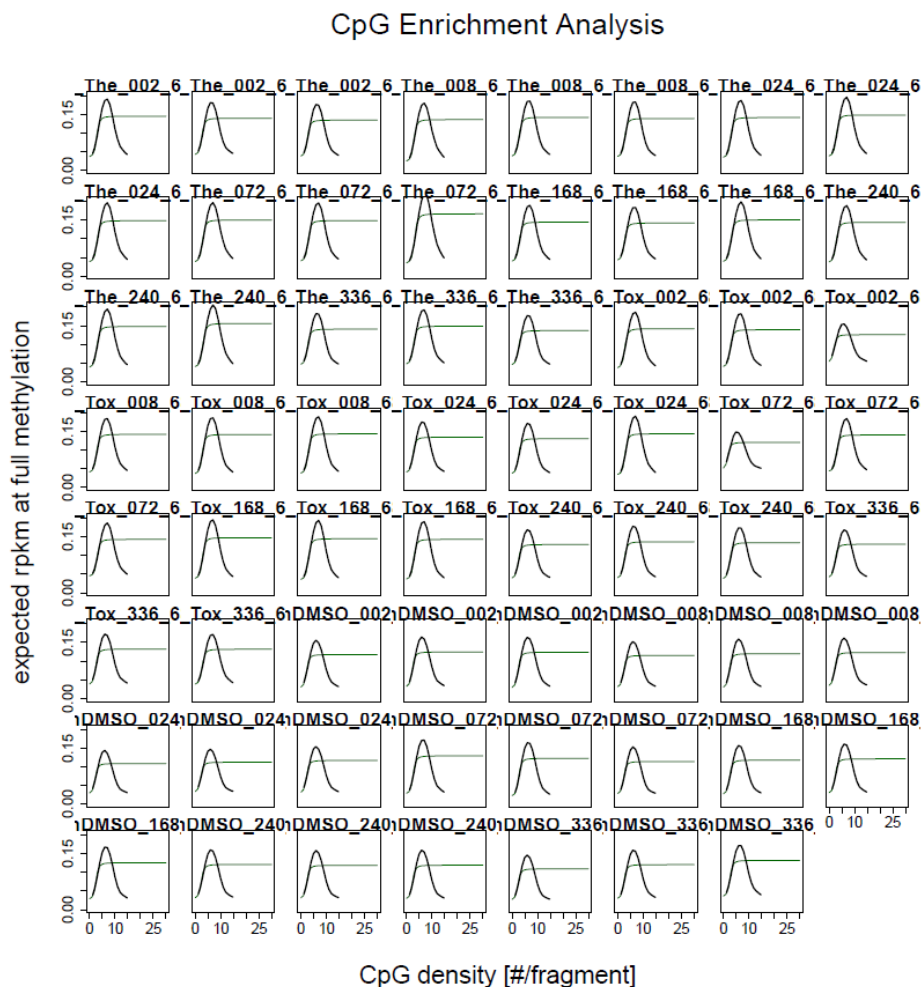


Figure S 2-1: The sample-wise CpG density-dependent enrichment profiles. The average enrichment profile of samples is depicted in black, and the fitted sigmoidal function is in green. Samples with flat profiles might indicate low enrichment efficiency or poor

agreement with the calibration data. In here, all samples show sufficient enrichment profiles

Supplementary tables

Table S 2-1: Differentially methylated genes between all EPI-treated and control samples.

Symbol	The average of log2FC values	The average of p-values	The average of adjusted p-values
<i>PIGG</i>	0.929599	2.30E-08	1.07E-05
<i>SMG6</i>	0.723932	1.28E-07	4.13E-05
<i>ADAP1</i>	0.637901	0.000161	0.008325
<i>MCF2L</i>	0.628397	2.79E-08	1.24E-05
<i>MAD1L1</i>	0.609735	4.15E-06	0.000586
<i>TCF25</i>	0.574223	2.92E-06	0.00045
<i>OSBPL2</i>	0.560973	3.63E-08	1.53E-05
<i>EHMT1</i>	0.557024	0.000117	0.006664
<i>KIF1A</i>	0.508996	4.44E-05	0.003327
<i>PPFIA1</i>	-0.51613	3.63E-07	9.23E-05
<i>PRDM15</i>	-0.51623	0.000103	0.006067
<i>HDAC4</i>	-0.54593	0.000119	0.006739
<i>POLE</i>	-0.54641	7.05E-05	0.004642
<i>HTT</i>	-0.57285	6.23E-07	0.000139
<i>LSP1</i>	-0.58915	0.000108	0.006307
<i>NCOR2</i>	-0.60021	1.08E-08	5.84E-06
<i>RNF213</i>	-0.62046	4.39E-08	1.78E-05
<i>MOK</i>	-0.63346	0.000139	0.007539
<i>MBTPS1</i>	-0.65155	0.000206	0.009931
<i>SUN1</i>	-0.6618	1.68E-05	0.001648
<i>SPG7</i>	-0.68089	3.48E-06	0.000513
<i>DYNC1H1</i>	-0.68341	2.05E-05	0.001904
<i>SDHA</i>	-0.68512	1.75E-05	0.001695
<i>SPTAN1</i>	-0.70132	0.000115	0.006575
<i>EIF3B</i>	-0.70674	3.67E-05	0.002903
<i>ANKRD11</i>	-0.72403	3.98E-05	0.003077
<i>SNHG14</i>	-0.72663	4.81E-06	0.000654
<i>PFKP</i>	-0.7734	6.49E-08	2.42E-05
<i>POLR2A</i>	-0.78209	4.87E-05	0.003559
<i>RGS12</i>	-0.83223	8.72E-06	0.001017
<i>DENND3</i>	-0.86963	1.10E-08	5.93E-06

<i>SEPTIN9</i>	-0.88978	1.93E-05	0.00182
<i>AGPAT3</i>	-0.91077	3.19E-09	2.19E-06
<i>ATP9B</i>	-1.23879	1.35E-08	6.99E-06
<i>IGF1R</i>	-1.39002	1.07E-09	8.94E-07

Table S 2-2: Differentially methylated genes between EPI therapeutic-treated and control samples.

Symbol	The average of log ₂ FC values	The average of p-values	The average of adj-p-values
<i>PIGG</i>	1.086588	9.42E-11	7.38E-08
<i>PALM</i>	1.028544	0.000176	0.007639
<i>ADAP1</i>	0.896716	2.80E-07	5.63E-05
<i>LAMA5</i>	0.880503	5.24E-06	0.000561
<i>TSC2</i>	0.800732	8.68E-07	0.000138
<i>DNM2</i>	0.775342	1.61E-09	8.19E-07
<i>MCF2L</i>	0.767292	2.26E-11	2.18E-08
<i>MAD1L1</i>	0.746182	8.77E-06	0.000832
<i>TCF25</i>	0.684485	1.09E-08	4.01E-06
<i>NCOR2</i>	0.683063	1.25E-05	0.001087
<i>GET4</i>	0.677513	5.74E-07	9.98E-05
<i>PRKCZ</i>	0.650161	4.55E-05	0.002857
<i>DPP9</i>	0.60892	4.88E-05	0.003008
<i>KIF1A</i>	0.608094	5.71E-05	0.003375
<i>SNHG14</i>	0.542938	7.11E-05	0.003971
<i>SMARCA4</i>	-0.50261	1.38E-05	0.001173
<i>NPHP4</i>	-0.53957	7.72E-07	0.000126
<i>DNMT1</i>	-0.56003	0.000146	0.006672
<i>TNK2</i>	-0.56569	2.30E-05	0.001718
<i>HDAC4</i>	-0.59993	0.000203	0.008475
<i>PRDM15</i>	-0.65886	1.20E-05	0.001054
<i>ANKRD11</i>	-0.66849	1.14E-06	0.000171
<i>SPG7</i>	-0.68964	2.34E-05	0.001741
<i>DENND3</i>	-0.70321	1.79E-05	0.001424
<i>HDLBP</i>	-0.70592	1.64E-06	0.000227
<i>RNF213</i>	-0.71511	1.70E-08	5.78E-06
<i>PKN1</i>	-0.72406	4.11E-07	7.66E-05
<i>ZC3H18</i>	-0.72933	8.70E-06	0.000827

<i>CTTN</i>	-0.74312	3.48E-05	0.002345
<i>NADSYN1</i>	-0.74643	2.09E-09	1.02E-06
<i>CHFR</i>	-0.74919	0.000224	0.009081
<i>SUN1</i>	-0.78903	6.66E-06	0.000673
<i>CCDC57</i>	-0.88915	0.000104	0.005215
<i>RGS12</i>	-0.8961	4.41E-05	0.002793
<i>SEPTIN9</i>	-0.90718	6.73E-12	7.74E-09
<i>ATP11A</i>	-0.95157	2.68E-05	0.001927
<i>SPTAN1</i>	-0.97206	4.34E-06	0.000485

Table S 2-3: Differentially methylated genes between EPI toxic-treated and control samples.

Symbol	The average of log ₂ FC values	The average of p-values	The average of adj-p-values
<i>SPG7</i>	0.844868	8.99E-06	0.002849
<i>EIF3B</i>	-0.51764	1.47E-05	0.004017
<i>BRD9</i>	-0.54008	2.46E-06	0.001116
<i>ATP11A</i>	-0.54375	1.55E-05	0.00417
<i>SUN1</i>	-0.59181	5.68E-07	0.00037
<i>SDHA</i>	-0.68544	6.08E-06	0.002158
<i>NCOR2</i>	-0.68822	2.10E-08	2.69E-05
<i>MAD1L1</i>	-0.72482	1.29E-06	0.000693
<i>PRDM15</i>	-0.74761	2.44E-06	0.001112
<i>LINC02188</i>	-0.75217	2.47E-05	0.005765
<i>CCDC187</i>	-0.77581	7.89E-06	0.0026
<i>ANKLE2</i>	-0.78812	5.10E-06	0.001901
<i>AGPAT3</i>	-0.83796	6.87E-06	0.002355
<i>EHMT1</i>	-0.84411	1.34E-06	0.000712
<i>PFKP</i>	-0.8911	3.23E-07	0.000239
<i>ANKRD11</i>	-0.90485	2.09E-05	0.005137
<i>DENND3</i>	-0.99421	7.61E-08	7.70E-05
<i>POLR2A</i>	-1.02903	1.89E-05	0.004781
<i>PPP6R2</i>	-1.03661	1.20E-11	4.85E-08

References

- [1] E.M. Martin, R.C. Fry, Environmental Influences on the Epigenome: Exposure-Associated DNA Methylation in Human Populations, *Annual Review of Public Health* 39 (2018) 309-333. <https://doi.org/10.1146/annurev-publhealth-040617-014629>
- [2] L.D. Moore, T. Le, G. Fan, DNA methylation and its basic function, *Neuropsychopharmacology* 38 (2013) 23-38. <https://doi.org/10.1038/npp.2012.112>
- [3] G. Lev Maor, A. Yearim, G. Ast, The alternative role of DNA methylation in splicing regulation, *Trends in Genetics* 31 (2015) 274-280. <https://doi.org/10.1016/j.tig.2015.03.002>
- [4] Z. Jin, Y. Liu, DNA methylation in human diseases, *Genes & Diseases* 5 (2018) 1-8. <https://doi.org/10.1016/j.gendis.2018.01.002>
- [5] R.R. Meehan, J.P. Thomson, A. Lentini, C.E. Nestor, S. Pennings, DNA methylation as a genomic marker of exposure to chemical and environmental agents, *Current Opinion in Chemical Biology* 45 (2018) 48-56. <https://doi.org/10.1016/j.cbpa.2018.02.006>
- [6] I. Rauluseviciute, F. Drabløs, M.B. Rye, DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis, *Clinical Epigenetics* 11 (2019) 193. <https://doi.org/10.1186/s13148-019-0795-x>
- [7] O. Taiwo, G.A. Wilson, T. Morris, S. Seisenberger, W. Reik, D. Pearce, S. Beck, L.M. Butcher, Methylome analysis using MeDIP-seq with low DNA concentrations, *Nature Protocols* 7 (2012) 617-636. <https://doi.org/10.1038/nprot.2012.012>
- [8] D. Beck, M. Ben Maamar, M.K. Skinner, Genome-wide CpG density and DNA methylation analysis method (MeDIP, RRBS, and WGBS) comparisons, *Epigenetics* (2021) 1-13. <https://doi.org/10.1080/15592294.2021.1924970>
- [9] L. Kuepfer, O. Clayton, C. Thiel, H. Cordes, R. Nudischer, L.M. Blank, V. Baier, S. Heymans, F. Caiment, A. Roth, D.A. Fluri, J.M. Kelm, J. Castell, N. Selevsek, R. Schlapbach, H. Keun, J. Hynes, U. Sarkans, H. Gmuender, R. Herwig, S. Niederer, J. Schuchhardt, M. Segall, J. Kleinjans, A model-based assay design to reproduce in vivo patterns of acute drug-induced toxicity, *Archives of Toxicology* 92 (2018) 553-555. <https://doi.org/10.1007/s00204-017-2041-7>
- [10] N. Selevsek, F. Caiment, R. Nudischer, H. Gmuender, I. Agarkova, F.L. Atkinson, I. Bachmann, V. Baier, G. Barel, C. Bauer, S. Boerno, N. Bosc, O. Clayton, H. Cordes, S. Deeb, S. Gotta, P. Guye, A. Hersey, F.M.I. Hunter, L. Kunz, A. Lewalle, M. Lienhard, J. Merken, J. Minguet, B. Oliveira, C. Pluess, U. Sarkans, Y. Schrooders, J. Schuchhardt, I. Smit, C. Thiel, B. Timmermann, M. Verheijen, T. Wittenberger, W. Wolski, A. Zerck, S. Heymans, L. Kuepfer, A. Roth, R. Schlapbach, S. Niederer, R. Herwig, J. Kleinjans, Network integration and modelling of dynamic drug responses at multi-omics levels, *Communications biology* 3 (2020) 573-573. <https://doi.org/10.1038/s42003-020-01302-8>
- [11] M. Verheijen, Y. Schrooders, H. Gmuender, R. Nudischer, O. Clayton, J. Hynes, S. Niederer, H. Cordes, L. Kuepfer, J. Kleinjans, F. Caiment, Bringing in vitro analysis closer to in vivo: Studying doxorubicin toxicity and associated mechanisms in 3D human microtissues with PBPK-based dose modelling, *Toxicology Letters* 294 (2018) 184-192. <https://doi.org/10.1016/j.toxlet.2018.05.029>
- [12] T.A. Down, V.K. Rakyen, D.J. Turner, P. Flicek, H. Li, E. Kulesha, S. Gräf, N. Johnson, J. Herrero, E.M. Tomazou, N.P. Thorne, L. Bäckdahl, M. Herberth, K.L. Howe, D.K. Jackson,

- M.M. Miretti, J.C. Marioni, E. Birney, T.J.P. Hubbard, R. Durbin, S. Tavaré, S. Beck, A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis, *Nature biotechnology* 26 (2008) 779-785. <https://doi.org/10.1038/nbt1414>
- [13] L. Chavez, J. Jozefczuk, C. Grimm, J. Dietrich, B. Timmermann, H. Lehrach, R. Herwig, J. Adjaye, Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage, *Genome research* 20 (2010) 1441-1450. <https://doi.org/10.1101/gr.110114.110>
- [14] G.A. Wilson, P. Dhami, A. Feber, D. Cortázar, Y. Suzuki, R. Schulz, P. Schär, S. Beck, Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers, *GigaScience* 1 (2012). <https://doi.org/10.1186/2047-217X-1-3>
- [15] J. Huang, V. Renault, J. Sengenès, N. Touleimat, S. Michel, M. Lathrop, J. Tost, MeQA: a pipeline for MeDIP-seq data quality assessment and analysis, *Bioinformatics* 28 (2012) 587-8. <https://doi.org/10.1093/bioinformatics/btr699>
- [16] M. Lienhard, S. Grasse, J. Rolff, S. Frese, U. Schirmer, M. Becker, S. Börno, B. Timmermann, L. Chavez, H. Sültmann, G. Leschber, I. Fichtner, M.R. Schweiger, R. Herwig, QSEA-modelling of genome-wide DNA methylation from sequencing enrichment experiments, *Nucleic acids research* 45 (2017) e44-e44. <https://doi.org/10.1093/nar/gkw1193>
- [17] G.A. Wilson, S. Beck, Computational analysis and integration of MeDIP-seq methylome data, in: *Next Generation Sequencing - Advances, Applications and Challenges*, 2016.
- [18] Y. Shi, W. Gong, X. Gong, P. Wang, X. Zhao, Genome-wide DNA methylation analysis of breast cancer MCF-7 / Taxol cells with MeDIP-Seq, *PLoS one* 15 (2020) e0241515. <https://doi.org/10.1371/journal.pone.0241515>
- [19] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* 25 (2009) 1754-60. <https://doi.org/10.1093/bioinformatics/btp324>
- [20] P. Danecek, J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies, H. Li, Twelve years of SAMtools and BCFtools, *Gigascience* 10 (2021). <https://doi.org/10.1093/gigascience/giab008>
- [21] R Core Team, R: A Language and Environment for Statistical Computing, in: R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [22] R.G. Cavalcante, M.A. Sartor, annotatr: genomic regions in context, *Bioinformatics* (Oxford, England) 33 (2017) 2381-2383. <https://doi.org/10.1093/bioinformatics/btx183>
- [23] K. Dreval, V. Tryndyak, A. de Conti, F.A. Beland, I.P. Pogribny, Gene Expression and DNA Methylation Alterations During Non-alcoholic Steatohepatitis-Associated Liver Carcinogenesis, *Frontiers in genetics* 10 (2019). <https://doi.org/10.3389/fgene.2019.00486>
- [24] K. Zhang, C. Li, J. Liu, X. Tang, Z. Li, DNA methylation alterations as therapeutic prospects in thyroid cancer, *Journal of Endocrinological Investigation* 42 (2019) 363-370. <https://doi.org/10.1007/s40618-018-0922-0>
- [25] G. Minotti, P. Menna, E. Salvatorelli, G. Cairo, L. Gianni, Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity, *Pharmacological reviews* 56 (2004) 185-229. <https://doi.org/10.1124/pr.56.2.6>
- [26] S.S. Mahmood, R.B. Patel, J. Butler, M. Vaduganathan, Epirubicin and long-term heart failure risk in breast cancer survivors, *European journal of heart failure* 20 (2018) 1454-1456. <https://doi.org/10.1002/ehf.1215>

- [27] J.E. Bates, R.M. Howell, Q. Liu, Y. Yasui, D.A. Mulrooney, S. Dhakal, S.A. Smith, W.M. Leisenring, D.J. Indelicato, T.M. Gibson, G.T. Armstrong, K.C. Oeffinger, L.S. Constone, Therapy-Related Cardiac Risk in Childhood Cancer Survivors: An Analysis of the Childhood Cancer Survivor Study, *Journal of Clinical Oncology* 37 (2019) 1090-1101. <https://doi.org/10.1200/jco.18.01764>
- [28] A. Ghigo, M. Li, E. Hirsch, New signal transduction paradigms in anthracycline-induced cardiotoxicity, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1863 (2016) 1916-1925. <https://doi.org/10.1016/j.bbamcr.2016.01.021>
- [29] N. Nguyen, T. Souza, J. Kleinjans, D. Jennen, Transcriptome analysis of long noncoding RNAs reveals their potential roles in anthracycline-induced cardiotoxicity, *Non-coding RNA Research* (2022). <https://doi.org/10.1016/j.ncrna.2022.01.002>
- [30] N. Nguyen, T. Souza, M.C.T. Verheijen, H. Gmuender, N. Selevsek, R. Schlapbach, J. Kleinjans, D. Jennen, Translational Proteomics Analysis of Anthracycline-Induced Cardiotoxicity From Cardiac Microtissues to Human Heart Biopsies, *Front Genet* 12 (2021) 695625. <https://doi.org/10.3389/fgene.2021.695625>
- [31] J. Zhou, W.-P. Yong, C.S. Yap, A. Vijayaraghavan, R.A. Sinha, B.K. Singh, S. Xiu, S. Manesh, A. Ngo, A. Lim, C. Ang, C. Xie, F.Y. Wong, S.J. Lin, W.K. Wan, I.B. Tan, H. Flotow, P. Tan, K.-H. Lim, P.M. Yen, L.K. Goh, An integrative approach identified genes associated with drug response in gastric cancer, *Carcinogenesis* 36 (2015) 441-451. <https://doi.org/10.1093/carcin/bgv014>
- [32] D. Beck, I. Sadler-Riggelman, M.K. Skinner, Generational comparisons (F1 versus F3) of vinclozolin induced epigenetic transgenerational inheritance of sperm differential DNA methylation regions (epimutations) using MeDIP-Seq, *Environmental Epigenetics* 3 (2017). <https://doi.org/10.1093/eeep/dvx016>
- [33] M. Dworkin, S. Xie, M. Saha, J. Thimmapuram, V. Kalavacharla, Analyses of methylomes of upland and lowland switchgrass (*Panicum virgatum*) ecotypes using MeDIP-seq and BS-seq, *BMC Genomics* 18 (2017) 851. <https://doi.org/10.1186/s12864-017-4218-0>
- [34] H. Heberle, G.V. Meirelles, F.R. da Silva, G.P. Telles, R. Minghim, InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams, *BMC Bioinformatics* 16 (2015) 169. <https://doi.org/10.1186/s12859-015-0611-3>
- [35] H. Mi, A. Muruganujan, D. Ebert, X. Huang, P.D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools, *Nucleic Acids Research* 47 (2018) D419-D426. <https://doi.org/10.1093/nar/gky1038>
- [36] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (2014) 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- [37] S. Andrews, FastQC: a quality control tool for high throughput sequence data, in: Babraham Institute, Babraham, UK, 2010.
- [38] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics* 32 (2016) 3047-3048. <https://doi.org/10.1093/bioinformatics/btw354>
- [39] D.R. Zerbino, P. Achuthan, W. Akanni, M R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C.G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O.G.

Izuogu, S.H. Janacek, T. Juettemann, J.K. To, M.R. Laird, I. Lavidas, Z. Liu, J.E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D.N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C.K. Ong, A. Parker, M. Patricio, H.S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S.E. Hunt, M. Kostadima, N. Langridge, F.J. Martin, M. Muffato, E. Perry, M. Ruffier, D.M. Staines, S.J. Trevanion, B.L. Aken, F. Cunningham, A. Yates, P. Flicek, *Ensembl 2018*, *Nucleic Acids Research* 46 (2017) D754-D761. <https://doi.org/10.1093/nar/gkx1098>

[40] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinformatics* 12 (2011) 323. <https://doi.org/10.1186/1471-2105-12-323>

[41] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nature methods* 9 (2012) 357-359. <https://doi.org/10.1038/nmeth.1923>

[42] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biology* 15 (2014) 550. <https://doi.org/10.1186/s13059-014-0550-8>

[43] C.T. Hang, J. Yang, P. Han, H.-L. Cheng, C. Shang, E. Ashley, B. Zhou, C.-P. Chang, Chromatin regulation by Brg1 underlies heart muscle development and disease, *Nature* 466 (2010) 62-67. <https://doi.org/10.1038/nature09130>

[44] T. Sakaguchi, M. Takefuji, N. Wettschureck, T. Hamaguchi, M. Amano, K. Kato, T. Tsuda, S. Eguchi, S. Ishihama, Y. Mori, Y. Yura, T. Yoshida, K. Unno, T. Okumura, H. Ishii, Y. Shimizu, Y.K. Bando, K. Ohashi, N. Ouchi, A. Enomoto, S. Offermanns, K. Kaibuchi, T. Murohara, Protein kinase N promotes stress-induced cardiac dysfunction through phosphorylation of myocardin-related transcription factor A and disruption of its interaction with actin, *Circulation* 140 (2019) 1737-1752. <https://doi.org/10.1161/CIRCULATIONAHA.119.041010>

[45] J. Huang, L. Chen, Y. Yao, C. Tang, J. Ding, C. Fu, H. Li, G. Ma, Pivotal role of regulator of G-protein signaling 12 in pathological cardiac hypertrophy, *Hypertension* 67 (2016) 1228-1236. <https://doi.org/10.1161/HYPERTENSIONAHA.115.06877>

[46] M. Hohl, M. Wagner, J.-C. Reil, S.-A. Müller, M. Tauchnitz, A.M. Zimmer, L.H. Lehmann, G. Thiel, M. Böhm, J. Backs, C. Maack, HDAC4 controls histone methylation in response to elevated cardiac load, *The Journal of Clinical Investigation* 123 (2013) 1359-1370. <https://doi.org/10.1172/JCI61084>

[47] C.N. Koyani, C. Trummer, N. Shrestha, S. Scheruebel, B. Bourgeois, I. Plastira, S. Kickmaier, H. Sourij, P.P. Rainer, T. Madl, W. Sattler, B. Pelzmann, E. Malle, D. von Lewinski, Saxagliptin but not sitagliptin inhibits CaMKII and PKC via DPP9 inhibition in cardiomyocytes, *Frontiers in Physiology* 9 (2018). <https://doi.org/10.3389/fphys.2018.01622>

[48] F. Cividini, B.T. Scott, J. Suarez, D.E. Casteel, S. Heinz, A. Dai, T. Diemer, J.A. Suarez, C.W. Benner, M. Ghassemian, W.H. Dillmann, Ncor2/PPAR α -dependent upregulation of MCub in the type 2 diabetic heart impacts cardiac metabolic flexibility and function, *Diabetes* 70 (2021) 665-679. <https://doi.org/10.2337/db20-0779>

[49] G. Benincasa, B.A. Maron, O. Affinito, M. D'Alto, M. Franzese, P. Argiento, C. Schiano, E. Romeo, P. Bontempo, P. Golino, L. Berrino, J. Loscalzo, C. Napoli, Circulating CD4+T/methylation signatures of network-oriented SOCS3, ITGAL, NFIC, NCOR2, PGK1 genes associate with hemodynamics in pulmonary arterial hypertension patients, *European Heart Journal* 42 (2021). <https://doi.org/10.1093/eurheartj/ehab724.1952>

[50] T. Brattelid, L.H. Winer, F.O. Levy, K. Liestøl, O.M. Sejersted, K.B. Andersson, Reference gene alternatives to Gapdh in rodent and human heart failure gene expression studies, *BMC Molecular Biology* 11 (2010) 22. <https://doi.org/10.1186/1471-2199-11-22>

[51] J.L. Horton, O.J. Martin, L. Lai, N.M. Riley, A.L. Richards, R.B. Vega, T.C. Leone, D.J. Pagliarini, D.M. Muoio, K.C. Bedi, Jr., K.B. Margulies, J.J. Coon, D.P. Kelly, Mitochondrial protein hyperacetylation in the failing heart, *JCI insight* 2 (2016) e84897. <https://doi.org/10.1172/jci.insight.84897>

[52] N.H. Banke, A.R. Wende, T.C. Leone, J.M. O'Donnell, E.D. Abel, D.P. Kelly, E.D. Lewandowski, Preferential oxidation of triacylglyceride-derived fatty acids in heart is augmented by the nuclear receptor PPAR α , *Circulation Research* 107 (2010) 233-241. <https://doi.org/10.1161/CIRCRESAHA.110.221713>

Chapter 3 | Transcriptomics and proteomics analysis in Anthracycline-induced cardiotoxicity investigation

Adapted from

Nhan Nguyen, Terezinha Souza, Jos Kleinjans, and Danyel Jennen. "*Transcriptome analysis of long noncoding RNAs reveals their potential roles in anthracycline-induced cardiotoxicity.*" *Non-coding RNA Research* 7 (2022): 106-113.

Nhan Nguyen, Terezinha Souza, Marcha CT Verheijen, Hans Gmuender, Nathalie Selevsek, Ralph Schlapbach, Jos Kleinjans, and Danyel Jennen. "*Translational proteomics analysis of anthracycline-induced cardiotoxicity from cardiac microtissues to human heart biopsies.*" *Frontiers in Genetics* 12 (2021): 1018.

Introduction

Anthracyclines (ANTs) are a group of well-known chemotherapeutic agents consisting of thousands of analogs; the commonly used analogs are doxorubicin (DOX), epirubicin (EPI), and idarubicin (IDA). DOX is an essential drug in the treatment of multiple cancer types such as acute lymphoblastic leukemia, nephroblastoma, diffuse large B-cell lymphoma, and Hodgkin lymphoma, in both children and adults [1,2]. EPI, as a derivative of DOX, has a similar spectrum of activity compared to DOX. IDA, a derivative of daunorubicin, has shown more potency in antitumor activity, especially, its efficacy in multidrug resistance compared to other ANTs [3,4].

Despite being widely used in cancer treatments, ANTs have been defined as cardiotoxic agents. Multiple cohort studies have shown that ANTs exposure dose-dependently increases the risk of cardiac disorder in cancer survivors [5,6]. To reduce the cardiac disorder risks, the maximum cumulative doses of DOX and EPI used are recommended to be 450 to 600 mg/m² or 900 mg/m², respectively [3,7]. While the maximum cumulative dose of IDA for cardiac safety has not been defined, patients treated with IDA cumulative doses in a range of 150 to 4080 mg/m² showed a low probability of cardiotoxicity [8]. In a retrospective study on acute myeloid leukemia and myelodysplasia patients, the probability of IDA-related cardiomyopathy was 5% at a cumulative IDA dose of 150 to 290 mg/m² [9]. Even though ANTs can lead to undesirable effects, therapeutic alternatives have not been offered in oncology. Hence, understanding ANT adverse mechanisms may assist clinical treatments and limit drug side effects.

Although a range of studies has investigated how ANTs induce heart toxicity, the distinct molecular mechanism leading to their adverse cardiac effects remains unclear. The conventional paradigm was that reactive oxygen species generated by ANTs cause damage to multiple cellular components. However, combinations of ANTs with multiple antioxidants have failed to promote cardioprotection [10], so ANTs' toxic mechanism has been considered complex and multifactorial [11]. Emerging perspectives have suggested that ANTs can modulate growth factor receptors (ErbB2 and ErbB4), β_2 adrenergic receptor (β_2 AR), and Toll-like receptors (TLR2 and TLR4). Therefore, ANTs may alter particular signaling pathways including PI3K, NF- κ B, and GATA4 pathways, and then impact cardiac immune functions, cardiac contractility, and cardiomyocyte survival [12]. Even though recent studies have updated explanations for ANT-induced cardiotoxicity, further investigation is still crucial to clarify the ANT cardiotoxic mechanisms.

According to known ANT-induced cardiotoxicity paradigms, studies with targeted approaches could help to examine and strengthen the prior toxicity regime. However, targeted approaches could not entirely explore all potential mechanisms. Recent modern technologies such as RNA-seq or proteomics methods can effectively profile the cellular molecules in a single observation. These advanced technologies are feasible and efficient screening methods to perceive multiple drug-induced alterations inside cells and describe the comprehensive adverse effects without prior knowledge. Consequently, diverse data analyses are needed to interpret high-dimensional data delivered from these high throughput technologies.

In this chapter 3, we explored different research angles and data analysis approaches to analyze transcriptomics and proteomics data from *in vitro* cardiac microtissues exposed to ANTs. We also compared the gene and protein expression in human *in vitro* samples to their expression in the cardiac biopsy samples from heart failure patients.

Experimental design and dataset

Human cardiac microtissue samples

The human cardiac microtissues (3D InSight™ Human Cardiac Microtissues from InSphero) were incubated with either a clinically therapeutic dose or a toxic dose of ANT analogs (DOX, EPI, IDA) using PBPK-based dose profiles for 2 weeks [13] (Table S 1-1). The therapeutic dose was based on the common clinical treatment dose, while the toxic dose was the IC₂₀ value based on the ATP production (cell viability) previously determined after one week of exposure [14]. ANTs were then dissolved in DMSO 0.1% as stock solutions before diluting to the particular drug concentrations calculated by the PBPK model [13,15]. Therefore, the DMSO concentration in the medium fluctuated, and the control samples were exposed to this fluctuating DMSO profile (Table S 1-1). During these 2 weeks of exposure, the microtissues were harvested in triplicate after 2, 8, 24, 72, 168, and 240 hours of ANTs exposure. The microtissues exposed to ANT therapeutic doses were also collected in triplicated after 336 hours of exposure.

Human cardiac biopsy samples

The cardiac biopsies were collected from heart failure patients. The investigation conforms to the principles outlined in the Declaration of Helsinki. The patient biopsies collection was approved by the Medical Ethics Committee of Maastricht University Medical Center. Informed consent has been obtained from all the subjects [14]. The participants consisted of heart failure patients who have no cancer history, cancer survivors who underwent chemotherapy with ANTs, and cancer survivors who underwent chemotherapy without ANTs (Table S 3-1).

Chapter 3.1: Transcriptome analysis of long noncoding RNAs reveals their potential roles in anthracycline-induced cardiotoxicity

Objectives of the study

Investigating transcriptome plays a key role in elucidating the underlying cellular mechanisms of action. Alongside the traditional protein-coding genes, long non-coding RNAs (lncRNAs), which are non-coding RNAs longer than 200 nucleotides, have recently gained widespread attention as new players in numerous cellular functions such as transcription regulation, mRNA stability, translation regulation, and post-translational modifications [16,17]. Several studies have indicated that lncRNAs can provide a detailed view of cardiac development and pathology. For instance, researchers have emphasized the important function of lncRNAs such as *H19*, *MALAT1*, and *MDNCR* at different stages of human cardiac development and heart disease [18], which inform the lncRNAs' abilities to become potential biomarkers as well as therapeutic targets in heart diseases [19]. Other studies have also conceded the lncRNA expression patterns related to medical treatment. An investigation in non-alcoholic fatty liver disease revealed the pharmaceutical mechanisms of berberine in both mRNAs and lncRNAs [20]. Another study in Ang II-treated cardiac fibroblasts also showed the significantly altered expression of lncRNAs, such as *NR024118* [21]. Therefore, lncRNAs can involve not only in disease pathology but also in the cell responses to drug treatment.

Several studies have investigated the role of lncRNA in ANTs' mechanisms. There are studies on genome-wide lncRNA profiles in DOX-resistant breast cancer cells [22], as well as on lncRNAs related to the pathological response in the DOX neoadjuvant chemotherapy in breast cancer [23]. A specific lncRNA, *H19*, has even been highlighted as a major mediator in breast cancer chemoresistance after DOX treatment [24]. Although researchers have explored the lncRNA functions in the ANT mechanisms of action, few initial studies have focused on the lncRNA functions in ANT-induced cardiotoxicity. A study in mice indicated that the up-regulation of lncRNA *FOXC2-AS1* can protect cardiomyocytes from DOX toxicity [25]. Another study showed that the down-regulation of the lncRNA cardiac hypertrophy-related factor (*CHRF*) can diminish cardiac dysfunction and injury [26]. These studies have advised on the critical role of lncRNAs in ANT-induced cardiotoxicity, even though they mostly focus on some targeted lncRNAs and are

limited to DOX treatment. Thus, the majority of lncRNAs related to ANT side effects still await further investigation.

In this chapter 3-1, we explored the transcriptomic-wide lncRNA profiles of human cardiac *in vitro* microtissues and biopsies under different ANT treatments. This outcome of this study could provide a broad view of lncRNA candidates that are influenced by ANT treatments. This study also offers an alternate angle in transcriptome analysis that focused on the expression of non-coding genes. This could expand and enlarged the advance of the transcriptomic-wide measurement.

Methods & Analysis procedures

RNA sequencing

Total RNA in each sample was isolated using Qiagen AllPrep DNA/RNA/miRNA Universal Kit (Cat #80224). Ribosomal RNAs were depleted by using the Illumina RiboZero Gold kit (Cat #MRZG12324), and then samples were prepared by the Lexogen SENSE total RNA library preparation kit (Cat #009.96). The RNA quality and quantity of the samples were checked by the Agilent 420 TapeStation and the Qubit™ before they were sequenced by an HiSeq2000 with 100bp paired-end reads [15].

The adapter sequences of the paired-end sequenced raw data were removed by using Trimmomatic version 0.36 [27]. The sequencing quality of samples was examined by FastQC version 0.11.7 [28], and summarized by MultiQC [29] before and after trimming the reads. The reads were mapped onto the Ensembl human genome reference, version GRCh38.p12, Ensembl Archive Release 93 [30] using RSEM version 1.3.1 [31], and Bowtie2 version 2.3.4.1 [32] with the paired-end option. After removing 2 samples (DOX_Tox_240_2 and IDA_The_240_3) with a low read count (<5 million), 145 *in vitro* samples were used for transcriptome-wide analysis. Further analysis was performed in R version 4.0.2 (released on 2020-06-22) [33] with some R visualization packages including the Upset [34], Tidyverse [35], and ggplot [36]. The human genome database from the Ensembl Biomart website (<https://m.ensembl.org/biomart>) [30] was used to annotate the gene type, gene name, and gene function.

Differentially expressed genes in the in vitro RNA-seq data

The raw read counts of all remaining *in vitro* samples were normalized using the DESeq2 R package [37]. The ImpulseDE2 package [38] performed a time-series differential expressed gene analysis between the ANT-treated and the control samples for 2 weeks. ImpulseDE2 has its internal DESeq2 normalization, thus, the

raw gene read counts from RSEM were used as input data. The function “runImpulseDE2” was applied to perform case-control analysis using gene read count from ANT-treated samples and control samples ($p\text{-adj} < 0.01$). We detected genes that were differentially expressed across all the ANT-treated conditions compared to control samples. The pathway analysis for these differentially expressed genes was performed by using ConsensusPathDB [39] with all detected Ensembl gene IDs as the background gene list. A list of lncRNAs, which are related to heart diseases, was extracted from LncRNADisease_v2.0 [40] using the queries “heart” and “cardi” (searching for “cardio”, and “cardiac”). The functions of the lncRNAs related to heart disease were also extracted from the LncTarD database, a manually-curated database of experimentally-supported functional lncRNA-target regulations in human diseases [41], with Ensembl IDs of differential expression lncRNAs as input.

Differentially expressed genes in the biopsies RNA-seq data

The biopsies ($n=31$) were divided into 3 batches to run the RNA sequencing (Table S 3-1) [14]. The patient characteristics recruited in batch 3 differed from those in batch 1 and 2; thus the biopsies in batch 3 were excluded (Figure S 3-1). The biopsy samples of batches 1 and 2 ($n=19$) were used for further analysis, including heart failure control patients ($n=8$), cancer survivors who were treated with ANTs ($n=9$), and cancer survivors who were treated without ANTs ($n=2$) (Table S 3-1) [14]. The read counts of the biopsy samples were normalized and analyzed by using the DESeq2 package [37] to identify differentially expressed genes between ANT-treated patients ($n=9$) and control subjects ($n=8$) ($p\text{-adj} < 0.01$).

Based on the metadata, there were 7 matched pairs between heart failure control patients and cancer survivors, who were treated with ANTs and developed heart failure as a treatment side-effect [14]. The differences in gene expression of these pairs’ subjects were used to calculate the log₂ fold change values in biopsy samples.

Results

A general view

A cluster tree of the *in vitro* transcriptome profiles demonstrated a clear separation between control and ANT-treated samples, in which all controls were grouped in one branch of the cluster tree (Figure 3-1A). Among ANT-treated conditions, most of the DOX and EPI-treated samples were specifically grouped in one sub-branch, whereas the IDA-treated samples were mainly grouped in

another sub-branch. By contrast, the transcriptome profiles of the biopsies data (n=19) did not demonstrate a clear distinction between the heart failure control patients and the heart failure patients who underwent ANT treatments (Figure 3-1B).

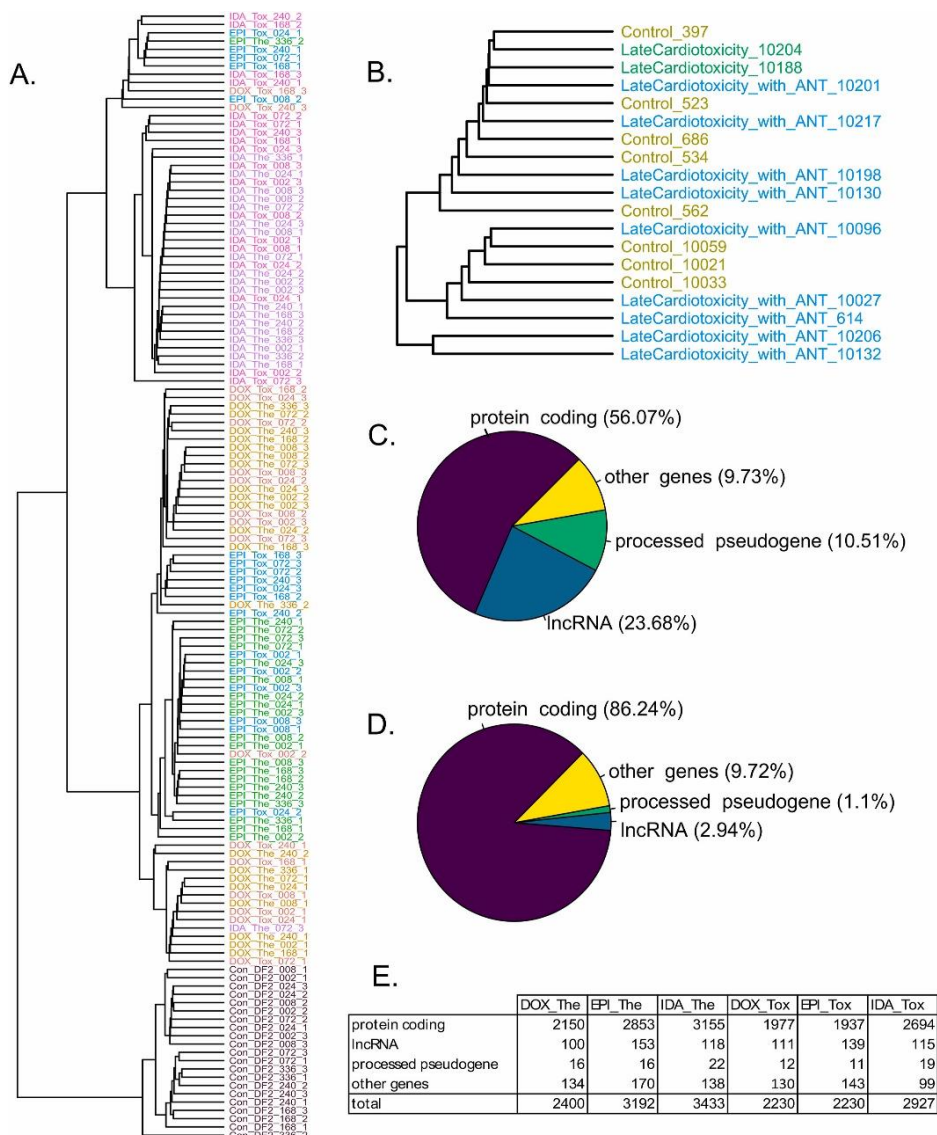


Figure 3-1: A general view of transcriptome profiles of the *in vitro* and biopsy samples. **(A)** The cluster tree of the *in vitro* samples' transcriptome profiles. **(B)** The cluster tree of the biopsy samples' transcriptome profiles. The biopsy samples included heart failure patients without a cancer history (Control) and heart failure patients who underwent cancer treatment with ANT (LateCardiotoxicity_with_ANT) and without ANT

(LateCardiotoxicity). The ending numbers are patient IDs (**C**) The gene type of expressed genes (the average read count across samples > 0 , annotated using the Ensembl database) in the *in vitro* samples. (**D**) Gene types of overlapped differentially expressed genes (545 genes) across all *in vitro* ANT-treated conditions compared to control samples. (**E**) The number of differentially expressed genes in each *in vitro* ANT-treated condition compared to control samples. $P\text{-adj} < 0.01$. Con_DF2: control samples; DOX: doxorubicin, EPI: epirubicin, IDA: idarubicin; The: therapeutic dose, Tox: toxic dose; 002, 008, 024, 072, 168, 240, 336 are corresponding exposure periods; ANT: anthracycline(s).

While the human genome contains 58,395 detectable genes, 31,910 genes were expressed in the *in vitro* cardiac tissues under ANT exposure conditions with the average read count across samples is larger than 0. According to the Ensembl database, these active genes belong to different gene types (Figure 3-1C). Protein-coding genes, as key components of the cellular mechanisms, contributed to a large proportion of genes that were expressed (56.07%). Notably, lncRNA genes, which lack the protein-coding ability, accounted for 23.68 % of genes that were expressed in both ANT-treated and control samples. The rRNA genes, which is a predominant RNA group, were rarely detected as ribo-depleted total RNA library preparation was used in this study.

Differentially expressed (DE) genes

By using the ImpulseDE2 tool on the *in vitro* data, DE genes (adjusted $p\text{-value} < 0.01$) were detected in each ANT-treated condition compared to the control condition during the 2 weeks of exposure. In total, 545 genes were consistently differentially expressed in all *in vitro* ANT-treated conditions compared to control. Most of these overlapping DE genes were protein-coding genes, while 16 of them (2.94%) were lncRNA genes (Figure 3-1D). In particular, each ANT-treated condition had over 100 lncRNA genes which were differentially expressed compared to the control samples (Figure 3-1E).

Pathway analysis revealed which cardiac functions the 545 overlapped DE genes are involved in. Three of the top 10 over representative pathways are heart disease pathways; it emphasizes that the effect of ANT treatment through these genes might facilitate cardiotoxicity and heart failure development (Figure 3-2A, Table S 3-2). A previous study using the same dataset with another differential gene expression analysis tool and other pathway databases also provided a similar outcome [14]. Although pathway analysis can capture a part of the ANT toxicity mechanism, its expository ability is mainly restricted to protein-coding genes (Figure 3-2B). Particularly, the 3 heart disease pathways of the top 10 over representative pathways are from the KEGG database and contain only protein-coding genes (Table S 3-2) [42]. This conventional approach neglects non protein-

coding genes, especially lncRNAs which were consistently differentially expressed between ANT-treated conditions and control (Figure 3-1D).

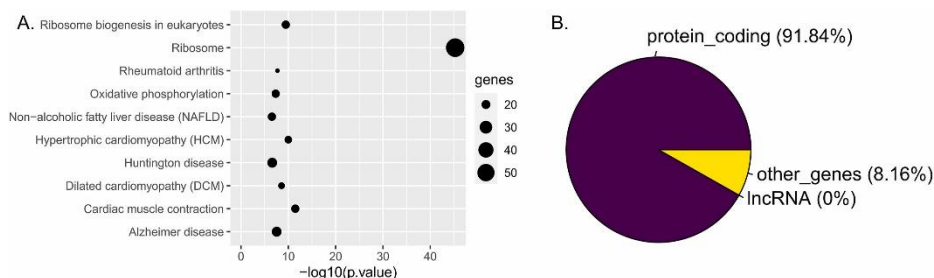


Figure 3-2: Pathway analysis outcomes from overlapped differentially expressed (DE) genes in the *in vitro* anthracycline-treated samples. **(A)** Top 10 over representative pathways of the 545 overlapped DE genes across *in vitro* ANT-treated conditions; **(B)** Gene type proportions of the overlapped DE genes that popped up in the pathway analysis.

In the biopsies data, 37 DE genes, including 5 lncRNA genes, were detected by DESeq2 between ANT-treated patients and control subjects. The pathway analysis did not reveal any connection between these 37 DE genes to heart function and heart disease (Table S 3-3). Of the 5 DE lncRNA genes in biopsies data, the *LINC00612* gene might be involved in acute myocardial infarction [43], while the *AL031280.1* gene was also differentially expressed in IDA-treated samples compared to control.

Differentially expressed lncRNAs

Sixteen lncRNA genes were consistently differentially expressed in all ANT-treated conditions while other lncRNAs were only differentially expressed under certain ANT-treated conditions (Table 3-1). For example, 6 lncRNAs were differentially expressed only in the DOX-treated conditions, while 13 lncRNAs were specifically differentially expressed in the EPI-treated conditions. The IDA-treated conditions had the highest number of drug-specific lncRNAs with 17 DE lncRNAs (Table 3-1). For the dose-specific effects, the *AL451123.1* lncRNA was differentially expressed in all ANT therapeutic-treated samples. The other 2 lncRNAs, *AL133453.1* and *BACE1-AS*, were differentially expressed in all ANT toxic-treated samples.

Table 3-1: Differentially expressed (DE) lncRNAs in different anthracycline (ANT) treatment conditions.

	DOX-treated condition	EPI-treated condition	IDA-treated condition
Numbers of DE lncRNA	6	13	17
lncRNAs	<i>AC025259.1</i> , <i>AL031985.3</i> , <i>FP671120.5</i> , <i>AD000090.1</i> , <i>HCP5</i> , <i>FTX</i>	<i>AC055811.1</i> , <i>N4BP2L2-IT2</i> , <i>AC078880.4</i> , <i>AC009264.1</i> , <i>AL132656.2</i> , <i>AC007262.2</i> , <i>LINC02503</i> , <i>AL354733.3</i> , <i>XIST</i> , <i>AC092828.1</i> , <i>AP000766.1</i> , <i>AC125257.1</i> , <i>AC007114.2</i>	<i>AC009779.2</i> , <i>AL162311.3</i> , <i>AC009133.2</i> , <i>ESRG</i> , <i>AC100803.3</i> , <i>AC104794.2</i> , <i>OVCH1-AS1</i> , <i>C4B-AS1</i> , <i>SNHG22</i> , <i>LINC02108</i> , <i>URB1-AS1</i> , <i>SNHG16</i> , <i>C4A-AS1</i> , <i>KCNQ1OT1</i> , <i>AC093495.1</i> , <i>AC018761.2</i> , <i>GABPB1-AS1</i>
Numbers of DE lncRNA	16		
lncRNAs	<i>AC006064.4</i> , <i>AC007009.1</i> , <i>RMRP</i> , <i>LINC00622</i> , <i>SNHG7</i> , <i>AC093866.1</i> , <i>SNHG29</i> , <i>H19</i> , <i>AC132217.1</i> , <i>AC124312.3</i> , <i>LINC01638</i> , <i>AC020909.3</i> , <i>AC106791.1</i> , <i>BDNF-AS</i> , <i>AC010680.5</i> , <i>PCAT19</i>		

Notes: DOX, doxorubicin; EPI, epirubicin; IDA, idarubicin. Each treatment condition consists of samples treated with corresponding drugs in the therapeutic dose and toxic dose.

The LncTarD and LncRNADisease_v2.0 databases were used to explore the association between the DE lncRNAs and heart disease. The LncTarD database collects functional lncRNA–target regulations in humans [41], while the LncRNADisease_v2.0 is a lncRNAs related diseases database [40]. Through these databases, the association of DE lncRNAs with heart disease was detected, especially the *H19* and *FTX* genes were mentioned in both databases (Table S 3-4, Table S 3-5). Furthermore, other studies also highlighted the potential causal relationships between the DE lncRNAs and heart diseases [44-48]. Based on this prior knowledge, a network was established to represent the relationship between DE lncRNAs and corresponding heart diseases (Figure 3-3). This network highlights some key lncRNAs, such as *H19* and *BDNF-AS*, which strongly connect to cardiac diseases.

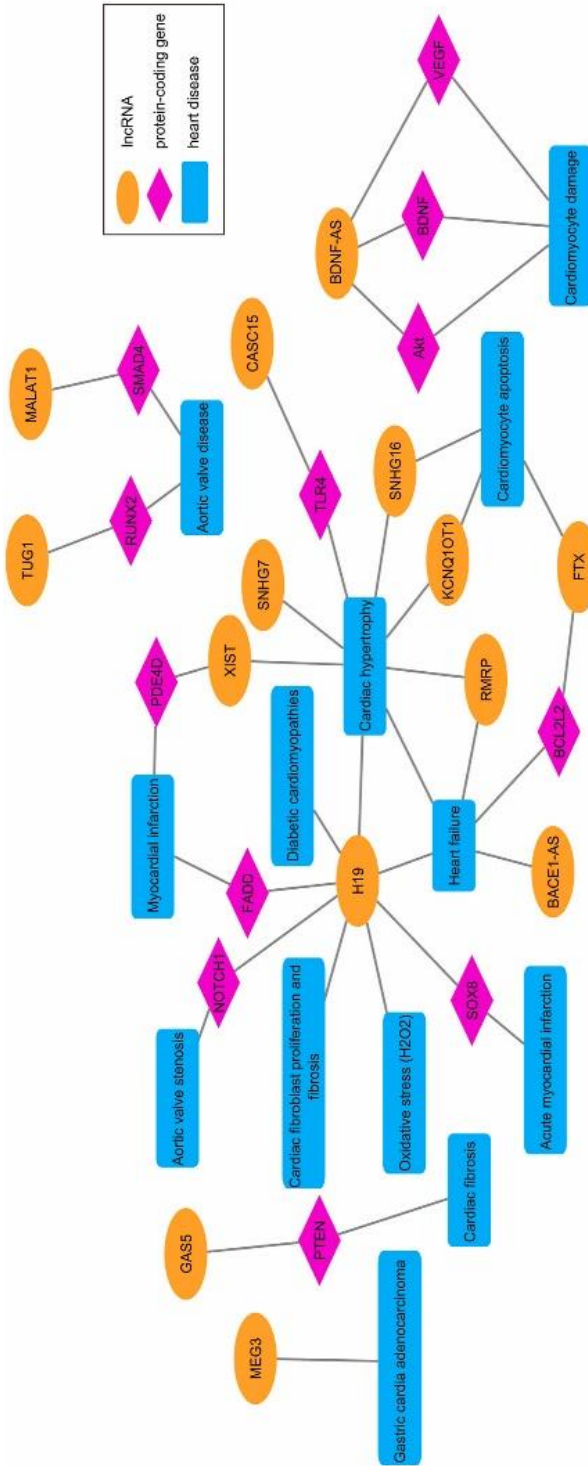


Figure 3-3: Network representing relationships between differentially expressed (DE) lncRNAs and corresponding heart diseases

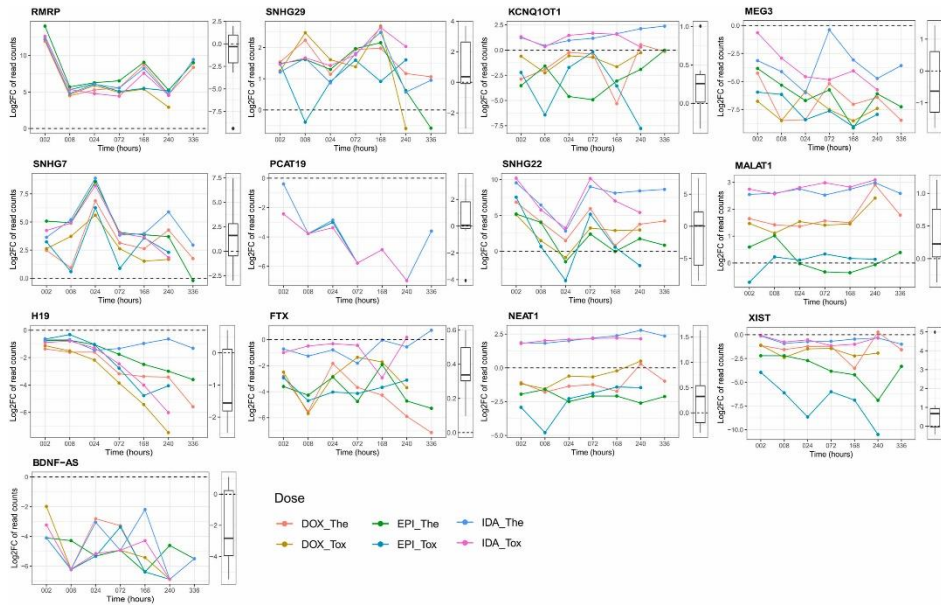


Figure 3-4: Log₂ fold change (log₂FC) gene expression in selected differentially expressed (DE) lncRNAs. For each lncRNA, log₂FC values over time between ANT-treated and control samples in the *in vitro* data are presented in the line chart, while the distribution of the log₂FC values between ANT-treated patients and control patients in matched pairs are represented in the boxplot. Con_DF2: control samples; DOX: doxorubicin, EPI: epirubicin, IDA: idarubicin; The: therapeutic dose, Tox: toxic dose; 002, 008, 024, 072, 168, 240, 336 are corresponding exposure periods in hours; ANT: anthracycline.

Of the 16 DE lncRNAs found in all ANT-treated conditions, some lncRNAs showed a remarkable alteration in their expressions (Figure 3-4). *SNHG29*, *SNHG7*, and *RMRP* were up-regulated in the *in vitro* ANT-treated conditions. Although these lncRNAs' expressions were not statistically different between patient groups in the biopsies data, the expression of *SNHG29* and *SNHG7* was also up-regulated in the ANT-treated patients compared to the control subjects (Figure 3-4). Other lncRNAs that are *H19*, *BDNF-AS*, and *PCAT19* were down-regulated in the *in vitro* ANT-treated conditions, especially, the expression of *H19* showed a clear distinction between ANT-treated conditions after a longer exposure time. Similarly, the expressions of *H19* and *BDNF-AS* genes were down-regulated in the ANT-treated patients (Figure 3-4).

Other lncRNAs were differentially expressed in specific *in vitro* ANT-treated conditions (Table 3-1). For instance, the lncRNA *XIST* was prominently down-regulated in the EPI-treated conditions, whereas the *FTX* gene was down-regulated in the DOX-treated conditions. The 2 lncRNAs, *SNHG22* and *KCNQ1OT1*, were only significantly up-regulated in the IDA-treated conditions

(Figure 3-4). Additionally, *MEG3* was significantly down-regulated in the DOX and EPI-treated conditions, while *MALAT1* was significantly up-regulated in the DOX and IDA-treated conditions. *NEAT1* is an interesting lncRNA, which increased expression in the IDA-treated condition and decreased expression in the EPI-treated condition, while its expression in the DOX-treated condition was not differentially expressed compared to the control samples (Figure 3-4). The expressions of these lncRNAs also differed between ANT-treated patients and control subjects in the biopsies data (Figure 3-4).

Discussion

This chapter 3-1 revealed more insights into the alterations of lncRNAs not only during DOX treatment but also during EPI and IDA treatment. We also depict the expression of the lncRNA candidates in *in vitro* samples to their expression in the human biopsies. By using the transcriptomic-wide approach, this study provided a broad view of lncRNA candidates that were influenced by ANT treatments.

The *in vitro* transcriptome profiles of the ANT-treated samples differed from those of control samples, especially the transcriptome profiles of ANT-treated samples were mainly grouped based on their drug treatment (Figure 3-1A). The gene expression profiles of the DOX-treated samples were similar to those of the EPI-treated samples, rather than those of the IDA-treated samples. This phenomenon is supported by the prior knowledge that EPI, as a derivative from DOX, can share a similar mechanism of action with DOX [3]. By contrast, IDA is an analog derived from daunorubicin and is more lipophilic than DOX [49]. This underlying chemical difference could possibly explain why IDA treatments cause different gene expression profiles in cardiac tissue compared to DOX and EPI treatments. Hence, the *in vitro* transcriptome analysis could indicate the subtle distinction in mechanism between specific ANTs. However, some of the therapeutic-treated samples were grouped with the toxic-treated samples, specifically for IDA and DOX treatment (Figure 3-1A).

In the *in vitro* experiment, although a part of DE genes were protein-coding genes, another part of the DE genes were lncRNA genes (Figure 3-1D-E). While the conventional pathway analysis can demonstrate how DE protein-coding genes relate to heart disease (Figure 3-2A), it neglects the role of DE lncRNAs in the ANT-induced cardiotoxicity, in which no lncRNA gene was recognized on the pathway databases (Figure 3-2B). Specifically, 16 lncRNA were consistently differentially expressed across all ANT-treated conditions compared to controls (Table 3-1); it suggests that the expression of these lncRNAs could be affected by

typical ANTs mechanisms. Therefore, clariFYing DE lncRNAs' functions could provide a better understanding of the ANT toxicity mechanism, especially when lncRNA genes have recently emerged as potential targets in ANT-related studies [22-25].

By using lncRNA databases and doing literature research, we inspected the functions of not only the 16 overlapping DE lncRNAs but also other lncRNAs, which were differentially expressed in particular ANT-treated conditions (Figure 3-3). Both lncRNA databases manifested the role of *H19*, one of the 16 overlapping DE lncRNAs, in several heart diseases (Table S 3-4, Table S 3-5). *H19* was down-regulated in all *in vitro* ANT-treated conditions as well as in the ANT-treated patients in the biopsy samples (Figure 3-4), possibly due to oxidative stress. This phenomenon corresponds with another observation that showed a reduction in the *H19* level as a response to oxidative stress (H_2O_2) in C-kit+ cardiac progenitor cells [44]. This study also indicated some conserved binding sites of *miR-675*, a miRNA generated from *H19*, on 3'UTR of *USP10*; and the *H19/miR-675/USP10* axis can suppress both p53 and p21 expression [44]. Furthermore, *H19* can suppress the activity of *miR-22-3p* by binding to a *miR-22-3p* binding site in the *KDM3A* gene. An *H19* upregulation can then improve cardiac performance, alleviate cardiac fibrosis, and decrease inflammation [45]. These studies have encouraged that *H19* might be a potential biomarker and therapeutic target for ANT-induced cardiotoxicity. However, the overexpression of *H19* can contribute to the chemoresistance of breast cancer cells [24]. This suggests that *H19* can be involved in the on-target toxicity of ANTs, in which down-regulated *H19* produces desired treatment response in tumors, but lead to cardiotoxicity in cardiac cells.

Other overlapping DE lncRNAs are also involved in heart diseases and cellular mechanisms. Both *H19* and *RMRP* engage in cardiac hypertrophy and heart failure progression [46], while a *BDNF-AS* downregulation can activate *BDNF*, *VEGF*, and *Akt*, and thus rescue hypoxia/reoxygenation-induced damage in cardiomyocyte [47] (Figure 3-3). Hence, the *BDNF-AS* was downregulated both in the *in vitro* experiment and in the ANT-treated patients (Figure 3-4), which suggests the recovery intention of cardiac tissue under ANT treatment. Another lncRNA, namely *SNHG7*, has also emerged as a novel regulator for cardiac hypertrophy (Figure 3-3). *SNHG7* was up-regulated in the *in vitro* ANT-treated samples as well as ANT-treated patients (Figure 3-4), while a study in neonatal rat cardiomyocytes revealed that the up-regulation of the *SNHG7* genes can stabilize *SDAD1* mRNA, and then facilitate cardiac hypertrophy [48]. Other lncRNAs, including *SNHG29* and *PCAT19*, have not been investigated in heart failure contexts, but they are

known as potential regulators in cell signaling pathways. *SNHG29* acts as a competing endogenous RNA, in which it sponges the *miR-223-3p* to regulate the *CTNND1* expression. Therefore, *SNHG29* can modulate the Wnt/ β -catenin signaling pathway via the *miR-223-3p/CTNND1* axis [50]. *PCAT19* also acts as a competing endogenous RNA, in which it sponges the *miR-182* to regulate the *PDK4*, and consequently modulates the glycolysis and mitochondrial respiration in laryngeal cancer cell lines [51]. Additionally, the *PCAT19* expression was negatively correlated with the p53 expression in non-small cell lung cancer patients, and the silencing of *PCAT19* elevated the p53 expression level in H1993 cells [52]. These aforementioned studies have demonstrated the abilities of the overlapping DE lncRNAs related to cellular functions and heart disease, as well as their potential for ANT-induced cardiotoxicity research.

Some lncRNAs were only differentially expressed in samples treated with particular ANT conditions. The *AL451123.1* was differentially expressed in all ANT therapeutic-treated samples, whereas the *AL133453.1* and *BACE1-AS* were differentially expressed in all ANT toxic-treated samples. The functions of *AL451123.1* and *AL133453.1* are still unclear, whereas *BACE1-AS* is known as a heart failure-related lncRNA [40] (Figure 3-3). The *BACE1-AS* up-regulation might increase the *BACE1* level and accumulate β -amyloid, which is *BACE1*'s product. This dysregulation of the *BACE1-AS/BACE1*/ β -amyloid axis could diminish the cardiomyocyte viability [53]. Similarly, other lncRNAs were differentially expressed in samples treated with particular ANT analogs. Of the 6 lncRNAs which were differentially expressed in DOX-treated samples, the *FTX* could inhibit apoptosis and reduce hypertrophy in cardiomyocytes [54,55], while *AD000090.1* was proposed to regulate hypoxic responses [56] (Figure 3-3). lncRNA *XIST* was only differentially expressed in EPI-treated samples and could promote the progression of cardiac hypertrophy resulting in heart failure disease [57] (Figure 3-3). *KCNQ1OT1* and *SNHG16* were notable lncRNAs, which were differentially expressed in IDA-treated samples. Several studies have advocated that these 2 lncRNAs may facilitate cardiomyocyte apoptosis and accelerate cardiac hypertrophy [58,59]. The lncRNA databases also recommend other heart disease-related lncRNAs, which were differentially expressed in particular ANT-treated conditions, including *MALAT1*, *MEG3*, *TUG1*, *GAS5*, *CASC15* (Figure 3-3, Table S 3-4, Table S 3-5). When heart disease-related lncRNAs were atypically affected by certain ANT-treated conditions, they could reveal the subtle difference in toxic mechanisms of specific ANT analogs.

Although the DE lncRNAs in specific *in vitro* ANT-treated conditions could release the mechanism of individual ANT analogs, it is an obstacle to confirm this knowledge in clinical application. In this chapter 3-1, the transcriptome profiles of the biopsies data (n=19) did not demonstrate a clear distinction between the heart failure control patients and the heart failure patient who underwent cancer treatments (Figure 3-1B). There were 37 DE genes between the ANT-treated patients and control groups; however, there was no clear relation between these DE genes and cardiac function. All participants were heart failure patients, so it could be that non-significant differences related to heart disease were found among them. Furthermore, cancer patients often underwent chemotherapy with multi-drug combinations to improve the treatment efficiency [60,61]. Most patients involved in this study had been treated with multiple ANT analogs and other anti-tumor drugs a long time ago (Table S 3-1). Although it is difficult to distinguish the effect of ANT analogs leading to cardiovascular disease in a clinical setting, we observed some similar changes in lncRNA expressions between the *in vivo* experiments and human biopsies (Figure 3-3).

In conclusion, this chapter 3-1 provided new insight into the transcriptome alterations related to ANT-induced cardiotoxicity, especially the differential expression of lncRNAs. While the conventional pathway analysis might not be able to capture the role of these DE lncRNAs in cellular mechanisms, recent research has acknowledged the involvement of these lncRNAs in heart disease progression (Figure 3-3). *H19* seems to be involved in both chemoresistances as well as cardiotoxicity, which suggests its participation in the on-target toxicity of ANTs. Some lncRNAs, including *H19*, *RMRP*, *BDNF-AS*, and *SNHG7*, could be targets for further research on the typical mechanisms of ANT-induced cardiotoxicity, while other lncRNAs could advocate the cardiac responses to certain ANT analogs and doses. Although the functions of some lncRNAs have been explored, further study is needed to investigate the functionalities of other unknown lncRNAs (Table 3-1) related to heart disease.

Chapter 3.2: Proteomics analysis of anthracycline-induced cardiotoxicity from cardiac microtissues to human heart biopsies

Objectives of the study

The change of the proteome would directly influence cellular function and survival. In addition, the protein biomarkers can be transposable to clinic applications. Some studies have focused on ANY-induced protein activities. For example, the inhibition of membrane-bound calcium-independent phospholipase A₂ (iPLA₂) by ANTs restricts the recuperative capacity of isolated adult rat cardiomyocytes [62]. Another study, also performed in rats, demonstrated that the upregulation of SIRT1, a member of the Sirtuin protein family, improves cardiac function after DOX treatment [63]. A further study has explored the changes in protein levels in association with mRNA levels in the mitochondria of rat hearts after DOX treatment. It suggests that the alterations from transcript to protein level can be early acute markers for cardiac-specific mitochondrial toxicity [64]. Thus, the alteration of protein profiles could reveal different aspects of ANTs-induced heart failure and thereby provide a better understanding of ANTs' mechanism of cardiotoxicity. However, researchers so far mainly focused on the proteome upon DOX treatment, while little attention has been paid to the protein expression induced by other ANTs such as EPI and IDA.

In this chapter 3-2, we investigated the proteome-wide profiles under ANTs treatment including DOX, EPI, and IDA in human cardiac *in vitro* microtissue as well as the proteome profiles of heart failure biopsies from ANTs treated patients. Hence, this study accommodates a deeper understanding of ANTs-induced cardiotoxicity on the proteomic level and suggests some new proteins as prominent targets for ANTs-related heart failure study.

Methods and analysis procedures

Sample preparation of in vitro cardiac microtissues

When the cardiac microtissues were exposed to ANT conditions for 2 weeks, the proteome was extracted and measured at 7 time points, from 2 to 336 hours, except for samples treated with the toxic dose of IDA. The IDA-treated samples were harvested until 168 hours of exposure because of substantial cell death at later time points. Furthermore, due to the limited amount of microtissue, the IDA-treated

samples after 2 and 24 hours of exposure had only 2 replicates. Thus, the proteomics *in vitro* testing generated 139 samples in total.

After collection, cardiac microtissues were resuspended in 100 μ l lysis buffer containing 8M Urea, 1 mM Dithiothreitol, 0.1M Ammonium bicarbonate, pH 7.8. After four freeze-thaw cycles, the samples were centrifuged at 16000xg for 15 min at 4° C and protein concentrations were assessed with the Qubit™ Protein Assay Kit (Invitrogen, Molecular Probes). Protein isolates were then submitted to in-solution digestion [65] or Filter Aided Sample Preparation (FASP) [66]. Protein digestions were stopped by adding formic acid to a final concentration of 1%. The peptides were cleaned up using Sep-Pak tC18 cartridges (Waters), according to the manufacturer's instructions, and eluted with 60% CAN and 0.1% formic acid (Sigma-Aldrich, USA).

ATP measurement

After the microtissue collection, the ATP content in microtissues was measured as described in the previous study [67]. In short, the ATP level of the microtissues (n=4) was assessed using Promega's CellTier Glo 3D (Cat #G9683) according to the manufacturer's instructions. The luminescence was evaluated after 30 min incubation with the luciferase reagent [67].

Sample preparation of human cardiac biopsies

Cardiac biopsies (n=21) were taken from heart failure patients including patients (n=11) who had no cancer history as control subjects, patients (n=8) who had cancer and received medications including ANTs, and patients (n=2) who had cancer and received medications without ANTs (Figure 1, Table S2). Five biopsies (2 ANT-treated patients and 3 control patients) were used in a pilot study to establish the sample preparation workflow including the sample collection, sample processing, and data measurements. The sample preparation workflow was then applied to the remaining biopsies.

Samples were lysed and digested using a Barocycler NEP2320 (Pressure BioSciences) at 33 °C [68]. Briefly, each sample 1 (\pm 0.1) mg was lysed with a buffer containing 8M Urea, 0.1M Ammonium bicarbonate, and complete protease inhibitor (Roche), in combination with a cycling program of 50 seconds of ultrahigh-pressure (45,000 p.s.i.) and 10 seconds of ambient pressure (total of 60 pressure cycles). Protein reduction and alkylation were performed with 10 mM tris(2-carboxyethyl)phosphine and 40mM iodoacetamide, respectively. Protein digestions were performed sequentially with Lys-C and trypsin using PCT with a cycling scheme of 50 seconds at 20,000 p.s.i. and 10 seconds at ambient pressure.

Lys-C digestion was carried out in 6 M urea for 45 cycles, whereas trypsin digestion was performed in 1.6 M urea for 90 cycles. Protein digestions were stopped by adding trifluoroacetic acid (TFA) to a final concentration of 1%. The peptides were cleaned up using reverse-phase cartridges Finissterre SPE C₁₈ (Wicom International AG) according to the manufacturer's instructions.

Mass spectrometry (MS) measurements

Digested peptides from cardiac microtissues and cardiac biopsies were submitted to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) coupled to a NanoLC-2D HPLC system (Eksigent, Dublin, CA) or EASY-nLC 1000 system (Thermo Fisher Scientific, Germany). Samples were loaded onto a self-made column (75 μm \times 150 mm) packed with reverse-phase C₁₈ material (ReproSil-Pur 120 C₁₈-AQ, 1.9 μm , Dr. Maisch HPLC GmbH) when coupled with the EASY-nLC 1000 system and onto an Easy-Spray Column (75 μm \times 500 mm) packed with reverse-phase C₁₈ material (Silica 100 \AA , 2 μm) when coupled with the NanoLC-2D HPLC system. Peptides are separated with a linear gradient of acetonitrile/water, containing 0.1 % formic acid, at a flow rate of 300 nl/min. A gradient from 5 to 30% acetonitrile in 60 minutes was used. The mass spectrometer was set to acquire full-scan MS spectra (300–1500 m/z) at 120 000 resolution at 200 m/z; the precursor automated gain control (AGC) target was set to 400 000. A charge-state screening was enabled, and precursors with +2 to +7 charge states and intensities >5000 were selected for tandem mass spectrometry (MS/MS). Ions were isolated by using the quadrupole mass filter with a 1.6 m/z isolation window. Wide quadrupole isolation was used, and the injection time was set to 50 ms. The AGC values for MS/MS analysis were set to 5000 and the maximum injection time was 300 ms. HCD fragmentations were performed at normalized collision energy (NCE) of 30%. MS/MS was detected in the ion trap in centroid mode. Precursor masses previously selected for MS/MS measurement were excluded from further selection for 25 s, and the exclusion window was set at 10 ppm.

Peptide/Protein quantification and preprocessing data

Raw MS data were processed using Genedata Expressionist[®] software v.11.0. In short, after noise reduction and normalization, LC-MS peaks were detected and their properties calculated (m/z and RT boundaries, m/z and RT center values, intensity). Individual peaks were grouped into clusters and MS/MS data associated with these clusters were annotated with MS/MS Ions Search (Mascot 2.6) using Peptide Tolerance: 10.0 ppm, MS/MS Tolerance: 0.50 Da, Max Missed Cleavages: 2, and database: Uniprot Swiss-Prot 29062016, Taxonomy Homo

sapiens (human). Results are validated by applying a threshold of 5% normalized False Discovery Rate (FDR). Protein interference was done based on peptide and protein annotations. Redundant proteins were ignored according to the Occam's razor principle, and at least 2 peptides were required for positive protein identification (shared peptides were ignored). Protein intensities were computed using the Hi3 method. A maximum of the top 3 peptides per protein (based on the average intensity across samples) was used in the calculation. If a peptide was identified in multiple charges (2+, 3+, 4+) and modification states (Carbamidomethyl (C), Deamidated (NQ) or Oxidation (M)), values were consolidated into a single peptide intensity. The volume of a peak was computed as the area under the intensity curve inside the peak region. The area under the intensity curve is subdivided into trapezoids at the data points according to the trapezoidal rule. Volume is robust to different scan rates and takes more information (data points) into account. The intensities were log₂ transformed.

To normalize the *in vitro* proteomic data, the log₂ transformed values of the control samples were shifted to the median of the medians determined by a reference group consisting of the proteins found in all these control samples. For every ANT treatment and for each time-point, the common protein set between the controls and the ANT-treated samples was determined. The median of the medians of the (in general 3) normalized control samples was determined using this common protein set between the controls and the treatment samples. The data from the samples of the ANT treatments were shifted to these medians.

To normalize the heart biopsies data, the log₂ transformed values of each sample were shifted to a common median which was determined using only the proteins present in all samples.

Data analysis procedure

A workflow was applied to investigate the proteomics data from human *in vitro* cardiac microtissues exposed to ANTs and cardiac biopsies deriving from heart failure patients (Figure 3-5). The log₂ transformed normalized values were standardized to 6 decimal digits. Thereupon, proteins that matched multiple UniProt IDs were removed.

The protein expression patterns in the *in vitro* and biopsies datasets were analyzed separately (Figure 3-5) using the weighted correlation network analysis (WGCNA) package for finding modules of highly correlated proteins [69] in R (version 3.5.3, released on 11th March 2019). Proteins that had over 50% missing values across samples were removed using the default filter of the

goodSamplesGenes function in WGCNA. Hierarchical clustering sample trees were built from the remaining proteins. Thereafter, clustering protein trees were established based on the similarity of protein expression profiles across samples using the adjacency function with a signed network. The protein clustering tree of the *in vitro* data (using power =2) had topological overlap measures $R^2 = 0.64$ and mean connectivity = 59.6. The protein clustering tree of the human biopsies data (using power =5) had similar features with topological overlap measures $R^2 = 0.73$ and mean connectivity = 56.9. Proteins in the branches of protein clustering trees were divided into modules (groups) named by colours using the cutreeDynamic function (distM = dissTOM, deepSplit = 3, pamRespectsDendro = FALSE, minClusterSize = 30). These modules were merged when the difference between their module eigengene profiles was less than 0.25. Per module, proteins with high module membership (≥ 0.8), which demonstrate the strong impact of these proteins in the module, were considered as weighted proteins.

Module eigengene values (the first principal component) and the principal component analysis (PCA) were used to define ANT-affected modules. The biological interpretation of ANT-affected modules was explored by performing an over-represented pathways analysis using ConsensusPathDB with p-value and q-value < 0.01 [39]. Simultaneously, a reference of all proteins associated with a query term “heart failure” (572 proteins) in the DisGeNET database, a collection of genes and variants associated with human diseases, version 7.0, released on 4th May 2020 [70], was used to recognize the association of detected proteins to heart failure.

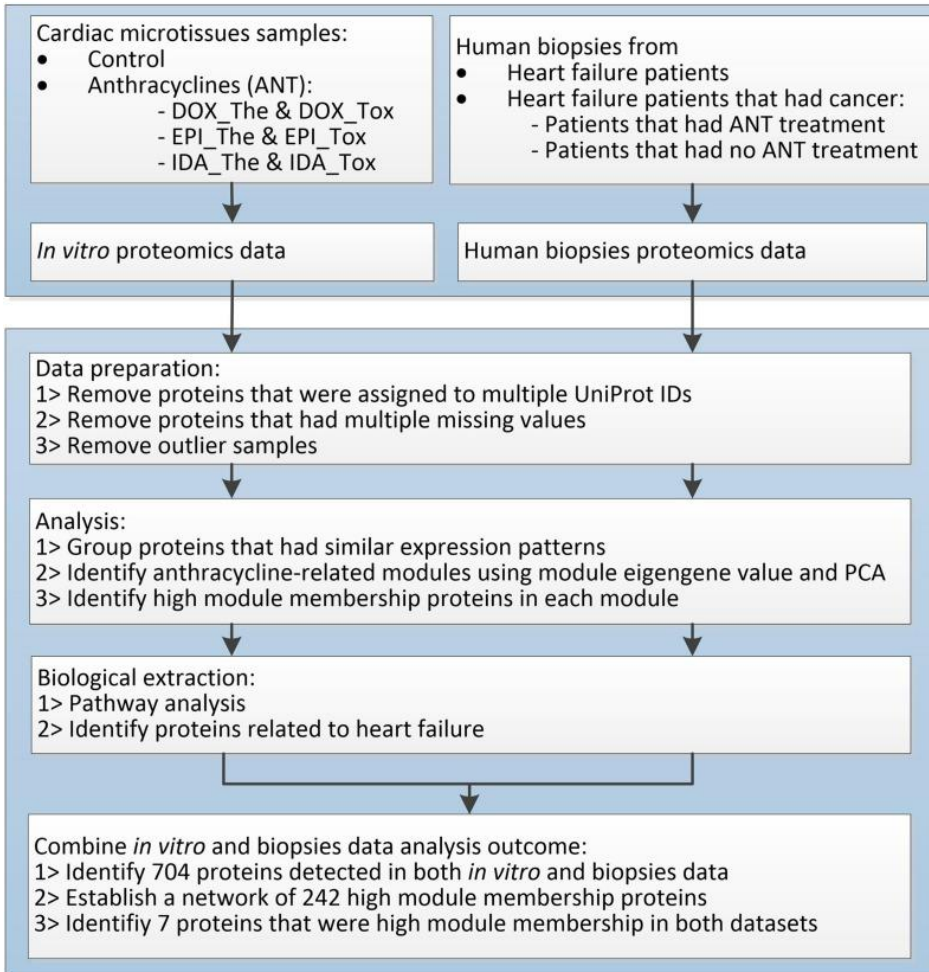


Figure 3-5: The proteomic data analysis workflow.

Proteins presented in both *in vitro* dataset and biopsies dataset were defined. The UniProt IDs of proteins, which were weighted proteins in at least one dataset, were used to build a protein network using the BisoGenet app with Homo sapiens species data, with non-adding connection, and other parameters following the default settings [71] in the Cytoscape version 3.7.1 [72]. The protein-protein interaction information is based on multiple sources comprised of the DIP, BIND, HPRD, MINT, Intact, and BioGrid databases [71]. Proteins, which were weighted proteins in both datasets, as well as the hub proteins, which have ≥ 30 connections in the protein network, were identified.

Results

In vitro data

In total, 2497 proteins were detected among 139 cardiac microtissue samples with 2327 proteins having a unique UniProt ID. After removing proteins with over 50% missing values across samples, the remaining 810 proteins were used to construct the hierarchical clustering samples tree (Figure 3-6A). The control samples were grouped in a separate branch, while samples treated by ANTs were clustered in another branch. Among the ANTs samples, IDA-treated samples converged at one sub-branch, whilst samples treated with DOX and EPI are in two closely connected sub-branches, except for the samples treated with the EPI toxic dose at late time points (168, 240, 336 hours exposure). Thus, the proteomic profiles demonstrate a clearer separation between the therapeutic and toxic dose compared to the transcriptome profile of the same ANT-treated samples. The ATP measurement also emphasized that ANT-induced mitochondrial dysfunction was dose-dependent, especially after long-time exposure. In particular, with the same ANT analog, the toxic-treated samples had lower ATP levels than the therapeutic-treated samples, except for the IDA-treated samples after 72 hours of exposure that under ANTs exposure (Figure 3-7).

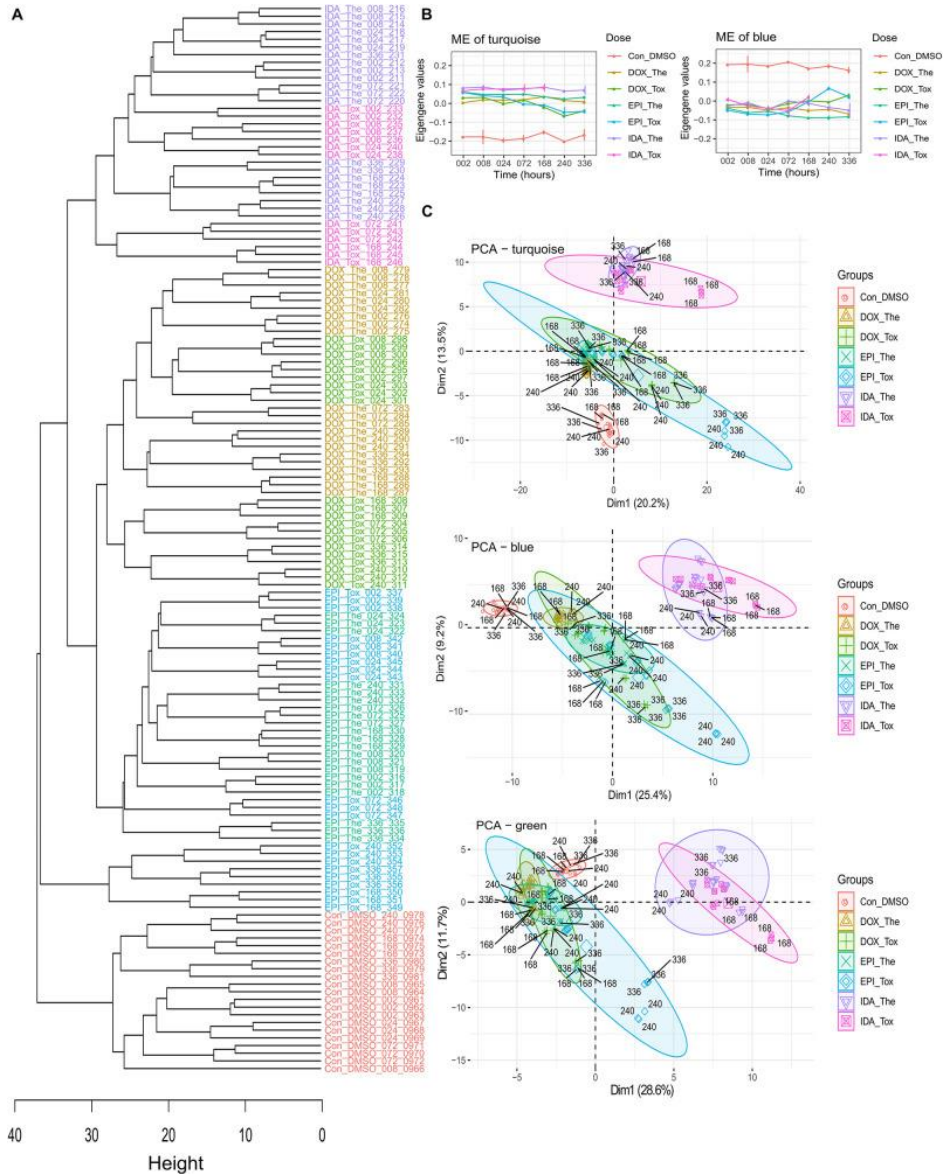


Figure 3-6: The in vitro cardiac microtissue proteomics data analysis. **(A)** The hierarchical clustering tree is based on the in vitro protein expression values. The control, epirubicin, doxorubicin, and idarubicin samples were colored red, blue, green, and purple respectively. **(B)** The module eigengene (ME) values in WGCNA modules. **(C)** The PCA of the modules turquoise, blue, and green. Drugs: DOX: doxorubicin, EPI: epirubicin, IDA: idarubicin. Treatment doses: The: the therapeutic dose, Tox: toxic dose. Exposure period: 002, 008, 024, 072, 168, 240, and 336: 2, 8, 24, 72, 168, 240, and 336 hours of exposure, respectively.

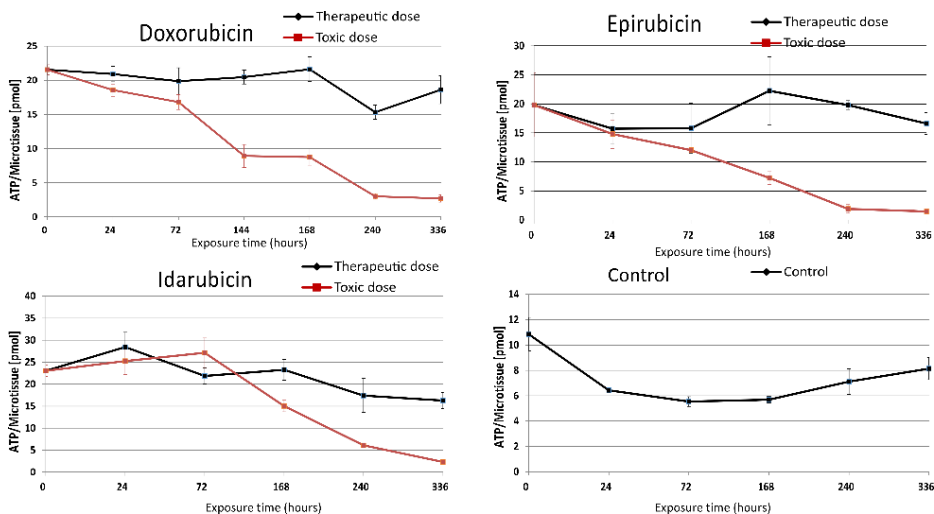


Figure 3-7: The ATP contents of microtissues in doxorubicin, epirubicin, and idarubicin-treated samples and controls.

These 810 proteins were categorized into 6 modules based on their expression profiles across ANT-treated conditions. The large modules were the turquoise and blue modules, which consisted of 262 and 194 proteins respectively, while other modules (brown, yellow, green, and red) consisted of smaller numbers of proteins (Table 3-2). In the turquoise and blue modules, their eigengene values (the first principle component) demonstrated distinctions in protein expression levels between ANT-treated samples and control samples, while this did not happen in other modules (Figure 3-6B, Figure S 2-1). However, the PCA plots indicated that protein profiles in the green module also showed a separation between ANT-treated samples and control samples (Figure 3-6C, Figure S 3-3). By using module eigengene values and PCA, we identified 3 protein modules (turquoise, green, and blue) in which the difference in protein expressions between ANTs and control conditions was evident. The high module membership proteins in the *in vitro* protein modules were identified as weighted proteins (Table 3-2).

Subsequently, the pathway analysis in ConsensusPathDB (p -value < 0.01, q -value < 0.01) indicated that the proteins in the turquoise and blue modules are mainly involved in translation and protein metabolism pathways. The pathway analysis performed on only weighted proteins in the turquoise and blue modules showed a similar outcome (Table S 3-7). While the proteins in the green module are related to the metabolism pathways, the number of weighted proteins in this module was small (3 proteins) and appeared non-overrepresented in any pathway

(Table S 3-7). Multiple proteins belonging to these modules are known as heart failure-related proteins in the DisGeNET database (Table 3-2).

Table 3-2: The number of proteins in WGCNA modules in the *in vitro* dataset.

	Total proteins	Turquoise module	Blue module	Brown module	Yellow module	Green module	Red module
Number of proteins	810	262	194	148	96	80	30
Number of proteins related to heart failure	101	25	14	12	4	7	0
Number of the high module membership proteins	-	10	8	19	3	3	0
Number of the high module membership proteins related to heart failure	-	1	0	3	0	0	0

Human biopsies data

In 21 cardiac biopsies from heart failure patients, there were 1639 proteins with a unique UniProt ID out of 1669 proteins that were detected. Five biopsies were used to establish the sample preparation workflow before this workflow was applied to the remaining 16 biopsy samples. Therefore, the first 5 biopsies (2 patients that had no cancers and 3 cancer survivors that had received ANT treatments) showed large distances to the latter 16 biopsy samples in the hierarchical clustering sample tree (Figure S 3-4). The latter 16 biopsy samples, including 8 patients that had no cancers as a control group, 6 patients that had undergone cancer treatment including ANTs, and 2 patients that had received cancer treatment without ANTs, were used for further analysis. After removing proteins with over 50% missing expression values across samples, the remaining 1602 proteins were used for further analysis. The hierarchical clustering sample tree built from the 16 samples could not define patient groups according to their medical history (Figure 3-8A). The PCA plot also confirmed that there was no apparent difference between cardiac biopsy samples taken from the different patient groups (Figure S 3-5).

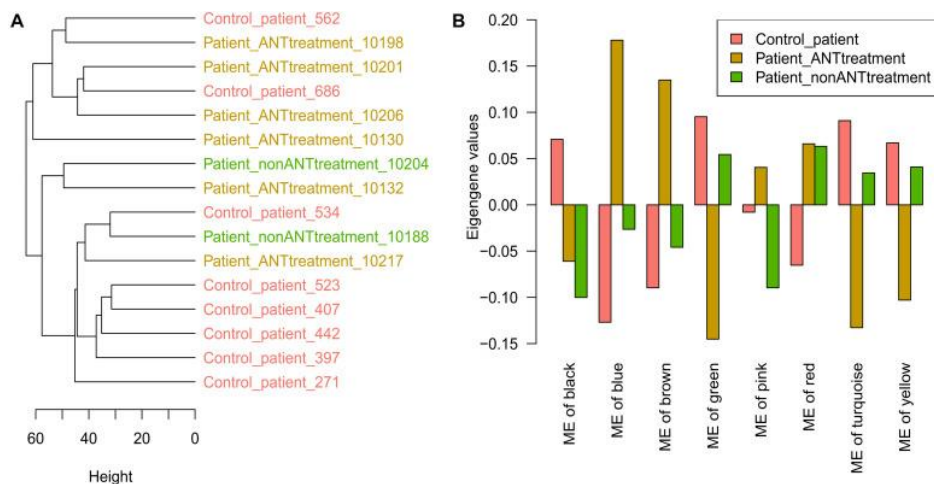


Figure 3-8: Human cardiac biopsies proteomics data analysis. **(A)** The hierarchical clustering tree is based on the protein expression values in biopsy samples. **(B)** The module eigengene (ME) values in WGCNA modules. Control_patient: heart failure patients with no cancer history; Patient_ANTtreatment: heart failure patients that had cancer treatment with anthracyclines (ANTs); Patient_nonANTtreatment: heart failure patients that had cancer treatment without ANTs; the number at the end of each patient indicates the biopsies sample ID.

The WGCNA analysis divided the 1602 proteins into 8 modules based on their expression pattern across the patient groups. Although these modules consisted of different protein sets, 6 of them were also named by the same colors as the *in vitro* data analysis. The large modules were the turquoise, blue, and brown modules, which consisted of 342, 330, and 320 proteins, respectively, while other modules (yellow, green, red, black, and pink) consisted of smaller numbers of proteins (Table 3-3). In the turquoise, blue, brown, yellow, and green modules, their eigengene values manifested the difference in protein expressions between control patients and patients who had undergone cancer treatment including ANTs treatment (Figure 3-8B). The PCA plot did not identify additional modules that showed differences in protein expression between ANT-treated and control groups (Figure S 3-5). Weighted proteins were defined from the high module membership proteins in each module (Table 3-3).

Table 3-3: The number of proteins in WGCNA modules in the human biopsies dataset.

	Total proteins	Turquoise module	Blue module	Brown module	Yellow module	Green module	Red module	Black module	Pink module
Number of proteins	1602	342	330	320	213	177	135	47	38
Number of proteins related to heart failure	159	22	22	21	18	9	10	7	2
Number of the high module membership proteins	-	115	95	65	54	53	41	13	5
Number of the high module membership proteins related to heart failure	-	11	8	8	4	3	5	1	1

The pathway analysis indicated that the proteins in these 5 selected modules from the biopsies data are related to different cellular mechanisms comprising metabolism, mitochondrial function, muscle contraction, and signaling pathways. The pathway analysis for only weighted proteins in these modules showed a similar outcome (Table S 3-8). According to the DisGeNET database, multiple proteins in these selected modules are known as heart failure-related proteins (Table 3-3).

Combining the in vitro and human biopsies data

Combining proteins used in the *in vitro* cardiac microtissue and human biopsies analysis (810 and 1602 proteins, respectively) resulted in 704 proteins present in both datasets. Of these, 242 proteins (34.4%) were weighed in WGCNA modules in at least one dataset, and 7 of them were weighted proteins in both datasets (Table 3-4). Of the 7 weighted proteins in both datasets, DECR1, SH3BGRL, and ATP5F1B have been recognized as heart failure-related proteins [70,73,74]. Although the role of ETFB in the late stages of heart failure is not clear yet, a cohort study showed that ETFB is strongly associated with chronic ANT-induced cardiotoxicity [75]. The remaining proteins, i.e. EEF1D, TIMM13, and PMPCB, are involved in transferring aminoacyl-tRNAs, regulating heat-shock response, and maintaining mitochondrial functions [76], but their roles in heart disease contexts have not been investigated.

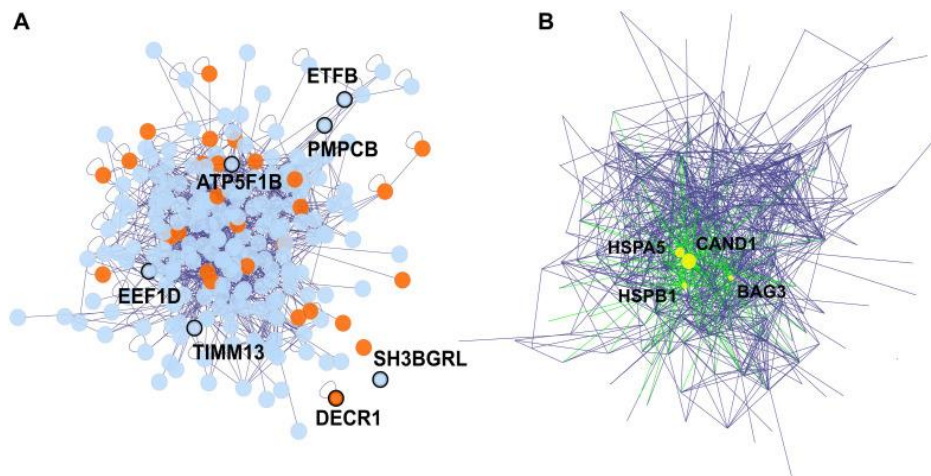


Figure 3-9: The protein-protein interaction network of proteins detected in both in vitro and biopsies datasets, and that were weighted proteins in at least one dataset. Edges represent the protein-protein interaction, and nodes represent proteins, 16 orphan nodes were removed from the network. (A) The 7 nodes with protein names in bold and showing thicker borders were weighted proteins in both in vitro and biopsies datasets. Proteins associated with heart failure (detected via DisGeNET), were highlighted in orange. (B) The 4 hub proteins (> 30 degrees of connection) of the protein-protein interaction network are shown as yellow nodes.

Table 3-4: The high module membership proteins in both in vitro and human biopsies datasets.

		ANTs – related modules in the in vitro data		
		Turquoise	Brown	Green
ANTs – related modules in the biopsies data	Turquoise	-	-	-
	Blue	SH3BGRL (O75368) EEF1D (P29692)	-	-
	Brown	TIMM13 (Q9Y5L4)	ATP5F1B (P06576) ETFB (P38117) DECR1 (Q16698)	PMPCB (O75439)
	Yellow	-	-	-
	Green	-	-	-

Notes: The protein names were retrieved from UniProt ID.

The 242 proteins which were weighted proteins in at least one dataset, were used to establish a protein-protein interaction network via the BisoGenet app in Cytoscape (Figure 3-9). In this network, 28 proteins (11.57%) are known as heart failure-related proteins (Figure 3-9A). We identified 4 hub proteins which are

nodes with a high degree of connectivity and are key connectors between proteins in the network: CAND1, HSPA5, HSPB1, and BAG3 (Figure 3-9B, Figure S 3-6). Literally, these hub proteins are essential components in the proteome: CAND1 contributed to the ubiquitin complexes, HSPA5 and HSPB1 are heat shock proteins involved in stress responses, while BAG3 is a co-chaperone for HSP70 and HSC70 chaperone proteins [76]. The alterations in these hub protein expressions under ANT treatment can systematically spread across the protein network, and lead to the cascade effect of ANTs.

Expression of the weighted proteins and the hub proteins

Of these 7 weighted proteins in both datasets and 4 hub proteins, their expression level varied over time across ANT treatment conditions in the *in vitro* data, as well as varied across patient groups in the biopsies data. Of these 11 proteins, 6 proteins i.e. ATP5F1B, EEF1D, ETFB, DECR1, HSPA5, and HSPB1, had high expression (\log_2 expression > 10) in all samples (Figure S 3-7). The \log_2 fold change (\log_2 FC) is the log ratio of protein expression in ANT treatments compared to the protein expression in equivalent control samples. In the *in vitro* testing, \log_2 FCs of these 6 proteins were used to manifest the up and down-regulated proteins over time under ANT effects after correcting for the correspondent baseline in control samples (Figure 3-10A). Regarding the human heart biopsy samples, the \log_2 FC of these 6 proteins represents differences between cancer patients who received ANT and non-ANT treatment versus non-cancer patients (Figure 3-10B).

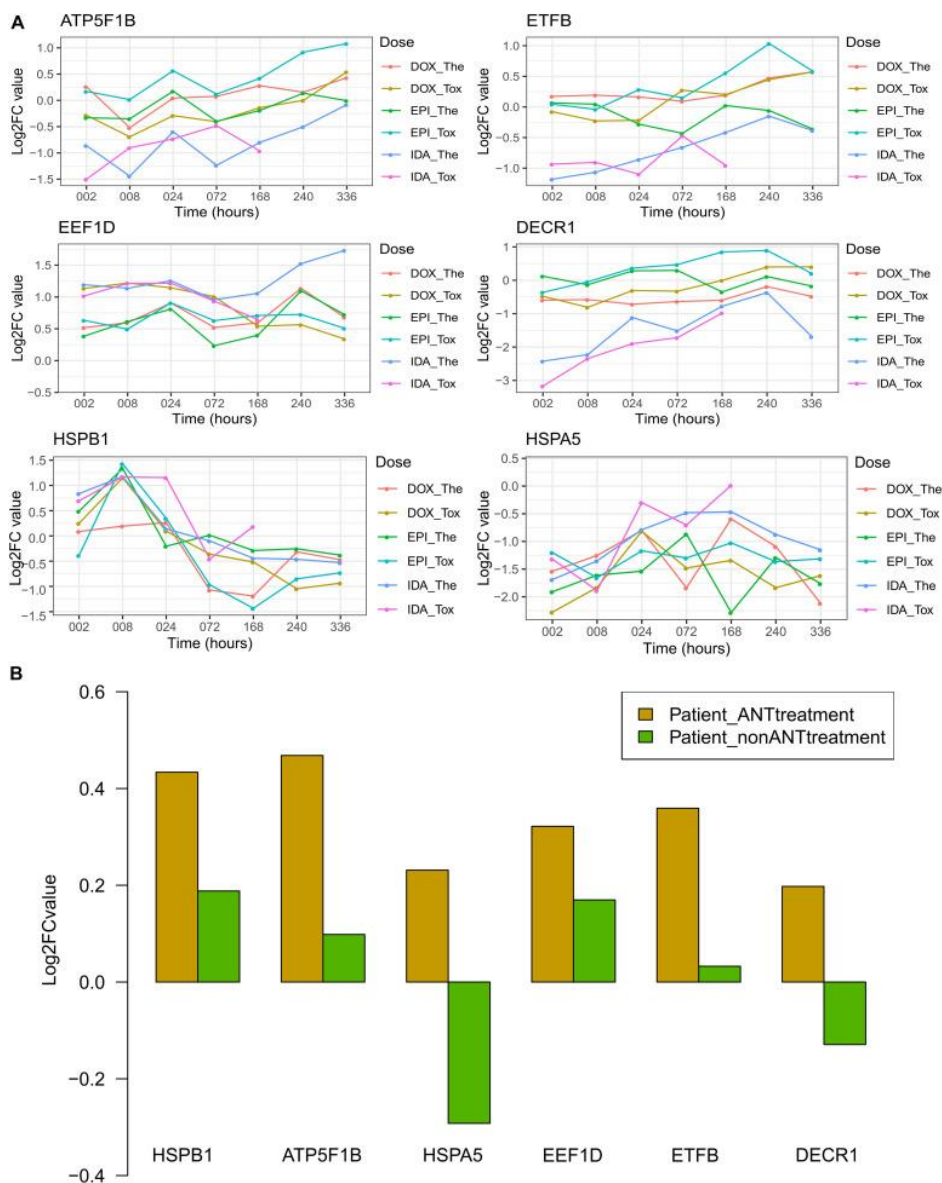


Figure 3-10: The log₂FC expression of selected proteins in the in vitro and human biopsies datasets. Protein names and its UniPort IDs: ATP5F1B (P06576), EEF1D (P29692), ETFB (P38117), DECR1 (Q16698), HSPA5 (P11021), HSPB1 (P04792). (A) The log₂FC value of in vitro data resulted from the comparison of protein expression levels in anthracyclines (ANTs) samples to control samples per time point. (B) The log₂FC value of human biopsies data resulted from the protein expression levels in heart failure patients having cancer therapy with or without ANTs compared to the heart failure patients with no cancer (control group). Patient_ANTtreatment: heart failure patients that had cancer treatment with ANTs; Patient_nonANTtreatment: heart failure patients that had cancer treatment without ANTs.

Generally, the *in vitro* protein expression in IDA-treated samples differed from DOX- and EPI-treated samples, especially for the 4 proteins i.e. ATP5F1B, ETFB, EEF1D, DECR1 (Figure 3-10A). The DECR1 and ETFB expression in the IDA-treated samples was downregulated at early time points, then increased and were closer to their expression in DOX and EPI-treated samples (Figure 3-10A). These 2 proteins show that the effect of ANT analogs could be diverse in short-term treatments, but become more convergent after long-term treatment. The log₂FCs of ATP5F1B in ANTs-treated samples increased at later time points (from 168 to 336 hours), despite their fluctuations at earlier time points in the *in vitro* data (Figure 3-10 A). Interestingly, this is in agreement with the higher expression levels of ATP5F1B in the heart biopsies taken from patients treated with ANTs compared to control patients and patients treated without ANTs (Figure 3-10B). In addition, although the log₂FCs of EEF1D differed between *in vitro* ANTs-treated samples, their log₂FCs were always positive (Figure 3-10A). Similarly, EEF1D was overexpressed in the patients treated with ANTs compared to control patients and patients treated without ANTs (Figure 3-10 B). For the other 4 proteins (ETFB, DECR1, HSPB1, and HSPA5), the *in vitro* data did not clearly show the same pattern as apparent in the human biopsies data (Figure 3-10).

Discussion

In this chapter 3-2, we investigated how ANTs alter the protein expressions and could influence cellular mechanisms and promote heart failure in human cardiac microtissue as well as in human cardiac biopsies from ANTs-treated patients. ANTs, as a drug family, can share a common mechanism of toxicity and reshape cardiac protein expressions. Furthermore, DOX, EPI, and IDA can have their variant adverse effects on cardiac tissue, which could also be captured by the protein expression alterations.

The *in vitro* samples were grouped according to their protein expression profiles in the clustering tree (Figure 3-6A). The DOX and EPI-treated samples shared a high similarity in protein expressions not only in the clustering tree but also in module eigengene values and PCA plots (Figure 3-6A-C). This confirms the prior knowledge that EPI is a derivative of DOX, and shares a similar mechanism with DOX, while IDA is an analog derived from daunorubicin, another ANT [3,4]. The lipophilic ability could be an underlying factor that leads to the differences in protein profiles, in which IDA has higher lipophilicity than its parent molecule, daunorubicin [77], while daunorubicin has higher lipophilicity than DOX [78]. Interestingly, a rat cardiomyoblast study showed the inverse correlation between the lipophilicity of ANT analogs and their toxicity [78].

The *in vitro* model, as a well-controlled system, seems to be able to capture subtle divergences across ANT analogs, the human biopsies could not be grouped by the protein profiles in accordance with the patients' medical history (Figure 3-8A). Possibly, all participants were heart failure patients, so they shared quite similar protein expression profiles. Furthermore, clinical treatments usually incorporate different ANT analogs with other drugs such as docetaxel, cytosine arabinoside, vincristine, etc., rather than using a single ANT as a monotherapy (Table S 3-1). This combined chemotherapy has been recommended because of its survival benefit and cost-effectiveness [60,61]. However, it also causes challenges for ANT-induced cardiotoxicity studies. Even though patient data is certainly complex and reflects multifactor treatments, the combination of *in vitro* and human biopsies data could still assist to project the outcome of an empirical experiment to clinical applications.

The ANT-affected protein groups were acquired according to protein expression across samples. In the *in vitro* data, 3 protein groups (turquoise, green, and blue) demonstrated not only how certain ANT analogs but also how different doses and time exposure can impact protein expressions (Figure 3-6B-C). In the PCA plots, samples from short-time exposure were grouped together, and samples treated with toxic doses and from longer exposures (168, 240, and 336 hours) were more dispersed (Figure 3-6C). An early clinical study showed that ANTs possibly cause myocardial damage after 24 hours following drug administration; however, patients may recover after 72 or 96 hours [79]. Related to this clinical phenomenon, the proteome observed in the *in vitro* ANTs samples at early time points could reflect the acute ANTs toxicity, while the proteome observed at later time points represented the intermediate and late toxicity. Thus, evaluating the proteome after long-term exposure such as 168 hours (1 week) might elucidate more relevant aspects of ANT chronic cardiotoxicity. This approach also has been proposed in another study that predicts ANT cardiotoxicity in patients before the obvious clinical symptoms develop using a serial assessment of the left ventricular function in 1-3 weeks after treatment [79,80]. In the biopsies, 5 protein groups (turquoise, blue, brown, yellow, and green modules) highlighted the difference in the protein expressions between the heart failure patients groups (Figure 3-8B).

A part of the ANTs-affected proteins was involved in cellular metabolisms in both the *in vitro* and cardiac biopsy samples (Table S 3-7, Table S 3-8). The ANTs-affected proteins in the *in vitro* dataset also belonged to the translation pathway (Table S 3-7). Research has indicated that DOX alters the transcription process via signaling factors such as the transcription factor NF- κ B, the transcription factor

GATA4, or through the PI3K-dependent signaling pathway [12]. Furthermore, the ANTs-affected proteins in the human biopsies dataset illustrated the impact of the ANT treatment on the mitochondrial function, muscle contraction, and signaling pathways in the long term (Table S 3-8). While mitochondrial dysfunction has been a longitudinal topic in investigating ANT-side effects [81], signaling pathways have especially emerged as a new paradigm of ANT-induced cardiotoxicity [12].

The 704 proteins were detected in both the *in vitro* and the biopsies datasets, and 7 of them were weighted proteins in both datasets (Figure 3-9, Table 3-4). These proteins suggest an extrapolation from *in vitro* outcomes to *in vivo*, in which ANTs demonstrated strong impacts in the 7 overlapping weighted proteins in both *in vitro* and biopsy samples. For instance, ATP5F1B, an ATP synthase subunit beta protein that belongs to the ATP synthase complex [76], was up-regulated in both *in vitro* and biopsy samples (Figure 3-10). It suggested compensation of the mitochondrial disfunction under ANT treatment. However, the ATP levels still decreased in all *in vitro* ANTs samples after long-term exposure (Figure 3-7). EEF1D, another weighted protein, is involved in transferring aminoacyl-tRNAs to the ribosome and regulating heat-shock-responsive genes [76]. This protein was triggered by ANTs and up-regulated from the beginning of the *in vitro* ANT treatments as well as in the ANTs-treated heart failure patients (Figure 3-10). The consistency in ATP5F1B and EEF1D's expression pattern between the *in vitro* samples and human cardiac biopsies suggests that these two proteins could be potential targets to predict ANT-induced cardiotoxicity.

The other 5 weighted proteins also play important roles in cellular function and are associated with heart failure. DECR1, an enzyme in the mitochondrial fatty acid beta-oxidation pathway (2,4-dienoyl-CoA reductase), has been known as a heart failure-related protein in the DisGeNET database [70]. The promoter of weighted protein SH3BGRL contains an NF- κ B binding site bounded, thus it can be regulated by the Rel/NF- κ B family [82]; while the activation of NF- κ B can be modulated by ANTs [12]. Furthermore, a clinical study showed that SH3BGRL was differentially expressed on the transcriptomic level between heart failure biopsies of non-ischemic cardiomyopathy patients and non-heart failure biopsies from unused cardiac transplant donors [83]. Even though it is unclear how SH3BGRL is directly involved in heart failure, SH3BGRL has evolved into a biomarker for identifying cardiotoxic agents and for diagnosing heart diseases [73]. The other weighted proteins, i.e. ETFB, PMPCB, and TIMM13, are mitochondrial proteins (Figure 6) and are mediated by TLR4 and NF- κ B activation, which can be triggered by ANT [12,84]. ETFB is involved in mitochondrial electron transfer and

is strongly associated with chronic anthracycline-induced cardiotoxicity [75]. PMPCB (aka β -MPP) belongs to the mitochondria proteases, while TIMM13 is a chaperone-like protein; both of them are essential for importing and modulating proteins in mitochondria [76,85,86].

The protein network represents inter-dependence interactions between heart failure-related proteins and other proteins (Figure 5). Four hub proteins in the network were CAND1, HSPA5, HSPB1, and BAG3 (Figure 5B), which serve as key bridges for protein-protein interactions. CAND1 is an important assembly factor in SCF E3 ubiquitin ligase complexes [87], which manipulate the turnover and function of the sarcomere proteins and the apoptosis signaling pathway in cardiomyocytes. The alterations in ubiquitin-proteasome (E3s) are associated with cardiac dysfunction [88]. Another hub protein, HSPA5, is a member of the heat shock protein family A (Hsp70). This protein maintains protein homeostasis and Ca²⁺ homeostasis in the endoplasmic reticulum, as well as activates the unfolded protein response pathway and induces autophagy. HSPA5 is a current target for protecting cardiomyocytes because it is an oxidative stress sensor and responder and can rescue cardiomyocytes from apoptosis [89]. The HSPA5 level under ANT treatment was fluctuated since early time exposures (Figure 3-10), while HSPA5 was expected to be gradually changed due to its long half-life (over 30 days) [89]. This suggests that HSPA5 expression was immediately impacted by ANTs, and could not protect cardiomyocytes, which lead to chronic injury of heart tissue. Another heat shock protein, i.e. HSPB1, participates in different cell functions including stress resistance [76]. HSPB1 was up-regulated in mice cardiomyocytes as a response to myocardial infarction and involved in repairing tissue damaged by inhibiting NF κ B inflammatory signaling [90]. In this study, HSPB1 levels were up-regulated in all *in vitro* ANT-treated samples in early time exposure (2-24 hours) (Figure 3-10A), which may indicate an acute response to ANT treatment as a stress responder. However, after 24 hours, the HSPB1 expression decreased in all ANT-treated samples (Figure 3-10A). The last hub protein is BAG3, which is a co-chaperone for HSP70 and HSC70 chaperone proteins and is involved in a wide range of different cell functions [76]. BAG3 has been reported to maintain cardiomyocyte function during proteotoxic stress and may become a target for heart failure therapy in the clinic [91,92].

This chapter 3-2 provided a broad picture of protein profiles under ANTs exposure as well as a proteomics analysis approach focused on the relevance of dedicated *in vitro* experiments to clinical data. The *in vitro* cardiac microtissues experiment showed a distinction of protein profiles between early and later

exposure, which suggests different cellular responses for acute and intermediate/chronic ANT toxicity. A part of the proteins shows similar expression patterns between the *in vitro* cardiac microtissues and human biopsies under ANT exposure. Some of these proteins belong to traditional ANT-affected pathways, while other proteins are involved in signaling pathways such as the NF- κ B signaling pathway, which is a prominent factor in the recent ANT-induced cardiotoxicity paradigms. While some detected proteins (SH3BGRL, HSPA5, and BAG3) are current cardiotoxicity biomarkers or targets of cardiac therapy, other proteins, such as ATP5F1B and EEF1D, could be potential biomarkers for cardiotoxicity and may reveal new insights into the proteome alterations caused by ANTs in the *in vitro* human cardiac tissues and translate to the patient situation.

General remarks

In chapter 2, we have explored the effects of EPI, a member of the ANT family, on DNA methylation status; subsequently, in this chapter 3, we investigated the influence of ANTs on transcriptome and proteome to suspect how ANTs affect cellular mechanisms and cause heart failure. In particular, we analyzed the transcriptomic and proteomic data derived from cardiac microtissues exposed to ANTs as well as from human cardiac biopsies collected from ANT-treated patients.

From the informatics perspective, investigating gene and protein expression plays a key role in elucidating the underlying drug mechanisms of action. Exploring lncRNAs (chapter 3-1) revealed that several lncRNAs demonstrated their potential roles in adverse ANT side effects related to heart disease progression. This extends the research boundary from only focusing on the expression of protein-coding genes to the expression of non-coding RNA genes. Furthermore, by combining the *in vitro* data and human cardiac biopsies data, we aimed to translate the alternation patterns from the *in vitro* experiments to the observations in clinical settings. Especially, the alterations of some candidate lncRNAs and proteins' expression in the *in vitro* cardiac tissue were also affirmed by similar changes in the human biopsies.

From a toxicology perspective, while ANT is an important chemotherapeutic family, its adverse effects can damage cardiomyocytes, cause cardiac dysfunction, and lead to heart failure. Although researchers have proposed several ANT-induced cardiotoxicity paradigms [3,10,12,93,94], full insight into the ANT mode of action is still missing; thus, further investigation is still crucial to define the ANTs cardiotoxic mechanisms. Especially, ANTs, as a drug family, can share a common mechanism of toxicity, but individual ANTs' analogs such as DOX, EPI, and IDA can have their variant adverse effects on cardiac tissue [95]. However, most studies have focused on DOX, the first generation of ANTs, whereas there is not much research on the second generation of ANTs including EPI and IDA [3]. Thus, investigating transcriptome and proteome under different ANT analogs' exposure could elaborate on the diversity in ANT-induced cardiotoxicity.

In this chapter 3, we analyzed both transcriptomic and proteomic profiles of the same *in vitro* microtissues exposed to DOX, EPI, and IDA. In comparison with transcriptome, proteome conveyed a clearer sample separation according to analogs and doses at the proteome rather than transcriptome (Figure 3-1, Figure 3-6). It could be that there are more disruptions in the transcriptome than in the proteome. We also attempted to link the *in vitro* outcome to the human biopsies

for both transcriptomic and proteomic analysis, but there is consistently non-vital difference in those molecular profiles between ANT-treated patients and control subjects. Corresponding to the *in vitro* transcriptome and proteome-wide analysis, several lncRNA and proteins were highlighted as potential biomarkers or targets for further ANT-induced mechanisms investigation. We also inspected their expression patterns in human biopsies under ANT exposure to pursue the transparency of the detected outcome from *in vitro* experiments to the clinical phenomenon.

Supplementary Materials

Data & code accessibility

The data presented in the study are deposited in the BioStudies repository (<http://www.ebi.ac.uk/biostudies>). The accession number of the RNA-seq data from DOX, EPI, and IDA *in vitro* samples are S-HECA₁₀, S-HECA₁₁, and S-HECA₁₂, respectively. The accession number of the RNA-seq data from biopsy samples is S-HECA₄₆₉. The accession number of the proteomics data is S-HECA₁₀₄ for *in vitro* proteomics data and S-HECA₅₀ for human biopsies proteomics data.

The code of the data analysis is available on Github for the RNA-seq analysis (https://github.com/NhanNguyenooo/lncRNA_ANT) and the proteomics analysis (https://github.com/NhanNguyenooo/Anthracycline_Protein_Analysis).

Supplementary figures

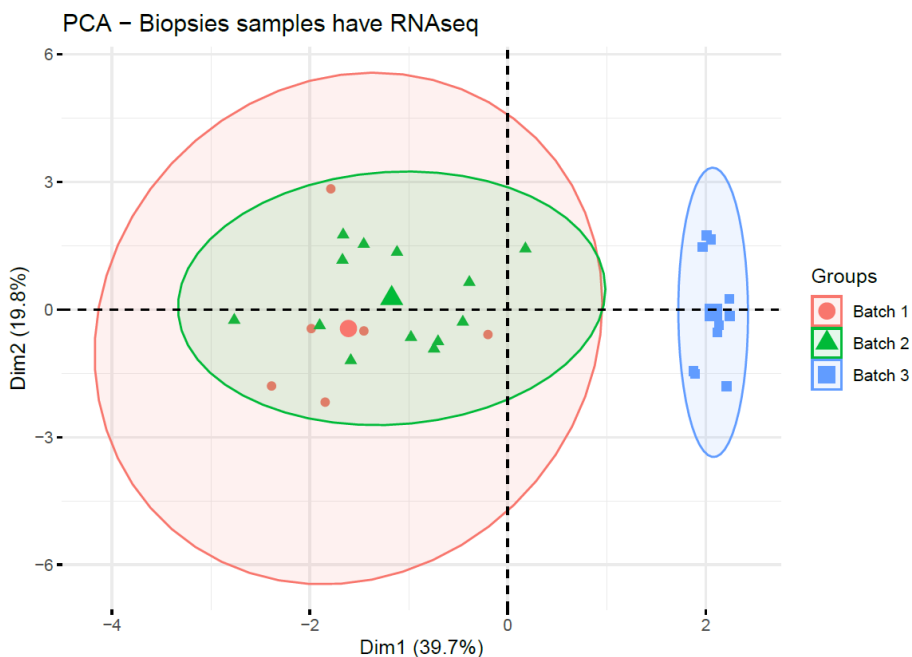


Figure S 3-1: The transcriptome profile of biopsy samples in batches in the PCA plot

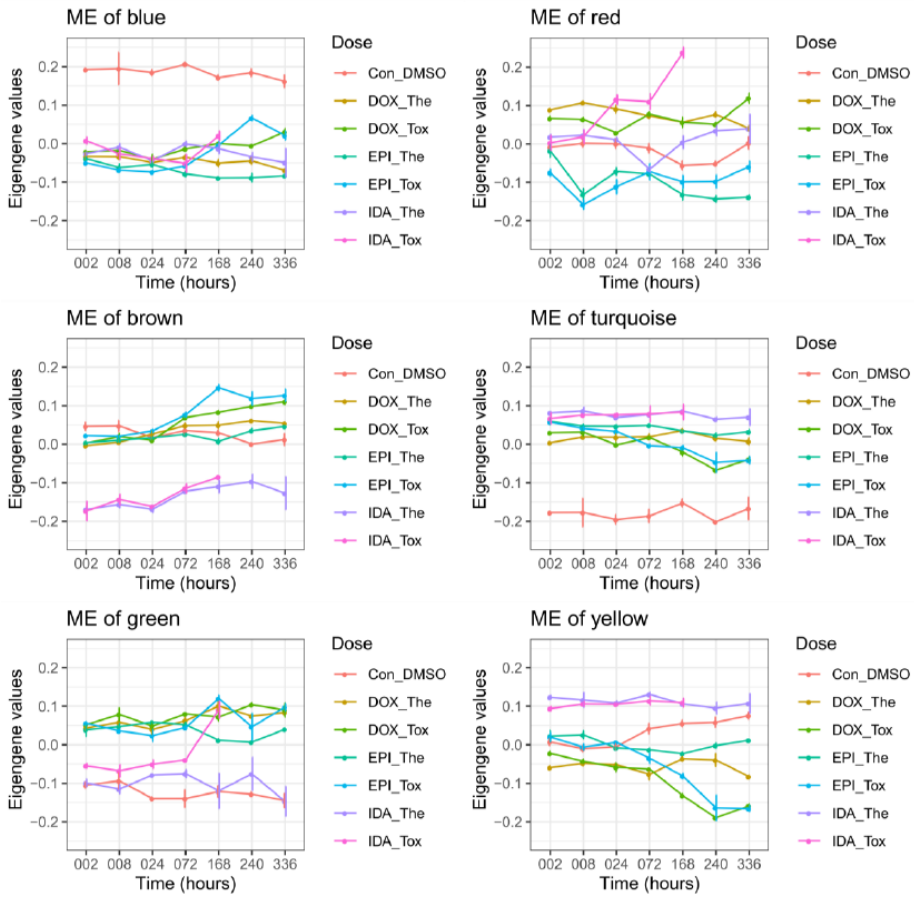


Figure S 3-2: Module eigengene (ME) values in all *in vitro* modules

Figure S 3-3: PCA plots in all *in vitro* modules. Please check Figure S2 in the supplementary materials are publicly available at the published proteomics article <https://doi.org/10.3389/fgene.2021.695625>

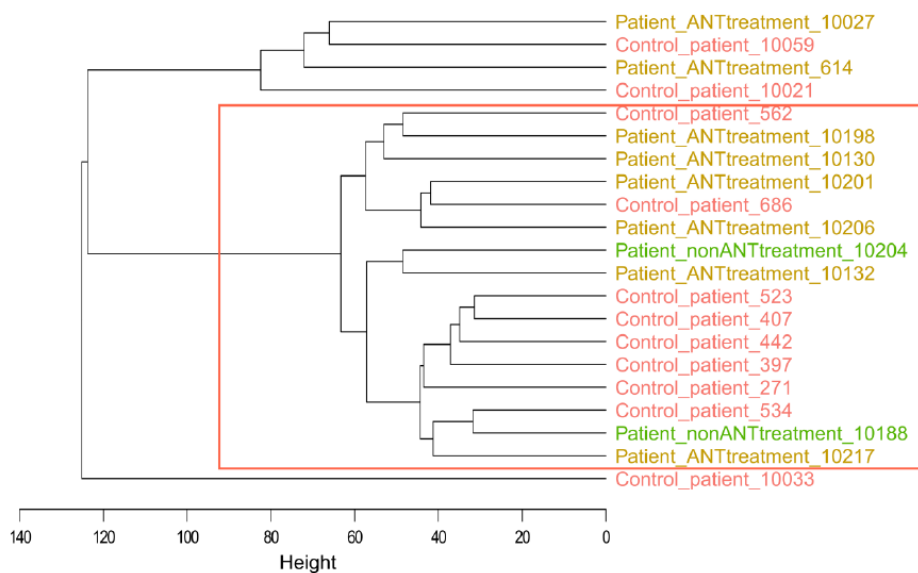


Figure S 3-4: Hierarchical clustering tree with all samples in the human cardiac biopsies dataset. The samples in the red rectangular belonged to one batch, while the rest of the samples belonging to another batch, were considered outliers and were removed for further analyses. Control_patient: heart failure patients with no cancer history; Patient_ANTtreatment: heart failure patients that had cancer treatment with anthracyclines (ANTs); Patient_nonANTtreatment: heart failure patients that had cancer treatment without ANTs; the number at the end of each patient indicates the biopsies sample ID.

Figure S 3-5: PCA plot from all proteins in the biopsies dataset and proteins in each biopsies module. Please check Figure S4 in the supplementary materials that are publicly available in the published proteomics article. <https://doi.org/10.3389/fgene.2021.695625>

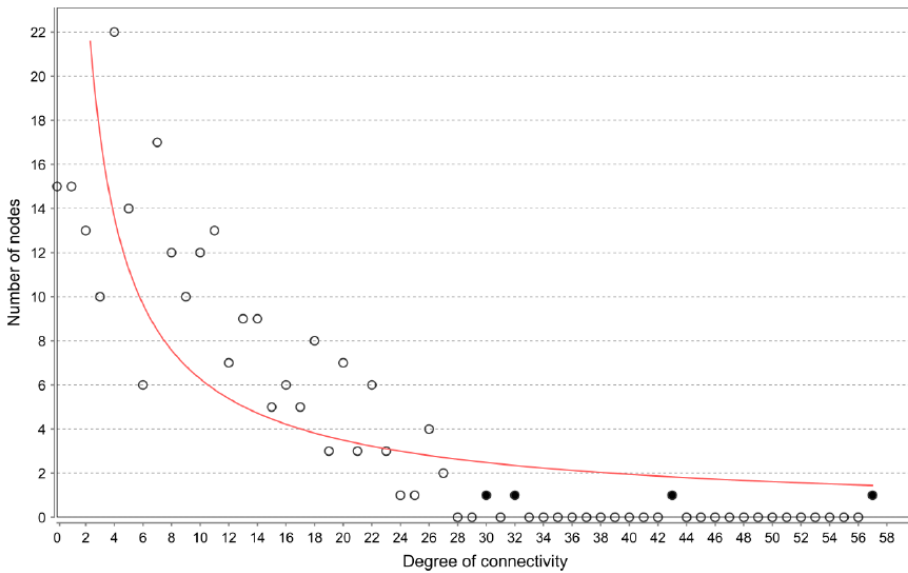


Figure S 3-6: Protein network analysis - the number of nodes with their degree of connection. Four hub proteins (degree of connection ≥ 30) were highlighted in black: CAND1 (Q86VP6) – 57 connections, HSPA5 (P11021) – 43 connections, HSPB1 (Po4792) – 32 connections, BAG3 (O95817) – 30 connections.

Figure S 3-7: The log expression of selected proteins in the *in vitro* and human biopsies datasets. Please check Figure S6 in the supplementary materials that are publicly available in the published proteomics article. <https://doi.org/10.3389/fgene.2021.695625>

Supplementary tables

Table S 3-1: The demographic and clinical information of the biopsy samples used in the transcriptomics and proteomics analysis [14]. This table is adapted from the metadata that is publicly available in the published transcriptomics article (<https://doi.org/10.1016/j.ncrna.2022.01.002>, Table S2) and proteomics article (<https://doi.org/10.3389/fgene.2021.695625>, Table S2).

SID	Matched to SID	RNA-seq batch	Type	Sex	BMI	Therapy
10188	271	2	Breast cancer	F	29	CYC
686	10201	2	Control	F	25	-
534	10132	2	Control	F	27	-
10201	686	2	Breast cancer	F	26	DOX, CYC, EVR, EXE, TXT
10132	534	2	Breast cancer	F	29	EPI, DOC

10130	562	2	Breast cancer	F	29	DOX, CYC, RTX, VCR, PDN
271	10188	-	Control	F	-	-
442	10204	-	Control	F	-	-
10206	397	2	Chronic lymphatic leukemia	M	23	DOX, CYC, VCR, PDN
10217	407	2	Adenocarcinoma uterus	F	28	DOX, CYC, 5-FU
407	10217	-	Control	F	-	-
10204	442	2	Breast cancer	F	27	CYC, letrozol
397	10206	2	Control	M	23	-
523	423	2	Control	F	26	-
10198	334	2	Breast cancer	F	30	ANT + hormonal therapy
562	10130	2	Control	F	29	-
10021	10027	1	Control	F	29	-
614	10059	1	Breast cancer	F	23	DOX, CYC, 5-FU
10059	614	1	Control	F	23	-
10027	10021	1	Breast cancer	F	30	DOX, EPI, TAM
10033	10096	1	Prostate cancer	M	35	- (local therapy only) (*)
10096	10033	1	Non-Hodgkin lymphoma	M	33	DOX, CYC, RTX, VCR, PDN
10220	245	3	Breast cancer	F	24	TAM
10239	237	3	Breast cancer	F	25	DOX, CYC, DOC, TAM
10244	729	3	Hodgkin lymphoma	M	25	extreme radiotherapy (mantelveld)
10247	346	3	Breast cancer	F	26	DOX, CYC, DOC, TAM
10252	150	3	Hodgkin lymphoma	M	27	DOX, bleomycine, vinblastin, dacarbazine
10284	396	3	Breast cancer	F	30	DOX, CYC, 5-FU, TAM
10285	679	3	Breast cancer	F	24	DOX, CYC, DOC
10315	488	3	Lymphogenic spread melanoma	F	27	Nivolumab (monoclonal antibody)
237	10239	3	Control	F	24	-
423	523	3	Breast cancer	F	25	DOX, EPI
488	10315	3	Control	F	36	-
617	10312	3	Control	F	38	-

Notes: SID, sample ID; Control, heart failure patients who do not have cancer history; F, female; M, male; 5-FU, 5-fluorouracil; ANT, anthracycline; CYC, cyclophosphamide; DOC,

docetaxel; DOX, doxorubicin; EPI, epirubicin; EVR, everolimus; EXE, exemestane; PDN, prednisone; RTX, rituximab; TAM, tamoxifen; TXT, taxotene; VCR, vincristine; -, not applicable or no information; (*) this patient was labeled as control sample in the metadata of the HeCaToS project.

Table S 3-2: The pathway analysis outcome of the differentially expressed (DE) genes in the *in vitro* samples. Please check Table S3 in the supplementary materials that are publicly available in the published article <https://doi.org/10.1016/j.ncrna.2022.01.002>

Table S 3-3: The pathway analysis outcome of the differentially expressed (DE) genes in the biopsy samples. Please check Table S4 in the supplementary materials that are publicly available in the published transcriptomics article <https://doi.org/10.1016/j.ncrna.2022.01.002>

Table S 3-4: The differentially expressed (DE) lncRNAs related to heart disease according to the LncTarD database [41].

Disease	lncRNA	Target	Regulation direction	Expression Pattern	Influenced Function	Regulatory Mechanism
Acute myocardial infarction	H19	SOX8	positive	Down-regulation	cell injury (-); PI3K/AKT/mTOR signaling pathway (+)	ceRNA (miR-139)
Aortic valve disease	TUG1	RUNX2	positive	Up-regulation	cell differentiation (+)	ceRNA (miR-204-5p)
Aortic valve disease	MALAT1	SMAD4	positive	Up-regulation	cell differentiation (+)	ceRNA (miR-204)
Aortic valve stenosis	H19	NOTCH1	negative	Up-regulation	biomineral tissue development(+)	transcriptional regulation
Cardiac fibrosis	GAS5	PTEN	positive	Down-regulation	cell proliferation (-); cell growth (-); fibrotic (-)	ceRNA (miR-21)
Cardiac hypertrophy	CASC15	TLR4	positive	Up-regulation	cardiac hypertrophy (+)	ceRNA (miR-432-5p)
Heart failure	FTX	BCL2L2	negative	Down-regulation	apoptosis process (+)	ceRNA (miR-29b-1-5p)
Myocardial infarction	XIST	PDE4D	positive	Up-regulation	apoptosis process (+); cell proliferation (-)	ceRNA (miR-130a-3p)
Myocardial infarction	H19	FADD	positive	Up-regulation	myocardial necrosis (-)	ceRNA (miR-103a-3p; miR-107)

Table S 3-5: The differentially expressed (DE) lncRNAs related to heart disease according to the LncRNADisease_v2.0 database validated by experimental methods [40].

LncRNA	Disease	Evidence
<i>H19</i>	Heart disease	qRT-PCR//Western blot
	Cardiac fibroblast proliferation and fibrosis	IHC//MTT assay//qRT-PCR//Western blot
	Diabetic cardiomyopathies	Luciferase reporter gene assay//qPCR//Western blot
	Cardiomyocyte hypertrophy	qPCR
<i>BACE1-AS</i>	Heart failure	IHC//RNA ISH
<i>FTX</i>	Cardiomyocyte apoptosis	Immunoblot//Pull-down assay//qRT-PCR
<i>MEG3</i>	Gastric cardia adenocarcinoma	qPCR//qRT-PCR

Table S 3-6: The list of the differentially expressed (DE) lncRNA in each treatment condition in the *in vitro* samples. Please check Table S7 in the supplementary materials that are publicly available in the published article <https://doi.org/10.1016/j.jncrna.2022.01.002>

Table S 3-7: The top 10 over-represented pathways of selected modules in the *in vitro* data (p-value<0.01, and q-value <0.01). Please check Table S3 in the supplementary materials that are available in the published article. <https://doi.org/10.3389/fgene.2021.695625>

Table S 3-8: The top 10 over-represented pathways of selected modules in the human biopsies proteomics data (p-value<0.01, and q-value <0.01). Please check Table S4 in the supplementary materials that are publicly available in the published transcriptomics article. <https://doi.org/10.3389/fgene.2021.695625>

References

- [1] W.H.O. WHO, World Health Organization model list of essential medicines for children: 7th list 2019, in: World Health Organization, 2019.
- [2] W.H.O. WHO, World Health Organization model list of essential medicines: 21st list 2019, in: World Health Organization, 2019.
- [3] G. Minotti, P. Menna, E. Salvatorelli, G. Cairo, L. Gianni, Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity, *Pharmacological reviews* 56 (2004) 185-229. <https://doi.org/10.1124/pr.56.2.6>
- [4] M.J. Kuffel, J.M. Reid, M.M. Ames, Anthracyclines and their C-13 alcohol metabolites: growth inhibition and DNA damage following incubation with human tumor cells in culture, *Cancer Chemotherapy and Pharmacology* 30 (1992) 51-57. <https://doi.org/10.1007/BF00686485>
- [5] D.A. Mulrooney, M.W. Yeazel, T. Kawashima, A.C. Mertens, P. Mitby, M. Stovall, S.S. Donaldson, D.M. Green, C.A. Sklar, L.L. Robison, W.M. Leisenring, Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: retrospective analysis of the Childhood Cancer Survivor Study cohort, *BMJ* 339 (2009) b4606. <https://doi.org/10.1136/bmj.b4606>
- [6] J.E. Bates, R.M. Howell, Q. Liu, Y. Yasui, D.A. Mulrooney, S. Dhakal, S.A. Smith, W.M. Leisenring, D.J. Indelicato, T.M. Gibson, G.T. Armstrong, K.C. Oeffinger, L.S. Constine, Therapy-Related Cardiac Risk in Childhood Cancer Survivors: An Analysis of the Childhood Cancer Survivor Study, *Journal of Clinical Oncology* 37 (2019) 1090-1101. <https://doi.org/10.1200/jco.18.01764>
- [7] M. Ryberg, D. Nielsen, T. Skovsgaard, J. Hansen, B.V. Jensen, P. Dombernowsky, Epirubicin cardiotoxicity: an analysis of 469 patients with metastatic breast cancer, *Journal of Clinical Oncology* 16 (1998) 3502-3508. <https://doi.org/10.1200/jco.1998.16.11.3502>
- [8] Electronic Medicines Compendium (EMC), Zavedos 5mg Capsules - Summary of Product Characteristics, in: Datapharm Ltd., Surrey, UK.
- [9] P. Anderlini, R.S. Benjamin, F.C. Wong, H.M. Kantarjian, M. Andreeff, S.M. Kornblau, S. O'Brien, B. Mackay, M.S. Ewer, S.A. Pierce, et al., Idarubicin cardiotoxicity: a retrospective study in acute myeloid leukemia and myelodysplasia, *Journal of Clinical Oncology* 13 (1995) 2827-34. <https://doi.org/10.1200/jco.1995.13.11.2827>
- [10] T. Šimůnek, M. Štěřba, O. Popelová, M. Adamcová, R. Hrdina, V. Geršl, Anthracycline-induced cardiotoxicity: Overview of studies examining the roles of oxidative stress and free cellular iron, *Pharmacological Reports* 61 (2009) 154-171. [https://doi.org/10.1016/S1734-1140\(09\)70018-0](https://doi.org/10.1016/S1734-1140(09)70018-0)
- [11] C.A. Geisberg, D.B. Sawyer, Mechanisms of Anthracycline cardiotoxicity and strategies to decrease cardiac damage, *Current Hypertension Reports* 12 (2010) 404-410. <https://doi.org/10.1007/s11906-010-0146-y>
- [12] A. Ghigo, M. Li, E. Hirsch, New signal transduction paradigms in anthracycline-induced cardiotoxicity, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1863 (2016) 1916-1925. <https://doi.org/10.1016/j.bbamcr.2016.01.021>
- [13] L. Kuepfer, O. Clayton, C. Thiel, H. Cordes, R. Nudischer, L.M. Blank, V. Baier, S. Heymans, F. Caiment, A. Roth, D.A. Fluri, J.M. Kelm, J. Castell, N. Selevsek, R. Schlapbach,

H. Keun, J. Hynes, U. Sarkans, H. Gmuender, R. Herwig, S. Niederer, J. Schuchhardt, M. Segall, J. Kleinjans, A model-based assay design to reproduce in vivo patterns of acute drug-induced toxicity, *Archives of Toxicology* 92 (2018) 553-555. <https://doi.org/10.1007/s00204-017-2041-7>

[14] M. Verheijen, Transcriptomics close to my heart: advanced models & methods for toxicogenomics research illustrated by anthracycline-induced cardiotoxicity. Chapter 4., in: *Toxicogenomic*, GROW school, Maastricht University, ProefschriftMaken Maastricht, 2019, pp. 79-109.

[15] M. Verheijen, Y. Schrooders, H. Gmuender, R. Nudischer, O. Clayton, J. Hynes, S. Niederer, H. Cordes, L. Kuepfer, J. Kleinjans, F. Caiment, Bringing in vitro analysis closer to in vivo: Studying doxorubicin toxicity and associated mechanisms in 3D human microtissues with PBPK-based dose modelling, *Toxicology Letters* 294 (2018) 184-192. <https://doi.org/10.1016/j.toxlet.2018.05.029>

[16] J.T.Y. Kung, D. Colognori, J.T. Lee, Long noncoding RNAs: past, present, and future, *Genetics* 193 (2013) 651-669. <https://doi.org/10.1534/genetics.112.146704>

[17] R.-W. Yao, Y. Wang, L.-L. Chen, Cellular functions of long noncoding RNAs, *Nature Cell Biology* 21 (2019) 542-551. <https://doi.org/10.1038/s41556-019-0311-8>

[18] S. Sweta, T. Dudnakova, S. Sudheer, A.H. Baker, R. Bhushan, Importance of Long Non-coding RNAs in the Development and Disease of Skeletal Muscle and Cardiovascular Lineages, *Frontiers in cell and developmental biology* 7 (2019) 228. <https://doi.org/10.3389/fcell.2019.00228>

[19] L. Hobuß, C. Bär, T. Thum, Long non-coding RNAs: At the heart of cardiac dysfunction?, *Frontiers in physiology* 10 (2019). <https://doi.org/10.3389/fphys.2019.00030>

[20] X. Yuan, J. Wang, X. Tang, Y. Li, P. Xia, X. Gao, Berberine ameliorates nonalcoholic fatty liver disease by a global modulation of hepatic mRNA and lncRNA expression profiles, *Journal of Translational Medicine* 13 (2015) 24. <https://doi.org/10.1186/s12967-015-0383-6>

[21] X.-Y. Jiang, Q.-L. Ning, Expression profiling of long noncoding RNAs and the dynamic changes of lncRNA-NR024118 and Cdkn1c in angiotensin II-treated cardiac fibroblasts, *International journal of clinical and experimental pathology* 7 (2014) 1325-1336.

[22] D.-X. He, G.-Y. Zhang, X.-T. Gu, A.-Q. Mao, C.-X. Lu, J. Jin, D.-Q. Liu, X. Ma, Genome-wide profiling of long non-coding RNA expression patterns in anthracycline-resistant breast cancer cells, *International journal of oncology* 49 (2016) 1695-1703. <https://doi.org/10.3892/ijo.2016.3665>

[23] Y. Zeng, G. Wang, C.-F. Zhou, H.-B. Zhang, H. Sun, W. Zhang, H.-H. Zhou, R. Liu, Y.-S. Zhu, LncRNA profile study reveals a three-LncRNA signature associated with the pathological complete response following neoadjuvant chemotherapy in breast cancer, *Frontiers in pharmacology* 10 (2019). <https://doi.org/10.3389/fphar.2019.00574>

[24] Q.-N. Zhu, G. Wang, Y. Guo, Y. Peng, R. Zhang, J.-L. Deng, Z.-X. Li, Y.-S. Zhu, LncRNA H19 is a major mediator of doxorubicin chemoresistance in breast cancer cells through a cullin4A-MDR1 pathway, *Oncotarget* 8 (2017) 91990-92003. <https://doi.org/10.18632/oncotarget.21121>

[25] S. Zhang, Y. Yuan, Z. Zhang, J. Guo, J. Li, K. Zhao, Y. Qin, C. Qiu, LncRNA FOXC2-AS1 protects cardiomyocytes from doxorubicin-induced cardiotoxicity through activation of

WNT1-inducible signaling pathway protein-1, *Bioscience, Biotechnology, and Biochemistry* 83 (2019) 653-658. <https://doi.org/10.1080/09168451.2018.1553606>

[26] L. Chen, K.-P. Yan, X.-C. Liu, W. Wang, C. Li, M. Li, C.-G. Qiu, Valsartan regulates TGF- β /Smads and TGF- β /p38 pathways through lncRNA CHRf to improve doxorubicin-induced heart failure, *Archives of Pharmacal Research* 41 (2018) 101-109. <https://doi.org/10.1007/s12272-017-0980-4>

[27] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (2014) 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>

[28] S. Andrews, FastQC: a quality control tool for high throughput sequence data, in: Babraham Institute, Babraham, UK, 2010.

[29] P. Ewels, M. Magnusson, S. Lundin, M. Källér, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics* 32 (2016) 3047-3048. <https://doi.org/10.1093/bioinformatics/btw354>

[30] D.R. Zerbino, P. Achuthan, W. Akanni, M R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C.G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O.G. Izuogu, S.H. Janacek, T. Juettemann, J.K. To, M.R. Laird, I. Lavidas, Z. Liu, J.E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D.N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C.K. Ong, A. Parker, M. Patricio, H.S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S.E. Hunt, M. Kostadima, N. Langridge, F.J. Martin, M. Muffato, E. Perry, M. Ruffier, D.M. Staines, S.J. Trevanion, B.L. Aken, F. Cunningham, A. Yates, P. Flicek, *Ensembl 2018*, *Nucleic Acids Research* 46 (2017) D754-D761. <https://doi.org/10.1093/nar/gkx1098>

[31] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinformatics* 12 (2011) 323. <https://doi.org/10.1186/1471-2105-12-323>

[32] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nature methods* 9 (2012) 357-359. <https://doi.org/10.1038/nmeth.1923>

[33] R.C. Team, R: A language and environment for statistical computing, in: Vienna, Austria, 2013.

[34] J.R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties, *Bioinformatics* 33 (2017) 2938-2940. <https://doi.org/10.1093/bioinformatics/btx364>

[35] H. Wickham, M. Averick, J. Bryan, W. Chang, L.D.A. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, Welcome to the Tidyverse, *Journal of Open Source Software* 4 (2019) 1686. <https://doi.org/10.21105/joss.01686>

[36] H. Wickham, M.H. Wickham, The ggplot package, in: 2007.

[37] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biology* 15 (2014) 550. <https://doi.org/10.1186/s13059-014-0550-8>

[38] D.S. Fischer, F.J. Theis, N. Yosef, Impulse model-based differential expression analysis of time course sequencing data, *Nucleic Acids Research* 46 (2018) e119-e119. <https://doi.org/10.1093/nar/gky675>

- [39] A. Kamburov, K. Pentchev, H. Galicka, C. Wierling, H. Lehrach, R. Herwig, ConsensusPathDB: toward a more complete picture of cell biology, *Nucleic acids research* 39 (2011) D712-D717. <https://doi.org/10.1093/nar/gkqu56>
- [40] Z. Bao, Z. Yang, Z. Huang, Y. Zhou, Q. Cui, D. Dong, LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases, *Nucleic Acids Res* 47 (2019) D1034-d1037. <https://doi.org/10.1093/nar/gky905>
- [41] H. Zhao, J. Shi, Y. Zhang, A. Xie, L. Yu, C. Zhang, J. Lei, H. Xu, Z. Leng, T. Li, W. Huang, S. Lin, L. Wang, Y. Xiao, X. Li, LncTarD: a manually-curated database of experimentally-supported functional lncRNA-target regulations in human diseases, *Nucleic Acids Research* 48 (2019) D118-D126. <https://doi.org/10.1093/nar/gkz985>
- [42] J. Du, Z. Yuan, Z. Ma, J. Song, X. Xie, Y. Chen, KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model, *Molecular BioSystems* 10 (2014) 2441-2447. <http://dx.doi.org/10.1039/C4MB00287C>
- [43] L. Li, Y. Cong, X. Gao, Y. Wang, P. Lin, Differential expression profiles of long non-coding RNAs as potential biomarkers for the early diagnosis of acute myocardial infarction, *Oncotarget* 8 (2017) 88613-88621. <https://doi.org/10.18632/oncotarget.20101>
- [44] B. Cai, W. Ma, C. Bi, F. Yang, L. Zhang, Z. Han, Q. Huang, F. Ding, Y. Li, G. Yan, Z. Pan, B. Yang, Y. Lu, Long noncoding RNA H19 mediates melatonin inhibition of premature senescence of c-kit⁺ cardiac progenitor cells by promoting miR-675, *J. Pineal Res* 61 (2016) 82-95. <https://doi.org/10.1111/jpi.12331>
- [45] B.-f. Zhang, J. Chen, H. Jiang, LncRNA H19 ameliorates myocardial ischemia-reperfusion injury by targeting miR-22-3P, *International Journal of Cardiology* 278 (2019) 224. <https://doi.org/10.1016/j.ijcard.2018.11.017>
- [46] S. Greco, G. Zaccagnini, A. Perfetti, P. Fuschi, R. Valaperta, C. Voellenkle, S. Castelvécchio, C. Gaetano, N. Finato, A.P. Beltrami, L. Menicanti, F. Martelli, Long noncoding RNA dysregulation in ischemic heart failure, *Journal of Translational Medicine* 14 (2016) 183. <https://doi.org/10.1186/s12967-016-0926-5>
- [47] R. Zhao, X. Wang, H. Wang, T. Yu, Q. Wang, X. Yang, J. Sun, Inhibition of long noncoding RNA BDNF-AS rescues cell death and apoptosis in hypoxia/reoxygenation damaged murine cardiomyocyte, *Biochimie* 138 (2017) 43-49. <https://doi.org/10.1016/j.biochi.2017.03.018>
- [48] L. Jing, S. Li, J. Wang, G. Zhang, Long non-coding RNA small nucleolar RNA host gene 7 facilitates cardiac hypertrophy via stabilization of SDA1 domain containing 1 mRNA, *J Cell Biochem* 120 (2019) 15089-15097. <https://doi.org/10.1002/jcb.28770>
- [49] D. Matyszevska, E. Nazaruk, R.A. Campbell, Interactions of anticancer drugs doxorubicin and idarubicin with lipid monolayers: New insight into the composition, structure and morphology, *Journal of Colloid and Interface Science* 581 (2021) 403-416. <https://doi.org/10.1016/j.jcis.2020.07.092>
- [50] L. Han, Z. Li, Y. Jiang, Z. Jiang, L. Tang, SNHG29 regulates miR-223-3p/CTNND1 axis to promote glioblastoma progression via Wnt/ β -catenin signaling pathway, *Cancer Cell International* 19 (2019) 345. <https://doi.org/10.1186/s12935-019-1057-x>
- [51] S. Xu, J. Guo, W. Zhang, LncRNA PCAT19 promotes the proliferation of laryngocarcinoma cells via modulation of the miR-182/PDK4 axis, *J Cell Biochem* 120 (2019) 12810-12821. <https://doi.org/10.1002/jcb.28552>

- [52] X. Zhang, Q. Wang, Y. Xu, B. Wang, C. Jia, L. Wang, H. Sun, H. Zhao, Z. Wang, Q. Zou, S. Sun, L. Zhang, lncRNA PCAT19 negatively regulates p53 in non-small cell lung cancer, *Oncology letters* 18 (2019) 6795-6800. <https://doi.org/10.3892/ol.2019.11041>
- [53] S. Greco, G. Zaccagnini, P. Fuschi, C. Voellenkle, M. Carrara, I. Sadeghi, C. Bearzi, B. Maimone, S. Castelvechchio, K. Stellos, C. Gaetano, L. Menicanti, F. Martelli, Increased BACE1-AS long noncoding RNA and β -amyloid levels in heart failure, *Cardiovascular Research* 113 (2017) 453-463. <https://doi.org/10.1093/cvr/cvx013>
- [54] B. Long, N. Li, X.-X. Xu, X.-X. Li, X.-J. Xu, D. Guo, D. Zhang, Z.-H. Wu, S.-Y. Zhang, Long noncoding RNA FTX regulates cardiomyocyte apoptosis by targeting miR-29b-1-5p and Bcl2l2, *Biochemical and Biophysical Research Communications* 495 (2018) 312-318. <https://doi.org/10.1016/j.bbrc.2017.11.030>
- [55] X. Yang, L. Tao, J. Zhu, S. Zhang, Long noncoding RNA FTX reduces hypertrophy of neonatal mouse cardiac myocytes and regulates the PTEN/PI3K/Akt signaling pathway by sponging microRNA-22, *Medical science monitor: international medical journal of experimental and clinical research* 25 (2019) 9609-9617. <https://doi.org/10.12659/MSM.919654>
- [56] G. Borchert, V. King, C. Francis, R. Langley, G. Daly, V. Pastukh, M. Gillespie, Novel Hypoxia Regulated Long Noncoding RNAs, in: A71. THE EPIGENOME, GENOME AND NON-CODING RNAs IN LUNG DISEASE, American Thoracic Society, 2018, pp. A2289-A2289.
- [57] L. Xiao, Y. Gu, Y. Sun, J. Chen, X. Wang, Y. Zhang, L. Gao, L. Li, The long noncoding RNA XIST regulates cardiac hypertrophy by targeting miR-101, *J Cell Physiol.* 234 (2019) 13680-13692. <https://doi.org/10.1002/jcp.28047>
- [58] L. Lai, Y. Xu, L. Kang, J. Yang, G. Zhu, LncRNA KCNQ1OT1 contributes to cardiomyocyte apoptosis by targeting FUS in heart failure, *Experimental and Molecular Pathology* 115 (2020) 104480. <https://doi.org/10.1016/j.yexmp.2020.104480>
- [59] D. Wang, B. Lin, W. Zhang, X. Wang, Up-regulation of SNHG16 induced by CTCF accelerates cardiac hypertrophy by targeting miR-182-5p/IGF1 axis, *Cell Biol Int* 44 (2020) 1426-1435. <https://doi.org/10.1002/cbin.11333>
- [60] G. von Minckwitz, Docetaxel/anthracycline combinations for breast cancer treatment, *Expert Opinion on Pharmacotherapy* 8 (2007) 485-495. <https://doi.org/10.1517/14656566.8.4.485>
- [61] M.J. Keating, T.L. Smith, K.B. McCredie, G.P. Bodey, E.M. Hersh, J.U. Gutterman, E. Gehan, E.J. Freireich, A four-year experience with anthracycline, cytosine arabinoside, vincristine and prednisone combination chemotherapy in 325 adults with acute leukemia, *Cancer* 47 (1981) 2779-2788. [https://doi.org/10.1002/1097-0142\(19810615\)47:12<2779::Aid-cnrcr2820471204>3.o.Co;2-o](https://doi.org/10.1002/1097-0142(19810615)47:12<2779::Aid-cnrcr2820471204>3.o.Co;2-o)
- [62] L. Swift, J. McHowat, N.J.C.r. Sarvazyan, Inhibition of membrane-associated calcium-independent phospholipase A2 as a potential culprit of anthracycline cardiotoxicity, *Cancer research* 63 (2003) 5992-5998.
- [63] D. Cappetta, G. Esposito, E. Piegari, R. Russo, L.P. Ciuffreda, A. Rivellino, L. Berrino, F. Rossi, A. De Angelis, K. Urbanek, SIRT1 activation attenuates diastolic dysfunction by reducing cardiac fibrosis in a model of anthracycline cardiomyopathy, *International Journal of Cardiology* 205 (2016) 99-110. <https://doi.org/10.1016/j.ijcard.2015.12.008>

- [64] G.C. Pereira, S.P. Pereira, F.B. Pereira, N. Lourenço, J.A. Lumini, C.V. Pereira, J.A. Bjork, J. Magalhães, A. Ascensão, M.R. Wieckowski, A.J. Moreno, K.B. Wallace, P.J. Oliveira, Early cardiac mitochondrial molecular and functional responses to acute Anthracycline treatment in Wistar rats, *Toxicological Sciences* 169 (2019) 137-150. <https://doi.org/10.1093/toxsci/kfz026>
- [65] N. Selevsek, M. Matondo, M.S. Carbayo, R. Aebersold, B. Domon, Systematic quantification of peptides/proteins in urine using selected reaction monitoring, *Proteomics* 11 (2011) 1135-1147. <https://doi.org/10.1002/pmic.201000599>
- [66] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nature Methods* 6 (2009) 359-362. <https://doi.org/10.1038/nmeth.1322>
- [67] M. Verheijen, M. Lienhard, Y. Schrooders, O. Clayton, R. Nudischer, S. Boerno, B. Timmermann, N. Selevsek, R. Schlapbach, H. Gmuender, S. Gotta, J. Geraedts, R. Herwig, J. Kleinjans, F. Caiment, DMSO induces drastic changes in human cellular processes and epigenetic landscape in vitro, *Scientific Reports* 9 (2019) 4641. <https://doi.org/10.1038/s41598-019-40660-0>
- [68] T. Guo, P. Kouvonen, C.C. Koh, L.C. Gillet, W.E. Wolski, H.L. Röst, G. Rosenberger, B.C. Collins, L.C. Blum, S. Gillissen, M. Joerger, W. Jochum, R. Aebersold, Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps, *Nature medicine* 21 (2015) 407-413. <https://doi.org/10.1038/nm.3807>
- [69] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinformatics* 9 (2008) 559. <https://doi.org/10.1186/1471-2105-9-559>
- [70] J. Piñero, J.M. Ramírez-Anguita, J. Saüch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, L.I. Furlong, The DisGeNET knowledge platform for disease genomics: 2019 update, *Nucleic Acids Research* 48 (2019) D845-D855. <https://doi.org/10.1093/nar/gkz1021>
- [71] A. Martin, M.E. Ochagavia, L.C. Rabasa, J. Miranda, J. Fernandez-de-Cossio, R. Bringas, BisoGenet: a new tool for gene network building, visualization and analysis, *BMC Bioinformatics* 11 (2010) 91. <https://doi.org/10.1186/1471-2105-11-91>
- [72] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T.J.G.r. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome research* 13 (2003) 2498-2504. <https://doi.org/10.1101/gr.1239303>
- [73] N. Narain, Rajin, R. Sarangarajan, V. Vishnudas, K., M. Kiebish, Andrew, Use of markers in the identification of cardiotoxic agents, in: World Intellectual Property Organization (WIPO), 2014.
- [74] H.R. Ali, C.R. Michel, Y.H. Lin, T.A. McKinsey, M.Y. Jeong, A.V. Ambardekar, J.C. Cleveland, R. Reisdorph, N. Reisdorph, K.C. Woulfe, K.S. Fritz, Defining decreased protein succinylation of failing human cardiac myofibrils in ischemic cardiomyopathy, *Journal of Molecular and Cellular Cardiology* 138 (2020) 304-317. <https://doi.org/10.1016/j.yjmcc.2019.11.159>
- [75] S. Ruiz-Pinto, G. Pita, M. Martín, T. Alonso-Gordoa, D.R. Barnes, M.R. Alonso, B. Herraiez, P. García-Miguel, J. Alonso, A. Pérez-Martínez, A.J. Cartón, F. Gutiérrez-Larraya, J.A. García-Sáenz, J. Benítez, D.F. Easton, A. Patiño-García, A. González-Neira, Exome array analysis identifies ETFB as a novel susceptibility gene for anthracycline-induced

cardiotoxicity in cancer patients, *Breast Cancer Research and Treatment* 167 (2018) 249-256. <https://doi.org/10.1007/s10549-017-4497-9>

[76] The UniProt Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Research* 47 (2018) D506-D515. <https://doi.org/10.1093/nar/gky1049>

[77] D. Matyszewska, The influence of charge and lipophilicity of daunorubicin and idarubicin on their penetration of model biological membranes – Langmuir monolayer and electrochemical studies, *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1862 (2020) 183104. <https://doi.org/10.1016/j.bbamem.2019.183104>

[78] K. Matyjaszczyk-Gwarda, T. Wójcik, M. Łukawska, S. Chlopicki, M. Walczak, Lipophilicity profiling of anthracycline antibiotics by microemulsion electrokinetic chromatography—effects on cardiotoxicity and endotheliotoxicity, *ELECTROPHORESIS* 40 (2019) 3108-3116. <https://doi.org/10.1002/elps.201900259>

[79] J.W. Singer, K.A. Narahara, J.L. Ritchie, G.W. Hamilton, J.W. Kennedy, Time- and dose-dependent changes in ejection fraction determined by radionuclide angiography after anthracycline therapy, *Cancer treatment reports* 62 (1978) 945-948.

[80] J.L. Ritchie, J.W. Singer, D. Thorning, S.G. Sorensen, G.W.J.C. Hamilton, Anthracycline cardiotoxicity: clinical and pathologic outcomes assessed by radionuclide ejection fraction, *Cancer* 46 (1980) 1109-1116. [https://doi.org/10.1002/1097-0142\(19800901\)46:5<1109::aid-cnrcr2820460506>3.o.co;2-b](https://doi.org/10.1002/1097-0142(19800901)46:5<1109::aid-cnrcr2820460506>3.o.co;2-b)

[81] A. Murabito, E. Hirsch, A. Ghigo, Mechanisms of Anthracycline-Induced Cardiotoxicity: Is Mitochondrial Dysfunction the Answer?, *Frontiers in cardiovascular medicine* 7 (2020) 35-35. <https://doi.org/10.3389/fcvm.2020.00035>

[82] S.M. Majid, A.S. Liss, M. You, H.R. Bose, The suppression of SH3BGRL is important for v-Rel-mediated transformation, *Oncogene* 25 (2006) 756-768. <https://doi.org/10.1038/sj.onc.1209107>

[83] M.M. Kittleson, K.M. Minhas, R.A. Irizarry, S.Q. Ye, G. Edness, E. Breton, J.V. Conte, G. Tomaselli, J.G.N. Garcia, J.M. Hare, Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure, *Physiological Genomics* 21 (2005) 299-307. <https://doi.org/10.1152/physiolgenomics.00255.2004>

[84] J. Schilling, L. Lai, N. Sambandam, C.E. Dey, T.C. Leone, D.P. Kelly, Toll-Like receptor-mediated inflammatory signaling reprograms cardiac energy metabolism by repressing peroxisome proliferator-activated receptor γ coactivator-1 signaling, *Circulation : Heart Failure* 4 (2011) 474-482. <https://doi.org/10.1161/CIRCHEARTFAILURE.110.959833>

[85] D.A. Bota, K.J.A. Davies, Protein degradation in mitochondria: implications for oxidative stress, aging and disease:: a novel etiological classification of mitochondrial proteolytic disorders, *Mitochondrion* 1 (2001) 33-49. [https://doi.org/10.1016/S1567-7249\(01\)00005-8](https://doi.org/10.1016/S1567-7249(01)00005-8)

[86] R.A. Gottlieb, A. Thomas, Mitophagy and Mitochondrial Quality Control Mechanisms in the Heart, *Current pathobiology reports* 5 (2017) 161-169. <https://doi.org/10.1007/s40139-017-0133-y>

[87] Nathan W. Pierce, J.E. Lee, X. Liu, Michael J. Sweredoski, Robert L.J. Graham, Elizabeth A. Larimore, M. Rome, N. Zheng, Bruce E. Clurman, S. Hess, S.-o. Shan,

Raymond J. Deshaies, Candi Promotes Assembly of New SCF Complexes through Dynamic Exchange of F Box Proteins, *Cell* 153 (2013) 206-215. <https://doi.org/10.1016/j.cell.2013.02.024>

[88] M.S. Willis, A. Bevilacqua, T. Pulinilkunnil, P. Kienesberger, M. Tannu, C. Patterson, The role of ubiquitin ligases in cardiac disease, *Journal of molecular and cellular cardiology* 71 (2014) 43-53. <https://doi.org/10.1016/j.yjmcc.2013.11.008>

[89] J. Wang, J. Lee, D. Liem, P. Ping, HSPA5 Gene encoding Hsp70 chaperone BiP in the endoplasmic reticulum, *Gene* 618 (2017) 14-23. <https://doi.org/10.1016/j.gene.2017.03.005>

[90] Y. Wang, J. Liu, Q. Kong, H. Cheng, F. Tu, P. Yu, Y. Liu, X. Zhang, C. Li, Y. Li, X. Min, S. Du, Z. Ding, L. Liu, Cardiomyocyte-specific deficiency of HSPB1 worsens cardiac dysfunction by activating NFκB-mediated leucocyte recruitment after myocardial infarction, *Cardiovascular Research* 115 (2018) 154-167. <https://doi.org/10.1093/cvr/cvy163>

[91] L.M. Judge, J.A. Perez-Bermejo, A. Truong, A.J. Ribeiro, J.C. Yoo, C.L. Jensen, M.A. Mandegar, N. Huebsch, R.M. Kaake, P.-L. So, D. Srivastava, B.L. Pruitt, N.J. Krogan, B.R. Conklin, A BAG3 chaperone complex maintains cardiomyocyte function during proteotoxic stress, *JCI insight* 2 (2017) e94623. <https://doi.org/10.1172/jci.insight.94623>

[92] A.M. Feldman, D.G. Tilley, W. Zhu, K. Khalili, W.J. Koch, Bag3 as a target for therapy of heart failure, in: <http://www.freepatentsonline.com/y2017/0016066.html>, TEMPLE UNIVERSITY OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION (Philadelphia, PA, US), United States, 2017.

[93] P. Menna, S. Recalcati, G. Cairo, G. Minotti, An introduction to the metabolic determinants of anthracycline cardiotoxicity, *Cardiovascular Toxicology* 7 (2007) 80-85. <https://doi.org/10.1007/s12012-007-0011-7>

[94] E. Gammella, F. Maccarinelli, P. Buratti, S. Recalcati, G. Cairo, The role of iron in anthracycline cardiotoxicity, *Front. Pharmacol.* 5 (2014). <https://doi.org/10.3389/fphar.2014.00025>

[95] S.A. Padia, Is Idarubicin the future of TACE?, *Radiology* 291 (2019) 809-810. <https://doi.org/10.1148/radiol.2019190789>

Chapter 4 | Transcription factor and target relations in drug-induced hepatotoxicity

Introduction

Focusing on regulatory relationships such as transcriptional factor (TF) – targets is a fundamental biological research topic. However, with the advancement of sequencing technologies and their increased efficiency, omics technologies can provide high-dimensional omics data and result in a long list of differentially expressed (DE) genes and proteins. The long list of genes/proteins burdens researchers with the information gathering issue, because it is difficult to extract and appraise all possible interactions between these genes/proteins. Especially, the increase of different databases also provides a wide range of TF-target relations. For instance, the hTFtarget - a comprehensive database for regulations of human transcription factors and their targets (2020) has a total of 1,319,123 TF-target relations [1]. The combination of the long list of candidate genes/proteins with a wide range of databases leads to difficulty to extract and appraise all possible interactions between these genes/proteins. It requires a dedicated tool to extract and filter the relative regulations to suggest strong candidate TFs for further investigation.

In this chapter 4, we first developed the RegOmics (Regulatory extraction for Omics data) tool to aid researchers with identifying TF - target relations and suggesting potential TF candidates in user-uploaded data. Thereafter, we analyzed the epigenomics, transcriptomics, and proteomics derived from hepatic microtissues exposed to Rifampicin (RIF). We also applied the RegOmics tool to obtain and refine the TF-target relations in these omics datasets.

Chapter 4.1: RegOmics tool – Regulatory omics information

Objectives of the study

Several tools are available for detecting TF- target regulations, such as some TF prediction tools which exploit DEGs: ChEA3 [2] and LISA [3]. However, they do not include TF-target regulation information for lncRNAs. Some lncRNA databases have provided lncRNAs information related to disease [4] or even the lncRNA-target relations in humans [5]. Although these lncRNA-target databases are mainly straightforward to use, most of them can only provide information for each lncRNA but not a group of input genes. Furthermore, few databases or tools are not working anymore, such as LncReg, a reference resource for lncRNA-associated regulatory networks published in 2015 [6], or TF2LncRNA which was launched in 2014 to identify common transcription factors for a list of lncRNA genes from CHIP-Seq data [7].

Furthermore, the modern high throughput technologies have supplied high-dimensional omics data with extensive gene and protein detections. Even after all the differential expression analysis and other data filtering, there remains a long list of interesting genes/proteins. These cause difficulties in information gathering to understand the inter-dependencies between these genes/proteins. Thus, it is a necessity for a tool/database that is capable of inquiring the TF-target relations of not only protein-coding genes but also lncRNA genes from omics data input.

In this chapter 4.1, we developed RegOmics (Regulatory extraction for Omics data) as a user-friendly tool for identifying TF - target relations and suggesting potential TF candidates in user-uploaded data. This tool is able to operate with the long list of genes/proteins derived from omic data. RegOmics can identify the TF-target relations that existed in the user input data, can rank these TF-target relations, and could suggest candidate TFs to researchers. The tool is available as a web-based version (https://nhannguyen.shinyapps.io/RegOmics_Rshinyapp/) and a source R code on Github (<https://github.com/NhanNguyenooo/RegOmics>).

Tool features

The RegOmics tool consists of two working options: R shiny website and R source code to run on local computers. R shiny website offers an online and user-friendly app interface that allows users to input the gene/protein files, click on the “Run tool” button, and see the outcome. Both the R shiny and R source code follow the

same schematic workflow (Figure 4-1), although the performance of the web-based version is limited by the size of the input files (<5MB).

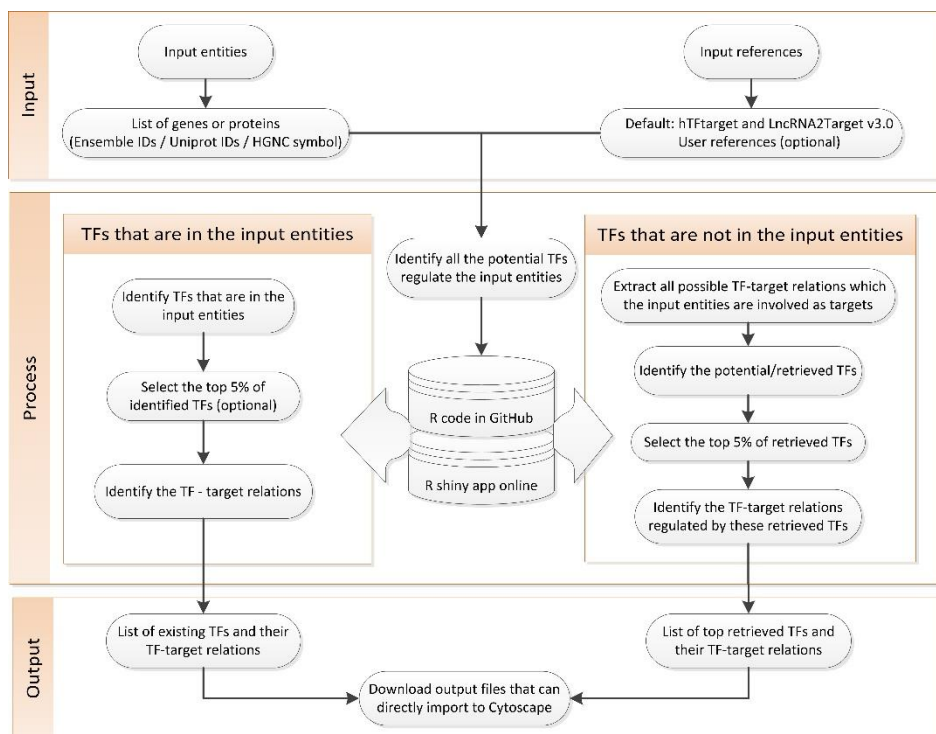


Figure 4-1: The RegOmics tool architecture

For the input files, the genomic annotation uses gene/protein names (HGNC Symbol) or Ensemb ID for genes and Uniport for proteins. We established a default TF-target reference using different databases, comprising hTFtarget database for human TF-target relations [1] and the latest release of the LncRNA2Target (version 3.0), which is a comprehensive database for target genes of lncRNAs in humans and mice [8]. For the lncRNA reference, we only used the TF-target relations acquired from and validated by conventional experiments including RT-qPCR, western blot, luciferase reporter assays, immunoprecipitation assays, and RNA pull-down assays. Users can also add their own TF-target preference by using the RegOmics runs at local computer.

The RegOmics tool first examines the input files and shows related information about their protein/genes, name, and IDs, annotation the IDs to gene/protein names (if needed), and the number of proteins/genes in each file. After that, according to the TF-target reference, the tool inspects TFs that are in the user

upload genes/proteins list. It also provides a downloadable file for the related TF-target relations that existed in the data (Figure 4-2).

RegOmics

Author: Nhan Nguyen, see [GitHub](#). Date: 20 October, 2022

How to use the tools

This webpage tool is suitable for small data (<5MB). Users upload input files and click to run the tool in the **Input and Run** section. The results will appear in on tabs in the **Outcome** section to see the result.

For bigger dataset, it is better to use the code in [GitHub](#) and run in a local computer. Users could check sample examples to used in the RegOmics tool at the folder "example" in [GitHub](#).

Input and Run

Upload files (txt/csv) here

Browse... 4 files

Upload complete

Click me to run the tool

Default regulatory database:

- The database of human transcription factor targets, Zhang, Q. et al. (2020) "hTFtarget: a comprehensive database for regulations of human transcription factors and their targets". Genomics, Proteomics & Bioinformatics, doi:10.1016/j.gpb.2019.09.006.
- The LncRNA2Target v3.0, Jiang, Q. et al. (2021) "LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse". Nucleic acids research, doi:10.1093/nar/gky1051

Outcome

Input samples

Samples used

Across TFs in the data

TF-target network

You uploaded 4 file(s):

samples	input_types	total_entities	pvalue	note
Protein_Rif_The_002.csv	-	708	FALSE	Could not detect the gene/protein identifier
Protein_Rif_The_002.txt	UNIPROT	708	TRUE	NA
Protein_Rif_The_008.csv	UNIPROT	698	TRUE	NA
Protein_Rif_The_008.txt	-	698	FALSE	Could not detect the gene/protein identifier

Reset the tool

Click me to reset the tool

Clear all the uploaded data and reset the R environment before the new run

Figure 4-2: The online RegOmics tool. (A) The default page provides general information about how to use the tool and the used regulatory databases. Users can upload the gene/protein files and click on the run button. The reset button is the end of the webpage, user can reset the tool before the next analysis.



Figure 4-3: The online RegOmics tool (continue). (B) In the outcome section, users can click on different tabs to see the corresponding results, such as: which uploaded files are used, the number of existing and retrieved TFs, their TF-target relations. User can download the results via the download button.

Furthermore, by tracking the regulatory relationship, we also identify the potential TFs that are not named in the uploaded gene/protein list but can regulate gene/protein in the uploaded list. For those potential TFs, we rank them based on the number of their target genes/proteins that are in the upload list. By default, we concentrate on the top 5% of them to only retrieve the potential TFs with the most TF-target connections; however, users can adjust this percentage selection. We also offer downloadable network and node files (.txt) presenting these identified TF-target relations to import in Cytoscape for visualization.

In conclusion, RegOmics enables researchers to easily find potential TFs from a set of genes/proteins of interest. In the R shiny website version, the tool provides a user-friendly tool to identify and select potential TF for further study (Figure 4-2). In addition, users can conveniently browse and retrieve TF-target regulatory relationships for interested TFs. In the R source code, we provide additional steps for graph illustrations that create networks to visualize the connection between selected TFs and targets. Users can also modify the default settings according to what they need.

Chapter 4.2: TF-target relations related to Rifampicin-induced hepatotoxicity from DNA methylation to transcriptome and proteome

Objectives of the study

Rifampicin (RIF) can inhibit bacterial RNA polymerase, without affecting the mammalian enzyme, leading to RNA synthesis suppression and bactericidal action. This antibiotic is administered orally and metabolized in the liver. It can be used to treat different bacterial infections, especially *Mycobacterium tuberculosis* [9]. In particular, the usage of RIF benefits tuberculosis patients by shortening the treatment duration [10]. While RIF is still an essential drug for the tuberculosis treatment, the emergence of RIF resistance causes a threat to control this disease [11]. Several researchers suggested that higher RIF doses could increase the cure rates and reduce treatment duration compared to the current benchmark [10]. However, RIF is a potential hepatotoxic agent [12]; further research about RIF-induced hepatotoxicity is needed to understand its adverse effects and whether high-dose RIF could be well tolerated in the human liver.

Although researchers have investigated the side effect of RIF, most studies concentrated on clinical markers [13]. Some studies attempted to explore the underlying RIF-induced hepatotoxicity mechanisms. For instance, researchers recognized that, via activating the hepatocyte PXR receptor, RIF could influentially induce some metabolic enzyme pathways such as the cytochrome P450 system [12]. Cytochrome P450, consisted of more than 30 related enzymes, mainly located on the endoplasmic reticulum membrane of hepatocytes and enterocytes. This enzyme group is accountable for drug oxidative metabolisms and endogenous substances such as prostaglandins, fatty acids, and steroids [14]. RIF also can affect the ABCB₁ transporter and P-glycoprotein transport system, which can interfere with the membrane transport activities [12,14].

In recent years, advanced omics technologies have provided a new approach to study drug toxicity; however, RIF-induced hepatotoxicity studies are still centralized on the metabolite level measured in human fluids [15,16] or on the protein level in mice [17]. In this study, we focused on the alteration of DNA methylation, transcriptome, and proteome of human hepatic microtissues under RIF exposure. Our primary aim is to identify potential targets and reveal underlying pathways for the RIF toxic mechanism of action.

Experimental design and dataset

The human hepatic microtissues (3D InSight™ Human Hepatic Microtissues from InSphero) were exposed to either a RIF clinically therapeutic or a toxic (IC₂₀) dose [18]. Every weekday, the sample medium was renewed 3 times corresponding to the hepatic drug interstitial concentration profile at 2, 8, and 24 hours calculated by the physiologically based pharmacokinetic (PBPK) modeling. RIF was dissolved in 0.1% DMSO, thus control samples were also exposed to similar DMSO concentrations (Table S 1-2). During these 2 weeks of exposure, the DNA from microtissues were collected in triplicates at 0, 72 and 168 hours of exposure, the RNA and protein from microtissue were collected in triplicates at 2, 8, 72, 168, and 240 hours of exposure. The RNA and protein from RIF therapeutic-treated samples were also harvested after 336 hours of exposure.

After DNA extraction, the methylated DNA fragments were isolated by anti-5-methylcytosine antibody and then paired-end sequenced (MeDIP-seq) with 50 bp read length [19]. The quality assessment demonstrated that the CpG density-dependent enrichment profiles among samples were sufficient (Figure S 4-1).

After RNA extraction, the total RNA in each sample was isolated using Qiagen AllPrep DNA/RNA/miRNA Universal Kit (Cat #80224). Ribosomal RNAs were depleted by using the Illumina RiboZero Gold kit (Cat #MRZG12324), and then samples were prepared by the Lexogen SENSE total RNA library preparation kit (Cat #009.96). The RNA quality and quantity of the samples were checked by the Agilent 420 TapeStation and the Qubit™ before they were sequenced by an Illumina HiSeq2000 with 100bp paired-end reads [20].

For the protein extraction, microtissues were resuspended in 100 ul lysis buffer containing 8M Urea, 1 mM Dithiothreitol, 0.1M Ammonium bicarbonate, pH 7.8. After four freeze-thaw cycles, the samples were centrifuged at 16000xg for 15 min at 4° C and protein concentrations were assessed with the Qubit™ Protein Assay Kit (Invitrogen, Molecular Probes). Protein isolates were then subjected to in-solution digestion [21] or Filter Aided Sample Preparation (FASP) [22]. Protein digestions were stopped by adding formic acid to a final concentration of 1%. The peptides were cleaned up using Sep-Pak tC18 cartridges (Waters) according to the manufacturer's instructions and eluted with 60% ACN and 0.1% formic acid (Sigma-Aldrich, USA).

Analysis procedure

MeDIP-seq data processing

The MeDIP-seq paired-end reads were aligned to human reference genome hg38 using Burrows-Wheeler Alignment tool (BWA) version 1.17 [23] and converted to .bam files using Samtools version 1.10 [24] in the Linux environment (Figure 2-1). Thereafter, the aligned MeDIP-seq data were processed using the QSEA package and human genome build hg38 with default window size (250 bases) excluding sex chromosomes [25] according to the description of the MeDIP-seq data analysis workflow in chapter 2. The DMRs were detected using generalized linear models (GLMs, p -value < 0.01) with pairwise comparisons between RIF-treated and control samples (Figure 2-1).

RNA sequencing data processing

The sequencing quality of samples was examined by FastQC version 0.11.7 [26], and summarized by MultiQC [27]. Thereafter, the reads were mapped onto the human transcriptome reference (Homo_sapiens.GRCh38.cdna.all.fa.gz) release 104, using Salmon version 1.5.2 [28] with the paired-end option. The annotation file that converts transcripts to genes was also generated from the used transcriptome reference. All samples had sufficient read counts (>25 million) and were used for differential expression (DE) analysis between either RIF therapeutic or toxic-treated samples compared to controls. We also utilized and adapted the R-ODAF to have a rigorous DE analysis, in which we increased the low read counts filtering genes (from >1 CPM of 75% of samples in the group to >5 CPM for all samples in the group) [29].

Proteomic data processing

The raw MS data were processed using Genedata Expressionist® software v.11.0 and annotated to the Uniprot Swiss-Prot 29062016 database with the taxonomy Homo sapiens (human). The protein intensities were computed using the Hi3 method. The \log_2 transformed values of the data were normalized using the median of median method following the similar pre-processing workflow in chapter 3.2. The DE protein analysis was performed in the HeCaToS project [19]. In this particular study, we only processed proteins that matched with unique Uniprot ID for further analysis.

Further analysis

Further analysis was performed in R version 4.0.5 (released on 31st March 2021) [30] with the developed RegOmics tool (chapter 4-1) and some R visualization

packages including Tidyverse [31], and ggplot [32]. Pathway analysis for DE genes and protein was performed using ConsensusPathDB [33] with its default setting and background list of genes and proteins. Cytoscape v3.9.1 was used to visualize the TF-target relations [34].

Results

General view

At the DNA methylation level, the MeDIP-seq data did not show a clear separation between control and RIF-treated samples in the PCA plot (Figure 4-3). However, there is a clear separation between RIF-treated samples and controls on the transcriptome and proteome level, in which the RIF therapeutic and toxic-treated samples were grouped together and distant from control samples (Figure 4-4A-B). In the established clustering trees, even though samples were consistently gathered together as treatment and control groups in both transcriptome and proteome, there were still differences in particular samples grouped by dose and time of exposure based on perceived gene and protein expression (Figure 4-4C). In particular, the RIF-treated samples on the proteomic layer were clustered into 2 groups with early time points of exposure (from 2 until 72 hours) and later time points of exposure (after 168 hours).

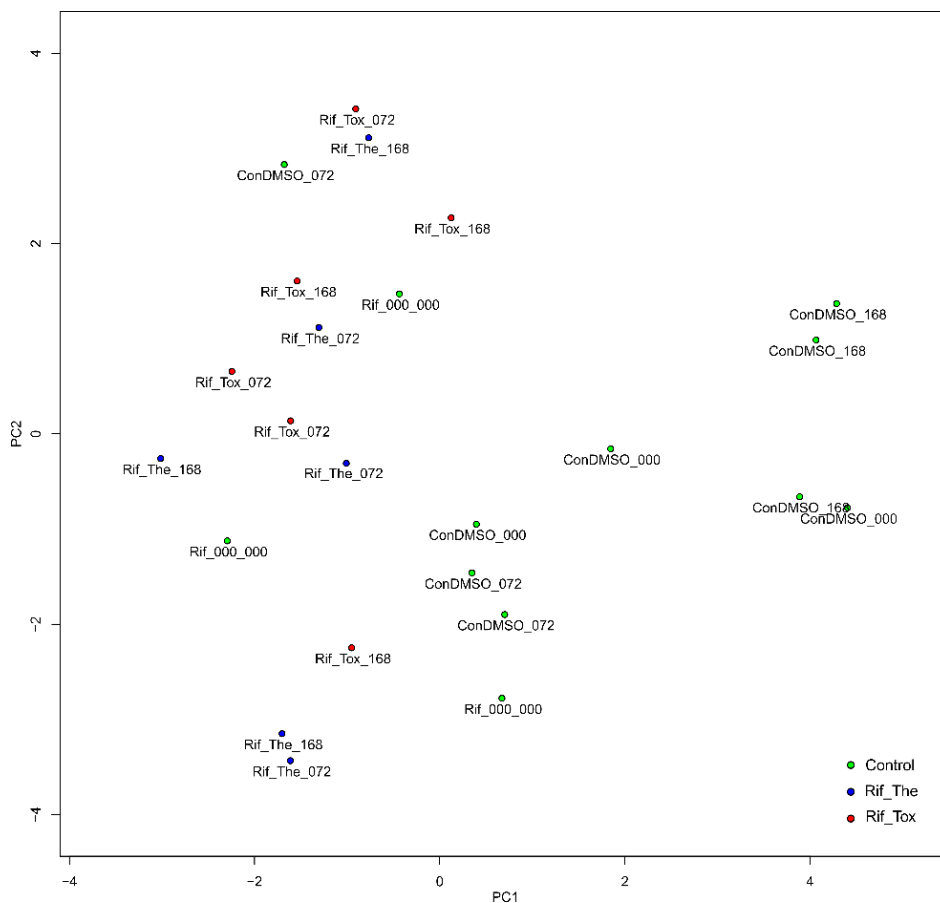


Figure 4-3: PCA plot of MeDIP-seq data from hepatic tissues exposed to RIF-treated samples and controls. **Note:** The microtissues had not been exposed to RIF at 0 hours of exposure, thus the RIF_000 could be considered as control samples.

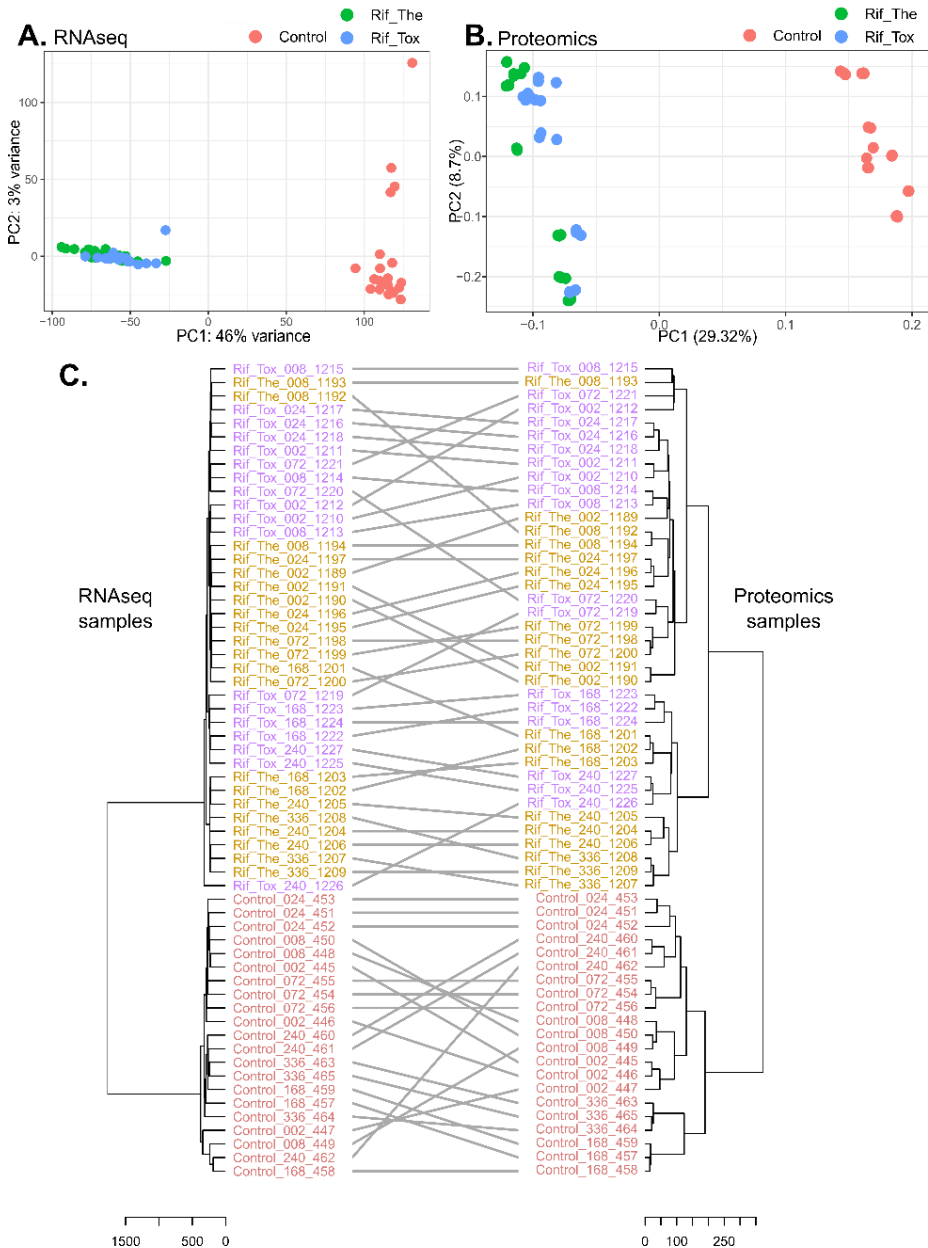


Figure 4-4: Overview of perceived transcriptome and proteome data from RIF-treated and control samples. (A) The PCA plot of RNA-seq data; (B) The PCA plot of proteomics data; (C) A comparison between the sample clustering tree between RNA-seq and proteomics data.

Single omics analyses

The MeDIP-seq data analysis suggested 13 and 1 differential methylated candidate genes under RIF therapeutic and toxic-treated conditions, respectively, compared to controls (Figure 4-5). Intriguingly, the only hypo-methylated gene in the RIF toxic-treated condition, LOC101928626, was hyper-methylated in the RIF therapeutic-treated condition. Depending on the annotation databases, this gene can be considered as a lncRNA, pseudogene or uncharacterized gene [35].

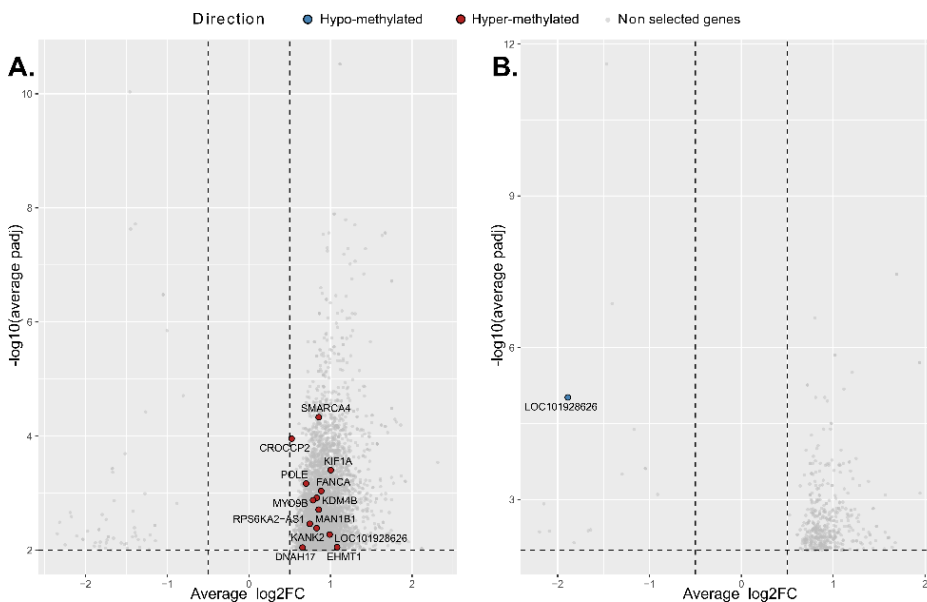


Figure 4-5: The volcano plots as the outcome from the MeDIP-seq analysis. Differential methylated genes in samples treated with RIF therapeutic dose (A) and toxic dose (B) compared to controls.

On the transcriptome level, the DE analysis with R-ODAF filters resulted in 4753 and 3514 DE genes for RIF therapeutic-treated and toxic-treated samples, respectively, compared to controls. While most of these DE genes are protein-coding genes, some of them are labeled as both protein-coding and lncRNA genes in the Ensemble database, such as 5 DE genes at RIF therapeutic-treated samples, i.e. *CABIN1*, *MAPK4*, *PHF1*, *SMIM10L1*, and *SNHG32*, and 3 DE genes at RIF toxic-treated sample, i.e. *MAPK4*, *PHF1*, and *SNHG32* [35]. The DE genes of each RIF treatment condition were used to perform pathway analysis (Figure 4-6A). However, even though most of these DE genes are protein-coding genes, a large part of them did not appear in the gene list of different pathway databases (53.2% and 33.6% of DE genes in RIF therapeutic and toxic-treated conditions, respectively).

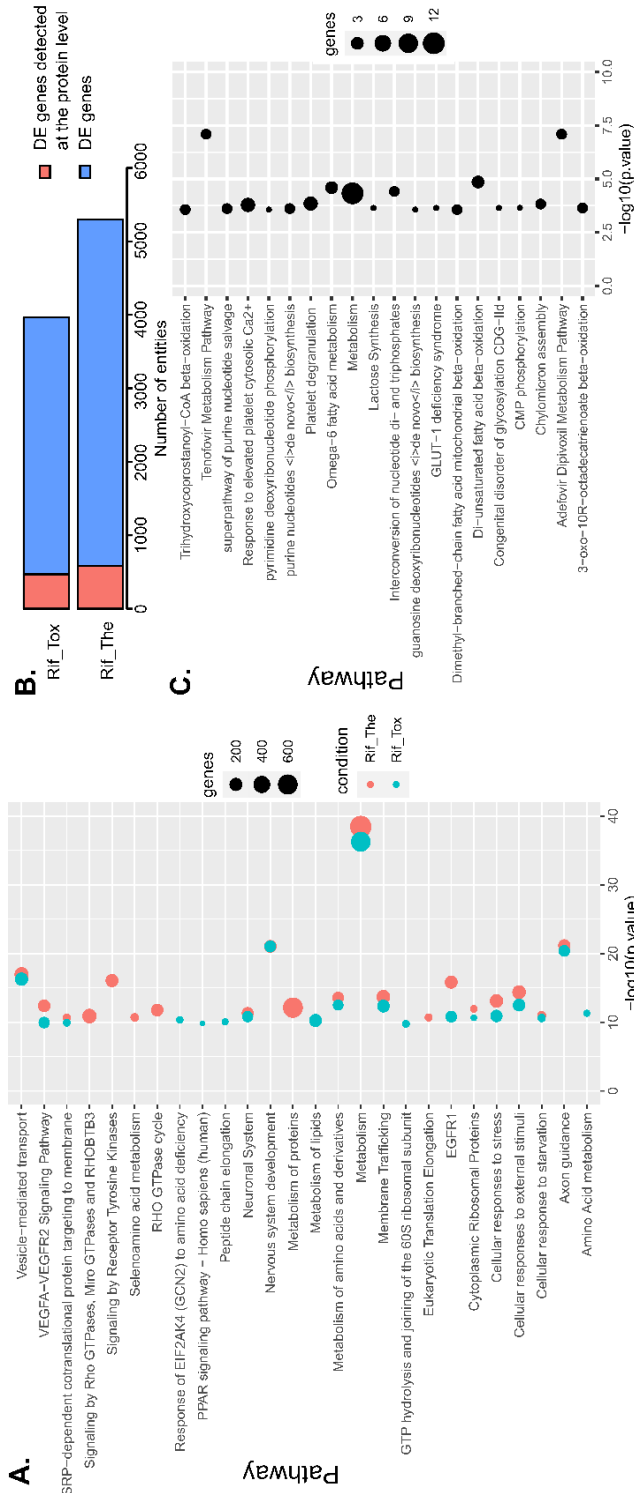


Figure 4-6: Pathway analyses for the selected genes and proteins. (A) Pathway analysis outcome of DE genes from RNA-seq data. (B) The number of detected DE genes at the proteomic level. (C) Pathway analysis outcome of DE proteins from proteomics data.

Although the majority of DE genes on the transcriptomic level can produce protein products, the proteins of a minor part of these genes (less than 14%) were detected on the proteomics level. In particular, only 582 and 471 DE genes on the transcriptomic level, for therapeutic and toxic-treated conditions respectively, were detected on the proteomic level (Figure 4-6B). Due to the dose and time of exposure, the proteomics DE analysis resulted in different DE proteins per RIF treatment condition. Only 11 DE proteins (q-value < 0.05) appeared in all RIF-treated samples compared to control, which are AK1 (P00568), APOA2 (P02652), ENO1 (P06733), PSAP (P07602), KRT7 (P08729), PDIA4 (P13667), NME2 (P22392), CCT6A (P40227), HSPE1 (P61604), PPIA (P62937), and SELENBP1 (Q13228). The protein pathway analysis was performed using these 11 proteins (Figure 4-6C).

TF-target relations across omics layers

Among the 3 aforementioned omics data, the transcriptome displayed the most comprehensive coverage of biomolecules. Thus, the TF-target relation analysis was initially performed on the transcriptomic data before being extended to the MeDIP-seq and proteomics data. On the transcriptome level, the default RegOmics database (chapter 4-1) revealed TF-target regulations for 90.2% and 90.7% DE genes of RIF therapeutic and toxic-treated samples compared to controls. There are 84 identical TFs among RIF-treated conditions (Figure 4-7). While many TFs were influenced by the therapeutic condition, only 6 TFs were strongly impacted by the toxic condition, including *ARNTL*, *ATF5*, *FOXA2*, *HSF2*, *KAT2B*, and *NFE2L2*.

	RIF therapeutic-treated condition		
DE genes	1421	3332	182
TFs	33	84	6
Targets	1218	2924	153
	RIF toxic- treated condition		

Figure 4-7: Number of the DE genes and TF-target relations among RIF-treated conditions on the transcriptome level. **Note:** some genes were both TF and targets, so they were counted for the number of TF and number of targets separately.

By conjoining the TF-target relations in transcriptomics with the accomplished MeDIP-seq data analysis, we recognized some differential methylated genes that were differentially expressed on the transcriptome level

and were also involved in multiple regulatory interactions (Figure 4-8). For instance, 4 genes, i.e. *SMARCA4*, *MAN1B1*, *POLE*, and *KIF1A* were hyper-methylated (Figure 4-3) and differentially expressed on the transcriptome level for the RIF therapeutic-treated condition. By contrast, the *LOC101928626* gene, which was differentially methylated in both RIF-treated conditions, was not differentially expressed on the transcriptome level. In particular, *SMARCA4* is both a transcription factor and target gene; and it was involved in more than 700 possible TF-target relations. Especially, *SMARCA4* can regulate the *MAN1B1* gene (Figure 4-8). While TFs that regulate *KIF1A* were not differentially expressed on the transcriptome level in RIF-treated samples, some TFs that regulate *MAN1B1* and *POLE* were differentially expressed in RIF-treated conditions (Figure 4-8).

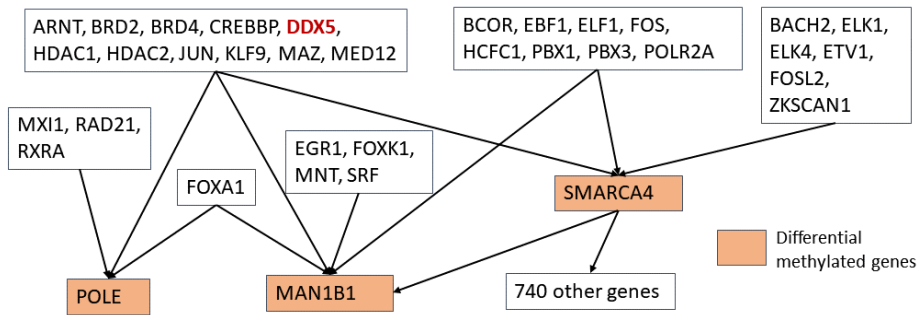


Figure 4-8: The TF-target relations of 3 hyper-methylated genes for RIF therapeutic-treated samples: *SMARCA4*, *MAN1B1*, and *POLE*. The gene *DDX5* is highlighted in red and is related to **Figure 4-9**.

The TF-target analysis was then expanded from transcriptome to proteome. While there were 11 detected DE proteins across all RIF-treated conditions, only using these 11 proteins could constrain the TF-target analysis power. Thus, we expanded the TF-target analysis to proteins that were detected in at least one RIF-treated condition and were DE on the transcriptome level. Furthermore, there are TFs that can regulate the DE genes or detected proteins but did not appear in the DE gene list on the transcriptomic level or were not detected on the proteomic level. These TFs were labeled as retrieved TFs (Table 4-1). Although there were 21 identical retrieved TFs across RIF-treated conditions, there was no overlapping in the top 5% retrieved TFs among treatments.

Table 4-1: The number of TFs in the transcriptome and proteome under RIF exposure

Data type	Input type	Treatment	Entities	Existing TFs in the input entities	Retrieved TFs	Top 5% of the retrieved TFs
DE genes at transcriptome	RNA-seq	The	4716	117	575	28
		Tox	3493	90	555	27
DE genes appeared in proteome data	Protein	The	582	1	431	22
		Tox	471	1	423	21
Number of identical TFs across input types and treatment conditions			-	1	21	0

Notes: The, therapeutic-treated samples; Tox, toxic-treated samples; TF, transcription factor; -, not applicable.

Out of the overlapped 84 TFs in the transcriptome (Figure 4-7), only 1 TF, namely DDX5, was detected in the proteomic data (Table 4-1). We used DDX5 as a case study for exploring the regulatory relationship in RIF treatment. DDX5 regulates the 3 hyper-methylated genes at the RIF therapeutic-treated condition as mentioned in the DNA methylation analysis (Figure 4-8) as well as were involved in other TF-target relations in the transcriptomics and proteomics (Figure 4-9). While there is no change in the DNA methylation status of DDX5 after RIF treatment, on the transcriptome level, DDX5 was differentially expressed in both therapeutic and toxic treatment and involved in multiple TF-target relations. One target of DDX5 is *SMARCA4* that was hyper-methylated at the RIF therapeutic-treated condition (Figure 4-8, Table S 4-1). At the RIF therapeutic-treated condition, *SMARCA4* was also hyper-methylated although its methylation status did not fulfill all the selected criterial of the MeDIP-seq data analysis (Table S 4-2). *SMARCA4* was differentially expressed at the transcriptome level in both RIF-treated conditions. Furthermore, some genes that are involved in DDX5 regulations were in the top 5% TFs in the therapeutic treatment, such as HDAC1 and FOXA1, or in both therapeutic and toxic treatment, such as MAZ, BRD4, and HDAC2. Expanding to the proteome, some of these genes were also detected on the proteome level (Figure 4-9).

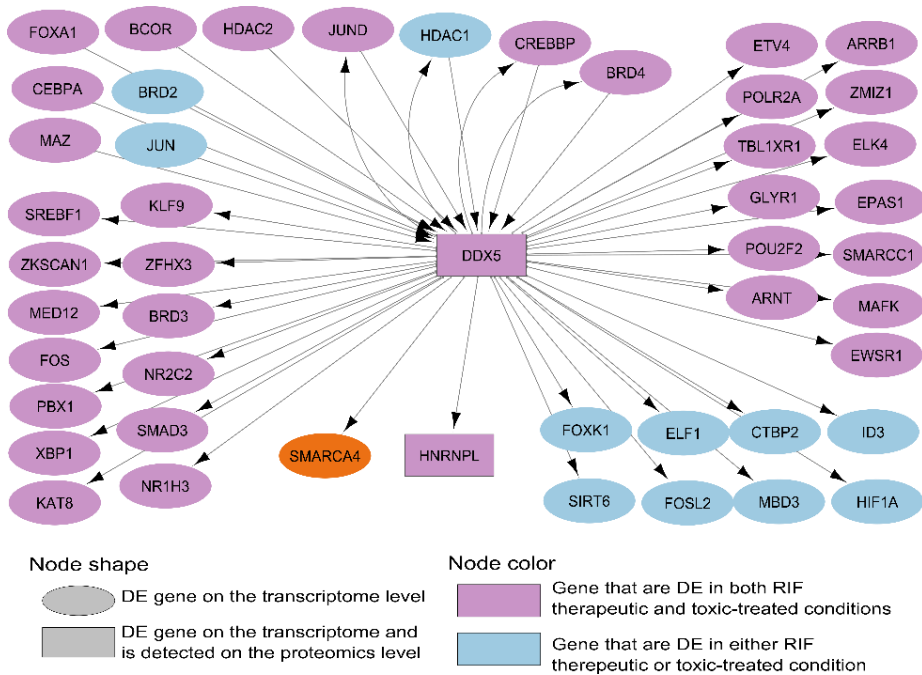


Figure 4-9: The TF-target network of *DDX5* observed in the RIF-treated sampled. The gene *SMARCA4* is highlighted in orange. *SMARCA4* was differentially expressed (DE) in both RIF-treated conditions on the transcriptome level and were differentially methylated in the RIF therapeutic-treated condition (**Figure 4-8**).

Thus, TF-target analysis highlighted some TF genes that can be the key factors to regulate the cellular process under RIF treatment. By exploring different omics data derived from the same RIF-treated samples, we identified candidate genes, such as *DDX5* and *SMARCA4*, and their regulatory relationships that spread through the methylome via the transcriptome to the proteome.

Discussion

This chapter 4-2 explored the molecular interruptions in hepatic microtissues due to RIF exposure, including DNA methylation modification, transcriptomic, and proteomic alterations. We identified not only the potential biomarkers in different omics layers but also inspected the interdependences between biomolecules by projecting the outcomes of each omics data analysis to the regulatory relationships. Some TF-target relations, such as *DDX5* - *SMARCA4*, were evident across 3 omics layers. Thus, the cross-omics analysis helped to filter the noises from single omics layers and strengthened the possibly important regulations under drug treatment. This study not only provided a global view of RIF-induced

hepatotoxicity at different omics layers but also prospected candidate genes and which regulations these genes could be involved.

In general, different -omics layers can represent different scales of RIF-induced effects. For instance, while the transcriptomic and proteomics data could clearly portray the difference between RIF-treated samples compared to control (Figure 4-4 A-B), the PCA plot of the MeDIP-seq data could not do this (Figure 4-3). RIF-treated samples were better differentiated from controls on the transcriptome and proteome level rather than on the methylome level. Together with another urinary metabolomics study [15], it seems that RIF has a clear tendency in interrupting the downstream products of the genome such as transcriptome, proteome, and perhaps even metabolome. Furthermore, the perceived RNA and protein expression derived even from the same samples could facilitate different sample grouping according to dose and time of exposure differences (Figure 4-4C). This suggests that single omics analyses may only depict a part of the biological status and may include some noise signals from that particular omics layer. Thus, relying on particular omics layers can obstruct the drug adverse effect investigation.

The MeDIP-seq data analysis revealed 13 hyper-methylated and 1 hypo-methylated genes under RIF therapeutic and toxic-treated conditions compared to controls, respectively (Figure 4-5). Although the LOC101928626 gene, which can be labeled as a lncRNA gene [35], was differentially methylated in both RIF therapeutic and toxic-treated conditions, there is not much prior knowledge about this gene. Similarly, other genes, such as *EHMT1*, *FANCA*, *KDM4B*, *MAN1B1*, and *SMARCA4*, also have not been deeply studied in the context of hepatotoxicity. For instance, the down-regulation of *KDM4* levels could contribute to liver fibrosis pathology [36]. While *EHMT1* seems to be related to non-alcoholic fatty liver disease [37], *FANCA* may be involved in the hepatocyte gene repair pathway [38]. *MAN1B1* and *SMARCA4* were differentially methylated on the DNA methylome level and even differential expressed on the transcriptome level. Nevertheless, there are not many studies about them and their regulation in hepatotoxicity (Figure 4-8), although some studies have focused on the role of *SMARCA4* in liver cancer [39].

The transcriptome and proteomics analysis unveiled DE genes and proteins that are involved in important cellular pathways such as metabolism and energy production (Figure 4-6). A metabolomics study demonstrated that anti-TB drugs including RIF might affect the tricarboxylic acid circulation (TCA cycle) pathway and the purine metabolic pathway [15]. These 2 pathways also appeared as the top

pathways of our proteomics analysis (Figure 4-6C). While the proteomic outcome in this study did not reveal any arginine or proline metabolic pathways as they were recognized in this previous metabolomics study [15], the outcome of the transcriptomic analysis emphasized the amino acid metabolism, the metabolism of other cell substances as well as the PPAR signaling pathway and translation elongation (Figure 4-6A). A toxicoproteomics study in mice also affirmed that RIF can impact the amino acid biosynthesis, ribosome, and PPAR signaling pathway [17].

Similar to the detected differential methylated genes, there is not much knowledge about how the detected DE genes and proteins are related to RIF-induced hepatotoxicity. Some DE genes could be related to hepatic disease. For instance, *SNHG32* generated *SNORD52* that promotes hepatocellular carcinoma tumorigenesis [40]. Some DE proteins are also known to be associated with drug-induced hepatotoxicity. For example, AK-1 is an NAD⁺-dependent deacetylases inhibitor; using this protein as a treatment could attenuate oxidative stress and alleviates carbon tetrachloride-induced hepatotoxicity [41]. ENO₁ is a multifunctional enzyme and is consistently up-regulated during liver injury under geniposide exposure [42]. According to a primary mouse hepatocyte study, *PDIA4* also changed its protein expression level under cyclosporin A exposure [43].

Pathway analysis can reveal the function of gene or protein groups but does not clearly state the relation between these entities. The TF-target analysis could fulfil this gap for a better understanding of the biological system. Especially, in the transcriptomic data of this study, pathway analysis only covered 53.2% and 33.6% of DE genes in RIF therapeutic and toxic-treated conditions, respectively; however, the default TF-target database from the RegOmics tool was able to cover more than 90% of DE genes for each RIF-treated condition compared to controls. By projecting the relations retrieved from transcriptomic data to the outcome of the MedIP-seq and proteomics data analysis, we were able to form and reinforce the regulations that appeared in different omics layers. This cross-omics analysis helped to refine the important TF-target relations from the massive relations that can be influenced by RIF treatment.

In particular, we disclosed 6 TFs, i.e. *ARNTL*, *ATF5*, *FOXA2*, *HSF2*, *KAT2B*, and *NFE2L2*, that are only differentially expressed in the RIF toxic condition. *ARNTL* (aryl-hydrocarbon receptor nuclear translocator-like) is a transcription factor that can bind to the promoter of *Sirt1*, a metabolic mediator. The accumulation of *ARNTL* protein in nucleus enhances the *Sirt1* transcription, and then regulates endoplasmic reticulum stress, lipogenesis, and hepatic inflammation in

nonalcoholic fatty liver disease [44]. Furthermore, ATF5 is a highly abundant liver-enriched transcription factor [45]. ATF5 participates in cellular differentiation and promotes cellular adaptation to stress by activating molecular chaperones, proteases, and prosurvival molecules [46]. It also promotes mitochondrial function and recovery from mitochondrial stress [47]. Other TFs are also associated with different hepatic processes, such as HSF2 can regulate aerobic glycolysis by suppressing FBP1 in hepatocellular carcinoma [48]. While NFE2L2 plays a role in hepatoprotection against lipotoxicity [49], FOXA2 can maintain apical MRP2 expression and prevent hyperbilirubinemia in acute liver failure [50]. Thus, TF-target analysis revealed candidate TFs that have crucial functions in regulatory units for RIF toxic mechanisms.

Among various TF candidates (Table 4-1), DDX5 also seems to play a central role in the RIF-induced interruptions. It can regulate SMARCA4, MAN1B1, and POLE that were hyper-methylated under RIF therapeutic treatment (Figure 4-8). DDX5 was differentially expressed on the transcriptome level and then was detected on the proteome level despite the finite protein detection. Apparently, DDX5 can regulate and be regulated by several genes that were DE on transcriptome levels under RIF treatment (Figure 4-9). However, there has not yet been much investigation on DDX5's function in hepatotoxicity. Based on previous studies, some of DDX5's targets have different functions in liver cells. For instance, CEBPA has specific regulatory functionalities for homeostatic and cell cycle in liver regeneration [51]. HDAC, which is histone deacetylase and can play an epigenetic role, might be involved in the carbon tetrachloride-induced hepatotoxicity; the increase of HDAC could reduce liver injury [52]. Other genes in co-regulated relationship with DDX5, such as JUND and HDAC1, are involved in oxidative stress, liver dysfunction, and the hepatocellular carcinoma process [53,54]. Other genes, which regulate DDX5, are also known for their critical functions in hepatotoxicity. In particular, KLF9, XBP1, SIRT6, and NR1H3 play important roles in arsenic trioxide, acetaminophen, and cadmium-induced hepatotoxicity, respectively [55-58]. Possibly, due to RIF treatment, DDX5's expression alters on the transcriptome level and triggers other genes; this consequently leads to RIF-induced hepatotoxicity. However, a wide range of possible TF-target relations without prior knowledge about regulatory pathways related to the particular treatment condition can create difficulties to select candidate genes/proteins for further investigations.

General remarks

In this chapter, we developed a TF-target detection tool (RegOmics) to extract the biological regulations from omics data. We also separately analyze the epigenome, transcriptome, and proteome of hepatic microtissues under RIF exposure. Thereafter, we employed the RegOmics tool to extract the biological regulations related to RIF-induced hepatotoxicity. We were able to inspect the potential regulations that were affected by RIF treatment.

From the bioinformatics perspective, we launched the RegOmics tool for TF-target extraction from extensive lists of genes and/or proteins. While several tools have been developed for TFs or TF-target detection, they focus on either protein-coding genes or lncRNAs. Besides, some of them are not currently available or user-friendly. Our RegOmics tool combines two TF-target databases to cover the regulations in both protein-coding and lncRNA genes. This tool has been built on R and available as R source code as well as R shiny app. The R shiny app offers a user-friendly interface for researchers to inspect the TF-target regulation in their data without coding skill requirements.

For the drug side effects, we explored the genome-wide DNA methylation modification as well as the transcriptome-wide and proteome-wide alterations under RIF treatment in *in vitro* human hepatic microtissues. Whereas RIF is known for its hepatotoxicity effects, there is little study on the underlying RIF toxic mechanism. The chapter 4-2 revealed several changes in DNA methylation, gene, and protein expression that seem to be associated with RIF toxic mechanism. However, there is insufficient prior knowledge about these genes' and proteins' function nor about how they could be involved in drug-induced hepatotoxicity. While this study suggested potential candidates for further RIF-induced hepatotoxicity studies, it also portrayed an urgent need for experimental validation.

Supplementary Materials

Data & code accessibility

The data presented in the study are deposited in the BioStudies repository (<http://www.ebi.ac.uk/biostudies>) with the accession numbers: S-HECA421, S-HECA410, S-HECA411, and S-HECA401 for methylome data; S-HECA157 and S-HECA158 for transcriptome data; S-HECA131 for proteome data.

The RegOmics tool is publicly available on R shiny app as a web-based version (<https://nhannguyen.shinyapps.io/RegOmics/>) and on Github as open source code (<https://github.com/NhanNguyenooo/RegOmics>).

The code of the data analysis is publicly available on Github (https://github.com/NhanNguyenooo/Rifampicin_omics_integration).

Supplementary figures

CpG Enrichment Analysis

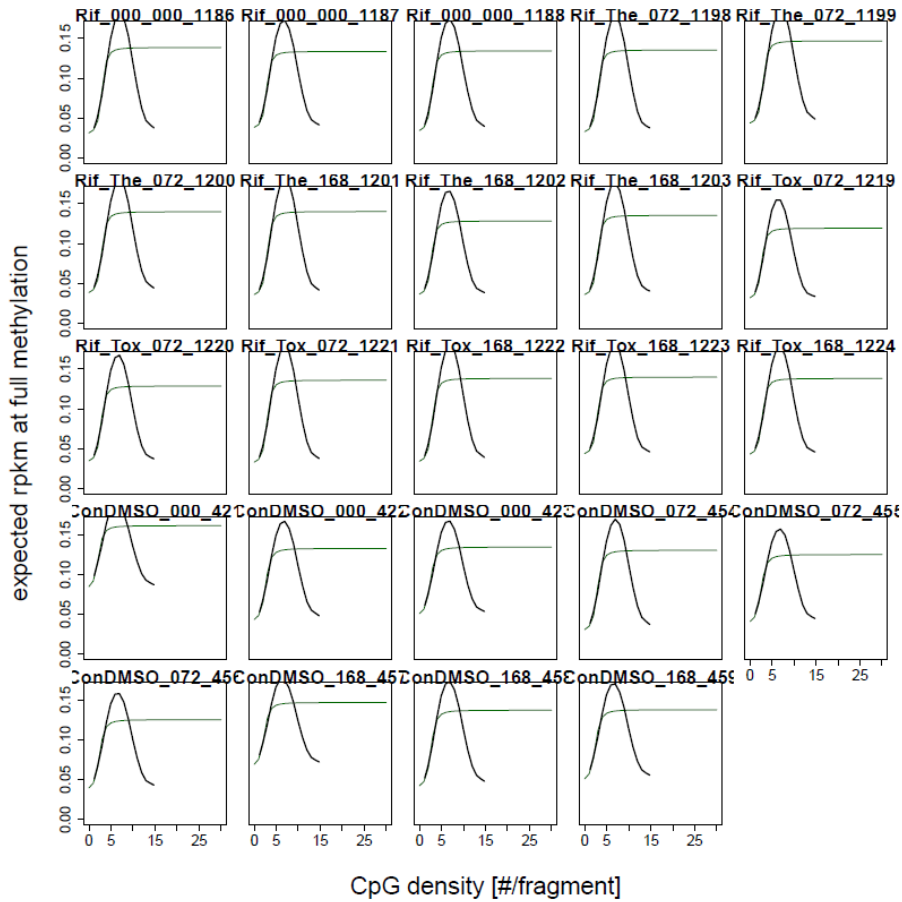


Figure S 4-1: The sample-wise CpG density-dependent enrichment profiles. The average enrichment profile of samples is depicted in black, and the fitted sigmoidal function is in green. Samples with flat profiles might indicate low enrichment efficiency or poor agreement with the calibration data. In here, all samples show sufficient enrichment profiles.

Supplementary tables

Table S 4-1: Differentially methylated genes between RIF therapeutic-treated and control samples.

Symbol	The average of log ₂ FC values	The average of p-values	The average of adj-p-values
<i>CROCCP2</i>	0.524027	1.46E-08	0.000111
<i>DNAH17</i>	0.657212	1.51E-05	0.009017
<i>EHMT1</i>	1.081145	1.47E-05	0.008888
<i>FANCA</i>	0.88627	4.01E-07	0.000917
<i>KANK2</i>	0.830295	4.30E-06	0.004101
<i>KDM4B</i>	0.832762	6.30E-07	0.00121
<i>KIF1A</i>	1.002727	1.04E-07	0.000394
<i>LOC101928626</i>	0.990407	6.40E-06	0.005324
<i>MAN1B1</i>	0.855219	1.31E-06	0.001949
<i>MYO9B</i>	0.787109	7.36E-07	0.001331
<i>POLE</i>	0.701396	2.48E-07	0.000681
<i>RPS6KA2-AS1</i>	0.747184	3.30E-06	0.003455
<i>SMARCA4</i>	0.856038	3.70E-09	4.67E-05

Table S 4-2: Differentially methylated genes between RIF toxic-treated and control samples.

Symbol	The average of log ₂ FC values	The average of p-values	The average of adj-p-values
<i>LOC101928626</i>	-1.88828	4.86E-11	9.60E-06
<i>SMARCA4*</i>	0.7260945	1.20E-07	0.002930994

Note: (*) This differentially methylated gene did not fulfill all the selected criterial of the MedIP-seq data analysis (**Figure 2-1**).

References

- [1] Q. Zhang, W. Liu, H.M. Zhang, G.Y. Xie, Y.R. Miao, M. Xia, A.Y. Guo, hTFtarget: A comprehensive database for regulations of human transcription factors and their targets, *Genomics Proteomics Bioinformatics* 18 (2020) 120-128. <https://doi.org/10.1016/j.gpb.2019.09.006>
- [2] A.B. Keenan, D. Torre, A. Lachmann, A.K. Leong, M.L. Wojciechowicz, V. Utti, K.M. Jagodnik, E. Kropiwnicki, Z. Wang, A. Ma'ayan, ChEA3: transcription factor enrichment analysis by orthogonal omics integration, *Nucleic Acids Res* 47 (2019) W212-w224. <https://doi.org/10.1093/nar/gkz446>
- [3] Q. Qin, J. Fan, R. Zheng, C. Wan, S. Mei, Q. Wu, H. Sun, M. Brown, J. Zhang, C.A. Meyer, X.S. Liu, Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data, *Genome Biology* 21 (2020) 32. <https://doi.org/10.1186/s13059-020-1934-6>
- [4] Z. Bao, Z. Yang, Z. Huang, Y. Zhou, Q. Cui, D. Dong, LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases, *Nucleic Acids Res* 47 (2019) D1034-d1037. <https://doi.org/10.1093/nar/gky905>
- [5] H. Zhao, J. Shi, Y. Zhang, A. Xie, L. Yu, C. Zhang, J. Lei, H. Xu, Z. Leng, T. Li, W. Huang, S. Lin, L. Wang, Y. Xiao, X. Li, LncTarD: a manually-curated database of experimentally-supported functional lncRNA-target regulations in human diseases, *Nucleic Acids Research* 48 (2019) D118-D126. <https://doi.org/10.1093/nar/gkz985>
- [6] Z. Zhou, Y. Shen, M.R. Khan, A. Li, LncReg: a reference resource for lncRNA-associated regulatory networks, *Database* 2015 (2015). <https://doi.org/10.1093/database/bavo83>
- [7] Q. Jiang, J. Wang, Y. Wang, R. Ma, X. Wu, Y. Li, TF2LncRNA: identifying common transcription factors for a list of lncRNA genes from ChIP-Seq data, *BioMed research international* 2014 (2014) 317642-317642. <https://doi.org/10.1155/2014/317642>
- [8] L. Cheng, P. Wang, R. Tian, S. Wang, Q. Guo, M. Luo, W. Zhou, G. Liu, H. Jiang, Q. Jiang, LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse, *Nucleic acids research* 47 (2019) D140-D144. <https://doi.org/10.1093/nar/gky1051>
- [9] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res* 46 (2018) D1074-d1082. <https://doi.org/10.1093/nar/gkx1037>
- [10] A.A. Abulfathi, E.H. Decloedt, E.M. Svensson, A.H. Diacon, P. Donald, H. Reuter, Clinical Pharmacokinetics and Pharmacodynamics of Rifampicin in Human Tuberculosis, *Clinical Pharmacokinetics* 58 (2019) 1103-1129. <https://doi.org/10.1007/s40262-019-00764-2>
- [11] R. Prasad, N. Gupta, A. Banka, Multidrug-resistant tuberculosis/rifampicin-resistant tuberculosis: Principles of management, *Lung India : official organ of Indian Chest Society* 35 (2018) 78-81. https://doi.org/10.4103/lungindia.lungindia_98_17
- [12] V. Ramappa, G.P. Aithal, Hepatotoxicity Related to Anti-tuberculosis Drugs: Mechanisms and Management, *Journal of clinical and experimental hepatology* 3 (2013) 37-49. <https://doi.org/10.1016/j.jceh.2012.12.001>

- [13] A. Ramachandran, R.G.J. Visschers, L. Duan, J.Y. Akakpo, H. Jaeschke, Mitochondrial dysfunction as a mechanism of drug-induced hepatotoxicity: current understanding and future perspectives, *Journal of clinical and translational research* 4 (2018) 75-100. <https://doi.org/10.18053/jctres.04.201801.005>
- [14] W.W. Yew, Clinically Significant Interactions with Drugs Used in the Treatment of Tuberculosis, *Drug Safety* 25 (2002) 111-113. <https://doi.org/10.2165/00002018-200225020-00005>
- [15] J. Cao, Y. Mi, C. Shi, Y. Bian, C. Huang, Z. Ye, L. Liu, L. Miao, First-line anti-tuberculosis drugs induce hepatotoxicity: A novel mechanism based on a urinary metabolomics platform, *Biochemical and Biophysical Research Communications* 497 (2018) 485-491. <https://doi.org/10.1016/j.bbrc.2018.02.030>
- [16] M. Combrink, D.T. Loots, I. du Preez, Metabolomics describes previously unknown toxicity mechanisms of isoniazid and rifampicin, *Toxicology Letters* 322 (2020) 104-110. <https://doi.org/10.1016/j.toxlet.2020.01.018>
- [17] J.-H. Kim, W.S. Nam, S.J. Kim, O.K. Kwon, E.J. Seung, J.J. Jo, R. Shresha, T.H. Lee, T.W. Jeon, S.H. Ki, H.S. Lee, S. Lee, Mechanism Investigation of Rifampicin-Induced Liver Injury Using Comparative Toxicoproteomics in Mice, *International journal of molecular sciences* 18 (2017) 1417. <https://doi.org/10.3390/ijms18071417>
- [18] L. Kuepfer, O. Clayton, C. Thiel, H. Cordes, R. Nudischer, L.M. Blank, V. Baier, S. Heymans, F. Caiment, A. Roth, D.A. Fluri, J.M. Kelm, J. Castell, N. Selevsek, R. Schlapbach, H. Keun, J. Hynes, U. Sarkans, H. Gmuender, R. Herwig, S. Niederer, J. Schuchhardt, M. Segall, J. Kleinjans, A model-based assay design to reproduce in vivo patterns of acute drug-induced toxicity, *Archives of Toxicology* 92 (2018) 553-555. <https://doi.org/10.1007/s00204-017-2041-7>
- [19] N. Selevsek, F. Caiment, R. Nudischer, H. Gmuender, I. Agarkova, F.L. Atkinson, I. Bachmann, V. Baier, G. Barel, C. Bauer, S. Boerno, N. Bosc, O. Clayton, H. Cordes, S. Deeb, S. Gotta, P. Guye, A. Hersey, F.M.I. Hunter, L. Kunz, A. Lewalle, M. Lienhard, J. Merken, J. Minguet, B. Oliveira, C. Pluess, U. Sarkans, Y. Schrooders, J. Schuchhardt, I. Smit, C. Thiel, B. Timmermann, M. Verheijen, T. Wittenberger, W. Wolski, A. Zerck, S. Heymans, L. Kuepfer, A. Roth, R. Schlapbach, S. Niederer, R. Herwig, J. Kleinjans, Network integration and modelling of dynamic drug responses at multi-omics levels, *Communications biology* 3 (2020) 573-573. <https://doi.org/10.1038/s42003-020-01302-8>
- [20] M. Verheijen, Y. Schrooders, H. Gmuender, R. Nudischer, O. Clayton, J. Hynes, S. Niederer, H. Cordes, L. Kuepfer, J. Kleinjans, F. Caiment, Bringing in vitro analysis closer to in vivo: Studying doxorubicin toxicity and associated mechanisms in 3D human microtissues with PBPK-based dose modelling, *Toxicology Letters* 294 (2018) 184-192. <https://doi.org/10.1016/j.toxlet.2018.05.029>
- [21] N. Selevsek, M. Matondo, M.S. Carbayo, R. Aebersold, B. Domon, Systematic quantification of peptides/proteins in urine using selected reaction monitoring, *Proteomics* 11 (2011) 1135-1147. <https://doi.org/10.1002/pmic.201000599>
- [22] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nature Methods* 6 (2009) 359-362. <https://doi.org/10.1038/nmeth.1322>
- [23] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* 25 (2009) 1754-60. <https://doi.org/10.1093/bioinformatics/btp324>

- [24] P. Danecek, J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies, H. Li, Twelve years of SAMtools and BCftools, *Gigascience* 10 (2021). <https://doi.org/10.1093/gigascience/giab008>
- [25] M. Lienhard, S. Grasse, J. Rolff, S. Frese, U. Schirmer, M. Becker, S. Börno, B. Timmermann, L. Chavez, H. Sülthmann, G. Leschber, I. Fichtner, M.R. Schweiger, R. Herwig, QSEA-modelling of genome-wide DNA methylation from sequencing enrichment experiments, *Nucleic acids research* 45 (2017) e44-e44. <https://doi.org/10.1093/nar/gkx1193>
- [26] S. Andrews, FastQC: a quality control tool for high throughput sequence data, in: Babraham Institute, Babraham, UK, 2010.
- [27] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics* 32 (2016) 3047-3048. <https://doi.org/10.1093/bioinformatics/btw354>
- [28] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, *Nature methods* 14 (2017) 417-419. [10.1038/nmeth.4197](https://doi.org/10.1038/nmeth.4197)
- [29] M. Verheijen, W. Tong, L. Shi, T.W. Gant, B. Seligman, F. Caiment, Towards the development of an omics data analysis framework, *Regulatory Toxicology and Pharmacology* 112 (2020) 104621. <https://doi.org/10.1016/j.yrtph.2020.104621>
- [30] R.C. Team, R: A language and environment for statistical computing, in: Vienna, Austria, 2013.
- [31] H. Wickham, M. Averick, J. Bryan, W. Chang, L.D.A. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, Welcome to the Tidyverse, *Journal of Open Source Software* 4 (2019) 1686. <https://doi.org/10.21105/joss.01686>
- [32] H. Wickham, M.H. Wickham, The ggplot package, in: 2007.
- [33] R. Herwig, C. Hardt, M. Lienhard, A. Kamburov, Analyzing and interpreting genome data at the network level with ConsensusPathDB, *Nature Protocols* 11 (2016) 1889-1907. <https://doi.org/10.1038/nprot.2016.117>
- [34] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T.J.G.r. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome research* 13 (2003) 2498-2504. <https://doi.org/10.1101/gr.1239303>
- [35] M. Safran, N. Rosen, M. Twik, R. BarShir, T.I. Stein, D. Dahary, S. Fishilevich, D. Lancet, The GeneCards Suite, in: I. Abugessaisa, T. Kasukawa (Eds.), *Practical Guide to Life Science Databases*, Springer Singapore, Singapore, 2021, pp. 27-56.
- [36] M. Kong, J. Wu, Z. Fan, B. Chen, T. Wu, Y. Xu, The histone demethylase Kdm4 suppresses activation of hepatic stellate cell by inducing MiR-29 transcription, *Biochemical and Biophysical Research Communications* 514 (2019) 16-23. <https://doi.org/10.1016/j.bbrc.2019.04.105>
- [37] C.J. Pirola, S. Sookoian, Epigenetics factors in nonalcoholic fatty liver disease, *Expert Review of Gastroenterology & Hepatology* (2020) 1-16. <https://doi.org/10.1080/17474124.2020.1765772>

- [38] Q.-S. Zhang, A. Tiyaboonchai, S. Nygaard, K. Baradar, A. Major, N. Balaji, M. Grompe, Induced Liver Regeneration Enhances CRISPR/Cas9-Mediated Gene Repair in Tyrosinemia Type 1, *Human Gene Therapy* 32 (2021) 294-301. <https://doi.org/10.1089/hum.2020.042>
- [39] S.Y. Kim, Q. Shen, K. Son, H.S. Kim, H.D. Yang, M.J. Na, E. Shin, S. Yu, K. Kang, J.S. You, K.-R. Yu, S.M. Jeong, E.K. Lee, Y.M. Ahn, W.S. Park, S.W. Nam, SMARCA4 oncogenic potential via IRAK1 enhancer to activate Gankyrin and AKR1B10 in liver cancer, *Oncogene* 40 (2021) 4652-4662. <https://doi.org/10.1038/s41388-021-01875-6>
- [40] C. Li, L. Wu, P. Liu, K. Li, Z. Zhang, Y. He, Q. Liu, P. Jiang, Z. Yang, Z. Liu, Y. Yuan, L. Chang, The C/D box small nucleolar RNA SNORD52 regulated by Upfi facilitates Hepatocarcinogenesis by stabilizing CDK1, *Theranostics* 10 (2020) 9348-9363. <https://doi.org/10.7150/thno.47677>
- [41] Z. Zhou, J. Qi, J.-W. Kim, M.-j. You, C.W. Lim, B. Kim, AK-1, a Sirt2 inhibitor, alleviates carbon tetrachloride-induced hepatotoxicity in vivo and in vitro, *Toxicology Mechanisms and Methods* 30 (2020) 324-335. <https://doi.org/10.1080/15376516.2020.1729915>
- [42] J. Wei, F. Zhang, Y. Zhang, C. Cao, X. Li, D. Li, X. Liu, H. Yang, L. Huang, Proteomic Investigation of Signatures for Geniposide-Induced Hepatotoxicity, *Journal of Proteome Research* 13 (2014) 5724-5733. <https://doi.org/10.1021/pr500719u>
- [43] W.F.P.M. Van den Hof, A. Van Summeren, A. Lommen, M.L.J. Coonen, K. Brauers, M. van Herwijnen, W.K.W.H. Wodzig, J.C.S. Kleinjans, Integrative cross-omics analysis in primary mouse hepatocytes unravels mechanisms of cyclosporin A-induced hepatotoxicity, *Toxicology* 324 (2014) 18-26. <https://doi.org/10.1016/j.tox.2014.06.003>
- [44] Y. Chen, X. Chen, J. Gao, C. Xu, P. Xu, Y. Li, Y. Zhu, C. Yu, Long noncoding RNA FLRL2 alleviated nonalcoholic fatty liver disease through Arntl-Sirt1 pathway, *The FASEB Journal* 33 (2019) 11411-11419. <https://doi.org/10.1096/fj.201900643RRR>
- [45] M. Pascual, M.J. Gómez-Lechón, J.V. Castell, R. Jover, ATF5 is a highly abundant liver-enriched transcription factor that cooperates with constitutive androstane receptor in the transactivation of CYP2B6: Implications in hepatic stress responses, *Drug Metabolism and Disposition* 36 (2008) 1063. <https://doi.org/10.1124/dmd.107.019380>
- [46] T.K. Sears, J.M. Angelastro, The transcription factor ATF5: role in cellular differentiation, stress responses, and cancer, *Oncotarget* 8 (2017) 84595-84609. <https://doi.org/10.18632/oncotarget.21102>
- [47] Christopher J. Fiorese, Anna M. Schulz, Y.-F. Lin, N. Rosin, Mark W. Pellegrino, Cole M. Haynes, The Transcription Factor ATF5 Mediates a Mammalian Mitochondrial UPR, *Current Biology* 26 (2016) 2037-2043. <https://doi.org/10.1016/j.cub.2016.06.002>
- [48] L.-N. Yang, Z.-Y. Ning, L. Wang, X. Yan, Z.-Q. Meng, HSF2 regulates aerobic glycolysis by suppression of FBP1 in hepatocellular carcinoma, *American journal of cancer research* 9 (2019) 1607-1621.
- [49] D.H. Lee, J.S. Park, Y.S. Lee, J. Han, D.-K. Lee, S.W. Kwon, D.H. Han, Y.-H. Lee, S.H. Bae, SQSTM1/p62 activates NFE2L2/NRF2 via ULK1-mediated autophagic KEAP1 degradation and protects mouse liver from lipotoxicity, *Autophagy* 16 (2020) 1949-1973. <https://doi.org/10.1080/15548627.2020.1712108>
- [50] S. Wang, R. Feng, S.S. Wang, H. Liu, C. Shao, Y. Li, F. Link, S. Munker, R. Liebe, C. Meyer, E. Burgermeister, M. Ebert, S. Dooley, H. Ding, H. Weng, FOXA2 prevents

hyperbilirubinaemia in acute liver failure by maintaining apical MRP2 expression, *Gut* (2022) gutjnl-2022-326987. <https://doi.org/10.1136/gutjnl-2022-326987>

[51] J.S. Jakobsen, J. Waage, N. Rapin, H.C. Bisgaard, F.S. Larsen, B.T. Porse, Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries, *Genome Res* 23 (2013) 592-603. <https://doi.org/10.1101/gr.146399.112>

[52] H. Lu, J. Wan, R. Jiang, J. Xie, X. Peng, L. Zhang, Sodium butyrate potentiates carbon tetrachloride-induced acute liver injury in mice, *Toxicology Mechanisms and Methods* 22 (2012) 648-655. <https://doi.org/10.3109/15376516.2012.716091>

[53] Y. Sun, H. Pan, S. Shen, Z. Xia, Z. Yu, C. Li, P. Sun, C. Xin, Alisma Shugan Decoction (ASD) ameliorates hepatotoxicity and associated liver dysfunction by inhibiting oxidative stress and p65/Nrf2/JunD signaling dysregulation in vivo, *Medical science monitor: international medical journal of experimental and clinical research* 26 (2020) e921738-e921738. <https://doi.org/10.12659/MSM.921738>

[54] K.R. Zahid, S. Yao, A.R.R. Khan, U. Raza, D. Gou, mTOR/HDAC1 crosstalk mediated suppression of ADH1A and ALDH2 links alcohol metabolism to hepatocellular carcinoma onset and progression in silico, *Frontiers in oncology* 9 (2019) 1000. <https://doi.org/10.3389/fonc.2019.01000>

[55] D. Yang, Z. Lv, H. Zhang, B. Liu, H. Jiang, X. Tan, J. Lu, R. Baiyun, Z. Zhang, Activation of the Nrf2 signaling pathway involving KLF9 plays a critical role in allicin resisting against Arsenic Trioxide-induced hepatotoxicity in rats, *Biological Trace Element Research* 176 (2017) 192-200. <https://doi.org/10.1007/s12011-016-0821-1>

[56] H. Ye, C. Chen, H. Wu, K. Zheng, B. Martín-Adrados, E. Caparros, R. Francés, L.J. Nelson, M. Gómez del Moral, I. Asensio, J. Vaquero, R. Bañares, M.A. Ávila, R.J. Andrade, M. Isabel Lucena, M.L. Martínez-Chantar, H.L. Reeves, S. Masson, R.S. Blumberg, J. Gracia-Sancho, Y.A. Nevzorova, E. Martínez-Naves, F.J. Cubero, Genetic and pharmacological inhibition of XBP1 protects against APAP hepatotoxicity through the activation of autophagy, *Cell Death & Disease* 13 (2022) 143. <https://doi.org/10.1038/s41419-022-04580-8>

[57] Y. Zhou, X. Fan, T. Jiao, W. Li, P. Chen, Y. Jiang, J. Sun, Y. Chen, P. Chen, L. Guan, Y. Wen, M. Huang, H. Bi, SIRT6 as a key event linking P53 and NRF2 counteracts APAP-induced hepatotoxicity through inhibiting oxidative stress and promoting hepatocyte proliferation, *Acta Pharmaceutica Sinica B* 11 (2021) 89-99. <https://doi.org/10.1016/j.apsb.2020.06.016>

[58] Z. Azarmehr, N. Ranji, Z. Khazaei Koochpar, H. Habibollahi, The effect of N-Acetyl cysteine on the expression of Fxr (Nr1h4), LXRα (Nr1h3) and Sirt1 genes, oxidative stress, and apoptosis in the liver of rats exposed to different doses of cadmium, *Molecular Biology Reports* 48 (2021) 2533-2542. <https://doi.org/10.1007/s11033-021-06300-5>

Chapter 5 | General discussion

Drug side effects are related to drug safety and thus comprise an important study subject in pharmacology. The recently developed high throughput technologies (HTPs) provide a wide range of omics data and help to understand the biological mechanisms underlying drug side effects. However, the outcome of HTPs as omics data demands compatible bioinformatics analyses. The flourishing of effective omics data analyses could advance adverse drug effect research.

The prominence of omics data requires new approaches and efficient data analyses that can leverage the outcome of advanced technologies to explore cutting-edge discoveries. This thesis has focused on exploring different data analysis approaches as well as research angles to interpret omics data. We analyzed individual omics data types from epigenomics (MeDIP-seq, chapter 2), transcriptomics (RNA-seq, chapter 3) to proteomics (MS, chapter 3). Furthermore, we also not limited our analysis scale to single omics layers, but also pursued cross-omics interpretation. Chapter 2 demonstrated some inter-relations in which the gene expression could be influenced by DNA methylation alterations. Chapter 3-1 addressed the importance of both coding and non-coding genes on the transcriptional level. It offers a critical perspective on the limitation of only studying the proteome (chapter 3-2) or the expression of protein-coding genes at the transcriptome. Further, we have emphasized the analysis of the TF-target relations in chapter 4 via the RegOmics tool. This approach serves as an alternative to various cross-omics analyses using mathematical and statistical methods. Thus, all these bioinformatics analyses and tools can serve as introductory and accessible workflows to interpret omics data. They can aid researchers to understand biological mechanisms and direct further studies.

Alongside exploring new data analyses, this thesis also demonstrated how we utilized these contemporary approaches to analyze omics data in particular drug side effects as case studies. We specifically focused on cardiotoxicity and hepatotoxicity, which are always the top prevalent drug side effects. In the pharmaceutical development process, drug toxicity encountered 31% attrition of drug candidates, in particular, cardiotoxicity and hepatotoxicity are the most common toxicology attrition issues [1]. These two drug toxicities have also been responsible for the 14% and 18% post-marketing withdrawal, respectively [2]. Therefore, estimating the cardiotoxic and hepatotoxic potential of drugs always plays important role in drug optimization [3] and drug regulation [4]. In this thesis, we explored the cardiotoxicity of anthracyclines, including doxorubicin, idarubicin, and epirubicin (chapter 2, chapter 3), and the hepatotoxicity of rifampicin (chapter 4).

As a response to modern drug development and regulatory reformation, this thesis distinctively illuminated drug-induced alterations at different molecular levels. Recent movements have suggested the importance of understanding molecular mechanisms in drug adverse effects research. Consequently, drug regulations gradually emphasize the assimilation from molecular interactions (cellular level) to organismal physiology (organism level) [4]. In this thesis, we observed, detected, and discussed the changes in DNA methylation, gene, and protein expression as well as their inter-dependencies under drug exposure. This also reflects the shift of adverse drug events research toward systems pharmacology. Comprehensive databases, including drug-affected mechanisms as system pharmacology networks, could aid the clinical toxicities prediction based on the established associations between drug toxicities and drug toxic mechanisms [4]. Alongside the rise of omics technologies, new fields of pharmaceutical science have been transpired including pharmacometabolomics (metabolite levels), pharmacotranscriptomics (gene expression levels), and pharmacoproteomics (protein expression levels) [5]. Hence, the outcome of different omics analyses in this thesis not only additionally contributed toward pharmacoproteomics and pharmacotranscriptomics, but also posed a debate about pharmacoepigentic (epigenetic modifications) and its importance in drug adverse effect investigation. The results in chapters 2 and 4 have stated the necessity of assessing the DNA methylation status under drug treatment, in this case, ANTs and RIF. Recently, pharmacoepigentics, which studies the epigenetic basis of variation in response to drugs, and pharmacoepigenticomics, which is the application of pharmacoepigentics on a genome-wide scale, have been promoted as new sub-disciplines in pharmacology [6]. These inexperienced pharmaco-omic studies embrace a lot of challenges as well as growth potentials to explore drug action and drug side effects.

To demonstrate omics analyses in this thesis, we used data derived from the HeCaToS project. This project generated multi-omics data using different HTPs from the same *in vitro/in vivo* samples that offers a comprehensive and synchronized view of the biological system. Besides all HTPs usage, other innovations have also been utilized in this project. Physiologically based pharmacokinetic (PBPK) modeling can be used to individualize the drug doses based on patient-specific factors [4]. The reverse PBPK modeling used in the HeCaToS project attempted to profile the drug concentrations in targeted tissue after oral administrations and, subsequently, to offer *in vitro* settings closed to situations in the reality. However, replicating these drug profiles on weekdays for 2 weeks did not have biological replications and did not fully reflect the

therapeutic dose treatment in clinics. In particular, ANTs administration is generally by injection or an intravenous infusion. Although the administration process can be daily for 2 weeks followed by an off week to permit the body to recover and regenerate new cells; this is only one example among numerous time and administration sequences used according to the cancer type and its stage [7]. Other dose regimes suggest DOX's dosing once every 1 to 4 weeks [8]. RIF is usually taken by oral route, but similar to ANTs, its dose regime can vary. For instance, RIF can be used with a single or twice-daily dose(s) for 2-4 days or months, according to diseases and patient types [9]. Besides, the controls were done on a different date compared to the drug-treated samples which could create some batch effects. The HeCaToS project also attempted to employ *in vitro/in vivo* extrapolation by doing *in vitro* experiments as well as collecting patient biopsies to understand the side effect of common drugs. The initial aim is to help to benchmark and translate *in vitro* perturbed mechanisms to *in vivo* human situations, even though we had realized its difficulties after analyzing the data in this thesis (chapter 3).

To explore the hidden aspects of biological systems, we could set up new experiments and generate new data but we also can use available data and change the bioinformatics analyses. This thesis sought the unfamiliar perspectives of omics data analyses. While we still performed the popular approaches such as differential expression analysis (per time point and over time course) or co-expression analysis, we did tackle certain concealed elements. We upgraded the differential methylation analysis for the MeDIP-seq data to suggest strong candidate genes (chapter 2). In chapter 3, we unveiled the roles of lncRNAs along with the predominant focus on protein-coding genes; we also attempted to relate the proteome of microtissues in the *in vitro* setting to the proteome of human biopsies. Being proficient in different single omics approaches serves as a fundamental foundation to accomplish cross-omics analyses. After improving individual omics analysis, we also experimented to link the outcome, i.e. candidate genes/proteins, of each omics analysis with others. Chapter 4 focused on regulatory units, i.e. TF-target relations, with their members were nominated as potential candidates in individual omics data analysis (MeDIP-seq / RNA-seq / proteomics). These selected regulatory units can provide better systematic views of drug toxic mechanisms and better hypotheses for experimental testing. The TF-target analysis was performed by our tool called RegOmics. In fact, the shortcomings of the HeCaToS study design can introduce biases in the outcomes of the bioinformatics analyses as well as limit the biological interpretations. However, the scheme of bioinformatics analyses in this thesis can be widely

applied to other circumstances. While the study design and the quality of the data can be changed due to the development of technologies, adequate data analysis procedures could remain.

For future research on omic data analysis, we could endeavour different approaches to integrative multi-omics data. As mention in chapter 1, there is a broad spectrum of diverse methods: from using prior knowledge to data-driven methods or the combination between them [10,11]. Although some methods have been popularly used such as interaction network, machine learning, or composite network methods, we could consider other bottom-up simulation methods such as agent based modeling or game theory. These simulation methods can build complex systems based on the information about the actions and interactions of entities that are involved in the system. These methods have been applied for ecology, cell-cell commutation, or kinetic interactions [12-14], but are not often used in omics data analysis. Besides, notwithstanding the archived TF-target networks linking 2 omics layers (Figure 4-8, Figure 4-9), further investigation on sophisticated data visualization approach for network analysis is still in need for highlighting the involement of ≥ 3 omics layers, presenting ranked elements, and demonstrating the dose and time of exposure in study design.

In addition to new data analysis approaches, this thesis also deepens the knowledge about ANTs and RIF-induced mechanisms. Despite some limitations, the HeCaToS project provides comprehensive multiple-omics data derived from the same samples. We produced new insights in cardio- and hepatotoxicity as well as identified several potential biomarkers for drug toxicity. Different bioinformatics approaches have been invested and tested in this thesis; further experimental validation is the key to verifying the competencies of these bioinformatics approaches.

References

- [1] F.P. Guengerich, Mechanisms of drug toxicity and relevance to pharmaceutical development, *Drug metabolism and pharmacokinetics* 26 (2011) 3-14. <https://doi.org/10.2133/dmpk.dmpk-10-rv-062>
- [2] I.J. Onakpoya, C.J. Heneghan, J.K. Aronson, Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature, *BMC Med* 14 (2016) 10. <https://doi.org/10.1186/s12916-016-0553-2>
- [3] I. Yusof, F. Shah, T. Hashimoto, M.D. Segall, N. Greene, Finding the rules for successful drug optimisation, *Drug Discovery Today* 19 (2014) 680-687. <https://doi.org/10.1016/j.drudis.2014.01.005>
- [4] S.-M. Huang, D.R. Abernethy, Y. Wang, P. Zhao, I. Zineh, The Utility of Modeling and Simulation in Drug Development and Regulatory Review, *Journal of Pharmaceutical Sciences* 102 (2013) 2912-2923. <https://doi.org/10.1002/jps.23570>
- [5] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Research* 46 (2017) D1074-D1082. <https://doi.org/10.1093/nar/gkx1037>
- [6] J. Peedicayil, Pharmacoeugenetics and Pharmacoeugenomics: An Overview, *Current drug discovery technologies* 16 (2019) 392-399. <https://doi.org/10.2174/1570163815666180419154633>
- [7] V. Priyanka, K. Anup, Anthracyclines, in: StatPearls Publishing, Treasure Island (FL), StatPearls [Internet], 2022.
- [8] D. Steven, P.C. Michael, Anthracycline Medications (Doxorubicin), in: Treasure Island (FL): StatPearls Publishing, StatPearls [Internet], 2021.
- [9] Mylan, Rifampicin 150 mg Capsules, in: Electronic Medicines Compendium (EMC), [Internet] <https://www.medicines.org.uk/>, 2021.
- [10] Y. Hasin, M. Seldin, A. Lusic, Multi-omics approaches to disease, *Genome Biology* 18 (2017) 83. <https://doi.org/10.1186/s13059-017-1215-1>
- [11] M.A. Wörheide, J. Krumsiek, G. Kastenmüller, M. Arnold, Multi-omics integration in biomedical research – A metabolomics-centric review, *Analytica Chimica Acta* 1141 (2021) 144-162. <https://doi.org/10.1016/j.aca.2020.10.038>
- [12] M. Archetti, K.J. Pienta, Cooperation among cancer cells: applying game theory to cancer, *Nature Reviews Cancer* 19 (2019) 110-117. <https://doi.org/10.1038/s41568-018-0083-7>
- [13] B. Zhang, D.L. DeAngelis, An overview of agent-based models in plant biology and ecology, *Annals of Botany* 126 (2020) 539-557. <https://doi.org/10.1093/aob/mcaa043>
- [14] S. Maestri, E. Merelli, M. Pettini, Agent-based models for detecting the driving forces of biomolecular interactions, *Scientific Reports* 12 (2022) 1878. <https://doi.org/10.1038/s41598-021-04205-8>

Impact paragraph

As aforementioned in chapter 5 – general discussion, this thesis firstly provides a brief of omics technologies currently available for drug side effect research and then explores different omics analysis approaches as well as the drugs-induced molecular alterations inside the cell. Our main aim is to form different data analyses, while the main theme of this thesis is seated within the drug side effects field because understanding the negative effects of drugs is essential for drug safety. Furthermore, exploring the drug toxic mechanism of known drugs could support the clinical usage of those drugs as well as aid the drug discovery process [1].

This thesis focused on the toxic mechanisms of anthracyclines (ANTs), including 3 common analogs: doxorubicin (DOX), epirubicin (EPI), and idarubicin (IDA), and rifampicin (RIF). ANTs are essential chemotherapeutic agents, but their adverse effects can lead to heart failure in cancer survivors. RIF is an important antibiotic for tuberculosis but can cause liver injury. Despite their adverse effects, these drugs are still popular today [2-4]. Although different research had tried to draft the toxic mechanisms of these drugs, their paradigms remain incomplete. In this thesis, we detected several potential targets and phenomena such as altered DNA methylated genes, altered genes, lncRNAs, and proteins' expression under these drug treatments. All these outcomes can be immediately used by risk assessors to evaluate the ANTs-induced cardiotoxicity and RIF-induced hepatotoxicity. These potential targets can be candidate biomarkers to diagnose the particular drug's side effects and help to understand the drug's toxic mechanism. Especially, while many studies mainly focused on the proteome and protein-coding genes, this thesis emphasized the other aspects of the biological system such as DNA methylation and lncRNAs. Epigenetic modifications and lncRNAs have been implicated in cellular processes, and can also be useful tools to reveal drug-induced adverse side effects.

The analysis approaches in this thesis are publicly available, a part of them has been published, and can already be used by other researchers to improvise and analyze the omics data. The usage of high throughput technologies resulting in omics data has been widely accepted in drug toxicity studies. It aids scientists to explore cutting-edge discoveries and promotes reforming the toxicity field. However, the rational recognition of modern omics technologies requires effective and transparent data analysis approaches. This urgent need has appeared not only in the academic research community but also among regulatory agencies and in the pharmaceutical industry [5]. This thesis portrays different analysis approaches for various omics data types and thus enriches the omics analysis toolbox. Our

omics analyses have been published and are accessible to a broad range of users. This also supports the appreciation of omics data and promotes the transparency of omics analysis. Furthermore, it ties to other movements to contribute to the use of omics data in drug development and drug regulatory.

Overall, this thesis discloses several points of interest about both bioinformatics and drug adverse effects perspectives. We displayed the substantial changes in DNA methylation, transcriptome, and proteome as well as parts of the molecular inter-dependences, specifically TF-target relations, beyond the limitation of the data and the need for findings validation. The established omics analysis approaches could take advantage of the modern omics technologies to demonstrate the molecular toxic mechanisms and head toward the advanced toxicology field.

References

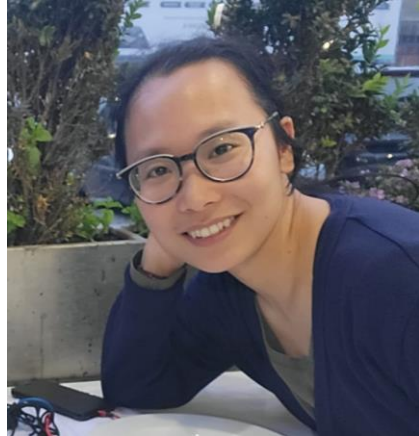
- [1] I. Yusof, F. Shah, T. Hashimoto, M.D. Segall, N. Greene, Finding the rules for successful drug optimisation, *Drug Discovery Today* 19 (2014) 680-687. <https://doi.org/10.1016/j.drudis.2014.01.005>
- [2] W.H.O. WHO, World Health Organization model list of essential medicines for children: 7th list 2019, in: World Health Organization, 2019.
- [3] W.H.O. WHO, World Health Organization model list of essential medicines: 21st list 2019, in: World Health Organization, 2019.
- [4] R. Prasad, N. Gupta, A. Banka, Multidrug-resistant tuberculosis/rifampicin-resistant tuberculosis: Principles of management, *Lung India : official organ of Indian Chest Society* 35 (2018) 78-81. https://doi.org/10.4103/lungindia.lungindia_98_17
- [5] S.-M. Huang, D.R. Abernethy, Y. Wang, P. Zhao, I. Zineh, The Utility of Modeling and Simulation in Drug Development and Regulatory Review, *Journal of Pharmaceutical Sciences* 102 (2013) 2912-2923. <https://doi.org/10.1002/jps.23570>

Addendum

About the author

Nhan Nguyen (Nguyễn Thị Nhân) was born on December 14th, 1993 in Ninh Son village, Chuong My district, Ha Tay province (later it merged into Hanoi), Vietnam.

In 2008, she entered the high school belonged to Hanoi University of Science, Vietnam National University with biology specializations. In 2011, Nhan joined the honor program for the bachelor's in Biology at Vietnam National University, where she studied different biological disciplines from biochemistry and molecular biology to ecology. Nhan graduated in 2015 and received the Excellent Achievement Award for her performance during her study.



In September 2016, Nhan started her master in systems biology at Maastricht University, thanks to the Maastricht University (UM) Holland-High Potential scholarship. In this master's program, she learned about programming and the application of mathematical and computational models to biological science. She did her master internship at MultiModal Molecular Imaging Institute (M4I) and received her master's degree in June 2018.

In July 2018, Nhan started working as a PhD student at the Department of Toxicogenomics, Maastricht University. Her PhD focused on analyzing the omics data from the HeCaToS (Hepatic and Cardiac Toxicity Systems modelling) project. During her PhD, Nhan was involved in different teaching activities as a tutor, writing assessor, and mentor.

From August 2022, Nhan works as a postdoc at the Helmholtz Centre for Infections Research (HZI), Hannover, Germany. Her project focuses on analyzing multiple omics data related to vaccine response in humans.

Extracurricular activities

Women in Data Science (WiDS) Datathon, Maastricht (2020, 2021, 2022)

"Inspire high school kids!" program, Maastricht University (2020 - 2022)

PhD representative, GROW School, Maastricht University (2018 - 2020)

Premium project - Honours programme for master students (2018)

Certificates

The University Teaching Qualification (UTQ), 2022.

The Postgraduate Education in Toxicology (PET) on Toxicogenomics, 2019.

Contacts

ORCID: [0000-0001-8720-1195](https://orcid.org/0000-0001-8720-1195)



List of publications

Scientific publication

Nhan Nguyen, Danyel Jennen, and Jos Kleinjans. "Omics technologies to understand drug toxicity mechanisms." *Drug Discovery Today* (2022): 103348. <https://doi.org/10.1016/j.drudis.2022.103348>

Nhan Nguyen, Matthias Lienhard, Ralf Herwig, Jos Kleinjans, and Danyel Jennen. "Epirubicin alters DNA methylation profiles related to cardiotoxicity". *Frontiers in Bioscience - Landmark* 27.6 (2022): 173. <https://doi.org/10.31083/j.fbl2706173>

Nhan Nguyen, Matthias Lienhard, Ralf Herwig, Jos Kleinjans, and Danyel Jennen. "A bioinformatics workflow to detect genes with DNA methylation alterations: a case study of analyzing MeDIP-seq data in cardiac microtissue exposed to epirubicin". *International Conference Proceedings by ACM* (2022): 63-69. <https://doi.org/10.1145/3510427.3510437>

Nhan Nguyen, Terezinha Souza, Jos Kleinjans, and Danyel Jennen. "Transcriptome analysis of long noncoding RNAs reveals their potential roles in anthracycline-induced cardiotoxicity." *Non-coding RNA Research* 7 (2022): 106-113. <https://doi.org/10.1016/j.ncrna.2022.01.002>

Nhan Nguyen, Terezinha Souza, Marcha CT Verheijen, Hans Gmuender, Nathalie Selevsek, Ralph Schlapbach, Jos Kleinjans, and Danyel Jennen. "Translational proteomics analysis of anthracycline-induced cardiotoxicity from cardiac microtissues to human heart biopsies." *Frontiers in Genetics* 12 (2021): 1018. <https://doi.org/10.3389/fgene.2021.695625>

Conference presentation

Related to this thesis

"RegmuNet: Interactive regulatory network based on prior knowledge and multiple omics data integration". Poster presentation at the 61st the Society of Toxicology (SOT) annual meeting, the United States, 2022.

"A bioinformatics workflow to detect genes with DNA methylation alterations: a case study of analyzing MeDIP-seq data in cardiac microtissue exposed to epirubicin". Oral presentation at the 12th International Conference on Bioscience, Biochemistry and Bioinformatics (ICBBB), Japan, 2022.

“Epirubicin alters the DNA methylation in cardiac microtissue”. Poster presentation at the 56th (virtual) Congress of the European Societies of Toxicology (EUROTOX), 2021.

“Transcriptome analysis of long noncoding RNAs reveals their potential roles in anthracycline-induced cardiotoxicity”. Poster presentation at the 11th (virtual) World Congress on Alternatives and Animal Use in the Life Sciences, 2021.

“Analysing DNA methylation alterations in cardiac microtissue exposed to epirubicin”. Poster presentation at the Netherlands Bioinformatics and Systems Biology (BioSB) conference, 2021.

“Systems toxicology II: lncRNA – a new player”. Oral presentation at the Postgraduate Education in Toxicology (PET) on Toxicogenomics, 2021.

“Anthracycline-induced cardiotoxicity: in vitro and in vivo proteomics investigations”. Poster presentation at the Netherlands Bioinformatics and Systems Biology (BioSB) conference, 2020.

Previous works

“A cell culture protocol for ^{13}C -Metabolic Flux Analysis on a microfluidics chip”. Master thesis presentation at the MaCSBio Science Day 2018.

“The LC_{50} evaluation of lead to *Danio rerio* in the embryo and larval stages”. Abstract at the sixth National Scientific Conference on Ecology and Biological Resources, Vietnam, 2015.

Acknowledgments

I would like to thank my supervision team, **Prof. dr. Jos Kleainjans** and **Dr. Danyel Jennen**, for their guidance over these four years. **Jos**, thank you for giving me the opportunity to do my PhD. Your thoughtful and pragmatic suggestions as well as your prompt responses helped me to move forward with my projects. My thesis can be finished on time thank to you. **Danyel**, thank you very much for giving me not only the freedom to do research but also challenges to grow; they made me become a more independent researcher. Your eyes to details always assure documents with good format and illustration.

Throughout these years, many scholars have given me advices and invaluable supports. I would like to thank **Terezinha**, who helped me to establish ideas and research directions at the beginning of my PhD, and **Yanick**, who generated the sequencing data which became a major part of my PhD work. I also would like to thank **Florian** and **Twan** with all of your critical questions and helpful suggestions. I learnt to become a better researcher from it. **Marcha**, **Tim** and **Maria**, thank you very much for all of your advices and aids which gave me different perspectives and considerations.

I would like to thank all of the former and current TGX staffs for your care, understanding and companionship. A special thank to **Christa** and **Rob**, you always welcome and help me with all paperwork and administration questions. **Simone**, **Jacco**, **Marcel**, **Theo**, **Rossella**, and **Ducan**, I enjoy the conversation with you a lot. **Bert**, after my master's program, it is a very nice surprise to be in the same department with you.

To all – especially, not mention yet - former and current PhD colleagues, it is lovely and delightful to know you and to spend time with you. Being with you brings not only joys but also shared moments to overcome hurdles. Special thanks to my “former” officemates: **Daniela** - we share a lot of time together, you are always positive and enjoying life; **Julia** – it is always nice to listen and see when you are talking. To my “current” officemates: **Jelmer** – you introduced me to cool kids' jargons and jokes; together with **Kaidi** and **Qingfeng**, we made our office becomes a foodies office. **Juan** – you are a special person in every aspects, and for sure, with a good heart. **Toni** and **Nikki**, your guys introduced me to all “expensive” stuffs and *coffee-made-by-people* ☺. To **Rajinder**, **Marta**, **Manon**, **Nicolaj**, **Somaieh**, **Milena**, **Na**, **Yueqin**, **Zhiling**, **Jiang**, and **Julian** thank you for all the chats and the good time we spent together.

I also would like to thank all other colleagues, I wish you all the best at work and in the future.

During my PhD, I had joint in different activities and committees. Thus, I would like to thank the **GROW office**, **PhD coordinators**, and **PhD representatives**. I also would like to thank the **Portfolio planning group** with their International Track of Medicine program, the **Team Schoolcontacten** with their "Inspire high school kids!" program, and of course, the **PhD dance community**. It was a pleasure to get involved in your activities and events.

Ik wil mijn Nederlandse vrienden bedanken, door jullie heb ik een tweede huis in Eijsden en Maastricht. **Dientje** en **Guy**, ik voel me altijd thuis in het huis van jullie; we deden veel dingen samen: bakken, koken, kletsen over van alles. Ik ga de reis naar Eijsden zeker elke zaterdag missen. **Cyrils** en **Gerry**, bedankt voor jullie hulp en steun; jullie hebt me veel geleerd over Nederland en het leven in Maastricht. **Hub**, mijn judoleraar en mijn vriend, bedankt voor je geduld en je vriendelijkheid; we hebben altijd een gezellige tijd samen met lekker eten en goede drankjes. Ook wil ik **Judoclub Tsukuri Maastricht** bedanken; het is altijd fijn om bij jullie te zijn.

*In English: "I would like to thank my Dutch friends, who made Eijsden and Maastricht becomes my seconded home. **Dientje** and **Guy**, I always feel like home at your place; we did a lot of thing together from baking, cooking, to discussing everything. I will, for sure, miss the trip to Eijsden every Saturday. **Cyrils** and **Gerry**, thank you for your help and support; you taught me a lot of information about the Netherland and the life in Maastricht. **Hub** - you are my judo teacher and my friend; thank you for your patience and your kindness; we have always a good time together with good food and good drinks. I also would like to thank the **Judoclub Tsukuri Maastricht**; it is always nice to be with you."*

Em/mình xin cảm ơn những người anh/chị/bạn/em Việt Nam: **VietSoc Maastricht - Hội Sinh viên Việt Nam ở Maastricht**, **SIVILI - Hội Sinh viên Việt Mam ở Liege**, và **Hội lưu học sinh Việt Nam 1**. Cuộc sống ở Hà Lan và quá trình học tập của em/mình không thể trọn vẹn nếu không có mọi người. Mình cũng muốn cảm ơn đến các bạn cấp 2, cấp 3 và bạn đại học của mình. Thời gian và khoảng cách địa lý không là trở ngại với chúng ta.

*In English: "I would like to thanks all of my Vietnamese friends: **VietSoc Maastricht group**, **SIVILI - Vietnamese students at Liege group**, and **Overseas Vietnamese student branch 1**. My Dutch and PhD journey could not be complete without you. I also would like to thank all of my Vietnamese friends from my middle school, high school, and bachelor program. Time and physical distance are not an obstacle to our friendship."*

Lời cuối cùng, con xin cảm ơn bố mẹ, em xin cảm ơn anh chị cùng các cháu của cô.

In English: "Finally, I would like to thank my family: my parents, my brother, my sister-in-law, and my niblings."