PhD-FSTC-2019-42
The Faculty of Sciences, Technology and Communication

The Faculty of Health, Medicine and Life Sciences

DISSERTATION

Defence held on 30/08/2019 in Maastricht, Netherlands

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN *BIOLOGIE*

AND

DOCTOR AT MAASTRICHT UNIVERSITY

by

Muhammad ALI
Born on 16 June 1989 in Gujranwala (Pakistan)

INTEGRATIVE NETWORK-BASED APPROACHES FOR MODELING HUMAN DISEASE

Dissertation defence committee

Prof. Dr. Antonio del Sol Mesa, dissertation supervisor, *Professor, Université du Luxembourg*

Prof. Dr. Jos Kleinjans, dissertation supervisor, *Professor, Maastricht University, Netherlands*

Prof. Dr. Jos Prickaerts, Chairman, *Professor, Maastricht University, Netherlands*

Prof. Dr. Jens Schwamborn, Vice Chairman, *Professor, Université du Luxembourg*

Prof. Dr. Jonathan Mill, *Professor, University of Exeter Medical School, United Kingdom*

Dr. Tim Vanmierlo, *University of Hasselt, Belgium*

INTEGRATIVE NETWORK-BASED APPROACHES FOR MODELING HUMAN DISEASE

Dissertation

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof. dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public on:
Friday 30 August 2019, at 10:00 hrs.

by

Muhammad **Ali**

This thesis was written under a joint degree program supervised by
Maastricht University and the University of Luxembourg

Supervisor

Prof. Dr. Jos Kleinjans

Co-supervisors

Dr. Daniel van den Hove
Dr. Ehsan Pishva

Assessment Committee

Prof. Dr. Jos Prickaerts (Chair, Maastricht University)
Prof. Dr. Jens Christian Schwamborn (University of Luxembourg)
Prof. Dr. Jorge Goncalves (University of Luxembourg)
Prof. Dr. Jonathan Mill (University of Exeter Medical School, United Kingdom)
Dr. Alfredo Ramirez (University of Bonn, Germany)
Dr. Tim Vanmierlo (Maastricht University)

This thesis was written under a joint degree program supervised by
Maastricht University and the University of Luxembourg

Declaration

I hereby declare that this dissertation has been written only by the undersigned and without any assistance from third parties. Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated herein.

Muhammad Ali,
Esch-sur-Alzette, Luxembourg
22 August 2019

Acknowledgements

Without the support of a number of people and institutions, it would not have been possible to write this dissertation. It is my pleasure to have the opportunity to express my gratitude to some of them here.

For my academic achievements, I would like to acknowledge my supervisor, Prof. Dr. Antonio del Sol for the opportunity to join his team and his constant support and guidance. I can say with certainty that four years of regular brain storming and critical scientific discussions constituted an invaluable training period. I would also like to acknowledge the help and support of my co-supervisor Prof. Dr. Jos Kleinjans. Their suggestions have helped to define my training as a Ph.D. student and the resulting Ph.D. thesis.

Especially, I am thankful to Prof. Dr. Daniel L.A. van den Hove for his constant and indispensable support as an advisor for my thesis work. His encouragement has been the biggest contribution towards the success in the execution of this thesis work.

I am grateful to Dr. Sascha Jung and Dr. Ehsan Pishva for their intellectual guidance and extending all the support despite their busy schedule. Few would have been possible without their scientific and moral support.

I would like to thank Prof. Jens C. Schwamborn and Prof. Jorge Goncalves for agreeing to take part in my CET and thesis review committees. Their valuables comments and suggestions have helped me a lot in shaping up my thesis and critically viewing my own research work.

I am especially grateful to the paranymphs Ghazi Al Jowf, Kyonghwan Choe, and Martin Bustelo, as well as the former and current members of the Computational Biology Group at LCSB and MHeNS at Maastricht University, for their enthusiasm, support, and making this an unforgettable journey.

I would also like to thank FNR Luxembourg, and JPND Maastricht, for funding and hosting me for four years. Furthermore, the help and support offered by the EPIAD consortium and the secretaries Ms. Nicole Senden and Ms. Christine Marszalek also deserves to be acknowledged.

On a social note, I would like to thank my family and friends, for their emotional support. Without their constant support, I would not have been able to achieve this milestone in my life.

Preface

Before you lies the dissertation "Integrative Network-Based Approaches For Modeling Human Disease", the basis of which is a reasearch work conducted throughout the duration of Ph.D. It has been written to fulfill the graduation requirements of Doctor of Biology at the University of Luxembourg and Maastricht University. I was engaged in researching and writing this dissertation from January to April 2019.

The conducted projects were undertaken under the supervision of Prof. Dr. Antonio del sol and Prof. Dr. Jos Kleinjans together with Dr. Daniel L.A. van den Hove. The addressed research questions were formulated together with my supervisor and co-supervisor. Different projects had different level of difficulties, but conducting extensive literature-review and thorough analyses have allowed me to answer the important biological questions described in this thesis. Fortunately, both of my supervisors, were always available and willing to answer my queries.

I would like to thank my supervisors for their excellent guidance and support during this process. I also wish to thank all of the advisers, without whose cooperation I would not have been able to conduct these analyses.

To my other colleagues at Computational Biology Group at Luxembourg University and Department of Psychiatry and Neuropsychology at Maastricht University: I would like to thank you for your wonderful cooperation as well. It was always helpful to bat ideas about my research around with you. I also benefitted from debating issues with my friends and family. If I ever lost interest, you kept me motivated. My family (especially my mother) and friends (Rehan, Raza, Adnan) deserve a particular note of thanks: your wise counsel and kind words have, as always, served me well.

I hope you enjoy your reading.

Abstract

The large-scale development of high-throughput sequencing technologies has allowed the generation of reliable omics data related to various regulatory levels. Moreover, integrative computational modeling has enabled the disentangling of a complex interplay between these interconnected levels of regulation by interpreting concomitant large quantities of biomedical information ('big data') in a systematic way. In the context of human disorders, network modeling of complex gene-gene interactions has been successfully used for understanding disease-related dysregulation and for predicting novel drug targets to revert the diseased phenotype.

Recent evidence suggests that changes at multiple levels of genomic regulation are responsible for the development and course of multifactorial diseases. Although existing computational approaches have been able to explain cell-type-specific and disease-associated transcriptional regulation, they so far have been unable to utilize available epigenetic data for systematically dissecting underlying disease mechanisms.

In this thesis, we first provided an overview of recent advances in the field of computational modeling of cellular systems, its major strengths and limitations. Next, we highlighted various computational approaches that integrate information from different regulatory levels to understand mechanisms behind the onset and progression of multifactorial disorders. For example, we presented INTREGNET, a computational method for systematically identifying minimal sets of transcription factors (TFs) that can induce desired cellular transitions with increased efficiency. As such, INTREGNET can guide experimental attempts for achieving effective *in vivo* cellular transitions by overcoming epigenetic barriers restricting the cellular differentiation potential. Furthermore, we introduced an integrative network-based approach for ranking Alzheimer's disease (AD)-associated functional genetic and epigenetic variation. The proposed approach explains how genetic and epigenetic variation can induce expression changes via gene-gene interactions, thus allowing for a systematic dissection of mechanisms underlying the onset and progression of multifactorial diseases like AD at a multi-omics level. We also showed that particular pathways, such as sphingolipids (SL) function, are significantly dysregulated in AD. In-depth integrative analysis of these SL-related genes reveals their potential as biomarkers and for SL-targeted drug development for AD. Similarly, in order to understand the functional consequences of *CLN3* gene mutation in

Batten disease (BD), we conducted a differential gene regulatory network (GRN)-based analysis of transcriptomic data obtained from an *in vitro* BD model and revealed key regulators maintaining the disease phenotype.

We believe that the work conducted in this thesis provides the scientific community with a valuable resource to understand the underlying mechanism of multifactorial diseases from an integrative point of view, helping in their early diagnosis as well as in designing potential therapeutic treatments.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

The remarkable development of high-throughput sequencing technologies has enabled the collection of a variety of "omic" modalities for various human diseases, generated at the whole genome-level, including genomic, epigenomic, transcriptomic, proteomic and metabolomic data. Computational analysis of such datasets has provided compelling evidence for various genetic, epigenetic and transcriptional changes to be associated with the onset and progression of various human disorders [102, 306]. Owing to the multifactorial nature of most of these disorders, recent advancements in computational disease modeling, by integrating regulatory information from different levels, provide a new framework for understanding the complex nature of human health and disease. For example, modelling of complex gene interaction networks has been very useful for disease modelling [143, 13, 182] and for disentangling the interplay between different regulatory layers [193, 93, 195]. These regulatory network models constitute the starting point for the identification of key regulatory circuits and motifs within large-scale interaction datasets built from genome-wide gene expression profiling, corresponding to the most influential interactions determining network stability, or triggering disease progression or differentiation [143, 13, 303, 197]. However, integrative network modelling approaches –i.e. linking different regulatory layers– [193, 93, 195, 104] are still scarce, which hampers the possibility of studying the crosstalk established among regulatory layers for determining a given phenotype or mediating phenotypic transitions [73]. As such, developing tailor-made computational models is a crucial step in understanding the contributions of genomic, epigenomic, and transcriptomic landscapes in cellular circuitry, lineage specification, and the onset and progression of human disease.

1.1 *In vitro* applications of computational disease modeling

The startling breakthrough of obtaining induced pluripotent stem cells (iPSCs) from differentiated fibroblasts by over-expressing a set of transcription factors (TFs) –usually referred to as cellular reprogramming– laid the foundation of *in vitro* human disease modeling and downstream applications [70]. iPSCs-based disease models have allowed the generation of patient-specific differentiated cell types, overcoming the gap between studies using animal-based disease models and pre-clinical therapeutic research [268, 256]. This disease-in-a-dish technology has provided new avenues for understanding functional dysregulation associated with diseases, discovering disease-related genes and promoting personalized medicine [140, 48]. Beside understanding disease mechanisms, these iPSCs-based disease models can be used for drug screening, in order to mitigate disease phenotypes by targeting particular pathological molecular mechanisms identified by analysing these models [268]. Owing to the complications in obtaining specialized cell types and tissue samples for experimental studies [97], researchers are relying on using iPSCs to generate more representative models for studying human disease. For example, a schematic illustration of generating an iPSCs-based neurological disease model is shown in Figure 1, where patient-specific differentiated cell types are obtained from somatic cells of a patient.

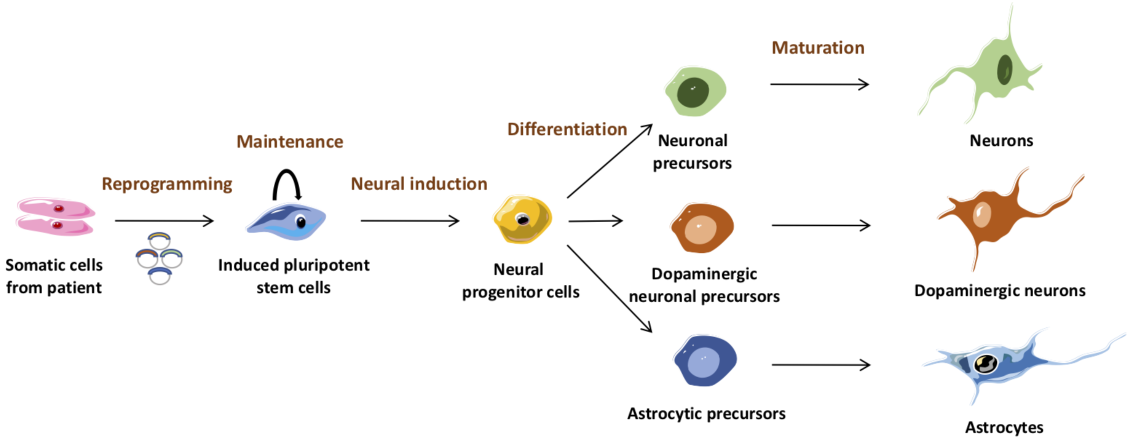


Figure 1: **Workflow of *in vitro* iPSCs-based disease modeling**

The traditional workflow of generating iPSCs-based disease models by reprogramming patient-specific somatic cells poses significant challenges in terms of time and resources [210, 305]. Similar to reprogramming, where one wants to differentiate iPSCs into a particular lineage and a specific mature cell type, trans-differentiation aims at obtaining the same cell type of interest without undergoing an intermediate pluripotent state. For example, researchers have been able to successfully achieve a directed conversion of human dermal fibroblasts into cardiac progenitors by over-expressing the TFs *ETS2* and *MESP1* [118], contributing to the paradigm of regenerative medicine for treating cardiovascular diseases. Although directed cellular conversions dramatically reduced the time required for obtaining a specific cellular disease model, the identification of efficient TFs, i.e. to achieve a successful conversion, remained a trial-and-error experimental process, limiting its utility and applicability. To this end, various computational methods have been developed to speed up this process by utilizing transcriptomic data sets and systematically predicting candidate TFs that can convert one fully differentiated cell type into another [59, 232, 198]. However, despite these developments, limited cellular conversion efficiency still represents a major problem that has not yet been solved by these methods, limiting the application of this technique in regenerative medicine.

Experimental evidence suggests that only including information on transcription, i.e. expression profiles, is insufficient for identifying a suitable set of TFs that can produce efficient cellular conversion, as it is the interplay of epigenetic and transcriptional regulation that mediates cellular conversion [191, 137, 244]. Dysregulation at these regulatory levels has been found to disrupt physiological cellular differentiation and lies at the core of many disorders [161, 295], requiring an *ex vivo* or *in vitro* application for the development of novel treatment strategies. For example, mesenchymal stem cells (MSCs) represent a rare stem cell type whose *in vitro* expansion is vital for obtaining sufficient amounts of cells for treating various heart- [5, 174, 226, 123], brain- [122, 178, 162] and wound healing- [300, 301] related disorders. However, progressive spontaneous differentiation and aging of MSCs may occur during expansion, both of which can be modulated by extrinsic epigenetic signals such as histone H3 acetylation, playing a key role in regulating these intricate processes [161]. Similarly, epigenetic mechanisms have been found to be crucial for regulating B-cell maturation and their dysregulation has been associated to the initiation and acceleration of multiple autoimmune diseases such as systemic lupus erythematosus [296, 297, 298] and rheumatoid arthritis [184, 95, 184]. Taken together, this evidence suggests that epigenetic mechanisms, along with other regulatory layers, play a crucial role in normal cellular differentiation processes. As such, generating computational disease models by integrating epigenetic and transcriptomic information can provide deeper insights into the underlying mechanisms, allowing us to predict specific external stimuli (e.g. TF over-expression or compound-based induction) that can overcome the epigenetic barriers restricting the differentiation potential of cells in different disorders.

1.2 Reconstruction of integrative cell-type-specific network models

Modeling human diseases by network-based approaches demands the reconstruction of reliable network models that are context-specific and explain the regulation of gene expression program. It has become increasingly clear that it is the cross-talk between transcriptomic and epigenetic layers that regulates gene expression programs across various human cell types [53, 274, 41]. In addition, epigenetic mechanisms, such as CpG DNA methylation [238, 164], histone modifications [67, 44] and chromatin accessibility [203, 68] have been shown to be an important factor in controlling and predicting the variability of gene expression signatures across different cell and tissue types. A few methods that acknowledge the importance of these different, but interconnected layers of regulation in controlling gene expression programs exist, all suggesting that integrating information from both layers to generate more precise network models of human cell and tissue types is the way forward [204, 245, 113, 175, 64]. Most of these methods rely on the integration of active enhancer information with transcriptomic profiles and position weight matrix (PWM)-based TF-binding predictions to link regulators with their target genes.

Although existing integrative methods for reconstructing network models for different cell and tissue types provide meaningful insights for understanding the underlying mechanisms of gene regulation, these approaches suffer from some important limitations. Foremost, these methodologies usually rely on histone modification marks for active enhancer identification (H3K27ac) to predict active enhancer regions and associate them to their target genes based on ad hoc criteria, such as the nearest gene or all genes within a defined range. Such enhancer annotations might lead to the inference of false-positive (and -negative) interactions as it has been widely known and also experimentally verified that enhancers do not necessarily act on the closest promoter, but may also bypass neighbouring genes to act on more distant genes along the same as well as a different chromosome [100, 109]. Secondly, these approaches rely on PWM-based predictions of TF bindings in regulatory regions to associate regulator TFs with their respective target genes.

Such PWM-based predictions might lead to the inference of many false-positive interactions due to the detection of false-positive motifs, as indicated by existing studies [313, 163]. Lastly, these methods lack systematic benchmarking of predictive network models against experimental cell-type-specific TF chromatin immunoprecipitation (ChIP) sequencing (Seq) data. These limitations suggest that there is still a need for a more sophisticated integrative computational method that relies only on experimental data from different regulatory levels to reconstruct reliable context-specific networks. Furthermore, systematic benchmarking of these reconstructed networks should be carried out to prove their context-specificity. Moreover, the application of such tailor-made integrative network models is yet to be explored in the context of predicting combinations of TFs that could produce highly efficient cellular conversions between two cell types of interest.

1.3 Network-based modeling in Alzheimer's disease

Variations at multiple levels of genomic regulation, including genetic aberrations (e.g. single nucleotide polymorphisms [SNPs]), epigenetic (e.g. DNA methylation) and gene expression changes, are known to be associated with various human diseases, including Alzheimer's disease (AD). Many studies exist that use information from an individual regulatory level to identify causal genes and understand the mechanisms underlying the pathogenesis of AD. For example, genome-wide association studies (GWAS) have successfully identified numerous susceptibility genes for AD [89, 286]. Similarly, a crucial role for changes in DNA methylation [290, 61] and gene expression levels [286, 121] has been observed in AD patients. Nevertheless, the heterogeneous and multifactorial nature of AD demands the integration of regulatory information from different omic levels in order to adequately capture the mechanisms underlying the onset and progression of this disease. Yet again, systematic analytical approaches for identifying multi-omics AD biomarkers to prioritize key genes are still scarce.

Apart from genome-wide hypothesis-generating approaches assessing different layers of regulation in an integrative fashion, similar multi-omics approaches might also be useful in studying existing hypothesis on the pathogenesis of AD, e.g. those on sphingolipid [190] and the tryptophan-kynurenine pathways [293, 98]. As such, an in-depth integrative analysis of genes involved in such pathways can help identifying causal genes, as well as generate testable hypotheses from analysed changes in associated gene expression and DNA methylation signatures. Such analyses have the potential to provide novel biomarkers and druggable targets in AD, and propose new disease modifying agents that can help in slowing down the progression or reverting the disease phenotype.

Taken together, recent evidence have suggested that changes at multiple levels of genomic regulation are responsible for the development and course of multifactorial diseases. Although existing computational approaches have been able to explain cell-type-specific and disease-associated transcriptional regulation, they so far have been unable to utilize available epigenetic and transcriptomic data for systematically dissecting underlying disease mechanisms. In order to bridge this gap in the literature, we have presented various computational approaches in this thesis that integrate information from different regulatory levels to understand mechanisms behind the onset and progression of multifactorial disorders. Thus, helping in their early diagnosis as well as providing avenues for designing more effective therapeutic treatment strategies.

1.4 Thesis outline

The research conducted in this thesis can be divided into five parts. CHAPTER 2 constitutes a concise overview of existing computational methods in the field of systems biology. Particular attention is paid to state-of-the-art gene regulatory network (GRN) based methods for instructive factors determination and human disease modeling. Along with the strengths, the limitations of these methods are highlighted, thereby providing avenues for the research conducted and described in the following chapters.

Owing to the limited cellular conversion efficiency and lack of integrative methods for predicting more efficient sets of instructive factors, CHAPTER 3 describes INTREGNET, an integrative computational method for systematically identifying reliable minimal sets of TFs that can induce desired cellular conversions with increased efficiency. The application of this method is demonstrated in an *in vitro* setting, where limited conversion efficiency is a crucial barrier for its application in regenerative medicine.

As explained above, the heterogeneous and multifactorial nature of AD requires the integration of regulatory information from different -omics levels in order to capture the underlying mechanisms behind the onset and progression of this disease. In CHAPTER 4, global multi-omics alterations in AD patients are identified by comparing genomic (gene aberration), epigenomic (DNA methylation) and transcriptomic data sets of 46 diseased patients with 32 age-matched controls.

CHAPTER 5 features an integrative exploration of specific neurobiological pathways known to be impaired in AD. A comprehensive analysis of gene expression and DNA methylation levels is performed for genes known to be associated with sphingolipid function. The identified key genes and their particular methylation signatures offer mechanistic insights into AD pathology and may act as potential biomarkers.

In vitro modeling of human diseases allows us to gain crucial insights into mechanisms underlying disorders, hence devising and optimizing new strategies for therapeutic intervention. CHAPTER 6 features the differential network-based analysis of transcriptomic data sets obtained from brain organoids that served as an *in vitro* model of Batten disease. This study focuses on identifying key genes and pathways that are disrupted during the course of this disease.

Chapter 2

Modeling of Cellular Systems: Application in Stem Cell Research and Computational Disease Modeling

Muhammad Ali [A], Antonio del Sol [A, C].

[A] Computational Biology Group, Luxembourg Centre for System Biomedicine (LCSB), University of Luxembourg, Luxembourg City, Luxembourg.

[C] Moscow Institute of Physics and Technology, Dolgoprudny, Moscow, Russia Federation.

Ali M., del Sol A. (2018) Modeling of Cellular Systems: Application in Stem Cell Research and Computational Disease Modeling. In: Alves Barbosa da Silva F., Carels N., Paes Silva Junior F. (eds) Theoretical and Applied Aspects of Systems Biology. *Computational Biology*, vol 27. Springer, Cham.

2.1 Abstract

The large-scale development of high-throughput sequencing technologies has allowed the generation of reliable omics data at different regulatory levels. Integrative computational models enable the disentangling of a complex interplay between these interconnected levels of regulation by interpreting these large quantities of biomedical information in a systematic way. In the context of human diseases, network modeling of complex gene-gene interactions has been successfully used for understanding disease-related dysregulation and for predicting novel drug targets to revert the diseased phenotype. Furthermore, these computational network models have emerged as a promising tool to dissect the mechanisms of developmental processes such as cellular differentiation, trans-differentiation, and reprogramming. In this chapter, we provide an overview of recent advances in the field of computational modeling of cellular systems, its major strengths and limitations. Particularly, attention is paid to highlight the impact of computational modeling in our understanding of stem cell biology and the complex multifactorial nature of human disorders and their treatment.

2.2 Introduction to Systems Biology

Systems biology is the integration of computational and experimental research to study the mechanisms underlying complex biological processes as integrated systems of many interacting components. Systems biology offers a holistic rather than reductionistic approach for understanding and controlling biological complexity, which arises due to the interconnected components working together in a synchronized fashion to maintain the phenotype of an organism. Systems biology-based approaches help us in exploring these systems at the level of a cell, tissue, organ, organism, as well as a population and an ecosystem. Characterization of these systems in their full complexity allows us to better understand the properties of the components involved and their static as well as the dynamic behaviour.

During the last decade, various experimental techniques have enabled the large-scale generation of high-throughput (HT) biological data across different levels of regulation. Among them, the ones which have been extensively used for modeling biological systems are, mutation detection by single nucleotide polymorphism (SNP) genotyping [273], gene expression quantification by messenger ribonucleic acid sequencing (RNA-seq) [11], identification of protein interactions with deoxyribonucleic acid (DNA) via chromatin immunoprecipitation sequencing (ChIP-Seq) [7], and quantification of different metabolite levels in the organism by HT metabolic screening [264]. The associated plethora of data has spurred the development of computational models, allowing the dissection of the complex mechanisms underlying different biological processes at different regulatory levels. This vast amount of data across different levels of a biological system has also opened a new gateway to integrate data from these different but interconnected layers to gain a deeper system-level understanding.

2.3 Computational Modeling of Cellular Systems

The complexity of biological systems can be broken down to an individual molecule or atom, but to study their overall effect on the system, we need to understand their interactions with each other and with other ongoing processes or pathways in the system. This is even crucial for understanding their role in the onset or progression of diseases such as cancer and Alzheimer's disease. Mathematical models of biological systems, which use efficient algorithms and data structures, enable researchers to investigate how complex regulatory processes are intertwined and how any perturbation in these processes can lead to the development of disease. Recent advancements in computational resources and large-scale generation of so-called "omics" data sets has led to model, visualize, and rationally perturb systems at different levels such as modeling and designing from an atomic resolution to cellular pathways and the analysis of guided alterations in systems and their propagation.

A computational model of a complex system can help us in understanding the behavior of that system by simulating its dynamics. Numerous computational models have been developed to address different kinds of processes – for example, flight simulator models [152], protein folding models [2], and artificial neural network models [1]. Moreover, computational modeling has emerged as a powerful and promising approach to investigate and manipulate biological systems. In particular, different categories of cellular processes have been modelled by using the computational models, such as gene regulation, signaling pathways, and metabolic processes [29]. However, modeling the biological system at a cellular level is a convoluted problem involving the challenging task of understanding the cellular dynamics and characterizing the underlying biological principles. Gaining a systems-level understanding of these intertwined cellular processes and their complex interconnections may serve as a critical foundation for developing therapeutic fronts where we anticipate that computational cellular modeling approaches will make a profound impact.

2.3.1 Gene Regulatory Networks

It is increasingly recognized that complex biological systems cannot be described in a reductionistic view. To understand the behavior of such a complex system, a deeper understanding of the different components of this system and their interactions with each other is required. This knowledge can help us in viewing the system under study as a network of components, which has a certain topology. This topological information is fundamental in constructing a realistic model to unlock the functions of the network. There are various types of biological networks, which have been extensively studied by researchers, such as gene regulatory networks (GRNs), protein-protein interaction (PPI) networks, signal transduction networks, and metabolic networks. In particular, GRNs are the on-off switches of a cell operating at the transcriptional level where two genes are connected to each other if the expression of one gene modulates the expression of another one by either activation or inhibition. A GRN can be represented by a directed graph where nodes represent the genes and directed edges among these nodes represent gene-gene interactions. As a simple example of a GRN, Figure 2 depicts the schematic illustration of core pluripotency transcription factors (TFs) that maintain the pluripotency potential of stem cells. *POU5F1*, *SOX2*, and *NANOG* have a positive self-regulation, while they also positively regulate each other.

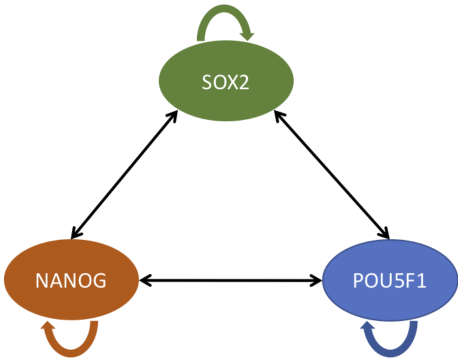


Figure 2: **Transcriptional core of pluripotency factors.** Schematic representation of the transcriptional regulation of core pluripotency factors.

Genes in a GRN are not independent from each other; rather they regulate each other and act collectively. This collective behavior can be observed by mRNA quantification obtained from a microarray or mRNA-Seq experiment where some genes are significantly upregulated, while others are downregulated, suggesting that upregulated genes might be the one inhibiting the downregulated genes. The connections among all the genes in a GRN cannot be inferred correctly by just relying on their mRNA levels or simple gene expression correlation-based methods but by integrating literature-based information stored in relevant repositories. These repositories, such as the MetaCore database from Clarivate Analytics and gene pathway studio [206], contain experimental evidence of gene-gene interactions where one gene regulates the expression of its target genes.

The topological analysis of a GRN can help in identifying some important genes in the network, such as those involved in network motifs. Network motifs are topological patterns that occur in real networks significantly more often than in randomized networks [192]. These patterns have been preserved over evolution on the expense of mutations that randomly change edges. Similarly, the detection of elementary circuits, which is the path starting from and ending in the same gene visiting each intermediate gene only once, has been associated with the stability of GRNs [262, 227]. These circuits can either have an even number of inhibitions hence called positive circuits (positive feedback loops) or an odd number of inhibitions, therefore called negative circuits. Moreover, the genes in the strongly connected components (SCCs) of a network – a subnetwork in which every gene is reachable from every other genes in that subnetwork through a direct path – are interconnected positive and negative circuits and usually considered to be the pivotal genes, maintaining the network phenotype.

GRNs play an important role in unravelling the molecular mechanisms underlying a particular biological process, such as cell cycle, apoptosis, and cellular differentiation. A paramount problem in modeling a GRN is to understand the dynamical interactions among the genes in the GRN, which collectively govern the behavior of the cell. Several methods have been proposed to date to infer GRNs from gene expression and epigenetic data [58, 211, 314, 175]. Although the goal is same, i.e. to model biological processes, available methods rely on different modeling formalisms, for example, logical models have been used to infer Boolean networks, probabilistic Boolean networks, and Petri nets. Furthermore, continuous models have also been introduced for the same purpose; prominent examples include continuous linear models and models of TF activity [26]. Computational methods for modeling GRNs have proved to be a promising bioinformatics application. In this chapter, we tried to explore the applications of GRN models in stem cell research and disease modeling.

2.4 Systems Biology of Stem Cells

A human body comprises different kinds of cells that are distinct in their structure as well as in function. These trillions of cells are largely containing the same genomic material and contain only a limited number of - approximately 400 [288] - distinct cell types. The different types of cells in the body and their structure perfectly suit the role they perform. For instance, kidney cells (hepatocytes) are completely different in structure and function from skin cells (fibroblasts). Interestingly, all these different kind of cell types in an adult organism actually originate from the same kind of precursor cells, called pluripotent stem cells. Pluripotent stem cells have the potential to give rise to any kind of fetal or adult cell type. Whereas stem cells have the potential to give rise to any kind of lineage at the embryonic developmental stage, this plasticity, i.e. pluripotency, is lost upon differentiation into a certain somatic cell type. Cell identity specification is considered to be determined by cell-specific gene expression programs, which represent highly complex processes tightly controlled at the transcriptional and epigenetic regulatory levels. In order for a cell-specific gene to be expressed during differentiation, the DNA corresponding to this gene and its distal regulatory elements must be in an accessible and active state. In this context, the cell-specific epigenetic landscape is hypothesized to account for the differences between heterogeneous cell fates.

2.4.1 The Generation of iPSCs

Recent advancements in molecular biology have enabled researchers to obtain induced pluripotent stem cells (iPSCs) from somatic cells by following a reliable cell conversion methodology – usually referred as cellular reprogramming. By following established protocols of applying a particular recipe of TFs into the medium of an *in vitro* somatic cells culture, iPSCs can be grown in culture and will have the same plasticity potential as those of stem cells found in embryos. The very first and a well-known example of cellular reprogramming is the conversion of mouse fibroblasts into iPSCs by introducing four TFs (*POU5F1*, *SOX2*, *MYC*, and *KLF4*) [270]. iPSCs provide a new framework to obtain a renewable source of healthy cells which can help in treating a wide spectrum of diseases, such as neurodegenerative and cardiovascular disorders. Nevertheless, a bottleneck in cellular reprogramming is the identification of effective reprogramming determinants, i.e specific TFs, that can trigger a transition between cellular phenotypes with high conversion efficiency and fidelity.

2.4.2 Transdifferentiation

Similar to reprogramming, where we want iPSCs to differentiate into a particular lineage and cell type, another approach to obtain the same cell type of interest without undergoing an intermediate pluripotent state is transdifferentiation. Transdifferentiation is the direct and irreversible conversion of one somatic cell type to another. Various examples of transdifferentiation have been reported in the literature where a defined TF recipe or a combination of TFs and microRNA (miRNA) or other small molecules was introduced in a somatic cell type culture and the desired mature cell type was obtained within days. For example, the TF *MYOD1* has been used to transdifferentiate mouse embryonic fibroblasts into myoblasts [46]. Since this first case reported in literature in 1990, there have been numerous examples of successful somatic cell conversions with defined factors and small molecules [34, 289, 218].

Interestingly, various computational methods have been reported to systematically predict the candidate TFs that can help in converting one fully differentiated cell type to another, and their predictions have been experimentally validated in a laboratory setting [198, 232, 129]. Transdifferentiation is emerging as a promising approach to directly transdifferentiate cells while avoiding the use of iPSCs to derive patient-specific cells. This remarkable potential of transdifferentiation is proving to be the most promising source of regenerative medicine for tissue regeneration and disease therapy. Nevertheless, an important roadblock to efficient transdifferentiation is the limited number of successful cellular conversions obtained so far, with low to intermediate efficiency. Furthermore, the role of changes in the epigenetic landscape for achieving an efficient transdifferentiation has not yet systematically explored.

2.5 Modeling Cellular Phenotypes and Conversions

In some modeling approaches, a cellular phenotype is modelled as a network of genes with a particular gene expression pattern and a unique stable steady state (attractor). Phenotypic transitions in such models are introduced by identifying the genes in the network that can destabilize this attractor and lead the system into another attractor. This concept has been applied to model diseases as a transition from a healthy phenotype to a diseased state, caused by a mutation or a chemical compound [62]. Moreover, it has also been applied in modeling cellular conversion [232] (reprogramming, differentiation, and transdifferentiation), where researchers first identify the attractors of two phenotypes (starting and destination cell types) and then pinpoint a minimal set of genes in the network's elementary circuits whose perturbation (up- or downregulation) led the attractor of the starting cell type to the attractor of the destination cell type [58, 211].

Modeling the cellular phenotype requires the inference of condition-specific GRNs. Literature suggests a number of different GRN inference methods, which rely on different underlying rationales, such as modeling formalism (Boolean and Bayesian) and different updating schemes (synchronous and asynchronous). Furthermore, there have been methods introducing the concept of contextualization, which is the removal of non-specific edges that are not compatible with the gene expression program of the cell type under consideration [58, 314]. Most of these methods rely only on gene expression data, but more recently, approaches using gene expression as well as epigenetic information have also been introduced [175]. Nevertheless, a bottleneck in the GRN inference problem is the benchmarking of inferred networks. Most of these methods rely on the interactions of a set of specific TFs in a particular cell type diagnosed by experimental ChIP-seq to validate the networks. Unfortunately, this benchmarking approach can only validate a part of the network as the complete benchmarking information, ChIP-seq for all the TFs in one cell type, is not available for even a single cell type. Moreover, ChIP-seq cannot be a perfect gold standard as some TF-DNA interactions might be incorrectly labeled as positives because TF binding does not necessarily indicate a functional interaction. Besides ChIP-seq, SNP data as well as random network inference has been used as a reference for the benchmarking of inferred networks [175], but none of these approaches offer a complete and systematic network inference validation. However, due to the consistent release of new TF ChIP-seq experiments by collaborating labs in the ENCODE consortium [52], the amount of available TF binding site profiles is steadily increasing, which might eventually mitigate the problem of missing data in the future. Furthermore, increasing number of genes and TFs perturbation experiments in the Gene Expression Omnibus database [50] may serve as an alternative approach for network validation as the gene expression profiling after gene over-expression or knock-down can provide authentic information about the functional gene-gene interactions.

2.6 Computational Disease Modeling

The advances in molecular biology have resulted in the establishment of fast and efficient protocols for generating iPSCs cells *in vitro*. This –cells-in-a-petri-dish– approach has allowed for sophisticated modeling of human disorders and uncovering the molecular basis of disease-related dysregulation. Moreover, the generation of patient-specific iPSCs-derived cell types possessing specific disease-related mutations provides an extremely viable *in vitro* system for the investigation of disease-associated perturbations and to apply drug screening. However, the complex nature of various human disorders, which often involve multiple dysregulated genes acting together, hinders our understanding of disease-specific impairments [205]. As such, dysregulated genes, in conjunction, initiate a cascade of failures, which causes malfunctioning at the systems level, resulting in specific disease phenotypes. Therefore, instead of investigating individual genes in a system, we may rather focus on their interactions as a channel to propagate disease-related perturbations. In this context, healthy and disease states can be represented as cellular network phenotypes with stable steady states, where a disease-specific perturbation shifts the steady state of a healthy network into the steady states of a disease network. Thus, the construction of complex regulatory interaction networks offers a new method for gaining a system-level understanding of disease pathology. These network-based models have proved to be a promising framework for identifying disease-related genes based on network topology [143]. For example, disease-gene-drug associations have already been predicted based on differential network analysis [314]. Furthermore, disease-gene relationships have also been reported based on the identification of disease-related subnetworks and prediction of network neighbours of disease-associated genes [83, 13].

2.6.1 Differential Network Analysis and Disease Models

There has been an increasing number of approaches exploring the associations between genes, drugs, and diseases. Some of them include the construction of data repositories where different compounds have been tested experimentally to associate drugs with genes and diseases, including the connectivity map [144] and gene perturbation atlas [302]. These approaches have provided immense help in linking drugs to their target genes, which has also benefited in drug repositioning based on particular gene expression signatures produced after drug perturbation. However, these approaches neglect the mechanisms underlying gene regulation and avoid the indirect targets of drugs. Moreover, only a limited set of drugs and cell types have been used to carry out these experiments, which implicitly means that these approaches cannot cover the entire spectrum of human diseases. In this regard, approaches relying on network pharmacology have proved to be promising in identifying candidate genes whose perturbation might lead to a desired therapeutic phenotype. Recently, there have been few reports relying on unique and differential network topologies to identify the differential regulatory mechanisms leading to a given pathology [314, 116, 195]. These approaches allow the building of condition-specific networks by collecting gene-gene interaction information from literature-curated resources and to predict target genes and drugs that could maximize the reversion from a disease phenotype to a healthy one. For example, by using the differential network-based approach, cyclosporine was predicted as a candidate drug to treat systemic lupus and rheumatoid arthritis. Surprisingly, this blindfold prediction was in agreement with existing literature, as cyclosporine has been successfully applied to treat these diseases [32, 294].

These findings suggest that network-based approaches hold a great potential to identify new disease-related genes and biomarkers for complex diseases. These approaches can uncover the regulatory mechanisms underlying disease pathologies by analysing the differences in gene regulatory interactions of condition-specific networks. Furthermore, *in silico* simulations to mimic the network response upon drug application can boost the quest of identifying a putative drug for therapeutic intervention. Nevertheless, a prominent limitation of cell reprogramming approaches is the availability of good-quality interactome maps. For only a limited number of human diseases, we are able to gather enough omics data to construct a reliable interactome, which can help in exploring the underlying disease mechanisms. In order to overcome this information gap, research teams throughout the world are profiling next-generation sequencing experiments to obtain high-quality interaction maps of specific human disorders [74, 165, 128], while other consortiums like Roadmap Epigenomics [23] and ENCODE [52] are striving to create reference human epigenomes and large-scale ChIP-seq profiling for different TFs across different cell types, respectively. Nonetheless, this information is still far from being complete and will require extensive future efforts to develop complete, high-quality, and noise-free interaction maps for all well-studied human diseases. We strongly believe that bridging this information gap will play a crucial role in the future of biomedical research.

Chapter 3

INTREGNET: Integrating epigenetic and transcriptional landscapes in a network-based model for increasing cellular conversion efficiency

Muhammad Ali [A], Sascha Jung [A], Antonio del Sol [A].

[A] Computational Biology Group, Luxembourg Centre for System Biomedicine (LCSB), University of Luxembourg, Luxembourg.

Ali M., Jung S., del Sol A. (2019) INTREGNET: Integrating epigenetic and transcriptional landscapes in a network-based model for increasing cellular conversion efficiency. *In preparation*.

3.1 Abstract

The design of novel strategies for cellular conversion by using a defined set of transcription factors (TFs) has shown promising applications in regenerative medicine. Nevertheless, the identification of TFs that can induce a desired transition with high conversion efficiency remains a significant challenge. Although computational approaches have been developed to guide cellular conversion experiments, they do not tackle the problem of conversion efficiency. In particular, these approaches do not take into account epigenetic regulatory effects when modeling cellular conversion, which is important for addressing the aforementioned problem. Here, we present INTREGNET, a computational method for systematically identifying minimal sets of TFs that can induce desired cellular transitions with increased efficiency. This method relies on the integration of transcriptomic and epigenetic information for reconstructing cell-type-specific core transcriptional regulatory networks (TRNs). Specifically, when applied to the induction of pluripotent stem cells (PSCs) from different somatic cells, INTREGNET was able to distinguish between more- and less-efficient TF combinations. Thus, INTREGNET can guide experimental attempts for achieving effective *in vivo* cellular transitions, where limited conversion efficiency is a crucial barrier for its application in regenerative medicine.

3.2 Introduction

Cell identity specification in multi-cellular organisms gives rise to hundreds of different cell types sharing the same genetic information through a complex process, which is tightly controlled at different regulatory levels. Established cell-type-specific gene expression programs are orchestrated by intricate and interconnected regulatory networks at the epigenetic and transcriptional level controlling homeostasis of differentiated or pluripotent cells [10, 76, 135, 253]. Epigenetic mechanisms, such as DNA methylation [124, 260] and histone modifications [265, 139], regulate chromatin accessibility [235] and constitute a epigenetic code that is recognized by transcriptional and epigenetic regulators, such as transcription factors (TFs), chromatin modifiers and remodelers [42, 106]. However, in order to specify cell identity, it has been shown that a small set of transcription factors, known as core TFs, is sufficient [31, 96, 199, 247]. Following this rationale, current strategies for inducing desired cellular conversions are based on the over-expression of a combination of exogenous TFs and have been used as a qualitative measure for evaluating the ability of core TFs to induce and enhance cellular transitions.

Over the past years, several computational methods have been developed to guide cellular conversion experiments. Early approaches relied on the identification of significant differences in transcriptomic or epigenetic profiles [59, 85, 60, 111], while more recent ones acknowledged the importance of gene regulatory networks (GRNs) to identify TFs that can induce desired cellular transitions, termed as instructive factors (IFs) [232, 198]. However, these methods are still unable to identify optimal combinations of IFs that more efficiently trigger such transitions. Experimental evidence suggests that gene expression is insufficient for determining efficient IFs, as it is the interplay of epigenetic and transcriptional regulation that mediates cellular conversions [191, 137, 244]. As such, cell-type-specificity is determined not only by the transcriptional, but also epigenetic program, characterized by accessible chromatin regions, active enhancers, and differential binding of regulators [204, 113, 245], which highlights the epigenetic reorganization required when converting one cell type into another. Altogether, this demonstrates the need for an integrative approach to create tailor-made computational models that are essential for predicting more efficient sets of IFs.

Epigenetic and/or transcriptomic dysregulation that disrupt the normal cellular differentiation process lies at the core of many diseases [161, 295], requiring an *ex vivo* or *in vitro* cell application for the development of novel treatment strategies. For example, mesenchymal stem cells (MSCs) represent a rare stem cell type whose *in vitro* expansion is vital for obtaining sufficient amounts of cells for treating various heart [5, 174, 226, 123], brain [122, 178, 162] and wound-healing [300, 301] related disorders. However, progressive spontaneous differentiation and aging of MSCs may occur during expansion, which can be modulated by extrinsic epigenetic signals such as histone H3 acetylation, playing a key role in regulating these intricate processes [161]. Similarly, an increasing amount of literature suggests epigenetic mechanisms to be crucial for regulating B-cell maturation and its dysregulation has been associated to the initiation and acceleration of multiple autoimmune diseases such as systemic lupus erythematosus [296, 298, 297] and rheumatoid arthritis [184, 95, 134]. Taken together, this evidence suggests that epigenetic mechanisms, along with other regulatory layers, play a crucial role in normal cellular differentiation. Therefore, reconstructing cell-type-specific network models by integrating epigenetic and transcriptomic information can provide deeper insights into underlying mechanisms, allowing us to predict specific external stimuli (e.g. TF over-expression) that can overcome the epigenetic barriers restricting the differentiation potential of cells in different disorders.

In the present work, we developed a novel computational method to predict efficient IFs for desired cellular conversions by reconstructing INtegrative Transcriptional REGulatory NETworks (INTREGNET) based on cell-type-specific transcriptomic and epigenetic data sets. We analyzed more than 7600 publicly available gene expression profiles to identify a set of core TFs across different human cell types and cell lines. Based on the integration of i) a set of candidate core TFs, ii) histone modifications, iii) chromatin accessibility, and iv) experimentally validated TF binding sites, we are able to reconstruct core transcriptional regulatory networks (TRNs) for 48 different human cell types and - lines. Furthermore, molecular interactions between TFs have been inferred by integrating protein-protein interaction (PPI) data. The reconstructed networks are cell-type-specific, encompassing interactions that are compatible with the corresponding epigenetic and transcriptomic state. Benchmarking against experimentally validated gold standard networks [209, 27, 82] verifies the cell-type-specificity of the reconstructed networks, preserving more than 95% of the gold-standard interactions. Further, these cell-type-specific networks were employed to build a Boolean-based model for predicting sets of instructive factors that induce desired cellular conversions. Results show that INTREGNET outperforms other state-of-the-art methods by predicting significantly more experimentally validated IFs. More importantly, INTREGNET is able to predict specific sets of IFs inducing cellular conversion events with increased efficiency. Thus, INTREGNET can provide a guidance to stem cell researchers to improve the efficiency of cellular conversion, which constitutes a long-standing problem in regenerative medicine and beyond.

3.3 Materials and methods

Cell-type-specific core TRNs were reconstructed by integrating transcriptomic and epigenetic profiles. An overview of INTREGNET's workflow is shown in Figure 3, while each individual step, i.e. epigenetic and transcriptomic data processing, network reconstruction, validation, and application, is described in detail in the remainder of this section.

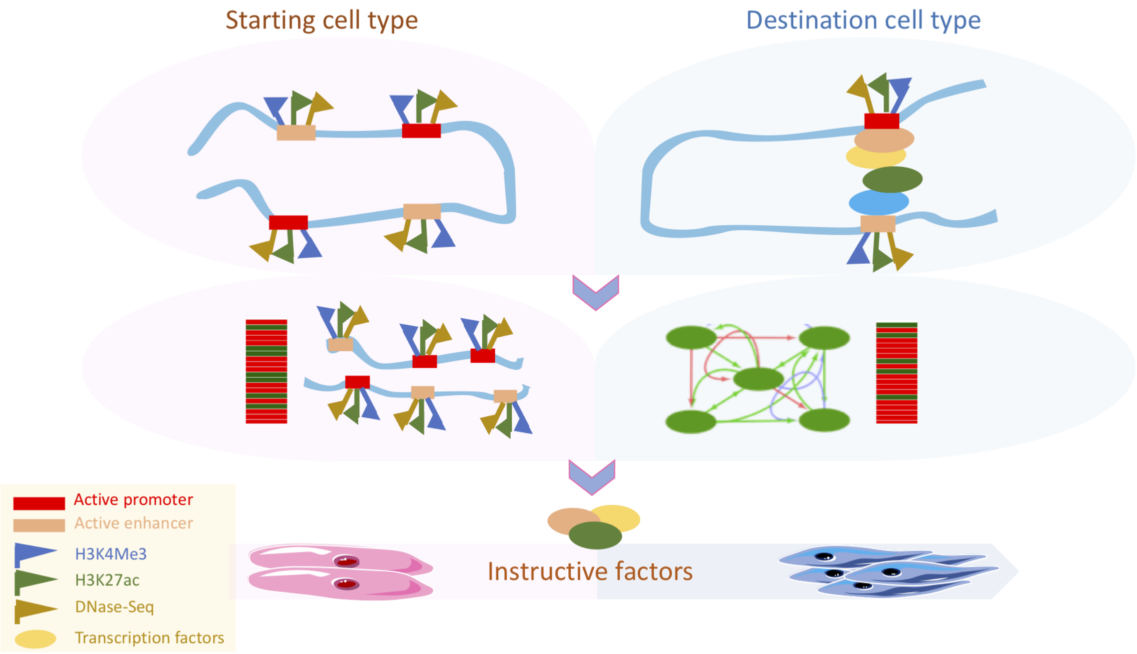


Figure 3: **Schematic workflow of INTREGNET.** INTREGNET utilizes epigenetic and transcriptomic profiles from the initial as well as final cell type. Epigenetic profiles help INTREGNET to characterize active promoter (H3K4me3) regions, active enhancer (H3K27ac) regions, and accessible genomic domains (DNase-seq). Transcriptomic profiles are used to identify uniquely and significantly expressed TFs. To predict a set of instructive factors for a desired cellular transition, first, the epigenetically active domains are characterized in starting and destination cell types. Next, the active regulatory domains in the destination cell type are integrated with TF ChIP-seq data to reconstruct a core regulatory network for the destination cell type. Here, enhancer and promoter regulation (green) is distinguished from enhancer-only (red) and promoter-only regulation (purple). Lastly, the core TRN of the destination cell type is integrated with gene expression signatures and active cis- and trans-regulatory elements from the starting cell type to predict instructive factors required for cellular conversion under consideration.

3.3.1 Identification of core TFs

Individual cell types and transcriptomic samples were characterized by a set of core TFs. Each sample was compared against a background of more than 7600 different samples of various cell types and cell lines included in Recount2 [51], a database of publicly available, uniformly processed RNA-seq data sets. Of note, all samples from The Cancer Genome Atlas (TCGA) and those containing the terms “cancer”, “disease”, and “single cell” in the title or description of their Gene Expression Omnibus [50] (GEO) entry were excluded prior to the analysis. GEO accessions of all considered RNA-seq samples can be found in supplementary Table S17. Transcription factors were then ranked based on the uniqueness of their expression in every individual sample using a modified version of the method proposed by D'Alessio et al. [59]. Generally, the approach consisted of three steps that were repeated for every transcription factor. First, given a single query sample, all data sets having a Pearson correlation of more than 0.75 with the query were excluded from the background. For this purpose, 30 ESC samples were randomly chosen from the Recount2 data set and their correlation with rest of the samples was computed iteratively. Every unique correlation score obtained was then used as a threshold for creating the confusion matrix, based on the annotation of all the ESC samples in Recount2 data set. This process was repeated for every selected ESC sample and F1 scores were computed against every correlation threshold. By plotting the F1 score against the respective correlation thresholds, we see that highest F1 score is obtained at a 0.75 correlation threshold (supplementary Figure S19). Subsequently, the uniqueness of each TF's expression in the query was assessed by comparing an idealized probability distribution, which contains 1 in place of the considered query sample and 0 otherwise, to the background distribution containing the expression of the TF in all samples. Finally, the background distribution is normalized by the sum of its elements and compared to the idealized distribution by means of Jensen-Shannon divergence (JSD). The 10 most unique TFs in each sample, i.e. having the highest JSD value, were selected as core TFs.

3.3.2 Reconstruction of cell-type-specific core TRNs

Based on the identified core TFs, transcriptional regulatory networks were reconstructed for various human cell types and cell lines, reflecting the coordinated action of transcription factors on their targets in a cell-specific manner. Regulatory relationships were reconstructed from transcription factor ChIP-seq experiments, the active promoter mark H3K4me3, the active enhancer mark H3K27ac and chromatin accessibility defined by DNase-seq, and are represented in a Boolean modeling framework. For that purpose, INTREGNET performed three steps. First, transcriptomic data was made compatible with the Boolean modeling framework. RefBool [127] was elected for discretizing the expression values in every sample individually, based on a universal transcriptomic reference for each gene. Unlike the proposed use of RefBool [127], a single p-value threshold was used for determining active and inactive genes. More specifically, when testing the null hypothesis that a gene is not expressed, p-values of less than 0.15 are considered to be significant, which leads to the rejection of the null hypothesis. Second, active proximal and distal regulatory regions were identified for every active TF. Promoters were defined based on the Ensembl promoter annotation file obtained from The Eukaryotic Promoter Database [223] (accessed March 23rd, 2018) and restricted to 1500bp upstream and 500bp downstream of the transcription start site (TSS). In order to assess whether the promoter region of a TF is active in a given cell type, corresponding H3K4me3 peaks were obtained from ENCODE [54] or Cistrome [186] and projected onto the region. A promoter region is considered to be active if it overlaps with at least one H3K4me3 peak. For enhancers, the GeneHancer database [80] (accessed April 6th, 2018) was leveraged to link active TFs to their known enhancer regions. All regions overlapping a cell-type-specific H3K27ac peak were considered to be active while inactive enhancer regions were discarded. More precisely, instead of considering the complete enhancer to be active, the region was truncated to the H3K27ac peak region. Finally, TF binding events were identified in active promoter and enhancer regions. Publicly available and uniformly processed TF ChIP-seq data sets, i.e. called peaks, were obtained from Cistrome [186], pooled and projected onto the active regulatory regions. Every binding event sharing one base pair with an active region constitutes a potential regulatory interaction. These interactions were further filtered by cell-type-specific accessible chromatin profiles (DNase-seq peaks) from ENCODE [54] and GEO [50], such that all remaining interactions overlap with at least one peak.

Following this strategy, a TRN was reconstructed among the set of core and non-core TFs in every sample. The selection of non-core TFs consisted of three steps. First, only those TFs were considered that were expressed in the cell type. Next, these expressed TFs were filtered based on their JSD ranks and only those TFs whose ranks were significantly lower than their average rank across all samples were selected. Finally, only those non-core TFs that were regulating at least one core TF and that were simultaneously being regulated by at least one core TF were kept in the network. The subsequently derived interactome constituted the core TRN of cell type under consideration.

ENCODE and GEO accessions of considered H3K27ac, H3K4me3, and DNase-seq experiments for every individual cell type/line are given in supplementary Table S9. All considered data sets were annotated to genome assembly GRCh38 or converted to GRCh38 by using the CrossMap tool [311].

3.3.3 Validation of reconstructed TRNs

To assess the cell-type-specificity of reconstructed core networks, a comparison with manually curated core networks from the literature, containing TF ChIP-seq validated interactions, was conducted. Here, the set of gold-standard networks was composed of human hepatocytes [209], embryonic stem cells (ESCs) [27] and two cancer cell lines (MCF7 and HepG2) [82], and compared at the level of the subset of TFs that were present in the reconstructed core TRNs.

In addition, we leveraged cell-type-specific TF ChIP-seq data to calculate the enrichment of experimentally validated interactions in the reconstructed TRNs. For this purpose, uniformly processed TF ChIP-seq peaks were gathered from ENCODE [54] and Cistrome [186]. In order to keep the data consistent between these two resources, all peak files were converted to GRCh38 genome assembly by using the CrossMap tool [311], if they were aligned to a different assembly. ENCODE and GEO accessions of all the considered TF ChIP-seq experiments are given in supplementary Table S10. Based on these datasets, two analyses were carried out. First, the fraction of TF interactions in promoter regions that were validated by a peak in the cell-type-specific ChIP-seq experiments was assessed. For that purpose, only those TFs in the networks were considered for which ChIP-seq data was available. Second, the number of false-positive interactions were quantified by counting the fraction of interactions in promoter regions that were not validated by a ChIP-seq peak from an experiment in the same cell type/line, but under different conditions. Of note, true negative and false-negative interactions cannot be reliably assessed and are therefore excluded from the assessment.

3.3.4 Inference of Boolean logic rules

The representation of the reconstructed TRNs in a Boolean modeling framework requires the inference of Boolean expressions that describe the relationship between all regulators of a single TF. Here, TFs can act cooperatively, e.g. by forming a complex, or competitively by sharing parts of their DNA binding motif. INTREGNET infers these connections by identifying all ChIP-seq peaks in the TRN that reciprocally overlap more than 62% using the intersectBed program from bedtools v2.22.1 with parameter `-loj -r -f 0.62`. As a result, an undirected network among TF binding sites is obtained in which the strongly connected components (SCCs) are assumed to represent the cooperative interactions of TFs. Here, SCCs were detected using the “clusters”-method of the R “igraph”-package (version 1.2.2). For determining the overlap threshold, a positive and negative gold-standard dataset of protein-protein interactions (PPIs) was assembled. The positive set consisted of 33 PPIs included in iRefIndex [237] that fulfill three requirements. First, ChIP-seq data was available in Cistrome [186]. Second, their interaction type has been classified as “direct interaction” (MI:0407) and, third, experimental validation had been conducted in humans. For the negative set, manual annotations in the Negatome 2.0 database [24] were obtained resulting in 72 true-negative PPIs for which ChIP-seq data was available in Cistrome. The percentage of overlap for all peaks of all gold-standard PPIs was assessed to assemble the positive and negative distributions. Two TFs with overlapping ChIP-seq binding sites are said to form a complex, if the probability of belonging to the positive distribution is higher than belonging to the negative distribution. Transcription factors predicted to form a complex were connected by an AND-gate, which represents the necessity of each individual subunit, while TFs with competing and non-overlapping binding sites are connected by an OR-gate. Finally, enhancers and the promoter region of a TF were incorporated into a single regulatory rule by forcing the regulation of the promoter and at least one enhancer, which corresponds to the connection of multiple enhancer regions by OR-gates and of the enhancers with the promoter by an AND-gate.

3.3.5 Prediction of efficient combinations of instructive factors

An algorithm recently developed in our lab (manuscript under preparation) was used for predicting optimal combination of instructive factors (IFs) to efficiently induce desired cellular transitions. In terms of the transcriptional regulatory network of the target cell type, the algorithm searches for the minimum combination of IFs whose perturbation can efficiently restore the gene expression program of the target cell. This corresponds to the state of the network in which all Boolean regulatory rules evaluate to true. By construction, this state is a steady state of the system, regardless of the imposed updating scheme. In order to identify the probability of all network states to reach this desired steady state, the model checker PRISM v4.4 [142] was employed. As PRISM is unable to handle Boolean logic rules, Boolean rules were transformed to equivalent polynomials by the following rules [136]: Given two TFs A and B in the Boolean TRN, the following relations hold.

$$\neg A \equiv 1 - A \tag{3.1}$$

$$A \vee B \equiv A + B - A \bullet B \tag{3.2}$$

$$A \wedge B \equiv A \bullet B \tag{3.3}$$

While the second rule states a valid transformation of Boolean rules into polynomials, it is impractical in the presence of multiple TFs with competing or non-overlapping binding sites within regulatory regions. By applying De Morgan's law ($A \vee B \equiv \neg(\neg A \wedge \neg B)$), a fourth rule can be derived that is easily adaptable to multiple TFs:

$$A \vee B \equiv \neg(\neg A \wedge \neg B) \equiv 1 - ((1 - A) \bullet (1 - B)) \tag{3.4}$$

With these transformations, a PRISM model of a discrete time markov chain (DTMC) is established in which each TF is a module that can change its state based on the evaluation of the polynomial expressions. During each step of the model, a single TF is selected individually and its state is updated, i.e. the DTMC obeys an asynchronous updating scheme. Finally, the property that has to be checked by PRISM can be stated as "eventually all TFs in the network are active". In PRISM syntax that corresponds to "$F(TF_1 + TF_2 + ... + TF_n = n)$". Invoking PRISM with the option "-v" returns all states with their corresponding probabilities of fulfilling the property.

Finally, candidate IFs are established by selecting TFs of the core TRN that do not have an active promoter or any active enhancer in the initial cell type. As a second step, the algorithm searches for the minimum set of TFs whose perturbation leads to the maximum number of gene expression changes of differentially expressed genes between initial and final cell types, while minimizing the number of epigenetic changes during this process. Like in the construction of core TRNs, promoter regions, defined by The Eukaryotic Promoter Database [223], and enhancer regions, defined by the GeneHancer database [80], are called active if they overlap with a cell-type-specific

H3K4me3 or H3K27ac peak, respectively. A score for each combination of candidate factors is set as the weighted average of all Boolean TRN states able to reach the desired steady state, i.e. in which all TFs are active. Here, the weight for each TF to be in state 0 or 1 is defined as the probability of observing a greater or lower expression value in the background distribution of RefBool [127], respectively. Consequently, the probability of being in a certain network state is defined as the product of the probabilities of being in the individual TF states. Combinations having a higher score are more favorable, and thus predicted to be more efficient, than low scoring combinations.

3.3.6 Validation of cellular conversion algorithm

To assess the predictive power of the proposed method for identifying an efficient combination of instructive factors, an extensive literature review was conducted to gather information about the experimentally confirmed cellular conversions where a particular set of instructive factors has been utilized for achieving a desired cellular transition. Interestingly, for some of the transitions, existing studies reported the conversion of an identical starting cell type to a similar destination cell type but utilized different combinations of instructive factors, thus yielding different cellular conversion efficiencies. For all the cellular transition examples which are in pairs (low and high efficiency) or only a single perturbation is reported (supplementary Table S10), the core TRNs representing the destination cell types were reconstructed. In addition to the destination core TRNs topologies, the epigenetic status of the constituent TFs and their gene expression values in the starting cell type were employed for predicting the sets of instructive factors.

3.3.7 Nomenclature of TFs

Throughout this study, the official HGNC gene symbols (e.g. *MYC* and *POU5F1*) have been used for representing the TFs and their targets.

3.4 Results

3.4.1 Reconstruction of cell-type-specific core TRNs

A small group of core TFs has been reported to predominantly control the gene expression program of embryonic stem cells [27] as well as cell types [209] and cell lines [82]. These sets of core TFs usually auto-regulate themselves and control the regulation of other TFs by making interconnected regulatory loops, hence forming a core network that determines cellular identity and function [209, 27, 82]. Experimental studies suggest that there is a complex interplay between transcriptional and epigenetic landscape that controls cell differentiation and lineage commitment [176]. Thus, it is important to consider the combined regulatory effect of these different but interconnected layers, while modeling the cellular phenotypes and their transitions. The approach we present here, connects both layers of regulation to reconstruct cell-type-specific core TRNs by integrating high throughput transcriptomic (RNA-seq) and epigenetic (DNase-seq, H3K4me3, and H3K27ac) data sets in a systematic way.

In order to obtain a core TRN for a cell type of interest, first a set of 10 core TFs are identified by using a modified version of the statistical measure introduced by D'Alessio et al. [59]. Next, all the neighboring non-core TFs are identified, which are strongly connected to the core TFs. Further, all the candidate TFs are discretized by using a modified version of RefBool [127] and only those TFs are kept in the network that are expressed according to their measured gene expression levels. Finally, regulatory interactions among the selected TFs in the network are obtained by integrating experimental TF ChIP-seq data with cell-type-specific active regulatory regions, which are identified by histone modification marks and chromatin accessibility data sets. Moreover, based on the regulators of every TF in the network, the joint regulatory effect of TFs is inferred for every individual regulatory region and the resulting network is represented in a Boolean modeling framework. The obtained core TRNs are highly cell-type-specific as they comprise only those interactions that are compatible with both epigenetic and transcriptomic layers of regulation. We used INTREGNET to reconstruct directed core TRNs for 48 cell types and cell lines. Every network has up to 33 TFs (on average 18 TFs), while every TF in the network has up to 30 regulators (on average 11 regulators per TF) and 44 active enhancers (on average 9 enhancers per TF).

3.4.2 Validation of the reconstructed core TRNs

A bottleneck in the reconstruction of TRNs is to systematically benchmark them in the presence of incomplete ground truth data. To this extent, the large-scale generation of high throughput ChIP-seq data for various TFs across different cell and tissue types has provided a framework for partial network validation [176]. We have reconstructed cell-type-specific core TRNs for 8 different cell type/lines and validated them by using 1044 TF ChIP-seq experiments obtained from ENCODE [54] and Cistrome [186]. Here, only cell types/lines containing more than 10 profiled TFs were considered for validation in order to cover a significant part of the network.

The proposed networks are benchmarked by assessing the enrichment of cell-type-specific experimentally validated TF ChIP-seq interactions in the core TRNs (see Figure 4). On the one hand, network interactions that have been validated by cell-type-specific TF ChIP-seq data are considered as true positives (TP). On the other hand, interactions that have been validated by TF ChIP-seq data, profiled in cell-types other than the one under consideration, are considered as false positives (FP). The enrichment of ChIP-seq validated TP interactions was on average four times higher in comparison to the FP interactions, which shows that interactions in the reconstructed TRNs are highly cell-type-specific.

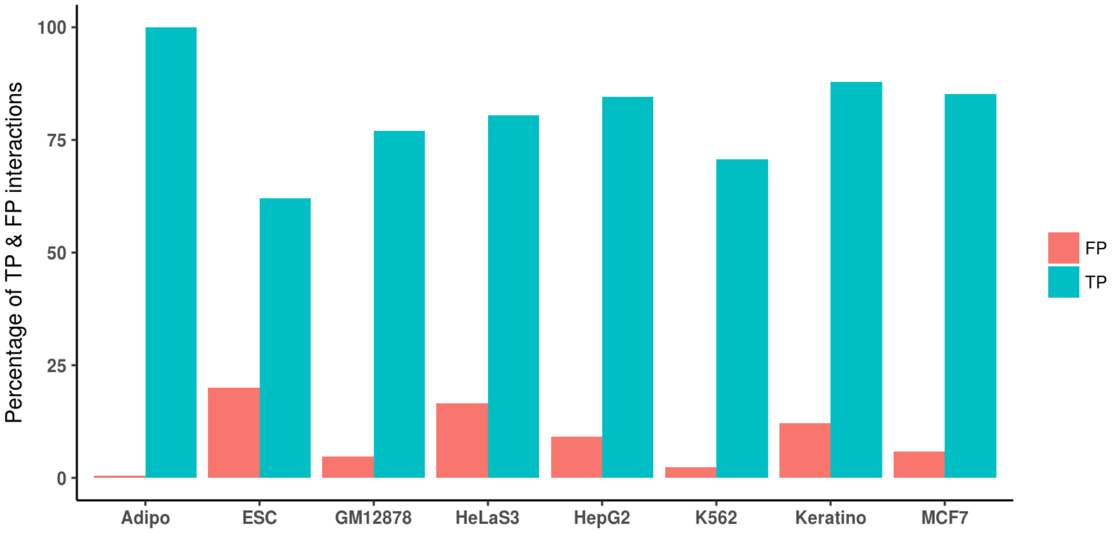


Figure 4: **Enrichment of cell-type-specific TF ChIP-seq data in reconstructed core TRNs.** Core transcriptional regulatory networks (TRNs) for different well-studied human cell types/lines have been benchmarked against cell-type-specific TF ChIP-seq data. True positives (TP) represent the interactions that are present in the reconstructed core TRNs and have been experimentally validated by cell-type-specific TF ChIP-seq data. Alternatively, interactions that have been validated by TF ChIP-seq data, profiled in cell-types other than the one under consideration, are considered as false positives (FP). Benchmarking is carried out for various primary cell types, e.g. adipocytes (Adipo), embryonic stem cells (ESC), keratinocytes (Keratino), and cancerous cell lines, e.g. GM12878, HeLaS3, HepG2, K562 and MCF7

.

We also validated the specificity of the reconstructed core TRNs by comparing them against experimentally verified gold-standard (GS) core networks. For this purpose, core networks of human embryonic stem cells (ESCs) [27], hepatocytes [209], and HepG2 and MCF7 cell lines [82] were curated from the literature and compared to the core TRNs reconstructed with INTREGNET (see Table 1). Here, an intersecting part of the GS and reconstructed core networks has been considered for the validation. As expected, the reconstructed networks for ESCs, hepatocytes and MCF7 cells are in complete agreement with respective GS networks, whereas only one interaction was missing in the hepatocytes network. Surprisingly, we inferred four new interactions for *HNF1A* and *FOXA2* in the hepatocytes reconstructed network that are missing in the respective GS network. Interestingly, all the newly inferred interactions have been validated by a previous TF knock-down study conducted in human hepatoma cells [276]. Overall, 95% of the interactions in the reconstructed networks are also present in the corresponding GS core networks, with all of the inferred interactions being experimentally validated (Table 1 and supplementary Table S16).

These results suggest that the reconstructed networks are in a good agreement with experimentally validated core networks and, therefore, can be considered as a starting point for the identification of IFs.

Cell type	GS int.	Infer int.	Matching	Non-matching GS	Unique infer	Infer validated	Overall validated
ESC[27]	9	9	9	0	0	0	100%
Hepatocytes[209]	13	12	8	2	4	4	85.71%
HepG2[82]	16	16	16	0	0	0	100%
MCF7[82]	13	4	4	0	0	0	100%

Table 1: **Benchmarking of reconstructed core TRNs** against the experimentally validated core networks. For the four well-characterized human cell types/lines, the reconstructed core networks were compared against their experimentally validated gold-standard core networks. Int. represents interactions (Int.) in gold-standard (GS) networks, whereas Infer Int. represent inferred (Infer.) interactions in the reconstructed networks.

Next, we assessed whether INTREGNET can distinguish cooperative and competitive regulation of TFs in the same regulatory region, represented in the form of a Boolean logic rule. As expected, INTREGNET was able to predict the complexes of TFs that have been experimentally verified in different human cell types. For examples, a complex of *POU5F1* and *SOX2*, along with *NANOG* has been shown to collectively regulate their own expression in ESC [27]. Interestingly, aside from predicting this complex, INTREGNET was also able to highlight its cooperative role in regulating many other TFs in the ESC network, as indicated in existing literature [27, 242]. Similarly, INTREGNET was able to predict a complex of *FLI1, TAL1* and *GATA2* that has been shown to regulate the expression of *FLI1* in blood stem cells [225], and a part of this complex (*TALI1-GATA2*) has also been experimentally verified using two-hybrid yeast assays [215]. Therefore, these findings suggest that representation of reconstructed TRNs in a Boolean modeling framework and inference of Boolean logic rules can offer a simple, yet powerful, approach to model the dynamics of regulatory networks.

3.4.3 Prediction of instructive factors for cellular conversions

The integration of epigenetic and transcriptional data enabled the reconstruction of cell-type-specific core TRNs, which recapitulate the genome-wide connectivity between core TFs and their cooperative or competitive regulatory effect on the enhancers and promoters of their respective targets. These reconstructed TRNs are able to provide a mechanistic insight into the global functions of these key regulators in controlling cell identity. Therefore, the underlying regulatory network information enabled us to prioritize an optimal combination of TFs in the core TRN, which are crucial to establish and maintain its cell-specific gene expression program. Moreover, by considering the epigenetic state and gene expression levels of those TFs in the starting cell type, we were able to faithfully predict a particular set of IFs required for any desired cellular transition among different cell types.

Final cell type	Initial cell type	Combination of IFs
Hepatocytes	Fibroblast	*HNF1A, HNF4A, ONECUT1, CEBPA, ATF5, PROX1, TP53-siRNA, MYC*
		HNF1A, HNF4A, FOXA3
		FOXA2, HNF4A, CEBPB, MYC
		FOXA2, HNF4A, CEBPB
iPSC	Hematopoietic stem cell	*POU5F2, SOX2, KLF4*
		POU5F2, SOX2
	Fibroblast	*POU5F2, SOX2 KLF4*
		POU5F1, SOX2
		POU5F1, SOX2, LIN28A, NANOG
	Keratinocyte	*POU5F1, SOX2*
		POU5F1, SOX2 KLF4
		POU5F1, SOX2 KLF4, MYC
	Neural stem cell	*POU5F1*
		POU5F1, KLF4
Neural stem cell	Hematopoietic stem cell	*SOX2*
	Foreskin fibroblast	*CBX2, HES1, ID1, TFAP2A, ZFP42, ZNF423*
		ZNF521
	Fibroblast (NHDF)	*SOX2, PAX6*
Neuron	Embryonic stem cell	*NEUROG2*
		POU3F2, ASCL1, MYT1L
	Fibroblast	*POU3F2, ASCL1, MYT1L*
		POU3F2, ASCL1, NEUROD1
Myoblast	Fibroblast	*MYOD1*
Melanocyte	Fibroblast	*MITF, PAX3, SOX10*
	Keratinocyte	*MITF, LEF1, SOX10, SOX9*
		MITF, LEF1, SOX10, SOX9, PAX3, SOX2
Adipocyte	Mesenchymal stem cell	*CEBPB*
		PPARG
		CEBPB, PPARG

Table 2: **Enrichment of predicted instructive factors (IFs)** in experimentally validated combinations. Predicted IFs are highlighted in red whereas TFs that were replaced by another validated IF are highlighted in blue.

Next, we asked whether INTREGNET can distinguish between a more and less efficient set of IFs required to obtain a desired cellular transition. A thorough comparison against different experimentally tested cellular transitions revealed that INTREGNET was able to successfully predict IFs in most cases. We examined a total of 32 cellular conversion experiments with defined factors and compared them with the predictions of INTREGNET and two former approaches, Mogrify [232] and d'Alessio [59]. In particular, we collected examples of cellular conversions to neural stem cells (NSC), hepatocytes, iPSCs, neurons, myoblasts, melanocytes and adipocytes from various initial cell types and assessed the enrichment of predicted IFs in the experimentally validated combinations. In most cases, i.e. for hepatocytes, myoblasts, NSCs and iPSCs, INTREGNET shows an increased enrichment compared to previous approaches (Figure 5A, Table 2). In particular, on average, more than 91% of IFs for inducing pluripotent stem cells were identified by INTREGNET when compared to 74% and 44% with Mogrify and d'Alessio, respectively. Notably, the predictions for iPSCs did not include *KLF4* in most cases but instead contained *PRDM14*, a TF that has been shown to replace *KLF4* while yielding higher conversion efficiency [45]. Most predicted combinations of IFs also contained *MYC*, one of the originally proposed inducers of pluripotency, while at the same time suggesting the over-expression of *MYCN*, another TF of the basic helix-loop-helix family that has been shown to contribute to the induction of pluripotency [202]. This suggests that INTREGNET not only predicts IFs of cellular conversions, but preferentially selects TFs yielding higher conversion efficiency. In contrast to INTREGNET's better performance in four distinct target cell types, almost no validated TFs have been predicted for adipocytes (0%), neurons (0%) and melanocytes (17%). Since INTREGNET only reconstructs core TRNs, we investigated the regulatory relationships of core TFs and known IFs to test the hypothesis that the predictions constituted upstream regulators of these TFs.

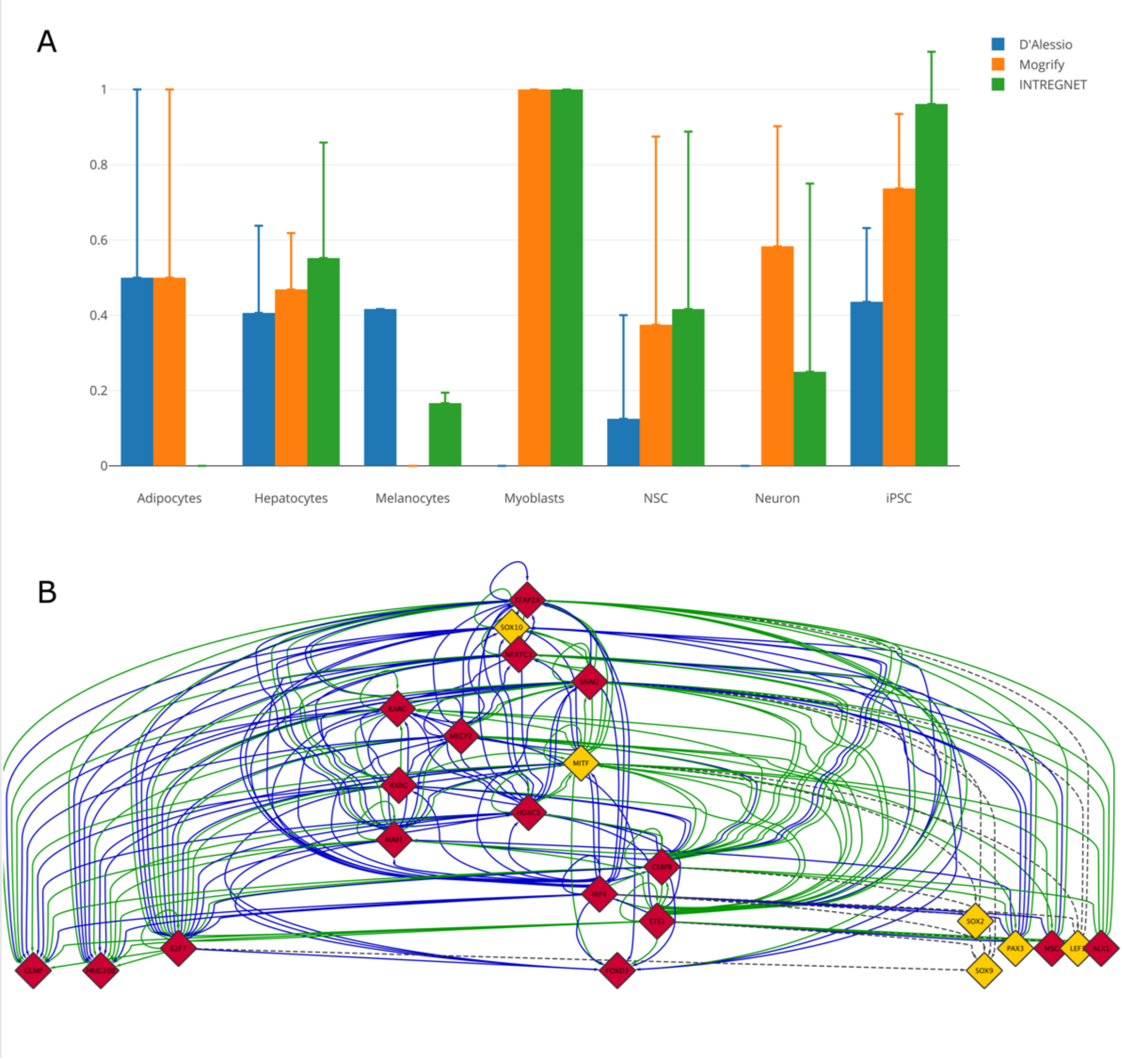


Figure 5: **INTREGNET performance.** A) Recovery of experimentally validated instructive factors (IFs) for seven final cell types. The bar heights corresponds to the average percentage of recovered factors. Error bars represent the standard deviation. B) Melanocyte core TRN including all experimentally validated IFs (yellow). Enhancer and promoter regulation (green) is distinguished from enhancer-only regulation (blue) and inferred promoter-only regulation for the IFs that were not in the reconstructed core TRN (black,dashed).

.

Indeed, we found that *SOX9*, *PAX3*, *SOX10*, *MITF*, *SOX2*, and *LEF1* were actively regulated in their promoter regions by at least one of the predicted IFs in melanocytes. For the conversion of fibroblasts, only two TFs, *IRF4* and *TFAP2A*, regulated the promoter regions of all known instructive factors and have been shown to co-regulate important loci for melanocyte differentiation [252]. In contrast, these factors have not been predicted when the initiating the conversion from keratinocytes, since they are already expressed. The known instructive factors were, thus, differentially regulated by predicted IFs. Particularly the TFs *SOX9*, *MITF* were regulated by at least one predicted IF and are able to propagate their effect to all validated conversion factors (Figure 5B). Of all known TFs, only *PAX3* has been predicted to induce the conversion of keratinocytes to melanocytes and is assumed to be an important co-factor of *MITF*, due to their co-localization [252]. Similarly, in adipocytes, we identified the promoter region of *CEBPB* to be actively regulated by the predicted IFs *ATF3, SMAD3, KLF11, E2F1* and *MITF*. In turn, previous studies already demonstrated that *CEBPB*, among other TFs, transcriptionally activates *PPARG* in adipocytes [153]. Unlike in melanocytes and adipocytes, no direct downstream regulatory effects could be identified for *ASCL1, POU3F2* and *MYT1L* in neurons. Despite the missing regulatory links of the predicted instructive factors and *ASCL1, POU3F2* and *MYT1L*, INTREGNET identifies combinations yielding increased conversion efficiency. Over-expression of *NEUROG2* in embryonic stem cells was reported to produce a nearly pure neuron population [310], while other combinations resulted in substantially lower conversion efficiency [218].

3.4.4 INTREGNET increases the efficiency of iPSC generation

With regard to our finding that INTREGNET identifies factors yielding increased conversion efficiencies, we investigated whether the scores assigned to combinations are prognostic for the efficiency of the cellular transition. Here, especially the ability of TF-based cellular conversions for reprogramming somatic cells into iPSCs has provided an avenue for obtaining patient-specific cell types that can help in modeling human diseases [173]. Taking into consideration the importance of converting somatic cells into iPSCs, and the wealth of data showing different conversion efficiencies depending on the combination of TFs and the initial cell population [88, 114], this classical reprogramming system serves as a suitable system to demonstrate the versatility of INTREGNET.

We leveraged a collection of publicly available experimental datasets measuring the efficiency of iPSC reprogramming with various combinations of factors in four distinct cell types [88, 114, 132, 280] and assessed whether INTREGNET can distinguish more and less efficient combinations, i.e. ranking more efficient conversions higher. For that purpose, a core iPSC network was reconstructed for computing the scores of perturbations as previously described (Figure 6). Before investigating the association between scores and cellular conversion efficiency, we inspected the network more closely to assess its descriptive and dynamic quality. Apparently, apart from

LIN28A, all known inducers of iPSC induction, i.e. *NANOG*, *MYC*, *POU5F1*, *SOX2*, *KLF4*, *PRDM14* and *MYCN*, are present in the network. Importantly, no perturbation of core TFs can propagate through the complete network, if it does not contain *POU5F1*. Therefore, the network model resembles previous experimental findings emphasizing that *POU5F1* is indispensable for the generation of iPSCs. In addition, the core TRN contains *FOXH1*, *TP53*, *ZNF423* and *MTA3*, which are known to play diverse roles in the conversion to pluripotent stem cells. For example, a previous study revealed that *FOXH1* significantly enhances iPSC conversion efficiency [269], whereas the roles of *ZNF423* and *MTA3* as transcriptional regulators are not yet understood. On the other hand, *TP53* is a known repressor of PSC induction [159, 312], but has been shown to play important roles in the maintenance of embryonic stem cells [159, 277].

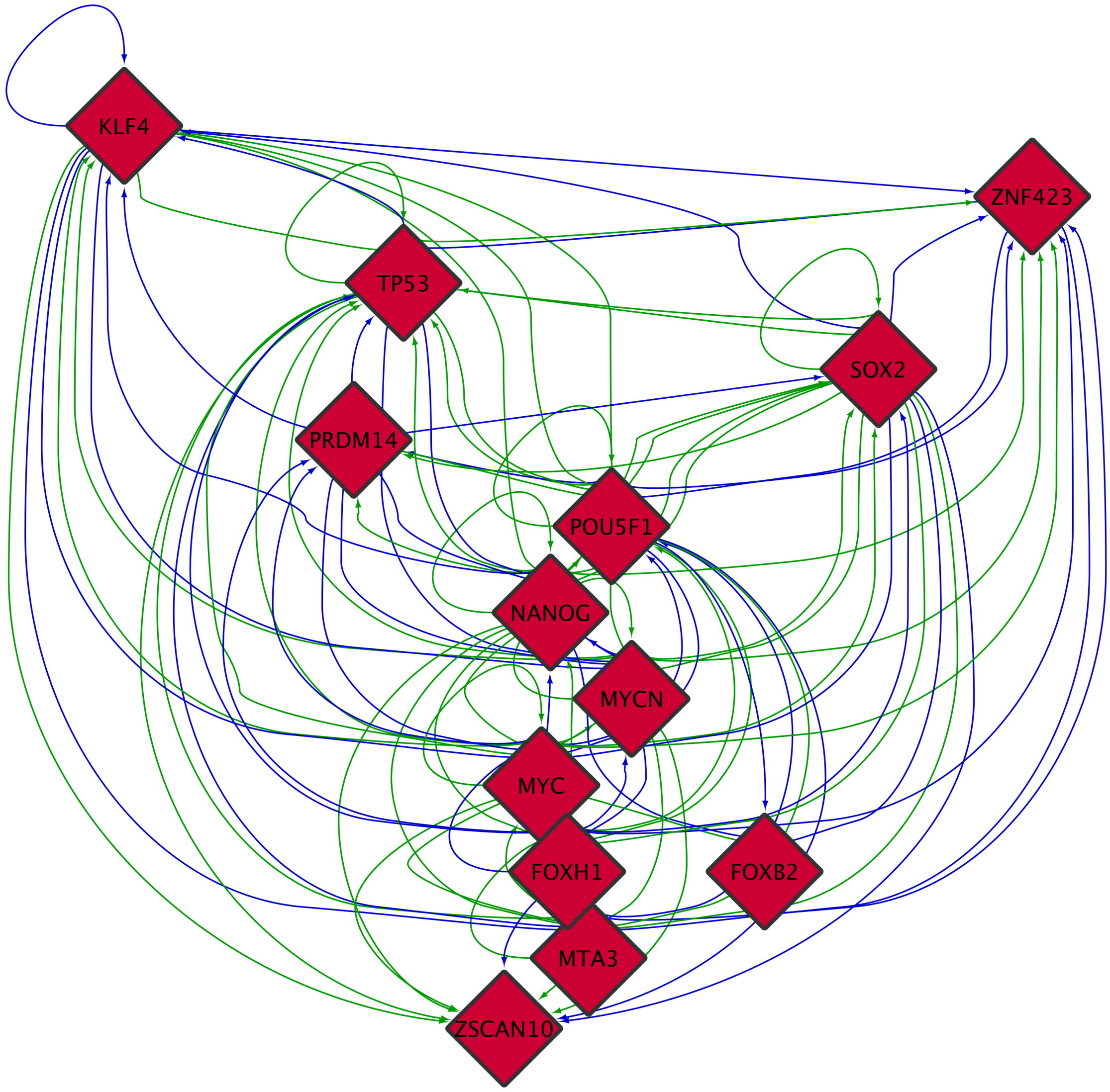


Figure 6: **Reconstructed core TRN of induced pluripotent stem cells.** Enhancer and promoter regulation (green) is distinguished from enhancer-only regulation (blue). No regulatory interaction has been inferred that only regulates the promoter region of a TF.

.

Due to the dual role of *TP53*, we examined whether the suppression of iPSC conversion is reflected in the dynamics of the network model. In particular, we identified that combinations including *TP53* yield lower scores than those not containing it (Wilcoxon-Mann-Whitney test, p-value $<2.2e^{-16}$). Of note, this result is valid regardless of the initial phenotype and, thus, resembles the experimental findings. The support for the qualitative and dynamic validity of the reconstructed iPSC network led us to investigate combinations of IFs yielding high and low conversion efficiency. We collected four conversion examples from different initial cell types, i.e. neural stem cells (NSCs) [132], hematopoietic stem cells (HSCs) [187], and newborn and adult fibroblasts (Fibroblast, NHDF) [114]. In contrast to the predictions described in the previous section, only known combinations were considered and ranked by their score. Strikingly, each dyad of more efficient combinations was correctly predicted, which provides evidence for INTREGNET's ability to predict more efficient combinations of IFs (Figure 7A). While the differences in scores appear to be negligible, the average number of steps taken until the effect of the perturbation propagated to the complete network model is substantially different. Notably, the scores obtained by INTREGNET are not only consistent within a given initial cell type but also across different initial conditions. *POU5F1, SOX2* and *KLF4* have been utilized for inducing PSCs from HSC as well as newborn and adult fibroblasts. Experimental evidence suggests that the reprogramming efficiencies, calculated as the percentage of formed iPSC colonies per 10^5 cells, is rather low for fibroblasts (newborn: 0.05, adult: 0.045) and is significantly elevated in HSCs (0.2). These results are confirmed by the ranking based on the scores obtained by INTREGNET (Figure 7A). Thus, the scores indicate that more plastic cells, such as neural and hematopoietic stem cells, achieve higher conversion efficiencies even though fewer factors are perturbed. Especially neural stem cells show higher epigenetic similarity with PSCs, thus requiring less restructuring of the chromatin for activating the transcriptional core network, which is reflected in a higher overlap of active enhancers specific to induced pluripotency (Figure 7B).

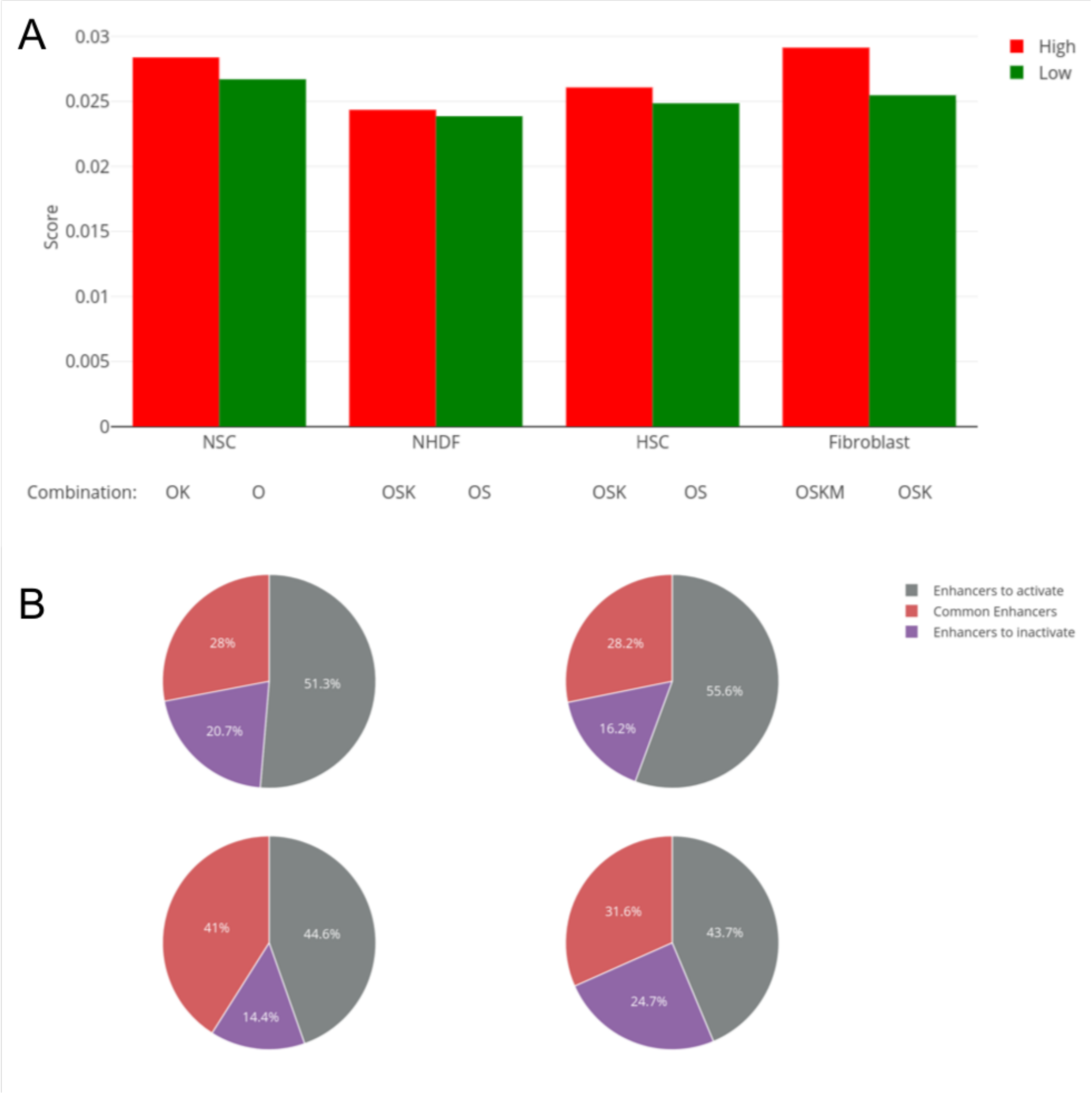


Figure 7: **Reprogramming efficiency for inducing PSCs.** A) Predicted reprogramming efficiency for inducing pluripotent stem cells from different initial cell types using several combinations of IFs (O = *POU5F1*, S = *SOX2*, K = *KLF4*, M = *MYC*). Two cell lines of adult (NHDF) and newborn fibroblasts (Fibroblast (BJ)) were included. Combinations with experimentally validated increased efficiency (red) were compared against low efficiency combinations. B) Distribution of enhancer landscape changes of HSCs (top-left), Fibroblasts (top-right), NSCs (bottom-left) and NHDFs (bottom-right) required for compatibility with iPSCs. Enhancer landscapes were restricted to TFs included in the core TRN and related TFs they regulated by INTREGNET.

.

3.5 Discussion

A major bottleneck in regenerative medicine is the efficiency of induced cellular conversions, which hampers the translation of therapeutic interventions into clinical applications. While computational methods have been developed to identify IFs of desired cellular conversions [59, 198, 243, 38], none of them is able to systematically prioritize IFs with increasing conversion efficiencies. In view of the interplay between epigenetic and transcriptional regulation to maintain and switch between cellular phenotypes [157, 150, 81], it is important to consider epigenetic information for predicting efficient IFs. Nevertheless, current computational approaches solely rely on transcriptional regulation, which constitutes an important limitation.

In this study, we developed INTREGNET, a computational framework that predicts efficient combinations of IFs for desired cellular conversions. The method is based on the systematic integration of epigenetic and transcriptomic information to reconstruct core TRNs, offering several advantages over current approaches. Firstly, it exclusively relies on experimental data for TRN reconstruction, which increases precision compared to position weight matrix-based methods that are not cell-type-specific. In particular, INTREGNET introduces cell-type-specificity by integrating information on TF ChIP-seq experiments, chromatin accessibility and active cis-regulatory elements to accurately reconstruct networks. Secondly, integration of protein-protein interaction (PPI) data allows for dissecting region-specific cooperative and competitive TF-binding, i.e. the joint effect of multiple TFs on the transcription of target genes. Considering these protein-protein interactions is critical for prioritizing more efficient combinations of IFs, exemplified by the complex formation of *SOX2* and *POU5F1* that is necessary for inducing pluripotent stem cells [242, 27]. Finally, the devised strategy for predicting efficient IFs actively incorporates differences in the epigenetic landscape between the initial and target cell type. Despite the specific combination of IFs, the amount of epigenetic restructuring required during reprogramming is a key determinant of cellular conversion efficiency [219]. INTREGNET accounts for these epigenetic landscape differences by penalizing the calculated efficiency of IFs with the amount of required restructuring.

Despite the advantages of INTREGNET, we acknowledge that it has certain limitations, suggesting potential future improvements. In this regard, a common problem of gene regulatory network reconstruction approaches is missing data. In particular, binding site information of certain TFs is currently unavailable, even though our method leverages a comprehensive compendium of over 11,000 publicly accessible TF ChIP-seq profiles. For example, *LIN28A* was identified as a core TF of iPSCs, but its binding sites have not been profiled. As a consequence, it cannot be contained in the core TRN and predicted as an IF for inducing PSCs. However, the amount of available

TF binding site profiles is steadily increasing, which eventually will mitigate this problem in the future. Moreover, the availability of additional epigenetic profiles, such as multiple histone modifications and chromatin conformation, will become greater in the future, opening the possibility of integrating them into the TRN. Another important limitation is that INTREGNET relies on bulk datasets. Indeed, transcriptomic and epigenetic heterogeneity in cellular populations can influence successful conversion due to the existence of different sub-populations exhibiting distinct conversion efficiencies [31]. In this regard, modeling core TRNs using single-cell data could allow the identification of sub-populations with the highest conversion propensity. Furthermore, single-cell data can help in devising novel experimental strategies for cellular conversion, such as initially priming cell populations and subsequently inducing the desired cell type conversion.

In principal, INTREGNET can be customized for applications for human disease modeling, in view of diseases as network perturbations from healthy to disease phenotype [62]. A core TRN reconstructed from different epigenetic and transcriptional profiles obtained from pathological cells might help in identifying causal TFs that establish or maintain the disease phenotype. Finally, *in silico* network perturbations can guide experimental efforts in pre-selecting a set of putative target TFs, whose perturbation induces the conversion into a healthy phenotype, with vast amounts of potential applications to personalized medicine.

To our knowledge, INTREGNET is one of the first approaches that aims at identifying highly efficient IFs based on the systematic integration of information linked to multiple regulatory levels, and is expected to find diverse applications in the field of regenerative medicine. In particular, considering the success of *in vivo* reprogramming in preclinical models, we believe INTREGNET to be a valuable tool for alleviating the impediment of low efficiency by guiding cellular conversion experiments.

Chapter 4

Identification of causal genes for Alzheimer’s disease using a network-based integrative analysis of genomic, epigenomic and transcriptomic data

Muhammad Ali [A,B], Roy Lardenoije [B], Janou A.Y. Roubroeks [B], Katie Lunnon [C], Diego Mastroeni [D], Paul D. Coleman [D], Jos Kleinjans [B], Antonio del Sol [A], Daniel L.A. van den Hove [B], Ehsan Pishva [B,C].

[A] Computational Biology Group, Luxembourg Centre for System Biomedicine (LCSB), University of Luxembourg, Luxembourg.

[B] School for Mental Health and Neuroscience (MHeNS), Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, the Netherlands.

[C] University of Exeter Medical School, University of Exeter, Exeter, UK.

[D] Biodesign Institute, Arizona State University, Tempe, AZ, US.

Ali M., et al. (2019) Identification of causal genes for Alzheimer’s disease using a network-based integrative analysis of genomic, epigenomic and transcriptomic data. *In preparation*.

4.1 Abstract

Recent evidence suggests that changes at multiple levels of genomic regulation, including those linked to genetic variation, DNA methylation, and gene expression, are involved in the development and course of Alzheimer's disease (AD). While the heterogeneous and multifactorial nature of AD requires the integration of regulatory information from different -omics levels in order to accurately capture the mechanisms underlying its pathogenesis, systematic analytical approaches for identifying multi-omics signatures of AD are still lacking. Here, we applied a novel approach for systematically integrating genomic (gene variation), epigenomic (DNA methylation) and transcriptomic data obtained from the middle temporal gyrus (MTG) of AD patients and age-matched controls. This method uses information about AD-associated genetic and epigenetic variation in upstream regulatory genes affecting intermediate (mediator) genes, which, through gene-gene interactions, in turn, affect proximal downstream genes evoking expression changes. In depth analysis of top-ranked genes revealed a strong connectivity between their subnetworks, providing important insights into interconnected dependence of these genes at different regulatory levels. Interestingly, some of the top-ranked genes (*ETS1, WT1, APP*) are well-known for their implication in the pathogenesis of AD, validating the potential of the proposed approach in recapitulating existing knowledge as well as in predicting novel candidate genes. Thus, the presented approach has the capacity to provide more insight in the underlying mechanisms of complex disorders like AD.

4.2 Introduction

Alzheimer's disease (AD), the most common form of dementia, affects about 30% of those aged over 85 years [79]. AD is classified as a neurodegenerative disease, impacting on a patient's brain integrity and functioning, eventually resulting in a progressive deterioration of cognitive capabilities [30]. Despite decades of research, an effective treatment for AD is still lacking. In recent years, numerous major pharmaceutical companies have terminated their drug development programs on AD, as related clinical drug trials failed, which is primarily attributed to the heterogeneous and yet unclear pathogenesis of AD [4, 279].

The remarkable development of high-throughput sequencing technologies has allowed the generation of great quantities of genomic, epigenomic and transcriptomic data for various human diseases that has allowed us to dissect the mechanisms behind the onset and progression of multifactorial diseases. As such, many studies have used information from an individual regulatory level to identify causal genes and understand the mechanisms underlying the pathophysiology of AD. For example, genome-wide association studies (GWAS) have successfully identified numerous susceptibility genes for AD [89]. Some prominent examples include SNPs associated to *APP* [125], *PSEN1* [130], and *PSEN2* [33] that have been implicated in early-onset of AD. Similarly, based on the crucial role of DNA methylation in cellular processes [214], including gene regulation [229], cellular differentiation [131] and genomic imprinting [221], there have been many studies linking changes in DNA methylation status to the pathogenesis of AD [290, 61]. For example, epigenetic alterations in the DNA methylation levels of *ANK1, BIN1*, and *RHBDF2* genes have been suggested to play a key role in the onset of AD [61]. Furthermore, analysis of genome-wide transcriptomic data sets from post-mortem brain tissue has unveiled various key genes in different biological pathways associated with AD, among which, for example, *TYROBP* and *SPI1*, have been implicated in the brain's immune response [286]. These findings highlight that changes associated with AD are not restricted to a particular regulatory layer and can be observed across genetic, epigenetic and transcriptomic levels in both brain and blood samples [147, 61, 179, 100, 109, 170].

Although various levels of genomic regulation, including DNA methylation, chromatin modifications and microRNAs (miRNAs), are known to be highly interconnected at the functional level [63], commonly used analytical approaches are usually restricted to analysing only one or two layers of molecular information in association with AD [61, 286, 107], and, moreover, are mostly restrained to correlations. Therefore, an integrative multi-omics systems biology approach to uncover the relative, interdependent contribution of various molecular layers in the development and course of AD is of utmost importance.

In recent years, several computational tools using network-based approaches have been developed in order to detect cancer-related genes by integrating information from different regulatory levels, i.e. genomic (genetic variation), epigenomic (DNA methylation), transcriptomic (gene expression), and proteomic levels. A few prominent examples include, DriverNet [14], HotNet2 [155], TieDIE [220], and NetICS [66] that use network diffusion algorithm to identify causative disease genes at epigenomic, transcriptomic and proteomic levels. However, so far, these approaches have not been applied to understand the mechanism underlying AD pathology and prioritize AD-associated genes.

In the present study, we have conducted a network-based integrative analysis of genetic, epigenetic and transcriptomic data sets derived from post-mortem middle temporal gyrus (MTG) tissue from AD patients and age-matched elderly controls (Lardenoije et al., Under Review). We have applied a bidirectional graph diffusion-based technique [66] to prioritize genes based on known AD-associated genetic variations, differential methylation and differential mRNA expression. This method uses a rank-aggregation technique for integrating diverse molecular data types within a directed functional interaction network. Our findings show a strong connection between sub-networks of top-ranked genes (*ETS1, TP63, ZNF217, WT1, IL15* and *APP*). The conducted analysis can explain how genetic and epigenetic variation can induce expression changes in other genes via gene-gene interactions. Furthermore, we used connectivity map database [145] to uncover functional connections between predicted top-ranked AD-associated genes and drugs that may revert their gene expression from AD towards the healthy phenotype. The drug enrichment analysis suggested a combination of levcycloserine and apramycin to be the most effective therapeutic treatment for AD in terms of normalizing AD-associated gene expression patterns.

Taken together, the conducted analyses allows a systematic dissection of mechanisms underlying the onset and progression of multifactorial diseases like AD at a multi-omics level, suggesting potential candidate genes and putative drugs that could be employed to target these genes. Thus, we are providing the scientific community with a novel approach that can pave the way for deconvoluting complex and multifactorial human diseases, hence fostering the development of novel treatment strategies.

4.3 Materials and methods

Multi-omics AD signatures within the MTG were identified by the integration of datasets from three different regulatory levels. Firstly, AD-associated SNPs identified by GWAS were retrieved from the International Genomics of Alzheimer's Project (IGAP) [147]. Secondly, methylation

(5-methyl-cytosine; 5mC) data were obtained from post-mortem MTG tissues of 46 AD patients and 32 elderly, non-demented controls. Lastly, for the same individuals, gene expression data within the MTG were obtained using Illumina Beadchip microarrays. An overview of our analysis pipeline is shown in Figure 8, while each individual step, i.e. the identification and annotation of SNPs, DNA methylation and gene expression data processing, gene-gene interaction network curation, and network diffusion analysis is described in detail in the remainder of this section.

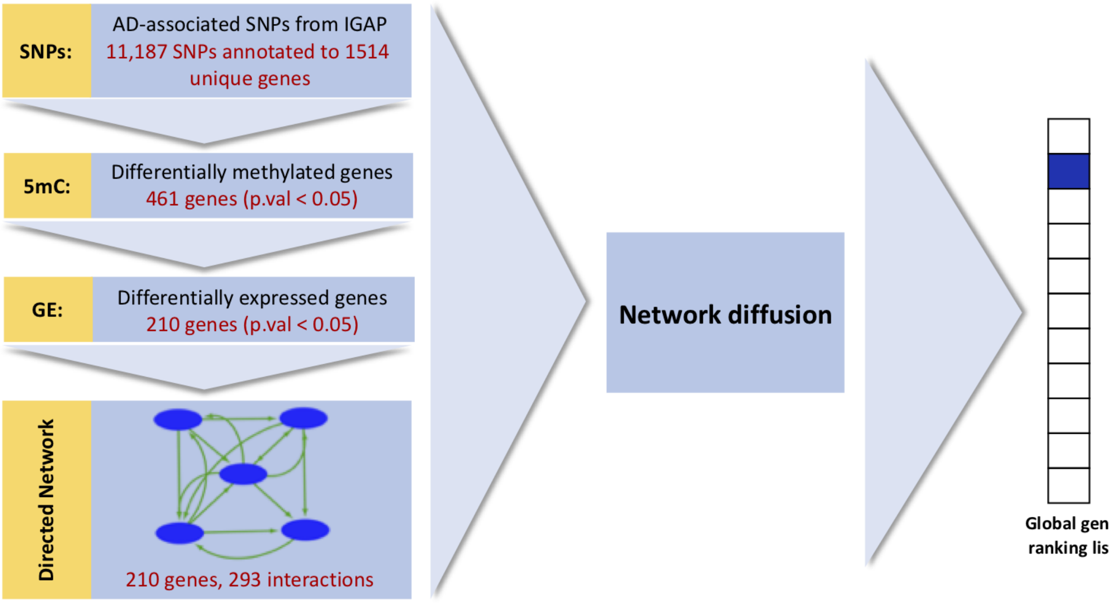

SNPs:
AD-associated SNPs from IGAP
11,187 SNPs annotated to 1514 unique genes
5mC:
Differentially methylated genes
461 genes (p.val < 0.05)
GE:
Differentially expressed genes
210 genes (p.val < 0.05)
Directed Network
210 genes, 293 interactions
Network diffusion
Global gene ranking list

Figure 8: **Schematic pipeline of the multi-omics approach used for ranking AD-associated genes.** AD-associated SNPs were obtained from a large, two-stage meta-analysis conducted in IGAP study. The combined analysis of two stages resulted in 11,187 SNPs showing moderate evidence of association (P-value ¡ 0.05) to AD. Annotating these SNPs to the human genome (hg19) resulted into 1,514 unique genes. Next, filtering significantly differentially methylated (P-value ¡ 0.05) probes for these 1,514 SNPs-associated genes resulted into 837 probes, annotated to 461 unique genes. Further filtering these genes for significant differential expression (P-value ¡ 0.05) resulted into 210 unique genes. Accordingly, we obtained 293 direct gene-gene interactions between these genes from the MetaCore database. By using the obtained network, and p-values of differentially methylated and expressed genes as an input for the network diffusion algorithm, we ranked these 210 genes based on their mediator effect.

4.3.1 Post-mortem tissue samples

The present study included donors from the Brain and Body Donation Program (BBDP) at the Banner Sun Health Research Institute (BHSRI), who signed an informed consent form approved by the institutional review board, including specific consent of using the donated tissue for future research [17, 18].

DNA was obtained from the MTG of 46 AD patients and 32 neurologically normal control BBDP donors stored at the Brain and Tissue Bank of the BSHRI (Sun City, Arizona, USA) [17, 18]. The organization of the BBDP allows for fast tissue recovery after death, resulting in an average post-mortem interval of only 2.8 hours for the included samples. A consensus diagnosis of AD or non-demented control was reached by following National Institutes of Health (NIH) AD Center criteria [18]. Comorbidity with any other type of dementia, cerebrovascular disorders, mild cognitive impairment (MCI), and presence of non-microscopic infarcts were applied as exclusion criteria. Detailed information about the BBDP has been reported elsewhere [17, 18].

4.3.2 SNP identification and annotation

AD-associated SNPs are identified by IGAP in a large, two-stage meta-analysis of GWAS in 25,580 AD and 48,466 control individuals of European ancestry [147]. According to the combined analysis of two stages, 11,632 SNPs showing moderate evidence of association (P-value $<$0.001) in stage 1 were followed up for subsequent association analysis in stage 2. We applied a P-value threshold of 0.05 to obtain those SNPs that have been found to be statistically significantly associated with AD in both stages of the IGAP study. Furthermore, we only kept those SNPs for which the direction of association (positive or negative) was the same and with the same "effect allele" in stage 1 and 2. Following these filtration criteria, we obtained 11,187 SNPs and annotated them to the hg19 genome by Homer `annotatePeaks.pl` in order to characterize their genomic annotation [105]. Annotating these SNPs to the nearest gene resulted into 1,514 unique genes.

4.3.3 Differential methylation (5mC) analysis

For differential methylation analysis, the 5mC data were obtained from an unpublished study from our group, where Illumina HM 450K arrays were used for quantifying the methylation status of 485,000 different human CpG sites. In view of the present study, we included those 5mC datasets for which corresponding gene expression profiles (see below) were also available, resulting in data derived from 46 AD patients and 32 age-matched controls. Pre-processing and analysis of the raw data sets was conducted in R (version 3.4.4) [272]. Raw IDAT files corresponding to the selected individuals were read into R using the wateRmelon "readEpic" function (version 1.20.3) [224]. The "pfilter" function from the wateRmelon package (version 1.18.0) [224] was used to filter data sets based on bead count and detection p-values. Background correction and normalization of the remaining probe data was performed by "preprocessNoob" function of minfi package (version 1.22.1) [9]. We used the MLML function within the MLML2R package [77] for estimating the proportion of uC, 5mC and 5hmC for each CpG site, based on the combined input signals from the BS and OxBS arrays. All of the cross-hybridizing probes and the probes that contained a SNP in the sequence were removed resulting into 407,922 probes to be considered for the differential methylation analysis [43].

Raw IDAT files corresponding to the selected individuals were loaded into R using minfi "read. metharray" function (version 1.22.1) [9] to make an RGset for computing the cell type composition of the samples by "estimateCellCounts" function of the same package. For estimating the cell composition, we used FlowSorted.DLPFC.450k package (version 1.18.0) [120] as the reference data for "NeuN_pos" cell composition of frontal cortex. The limma package (version 3.32.10) [241] was used to perform linear regression in order to test the relationship between the beta values of the probes and the diagnosis of AD. The used regression model considered beta values as outcome, AD diagnosis as predictor, and age, gender, and neuronal cell proportion as covariates. In order to identify significantly differentially methylated probes (DMPs), probes with unadjusted P-value less than 0.05 were considered for further analysis. Resulting probes were annotated using Illumina human UCSC annotation. We took the most significant probe as a representative of the methylation status of a gene. Filtering the 1,514 SNPs-associated genes for DMPs resulted in 837 probes, annotated to 461 unique genes.

4.3.4 Differential gene expression analysis

For differential gene expression analysis, Illumina HumanHT-12 v4 beadchip arrays were used. The brain tissue sample used for RNA extraction was identical as used for the methylation study. Pre-processing and analysis of the raw data sets was conducted in R (version 3.4.4) [272]. Raw expression data was log-transformed and quantile-quantile normalized. For computing the cell composition, the Neun_pos cell percentage was calculated from the methylation data. The same regression model used for assessing methylation was applied to the expression data where the effects of age, gender and cell type composition were regressed out using limma. Genes having nominal P-value less than 0.05 were included for further analysis. Filtering the remaining 461 genes for differential expression resulted in 210 genes for downstream analysis.

4.3.5 Gene-gene interaction network

We used MetaCore (Clarivate Analytics) to obtain directed functional interactions between genes both known to be genetically associated with AD, and differentially methylated and expressed according to our differential methylation and expression analyses, respectively. The MetaCore database contains a collection of manually curated and experimentally validated direct gene-gene interactions based on existing literature. This high level of manual curation ensures the creation of highly confident interaction network maps. In order to obtain a set of directed functional interactions among the selected genes, our analysis was restricted to "Functional interactions", "Binding interactions", and "Low trust interactions". The interaction network obtained from MetaCore contains a variety of different interaction types, including (transcriptional) regulation, (de)phosphorylation, binding, and influence on expression. The obtained interactions are directed, i.e. the source and target genes are known. Furthermore, the information about the interaction type (activation or inhibition) is also given where available.

4.3.6 Network-based integration analysis

A network diffusion-based algorithm was employed to understand the functional implications of genetic variations at both epigenomic and transcriptomic levels [66]. Functional gene-gene interaction network, AD-associated genetic variations, as well as the differentially methylated and differentially expressed genes with respective p-values, were provided as inputs to the network-diffusion algorithm. The underlying hypothesis is that genetic variation in upstream genes affect intermediate (mediator) genes, which in turn affect proximal downstream genes evoking significant expression changes. The diffusion of information from upstream aberrant genes towards mediator genes and, eventually, in downstream differentially expressed genes, relies on the directionality of the provided functional network interactions. Accordingly, network diffusion was used to obtain a ranked list of AD-associated genes based on their potential being a mediator gene and evoke changes in gene expression.

4.3.7 Drug enrichment analysis

Connectivity map [145] was used to check for the enrichment of drug target genes in the subnetworks of top ten ranked mediator genes identified by the network diffusion. For every subnetwork, we obtained a signature of up- and downregulated probes based on the fold changes of the respective genes in the differential expression analysis. The obtained signatures were used as an input for the connectivity map to identify drugs that are known to induce opposite gene expression profiles. As such, Connectivity map gives a ranked list of drugs based on their enrichment in the provided query gene expression signature. The most negatively enriched (correlated) drugs, i.e. those inversely correlating to the diseased (AD) gene expression signature, were chosen as candidate drugs.

4.4 Results

4.4.1 Prediction of AD-associated genes by network diffusion

Existing studies relying on single regulatory levels have been able to identify AD-associated abnormalities at the genomic [89, 125, 130, 33], epigenomic [290, 61] and transcriptomics level [286]. However, a thorough understanding of the cross-talk between these interconnected layers of regulation is still missing which is essential for uncovering the mechanisms underlying the pathogenesis of AD. Therefore, here we used a network diffusion approach that integrates regulatory information from genomic, epigenomic and transcriptomic layers to rank key genes based on their ability to evoke disease-associated transcriptional changes. Based on the input information from different regulatory levels and functional gene-gene interaction networks, the genes are ranked according to their predicted involvement in AD. The top thirty ranked genes that are predicted to be associated with AD are shown in the Table 3.

Rank	Genes	Rank	Genes	Rank	Genes
1	*ETS1*	11	*MTA3*	21	*MN1*
2	*TP63*	12	*CAV1*	22	*PLAGL1*
3	*ZNF217*	13	*OPRM1*	23	*SATB2*
4	*WT1*	14	*NR1H3*	24	*CLU*
5	*IL15*	15	*PRKD1*	25	*HLX*
6	*FHL2*	16	*ACTB*	26	*WWOX*
7	*APP*	17	*ZFPM2*	27	*HDAC9*
8	*EPAS1*	18	*ARHGEF7*	28	*RGS4*
9	*SMARCA2*	19	*CUX2*	29	*ETV6*
10	*RXRA*	20	*OLIG2*	30	*ABCA1*

Table 3: **Top 30 ranked AD-associated genes identified by network diffusion.**

4.4.2 Subnetwork of top-ranked AD-associated genes

With regard to our finding that network diffusion identifies key genes associated to AD pathogenesis, we investigated whether the predicted genes are isolated or densely connected to each other via regulatory interactions. A graphical illustration of the directed functional interactions used as an input for the network diffusion showed that all the top-ranked mediator genes either directly regulate each other or regulate upstream genes that are significantly differentially methylated and regulate other mediator genes. An illustration of gene-gene interactions in the form of a subnetwork of the top-ranked genes is shown in Figure 9.

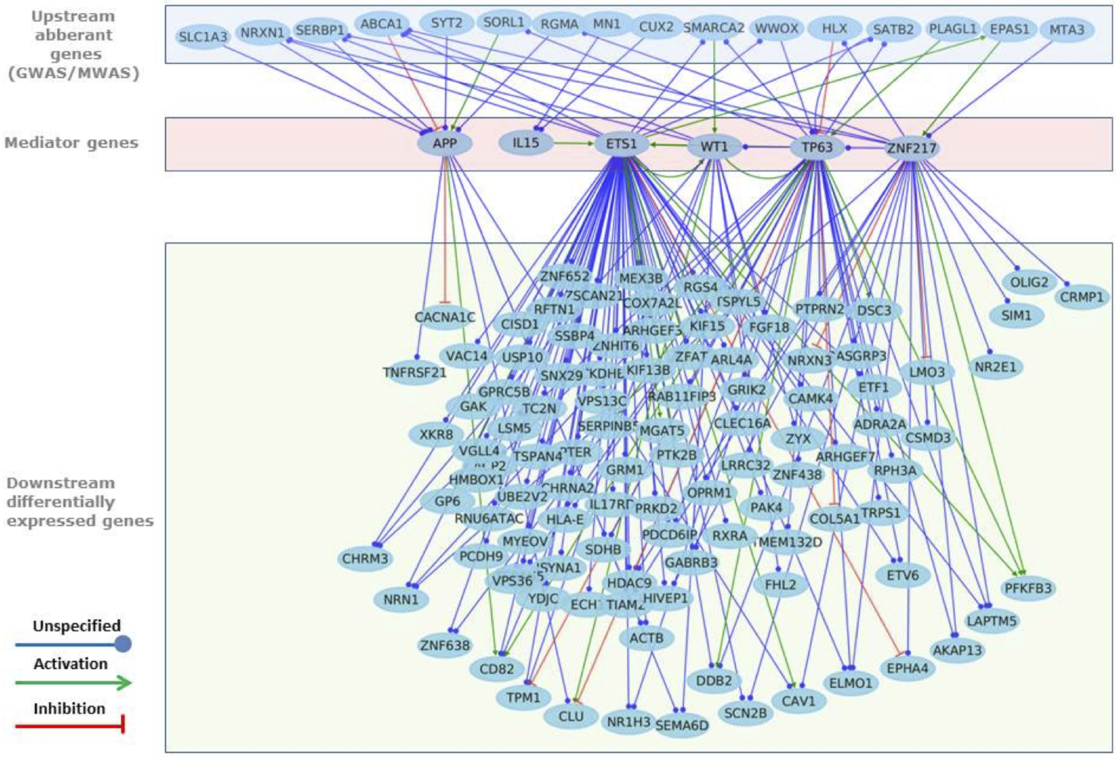


Figure 9: **Schematic illustration of top-ranked AD-associated genes subnetwork.**

As described earlier, *ETS1* is one of the most important mediator genes that has been implicated in AD [121] and it is ranked first according to our predictions. It is one of the most highly interconnected genes in the network that is regulated by three other top-ranked upstream genes (*TP63, WT1* and *IL15*), which are significantly differentially methylated. Furthermore, genetic variation in *ETS1* is known to be implicated in various neurodegenerative diseases including AD [188, 169, 239]. Interestingly, this mediator gene has the highest out-degree of 83, which means this gene regulates approximately one-third of all significantly differentially expressed downstream genes in the network. These results highlight the ability of this mediator gene to evoke changes in the gene expression program once perturbed by genetic variation and/or differential methylation. The individual subnetworks of the top-ranked mediator genes are shown in Figure 10 which allow us to take a thorough look into the epigenetic and transcriptional regulation of these genes.

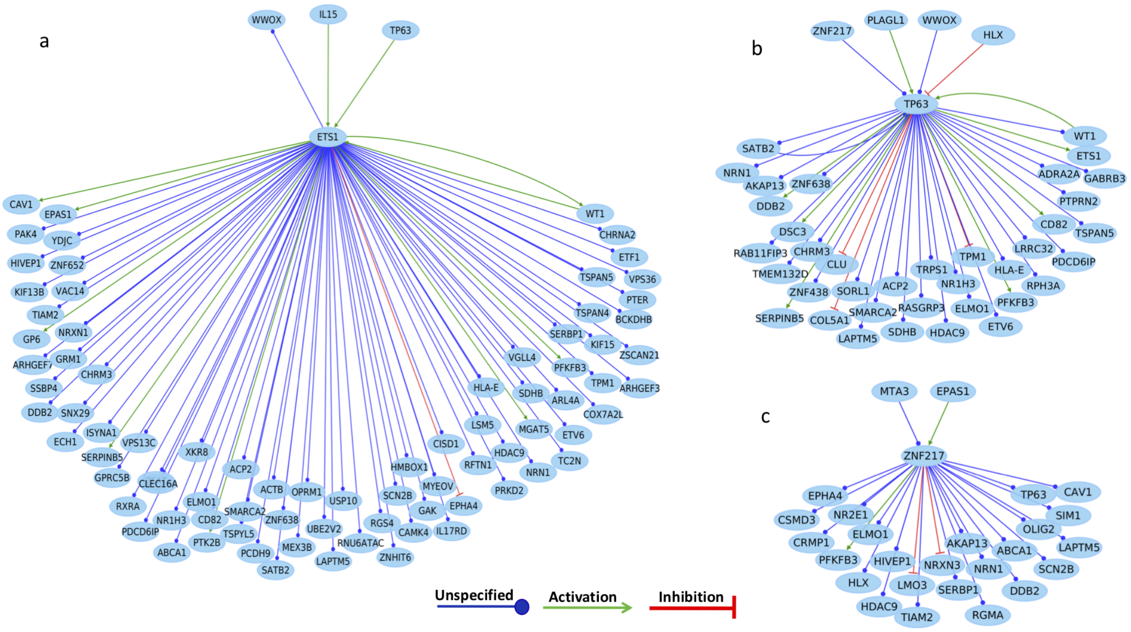


Figure 10: **Top ranked AD-associated genes subnetwork.** a) *ETS1*, b) *TP63*, and c) *ZNF217* subnetworks.

4.4.3 *WT1* as a mediator gene

As a proof of concept, we examined Wilms tumor suppressor (*WT1*), a top-ranked mediator gene that is predicted to be implicated in AD pathogenesis, in more detail. *WT1* has been known to be involved in different cellular processes including proliferation, differentiation, and apoptosis. Laser confocal microscopy and gene expression analysis of cultured hippocampal neurons from a mouse model for AD revealed a strong correlation between *WT1* expression and apoptosis induced by amyloid beta exposure(Abeta) [169]. In the same study, Lovell and co-workers observed that a reduction in *WT1* expression levels by blocking its transcription using an antisense oligonucleotide was accompanied by a significant decrease of neuronal apoptosis in Abeta-treated cultures. This study confirms the key role of *WT1* in mediating neuronal degeneration associated with the pathogenesis of AD. These observations are in line with our differential methylation and expression analyses where a significant hypomethylation (logFC: -0.006, P-value: 0.022) has been observed in the gene body, concomitant with significantly increased expression levels of this gene (logFC: 0.026, P-value: 0.033) in AD patients. We further examined this mediator gene in more detail to analyze upstream and downstream genes within its subnetwork and dissected their involvement in the pathogenesis of AD. A graphical illustration of the *WT1* subnetwork is shown in Figure

11.

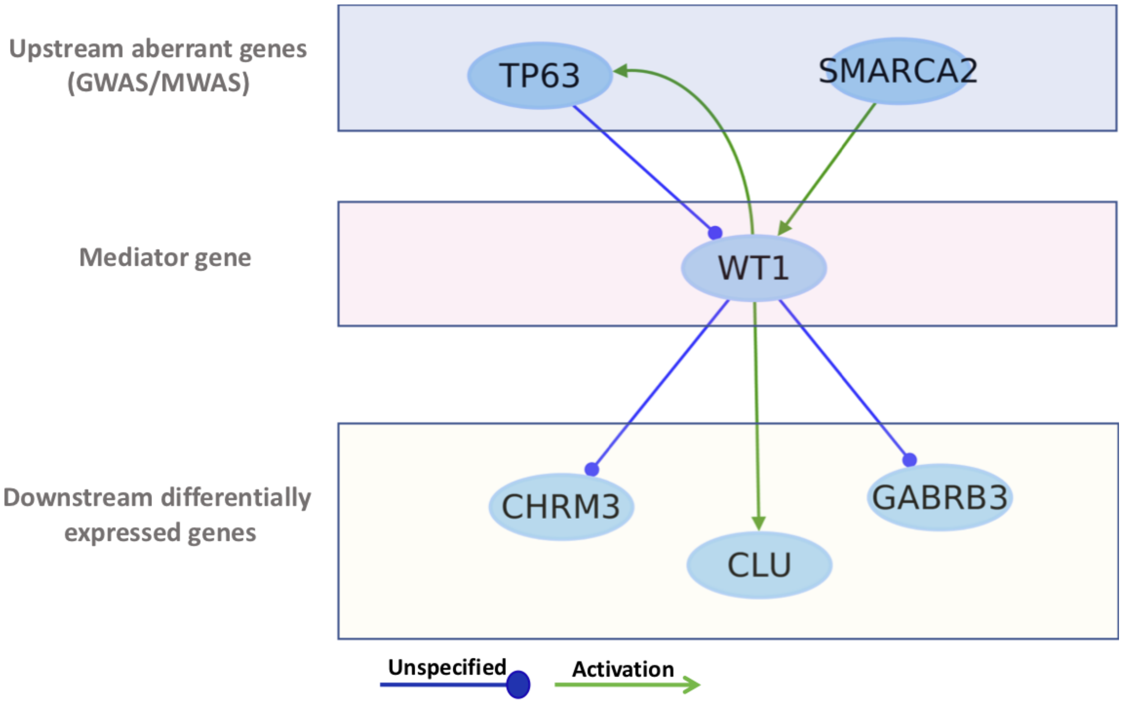

Upstream aberrant genes (GWAS/MWAS)
Mediator gene
Downstream differentially expressed genes
TP63
SMARCA2
WT1
CHRM3
CLU
GABRB3
Unspecified
Activation

Figure 11: **Subnetwork of the AD-associated mediator gene *WT1*.**

The network representation shows that *WT1* has two upstream regulators (*TP63* and *SMARCA2*) known for AD-associated genetic variation and displaying significant differential methylation. Interestingly, *TP63* is also among the five downstream (significantly differentially expressed) genes of *WT1* (*TP63, CLU, GABRB3*, and *CHRM3*). Our differential methylation and expression analysis of these upstream regulators revealed that *SMARCA2* was downregulated (logFC: -0.18, P-value: 4.73^{e-005}) and hypermethylated (logFC: 0.051, P-value: 0.0002) in AD patients, while *TP63* was upregulated (logFC: 0.035, P-value: 0.015) and hypomethylated (logFC: -0.18, P-value: 0.025). Experimental evidence that revealed the physical interaction of *TP63* with *WT1* [249] and positive regulation of *TP63* by *WT1* [160] suggests that they are linked by means of a positive feedback loop. Altogether, the activation of *WT1* by *TP63* overexpression could explain the relative high expression levels of *WT1* in AD patients in comparison to the healthy controls. Of note, both of these *WT1* upstream regulators (*TP63* and *SMARCA2*) have been reported to be crucial for normal neuronal cellular processes and perturbations in these genes are associated with various nervous system disorders including AD [36, 188, 156, 250].

Apart from predicting well-studied AD associated genes, network diffusion was able to identify more novel genes associated with this disease. For example, variants in clusterin (*CLU*), a *WT1* downstream gene, has been associated with AD [146, 100] and high expression levels of this gene have recently been observed in brain regions with plaque pathology [194]. These findings are in line with our analysis, as we observed an increase in expression of this gene (logFC: 0.052, P-value: 0.020) and a representative hypomethylated probe (logFC: -0.001, P-value: 0.034) within 200 bp of the transcription start site (TSS) region. Similarly, the observed lower expression levels (logFC: -0.35, P-value: 0.001) of GABA-Alpha receptor subunit beta-3 (*GABRB3*), concomitant with hypermethylation in the gene body (logFC: 0.028, P-value: 0.017) are in line with an existing study that found lower levels of *GABRB3* mRNA in the AD hippocampus [240, 196], suggesting an altered functional profile of this receptor in AD. Furthermore, we noticed downregulation of the cholinergic receptor *CHRM3* (logFC: -0.48, P-value: 0.0009), accompanied by hypermethylation of this gene (logFC: 0.045, P-value: 0.002) positioned downstream from *WT1*. Although cholinergic neurotransmitter pathway is well-known for its crucial role in the progression of AD [40], it was until recently that the involvement of cholinergic receptors muscarinic (CHRM) has been associated to AD [40]. Altogether, these findings suggest that genomic (genetic variation) and epigenomic (differential methylation) changes in upstream regulators disrupt WT1 activity in such a way that it evokes changes in the expression levels of its downstream genes. Series of synchronized changes at different regulatory levels like these are hypothesized to perturb normal cellular function during the development and course of AD.

4.4.4 Drug targets in mediator gene subnetwork

Connectivity map [145] was used as a reference database for discovering functional connections among predicted top-ranked AD-associated mediator genes and drug actions. The subnetworks of mediator genes were analysed for their enrichment in drug targets based on the similarity of drug-induced gene expression profiles available in the Connectivity map database. The details of this drug enrichment analysis are summarized in Table 4.

As we have used the discretized disease gene expression signatures for querying the Connectivity map, we are interested in drugs that produce the most negative correlated gene expression profile when compared to our query signature. We assume that drugs that are able to produce the exact opposite gene expression profile could be the potential candidates for reverting the diseases (AD-

Gene	Upregulated probes	Downregulated probes	Most enriched drug	Enrichment score	P-value
ETS1	68	73	levcycloserine	-0.894	0.00024
TP63	44	37	isoxicam	-0.828	0.00038
ZNF217	34	14	ginkgolide A	-0.824	0.00185
WT1	23	21	gentamicin	-0.805	0.00280
IL15	3	2	ondansetron	-0.957	0.00001
FHL2	8	6	thioguanosine	-0.883	0.00046
APP	16	18	cefuroxime	-0.934	0.00002
EPAS1	15	13	apramycin	-0.861	0.00068
SMARCA2	11	14	CP-944629	-0.851	0.00001
RXRA	12	5	chlorambucil	-0.881	0.00001

Table 4: **Drug enrichment analysis in mediator gene subnetworks from the connectivity map.**

associated) gene expression program towards a healthy phenotype. As such, these compounds may form targets for drug repurposing. Interestingly, the top-ranked drug for the *EPAS1* query signature named apramycin also turns out to be ranked third for the *RXRA* signature with an enrichment score of -0.834 and a P-value of 0.00137. As *ETS1* is the top-ranked mediator gene in our analysis and it also contains one-third of the dysregulated genes in its subnetwork, a drug that could revert the gene expression profile for genes in the *ETS1* subnetwork might be very effective in obtaining a transition from an AD-associated towards a healthy phenotype. In fact, the drug enrichment analysis suggests a combination of levcycloserine and apramycin to be the most effective therapeutic treatment for AD in terms of normalizing AD-associated gene expression patterns as it holds the potential to revert the maximal gene expression program by a minimal number of candidate drugs. Notably, cycloserine, which is a partial glycine agonist that exhibits its activity by binding the N-methyl-d-aspartate (NMDA) receptor, has been found to significantly improve implicit memory [251] and cognitive function [279] in AD patients.

4.5 Discussion

In view of the interplay between genomic, epigenomic and transcriptomic dysregulation in AD, in the present study, we applied a novel approach for prioritizing AD-associated genes (i.e. genetic variation) based upon AD-linked variation at the epigenomic and transcriptomic level. To this end, by making use of an integrative graph-diffusion based method [66], we have integrated information from different molecular regulatory levels into a directed functional gene-gene interaction network. This method uses information about AD-associated genetic and epigenetic variation in upstream regulatory genes affecting intermediate (mediator) genes, which, through gene-gene interactions, in turn, affect proximal downstream genes evoking expression changes. As such, this approach ranks genes within such gene-gene interaction networks, based on their potential to evoke downstream changes. Some of the most prominent candidate genes include *ETS1, WT1* and *APP* genes, which are all known to be involved in various neuronal cellular processes, while expression changes of these genes have been implicated in the course of AD [121, 169, 208, 181].

A thorough review of the existing literature suggests that all the top-ranked genes have been associated with AD progression. For example, consistent with our differential expression analysis, *ETS1* has been shown to be upregulated in AD brains and has been associated with reactive microglia and Aβ deposition [121]. Similarly, *TP63*, a member of p53 family of transcription factors, has been shown to regulate adult neural precursor and newly born neurons [36], and may have a neuroprotective role by regulating synaptic gene expression [188]. Although the role of *ZNF217* is not yet fully understood in view of neurodegenerative diseases like AD, experimental evidence suggests that the miR-200/*ZNF217* axis may represent a regulatory mechanism mediating the development of AD [291]. Moreover, there is compelling evidence suggesting that *WT1, IL15, APP*, and *EPAS1* play crucial roles in neuronal degeneration and AD [169, 239, 208, 181, 258]. Interestingly, similar evidence is available for the observed changes in the methylation levels [119]. For example, in accordance with our differential methylation analysis, a higher methylation level of the *APP* gene has been reported as an AD-specific phenomenon [119].

In addition to the required multi-omics datasets, the gene-gene interaction information used by network diffusion allows unravelling the dynamics of (dys)regulation in the network. For example, *WT1* has been ranked fourth as an AD-associated mediator gene, controlled by two upstream genes and directly regulating four downstream genes. Interactions within the *WT1* subnetwork suggest its upregulation by a positive feedback loop involving *TP63*, that has been experimentally verified in one direction [249, 160]. This upregulation might be due to genetic variation or hypomethylation in the gene body of *TP63*, or both, thereby promoting *WT1* expression, and, as a result, changes in the expression levels of its downstream genes (*TP63, CLU, GABRB3*, and *CHRM3*). Interestingly, experimental evidence has linked the upregulation of *WT1* and its downstream targets (*CLU* and *TP63*) with neurodegeneration in nervous system disorders, including AD [169, 36, 194]. Similarly, downregulation of its downstream genes (*GABRB3* and *CHRM3*) has been implicated in cognitive decline in AD patients [240, 196, 40, 200]. Taken together, this integrative analysis provides insight on how changes in DNA methylation levels and genetic variation can lead to transcriptional changes via gene-gene interactions, hence potentially explaining the diseased state.

We have shown that the conducted analysis not only identifies disease-related multi-omics signatures and key genes, but also has the ability to predict putative drugs that could revert the disease phenotype. Connectivity map [145] was used as a reference database for linking subnetworks of mediator genes to drugs that have been shown to produce opposite gene expression profiles. A systematic drug enrichment analysis led to the prediction of levcycloserine and apramycin as the most promising existing drugs for reverting the observed AD-associated gene expression profiles. Interestingly, cycloserine treatment has been found to significantly improve implicit memory [251] and cognitive function [279] in AD patients, suggesting the potential of the proposed approach in recapitulating previously-known drugs as well as predicting novel candidates.

Despite being able to prioritize AD-associated causal genes by systematically integrating multi-omics data onto a functional gene-gene interaction network, we acknowledge that the utilized approach has certain limitations, providing avenues for future improvements. For example, network diffusion can investigate the mediator effects of only those genes that are present in the gene interaction network. This highlights the problem of missing data in the literature, as currently, the well-curated and experimentally proven gene-gene interaction maps are not covering the whole spectrum of human genes, rather they are more enriched towards well-studied transcription factors and genes. As such, these results may be biased towards such well-studied, hence highly connected, genes in the network. This bias might arise due to their high connectivity, which contributes to higher chances of finding various differentially methylated or differentially expressed gene in their network neighbourhood. Furthermore, as the reference database used for drug enrichment analysis is comprised of a selected number of drugs profiled on only a few cell lines, most of which are cancerous, this may limit the possibility of finding an optimal drug for reverting the disease-related gene expression pattern. However, decreasing expression profiling costs and an increasing number of such resources [117, 201], will eventually mitigate this problem in the future.

In conclusion, the conducted analysis offers a novel approach for integrating information from different levels of regulation in order to detect and rank AD-associated genetic variation based on its functional significance and gene-gene interaction capability at the transcriptional regulatory level. Such analysis will find its applications in predicting potentially causal genes for other human pathologies where individual datasets are available from different -omics levels. Thus, we are providing the scientific community with a novel approach that can pave the way for deconvoluting complex and multifactorial human diseases, hence fostering the development of novel treatment strategies.

Chapter 5

The role of altered sphingolipid function in Alzheimer's disease; a gene regulatory network-based approach

Muhammad Ali [A,B], Caterina Giovagnoni [B], Lars M.T. Eijssen [B], Roy Lardenoije [B], Janou A.Y. Roubroeks [B], Ehsan Pishva [B], Diego Mastroeni [C], Paul D. Coleman [C], Pilar Martinez-Martinez [B], Jos Kleinjans [B], Antonio del Sol [A], Daniel L.A. van den Hove [B].

[A] Computational Biology Group, Luxembourg Centre for System Biomedicine (LCSB), University of Luxembourg, Luxembourg.

[B] School for Mental Health and Neuroscience (MHeNS), Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, the Netherlands.

[C] Biodesign Institute, Arizona State University, Tempe, AZ, US.

Ali M., et al. (2019) The role of altered sphingolipid function in Alzheimer's disease; a gene regulatory network-based approach. *In preparation*.

5.1 Abstract

Sphingolipids (SLs) are bioactive lipids involved in many physiological pathways. They show an altered metabolism in several central nervous system (CNS) disorders such as Alzheimer's disease (AD). The pathophysiology of AD is still not fully understood. Recent evidence suggests that epigenetic dysregulation plays a crucial role in the disease. In the present study, we examined if genes associated with SL signaling present transcriptional and epigenetic variation in the AD brain. Combining transcriptomic and epigenetic data of SL-related genes from 46 AD and 32 healthy individuals, among 252 SL-related genes assessed, we found 103 genes to be significantly differentially expressed in AD, i.e. indicating a profound enrichment of SL-related gene expression in AD. Additionally, analysis of methylation data revealed *PTGIS, GBA*, and *ITGB2* to be differentially hydroxymethylated and *PLA2G6* to display differential levels of unmodified cytosine in AD. In order to evaluate how SLs influence the disease, we performed a Gene Regulatory Network (GRN) analysis, by reconstructing phenotype-specific, i.e. AD and healthy control, networks. Subsequently, the reconstructed disease network was employed to identify novel perturbation candidates whose alterations hold the potential to revert the gene expression program from an AD towards a healthy state. In particular, we identified *CAV1, TNF* and *IL4* to be the most influential gene combination in the AD network, as a perturbation of these three genes has the potential to revert the expression levels of 41 SL-related genes in the network. This multifactorial epigenetic-transcriptional approach highlights the importance of changes in SL function and related molecules in AD. Moreover, although the genes highlighted are not necessarily responsible for the development and course of the disease, identifying specific dysregulated SL-related genes and their downstream effects will provide a starting point to characterize possible AD biomarkers and guiding the development of new therapeutic approaches.

5.2 Introduction

Alzheimer's disease (AD) is the most common age-related neurodegenerative disorder representing one of the main causes of dementia worldwide [133]. Nowadays, the increasing prevalence of individuals affected by AD and the lack of an effective therapy makes AD to be one of the most challenging diseases in the world [16]. This progressive disease is characterized, amongst others, by initial short-term memory loss and subsequent, language problems, changes in personality, and apathy [16]. The parthenogenesis of AD is still not fully understood, but likely involves both genetic and environmental factors [281]. AD is histologically characterized by the progressive over-production and accumulation of amyloidβ (Aβ) peptide and hyperphosphorylated tau that lead to the formation of extracellular senile plaques and intracellular neurofibrillary tangles, respectively [65]. The concomitant neurotoxicity causes e.g. activation of inflammatory pathways, region-specific synaptic and neuronal degeneration, with huge downstream effects on the physiology of the central nervous system (CNS). Inflammation, metabolic dysfunction, and dysregulation in cell cycle control are just a few of the molecular pathophysiological signs of AD that have been discovered so far and identified as the main etiological causes of neurodegeneration [236].

Increasing evidence suggests that a combination of genetic, epigenetic and environmental factors is contributing to AD progression and associated cognitive impairment. Recently, the role of sphingolipids (SL) garnered more attention in this respect [110]. A clear link between lipids and AD was first reported in 1993 when researchers demonstrated the binding between apolipoprotein E (*APOE4*) and Aβ, concomitant with the increased frequency of the APOE type 4 allele observed in AD patients [56, 266]. More recent evidence has further strengthened the notion of altered SL metabolism in AD [101, 56]. Sphingolipids are complex molecules composed of a backbone of sphingoid bases and a set of aliphatic amino alcohols. A very common SL with an R group consisting of a hydrogen atom only is a ceramide. Ceramide undergoes post-transcriptional modification to form more complex SLs highly abundant in the CNS. Besides their role in building up the cell membrane, they have a variety of bioactive functions regulating different physiological processes such as the cell cycle, differentiation, and regulating synapse structure and function [285].

Multiple factors such as early development, but also environmental stimuli, including nutritional factors and drugs are known to affect SL homeostasis through epigenetic mechanisms regulating the expression levels of SL-associated genes [65]. In the brain, the delicate balance of SL species is absolutely necessary for normal neuronal function, as several brain disorders are known to be caused by dysregulation in e.g. SL metabolism [212]. Interestingly, different metabolic and lipidomic analyses have shown an altered SL metabolism in early AD that contributes to the progression of pathology, by impacting upon Aβ production and tau phosphorylation [189]. In particular, there is evidence that gangliosides, a class of glycosylated SLs, contribute to the initiation and progression of AD by facilitating plaque formation [66]. These studies underscore the importance of SLs in AD onset and progression as well as the need to understand their dysregulation from an integrative point of view [189, 292]. Therefore, a better understanding of the relationship between epigenetic and transcriptomic processes in regulating SL function is of utmost importance for elucidating the underlying role of SLs in AD pathology and the potential development of novel SL-targeted AD therapeutics.

In the present study, we examined SL-related genes from an epigenetic-transcriptional point of view, to further understand the involvement of downstream SL (dys)function in AD. The overarching hypothesis was to identify if, and if so, to which extent SL genes are disrupted at the methylomic and transcriptomic level in AD. To explore this hypothesis, we first identified a set of 252 SL-associated genes based on manually selected Gene Ontology (GO) terms. Transcriptomic analyses showed a profound enrichment of SL-related differentially expressed genes in AD. The conducted epigenetic data analyses revealed *PTGIS, GBA*, and *ITGB2* to be differentially hydroxymethylated and *PLA2G6* to display differential levels of unmodified cytosine in AD. Furthermore, to evaluate how SLs influence the disease, we performed a Gene Regulatory Network (GRN) analysis. The reconstructed networks were employed for *in silico* perturbation analysis and identified *CAV1, TNF* and *IL4* to be the most influential gene combination in the AD network. Taken together, these findings confirmed the initial hypothesis that SL metabolism is significantly altered in AD. Furthermore, the identification of dysregulated SL-related genes and systematic dissection of their downstream effects by *in silico* network perturbation analysis, revealed the potential of this approach to predict diagnostic biomarkers as well as aid in the development of novel SL-targeted AD therapeutics.

5.3 Materials and methods

5.3.1 Identification of sphingolipid pathway associated genes

A gene set involved in sphingolipid function was created by extracting the genes annotated with a list of relevant manually selected Gene Ontology (GO) terms (see supplementary Table 12) [55]. The genes connected to these terms were extracted by using the WikiPathways plugin for

PathVisio, which allowed to save all elements connected to a GO term of interest in an xml type file format (gpml format) [257, 141]. This plugin requires the GO ontology file (‘go.obo’ from geneontology.org; downloaded Nov. 17th, 2018) and a bridgeDb file with gene identifier mappings (‘Hs_Derby_Ensembl_91.bridge’ from www.pathvisio.org in this case) [283]. Thereafter, an R script was used to extract all contributing genes (as identified by their HGNC symbols) from the gpml files for each term. Subsequently, all information per gene was combined, by merging all GO terms from the selection by which the gene is annotated. Furthermore, a basic ‘tree-like’ textual display of the terms in the selection was generated, to support interpretation (Supplementary Table 13, 14, 15). In the end, this procedure resulted in a gene set consisting of 252 SL-related genes that were assessed in downstream applications.

5.3.2 Post-mortem tissue samples

This study makes use of brain tissue from donors of the Brain and Body Donation Program (BBDP) at the Banner Sun Health Research Institute (BHSRI), who signed an informed consent form approved by the institutional review board, including specific consent of using the donated tissue for future research [17, 18].

DNA was obtained from the middle temporal gyrus (MTG) of 46 AD patients and 32 neurologically normal control BBDP donors stored at the Brain and Tissue Bank of the BSHRI (Sun City, Arizona, USA) [17, 18]. The organization of the BBDP allows for fast tissue recovery after death, resulting in an average post-mortem interval of only 2.8 hours for the included samples. A consensus diagnosis of AD or non-demented control was reached by following National Institutes of Health (NIH) AD Center criteria [18]. Comorbidity with any other type of dementia, cerebrovascular disorders, mild cognitive impairment (MCI), and presence of non-microscopic infarcts was applied as exclusion criteria. Detailed information about the BBDP has been reported elsewhere [17, 18].

5.3.3 Differential (hydroxy)methylation analysis

For differential DNA methylation (5-methylcytosine, 5mC), hydroxymethylation (5-hydroxymethyl cytocine, 5hmC) and unmethylation (unmethylatedcytosine, uC) analysis, data was obtained from an unpublished study from our group, where Illumina HM 450K arrays were used for quantifying methylation status of 485,000 different human CpG sites. We only considered 5mC, 5hmC, and uC datasets related to 46 AD patients and 32 controls for which the corresponding gene expression profiles were also available. Preprocessing and analysis of the raw datasets was conducted in R (version 3.4.4) [272]. Raw IDAT files corresponding to the selected individuals were read into R using the wateRmelon “readEpic” function (version 1.20.3) [224]. The “pfilter” function from the wateRmelon package (version 1.18.0) [224] was used to filter datasets based on bead count and detection of p-values. Background correction and normalization of the remaining probe data was performed by using the “preprocessNoob” function of minfi package (version 1.22.1) [9]. Beta values for the probes were obtained by the “getBeta” function of the minfi package. We used the MLML function within the MLML2R package [77] for estimating the proportion of uC, 5mC and 5hmC for each CpG, based on the combined input signals from the bisulfite (BS) and oxidative BS (oxBS) arrays. All of the cross-hybridizing probes and the probes that contained a SNP in the sequence were removed resulting into 407,922 probes to be considered for the differential methylation analysis [43].

Raw IDAT files corresponding to the selected individuals were loaded into R using the minfi “read.metharray” function (version 1.22.1) [9] to generate an RGset for computing the cell type composition of the samples by using the “estimateCellCounts” function of the same package. For estimating the cell composition, we used the FlowSorted.DLPFC.450k package (version 1.18.0) [120] as the reference data for “NeuN_pos” cell composition within the frontal cortex. The limma package (version 3.32.10) [241] was used to perform linear regression in order to test the relationship between the beta values of the probes and the diagnosis of AD. The used regression model considered beta values as outcome, AD diagnosis as predictor, and age, gender, and neuronal cell proportion as covariates. In order to identify significantly differentially methylated probes (DMPs), false discovery rate (FDR) correction for multiple testing was applied where unadjusted *P*-values were corrected for those 252 genes belonging to the sphingolipid pathway. Accordingly, all probes with *P*-value less than 0.0002 were considered as statistically significant in terms of displaying differential levels of methylated, hydroxymethylated or non-methylated cytosine, and considered for further analysis. Resulting probes were annotated using Illumina human UCSC annotation. In order to identify differentially methylated genes, we took the most significant probe as a representative of the methylation status of a gene.

5.3.4 Differential gene expression analysis

For differential gene expression analysis, Illumina HumanHT-12 v4 beadchip expression array data for the same MTG samples was obtained from another unpublished study. Preprocessing and analysis of the raw datasets was conducted in R (version 3.4.4) [272]. Raw expression data was log-transformed and quantile-quantile normalized. For computing the cell type composition, the Neun_pos cell percentage was calculated from the methylation data. The same regression model used for assessing methylation was applied to the expression data where the effects of age, gender and cell type composition were regressed out using limma. The unadjusted *P*-value obtained from limma were FDR-adjusted for only the set of 252 genes in the sphingolipid pathway and only the genes with adj.*P*.value <0.05 were considered as statistically significantly differentially expressed.

5.3.5 Gene-gene interaction network

We used Pathway Studio [206] to obtain directed functional interactions between the genes belonging to sphingolipid associated pathway. The Pathway Studio database contains a collection of literature-curated and experimentally validated direct gene-gene interactions. The high level of literature curation ensures the creation of highly confident interaction network maps. In order to obtain a set of directed functional regulatory interactions among the selected genes, our analysis was restricted to interactions belonging to categories of “Expression”, “Regulation”, “Direct Regulation”, “Promoter Binding”, and “Binding”. The obtained interactions are directed, i.e. the source and target genes are known. Furthermore, the information about the interaction type (activation or inhibition) is also given where available.

5.3.6 *In silico* network simulation analysis for phenotypic reversion

The differential network topology allowed us to identify common and phenotype-specific positive and negative elementary circuits, i.e. a network path which starts and ends at the same node with all the intermediate nodes being traversed only once. These circuits have been shown to play a significant role in maintaining network stability [94] and the existence of these circuits is considered to be a necessary condition for having stable steady states [275]. Considering the importance of these circuits, it has been shown that perturbation of genes in positive circuits induces a phenotypic transition [58]. Furthermore, the differential network topology also aids in identifying differential regulators of the genes common to both phenotype-specific networks. Altogether, the differential regulators and genes in the elementary circuits constitute an optimal set of candidate genes for network perturbation as they are able to revert most of the gene expression program upon perturbation. Identification of network perturbation candidates was carried out by using the Java implementation proposed by Zickenrott and colleagues [314]. The same Java implementation was used to perform a network simulation analysis by perturbing multi-target combinations of up to three network perturbation candidate genes. The used algorithm provided a ranked list of single- and multi-genes combinations (maximally 3 genes) and their scores, which represent the number of other genes within the network whose expression is predicted to be reverted upon the chosen perturbation. Generally, a single- or multi-gene perturbation combination obtaining a high score is indicative of its ability to regulate the expression of a large subset of downstream genes, hence playing a crucial role in the maintenance and stability of the phenotype under consideration.

5.4 Results

5.4.1 Transcriptome analysis of sphingolipid genes

In order to identify SL-associated genes, we used the gene ontology (GO) terms and WikiPathways plugin [257] for PathVisio [141] to convert each GO terms of interest into a tree-like pathway diagram. By removing the genes belonging to irrelevant families and keeping only the ones related to sphingolipid GO terms, we identified 252 genes to be involved in this pathway. Next, information on the expression of these 252 genes within the MTG was extracted from available microarray data performed on brain tissue derived from AD patients and age-match elderly controls. The genome-wide differential expression analysis (DEA) of the transcriptomic data resulted in 7,776 genes to be significantly (FDR corrected P-value <0.05) differentially expressed (up- and down-regulated) when comparing AD patients and age-matched controls. By applying multiple correction for the number of SL-associated genes, we found 103 out of a total of 252 genes to be significantly differentially expressed (up- and downregulated) (see Table 5 for the top 30 differentially expressed genes and supplementary Table 11 for a complete overview).

GeneName	logFC	FDR_adj_Pval	GeneName	logFC	FDR_adj_Pval
STS	-0.221	0.000001	PPM1L	-0.139	0.000538
ARSG	-0.148	0.000011	SMO	0.331	0.000538
EZR	0.631	0.000017	VAPA	-0.279	0.000538
ALOX12B	-0.195	0.000033	ST8SIA2	-0.11	0.000538
ST6GALNAC5	-0.921	0.000033	ELOVL4	-0.633	0.000538
B3GALNT1	-0.387	0.000033	CDH13	-0.654	0.000571
GLTP	0.484	0.000111	RFTN1	-0.382	0.000624
CLN8	0.269	0.000163	EHD2	0.327	0.00075
CD8A	-0.122	0.000179	ST8SIA5	-0.319	0.000784
MAL2	-0.994	0.000196	PRKD1	0.301	0.000784
TFPI	0.241	0.000226	AGK	-0.43	0.000784
CSNK1G2	0.347	0.000272	ATP1A1	-0.553	0.000932
RFTN2	0.529	0.000333	ANXA2	0.369	0.000932
KDSR	0.372	0.000388	GBA	-0.206	0.001177
P2RX7	0.466	0.000528	CLIP3	-0.256	0.001177

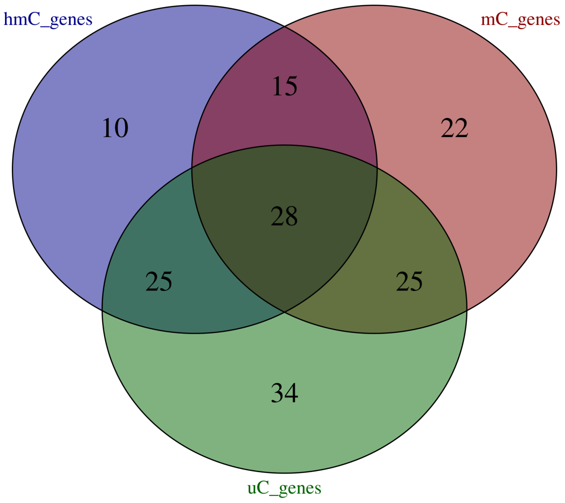
Table 5: **Differentially expressed genes in sphingolipid metabolism pathway.** A list of top significantly differentially expressed genes (FDR adjusted *P*-value <0.05) when comparing AD and control samples.

5.4.2 The SL pathway is significantly dysregulated in AD

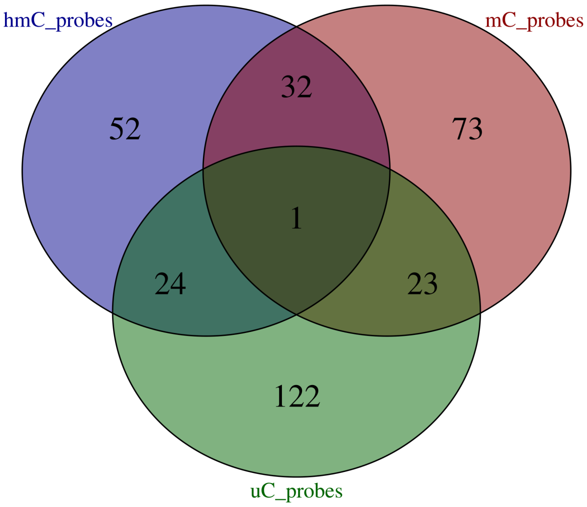
At the very outset, we sought to determine whether SL-function-associated genes were differentially expressed in AD patients in comparison to healthy controls. The genome-wide differential gene expression analysis revealed 24.5% (7,776 out of 31,726 genes) of the genes to be significantly differentially expressed (adj.*p*.val <0.05) between AD and control samples. However, out of the 252 pre-identified SL pathway-associated genes, 103 were found to be significantly differentially expressed, i.e. 40,87%, indicating a profound enrichment of dysregulated genes linked to SL. In line with existing literature [103, 172], this confirms our initial hypothesis that dysregulated SL function represents a key feature affected during the development and course of AD.

5.4.3 Differentially methylated genes are shared across different methylation levels

Overall, 109, 129, and 170 probes displayed nominally significant (unadjusted *P*-value <0.05) levels of 5-mC, 5-hmC, and uC, respectively. These CpG sites were associated to 78, 90, and 112 unique genes, respectively. Interestingly, we see a higher overlap between (h)mC and uC (a Venn diagram representing the overlap of genes across different level is shown in Figure 12a. Similarly, the overlap of particular probes across different epigenetic levels is depicted in Figure 12b. Notably, there were 28 genes that were both significantly differentially methylated, hydroxymethylated as well as displaying different levels of unmodified cytosine (Figure 12a), representing consistent nominal differences observed across all levels of methylation. This highlights the robust and multifaceted interconnection between AD and SLs.



(a)



(b)

Figure 12: **Venn diagram of differentially (hydroxy)methylated SL genes and probes.** a) Overlap between genes across different levels (hmC, mC, uC) that are nominally significant (unadjusted *P*-value <0.05). b) Overlap between nominally significant probes across all three levels.

In order to highlight the most relevant (hydroxy)methylation changes, we subsequently corrected the obtained hits for multiple testing. As such, after correction, four CpG sites still displayed differential levels of methylated or unmodified cytosine (*P*-value <0.0002) when comparing AD and control samples. More specifically, three probes, associated to *PTGIS, GBA*, and *ITGB2*, were differentially hydroxymethylated whereas, one probe, associated to PLA2G6, showed different levels of unmodified cytosine (see Table 6).

Gene Name	Probe Name	mC	hmC	uC	*P*-value	logFC
PLA2G6	cg22326681			x	0.000724	-0.027
PTGIS	cg07612655		x		0.00008	0.037
GBA	cg19257864		x		0.00017	0.036
ITGB2	cg18012089		x		0.000472	0.061

Table 6: **Differentially (hydroxy)methylated genes in SL metabolism pathway.** Significantly differentially methylated (mC), hydroxymethylated (hmC) and unmodified cytosine (uC) probes between AD and controls.

5.4.4 Gene regulatory network analysis

In order to gain a deeper understanding of SL-associated dysregulations at a systems-level, we conducted a differential gene regulatory network (GRN) based analysis to reconstruct two context-specific networks, representing the AD and healthy phenotypes. The employed GRN inference approach [314] relies on Booleanized differential gene expression data and a prior knowledge network (PKN) of gene-gene interactions to reconstruct context-specific networks. The reconstructed AD network comprised 110 genes and 307 interactions (Figure 13a), whereas the healthy phenotype network consisted of 119 genes and 280 interactions (Figure 13b). Although the number of initially identified SL-associated genes was much higher when compared to the number of genes present in the networks, the dependence on experimentally validated manually curated interactions from Pathway Studio [206] suggests that not all of these genes necessarily interact with each other. Interestingly, the enrichment of significantly differentially expressed genes in SL-associated genes was even higher for the genes present in the networks, as 59 out of these 124 genes (47.58%) were significantly differentially expressed (adjusted FDR <0.05). This verifies the reliability of the applied network reconstruction approach such that most of the significantly differentially expressed genes are kept in the networks during the filtration of interactions that were not context-specific. The differential network analysis of AD and healthy phenotypes highlighted important SL-associated genes (e.g. *EZR, ITGB2, NOS1, NOS3, SRC, S1PR3, SPHK1, TGFBR2*) that seem to have a prominent role in the development and course of AD [267, 309].

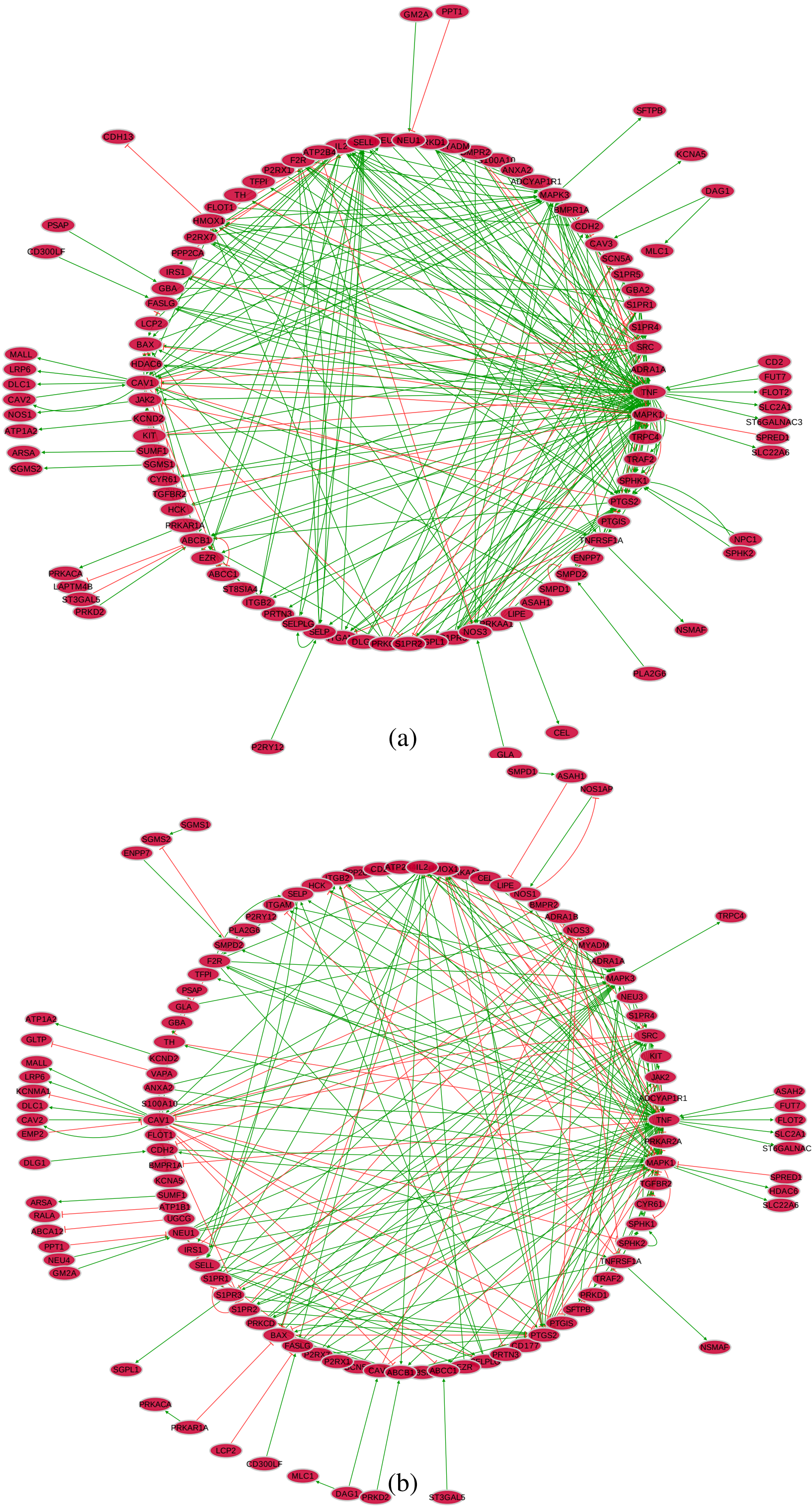


Figure 13: **GRN of SL metabolism diseased and control phenotypes.** a) Gene regulatory network representing the diseased phenotype containing 110 nodes (transcription factors and genes) and 307 interactions; b) network representing the healthy phenotype containing 119 nodes and 280 interactions. Green arrowhead lines in the network represent positive interactions, i.e. activation (253 and 205 in the disease and control phenotype networks, respectively), while the red ones represent negative interactions, i.e. inhibition (54 and 75 in the respective phenotypes).

5.4.5 *In silico* network perturbation analysis

In light of viewing diseases as network perturbations [62, 3], we performed *in silico* network perturbations to identify the most influential combination of genes in the GRN representing the AD phenotype. The network perturbation analysis highlighted the governing role of the perturbation candidates in the GRN and revealed that a three-genes perturbation combination, consisting of *TNF, IL2*, and *MAPK3*, has the potential to revert the expression levels of 41 genes in the network from a diseased towards a healthy phenotype (see Table 7). Similarly, *CAV1, S1PR2*, and *TNF* represented another strong combination, which comprised of two significantly differentially expressed genes (*CAV1* and *S1PR2*) and was found to revert the expression level of 38 other genes in the network, making it an ideal candidate for experimental validation. Importantly, caveolin 1 (*CAV1*), which seems to be one of the most important perturbation candidates, is found to be upregulated in AD patients, consistent with existing studies that associated its elevated expression level to cerebral amyloid angiopathy in AD [282, 87]. As caveolin is a cholesterol-binding membrane protein, its upregulation in AD patients might cause alterations of cholesterol distribution in the plasma membrane, in line with existing studies that validated the notion of dysregulated cholesterol homeostasis in AD [87]. Similarly, *S1PR1* and *S1PR2*, encoding for sphingolipid receptors, also constitute a potent perturbation combination, as these receptors are known to critically regulate many physiological and pathophysiological processes [171]. In addition, they have been reported to modulate the activity of β-site APP cleaving enzyme-1 (*BACE1*) in neurons [271], which is known as a rate limiting enzyme for amyloid-β peptide (Aβ) production, suggesting its therapeutic potential in AD.

Although the genes signatures identified are not necessarily responsible for disease onset and progression, they are able to revert most of the diseased gene expression program upon perturbation, suggesting a prominent role of predicted genes in the establishment of the disease phenotype. Taken together, the *in silico* network perturbation analysis highlights novel candidates that could serve as potential targets for therapeutic intervention in AD.

Rank	Pert score	Gene combo	Signif.	Rank	Pert score	Gene combo	Signif.
1	41	*TNF,IL2,MAPK3*	0	16	36	*CAV1,S1PR1,TNF*	1
2	39	*CAV1,F2R,TNF*	1	17	36	*CAV1,FLOT1,TNF*	1
3	38	*F2R,S1PR2,TNF*	1	18	35	*SPHK1,F2R,TNF*	1
4	38	*F2R,JAK2,TNF*	0	19	35	*SPHK1,CAV1,TNF*	2
5	38	*CAV2,TNF,MAPK3*	0	20	35	*PRKAR1A,F2R,TNF*	1
6	38	*CAV1,TNF,MAPK3*	1	21	35	*PRKAR1A,CAV1,TNF*	2
7	38	*CAV1,S1PR2,TNF*	2	22	35	*JAK2,TNF,TNFRSF1A*	1
8	38	*CAV1,JAK2,TNF*	1	23	35	*FLOT1,JAK2,TNF*	0
9	37	*TH,CAV1,TNF*	1	24	35	*F2R,TNF,TNFRSF1A*	1
10	37	*JAK2,TNF,MAPK3*	0	25	35	*CDH2,CAV1,TNF*	1
11	37	*CAV2,TNF,IL2*	0	26	35	*CAV1,TNF,TNFRSF1A*	2
12	37	*CAV1,TNF,IL2*	1	27	34	*TH,TNF,MAPK3*	0
13	36	*TH,JAK2,TNF*	0	28	34	*SPHK2,F2R,TNF*	1
14	36	*F2R,TNF,MAPK3*	0	29	34	*SPHK2,CAV1,TNF*	2
15	36	*F2R,S1PR1,TNF*	0	30	34	*SPHK1,TNF,MAPK3*	1

Table 7: **SL metabolism network perturbation analysis.** Top 30 key candidate genes combinations identified by *in silico* network perturbation analysis. Rank represents the importance of a given combination of genes. Perturbation score (pert score) represents the total number of genes whose gene expression is reverted upon inducing a perturbation of a given gene combination (gene combo). The number of perturbation candidates that are significantly differentially expressed in the gene combinations are represented in the significance (signif.) column.

5.5 Discussion

In-depth integrative analyses of particular pathways, as performed for the SL pathway in the present study, could aid in obtaining more insight in the yet unclear pathogenesis of AD. Although developments in high-throughput sequencing technologies and computational analysis of obtained datasets have enhanced our knowledge about genes causal to AD, the mechanisms underlying dysregulation of such specific pathways are yet to be explored. A comprehensive characterization of these pathways demands the integrative analysis of various interconnected layers of regulation that have been overlooked and/or understudied so far. Such an explorative study holds the potential to provide more insight into the mechanisms behind dysregulation of specific pathways as seen in AD, thus providing avenues for e.g. designing more effective therapeutic treatment strategies.

Our results reveal an alteration of SL gene function at different levels of DNA methylation. Methylation data showing that the integrin subunit beta 2 (*ITGB2*) gene was significantly hydroxymethylated and upregulated in AD patients are in line with an existing study reporting the high expression level of this gene in mouse models of AD [53]. Furthermore, *PLA2G6*, displaying differential levels of unmodified cytosine, is well known for its implication in neurodegenerative disorders, including AD [55]. Similarly, genetic variation in the gene encoding glucosylceramidase beta (*GBA*) has been suggested to influence the risk of dementia in Parkinson's disease [52]. Interestingly, *de novo* genetic variation in the prostaglandin 2 synthase (*PTGIS*) gene has been suggested to contribute to neurodevelopmental disorders, such as childhood onset schizophrenia [62], whereas there is no supporting literature on the influence of this gene to neurodegeneration to date.

Owing to the alterations in the levels of expression and methylation of SL-associated genes in AD, and the possibility of using them as biomarkers for AD [190], we aimed at bridging the gap in the literature by conducting an integrative analysis of genes involved in SL function. To this end, by systematically identifying significantly differentially expressed genes in post-mortem MTG tissue derived from AD patients and age-matched elderly controls, we reconstructed phenotype-specific, i.e. AD and healthy control, networks. Subsequently, the reconstructed disease network was employed to identify novel perturbation candidates whose alterations hold the potential to revert the gene expression program from an AD towards a healthy state. Further, overlaying the differential methylation data allowed us to explain the observed changes in the expression levels of these genes during the onset of AD. Some of the most prominent predicted candidate genes include *CAV1, S1PR1/2* and *SPHK1*, which are all known to be involved in various neuronal processes, while expression changes of these genes have been implicated in the progression of AD [282, 87, 171, 267, 154].

Notably, *ARSG*, coding for arylsulfatase G, is the most significantly differentially expressed SL-associated gene in AD. *ARSG*, a member of the family of the sulfatases, is involved in hormone biosynthesis, in the modulation of different cellular pathways, including the degradation of macromolecules. The loss of sulfatase activity has been linked to various pathological conditions such as lysosomal storage disorders, cancer and neurodevelopmental dysfunction [84]. Here, for the first time, we are able to link the dysregulation of *ARSG* to AD. In line with previous analysis, *EZR*, which encodes for the membrane protein Ezrin, is profoundly increased in AD [309]. Moreover, genes like *ALOX12B, P2RX7* and *ST6GALNAC5* have already been implicated in AD or other neurodegenerative disorders [180, 177, 284]. Interestingly, other genes, such as *CLN8, ARSG*, and *B3GALNT1*, although associated to neurological dysfunction, had not yet been directly linked to AD [168].

Although our analyses identified novel and pre-identified AD candidate genes, the moderate sample size might have limited the detectable changes in AD samples. This limitation can be seen as an opportunity for conducting more diverse studies including wide-range of analyses in other brain regions to further investigate the role of SL function in AD. Nevertheless, our results provide a clear evidence about the involvement of SLs and related molecules in AD, highlighting the diagnostic and SL-targeted drug-development potential of predicted genes. Even though the reported genes and epigenetic modifications are not predictive signs of the disease progression, our data can serve as a starting point to fill the wide gap of knowledge concerning the role of SLs in AD. Thus, SL function and associated molecules dysregulated in AD could aid in the development of new therapeutic approaches.

Chapter 6

A network-based approach for the identification of Batten disease-specific dysregulation using induced pluripotent stem cell (iPSC)-derived cerebral organoids

Muhammad Ali [A,B], Gemma Gomez Giro [A], Daniel L.A. van den Hove [B], Antonio del Sol [A], Jens C. Schwamborn [A].

[A] Computational Biology Group, Luxembourg Centre for System Biomedicine (LCSB), University of Luxembourg, Luxembourg.

[B] School for Mental Health and Neuroscience (MHeNS), Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, the Netherlands.

6.1 Abstract

Mutations in the *CLN3* gene have been associated with juvenile neuronal ceroid lipofuscinoses (JNCL), the most prevalent form of Batten disease, a lysosomal disease that causes neurodegeneration in children. The early onset of JNCL is characterized by vision loss, followed by progressive deterioration of motor skills and seizures, eventually leading to death at adult age. The limited knowledge of *CLN3* function and scarcity of an adequate disease model has significantly hampered our understanding of disease-specific dysregulation at e.g. the gene expression level. In particular, the reconstruction and analysis of molecular networks that explain transcriptional dysregulation are yet to be explored. In order to understand the functional consequences of a particular mutation in the *CLN3* gene and to identify genes and pathways compromised, we have generated an early human neurodevelopmental model of Batten disease, using isogenic human induced pluripotent stem cell (iPSC)-derived cerebral organoids. The functional characterization of this *in vitro* model showed the presence of disease-specific lipofuscin storage material and lysosomal enzyme dysregulation, highlighting the potential of iPSC-derived *CLN3* mutant organoids to recapitulate disease-specific features. Moreover, differential gene regulatory network (GRN)-based analysis of transcriptomic data obtained from control and disease organoids revealed key regulators maintaining the disease phenotype. Furthermore, pathway enrichment analysis conducted on the disease network showed that genes significantly dysregulated in the network are associated with molecular pathways related to development, validating the potential of this systems-level approach to identify key genes and associated molecular mechanism implicated in Batten disease.

6.2 Introduction

Juvenile neuronal ceroid lipofuscinoses (JNCL), the most prevalent form of Batten disease, is a rare and fatal lysosomal storage disorder (LSD), mostly affecting children and young adults [248]. JNCL typically begins with progressive loss of sight between four and eight years of age, due to retinal degeneration. The clinical course progresses around the age of 10-12 years, with loss of motor coordination and mental decline, often worsened by seizure episodes. These symptoms might be accompanied by behavioural abnormalities such as anxiety and aggression [138]. Moreover, there is also evidence of pathology outside the central nervous system (CNS), more specifically in the cardiovascular [216] and immune system [39]. The disease inexorably leads to death during the second or third decade of life and there are unfortunately no established treatments to date that can stop, reverse, or prevent this disease.

JNCL is caused by recessively inherited mutations in the *CLN3* gene (NG_008654.2), which is located on chromosome 16p12.1 (NC_000016.10). The *CLN3* gene encodes a predicted 438 amino acid protein with a molecular mass of 48kDa. The CLN3 protein (Q13286) is predicted to be a transmembrane protein in the lysosome [57]. Low expression levels and the unavailability of specific antibodies for the CLN3 protein make it difficult to elucidate its precise cellular function. Over the years, *CLN3* has been linked to a vast number of cellular processes, including lysosomal pH regulation, autophagy, endocytosis, trans-Golgi protein transport, cell migration, morphology, proliferation, and apoptosis. In neurons, *CLN3* seems to reside in synaptic vesicles, suggesting a role in synaptic transmission [37]. The CLN3 protein does not share fundamental homology with other proteins and yet it is highly conserved across species. Therefore, animal models have constituted the first and most important source to gain more insight into the exact function of this protein. However, the recent advancements in iPSC-based disease modeling, such as *in vitro* development of human brain organoids, provide us an opportunity to study neurodevelopmental and neurodegenerative diseases in more detail.

Organoids are three-dimensional (3D) structures originating from stem cells and relying on their intrinsic ability to self-organize and form complex structures, when provided with a support matrix and in the presence of suitable exogenous factors. These structures are capable of forming heterogeneous tissue-specific cells, of maintaining gene-gene, cell-cell and cell-matrix interactions, and of recapitulating a large number of physiological functions of the organ they model [49].

A gene regulatory networks (GRNs)-based approach can be employed to gain a deeper understanding of Batten disease-associated dysregulation from a systems point of view. GRNs have been extensively studied for gaining a systems-level understanding of disease-related dysregulation and its underlying mechanisms. These network-based diseased models have been used to predict disease-associated genes and sub-networks [35, 231], while different network topological properties, such as neighbourhood connectivity [234] and Betweenness centrality [126], have been used to predict gene-disease associations.

In the present study, we made use of transcriptomic data from an early human neurodevelopmental model of Batten disease, using isogenic human induced pluripotent stem cell (iPSC)-derived cerebral organoids, in order to study the contribution of the `c.1054C>T` mutation in the *CLN3* gene to brain formation and to the pathophysiology of Batten disease in general. The functional characterization of this *in vitro* disease model [90] showed the presence of disease-specific storage material in iPSC-derived *CLN3* mutant organoids, thus recapitulating disease features by introducing a disease-causing mutation in the *CLN3* gene. In order to gain a deeper understanding of Batten disease-related dysregulation at a systems-level, we utilized a differential GRN inference approach presented by Zickenrott et al. [314], to reconstruct phenotype-specific networks representing the diseased (mutant) and healthy (wild-type) phenotypes. By employing an *in silico* network perturbation analysis on the reconstructed phenotype-specific network, we predicted novel candidate genes that maintain the disease phenotype. Interestingly, pathway enrichment analysis conducted on the network showed that the genes in the network are significantly dysregulated in molecular pathways related to development, a hallmark of Batten disease. Altogether, our data suggest that the mutation in the *CLN3* gene causes the accumulation of pathological storage material and lysosomal enzyme dysregulation at the early stages of brain development, reflected by changes at the transcriptomic level.

6.3 Materials and methods

This chapter is based on a joint work conducted in collaboration with Prof. Schwamborn's lab at the University of Luxembourg. The development of *in vitro* Batten disease model, its functional characterization, and RNA-seq data sampling were done by our collaborators [90] while computational analyses including *in silico* disease modeling, network perturbation, and pathway enrichment were carried out by us.

6.3.1 Insertion of *CLN3*$^{\mathrm{Q352X}}$ mutation in iPSCs

The characterized Gibco (Cat no. A13777) episomal human iPSC line was established as a control line to conduct the genome editing. By using the CRISPR/Cas9 genome editing technology, a `c.1054C>T` genomic mutation was introduced in the *CLN3* gene based on an in-house developed protocol [8]. As a result, we obtained two lines having the same genetic background, one representing the control and another the mutant (*CLN3*$^{\mathrm{Q352X}}$) phenotypes. We created 6 replicates (samples) per line which were all grown in the same conditions and they were all at the same time-point of development.

6.3.2 Generation and culture of human cerebral organoids

Human whole brain organoids were derived from hiPSCs following the Lancaster and Knoblich, 2014 protocol [148]. All of the required growth mediums were purchased from Invitrogen, Gent, Belgium, unless otherwise specified. A total of 9,000 cells per well were seeded into a 96-ultra low adhesion plate (VWR) in embryoid body (EB) formation medium, combining DMEM-F12 (Invitrogen) with 20% KO-Serum Replacement (Invitrogen), 3% FBS (Invitrogen), 1% GlutaMax (Life Technologies), 1% NEAA (Thermo) and 0.0007% 2-Mercaptoethanol (Merck) and supplemented with Y-27632 (Merck Millipore) and bFGF(PreproTech) at a final concentration of 4 ng/mL. Embryoid bodies (EBs) were kept in this media for six days, after which the medium was replaced by Neural induction medium made of DMEM-F12 (Invitrogen) with 1% N2 supplement (Invitrogen), 1% GlutaMax (Life Technologies), 1% NEAA (Thermo) and 1% Heparin (Sigma) (final concentration 1 μg/mL).

EBs were kept in this medium until the eleventh day, after which they were transferred into Matrigel (Corning) droplets and cultured in 24-well-plates under differentiation medium conditions. Cerebral organoid differentiation medium consisted of DMEM-F12 (Invitrogen) and Neurobasal (Invitrogen) media in a 1:1 ratio supplemented with N2 (Invitrogen), 1% GlutaMax (Life Technologies), 0.5% NEAA (Thermo), 1% Penicillin/Streptomycin (Invitrogen) and 0.025% Insulin (Sigma). The first four days of differentiation, medium was supplemented with B27 without vitamin A (Life Technologies) and organoids were kept in static conditions. Later, organoid plates were placed on an orbital shaker (IKA), rotating at 80 rpm, and cultured in differentiation media containing B27 with vitamin A (Life Technologies). Media was exchanged every second or third day and cerebral organoids were kept in culture for another 55 days after the start of differentiation (day 11).

6.3.3 Isolation of RNA samples

Total RNA was isolated from cerebral organoids using the RNeasy Mini Kit (Qiagen) following the manufacturer's instructions. An on-column DNAse digestion step was included in the protocol and performed with RNase-Free DNase Set (Qiagen). Five samples per condition were taken as replicates where every sample consisted of a pool of three organoids. RNA concentration was spectrophotometrically determined using NanoDrop (ND 2000). Library preparation for sequencing was done with 1 μg of total RNA using the TruSeq mRNA Stranded Library Prep Kit (Illumina) according to manufacturer's protocol. Briefly, the mRNA pull down was done using magnetic beads with an oligodT primer. To preserve strand information, the second strand synthesis was done with incorporation of dUTP so that during PCR amplification only the first strand was amplified. The libraries were quantified using the Qubit dsDNA HS assay kit (Thermofisher) and size distribution was determined using the Agilent 2100 Bioanalyzer. Pooled libraries were sequenced on NextSeq500 using the manufacturer's instructions.

6.3.4 RNA-Seq data processing and analysis

Illumina NextSeq single-end reads were filtered by using BBDuk (`trimq=10 qtrim=r ktrim=r k=23 mink=11 hdist=1 tpetbominlen=40`; http://jgi.doe.gov/data-and-tools/bb-tools/) to remove illumina adapters, PhiX library adapters, and to quality trim the reads. FastQC [6] was used to check the quality of the reads in order to assure that only high-quality reads were kept for subsequent analysis. Resulting reads were mapped to human GRCh37 genome by using tophat (version 2.1.1) [278] (`library-type=fr-secondstrand`) and Bowtie2 (version 2.3.2.0). Obtained alignment files were sorted by using samtools (version 1.6-5) [158] and the statistics of the alignment rate were obtained by using samtoolsflagstat. Cufflinks (version 2.2.1) [278] was used to quantify the transcripts and resulting expression values per gene were obtained in FPKM (fragments per kb per million reads). Differential expression analysis between the wild-type and mutant samples was conducted by using the cuffdiff program from the cufflinks tool. Only significantly differentially expressed genes with an absolute log2 fold change greater than 1 were considered for subsequent analysis.

6.3.5 Gene Regulatory Network (GRN) reconstruction

For the set of significantly differentially expressed genes when comparing mutant and wild-type samples, experimentally validated direct gene-gene interactions were retrieved from MetaCore (Clarivate Analytics). The interaction types belonging to categories “Transcription regulation” and “Binding” were kept in the prior knowledge network (PKN) from MetaCore. The differential network inference method proposed by Zickenrott et al. [314] was used to prune the network edges (interactions) which were not compatible with the discretized gene expression program of the respective phenotype. Briefly, this method uses discretized differential gene expression data and infers two networks representing the mutant (disease) and wild-type (healthy) phenotypes as steady states. Some of the interactions derived from MetaCore have an unspecified regulatory effect, as the exact mechanism of regulation is not known in those cases. The proposed algorithm infers the regulatory effect (activation or inhibition) for such unspecified interactions based on the given gene expression pattern.

6.3.6 Identification of network perturbation candidates

The differential network topology allowed us to identify common and phenotype-specific positive and negative elementary circuits, i.e. a network path which starts and ends at the same node with all the intermediate nodes being traversed only once. These circuits have been shown to play a significant role in maintaining network stability [94] and the existence of these circuits is considered to be a necessary condition for having a stable steady (network) state [275]. Considering the importance of these circuits, it has been shown that perturbation of genes in the positive circuits induces a phenotypic transition [58]. Furthermore, the differential network topology also aids in identifying the differential regulators of the genes, which are common to both phenotype-specific networks. Altogether, the differential regulators and genes in the elementary circuits constitute an optimal set of candidate genes for network perturbation as they are able to revert most of the gene expression program upon perturbation. Identification of network perturbation candidates was carried out by using the Java implementation proposed by Zickenrott et al. [314].

6.3.7 *In silico* network simulation analysis for phenotype reversion

The Java implementation from Zickenrott et al. [314] was used to perform the network simulation analysis by perturbing multi-target combinations of up to four candidate genes identified in the previous step. The used algorithm gives a ranked list of single- and multi-gene(s) combinations (4 genes maximally) and their scores, which represent the number of genes whose expression is being reverted upon inducing the chosen perturbation. If a single- or multi-gene(s) perturbation combination obtains a high score, it is indicative of its ability to regulate the expression of a large number of downstream genes, hence playing a crucial role in the maintenance and stability of the phenotype under consideration.

6.3.8 Gene and pathway enrichment analysis

MetaCore (Clarivate Analytics) and EnrichNet [91] were used to conduct gene ontology (GO) and pathway enrichment analysis. The set of upregulated genes in the diseased network were used to identify the most over-represented biological processes and molecular functions associated with the genes in the network.

6.4 Results

6.4.1 Whole transcriptome analysis reveals impaired development in *CLN3*$^{\mathrm{Q352X}}$ cerebral organoids

Although there are a number of studies describing mechanism dysregulated in Batten disease, they do not provide insight into underlying transcriptional dysregulation in the pathologic brain compared to the healthy state. Therefore, we performed bulk RNA-seq analysis in our organoid model system for Batten disease to determine whether developmental differences were reflected in the gene expression profiles of the organoids. The differential expression analysis (DEA) of the transcriptomic data resulted in 972 genes to be significantly (Benjamini Hochberg corrected *P*-value $<$0.05 and logFC $>$1) differentially expressed (up- and downregulated) between the Control and the *CLN3*$^{\mathrm{Q352X}}$ mutant phenotypes (see Figure 14).

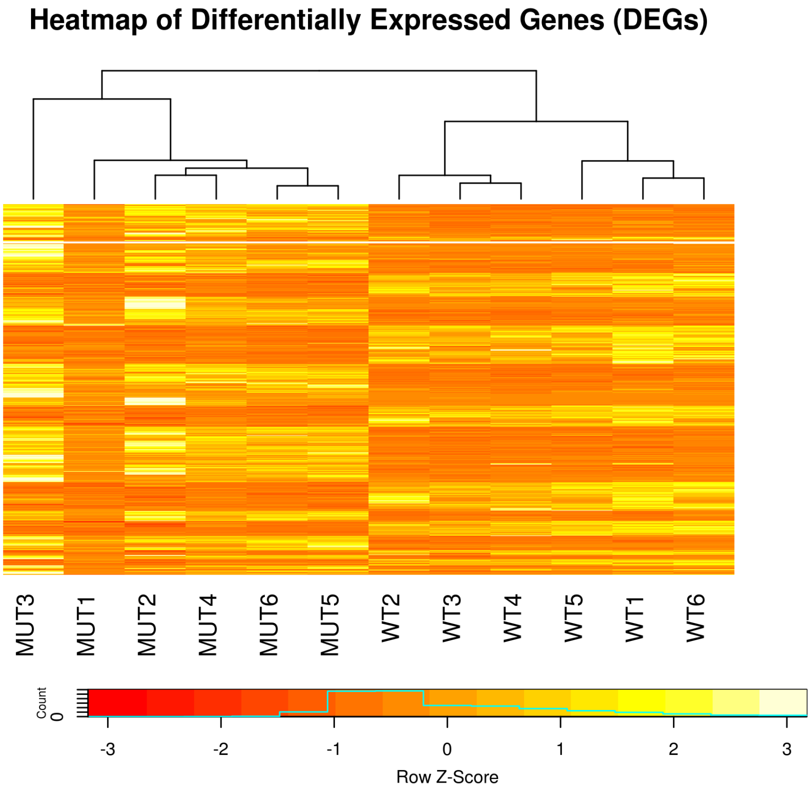
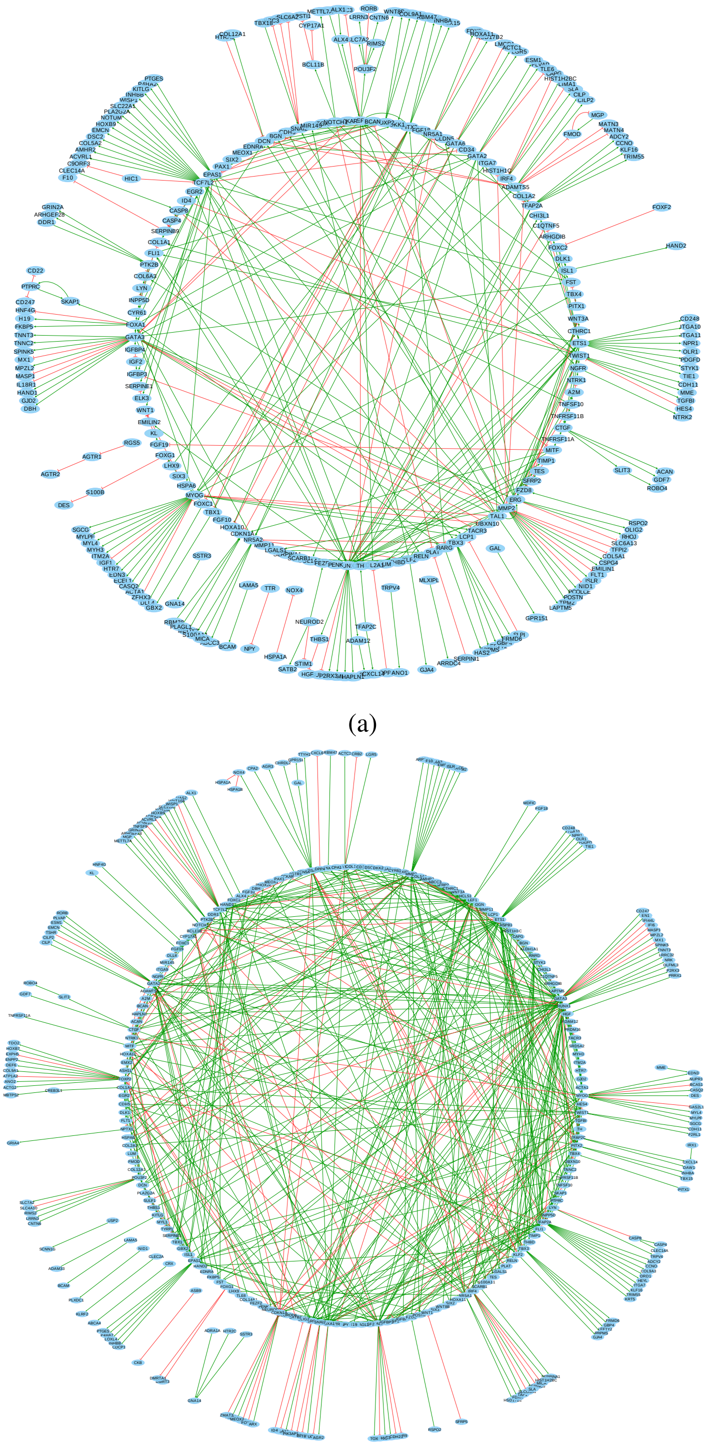


Figure 14: **Heatmap showing the clustering of differentially expressed genes between control (healthy) and *CLN3*$^{\mathrm{Q352X}}$ (mutant) brain organoids**

Gene regulatory network (GRN) analysis

In order to gain a deeper understanding of Batten disease-related dysregulation at a systems-level, we employed a differential GRN-based approach to reconstruct phenotype-specific networks representing the *CLN3*$^{\mathrm{Q352X}}$-diseased (mutant) and control (healthy) phenotypes. The employed GRN inference approach by Zickenrott et al. [314] relies on discretized differential gene expression data and a prior knowledge network (PKN) of interactions to reconstruct phenotype-specific networks. The reconstructed *CLN3*$^{\mathrm{Q352X}}$-diseased network comprised 353 genes and 641 interactions, whereas the control healthy network contained 298 genes and 399 interactions (see Figure 15a,15b).

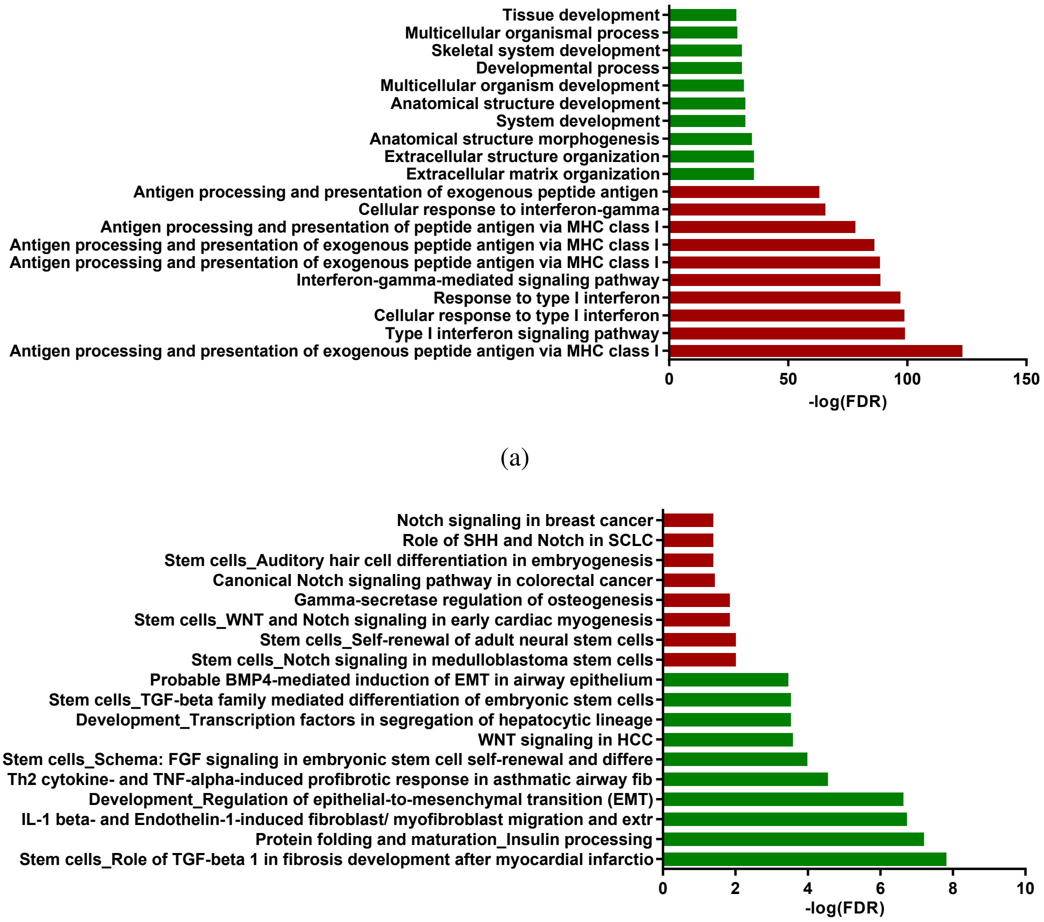


(a)

(b)

Figure 15: **GRN representing Batten diseased and control phenotypes.** a) Gene regulatory network representing the healthy phenotype contained 298 nodes (transcription factors and genes) and 399 interactions; b) GRN representing the diseased phenotype and contains 353 nodes and 641 interactions. Green arrowhead lines in the network represent positive interactions, i.e. activation (292 for the Control and 520 in the *CLN3* mutant), while the red ones represent negative interactions, i.e. inhibition (107 and 121 respectively in the two phenotypes).

Interestingly, GO analysis of the *CLN3*$^{\mathrm{Q352X}}$ diseased network revealed that most of the upregulated genes in the network are significantly enriched in cellular processes related to development. Some prominent examples are *PAX5, TBX15*, and *HAND1* genes, that are well known for their key roles in B cell [185], skeletal [255], and cardiovascular development [183]. Similarly, the downregulated genes, such as human leukocyte antigen (*HLA*) genes, were targeting biological processes and pathways related to the immune response and antigen processing and presentation (see Figure 16a). Moreover, pathway enrichment analysis conducted on the network showed that the genes in the network are significantly dysregulated in molecular pathways related to stem cells and development (see Figure 16b). Some prominent examples include *NOTCH1* [222], *WNT3A* [166], and *HES4* [72] genes, and they have been shown to play important roles in various developmental processes.



(a)

(b)

Figure 16: **Gene and pathway enrichment.** a) Gene enrichment analysis of *CLN3*$^{\mathrm{Q352X}}$ network (top). Genes which were upregulated in the disease phenotype indicated a significant enrichment of cellular processes highlighted in green, while processes associated with downregulated genes are depicted in red. b) Pathway enrichment analysis of the *CLN3*$^{\mathrm{Q352X}}$ network (bottom), upregulated pathways are highlighted in green, while pathways associated to downregulated genes are marked in red.

Additionally, evaluation of gene expression data outside the disease network indicated expression changes related to cortical neuron morphogenesis and central nervous system development and highlighted decreased expression of associated genes, such as *FOXG1* [99], *FEFZ2* [71], *CTIP2* [207], *SATB2* [28], *TBR1* [69] or *NEUROD2* [213] in *CLN3*$^{\mathrm{Q352X}}$ mutant cerebral organoids (see Figure 17). This suggests that alterations in development and cortical neuronal specification may occur during early development in our isogenic *CLN3*$^{\mathrm{Q352X}}$ organoids, compared to the control.

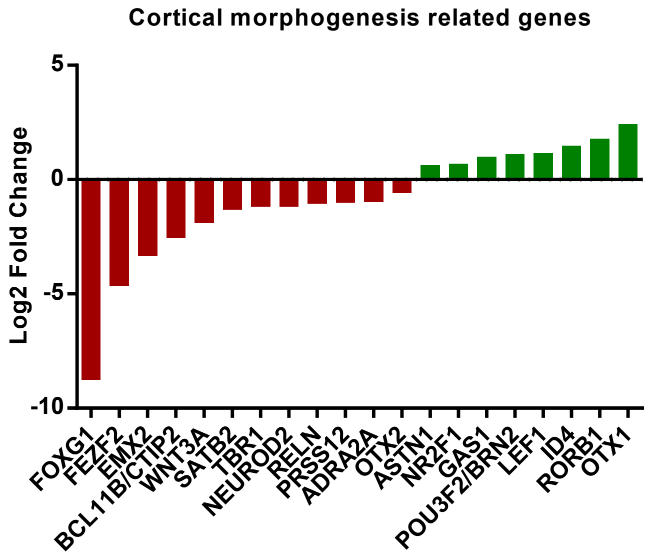


Figure 17: **Log 2 fold change expression values for genes related to brain development and cortical morphogenesis showed a downregulation in mutant samples in comparison to control for most of the genes.**

6.4.2 Lysosome enzyme expression is altered in *CLN3*$^{\text{Q352X}}$ cerebral organoids and lipofuscin storage material is present

Following up on our RNA-seq analysis, we sought out connections between our expression data and pathways that are especially relevant in JNCL, coupling the findings to molecular and biochemical analyses. Thus, we screened our dataset for genes that were differentially expressed between our control and *CLN3*$^{\text{Q352X}}$ mutant organoids and were related to lysosomal and vesicle-mediated transport pathways. The list of genes belonging to these pathways was extracted from Pathcards [19], an integrated database of human biological pathways and their annotations.

Among the differentially expressed genes related to lysosomal pathways, we found several lysosomal enzymes, such as *TPP1/CLN2*, a soluble serine protease in the lysosome, or cathepsins, like cysteine proteases *CTSC, CTSK* and *CTSZ* (see Figure 18). Increased amounts of TPP1 protein have been described in various pathological conditions such as neurodegenerative lysosomal storage disorders, inflammation, cancer and aging [92]. Moreover, *TPP1* has been reported to interact with the *CLN3* gene [287]. Consistent with existing reports, we could also find an increased amount of *TPP1* in our *CLN3*$^{\mathrm{Q352X}}$ cerebral organoids, compared to the control brain organoids.

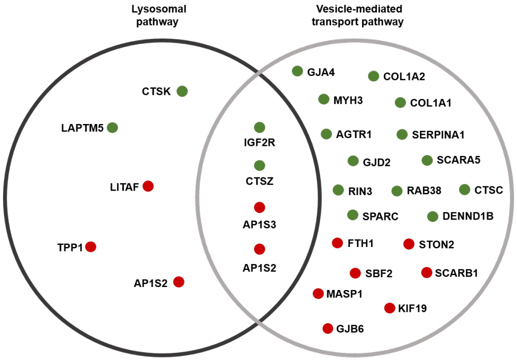


Figure 18: **Lysosome enzyme expression is altered in *CLN3*$^{\mathrm{Q352X}}$ cerebral organoids.** Venn diagram showing the differentially expressed genes related to lysosomal and vesicle-transport pathways that are differentially up- (green) or downregulated (red) in the *CLN3*$^{\mathrm{Q352X}}$ mutant brain organoids.

Importantly, the functional characterization of cerebral organoids by autofluorescence (confocal laser excitation) and ultrastructural analysis (transmission electron microscopy) showed that developed *in vitro* batten disease model recapitulates important disease hallmarks, such as increased autophagic vacuoles and presence of intracytoplasmic and electron dense storage material in the organoid cultures [90] (unpublished work. Data not shown here). The carried out functional characterization also reinforces the idea that the pathogenesis of JNCL is associated with alterations in lysosomal compartments that might start with dysregulations at the transcriptional level.

6.4.3 *In silico* network perturbation analysis

In view of diseases as network perturbations [62, 3], we performed *in silico* network perturbations to identify the most influential genes in the diseased network. The network perturbation analysis highlighted the governing role of the perturbation candidates in the GRN. In this regard, simulation of single transcription factor perturbations revealed that *FOXA1, TAL1, GATA3, ETS1*, and *RUNX1* play an important role in maintaining the diseased phenotype network, i.e. leading to a significant reversion of the pathological gene expression program upon perturbation. Existing literature suggests a crucial role of these transcription factors (TFs) in various human developmental processes. For example, *FOXA1* has been widely known to be involved in the development of T-cell [151], midbrain dopaminergic neurons [78], and mammary gland and prostate [22]. Similarly, other TFs such as *TAL1, GATA3, ETS1*, and *RUNX1* have been reported to play important roles in the development of hematopoietic [230, 308, 115], immune [15], and cardiovascular systems [86]. Furthermore, dysfunction of these genes has been associated with hematological malignancies [230, 263], congenital heart defects [304] and cancers [124, 47]. Considering their topological characteristics and key roles in the developmental processes, they constitute ideal candidates for perturbation. A perturbation combination of *TAL1, GATA3, ETS1*, and *RUNX1* reverted the gene expression state of 118 other genes in the diseased network. Although the predicted genes are not necessarily responsible for disease onset and progression, they are able to revert most of the diseased gene expression program upon perturbation (Table 8). These finding suggests that the predicted genes might play a crucial role in the establishment of the disease phenotype.

	Single-gene perturbation				Multi-gene perturbations
Rank	Score	Gene	Rank	Score	Genes
1	82	*FOXA1*	1	118	*TAL1, GATA3, ETS1, RUNX1*
2	69	*TAL1*	2	116	*FOXA1, MYOG, MMP2, GATA3*
3	67	*LEF1*	3	114	*FOXA1, MMP2, GATA3, ETS1*
4	67	*GATA3*	4	114	*FOXA1, GATA3, ETS1, RUNX1*
5	66	*MMP2*	5	113	*FOXA1, MYOG, GATA3, ETS1*
6	65	*RUNX1*	6	111	*FOXA1, MYOG, GATA3, RUNX1*
7	65	*ETS1*	7	111	*FOXA1, MMP2, GATA3, RUNX1*
8	64	*GATA6*	8	110	*MYOG, FOXA1, MMP2, RUNX1*
9	63	*MYOG*	9	110	*MYOG, FOXA1, GATA3, GATA2*
10	63	*IRF4*	10	110	*FOXA1, MMP2, GATA3, GATA2*

Table 8: **Top 10 key candidate genes from single- and multi-gene network perturbation simulation analysis.** Genes are ranked based on their score. The score represents the number of genes whose discretized expression is reverted (shifted from the pathologic towards the healthy phenotype) upon *in silico* perturbation. The scores obtained for different candidate genes are a qualitative measure of their ability to revert the disease phenotype.

6.5 Discussion

The development of an *in vitro* Batten disease model and associated transcriptomic analyses in the context of genetic variation in *CLN3* are, to our knowledge, non-existent in humans to date. In order to bridge this knowledge gap, we report a systems-level study utilizing the transcriptomic data from an *in vitro* Batten disease model harboring the *CLN3*$^{\mathrm{Q352X}}$ mutation and the phenotypic hallmarks of this disease. We identified significant changes in the gene expression levels of important genes that are associated with development and differentiation. This highlights the potential of created isogenic cell line to recapitulate disease features as a consequence of introducing a disease-causing mutation in the *CLN3* gene. To our knowledge, the conducted study constitutes a first attempt to generate a computational disease model of Batten disease, employed for investigating the contributions of `c.1054C>T` mutation in the *CLN3* gene to brain formation and to the pathophysiology of Batten disease in general.

Additionally, we were able to describe lysosomal alterations already happening at the transcriptional level, concomitant with the differential expression of genes that govern lysosomal and vesicle-mediated pathways. To this end, we report a downregulation in *TPP1* peptidase in our *CLN3*$^{\mathrm{Q352X}}$ mutant organoids. Notably, we observed an increase at the protein level. *TPP1* has been shown to be involved in the initial degradation of subunit c when adding both purified TPP1 and soluble lysosomal fractions, containing various proteinases, to mitochondrial fractions, which normally results in rapid degradation of subunit c, but not in the presence of a *TPP1* inhibitor or when the enzyme is non-functional, as in CLN2 disease [75]. We hypothesize that the sustained increase in *TPP1* levels might be a cellular response to degrade extra subunit c of mitochondrial ATP synthase (SCMAS) starting to accumulate in the lysosomes due to *CLN3* deficiency, while the expression levels may change rapidly in response to the cellular demands. Another altered lysosomal enzyme in our *CLN3*$^{\mathrm{Q352X}}$ mutant organoids was *CTSD*, aspartic protease especially abundant in neuronal lysosomes. *CTSD* was also found inducing lysosomal storage material in mouse CNS neurons that presented a deficiency in this enzyme [72]. We reported a decrease in protein levels of *CTSD*, which might be compensated by the cells at the transcriptional level by upregulating several other cathepsin genes, such as *CTSC* or *CTSZ*.

Interestingly, GO analysis of the *CLN3*$^{\mathrm{Q352X}}$ diseased network revealed that most of the upregulated genes in the network are significantly enriched in cellular processes related to development. Some prominent examples are *PAX5, TBX15*, and *HAND1* genes, that are well known for their key roles in B cell [185], skeletal [255], and cardiovascular development [183]. Similarly, the downregulated genes, such as human leukocyte antigen (*HLA*) genes, were targeting biological processes and pathways related to the immune response and antigen processing and presentation

The DEA of the transcriptomic data obtained from control and *CLN3*$^{\mathrm{Q352X}}$ mutant organoids provided additional insight into Batten disease-associated dysregulation at the gene-expression level. The GRN analysis provided a systems-level view of this dysregulation and revealed the underlying key genes maintaining the disease phenotype. The cellular processes and pathway enrichment analysis of upregulated genes in the disease phenotype network showed a strong association of these genes with developmental processes and pathways (see supplementary Figure 20). Some prominent examples are *PAX5, TBX15*, and *HAND1* genes, that are well known for their key roles in B cell [185], skeletal [255], and cardiovascular development [183]. The enrichment analysis suggested skeletal system development to be one of the several processes that is significantly disrupted in this disease. The normal outcome of this process is the development of the skeleton over time, from its formation until becoming a mature structure [21], however, this process is significantly affected in Batten disease. As evident from existing studies, deposition of lysosomal residual bodies, the end products of prelysosomal and intralysosomal degradation of cellular constituents, is ubiquitous and affect skeletal muscles in Batten disease [233, 216]. Surprisingly, the TGF-beta, Wnt and BMP signaling pathways that were found to be significantly associated with the diseased network are widely known for their fundamental roles in embryonic skeletal development and postnatal bone homeostasis [299, 167]. Similarly, other developmental processes such as tissue development, multicellular organism development and extracellular matrix (ECM) organization were significantly enriched concomitant with signaling pathways regulating stem cell differentiation and epithelial-to-mesenchymal transition. Interestingly, there is compelling evidence suggesting major changes in the expression of numerous ECM molecules in nervous system-related disorders, such as multiple sclerosis [261], Alzheimer disease, and Parkinson disease [25, 254]. Furthermore, various disease models of nervous system disorders and LSDs share common features, such as neuro-inflammation and neuro-degeneration [12, 20]. Taken together, these results suggest the dysregulation of developmental pathways and processes in the Batten disease model, consistent with existing literature explaining the phenotypic characteristics of this disorder throughout its course.

Although the development of *in vitro* Batten disease model and GRN-based modeling of transcriptomic data provided insights into key developmental pathways affected by the disease, we acknowledge that the presented approach has some important limitations. Foremost, *in vitro* cultured cerebral organoids present variable shapes and features, unlike that of a mature human brain.

Moreover, they lack surrounding tissues that are important for the interplay of neural and non-neural tissue cross talk, such as meninges, bones and vasculature [149]. Due to these factors, organoids showed marked variability, particularly between preparations. To account for these variations, controls as well as mutant organoids were prepared at the same time, grown in the same medium, and organoids from at least three different independent derivations were taken per experiment. It is also important to be aware that the transcriptomic data analysed in this study was profiled by bulk RNA-seq, which has its own limitation due to the heterogeneity caused by diverse cell types in the brain [217]. To this end, the reliance on literature-derived interaction networks and further contextualization of these networks with discretized gene expression data maximizes the perseverance of context-specific interactions in the reconstructed GRN models, while removing the noise in the data by filtering incompatible interactions. However, a more sophisticated study, assessing different neural cell types in isolation or performing single-cell RNA-seq profiling would greatly improve the power of this analysis to detect significant disease-associated changes [246]. Furthermore, as the *in silico* network perturbation analysis revealed *FOXA1, TAL1, GATA3, ETS1*, and *RUNX1* to be the key regulators maintaining the Batten diseased phenotype, a complementary random perturbation analysis could aid in assessing the significance of these predictions. In addition, an experimental validation of these predictions by TF knock-down or over-expression would greatly benefit in understanding the specific contributions of these TFs to the disease outcome.

Altogether, our data suggests that the introduction of the `c.1054C>T` mutation in the *CLN3* gene causes the accumulation of pathological storage material and lysosomal enzyme dysregulation at the early stages of brain development, which can be modelled with cerebral organoids. Furthermore, gene expression profiling on control and *CLN3*$^{\mathrm{Q352X}}$ mutant organoids allowed us to characterize transcriptional changes that arise as a consequence of this mutation. We believe the development of this Batten disease *in vitro* model system and generation of corresponding transcriptomic profiles will provide the scientific community with a valuable resource to further dissect its underlying mechanism, helping in its early diagnosis as well as in designing potential therapeutic treatments.

Chapter 7

General Discussion

The large-scale development of high-throughput sequencing technologies has allowed the generation of reliable omics data at different regulatory levels. Integrative computational models enable disentangling the complex interplay between these interconnected levels of regulation by assessing these large quantities of biomedical information in a systematic way. However, modeling human diseases by computational approaches demands the reconstruction of reliable network models that are context-specific and encapsulate the regulatory gene expression program. For example, it has become increasingly clear that it is the cross-talk between epigenetic and transcriptomic layers that regulates gene expression programs across various human cell types [53, 274, 41]. Although existing integrative methods for reconstructing network models provide meaningful insights for understanding the underlying mechanisms of gene regulation, they suffer from some important limitations. First and foremost, these methodologies usually rely on histone modification marks for active enhancer identification (H3K27ac) to predict active enhancer regions and associate them to their target genes based on ad hoc criteria, such as the nearest gene or all genes within a defined range. Such enhancer annotations might lead to the inference of false-positive (and -negative) interactions as it has been shown that enhancers do not necessarily act on the closest promoter, but can bypass neighboring genes to act on more distant genes along the same as well as a different chromosome [100, 109]. Secondly, these approaches rely on position weight matrix (PWM)-based predictions of transcription factor (TF) binding in regulatory regions to associate regulator TFs with their respective target genes. Such PWM-based predictions might lead to the inference of many false-positive interactions due to the detection of false-positive motifs, as indicated by existing studies [313, 163]. Lastly, these methods lack systematic benchmarking of predictive network models against experimental cell-type-specific TF chromatin immunoprecipitation-sequencing (ChIP-seq) data.

These limitations suggest the need for more sophisticated integrative computational methods that rely only on experimental data from different regulatory levels to reconstruct reliable cell-type-specific networks. As such, these network models can help us in addressing the fundamental biological questions related to cell-type-specific and disease-associated transcriptional regulation. Moreover, the application of such tailor-made integrative network models is yet to be explored in the context of epigenetic and transcriptomic dysregulation that play a crucial role in normal cellular differentiation processes and lies at the core of many disorders [161, 295]. Therefore, reconstructing cell-type-specific network models by integrating epigenetic and transcriptomic information can provide deeper insights into underlying mechanisms, e.g. allowing us to predict specific external stimuli (e.g. TF over-expression) that can overcome epigenetic barriers restricting the differentiation potential of cells.

In order to address the aforementioned limitations, we developed INTREGNET, a computational framework that reconstruct cell-type-specific core transcriptional regulatory networks (TRNs) for various human cell types and cell lines. Chapter 3 provides a concise overview of this approach. The reconstructed networks allowed us to understand cell-specific regulation of TFs at the epigenetic and transcriptomic level, thus enabling us to predict efficient combinations of instructive factors (IFs) for desired cellular conversions between any two cell types of interests. This method is based on the systematic integration of epigenetic and transcriptomic information to reconstruct core TRNs, offering several advantages over current approaches. Firstly, it exclusively relies on experimental data for TRN reconstruction, which increases precision compared to PWM-based methods that are not cell-type-specific. In particular, INTREGNET introduces cell-type-specificity by integrating information on TF ChIP-seq experiments, chromatin accessibility and active cis-regulatory elements to accurately reconstruct networks. Secondly, integration of protein-protein interaction (PPI) data allows for dissecting region-specific cooperative and competitive TF-binding, i.e. the joint effect of multiple TFs on the transcription of target genes. Considering these protein-protein interactions are critical for prioritizing more efficient combinations of IFs, exemplified by the complex formation of *SOX2* and *POU5F1* that is necessary for inducing pluripotent stem cells [242, 27]. Finally, the devised strategy for predicting efficient IFs actively incorporates differences in the epigenetic landscape between the initial and target cell type. Despite the specific combination of IFs, the amount of epigenetic restructuring required during reprogramming is a key determinant of cellular conversion efficiency [219]. INTREGNET accounts for these epigenetic landscape differences by penalizing the calculated efficiency of IFs with the amount of required restructuring.

In principal, INTREGNET can be customized for applications for human disease modeling, in view of diseases as network perturbations from healthy to disease phenotype [62]. A core TRN reconstructed from different epigenetic and transcriptional profiles obtained from pathological cells might help in identifying causal TFs that establish or maintain the disease phenotype. Finally, *in silico* network perturbations can guide experimental efforts in pre-selecting a set of putative target TFs, whose perturbation induces the conversion into a healthy phenotype, with vast amounts of potential applications to personalized medicine. To our knowledge, INTREGNET is one of the first approaches that aims at identifying highly efficient IFs based on the systematic integration of information linked to multiple regulatory levels, and is expected to find diverse applications in the field of regenerative medicine. In particular, considering the success of *in vivo* reprogramming in preclinical models, we believe INTREGNET to be a valuable tool for alleviating the impediment of low efficiency by guiding cellular conversion experiments.

The remarkable development of high-throughput sequencing technologies has allowed the generation of great quantities of genomic, epigenomic and transcriptomic data for various human diseases that has allowed us to dissect the mechanisms behind the onset and progression of multifactorial diseases. As such, many studies have used information from an individual regulatory level to identify causal genes and understand the mechanisms underlying the pathophysiology of Alzheimer's disease (AD). For example, genome-wide association studies (GWAS) have successfully identified numerous susceptibility genes for AD [89, 125, 130, 33]. Similarly, based on the crucial role of DNA methylation in cellular processes [214], including gene regulation [229], cellular differentiation [131] and genomic imprinting [221], there have been many studies linking changes in DNA methylation status to the pathogenesis of AD [259, 290, 61]. Furthermore, analysis of genome-wide transcriptomic data sets from post-mortem brain tissue has unveiled various key genes in different biological pathways associated with AD [286]. These findings highlight that changes associated with AD are not restricted to a particular regulatory layer and can be observed across genetic, epigenetic and transcriptomic levels [147, 61, 179, 100, 109, 170]. Although various levels of genomic regulation, including DNA methylation, chromatin modifications and microRNAs (miRNAs), are known to be highly interconnected at the functional level [63], commonly used analytical approaches are usually restricted to analyzing only one or two layers of molecular information in association with AD [61, 286, 107], and, moreover, are mostly restrained to correlations. Therefore, an integrative multi-omics systems biology approach to uncover the relative, interdependent contribution of various molecular layers in the development and course of AD is of utmost importance.

In view of the interplay between genomic, epigenomic and transcriptomic dysregulation in AD, in the study described in Chapter 4, we applied a novel approach for prioritizing AD-associated genes (i.e. genetic variation) based upon AD-linked variation at the epigenomic and transcriptomic level. To this end, by making use of an integrative graph-diffusion based method [66], we have integrated information from different molecular regulatory levels into a directed functional gene-gene interaction network. The proposed method uses information about AD-associated genetic and epigenetic variation in upstream regulatory genes affecting intermediate (mediator) genes, which, through gene-gene interactions, in turn, affect proximal downstream genes evoking expression changes. As such, this approach ranks genes within such gene-gene interaction networks, based on their potential to evoke downstream changes. Some of the most prominent candidate genes include *ETS1, WT1* and *APP* genes, which are all known to be involved in various neuronal cellular processes, while expression changes of these genes have been implicated in the course of AD [121, 169, 208, 181].

We have also shown that the approach presented in Chapter 4 not only identifies disease-related multi-omics signatures and key genes, but also has the ability to predict putative drugs that could revert the disease phenotype. Connectivity map [145] was used as a reference database for linking subnetworks of mediator genes to drugs that have been shown to produce opposite gene expression profiles. A systematic drug enrichment analysis led to the prediction of levcycloserine and apramycin as the most promising existing drugs for reverting the observed AD-associated gene expression profiles. Interestingly, cycloserine treatment has been found to significantly improve implicit memory [251] and cognitive function [279] in AD patients, suggesting the potential of the proposed approach in recapitulating previously-known drugs as well as predicting novel candidates.

In conclusion, the conducted analysis offers a novel approach for integrating information from different levels of regulation in order to detect and rank AD-associated genetic variation based on their functional significance. Such analysis will find its applications in predicting potentially causal genes for other human pathologies where individual datasets are available from different -omics levels. Thus, we are providing the scientific community with a novel approach that can pave the way for deconvoluting complex and multifactorial human diseases, hence fostering the developmental of novel treatment strategies.

Although developments in high-throughput sequencing technologies and computational analysis of obtained datasets have enhanced our knowledge about AD causal genes, the mechanisms underlying dysregulation of implicated pathways are yet to be explored. A comprehensive characterization of these pathways demands the integrative analysis of various interconnected layers of regulation that have been overlooked and/or understudied so far. In-depth integrative analyses of such pathways, as performed for the sphingolipid (SL) pathway in the Chapter 5, could aid in obtaining more insight in the yet unclear pathogenesis of AD, thus providing avenues for designing more effective therapeutic treatment strategies.

Owing to significant alterations in the expression and methylation levels of SL-associated genes in AD, and the possibility of using them as a biomarkers [190], we aimed at conducting an integrative analysis focused only on the genes involved in SL function. Some of the most prominent candidate genes predicted to underlie SL dysregulation include *CAV1, S1PR1/2* and *SPHK1*, which are all known to be implicated in the development and course of AD [282, 87, 171, 267, 154].

Of note, a similar analysis (unpublished observations; data not shown) was conducted on genes associated to tryptophan (TRP), more specifically the TRP-kynurenine (KYN) pathway, implicated in AD [307, 228, 112]. Unlike in the case of SL, the integrative epigenetic and transcriptomic analysis found no significant disease-associated changes at the network level. Considering the fact that the TRP-KYN pathway mainly reflects a metabolic cascade, it is not surprising that a GRN and associated integrative analyses are not that successful, as genes involved in metabolic cascades are not expected to highly interact with each other at the epigenetic and transcriptomic level. To conclude, this analysis could serve as a negative control and indirectly validate our findings of significant changes in SL metabolism in AD, where strong interconnectivity of involved genes was observed at the epigenetic and transcriptional regulatory levels.

Development of *in vitro* disease models and analyses of profiled transcriptomic datasets to attain systems-level understanding of disease-associated dysregulation provide avenues for pre-clinical validation of potential cell therapy applications. Such analyses are very scarce for rare neurological disorders such as Batten disease. In particular, the development of an *in vitro* Batten disease model and comparative transcriptomic analyses in the context of genetic variation in *CLN3* are very limited and, to our knowledge, non-existent in humans to date. In order to bridge this gap in the literature, we report a systems-level study utilizing the transcriptomic data from an *in vitro* Batten disease model harboring the *CLN3*$^{\mathrm{Q352X}}$ mutation and the phenotypic hallmarks of this disease.

The differential expression analysis (DEA) of the transcriptomic data obtained from control and *CLN3*$^{\mathrm{Q352X}}$ mutant organoids provided additional insight into Batten disease-associated dysregulation at the gene-expression level. The GRN analysis provided a systems-level view of this dysregulation and revealed the underlying key genes maintaining the disease phenotype. The cellular processes and pathway enrichment analysis of upregulated genes in the disease phenotype network showed a strong association of these genes with developmental processes and pathways. Some prominent examples are *PAX5, TBX15*, and *HAND1* genes, that are well known for their key roles in B cell [185], skeletal [255], and cardiovascular development [183]. The enrichment analysis suggested skeletal system development to be one of the several processes that is significantly disrupted in this disease. The normal outcome of this process is the development of the skeleton over time, from its formation until becoming a mature structure [21], however, this process is significantly affected in Batten disease. As evident from existing studies, deposition of lysosomal residual bodies, the end products of prelysosomal and intralysosomal degradation of cellular constituents, is ubiquitous and affect skeletal muscles in Batten disease [233, 216]. Surprisingly, the TGF-beta, Wnt and BMP signaling pathways that were found to be significantly associated with the diseased network are widely known for their fundamental roles in embryonic skeletal development and postnatal bone homeostasis [299, 167]. Similarly, other developmental processes such as tissue development, multicellular organism development and extracellular matrix (ECM) organization were significantly enriched concomitant with signaling pathways regulating stem cell differentiation and epithelial-to-mesenchymal transition. Interestingly, there is compelling evidence suggesting major changes in the expression of numerous ECM molecules in nervous system-related disorders, such as multiple sclerosis [261], Alzheimer disease, and Parkinson disease [25, 254]. Furthermore, various disease models of nervous system disorders and LSDs share common features, such as neuro-inflammation and neuro-degeneration [12, 20]. Taken together, these results suggest the dysregulation of developmental pathways and processes in the Batten disease model, consistent with existing literature explaining the phenotypic characteristics of this disorder throughout its course. Altogether, our data suggests that the introduction of the `c.1054C>T` mutation in the *CLN3* gene causes the accumulation of pathological storage material and lysosomal enzyme dysregulation at the early stages of brain development, which can be modelled with cerebral organoids. Furthermore, gene expression profiling on control and *CLN3*$^{\mathrm{Q352X}}$ mutant organoids allowed us to characterize transcriptional changes that arise as a consequence of this mutation. We believe the development of this Batten disease *in vitro* model system and generation of corresponding transcriptomic profiles will provide the scientific community with a valuable resource to further dissect its underlying mechanism, helping in its early diagnosis as well as in designing potential therapeutic treatments.

7.1 Current challenges and future perspectives

The most important limitations encountered in computational approaches for disease modeling are discussed in Chapter 2, and the studies described in this thesis are also subject to some of those limitations. One of the most important limitation that all computational network-based approaches suffer from is the validation of reconstructed network models. The currently used gold-standard for network validation concerns cell-type-specific TF ChIP-seq data, however, due to a large number of known human TFs and various cell types, ChIP-seq profiling for every TF is far from being complete. Even though INTREGNET, a method described in Chapter 3, leverages a comprehensive compendium of over 11,000 publicly accessible TF ChIP-seq profiles from the Cistrome database [186], we still run into the problem of missing data. For example, *LIN28A* was identified as a core TF of induced pluripotent stem cells (iPSCs), but its binding sites have not been profiled by ChIP-seq in any human cell type or cell line. As INTREGNET relies on TF ChIP-seq data to reconstruct core TRN models, it cannot be contained in the core TRN and predicted as an IF for inducing PSCs. However, the amount of available TF binding site profiles is steadily increasing, which eventually will mitigate this problem in the future. Moreover, the availability of additional epigenetic profiles, such as multiple histone modifications and chromatin conformation, will become greater in the future, opening the possibility of integrating them into the TRN.

Similarly, another important limitation of INTREGNET is reliance on bulk datasets. Indeed, transcriptomic and epigenetic heterogeneity in cellular populations can influence successful conversion due to the existence of different sub-populations exhibiting distinct conversion efficiencies [31]. In this regard, modeling core TRNs using single-cell data could allow the identification of sub-populations with the highest conversion propensity. Furthermore, single-cell data can help in devising novel experimental strategies for cellular conversion, such as initially priming cell populations and subsequently inducing the desired cell type conversion.

Another prominent limitation of network-based modeling approaches is their reliance on literature-derived gene-gene interaction networks. Although these prior knowledge networks (PKN) help us in understanding the transcriptional regulation of genes, they are far from being complete. As described in Chapter 4, despite being able to prioritize AD-associated genes by systematically integrating multi-omics data onto a functional gene-gene interaction network, we acknowledge that the presented approach has certain limitations, providing avenues for future improvements. For example, the employed network diffusion approach can investigate the mediator effects of only those genes that are present in the gene interaction network. This highlights the problem of missing data in the literature, as currently, the well-curated and experimentally proven gene-gene interaction maps are not covering the whole spectrum of human genes, rather they are more enriched towards well-studied TFs and genes. As such, these results may be biased towards such well-studied, hence highly connected, genes in the network. This bias might arise due to their high connectivity, which contributes to higher chances of finding various differentially methylated or differentially expressed gene in their network neighbourhood. However, decreasing expression profiling costs and an increasing number of gene knock-down and over-expression experiments in data bases like gene perturbation atlas (GPA) [302] and gene expression omnibus (GEO) [50], will eventually help towards completing the functional interaction maps.

The inference of causality is another important limitation inherent to integrative studies considering the epigenetic and transcriptional data for understanding disease-related dysregulation. It is impossible to say whether epigenetic and transcriptional differences detected between AD and control individuals represent a cause or consequence of pathology. However, unraveling the causal or consequential relationship of these changes is now possible by the help of *in vitro* (or even *in vivo*) studies where epigenetic editing or transcriptional knock-down and over-expression experiments can help us in understanding their contributions to the disease.

One of the most critical limitations confronted in epigenetic studies are small to moderate sample sizes, limiting their potential to detect significant changes. This is exemplified by some of the very high *p*-values reported for differentially methylated genes in Chapter 4. Also the results of Chapter 5 should be interpreted with caution, as only three probes survived correction for multiple testing in the differential methylation analysis. Although analyses described in Chapter 5 identified novel and pre-identified SL-related genes based on their epigenetic and transcriptional changes, the moderate sample size might have limited detectable changes in AD samples. This limitation can be seen as an opportunity for conducting more diverse studies including wide-range of analyses in other brain regions to further investigate the role of SL function in AD. Nevertheless, our results provide a clear evidence about the involvement of SLs and related molecules in AD, highlighting the diagnostic and SL-targeted drug-development potential of predicted genes. Even though the reported genes and epigenetic modifications are not predictive signs of disease progression, our data can serve as a starting point to further investigate the role of SLs in AD. Thus, exploring SL function and associated molecules dysregulated in AD could aid in the development of new therapeutic approaches.

Although the development of *in vitro* Batten disease model and GRN-based modeling of transcriptomic data provided insights into key developmental pathways affected by the disease, we acknowledge that the approach presented in Chapter 6 has some important limitations. Foremost, *in vitro* cultured cerebral organoids present variable shapes and features, unlike that of a mature human brain. Moreover, they lack surrounding tissues that are important for the interplay of neural and non-neural tissue cross talk, such as meninges, bones and vasculature [149]. Due to these factors, organoids showed marked variability, particularly between preparations. To account for these variations, controls as well as mutant organoids were prepared at the same time, grown in the same medium, and organoids from at least three different independent derivations were taken per experiment.

It is also important to be aware that the transcriptomic data analysed in this study was profiled by bulk RNA-seq, which has its own limitation due to the heterogeneity caused by diverse cell types in the brain [217]. To this end, the reliance on literature-derived interaction networks and further contextualization of these networks with discretized gene expression data maximizes the perseverance of context-specific interactions in the reconstructed GRN models, while removing the noise in the data by filtering incompatible interactions. However, a more sophisticated study, assessing different neural cell types in isolation or performing single-cell RNA-seq profiling would greatly improve the power of this analysis to detect significant disease-associated changes [246].

Furthermore, as the *in silico* network perturbation analysis conducted in Chapter 6 revealed that *FOXA1, TAL1, GATA3, ETS1*, and *RUNX1* play a crucial role in maintaining the Batten diseased phenotype, an experimental validation of these predictions by TF knock-down or over-expression would greatly benefit in understanding the specific contributions of these TFs to the disease outcome.

Even though the research conducted in this thesis covers a wide range of cell types, tissues, diseases, and techniques, the current status of our understanding of epigenetics and transcriptomic cross-talk in regulating normal cellular processes and their dysregulation in various disorders is far from being complete. Thus, while solutions can be offered to address the specific limitations described above, a more radical shift in integrative computational approaches is required to truly mature this emerging field. Although advances in sequencing technologies have made it easier to generate and share high-quality multi-omics datasets, the field still seems to lag behind in how to deal with these datasets and interpret the findings of the integrative analyses. By harnessing the computational power of the present era and advances in integrative modeling and machine learning, it may be possible to generate predictive models that may aid in the diagnosis and prognosis of multifactorial human disorders [108], thereby fostering the development of novel therapeutic strategies that could make personalized medicine a reality.

Chapter 8

Valorization

The multifactorial nature of neurodevelopmental disorders, like Batten disease, or age-related disorders, such as Alzheimer's disease (AD), requires the generation and integrative analysis of biological data from different regulatory levels (genomics, epigenomics, and transcriptomics) to advance our understanding of the underlying mechanisms. A deep and thorough understanding of these multi-layered mechanisms at systems-level is the key to deconvolute the complexity of human pathologies, hence fostering the development of novel and effective treatment strategies.

Despite recent advances in next-generation sequencing technologies and the development of novel computational modeling approaches that shed more light on disease processes, we are still far from completely characterizing the disease-causative agents and finding a definite cure for most human pathologies, including AD. This highlights the need for explorative studies that utilize multi-level regulatory information to decipher the underlying mechanisms controlling normal gene expression regulation and their dysregulation in human disorders. In order to meet this challenge, the research presented in this thesis aims to further accelerate research in the field of computational disease modeling. Though it is unlikely that the work carried out in this thesis will have a direct impact on society in the short run, the approaches introduced here will definitely guide future studies, bringing the existing knowledge one step closer to its applications in disease intervention.

All in all, the research presented in this thesis highlights the potential of computational disease modeling and integrative multi-omics analysis for dissecting human disorders and proposing rational therapeutic strategies. For example, the approach presented in chapter 3 may find its application in facilitating experimental attempts for treating human developmental disorders, that arise due to a disruption in the normal cellular differentiation process [161, 295]. To this end, the proposed method (INTREGNET) is able to predict specific sets of instructive factors (IFs) that can induce desired cellular conversion events with increased efficiency, hence overcoming a longstanding problem in regenerative medicine hampering the translation of therapeutic interventions into clinical applications.

The research work described in chapters 4 and 5, focused on AD, offers novel insights into epigenetic and transcriptomic dysregulation by comparing multi-omics datasets from patients and healthy controls. Different markers identified at the genome-wide level, as well as by zooming in on sphingolipid metabolism, can be further tested for their potential as diagnostic markers or as putative drug targets. Furthermore, expanding existing knowledge about the involvement of different regulatory layers in AD-associated dysregulation is already a merit on itself, as such a deeper understanding of underlying mechanisms is vital for the development of novel therapeutic intervention strategies.

The final scientific efforts described in chapter 6 are directed towards generating a computational model of Batten disease in order to understand the functional consequences of a particular mutation in the *CLN3* gene, and to identify genes and pathways compromised in this human neurodevelopmental disorder. The conducted gene regulatory network (GRN) and *in-silico* gene perturbation analyses revealed key driver genes in maintaining the diseased phenotype network, i.e. leading to a significant reversion of the pathological gene expression program upon perturbation. The reported findings not only highlight the potential of employed systems-level approaches to identify relevant genes and associated molecular mechanisms implicated in Batten disease, but also provide a prediction of putative candidate genes that might be the drivers of disease-related dysregulation. We believe this study has a direct impact on the society as it provides the scientific community with a very a first *in vitro* and *in silico* *CLN3*$^{\mathrm{Q352X}}$ mutation Batten disease model, as well as the fact that it identifies key genes to be experimentally validated for their potential as an early diagnostic marker or target for designing potential therapeutic treatment strategies.

Taken together, the research work conducted in this thesis may have a substantial impact on our society, providing the scientific community with novel approaches to develop computational disease models and dissect their underlying mechanisms. These computational models can help us unlock the biological systems [29], as well as devise new intervention strategies to halt the progression of human disorders or cure them. Finally, after going through four years of extensive training and hands-on practical experience, I am confident in saying that my efforts have allowed me to explore the computational disease modeling field in depth, also identify associated gaps and weaknesses in existing knowledge and approaches. During the last year of my project, I have stretched my skills beyond the vigorous foundation provided by my supervisors to meet the requirements to advance in this field. The expertise I have gained throughout my Ph.D. trajectory has enabled me to design my own studies and write grant proposals, which means I am now ready to make a real impact on society as an independent researcher.

Chapter 9

Summary

The remarkable development of high-throughput sequencing technologies has allowed the generation of great quantities of genomic, epigenomic and transcriptomic data for various human diseases that has allowed us to dissect the mechanisms behind the onset and progression of multifactorial diseases. Owing to the multifactorial nature of most human disorders, recent advancements in computational disease modeling, by integrating regulatory information from different levels, provide a new framework for understanding the complex nature of human health and disease. For example, modelling of complex gene interaction networks has been very useful for disease modelling [143, 13, 182] and for disentangling the interplay between different regulatory layers [193, 93, 195]. However, integrative network modelling approaches –i.e. linking different regulatory layers– [193, 93, 195, 104] are still scarce, which hampers the possibility of studying the crosstalk established among regulatory layers for determining a given phenotype or mediating phenotypic transitions [73]. As such, developing tailor-made computational models is a crucial step in understanding the contributions of genomic, epigenomic, and transcriptomic landscapes in cellular circuitry, lineage specification, and the onset and progression of human disease.

In order to bridge the gaps in the literature, we report integrative systems-level approaches to dissect the underlying disease mechanisms, helping in their early diagnosis as well as in designing potential therapeutic treatments. The research conducted in this thesis can be divided into five parts. CHAPTER 2 constitutes a concise overview of existing computational methods in the field of systems biology. Particular attention is paid to state-of-the-art gene regulatory network (GRN) based methods for instructive factors (IFs) determination and human disease modeling. Along with the strengths, the limitations of these methods are highlighted, thereby providing avenues for the research conducted and described in the following chapters.

Due to a clear lack of integrative methods for predicting more efficient sets of instructive factors, CHAPTER 3 describes INTREGNET, an integrative computational method for systematically identifying reliable minimal sets of IFs that can induce desired cellular conversions with increased efficiency. The application of this method is demonstrated in an *in vitro* setting, where limited conversion efficiency is a crucial barrier for its application in regenerative medicine.

As explained above, the heterogeneous and multifactorial nature of human disorders, such as Alzheimer's disease (AD), requires the integration of regulatory information from different -omics levels in order to capture the underlying mechanisms behind the onset and progression of this disease. In CHAPTER 4, global multi-omics alterations in AD patients are identified by comparing genomic (gene aberration), epigenomic (DNA methylation) and transcriptomic data sets of 46 diseased patients with 32 age-matched controls.

CHAPTER 5 features an integrative exploration of specific neurobiological pathways known to be impaired in AD. A comprehensive analysis of gene expression and DNA methylation levels is performed for genes known to be associated with sphingolipid function. The identified key genes and their particular methylation signatures offer mechanistic insights into AD pathology and may act as potential biomarkers.

In vitro modeling of human diseases allows us to gain crucial insights into mechanisms underlying disorders, hence devising and optimizing new strategies for therapeutic intervention. CHAPTER 6 features the differential network-based analysis of transcriptomic data sets obtained from brain organoids that served as an *in vitro* model of Batten disease. This study focuses on identifying key genes and pathways that are disrupted during the course of this disease.

In conclusion, we believe that the work conducted in this thesis provides the scientific community with a valuable resource to understand the underlying mechanism of multifactorial diseases from an integrative point of view, helping in their early diagnosis as well as in designing potential therapeutic treatments.

Chapter 10

Curriculum Vitae

Muhammad Ali was born on the 16th of June, 1989 in Gujranwala, Pakistan. He grew up and went to school in his home town. After finishing his bachelor's in Bioinformatics from Government College University, Faisalabad, Pakistan in 2011, he joined the University of Saarland, Germany for a Masters in Bioinformatics. During his master's degree, the scientific areas which grabbed his interest were modern methods in drug discovery and gene regulatory network (GRN) analysis. For his master thesis, he worked in the lab of Prof. Volkhard Helms and characterized the biochemical and biophysical properties of protein-protein interaction interface residues. As GRN modeling and analyses were among his favorite areas of interest, after completing his master's degree, he applied for a doctoral position in the computational biology group of Prof. Antonio del Sol at the University of Luxembourg. Luckily, he got this most-awaited opportunity to join Prof. Antonio del Sol's research group and excel in this interesting field of science. Fortunately, during early months of his Ph.D., his supervisor (Prof. Antonio del Sol) got an EU Joint Programme – Neurodegenerative Disease Research (JPND) grant on Alzheimer's disease (AD) epigenetic analyses for biomarker identification, originally proposed by Dr. Daniel van den Hove from Maastricht University. Based on Muhammad's interests, his supervisor allowed him to work on this grant together with scientific personnel from Maastricht University. That was the time when the challenging era of Muhammad's doctoral degree started. He did multiple projects during the course of his Ph.D. degree together with very kind and supportive promoters from Maastricht University (Dr. van den Hove and Dr. Pishva), as well as expert advisers from the University of Luxembourg (Dr. Angarica and Dr. Jung). Overall, the aim of his doctoral degree was to develop network-based approaches for modeling human disease. In addition to conducting his research jointly at the University of Luxembourg and Maastricht University, he presented his research at numerous national and international platforms provided by the Epi-AD consortium under the framework of JPND grant. These opportunities broadened Muhammad's exposure to the scientific world and gave him the confidence to be an independent researcher. During his Ph.D., Muhammad has also been a tutor and practical supervisor in several Bachelor courses at the University of Luxembourg. Next to teaching others, Muhammad also kept on expanding his own skills by taking courses and workshops in statistics, team and project management, and scientific writing to evolve his research and aptitude. After successfully defending his doctoral degree thesis, Muhammad has planned to join the biomedical data science group of Dr. Enrico Glaab at the University of Luxembourg, to further expand his work on human disease modeling and integrative GRN analyses through the implementation of machine learning and systems biology approaches.

Appendix A

List of Abbreviations

3D	Three-dimensional
AD	Alzheimer’s disease
APOE4	apolipoprotein E
Aβ	Amyloid β
BACE1	β-site APP cleaving enzyme-1
BBDP	Brain and Body Donation Program
BHSRI	Banner Sun Health Research Institute
CAV1	Caveolin 1
ChIP-Seq	Chromatin Immunoprecipitation sequencing
CHRM	cholinergic receptors muscarinic
CLU	clusterin
CNS	central nervous system
CTSD	cathepsin D
DEA	differential expression analysis
DEGs	differentially expressed genes
DMPs	differentially methylated probes
DNA	deoxyribonucleic acid
DNase-Seq	deoxyribonuclease sequencing
DTMC	discrete time markov chain
ECM	extracellular matrix

ESC	embryonic stem cells
FDR	false discovery rate
FP	false positives
FPKM	fragments per kb per million reads
GABRB3	GABA-Alpha receptor subunit beta-3
GBA	glucosylceramidase Beta gene
GEO	Gene Expression Omnibus
GO	gene ontology
GRNs	gene regulatory networks
GS	Gold-standard
GWAS	genome-wide association studies
hmC	hydroxymethylated cytosine
HSCs	hematopoietic stem cells
IFs	instructive factors
IGAP	International Genomics of Alzheimer's Project
INTREGNET	INtegrative Transcriptional REGulatory NETworks
iPSCs	induced pluripotent stem cells
ITGB2	Integrin subunit beta 2
JNCL	juvenile neuronal ceroid lipofuscinosis
JSD	Jensen-Shannon divergence
LSDs	lysosomal storage disorders
mC	methylated cytosine
MCI	mild cognitive impairment
miRNA	microRNA
MSCs	mesenchymal stem cells
MTG	middle temporal gyrus
NCLs	neuronal ceroid lipofuscinoses
NIH	National Institutes of Health
NMDA	N-methyl-d-aspartate
NSCs	neural stem cells
PKN	prior knowledge network

PPI	protein-protein interaction
PSCs	pluripotent stem cells
PTGIS	prostaglandin 2 synthase
PWM	position weight matrix
RNA	ribonucleic acid
RNA-seq	ribonucleic acid sequencing
SCMAS	subunit c of mitochondrial ATP synthase
Sls	Sphingolipids
SNP	single nucleotide polymorphisms
SSCs	strongly connected components
TCGA	The Cancer Genome Atlas
TFs	transcription factors
TP	true positives
TRNs	transcriptional regulatory networks
TSS	transcription start site
uC	unmethylated cytosine
VPA	valproic acid
WT1	Wilms tumor suppressor

Appendix B

Scientific output

Major parts of this thesis are based upon work that has either been published or is in preparation for submission with the candidate as first author. In addition, the candidate has co-authored several publications of which minor parts are incorporated in the thesis. The full list of scientific outputs is listed below:

B.1 Publications in peer-review journals

- **Ali M.**, del Sol A. (2018) Modeling of Cellular Systems: Application in Stem Cell Research and Computational Disease Modeling. In: Alves Barbosa da Silva F., Carels N., Paes Silva Junior F. (eds) Theoretical and Applied Aspects of Systems Biology. Computational Biology, vol 27. Springer, Cham

B.2 Submissions in peer-review journals

- Lardenoije R. et al. (2019) The Alzheimer's disease DNA (hydroxy)methylome in the brain and blood. *Clinical Epigenetics*.
- Jung S., Ali M., and del Sol A. (2019) Methods in Epigenetics-based Systems Biology and their Applications. *Elsevier: Epigenetics Methods*.

B.3 Manuscripts in preparation

- **Ali M.**, et al. (2019) INTREGNET: Integrating epigenetic and transcriptional landscapes in a network-based model for increasing cellular conversion efficiency. *In preparation.*
- **Ali M.**, et al. (2019) Identification of causal genes for Alzheimer's disease using a network-based integrative analysis of genomic, epigenomic and transcriptomic data. *In preparation.*
- **Ali M.**, et al. (2019) The role of altered sphingolipid function in Alzheimer's disease; a gene regulatory network-based approach. *In preparation.*
- Giro G. G., et al. (2019) Modeling Juvenile Neuronal Ceroid Lipofuscinosis by genome editing in human induced pluripotent stem cells and cerebral organoids. *In preparation.*
- Giesert F., et al. (2019) Unique gene activity changes in ventral midbrain precede dopaminergic neuron degeneration in different PD models. *In preparation.*
- ENCODE-DREAM Consortium, **Ali M.**, et al. (2019) Systematic evaluation of multimodal approaches to predict *in vivo* DNA binding landscapes of regulatory proteins across cell types. *In preparation.*

B.4 Oral presentations in scientific conferences, symposia and workshops

- An Integrative Approach for Network inference from Epigenetics and Transcriptomics data (2017). *2nd Annual EPI-AD meeting.* London, UK.
- Reconstructing cell-type-specific networks by integrating multi-omics datasets (2018). *Dutch Neuroscience Meeting.* Lunteren, Netherlands.
- A computational approach for the identification of highly efficient instructive factors (2018). *3rd Annual EPI-AD meeting.* Barcelona, Spain.
- Neuroepigenetic: a life span perspective (2018). *EPI-AD/EURON Workshop.* Barcelona, Spain.

Appendix C

Supplementary figures and tables

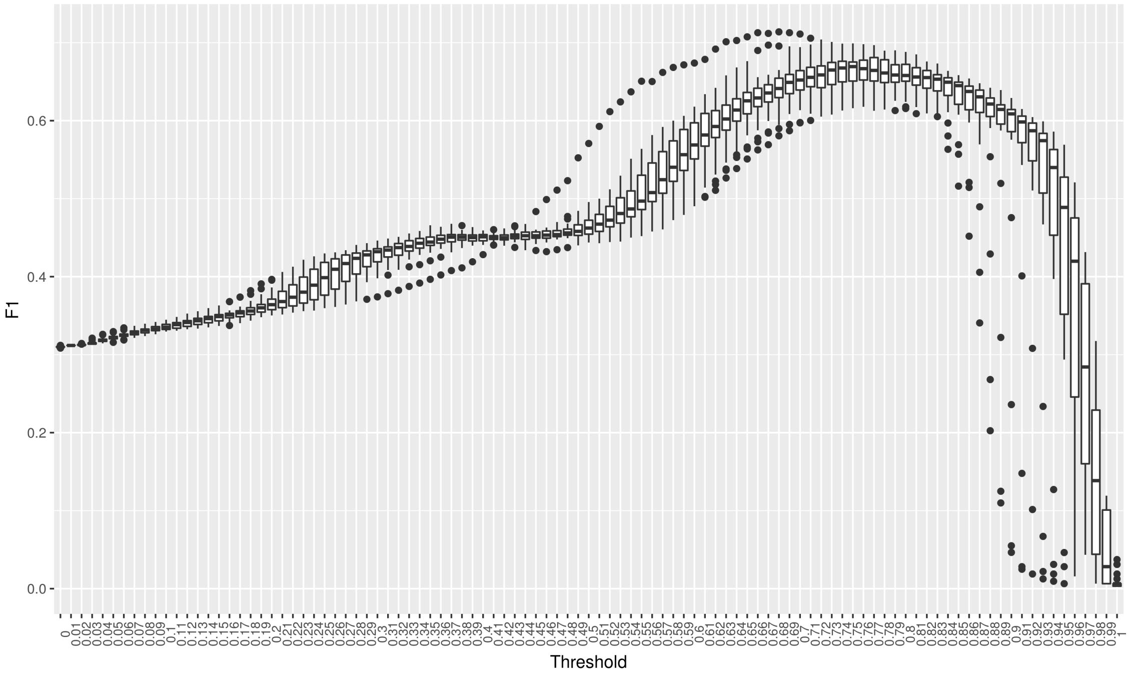


Figure 19: **Optimal correlation threshold** was considered to be 0.75. All samples having correlation higher than this were discarded from the background to compute JSD.

.

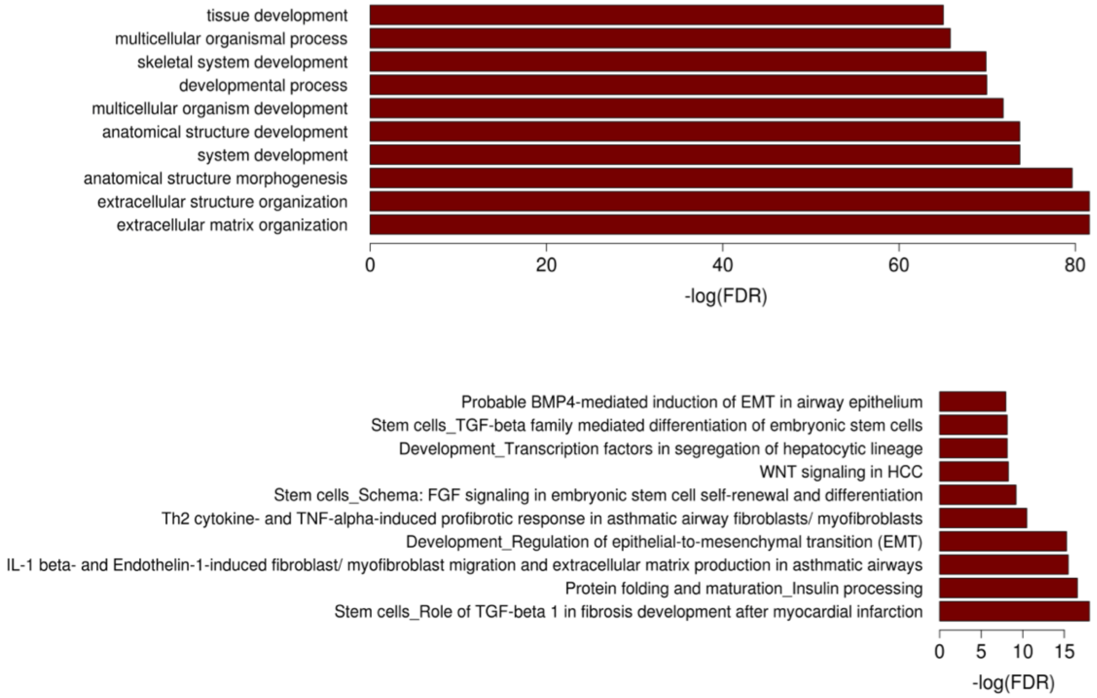


Figure 20: **Gene ontology enrichment analysis.** A) Gene enrichment analysis of disease network (top). Genes which are up-regulated in disease phenotype are significantly enriched in cellular processes associated to development. B) Pathway enrichment analysis of disease network (bottom). Genes which are up-regulated in disease phenotype are significantly enriched in developmental pathways.

.

Table 9: **Accession numbers of RNA-seq, DNase-Seq, H3K27ac and H3K4me3 histone marks** for all sample considered for reconstructing TRNs. The samples are taken from GEO, ENCODE and IHEC

Cell type/lines	RNA-Seq	DNase-Seq	H3K27ac	H3K4e3
A549	GSM1573117	ENCSR000ELW	ENCSR000AUI	ENCSR000DPD
Adipocytes	GSM1543671	GSM1443801	GSM1443807	ENCSR367VRA
BJ Fibroblast	GSM1510127	ENCSR000EME	GSM2401449	ENCSR000DQH
Cardiomyocytes	GSM1925978	GSE85630	GSM2280036	GSM2280016
ESCs	GSM1088317	ENCSR794OFW	ENCSR880SUY	ENCSR019SQX
Foreskin fibroblasts	GSM1588051	ENCSR251UPG	ENCSR822ZIG	ENCSR813CFB
GM12878	GSM754335	ENCSR000EJD	ENCSR000AKC	ENCSR057BWO
H9 ESCs	GSM1552696	ENCSR915BSC	ENCSR876RGF	ENCSR043VGU
HEK293	GSM1513689	GSM2392668	ENCSR000FCH	ENCSR000DTU
HeLaS3	ERR380552	ENCSR959ZXU	ENCSR000AOC	ENCSR340WQU
Hepatocytes	GSM1306654	ENCSR364MFN	ENCSR507UDH	ENCSR442ZOI
HepG2	GSM984650	ENCSR149XIL	ENCSR000AMO	ENCSR000AMP
CD34+ CMP	GSM976976	ENCSR468ZXN	ENCSR891KSP	ENCSR681HMF
HUVEC	GSM1273487	ENCSR000EOQ	ENCSR000ALB	ENCSR000AKN
iPSCs	GSM1088317	ENCSR261SMF	ENCSR875QDS	ENCSR263ELQ
K562	GSM1641262	ENCSR921NMD	ENCSR000AKP	ENCSR668LDD
Keratinocytes	GSM869035	ENCSR724CND	ENCSR666TFS	ENCSR703DFH
MCF7	GSM1817678	ENCSR000EPH	ENCSR752UOD	ENCSR985MIB
Melanocytes	GSM819489	ENCSR434OBM	ENCSR693VHX	ENCSR350JZR
Myoblasts	GSM1412725	ENCSR000EOO	ENCSR000ANF	ENCSR000ANK
Neuron	GSM1422448	ENCSR626RVD	ENCSR905TYC	ENCSR849YFO
NHDF	GSM1194807	ENCSR000EPO	ENCSR000APN	ENCSR000APR
NSCs	GSM1057334	ENCSR278FVO	ENCSR799SRL	ENCSR956CTX
Osteoblasts	GSM1333383	ENCSR000ELJ	ENCSR000APH	ENCSR000ATH
B cell	GSM1576394	ENCSR381PXW	ENCSR191ZQT	ENCSR939UQD
T cell	GSM1447398	ENCSR414IHC	ENCSR222QLW	ENCSR395YXN
Astrocyte	GSM1521786	ENCSR000EPM	ENCSR000AOQ	ENCSR000AOU
Lung Fibroblasts	GSM759890	ENCSR000EPR	ENCSR000AMR	ENCSR000AMW
HMECs	GSM721141	ENCSR000ENV	ENCSR000ALW	ENCSR000AML
Myotubes	GSM1412733	ENCSR000EOP	ENCSR000ANV	ENCSR000ANZ
SMCs	GSM1528677	GSM1024769	ENCSR210ZPC	ENCSR515PKY
Myotube	ENCFF320IDT	ENCFF026BDV	ENCFF345MCA	ENCFF044SEF
HMECs	ENCFF380GBC	ENCFF710XFX	ENCFF292XKK	ENCFF113WKS
Cardiac muscle cells	ENCFF888LPS	ENCFF054OJL	ENCFF214RHU	ENCFF190ZIS
Trophoblast	ENCFF342LYI	ENCFF334BDK	ENCFF698NII	ENCFF449TRE
Endodermal cells	ENCFF237ZQX	ENCFF168NOO	ENCFF587KQG	ENCFF385NQA
Mesendoderm (hESC)	ENCFF466QUZ	ENCFF993ETK	ENCFF318GQT	ENCFF293JHB
MSCs	ENCFF290OQE	ENCFF911JWG	ENCFF196AMI	ENCFF289UTW
NPCs	ENCFF789VZB	ENCFF315NGA	ENCFF874YBQ	ENCFF907ZOS
NPC (from H9)	ENCFF672VVX	ENCFF699MIZ	ENCFF779WYN	ENCFF076HNX
Astrocyte	ENCFF256APB	ENCFF558EUY	ENCFF040LCK	ENCFF254FYG
Monocyte CD14+	ENCFF299BIL	ENCFF581KXE	ENCFF039XWV	ENCFF640ZHV
Natural killer cell	ENCFF036GDL	ENCFF628EFJ	ENCFF240LSH	ENCFF505EGX
Megakaryocyte	S004BT	S004BT	S004BT	S004BT
Erythroblast	S002S3	S002S3	S002S3	S002S3
Monocyte CD16-	C005PS	C005PS	C005PS	C005PS
CD34+ CMPs	ENCFF690QPA	ENCFF846OZD	ENCFF660GJX	ENCFF020JLV
OCI-LY7	ENCFF773MOU	ENCFF190VGB	ENCFF929NXZ	ENCFF111JSX

Table 10: **Cellular conversion examples** with reported efficiency.

Initial cell type	Instructive factors (IFs)	Final cell type	Efficiency
HSC	SOX2	NSC	Low
Adult Foreskin	HNF1A,HNF4A,ONECUT1,CEBPA,ATF5,PROX1,TP53-siRNA,MYC	Hepatocytes	High
Fetal Limb Fibro	HNF1A,HNF4A,FOXA3	Hepatocytes	Low
Forehead Fibro	FOXA2,HNF4A,CEBPB,MYC	Hepatocytes	High
Forehead Fibro	FOXA2,HNF4A,CEBPB	Hepatocytes	Low
HSC (CD33+ cord blood cells)	POU5F1,SOX2,KLF4	H1ESC	High
HSC (CD33+ cord blood cells)	POU5F1,SOX2	H1ESC	Low
NHDF (Also in BJ)	POU5F1,SOX2,KLF4	H1ESC	High
NHDF (Also in BJ)	POU5F1,SOX2	H1ESC	Low
ForeskinFibro	KLF4,SOX2,POU5F1,MYC	ESC	High
ForeskinFibro	KLF4,SOX2,POU5F1	ESC	Low
ForeskinFibro (Also IMR90)	LIN28A,SOX2,POU5F1,NANOG	ESC	Low
Keratino	POU5F1,SOX2,KLF4	H1ESC	High
Keratino	POU5F1,SOX2,KLF4,MYC	H1ESC	Low
Keratino	POU5F1,SOX2	H1ESC	Low
NSC	POU5F1,KLF4	H9ESC	High
NSC	POU5F1	H9ESC	Low
Keratino	POU5F1,SOX2,KLF4,MYC	H1ESC	High
Keratino	POU5F1,SOX2,KLF4	H1ESC	Low
ForeskinFibro	CBX2,HES1,ID1,TFAP2A,ZFP42,ZNF423	NSC	Low
Fetal Fibro	ZNF521	NSC	High
NHDF	SOX2,PAX6	NSC	Low
H1 ESC (or iPSC)	NEUROG2/NEUROD1	Excitatory Neuron	High
H9ESC	POU3F2,ASCL1,MYT1L	Neuronal Cells	Low
ForeskinFibro & Fetal Fibro	POU3F2,ASCL1,NEUROD1	Neuronal Cells	High
ForeskinFibro & Fetal Fibro	POU3F2,ASCL1	Neuronal Cells	Low
ForeskinFibro & Fetal Fibro	POU3F2,ASCL1,MYT1L	Neuronal Cells	Low
Adult Lung Fibro	ASCL1, POU3F2, MYT1L	Neuron	NA
NHDF	MYOD1	Myoblasts	Low
Fetal Dermal Fibro	MITF,PAX3,SOX10	Melanocytes	NA
Keratino (HaCaT & MET-4)	MITF,LEF1,SOX10,SOX9	Melanocytes	NA
Keratino (HaCaT & MET-4)	MITF,PAX3,LEF1,SOX10,SOX9,SOX2	Melanocytes	NA
Neonatal ForeskinFibro	TP63,KLF4	Keratino	NA
HUVEC	GFI1,RUNX1,SPI1,FOSB	HSC	NA
MSCs (derived from iPSC)	CEBPB	Adipocytes	NA
MSCs (derived from iPSC)	PPARG	Adipocytes	NA
MSCs (derived from iPSC)	PPARG, CEBPB	Adipocytes	NA

Table 11: **Differential expression analysis results for SL genes**.

GeneName	logFC	Pval	GeneName	logFC	FDR_adj_Pval
STS	-0.221	0.000001	DAG1	0.12	0.149213
ARSG	-0.148	0.000011	SGPL1	0.09	0.15077
EZR	0.631	0.000017	GM2A	0.08	0.163169
ALOX12B	-0.195	0.000033	SGPP1	-0.057	0.174859
SIAT7E	-0.921	0.000033	CD177	0.02	0.177314
B3GALNT1	-0.387	0.000033	SPNS3	0.026	0.177586
GLTP	0.484	0.000111	FUT3	-0.019	0.183127
CLN8	0.269	0.000163	NEU3	0.022	0.197916
CD8A	-0.122	0.000179	LRP8	0.089	0.197916
MAL2	-0.994	0.000196	S1PR5	0.198	0.222959
TFPI	0.241	0.000226	ALDH3B1	0.026	0.226414
CSNK1G2	0.347	0.000272	ARSI	-0.015	0.253376
RFTN2	0.529	0.000333	NOS1AP	0.157	0.253376
KDSR	0.372	0.000388	SELP	0.031	0.253376
P2RX7	0.466	0.000528	FLOT1	-0.166	0.253376
PPM1L	-0.139	0.000538	ITGAM	0.121	0.253376
SMO	0.331	0.000538	SAMD8	-0.043	0.261887
VAPA	-0.279	0.000538	DEGS2	-0.02	0.262364
ST8SIA2	-0.11	0.000538	MYO1A	0.017	0.262364
ELOVL4	-0.633	0.000538	NEU4	0.201	0.262364
CDH13	-0.654	0.000571	ADD2	-0.018	0.262777
RFTN1	-0.382	0.000624	SCN5A	0.01	0.262777
EHD2	0.327	0.00075	NSMAF	0.054	0.262777
ST8SIA5	-0.319	0.000784	ARSF	-0.071	0.262777
PRKD1	0.301	0.000784	IL2	0.016	0.262777
AGK	-0.43	0.000784	KCNA5	-0.15	0.262777
ATP1A1	-0.553	0.000932	CAV3	0.021	0.262777
ANXA2	0.369	0.000932	IRS1	-0.073	0.262777

GBA	-0.206	0.001177	SPTLC1	-0.107	0.290484
CLIP3	-0.256	0.001177	FXYD1	-0.094	0.305446
PPT1	-0.303	0.001177	BAX	0.041	0.305446
NEU1	-0.237	0.001252	P2RX1	-0.024	0.306597
PPP2R1A	-0.416	0.001258	B3GALT2	-0.142	0.306816
BVES	0.127	0.001258	SIAT7A	0.121	0.307224
S1PR3	0.426	0.001297	KIF18A	0.017	0.310944
NOS3	0.428	0.001391	ST8SIA4	-0.037	0.324424
B4GALT6	-0.462	0.001518	GALC	-0.041	0.324424
ITGB8	0.196	0.001653	LIPE	0.071	0.324424
AKAP6	-0.322	0.002434	MAG	-0.215	0.33563
SRC	-0.132	0.002434	CD300LF	0.026	0.33563
TNFRSF1A	0.402	0.002806	RANGRF	0.084	0.340824
DLC1	0.604	0.002806	DEGS1	-0.071	0.357318
KCND2	-0.367	0.002806	FASLG	0.014	0.364778
ATP1B1	-0.768	0.002806	A3GALT2	0.011	0.383884
PPP2CA	-0.326	0.002806	PLA2G15	-0.048	0.383884
SERINC3	-0.233	0.002806	PEMT	-0.052	0.383884
CLN6	-0.169	0.002806	SGPP2	-0.079	0.383884
FUT7	0.076	0.002839	CHRNA3	0.014	0.383884
ITGB2	0.595	0.003333	ACER3	0.076	0.383884
ARV1	-0.297	0.003869	ALOXE3	-0.021	0.383884
SLC2A1	0.339	0.003895	LCP2	-0.017	0.384951
PRKD2	0.126	0.003925	TH	-0.021	0.384951
LRP6	0.108	0.003933	SLC22A6	-0.013	0.387013
LAPTM4B	-0.344	0.003933	ALDH3B2	0.01	0.387741
PLA2G6	-0.103	0.004259	CEL	0.099	0.409211
NOS1	-0.114	0.004742	CLIP1	-0.067	0.413476
ATP1B3	0.236	0.00488	KCNMA1	0.063	0.414249
KIT	-0.397	0.004889	S1PR1	0.131	0.418623

PSAPL1	0.054	0.004889	ALDH3A2	0.051	0.428702
ATP2B4	-0.242	0.00542	EFNA5	0.016	0.46573
PRKD3	0.11	0.005787	COL4A3BP	-0.055	0.473057
PLEKHA8	0.063	0.007247	FUT5	0.009	0.489551
ST3GAL5	-0.222	0.007779	CERKL	-0.01	0.489551
DOCK2	0.307	0.00832	UGT8	0.104	0.490274
MAPK1	-0.185	0.008686	SIAT7A	-0.019	0.495334
MYADM	-0.22	0.010966	REEP2	0.078	0.495334
ENPP7	0.06	0.011654	TRAF2	0.023	0.507386
SPHK2	0.189	0.012052	SMPD2	0.012	0.524609
TGFBR2	0.408	0.012325	ELOVL2	-0.058	0.536481
ASAH1	-0.049	0.012325	ARSJ	0.015	0.548505
ST8SIA3	-0.427	0.012482	NAGA	-0.019	0.549817
HMOX1	0.229	0.012492	P2RY12	-0.108	0.554417
CLN3	0.08	0.013016	SUMF1	0.035	0.560218
TRPC4	0.054	0.013176	JAK2	-0.044	0.577173
DLG1	0.102	0.013176	CYR61	0.107	0.577173
ELOVL6	-0.154	0.0139	PLLP	-0.099	0.577173
MYOF	0.301	0.015835	FA2H	-0.081	0.595061
CD2	0.052	0.016233	SERINC2	0.011	0.598642
RTN4R	-0.19	0.016656	FAM57B	0.016	0.600442
ORMDL2	0.078	0.018476	SERINC5	-0.01	0.61055
ARSA	0.097	0.020241	TEX2	-0.03	0.615876
SIAT7C	0.158	0.020241	ABCB1	0.083	0.619912
ADRA1A	0.022	0.020241	ARSE	0.01	0.626264
S100A10	0.354	0.021262	TNF	0.011	0.626264
SPHK1	0.038	0.021262	CAV2	0.035	0.628475
SMPD1	-0.096	0.022814	ATP1A2	0.069	0.628475
SERINC1	-0.414	0.023649	ABCA12	0.005	0.664369
BMPR2	-0.253	0.023649	HDAC6	-0.042	0.673743

PACSIN2	0.211	0.026042	S1PR4	0.015	0.673743
PRKACA	-0.039	0.026238	SELL	0.018	0.701221
PSAP	-0.113	0.026651	PRKCD	-0.034	0.708357
RALA	0.125	0.026726	MALL	0.032	0.710837
PRKAR1A	-0.353	0.029255	ST8SIA1	0.027	0.710837
ELOVL3	-0.036	0.030793	GBA2	0.017	0.715851
ADRA1B	-0.291	0.030962	CDH2	-0.027	0.722193
ARSK	-0.134	0.033417	SPTLC3	-0.008	0.765263
SGMS2	0.036	0.035346	PLVAP	-0.009	0.775108
CAV1	0.29	0.035346	MLC1	0.018	0.798132
FUT6	0.15	0.036918	HCK	0.025	0.799639
S1PR2	-0.044	0.038326	GAL3ST1	-0.018	0.801166
ORMDL3	0.167	0.046252	ARSB	0.008	0.805402
SELPLG	0.077	0.046424	MAL	0.042	0.805402
NPC1	0.211	0.047133	UGCG	-0.027	0.805976
GLA	0.081	0.051585	EMP2	0.005	0.805976
BMPR1A	0.127	0.055808	ACER1	0.003	0.805976
SMPD4	0.091	0.055808	SPRED1	0.027	0.813718
B4GALT3	0.119	0.055808	PTGS2	0.034	0.830598
ASAH2	-0.016	0.064646	HTRA2	0.008	0.865576
SPNS1	0.211	0.064735	F2R	0.013	0.865576
PTGIS	-0.043	0.065345	ACER2	0.003	0.878838
ADCYAP1R1	-0.028	0.065596	HEXA	-0.003	0.896144
ALDH5A1	-0.173	0.071878	CMTM8	-0.017	0.896144
PRTN3	-0.197	0.071878	FLOT2	0.01	0.896144
B3GALT4	0.079	0.073885	B3GALT1	0.002	0.932972
ELOVL7	0.172	0.087444	ESYT1	-0.005	0.945216
CERK	0.092	0.100127	ARSD	-0.005	0.954864
ABCC1	-0.02	0.100127	ARSH	-0.001	0.975073
ORMDL1	-0.084	0.100127	B4GALNT1	0.006	0.975637

GPR6	-0.125	0.106324	ST8SIA6	0.001	0.97661
MAPK3	0.117	0.10701	PLTP	-0.009	0.979097
SPTLC2	0.072	0.10886	SGMS1	-0.001	0.990524
SPNS2	0.108	0.122663	SMPDL3B	0	0.992681
MARVELD1	0.044	0.122663	SFTPB	0	0.992681
ELOVL1	0.115	0.122663	NEU2	0	0.996678
GLB1	0.087	0.138449	PRKAA1	0	0.996678
HEXB	-0.146	0.140718	PRKAR2A	0	0.996678

Table 12: **List of included manually selected GO terms**, ordered by each subtree of the GO (Biological Process, Cellular Component, and Molecular Function). Two terms were excluded before proceeding, as they are too generic (Lipid metabolic process, and Membrane raft).

BP (Biological Process)	MF (Molecular Function)	CC (Cellular Component)
caveola_assembly.gpml	ceramide_binding.gpml	caveola.gpml
ceramide_biosynthetic_process.gpml	glycosphingolipid_binding.gpml	plasma_membrane_raft.gpml
ceramide_catabolic_process.gpml	sphingolipid_binding.gpml	SPOTS_complex.gpml
ceramide_metabolic_process.gpml	sphingolipid_transporter_activity.gpml	
ceramide_transport.gpml	sphingosine-1-phosphate_receptor_activity.gpml	
galactosylceramide_metabolic_process.gpml		
ganglioside_biosynthetic_process.gpml		
ganglioside_catabolic_process.gpml		
ganglioside_metabolic_process.gpml		
glucosylceramide_metabolic_process.gpml		
glycosphingolipid_biosynthetic_process.gpml		
glycosphingolipid_catabolic_process.gpml		
glycosphingolipid_metabolic_process.gpml		
glycosylceramide_biosynthetic_process.gpml		
glycosylceramide_catabolic_process.gpml		
glycosylceramide_metabolic_process.gpml		
membrane_raft_assembly.gpml		
membrane_raft_distribution.gpml		
membrane_raft_localization.gpml		
membrane_raft_organization.gpml		
membrane_raft_polarization.gpml		
negative_regulation_of_sphingolipid_biosynthetic_process.gpml		
phytosphingosine_metabolic_process.gpml		
plasma_membrane_raft_assembly.gpml		
plasma_membrane_raft_organization.gpml		
positive_regulation_of_ceramide_biosynthetic_process.gpml		
positive_regulation_of_sphingolipid_biosynthetic_process.gpml		
protein_transport_into_membrane_raft.gpml		
protein_transport_into_plasma_membrane_raft.gpml		
regulation_of_ceramide_biosynthetic_process.gpml		
regulation_of_sphingolipid_biosynthetic_process.gpml		
sphinganine_metabolic_process.gpml		
sphingoid_biosynthetic_process.gpml		
sphingoid_metabolic_process.gpml		
sphingolipid_biosynthetic_process.gpml		
sphingolipid_catabolic_process.gpml		
sphingolipid_mediated_signaling_pathway.gpml		
sphingolipid_metabolic_process.gpml		
sphingomyelin_biosynthetic_process.gpml		
sphingomyelin_catabolic_process.gpml		
sphingomyelin_metabolic_process.gpml		
sphingosine-1-phosphate_receptor_signaling_pathway.gpml		
sphingosine_biosynthetic_process.gpml		
sphingosine_metabolic_process.gpml		

Table 13: **Textual representations of the subtree of the Biological Process;** GO tree containing the selected sphingolipid related terms. Parent-child dependency between terms is indicated by indentation. An asterisk '*' marks each but the first occurrence of a term that is present multiple times in the subtree.

GO:0006665 sphingolipid metabolic process		
GO:0006665 sphingolipid metabolic process		
GO:0006672 ceramide metabolic process		
	GO:0006677 glycosylceramide metabolic process	
		GO:0006681 galactosylceramide metabolic process
		GO:0006678 glucosylceramide metabolic process
		GO:0046477 glycosylceramide catabolic process
		GO:0046476 glycosylceramide biosynthetic process
	GO:0046514 ceramide catabolic process	
		GO:0006689 ganglioside catabolic process
		GO:0046477 glycosylceramide catabolic process *
	GO:0046513 ceramide biosynthetic process	
		GO:0001574 ganglioside biosynthetic process
		GO:0046476 glycosylceramide biosynthetic process *
	GO:0001573 ganglioside metabolic process	
		GO:0001574 ganglioside biosynthetic process *
		GO:0006689 ganglioside catabolic process *
GO:0006684 sphingomyelin metabolic process		
	GO:0006685 sphingomyelin catabolic process	
	GO:0006686 sphingomyelin biosynthetic process	
GO:0006687 glycosphingolipid metabolic process		
	GO:0006677 glycosylceramide metabolic process *	
		GO:0006681 galactosylceramide metabolic process *
		GO:0006678 glucosylceramide metabolic process *
		GO:0046477 glycosylceramide catabolic process *
		GO:0046476 glycosylceramide biosynthetic process *
	GO:0001573 ganglioside metabolic process *	
		GO:0001574 ganglioside biosynthetic process *
		GO:0006689 ganglioside catabolic process *
	GO:0046479 glycosphingolipid catabolic process	
		GO:0006689 ganglioside catabolic process *
		GO:0046477 glycosylceramide catabolic process *
	GO:0006688 glycosphingolipid biosynthetic process	
		GO:0001574 ganglioside biosynthetic process *
		GO:0046476 glycosylceramide biosynthetic process *
GO:0046519 sphingoid metabolic process		
	GO:0006667 sphinganine metabolic process	
	GO:0046520 sphingoid biosynthetic process	
		GO:0046512 sphingosine biosynthetic process
	GO:0006670 sphingosine metabolic process	
		GO:0046512 sphingosine biosynthetic process *
GO:0030149 sphingolipid catabolic process		
	GO:0006685 sphingomyelin catabolic process *	
	GO:0046514 ceramide catabolic process *	
		GO:0006689 ganglioside catabolic process *
		GO:0046477 glycosylceramide catabolic process *
	GO:0046479 glycosphingolipid catabolic process *	
		GO:0006689 ganglioside catabolic process *
		GO:0046477 glycosylceramide catabolic process *
GO:0030148 sphingolipid biosynthetic process		
	GO:0046520 sphingoid biosynthetic process *	
		GO:0046512 sphingosine biosynthetic process *
	GO:0046513 ceramide biosynthetic process *	
		GO:0001574 ganglioside biosynthetic process *
		GO:0046476 glycosylceramide biosynthetic process *
	GO:0006686 sphingomyelin biosynthetic process *	
	GO:0006688 glycosphingolipid biosynthetic process *	
		GO:0001574 ganglioside biosynthetic process *
		GO:0046476 glycosylceramide biosynthetic process *
GO:0090153 regulation of sphingolipid biosynthetic process		
GO:2000303 regulation of ceramide biosynthetic process		
	GO:2000304 positive regulation of ceramide biosynthetic process	
GO:0090154 positive regulation of sphingolipid biosynthetic process		
	GO:2000304 positive regulation of ceramide biosynthetic process *	
GO:0090155 negative regulation of sphingolipid biosynthetic process		
GO:0090520 sphingolipid mediated signaling pathway		
GO:0003376 sphingosine-1-phosphate signaling pathway		
GO:0031579 membrane raft organization		
GO:0044857 plasma membrane raft organization		
	GO:0044854 plasma membrane raft assembly	
		GO:0070836 caveola assembly
GO:0031580 membrane raft distribution		
	GO:0001766 membrane raft polarization	
GO:0001765 membrane raft assembly		
	GO:0044854 plasma membrane raft assembly	
		GO:0070836 caveola assembly
GO:0006629 lipid metabolic process		
GO:0051665 membrane raft localization		
GO:0031580 membrane raft distribution *		
	GO:0001766 membrane raft polarization *	
GO:0032596 protein transport into membrane raft		
GO:0044861 protein transport into plasma membrane raft		

Table 14: **Textual representations of the subtree of the Cellular Component;** GO tree containing the selected sphingolipid related terms. Parent-child dependency between terms is indicated by indentation. An asterisk '*' marks each but the first occurrence of a term that is present multiple times in the subtree.

GO:0035339 SPOTS complex		
GO:0045121 membrane raft		
	GO:0044853 plasma membrane raft	
		GO:0005901 caveola

Table 15: **Textual representations of the subtree of the Molecular Function;** GO tree containing the selected sphingolipid related terms. Parent-child dependency between terms is indicated by indentation. An asterisk '*' marks each but the first occurrence of a term that is present multiple times in the subtree.

GO:0046625 sphingolipid binding	
	GO:0043208 glycosphingolipid binding
	GO:0097001 ceramide binding
GO:0046624 sphingolipid transporter activity	
GO:0038036 sphingosine-1-phosphate receptor activity	

Table 16: **Benchmarking of inferred networks against gold-standard core networks.**

Cell type	Gold-Standard Network		Inferred Network		Match
	Source	Target	Source	Target	
ESC	NANOG	NANOG	NANOG	NANOG	Yes
ESC	NANOG	POU5F1	NANOG	POU5F1	Yes
ESC	NANOG	SOX2	NANOG	SOX2	Yes
ESC	POU5F1	NANOG	POU5F1	NANOG	Yes
ESC	POU5F1	POU5F1	POU5F1	POU5F1	Yes
ESC	POU5F1	SOX2	POU5F1	SOX2	Yes
ESC	SOX2	NANOG	SOX2	NANOG	Yes
ESC	SOX2	POU5F1	SOX2	POU5F1	Yes
ESC	SOX2	SOX2	SOX2	SOX2	Yes
Hepatocyte	ONECUT1	HNF4A	ONECUT1	HNF4A	Yes
Hepatocyte	ONECUT1	ONECUT1	ONECUT1	ONECUT1	Yes
Hepatocyte	FOXA2	FOXA2	FOXA2	FOXA2	Yes
Hepatocyte	HNF4A	FOXA2	HNF4A	FOXA2	Yes
Hepatocyte	HNF4A	HNF1A	HNF4A	HNF1A	Yes
Hepatocyte	HNF4A	HNF4A	HNF4A	HNF4A	Yes
Hepatocyte	HNF1A	HNF1A	HNF1A	HNF1A	Yes
Hepatocyte	HNF1A	HNF4A	HNF1A	HNF4A	Yes
Hepatocyte	CREB1	CREB1	HNF1A	ONECUT1	No
Hepatocyte	CREB1	FOXA2	HNF1A	FOXA2	No
Hepatocyte	USF1	ONECUT1	FOXA2	HNF1A	No
Hepatocyte	HNF4A	USF1	FOXA2	HNF4A	No
Hepatocyte	ONECUT1	FOXA2			No
HepG2	HNF4A	CEBPB	HNF4A	CEBPB	Yes
HepG3	HNF4A	FOXA1	HNF4A	FOXA1	Yes
HepG4	HNF4A	FOXA2	HNF4A	FOXA2	Yes
HepG5	HNF4A	HNF4A	HNF4A	HNF4A	Yes
HepG6	FOXA2	CEBPB	FOXA2	CEBPB	Yes

HepG7	FOXA2	FOXA1	FOXA2	FOXA1	Yes
HepG8	FOXA2	FOXA2	FOXA2	FOXA2	Yes
HepG9	FOXA2	HNF4A	FOXA2	HNF4A	Yes
HepG10	FOXA1	CEBPB	FOXA1	CEBPB	Yes
HepG11	FOXA1	FOXA1	FOXA1	FOXA1	Yes
HepG12	FOXA1	FOXA2	FOXA1	FOXA2	Yes
HepG13	FOXA1	HNF4A	FOXA1	HNF4A	Yes
HepG14	CEBPB	CEBPB	CEBPB	CEBPB	Yes
HepG15	CEBPB	FOXA1	CEBPB	FOXA1	Yes
HepG16	CEBPB	FOXA2	CEBPB	FOXA2	Yes
HepG17	CEBPB	HNF4A	CEBPB	HNF4A	Yes
MCF7	ESR1	ESR1	ESR1	ESR1	Yes
MCF8	ESR1	FOXA1	ESR1	FOXA1	Yes
MCF9	FOXA1	ESR1	FOXA1	ESR1	Yes
MCF10	FOXA1	FOXA1	FOXA1	FOXA1	Yes
MCF11	ESR1	FOSL2			No
MCF12	ESR1	JUND			No
MCF13	FOSL2	ESR1			No
MCF14	FOSL2	FOSL2			No
MCF15	FOSL2	FOXA1			No
MCF16	FOSL2	JUND			No
MCF17	JUND	ESR1			No
MCF18	JUND	FOSL2			No
MCF19	JUND	FOXA1			No

Table 17: **Accession numbers of RNA-seq sample considered as the background for JSD computation.** The samples are taken from GEO [50], ENCODE [54] and IHEC consortiums.

Accession ID	Accession ID	Accession ID	Accession ID	Accession ID	Accession ID
GSM417715	GSM1156942	GSM1158474	GSM1519568	GSM1695867	GSM1023079
GSM417716	GSM1156943	GSM1158475	GSM1519569	GSM1695868	GSM1023080
GSM453868	GSM1156944	GSM1158476	GSM1519570	GSM1695869	GSM1023081
GSM453869	GSM1156945	GSM1158477	GSM1519571	GSM1695905	GSM1023082
GSM485364	GSM1156946	GSM1158478	GSM1521768	GSM1695906	GSM1023083
GSM485365	GSM1156947	GSM1158479	GSM1521769	GSM1695907	GSM1023084
GSM485366	GSM1156948	GSM1158480	GSM1521770	GSM1695908	GSM1023085
GSM485367	GSM1156949	GSM1158481	GSM1521771	GSM1695909	GSM1023086
GSM485368	GSM1156950	GSM1158482	GSM1521772	GSM1695910	GSM1023087
GSM485369	GSM1156951	GSM1158483	GSM1521773	GSM1695911	GSM1030556
GSM485370	GSM1156952	GSM1158484	GSM1521774	GSM1695912	GSM1030557
GSM485371	GSM1156953	GSM1158485	GSM1521775	GSM1695913	GSM1033470
GSM485372	GSM1156954	GSM1158486	GSM1521776	GSM1697912	GSM1033471
GSM485373	GSM1156955	GSM1158487	GSM1521777	GSM1697913	GSM1033472
GSM485374	GSM1156956	GSM1158488	GSM1521778	GSM1701465	GSM1033473
GSM485375	GSM1156957	GSM1158489	GSM1521779	GSM1701466	GSM1033474
GSM485376	GSM1156958	GSM1158490	GSM1521780	GSM1701467	GSM1037852
GSM485377	GSM1156959	GSM1158491	GSM1521781	GSM1701468	GSM1037853
GSM485378	GSM1156960	GSM1158492	GSM1521782	GSM1701469	GSM1037854
GSM485379	GSM1156961	GSM1158493	GSM1521783	GSM1701470	GSM1037855
GSM485380	GSM1156962	GSM1158494	GSM1521784	GSM1701471	GSM1037856
GSM485381	GSM1156963	GSM1158495	GSM1521785	GSM1701472	GSM1053748
GSM485382	GSM1156964	GSM1158496	GSM1521786	GSM1701473	GSM1053749
GSM485383	GSM1156965	GSM1158497	GSM1524370	GSM1701474	GSM1053750
GSM485384	GSM1156966	GSM1158498	GSM1524371	GSM1701475	GSM1053751
GSM485385	GSM1156967	GSM1158499	GSM1524869	GSM1701476	GSM1053752
GSM485386	GSM1156968	GSM1158500	GSM1524870	GSM1701478	GSM1053753

GSM485387	GSM1156969	GSM1158501	GSM1524871	GSM1701479	GSM1053764
GSM485388	GSM1156970	GSM1158502	GSM1524872	GSM1704298	GSM1053765
GSM485389	GSM1156971	GSM1158503	GSM1524873	GSM1704299	GSM1053766
GSM485390	GSM1156972	GSM1158504	GSM1524874	GSM1704300	GSM1053767
GSM485391	GSM1156973	GSM1158505	GSM1524875	GSM1704301	GSM1053768
GSM485392	GSM1156974	GSM1158506	GSM1524876	GSM1704302	GSM1053769
GSM485393	GSM1156975	GSM1158507	GSM1527072	GSM1704303	GSM1053770
GSM485394	GSM1156976	GSM1158508	GSM1527073	GSM1704304	GSM1053771
GSM485395	GSM1156977	GSM1158509	GSM1527074	GSM1704305	GSM1053772
GSM485396	GSM1156978	GSM1158510	GSM1527075	GSM1704839	GSM1053773
GSM485397	GSM1156979	GSM1158511	GSM1527076	GSM1704840	GSM1053774
GSM485398	GSM1156980	GSM1158512	GSM1527077	GSM1704841	GSM1053775
GSM485399	GSM1156981	GSM1158513	GSM1528672	GSM1704842	GSM1053776
GSM485400	GSM1156982	GSM1158514	GSM1528673	GSM1704843	GSM1053777
GSM485401	GSM1156983	GSM1158515	GSM1528674	GSM1704844	GSM1053778
GSM485402	GSM1156984	GSM1158516	GSM1528675	GSM1704845	GSM1053779
GSM485403	GSM1156985	GSM1158517	GSM1528676	GSM1704846	GSM1053780
GSM485404	GSM1156986	GSM1158518	GSM1528677	GSM1704847	GSM1053781
GSM485405	GSM1156987	GSM1158519	GSM1528678	GSM1704848	GSM1053782
GSM485406	GSM1156988	GSM1158520	GSM1528679	GSM1704849	GSM1053783
GSM485407	GSM1156989	GSM1158521	GSM1528680	GSM1704850	GSM1053784
GSM485408	GSM1156990	GSM1158522	GSM1528681	GSM1704851	GSM1053785
GSM485410	GSM1156991	GSM1158523	GSM1528682	GSM1704852	GSM1053786
GSM485411	GSM1156992	GSM1158524	GSM1528683	GSM1704853	GSM1053787
GSM485412	GSM1156993	GSM1158525	GSM1528684	GSM1704854	GSM1053788
GSM485413	GSM1156994	GSM1158526	GSM1528685	GSM1704855	GSM1053789
GSM485414	GSM1156995	GSM1158527	GSM1529688	GSM1704856	GSM1053790
GSM485415	GSM1156996	GSM1158528	GSM1529689	GSM1704857	GSM1053791
GSM485416	GSM1156997	GSM1158529	GSM1529690	GSM1704858	GSM1053792
GSM485417	GSM1156998	GSM1158530	GSM1529691	GSM1707595	GSM1053793

GSM485418	GSM1156999	GSM1158531	GSM1529692	GSM1707596	GSM1053794
GSM485419	GSM1157000	GSM1158532	GSM1532279	GSM1707597	GSM1053795
GSM485420	GSM1157001	GSM1158533	GSM1532280	GSM1707598	GSM1053796
GSM485421	GSM1157002	GSM1158534	GSM1532281	GSM1712007	GSM1053797
GSM485422	GSM1157003	GSM1158535	GSM1532282	GSM1712008	GSM1053798
GSM485423	GSM1157004	GSM1158536	GSM1532283	GSM1712009	GSM1053799
GSM485424	GSM1157005	GSM1158537	GSM1532284	GSM1712010	GSM1053800
GSM485425	GSM1157006	GSM1158538	GSM1532285	GSM1712011	GSM1057332
GSM485426	GSM1157007	GSM1158539	GSM1532286	GSM1712012	GSM1057333
GSM485427	GSM1157008	GSM1158540	GSM1533239	GSM1712013	GSM1057334
GSM485428	GSM1157009	GSM1158541	GSM1533240	GSM1712014	GSM1207205
GSM485429	GSM1157010	GSM1158542	GSM1533241	GSM1712015	GSM1207206
GSM485430	GSM1157011	GSM1158543	GSM1533242	GSM1712016	GSM1207207
GSM485431	GSM1157012	GSM1158544	GSM1533243	GSM1712017	GSM1060352
GSM485432	GSM1157013	GSM1158545	GSM1533244	GSM1712018	GSM1060353
GSM485433	GSM1157014	GSM1158546	GSM1533245	GSM1712019	GSM1060354
GSM485434	GSM1157015	GSM1158547	GSM1533246	GSM1712020	GSM1060355
GSM485435	GSM1157016	GSM1158548	GSM1533247	GSM1712021	GSM1060356
GSM485436	GSM1157017	GSM1158549	GSM1533248	GSM1712022	GSM1060357
GSM485437	GSM1157018	GSM1158550	GSM1533249	GSM1712023	GSM1060358
GSM485438	GSM1157019	GSM1158551	GSM1533250	GSM1712024	GSM1060359
GSM485439	GSM1157020	GSM1158552	GSM1533257	GSM1714397	GSM1060360
GSM485440	GSM1157021	GSM1158553	GSM1533258	GSM721696	GSM1062234
GSM485441	GSM1157022	GSM1158554	GSM1533259	GSM721697	GSM1062235
GSM485442	GSM1157023	GSM1158555	GSM1533260	GSM721698	GSM1062236
GSM485443	GSM1157024	GSM1158556	GSM1533261	GSM721699	GSM1062237
GSM485444	GSM1157025	GSM1158557	GSM1694954	GSM721700	GSM1062238
GSM485445	GSM1157026	GSM1158558	GSM1694956	GSM721701	GSM1062239
GSM485446	GSM1157027	GSM1158559	GSM1694957	GSM1717523	GSM1062240
GSM485447	GSM1157028	GSM1158560	GSM1694958	GSM1717524	GSM1062241

GSM485448	GSM1157029	GSM1158561	GSM1694959	GSM1717525	GSM1062242
GSM485449	GSM1157030	GSM1158562	GSM1694960	GSM1717526	GSM1062243
GSM485450	GSM1157031	GSM1158563	GSM1694961	GSM1717527	GSM1062244
GSM485451	GSM1157032	GSM1158564	GSM1694962	GSM1717528	GSM1062245
GSM485452	GSM1157033	GSM1158565	GSM1694963	GSM1717529	GSM1062246
GSM485453	GSM1157034	GSM1158566	GSM1694964	GSM1717530	GSM1062247
GSM485454	GSM1157035	GSM1158567	GSM1694965	GSM1717531	GSM1062248
GSM485455	GSM1157036	GSM1158568	GSM1694966	GSM1717532	GSM1062249
GSM485456	GSM1157037	GSM1158569	GSM1694967	GSM1717533	GSM1062250
GSM485457	GSM1157038	GSM1158570	GSM1694969	GSM1717534	GSM1062251
GSM485458	GSM1157039	GSM1158571	GSM1694968	GSM1717535	GSM1063280
GSM485459	GSM1157040	GSM1158572	GSM1694970	GSM1717536	GSM1063281
GSM485460	GSM1157041	GSM1158573	GSM1694971	GSM1717537	GSM1063282
GSM485461	GSM1157042	GSM1158574	GSM1694972	GSM1717538	GSM1063283
GSM485462	GSM1157043	GSM1158575	GSM1694973	GSM1717539	GSM1063284
GSM485463	GSM1157044	GSM1158576	GSM1694974	GSM1717540	GSM1063285
GSM485464	GSM1157045	GSM1158577	GSM1694975	GSM1717541	GSM1063286
GSM485465	GSM1157046	GSM1158578	GSM1694976	GSM1717542	GSM1063287
GSM485466	GSM1157047	GSM1158579	GSM1694977	GSM1717543	GSM1063288
GSM485467	GSM1157048	GSM1158580	GSM1694978	GSM1717544	GSM1063289
GSM485468	GSM1157049	GSM1158581	GSM1694979	GSM1717545	GSM1063290
GSM485469	GSM1157050	GSM1158582	GSM1694980	GSM1717546	GSM1063291
GSM485470	GSM1157051	GSM1158583	GSM1694981	GSM1717547	GSM1063292
GSM485471	GSM1157052	GSM1158584	GSM1694982	GSM1717548	GSM1063293
GSM485472	GSM1157053	GSM1158585	GSM1694983	GSM1717549	GSM1063294
GSM485473	GSM1157054	GSM1158586	GSM1536176	GSM1717550	GSM1063295
GSM485474	GSM1157055	GSM1158587	GSM1536177	GSM1717551	GSM1063296
GSM485475	GSM1157056	GSM1158588	GSM1536178	GSM1717552	GSM1063297
GSM485476	GSM1157057	GSM1158589	GSM1536179	GSM1717553	GSM1063298
GSM485477	GSM1157058	GSM1158590	GSM1536180	GSM1717554	GSM1063299

GSM485478	GSM1157059	GSM1158591	GSM1536181	GSM1717555	GSM1063300
GSM485479	GSM1157060	GSM1158592	GSM1536182	GSM1717556	GSM1063301
GSM485480	GSM1157061	GSM1158593	GSM1536183	GSM1717557	GSM1063302
GSM485481	GSM1157062	GSM1158594	GSM1536184	GSM1717558	GSM1063303
GSM485482	GSM1157063	GSM1158595	GSM1536185	GSM1717559	GSM1063304
GSM485483	GSM1157064	GSM1158596	GSM1536186	GSM1717560	GSM1064826
GSM485484	GSM1157065	GSM1158597	GSM1536187	GSM1717561	GSM1064829
GSM485485	GSM1157066	GSM1158598	GSM1536188	GSM1717562	GSM1196045
GSM485486	GSM1157067	GSM1158599	GSM1536189	GSM1717563	GSM1065157
GSM485487	GSM1157068	GSM1158600	GSM1536190	GSM1717564	GSM1065160
GSM485488	GSM1157069	GSM1158601	GSM1536191	GSM1717565	GSM1065161
GSM485489	GSM1157070	GSM1158602	GSM1536192	GSM1717566	GSM1065162
GSM485490	GSM1157071	GSM1158603	GSM1536193	GSM1717567	GSM1065917
GSM485491	GSM1157072	GSM1158604	GSM1536247	GSM1717568	GSM1065918
GSM485492	GSM1157073	GSM1158605	GSM1536248	GSM1717569	GSM1065919
GSM485493	GSM1157074	GSM1158606	GSM1536249	GSM1717570	GSM1065920
GSM485494	GSM1157075	GSM1158607	GSM1536250	GSM1717571	GSM1065921
GSM485495	GSM1157076	GSM1158608	GSM1536429	GSM1717572	GSM1065922
GSM485496	GSM1157077	GSM1158609	GSM1536430	GSM1717573	GSM1065923
GSM485497	GSM1157078	GSM1158610	GSM1536431	GSM1717574	GSM1065924
GSM485498	GSM1157079	GSM1158611	GSM1536432	GSM1717575	GSM1065925
GSM485499	GSM1157080	GSM1158612	GSM1536433	GSM1717576	GSM1065926
GSM485500	GSM1157081	GSM1158613	GSM1536434	GSM1717577	GSM1065927
GSM485501	GSM1157082	GSM1158614	GSM1536435	GSM1717578	GSM1065928
GSM485502	GSM1157083	GSM1158615	GSM1536436	GSM1717579	GSM1065929
GSM485503	GSM1157084	GSM1158616	GSM1536437	GSM1717580	GSM1065930
GSM485504	GSM1157085	GSM1158617	GSM1536438	GSM1717581	GSM1065931
GSM485505	GSM1157086	GSM1158618	GSM1537303	GSM1717582	GSM1065932
GSM485506	GSM1157087	GSM1158619	GSM1537304	GSM1717583	GSM1076106
GSM485507	GSM1157088	GSM1158620	GSM1543665	GSM1717584	GSM1076107

GSM485508	GSM1157089	GSM1158621	GSM1543667	GSM1717585	GSM1076108
GSM485509	GSM1157090	GSM1158622	GSM1543670	GSM1717586	GSM1088317
GSM485510	GSM1157091	GSM1158623	GSM1543671	GSM1717587	GSM1088318
GSM485511	GSM1157092	GSM1158624	GSM1545029	GSM1717588	GSM1088319
GSM485512	GSM1157093	GSM1158625	GSM1545030	GSM1717589	GSM1088201
GSM485513	GSM1157094	GSM1158626	GSM1545031	GSM1717590	GSM1088202
GSM485514	GSM1157095	GSM1158627	GSM1545032	GSM1717591	GSM1088203
GSM485515	GSM1157096	GSM1158628	GSM1545033	GSM1717592	GSM1088204
GSM485516	GSM1157097	GSM1364030	GSM1545034	GSM1717593	GSM1088205
GSM485517	GSM1157098	GSM1364031	GSM1545035	GSM1717594	GSM1088206
GSM485518	GSM1157099	GSM1364032	GSM1545036	GSM1717595	GSM1088207
GSM485519	GSM1157100	GSM1364033	GSM1546371	GSM1717596	GSM1088208
GSM485520	GSM1157101	GSM1364034	GSM1546372	GSM1717597	GSM1088209
GSM485521	GSM1157102	GSM1364035	GSM1547996	GSM1717598	GSM1088210
GSM485522	GSM1157103	GSM1364036	GSM1547997	GSM1717599	GSM1088211
GSM485523	GSM1157104	GSM1364037	GSM1547998	GSM1717600	GSM1088212
GSM485524	GSM1157105	GSM1364038	GSM1547999	GSM1717601	GSM1088213
GSM432598	GSM1157106	GSM1364039	GSM1548000	GSM1717602	GSM1088214
GSM432600	GSM1157107	GSM1364040	GSM1548001	GSM1717603	GSM1088215
GSM432601	GSM1157108	GSM1364041	GSM1548002	GSM1717604	GSM1088216
GSM432602	GSM1157109	GSM1364042	GSM1548003	GSM1717605	GSM1088217
GSM432603	GSM1157110	GSM1364043	GSM1548004	GSM1717606	GSM1088218
GSM432604	GSM1157111	GSM1364044	GSM1548005	GSM1717607	GSM1088219
GSM432605	GSM1157112	GSM1368999	GSM1548006	GSM1717608	GSM1088220
GSM432606	GSM1157113	GSM1369000	GSM1548007	GSM1717609	GSM1088221
GSM432607	GSM1157114	GSM1369001	GSM1548008	GSM1717610	GSM1088222
GSM432608	GSM1157115	GSM1369002	GSM1548009	GSM1717611	GSM1088223
GSM432609	GSM1157116	GSM1369003	GSM1548010	GSM1717612	GSM1088224
GSM424320	GSM1157181	GSM1369004	GSM1548011	GSM1717613	GSM1088225
GSM424321	GSM1157182	GSM1369005	GSM1548012	GSM1717614	GSM1088226

GSM424322	GSM1157183	GSM1369006	GSM1548013	GSM1717615	GSM1088227
GSM424323	GSM1157184	GSM1369007	GSM1548014	GSM1717616	GSM1088228
GSM424324	GSM1157185	GSM1369008	GSM1550090	GSM1717617	GSM1088229
GSM424325	GSM1157186	GSM1369009	GSM1550091	GSM1717618	GSM1088230
GSM424326	GSM1157187	GSM1369010	GSM1550092	GSM1717619	GSM1088231
GSM424327	GSM1157188	GSM1369011	GSM1550093	GSM1717620	GSM1088232
GSM424328	GSM1157189	GSM1369012	GSM1550094	GSM1717621	GSM1088233
GSM424329	GSM1157190	GSM1369013	GSM1550095	GSM1717622	GSM1088234
GSM424330	GSM1157191	GSM1369014	GSM1550096	GSM1717623	GSM1088235
GSM424331	GSM1157192	GSM1369015	GSM1550097	GSM1717624	GSM1088236
GSM424332	GSM1157193	GSM1369016	GSM1550098	GSM1717625	GSM1088237
GSM424333	GSM1157194	GSM1369017	GSM1550099	GSM1717626	GSM1088238
GSM424334	GSM1157195	GSM1369018	GSM1550100	GSM1717627	GSM1088239
GSM424335	GSM1157196	GSM1369182	GSM1550101	GSM1717628	GSM1088241
GSM424336	GSM1157197	GSM1369183	GSM1550102	GSM1717629	GSM1088242
GSM424337	GSM1157198	GSM1369184	GSM1550103	GSM1717630	GSM1088243
GSM424338	GSM1157199	GSM1369185	GSM1550104	GSM1717631	GSM1088244
GSM424339	GSM1157200	GSM1369187	GSM1550105	GSM1717632	GSM1088245
GSM424340	GSM1157201	GSM1369188	GSM1550106	GSM1717633	GSM1088246
GSM424341	GSM1157202	GSM1369189	GSM1550107	GSM1717634	GSM1088247
GSM424342	GSM1157203	GSM1369190	GSM1550108	GSM1717635	GSM1088248
GSM424343	GSM1157204	GSM1369191	GSM1550109	GSM1717636	GSM1088249
GSM424344	GSM1157205	GSM1369192	GSM1550110	GSM1717637	GSM1088250
GSM424345	GSM1157206	GSM1369193	GSM1550111	GSM1717638	GSM1088251
GSM424346	GSM1157207	GSM1369194	GSM1550112	GSM1717639	GSM1088252
GSM424347	GSM1157208	GSM1369195	GSM1550113	GSM1717640	GSM1088253
GSM424348	GSM1157209	GSM1369196	GSM1550114	GSM1717641	GSM1088254
GSM424349	GSM1157210	GSM1369197	GSM1415906	GSM1717642	GSM1088255
GSM424350	GSM1157211	GSM1369198	GSM1415907	GSM1717643	GSM1088256
GSM424351	GSM1157212	GSM1369199	GSM1415908	GSM1717644	GSM1088257

GSM424352	GSM1157213	GSM1369200	GSM1415909	GSM1717645	GSM1088258
GSM424353	GSM1157214	GSM1369201	GSM1415910	GSM1717646	GSM1088259
GSM424354	GSM1157215	GSM1369202	GSM1415911	GSM1717647	GSM1088260
GSM424355	GSM1157216	GSM1369203	GSM1551308	GSM1717648	GSM1088261
GSM424356	GSM1157217	GSM1369204	GSM1551309	GSM1717649	GSM1088262
GSM424357	GSM1157218	GSM1369205	GSM1551310	GSM1717650	GSM1088263
GSM424358	GSM1157219	GSM1369206	GSM1552693	GSM1717651	GSM1088264
GSM424359	GSM1157220	GSM1369207	GSM1552694	GSM1717652	GSM1088265
GSM424360	GSM1157221	GSM1369208	GSM1552695	GSM1717653	GSM1088266
GSM480870	GSM1157222	GSM1369209	GSM1552696	GSM1717654	GSM1088267
GSM480871	GSM1157223	GSM1369210	GSM1552807	GSM1717655	GSM1088268
GSM480872	GSM1157224	GSM1369211	GSM1552808	GSM1717656	GSM1088269
GSM480873	GSM1157225	GSM1369212	GSM1552809	GSM1717657	GSM1088270
GSM517435	GSM1157226	GSM1369213	GSM1552810	GSM1717658	GSM1088271
GSM517437	GSM1157227	GSM1369214	GSM1552811	GSM1717659	GSM1088272
GSM517438	GSM1157228	GSM1369215	GSM1552812	GSM1717660	GSM1088273
GSM517439	GSM1157229	GSM1369216	GSM1592570	GSM1717661	GSM1088274
GSM517441	GSM1157230	GSM1369217	GSM1592571	GSM1717662	GSM1088275
GSM517442	GSM1157231	GSM1369218	GSM1592572	GSM1717663	GSM1088276
GSM501716	GSM1157232	GSM1369219	GSM1592573	GSM1717664	GSM1088277
GSM484893	GSM1157233	GSM1369220	GSM1553085	GSM1717665	GSM1088278
GSM484894	GSM1157234	GSM1369221	GSM1553086	GSM1717666	GSM1088279
GSM484895	GSM1157235	GSM1369222	GSM1553087	GSM1717667	GSM1091810
GSM484896	GSM1157236	GSM1369223	GSM1553088	GSM1717668	GSM1091811
GSM484897	GSM1157237	GSM1369224	GSM1553089	GSM1717669	GSM1093060
GSM484898	GSM1157238	GSM1369225	GSM1553090	GSM1717670	GSM1093061
GSM484899	GSM1157239	GSM1370364	GSM1553091	GSM1717671	GSM1095135
GSM484900	GSM1157240	GSM1372330	GSM1553092	GSM1717672	GSM1095139
GSM484901	GSM1157241	GSM1372331	GSM1553093	GSM1717673	GSM1095140
GSM484902	GSM1157242	GSM1372332	GSM1553094	GSM1717674	GSM1095141

GSM484903	GSM1157243	GSM1372333	GSM1553095	GSM1717675	GSM1097887
GSM484904	GSM1157244	GSM1372334	GSM1553096	GSM1717676	GSM1097888
GSM484905	GSM1157245	GSM1372335	GSM1553097	GSM1717677	GSM1098196
GSM484906	GSM1157246	GSM1372336	GSM1553098	GSM1717678	GSM1098197
GSM494809	GSM1157247	GSM1372337	GSM1553099	GSM1717679	GSM1098198
GSM494810	GSM1157248	GSM1372338	GSM1553100	GSM1717680	GSM1098199
GSM530678	GSM1157249	GSM1372339	GSM1553101	GSM1717681	GSM1098200
GSM475204	GSM1157250	GSM1372340	GSM1553102	GSM1717682	GSM1098201
GSM475205	GSM1157251	GSM1372341	GSM1553103	GSM1717683	GSM1098202
GSM475206	GSM1157252	GSM1372342	GSM1553104	GSM1717684	GSM1098203
GSM475207	GSM1157253	GSM1372343	GSM1553105	GSM1717685	GSM1098204
GSM475208	GSM1157254	GSM1372344	GSM1553106	GSM1717686	GSM1098205
GSM475209	GSM1157255	GSM1372345	GSM1553107	GSM1717687	GSM1098206
GSM546438	GSM1157256	GSM1372346	GSM1553108	GSM1717688	GSM1098207
GSM546439	GSM1157257	GSM1371574	GSM1553109	GSM1717689	GSM1098208
GSM546440	GSM1157258	GSM1371576	GSM1553110	GSM1717690	GSM1098209
GSM546441	GSM1157259	GSM1371577	GSM1553111	GSM1717691	GSM1098210
GSM546442	GSM1157260	GSM1371580	GSM1915560	GSM1717692	GSM1098211
GSM546443	GSM1157261	GSM1371583	GSM1915561	GSM1717693	GSM1098212
GSM546444	GSM1157262	GSM1371584	GSM1915562	GSM1717694	GSM1098213
GSM563061	GSM1157263	GSM1375212	GSM1915563	GSM1717695	GSM1098214
GSM574244	GSM1157264	GSM1375213	GSM1915564	GSM1717696	GSM1098215
GSM597207	GSM1157265	GSM1376804	GSM1915565	GSM1717697	GSM1098216
GSM597208	GSM1157266	GSM1376805	GSM1915566	GSM1717698	GSM1098217
GSM597209	GSM1157267	GSM1376806	GSM1915567	GSM1717699	GSM1098218
GSM597210	GSM1157268	GSM1376807	GSM1915568	GSM1717700	GSM1098219
GSM597211	GSM1157269	GSM1376808	GSM1915569	GSM1717701	GSM1098220
GSM601403	GSM1157270	GSM1376809	GSM1915570	GSM1717702	GSM1098221
GSM601404	GSM1157271	GSM1376810	GSM1915571	GSM1717703	GSM1098222
GSM601405	GSM1157272	GSM1376811	GSM1915572	GSM1717704	GSM1098223

GSM601406	GSM1157273	GSM1377536	GSM1915573	GSM1717705	GSM1098224
GSM601407	GSM1157274	GSM1377537	GSM1915574	GSM1717706	GSM1098225
GSM601408	GSM1157275	GSM1378372	GSM1915575	GSM1717707	GSM1098226
GSM602557	GSM1157276	GSM1378373	GSM1915576	GSM1717708	GSM1098227
GSM602559	GSM1157277	GSM1380867	GSM1915577	GSM1717709	GSM1098228
GSM602561	GSM1157278	GSM1380868	GSM1553412	GSM1717710	GSM1098229
GSM602563	GSM1157279	GSM1381226	GSM1553413	GSM1717711	GSM1098230
GSM602565	GSM1157280	GSM1381227	GSM1553414	GSM1717712	GSM1098231
GSM602567	GSM1157281	GSM1381228	GSM1553415	GSM1717713	GSM1098232
GSM602569	GSM1157282	GSM1381229	GSM1554463	GSM1717714	GSM1098233
GSM602571	GSM1157283	GSM1381230	GSM1554464	GSM1724087	GSM1098234
GSM602573	GSM1157284	GSM1381231	GSM1554465	GSM1724088	GSM1098235
GSM602575	GSM1157285	GSM1381984	GSM1554466	GSM1724089	GSM1098236
GSM602577	GSM1157286	GSM1381985	GSM1554467	GSM1724090	GSM1098237
GSM602579	GSM1157287	GSM1381986	GSM1554468	GSM1724091	GSM1098238
GSM602581	GSM1157288	GSM1381987	GSM1556288	GSM1724092	GSM1098239
GSM602583	GSM1157289	GSM1381988	GSM1556289	GSM1726439	GSM1098240
GSM602585	GSM1157290	GSM1381989	GSM1556290	GSM1726440	GSM1098241
GSM602587	GSM1157291	GSM1381990	GSM1556291	GSM1726441	GSM1098242
GSM602589	GSM1157292	GSM1381991	GSM1556292	GSM1726442	GSM1098243
GSM602591	GSM1157293	GSM1381992	GSM1556293	GSM1726443	GSM1098244
GSM602593	GSM1157294	GSM1381993	GSM1556294	GSM1726444	GSM1098245
GSM602595	GSM1157295	GSM1381994	GSM1556295	GSM1726445	GSM1098246
GSM614544	GSM1157296	GSM1381995	GSM1556296	GSM1726446	GSM1098247
GSM614545	GSM1157297	GSM1381996	GSM1556297	GSM1726447	GSM1098248
GSM651905	GSM1157298	GSM1381997	GSM1556298	GSM1726448	GSM1098249
GSM651906	GSM1157299	GSM1381998	GSM1556299	GSM1726449	GSM1098250
GSM651907	GSM1157300	GSM1381999	GSM1558381	GSM1726450	GSM1098251
GSM651908	GSM1157301	GSM1382000	GSM1558415	GSM1726451	GSM1098252
GSM1241350	GSM1157302	GSM1382001	GSM1558416	GSM1726452	GSM1098253

GSM1241351	GSM1157303	GSM1382002	GSM1559439	GSM1726453	GSM1098254
GSM1241352	GSM1157304	GSM1382003	GSM1559440	GSM1726454	GSM1098255
GSM1241353	GSM1157305	GSM1382004	GSM1559441	GSM1726455	GSM1098256
GSM1241354	GSM1157306	GSM1382005	GSM1559442	GSM1726456	GSM1098257
GSM1241355	GSM1157307	GSM1382006	GSM1559443	GSM1726457	GSM1098258
GSM1241356	GSM1157308	GSM1382007	GSM1559444	GSM1726458	GSM1098259
GSM1241357	GSM1157309	GSM1382008	GSM1560720	GSM1726459	GSM1098260
GSM1241358	GSM1157310	GSM1382009	GSM1560721	GSM1726460	GSM1098261
GSM1241359	GSM1157311	GSM1382010	GSM1560722	GSM1726461	GSM1098262
GSM1241360	GSM1157312	GSM1382011	GSM1560723	GSM1726462	GSM1098263
GSM1241361	GSM1157313	GSM1382012	GSM1560866	GSM1726463	GSM1098264
GSM1241362	GSM1157314	GSM1382013	GSM1560867	GSM1726464	GSM1098265
GSM1241363	GSM1157315	GSM1382014	GSM1560868	GSM1726465	GSM1098266
GSM1241364	GSM1157316	GSM1382015	GSM1560869	GSM1726466	GSM1098267
GSM1241365	GSM1157317	GSM1382016	GSM1561610	GSM1726467	GSM1098268
GSM1241366	GSM1157319	GSM1382017	GSM1561611	GSM1726468	GSM1098269
GSM1241367	GSM1157320	GSM1382018	GSM1561612	GSM1726469	GSM1098270
GSM1241368	GSM1157321	GSM1382019	GSM1561613	GSM1726470	GSM1098271
GSM1241369	GSM1157322	GSM1382020	GSM1561617	GSM1726471	GSM1098272
GSM1241370	GSM1157323	GSM1382021	GSM1561619	GSM1726472	GSM1098273
GSM1241371	GSM1157324	GSM1382022	GSM1561620	GSM1726473	GSM1098274
GSM1241372	GSM1157325	GSM1382023	GSM1561621	GSM1726474	GSM1098275
GSM1241373	GSM1157326	GSM1382024	GSM1561622	GSM1726475	GSM1098276
GSM1241374	GSM1157327	GSM1382025	GSM1561623	GSM1726476	GSM1098277
GSM1241375	GSM1157328	GSM1382026	GSM1561624	GSM1807973	GSM1098278
GSM1241376	GSM1157329	GSM1382027	GSM1561625	GSM1807984	GSM1098279
GSM1241377	GSM1157330	GSM1382028	GSM1561626	GSM1807985	GSM1098280
GSM1241378	GSM1157331	GSM1382029	GSM1561627	GSM1807986	GSM1098281
GSM1241379	GSM1157332	GSM1382030	GSM1561628	GSM1807987	GSM1098282
GSM1241380	GSM1157333	GSM1382031	GSM1561629	GSM1807974	GSM1098283

GSM1241381	GSM1157334	GSM1382032	GSM1561630	GSM1807975	GSM1098284
GSM1241382	GSM1157335	GSM1382033	GSM1561631	GSM1807976	GSM1098285
GSM1241383	GSM1157336	GSM1382034	GSM1561632	GSM1807977	GSM1098286
GSM1241384	GSM1157337	GSM1382035	GSM1561633	GSM1807978	GSM1098287
GSM1241385	GSM1157338	GSM1382036	GSM1561634	GSM1807979	GSM1098288
GSM1241386	GSM1157339	GSM1382037	GSM1561635	GSM1807980	GSM1098289
GSM1241387	GSM1157340	GSM1382453	GSM1561636	GSM1807981	GSM1098290
GSM1243306	GSM1157341	GSM1383903	GSM1561637	GSM1807982	GSM1098291
GSM1243307	GSM1157342	GSM1383904	GSM1561638	GSM1807983	GSM1098292
GSM1243308	GSM1157343	GSM1383905	GSM1561639	GSM1807988	GSM1098293
GSM1243309	GSM1157344	GSM1383906	GSM1561640	GSM1807989	GSM1098294
GSM1242494	GSM1157345	GSM1383907	GSM1561642	GSM1807990	GSM1098295
GSM1242495	GSM1157346	GSM1383908	GSM1561643	GSM1807991	GSM1098296
GSM1242496	GSM1157347	GSM1383909	GSM1561644	GSM1807992	GSM1098297
GSM1242497	GSM1157348	GSM1383910	GSM1561645	GSM1807993	GSM1098298
GSM1242498	GSM1157349	GSM1383911	GSM1561646	GSM1807994	GSM1098299
GSM1242499	GSM1157350	GSM1383912	GSM1561647	GSM1807995	GSM1098300
GSM1242500	GSM1157351	GSM1383913	GSM1561648	GSM1807996	GSM1098301
GSM1242501	GSM1157352	GSM1383914	GSM1561649	GSM1807997	GSM1098302
GSM1242502	GSM1157353	GSM1383915	GSM1560004	GSM1807998	GSM1098303
GSM1242503	GSM1157354	GSM1383916	GSM1560005	GSM1807999	GSM1098304
GSM1242510	GSM1157355	GSM1383917	GSM1560006	GSM1808000	GSM1098305
GSM1245898	GSM1157356	GSM1383918	GSM1560010	GSM1808001	GSM1098306
GSM1245899	GSM1157357	GSM1386272	GSM1560011	GSM1808002	GSM1098307
GSM1245900	GSM1157358	GSM1386273	GSM1560012	GSM1808003	GSM1098308
GSM1245901	GSM1157359	GSM1386274	GSM1560019	GSM1808004	GSM1098309
GSM1246806	GSM1157360	GSM1386275	GSM1560020	GSM1808005	GSM1098310
GSM1246807	GSM1157361	GSM1386276	GSM1560021	GSM1808006	GSM1098311
GSM1246808	GSM1157362	GSM1386277	GSM1566740	GSM1808007	GSM1098312
GSM1246809	GSM1157363	GSM1386278	GSM1566741	GSM1808008	GSM1098313

GSM1246810	GSM1157364	GSM1386279	GSM1566742	GSM1808009	GSM1098314
GSM1246811	GSM1157365	GSM1386280	GSM1566743	GSM1808010	GSM1098315
GSM1246812	GSM1157366	GSM1386281	GSM1566744	GSM1808011	GSM1098316
GSM1246813	GSM1157367	GSM1386282	GSM1566745	GSM1808012	GSM1098317
GSM1246814	GSM1157368	GSM1386283	GSM1566746	GSM1808013	GSM1098318
GSM1246815	GSM1157369	GSM1386284	GSM1566747	GSM1808014	GSM1098319
GSM1246816	GSM1157370	GSM1386285	GSM1566748	GSM1808015	GSM1098320
GSM1246817	GSM1157371	GSM1386286	GSM1566749	GSM1808016	GSM1098321
GSM1246818	GSM1157372	GSM1386287	GSM1566750	GSM1808017	GSM1098322
GSM1246819	GSM1157373	GSM1394656	GSM1566751	GSM1808018	GSM1098323
GSM1246820	GSM1157374	GSM1394657	GSM1566752	GSM1808019	GSM1098324
GSM1246821	GSM1157375	GSM1395289	GSM1566753	GSM1808020	GSM1098325
GSM1246822	GSM1157376	GSM1395290	GSM1709930	GSM1808021	GSM1098326
GSM1246823	GSM1157377	GSM1395291	GSM1709931	GSM1808022	GSM1098327
GSM1246824	GSM1157378	GSM1395292	GSM1709932	GSM1808023	GSM1098328
GSM1246825	GSM1157379	GSM1395293	GSM1709933	GSM1808024	GSM1098329
GSM1254205	GSM1157380	GSM1395294	GSM1709934	GSM1808025	GSM1098330
GSM1259263	GSM1157381	GSM1395295	GSM1709935	GSM1808026	GSM1098331
GSM1259264	GSM1157382	GSM1395296	GSM1709936	GSM1808027	GSM1098332
GSM1260479	GSM1157383	GSM1395297	GSM1709937	GSM1808028	GSM1098333
GSM1260481	GSM1157384	GSM1395298	GSM1567911	GSM1808029	GSM1098334
GSM1260483	GSM1157385	GSM1395299	GSM1567912	GSM1808030	GSM1098335
GSM1260485	GSM1157386	GSM1395300	GSM1567913	GSM1808031	GSM1098336
GSM1260487	GSM1157387	GSM1395301	GSM1567914	GSM1808032	GSM1098337
GSM1260489	GSM1157388	GSM1395302	GSM1567915	GSM1808033	GSM1098338
GSM1260491	GSM1157389	GSM1395303	GSM1567916	GSM1808034	GSM1098339
GSM1260493	GSM1157390	GSM1395304	GSM1567917	GSM1808035	GSM1098340
GSM1260495	GSM1157391	GSM1395305	GSM1567918	GSM1808036	GSM1098341
GSM1260497	GSM1157392	GSM1395306	GSM1567919	GSM1808037	GSM1098342
GSM1260499	GSM1157393	GSM1395307	GSM1567920	GSM1808038	GSM1098343

GSM1260501	GSM1157394	GSM1395308	GSM1567921	GSM1808039	GSM1098344
GSM1260503	GSM1157395	GSM1395309	GSM1567922	GSM1808040	GSM1098345
GSM1260505	GSM1157396	GSM1395311	GSM1567923	GSM1808041	GSM1098346
GSM1260507	GSM1157397	GSM1395312	GSM1567924	GSM1808042	GSM1098347
GSM1260509	GSM1157398	GSM1395313	GSM1567925	GSM1808043	GSM1098348
GSM1260511	GSM1157399	GSM1395314	GSM1567926	GSM1808044	GSM1098349
GSM1260513	GSM1157400	GSM1395315	GSM1567927	GSM1808045	GSM1098350
GSM1260515	GSM1157401	GSM1395316	GSM1567928	GSM1808046	GSM1098351
GSM1260517	GSM1157402	GSM1396537	GSM1567929	GSM1808047	GSM1098352
GSM1260519	GSM1157403	GSM1396538	GSM1567930	GSM1808048	GSM1098353
GSM1260521	GSM1157404	GSM1396539	GSM1567931	GSM1808049	GSM1098354
GSM1260523	GSM1157405	GSM1396585	GSM1567932	GSM1808050	GSM1098355
GSM1260525	GSM1157406	GSM1396586	GSM1567933	GSM1808051	GSM1098356
GSM1260527	GSM1157407	GSM1396587	GSM1567934	GSM1808052	GSM1098357
GSM1260529	GSM1157408	GSM1396590	GSM1567935	GSM1808053	GSM1098358
GSM1260531	GSM1157409	GSM1396591	GSM1567936	GSM1808054	GSM1098359
GSM1260533	GSM1157410	GSM1396593	GSM1567937	GSM1808055	GSM1098360
GSM1260535	GSM1157411	GSM1396595	GSM1567938	GSM1808056	GSM1098361
GSM1260537	GSM1157412	GSM1396598	GSM1567939	GSM1808057	GSM1098362
GSM1260539	GSM1157413	GSM1396600	GSM1567940	GSM1808058	GSM1098363
GSM1260541	GSM1157414	GSM1396601	GSM1567941	GSM1808059	GSM1098364
GSM1260543	GSM1157415	GSM1396602	GSM1567942	GSM1808060	GSM1098365
GSM1260545	GSM1157416	GSM1396603	GSM1567943	GSM1808061	GSM1098366
GSM1260547	GSM1157417	GSM1396604	GSM1567944	GSM1808062	GSM1098367
GSM1260549	GSM1157418	GSM1396605	GSM1567945	GSM1808063	GSM1098368
GSM1260551	GSM1157419	GSM1396606	GSM1568709	GSM1808064	GSM1098369
GSM1260553	GSM1157420	GSM1396607	GSM1568710	GSM1808065	GSM1098370
GSM1260555	GSM1157421	GSM1396608	GSM1568711	GSM1808066	GSM1098371
GSM1260557	GSM1157422	GSM1396609	GSM1568712	GSM1808718	GSM1098372
GSM1260559	GSM1157423	GSM1397514	GSM1571055	GSM1808719	GSM1098373

GSM1260561	GSM1157424	GSM1397515	GSM1571056	GSM1816172	GSM1098374
GSM1260563	GSM1157425	GSM1397516	GSM1571057	GSM1816174	GSM1098375
GSM1260565	GSM1157426	GSM1397742	GSM1571058	GSM1816175	GSM1098376
GSM1260567	GSM1157427	GSM1399180	GSM1571059	GSM1816176	GSM1098377
GSM1260569	GSM1157428	GSM1399181	GSM1571060	GSM1816177	GSM1098378
GSM1260571	GSM1157429	GSM1399182	GSM1571061	GSM1817212	GSM1098379
GSM1260573	GSM1157430	GSM1399183	GSM1571062	GSM1817213	GSM1098380
GSM1260575	GSM1157431	GSM1399184	GSM1571063	GSM1817214	GSM1098381
GSM1260577	GSM1157432	GSM1399185	GSM1571064	GSM1817215	GSM1098382
GSM1260579	GSM1157433	GSM1399186	GSM1571065	GSM1817216	GSM1098383
GSM1260581	GSM1157434	GSM1399187	GSM1571066	GSM1817217	GSM1098384
GSM1260583	GSM1157435	GSM1399188	GSM1571067	GSM1817678	GSM1098385
GSM1260585	GSM1157436	GSM1399189	GSM1571068	GSM1829628	GSM1098386
GSM1335668	GSM1157437	GSM1399190	GSM1571069	GSM1830134	GSM1098387
GSM1335670	GSM1157438	GSM1399191	GSM1571070	GSM1830135	GSM1098388
GSM1335672	GSM1157439	GSM1399192	GSM1571071	GSM1830136	GSM1098389
GSM1335674	GSM1157440	GSM1399193	GSM1571072	GSM1830137	GSM1098390
GSM1335676	GSM1157441	GSM1399196	GSM1571073	GSM1830782	GSM1098391
GSM1335678	GSM1157442	GSM1399197	GSM1571074	GSM1830783	GSM1098392
GSM1335680	GSM1157443	GSM1399198	GSM1571075	GSM1830784	GSM1098393
GSM1335682	GSM1157444	GSM1399199	GSM1571076	GSM1830785	GSM1098394
GSM1335684	GSM1157445	GSM1399200	GSM1571077	GSM1830786	GSM1098395
GSM1335686	GSM1157446	GSM1399201	GSM1571078	GSM1830787	GSM1098572
GSM1335688	GSM1157447	GSM1399202	GSM1571079	GSM1830788	GSM1098573
GSM1335690	GSM1157448	GSM1399203	GSM1571080	GSM1830789	GSM1098574
GSM1335692	GSM1157449	GSM1399204	GSM1571081	GSM1836551	GSM1098575
GSM1335694	GSM1157450	GSM1399205	GSM1571082	GSM1836552	GSM1100205
GSM1335696	GSM1157451	GSM1399206	GSM1571083	GSM1836553	GSM1100206
GSM1335698	GSM1157452	GSM1399207	GSM1571084	GSM1836554	GSM1100295
GSM1335702	GSM1157453	GSM1399208	GSM1571085	GSM1836555	GSM1100296

GSM1335704	GSM1157454	GSM1399209	GSM1571086	GSM1836556	GSM1100297
GSM1335706	GSM1157455	GSM1399210	GSM1571087	GSM1836573	GSM1100298
GSM1335708	GSM1157456	GSM1400982	GSM1571088	GSM1836574	GSM1100299
GSM1335710	GSM1157457	GSM1400983	GSM1571089	GSM1836575	GSM1100300
GSM1335712	GSM1157458	GSM1401303	GSM1571090	GSM1836576	GSM1100301
GSM1335714	GSM1157459	GSM1401320	GSM1571091	GSM1836577	GSM1100302
GSM1335716	GSM1157460	GSM1401321	GSM1571092	GSM1836578	GSM1100303
GSM1335718	GSM1157461	GSM1401324	GSM1571093	GSM1836579	GSM1100304
GSM1335720	GSM1157462	GSM1401325	GSM1571094	GSM1836580	GSM1100305
GSM1335722	GSM1157463	GSM1401326	GSM1571095	GSM1836581	GSM1100306
GSM1335724	GSM1157464	GSM1401327	GSM1571096	GSM1836582	GSM1100307
GSM1335726	GSM1157465	GSM1401328	GSM1571097	GSM1836583	GSM1100308
GSM1335728	GSM1157466	GSM1401329	GSM1571098	GSM1836584	GSM1101966
GSM1335730	GSM1157467	GSM1401330	GSM1571099	GSM1836622	GSM1101967
GSM1335732	GSM1157468	GSM1401331	GSM1571100	GSM1836623	GSM1101968
GSM1335734	GSM1157469	GSM1401332	GSM1571101	GSM1836624	GSM1101969
GSM1335736	GSM1157470	GSM1401333	GSM1571102	GSM1836625	GSM1101970
GSM1335738	GSM1157471	GSM1401334	GSM1571103	GSM1836626	GSM1101971
GSM1335740	GSM1157472	GSM1401335	GSM1571104	GSM1836627	GSM1101972
GSM1335742	GSM1157473	GSM1401336	GSM1571105	GSM1836628	GSM1101973
GSM1335744	GSM1157474	GSM1401337	GSM1571106	GSM1836629	GSM1101974
GSM1335746	GSM1157475	GSM1401338	GSM1571107	GSM1836630	GSM1101975
GSM1335748	GSM1157476	GSM1401339	GSM1571108	GSM1836631	GSM1101976
GSM1335750	GSM1157477	GSM1401340	GSM1571109	GSM1836632	GSM1101977
GSM1335752	GSM1157478	GSM1401341	GSM1571110	GSM1836633	GSM1104010
GSM1335754	GSM1157479	GSM1401342	GSM1571111	GSM1836634	GSM1104011
GSM1335756	GSM1157480	GSM1401343	GSM1571112	GSM1836635	GSM1104012
GSM1261033	GSM1157541	GSM1401344	GSM1571113	GSM1836636	GSM1104013
GSM1261034	GSM1157542	GSM1401345	GSM1571114	GSM1842233	GSM1104014
GSM1261035	GSM1157543	GSM1401346	GSM1571115	GSM1842234	GSM1104015

GSM1261651	GSM1157544	GSM1401347	GSM1571116	GSM1842235	GSM1104016
GSM1261652	GSM1157545	GSM1401348	GSM1571117	GSM1842236	GSM1104017
GSM1261653	GSM1157546	GSM1401349	GSM1571118	GSM1842237	GSM1104129
GSM1261654	GSM1157547	GSM1401350	GSM1571119	GSM1842238	GSM1104130
GSM1266739	GSM1157548	GSM1401351	GSM1571120	GSM1842239	GSM1104131
GSM1266740	GSM1157549	GSM1401352	GSM1571121	GSM1842240	GSM1105766
GSM1266741	GSM1157550	GSM1401353	GSM1571122	GSM1842241	GSM1105767
GSM1266742	GSM1157551	GSM1401354	GSM1571123	GSM1842242	GSM1105768
GSM1266743	GSM1157552	GSM1401355	GSM1571124	GSM1842243	GSM1105769
GSM1266744	GSM1157553	GSM1401356	GSM1571125	GSM1842244	GSM1105770
GSM1266745	GSM1157554	GSM1401357	GSM1576159	GSM1843468	GSM1105771
GSM1266746	GSM1157555	GSM1401358	GSM1576391	GSM1843469	GSM1105772
GSM1255335	GSM1157556	GSM1401359	GSM1576392	GSM1843471	GSM1105773
GSM1255336	GSM1157557	GSM1401360	GSM1576393	GSM1843472	GSM1105774
GSM1273487	GSM1157558	GSM1401361	GSM1576394	GSM1403191	GSM1105775
GSM1273488	GSM1157559	GSM1401362	GSM1576395	GSM1847138	GSM1105776
GSM1273672	GSM1157560	GSM1401363	GSM1576396	GSM1847139	GSM1105777
GSM1273673	GSM1157561	GSM1401364	GSM1576397	GSM1847140	GSM1105778
GSM1273674	GSM1157562	GSM1401365	GSM1576398	GSM1847141	GSM1105779
GSM1273675	GSM1157563	GSM1401366	GSM1576399	GSM1847142	GSM1105780
GSM1273676	GSM1157564	GSM1401367	GSM1576400	GSM1847143	GSM1105781
GSM1273677	GSM1157565	GSM1401368	GSM1576401	GSM1857483	GSM1105782
GSM1277968	GSM1157566	GSM1401377	GSM1576402	GSM1857484	GSM1105783
GSM1277969	GSM1157567	GSM1401378	GSM1576403	GSM1857485	GSM1105784
GSM1277970	GSM1157568	GSM1401379	GSM1576404	GSM1865616	GSM1105785
GSM1277971	GSM1157569	GSM1401380	GSM1576405	GSM1865617	GSM1105786
GSM1277972	GSM1157570	GSM1402482	GSM1576406	GSM1865618	GSM1105787
GSM1277973	GSM1157571	GSM1402483	GSM1576407	GSM1865619	GSM1105788
GSM1277974	GSM1157572	GSM1402484	GSM1576408	GSM1865620	GSM1105789
GSM1277975	GSM1157573	GSM1402485	GSM1576409	GSM1865621	GSM1105790

GSM1277976	GSM1157574	GSM1402486	GSM1576410	GSM1865622	GSM1105791
GSM1278007	GSM1157575	GSM1402487	GSM1576411	GSM1865623	GSM1105792
GSM1278330	GSM1157576	GSM1402488	GSM1576412	GSM1865624	GSM1105793
GSM1278331	GSM1157577	GSM1402489	GSM1576413	GSM1865625	GSM1105794
GSM1279702	GSM1157578	GSM1402490	GSM1576414	GSM1865626	GSM1105795
GSM1279703	GSM1157579	GSM1402491	GSM1576415	GSM1865627	GSM1105796
GSM1279746	GSM1157580	GSM1402492	GSM1576416	GSM1865628	GSM1105797
GSM1279747	GSM1157581	GSM1402493	GSM1576417	GSM1865629	GSM1105798
GSM1279748	GSM1157582	GSM1402494	GSM1576418	GSM735419	GSM1105799
GSM1282320	GSM1157584	GSM1402495	GSM1576419	GSM735420	GSM1105800
GSM1282321	GSM1157585	GSM1402496	GSM1576420	GSM735421	GSM1105801
GSM1282322	GSM1157586	GSM1402497	GSM1576421	GSM735422	GSM1105802
GSM1282323	GSM1157587	GSM1402579	GSM1576422	GSM735423	GSM1105803
GSM1282324	GSM1157588	GSM1406028	GSM1576423	GSM1872828	GSM1105804
GSM1282325	GSM1157589	GSM1406029	GSM1576424	GSM1872829	GSM1105805
GSM1282326	GSM1157590	GSM1406030	GSM1576425	GSM1872830	GSM1105806
GSM1282327	GSM1157591	GSM1406031	GSM1576426	GSM1872831	GSM1105807
GSM1282328	GSM1157592	GSM1406032	GSM1576427	GSM1872833	GSM1105808
GSM1282329	GSM1157593	GSM1406318	GSM1576428	GSM1872834	GSM1105809
GSM1282330	GSM1157594	GSM1406320	GSM1576429	GSM1872836	GSM1105810
GSM1378014	GSM1157595	GSM1406321	GSM1576430	GSM1872837	GSM1105811
GSM1378015	GSM1157596	GSM1406322	GSM1576431	GSM1872838	GSM1105812
GSM1378016	GSM1157597	GSM1406323	GSM1576432	GSM1872839	GSM1105813
GSM1378017	GSM1157598	GSM1406324	GSM1576433	GSM1872840	GSM1105814
GSM1378018	GSM1157599	GSM1406325	GSM1576434	GSM1872841	GSM1105815
GSM1378019	GSM1157600	GSM1406326	GSM1576435	GSM1872842	GSM1105816
GSM1378021	GSM1157601	GSM1406327	GSM1576436	GSM1872843	GSM1105817
GSM1378022	GSM1157602	GSM1406328	GSM1576437	GSM1872844	GSM1105818
GSM1378023	GSM1157603	GSM1406329	GSM1576438	GSM1872845	GSM1105819
GSM1378024	GSM1157604	GSM1406330	GSM1576439	GSM1872846	GSM1105820

GSM1378025	GSM1157605	GSM1406331	GSM1576440	GSM1872847	GSM1105821
GSM1378026	GSM1157606	GSM1406332	GSM1576441	GSM1872848	GSM1105822
GSM1282850	GSM1157607	GSM1406333	GSM1576442	GSM1872849	GSM1105823
GSM1289096	GSM1157608	GSM1406334	GSM1576443	GSM1872851	GSM1105824
GSM1289414	GSM1157609	GSM1406335	GSM1576444	GSM1872852	GSM1105825
GSM1289415	GSM1157610	GSM1406337	GSM1576445	GSM1872853	GSM1105826
GSM1290216	GSM1157611	GSM1409687	GSM1576446	GSM1872854	GSM1105827
GSM1290218	GSM1157612	GSM1409688	GSM1577755	GSM1872856	GSM1105828
GSM1290015	GSM1157613	GSM1409689	GSM1577756	GSM1872857	GSM1105829
GSM1290016	GSM1157614	GSM1409690	GSM1577757	GSM1872858	GSM1105830
GSM1290017	GSM1157615	GSM1409691	GSM1577758	GSM1872859	GSM1105831
GSM1290018	GSM1157616	GSM1409692	GSM1577759	GSM1872860	GSM1105832
GSM1293558	GSM1157617	GSM1409693	GSM1577760	GSM1872861	GSM1105833
GSM1293559	GSM1157618	GSM1409694	GSM1577761	GSM1872862	GSM1105834
GSM1293560	GSM1157619	GSM1409695	GSM1577762	GSM1872863	GSM1105835
GSM1293561	GSM1157620	GSM1409696	GSM1577763	GSM1872864	GSM1105836
GSM1293562	GSM1157621	GSM1409697	GSM1577764	GSM1872865	GSM1105837
GSM1293563	GSM1157622	GSM1409698	GSM1577738	GSM1872866	GSM1105838
GSM1293564	GSM1157623	GSM1409699	GSM1577739	GSM1872867	GSM1105839
GSM1293565	GSM1157624	GSM1409700	GSM1577740	GSM1872869	GSM1105840
GSM1293566	GSM1157625	GSM1409701	GSM1577741	GSM1872870	GSM1105841
GSM1293567	GSM1157626	GSM1409702	GSM1577742	GSM1872872	GSM1105842
GSM1293568	GSM1157627	GSM1409703	GSM1577743	GSM1872873	GSM1105843
GSM1293569	GSM1157628	GSM1409704	GSM1577744	GSM1872874	GSM1105844
GSM1293570	GSM1157629	GSM1409705	GSM1414746	GSM1872876	GSM1105845
GSM1293571	GSM1157630	GSM1409706	GSM1414747	GSM1872877	GSM1105846
GSM1293572	GSM1157631	GSM1409707	GSM1414748	GSM1872878	GSM1105847
GSM1293573	GSM1157632	GSM1409708	GSM1414749	GSM1872879	GSM1105848
GSM1293574	GSM1157633	GSM1409709	GSM1414750	GSM1872880	GSM1105849
GSM1293575	GSM1157634	GSM1412698	GSM1414751	GSM1872881	GSM1105850

GSM1293576
GSM1293577
GSM1293578
GSM1293579
GSM1293580
GSM1293581
GSM1293741
GSM1293742
GSM1293743
GSM1293744
GSM1293745
GSM1293746
GSM1293747
GSM1293748
GSM1293749
GSM1293750
GSM1294387
GSM1294388
GSM1294389
GSM1294390
GSM1294391
GSM1294392
GSM1294393
GSM1294394
GSM1294395
GSM1294396
GSM1294397
GSM1294398
GSM1295103
GSM1295104
GSM1157635
GSM1157636
GSM1157637
GSM1157638
GSM1157639
GSM1157641
GSM1157642
GSM1157643
GSM1157644
GSM1157645
GSM1157646
GSM1157647
GSM1157648
GSM1157649
GSM1157650
GSM1157651
GSM1157652
GSM1157653
GSM1157654
GSM1157655
GSM1157656
GSM1157657
GSM1157658
GSM1157659
GSM1157660
GSM1157661
GSM1157662
GSM1157663
GSM1157664
GSM1157665
GSM1412699
GSM1412700
GSM1412701
GSM1412702
GSM1412703
GSM1412704
GSM1412705
GSM1412706
GSM1412707
GSM1412708
GSM1412709
GSM1412710
GSM1412711
GSM1412712
GSM1412713
GSM1412714
GSM1412715
GSM1412716
GSM1412717
GSM1412718
GSM1412719
GSM1412720
GSM1412721
GSM1412722
GSM1412723
GSM1412724
GSM1412725
GSM1412726
GSM1412727
GSM1412728
GSM1581661
GSM1581662
GSM1581663
GSM1581664
GSM1581665
GSM1581666
GSM1585606
GSM1585607
GSM1585608
GSM1585609
GSM1585610
GSM1585611
GSM1585612
GSM1585613
GSM1585614
GSM1585615
GSM1587421
GSM1587422
GSM1587423
GSM1587424
GSM1587425
GSM1587426
GSM1587427
GSM1587428
GSM1588051
GSM1588052
GSM1588053
GSM1588054
GSM1588055
GSM1588056
GSM1872882
GSM1872883
GSM1872885
GSM1872992
GSM1872993
GSM1872994
GSM1872995
GSM1872996
GSM1872997
GSM1872998
GSM1872999
GSM1873000
GSM1874590
GSM1874591
GSM1874592
GSM1876343
GSM1876344
GSM1886913
GSM1886914
GSM1886915
GSM1886916
GSM1886917
GSM1886918
GSM1886923
GSM1886924
GSM1886925
GSM1886926
GSM1886927
GSM1886928
GSM1888331
GSM1105851
GSM1105852
GSM1105853
GSM1105854
GSM1105855
GSM1105856
GSM1105857
GSM1105858
GSM1105859
GSM1105860
GSM1105861
GSM1105862
GSM1105863
GSM1105864
GSM1111646
GSM1111647
GSM1111648
GSM1111649
GSM1111650
GSM1111651
GSM1111652
GSM1111653
GSM1111654
GSM1111655
GSM1111656
GSM1111657
GSM1111658
GSM1111659
GSM1111660
GSM1111661

GSM1295105	GSM1157666	GSM1412729	GSM1598127	GSM1888332	GSM1113312
GSM1296624	GSM1157667	GSM1412730	GSM1598128	GSM1888333	GSM1113313
GSM1296629	GSM1157668	GSM1412731	GSM1598129	GSM1888652	GSM1113314
GSM1297576	GSM1157669	GSM1412732	GSM1598130	GSM1888653	GSM1113315
GSM1297577	GSM1157670	GSM1412733	GSM1598131	GSM1888654	GSM1113316
GSM1297578	GSM1157671	GSM1412734	GSM1598132	GSM1888655	GSM1113317
GSM1297579	GSM1157672	GSM1412735	GSM1598133	GSM1888656	GSM1113318
GSM1297580	GSM1157673	GSM1414929	GSM1598134	GSM1888657	GSM1113319
GSM1297581	GSM1157674	GSM1414930	GSM1599009	GSM1888658	GSM1113320
GSM1297582	GSM1157675	GSM1414931	GSM1599010	GSM1888659	GSM1113322
GSM1297583	GSM1157676	GSM1414932	GSM1599120	GSM1888660	GSM1113323
GSM1297584	GSM1157677	GSM1414933	GSM1599121	GSM1888661	GSM1113324
GSM1297585	GSM1157678	GSM1414934	GSM1599122	GSM1888662	GSM1113325
GSM1297586	GSM1157679	GSM1414935	GSM1599123	GSM1888663	GSM1113326
GSM1297587	GSM1157680	GSM1414936	GSM1599124	GSM1888664	GSM1113327
GSM1297588	GSM1157681	GSM1414937	GSM1599125	GSM1888665	GSM1113328
GSM1297589	GSM1157682	GSM1414938	GSM1599126	GSM1888666	GSM1113329
GSM1297590	GSM1157683	GSM1414939	GSM1599127	GSM1888667	GSM1113330
GSM1297506	GSM1157684	GSM1414940	GSM1599128	GSM1888668	GSM1113331
GSM1297507	GSM1157685	GSM1414941	GSM1602977	GSM1888669	GSM1113332
GSM1297508	GSM1157686	GSM1414942	GSM1602978	GSM1888670	GSM1113333
GSM1298379	GSM1157687	GSM1414943	GSM1602979	GSM1888671	GSM1113334
GSM1298380	GSM1157688	GSM1414944	GSM1602980	GSM1888672	GSM1113335
GSM1298381	GSM1157689	GSM1414945	GSM1602981	GSM1888673	GSM1113336
GSM1302027	GSM1157690	GSM1414946	GSM1602982	GSM1888674	GSM1113337
GSM1302028	GSM1157691	GSM1414947	GSM1602983	GSM1888675	GSM1113338
GSM1302029	GSM1157692	GSM1414948	GSM1602984	GSM1888676	GSM1113339
GSM1302030	GSM1157693	GSM1414949	GSM1602985	GSM1888677	GSM1113340
GSM1302031	GSM1157694	GSM1414950	GSM1602986	GSM1888678	GSM1113341
GSM1302032	GSM1157695	GSM1414951	GSM1602987	GSM1888679	GSM1113342

GSM1304777	GSM1157696	GSM1414952	GSM1602988	GSM1888680	GSM1113343
GSM1304778	GSM1157697	GSM1414953	GSM1602989	GSM1888681	GSM1113344
GSM1304779	GSM1157698	GSM1414954	GSM1602990	GSM1890005	GSM1113345
GSM1304780	GSM1157699	GSM1414955	GSM1602991	GSM1900662	GSM1113346
GSM1304781	GSM1157700	GSM1414956	GSM1602992	GSM1900663	GSM1113347
GSM1304782	GSM1157701	GSM1414957	GSM1602993	GSM1900664	GSM1113348
GSM1304783	GSM1157702	GSM1414958	GSM1602994	GSM1900665	GSM1113349
GSM1304784	GSM1157703	GSM1414959	GSM1602995	GSM1900666	GSM1113350
GSM1304785	GSM1157704	GSM1414960	GSM1602996	GSM1900667	GSM1113351
GSM1304786	GSM1157705	GSM1414961	GSM1602997	GSM1900668	GSM1113352
GSM1304787	GSM1157706	GSM1414962	GSM1602998	GSM1900669	GSM1113353
GSM1304788	GSM1157707	GSM1414963	GSM1602999	GSM1900670	GSM1113354
GSM1304789	GSM1157708	GSM1414964	GSM1603000	GSM1900671	GSM1113355
GSM1304790	GSM1157709	GSM1414965	GSM1603001	GSM1901303	GSM1113356
GSM1304791	GSM1157710	GSM1414966	GSM1603002	GSM1901304	GSM1113357
GSM1306652	GSM1157711	GSM1414967	GSM1603003	GSM1901305	GSM1113358
GSM1306653	GSM1157712	GSM1414968	GSM1603004	GSM1901306	GSM1113359
GSM1306654	GSM1157713	GSM1414969	GSM1603005	GSM1901307	GSM1113360
GSM1306655	GSM1157714	GSM1414970	GSM1603006	GSM1901308	GSM1113361
GSM1306656	GSM1157715	GSM1414971	GSM1603007	GSM1901309	GSM1113362
GSM1306657	GSM1157716	GSM1414972	GSM1603008	GSM1901310	GSM1113363
GSM1306659	GSM1157717	GSM1414973	GSM1603009	GSM1901311	GSM1113364
GSM1306651	GSM1157718	GSM1414974	GSM1603010	GSM1901312	GSM1113365
GSM1093229	GSM1157719	GSM1414975	GSM1603011	GSM1901313	GSM1113366
GSM1093230	GSM1157720	GSM1414976	GSM1603012	GSM1901314	GSM1113367
GSM1093231	GSM1157721	GSM1414977	GSM1603013	GSM1901315	GSM1113368
GSM1093232	GSM1157722	GSM1414979	GSM1603014	GSM1901316	GSM1113369
GSM1093233	GSM1157723	GSM1414980	GSM1603015	GSM1901317	GSM1113371
GSM1093234	GSM1157724	GSM1415126	GSM1603016	GSM1901318	GSM1113372
GSM1093235	GSM1157725	GSM1415127	GSM1603017	GSM1901319	GSM1113373

GSM1093236	GSM1157726	GSM1415128	GSM1603018	GSM1901320	GSM1113374
GSM1093237	GSM1157727	GSM1415129	GSM1603019	GSM1901325	GSM1113375
GSM1093238	GSM1157728	GSM1415130	GSM1603020	GSM1901326	GSM1113376
GSM1312705	GSM1157729	GSM1415131	GSM1603021	GSM1901327	GSM1113377
GSM1312706	GSM1157730	GSM1415132	GSM1603022	GSM1901328	GSM1113378
GSM1312707	GSM1157731	GSM1415133	GSM1603023	GSM1901333	GSM1113379
GSM1312708	GSM1157732	GSM1415134	GSM1603024	GSM1901334	GSM1113380
GSM1312709	GSM1157733	GSM1415135	GSM1603025	GSM1901335	GSM1113381
GSM1312710	GSM1157734	GSM1415136	GSM1603026	GSM1901336	GSM1113382
GSM1312711	GSM1157735	GSM1415137	GSM1603027	GSM1901337	GSM1113383
GSM1312712	GSM1157736	GSM1415138	GSM1603028	GSM1901338	GSM1113384
GSM1312713	GSM1157737	GSM1415139	GSM1603029	GSM1901339	GSM1113385
GSM1312714	GSM1157738	GSM1415140	GSM1603030	GSM1901340	GSM1113386
GSM1312715	GSM1157739	GSM1415141	GSM1603031	GSM1901341	GSM1113387
GSM1312716	GSM1157740	GSM1415142	GSM1603032	GSM1901342	GSM1113388
GSM1312717	GSM1157741	GSM1415143	GSM1603033	GSM1901343	GSM1113389
GSM1312718	GSM1157742	GSM1415144	GSM1603034	GSM1901344	GSM1113390
GSM1312719	GSM1157743	GSM1415145	GSM1603035	GSM1901345	GSM1113391
GSM1312720	GSM1157744	GSM1415146	GSM1603036	GSM1901346	GSM1113392
GSM1312721	GSM1157745	GSM1415147	GSM1603037	GSM1901347	GSM1113393
GSM1312722	GSM1157746	GSM1415148	GSM1603038	GSM1906585	GSM1113394
GSM1312723	GSM1157747	GSM1415149	GSM1603039	GSM1906586	GSM1113395
GSM1312724	GSM1157748	GSM1416801	GSM1603040	GSM1908039	GSM1113396
GSM1312725	GSM1157749	GSM1416804	GSM1603041	GSM1908040	GSM1113397
GSM1312726	GSM1157750	GSM1420579	GSM1603042	GSM1908041	GSM1113398
GSM1312727	GSM1157751	GSM1422445	GSM1603043	GSM1908042	GSM1113399
GSM1312728	GSM1157752	GSM1422446	GSM1603044	GSM1908043	GSM1113400
GSM1312729	GSM1157753	GSM1422447	GSM1603045	GSM1908044	GSM1113401
GSM1312730	GSM1157754	GSM1422448	GSM1603046	GSM1908045	GSM1113402
GSM1312731	GSM1157755	GSM1857097	GSM1603047	GSM1908046	GSM1113403

GSM1312732	GSM1157756	GSM1857098	GSM1603048	GSM1908047	GSM1113404
GSM1312733	GSM1157757	GSM1425760	GSM1604265	GSM1915044	GSM1113405
GSM1312734	GSM1157758	GSM1425761	GSM1604266	GSM1915045	GSM1113406
GSM1312735	GSM1157759	GSM1425762	GSM1604267	GSM1915046	GSM1113407
GSM1312736	GSM1157760	GSM1425763	GSM1608261	GSM1915050	GSM1113408
GSM1312737	GSM1157761	GSM1425764	GSM1608262	GSM1915051	GSM1113409
GSM1312738	GSM1157762	GSM1425765	GSM1608263	GSM1915052	GSM1113410
GSM1312739	GSM1157763	GSM1425766	GSM1608264	GSM1917073	GSM1113411
GSM1312740	GSM1157764	GSM1425767	GSM1608265	GSM1917074	GSM1113412
GSM1312741	GSM1157765	GSM1425771	GSM1608266	GSM1917075	GSM1113413
GSM1312742	GSM1157766	GSM1425772	GSM1608267	GSM1917076	GSM1113415
GSM1312743	GSM1157767	GSM1425773	GSM1608282	GSM1917077	GSM1113416
GSM1312744	GSM1157768	GSM1425774	GSM1608283	GSM1917078	GSM1113417
GSM1312745	GSM1157769	GSM1425775	GSM1608284	GSM1918964	GSM1113418
GSM1312746	GSM1157770	GSM1425776	GSM1609427	GSM1918965	GSM1113419
GSM1312747	GSM1157771	GSM1425777	GSM1609428	GSM1918966	GSM1113420
GSM1312748	GSM1157772	GSM1425778	GSM1609429	GSM1918967	GSM1113421
GSM1312749	GSM1157773	GSM1425779	GSM1609430	GSM1918968	GSM1119582
GSM1313402	GSM1157774	GSM1425780	GSM1609431	GSM1918969	GSM1119581
GSM1313403	GSM1157775	GSM1425781	GSM1609432	GSM1925959	GSM1119583
GSM1314181	GSM1157776	GSM1425782	GSM1609433	GSM1925960	GSM1126516
GSM1314182	GSM1157777	GSM1425783	GSM1609434	GSM1925961	GSM1126517
GSM1314482	GSM1157778	GSM1432452	GSM1609435	GSM1925962	GSM1126518
GSM1314483	GSM1157779	GSM1432453	GSM1609436	GSM1925963	GSM1126519
GSM1314708	GSM1157780	GSM1432454	GSM1609437	GSM1925964	GSM1126520
GSM1314709	GSM1157781	GSM1432455	GSM1609438	GSM1925965	GSM1129239
GSM1314710	GSM1157782	GSM1432456	GSM1609439	GSM1925966	GSM1129240
GSM1314711	GSM1157783	GSM1432457	GSM1609440	GSM1925967	GSM1129241
GSM1314712	GSM1157784	GSM1432458	GSM1609441	GSM1925968	GSM1129242
GSM1314713	GSM1157785	GSM1432459	GSM1609442	GSM1925969	GSM1129243

GSM1314714	GSM1157786	GSM1432460	GSM1609443	GSM1338794	GSM1129244
GSM1314715	GSM1157787	GSM1432461	GSM1609444	GSM1338795	GSM1129245
GSM1314716	GSM1157788	GSM1432462	GSM1612313	GSM1338796	GSM1131155
GSM1314717	GSM1157789	GSM1432463	GSM1612314	GSM1338797	GSM1131156
GSM1314718	GSM1157790	GSM1432464	GSM1612315	GSM1338798	GSM1131186
GSM1314719	GSM1157791	GSM1432465	GSM1612316	GSM1338799	GSM1131187
GSM1315608	GSM1157792	GSM1434984	GSM1612317	GSM1338800	GSM1131188
GSM1315621	GSM1157793	GSM1434985	GSM1612318	GSM1338801	GSM1131189
GSM1315625	GSM1157794	GSM1435495	GSM1614703	GSM1338802	GSM1131190
GSM1315635	GSM1157795	GSM1435496	GSM1614705	GSM1338803	GSM1131191
GSM1315639	GSM1157796	GSM1435497	GSM1614706	GSM1338804	GSM1131192
GSM1315644	GSM1157797	GSM1435498	GSM1614707	GSM1338805	GSM1131193
GSM1315645	GSM1157798	GSM1435499	GSM1618311	GSM1338806	GSM1131194
GSM1315646	GSM1157799	GSM1435500	GSM1618312	GSM1338807	GSM1131195
GSM1315647	GSM1157800	GSM1435501	GSM1618313	GSM1338808	GSM1131196
GSM1315649	GSM1157801	GSM1435502	GSM1618314	GSM1338809	GSM1131197
GSM1315651	GSM1157802	GSM1435504	GSM1618315	GSM1338810	GSM1131743
GSM1315652	GSM1157803	GSM1435506	GSM1618316	GSM1338811	GSM1131744
GSM1315653	GSM1157804	GSM1435507	GSM1618317	GSM1338812	GSM1131745
GSM1315654	GSM1157805	GSM1435508	GSM1618318	GSM1338813	GSM1131746
GSM1315655	GSM1157806	GSM1435509	GSM1618319	GSM1338814	GSM1131747
GSM1315656	GSM1157807	GSM1435510	GSM1618320	GSM1338815	GSM1132418
GSM1315658	GSM1157808	GSM1435511	GSM1618321	GSM1943688	GSM1132419
GSM1315659	GSM1157809	GSM1435512	GSM1618322	GSM1943689	GSM1132420
GSM1315660	GSM1157810	GSM1435513	GSM1619134	GSM1943690	GSM1132421
GSM1315663	GSM1157811	GSM1435813	GSM1619135	GSM1943691	GSM1132422
GSM1315664	GSM1157812	GSM1435814	GSM1619136	GSM1943692	GSM1132423
GSM1315665	GSM1157813	GSM1435815	GSM1619137	GSM1943693	GSM1132424
GSM1315668	GSM1157814	GSM1435816	GSM1619138	GSM1943694	GSM1132425
GSM1315669	GSM1157815	GSM1435817	GSM1619139	GSM1939326	GSM1132426

GSM1315675	GSM1157817	GSM1435818	GSM1619140	GSM1939327	GSM1132427
GSM1315676	GSM1157818	GSM1435819	GSM1619141	GSM1939328	GSM1132428
GSM1315677	GSM1157819	GSM1435820	GSM1619142	GSM1939329	GSM1133247
GSM1315680	GSM1157820	GSM1435821	GSM1619143	GSM1939330	GSM1133248
GSM1315681	GSM1157821	GSM1435822	GSM1619144	GSM1939331	GSM1133249
GSM1315682	GSM1157822	GSM1435823	GSM1619145	GSM1939332	GSM1133250
GSM1315683	GSM1157823	GSM1435824	GSM1619146	GSM1939333	GSM1133251
GSM1315684	GSM1157824	GSM1435825	GSM1619147	GSM1939334	GSM1133660
GSM1315691	GSM1157825	GSM1435826	GSM1619148	GSM1925973	GSM1133661
GSM1315704	GSM1157826	GSM1436135	GSM1619149	GSM1925974	GSM1133662
GSM1315705	GSM1157827	GSM1436136	GSM1619150	GSM1925975	GSM1133663
GSM1315707	GSM1157828	GSM1436137	GSM1619151	GSM1925976	GSM1133664
GSM1315708	GSM1157829	GSM1436138	GSM1619152	GSM1925977	GSM1133665
GSM1315709	GSM1157830	GSM1436351	GSM1619153	GSM1925978	GSM1133666
GSM1315710	GSM1157831	GSM1436352	GSM1619154	GSM1925979	GSM1133667
GSM1315711	GSM1157832	GSM1436353	GSM1619155	GSM1955072	GSM1133668
GSM1315712	GSM1157833	GSM1436354	GSM1619156	GSM1955073	GSM1133669
GSM1315713	GSM1157834	GSM1438894	GSM1619157	GSM1955074	GSM1133670
GSM1315714	GSM1157835	GSM1438895	GSM1619158	GSM1955075	GSM1133671
GSM1315715	GSM1157836	GSM1438896	GSM1619159	GSM1955076	GSM1133672
GSM1315716	GSM1157837	GSM1438897	GSM1619160	GSM1955077	GSM1133673
GSM1315717	GSM1157838	GSM1440487	GSM1619161	GSM1955078	GSM1133674
GSM1315718	GSM1157839	GSM1440488	GSM1619162	GSM1955079	GSM1133675
GSM1315719	GSM1157840	GSM1440489	GSM1619163	GSM1955080	GSM1133676
GSM1315720	GSM1157841	GSM1440490	GSM1619164	GSM1955081	GSM1133677
GSM1315721	GSM1157842	GSM1440491	GSM1619165	GSM1955082	GSM1133678
GSM1315722	GSM1157843	GSM1440492	GSM1619166	GSM1955083	GSM1133679
GSM1315723	GSM1157844	GSM1440493	GSM1619167	GSM1955084	GSM1133680
GSM1315724	GSM1157845	GSM1440494	GSM1619168	GSM1955085	GSM1133681
GSM1315725	GSM1157846	GSM1440495	GSM1619169	GSM1955086	GSM1133682

GSM1315726	GSM1157848	GSM1440496	GSM1619170	GSM1955087	GSM1133683
GSM1315727	GSM1157849	GSM1440497	GSM1619171	GSM1955088	GSM1133684
GSM1315728	GSM1157850	GSM1440498	GSM1619172	GSM1955089	GSM1142684
GSM1315729	GSM1157851	GSM1440499	GSM1619173	GSM1955090	GSM1142685
GSM1315730	GSM1157852	GSM1440500	GSM1619174	GSM1955091	GSM1142686
GSM1315731	GSM1157853	GSM1440501	GSM1619175	GSM1960355	GSM1142687
GSM1315732	GSM1157854	GSM1440502	GSM1619176	GSM1960356	GSM1153501
GSM1315733	GSM1157855	GSM1440503	GSM1619177	GSM1960357	GSM1153507
GSM1315734	GSM1157856	GSM1440610	GSM1619178	GSM1960706	GSM1153509
GSM1315735	GSM1157857	GSM1440611	GSM1619179	GSM1960707	GSM1153510
GSM1315736	GSM1157858	GSM1443819	GSM1619180	GSM742937	GSM1153512
GSM1315738	GSM1157859	GSM1443820	GSM1619181	GSM742938	GSM1153513
GSM1315739	GSM1157860	GSM1443821	GSM1619182	GSM742939	GSM1153528
GSM1315745	GSM1157925	GSM1443822	GSM1619183	GSM742940	GSM1153529
GSM1315751	GSM1157926	GSM1443823	GSM1619184	GSM742941	GSM1228810
GSM1315767	GSM1157927	GSM1443824	GSM1619185	GSM742942	GSM1228811
GSM1315768	GSM1157928	GSM1443825	GSM1619186	GSM742943	GSM1153916
GSM1315772	GSM1157929	GSM1443826	GSM1619187	GSM742944	GSM1153917
GSM1315774	GSM1157930	GSM1443827	GSM1619188	GSM742945	GSM1155149
GSM1315779	GSM1157931	GSM1443829	GSM1619189	GSM742946	GSM1155150
GSM1315780	GSM1157932	GSM1444166	GSM1619190	GSM742947	GSM1155151
GSM1315781	GSM1157933	GSM1444171	GSM1619191	GSM742948	GSM1155152
GSM1315782	GSM1157934	GSM1444180	GSM1619192	GSM742949	GSM1155153
GSM1315783	GSM1157935	GSM1444185	GSM1619193	GSM742950	GSM1155154
GSM1315784	GSM1157936	GSM1446338	GSM1619194	GSM742952	GSM1155155
GSM1315785	GSM1157937	GSM1446339	GSM1619195	GSM749465	GSM1155156
GSM1315786	GSM1157938	GSM1446340	GSM1619196	GSM749466	GSM1155157
GSM1317868	GSM1157939	GSM1446341	GSM1619197	GSM749467	GSM1155158
GSM1317869	GSM1157940	GSM1446342	GSM1619198	GSM749468	GSM1155159
GSM1317870	GSM1157941	GSM1446343	GSM1619199	GSM747470	GSM1155160

GSM1317871	GSM1157942	GSM1446344	GSM1619200	GSM747471	GSM1155161
GSM1317872	GSM1157943	GSM1446345	GSM1619201	GSM747472	GSM1155162
GSM1317873	GSM1157944	GSM1446880	GSM1619202	GSM747473	GSM1155163
GSM1317874	GSM1157945	GSM1446881	GSM1619203	GSM747474	GSM1155164
GSM1317875	GSM1157946	GSM1446882	GSM1619204	GSM747475	GSM1155165
GSM1317876	GSM1157947	GSM1446883	GSM1619205	GSM747476	GSM1155166
GSM1317877	GSM1157948	GSM1446884	GSM1619206	GSM747477	GSM1155167
GSM1319846	GSM1157949	GSM1446885	GSM1619207	GSM747478	GSM1155168
GSM1319847	GSM1157950	GSM1446886	GSM1619208	GSM747479	GSM1155370
GSM1319848	GSM1157951	GSM1446887	GSM1619209	GSM747480	GSM1155371
GSM1319849	GSM1157952	GSM1447395	GSM1619210	GSM1973958	GSM1155372
GSM1319850	GSM1157953	GSM1447396	GSM1619211	GSM1973959	GSM1155373
GSM1319851	GSM1157954	GSM1447397	GSM1619212	GSM1973960	GSM1155374
GSM1319852	GSM1157955	GSM1447398	GSM1619213	GSM1973961	GSM1155375
GSM1319853	GSM1157956	GSM1447399	GSM1619214	GSM1973962	GSM1155376
GSM1323528	GSM1157957	GSM1447400	GSM1619215	GSM1973963	GSM1155377
GSM1323529	GSM1157958	GSM1447401	GSM1619216	GSM1974764	GSM1155378
GSM1323530	GSM1157959	GSM1447402	GSM1619217	GSM1974765	GSM1155379
GSM1323531	GSM1157960	GSM1447403	GSM1619218	GSM1974766	GSM1155380
GSM1325496	GSM1157961	GSM1447404	GSM1619219	GSM1977027	GSM1155381
GSM1325497	GSM1157962	GSM1447405	GSM1619220	GSM1977028	GSM1155382
GSM1326407	GSM1157963	GSM1447406	GSM1619221	GSM1977029	GSM1155383
GSM1326408	GSM1157964	GSM1462858	GSM1619222	GSM1977030	GSM1155384
GSM1326409	GSM1157965	GSM1462859	GSM1619223	GSM1977031	GSM1155385
GSM1326410	GSM1157966	GSM1462860	GSM1619224	GSM1977032	GSM1155386
GSM1326411	GSM1157967	GSM1462861	GSM1619225	GSM1977033	GSM1162717
GSM1326412	GSM1157968	GSM1462862	GSM1619226	GSM1977034	GSM1162718
GSM1326569	GSM1157969	GSM1462863	GSM1619227	GSM1977035	GSM1162719
GSM1326570	GSM1157970	GSM1464095	GSM1619228	GSM1977036	GSM1162720
GSM1326571	GSM1157971	GSM1464101	GSM1619229	GSM1977037	GSM1162721

GSM1326572	GSM1157972	GSM1466233	GSM1619230	GSM1977038	GSM1162722
GSM1326573	GSM1157973	GSM1466234	GSM1619231	GSM1977039	GSM1162723
GSM1326574	GSM1157974	GSM1466235	GSM1619232	GSM1977040	GSM1162724
GSM1326575	GSM1157975	GSM1466236	GSM1619233	GSM1977041	GSM1162725
GSM1326576	GSM1157976	GSM1466237	GSM1619234	GSM1977042	GSM1162726
GSM1326577	GSM1157977	GSM1466238	GSM1619235	GSM1977043	GSM1162727
GSM1326578	GSM1157978	GSM1466239	GSM1619236	GSM1977044	GSM1162728
GSM1326579	GSM1157979	GSM1466240	GSM1619237	GSM1977045	GSM1162729
GSM1326580	GSM1157980	GSM1466241	GSM1619238	GSM1977046	GSM1162730
GSM1327170	GSM1157981	GSM1466242	GSM1619239	GSM1977047	GSM1162731
GSM1327171	GSM1157982	GSM1574593	GSM1619240	GSM1977399	GSM1162732
GSM1327339	GSM1157983	GSM1574594	GSM1619241	GSM1977400	GSM1163070
GSM1327340	GSM1157984	GSM1574595	GSM1619242	GSM1977401	GSM1163071
GSM1327341	GSM1157985	GSM1574596	GSM1619243	GSM1977402	GSM1163072
GSM1327342	GSM1157986	GSM1466905	GSM1619244	GSM1977403	GSM1166072
GSM1327343	GSM1157987	GSM1466906	GSM1623140	GSM1977404	GSM1166073
GSM1327344	GSM1157988	GSM1466907	GSM1623141	GSM1977406	GSM1166074
GSM1327874	GSM1157989	GSM1479433	GSM1623142	GSM1977407	GSM1166084
GSM1327875	GSM1157990	GSM1479438	GSM1623143	GSM1977410	GSM1166085
GSM1327876	GSM1157991	GSM1479439	GSM1623144	GSM1977411	GSM1166086
GSM1327877	GSM1157992	GSM1479440	GSM1623145	GSM1977412	GSM1166090
GSM1322274	GSM1157993	GSM1479441	GSM1623146	GSM1977413	GSM1166091
GSM1328790	GSM1157994	GSM1479442	GSM1623147	GSM1977414	GSM1166092
GSM1328792	GSM1157995	GSM1479499	GSM1623148	GSM1977415	GSM1166097
GSM1328794	GSM1157996	GSM1479500	GSM1623149	GSM1977416	GSM1166098
GSM1328796	GSM1157997	GSM1479501	GSM1625957	GSM1977417	GSM1166099
GSM1332750	GSM1157998	GSM1479502	GSM1625958	GSM1977418	GSM1166100
GSM1332751	GSM1157999	GSM1479503	GSM1625959	GSM1977420	GSM1166105
GSM1333067	GSM1158000	GSM1479505	GSM1625960	GSM1977421	GSM1166106
GSM1333068	GSM1158001	GSM1479506	GSM1625961	GSM1977422	GSM1166107

GSM1333069	GSM1158002	GSM1479507	GSM1625962	GSM1978251	GSM1166108
GSM1333110	GSM1158003	GSM1479508	GSM1625963	GSM1978252	GSM1166113
GSM1333111	GSM1158004	GSM1479509	GSM1625964	GSM1978253	GSM1166114
GSM1333112	GSM1158005	GSM1479510	GSM1625965	GSM1978254	GSM1166115
GSM1333113	GSM1158006	GSM1479512	GSM1625966	GSM1978255	GSM1166116
GSM1333378	GSM1158007	GSM1479520	GSM1626439	GSM1978256	GSM1166121
GSM1333379	GSM1158008	GSM1479521	GSM1626440	GSM1978257	GSM1166122
GSM1333380	GSM1158009	GSM1479522	GSM1626441	GSM752696	GSM1166123
GSM1333381	GSM1158010	GSM1479523	GSM1626442	GSM752697	GSM1166124
GSM1333382	GSM1158011	GSM1479524	GSM1626443	GSM752698	GSM1166128
GSM1333383	GSM1158012	GSM1479526	GSM1626444	GSM752702	GSM1166129
GSM1333384	GSM1158013	GSM1481718	GSM1626445	GSM752703	GSM1166130
GSM1333385	GSM1158014	GSM1482932	GSM1626446	GSM752704	GSM1173802
GSM1333386	GSM1158015	GSM1482933	GSM1626447	GSM752705	GSM1173803
GSM1333387	GSM1158016	GSM1482934	GSM1626448	GSM752706	GSM1173804
GSM1333388	GSM1158017	GSM1482935	GSM1626449	GSM752707	GSM1173805
GSM1333389	GSM1158018	GSM1482936	GSM1626450	GSM752708	GSM1173806
GSM1333390	GSM1158019	GSM1482937	GSM1626451	GSM754335	GSM1173807
GSM1333391	GSM1158020	GSM1482938	GSM1626452	GSM2027504	GSM1173808
GSM1333392	GSM1158021	GSM1482939	GSM1626453	GSM2027505	GSM1174472
GSM1333393	GSM1158022	GSM1482940	GSM1626454	GSM2027506	GSM1184591
GSM1333394	GSM1158023	GSM1482941	GSM1626455	GSM2027507	GSM1184593
GSM1333395	GSM1158024	GSM1482942	GSM1626456	GSM2027508	GSM1184595
GSM1333396	GSM1158025	GSM1482943	GSM1626457	GSM2027509	GSM1184597
GSM1333397	GSM1158026	GSM1482944	GSM1626458	GSM2027510	GSM1184599
GSM1333398	GSM1158027	GSM1482945	GSM1626459	GSM2027511	GSM1184601
GSM1333399	GSM1158028	GSM1482946	GSM1626460	GSM2027512	GSM1185603
GSM1333400	GSM1158029	GSM1482947	GSM1626461	GSM2027513	GSM1185604
GSM1333401	GSM1158030	GSM1482948	GSM1626462	GSM2027514	GSM1185605
GSM1333402	GSM1158031	GSM1482949	GSM1626463	GSM2027515	GSM1185606

GSM1333403	GSM1158032	GSM1482950	GSM1626464	GSM2027516	GSM1185607
GSM1333404	GSM1158033	GSM1482951	GSM1626465	GSM2027517	GSM1185608
GSM1333405	GSM1158034	GSM1482952	GSM1626466	GSM2027518	GSM1185609
GSM1333406	GSM1158035	GSM1482953	GSM1626467	GSM2027519	GSM1185610
GSM1333407	GSM1158036	GSM1482954	GSM1626468	GSM2027520	GSM1185611
GSM1333408	GSM1158037	GSM1482955	GSM1626469	GSM2027521	GSM1185612
GSM1334287	GSM1158038	GSM1482956	GSM1626470	GSM2027522	GSM1185613
GSM1334288	GSM1158039	GSM1482957	GSM1626471	GSM2027523	GSM1185614
GSM1334289	GSM1158040	GSM1482958	GSM1626472	GSM2027524	GSM1185615
GSM1334293	GSM1158041	GSM1482959	GSM1626473	GSM2027525	GSM1185616
GSM1334294	GSM1158042	GSM1482960	GSM1626474	GSM2027526	GSM1185617
GSM1334295	GSM1158043	GSM1482961	GSM1626475	GSM2027527	GSM1185618
GSM1334330	GSM1158044	GSM1482962	GSM1626476	GSM2027528	GSM1185619
GSM1334331	GSM1158045	GSM1482963	GSM1626477	GSM2027529	GSM1187136
GSM1308994	GSM1158046	GSM1482964	GSM1626478	GSM2027530	GSM1187137
GSM1338133	GSM1158047	GSM1489558	GSM1626479	GSM2027531	GSM1187142
GSM1338134	GSM1158048	GSM1489559	GSM1626480	GSM2027532	GSM1193393
GSM1338135	GSM1158049	GSM1489560	GSM1626481	GSM2027533	GSM1193394
GSM1338136	GSM1158050	GSM1489561	GSM1626482	GSM2028114	GSM1193395
GSM1338137	GSM1158051	GSM1489562	GSM1626483	GSM2028115	GSM1193396
GSM1338138	GSM1158052	GSM1489563	GSM1626484	GSM2028120	GSM1193397
GSM1338139	GSM1158053	GSM1489564	GSM1626485	GSM2028121	GSM1194676
GSM1338140	GSM1158054	GSM1489565	GSM1626486	GSM2028122	GSM1194677
GSM1338141	GSM1158055	GSM1489566	GSM1626487	GSM2028123	GSM1194678
GSM1338142	GSM1158056	GSM1489567	GSM1626488	GSM2029382	GSM1194682
GSM1338759	GSM1158057	GSM1489568	GSM1626489	GSM2029383	GSM1194687
GSM1338760	GSM1158058	GSM1489569	GSM1626490	GSM2029384	GSM1194693
GSM1338764	GSM1158059	GSM1489570	GSM1626491	GSM2029385	GSM1194694
GSM1338768	GSM1158060	GSM1489571	GSM1626492	GSM2029386	GSM1194804
GSM1345809	GSM1158061	GSM1489572	GSM1626493	GSM2029387	GSM1194805

GSM1345810	GSM1158062	GSM1489573	GSM1626494	GSM2029388	GSM1194806
GSM1345811	GSM1158063	GSM1489574	GSM1626495	GSM761684	GSM1194807
GSM1345812	GSM1158064	GSM1489575	GSM1626496	GSM761685	GSM1194808
GSM1345813	GSM1158065	GSM1489576	GSM1626497	GSM758634	GSM1194809
GSM1345814	GSM1158066	GSM1489577	GSM1626498	GSM758635	GSM1194810
GSM1345815	GSM1158067	GSM1489578	GSM1626499	GSM758636	GSM1194811
GSM1345816	GSM1158068	GSM1489579	GSM1626500	GSM759885	GSM1194812
GSM1345817	GSM1158069	GSM1489580	GSM1626501	GSM759886	GSM1194813
GSM1345818	GSM1158070	GSM1489581	GSM1626502	GSM759887	GSM1194814
GSM1345819	GSM1158071	GSM1489582	GSM1626503	GSM759889	GSM1196950
GSM1345820	GSM1158072	GSM1489583	GSM1626504	GSM759890	GSM1196951
GSM1345821	GSM1158073	GSM1489584	GSM1626505	GSM759891	GSM1196952
GSM1345822	GSM1158074	GSM1489585	GSM1626506	GSM759892	GSM1196953
GSM1345823	GSM1158075	GSM1489586	GSM1626507	GSM759893	GSM1196954
GSM1345824	GSM1158076	GSM1489587	GSM1626508	GSM2046873	GSM1196955
GSM1345826	GSM1158077	GSM1489588	GSM1626509	GSM2046874	GSM1196956
GSM1348980	GSM1158078	GSM1489589	GSM1626510	GSM2046875	GSM1196957
GSM1348981	GSM1158079	GSM1489590	GSM1626511	GSM2046876	GSM1196958
GSM1348982	GSM1158080	GSM1489591	GSM1626512	GSM2046877	GSM1196959
GSM1348983	GSM1158081	GSM1489592	GSM1626513	GSM2046878	GSM1196575
GSM1354448	GSM1158082	GSM1489593	GSM1626514	GSM764210	GSM1196578
GSM1354449	GSM1158083	GSM1489594	GSM1626515	GSM764211	GSM1196584
GSM1354450	GSM1158084	GSM1489595	GSM1626516	GSM764212	GSM1202460
GSM1354451	GSM1158085	GSM1489596	GSM1626517	GSM793363	GSM1202461
GSM1354452	GSM1158086	GSM1489597	GSM1626518	GSM793364	GSM1202462
GSM1354453	GSM1158087	GSM1489598	GSM1631719	GSM793365	GSM1202463
GSM1354454	GSM1158088	GSM1489599	GSM1631720	GSM793366	GSM1202464
GSM1354455	GSM1158089	GSM1489600	GSM1631721	GSM793367	GSM1202465
GSM1354456	GSM1158090	GSM1489601	GSM1631881	GSM793368	GSM1202466
GSM1354457	GSM1158091	GSM1489602	GSM1631882	GSM793369	GSM1202467

GSM1354458	GSM1158092	GSM1489603	GSM1631883	GSM793370	GSM1202468
GSM1354459	GSM1158093	GSM1489604	GSM1631884	GSM793371	GSM1202469
GSM1354460	GSM1158094	GSM1489605	GSM1631885	GSM793372	GSM1202470
GSM1354461	GSM1158095	GSM1489606	GSM1631490	GSM793373	GSM1202471
GSM1354462	GSM1158096	GSM1489607	GSM1631491	GSM793374	GSM1202569
GSM1354463	GSM1158097	GSM1489608	GSM1631492	GSM793375	GSM1202570
GSM1354464	GSM1158098	GSM1489609	GSM1631493	GSM793376	GSM1202571
GSM1354465	GSM1158099	GSM1489610	GSM1631494	GSM1115019	GSM1202572
GSM1354466	GSM1158100	GSM1489611	GSM1631495	GSM1115020	GSM1202573
GSM1354841	GSM1158101	GSM1489612	GSM1631496	GSM1115021	GSM1202574
GSM1354843	GSM1158102	GSM1489613	GSM1631497	GSM1115022	GSM1202575
GSM1354845	GSM1158103	GSM1489614	GSM1631498	GSM1115023	GSM1202576
GSM1354847	GSM1158104	GSM1489615	GSM1631499	GSM1115024	GSM1202577
GSM1354849	GSM1158105	GSM1489616	GSM714814	GSM767949	GSM1202578
GSM1354852	GSM1158106	GSM1489617	GSM1633701	GSM767950	GSM1202579
GSM1354853	GSM1158107	GSM1489618	GSM1633702	GSM767951	GSM1202580
GSM1354855	GSM1158108	GSM1489619	GSM1641319	GSM800443	GSM1202581
GSM1354857	GSM1158109	GSM1489620	GSM1641320	GSM800445	GSM1202582
GSM1357994	GSM1158110	GSM1489621	GSM1641321	GSM799164	GSM1202583
GSM1357995	GSM1158111	GSM1489622	GSM1641322	GSM799165	GSM1202584
GSM1357996	GSM1158112	GSM1489623	GSM1641323	GSM799166	GSM1203305
GSM1357997	GSM1158113	GSM1489624	GSM1641324	GSM799167	GSM1203306
GSM1357998	GSM1158114	GSM1489625	GSM1641325	GSM804340	GSM1203307
GSM1358004	GSM1158115	GSM1492937	GSM1641326	GSM804341	GSM1203308
GSM1359512	GSM1158116	GSM1492939	GSM1641327	GSM804342	GSM1203309
GSM1359514	GSM1158117	GSM1492941	GSM1641328	GSM804343	GSM1203310
GSM1361091	GSM1158118	GSM1495400	GSM1641329	GSM804345	GSM1203311
GSM1361093	GSM1158119	GSM1495401	GSM1641330	GSM808734	GSM1203312
GSM1361095	GSM1158120	GSM1495402	GSM1641331	GSM808735	GSM1203313
GSM1361097	GSM1158121	GSM1495403	GSM1641332	GSM811624	GSM1203314

GSM1361099	GSM1158122	GSM1495404	GSM1641333	GSM811625	GSM1203315
GSM1361101	GSM1158123	GSM1495405	GSM1641334	GSM811626	GSM1203316
GSM1361974	GSM1158124	GSM1495406	GSM1641335	GSM811627	GSM1203317
GSM1361975	GSM1158125	GSM1495414	GSM1641336	GSM811628	GSM1203318
GSM1361976	GSM1158126	GSM1495415	GSM1641337	GSM811629	GSM1203319
GSM1361977	GSM1158127	GSM1308998	GSM1641338	GSM811630	GSM1203320
GSM1361978	GSM1158128	GSM1498119	GSM1641339	GSM811631	GSM1203321
GSM1361979	GSM1158129	GSM1498120	GSM1641340	GSM819489	GSM1203322
GSM1361980	GSM1158130	GSM1498121	GSM1641341	GSM819490	GSM1203323
GSM1361981	GSM1158131	GSM1498122	GSM1641342	GSM821030	GSM1203324
GSM1361982	GSM1158132	GSM1498123	GSM1645000	GSM821031	GSM1203325
GSM1361983	GSM1158133	GSM1498124	GSM1645001	GSM821032	GSM1203326
GSM1361984	GSM1158134	GSM1498125	GSM1645002	GSM821033	GSM1203327
GSM1361985	GSM1158135	GSM1498126	GSM1645003	GSM821034	GSM1203328
GSM1361986	GSM1158136	GSM1498127	GSM1647922	GSM821035	GSM1203329
GSM1361987	GSM1158137	GSM1498128	GSM1647923	GSM821036	GSM1203330
GSM1361988	GSM1158138	GSM1498129	GSM1647924	GSM821037	GSM1203331
GSM1361989	GSM1158139	GSM1498130	GSM1647925	GSM821038	GSM1203332
GSM1361990	GSM1158140	GSM1499784	GSM1647926	GSM821039	GSM1203333
GSM1361991	GSM1158141	GSM1499785	GSM1647927	GSM821040	GSM1203334
GSM1361992	GSM1158142	GSM1499786	GSM1647928	GSM821041	GSM1203335
GSM1361993	GSM1158143	GSM1501174	GSM1647929	GSM823383	GSM1203336
GSM1361994	GSM1158144	GSM1503677	GSM1647930	GSM830389	GSM1203337
GSM1361995	GSM1158145	GSM1503678	GSM1647931	GSM830390	GSM1203338
GSM1361996	GSM1158146	GSM1503679	GSM1647932	GSM830391	GSM1203339
GSM1361997	GSM1158147	GSM1503680	GSM1647933	GSM830392	GSM1203340
GSM1361998	GSM1158148	GSM1503681	GSM1647934	GSM830393	GSM1203341
GSM1361999	GSM1158149	GSM1503682	GSM1647935	GSM830394	GSM1203342
GSM1362000	GSM1158150	GSM1503683	GSM1647936	GSM830395	GSM1203343
GSM1362001	GSM1158151	GSM1503684	GSM1647937	GSM830396	GSM1203344

GSM1362002	GSM1158152	GSM1503685	GSM1647938	GSM830397	GSM1203345
GSM1362003	GSM1158153	GSM1503686	GSM1647939	GSM830398	GSM1203346
GSM1362004	GSM1158154	GSM1503687	GSM1647940	GSM830399	GSM1203347
GSM1362005	GSM1158156	GSM1503688	GSM1647941	GSM830400	GSM1204876
GSM1362006	GSM1158157	GSM1503689	GSM1647942	GSM830401	GSM1204877
GSM1362007	GSM1158158	GSM1503690	GSM1647943	GSM830402	GSM1204878
GSM1362008	GSM1158159	GSM1503691	GSM1647944	GSM830403	GSM1204879
GSM1362009	GSM1158160	GSM1503692	GSM1647945	GSM830404	GSM1204880
GSM1362010	GSM1158162	GSM1503693	GSM1647946	GSM830405	GSM1204881
GSM1362011	GSM1158163	GSM1503694	GSM1647947	GSM830448	GSM1206234
GSM1362012	GSM1158164	GSM1503695	GSM1647948	GSM830449	GSM1206235
GSM1362013	GSM1158165	GSM1503696	GSM1647949	GSM830450	GSM1206236
GSM1362014	GSM1158166	GSM1503697	GSM1647952	GSM830451	GSM1206237
GSM1362015	GSM1158167	GSM1503698	GSM1647954	GSM830452	GSM1206238
GSM1362016	GSM1158168	GSM1503699	GSM1647956	GSM830453	GSM1206239
GSM1362017	GSM1158169	GSM1608005	GSM1647959	GSM830454	GSM1206240
GSM1362018	GSM1158170	GSM1608006	GSM1647962	GSM830455	GSM1206242
GSM1362019	GSM1158171	GSM1608007	GSM1647964	GSM830456	GSM1206243
GSM1362020	GSM1158172	GSM1608008	GSM1647966	GSM830457	GSM1207643
GSM1362021	GSM1158173	GSM1608009	GSM1649191	GSM835231	GSM1207644
GSM1362022	GSM1158174	GSM1608010	GSM1649192	GSM835232	GSM1207645
GSM1362023	GSM1158175	GSM1608011	GSM1649193	GSM835233	GSM1207646
GSM1362024	GSM1158176	GSM1608012	GSM1649194	GSM838064	GSM1207647
GSM1362025	GSM1158177	GSM1608013	GSM1649195	GSM838066	GSM1207648
GSM1362026	GSM1158178	GSM1608014	GSM1649196	GSM838068	GSM1207649
GSM1362027	GSM1158179	GSM1608015	GSM1649197	GSM838070	GSM1207650
GSM1362028	GSM1158180	GSM1608016	GSM1649198	GSM838072	GSM1207651
GSM1362029	GSM1158245	GSM1608017	GSM1649199	GSM838074	GSM1207652
GSM1362030	GSM1158246	GSM1608018	GSM1649200	GSM838076	GSM1207653
GSM1362031	GSM1158247	GSM1608019	GSM1649201	GSM838078	GSM1207654

GSM1362032	GSM1158248	GSM1608020	GSM1649202	GSM838080	GSM1207659
GSM1362033	GSM1158249	GSM1608063	GSM1649203	GSM838082	GSM1207660
GSM1362034	GSM1158250	GSM1608064	GSM1649204	GSM838084	GSM1207661
GSM1362035	GSM1158251	GSM1608065	GSM1649205	GSM838086	GSM1207662
GSM1362036	GSM1158252	GSM1608066	GSM1649206	GSM838088	GSM1208968
GSM1362037	GSM1158253	GSM1504073	GSM1649207	GSM838090	GSM1208969
GSM1362038	GSM1158254	GSM1504074	GSM1649208	GSM838092	GSM1208970
GSM1362039	GSM1158255	GSM1504075	GSM1649209	GSM838094	GSM1208971
GSM1362040	GSM1158256	GSM1504076	GSM1649210	GSM838096	GSM1208972
GSM1362041	GSM1158257	GSM1505565	GSM1649211	GSM838098	GSM1215102
GSM1362042	GSM1158258	GSM1505566	GSM1649212	GSM838100	GSM1215103
GSM1362043	GSM1158259	GSM1505567	GSM1649213	GSM838102	GSM1215104
GSM1362044	GSM1158260	GSM1505568	GSM1649214	GSM838104	GSM1215105
GSM1362045	GSM1158261	GSM1505569	GSM1657075	GSM838106	GSM1215106
GSM1362046	GSM1158262	GSM1505570	GSM1657076	GSM838108	GSM1215136
GSM1362047	GSM1158263	GSM1505571	GSM1657077	GSM838109	GSM1215137
GSM1362048	GSM1158264	GSM1505572	GSM1658371	GSM838112	GSM1216753
GSM1362049	GSM1158265	GSM1505573	GSM1658372	GSM838114	GSM1216754
GSM1362050	GSM1158266	GSM1505574	GSM1658373	GSM838116	GSM1216755
GSM1362051	GSM1158267	GSM1505575	GSM1658374	GSM838118	GSM1216756
GSM1362052	GSM1158268	GSM1505576	GSM1658375	GSM838120	GSM1216757
GSM1362053	GSM1158269	GSM1505577	GSM1658376	GSM838122	GSM1216758
GSM1362054	GSM1158270	GSM1505578	GSM1658378	GSM838124	GSM1216759
GSM1362055	GSM1158271	GSM1505579	GSM1658379	GSM839747	GSM1216760
GSM1362056	GSM1158272	GSM1505580	GSM1658380	GSM841726	GSM1216761
GSM1362057	GSM1158273	GSM1505581	GSM1658381	GSM841727	GSM1216762
GSM1362058	GSM1158274	GSM1505582	GSM1658382	GSM841728	GSM1216763
GSM1362059	GSM1158275	GSM1505583	GSM1658383	GSM841729	GSM1216764
GSM1362060	GSM1158276	GSM1505584	GSM1658384	GSM856868	GSM1216765
GSM1362061	GSM1158277	GSM1505585	GSM1658385	GSM856869	GSM1216766

GSM1362062	GSM1158278	GSM1505586	GSM1658386	GSM856870	GSM1216767
GSM1362063	GSM1158279	GSM1505587	GSM1658387	GSM856871	GSM1216768
GSM1362064	GSM1158280	GSM1505588	GSM1658388	GSM856880	GSM1216769
GSM1362065	GSM1158281	GSM1505589	GSM1658389	GSM856881	GSM1216770
GSM1362066	GSM1158282	GSM1505594	GSM1658390	GSM864320	GSM1216771
GSM1362067	GSM1158283	GSM1505595	GSM1658391	GSM865291	GSM1216772
GSM1362068	GSM1158284	GSM1505596	GSM1658392	GSM865292	GSM1216773
GSM1362069	GSM1158285	GSM1505597	GSM1658393	GSM865293	GSM1216774
GSM1362070	GSM1158286	GSM1505598	GSM1658394	GSM865294	GSM1216775
GSM1362071	GSM1158287	GSM1505599	GSM1658395	GSM865295	GSM1216776
GSM1362072	GSM1158288	GSM1505600	GSM1658396	GSM865296	GSM1216777
GSM1362073	GSM1158289	GSM1505601	GSM1658397	GSM865297	GSM1216778
GSM1362074	GSM1158290	GSM1505602	GSM1658398	GSM865298	GSM1216779
GSM1362075	GSM1158291	GSM1505603	GSM1658399	GSM865299	GSM1216780
GSM1362076	GSM1158292	GSM1505604	GSM1658400	GSM865300	GSM1216781
GSM1362077	GSM1158293	GSM1505605	GSM1659004	GSM869033	GSM1216782
GSM1362078	GSM1158294	GSM1505606	GSM1659005	GSM869034	GSM1216783
GSM1362079	GSM1158295	GSM1505607	GSM1659006	GSM869035	GSM1216784
GSM1362080	GSM1158296	GSM1505608	GSM1659010	GSM883916	GSM1216785
GSM1362081	GSM1158297	GSM1505609	GSM1659543	GSM883917	GSM1216786
GSM1362082	GSM1158298	GSM1505610	GSM1659544	GSM883918	GSM1216787
GSM1362083	GSM1158299	GSM1505611	GSM1659545	GSM883919	GSM1216788
GSM1362084	GSM1158300	GSM1505612	GSM1659549	GSM898966	GSM1216789
GSM1362085	GSM1158301	GSM1505613	GSM1659550	GSM898967	GSM1216790
GSM1362086	GSM1158302	GSM1505614	GSM1659551	GSM898968	GSM1216791
GSM1362087	GSM1158303	GSM1641262	GSM1659552	GSM898969	GSM1216792
GSM1362088	GSM1158304	GSM1641263	GSM1659553	GSM898970	GSM1216793
GSM1362089	GSM1158305	GSM1641264	GSM1659554	GSM898971	GSM1216794
GSM1362090	GSM1158306	GSM1641271	GSM1665183	GSM898972	GSM1216795
GSM1362091	GSM1158307	GSM1641274	GSM1665184	GSM898973	GSM1216796

GSM1362092	GSM1158308	GSM1641275	GSM1665185	GSM1099813	GSM1216797
GSM1362093	GSM1158309	GSM1641276	GSM1665186	GSM1099814	GSM1216798
GSM1362094	GSM1158310	GSM1641277	GSM1665187	GSM1099815	GSM1216799
GSM1362095	GSM1158311	GSM1641278	GSM1665188	GSM1099816	GSM1216800
GSM1362096	GSM1158312	GSM1641279	GSM1665189	GSM907013	GSM1216801
GSM1362097	GSM1158313	GSM1505821	GSM1665190	GSM907014	GSM1216802
GSM1362098	GSM1158314	GSM1505822	GSM1665191	GSM907015	GSM1216803
GSM1362099	GSM1158315	GSM1505823	GSM1665192	GSM907016	GSM1216804
GSM1362100	GSM1158316	GSM1505825	GSM1665193	GSM907017	GSM1216805
GSM1362101	GSM1158317	GSM1505826	GSM1665194	GSM907018	GSM1216806
GSM1362102	GSM1158318	GSM1505828	GSM1665195	GSM916961	GSM1216807
GSM1362103	GSM1158319	GSM1505831	GSM1665196	GSM916962	GSM1216808
GSM1362104	GSM1158320	GSM1505834	GSM1665197	GSM916963	GSM1216809
GSM1362105	GSM1158321	GSM1505835	GSM1665198	GSM925605	GSM1216810
GSM1362106	GSM1158322	GSM1505837	GSM1665910	GSM925606	GSM1216811
GSM1362107	GSM1158323	GSM1505839	GSM1665911	GSM925607	GSM1216812
GSM1362108	GSM1158324	GSM1505840	GSM1665912	GSM925608	GSM1216813
GSM1362109	GSM1158325	GSM1505841	GSM1665913	GSM925613	GSM1216814
GSM1362110	GSM1158326	GSM1505842	GSM1665914	GSM925614	GSM1216815
GSM1362111	GSM1158327	GSM1505843	GSM1665915	GSM927073	GSM1216816
GSM1362112	GSM1158328	GSM1505846	GSM719425	GSM927074	GSM1216817
GSM1362113	GSM1158329	GSM1505847	GSM719427	GSM937708	GSM1216818
GSM1156797	GSM1158330	GSM1505848	GSM1677846	GSM937709	GSM1216819
GSM1156798	GSM1158331	GSM1505849	GSM1677847	GSM937710	GSM1216820
GSM1156799	GSM1158332	GSM1505850	GSM1677848	GSM937711	GSM1216821
GSM1156800	GSM1158333	GSM1505851	GSM1677849	GSM937712	GSM1216822
GSM1156801	GSM1158334	GSM1505854	GSM1678785	GSM937713	GSM1216823
GSM1156802	GSM1158335	GSM1505855	GSM1678786	GSM947444	GSM1216825
GSM1156803	GSM1158336	GSM1505856	GSM1678787	GSM947446	GSM1216826
GSM1156804	GSM1158337	GSM1505857	GSM1678788	GSM949822	GSM1216827

GSM1156805	GSM1158338	GSM1505860	GSM1678789	GSM949823	GSM1216828
GSM1156806	GSM1158339	GSM1505861	GSM1678790	GSM949825	GSM1216829
GSM1156807	GSM1158340	GSM1508256	GSM1678797	GSM949826	GSM1216830
GSM1156808	GSM1158341	GSM1508257	GSM1678798	GSM949827	GSM1216831
GSM1156809	GSM1158342	GSM1508258	GSM1678799	GSM949828	GSM1216832
GSM1156810	GSM1158343	GSM1508259	GSM1678800	GSM949829	GSM1216833
GSM1156811	GSM1158344	GSM1508260	GSM1678801	GSM949830	GSM1216834
GSM1156812	GSM1158345	GSM1508261	GSM1678802	GSM949831	GSM1216835
GSM1156813	GSM1158346	GSM1508262	GSM1679648	GSM949832	GSM1216836
GSM1156814	GSM1158347	GSM1508263	GSM1679649	GSM949833	GSM1216837
GSM1156815	GSM1158348	GSM1508264	GSM1679650	GSM949834	GSM1216838
GSM1156816	GSM1158349	GSM1508948	GSM1679651	GSM949835	GSM1216839
GSM1156817	GSM1158350	GSM1508949	GSM1679652	GSM949836	GSM1216840
GSM1156818	GSM1158351	GSM1508950	GSM1679653	GSM949837	GSM1216841
GSM1156819	GSM1158352	GSM1508951	GSM1679654	GSM949838	GSM1217954
GSM1156820	GSM1158353	GSM1508952	GSM1679655	GSM949839	GSM1217956
GSM1156821	GSM1158354	GSM1508953	GSM1679656	GSM949840	GSM1217958
GSM1156822	GSM1158355	GSM1509262	GSM1679657	GSM949841	GSM1217960
GSM1156823	GSM1158356	GSM1509265	GSM1679658	GSM949842	GSM1217961
GSM1156824	GSM1158357	GSM1509511	GSM1679659	GSM949843	GSM1219135
GSM1156825	GSM1158358	GSM1509512	GSM1679660	GSM949844	GSM1219136
GSM1156826	GSM1158359	GSM1509513	GSM1679661	GSM949845	GSM1224490
GSM1156827	GSM1158360	GSM1509514	GSM1679662	GSM955424	GSM1224491
GSM1156828	GSM1158361	GSM1510127	GSM1679663	GSM955160	GSM1224492
GSM1156829	GSM1158362	GSM1510128	GSM1679664	GSM955161	GSM1224493
GSM1156830	GSM1158363	GSM1510129	GSM1679665	GSM953381	GSM1224494
GSM1156831	GSM1158364	GSM1510130	GSM1679666	GSM953382	GSM1224495
GSM1156832	GSM1158365	GSM1510131	GSM1679667	GSM953383	GSM1224496
GSM1156834	GSM1158366	GSM1510132	GSM1679668	GSM953384	GSM1224497
GSM1156835	GSM1158367	GSM1510133	GSM1679669	GSM957471	GSM1224498

GSM1156836	GSM1158368	GSM1510134	GSM1679670	GSM957472	GSM1224499
GSM1156837	GSM1158369	GSM1510136	GSM1679671	GSM957473	GSM1226157
GSM1156838	GSM1158370	GSM1510137	GSM1679672	GSM957474	GSM1226158
GSM1156839	GSM1158371	GSM1510138	GSM1679673	GSM957475	GSM1226159
GSM1156840	GSM1158372	GSM1510139	GSM1679674	GSM970928	GSM1226160
GSM1156841	GSM1158373	GSM1510140	GSM1679675	GSM970929	GSM1226161
GSM1156842	GSM1158374	GSM1510141	GSM1679676	GSM970930	GSM1226162
GSM1156843	GSM1158375	GSM1511115	GSM1679677	GSM976973	GSM1226163
GSM1156844	GSM1158376	GSM1511116	GSM1679678	GSM976974	GSM1226164
GSM1156845	GSM1158377	GSM1511117	GSM1679679	GSM976975	GSM1226165
GSM1156846	GSM1158378	GSM1511118	GSM1679680	GSM976976	GSM1226166
GSM1156847	GSM1158379	GSM1511119	GSM1679681	GSM976977	GSM1226167
GSM1156848	GSM1158380	GSM1511120	GSM1679682	GSM976978	GSM1226168
GSM1156849	GSM1158381	GSM1511873	GSM1679683	GSM976979	GSM1228034
GSM1156850	GSM1158382	GSM1511874	GSM1679684	GSM976980	GSM1228035
GSM1156851	GSM1158383	GSM1511875	GSM1679685	GSM976981	GSM1228036
GSM1156852	GSM1158384	GSM1511876	GSM1679686	GSM976982	GSM1228037
GSM1156853	GSM1158385	GSM1511877	GSM1679687	GSM976983	GSM1228038
GSM1156854	GSM1158386	GSM1511878	GSM1679688	GSM976984	GSM1228039
GSM1156855	GSM1158387	GSM1511879	GSM1679689	GSM976985	GSM1228202
GSM1156856	GSM1158388	GSM1511880	GSM1679690	GSM976986	GSM1228203
GSM1156857	GSM1158389	GSM1513187	GSM1679691	GSM976987	GSM1228204
GSM1156858	GSM1158390	GSM1513188	GSM1679692	GSM976988	GSM1228205
GSM1156859	GSM1158391	GSM1513189	GSM1679693	GSM976989	GSM1228206
GSM1156860	GSM1158392	GSM1513190	GSM1679694	GSM978969	GSM1228207
GSM1156861	GSM1158393	GSM1513191	GSM1679695	GSM978970	GSM1228208
GSM1156862	GSM1158394	GSM1513192	GSM1679696	GSM992931	GSM1228209
GSM1156863	GSM1158395	GSM1513193	GSM1679697	GSM992932	GSM1228210
GSM1156864	GSM1158396	GSM1513194	GSM1679698	GSM992933	GSM1228211
GSM1156865	GSM1158397	GSM1513195	GSM1679699	GSM992934	GSM1228212

GSM1156866	GSM1158398	GSM1513196	GSM1679700	GSM997544	GSM1228213
GSM1156867	GSM1158399	GSM1513197	GSM1679701	GSM997545	GSM1228214
GSM1156868	GSM1158400	GSM1513198	GSM1679702	GSM997546	GSM1228215
GSM1156869	GSM1158401	GSM1513199	GSM1679703	GSM995300	GSM1228216
GSM1156870	GSM1158402	GSM1513200	GSM1679704	GSM995301	GSM1228217
GSM1156871	GSM1158403	GSM1513201	GSM1679705	GSM995302	GSM1228218
GSM1156872	GSM1158404	GSM1513202	GSM1679706	GSM995303	GSM1228219
GSM1156873	GSM1158405	GSM1513203	GSM1679707	GSM995304	GSM1229066
GSM1156874	GSM1158406	GSM1513204	GSM1679708	GSM990765	GSM1229067
GSM1156875	GSM1158407	GSM1513205	GSM1679709	GSM990766	GSM1229068
GSM1156876	GSM1158408	GSM1513206	GSM1679710	GSM990768	GSM1229069
GSM1156877	GSM1158409	GSM1513207	GSM1679711	GSM990769	GSM1229070
GSM1156878	GSM1158410	GSM1513208	GSM1679712	GSM990770	GSM1229071
GSM1156879	GSM1158411	GSM1513209	GSM1679713	GSM990771	GSM1229072
GSM1156880	GSM1158412	GSM1513210	GSM1679714	GSM990772	GSM1229103
GSM1156881	GSM1158413	GSM1513211	GSM1679715	GSM990773	GSM1229104
GSM1156882	GSM1158414	GSM1513212	GSM1679716	GSM990774	GSM1229105
GSM1156883	GSM1158415	GSM1513213	GSM1679717	GSM990775	GSM1229106
GSM1156884	GSM1158416	GSM1513214	GSM1679718	GSM990767	GSM1229107
GSM1156885	GSM1158417	GSM1513215	GSM1679719	GSM1002540	GSM1229108
GSM1156886	GSM1158418	GSM1513216	GSM1679720	GSM1002541	GSM1229109
GSM1156887	GSM1158419	GSM1513217	GSM1681901	GSM1002542	GSM1229110
GSM1156888	GSM1158420	GSM1513218	GSM1681902	GSM1002543	GSM1229111
GSM1156889	GSM1158421	GSM1513219	GSM1681903	GSM1002544	GSM1229112
GSM1156890	GSM1158422	GSM1513220	GSM1681904	GSM1002545	GSM1229113
GSM1156891	GSM1158423	GSM1513221	GSM1681905	GSM1002546	GSM1229114
GSM1156892	GSM1158424	GSM1513222	GSM1681906	GSM1002547	GSM1229116
GSM1156893	GSM1158425	GSM1513223	GSM1681907	GSM1002548	GSM1229117
GSM1156894	GSM1158426	GSM1513224	GSM1681908	GSM1002549	GSM1229118
GSM1156895	GSM1158427	GSM1513225	GSM1681909	GSM1002550	GSM1229119

GSM1156896	GSM1158428	GSM1513226	GSM1681910	GSM1002551	GSM1229120
GSM1156897	GSM1158429	GSM1513227	GSM1682266	GSM1002552	GSM1229121
GSM1156898	GSM1158430	GSM1513228	GSM1682267	GSM1002553	GSM1229123
GSM1156899	GSM1158431	GSM1513229	GSM721141	GSM1005575	GSM1229124
GSM1156900	GSM1158432	GSM1513230	GSM721123	GSM1006724	GSM1229125
GSM1156901	GSM1158433	GSM1513231	GSM721124	GSM1006725	GSM1229126
GSM1156902	GSM1158434	GSM1513232	GSM721125	GSM1005513	GSM1229127
GSM1156903	GSM1158435	GSM1513233	GSM721126	GSM1011896	GSM1233280
GSM1156904	GSM1158436	GSM1513234	GSM1686546	GSM1011897	GSM1233281
GSM1156905	GSM1158437	GSM1513235	GSM1686547	GSM1011898	GSM1233282
GSM1156906	GSM1158438	GSM1513236	GSM1686548	GSM1013679	GSM1233283
GSM1156907	GSM1158439	GSM1513237	GSM1687384	GSM1013682	GSM1233284
GSM1156908	GSM1158440	GSM1513238	GSM1687385	GSM1013684	GSM1233285
GSM1156909	GSM1158441	GSM1513239	GSM1687386	GSM1013686	GSM1233286
GSM1156910	GSM1158442	GSM1513240	GSM1687387	GSM1013688	GSM1233287
GSM1156911	GSM1158443	GSM1513241	GSM1693049	GSM1013692	GSM1233288
GSM1156912	GSM1158444	GSM1513242	GSM1693051	GSM1013693	GSM1233289
GSM1156913	GSM1158445	GSM1513243	GSM1693052	GSM1013695	GSM1233290
GSM1156914	GSM1158446	GSM1513244	GSM1693053	GSM1013697	GSM1233291
GSM1156915	GSM1158447	GSM1513245	GSM1693054	GSM1018004	ERR169802
GSM1156916	GSM1158448	GSM1513246	GSM1693055	GSM1018005	ERR169803
GSM1156917	GSM1158449	GSM1513247	GSM1694663	GSM1020212	ERR358486
GSM1156918	GSM1158450	GSM1513248	GSM1694664	GSM1020213	ERR380549
GSM1156919	GSM1158451	GSM1513249	GSM1694665	GSM1020214	ERR380552
GSM1156920	GSM1158452	GSM1513250	GSM1694666	GSM1020215	GSM1563053
GSM1156921	GSM1158453	GSM1513251	GSM1695162	GSM1020216	GSM1563054
GSM1156922	GSM1158454	GSM1513252	GSM1695197	GSM1023059	GSM1573117
GSM1156923	GSM1158455	GSM1513253	GSM1695198	GSM1023060	GSM984650
GSM1156924	GSM1158456	GSM1513254	GSM1695199	GSM1023061	ENCFF320IDT
GSM1156925	GSM1158457	GSM1513255	GSM1695850	GSM1023062	ENCFF380GBC

GSM1156926	GSM1158458	GSM1513256	GSM1695851	GSM1023063	ENCFF888LPS
GSM1156927	GSM1158459	GSM1513257	GSM1695852	GSM1023064	ENCFF342LYI
GSM1156928	GSM1158460	GSM1513258	GSM1695853	GSM1023065	ENCFF237ZQX
GSM1156929	GSM1158461	GSM1513689	GSM1695854	GSM1023066	ENCFF466QUZ
GSM1156930	GSM1158462	GSM1517598	GSM1695855	GSM1023067	ENCFF290OQE
GSM1156931	GSM1158463	GSM1517599	GSM1695856	GSM1023068	ENCFF789VZB
GSM1156932	GSM1158464	GSM1517600	GSM1695857	GSM1023069	ENCFF672VVX
GSM1156933	GSM1158465	GSM1517601	GSM1695858	GSM1023070	ENCFF256APB
GSM1156934	GSM1158466	GSM1519560	GSM1695859	GSM1023071	ENCFF299BIL
GSM1156935	GSM1158467	GSM1519561	GSM1695860	GSM1023072	ENCFF036GDL
GSM1156936	GSM1158468	GSM1519562	GSM1695861	GSM1023073	S004BT
GSM1156937	GSM1158469	GSM1519563	GSM1695862	GSM1023074	S002S3
GSM1156938	GSM1158470	GSM1519564	GSM1695863	GSM1023075	C005PS
GSM1156939	GSM1158471	GSM1519565	GSM1695864	GSM1023076	ENCFF690QPA
GSM1156940	GSM1158472	GSM1519566	GSM1695865	GSM1023077	ENCFF773MOU
GSM1156941	GSM1158473	GSM1519567	GSM1695866	GSM1023078	

References

[1] S. Agatonovic-Kustrin and R. Beresford. Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22(5):717 – 727, 2000.

[2] U. Ahluwalia, N. Katyal, and S. Deep. Models of protein folding. *Journal of Proteins & Proteomics*, 3(2), 2013.

[3] M. Ali and A. del Sol. *Modeling of Cellular Systems: Application in Stem Cell Research and Computational Disease Modeling*, pages 129–138. Springer International Publishing, Cham, 2018.

[4] K. Alkadhi and J. Eriksen. The complex and multifactorial nature of alzheimer's disease. *Curr Neuropharmacol*, 9(4):586–586, Dec 2011.

[5] L. C. Amado, A. P. Saliaris, K. H. Schuleri, M. St. John, J.-S. Xie, S. Cattaneo, D. J. Durand, T. Fitton, J. Q. Kuang, G. Stewart, S. Lehrke, W. W. Baumgartner, B. J. Martin, A. W. Heldman, and J. M. Hare. Cardiac repair with intramyocardial injection of allogeneic mesenchymal stem cells after myocardial infarction. *Proceedings of the National Academy of Sciences*, 102(32):11474–11479, 2005.

[6] S. Andrews. Fastqc: a quality control tool for high throughput sequence data. available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. *online*, 2010.

[7] C. Angelini and V. Costa. Understanding gene regulatory mechanisms by integrating chip-seq and rna-seq data: statistical solutions to biological problems. *Front Cell Dev Biol*, 2:51, 2014.

[8] J. Arias-Fuenzalida, J. Jarazo, X. Qing, J. Walter, G. Gomez-Giro, S. L. Nickels, H. Zaehres, H. R. Schöler, and J. C. Schwamborn. Facs-assisted crispr-cas9 genome editing facilitates parkinson's disease modeling. *Stem Cell Reports*, 9(5):1423–1431, Oct 2017.

[9] M. J. Aryee, A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, and R. A. Irizarry. Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium dna methylation microarrays. *Bioinformatics*, 30(10):1363–1369, May 2014.

[10] A. Avgustinova and S. A. Benitah. Epigenetic control of adult stem cell function. *Nat Rev Mol Cell Biol*, 17(10):643–658, Oct 2016.

[11] A. Ay and D. N. Arnosti. Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit Rev Biochem Mol Biol*, 46(2):137–151, 2011.

[12] B. A. Bahr and J. Bendiske. The neuropathogenic contributions of lysosomal dysfunction. *Journal of Neurochemistry*, 83(3):481–489, 2002.

[13] A.-L. Barabasi, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nat Rev Genet*, 12(1):56–68, Jan 2011.

[14] A. Bashashati, G. Haffari, J. Ding, G. Ha, K. Lui, J. Rosner, D. G. Huntsman, C. Caldas, S. A. Aparicio, and S. P. Shah. Drivernet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol*, 13(12):R124–R124, Dec 2012.

[15] A. G. Bassuk and J. M. Leiden. The role of ets transcription factors in the development and function of the mammalian immune system. volume 64 of *Advances in Immunology*, pages 65 – 104. Academic Press, 1997.

[16] F. Bature, B.-a. Guinn, D. Pang, and Y. Pappas. Signs and symptoms preceding the diagnosis of alzheimer's disease: a systematic scoping review of literature from 1937 to 2016. *BMJ Open*, 7(8), 2017.

[17] T. G. Beach, C. H. Adler, L. I. Sue, G. Serrano, H. A. Shill, D. G. Walker, L. Lue, A. E. Roher, B. N. Dugger, C. Maarouf, A. C. Birdsill, A. Intorcia, M. Saxon-Labelle, J. Pullen, A. Scroggins, J. Filon, S. Scott, B. Hoffman, A. Garcia, J. N. Caviness, J. G. Hentz,

E. Driver-Dunckley, S. A. Jacobson, K. J. Davis, C. M. Belden, K. E. Long, M. Malek-Ahmadi, J. J. Powell, L. D. Gale, L. R. Nicholson, R. J. Caselli, B. K. Woodruff, S. Z. Rapscak, G. L. Ahern, J. Shi, A. D. Burke, E. M. Reiman, and M. N. Sabbagh. Arizona study of aging and neurodegenerative disorders and brain and body donation program. *Neuropathology*, 35(4):354–389, Aug 2015.

[18] T. G. Beach, L. I. Sue, D. G. Walker, A. E. Roher, L. Lue, L. Vedders, D. J. Connor, M. N. Sabbagh, and J. Rogers. The sun health research institute brain donation program: description and experience, 1987-2007. *Cell Tissue Bank*, 9(3):229–245, Sep 2008.

[19] F. Belinky, N. Nativ, G. Stelzer, S. Zimmerman, T. Iny Stein, M. Safran, and D. Lancet. Pathcards: multi-source consolidation of human biological pathways. *Database (Oxford)*, 2015:bav006, Feb 2015.

[20] C. M. Bellettato and M. Scarpa. Pathophysiology of neuropathic lysosomal storage disorders. *Journal of Inherited Metabolic Disease*, 33(4):347–362, 2010.

[21] A. D. Berendsen and B. R. Olsen. Bone development. *Bone*, 80:14–18, Nov 2015.

[22] G. M. Bernardo and R. A. Keri. Foxa1: a transcription factor with parallel functions in development and cancer. *Bioscience Reports*, 32(2):113–130, 2012.

[23] B. E. Bernstein, J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. A. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham, M. Hirst, E. S. Lander, T. S. Mikkelsen, and J. A. Thomson. The nih roadmap epigenomics mapping consortium. *Nat Biotech*, 28(10):1045–1048, 2010.

[24] P. Blohm, G. Frishman, P. Smialowski, F. Goebels, B. Wachinger, A. Ruepp, and D. Frishman. Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res*, 42(Database issue):D396–D400, Jan 2014.

[25] D. Bonneh Barkay and C. A. Wiley. Brain extracellular matrix in neurodegeneration. *Brain Pathol*, 19(4):573–585, Oct 2009.

[26] A.-L. Boulesteix and K. Strimmer. Predicting transcription factor activities from combined

analysis of microarray and chip data: a partial least squares approach. *Theor Biol Med Model*, 2:23–23, 2005.

[27] L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–956, Sep 2005.

[28] O. Britanova, C. de Juan Romero, A. Cheung, K. Y. Kwan, M. Schwark, A. Gyorgy, T. Vogel, S. Akopov, M. Mitkovski, D. Agoston, N. Sestan, Z. Molnár, and V. Tarabykin. Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron*, 57(3):378–392, Feb 2008.

[29] G. W. Brodland. How computational models can help unlock biological systems. *Seminars in Cell & Developmental Biology*, 47-48(Supplement C):62 – 73, 2015.

[30] A. E. Budson and P. R. Solomon. New diagnostic criteria for alzheimer's disease and mild cognitive impairment for the practical neurologist. *Practical Neurology*, 12(2):88–96, 2012.

[31] Y. Buganim, E. Itskovich, Y.-C. Hu, A. Cheng, K. Ganz, S. Sarkar, D. Fu, G. G. Welstead, D. Page, and R. Jaenisch. Direct reprogramming of fibroblasts into embryonic sertoli-like cells by defined factors. *Cell Stem Cell*, 11(3):373–386, Sep 2012.

[32] D. Caccavo, B. Laganà, A. P. Mitterhofer, G. M. Ferri, A. Afeltra, A. Amoroso, and L. Bonomo. Long-term treatment of systemic lupus erythematosus with cyclosporin a. *Arthritis & Rheumatism*, 40(1):27–35, 1997.

[33] Y. Cai, S. S. A. An, and S. Kim. Mutations in presenilin 2 and its implications in alzheimer's disease and other dementia-associated disorders. *Clin Interv Aging*, 10:1163–1172, Jul 2015.

[34] M. Caiazzo, S. Giannelli, P. Valente, G. Lignani, A. Carissimo, A. Sessa, G. Colasante, R. Bartolomeo, L. Massimino, S. Ferroni, C. Settembre, F. Benfenati, and V. Broccoli. Direct conversion of fibroblasts into functional astrocytes by defined transcription factors. *Stem Cell Reports*, 4(1):25–36, 2015.

[35] S. E. Calvano, W. Xiao, D. R. Richards, R. M. Felciano, H. V. Baker, R. J. Cho, R. O. Chen, B. H. Brownstein, J. P. Cobb, S. K. Tschoeke, C. Miller-Graziano, L. L. Moldawer, M. N. Mindrinos, R. W. Davis, R. G. Tompkins, S. F. Lowry, I. Program, and H. R. to Injury Large Scale Collaborative Research. A network-based analysis of systemic inflammation in humans. *Nature*, 437(7061):1032–1037, 2005.

[36] G. I. Cancino, A. P. Yiu, M. P. Fatt, C. B. Dugani, E. R. Flores, P. W. Frankland, S. A. Josselyn, F. D. Miller, and D. R. Kaplan. p63 regulates adult neural precursor and newly born neuron survival to control hippocampal-dependent behavior. *J Neurosci*, 33(31):12569–12585, Jul 2013.

[37] J. Carcel-Trullols, A. D. Kovács, and D. A. Pearce. Cell biology of the ncl proteins: What they do and don't do. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1852(10, Part B):2242 – 2255, 2015.

[38] R. Chang, R. Shoemaker, and W. Wang. Systematic search for recipes to generate induced pluripotent stem cells. *PLOS Computational Biology*, 7(12):1–13, 12 2011.

[39] S. Chattopadhyay, M. Ito, J. D. Cooper, A. I. Brooks, T. M. Curran, J. M. Powers, and D. A. Pearce. An autoantibody inhibitory to glutamic acid decarboxylase in the neurodegenerative disorder batten disease. *Human molecular genetics*, 11 12:1421–31, 2002.

[40] L. Y. Chee and A. Cumming. Polymorphisms in the cholinergic receptors muscarinic (chrm2 and chrm3) genes and alzheimer's disease. *Avicenna J Med Biotechnol*, 10(3):196–199, 2018.

[41] L. Chen, B. Ge, F. P. Casale, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yan, K. Kundu, S. Ecker, A. Datta, D. Richardson, F. Burden, D. Mead, A. L. Mann, J. M. Fernandez, S. Rowlston, S. P. Wilder, S. Farrow, X. Shao, J. J. Lambourne, A. Redensek, C. A. Albers, V. Amstislavskiy, S. Ashford, K. Berentsen, L. Bomba, G. Bourque, D. Bujold, S. Busche, M. Caron, S.-H. Chen, W. Cheung, O. Delaneau, E. T. Dermitzakis, H. Elding, I. Colgiu, F. O. Bagger, P. Flicek, E. Habibi, V. Iotchkova, E. Janssen-Megens, B. Kim, H. Lehrach, E. Lowy, A. Mandoli, F. Matarese, M. T. Maurano, J. A. Morris, V. Pancaldi, F. Pourfarzad, K. Rehnstrom, A. Rendon, T. Risch, N. Sharifi, M.-M. Simon, M. Sultan, A. Valencia, K. Walter, S.-Y. Wang, M. Frontini, S. E. Antonarakis, L. Clarke, M.-L. Yaspo,

S. Beck, R. Guigo, D. Rico, J. H. A. Martens, W. H. Ouwehand, T. W. Kuijpers, D. S. Paul, H. G. Stunnenberg, O. Stegle, K. Downes, T. Pastinen, and N. Soranzo. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell*, 167(5):1398–1414.e24, Nov 2016.

[42] T. Chen and S. Y. R. Dent. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*, 15(2):93–106, Feb 2014.

[43] Y.-a. Chen, M. Lemire, S. Choufani, D. T. Butcher, D. Grafodatskaya, B. W. Zanke, S. Gallinger, T. J. Hudson, and R. Weksberg. Discovery of cross-reactive probes and polymorphic cpgs in the illumina infinium humanmethylation450 microarray. *Epigenetics*, 8(2):203–209, Feb 2013.

[44] X. Cheng. Structural and functional coordination of dna and histone methylation. *Cold Spring Harb Perspect Biol*, 6(8):10.1101/cshperspect.a018747 a018747, Aug 2014.

[45] N.-Y. Chia, Y.-S. Chan, B. Feng, X. Lu, Y. L. Orlov, D. Moreau, P. Kumar, L. Yang, J. Jiang, M.-S. Lau, M. Huss, B.-S. Soh, P. Kraus, P. Li, T. Lufkin, B. Lim, N. D. Clarke, F. Bard, and H.-H. Ng. A genome-wide rnai screen reveals determinants of human embryonic stem cell identity. *Nature*, 468:316 EP –, Oct 2010.

[46] J. Choi, M. L. Costa, C. S. Mermelstein, C. Chagas, S. Holtzer, and H. Holtzer. Myod converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc Natl Acad Sci U S A*, 87(20):7988–7992, 1990.

[47] J. Chou, S. Provot, and Z. Werb. Gata3 in development and cancer differentiation: cells gata have it! *J Cell Physiol*, 222(1):42–49, Jan 2010.

[48] Y. S. Chun, K. Byun, and B. Lee. Induced pluripotent stem cells and personalized medicine: current progress and future perspectives. *Anat Cell Biol*, 44(4):245–255, Dec 2011.

[49] H. Clevers. Modeling development and disease with organoids. *Cell*, 165(7):1586–1597, Jun 2016.

[50] E. Clough and T. Barrett. The gene expression omnibus database. *Methods Mol Biol*, 1418:93–110, 2016.

[51] L. Collado-Torres, A. Nellore, K. Kammers, S. E. Ellis, M. A. Taub, K. D. Hansen, A. E. Jaffe, B. Langmead, and J. T. Leek. Reproducible rna-seq analysis using recount2. *Nature Biotechnology*, 35:319 EP –, Apr 2017.

[52] E. P. Consortium. The encode (encyclopedia of dna elements) project. *Science*, 306(5696):636–640, 2004.

[53] R. E. Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–330, Feb 2015.

[54] T. E. P. Consortium. A user's guide to the encyclopedia of dna elements (encode). *PLOS Biology*, 9(4):1–21, 04 2011.

[55] T. G. O. Consortium. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res*, 47(D1):D330–D338, Jan 2019.

[56] E. Corder, A. Saunders, W. Strittmatter, D. Schmechel, P. Gaskell, G. Small, A. Roses, J. Haines, and M. Pericak-Vance. Gene dose of apolipoprotein e type 4 allele and the risk of alzheimer's disease in late onset families. *Science*, 261(5123):921–923, 1993.

[57] S. L. Cotman and J. F. Staropoli. The juvenile batten disease protein, cln3, and its role in

regulating anterograde and retrograde post-golgi trafficking. *Clin Lipidol*, 7(1):79–91, Feb 2012.

[58] I. Crespo, T. M. Perumal, W. Jurkowski, and A. del Sol. Detecting cellular reprogramming determinants by differential stability analysis of gene regulatory networks. *BMC Syst Biol*, 7:140–140, 2013.

[59] A. C. D'Alessio, Z. P. Fan, K. J. Wert, P. Baranov, M. A. Cohen, J. S. Saini, E. Cohick, C. Charniga, D. Dadon, N. M. Hannett, M. J. Young, S. Temple, R. Jaenisch, T. I. Lee, and R. A. Young. A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Reports*, 5(5):763–775, Nov 2015.

[60] F. P. Davis and S. R. Eddy. Transcription factors that convert adult cell identity are differentially polycomb repressed. *PLOS ONE*, 8(5):1–8, 05 2013.

[61] P. L. De Jager, G. Srivastava, K. Lunnon, J. Burgess, L. C. Schalkwyk, L. Yu, M. L. Eaton, B. T. Keenan, J. Ernst, C. McCabe, A. Tang, T. Raj, J. Replogle, W. Brodeur, S. Gabriel, H. S. Chai, C. Younkin, S. G. Younkin, F. Zou, M. Szyf, C. B. Epstein, J. A. Schneider, B. E. Bernstein, A. Meissner, N. Ertekin-Taner, L. B. Chibnik, M. Kellis, J. Mill, and D. A. Bennett. Alzheimer's disease: early alterations in brain dna methylation at ank1, bin1, rhbdf2 and other loci. *Nature Neuroscience*, 17:1156 EP –, Aug 2014.

[62] A. del Sol, R. Balling, L. Hood, and D. Galas. Diseases as network perturbations. *Current Opinion in Biotechnology*, 21(4):566 – 571, 2010.

[63] D. L. V. den Hove, K. Kompotis, R. Lardenoije, G. Kenis, J. Mill, H. W. Steinbusch, K.-P. Lesch, C. P. Fitzsimons, B. D. Strooper, and B. P. Rutten. Epigenetically regulated micrornas in alzheimer's disease. *Neurobiology of Aging*, 35(4):731 – 745, 2014.

[64] P. Dhingra, A. Martinez-Fundichely, A. Berger, F. W. Huang, A. N. Forbes, E. M. Liu, D. Liu, A. Sboner, P. Tamayo, D. S. Rickman, M. A. Rubin, and E. Khurana. Identification of novel prostate cancer drivers using regnetdriver: a framework for integration of genetic and epigenetic alterations with tissue-specific regulatory network. *Genome Biology*, 18(1):141, Jul 2017.

[65] G. Di Fede, M. Catania, E. Maderna, R. Ghidoni, L. Benussi, E. Tonoli, G. Giaccone, F. Moda, A. Paterlini, I. Campagnani, S. Sorrentino, L. Colombo, A. Kubis, E. Bistaffa, B. Ghetti, and F. Tagliavini. Molecular subtypes of alzheimer's disease. *Sci Rep*, 8(1):3269–3269, Feb 2018.

[66] C. Dimitrakopoulos, S. K. Hindupur, L. Häfliger, J. Behr, H. Montazeri, M. N. Hall, and N. Beerenwinkel. Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics*, 34(14):2441–2448, Jul 2018.

[67] X. Dong and Z. Weng. The correlation between histone modifications and gene expression. *Epigenomics*, 5(2):113–116, Apr 2013.

[68] Z. Duren, X. Chen, R. Jiang, Y. Wang, and W. H. Wong. Modeling gene regulation from paired expression and chromatin accessibility data. *Proceedings of the National Academy of Sciences*, 114(25):E4914–E4923, 2017.

[69] N. D. Dwyer and D. D. M. O'Leary. Tbr1 conducts the orchestration of early cortical development. *Neuron*, 29(2):309–311, Feb 2001.

[70] A. D. Ebert, P. Liang, and J. C. Wu. Induced pluripotent stem cells as a disease modeling and drug screening platform. *Journal of Cardiovascular Pharmacology*, 60(4), 2012.

[71] M. J. Eckler and B. Chen. Fez family transcription factors: controlling neurogenesis and cell fate in the developing mammalian nervous system. *Bioessays*, 36(8):788–797, Aug 2014.

[72] W. El Yakoubi, C. Borday, J. Hamdache, K. Parain, H. T. Tran, K. Vleminckx, M. Perron, and M. Locker. Hes4 controls proliferative properties of neural stem cells during retinal ontogenesis. *Stem Cells*, 30(12):2784–2795, Dec 2012.

[73] V. Espinosa Angarica and A. del Sol. Modeling heterogeneity in the pluripotent state: A promising strategy for improving the efficiency and fidelity of stem cell differentiation. *BioEssays*, 38(8):758–768, 2016.

[74] R. M. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M. D. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore,

S. Zhang, O. Ornatsky, Y. V. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu, M. Abu-Farha, J.-P. Lambert, H. S. Duewel, I. I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S.-L. Adams, M. F. Moran, G. B. Morin, T. Topaloglou, and D. Figeys. Large-scale mapping of human protein–protein interactions by mass spectrometry. *Molecular Systems Biology*, 3(1), 2007.

[75] J. Ezaki, M. Takeda-Ezaki, and E. Kominami. Tripeptidyl peptidase i, the late infantile neuronal ceroid lipofuscinosis gene product, initiates the lysosomal degradation of subunit c of atp synthasel. *The Journal of Biochemistry*, 128(3):509–516, 2000.

[76] E. Ezhkova, W.-H. Lien, N. Stokes, H. A. Pasolli, J. M. Silva, and E. Fuchs. Ezh1 and ezh2 cogovern histone h3k27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes & Development*, 25(5):485–498, 2011.

[77] K. S. F., M.-G. M. Jose, D.-R. Arce, B. Jose, and T.-P. Maria. *sagmb*, volume 18, chapter MLML2R: an R package for maximum likelihood estimation of DNA methylation and hydroxymethylation proportions. Springer International Publishing, 2019 2019.

[78] A. L. M. Ferri, W. Lin, Y. E. Mavromatakis, J. C. Wang, H. Sasaki, J. A. Whitsett, and S.-L. Ang. Foxa1 and foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. *Development*, 134(15):2761–2769, 2007.

[79] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P. R. Menezes, E. Rimmer, M. Scazufca, and A. D. International. Global prevalence of dementia: a delphi consensus study. *Lancet*, 366(9503):2112–2117, Dec 2005.

[80] S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, and D. Cohen. Genehancer: genome-wide integration of enhancers and target genes in genecards. *Database (Oxford)*, 2017:bax028, Apr 2017.

[81] N. Folguera-Blasco, E. Cuyàs, J. A. Menéndez, and T. Alarcón. Epigenetic regulation of cell fate reprogramming in aging and disease: A predictive computational model. *PLoS Comput Biol*, 14(3):e1006052, Mar 2018.

[82] M. Fournier, G. Bourriquen, F. C. Lamaze, M. C. Côté, É. Fournier, C. Joly-Beauparlant, V. Caron, S. Gobeil, A. Droit, and S. Bilodeau. Foxa and master transcription factors recruit mediator and cohesin to the core transcriptional regulatory circuitry of cancer cells. *Scientific Reports*, Oct 2016.

[83] L. Franke, H. van Bakel, L. Fokkens, E. D. de Jong, M. Egmont-Petersen, and C. Wijmenga. Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *The American Journal of Human Genetics*, 78(6):1011 – 1025, 2006.

[84] M.-A. Frese, S. Schulz, and T. Dierks. Arylsulfatase g, a novel lysosomal sulfatase. *Journal of Biological Chemistry*, 283(17):11388–11395, 2008.

[85] J.-D. Fu, N. Stone, L. Liu, C. Spencer, L. Qian, Y. Hayashi, P. Delgado-Olguin, S. Ding, B. Bruneau, and D. Srivastava. Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. *Stem Cell Reports*, 1(3):235–247, Sep 2013.

[86] Z. Gao, G. H. Kim, A. C. Mackinnon, A. E. Flagg, B. Bassett, J. U. Earley, and E. C. Svensson. Ets1 is required for proper migration and differentiation of the cardiac neural crest. *Development*, 137(9):1543–1551, 2010.

[87] S. B. Gaudreault, D. Dea, and J. Poirier. Increased caveolin-1 expression in alzheimer's disease brain. *Neurobiology of Aging*, 25(6):753 – 759, 2004.

[88] A. Giorgetti, N. Montserrat, T. Aasen, F. Gonzalez, I. Rodríguez-Pizà, R. Vassena, A. Raya, S. Boué, M. J. Barrero, B. A. Corbella, M. Torrabadella, A. Veiga, and J. C. I. Belmonte. Generation of induced pluripotent stem cells from human cord blood using oct4 and sox2. *Cell Stem Cell*, 5(4):353 – 357, 2009.

[89] M. Giri, M. Zhang, and Y. Lü. Genes associated with alzheimer's disease: an overview and current status. *Clin Interv Aging*, 11:665–681, May 2016.

[90] G. G. Giro, J. Arias-Fuenzalida, J. J., D. Zeuschner, M. Ali, S. Bolognin, R. Halder, C. Jäger, H. Zaheres, A. del Sol, H. R. Schöler, and J. C. Schwamborn. Modeling juvenile neuronal ceroid lipofuscinosis by genome editing in human induced pluripotent stem cells and cerebral organoids. *Under preparation*, Aug 2019.

[91] E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, and A. Valencia. Enrichnet: network-based gene set enrichment analysis. *Bioinformatics*, 28(18):i451–i457, Sep 2012.

[92] A. A. Golabek and E. Kida. *Tripeptidyl-peptidase I in health and disease*, volume 387, pages 1091–1099. Springer International Publishing, 2006.

[93] E. Gonçalves, J. Bucher, A. Ryll, J. Niklas, K. Mauch, S. Klamt, M. Rocha, and J. Saez-Rodriguez. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Mol. BioSyst.*, 9:1576–1583, 2013.

[94] J.-L. Gouze. Positive and negative circuits in dynamical systems. *Journal of Biological Systems*, 06(01):11–15, 1998.

[95] A. M. Grabiec and K. A. Reedquist. The ascent of acetylation in the epigenetics of rheumatoid arthritis. *Nature Reviews Rheumatology*, 9:311 EP –, Feb 2013.

[96] T. Graf and T. Enver. Forcing cells to change lineages. *Nature*, 462:587 EP –, Dec 2009.

[97] W. E. Grizzle, W. C. Bell, and K. C. Sexton. Issues in collecting, processing and storing human tissues and associated information to support biomedical research. *Cancer Biomark*, 9(1-6):531–549, 2010.

[98] E. Gulaj, K. Pawlak, B. Bien, and D. Pawlak. Kynurenine and its metabolites in alzheimer's disease patients. *Advances in Medical Sciences*, 55(2):204 – 211, 2010.

[99] C. Hanashima, S. C. Li, L. Shen, E. Lai, and G. Fishell. Foxg1 suppresses early cortical cell fate. *Science*, 303(5654):56–59, 2004.

[100] D. Harold, R. Abraham, P. Hollingworth, R. Sims, A. Gerrish, M. L. Hamshere, J. S. Pahwa, V. Moskvina, K. Dowzell, A. Williams, N. Jones, C. Thomas, A. Stretton, A. R. Morgan, S. Lovestone, J. Powell, P. Proitsi, M. K. Lupton, C. Brayne, D. C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. Morgan, K. S. Brown, P. A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A. D. Smith, S. Love, P. G. Kehoe, J. Hardy, S. Mead, N. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, B. Schürmann, R. Heun, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frölich, H. Hampel, M. Hüll, D. Rujescu, A. M. Goate, J. S. K. Kauwe, C. Cruchaga, P. Nowotny, J. C. Morris, K. Mayo,

K. Sleegers, K. Bettens, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, G. Livingston, N. J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C. E. Shaw, M. Tsolaki, A. B. Singleton, R. Guerreiro, T. W. Mühleisen, M. M. Nöthen, S. Moebus, K.-H. Jöckel, N. Klopp, H.-E. Wichmann, M. M. Carrasquillo, V. S. Pankratz, S. G. Younkin, P. A. Holmans, M. O'Donovan, M. J. Owen, and J. Williams. Genome-wide association study identifies variants at clu and picalm associated with alzheimer&#39;s disease. *Nature Genetics*, 41:1088 EP –, Sep 2009.

[101] T. Hartmann, J. Kuchenbecker, and M. O. W. Grimm. Alzheimer's disease: the lipid connection. *Journal of Neurochemistry*, 103(s1):159–170, 2007.

[102] Y. Hasin, M. Seldin, and A. Lusis. Multi-omics approaches to disease. *Genome Biology*, 18(1):83, May 2017.

[103] N. J. Haughey, V. V. R. Bandaru, M. Bae, and M. P. Mattson. Roles for dysfunctional sphingolipid metabolism in alzheimer's disease neuropathogenesis. *Biochim Biophys Acta*, 1801(8):878–886, Aug 2010.

[104] S. R. Hegde, K. Pal, and S. C. Mande. Differential enrichment of regulatory motifs in the composite network of protein-protein and gene regulatory interactions. *BMC Systems Biology*, 8(1):26, Feb 2014.

[105] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell*, 38(4):576–589, May 2010.

[106] S. Heinz, C. E. Romanoski, C. Benner, and C. K. Glass. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*, 16(3):144–154, Mar 2015.

[107] H. Hengel, A. Magee, M. Mahanjah, J.-M. Vallat, R. Ouvrier, M. Abu-Rashid, J. Mahamid, R. Schüle, M. Schulze, I. Krägeloh-Mann, P. Bauer, S. Züchner, R. Sharkia, and L. Schöls. Cntnap1 mutations cause cns hypomyelination and neuropathy with or without arthrogryposis. *Neurol Genet*, 3(2):e144–e144, Mar 2017.

[108] L. B. Holder, M. M. Haque, and M. K. Skinner. Machine learning for epigenetics and future medical applications. *Epigenetics*, 12(7):505–514, 2017.

[109] P. Hollingworth, D. Harold, R. Sims, A. Gerrish, J.-C. Lambert, M. M. Carrasquillo, R. Abraham, M. L. Hamshere, J. S. Pahwa, V. Moskvina, K. Dowzell, N. Jones, A. Stretton, C. Thomas, A. Richards, D. Ivanov, C. Widdowson, J. Chapman, S. Lovestone, J. Powell, P. Proitsi, M. K. Lupton, C. Brayne, D. C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. S. Brown, P. A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A. D. Smith, H. Beaumont, D. Warden, G. Wilcock, S. Love, P. G. Kehoe, N. M. Hooper, E. R. L. C. Vardy, J. Hardy, S. Mead, N. C. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, E. Rüther, B. Schürmann, R. Heun, H. Kölsch, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frölich, H. Hampel, J. Gallacher, M. Hüll, D. Rujescu, I. Giegling, A. M. Goate, J. S. K. Kauwe, C. Cruchaga, P. Nowotny, J. C. Morris, K. Mayo, K. Sleegers, K. Bettens, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, G. Livingston, N. J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C. E. Shaw, M. Tsolaki, A. B. Singleton, R. Guerreiro, T. W. Mühleisen, M. M. Nöthen, S. Moebus, K.-H. Jöckel, N. Klopp, H.-E. Wichmann, V. S. Pankratz, S. B. Sando, J. O. Aasly, M. Barcikowska, Z. K. Wszolek, D. W. Dickson, N. R. Graff-Radford, R. C. Petersen, t. A. D. N. Initiative, C. M. van Duijn, M. M. B. Breteler, M. A. Ikram, A. L. DeStefano, A. L. Fitzpatrick, O. Lopez, L. J. Launer, S. Seshadri, C. consortium, C. Berr, D. Campion, J. Epelbaum, J.-F. Dartigues, C. Tzourio, A. Alpérovitch, M. Lathrop, E. consortium, T. M. Feulner, P. Friedrich, C. Riehle, M. Krawczak, S. Schreiber, M. Mayhaus, S. Nicolhaus, S. Wagenpfeil, S. Steinberg, H. Stefansson, K. Stefansson, J. Snædal, S. Björnsson, P. V. Jonsson, V. Chouraki, B. Genier-Boley, M. Hiltunen, H. Soininen, O. Combarros, D. Zelenika, M. Delepine, M. J. Bullido, F. Pasquier, I. Mateo, A. Frank-Garcia, E. Porcellini, O. Hanon, E. Coto, V. Alvarez, P. Bosco, G. Siciliano, M. Mancuso, F. Panza, V. Solfrizzi, B. Nacmias, S. Sorbi, P. Bossù, P. Piccardi, B. Arosio, G. Annoni, D. Seripa, A. Pilotto, E. Scarpini, D. Galimberti, A. Brice, D. Hannequin, F. Licastro, L. Jones, P. A. Holmans, T. Jonsson, M. Riemenschneider, K. Morgan, S. G. Younkin, M. J. Owen, M. O'Donovan, P. Amouyel, and J. Williams. Common variants at abca7, ms4a6a/ms4a4e, epha1, cd33 and cd2ap are associated with alzheimer&#39;s disease. *Nature Genetics*, 43:429 EP –, Apr

2011.

[110] C. R. Horres and Y. A. Hannun. The roles of neutral sphingomyelinases in neurological pathologies. *Neurochemical Research*, 37(6):1137–1149, Jun 2012.

[111] P.-S. Hou, C.-Y. Chuang, C.-H. Yeh, W. Chiang, H.-J. Liu, T.-N. Lin, and H.-C. Kuo. Direct conversion of human fibroblasts into neural progenitors using transcription factors enriched in human esc-derived neural progenitors. *Stem Cell Reports*, 8(1):54–68, Jan 2017.

[112] Y. Hou, S. Lautrup, S. Cordonnier, Y. Wang, D. L. Croteau, E. Zavala, Y. Zhang, K. Moritoh, J. F. O'Connell, B. A. Baptiste, T. V. Stevnsner, M. P. Mattson, and V. A. Bohr. Nad+ supplementation normalizes key alzheimer's features and dna damage responses in a new ad mouse model with introduced dna repair deficiency. *Proceedings of the National Academy of Sciences*, 115(8):E1876–E1885, 2018.

[113] M. Huang, Y. Chen, M. Yang, A. Guo, Y. Xu, L. Xu, and H. P. Koeffler. dbcorc: a database of core transcriptional regulatory circuitries modeled by h3k27ac chip-seq signals. *Nucleic Acids Research*, 46(D1):D71–D77, 2018.

[114] D. Huangfu, K. Osafune, R. Maehr, W. Guo, A. Eijkelenboom, S. Chen, W. Muhlestein, and D. A. Melton. Induction of pluripotent stem cells from primary human fibroblasts with only oct4 and sox2. *Nature Biotechnology*, 26:1269 EP –, Oct 2008.

[115] M. Ichikawa, T. Asai, S. Chiba, M. Kurokawa, and S. Ogawa. Runx1/aml-1 ranks as a master regulator of adult hematopoiesis. *Cell Cycle*, 3(6):720–722, 2004.

[116] T. Ideker and N. J. Krogan. Differential network biology. *Molecular Systems Biology*, 8(1), 2012.

[117] F. Iorio, R. Bosotti, E. Scacheri, V. Belcastro, P. Mithbaokar, R. Ferriero, L. Murino, R. Tagliaferri, N. Brunetti-Pierri, A. Isacchi, and D. di Bernardo. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci U S A*, 107(33):14621–14626, Aug 2010.

[118] J. F. Islas, Y. Liu, K.-C. Weng, M. J. Robertson, S. Zhang, A. Prejusa, J. Harger, D. Tikhomirova, M. Chopra, D. Iyer, M. Mercola, R. G. Oshima, J. T. Willerson, V. N.

Potaman, and R. J. Schwartz. Transcription factors ets2 and mesp1 transdifferentiate human dermal fibroblasts into cardiac progenitors. *Proc Natl Acad Sci U S A*, 109(32):13016–13021, Aug 2012.

[119] A. Iwata, S. Tsuji, K. Iwamoto, K. Nagata, T. Saido, H. Hatsuta, S. Murayama, A. Tamaoka, H. Takuma, and M. Bundo. Altered CpG methylation in sporadic Alzheimer's disease is associated with APP and MAPT dysregulation. *Human Molecular Genetics*, 23(3):648–656, 09 2013.

[120] A. E. Jaffe, Y. Gao, R. Tao, T. M. Hyde, D. R. Weinberger, and J. E. Kleinman. The methylome of the human frontal cortex across development. *bioRxiv*, 2014.

[121] N. Jantaratnotai, A. Ling, J. Cheng, C. Schwab, P. McGeer, and J. McLarnon. Upregulation and expression patterns of the angiogenic transcription factor ets-1 in alzheimer's disease brain. *J Alzheimers Dis*, 37(2):367 – 77, 2013.

[122] H. K. Jin, J. E. Carter, G. W. Huntley, and E. H. Schuchman. Intracerebral transplantation of mesenchymal stem cells into acid sphingomyelinase–deficient mice delays the onset of neurological abnormalities and extends their life span. *The Journal of Clinical Investigation*, 109(9):1183–1191, 5 2002.

[123] S. J. Joggerst and A. K. Hatzopoulos. Stem cell therapy for cardiac repair: benefits and barriers. *Expert Reviews in Molecular Medicine*, 11:e20, 2009.

[124] P. A. Jones. Functions of dna methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, 13(7):484–492, Jul 2012.

[125] C. D. Jonghe, C. W. Esselens, S. Kumar-Singh, K. Craessaerts, S. Serneels, F. Checler, W. Annaert, C. V. Broeckhoven, and B. D. Strooper. Pathogenic app mutations near the gamma-secretase cleavage site differentially affect abeta secretion and app c-terminal fragment stability. *Human molecular genetics*, 10 16:1665–71, 2001.

[126] P. F. Jonsson and P. A. Bates. Global topological features of cancer proteins in the human interactome. *Bioinformatics*, 22(18):2291–2297, Sep 2006.

[127] S. Jung, A. Hartmann, and A. del Sol. Refbool: a reference-based algorithm for discretizing gene expression data. *Bioinformatics*, 33(13):1953–1962, 2017.

[128] L. S. Kaltenbach, E. Romero, R. R. Becklin, R. Chettier, R. Bell, A. Phansalkar, A. Strand, C. Torcassi, J. Savage, A. Hurlburt, G.-H. Cha, L. Ukani, C. L. Chepanoske, Y. Zhen, S. Sahasrabudhe, J. Olson, C. Kurschner, L. M. Ellerby, J. M. Peltier, J. Botas, and R. E. Hughes. Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genet*, 3(5), 2007.

[129] U. S. Kamaraj, J. Gough, J. M. Polo, E. Petretto, and O. J. L. Rackham. Computational methods for direct cell conversion. *Cell Cycle*, 15(24):3343–3354, 2016.

[130] R. J. r. Kelleher and J. Shen. Presenilin-1 mutations and alzheimer's disease. *Proc Natl Acad Sci U S A*, 114(4):629–631, Jan 2017.

[131] D. A. Khavari, G. L. Sen, and J. L. Rinn. Dna methylation and epigenetic control of cellular differentiation. *Cell Cycle*, 9(19):3880–3883, 2010.

[132] J. B. Kim, B. Greber, M. J. Araúzo-Bravo, J. Meyer, K. I. Park, H. Zaehres, and H. R. Schöler. Direct reprogramming of human neural stem cells by oct4. *Nature*, 461:649 EP –, Aug 2009.

[133] S. Y. Kim, J.-S. Lim, I. G. Kong, and H. G. Choi. Hearing impairment and the risk of neurodegenerative dementia: A longitudinal follow-up study using a national sample cohort. *Scientific Reports*, 8(1):15266, 2018.

[134] K. Klein and S. Gay. Epigenetics in rheumatoid arthritis. *Current Opinion in Rheumatology*, 27(1), 2015.

[135] M. Ko, H. S. Bandukwala, J. An, E. D. Lamperti, E. C. Thompson, R. Hastie, A. Tsangaratou, K. Rajewsky, S. B. Koralov, and A. Rao. Ten-eleven-translocation 2 (tet2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proceedings of the National Academy of Sciences*, 108(35):14566–14571, 2011.

[136] K. Kobayashi and K. Hiraishi. Verification and optimal control of context-sensitive probabilistic boolean networks using model checking and polynomial optimization. *Scientific-WorldJournal*, 2014:968341–968341, Jan 2014.

[137] R. P. Koche, Z. D. Smith, M. Adli, H. Gu, M. Ku, A. Gnirke, B. E. Bernstein, and A. Meissner. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell*, 8(1):96 – 105, 2011.

[138] A. KOHLSCHUTTER, R. LAABS, and M. ALBANI. Juvenile neuronal ceroid lipofuscinosis (jncl): Quantitative description of its clinical variability. *Acta Paediatrica*, 77(6):867–872, 1988.

[139] T. Kouzarides. Chromatin modifications and their function. *Cell*, 128(4):693 – 705, 2007.

[140] S. Kumar, J. Blangero, and J. E. Curran. *Induced Pluripotent Stem Cells in Disease Modeling and Gene Identification*, pages 17–38. Springer New York, New York, NY, 2018.

[141] M. Kutmon, M. P. van Iersel, A. Bohler, T. Kelder, N. Nunes, A. R. Pico, and C. T. Evelo. Pathvisio 3: an extendable pathway analysis toolbox. *PLoS Comput Biol*, 11(2):e1004085–e1004085, Feb 2015.

[142] M. Kwiatkowska, G. Norman, and D. Parker. Prism: Probabilistic symbolic model checker. *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 200–204, 2002.

[143] K. Lage, E. O. Karlberg, Z. M. Størling, P. Í. Ólason, A. G. Pedersen, O. Rigina, A. M. Hinsby, Z. Tümer, F. Pociot, N. Tommerup, Y. Moreau, and S. Brunak. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology*, 25:309 EP –, Mar 2007.

[144] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935, 2006.

[145] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The

connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935, 2006.

[146] J.-C. Lambert, S. Heath, G. Even, D. Campion, K. Sleegers, M. Hiltunen, O. Combarros, D. Zelenika, M. J. Bullido, B. Tavernier, L. Letenneur, K. Bettens, C. Berr, F. Pasquier, N. Fiévet, P. Barberger-Gateau, S. Engelborghs, P. De Deyn, I. Mateo, A. Franck, S. Helisalmi, E. Porcellini, O. Hanon, t. E. A. D. I. Investigators, M. M. de Pancorbo, C. Lendon, C. Dufouil, C. Jaillard, T. Leveillard, V. Alvarez, P. Bosco, M. Mancuso, F. Panza, B. Nacmias, P. Bossù, P. Piccardi, G. Annoni, D. Seripa, D. Galimberti, D. Hannequin, F. Licastro, H. Soininen, K. Ritchie, H. Blanché, J.-F. Dartigues, C. Tzourio, I. Gut, C. Van Broeckhoven, A. Alpérovitch, M. Lathrop, and P. Amouyel. Genome-wide association study identifies variants at clu and cr1 associated with alzheimer&#39;s disease. *Nature Genetics*, 41:1094 EP –, Sep 2009.

[147] J. C. Lambert, C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, A. L. DeStafano, J. C. Bis, G. W. Beecham, B. Grenier-Boley, G. Russo, T. A. Thorton-Wells, N. Jones, A. V. Smith, V. Chouraki, C. Thomas, M. A. Ikram, D. Zelenika, B. N. Vardarajan, Y. Kamatani, C. F. Lin, A. Gerrish, H. Schmidt, B. Kunkle, M. L. Dunstan, A. Ruiz, M. T. Bihoreau, S. H. Choi, C. Reitz, F. Pasquier, C. Cruchaga, D. Craig, N. Amin, C. Berr, O. L. Lopez, P. L. De Jager, V. Deramecourt, J. A. Johnston, D. Evans, S. Lovestone, L. Letenneur, F. J. Morón, D. C. Rubinsztein, G. Eiriksdottir, K. Sleegers, A. M. Goate, N. Fiévet, M. W. Huentelman, M. Gill, K. Brown, M. I. Kamboh, L. Keller, P. Barberger-Gateau, B. McGuiness, E. B. Larson, R. Green, A. J. Myers, C. Dufouil, S. Todd, D. Wallon, S. Love, E. Rogaeva, J. Gallacher, P. St George-Hyslop, J. Clarimon, A. Lleo, A. Bayer, D. W. Tsuang, L. Yu, M. Tsolaki, P. Bossù, G. Spalletta, P. Proitsi, J. Collinge, S. Sorbi, F. Sanchez-Garcia, N. C. Fox, J. Hardy, M. C. Deniz Naranjo, P. Bosco, R. Clarke, C. Brayne, D. Galimberti, M. Mancuso, F. Matthews, E. A. D. I. (EADI), G. Disease, E. R. in Alzheimer's, A. D. G. Consortium, C. f. H. Epidemiology, A. R. in Genomic, S. Moebus, P. Mecocci, M. Del Zompo, W. Maier, H. Hampel, A. Pilotto, M. Bullido, F. Panza, P. Caffarra, B. Nacmias, J. R. Gilbert, M. Mayhaus, L. Lannefelt, H. Hakonarson, S. Pichler, M. M. Carrasquillo, M. Ingelsson, D. Beekly, V. Alvarez, F. Zou, O. Valladares, S. G. Younkin, E. Coto, K. L. Hamilton-Nelson, W. Gu, C. Razquin, P. Pastor,

I. Mateo, M. J. Owen, K. M. Faber, P. V. Jonsson, O. Combarros, M. C. O'Donovan, L. B. Cantwell, H. Soininen, D. Blacker, S. Mead, T. H. J. Mosley, D. A. Bennett, T. B. Harris, L. Fratiglioni, C. Holmes, R. F. de Bruijn, P. Passmore, T. J. Montine, K. Bettens, J. I. Rotter, A. Brice, K. Morgan, T. M. Foroud, W. A. Kukull, D. Hannequin, J. F. Powell, M. A. Nalls, K. Ritchie, K. L. Lunetta, J. S. Kauwe, E. Boerwinkle, M. Riemenschneider, M. Boada, M. Hiltuenen, E. R. Martin, R. Schmidt, D. Rujescu, L. S. Wang, J. F. Dartigues, R. Mayeux, C. Tzourio, A. Hofman, M. M. Nöthen, C. Graff, B. M. Psaty, L. Jones, J. L. Haines, P. A. Holmans, M. Lathrop, M. A. Pericak-Vance, L. J. Launer, L. A. Farrer, C. M. van Duijn, C. Van Broeckhoven, V. Moskvina, S. Seshadri, J. Williams, G. D. Schellenberg, and P. Amouyel. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for alzheimer's disease. *Nat Genet*, 45(12):1452–1458, Dec 2013.

[148] M. A. Lancaster and J. A. Knoblich. Generation of cerebral organoids from human pluripotent stem cells. *Nat Protoc*, 9(10):2329–2340, Oct 2014.

[149] M. A. Lancaster, M. Renner, C.-A. Martin, D. Wenzel, L. S. Bicknell, M. E. Hurles, T. Homfray, J. M. Penninger, A. P. Jackson, and J. A. Knoblich. Cerebral organoids model human brain development and microcephaly. *Nature*, 501(7467):373–379, Sep 2013.

[150] J. M. LaSalle, W. T. Powell, and D. H. Yasui. Epigenetic layers and players underlying neurodevelopment. *Trends Neurosci*, 36(8):460–470, Aug 2013.

[151] C.-I. Lau, D. C. Yánez, A. Solanki, E. Papaioannou, J. I. Saldaña, and T. Crompton. Foxa1 and foxa2 in thymic epithelial cells (tec) regulate medullary tec and regulatory t-cell maturation. *Journal of Autoimmunity*, 93:131 – 138, 2018.

[152] O. A. Ledyankina and S. A. Mikhailov. Composite model of a research flight simulator for a helicopter with the hingeless main rotor. *Russian Aeronautics*, 59(4):495–499, 2016.

[153] J.-E. Lee and K. Ge. Transcriptional and epigenetic regulation of pparγ expression during adipogenesis. *Cell & Bioscience*, 4(1):29, May 2014.

[154] J. Y. Lee, S. H. Han, M. H. Park, B. Baek, I.-S. Song, M.-K. Choi, Y. Takuwa, H. Ryu, S. H. Kim, X. He, E. H. Schuchman, J.-S. Bae, and H. K. Jin. Neuronal sphk1 acetylates cox2

and contributes to pathogenesis in a model of alzheimer's disease. *Nature Communications*, 9(1):1479, 2018.

[155] M. D. M. Leiserson, F. Vandin, H.-T. Wu, J. R. Dobson, J. V. Eldridge, J. L. Thomas, A. Papoutsaki, Y. Kim, B. Niu, M. McLellan, M. S. Lawrence, A. Gonzalez-Perez, D. Tamborero, Y. Cheng, G. A. Ryslik, N. Lopez-Bigas, G. Getz, L. Ding, and B. J. Raphael. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet*, 47(2):106–114, Feb 2015.

[156] A.-M. Lepagnol-Bestel, G. Maussion, M. Simonneau, P. Gorwood, J.-M. Moalic, Y. Loe-Mie, A. Doron-Faigenboim, T. Pupko, H. Delacroix, L. Aggerbeck, and S. Imbeaud. SMARCA2 and other genome-wide supported schizophrenia-associated genes: regulation by REST/NRSF, network organization and primate-specific evolution. *Human Molecular Genetics*, 19(14):2841–2857, 05 2010.

[157] B. Li, M. Carey, and J. L. Workman. The role of chromatin during transcription. *Cell*, 128(4):707–719, Feb 2007.

[158] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and . G. P. D. P. Subgroup. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, Aug 2009.

[159] M. Li, Y. He, W. Dubois, X. Wu, J. Shi, and J. Huang. Distinct regulatory mechanisms and functions for p53-activated and p53-repressed dna damage response genes in embryonic stem cells. *Molecular Cell*, 46(1):30 – 42, 2012.

[160] X. Li, S. Ottosson, S. Wang, E. Jernberg, L. Boldrup, X. Gu, K. Nylander, and A. Li. Wilms' tumor gene 1 regulates p63 and promotes cell proliferation in squamous cell carcinoma of the head and neck. *BMC Cancer*, 15:342–342, May 2015.

[161] Z. Li, C. Liu, Z. Xie, P. Song, R. C. H. Zhao, L. Guo, Z. Liu, and Y. Wu. Epigenetic dysregulation in mesenchymal stem cell aging and spontaneous differentiation. *PLOS ONE*, 6(6):1–9, 06 2011.

[162] W. Liao, J. Xie, J. Zhong, Y. Liu, L. Du, B. Zhou, J. Xu, P. Liu, S. Yang, J. Wang, Z. Han,

and Z. C. Han. Therapeutic effect of human umbilical cord multipotent mesenchymal stromal cells in a rat model of stroke. *Transplantation*, 87(3), 2009.

[163] A. Lihu and S. Holban. A review of ensemble methods for de novo motif discovery in ChIP-Seq data. *Briefings in Bioinformatics*, 16(6):964–973, 04 2015.

[164] D. H. K. Lim and E. R. Maher. Dna methylation: a form of epigenetic control of gene expression. *The Obstetrician & Gynaecologist*, 12(1):37–42, 2010.

[165] J. Lim, T. Hao, C. Shaw, A. J. Patel, G. Szabó, J.-F. Rual, C. J. Fisk, N. Li, A. Smolyar, D. E. Hill, A.-L. Barabási, M. Vidal, and H. Y. Zoghbi. A protein&#x2013;protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration. *Cell*, 125(4):801–814, 2006.

[166] X. Lim and R. Nusse. Wnt signaling in skin development, homeostasis, and disease. *Cold Spring Harb Perspect Biol*, 5(2):a008029, 2013.

[167] F. Liu, S. Kohlmeier, and C.-Y. Wang. Wnt signaling and skeletal development. *Cell Signal*, 20(6):999–1009, Jun 2008.

[168] L. Lonka, A. Aalto, O. Kopra, M. Kuronen, Z. Kokaia, M. Saarma, and A.-E. Lehesjoki. The neuronal ceroid lipofuscinosis cln8 gene expression is developmentally regulated in mouse brain and up-regulated in the hippocampal kindling model of epilepsy. *BMC Neurosci*, 6:27–27, Apr 2005.

[169] M. A. Lovell, C. Xie, S. Xiong, and W. R. Markesbery. Wilms' tumor suppressor (wt1) is a mediator of neuronal degeneration associated with the pathogenesis of alzheimer's disease. *Brain Research*, 983(1):84 – 96, 2003.

[170] K. Lunnon, R. Smith, E. Hannon, P. L. De Jager, G. Srivastava, M. Volta, C. Troakes, S. Al-Sarraj, J. Burrage, R. Macdonald, D. Condliffe, L. W. Harries, P. Katsel, V. Haroutunian, Z. Kaminsky, C. Joachim, J. Powell, S. Lovestone, D. A. Bennett, L. C. Schalkwyk, and J. Mill. Methylomic profiling implicates cortical deregulation of ank1 in alzheimer disease. *Nature Neuroscience*, 17:1164 EP –, Aug 2014.

[171] M. Maceyka, K. B. Harikumar, S. Milstien, and S. Spiegel. Sphingosine-1-phosphate signaling and its role in disease. *Trends Cell Biol*, 22(1):50–60, Jan 2012.

[172] J. S. Malamon and A. Kriete. Integrated systems approach reveals sphingolipid metabolism pathway dysregulation in association with late-onset alzheimer's disease. *Biology (Basel)*, 7(1):16, Feb 2018.

[173] N. Malik and M. S. Rao. A review of the methods for human ipsc derivation. *Methods Mol Biol*, 997:23–33, 2013.

[174] A. A. Mangi, N. Noiseux, D. Kong, H. He, M. Rezvani, J. S. Ingwall, and V. J. Dzau. Mesenchymal stem cells modified with akt prevent remodeling and restore performance of infarcted hearts. *Nature Medicine*, 9:1195 EP –, Aug 2003.

[175] D. Marbach, D. Lamparter, G. Quon, M. Kellis, Z. Kutalik, and S. Bergmann. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat Meth*, 13(4):366–370, Apr 2016.

[176] G. K. Marinov, A. Kundaje, P. J. Park, and B. J. Wold. Large-scale quality analysis of published chip-seq data. *G3: Genes, Genomes, Genetics*, 4(2):209–223, 2014.

[177] E. Martin, M. Amar, C. Dalle, I. Youssef, C. Boucher, C. Le Duigou, M. Brückner, A. Prigent, V. Sazdovitch, A. Halle, J. M. Kanellopoulos, B. Fontaine, B. Delatour, and C. Delarasse. New role of p2x7 receptor in an alzheimer's disease mouse model. *Molecular Psychiatry*, 24(1):108–125, 2019.

[178] G. Martino, R. J. M. Franklin, A. B. Van Evercooren, D. A. Kerr, and t. S. C. i. M. S. S. C. Group. Stem cell transplantation in multiple sclerosis: current status and future prospects. *Nature Reviews Neurology*, 6:247 EP –, Apr 2010.

[179] S. J. Marzi, S. K. Leung, T. Ribarska, E. Hannon, A. R. Smith, E. Pishva, J. Poschmann, K. Moore, C. Troakes, S. Al-Sarraj, S. Beck, S. Newman, K. Lunnon, L. C. Schalkwyk, and J. Mill. A histone acetylome-wide association study of alzheimer's disease identifies disease-associated h3k27ac differences in the entorhinal cortex. *Nature Neuroscience*, 21(11):1618–1627, 2018.

[180] R. Mashima and T. Okuyama. The role of lipoxygenases in pathophysiology; new insights and future perspectives. *Redox Biol*, 6:297–310, Aug 2015.

[181] T. Matsui, M. Ingelsson, H. Fukumoto, K. Ramasamy, H. Kowa, M. P. Frosch, M. C. Irizarry, and B. T. Hyman. Expression of app pathway mrnas and proteins in alzheimer's disease. *Brain Research*, 1161:116 – 123, 2007.

[182] I. Maze, L. Shen, B. Zhang, B. A. Garcia, N. Shao, A. Mitchell, H. Sun, S. Akbarian, C. D. Allis, and E. J. Nestler. Analytical tools and current challenges in the modern era of neuroepigenomics. *Nature Neuroscience*, 17:1476 EP –, Oct 2014.

[183] D. G. McFadden, A. C. Barbosa, J. A. Richardson, M. D. Schneider, D. Srivastava, and E. N. Olson. The hand1 and hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development*, 132(1):189–201, 2005.

[184] F. Meda, M. Folci, A. Baccarelli, and C. Selmi. The epigenetics of autoimmunity. *Cell Mol Immunol*, 8(3):226–236, May 2011.

[185] J. Medvedovic, A. Ebert, H. Tagoh, and M. Busslinger. Pax5 a master regulator of b cell development and leukemogenesis. volume 111 of *Advances in Immunology*, pages 179 – 206. Academic Press, 2011.

[186] S. Mei, Q. Qin, Q. Wu, H. Sun, R. Zheng, C. Zang, M. Zhu, J. Wu, X. Shi, L. Taing, T. Liu, M. Brown, C. A. Meyer, and X. S. Liu. Cistrome data browser: a data portal for chip-seq and chromatin accessibility data in human and mouse. *Nucleic Acids Research*, 45(D1):D658–D662, 2017.

[187] X. Meng, A. Neises, R.-J. Su, K. J. Payne, L. Ritter, D. S. Gridley, J. Wang, M. Sheng, K.-H. W. Lau, D. J. Baylink, and X.-B. Zhang. Efficient reprogramming of human cord blood cd34+ cells into induced pluripotent stem cells with oct4 and sox2 alone. *Molecular Therapy*, 20(2):408 – 416, 2012.

[188] P. Merlo, B. Frost, S. Peng, Y. J. Yang, P. J. Park, and M. Feany. p53 prevents neurodegeneration by regulating synaptic genes. *Proc Natl Acad Sci U S A*, 111(50):18055–18060, Dec 2014.

[189] M. M. Mielke, N. J. Haughey, V. V. R. Bandaru, H. Zetterberg, K. Blennow, U. Andreasson, S. C. Johnson, C. E. Gleason, H. M. Blazel, L. Puglielli, M. A. Sager, S. Asthana, and

C. M. Carlsson. Cerebrospinal fluid sphingolipids, b-amyloid, and tau in adults at risk for alzheimer's disease. *Neurobiol Aging*, 35(11):2486–2494, Nov 2014.

[190] M. M. Mielke and C. G. Lyketsos. Alterations of the sphingolipid pathway in alzheimer's disease: new biomarkers and treatment targets? *Neuromolecular Med*, 12(4):331–340, Dec 2010.

[191] T. S. Mikkelsen, J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B. E. Bernstein, R. Jaenisch, E. S. Lander, and A. Meissner. Dissecting direct reprogramming through integrative genomic analysis. *Nature*, 454:49 EP –, May 2008.

[192] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[193] J. Min Lee, E. P. Gianchandani, J. A. Eddy, and J. A. Papin. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLOS Computational Biology*, 4(5):1–20, 05 2008.

[194] J. S. Miners, P. Clarke, and S. Love. Clusterin levels are increased in alzheimer's disease and influence the regional distribution of aβ. *Brain Pathology*, 27(3):305–313, 2017.

[195] K. Mitra, A.-R. Carvunis, S. K. Ramesh, and T. Ideker. Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet*, 14(10):719–732, Oct 2013.

[196] K. Mizukami, D. R. Grayson, M. D. Ikonomovic, R. Sheffield, and D. M. Armstrong. Gabaa receptor β2 and β3 subunits mrna in the hippocampal formation of aged human brain with alzheimer-related neuropathology. *Molecular Brain Research*, 56(1):268 – 272, 1998.

[197] V. Moignard, S. Woodhouse, L. Haghverdi, A. J. Lilly, Y. Tanaka, A. C. Wilkinson, F. Buettner, I. C. Macaulay, W. Jawaid, E. Diamanti, S.-I. Nishikawa, N. Piterman, V. Kouskoff, F. J. Theis, J. Fisher, and B. Gottgens. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat Biotech*, 33(3):269–276, Mar 2015.

[198] S. Morris, P. Cahan, H. Li, A. Zhao, A. San?Roman, R. Shivdasani, J. Collins, and G. Daley. Dissecting engineered cell types and enhancing cell fate conversion via cellnet. *Cell*, 158(4):889–902, Aug 2014.

[199] S. A. Morris and G. Q. Daley. A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell Research*, 23:33 EP –, Jan 2013.

[200] E. J. Mufson, S. E. Counts, S. E. Perez, and S. D. Ginsberg. Cholinergic system during the progression of alzheimer's disease: therapeutic implications. *Expert Rev Neurother*, 8(11):1703–1718, Nov 2008.

[201] A. Musa, L. S. Ghoraie, S.-D. Zhang, G. Glazko, O. Yli-Harja, M. Dehmer, B. Haibe-Kains, and F. Emmert-Streib. A review of connectivity map and computational approaches in pharmacogenomics. *Brief Bioinform*, 19(3):506–523, Jan 2017.

[202] M. Nakagawa, M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, and S. Yamanaka. Generation of induced pluripotent stem cells without myc from mouse and human fibroblasts. *Nature Biotechnology*, 26:101 EP –, Nov 2007.

[203] A. Natarajan, G. G. Yardimci, N. C. Sheffield, G. E. Crawford, and U. Ohler. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res*, 22(9):1711–1722, Sep 2012.

[204] S. Neph, A. B. Stergachis, A. Reynolds, R. Sandstrom, E. Borenstein, and J. A. Stamatoyannopoulos. Circuitry and dynamics of human transcription factor regulatory networks. *Cell*, 150(6):1274–1286, 2017/08/24 2012.

[205] T. C. G. A. R. Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455:1061 EP –, Sep 2008.

[206] A. Nikitin, S. Egorov, N. Daraselia, and I. Mazo. Pathway studio—the analysis and navigation of molecular networks. *Bioinformatics*, 19(16):2155–2157, 2003.

[207] K. Nikouei, A. B. Muñoz-Manchado, and J. Hjerling-Leffler. Bcl11b/ctip2 is highly expressed in gabaergic interneurons of the mouse somatosensory cortex. *Journal of Chemical Neuroanatomy*, 71:1 – 5, 2016.

[208] R. J. O'Brien and P. C. Wong. Amyloid precursor protein processing and alzheimer's disease. *Annu Rev Neurosci*, 34:185–204, 2011.

[209] D. T. Odom, R. D. Dowell, E. S. Jacobsen, L. Nekludova, P. A. Rolfe, T. W. Danford, D. K. Gifford, E. Fraenkel, G. I. Bell, and R. A. Young. Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol*, 2:2006.0017–2006.0017, May 2006.

[210] M. Ohnuki and K. Takahashi. Present and future challenges of induced pluripotent stem cells. *Philos Trans R Soc Lond B Biol Sci*, 370(1680):20140367–20140367, Oct 2015.

[211] S. Okawa, V. E. Angarica, I. Lemischka, K. Moore, and A. del Sol. A differential network analysis approach for lineage specifier prediction in stem cell subpopulations. *NPJ Syst Biol Appl*, 1:15012, 2015.

[212] A. S. B. Olsen and N. J. Færgeman. Sphingolipids: membrane microdomains in brain development, function and neurological diseases. *Open Biol*, 7(5):170069, May 2017.

[213] J. M. Olson, A. Asakura, L. Snider, R. Hawkes, A. Strand, J. Stoeck, A. Hallahan, J. Pritchard, and S. J. Tapscott. Neurod2 is necessary for development and survival of central nervous system neurons. *Developmental Biology*, 234(1):174 – 187, 2001.

[214] S. K. T. Ooi, A. H. O'Donnell, and T. H. Bestor. Mammalian cytosine methylation at a glance. *Journal of Cell Science*, 122(16):2787–2791, 2009.

[215] H. Osada, G. Grutz, H. Axelson, A. Forster, and T. H. Rabbitts. Association of erythroid transcription factors: complexes involving the lim protein rbtn2 and the zinc-finger protein gata1. *Proc Natl Acad Sci U S A*, 92(21):9585–9589, Oct 1995.

[216] J. R. Ostergaard. Juvenile neuronal ceroid lipofuscinosis (batten disease): current insights. *Degener Neurol Neuromuscul Dis*, 6:73–83, Aug 2016.

[217] F. Ozsolak and P. M. Milos. Rna sequencing: advances, challenges and opportunities. *Nat Rev Genet*, 12(2):87–98, Feb 2011.

[218] Z. P. Pang, N. Yang, T. Vierbuchen, A. Ostermeier, D. R. Fuentes, T. Q. Yang, A. Citri, V. Sebastiano, S. Marro, T. C. Südhof, and M. Wernig. Induction of human neuronal cells by defined transcription factors. *Nature*, 476:220 EP –, May 2011.

[219] B. Papp and K. Plath. Reprogramming to pluripotency: stepwise resetting of the epigenetic landscape. *Cell Res*, 21(3):486–501, Mar 2011.

[220] E. O. Paull, D. E. Carlin, M. Niepel, P. K. Sorger, D. Haussler, and J. M. Stuart. Discovering causal pathways linking genomic events to transcriptional states using tied diffusion through interacting events (tiedie). *Bioinformatics*, 29(21):2757–2764, Nov 2013.

[221] M. Paulsen and A. C. Ferguson-Smith. Dna methylation in genomic imprinting, development, and disease. *The Journal of Pathology*, 195(1):97–110, 2001.

[222] A. L. Penton, L. D. Leonard, and N. B. Spinner. Notch signaling in human development and disease. *Semin Cell Dev Biol*, 23(4):450–457, Jun 2012.

[223] R. C. Perier, V. Praz, T. Junier, C. Bonnard, and P. Bucher. The eukaryotic promoter database (epd). *Nucleic Acids Res*, 28(1):302–303, Jan 2000.

[224] R. Pidsley, C. C. Y Wong, M. Volta, K. Lunnon, J. Mill, and L. C. Schalkwyk. A data-driven approach to preprocessing illumina 450k methylation array data. *BMC Genomics*, 14:293–293, May 2013.

[225] J. E. Pimanda, K. Ottersbach, K. Knezevic, S. Kinston, W. Y. I. Chan, N. K. Wilson, J.-R. Landry, A. D. Wood, A. Kolb-Kokocinski, A. R. Green, D. Tannahill, G. Lacaud, V. Kouskoff, and B. Göttgens. Gata2, fli1, and scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proceedings of the National Academy of Sciences*, 104(45):17692–17697, 2007.

[226] M. F. Pittenger and B. J. Martin. Mesenchymal stem cells and their potential as cardiac therapeutics. *Circulation Research*, 95(1):9–20, 2004.

[227] E. Plahte, T. Mestl, and S. W. Omholt. Feedback loops, stability and multistationarity in dynamical systems. *Journal of Biological Systems*, 03(02):409–413, 1995.

[228] I. Plangár, D. Zádori, P. Klivényi, J. Toldi, and L. Vécsei. Targeting the kynurenine pathway-related alterations in alzheimer's disease: a future therapeutic strategy. *Journal of Alzheimer's disease : JAD*, 24 Suppl 2:199–209, 2011.

[229] A. R. Poetsch and C. Plass. Transcriptional regulation by dna methylation. *Cancer Treatment Reviews*, 37:S8–S12, Jan 2011.

[230] C. Porcher, H. Chagraoui, and M. S. Kristiansen. Scl/tal1: a multifaceted regulator from blood development to disease. *Blood*, 129(15):2051–2060, 2017.

[231] M. A. Pujana, J.-D. J. Han, L. M. Starita, K. N. Stevens, M. Tewari, J. S. Ahn, G. Rennert, V. Moreno, T. Kirchhoff, B. Gold, V. Assmann, W. M. ElShamy, J.-F. Rual, D. Levine, L. S. Rozek, R. S. Gelman, K. C. Gunsalus, R. A. Greenberg, B. Sobhian, N. Bertin, K. Venkatesan, N. Ayivi-Guedehoussou, X. Solé, P. Hernández, C. Lázaro, K. L. Nathanson, B. L. Weber, M. E. Cusick, D. E. Hill, K. Offit, D. M. Livingston, S. B. Gruber, J. D. Parvin, and M. Vidal. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature Genetics*, 39:1338 EP –, Oct 2007.

[232] O. J. L. Rackham, J. Firas, H. Fang, M. E. Oates, M. L. Holmes, A. S. Knaupp, T. F. Consortium, H. Suzuki, C. M. Nefzger, C. O. Daub, J. W. Shin, E. Petretto, A. R. R. Forrest, Y. Hayashizaki, J. M. Polo, and J. Gough. A predictive computational framework for direct reprogramming between human cell types. *Nature Genetics*, 48:331 EP –, Jan 2016.

[233] J. Radke, R. Koll, E. Gill, L. Wiese, A. Schulz, A. Kohlschütter, M. Schuelke, C. Hagel, W. Stenzel, and H. H. Goebel. Autophagic vacuolar myopathy is a common feature of cln3 disease. *Ann Clin Transl Neurol*, 5(11):1385–1393, Oct 2018.

[234] E. Ramadan, S. Alinsaif, and M. R. Hassan. Network topology measures for identifying disease-gene association in breast cancer. *BMC Bioinformatics*, 17(7):274, Jul 2016.

[235] O. J. Rando. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Current Opinion in Genetics & Development*, 22(2):148 – 155, 2012.

[236] J. Raskin, J. Cummings, J. Hardy, K. Schuh, and R. A. Dean. Neurobiology of alzheimer's disease: Integrated molecular, physiological, anatomical, biomarker, and cognitive dimensions. *Curr Alzheimer Res*, 12(8):712–722, Oct 2015.

[237] S. Razick, G. Magklaras, and I. M. Donaldson. irefindex: A consolidated protein interaction database with provenance. *BMC Bioinformatics*, 9(1):405, Sep 2008.

[238] A. Razin and B. Kantor. *DNA Methylation in Epigenetic Control of Gene Expression*, pages 151–167. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[239] M. Rentzos, M. Zoga, G. P. Paraskevas, E. Kapaki, A. Rombos, C. Nikolaou, A. Tsoutsou, and D. Vassilopoulos. Il-15 is elevated in cerebrospinal fluid of patients with alzheimer's disease and frontotemporal dementia. *Journal of Geriatric Psychiatry and Neurology*, 19(2):114–117, 2006.

[240] R. A. Rissman and W. C. Mobley. Implications for treatment: Gabaa receptors in aging, down syndrome and alzheimer's disease. *J Neurochem*, 117(4):613–622, May 2011.

[241] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res*, 43(7):e47–e47, Apr 2015.

[242] A. Rizzino. The sox2-oct4 connection: Critical players in a much larger interdependent network integrated at multiple levels. *Stem Cells*, 31(6):1033–1039, Jun 2013.

[243] S. Ronquist, G. Patterson, L. A. Muir, S. Lindsly, H. Chen, M. Brown, M. S. Wicha, A. Bloch, R. Brockett, and I. Rajapakse. Algorithm for cellular reprogramming. *Proceedings of the National Academy of Sciences*, 2017.

[244] M. S. Roost, R. C. Slieker, M. Bialecka, L. van Iperen, M. M. Gomes Fernandes, N. He, H. E. D. Suchiman, K. Szuhai, F. Carlotti, E. J. P. de Koning, C. L. Mummery, B. T. Heijmans, and S. M. Chuva de Sousa Lopes. Dna methylation and transcriptional trajectories during human development and reprogramming of isogenic pluripotent stem cells. *Nature Communications*, 8(1):908, 2017.

[245] V. Saint-André, A. J. Federation, C. Y. Lin, B. J. Abraham, J. Reddy, T. I. Lee, J. E. Bradner, and R. A. Young. Models of human core transcriptional regulatory circuitries. *Genome Res*, 26(3):385–396, Mar 2016.

[246] A.-E. Saliba, A. J. Westermann, S. A. Gorski, and J. Vogel. Single-cell rna-seq: advances and future challenges. *Nucleic Acids Res*, 42(14):8845–8860, Aug 2014.

[247] I. Sancho-Martinez, S. H. Baek, and J. C. Izpisua Belmonte. Lineage conversion methodologies meet the reprogramming toolbox. *Nature Cell Biology*, 14:892 EP –, Sep 2012.

[248] M. Sara, W. Ruth, and H. Goebel. *The Neuronal Ceroid Lipofuscinoses (Batten Disease) (2 ed.)*, chapter CLN3, pages 295–324. Oxford University Press, 2011.

[249] V. Scharnhorst, P. Dekker, A. J. van der Eb, and A. G. Jochemsen. Physical interaction between wilms tumor 1 and p73 proteins modulates their functions. *Journal of Biological Chemistry*, 275(14):10202–10211, 2000.

[250] Y. Schirer, A. Malishkevich, Y. Ophir, J. Lewis, E. Giladi, and I. Gozes. Novel marker for the onset of frontotemporal dementia: early increase in activity-dependent neuroprotective protein (adnp) in the face of tau mutation. *PLoS One*, 9(1):e87383–e87383, Jan 2014.

[251] B. L. Schwartz, S. Hashtroudi, R. L. Herting, P. Schwartz, and S. I. Deutsch. d-cycloserine enhances implicit memory in alzheimer patients. *Neurology*, 46 2:420–4, 1996.

[252] H. E. Seberg, E. Van Otterloo, and R. A. Cornell. Beyond mitf: Multiple transcription factors directly regulate the cellular phenotype in melanocytes and melanoma. *Pigment Cell & Melanoma Research*, 30(5):454–466, 2017.

[253] G. L. Sen, J. A. Reuter, D. E. Webster, L. Zhu, and P. A. Khavari. Dnmt1 maintains progenitor function in self-renewing somatic tissue. *Nature*, 463(7280):563–567, Jan 2010.

[254] H. Shimizu, M. Ghazizadeh, S. Sato, T. Oguro, and O. Kawanami. Interaction between beta-amyloid protein and heparan sulfate proteoglycans from the cerebral capillary basement membrane in alzheimer's disease. *Journal of Clinical Neuroscience*, 16(2):277–282, Feb 2009.

[255] M. K. Singh, M. Petry, B. Haenig, B. Lescher, M. Leitges, and A. Kispert. The t-box transcription factor tbx15 is required for skeletal development. *Mechanisms of Development*, 122(2):131 – 144, 2005.

[256] V. K. Singh, M. Kalsan, N. Kumar, A. Saini, and R. Chandra. Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. *Frontiers in Cell and Developmental Biology*, 3:2, 2015.

[257] D. N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Mélius, E. Cirillo, S. L. Coort, D. Digles, F. Ehrhart, P. Giesbertz, M. Kalafati, M. Martens, R. Miller, K. Nishida, L. Rieswijk, A. Waagmeester, L. M. T. Eijssen, C. T. Evelo, A. R. Pico, and E. L. Willighagen. Wikipathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res*, 46(D1):D661–D667, Jan 2018.

[258] M. Smeyne, P. Sladen, Y. Jiao, I. Dragatsis, and R. J. Smeyne. Hif1a is necessary for exercise-induced neuroprotection while hif2a is needed for dopaminergic neuron survival in the substantia nigra pars compacta. *Neuroscience*, 295:23–38, Jun 2015.

[259] A. R. Smith, R. G. Smith, E. Pishva, E. Hannon, J. A. Y. Roubroeks, J. Burrage, C. Troakes, S. Al-Sarraj, C. Sloan, J. Mill, D. L. van den Hove, and K. Lunnon. Parallel profiling of dna methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in alzheimer's disease. *Clinical Epigenetics*, 11(1):52, Mar 2019.

[260] Z. D. Smith and A. Meissner. Dna methylation: roles in mammalian development. *Nat Rev Genet*, 14(3):204–220, Mar 2013.

[261] R. Sobel. The extracellular matrix in multiple sclerosis: an update. *Brazilian Journal of Medical and Biological Research*, 34:603 – 609, 05 2001.

[262] S. Soliman. A stronger necessary condition for the multistationarity of chemical reaction networks. *Bulletin of Mathematical Biology*, 75(11):2289–2303, 2013.

[263] R. Sood, Y. Kamikubo, and P. Liu. Role of runx1 in hematological malignancies. *Blood*, 129(15):2070–2082, Apr 2017.

[264] S. Stolyar, S. Van Dien, K. L. Hillesland, N. Pinel, T. J. Lie, J. A. Leigh, and D. A. Stahl. Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology*, 3(1), 2007.

[265] B. D. Strahl and C. D. Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–45, Jan 2000.

[266] W. J. Strittmatter, A. M. Saunders, D. Schmechel, M. Pericak-Vance, J. Enghild, G. S. Salvesen, and A. D. Roses. Apolipoprotein e: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial alzheimer disease. *Proc Natl Acad Sci U S A*, 90(5):1977–1981, Mar 1993.

[267] D. B. Swartzlander, N. E. Propson, E. R. Roy, T. Saito, T. Saido, B. Wang, and H. Zheng. Concurrent cell type-specific isolation and profiling of mouse brains in inflammation and alzheimer's disease. *JCI Insight*, 3(13):e121109, Jul 2018.

[268] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5):861–872, Nov 2007.

[269] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, A. Sasaki, M. Yamamoto, M. Nakamura, K. Sutou, K. Osafune, and S. Yamanaka. Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nature Communications*, 5:3678 EP –, Apr 2014.

[270] K. Takahashi and S. Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676, 2006.

[271] N. Takasugi, T. Sasaki, K. Suzuki, S. Osawa, H. Isshiki, Y. Hori, N. Shimada, T. Higo, S. Yokoshima, T. Fukuyama, V. M.-Y. Lee, J. Q. Trojanowski, T. Tomita, and T. Iwatsubo. Bace1 activity is modulated by cell-associated sphingosine-1-phosphate. *J Neurosci*, 31(18):6850–6857, May 2011.

[272] R. C. Team. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. available online at https://www.r-project.org/. *online*, 2018.

[273] S. Teng, T. Madej, A. Panchenko, and E. Alexov. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys J*, 96(6):2178–2188, 2009.

[274] A. E. Teschendorff, M. Widschwendter, and Y. Jiao. A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinformatics*, 30(16):2360–2366, 05 2014.

[275] R. Thomas. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. In J. Della Dora, J. Demongeot, and B. Lacolle, editors, *Numerical Methods in the Study of Critical Phenomena*, pages 180–193, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.

[276] Y. Tomaru, M. Nakanishi, H. Miura, Y. Kimura, H. Ohkawa, Y. Ohta, Y. Hayashizaki, and M. Suzuki. Identification of an inter-transcription factor regulatory network in human hepatoma cells by matrix rnai. *Nucleic Acids Res*, 37(4):1049–1060, Mar 2009.

[277] A. Tovy, A. Spiro, R. McCarthy, Z. Shipony, Y. Aylon, K. Allton, E. Ainbinder, N. Furth, A. Tanay, M. Barton, and M. Oren. p53 is essential for dna methylation homeostasis in naïve embryonic stem cells, and its loss promotes clonal heterogeneity. *Genes Dev*, 31(10):959–972, May 2017.

[278] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nat Protoc*, 7(3):562–578, Mar 2012.

[279] G. E. Tsai, W. E. Falk, J. Gunther, and J. T. Coyle. Improved cognition in alzheimer's disease with short-term d-cycloserine treatment. *American Journal of Psychiatry*, 156(3):467–469, 1999.

[280] J. Utikal, N. Maherali, W. Kulalert, and K. Hochedlinger. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci*, 122(19):3502–3510, Oct 2009.

[281] C. Van Cauwenberghe, C. Van Broeckhoven, and K. Sleegers. The genetic landscape of alzheimer disease: clinical implications and perspectives. *Genet Med*, 18(5):421–430, May 2016.

[282] Z. K. Van Helmond, J. S. Miners, E. Bednall, K. A. Chalmers, Y. Zhang, G. K. Wilcock, S. Love, and P. G. Kehoe. Caveolin-1 and -2 and their relationship to cerebral amyloid angiopathy in alzheimer's disease. *Neuropathology and Applied Neurobiology*, 33(3):317–327, 2007.

[283] M. P. van Iersel, A. R. Pico, T. Kelder, J. Gao, I. Ho, K. Hanspers, B. R. Conklin, and C. T. Evelo. The bridgedb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics*, 11(1):5, Jan 2010.

[284] S. Vandermeersch, J. Vanbeselaere, C. P. Delannoy, A. Drolez, C. Mysiorek, Y. Guérardel, P. Delannoy, and S. Julien. Accumulation of gd1a ganglioside in mda-mb-231 breast cancer cells expressing st6galnac v. *Molecules*, 20(4):6913–6924, Apr 2015.

[285] C. Verderio, M. Gabrielli, and P. Giussani. Role of sphingolipids in the biogenesis and

biological activity of extracellular vesicles. *Journal of Lipid Research*, 59(8):1325–1340, 2018.

[286] J. Verheijen and K. Sleegers. Understanding alzheimer disease at the interface between genetics and transcriptomics. *Trends in Genetics*, 34(6):434 – 447, 2018.

[287] J. Vesa, M. H. Chin, K. Oelgeschläger, J. Isosomppi, E. C. DellAngelica, A. Jalanko, and L. Peltonen. Neuronal ceroid lipofuscinoses are connected at molecular level: interaction of cln5 protein with cln2 and cln3. *Mol Biol Cell*, 13(7):2410–2420, Jul 2002.

[288] M. K. Vickaryous and B. K. Hall. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biological Reviews*, 81(3):425–455, 2006.

[289] T. Vierbuchen, A. Ostermeier, Z. P. Pang, Y. Kokubu, T. C. Südhof, and M. Wernig. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, 463(7284):1035–1041, 2010.

[290] J. Wang, J.-T. Yu, M.-S. Tan, T. Jiang, and L. Tan. Epigenetic mechanisms in alzheimer's disease: Implications for pathogenesis and therapy. *Ageing Research Reviews*, 12(4):1024 – 1041, 2013.

[291] J. Wang, T. Zhou, T. Wang, and B. Wang. Suppression of lncrna-atb prevents amyloid-β-induced neurotoxicity in pc12 cells via regulating mir-200/znf217 axis. *Biomedicine & Pharmacotherapy*, 108:707 – 715, 2018.

[292] W. Wang, P. T. Toran, R. Sabol, T. J. Brown, and B. M. Barth. Epigenetics and sphingolipid metabolism in health and disease. *Int J Biopharm Sci*, 1(2):105, Oct 2018.

[293] D. F. Weaver, A. Meek, C. Barden, M. Reed, M. Taylor, Y. Wang, M. Brant, P. Stafford, B. Kelly, and E. C. Diez. Alzheimer's disease as a disorder of tryptophan metabolism. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 13(7):P1267, Jul 2017.

[294] G. A. Wells, D. Haguenauer, B. Shea, M. E. Suarez-Almazor, V. Welch, P. Tugwell, and J. Peterson. Cyclosporine for treating rheumatoid arthritis. *Cochrane Database of Systematic Reviews*, 2(2), 1998.

[295] H. Wu, Y. Deng, Y. Feng, D. Long, K. Ma, X. Wang, M. Zhao, L. Lu, and Q. Lu. Epigenetic regulation in b-cell maturation and its dysregulation in autoimmunity. *Cell Mol Immunol*, 15(7):676–684, Jul 2018.

[296] H. Wu, S. Fu, M. Zhao, L. Lu, and Q. Lu. Dysregulation of cell death and its epigenetic mechanisms in systemic lupus erythematosus. *Molecules*, 22(1):30, Dec 2016.

[297] H. Wu, M. Zhao, C. Chang, and Q. Lu. The real culprit in systemic lupus erythematosus: abnormal epigenetic regulation. *Int J Mol Sci*, 16(5):11013–11033, May 2015.

[298] H. Wu, M. Zhao, A. Yoshimura, C. Chang, and Q. Lu. Critical link between epigenetics and transcription factors in the induction of autoimmunity: a comprehensive review. *Clinical Reviews in Allergy & Immunology*, 50(3):333–344, Jun 2016.

[299] M. Wu, G. Chen, and Y.-P. Li. Tgf-b and bmp signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res*, 4:16009–16009, Apr 2016.

[300] Y. Wu, L. Chen, P. G. Scott, and E. E. Tredget. Mesenchymal stem cells enhance wound healing through differentiation and angiogenesis. *STEM CELLS*, 25(10):2648–2659, 2007.

[301] Y. Wu, R. C. H. Zhao, and E. E. Tredget. Concise review: bone marrow-derived stem/progenitor cells in cutaneous repair and regeneration. *Stem Cells*, 28(5):905–915, May 2010.

[302] Y. Xiao, Y. Gong, Y. Lv, Y. Lan, J. Hu, F. Li, J. Xu, J. Bai, Y. Deng, L. Liu, G. Zhang, F. Yu, and X. Li. Gene perturbation atlas (gpa): a single-gene perturbation repository for characterizing functional mechanisms of coding and non-coding genes. *Sci Rep*, 5(10889), 2015.

[303] H. Xu, Y.-S. Ang, A. Sevilla, I. R. Lemischka, and A. Ma'ayan. Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLOS Computational Biology*, 10(8):1–14, 08 2014.

[304] M. Ye, C. Coldren, X. Liang, T. Mattina, E. Goldmuntz, D. W. Benson, D. Ivy, M. B. Perryman, L. A. Garrett-Sinha, and P. Grossfeld. Deletion of ets-1, a gene in the jacobsen syndrome critical region, causes ventricular septal defects and abnormal ventricular morphology in mice. *Hum Mol Genet*, 19(4):648–656, Feb 2010.

[305] Z. Ye, B.-K. Chou, and L. Cheng. Promise and challenges of human ipsc-based hematologic disease modeling and treatment. *International Journal of Hematology*, 95(6):601–609, Jun 2012.

[306] E. Younesi and M. Hofmann-Apitius. From integrative disease modeling to predictive, preventive, personalized and participatory (p4) medicine. *EPMA J*, 4(1):23–23, Nov 2013.

[307] D. Zádori, P. Klivényi, I. Plangár, J. Toldi, and L. Vécsei. Endogenous neuroprotection in chronic neurodegenerative disorders: with particular regard to the kynurenines. *J Cell Mol Med*, 15(4):701–717, Apr 2011.

[308] N. Zaidan and K. Ottersbach. The multi-faceted role of gata3 in developmental haematopoiesis. *Open Biology*, 8(11):180152, 2018.

[309] Q. Zhang, C. Ma, M. Gearing, P. G. Wang, L.-S. Chin, and L. Li. Integrated proteomics and network analysis identifies protein hubs and network alterations in alzheimer's disease. *Acta Neuropathol Commun*, 6(1):19–19, Mar 2018.

[310] Y. Zhang, C. Pak, Y. Han, H. Ahlenius, Z. Zhang, S. Chanda, S. Marro, C. Patzke, C. Acuna, J. Covy, W. Xu, N. Yang, T. Danko, L. Chen, M. Wernig, and T. C. Sudhof. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron*, 78(5):785 – 798, 2013.

[311] H. Zhao, Z. Sun, J. Wang, H. Huang, J.-P. Kocher, and L. Wang. Crossmap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*, 30(7):1006–1007, 2014.

[312] Y. Zhao, X. Yin, H. Qin, F. Zhu, H. Liu, W. Yang, Q. Zhang, C. Xiang, P. Hou, Z. Song, Y. Liu, J. Yong, P. Zhang, J. Cai, M. Liu, H. Li, Y. Li, X. Qu, K. Cui, W. Zhang, T. Xiang, Y. Wu, Y. Zhao, C. Liu, C. Yu, K. Yuan, J. Lou, M. Ding, and H. Deng. Two supporting factors greatly improve the efficiency of human ipsc generation. *Cell Stem Cell*, 3(5):475 – 479, 2008.

[313] A. Zia and A. M. Moses. Towards a theoretical understanding of false positives in dna motif finding. *BMC Bioinformatics*, 13:151–151, Jun 2012.

[314] S. Zickenrott, V. E. Angarica, B. B. Upadhyaya, and A. del Sol. Prediction of disease-gene-drug relationships following a differential network analysis. *Cell Death Dis*, 7:e2040, 2016.