Small RNA Dynamics in Cholinergic Systems

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

Vorgelegt beim Fachbereich 14
der Johann Wolfgang von Goethe-Universität
in Frankfurt am Main

von
Sebastian Lobentanzer
aus Schlüchtern

Frankfurt 2019-2020
(D30)

Dissertation vorgelegt...

Small RNA dynamics in cholinergic systems

Abstract

Quisque facilisis erat a dui. Nam malesuada ornare dolor. Cras gravida, diam sit amet rhoncus ornare, erat elit consectetuer erat, id egestas pede nibh eget odio. Proin tincidunt, velit vel porta elementum, magna diam molestie sapien, non aliquet massa pede eu diam. Aliquam iaculis. Fusce et ipsum et nulla tristique facilisis. Donec eget sem sit amet ligula viverra gravida. Etiam vehicula urna vel turpis. Suspendisse sagittis ante a urna. Morbi a est quis orci consequat rutrum. Nullam egestas feugiat felis. Integer adipiscing semper ligula. Nunc molestie, nisl sit amet cursus convallis, sapien lectus pretium metus, vitae pretium enim wisi id lectus. Donec vestibulum. Etiam vel nibh. Nulla facilisi. Mauris pharetra. Donec augue. Fusce ultrices, neque id dignissim ultrices, tellus mauris dictum elit, vel lacinia enim metus eu nunc.

Contents

1 Introduction 1
1.1 Cholinergic Systems . . . 1
1.2 Cholinergic Aspects of Physiology and Disease . . . 3
1.2.1 Alzheimer's Disease . . . 3
1.2.2 Schizophrenia and Bipolar Disorder . . . 5
1.2.3 Immunity . . . 6
1.2.4 Neuroinflammation . . . 7
1.2.5 Stroke . . . 9
1.2.6 Circadian Aspects of Cholinergic Systems . . . 10
1.2.7 Neurokines . . . 12
1.3 Transcriptional Connectomics . . . 14
1.3.1 Transcription Factors . . . 15
1.3.2 microRNAs . . . 16
1.3.3 Transfer RNA Fragments . . . 18
1.4 Nested Multimodal Transcriptional Interactions - The Need for Connectomics . . . 19

2 miRNeo: Creation of a Comprehensive Connectomics Database 22
2.1 Implementation . . . 23
2.1.1 Neo4j: A Graph-Based Infrastructure . . . 24
2.1.2 High-throughput Database Generation . . . 25
2.1.3 Maintenance and Quality Control . . . 25
2.2 Materials . . . 26
2.2.1 Gene Annotation . . . 26
2.2.2 microRNA Annotation . . . 27
2.2.3 Transcription Factor Targeting . . . 27
2.2.4 microRNA Interactions . . . 28
2.2.5 Filtering of Aggregated Prediction Scores . . . 30
2.2.6 De-novo Prediction of tRF Targeting . . . 30
2.2.7 microRNA Primate Specificity . . . 31
2.3 miRNeo Usage . . . 33
2.4 Statistical Approach to Transcriptional Connectomics . . . 37
2.4.1 Permutation . . . 37
2.4.2 Gene Set Enrichment Analysis . . . 38

3 microRNA Dynamics in Cholinergic Differentiation of Human Neuronal Cells 39
3.1 Neuronal Transcriptomes - Background . . . 39
3.2 Cortical Single-Cell RNA Sequencing . . . 41
3.2.1 Single-cell Dataset Processing . . . 41
3.2.2 microRNA and Transcription Factor Targeting Predictions . . . 42
3.2.3 Single-cell Expression of Cholinergic and Neurokine Transcripts . . . 42
3.2.4 Nested Regulatory Networks of miRNAs and Transcription Factors in Single Cholinergic Cells . . . 42
3.2.5 Transcript Clustering Based On Expression . . . 42
3.2.6 Co-Expression of Functional Groups of Cholinergic Transcripts . . . 44

3.3 The Cellular Model . . . 46
3.3.1 The SH-SY5Y Neuroblastoma Cell Line . . . 46
3.3.2 The LA-N Neuroblastoma Cell Lines . . . 46
3.3.3 Culture . . . 47
3.3.4 Differentiation . . . 47
3.3.5 RNA Isolation . . . 48
3.4 Small RNA Sequencing and Differential Expression Analysis . . . 50
3.4.1 Sequencing . . . 50
3.4.2 Sequence Alignment . . . 51
3.4.3 Differential Expression Analysis - R/DESeq2 . . . 52
3.4.4 microRNA Dynamics in CNTF-mediated Cholinergic Differentiation of LA-N-2 and LA-N-5 . . . 52
3.5 microRNA Family Gene Ontology Enrichment . . . 58
3.5.1 microRNA Family Enrichment . . . 58
3.5.2 Creation of miRNA Family Gene Target Sets . . . 58
3.5.3 GO Analysis of Target Sets . . . 58
3.5.4 Large Scale GO Term Curation . . . 59
3.6 Whole Genome miRNA→Gene Network Generation . . . 60
3.7 Application to Schizophrenia and Bipolar Disorder . . . 63
3.7.1 Analysed Datasets . . . 63
3.7.2 Microarray Quality Control and Data Preparation . . . 63
3.7.3 Differential Expression Meta-Analysis . . . 65
3.7.4 Sexual Dimorphism in Schizophrenia and Bipolar Disorder . . . 66
3.7.5 Combination of Disease Data and Cell Culture . . . 68
3.7.6 miR-125b-5p Acetylcholinesterase Targeting Assays . . . 70
3.7.7 hsa-miR-125b-5p Targets Acetylcholinesterase . . . 70
3.7.8 Cholinergic/Neurokine Mechanisms in Web-Available RNA Sequencing Experiments . . . 71

4 Dynamics Between Small and Large RNA in the Blood of Stroke Victims 72
4.1 RNA Sequencing, Differential Expression, and Descriptive Methods . . . 73
4.1.1 The PREDICT Cohort . . . 73
4.1.2 Clinical Parameters Collected in the PREDICT Study . . . 73
4.1.3 Sample Collection, RNA Isolation, and Sequencing . . . 73
4.1.4 RNA Sequencing Alignment . . . 74
4.1.5 Quality Control and Filtering . . . 74
4.1.6 RNA Sequencing Differential Expression Analysis . . . 74
4.1.7 Gene Ontology Analyses . . . 74
4.1.8 Homology Computation Among tRNA Fragments . . . 75
4.1.9 t-Distributed Stochastic Neighbour Embedding . . . 75
4.1.10 Cholinergic Association of Small RNA Species . . . 75
4.2 Descriptive Analysis of RNA Dynamics in Blood After Stroke . . . 77
4.2.1 Differential Expression of Large RNA . . . 77
4.2.2 Gene Ontology Analyses of Differentially Expressed Genes . . . 77
4.2.3 Differential Expression of small RNA . . . 78
4.2.4 Homology Among tRNA Fragments . . . 79
4.2.5 Cholinergic Association of Small RNA Species . . . 80
4.3 Blood Compartments of Cholinergic Systems and Small RNA Species . . . 81
4.3.1 Large RNA Regulatory Circuits in Tissues of the Blood . . . 81
4.3.2 An Atlas of Small RNA Expression in Cell Types of the Blood . . . 81

4.3.3 Definition of Presence and Absence of Lowly Expressed smRNA Molecules . . . 82
4.3.4 Large RNA Expression Patterns Identify Cholinergic Systems in CD14$^+$ Monocytes . . . 84
4.3.5 Identification of Functional Enrichment of smRNA Expression in Blood-Borne Cells . . . 84
4.3.6 Expression Patterns of Differentially Expressed and Cholinergic-Associated smRNAs . . . 85
4.4 Regulatory Circuits of Small RNA and Transcription Factors in CD14$^+$ Monocytes . . . 88
4.4.1 Comprehensive Circuit Network Creation . . . 88
4.4.2 Gene Ontology Analyses of TF→Gene Networks of CD14$^+$ Monocytes . . . 88
4.4.3 Dichotomy of Small RNA Targeting of Transcription Factors in CD14$^+$ Monocytes . . . 89
4.4.4 Gradual Shift in Control Over Transcription Factors by miRNAs and tRFs . . . 89
4.4.5 Dichotomous Transcriptomic Footprints of Transcription Factors in CD14$^+$ Monocytes . . . 91
4.5 Feedforward Loops of Small and Large RNA . . . 94
4.5.1 Feedforward Loop Creation . . . 94
4.5.2 Visualisation and Modularisation . . . 95
4.5.3 Module-specific Functions via GO Analysis . . . 95
4.5.4 Feedforward Loop Network of CD14$^+$ Monocytes . . . 95
4.5.5 Transcript Clustering by smRNA:TF:gene Feedforward Loops Increases Informative Resolution of Gene Set Enrichment Analyses . . . 106

5 Discussion **111**
5.1 Methods . . . 112
5.1.1 Transcriptional Interactions: *miRNeo* . . . 112
5.1.2 RNA Sequencing . . . 114
5.1.3 Statistical Analyses of Network Interactions . . . 115
5.1.4 Cholinergic Cellular Models: LA-N-2 and LA-N-5 . . . 117
5.1.5 Stroke Patient Blood Samples . . . 118
5.1.6 Gene Ontology Analyses . . . 119
5.1.7 Feedforward Loop Analyses . . . 119
5.2 A Mechanistic Perspective of Transcriptional Interactions . . . 122
5.2.1 Analysis of Small RNA Dynamics via RNA-sequencing and Bioinformatics . . . 122
5.2.2 The Cholinergic/Neurokine Interface . . . 124
5.2.3 Molecular Biology of Feedforward Loops . . . 128
5.3 Small RNA Therapeutics and Pharmacology . . . 132

Bibliography **133**

List of Figures **169**

A Transcription Factor Regulatory Circuits - Tissue Types **171**

B List of Primate-Specific Homologues of Human microRNAs **172**

C microRNA Differential Expression in LA-N-2 and LA-N-5 **173**

D List of GO Terms from Analysis of Differentially Expressed Large RNA in Stroke 174

E Examples of Presence/Absence Definition of Small RNA 175

Preamble

This dissertation comprises three main chapters, which are written in a combined *methods-results-discussion* style. In the first main part, chapter two, I address the creation, maintenance, and usage of the database designed for assessing transcriptional interactions in the experimental parts. Since the creation process in itself is methodical, distinction between method and result can often not be implemented in a clear, »journal-style« manner. In chapters three and four however, that are concerned with experimental application of transcriptional interactions in cholinergic differentiation and disease, the manuscript will be consistently structured to visually distinguish the methods from results and discussion. Method-related paragraphs will be set in sans-serif font style, while non-method parts will be set in serif font. Because of the dimensions of this dissertation and the diverging topics, the non-method parts of each chapter will be in the style of combined results and discussion, to keep the immediate discussion close to the related results. Finally, there will be dedicated chapters for more broad and generalised discussion and conclusions.

At the date of submission, the majority of the contents of chapter three, the main experimental work of this dissertation, as well as most of the ideas developed in chapter two, have been published in a peer-reviewed journal. [1] Chapter four serves to illustrate my contributions to another manuscript that is at the moment submitted to XX, awaiting response.[?] That manuscript diverges from the contents described in chapter four mainly by the additional experiments concerned with validation of detected tRNA fragments. The development and usage of the database described in chapter two is invited for closer explanation by Cell Protocols and pending submission.[?]

This dissertation features content boxes for general information pertaining to a specific aspect (e.g., cholinergic genes). For very large or very small numbers, the scientific exponential notation (»E notation«) is used; e.g., 4.7E-05 reads as 4.7×10^{-5}, or 0.000047.

Data Availability

This is the dedication.

»Ever tried. Ever failed. No matter.
Try again. Fail again.
Fail better.«

Simon Beckett

Acknowledgments

Lorem ipsum consectetuer adipiscing elit. Morbi commodo, ipsum sed pharetra gravida, orci magna rhoncus neque, id pulvinar odio lorem non turpis. Nullam sit amet enim. Suspendisse id velit vitae ligula volutpat condimentum. Aliquam erat volutpat. Sed quis velit. Nulla facilisi. Nulla libero. Vivamus pharetra posuere sapien. Nam consectetuer. Sed aliquam, nunc eget euismod ullamcorper, lectus nunc ullamcorper orci, fermentum bibendum enim nibh eget ipsum. Donec porttitor ligula eu dolor. Maecenas vitae nulla consequat libero cursus venenatis. Nam magna enim, accumsan eu, blandit sed, blandit a, eros.

Abbreviations

ABC ATP binding cassette

ACh acetylcholine

AChE acetylcholinesterase (protein)

AD Alzheimer's Disease

Ago argonaute (protein)

ALD adrenoleukodystrophy

API application programming interface

BD Bipolar Disorder

CA cholinergic-associated

CAGE 5' cap analysis of gene expression

cAMP cyclic adenosine monophosphate

CARS compensatory anti-inflammatory response syndrome

CD cluster of differentiation

ChAT choline acetyltransferase (protein)

ChIP chromatin immunoprecipitation

CIDS CNS injury-induced immunodepression syndrome

CNS central nervous system

DAG directed acyclic graph

DAMP damage-associated molecular pattern

DE differentially expressed

DMEM Dulbecco's modified eagle medium

FCS fetal calf serum

FDR false discovery ratio

FFL feedforward loop

GEO Gene Expression Omnibus (NCBI)

GO Gene Ontology

gp130 see IL6ST (gene)

HLA-DR monocyte human leukocyte antigen isotype DR

iPSC induced pluripotent stem cell

IRF interferon regulatory factor

KO knockout

LA-N-2 human neuroblastoma cell line (female)

LA-N-5 human neuroblastoma cell line (male)

LBP lipopolysaccharide binding protein

LDT laterodorsal tegmentum (Ch6)

LFC $\log_2$ fold change

LPS lipopolysaccharide

MBL mannan-binding lectin

MCAO middle cerebral artery occlusion

miRNA microRNA

mRS modified Rankin Scale (clinical score of stroke severity)

NCBI National Center for Biotechnology Information

ND10 nuclear domain 10

nt nucleotide

OR odds ratio

PBG parabigeminal nucleus (Ch8)

PBS phosphate buffered saline

PCA principal component analysis

RT-qPCR real-time quantitative polymerase chain reaction

PD Parkinson's Disease

PPN pedunculo-pontine nucleus (Ch5)

REM rapid eye movement

RIN RNA integrity number (RNA quality measure)

RISC RNA-induced silencing complex

RPMI1640 Roswell Park Memorial Institute medium

SCN suprachiasmatic nuclei

SCZ Schizophrenia

RNA-seq RNA sequencing

SG significant gene (as in »differentially expressed«)

SIRS systemic inflammatory response syndrome

smRNA small non-coding RNA

SQL structured query language

TF transcription factor

tiRNA transfer RNA half

TPM transcripts per million

tRF transfer RNA fragment

tRNA transfer RNA

t-SNE t-distributed stochastic neighbour embedding

UTR untranslated region

vAChT vesicular acetylcholine transporter (protein; from SLC18A3 gene)

VP vascular permeability

WT wild type

Gene Symbols

ACHE acetylcholinesterase

ACLY ATP citrate lyase

AIF1 allograft inflammatory factor 1 (microglia marker protein)

AKT Serine/Threonine Kinase 1 (also known as Protein Kinase B)

ANG angiogenin

BAD BCL-2-associated agonist of cell death

BCL-2 B cell lymphoma 2

BDNF brain-derived neurotrophic factor

BMAL1 brain and muscle ARNT-like protein 1 (also known as ARNTL)

CHAT choline acetyltransferase

CHRNA7 nicotinic acetylcholine receptor subunit $\alpha 7$

CLOCK circadian locomotor output cycles kaput

CNTF ciliary neurotrophic factor

CNTFR ciliary neurotrophic factor receptor (soluble)

CRY cryptochrome

ERK extracellular-signal-regulated kinase

GFAP glial fibrillary acidic protein (central astrocyte marker)

IFN interferon

IL interleukin

IL6R interleukin 6 receptor (soluble)

IL6ST interleukin 6 signal transducer (membrane bound; also known as gp130)

ISG IFN-stimulated gene

JAK janus kinase

JNK JUN N-terminal kinase

LIF leukaemia inhibitory factor

LIFR leukaemia inhibiting factor receptor (soluble)

MAPK mitogen-activated protein kinase

MCL1 myeloid leukaemia cell differentiation protein 1

MHC major histocompatibility complex

MyD88 myeloid differentiation primary response 88

NF-κB nuclear factor 'kappa-light-chain-enhancer' of activated B-cells

NGF nerve growth factor

NGFR nerve growth factor receptor (also known as p75)

NPAS2 neuronal PAS domain protein 2

NR1D1 nuclear receptor subfamily 1; group D; member 1 (also known as Rev-Erbα)

NR1F1 nuclear receptor subfamily 1; group F; member 1 (also known as RORα)

NTRK1 neurotrophic receptor tyrosine kinase 1

NTRK2 neurotrophic receptor tyrosine kinase 2

OLIG1 oligodendrocyte transcription factor 1

PER period

PI3K phosphoinositide 3-kinase

RBFOX3 RNA-binding Fox-1 homolog 3 (neuronal marker gene; also known as NeuN)

RORA RAR-related orphan receptor α (see NR1F1)

SLC18A3 vesicular acetylcholine transporter (official gene symbol)

SST somatostatin

STAT signal transducer and activator of transcription

TGF transforming growth factor

TLR toll-like receptor

TNF tumour necrosis factor

TYK tyrosine kinase

VEGF vascular endothelial growth factor

VIP vasoactive intestinal peptide

1 Introduction

1.1. Cholinergic Systems

Nary a process in the mammalian body can commence without participation of cholinergic systems. Acetylcholine (ACh) was chemically and pharmacologically described by Henry Dale more than 100 years ago.[2] A short time later, Otto Loewi published the first proof of signal transmission by small molecules: he transferred physiological solutions from electrically stimulated frog hearts to naive hearts and observed their reactions; the solution that provoked a parasympathetic response he proposed to contain a »vagus substance«.[3] Finally, in 1929, Henry Dale completed the picture by isolating acetylcholine from mammalian tissue and identifying it as the molecule responsible for the parasympathetic response.[4] Dale and Loewi's joint effort in »Discoveries Relating to Chemical Transmission of Nerve Impulses« was rewarded with the »Nobel Prize in Physiology or Medicine« in 1936.

Although we have learned much about cholinergic systems in these past 100 years, our understanding of the mammalian nervous system still is fairly limited. Even when disregarding peripheral nervous systems, the complexity of cholinergic transmission is immense, and a myriad functions have been attributed to cholinergic circuits in the central nervous system (CNS). Central nervous projections of cholinergic fibres were extensively mapped by Marek-Marsel Mesulam and others in the 1980s,[5,6] with a majority of long projection neurons originating in one of the eight cholinergic nuclei, Ch1-Ch8. While many of these anatomical structures have been filled with meaning by associations with both rudimentary as well as higher brain functions, there are still as many cholinergic pathways whose function is entirely unclear (Figure 1.1, from Lobentanzer *et al.*[1]). This holds particularly

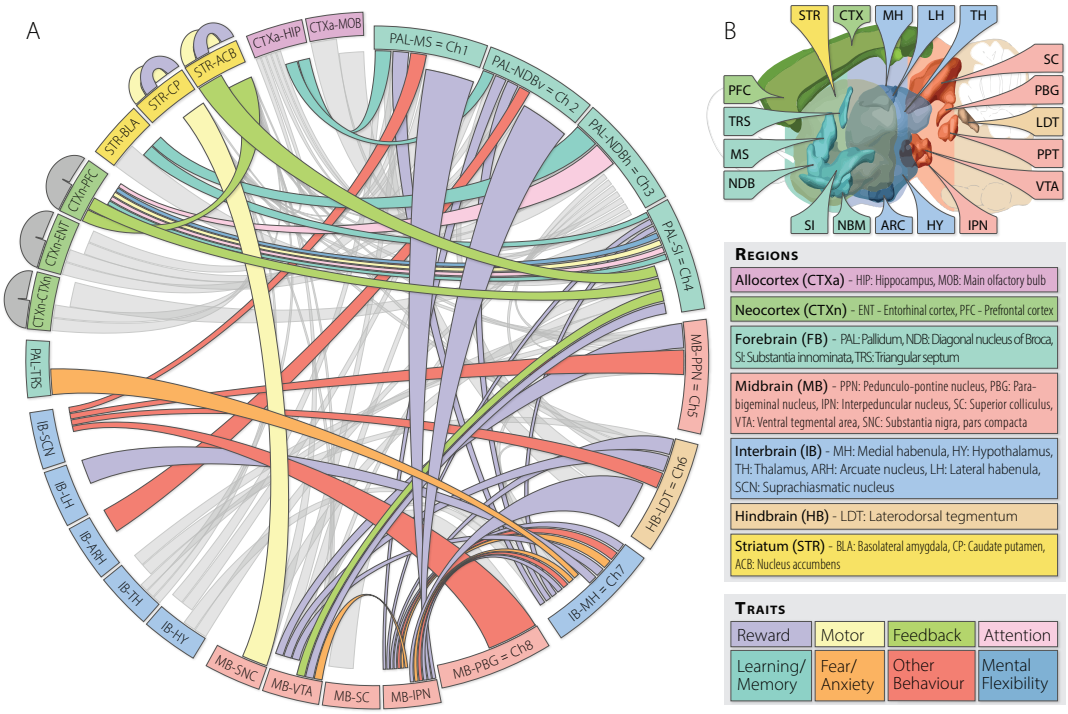

Figure 1.1: Cholinergic Projections in the CNS. Cholinergic systems are implicated in many diverse functional categories. **A)** The bulk of cholinergic projection neurons stems from one of the eight cholinergic nuclei, Ch1-Ch8 (right side of ideogram). They innervate wide areas of the mammalian CNS, and in turn receive incoming connections from all around the brain (left side of ideogram). Efferent connections are indicated by a small gap between ideogram and connector, in the first clockwise half of each ideogram component, afferent connection by a large gap, in the second half. A number of projections has been associated with specific functions, as implicated by the colours of the connectors. Two populations of cholinergic interneurons have been identified, in the striatum and the neocortex (outside of ideogram). **B)** Brain region and trait colour legend for A).

true for the only recently discovered cortical cholinergic interneurons, which, in comparison to their projecting counterparts, are very small and numerically vastly inferior to other neuron types in the cortex. Thus, their detection and analysis with current methods is challenging.

The histological definition of what constitutes a cholinergic neuron is not without debate. The staining procedures established in the 1970s utilised monoclonal antibodies against acetylcholinesterase (AChE),[7] whose association with cholinergic neurons is not definitive, as it can be expressed post-synaptically as well (for an overview of genes of the cholinergic systems, see Box 1). Later on, developments in horseradish peroxidase systems allowed immunohistochemistry on choline acetyltransferase (ChAT), which is a more immediate marker of cholinergic neurons,[5] albeit much more lowly expressed than AChE. However, AChE-based staining still was consistently used in addition to ChAT staining,[6] sometimes without much differentiation. Recently, single-cell RNA sequencing (RNA-seq) allows a more detailed appreciation of the transcriptional diversity of neurons, and enables a clearer distinction between cholinergic and non-cholinergic neurons expressing AChE (see Section 3.2).

1.2. Cholinergic Aspects of Physiology and Disease

Cholinergic systems are integral for a myriad physiological functions, and as such they are critically involved in aetiologies and phenotypes of a number of central and peripheral diseases. Of interest to this dissertation are the cholinergic aspects of degenerative and non-degenerative central nervous diseases (such as Alzheimer's Disease, Bipolar Disorder, Schizophrenia), ischemic conditions in stroke, and peripheral modulation of immune responses, particularly in the context of the aforementioned diseases.

1.2.1 Alzheimer's Disease

Alzheimer's Disease (AD) was characterised by Alois Alzheimer in 1906 and later named after him by his colleague and mentor, Emil Kraepelin.[8] AD is a progressive neurodegenerative disease, its main risk factor is age, and it is estimated to make up 60-70% of all dementia cases. The disease incidence and progression distinctly differ between the sexes;[9] generally, women are affected more often. Unlike the very rare familial form (that can affect patients in their fifties), spontaneous AD usually only begins to manifest symptomatically in the 6th to 7th life decade. As a result of the demographic change in most western countries, patient numbers, and thus, medical efforts, are expected to more than double in size until the year 2050. The cognitive decline associated with AD is progressive and ultimately leads to exhaustive care dependency; there is no cure.

The pathological hallmarks of AD are two types of atypical protein aggregates, inter-cellular amyloid β »plaques«, and intra-cellular »neurofibrillary tangles« composed of hyper-phosphorylated tau-protein. Often, pathological aggregation of these proteins begins decades before the onset of

Box 1: The Cholinergic Genes

Acetylcholine is synthesised from acetyl-CoA - supplied by **ATP citrate lyase** (*ACLY*) - and choline via enzymatic catalysis by **choline acetyltransferase** (ChAT from the *CHAT* gene). It is then packed into vesicles by the **vesicular acetylcholine transporter** (vAChT from the *SLC18A3* gene). After release into the synaptic cleft, it binds to a variety of **nicotinic and muscarinic receptors** (*CHRNx*, 16 subunits, and *CHRMx*, 5 subtypes). Of those, the nicotinic receptors form heteropentameric or, seldom, homopentameric ion channels, while the muscarinic receptors are monomeric G protein-coupled transmembrane receptors. The human possesses a duplicate of the nicotinic α7 receptor, **dupα7** (*CHRFAM7A*), which cannot bind ACh and supposedly acts as a dominant negative regulator of the α7 homomeric receptor. Termination of the signal is mainly achieved by **acetylcholinesterase** (AChE from the *ACHE* gene), one of the fastest enzymes known, with a theoretical rate of 25 000 molecules per second. AChE tetramers are usually tethered to cell membranes in the synaptic vicinity by the **proline-rich membrane anchor** (*PRIMA1*) or **collagen Q** (*COLQ*) peptides. Complementary to the mostly residual AChE is the circulatory **butyryl cholinesterase** (BChE from the *BCHE* gene), which can also nonspecifically degrade ACh. After degradation, residual choline is reimported into cells via the **high affinity choline uptake transporter** (HACU, from the *SLC5A7* gene).

symptoms. However, there have also been numerous cases of cognitively healthy subjects showing high amounts of protein aggregates. These inconsistencies and the unclear causality of pathology and symptoms have led to a redirection of scientific efforts to processes unrelated to amyloid and tau, such as neuroinflammation (Section 1.2.4).

Cholinergic systems have long been associated with AD, as evidenced by the cholinergic hypothesis that was posed in the 1980s. To cite from Bartus *et al.*, 1982: [10]

> »We have been guided by three deductive requirements that must be satisfied if the cholinergic hypothesis is to deserve continued attention: (i) specific dysfunctions in cholinergic markers should be found in the brains of subjects suffering from age-related memory loss, (ii) artificial disruption of central cholinergic function in young subjects should induce behavioural impairments that mimic the cognitive loss found naturally in aged subjects, and (iii) appropriately enhancing central cholinergic activity in aged subjects should significantly reduce age-related cognitive deficits.«

Although many reports substantiate all three deductive prerequisites, the cholinergic hypothesis has in the last decades been overshadowed by alternative hypotheses, particularly, amyloid-related theories. However, the therapeutic approaches developed along the lines of preventing amyloid β aggregation or otherwise clearing the plaques or soluble aggregates have not been successful in alleviating the cognitive decline in patients. Thus, pro-cholinergic intervention by means of AChE inhibition still makes up the majority of approved drugs. The monotherapeutic approach of AChE inhibition is based on a premise that seems simple in light of the enormous complexity surrounding the interplay of the billions of neurons in the process of memory formation and recall, and in fact, pro-cholinergic therapy has only been shown to alleviate symptoms or delay their onset; a reversal of cognitive losses has so far not been achieved by any drug regime. As such, even in regard only to the cholinergic systems, novel approaches are sorely needed.

On the other hand, the cholinergic deficit in AD is not purely symptomatic. There is considerable debate whether the pathology originates from the entorhinal cortex or the basal forebrain. As Heiko and Eva Braak have shown, [11] AD pathology follows a characteristic brain region distribution process that can be stratified into stages and starts in the entorhinal region. The early stages of pathology by far precede the onset of symptoms. Taylor Schmitz and colleagues substantiate the cholinergic origin of neurodegeneration in their longitudinal *in vivo* imaging studies: [12,13] In cognitively normal and impaired human subjects, basal forebrain volume predicted longitudinal entorhinal degeneration, but not *vice versa*. As such, the cholinergic basal forebrain dysfunction precedes as well as predicts pathology in other affected areas and cognitive deficits. [12] Additionally, the spread of Alzheimer's pathology in the longitudinal progression of the disease reflects the spatial topography of basal forebrain cholinergic projections. [13]

1.2.2 Schizophrenia and Bipolar Disorder

Cognitive deficits can also occur without neuron death. The earliest description of what we today call Schizophrenia (SCZ) was coined by Emil Kraepelin: »*dementia praecox*«, premature dementia.[14] Cognitive deficits are an integral but often overlooked part of the clinical picture of SCZ, which is often dominated by the more impressive positive symptoms such as hallucinations and paranoia. Likewise, Bipolar Disorder (BD) can present with cognitive impairments. The cognitive impairments affecting both SCZ and BD patients involve diminished problem solving capabilities and reduced intelligence, and are more pronounced in SCZ than in BD.[15] They have been connected to cholinergic dysfunction[16,17] and the sum of anticholinergic medications.[18,19] A human polymorphism in the $\alpha 5$ nicotinic receptor subunit predicts a higher propensity for smoking and SCZ, showing parallel manifestations in engineered mice[20] and rats.[21] Correspondingly, cholinergic stimulation can improve cognition[22,23,24] and mood,[25] but can on the other hand provoke schizotypic behaviour in AD patients.[26]

SCZ and BD clinically present with clear sexual dimorphisms. Compared to women, men have a higher SCZ prevalence with an odds ratio (OR) of 1.4, are affected earlier (at 15-25 as compared to 25-35 years of age), and face a worse prognosis.[27] Cholinergic participation also appears sex-dependent: Male SCZ patients more often self-medicate by smoking (7.2 versus 3.3 weighted average OR with 90% lifetime prevalence).[28] BD, on the other hand, affects men and women almost equally, with an OR of ~1. However, women make up 80-90% of so-called »rapid cyclers«, a subgroup of patients showing short intervals between manic and depressive phases which is associated with a worse prognosis.[29] Additionally, major depressive disorder, which is a prerequisite for BD diagnosis, more often affects women (OR = 2).[30]

Psychiatric genomics has recently identified a high amount of shared heritability between SCZ and BD.[31] Likewise, transcriptomic analyses have shown a high correlation (71%) between the transcriptional perturbations in the two diseases.[32] Clinical as well as molecular pathology intensifies from BD to SCZ, suggesting the two lie on different points of a shared spectrum. However, their genetic origins are tremendously complex. Multiple disease-relevant markers have been identified by GWAS (genome-wide association studies), even able to distinguish between several sub-phenotypes of each disease.[33] These markers are found in neurotransmitter receptors (e.g., dopaminergic, glutamatergic, cholinergic), scaffolding proteins (DISC: »disrupted in schizophrenia«), multiple transcription factors (TFs), microRNAs (miRNAs), and non-coding regions without known function.[34,35,36]

Considering this complex disease aetiology, it is not surprising that there are no »designer« drugs available against SCZ and BD. All available therapeutic options have been empirically identified, starting with the first antipsychotic, chlorpromazine, synthesised in 1950 by the French pharmaceutical company Rhône-Poulenc. Originally developed in a series dedicated to the search for antihistaminics, it was soon recognised for its antipsychotic potential, and widely prescribed only few years later. Many

other neuroleptic compounds have been derived from chlorpromazine, and through binding affinity assays, their receptor profiles were established. Most compounds with antipsychotic properties have a wide spectrum of different receptor activities, but most early drugs were strong antagonists of the D_2 dopamine receptor. Thus, the »dopaminergic hypothesis« of SCZ was formulated. However, aetiology as well as therapeutic principles are unclear to this day, and most antipsychotic substances still are very »dirty drugs«. In fact, newer developments leading to the discovery of the second generation (»atypical«) neuroleptic substances, starting with clozapine, have created molecules with an even wider spectrum of interactions und thus less specificity towards a single therapeutic mechanism of action. Similarly, the archetypal »mood stabiliser« Lithium, that has been found to ameliorate depressive as well as manic phases of BD, likely influences a wide variety of neuronal functions via mechanisms yet unclear.[37] This general development is contrary to pharmaceutical research practice, where most developments aim for a higher specificity. It is thus very likely that pharmacological therapy of SCZ and BD requires an approach consisting of multiple pharmacodynamic angles, to account for the multigenic disruption.

1.2.3 Immunity

Aside from its vast neuronal functions, ACh also is highly relevant in immune cells, recently reviewed by Fujii and colleagues.[38] The first to isolate ACh from an animal organ, Dale and Dudley,[4] used the spleens of oxen and horses. The spleen receives sympathetic, but only very sparse parasympathetic innervation, and as such, the large amounts of ACh found by Dale and Dudley had to have come from immune cells. Indeed, nearly all mammalian immune cells express cholinergic components, most importantly, B- and T-cells, monocytes/macrophages, and dendritic cells. While ACh is physicochemically rather stable, it is extremely susceptible to enzymatic degradation, and cholinesterases are ubiquitarily distributed and cleave ACh with stunning efficiency, reducing its diffusion range to few millimetres. As a result, ACh has to be supplied synaptically or, at most, in paracrine fashion. ChAT activity has been confirmed in B- and T-cells, which both contain significant amounts of ACh, although T-cells generally possess higher amounts. Additionally, in peripheral cells ACh can be synthesised by the mitochondrial carnitine acetyltransferase. In addition to B- and T-cells, *CHAT* mRNA has been found in macrophages and dendritic cells. ChAT expression and ACh synthesis can be induced by various immune mediators, such as lipopolysaccharide (LPS) and other toll-like receptor (TLR) agonists.

In addition to ACh synthesis, all of the aforementioned cell types can receive cholinergic signals. They express all muscarinic receptors as well as a selection of nicotinic receptor subunits, and the signal-terminating esterases. Although it is not completely clear how the parasympathetic signal reaches the immune cells, cholinergic activation as a result of inflammation can dampen the immune response in what is described as the »cholinergic anti-inflammatory reflex«.[39] This reflex loop is designed to protect the body from pathogens and inflammation, but also from the harmful effects of

immune stimulation. Upon afferent signalling through the afferent vagus nerve and humoral components, the brain releases humoral (via the hypothalamic-pituitary-adrenal axis) and neuronal (via the sympathetic and parasypathetic autonomous fibres) anti-inflammatory signals. The spleen has been identified as a pivotal organ in this response. Since none of the immune organs receives parasympathetic innervation, it has been proposed that the cholinergic activation is generated locally, with the help of sympathetic signalling to the organ.[40]

A special role among cellular ACh receptors is occupied by the $\alpha 7$ nicotinic receptor subunit. Previously thought to exclusively form homopentameric ion-channel receptors, its functional characteristics have recently been extended. It has been found to form heteropentamers with $\beta 2$ subunits, akin to the prominent $\alpha 4\beta 2$ receptors in the brain, and an expression in immune cells also is likely. A hybrid duplication of the *CHRNA7* gene with *FAM7*, *CHRFAM7A*, is translated to a functional protein (dup$\alpha 7$). However, it seems to lack ACh binding ability, and thus is thought to act as a dominant negative regulator of $\alpha 7$ receptor function. In addition to its ionotropic function, mainly by means of Ca^{2+} transduction, the $\alpha 7$ receptor has been found to possess G-protein coupled metabotropic effects that can extend the duration of cholinergic activation. The $\alpha 7$ receptor can also, independently of Ca^{2+}, activate the JAK2/STAT3 pathway (see Section 1.2.7) in macrophages, leading to suppression of NF-κB signalling.

On the other hand, cholinergic activation via M_1 and/or M_5 muscarinic receptors can lead to a positive immune response. The difference between muscarinic and nicotinic immune-signalling is elucidated by transgenic receptor knockout (KO) animals: Splenar cells from selective M_1/M_5-KO mice secreted significantly lower amounts of the neuromodulators tumour necrosis factor (TNF)-α, interferon (IFN)-γ, and interleukin (IL)-6 than those from wild type (WT) mice. Conversely, antigen-stimulated splenar cells from $\alpha 7$-KO mice produced significantly greater amounts of TNF-α, IFN-γ, and IL-6 than WT. In summary, the effects of cholinergic stimulation of the immune system is bidirectional and strongly context-dependent, and specific pharmacological intervention can shift homeostasis in both pro- as well as anti-inflammatory directions.

1.2.4 Neuroinflammation

Neurodegenerative as well as non-degenerative neurologic diseases are increasingly being associated with immunologic phenomena, prompting the need for integrative and translational approaches.[41] Transient and chronic inflammatory events can influence neuronal function and even survival in a dramatic fashion. Further, failure to resolve the acute inflammatory states may lead to maladaptive states, cases of »frustrated resolution«,[42] in which the goal of adaptive immunity is not met. As was recently shown,[43] resolution of inflammation is not just the »phasing-out« of inflammatory events, but rather bridges the gap between innate and adaptive immunity. While many acute-phase T_H1-type cytokines may have evolved to drive inflammation, their protracted influence may derail these post-inflammatory events and thus lead to maladaptive responses and chronic inflammation. T_H1-

type cytokines include TNFs, IFNs, IL-1β and IL-6, and downstream mediators discussed in this context are manifold: phosphoinositide 3-kinase (PI3K); cyclic adenosine monophosphate (cAMP); myeloid leukaemia cell differentiation protein 1 (MCL1); the complex of B cell lymphoma 2 (BCL-2), Serine/Threonine Kinase 1 (AKT), and BCL-2-associated agonist of cell death (BAD); all variants of mitogen-activated protein kinase (MAPK), i.e., extracellular-signal-regulated kinase (ERK) 1 and 2, JUN N-terminal kinase (JNK), and p38 MAPK; and the NF-κB pathway. This list is not comprehensive, for a more detailed overview, see Fullerton & Gilroy.[42]

While none of the cardinal symptoms of inflammation are easily assessed in a CNS context, Virchow's fifth cardinal sign, *functio laesa*, is particularly difficult to tie to chronic neuroinflammation. Complex behavioural syndromes such as the functional deficits accompanying neurologic diseases may be influenced by protracted, maladaptive immunity, but the affected areas, brain structures, and timelines cannot be measured with current methods in neuroimaging. Only recently, it has become known that the brain is not immunologically pristine, but rather possesses a very specialised immune system, showing grave distinctions from, but also overlap with, peripheral immune systems.[44,45] The mechanisms of immune privilege of the brain are constantly being refined; there is crosstalk between brain and periphery with blood-to-brain and brain-to-blood messaging,[40] and even migration of immune cells into the CNS, mainly as a response to sustained inflammation.

The first line of defence in CNS tissues are microglia, which in their physiological state are resident ramified monocytes, and upon antigen sensing can produce an immediate native immune reaction. Further, the nascent immune system of the CNS comprises similar cells as the peripheral systems (T-cells, B-cells, NK-cells, dendritic cells), albeit with significant differences: antigen presenting cells express significantly fewer major histocompatibility complex (MHC) I and II molecules (which can however be induced by cytokine release upon inflammation); and the endocrine conditions (secretion of immune mediators from neurons) entail a more rapid response (seconds instead of days) with shorter duration of inflammation than in the periphery.[45]

Under the surveillance of resident microglia, there is constitutive and inducible migration of immune cells between CNS tissues and periphery, in a loop of infiltration and drainage. Starting at antigen presentation inside the CNS, activated immune cells leave the CNS through one of two routes, both ending at the deep cervical lymph nodes: either through the cribiform plate into the nasal mucosa, or into the meningeal lymphatic vessels accompanying the sagittal and transversal sinuses in the *dura mater*. Following an immunological stimulus, activated immune cells in the deep cervical lymph nodes facilitate a secondary immune response and protect nervous tissue through secretion of cytokines[46] and re-migration of secondary immune cells to the CNS. Regulatory cytokines include IL-1 and IL-6, CCLs and CXCLs, leukaemia inhibitory factor (LIF), and epidermal and fibroblast growth factors. Re-migrating T cells can utilise various adhesion molecules expressed by endothelia along the blood-brain-barrier to cross into the CNS in a controlled fashion.[45]

1.2.5 Stroke

Stroke is a medical emergency in which reduced blood flow leads to massive neuron death in the brain. There are two types of stroke: Haemorrhagic stroke, which is caused by bleeding due to a rupture of brain vessels, and ischemic stroke, misperfusion of a brain region caused by a clot in a cranial artery. Ischemic stroke makes up the vast majority of all strokes, about 90%. Stroke is currently the second most frequent cause of death in developed countries, only second to coronary artery disease. Those who survive the stroke are in many cases permanently and severely disabled. The prognosis of stroke patients is mainly dependent on clinical complications during initial care.

Infections are a leading cause of death in stroke patients, the CNS injury itself is an independent risk factor for the development of life-threatening infection. The most frequent complications accompanying stroke are fever and pneumonia, the fever in turn being most often caused by the infection. Affecting more than 20% of stroke patients, pneumonia is the most common serious post-stroke complication, featuring a mortality rate of more than 30%.[47] Stroke patients demonstrate a significant immunosuppression, resulting in lower count and functionality of immune cells. The ability of monocytes to synthesise cytokines is drastically reduced, a finding that has been reproduced in animal models of cerebral ischemia.

T cell activation and proliferation in the deep cervical lymph nodes is elevated following CNS injury, implicating that the drainage of immune cells from the site of injury plays an important role in immune system stimulation.[46] Depletion of those T cell leads to neuron death, suggesting that this naive response is favourable in stroke and similar conditions. However, the specifics of activation and migration, and the mediators (cytokines, antibodies) influencing the post-injurious response are still largely unclear.[44]

The immunological situation after stroke is dominated by two opposing factors, the pro-inflammatory bodily response to injury and, often, infection, and the counter-regulatory immunodepressive response including the cholinergic anti-inflammatory reflex (see Section 1.2.3). The inflammatory response can become pathologic in the case of excess stimulation, which can result in »systemic inflammatory response syndrome« (SIRS), which in extreme cases can lead to shock and organ failure. As a counterbalancing measure, the body responds with a »compensatory anti-inflammatory response syndrome« (CARS), designed to allow fighting the infection while also protecting the body from excessive immune stimulation. However, in CNS injury, the anti-inflammatory component may overwhelm inflammatory processes, leading to a pathological »CNS injury-induced immunodepression syndrome«, CIDS.

In addition to the humoral and neurohumoral immunomodulatory pathways described above, the brain can directly steer immune processes by the release of cytokines. This may be particularly impactful in CNS injury, where blood-brain-barrier and homeostasis are disrupted. Contrary to the selective uptake of substances into the CNS, export from the CNS is mostly instantaneous and not

tightly controlled. Among the cytokines found in circulation after stroke are transforming growth factor (TGF)-β, IL-1β, IL-6, and TNF-α. So far, it is unclear which of the immunomodulatory axes (humoral, neurohumoral, or direct release of cytokines from the brain) contributes to CIDS, and how they relate to each other. As a consequence, it is also unclear if directed intervention against CIDS would be beneficial, or even feasible.[47]

The pro-inflammatory acute and sub-acute, and the anti-inflammatory chronic phases of stroke development are mediated by cellular components of the immune system. Particularly important for the homeostasis of debris removal and vascular stability are circulatory and stationary monocytes (in the brain called microglia).[48] Monocytes differentiate into macrophages upon inflammatory stimuli; the two diametrically opposite phenotypes are the pro-inflammatory M1 type and the anti-inflammatory M2 type (unfortunately harbouring confusion potential with muscarinic receptors). Monocytes, in humans, show similar classes, that can be distinguished by expression of several clusters of differentiation (CDs), mainly, CD14 and CD16. The main pro-inflammatory phenotype expresses mainly CD14 and no CD16 ($CD14^{++}CD16^{-}$), while the anti-inflammatory phenotype expresses less CD14 and high amounts of CD16 ($CD14^{+}CD16^{++}$). There is an immediate phenotype ($CD14^{++}CD16^{+}$), which is closer in function to the pro-inflammatory phenotype. Rodents also possess pro- and anti-inflammatory phenotypes, but no intermediate phenotype.[48]

After stroke, monocytes undergo complex regulation that parallels acute and chronic phases. In acute and sub-acute phases, pro-inflammatory monocytes are elevated in blood and brain, whereas anti-inflammatory monocytes are decreased. This partly reverses in the chronic phase; in the brain, anti-inflammatory monocytes »take over«, while the numbers of pro-inflammatory monocytes decrease. Whether this is conveyed through cross differentiation from one phenotype to the other, or by migration, is yet unclear.[48] In blood, both phenotypes are decreased in the chronic phase, while the bone marrow increases monocyte production. In regard to the cholinergic inflammatory reflex, it seems important to note that the spleen, which acts as a reservoir for monocytes in the periphery,[49] has significant effects on stroke recovery. Stroke leads to contraction of the spleen, and a subsequent reduction in the number of monocytes in the spleen, and an increase in the brain.[50] Splenectomy, 2 weeks before permanent middle cerebral artery occlusion in rats, ameliorated acute pathology and reduced the number of brain macrophages after the infarction.[51] Influences on medium- to long-term recovery, however, were not tested.

1.2.6 Circadian Aspects of Cholinergic Systems

Cholinergic systems and psychiatric diseases share another common theme: regulation of circadian time and sleep patterns. Cholinergic nuclei have been associated with the resetting of the circadian clock in the suprachiasmatic nuclei (SCN). Retrograde tracing from the SCN[52] has identified basal forebrain nuclei (Ch1 & Ch4), as well as the pedunculo-ponitine nucleus (PPN, Ch5), laterodorsal tegmentum (LDT, Ch6), and the parabigeminal nucleus (parabigeminal nucleus (PBG), Ch8)

as regulatory input to the SCN (compare also Figure 1.1). Basal forebrain cholinergic neurons are active in wakefulness and in the rapid eye movement (REM) phase of sleep,[53] and optogenetic activation of PPN and LDT cholinergic neurons (channelrhodopsin 2 under the ChAT promoter) during non-REM sleep was sufficient to induce REM sleep in mice.[54] Basal forebrain projections to other brain regions seem to functionally diverge from the projections to the SNC. In a study analysing prefrontal and hippocampal cholinergic activities, the increase in tonic ACh release during REM sleep was contingent on subsequent wakefulness,[55] and thus may convey a stronger »wake-up« signal than projections to the SCN alone.

Muscarinic receptors M1 and M3 are essential for REM sleep: REM-sleep is completely abolished in combined M1/M3 receptor KO mice.[56] Arousal-induced phase shifts induced by activation of Ch4 cholinergic neurons projecting to the SCN were blocked in animals pretreated with (anti-muscarinergic) atropine injections to the SCN, demonstrating that cholinergic activity at muscarinic receptors in the SCN is necessary for arousal-induced phase shifting.[57] However, in their atropine perfusion experiment (locally via injection), the authors did not preclude cholinergic influences from the other nuclei.

In parallel, psychiatric diseases regularly exhibit symptoms of disturbed circadian rhythm. In the cholinergic-catecholaminergic imbalance hypothesis of BD,[16] the imbalance follows variable transcriptionally regulated rhythms, and affected individuals exhibit decreased REM latency (the duration from onset of sleep to the first REM phase), which can be modulated by muscarinic agonists/antagonists.[58] Conversely, sleep deprivation exerts short-term antidepressant effects,[59] reduced cortical ACh levels,[60] and vast transcriptional changes in basal forebrain cholinergic neurons.[61]

While the SCN regulates circadian timing of the organism, individual cellular timings are controlled by a group of transcriptional activators and de-activators, called clock genes. The autoregulatory feedback loop thus created oscillates between day and night timing, under the influence of external factors. How exactly the individual cellular clocks are synchronised by the SCN is still unclear.[62] The first molecular circadian controller, circadian locomotor output cycles kaput (CLOCK), was identified by Joseph Takahashi and colleagues via mutagenesis screening in mice in 1997.[63] The transcription factors CLOCK and brain and muscle ARNT-like protein 1 (BMAL1) form heterodimers and bind to E-box elements in the promoters of period (PER) 1 and 2, and cryptochrome (CRY) 1 and 2, which lead to negative feedback regulation; or to E-box elements in the promoters of NR1D1 (giving rise to the Rev-Erbα protein) and NR1F1 (giving rise to RORα), which compete for the ROR element in the BMAL1 promoter. RORα induces BMAL1 expression, while Rev-Erbα represses it, thus leading to an oscillating expression pattern. In neurons, CLOCK can be substituted by its paralogue NPAS2.

1.2.7 Neurokines

In comparison to the widely studied cholinergic projection neurons originating in the basal forebrain (Ch1-Ch4) that are known to depend on a retrograde survival signal by means of nerve growth factor (NGF), trophic influences on other cholinergic populations such as the cortical interneurons are unclear. NGF was described by Rita Levi-Montalcini in the 1950s as the first known instance of trophic peptides required for the survival of sympathetic ganglia.[64] The group of neurotrophic substances since discovered (most prominently, the brain-derived neurotrophic factor BDNF) are commonly referred to as »neurotrophins«. They convey their trophic effects through a family of transmembrane receptors; NGF binds to neurotrophic receptor tyrosine kinase 1 (NTRK1) with high affinity, BDNF binds to neurotrophic receptor tyrosine kinase 2 (NTRK2) with high affinity. However, both also bind to a third receptor, nerve growth factor receptor (NGFR), which is also known as p75, although with low affinity. NGFR function is complex, depending on the context it seems to be able to suppress as well as enhance the primary neurotrophic signal mediated by NTRK1/2. The dependence of basal forebrain cholinergic neurons on retrograde NGF signalling was discovered in the 1980s.[65]

A second group of trophic peptides with cholinergic implications are the so-called »neurokines«; the name results from the fact that this particular subgroup of cytokines has been associated with neuronal function in the central and peripheral nervous systems. Most prominently, they include the ciliary neurotrophic factor (CNTF), LIF, and IL-6, all of which coincidentally have been known by the acronym CDF. In the late 1980s, two groups of scientists (McManaman[66] and Rao[67]) independently identified proteins in extracts of muscle fibre that induced a differentiation of neurons towards a cholinergic type, and thus termed these proteins »choline acetyltransferase development factor« or »cholinergic differentiation factor« (both abbreviated CDF). Only later, through sequencing of the peptides, it became known that they had in fact discovered two distinct neurokines, LIF (Rao) and CNTF (McManaman, personal communication). IL-6, on the other hand, is abbreviated CDF for an entirely different reason: in this case it is short for »CTL (cytolytic T lymphocyte) differentiation factor«.

CNTF, LIF, and IL-6 convey their impact on neuronal activity through a partly redundant neurokine receptor pathway.[29] There are two basic types of neurokine receptors: soluble and transmembrane. The primary receptors for CNTF (CNTFR) and IL-6 (IL6R) are soluble proteins that are secreted into the extracellular space and, upon binding of a neurokine, bind to transmembrane receptor dimers on the cell surface. These transmembrane receptors are the LIF receptor (LIFR) and the »interleukin 6 signal transducer« (IL6ST), which is also known as gp130. Due to the latter's predominance, neurokines are also referred to as gp130 receptor family cytokines[68]. Every neurokine has its preferred constellation of soluble and transmembrane receptors: CNTF binds to the soluble CNTF receptor and a dimer consisting of one gp130 and one LIFR protein; IL-6 binds to the soluble IL6R and a dimer of two units of gp130; LIF does not usually bind a soluble receptor but rather binds

immediately to a dimer comprising one of each gp130 and LIFR; however, there are significant redundancy, pleiotropy, and crosstalk between those systems.[69,68,70] Notably, a membrane bound IL-6 receptor is also expressed on several select cell types: macrophages, neutrophils, some types of T-cells, and hepatocytes.[?]

All receptor constellations result in a main effect of activation of the JAK/STAT cascade (Fig. 1.2). More specifically, neurokines can activate janus kinases (JAKs) 1 and 2 or the homologous tyrosine kinase (TYK) 2, and, successively, STAT (»signal transducer and activator of transcription«) isoforms 1, 3, 5A, and 5B, which then convey a multitude of cellular effects (e.g. in immunity or differentiation) through transcriptional activation. The STAT cascade is inherently self-limiting in that it usually leads to expression of transcription factors that serve as repressors of the STAT genes by SOCS (suppressors of cytokine signalling), PIAS (protein inhibitors of activated STATs), and PTPs (protein tyrosine phosphatases).[69]

remove miRs to later?

Neurokines, particularly IL-6, may serve as a link between the immunological and cholinergic aspects of physiologic or disease processes. Since IL-6 is implicated in neurodegenerative, psychiatric, and injurious CNS diseases (Section 1.2), which all also possess a cholinergic facet, it makes sense to not only see it in the light of an immunomodulator, but also as a potential influence on neuronal function in cholinergic systems. A third of basal IL-6 levels are generated in adipose tissue in healthy humans,[71] and central (fatty) obesity increases risk for AD about 3-fold.[72,73] SCZ and BD also are associated with obesity, although causality still is unclear. While obesity itself is more predominant in SCZ than in BD, obesity in BD patients is associated with decreased global cognitive ability as well as with poorer performance on individual tests of processing speed, reasoning/problem-solving, and sustained attention.[74] Low-grade chronic inflammation is recognised in obesity[75] as well as in neurodegenerative[76] and non-degenerative psychiatric diseases.[77,78] Sleep disturbance in animal models of mood disorders is accompanied by elevation in blood levels of IL-1, IL-6, and TNF-α.[79] Additionally, it has been shown that LIF can lead to a catecholaminergic-to-cholinergic neurotransmitter switch in peripheral neurons in a mouse model of protracted inflammation accompanying collagen-induced arthritis.[80] Though being a marginal phenomenon, it is not unthinkable that similar processes in central nervous cell may contribute to a disruption of homeostasis of cholinergic systems, and thus, to disease.

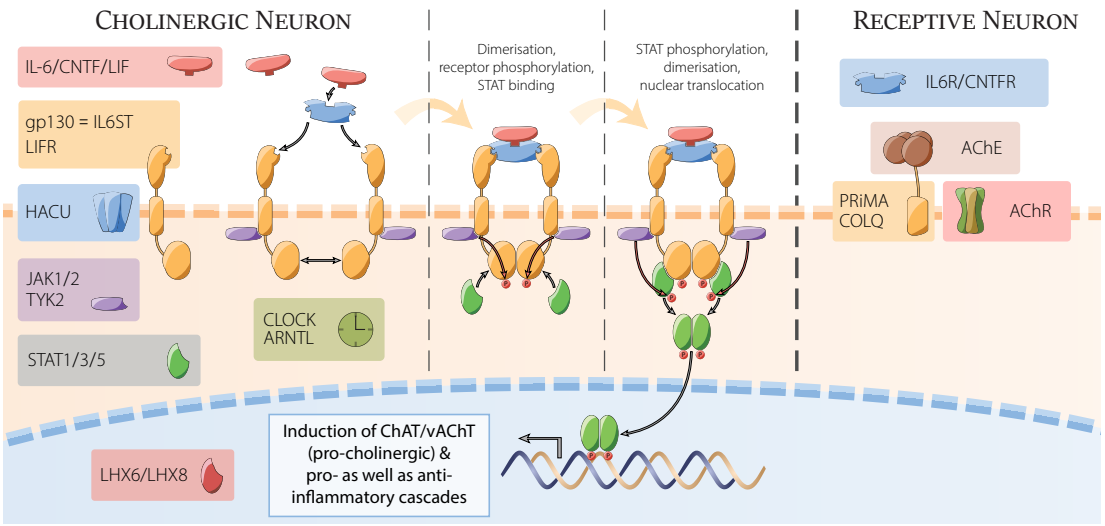

Figure 1.2: The Neurokine Pathway. The neurokines, such as CNTF, LIF, and IL-6, signal through a combination of soluble and membrane-bound receptors. Activation of a transmembrane neurokine receptor is usually followed by JAK recruitment and phosphorylation, and successively by STAT activation and translocation to the nucleus.

1.3. Transcriptional Connectomics

The term »connectomics« is not strictly limited to one scientific discipline; it is frequently used when the studied matter is defined by complex relationships between interaction partners. The most frequent use outside of transcriptional matters is neuronal connectomics, i.e., the relationships and projections between brain regions. In this dissertation, connectomics generally refers to epi-transcriptional interaction, the processes surrounding protein-coding gene expression. For the sake of simplicity, in this dissertation all descriptions of genomics and transcriptomics matters, of genes and their small RNA regulators, are to be seen in the context of *Homo sapiens*, unless explicitly stated otherwise.

No matter their location, cholinergic neurons are defined by their ability to synthesise ACh and release it to neighbouring cells to a certain effect. To fulfil this task, two particular proteins are essential: the choline acetyltransferase (ChAT) to synthesise ACh from choline and acetyl-Coenzyme A, and the vesicular acetylcholine transporter (vAChT, official gene symbol SLC18A3), which concentrates ACh in vesicles for later release. A notable genetic feature connects these two proteins beyond their functional association: the small *SLC18A3* gene - only 2420 nucleotides (nt) in size - sits inside the first intron of the CHAT gene and thus is already included in its primary transcript, and is subject to the CHAT promoter. However, oftentimes the (mature) transcript levels of CHAT and SLC18A3 mRNA seem to be independently regulated; from the perspective of the organism, the possibility of differential regulation between these two genes makes sense. Since SLC18A3 apparently does not possess its own promoter, this differential regulation has to be conveyed epigenetically.

This dissertation deals in large parts with approaches aiming to decipher these interactions; and while its primary topic revolves around cholinergic systems, the methods described in the following

are designed to be applicable to the entirety of the genome/epigenome. Four particular types of cellular actors are subjects of these methods and therefore will be briefly introduced: genes in the classical sense as the conveyors of cellular function by encoding for proteins; TFs, a subclass of protein coding genes that are able to regulate the expression of other genes; miRNAs, a class of small non-coding RNA (smRNA) that has been known for approximately two decades and is reasonably well described functionally and mechanistically; and transfer RNA fragments (tRFs), a second class of regulatory smRNA that has only recently been rediscovered and is significantly less well described regarding its functionality.

is this the right place?

Naturally, there are multiple additional epigenetic regulatory mechanisms that are not subject to the herein described methods, some of which closely interact with small RNA function. For instance, long non-coding RNAs are a large, novel class of RNA that is poorly characterised as of yet, but has been shown in several instances to interfere with gene expression via RNA-binding protein interactions, or with miRNA function via sponging of miRNA molecules. Other epigenetic processes such as DNA methylation or histone modifications are also known to significantly influence gene expression; however, their effects are in most cases not catalogued in a comprehensive fashion and thus are not amenable to whole-genome bioinformatics analyses.

1.3.1 Transcription Factors

Transcription factors (TFs) were among the first intracellular regulatory mechanisms to be discovered (the earliest article referencing the term »transcription factor« in its title on PubMed was published in 1972). TFs commonly translocate from the cytosol into the nucleus upon activation (often by phosphorylation), where they bind specific DNA sequences that usually range in size from 6 to 12 nt. The regions containing these binding sites (about 100 - 1000 nt in size) determine the effect upon binding, which can be one of two main modes: either a promoter, leading to an increased activity of transcription in the downstream vicinity of the binding site, or a repressor, having the opposite effect.

There exists a vast body of knowledge on TF-interactions with genes, mostly due to the long period of time since their discovery and the multitude of scientific publications, most often studying single TFs and their interactions with few genes, but cumulatively curated by several organisations. One of the currently largest curations of TF data, TRANSFAC, saw its original release in 1988. While these curation efforts can be extensive, they may present with serious bias towards particular TFs that may hold more scientific interest and thus are published far more frequently than others. Recently, comprehensive efforts have extended the available data significantly. Driven by the advent of RNA-seq, computational approaches have become able to not only comprehensively predict TF-gene interactions, but to do so in a highly tissue-specific manner (see Section 2.2.3). The human body is estimated to express up to 2600 distinct DNA-binding proteins, most of them presumed TFs,[81] although other studies give lower estimates.

1.3.2 microRNAs

The first endogenous »small RNA with antisense complementarity« was described in 1993,[82] but microRNAs (miRNAs) were only recognised as a distinct regulatory class of molecules in the early 2000s. They are typically between 18 and 22 nt-long, single stranded RNA fragments, and their function is now largely undisputed: miRNAs serve as targeting molecules for a protein complex whose primary purpose is to repress translation of mRNA, and, in some cases, lead to mRNA degradation. The complex, therefore, is called RNA-induced silencing complex (RISC); central to its function is the family of argonaute (Ago) proteins, which can bind the mature miRNA and orient it for interaction with its targets. Guidance of RISC to the target mRNA is generally mediated via sequence complementarity between miRNA and the targeted mRNA. Specifically, a »seed« region, usually bases 2-8 on the miRNA, is mainly responsible for the interaction; in case of perfect complementarity of this seed to the mRNA sequence, the interaction is considered »canonical«.

In early miRNA research, the 3' untranslated region (UTR) of the mRNA was believed to contain most miRNA binding sites due to its greater accessibility (i.e., the lack of active ribosomes); however, cumulative recent reports suggest that binding inside the coding region of the mRNA is a regular occurrence(cite). The rules governing miRNA binding to target sequences show considerable flexibility; a recent study shows about 30% of analysed relationships to be of »non-canonical« nature. In those cases, seed pairing with the mRNA is often imperfect. To ameliorate this loss of stability, compensation occurs typically by a secondary complementary structure after a small gap of non-complementary bases, leading to a »bridge«-type constellation. This flexibility has implications in applications involving targeting algorithms; those that consider only the seed region are more prone to false negatives than models that consider, for instance, the free energy of the whole molecule (see Section 2.2.4).

figure?

Biogenesis

miRNAs, similar to coding genes, are transcribed from loci on the genome, many inside introns or even exons of coding genes.[83] The primary transcript (primary miRNA or pri-miRNA) typically contains a hairpin-like structure that usually results in a double-stranded molecule because of internal complementarity, and can contain up to six mature miRNAs. This hairpin structure is recognised by the DGCR8 protein (DiGeorge Syndrome Critical Region 8, in invertebrates called »Pasha«); the complex then associates with the RNA-cleaving protein »Drosha«, which removes bases on the opposite side of the hairpin, creating a miRNA precursor (or pre-miRNA), which is subsequently exported from the nucleus by the shuttle protein Exportin-5. In a final step in the cytosol, the ribonuclease »Dicer« removes the loop joining the 3' and 5' arms of the pre-miRNA, resulting in a duplex of mature miRNA, about 20 nt long. Initially, it was thought to contain only one active

miRNA, resulting in a designation of »miRNA*« for the complementary strand (commonly, the strand with lower expression). However, this notion has been disproven, and to reflect the possibility of both strands performing miRNA functions, nomenclature has changed to specify the arm of the pre-miRNA from which the mature form originates (suffix »-3p« for the 3' arm, and »-5p« for the 5' arm).

miRNA genes, in the same way as protein coding genes, can be subject to promoters and repressors, adding another layer of expression control by TFs. However, these TF-miRNA relationships are far less well described than common coding gene interactions, because miRNAs due to their shortness are not amenable to many standard gene expression assay forms. Estimation of the number of distinct gene targets of any one miRNA varies widely; however, it is generally accepted to not be less than several dozen targets per miRNA, and up to thousands of genes per miRNA (although that estimate may be overenthusiastic).

Organisation and Curation

miRNAs are organised and curated by means of a periodically updated web-based platform, miRBase.[84] For *Homo sapiens*, miRBase v21 contains 2588 mature miRNAs from 1881 precursors. Evolutionarily, the miRNA repertoire has grown from rodents to primates, resulting in a number of primate-specific miRNAs that may convey additional function. miRNA nomenclature is organised[85] in a way that assigns evolutionarily conserved miRNAs the same designation (number) in all species in which they are expressed. In their full names, a prefix stating the organism of origin is added; for example, hsa-miR-125b-5p (for *Homo sapiens*) and mmu-miR-125b-5p (for *Mus musculus*) share the same sequence and most of their functionalities.

miRNAs are subcategorised in families (designated »mir« with lowercase »r«) by their genomic origin and phylogenetic homology aspects. As the annotation itself, family affiliations are in flux and change with each miRBase version. miRBase v21 lists 151 distinct miRNA families with 721 individual members in total. The remaining 1867 miRNAs do not (yet) belong to a larger family; the majority (80%) of those is newly discovered, as indicated by a 4-digit designation number.

Disease Association

miRNAs have been associated with a number of CNS diseases, including AD, Parkinson's Disease (PD), BD, and SCZ. However, the largest contribution since their discovery by far has been made by cancer research; of the approximately 90 000 publications found on PubMed with the term miRNA, about 42 000 involve cancer (search term »miRNA AND cancer«). In comparison, »miRNA AND Alzheimer's Disease« results in about 600 hits, while a search for »miRNA AND Schizophrenia« yields just 363 publications (as of October 2019).

In AD, several groups of miRNAs have been found to show characteristic perturbations before the

onset of symptoms, which makes them interesting biomarker candidates.[86] Some miRNAs have been extensively studied in a variety of contexts, most prominently hsa-miR-132-3p. Among its targets are several key neuronal regulators (e.g. FOXP2, FOXO3, P300, MeCP2), and it is in turn controlled by many pivotal neuronal elements (e.g. REST, ERK1/2, CREB); this presents an explanation for the many physiological and pathological situations that miR-132-3p has been found to play a role in. Its functions include the control of neuronal survival/apoptosis, migration and neurite extension, neuronal differentiation, and synaptic plasticity.

miRNAs are able to fulfil their regulatory purpose in a context- and cell-type-dependent manner,[87] such that the perturbation of one single miRNA may provide different functional outcomes in different tissues (e.g., glial cells and neurons), or different stages of disease. However, this »jack-of-all-trades« behaviour also poses significant problems in establishing miRNAs as pharmacological targets: In the case of antagonising or mimicking an existing miRNA, the amount of off-target effects would not only be enormous, the entire definition of an off-target effect would continuously change between tissues and during the course of the disease. For this reason, the design of custom oligonucleotides with limited capabilities may be preferable in the development of therapeutics based on RNA interference (See also Section 5.3).

1.3.3 Transfer RNA Fragments

Transfer RNA (tRNA) breakdown products have been known for decades, with first descriptions in the 1970s; back then, they were associated with a higher turnover of tRNA in cancer cells,[88] and proposed as urine-based biomarkers for certain malignancies.[89] However, their genesis was attributed to random processes, and due to lacking molecular biology characterisation techniques, interest in those fragments quickly faded. It was not until recently that studies have shown tRNA to be a major source of stable expression of small noncoding RNA[90,91] in most mammalian tissues. Indeed, replicating the reports from the 1970s, but now in the form of comprehensive small RNA analysis of human biofluids,[92] tRNA breakdown products are the dominant form of small RNA in secreted fluids, such as urine and bile, and make up large parts of other bodily fluids as well. They exist in two major forms: transfer RNA halves (tiRNAs), and the smaller transfer RNA fragments (tRFs). tiRNAs derive from either end of the tRNA, and are created by angiogenin cleavage at the anticodon loop.[93,94] Smaller fragments are derived from the 3' and 5' ends of the tRNA (3'-tRF/5'-tRF) or internal tRNA parts (i-tRF), respectively, and may incorporate into Ago protein complexes and act like miRNAs to suppress their targets.[95,96]

However, there is considerable controversy about the generalisation of tRF functions, as distinct publications discover very different and sometimes opposing mechanisms of action for their respective fragments. An obvious assumption is the miRNA-like functionality, at least for those tRFs that are in the length range of miRNAs. There have been several instances of tRFs proven to act as miRNA-like suppressors of translation in a RISC-associated manner,[96] and of Dicer playing a large

part in their biogenesis.[90] There are even instances of small RNA molecules previously mislabeled miRNAs that have been discovered to actually be tRNA-derived, such as miR-1280.[97]

On the other hand, multiple groups have identified tRFs to function not in an antisense-complementary manner, but by homology aspects. A valine-derived tRF was found to regulate translation by competing with mRNA directly at the binding site at the initiation complex and thereby displacing the original mRNA, leading to its translational repression.[98] Others have found multiple classes of tRFs derived from glutamine, aspartate, glycine, and tyrosine tRNAs, that displace multiple oncogenic transcripts from an RNA-binding protein (YBX1), conveying tumour-suppressive activity.[99] Most counterintuitive is the recent finding of a tRF proven to bind to several ribosomal protein mRNAs and *enhancing* their translation, and, when specifically inhibited, leading to apoptosis in rapidly dividing cells.[100]

There is no consistent nomenclature yet to describe and organise tRFs, which are by nature more heterogeneous than miRNAs; while only 61 mature tRNAs are required in a cell to achieve a one-to-one »codon→amino acid« translation, one tRNA molecule can be the origin of several hundred distinct tRF molecules. Additionally, the amount of human tRNA genes is estimated at 500-600,[101] and there are many more pseudo-tRNA genes. To communicate the identity of individual tRFs, multiple approaches are common in current literature; most prominently, tRFs are tied to the parent tRNA and the amino acid carried by this tRNA. To illustrate: The 22-nt LeuCAG3′ tRF (meaning: a fragment of 22 bases starting at the 3' end of the leucine-carrying tRNA with anticodon »CAG«) was shown to play an important role in regulating ribosome biogenesis.[100] Since there is no repository of the likes of miRBase yet, this approach can be cumbersome for replication purposes, and explicit statement of the exact sequence of each fragment is a must in publication. In fact, since the aforementioned paper does not mention the sequence explicitly, there exist 6 distinct possibilities of fragments fitting this description. While manageable on this small scale, this system prohibits efficient analysis of larger sets of tRFs that cannot be individually controlled. For this reason, the approach of Loher and colleagues[102] may be preferable: they propose the generation of a »license plate« based on the sequence of the fragment directly, composed of the prefix »tRF«, the length of the fragment, and a custom oligonucleotide string encoding (e.g., »B3« codes for »AAAGT«). This way, tRF names are unique and unmistakably linked to the sequence, nomenclature is species-independent, and tRNA origin can be quickly determined by sequence lookup.

1.4. Nested Multimodal Transcriptional Interactions - The Need for Connectomics

The ultimate aim of transcriptional connectomics is the combination of all interacting cellular components in a model that satisfactorily explains our real-life observations and is able to predict the functional outcome of a modification of one of these players. Even in the simplified case of only

studying the interactions between coding genes, TFs, miRNAs, and tRFs, the complexity of the required model exceeds our current capabilities by far. The more we know about the functioning of these intertwined systems, the more we understand how much there is still to learn.

For instance, only recently has it become clear how complex transcriptional regulation by means of TFs really is, and, incidentally, the two systems studied foremost in this dissertation (nerve and immune cells) are the two most transcriptionally complex systems in any mammal.[103] Through study of comprehensive genomic information of 394 tissue types in approximately 1000 human primary cell, tissue, and culture samples (from the FANTOM5 consortium) it was estimated that the mean number of active TFs towards any given gene is highest in immune (12 TFs per gene) and nervous cells (10 TFs per gene), and that any one TF in nervous and immune cells controls expression of a mean of 175 and 160 genes, respectively (see also Section 2.2.3).[103]

Similarly, it has been found that miRNAs, particularly in the nervous system, possess a much higher tissue specificity than coding genes, resulting in an expression landscape that varies widely between individual neuron types that are in close proximity in the brain. With the exception of single cell RNA-seq, no modern analysis method is capable of a resolution appropriate for accurate characterisation of these expression patterns, resulting in extinction of the signal of miRNAs that are not expressed consistently across cell types (similar to »housekeeping« genes) because of statistical interference. Very recent studies show that miRNA:gene co-expression networks are tightly linked to cell types in the nervous system, and that groups of miRNAs as functional modules associate with particular phenotypes in developmental and mature states.[104] This functional association with cell phenotype was found in quality comparable to the expression patterns of TFs, yet in quantity conveys smaller impact and thus is thought to be a fine-tuning mechanism, subtle and precise in purpose.

Another aspect of the tissue specificity of CNS-associated miRNAs is the high likelihood of underrepresentation or even non-discovery of those very specifically expressed miRNAs. Adding to the problem is the experimental bias towards rodent models when it comes to thorough studies of the CNS, where human or other primate samples are a rarity compared to rats or mice. Assessments of the numbers of yet unknown novel primate- and tissue specific miRNAs estimate their magnitude in the thousands,[105] resulting in an effective doubling of currently known miRNAs.

These high numbers of potentially interacting players present computational challenges: Approximating the number of expressed genes in a human cell at 20 000, the number of TFs at a low 500, and an actual number of interactions per TF at 10, the total possible interactions C are given by

$$C = \frac{500!}{10!(500-10)!} \cdot 20\,000$$

which practically equals infinity. This is without accounting for different tissue types or cell states (e.g., differentiation or disease). Similarly, the amount of mature miRNAs (2588 in miRBase v21) and their ability to target even more distinct transcripts than TFs with one single molecule present

immense computational requirements for even listing all possible or actual relationships. An interaction table describing targeting of genes by miRNAs in one type of tissue has $2588 \cdot 20\,000 \approx 50$ million individual fields.

Combining the different modes of transcriptional interaction presents additional challenges. A simple model system to visualise (in only one type of cell) the interaction of TFs targeting genes, and of miRNAs targeting genes as well as TFs, contains about 20 000 genes (a subset of which of the size of about 2000 are TFs), 2588 mature miRNAs, and a total of $2588 \cdot 20\,000 + 2000 \cdot 20\,000 \approx 90\,000\,000$ potential interactions. In standard application scenarios, such as the generation of an interaction network around a group of genes (e.g., the cholinergic genes), the processing requirements grow linearly with each added interaction partner, and exponentially with every regulatory layer that is added.

Practically, this information has to be provided, gathered, and integrated, which further multiplies the amount of storage and processing power required. miRWalk 2.0, a collection of miRNA interaction data, has collected 12 of the most popular miRNA-targeting prediction datasets, each of which has their strengths and weaknesses (see 2.2.4). Experimentally validated interactions (e.g. as collected in DIANA TarBase or miRTarBase) are gold standard, but far from comprehensive and strictly speaking only relevant for the cellular context in which the experiment was originally performed; there are also different evidence qualities to be accounted for, depending on the type of experiment performed. Ideally, all of these data are still accessible when performing the analysis, so a database created for this purpose should be able to incorporate all this information without any data loss while still remaining feasible in terms of computation time as well as space and working memory requirements.

This dissertation will first describe the creation of such a database and what has been learned during its various stages, and then go on to apply the database to different biological problems from real world experiments, such as the cholinergic differentiation of human male and female cultured neuronal cells, and the blood of stroke victims.

more extensive description of content?

»Wir sehen in der Natur nie etwas als Einzelheit, sondern wir sehen alles in Verbindung mit etwas anderem, das vor ihm, neben ihm, hinter ihm, unter ihm und über ihm sich befindet.«

Johann Wolfgang von Goethe

2

miRNeo: Creation of a Comprehensive Connectomics Database

Natural philosophy, as represented by the thought of Johann Wolfgang von Goethe, wants to holistically describe nature and explain and interpret its particular mechanisms. Although natural philosophy is the predecessor of modern, empirical science, its concepts and approaches are still valuable in today's data driven world. As the data we collect grows to dimensions that can only be interpreted with the aid of computers, functional reductionism becomes a valuable paradigm: By studying the facets of nature, we strive to understand it as a whole. Similarly, we regularly encounter Goethe's paraphrase of »all things are connected« in neuro-immunology, and in transcriptional connectomics.

Bioinformatic support in connectomics is indispensable, which can be seen by the sheer multitude of possible interactions between the participating factors. However, when I began working on this project (October 2015), there was no integrative database available for this purpose. Earlier that year, miRWalk 2.0 had been published, for the first time providing a relatively comprehensive source of predicted as well as experimentally validated miRNA targeting data [106] (see 1.3.2). One year later, Marbach's »regulatory circuits« were published, [103] enabling analysis of comprehensive TF→gene relationships in 394 human tissues (see Section 1.3.1). These collections (as well as the data they were derived from) are the basis of the database further called *miRNeo*, the development of which will be described in the following chapter.

Since a large part of the scientific progress of this dissertation deals with practical problems of multimodal connectomics, I will begin by describing the infrastructure that makes effective computation

of these problems possible. After this technical description of database structure and creation, I will explain the types and organisation of its content. The remainder of the chapter will then deal with the application of this infrastructure to real-world problems in transcriptional connectomics, and the statistical approaches suited to this special case.

2.1. Implementation

For any biological question to be asked in a bioinformatics setting, the effectiveness of the computational query determines the practicality of the approach. Because resources (i.e., processing power, storage, and working memory) are limited, the database that is queried should be organised in a way that facilitates retrieval of the desired information without excess processing of useless information, for instance, reporting the absence of a connection. In the simplified case of only miRNAs interacting with genes in one direction (miRNA→gene), this means retrieval of only those interactions relevant for the queried genes or miRNAs.

Traditional table-based approaches (also known as relational databases) such as SQL (»Structured Query Language«) cannot provide such an implementation, since individual entries for genes and miRNAs (rows and columns) have to be accessed in their entirety, whether there is a connection between gene and miRNA (1) or not (0). Additionally, adding layers to these interactions (e.g., distinct prediction algorithms, tissues, or the interaction between TFs and genes) require the addition of entire tables the same size as the database, which is detrimental to effective use of space; and more complex queries also necessitate the transfer of information between those distinct tables (in SQL typically via a JOIN command), which claims additional working memory and processing time. Overall, the so-called »many-to-many« organisation of data does not lend itself to representation in a relational database.

Figure to explain tables?

The actual performance is determined by the processing power of the machine it is running on and several structural properties, such as organisation, indexing, monotony, and of course the size of the database; therefore, an estimation of processing time for queries is bound to be inaccurate. However, processing times typically do not vary on the scale of orders of magnitude, and thus general estimations can be made. Well optimised SQL databases with a size of 5 to 10 GB on disk usually require tens of minutes if not hours to complete one single complex query;[107] *miRNeo* in its current form takes up approximately 15 GB of storage. Since one analysis typically consists of several hundreds (and, in the case of permutation analyses, several hundreds of thousands) of these queries, processing times in SQL implementation are too long to be practically useful. (It seems important to note that, as of 2018, SQL also offers a graph-based organisation in addition to the traditional, relational layout. These two are separate systems, and not to be confused. The advantages of Neo4j as explained in the following should be seen from the perspective of 2015, when the database was established, and when there was no graph-based SQL implementation.)

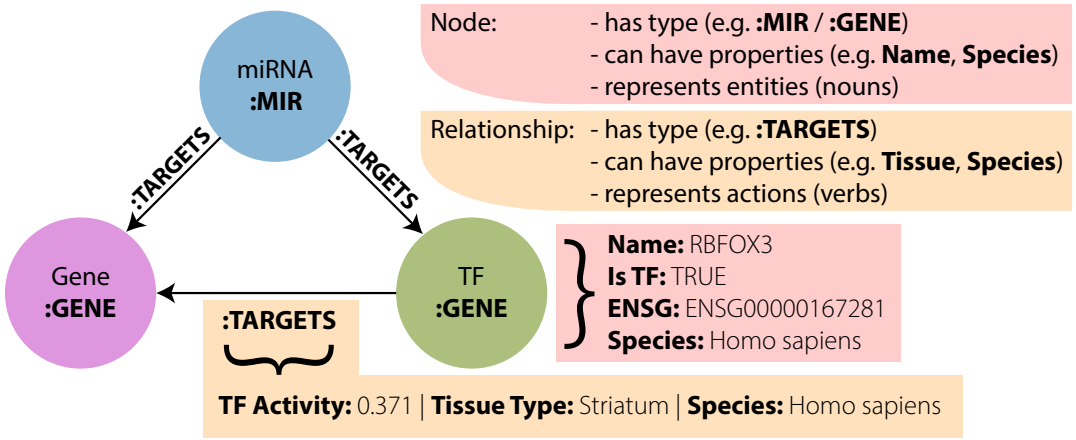

Figure 2.1: Organisation of a graph database. A graph consists of two basic building blocks: *Nodes*, representing entities, and *edges*, representing connections between entities. Each database entry (node or edge) is an instance of a particular *type* and can possess an arbitrary amount of *properties* detailing its specifics.

2.1.1 Neo4j: A Graph-Based Infrastructure

To query and display biological data that are organised in a network-like structure (many-to-many), a database that lends itself to the efficient processing and storage of network data is optimal. »Neo4j« utilises a database structure that is built on the save and recall of data points in *nodes* and *edges*, which represent entities (nodes) and relationships between those entities (edges); both nodes and edges can have any number of attributes and a unique property called »type«, usually describing the class of the entry (such as *gene* or *miRNA*). This database organisation replicates the network-like structure of the biological data studied (Fig. 2.1). Neo4j combines this network-like data structure with an efficient indexing system for quickly finding the entries queried for, and then »walks« along the edges of the nodes that have been found, thus only searching and returning the data that is relevant to the current query. Theoretically, this makes the database more likely to be efficient in the setting of transcriptional interactions, an estimation that turned out to be true.

Depending on the input, these queries can also be rather large; however, the main pitfall of tabular databases such as SQL is circumvented: there is no need to process entire rows or columns of the table to make sure that the query is satisfied in its entirety. This is particularly useful in a setting of sparse information. To illustrate: only 30 of the 2588 miRNAs target a specific gene, which is common; a relational database, after finding the index of the queried gene, would have to search 2588 fields for 1/0. The graph database, on the other hand, has to execute only 30 searches (or, more accurately, 30 »walks« along the edges connected to the indexed node). In practice, even in the very first prototype implementations, this accelerated standard-case computations immensely, and was even able to accommodate advanced approaches in situations that had been inaccessible in the tabular implementation.

2.1.2 High-throughput Database Generation

Neo4j provides several API (»application programming interface«) possibilities in implementation. For the purpose of entering large amounts of data into the database at once, the Java implementation is superior to the other forms in that it provides a batch processing mode via its `BatchInserter` class. I thus wrote a custom Java program for the purpose of creating an initial state of the database from the largest set of data, the complete miRWalk 2.0 content with 12 algorithms and validated interactions. The downloaded data was organised in a plain text based file format, with one text file for each miRNA, totalling in size about 6 GB (for *H. sapiens*). The database was set up in a way that allows only one node for each individual miRNA and gene entered to avoid duplications, using the commands

- `createDeferredConstraint()`
- `assertPropertyIsUnique()`
- `createDeferredSchemaIndex()`

of the Neo4j Java package. This approach made sure to create only one node for each miRNA (type: MIR) and gene (type: GENE) in the data, which is essential for proper functioning of the database. Each of these nodes received several properties to store individual data, such as the various gene/miRNA identifiers, miRNA sequence, and species.

Between those basic nodes, the batch insertion process created edges for each relationship that was found in the original data, assigning a type identifier to each edge detailing the origin of this interaction (type: name of the prediction algorithm or »VALIDATED« for experimental data). Thus, while the nodes for genes and miRNAs themselves are unique, an arbitrary number of relationships can exist between any two nodes, depending on how many interactions they share.

2.1.3 Maintenance and Quality Control

All additional datasets, such as the TF regulatory circuits or tRF targeting predictions, were entered into *miRNeo* using the regular operation mode. Testing was also performed in regular operation, with manual as well as automated tests to assert the correct transfer of information from raw data to the graph database, and to avoid unpredictable behaviour. At times, conflicts had to be resolved manually, for instance when miRNA names conflicted between old »miRNA*« and new »3p/5p« notation; all manual edits are documented in the code, which was published alongside Lobentanzer *et al*.[1]

Except for the rapid import of large amounts of data in creation of a database, the Java implementation of Neo4j does not offer many advantages over the native R implementation, »RNeo4j«. Thus, after creation and a short period of experimentation with graphical user interfaces, I abandoned the Java program in favour of the more flexible R programming. While Java is an object-based programming language, whose benefits lie in extreme flexibility in regards to platform and purpose, high

modularity, and speedy processing, R as a procedural language is the work horse of modern bioinformatics. Its procedural design (the division of data and functions that operate on that data) facilitates the transfer of approaches between distinct datasets, and the enormous vibrant community of data scientists using R provides a wealth of third party packages to tackle almost any bioinformatic task. In the remainder of this dissertation, all analyses are performed in R, unless specifically stated otherwise.

2.2. Materials

All materials used in the creation of *miRNeo* have been acquired from resources that are non-commercial, web-available, and open-source (in the case of code). All properties and relationships derived from this data were entered into *miRNeo* as either nodes, properties of nodes, edges, or properties of edges.

2.2.1 Gene Annotation

Even though »regular« protein coding genes have been known for a long time, there is no consensus yet about their nomenclature and organisation. Complicated by newly discovered functions and properties of phylogenetic nature, the scientific representation of the human genome is in constant flux. Several large organisations strive to provide a robust annotation of the human gene catalog, but also in many cases contradict one another. There are three nomenclature systems that are of high importance in modern genomics:

- The traditional naming system of acronyms (e.g. CHAT) and fantasy-names (such as »Sonic Hedgehog«), also occasionally called »gene symbol«, is still widely popular because of its accessibility to humans, but is also not particularly robust because of a high amount of synonyms with high confusion potential (see e.g. Section 1.2.7 on CDF) and instances of genes without names having to carry unwieldy systematic names. Gene symbols are largely, but not exclusively, curated by the HUGO Genome Nomenclature Consortium (HGNC), under the roof of the Human Genome Organisation (HUGO). As such, there also exists an »official« HGNC symbol for many genes, but these are not consistently used throughout the literature.
- The American National Center for Biotechnology Information (NCBI), a branch of the National Institute of Health (NIH), curates and hosts a multitude of biological and medical data, and for the organisation of gene information uses its own systematic nomenclature termed »Entrez« ID. Entrez is a molecular biology database that integrates many aspects of biology and medicine in a gene-centred manner, and therefore Entrez IDs are useful to quickly connect a gene to its function, nucleotide sequence, or associated diseases. Entrez IDs are regular integers without additional characters.
- Akin to the NCBI effort, ENSEMBL is a project of the European Bioinformatics Institute (EBI) as part of the European Molecular Biology Laboratory (EMBL). Compared to the Entrez database,

it is more focused on study and maintenance of the genome itself, and therefore has a more intricate nomenclature that allows for differentiation of, for example, genes and their various transcript isoforms (ENSEMBL IDs carry character prefixes for class identification, e.g., ENSG for genes, ENST for transcripts).

All of these are being used on a regular basis in many publications, and, often, they are used exclusively. As a result, the end user of the published data has to have access to all possible annotation forms, or, at least, a means to translate one into the other; often, this also introduces conflicts. For this reason, all ID types were entered into *miRNeo* upon creation or during maintenance, for convenience and to minimise analysis prolongations due to conflict resolution.

2.2.2 microRNA Annotation

miRBase provides a consistent annotation for miRNAs. Due to their relatively recent discovery, there still are major changes from version to version; the syntax, however, is stable. In addition to the miRNA »names« that are composed of species, the string »miR«, pre-miRNA designation number, and strand origin (not in all cases!), such as »hsa-miR-125b-5p«, miRBase provides IDs for pre-miRNA molecules (also called ancestors) termed »MIID«, and IDs for mature miRNA molecules termed »MIMAT«. However, in practice, these are rarely used. Similarly, miRNA families are annotated using the »MIPF« ID.

2.2.3 Transcription Factor Targeting

The FANTOM5 project has applied 5' cap analysis of gene expression (CAGE) to a large number of human samples from diverse tissues to determine the accurate 5' ends of each transcript.[108] Knowledge of this fact enables accurate prediction of transcription factor binding sites likely to control a transcript's expression. Marbach and colleagues used this information in combination with detailed human gene expression data to derive a complex interaction network of TFs and genes (»regulatory circuits«), and in doing so aggregated samples with similar expression patterns and origins into 394 fictional tissues.[103] For every tissue, each TF was assigned transcriptional activities towards all genes that it supposedly targets (with the sum of all activities in any given tissue being 1). Marbach and colleagues have shown that the cumulative transcriptional activities towards any given gene correlate well with the actual gene expression in corresponding samples from an independent repository.

Even in its fifth iteration, FANTOM data is not entirely comprehensive, which came to my attention due to a cholinergic anomaly: The 5' CAGE peaks of the *CHAT* and *CHRNA7* (the nicotinic $\alpha 7$ receptor subunit) genes in raw FANTOM5 brain tissue data do not pass the expression threshold, and therefore are not included in, e.g., Marbach's »regulatory circuits«. Both are critically important not only for neuronal cholinergic systems, but also for the non-neuronal aspect of immune processes. For instance, macrophages have been shown to produce ACh via ChAT, and the $\alpha 7$ homomeric ACh

receptor conveys direct immune suppression by its expression on monocytes.[38] Paradoxically, the CAGE peak of *SLC18A3*, which lies in the first intron of *CHAT*, crosses the threshold and therefore is included in the data. Unfortunately, I was not able to remedy these circumstances even upon personal communication with Daniel Marbach (author of »regulatory circuits«) and Hideya Kawaji of the FANTOM5 consortium, although the latter acknowledged the possibility of a gene annotation deficit leading to misattribution of the *CHAT* signal to *SLC18A3* due to the closeness of their 5' ends. Thus, it seems viable to substitute *SLC18A3* targeting data for the absent *CHAT* data in certain situations.

The entire collection of transcriptional activities in all tissues was downloaded from the project's web page,[103] and neuronal and immune tissues were manually curated and entered into *miRNeo*. The collected data comprises 33 neuronal tissues and 26 immune cell tissues (Appendix A), and 1 130 196 TF→gene relationships in total (not all 394 tissues were entered due to the time requirements).

2.2.4 microRNA Interactions

The content of miRWalk 2.0 is freely available online;[109] however, there is no option of downloading the complete set. The targeting data thus was downloaded per miRNA using a custom crawler, with standard options for all 12 prediction algorithms (miRWalk, miRDB, PITA, MicroT4, miRMap, RNA22, miRanda, miRNAMap, RNAhybrid, miRBridge, PICTAR2, and TargetScan) in plain text format. For experimentally validated interactions, the main sources were DIANA TarBase[110] and miRTarBase,[111] both of which offer complete download options. As of 2019, the 3.0 version of miRWalk allows complete species downloads; however, the developers have abandoned their third party algorithm plurality reducing the number of available alternatives from 12 to 4, which can be considered a significant disadvantage:

While sequence complementarity, particularly of the »seed«-region, is the primary paradigm of miRNA-mRNA interaction, prediction algorithms vary widely in their implementation, general purpose, and approach to interaction prediction (for a comprehensive review of approaches and rules, see Yue *et al.*[112]). A large group of available options utilise sequence conservation aspects to increase candidate viability (such as miRanda, PicTar, TargetScan, and microT4). Others, such as RNA22 and PITA, utilise biophysical aspects such as free energy of binding or the accessibility of target sites due to secondary RNA structures as prediction arguments. All of these approaches have their up- and downsides, e.g. considering their general precision and sensitivity, or their adequate prediction of particular cases, such as multiple site targeting. Thus, it has been proposed to use a combination of complementary approaches instead of only one algorithm per analysis.[113] For this reason, I may have preferred the 2.0 version of miRWalk, even if 3.0 had been available at the time.

One advantage of the collection of all data in a quickly accessible database is the opportunity to compare the different approaches to target prediction. A statistical evaluation of the collected inter-

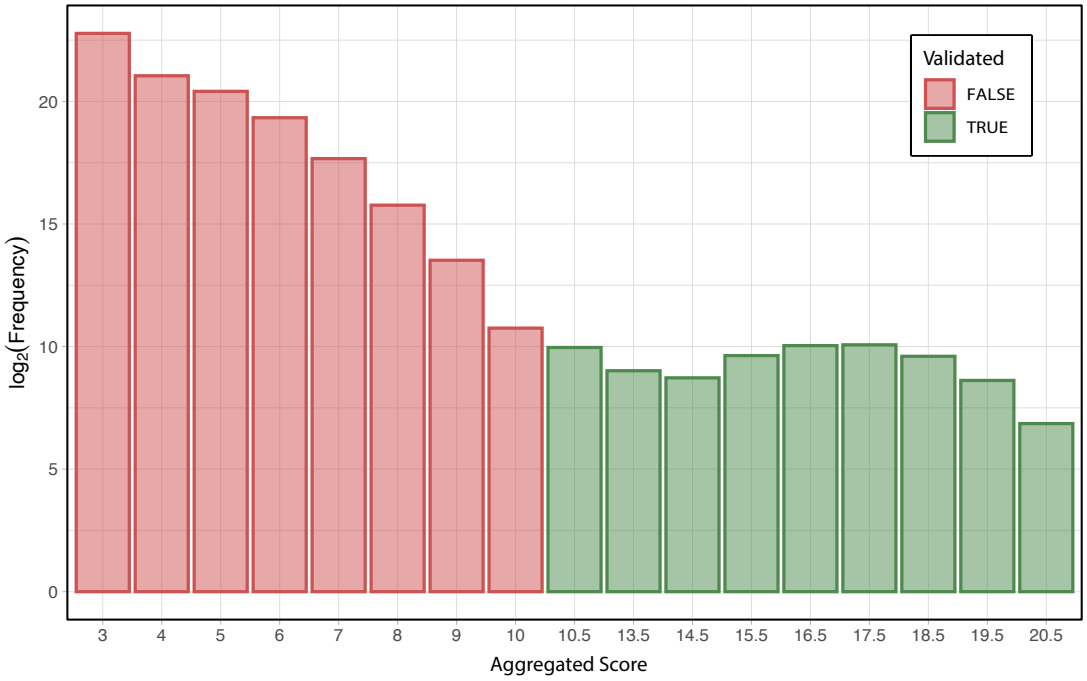

Figure 2.2: Histogram of miRNA→gene score distribution. Aggregation of individual algorithms yields a score range of 3 to 10 per individual miRNA→gene interaction. In case of additional existence of experimental validation (evidence level high) for any predicted interaction, score is increased by 10.5. The distribution shows a sharp decrease in predicted interactions towards higher scores, and a maximum of validated interactions at prediction scores 6 and 7.

action data from miRWalk 2.0 showed vast differences in general prediction quantity (Table 2.1) as well as prediction accuracy and sensitivity when compared to the validated subset of data (Table 2.2). Since the ground truth is not known, this is an additional argument for the combination of multiple algorithms instead of the use of a single set. Apart from RNAhybrid and miRBridge, all algorithms presented reasonable base hit frequencies and increases in the validated test set. While miRBridge already has the lowest positive frequency of all the algorithms, it is the only one to achieve a negative score in the validated test set. On the other hand, RNAhybrid has a vastly higher base hit frequency than the second highest scoring algorithm (by more than 300%), making it very likely to produce false positive results, and less valuable in the aggregation scoring system. The remaining 10 algorithms were included in *miRNeo* targeting data. For ease of use, an additional relationship type was created from the aggregated single algorithm hits of any miRNA→gene relationship, with the sum of algorithms predicting the interaction as a score variable. This yields a theoretical score range from 3 to 10 (miRNA→gene relationships with only one or two hits were ignored for the sake of space). To account for experimentally validated interactions, each miRNA→gene relationship that was supported by strong evidence of interaction was modified by addition of 10.5 score points (a half point for quick identification of a validated relationship), extending the maximum score to 20.5 points. The resulting optimised graph contains 11 687 931 human miRNA→gene targeting relationships with a distinct score distribution (Fig. 2.2). In comparison, only 6146 miRNA→gene relationships are experimentally validated with »strong« evidence.

algorithm	hit frequency
RNAHYBRID	71.62%
MIRMAP	19.90%
MIRWALK	19.74%
TARGETSCAN	16.33%
RNA22	12.34%
MICROT4	11.81%
MIRANDA	10.65%
PITA	4.90%
MIRDB	1.17%
MIRNAMAP	0.75%
PICTAR2	0.62%
MIRBRIDGE	0.15%

Table 2.1: Prediction algorithms ordered by the fraction of all possible interactions they predict as being real (positive rate). Different algorithms display a wide variation of hit rates in the entirety of predicted interactions between any miRNA and gene. Red: excluded from analysis.

algorithm	validated hit frequency	hit rate increase
PICTAR2	6.98%	1129.40%
MIRDB	9.80%	838.43%
MIRANDA	51.73%	485.94%
TARGETSCAN	70.63%	432.51%
MIRNAMAP	3.10%	410.95%
PITA	15.57%	317.20%
MICROT4	32.60%	276.10%
MIRMAP	53.86%	270.65%
MIRWALK	50.95%	258.15%
RNA22	22.51%	182.38%
RNAHYBRID	90.47%	126.32%
MIRBRIDGE	0.01%	0.00%

Table 2.2: Prediction algorithms ordered by their increase in true positive rate when considering only validated interactions. The hit rate increase when comparing experimentally validated interactions with the entire predicted data (Table 2.1) is also subject to strong variation. Hit rate increase is the increase of hit rate if only considering validated data as opposed to all predicted interactions. None of the studied algorithms unite a good precision (hit rate increase) and coverage (validated hit frequency).

2.2.5 Filtering of Aggregated Prediction Scores

For the estimation of the »true« miRNA→gene interactions in the predicted-only data in *miRNeo*, two premises are relevant: First, the enormous amount of hits with a score of 3 in all likelihood is an over-estimation, and second, the amount of currently validated interactions can be but a small fraction of »true« interactions. Assuming the truth lies on the axis between these two extremes (i.e., at some score value inside the *miRNeo* interactions), the true amount of human miRNA→gene interactions must approximately fall within the range of 2^{10} to 2^{20}. Looking at the score distribution of all *miRNeo* interactions (Fig. 2.2), the maximum amount of validated interactions is predicted by a combination of 6 or 7 algorithms (i.e., a score of 16.5 or 17.5). Thus, to approximate the true state, I chose to apply a low-cut filter to *miRNeo* queries at a minimum score of 6. This is the standard case referred to as »*miRNeo* query« in the remainder of this dissertation. In some cases, such as the graphical analysis of whole-genome miRNA targeting (see e.g. Section 3.6), the score threshold was raised to 7 to circumvent computational limitations.

2.2.6 De-novo Prediction of tRF Targeting

Due to the recency of their (re-)discovery, no comprehensive interaction sources exist for transfer RNA fragments. There have been documented cases of miRNA-like behaviours of distinct RNA fragments,[90,96] justifying an attempt to predict interactions in a comprehensive manner. Of the available options for nucleotide interaction prediction algorithms, TargetScan[114] seems particularly suited for this task because it provides the option of evaluating the evolutionary conservation of target

sites in the putatively targeted genes, thereby providing an additional layer of security: The sequence of 3' UTRs is evolutionarily less stable than the coding part of genes; thus, high conservation of the binding site may indicate evolutionary pressure to keep up the interaction with the fragment, making an actual function of the interaction more likely. TargetScan also presents with reasonable sensitivity and specificity as confirmed by an independent group,[115] and through an additional algorithm allows the attribution of a score based on the branch length (on the species tree) of conserved targeting.[116]

miRNA-like behaviour implies the existence of a region on the tRF similar to a miRNA »seed«, and TargetScan also expects a seed as input to its targeting algorithm. Since there has been no definitive answer to the question as to where the seed region in tRFs may be, it is safest to assume nothing and explore all possibilities, i.e., simulate every possible seed position for interaction discovery. For this purpose, all discovered sequences of tRFs (exceeding a base mean expression of 10 counts) were chopped into 7-nt pieces (7mers), which is the length of miRNA seeds, and statistically improbable enough to appear in the genome at random; the average length of a human 3' UTR is 800 nt, so the probability of finding any 7mer randomly in any one 3' UTR is $p = \frac{800}{4^7} = 0.049$, which agrees with the 5% false discovery ratio (FDR) convention.

Describe Targetscan process

2.2.7 microRNA Primate Specificity

During the course of evolution, higher organisms typically attained more complexity in a variety of functional categories. The CNS as the system of highest complexity underwent several drastic developments from invertebrates to lower mammals to higher mammals still. miRNAs are no exception. While many miRNAs are functionally as well as literally conserved in all mammals, primates in particular have gained a substantial amount of novel miRNAs whose function is in large parts elusive. Due to the restrictions on experimentation on higher mammals, particularly primates, many of those miRNAs can only be studied observationally, or by transgenic experiments in rodents. A cholinergic example of a gain-of-function in higher mammalian miRNA regulation is the vesicular acetylcholine transporter, SLC18A3. As described in Section 2.2.3, the SLC18A3 gene is situated in the first intron of CHAT, and thus is always primarily co-expressed with the latter. However, a primate-specific miRNA, miR-298, targets the 3' UTR of SLC18A3.[117] Thus, the primate neuron has gained a mechanism of independent SLC18A3/CHAT regulation that the mouse, for example, does not possess. It is easily imagined that such a gain of neuronal flexibility, in many instances, can aid the development of a more effective brain. However, the primate specificity of miRNAs is not yet consensus, and thus not found in annotation databases such as miRBase, even though they list all miRNAs discovered in any species. To get an impression of the amount of possible gain of function, I performed a review of miRNAs expressed in a representative variety of annotated species. From hereon out, largely method-related paragraphs will be set in sans-serif font face.

Species Selection

The tested species were selected from miRBase v21. Some of the available species are severely limited in the extent of miRNA annotation, likely because of a research bias. Therefore, only the most well-annotated species were selected. These are (number of annotated primary and mature miRNAs in brackets):

- *Homo sapiens* (human; 1881, 2588)
- *Gorilla gorilla* (gorilla; 352, 357)
- *Pan troglodytes* (chimp; 655, 587)
- *Pongo pygmaeus* (orangutan; 642, 660)
- *Macaca mulatta* (rhesus macaque; 619, 914)
- *Bos taurus* (cow; 808, 793)
- *Canis familiaris* (dog; 502, 435)
- *Mus musculus* (mouse; 1193, 1915)
- *Rattus norvegicus* (rat; 495, 765)

The first four species belong to the hominid group; the first five are primates. It is likely that these collections are not complete, with the degree of completeness depending on the amount of research performed on the species (as demonstrated, e.g., by the difference between mouse and the other non-primates). This places considerable difficulty on asserting primate specificity of miRNA, and in turn on assertion of the effects of evolution on the miRNA regulatory system.

Single miRNA Inter-Species Homology Computation

To determine the homology of miRNAs between the studied species, reference genomes were downloaded from the respective sources and analysed phylogenically, using the genomic coordinates provided by miRBase. Sequence homology was determined via dynamic programming using the Smith-Waterman algorithm.[118] Briefly, this algorithm can be used to determine the similarity of two genomic sequences, based on a scoring system rewarding matches and penalising mismatches. Smith and Waterman extended the original approach by Needleman and Wunsch,[119] which is used to compare two complete sequences. Both algorithms rate an alignment by dynamic scoring inside a 2D-matrix, with the sequences to be compared as the x- and y-axes (one letter per cell). By a change in the scoring system, the Smith-Waterman algorithm finds the best local alignments, instead of comparing the two sequences in their entirety. In the case of miRNAs, this behaviour is useful because, between species, there are frequent additions or deletions of single nt on both ends of the homologous miRNA. Genomes were procured from the following sources:

- *Homo sapiens*: GRCh38 (NCBI)
- *Gorilla gorilla*: gorGor3 (UCSC)
- *Pan troglodytes*: panTro4 (UCSC)
- *Pongo pygmaeus*: PPYG2 (Ensembl)
- *Macaca mulatta*: rheMac3 (UCSC)
- *Bos taurus*: bosTau6 (UCSC)
- *Canis familiaris*: canFam3 (UCSC)
- *Mus musculus*: mm10 (UCSC)
- *Rattus norvegicus*: rn5 (UCSC)

Using the genome coordinates provided by miRBase, the genomic sequences of miRNAs and pre-miRNAs of each species were determined. Using the Smith-Waterman algorithm, all identified homologs of human miRNAs were subjected to homology scoring, and score results were visualised as a heatmap.

Inter-Species Distribution of miRNAs

The inter-species relationships of annotated miRNAs do not follow a simple evolutionary distribution from less complex to more complex organisms, but rather seem to partially result from parallel development (Fig. 2.3). Taking into account the high probability of missing annotations in several species (particularly hominids), it seems prudent to define primate specificity of miRNAs not by presence in primates, but rather by absence of the miRNAs in non-primate species (also excluding miRNAs *only* annotated in human). Thus, primate specificity of a human miRNA is assumed if the miRNA is expressed in at least one primate species, and absent from all non-primate species in this roster. This definition yields a list of 377 primary and 350 mature putative "primate specific" miRNAs in miRBase v21 (Appendix B). Judging from recent analyses,[105] there probably exist many more. The primate-specificity attribute was entered into *miRNeo* as miRNA node property.

elaborate?

2.3. miRNeo Usage

Neo4j uses a language (called »Cypher«) akin to SQL, which utilises keyphrases to issue commands, but combines it with a semi-graphical syntax to account for the graph-based layout of the data. In the following, I will describe its basic usage and the advantages it provides in the matter of transcriptional connectomics. The basic »finder« function (similar to SELECT in SQL) is called MATCH in Cypher, and, when combined with the semi-graphical syntax, can be used to identify nodes or more complex patterns in the database. The graphical syntax consists of two main building blocks that represent the basic types of data inside the database: nodes as regular brackets »`( )`« and edges between nodes as a construct of hyphens and box brackets, that can also have a direction indicated by the greater sign »`( )-[ ]->( )`«. To specify the elements to be found, attributes of nodes and/or edges can be filtered by using curly brackets in the node definition, or the WHERE clause. To be returned, elements need to be assigned arbitrary variable names:

Listing 2.1: MATCH

```
1 MATCH (gene:GENE {species: 'HSA'})
2 WHERE gene.name = 'CHAT'
3 RETURN gene
```

Query 2.1 identifies a node (arbitrarily designated »gene«) with type GENE (indicated by the colon), with attributes »species« (HSA, i.e. *H. sapiens*) and »name« (CHAT), and returns the node

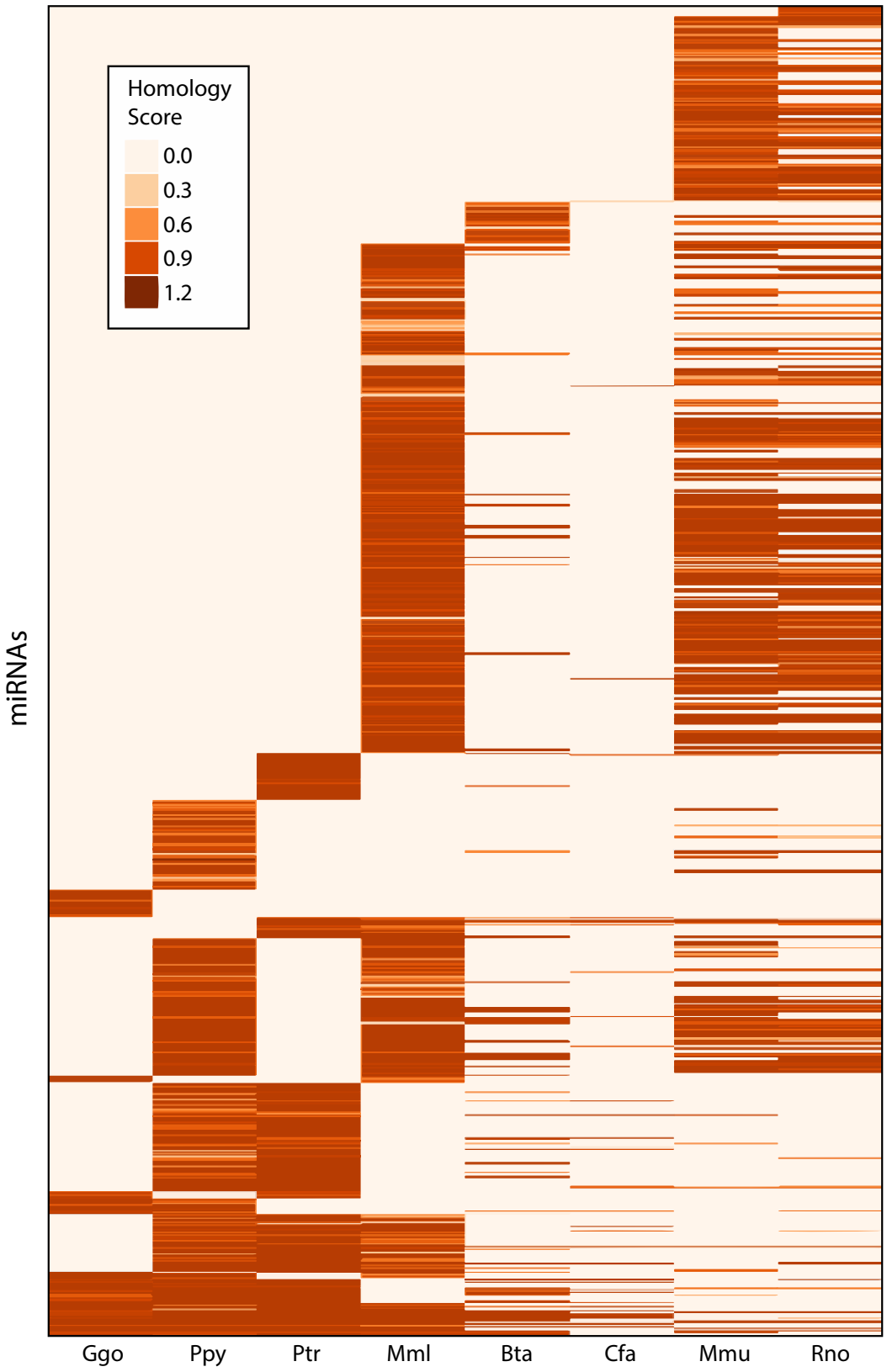

Figure 2.3: Homologues of Human microRNAs in Primate- and Non-Primate-Species. Homology to human miRNAs was determined by Smith-Waterman local alignment for each homologous miRNA of 8 species. Homology scores were visualised on a heatmap, each column represents the homology to human of the miRNAs of the respective species. The heatmap is ordered from bottom to top by the amount of miRNA homologues in primates. The miRNAs at the very bottom are shared by human as well as all four primate species, followed by the miRNAs shared by three primate species, and so on. Ggo: *Gorilla gorilla*, Ppy: *Pongo pygmaeus* (Orangutan), Ptr: *Pan troglodytes* (Chimp), Mmt: *Macaca mulatta* (Rhesus macaque), Bta: *Bos taurus* (Cow), Cfa: *Canis familaris* (Dog), Mmu: *Mus musculus* (Mouse), Rno: *Rattus norvegicus* (Rat).

with all its attributes. Since the nodes of type GENE are restrained, there can only be one gene of species *H. sapiens* with this name in the database, and thus, only one data point will be returned. The graphical syntax further allows for pattern matching of, for instance, miRNA→gene relationships:

Listing 2.2: Patterns

```
MATCH (mir:MIR)-[rel:TARGETS]->(gene:GENE {species: 'HSA'})
WHERE gene.name = 'CHAT'
RETURN mir, rel, gene
```

Query 2.2, similar to query 2.1, starts by identifying the node of species HSA with the name CHAT, and proceeds to look for miRNA→gene relationship edges arriving at this node; the relationships have to be of the type TARGETS (the pre-aggregated score-based accumulation of targeting). As soon as no further edges are found, the process terminates and returns all found miRNAs (»mir«), relationships (»rel«), and genes (»gene«) in discrete form, including all their attributes, such as the ENSG and Entrez IDs, the MIMAT IDs for all found miRNAs, or the score value of their targeting relationship. In this query, since there is a constraint on genes, the only gene returned is *CHAT*. However, Cypher is not limited to filtering on unique attributes; it allows for query and return of as many data points as are needed. For example, if one is interested in all miRNA→gene interactions in the cholinergic system, the query may look as follows:

Listing 2.3: Filtering

```
MATCH (mir:MIR)-[rel:TARGETS]->(gene:GENE {species: 'HSA'})
WHERE gene.name IN {cholinergic_genes}
RETURN mir, rel, gene
```

The effectiveness of graph-based databases becomes clear in this approach: Query 2.3 is processed starting at a user-defined filter, the list of cholinergic genes as an input (containing *CHAT*, *SLC18A3*, cholinergic receptor genes, acetylcholinesterase, etc). In a first step, all nodes are found that fulfil the criteria: type GENE, from species *H. sapiens*, that are in the list of names given. Since the gene nodes are indexed, this only requires milliseconds. Then, through the connection of edges to these nodes, it finds all miRNA nodes that have a miRNA→gene relationship towards any of the cholinergic genes. By using the gene nodes as starting point, the query can end as soon as no other edges fulfilling these criteria are found on any of the nodes. In comparison, to satisfy this query in a relational database, the rows representing these cholinergic genes would have to be assessed in their entirety, not only in those columns that represent an extant relationship, thus prolonging execution.

The database then returns all miRNA→gene relationships in this set, representing the network of cholinergic miRNA regulators, including all of their attributes. The advantages of graph-based data do not end there; say one wants to return only »master« regulators of cholinergic systems, defined as

miRNAs that target at least 4 of the genes in the cholinergic set. In a relational database, this would have to be done post-hoc, by aggregation of relationships and removal of any results that do not exceed this threshold. This requires storage of the entire result in memory, and additional computational steps that can be very taxing depending on the size of the result table. In Cypher, this can be done during the query (code comments indicated by »//« explain single steps):

Listing 2.4: Two-stage Filtering

```
MATCH (gene:GENE {species: 'HSA'})
WHERE gene.name IN {cholinergic_genes}
WITH gene //the found genes are used as input for the second query
MATCH (mir:MIR)-[rel:TARGETS]->(gene)
WHERE count(rel) >= 4
RETURN mir, rel, gene
```

Query 2.4 essentially proceeds in the same way as query 2.3 in that it identifies the gene nodes filtered for and looks for the miRNAs connected to those nodes by TARGETS-type relationships; however, in the second step (which is performed per gene node as returned by the WITH clause), it returns only those patterns that have at least 4 incoming miRNA→gene relationships. Query 2.4 only requires little additional processing compared to query 2.3, and thus does not require nearly as much time as the post-hoc filtering required in a relational database query. This filtering can be applied in many stages, and in many forms, such as sums, averages, maximum and minimum, or other combinations of arithmetic and logical classifiers. Additionally, the patterns can be extended to represent complex relationships inside the graph. For instance, the following query 2.5 was used to find miRNAs that regulate any given gene in the database, and, simultaneously, affect TFs that are involved in regulation of this same gene (this type of interaction is called feedforward loop, see also Section 4.5).

Listing 2.5: Feedforward Loop Identification

```
MATCH (gene:GENE) //find gene
WHERE gene.id = ID //by identifier (Entrez)
WITH gene //use as input for next step
MATCH (tf:GENE {species: 'HSA', tf:TRUE})-[rel]->(gene)
//find TFs targeting that gene
WHERE type(rel) IN {tissue_types} //TFs only from specific tissues
//for instance, CNS cell types (Appendix A)
WITH gene, rel, tf //use as input for next step
MATCH (tf)-[rel]->(gene)<-[rel_m1:TARGETS]-(mir:MIR {species:
    'HSA'})-[rel_m2:TARGETS]->(tf)
```

```
//find miRNAs that target gene and gene-targeting TF at the same time
WHERE rel_m1.score > 5 AND rel_m2.score > 5
//low-cut filter at a minimum cumulative score of 6
RETURN gene, tf, rel, type(rel) AS tissue, mir, rel_m1, rel_m2
```

This analysis can be performed in real time, on the whole genome and miRnome, and merely takes seconds for one iteration, a performance unimaginable in a relational database approach; advanced statistical approaches such as permutation only become viable at this timescale.

2.4. Statistical Approach to Transcriptional Connectomics

The enormous amounts of data generated by modern molecular biology methods, such as RNA-seq and bioinformatics, present new challenges to statistical methodology. A major objective in the analysis of large datasets is a robust statistical representation of the distribution of this data. Traditionally used approaches such as Student's t-test are not automatically applicable to the intermediary results of these modern methods, because the premise of a normal distribution often does not hold, or has to be proven first. This section will describe the statistical problems encountered in the analysis of intermediary data produced by *miRNeo*; the statistical properties of large count data directly generated by RNA-seq will be discussed in Sections 3.4.3 and 5.1.2.

2.4.1 Permutation

The evaluation of comprehensive prediction datasets regarding miRNA→gene interactions on a genome scale is statistically challenging. Molecular interaction studies have explored only a minority of all possible targeting relationships, and as such, the ground truth of miRNA→gene interaction is unknown (see Section 2.2.4). Since there is no negative interaction data, validated interactions can only be defined in the positive space. Additionally, the various prediction algorithms also heavily diverge in their predictions, which leads to the question of how to approach the estimation of false discovery ratio (FDR) while simultaneously avoiding high false negative rates.

One possible approach that can aid in identification of the most pertinent effects in this case is random permutation. In this approach, the result of an analysis (e.g., a numeric targeting score of a miRNA→gene interaction, or a Spearman correlation between two gene sets) is compared to a null distribution that was generated from an iterative analysis similar to the initial one, but with randomised input (e.g., a group of miRNAs of the same size as the original set, randomly selected from all miRNAs, or the gene sets from the original analysis with randomly scrambled group affiliations). This permutation of the analysis is performed many times (usually between 10 000 and 1 000 000 iterations, depending on the context), and results in a distribution of possible outcomes that can be arranged from lowest to highest, often resulting in a normal (or »normal-like«) distribution, thus facilitating the estimation of confidence intervals, and, similarly, p-values for the »real« result.

A positive side-effect of performing a permutation analysis on a base collection of data, such as *miRNeo*, is the automatic correction of inherent biases. For instance, should a particular gene by its genetic structure invite a large amount of false positive predictions as to the miRNA→gene interactions towards it, these will be

present in the test as well as in the permutation comparison, and thus cancel out and yield a high p-value for this interaction, effectively transforming the false positive into a true negative. For further discussion, see Section 5.1.3.

2.4.2 Gene Set Enrichment Analysis

The objective of gene set enrichment is the identification of statistically over-represented entities in a dataset. The standard use case in biomedicine is the Gene Set Enrichment Analysis (GSEA), that is used to identify the most important classes of genes in large datasets, such as the ones produced by RNA-seq. Briefly, the analysis follows these steps: the studied genes are scored by a certain method, such as p-values from differential expression analysis, which enables the identification of a relevant subgroup, the test set (e.g., the 100 genes with lowest p-values). This test set is then compared to a background of genes (usually, all detected genes, or a large amount of genes from the entire dataset) by a statistical method fit to determine their enrichment in pre-defined categories. Often, ontological categories are used, such as the »biological process« type of Gene Ontology (GO), or KEGG pathways.

For each of these categories, the method tests for a representation of genes in the test set exceeding the frequency statistically expected by random sampling from the background of genes; thus enabling an estimation of the functionality these test set genes may inhabit in the process that is studied. Statistical approaches often employed in gene set enrichment are Kolmogorov-Smirnov statistics, permutations, or, more generally, hypergeometric tests such as Fisher's exact test. There are a wide variety of software solutions available for the implementation of gene set enrichment testing.

Gene Ontology curates an enormous catalogue of coding gene products and their functions. At the current time, GO hosts 7 330 378 annotations (2 836 377 for »biological process«, 2 289 165 for »molecular function«, and 2 204 836 for »cellular component«), subdividing 1 405 197 individual gene products from 4493 species (205 with more than 1000 annotations) into 44 733 ontological terms (29 457 »biological process«, 11 093 »molecular function«, and 4183 »cellular component« terms). The individual GO categories are organised in a hierarchical manner, more specifically, a directed acyclic graph (DAG). Each branch of the DAG tree contains related terms, progressing from the most general terms (top) to the most specific ones (at the bottom).

Whenever a GO analysis is described in chapters three and four, it means a gene set enrichment analysis performed on a particular subset of genes (that may e.g. be the targets of a group of miRNAs) towards the elucidation of their biological function, i.e., the »biological process« category of GO annotation. For further discussion, see Section 5.1.6.

One of the difficulties in understanding the brain is that it is like nothing so much as a lump of porridge.

Richard L. Gregory

3

microRNA Dynamics in Cholinergic Differentiation of Human Neuronal Cells

This chapter will discuss the current state of knowledge on brain transcriptomics, generally and in the specific case of cholinergic neurons in the CNS, and then go on to explain the steps we undertook to elucidate small RNA processes in central cholinergic systems. First, our aim was to clarify co-expression patterns of central cholinergic neurons, which required analysis of transcriptome data in single-cell resolution. Based on this information, we selected two human models of cholinergic neuronal differentiation and established a differentiation protocol amenable to RNA extraction and successive molecular biology assays, most importantly, RNA-seq. The expression patterns so obtained were then used to perform bioinformatics analyses using the database introduced in Chapter 2, *miRNeo*.

3.1. Neuronal Transcriptomes - Background

The mammalian brain requires a constant supply of oxygen and nutrients, because it does not provide storage for either. Though it only makes up approximately 2% of the entire human body mass, its energy expenditure is around 20% of the whole.[120] For this reason, each square millimetre of brain tissue (except for the ventricles) is infiltrated by hundreds of capillaries.[121] Since the blood-brain-barrier is essentially provided by supporting glia cells surrounding all capillaries from the »inside« (see Fig. 3.1, modified from Lobentanzer & Klein[122]), neurons numerically constitute only a minority of brain tissues (but burn two thirds of its energy).

Until very recently, studies aiming to clarify the transcriptional profiles of neurons applied either

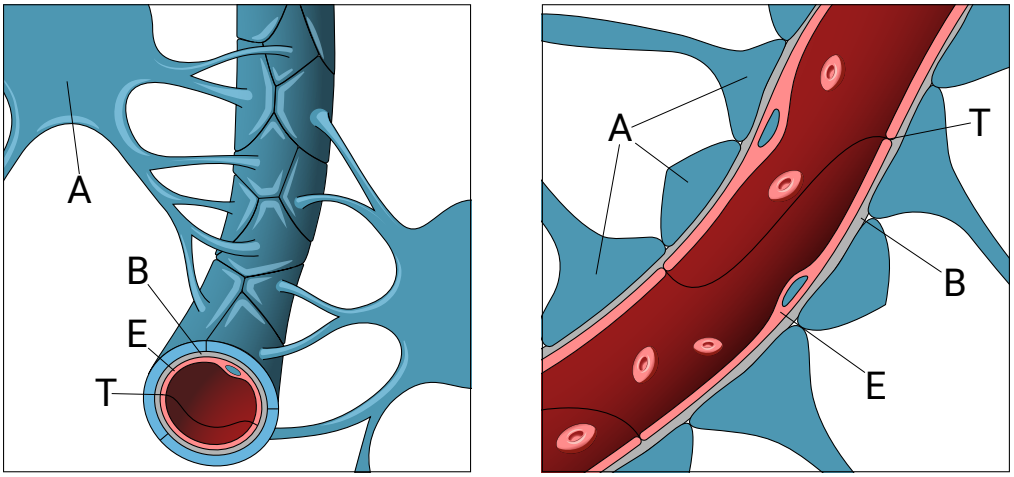

Figure 3.1: Schematic display of the blood-brain-barrier. The blood-brain-barrier surrounds virtually every capillary in the CNS. A: Astrocyte, B: Basal Membrane, E: Endothelial Cell, T: Tight Junction. Modified from Lobentanzer & Klein, 2019. [122]

microarray technology or RNA-seq (also known as deep sequencing or next generation sequencing). For these methods, several cubic millimetres of brain tissue are required at the least; often, cubic centimetres are used. In contrast, the diameter of neuronal somata is usually in the micrometre range. Thus, the resolution of the method and the actual cellular resolution differ by a factor of approximately 1000. Additionally, even among the neuronal population, there is considerable heterogeneity and transcriptomic plurality; single brain regions rarely consist of less than 30 different neuron types, tightly packed next to each other, each with their own transcriptional identity. [123,124,125,126] Newest studies, deciphering the murine nervous system by sequencing of 500 000 individual cells, show that neuron diversity is very similar regardless of brain region. [127] These circumstances hold true for any mammal, and most of our knowledge stems from the analysis of our favourite research animal, the mouse. In humans, the diversity is only exacerbated; in fact, the elevation in CNS complexity, which is only made possible by enhanced transcriptional control, may be the reason for our superior cognitive abilities.

Cholinergic neurons always constitute a minority in any neuronal population, sometimes to extremes. Most tissues are dominated by few neuron types, such as pyramidal cells in the cortex. The most common neurotransmitter types are GABAergic (inhibitory) and glutamatergic (excitatory), each with several subtypes. It is estimated that more than 80% of cortical neurons are excitatory, and more than 90% of synapses release glutamate. [120] There are two major cholinergic regions in the mammalian brain: the striatum is fairly well-populated with rather large cholinergic interneurons, and the basal forebrain holds a large amount of (smaller) cholinergic projection neurons (compare Fig. 1.1). However, in transcriptomic analyses, these tissues are seldom used, maybe due to lack of scientific interest, or because they are notoriously hard to access (the basal forebrain is small and deeply imbedded in the midbrain). The cortex, particularly the neocortex, is most often the tissue of choice in these

studies, due to its scientific interest and accessibility. Though it contains only a miniscule amount of cholinergic interneurons whose transcriptional identity is still a matter of debate, several of the recent single-cell RNA-seq approaches have independently identified cholinergic interneurons in cortical regions (see Fig. 3.2).

3.2. Cortical Single-Cell RNA Sequencing

The impact of transcriptional dynamics on any disease depends on co-expression of the relevant genes in the affected cell. Selection of a model therefore has to take co-expression into account. In particular, if neurokines are to possess any relevance for cholinergic properties of central nervous cells, the cells in question would have to express molecular machinery required to receive neurokine signals. The advent of single-cell RNA-seq for the first time enables the resolution of gene expression on a cellular basis, and thus the disentangling of spatially close individual neuron types (and other, non-neuronal CNS cells); most of this information is lost in RNA-seq performed on brain homogenate, even of a small biopsy. Differences in genes are reduced to the universally expressed »housekeeping« genes, save the most extreme perturbations. In miRNAs, this circumstance is only exacerbated, in parallel to their even more tissue-specific expression.

3.2.1 Single-cell Dataset Processing

To provide a detailed tally of transcriptional subtypes in the CNS, publicly available single-cell RNA-seq datasets of suitable tissues were analysed towards their cholinergic properties. All studies that were available at the time focused on some subsection of the cortex (visual or somatosensory) or the hippocampus. Additionally, the data provided by those studies was in some cases pre-aggregated to represent classes of single neurons with similar transcriptomes (Fig. 3.2 A&B [124,125]); in other cases, every single neuron was represented (Fig. 3.2 C&D [123,126]).

An important quality-related parameter of a single-cell RNA-seq experiment is the sequencing depth achieved per single sequenced cell. Some of the screened datasets do not provide sufficient depth to resolve genes with medium expression, which includes our primary cholinergic markers *CHAT* and *SLC18A3*. The datasets which did provide adequate sequencing depth were filtered for their expression of these markers, and additionally characterised by their expression of common markers for cell types to be expected in the CNS. Raw data were downloaded from their respective sources and imported into the R environment, where they were converted into similar format. The numeric expression values of each dataset were normalised to transcripts per million (TPM) to allow comparison (with counts n and transcript length ℓ of gene A and all genes i per sample):

$$TPM_A = \frac{\frac{n_A}{\ell_A}}{\sum_i \frac{n_i}{\ell_i}} \times 10^6$$

For graphical display, TPM were further normalised to a range of 0-1. The transcripts of interest were filtered from each dataset and plotted as heatmaps. Plotted were only samples that expressed *CHAT*, *SLC18A3* (also known as vAChT), and/or *SLC5A7* (also known as HACU).

3.2.2 microRNA and Transcription Factor Targeting Predictions

Making use of the information aggregated in *miRNeo*, the genes identified as being expressed in cholinergic neurons were subjected to permutation targeting analyses of miRNAs and TFs. Genes were assumed to be expressed in cholinergic neurons if they were expressed in more than one individual sample in all single-cell RNA-seq datasets (Fig. 3.2 A-D). The TFs identified as active towards cholinergic genes in cholinergic neurons were additionally subjected to another round of miRNA targeting permutation analysis. Targeting of genes with random selections of miRNAs and TFs were permuted 100 000 times to estimate FDR. Statistical significance of the miRNA→gene or TF→gene interactions was assumed at FDR < 0.05.

3.2.3 Single-cell Expression of Cholinergic and Neurokine Transcripts

The identified samples provide an overview of potentially cholinergic cells in the sampled brain regions, and allow an assessment of the functional type and gene co-expression patterns in central cholinergic cells (Fig. 3.2). Most cells identified as cholinergic by this definition expressed the general neuronal marker *RBFOX3*, also known by its trivial name NeuN, but not the microglial marker *AIF1*. Few cells (or clusters of cells) expressed non-neuronal markers such as *GFAP* (astrocytes) or *OLIG1* (oligodendrocytes), hinting at sparse non-neuronal cholinergic functions. In agreement with our findings, cells or clusters identified as cholinergic by the authors of the respective studies [124,125] (also by personal communication with Peter Lönnerberg) were classified as interneurons and co-expressed a number of known phenotypic neuronal markers, such as *somatostatin (SST)* and *vasoactive intestinal peptide (VIP)*.

The identified cholinergic cells also revealed a constant co-expression with neurokine-related genes, particularly the transmembrane neurokine receptors LIFR and IL6ST, demonstrating a capacity to receive and process neurokine signals. In contrast, the high affinity receptor for NGF, *NTRK1*, is not co-expressed in mature (NeuN-positive) cholinergic neurons in the analysed regions, fundamentally distinguishing these cells from the basal forebrain cholinergic projection neurons.

3.2.4 Nested Regulatory Networks of miRNAs and Transcription Factors in Single Cholinergic Cells

Permutation targeting analyses revealed a nested regulatory interaction between 72 primate-specific miRNAs, 216 conserved miRNAs, and 18 TFs towards cholinergic genes expressed in cholinergic neurons (Fig. 3.2 E). TFs targeting cholinergic genes were in turn targeted by 49 conserved and 20 primate-specific miRNAs that also targeted cholinergic genes directly.

3.2.5 Transcript Clustering Based On Expression

Hierarchic clustering was applied to expression data to identify functional grouping of transcripts and cells based on co-expression. Initially, samples (i.e., single cells, pre-aggregated clusters of cells, or brain regions) are compared using a similarity- or distance-matrix (where similarity = 1 - distance). The similarity measure is based on

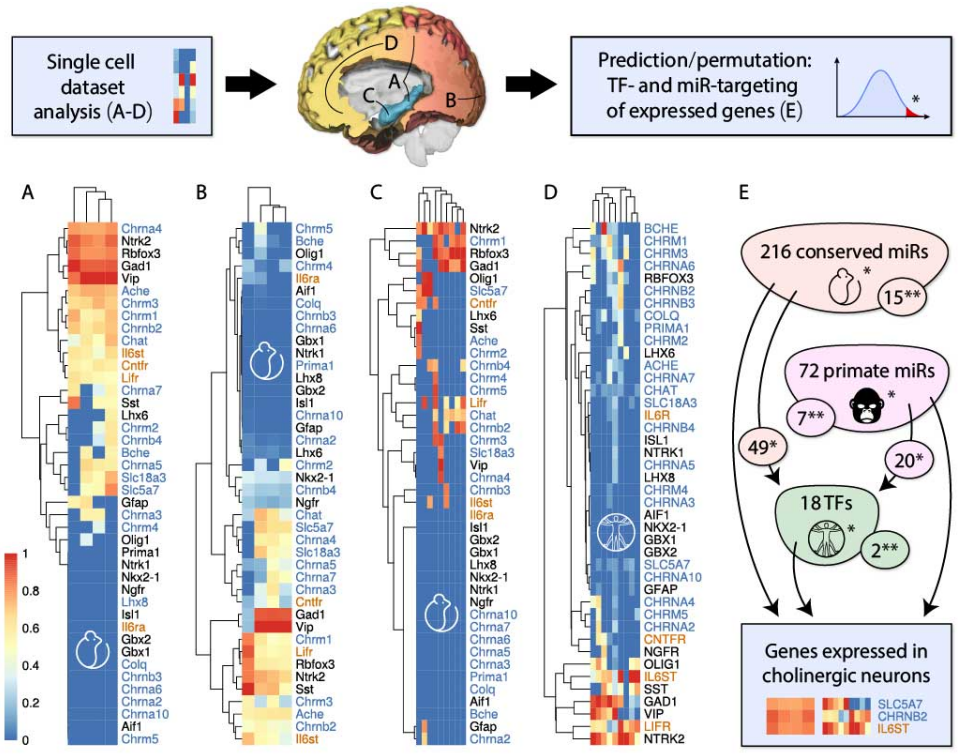

Figure 3.2: Single-Cell Sequencing of CNS Tissues. Expression patterns of cholinergic and cholinergic-related genes were analysed using web-available single-cell sequencing datasets. Expression was normalised to reflect a span between 0 and 1. **A)** Clustered single-cell sequences from transgenic mouse somatosensory cortex and hippocampus. [124] **B)** Clustered single-cell sequences from transgenic mouse visual cortex. [125] **C)** Single-nucleus sequencing of adult mouse hippocampus. [126] **D)** Single-cell sequencing of the human developing neocortex. [123] **E)** Nested regulatory circuits comprised of evolutionarily conserved and primate-specific miRNAs and transcription factors active towards genes expressed in cholinergic neurons were identified by permutation targeting analysis via *miRNeo*. *: $p < 0.05$, **: $p < 0.001$.

a computation according to the method used. For instance, Euclidean distance between two gene expression vectors (i.e., samples) of length n is the distance between points p and q in n-dimensional space, defined by:

$$d_E(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

Applying this measure to all pairwise combinations of samples results in a dissimilarity matrix that can be converted to a hierarchy using one of several clustering algorithms. Generally, samples are grouped by their similarity. Initially, each sample is assigned to its own cluster, and then, cluster number is iteratively reduced by joining the closest clusters. This results in a hierarchic tree of samples, that can be »cut« at any height to yield an arbitrary number of clusters. In biological analyses, the method after Ward (in R, »Ward.D2«) is often used. [128]

Due to the structure of the data (small number of entities compared to whole genome analysis, repetition of zeroes in individual samples), the Bray-Curtis dissimilarity [129] is superior to Euclidean distance. Bray-Curtis dissimilarity is defined as:

$$d_{BC}(p,q) = \frac{2C_{p,q}}{S_p + S_q}$$

Where C is the sum of the lesser expression values common to both vectors p and q, and S is the total number of transcripts expressed in each sample (i.e., values greater than zero in each vector). Based on this measure, the samples were clustered according to their cholinergic gene expression levels using Ward's method to yield

five separate clusters. Intermediary clustering results (not shown) revealed a uniform distribution of ATP citrate lyase (ACLY), yielding no additional information; thus, it was removed. Also removed for the purpose of clustering were the non-neuronal nicotinic receptor subunits $\alpha 1$, $\beta 1$, γ, δ, and ϵ.

3.2.6 Co-Expression of Functional Groups of Cholinergic Transcripts

Hierarchic clustering of cholinergic transcripts in each of the datasets revealed a grouping of cholinergic transcripts according to their biological function. Table 3.1 shows considerable uniformity in two single-cell mouse datasets, which diverge substantially from the brain-region- and TF-based human set. Generally, clustering shows separation of at least 3 groups of cells, one of which is the classic *cholinergic* neuron with genes for synthesis and transport of acetylcholine. Due to the frequent co-expression of *CHAT* and *SLC18A3* in neurons, it is safe to assume the *SLC18A3* as a viable substitute for *CHAT* expression and clustering in the FANTOM5 data of Marbach *et al.* [103] (for more details, see Section 2.2.3). In the single-cell datasets, the *CHAT* gene is expressed in parallel with the two cholinergic transporters, without exemption. The other groups could be described as *receptive* neuron (not cholinergic as the aforementioned, but different types of cholinergic receptors and esterase) and other, rather specialised groups, probably comprising various glial cells. These last, specialised groups are not very visible in the human dataset, which lacks the single cell resolution of the mouse datasets and therefore includes glial cells in every sample of any region. Therefore, differences in cholinergic gene expression patterns derived from Marbach *et al.* are likely the result of the numbers and dominant types of cholinergic neurons in the respective regions.

Functional stratification of cholinergic genes is also visible in a dendrogram of gene clusters from all four analysed single-cell sequencing datasets (Fig. 3.3). While there is variability in the composition of receptor subunits (which is to be expected regarding the different sampled brain regions), the core cholinergic genes (such as *CHAT*, *SLC18A3*, *SLC5A7*, and *ACHE*) associate similarly in all datasets. Notably, the distinction between a *cholinergic* and a *cholinoceptive* neuron is always visible by a grouping of, on one hand the synthesis, vesicular packaging, and reuptake of ACh, and on the other hand, cholinergic receptors and signal termination by AChE.

cluster	Zeisel et al	Tasic et al	Marbach et al
I	Ache, Chrm1, Chrm2, Chrm3, Chrm4, Chrna4, Chrna5, Chrna7, Chrnb2	Ache, Chrm1, Chrm2, Chrm3, Chrm4, Chrna2, Chrna4, Chrna5, Chrna7, Chrnb2, Chrnb4	CHRM1, CHRM2, CHRM5, CHRNA2, CHRNA4, CHRNA6, CHRNB2, CHRNB3
Ib			ACHE, BCHE, CHRNA3, CHRNB4, PRIMA1
Ic			CHRM3
Id			CHRNA5, CHRNA9
II	Chat, Chrnb4, Slc18a3, Slc5a7	Chat, Chrm5, Chrna3, Slc18a3, Slc5a7	SLC18A3, SLC5A7
III	Chrm5, Chrna10, Chrna3	Chrna10	
IV	Bche, Prima1	Bche, Prima1	
V	Chrna2, Chrna6, Chrnb3	Chrna6, Chrnb3	

Table 3.1: Cholinergic Transcript Clusters According to Cell Type vs. Brain Region. The two transgenic mouse datasets from Zeisel *et al.* [124] and Tasic *et al.* [125] show high similarity in transcript distribution. With high likeliness, cluster I is a group of postsynaptically cholinergic, »receptive« cells. Cluster II represents the classic *cholinergic* neuron, with synthesis, vesicular packaging and ACh-reuptake genes. The transcription factor-based dataset of Marbach *et al.* [103] depends on whole brain regions instead of single cells to determine similarity, and thus yields distinctly different classification. However, it also distinguishes between cholinergic synthesis (with *SLC18A3* as a substitute for *CHAT* expression) and cholinoceptive functions.

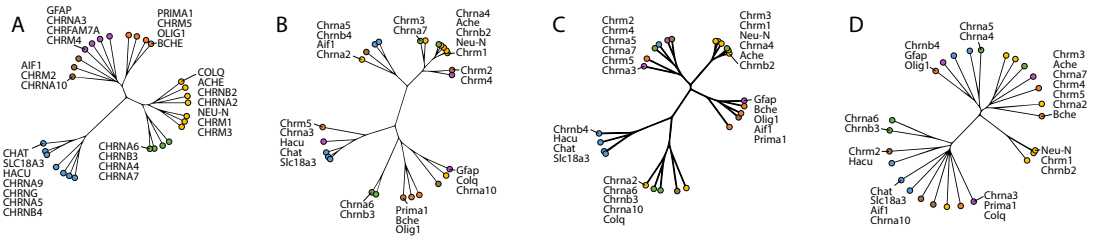

Figure 3.3: Clusters of Cholinergic Genes in Single-Cell Sequencing. Cholinergic genes were clustered using Bray-Curtis dissimilarity in four public data sets of single-cell sequencing. The displayed dendrograms visualise the distance between the genes across all samples. Gene clusters were coloured by grouping in Darmanis *et al.* [123] (A). Notably, genes clustered according to their biological function, for instance, CHAT, vAChT and HACU always are closely associated (blue), as are the genes comprising the putative »cholinoceptive« neuron (yellow). **A)** Single-cell sequencing of the human developing neocortex. [123] **B)** Clustered single-cell sequences from transgenic mouse visual cortex. [125] **C)** Clustered single-cell sequences from transgenic mouse somatosensory cortex and hippocampus. [124] **D)** Single-nucleus sequencing of adult mouse hippocampus. [126] Marker genes are: NeuN - neurons; OLIG1 - oligodendrocytes; AIF1 - microglia; GFAP - astrocytes.

3.3. The Cellular Model

We selected two mono-cultures of human neuronal cells for subsequent experiments: LA-N-2 and LA-N-5. During the selection process, multiple options were considered. Multicellular models would, in principle, allow disentanglement of the functions of distinct cell types, for instance glia and neurons. This could be achieved by *in vivo* or *ex vivo* approaches in rodents. However, our diseases of interest (Section 1.2) display a noticeable lack of transferability from lower mammals to human. Alternatively, co-cultures of human cells in mono-layer or as 3D-culture have been proposed, but these still lack experimental stability.

3.3.1 The SH-SY5Y Neuroblastoma Cell Line

A prominent example of human neuronal cell culture used in the identification and elucidation of cholinergic processes is the immortalised neuroblastoma cell line SH-SY5Y.[130] Derived from its parent line SK-N-SH, an adrenergic neuroblastoma,[131] it expresses ample amounts of *ACHE*, and thus had become a work horse in many cholinergic fields, such as Alzheimer's Disease (which is treated with AChE inhibitors), pesticide development, and warfare. However, in spite of its usefulness for processes involving *ACHE*, it turned out a less than optimal choice for the study of molecular events surrounding *CHAT* and *SLC18A3*, as it barely expresses both genes, and cannot be coerced to elevate *CHAT* expression by the usual differentiation techniques (own experimentation, data not shown). Thus, for the questions asked in this chapter of the dissertation, SH-SY5Y does not qualify as adequate representation of a »cholinergic neuron«.

3.3.2 The LA-N Neuroblastoma Cell Lines

Following the elimination of SH-SY5Y as a suitable subject, a literature search for candidates representing a cholinergic neuronal transcriptome revealed, among others, representatives of the LA-N neuroblastoma cell lines developed by R.C. Seeger around 1980.[132,133] Neuroblastoma is a form of neuronal cancer often affecting small children, and, consequently, the two cell lines used in my experiments are immortalised biopsies of a 3 year old girl (LA-N-2[132]) and of a 4 month old boy (LA-N-5[133]). The decision to use LA-N-2 as initial cellular model was influenced by three factors: it is well described in literature, although most studies had been published in the 1980s and 90s; it expresses substantial amounts of *CHAT* and *SLC18A3*; and it responds to neurokine-mediated differentiation by assuming a neuronal morphology accompanied by further elevation of *CHAT* and *SLC18A3* expression. LA-N-5 was not nearly as well described as LA-N-2, but later added to the experimental roster because of the complementary sex and hints towards cholinergic differentiation under retinoic acid.[134]

3.3.3 Culture

LA-N-2 and LA-N-5 are very similar in their culture requirements. They have comparatively high duplication times, which can be lowered by using certain conditions that affect medium composition, nutrition, and CO_2 content. The cells were acquired at DSMZ (Braunschweig, Germany), which recommends keeping them in a 50:50 mixture of Dulbecco's modified eagle medium (DMEM) and Roswell Park Memorial Institute medium (RPMI1640), with 20% fetal calf serum (FCS) added. Sometimes, recommendations also suggest Leibovitz's L-15 medium, which is specifically designed for low CO_2 conditions, and others have suggested increased CO_2 levels inside the incubator. A combination of the DSMZ-recommended medium with 8% CO_2 atmosphere inside a 37°C incubator accelerated growth to a degree that the cells could be split 1:3 to 1:4 in a weekly cycle. This protocol was used for all further experiments, which were performed between splits 2 to 8 after thawing of a batch from -80°C. All handling during maintenance and experimentation was performed under a laminar flow hood.

3.3.4 Differentiation

Neuronal differentiation of neuroblastoma cell lines has been performed in many instances, utilising a wide variety of differentiation agents such as the very general retinoic acid or 5-bromo-uracil, or very specific reagents, such as the neurokines IL-6 and CNTF. LA-N cells have also been described to react to a selection of these substances; however, due to our elevated interest in neurokine mechanisms, we opted for a neurokine-based differentiation protocol. In personal communication, James McManaman revealed that the »*CHAT* development factor« that he had discovered [66] was, in fact, CNTF, which had never been published. Additionally, of the neurokines used for differentiation purposes, CNTF is best described in literature and easily acquired in dried form from Merck (formerly SigmaAldrich, Darmstadt, Germany). CNTF was resuspended in pure water to a concentration of 25 µg ml^{-1} and stored for experimentation in aliquots at -20°C.

LA-N cells are very sensitive to repeated temperature changes (or other handling-related disturbances), which resulted in increased amounts of apoptotic cells following repeated removal from the incubator after seeding or medium changes during the experiment (Lobentanzer, not published). For this reason, the differentiation reagent was only added once, 24h after initial seeding of the cells, and further disturbances avoided until the time of lysis. For the maximum duration of the experiments, 120h from seeding until lysis, the initially supplied medium was sufficient for survival.

Differentiation was performed in regular growth medium without changes in FCS content, and CNTF was added to the medium after an initial growth period of 24h. Cells were seeded into 12-well plates at approximately 200 000 cells/well, with 1 ml of growth medium. To determine the optimal amount of CNTF for differentiation, time-dose experiments were perfomed for both cell lines in a range from 1 ng ml^{-1} to 100 ng ml^{-1} for several time points during four days. Here, we discovered the first pharmacological difference between LA-N-2 and LA-N-5: the maximum of their cholinergic response to neurokine stimulation (i.e., an elevation in *CHAT* and *SLC18A3* transcription) occurs at different concentrations of CNTF. While LA-N-2 cells respond most strongly to 100 ng ml^{-1}, LA-N-5 cells show an »inverted u«-type dose response with a maximum around 10 ng ml^{-1} CNTF (Fig. 3.4 A). James McManaman, who studied LA-N differentiation thoroughly in the 1990s, [135] believes both lines to respond in an »inverted u«-type manner (personal communication); thus, in all likelihood the LA-N-2 response would also diminish at CNTF concentrations significantly higher than 100 ng ml^{-1}. Also, CNTF concentrations could likely be significantly lowered by removal of the high amount of FCS in the medium, however, that would require the use of a special serum-free medium, which would have to be established up front, and

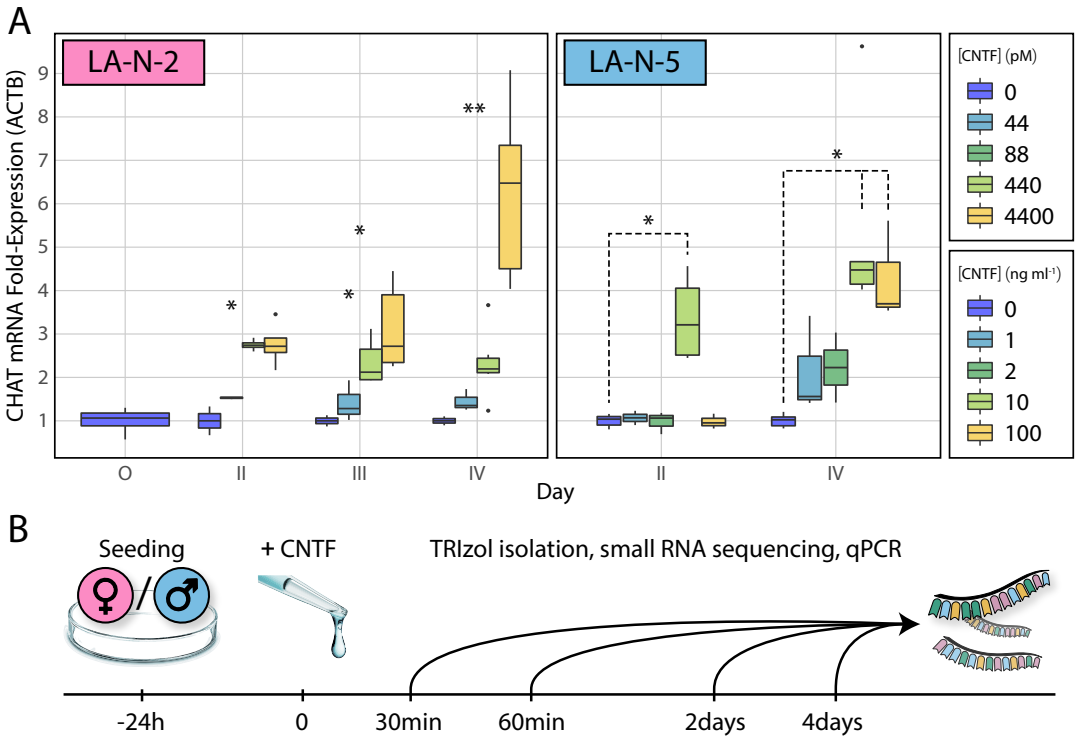

Figure 3.4: Time-dose curve of CNTF-mediated differentiation of LA-N-2 and LA-N-5. A) Cells were stimulated with varying doses of CNTF, and lysed at various time points to determine CHAT mRNA levels via qPCR. Expression ($\Delta\Delta C_t$) was normalised to housekeeping genes (ACTB, GAPDH, RPLP0) and to control sample without CNTF to determine fold-changes. LA-N-5 reacts strongest to a concentration of 10 ng ml^{-1}, while LA-N-2 reacts strongest to 100 ng ml^{-1}. *: $p < 0.05$, **: $p < 0.001$ **B)** Cells were seeded at ~2E05 cells/well in a 12-well-plate. After 24h, CNTF was added to the existing medium as quickly as possible to avoid disturbance. Cells were lysed *in situ* at time points 30 minutes, 60 minutes, 48 hours, and 96 hours using TRIzol for downstream RNA processing.

may have other, unforeseen consequences. Regardless, CNTF concentrations around 100 ng ml^{-1} (i.e., pico- to nano-molar) still are well within the physiological range of concentrations that the mammalian brain is able to reach by paracrine secretion via, e.g., astrocytes.[136]

vacht qpcr figure?

To study the small RNA dynamics following CNTF exposure of LA-N-2 and LA-N-5, the experiment was stopped at 4 time points and the cells were quickly lysed *in situ* to preserve total RNA in that state: for the quick, immediate-early-like phase, at 30 and 60 minutes after the addition of CNTF, and, for the long-term effects of differentiation, at 48 and 96 hours after the addition of CNTF (Fig. 3.4 B, from Lobentanzer *et al.*[1]). Each time point was controlled by a pseudo-treated (using pure water) culture from the same batch that had been seeded at the same time as the experimental group. In the final series used for the parallel sequencing of LA-N-2 and LA-N-5, all experiments were carried out in quadruplicates.

3.3.5 RNA Isolation

Total RNA was isolated using TRIzol (ThermoFisher Scientific), essentially as suggested by the manufacturer, with slight changes to the protocol to enrich small RNA species. The cells, growing in a monolayer in 12-well-plates, were cleared of medium, washed two times with 500 µl of cell culture grade phosphate buffered saline (PBS) (Gibco), and immediately suspended in 1 ml of TRIzol, pipetting up and down until visibly dissolved. After incubation for 5 minutes at room temperature, the samples were stored in -20°C for short periods of time until

RNA isolation.

TRIzol-suspended lysates (1 ml) were added to RNA-separation centrifuge tubes (PhaseMaker Tubes, ThermoFisher Scientific), adding 200 µl of pure chloroform and mixing vigorously for 15 seconds. After two minutes, the mixture was centrifuged at 12 000 g and 4°C for 15 minutes, and the upper, watery phase containing the RNA was extracted. This was mixed with approximately 2 parts of pure ethanol and incubated for 10 minutes at room temperature to precipitate the RNA. The precipitate was spun at 12 000 g and 4°C for another 10 minutes, and the supernatant discarded. The pellet was washed with 85% ethanol (vortexed briefly) and centrifuged again for 5 minutes at 7500 g and 4°C.

After the final centrifugation step, the samples were transferred to the laminar flow hood, and air dried after removal of most of the supernatant via micropipettors. The pellet was allowed to dry almost until completion and resuspended in 30 µl to 50 µl pure RNase-free water. RNA concentration was measured at a Nanodrop 2000 instrument (ThermoFisher Scientific) and samples were diluted to a uniform concentration of 100 ng μl^{-1}. Finally, RNA samples were aliquoted according to later purpose and stored at -80°C.

RNA quality was determined by analysis on a 2100 Bioanalyzer instrument (Agilent) using a nano chip and 1 µl of sample; RNA integrity number (RIN) was near optimal for all samples (>9).

3.4. Small RNA Sequencing and Differential Expression Analysis

For the detection and analysis of small RNA species, RNA-seq is the current gold standard method. It allows the mapping of a comprehensive transcriptome and thus is vastly superior to small scale and consecutive methods such as real-time quantitative polymerase chain reaction (RT-qPCR), and even the larger scale microarrays. Microarrays, while also potentially allowing a »snapshot« of entire transcriptomes, are limited by the predetermined sequences on the chip. RNA-seq, on the other hand, is not biased towards any structural property of the sample; this is particularly important in the analysis of small RNA species, since their sequences are very variable (tRFs) and still not completely catalogued (miRNAs). Assuming an adequate sequencing depth (at least about one million reads/sample), RNA-seq allows a comparison of all expressed small RNA species at once, which is immensely helpful when dealing with processes on the combinatorial scale of miRNA regulation.

3.4.1 Sequencing

For small RNA sequencing, the aliquoted samples were shipped on dry ice to the cooperating institute at the Hebrew University of Jerusalem, the Silberman Institute of Molecular Biology, the laboratory of Prof. Hermona Soreq. 600 ng of total RNA per sample were prepared for sequencing using the NEBNext Small RNA Library Prep Set for Illumina (New England BioLabs). The libraries were multiplexed with coloured barcodes, allowing for sequencing of all 48 samples on one chip. Briefly, this includes ligation of sequencing adapters to both 3' and 5' ends of all (single-stranded) RNA fragments in the sample, followed by 12-15 cycles of reverse transcription to form the RNA library. Ligated and amplified libraries were then size selected via gel electrophoresis on a 6% polyacrylamide gel. The band representing small RNA species on the gel was excised and prepared for loading onto the sequencing chip. After loading, the chip was sequenced in a NextSeq 550 series instrument (Illumina) with a read length of 80 nucleotides (nt), single-end.

The quantity of reads per sample was determined by analysis of the raw fastq files. The read count across all samples before filtering was 7.8E06 $\pm_{SD}$ 2.5E06, read count after quality filter and adapter removal was 6.8E06 $\pm_{SD}$ 2.2E06 (n = 48); a mean of 87% of reads remained after filtering, exceeding the recommended minimum amount ($\sim$1 million) by 4- to 12-fold (Fig. 3.5 A). Overall, $\sim$326 million reads remained to be passed down to subsequent analyses. Sequencing quality was determined by analysis of the raw reads using the FastQC software.[137] Even before adapter removal and quality filtering, FastQC detected no »reads of poor quality« in any sample. Fig. 3.5 B gives a representative example of read quality per base (Sample 1).

Raw reads were adapter-trimmed and quality filtered using the flexbar software[138] with parameters

```
-a adapters.fa -q TAIL -qf sanger -qw 4
-min-read-length 16 -n 1 --zip-output GZ
```

The sequence used in the *adapters.fa* file, as recommended by the manufacturer, was

AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

Paired-end sequencing still is superfluous in small RNA-seq, because none of the common alignment pipelines can use the second (reverse) read, and manual paired alignment does not yield nearly as much benefit as the

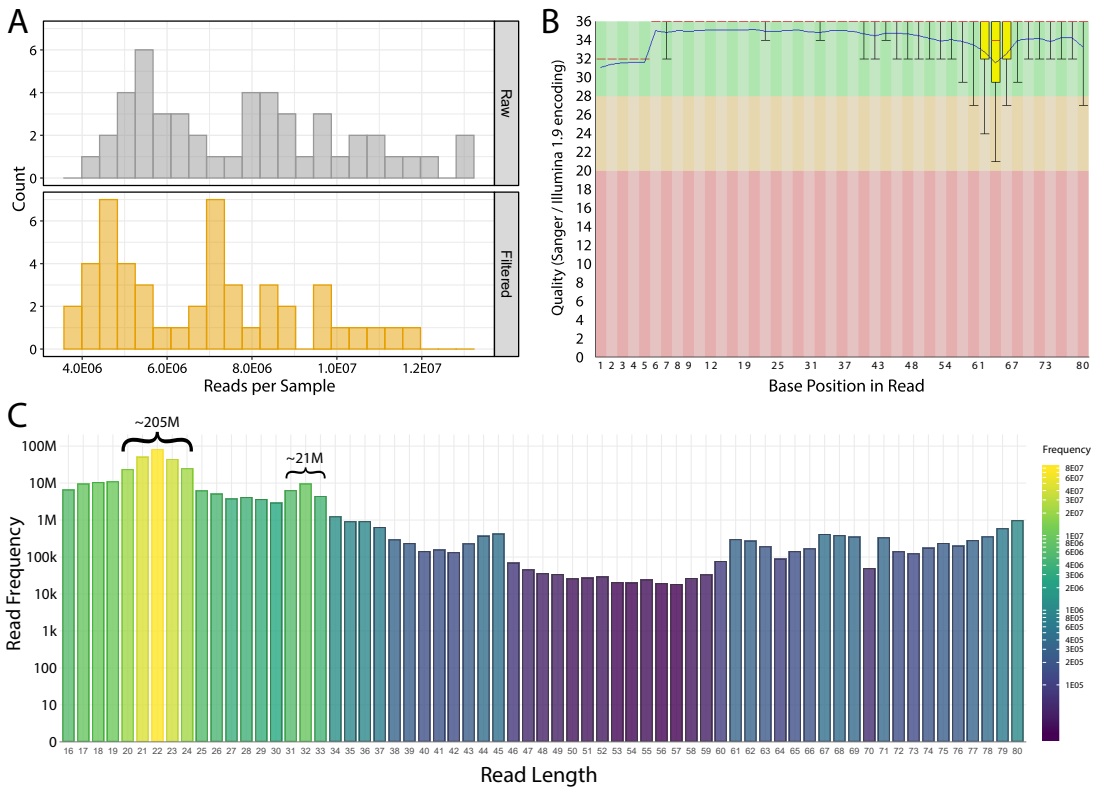

Figure 3.5: Small RNA Sequencing - Read Count, Quality, and Length. All samples provided near optimal quality. **A)** Per sample read count had a mean of 7.8E06 $\pm_{SD}$ 2.5E06 in raw samples (top) and 6.8E06 $\pm_{SD}$ 2.2E06 after quality filtering and adapter removal (bottom). 87% of reads were retained after filtering, with samples spanning read count values between 4 and 12 million. **B)** Representative example of quality score per base position in the sequencing (FastQC output of sample 1). Quality scores are always near the optimum, with a characteristic slight dip around nt 65. This occurs in all samples and is likely a technical result of the sequencing process. Possibly, it reflects the most common adapter ligation position after size selection of the RNA pool. **C)** Read length was determined for every one of the ~326 million reads. Nearly 80 million reads have a length of 22 nt, and the peak from 21 to 24 nt comprises ~205 million reads. This represents the bulk of miRNAs, and probably a significant amount of tRFs. The second peak, from 31 to 33 nt, still comprises ~21 million reads; these in all likelihood represent the longer tiRNAs. The reads above a length of 33 nt only sum up to an amount of ~6 million, and may contain RNA of viral origin, or even mature tRNAs.

depth increase in single-end sequencing (the read count per sample effectively doubles). 80 nt is the maximum read length possible in our small RNA workflow, and is excessive for the analysis of miRNAs. For transfer RNA fragments, however, a longer read can yield a more complete picture of expression, since the longer tiRNAs can easily reach 40 nt in length. Indeed, the read length distribution after adapter removal shows a significant amount of small RNA species exceeding the length possible for miRNAs (Fig. 3.5 C).

align longer reads to genome? viral?

discuss in text?

3.4.2 Sequence Alignment

For the alignment of miRNA sequences, parts of the miRExpress 2.0 [139] pipeline were used according to the documentation. First, a lookup table for the current miRBase version 21 was created as per the instructions of the authors. The alignment was then performed using the commands *Raw_data_parse*, *statistics_reads*, *alignmentSIMD*, and *analysis*; *Trim_adaptor* was skipped because the adapters had already been trimmed in the quality filtering step. Additionally, since miRExpress is not accepting of sequences of any length, the raw data was length filtered to include only reads up to a length of 25 nt before input into miRExpress. Thus, raw reads were aligned to the miRnome provided by miRBase v21, yielding count tables of mature miRNAs and miRNA precursors for each

sample. In total, 1913 mature miRNAs from miRBase v21 were discovered in the data.

3.4.3 Differential Expression Analysis - R/DESeq2

To determine the effect and dynamics of CNTF-mediated differentiation of LA-N-2 and LA-N-5, the expression state of each measured time point was compared to the respective control using the established R package *DESeq2*.[140] *DESeq2* determines differential expression (for gene i and sample j) in count-based data by application of a linear regression model to a negative binomial distribution based on a fitted mean μ_{ij} and a gene-specific dispersion value α_i. The mean is derived using a sample-specific »size factor«, s_j, and a parameter q_{ij} proportional to the expected true concentration of RNA fragments in the sample. The *DESeq2* differential expression pipeline is composed of the following commands:

- `estimateSizeFactors()` to estimate s_j
- `estimateDispersion()` to estimate α_i
- `nbinomWaldTest()` application of a generalised linear model to determine log-fold changes and statistics via the Wald test, using $\mu_{ij} = s_j q_{ij}$ and $log_2(q_{ij}) = x_j \beta_i$.

The Wald test, named after Abraham Wald,[141] is an approach to hypothesis testing that measures the distance between the tested unrestricted estimate and the null hypothesis, using the precision as a weighting factor. The larger the distance between tested values and the null, the more likely the measured values are »true«. RNA-seq data can be modelled using binomial distributions,[142] such as the Poisson distribution, and the difference between two Poisson means (e.g., »treated« vs »control«) can be tested by generalised linear models based on the distributions directly (Poisson regression), Fisher's exact test, or the likelihood ratio test. However, comparative analysis has shown that the Wald test on log-transformed data provides statistical power superior to these other methods,[143] particularly in lowly expressed fragments. The design formula for the linear regression was applied to LA-N-2 and LA-N-5 separately as a simple factor combination of condition and time point:

$$y \sim condition\_time$$

The heteroskedastic nature of RNA-seq count data (variance is much higher in low-count features) brings statistical problems. To reduce the noise introduced by the high variance in low-count genes while preserving large, »real« differences, the authors propose the »shrinkage« of log-fold changes to avoid arbitrary low-cut filtering at a predefined expression (count) value. Multiple variants are available; for miRNA data, the adaptive algorithm »apeglm«[144] (adaptive t prior shrinkage estimator) yielded sensible results (see Fig. 3.6).

3.4.4 microRNA Dynamics in CNTF-mediated Cholinergic Differentiation of LA-N-2 and LA-N-5

Differential expression analysis performed in this manner yielded 490 differentially expressed (DE) miRNAs across all groups, with characteristic distributions between cell lines and time points. The raw data and processed counts were deposited to NCBI Gene Expression Omnibus (GEO), accession GSE132951. An earlier sequencing experiment (deposited as GSE120520), which was similar in principle, but only comprised three biological replicates and only LA-N-2, reproduced 80% of DE miRNAs in the newer LA-N-2 samples. Considering the general reproducibility of RNA-seq and

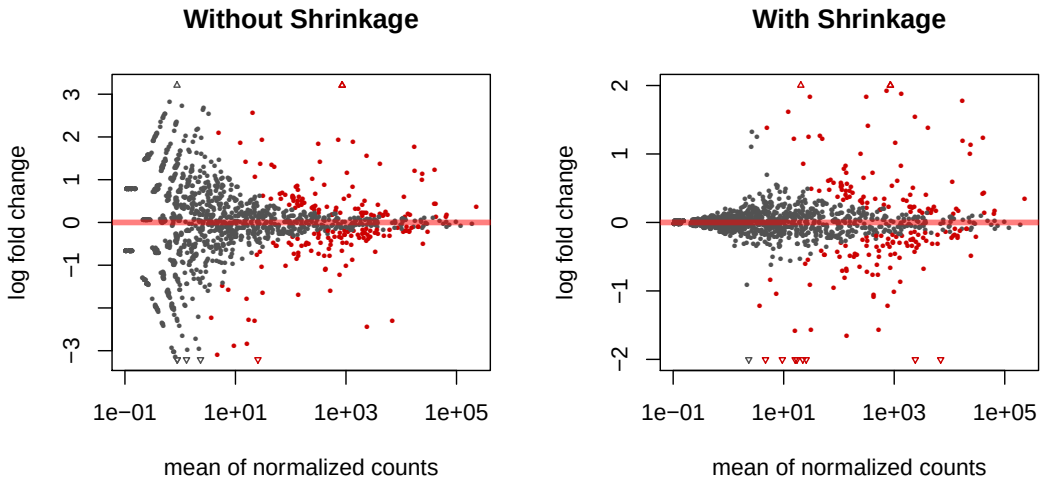

Figure 3.6: MD Plot Shrinkage Comparison. A mean-difference plot (MD Plot) is a plot of log-intensity ratios (differences, »M-values«) versus log-intensity averages (means, »A-values«); it is synonymous with »MA Plot«. The *DESeq2* function *plotMD* shows the log fold changes attributable to a given variable over the mean of normalised counts for all the samples in the data set. Points will be coloured red if the adjusted p value is less than 0.1. Points which fall out of the window are plotted as open triangles pointing either up or down. The left plot is generated from the standard linear model, the plot on the right is corrected by the »apeglm« algorithm [144] to reduce noise in the low-count fragments (data from LA-N-2 CNTF vs control on day 4).

the lower replicate number, 80% is an excellent substantiation of the result. About 25% of miRNAs predicted in single-cell permutation targeting analysis (see Fig. 3.2 E) were found DE in LA-N-2 and LA-N-5 (Fig. 3.7 A) in all three groups, i.e., conserved, primate-specific, and TF-targeting miRNAs.

Differential Expression in Both Cell Lines

114 mature miRNAs were detected as DE in both cell lines, with some changes similar in both, while others were inverted (Fig. 3.7 B). In both cases, however, count-change values (see Box 2) correlated highly between the two cell lines (similar: 76 miRNAs, Spearman's $\rho = 0.9066$, p < 2.2E-16; inverted: 38 miRNAs, $\rho = 0.9294$, p < 2.2E-16).

Differential Expression Along The Timeline

For consistency, from hereon out, time points 30 minutes and 60 minutes will be termed »early«, while 2 days and 4 days will be referred to as »late«. Differential expression was detected in all groups, lending credibility to the rapid changes in expression needed for a miRNA response of the »immediate-early« type. However, the response to long-term CNTF stimulation was larger in miRNA numbers as well as effect sizes (Fig. 3.8 A&B). Of all early perturbed miRNAs, only 3 and 13 miRNAs were exclusively perturbed immediate-early-like in LA-N-2 and LA-N-5, respectively; all others were still DE after 2 and/or 4 days. In LA-N-2, the late time points at 2 and, particularly, 4 days showed the greatest perturbation; in LA-N-5, the picture was more complex (Fig. 3.8 C&D). However, generally, there were large similarities as well as exclusivities between the time points 2 and 4 days in both cell

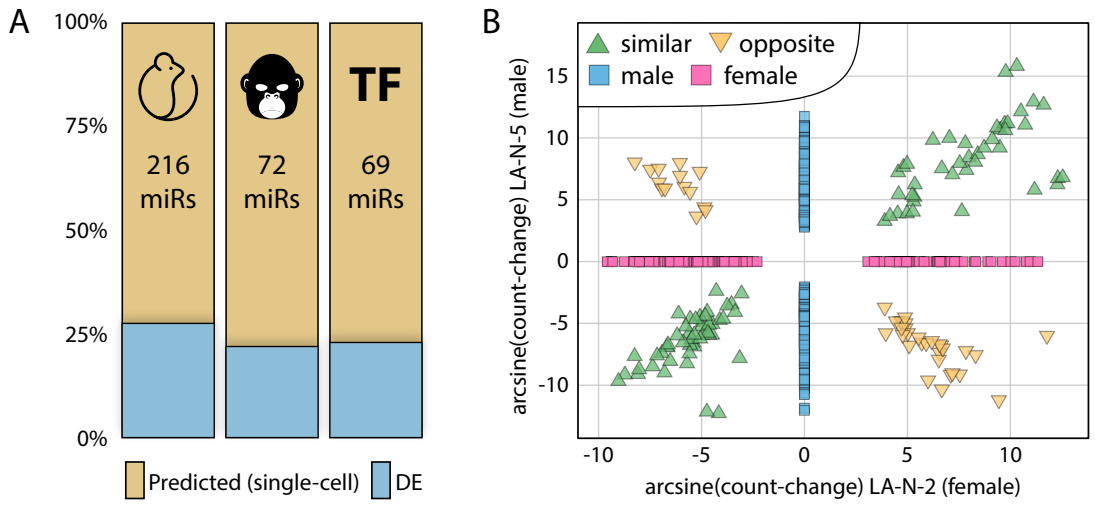

Figure 3.7: Differentially Expressed microRNAs in LA-N-2 and LA-N-5. A) Differential expression in sequencing of LA-N-2 and LA-N-5 identified approximately 25% of miRNAs predicted by permutation analysis of single cells expressing cholinergic markers (compare Figure 3.2 E) in all miRNA subgroups (evolutionarily conserved, primate-specific, and TF-targeting). **B)** Differentially expressed miRNAs in LA-N-2 and LA-N-5 can be stratified into four groups by their expression changes: miRNAs only differentially expressed in LA-N-2 (pink) or LA-N-5 (blue), miRNAs changed in similar direction in both cell lines (green) and miRNAs changed inversely between cell lines (yellow). Notably, the changes in expression as measured by count-change in similarly as well as inversely changed miRNAs both correlate well.

lines. When comparing early and late time points between LA-N-2 and LA-N-5 directly, similarly complex patterns emerged (Fig. 3.8 E&F). Particularly at late time points (Fig. 3.8 F), every possible combination of overlap exists. 24 miRNAs were DE in all late conditions; 107 miRNAs were DE only in LA-N-2, and 269 miRNAs were DE only in LA-N-5.

something special?

compare most important targets early/late

Differential Expression Between LA-N-2 and LA-N-5

While there was considerable intersection in DE miRNAs between the cell lines, a substantial amount of miRNAs was only DE in one of the two lines. Generally, response to CNTF was higher in the male-

Box 2: The count-change metric

The frequently used log-fold change metric is not ideally suited for assessing the potential effect of expression changes for individual miRNAs because it does not reflect mean expression levels. To determine the absolute change in expression, the count-change metric was introduced, a combination of base mean expression and log-fold change, to weigh DE miRNAs against one another. The count-change is defined as follows:

$$CC = (BM \cdot 2^{LFC}) - BM$$

CC: count-change, BM: base mean expression, LFC: log-2-fold-change.
Importantly, by using the base mean expression, count-change correlates directly with sequencing depth. Generalisation, e.g. comparison between two individual experiments, is therefore not straightforward. A normalisation to raw reads would enhance comparability, however, other effects such as fragment distribution and quality aspects may also play a significant role.

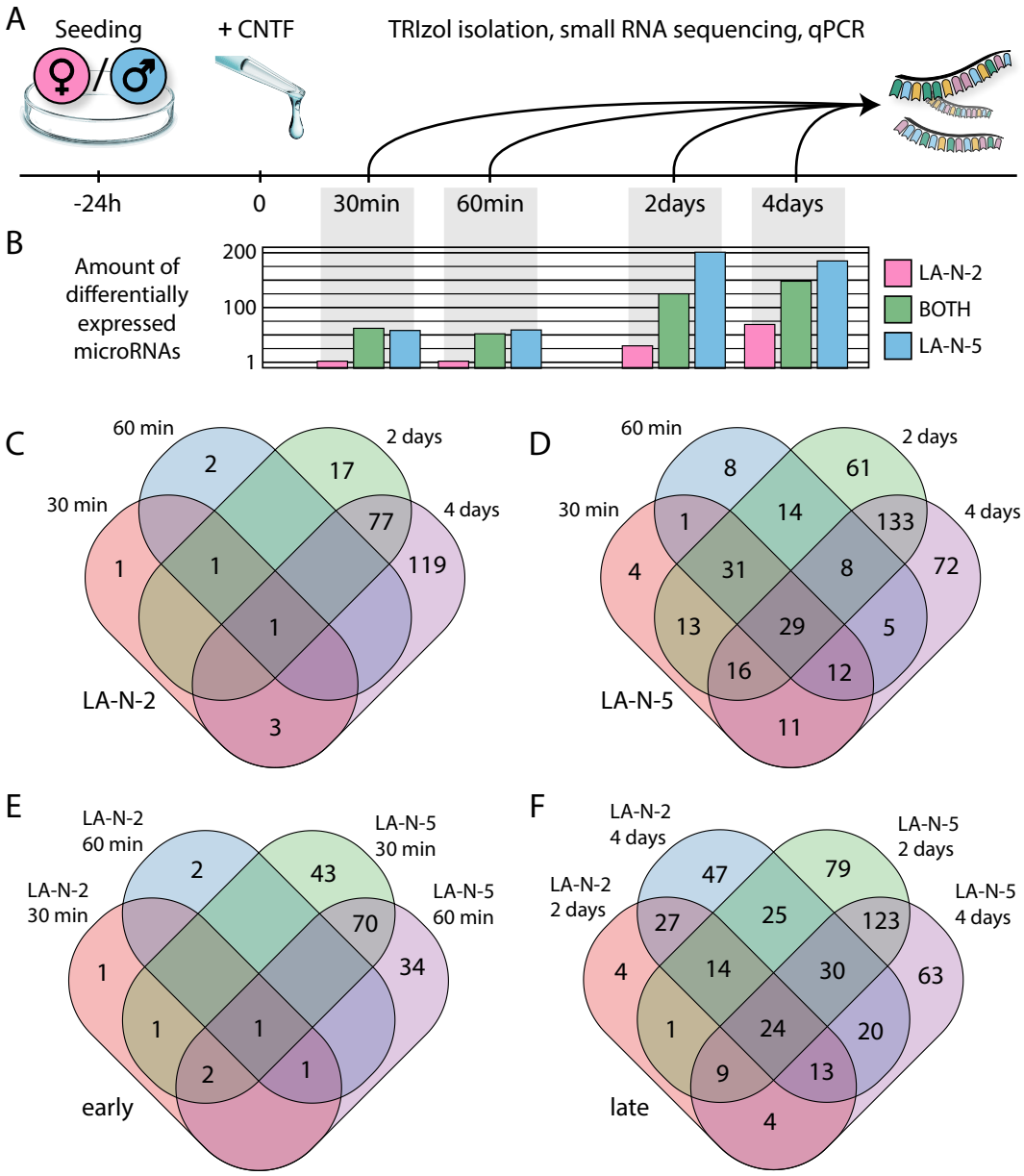

Figure 3.8: LA-N-2 / LA-N-5 Timeline and Differential Expression. A) Experimental timeline of CNTF differentiation. **B)** Bar plot of differentially expressed (DE) miRNAs per time point, divided by cell line where differential expression was measured (LA-N-2 only, LA-N-5 only, or both). **C)** Venn diagram of DE miRNAs in LA-N-2, divided by time point. Few early DE miRNAs, and continually more the longer differentiation lasts. **D)** Venn diagram of DE miRNAs in LA-N-5, divided by time point. Similar in pattern to **C**, but more pronounced in numbers. **E)** Intersection of early time points in LA-N-2 and LA-N-5. Despite the low differential expression in LA-N-2, there is overlap. **F)** Intersection of late time points in LA-N-2 and LA-N-5. Overlap is pronounced and complex, however, there are also cell line-exclusive miRNAs.

originated LA-N-5 cells; however, there were also miRNAs found DE only in the female LA-N-2 (compare Fig. 3.8). Thus, not all of the differences in miRNA expression can be attributed to a higher sensitivity in LA-N-5. Similarly, LA-N-5 shows a »non-significant trend« toward higher count-change values (mean of absolute count-change across all DE time points, 20 907 versus 3066, Welch two-sample t test, p = 0.08).

The influence of genotype on the differentiating effect of CNTF was determined via a statistical interaction design in the *DESeq2* Wald test. Briefly, by including an interaction term in the linear regression formula, the effect of the condition (CNTF or control at each time point) between the two genotypes can be isolated:

$$y \sim condition + genotype + condition : genotype$$

Using the interaction term *condition* : *genotype*, miRNAs that reacted significantly different to CNTF stimulation in LA-N-5 compared to LA-N-2 were determined. Of note, the sexual dimorphism becomes more pronounced over the course of differentiation. While there is no significant difference between LA-N-2 and LA-N-5 at 30 minutes and only one miRNA DE at 60 minutes, numbers increase at 2 days and reach a maximum at 4 days, with significant overlap (Fig. 3.9 A). Although not all miRNAs found in this manner belong to the group of miRNAs with inverted expression between LA-N-2 and LA-N-5, several show significant differential regulation between the male and female cellular models (e.g., hsa-miR-615-3p, Fig. 3.9 B). To further examine the effect of genotype on the small RNA response to CNTF, the regular differential expression results (Section 3.4.4) were intersected with the interaction term for the late time points. This resulted in a complex pattern of intersecting miRNAs, in both cell lines (Fig. 3.9 C&D). Again, all possible overlaps between any two groups exist; 37 and 36 miRNAs are found in all four groups of LA-N-2 and LA-N-5, respectively. Among those, 16 mature miRNAs belong to all sets. All pertinent sets of miRNAs can be found in Appendix C.

target, GO for which?

While this descriptive analysis shows significant differences in small RNA expression in response to neurokine-mediated cholinergic differentiation of these two cell lines, the functional implications are much less clear. In the following, approaches to discerning cellular function, rather than just level changes, will be explored, via unbiased analysis as well as in a system-targeted manner.

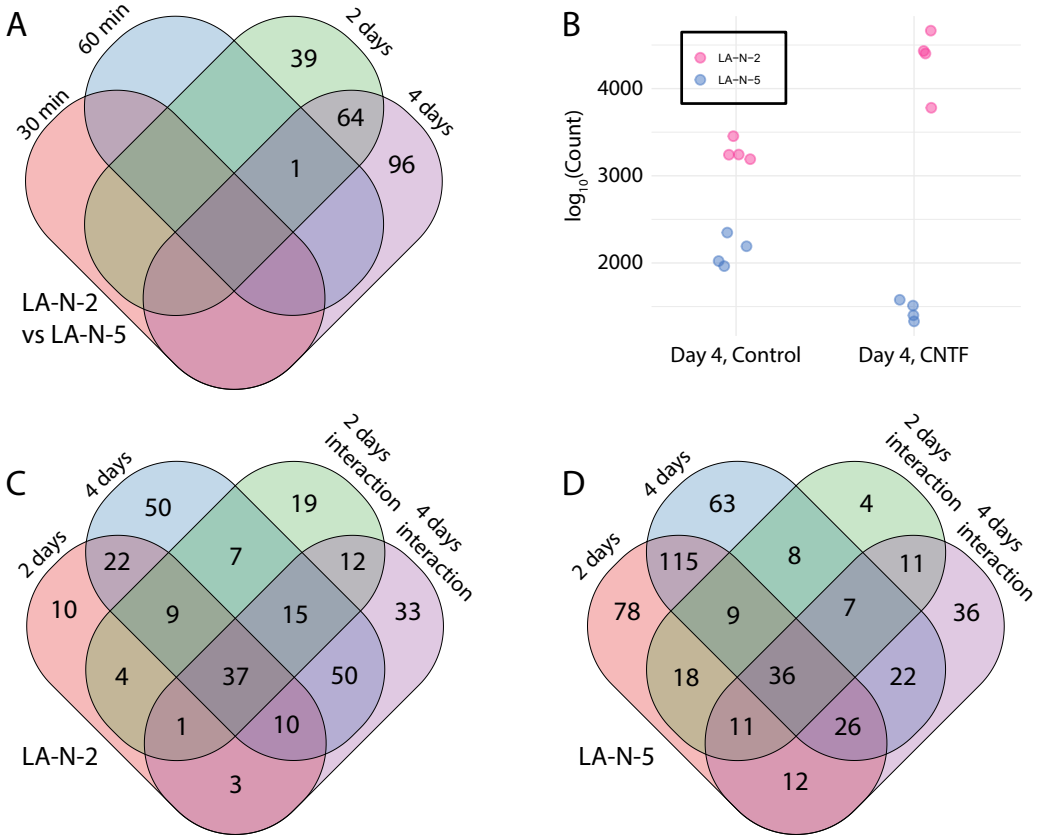

Figure 3.9: miRNAs Differentially Expressed Between LA-N-2 and LA-N-5. Application of a design model formula which includes an interaction term enables display of the influence of the male or female genotype on differential miRNA expression. **A)** Venn diagram of miRNAs differentially expressed *between* LA-N-2 and LA-N-5 at all four time points. **B)** Counts plot of normalised raw expression values of hsa-miR-615-3p. Exemplary of a high influence of genotype on the differential expression caused by CNTF differentiation, hsa-miR-615-3p is more highly expressed in the female LA-N-2 and elevated after four days of CNTF-induced differentiation, while in the male LA-N-5, it is expressed slightly lower and suppressed upon differentiation. **C)** Venn diagram comparing late differential expression in LA-N-2 with late time points of differential expression between LA-N-2 and LA-N-5. All possible combinations exist, however, there are miRNAs affected by genotype that are not differentially expressed in the simple model. **D)** Venn diagram comparing late differential expression in LA-N-5 with late time points of differential expression between LA-N-2 and LA-N-5. Essentially similar to **C**, but with partly higher quantities of DE miRNAs.

3.5. microRNA Family Gene Ontology Enrichment

A significant drawback of the recency of the discovery of regulatory small RNAs is the lack of comprehensive functional annotation. While protein coding genes are well annotated and neatly organised into an enormous amount of ontological categories (see Section 2.4.2), miRNAs have only been anecdotally associated with specific functions in the cell. Additionally, the functional roles of protein coding genes are much more limited than those of miRNAs; the number of potential functions of any miRNA correlates with the number of mRNA targets this miRNA has, and is also highly context-dependent (e.g. regarding cell type, cell state, disease). Thus, to systematically screen a large amount of miRNAs and families, we had to turn to an indirect approach: the GO analysis of targeted genes.

3.5.1 microRNA Family Enrichment

how many DE miRs in families?

To categorise and systematise the sexual dimorphism of CNTF differentiation of LA-N cells, statistically overrepresented miRNA families in the differential expression datasets were determined. Of the 151 miRNA families listed in miRBase v21, members of 71 families are DE in LA-N-2 and LA-N-5. Enrichment of male, female, and ubiquitously DE miRNAs in these families was determined by hypergeometric enrichment via Fisher's exact test for each of the families. The targets of all individual miRNAs in the enriched families were determined via *miRNeo* query.

3.5.2 Creation of miRNA Family Gene Target Sets

GO analysis of the targets of a single miRNA is challenging, because the analysis requires a weighted scoring system of input genes. For single miRNAs, the options for scoring are limited to the aggregated targeting score or permutation p-values. Using families enables the introduction of a further scoring method: the aggregation of individual family members targeting the same gene. The reasoning behind this approach is to determine a general functional »area« of biological process that the miRNA family in question operates in. To account for the possibility of multiple areas being affected by a family, the test set of genes in any GO enrichment analysis should not be too small (i.e., rather the top 100 genes than the top 10).

Following this reasoning, the targets of all miRNAs in each family were determined via *miRNeo* query. For each family, genes were ranked by their cumulative targeting score ρ from all family members. For gene i and number of miRNAs in family x, gene score ρ is calculated from individual miRNA→gene scores s:

$$\rho_i = \sum_{n=1}^{x} s_{ni}$$

3.5.3 GO Analysis of Target Sets

The gene target sets of individual miRNA families were ordered decreasingly by their cumulative score ρ and subjected to GO analysis via the R package *topGO*.[145] Briefly, *topGO* analysis extends the basic hypergeometric approach of GO enrichment analysis by de-correlating the DAG structure of GO annotation (see Section 2.4.2), allowing a weighted correction for the interdependency of neighbouring GO nodes. If a gene is found in both

the parent node (more general) and the child node (more specific), the less specific parent node gene is weighted less; in this way, the most specific node of each hierarchical branch can be found without confounding the result with less specific terms. While GO analysis always is subject to interpretation by the researcher, this weighted algorithm has been shown to reduce false positives while retaining a high true positive ratio.

topGO analysis was performed using the classic (i.e., Fisher's exact test) as well as weighted methods for comparison, however, to determine significance, the p-values calculated by the enhanced weighted algorithm were used. FDR was controlled at 5%. As recommended by the authors,[145] the ordered list of gene targets up to the 3000th position was used as a background for the analysis; the test set in each case was the top 10% of targeted genes.

Gene Targeting of Enriched Families

Five families were enriched in both male and female cells, and 12 families in only one of the two cell lines (Fig. 3.10 A, left side). The size range of enriched families was substantial, from small families with only 4 mature members to extensive families with dozens of mature miRNAs. Of note, the amount of family members in any miRNA family did not correlate with the absolute amount of targets predicted (Fig. 3.10 A). Rather, the influence of individual miRNAs was the main factor determining the size of the gene target network. However, those families that were enriched in only one cell line presented with significantly smaller target sets than those that were found DE in both (mean targeted genes per miRNA 217 versus 378, Welch two-sample t test, $p = 0.001$). Relative to family size, 4 of the enriched families targeted less genes than all others: mir-10 ($p = 0.016$), mir-192 ($p = 0.042$), mir-379 ($p = 0.011$), and mir-515 ($p < 0.001$). Hypothetically, the spectrum of target amounts may correlate with the degree of functional specification of distinct miRNA families: on one end, broadly acting families such as let-7 with sex-independent function, on the other, families with a narrow target profile, such as mir-10, whose restricted function can associate with sex-specific effects.

which/how many of the pertinent mirs above are in families?

3.5.4 Large Scale GO Term Curation

The GO analysis performed in this manner for all 17 enriched families resulted in a list of 737 distinct GO terms related to any of the families. To generate an overview of functional implications of the individual families, the GO terms were filtered and aggregated manually. Terms not relating to CNS- or immune-function were removed, and the remaining terms were sorted into one of 21 categories (Fig. 3.10 B). Generally, the 17 miRNA-families associate with neurodevelopment and neural plasticity, diverse immune functions, cell cycle control, and sex. Most families present with association to general ontological categories such as neurodevelopment or sex, while more specific categories show a sparser distribution.

Only two families associate significantly with neurokine-related function, mir-10 and mir-199. Both are, as many others, involved in neurodevelopment- and sex-related function, but both also

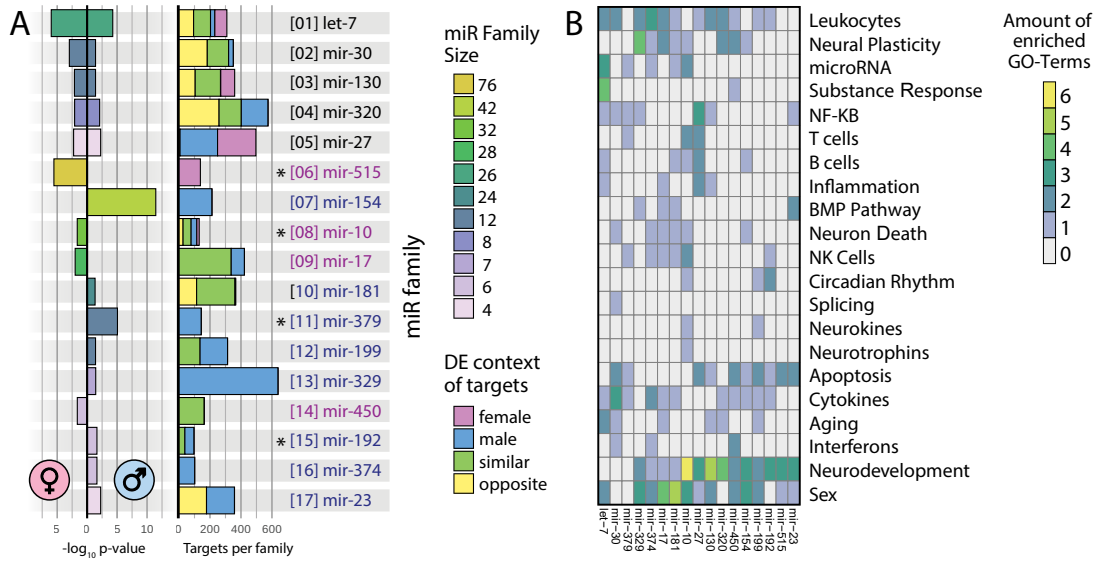

Figure 3.10: miRNA Families Enriched in Differential Expression and their Ontological Associations. 17 miRNA families were enriched significantly in the DE miRNAs following CNTF-mediated differentiation of LA-N-2 and LA-N-5 (Fisher's exact test, $p < 0.05$). **A, left side)** Bar plot of p-values of enriched families, ordered by family size; family size encoded by colour. **A, right side)** Stacked bar plot of the number of gene targets per family. Bars are divided by the DE pattern between LA-N-2 and LA-N-5 of each individual family member. DE context (encoded by colour) varies from detection in all categories (such as let-7 or mir-10) to detection only in one cell line (such as mir-515 or mir-154). Four families show significantly less target genes than all other families in relation to their size (denoted by asterisks). **B)** Gene Ontology enrichment analysis of gene targets of all enriched families via 737 distinct terms curated into CNS- or immunity-related categories. The miRNA families mir-10 and mir-199 show association with neurokines and circadian rhythm.

show the very specific association with circadian rhythm. Family mir-10 additionally is implicated in control of neurotrophin-related mechanisms, and in several blood-borne immune cells, such as T-, B-, and NK-cells.

3.6. Whole Genome miRNA→Gene Network Generation

A common approach to complex network relationships is physical modelling. A complex graph (with directed and weighted edges) can be coerced to self-organise by application of a force-directed layout. In this process (also known as spatialisation), the network, defined only by its nodes and edges, is transformed into a map, usually in two dimensions. An important prerequisite is the scale-free topology of the network, a structure that transcriptional connectomes usually present with.[146] A force-directed layout transforms a network by simulating a gravitational system, or a system of magnetic nodes connected by springs, in which the nodes repel each other, but edges between two nodes pull them towards each other. By manipulation of multiple physical attributes of the model, a mapped representation of the network's organisation can be produced. As a result, nodes (i.e., genes, TFs, and miRNAs) with close interaction are mapped in close proximity, while nodes with low interaction are far apart. Similarly, nodes with pivotal function in the network (»hubs«) gravitate towards the centre of the map, while »less important« nodes are shifted towards the fringes.

The network comprising all DE members of the 17 enriched miRNA families and 12 495 targeted genes as

determined via *miRNeo* query was subjected to force-directed mapping using the Java-based software Gephi 0.9 and its primary force-directed algorithm, ForceAtlas2.[147] Gephi, and ForceAtlas2, are designed to generally handle graphs with up to 10 000 unique relationships; however, the standard *miRNeo* query resulted in a network with ~160 000 edges. To reach a computationally manageable number of relationships, the score threshold was raised to a minimum of 7, which resulted in a network of 46 937 unique edges. The resulting network was exported as a vector graph and manually edited in Adobe Illustrator to further enhance its readability (Fig. 3.11).

The resulting transcriptional connectome map illustrates the functional compartmentalisation of miRNA→gene interactions. miRNAs of distinct families are frequently found in close proximity to one another, most often forming one or two clusters. In the case of two clusters forming, the clusters are usually representative of the two complementary strands of the pre-miRNA(s), since 3' and 5' variants of any pre-miRNA usually possess fundamentally different seed sequences, and thus, targets. The let-7 family is distinguished by its removal from the bulk of other interactions, possibly representing a particularly specialised set of functions, at least for the 5' variants of the bottom cluster. Families with predominant differential expression in one of the two cell lines (sexes) inhabit different sides of the main graph and show little intermingling, pointing towards sexually dimorphic gene target distribution. The two neurokine-associated families, mir-10 and mir-199, are located near the centre of the graph, in two strand specific clusters (»[08a]&[12a]« and »[08b]&[12b]«).

evolutionary? discussion?

To gather more detailed information than grouping of miRNAs with similar function, such as direct miRNA→gene interaction, the size of the studied networks must be reduced. For each family affected by CNTF-differentiation, a single graph was created, laid out by application of ForceAtlas2, and analysed for critical nodes. The distinct families and their gene targets yield immensely diverse graph layouts, that here cannot be described in their entirety. However, the entire collection of graphs in interactive visual form is accessible at https://slobentanzer.github.io/cholinergic-neurokine. Due to an elevated interest, the cholinergic/neurokine miRNA interface and the families mir-10 and mir-199 will be described in more detail, and in conjunction with sex-specific perturbations in neurologic diseases.

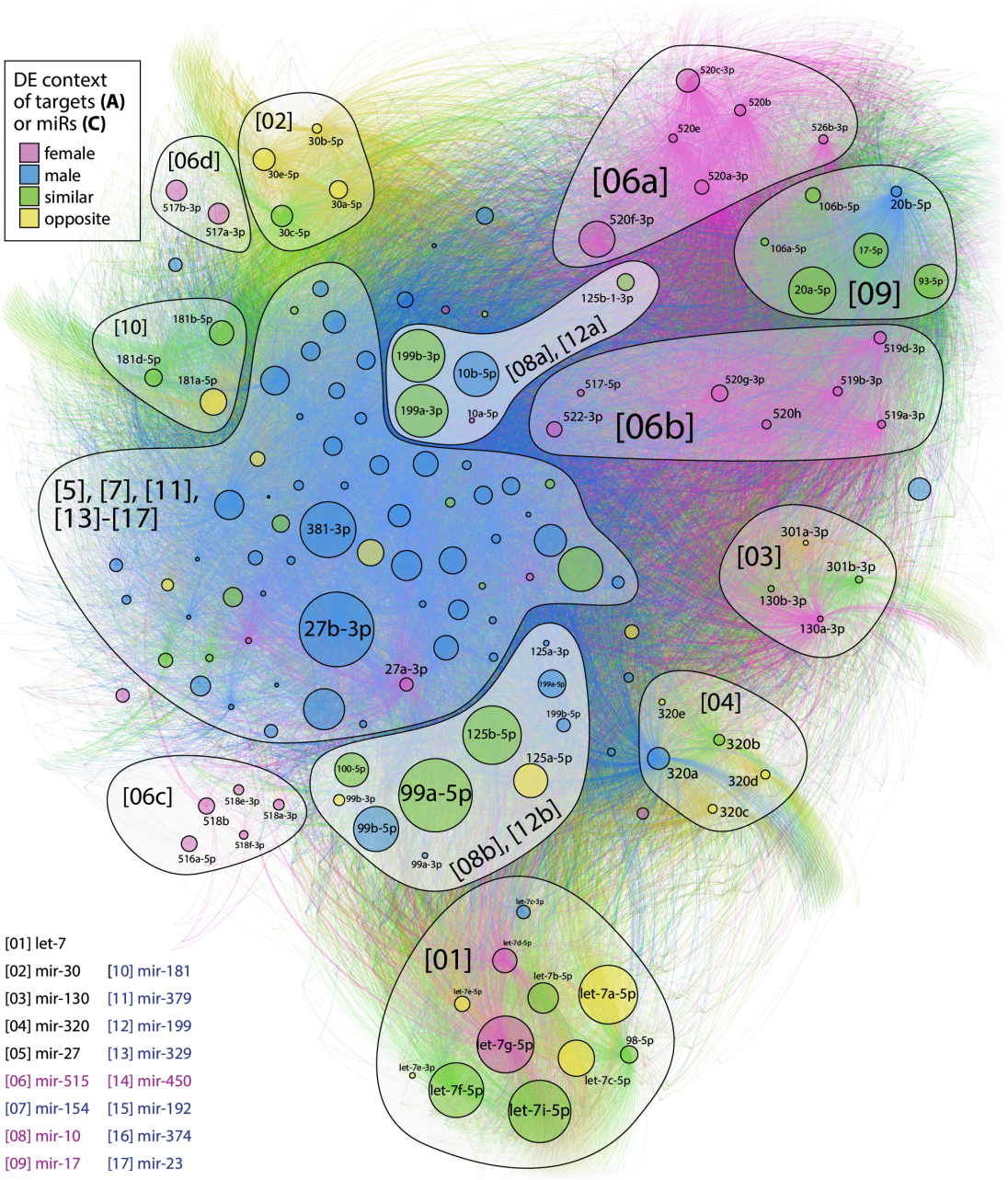

Figure 3.11: Full Connectome of LA-N-2 and LA-N-5 Differentially Expressed miRNA Families. The network of miRNA families and their 12 495 targeted genes self-organises into a connectome map with 46 937 unique edges. miRNA node size scaled by absolute count-change, nodes coloured by DE context. Numbers in brackets denote miRNA families, gene nodes have minimal size. By application of a force-directed layout, the miRNA families visibly self-segregate into clusters. The let-7 family, male-biased and female-biased clusters take up major parts of the network. Families mir-10 and mir-199, with neurokine association, form two mixed, sexually dimorphic clusters near the centre of the map (lighter shade).

3.7. Application to Schizophrenia and Bipolar Disorder

A comprehensive structural analysis of perturbations on a genome scale is hardly possible without heavy truncation of results or dimensionality reduction methods. Truncation is commonly performed by ranking perturbations by their p-values in ascending order and only regarding the highest ranked entries, which often amounts to less than ten individual transcripts. On the other hand, commonly used dimensionality reduction techniques include principal component analysis (PCA), t-distributed stochastic neighbour embedding (t-SNE), and clustering/stratification approaches. While truncation enables human-readable presentation of results, in principle it does not lend itself to complex polygenic events such as neurologic disease. Common dimensionality reduction techniques are useful in providing structural overview of a high-dimensional dataset, but give little insight into causal relationships of single entities. We thus aimed to find an alternative approach to dimensionality reduction which conserves the internal relationships inherent to the data, and which profits from the network organisation of our input data.

For the application of *miRNeo* data to real-world problems, suitable psychiatric and neurologic disease datasets were sought in the common repositories ArrayExpress, NCBI GEO, and Synapse. Among the datasets with agreeable quality, SCZ and BD were the only diseases with sample amounts that allowed a statistically valid analysis of sexual dimorphisms. While many neurologic disease studies are simply limited in their number of subjects, autism presented a different issue: the majority of donors were males (more than 90%). Direct analysis of miRNA expression patterns was not possible, because very few studies study miRNAs directly, yet. Thus, studies on mRNA were substituted to infer on miRNA dynamics.

3.7.1 Analysed Datasets

Twelve datasets including 1361 subjects were downloaded from their repositories (Table 3.2). Data of DLPFC RNA-seq of 579 SCZ patients and controls was obtained from the Common Mind Consortium (http://www.synapse.org/CMC). To address the diverse origins and technological aspects of data, care was taken to appropriately unify and normalise the data. The data preparation and meta analyses were performed essentially as described by Gandal and colleagues.[32] Samples of brain regions not consistent with the research question (e.g., cerebellum), or from patients with diseases other than SCZ or BD, were removed from datasets on a case-by-case basis. RNA-seq datasets were used to individually confirm the perturbations found in the meta-analysis of microarray studies.

3.7.2 Microarray Quality Control and Data Preparation

Read-In and Normalisation

Illumina datasets were read, $\log_2$-transformed, and quantile-normalised using R/lumi.[160] Affymetrix datasets were read and RMA-normalised ($\log_2$-transformed, background corrected, quantile-normalised) using R/affy.[161] Affymetrix data were additionally corrected for 3'/5' bias using the *AffyRNAdeg()* function (not available for other

source	accession	publication	technology	subjects/samples	disease
NCBI GEO	GSE35978	Chen *et al.* [148]	microarray	150/312	SCZ & BD
NCBI GEO	GSE12649	Iwamoto *et al.* [149]	microarray	102/102	SCZ & BD
NCBI GEO	GSE53987	Lanz *et al.* [150]	microarray	76/205	SCZ & BD
NCBI GEO	GSE17612	Maycox *et al.* [151]	microarray	51/51	SCZ
NCBI GEO	GSE21138	Narayan *et al.* [152]	microarray	30/59	SCZ
NCBI GEO	GSE5392	Ryan *et al.* [153]	microarray	82/82	BD
NCBI GEO	GSE80655	Ramaker *et al.* [154]	RNA-seq	96/281	SCZ & BD
NCBI GEO	GSE106589	Hoffman *et al.* [155]	RNA-seq	94/94	SCZ
NCBI GEO	GSE68559	Webb *et al.* [156]	RNA-seq	10/98	NA
NCBI GEO	GSE96659	Fontenot *et al.* [157]	RNA-seq	5/209	NA
NCBI GEO	GSE45642	Li *et al.* [158]	RNA-seq	86/670	NA
Synapse	CMC	Gulyás-Kovács *et al.* [159]	RNA-seq	579/579	SCZ & BD

Table 3.2: Data Sources for Microarray and RNA-seq Analyses.

chip manufacturers). All available biological (e.g., sex, age) and technical (e.g., batch, RIN, post-mortem interval) covariates were collected and used for the analysis. Individual correlations of case-control status S with any covariate C were assessed using a linear model (R/lm) with formula $C \sim S$; statistical significance was determined via ANOVA (R/anova). If necessary, case-control samples were balanced to eliminate significant covariate correlations with case-control status (all p > 0.05).

Outliers

Outlier removal was performed using the method proposed by Oldham, Langfelder & Horvath.[162] Briefly, the (dis-)similarity matrix of samples is transformed into a signed, weighted correlation network. Network adjacency (a) of samples (nodes) S_i and S_j is defined as:

$$a_{ij} = \left(\frac{cor(S_i, S_j) + 1}{2} \right)^2$$

As such, the connectivity between samples can be measured by the standardised connectivity (*Z.K*), which describes the strength of correlation between any given node and all other nodes in the network. As proposed by Oldham *et al.*[162], outliers were removed if their Z-score was below the threshold of *Z.K* = -2.

Annotation

To enable comparison between datasets of diverse technical origin, probes were annotated using ENSEMBL gene identifiers using R/biomaRt.[163] To maintain comparability with the analysis by Gandal *et al.*,[32] the same version of ENSEMBL DB (v75, Feb 2014) was used. Probes were collapsed onto single genes using the *collapseRows()* function of R/WGCNA,[164] using the maximum mean signal across all probes per gene. Of note, information loss occured by multiple collapsing of probes and integration of datasets, which can only be performed using the genes common to all datasets (i.e., represented by microarray probes). The final gene set encompassed 12 391 individual genes, with several notable cholinergic/neurokine exceptions (CHRNA7, CHRM1, LHX8, CHKB, PRIMA1, CNTF). Missing genes result from annotation deficits between different probe sets, cannot be comprehensively manually controlled on a genome scale, and cannot be re-introduced at this stage.

3.7.3 Differential Expression Meta-Analysis

The individual experimental datasets were each corrected for covariate influences by multiple regression based on all available biological and technical covariates. Briefly, the linear regression model was solved using matrix algebra operations. In matrix form, a linear regression model of observations Y (i.e., gene expression levels), independent variables X (i.e., covariates), coefficients β, and error terms ε can be described as:

$$Y = X\beta + \varepsilon$$

As a consequence, the residual sums of squares can be expressed as the cross product:

$$RSS = (Y - X\beta)^T (Y - X\beta)$$

Then, the coefficients $\hat{\beta}$ can be estimated by solving the derivative:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Coefficients were estimated for all relevant technical and biological covariates (e.g. post-mortem interval, RIN, sex, age) and used to regress covariate influence on gene expression levels:

$$Y_{new} = Y - (X\hat{\beta})^T$$

After covariate regression, differential expression was calculated across all datasets for each disease group using a linear mixed model with a fixed effect for each study and case-control status (»group«), and a random effect for each individual subject. Computation was performed in R, using R/nlme, [165] with parameters

$$fixed = \sim group + study \text{ and } random = \sim 1|subject$$

This yielded an array of log-fold changes between cases and controls for each gene and disease. To determine statistical significance, 10 000 permutations of the mixed-model regression were performed for each use case, randomly assigning case-control status. The resulting null distributions were used to determine FDR, with threshold for significance at 0.05.

Sex-Specific Meta-Analysis

Samples of all datasets were split between males and females (cases as well as controls), and individually subjected to the same procedure as the sex-independent data: covariate regression, differential expression via a linear mixed model, and estimation of statistical significance via permutation testing.

Transcriptome Correlation

Correlation of disease transcriptomes was performed by using Spearman's rank correlation coefficient. Spearman's ρ was determined between SCZ and BD sex-independently as well as separately in males and females.

Most Diverging Genes

Genes were ranked by their divergence between any two compared datasets, sex-independent data of SCZ and BD, and any meaningful combination of sex-dependent data in SCZ, BD, males, and females. The divergence δ of any gene G between datasets i and j was defined as:

$$\delta = LFC(G)_i - LFC(G)_j$$

Where positive values of δ indicate a positive bias of G towards dataset i; LFC: $\log_2$ fold change.

3.7.4 Sexual Dimorphism in Schizophrenia and Bipolar Disorder

Sex-independent correlation replicated the finding of Gandal *et al*,[32] with Spearman's $\rho = 0.7100$ ($p < 0.001$). However, diverging from the established annotation of ENSEMBL v75 to later versions of the database significantly altered the correlation coefficient, leading to lower correlation in all tested cases. Comparing the sex-independent data with only male or female subjects (all with ENSEMBL v75 annotation), those also show lower general correlation between SCZ and BD: in females, correlation was $\rho = 0.6150$ ($p < 0.001$), in males, $\rho = 0.5783$ ($p < 0.001$). While it is possible that these variations are caused by structural properties of the data unrelated to sexual dimorphism, such as the loss of power due to the reduction in size, the consistently lower correlation in sex-specific subsets also may indicate an averaging effect between male and female patients, leading to a higher correlation in spite of significant sexual dimorphism.

To address the potential differences between male and female brain transcriptomes, which may reflect the observed clinical dimorphism, we subjected the 100 most-diverging genes between any two datasets to GO enrichment analysis (Fig. 3.12, from Lobentanzer *et al.* [1]) in hopes of identifying the most discriminating molecular pathways between SCZ and BD, and afflicted males and females. Sex-independently, the most-diverging pathways between SCZ and BD principally involved mechanisms of inflammation and immunity (e.g., "acute inflammatory response," $p = 0.003$; "cellular response to cytokine stimulus," $p = 0.01$).

Differences in Sexual Dimorphism Between SCZ and BD

Computation of diverging pathways between males and females in each disease indicated a larger divergence between sexes in SCZ than in BD. SCZ-biased genes of males and females showed no overlapping GO terms (Fig. 3.12 A), but BD-biased genes of males and females showed large GO term overlap, particularly in inflammatory components (Fig. 3.12 B). Notably, specific components of neurokine signalling were elevated in both males (IL-6, $p = 0.007$) and females (JAK/STAT, $p = 0.01$) with BD.

Overlap of Male-Biased Genes Between SCZ and BD

Shared transcriptional properties of SCZ and BD were identifiable only in male-biased genes. While female-biased SCZ genes showed no implications in CNS processes, male-biased SCZ and BD genes overlapped in functions concerning inflammation and immunity (Fig. 3.12 C). Female-biased BD genes were associated with CNS function and development (Fig. 3.12 D).

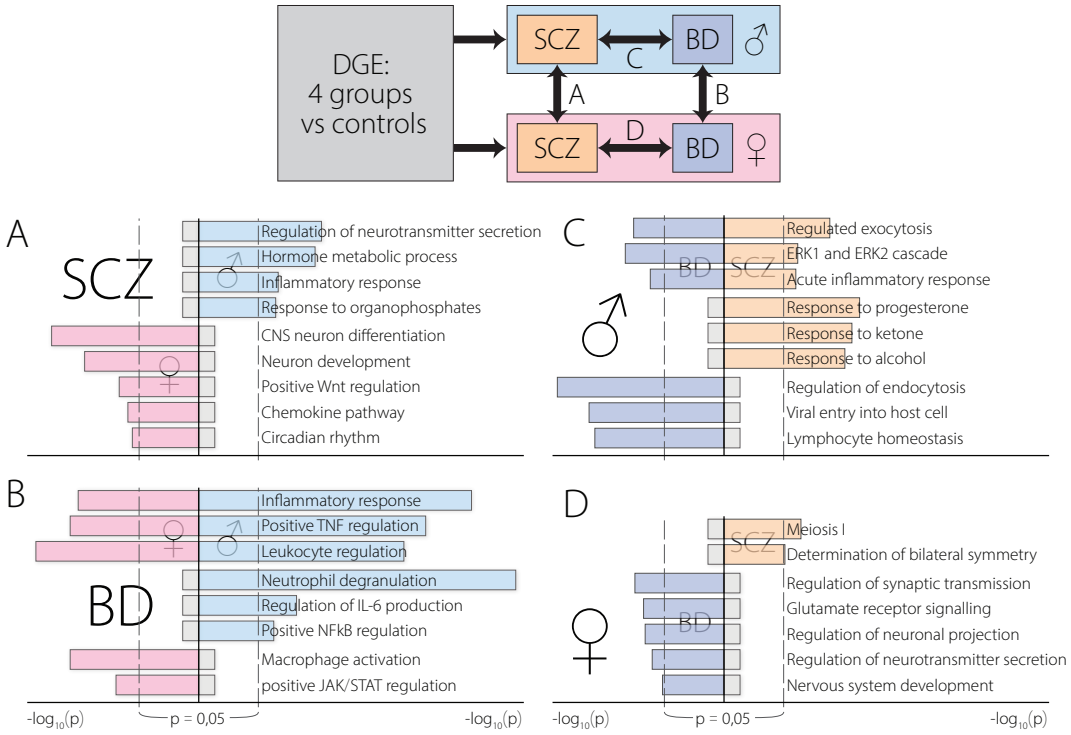

Figure 3.12: GO Enrichment of Diverging Genes. Results of differential gene expression were dually compared: SCZ versus BD and male versus female (indicated by colours). GO enrichment of the top 100 distinguishing genes in one dimension was compared with the other for each pair of combinations. **A)** SCZ-biased genes diverge between males and females. **B)** BD-biased genes share immunological ontology in both males and females. **C)** Male-biased genes share immunological ontology in BD and SCZ. **D)** Female-biased genes diverge between SCZ and BD.

Specificity of Ontological Terms

When comparing different areas of biological function, such as neurotransmission, immunity, and inflammation, the different areas notably diverged in the specificity of identified terms. Considering the functionality of the applied method (*topGO*, see Section 3.5.3) to find the most specific node in any branch of the DAG tree while disregarding its (less specific) parent nodes, this may indicate a difference in the magnitude of perturbation in the different systems. For instance, the GO terms indicating neurotransmission as affected system were much less specific than those indicating immunity-related processes. While significant neurotransmission-related terms failed to implicate specific neuron types or neurotransmitters (e.g., GO:0021953, »CNS neuron differentiation«; GO:0046928, »Regulation of neurotransmitter secretion«), immunity-related terms were very specific towards reg-

ulatory subsystems, and regularly implicated neurokine mechanisms (e.g., GO:0032675, »Regulation of interleukin-6 production«; GO:0046427, »Positive regulation of JAK/STAT cascade«).

3.7.5 Combination of Disease Data and Cell Culture

To implement the proposed complexity reduction technique, we applied a reductionist approach to the comprehensive network generated from perturbed miRNA families and their targeted genes (Fig. 3.11), based on the unbiased analysis of sexual dimorphism in SCZ and BD, which implicated processes of neuronal, immunological, and circadian origin (Figure 3.12). To merge these results with the implications of cholinergic cell culture, we added genes implicated in neurokine signaling and circadian rhythm to the list of cholinergic genes (see Box 1). Returning to the collection of web-available patient data, we subjected this limited set of 76 genes and their 18 neuronal TFs to differential expression analysis.

The comprehensive network was then filtered in multiple consecutive steps. (I) Permutation analysis of comprehensive miRNA targeting data specific for genes expressed in cholinergic neurons (Fig. 3.2) yielded a list of miRNA candidates that shows overlap with (II) miRNAs DE in our two models of neurokine-induced cholinergic differentiation (Fig. 3.7 A). (III) We included only families of miRNAs we found to be enriched in differential expression (Fig. 3.10). Sixty-nine miRNAs from 12 families passed this filtering process and were consecutively assembled in a force-directed network with the 94 genes of the previously compiled list. As a »spike-in«, we added miR-132-3p (DE in LA-N-5 cells), a miRNA which controls cholinergic processes [166,167] and is known for its function in neurons [168] and immunity [169] and its perturbation in disease. [170] The resulting network (Fig. 3.13 A, from Lobentanzer *et al.* [1]) shows high structural homology to the comprehensive network shown in Figure 3.11. The miRNA families in this reduced network show spatial organisation similar to the comprehensive network.

In agreement with their localisation in the comprehensive network, miRNA families mir-10 and mir-199 inhabit a central role in the resulting interactome. Most-targeted genes in this network (as indicated by their size) are the circadian regulators CLOCK and RORA. While CLOCK is located centrally, next to mir-10/199 miRNAs, RORA shows closeness to the mir-30/515 families. Generally, genes with larger cellular influence, such as transcription factors (STAT3, CLOCK) or TGF-β ligands (BMP family genes) are frequently targeted by miRNAs, while more specific transcripts, such as the cholinergic receptor genes or neurokines, are targeted more selectively.

more details?

More so than the spiked-in miR-132-3p, mir-10/199 miRNAs target cholinergic genes, for instance, the neuronal nicotinic α4β2 and muscarinic M1 receptors (mir-125) and HACU (miR-199). In addition, they target neurokine genes, such as the transmembrane neurokine receptor LIFR or STAT3, and circadian regulators (e.g., CLOCK and RORA). The two families react highly sexually dimorphic to CNTF-mediated differentiation; some are detected as DE only in one cell line, others exhibit inverted changes between cell lines. The 3p-variant of miR-125a distinguishes itself from the

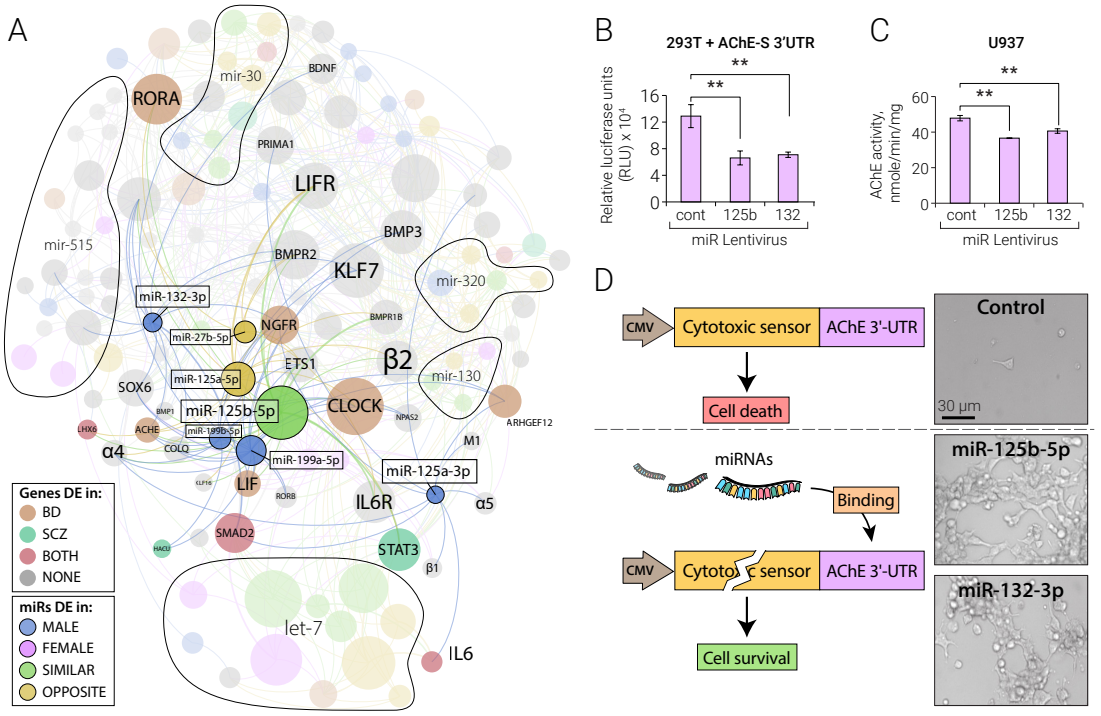

Figure 3.13: The cholinergic/neurokine interface. A) The miRNA families mir-10 and mir-199 pose a sexually dimorphic interface of cholinergic, neurokine, and circadian regulation by targeting nicotinic/ muscarinic (e.g., a4b2 and M1) and neurokine receptors, transcriptional regulators of cholinergic differentiation (LHX and STAT) and circadian rhythm (CLOCK and RORA), the AChE and the AChE linker proteins PRIMA1/COLQ, and high-affinity choline uptake (HACU). Members of mir-10/199 families, spike-in miR-132-3p, and their targeted genes are shown in colour, and other miRNA families that passed the multiple filtering are indicated as areas. miRNA node size corresponds to count-change and gene node size to connectivity; colour and thicker edges indicate the DE context and experimentally validated connections. **B–D** Validation experiments of AChE targeting by miR-125b-5p, with miR-132-3p as a positive control. **B)** Lentiviral expression of miR-132 and miR-125b suppresses luciferase fused to the 3' UTR of AChE in HEK293T cells. Error bars indicate SE. **C)** Lentiviral expression of miR-132 and miR-125b suppresses the endogenous AChE hydrolytic activity of U937 cells with similar efficacy. Error bars indicate SE. **D)** Life/death assay of stably transfected HEK293T cells carrying the AChE 3' UTR fused to a cytotoxic sensor and co-transfected with miR-125b-5p, miR-132-3p, or control plasmids. Cells survive in case of binding of miR-132-3p and miR-125-5p to the 3' UTR.

bulk of mir-10/199 miRNAs by exclusively targeting M1, α5 and β1 receptors, and IL-6, and thus is slightly removed from the centre of the network.

Describe other members?

The miRNA with most targets in this reduced interactome is miR-125b-5p, also displaying most experimentally validated interactions with neurokine genes (miRTarBase accessions: IL-6, MIRT022105; IL-6R, MIRT006844; JAK2, MIRT734987; LIF, MIRT001037; LIFR, MIRT732494; STAT3, MIRT005006). miR-125b-5p also is the most perturbed miRNA (in this interactome) upon CNTF-mediated differentiation (highest absolute count-change), and the only member of mir-10/mir-199 changed in similar direction in both cell lines (up-regulated). miR-125b-5p also targets multiple other inflammation-related genes (e.g., TNF, MIRT733472; IRF4, MIRT004534) and 5-lipoxygenase, which can influence inflammatory processes via production of eicosanoids.[171] miR-125b-5b has been directly associated with cytokine-mediated inflammation, as its over-expression increased the expression of TNF-α, IL-1β, and IL-6, and markedly decreased IκB-α.[172]

A notable intersection of spike-in miR-132-3p and miR-125b-5p is the ACHE, an interaction

which had not been validated for miR-125b-5p, but is known for miR-132-3p. [169,166] Using miR-132-3p as a positive control, we performed ACHE-mRNA binding assays in validation of the predicted targeting by miR-125b-5p.

3.7.6 miR-125b-5p Acetylcholinesterase Targeting Assays

We performed three independent cell culture assays to confirm *ACHE* mRNA targeting by hsa-miR-125b-5p: luciferase suppression, AChE protein activity, and a cell death assay with a cytotoxic sensor. The 3' UTR of human *ACHE* mRNA [173] was cloned into the microRNA Target Selection System plasmid (System Biosciences, CA, USA) multiple cloning site, using EcoRI and NotI restriction enzymes (New England Biolabs). All plasmids were verified by DNA sequencing. For luciferase assays, HEK293T cells were transfected with miRNA Target Selection-AChE-3' UTR, and selected in the presence of Puromycin for 3 weeks. Stably transfected HEK293T (293T-AChE 3' UTR) cells were grown on 12-well plates and infected with lentiviruses expressing miR-125b-5p, miR-132-3p or a negative control sequence. After 48 hours incubation, cells were analysed using the Dual Luciferase Assay kit (Promega, WI USA) and Luciferase activity was measured using an Envision luminescent plate reader (Perkin-Elmer, Waltham, MA), essentially as previously described by Hanin *et al.* [174] For each reporter construct, renilla luciferase activity was normalized according to that of the firefly. Normalised activity after infection with miR-132-3p or miR-125b-5p was expressed as relative to that obtained after infection with the same plasmid with miRNA negative control. To show effects of changes in this miRNA's levels on real-life protein activities, we performed an AChE hydrolytic activity assay following infection of human monocyte-like U937 cells with hsa-miR-125b-5p, miR-132-3p or a negative control lentiviral vector. AChE hydrolytic activity levels were assessed by kinetic measurements of the hydrolysis rates of 1 mM acetylthiocholine (ATCh, Sigma) at room temperature, following 20 min incubation with and without 50 µM tetraisopropyl pyrophosphoramide (iso-OMPA, Sigma), a specific inhibitor of butyrylcholinesterase, to selectively assay for AChE-specific or total cholinesterase activity. For the life/death assay, stably transfected HEK293T cells were infected with lentiviruses expressing miR-125b-5p, miR-132-3p or a negative control sequence. 72 hours post-infection, a cytotoxic reporter fused to AChE 3' -UTR was added to the media and cells were kept for an additional 5 days to assess their viability. For all cell culture assays, statistical significance was determined using ANOVA with correction for multiple testing. Each sample was assayed in at least 3 biological replicates, and in all cases, hsa-miR- 132-3p served as a positive control.

3.7.7 hsa-miR-125b-5p Targets Acetylcholinesterase

In all tested conditions, miR-125b-5p suppressed *ACHE* mRNA with equal potency as the positive control miR-132-3p (Fig. 3.13 B-D). Towards mRNA expression (luciferase) and functionality (cytotoxic sensor) as well as on protein level (AChE activity), miR-125b-5p demonstrated its interaction with ACHE mRNA 3' UTR. Luciferase units after miR-125b-5p transfection were approximately halved, indicating significant transcript degradation of the *ACHE* 3'UTR.

3.7.8 Cholinergic/Neurokine Mechanisms in Web-Available RNA Sequencing Experiments

To include recent developments in methodology, we analysed several recent RNA-seq studies addressing related questions. In a study of post-mortem brain transcriptome profiling of psychiatric disorders,[154] we found a down-regulation of IL-6, LIF, and several cholinergic receptors (M2, M4, $\alpha 4$, $\beta 2$, $\alpha 7$), with sex-specific differences (males had significantly higher levels of neurokines than females). These changes were visible only in SCZ patients, not in BD or major depressive disorder. In a study of induced pluripotent stem cells (iPSCs) of SCZ patients and controls that were induced to show a neuronal phenotype,[155] we found an up-regulation of *CHAT* in SCZ-derived iPSCs, and a down-regulation of IL6R and the nicotinic $\alpha 6$ subunit. In this study, SCZ males showed a higher expression of the *SLC18A3* and lower expression of nicotinic subunits α 2, 7, and 9, and $\beta 3$. In a study of differentiated human neuronal progenitor cells,[157] a knockdown of the circadian transcriptional controller CLOCK resulted in up-regulation of LIF and simultaneous down-regulation of neurokine transmembrane receptors LIFR and IL6ST, accompanied by slight bi-directional changes in several cholinergic receptors.

I know words. I have the best words.

Donald Trump

4

Dynamics Between Small and Large RNA in the Blood of Stroke Victims

Stroke is a dramatic incision into bodily homeostasis and affects a multitude of organ functions, first and foremost the brain. The immediate actions upon stroke are focused on preserving as much functional tissue as possible, so as to alleviate the cognitive damages resulting from neuron death. After this initial period of few hours, longer-lasting processes determine the health and recovery of the patient. Many of these later events are related to immunity. The greatest danger to the patient after survival of the initial period are infections, such as pneumonia, usually between one and two weeks after the infarction. Pneumonia is often facilitated by aspiration of liquids or solids when the swallowing mechanism is impaired as a consequence of the cerebral damage. However, as introduced in Section 1.2.5, stroke-related immunodepression can play a role in post-stroke survival, and has been shown to have an impact on the transcriptome of blood-borne immune cells, at least for protein coding genes. The role of short RNA transcripts, and particularly of transfer RNA fragments, is much less clear. We thus opted to analyse the blood of stroke victims taken upon hospitalisation and screen it for changes in small and large RNA expression.

4.1. RNA Sequencing, Differential Expression, and Descriptive Methods

4.1.1 The PREDICT Cohort

The patient collective for the present study was recruited from a prospective, international, multi-center study with 11 study sites in Germany and Spain, led and approved by the neurologic department of Charité Berlin (www.clinicaltrials.gov, NCT01079728).[175] The study, called PREDICT, screened 484 stroke patients for clinical attributes and conventional biomarkers, with daily measurements in the first 5 days after stroke, and a three months follow-up. From these patients, a representative cohort of 49 patients were selected for blood small RNA sequencing.

4.1.2 Clinical Parameters Collected in the PREDICT Study

Stroke patients were assessed daily for the duration of hospitalisation, at least until four days after admission. Blood-based biomarkers that were measured at least once during this period include: monocyte human leukocyte antigen isotype DR (HLA-DR); interleukins IL-6, IL-8 and IL-10; IL-10 levels after 24h *in vitro* stimulation with lipopolysaccharide; lipopolysaccharide binding protein (LBP); mannan-binding lectin (MBL); and TNF-α. Also recorded were the time between admission and the collection of the blood sample, and the modified Rankin Scale (mRS). This scale is a rough categorisation of the severity of stroke, with 0 referring to no symptoms, and 6 signifying death. Scores 1-2 describe slight neurological deficits, 3 requires frequent help because of medium level deficits, 4 requires constant assistance with daily tasks, and 5 requires stationary care.

4.1.3 Sample Collection, RNA Isolation, and Sequencing

Blood was collected into RNA stabilising tubes (Tempus Blood RNA tubes, Applied Biosystems) on each day of hospitalisation, and we subjected blood samples collected on the second day to small and large RNA-sequencing. For sequencing, we only considered samples from patients with modified Rankin Scale (mRS) values of 3 and below at discharge from the hospital, to exclude very severe cases of stroke, leaving n=240 relevant cases. The time from stroke occurrence to blood withdrawal varied between 0.94 to 2.63 days, with an average of 1.98 days. Blood samples from age- and ethnicity-matched healthy controls were obtained at matched circadian time from donors with ethical approvals from institutional review boards (ZenBio, North Carolina, USA).

RNA was extracted from 3 ml of whole blood of all 484 PREDICT patients using the Tempus Spin RNA isolation kit (Invitrogen, Thermo Fisher Scientific, Waltham MA, USA). RNA quality was determined by RNA gel for all samples and by Bioanalyzer 6000 (Agilent, Santa Clara CA, USA) for samples selected for RNA-sequencing, which showed high RNA quality with a median RIN of 8.8 (lowest RIN 7.9, highest RIN 9.9). We used 600 ng total RNA of 49 samples for small RNA library construction (NEBNext Multiplex Small RNA library prep set for Illumina, New England Biolabs, Ipswich MA, USA) and selected 24 out of the 49 short RNA-sequenced samples for PolyA-selected mRNA sequencing. These libraries were prepared from 1000 ng total RNA using the TruSeq RNA library preparation kit (Illumina, San Diego CA, USA) and were sequenced on the Illumina NextSeq 500 platform at the Hebrew University's Center for Genomic Technologies.

4.1.4 RNA Sequencing Alignment

Small RNA species were aligned after quality filtering using flexbar and miRExpress 2.0, as described in Section 3.4.2. Additionally, to assess tRF expression, small RNA reads were aligned to the exclusive tRNA space using the MINTmap pipeline.[102] Briefly, this pipeline compares short RNA sequencing reads with a collection of sequences determined to only be contained inside mature tRNAs, without confounding from the many tRNA lookalikes in the human genome, e.g., in pseudogenes. The two RNA species were united into one expression matrix containing both miRNA and tRF expression.

Large RNA species were aligned to the human transcriptome (ENSEMBL transcriptome *Homo sapiens* GRCh38 release 79) using the fast dual-phase parallel inference algorithm *Salmon*.[176] Briefly, the method combines an »online« fragment mapping utilising continuous updating of a Bayesian prior with an »offline« phase that determines fragment quantities by application of the Bayesian model determined before via a standard expectation maximisation (EM) algorithm or a variable Bayesian EM. Additionally, the pipeline corrects for multiple typical biases in sequencing, such as position-specific biases, sequence-specific 3' and 5' end biases, fragment GC content bias, and fragment length distribution. The resulting quantified fragments were imported into R using the rsubreads package.[177]

4.1.5 Quality Control and Filtering

Raw and processed reads were quality-controlled using FastQC, as described in Section 3.4.1, with no samples falling below acceptable thresholds. Small and large RNA alignments were batch-corrected followed by analysis of inter-sample relationships via the method proposed by Oldham *et al.*[162] (as described in Section 3.7.2). We excluded no large RNA samples and one small RNA sample (»11_40044_S12«).

4.1.6 RNA Sequencing Differential Expression Analysis

Quantified reads were subjected to differential expression analysis using DESeq2, essentially as described in Section 3.4.3. Small RNA species were analysed together by combining count tables for miRNAs and tRFs, large RNAs were analysed separately. Both datasets were corrected for covariates *subject age* and *batch*. Correction for patient sex was not necessary because all patients in the final analyses were male. LFC values were shrunk using *apeglm* as described in Section 3.4.3, at an alpha level of *0.1*.

4.1.7 Gene Ontology Analyses

We performed GO analyses on the set of DE transcripts, using different ranking methods. GO analyses were performed using R/topGO as described in Section 3.5.3 using the weighted method.

Ranking by P-Value

Transcripts were ranked by p-value, and different test sets were tested against the background of the topmost two thousand transcripts. We tested the set of all DE transcripts (adjusted p-value < 0.05) and the separate sets of positively and negatively regulated transcripts. Additionally, for each test group, the criterion of LFC > 1.4 was applied and re-tested.

Ranking By Count-Change

Alternatively, transcripts were ranked by count-change, and the top 100 significantly DE transcripts were tested against the background of the first two thousand transcripts. Similarly to the p-value ranking, test sets comprised all transcripts as well as only negatively or positively regulated transcripts.

Visualisation of Results

While GO enrichment analysis can be informative, interpretation and visualisation of its results is not standardised and often limited to presentation of top *X* terms by p-value. R/gsoap [178] is an analysis tool proposed to aid in interpretation of GO enrichment results via t-distributed stochastic neighbour embedding (t-SNE) display of similarity of terms based on the amount of shared significant genes. GO enrichment results were processed to fulfil *gsoap* input criteria and visualised using *ggplot2*. [179]

4.1.8 Homology Computation Among tRNA Fragments

Transfer RNA fragment origin can be ambiguous, even in fragments derived from tRNA-exclusive space. To assess sequence-based relationships between tRFs, all detected fragments were subjected to pairwise homology analysis using local Smith-Waterman alignment (*pairwiseAlignment* function of the R/Biostrings package), and scores were transformed into a distance matrix to enable clustering and visualisation of relationships. t-SNE was employed to visualise tRF homologies in a 2D space.

4.1.9 t-Distributed Stochastic Neighbour Embedding

SNE (Stochastic Neighbour Embedding) replaces Euclidian distances between data points with conditional probabilities that represent similarities. The Gaussian distribution used in SNE to represent the probability density for any given data point in the low-dimensional space is replaced by a Student's t-Distribution in the updated t-SNE algorithm. In combination with the use of a symmetrised function with simpler gradients, this alleviates problems with optimisation of the cost function that is used to create forces between points on the low-dimensional map. [180] The superiority of t-SNE with random initialisation remains subject of debate, and some advocate the use of the newer UMAP algorithm [181], although most of the discussion is centered around analysis of single-cell RNA-seq and preservation of global structure in the lower-dimensional visualisation. [182]

t-SNE was used in a variety of applications to reduce the dimensionality of high-dimensional data, for instance, the amino acid origin of tRFs, or the association of tRFs with distinct cell types in the blood. t-SNE analyses were performed in R, using the Rtsne package. [183] t-SNE requires, apart from the input data, a parameter called *perplexity*, which determines the weighting of local as opposed to global effects in the data. So far, there are no strict rules governing the selection of a perplexity value, other than that the perplexity cannot exceed the number of individual data points. Since different perplexities can give widely varying results, which can sometimes be misleading, the resulting maps have to be screened with a range of perplexities to assess their robustness.

4.1.10 Cholinergic Association of Small RNA Species

To determine association of distinct smRNAs with cholinergic transcripts, we analysed the multiple-targeting relationships of each distinct smRNA towards our curated list of cholinergic-associated (CA) transcripts. We first

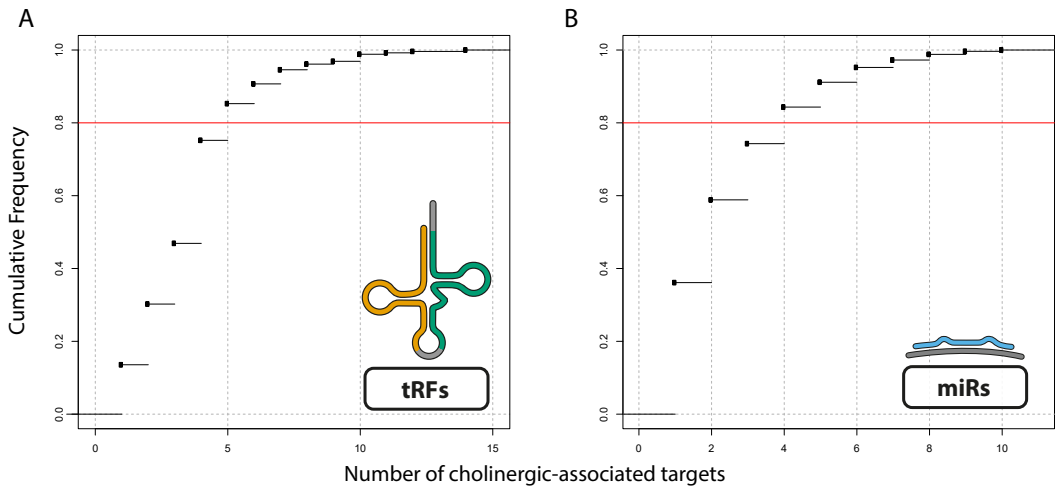

Figure 4.1: Cholinergic-associated Small RNA ECDF Curves. Cholinergic association was tested using *miRNeo* targeting data of miRNAs and tRFs. To assess the best-suited threshold for defining cholinergic association, empirical cumulative density functions were calculated for the number of cholinergic-associated (CA) genes targeted by each unique smRNA. **A)** Cumulative frequency of number of CA genes targeted by tRFs. Threshold of 80% (red line) is passed at five CA genes targeted. **B)** Cumulative frequency of number of CA genes targeted by miRNAs. Threshold of 80% (red line) is passed at four CA genes targeted.

created complete targeting data of all DE smRNAs towards all CA transcripts, which we then successively filtered for multiple targeting behaviours. To assess the base level of multiple targeting of cholinergic transcripts, we utilised empirical cumulative density functions of the number of individual cholinergic targets of each miRNA and tRF (Figure 4.1). We assumed 80% to be a robust threshold of cholinergic targeting, and for diverging numbers between miRNAs and tRFs chose to use the higher (more stringent) threshold. smRNAs above this threshold (i.e., smRNAs targeting at least as many cholinergic transcripts as the threshold value) were considered CA.

4.2. Descriptive Analysis of RNA Dynamics in Blood After Stroke

4.2.1 Differential Expression of Large RNA

At an alpha level of 0.05, we detected 694 differentially expressed (DE) long transcripts, 204 of them up- and 490 down-regulated (Figure 4.3 A). 18 of the up-regulated and 109 of the down-regulated transcripts exceeded the common $\log_2$ fold change (LFC) threshold of 1.4. To determine the most-impacted pathways, we performed GO analyses.

4.2.2 Gene Ontology Analyses of Differentially Expressed Genes

Ranking of all transcripts (regardless of direction of regulation) by their differential expression p-value resulted in GO terms mainly related to circulatory system processes (p = 0.018) and immunity (Figure 4.2 A). Most notable immune-related terms included cytokine-mediated pathways (p = 2.4E−04), response to IFNs α (p = 0.013) and β (p = 1.2E−03), regulation of JAK/STAT cascade (p = 0.013), response to LPS (p = 0.025), and macrophage activation (p = 0.026). Limiting the test set to transcripts with LFC above 1.4 increased sensitivity towards immune processes, yielding lower p-values for the enrichment of of positive (1.7E−04) and negative regulation of cytokine production (5.7E−04), type I interferon production (3.9E−04), response to bacterium (5.9E−04), innate immune response (2.0E−03), response to organophosphorous (2.3E−03), cytoplasmatic pattern recognition receptor signalling pathway (2.8E−03), and response to LPS (9.1E−03).

Up-regulated transcripts pertained to circulatory system processes, such as platelet degranulation (1.2E−03) and aggregation (0.02), and sprouting angiogenesis (4.8E−03), but also antigen processing and presentation (4.5E−03). Test set limitation to LFC above 1.4 did not increase sensitivity towards those terms, but presented essentially similar results. Up-regulated genes as such may be indicative of the bodily response to blood flow disruption and ischaemia caused by the stroke.

Down-regulated transcripts were enriched in terms involving response to IFN α (1.3E−03) and β (3.1E−04), response to LPS (1.5E−03), rhythmic process (2.5E−03), positive regulation of T cell proliferation (4.3E−03), positive regulation of JAK-STAT cascade (0.015), and cellular response to IL-1 (0.019). Test set limitation to LFC above 1.4 again increased sensitivity towards immune-related terms, but without changing the general pattern. Thus, down-regulated genes in all likelihood represent the post-stroke immunodepression, which can exacerbate into CIDS (see Section 1.2.5). The terms involving INF, IL-1, LPS, and JAK-STAT also indicate an important role for cytokine signalling in these processes.

As a cross-check, DE transcripts were ranked by count-change, and re-analysed (Figure 4.2 B). The top 100 changed transcripts, without regard to direction (absolute count-change) yielded terms implying response to IFN α (3.8E−04), β (1.1E−04), and γ (1.4E−04), mitochondrial organisa-

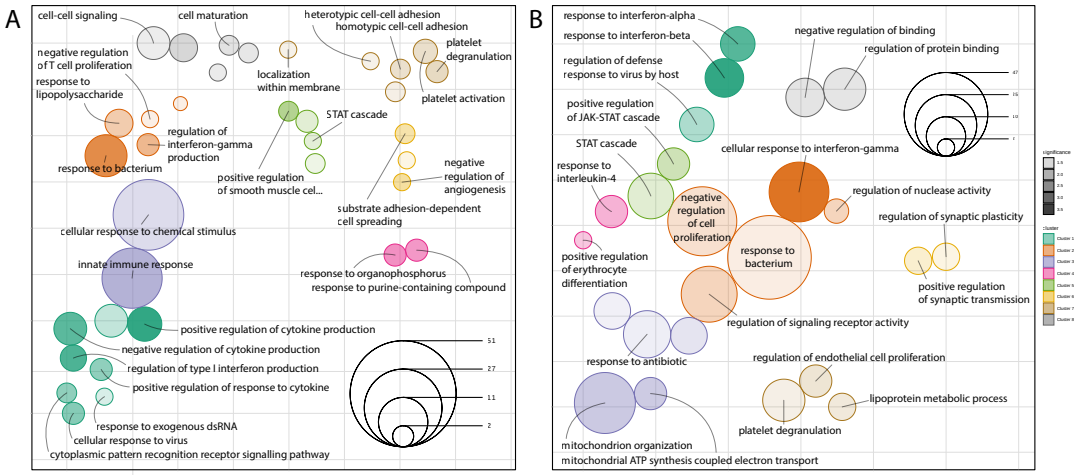

Figure 4.2: Large RNA Differential Expression Gene Ontology Enrichment. GO terms from enrichment analysis were projected on a 2D map via t-distributed stochastic neighbour embedding (t-SNE) analysis of their shared significant genes. LEGEND **A)** t-SNE visualisation of GO terms of DE genes with LFC > 1.4 shows 36 terms in eight clusters. Immunological terms (left-hand side) and circulatory terms (right-hand side) are predominant. **B)** t-SNE visualisation of 24 GO terms of top 100 genes measured by count-change corroborate immunological changes in stroke patient blood differential gene expression.

tion (5.6E−03) and ATP synthesis (6.9E−03), response to IL-4 (6.9E−03), positive regulation of JAK-STAT cascade (8.8E−03), response to antibiotic (0.044), and platelet degranulation (0.045). The top 100 up-regulated transcripts yielded terms involving platelet degranulation (3.8E−03), mitochondrial ATP synthesis (4.2E−03), response to xenobiotic stimulus (0.017), platelet aggregation (0.013), and response to antibiotic (0.016), while the top 100 down-regulated transcripts were associated with inflammatory response (1.3E−04), regulation of apoptosis (1.8E−04), cytokine secretion (6.8E−04), antigen processing and presentation (1.2E−03), regulation of lymphocyte apoptosis (2.3E−03) and proliferation (2.7E−03), response to antibiotic (5.2E−03), leukocyte homeostasis (7.6E−03), response to IL-1 (7.6E−03), and many more immune-specific processes. This corroborates the previous findings that up-regulated transcripts represent the response to circulatory system damage, and down-regulated transcripts indicate a cytokine-mediated immunodepression. For a full list of all terms from these analyses, see Appendix D.

DE genes?

4.2.3 Differential Expression of small RNA

In the simultaneous co-analysis of miRNAs and tRFs, we detected 420 DE miRNAs and 143 DE tRFs (adjusted p-value < 0.05, Figure 4.3 B&C). 63% of miRNAs (265) were down-regulated, while 87% of tRFs (124) were up-regulated. tRFs were mainly derived from the 3' end (3'-tRFs, 87) or from internal tRNA regions (i-tRFs, 48), while the tRFs from 5' ends (5'-tRFs) were in the minority (6). The amino acid distribution was shifted in favour of alanine- (35), glycine- (28), and proline-carrying (12) tRNAs (Figure 4.3 D). 30 of the 35 alanine-associated tRFs were 3'-tRFs, and all of those were up-regulated, indicating non-random generation of these fragments.

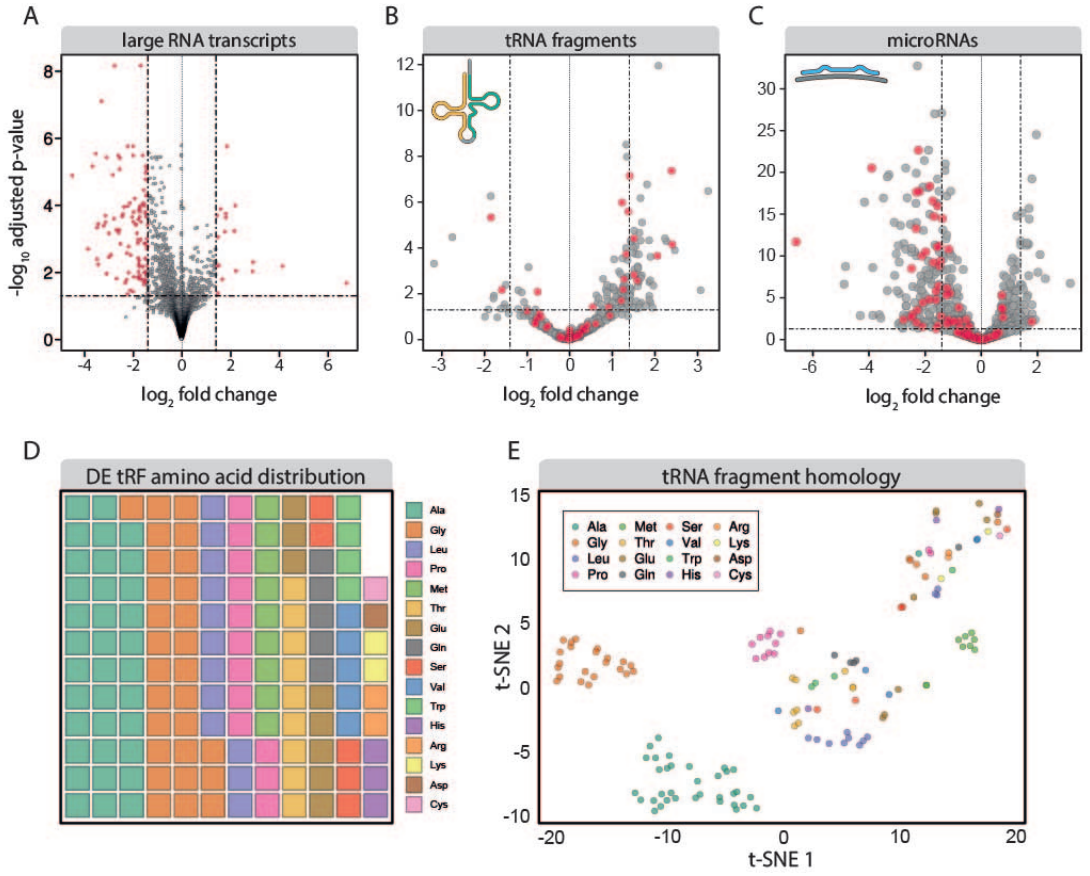

Figure 4.3: Small and Large RNA Differential Expression and tRF Properties. **A)** Differential expression analysis reveals multiple large RNA transcripts changed in patient blood after stroke. The majority of differentially expressed (DE) transcripts above a LFC threshold of 1.4 (red) are down-regulated. **B)** Blood-borne tRNA fragments (tRFs) also change after stroke. Unlike large RNA and miRNAs, the majority of DE tRFs are up-regulated. Cholinergic-associated (CA) tRFs in red. **C)** Blood-borne miRNAs are heavily influenced by the events following stroke. Like large RNA transcripts, miRNAs are also overwhelmingly down-regulated. CA miRNAs in red. **D)** The distribution of amino acid origin among the DE tRFs is non-random and biased towards the amino acids alanine, glycine, leucine, proline, and methionine. Each square represents one DE tRF, colour denotes amino acid origin. **E)** t-distributed stochastic neighbour embedding (t-SNE) of pairwise fragment homology by local Smith-Waterman alignment shows clustering of the dominant amino acid groups of tRFs. Clear clusters can be observed for tRFs derived from tRNA carrying alanine, glycine, leucine, proline, and methionine.

4.2.4 Homology Among tRNA Fragments

Using pairwise homology among all DE tRFs, visualised via t-SNE (see Section 4.1.9), we identified clusters of highly similar fragments, that correlate with their amino-acid origin, i.e., the amino acid which is carried by the respective parent tRNA (Figure 4.3 E). This relationship persisted across distinct individual tRNAs coding for the same amino acid, and was particularly pronounced in tRNAs associated with alanine, glycine, leucine, proline, and methionine. This further indication of non-random generation of tRNA fragments shorter than tiRNAs leaves an open question about their biogenesis, particularly, which nucleases are responsible for the generation of the mature transcripts, and if 3'-tRF generation is dependent or independent from tiRNA generation by angiogenin.

4.2.5 Cholinergic Association of Small RNA Species

The association of smRNA species to distinct systems or pathways is not trivial because of the multiple-targeting nature of these RNAs. For the purpose of the following analyses, we defined a small RNA as being associated with cholinergic processes by the positive association of the smRNA with a number of cholinergic-associated (CA) large transcripts. We do not assess the question whether this small RNA also targets other systems equally, or even if it targets cholinergic transcripts with greater likelihood than a random selection of genes. For this reason, we select a fairly high threshold for the definition of a CA smRNA, which is the targeting of at least 5 CA transcripts (above 80% on the empirical cumulative density function of cholinergic targeting, see Section 4.1.10).

Following this definition, we detected 52 CA miRNAs (90% down-regulated, 5 up and 47 down), and 18 CA tRFs (83% up-regulated, 15 up and 3 down). Above an LFC threshold of 1.4, we found 33 CA miRNAs (97% down-regulated, 1 up and 32 down), and 9 CA tRFs (78% up-regulated, 7 up and 2 down). CA smRNAs are marked in red in Figure 4.3 B&C.

amino acids?

4.3. Blood Compartments of Cholinergic Systems and Small RNA Species

To address the shortcomings of whole-blood RNA sequencing, which is more representative of the clinical setting, but less specific regarding cellular compartments, we consulted third party datasets to assess RNA species distribution in different cellular and non-cellular compartments of the blood. For small RNA species, we re-analysed a published dataset of small RNA-seq of 450 human samples from various blood tissues; [184] for large RNA transcripts, we utilised the tissue specificity of Marbach's regulatory circuits. [103] The large RNA information was used to identify blood cell types with cholinergic transcriptional activity, which was then used to zoom in into small RNA expression subsets related to cholinergic processes.

4.3.1 Large RNA Regulatory Circuits in Tissues of the Blood

To evaluate the cell type distribution of cholinergic genes in blood tissue types, we utilised the expression patterns derived from cumulative transcription factor activity of Marbach's regulatory circuits. [103] As shown by the authors, the cumulative activity of all transcription factors towards one gene describe well the actual expression of that gene in the respective tissue type. To maximise comparability to the parallel analyses of small RNA species (Section 4.3.2), blood cell types (i.e., »regulatory circuits«) were selected to reflect the cell type selection of Juzenas *et al.* [184] based on similar markers of the »cluster of differentiation« family of genes. These were: $CD4^+$ T-helper cells, $CD8^+$ cytotoxic T-cells, $CD14^+$ monocytes, $CD15^+$ neutrophils, $CD19^+$ B-cells, $CD56^+$ natural killer cells, and, for comparison, whole blood. For the sake of simplicity, genes were considered »present« in each blood tissue type if at least one TF showed significant activity towards the gene.

TF activities were collected for all of the tissues and aggregated across all TFs per gene by summing. The resulting table of 15 032 genes in the seven tissues was used as input for the *Rtsne* function. [183] t-SNE was computed using a range of perplexities and visualised as 2D map with a perplexity of 49 using R/ggplot2. [179]

4.3.2 An Atlas of Small RNA Expression in Cell Types of the Blood

To evaluate the cell type distribution of our small RNA molecules, we analysed a dataset deposited by Juzenas *et al.*, [184] who separated and sequenced 450 samples comprising seven types of individual blood cell types (characterised by »cluster of differentiation«-type membrane-bound receptors), serum, exosomes, and whole blood. The individual blood cell types comprised $CD4^+$ T-helper cells, $CD8^+$ cytotoxic T-cells, $CD14^+$ monocytes, $CD15^+$ neutrophils, $CD19^+$ B-cells, $CD56^+$ natural killer cells, and $CD235a^+$ erythrocytes (the only distinct cell type not available in Marbach's regulatory circuits, since mature erythrocytes do not transcribe). Starting from the raw data deposited on NCBI GEO, we controlled the quality, applied quality-based filtering, and aligned the 450 samples to miRNA and tRF sequences, as described in Section 4.1.4. The original publication did not offer statistical analyses because of a failure in the spike-in procedure, and defined presence of a small RNA by a measure of at least five counts in 85% of samples. However, since this definition relies heavily on sequencing depth, and depth can vary widely even in methodically robust sequencing experiments depending on a large number of variables (see Figure 3.5 C), we defined our own test for descriptive analysis of presence or absence of lowly expressed small RNAs in each of the sample types (Section 4.3.3).

4.3.3 Definition of Presence and Absence of Lowly Expressed smRNA Molecules

This definition comprises estimation of a log-normal distribution from a small RNA expression profile, and a statistical test to refute the null hypothesis that the distribution is in fact log-normal. The danger of evaluating true expression of lowly expressed smRNA molecules by a count-based threshold is the possibility of random reads resulting from degradation products of highly expressed RNA with similar sequence, and the amplification of noise. Both problems are exacerbated by an increase in sequencing depth. In today's RNA-seq technology, most chips can accommodate only a limited amount of samples compared to the amount of reads that can be generated. While this is not as problematic in cases of longer inserts and paired design, which is usually employed in large RNA-seq, in small RNA-seq this can lead to enormous overheads of reads. It is not uncommon to receive tens of millions of reads for each sample, which exceeds the recommended amount (of at least one million) by large margins.

Thus, there is the need to distinguish between degradation products of highly expressed RNA molecules or amplified noise and legitimate lowly expressed smRNA molecules (even more so since one of the smRNA species is a product of non-random tRNA degradation). The central assumption for our proposed method is: The expression pattern of legitimate smRNA molecules follows, as is common in biology, a normal distribution of some kind, or, for the discrete case, a normal poisson distribution. On the other hand, degradation products or noise would rather follow other, »non-biological« distributions, such as a uniform distribution or a monotonously decreasing power-law distribution such as the Pareto distribution. Thus, we chose to statistically test each smRNA in each tissue type for the adherence to this criterion, by comparing the measured counts with a distribution function estimated based on the mean and standard deviation of the measured counts. During testing, we found the log-normal distribution to give the best classification results.

The distribution mean and standard deviation of the expression values per cell type and smRNA were estimated using the *fitdist* function of the R/fitdistrplus package. [185] The count distribution was then tested against a log-normal distribution with the estimated mean and standard deviation via the R implementation of the Kolmogorov-Smirnov test, with a cutoff of 0.1. The small RNA was defined as present if the test failed to reject the null hypothesis (see Appendix E for numerous examples).

Analysis of Expression Patterns and Establishment of Virtual Tissues

The distribution of smRNA expression across the different cell types was used to assign eight functional compartments (i.e., »virtual tissues«) to the entirety of detected fragments such that each smRNA was sorted into one of the tissue classes. Ideally, these classes would be unambiguous, i.e., there would be no overlap of smRNA molecules between the classes. Eight classes were created via hierarchical clustering of miRNA and tRF expression separately (Figure 4.4), and then used in combination with t-SNE applied to the entire expression matrix, to visualise the compartmentalisation of smRNAs in these virtual tissues. The samples taken from stroke patients in the PREDICT study were sequenced from whole blood, which precludes direct information about tissue distribution. Thus, the two-dimensional maps from t-SNE visualisation were used to, first, explore the tissue association of smRNAs differentially expressed in whole blood samples of stroke patients, and second, examine the potential impact of cholinergic-associated smRNAs in these tissues.

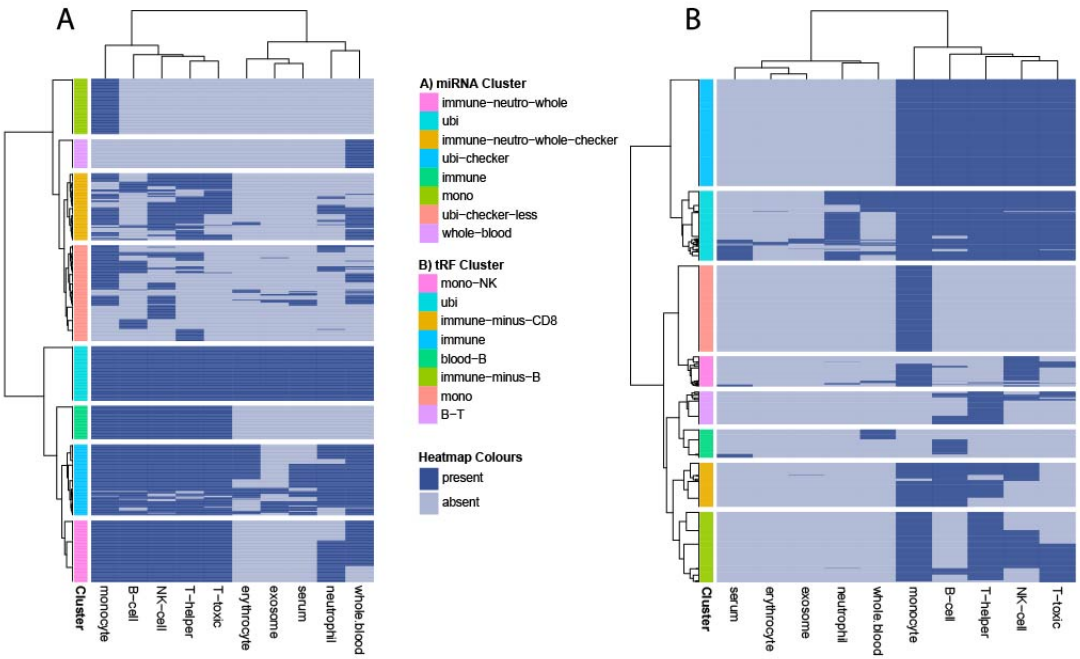

Figure 4.4: Functional Characterisation of Hierarchical Clusters in Blood Cell Small RNA Expression. Information on presence/absence of miRNAs and tRFs in the tissue types analysed in Juzenas *et al.* [184] were hierarchically clustered into 8 clusters using the Ward method, [186] and plotted on a heatmap (single smRNAs on the y-axis, tissue types on the x-axis). To assign meaning to these clusters, manual inspection was followed by annotation of enrichment in tissue types. Complex combinations were approximated by their most prominent features. **A)** Clusters of miRNA presence/absence in blood cell compartments. Clearest cluster association was shown by miRNAs expressed only in monocytes (»mono«), in all blood-borne immune cells except neutrophils (»immune«), ubiquitously without exception (»ubi«), and only in whole blood (i.e., in none of the single compartments (»whole-blood«). **B)** Clusters of tRF presence/absence in blood cell compartments. Clearest cluster association was shown by tRFs expressed only in monocytes (»mono«), and in all blood-borne immune cells except neutrophils (»immune«). The other tissue-related clusters were not as clear as in the miRNA expression data, indicating a looser association to cell type of tRNA-derived smRNAs.

4.3.4 Large RNA Expression Patterns Identify Cholinergic Systems in CD14$^+$ Monocytes

The expression patterns of 15 032 large RNA molecules in blood-borne immune cells were visualised in a t-SNE-derived 2D map (Figure 4.5 A). More than half of all transcripts show highest expression in whole blood (7533, not shown), so subsequent analyses were performed on the set of six tissues, without the whole blood compartment. In this set (14 280 genes), most transcripts show highest expression in CD14$^+$ monocytes (9125 transcripts), followed by CD19$^+$ B-cells (1176) and CD15$^+$ neutrophils (1166). Remaining are CD4$^+$ T-helper cells (1092), CD56$^+$ NK-cells (948), and CD8$^+$ cytotoxic T-cells (773). When filtered for cholinergic genes, there is visible enrichment of core cholinergic transcripts in a spatial sub-compartment of CD14$^+$ monocytes (Figure 4.5 B). Considering the different monocyte phenotypes (pro- and anti-inflammatory, see Section 1.2.5), and their implied transcriptomic differences, which most likely are brought on by divergent TF activity, this compartmentalisation of cholinergic transcripts inside one spatial sub-compartment may indicate a cholinergic »preference« in favour of one particular monocyte phenotype.

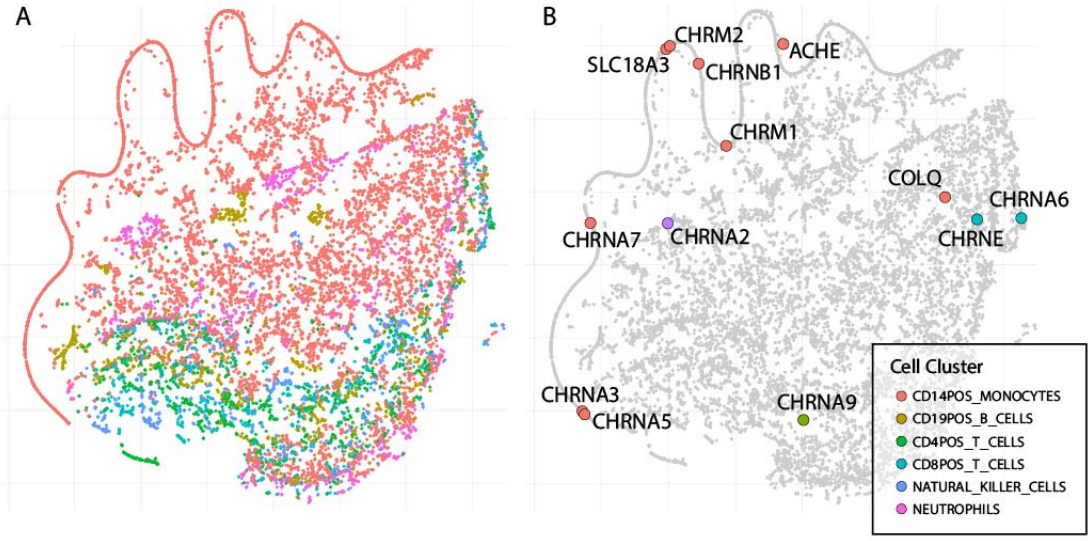

Figure 4.5: Large RNA Expression Patterns in Blood-Borne Cells. Expression derived from transcriptional activity in blood-borne cell types in the Marbach dataset [103] was visualised via t-SNE. The input matrix comprised all 14 280 detected genes in 6 types of blood-borne immune cells. Genes were plotted on the first two t-SNE dimensions and coloured by the cell type of their highest expression, i.e., the highest cumulative transcriptional activity of all active TFs. **A)** Complete t-SNE shows a gradient of expression across the different cell types, with much expression in CD14$^+$ monocytes and T-cells. **B)** Highlighting of cholinergic core genes reveals an enrichment in close compartments of CD14$^+$ monocytes. The vesicular ACh-transporter SLC18A3 can serve as substitute for the main cholinergic marker, CHAT, as discussed in Section 2.2.3.

4.3.5 Identification of Functional Enrichment of smRNA Expression in Blood-Borne Cells

To date, there is no comprehensive expression catalogue of smRNA species expression in the tissue types of the human body that is comparable to what has been achieved in the description of large

RNA. To classify the detected smRNAs in a manner specific to tissues in human blood, we utilised a dataset published by Juzenas *et al.*,[184] who describe miRNA expression in a variety of blood tissues. We re-analysed the publicly deposited data for miRNA and tRF expression, and developed our own method of defining »presence« of the smRNA in each tissue type based on the evaluation of a log-normal distribution model (instead of using a simple count threshold, see Section 4.3.2 for details).

Using these presence/absence data, we first utilised hierarchical clustering to establish »virtual tissues« that could be assigned to each smRNA (Figure 4.4) for later evaluation in the stroke patient sequencing. Both miRNAs as well as tRFs showed a number of smRNAs clearly associated with several compartments, whereas other compartments and smRNAs were distributed in a more complex manner. The ten tissue types of the Juzenas *et al.*[184] study were equally parted into two five-tissue superclusters by the expression patterns of both smRNA species (Figures 4.4 A&B, x-axis). These two clusters distinguish immune from non-immune compartments in the blood, but for one notable exception: while the »immune supercluster« comprises monocytes, T-cells, B-cells, and NK-cells, the »non-immune supercluster« contains neutrophils in addition to erythrocytes and the non-cellular tissues serum, exosomes, and whole blood. Notably, the neutrophil samples cluster closest to the whole blood compartment in both smRNA species.

Two distinct virtual tissues showed high consistency in both smRNA species: a virtual tissue containing only $CD14^+$ monocytes and another tissue comprising all studied cellular immune components except neutrophils (i.e., monocytes, B-cells, both types of T-cells, and NK-cells). miRNAs (Figure 4.4 A), in addition, yield clear clusters for miRNAs expressed in whole blood, and for miRNAs expressed ubiquitously without exception. In tRFs (Figure 4.4 B), the general picture is more complex, as the clusters are often mixed.

4.3.6 Expression Patterns of Differentially Expressed and Cholinergic-Associated smRNAs

Similarly to the visualisation of large RNA molecules, the expression patterns of 600 miRNAs and 1671 tRFs in ten tissues of the blood were visualised in t-SNE-derived 2D maps (Figure 4.6 A&B). In the initial visualisation, the forming of multiple clusters according to some virtual tissues can be observed, while other virtual tissues are visibly more dispersed. Clearest clusters are formed in both cases by ubiquitously expressed smRNAs (»ubi«), smRNAs expressed only in monocytes (»mono«), and smRNAs equally expressed in all immune-related blood-borne cells except for neutrophils (»immune«). Examination of DE smRNAs on this 2D map shows a further parallel between miRNAs and tRFs (Figure 4.6 C&D): differential expression after stroke takes place in all compartments of the blood, and highest changes in transcript amount (as measured by count-change) are observed in ubiquitously expressed smRNAs. Similarly, cholinergic-associated (CA) miRNAs and tRFs (Figure 4.6 E&F) are observed in all compartments, but the most highly differentially regulated CA smRNAs are expressed in all blood compartments alike. Notably, whole blood does not play a role in DE

miRNAs, which may indicate lower relative importance of non-cellular blood compartments in terms of classification. In other words, most smRNAs that are found in non-cellular compartments are found in the cellular compartments as well, making them irrelevant for classification (however, their biological function in these non-cellular compartments remains a matter of interest).

On the other hand, smRNAs that are ubiquitously expressed are also detected in differential expression with high frequency and perturbation (see Figure 4.6 C&D). In addition to a putatively significant biological function of these smRNAs, this may also indicate a covariation of broadness of expression with detection in whole blood differential expression. Presenting an important limitation, this possibility cannot be assessed in the present data, because it requires a stratification of blood tissues prior to sequencing, for instance via fluorescence-assisted or magnetic-activated cell sorting (FACS/MACS). This issue is further discussed in Section 5.1.5.

Figure 4.6: Small RNA Expression Patterns in Blood-Borne Cells. Two-dimensional expression maps were created using t-SNE on the full numeric expression data derived from re-analysis of the Juzenas *et al.* [184] data set for miRNAs and tRFs separately. Single smRNAs (points) were coloured by the virtual tissues derived from the cluster heatmap analysis (Figure 4.4). Node size reflects absolute count change in **C**, **D**, **E**, and **F**. Shown are full data, differentially expressed (DE) smRNAs, and cholinergic-associated (CA) smRNAs for each species. **A)** Full t-SNE visualisation of miRNA expression. The largest 2D-associative clusters are comprised of the clearest presence/absence virtual tissues, monocytes (yellow) and ubiquitously expressed (orange). Smaller clusters can be identified for the tissue of all immune cells except neutrophils (green) and the complex cluster of immune cells including neutrophils and whole blood (turquoise). **B)** Full t-SNE visualisation of tRF expression. The largest 2D-associative clusters are, as in miRNAs, comprised of the clearest presence/absence virtual tissues, monocytes (brown) and immune cells except neutrophils (pink). A smaller cluster can be identified for ubiquitously expressed tRFs (orange). **C)** miRNAs DE after stroke are ubiquitously expressed in all virtual tissues. Highest differential expression is seen in the »ubi« cluster. **D)** Likewise, tRFs DE after stroke are ubiquitously expressed in all virtual tissues, and highest differential expression is seen in the »ubi« cluster. **E)** CA miRNAs are enriched in the lower quadrants of the 2D map, particularly in the clusters associated with ubiquitous expression (»ubi«, »ubi-checker«). **F)** CA tRFs show a similar distribution, skewed towards virtual tissues with ubiquitous expression. This may indicate covariation of detection with broadness of expression (see text and Section 5.1.5).

4.4. Regulatory Circuits of Small RNA and Transcription Factors in CD14$^+$ Monocytes

The clear separation of CD14-biased smRNAs (Figure 4.4) and the cholinergic importance of CD14$^+$ monocytes as shown by large RNA t-SNE (Figure 4.5) merit a detailed analysis of these cells in terms of their transcriptomic interactions. We thus created the whole-transcriptome network of transcription factors, miRNAs, tRFs, and target genes in CD14$^+$ monocytes using *miRNeo*, and analysed it in different ways.

4.4.1 Comprehensive Circuit Network Creation

The comprehensive transcriptomic network in CD14$^+$ monocytes was created in a two-step process of *miRNeo* targeting. First, the complete TF→gene network was created from the targeting data derived from Marbach *et al.* [103], yielding a CD14-specific network comprising 616 TFs with activity towards 13 447 transcripts, in 318 731 unique interactions. Second, this network was then subjected to successive *miRNeo* targeting of all transcripts in the network by miRNAs and tRFs.

For each node fulfilling an active role in this network (i.e., miRNAs, tRFs, and TFs), an activity parameter was computed. The activity of each node is hereby defined as the sum of all scores of each of its targeting relationships. In the case of miRNAs, the score is the summary score introduced in Section 2.2.4; for tRFs, it is the score calculated with the BL-PCT method (see Section 2.2.6); and for TFs, it is the transcriptional activity given by Marbach and colleagues. [103] Activities were normalised, for each biotype separately, by scaling the calculated values v onto a range between 0 and 1, using

$$v_{i,norm} = \frac{v_i}{max(v)}$$

with $max(v)$ being the maximum of all scores in this biotype category, and all $v > 0$. The activity of each relationship determined the weight of the edge between the two connected nodes.

The network was visualised in gephi, [147] omitting all non-TF genes, and using ForceAtlas2 to generate a force-directed 2D map of smRNA→TF interactions in CD14$^+$ monocytes. Network modularity was calculated using the function included in gephi, [190] with a resolution of 2.0, to yield two distinct modularity classes of predominant regulation by either miRNAs or tRFs. The associations of TFs to the tRF- and miRNA-regulated modules were used to perform subsequent analyses of the distinct modules.

4.4.2 Gene Ontology Analyses of TF→Gene Networks of CD14$^+$ Monocytes

The TF→gene networks of each of the two modules derived from smRNA species association (miRNAs versus tRFs) were analysed using topGO [145] essentially as described in Section 3.5.3. Genes were ordered according to the cumulative activity of TF targeting of each gene in CD14$^+$ monocytes. To display a range of top genes, transcript background was iterated in five equal steps from 1000 transcripts to the maximum size of target transcripts in each network (12 927 for miRNA-targeted TFs, 12 904 for tRF-targeted TFs). The test set was the top 10% of transcripts for each background size. GO terms were collected and screened for multiple entries among the sets. The most prevalent terms were used to infer the functional roles of miRNA- and tRF-targeted transcription factors. We determined the overlap of GO terms between both smRNA species as well as the terms exclusive to either.

4.4.3 Dichotomy of Small RNA Targeting of Transcription Factors in CD14$^+$ Monocytes

Organisation of the smRNA→TF network via a force-directed algorithm resulted in visible clustering of two distinct subnetworks, that are governed by miRNAs and tRFs, respectively (Figure 4.7). Inside this network, 10 TFs were found DE in patient blood after stroke (Figure 4.7 A). Calculation of modularity clearly divided the network into TFs primarily influenced by miRNAs and TFs primarily influenced by tRFs (Figure 4.7 B). Based on these two sets of TFs, two distinct TF→gene networks were created: 289 miRNA-biased TFs with 152 649 unique TF→gene targeting relationships, and 280 tRF-biased TFs with 163 641 unique TF→gene targeting relationships.

4.4.4 Gradual Shift in Control Over Transcription Factors by miRNAs and tRFs

356 TFs were detected in the stroke patient blood sequencing experiment. It is notable that, although the complete graph shows clear segregation between miRNA-targeted and tRF-targeted transcripts, merely 106 of those TFs are targeted by only one of the two smRNA species (48 only by miRNAs and 58 only by tRFs), and 55 are supposedly not at all targeted by any smRNA present in CD14$^+$ cells. The remaining 195 TFs are putative targets of both smRNA species (Figure 4.7 C).

At an alpha level of 0.1 for the differential expression between stroke patients and controls, 26 of these TFs remain, also showing a gradual pattern of targeting by miRNAs and tRFs (Figure 4.7 D). Six of these transcription factors are implicated in the control of cholinergic core or receptor genes (marked with a »C«). It is notable that a number of TFs show no indication of being a target of either smRNA species present in CD14$^+$ cells under the premises of our targeting approach (Figure 4.7 E). Considering the multiple-targeting behaviour of smRNAs, and the general experience that non-targeted genes are uncommon, this finding is interesting in itself, particularly since it involves well-described regulators of immunological processes, such as *STAT2* and *ELF1*.

The *STAT* family of transcription factors is an interesting example in this analysis. The cholinergic/neurokine interface is facilitated by JAK/STAT signalling, in which neurokine receptors can activate the pathway through STAT1, STAT3, and STAT5A/B phosphorylation.[1] There are three differentially expressed STATs in our data set: STAT1 is down-regulated (highest absolute count-change of all TFs), targeted preferentially by tRFs (tRF fraction = 82.1%), and directly associated with cholinergic genes in CD14$^+$ monocytes (»C«); STAT5B is up-regulated (second highest count-change of all TFs), preferentially targeted by miRNAs (tRF fraction = 37.5%), and not directly associated with cholinergic genes in CD14$^+$ cells; and STAT2 is also down-regulated (third highest absolute count-change of all TFs), does not directly associate with cholinergic genes, and additionally is not predicted to be targeted by any smRNA present in CD14$^+$ monocytes, although it is expressed and induced by interferons in these cells.[187]

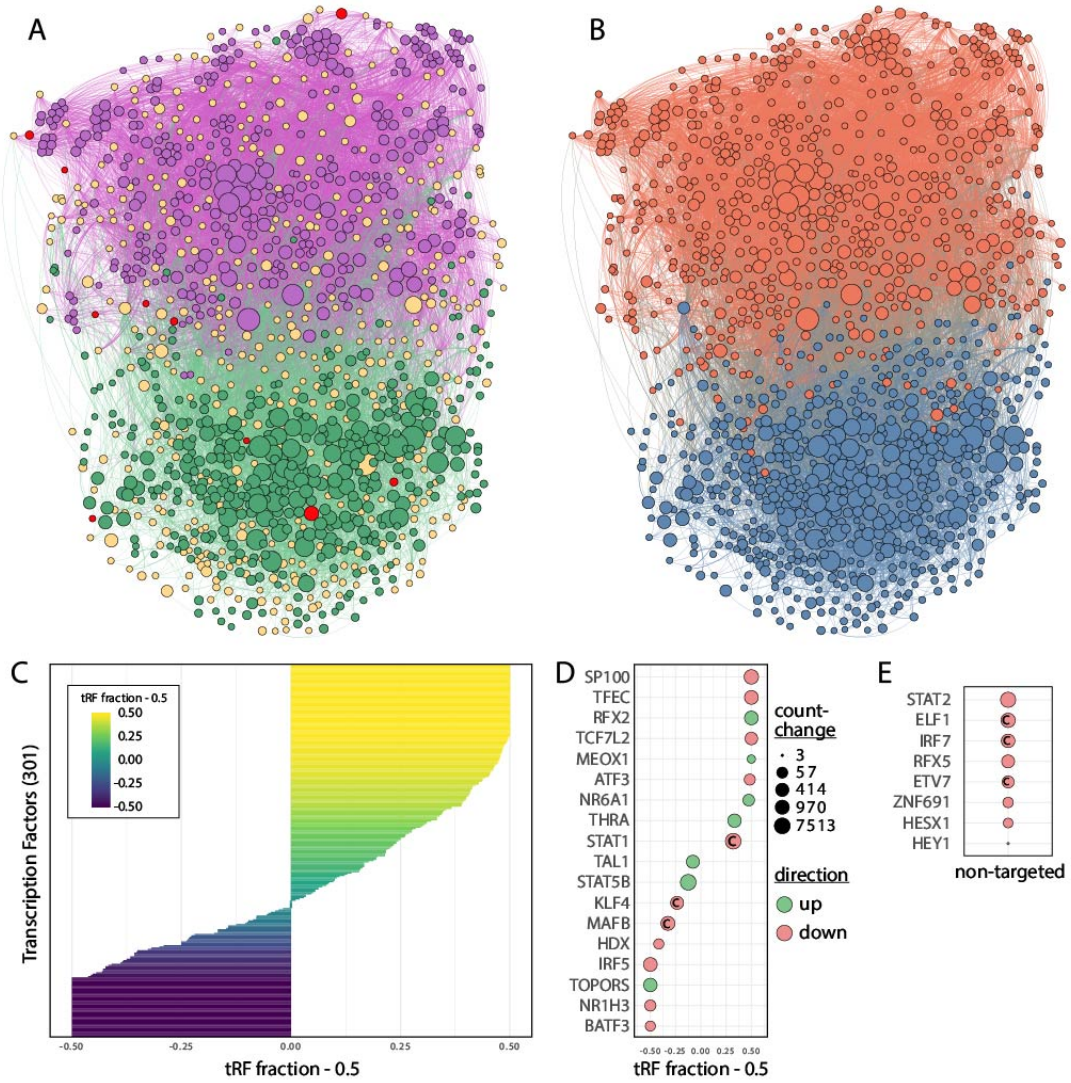

Figure 4.7: Small RNA Targeting of Transcription Factors in CD14$^+$ Monocytes. The two-dimensional map of all TFs active in CD14$^+$ monocytes and their smRNA controllers was created via *miRNeo* targeting and visualisation via force-directed algorithm. Node size is determined by activity (see Section 4.4.1). **A)** Nodes coloured by biotype: miRNAs - green, tRFs - purple, TFs - yellow, differentially expressed (DE) TFs - red. TFs targeted mainly by tRFs segregate visually from TFs targeted mainly by miRNAs. Both sets contain DE TFs, indicating complementary function. **B)** The network was divided into two modules by network connectivity measures. Node colour denotes modularity class association. Network modularity largely reflects TF targeting of tRFs (orange) versus miRNAs (blue). **C)** Bar graph displays the fraction of smRNAs of each species targeting each of the TFs (on the y-axis, 301 TFs). A tRF fraction of 1 means all smRNAs targeting the TF are tRFs, 0 means all are miRNAs. Displayed on x axis is »tRF fraction - 0.5« to center on 50:50 targeting by both species. 48 TFs are solely targeted by miRNAs (leftmost) and 58 solely by tRFs (rightmost). **D)** Similarly, TFs differentially expressed in the blood of stroke victims show a distribution on the miRNA-tRF gradient. Point size denotes count-change, point colour denotes direction of change; a »C« denotes the TF as targeting cholinergic core and receptor genes. **E)** Notably, there is also a set of TFs that are supposedly not targeted by any smRNA present in CD14$^+$ cells.

4.4.5 Dichotomous Transcriptomic Footprints of Transcription Factors in CD14$^+$ Monocytes

To determine the putative effect of TF regulation by each smRNA species, we evaluated the potential impact of the TFs most targeted by either miRNAs or tRFs (i.e., the two modules from Figure 4.7 B). The top 10% of TF targets in CD14$^+$ monocytes (derived from Marbach *et al.* [103]) were subjected to iterative GO analysis (see Section 4.4.2). Assuming a general effect of repression in tRF-targeted TFs (because the majority of DE tRFs are up-regulated), and a general de-repression in the set of miRNA-targeted TFs (because most DE miRNAs are down-regulated), the putative functional effects of changes in smRNA levels can be described by GO enrichment analysis of these two test sets. Although this is a very rough categorisation, it may help in classifying the areas of influence shared between the two smRNA species, or exclusive to either.

GO Term Overlap Between miRNA- and tRF-targeted Transcription Factors

If we assume the functions associated with TF→gene interaction (the »footprint«) in each subnetwork under miRNA or tRF control as an indication of the sphere of (most) influence of this smRNA species, the GO terms associated with both can give an indication of their overlapping functions. Further assuming the simplified scenario of dominating de-repression in miRNA-controlled transcripts, and dominating repression in tRF-controlled transcripts, this set of overlapping function is the set where a homeostasis is met by the cooperation of miRNAs and tRFs, or where, upon perturbations such as stroke, a shift in the balance between the two smRNA species can alter the physiological response to the stimulus.

We found 39 significant GO terms to overlap between multiple sets of miRNA- and tRF-targeted transcription factors, almost exclusively comprised of immunity-related terms. Seven terms were found with adjusted p-value < 0.001, namely: neutrophil chemotaxis (p = 1.3E-04), regulation of myeloid leukocyte differentiation (2.6E-04), positive regulation of cold-induced thermogenesis (2.8E-04), negative regulation of ERK1 and ERK2 cascade (3.0E-04), regulation of type 2 immune response (4.9E-04), regulation of antigen receptor-mediated signalling pathway (5.0E-04), and negative regulation of IFN-γ production (5.9E-04). Further terms included positive regulation of CD4$^+$, alpha-beta T cell activation (p = 0.0013), monocyte chemotaxis (0.0018), negative regulation of immune response (0.0021), response to hypoxia (0.0041), positive regulation of cytokine secretion (0.0049), and parasympathetic nervous system development (0.0051).

Functions Distinguishing Between miRNA- and tRF-targeted Transcription Factors

After removal of the overlapping GO terms between miRNA- and tRF-targeted transcription factors, the remaining miRNA- and tRF-associated sets were examined to assess their differences. In the following, only terms which had been found in at least two steps of the five-step iterative process are considered. Terms found in both sets that were not identical but very similar were also removed from the analysis.

Transcription factors from the module targeted preferentially by miRNAs (Figure 4.7 B, blue) are active towards genes that implicate the following biological processes; several terms showed adjusted p-values below 0.001: response to TNF ($p = 1.2E-04$), erythrocyte differentiation (2.6E-04), cellular response to cytokine stimulus (3.5E-04), positive regulation of cytokine production (3.8E-04), positive regulation of myeloid cell differentiation (3.9E-04), regulation of IL-12 production (4.9E-04), positive regulation of leukocyte chemotaxis (6.5E-04), regulation of cellular response to insulin (6.6E-04), negative regulation of T cell mediated immunity (8.1E-04). Other terms include: regulation of macrophage activation (0.0021), response to LPS (0.0045), negative regulation of haematopoiesis (0.006), regulation of production of interleukins 1, 6, 12, 13, and 17 (all $p < 0.005$). These processes may, simply, be seen as amplified, since the general down-regulation of miRNAs would lead to a de-repression of their targets. However, in many cases of »regulation«, no direction is implied.

The processes regulated exclusively by miRNA-regulated TFs in this scenario thus are mainly related to pro-inflammatory events, more specifically, innate immune response, mediated by interferons and pro-inflammatory interleukins. An amplification of pro-inflammatory innate responses is contrasted by a reduction in haematopoiesis and T cell-mediated reactions.

Transcription factors from the module targeted preferentially by tRFs (Figure 4.7 B, orange) are active towards genes that implicate the following processes; several terms showed adjusted p-values below 0.001: negative regulation of apoptotic process (1.3E-04), negative regulation of coagulation (1.9E-04), positive regulation of haematopoiesis (1.9E-04), regulation of IFN-γ production (2.5E-04), positive regulation of angiogenesis (2.7E-04), IL-4 production (4.1E-04), nuclear pore organisation (4.5E-04), monocyte differentiation (5.0E-04), leukocyte migration (5.2E-04), and T cell cytokine production (6.4E-04). Other terms include: negative regulation of NIK/NF-κB signalling (0.0016), macrophage differentiation (0.0030), lymphocyte activation involved in immune response (0.0032), negative regulation of leukocyte mediated immunity (0.0043), negative regulation of neuron death (0.0052), monocyte differentiation (0.0062), negative regulation of insulin receptor signalling pathway (0.013), natural killer cell activation (0.017), positive regulation of STAT cascade (0.019), sensory perception of pain (0.026), $CD8^+$, alpha-beta T cell activation (0.043), production of interleukins 2, 4, 6 (all $p < 0.005$). These processes, as opposed to the miRNA-associated processes, may be seen as attenuated, since a strong trend towards up-regulation is seen in tRFs; this again holds

true only for terms where a direction is implicit or explicitly described.

The processes regulated exclusively by tRF-targeted TFs refer to apoptosis, coagulation, angiogenesis, myeloid leukocyte regulation, and interleukin/STAT signalling. Processes amplified via the putative de-repression include apoptosis, coagulation, NIK/NF-κB signalling, leukocyte activity and differentiation, and insulin-mediated signalling. Negatively impacted processes include haematopoiesis, angiogenesis, STAT signalling, and IL-2 production.

Immediately comparing miRNA- and tRF-associated processes, there are multiple functional overlaps even in the set curated to show only exclusive terms for either smRNA species. Specifically, the perturbations in both species seem to up-regulate innate immune response, particularly via INFs, TNF, interleukins, and myeloid leukocytes. Additionally, the perturbations in both species seem to have an additive suppressive effect on haematopoiesis and T cell activation.

4.5. Feedforward Loops of Small and Large RNA

To delve deeper into the transcriptional cooperation between small RNAs, transcription factors, and the genes they target, feedforward loops including all three actors can be of analytical use. Briefly, a feedforward loop (FFL) describes a constellation of three entities (X, Y, and Z), in which one entity (X) has control over an intermediate (Y), and both control the outcome of the ultimate (Z).[188] Since data on TF control over smRNAs is scarce, only cases of X = smRNA, Y = TF, Z = gene can be realistically evaluated. FFLs can be further distinguished: a coherent FFL describes the case of regulation by X and Y towards Z in a similar direction (i.e., amplification, Figure 4.8 A), while in an incoherent FFL, X and Y influence Z in opposite directions (attenuation, Figure 4.8 B). While the latter is unintuitive at first sight, it can serve a multitude of meaningful functions in a cellular context, such as noise reduction, reduction of cross-contamination, or increase of temporal resolution.[189]

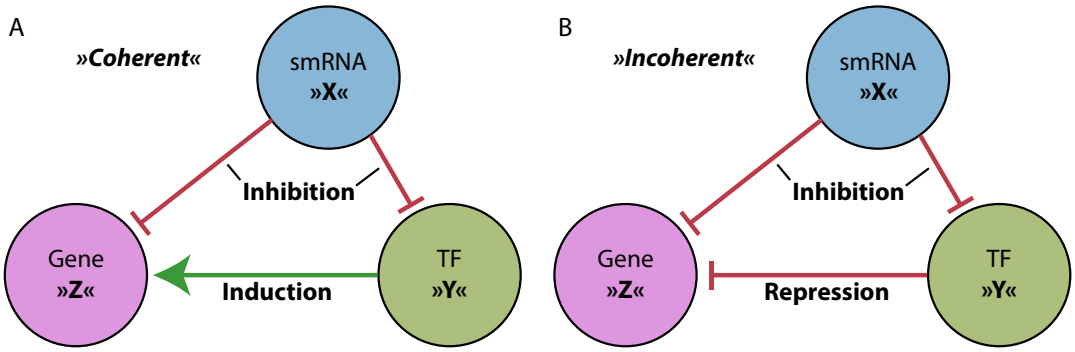

Figure 4.8: Small RNA Feedforward Loop Theory. The figure represents the two cases of smRNA FFLs most accessible to current analysis methods, i.e., the cases of smRNA (**X**) targeting of TF (**Y**) and ultimate gene (**Z**). **A)** Basic coherent smRNA FFL. The smRNA (**X**) inhibits both the TF as well as the ultimate gene target. Since the TF (**Y**) is an activator of target gene (**Z**) expression, smRNA regulation has the same direct and indirect effect on ultimate gene expression. **B)** Basic incoherent smRNA FFL. The smRNA (**X**) has a direct suppressive effect on the target gene (**Z**), but the simultaneous suppression of the TF (**Y**), which in turn represses the target gene (**Z**), leads to elevation of target gene expression, ameliorating or even inverting the direct effect.

In the following, the principle of feedforward loops will be applied to transcriptomic and ontological analyses of gene expression in blood-borne cells after stroke. Pathways of co-regulated transcripts and their TFs will be identified by modularisation of a comprehensive FFL network in $CD14^+$ monocytes. The additional information leveraged from FFL-implied pathways will be ordered and interpreted based on previous studies on the identified pivotal actors.

4.5.1 Feedforward Loop Creation

Starting from the set of differentially expressed TFs ($p < 0.05$) active in $CD14^+$ monocytes (as seen in Figure 4.7), single feedforward loops (FFLs) of smRNAs (miRNA or tRF), TFs, and genes were detected using *miRNeo* (as described in Query 2.5, Section 2.3). FFLs were created $CD14^+$ monocyte-specific by using only TF activity from these cells and by removal of any smRNAs not detected in $CD14^+$ cells in the Juzenas *et al.*[184] data set. Ad-

ditionally, because of the high amount of TF→gene relationships in $CD14^+$ cells, TF relationships were filtered for the 10% with highest activity.

4.5.2 Visualisation and Modularisation

The network of all smRNAs, TFs, and genes included in these FFLs was visualised in gephi [147] as a two-dimensional force-directed map, using the ForceAtlas2 algorithm. At initial network creation, tRFs were represented by the seeds included in their sequences, which were later associated with the mature tRFs. Using a community detection algorithm [190], the network was sub-classified into five module classes (using edge weights and a resolution of 1.5, Modularity coefficient = 0.482).

4.5.3 Module-specific Functions via GO Analysis

The module classification was reimported into R, and using R/topGO, [145] the functions of individual submodules were assessed by testing the significantly differentially expressed genes (adjusted $p < 0.05$) from each module against a background of 2000 randomly selected genes. Significant terms were manually screened, and differentially expressed genes were extracted from the test data for relevant terms.

4.5.4 Feedforward Loop Network of $CD14^+$ Monocytes

The complete FFL network of TFs DE in stroke patient blood ($p < 0.05$) was created using *miRNeo* and visualised in gephi. In total, 195 043 unique FFLs were discovered, 193 803 containing miRNAs, and only 1240 containing tRFs. After filtering of the top 10% of TF→gene relationships by activity, 19 309 miRNA- and 169 tRF-FFLs remained. These FFLs constitute a network of 2628 nodes and 22 456 edges (Figure 4.9). Community detection [190] resulted in a sub-classification of nodes into five distinct modules, which were subsequently analysed for their functions using GO.

The reasoning behind this approach is to explain in more detail the findings of GO enrichment analysis of the differentially expressed large transcripts (compare Section 4.2.2), and to more closely define the pathways implicated in the context of smRNA regulatory modules. In applying feedforward loops, there may be an increase in identification of biologically relevant pathways, as opposed to network creation by TF→gene and smRNA→gene relationships alone. Notably, there was no cross-talk of modules across significant GO terms; if a term was found significant in GO enrichment analysis of a module, all genes associated with the term were located in the same module. The following paragraphs will attempt to interpret the terms associated with each module to shed some light on the distribution of their functions and possible inter-module cooperations.

Particular attention was paid to GO terms indicating a direction (e.g., »positive regulation«), which in combination with the direction of differential expression of the implicated genes may give an indication of the tangible effect of module gene regulation after stroke. When writing of »module genes«, differential expression (adjusted p-value < 0.05) of these genes in stroke patient blood is always implied.

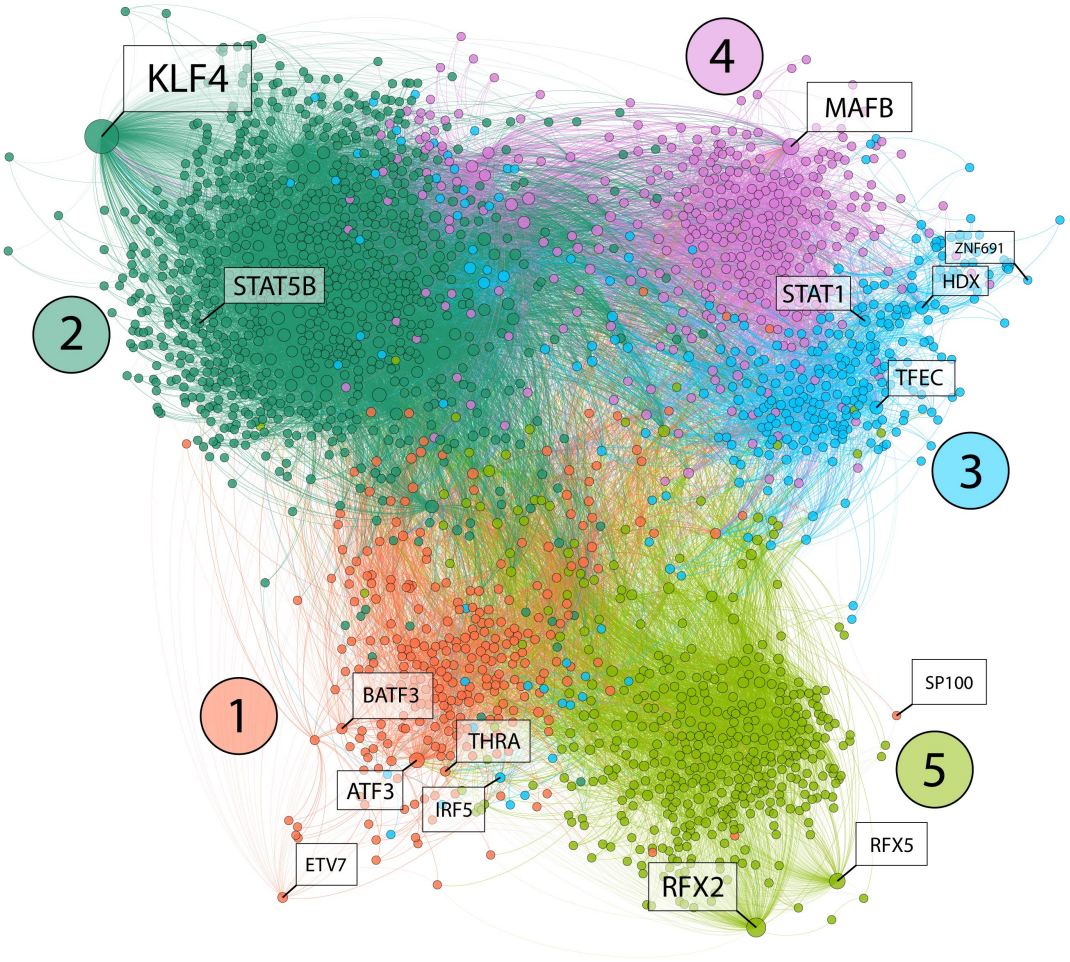

Figure 4.9: Complete Feedforward Loop Network of Differentially Expressed Transcription Factors in CD14$^+$ Monocytes. Feedforward loops (FFLs) were created by *miRNeo* query of transcriptional interaction of transcription factors (TFs), miRNAs, and tRFs towards all genes, but limited to FFLs of which the regulatory elements (TFs and smRNAs) were present and active in CD14$^+$ monocytes, and additionally filtered for the top 10% TF$\rightarrow$gene relationships by activity. The resulting network was assembled in a force-directed 2D-projection and analysed to yield modularity information. The algorithm identified 5 distinct modules of small and large RNA FFLs, indicated by colours and numbers in circles. TFs differentially expressed in stroke patient blood are marked with a name tag (node size denotes differential expression count-change of TFs).

Module One

GO enrichment analysis of significantly DE genes from module one resulted in 16 GO terms from 35 DE genes. The most significant biological process governed by module one is »negative regulation of transcription« with eight significant genes (SGs) and an adjusted p-value of 1.6E-04. The second- and third-most significant terms indicate an influence on apoptotic processes (p = 0.0021 and 0.0049) with three SGs, which are a subset of genes from the first term, *SP100*, *SKIL*, and *ATF3*. Further, module one genes are implicated in positive regulation of GTPase activity (5 SGs, p = 0.0069), epithelial cell differentiation (5 SGs, p = 0.0088), cellular response to IFN-γ (2 SGs, p = 0.024), platelet degranulation (2 SGs, p = 0.033), and cellular response to IL-1 (2 SGs, p = 0.049).

Apart from *SP100* (major constituent of PML-SP100 bodies, see module three), *SKIL* (part of SMAD pathway, regulating cell growth and differentiation), and *ATF3* (member of the broadly acting CREB family of transcription factors), frequently implicated genes include *CCL2* (chemokine ligand for CCR2, exhibits chemotactic activity for monocytes [191]) and *CAMK1* (broadly acting calmodulin-dependent protein kinase, involved e.g. in the ERK cascade). All above SGs are down-regulated in stroke patient blood (own results).

SP100 is induced by IFNs, presents with potent antiviral and tumour suppressor effects, and has been shown to induce apoptosis via the extrinsic pathway together with *FLASH* or caspase-2 and p53 in *PML* nuclear bodies (see module three). [192] A reduction of *SP100* expression would thus likely lead to inhibition of apoptotic events. Histone deacetylase (*HDAC*) inhibition (compare module five) suppresses the IFN-mediated *SP100* up-regulation. [193] Recently, *SP100*, *STAT1* (compare module three), and *KLF4* (compare module two) were found to be co-elevated in the peripheral monocytes of tobacco-smoking (but not in non-smoking) HIV-positive patients and associated with an increased depressive index. [194]

SKIL, also known as SnoN, is known as a pro-apoptotic mediator that can bind and activate p53, and is targeted by mir-30 family members, which can ameliorate its apoptotic effect. [195] Conversely, Smad3, a pro-apoptotic protein, is repressed in synergy by *STAT3* and the *SKIL* paralog c-Ski. [196] A recent study has identified *SKIL* at the heart of a feedforward loop including *EGR1* and hsa-miR-124-3p in the blood of schizophrenia patients, which showed a down-regulation of *EGR1* and *SKIL*, and concomitant up-regulation of miR-124-3p. [197] However, in post-stroke whole blood, *EGR1* and miR-124-3p are unchanged (own results).

ATF3 also shows a pro-apoptotic effect, it can be activated by stress mediator p38 MAPK; [198] reduction of *ATF3* in a cell model via siRNA interference reduced apoptosis. [199] *ATF3* is a key regulator of macrophage IFN responses, it represses pro-inflammatory pathways (e.g. IL-6, TLR), and is also induced by IFNs. Correspondingly, *ATF3*-deficient mice are more susceptible to endotoxin-mediated shock by cytokine overproduction. *ATF3* and *CXCL10* (compare module two) are co-induced by IFNs. [200] *ATF3* sustains *STAT3* phosphorylation through inhibition of phosphatases,

and thus amplifies IL6-gp130-STAT3 signalling.[201] *ATF3* down-regulates *ACHE* expression during stress by binding to the consensus recognition site of cyclic-AMP responsive element binding (CREB) proteins, »TGACGTCA«. In this case, a non-enzymatic, pro-apoptotic function of *ACHE* is stipulated.[202] Similarly, *ATF3* attenuates hypoxia-induced apoptosis by down-regulating the expression of the pro-apoptotic factor carboxyl-terminal modulator protein (CTMP), via binding to the ATF/CREB site in the CTMP-promoter.[203]

Module one genes, which decrease translational activity, are in the majority down-regulated, which may indicate a positive influence on transcriptional mechanisms, in line with the facilitation of a response to the insult. Given the uncertainty of whether the transcription factors of module one act as activators or repressors (FANTOM5 data only supplies binding probability, not mode of action), the real picture likely is more complicated, with mixed outcomes in expression regulation. A major component of module one processes is apoptotic signalling, which the top three (down-regulated) genes, *SP100*, *SKIL*, and *ATF3*, are all involved in. All three genes are oncogenes, and as such involved in cell cycle control. While module one genes present a complex regulatory picture, the function of the three most-involved genes may be summarised as pro-apoptotic; their consistent down-regulation in patient blood after stroke thus implies an inhibition of apoptotic processes inside module one. Additionally, *SP100* and *ATF3* present with clear ties to cholinergic processes, in their involvement with neurokines and STATs, association with nicotine consumption, and the attenuation of *ACHE* expression.

Module Two

Module two, being the largest module, predictably also offers most process-related terms (50 GO terms, 34 SGs), of which most are related to immunity or basic processes of cell physiology. For the sake of clarity, immunity-related terms will be explored first. Module two genes most significantly participate in regulation of response to cytokine stimulus (4 SGs, p = 2.9E-04), response to LPS (5 SGs, p = 5.6E-04), inflammatory response (6 SGs, p = 0.0015), and T- and B-cell differentiation (2 SGs each, p = 0.036 and 0.041). The highest DE genes from module two involved in these processes are *SORL1* (a known regulator of neurokine signalling), *STAT5B*, the IL-10 receptor α, *PLXNB2* (involved in cell migration), *JAK2*, the transcription factor *KLF4* (with many broad functions, but implicated in MAPK/ERK cascades and IL-4 mediated macrophage differentiation[204]), *PTPN2* (phosphatase involved in dephosphorylating JAKs and STATs), and *CXCL10* (pro-inflammatory chemokine ligand implicated in response to brain injury by activating microglia).

Non-immune physiological terms refer to regulation of cell migration (8 SGs, p = 1.4E-04), negative regulation of MAPK cascade (4 SGs, p = 4.7E-04), cellular response to peptide (5 SGs, p = 9.5E-04), nucleus organisation (3 SGs, p = 0.0010), erythrocyte differentiation (3 SGs, p = 0.0010), negative regulation of cell adhesion (4 SGs, p = 0.0015), regulation of ERK cascade (3 SGs, p = 0.0085), regulation of angiogenesis (3 SGs, p = 0.013), regulation of cold-induced thermogenesis (2 SGs, p =

0.031) and several more basic processes. Many of these processes appear to be vital to the immune functions described above, as they share many of the most highly DE genes, such as *SORL1*, *STAT5B*, *PLXNB4*, *KLF4*, and *PTPN2*.

Module two includes several genes directly involved in the JAK/STAT immune response, such as *IL10RA*, *JAK2*, *STAT5B*, and *PTPN2*; module two genes are mostly down-regulated except for *SORL1* (which shows dramatic increase in terms of count-change) and *STAT5B* (own results).

SORL1 gives rise to the protein SorLA, synonymous with LR11, a transmembrane receptor that can interact with a wide variety of ligands intra- as well as extra-cellularly.[205] SorLA is primarily known as the neuronal ApoE4 receptor, and thus widely associated with AD risk. Among other functions, it regulates APP trafficking and processing, and is significantly decreased in the AD-vulnerable regions of late-onset AD patients.[205] However, in addition to its well-studied neuronal roles, it is highly expressed on the surface of monocytes and macrophages, and additionally up-regulated in acute myeloid leukaemia.[206] SORL1 binds many immune-related ligands such as the neurokine IL-6 and soluble neurokine receptors IL6R and CNTFR, mediating their cellular uptake. It associates with the transmembrane IL-6 receptor and reduces downstream effects via a reduction in STAT3 phosphorylation.[207] Conversely, while decreasing *cis* signalling as just described, SORL1 may increase *trans* signalling, i.e., IL6 availability in the blood stream which then binds to the soluble IL6 receptor and can affect any cell possessing the ubiquitously distributed gp130.[208] It has been shown that overexpression of a soluble gp130 form can effectively suppress inflammation mediated by the soluble IL6 receptor without interfering with the function of the transmembrane IL6 receptor.[208]

PLXNB2 can promote inflammation via activation of the NF-κB pathway.[209] In addition, it prominently mediates a plethora of the functions of angiogenin (ANG): it is the receptor for ANG in physiological and pathological cells; ANG acts though *PLXNB2* on cell proliferation; *PLXNB2* modifies ANG RNA-processing activity and cell type specificity (see Section 1.3.3); and *PLXNB2* mediates neuroprotective effects of ANG.[210] Moreover, *PLXNB2*-deficient macrophages showed greater mobility and would healing capabilities than WT cells.[211] A recent study shows its close association to inflammation-related circulatory events: upon inflammatory stimulation (using TNF-α, IL-1β, IL-4, and IL-6) of murine and human cells, endothelium-derived extracellular vesicles carrying miRNAs were released and taken up by monocytes causing decreased monocytic *PLXNB2* levels and increased splenic monocyte mobilisation.[212] Twelve miRNAs were enriched in these vesicles, most of which show high differential expression in stroke patient blood (own results in brackets); these are hsa-miRs: -632 (not detected in stroke), -126-3p (LFC = -1.12, p = 3.0E-11), 151a-3p (LFC = -2.37, p = 5.6E-22), -26b-5p (LFC = -1.17, p = 5.0E-05), -126-5p (LFC = -2.23, p = 5.2E-07), -7a-5p (LFC = -0.54, p = 0.03), -1972 (not detected), -15b-5p (LFC = -0.85, p = 5.2E-07), -23a-3p (LFC = -1.68, p = 9.9E-16), -374b-5p (LFC = -1.85, p = 1.3E-05), -23b-3p (LFC = -1.73, p = 1.2E-17), and -7b-5p (LFC = 0.73, p = 4.1E-06).

KLF4 has gained much attention since it was discovered to be one of the four factors for induction

of pluripotent stem cells (iPSCs, induced by OCT3/4, SOX2, MYC, and KLF4).[213] In macrophages (often using the model RAW264.7), *KLF4* controls activation in response to LPS stimulation by regulating, among others, NF-κB, TGF-β, IL-1β, and HMGB1, at least partly through Smad3 inhibition (compare module one).[214,215,216] In monocytes, *KLF4* regulates differentiation towards a pro-inflammatory type of resident monocytes, such that *KLF4* KO mice completely lacked inflammatory monocytes in blood and spleen.[217,218,219]

The protein tyrosine phosphatase *PTPN2* is implied in IL-1β-mediated inflammation. Mice deficient of *PTPN2* die few weeks after birth because of anaemia, colitis, and severe systemic inflammation. Macrophages depleted of *PTPN2* show excessive inflammasome activation. *PTPN2* reduction leads to more general inflammation, but also less tumour susceptibility, likely because of a more efficient eradication of oncogenic cells.[220] Increased inflammatory cascades following *PTPN2* reduction are likely caused by the decrease in *STAT1* and *STAT3* dephosphorylation, and the concomitant increase in STAT signalling (compare module five, *HDAC7*).[221]

In summary, module two genes seem to be responsible for facilitating an adequate immune response by regulating basic processes in order to enable immune cells to fulfil their functions (e.g., response to cytokines and cellular mobility). *SORL1* up-regulation suppresses STAT3 activation, but putatively increases IL-6 availability in the blood stream, thereby pushing a whole-body immune activation. *PLXNB2* reproducibly is reduced in response to inflammatory signalling, which leads to higher-functioning monocytic cells. *PTPN2* reduction likewise is associated with a higher-functioning immune system and increased inflammatory cascades via an inhibition of STAT inactivation. *KLF4* reduction may indicate a pro-differentiation signal, generating mature immune cells to interfere with the infarction. In addition, *PLXNB2* interferes with angiogenin (ANG) function, and thus may directly impact tRF generation.

Module Three

Module three, with 14 GO terms and 13 SGs, appears similar in principle to module two, although much fewer basic physiological terms are involved. This is explained by the smaller size of the test set; the ontologies for basic terms include substantially more genes, and the likelihood of significant enrichment in the test set is reciprocal to test set size. The most significant biological processes of module two genes involve cellular response to IFN-γ (3 SGs, p = 1.6E-04) and positive regulation of transcription (5 SGs, p = 9.1E-04), as opposed to negative regulation of transcription found in module one. Further, module three genes are involved in the cytokine-mediated signalling pathway (4 SGs, p = 0.0018), regulation of IFN production (2 SGs, p = 0.0034), JAK-STAT cascade (2 SGs, p = 0.0048), negative regulation of angiogenesis (2 SGs, p = 0.0056), regulation of innate immune response (2 SGs, p = 0.035), and positive regulation of cytokine production (2 SGs, p = 0.048). Most frequently occurring genes are *STAT1*, *PLSCR1* (implicated in amplification of IFN response), *PML* (also associated with IFN- and TNF-responses), and *IRF5* (implicated in TLR7/8-induced induc-

tion of IFNs and other pro-inflammatory cytokines). All SGs of module three are down-regulated in stroke patient blood (own results).

Phospholipid scramblase 1 (*PLSCR1*) is an oncogene implicated in cell cycle control, apoptosis, and mediation of antiviral response. While its mechanism of action is yet unexplained, the mature protein localises to the nucleus and has been shown to bind DNA.[222] While it does not induce apoptosis on its own, its overexpression has inhibitory effects on several cell cycle controllers and anti-apoptotic proteins such as Bcl-2.[223] *PLSCR1* participates in the antiviral response by potentiating IFN activity, which increases expression of a subset of IFN-stimulated genes (ISGs), including STAT1.[224] It is expressed in human macrophages and monocytes, in which it is increased in systemic inflammatory conditions, and seems to also contribute to pro-thrombotic conditions.[225,226]

Promyelocytic leukaemia protein (*PML*) is, together with *SP100* (see module one), the major constituent of PML nuclear bodies, that are also known as nuclear domain 10 (ND10). These small nuclear organelles are known for their peculiar and enigmatic function in antiviral response, which they seem to convey by a wide variety of molecular functions, from chromatin modification to physical trapping of virus particles. Recently, however, they have also been connected to a direct regulation of innate immune responses. IFN therapy increases the expression of both *PML* and *SP100*, and enhances their antiviral activity.[227] Correspondingly, *PML* depletion reduces the capacity of IFNs to interfere with viral infection.[228] PML has the capacity to modulate different stages of the pathway from IFNs through ISGs to the activation of STATs by associating with and stabilising transcription factor complexes, for instance by binding directly to *STAT1* and interferon regulatory factors (IRFs) (see below).[229] The production of IL-1β and IL-6 is significantly reduced in *PML*-deficient cells;[230] and correspondingly, *PML*-deficient mice are resistant to LPS-mediated lethality.[231] This directly links to an epigenetic switch that leads to cellular transformation: inflammation, through NF-κB, activates Lin28 and rapidly reduces let-7 miRNA levels. This leads to de-repression of the IL-6 mRNA, and the increased IL-6 peptide conveys cellular transformation through *STAT3* activation, as well as positive feedback by inducing NF-κB synthesis.[232]

IRF5, a well-studied member of the Interferon Regulatory Factor transcription factor family, is an important element of most blood-borne immune cell types, particularly of macrophages and pro-inflammatory monocytes.[233] It has been demonstrated that *IRF5* is involved in transcriptional induction of IL-6, IL-12, and TNF-α mediated by the toll-like receptor (TLR)/myeloid differentiation primary response 88 (MyD88) complex. *IRF5*-deficient mice also show resistance to LPS-mediated lethality (compare PML/SP100).[234] While KO-mice showed severely impaired production of the aforementioned cytokines, IFN-α production was not impaired; the responsible interferon-stimulated response elements have yet to be determined.[234] *IRF5* is highly expressed in monocytes and macrophages as well as B cells and dendritic cells, and its expression is induced by a pro-inflammatory environment.[235] On the spectrum of macrophage states after differentiation, *IRF5* (together with *IRF1* and *IRF8*) is involved in the commitment to a pro-inflammatory (M1) phenotype (as

opposed to the »wound healing« M2 type).[236] The M1 phenotype is brought about by increased activity of NF-κB and STAT1, whereas the M2 counterpart is induced by IL-4-, IL-10-, and IL-13-mediated STAT3 and STAT6 signalling.[237] Molecular competition for MyD88 binding between IRF5 and IRF4 is crucial for the determination of pro- (*IRF5*) or anti-inflammatory (*IRF4*) differentiation.[238] Thus, a down-regulation of IRF5 together with unchanged levels of IRF4, as is the case in stroke patient blood (own results), would lead to anti-inflammatory conditions in monocyte and macrophage populations. M1→M2 macrophage transition supports resolution of inflammation and tissue healing; reduction of IRF5 expression in monocytes and macrophages via siRNA interference improved healing after myocardial infarctions and skin wounds in mice, in parallel with a reduction of the pro-inflammatory cytokines IL-1β, IL-6, and TNF-α.[239]

Module three genes (implicated in positive regulation of transcription, but all down-regulated) appear to act in partial opposition to module one genes (implicated in negative regulation of transcription, and all down-regulated), thus possibly being part of homeostatic events surrounding gene transcription in response to inflammatory events. Modules one and three also share GO terms involved in responses to, and production of, interferons and interleukins.

In summary, module three seems to be representative of immune suppression via several mechanisms; mediation of pro-inflammatory signalling via INFs is repressed, and IRF5 suppression may induce an anti-inflammatory, pro-resolving differentiation of monocytes and macrophages; cytokines and STAT signalling are likewise down-regulated; ND10 nuclear bodies are repressed. On the other hand, pro-apoptotic signalling may be de-repressed via PLSCR1 reduction.

Module Four

Module four, with 12 significant GO terms from 7 SGs, is a fairly small module, which is nevertheless highly associated with immune processes; most significantly, genes of module four convey positive regulation of innate immune response (2 SGs, $p = 0.0039$) and positive regulation of cytokine production (2 SGs, $p = 0.0075$); more accurately, IL-1 production (1 SG, $p = 0.038$). Further, module four genes show association with negative regulation of myeloid cell differentiation (1 SG, $p = 0.038$), positive regulation of response to cytokine (1 SG, $p = 0.038$), myeloid cell homeostasis (1 SG, $p = 0.042$), positive regulation of inflammatory response (1 SG, $p = 0.045$), and negative regulation of myeloid leukocyte differentiation (1 SG, $p = 0.049$). The most prevalent SGs in these terms are *GBP5* (activator of the NLRP3 inflammasome assembly), *MAFB* (transcription factor required for monocyte differentiation), and *ZBP1* (cytoplasmic DNA-sensor which activates downstream IFN production in activated macrophages). All above SGs are down-regulated in stroke patient blood (own results).

GBP5 (for Guanine Nucleotide Binding Protein) is a molecular marker for the classically activated, pro-inflammatory M1 type macrophage,[240] and a critical factor for the assembly of the *NLRP3* inflammasome.[241] *GBP5* is strongly induced by IFNs through NF-κB, and in turn induces expression of IFNs and pro-inflammatory cytokines such as IL-6 and TNF-α.[242] Consequently, it plays a critical

role in response to viral infection, e.g. by Influenza or HIV, as well as diverse other pathogens.[242,243] In mice, miR-21-5p inhibition led to an increase in macrophage GBP5, with a concomitant increase in TNF-α and a decrease in the anti-inflammatory IL-10.[244] However, in stroke patient blood, both miR-21-5p and GBP are down-regulated (own results).

MAFB is a transcription factor with critical roles in macrophage differentiation and function, and specific for mononuclear phagocytes in all cells of the haematopoietic system.[245] Macrophages deficient in *MAFB* and *MAF* reacquire the ability for self-renewal, but only upon concomitant up-regulation of two pluripotent stem cell-inducing factors, *KLF4* and *MYC* (compare module two).[246] However, in stroke patient blood (own results), *KLF4* is reduced while *MAF* and *MYC* are unchanged. After induction of ischemic stroke, macrophage-specific *MAFB*-deficient mice showed excessive sterile inflammation, likely due to a failure in clearing of damage-associated molecular patterns (DAMPs).[247] The authors found *MAFB* to be a critical controller of the macrophage scavenger receptor 1 (MSR1), which in turn is essential for the clearance of DAMPs after ischemic stroke, and is also down-regulated in stroke patient blood (own results, LFC = -3.54, p = 2.9E-06). Additionally, retinoic acid receptor agonist *Am80* increased *MAFB* and *MSR1* expression in the delayed phase of ischemic stroke, and had ameliorating effects on stroke pathology.[247] *miRNeo* query of retinoic acid receptor interactions in $CD14^+$ cells indicates *RARA*, the α retinoic receptor, as most active towards MAFB (lower activity also exhibited by *RARG* and *RXRA*). *RARA* also targets STATs 1, 5A, and 5B, as well as *ACHE* and *IL6ST* (also known as gp130) in these cells.

An *ex vivo* experiment of *MAFB* inhibition by siRNA in human $CD14^+$ monocytes found an elevation in IRF3 phosphorylation and concomitant increase in INF-α and INF-β production.[248] *MAFB* also seems responsible for direct regulation of all genes of the C1q complement complex, which activates the classical component pathway.[249] Thereby, and by regulation of the Axl protein, *MAFB* is essential for efferocytosis, the phagocytosis of apoptotic cells *in vivo*.[250] *MAFB* is regulated by diverse pathways; immunological mediators (e.g. cytokines), lipid metabolism, and miRNAs.[245] Two miRNAs experimentally identified as regulating *MAFB*, miR-152[251] and miR-155[252], are significantly down-regulated in stroke patient blood (own results; hsa-miR-152-3p: LFC = -1.97, p = 2.0E-11; hsa-miR-155-5p: LFC = -0.69, p = 6.5E-04). miR-155 has additionally been shown to induce pro-inflammatory macrophages, while *MAFB* itself promotes anti-inflammatory M2-type macrophage differentiation.[100] In $CD14^+$ monocytes, MAFB interacts with *CHRNA6*, *IL6*, the *IL6* receptor, and *STAT1* (via *miRNeo*, from Marbach *et al.*[103] regulatory circuits). In summary, *MAFB* reduction after stroke may contribute to a shift towards pro-inflammatory M1-type macrophages, prolonged activity of which may inhibit efferocytosis and the M2-mediated healing process in the delayed phase.

Z-DNA binding protein 1 (*ZBP1*, also known as *DAI*) is an IFN-induced cytoplasmic sensor of DAMPs that positively mediates various forms of programmed cell death, general pro-inflammatory events (such as cytokine production), and *NLRP3* inflammasome assembly; however, its triggering

ligands or molecular patterns are still unclear.[253] More recently, it was shown that *ZBP1* is necessary for type-I and type-II IFN-mediated necroptosis (a programmed form of necrosis),[254] by sensing viral as well as endogenous RNA (in addition to DNA), possibly in the unusual Z-conformation.[255]

Murine macrophages lacking the MyD88 TLR adapter protein were able to undergo apoptosis via redundant TLR pathways mediated by *ZBP1*. In contrast, KO of the type-I INF receptor, as well as of its downstream effectors *STAT1* and *IRF9*, abolished the macrophages' ability to undergo TLR-mediated MyD88-independent apoptosis.[256] Moreover, de-novo transcription and protein synthesis (compare »regulation of transcription« in modules one and three), as well as JAK1/STAT1 transcriptional activation are required for IFN-induced necroptosis through *ZBP1* signalling.[254] *ZBP1*-deficient mice (via CRISPR/Cas-9 KO) were significantly protected from acute IFN-mediated systemic inflammatory response syndrome (SIRS),[254] making the *ZBP1* down-regulation in stroke patient blood (own results, LFC = -1.70, p = 0.001) a logical bodily response to prevent CNS injury-induced immunodepression syndrome (CIDS), and a possible component of the counter-regulation, compensatory anti-inflammatory response syndrome (CARS) (see Section 1.2.5). Notably, NF-κB signalling blocks type-II IFN-induced necroptosis (compare modules two and three).[257]

In summary, module four constitutes a small, largely anti-inflammatory module, conveying an attenuation of inflammatory processes by limiting the sensing of inflammatory stimuli (*ZBP1*), activation of inflammatory pathways, inflammasome assembly and cytokine production (*GBP5*), and clearance of DAMPs in the infarct area (*MAFB*). However, GO terms also indicate a positive influence on differentiation of monocytes. For instance, the decrease in *MAFB* expression may promote a delayed phenotypic shift towards M1-type macrophages, which may have a deleterious effect on patient recovery.

Module Five

Module five is a medium-sized module (29 GO terms from 18 SGs) and highly involved with stroke-relevant processes, most importantly, vascular permeability. Most terms relate to basic molecular functions of the cell, such as regulation of cell-substrate adhesion (3 SGs, p = 0.0037), cellular component maintenance (2 SGs, p = 0.0052), membrane depolarisation (2 SGs, p = 0.0052), regulation of protein tyrosine kinase activation (2 SGs, p = 0.0063), positive regulation of small molecule metabolism (2 SGs, p = 0.0063), fatty acid metabolic process (3 SGs, p = 0.0064), positive regulation of lipid metabolic process (2 SGs, p = 0.010), regulation of cytokine production (3 SGs, p = 0.022), response to hypoxia (2 SGs, p = 0.027), and several more. The genes most highly implicated in these processes are *SRC*, *ABCD1*, and *HDAC7*. Module five SG expression is mixed in stroke patient blood; of the most prevalent genes, *SRC* and *ABCD1* are down-regulated, and *HDAC7* is up-regulated.

Src, from the *SRC* gene, is a well-studied kinase from the larger family of Src kinases implicated in diverse physiological processes. Increased vascular permeability (VP), i.e., a loss of blood-brain-barrier function, can be induced by TNF-α, IL-1β, IL-6, and vascular endothelial growth factor

(VEGF). Of relevance, Src controls VEGF expression after ischemic stroke, thereby modulating VP. Reduction of Src (but not Src family kinase Fyn) activity in mice via complete congenital KO or pharmacologic inhibition (using the synthetic inhibitor PP1) resulted in a reduction of permanent middle cerebral artery occlusion (MCAO)-induced stroke volume of 50% and up to 70%, respectively. This beneficial effect was associated with a prevention of VP.[258] In these experiments, Src inhibition seemed to restore perfusion and oxygenation in the penumbra (the region adjacent to the infarct), protecting the surrounding tissue. Later, these VEGF-modulating effects of Src on VP and stroke recovery were replicated in rats with transient infarction.[259] In another set of experiments, it was shown that the neuroprotective effect of ischemic postconditioning in mice probably depends on Src kinase activation.[260] In a human cellular model, Src was up-regulated upon inflammatory TNF-α signalling and conveyed increases in VP in a p38 MAPK-dependent manner (compare module one, *ATF3*).[261] Src family kinases can physically bind to gp130, phosphorylate STAT3,[262] and are intricately involved with TLR-CD14 signalling, in an »important but not obligatory« manner.[263]

ABCD1 is a member of the ATP binding cassette (ABC) superfamily of active transporters, responsible for the transport of very long chain fatty acids, and implicated in inflammatory processes by its influence on peroxisomal β-oxidation.[264] It is best known for its role in adrenoleukodystrophy (ALD), where an *ABCD1* deficiency leads to microvascular perfusion anomalies in the brain.[265] The working hypothesis states that *ABCD1*-related up-regulation of adhesion molecules in the endothelium causes higher leukocyte interaction and thus impairs blood flow in the capillaries. Notably, carriers of the ApoE4 allele are more significantly affected by ALD (compare module two, *SORL1*).[266] Most molecular changes caused by the deficiency are likely a result of impaired fatty acid transport into peroxisomes. *Abcd1*-deficient mice show a dramatically altered brain region-specific phospholipid profile.[264]

HDAC7 (histone deacetylase 7) is a protein responsible for deacetylating histones and non-histone proteins, with regulatory implications in cell proliferation, apoptosis, differentiation, and migration.[267] *Hdac7*$^{-/-}$ mice are embryonic lethal because of an angiogenetic failure leading to rupture of blood vessels.[268] It has been shown that the HDAC7 protein directly interacts with STAT3, deacetylation of which interferes with its ability to dimerise, thus impairing its functionality.[269,267] *HDAC7* was also found to induce an apoptotic gene program and c-Myc suppression upon ectopic expression in human SD-1 cells.[270] *HDAC7* additionally suppresses macrophage genes relevant for phagocytosis, TLRs, interleukins and TNF pathway genes by acting as a transcriptional repressor of myocyte enhancer factor *MEF2C*.[271] In human C10 cells reprogrammed to macrophages, *HDAC7* expression significantly inhibited the expression of TNF-α, IL-1β, and IL-6.[271] VEGF stimulates HDAC7 phosphorylation, leading to its aggregation in the cytosol and subsequent suppression of angiogenesis mediated by matrix metalloproteases (MMT) 10 and 14, which degrade proteins essential for vessel integrity.[272,273] *HDAC7* can also directly associate with *RARA* (see module four, *MAFB*) to form a repressor complex that inhibits, among others, miR-10a.[274] This inhibition led to de-repression

of pro-inflammatory signalling via *GATA6* and *VCAM1*, with negative impact on VP in vascular endothelia.

In summary, module five genes seem to be primarily associated with vascular permeability and the integrity of the blood-brain-barrier. Reduction of *SRC* may be beneficial in regard to stroke volume, whereas the effects of *ABCD1* decrease on vessel integrity are less clear. Secondarily, down-regulation of *SRC* and *ABCD1* may have a negative impact on lipid metabolism, with consequences for lipid-mediated signalling. *HDAC7* up-regulation, on the other hand, may constitute a pro-apoptotic signal.

4.5.5 Transcript Clustering by smRNA:TF:gene Feedforward Loops Increases Informative Resolution of Gene Set Enrichment Analyses

In principle, the analysis of separate modules represents changes brought on by DE TFs in stroke patient blood, focused on $CD14^+$ cells by analysing feedforward loops comprising smRNAs and TF-interactions specific to $CD14^+$ cells. An advantage of the separation via FFL association is a reduction in dimensionality, which may make the implications of each module more accessible than a complete list of all changed genes or TFs in the experiment, and may also increase the informative resolution of the GO enrichment analyses. Visual inspection of Figure 4.9 indicates a partition between modules two and four on the top, and modules one and five on the bottom of the map. The role of module three is visually less clear. Module three is the module most »intermingling« with the other modules, particularly in the case of *IRF5*, which spatially is closer to module one and five than it is to three (see Figure 4.9, bottom). More generally, there are two zones of module three intermingling, between modules one and five, and between modules two and four.

Visual representation of the GO terms derived from non-modularised analysis versus the terms collected from module analyses clearly shows the proposed advantage regarding informative resolution of the method (Figure 4.10). The non-modularised version shows less terms, that are also less specific (Figure 4.10 A, reproduced from Figure 4.2), while the module-derived terms are higher in number, and also refer to more specific processes (Figure 4.10 B). Intriguingly, this t-SNE visualisation of GO terms shows a central cluster around module one, comprising terms from all other modules. Since the similarity of terms is calculated via their shared genes, this may indicate that processes involving co-operation of modules are grouped in the center, while module-specific processes cluster at the edges. Correspondingly, module three terms cluster on the left side in close vicinity of this central cluster, while module two and module five terms are located in opposite areas on the graph.

The most prevalent themes of module function are inflammatory events, in particular those mediated by TLR/MyD88-associated pathways, IFN responses, and interleukin and TNF-α signalling. Many of these innate immune pathways are important in sterile inflammation as well as in response to viral infection. This is in concert with the fact that the most prevalent genes identified by module GO analysis are pivotal regulators of immunity. Other themes implied by the analysis are apop-

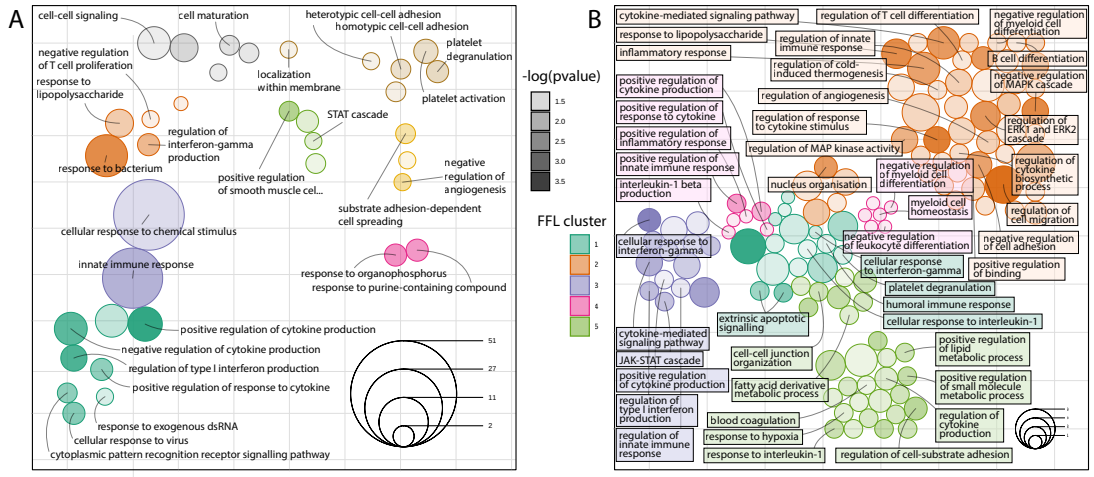

Figure 4.10: FFL module Gene Ontology Enrichment Increases Informative Resolution. A comparison between GO enrichment performed on (A) differentially expressed (DE) genes and (B) DE genes classified by FFL network modules shows similar biological processes but increased information depth. Size denotes number of significant genes in term, colour indicates p-value (all $p < 0.05$). **A)** DE genes with absolute aclfc > 1.4 are enriched in processes of immunity and circulation. Cluster colour derived from t-SNE similarity. Reproduction of Figure 4.2. **B)** t-SNE mapping of GO terms derived from analysis of single FFL modules shows details of involvement with immune processes and basic cellular function. Colour indicates original FFL module (see text). The middle cluster, comprised of parts from all modules surrounding module 1, may signify an »area of cooperation« between the modules.

totic/necroptotic events, transcriptional activation, lipid metabolism and signalling, and blood vessel integrity. In most cases, the participating forces are representative of activation as well as inhibition, again underscoring the homeostatic aspect of these regulatory processes.

Inflammation as a common theme is represented in all modules, with diverging implications. Modules one, two, and five may be described as pro-inflammatory modules, while anti-inflammatory properties dominate in modules three and four. The module topology of the two-dimensional force-directed map of FFLs (see Figure 4.9) replicates this pro- versus anti-inflammatory dichotomy, indicating that the innate immune response may be the defining factor in FFL modularisation (Figure 4.11). Module one conveys pro-inflammatory signals via the de-repression of IFN response through *ATF3*, module two via the increase in IL6 availability through *SORL1*, the decrease in STAT dephosphorylation by *PTPN2*, and the decrease in *PLXNB2*, which may be mediated by smRNA-carrying vesicles and is associated with splenic monocyte mobilisation. In module five, *HDAC7* in association with *RARA* may repress miR-10a, which in turn de-represses pro-inflammatory mediators. Modules three and four, on the other hand, may be described as largely anti-inflammatory. Module three presents inhibition of IFN signalling via the suppression of *PLSCR1* and *STAT1*, *PML/SP100* nuclear bodies, and *IRF5*. Down-regulation of module four genes may convey attenuation of inflammation through *GBP5*, *MAFB* (and the related *MSR1*), and *ZBP1* suppression. Attenuation of inflammation after stroke may be a bodily response to prevent systemic inflammatory response syndrome (SIRS), however, exaggerated response or external intervention harbours the danger of compensatory anti-inflammatory response syndrome (CARS) and CNS injury-induced immunodepression syndrome (CIDS).

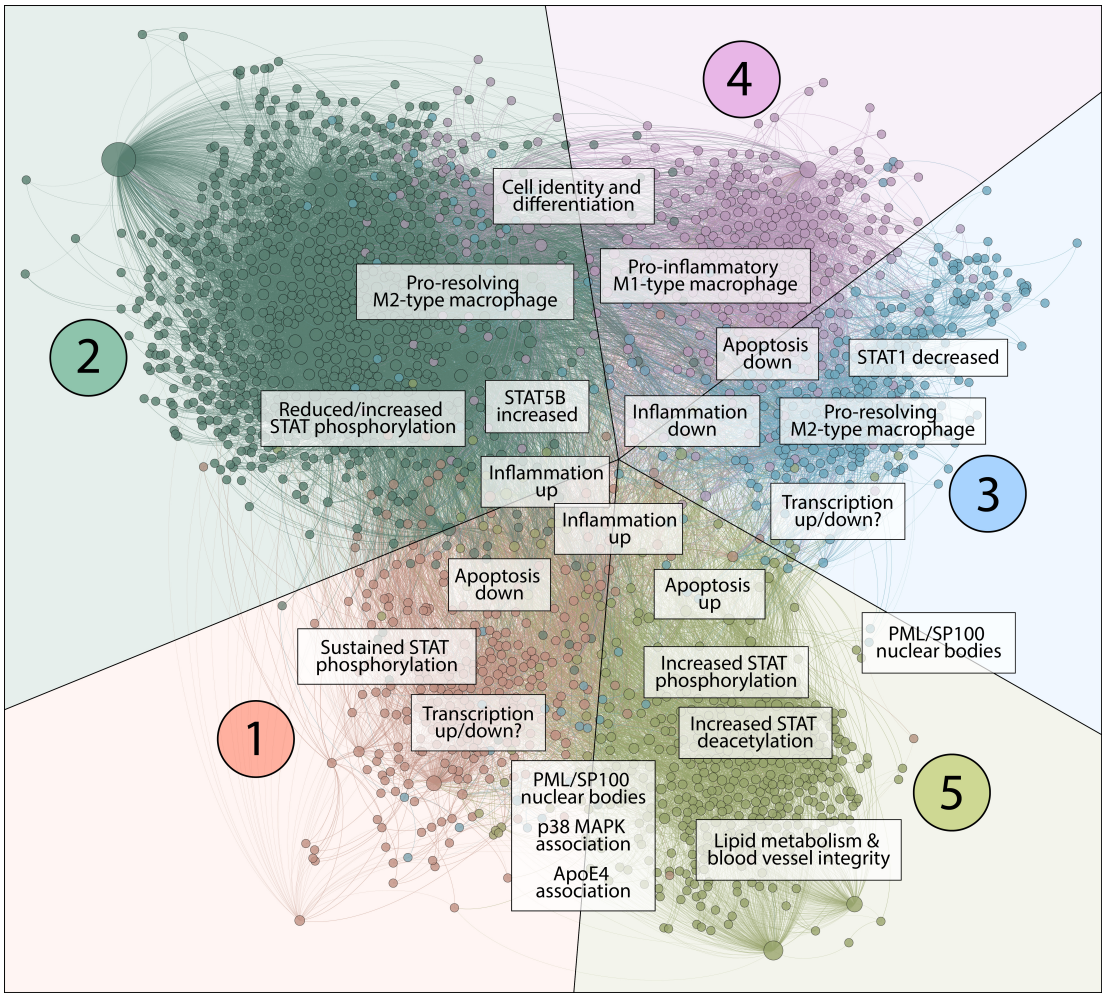

Figure 4.11: Complete Feedforward Loop Network of Differentially Expressed Transcription Factors in CD14$^+$ Monocytes, Annotated. A reproduction of Figure 4.9 annotated with most pertinent molecular functions as identified via individual module GO analysis. Distance to the center of the graph of each annotation resembles the distribution of the category across all modules (closer to the center - more common functional class), and functions shared by two modules are indicated by their annotation across the intersecting lines. The most common functional classes refer to inflammation, apoptosis, and STAT regulation.

Influences on apoptotic/necroptotic events are also distributed between several modules. Module one conveys anti-apoptotic signals via reduction of *SP100*, in agreement with *PML* reduction in module three. Additionally, *PLSCR1* down-regulation in module three may have an indirect anti-apoptotic effect through de-repression of anti-apoptotic proteins such as Bcl-2. *ZBP1* in module four is a DAMP sensor that can induce necroptosis independently of MyD88, however, it is suppressed in stroke patient blood and requires STAT1 signalling (which is also suppressed), and de-novo protein synthesis. Transcriptional activation is regulated by genes in modules one and three, but the final effect of their competition cannot be assessed with certainty. *HDAC7* up-regulation in module five, on the other hand, may induce a pro-apoptotic program.

Lipid metabolism and blood vessel integrity are implied specifically in module five. Peroxisomal β-oxidation of very long chain fatty acids is reduced through the inhibition of *ABCD1*, with consequences in lipid mediator signalling (e.g., regulation of *MAFB*) and inflammation. However, the bottom line effect of module five gene reactions to stroke in regard to blood vessel stability, i.e., the question if *SRC*, *ABCD1*, and *HDAC7* regulation are beneficial or detrimental to vascular integrity, is still largely unclear and a worthwhile subject for further studies.

Another common theme of module gene processes is the involvement and regulation of signal transducers and activators of transcription, STATs. Many SGs are inducers of STATs (*PLSCR1*, *IRF5*, *RARA*, *MAFB*), binding partners (*SKIL*, *PML*), or involved in activation or deactivation via phosphorylation/dephosphorylation and acetylation/deacetylation (*ATF3*, *SORL1*, *PTPN2*, *SRC*, *HDAC7*). Thus, homeostasis of STAT transcription and activation seems to play an important role in the observed processes. Of note, STAT1 is decreased in stroke patient blood, but STAT3 is not differentially regulated. However, effects mediated by modulators of activity, such as sustained STAT3 phosphorylation via *ATF3* (module one), reduced STAT3 phosphorylation via *SORL1* (module two), increased STAT3 phosphorylation via *PTPN2* (module two) and *SRC* (module five), or STAT3 deacetylation via *HDAC7* (module five) can have dramatic impacts on cellular function without a change in STAT3 expression. Additionally, the studies cited on STAT control via phosphorylation and acetylation are not comprehensive, so some of the implied proteins may interact with STATs other than STAT3. In stroke patient blood, changes are seen in STAT1 (down-regulated), STAT2 (down-regulated), and STAT5B (up-regulated).

Among module functions is also the control of blood cell differentiation, particularly of monocytes and macrophages. Both cell types can be driven towards a pro- or an anti-inflammatory phenotype via expression and activation of certain mediators. *KLF4* (module two) drives monocytes towards a pro-inflammatory phenotype, and *IRF5* (module three) mediates commitment to the pro-inflammatory M1-phenotype of macrophages in response to IFN signalling. Their concomitant down-regulation may thus produce an anti-inflammatory, pro-resolving phenotype of monocytic immune cells. Additionally, *IRF5*-mediated M1-commitment is conveyed by STAT1 signalling, which is impaired through *STAT1* down-regulation, whereas M2-commitment is mediated by STAT3 and

STAT6 signalling, both of which are unimpaired. On the other hand, *MAFB* (module four) expression drives commitment to the anti-inflammatory M2 type, and thus, its down-regulation may convey an opposite signal to *KLF4* and *IRF5* regulation.

A number of module SGs also show cholinergic association: apart from the cholinergic/neurokine interface connecting IL6, gp130, and JAK/STAT to cholinergic properties of immune cells and neurons,[1] *ATF3* down-regulates *ACHE* which interferes with its stipulated non-enzymatic, pro-apoptotic function; *RARA* induces *STAT1*, gp130, and *ACHE*, and *MAFB* induces *CHRNA6*, *IL6*, *IL6R*, and *STAT1*. Other module cross-associations include module one/module three cooperation in regard to PML/SP100 nuclear bodies, *SP100* and *STAT1* co-elevation in monocytes of tobacco-smoking HIV-positive patients, and their antagonism in regard to the activation or deactivation of transcription. Module two and module four both have an impact on cell identity and differentiation via pluripotency-associated *KLF4* and *MAFB* signalling. Module one and module five are associated via the induction of SGs by p38 MAPK (*ATF3* and *SRC*), the ApoE4-association of *SORL1* and *ABCD1*, the support of *HDAC7* in IFN-mediated *SP100* up-regulation, and their opposite effect on apoptotic signalling. All of these associations (1-3, 1-5, 2-4) can be retraced in the FFL module visualisation (Figure 4.11).

If the human brain were so simple that we could understand it,
we would be so simple that we couldn't.

Emerson M. Pugh

5 Discussion

The general objective of this dissertation was to facilitate an understanding of the complex processes surrounding transcriptional interactions in mammalian cells. The multi-leveled relationships of only two interacting molecule species make analysis and interpretation inaccessible without the help of software, and even the output of software analyses can be overwhelming by the sheer multitude of genes that are involved. Additionally, while this dissertation started with the objective of illuminating cholinergic processes exclusively in neurons, it naturally gravitated towards immunology in all of the different foci: the studied degenerative and non-degenerative psychiatric diseases, as well as stroke, all have significant immunological components. Arguably, immune cells and the central nervous system are those two mammalian tissues that offer the greatest challenges to the life sciences in terms of complexity.

Historically, the areas of immunology and neuroscience research have little in common, and translational advances have been few. However, as Robert Dantzer illustrates in his recent review,[40] the two disciplines have much to learn from each other, and bringing them closer together is all but necessary. Not by coincidence, the description of brain-to-immune-signalling in that review is predominated by cholinergic implications in immunity; and the best-studied immune-to-brain-signalling molecules are the endogenous pyrogens IL-1β, IL-6, TNF-α, and IFN-α, which in this dissertation also are frequently implicated. However, current research barely scratches the surface of neuro-immune communication; currently, interactions are mainly studied at the level of cell-to-cell or protein-to-protein (also including transcriptional processes induced by proteins). An enormous amount of transcriptional interactions are still almost completely in the dark, including but not limited to small RNA regulatory processes. The main hurdle of integrating an additional regulatory process onto an already complex and incompletely understood subject such as neuro-immune communica-

tion is the resulting exponential increase in complexity.

This brings us back to the initial objective of this dissertation, the facilitation of an understanding of transcriptional interactions. Since the data generated by modern life-science technologies is not comprehensible to humans in its raw form, and even after statistical analyses often remains overwhelming, dimensionality reduction is a logical step to further the comprehension of the science by the scientist. Indeed, most approaches described in this dissertation result in reduction of dimensionality, and a common train of thought behind the distinct analysis steps undertaken was governed by the idea of a »smart« dimensionality reduction, as opposed to, for instance, exclusively looking at the miRNA→gene relationship with the lowest p-value.

To this end, I had to develop a computational basis for the assessment of transcriptional interactions in a manner that is applicable in practical research, i.e., that can generate results for these complex interactions in a matter of seconds to hours. I also had the fortune to be able to apply these methods to a range of relevant biological data, including the ones discussed in-depth in this dissertation: the cellular model of human male and female cholinergic neurons, and the blood of stroke patients. All undertaken analyses are subject to a wide variety of limitations, the most important of which will be discussed in the following.

For the sake of clarity, the discussion will be split into parts: first, the methodological and technical aspects; second, the bio-mechanistic perspective and basic molecular biology implications; and third, the physiologic, pathologic, and medical/therapeutic inferences.

5.1. Methods

5.1.1 Transcriptional Interactions: *miRNeo*

The comprehensive (however justified this term may be) analysis of transcriptional interactions seems to be, for the moment, a rather marginal endeavour. At the beginning of my work on this dissertation, a database such as *miRNeo*, even in its most basic form, was not available. Recently, some efforts have been published,[275,276] including one which has necessitated a name change of my database, which was previously called *miRNet*.[275] The premise of the approach is simple: for a biological network that is structured in the way of interaction partners connected by molecular interactions, build a database that models interaction partners as nodes of a network, and their interactions as its edges. The technical implementations, however, diverge.

miRNeo follows the philosophy of modelling the studied networks as closely as possible in the raw database, to keep data recall at a minimum in terms of storage and processing power requirements. Neo4j seemed like a fitting platform for its implementation, since it is focused around building large networks with flexible computational requirements and possesses an infrastructure for process optimisation. Additionally, it can be integrated into development environments common in bioinfor-

matics, such as Java and R. Most of the work presented in this dissertation has been performed on Neo4j version 3.0, however, the release of Neo4j version 4.0 was just announced and promises to bring further improvement in terms of handling and performance.[277]

The main drawback of graph database integration into biological applications is the difference in infrastructure to virtually all other data, which is in tabular format. The effort of transitioning data into a dedicated graph format is not justified for simple questions, such as the gene targets of a single miRNA. The practical creation of *miRNeo* from raw data in its current extent, without accounting for development time, would take up the majority of a month in computational time on a standard 16-core personal computer. However, nested analyses with multiple levels, and dynamic analyses with multiple steps in which the analysis in the next step depends on the result of the previous, necessitate computationally efficient implementation, and *miRNeo* was able to handle all complex questions that presented themselves during my work. The most computationally intensive questions were the comprehensive whole-genome feedforward loop analyses, which nevertheless were completed in a matter of hours.

The most important limitation of *miRNeo* is the sum of limitations that apply to the raw data *miRNeo* is created from. Small RNA targeting is immensely complex, and small RNA expression is even more tissue-specific than transcription factor expression.[104] Thus, all results from predictions, be they based on complementarity, evolutionary conservation, or physical modelling, and even experimentally validated interactions can currently not be seen as certain indications of an actual interaction in different contexts, making validations indispensible. However, complex multi-layer interactions are nearly impossible to validate, making this area of research highly dependent on inferral from circumstantial evidence. In 2017, Kenneth Kosik introduced an experimental model of miRNA interactions at a conference for non-coding RNA in neurodegenerative disease.[278] The study included the successive knockout of each one of a set of 11 miRNAs, and observing the cellular phenotype for each of the resulting cultures in a high throughput setting; he gave the cost of these experiments to be in the million dollar range. According to Kosik, the knockout of *each one* of the 11 miRNAs led to a loss of the particular phenotype, which implied that all 11 cooperated to govern the molecular basis of the phenotype. However, this kind of experimentation cannot be applied to all open questions simply for economical reasons, and additionally, there still has been no publication of the study in a peer-reviewed journal as of now (April of 2020).[279]

update when submitting

Similarly, in transcription factor interactions, the shortcomings of raw data may transfer into the database. A very pertinent example of FANTOM5 misannotation is the controversy around the promoters of *CHAT* and *SLC18A3* (see also Section 2.2.3). Since, by the statement of a FANTOM5 scientist, it is possible that the 5'-peaks of the two genes may have been confused because they lie in such close vicinity, it is not possible to distinguish between *CHAT* and *SLC18A3* in this data, or even state with certainty which of the two is implied in an analysis. However, as the immune cell data underlying Figure 4.5 shows, it is very feasible that the *SLC18A3* signal in reality refers to *CHAT*

expression, because blood-borne immune cells do not require a vesicular transporter, but have been proven to express *CHAT*. An advantage in modelling transcription factor interactions, however, is that in using FANTOM5 data and secondary sources such as Marbach *et al.* [103] we are one step further than we are in small RNA analyses: we can differentiate between interactions in the different cell types of the human body. In extension, we can also infer on small RNA regulation in a cell type-specific manner by applying our knowledge on transcription factor interactions in these cell types, as we have attempted in the analysis of blood cell-specific networks in stroke (Section 4.3).

But even simpler shortcomings of the raw data in *miRNeo* must be acknowledged: for instance, the annotation of biologically active molecules, be it DNA, RNA, or proteins, is always in flux, which together with the multiple institutions handling annotation leads to foreseeable deficits in translation between one set of data to the other. Particularly in whole-genome analyses (or likewise whole-miRnome etc.), individual control of every nomenclature deficit that results in loss of information (e.g., gene identifiers are different in experimental data and database) is not possible. For this reason, I integrated several identifiers (e.g. Entrez, HGNC, ENSEMBL) with failsafe mechanisms for the identification of as many molecules as possible. Still, there has been loss in several of the analyses. For miRNAs, the newer version 22 of miRBase annotation has not yet been implemented, since it might have caused compatibility issues with previous results.

Consequently, the more complex assessments suffer from a combination of single shortcomings of the raw data. For instance, feedforward loops are only feasible in a very particular constellation (see also Section 4.5): because we lack information on TF→smRNA interactions, smRNA species have to be at the center of the loop (X), and transcription factors have to assume the role of controlling while being controlled (Y). Associations of the kind TF→smRNA are still in the stage of anecdotal evidence, for instance the HDAC7/RARA/miR-10a circuit. [274]

5.1.2 RNA Sequencing

By 2020, RNA-seq has by and large outgrown the teething troubles of its initial technological phase. However, the method itself brings with it inherent, largely mathematical problems. A lack of reproducibility was a great concern in the initial periods of RNA-seq, but the reproduction rate of cholinergic cell culture (Section 3.4.1) as well as the validations performed on small RNA during the studies of stroke patient blood (data not shown in this dissertation, but in the associated publication[?]) show an agreeable reliability of the method in our hands.

Unrelatedly, the multiple testing problematic still remains a pertinent issue of modern molecular biology. It is now more impressive to find a negative result in one of these analyses (e.g., no differential expression in a reasonably powered sequencing experiment) than it is to find actual differences. The question of where to place the threshold of significance, or whether to use such a threshold at all, is still a matter of very lively debate among scientists of many disciplines; additionally, consensus thresholds vary between fields or even between different kinds of assay. This dissertation, in general, follows the

philosophy of balance between limiting false positives (by monitoring FDR) and identifying »real« changes (by monitoring adequate powering and effect sizes). In the limited area of RNA-seq, this is still manageable, because standard approaches in the form of multiple correction for differential expression analysis (e.g. in DESeq2,[140] which essentially uses Benjamini-Hochberg correction) and power analysis packages (I used R/powsimR[280]) already exist; in the extended graph-based network analyses, the matter is more complicated (see below in Section 5.1.3).

For sequencing, particularly of small RNA, there remain open questions about the nature of detection. For instance, the alignment from raw sequencing reads of the two different smRNA species surveyed in Chapter 4 is handled by two separate software solutions, each tailored exactly to the biological nature of the respective smRNA species. I used miRExpress, version 2,[139] to align miRNA sequences, and MINTmap[102] for the tRNA fragments (more specifically, only the fragments »exclusive to the tRNA space«). Procedurally, there is no argument or consensus against analysing these two species separately, and unifying results afterwards. The main effect of concatenation of the two count tables for joint analysis in DESeq2 differential expression analysis is a loss in sensitivity, because multiple testing has to be corrected in relation to the number of unique analytes. However, since the hypotheses assumed before analysis included modes of cooperation between the two distinct species, I decided to test them together rather than apart from each other. The inspection of MA plots (see e.g. Figure 3.6), effect size distribution, and the comparison to separate analyses for both species all indicated the joint approach to be feasible. The loss in specificity was mild in miRNAs, joint analysis reproduced 98.4% of detected and 83% of differentially expressed miRNAs (FDR < 0.05), and 98.3% of differentially expressed miRNAs with high effect size (absolute LFC > 1.4). For tRFs, the loss was greater; joint analysis reproduced 96.1% of detected, but only 20.5% of differentially expressed tRFs. This can be explained by a high number of very lowly expressed tRFs compared to miRNAs: tRFs with high differential expression effect size (LFC > 1.4) were reproduced at 52.1%; and reproduction in the top 5 percentile by count-change was 92.3%. Notably, this 95th percentile count-change cutoff value is at an absolute count-change of 28.5, meaning the loss of differentially expressed tRFs compared to separate analysis happened in the very low expression range, which is more desirable than it is a problem. An alternative solution would have been the truncation of low-count analytes before differential expression analysis; however, the threshold for such a step is always arbitrary, and thus truncation is no longer recommended.[144]

5.1.3 Statistical Analyses of Network Interactions

There is much less consensus when it comes to statistical interpretation of network analyses. This may in part be a result of network analyses being relatively uncommon compared to, for instance, sequencing experiments, and thus a lack of community consensus on how to approach certain problems. Often, a »network analysis« is a very confined, ultimate visualisation of the impact of one or few miRNAs with several genes that seem pertinent to the publication. Thus, there is often no need

to characterise the statistical relevance of the shown relationships, as they serve as a hypothetical, or an illustration of a proposed pathway. In contrast, most network analyses presented in this dissertation are intermediary steps, the results of which are supposed to aid in identification of pertinent factors in the molecular interactions studied. As a result, there is a need to measure the relevance of each component of the network, as well as the validity of its message as a whole. Those can be approached in different ways:

The validity of single network components is subject to various pre-existing properties. To make this discussion more tangible, I will give an example of the most common single component in my analyses: a miRNA→gene interaction. A first measure of its validity can be the existence of a validation experiment of this interaction. This is reflected in the scoring inside the database. The limitations as discussed in Section 5.1.1, e.g. in regard to tissue specificity, still apply. Below this highest level of stringency are the predicted interactions. As has been discussed previously by others, miRNA→gene relationship predictions are best seen in comparison to other models, and most valuable predictions are the ones that several prediction algorithms agree upon, particularly if these algorithms are based on modelling different aspects of miRNA→gene binding.[113] For this reason, I implemented the largest collection of algorithms I could find at the time (miRWalk 2.0[106]), supplemented by other sources as they became available, and also statistically evaluated the performance of all included algorithms to select a suitable subset for the summation of scores (Section 2.2.4). These steps were undertaken to minimise the risk of bias by using as many sources as possible, while still retaining an amount of flexibility in analysis. For instance, if the resulting gene network was too large for sensible analysis, its size could be easily decreased by elevation of the score threshold, thus making the analysis more stringent.

The scoring threshold as used for miRNA interactions is not available for tRFs, because there are no prediction datasets available yet. The prediction was performed in-house on miRNA-like seeds of each detected differentially expressed tRF, which brings two important limitations: it assumes a miRNA-like functionality of tRFs, disregarding other interactions principles that have been found (see Section 1.3.3); and it limits the prediction sources to one, TargetScan.[114] TargetScan was selected for its approach of measuring evolutionary conservation of putative target sites, a measure that is very valuable in the case of tRFs, because we know little about any other parameters that could be relevant for the targeting. Since tRNAs and their fragments have been part of mammalian cells for a long time, evolutionary conservation of target sites in the genetically flexible 3' UTRs are a significant measure of functionality.[116] Thus, for tRFs, the aggregation score of multiple algorithms was substituted with the conservation score generated by branch length (BL) and probability of conserved targeting (P_{CT}).[116]

However, these cautionary steps still cannot preclude any and all possible biases that may be inherent to prediction models, and thus, statistical analyses on the basis of this general procedure are desirable. The simplest, most practical, although computationally intensive solution to inherent database

biases is permutation (see Section 2.4.1). Briefly, a null distribution of the measured parameter (e.g., miRNA→gene interaction score) is generated by iterative analysis of randomly permuted datasets of the same size as the »real« dataset to be tested. The location of the »real« score inside this null distribution then gives the »extremity« of that real result, and thus the likelihood of the result being as extreme by chance, which equals FDR. However, this approach requires a defined set of analytes that present with a measurable attribute (e.g., multiple miRNAs targeting one gene, or a defined set of target genes for any one miRNA). An additional measure to ensure robustness of this approach is the iteration across a range of parameters, for instance, a sliding score cutoff. Results staying »significant« across a range of different cutoffs may be an indication of their robustness. However, for each level of iterations added, computation time also increases. Since permutation approaches usually require tens to hundreds of thousands of iterations, the computational requirements can be considerable.

The advantage of permutations, as opposed to the disadvantage of having to test a set of multiple entries, is its scalability. Thus, permutations can also be used to assess the validity of the entire network, for instance by random permutation of case-control status (applicable in patient scenarios, as in Chapter 4), or of another attribute (such as sex, as in Chapter 3).

5.1.4 Cholinergic Cellular Models: LA-N-2 and LA-N-5

The decision to use a human cellular model of cholinergic neuronal cells was driven by two main factors: I) *In vivo* experimentation, i.e., animal-based research, is not reliable as a model of complex human psychiatric diseases. II) Particular to RNA-based mechanisms, the difference between rodent and human genes is enormous, for instance in 3' UTRs that are not subject to the same evolutionary pressure as coding regions, which leads to low transferability of the application of hypothetical therapeutic oligonucleotides. Furthermore, so-called »3D cell cultures« or co-cultures of human cells of different types (e.g., neurons with astrocytes) are still not as developed as is necessary for stable experimentation, particularly in the case of cholinergic systems; thus, we opted for mono-culture of neuronal cell lines. Even in traditionally cholinergic research, the number of human neuronal models representative of actual cholinergic neurons is very low; the popular cell line SH-SY5Y had to be excluded after in-depth experimentation because it fails to express sensible amounts of the main cholinergic marker *CHAT* and the vesicular transporter *SLC18A3* even upon differentiation (own results).

LA-N cells, although their maintenance is not as straightforward as that of many of the workhorses of human cell culture, were the logical conclusion to the search for an adequate model. The main argument for their selection was their expression of *CHAT* and *SLC18A3* in their natural state as well as an impressive induction of these genes via stimulation by neurokines. This expression of *CHAT* is a pivotal factor in the studies of cholinergic neurons, because there is much confusion about what constitutes a cholinergic cell in the CNS, and about the properties of the different cholinergic populations in the different brain regions. By definition, cholinergic neurons must be able to syn-

thesise (*CHAT*) and release (*SLC18A3*) ACh, all other defining properties are optional (for instance, the defining property of basal forebrain neurons to receive the retrograde NGF signal via *NTRK1*). An additional benefit was the availability of LA-N cells of both sexes, which aided in studying sex differences in Lobentanzer *et al.* [1]

The decision for a mono-culture of immortalised human neuronal cells naturally introduced limitations common to all similar models: the cells are derived from tumour cells and do not resemble a physiological cell any more, otherwise they would not have lost senescence. All biological implications derived from these models have to be interpreted based on this central limitation. However, interpretation is based on the assumption that basic molecular processes, such as the control of mRNA by miRNAs, are still intact. Additionally, the cell communities generated *in vitro* that are the basis of bulk sequencing analyses are very homogeneous, more so than any tissue derived from a living multi-cellular organism. While this enables bulk sequencing without »dilution« by supporting CNS-derived cells, as would be the case in patient brain region bulk sequencing, it also introduces a »non-natural« homogeneity in the RNA derived from the lysis of these cells and therefore harbours the danger of exacerbated sensitivity in differential expression. This may lead to the false positive identification of differentially expressed smRNAs, particularly those with a low base expression or small changes. However, controlling of effect sizes (e.g. a cutoff for absolute LFC or analysis of count-change) is well suited to prevent many irrelevant false positives.

5.1.5 Stroke Patient Blood Samples

The main limitations in the sequencing of post-stroke patient blood are the lack in cell type specificity of RNA generation and the patient collective composition, which resulted in exclusion of females because of imbalancing and thus contains only male patients. Additionally, the controls by principle have to be external, since stroke is an unexpected event and thus there is virtually no possibility of attaining a blood samples of patients pre-stroke except in very specific clinical conditions. As a result, control samples are healthy volunteers, that have to be matched in a manner that reproduces the patient collective as closely as possible, which unfortunately often cannot be more specific than matching for sex and age. Strictly speaking, the results of analyses based on the differential expression profile thus can only be immediately applied to male patients. However, the tissue-specific analyses performed on small RNAs as well as on transcription factor interactions both are based on large subject collectives representative of males and females, and thus the $CD14^{+}$-related analyses, such as the FFL analysis in Section 4.5.4, can in large parts be applied to both sexes.

Section 4.3 exclusively deals with the shortcomings of sequencing whole blood instead of isolated cellular components. Whether this method of extrapolation serves the purpose of clarifying the tissues involved in smRNA stroke response remains to be clarified and is discussed below (Section **??**). However, translational approaches like this one can also aid in studying the transferability of whole blood results (which may be more common in the clinical setting) to tissue-specific analyses, which

often require complex and costly purification steps in acquiring the isolated cell populations.

5.1.6 Gene Ontology Analyses

Some of the approaches to dimensionality reduction presented here rely heavily on the analysis of enrichment of genes in ontological categories of biological processes. The primary purpose of using GO enrichment as a tool is to deal with the reality of not being able to know all functions of any given gene. Strictly speaking, here also apply the same limitations as to other datasets of curated information: the quality of results depends on the quality of annotation in the raw ontology collection. For some lesser known proteins, these annotations may well be far from complete, and thus result in presentation bias towards the interpreting scientist. Similarly, by interpreting the list of ontological terms yielded by an analysis, terms may be selected or discarded based on their perceived relevance to the topic of study; inadequate knowledge about all participating processes may lead to the dismissal of relevant terms, constituting confirmation bias. To try and avoid large amounts of this form of confirmation bias, GO terms were presented in comprehensive form or as curated form of all available data as much as possible (see for instance Figure 3.10 B and Section 4.5.4). Additionally, the visual display of GO terms as a t-SNE projection, where distance is based on the amount of shared genes between the terms (using R/gsoap [178]), can aid in identifying the underlying categories of processes, and their relationships to one another. This is aided by the weighing functionality of R/topGO analysis, [145] whereby less specific parent GO terms are dismissed in favour of the more specific child nodes in the DAG graph (see also Sections 2.4.2 and 3.5.3).

Ontology analyses of small RNA targeting of protein-coding transcripts, projection

5.1.7 Feedforward Loop Analyses

Feedforward loop analysis brings together most of the issues discussed above. I) It is based on the aggregation of several types of molecular targeting relationships and thus is subject to their individual limitations. II) It is applied to the results of differential expression in stroke patient blood, and thus is also influenced by sequencing-related issues. And III) The modules yielded by FFL stratification are then scrutinised with the help of gene ontological analysis to find gene collectives relevant to stroke.

Regarding I); the results of concatenation of these individual relationship types are unknown and difficult to measure. Every kind of influence on the validity of results is thinkable: the insecurities from each individual method might be additive, or even super-additive, making the end result more unreliable in consequence. Alternatively, the processes being firmly rooted in the biological reality of the cell might also have a corrective function on the end result, effectively acting as a filter that removes »illogical« circuits from the output, thus increasing its validity. There is no measure for answering this question as of now. However, the circuits and their ontological associations gathered in Chapter 4 make sense from a biological perspective, and, maybe more importantly, they make more

biological sense than an observation of only differentially expressed genes, or only of smRNAs. In short, although the evidence is circumstantial, the message gained from FFL analysis may be larger than the sum of its parts. A more detailed discussion of the individual findings is held in the individual module descriptions of Section 4.5.4 and Section 4.5.5.

Regarding II); sequencing-related issues need to be primarily controlled in the process of differential expression analysis. During feedforward loop analysis, results from differential expression are fixed, and thus cannot do much harm if they are used in a descriptive way, as is the case in my analyses. Should they be used beyond that, for instance, to weigh targeting relationships by the differential expression of their participating factors, more care has to be taken that the model applied makes sense. Although the approach is feasible in principle, the practical application in FFL analysis is not trivial. Generally, it makes sense to compare the expression levels of smRNAs and their target genes, for two main reasons:

One, any interaction probably only has practical relevance in the cell if the expression levels (or rather, the change in expression), is on the same scale for both smRNA and mRNA. Of note, the count-change measure introduced in Lobentanzer *et al.* [1] is much more suited to this assessment than the commonly used LFC, because the latter does not relate to expression levels at all. However, until the sequencing of small and large RNAs can be routinely done in the same experiment (i.e., on the same microfluidic chip), the comparison of base mean expression (or count-change) between small and large RNA will always be very approximate, because it is dependent on sequencing depth of the individual experiment.

And two, considering miRNA-like interactions, those interactions will be particularly interesting where the smRNA is regulated inversely to the target mRNA. However, this concerns only the theoretical interaction of two isolated partners, whereas the smRNA→gene interactions in live cells are layered and only the strongest single relationships will have a chance of prevailing against the regulatory »chaos« that is an actual cell. This alone precludes actual analysis of differential expression influence on smRNA targeting (apart from complex mathematical models which remain to be established), without even considering the third interaction partner, transcription factors. Those introduce another element of uncertainty: although we know more about their tissue specific activities through efforts such as Marbach *et al.* [103], we do not specifically know which transcription factor acts on a promoter or repressor towards which genes in which tissues. Mathematical structural models able to predict these interactions in a cellular, whole-genome context are desirable, but not yet a reality.

Regarding III); the main criticism of proceeding from comprehensive FFLs through modularisation to GO analysis is the arbitrary nature of network modularisation. Modularisation itself is a purely mathematical process of describing the interconnectedness of nodes in the network, implemented by Blondel *et al.* (»Fast unfolding of communities in large networks«).[190] The choice of resolution that yielded five communities in my analysis thus was largely arbitrary, selected in a way

that correlated module identity with the visual clusters created by force-directed organisation of the FFL network. However, since the main purpose of modularisation is a reduction in dimensionality that facilitates human understanding, every possible resolution that results in a manageable number of modules may be seen as »reasonable«. A standardisation of these procedures is nevertheless desirable and will be subject of further studies.

5.2. A Mechanistic Perspective of Transcriptional Interactions

This dissertation is aimed at elucidating epi-transcriptional processes surrounding the expression of cholinergic genes. To this end, a framework was developed to assess interactions between players on the field of RNA-related processes in mammalian cells. This framework was then applied to state-of-the-art measurements of RNA levels, i.e., RNA-sequencing. In the following paragraphs, an assessment will be held on the outcomes of my efforts in clarifying »Small RNA Dynamics in Cholinergic Systems«.

5.2.1 Analysis of Small RNA Dynamics via RNA-sequencing and Bioinformatics

Small RNAs and the mechanisms by which they control the expression of coding genes have fascinated researchers since their discovery around the turn of the millennium. Much of the pioneering work has been done on miRNAs, but with tRFs, a new class of regulatory small RNA is increasingly being investigated in physiological and pathological contexts. During the current, early phase of small RNA studies, research has often assumed a limited perspective; many publications study interaction between few partners, in most cases reducing focus to one miRNA and one targeted gene. However, during the initial phase of my work on transcriptional interactions, it quickly became apparent that an integrative perspective is crucial. First and foremost, this involves information on the coding genes that are the supposed targets of small RNA intervention, but also the workings of transcription factors, which shape the phenotype of the cell, and, relatedly, tissue specificity of all of the aforementioned processes.

As such, a comprehensive integrative model of smRNA interactions did not exist when I started to work on this dissertation. More recently, there have been developments of integrative databases which model miRNA→gene interaction, one of them also including tissue specificity and transcription factors. The most extensive efforts in my view are mirDIP, miRWalk 3.0, and miRNet (causing the name change of my own database). Thus, they will be briefly reviewed and compared to my own work in the following.

mirDIP 4.1[276] and miRWalk 3.0[281] are similar in their focus on miRNA→gene interactions. To this end, both collected and integrated third-party data into their database. Both offer public access through a browser-based interface and database downloads. In addition, mirDIP offers integration into development environments via Java, R, and Python APIs. The main difference between the two is their data aggregation approach. While the mirDIP team collected all resources available (75 different sources[276]), the miRWalk developers reduced their source count between versions 2.0 and 3.0, from 12 sources to 4.[281] Instead of combinatorial power, the authors of miRWalk 3.0 rely on a sin-

gle algorithm as core principle of miRNA→gene targeting, TarPmiR.[282] Briefly, TarPmiR utilises machine learning (random forest) to identify miRNA binding sites by characteristics learned from photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) sequencing results. A comparison to other prediction algorithms indicated superior performance in the authors' hands.[282] However, it also shows how incomplete these approaches still are: the average recall of TarPmiR in the initial publication was 0.543, and the precision was merely 0.181 (or 0.191, the numbers in the manuscript conflict), indicating a high number of false positives. In light of these numbers, reliance on any one algorithm still remains statistically inferior to the combination of predictions based on different modelling techniques.[113] In light of the reliability of prediction algorithms, which ranges from very low to medium at best, stringent assessment of statistical properties of these data collections is necessary. However, the authors of miRWalk 3.0 have not statistically evaluated the performance of their database in the most recent publication.[281]

At the other extreme, mirDIP 4.1 includes 30 publicly available sources, selected from a review of 75 sources, to yield a total amount of 150 million targeting predictions.[276] Of note, due to performance issues (space requirements and query times), the database only supports miRNA:gene interactions, without additional information such as mRNA binding site, and does not work with gene identifiers other than HUGO symbol (which may be ambiguous). For integration of all third-party datasets, the authors normalised confidence levels of predictions inside each dataset to yield a score between 0 and 1, and then ranked each prediction dataset based on a benchmarking procedure (using experimentally validated interactions). These ranks were then used to calculate the confidence of miRNA→gene relationships via an integrative scoring, similar to the score in *miRNeo*, but with the addition of a weight for each prediction dataset. The addition of a weight may be beneficial the more source datasets are used, to differentiate between different qualities of source material. What I did by excluding the two poor performance datasets (Section 2.2.4) equals a simplified weighing procedure (with score of 1 for included datasets and 0 for dropped). A comparison of *miRNeo* accuracy compared to the mirDIP 4.1 data using the benchmarking data will be interesting, but has not been performed yet. Once the mirDIP data is integrated in *miRNeo*, the benefit of the weighing procedure can be evaluated.

miRNeo is not designed to compete with these types of database; on the contrary, *miRNeo* relies on a combination of publicly available datasets to enable accurate prediction[113] and to be able to derive test statistics from comparison of different source materials. Rather, *miRNeo* is designed to be efficient in managing complex computations on RNA-based interactions so as to enable the study of complex relationships and biological mechanisms such as feedforward loops.

As such, miRNet[275] is closest in functionality to *miRNeo*, as it provides interaction data on transcription factors as well. Very recently, it seems to have been updated to version 2.0, which allows study of transcription factors and feedforward loops, but there has been no publication detailing the results as of yet (April 2020). Its main »advantage« (see below) over *miRNeo* for users is that it pro-

vides an easy-to-use web-based interface for analyses; its main downside is that it practically includes only two main miRNA targeting sources, TarBase and miRecords. miRNA:TF data were collected from a dedicated source, TransmiR, which includes only manually curated interactions, and thus likely underestimates the true interactions by orders of magnitude. Additionally, the new version of miRNet seems to diverge significantly from the original description,[275] and the only way of evaluating the database is by the very limited »About«-section on the webpage, which unfortunately features several inconsistencies. For instance, the authors state that »miRNA to TF interaction data were collected from TransmiR 2.0«, however, TransmiR is a TF→miRNA database, which is critically and fundamentally different in its implications. In addition, a curated database such as TransmiR cannot currently be designated comprehensive, by their own description it includes »2,852 TF-miRNA regulations from 1,045 publications«, and very limited tissue-specific information. However, in the miRNet web application, tissues can be selected for TF interactions, regardless of target type. How that is possible is not explained. In summary, while the idea behind miRNet may be similar to *miRNeo*, function and performance cannot currently be assessed without additional information on what exactly miRNet does, and what data it is based on. It may even be dangerous to present researchers with such an easily accessible tool, a »black box«, which from the input of only a few gene names generates complex analyses without requiring any understanding from the researcher performing the analyses. Maybe even more questionable is the lack of transparency as to how the results are generated.

In summary, *miRNeo* as an integrative approach to small RNA dynamics is a valuable addition to the repertoire of the study of transcriptional interactions. It collects several resources for targeting of genes by small RNAs, which have been statistically evaluated as to their performance; it integrates this targeting data with tissue-specific TF→gene targeting information from 394 human tissues, based on the FANTOM5 dataset; and it provides, through its graph-based infrastructure, high computational performance for the assessment of complex relationships in RNA interaction. From these vantage points, it is the most complete integrative transcriptional interaction database to date. Its main limitation in this context is the limited availability due to the lack of a web-interface or public R package, and the high level of knowledge it requires for usage.

5.2.2 The Cholinergic/Neurokine Interface

Multiple lines of orthogonal evidence confirm the significance of neurokines for cholinergic processes, and imply a cooperation between cholinergic and neurokine systems in health as well as in disease. Earliest descriptions of neurokines, in the late 1980s, have tied them to cholinergic differentiation, which was the reason for adopting the LA-N cell models for experimental work in this dissertation. The bioinformatic analyses of these experiments identified two miRNA families, mir-10 and mir-199, to inhabit a pivotal role in interfacing between cholinergic and neurokine genes, and transcriptomic analyses of single cell data from murine and human CNS demonstrate a co-expression of cholinergic

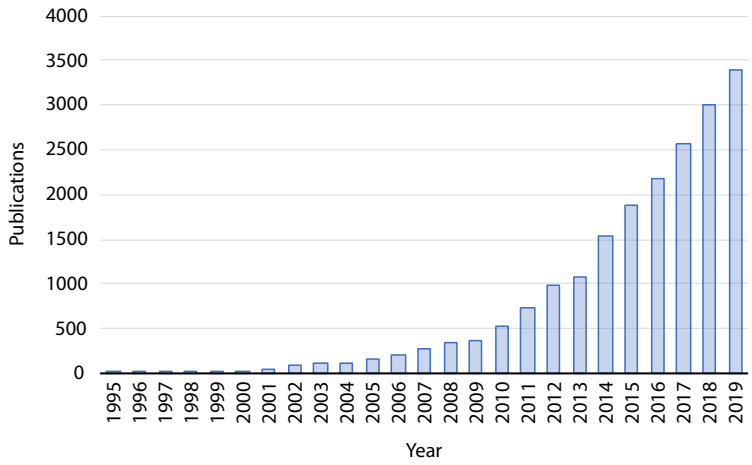

Figure 5.1: Number of Publications on Neuroinflammation by Year. Data were downloaded from https://www.ncbi.nlm.nih.gov/pubmed on the first of April 2020.

markers and neurokine receptors.[1] Thus, an assessment of cholinergic neuron functionality, be it in health or in neurological disease, has to take into account these para- and endocrine influences, particularly from the neurokine and neurotrophic factor families.

More generally, it is very likely that most neuroscientific endeavours would benefit from integrating other aspects of life-science, in particular, endocrinology and immunology. As recent literature shows, many classifications of diseases originally thought neurologic are currently being revised, often resulting in inclusion of immunological aspects; first and foremost, the term »neuroinflammation« has seen a rise in popularity by 806% in the last decade (PubMed search results of publications between 1900 and 2010: 2414, between 2010 and 2020: 19465, Figure 5.1). Consequently, the scientific community has much to gain from cooperation between the neurologic and immunologic branches of research.

Importantly, the interaction between neurokine and cholinergic systems is not unidirectional; both systems control manifold properties of the mammalian body, and thus, communication between the two systems takes place in various ways, cell types, and organs of the body. Arguably, the most immediate form of this communication is the elicitation of cholinergic properties in neuronal cells by neurokine signals. It has been shown by myself and others that isolated neuronal cells express more choline acetyltransferase (CHAT) and vesicular acetylcholine transporter (SLC18A3) upon neurokine stimulation, and stimulation by leukaemia inhibitory factor (LIF) results in catecholaminergic→cholinergic transformation of sympathetic neurons *in vivo*. In theory, this type of interaction between neurokine and cholinergic systems requires only one type of cell (if the neuron in question were able to synthesise and release a neurokine); however, *in vivo*, it likely involves at least two types of cells: the neuron receptive of the neurokine signal, and a regulatory cell which releases the neurokine. While the regulatory cell types releasing IL-6, the most studied neurokine by far, are already well described, the cellular sources of the other, lesser-studied neurokines are still enigmatic, particularly in the CNS. Similarly, the differences in effect on the stimulated cells by the different neurokines, which

in all likelihood are rather subtle, have not been studied as of yet. By the rather unique combination of soluble and membrane-bound receptors, which cooperate in a fashion unique to each individual member of the gp130 family, neurokines present a tremendously complex regulatory mechanism.

Conversely, cholinergic systems can influence neurokines, however, this side of the interaction is much less clear and subject to considerable controversy. While the definition of »cholinergic« is relatively simple in the transcriptional context, a clear definition of neurokine tissues is all but impossible. A cholinergic neuron by definition is characterised by its expression of CHAT and SLC18A3, without which it would not be able to transmit a cholinergic signal; in the case of non-neuronal cholinergic systems, it is admittedly more complex. In neurokine systems, however, most cell types that may be considered as candidates fulfil a wide range of functions, using multiple and diverse messenger molecules; mainly, this involves tissues of the immune system. As such, the »target cell« of cholinergic→neurokine interaction may be different depending on the context, which complicates the analysis of clinical and experimental data. The most prominent instance of cholinergic→neurokine interaction is the so-called cholinergic anti-inflammatory reflex, coined and publicised by the work of Kevin Tracey.[?] While Tracey's work is not specifically aimed at influences on neurokine systems, the anti-inflammatory properties of vagal activation extend to neurokines, as can be seen by the suppression of IL-6-mediated effects of LPS by vagal activation.[?] However, while there has been proof of the anti-inflammatory effect of vagal activation, its mechanism still is a matter of debate. Since ACh is rapidly degraded by circulating esterases, an endocrine functionality is out of the question. However, the spleen, as the major organ target of the immunosuppression, is not parasympathetically innervated (or very sparsely, as some results suggest). The current working hypothesis involves a participation of sympathetic mediation of the vagal signal through the splenic nerve, where it activates β2 adrenergic receptors on $ChAT^+$ T cells, which in turn release the ACh required for cholinergic suppression of regulatory T cells. Influences on immune cells by direct cholinergic signalling via ACh are further complicated by the availability of different receptor types and subtypes. For instance, the activation of homopentameric $\alpha 7$ nicotinic receptors causes a suppression of inflammatory processes; incidentally, this can also happen in part due to activation of JAK2/STAT3 activation.[?] Conversely, an activation of muscarinic receptors often is associated with immune stimulation.[?]

check

The identification of the two interfacing families, mir-10 and mir-199, adds another piece of circumstantial evidence to the complex picture of cholinergic/neurokine interaction (Figure 5.2).

Hypothesis: cholinergic and neurokine systems intermingle significantly in the cns, affecting physiological as well as pathogenic (pathologic?) processes. Multiple angles reject null (orthogonal evidence), application to disease

A pivotal mediator of cholinergic/neurokine interaction, and of neurokine function in general, is the JAK/STAT pathway, which is immediately tied to gp130 activation. Importantly, JAK/STAT signalling is exclusive to neither cholinergic nor neurokine processes, but is critically important in both.

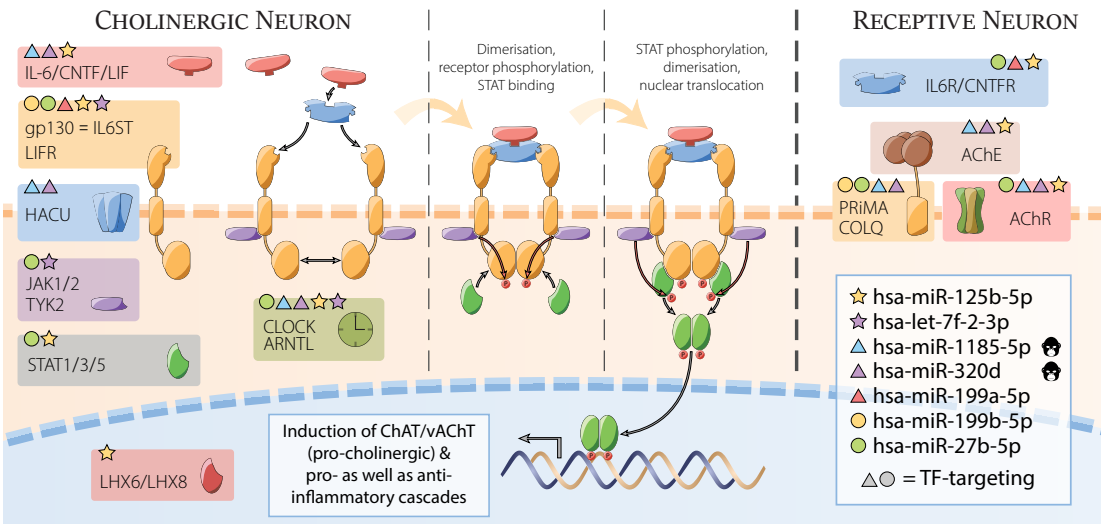

Figure 5.2: The Cholinergic/Neurokine miRNA Interface. The neurokines, such as CNTF, LIF, and IL-6, signal through a combination of soluble and membrane-bound receptors. Activation of a transmembrane neurokine receptor is usually followed by JAK recruitment and phosphorylation, and successively by STAT activation and translocation to the nucleus. Gp130-family neurokine, cholinergic, and circadian signalling pathways are controlled by primate-specific and evolutionarily conserved miRNAs. miRNA targeting of individual genes (indicated by coloured symbols) yields complex transcriptional interactions. Several miRNAs directly targeting the cholinergic pathway also target TFs controlling this pathway (circles and triangles). miR-125 and miR-199 are members of the mir-10 family.

Neurokines can activate tyrosine kinases JAK1/2 and TYK2, as well as STATs 1, 3, 5A, and 5B.[69] Which of these leads to pro-cholinergic differentiation of neurons is still unclear and an interesting topic for further research. Some work has been done on the distinction of effects between different members of the JAK and STAT families, mainly in immune cells. For instance, STAT1 activation in phagocytic monocytes leads to differentiation towards the M1 type pro-inflammatory macrophage, while STAT3 activation favours the generation of anti-inflammatory M2 type macrophages.[?] The broad expression and wide-reaching functionalities of the JAK/STAT pathway bring with them an important caveat of all matters they were implicated in in the course of this dissertation, particularly in regard to gene ontology analyses: due to their importance in many processes in mammalian cells, they may be overrepresented in annotation, for instance in the ontology catalogues, such that these harbour an implicit bias for finding associations to JAK/STAT-related processes. Although there are measures in place in the analysis process that are supposed to suppress false identification, e.g. the weighing in R/topGO, there is no way of guaranteeing the absence of false positives in these results. However, since JAK/STAT mechanisms were implied with high frequency and in various independent analyses, there is a high level of confidence in their relevance to the studied phenomena.

Neurokines are implicated in a range of diseases that have previously been associated with cholinergic systems and their dysfunction, particularly in the context of neuroinflammation, and particularly regarding IL-6. Whether this is a result of research bias, or of IL-6 actually being more relevant for disease processes, cannot at the moment be determined. AD, SCZ, BD, stroke,

LAN as model system

co-expression in single cell

cell model, chat anomaly, regulation of expression of these two, induction, low vs high control genes

broadly acting vs specific mir families

5.2.3 Molecular Biology of Feedforward Loops

Small RNA feedforward loops are a mechanistically feasible epigenetic controller of transcription, and the existence of biologically relevant FFLs has been convincingly shown. However, this evidence is still anecdotal, and thus, quantitative estimations of the extent of this phenomenon cannot with certainty be made. Hypothetically, feedforward loops affect a significant portion of all miRNA→gene relationships, as can be seen in the FFL module analysis in Section 4.5. Intriguingly, tRF→gene feedforward loops (with a miRNA-like mechanism) are predicted in significantly smaller numbers. The stroke-relevant genes identified in Section 4.2.1 are involved in 3.5% of miRNA FFLs (681 FFLs) and 11% of tRF FFLs (21 FFLs). Thus, the low number of identified tRF-FFLs in stroke is not a stroke-specific observation, but rather a consequence of a low number of tRF-FFLs overall. Whether this is a result of the still inaccurate prediction or a real difference between these two smRNA species cannot be answered by my analyses. It is, however, an interesting question for future research.

Hypothetically, if the low number of tRF-FFLs is not an artefact, but rather a representation of real epigenetic state, the question then arises, »What may the reason for this discrepancy be?« Generally observed, tRF→gene interaction is present and not significantly less so than miRNA→gene interaction, for instance in the FFL module analysis in Section 4.5. For comparison, the network that is the basis for FFL analysis in Figure 4.9 contains 481 miRNAs and 344 tRFs, but 681 miRNA-FFLs and only 21 tRF-FFLs. This discrepancy may carry biological significance, and two possible explanations come to mind: 1) tRFs may preferentially target genes that represent the ultimate stage of gene expression, and show less direct tRF→TF interaction; or 2) tRFs show tRF→gene and tRF→TF interactions in similar extent as miRNAs, but the target sets do not overlap, i.e., targeted non-TF transcripts and TF transcripts do not form meaningful FFLs in the case of tRF targeting.

To determine the most likely answer based on my data, I calculated the ratio of smRNA→TF to smRNA→gene (excluding TFs) interactions for both smRNA species in the raw FFL network data of Figure 4.9. The ratio was similar in both species, around 10% (miRNAs: 10.95%, 6491 / 59 269; tRFs: 9.37%, 17 542 / 187 124), indicating that assumption #1 may not hold. It follows that miRNAs and tRFs target TFs and non-TF genes in comparable amounts, but that the targeting in the case of miRNAs shows significantly more overlap between TFs and non-TF genes, leading to significantly more FFLs, in agreement with hypothesis #2. This argument is only strengthened by the fact that the absolute number of tRF→gene interactions was significantly higher as compared to miRNA→gene (as a result of the score-based thresholding procedure in miRNA analysis).

Another aspect of FFL theory is the coherence of loops. While there are feasible roles for coherent

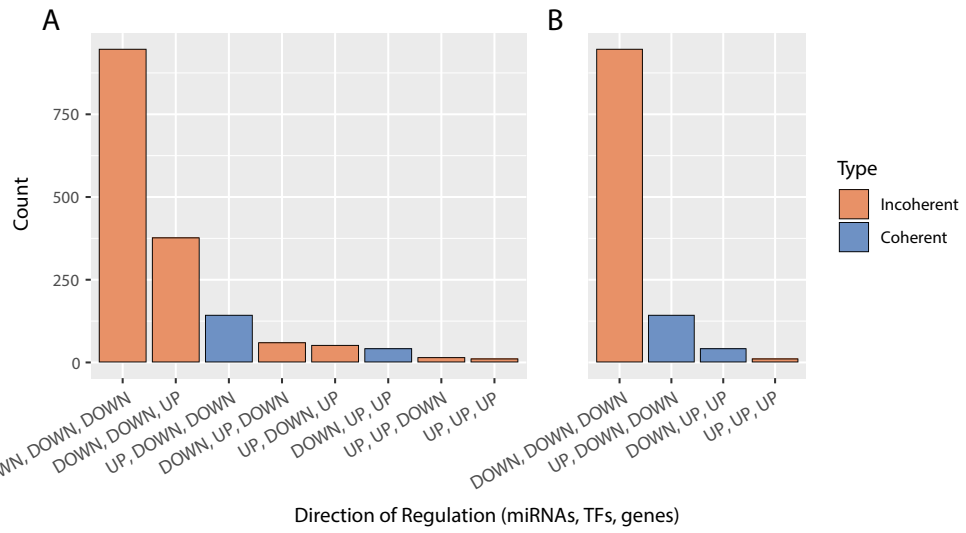

Figure 5.3: Coherence of miRNA Feedforward Loops. Individual FFLs were classified based on the direction of regulation of each of their components (miRNAs, TFs, and non-TF genes) in the blood of stroke patients, as determined via RNA-seq. Barplots represent the count of each class of FFL, colour denotes coherence. Incoherent FFLs dominate quantitatively. **A)** Barplot of all possible types of FFLs. **B)** Barplots of FFLs only with coherent TF→gene relationships (both either up- or down-regulated).

as well as incoherent smRNA:TF:gene loops, their implications may diverge depending on the cellular context. Just looking at summary statistics, there is an implied difference between the two smRNA species: while miRNAs in the majority are down-regulated, tRFs in the majority are up-regulated (see Figure 4.3). Combined with the preferential down-regulation of mRNA and the putative antagonistic role of both smRNA species, the general role of tRFs agrees with coherent FFLs, while the general role of miRNAs seems incoherent. Computing coherence on an individual FFL level indeed shows a high number of incoherent FFLs among all miRNA FFLs, mainly of the type »all down-regulated« (Figure 5.3). Likewise, all 21 detected tRF FFLs were of the type »incoherent«. However, due to the very low number of tRF FFLs, this finding in all likelihood is not representative. The few detected tRF FFLs may also be false positives.

Regardless of their individual biological significance, FFLs can be used as a tool to gain insight on transcriptional processes. FFLs may identify tightly connected processes, and allow stratification of large data, which is one of the main problems in descriptive bioinformatics. The approach shown in Section 4.5 is an attempt at dimensionality reduction that retains as much of the original data structure as possible, while allowing human interpretation. As is demonstrated by the comparison between GO analyses on the whole set of data and the individual clusters as defined by FFL analysis (Figure 4.10), the latter allows deeper insight into the biological processes affected by stroke through its increased resolution.

In this manner, the most pathologically pertinent coding transcripts in the blood of stroke victims were identified and classified, as well as their associations with biological processes involved in pathogenesis of or response to the infarction. The main biological pathways identified in this context are in-

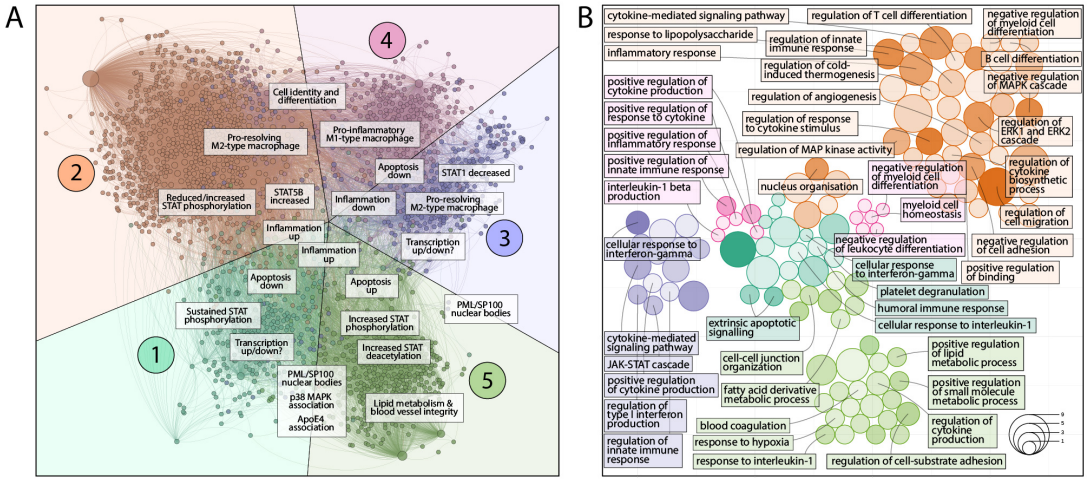

Figure 5.4: Comparison of Annotated Feedforward Loop Network and t-SNE Visualisation of Module GO Terms. A) Reproduction of Figure 4.11, colours have been adapted to allow module comparison between **A** and **B**. Displayed is the network of all $CD14^{+}$-specific FFLs, stratified into five modules, and overlaid with module-specific GO analysis results. **B)** Reproduction of Figure 4.10 B. Displayed is the t-SNE-based visualisation of all module-specific GO terms from the $CD14^{+}$ FFL-network (**A**), coloured by module. The distance between nodes is based on amount of shared genes between terms, depth of colour represents significance level, size of node represents number of genes in term.

volved in immunity and inflammation, cell death, regulation of transcription, STAT signalling, lipid metabolism, and blood vessel integrity. A comparison between annotated FFL modules and t-SNE visualisation of module GO terms summarises the implications drawn from this bioinformatically supported clinical study (Figure 5.4). Of note, the number of molecules comprising each module correlates with the number of GO terms identified for the respective module, as seen by the comparison between the largest module (two) and the smallest modules (three and four). This can be explained by the comparison of the number of DE genes in each module and the absolute number of genes making up any GO term (i.e., the *successes*). Hypergeometric enrichment p-values are dependent on the size of the set of successes, and thus, the likelihood of identifying a large (i.e., less specific) GO term with a comparatively small number of test set genes is very low, which is not the case for smaller, highly specific GO terms. Consequently, larger modules (test sets) have available to them a larger number of GO terms that can potentially be enriched.

Comparing the topography between Figure 5.4 A and B, the location of most modules appears as a mirror image. This could be interpreted as a confirmation of the general feasibility of the approach: the similarity of transcripts as determined by their participation in closely related FFLs is paralleled by the similarity of module GO terms as determined by the genes shared between the GO terms. However, a significant difference between the two visualisations seems the central position of module one in Figure 5.4 B, which may indicate a central relevance of module one transcripts in the studied processes. Indeed, module one GO terms appear to function as a nucleation point for related terms from the other modules (central cluster of Figure 5.4 B), which may be used as an indicator of a focus point for further studies.

There are several lines of investigation that could be based on the present results: 1) As described

above, the cluster surrounding module one GO terms could be dissected as to the implications of the genes relevant to these terms. These genes may represent a »cooperative set«, which mediates between the distinct modules and their influences on the biological processes in question. As such, it may be interesting to »zoom in« into the network of only those genes defining this central cluster, to identify pathways sitting at the epicentre of transcriptional response to stroke.

2) Feedforward loops as an abstract classification method may be helpful in the differentiation of inductive versus repressive behaviours of single TF→gene relationships. As discussed several times in the course of this dissertation, the current comprehensive data on TF→gene interaction does not allow prediction of the direction of regulatory influence of the TF over the gene. Even in manually collected data, such as TRANSFAC, the interaction of TF and gene is often described in a »yes/no« fashion, with the added limitation of tissue-specific information that is not easily transferred. This is mainly owed to the fact that most TF:gene interactions are found via binding assays such as chromatin immunoprecipitation (ChIP) sequencing and related variants. The combination of smRNA:TF:gene FFLs with interventional experiments (i.e., yielding regulatory output) may serve as methodical support to determine the direction of regulation in individual TF→gene interactions. The application of FFLs (from prediction or web-available datasets) to experiments (also from web-available datasets) may aid in detecting regulatory circuits, and a meta-analysis of these circuits across multiple different experiments may be used to calculate likelihood data of positive or negative regulatory interaction between any TF and gene. Such an approach may be a cost-effective data-mining alternative to painstaking single-experiment molecular biology.

3) Similarly, FFLs can aid in the classification of small RNA species and their families and subfamilies in a functional manner. The participation in FFLs from a comprehensive dataset can be mathematically transformed into a similarity- or distance-matrix, and the information so gained on relationships among smRNAs can be used for the stratification and analysis of relationships between individual smRNAs. This classification can serve as an independent comparative dataset, complementing the traditional strata derived from phylogenetic studies.

4) There is need for the development of statistical frameworks in the analysis of feedforward loops. In particular, a measure for the relative importance of each FFL in a network would be helpful for the dissemination of the network and its functions. Possible components of a mathematical description of significance in this setting may be the differential expression of FFL components, the strength of the interaction between FFL components, or network-specific parameters such as centrality. A formal definition of such an »importance measure« is not a trivial task and will require extensive comparison and validation.

5.3. Small RNA Therapeutics and Pharmacology

Extant approaches, methods, diseases, PCSK9, asthma, using small RNA antisense as substitute for single-target small molecules, reduce off-target effects, side effects of a different kind, what are off-target-effects in miRNA therapy?

Transcriptomics as basis for selection and design of antisense therapy, combinatorial, compare dirty drugs from psychiatric disorders, serendipity impossible, determinant is the sequence as opposed to functional groups that can be iteratively modified (only 4 building blocks)

elusive small molecular drugs, e.g. for IRF5[283]

immunity: sustained central inflammation and delayed resolution, cholinergic signalling in immune cells, neurokines, TH1 type cytokines, cell exchange with CNS

stroke: monocytes from acute and subacute to chronic - cross differentiation, migration, controlled by ?

Monocytosis caused by either congenital ApoE KO or repeated LPS injections led to an imbalance between pro-inflammatory and wound healing monocyte types, in which increased pro-inflammatory monocytes hampered the transition to an anti-inflammatory healing microenvironment, in a murine model of myocardial infarction.[284]

many transcription factors show highly context-dependent actions, eg MAF family[245]

sexual differences

Bibliography

[1] Lobentanzer S, Hanin G, Klein J, & Soreq H. Integrative Transcriptomics Reveals Sexually Dimorphic Control of the Cholinergic/Neurokine Interface in Schizophrenia and Bipolar Disorder. *CellReports*, pp. 1–19 (2019). doi:10.1016/j.celrep.2019.09.017.
URL https://doi.org/10.1016/j.celrep.2019.09.017

[2] Dale H H. THE ACTION OF CERTAIN ESTERS AND ETHERS OF CHOLINE, AND THEIR RELATION TO MUSCARINE. *Journal of Pharmacology and Experimental Therapeutics*, 6(2) (1914).

[3] Loewi O. Über humorale Übertragbarkeit der Herznervenwirkung. *Pflügers Arch. Ges. Physiol.*, 189:239–242 (1921).

[4] Dale H H & Dudley H W. THE PRESENCE OF HISTAMINE AND ACETYLCHOLINE IN THE SPLEEN OF THE OX AND THE HORSE. *J. Physiol.*, 68:97 (1929).
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1402860/pdf/jphysiol01676-0019.pdf

[5] Mesulam M M, Mufson E J, Levey A I, & Wainer B H. Atlas of cholinergic neurons in the forebrain and upper brainstem of the macaque based on monoclonal choline acetyltransferase immunohistochemistry and acetylcholinesterase histochemistry. *Neuroscience*, 12(3):669–686 (1984). doi:10.1016/0306-4522(84)90163-5.

[6] Mesulam M M & Geula C. Nucleus basalis (Ch4) and cortical cholinergic innervation in the human brain: Observations based on the distribution of acetylcholinesterase and choline acetyltransferase. *The Journal of Comparative Neurology*, 275(2):216–240 (1988). doi:10.1002/cne.902750205.
URL http://www.ncbi.nlm.nih.gov/pubmed/3220975; http://doi.wiley.com/10.1002/cne.902750205

[7] Mesulam M M & Van Hoesen G W. Acetylcholinesterase-rich projections from the basal forebrain of the rhesus monkey to neocortex. *Brain Research*, 109(1):152–157 (1976). doi:10.1016/0006-8993(76)90385-1.
URL https://www.sciencedirect.com/science/article/abs/pii/0006899376903851?via{%}3Dihub; https://linkinghub.elsevier.com/retrieve/pii/0006899376903851

[8] Berrios G E. Alzheimer's disease: A conceptual history. *International Journal of Geriatric Psychiatry*, 5(6):355–365 (1990). doi:10.1002/gps.930050603.
URL http://doi.wiley.com/10.1002/gps.930050603

[9] Miech R A, Breitner J C, Zandi P P, Khachaturian A S, Anthony J C, & Mayer L. Incidence of AD may decline in the early 90s for men, later for women: The Cache County study. *Neurology*, 58(2):209–218 (2002). doi:10.1212/WNL.58.2.209.

[10] Bartus R, Dean R, Beer B, & Lippa A. The cholinergic hypothesis of geriatric memory dysfunction. *Science*, 217(4558):408–414 (1982). doi:10.1126/science.7046051.
URL http://www.sciencemag.org/cgi/doi/10.1126/science.7046051

[11] Braak H & Braak E. Staging of alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, 16(3):271–278 (1995). doi:10.1016/0197-4580(95)00021-6.
URL https://www.sciencedirect.com/science/article/abs/pii/0197458095000216?via{%}3Dihub

[12] Schmitz T W & Nathan Spreng R. Basal forebrain degeneration precedes and predicts the cortical spread of Alzheimer's pathology. *Nature Communications*, 7:1–13 (2016). doi:10.1038/ncomms13249.

[13] Schmitz T W, Mur M, Aghourian M, Bedard M A, & Spreng R N. Longitudinal Alzheimer's Degeneration Reflects the Spatial Topography of Cholinergic Basal Forebrain Projections. *Cell Reports*, 24(1):38–46 (2018). doi:10.1016/j.celrep.2018.06.001.
URL https://doi.org/10.1016/j.celrep.2018.06.001

[14] Kraepelin E. *Psychiatrie. Ein Lehrbuch für Studirende und Aerzte.* Leipzig Barth, edition 8, edition (1913).

[15] Bortolato B, Miskowiak K, Vieta E, Köhler C, & Carvalho A F. Cognitive dysfunction in bipolar disorder and schizophrenia: a systematic review of meta-analyses. *Neuropsychiatric Disease and Treatment*, 11:3111 (2015). doi:10.2147/NDT.S76700.
URL https://www.dovepress.com/cognitive-dysfunction-in-bipolar-disorder-and-schizophrenia-a-systemat-peer-re

[16] Van Enkhuizen J, Janowsky D S, Olivier B, Minassian A, Perry W, Young J W, & Geyer M A. The catecholaminergic-cholinergic balance hypothesis of bipolar disorder revisited. *European Journal of Pharmacology*, 753:114–126 (2015). doi:10.1016/j.ejphar.2014.05.063.
URL http://dx.doi.org/10.1016/j.ejphar.2014.05.063

[17] Smucny J & Tregellas J R. Targeting neuronal dysfunction in schizophrenia with nicotine: Evidence from neurophysiology to neuroimaging. *Journal of Psychopharmacology*, 31(7):801–811 (2017). doi:10.1177/0269881117705071.

[18] Gray S L, Anderson M L, Dublin S, Hanlon J T, Hubbard R, Walker R, Yu O, Crane P K, & Larson E B. Cumulative Use of Strong Anticholinergics and Incident Dementia. *JAMA Internal Medicine*, 175(3):401 (2015). doi:10.1001/jamainternmed.2014.7663.
URL http://archinte.jamanetwork.com/article.aspx?doi=10.1001/jamainternmed.2014.7663

[19] Eum S, Hill S K, Rubin L H *et al.* Cognitive burden of anticholinergic medications in psychotic disorders. *Schizophrenia Research*, 190:129–135 (2017). doi:10.1016/j.schres.2017.03.034.
URL http://www.ncbi.nlm.nih.gov/pubmed/28390849; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5628100; http://linkinghub.elsevier.com/retrieve/pii/S0920996417301718; https://linkinghub.elsevier.com/retrieve/pii/S0920996417301718

[20] Koukouli F, Rooy M, Tziotis D *et al.* Nicotine reverses hypofrontality in animal models of addiction and schizophrenia. *Nature Medicine*, 23(3):347–354 (2017). doi:10.1038/nm.4274.
URL http://dx.doi.org/10.1038/nm.4274

[21] Forget B, Scholze P, Langa F, Mourot A, Faure P, & Maskos U. Article A Human Polymorphism in CHRNA5 Is Linked to Relapse to Nicotine Seeking in Transgenic Rats Article A Human Polymorphism in CHRNA5 Is Linked to Relapse to Nicotine Seeking in Transgenic Rats. *Current Biology*, pp. 1–10 (2018). doi:10.1016/j.cub.2018.08.044.
URL https://doi.org/10.1016/j.cub.2018.08.044

[22] Sacco K A, Bannon K L, & George T P. Nicotinic receptor mechanisms and cognition in normal states and neuropsychiatric disorders. *Journal of Psychopharmacology*, 18(4):457–474 (2004). doi:10.1177/0269881104018004003.
URL http://www.ncbi.nlm.nih.gov/pubmed/15582913; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1201375; http://journals.sagepub.com/doi/10.1177/0269881104018004003

[23] Rowe A R, Mercer L, Casetti V, Sendt K V, Giaroli G, Shergill S S, & Tracy D K. Dementia praecox redux: A systematic review of the nicotinic receptor as a target for cognitive symptoms of schizophrenia. *Journal of Psychopharmacology*, 29(2):197–211 (2015). doi:10.1177/0269881114564096.
URL http://journals.sagepub.com/doi/10.1177/0269881114564096

[24] Lewis A S, van Schalkwyk G I, & Bloch M H. Alpha-7 nicotinic agonists for cognitive deficits in neuropsychiatric disorders: A translational meta-analysis of rodent and human studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 75:45–53 (2017). doi:10.1016/j.pnpbp.2017.01.001.

URL http://www.ncbi.nlm.nih.gov/pubmed/28065843; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5446073; https://linkinghub.elsevier.com/retrieve/pii/S0278584616304353

[25] Higley M J & Picciotto M R. Neuromodulation by acetylcholine: Examples from schizophrenia and depression. *Current Opinion in Neurobiology*, 29:88–95 (2014). doi:10.1016/j.conb.2014.06.004.
URL http://dx.doi.org/10.1016/j.conb.2014.06.004

[26] Değirmenci Y & Keçeci H. Visual Hallucinations Due to Rivastigmine Transdermal Patch Application in Alzheimer's Disease; The First Case Report. *International Journal of Gerontology*, 10(4):240–241 (2016). doi:10.1016/j.ijge.2015.10.010.

[27] Leger M & Neill J C. A systematic review comparing sex differences in cognitive function in schizophrenia and in rodent models for schizophrenia, implications for improved therapeutic strategies. *Neuroscience & Biobehavioral Reviews*, 68:979–1000 (2016). doi:10.1016/j.neubiorev.2016.06.029.
URL http://www.ncbi.nlm.nih.gov/pubmed/27344000; http://linkinghub.elsevier.com/retrieve/pii/S0149763415302712

[28] de Leon J & Diaz F J. A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophrenia Research*, 76(2-3):135–157 (2005). doi:10.1016/j.schres.2005.02.010.
URL https://linkinghub.elsevier.com/retrieve/pii/S0920996405000757

[29] Berger M, (Editor) *Psychische Erkrankungen*. Berger, Mathias (2014).

[30] Bekker M H & van Mens-Verhulst J. Anxiety Disorders: Sex Differences in Prevalence, Degree, and Background, But Gender-Neutral Treatment. *Gender Medicine*, 4:S178–S193 (2007). doi:10.1016/S1550-8579(07)80057-X.
URL https://linkinghub.elsevier.com/retrieve/pii/S155085790780057X

[31] Anttila V, Bulik-Sullivan B, Finucane H K *et al.* Analysis of shared heritability in common disorders of the brain. *Science*, 360(6395):eaap8757 (2018). doi:10.1126/science.aap8757.
URL http://www.sciencemag.org/lookup/doi/10.1126/science.aap8757

[32] Gandal M J, Haney J R, Parikshak N N *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, 359(6376):693–697 (2018). doi:10.1126/science.aad6469.
URL http://science.sciencemag.org/content/359/6376/693; http://www.sciencemag.org/lookup/doi/10.1126/science.aad6469

[33] Ruderfer D M, Ripke S, McQuillin A *et al.* Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell*, 173(7):1705–1715.e16 (2018). doi:10.1016/j.cell.2018.05.046.
URL https://linkinghub.elsevier.com/retrieve/pii/S0092867418306585

[34] Harrison P J. Recent genetic findings in schizophrenia and their therapeutic relevance. *Journal of Psychopharmacology*, 29(2):85–96 (2015). doi:10.1177/0269881114553647.
URL http://www.ncbi.nlm.nih.gov/pubmed/25315827; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4361495; http://journals.sagepub.com/doi/10.1177/0269881114553647

[35] Henriksen M G, Nordgaard J, & Jansson L B. Genetics of Schizophrenia: Overview of Methods, Findings and Limitations. *Frontiers in Human Neuroscience*, 11:322 (2017). doi:10.3389/fnhum.2017.00322.
URL http://journal.frontiersin.org/article/10.3389/fnhum.2017.00322/full

[36] Kanazawa T, Bousman C A, Liu C, & Everall I P. Schizophrenia genetics in the genome-wide era: a review of Japanese studies. *npj Schizophrenia*, 3(1):27 (2017). doi:10.1038/s41537-017-0028-2.
URL http://www.nature.com/articles/s41537-017-0028-2

[37] Malhi G S, Tanious M, Das P, Coulston C M, & Berk M. Potential Mechanisms of Action of Lithium in Bipolar Disorder. *CNS Drugs*, 27(2):135–153 (2013). doi:10.1007/s40263-013-0039-0.
URL http://link.springer.com/10.1007/s40263-013-0039-0

[38] Fujii T, Mashimo M, Moriwaki Y, Misawa H, Ono S, Horiguchi K, & Kawashima K. Physiological functions of the cholinergic system in immune cells. *Journal of Pharmacological Sciences*, 134(1):1–21 (2017). doi:10.1016/j.jphs.2017.05.002.
URL https://linkinghub.elsevier.com/retrieve/pii/S1347861317300695

[39] Pavlov V A & Tracey K J. Neural regulation of immunity: Molecular mechanisms and clinical translation. *Nature Neuroscience*, 20(2):156–166 (2017). doi:10.1038/nn.4477.

[40] Dantzer R. Neuroimmune interactions: From the brain to the immune system and vice versa. *Physiological Reviews*, 98(1):477–504 (2018). doi:10.1152/physrev.00039.2016.

[41] Lurie D I. An Integrative Approach to Neuroinflammation in Psychiatric disorders and Neuropathic Pain. *Journal of Experimental Neuroscience*, 12:117906951879363 (2018). doi:10.1177/1179069518793639.
URL http://journals.sagepub.com/doi/10.1177/1179069518793639

[42] Fullerton J N & Gilroy D W. Resolution of inflammation: A new therapeutic frontier. *Nature Reviews Drug Discovery*, 15(8):551–567 (2016). doi:10.1038/nrd.2016.39.
URL http://dx.doi.org/10.1038/nrd.2016.39

[43] Newson J, Stables M, Karra E, Arce-Vargas F, Quezada S, Motwani M, Mack M, Yona S, Audzevich T, & Gilroy D W. Resolution of acute inflammation bridges the gap between innate and adaptive immunity. *Blood*, 124(11):1748–1764 (2014). doi:10.1182/blood-2014-03-562710.
URL http://www.ncbi.nlm.nih.gov/pubmed/25006125; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4383794; https://ashpublications.org/blood/article/124/11/1748/33037/Resolution-of-acute-inflammation-bridges-the-gap

[44] Louveau A, Harris T H, & Kipnis J. Revisiting the Mechanisms of CNS Immune Privilege. *Trends in Immunology*, 36(10):569–577 (2015). doi:10.1016/j.it.2015.08.006.
URL http://dx.doi.org/10.1016/j.it.2015.08.006

[45] Negi N & Das B K. CNS: Not an immunoprivilaged site anymore but a virtual secondary lymphoid organ. *International Reviews of Immunology*, 37(1):57–68 (2018). doi:10.1080/08830185.2017.1357719.
URL http://dx.doi.org/10.1080/08830185.2017.1357719; https://www.tandfonline.com/doi/full/10.1080/08830185.2017.1357719

[46] Walsh J, Hendrix S, Boato F *et al.* MHCII-independent CD4+ T cells protect injured CNS neurons via IL-4. *Journal of Clinical Investigation*, 125(2):699–714 (2015). doi:10.1172/JCI76210DS1.

[47] Meisel C, Schwab J M, Prass K, Meisel A, & Dirnagl U. Central nervous system injury-induced immune deficiency syndrome. *Nature reviews. Neuroscience*, 6(10):775–86 (2005). doi:10.1038/nrn1765.
URL http://www.ncbi.nlm.nih.gov/pubmed/16163382

[48] ElAli A & Jean LeBlanc N. The Role of Monocytes in Ischemic Stroke Pathobiology: New Avenues to Explore. *Frontiers in Aging Neuroscience*, 8(FEB):1–7 (2016). doi:10.3389/fnagi.2016.00029.
URL http://journal.frontiersin.org/Article/10.3389/fnagi.2016.00029/abstract

[49] Swirski F K, Nahrendorf M, Etzrodt M *et al.* Identification of Splenic Reservoir Monocytes and Their Deployment to Inflammatory Sites. *Science*, 325(5940):612–616 (2009). doi:10.1126/science.1175202.
URL https://www.sciencemag.org/lookup/doi/10.1126/science.1175202

[50] Kim E, Yang J, Beltran C D, & Cho S. Role of Spleen-Derived Monocytes/Macrophages in Acute Ischemic Brain Injury. *Journal of Cerebral Blood Flow & Metabolism*, 34(8):1411–1419 (2014). doi:10.1038/jcbfm.2014.101.
URL http://journals.sagepub.com/doi/10.1038/jcbfm.2014.101

[51] Ajmo C T, Vernon D O L, Collier L, Hall A A, Garbuzova-Davis S, Willing A, & Pennypacker K R. The spleen contributes to stroke-induced neurodegeneration. *Journal of Neuroscience Research*, 86(10):2227–2234 (2008). doi:10.1002/jnr.21661.
URL http://doi.wiley.com/10.1002/jnr.21661

[52] Bina K G, Rusak B, & Semba K. Localization of cholinergic neurons in the forebrain and brainstem that project to the suprachiasmatic nucleus of the hypothalamus in rat. *Journal of Comparative Neurology*, 335(2):295–307 (1993). doi:10.1002/cne.903350212.

[53] Xu M, Chung S, Zhang S *et al.* Basal forebrain circuit for sleep-wake control. *Nature Neuroscience*, 18(11):1641–1647 (2015). doi:10.1038/nn.4143.
URL http://www.ncbi.nlm.nih.gov/pubmed/26457552; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5776144; http://www.nature.com/articles/nn.4143

[54] Van Dort C J, Zachs D P, Kenny J D *et al.* Optogenetic activation of cholinergic neurons in the PPT or LDT induces REM sleep. *Proceedings of the National Academy of Sciences*, 112(2):584–589 (2015). doi:10.1073/pnas.1423136112.
URL http://www.pnas.org/lookup/doi/10.1073/pnas.1423136112

[55] Teles-Grilo Ruivo L M, Baker K L, Conway M W, Kinsley P J, Gilmour G, Phillips K G, Isaac J T, Lowry J P, & Mellor J R. Coordinated Acetylcholine Release in Prefrontal Cortex and Hippocampus Is Associated with Arousal and Reward on Distinct Timescales. *Cell Reports*, 18(4):905–917 (2017). doi:10.1016/j.celrep.2016.12.085.
URL http://linkinghub.elsevier.com/retrieve/pii/S2211124716318071

[56] Niwa Y, Kanda G N, Yamada R G *et al.* Muscarinic Acetylcholine Receptors Chrm1 and Chrm3 Are Essential for REM Sleep. *Cell Reports*, 24(9):2231–2247.e7 (2018). doi:10.1016/j.celrep.2018.07.082.
URL https://doi.org/10.1016/j.celrep.2018.07.082

[57] Yamakawa G R, Basu P, Cortese F, MacDonnell J, Whalley D, Smith V M, & Antle M C. The cholinergic forebrain arousal system acts directly on the circadian pacemaker. *Proceedings of the National Academy of Sciences*, 113(47):13498–13503 (2016). doi:10.1073/pnas.1610342113.
URL http://www.pnas.org/lookup/doi/10.1073/pnas.1610342113

[58] Ising M, Lauer C, Holsboer F, & Modell S. The Munich vulnerability study on affective disorders: premorbid neuroendocrine profile of affected high-risk probands. *Journal of Psychiatric Research*, 39(1):21–28 (2005). doi:10.1016/j.jpsychires.2004.04.009.
URL https://linkinghub.elsevier.com/retrieve/pii/S0022395604000585

[59] Wu J C & Bunney W E. The biological basis of an antidepressant response to sleep deprivation and relapse: review and hypothesis. *American Journal of Psychiatry*, 147(1):14–21 (1990). doi:10.1176/ajp.147.1.14.
URL http://psychiatryonline.org/doi/abs/10.1176/ajp.147.1.14

[60] Boonstra T W, Stins J F, Daffertshofer A, & Beek P J. Effects of sleep deprivation on neural functioning: an integrative review. *Cellular and Molecular Life Sciences*, 64(7-8):934–946 (2007). doi:10.1007/s00018-007-6457-8.
URL http://link.springer.com/10.1007/s00018-007-6457-8

[61] Nikonova E V, Gilliland J D, Tanis K Q *et al.* Transcriptional Profiling of Cholinergic Neurons From Basal Forebrain Identifies Changes in Expression of Genes Between Sleep and Wake. *Sleep*, 40(6):16–20 (2017). doi:10.1093/sleep/zsx059.
URL https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/zsx059

[62] Balsalobre A. Clock genes in mammalian peripheral tissues. *Cell and Tissue Research*, 309(1):193–199 (2002). doi:10.1007/s00441-002-0585-0.

[63] King D P, Zhao Y, Sangoram A M *et al.* Positional Cloning of the Mouse Circadian Clock Gene. *Cell*, 89(4):641–653 (1997). doi:10.1016/S0092-8674(00)80245-7.
URL http://linkinghub.elsevier.com/retrieve/pii/S0030665708702269; https://linkinghub.elsevier.com/retrieve/pii/S0092867400802457

[64] Levi-Montalcini R & Booker B. Destruction of the sympathetic ganglia in mammals by an antiserum to a nerve-growth protein. *Proceedings of the National Academy of Sciences*, 46(3):384–391 (1960). doi:10.1073/pnas.46.3.384.
URL http://www.ncbi.nlm.nih.gov/pubmed/16578497; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC222845; http://www.pnas.org/cgi/doi/10.1073/pnas.46.3.384

[65] Hefti F. Nerve growth factor promotes survival of septal cholinergic neurons after fimbrial transections. *Journal of Neuroscience*, 6(8):2155–2162 (1986).

[66] McManaman J L, Crawford F G, Stewart S S, & Appel S H. Purification of a Skeletal Muscle Polypeptide Which Stimulates Choline Acetyltransferase Activity in Cultured Spinal Cord Neurons. *Journal of Biological Chemistry*, 263(12):5890–5897 (1988).

[67] Rao M S, Patterson P H, & Landis S C. Multiple cholinergic differentiation factors are present in footpad extracts: comparison with known cholinergic factors. *Development (Cambridge, England)*, 116(3):731–44 (1992).
URL http://www.ncbi.nlm.nih.gov/pubmed/1289063

[68] White U A & Stephens J M. The gp130 receptor cytokine family: regulators of adipocyte development and function. *Current pharmaceutical design*, 17(4):340–6 (2011). doi:10.2174/138161211795164202.
URL http://www.ncbi.nlm.nih.gov/pubmed/21375496; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3119891

[69] Rawlings J S. The JAK/STAT signaling pathway. *Journal of Cell Science*, 117(8):1281–1283 (2004). doi:10.1242/jcs.00963.
URL http://jcs.biologists.org/cgi/doi/10.1242/jcs.00963

[70] Nathanson N M. Regulation of neurokine receptor signaling and trafficking. *Neurochemistry International*, 61(6):874–878 (2012). doi:10.1016/j.neuint.2012.01.018.
URL https://linkinghub.elsevier.com/retrieve/pii/S0197018612000307

[71] Mohamed-Ali V, Goodrick S, Rawesh A, Katz D R, Miles J M, Yudkin J S, Klein S, & Coppack S W. Subcutaneous Adipose Tissue Releases Interleukin-6, But Not Tumor Necrosis Factor-α, in Vivo 1. *The Journal of Clinical Endocrinology & Metabolism*, 82(12):4196–4200 (1997). doi:10.1210/jcem.82.12.4450.
URL https://academic.oup.com/jcem/article-lookup/doi/10.1210/jcem.82.12.4450; http://www.ncbi.nlm.nih.gov/pubmed/9398739

[72] Gustafson D, Lissner L, Bengtsson C, Bjorkelund C, & Skoog I. A 24-year follow-up of body mass index and cerebral atrophy. *Neurology*, 63(10):1876–1881 (2004). doi:10.1212/01.WNL.0000141850.47773.5F.
URL http://www.ncbi.nlm.nih.gov/pubmed/15557505; http://www.neurology.org/cgi/doi/10.1212/01.WNL.0000141850.47773.5F

[73] Profenno L A, Porsteinsson A P, & Faraone S V. Meta-Analysis of Alzheimer's Disease Risk with Obesity, Diabetes, and Related Disorders. *Biological Psychiatry*, 67(6):505–512 (2010). doi:10.1016/j.biopsych.2009.02.013.
URL https://www.sciencedirect.com/science/article/abs/pii/S0006322309002261; https://linkinghub.elsevier.com/retrieve/pii/S0006322309002261

[74] Depp C A, Strassnig M, Mausbach B T *et al.* Association of obesity and treated hypertension and diabetes with cognitive ability in bipolar disorder and schizophrenia. *Bipolar Disorders*,

16(4):422–431 (2014). doi:10.1111/bdi.12200.
URL http://doi.wiley.com/10.1111/bdi.12200

[75] Eder K, Baffy N, Falus A, & Fulop A K. The major inflammatory mediator interleukin-6 and obesity. *Inflammation Research*, 58(11):727–736 (2009). doi:10.1007/s00011-009-0060-4.
URL http://link.springer.com/10.1007/s00011-009-0060-4; http://www.ncbi.nlm.nih.gov/pubmed/19543691

[76] Heppner F L, Ransohoff R M, & Becher B. Immune attack: the role of inflammation in Alzheimer disease. *Nature Reviews Neuroscience*, 16(6):358–372 (2015). doi:10.1038/nrn3880.
URL https://www.nature.com/articles/nrn3880; http://www.nature.com/articles/nrn3880

[77] Kirkpatrick B & Miller B J. Inflammation and Schizophrenia. *Schizophrenia Bulletin*, 39(6):1174–1179 (2013). doi:10.1093/schbul/sbt141.
URL https://academic.oup.com/schizophreniabulletin/article-lookup/doi/10.1093/schbul/sbt141

[78] Takao K, Kobayashi K, Hagihara H *et al.* Deficiency of Schnurri-2, an MHC Enhancer Binding Protein, Induces Mild Chronic Inflammation in the Brain and Confers Molecular, Neuronal, and Behavioral Phenotypes Related to Schizophrenia. *Neuropsychopharmacology*, 38(8):1409–1425 (2013). doi:10.1038/npp.2013.38.
URL http://www.nature.com/articles/npp201338

[79] Hodes G E, Kana V, Menard C, Merad M, & Russo S J. Neuroimmune mechanisms of depression. *Nature Neuroscience*, 18(10):1386–1393 (2015). doi:10.1038/nn.4113.

[80] Stangl H, Springorum H R, Muschter D, Grässel S, & Straub R H. Catecholaminergic-to-cholinergic transition of sympathetic nerve fibers is stimulated under healthy but not under inflammatory arthritic conditions. *Brain, Behavior, and Immunity*, 46:180–191 (2015). doi:10.1016/j.bbi.2015.02.022.
URL http://dx.doi.org/10.1016/j.bbi.2015.02.022

[81] Babu M M, Luscombe N M, Aravind L, Gerstein M, & Teichmann S A. Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology*, 14(3):283–291 (2004). doi:10.1016/j.sbi.2004.05.004.
URL https://linkinghub.elsevier.com/retrieve/pii/S0959440X04000788

[82] Lee R C, Feinbaum R L, & Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854 (1993). doi:

10.1016/0092-8674(93)90529-Y.
URL https://linkinghub.elsevier.com/retrieve/pii/009286749390529Y

[83] Rodriguez A. Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Research*, 14(10a):1902–1910 (2004). doi:10.1101/gr.2722704.
URL http://www.genome.org/cgi/doi/10.1101/gr.2722704

[84] Kozomara A, Birgaoanu M, & Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1):D155–D162 (2019). doi:10.1093/nar/gky1141.
URL https://academic.oup.com/nar/article/47/D1/D155/5179337

[85] Ambros V, Bartel B, Bartel D P *et al.* A uniform system for microRNA annotation. *RNA (New York, N.Y.)*, 9(3):277–9 (2003). doi:10.1261/rna.2183803.
URL http://www.ncbi.nlm.nih.gov/pubmed/12592000; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1370393

[86] Salta E & De Strooper B. microRNA-132: a key noncoding RNA operating in the cellular phase of Alzheimer's disease. *The FASEB Journal*, 31(2):424–433 (2017). doi:10.1096/fj.201601308.
URL http://www.fasebj.org/doi/10.1096/fj.201601308

[87] Lu L F, Gasteiger G, Yu I S *et al.* A Single miRNA-mRNA Interaction Affects the Immune Response in a Context- and Cell-Type-Specific Manner. *Immunity*, 43(1):52–64 (2015). doi:10.1016/j.immuni.2015.04.022.
URL http://www.ncbi.nlm.nih.gov/pubmed/26163372; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4529747; https://linkinghub.elsevier.com/retrieve/pii/S1074761315002551

[88] Borek E, Baliga B S, Gehrke C W, Kuo C W, Belman S, Troll W, & Waalkes T P. High Turnover Rate of Transfer RNA in Tumor Tissue. *CANCER RESEARCH*, 37:3362–3366 (1977).
URL https://cancerres.aacrjournals.org/content/37/9/3362.full-text.pdf

[89] Speer J, Gehrke C W, Kuo K C, Waalkes T P, & Borek E. tRNA breakdown products as markers for cancer. *Cancer*, 44(6):2120–2123 (1979). doi:10.1002/1097-0142(197912)44:6<2120::AID-CNCR2820440623>3.0.CO;2-6.
URL http://www.ncbi.nlm.nih.gov/pubmed/509391; http://doi.wiley.com/10.1002/1097-0142{%}28197912{%}2944{%}3A6{%}3C2120{%}3A{%}3AAID-CNCR2820440623{%}3E3.0.CO{%}3B2-6

[90] Cole C, Sobala A, Lu C, Thatcher S R, Bowman A, Brown J W, Green P J, Barton G J, & Hutvagner G. Filtering of deep sequencing data reveals the existence of abundant

Dicer-dependent small RNAs derived from tRNAs. *RNA*, 15(12):2147–2160 (2009). doi:10.1261/rna.1738409.
URL http://www.ncbi.nlm.nih.gov/pubmed/19850906; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2779667; http://rnajournal.cshlp.org/cgi/doi/10.1261/rna.1738409

[91] Lee Y S, Shibata Y, Malhotra A, & Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & development*, 23(22):2639–49 (2009). doi:10.1101/gad.1837609.
URL http://www.ncbi.nlm.nih.gov/pubmed/19933153; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2779758

[92] Godoy P M, Bhakta N R, Barczak A J *et al.* Large Differences in Small RNA Composition Between Human Biofluids. *Cell Reports*, 25(5):1346–1358 (2018). doi:10.1016/j.celrep.2018.10.014.
URL https://doi.org/10.1016/j.celrep.2018.10.014; https://linkinghub.elsevier.com/retrieve/pii/S2211124718315778

[93] Yamasaki S, Ivanov P, Hu G f, & Anderson P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *The Journal of Cell Biology*, 185(1):35–42 (2009). doi:10.1083/JCB.200811106.
URL http://jcb.rupress.org/content/185/1/35.long

[94] Ivanov P, Emara M M, Villen J, Gygi S P, & Anderson P. Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Molecular Cell*, 43(4):613–623 (2011). doi:10.1016/j.molcel.2011.06.022.
URL https://www.sciencedirect.com/science/article/pii/S1097276511005247?via{%}3Dihub; https://linkinghub.elsevier.com/retrieve/pii/S1097276511005247

[95] Burroughs A M, Ando Y, de Hoon M L, Tomaru Y, Suzuki H, Hayashizaki Y, & Daub C O. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biology*, 8(1):158–177 (2011). doi:10.4161/rna.8.1.14300.
URL http://www.tandfonline.com/doi/abs/10.4161/rna.8.1.14300

[96] Kumar P, Anaya J, Mudunuri S B, & Dutta A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biology*, 12(1):78 (2014). doi:10.1186/s12915-014-0078-0.
URL http://www.ncbi.nlm.nih.gov/pubmed/25270025; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4203973; http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-014-0078-0

[97] Huang B, Yang H, Cheng X *et al.* tRF/miR-1280 Suppresses Stem Cell–like Cells and Metastasis in Colorectal Cancer. *Cancer Research*, 77(12):3194–3206 (2017). doi:10.1158/0008-5472.CAN-16-3146.
URL http://cancerres.aacrjournals.org/; http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-16-3146

[98] Gebetsberger J, Wyss L, Mleczko A M, Reuther J, & Polacek N. A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biology*, 14(10):1364–1373 (2017). doi:10.1080/15476286.2016.1257470.
URL https://www.tandfonline.com/action/journalInformation?journalCode=krnb20; https://www.tandfonline.com/doi/full/10.1080/15476286.2016.1257470

[99] Goodarzi H, Liu X, Nguyen H C, Zhang S, Fish L, & Tavazoie S F. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. *Cell*, 161(4):790–802 (2015). doi:10.1016/j.cell.2015.02.053.
URL https://www.sciencedirect.com/science/article/pii/S0092867415003189?via{%}3Dihub; https://linkinghub.elsevier.com/retrieve/pii/S0092867415003189

[100] Kim H K, Fuchs G, Wang S *et al.* A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature*, 552(7683):57 (2017). doi:10.1038/nature25005.
URL http://www.nature.com/doifinder/10.1038/nature25005

[101] Parisien M, Wang X, & Pan T. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biology*, 10(12):1853–1867 (2013). doi:10.4161/rna.27361.
URL http://www.ncbi.nlm.nih.gov/pubmed/24448271; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3917988; http://www.tandfonline.com/doi/abs/10.4161/rna.27361

[102] Loher P, Telonis A G, & Rigoutsos I. MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Scientific Reports*, 7(1):41184 (2017). doi:10.1038/srep41184.
URL http://dx.doi.org/10.1038/srep41184; http://www.nature.com/articles/srep41184

[103] Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, & Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nature Methods*, 13(4):366–370 (2016). doi:10.1038/nmeth.3799.
URL http://www.ncbi.nlm.nih.gov/pubmed/26950747; http://www.nature.com/articles/nmeth.3799; http://regulatorycircuits.org

[104] Nowakowski T J, Rani N, Golkaram M *et al.* Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nature Neuroscience*,

21(12):1784–1792 (2018). doi:10.1038/s41593-018-0265-3.
URL http://www.ncbi.nlm.nih.gov/pubmed/30455455; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6312854; http://www.nature.com/articles/s41593-018-0265-3

[105] Londin E, Loher P, Telonis A G *et al.* Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proceedings of the National Academy of Sciences*, 112(10):E1106–E1115 (2015). doi:10.1073/pnas.1420955112.
URL http://www.pnas.org/lookup/doi/10.1073/pnas.1420955112

[106] Dweep H & Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nature Methods*, 12(8):697–697 (2015). doi:10.1038/nmeth.3485.
URL http://www.nature.com/doifinder/10.1038/nmeth.3485

[107] Chaudhuri S, Narasayya V, & Ramamurthy R. Estimating Progress of Execution for SQL Queries (2004).
URL https://www.microsoft.com/en-us/research/publication/estimating-progress-of-execution-for-sql-queries/?from=https{%}3A{%}2F{%}2Fresearch.microsoft.com{%}2Fapps{%}2Fpubs{%}2F{%}3Fid{%}3D76556

[108] Hon C c, Ramilowski J A, Harshbarger J *et al.* An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature Publishing Group*, 543(7644):199–204 (2017). doi:10.1038/nature21374.
URL http://dx.doi.org/10.1038/nature21374

[109] Dweep H & Gretz N. miRWalk2 web page.

[110] Karagkouni D, Paraskevopoulou M D, Chatzopoulos S *et al.* DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Research*, 46(D1):D239–D245 (2018). doi:10.1093/nar/gkx1141.
URL http://academic.oup.com/nar/article/46/D1/D239/4634010

[111] Chou C H, Shrestha S, Yang C D *et al.* miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research*, 46(D1):D296–D302 (2018). doi:10.1093/nar/gkx1067.
URL http://www.ncbi.nlm.nih.gov/pubmed/29126174; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5753222; http://academic.oup.com/nar/article/46/D1/D296/4595852

[112] Yue D, Liu H, & Huang Y. Survey of Computational Algorithms for MicroRNA Target Prediction. *Current Genomics*, 10(7):478–492 (2009). doi:10.2174/138920209789208219.

URL http://www.ncbi.nlm.nih.gov/pubmed/20436875; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2808675; http://www.eurekaselect.com/openurl/content.php?genre=article{&}issn=1389-2029{&}volume=10{&}issue=7{&}spage=478

[113] Witkos T M, Koscianska E, & Krzyzosiak W J. Practical Aspects of microRNA Target Prediction. *Current molecular medicine*, 11(2):93–109 (2011). doi:10.2174/156652411794859250.

[114] Friedman R C, Farh K K H, Burge C B, & Bartel D P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105 (2009). doi:10.1101/gr.082701.108.
URL http://www.ncbi.nlm.nih.gov/pubmed/18955434; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2612969; http://genome.cshlp.org/cgi/doi/10.1101/gr.082701.108

[115] Alexiou P, Maragkakis M, Papadopoulos G L, Reczko M, & Hatzigeorgiou A G. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, 25(23):3049–3055 (2009). doi:10.1093/bioinformatics/btp565.
URL http://www.ncbi.nlm.nih.gov/pubmed/19789267; https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp565

[116] Agarwal V, Bell G W, Nam J W, & Bartel D P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4 (2015). doi:10.7554/eLife.05005.
URL https://elifesciences.org/articles/05005

[117] Soreq H. Checks and balances on cholinergic signaling in brain and body function. *Trends in Neurosciences*, 38(7):448–458 (2015). doi:10.1016/j.tins.2015.05.007.
URL http://dx.doi.org/10.1016/j.tins.2015.05.007

[118] Smith T & Waterman M. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197 (1981). doi:10.1016/0022-2836(81)90087-5.
URL https://www.sciencedirect.com/science/article/pii/0022283681900875?via{%}3Dihub; https://linkinghub.elsevier.com/retrieve/pii/0022283681900875

[119] Needleman S B & Wunsch C D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453 (1970). doi:10.1016/0022-2836(70)90057-4.
URL https://www.sciencedirect.com/science/article/pii/0022283670900574?via{%}3Dihub

[120] Raichle M E & Gusnard D A. Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences*, 99(16):10237–10239 (2002). doi:10.1073/pnas.172399499.
URL http://www.pnas.org/cgi/doi/10.1073/pnas.172399499

[121] Bohn K A, Adkins C E, Mittapalli R K, Terrell-Hall T B, Mohammad A S, Shah N, Dolan E L, Nounou M I, & Lockman P R. Semi-automated rapid quantification of brain vessel density utilizing fluorescent microscopy. *Journal of Neuroscience Methods*, 270:124–131 (2016). doi:10.1016/j.jneumeth.2016.06.012.
URL http://www.ncbi.nlm.nih.gov/pubmed/27321229; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4981522; https://linkinghub.elsevier.com/retrieve/pii/S0165027016301339

[122] Lobentanzer S & Klein J. Zentrales und Peripheres Nervensystem. In Wichmann & Fromme, (Editors) *Handbuch für Umweltmedizin*, chapter 11. ecomed Medizin, erg. lfg. edition (2019).

[123] Darmanis S, Sloan S A, Zhang Y, Enge M, Caneda C, Shuer L M, Hayden Gephart M G, Barres B A, & Quake S R. A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences*, 112(23):201507125 (2015). doi:10.1073/pnas.1507125112.
URL http://www.pnas.org/content/112/23/7285.abstract

[124] Zeisel a, Manchado a B M, Codeluppi S *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, 347(6226):1138–42 (2015). doi:10.1126/science.aaa1934.
URL http://science.sciencemag.org.docelec.univ-lyon1.fr/content/347/6226/1138.abstract

[125] Tasic B, Menon V, Nguyen T N T *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*, advance on(January):1–37 (2016). doi:10.1038/nn.4216.
URL http://dx.doi.org/10.1038/nn.4216

[126] Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi I, Trombetta J J, Hession C, Zhang F, & Regev A. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science*, 353(6302):925–928 (2016). doi:10.1126/science.aad7038.
URL http://www.sciencemag.org/lookup/doi/10.1126/science.aad7038

[127] Zeisel A, Hochgerner H, Lönnerberg P *et al.* Molecular Architecture of the Mouse Nervous System. *Cell*, 174(4):999–1014.e22 (2018). doi:10.1016/j.cell.2018.06.021.
URL http://www.ncbi.nlm.nih.gov/pubmed/30096314; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6086934; https://linkinghub.elsevier.com/retrieve/pii/S009286741830789X

[128] Murtagh F & Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, 31(3):274–295 (2014).

doi:10.1007/s00357-014-9161-z.
URL http://link.springer.com/10.1007/s00357-014-9161-z

[129] Bray J R & Curtis J T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4):325–349 (1957). doi:10.2307/1942268.
URL http://doi.wiley.com/10.2307/1942268

[130] Biedler J L, Roffler-Tarlov S, Schachner M, & Freedman L S. Multiple Neurotransmitter Synthesis by Human Neuroblastoma Cell Lines and Clones. *Cancer Res.*, 38(11_Part_1):3751–3757 (1978).
URL http://cancerres.aacrjournals.org/content/38/11{_}Part{_}1/3751.short

[131] Biedler J L, Helson L, & Spengler B A. Morphology and Growth, Tumorigenicity, and Cytogenetics of Human Neuroblastoma Cells in Continuous Culture. *Cancer Research*, 33(11):2643–2652 (1973).

[132] Seeger R C, Rayner S A, Laug W E, Neustein H B, & Benedict W F. Morphology, Growth, Chromosomal Pattern, and Fibrinolytic Activity of two New Human Neuroblastoma Cell Lines. *Cancer Research*, 37(5):1364–1371. (1977).

[133] Seeger R C, Danon Y L, Rayner S A, & Hoover F. Definition of a Thy-1 determinant on human neuroblastoma, glioma, sarcoma, and teratoma cells with a monoclonal antibody. *Journal of immunology (Baltimore, Md. : 1950)*, 128(2):983–9 (1982).
URL http://www.ncbi.nlm.nih.gov/pubmed/6172518

[134] Hill D P & Robertson K a. Characterization of the cholinergic neuronal differentiation of the human neuroblastoma cell line LA-N-5 after treatment with retinoic acid. *Developmental Brain Research*, 102(1):53–67 (1997). doi:10.1016/S0165-3806(97)00076-X.
URL http://www.ncbi.nlm.nih.gov/pubmed/9298234; https://linkinghub.elsevier.com/retrieve/pii/S016538069700076X

[135] McManaman J L & Crawford F G. Skeletal Muscle Proteins Stimulate Cholinergic Differentiation of Human Neuroblastoma Cells. *Journal of neurochemistry*, pp. 258–266 (1991).

[136] Sun M, Liu H, Min S, Wang H, & Wang X. Ciliary neurotrophic factor-treated astrocyte-conditioned medium increases the intracellular free calcium concentration in rat cortical neurons. *Biomedical Reports*, 4(4):417–420 (2016). doi:10.3892/br.2016.602.
URL https://www.spandidos-publications.com/; https://www.spandidos-publications.com/10.3892/br.2016.602

[137] Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, & Wingett S. *FastQC*. Babraham, UK (2012).

[138] Roehr J T, Dieterich C, & Reinert K. Flexbar 3.0 – SIMD and multicore parallelization. *Bioinformatics*, 33(18):2941–2942 (2017). doi:10.1093/bioinformatics/btx330.
URL https://academic.oup.com/bioinformatics/article/33/18/2941/3852078

[139] Wang W C, Lin F M, Chang W C, Lin K Y, Huang H D, & Lin N S. miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics*, 10:328 (2009). doi:10.1186/1471-2105-10-328.
URL https://www.ncbi.nlm.nih.gov/pubmed/19821977

[140] Love M I, Huber W, & Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550 (2014). doi:10.1186/s13059-014-0550-8.
URL http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8

[141] Wald A. Contributions to the Theory of Statistical Estimation and Testing Hypotheses. *The Annals of Mathematical Statistics*, 10(4):299–326 (1939). doi:10.1098/rsta.1937.0005.
URL http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.1937.0005

[142] Bullard J H, Purdom E, Hansen K D, & Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, 11(1):94 (2010). doi:10.1186/1471-2105-11-94.
URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-94

[143] Chen Z, Liu J, Ng H, Nadarajah S, Kaufman H L, Yang J Y, & Deng Y. Statistical methods on detecting differentially expressed genes for RNA-seq data. *BMC Systems Biology*, 5(Suppl 3):S1 (2011). doi:10.1186/1752-0509-5-S3-S1.
URL http://www.ncbi.nlm.nih.gov/pubmed/22784615; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3287564; http://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-5-S3-S1

[144] Zhu A, Ibrahim J G, & Love M I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*, 35(12):2084–2092 (2019). doi:10.1093/bioinformatics/bty895.
URL https://academic.oup.com/bioinformatics/article/35/12/2084/5159452

[145] Alexa A, Rahnenfuhrer J, & Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22(13):1600–1607 (2006). doi:10.1093/bioinformatics/btl140.
URL http://www.ncbi.nlm.nih.gov/pubmed/16606683; https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btl140

[146] Broido A D & Clauset A. Scale-free networks are rare. *Nature Communications*, 10(1):1017 (2019). doi:10.1038/s41467-019-08746-5.
URL http://www.nature.com/articles/s41467-019-08746-5

[147] Jacomy M, Venturini T, Heymann S, & Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6):1–12 (2014). doi:10.1371/journal.pone.0098679.

[148] Chen C, Cheng L, Grennan K, Pibiri F, Zhang C, Badner J A, Members of the Bipolar Disorder Genome Study (BiGS) Consortium, Gershon E S, & Liu C. Two gene co-expression modules differentiate psychotics and controls. *Molecular psychiatry*, 18(12):1308–14 (2013). doi:10.1038/mp.2012.146.
URL http://www.ncbi.nlm.nih.gov/pubmed/23147385; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4018461

[149] Iwamoto K, Bundo M, & Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Human molecular genetics*, 14(2):241–53 (2005). doi:10.1093/hmg/ddi022.
URL http://www.ncbi.nlm.nih.gov/pubmed/15563509

[150] Lanz T A, Joshi J J, Reinhart V, Johnson K, Grantham L E, & Volfson D. STEP levels are unchanged in pre-frontal cortex and associative striatum in post-mortem human brain samples from subjects with schizophrenia, bipolar disorder and major depressive disorder. *PloS one*, 10(3):e0121744 (2015). doi:10.1371/journal.pone.0121744.
URL http://www.ncbi.nlm.nih.gov/pubmed/25786133; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4364624

[151] Maycox P R, Kelly F, Taylor A *et al.* Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Molecular psychiatry*, 14(12):1083–94 (2009). doi:10.1038/mp.2009.18.
URL http://www.ncbi.nlm.nih.gov/pubmed/19255580

[152] Narayan S, Tang B, Head S R, Gilmartin T J, Sutcliffe J G, Dean B, & Thomas E A. Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain research*, 1239:235–48 (2008). doi:10.1016/j.brainres.2008.08.023.
URL http://www.ncbi.nlm.nih.gov/pubmed/18778695; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2783475

[153] Ryan M M, Lockstone H E, Huffaker S J, Wayland M T, Webster M J, & Bahn S. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations

in synaptic genes. *Molecular psychiatry*, 11(10):965–78 (2006). doi:10.1038/sj.mp.4001875.
URL http://www.ncbi.nlm.nih.gov/pubmed/16894394

[154] Ramaker R C, Bowling K M, Lasseigne B N *et al.* Post-mortem molecular profiling of three psychiatric disorders. *Genome Medicine*, 9(1):1–12 (2017). doi:10.1186/s13073-017-0458-5.

[155] Hoffman G E, Hartley B J, Flaherty E, Ladran I, Gochman P, Ruderfer D M, Stahl E A, Rapoport J, Sklar P, & Brennand K J. Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nature Communications*, 8(1) (2017). doi:10.1038/s41467-017-02330-5.
URL http://dx.doi.org/10.1038/s41467-017-02330-5

[156] Webb A, Papp A C, Curtis A *et al.* RNA sequencing of transcriptomes in human brain regions: Protein-coding and non-coding RNAs, isoforms and alleles. *BMC Genomics*, 16(1):1–16 (2015). doi:10.1186/s12864-015-2207-8.
URL http://dx.doi.org/10.1186/s12864-015-2207-8

[157] Fontenot M R, Berto S, Liu Y, Werthmann G, Douglas C, Usu N, Gleason K, Tamminga C A, Takahashi J S, & Konopka G. Novel transcriptional networks regulated by clock in human neurons. *Genes and Development*, 31(21):2121–2135 (2017). doi:10.1101/gad.305813.117.

[158] Li G, Klein J, & Zimmermann M. Pathophysiological Amyloid Concentrations Induce Sustained Upregulation of Readthrough AChE Mediating Anti-Apoptotic Effects. *Neuroscience*, 240:349–60 (2013). doi:10.1016/j.neuroscience.2013.02.040.
URL http://www.ncbi.nlm.nih.gov/pubmed/23485809

[159] Gulyás-Kovács A, Keydar I, Xia E, Fromer M, Hoffman G, Ruderfer D, Sachidanandam R, & Chess A. Unperturbed expression bias of imprinted genes in schizophrenia. *Nature Communications*, 9(1):2914 (2018). doi:10.1038/s41467-018-04960-9.
URL https://www.biorxiv.org/content/early/2018/05/24/329748?{%}3Fcollection=; http://www.nature.com/articles/s41467-018-04960-9

[160] Du P, Kibbe W A, & Lin S M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, 24(13):1547–1548 (2008). doi:10.1093/bioinformatics/btn224.
URL http://www.ncbi.nlm.nih.gov/pubmed/18467348; https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btn224

[161] Gautier L, Cope L, Bolstad B M, & Irizarry R A. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3):307–315 (2004). doi:10.1093/bioinformatics/btg405.
URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btg405

[162] Oldham M C, Langfelder P, & Horvath S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Systems Biology*, 6(1):63 (2012). doi:10.1186/1752-0509-6-63.
URL http://www.ncbi.nlm.nih.gov/pubmed/22691535; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3441531; http://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-6-63

[163] Durinck S, Spellman P T, Birney E, & Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8):1184–1191 (2009). doi:10.1038/nprot.2009.97.
URL http://www.ncbi.nlm.nih.gov/pubmed/19617889; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3159387; http://www.nature.com/articles/nprot.2009.97

[164] Langfelder P & Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9 (2008). doi:10.1186/1471-2105-9-559.

[165] Pinheiro J, Bates D, DebRoy S, Sarkar D, & R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models* (2019). R package version 3.1-142.
URL https://CRAN.R-project.org/package=nlme

[166] Shaltiel G, Hanan M, Wolf Y, Barbash S, Kovalev E, Shoham S, & Soreq H. Hippocampal microRNA-132 mediates stress-inducible cognitive deficits through its acetylcholinesterase target. *Brain structure function*, 218(1):1–5 (2013). doi:10.1007/s00429-011-0376-z.
URL http://www.ncbi.nlm.nih.gov/pubmed/22246100

[167] Hanin G, Yayon N, Tzur Y *et al.* miRNA-132 induces hepatic steatosis and hyperlipidaemia by synergistic multitarget suppression. *Gut*, 67(6):1124–1134 (2018). doi:10.1136/gutjnl-2016-312869.
URL http://gut.bmj.com/lookup/doi/10.1136/gutjnl-2016-312869; http://www.ncbi.nlm.nih.gov/pubmed/28381526; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5969364

[168] Mellios N, Sugihara H, Castro J *et al.* miR-132, an experience-dependent microRNA, is essential for visual cortex plasticity. *Nature Neuroscience*, 14(10):1240–1242 (2011). doi:10.1038/nn.2909.
URL http://www.nature.com/articles/nn.2909

[169] Shaked I, Meerson A, Wolf Y, Avni R, Greenberg D, Gilboa-Geffen A, & Soreq H. MicroRNA-132 potentiates cholinergic anti-inflammatory signaling by targeting acetyl-

cholinesterase. *Immunity*, 31(6):965–973 (2009). doi:10.1016/j.immuni.2009.09.019.
URL http://www.ncbi.nlm.nih.gov/pubmed/20005135

[170] Pichler S, Gu W, Hartl D, Gasparoni G, Leidinger P, Keller A, Meese E, Mayhaus M, Hampel H, & Riemenschneider M. The miRNome of Alzheimer's disease: consistent downregulation of the miR-132/212 cluster. *Neurobiology of Aging*, 50:167.e1–167.e10 (2017). doi:10.1016/j.neurobiolaging.2016.09.019.
URL https://linkinghub.elsevier.com/retrieve/pii/S0197458016302330

[171] Busch S, Auth E, Scholl F, Huenecke S, Koehl U, Suess B, & Steinhilber D. 5-Lipoxygenase Is a Direct Target of miR-19a-3p and miR-125b-5p. *The Journal of Immunology*, 194(4):1646–1653 (2015). doi:10.4049/jimmunol.1402163.
URL http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1402163

[172] Zhang J, Qu P, Zhou C, Liu X, Ma X, Wang M, Wang Y, Su J, Liu J, & Zhang Y. MicroRNA-125b is a key epigenetic regulatory factor that promotes nuclear transfer reprogramming. *Journal of Biological Chemistry*, 292(38):15916–15926 (2017). doi:10.1074/jbc.M117.796771.

[173] Soreq H, Ben-Aziz R, Prody C A, Seidman S, Gnatt A, Neville L, Lieman-Hurwitz J, Lev-Lehman E, Ginzberg D, & Lipidot-Lifson Y. Molecular cloning and construction of the coding region for human acetylcholinesterase reveals a G + C-rich attenuating structure. *Proceedings of the National Academy of Sciences*, 87(24):9688–9692 (1990). doi:10.1073/pnas.87.24.9688.
URL http://www.pnas.org/cgi/doi/10.1073/pnas.87.24.9688

[174] Hanin G, Shenhar-Tsarfaty S, Yayon N *et al.* Competing targets of microRNA-608 affect anxiety and hypertension. *Human Molecular Genetics*, 23(17):4569–4580 (2014). doi:10.1093/hmg/ddu170.
URL https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddu170; http://www.ncbi.nlm.nih.gov/pubmed/24722204; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4119407

[175] Hoffmann S, Harms H, Ulm L *et al.* Stroke-induced immunodepression and dysphagia independently predict stroke-associated pneumonia – The PREDICT study. *Journal of Cerebral Blood Flow & Metabolism*, 37(12):3671–3682 (2017). doi:10.1177/0271678X16671964.
URL https://www.ncbi.nlm.nih.gov/pubmed/27733675; http://journals.sagepub.com/doi/10.1177/0271678X16671964

[176] Patro R, Duggal G, Love M I, Irizarry R A, & Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4):417–419 (2017). doi:

10.1038/nmeth.4197.
URL http://www.nature.com/articles/nmeth.4197

[177] Liao Y, Smyth G K, & Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, 47(8):e47–e47 (2019). doi:10.1093/nar/gkz114.
URL https://academic.oup.com/nar/article/47/8/e47/5345150

[178] Tokar T, Pastrello C, & Jurisica I. GSOAP: a tool for visualization of gene set over-representation analysis. *Bioinformatics* (2020). doi:10.1093/bioinformatics/btaa001.
URL https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btaa001/5715574

[179] Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York (2016).

[180] van der Maaten L & Hinton G. Visualizing Data using t-SNE. *Journal of Machine Learning research*, 9(1) (2008).

[181] McInnes L, Healy J, & Melville J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (2018).

[182] Becht E, McInnes L, Healy J, Dutertre C A, Kwok I W H, Ng L G, Ginhoux F, & Newell E W. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, 37(1):38–44 (2019). doi:10.1038/nbt.4314.
URL http://www.nature.com/articles/nbt.4314

[183] Krijthe J H. *Rtsne: T-Distributed Stochastic Neighbor Embedding using Barnes-Hut Implementation* (2015). R package version 0.15.
URL https://github.com/jkrijthe/Rtsne

[184] Juzenas S, Venkatesh G, Hübenthal M *et al.* A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Research*, 45(16):9290–9301 (2017). doi:10.1093/nar/gkx706.

[185] Delignette-Muller M L & Dutang C. fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34 (2015).
URL http://www.jstatsoft.org/v64/i04/

[186] Ward J H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244 (1963). doi:10.1080/01621459.1963.10500845.
URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845

[187] Lehtonen A, Matikainen S, & Julkunen I. Interferons up-regulate STAT1, STAT2, and IRF family transcription factor gene expression in human peripheral blood mononuclear cells and macrophages. *The Journal of Immunology*, 159(2):794 LP – 803 (1997).
URL http://www.jimmunol.org/content/159/2/794.abstract

[188] Reeves G T. The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback. *Journal of Biological Engineering*, 13(1):62 (2019). doi: 10.1186/s13036-019-0190-3.
URL https://jbioleng.biomedcentral.com/articles/10.1186/s13036-019-0190-3

[189] Lai X, Wolkenhauer O, & Vera J. Understanding microRNA-mediated gene regulatory networks through mathematical modelling. *Nucleic Acids Research*, 44(13):6019–6035 (2016). doi:10.1093/nar/gkw550.
URL https://academic.oup.com/nar/article/44/13/6019/2457646/Understanding-microRNA-mediated-gene-regulatory; https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw550

[190] Blondel V D, Guillaume J L, Lambiotte R, & Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment* (2008). doi: 10.1088/1742-5468/2008/10/P10008.
URL http://arxiv.org/abs/0803.0476; http://dx.doi.org/10.1088/1742-5468/2008/10/P10008

[191] Zhang Y J, Rutledge B J, & Rollins B J. Structure/activity analysis of human monocyte chemoattractant protein-1 (MCP-1) by mutagenesis. Identification of a mutated protein that inhibits MCP-1-mediated monocyte chemotaxis. *The Journal of biological chemistry*, 269(22):15918–24 (1994).
URL http://www.ncbi.nlm.nih.gov/pubmed/8195247

[192] Sanchez-Pulido L, Valencia A, & Rojas A M. Are promyelocytic leukaemia protein nuclear bodies a scaffold for caspase-2 programmed cell death? *Trends in Biochemical Sciences*, 32(9):400–406 (2007). doi:10.1016/j.tibs.2007.08.001.
URL https://linkinghub.elsevier.com/retrieve/pii/S0968000407001892

[193] Vlasáková J, Nováková Z, Rossmeislová L, Kahle M, Hozák P, & Hodný Z. Histone deacetylase inhibitors suppress IFNα-induced up-regulation of promyelocytic leukemia protein. *Blood*, 109(4):1373–1380 (2007). doi:10.1182/blood-2006-02-003418.
URL https://ashpublications.org/blood/article/109/4/1373/23413/Histone-deacetylase-inhibitors-suppress

[194] Lorenz D R, Misra V, & Gabuzda D. Transcriptomic analysis of monocytes from HIV-positive men on antiretroviral therapy reveals effects of tobacco smoking on interferon and stress response systems associated with depressive symptoms. *Human Genomics*, 13(1):59 (2019). doi:10.1186/s40246-019-0247-x.
URL https://humgenomics.biomedcentral.com/articles/10.1186/s40246-019-0247-x

[195] Kim J O, Park J H, Kim T *et al.* A novel system-level approach using RNA-sequencing data identifies miR-30-5p and miR-142a-5p as key regulators of apoptosis in myocardial infarction. *Scientific Reports*, 8(1):14638 (2018). doi:10.1038/s41598-018-33020-x.
URL http://www.nature.com/articles/s41598-018-33020-x

[196] Makino Y, Yoon J H, Bae E, Kato M, Miyazawa K, Ohira T, Ikeda N, Kuroda M, & Mamura M. Repression of Smad3 by Stat3 and c-Ski/SnoN induces gefitinib resistance in lung adenocarcinoma. *Biochemical and Biophysical Research Communications*, 484(2):269–277 (2017). doi:10.1016/j.bbrc.2017.01.093.
URL https://linkinghub.elsevier.com/retrieve/pii/S0006291X17301432

[197] Xu Y, Yue W, Yao Shugart Y *et al.* Exploring Transcription Factors-microRNAs Co-regulation Networks in Schizophrenia. *Schizophrenia Bulletin*, 42(4):1037–1045 (2016). doi:10.1093/schbul/sbv170.
URL https://academic.oup.com/schizophreniabulletin/article-lookup/doi/10.1093/schbul/sbv170

[198] Song H M, Park G H, Eo H J, & Jeong J B. Naringenin-Mediated ATF3 Expression Contributes to Apoptosis in Human Colon Cancer. *Biomolecules & Therapeutics*, 24(2):140–146 (2016). doi:10.4062/biomolther.2015.109.
URL http://www.biomolther.org/journal/DOIx.php?id=10.4062/biomolther.2015.109

[199] Sun M M, Wang Y C, Li Y, Guo X D, Chen Y M, & Zhang Z Z. Effect of ATF3-deletion on apoptosis of cultured retinal ganglion cells. *International journal of ophthalmology*, 10(5):691–695 (2017). doi:10.18240/ijo.2017.05.05.
URL http://www.ijo.cn/gjyken/ch/reader/view{_}abstract.aspx?file{_}no=20170505{&}flag=1; http://www.ncbi.nlm.nih.gov/pubmed/28546922; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5437453

[200] Labzin L I, Schmidt S V, Masters S L *et al.* ATF3 Is a Key Regulator of Macrophage IFN Responses. *The Journal of Immunology*, 195(9):4446–4455 (2015). doi:10.4049/jimmunol.1500204.
URL http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1500204

[201] Glal D, Sudhakar J N, Lu H H, Liu M C, Chiang H Y, Liu Y C, Cheng C F, & Shui J W. ATF3 Sustains IL-22-Induced STAT3 Phosphorylation to Maintain Mucosal Immunity Through Inhibiting Phosphatases. *Frontiers in Immunology*, 9 (2018). doi:10.3389/fimmu.2018.02522.
URL https://www.frontiersin.org/article/10.3389/fimmu.2018.02522/full

[202] Heinrich R, Hertz R, Zemel E, Mann I, Brenner L, Massarweh A, Berlin S, & Perlman I. ATF3 Regulates the Expression of AChE During Stress. *Frontiers in Molecular Neuroscience*, 11 (2018). doi:10.3389/fnmol.2018.00088.
URL http://journal.frontiersin.org/article/10.3389/fnmol.2018.00088/full

[203] Huang C Y, Chen J J, Wu J S, Tsai H D, Lin H, Yan Y T, Hsu C Y, Ho Y S, & Lin T N. Novel Link of Anti-apoptotic ATF3 with Pro-apoptotic CTMP in the Ischemic Brain. *Molecular Neurobiology*, 51(2):543–557 (2015). doi:10.1007/s12035-014-8710-0.
URL http://link.springer.com/10.1007/s12035-014-8710-0

[204] Ghaleb A M & Yang V W. Krüppel-like factor 4 (KLF4): What we currently know. *Gene*, 611:27–37 (2017). doi:10.1016/j.gene.2017.02.025.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378111917301142

[205] Yin R H, Yu J T, & Tan L. The Role of SORL1 in Alzheimer's Disease. *Molecular Neurobiology*, 51(3):909–918 (2015). doi:10.1007/s12035-014-8742-5.
URL http://link.springer.com/10.1007/s12035-014-8742-5

[206] Sakai S, Nakaseko C, Takeuchi M *et al.* Circulating soluble LR11/SorLA levels are highly increased and ameliorated by chemotherapy in acute leukemias. *Clinica Chimica Acta*, 413(19-20):1542–1548 (2012). doi:10.1016/j.cca.2012.06.025.
URL https://linkinghub.elsevier.com/retrieve/pii/S0009898112003300

[207] Larsen J V & Petersen C M. SorLA in Interleukin-6 Signaling and Turnover. *Molecular and Cellular Biology*, 37(11) (2017). doi:10.1128/MCB.00641-16.
URL http://mcb.asm.org/lookup/doi/10.1128/MCB.00641-16

[208] Rabe B, Chalaris A, May U, Waetzig G H, Seegert D, Williams A S, Jones S A, Rose-John S, & Scheller J. Transgenic blockade of interleukin 6 transsignaling abrogates inflammation. *Blood*, 111(3):1021–1028 (2008). doi:10.1182/blood-2007-07-102137.
URL https://ashpublications.org/blood/article/111/3/1021/25448/Transgenic-blockade-of-interleukin-6

[209] Zhang C, Xiao C, Dang E *et al.* CD100–Plexin-B2 Promotes the Inflammation in Psoriasis by Activating NF-κB and the Inflammasome in Keratinocytes. *Journal of Investigative Der-*

matology, 138(2):375–383 (2018). doi:10.1016/j.jid.2017.09.005.
URL https://linkinghub.elsevier.com/retrieve/pii/S0022202X17329494

[210] Yu W, Goncalves K A, Li S *et al.* Plexin-B2 Mediates Physiologic and Pathologic Functions of Angiogenin. *Cell*, 171(4):849–864.e25 (2017). doi:10.1016/j.cell.2017.10.005.
URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417311893

[211] Roney K E, O'Connor B P, Wen H *et al.* Plexin-B2 Negatively Regulates Macrophage Motility, Rac, and Cdc42 Activation. *PLoS ONE*, 6(9):e24795 (2011). doi:10.1371/journal.pone.0024795.
URL https://dx.plos.org/10.1371/journal.pone.0024795

[212] Akbar N, Digby J E, Cahill T J *et al.* Endothelium-derived extracellular vesicles promote splenic monocyte mobilization in myocardial infarction. *JCI Insight*, 2(17) (2017). doi:10.1172/jci.insight.93344.
URL https://insight.jci.org/articles/view/93344

[213] Takahashi K & Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4):663–676 (2006). doi:10.1016/j.cell.2006.07.024.
URL https://linkinghub.elsevier.com/retrieve/pii/S0092867406009767

[214] Feinberg M W, Cao Z, Wara A K, Lebedeva M A, SenBanerjee S, & Jain M K. Kruppel-like Factor 4 Is a Mediator of Proinflammatory Signaling in Macrophages. *Journal of Biological Chemistry*, 280(46):38247–38258 (2005). doi:10.1074/jbc.M509378200.
URL http://www.jbc.org/lookup/doi/10.1074/jbc.M509378200

[215] Liu J, Liu Y, Zhang H, Chen G, Wang K, & Xiao X. KLF4 PROMOTES THE EXPRESSION, TRANSLOCATION, AND RELEASE OF HMGB1 IN RAW264.7 MACROPHAGES IN RESPONSE TO LPS. *Shock*, p. 1 (2008). doi:10.1097/shk.0b013e318162bef7.
URL http://journals.lww.com/00024382-900000000-99762

[216] Liu J, Yang T, Liu Y, Zhang H, Wang K, Liu M, Chen G, & Xiao X. Krüppel-like factor 4 inhibits the expression of interleukin-1 beta in lipopolysaccharide-induced RAW264.7 macrophages. *FEBS Letters*, 586(6):834–840 (2012). doi:10.1016/j.febslet.2012.02.003.
URL http://doi.wiley.com/10.1016/j.febslet.2012.02.003

[217] Feinberg M W, Wara A K, Cao Z *et al.* The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *The EMBO Journal*, 26(18):4138–4148 (2007). doi:10.1038/sj.emboj.7601824.
URL http://emboj.embopress.org/cgi/doi/10.1038/sj.emboj.7601824

[218] Alder J K, Georgantas R W, Hildreth R L, Kaplan I M, Morisot S, Yu X, McDevitt M, & Civin C I. Kruppel-Like Factor 4 Is Essential for Inflammatory Monocyte Differentiation In Vivo. *The Journal of Immunology*, 180(8):5645–5652 (2008). doi:10.4049/jimmunol.180.8.5645.
URL http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.180.8.5645

[219] Kurotaki D, Osato N, Nishiyama A *et al.* Essential role of the IRF8-KLF4 transcription factor cascade in murine monocyte differentiation. *Blood*, 121(10):1839–1849 (2013). doi:10.1182/blood-2012-06-437863.
URL https://ashpublications.org/blood/article/121/10/1839/31201/Essential-role-of-the-IRF8KLF4-transcription

[220] Spalinger M R, Manzini R, Hering L *et al.* PTPN2 Regulates Inflammasome Activation and Controls Onset of Intestinal Inflammation and Colon Cancer. *Cell Reports*, 22(7):1835–1848 (2018). doi:10.1016/j.celrep.2018.01.052.
URL https://linkinghub.elsevier.com/retrieve/pii/S2211124718301013

[221] Kim M, Morales L, Jang I S, Cho Y Y, & Kim D. Protein Tyrosine Phosphatases as Potential Regulators of STAT3 Signaling. *International Journal of Molecular Sciences*, 19(9):2708 (2018). doi:10.3390/ijms19092708.
URL http://www.mdpi.com/1422-0067/19/9/2708

[222] Ben-Efraim I, Zhou Q, Wiedmer T, Gerace L, & Sims P J. Phospholipid Scramblase 1 Is Imported into the Nucleus by a Receptor-Mediated Pathway and Interacts with DNA † , ‡. *Biochemistry*, 43(12):3518–3526 (2004). doi:10.1021/bi0356911.
URL https://pubs.acs.org/doi/10.1021/bi0356911

[223] Huang Y, Zhao Q, Zhou C X, Gu Z M, Li D, Xu H Z, Sims P J, Zhao K W, & Chen G Q. Antileukemic roles of human phospholipid scramblase 1 gene, evidence from inducible PLSCR1-expressing leukemic cells. *Oncogene*, 25(50):6618–6627 (2006). doi:10.1038/sj.onc.1209677.
URL http://www.nature.com/articles/1209677

[224] Dong B, Zhou Q, Zhao J *et al.* Phospholipid Scramblase 1 Potentiates the Antiviral Activity of Interferon. *Journal of Virology*, 78(17):8983–8993 (2004). doi:10.1128/JVI.78.17.8983-8993.2004.
URL http://jvi.asm.org/cgi/doi/10.1128/JVI.78.17.8983-8993.2004

[225] Suzuki E, Amengual O, Atsumi T *et al.* Increased expression of phospholipid scramblase 1 in monocytes from patients with systemic lupus erythematosus. *The Journal of rheumatology*, 37(8):1639–45 (2010). doi:10.3899/jrheum.091420.
URL http://www.jrheum.org/lookup/doi/10.3899/jrheum.091420; http://www.ncbi.nlm.nih.gov/pubmed/20516018

[226] Herate C, Ramdani G, Grant N J, Marion S, Gasman S, Niedergang F, Benichou S, & Bouchet J. Phospholipid Scramblase 1 Modulates FcR-Mediated Phagocytosis in Differentiated Macrophages. *PLOS ONE*, 11(1):e0145617 (2016). doi:10.1371/journal.pone.0145617.
URL http://dx.plos.org/10.1371/journal.pone.0145617

[227] Regad T, Saib A, Lallemand-Breitenbach V, Pandolfi P P, de Thé H, & Chelbi-Alix M K. PML mediates the interferon-induced antiviral state against a complex retrovirus via its association with the viral transactivator. *The EMBO journal*, 20(13):3495–505 (2001). doi:10.1093/emboj/20.13.3495.
URL http://emboj.embopress.org/cgi/doi/10.1093/emboj/20.13.3495; http://www.ncbi.nlm.nih.gov/pubmed/11432836; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC125516

[228] Chee A V, Lopez P, Pandolfi P P, & Roizman B. Promyelocytic Leukemia Protein Mediates Interferon-Based Anti-Herpes Simplex Virus 1 Effects. *Journal of Virology*, 77(12):7101–7105 (2003). doi:10.1128/JVI.77.12.7101-7105.2003.
URL http://jvi.asm.org/cgi/doi/10.1128/JVI.77.12.7101-7105.2003

[229] Chen Y, Wright J, Meng X, & Leppard K N. Promyelocytic Leukemia Protein Isoform II Promotes Transcription Factor Recruitment To Activate Interferon Beta and Interferon-Responsive Gene Expression. *Molecular and Cellular Biology*, 35(10):1660–1672 (2015). doi:10.1128/MCB.01478-14.
URL http://mcb.asm.org/lookup/doi/10.1128/MCB.01478-14

[230] Lo Y H, Huang Y W, Wu Y H *et al.* Selective inhibition of the NLRP3 inflammasome by targeting to promyelocytic leukemia protein in mouse and human. *Blood*, 121(16):3185–94 (2013). doi:10.1182/blood-2012-05-432104.
URL http://www.ncbi.nlm.nih.gov/pubmed/23430110

[231] Lunardi A, Gaboli M, Giorgio M *et al.* A Role for PML in Innate Immunity. *Genes & Cancer*, 2(1):10–19 (2011). doi:10.1177/1947601911402682.
URL http://gan.sagepub.com/lookup/doi/10.1177/1947601911402682

[232] Iliopoulos D, Hirsch H A, & Struhl K. An Epigenetic Switch Involving NF-κB, Lin28, Let-7 MicroRNA, and IL6 Links Inflammation to Cell Transformation. *Cell*, 139(4):693–706 (2009). doi:10.1016/j.cell.2009.10.014.
URL https://linkinghub.elsevier.com/retrieve/pii/S0092867409013026

[233] Li D, De S, Li D, Song S, Matta B, & Barnes B J. Specific detection of interferon regulatory factor 5 (IRF5): A case of antibody inequality. *Scientific Reports*, 6(1):31002 (2016). doi:

10.1038/srep31002.
URL http://www.nature.com/articles/srep31002

[234] Takaoka A, Yanai H, Kondo S *et al.* Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature*, 434(7030):243–249 (2005). doi:10.1038/nature03308.
URL http://www.nature.com/articles/nature03308

[235] Krausgruber T, Blazek K, Smallie T, Alzabin S, Lockstone H, Sahgal N, Hussell T, Feldmann M, & Udalova I A. IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. *Nature Immunology*, 12(3):231–238 (2011). doi:10.1038/ni.1990.
URL http://www.nature.com/articles/ni.1990

[236] Chistiakov D A, Myasoedova V A, Revin V V, Orekhov A N, & Bobryshev Y V. The impact of interferon-regulatory factors to macrophage differentiation and polarization into M1 and M2. *Immunobiology*, 223(1):101–111 (2018). doi:10.1016/j.imbio.2017.10.005.
URL http://dx.doi.org/10.1016/j.imbio.2017.10.005

[237] Wang N, Liang H, & Zen K. Molecular Mechanisms That Influence the Macrophage M1/M2 Polarization Balance. *Frontiers in Immunology*, 5 (2014). doi:10.3389/fimmu.2014.00614.
URL http://journal.frontiersin.org/article/10.3389/fimmu.2014.00614/abstract

[238] Negishi H, Fujita Y, Yanai H, Sakaguchi S, Ouyang X, Shinohara M, Takayanagi H, Ohba Y, Taniguchi T, & Honda K. Evidence for licensing of IFN- -induced IFN regulatory factor 1 transcription factor by MyD88 in Toll-like receptor-dependent gene induction program. *Proceedings of the National Academy of Sciences*, 103(41):15136–15141 (2006). doi:10.1073/pnas.0607181103.
URL http://www.pnas.org/cgi/doi/10.1073/pnas.0607181103

[239] Courties G, Heidt T, Sebas M *et al.* In vivo silencing of the transcription factor IRF5 reprograms the macrophage phenotype and improves infarct healing. *Journal of the American College of Cardiology*, 63(15):1556–66 (2014). doi:10.1016/j.jacc.2013.11.023.
URL http://www.ncbi.nlm.nih.gov/pubmed/24361318; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3992176

[240] Fujiwara Y, Hizukuri Y, Yamashiro K, Makita N, Ohnishi K, Takeya M, Komohara Y, & Hayashi Y. Guanylate-binding protein 5 is a marker of interferon-γ-induced classically activated macrophages. *Clinical & Translational Immunology*, 5(11):e111 (2016). doi:10.1038/cti.2016.59.
URL http://doi.wiley.com/10.1038/cti.2016.59

[241] Shenoy A R, Wellington D A, Kumar P, Kassa H, Booth C J, Cresswell P, & MacMicking J D. GBP5 Promotes NLRP3 Inflammasome Assembly and Immunity in Mammals. *Science*,

336(6080):481–485 (2012). doi:10.1126/science.1217141.
URL https://www.sciencemag.org/lookup/doi/10.1126/science.1217141

[242] Feng J, Cao Z, Wang L, Wan Y, Peng N, Wang Q, Chen X, Zhou Y, & Zhu Y. Inducible GBP5 Mediates the Antiviral Response via Interferon-Related Pathways during Influenza A Virus Infection. *Journal of Innate Immunity*, 9(4):419–435 (2017). doi:10.1159/000460294.
URL https://www.karger.com/Article/FullText/460294

[243] Krapp C, Hotter D, Gawanbacht A *et al.* Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity. *Cell Host & Microbe*, 19(4):504–514 (2016). doi:10.1016/j.chom.2016.02.019.
URL https://linkinghub.elsevier.com/retrieve/pii/S1931312816300609

[244] Corsetti P P, de Almeida L A, Gonçalves A N A, Gomes M T R, Guimarães E S, Marques J T, & Oliveira S C. miR-181a-5p Regulates TNF-α and miR-21a-5p Influences Gualynate-Binding Protein 5 and IL-10 Expression in Macrophages Affecting Host Control of Brucella abortus Infection. *Frontiers in Immunology*, 9 (2018). doi:10.3389/fimmu.2018.01331.
URL https://www.frontiersin.org/article/10.3389/fimmu.2018.01331/full

[245] Hamada M, Tsunakawa Y, Jeon H, Yadav M K, & Takahashi S. Role of MafB in macrophages. *Experimental Animals*, 69(1):1–10 (2020). doi:10.1538/expanim.19-0076.
URL https://www.jstage.jst.go.jp/article/expanim/69/1/69{_}19-0076/{_}article

[246] Aziz A, Soucie E, Sarrazin S, & Sieweke M H. MafB/c-Maf Deficiency Enables Self-Renewal of Differentiated Functional Macrophages. *Science*, 326(5954):867–871 (2009). doi:10.1126/science.1176056.
URL ???; https://www.sciencemag.org/lookup/doi/10.1126/science.1176056

[247] Shichita T, Ito M, Morita R, Komai K, Noguchi Y, Ooboshi H, Koshida R, Takahashi S, Kodama T, & Yoshimura A. MAFB prevents excess inflammation after ischemic stroke by accelerating clearance of damage signals through MSR1. *Nature Medicine*, 23(6):723–732 (2017). doi:10.1038/nm.4312.
URL http://dx.doi.org/10.1038/nm.4312

[248] Liu T M, Wang H, Zhang D N, & Zhu G Z. Transcription Factor MafB Suppresses Type I Interferon Production by CD14+ Monocytes in Patients With Chronic Hepatitis C. *Frontiers in Microbiology*, 10 (2019). doi:10.3389/fmicb.2019.01814.
URL https://www.frontiersin.org/article/10.3389/fmicb.2019.01814/full

[249] Tran M T N, Hamada M, Jeon H *et al.* MafB is a critical regulator of complement component C1q. *Nature Communications*, 8(1):1700 (2017). doi:10.1038/s41467-017-01711-0.
URL http://www.nature.com/articles/s41467-017-01711-0

[250] Sato M, Shibata Y, Inoue S *et al.* MafB enhances efferocytosis in RAW264.7 macrophages by regulating Axl expression. *Immunobiology*, 223(1):94–100 (2018). doi:10.1016/j.imbio.2017.10.007.
URL https://linkinghub.elsevier.com/retrieve/pii/S0171298517301420

[251] Tozaki-Saitoh H, Masuda J, Kawada R, Kojima C, Yoneda S, Masuda T, Inoue K, & Tsuda M. Transcription factor MafB contributes to the activation of spinal microglia underlying neuropathic pain development. *Glia*, 67(4):729–740 (2019). doi:10.1002/glia.23570.
URL http://doi.wiley.com/10.1002/glia.23570

[252] Jablonski K A, Gaudet A D, Amici S A, Popovich P G, & Guerau-de Arellano M. Control of the Inflammatory Macrophage Transcriptional Signature by miR-155. *PLOS ONE*, 11(7):e0159724 (2016). doi:10.1371/journal.pone.0159724.
URL https://dx.plos.org/10.1371/journal.pone.0159724

[253] Kuriakose T & Kanneganti T D. ZBP1: Innate Sensor Regulating Cell Death and Inflammation. *Trends in Immunology*, 39(2):123–134 (2018). doi:10.1016/j.it.2017.11.002.
URL https://linkinghub.elsevier.com/retrieve/pii/S1471490617302120

[254] Yang D, Liang Y, Zhao S, Ding Y, Zhuang Q, Shi Q, Ai T, Wu S Q, & Han J. ZBP1 mediates interferon-induced necroptosis. *Cellular & Molecular Immunology*, pp. 1–13 (2019). doi:10.1038/s41423-019-0237-x.
URL http://dx.doi.org/10.1038/s41423-019-0237-x; http://www.nature.com/articles/s41423-019-0237-x

[255] Maelfait J, Liverpool L, Bridgeman A, Ragan K B, Upton J W, & Rehwinkel J. Sensing of viral and endogenous RNA by ZBP1/DAI induces necroptosis. *The EMBO journal*, 36(17):2529–2543 (2017). doi:10.15252/embj.201796476.
URL https://onlinelibrary.wiley.com/doi/abs/10.15252/embj.201796476; http://www.ncbi.nlm.nih.gov/pubmed/5579359; http://www.ncbi.nlm.nih.gov/pubmed/28716805; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5579359

[256] Kuriakose T, Man S M, Subbarao Malireddi R K, Karki R, Kesavardhana S, Place D E, Neale G, Vogel P, & Kanneganti T D. ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Science Immunology*, 1(2):aag2045–aag2045 (2016). doi:10.1126/sciimmunol.aag2045.
URL https://immunology.sciencemag.org/lookup/doi/10.1126/sciimmunol.aag2045

[257] Thapa R J, Basagoudanavar S H, Nogusa S, Irrinki K, Mallilankaraman K, Slifker M J, Beg A A, Madesh M, & Balachandran S. NF- B Protects Cells from Gamma Interferon-Induced RIP1-Dependent Necroptosis. *Molecular and Cellular Biology*, 31(14):2934–2946 (2011).

doi:10.1128/MCB.05445-11.
URL http://mcb.asm.org/cgi/doi/10.1128/MCB.05445-11

[258] Paul R, Zhang Z G, Eliceiri B P, Jiang Q, Boccia A D, Zhang R L, Chopp M, & Cheresh D A. Src deficiency or blockade of Src activity in mice provides cerebral protection following stroke. *Nature Medicine*, 7(2):222–227 (2001). doi:10.1038/84675.

[259] Zan L, Zhang X, Xi Y, Wu H, Song Y, Teng G, Li H, Qi J, & Wang J. Src regulates angiogenic factors and vascular permeability after focal cerebral ischemia–reperfusion. *Neuroscience*, 262(2):118–128 (2014). doi:10.1016/j.neuroscience.2013.12.060.
URL https://linkinghub.elsevier.com/retrieve/pii/S0022202X15370834; https://linkinghub.elsevier.com/retrieve/pii/S0306452213010932

[260] Kumar A, Jaggi A S, & Singh N. Pharmacological investigations on possible role of Src kinases in neuroprotective mechanism of ischemic postconditioning in mice. *International Journal of Neuroscience*, 124(10):777–786 (2014). doi:10.3109/00207454.2013.879869.
URL http://www.tandfonline.com/doi/full/10.3109/00207454.2013.879869

[261] Adam A P, Lowery A M, Martino N, Alsaffar H, & Vincent P A. Src Family Kinases Modulate the Loss of Endothelial Barrier Function in Response to TNF-α: Crosstalk with p38 Signaling. *PLOS ONE*, 11(9):e0161975 (2016). doi:10.1371/journal.pone.0161975.
URL https://dx.plos.org/10.1371/journal.pone.0161975

[262] Brown M T & Cooper J A. Regulation, substrates and functions of src. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1287(2-3):121–149 (1996). doi:10.1016/0304-419X(96)00003-0.
URL https://linkinghub.elsevier.com/retrieve/pii/0304419X96000030

[263] Okutani D, Lodyga M, Han B, & Liu M. Src protein tyrosine kinase family and acute inflammatory responses. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 291(2):L129–L141 (2006). doi:10.1152/ajplung.00261.2005.
URL https://www.physiology.org/doi/10.1152/ajplung.00261.2005

[264] Hama K, Fujiwara Y, Morita M *et al.* Profiling and Imaging of Phospholipids in Brains of Abcd1 -Deficient Mice. *Lipids*, 53(1):85–102 (2018). doi:10.1002/lipd.12022.
URL http://doi.wiley.com/10.1002/lipd.12022

[265] Lauer A, Da X, Hansen M B *et al.* ABCD1 dysfunction alters white matter microvascular perfusion. *Brain*, 140(12):3139–3152 (2017). doi:10.1093/brain/awx262.
URL https://academic.oup.com/brain/article/140/12/3139/4608962

[266] Orchard P J, Markowski T W, Higgins L, Raymond G V, Nascene D R, Miller W P, Pierpont E I, & Lund T C. Association between APOE4 and biomarkers in cerebral adrenoleukodystrophy. *Scientific Reports*, 9(1):7858 (2019). doi:10.1038/s41598-019-44140-3.
URL http://www.nature.com/articles/s41598-019-44140-3

[267] Lei Y, Liu L, Zhang S *et al.* Hdac7 promotes lung tumorigenesis by inhibiting Stat3 activation. *Molecular Cancer*, 16(1):170 (2017). doi:10.1186/s12943-017-0736-2.
URL https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-017-0736-2

[268] Chang S, Young B D, Li S, Qi X, Richardson J A, & Olson E N. Histone Deacetylase 7 Maintains Vascular Integrity by Repressing Matrix Metalloproteinase 10. *Cell*, 126(2):321–334 (2006). doi:10.1016/j.cell.2006.05.040.
URL https://linkinghub.elsevier.com/retrieve/pii/S0092867406008130

[269] Peixoto P, Blomme A, Costanza B *et al.* HDAC7 inhibition resets STAT3 tumorigenic activity in human glioblastoma independently of EGFR and PTEN: new opportunities for selected targeted therapies. *Oncogene*, 35(34):4481–4494 (2016). doi:10.1038/onc.2015.506.
URL http://www.nature.com/articles/onc2015506

[270] Barneda-Zahonero B, Collazo O, Azagra A *et al.* The transcriptional repressor HDAC7 promotes apoptosis and c-Myc downregulation in particular types of leukemia and lymphoma. *Cell Death & Disease*, 6(2):e1635–e1635 (2015). doi:10.1038/cddis.2014.594.
URL http://www.nature.com/articles/cddis2014594

[271] Barneda-Zahonero B, Román-González L, Collazo O *et al.* HDAC7 Is a Repressor of Myeloid Genes Whose Downregulation Is Required for Transdifferentiation of Pre-B Cells into Macrophages. *PLoS Genetics*, 9(5):e1003503 (2013). doi:10.1371/journal.pgen.1003503.
URL https://dx.plos.org/10.1371/journal.pgen.1003503

[272] Miano J M & Berk B C. HDAC7 supports vascular integrity. *Nature Medicine*, 12(9):997–998 (2006). doi:10.1038/nm0906-997.

[273] Ha C H, Jhun B S, Kao H Y, & Jin Z G. VEGF Stimulates HDAC7 Phosphorylation and Cytoplasmic Accumulation Modulating Matrix Metalloproteinase Expression and Angiogenesis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 28(10):1782–1788 (2008). doi:10.1161/ATVBAHA.108.172528.
URL https://www.ahajournals.org/doi/10.1161/ATVBAHA.108.172528

[274] Lee D Y, Lin T E, Lee C I, Zhou J, Huang Y H, Lee P L, Shih Y T, Chien S, & Chiu J J. MicroRNA-10a is crucial for endothelial response to different flow patterns via interaction of retinoid acid receptors and histone deacetylases. *Proceedings of the National Academy of*

Sciences, 114(8):2072–2077 (2017). doi:10.1073/pnas.1621425114.
URL http://www.pnas.org/lookup/doi/10.1073/pnas.1621425114

[275] Fan Y, Siklenka K, Arora S K, Ribeiro P, Kimmins S, & Xia J. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Research*, 44(W1):W135–W141 (2016). doi:10.1093/nar/gkw288.
URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw288

[276] Tokar T, Pastrello C, Rossos A E M, Abovsky M, Hauschild A C, Tsay M, Lu R, & Jurisica I. mirDIP 4.1—integrative database of human microRNA target predictions. *Nucleic Acids Research*, 46(D1):D360–D370 (2018). doi:10.1093/nar/gkx1144.
URL http://academic.oup.com/nar/article/46/D1/D360/4670951

[277] Neo4j I. Introducing Neo4j 4.0 (2020).
URL https://neo4j.com/press-releases/announcing-neo4j-4-0/

[278] Kosik K S. *High Throughput Profiling of the microRNA - mRNA Interactome in the Developing Human Brain* (2017). Non-coding RNAs in Nervous System Development, Plasticity and Disease, SPP1738, June 21-24, Marburg, Germany.

[279] Kosik K S. Publication List (2020).
URL https://ken-kosik.mcdb.ucsb.edu/publications

[280] Vieth B, Ziegenhain C, Parekh S, Enard W, & Hellmann I. powsimR: power analysis for bulk and single cell RNA-seq experiments. *Bioinformatics*, 33(21):3486–3488 (2017). doi:10.1093/bioinformatics/btx435.
URL https://academic.oup.com/bioinformatics/article/33/21/3486/3952669

[281] Sticht C, De La Torre C, Parveen A, & Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLOS ONE*, 13(10):e0206239 (2018). doi:10.1371/journal.pone.0206239.
URL http://dx.plos.org/10.1371/journal.pone.0206239

[282] Ding J, Li X, & Hu H. TarPmiR: a new approach for microRNA target site prediction. *Bioinformatics*, 32(18):2768–2775 (2016). doi:10.1093/bioinformatics/btw318.
URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw318

[283] Almuttaqi H & Udalova I A. Advances and challenges in targeting IRF5, a key regulator of inflammation. *The FEBS Journal*, 286(9):1624–1637 (2019). doi:10.1111/febs.14654.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/febs.14654

[284] Panizzi P, Swirski F K, Figueiredo J L, Waterman P, Sosnovik D E, Aikawa E, Libby P, Pittet M, Weissleder R, & Nahrendorf M. Impaired Infarct Healing in Atherosclerotic Mice With Ly-6ChiMonocytosis. *Journal of the American College of Cardiology*, 55(15):1629–1638 (2010). doi:10.1016/j.jacc.2009.08.089.
URL https://linkinghub.elsevier.com/retrieve/pii/S0735109710004547

List of Figures

1.1 Cholinergic Projections. . . . 2
1.2 The Neurokine Pathway. . . . 14

2.1 Graph database organisation. . . . 24
2.2 Graph database organisation. . . . 29
2.3 microRNA Species Homology. . . . 34

3.1 The blood-brain-barrier. . . . 40
3.2 Single-Cell Sequencing of CNS Tissues. . . . 43
3.3 Clusters of Cholinergic Transcripts in Single-Cell Sequencing. . . . 45
3.4 LA-N-2 and LA-N-5, Time-dose Curve and Differentiation Timeline. . . . 48
3.5 Small RNA Sequencing - Read Count, Quality, and Length. . . . 51
3.6 MD Plot Shrinkage Comparison. . . . 53
3.7 Differentially Expressed microRNAs in LA-N-2 and LA-N-5. . . . 54
3.8 Timeline Differential Expression. . . . 55
3.9 miRNAs Differentially Expressed Between LA-N-2 and LA-N-5. . . . 57
3.10 Differential Expression miRNA Family Enrichment . . . 60
3.11 LA-N-2 / LA-N-5 Full Connectome. . . . 62
3.12 GO Enrichment of Diverging Genes. . . . 67
3.13 The cholinergic/neurokine interface. . . . 69

4.1 Cholinergic-associated Small RNA ECDF Curves. . . . 76
4.2 Large RNA Differential Expression Gene Ontology Enrichment. . . . 78
4.3 Small and Large RNA Differential Expression and tRF Properties. . . . 79
4.4 Functional Characterisation of Hierarchical Clusters in Blood Cell Small RNA Expression. . . . 83
4.5 Large RNA Expression Patterns in Blood-Borne Cells. . . . 84
4.6 Small RNA Expression Patterns in Blood-Borne Cells. . . . 87
4.7 Small RNA Targeting of Transcription Factors in $CD14^{+}$ Monocytes. . . . 90
4.8 Small RNA Feedforward Loop Theory. . . . 94
4.9 Complete Feedforward Loop Network of Differentially Expressed Transcription Factors in $CD14^{+}$ Monocytes. . . . 96

4.10 FFL module Gene Ontology Enrichment. 107
4.11 Complete Feedforward Loop Network of Differentially Expressed Transcription Factors in $CD14^{+}$ Monocytes, Annotated. 108

5.1 Number of Publications on Neuroinflammation by Year. 125
5.2 The Cholinergic/Neurokine miRNA Interface. 127
5.3 Coherence of miRNA Feedforward Loops. 129
5.4 Comparison of Annotated Feedforward Loop Network and t-SNE Visualisation of Module GO Terms. 130

A

Transcription Factor Regulatory Circuits - Tissue Types

B

List of Primate-Specific Homologues of Human microRNAs

C

microRNA Differential Expression in LA-N-2 and LA-N-5

D

List of GO Terms from Analysis of Differentially Expressed Large RNA in Stroke

E

Examples of Presence/Absence Definition of Small RNA