# Small RNA dynamics
# in cholinergic systems

Thesis advisor: Professor Jochen Klein                    Sebastian Lobentanzer

# Small RNA dynamics in cholinergic systems

## Abstract

Lorem ipsum dolor sit amet, 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.

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

THIS IS THE DEDICATION.

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

Simon Beckett

# Acknowledgments

Thanks are due, for every scientist is not only standing on the shoulders of giants, but also on those of very real persons, without whom this dissertation would not have been possible. 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.

# 0
# Introduction

## 0.1   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[?]. 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«[?]. 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[?]. Dale and Loewi's joint effort in »Discoveries in 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 of functions have been attributed to cholinergic circuits in the central nervous system (CNS). Central nervous projections of cholinergic fibres were extensively mapped by M. Marsel Mesulam and others in the 1980s[?], 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 ??, from[?]).
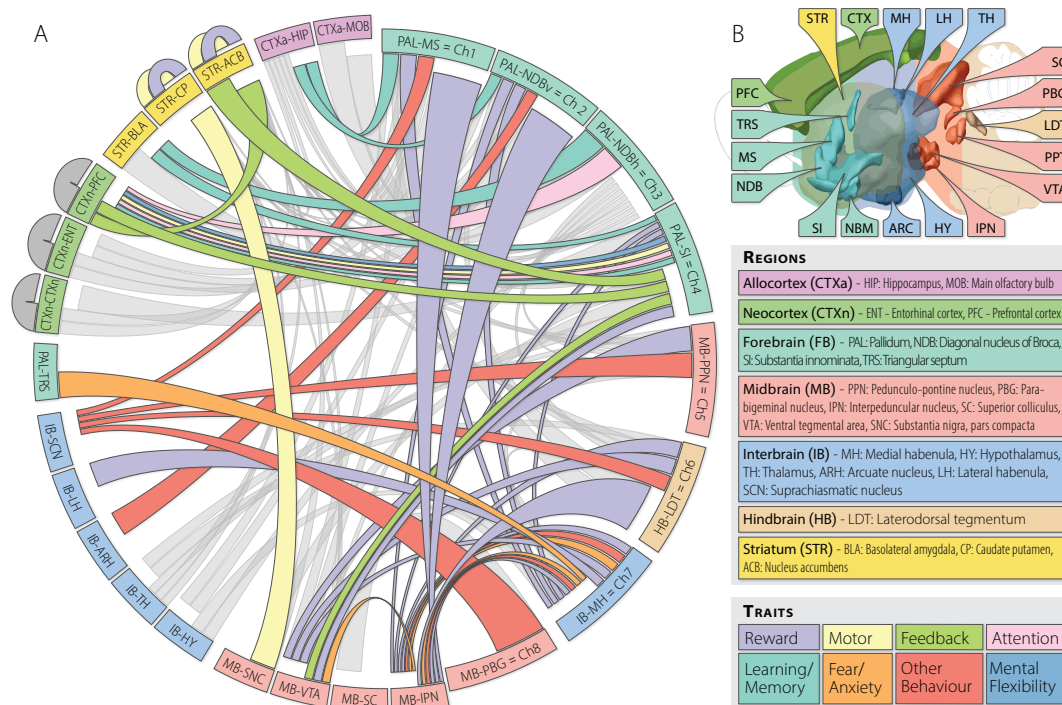
**Figure 1:** This is a figure that floats inline and here is its caption.

This holds particularly 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.

Disease? Subsection for disease?

### 0.1.1 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 neurotrophic 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[?], and the dependence of basal forebrain cholinergic neurons on retrograde NGF signalling was discovered in the 1980s[?].

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), leukaemia inhibitory factor (LIF), and interleukin 6 (IL-6), all of which coincidentally have been known under the acronym CDF. In the end of the 1980s, two groups of scientists (McManaman[?] and Rao[?]) 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 abbre-

viated 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. 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 »IL-6 signal transducer« IL6ST, also known as gp130. Every neurokine has its preferred constellation of soluble and transmembrane receptors: CNTF bind 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 from 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 is significant redundancy and crosstalk between those systems[?][?].

All receptor constellations result in a main effect of activation of the JAK/STAT cascade. More specifically, neurokines can activate janus kinases (JAK) 1 and 2 or the homologous tyrosine kinase (TYK) 2, and successively »signal transducer and activator of transcription« (STAT) 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 (XXX).

FIGURE STA1

## 0.2   Transcriptional Connectomics

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 capacity, 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 (2 420 nucleobases) 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* 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 subject of these methods and therefore will be briefly introduced: genes in the classical sense as the conveyors of cellular function by encoding for proteins; transcription factors, a subclass of protein coding genes that are able to regulate the expression of other genes; microRNAs (miRNAs), a class of small non-coding RNA 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 small non-coding RNA that has only recently been rediscovered and is significantly less well described regarding its functionality.

Where to put:

Distinguish neuronal connectomics from transcriptional connectomics

For the sake of simplicity, all descriptions of genomics and transcriptomics matters, genes, miRNAs, and tRFs in this dissertation are to be seen in the context of Homo sapiens, unless explicitly stated otherwise.

### 0.2.1  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 nucleobases. The regions containing these binding sites (about 100 - 1 000 bases 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 might 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 sequencing, computational approaches have become able to not only comprehensively predict TF-gene interactions, but to do so in a highly tissue-specific manner (see **??**). The human body is estimated to express up to 2 600 distinct DNA-binding proteins, most of them presumed TFs[?], although other studies give lower estimates.

## 0.2.2 MICRORNAs

While the first endogenous »small RNA with antisense complementarity« was described in 1993[?], miRNAs were only recognised as a distinct regulatory class of molecules in the early 2000s. They are typically between 18 and 22 nucleobase-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 inter-action 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 complementar-ity 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 con-tain 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. 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« na-ture. In those cases, seed pairing with the mRNA is often imperfect. To ameliorate this loss of sta-bility, compensation occurs typically by a secondary complementary structure after a small gap of non-complementary bases, leading to a »bridge«-type constellation. FIGURE? 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 ??).

miRNAs, similar to coding genes, are transcribed from loci on the genome, many inside introns or even exons of coding genes[?]. 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 ex-ported from the nucleus by the shuttle protein Exportin-5. In a final step in the cytosol, the RNAse »Dicer« removes the loop joining the 3' and 5' arms of the pre-miRNA, resulting in a duplex of mature miRNA, about 20 nucleotides 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).

miRNAs are organised and curated by means of a periodically updated web-based platform, miR-Base[?]. For Homo sapiens, miRBase v21 contains 2 588 mature miRNAs from XXX 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[?] 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.

Disease?

miRNA genes, in the same way as protein coding genes, can also 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 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 might be overenthusiastic).

Prediction?

### 0.2.3 TRANSFER RNA FRAGMENTS

Self-plagiarised: Recent studies have shown transfer RNA (tRNA) to be a major source of small non-coding RNA[?][?], including tRNA halves (tiRNAs) and smaller tRNA fragments (tRFs). tiRNAs derive from either end of the tRNA, and are created by angiogenin cleavage at the anticodon loop[?][?]. 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[?][?].

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

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[?] . 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 tumor-suppressive activity[?] . 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[?] .

History?

There is no consistent nomenclature yet to describe and organise tRFs, which are by nature more heterogeneous than miRNAs; considering their biogenesis, one tRNA molecule can be the origin of several hundred distinct tRF molecules. Multiple approaches are common in current literature, most prominently tRFs are tied to the parent tRNA and the amino acid coded for by this tRNA. For example, the 22-nucleotide-long 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[?] . 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. For this reason, the approach of Loher and colleagues[?] might 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« stands 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.

Disease?

? Levels of tRFs may be modulated even more rapidly than levels of miRs, since tRNA molecules are very abundant in the cell and generation of mature tRFs requires only enzymatic degradation of tRNA but no de-novo transcription of the molecule in the nucleus (citation).

## 0.2.4  Nested Multimodal Transcriptional Interactions - The Need for Connectomics

For every scientist trying to explain the cooperation between only these few factors in light of how genotype results in phenotype, or how dysregulation of any one factor leads to loss of homeostasis and eventually disease, multiple levels of obstacles have to be overcome. The ultimate aim of any such approach is the generation of a robust model for the studied phenomenon; the theoretical and practical hurdles to be surmounted to reach this goal are many. Theoretically, the more we know about the functioning of these intertwined systems, the more we understand how much there is still to learn.

For example, only recently it has 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. Through study of comprehensive genomic information of 394 tissue types in approximately 1 000 human primary cell, tissue, and culture samples (from the FANTOM5 consortium) it was estimated that genes in immune and nervous cells are controlled by a mean of 12 and 10 TFs each, and that TFs in nervous and immune cells influence expression of a mean of 175 and 160 genes, respectively (see ??).

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 sequencing, 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 miRs as functional modules associate with particular phenotypes in developmental and mature states[?]. 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 for 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[?], resulting in an effective doubling of currently known miRNAs.

These high numbers of potentially interacting players present computational challenges: If estimating the number of expressed genes in a human cell at 20 000 (and the number of TFs at a low 1 000), this makes for an estimated minimum of 200 000 »real« interactions in the possible $C = \frac{1\,000!}{10!(1\,000-10)!} \cdot 20\,000$, 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 miRs (2 588 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 $2\,588 \cdot 20\,000 = 51\,760\,000$ individual fields.

Combining all aspects 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 2 000 are TFs), 2 588 mature miRNAs, and a total of $2\,588 \cdot 20\,000 + 2\,000 \cdot 20\,000 = 91\,760\,000$ 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 **??**). 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 male and female cultured cells, or the blood of stroke victims.

*»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

# 1

# *miRNet*: Creation of a Comprehensive Connectomics Database

The need for bioinformatical support in connectomics is immediately obvious from 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[?] (see ??). One year later, the »regulatory circuits« enabling analysis of comprehensive TF-gene relationships in 394 human tissues were published[?] (see ??). These collections (as well as the data they were aggregated from) are the basis of the database further called *miRNet*, the development of which will be described in the following chapter.

## 1.1   Materials

All materials used in the creation of *miRNet* have been acquired from resources that are non-commercial, web-available, and open-source (in the case of code).

### 1.1.1   Gene Annotation

Even though »regular« protein coding genes have been know 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 mainly acronyms and fantasy-names (e.g. ChAT) is still widely popular because of its accessibility to us humans, but is also not particularly robust because of numerous synonym-constructs and instances of genes without names having to carry unwieldy systematic names.

- The American 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-centered 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 *miRNet* upon creation or during maintenance, for convenience and to minimise analysis time due to conflict resolution.

### 1.1.2 Transcription Factor Targeting

The FANTOM5 project has applied CAGE (5' cap analysis of gene expression) to a large number of human samples from diverse tissues to determine the accurate 5' ends of each transcript[?]. Knowledge of this fact enables accurate prediction of promoters 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[?]. 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); and 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

α7 receptor subunit) genes in raw FANTOM5 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 α7 homomeric ACh receptor conveys direct immune suppression by expression on monocytes . 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.

The entire collection of transcriptional activities in all tissues was downloaded[?] , and neuronal and immune tissues were entered into *miRNet*. The collected data comprises XX neuronal tissues and XX immune cell tissues (Appendix **??**), and XX TF-gene relationships in total.

### 1.1.3 microRNA Interactions

The content of miRWalk 2.0 is freely available online[?] ; however, there is no option of downloading the complete set. The targeting data thus was downloaded per miRNA 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[?] and miRTarBase[?] , 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 (see below).

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[?] ). 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[?] .

A statistical evaluation of the collected interaction data from miRWalk 2.0 showed vast differences in general prediction quantity (Table **??**) as well as prediction accuracy and sensitivity when compared to the validated subset of data (Table **??**). Since the ground truth is not known, this is an additional

| 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 1.1:** 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.

argument for the combination of multiple algorithms instead of the use of a single set. Apart from RNAhybrid and miRBridge, all algorithms were presented reasonable base hit frequencies and increases in the validated test set. Therefore, the remaining 10 algorithms were included in *miRNet* 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 1 to 10.

The collected (human) data comprises XX miRNA-gene targeting predictions (all 12 algorithms) and XX experimentally validated interactions (XX with evidence type »strong«).

### 1.1.4 DE-NOVO PREDICTION OF tRF INTERACTION

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[?][?], justifying an attempt to predict interactions in a comprehensive manner. Of the available options for nucleotide interaction prediction algorithms, TargetScan[?] 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 might 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[?], and through an additional algorithm allows the attribution of a score based on the branch length (on the species tree) of conserved targeting[?].

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 defini-

| 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 1.2:** 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 **??**) 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).

tive answer to the question as to where the seed region in tRFs might 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 were chopped into 7-base pieces (7mers), which is the lenght of miRNA seeds, and statistically improbable enough to appear in the genome at random; the average length of a human 3' UTR is 800 bases, so the probability of finding any 7mer randomly in any one 3' UTR is $p = \frac{800}{4^7} = 0.049$.

Describe Targetscan process

## 1.2 IMPLEMENTATION

For any biological question to be asked, the effectiveness of the bioinformatical query determines the practicality of the approach. Since computational resources 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. In the simple 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 miRs (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, such as distinct algorithms or the interaction between TFs and genes, require the addition of whole tables of similar size to the database, which is detrimental to effective use of space; and queries then also necessitate the transfer of information between those distinct tables (in SQL typically via a JOIN command), that 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.

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[?]; *miRNet* 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.

### 1.2.1 NEO4J: A GRAPH-BASED INFRASTRUCTURE

To display and query 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«, typically used to describe the class of the entry (such as »gene« or »miRNA«). This database organisation replicates the network-like structure of the biological data studied. Theoretically, this makes the database more likely to be efficient in the setting of transcriptional interactions, an estimation that turned out to be true. Neo4j combines the 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.

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. As soon as a queried node does not return any more edges that fit the query, the process can terminate and resources are freed for the next step. In practice, even in the very first prototype implementations, this accelerated standard case computations approximately thousand-fold, and was even able to accommodate advanced approaches in situations that were inaccessible in the tabular implementation.

### 1.2.2 HIGH-THROUGHPUT DATABASE GENERATION

Neo4j provides several API 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 `createDeferredConstraint()`, `assertPropertyIsUnique()`, and `createDeferredSchemaIndex()` commands 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, origin of data, 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.

### 1.2.3 Maintenance and Quality Control

All additional datasets, such as the TF regulatory circuits or tRF targeting predictions, were entered into *miRNet* 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, e.g. when miRNA names conflicted between old »\*« and new 3p/5p notation; all manual edits are documented in the code.

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. However, the entire Java-based code used in creation and maintenance is available in the code repository accompanying my first manuscript[?].

### 1.2.4 Transcription Factor Regulatory Circuits of Nervous and Immune Cells

Appendix **??**

### 1.2.5 De-Novo Prediction of Transfer RNA Fragment Targeting

### 1.2.6 Accumulation of microRNA targeting

## 1.3 Usage

### 1.3.1 Cypher Query Language

Neo4j uses a »language« akin to SQL, utilising keyphrases to issue commands, but combines it with a semi-graphical syntax to account for the graph-based layout of the data. For instance, 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

```
    MATCH (m:MIR)
WHERE m.name = 'hsa-miR-125b-5p'
RETURN m
    MATCH (m:MIR)-[r:TARGETS]->(g:GENE)
WHERE g.name IN {cholinergic_genes}
RETURN m, r, g
```

## 1.4 Statistical Approach to Connectomics

Lorem ipsum dolor sit amet, 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.

*This is some random quote to start off the chapter.*

Firstname lastname

# 2

# Single-cell sequencing?

Lorem ipsum dolor sit amet, 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.

## 2.1 Background

# 3

# microRNA Dynamics in Cholinergic Differentiation of Human Neuronal Cells

LOREM IPSUM DOLOR SIT AMET, 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.

## 3.1 BACKGROUND? - INTRO?

## 3.2 THE CELLULAR MODEL

?

*Nulla facilisi. In vel sem. Morbi id urna in diam dignis-*
*sim feugiat. Proin molestie tortor eu velit. Aliquam erat*
*volutpat. Nullam ultrices, diam tempus vulputate egestas,*
*eros pede varius leo.*

Quoteauthor Lastname

# 4

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

Lorem ipsum dolor sit amet, 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.

*This is some random quote to start off the chapter.*

Firstname lastname

# 5

# Discussion

Lorem ipsum dolor sit amet, 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.

## 5.1 Methods

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.

Proin at eros non eros adipiscing mollis. Donec semper turpis sed diam. Sed consequat ligula nec tortor. Integer eget sem. Ut vitae enim eu est vehicula gravida. Morbi ipsum ipsum, porta nec,

tempor id, auctor vitae, purus. Pellentesque neque. Nulla luctus erat vitae libero. Integer nec enim. Phasellus aliquam enim et tortor. Quisque aliquet, quam elementum condimentum feugiat, tellus odio consectetuer wisi, vel nonummy sem neque in elit. Curabitur eleifend wisi iaculis ipsum. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. In non velit non ligula laoreet ultrices. Praesent ultricies facilisis nisl. Vivamus luctus elit sit amet mi. Phasellus pellentesque, erat eget elementum volutpat, dolor nisl porta neque, vitae sodales ipsum nibh in ligula. Maecenas mattis pulvinar diam. Curabitur sed leo.

Nulla facilisi. In vel sem. Morbi id urna in diam dignissim feugiat. Proin molestie tortor eu velit. Aliquam erat volutpat. Nullam ultrices, diam tempus vulputate egestas, eros pede varius leo, sed imperdiet lectus est ornare odio. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Proin consectetuer velit in dui. Phasellus wisi purus, interdum vitae, rutrum accumsan, viverra in, velit. Sed enim risus, congue non, tristique in, commodo eu, metus. Aenean tortor mi, imperdiet id, gravida eu, posuere eu, felis. Mauris sollicitudin, turpis in hendrerit sodales, lectus ipsum pellentesque ligula, sit amet scelerisque urna nibh ut arcu. Aliquam in lacus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Nulla placerat aliquam wisi. Mauris viverra odio. Quisque fermentum pulvinar odio. Proin posuere est vitae ligula. Etiam euismod. Cras a eros.

Nunc auctor bibendum eros. Maecenas porta accumsan mauris. Etiam enim enim, elementum sed, bibendum quis, rhoncus non, metus. Fusce neque dolor, adipiscing sed, consectetuer et, lacinia sit amet, quam. Suspendisse wisi quam, consectetuer in, blandit sed, suscipit eu, eros. Etiam ligula enim, tempor ut, blandit nec, mollis eu, lectus. Nam cursus. Vivamus iaculis. Aenean risus purus, pharetra in, blandit quis, gravida a, turpis. Donec nisl. Aenean eget mi. Fusce mattis est id diam. Phasellus faucibus interdum sapien. Duis quis nunc. Sed enim.

Pellentesque vel dui sed orci faucibus iaculis. Suspendisse dictum magna id purus tincidunt rutrum. Nulla congue. Vivamus sit amet lorem posuere dui vulputate ornare. Phasellus mattis sollicitudin ligula. Duis dignissim felis et urna. Integer adipiscing congue metus. Nam pede. Etiam non wisi. Sed accumsan dolor ac augue. Pellentesque eget lectus. Aliquam nec dolor nec tellus ornare venenatis. Nullam blandit placerat sem. Curabitur quis ipsum. Mauris nisl tellus, aliquet eu, suscipit eu, ullamcorper quis, magna. Mauris elementum, pede at sodales vestibulum, nulla tortor congue massa, quis pellentesque odio dui id est. Cras faucibus augue.

Suspendisse vestibulum dignissim quam. Integer vel augue. Phasellus nulla purus, interdum ac, venenatis non, varius rutrum, leo. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Duis a eros. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Fusce magna mi, porttitor quis, convallis eget, sodales ac, urna. Phasellus luctus venenatis magna. Vivamus eget lacus. Nunc tincidunt convallis tortor. Duis eros mi, dictum vel, fringilla sit amet, fermentum id, sem. Phasellus nunc enim, faucibus ut, laoreet in, consequat id, metus. Vivamus dignissim. Cras lobortis tempor velit. Phasellus nec diam ac nisl lacinia tristique. Nullam nec metus id mi dictum dignissim. Nullam quis wisi non sem lobortis condimentum. Phasel-

lus pulvinar, nulla non aliquam eleifend, tortor wisi scelerisque felis, in sollicitudin arcu ante lacinia leo.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vestibulum tortor quam, feugiat vitae, ultricies eget, tempor sit amet, ante. Donec eu libero sit amet quam egestas semper. Aenean ultricies mi vitae est. Mauris placerat eleifend leo. Quisque sit amet est et sapien ullamcorper pharetra. Vestibulum erat wisi, condimentum sed, commodo vitae, ornare sit amet, wisi. Aenean fermentum, elit eget tincidunt condimentum, eros ipsum rutrum orci, sagittis tempus lacus enim ac dui. Donec non enim in turpis pulvinar facilisis. Ut felis.

Cras sed ante. Phasellus in massa. Curabitur dolor eros, gravida et, hendrerit ac, cursus non, massa. Aliquam lorem. In hac habitasse platea dictumst. Cras eu mauris. Quisque lacus. Donec ipsum. Nullam vitae sem at nunc pharetra ultricies. Vivamus elit eros, ullamcorper a, adipiscing sit amet, porttitor ut, nibh. Maecenas adipiscing mollis massa. Nunc ut dui eget nulla venenatis aliquet. Sed luctus posuere justo. Cras vehicula varius turpis. Vivamus eros metus, tristique sit amet, molestie dignissim, malesuada et, urna.

Cras dictum. Maecenas ut turpis. In vitae erat ac orci dignissim eleifend. Nunc quis justo. Sed vel ipsum in purus tincidunt pharetra. Sed pulvinar, felis id consectetuer malesuada, enim nisl mattis elit, a facilisis tortor nibh quis leo. Sed augue lacus, pretium vitae, molestie eget, rhoncus quis, elit. Donec in augue. Fusce orci wisi, ornare id, mollis vel, lacinia vel, massa.

# 6
## Conclusion

Lorem ipsum dolor sit amet, 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.

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.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vestibulum tortor quam, feugiat vitae, ultricies eget, tempor sit amet, ante. Donec eu libero sit amet quam egestas semper. Aenean ultricies mi vitae est. Mauris placerat eleifend leo. Quisque sit amet est et sapien ullamcorper pharetra. Vestibulum erat wisi, condimentum sed, commodo vitae, ornare sit amet, wisi. Aenean fermentum, elit eget tincidunt condimentum, eros ipsum rutrum orci, sagittis tempus lacus enim ac dui. Donec non enim in turpis pulvinar facilisis. Ut felis.

Cras sed ante. Phasellus in massa. Curabitur dolor eros, gravida et, hendrerit ac, cursus non, massa. Aliquam lorem. In hac habitasse platea dictumst. Cras eu mauris. Quisque lacus. Donec ipsum. Nullam vitae sem at nunc pharetra ultricies. Vivamus elit eros, ullamcorper a, adipiscing sit amet, porttitor ut, nibh. Maecenas adipiscing mollis massa. Nunc ut dui eget nulla venenatis aliquet. Sed luctus posuere justo. Cras vehicula varius turpis. Vivamus eros metus, tristique sit amet, molestie dignissim, malesuada et, urna.

Cras dictum. Maecenas ut turpis. In vitae erat ac orci dignissim eleifend. Nunc quis justo. Sed vel ipsum in purus tincidunt pharetra. Sed pulvinar, felis id consectetuer malesuada, enim nisl mattis elit, a facilisis tortor nibh quis leo. Sed augue lacus, pretium vitae, molestie eget, rhoncus quis, elit. Donec in augue. Fusce orci wisi, ornare id, mollis vel, lacinia vel, massa.

Lorem ipsum dolor sit amet, 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.

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.

# A

# Transcription Factor Regulatory Circuits - Tissue Types