

Networks within networks

Probing the neuronal and molecular underpinnings of language-related disorders using human cell models

MIDAS ANIJS

Funding body

his research was funded by the Max Planck Society for the Advancement of Science (www.mpg.de/en) in the form of an IMPRS for Language Sciences PhD Fellowship (2015-2019) awarded to Midas Anijs.

International Max Planck Research School (IMPRS) for Language Sciences

The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics

Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

© 2023, Midas Anijs

ISBN: 978-94-92910-53-0

Cover design by Damiët Anijs

Lay-out by Midas Anijs

Printed and bound by Ipskamp Drukkers, Enschede

All rights reserved. No part of this book may be reproduced, distributed, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission of the author. The research reported in this thesis was conducted at the Max Planck Institute for Psycholinguistics, in Nijmegen, the Netherlands

Networks within networks

Probing the neuronal and molecular underpinnings of language-related disorders using human cell models

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens het besluit van het college voor promoties
in het openbaar te verdigen op woensdag 31 januari 2024
om 12.30 precies

door
Midas Vincent Anijs
geboren op 10 juli 1992
te Nijmegen

Promotoren

Prof. dr. S.E. Fisher

Prof. dr. S.C. Vernes (University of St. Andrews, Verenigd Koninkrijk)

Manuscriptcommissie

Prof. dr. E.J.M. Storkebaum (voorzitter)

Prof. dr. E.B.E. Becker (University of Oxford, Verenigd Koninkrijk)

Dr. D. Schubert

Networks within networks

Probing the neuronal and molecular underpinnings of language-related disorders using human cell models

Dissertation

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.M. Sanders,
according to the decision of the Doctorate Board
to be defendend in public on Wednesday, January 31, 2024
at 12.30 pm

by
Midas Vincent Anijs
born on July 10, 1992
in Nijmegen

Supervisors

Prof. dr. S.E. Fisher

Prof. dr. S.C. Vernes (Univesrity of St Andrews, United Kingdom)

Manuscript committee

Prof. dr. E.J.M. Storkebaum (chair)

Prof. dr. E.B.E. Becker (University of Oxford, United Kingdom)

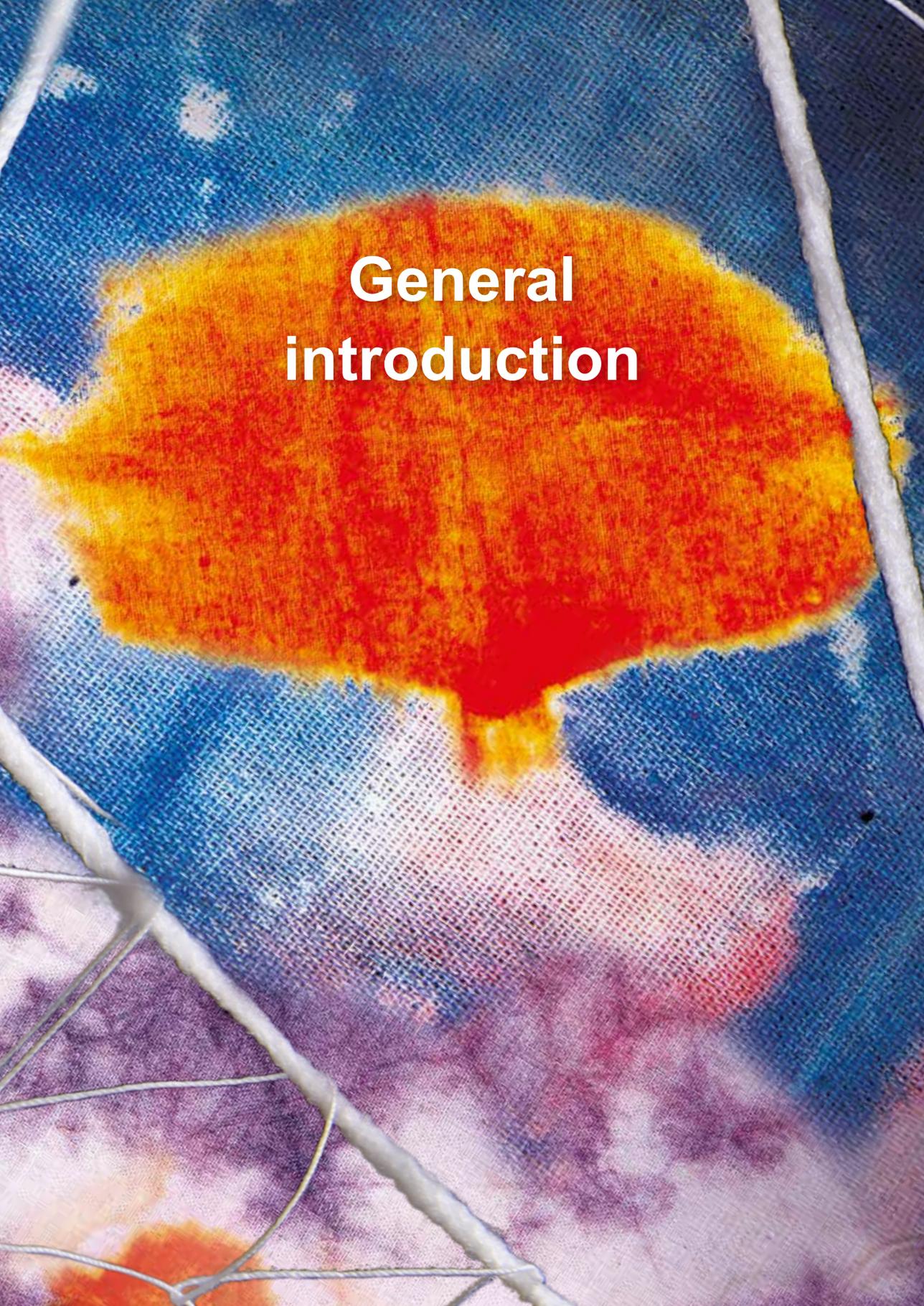
Dr. D. Schubert

Contents

Chapter 1	General introduction	9
Chapter 2	Hyperkinetic stereotyped movements in a boy with biallelic <i>CNTNAP2</i> variants	39
Chapter 3	Targeting multiple exons of <i>CNTNAP2</i> in human neurons suggests isoform-specific deficit in cilia assembly	53
Chapter 4	Characterizing neurodevelopmental microRNA networks and their response to system perturbation by loss of <i>CNTNAP2</i> in a human neuronal network model	97
Chapter 5	<i>ARHGEF39</i> , a gene implicated in developmental language disorder, activates RHOA and is involved in cell de-adhesion and neural progenitor cell proliferation	123
Chapter 6	General discussion	155
Appendices	English summary	175
	Nederlandse samenvatting	178
	Acknowledgements	181
	Curriculum vitae	183
	Publications	184

1





**General
introduction**

CHAPTER 1

General introduction

1.1 Human language and language disorders

Language is a fascinating trait, as it provides humans with an ability for complex communication that is unparalleled in the animal kingdom. Most children acquire speech and language skills in the first few years of life in a way that seems effortless and without the need for formal instruction. Therefore, many efforts have been made to understand the natural origins of this human instinctive tendency to speak and the properties of the human brain that make it capable of language. The field of the genetics of language aims to identify the molecular elements that underpin these properties and those that can explain variation in speech and language abilities. The functional investigation of this collection of molecular elements provides a causal basis for the biological processes that make the human brain capable of the comprehension and production of human speech and language.

Developmental disorders that involve speech or language impairments display complex profiles of impairment. Some disorders disproportionately impact language function against a background of preserved general cognition, whereas other disorders affect communication abilities as part of a broader set of cognitive deficits, such as intellectual disability (ID) and autism spectrum disorder (ASD). Further, the profiles of ‘language-specific’ disorders highlight different aspects of language and may tell something about the relationship between them. For example, developmental language disorder (DLD), previously referred to as specific language impairment, is a condition in which the language development of children does not follow its usual course despite typical development in other areas¹. Children with DLD display language impairments in multiple areas, including acquisition of vocabulary and syntax. On the other hand, childhood apraxia of speech (CAS) is defined by a primary deficit in programming the rapid motor sequences involved in speech production. Developmental dyslexia (DD) is characterized by difficulties with written language, while spoken language seems overtly normal. Studying the genes that are related to these speech and language disorders will facilitate a better understanding of the causes of these impairments and what may be required for the normal development of language.

The molecular architecture of language disorders is complex and in most cases involves contributions from multiple genes or genetic risk factors². So far, dozens of candidate language-related genes have been identified (Table 1). This forms only the beginning of identifying the genetic architecture of language disorders, as these genes are active in diverse pathways and at diverse timepoints. The identification of new candidate genes is further complicated by the fact that the phenotypic consequences of mutations or risk factors can be modified by genetic or environmental interactions. This may need to be considered when validating potentially pathogenic mutations. Another challenge for mapping genetic influences onto neurodevelopmental disorders is formed by complexity at the behavioural level, as the boundaries between disorders are not always clear. Variants in some genes contribute to the susceptibility for multiple disorders³ (Table 1). Speech and language disorders are frequently found to be co-morbid with each other⁴. Language impairment also co-occurs with developmental disorders that impact a broader range of cognitive functions, such as ASD⁵. This overlap between behavioural categories is expected since speech, language and reading have a considerable cognitive overlap and are also associated with cognitive systems that are not specific to language. This means that connecting genes to behaviour is complicated by intermediate phenotypes

at the cognitive and neurobiological level that sometimes overlap between disorders and behavioural categories.

In addition to shedding light on the molecular bases of these disorders, the functional investigation of disease-related genes can help tease out differences between disorders at a neurobiological and clinical level, as well as elucidate new links between disorders. In the next section, I will discuss three genes that each help to address these complexities and that form entry points to study gene networks and human neuronal features that contribute to the ability of language.

Table 1. Overview of candidate language-related genes *

Gene	Disorder	Common variation	Rare variation	Functional validation of genetic variant
ARHGEF19	DLD	GWAS ⁶		
ARHGEF39	DLD	GWAS ⁷		miR target regulation ⁷
ATP2C2	DLD	Linkage, association ^{8,9}	Part of larger deletion ¹⁰	
CMIP	DLD	Linkage, association ^{9,11,12}		
CNTNAP2	DLD, CDD	Linkage ¹³ , association ^{11,14-16}	SNV/indel ¹⁷⁻²⁰	Gene expression ^{21,22} , axon outgrowth ²³ , progenitor overgrowth ²⁴
NFXL1	DLD	Association ²⁵		
NOP9	DLD	GWAS ⁶		
ZNF277	DLD		SNV/indel ²⁶	
CHD3	CAS, NDD		SNV/indel ^{27,28}	
ERC1	CAS, DLD		Part of larger deletion ²⁹	
FOXP1	ASD, ID		SNV/indel ^{18,30-33}	DNA binding, subcellular localization ³⁴
FOXP2	CAS	Linkage	SNV/indel ³⁵⁻³⁹ , translocation breakpoint ⁴⁰⁻⁴⁴ , part of larger deletion ⁴⁴⁻⁵²	DNA binding, subcellular localization ^{34,53,54}
FOXP4	CAS, NDD		SNV/indel ⁵⁵	
GRIN2B	DD, ASD	Association ⁵⁶	SNV/indel ¹⁸	
GRIN2A	CAS		SNV/indel ⁵⁷⁻⁶⁰ , Part of larger deletion ⁵⁸⁻⁶⁰	
KMT2D	CAS		SNV/indel ²⁷	

Gene	Disorder	Common variation	Rare variation	Functional validation of genetic variant
OXR1	CAS		SNV/indel ²⁷	
SCN9A	CAS		SNV/indel ²⁷	
SETBP1	CAS	Association ⁶¹	SNV/indel ^{27,62}	
SETD1A	CAS		SNV/indel ²⁷	
SRPX2	CAS		SNV/indel ²⁷	
WDR5	CAS		SNV/indel ²⁷	
DCDC2	DD	Linkage ^{63,64} , association ^{11,12,63,65-70}	Part of larger duplication ⁷¹	Gene expression ^{72,73}
DYX1C1	DD	Linkage ⁷⁴⁻⁷⁶ , association ⁷⁷⁻⁸²	Translocation breakpoint ^{83,84}	Gene expression ^{85,86}
GCFC2	DD	Linkage ⁸⁷		
KIAA0319	DD	Linkage ^{64,88} , association ^{11,12,89-95}	Part of larger duplication ⁷¹	Gene expression ^{96,97}
LOC388780	DD	GWAS ⁹⁸		
MRPL19	DD	Linkage ⁸⁷ , association ⁸		
ROBO1	DD	Linkage ⁹⁹⁻¹⁰¹ , association ^{93,102,103}	SNV/indel ¹⁰⁴ , translocation breakpoint ¹⁰⁵	TF target regulation ^{104,106}
S100B	DD	Association ⁶⁶	Part of larger deletion ¹⁰⁷	
VEPH1	DD	GWAS ⁹⁸		
ROBO2	Expressive vocabulary	GWAS ¹⁰⁸		
SCN11A	Social communication	GWAS ¹⁰⁹		
PLCL1		PheWAS ¹¹⁰		

* This overview is based on literature up to 2021. Recent large GWASs have contributed tens of new candidate genes for reading disability and language skills.¹¹¹⁻¹¹³

Abbreviations: GWAS = genome-wide association study, indel = insertion-deletion, miR = microRNA, PheWAS = phenome-wide association study, TF = transcription factor

1.2 Genetics of developmental speech and language disorders

1.2.1 FOXP2 – the first gene implicated in a monogenic speech and language disorder

The first example of a gene that was identified as mutated in a monogenic form of speech and language disorder is *FOXP2*. This gene was first implicated via a single causal

mutation in the KE family, where half of the thirty members suffer from CAS, as well as an unrelated child with a similar disorder who carried a chromosomal translocation³⁵. As such, the KE family presented a notable rare exception to the genetic complexity of most cases of speech and language disorder. *FOXP2* belongs to the winged helix/forkhead-box class of transcription factors that contain a characteristic DNA-binding domain called the forkhead box¹¹⁴. Nomenclature guidelines recommend species-specific capitalisation for Fox proteins, i.e. *FOXP2* for human, *Foxp2* for mouse and *FoxP2* for other species and when describing multiple species¹¹⁵. Studying *FoxP2* and its targets has helped to uncover neural pathways, cellular processes, and molecular networks that contribute to speech and language, and disorders thereof¹¹⁶. Neural expression patterns of *FOXP2* coincided with known sites of structural and functional abnormality in brain regions involved in speech, language and motor control, supporting roles of this gene in the development of motor control pathways important for speech¹¹⁷. The identification of a gene related to the development of speech and language represented a breakthrough, as it led to targeted functional genomics studies and the design of new animal models to study the neurobiological underpinnings of speech and language.

Following the initial studies of the KE family, more mutations in other families and unrelated individuals have been identified that confirmed a role for *FOXP2* in speech and language development^{36-38,118-120}. However, mutations in *FOXP2* are rare, with one study estimating that they explain fewer than ~2% of CAS cases³⁷, and the gene has not been robustly implicated in other speech/language disorders. This means that much of the genetic architecture underlying such disorders remains unknown, which limits our ability to understand the causal mechanisms involved. Technological advances in next-generation (NGS) sequencing have accelerated this search by powering the discovery of rare and *de novo* variants in a range of different disorders, including those that affect neurodevelopment, such as ASD and ID¹²¹. NGS methods have recently been successfully applied to DLD¹²² and CAS²⁷, leading to the identification of new candidate genes for follow-up studies. Table 1 shows that the identification of rare variants has been instrumental to find most of the candidate language-related genes known thus far. Rare gene variants are expected to have a large effect size, which makes them valuable candidates for functional characterization in model systems and to expose new aspects of the biology underlying speech and language. At the same time, monogenic disorders only explain a subset of cases, and complex genetic interactions may be missed by focusing on rare variants with large effect sizes, which emphasizes the need for complementary approaches for unravelling the neurogenetic pathways underlying speech and language disorders.

1.2.2 *CNTNAP2* – a functional link between complex disorders

Several types of evidence have implicated the *CNTNAP2* gene in language-related disorders. Mutations in *CNTNAP2* form a genetic link between multiple complex disorders¹²³. As such, *CNTNAP2* is an outstanding candidate for studying both the genetic and phenotypic complexity of language-related neurodevelopment.

CNTNAP2 was initially linked to language-related disorders as one of the first genes that was targeted and regulated by *FOXP2*¹⁵. However, rare mutations in *CNTNAP2* have been related to broader cognitive impairments than CAS. Homozygous mutations in *CNTNAP2* cause a syndrome called CASPR2 deficiency disorder (CDD), that is characterized by

severe intellectual disability, early-onset drug-resistant epilepsy, reduced or absent language, and in some cases focal cortical dysplasia^{19,20}. Rare heterozygous mutations in *CNTNAP2* have further been described in individuals with a range of complex phenotypes disorders, including language impairment, intellectual disability, autistic characteristics and seizures, but are sometimes also found in non-affected individuals, making their causal relevance difficult to interpret^{123,124}. This broad range of phenotypic consequences suggest that *CNTNAP2* mutations can affect multiple processes to varying extents and that *CNTNAP2* may have several genetic interactions that could modify the effects of mutations. Therefore, in-depth functional investigations are needed into the molecular network of *CNTNAP2* and the neurodevelopmental functions it relates to.

Common variants in *CNTNAP2* have been associated with language-related measures across different cohorts, such as individuals with DLD^{11,15}, ASD¹²⁵⁻¹²⁷, DD¹⁴ and typically developing individuals^{16,128}. Neuroimaging further suggests that *CNTNAP2* polymorphisms may affect activation in functional brain networks during language processing tasks¹²⁹⁻¹³². However, it should be noted that these studies rely on small sample sizes and that they are not independently replicated. The need for larger sample sizes is emphasized by a study with 6,165 subjects that did not replicate the association between SNPs in *CNTNAP2* with decreased integrity of white matter tracts¹³³, which had been reported in smaller studies^{127,134,135}. A targeted variant study with 1,700 participants replicated the association between rs7794745 and reduced grey matter volume in the left superior occipital sulcus¹³⁶, but none of the other reported earlier associations with grey matter volume¹³⁷. This demonstrates that large cohorts are needed to discover more precisely the impact of common *CNTNAP2* variants on the structure and function of brain networks, including those that may be involved in language deficits. Functional studies can then help to elucidate the molecular mechanisms and neuronal features which mediate the roles of CASPR2 in these complex phenotypes.

1.2.3 *ARHGEF39* – a functional entry point from common genetic variation

The complex genetic architecture of speech and language impairments means that variants with small effect sizes are also an important part of the molecular basis of these disorders. The study of common variants has led to the identification of new candidate language-related genes (Table 1). Several language-related genes so far have been identified by classical linkage screens in family-based cohorts with speech, language, and reading disorders and complementary targeted association studies^{138,139}. However, this targeted approach based on collections of small families has been too focused to produce major advances in explaining the heritability of language-related disorders. Genome-wide association studies (GWAS) have much greater potential to discover new language-associated variants. Such studies systematically test millions of common variants in the human genome for association with disorder status or quantitative measures of relevant phenotypes. This means that GWASs need to account for large amounts of multiple testing, and since the effect sizes of common variants are expected to be small, GWASs require thousands of subjects to reach statistical power. It has been challenging for language-related GWAS to achieve these cohort sizes¹⁴⁰. Data collection for language-related traits is challenging, as the reliable assessment of children's language abilities is time consuming and these phenotypes show developmental variation which complicate the

selection of suitable measures¹⁴¹. Recent GWASs for with cohorts in the tens of thousands rather than hundreds and thousands show that it is possible to buffer these imperfections in phenotypic characterization and to find additional common variants that contribute to the development of speech and language¹¹¹⁻¹¹³.

After a significant association is detected between a common genetic variant and a trait, it is often difficult to determine if this association reflects a causal relationship that provides new biological insight. The significant variant could be in linkage disequilibrium with an actually functional variant or be involved in interactions with other genetic variants. Apart from these considerations about confounding and effect modification, the impact of an associated variant on gene function is usually not directly clear. When a variant is in a coding region, it is relatively easy to predict if this will result in an amino-acid change and how the variant potentially affects protein function. However, often variants are in non-coding regions for which the functional impact is more difficult to predict and to experimentally validate. Nonetheless, progress has been made in assessing the biological roles of variants outside coding regions in the context of speech and language disorders. Analysis of known microRNA sites in 3'UTR sequence data in the exomes of a cohort of children with DLD identified the SNP rs72727021. The alternative allele for rs72727021, which has a population frequency of 4.7%, was carried by 23% of the DLD children in the cohort and was significantly associated with reduced performance on a non-word repetition task. It was predicted that this SNP could affect microRNA binding in a gene called *ARHGEF39*. Functional assays in cell models confirmed that the risk allele of the SNP disrupted regulation of *ARHGEF39* by microRNAs and led to an increase in gene expression. An association between the risk allele and increased *ARHGEF39* expression was also observed in expression quantitative trait locus data from post-mortem human brain⁷. Advances in the functional annotation of the non-coding genome are not just helpful for functional validation, but also indicate new molecular mechanisms for the biological effects of genetic variants.

1.3 The importance of molecular networks for understanding gene function

There is not a straightforward relationship between individual genes and specific aspects of human cognition, but instead there is a complex, multi-layered biological system that links genes and cognition^{2,142}. In order to understand this complexity, we need to consider genes and their products in molecular networks. This means that a gene does not function on its own, but that we need to take into account regulatory factors, interaction partners and downstream targets if we aim to understand the role of a gene in processes that shape the development and function of the brain. Molecular networks are not static, but changes in their composition and configuration produce multiple states that reflect different functional phenotypes at different levels, such as cell type, circuit, region, brain and organism, across development^{142,143}. We can discover connections between the heterogeneous genetic risk factors for cognitive disorders, and identify key molecules and mechanisms, by investigating these networks and relevant states.

The development of high-throughput molecular techniques has enabled comprehensive analyses of molecular networks across genotype, cell type, brain region, developmental stage and/or species. First, gene/protein expression data (e.g. transcriptomic and proteomic data) show the molecular network components that are present in the cellular

or tissue system that is being analysed. The condition-specific abundance of gene products helps to associate these genes to functional phenotypes, and co-expression clusters help to predict which gene products may interact¹⁴⁴. Second, the molecular network can be further structured by identifying the regulatory interactions between components of the molecular network. Several types of physical interactions can structure molecular networks, for example interactions of proteins with DNA, proteins with proteins, and RNA molecules with RNA molecules. These different types of interactions are relevant, because gene regulatory networks are formed by interactions across multiple molecular levels: genome, transcriptome and proteome¹⁴³. The next parts of this section show how, at each of these levels, gene regulatory networks can contribute to understanding the function of language-related genes in brain development.

1.3.1 Protein-DNA interactions: Transcription factor networks

Transcription factors play an important role in organizing the diversity of gene expression patterns in different cell types and for different processes¹⁴⁵. Transcription factors are proteins that bind to DNA, typically at a specific consensus sequence, and regulate the transcription of genes in an activating and/or repressing role¹⁴⁶. These regulatory interactions create gene regulatory networks in which subsets of transcription factors form patterns, such as auto-regulation, feed-forward loops, and multi-input modules, thereby facilitating cell fate decision making and stability in the regulated biological processes^{147,148}. In order to understand the functions of a transcription factor, it is important to know where it binds and understand the resulting effects on gene expression and subsequent phenotypes.

Identifying the downstream target genes of the transcription factor FOXP2 led to the identification of molecular networks that are informing our understanding of the neurobiology of speech and language¹⁴⁹. The downstream targets of FOXP2 indicate the fundamental biological pathways that this protein regulates during neurodevelopment. Genome-wide gene binding of FOXP2 has been measured by chromatin immunoprecipitation (ChIP) experiments in human foetal tissue¹⁵⁰, human neuroblastoma cells¹⁵¹, and embryonic mouse brain¹⁵². The FOXP2 protein-DNA networks that were identified in these studies were enriched for genes associated with cell signalling, cell migration, neurogenesis, neuronal differentiation, neurite development, and axon guidance¹⁵⁰⁻¹⁵². The influence of FoxP2 on such pathways has been confirmed with functional studies in mouse models and immortalized cell lines. For example, loss-of-function mutations in *Foxp2* result in decreased neurite lengths in mouse striatal neurons¹⁵² and in cortical neurons¹⁵³, whereas modification of mouse *Foxp2* with two human-specific amino acid substitutions leads to increased dendrite lengths of medium spiny neurons¹⁵⁴. The FOXP2 target network has also suggested functional genetic links between multiple disorders. Studies of FOXP2 target genes, such as *CNTNAP2*¹⁵, *SRPX2*¹⁵⁵, *MET*¹⁵⁶, *DISC1*¹⁵⁷, and *MEF2C*¹⁵⁸, show that FOXP2-regulated processes may play roles in various neurodevelopmental disorders, including autism, schizophrenia and speech/language disorder.

The upstream regulators of FOXP2 are highly relevant for its roles in disease, as they determine where and when FOXP2 can affect the functions that were identified through its downstream targets. In transcription factor networks, upstream and downstream

interactions are often connected in loops, creating sophisticated mechanisms for stability and cell fate decision making¹⁴⁸. The positive autoregulatory function of FOXP2 is an example of such a loop¹⁵⁹. The ability of a transcription factor to self-sustain its levels enables rapid amplification of a protein and a quick transition to a new state, meanwhile the feedforward loop keeps a cell committed to this specific state until the feedforward loop is disrupted¹⁶⁰. Regulation of FOXP2 by WNT pathway genes provides more developmental insight into the contribution of FOXP2 to switches in cell state, as this pathway is important for neurogenesis and differentiation. FoxP2 is activated by Lef1 in the zebrafish brain, via highly conserved LEF/TCF binding regions¹⁶¹, and by PAX6¹⁶². LEF1 is activated by canonical WNT/ β -catenin signalling, and is involved in neurogenesis and cell fate specification¹⁶¹. PAX6 is a transcription factor that controls the identity and differentiation of neural progenitor cells^{163,164}. Further, LEF1, β -catenin, and PAX6 form a pathway that initiates a neurogenic transcription factor cascade¹⁶⁵. Activation of FOXP2 by PAX6 suggests that FOXP2 may play a role in the differentiation of neural progenitor cells into neurons. This is in line with a finding that FOXP2 drives the differentiation of early neurons into deep layer neurons under control of WNT3¹⁶⁶. Being activated by both LEF1 and PAX6, FOXP2 is part of this neurogenic transcription factor network. Interestingly, expression data indicate that FOXP2 leads to downregulation of WNT pathway genes in a fibroblast cell line, and that β -catenin and FOXP2 physically interact¹⁶⁷. Thus, FOXP2 may create a feedback loop that could have important functional consequences for WNT/ β -catenin signalling, neurogenesis, and cell fate specification in the brain, but this hypothesis needs further investigation¹⁴⁹. Overall, both downstream targets and upstream regulators contribute to understanding FOXP2-related functional mechanisms during brain development. Downstream targets identify specific biological pathways and processes, whereas upstream regulators help to identify the developmental contexts in which FOXP2 molecular networks are active.

1.3.2 Protein-protein interaction networks

Protein-protein interactions are essential for every biological process and tracing these interactions helps to group proteins into functional pathways¹⁶⁸. Protein-protein interaction networks therefore provide an essential layer of topology for the molecular network. Protein-protein interactions often lack the directionality of DNA-protein interactions¹⁶⁹. A protein-DNA interaction resulting in transcriptional regulation allows a separation in upstream regulators and downstream targets, but that distinction is not as clear for many protein-protein interactions. This makes it harder to infer mechanisms from a diagram of protein-protein interactions. Studying protein-protein interactions regularly leads to the identification of novel biological processes in which a protein of interest is involved. Large-scale proteomic screens are very useful to refine molecular networks and can yield novel functional insights about a protein.

Proteomic screens and the subsequent validation of protein-protein interactions have enriched the understanding of FOXP2 transcriptional networks and provided new information regarding the functionality of the protein. An affinity purification-mass spectrometry screen identified several brain-expressed transcription factors that physically interact with FOXP2, extending its molecular network³⁴. These transcription factors (NR2F1, NR2F2, SATB1, SATB2, and SOX5) are co-expressed in varying compositions

in the cerebral cortex and cerebellum³⁴, which may contribute to diversification of FOXP2 transcriptional activity. Studying the consequences of these interactions can help to understand when and how some targets are regulated by FOXP2 in different cell types. Protein interaction studies have further helped to reveal the mechanisms by which FOXP2 affects the chromatin state. FOXP2 interacts with CTBP1 and CTBP2¹⁷⁰, which belong to a complex that represses gene expression through histone modification and that is recruited by DNA-binding proteins¹⁷¹. FOXP proteins also interact with several proteins in the nucleosome remodelling and histone deacetylase (NuRD) complex and may be involved in the recruitment of the NuRD complex to the DNA¹⁷². It has further been reported that FOXP2 can mediate chromatin accessibility without directly binding to DNA, through physical interaction with NFIA and NFIB¹⁷³. This means that FOXP2 can also regulate gene expression at genomic target sites with an NFI motif instead of a FOX motif, affecting additional downstream pathways. Such examples show that protein-protein interaction studies are essential to understand and uncover the multiple functions of a single protein.

Proteomic screens also led to the elucidation of novel functions for CASPR2, the protein product of *CNTNAP2*. CASPR2 is most prominently known as a transmembrane cell adhesion molecule. It was first studied in the peripheral nervous system where it interacts extracellularly with Contactin-2 to form a neuron-glia cell adhesion complex¹⁷⁴, and clusters potassium channels at the juxtaparanode of myelinated axons via cytoplasmic protein-protein interactions¹⁷⁵. However, reverse genetic experiments in mice and primary neurons have also found effects of loss of *CNTNAP2* in the central nervous system, where it affects phenotypes such as synaptic function and dendritic branching¹⁷⁶⁻¹⁷⁸. Protein-protein interaction studies were important for investigating mechanisms by which CASPR2 affects these functions. An affinity purification-mass spectrometry study of CASPR2 interactors confirmed protein-protein interactions that were studied in the peripheral nervous system in the synaptic subfractions from cerebral cortex and hippocampus¹⁷⁹. The same study also identified a short protein isoform of CASPR2 that lacks the extracellular domain. This means that CASPR2 functions are best viewed in light of interactions in multiple isoform-specific molecular networks. The exact contributions of the different isoforms remain to be clarified.

Domain-specific proteomic screens have found additional interactors of CASPR2 that help to identify mechanisms for neuronal phenotypes associated with *CNTNAP2* mutations. A yeast two-hybrid screen for the CASPR2 intracellular domain identified CASK as a CASPR2 interaction partner¹⁸⁰. CASK is a scaffold protein that coordinates signal transduction pathways at the cytoskeleton that regulate neural development and synaptic function¹⁸¹. CASK is mislocalized in *CNTNAP2* knockout neurons and overexpression of CASK rescues reduced dendritic arborization in *CNTNAP2* knockout neurons, indicating that CASK functions downstream of *CNTNAP2* in relation to this phenotype¹⁸⁰. Affinity purification-mass spectrometry for the extracellular domain of Caspr2 in synaptic fractions of the cerebellum detected new interaction partners, including the Ca²⁺ channel IP₃R1¹⁸². Caspr2 and IP₃R1 are both needed for normal dendritic development in Purkinje cells and Caspr2 suppresses IP₃R1-induced morphological changes in cell lines, suggesting that IP₃R1-mediated calcium signalling plays a role in *CNTNAP2* phenotypes¹⁸². This work thus shows how protein-protein interaction studies can not only provide necessary elements for understanding the mechanisms behind specific phenotypes, but also help uncover

new aspects of gene function such as the roles of different isoforms.

1.3.3 RNA-RNA interactions: MicroRNA networks

MicroRNA (miRNA) networks have emerged as regulatory mechanisms at the post-transcriptional level with relevance for virtually every biological processes, as about 60% of human protein-coding genes are conserved targets of miRNAs¹⁸³. In the mammalian brain, miRNAs display complex spatio-temporal expression patterns and have the potential to regulate thousands of target genes^{185,186}. Prominent regulatory roles have been identified for miRNAs in several neurodevelopmental processes so far, such as cell fate determination, migration, neuronal polarization, and synapse formation¹⁸⁷. The developmental expression trajectories of miRNAs and their target genes underwent evolutionary changes resulting in new configurations of gene regulatory networks that may have contributed to the rapid evolution of the human brain¹⁸⁸. A role for miRNAs in remodelling cortical development suggests that human-specific cognitive abilities, such as language, may partly be traced to specific miRNA-mRNA interactions. MiRNAs have been shown to create an extra layer of connectivity in transcription factor networks by the formation of additional co-regulatory loops^{147,189}. MiRNA networks thus form an important extension of known regulatory pathways and investigating miRNA-mRNA interactions may help us understand the biological mechanisms that are affected by language-related molecular networks.

Exploring the miRNAs regulated downstream of FOXP2 has revealed molecular effectors that contribute to FOXP2-related phenotypes. Several miRNAs with prominent neurodevelopmental roles have been identified as regulatory targets of FOXP2, including miR-137 and miR-9¹⁵². MiR-137 is one of the most studied microRNAs, as it has been associated with schizophrenia^{190,191}. Misregulation of miR-137 can have consequences for neurodevelopment through its role in neural differentiation¹⁹² and synaptic transmission and plasticity^{193,194}, making it well placed to contribute to FOXP2-related phenotypes. Target genes of miR-137 may also be parts of the molecular network of FOXP2. Several validated target genes of miR-137 have also been implicated in neurodevelopmental disorder, such as schizophrenia¹⁹⁵ and ASD¹⁹⁶. MiR-9 shows how miRNAs form an additional regulatory layer. MiR-9 is one of the most highly expressed miRNAs in the brain and is involved in neuronal progenitor maintenance, neurogenesis, and differentiation during neurodevelopment¹⁹⁷. Studies in mouse, human and songbird systems found that miR-9 represses *FoxP2*¹⁹⁸⁻²⁰⁰. Given that FOXP2 activates miR-9 during neurite outgrowth¹⁵², a negative feedback loop is formed between FOXP2 and miR-9 which allows for tight control of FOXP2 protein production¹⁸⁹. It has been demonstrated that miR-9 exerts this control on *Foxp2* in the mouse embryonic cortex, as *Foxp2* is expressed more strongly when miR-9 binding sites in the 3'UTR are mutated¹⁹⁸. In zebra finch songbirds, miR-9 is upregulated and *FoxP2* downregulated in Area X of males singing undirected songs, indicating that controlled expression of miR-9 and *FoxP2* may be involved in the function of circuits underlying social vocal behaviour¹⁹⁹. MiRNA networks thus help to extend molecular networks and can identify potential functional links between disorders.

1.4 Models to study the functions of language-related genes

1.4.1 Multiple approaches for studying the roles of language-related gene function in the brain

In order to understand the contribution of molecular networks to the development of speech and language, the functional properties of language-related genes need to be studied at multiple levels. The biochemical properties of proteins determine how gene products operate in the cell, and how they affect cellular processes. These cellular processes support the development and function of neural circuits. The activity in these neural circuits influences the phenotypes observed at levels of behaviour. Multiple types of models make it possible to assess the effects of individual genes or molecular networks on these multiple layers that connect genes to language-related behaviours. Each model comes with its own set of advantages and limitations, which will be discussed below.

Assays in cell lines are key to study molecular mechanisms, as they are well suited to assess the properties of proteins and functional impacts of mutations. For example, assays in human embryonic kidney (HEK) 293 cells showed that the single amino-acid substitution of the KE family resulted in a form of the FOXP2 protein that was unable to bind target DNA at its usual consensus site and that severely interfered with gene regulation. The functional consequences of this small change to the protein sequence help to explain at a molecular level how this mutation may lead to the severe CAS phenotype⁵⁴. Such *in vitro* systems have been used to characterize six other independent rare FOXP2 variants for their subcellular localization, transcriptional regulation activity and protein-protein interactions⁵³. Two mutations showed disrupted nuclear localization, lack of transcriptional regulation activity, and reduced dimerization with other FOXP proteins, similar to the variant found in the KE family. The other FOXP2 variants investigated in that study acted like wild-type protein suggesting that some of these mutations may be incidental to the phenotype or that they have effects that are not detectable by these assays. Neuroblastoma cell lines have further been used to study the effect of FOXP2 on cellular processes, showing that its expression promotes retinoic acid induced neurite outgrowth and reduces cell migration²⁰¹. The controlled and simplified nature of cell lines makes them ideal to characterize fundamental properties and individual mutations, but they also have important limitations. The immortalization process involved in creating a cell line introduces drastic changes in karyotype, cell cycle, and morphology. This means that these cells cannot be considered normal cells, such that even cell lines that are derived from brain cells are not equivalent to neurons found *in vivo*. Further, cell line cultures are usually homogenous monolayers that lack the structure and complexity of brain tissue. These factors limit how informative observations in these models can be for understanding neurodevelopment in living organisms.

Animal models provide the unique opportunity to manipulate and study the neurogenetics of a living and behaving organism. This creates the possibility to study the functions of genes in different brain regions, neural circuits, developmental stages, and behavioural contexts. Histological studies can reveal the spatial and temporal distribution of gene products in the developing brain and identify brain areas of interest. Physiological studies can measure the effects of a gene knockout on the activity and connectivity in single neurons or a network of neurons, to give insight into how brain areas function together.

The role of a gene in the development of these neural circuits can be further studied with conditional knockouts that remove a gene after a certain time point and/or only in a certain area. Animal models have the added benefit that they allow the study of effects of genes on behaviours in a controlled way.

Employing diverse species as animal models further allows to investigate more ultimate questions about gene function by integrating evolutionary difference and convergence from the genetic and phenotypic level. However, evolutionary differences also mean that findings from animal models carry a translational challenge, as non-human animals do not speak. In mice, it has been proposed that deficits in motor skill learning in *Foxp2* mutant mice may reflect the deficits in orofacial motor control in affected human individuals²⁰². Knockdown experiments have shown that songbirds with experimentally reduced amounts of *FoxP2* in the Area X song nucleus during song development show incomplete and inaccurate vocal imitation, which has been argued to be reminiscent of the speech sequencing problems seen in humans carrying *FOXP2* mutations²⁰³. However, it is not straightforward to map human traits, such as speech and language, onto animal behaviour, and vice versa. In addition, species-specific aspects of human brain development may mean that neurobiological phenotypes that are important for language could be missed in animal models. The evolution of human language has been influenced by changes in the genome, transcriptome, and neurodevelopment. Mouse and rat *Cntnap2* knockout models show that mutations in a highly conserved gene can produce marked differences in behaviour and seizure phenotypes between species, with neither model fully recapitulating the full spectrum of human CDD symptoms²⁰⁴. These discrepancies emphasize that cross-species comparisons are important to understand gene function, and illustrate the importance of establishing human-specific neurodevelopmental models to most effectively study neuronal features that are affected by mutations in language-related genes.

1.4.2 Human-specific features of the genome, transcriptome, and neurodevelopment

Human stem cell-derived neuronal models are a relatively recent addition to the toolkit for studying language-related gene function. These models recapitulate the trajectory of early brain development *in vitro* and give rise to multiple cell types with varying levels of self-organisation. In this way, they combine the experimental control of cell lines with the improved complexity and structure of animal models. In addition, these models account for several human-specific features: at the genomic, transcriptomic, and developmental level.

Marked differences exist between the genomes of humans and our closest relatives, chimpanzees. Human-specific nucleotide changes constitute around 1.2% of our genome and larger insertions and deletions about 3%^{205,206}. Additional complex human-specific genetic changes have started being resolved, as long-read sequencing and reference-free assemblies help to systematically identify structural variants and to assign variants better to specific lineages^{207,208}.

Most human-specific genetic changes are located in non-protein coding regions of the genome with only a small fraction of changes altering amino acid sequences²⁰⁵. Variants in non-coding sequences that may regulate gene expression have indeed been found to have

accumulated in specific patterns on the human lineage, suggesting that these changes gave a selective advantage^{210,211}. More specifically, transcription factor binding sites have been shown to have signs of positive selection during human evolution^{212,213}. Positive selection has also been detected for FOXP2 binding sites in European genomes with an especially strong signal near *CNTNAP2*²¹⁴. This suggests that subtle genomic differences have reconfigured molecular networks by changing DNA-protein interactions that may have led to human-specific changes in gene expression during brain development.

Differences in gene expression patterns have been observed in homologous brain regions and cell types of humans and non-human primates²¹⁵⁻²²³. Candidate genes with divergent expression may influence evolved human traits²²¹. However, the interpretation of divergence in gene expression between species is complex. Not all expression changes will have a phenotypic effect²²⁴. Patterns of human-specific differential expression also vary across brain regions^{218,221,222}. This indicates that these regions relate to specific functions, but region-specific differences also include artefacts resulting from changes in cell type composition^{223,225}. Additionally, the brains of humans and non-human primates do not develop at the same speed. Gene expression data from developing brains show divergence in developmental trajectories in humans, including prolonged expression of subsets of genes involved in neural and synaptic development^{219,220,225}. Recent technical advancements have increased the ability to reliably detect these complex spatiotemporal patterns of divergence in gene expression. Single cell RNA sequencing improves the detection cell type-specific expression changes, of which two thirds are missed by bulk sequencing techniques²²³. These data indicate that some cell types, such as astrocytes and oligodendrocyte progenitors, show accelerated transcriptional divergence in the human lineage compared to neurons²²³. Single-cell transcriptome atlases also provide new insights in gene function. For example, a recent study of the dorsolateral prefrontal cortex of human, chimpanzee, macaque and marmoset found that FOXP2 has a human-specific expression pattern and regulatory network in microglia²²⁶. Finally, stem cell-derived models gives access to great ape brain tissue and a window into the developmental gene expression differences between humans and non-human primates^{227,228}. A comparison of human, chimpanzee and macaque brain organoids confirmed that human neuronal development occurs at a slower pace and that human-specific gene expression changes may be specific to certain cell states on the progenitor-to-neuron trajectory²²⁹. These species-specific transcriptome differences are thus also reflected in human neuronal models and affect processes that are likely relevant when investigating the functions of language-related genes.

Development is an area that might be particularly sensitive to evolutionary changes. A slight alteration in timing, intensity or location of events during development can result in significantly different outcomes. The expansion of the neocortex is one of the most marked species-specific aspects of the human brain and may have affected human cognitive abilities²³⁰. Theories for the cellular mechanism underlying this expansion suggest that it is due to an increased number of neural progenitor cells (NPCs)²³¹. One explanation for this expansion could be human-specific properties in the mode of cell division of NPCs. The mode of cell division determines the ability of NPCs for self-amplification and thus their pool size. It has been shown that human radial glia undergo a larger amount of non-vertical division than mouse radial glia cells²³². Species-specific NPC behaviour has also been detected in multiple stem cell-derived models²³³. The mode of cell division

is regulated via mitotic spindle orientation by genes such as ASPM and MCPH1^{234,235}, which appear positively selected during human evolution²³⁶⁻²³⁸ and are implicated in human microcephaly^{239,240}. Further, the human cortex contains specialized populations of progenitor cells that are virtually absent in rodents, such as basal or outer radial glia progenitors²⁴¹. It has been shown that a human-specific gene ARHGAP11B amplifies the proliferation of this type of progenitor cells and is capable of causing neocortex folding in the mouse^{242,243}. Expression of ARHGAP11B in chimpanzee organoids doubles the number of cycling basal progenitors, providing insight into how this human-specific gene may have contributed to neocortex expansion during recent evolution²⁴⁴. NPC-related development and interactions with human-specific genes may thus be particularly relevant aspects to study when investigating human neuronal features that contribute to language. Such human-specific aspects of neurodevelopment can be taken into account if human neuronal models are employed for these studies.

1.4.3 Human neuronal models for investigating neurodevelopmental disorder

The field of human neuronal models has been rapidly evolving and there is a wide choice of models that can suit the study of language-related gene function. Several differentiation protocols have been developed to generate neuronal cultures from human cells²⁴⁵. I will briefly outline the main options for these methods. The selected method determines which research questions can be answered, and thus choices must be made. First, a protocol can be selected based on which region of the developing brain and which cell types and developmental steps should be included. Secondly, the source of cells for a differentiation model has genetic and epigenetic implications for the research project. Additional practical considerations include the time it takes to generate these cultures and reproducibility.

The generation of *in vitro* neuronal models can be divided into two main methods. The first method, termed trans-differentiation or direct conversion, directly induces neurons from somatic cells by overexpressing specific combinations of transcription factors and microRNAs²⁴⁶. These methods rapidly and robustly generate homogeneous cell populations, but they do not faithfully model neuronal development as several intermediate steps are missed. Co-culturing with cells from a different source is needed to include multiple cell types in direct conversion cultures²⁴⁶, such as astrocytes that are necessary for the development of mature synapses in neuronal networks²⁴⁷. The second method relies on human pluripotent stem cells, which are differentiated into heterogeneous cultures and can recapitulate different *in vivo* developmental stages. For this, human embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) can be used. These cells can be differentiated in two-dimensional and three-dimensional cultures. Two-dimensional cultures can be generated by adding growth factors or small molecules to the medium to generate NPCs, which can then be further differentiated into neurons²⁴⁸⁻²⁵⁰. Three-dimensional cultures, also called organoids, can be classified in less directed and highly directed protocols. Cells are grown in suspension and, over a period lasting from weeks to months, form structures that capture more of the layered architecture and cellular micro-environment of *in vivo* brain development due to the combination of external growth factor patterning, and intrinsic and environmental cues^{251,252}. The protocol becomes more directed when further signalling molecules and growth

factors are added to the medium to guide the culture to differentiate into a specific brain region. Differentiation protocols thus vary in their source material, level of directed differentiation and complexity and architecture of the generated cultures.

The selection of an adequate differentiation protocol involves several considerations. First, the brain region one intends to model or the expression pattern of a gene of interest will influence the type of model. Several conversion and differentiation protocols have been developed with varying specificities for cell type and brain region ²⁵³⁻²⁵⁵. For example, it is possible to directly induce excitatory ²⁵⁶ and inhibitory neurons ²⁵⁷, and directed differentiation protocols can be specified to generate forebrain ²⁵⁸, midbrain ²⁵⁹ and cerebellar organoids ²⁶⁰. It is also important to consider the complexity of a research question. In order to investigate cell-autonomous disease processes, it may be sufficient to study a homogeneous culture of induced neurons with a very specific identity. However, if the hypothesized disease mechanism includes multiple developmental stages and cell types, a directed differentiation protocol might be more suitable to capture cell-cell interactions of a neuronal network. If the micro-environment and high levels of connectivity are particularly important, a three-dimensional protocol might be most suitable.

Second, the source of cells presents choices about genetic and epigenetic background in a model. ESCs are considered the gold standard for pluripotency, which makes them an ideal source for studying early brain development ^{261,262}. However, it is difficult to predict whether background mutations found in an embryonic cell line are potentially pathogenic, because these cell lines never developed into a human being ²⁶³. Here, hiPSCs offer the possibility to capture the genetic signature of a living individual and recapitulate that individual's neurodevelopment *in vitro*, as somatic cells can be reprogrammed or reverse-differentiated into an ESC-like state by introduction of key transcription factors ²⁶⁴. It should be noted that hiPSC clones sometimes carry residual DNA methylation signatures from somatic cells or somatic mosaic mutations from founder cells, which can be sources for variation and may require additional replicates ^{265,266}. For somatic cells that are trans-differentiated into neurons, this retention of epigenetic signatures is even stronger and reflects donor age ^{267,268}, which is a disadvantage for modelling early developmental disorders. Background genetic variation between cell lines has been a major confounding factor in experiments ²⁶⁹. Thanks to genome engineering techniques such as CRISPR/Cas9, this issue for modelling disorders can be addressed by creating experimental conditions that share a single genetic background except for an engineered mutation of interest ²⁷⁰. Using isogenic pairs, a single gene or isoform can be studied in a control ESC or iPSC line to answer more fundamental questions about gene function. Alternatively, a disease-causing mutation can be rescued to answer more disease-specific questions.

Finally, there are more practical considerations. A more complex culture is generally more variable and will require a higher number of replicates and careful analysis to find an effect ^{255,271}. More complex cultures also come at a higher cost in terms of resources and time that is needed to generate these cultures, which reduces scalability. Next to the neurobiological and genetic research questions, this trade-off between homogeneity and complexity is thus highly relevant when designing a feasible project. The right model can be selected by taking into account the expression pattern of language-related genes,

the complexity of the neurobiological question, the type of mutation to model and the available time and resources.

1.5 Aims and scope of this thesis

The biological basis of language cannot be understood without investigating the molecular and neurobiological substrates that contribute to human cognition. Genetic studies have led to the identification of new candidate language-related genes. Insight into the molecular networks of the products that these genes encode can help explain which aspects of the human brain are needed for normal language development, and potentially identify common features of different language-related genes. However, there is a need for appropriate models to investigate the molecular and cellular effects of genetic variants on the human nervous system. The aim of this thesis is to understand how language-related genes and gene networks influence human neuronal features that contribute to the development of language-related traits.

In **chapter 2**, I describe the clinical diversity of CASPR2 deficiency disorder (CDD) phenotypes. I report the special case of a boy with hyperkinetic stereotyped movements that are unlike other dyskinesias and have not been reported in the case of CDD. This individual inherited a missense mutation and partial duplication in *CNTNAP2* from his father and mother, respectively. These mutations suggest a layered molecular mechanism underlying the phenotype, as they affect the protein itself and sites for transcriptional regulation by FOXP2. The varying functional implications of mutations in *CNTNAP2* expand our understanding of CDD and emphasize the importance of screening regulatory regions in patients with CDD-suggestive phenotypes and atypical neurological symptoms.

In **chapter 3**, I investigate how mutations in *CNTNAP2* can lead to varying clinical outcomes and if a previously unstudied short isoform plays a role in this. To this end, I measure the effects of multiple mutations in *CNTNAP2* on cell type composition, synapses, and global gene expression of human ESC-derived neuronal networks. I demonstrate that homozygous loss of CASPR2 affects the proportion of neural precursor cells, the number of excitatory synapses, and gene expression for several biological processes, whereas heterozygous loss only produces a subset of the gene expression changes. Gene expression data further reveal that the short isoform of *CNTNAP2* is expressed early in development at high levels and that its loss specifically disrupts the expression of cilia assembly genes. Further investigation of a role for *CNTNAP2* in cilia function during early brain development could help clarify its roles in neuronal migration and cortical lamination, and provide an explanation for clinical variation between cases with mutations that affect different isoforms.

In **chapter 4**, I study how neurodevelopmental perturbations caused by loss of *CNTNAP2* affect the miRNA regulatory system in human neuronal networks. First, I characterize the developmental expression pattern of microRNAs and their targets in the wild-type neuronal differential model from chapter 3. Second, I use this baseline to explore how miRNA networks change when neurodevelopment is disrupted in a model for *CNTNAP2* deficiency disorder. Understanding perturbations of miRNA networks in this model offers further insight into the molecular networks that are related to *CNTNAP2* and identifies genomic targets that may allow further characterization of the clinical heterogeneity

observed in patients.

In **chapter 5**, I investigate the function of *ARHGEF39*, a gene showing association with developmental language disorder involving a polymorphism in the 3'UTR that disrupts post-transcriptional regulation by microRNA. Here, I show that *ARHGEF39* activates RHOA, a member of the Rho family of GTPases that regulate intracellular actin dynamics, and that overexpression of *ARHGEF39* leads to cell de-adhesion. Single cell RNA-sequencing data show that *ARHGEF39* is a marker gene for proliferating progenitor cells, and that its expression is associated with cell division and other RHOA regulating genes. The activation of RHOA and expression of *ARHGEF39* in proliferating neural progenitor cells suggest that *ARHGEF39* has a role in cortical size development. Further investigation of the effect of *ARHGEF39* expression levels on neural progenitor cell division and attachment could help to understand its role in neocortical development and may provide an explanation for the contribution of altered *ARHGEF39* expression levels to the risk for language impairments.

Finally, in **chapter 6**, I summarize the research of the thesis, and discuss how results from these studies have expanded our knowledge about human neurobiological processes that are affected by *CNTNAP2*, *ARHGEF39*, and microRNA networks, as well as putting forward ideas for further research into molecular and neurodevelopmental mechanisms that contribute to the development of language-related traits.

References

- 1 Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T. & the, C.-c. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J Child Psychol Psychiatry* 58, 1068-1080, doi:10.1111/jcpp.12721.
- 2 Fisher, S. E. (2006). Tangled webs: tracing the connections between genes and cognition. *Cognition* 101, 270-297, doi:10.1016/j.cognition.2006.04.004.
- 3 Geschwind, D. H. & Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science* 349, 1489-1494, doi:10.1126/science.aaa8954.
- 4 Pennington, B. F. & Bishop, D. V. (2009). Relations among speech, language, and reading disorders. *Annu Rev Psychol* 60, 283-306, doi:10.1146/annurev.psych.60.110707.163548.
- 5 Bishop, D. V. (2010). Overlaps between autism and language impairment: phenomimicry or shared etiology? *Behav Genet* 40, 618-629, doi:10.1007/s10519-010-9381-x.
- 6 Nudel, R. et al. (2014). Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes Brain Behav* 13, 418-429, doi:10.1111/gbb.12127.
- 7 Devanna, P. et al. (2018). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol Psychiatry* 23, 1375-1384, doi:10.1038/mp.2017.30.
- 8 Eicher, J. D. & Gruen, J. R. (2015). Language impairment and dyslexia genes influence language skills in children with autism spectrum disorders. *Autism Res* 8, 229-234, doi:10.1002/aur.1436.
- 9 Newbury, D. F. et al. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet* 85, 264-272, doi:10.1016/j.ajhg.2009.07.004.
- 10 Smith, A. W., Holden, K. R., Dwivedi, A., Dupont, B. R. & Lyons, M. J. (2015). Deletion of 16q24.1 supports a role for the ATP2C2 gene in specific language impairment. *J Child Neurol* 30, 517-521, doi:10.1177/0883073814545113.
- 11 Newbury, D. F. et al. (2011). Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. *Behav Genet* 41, 90-104, doi:10.1007/s10519-010-9424-3.
- 12 Scerri, T. S. et al. (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biol Psychiatry* 70, 237-245, doi:10.1016/j.biopsych.2011.02.005.
- 13 Villanueva, P. et al. (2011). Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. *Eur J Hum Genet* 19, 687-695, doi:10.1038/ejhg.2010.251.
- 14 Peter, B. et al. (2011). Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *J Neurodev Disord* 3, 39-49, doi:10.1007/s11689-010-9065-0.

- 15 Vernes, S. C. *et al.* (2008). A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359, 2337-2345, doi:10.1056/NEJMoa0802828.
- 16 Whitehouse, A. J., Bishop, D. V., Ang, Q. W., Pennell, C. E. & Fisher, S. E. (2011). CNTNAP2 variants affect early language development in the general population. *Genes Brain Behav* 10, 451-456, doi:10.1111/j.1601-183X.2011.00684.x.
- 17 Al-Murrani, A., Ashton, F., Aftimos, S., George, A. M. & Love, D. R. (2012). Amino-Terminal Microdeletion within the CNTNAP2 Gene Associated with Variable Expressivity of Speech Delay. *Case Rep Genet* 2012, 172408, doi:10.1155/2012/172408.
- 18 O’Roak, B. J. *et al.* (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 43, 585-589, doi:10.1038/ng.835.
- 19 Rodenas-Cuadrado, P. *et al.* (2016). Characterisation of CASPR2 deficiency disorder—a syndrome involving autism, epilepsy and language impairment. *BMC Med Genet* 17, 8, doi:10.1186/s12881-016-0272-8.
- 20 Strauss, K. A. *et al.* (2006). Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med* 354, 1370-1377, doi:10.1056/NEJMoa052773.
- 21 Chiocchetti, A. G. *et al.* (2015). Variants of the CNTNAP2 5' promoter as risk factors for autism spectrum disorders: a genetic and functional approach. *Mol Psychiatry* 20, 839-849, doi:10.1038/mp.2014.103.
- 22 Falivelli, G. *et al.* (2012). Inherited genetic variants in autism-related CNTNAP2 show perturbed trafficking and ATF6 activation. *Hum Mol Genet* 21, 4761-4773, doi:10.1093/hmg/dds320.
- 23 Canali, G. *et al.* (2018). Genetic variants in autism-related CNTNAP2 impair axonal growth of cortical neurons. *Hum Mol Genet* 27, 1941-1954, doi:10.1093/hmg/ddy102.
- 24 De Jong, J. O. *et al.* (2021). Cortical overgrowth in a preclinical forebrain organoid model of CNTNAP2-associated autism spectrum disorder. *Nat Commun* 12, 4087, doi:10.1038/s41467-021-24358-4.
- 25 Villanueva, P. *et al.* (2015). Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. *PLoS Genet* 11, e1004925, doi:10.1371/journal.pgen.1004925.
- 26 Ceroni, F. *et al.* (2014). Homozygous microdeletion of exon 5 in ZNF277 in a girl with specific language impairment. *Eur J Hum Genet* 22, 1165-1171, doi:10.1038/ejhg.2014.4.
- 27 Eising, E. *et al.* (2019). A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol Psychiatry* 24, 1065-1078, doi:10.1038/s41380-018-0020-x.
- 28 Snijders Blok, L. *et al.* (2018). CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. *Nat Commun* 9, 4619, doi:10.1038/s41467-018-06014-6.
- 29 Thevenon, J. *et al.* (2013). 12p13.33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech. *Eur J Hum Genet* 21, 82-88, doi:10.1038/ejhg.2012.116.
- 30 Bacon, C. & Rappold, G. A. (2012). The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Hum Genet* 131, 1687-1698, doi:10.1007/s00439-012-1193-z.
- 31 Hamdan, F. F. *et al.* (2010). De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *Am J Hum Genet* 87, 671-678, doi:10.1016/j.ajhg.2010.09.017.
- 32 Sollis, E. *et al.* (2016). Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. *Hum Mol Genet* 25, 546-557, doi:10.1093/hmg/ddv495.
- 33 Lozano, R., Vino, A., Lozano, C., Fisher, S. E. & Deriziotis, P. (2015). A de novo FOXP1 variant in a patient with autism, intellectual disability and severe speech and language impairment. *Eur J Hum Genet* 23, 1702-1707, doi:10.1038/ejhg.2015.66.
- 34 Estruch, S. B. *et al.* (2018). Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. *Hum Mol Genet* 27, 1212-1227, doi:10.1093/hmg/ddy035.
- 35 Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519-523, doi:10.1038/35097076.
- 36 Laffin, J. J. *et al.* (2012). Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. *Genet Med* 14, 928-936, doi:10.1038/gim.2012.72.
- 37 MacDermot, K. D. *et al.* (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet* 76, 1074-1080, doi:10.1086/430841.
- 38 Turner, S. J. *et al.* (2013). Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. *Am J Med Genet A* 161A, 2321-2326, doi:10.1002/ajmg.a.36055.
- 39 Horn, D. *et al.* (2010). Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. *Hum Mutat* 31, E1851-1860, doi:10.1002/humu.21362.
- 40 Shriberg, L. D. *et al.* (2006). Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *J Speech Lang Hear Res* 49, 500-525, doi:10.1044/1092-4388(2006/038).

- 41 Lai, C. S. *et al.* (2000). The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am J Hum Genet* 67, 357-368, doi:10.1086/303011.
- 42 Tomblin, J. B. *et al.* (2009). Language features in a mother and daughter of a chromosome 7;13 translocation involving FOXP2. *J Speech Lang Hear Res* 52, 1157-1174, doi:10.1044/1092-4388(2009/07-0162).
- 43 Patel, C. *et al.* (2011). Translocation breakpoint at 7q31 associated with tics: further evidence for IMMP2L as a candidate gene for Tourette syndrome. *Eur J Hum Genet* 19, 634-639, doi:10.1038/ejhg.2010.238.
- 44 Feuk, L. *et al.* (2006). Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *Am J Hum Genet* 79, 965-972, doi:10.1086/508902.
- 45 Lennon, P. A. *et al.* (2007). Deletion of 7q31.1 supports involvement of FOXP2 in language impairment: clinical report and review. *Am J Med Genet A* 143A, 791-798, doi:10.1002/ajmg.a.31632.
- 46 Zeesman, S. *et al.* (2006). Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. *Am J Med Genet A* 140, 509-514, doi:10.1002/ajmg.a.31110.
- 47 Zilina, O. *et al.* (2012). Maternally and paternally inherited deletion of 7q31 involving the FOXP2 gene in two families. *Am J Med Genet A* 158A, 254-256, doi:10.1002/ajmg.a.34378.
- 48 Rice, G. M. *et al.* (2012). Phenotype of FOXP2 haploinsufficiency in a mother and son. *Am J Med Genet A* 158A, 174-181, doi:10.1002/ajmg.a.34354.
- 49 Adegbola, A. A. *et al.* (2015). Monoallelic expression of the human FOXP2 speech gene. *Proc Natl Acad Sci U S A* 112, 6848-6854, doi:10.1073/pnas.1411270111.
- 50 Nazaryan, L. *et al.* (2014). The strength of combined cytogenetic and mate-pair sequencing techniques illustrated by a germline chromothripsis rearrangement involving FOXP2. *Eur J Hum Genet* 22, 338-343, doi:10.1038/ejhg.2013.147.
- 51 Palka, C. *et al.* (2012). Mosaic 7q31 deletion involving FOXP2 gene associated with language impairment. *Pediatrics* 129, e183-188, doi:10.1542/peds.2010-2094.
- 52 Utine, G. E. *et al.* (2014). Etiological yield of SNP microarrays in idiopathic intellectual disability. *Eur J Paediatr Neurol* 18, 327-337, doi:10.1016/j.ejpn.2014.01.004.
- 53 Estruch, S. B., Graham, S. A., Chinnappa, S. M., Deriziotis, P. & Fisher, S. E. (2016). Functional characterization of rare FOXP2 variants in neurodevelopmental disorder. *J Neurodev Disord* 8, 44, doi:10.1186/s11689-016-9177-2.
- 54 Vernes, S. C. *et al.* (2006). Functional genetic analysis of mutations implicated in a human speech and language disorder. *Hum Mol Genet* 15, 3154-3167, doi:10.1093/hmg/ddl392.
- 55 Snijders Blok, L. *et al.* (2021). Heterozygous variants that disturb the transcriptional repressor activity of FOXP4 cause a developmental disorder with speech/language delays and multiple congenital abnormalities. *Genet Med* 23, 534-542, doi:10.1038/s41436-020-01016-6.
- 56 Ludwig, K. U. *et al.* (2010). Variation in GRIN2B contributes to weak performance in verbal short-term memory in children with dyslexia. *Am J Med Genet B Neuropsychiatr Genet* 153B, 503-511, doi:10.1002/ajmg.b.31007.
- 57 Lesca, G. *et al.* (2012). Epileptic encephalopathies of the Landau-Kleffner and continuous spike and waves during slow-wave sleep types: genomic dissection makes the link with autism. *Epilepsia* 53, 1526-1538, doi:10.1111/j.1528-1167.2012.03559.x.
- 58 Lesca, G. *et al.* (2013). GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat Genet* 45, 1061-1066, doi:10.1038/ng.2726.
- 59 Carvill, G. L. *et al.* (2013). GRIN2A mutations cause epilepsy-aphasia spectrum disorders. *Nat Genet* 45, 1073-1076, doi:10.1038/ng.2727.
- 60 Dimassi, S. *et al.* (2014). A subset of genomic alterations detected in rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. *Epilepsia* 55, 370-378, doi:10.1111/epi.12502.
- 61 Kornilov, S. A. *et al.* (2016). Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population. *Pediatrics* 137, doi:10.1542/peds.2015-2469.
- 62 Hildebrand, M. S. *et al.* (2020). Severe childhood speech disorder: Gene discovery highlights transcriptional dysregulation. *Neurology* 94, e2148-e2167, doi:10.1212/WNL.0000000000009441.
- 63 Schumacher, J. *et al.* (2006). Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am J Hum Genet* 78, 52-62, doi:10.1086/498992.
- 64 Deffenbacher, K. E. *et al.* (2004). Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. *Hum Genet* 115, 128-138, doi:10.1007/s00439-004-1126-6.
- 65 Meng, H. *et al.* (2005). DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci U S A* 102, 17053-17058, doi:10.1073/pnas.0508591102.
- 66 Matsson, H. *et al.* (2015). Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. *J*

- Hum Genet* 60, 399-401, doi:10.1038/jhg.2015.37.
- 67 Eicher, J. D. *et al.* (2015). The DYX2 locus and neurochemical signaling genes contribute to speech sound disorder and related neurocognitive domains. *Genes Brain Behav* 14, 377-385, doi:10.1111/gbb.12214.
- 68 Zhong, R. *et al.* (2013). Meta-analysis of the association between DCDC2 polymorphisms and risk of dyslexia. *Mol Neurobiol* 47, 435-442, doi:10.1007/s12035-012-8381-7.
- 69 Marino, C. *et al.* (2012). DCDC2 genetic variants and susceptibility to developmental dyslexia. *Psychiatr Genet* 22, 25-30, doi:10.1097/YPG.0b013e32834acdb2.
- 70 Zhang, Y. *et al.* (2016). Association of DCDC2 Polymorphisms with Normal Variations in Reading Abilities in a Chinese Population. *PLoS One* 11, e0153603, doi:10.1371/journal.pone.0153603.
- 71 Siggberg, L. *et al.* (2011). Familial 6p22.2 duplication associates with mild developmental delay and increased SSADH activity. *Am J Med Genet B Neuropsychiatr Genet* 156B, 448-453, doi:10.1002/ajmg.b.31180.
- 72 Cicchini, G. M., Marino, C., Mascheretti, S., Perani, D. & Morrone, M. C. (2015). Strong motion deficits in dyslexia associated with DCDC2 gene alteration. *J Neurosci* 35, 8059-8064, doi:10.1523/JNEUROSCI.5077-14.2015.
- 73 Meng, H. *et al.* (2011). A dyslexia-associated variant in DCDC2 changes gene expression. *Behav Genet* 41, 58-66, doi:10.1007/s10519-010-9408-3.
- 74 Chapman, N. H. *et al.* (2004). Linkage analyses of four regions previously implicated in dyslexia: confirmation of a locus on chromosome 15q. *Am J Med Genet B Neuropsychiatr Genet* 131B, 67-75, doi:10.1002/ajmg.b.30018.
- 75 Schulte-Korne, G. *et al.* (1998). Evidence for linkage of spelling disability to chromosome 15. *Am J Hum Genet* 63, 279-282, doi:10.1086/301919.
- 76 Grigorenko, E. L. *et al.* (1997). Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am J Hum Genet* 60, 27-39.
- 77 Dahdouh, F. *et al.* (2009). Further evidence for DYX1C1 as a susceptibility factor for dyslexia. *Psychiatr Genet* 19, 59-63, doi:10.1097/YPG.0b013e32832080e1.
- 78 Bates, T. C. *et al.* (2010). Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation. *Mol Psychiatry* 15, 1190-1196, doi:10.1038/mp.2009.120.
- 79 Wigg, K. G. *et al.* (2004). Support for EKN1 as the susceptibility locus for dyslexia on 15q21. *Mol Psychiatry* 9, 1111-1121, doi:10.1038/sj.mp.4001543.
- 80 Brkanac, Z. *et al.* (2007). Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. *Am J Med Genet B Neuropsychiatr Genet* 144B, 556-560, doi:10.1002/ajmg.b.30471.
- 81 Marino, C. *et al.* (2007). Association of short-term memory with a variant within DYX1C1 in developmental dyslexia. *Genes Brain Behav* 6, 640-646, doi:10.1111/j.1601-183X.2006.00291.x.
- 82 Tran, C. *et al.* (2013). A family-based association analysis and meta-analysis of the reading disabilities candidate gene DYX1C1. *Am J Med Genet B Neuropsychiatr Genet* 162B, 146-156, doi:10.1002/ajmg.b.32123.
- 83 Nopola-Hemmi, J. *et al.* (2000). Two translocations of chromosome 15q associated with dyslexia. *J Med Genet* 37, 771-775, doi:10.1136/jmg.37.10.771.
- 84 Taipale, M. *et al.* (2003). A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* 100, 11553-11558, doi:10.1073/pnas.1833911100.
- 85 Tapia-Paez, I., Tammimies, K., Massinen, S., Roy, A. L. & Kere, J. (2008). The complex of TFII-I, PARP1, and SFPQ proteins regulates the DYX1C1 gene implicated in neuronal migration and dyslexia. *FASEB J* 22, 3001-3009, doi:10.1096/fj.07-104455.
- 86 Tammimies, K. *et al.* (2012). The rs3743205 SNP is important for the regulation of the dyslexia candidate gene DYX1C1 by estrogen receptor beta and DNA methylation. *Mol Endocrinol* 26, 619-629, doi:10.1210/me.2011-1376.
- 87 Anthoni, H. *et al.* (2007). A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Hum Mol Genet* 16, 667-677, doi:10.1093/hmg/ddm009.
- 88 Francks, C. *et al.* (2004). A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am J Hum Genet* 75, 1046-1058, doi:10.1086/426404.
- 89 Cope, N. *et al.* (2005). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet* 76, 581-591, doi:10.1086/429131.
- 90 Harold, D. *et al.* (2006). Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Mol Psychiatry* 11, 1085-1091, 1061, doi:10.1038/sj.mp.4001904.
- 91 Luciano, M. *et al.* (2007). A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability. *Biol Psychiatry* 62, 811-817, doi:10.1016/j.biopsych.2007.03.007.
- 92 Paracchini, S. *et al.* (2008). Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *Am J Psychiatry* 165, 1576-1584, doi:10.1176/appi.ajp.2008.07121872.
- 93 Mascheretti, S. *et al.* (2014). KIAA0319 and ROBO1: evidence on association with reading and pleiotropic

- effects on language and mathematics abilities in developmental dyslexia. *J Hum Genet* 59, 189-197, doi:10.1038/jhg.2013.141.
- 94 Rice, M. L., Smith, S. D. & Gayan, J. (2009). Convergent genetic linkage and associations to language, speech and reading measures in families of probands with Specific Language Impairment. *J Neurodev Disord* 1, 264-282, doi:10.1007/s11689-009-9031-x.
- 95 Zhao, H., Chen, Y., Zhang, B. P. & Zuo, P. X. (2016). KIAA0319 gene polymorphisms are associated with developmental dyslexia in Chinese Uyghur children. *J Hum Genet* 61, 745-752, doi:10.1038/jhg.2016.40.
- 96 Paracchini, S. *et al.* (2006). The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 15, 1659-1666, doi:10.1093/hmg/ddl089.
- 97 Dennis, M. Y. *et al.* (2009). A common variant associated with dyslexia reduces expression of the KIAA0319 gene. *PLoS Genet* 5, e1000436, doi:10.1371/journal.pgen.1000436.
- 98 Gialluisi, A. *et al.* (2021). Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. *Mol Psychiatry* 26, 3004-3017, doi:10.1038/s41380-020-00898-x.
- 99 Nopola-Hemmi, J. *et al.* (2001). A dominant gene for developmental dyslexia on chromosome 3. *J Med Genet* 38, 658-664, doi:10.1136/jmg.38.10.658.
- 100 Fisher, S. E. *et al.* (2002). Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nat Genet* 30, 86-91, doi:10.1038/ng792.
- 101 Stein, C. M. *et al.* (2004). Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *Am J Hum Genet* 74, 283-297, doi:10.1086/381562.
- 102 Bates, T. C. *et al.* (2011). Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. *Behav Genet* 41, 50-57, doi:10.1007/s10519-010-9402-9.
- 103 Tran, C. *et al.* (2014). Association of the ROBO1 gene with reading disabilities in a family-based analysis. *Genes Brain Behav* 13, 430-438, doi:10.1111/gbb.12126.
- 104 Massinen, S. *et al.* (2016). Genomic sequencing of a dyslexia susceptibility haplotype encompassing ROBO1. *J Neurodev Disord* 8, 4, doi:10.1186/s11689-016-9136-y.
- 105 Hannula-Jouppi, K. *et al.* (2005). The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* 1, e50, doi:10.1371/journal.pgen.0010050.
- 106 Lamminmaki, S., Massinen, S., Nopola-Hemmi, J., Kere, J. & Hari, R. (2012). Human ROBO1 regulates interaural interaction in auditory pathways. *J Neurosci* 32, 966-971, doi:10.1523/JNEUROSCI.4007-11.2012.
- 107 Poelmans, G. *et al.* (2009). Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. *Am J Med Genet B Neuropsychiatr Genet* 150B, 140-147, doi:10.1002/ajmg.b.30787.
- 108 St Pourcain, B. *et al.* (2014). Common variation near ROBO2 is associated with expressive vocabulary in infancy. *Nat Commun* 5, 4831, doi:10.1038/ncomms5831.
- 109 St Pourcain, B. *et al.* (2014). Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Mol Autism* 5, 18, doi:10.1186/2040-2392-5-18.
- 110 Namjou, B. *et al.* (2014). Phenome-wide association study (PheWAS) in EMR-linked pediatric cohorts, genetically links PLCL1 to speech language development and IL5-IL13 to Eosinophilic Esophagitis. *Front Genet* 5, 401, doi:10.3389/fgene.2014.00401.
- 111 Doust, C. *et al.* (2022). Discovery of 42 genome-wide significant loci associated with dyslexia. *Nat Genet* 54, 1621-1629, doi:10.1038/s41588-022-01192-y.
- 112 Eising, E. *et al.* (2022). Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proc Natl Acad Sci U S A* 119, e2202764119, doi:10.1073/pnas.2202764119.
- 113 Price, K. M. *et al.* (2022). Hypothesis-driven genome-wide association studies provide novel insights into genetics of reading disabilities. *Transl Psychiatry* 12, 495, doi:10.1038/s41398-022-02250-z.
- 114 Hannenhalli, S. & Kaestner, K. H. (2009). The evolution of Fox genes and their role in development and disease. *Nat Rev Genet* 10, 233-240, doi:10.1038/nrg2523.
- 115 Kaestner, K. H., Knochel, W. & Martinez, D. E. (2000). Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev* 14, 142-146.
- 116 Fisher, S. E. & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends Genet* 25, 166-177, doi:10.1016/j.tig.2009.03.002.
- 117 Lai, C. S., Gerrelli, D., Monaco, A. P., Fisher, S. E. & Copp, A. J. (2003). FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* 126, 2455-2462, doi:10.1093/brain/awg247.
- 118 Reuter, M. S. *et al.* (2017). FOXP2 variants in 14 individuals with developmental speech and language

- disorders broaden the mutational and clinical spectrum. *J Med Genet* 54, 64-72, doi:10.1136/jmedgenet-2016-104094.
- 119 Morgan, A., Fisher, S. E., Scheffer, I. & Hildebrand, M. in *GeneReviews*((R)) (eds M. P. Adam et al.) (University of Washington, Seattle, 1993).
- 120 Morison, L. D. et al. (2023). In-depth characterisation of a cohort of individuals with missense and loss-of-function variants disrupting FOXP2. *J Med Genet* 60, 597-607, doi:10.1136/jmg-2022-108734.
- 121 Hoischen, A., Krumm, N. & Eichler, E. E. (2014). Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nat Neurosci* 17, 764-772, doi:10.1038/nn.3703.
- 122 Chen, X. S. et al. (2017). Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Sci Rep* 7, 46105, doi:10.1038/srep46105.
- 123 Rodenas-Cuadrado, P., Ho, J. & Vernes, S. C. (2014). Shining a light on CNTNAP2: complex functions to complex disorders. *Eur J Hum Genet* 22, 171-178, doi:10.1038/ejhg.2013.100.
- 124 Toma, C. et al. (2018). Comprehensive cross-disorder analyses of CNTNAP2 suggest it is unlikely to be a primary risk gene for psychiatric disorders. *PLoS Genet* 14, e1007535, doi:10.1371/journal.pgen.1007535.
- 125 Anney, R. et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 21, 4781-4792, doi:10.1093/hmg/dds301.
- 126 Alarcon, M. et al. (2008). Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet* 82, 150-159, doi:10.1016/j.ajhg.2007.09.005.
- 127 Ross, L. A. et al. (2017). Common variation in the autism risk gene CNTNAP2, brain structural connectivity and multisensory speech integration. *Brain Lang* 174, 50-60, doi:10.1016/j.bandl.2017.07.005.
- 128 Zhu, B. et al. (2017). Associations between the CNTNAP2 gene, dorsolateral prefrontal cortex, and cognitive performance on the Stroop task. *Neuroscience* 343, 21-29, doi:10.1016/j.neuroscience.2016.11.021.
- 129 Whalley, H. C. et al. (2011). Genetic variation in CNTNAP2 alters brain function during linguistic processing in healthy individuals. *Am J Med Genet B Neuropsychiatr Genet* 156B, 941-948, doi:10.1002/ajmg.b.31241.
- 130 Folia, V., Forkstam, C., Ingvar, M., Hagoort, P. & Petersson, K. M. (2011). Implicit Artificial Syntax Processing: Genes, Preference, and Bounded Recursion. *Biolinguistics* 5, 28.
- 131 Kos, M. et al. (2012). CNTNAP2 and language processing in healthy individuals as measured with ERPs. *PLoS One* 7, e46995, doi:10.1371/journal.pone.0046995.
- 132 Koeda, M. et al. (2015). Interaction effect between handedness and CNTNAP2 polymorphism (rs7794745 genotype) on voice-specific frontotemporal activity in healthy individuals: an fMRI study. *Front Behav Neurosci* 9, 87, doi:10.3389/fnbeh.2015.00087.
- 133 Jahanshad, N. et al. (2017). Do Candidate Genes Affect the Brain's White Matter Microstructure? Large-Scale Evaluation of 6,165 Diffusion MRI Scans. *bioRxiv*, 107987, doi:10.1101/107987.
- 134 Clemm von Hohenberg, C. et al. (2013). CNTNAP2 polymorphisms and structural brain connectivity: a diffusion-tensor imaging study. *J Psychiatr Res* 47, 1349-1356, doi:10.1016/j.jpsychires.2013.07.002.
- 135 Gupta, C. N. et al. (2015). Genetic markers of white matter integrity in schizophrenia revealed by parallel ICA. *Front Hum Neurosci* 9, 100, doi:10.3389/fnhum.2015.00100.
- 136 Udden, J., Snijders, T. M., Fisher, S. E. & Hagoort, P. (2017). A common variant of the CNTNAP2 gene is associated with structural variation in the left superior occipital gyrus. *Brain Lang* 172, 16-21, doi:10.1016/j.bandl.2016.02.003.
- 137 Tan, G. C., Doke, T. F., Ashburner, J., Wood, N. W. & Frackowiak, R. S. (2010). Normal variation in fronto-occipital circuitry and cerebellar structure with an autism-associated polymorphism of CNTNAP2. *Neuroimage* 53, 1030-1042, doi:10.1016/j.neuroimage.2010.02.018.
- 138 Carrion-Castillo, A., Franke, B. & Fisher, S. E. (2013). Molecular genetics of dyslexia: an overview. *Dyslexia* 19, 214-240, doi:10.1002/dys.1464.
- 139 Reader, R. H., Covill, L. E., Nudel, R. & Newbury, D. F. (2014). Genome-Wide Studies of Specific Language Impairment. *Curr Behav Neurosci Rep* 1, 242-250, doi:10.1007/s40473-014-0024-z.
- 140 Graham, S. A. & Fisher, S. E. (2015). Understanding Language from a Genomic Perspective. *Annu Rev Genet* 49, 131-160, doi:10.1146/annurev-genet-120213-092236.
- 141 Rice, M. & Tager-Flusberg, H. in *Neurophenotypes Innovations in Cognitive Neuroscience* (eds Vinoth Jagaroo & Susan L. Santangelo) Ch. Chapter 12, 227-243 (Springer US, 2016).
- 142 Grant, S. G. (2003). Systems biology in neuroscience: bridging genes to cognition. *Curr Opin Neurobiol* 13, 577-582, doi:10.1016/j.conb.2003.09.016.
- 143 Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet* 16, 441-458, doi:10.1038/nrg3934.
- 144 Carter, S. L., Brechbuhler, C. M., Griffin, M. & Bond, A. T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 20, 2242-2250, doi:10.1093/bioinformatics/bth234.

- 145 Lee, T. I. & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* 152, 1237-1251, doi:10.1016/j.cell.2013.02.014.
- 146 Lambert, S. A. *et al.* (2018). The Human Transcription Factors. *Cell* 172, 650-665, doi:10.1016/j.cell.2018.01.029.
- 147 Gerstein, M. B. *et al.* (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91-100, doi:10.1038/nature11245.
- 148 Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet* 8, 450-461, doi:10.1038/nrg2102.
- 149 Den Hoed, J., Devaraju, K. & Fisher, S. E. (2021). Molecular networks of the FOXP2 transcription factor in the brain. *EMBO Rep* 22, e52803, doi:10.15252/embr.202152803.
- 150 Spiteri, E. *et al.* (2007). Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am J Hum Genet* 81, 1144-1157, doi:10.1086/522237.
- 151 Vernes, S. C. *et al.* (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am J Hum Genet* 81, 1232-1250, doi:10.1086/522238.
- 152 Vernes, S. C. *et al.* (2011). Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet* 7, e1002145, doi:10.1371/journal.pgen.1002145.
- 153 Druart, M., Groszer, M. & Le Magueresse, C. (2020). An Etiological Foxp2 Mutation Impairs Neuronal Gain in Layer VI Cortico-Thalamic Cells through Increased GABA(B)/GIRK Signaling. *J Neurosci* 40, 8543-8555, doi:10.1523/JNEUROSCI.2615-19.2020.
- 154 Enard, W. *et al.* (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137, 961-971, doi:10.1016/j.cell.2009.03.041.
- 155 Roll, P. *et al.* (2010). Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Hum Mol Genet* 19, 4848-4860, doi:10.1093/hmg/ddq415.
- 156 Mukamel, Z. *et al.* (2011). Regulation of MET by FOXP2, genes implicated in higher cognitive dysfunction and autism risk. *J Neurosci* 31, 11437-11442, doi:10.1523/JNEUROSCI.0181-11.2011.
- 157 Walker, R. M. *et al.* (2012). The DISC1 promoter: characterization and regulation by FOXP2. *Hum Mol Genet* 21, 2862-2872, doi:10.1093/hmg/dds111.
- 158 Chen, Y. C. *et al.* (2016). Foxp2 controls synaptic wiring of corticostriatal circuits and vocal communication by opposing Mef2c. *Nat Neurosci* 19, 1513-1522, doi:10.1038/nn.4380.
- 159 Becker, M., Devanna, P., Fisher, S. E. & Vernes, S. C. (2018). Mapping of Human FOXP2 Enhancers Reveals Complex Regulation. *Front Mol Neurosci* 11, 47, doi:10.3389/fnmol.2018.00047.
- 160 Bateman, E. (1998). Autoregulation of eukaryotic transcription factors. *Prog Nucleic Acid Res Mol Biol* 60, 133-168, doi:10.1016/s0079-6603(08)60892-2.
- 161 Bonkowsky, J. L. *et al.* (2008). Domain-specific regulation of foxP2 CNS expression by lef1. *BMC Dev Biol* 8, 103, doi:10.1186/1471-213X-8-103.
- 162 Coutinho, P. *et al.* (2011). Discovery and assessment of conserved Pax6 target genes and enhancers. *Genome Res* 21, 1349-1359, doi:10.1101/gr.124115.111.
- 163 Gotz, M., Stoykova, A. & Gruss, P. (1998). Pax6 controls radial glia differentiation in the cerebral cortex. *Neuron* 21, 1031-1044, doi:10.1016/s0896-6273(00)80621-2.
- 164 Thakurela, S. *et al.* (2016). Mapping gene regulatory circuitry of Pax6 during neurogenesis. *Cell Discov* 2, 15045, doi:10.1038/celldisc.2015.45.
- 165 Gan, Q. *et al.* (2014). Pax6 mediates ss-catenin signaling for self-renewal and neurogenesis by neocortical radial glial stem cells. *Stem Cells* 32, 45-58, doi:10.1002/stem.1561.
- 166 Kraushar, M. L. *et al.* (2015). Thalamic WNT3 Secretion Spatiotemporally Regulates the Neocortical Ribosome Signature and mRNA Translation to Specify Neocortical Cell Subtypes. *J Neurosci* 35, 10911-10926, doi:10.1523/JNEUROSCI.0601-15.2015.
- 167 Richter, G. *et al.* (2021). beta-catenin regulates FOXP2 transcriptional activity via multiple binding sites. *Febs j* 288, 3261-3284, doi:10.1111/febs.15656.
- 168 Stelzl, U. *et al.* (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122, 957-968, doi:10.1016/j.cell.2005.08.029.
- 169 Silverbush, D. & Sharan, R. (2019). A systematic approach to orient the human protein-protein interaction network. *Nat Commun* 10, 3015, doi:10.1038/s41467-019-10887-6.
- 170 Li, S., Weidenfeld, J. & Morrissey, E. E. (2004). Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Mol Cell Biol* 24, 809-822, doi:10.1128/MCB.24.2.809-822.2004.
- 171 Shi, Y. *et al.* (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* 422, 735-738, doi:10.1038/nature01550.

- 172 Chokas, A. L. *et al.* (2010). Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin-6. *J Biol Chem* 285, 13304-13313, doi:10.1074/jbc.M109.088468.
- 173 Hickey, S. L., Berto, S. & Konopka, G. (2019). Chromatin Decondensation by FOXP2 Promotes Human Neuron Maturation and Expression of Neurodevelopmental Disease Genes. *Cell Rep* 27, 1699-1711 e1699, doi:10.1016/j.celrep.2019.04.044.
- 174 Traka, M. *et al.* (2003). Association of TAG-1 with Caspr2 is essential for the molecular organization of juxtaparanodal regions of myelinated fibers. *J Cell Biol* 162, 1161-1172, doi:10.1083/jcb.200305078.
- 175 Poliak, S. *et al.* (2003). Juxtaparanodal clustering of Shaker-like K⁺ channels in myelinated axons depends on Caspr2 and TAG-1. *J Cell Biol* 162, 1149-1160, doi:10.1083/jcb.200305018.
- 176 Penagarikano, O. *et al.* (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147, 235-246, doi:10.1016/j.cell.2011.08.040.
- 177 Anderson, G. R. *et al.* (2012). Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc Natl Acad Sci U S A* 109, 18120-18125, doi:10.1073/pnas.1216398109.
- 178 Varea, O. *et al.* (2015). Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. *Proc Natl Acad Sci U S A* 112, 6176-6181, doi:10.1073/pnas.1423205112.
- 179 Chen, N. *et al.* (2015). Interaction proteomics of canonical Caspr2 (CNTNAP2) reveals the presence of two Caspr2 isoforms with overlapping interactomes. *Biochim Biophys Acta* 1854, 827-833, doi:10.1016/j.bbapap.2015.02.008.
- 180 Gao, R. *et al.* (2018). CNTNAP2 stabilizes interneuron dendritic arbors through CASK. *Mol Psychiatry* 23, 1832-1850, doi:10.1038/s41380-018-0027-3.
- 181 Hsueh, Y. P. (2006). The role of the MAGUK protein CASK in neural development and synaptic function. *Curr Med Chem* 13, 1915-1927, doi:10.2174/092986706777585040.
- 182 Argent, L. *et al.* (2020). Caspr2 interacts with type 1 inositol 1,4,5-trisphosphate receptor in the developing cerebellum and regulates Purkinje cell morphology. *J Biol Chem* 295, 12716-12726, doi:10.1074/jbc.RA120.012655.
- 183 Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19, 92-105, doi:10.1101/gr.082701.108.
- 184 Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297, doi:10.1016/s0092-8674(04)00045-5.
- 185 Boudreau, R. L. *et al.* (2014). Transcriptome-wide discovery of microRNA binding sites in human brain. *Neuron* 81, 294-305, doi:10.1016/j.neuron.2013.10.062.
- 186 Nowakowski, T. J. *et al.* (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat Neurosci* 21, 1784-1792, doi:10.1038/s41593-018-0265-3.
- 187 Rajman, M. & Schrantz, G. (2017). MicroRNAs in neural development: from master regulators to fine-tuners. *Development* 144, 2310-2322, doi:10.1242/dev.144337.
- 188 Somel, M. *et al.* (2011). MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol* 9, e1001214, doi:10.1371/journal.pbio.1001214.
- 189 Zhang, H. M. *et al.* (2015). Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Brief Bioinform* 16, 45-58, doi:10.1093/bib/bbt085.
- 190 Schizophrenia Psychiatric Genome-Wide Association Study, C. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 43, 969-976, doi:10.1038/ng.940.
- 191 Pardin, A. F. *et al.* (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 50, 381-389, doi:10.1038/s41588-018-0059-2.
- 192 Sun, G. *et al.* (2011). miR-137 forms a regulatory loop with nuclear receptor TLX and LSD1 in neural stem cells. *Nat Commun* 2, 529, doi:10.1038/ncomms1532.
- 193 Olde Loohuis, N. F. *et al.* (2015). MicroRNA-137 Controls AMPA-Receptor-Mediated Transmission and mGluR-Dependent LTD. *Cell Rep* 11, 1876-1884, doi:10.1016/j.celrep.2015.05.040.
- 194 He, E. *et al.* (2018). MIR137 schizophrenia-associated locus controls synaptic function by regulating synaptogenesis, synapse maturation and synaptic transmission. *Hum Mol Genet* 27, 1879-1891, doi:10.1093/hmg/ddy089.
- 195 Kwon, E., Wang, W. & Tsai, L. H. (2013). Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Mol Psychiatry* 18, 11-12, doi:10.1038/mp.2011.170.
- 196 Devanna, P. & Vernes, S. C. (2014). A direct molecular link between the autism candidate gene RORA and the schizophrenia candidate MIR137. *Sci Rep* 4, 3994, doi:10.1038/srep03994.
- 197 Radhakrishnan, B. & Alwin Prem Anand, A. (2016). Role of miRNA-9 in Brain Development. *J Exp Neurosci*

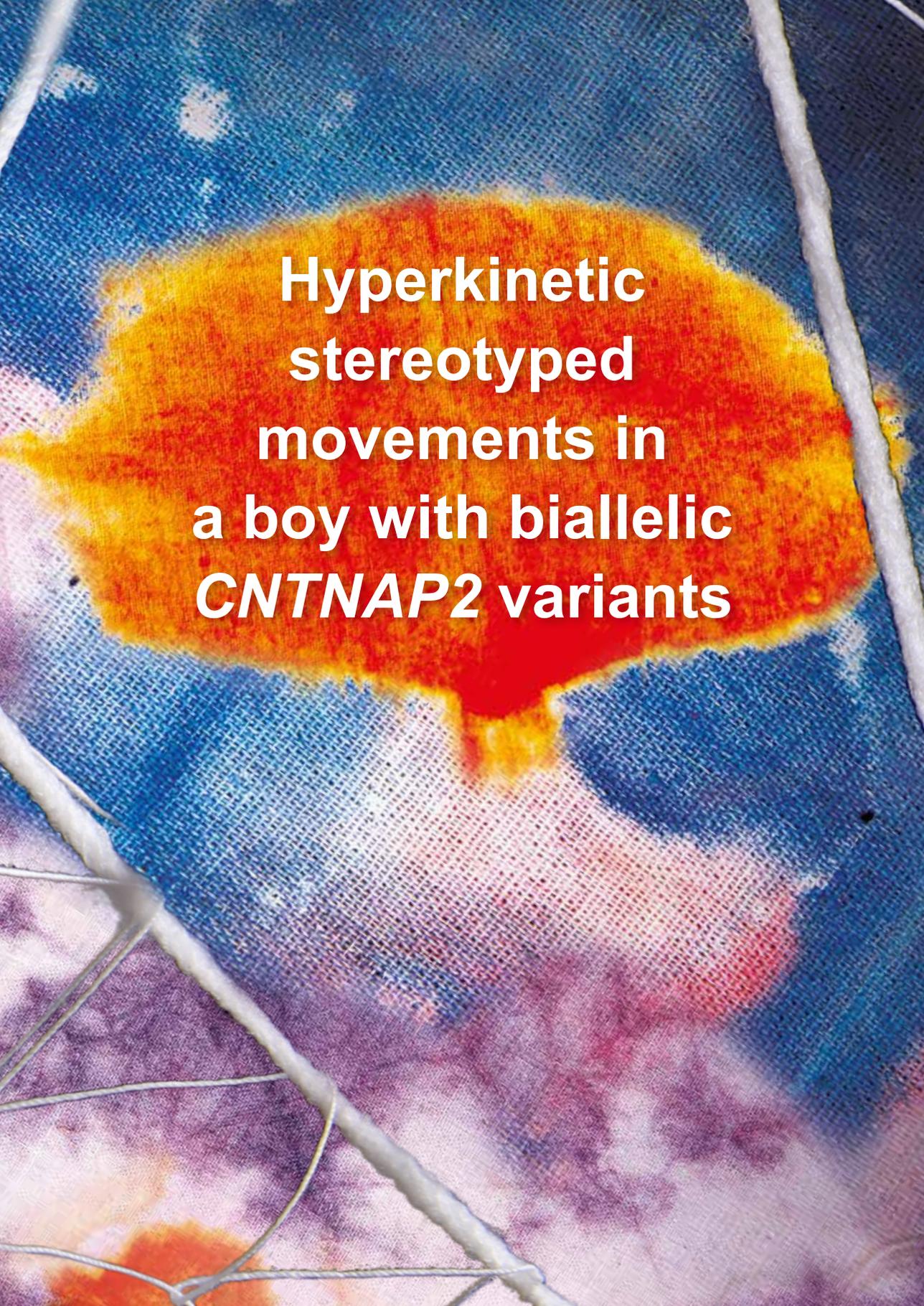
- 10, 101-120, doi:10.4137/JEN.S32843.
- 198 Clovis, Y. M., Enard, W., Marinaro, F., Huttner, W. B. & De Pietri Tonelli, D. (2012). Convergent repression of Foxp2 3'UTR by miR-9 and miR-132 in embryonic mouse neocortex: implications for radial migration of neurons. *Development* 139, 3332-3342, doi:10.1242/dev.078063.
- 199 Shi, Z. *et al.* (2013). miR-9 and miR-140-5p target FoxP2 and are regulated as a function of the social context of singing behavior in zebra finches. *J Neurosci* 33, 16510-16521, doi:10.1523/JNEUROSCI.0838-13.2013.
- 200 Fu, L. *et al.* (2014). Multiple microRNAs regulate human FOXP2 gene expression by targeting sequences in its 3' untranslated region. *Mol Brain* 7, 71, doi:10.1186/s13041-014-0071-0.
- 201 Devanna, P., Middelbeek, J. & Vernes, S. C. (2014). FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Front Cell Neurosci* 8, 305, doi:10.3389/fncel.2014.00305.
- 202 Groszer, M. *et al.* (2008). Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr Biol* 18, 354-362, doi:10.1016/j.cub.2008.01.060.
- 203 Haesler, S. *et al.* (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol* 5, e321, doi:10.1371/journal.pbio.0050321.
- 204 Thomas, A. M., Schwartz, M. D., Saxe, M. D. & Kilduff, T. S. (2017). Cntnap2 Knockout Rats and Mice Exhibit Epileptiform Activity and Abnormal Sleep-Wake Physiology. *Sleep* 40, zsw026-zsw026, doi:10.1093/sleep/zsw026.
- 205 Chimpanzee, S. & Analysis, C. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69-87, doi:10.1038/nature04072.
- 206 Suntsova, M. V. & Buzdin, A. A. (2020). Differences between human and chimpanzee genomes and their implications in gene expression, protein functions and biochemical properties of the two species. *BMC Genomics* 21, 535, doi:10.1186/s12864-020-06962-8.
- 207 Kronenberg, Z. N. *et al.* (2018). High-resolution comparative analysis of great ape genomes. *Science* 360, doi:10.1126/science.aar6343.
- 208 Pollen, A. A., Kilik, U., Lowe, C. B. & Camp, J. G. (2023). Human-specific genetics: new tools to explore the molecular and cellular basis of human evolution. *Nat Rev Genet*, 1-25, doi:10.1038/s41576-022-00568-4.
- 209 King, M. C. & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* 188, 107-116, doi:10.1126/science.1090005.
- 210 Prabhakar, S., Noonan, J. P., Paabo, S. & Rubin, E. M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science* 314, 786, doi:10.1126/science.1130738.
- 211 Pollard, K. S. *et al.* (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2, e168, doi:10.1371/journal.pgen.0020168.
- 212 Arbiza, L. *et al.* (2013). Genome-wide inference of natural selection on human transcription factor binding sites. *Nat Genet* 45, 723-729, doi:10.1038/ng.2658.
- 213 Zhang, X., Fang, B. & Huang, Y. F. (2023). Transcription factor binding sites are frequently under accelerated evolution in primates. *Nat Commun* 14, 783, doi:10.1038/s41467-023-36421-3.
- 214 Ayub, Q. *et al.* (2013). FOXP2 targets show evidence of positive selection in European populations. *Am J Hum Genet* 92, 696-706, doi:10.1016/j.ajhg.2013.03.019.
- 215 Gu, J. & Gu, X. (2003). Induced gene expression in human brain after the split from chimpanzee. *Trends Genet* 19, 63-65, doi:10.1016/s0168-9525(02)00040-9.
- 216 Hsieh, W. P., Chu, T. M., Wolfinger, R. D. & Gibson, G. (2003). Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* 165, 747-757, doi:10.1093/genetics/165.2.747.
- 217 Caceres, M. *et al.* (2003). Elevated gene expression levels distinguish human from non-human primate brains. *Proc Natl Acad Sci U S A* 100, 13030-13035, doi:10.1073/pnas.2135499100.
- 218 Khaitovich, P. *et al.* (2004). Regional patterns of gene expression in human and chimpanzee brains. *Genome Res* 14, 1462-1473, doi:10.1101/gr.2538704.
- 219 Somel, M. *et al.* (2009). Transcriptional neoteny in the human brain. *Proc Natl Acad Sci U S A* 106, 5743-5748, doi:10.1073/pnas.0900544106.
- 220 Liu, X. *et al.* (2012). Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res* 22, 611-622, doi:10.1101/gr.127324.111.
- 221 Sousa, A. M. M. *et al.* (2017). Molecular and cellular reorganization of neural circuits in the human lineage. *Science* 358, 1027-1032, doi:10.1126/science.aan3456.
- 222 Xu, C. *et al.* (2018). Human-specific features of spatial gene expression and regulation in eight brain regions. *Genome Res* 28, 1097-1110, doi:10.1101/gr.231357.117.
- 223 Khrameeva, E. *et al.* (2020). Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res* 30, 776-789, doi:10.1101/gr.256958.119.
- 224 Khaitovich, P., Enard, W., Lachmann, M. & Paabo, S. (2006). Evolution of primate gene expression. *Nat Rev*

- Genet* 7, 693-702, doi:10.1038/nrg1940.
- 225 Zhu, Y. *et al.* (2018). Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 362, doi:10.1126/science.aat8077.
- 226 Ma, S. *et al.* (2022). Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* 377, eabo7257, doi:10.1126/science.abo7257.
- 227 Mora-Bermudez, F. *et al.* (2016). Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *Elife* 5, e18683, doi:10.7554/eLife.18683.
- 228 Pollen, A. A. *et al.* (2019). Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* 176, 743-756 e717, doi:10.1016/j.cell.2019.01.017.
- 229 Kanton, S. *et al.* (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* 574, 418-422, doi:10.1038/s41586-019-1654-9.
- 230 Geschwind, D. H. & Rakic, P. (2013). Cortical evolution: judge the brain by its cover. *Neuron* 80, 633-647, doi:10.1016/j.neuron.2013.10.045.
- 231 Florio, M. & Huttner, W. B. (2014). Neural progenitors, neurogenesis and the evolution of the neocortex. *Development* 141, 2182-2194, doi:10.1242/dev.090571.
- 232 LaMonica, B. E., Lui, J. H., Hansen, D. V. & Kriegstein, A. R. (2013). Mitotic spindle orientation predicts outer radial glial cell generation in human neocortex. *Nat Commun* 4, 1665, doi:10.1038/ncomms2647.
- 233 Otani, T., Marchetto, M. C., Gage, F. H., Simons, B. D. & Livesey, F. J. (2016). 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. *Cell Stem Cell* 18, 467-480, doi:10.1016/j.stem.2016.03.003.
- 234 Fish, J. L., Kosodo, Y., Enard, W., Paabo, S. & Huttner, W. B. (2006). Aspm specifically maintains symmetric proliferative divisions of neuroepithelial cells. *Proc Natl Acad Sci U S A* 103, 10438-10443, doi:10.1073/pnas.0604066103.
- 235 Gruber, R. *et al.* (2011). MCPH1 regulates the neuroprogenitor division mode by coupling the centrosomal cycle with mitotic entry through the Chk1-Cdc25 pathway. *Nat Cell Biol* 13, 1325-1334, doi:10.1038/ncb2342.
- 236 Zhang, J. (2003). Evolution of the Human *ASPM* Gene, a Major Determinant of Brain Size. *Genetics* 165, 2063.
- 237 Evans, P. D. *et al.* (2005). *Microcephalin*, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans. *Science* 309, 1717.
- 238 Mekel-Bobrov, N. *et al.* (2005). Ongoing Adaptive Evolution of *ASPM*, a Brain Size Determinant in *Homo sapiens*. *Science* 309, 1720.
- 239 Jackson, A. P. *et al.* (2002). Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am J Hum Genet* 71, 136-142, doi:10.1086/341283.
- 240 Bond, J. *et al.* (2002). *ASPM* is a major determinant of cerebral cortical size. *Nat Genet* 32, 316-320.
- 241 Johnson, M. B. *et al.* (2015). Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci* 18, 637-646, doi:10.1038/nn.3980.
- 242 Florio, M. *et al.* (2015). Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. *Science* 347, 1465-1470, doi:10.1126/science.aaa1975.
- 243 Xing, L. *et al.* (2021). Expression of human-specific *ARHGAP11B* in mice leads to neocortex expansion and increased memory flexibility. *Embo j* 40, e107093, doi:10.15252/emboj.2020107093.
- 244 Fischer, J. *et al.* (2022). Human-specific *ARHGAP11B* ensures human-like basal progenitor levels in hominid cerebral organoids. *EMBO Rep* 23, e54728, doi:10.15252/embr.202254728.
- 245 Mertens, J., Marchetto, M. C., Bardy, C. & Gage, F. H. (2016). Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. *Nat Rev Neurosci* 17, 424-437, doi:10.1038/nrn.2016.46.
- 246 Zhang, Y. *et al.* (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78, 785-798, doi:10.1016/j.neuron.2013.05.029.
- 247 Clarke, L. E. & Barres, B. A. (2013). Emerging roles of astrocytes in neural circuit development. *Nat Rev Neurosci* 14, 311-321, doi:10.1038/nrn3484.
- 248 Brennand, K. J. *et al.* (2011). Modelling schizophrenia using human induced pluripotent stem cells. *Nature* 473, 221-225, doi:10.1038/nature09915.
- 249 Shi, Y., Kirwan, P. & Livesey, F. J. (2012). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat Protoc* 7, 1836-1846, doi:10.1038/nprot.2012.116.
- 250 Gunhanlar, N. *et al.* (2018). A simplified protocol for differentiation of electrophysiologically mature neuronal networks from human induced pluripotent stem cells. *Mol Psychiatry* 23, 1336-1344, doi:10.1038/mp.2017.56.
- 251 Lancaster, M. A. *et al.* (2013). Cerebral organoids model human brain development and microcephaly. *Nature* 501, 373-379, doi:10.1038/nature12517.
- 252 Mariani, J. *et al.* (2015). *FOXG1*-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in

- Autism Spectrum Disorders. *Cell* 162, 375-390, doi:10.1016/j.cell.2015.06.034.
- 253 Ichida, J. K. & Kiskinis, E. (2015). Probing disorders of the nervous system using reprogramming approaches. *Embo j* 34, 1456-1477, doi:10.15252/embj.201591267.
- 254 Srikanth, P. & Young-Pearse, T. L. (2014). Stem cells on the brain: modeling neurodevelopmental and neurodegenerative diseases using human induced pluripotent stem cells. *J Neurogenet* 28, 5-29, doi:10.3109/01677063.2014.881358.
- 255 Kelava, I. & Lancaster, M. A. (2016). Stem Cell Models of Human Brain Development. *Cell Stem Cell* 18, 736-748, doi:10.1016/j.stem.2016.05.022.
- 256 Miskinyte, G. *et al.* (2017). Direct conversion of human fibroblasts to functional excitatory cortical neurons integrating into human neural networks. *Stem Cell Res Ther* 8, 207, doi:10.1186/s13287-017-0658-3.
- 257 Sun, A. X. *et al.* (2016). Direct Induction and Functional Maturation of Forebrain GABAergic Neurons from Human Pluripotent Stem Cells. *Cell Rep* 16, 1942-1953, doi:10.1016/j.celrep.2016.07.035.
- 258 Kadoshima, T. *et al.* (2013). Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc Natl Acad Sci U S A* 110, 20284-20289, doi:10.1073/pnas.1315710110.
- 259 Jo, J. *et al.* (2016). Midbrain-like Organoids from Human Pluripotent Stem Cells Contain Functional Dopaminergic and Neuromelanin-Producing Neurons. *Cell Stem Cell* 19, 248-257, doi:10.1016/j.stem.2016.07.005.
- 260 Muguruma, K., Nishiyama, A., Kawakami, H., Hashimoto, K. & Sasai, Y. (2015). Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Rep* 10, 537-550, doi:10.1016/j.celrep.2014.12.051.
- 261 Puri, M. C. & Nagy, A. (2012). Concise review: Embryonic stem cells versus induced pluripotent stem cells: the game is on. *Stem Cells* 30, 10-14, doi:10.1002/stem.788.
- 262 Ardhanareeswaran, K., Mariani, J., Coppola, G., Abyzov, A. & Vaccarino, F. M. (2017). Human induced pluripotent stem cells for modelling neurodevelopmental disorders. *Nat Rev Neurol* 13, 265-278, doi:10.1038/nrneurol.2017.45.
- 263 Merkle, F. T. *et al.* (2022). Whole-genome analysis of human embryonic stem cells enables rational line selection based on genetic variation. *Cell Stem Cell* 29, 472-486 e477, doi:10.1016/j.stem.2022.01.011.
- 264 Takahashi, K. *et al.* (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861-872, doi:10.1016/j.cell.2007.11.019.
- 265 Bar, S. & Benvenisty, N. (2019). Epigenetic aberrations in human pluripotent stem cells. *Embo j* 38, doi:10.15252/embj.2018101033.
- 266 Abyzov, A. *et al.* (2012). Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* 492, 438-442, doi:10.1038/nature11629.
- 267 Mertens, J. *et al.* (2015). Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell* 17, 705-718, doi:10.1016/j.stem.2015.09.001.
- 268 Huh, C. J. *et al.* (2016). Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts. *Elife* 5, doi:10.7554/eLife.18648.
- 269 Choi, J. *et al.* (2015). A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. *Nat Biotechnol* 33, 1173-1181, doi:10.1038/nbt.3388.
- 270 Bassett, A. R. (2017). Editing the genome of hiPSC with CRISPR/Cas9: disease models. *Mamm Genome* 28, 348-364, doi:10.1007/s00335-017-9684-9.
- 271 Gordon, A. & Geschwind, D. H. (2020). Human in vitro models for understanding mechanisms of autism spectrum disorder. *Mol Autism* 11, 26, doi:10.1186/s13229-020-00332-7.

2



The background is a textured fabric, possibly canvas, with a vibrant, abstract painting. The central focus is a large, irregular shape in shades of orange, yellow, and red, resembling a flame or a sun. This shape is set against a background of blue and purple hues. The overall composition is dynamic and colorful.

**Hyperkinetic
stereotyped
movements in
a boy with biallelic
CNTNAP2 variants**

CHAPTER 2

This chapter has been adapted from:

Scala, M.* , Anijs, M.*, Battini, R., Madia, F., Capra, V., Scudieri, P., Verrotti, A., Zara, F., Minetti, C., Vernes, SC., and Striano, P. (2021). Hyperkinetic stereotyped movements in a boy with biallelic CNTNAP2 variants. *Ital J Pediatr.* 12;47(1):208, doi: 10.1186/s13052-021-01162-w

* These authors have contributed equally

BACKGROUND Heterozygous variants in CNTNAP2 have been implicated in a wide range of neurological phenotypes, including intellectual disability (ID), epilepsy, autistic spectrum disorder (ASD), and impaired language. However, heterozygous variants can also be found in unaffected individuals. Biallelic CNTNAP2 variants are rarer and cause a well-defined genetic syndrome known as CASPR2 deficiency disorder, a condition characterised by ID, early-onset refractory epilepsy, language impairment, and autistic features.

RESULTS A 7-year-old boy presented with hyperkinetic stereotyped movements that started during early infancy and persisted over childhood. Abnormal movements consisted of rhythmic and repetitive shaking of the four limbs, with evident stereotypic features. Additional clinical features included ID, attention deficit-hyperactivity disorder (ADHD), ASD, and speech impairment, consistent with CASPR2 deficiency disorder. Whole-genome array comparative genomic hybridization detected a maternally inherited 0.402 Mb duplication, which involved intron 1, exon 2, and intron 2 of CNTNAP2 (c.97+?_209-?dup). The affected region in intron 1 contains a binding site for the transcription factor FOXP2, potentially leading to abnormal CNTNAP2 expression regulation. Sanger sequencing of the coding region of CNTNAP2 also identified a paternally-inherited missense variant c.2752C>T, p.(Leu918Phe).

CONCLUSIONS This case expands the molecular and phenotypic spectrum of CASPR2 deficiency disorder, suggesting that Hyperkinetic stereotyped movements may be a rare, yet significant, clinical feature of this complex neurological disorder. Furthermore, the identification of an in-frame, largely non-coding duplication in CNTNAP2 points to a sophisticated underlying molecular mechanism, likely involving impaired FOXP2 binding.

Hyperkinetic stereotyped movements in a boy with biallelic *CNTNAP2* variants

2.1 Introduction

CNTNAP2 is located on chromosome 7q35, consists of 24 exons, and spans 2.3 Mb, making it a large physical target for disruptive variants. Most of the reported affected individuals carry heterozygous disruptions of *CNTNAP2* and usually display some combination of core phenotypes, including intellectual disability (ID), seizures, autistic features, and impaired language¹. Of note, heterozygous variants in *CNTNAP2* predicted to be deleterious have also been identified in unaffected subjects^{1,2}. Biallelic *CNTNAP2* variants are rare and have only been observed in patients with severe ID, early-onset drug-resistant epilepsy, limited or absent language, and autistic phenotypes³⁻⁷. These symptoms are collectively described as CASPR2 deficiency disorder³.

CASPR2, the protein product of *CNTNAP2*, is a transmembrane cell adhesion molecule from the neurexin family that is widely expressed throughout the brain^{1,8}. CASPR2 localises to juxtaparanodes of myelinated axons, where it is involved in neuron-glia interactions, and mediates the clustering of potassium channels via interaction with contactin-2 (also known as TAG-1)^{9,10}. CASPR2 is also localised to the synapse where it is involved in several additional processes, such as neuronal migration, neurite development and synapse maturation, stability, and function¹¹⁻¹⁵.

We report a patient that carries compound heterozygous variants in *CNTNAP2* including a missense variant and an intragenic duplication that were inherited from the father and mother, respectively. In addition to the common features found in CASPR2 deficiency disorder (ID, ADHD, ASD, and speech impairment), the boy presented with peculiar hyperkinetic stereotyped movements, expanding the molecular and phenotypic spectrum of CDD.

2.2 Methods

2.2.1 Array-CGH and validation of relative copy numbers in introns and exons

Whole genome array-CGH (aCGH) was performed on genomic DNA extracted from peripheral blood using the Agilent 180K platform, with an average resolution of approx. 40kb under optimal conditions. A variation in the number of copies was defined by a shift from the normal value of at least three consecutive probes for deletion and five for amplifications. The data was analysed using the Genomic Workbench software (Agilent) with human genomic DNA of the same sex as control (Agilent). Deletions and/or amplifications reported in the Database of Genomic Variants (<http://projects.tcag.ca/variation/>) were excluded (list available on request).

In order to test involvement of coding sequences in the maternal duplication, relative copy numbers were determined by real-time quantitative PCR (RT-qPCR) using iQ SYBR Green SuperMix (BioRad). Primers were designed for intron 1 (positive control), exon 2, intron 2 (95bp downstream of exon 2) and intron 2 (60kb downstream of exon 2, reference).

The efficiency of primers was tested via five-fold serial dilution of human genomic DNA (EMD Millipore). Melting curve analysis was performed to assess the specificity of the amplification. Data analysis was performed using CFX manager software (BioRad). Differential copy numbers were determined by transforming Cq values to fold gene expression while taking into account the primer efficiency and subsequently normalizing for the reference. The sequences, amplicon sizes and efficiencies for the primers used are listed in the table below:

Target	Forward primer	Reverse primer	Amplicon size (bp)	Primer efficiency
CNTNAP2 intron 1	CTGTTTCTCTGCATGCTGAC	GTATCCATACGGTAATGACACG	210	88%
CNTNAP2 exon 2	CAGAATTGCCTAAATTCCTTTC	CAGAGACAAGTGGCTCATCAC	164	75%
CNTNAP2 intron 2	GCAGACACCAGAAATCACTC	GACGTACTATCTCTAACTTCC	125	88%
CNTNAP2 intron 2 (ref)	CCCACATAGAATGGACACGTAGAA	CCTCCCCTGTGTATTGCG	100	98%

2.2.2 Sanger sequencing

All exons of *CNTNAP2* were screened using PCR and Sanger sequencing. PCR amplification was performed using Taq DNA polymerase kit (Invitrogen) as per supplier's instructions. PCR products were purified from gel with the Wizard SV gel clean-up kit (Promega) and Sanger sequenced to identify *CNTNAP2* mutations. The following primers were used for amplification and sequencing:

Target	Forward primer	Reverse primer	Amplicon size (bp)
CNTNAP2 exon 1	CAGCCCATCTCCCTCAAGA	GCTGGGTTTCGAGTTTGCT	353
CNTNAP2 exon 2	GGAGTCCTCTTTGTCTTCTCC	GAGTGATTCTGGTCTGCC	630
CNTNAP2 exon 3	GCACTGCCAAGACCAATTAAGA	TCTTTGCTTCTGCCAATGA	322
CNTNAP2 exon 4	TCACAAGCCCTACCATTGGA	ACATGCAAATAGAGACACAACCT	442
CNTNAP2 exon 5	AGAGGACTGTCAATTTCTCAAGA	ACGGAAGAATAGAAGTACAGT	362
CNTNAP2 exon 6	TCCCAGGTTAACTCGAATGGA	GCCTGGATAGCATGGTTCCT	573
CNTNAP2 exon 7	TGCCATAGATTTGGAGGCA	TGCGGGTGAAAATCCTTACC	307
CNTNAP2 exon 8	AGGCTGTGCTTCAAACCTTGT	ACCTAATCCTGAGCGTGAACA	474

Target	Forward primer	Reverse primer	Amplicon size (bp)
CNTNAP2 exon 9	GGAAATTGTGTTTCAGCTGGGT	TGCTCAGTGGGAATTACAGTTTGT	352
CNTNAP2 exon 10	TGAGTGACAGTAGACCCAG	TTGGACAACGTGAGGTAGGA	457
CNTNAP2 exon 11	CGCTTGGCACCTTCTTTCAT	TGTGGATTATTACAGTTGCTGA	377
CNTNAP2 exon 12	TCTGGGGAGCCATTTGTTCT	TGGGCTGAACCTTGTCACTT	467
CNTNAP2 exon 13	CTGTTCTACACCAGCTCAGTAA	CCCAGAAAACAAGCCCAATGAT	478
CNTNAP2 exon 14	GGGTGTAAGTGTGGCAGTCT	TCTAAATGCATACCCTTGTCGC	436
CNTNAP2 exon 15	TGTCTAATGCAGCCTCCTCA	AATCTCGGCTCCTGTACTGG	365
CNTNAP2 exon 16	CATGACTAGGCTGATCAGGGT	TGCTTCCCTGAGAGCATCC	426
CNTNAP2 exon 17	CCATTGATTTTGCCATCGACC	GGCCAACACCTTTACTTTTGG	494
CNTNAP2 exon 18	GCTATGCAGTGCATCTCCT	TCCACCTTACCTTTGTTGCA	348
CNTNAP2 exon 19	GGTATCGGCATCAGACCTCT	GTCTGCCCAAATGTAAATGGC	339
CNTNAP2 exon 20	AGCAGGAATTGAGGGGATGT	CCCTCAAACAAAACCAATGGC	301
CNTNAP2 exon 21	ACAGGGTAGAGACGTGCTTC	TGCCCAGCCTAATCACAATG	398
CNTNAP2 exon 22	ACAAGCATTCAAAGACAGGTATG	TCCATTCCATAGTCCCAAGAGA	393
CNTNAP2 exon 23	CTCGTCTGTCTGTGGAAGTAGA	TCCATAGTTGAGTAGCCCCA	423
CNTNAP2 exon 24	GTGTCTGACGGAGCTGTAGT	TCCTCCCTATCCCATAGCCA	317

2.3 Results

2.3.1 Case presentation

This case was a 7-year-old boy without a family history of neurodevelopmental disability, born at term to nonconsanguineous healthy parents (Fig. 1A) following a twin pregnancy complicated by intrauterine growth restriction and preeclampsia. His dizygotic twin brother was healthy at birth but was diagnosed with absence epilepsy during infancy. The neonatal course was characterised by feeding difficulties leading to failure to thrive. At 3 months of age, recurrent episodes of crying associated with semi-continuous, repetitive jerky movements of upper and lower limbs were observed, which were diagnosed as hyperkinetic stereotyped movements (Supplementary Video 1).

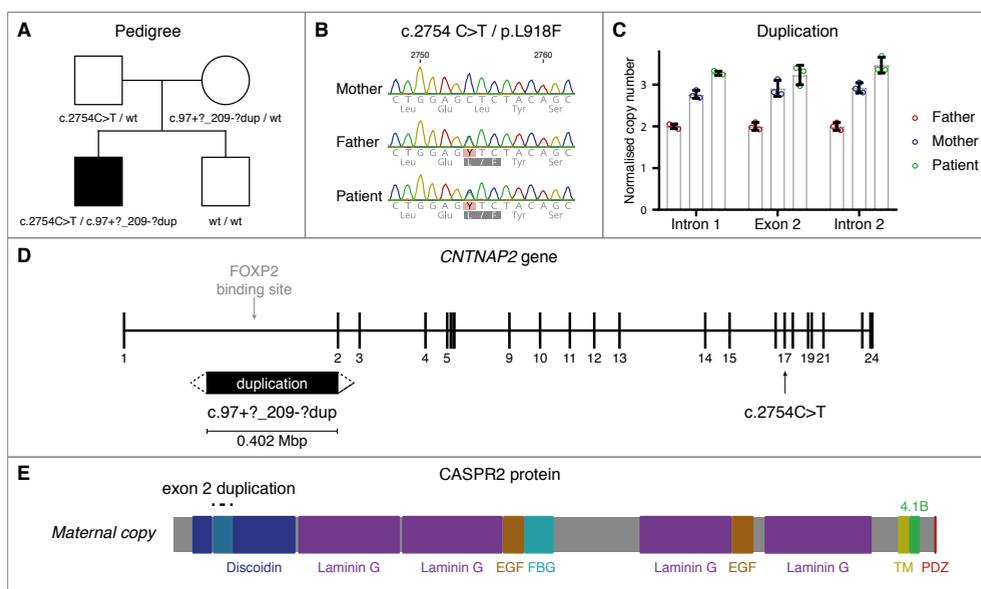


Figure 1 Genetic findings in the reported patient. (A) Pedigree of the family showing the affected patient carrying compound heterozygous variants in *CNTNAP2*: the paternally inherited missense c.2752C>T, p.(Leu918Phe) (ENST00000361727.3) in exon 17 and the maternally inherited c.97+?_209-?dup duplication in intron 1 (for which the exact breakpoints have not been mapped). (B) Sanger sequencing traces showing the heterozygous single nucleotide variant in the proband and his father, consisting of leucine to phenylalanine substitution. (C) Confirmation of exon 2 duplication. RT-qPCR from genomic DNA shows that the patient and his mother have one extra copy of exon 2 as compared to the father. The duplication breakpoint is located after exon 2 since the first ~200bp of intron 2 are still duplicated. Three technical replicates per condition. (D) Localization of inherited mutations on *CNTNAP2*. The maternally inherited duplication begins in the first intron and overlaps a binding site for the transcription factor FOXP2. Approximate boundaries of the duplication could be determined from the array-CGH and RT-qPCR (see methods), indicating that the duplication also involves exon 2 and part of intron 2. The paternally inherited mutation is located in exon 17. (E) Predicted consequences of *CNTNAP2* variants at the protein level. Exon 2 encodes the first part of the discoidin domain, the sequence of which is duplicated by the maternally inherited c.97+?_209-?dup variant. The paternally inherited p.(Leu918Phe) variant results in an amino acid change in exon 17, which encodes the third laminin G domain of CASPR2. Both domains belong to the extracellular region of CASPR2 which facilitates protein-protein interactions. Discoidin = Discoidin homology domain, Laminin G = Laminin G domain, EGF = EGF-like domain, FBG = Fibrinogen-like region, TM = transmembrane domain, 4.1B = Protein 4.1 binding domain, PDZ = PDZ interaction domain.

These paroxysmal hyperkinetic-dyskinetic episodes recurred periodically without any trigger, lasted 10-12 hours, and mimicked infantile colics. Their frequency was temporarily reduced when anti-reflux formulas and ranitidine were administered for concomitant recurrent vomiting. Electroencephalogram and brain magnetic resonance imaging were unremarkable.

During his first years of life, the patient displayed ADHD and autistic traits. The boy also featured a phonological processing deficiency and clumsiness in gross and fine movements was also observed. His neuropsychological evaluation (WPPSI-III at age 6 years) revealed a global IQ of 71 (verbal, 86; performance, 68; processing, 55). The movement disorder persisted during the following years through non-triggered fast, high-amplitude, rhythmic, continuous and repetitive shaking involving the four limbs with stereotypic features (Supplementary Videos 2-3). His stereotypic movement disorder had been misdiagnosed as tics before, which indeed tend to appear from age 6-7 years¹⁶. However, these repetitive movements already started before the age of 3 and consisted of intense patterns of movement that ran longer than tics and were more bilateral than tics. Several medications, including carbamazepine, valproate, gabapentin, levodopa, flunarizine, benzodiazepines, did not improve his symptoms. The boy is currently receiving clonazepam, 0.2 mg/Kg in two daily doses.

2.3.2 Genetic findings

Written informed consent was obtained from the patient's parents. Whole-genome array-CGH on peripheral blood genomic DNA of the family quartet revealed a maternally inherited *CNTNAP2* variant c.97+?_209-?dup in the proband. Real-time quantitative PCR confirmed that this 0.402 Mbp duplication involved part of intron 1 and exon 2, with a breakpoint within intron 2 (Fig. 1C). This duplication is not predicted to introduce a frameshift in the CASPR2 protein. The duplication was not inherited by the sibling.

PCR amplification and Sanger sequencing was performed to screen *CNTNAP2* exons in the family, revealing a paternally inherited missense variant c.2752C>T, p.(Leu918Phe) in the proband (Fig. 1B). This variant is absent from the gnomAD dataset (<https://gnomad.broadinstitute.org/variant/>) and affects a conserved residue within the Laminin G-like (LG) 3 domain (GERP = 5.49). It is reported in ClinVar (allele ID 924741, <https://www.ncbi.nlm.nih.gov/clinvar/variation/949154/>) and predicted damaging by *in silico* tools (e.g., DANN, 0.9977; Mutation Taster, 1; CADD, 24.1). Both variants were absent in the proband's sibling. Further details are available in Supplementary Methods.

2.4 Discussion

The range of clinical symptoms found in patients harboring biallelic *CNTNAP2* variants are collectively described as CASPR2 deficiency disorder. Dyskinetic features are not included amongst the classic neurological manifestations of this condition^{3,4,6,7,19,20} (Table 1). In this patient, we observed hyperkinetic stereotyped movements consisting of continuous, repetitive and rhythmic shaking of the four limbs, which became evident during the first months of life and persisted over the years. These abnormal movements did not resemble generalised dystonia, paroxysmal non-kinesigenic dyskinesia, or paroxysmal kinesigenic dyskinesia, but rather represent a novel neurological manifestation of CASPR2 deficiency

Table 1 Physical, neurodevelopmental and behavioural features of patients identified with homozygous CNTNAP2 mutations resulting in CASPR2 deficiency disorder

	18	2	3	8	1	2	1	2	1
Patients	18	2	3	8	1	2	1	2	1
Mutation	Jackman, et al. (2009) ¹⁷ Strauss, et al. (2006) ⁵	Zweier, et al. (2009) ⁶ Orrico, et al. (2001) ¹⁸	Watson, et al. (2014) ⁷	Rodenas-Cuadrado, et al. (2016) ³	Smogavec, et al. (2016) ⁴	Riccardi, et al. (2019) ¹⁹	Mittal, et al. (2021) ²⁰	This study	
DNA variant	c.3709delG	CNV del ex 2-9	CNV del ex 3	CNV del ex 2-3	various	c.985delA	c.1977_1989del113	c.97+?_209-?dup (mat), c.2752C>T (pat)	
Effect	Early STOP codon	In frame deletion	Frameshift, early STOP codon	Frameshift, early STOP codon	various	Frameshift, early STOP codon	Frameshift, early STOP codon	In frame duplication (mat) and missense mutation (pat)	
Resulting protein	I1253X	del.33-500	A156X	L39X	various	S329Vfs*28	V660Ffs*9	dup.33-70 (mat), L918T (pat)	
Pregnancy	NR	Normal	Normal	Normal	Normal	Normal	Normal	Complicated twin pregnancy	
Intellectual disability	Severe (100%)	Severe	Severe	Severe	Moderate-Severe	Moderate	Severe	Moderate	
Psychomotor delay	+	Mild	+	+	+	+	+	+	
Gross motor delay	+	Mild	+	-	+	+	+	+	
Ataxic cerebral palsy	NR	NR	+	-	NR	NR	NR	-	
Dysmorphic	NR	+	Mild	+	-	Mild	+	-	
Low forehead/hairline	NR	+	NR	+	NR	-	+	-	

	Arched palate	NR	-	NR	+	NR	-	-	-	-	-	-
	Dental malocclusion	NR	-	NR	+	NR	-	-	-	-	-	-
	Obesity	NR	-	NR	+	NR	-	-	-	-	-	-
Behaviour	Autism	+	-	-	+	NR	-	+	+	+	+	+
	Autistic features	+	+	+	+	+	+	+	+	+	+	+
	Motor stereo-types	NR	+	+	+	+	+	+	+	-	+	+
	Communicative impairment	+	+	+	+	+	+	+	+	+	+	+
	Behavioural problems	+	+	+	+	+	+	+	+	+	+	+
Epilepsy	ADHD	+	NR	restlessness	-	NR	NR	NR	NR	NR	NR	+
	Seizures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Age onset (months)	13-20	22-30	24	20-36	14-36	14	14	36	36	3	3
Language	Language impairment/regression	+	+	+	+	+	+	+	+	+	+	+
	Limited/absent language	+	+	+	+	+	+	+	+	+	+	+
Neurological	Cortical dysplasia	+		NR	-	+	-	-	-	-	-	-
	Cerebellar hypoplasia	-	+/-	NR	+/-	+	-	-	-	-	-	-

NR = Not reported

disorder. This dyskinetic phenotype may be related to the role of CASPR2 in facilitating nerve conduction and synaptic connectivity, particularly given its widespread expression in the central and peripheral nervous system^{1,10,15}. In patients harboring biallelic variants in *CNTNAP2*, brain MRI may be normal, such as in our case, or show cortical dysplasia (cortical dysplasia-focal epilepsy syndrome, CDFES)^{3,4,6}. Additional findings include cerebellar abnormalities (vermian hypoplasia or atrophy) and nonspecific white matter abnormalities^{3,4}.

The c.97+?_209-?dup and p.(Leu918Phe) variants were inherited from unaffected parents, supporting the incomplete penetrance of heterozygous *CNTNAP2* variants^{1,2}. The maternally inherited variant (c.97+?_209-?dup) involved a complete duplication of exon 2 (99 bp) (Fig. 1D), which may negatively impact protein folding and/or protein-protein interactions. Indeed, exon 2 encodes part of the discoidin homology domain, which is found in the extracellular portion CASPR2 and is known to mediate protein-protein interactions, pointing to a potential perturbation of this mechanism in the patient (Fig. 1E). The paternally inherited missense variant p.(Leu918Phe) is also in the extracellular portion of CASPR2, but is found in the third LG domain (Fig. 1E). Missense variants in the extracellular portion of CASPR2 have been shown to impair interactions with contactin-2 and affect axon growth in cortical neurons¹³. This effect is likely related to the trans-synaptic bridge formed in neurons by the interaction of CASPR2 and contactin-2 that contributes to synaptic organization and synaptic transmission²¹. It would therefore be of value to determine if the variants identified in this patient have functional consequences for protein-protein interactions, or for synaptic or juxtapanodal organisation and signalling. Such functional studies may shed light on why heterozygous *CNTNAP2* variants show incomplete penetrance and why biallelic mutations together produce a range of phenotypes, including the novel motor phenotype described herein.

The maternal duplication also has the potential to alter the regulation of *CNTNAP2*. The transcription factor FOXP2 targets a binding site in intron 1 of *CNTNAP2* and regulates its expression. Mutations in *FOXP2* are a monogenic cause of childhood apraxia of speech (SPCH1, OMIM #602081)²² and *CNTNAP2* is functionally implicated in the aetiology of this condition as part of the downstream network of FOXP2 target genes²³. This is supported by a genetic association of *CNTNAP2* variants with phonological memory performance in children with specific language impairment²³. Duplication of the FOXP2 target site in intron 1 of the maternally inherited allele could alter FOXP2-mediated regulation of *CNTNAP2* and lead to aberrant CASPR2 levels, a hypothesis that needs functional testing in cell or animal models. In this way, it is possible that the intronic duplication could contribute to neurodevelopmental or speech phenotypes in patients that are related to both FOXP2 and *CNTNAP2*, such as phonological processing.

This study has two limitations that could be addressed in future research. First, a functional characterization of the two detected *CNTNAP2* variants will be necessary to test their impact on protein structure and gene function. For example, co-immunoprecipitation experiments could help assess the effects of these coding mutations on protein-protein interactions with contactin-2 and a luciferase assay could be used to test if duplication of the FOXP2 target site leads to stronger repression of *CNTNAP2* in cells. Second, the presence of additional variants with a potential modifier effect on the clinical phenotype could be investigated more thoroughly with Next Generation Sequencing (NGS) techniques, such

as whole exome sequencing (WES) or whole genome sequencing (WGS), as array-CGH is limited in its resolution.

This case of a boy with hyperkinetic stereotyped movements and biallelic *CNTNAP2* mutations expands our knowledge about *CNTNAP2*-related disorders. It presents a new, rare neurological manifestation for CDD and posits a remarkable molecular mechanism in which coding and non-coding *CNTNAP2* mutations could contribute to the observed phenotypes. Accordingly, we suggest screening *CNTNAP2* regulatory regions in patients with a CDD-suggestive phenotype even if a single heterozygous *CNTNAP2* variant has been identified or if atypical neurological phenotypes are also present. This will lead to better diagnoses that can improve the management of patients with these disorders. This, together with functional studies of the consequences of the identified mutations, will advance our scientific understanding of disease genes like *CNTNAP2*.

References

- Rodenas-Cuadrado, P., Ho, J. & Vernes, S. C. (2014). Shining a light on CNTNAP2: complex functions to complex disorders. *Eur J Hum Genet* 22, 171-178, doi:10.1038/ejhg.2013.100.
- Toma, C. *et al.* (2018). Comprehensive cross-disorder analyses of CNTNAP2 suggest it is unlikely to be a primary risk gene for psychiatric disorders. *PLOS Genetics* 14, e1007535, doi:10.1371/journal.pgen.1007535.
- Rodenas-Cuadrado, P. *et al.* (2016). Characterisation of CASPR2 deficiency disorder--a syndrome involving autism, epilepsy and language impairment. *BMC Med Genet* 17, 8, doi:10.1186/s12881-016-0272-8.
- Smogavec, M. *et al.* (2016). Eight further individuals with intellectual disability and epilepsy carrying biallelic CNTNAP2 aberrations allow delineation of the mutational and phenotypic spectrum. *J Med Genet* 53, 820-827, doi:10.1136/jmedgenet-2016-103880.
- Strauss, K. A. *et al.* (2006). Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med* 354, 1370-1377, doi:10.1056/NEJMoa052773.
- Zweier, C. *et al.* (2009). CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. *Am J Hum Genet* 85, 655-666, doi:10.1016/j.ajhg.2009.10.004.
- Watson, C. M. *et al.* (2014). Diagnostic whole genome sequencing and split-read mapping for nucleotide resolution breakpoint identification in CNTNAP2 deficiency syndrome. *Am J Med Genet A* 164A, 2649-2655, doi:10.1002/ajmg.a.36679.
- Gordon, A. *et al.* (2016). Expression of Cntnap2 (Caspr2) in multiple levels of sensory systems. *Mol Cell Neurosci* 70, 42-53, doi:10.1016/j.mcn.2015.11.012.
- Poliak, S. *et al.* (2001). Localization of Caspr2 in myelinated nerves depends on axon-glia interactions and the generation of barriers along the axon. *J Neurosci* 21, 7568-7575, doi:10.1523/JNEUROSCI.21-19-07568.2001.
- Poliak, S. *et al.* (2003). Juxtaparanodal clustering of Shaker-like K⁺ channels in myelinated axons depends on Caspr2 and TAG-1. *J Cell Biol* 162, 1149-1160, doi:10.1083/jcb.200305018.
- Penagarikano, O. *et al.* (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147, 235-246, doi:10.1016/j.cell.2011.08.040.
- Anderson, G. R. *et al.* (2012). Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc Natl Acad Sci U S A* 109, 18120-18125, doi:10.1073/pnas.1216398109.
- Canali, G. *et al.* (2018). Genetic variants in autism-related CNTNAP2 impair axonal growth of cortical neurons. *Hum Mol Genet* 27, 1941-1954, doi:10.1093/hmg/ddy102.
- Varea, O. *et al.* (2015). Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. *Proc Natl Acad Sci U S A* 112, 6176-6181, doi:10.1073/pnas.1423205112.
- Lazaro, M. T. *et al.* (2019). Reduced Prefrontal Synaptic Connectivity and Disturbed Oscillatory Population Dynamics in the CNTNAP2 Model of Autism. *Cell Rep* 27, 2567-2578 e2566, doi:10.1016/j.celrep.2019.05.006.
- Zinner, S. H. & Mink, J. W. (2010). Movement disorders I: tics and stereotypies. *Pediatr Rev* 31, 223-233, doi:10.1542/pir.31-6-223.
- Jackman, C., Horn, N. D., Molleston, J. P. & Sokol, D. K. (2009). Gene associated with seizures, autism, and hepatomegaly in an Amish girl. *Pediatr Neurol* 40, 310-313, doi:10.1016/j.pediatrneurol.2008.10.013.

- 18 Orrico, A. *et al.* (2001). Possible case of Pitt-Hopkins syndrome in sibs. *Am J Med Genet* 103, 157-159, doi:10.1002/ajmg.1523.
- 19 Riccardi, F., Urquhart, J., McCullagh, G., Lawrence, P. & Douzgou, S. (2019). A patient with a novel CNTNAP2 homozygous variant: further delineation of the CASPR2 deficiency syndrome and review of the literature. *Clin Dysmorphol* 28, 66-70, doi:10.1097/MCD.0000000000000259.
- 20 Mittal, R., Kumar, A., Ladda, R., Mainali, G. & Aliu, E. (2021). Pitt Hopkins-Like Syndrome 1 with Novel CNTNAP2 Mutation in Siblings. *Child Neurol Open* 8, 2329048X211055330, doi:10.1177/2329048X211055330.
- 21 Lu, Z. *et al.* (2016). Molecular Architecture of Contactin-associated Protein-like 2 (CNTNAP2) and Its Interaction with Contactin 2 (CNTN2). *J Biol Chem* 291, 24133-24147, doi:10.1074/jbc.M116.748236.
- 22 Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519-523, doi:10.1038/35097076.
- 23 Vernes, S. C. *et al.* (2008). A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359, 2337-2345, doi:10.1056/NEJMoa0802828.

Supplementary material

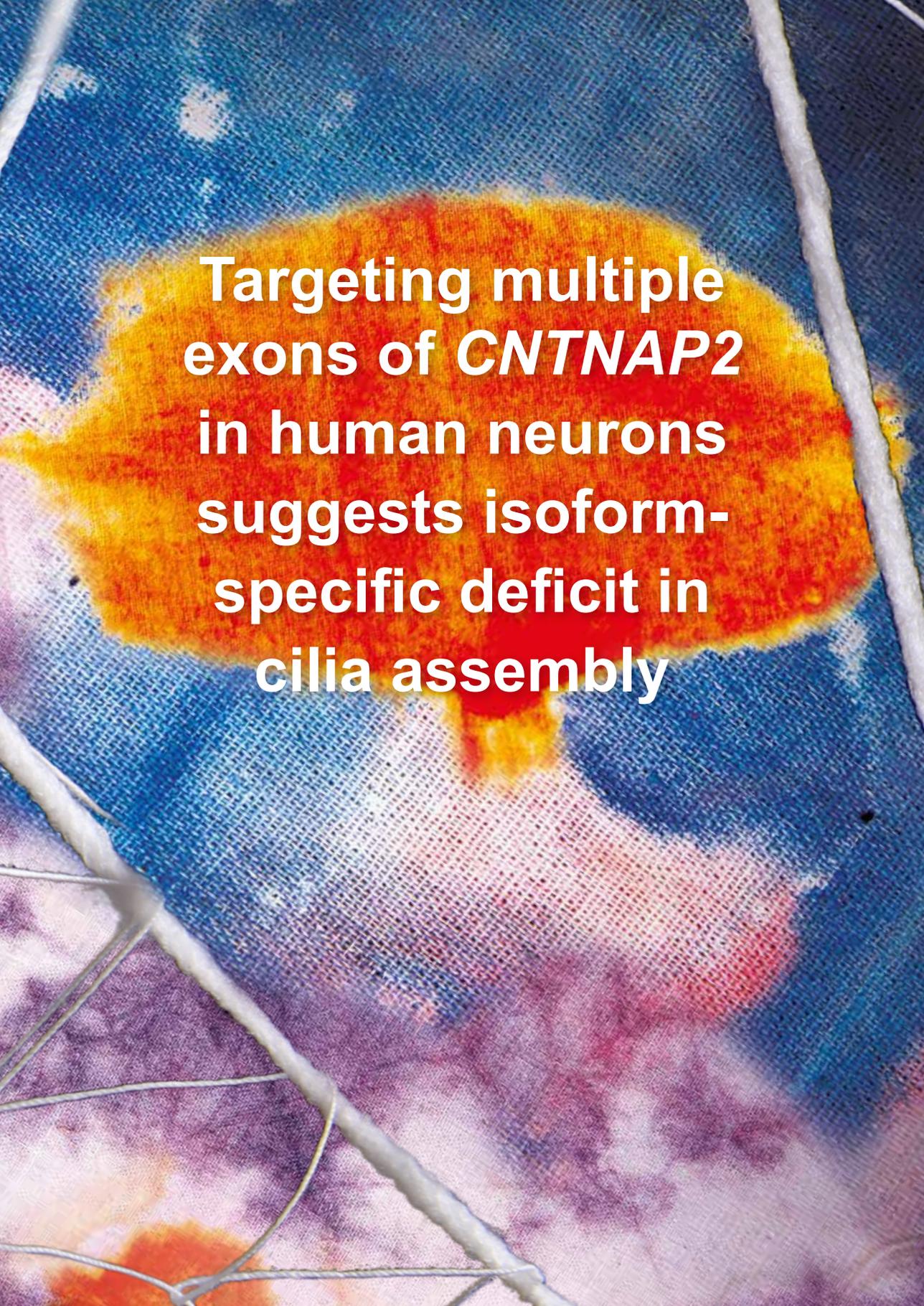
Supplementary video 1 (Available online at: <https://ijponline.biomedcentral.com/articles/10.1186/s13052-021-01162-w#Sec2>). This video shows the patient at 3 months of age. Videos were taken in a home environment while the child was playing with his parents. Hyperkinetic stereotyped movements can be observed at 11-16 sec., 19-26 sec, 30-38 sec.

Supplementary video 2 (Available online at: <https://ijponline.biomedcentral.com/articles/10.1186/s13052-021-01162-w#Sec2>). This video shows the patient at 6 years of age. Videos were taken in a home environment while the child was seated on mother's knees. Non-triggered fast, high-amplitude, rhythmic, continuous, and repetitive shaking involving the four limbs with stereotypic features can be observed during the whole video.

Supplementary video 1 (Available online at: <https://ijponline.biomedcentral.com/articles/10.1186/s13052-021-01162-w#Sec6>). This video shows the patient at 6 years of age. Videos were taken in a home environment while the child is eating, next to the unaffected sibling. Non-triggered fast, high-amplitude, rhythmic, continuous, and repetitive shaking involving the four limbs with stereotypic features can be observed at 13-20 sec., 30-42 sec., 50 sec.-1:06 min., 1:20-1:50 min.

3





**Targeting multiple
exons of *CNTNAP2*
in human neurons
suggests isoform-
specific deficit in
cilia assembly**

CHAPTER 3

Mutations in *CNTNAP2* are related to a broad range of neurodevelopmental disorders with different degrees of severity. Heterozygous variants in *CNTNAP2* have been involved in a wide range of neurological phenotypes, including intellectual disability, epilepsy, autistic spectrum disorder, schizophrenia and impaired language, but they are also found in unaffected individuals. Homozygous mutations are rarer, and cause a severe syndrome (CASPR2 deficiency disorder) characterized by intellectual disability, epileptic seizures, language impairments and autistic features. This study aims to understand the mechanistic underpinnings of the heterogeneity of *CNTNAP2*-related phenotypes by studying the impact of mutations in different *CNTNAP2* isoforms in a model for human neurodevelopment.

Here, three *CNTNAP2* mutant conditions in human neuronal network cultures were characterized: heterozygous and homozygous loss of the long isoform and homozygous loss of both isoforms. Gene expression, cell type composition and synapse counts were measured.

This study shows that homozygous loss of CASPR2 affects the differentiation of neural precursor cells, the number of excitatory synapses and gene expression for several biological processes, whereas heterozygous loss only produces a subset of the gene expression changes. This study also shows that the short isoform of *CNTNAP2* is expressed at high levels early in development and its expression drops substantially in later stages. Loss of expression of this short isoform specifically disrupts the expression of cilia assembly genes.

The limited phenotypes observed in the heterozygous CASPR2 condition are consistent with the lower risk of disorder in individuals carrying heterozygous *CNTNAP2* mutations. In contrast, mutations affecting the short isoform seem to be more disruptive than those affecting the rest of the gene. Further investigation of a role for CASPR2 in cilia function during early brain development could help to understand patient phenotypes relating to neuronal migration and cortical lamination and provide an explanation for clinical variation between cases with mutations that affect different isoforms.

Targeting multiple exons of *CNTNAP2* in human neurons suggests isoform-specific deficit in cilia assembly

3.1 Introduction

Homozygous mutations in the contactin-associated protein-like 2 gene (*CNTNAP2*) lead to a severe disorder, called CASPR2 deficiency disorder (CDD), that is characterized by intellectual disability, early-onset epilepsy, impaired language and autistic features¹. The broad potential relevance of mutations in *CNTNAP2* is further highlighted by links to several complex neurological disorders, such as childhood apraxia of speech, intellectual disability, autism spectrum disorder (ASD), epilepsy, and schizophrenia². *CNTNAP2* is not the only human gene in which genetic variation can contribute to risk for multiple neuropsychiatric phenotypes³. Such genetic overlap suggests shared pathophysiology between those disorders but could also indicate that different mutations in this gene may lead to divergent clinical outcomes. Given the clinical heterogeneity of patients with *CNTNAP2* mutations, it is important to functionally characterize different types of mutations in *CNTNAP2* to understand potential mechanisms underlying this heterogeneity.

CASPR2, the protein product of *CNTNAP2*, is a single-pass transmembrane cell adhesion protein that is widely expressed throughout the brain⁴. The large extracellular domain at the N-terminal side of CASPR2 binds an adhesion molecule contactin-2 at two cellular locations: it mediates neuron-glia interaction at the juxtaparanode of myelinated axons and forms a molecular bridge across the synaptic cleft in both inhibitory and excitatory neurons^{5,6}. The small C-terminal intracellular region of CASPR2 is required for the clustering of voltage-gated potassium channels that are important for conduction of action potentials⁷. At the plasma membrane, the intracellular region of CASPR2 plays an organizing role by binding cytoskeletal structural proteins, such as CASK and protein 4.1B^{8,9}. Other interactions with the intracellular region of CASPR2 can affect the subcellular localisation of CASPR2^{10,11}.

In mice, homozygous disruption of the long isoform of *Cntnap2* leads to social deficits, repetitive behaviours and seizures¹². These behavioral changes are accompanied by signs of dysfunction in the neurons of these mice, such as deficits in synapse development and function, decreased dendritic arborization, reduced inhibition, axonal excitability deficits, and altered network synchrony^{8,12-18}. Notably, these studies assessed consequences of *Cntnap2* loss in knockout mice that targeted the long isoform but still express a short isoform of *Caspr2*¹⁹. The short isoform contains the full intracellular and transmembrane sequence but lacks most of the extracellular domain¹⁹. The intracellular domains of *CNTNAP2* were strongly affected in a cohort of Amish children with cortical dysplasia-focal epilepsy (CDFE) syndrome that carry a homozygous frameshift mutation resulting in a premature stop codon in this part of *CNTNAP2*²⁰. It is possible that loss of the short isoform contributes to specific aspects of this particularly severe case of CDD. Alternative isoform expression is an important source of molecular diversity and functional complexity, especially in the central nervous system²¹. Therefore, understanding the isoform-specific functions of *CNTNAP2* may help to understand the pleiotropic effects of this gene.

In this study, we applied CRISPR/Cas9 gene editing to human neuronal network cultures to investigate the neurobiological consequences of heterozygous and homozygous loss of the long isoform and the potential additive effects of loss of the short isoform. So far, little is known about the function of *CNTNAP2* in human neurons. Recent studies in human cell models found changes in neural precursor cell proliferation, neuronal differentiation and neural circuit activity in patient-derived cells^{22,23}. This is the first study that looks at multiple types of *CNTNAP2* mutations in an isogenic human background, which allows a precise comparison of heterozygous, homozygous and isoform-specific mutations. The effects of these mutations were compared by measuring cell type composition, synapse numbers, and transcriptomics to determine how mutations in *CNTNAP2* can lead to multiple phenotypic outcomes.

3.2 Methods

3.2.1 Embryonic stem cells

H1-ESCs (WiCell, Madison WI, lot no.#WB16217) were cultured and maintained according to methods recommended by WiCell. Briefly, cells were cultured on matrigel-coated plates in mTeSR1 (Stem Cell Technologies, catalogue # 85850). Cells were passaged using versene (Life Technologies, catalogue # 15040066).

3.2.2 CRISPR-Cas9

Guide RNA (gRNA) sequences targeting exon 1 and exon 22 of *CNTNAP2* were selected using the Zhang laboratory CRISPR design tool and CRISPOR^{24,25}. These were ordered as oligonucleotides and cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene, catalogue # 62988). CRISPR editing was performed as described earlier²⁶. Briefly, hESCs were pretreated with mTeSR1 supplemented with 10 μ M Y-27632 (Sigma) for 30 minutes before dissociation with Accutase (Sigma, catalogue # SCR005). Cells were resuspended in 3 ml mTeSR1 and strained before counting. 2×10^5 cells per nucleofection were pelleted at 100 g for 3 minutes and carefully resuspended in 20 μ l supplemented P3 Primary Cell Nucleofector solution (Lonza, catalogue #V4XP-3012) with 2.1 μ g of plasmid DNA. The suspension was transferred into a cuvette and electoporated with program CB-150 in the Amaxa 4D Nucleofector (Lonza). 80 μ l mTeSR1 supplemented with 10 μ M Y-27632 was added to the cuvette before transferring its contents to the well of a matrigel-coated 24-wells plate. Cells were incubated at 37 °C in the presence of 5% CO₂ and 5% O₂. Medium was replaced after 24 hours for mTeSR1 without Y-27632 and with 0.5 μ g/ml puromycin. At 48 hours after nucleofection, puromycin was removed and cells were seeded at a clonal dilution of 25,000 cells in a 10 cm dish. Bulk DNA from remaining cells was harvested with the QuickExtract solution (Immunesource, catalogue # QE0905T) to estimate editing efficiency. After approximately one week, clones radiating from a single cell were manually picked into a matrigel-coated 96-wells plate under a dissection microscope. After clones were confluent, they were dissociated with accutase and split 1 in 2 into a new 96-wells plate. DNA from the other half was harvested with QuickExtract solution.

Editing efficiency was estimated on bulk gDNA using a T7E1 enzyme according to manufacturer's instructions (New England Biolabs, catalogue # M0302S). Briefly, 200 ng

of PCR product was annealed and subsequently potential mismatches caused by edited fragments were cut during a 15-minute incubation at 37 °C with 1 µl T7E1 in 20 µl reaction volume. The reaction was quenched by addition of 1.5 µl 0.25 M EDTA. The DNA products were tested on an agarose gel (2% w/v). Percentage gene modification was calculated by densitometry of the cleaved and uncleaved fragments, using the following formula: % gene modification = $100 \times (1 - (1 - \text{fraction cleaved})^{1/2})$.

CRISPR-Cas9 edited clones were screened for indels by PCR amplification and Sanger sequencing. Genome editing was quantified from the sequencing traces using the TIDE decomposition algorithm²⁷. The ICE tool was used for bulk processing and visualisation²⁸. Clones with frameshift-inducing indels were expanded and frozen for later use.

3.2.3 Whole genome sequencing

Genomic DNA was extracted from frozen pellets of human embryonic stem cells using the NucleoSpin RapidLyse Kit (Machery-Nagel). The quality of genomic DNA was determined on agarose gel (1% w/v) and the quantity was measured with a Qubit Fluorometer (Invitrogen). Whole genome sequencing was performed on the HiSeq-PE150 platform by Novogene (Hong Kong).

3.2.3.1 SNV and indel analysis

The Best Practices GATK Pipeline was implemented to call SNVs and indels in the four sequenced samples. The reads were aligned to the human reference genome GRCh37 (human_g1k_v37_decoy) with Burrows-Wheeler Algorithm (BWA version 0.7.17-r1188). Duplicate reads were marked with Picard (version 1.134). The consequent steps – sorting, reordering of the reads, realigning around indels, base quality recalibration and variant calling were performed with Genome Analysis Toolkit (GATK version 3.7).

Following standard practices in the field²⁹, sequential filtering steps were performed (Table 1) to obtain the set of high quality variant calls for each sample. Only the calls that passed variant quality score recalibration were taken forward (VRSQ>99%). Next, all variant calls in sites with excessive read depth ($DP > DP + 4 \times \sqrt{DP}$) as in³⁰ and in low complexity regions were removed. Calls with low read depth (<8) and poor genotype quality (<20) were removed as well. Finally, calls were limited to the coding sequences of genes expressed with at least 10 reads in one of the four stages of wild-type model development that was assayed with RNA-seq. Further, all calls that were shared between all samples and then calls that were found in the reference (wild-type) and parental sample were removed.

High quality filtered variant calls were annotated with snpEff, using database GRCh37.75, dbSNP build 147 (<http://www.ncbi.nlm.nih.gov/SNP/>), the Exome Variant Server (<http://evs.gs.washington.edu/EVS>), and the 1000 Genomes Project³¹. Subsets of variants with high impact were extracted using SnpSift to be examined more thoroughly in all pairwise sample intersections. Potential unanticipated target sites of editing by the CRISPR-Cas9 system were predicted by two different methods^{24,25} and pooled into a list of intervals for each round of editing. All variants within 300bp of these sites were extracted with GATK SelectVariants and manually evaluated for each sample and pairwise sample

intersections.

Table 1 Statistics of variant filtering (counts of variants remaining after each step)

Filters applied	WT	HET	HOM	DUAL
FAIL	4,242,787	4,238,663	4,232,605	4,238,663
DP>DP+4*sqrt(DP)	4,169,572	4,144,256	4,160,687	4,144,256
Low complexity	3,513,064	3,492,238	3,512,682	3,543,806
Low depth (<8)	3,505,913	3,485,396	3,494,672	3,532,173
GQ <20	3,502,545	3,482,201	3,389,468	3,528,660
Common to all	191,919	171,609	178,834	217,967
Coding + expressed	14,194	12,042	13,167	27,411
<i>De novo</i>		5,833	6,073	15,860

^a For the WT sample, this step removed all the calls that were removed collectively from the three other samples, leaving calls exclusive to WT.

3.2.3.2 Structural variants analysis

SVs were called with FREEC software³², as well as two ensemble callers – Smoove version 0.2.5 (<https://github.com/brentp/smoove>) and sv-callers version 1.1.1 with callers: dely, manta, gridss with default settings from each tool³³. Only SV calls on autosomes and chromosomes X, Y, and M that passed each respective tool quality cut-off were taken forward for analysis. The calls from FREEC that overlapped or were in proximity of each other and had the same type were merged using bedtools merge and converted to vcf with SURVIVOR version 1.0.7. All edited samples SVs were filtered against the wild-type sample with SURVIVOR, set at 50% overlap. Furthermore, for the DUAL sample, SVs that were present in a parental line, HOM, were filtered out as well. AnnotSV (<https://lbgf.fr/AnnotSV/runjob,version 2.4>) webserver was used to annotate each callset.

3.2.3.3 Structural variant quality control

The list of problematic genomic regions was compiled from four sources: genomic superSups (UCSC browser, hg19), centro- and telomeres (UCSC browser, hg19), blacklisted regions from ENCODE project³⁴, list of regions to exclude for lumpy-sv caller (<https://github.com/hall-lab/speedseq/blob/master/annotations/ceph18.b37.lumpy.exclude.2014-01-15.bed>). The four sources have been merged to a non-redundant set of 12,774 segments. All variant SV sets were filtered against this list of regions. Next, the SV call sets of each sample were further split into unbalanced SVs, e.g. deletions and duplications or other (e.g. insertions, inversions). Duphold (version 0.2.5) was run for deletions and duplications and filtering as recommended by the developer was performed producing a set of high quality CNVs.

3.2.3.4 Off-target analysis

Potential unanticipated target sites of editing by the CRISPR-Cas9 system were predicted by two different methods^{24,25} and pooled into a list of intervals for each round of editing (Table S2). All high quality SNV and indel variants within 300bp of these sites were extracted with GATK SelectVariants and manually evaluated for each sample and pairwise sample intersections. In each line, variants in the final set of structural variants that fell into potential off-target interval were called separately and examined manually.

3.2.4 Generation of NPCs

NPCs were generated as described in³⁵ with slight modifications to adapt the protocol for feeder-independent ESC cultures. H1-ESCs were dissociated using versene for 7 min at room temperature. EBs were generated by transferring dissociated ESCs to non-adherent plates in mTeSR1 with 10 μ M Y-27632 on a shaker in an incubator at 37 °C/5% CO₂. EBs were grown for 2 days in mTeSR1. On day 2 (d2), half of the medium was replaced with mTeSR1 without Y-27632. The medium was transitioned into neural induction medium (DMEM/F12, 1% N2 supplement (Thermo Fisher Scientific), 2 μ g/ml heparin (Sigma-Aldrich) and 1% penicillin/streptomycin) by replacing half of the medium at d3 and cultured for another 4 days in suspension (d4–d7) with medium replacements at d4 and d6. For generation of neural precursor cells (NPCs), EBs were slightly dissociated at d8 by trituration and plated onto laminin-coated 10 cm dishes (20 μ g/ml laminin (Sigma-Aldrich) in DMEM for 30 min at 37 °C), initially using neural induction medium (d8–15), and then from d16 in NPC medium (DMEM/F12, 1% N2 supplement, 2% B27-RA supplement (Thermo Fisher Scientific), 1 μ g/ml laminin, 20 ng/ml basic fibroblast growth factor (Merck-Millipore, Darmstadt, Germany) and 1% penicillin/streptomycin). On d16, cells were considered pre-NPCs (passage 1) and able to be passaged (1:4) and cryopreserved when confluent. Between passage 5 and 11, cells were considered NPCs and used for neural differentiation.

Before neural differentiation, NPCs were purified by fluorescence activated cell sorting (FACS), as described in³⁶. Cells were dissociated with accutase for 1-2 minutes at 37 °C and washed with PBS and resuspended in 100ul PBS with 2%FBS per 10cm dish used. Cells were labelled by incubating with fluorochrome-conjugated antibodies for 30 minutes on ice (see table S5 for antibodies and their dilutions). Cells were washed twice and resuspended in 200 μ l PBS with 2%FBS. They were passed through a 70 μ m cell strainer, before being sorted at approximately 85000 cells/cm² in a plate with NPC medium with a FACSAria (BD Biosciences) with 100 μ m nozzle.

3.2.5 Neural differentiation

NPCs (passages 5–11) were plated on sterile coverslips in 12-well plates that were coated with poly-L-ornithine (Sigma-Aldrich) for 1 h at room temperature. Coated coverslips were washed 3 times with sterile water and dried for 30 min. Subsequently, a 75 μ l drop of laminin solution (50 μ g/ml in water) was placed in the middle of each coverslip, incubated for 30 min at 37 °C/5% CO₂ and then replaced with a 75 μ l drop of DMEM until plating of NPCs. Immediately before plating, NPCs were washed with DPBS and dissociated with accutase. Cells were seeded at approximately 100,000 cells per coverslip. In the second differentiation experiment, the seeding density was adjusted because HOM networks were

more sparse in the first experiment. NPC growth rate was calculated for each line from cell counts at moments of passaging. WT, HOM and DUAL cultures had an average growth rate of 1.3 per day, compared to an average growth rate of 1.4 per day for WT and HET in the first differentiation experiment. Seeding density was therefore increased with a factor 1.9 to adjust for the estimated growth rate and have the cultures reach similar densities at 4 days after plating. To plate the cells, a 75 μ l drop of NPC cell suspension was placed on the laminin-coated spot for 1 h to let the NPCs attach on the coverslips. Afterwards, the well was filled with 1ml neural differentiation medium (Neurobasal medium, 1% N2 supplement, 2% B27-RA supplement, 1% minimum essential medium/non-essential amino acid, 20 ng/ml brain-derived neurotrophic factor (ProSpec Bio, Rehovot, Israel), 20 ng/ml glial cell-derived neurotrophic factor (ProSpec Bio), 1 μ M dibutyryl cyclic adenosine monophosphate (Sigma Aldrich), 200 μ M ascorbic acid (Sigma Aldrich), 2 μ g/ml laminin and 1% penicillin/streptomycin). After 1 h, 425 μ l of neural differentiation medium was added to each well. Cells were refreshed with medium 3 times per week. During weeks 1–4, medium was fully refreshed. After 4 weeks of neural differentiation, only half of the volume of medium per well was refreshed. Live cell calcium imaging and confocal imaging were performed between 8 and 10 weeks after plating of NPCs.

3.2.6 Immunocytochemistry

After 8 weeks in culture, cells were fixed using 4% paraformaldehyde solution for 15 minutes at room temperature and permeabilised in wash solution (0.1% Triton X-100 in PBS). Antibodies were diluted in blocking solution (1% Fish Gelatine, 0.1% Triton X-100, 5% BSA in PBS). Cells were incubated at 4°C overnight with primary antibodies (see table S5 for antibodies and their dilutions). Cells were incubated with secondary antibodies for 1 h at room temperature. Stainings for PSD-95 were performed with a biotinylated secondary antibody incubation in between to amplify the signal. Nuclei were stained in a final incubation with Hoechst (Sigma). Cells were imaged on a Zeiss LSM800 confocal microscope. Images were captured using ZEN software (Zeiss), and analysed using Fiji³⁷. Synapses were quantified using the SynQuant plug-in³⁸. Statistical analyses were carried out using a Kruskal-Wallis test and Dunn's test for multiple comparisons with Benjamini-Hochberg correction. Batch effects were assessed by a Wilcoxon rank-sum test. Data are expressed as the mean \pm 95% confidence interval of the mean.

3.2.7 RNA-seq

Cells were lysed in RLT+ buffer from the RNeasy plus Mini kit (Qiagen). RNA was extracted using an adapted kit protocol, in which RNA is precipitated in 1.5 volumes of 100% ethanol instead of 1 volume of 70% ethanol. The quality and quantity of total input mRNA was determined on an Agilent BioAnalyzer 2100 using Agilent RNA 6000 Nano kit Samples meeting the quality criteria (\geq 200 ng total RNA, RIN \geq 8 and 28S/18S \geq 1) were shipped to the Beijing Genomics Institute (BGI) in dry ice for RNA-sequencing.

RNA-seq reads were mapped using STAR v2.7³⁹ to the library of human transcriptome sequences obtained from ENSEMBL93 (GRCh38.p12). Read counts per gene were obtained by quantmode in STAR. Transcript-level counts to quantify the expression of *CNTNAP2* more accurately were generated using RSEM v1.3.3⁴⁰. Differentially expressed genes were

identified using DESeq2⁴¹, in which NPC batch was added as a factor in the design formula to account for the technical variation. Pathway analysis was performed using MetaScape⁴².

3.2.8 Clustering to BrainSpan

Developmental Transcriptome Dataset was obtained from BrainSpan (RNA-Seq Gencode v10 summarized to genes)⁴³. Experimental RNA-seq data was converted to FPKM. Genes that were expressed (FPKM > 5) in at least 20% of BrainSpan samples and in any of the experimental samples were selected. Samples were subsetted to samples of prenatal age and cortical origin. Age groups with less than three samples were removed (i.e. 25, 26 and 35 post-conceptual weeks). Pairwise Spearman rank correlation coefficients were calculated between cortical BrainSpan samples and experimental samples. Correlation per age was expressed as mean of multiple correlation coefficients.

3.2.9 Cell type deconvolution

Cell type composition scores were computed using BSeq-SC package that implements CIBERSORT^{44,45}. CIBERSORT uses linear support vector regression to estimate the cellular composition of each bulk sample based on the expression profiles of a set of reference cell types. The reference set was constructed based on single-cell RNA-seq data from human mid-gestational neocortex.⁴⁶ Significantly enriched genes for each cluster were selected as marker genes for the deconvolution. Gene counts in all input datasets were expressed as transcripts per million.

3.2.10 Western blotting

Frozen cell pellets were lysed in lysis buffer (0.1 M Tris, 150 mM NaCl, 10 mM EDTA, 0.2% Triton X-100, 1% PMSF, protease inhibitor cocktail) at 4 °C for 10 min and centrifuged at 10,000 g for 30 min at 4 °C, allowing cell debris to be pelleted and discarded. Western blotting was performed as described previously⁴⁷. Proteins were detected using primary antibodies for 60 min at room temperature or at 4 °C overnight. Secondary antibodies were applied for 60 min at room temperature. Antibodies were used as follows: CASPR2 (mouse monoclonal, catalogue # 165384, manufacturer NovoPro, Shanghai, China) at 1/2000 concentration, B-actin (mouse monoclonal, catalogue #A5441, manufacturer Sigma-Aldrich) at 1/5000 concentration, secondary anti-mouse HRP-conjugated antibody (goat, catalogue #1706516, Bio-Rad Laboratories, Hercules, CA, USA) at 1/5000. The reaction was developed using SuperSignal™ West Femto Maximum Sensitivity Substrate kit (catalogue #34096, ThermoFisher Scientific) for CASPR2 and ECL Prime (Sigma-Aldrich).

3.3 Results

3.3.1 Establishing human neuronal network cultures as a model to study *CNTNAP2* function

Human embryonic stem cell (hESC) *CNTNAP2* knockout lines were generated to model the functions of the gene and its isoforms in developing human neurons. I targeted the

CNTNAP2 gene locus in two sequential rounds of CRISPR/Cas9 editing (Fig. 1A). First, exon 1 was targeted for nonhomologous end joining repair to create indels that disrupt the long isoform. This long isoform knockout is similar to the *Cntnap2* null mouse model in which the first exon was also targeted¹³. A clone containing a homozygous knockout mutation of the long isoform was subsequently used to create a dual knockout line that lacks all protein isoforms. The dual isoform knockout was created using a guide RNA (gRNA) targeting exon 22 in this second round of editing (Fig. 1A-B). Both gRNAs were shown to be capable of editing via T7E1 mismatch cleavage assay on bulk DNA from the edited population of cells (Fig. S1A) and Sanger sequencing of single-cell derived clones (Fig. S1B). Clones were selected if they carried a frameshift mutation or deletion of the starting codon in the targeted transcript, according to their decomposed sequencing traces. (Fig. S1C). The experimental design consisted of four conditions (Fig. 1B): wild-type (WT), heterozygous long isoform knockout (HET), homozygous long isoform knockout (HOM) and the dual isoform homozygous knockout (DUAL). Next, the genomic stability of four CRISPR/Cas9 edited clones was examined by whole genome sequencing. *De novo* variants were called by comparing mutant cell lines to their previous generation, i.e. HET/HOM to WT, and DUAL to HOM (Fig. 1B). On-target mutations in *CNTNAP2* could be confirmed in the list of *de novo* indels (Table S1). We did not find evidence for off-target mutations at predicted sites. None of the *de novo* indels were within, or in proximity to, the potential off-target sites that were predicted using the gRNA sequences (allowing for up to four mismatches) (Table S2). We also did not find overlap between predicted off-target sites and *de novo* SNVs and SV breakpoints across the lines. This suggests that most of the observed genomic background variation results from different sources than CRISPR/Cas9 editing, such as environmental stress during cell culture⁴⁹.

To assess the potential impact of background variation in the genomes of these lines, the analysis was focussed on coding variants in genes that were expressed in embryonic stem cells and during *in vitro* neural differentiation. The load of indels and SNVs was similar in HET and HOM, but higher in DUAL (Table 2). All lines had similar numbers of *de novo* SVs. The number of indels, SNVs and SVs were comparable to those reported in similar

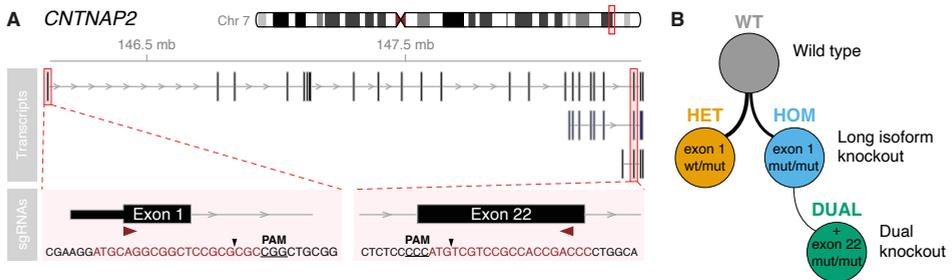


Figure 1 Development of isogenic *CNTNAP2* knockout hESCs. (A) Diagram of *CNTNAP2* gene with protein-coding transcripts (from top to bottom: ENST00000361727, ENST00000628930, ENST00000463592) and gRNAs targeting exon 1 and exon 22 to knockout the long transcript isoform or all transcript isoforms, respectively. Transcript colour indicates transcript support level (TSL)⁴⁸. Black is TSL1, blue is TSL2. (B) Pedigree of cell lines created. Heterozygous and homozygous long isoform knockouts were generated by targeting exon 1. Dual knockout was generated by an additional edit in exon 22 of the homozygous long isoform knockout clone. Thickness of connecting lines represents CRISPR/Cas9 editing efficiency (see figure S1B).

experiments^{29,50}. A subset of *de novo* variants was predicted as potentially pathogenic by SnpEff or AnnotSV (Table S1). Most genes with high impact variants are shared between conditions (87% of genes with SNV and indels and 81% of genes with SVs) (Fig. S2). Many high impact variants were found in genes that are known for their redundant sequences and being prone to false positive calls, such as mucins, zinc-finger, solute carrier protein, and olfactory receptor genes^{29,51} (Table S1). In agreement with this finding, potentially confounding mutations are not enriched in differential gene expression data from the same cell lines (Fig. S2).

Table 2 Number of indels, single nucleotide variants (SNVs) and structural variants (SVs) detected by whole genome sequencing over filtering steps.

sample	type	total	in dbSNP/DGV ^a	de novo	high impact ^b
WT	indels	2136	1829 (85.63%)	-	74 (3.46%)
HET	indels	1548	1369 (88.44%)	791 (51.10%)	59 (3.81%)
HOM	indels	2150	1914 (89.02%)	940 (43.72%)	76 (3.53%)
DUAL	indels	3417	2499 (73.13%)	1771 (51.83%)	142 (4.16%)
WT	SNVs	12058	11903 (98.71%)	-	44 (0.36%)
HET	SNVs	10494	10329 (98.43%)	5042 (48.05%)	33 (0.31%)
HOM	SNVs	11017	10860 (98.57%)	5133 (46.59%)	39 (0.35%)
DUAL	SNVs	23994	18002 (75.03%)	14089 (58.72%)	130 (0.54%)
WT	SVs	143	123 (86.01%)	-	9 (6.29%)
HET	SVs	154	138 (89.61%)	42 (27.27%)	20 (12.99%)
HOM	SVs	141	125 (88.65%)	37 (26.24%)	13 (9.22%)
DUAL	SVs	134	117 (87.31%)	20 (14.93%)	9 (6.72%)

^a Number and percentage (of total) variants from WGS that could also be found in databases for common variants: the Single Nucleotide Polymorphism database (dbSNP) for indels and SNVs and the Database of Genomic Variants (DGV) for SVs. ^b Number and percentage (of total) variants that were predicted to have functional impact by SnpEff (High) for indels and SNVs and by AnnotSV (ranking \geq class 4) for SVs

Wild-type and edited hESCs were differentiated to neuronal networks that contain a mix of neurons and astrocytes. The hESCs were first differentiated to neural precursor cells (NPCs), that then produced neurons and astrocytes in the neuronal networks via an 8-week process of neural differentiation³⁵. The maturity and relative cell type composition of this model were assessed by RNA-sequencing at four stages of its development: hESCs, NPCs, and neuronal network cultures at 4 weeks and 8 weeks differentiation. First, the developmental age of the cultures at each of these four stages was estimated by pairwise correlations with a set of postmortem human brain transcriptomes from BrainSpan⁴³. The hESC and NPC stages correlated most strongly with early age (8-9 post-conceptual weeks).

The highest correlation between neuronal network cultures and human neocortical tissue was found for mid-gestation (Fig. 2A). In this period (~16 post-conceptual weeks), the major germinal zones are formed, and the upper cortical layers start to develop⁵². This means that the transcriptome of this *in vitro* model reflects a crucial developmental stage in which neuronal specification and the initial assembly of neocortical neural circuits take place.

The estimated developmental age of this model also predicts the cell types that are reflected in these neuronal networks. I explored the relative contributions of these cell types using cell type decomposition analysis of the different stages in this model. Cell type composition scores were calculated with CIBERSORT based on a reference panel of marker genes from a single cell RNA-sequencing dataset of mid-gestation foetal cortex, because it matched the estimated developmental maturity of our model^{45,46}. Composition scores for progenitor cells were highest in the NPC developmental stage of our model and the composition scores for excitatory neurons increased in neuronal network cultures (Fig. 2B) The signal for glial cells increased at eight weeks of neuronal differentiation (Fig.

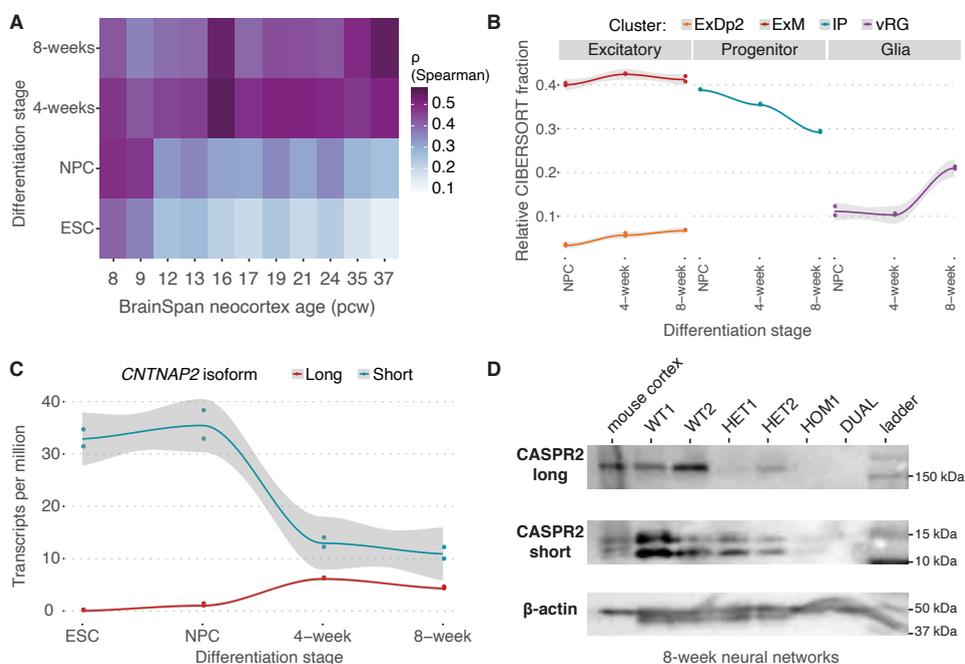


Figure 2 Cell model identity and *CNTNAP2* expression in cell model. (A) Correlation heatmap of gene expression between developmental stages of *in vitro* neuronal networks and *ex vivo* neocortical samples from BrainSpan pcw = post-conceptual weeks. (B) Graphs represent estimated relative fraction of cell types from transcriptomic data. ExDp2 = Excitatory deep-layer neurons, ExM = maturing excitatory neurons, IP = intermediate progenitors, vRG = ventricular radial glial cells. (C) *CNTNAP2* isoform expression across developmental stages of neuronal network model. (D) Western blot for short and long isoform of CASPR2 in 8-weeks neuronal networks. Lysate of adult mouse cortex was used as a positive control. β -actin serves as a loading control. WT1 and WT2 are lysates from 8-week neuronal network cultures generated from separate NPC differentiations from the same wild-type ESC line. HET1, HET2, HOM1 and DUAL are lysates from 8-week neuronal network cultures generated from separate CRISPR-edited ESC lines.

2B). Astrocytes are an essential factor for neural circuit development⁵³, and they indeed begin to emerge during mid-foetal periods in human foetal brain transcriptomes⁵⁴. The transcriptome of this model follows expected developmental trajectories and shows resemblance to mid-gestation cortex.

A final, crucial step in assessing the applicability of this model for studying the human neuronal function of *CNTNAP2* was looking at the expression of long and short isoforms. RNA-sequencing from the lines showed that the short isoform is expressed at high levels in hESCs and that it peaks in NPCs. The expression of the short isoform decreases with the initiation of neuronal differentiation, whereas the expression of the long isoform starts to increase at that stage (Fig. 2C). The onset of the long isoform roughly corresponds with mouse data in which the full-length protein is first detected at day 14 of embryonic development¹². In human foetal samples, the long isoform of CASPR2 is expressed in cortical areas at late mid-gestation⁵⁵, but the onset of expression has not been measured. In mature 8-week differentiated neuronal networks, both the long isoform (148kDa) and short isoform (12kDa) were detected at the protein level with a western blot (Fig. 2D, S3). The reduction of CASPR2 in the mutant conditions confirmed that our CRISPR-Cas9 manipulation worked. The reduction in short CASPR2 expression observed in the HOM condition further highlights that some of the 12kDa is made up by the intracellular domain that remains after ectodomain shedding of the long isoform^{56,57}.

3.3.2 Homozygous knockout of CASPR2 leads to reduction in NPCs

To explore the effect of CASPR2 on neuronal network development, mature 8-week old neuronal network cultures were generated from wild-type and mutant cell lines. First, we investigated the effect of heterozygous and homozygous mutations in exon 1 that only affect the long isoform. Four seeding replicates with at least three wells per assay were used per NPC differentiation. Two NPC differentiations were used from the WT ESC line (WT1 and WT2) and one NPC differentiation for each mutant ESC line (HET1, HET2 and HOM1). In addition, the HOM long isoform knockout and DUAL knockout lines were compared with the WT lines to study additional effects of mutations in the short isoform. In the second experiment, the plating density of NPCs was slightly adjusted for their estimated growth rate in order to have neuronal networks of more equal density (see Methods). Again, four seeding replicates with at least three wells per assay were used per NPC differentiation. The same NPC differentiation as before was used for HOM, but at a higher density (HOM1.2). New NPC differentiations were generated for WT (WT3) and DUAL (DUAL).

The proportions of cell types in the 8-week old neuronal networks were assessed by immunofluorescent (IF) staining of marker proteins for NPCs, neurons and astrocytes (Fig. 3A, B). HOM and DUAL had a significant reduction in NPCs compared to WT, as measured by the percentage of SOX2⁺ nuclei (Fig. 3C). This reduction in the number of NPCs did not result in any change in number of mature neurons in HOM and DUAL, based on the stable percentages of NeuN⁺ nuclei (Fig. 3D). These percentages are comparable to previously published data about this differentiation protocol³⁵. However, a small but significant reduction of mature neurons was registered in HET. The proportion of astrocytes did not differ across conditions, but the normalized GFAP⁺ area was significantly elevated in the second differentiation experiment compared to the first experiment (Fig. 3E).

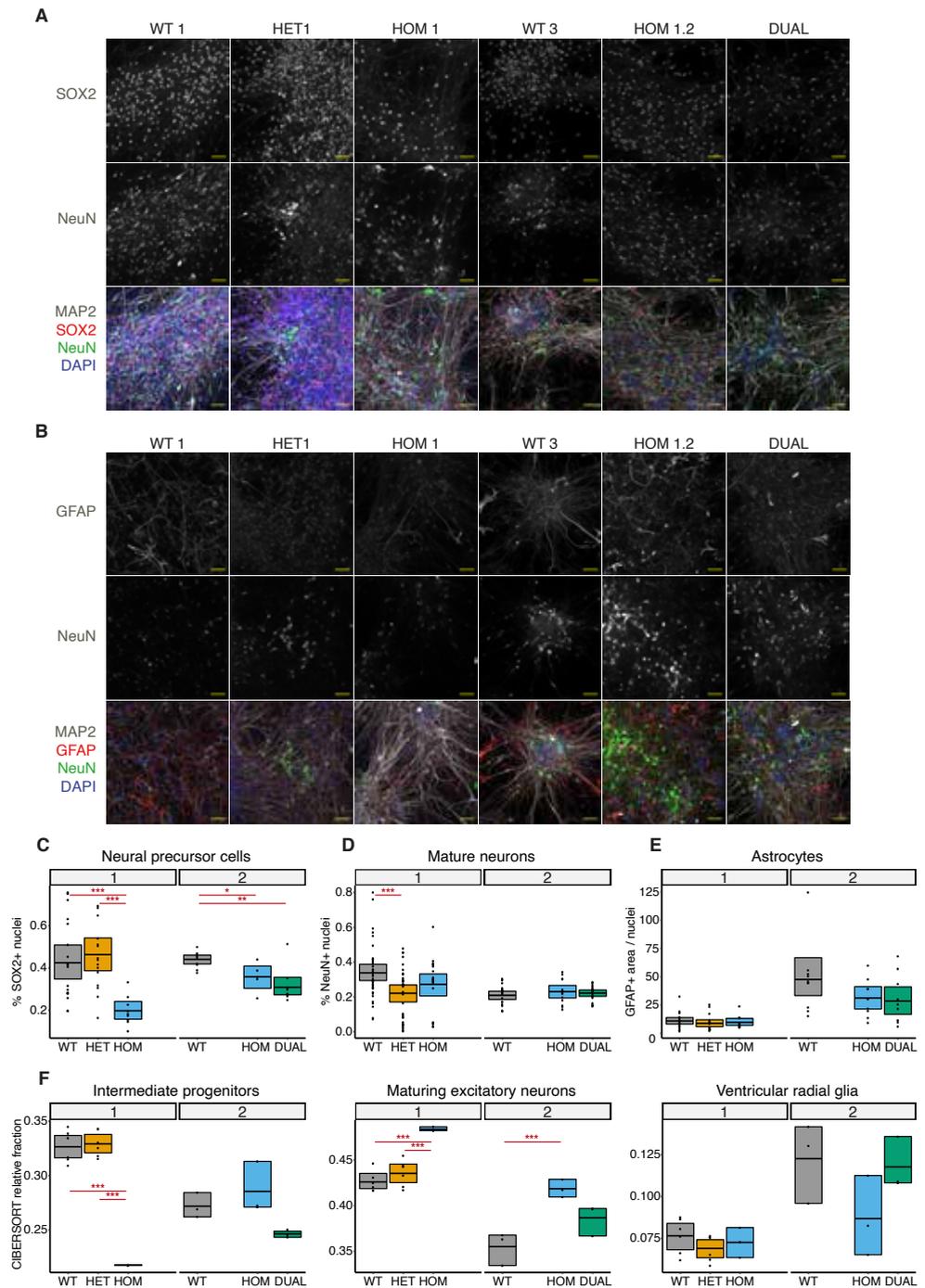


Figure 3 Cell type proportions in *CNTNAP2* knockout neurons. (A) Representative images showing SOX2 (top)
(legend continues on next page)

The immunofluorescent staining data was complemented with a cell type decomposition analysis from the transcriptomes of the neuronal networks. This analysis took into account the expression levels of hundreds of cell type markers at once that were used to construct a reference panel from single cell RNA-sequencing data (Fig. S2A). The reduction of SOX2⁺ NPCs, that was observed in HOM and DUAL lines via IF staining, was supported by a reduced composition score for intermediate progenitor cells in HOM in the transcriptomic data of the first differentiation experiment, but the differences for HOM and DUAL in the second differentiation experiment were not statistically significant (Fig. 3F). The composition score for maturing excitatory neurons showed a significant increase in the HOM condition, which was not observed in the IF staining (Fig. 3G). Possibly, this increase in the proportion of maturing neurons indicated an increase in neuronal differentiation or maturation that was not reflected in staining of fully mature neurons. However, the increase in the DUAL condition was not statistically significant. The composition scores for ventricular radial glial cells were consistent with the IF data for astrocytes, in which the proportions did not change across conditions but increased in the second differentiation experiment compared to the first differentiation experiment (Fig. 3H). Astrocytes were not included in the reference dataset, but ventricular radial glia can give rise to astrocytes and are therefore an approximation^{46,58}. Loss of CASPR2 did not affect the proportion for any of the other cell types in the cell type deconvolution analysis (Fig. S2B).

3.3.3 Knockout of *CNTNAP2* affects numbers and structure of excitatory synapses

Previous studies with mouse neurons have indicated that *Cntnap2* is involved in the formation and stabilisation of dendritic spines^{14,15,18}. The effect of *CNTNAP2* mutation on synapses was tested via immunofluorescent co-labelling of pre- and post-synaptic sites. The pre-synapse was labelled with Synapsin (SYN1/2) and the post-synapse with PSD-95 for excitatory synapses and with gephyrin (GPHN) for inhibitory synapses. Overlapping pre- and post-synaptic puncta in a MAP2 positive area were counted as a synapse using an automated algorithm. The samples were analysed per differentiation experiment, because of the technical variation in astrocyte abundance described earlier. A significant reduction in the number of excitatory synapses was detected in the DUAL knockout condition, but not for the other mutant conditions (Fig. 4B). The number of inhibitory synapses did not differ across conditions, but showed the same technical variation as measured for astrocytes (Fig. 5B).

and NeuN (middle) labelling. Merged image (bottom) is counterstained with MAP2 labelling and DAPI. Scale bar is 50µm. (B) Representative images showing GFAP (top) and NeuN (middle) labelling. Merged image (bottom) is counterstained with MAP2 labelling and DAPI. Scale bar 50µm. (C-E) Quantification of cell type markers relative to number of DAPI stained nuclei. (C) Percentage of SOX2 positive nuclei (n = 5-10). (D) Percentage of NeuN positive nuclei (n = 14-20). (E) GFAP area per image normalized to number of nuclei (n = 8-11). Each dot represents one image. Boxplots represent mean with 95% confidence interval. *p<0.05, *p<0.01, ***p<0.001, Kruskal-Wallis test and Dunn's test for multiple comparisons. (F-H) Graphs represent estimated of cell type abundance from transcriptomic data. (F) Relative fraction of intermediate progenitors. (G) Relative fraction of maturing excitatory neurons. (H) Relative fraction of ventricular radial glia cells. Boxplots represent mean with 95% confidence interval. *p<0.05, *p<0.01, ***p<0.001, ANOVA test and Tukey HSD's test for multiple comparisons. WT = wildtype, HET = heterozygous long isoform knockout, HOM = homozygous long isoform knockout, DUAL = dual isoform knockout.

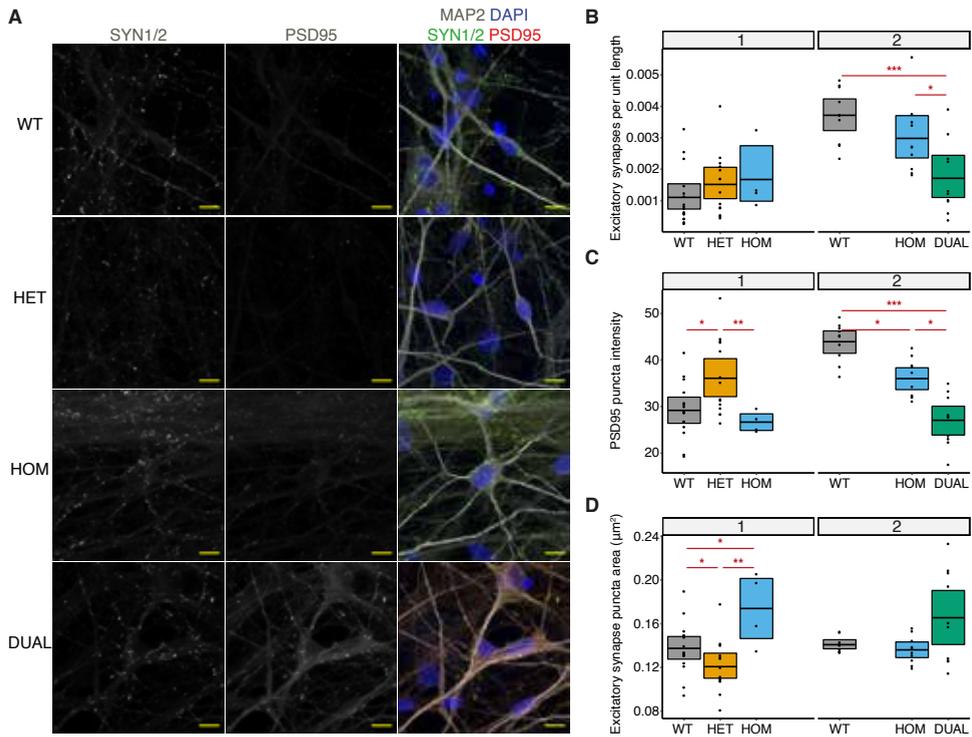


Figure 4 Excitatory synapses in *CNTNAP2* knockout neurons. (A) Representative images showing pre-synaptic SYN1/2 (top) and post-synaptic PSD95 (middle) labelling. Merged image (bottom) is counterstained with MAP2 labelling and DAPI. Scale bar is 10µm. (B) Quantification of overlapping SYN1/2 and PSD95 puncta per unit length of MAP2 split per differentiation experiment. (C) Quantification of intensity of PSD95 puncta staining. (D) Quantification of PSD95 puncta size. Each dot represents one image. Boxplots represent mean with 95% confidence interval. * $p < 0.05$, * $p < 0.01$, *** $p < 0.001$, Kruskal-Wallis test and Dunn's test for multiple comparisons.

PSD-95 is a known interaction partner for both isoforms of CASPR2¹⁹. At the juxtaparanode, CASPR2 is required for the accumulation of PSD-95⁷. To see if loss of CASPR2 affects the structure of excitatory post-synapses, the size and intensity of post-synaptic puncta were also quantified. Heterozygous and homozygous mutant conditions showed opposite effects. PSD95 intensity was increased in synapses in the heterozygous knockout but decreased in the homozygous knockout and significantly further in the dual knockout (Fig. 4C). The size of PSD95 puncta was decreased in the heterozygous knockout, but not consistently increased in the homozygous knockout (Fig. 4D). Intensity and area of GPHN puncta were consistent across conditions (Fig. 5C, D).

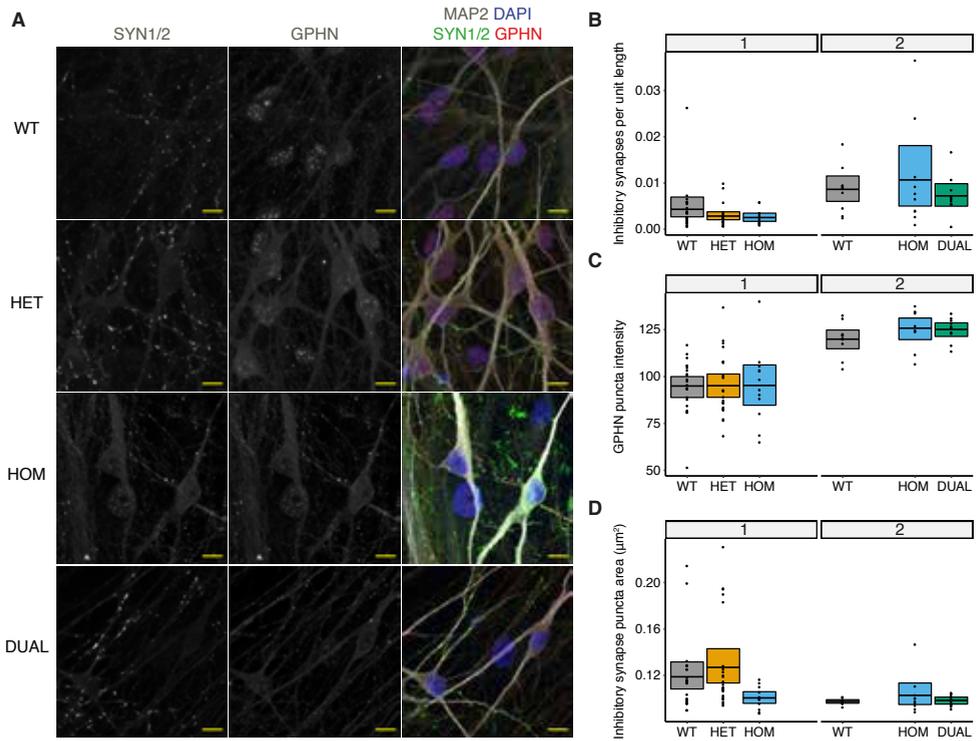


Figure 5. Inhibitory synapses in *CNTNAP2* knockout neurons. (A) Representative images showing pre-synaptic SYN1/2 (top) and post-synaptic GPHN (middle) labelling. Merged image (bottom) is counterstained with MAP2 labelling and DAPI. Scale bar is 10µm (B) Quantification of overlapping puncta per unit length of MAP2 split per differentiation experiment. (C) Quantification of intensity of GPHN puncta staining. (D) Quantification of GPHN puncta size. Each dot represents one image. Boxplots represent mean with 95% confidence interval. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Kruskal-Wallis test and Dunn's test for multiple comparisons.

3.3.4 Transcriptome analysis uncovers biological processes affected by loss of *CASPR2*

To evaluate the transcriptome changes associated with the phenotypes described above, bulk RNA sequencing was performed on the 8-week old neuronal networks. Principal component analysis (PCA) of the RNA-seq data shows a clear separation in the transcriptional pattern between the genotype groups, with 58% of the variance explained by principal component 1 (fig. 6A). In this graph, the heterozygous long isoform knockout is relatively close to the wild-type condition, whereas the homozygous long isoform knockout and dual knockout cluster further away. These distances are also reflected in the number of differentially expressed genes (DEGs) that were detected in the comparisons between WT and each of the mutant conditions (fig. 6B). NPC batch was added as a factor in the design formula of the DEG analysis to account for the technical variation in WT and HOM samples (see Methods). Next, we compared the three WT-mutant comparisons. The set of DEGs from WT vs. HET was smallest (633 genes). The WT vs. HOM (3999 genes) and

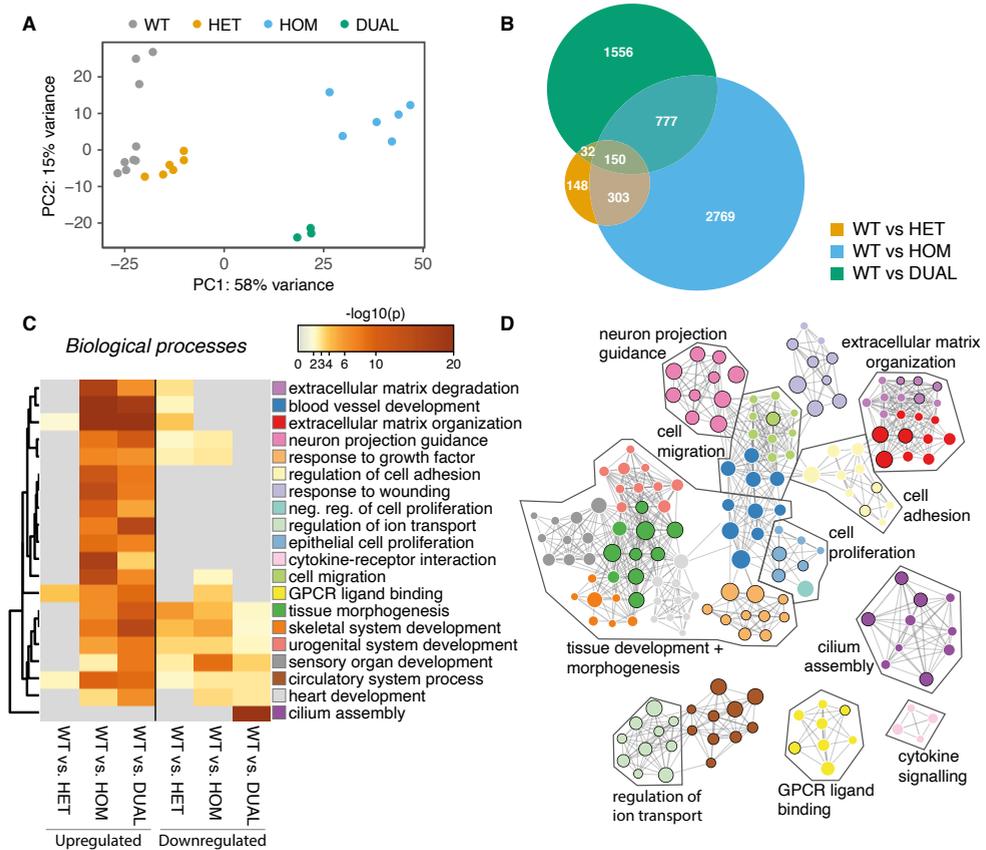


Figure 6 Differential gene expression in *CNTNAP2* knockout neurons. (A) Principal component analysis displays clustering of samples per condition and distance between conditions. (B) Euler diagram represents sets of differentially expressed genes (fold change >2 and FDR-corrected $p < 0.05$) per mutant-wildtype comparison. (C) Heatmap showing the top enrichment clusters for GO Biological processes, KEGG Pathways and Reactome Gene Sets, one row per cluster, using a discrete colour scale to represent statistical significance. Gray colour indicates a lack of significance. DEG lists are split into upregulated and downregulated. (D) Enrichment network visualisation for results from the six gene lists. Node colours indicate cluster and node size reflects number of DEGs in GO. Nodes containing CASPR2 candidate interaction partners (Table S3) are marked by black borders. Edge thickness represents the similarity score between GOs. More general labels were added manually.

WT vs. DUAL comparisons (2515 genes) were much larger and had the largest overlap of DEGs (777 genes). Still, the majority of DEGs were specific to each particular comparison. Among these comparison-specific DEGs, there were DEGs that reflect pathways that are affected by having a different isoform.

Gene ontology (GO) enrichment analysis was performed for the DEGs from each comparison. The DEG lists were split into upregulated and downregulated genes to facilitate biological interpretation of the effects. All mutant conditions showed enrichment for extracellular matrix binding and GPCR ligand binding among upregulated genes and for various categories related to tissue morphogenesis among downregulated genes (fig. 6C). GO clusters for neuron projection guidance, response to growth factor, regulation of cell adhesion, regulation of cell migration, cell proliferation, ion transport and cytokine signalling, were overrepresented in the upregulated DEGs from HOM and DUAL. Since these pathways are active in specific cell types, these enriched GO clusters could be related to changes in cell type composition in HOM and DUAL: a decrease in NPCs and increase in maturing neurons (Fig. 3C,F). However, downregulated DEGs from DUAL were specifically related to cilium assembly, suggesting a contribution of the short isoform to these processes. Whereas most GO category clusters in this analysis were connected via shared DEGs in a network diagram (fig. 6F), cilium assembly was separated from other clusters.

To explore why loss of both isoforms of CASPR2 leads to this cilium assembly gene expression phenotype, information about the candidate interaction partners of CASPR2 was collated from multiple studies and annotated with the putative interacting isoforms where possible (Table S3). By intersecting this list of potential downstream effectors of CASPR2 with the DEGs that are related to cilium assembly, it was discovered that Tectonic family member 1 (TCTN1) could play a central role in this phenotype. TCTN1 has been discovered as CASPR2 interaction partner in a yeast two-hybrid screen for the C-terminal domain of all isoforms of CASPR2⁸. TCTN1 is part of the tectonic-like complex at the transition zone of primary cilia. The transition zone acts as a barrier for diffusion of proteins between the cilia and plasma membranes during ciliogenesis⁵⁹. The transition zone interactome has been reviewed by multiple authors^{60,61}. Almost half of the genes encoding these transition zone proteins were significantly downregulated and a large majority showed a downward trend in DUAL (Fig. 7A). As shown on a schematic of the transition zone interactome (Fig. 7B) significantly downregulated genes are located across the entire transition zone protein complex. A few of the most significant DEGs cluster around TCTN1, such as TCTN2, TMEM231 and TMEM67 (Fig. 7B). Given that the short isoform is most highly expressed at the NPC stage (Fig. 2D), loss of the short isoform, and any associated effects on cilia, may be most impactful at this stage. Examining the expression of these transition zone DEGs, it became apparent that almost all of them have their peak or onset of expression at the NPC stage (Fig. 7C). These data indicate that the relevant DEGs are co-expressed with the short isoform of CASPR2 and supports the hypothesis that this isoform may play a role in cilium assembly through proteins at the transition zone.

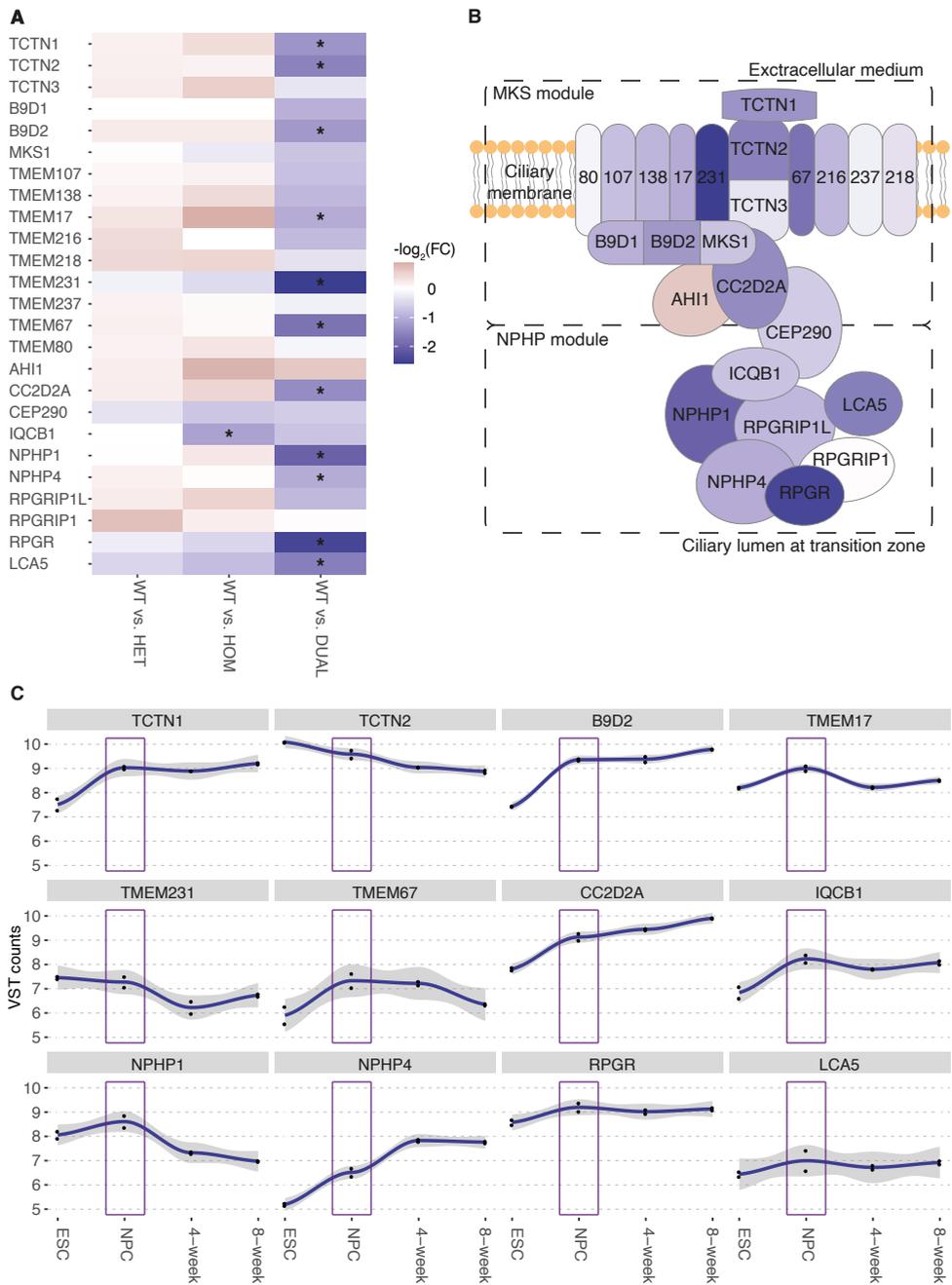


Figure 7 Gene expression effects on ciliary transition zone genes. (A) Heatmap of \log_2 fold changes across mutant conditions. Red is higher expressed and blue is lower expressed in the mutant condition. Significant differentially expressed genes are marked with an asterisk (fold change > 2 and adjusted p-value < 0.05).

(legend continues on next page)

3.4 Discussion

Mutations in *CNTNAP2* are related to a broad range of patient phenotypes with different degrees of severity ². Genome engineering and human cell models are essential tools to gain insight into the mechanisms underlying these disorders and help explain this phenotypic diversity. In this study, human neuronal network cultures were generated with different *CNTNAP2* knockout mutations to compare the severity of heterozygous and homozygous loss of the long isoform and to study the potential additive effects of homozygous loss of a short protein isoform that lacks the extracellular domain. Comparison of neuronal network cultures at the level of cell type composition, synapse numbers, and molecular pathways provides a picture in which severity of the phenotype increases with the number of alleles and isoforms that are disrupted. Heterozygous loss of the long isoform produced relatively minor gene expression changes and no effect on cell type composition or synapses. Homozygous loss of the long isoform resulted in larger transcriptomic changes and a reduction in the number of NPCs. Those network cultures with dual isoform *CNTNAP2* knockout were most severely affected. These had fewer excitatory synapses and showed specific transcriptomic changes that indicated a defect in cilia assembly, in addition to the reduction in NPCs. Due to the limited number of repeats in this study, these changes need to be interpreted with some caution and the hypothesised effects will require further validation in the future.

3.4.1 Cell type composition

The development of different cell types in correct proportions is important for the proper organisation and functioning of neuronal networks. Cell type composition can be affected by multiple processes, such as proliferation, differentiation and cell death. A recent study found that forebrain organoids with homozygous mutations in *CNTNAP2* are larger in size driven by an increased proliferation of NPCs ²². The significantly larger pool of NPCs that was quantified in early-stage forebrain organoids is contrary to the data in this study. This discrepancy could be explained by differences in these two model systems since our study uses two-dimensional neuronal network cultures. Effects on NPC proliferation could be more pronounced in an organoid model, as the slower maturation rate gives more time to develop such a phenotype or the phenotype may depend on the more complex micro-environment of three-dimensional cultures ⁶². The 2D and 3D cell models show consistent results in the significant upregulation of similar genes related to neurogenesis (Table S4) ²², which also affects cell type composition. An increase in neuronal differentiation of NPCs could lead to a decrease in NPCs in the absence of increased proliferation. This means that homozygous loss of *CASPR2* affects the cell type composition balance between NPCs and neurons during mid-foetal cortex development, which may affect neocortical circuit formation.

Although *CASPR2* is involved in neuron-glia interactions ⁶³, a difference in normalized GFAP⁺ area was not detected across experimental conditions, suggesting that *CASPR2*

(B) Schematic drawing of the transition zone (adapted from ⁶¹) shows the spatial distribution of all genes described in the heatmap. Proteins are coloured with the \log_2 fold changes in the WT vs. DUAL comparison. (C) Gene expression profiles of significant DEGs in the ciliary transition zone during developmental stages of neuronal network model with variance stabilizing transformed counts. NPC stage is highlighted with purple boxes.

loss does not affect astrocyte composition of the networks. This is in line with histological experiments in the hippocampus, medial prefrontal cortex and striatum of adult *Cntnap2* knockout mice that also did not detect changes in the number or size of astrocytes^{13,64}. However, there was a significant difference in normalized GFAP⁺ area between experiments. The adaptations to NPC seeding density may have led to higher astrocyte density. In monolayer cultures, astrocyte-to-astrocyte contact promotes astrocyte maturation, which leads to astrocytes with more extended processes that cover a larger area⁶⁵. These batch effects mean that the study had less power to detect condition-specific differences in GFAP⁺ area. As astrocytes contribute to synapse development, these batch effects may also affect the power of assays of network structure and function⁵³.

3.4.2 Synapse numbers and structure

Loss of CASPR2 in mouse neurons leads to decreased numbers of excitatory and inhibitory synapses both *in vivo* and *in vitro*^{14-16,18,66,67}. This study is among the first to track the effect of CASPR2 mutation on synapses in human cultures, thanks to a differentiation protocol that produces electrophysiologically mature human neurons³⁵. Given the low proportion of inhibitory neurons that are generated with this protocol, it was not possible to observe differences in the number of inhibitory synapses in this system. Using PSD-95 labelling, a significant decrease in number and intensity of excitatory synapses was detected in the DUAL condition, and a decrease in intensity in the HOM condition in the second differentiation for which the NPC seeding density was adapted to growth rate. Variation in the first differentiation, such as the increase in PSD-95 intensity in HET, could be related to varying density between conditions. PSD-95 is required for the molecular organisation of the post-synaptic density, and a decrease is associated with increased synaptic turnover⁶⁸. Electron microscopy of neurons in *Cntnap2* knockout mouse brain slices has shown an increased number of perforated post-synaptic densities, which are associated with increased synaptic turnover¹⁸. Lower levels of PSD-95 may also affect synaptic transmission, as PSD-95 is essential for the correct localisation of AMPA receptors⁶⁹. *Cntnap2* knockout mouse neurons indeed show reduced expression and mislocalisation of AMPA receptors, which leads to reduced synaptic transmission^{14,70}. PSD-95 is a protein-protein interaction partner of CASPR2. The decrease of PSD-95 identified here may therefore present a step in the mechanism between loss of Caspr2 and lower synapse numbers and reduced synaptic transmission.

3.4.3 Molecular networks

Looking at the gene expression profile of mature neuronal networks, we found support for biological processes that are disrupted by loss of CASPR2 and generated new hypotheses for processes that are related to CASPR2 function in neurons. Previous transcriptome studies in human patient-derived cell models have indicated several pathways that may be affected by disruptions of the gene. DEGs in induced neurons with a heterozygous deletion in *CNTNAP2* were most significantly enriched for ontologies related to extracellular matrix organisation and tissue morphogenesis²³, which mirrors our findings in the HET condition. DEGs in cortical organoids with a homozygous missense mutation in exon 22 affecting both isoforms of *CNTNAP2* further identified enrichment for ontologies related to axon guidance, neuronal differentiation and cell proliferation²².

These ontologies are also reflected in this model (Table S4), supporting putative functions of CASPR2 in these pathways in developing human neurons. This study additionally suggested a novel pathway that was only identified in gene ontology analyses when both isoforms of CASPR2 were lost. When both the long and short isoform were disrupted, cilia assembly pathways were significantly disrupted, suggesting this pathway is related more directly to the function of the short isoform.

Primary cilia are non-motile, microtubule-based organelles that sense extracellular cues and integrate multiple important signalling pathways during brain development, which gives them an essential role in processes such as neuronal cell fate, migration and differentiation⁷¹. Given this wide array of functions of the primary cilium, it is important to consider how CASPR2 may interact with the ciliary proteome to understand the functional relevance of primary cilia in CASPR2-related pathways. A yeast-two-hybrid screening experiment previously identified TCTN1 as a direct interaction partner of the CASPR2 intracellular domain⁸. TCTN1 is found in the ciliary transition zone – a region at the base of the cilium that forms a gate to regulate the transport of proteins⁶⁰. Out of 25 known transition zone genes, 11 of these were significantly downregulated in the DUAL cells including TCTN1 and TCTN2. Knockout studies have shown that TCTN1 and TCTN2 are indispensable for neuronal ciliogenesis^{59,72}. Both TCTN1 and TCTN2 are members of the tectonic-like complex and regulate the localisation of a protein called ARL13B to the ciliary membrane^{59,73}. The tectonic-like complex and ARL13B have relevance for neural development, as ARL13B is important for Sonic hedgehog signalling and radial migration of neurons and interneurons, as shown in a knockout mouse model^{74,75}. A similar effect on the layer position of neurons can also be observed after knockdown of TCTN2⁷⁶. Through its interaction with TCTN1 and the tectonic-like complex, loss of the short isoform of CASPR2 may thus affect the localisation of ARL13B, leading to defects in neuronal ciliogenesis and neurodevelopment. Future experiments may look at the localisation of these proteins and the magnitude and timing of defects in primary cilia assembly.

Disruptions in neuronal ciliogenesis can lead to focal cortical dysplasia, as was shown in the brains of patients and mouse models where ciliogenesis was disrupted by mutations in MTOR⁷⁷. Cortical dysplasia is a remarkable feature of the first cases of CASPR2 deficiency disorder, described in a group of Old Order Amish children²⁰. This phenotype manifested as an increase in cell density in the temporal lobe, observed via MRI scans and histological stains of the brains of patients with homozygous *CNTNAP2* mutations. This phenotype could be specific to these Old Order Amish children, as they are the only individuals that have been reported with a homozygous mutation that affects the short isoform. Cortical dysplasia has generally not been found in individuals with other bi-allelic *CNTNAP2* mutations that affect only the long isoform^{1,20,78,79}. Hypoplasia was observed in the cerebellum of two separately reported cases, but this is arguably a different symptom as it is a decrease instead of an increase in cell density and affects a different brain region^{1,79}. The widely used CASPR2 knockout mouse carries a frameshift deletion in exon 1 that knocks out the long isoform, but does not affect the short isoform. The intact expression of the short isoform, and its role in primary cilia function, as uncovered here, may explain why the cortical dysplasia phenotype characteristic of the Old Order Amish patients is not observed in the mouse model^{17,80}. This highlights the importance of exploring the molecular properties of genes in order to understand language-related disorder aetiology and to be able to model them appropriately in animal systems.

3.4.4 Clinical implications and future research

The phenotypic profile of the different *CNTNAP2* mutant conditions in this study have implications for assessing their contributions to disease risk. Patient mutations in *CNTNAP2* have varying outcomes and heterozygous mutations can be found in unaffected individuals². It is possible that the position of the mutation makes a difference. This study found that heterozygous mutations in the long isoform had a relatively small effect on global gene expression in our neuronal model, which may not always be sufficient to cause disorder on its own. This supports the view that heterozygous *CNTNAP2* mutations may not contribute to disease risk for several neuropsychiatric disorders, as concluded in a comprehensive genetic analysis of SNPs, CNVs and rare pathogenic variants in *CNTNAP2*⁸¹. Together with our findings, these data suggest that homozygous mutations in *CNTNAP2* or accompanying mutations in other genes could lead to more severe effects related to neurodevelopment. The results highlight the importance of screening heterozygous patients for additional mutations, as well as consideration of the downstream molecular effects of heterozygous loss of *CASPR2* as a modulating factor in backgrounds of increased risk. The results in this study further suggest that especially mutations that affect the short isoform may need extra attention, as they may be more severe due to an earlier developmental impact and potential ciliary defect. Further research into the roles of *CNTNAP2* in cilia function during early brain development is needed to understand its contributions to neuronal migration and cortical lamination. This study shows that human neurons and gene editing provide a versatile platform for discovering new functions of genes and their isoforms. The approach can be extended by applying to more complex human models that allow the tracking of neuronal migration in real time^{82,83}. By studying the isoform-specific functions of *CASPR2* and combining the outcomes from multiple experimental models, we will gain an advanced understanding of the multiplicity of molecular mechanisms underlying the clinical complexity of *CNTNAP2*-related disease.

References

- 1 Rodenas-Cuadrado, P. *et al.* (2016). Characterisation of *CASPR2* deficiency disorder--a syndrome involving autism, epilepsy and language impairment. *BMC Med Genet* 17, 8, doi:10.1186/s12881-016-0272-8.
- 2 Rodenas-Cuadrado, P., Ho, J. & Vernes, S. C. (2014). Shining a light on *CNTNAP2*: complex functions to complex disorders. *Eur J Hum Genet* 22, 171-178, doi:10.1038/ejhg.2013.100.
- 3 Geschwind, D. H. & Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science* 349, 1489-1494, doi:10.1126/science.aaa8954.
- 4 Gordon, A. *et al.* (2016). Expression of *Cntnap2* (*Caspr2*) in multiple levels of sensory systems. *Mol Cell Neurosci* 70, 42-53, doi:10.1016/j.mcn.2015.11.012.
- 5 Pinatel, D. *et al.* (2015). Inhibitory axons are targeted in hippocampal cell culture by anti-*Caspr2* autoantibodies associated with limbic encephalitis. *Front Cell Neurosci* 9, 265, doi:10.3389/fncel.2015.00265.
- 6 Poliak, S. *et al.* (2001). Localization of *Caspr2* in myelinated nerves depends on axon-glia interactions and the generation of barriers along the axon. *J Neurosci* 21, 7568-7575, doi:10.1523/JNEUROSCI.21-19-07568.2001.
- 7 Horresh, I. *et al.* (2008). Multiple molecular interactions determine the clustering of *Caspr2* and *Kv1* channels in myelinated axons. *J Neurosci* 28, 14213-14222, doi:10.1523/JNEUROSCI.3398-08.2008.
- 8 Gao, R. *et al.* (2018). *CNTNAP2* stabilizes interneuron dendritic arbors through *CASK*. *Mol Psychiatry* 23, 1832-1850, doi:10.1038/s41380-018-0027-3.
- 9 Denisenko-Nehrbass, N. *et al.* (2003). Protein 4.1B associates with both *Caspr/paranodin* and *Caspr2* at paranodes and juxtaparanodes of myelinated fibres. *Eur J Neurosci* 17, 411-416, doi:10.1046/j.1460-9568.2003.02441.x.
- 10 Gao, R. *et al.* (2020). *CNTNAP2* is targeted to endosomes by the polarity protein *PAR3*. *Eur J Neurosci* 51, 1074-1086, doi:10.1111/ejn.14620.

- 11 Bel, C., Oguievetskaia, K., Pitaval, C., Goutebroze, L. & Faivre-Sarrailh, C. (2009). Axonal targeting of Caspr2 in hippocampal neurons via selective somatodendritic endocytosis. *J Cell Sci* 122, 3403-3413, doi:10.1242/jcs.050526.
- 12 Penagarikano, O. *et al.* (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147, 235-246, doi:10.1016/j.cell.2011.08.040.
- 13 Poliak, S. *et al.* (2003). Juxtaparanodal clustering of Shaker-like K⁺ channels in myelinated axons depends on Caspr2 and TAG-1. *J Cell Biol* 162, 1149-1160, doi:10.1083/jcb.200305018.
- 14 Varea, O. *et al.* (2015). Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. *Proc Natl Acad Sci U S A* 112, 6176-6181, doi:10.1073/pnas.1423205112.
- 15 Gdalyahu, A. *et al.* (2015). The Autism Related Protein Contactin-Associated Protein-Like 2 (CNTNAP2) Stabilizes New Spines: An *In Vivo* Mouse Study. *PLoS One* 10, e0125633, doi:10.1371/journal.pone.0125633.
- 16 Jurgensen, S. & Castillo, P. E. (2015). Selective Dysregulation of Hippocampal Inhibition in the Mouse Lacking Autism Candidate Gene CNTNAP2. *J Neurosci* 35, 14681-14687, doi:10.1523/JNEUROSCI.1666-15.2015.
- 17 Scott, R. *et al.* (2019). Loss of Cntnap2 Causes Axonal Excitability Deficits, Developmental Delay in Cortical Myelination, and Abnormal Stereotyped Motor Behavior. *Cereb Cortex* 29, 586-597, doi:10.1093/cercor/bhx341.
- 18 Lazaro, M. T. *et al.* (2019). Reduced Prefrontal Synaptic Connectivity and Disturbed Oscillatory Population Dynamics in the CNTNAP2 Model of Autism. *Cell Rep* 27, 2567-2578 e2566, doi:10.1016/j.celrep.2019.05.006.
- 19 Chen, N. *et al.* (2015). Interaction proteomics of canonical Caspr2 (CNTNAP2) reveals the presence of two Caspr2 isoforms with overlapping interactomes. *Biochim Biophys Acta* 1854, 827-833, doi:10.1016/j.bbapap.2015.02.008.
- 20 Strauss, K. A. *et al.* (2006). Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med* 354, 1370-1377, doi:10.1056/NEJMoa052773.
- 21 Raj, B. & Blencowe, B. J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron* 87, 14-27, doi:10.1016/j.neuron.2015.05.004.
- 22 De Jong, J. O. *et al.* (2021). Cortical overgrowth in a preclinical forebrain organoid model of CNTNAP2-associated autism spectrum disorder. *Nat Commun* 12, 4087, doi:10.1038/s41467-021-24358-4.
- 23 Flaherty, E. *et al.* (2017). Patient-derived hiPSC neurons with heterozygous CNTNAP2 deletions display altered neuronal gene expression and network activity. *NPJ Schizophr* 3, 35, doi:10.1038/s41537-017-0033-5.
- 24 Hsu, P. D. *et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31, 827-832, doi:10.1038/nbt.2647.
- 25 Concordet, J. P. & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* 46, W242-W245, doi:10.1093/nar/gky354.
- 26 Ran, F. A. *et al.* (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8, 2281-2308, doi:10.1038/nprot.2013.143.
- 27 Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42, e168, doi:10.1093/nar/gku936.
- 28 Hsiao, T. *et al.* (2019). Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*, 251082, doi:10.1101/251082.
- 29 Budde, J. P. *et al.* (2017). Precision genome-editing with CRISPR/Cas9 in human induced pluripotent stem cells. *bioRxiv*, 187377, doi:10.1101/187377.
- 30 Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30, 2843-2851, doi:10.1093/bioinformatics/btu356.
- 31 Clarke, L. *et al.* (2017). The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res* 45, D854-D859, doi:10.1093/nar/gkw829.
- 32 Boeva, V. *et al.* (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28, 423-425, doi:10.1093/bioinformatics/btr670.
- 33 Kuzniar, A. *et al.* (2020). sv-callers: a highly portable parallel workflow for structural variant detection in whole-genome sequence data. *PeerJ* 8, e8214, doi:10.7717/peerj.8214.
- 34 Amemiya, H. M., Kundaje, A. & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* 9, 9354, doi:10.1038/s41598-019-45839-z.
- 35 Gunhanlar, N. *et al.* (2018). A simplified protocol for differentiation of electrophysiologically mature neuronal networks from human induced pluripotent stem cells. *Mol Psychiatry* 23, 1336-1344, doi:10.1038/mp.2017.56.
- 36 Yuan, S. H. *et al.* (2011). Cell-surface marker signatures for the isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. *PLoS One* 6, e17540, doi:10.1371/journal.pone.0017540.

- 37 Schindelin, J. *et al.* (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676-682, doi:10.1038/nmeth.2019.
- 38 Wang, Y. *et al.* (2020). SynQuant: an automatic tool to quantify synapses from microscopy images. *Bioinformatics* 36, 1599-1606, doi:10.1093/bioinformatics/btz760.
- 39 Dobin, A. *et al.* (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21, doi:10.1093/bioinformatics/bts635.
- 40 Li, B. & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323, doi:10.1186/1471-2105-12-323.
- 41 Love, M. I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550, doi:10.1186/s13059-014-0550-8.
- 42 Zhou, Y. *et al.* (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10, 1523, doi:10.1038/s41467-019-09234-6.
- 43 Kang, H. J. *et al.* (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483-489, doi:10.1038/nature10523.
- 44 Baron, M. *et al.* (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* 3, 346-360 e344, doi:10.1016/j.cels.2016.08.011.
- 45 Newman, A. M. *et al.* (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12, 453-457, doi:10.1038/nmeth.3337.
- 46 Polioudakis, D. *et al.* (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* 103, 785-801 e788, doi:10.1016/j.neuron.2019.06.011.
- 47 Vernes, S. C. *et al.* (2006). Functional genetic analysis of mutations implicated in a human speech and language disorder. *Hum Mol Genet* 15, 3154-3167, doi:10.1093/hmg/ddl392.
- 48 Yates, A. *et al.* (2016). Ensembl 2016. *Nucleic Acids Res* 44, D710-716, doi:10.1093/nar/gkv1157.
- 49 Kuijk, E. *et al.* (2020). The mutational impact of culturing human pluripotent and adult stem cells. *Nat Commun* 11, 2493, doi:10.1038/s41467-020-16323-4.
- 50 Smith, R. H. *et al.* (2020). Genome-Wide Analysis of Off-Target CRISPR/Cas9 Activity in Single-Cell-Derived Human Hematopoietic Stem and Progenitor Cell Clones. *Genes (Basel)* 11, doi:10.3390/genes11121501.
- 51 Field, M. A. *et al.* (2019). Recurrent misscalling of missense variation from short-read genome sequence data. *BMC Genomics* 20, 546, doi:10.1186/s12864-019-5863-2.
- 52 Nowakowski, T. J., Pollen, A. A., Sandoval-Espinosa, C. & Kriegstein, A. R. (2016). Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. *Neuron* 91, 1219-1227, doi:10.1016/j.neuron.2016.09.005.
- 53 Clarke, L. E. & Barres, B. A. (2013). Emerging roles of astrocytes in neural circuit development. *Nat Rev Neurosci* 14, 311-321, doi:10.1038/nrn3484.
- 54 Li, M. *et al.* (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362, doi:10.1126/science.aat7615.
- 55 Johnson, M. B. *et al.* (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* 62, 494-509, doi:10.1016/j.neuron.2009.03.027.
- 56 Martin-de-Saavedra, M. D. *et al.* (2022). Shed CNTNAP2 ectodomain is detectable in CSF and regulates Ca(2+) homeostasis and network synchrony via PMCA2/ATP2B2. *Neuron* 110, 627-643.e629, doi:10.1016/j.neuron.2021.11.025.
- 57 Zhang, J. *et al.* (2023). CNTNAP2 intracellular domain (CICD) generated by -secretase cleavage improves autism-related behaviors. *Signal Transduct Target Ther* 8, 219, doi:10.1038/s41392-023-01431-6.
- 58 deAzevedo, L. C. *et al.* (2003). Cortical radial glial cells in human fetuses: depth-correlated transformation into astrocytes. *J Neurobiol* 55, 288-298, doi:10.1002/neu.10205.
- 59 Garcia-Gonzalo, F. R. *et al.* (2011). A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat Genet* 43, 776-784, doi:10.1038/ng.891.
- 60 Garcia-Gonzalo, F. R. & Reiter, J. F. (2017). Open Sesame: How Transition Fibers and the Transition Zone Control Ciliary Composition. *Cold Spring Harb Perspect Biol* 9, doi:10.1101/cshperspect.a028134.
- 61 Goncalves, J. & Pelletier, L. (2017). The Ciliary Transition Zone: Finding the Pieces and Assembling the Gate. *Mol Cells* 40, 243-253, doi:10.14348/molcells.2017.0054.
- 62 Gordon, A. & Geschwind, D. H. (2020). Human in vitro models for understanding mechanisms of autism spectrum disorder. *Mol Autism* 11, 26, doi:10.1186/s13229-020-00332-7.
- 63 Traka, M. *et al.* (2003). Association of TAG-1 with Caspr2 is essential for the molecular organization of juxtaparanodal regions of myelinated fibers. *J Cell Biol* 162, 1161-1172, doi:10.1083/jcb.200305078.
- 64 Cope, E. C. *et al.* (2016). Immature Neurons and Radial Glia, But Not Astrocytes or Microglia, Are Altered in Adult Cntnap2 and Shank3 Mice, Models of Autism. *eNeuro* 3, doi:10.1523/ENEURO.0196-16.2016.
- 65 Li, J. *et al.* (2019). Astrocyte-to-astrocyte contact and a positive feedback loop of growth factor signaling

- regulate astrocyte maturation. *Glia* 67, 1571-1597, doi:10.1002/glia.23630.
- 66 Anderson, G. R. *et al.* (2012). Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc Natl Acad Sci U S A* 109, 18120-18125, doi:10.1073/pnas.1216398109.
- 67 Antoine, M. W., Langberg, T., Schnepel, P. & Feldman, D. E. (2019). Increased Excitation-Inhibition Ratio Stabilizes Synapse and Circuit Excitability in Four Autism Mouse Models. *Neuron* 101, 648-661 e644, doi:10.1016/j.neuron.2018.12.026.
- 68 Chen, X. *et al.* (2011). PSD-95 is required to sustain the molecular organization of the postsynaptic density. *J Neurosci* 31, 6329-6338, doi:10.1523/JNEUROSCI.5968-10.2011.
- 69 Chen, X. *et al.* (2015). PSD-95 family MAGUKs are essential for anchoring AMPA and NMDA receptor complexes at the postsynaptic density. *Proc Natl Acad Sci U S A* 112, E6983-6992, doi:10.1073/pnas.1517045112.
- 70 Kim, J. W. *et al.* (2019). Pharmacological modulation of AMPA receptor rescues social impairments in animal models of autism. *Neuropsychopharmacology* 44, 314-323, doi:10.1038/s41386-018-0098-5.
- 71 Gumez-Gamboa, A., Coufal, N. G. & Gleeson, J. G. (2014). Primary cilia in the developing and mature brain. *Neuron* 82, 511-521, doi:10.1016/j.neuron.2014.04.024.
- 72 Sang, L. *et al.* (2011). Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell* 145, 513-528, doi:10.1016/j.cell.2011.04.019.
- 73 Shi, X. *et al.* (2017). Super-resolution microscopy reveals that disruption of ciliary transition-zone architecture causes Joubert syndrome. *Nat Cell Biol* 19, 1178-1188, doi:10.1038/ncb3599.
- 74 Higginbotham, H. *et al.* (2012). Arl13b in primary cilia regulates the migration and placement of interneurons in the developing cerebral cortex. *Dev Cell* 23, 925-938, doi:10.1016/j.devcel.2012.09.019.
- 75 Higginbotham, H. *et al.* (2013). Arl13b-regulated cilia activities are essential for polarized radial glial scaffold formation. *Nat Neurosci* 16, 1000-1007, doi:10.1038/nn.3451.
- 76 Guo, J. *et al.* (2015). Developmental disruptions underlying brain abnormalities in ciliopathies. *Nat Commun* 6, 7857, doi:10.1038/ncomms8857.
- 77 Park, S. M. *et al.* (2018). Brain Somatic Mutations in MTOR Disrupt Neuronal Ciliogenesis, Leading to Focal Cortical Dyslamination. *Neuron* 99, 83-97 e87, doi:10.1016/j.neuron.2018.05.039.
- 78 Smogavec, M. *et al.* (2016). Eight further individuals with intellectual disability and epilepsy carrying bi-allelic CNTNAP2 aberrations allow delineation of the mutational and phenotypic spectrum. *J Med Genet* 53, 820-827, doi:10.1136/jmedgenet-2016-103880.
- 79 Zweier, C. *et al.* (2009). CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. *Am J Hum Genet* 85, 655-666, doi:10.1016/j.ajhg.2009.10.004.
- 80 Deemyad, T. *et al.* (2022). Lateralized Decrease of Parvalbumin+ Cells in the Somatosensory Cortex of ASD Models Is Correlated with Unilateral Tactile Hypersensitivity. *Cereb Cortex* 32, 554-568, doi:10.1093/cercor/bhab233.
- 81 Toma, C. *et al.* (2018). Comprehensive cross-disorder analyses of CNTNAP2 suggest it is unlikely to be a primary risk gene for psychiatric disorders. *PLoS Genet* 14, e1007535, doi:10.1371/journal.pgen.1007535.
- 82 Xiang, Y. *et al.* (2017). Fusion of Regionally Specified hPSC-Derived Organoids Models Human Brain Development and Interneuron Migration. *Cell Stem Cell* 21, 383-398 e387, doi:10.1016/j.stem.2017.07.007.
- 83 Qian, X. *et al.* (2020). Sliced Human Cortical Organoids for Modeling Distinct Cortical Layer Formation. *Cell Stem Cell* 26, 766-781 e769, doi:10.1016/j.stem.2020.02.002.

Supplementary material

Figure S1. CRISPR/Cas9 editing efficiency.

Figure S2. Overlaps among genes with predicted high impact mutations.

Figure S3. Protein expression of CASPR2 in 8-week neuronal networks and NPCs.

Figure S4. Deconvolution of bulk RNA-sequencing with single cell data.

Table S1. High impact *de novo* indels, SNVs and SVs in WT, HET, HOM and DUAL hESCs.

Table S2. Potential off-target sites for CRISPR gRNAs.

Table S3. Candidate protein interaction partners of CASPR2.

Table S4. Top 150 enriched Gene Ontology categories for DEGs.

Table S5. Antibodies and dilutions for FACS and immunocytochemistry.

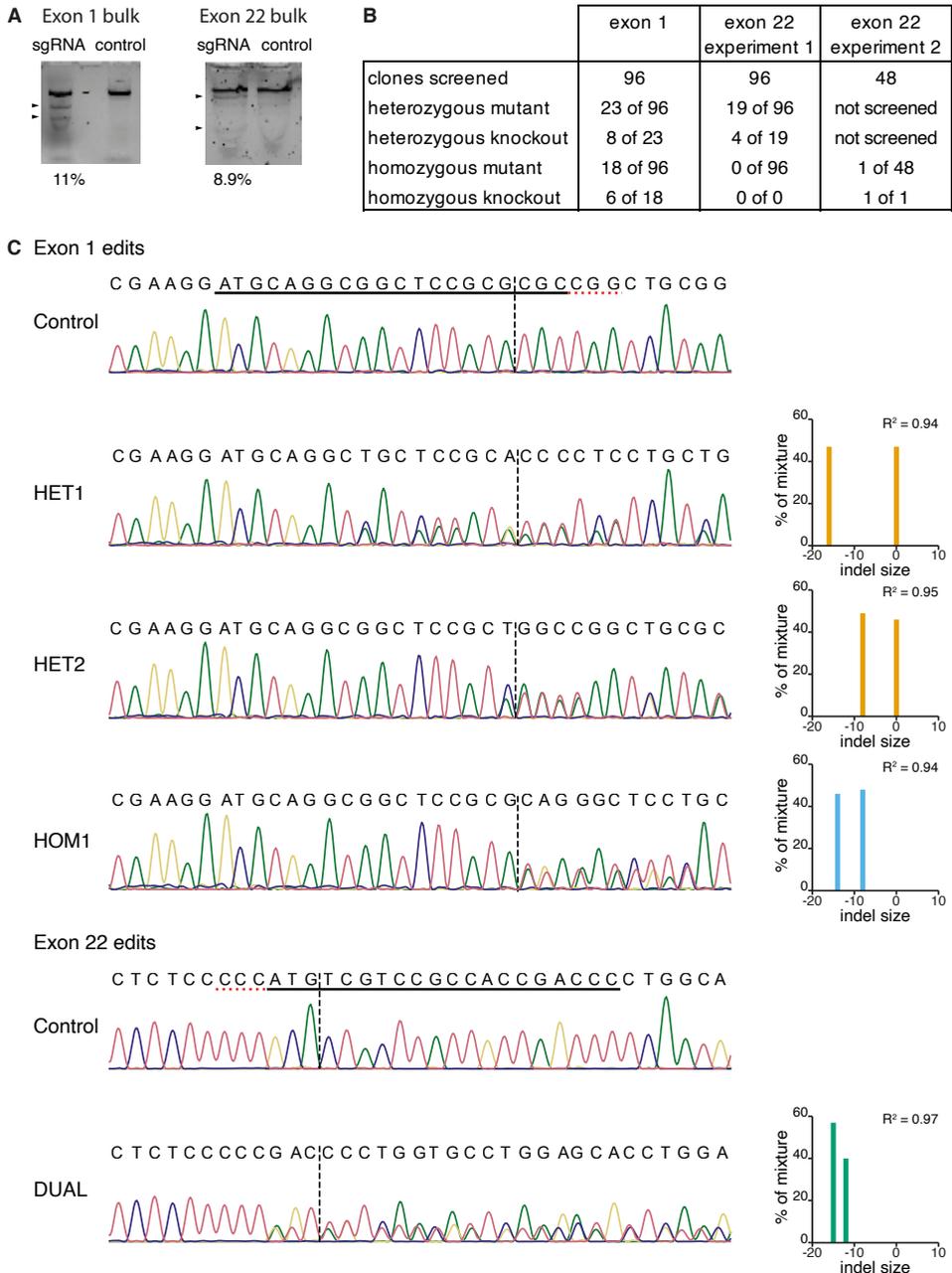


Figure S1 CRISPR/Cas9 editing efficiency. (A) T7E1 assay showing editing efficiency of sgRNAs on ESC bulk gDNA. (B) Results of clone screening by Sanger sequencing. Mutants are all clones with indels, knockouts are only those indels that cause frameshifts or disrupt a start codon. (C) Sanger sequencing traces of selected clones. Inferred indel size and frequencies are plotted on the side.

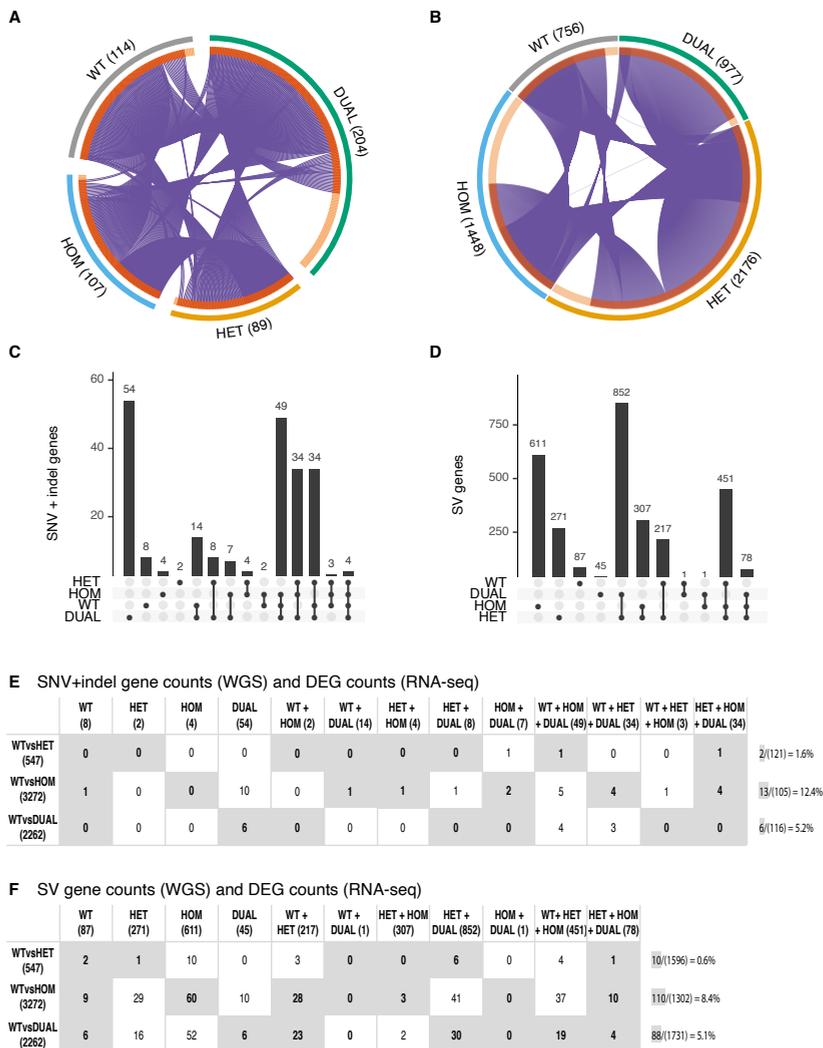
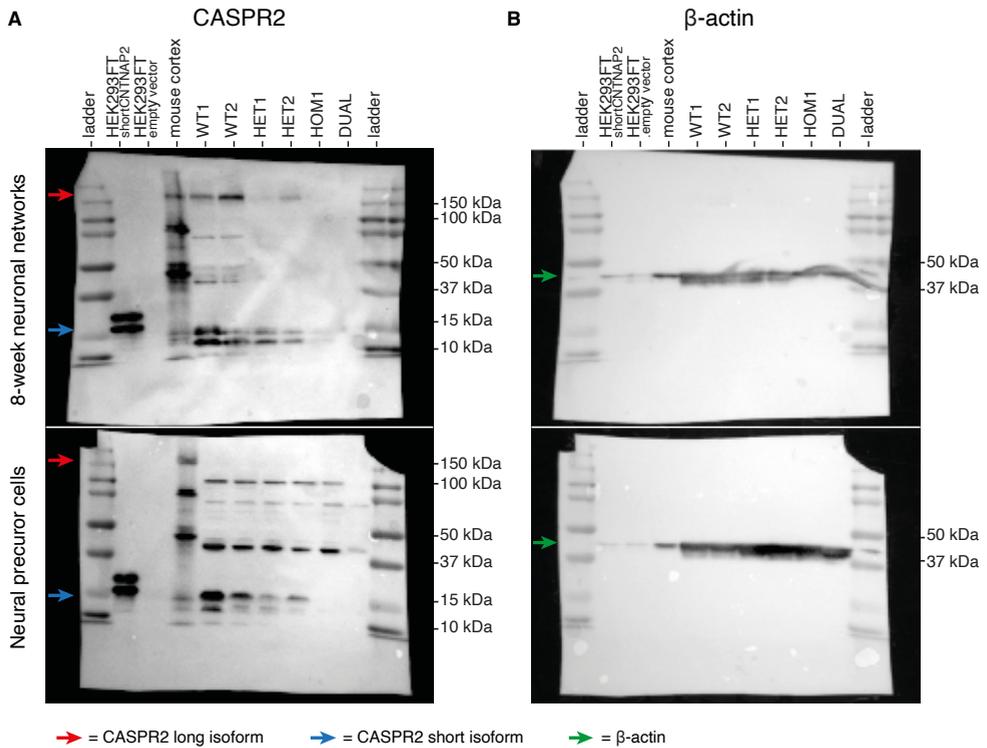


Figure S2 Overlaps among genes with predicted high impact mutations. (A + B) Circos plots showing overall overlap in genes with high impact SNVs or indels (A) and SVs (B) between sequenced ESC lines. Purple curves link identical genes between lists. Genes that hit multiple lists are coloured in dark orange, and genes unique to a list are shown in light orange. (C + D) Upset plots showing detailed overlap in genes with high impact SNVs or indels (C) and SVs (D) between sequenced ESC lines. Intersections between sets are organised in a matrix layout and the number of genes per intersection is plotted on the y-axis. Intersections with 0 genes are omitted. (E + F) Tables showing overlap between genes with high impact SNVs or indels (E) and SVs (F) and differentially expressed genes (DEG) for each WT-mutant comparison at 8 weeks of neural differentiation. Genes with predicted high impact variants are sorted in columns based on sets from C and D and intersected with differentially expressed genes in each row per WT-mutant comparison. Numbers between brackets indicate total number of genes per column or row. Numbers in squares are the shared genes between WGS calls and DGE hits. Squares in grey highlight potential confounding mutations per WT-mutant comparison. Percentages indicate DGE hits per total number of potentially confounding mutated genes.



C Prediction of post-translational modification on CASPR2 short isoform (MuSiteDeep)

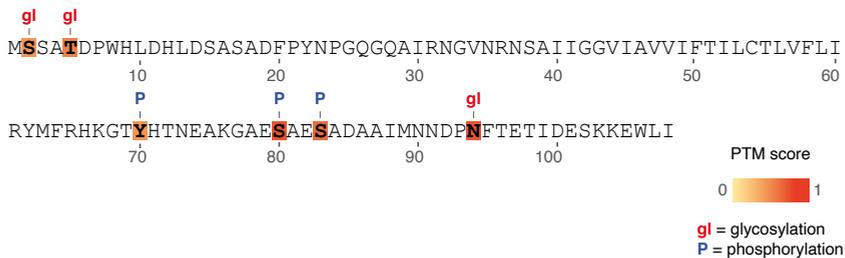


Figure S3 Protein expression of CASPR2 in 8-week neuronal networks and NPCs. (A) Western blot of CASPR2 expression in 8-week neuronal networks and NPCs. Red arrow indicates long isoform band around 150kDa. The intensity of the long isoform is decreased in HET 8-week neural networks and absent in HOM and DUAL samples. The long isoform is not detected in NPCs. Blue arrow indicates short isoform around 15 kDa. The isoform is expressed as a double band and is absent in the DUAL sample. HEK293FT cells transfected with a vector overexpressing the short isoform of CASPR2 are used as a positive control. The bands run slightly higher, possibly due to post-translational modifications (see panel C). Extra bands between the short and long isoform are likely the result of the polyclonal antibody reacting to non-specific epitopes. (B) Western blot of beta-actin.

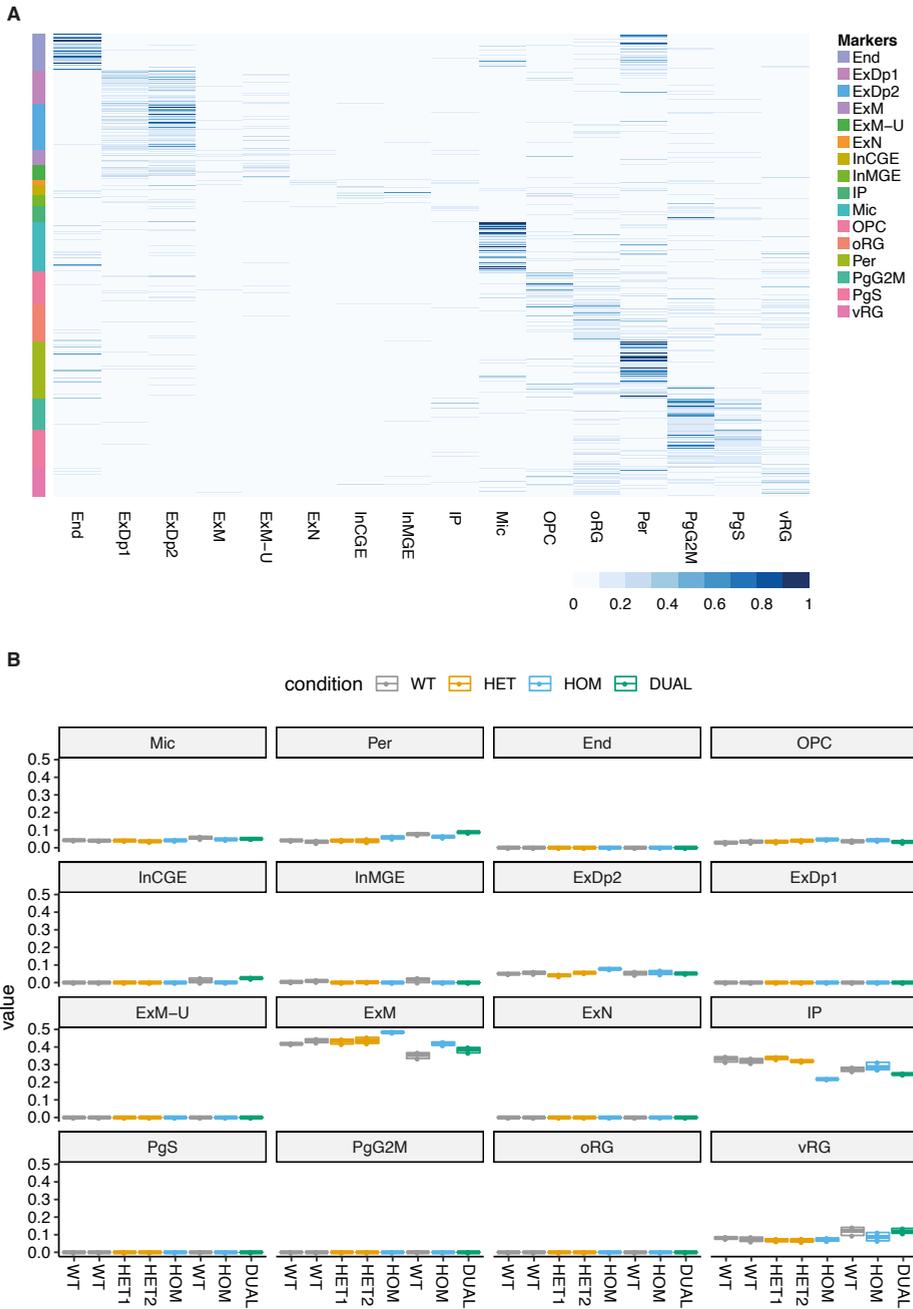


Figure S4 Deconvolution of bulk RNA-sequencing with single cell data. (A) Basis matrix of gene expression reference profiles from human neocortical cell type derived from single cell RNA-seq data (Polioudakis et al. 2019). (B) Relative fraction of all cell types in single cell data set for each sample.

Table S1 High impact de novo indels, SNVs and SVs in WT, HET, HOM and DUAL hESCs.

Not printed due to size. Full table is available at: <https://owncloud.gwdg.de/index.php/s/M5Ebp1h1tL5MGpn>

High impact de novo indels (358 rows) and SNVs (249 rows)

Type	Line	Chr	Position	REF	ALT	Gene	Effect

High impact de novo SVs (52 rows)

SV type	Line	SV chrom	SV start	SV end	SV length	AnnotSV. ranking	Genes

Table S2 Potential off-target sites for CRISPR gRNAs.

Not printed due to size. Full table is available at: <https://owncloud.gwdg.de/index.php/s/M5Ebp1h1tL5MGpn>

Predicted targets for Exon 1 gRNA 1(ATGCAGGCGGCTCCGCGCCGG) (144 rows)

prediction tool	chr	strand	position	sequence	# mismatches	Offtarget Score	gene/ locusDesc

Predicted targets for Exon 22 gRNA 1(GGGTCGGTGCGGACGACATGGG) (85 rows)

prediction tool	chr	strand	position	sequence	# mismatches	Offtarget Score	gene/ locusDesc

Table S3 Candidate protein interaction partners of CASPR2.

SYMBOL	Publication	Experimental evidence code (BioGRID system)	Interaction detail	Putative isoform interaction
ADAM22	Chen (2015)	Affinity Capture-MS	C-terminal antibody	all isoforms
CKMT1B	Chen (2015)	Affinity Capture-MS	C-terminal antibody	all isoforms
KCNAB2	Chen (2015)	Affinity Capture-MS	C-terminal antibody	all isoforms
LGI1	Chen (2015)	Affinity Capture-MS	C-terminal antibody	all isoforms
MPP3	Chen (2015)	Affinity Capture-MS	C-terminal antibody	all isoforms
PSME3	Chen (2015)	Affinity Capture-MS	C-terminal antibody	all isoforms
TPI1	Chen (2015)	Affinity Capture-MS	C-terminal antibody	all isoforms
KRT222	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
KIAA1755	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
ZNF410	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
ATP1B1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
HTRA1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
FLNB	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
PRRG3	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
TCTN1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
ANOS1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
BCR	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
BPGM	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
C2CD5	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
CACYBP	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
CIRBP	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
CKAP5	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
CNOT7	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
CNRIP1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
COPS5	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
CUL1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
DUS4L	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
EPRS	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
ERAL1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
FARSB	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms

SYMBOL	Publication	Experimental evidence code (BioGRID system)	Interaction detail	Putative isoform interaction
HNRNPLL	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
HSPA5	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
IMMT	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
LYSMD2	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
MARCKSL1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
MPP6	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
N4BP2L2	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
PARD3	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
PCCA	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
PCNA	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
PGM1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
POLR2B	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
POMP	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
RPRD2	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
RPS20	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
RSL24D1	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
STX8	Gao (2018)	Two-hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
DLG4	Rasband (2002), Chen (2015)	Co-localization, Affinity Capture-MS	C-terminal antibody	all isoforms
CASK	Spiegel (2002), Gao (2018)	Reconstituted Complex, Two- hybrid	CNTNAP2 C-terminal region (aa 1284–1331)	all isoforms
ITPR1	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
SYN1	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
MAP1A	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
MAP1B	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
MAP2	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
MAP4	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
MYH10	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform

SYMBOL	Publication	Experimental evidence code (BioGRID system)	Interaction detail	Putative isoform interaction
TUBA1B	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
TUBB2A	Argent (2020)	Affinity Capture-MS	CNTNAP2-ECD coated beads	long isoform
LGI3	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
CLU	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
KCNA3	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
LGI4	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
ADAM11	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
ADAM23	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
DLG1	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
DLG2	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
KCNA1	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
KCNA4	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
KCNA6	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
KCNAB1	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
LGI2	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
MPP2	Chen (2015)	Affinity Capture-MS	C-terminal antibody	long isoform
KCNA2	Rasband (2002), Chen (2015)	Affinity Capture- Western, Affinity Capture-MS	C-terminal antibody	long isoform
CNTN1	Rubio- Marrero (2016)	Co-crystal structure	CNTNAP2-ECD	long isoform
MEOX2	Corominas (2014), Rolland (2014)	Two-hybrid		
EPB41L3	Denisenko- Nehrbass (2003)	Reconstituted Complex, Two- hybrid		

SYMBOL	Publication	Experimental evidence code (BioGRID system)	Interaction detail	Putative isoform interaction
CANX	Falivelli (2012)	Affinity Capture-Western		
GORASP2	Luck (2020)	Two-hybrid		
POLR2G	Luck (2020)	Two-hybrid		
CTR9	Nakayama (2002)	Two-hybrid		
MACF1	Nakayama (2002)	Two-hybrid		
ZMIZ1	Nakayama (2002)	Two-hybrid		
CPE	Oiso (2009)	Affinity Capture-Western, Reconstituted Complex, Two-Hybrid		
IQCB1	Sang (2011)	Affinity Capture-MS		
CNTN2	Traka (2003)	Affinity Capture-Western		

Table S4 Top 150 enriched Gene Ontology categories for DEGs.

Full table, including log p-value for enrichment per comparison, is available at:
<https://owncloud.gwdg.de/index.php/s/M5Ebp1h1tL5MGpn>

Category	GO	Description	#GeneInGO AndHitList	%InGO	Log (q-value)
GO Biological Processes	GO:0030198	extracellular matrix organization	159	8.36	-56.52
GO Biological Processes	GO:0043062	extracellular structure organization	170	8.94	-56.14
GO Biological Processes	GO:0001568	blood vessel development	215	11.30	-54.18
Reactome Gene Sets	R-HSA-1474244	Extracellular matrix organization	132	6.94	-52.80
GO Biological Processes	GO:0048514	blood vessel morphogenesis	190	9.99	-47.98
GO Biological Processes	GO:0048729	tissue morphogenesis	195	10.25	-45.36
GO Biological Processes	GO:0048598	embryonic morphogenesis	181	9.52	-44.47
GO Biological Processes	GO:0035082	axoneme assembly	50	4.51	-42.59
GO Biological Processes	GO:0044782	cilium organization	116	10.46	-42.37
GO Biological Processes	GO:0003341	cilium movement	50	4.51	-41.41
GO Biological Processes	GO:0001525	angiogenesis	161	8.46	-40.37
GO Biological Processes	GO:0001501	skeletal system development	156	8.20	-38.53
Reactome Gene Sets	R-HSA-500792	GPCR ligand binding	132	6.94	-38.29
GO Biological Processes	GO:0048568	embryonic organ development	143	7.52	-37.40
GO Biological Processes	GO:0060271	cilium assembly	106	9.56	-36.73
GO Biological Processes	GO:0001578	microtubule bundle formation	54	4.87	-35.70
GO Biological Processes	GO:0003013	circulatory system process	157	8.25	-35.89
GO Biological Processes	GO:0006935	chemotaxis	167	8.78	-35.15
GO Biological Processes	GO:0042330	taxis	167	8.78	-34.96
GO Biological Processes	GO:0002009	morphogenesis of an epithelium	159	8.36	-34.64
GO Biological Processes	GO:0008015	blood circulation	152	7.99	-34.07
GO Biological Processes	GO:0007423	sensory organ development	163	8.57	-33.97
GO Biological Processes	GO:0008285	negative regulation of cell proliferation	178	9.36	-33.21
GO Biological Processes	GO:0030155	regulation of cell adhesion	178	9.36	-32.84
GO Biological Processes	GO:0009611	response to wounding	172	9.04	-32.56
GO Biological Processes	GO:0042060	wound healing	153	8.04	-32.49
GO Biological Processes	GO:0040017	positive regulation of locomotion	157	8.25	-32.48
GO Biological Processes	GO:2000147	positive regulation of cell motility	152	7.99	-32.38

GO Biological Processes	GO:0030335	positive regulation of cell migration	148	7.78	-32.23
GO Biological Processes	GO:0001655	urogenital system development	114	5.99	-31.91
GO Biological Processes	GO:0050673	epithelial cell proliferation	128	6.73	-31.83
GO Biological Processes	GO:0070848	response to growth factor	183	9.62	-31.75
GO Biological Processes	GO:0071363	cellular response to growth factor stimulus	178	9.36	-31.48
GO Biological Processes	GO:0051272	positive regulation of cellular component movement	153	8.04	-31.31
GO Biological Processes	GO:0050900	leukocyte migration	120	6.31	-30.15
GO Biological Processes	GO:0044057	regulation of system process	157	8.25	-29.70
GO Biological Processes	GO:0072001	renal system development	103	5.42	-29.59
GO Biological Processes	GO:0001822	kidney development	100	5.26	-28.82
GO Biological Processes	GO:0048562	embryonic organ morphogenesis	100	5.26	-28.36
GO Biological Processes	GO:0030031	cell projection assembly	119	10.73	-27.65
KEGG Pathway	hsa04060	Cytokine-cytokine receptor interaction	86	4.52	-27.81
GO Biological Processes	GO:0007018	microtubule-based movement	78	7.03	-27.19
GO Biological Processes	GO:0120031	plasma membrane bounded cell projection assembly	116	10.46	-26.98
GO Biological Processes	GO:0007507	heart development	148	7.78	-26.64
GO Biological Processes	GO:0050678	regulation of epithelial cell proliferation	109	5.73	-26.64
GO Biological Processes	GO:1901342	regulation of vasculature development	106	5.57	-25.86
Reactome Gene Sets	R-HSA-216083	Integrin cell surface interactions	49	2.58	-25.79
GO Biological Processes	GO:0045596	negative regulation of cell differentiation	168	8.83	-25.59
Reactome Gene Sets	R-HSA-1474228	Degradation of the extracellular matrix	63	3.31	-25.58
GO Biological Processes	GO:0043269	regulation of ion transport	162	8.52	-24.78
Reactome Gene Sets	R-HSA-1474290	Collagen formation	51	2.68	-24.57
Reactome Gene Sets	R-HSA-373076	Class A/1 (Rhodopsin-like receptors)	89	4.68	-24.53
GO Biological Processes	GO:0060485	mesenchyme development	94	4.94	-24.53
KEGG Pathway	hsa05200	Pathways in cancer	141	7.41	-24.47
GO Biological Processes	GO:0043408	regulation of MAPK cascade	164	8.62	-24.19
GO Biological Processes	GO:0070286	axonemal dynein complex assembly	26	2.34	-23.51
GO Biological Processes	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	167	8.78	-23.53
GO Biological Processes	GO:0043410	positive regulation of MAPK cascade	133	6.99	-23.48
GO Biological Processes	GO:0007389	pattern specification process	119	6.26	-23.44
GO Biological Processes	GO:0048608	reproductive structure development	118	6.20	-23.40

GO Biological Processes	GO:0031589	cell-substrate adhesion	106	5.57	-23.38
GO Biological Processes	GO:0050878	regulation of body fluid levels	125	6.57	-23.19
GO Biological Processes	GO:0061458	reproductive system development	118	6.20	-23.00
GO Biological Processes	GO:0001667	ameboidal-type cell migration	115	6.05	-22.97
GO Biological Processes	GO:0001101	response to acid chemical	102	5.36	-22.95
KEGG Pathway	hsa04080	Neuroactive ligand-receptor interaction	100	5.26	-22.92
GO Biological Processes	GO:0048732	gland development	120	6.31	-22.84
Reactome Gene Sets	R-HSA-2022090	Assembly of collagen fibrils and other multimeric structures	40	2.10	-22.75
GO Biological Processes	GO:0045765	regulation of angiogenesis	94	4.94	-22.62
GO Biological Processes	GO:0045785	positive regulation of cell adhesion	111	5.84	-22.39
GO Biological Processes	GO:0009617	response to bacterium	127	6.68	-22.05
GO Biological Processes	GO:0090596	sensory organ morphogenesis	87	4.57	-21.99
KEGG Pathway	hsa04151	PI3K-Akt signaling pathway	102	5.36	-21.83
GO Biological Processes	GO:0048762	mesenchymal cell differentiation	79	4.15	-21.63
GO Biological Processes	GO:0019932	second-messenger-mediated signaling	113	5.94	-21.48
GO Biological Processes	GO:0060562	epithelial tube morphogenesis	97	5.10	-21.47
GO Biological Processes	GO:0001503	ossification	110	5.78	-21.34
GO Biological Processes	GO:0051046	regulation of secretion	157	8.25	-21.28
GO Biological Processes	GO:0043009	chordate embryonic development	149	7.83	-21.17
KEGG Pathway	ko04060	Cytokine-cytokine receptor interaction	68	3.58	-21.12
GO Biological Processes	GO:0008217	regulation of blood pressure	66	3.47	-20.99
GO Biological Processes	GO:0003007	heart morphogenesis	82	4.31	-20.75
GO Biological Processes	GO:0007610	behavior	143	7.52	-20.75
GO Biological Processes	GO:0003006	developmental process involved in reproduction	148	7.78	-20.73
GO Biological Processes	GO:0045165	cell fate commitment	83	4.36	-20.61
GO Biological Processes	GO:0009792	embryo development ending in birth or egg hatching	150	7.89	-20.54
GO Biological Processes	GO:0048880	sensory system development	107	5.63	-20.41
Reactome Gene Sets	R-HSA-3000171	Non-integrin membrane-ECM interactions	38	2.00	-20.41
GO Biological Processes	GO:0007187	G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	77	4.05	-20.32
Reactome Gene Sets	R-HSA-1442490	Collagen degradation	38	2.00	-20.03
GO Biological Processes	GO:0150063	visual system development	105	5.52	-19.92

KEGG Pathway	ko04151	PI3K-Akt signaling pathway	93	4.89	-19.81
GO Biological Processes	GO:0099536	synaptic signaling	112	10.43	-19.05
Reactome Gene Sets	R-HSA-1650814	Collagen biosynthesis and modifying enzymes	40	2.10	-19.80
GO Biological Processes	GO:0010817	regulation of hormone levels	126	6.62	-19.72
GO Biological Processes	GO:0099537	trans-synaptic signaling	111	10.34	-18.97
GO Biological Processes	GO:0001763	morphogenesis of a branching structure	69	3.63	-19.63
GO Biological Processes	GO:0001654	eye development	104	5.47	-19.58
GO Biological Processes	GO:0061061	muscle structure development	145	7.62	-19.55
Reactome Gene Sets	R-HSA-8948216	Collagen chain trimerization	32	1.68	-19.49
GO Biological Processes	GO:0048771	tissue remodeling	63	3.31	-19.45
GO Biological Processes	GO:0007268	chemical synaptic transmission	109	10.15	-18.42
GO Biological Processes	GO:0098916	anterograde trans-synaptic signaling	109	10.15	-18.42
GO Biological Processes	GO:0072006	nephron development	57	3.00	-19.09
GO Biological Processes	GO:0071396	cellular response to lipid	128	6.73	-19.03
GO Biological Processes	GO:0043583	ear development	74	3.89	-18.90
GO Biological Processes	GO:0045055	regulated exocytosis	159	8.36	-18.89
GO Biological Processes	GO:0046649	lymphocyte activation	137	7.20	-18.88
GO Biological Processes	GO:0050865	regulation of cell activation	122	6.41	-18.88
GO Biological Processes	GO:0050679	positive regulation of epithelial cell proliferation	67	3.52	-18.87
GO Biological Processes	GO:0061138	morphogenesis of a branching epithelium	65	3.42	-18.84
KEGG Pathway	hsa04510	Focal adhesion	73	3.84	-18.81
GO Biological Processes	GO:0043010	camera-type eye development	93	4.89	-18.80
GO Biological Processes	GO:1904018	positive regulation of vasculature development	66	3.47	-18.78
GO Biological Processes	GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	96	5.05	-18.73
Reactome Gene Sets	R-HSA-3000178	ECM proteoglycans	41	2.16	-18.72
GO Biological Processes	GO:0003012	muscle system process	109	5.73	-18.65
GO Biological Processes	GO:0051271	negative regulation of cellular component movement	89	4.68	-18.55
KEGG Pathway	ko04080	Neuroactive ligand-receptor interaction	78	4.10	-18.54
GO Biological Processes	GO:0003002	regionalization	93	4.89	-18.40
GO Biological Processes	GO:0007420	brain development	162	8.52	-18.34
GO Biological Processes	GO:0001819	positive regulation of cytokine production	104	5.47	-18.33
Reactome Gene Sets	R-HSA-109582	Hemostasis	131	6.89	-18.27

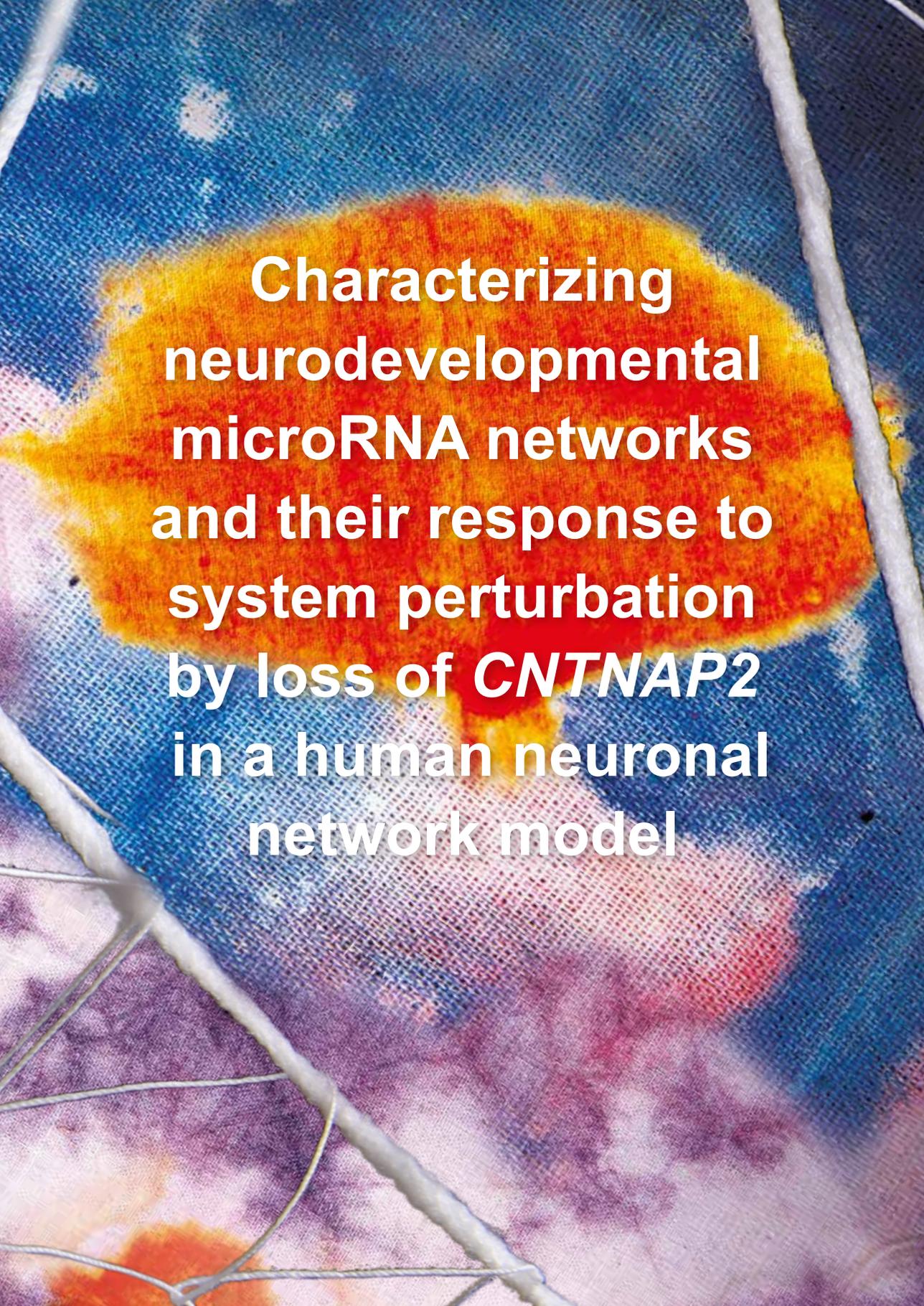
KEGG Pathway	ko04510	Focal adhesion	69	3.63	-18.24
GO Biological Processes	GO:0061448	connective tissue development	79	4.15	-18.07
GO Biological Processes	GO:1903522	regulation of blood circulation	84	4.42	-18.07
KEGG Pathway	ko04512	ECM-receptor interaction	42	2.21	-18.05
GO Biological Processes	GO:0010942	positive regulation of cell death	149	7.83	-18.02
KEGG Pathway	hsa04512	ECM-receptor interaction	43	2.26	-17.96
GO Biological Processes	GO:0045766	positive regulation of angiogenesis	60	3.15	-17.80
GO Biological Processes	GO:0090130	tissue migration	86	4.52	-17.70
GO Biological Processes	GO:0051216	cartilage development	65	3.42	-17.67
GO Biological Processes	GO:0090132	epithelium migration	85	4.47	-17.67
GO Biological Processes	GO:0010035	response to inorganic substance	125	6.57	-17.63
GO Biological Processes	GO:0000904	cell morphogenesis involved in differentiation	160	8.41	-17.44
KEGG Pathway	hsa05165	human papillomavirus infection	93	4.89	-17.43
GO Biological Processes	GO:0048545	response to steroid hormone	91	4.78	-17.22
GO Biological Processes	GO:0060537	muscle tissue development	100	5.26	-17.19
GO Biological Processes	GO:0007188	adenylate cyclase-modulating G protein-coupled receptor signaling pathway	69	3.63	-17.17
GO Biological Processes	GO:2000027	regulation of animal organ morphogenesis	76	4.00	-17.09
GO Biological Processes	GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	75	3.94	-17.07
GO Biological Processes	GO:0009636	response to toxic substance	116	6.10	-17.05
GO Biological Processes	GO:0090287	regulation of cellular response to growth factor stimulus	82	4.31	-17.04
GO Biological Processes	GO:0002521	leukocyte differentiation	112	5.89	-16.98
GO Biological Processes	GO:2000146	negative regulation of cell motility	80	4.21	-16.97
GO Biological Processes	GO:0006936	muscle contraction	92	4.84	-16.90
GO Biological Processes	GO:0010631	epithelial cell migration	83	4.36	-16.87
GO Biological Processes	GO:0006816	calcium ion transport	100	5.26	-16.86
GO Biological Processes	GO:0001704	formation of primary germ layer	49	2.58	-16.79
GO Biological Processes	GO:0022407	regulation of cell-cell adhesion	98	5.15	-16.64

Table S5 Antibodies and dilutions for FACS and immunocytochemistry.

Application	Name	Clone	Cat. No.	Host species	Concentration (v/v)
FACS	FITC Mouse Anti-Human CD44	G44-26	BD Biosciences, 560977	Mouse	1 in 20
FACS	APC Mouse Anti-Human CD184	ML5	BD Biosciences, 560936	Mouse	1 in 10
FACS	PE Mouse Anti-Human CD271	C40-1457	BD Biosciences, 560927	Mouse	1 in 10
FACS	PE-Cy7 Mouse Anti-Human	12G5	BD Biosciences, 561646	Mouse	1 in 40
ICC	Syn1/2		Synaptic Systems, 106 004	Guinea pig	1 in 1000
ICC	Gephyrin	mAb7a	Synaptic Systems, 147011	Mouse	1 in 250
ICC	PSD-95	K28/43	NeuroMab, 75-028	Mouse	1 in 250
ICC	SOX2		Millipore, AB5603	Rabbit	1 in 1000
ICC	NeuN	1B7	Abcam, ab104224	Mouse	1 in 1000
ICC	GFAP		EnCor, CPCA-GFAP	Chicken	1 in 1000
ICC	MAP2		EnCor, GPCA-MAP2A/B	Guinea pig	1 in 1000
ICC	MAP2		EnCor, CPCA-MAP2	Chicken	1 in 1000

4



The background is an abstract, textured composition. It features a large, irregular shape in the center, filled with a gradient of colors from yellow to red. This shape is set against a blue background with a fine, grid-like texture. Several white, irregular lines and shapes are scattered across the image, some resembling veins or cracks. The overall appearance is that of a complex, multi-layered graphic design.

**Characterizing
neurodevelopmental
microRNA networks
and their response to
system perturbation
by loss of *CNTNAP2*
in a human neuronal
network model**

CHAPTER 4

ABSTRACT

MicroRNAs (miRNAs) are key regulators of several neurodevelopmental processes, such as cell fate determination, migration, neurite outgrowth and synapse development. As a result, miRNA variants and expression changes have been implicated in the pathophysiology of several neurodevelopmental disorders. Additionally, miRNAs can function to stabilise relevant molecular networks against perturbations by potential causes of developmental disorders. To validate the pathophysiological and stabilising effects of miRNA networks, there is a need for human experimental models. Here, the expression of miRNAs is measured during the development of human embryonic stem cell-derived neuronal networks. The data show that miRNA networks respond to neurodevelopmental perturbation caused by loss of *CNTNAP2*, a gene involved in neurodevelopmental disorders. MiRNAs cluster in developmental expression profiles that regulate expression of cortical biological processes, such as morphogenesis, cell-cell contact, neuronal system development and cell projection organisation. Loss of *CNTNAP2* leads to widespread miRNA expression changes that are partially explained by changes in cell type composition of the cultures, but that also seem to be directed at mitigating deficits in neurite outgrowth. These results demonstrate that miRNAs can play a role in modulating disease risk. Future studies may search for compound mutations in candidate miRNAs and their targets to explain clinical variation in patients with neurodevelopmental disorders. Human cell models will be well positioned to validate the effects of these mutations and dive deeper into the structure of miRNA networks to identify more candidate loci.

Characterizing neurodevelopmental microRNA networks and their response to system perturbation by loss of *CNTNAP2* in a human neuronal network model

4.1 Introduction

MicroRNAs (miRNAs) are important regulators of gene expression. MiRNAs are short non-coding RNA molecules of 21-24 nucleotides in length that target the transcripts of multiple target genes via complementary binding to a core motif of 7-8 base pairs, typically in the 3'UTR, resulting in degradation of the transcript or blockade of translation¹. MiRNAs display complex spatio-temporal expression patterns in the mammalian brain and have the potential to regulate thousands of target genes^{2,3}. As such, miRNAs have emerged as key regulators of several neurodevelopmental processes, such as cell fate determination, migration, neurite outgrowth and synapse development⁴. This indicates that miRNA dysregulation could contribute to pathophysiological changes in neurodevelopment. As members of regulatory networks miRNAs may also function to stabilise relevant molecular networks against perturbations by potential causes of developmental disorders^{5,6}.

MiRNAs have been implicated in several neurodevelopmental disorders by evidence from both genetic association and gene expression studies. For example, *MIR137HG*, the host gene for miR-137, coincides with a replicated, genome-wide significant risk locus for schizophrenia^{7,8}. The risk allele of the polymorphism at this site is associated with upregulation of miR-137 in human brain tissue^{9,10}. Disorder-related variants may also affect miRNA-target binding. Functional variants affecting miRNA binding sites have been identified in cohorts of children with neurodevelopmental disorders including Tourette syndrome¹¹, developmental language disorder¹² and intellectual disability¹³. Patient-control studies have detected significant miRNA expression differences in post-mortem brains for disorders such as autism, schizophrenia and bipolar disorder^{14,15}. However, these studies can only demonstrate a correlation between miRNAs and disease. It is unclear whether these expression differences reflect a cause or a consequence of the disorder. To validate the pathophysiological effects of miRNAs, there is a need for human experimental models.

Human cell models offer a valuable platform to study the functions of miRNAs due to their amenability to molecular perturbations and capability to recapitulate species-specific miRNA biology. At least 14 miRNAs have been identified that are specific to the human genome and more than 100 miRNAs are not conserved beyond primates^{16,17}. In addition, miRNA expression patterns are highly divergent between species. A study of human, chimpanzee and macaque brains found that 40.6% of miRNAs were differentially expressed between at least two species in one or more brain regions¹⁸. Further, the developmental expression trajectories of miRNAs and their target genes underwent evolutionary changes resulting in new configurations of gene regulatory networks that may have contributed to the rapid evolution of the human brain¹⁹. Human cell models recapitulate species-specific differences in cortical cell gene expression²⁰, and can help to

dissect the contribution of miRNA networks to neurodevelopmental disorders.

This study aims to characterize developmental changes in miRNA expression in an *in vitro* human neuronal network model and leverage this data to uncover roles of miRNA regulatory networks in response to neurodevelopmental perturbation caused by loss of *CNTNAP2*. First, the development of miRNA regulatory networks during neuronal differentiation is characterized via small RNA-sequencing and expression correlations with target mRNAs. Second, we explore how miRNA networks change in this *in vitro* model when neurodevelopment is perturbed by mutations in *CNTNAP2*. Homozygous mutations in *CNTNAP2* lead to a severe syndrome, called CASPR2 deficiency disorder, that is characterized by intellectual disability, epileptic seizures, language impairments and autistic features^{21,22}. Genetic experiments with *CNTNAP2* have highlighted roles in multiple neurodevelopmental processes, such as neurite outgrowth and synaptic connectivity²³⁻²⁷. Still, it is currently not clear how mutations in *CNTNAP2* lead to this broad set of clinical outcomes that varies between patients. Understanding perturbations of miRNA networks in this model offers further insight into the molecular networks that are related to *CNTNAP2* and offers genomic targets to allow further characterization of the clinical heterogeneity observed in patients.

4.2 Methods

4.2.1 Cell culture and CRISPR/Cas9 genome editing

The procedures for stem cell culture, neuronal differentiation and CRISPR/Cas9 genome editing are described in Chapter 3.

4.2.2 RNA-seq

Cells were lysed in RLT+ buffer from the RNeasy plus Mini kit (Qiagen). RNA was extracted using an adapted kit protocol for the extraction of total RNA including small RNAs. RNA is precipitated in 1.5 volumes of 100% ethanol instead of 1 volume of 70% ethanol. The quality and quantity of total input mRNA was determined on an Agilent BioAnalyzer 2100 using Agilent RNA 6000 Nano kit Samples meeting the quality criteria (≥ 200 ng total RNA, RIN ≥ 8 and 28S/18S ≥ 1) were shipped to the Beijing Genomics Institute (BGI) in dry ice for smallRNA-sequencing and mRNA-sequencing.

4.2.3 RNA-seq analysis

MiRNAs were aligned to miRbase using Chimira²⁸. mRNA-seq reads were mapped using STAR v2.7²⁹ to the library of human transcriptome sequences obtained from ENSEMBL93 (GRCh38.p12). Read counts per gene were obtained by quantmode in STAR.

4.2.3.1 Clustering on expression pattern with BioLayout

MiRNAs were clustered based on expression pattern using BioLayout version 3.4³⁰. Only miRNAs that were expressed with at least 10 reads in one sample were included in this analysis. Minimum Pearson correlation threshold was set at 0.85 to form a co-expression

network. The MCL algorithm was run with inflation at 2.2 to create the expression pattern clusters.

5.2.3.2 Correlation with BrainSpan

MiRNAs with at least 10 reads in 80% of BrainSpan samples were selected. Pairwise Spearman rank correlations were calculated between BrainSpan samples and experimental samples. Wilcoxon's rank-sum test was used to assess if a spatial category of interest had significantly higher Spearman correlations than the background of all pairwise correlations.

4.2.3.3 Differential gene expression analysis

DGE analysis for mRNA and miRNA data was performed using DESeq2. Differentially expressed genes were identified using DESeq2³¹.

4.2.3.5 Motif activity prediction with ISMARA

ISMARA was run with the option for miRNAs. It assumes that the signal at each promoter p is a linear function of its binding sites N_{pm} , whereas c is a term reflecting the average activity of promoter p across the samples, c_s reflects the total expression in sample s and the A_{ms} are the (unknown) activities of each motif m in each samples³².

$$E_{ps} = \sum_m N_{pm} \times A_{ms} + c_p + \bar{c}_s$$

MiRNA sites were annotated using TargetScan 7 using preferential conservation scoring (PCT), which assigns target scores for 86 conserved miRNA seed families to all RefSeq transcripts^{33,34}. Target score for each promoter was calculated by averaging all the target scores over the transcripts associated with the promoter, so gene-level counts could be used for the model.

Motifs with a Z-value > 1.96 were considered as significantly active and their activity scores were Spearman rank correlated with the expression data of associated miRs. Targets for miRNAs with a correlation p-value < 0.05 were selected for pathway analysis and interactome meta-analysis with MetaScape. Pairwise overlaps between targets were calculated with BioConductor package GeneOverlap (v1.32.2). Differential activity for WTvsHOM and WTvsDUAL was measured with a Student's t-test.

4.2.3.6 Pathway analysis and interactome meta-analysis

Pathway analysis was performed on target lists using MetaScape with default parameters³⁵. For WT-mutant, target lists were filtered for differentially expressed genes. An interactome meta-analysis was run to detect biochemical complexes and signal transduction components with higher specificity.

4.3 Results

4.3.1 Developmental miRNA expression changes in an *in vitro* model of neuronal differentiation

First, the global expression of miRNAs over four stages of a neuronal network differentiation protocol was explored with small RNA-sequencing. The human neuronal model involves differentiation of embryonic stem cells (ESCs) to neural precursor cells (NPCs), which then give rise to neurons and astrocytes in neuronal networks that are developed in an 8-week process of neural differentiation³⁶. Samples from ESCs, NPCs, and neuronal networks of 4 weeks and 8 weeks were sequenced. When looking at differences in miRNA expression across these stages, it was apparent that the largest expression changes occur between ESCs and NPCs, and again between NPCs and differentiated neurons (Fig. 1A). By contrast, the 4-week and 8-week old neuronal networks clustered together, indicating their highly similar miRNA expression profiles. These contrasting expression profiles show that different miRNA networks are active at each stage of the human neuronal model.

MiRNAs control several processes during neurodevelopment that happen in a specific sequence and at specific time points^{4,39}. Therefore, we set out to annotate all expressed miRNAs based on their developmental expression patterns, which indicate at which stages these molecules can be active. A Markov clustering algorithm was applied in order to group 1197 miRNAs based on their expression patterns across model stages³⁰. The analysis identified 11 pattern clusters that demonstrate distinct patterns over neuronal differentiation (Fig. 1B). These clusters show overlap with known sets of miRNAs that play roles in different stages of neurodevelopment^{4,37,38} (Fig. 1C). Cluster01 (280 miRNAs) showed an upward trajectory through neuronal differentiation and contains highly expressed miRNAs that are important throughout neuronal development, such as let-7, miR-9, and miR-137⁴. Similarly, cluster04 (135 miRNAs) peaked in differentiated neuronal networks, and highly expressed miRNAs in this cluster are involved with the development of mature neurons; miR-132 and miR-134 regulate synapse function and miR-218 is involved in neuron fate specification⁴. Cluster03 (194 miRNAs) and cluster05 (106 miRNAs) had peaks earlier in development of the neuronal model, and I saw that miRNAs in this cluster also regulate processes early in development. From cluster03, both miR-92b and miR-219 are known to regulate the differentiation of intermediate progenitors and glial cells, respectively³⁸. Cluster05 miRNAs showed highest expression in early stages, and several miRNAs of the miR-17/92 cluster were present in cluster05 and are known to regulate a wide spectrum of early developmental processes⁴⁰. Other miRNAs of the miR-17/92 cluster were in cluster02 (278 miRNAs) that also contains several highly expressed miRNAs that are specific to embryonic stem cells³⁷. Thus, these expression cluster annotations provide a data-driven functional categorization of miRNAs in the human neuronal network model.

MiRNA expression at the stages of this neuronal model were further compared to miRNA expression data from post-mortem human brains in BrainSpan. The BrainSpan samples were all collected postnatally, and therefore all brain regions were analysed as one time point. The most significant expression correlation was found with cortex, a correlation that strongly increased during the maturation of the neuronal model (Fig. 1D). The observed cortical correlation suggests that patterns of miRNA expression are region-specific in

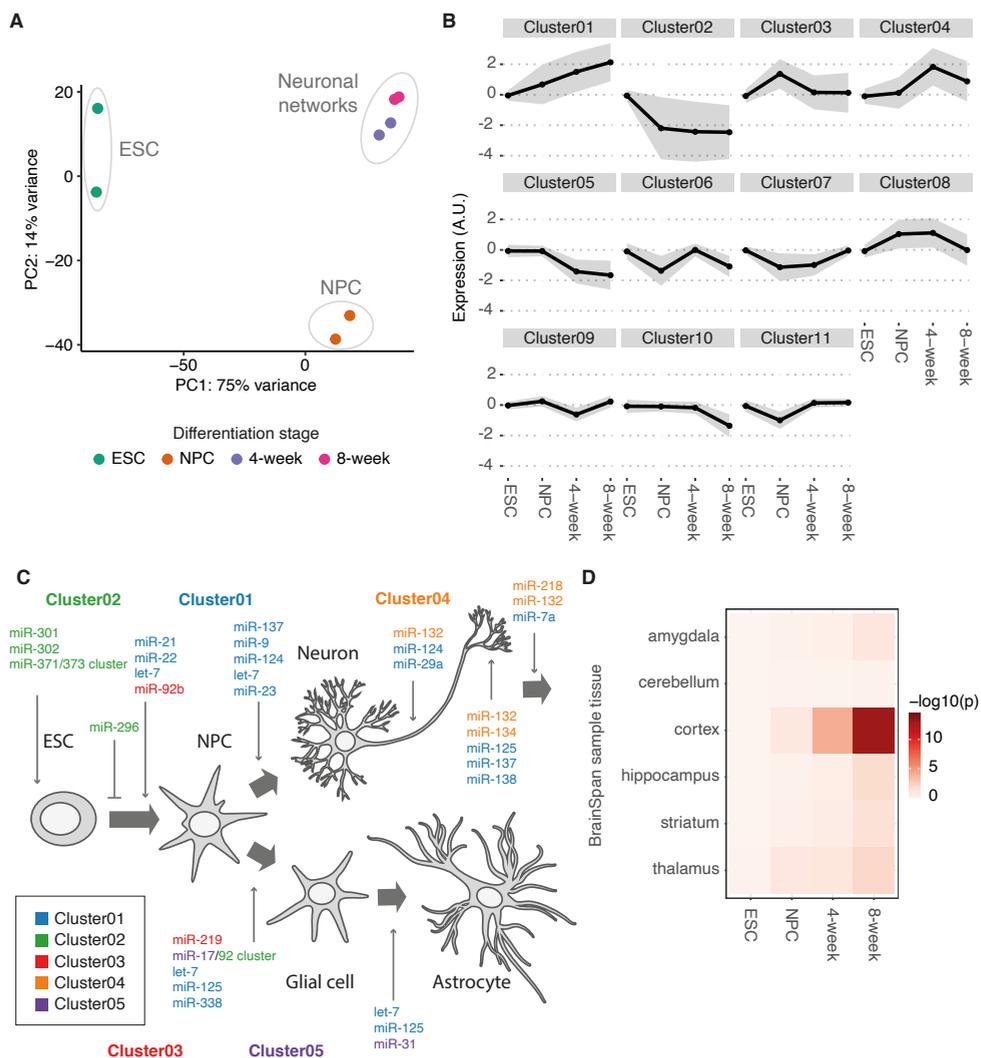


Figure 1 (A) Principal component analysis displays clustering of samples per differentiation stage based on miRNA expression profiles and distance between conditions. (B) Markov clustering organises miRNAs by expression pattern. For each cluster, average expression change for miRNAs is indicated by a black line, with the s.d. range indicated in grey. (C) The most highly expressed miRNAs of cluster01 to cluster05 are linked to functions that overlap with known developmental processes (miRNA functions are reviewed in ^{4,37,38}). Thick arrows indicate differentiation steps between cell types. The rightmost thick arrow indicates neuron fate specification (miR-218, miR132, miR7a). Other processes depicted are ESC maintenance (miR-301, miR-302 and miR-371/373 cluster), axon development (miR-132, miR-124, miR-29a), and spine maturation and synapse function (miR-132, miR-134, miR-125, miR-137, miR-138). (D) Correlation heatmap of miRNA expression between developmental stages of *in vitro* neuronal networks and *ex vivo* neocortical samples from BrainSpan.

the human brain and that this model most closely resembles the cortex. The previously generated transcriptomic dataset (Chapter 3 of this thesis) also allowed to assess the temporal correlation pattern of mRNA expression in the model which most closely matched cortical tissues, albeit with a much lower enrichment of correlation (Fig. S1). Possibly, mRNA expression contains more transcriptional noise than miRNA expression or conversely the larger number of post-mortem samples may have introduced more clinical variation in the dataset.

4.3.2 Correlations between miRNA and mRNA target expression indicate potential miRNA regulated pathways

In order to prioritize miRNAs with strong regulation and identify potentially miRNA-regulated processes, the representation of miRNA target motifs in mRNA expression was assessed at each developmental stage. This was done using an algorithm called ISMARA (Integrated System for Motif Activity Response Analysis) that infers the activity of miRNA target motifs by fitting a linear model that explains the observed gene expression signals in terms of the number of binding sites and unknown motif activity³². Each motif is then ranked based on a Z-score, which summarises the importance of the motif for explaining expression variation across the samples.

Using ISMARA, the activity for 208 miRNAs (106 motifs) could be predicted. 17 motifs had an inferred activity of more than two standard deviations away from 0 (Z-value above 1.96), which associated with 57 miRNAs (Table 1). 27 of these 57 miRNAs had a significant correlation between predicted activity and miRNA expression ($p < 0.05$) (Fig. S2).

The 27 significantly correlated miRNAs were distributed over 10 motifs, mostly found in cluster01 (11 miRNAs) and cluster02 (10 miRNAs) (Fig. S2). One motif was associated with cluster04 (UGUGCUU, miR-218). Three motifs were associated with miRNAs that were in multiple expression clusters: UGGUCCC and GUGCAAA were associated with miRNAs in cluster02 and cluster05. AUUGCAC was associated with miRNAs in cluster02, cluster05 and cluster03, all of which have a peak at the ESC or NPC stage. Predicted target mRNAs often drove the activity for multiple motifs (Fig. 2A). Pairwise overlaps showed that targets are often shared by motifs that associate to miRNAs from the same expression cluster (Fig. 2B). Notably, GGCAGUG activity was positively correlated to the expression of miR-34a-5p, whereas the other motifs with cluster01 miRNAs had a negative correlation.

Table 1 Most active predicted miRNA-associated motifs during model development

Motif	Z-value	Associated miRNAs ¶
UGUGCUU	5.49	miR-218 (**)
AAGGCAC	4.70	miR-124-3p.1
CCCUGAG	4.21	miR-125a-5p (*), miR-125b-5p (*)
CUUUGGU	3.80	miR-9-5p (*)
GGCAGUG	3.58	miR-34a-5p (*), miR-34c-5p, miR-449a, miR-449b-5p
AGCACCA	3.10	miR-29a-3p, miR-29b-3p, miR-29c-3p
AAGUGCU	2.95	miR-302a-3p (*), miR-302b-3p (*), miR-302c-3p.1 (*), miR-302d-3p (*), miR-302e, miR-372-3p, miR-373-3p (*), miR-520a-3p (*), miR-520c-3p, miR-520d-3p (*), miR-520e
UCCAGUU	2.85	miR-145-5p, miR-5195-3p
AUUGCAC	2.75	miR-25-3p, miR-32-5p (*), miR-363-3p (**), miR-367-3p, miR-92a-3p (*), miR-92b-3p (*)
UAUUGCU	2.70	miR-137
AACACUG	2.68	miR-141-3p, miR-200a-3p
UUGGCAC	2.66	miR-1271-5p, miR-96-5p
GAGGUAG	2.58	let-7a-5p (*), let-7b-5p (*), let-7c-5p (*), let-7d-5p (*), let-7e-5p (*), let-7f-5p (*), let-7g-5p (*), let-7i-5p (*), miR-4458, miR-4500 (*), miR-98-5p (*)
GCUGGUG	2.33	miR-138-5p
UCAAGUA	2.29	miR-1297, miR-26a-5p, miR-26b-5p (*)
UGGUCCC	2.09	miR-133a-3p.1 (*)
GUGCAAA	2.04	miR-19a-3p (**), miR-19b-3p (*)

¶ Stars indicate significant miRNA motif activity-expression correlation: * $p < 0.05$, ** $p < 0.01$.

Pathway analysis showed that the targets of cluster01-associated motifs were mostly related to early morphogenic processes (e.g., embryonic morphogenesis, tissue morphogenesis, tube morphogenesis). The targets of cluster02-associated motifs were mostly related to the organization of cell-cell contact (e.g., cell-cell adhesion, cell junction organization, modulation of chemical synapses) and general brain development (e.g. head development, neuronal system). Some ontologies related to cell morphology seemed to be regulated by miRNAs throughout development (e.g. regulation of cell projection organization, positive regulation of cellular component biogenesis and Rho GTPase cycle). These ontologies could give further indication of the developmental function of the miRNAs in the different expression clusters. Early morphogenic processes may need to be repressed along neuronal differentiation (cluster01), whereas genes related to neuron development and synapses may need to be repressed at the stem cell stage (cluster02). All of these ontologies related to the targets of developmentally active miRNA motifs present candidate neurodevelopmental phenotypes in which the pathological role of miRNA regulation could be further investigated.

4.3.3 *MiRNA expression differences in a model for CNTNAP2 deficiency disorder*

Having analysed the developmental expression patterns of miRNAs and their functional implications, the next step was to apply this knowledge to a knockout model of *CNTNAP2* to understand the role of miRNA regulatory networks that may surround *CNTNAP2* function. Two types of *CNTNAP2* knockout were used: a long isoform only knockout (HOM) and a dual isoform knockout (DUAL) that lacks all protein-coding isoforms (see Chapter 3 of this thesis for details). These were created by targeting the *CNTNAP2* gene locus in two sequential rounds of CRISPR/Cas9 editing in hESCs. In the first round, exon 1 was targeted to create indels that disrupt the long isoform. Subsequently, the HOM line was used to create a dual knockout line that lacks all protein isoforms by targeting exon 22, producing the DUAL line.

The miRNA expression of wild-type (WT), HOM and DUAL cell networks at 8 weeks of neuronal differentiation showed a clear clustering based on the experimental condition (Fig. 3A). The *CNTNAP2* mutant conditions had 280 upregulated and 189 downregulated miRNAs, of which 112 and 37, respectively, were shared between the HOM and DUAL conditions (Fig. 3B). The DUAL condition had a larger set of misregulated miRNAs, which could be indicative of additional functions of the short isoform. There was no overlap of downregulated and upregulated miRNAs between mutant conditions, showing that miRNA expression changes overall went in similar directions in the *CNTNAP2* mutant cultures.

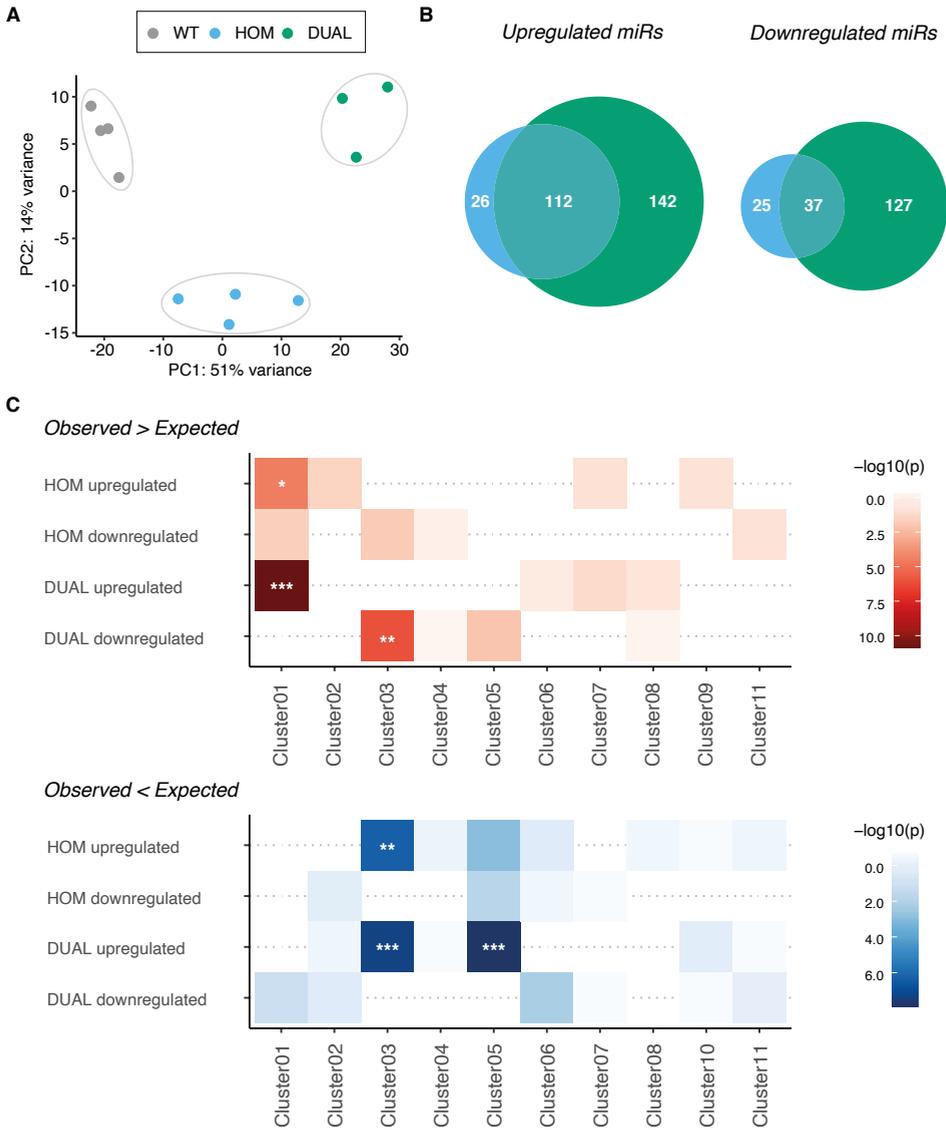


Figure 3 Differential expression of miRNAs between wild-type and *CNTNAP2* mutant mature neural networks. (A) Principal component analysis displays clustering of samples per condition and distance between conditions. (B) Euler diagrams represent sets of upregulated and downregulated miRNAs (fold change >2 and FDR-corrected $p < 0.05$) per mutant-wildtype comparison. (C) Heatmap showing enrichment for differentially expressed miRNAs in each expression pattern cluster. Top panel displays overrepresentation (observed > expected) and bottom panel underrepresentation (observed < expected) of each miR set per cluster. Colours represent $-\log_{10}(p)$ of FDR-corrected p-values from Fisher's exact test.

In order to understand the role of these differentially expressed miRNAs, enrichment of differentially expressed miRNAs was investigated across the developmental expression pattern clusters. Upregulated miRNAs in HOM and DUAL were significantly overrepresented in cluster01, a cluster related to overall neuronal differentiation, and significantly underrepresented in cluster03, a cluster related to early neuronal differentiation (Fig. 3C). In addition, miRNAs that were downregulated in DUAL were significantly overrepresented in cluster03, and miRNAs upregulated in DUAL were significantly underrepresented in cluster05, a cluster that is also related to early differentiation. This suggests that both conditions have an increase in miRNAs related to neuronal development and that the DUAL predominantly leads to additional miRNA expression changes in early development.

Table 2 Most active miRNA-associated motifs, comparing WT, HOM and DUAL. Motifs and correlated miRNAs that are detected in the WT-mutant and not in the comparison across developmental stages of WT neuronal network differentiation are highlighted in bold.

Motif name	Z-value	Associated miRNAs ¶
AAGGCAC	5.88	miR-124-3p.1
CUUUGGU	4.80	miR-9-5p (***)
UAUUGCU	2.96	miR-137
ACAGUAC	2.70	miR-101-3p.1
AGCAGCA	2.64	miR-15a-5p (***), miR-15b-5p, miR-16-5p (*), miR-195-5p, miR-424-5p, miR-497-5p (***), miR-6838-5p
AGCACCA	2.62	miR-29a-3p, miR-29b-3p, miR-29c-3p
UGCAUAG	2.53	miR-153-3p
AGCCCUU	2.39	miR-129-1-3p, miR-129-2-3p
AGUGCAA	2.37	miR-130a-3p (**), miR-130b-3p, miR-301a-3p (**), miR-301b-3p, miR-3666, miR-4295, miR-454-3p
AAAGUGC	2.31	miR-106a-5p, miR-106b-5p, miR-17-5p, miR-20a-5p, miR-20b-5p, miR-519d-3p, miR-526b-3p, miR-93-5p
UCACAUU	2.26	miR-130a-5p (**), miR-23a-3p (***), miR-23b-3p (**), miR-23c (***)
UGUGCUU	2.24	miR-218-5p
ACAUUCA	2.01	miR-181a-5p, miR-181b-5p (*), miR-181c-5p, miR-181d-5p (**), miR-4262
GAGGUAG	2.00	let-7a-5p (*), let-7b-5p (**), let-7c-5p, let-7d-5p (**), let-7e-5p, let-7f-5p (*), let-7g-5p, let-7i-5p, miR-4458, miR-4500, miR-98-5p (*)
UAAGGCA	1.96	miR-124-3p.2, miR-506-3p
UUGGCAA	1.96	miR-182-5p

¶ Stars indicate significant miRNA motif activity-expression correlation: * p < 0.05, ** p < 0.01, *** p < 0.001.

4.3.4 Differences in predicted miRNA activity in a model for *CNTNAP2* deficiency disorder

Having assessed changes in miRNA expression, we sought to investigate changes in miRNA activity in the mutant conditions. ISMARA was applied to mRNA-seq data from the same samples and detected 16 motifs with a Z-value above 1.96 (Table 2). 9 of these motifs did not pass the threshold in the motif activity analysis across the developmental stages in wild-type neuronal network differentiation. The 16 motifs are associated with 58 miRNAs. 17 of these 58 miRNAs had a significant correlation between predicted activity and miRNA expression ($p < 0.05$) and all correlations were negative (Fig. S3). The significantly correlated miRNAs were distributed over 6 motifs, of which 4 motifs were detected in the WT-mutant and not in the comparison across developmental stages of WT neuronal network differentiation. 11 of these 17 miRNAs were not detected in the analysis across developmental stages of WT neuronal network differentiation (Table 2). Most of the significantly correlated miRNAs were in cluster01 (10 miRNAs), but most motifs are associated to miRNAs in multiple clusters (Table 3). In Table 3, the directions of differences in miRNA expression and inferred motif activity are listed. For inferred motif activity differences that did not pass the significance threshold, the direction is mentioned between brackets.

In order to focus my pathway analysis on the strongest regulatory effects, I filtered the target genes for those that were differentially expressed between wild-type and mutant conditions. There was some overlap in the target lists (Fig. 4A), but it did not group very clearly per developmental expression pattern cluster in contrast to the developmental dataset (Fig. 4B). Instead, I found that motifs with upregulated miRNAs (GAGGUAG, ACAUUA, UCACAUU) and motifs with downregulated miRNAs (AGCAGCA, AGUGCAA, CUUUGGU) made up separate clusters,

In order to find the biochemical complexes and signal transduction components that could be represented in these miRNA target lists, an interactome (313 nodes) was built and annotations were provided for the most densely-connected complexes that were detected by the MCODE algorithm (Fig. 4C). The largest MCODE cluster was neuron projection development. *In vitro* studies have shown that knockdown of *CNTNAP2* in neurons leads to decreased neurite outgrowth^{24,41}. Interestingly, most genes in the neurite outgrowth MCODE cluster are targeted by downregulated miRNAs and are therefore subject to decreased repression. A subcluster of neurite outgrowth-related genes that is targeted by upregulated miRNAs is involved in ephrin-mediated growth cone collapse (EPHA3, EPHA4 and EPHA7)⁴². Together these findings suggest that the miRNA expression changes by themselves reflect an increase in neurite outgrowth. Repression by miRNAs also seems to be lifted for neurite outgrowth related signalling pathways such as receptor tyrosine kinase and RhoGTPase signalling. This may indicate that miRNAs expression changes reflect a regulatory mechanism that mitigates decreased neurite outgrowth due to *CNTNAP2* mutations.

Table 3 Direction of activity and expression changes in significant miRNA motif activity-expression correlations.

motif	activity WT-HOM	activity WT-DUAL	miRNA	cor. sig. ¶	miRNA cluster	miRNA WT-HOM	miRNA WT-DUAL
CUUUGGU	n.s. (UP)	UP	hsa-miR-9-5p	***	cluster 1	DOWN	DOWN
AGCAGCA	UP	n.s. (UP)	hsa-miR-15a-5p	***	cluster 3	n.s.	n.s.
			hsa-miR-16-5p	*	cluster 3	n.s.	n.s.
			hsa-miR-497-5p	***	cluster 2	n.s.	n.s.
AGUGCAA	n.s. (UP)	UP	hsa-miR-130a-3p	**	cluster 5	n.s.	DOWN
			hsa-miR-301a-3p	**	cluster 8	n.s.	DOWN
UCACAUU	DOWN	DOWN	hsa-miR-23a-3p	***	cluster 1	UP	UP
			hsa-miR-23b-3p	**	cluster 1	UP	UP
			hsa-miR-23c	***	cluster 1	UP	UP
ACAUUCA	n.s. (DOWN)	n.s. (DOWN)	hsa-miR-181b-5p	*	cluster 8	n.s.	n.s.
			hsa-miR-181d-5p	**	cluster 1	n.s.	n.s.
GAGGUAG	n.s. (DOWN)	n.s. (DOWN)	hsa-let-7a-5p	*	cluster 1	n.s.	n.s.
			hsa-let-7b-5p	**	cluster 1	UP	UP
			hsa-let-7d-5p	**	cluster 1	n.s.	UP
			hsa-let-7f-5p	*	cluster 1	n.s.	n.s.
			hsa-miR-98-5p	*	cluster 1	n.s.	n.s.

¶ Stars indicate significant miRNA motif activity-expression correlation: * $p < 0.05$, ** $p < 0.01$ *** $P < 0.001$.

4.4 Discussion

This study characterized the development of miRNA regulatory networks in an *in vitro* human neuronal network model and the response of these miRNA networks to neurodevelopmental perturbation by knockout mutations in *CNTNAP2*. The results showed that miRNA expression is highly dynamic during development and can be classified into several developmental profiles that contain known sets of miRNAs with specific functions. Motif-based analysis of predicted targets highlighted miRNA-regulated biological processes that are likely to be regulated at distinct stages throughout development. MiRNA networks change after neurodevelopmental perturbation caused by loss of *CNTNAP2*. Neuronal networks generated from *CNTNAP2* wild-type and mutant lines showed strong differences in miRNA expression. Particularly affected were miRNAs that correlated with neural precursors (cluster03) and neuronal differentiation (cluster01). Differentially expressed targets of active miRNAs showed enrichment for pathways including neuron projection development, supporting a role for miRNAs in this known phenotype of *CNTNAP2* knockout neurons. The direction in which the activity and expression of associated miRNAs changes suggest that miRNA regulatory networks adapt in order to counteract this developmental phenotype.

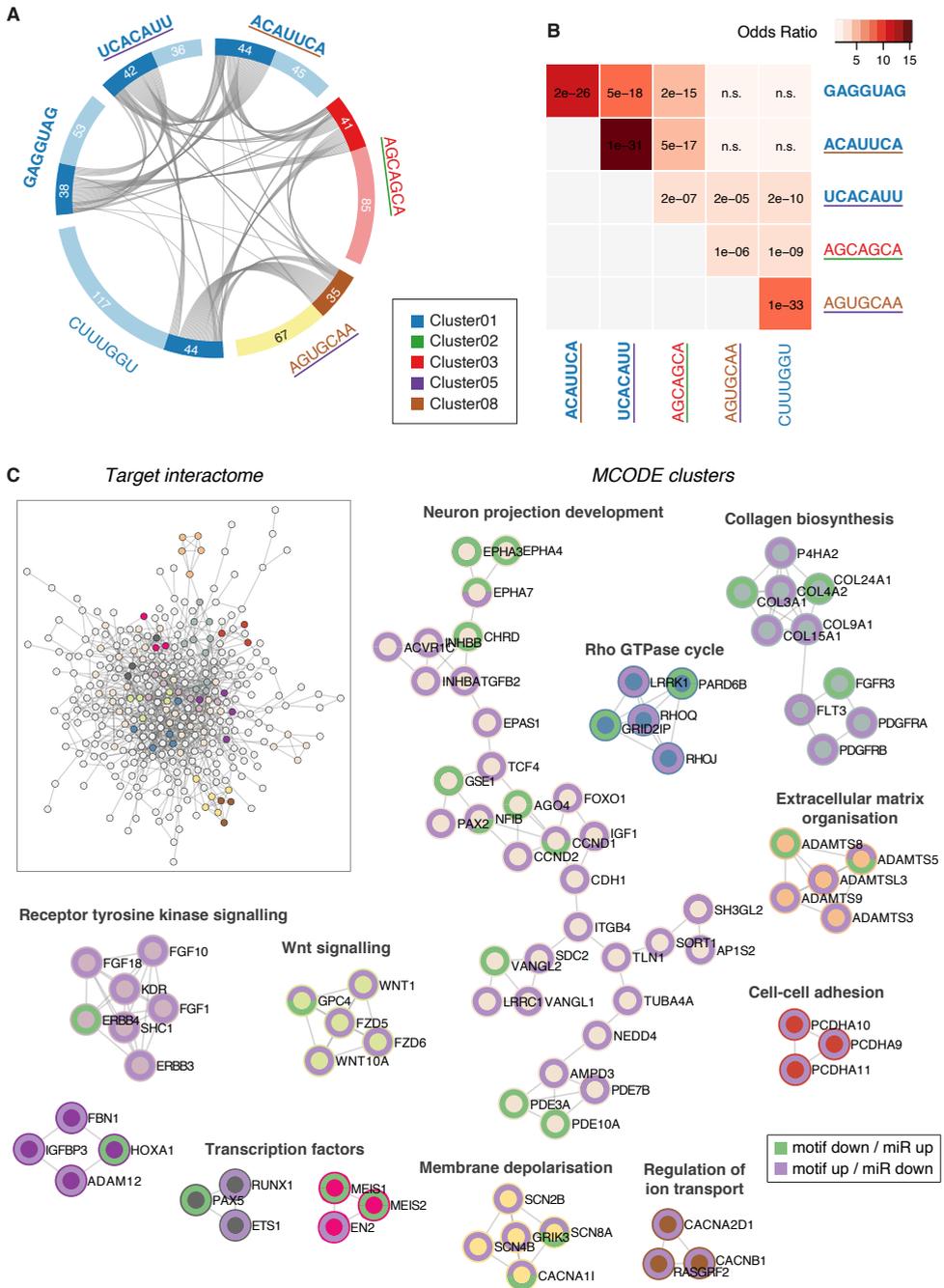


Figure 4 MiRNA targets of active miRNAs in WT-HOM-DUAL neuronal networks. (A) Motif-associated differentially expressed genes are shared between multiple motifs. (B) Pairwise overlap reveals clusters based on direction of differential miRNA expression (C) Interactome meta-analysis of target genes that drive motif activity.

The developmental expression profiles of miRNAs in this *in vitro* human neuronal network model suggest that this may represent an appropriate model for studying neurons in the cerebral cortex, in which the expression and activity of miRNAs can be studied to understand the roles of miRNA regulation in cortical development. Hierarchical clustering of post-mortem samples shows that cortical samples have specific miRNA expression patterns compared to those from other brain regions¹⁸. Differentiated samples from this human neuronal network model have a strong and specific miRNA expression correlation with human cortex, suggesting that the model recapitulates cell types and biological processes that are characteristic for cortical development. Several miRNAs are known to be involved in the control of cell identity and regulation of processes during neural development^{4,38}. When miRNAs are clustered based on their developmental expression profile in the current model, these well-studied miRNAs cluster in profiles that match their known functions and cell type specificity. This confirms that developmental expression trajectories from the brain are captured in this model, making it possible to functionally annotate less studied miRNAs that are expressed in the model. The main expression profile clusters were neuronal differentiation (cluster01), embryonic stem cells (cluster02), neural progenitor cells (cluster03), neuronal maturation (cluster04), and early development (cluster05). These data reveal the structure of miRNA expression patterns in *in vitro* neuronal network development and suggest when miRNAs function during cortical development.

The activities of miRNAs can be inferred from the expression of their target genes, which provides additional functional insight into the biological processes that are targeted by miRNA regulation. Most targets contain motifs of multiple active miRNAs, and that the most active miRNAs are involved in neuronal differentiation (cluster01) and embryonic stem cells (cluster02). Cluster-specific enriched biological pathways in this study have also been identified in earlier studies that examined the targets of differentially expressed miRNA targets between two differentiation conditions. A comparison of human induced pluripotent stem cells and NPCs found that miRNAs which repress Wnt signaling were upregulated during differentiation⁴³, which is part of the early morphogenic processes that were enriched in the regulated targets of active cluster01-miRNAs. A comparison of neural stem cells from E11 and E13 rat embryos identified neuronal differentiation to be under stronger repression in younger stem cells⁴⁴, which is also enriched in the regulated targets of active cluster02-miRNAs. The cluster-specific enrichment and matching contrast-based studies show that these processes are likely to be regulated by miRNAs in specific cell types during cortical development. In addition, the analysis presented here could also identify miRNA-regulated processes that are regulated across multiple stages in the neuronal network model. Processes such as cell projection organisation and RhoGTPase cycle may require regulation by miRNAs at multiple stages during cortical development. Functional analysis of rare CNVs in ASD emphasizes that these pathways can play roles in the pathogenesis of neurodevelopmental disorder⁴⁵. The detection of highly active miRNAs thus helps to prioritize molecular pathways that may be particularly vulnerable to gene expression changes during cortical development.

Perturbation of *CNTNAP2* gene function dramatically affected the expression of miRNAs in the neuronal cell model. Comparing *CNTNAP2* wild-type with mutant neuronal networks revealed 280 significantly upregulated and 189 significantly downregulated miRNAs in both mutant lines combined. miRNAs that were upregulated in mutant lines were

significantly overrepresented in neuronal differentiation (cluster01) and significantly underrepresented in neural precursor cells (cluster03) in both HOM and DUAL conditions. This indicates an increase in neuronal differentiation in *CNTNAP2* knockout neurons. Additionally, downregulated miRNAs were significantly overrepresented in neural progenitors (cluster03) in the DUAL condition, suggesting that NPCs are affected more severely when both the short and long isoform of *CNTNAP2* are knocked out. A role for the short isoform in NPCs is further supported by the short isoform expression pattern that peaks in NPCs and the DUAL-specific cilia assembly phenotype, which was previously described (Chapter 3). This previously described work further shows that *CNTNAP2* mutation results in a decrease in neural progenitor cell proportions and an increase in neuronal differentiation, via immunofluorescent staining and decomposed mRNA-seq data (Chapter 3) in both HOM and DUAL conditions. Differences in cell-type composition may contribute to the observed differential expression in this model. MiRNA expression is highly cell type specific^{3,46} and therefore changes in cell type composition could drive miRNA expression changes. Additionally, there may also be cell-autonomous effects on miRNA expression that result from loss of *CNTNAP2* function. Single cell miRNA expression profiling^{47,48} would help to determine which changes are due to cell type composition changes versus cell-autonomous effects in future experiments. The expression cluster annotation of differentially expressed miRNAs indicate that mutations in *CNTNAP2* lead to neurodevelopmental perturbation at the NPC and neuronal differentiation stage of the neuronal network model. The further identification of miRNAs that are directly related to *CNTNAP2* function should focus on the cell types that are associated to these developmental stages.

To understand how miRNAs are involved in *CNTNAP2* function, this study can point to a small number of miRNAs. Motif-based analysis of target mRNAs predicted the miRNAs with the highest differential activity between wild-type and mutant conditions. Candidate miRNAs were further filtered by correlating motif activity with miRNA expression: the activity of 6 motifs significantly correlated with the expression of 17 miRNAs. Motifs with upregulated miRNAs and motifs with downregulated miRNAs helped to identify separate clusters of target genes. This suggests not only that target genes can be combinatorially affected by multiple miRNAs, but also that miRNAs regulate these targets in the same direction. The group of differentially expressed target genes with differential motif activity contained a large cluster of interacting genes involved in neurite outgrowth and smaller clusters that were annotated for related signalling pathways, such as receptor tyrosine kinase and RhoGTPase signalling^{49,50}. Deficits in dendritic development have been documented as a phenotype associated with loss of *CNTNAP2* in several neuronal cell types, including pyramidal neurons, interneurons and Purkinje cells^{24,26,27,51}. MiRNA regulation may play a role in this neurite outgrowth phenotype, as also supported by the enrichment I found for cell projection organisation in the pathway analysis of active miRNA targets during neuronal network development. When miRNA expression is increased, this has the functional consequence of more target gene repression in the cell and conversely, repression of targets is lifted when miRNA expression is decreased. In the *CNTNAP2* knockout networks, miRNAs that target genes that promote neurite outgrowth are downregulated, whereas miRNAs that target genes that inhibit neurite outgrowth are upregulated. This suggests that miRNA regulation is adapted to counteract the neurite outgrowth phenotype of *CNTNAP2* knockout neurons. Future experiments

can investigate how these candidate miRNAs affect neurite outgrowth more generally, and how modulated expression of these molecules may affect the outcomes of *CNTNAP2* mutations that cause neurodevelopmental disorder.

A deeper understanding of miRNA regulatory networks can improve the genetic diagnosis of neurodevelopmental disorders, as these networks lead us to mechanisms that provide robustness to biological systems. Integrating miRNAs and the binding sites from established miRNA regulatory networks into genomic pipelines may identify additional mutations that explain clinical variation between patients with similar coding mutations⁵². For example, individuals with heterozygous mutations in *CNTNAP2* that present with CASPR2 deficiency disorder symptoms may carry compound mutations in miRNAs that regulate neurite outgrowth. Functional validation of these potentially pathogenic variants will remain essential, since miRNA-mRNA networks are constructed based on imperfect predictions³⁴. Current developments improve the reliability of construction of miRNA-mRNA networks. Increasing availability of physical interaction data from AGO2-HITS-CLIP, a high throughput method to reveal footprints of the contacts between RNA-induced silencing complex and RNA, from relevant tissues can lead to more accurate predictions of miRNA-mRNA binding^{34,53}. Single cell expression data can provide further insight into the dynamic configurations of these networks across cell types and developmental stages³. Following such developments, the number of candidate sites may be reduced further in future investigations. Given their amenability for experimental manipulation, scalability, and ability to model relevant phenotypes, human cell models provide a promising platform for the functional validation of potentially pathogenic variants in miRNA regulatory networks.

This study has shown that miRNA expression patterns of *in vitro* human neuronal networks resemble those in the human cortex and that developmental patterns of miRNA and target expression highlight biological processes that may require control by miRNAs during cortical development. This study has further shown that miRNA networks change in terms of expression and regulatory activity, which could indicate an adaptation to counterbalance neurodevelopmental phenotypes resulting from loss of *CNTNAP2*. Human cell models will thus allow researchers to both validate and identify new candidates in miRNA networks to investigate the modulating role of miRNAs in neurodevelopment and associated genetic disorders.

References

- 1 Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297, doi:10.1016/s0092-8674(04)00045-5.
- 2 Boudreau, R. L. *et al.* (2014). Transcriptome-wide discovery of microRNA binding sites in human brain. *Neuron* 81, 294-305, doi:10.1016/j.neuron.2013.10.062.
- 3 Nowakowski, T. J. *et al.* (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat Neurosci* 21, 1784-1792, doi:10.1038/s41593-018-0265-3.
- 4 Rajman, M. & Schratt, G. (2017). MicroRNAs in neural development: from master regulators to fine-tuners. *Development* 144, 2310-2322, doi:10.1242/dev.144337.
- 5 Hornstein, E. & Shomron, N. (2006). Canalization of development by microRNAs. *Nat Genet* 38 Suppl, S20-24, doi:10.1038/ng1803.
- 6 Ebert, M. S. & Sharp, P. A. (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell* 149, 515-524, doi:10.1016/j.cell.2012.04.005.
- 7 Schizophrenia Psychiatric Genome-Wide Association Study, C. (2011). Genome-wide association study

- identifies five new schizophrenia loci. *Nat Genet* 43, 969-976, doi:10.1038/ng.940.
- 8 Pardinas, A. F. *et al.* (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 50, 381-389, doi:10.1038/s41588-018-0059-2.
 - 9 Siegart, S. *et al.* (2015). The schizophrenia risk gene product miR-137 alters presynaptic plasticity. *Nat Neurosci* 18, 1008-1016, doi:10.1038/nn.4023.
 - 10 He, E. *et al.* (2018). MIR137 schizophrenia-associated locus controls synaptic function by regulating synaptogenesis, synapse maturation and synaptic transmission. *Hum Mol Genet* 27, 1879-1891, doi:10.1093/hmg/ddy089.
 - 11 Abelson, J. F. *et al.* (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* 310, 317-320, doi:10.1126/science.1116502.
 - 12 Devanna, P. *et al.* (2018). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol Psychiatry* 23, 1375-1384, doi:10.1038/mp.2017.30.
 - 13 Devanna, P., van de Vorst, M., Pfundt, R., Gilissen, C. & Vernes, S. C. (2018). Genome-wide investigation of an ID cohort reveals de novo 3'UTR variants affecting gene expression. *Hum Genet* 137, 717-721, doi:10.1007/s00439-018-1925-9.
 - 14 Wu, Y. E., Parikshak, N. N., Belgard, T. G. & Geschwind, D. H. (2016). Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nat Neurosci* 19, 1463-1476, doi:10.1038/nn.4373.
 - 15 Moreau, M. P., Bruse, S. E., David-Rus, R., Buyske, S. & Brzustowicz, L. M. (2011). Altered microRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder. *Biol Psychiatry* 69, 188-193, doi:10.1016/j.biopsych.2010.09.039.
 - 16 Berezikov, E. *et al.* (2006). Diversity of microRNAs in human and chimpanzee brain. *Nat Genet* 38, 1375-1377, doi:10.1038/ng1914.
 - 17 Hu, H. Y. *et al.* (2011). MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet* 7, e1002327, doi:10.1371/journal.pgen.1002327.
 - 18 Sousa, A. M. M. *et al.* (2017). Molecular and cellular reorganization of neural circuits in the human lineage. *Science* 358, 1027-1032, doi:10.1126/science.aan3456.
 - 19 Somel, M. *et al.* (2011). MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol* 9, e1001214, doi:10.1371/journal.pbio.1001214.
 - 20 Pollen, A. A. *et al.* (2019). Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* 176, 743-756 e717, doi:10.1016/j.cell.2019.01.017.
 - 21 Rodenas-Cuadrado, P. *et al.* (2016). Characterisation of CASPR2 deficiency disorder--a syndrome involving autism, epilepsy and language impairment. *BMC Med Genet* 17, 8, doi:10.1186/s12881-016-0272-8.
 - 22 Smogavec, M. *et al.* (2016). Eight further individuals with intellectual disability and epilepsy carrying biallelic CNTNAP2 aberrations allow delineation of the mutational and phenotypic spectrum. *J Med Genet* 53, 820-827, doi:10.1136/jmedgenet-2016-103880.
 - 23 Penagarikano, O. *et al.* (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147, 235-246, doi:10.1016/j.cell.2011.08.040.
 - 24 Anderson, G. R. *et al.* (2012). Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc Natl Acad Sci U S A* 109, 18120-18125, doi:10.1073/pnas.1216398109.
 - 25 Varea, O. *et al.* (2015). Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. *Proc Natl Acad Sci U S A* 112, 6176-6181, doi:10.1073/pnas.1423205112.
 - 26 Gao, R. *et al.* (2018). CNTNAP2 stabilizes interneuron dendritic arbors through CASK. *Mol Psychiatry* 23, 1832-1850, doi:10.1038/s41380-018-0027-3.
 - 27 Lazaro, M. T. *et al.* (2019). Reduced Prefrontal Synaptic Connectivity and Disturbed Oscillatory Population Dynamics in the CNTNAP2 Model of Autism. *Cell Rep* 27, 2567-2578 e2566, doi:10.1016/j.celrep.2019.05.006.
 - 28 Vitsios, D. M. & Enright, A. J. (2015). Chimira: analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics* 31, 3365-3367, doi:10.1093/bioinformatics/btv380.
 - 29 Dobin, A. *et al.* (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21, doi:10.1093/bioinformatics/bts635.
 - 30 Freeman, T. C. *et al.* (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput Biol* 3, 2032-2042, doi:10.1371/journal.pcbi.0030206.
 - 31 Love, M. I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550, doi:10.1186/s13059-014-0550-8.
 - 32 Balwierz, P. J. *et al.* (2014). ISMAR: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res* 24, 869-884, doi:10.1101/gr.169508.113.

- 33 Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19, 92-105, doi:10.1101/gr.082701.108.
- 34 Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, doi:10.7554/eLife.05005.
- 35 Zhou, Y. *et al.* (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10, 1523, doi:10.1038/s41467-019-09234-6.
- 36 Gunhanlar, N. *et al.* (2018). A simplified protocol for differentiation of electrophysiologically mature neuronal networks from human induced pluripotent stem cells. *Mol Psychiatry* 23, 1336-1344, doi:10.1038/mp.2017.56.
- 37 Gangaraju, V. K. & Lin, H. (2009). MicroRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol* 10, 116-125, doi:10.1038/nrm2621.
- 38 Zolboot, N., Du, J. X., Zampa, F. & Lippi, G. (2021). MicroRNAs Instruct and Maintain Cell Type Diversity in the Nervous System. *Front Mol Neurosci* 14, 646072, doi:10.3389/fnmol.2021.646072.
- 39 Barca-Mayo, O. & De Pietri Tonelli, D. (2014). Convergent microRNA actions coordinate neocortical development. *Cell Mol Life Sci* 71, 2975-2995, doi:10.1007/s00018-014-1576-5.
- 40 Mogilyansky, E. & Rigoutsos, I. (2013). The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ* 20, 1603-1614, doi:10.1038/cdd.2013.125.
- 41 Canali, G. *et al.* (2018). Genetic variants in autism-related CNTNAP2 impair axonal growth of cortical neurons. *Hum Mol Genet* 27, 1941-1954, doi:10.1093/hmg/ddy102.
- 42 Lowery, L. A. & Van Vactor, D. (2009). The trip of the tip: understanding the growth cone machinery. *Nat Rev Mol Cell Biol* 10, 332-343, doi:10.1038/nrm2679.
- 43 Kulcenty, K. *et al.* (2019). MicroRNA Profiling During Neural Differentiation of Induced Pluripotent Stem Cells. *Int J Mol Sci* 20, doi:10.3390/ijms20153651.
- 44 Nielsen, J. A., Lau, P., Maric, D., Barker, J. L. & Hudson, L. D. (2009). Integrating microRNA and mRNA expression profiles of neuronal progenitors to identify regulatory networks underlying the onset of cortical neurogenesis. *BMC Neurosci* 10, 98, doi:10.1186/1471-2202-10-98.
- 45 Pinto, D. *et al.* (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368-372, doi:10.1038/nature09146.
- 46 Jovicic, A. *et al.* (2013). Comprehensive expression analyses of neural cell-type-specific miRNAs identify new determinants of the specification and maintenance of neuronal phenotypes. *J Neurosci* 33, 5127-5137, doi:10.1523/JNEUROSCI.0600-12.2013.
- 47 Hucker, S. M. *et al.* (2021). Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. *Nat Commun* 12, 4316, doi:10.1038/s41467-021-24611-w.
- 48 Wang, N. *et al.* (2019). Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. *Nat Commun* 10, 95, doi:10.1038/s41467-018-07981-6.
- 49 Kiryushko, D., Berezin, V. & Bock, E. (2004). Regulators of neurite outgrowth: role of cell adhesion molecules. *Ann N Y Acad Sci* 1014, 140-154, doi:10.1196/annals.1294.015.
- 50 Govek, E. E., Newey, S. E. & Van Aelst, L. (2005). The role of the Rho GTPases in neuronal development. *Genes Dev* 19, 1-49, doi:10.1101/gad.1256405.
- 51 Argent, L. *et al.* (2020). Caspr2 interacts with type 1 inositol 1,4,5-trisphosphate receptor in the developing cerebellum and regulates Purkinje cell morphology. *J Biol Chem* 295, 12716-12726, doi:10.1074/jbc.RA120.012655.
- 52 Wanke, K. A., Devanna, P. & Vernes, S. C. (2018). Understanding Neurodevelopmental Disorders: The Promise of Regulatory Variation in the 3'UTRome. *Biol Psychiatry* 83, 548-557, doi:10.1016/j.biopsych.2017.11.006.
- 53 Moore, M. J. *et al.* (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat Protoc* 9, 263-293, doi:10.1038/nprot.2014.012.

Supplementary material

Figure S1. Correlation heatmap of mRNA gene expression between developmental stages of *in vitro* neuronal networks and *ex vivo* neocortical samples from BrainSpan

Figure S2. Scatter plots of motif activity and miRNA expression for significant miRNA motif activity-expression correlations from RNA expression analysis comparing across developmental stages of WT neuronal network differentiation

Figure S3. Scatter plots of motif activity and miRNA expression for significant miRNA motif activity-expression correlations from RNA expression analysis comparing WT and *CNTNAP2* mutant neuronal network cultures

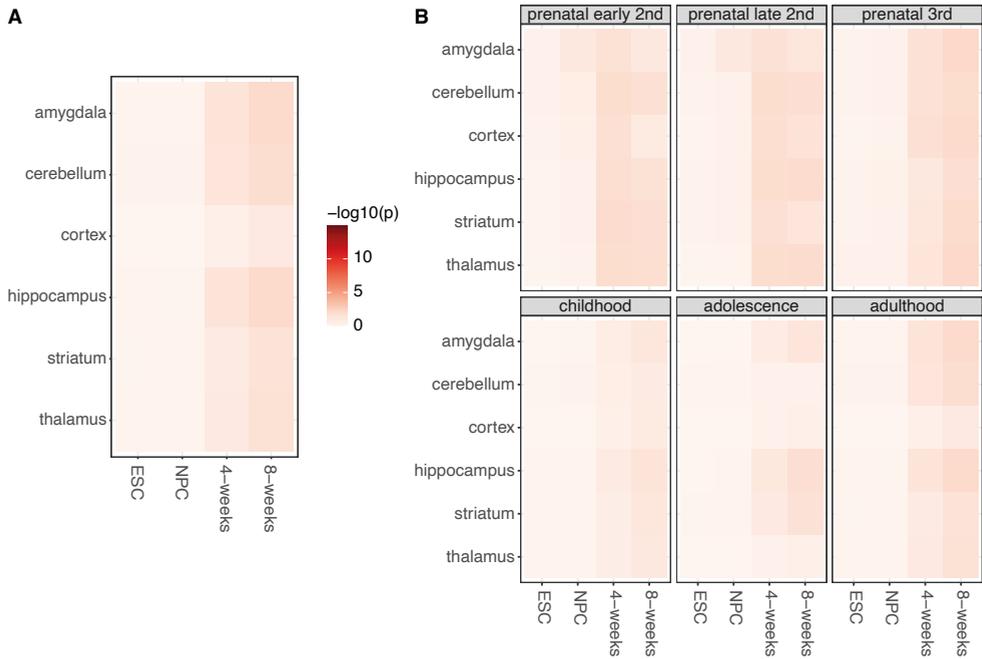


Figure S1 (A) Correlation heatmap of mRNA gene expression between developmental stages of *in vitro* neuronal networks and *ex vivo* neocortical samples from BrainSpan. (B) Correlation heatmaps split by developmental stage: prenatal early 2nd = 8-17 pcw, prenatal late 2nd = 19-24 pcw, prenatal 3rd = 25-37 pcw, childhood = 4mos – 13 yrs, adolescence = 15-21 yrs, adulthood = 23-40 yrs.

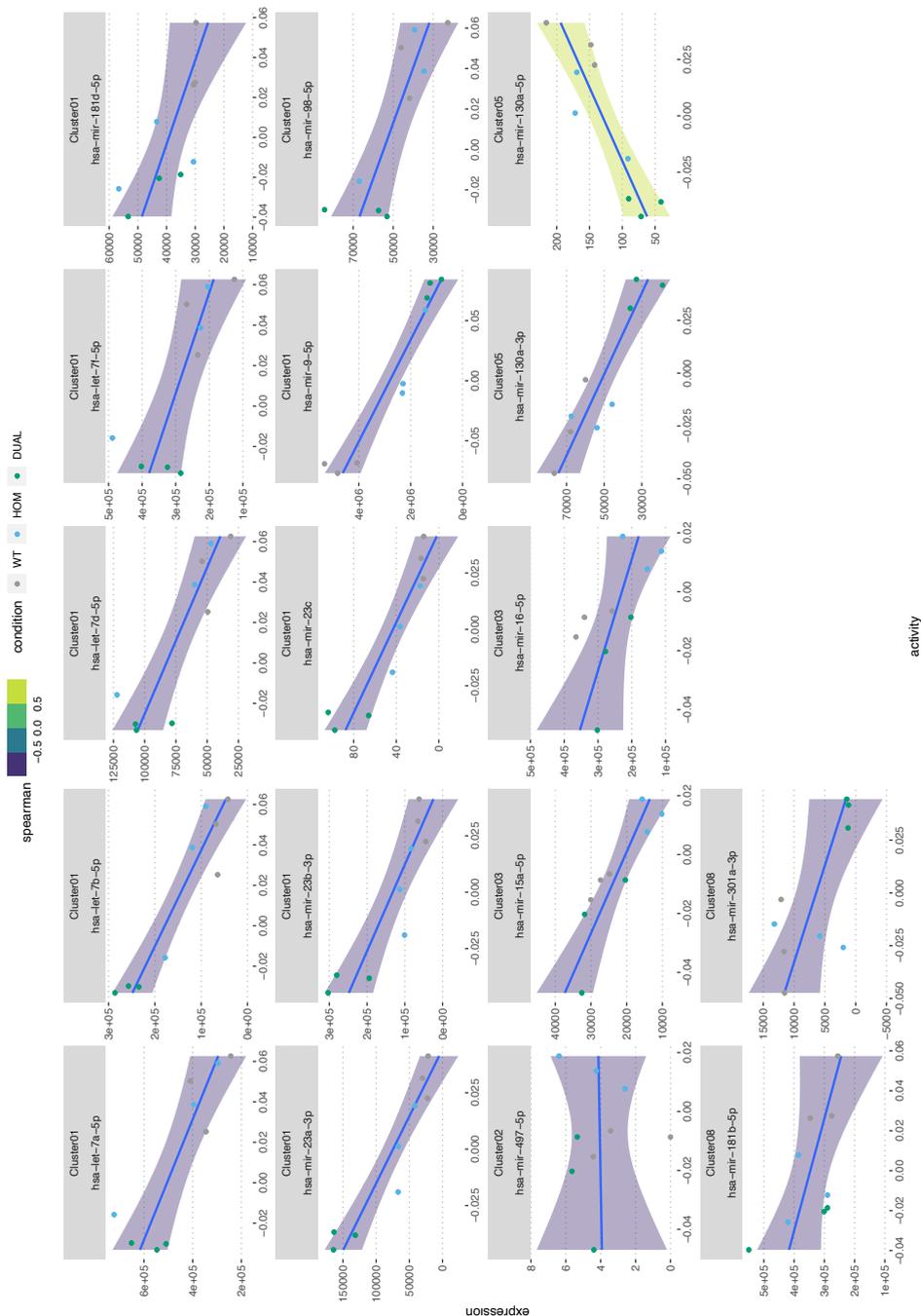
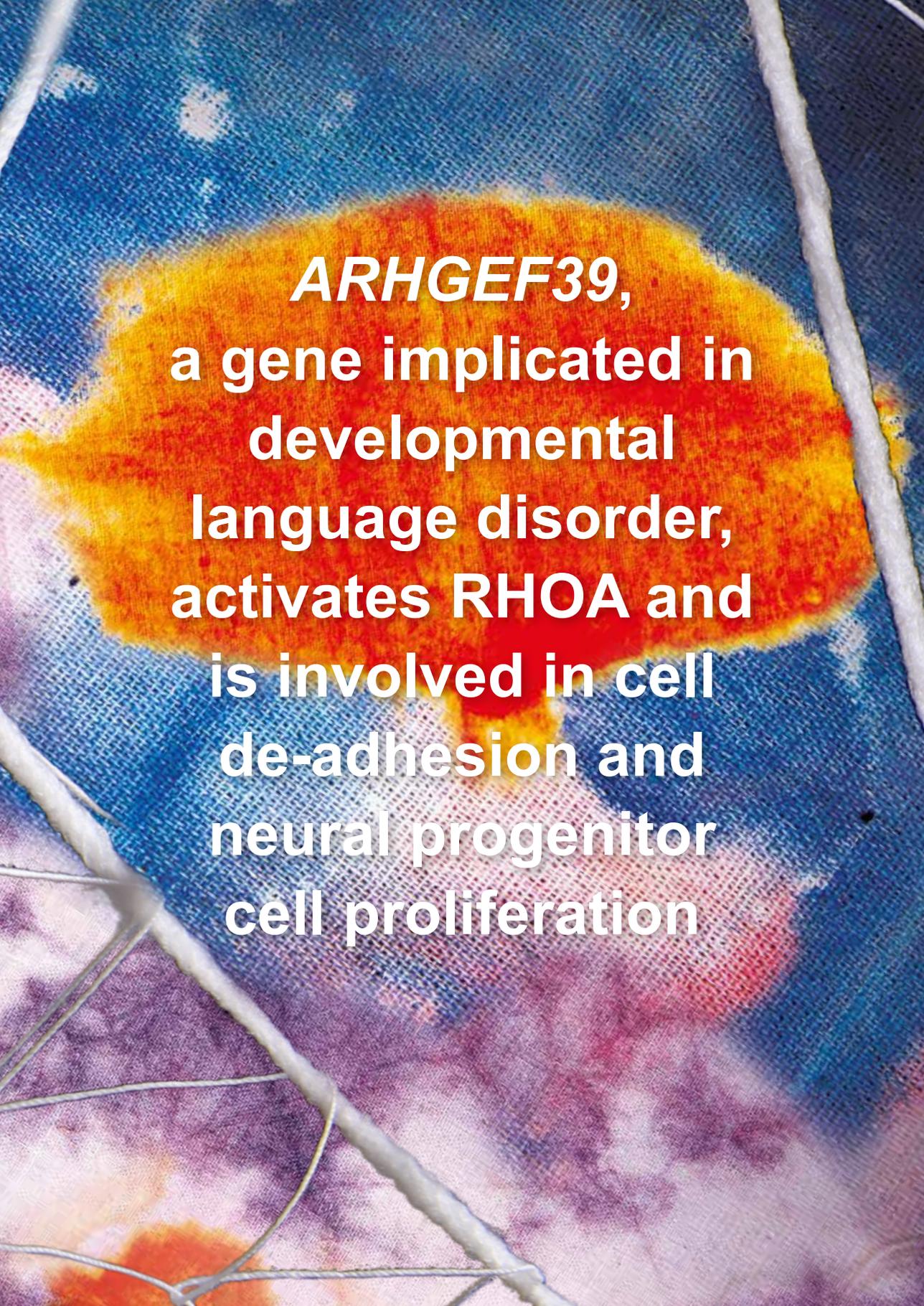


Figure S3 Scatter plots of motif activity and miRNA expression for significant miRNA motif activity-expression correlations from RNA expression analysis comparing WT and *CNTNAP2* mutant neuronal network cultures

5





***ARHGEF39*,**
a gene implicated in
developmental
language disorder,
activates RHOA and
is involved in cell
de-adhesion and
neural progenitor
cell proliferation

CHAPTER 5

This chapter has been published as:

Anijs, M., Devanna, P., and Vernes, SC. (2022) ARHGEF39, a Gene Implicated in Developmental Language Disorder, Activates RHOA and Is Involved in Cell De-Adhesion and Neural Progenitor Cell Proliferation. *Front Mol Neurosci. 15*, 941494, doi: 10.3389/fnmol.2022.941494

ARHGEF39 was previously implicated in developmental language disorder (DLD) via a functional polymorphism that can disrupt post-transcriptional regulation by microRNAs. *ARHGEF39* is part of the family of Rho guanine nucleotide exchange factors (RhoGEFs) that activate small Rho GTPases to regulate a wide variety of cellular processes. However, little is known about the function of *ARHGEF39*, or how its function might contribute to neurodevelopment or related disorders. Here, we explore the molecular function of *ARHGEF39* and show that it activates the Rho GTPase RHOA and that high *ARHGEF39* expression in cell cultures leads to an increase of detached cells. To explore its role in neurodevelopment, we analyse published single cell RNA-sequencing data and demonstrate that *ARHGEF39* is a marker gene for proliferating neural progenitor cells and that it is co-expressed with genes involved in cell division. This suggests a role for *ARHGEF39* in neurogenesis in the developing brain. The co-expression of *ARHGEF39* with other RHOA-regulating genes supports RHOA as substrate of *ARHGEF39* in neural cells, and the involvement of RHOA in neuropsychiatric disorders highlights a potential link between *ARHGEF39* and neurodevelopment and disorder. Understanding the GTPase substrate, co-expression network, and processes downstream of *ARHGEF39* provide new avenues for exploring the mechanisms by which altered expression levels of *ARHGEF39* may contribute to neurodevelopment and associated disorders.

ABSTRACT

ARHGEF39, a gene implicated in developmental language disorder, activates RHOA and is involved in cell de-adhesion and neural progenitor cell proliferation

5.1 Introduction

Developmental speech and language disorders are highly heritable, with most cases showing complex multifactorial inheritance ¹. This complex genetic aetiology makes the identification of risk genes challenging, but investigating the biological function of these genes offers an important gateway for understanding the biological basis of human speech and language ². Previously, *ARHGEF39* was identified as a potential risk factor for a syndrome of unexplained language problems in children known as developmental language disorder (DLD) ³. At the time of that study, this condition was labelled as specific language impairment, but DLD is now the recommended terminology ⁴. *ARHGEF39* was implicated following a screen of non-coding 3'UTR sequences for variation that could disrupt microRNA (miR) binding sites in children with DLD ³. A single nucleotide polymorphism in the *ARHGEF39* 3'UTR (rs72727021) was associated with a quantitative measure of language impairment (non-word repetition) and functional assays in cell models showed that the risk allele disrupted regulation of *ARHGEF39* by miR-215. Expression quantitative trait loci data further indicated that the DLD-associated allele was associated with higher expression of *ARHGEF39* in post-mortem human brain ³. However, little is known about the biological function of *ARHGEF39*, the role of this gene in neurodevelopmental processes, or how variation in these processes may contribute to human language development or disorder.

ARHGEF39 is one of 82 Rho guanine nucleotide exchange factors (RhoGEFs) in the human genome ⁵. RhoGEFs initiate the activation of RhoGTPases by stimulating them to bind GTP instead of GDP ⁶. RhoGTPases are involved in every cellular process that requires cytoskeletal reorganization ^{6,7}. The most extensively characterized RhoGTPases, CDC42, RAC and RHO, stimulate the re-organisation of the cytoskeleton into distinct cellular structures upon activation: filopodia, lamellipodia, and focal adhesions, respectively ⁸. In neurodevelopment, CDC42, RAC and RHO each have specific contributions to various processes, such as neurite outgrowth, axon pathfinding, and dendritic spine development via their effects on the cytoskeleton, membrane trafficking and microtubule dynamics ⁹. Each RhoGEF controls these processes by activating one or more of the RhoGTPases, meaning that the RhoGEFs substrate specificity is deterministic of its biological function. Aberrant RhoGTPase signalling is implicated in multiple neurodevelopmental disorders. Genes in RhoGTPase signalling pathways are enriched in rare CNVs associated with autism as well as in GWAS hits for schizophrenia and bipolar disorder ^{10,11}. Specific RhoGEFs have been implicated in neurodevelopmental disorders, such as language impairment (*ARHGEF19*), intellectual disability (*ARHGEF6*, *ARHGEF2*), and moderate intellectual disability with speech delay (*ARHGEF9*) ¹²⁻¹⁵. These findings highlight the potential importance of RhoGTPase signalling and its regulation in neurodevelopmental disorders.

Most studies on *ARHGEF39* have investigated its role in cancer, while the developmental and neurobiological roles of *ARHGEF39* have received limited attention, leaving open questions about its contribution to neurodevelopmental phenotypes and disorders. In hepatocellular and lung cancer, increased expression of *ARHGEF39* has been reported as a prognostic factor for tumour size and patient survival¹⁶⁻¹⁹. Furthermore, overexpression of *ARHGEF39* leads to increased proliferation, migration and invasion of cancer cells^{16,19-21}. These studies demonstrate a molecular link between *ARHGEF39* and cancer cell phenotypes. However, to understand the role of *ARHGEF39* in neurodevelopment or related disorders, we need to understand its molecular function, and study its expression patterns and potential molecular interactions in relevant models. In this study, we aim to do this by investigating the substrate specificity of *ARHGEF39* and consequences of its overexpression. We also utilise existing single cell RNA-seq datasets to determine the expression of *ARHGEF39* in the developing brain and uncover neurodevelopmental processes implicated in its function.

5.2 Methods

5.2.1 Cell culture and transfection

Biosensor assays were performed in human HEK293FT cells. Cells were obtained from ThermoFisher and were routinely screened for mycoplasma contamination. All experiments were carried out using cells grown in Dulbecco's modified Eagle's media (Invitrogen) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and 2 mM penicillin/streptomycin. Cells were maintained at 37°C in the presence of 5% CO₂. Transfections were performed using GeneJuice (Novagen) following the manufacturer's instructions.

5.2.2 Expression vectors

Several expression vectors were used for these experiments. A vector expressing *ARHGEF39* (pcDNA3.1-*ARHGEF39*) was obtained from NovoPro (catalogue # 718357). Plasmids encoding second generation FRET-based RhoGTPase biosensor were obtained from Addgene: pTriEx4-Rac1-2G (#66110), pTriExRhoA2G (#40176), pTriEx4-Cdc42-2G (#68814)²²⁻²⁴. *ARHGDIA* was PCR amplified from cDNA from SH-SY5Y cells with the following primers: *ARHGDIA_HindIII_Fw* TTACTAAGCTTATGGCTGAGCAGGAGCCACAG and *ARHGDIA_KpnI_Rv* TTACTGGTACCGTCCTCCAGTCCTTCTTGATG. The PCR product was cloned into the pcDNA3.1 expression vector using HindIII and KpnI restriction sites to create pcDNA3.1-*ARHGDIA*. The sequence was confirmed by Sanger sequencing.

5.2.3 FRET-based RhoGTPase biosensor assay

HEK293FT cells were seeded in poly-D-lysine-coated glass-bottom 96-well plates at 20,000 cells per well and were allowed to adhere overnight. Cells were transfected with 20 ng of pTriEx4-Rac1-2G, pTriExRhoA2G, or pTriEx4-Cdc42-2G and 10 ng of pcDNA3.1-*ARHGDIA* and 0, 20, 60 or 100 ng of pcDNA3.1-*ARHGEF39*. The minimum amount of *ARHGDIA* that was needed to increase the dynamic range of the assay was determined in a dose-response experiment (Fig. S1). pcDNA3.1-empty was used as filler to keep total

DNA content constant across conditions. Medium was changed to FluoroBrite DMEM (ThermoFisher) with 10% fetal bovine serum one hour before reading the plate. Images of cells were captured 48 hours after transfection.

Biosensor assays were performed on a Tecan Infinite M200 PRO plate reader with a temperature-controlled incubation chamber at 37 °C and 5% CO₂. Excitation wavelength was set at 453 nm and an emission scan was read from the bottom of each well between 487 and 600 nm with a 1 nm step size. For each experiment (n=3), three wells per RhoGTPase:ARHGEF39 ratio were measured. Background fluorescence was measured in nine untransfected wells. Average background values were subtracted from the raw intensity values and spectra were normalized by area in a|e UV-Vis-IR Spectral Software (version 2.2, <http://www.fluortools.com/software/ae>). Subsequently, the ratio between 528nm and 492nm ratio was calculated to determine RhoGTPase activation. Significant differences between groups were calculated using an ANOVA test followed by *post hoc* Tukey HSD test.

5.2.4 Cell adhesion

HEK293FT cells were seeded in 12-wells plates at 100,000 cells per well and were allowed to adhere overnight. Cells were transfected with 500 ng pcDNA3.1-ARHGEF39 or pcDNA3.1-empty. Cells were counted 48 hours after transfection. Total culture medium was removed, centrifuged at 200xg for 3 minutes and resuspended in 100 ul of Dulbecco's PBS (Sigma) to count the number of floating cells. Attached cells were detached from the plate with 0.25% trypsin-EDTA (Invitrogen) and resuspended in 1 ml of Dulbecco's PBS after centrifugation (Sigma). Cells were stained with 0.4% Trypan Blue (BioRad) and counted with the TC20 automated cell counter (BioRad) according to the manufacturer's instructions. Significant differences in total cell counts and viability percentages were calculated with a two-sided t-test. Viability percentages were arcsine transformed before statistical testing.

5.2.5 Analysis of ARHGEF39 in scRNA-seq data

Expression matrix and meta file of ²⁵⁻²⁷ were downloaded from <https://github.com/jeremysimon/MouseCortex>, <http://solo.bmap.ucla.edu/shiny/webapp/> and Gene Expression Omnibus (GSE120046), respectively. These datasets were processed using Seurat v3.2.2 ²⁸. Clusters were selected if ARHGEF39 expression value was larger than 0 in at least 10% of cells. For each of the selected clusters, cells were grouped in ARHGEF39-positive and ARHGEF39-negative cells for a differential gene expression analysis. The sizes of ARHGEF39-positive groups can be found in Table S1. The smallest group was 70 ARHGEF39-positive cells for Loo_SVZ3. Differential gene expression analysis was performed with edgeR v3.28.1 ²⁹, using genewise negative binomial general linear models (glmFit) and likelihood ratio tests for the model (glmLRT). We used a cut-off value of FDR-corrected p-value < 0.01. GO enrichment analysis was performed on the basis of these DEGs and marker genes for radial glia, cluster 2 (E14.5) ²⁵ and cycling progenitors (G2/M phase) ²⁶ by Metascape ³⁰. Metascape was set to use gene sets from Gene Ontology and Reactome with default parameters. The union of all genes expressed in at least 10% of a cluster in the Loo et al, Zhou et al and Polioudakis et al. datasets (9360) was used as list of background genes. Overlap between DEG lists was statistically assessed with Fisher's

exact tests using the R package GeneOverlap v3.15.

5.3 Results

5.3.1 ARHGEF39 activates RHOA GTPase

Given their distinct roles in remodelling the cytoskeleton, determining the RhoGTPase activation specificity of ARHGEF39 can provide a first clue in understanding its downstream functions. To test if CDC42, RAC1 or RHOA could be activated by ARHGEF39, we used second-generation genetically encoded Förster Resonance Energy Transfer (FRET) biosensors for each of the RhoGTPases of interest (CDC42, RAC1 and RHOA)²²⁻²⁴. These biosensor molecules contain the RhoGTPase of interest, a RhoGTPase binding domain (RBD) and two fluorophores. The excited donor fluorophore (mTFP) transfers energy to an acceptor fluorophore (Venus) that emits at a characteristic wavelength when the two fluorophores are brought in a close configuration as a result of the activated RhoGTPase binding the RBD (Fig. 1A). We chose a widely used and highly tractable human cell line (HEK293FT cells) as a model for these tests as they focus on the general properties of the molecular interaction of ARHGEF39 with RhoGTPases rather than a cell type specific function. A plasmid overexpressing ARHGEF39 was co-transfected to HEK293FT cells in increasing quantities with a uniform amount of biosensor to test for RhoGTPase activation. Activation of CDC42 or RAC1 were not observed at any ratio. In contrast, RHOA

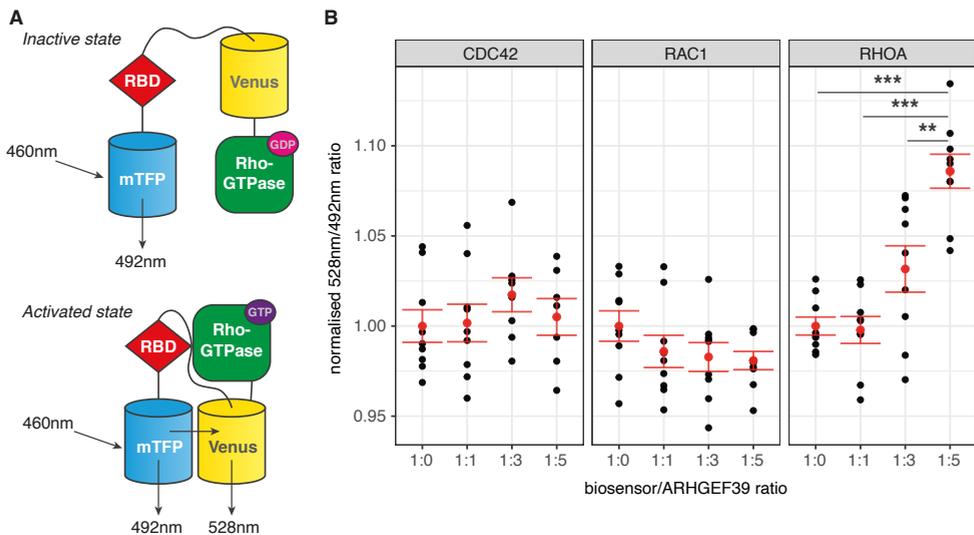


Figure 1 (A) Diagram of FRET-based Rho GTPase activity biosensor. If the RhoGTPase is in an inactive GDP-bound state only a cyan fluorescent protein (mTFP) is excited. When the RhoGTPase is in an activated GTP-bound state, it is able to bind a RhoGTPase binding domain (RBD) in the biosensor. This induces a conformational change that enables a cyan fluorescent protein (mTFP) to excite a yellow fluorescent protein (Venus) through FRET. Activity is measured as the ratio between emissions at 528nm and 492nm. (B) Normalised ratio of emissions at 528 and 492nm from RhoGTPase biosensors. Multiple ratios of biosensor vs ARHGEF39 co-transfection were used: 1:0 (no ARHGEF39 control), 1:1, 1:3 and 1:5. Mean and standard error are indicated in red. p-values are determined with a Tukey HSD test after ANOVA. *** indicates $p < 0.001$ ** $p < 0.01$

showed significant activation when ARHGEF39 was overexpressed at a 1:5 transfection ratio to the biosensor ($p = 0.00001$) (Fig. 1B).

5.3.2 Overexpression of ARHGEF39 increases the amount of cells in suspension

During our RhoGTPase activation experiments, we made an unexpected observation. In HEK293FT cell cultures that overexpressed ARHGEF39, more cells were observed floating in the culture media compared to cells that were transfected with an empty vector or EGFP control (Fig. 2A). This increase of cells in suspension could have multiple explanations. Given that ARHGEF39 activates RHOA (Figure 1), this effect could be related to the established role of RHOA in the assembly of cell-matrix interaction via focal adhesions^{8,31}. Alternatively, it could indicate that high concentrations of ARHGEF39 are toxic and cause cell death driving more cells into suspension. To differentiate between these possibilities, we quantified the number of cells in suspension vs. attached cells and assessed cell viability to determine whether the cells had only detached or if they had also died.

The visual observation that an increased number of cells were in suspension after overexpression of ARHGEF39 was confirmed by automated cell counting ($p = 0.0001$) (Fig. 2B). We also counted the attached cells and found no significant difference between conditions ($p = 0.96$). Total cell counts were not significantly different between conditions ($p = 0.59$) as the increased number of cells in suspension made up a very small proportion of the total (~0.5% of the control cells and ~8.4% of the ARHGEF39 transfected cells). Trypan blue staining demonstrated that viability of the attached cells was high (>90%) in both conditions and no significant difference was observed (Fig. 2C). Floating cells showed poor viability (~42%) in the empty vector control condition for the small proportion of floating cells that could be found. In contrast, floating cells were largely viable (~78%) in the ARHGEF39 overexpressing condition, which represents a substantial and significant increase compared to the control condition ($p = 0.004$). This suggests that high concentrations of ARHGEF39 are not toxic or causing cell death, rather, it is more likely that overexpression of ARHGEF39 leads to cell de-adhesion.

5.3.3 ARHGEF39 is expressed in multiple cell types and marks out neural progenitor cells

To understand its potential contributions to neurodevelopmental processes, we investigated ARHGEF39 expression in the developing brain using publicly available single cell RNA-sequencing (scRNA-seq) data. Large scRNA-seq datasets are available that describe the transcriptomes of individual cells. Unsupervised clustering methods applied to such data can group these cells based on similarity, after which the clusters are annotated for their properties such as cell type or state³². This provides a powerful resource that we have leveraged to understand the cell types in which *ARHGEF39* is found in the brain and, by exploring the co-expressed genes, to suggest which pathways are active when *ARHGEF39* is expressed. We explored three recent scRNA-seq datasets that detail the development of mouse²⁵ and human neocortex^{26,27}.

We first explored the marker genes of specific cell types that had been identified in the published cluster analyses. Here, a marker gene was defined as a gene that was significantly enriched in a cell type specific cluster during differential gene expression analysis when comparing the cluster with all other cells in the dataset. In the mouse and

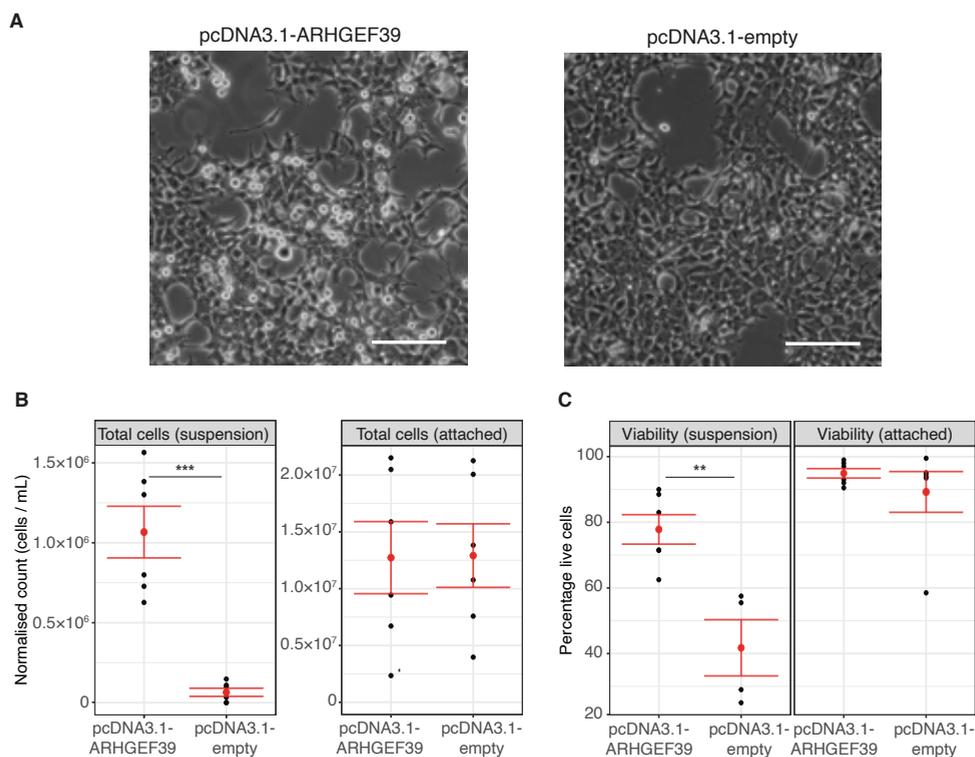


Figure 2 (A) Representative images of cells after transfection with an ARHGEF39 overexpressing vector (pcDNA3.1-ARHGEF39) or empty vector control (pcDNA3.1-empty). Scale bar indicates 500 μ m. (B) Counts of cells in suspension and attached cells 48 hours after transfection with pcDNA3.1-ARHGEF39 and pcDNA3.1-empty. Concentrations for cells in suspension and attached cells were measured in 100 μ l and 1ml resuspension volumes, respectively (See methods). Mean and standard error are indicated in red. (C) Viability of cells in suspension and attached cells 48 hours after transfection with pcDNA3.1-ARHGEF39 and pcDNA3.1-empty. p-values are determined with two-sided t-test *** indicates $p < 0.001$, and ** $p < 0.01$

one of the human datasets, *ARHGEF39* was identified as a marker gene for clusters of neural progenitor cells^{25,26}. In the embryonic mouse study, *Arhgef39* was classed as one of eight marker genes for one (cluster 2) of the four subpopulations of radial glia identified²⁵. In the mid-gestation human cortex study, *ARHGEF39* was among the 133 most significant differentially expressed genes for cycling progenitor cells in G2/M phase of the cell cycle²⁶. These data suggest that *ARHGEF39* marks out specific populations of neural progenitor cells in mouse and human cortical development. Given the G2/M annotation of the human cluster, the function of ARHGEF39 may be related to this cell cycle phase

Next, to identify all cell types in the developing brain in which ARHGEF39 could be found, we sought to identify the clusters in which *ARHGEF39* was reliably expressed in each dataset. Expression of ARHGEF39 in at least 10% of cells of a cluster was used as a threshold to identify positive cell types, and this identified nine ARHGEF39 positive clusters out of 71 clusters total (Table S1). Neural progenitor cell clusters had the highest percentage of *ARHGEF39*-positive cells across all datasets supporting its status as a neural

Table 1 Cell clusters in which at least 10% of cells express ARHGEF39. Clusters where ARHGEF39 was previously identified as a marker gene are denoted with an asterisk

Reference	Species	Age	ARHGEF39 ⁺ clusters	# cells	% pos.
Loo, et al. (2019) ²⁵	Mouse	E14.5-birth	Radial glia, cluster 2 (E14.5) *	334	45.5%
			Subventricular zone, proliferating (E14.5)	315	22.2%
			Ganglionic eminences (Birth)	421	17.1%
			Ganglionic eminences (E14.5)	762	10.4%
Polioudakis, et al. (2019) ²⁶	Human	GW17-18	Cycling progenitors (G2/M phase) *	695	20.3%
Fan, et al. (2020) ²⁷	Human	GW7-28	Neural progenitor cells	1333	30.3%
			Early (GW7-9)	1392	22.0%
			Immune cells	511	15.3%
			Cortical excitatory neurons	2065	10.4%

progenitor cell marker (Table 1). A role in neural development was further supported by bulk RNA-sequencing data from BrainSpan³³ and PsychENCODE³⁴ which showed a prenatal enrichment for *ARHGEF39* (Fig. S2). In the PsychENCODE data, *ARHGEF39* was also assigned to co-expression module ME5, which is enriched for gene expression associated to radial glia and neural progenitor cells³⁴. Other cell types that met the 10% threshold in the scRNA-seq datasets were cortex-adjacent ganglionic eminences in the developing mouse dataset²⁵, and cortical excitatory neurons and immune cells (containing microglia, macrophages and T cells) in one of the developing human datasets²⁷. These data show that, while strongly enriched in neural progenitor cells, *ARHGEF39* is also present in multiple cell types in the developing brain.

5.3.4 *ARHGEF39* is co-expressed with a core set of genes in the developing brain

Next, we used these scRNA-seq data to investigate how *ARHGEF39* expression may affect the molecular state of the cells. Exploiting the cell-to-cell heterogeneity within clusters, we divided each cluster into *ARHGEF39*-positive and *ARHGEF39*-negative cells and performed differential gene expression analysis between these groups to find genes that correlate with *ARHGEF39* expression. The numbers of differentially expressed genes (DEGs) for each cluster are in fig. 3A. In the mouse radial glia cluster and human cycling progenitors cluster, where *ARHGEF39* was identified as a marker gene, *ARHGEF39* was the only DEG identified. The overall expression of *ARHGEF39*-positive cells and *ARHGEF39*-negative cells are very similar within these clusters, indicating that expression of *ARHGEF39* did not correlate with a specific cell type or cell state within these clusters. In the next step of our analysis, we took all other marker genes that were defined for radial glia 2 and cycling progenitors in G2/M phase, as *ARHGEF39* likely correlates with a gene expression signature for these clusters as a whole^{25,26}.

Strong overlap between the lists of DEGs and the lists of marker genes show that there

is a core set of 46 genes that are co-expressed with ARHGEF39 (Fig. 3A; Fig. S3; Table S3). Similarly to ARHGEF39, all these genes are in co-expression modules that are enriched for prenatal gene expression that is associated with radial glia and neural progenitor cells in the PsychENCODE human developmental transcriptome³⁴. 44 out of 46 core set genes are assigned to the same co-expression module as ARHGEF39, supporting the co-expression observed in single cell data (Table S3). A large number of DEGs (1175) were identified in the cluster of early cells from the Fan et al. 2020 dataset compared to the other clusters. The core set of ARHGEF39 co-expressed genes is also detected in this cluster, but 79.4% of these DEGs did not overlap with other clusters. This can be explained by the type of annotation for this cluster. The cluster of early cells is not defined by cell type, but by gestational age (week 7 to 9)²⁷. As the early cell cluster contains a diversity of cell types, it also contains cell types that are ARHGEF39-negative. The non-overlapping DEGs are likely markers for these ARHGEF39-negative cell types that are included in the age-based early cell cluster but not in more cell type-based clusters. The core set of overlapping DEGs is present in all clusters and they likely correlate with a cell state in which ARHGEF39 is expressed. (Fig. 3A; Fig. S3).

5.3.5 ARHGEF39 expression-related genes are involved in cell cycle processes

To identify the biological processes that are associated with ARHGEF39 expression we performed a gene ontology (GO) enrichment analysis on the genes within each of the clusters listed in Table S2 (Fig 3B, Table S3). Out of the top 20 enriched terms, 14 terms were directly related to cell cycle processes with three of the other terms (positive regulation of transferase activity, microtubule and protein-DNA complex assembly) still closely clustering to these 14 terms. A number of categories were significantly enriched across all clusters analysed, the most significant of which was 'cell division' (adjusted p-value = 10^{-60}). This is in line with ARHGEF39 as a marker gene for cycling progenitors in G2/M phase. 'G2/M transition' is the most widely and most significantly represented cell cycle phase in these gene lists (adjusted p-value = 10^{-21}). However, the function of ARHGEF39 may not be limited to this phase as several of the involved genes overlap with other cell cycle phases, such as 'mitotic G1 phase and G1/S transition' (adjusted p-value = 10^{-12}) (Table S3). DEGs from the cortical excitatory neuron cluster was not included in the GO enrichment analysis because this cluster produced too few DEGs (N=2). A reason for this low number of DEGs might be that the majority of cells in this cluster of cortical excitatory neurons are post-mitotic. In this cluster, in addition to ARHGEF39, the only other DEG was CDC25C (*Cell Division Cycle 25C*), which is an important cell cycle regulator involved in G2/M progression (Table S2), which further supporting the link between ARHGEF39 and cell cycle.

5.3.6 ARHGEF39 is co-expressed with RhoGEFs and RhoGAPs that predominantly regulate RHOA

Since RhoGTPase regulation is a dynamic and complex process orchestrated by multiple proteins, we also looked for RhoGEF and RhoGAP family members that may function in a regulatory manner with ARHGEF39. Several RhoGEF and RhoGAP family members (N=17) were identified as DEGs (Table S2) and these were all upregulated in ARHGEF39-positive cells (Fig. S2). A short list of RhoGEFs/RhoGAPs that were differentially expressed in at least two of the clusters was integrated with substrate specificity data from a comprehensive

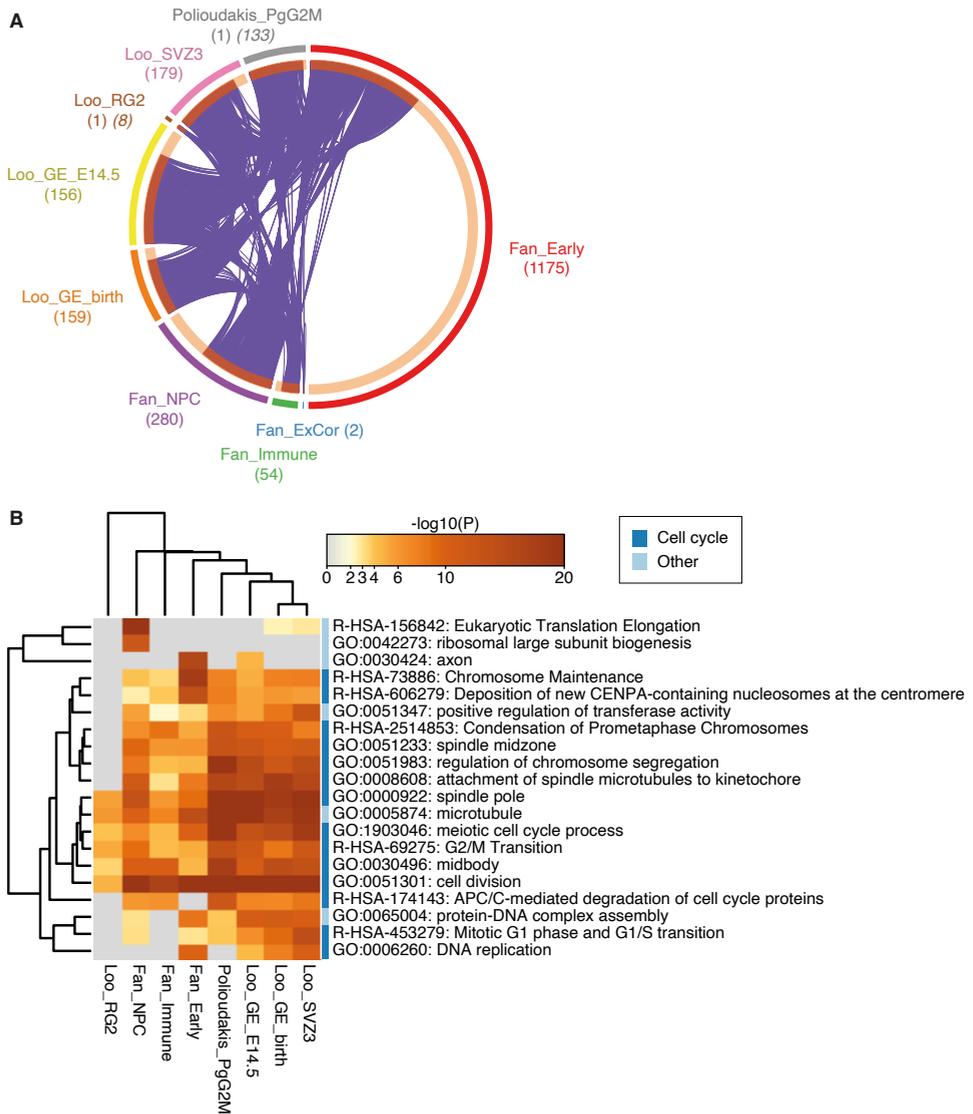


Figure 3 Differential gene expression analysis between ARHGEF39-positive and ARHGEF39-negative cells in ARHGEF39 expressing cell clusters. (A) Overlap between lists of DEGs. Purple curves link identical genes between lists. Genes that hit multiple lists are coloured in dark orange, and genes unique to a list are shown in light orange. Counts of DEGs are between brackets. For Loo_RG2 and Polioudakis_PgG2M, the number of cluster marker genes is added in oblique type. (B) Heatmap of top 20 enriched terms across lists of DEGs, coloured by p-values. Grey indicates a lack of significance. Dendrograms indicate similarity between terms (rows) and lists of DEGs (columns). Terms that are directly related to cell cycle processes are marked with a dark blue bar, other terms are marked with a light blue bar.

FRET-based RhoGTPase activity screen³⁵ to predict if they target the same substrates as ARHGEF39 (See Table 2). For the five RhoGEFs/RhoGAPs with significant results in the substrate specificity assays, four were shown to regulate RHOA, suggesting that RHOA is the most actively regulated RhoGTPase in cell states where ARHGEF39 is expressed. As such, these scRNA-seq data and previously published substrate specificity assays further support our findings from the biosensor assays (Fig. 1B) that RHOA is a substrate of ARHGEF39.

Table 2 RhoGEF and RhoGAP family members that are differentially expressed in at least two clusters. Substrates are listed from a family-wide characterization of substrate specificities of RhoGEFs and RhoGAPs in³⁵. Not all RhoGEFs/RhoGAPs showed significant RhoGTPase activity upregulation or downregulation in this screen and these are indicated with n.s. (not significant).

Protein	Family	Substrate	Clusters	# of clusters
ARHGEF39	RhoGEF	n.s.	Fan_early, Fan_NPC, Polioudakis_PgG2M, Loo_GE_E14.5, Loo_GE_birth, Loo_SVZ3, Fan_Immune, Fan_ExCor, Loo_RG2	9
ECT2	RhoGEF	RHOA	Fan_early, Fan_NPC, Polioudakis_PgG2M, Loo_GE_E14.5, Loo_GE_birth, Loo_SVZ3	6
ARHGAP11A	RhoGAP	RHOA	Fan_early, Fan_NPC, Polioudakis_PgG2M, Loo_GE_E14.5, Loo_GE_birth, Loo_SVZ3	6
RACGAP1	RhoGAP	RAC1	Fan_early, Fan_NPC, Polioudakis_PgG2M, Loo_GE_E14.5, Loo_GE_birth, Loo_SVZ3	6
ARHGAP19	RhoGAP	RHOA	Fan_early, Fan_NPC, Loo_GE_E14.5, Loo_GE_birth, Loo_SVZ3	5
DEPDC1	RhoGAP	n.s.	Fan_early, Fan_NPC, Polioudakis_PgG2M, Fan_Immune	4
ARHGAP11B	RhoGAP	RHOA, CDC42	Fan_early, Fan_NPC, Polioudakis_PgG2M	3
DEPDC1B	RhoGAP	n.s.	Fan_early, Fan_NPC, Polioudakis_PgG2M	3
OPHN1	RhoGAP	n.s.	Fan_early, Fan_NPC	2

5.4 Discussion

ARHGEF39 was implicated in specific language impairment via a functional polymorphism in its 3'UTR that disrupted post-transcriptional expression regulation by microRNAs³. In this study, we identify RHOA as a substrate and downstream effector of ARHGEF39. We show that overexpression of ARHGEF39 disrupts cell adhesion. In the developing cortex, we report that *ARHGEF39* acts a marker for proliferating neural progenitor cells and is significantly co-expressed with genes involved in cell division. RHOA activity, cell de-adhesion, cell division, and neural progenitor cells present new avenues to explore how changes in *ARHGEF39* may contribute to neural development and to language disorder.

The direct activation of RHOA by ARHGEF39 is a novel finding. A family-wide screen of RhoGTPase activation by RhoGEFs using biosensors in HEK293T cells previously did not detect any significant activity of RHOA, RAC1 or CDC42 by ARHGEF39³⁵. This previous study used a different ratio of ARHGEF39, inhibitor and biosensor. We optimised the sensitivity of this assay by determining the lowest effective dose of ARHGDI1 to inhibit activation of the biosensor by endogenous RhoGTPases (Fig. S1). A significant effect on RHOA activation was only observed at the highest ratio (1:5) and not at the lower ratios (1:1 or 1:3). The increased sensitivity due to lower inhibition of the biosensors allowed us to uncover the effect of ARHGEF39 on RHOA using RhoGTPase biosensors. Another previous study exploring the interaction partners of ARHGEF39 identified RAC1, but not RHOA (or CDC42) in a pulldown assay from lung cancer cells overexpressing ARHGEF39¹⁶. A recent study has added that ARHGEF39 is necessary for RAC1 activation during migration of lung cancer cells in response to growth factors¹⁹, but this study did not investigate any potential activation of RHOA. Our study showed that ARHGEF39 directly activates RHOA protein by using FRET-based RhoGTPase activity biosensors that measure activation in living cells. RAC1 activation was not detected, but this could be related to the lower propensity to migrate that HEK293FT cells have compared to lung cancer cells. It has been established that RHOA and RAC1 are mutually inhibitory RhoGTPases during cell migration, and that the activation of RAC1 is preceded by a brief peak in activation of RHOA at the leading edge that initiates protrusion³⁶. This could point to a mechanism by which ARHGEF39 is indirectly involved in the activation of RAC1 by first activating RHOA in migrating cells, but this remains to be tested. The potential relevance of the ARHGEF39-RHOA pathway to neurodevelopment was highlighted by the co-expression of ARHGEF39 with cell division pathways and with other RHOA-regulating proteins in the developing brain (see further discussion, below).

Expression of ARHGEF39 has been shown to promote cell proliferation in cancer cells^{16,20,21}. This role in cell proliferation is also supported for neural cells via our analyses of transcriptome-wide survey of *ARHGEF39*-associated gene expression across cell types in brain development. Combining the *ARHGEF39*-associated genes from multiple cell types presented a core network of genes that are involved in G2/M phase transition. During cell division, RhoGTPase activation is tightly regulated. RHOA and CDC42 are required for specific steps of remodelling the actin and microtubule cytoskeleton, whereas RAC1 must remain inactive during the entire process³⁷. In particular, active RHOA is required during cell rounding, a process in which rigidity of the cell cortex increases and focal adhesions are disassembled³⁸. ARHGEF39 mediated activation of RHOA and subsequent increases in cell rounding could explain the increases in cell detachment observed when ARHGEF39 is overexpressed in HEK293FT cells. Although not yet directly tested, this theory is supported by high-resolution microscopy in COS-7 cells revealing that ARHGEF39 is localised to the plasma membrane and focal adhesions³⁵ - regions where RHOA activity is important for cell rounding³⁸. These findings from cell lines should be validated in neural cells to assess the importance of these processes in neurodevelopment. Notably, several genes in the co-expression network of *ARHGEF39* in neural cell types have functions related to mitotic cell rounding during cell division. ECT2 locally activates RHOA during G2 and M phase and is necessary for proper cell rounding and formation of the mitotic spindle and contractile ring³⁹⁻⁴². RACGAP1 regulates cytokinesis by inactivating RAC1, recruitment of ECT2 and indirect activation of RHOA^{40,42-45}. DEPDC1B promotes disassembly of focal adhesions by

displacing RHOA and makes RHOA available for other processes such as cortical stiffening⁴⁶. Future research will be needed to determine any direct or indirect interactions between ARHGEF39 and these co-expressed genes, and how they form molecular pathways that affect the function of neural cell types. Taken together, these data propose a mechanism by which ARHGEF39 and some of its co-expressed genes converge on RHOA activation and cell division to contribute to neurodevelopment.

We have shown that *ARHGEF39* is enriched in proliferating neural progenitor cells during cortical development. It would be of interest to study the molecular pathways and biological processes mediated by ARHGEF39 in these cell populations to understand its role in neurodevelopment, particularly since changes in RhoGTPase activation have emerged as a molecular hub in various neurodevelopmental disorders^{11,47}. 16p11.2 deletion syndrome is a neurodevelopmental disorder that is characterized by a form of childhood apraxia of speech^{48,49}. Increased RHOA activity is a common feature of cellular and animal models for 16p11.2 deletion syndrome⁵⁰⁻⁵³. In human organoid models with 16p11.2 deletions, increased RHOA activation is observed alongside changes in proliferation, cell adhesion and migration⁵³. In the 16p11.2 deletion mouse model, altered cortical progenitor proliferation leads to an aberrant cortical cytoarchitecture that is characterized by a reduced number of upper layer neurons and increase in layer VI neurons⁵⁴. Focal cortical abnormalities have further been observed in individuals with 16p11.2 deletions by MRI⁵⁵. It would be of interest to determine if mutations in *ARHGEF39* also lead to changes in cortical cytoarchitecture via the predicted changes in RHOA activity and cell division in mice or in humans.

From an evolutionary perspective, the increased proliferative capacity of human neural progenitor cells is considered important for the development of higher cognitive abilities, because of its role in the evolutionary expansion of the neocortex⁵⁶. Human-specific gene *ARHGAP11B* is a prominent member in the network of *ARHGEF39*-associated genes, together with its ancient paralog ARHGAP11A. ARHGAP11B promotes the proliferation and delamination of radial glia cells, which may contribute to neocortical expansion⁵⁷. In order to explore the potential role of *ARHGEF39* in language-related neurodevelopment, it would be relevant to further study the cooperation of ARHGEF39 with ARHGAP11A and ARHGAP11B in neural progenitor cells, their role in cortical development, and how this may have changed over human evolution.

Together, these new insights on the molecular and cellular context of ARHGEF39 provide the bases for defining the role of ARHGEF39 in neurodevelopment. Future research into the effect of ARHGEF39 overexpression in neural progenitor cells on cell division, cell attachment and its co-expression network will be the next step for understanding the neurodevelopmental mechanisms that may be affected by *ARHGEF39* variants.

References

- 1 Reader, R. H., Covill, L. E., Nudel, R. & Newbury, D. F. (2014). Genome-Wide Studies of Specific Language Impairment. *Curr Behav Neurosci Rep* 1, 242-250, doi:10.1007/s40473-014-0024-z.
- 2 Deriziotis, P. & Fisher, S. E. (2017). Speech and Language: Translating the Genome. *Trends Genet* 33, 642-656, doi:10.1016/j.tig.2017.07.002.
- 3 Devanna, P. *et al.* (2018). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol Psychiatry* 23, 1375-1384, doi:10.1038/mp.2017.30.

- 4 Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T. & and the, C.-c. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J Child Psychol Psychiatry* 58, 1068-1080, doi:10.1111/jcpp.12721.
- 5 Fort, P. & Blangy, A. (2017). The Evolutionary Landscape of Dbl-Like RhoGEF Families: Adapting Eukaryotic Cells to Environmental Signals. *Genome Biol Evol* 9, 1471-1486, doi:10.1093/gbe/evx100.
- 6 Etienne-Manneville, S. & Hall, A. (2002). Rho GTPases in cell biology. *Nature* 420, 629-635, doi:10.1038/nature01148.
- 7 Rossman, K. L., Der, C. J. & Sondek, J. (2005). GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat Rev Mol Cell Biol* 6, 167-180, doi:10.1038/nrm1587.
- 8 Nobes, C. D. & Hall, A. (1995). Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* 81, 53-62, doi:10.1016/0092-8674(95)90370-4.
- 9 Govek, E. E., Newey, S. E. & Van Aelst, L. (2005). The role of the Rho GTPases in neuronal development. *Genes Dev* 19, 1-49, doi:10.1101/gad.1256405.
- 10 Pinto, D. *et al.* (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368-372, doi:10.1038/nature09146.
- 11 Zhao, Z. *et al.* (2015). Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. *Mol Psychiatry* 20, 563-572, doi:10.1038/mp.2014.82.
- 12 Nudel, R. *et al.* (2014). Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes Brain Behav* 13, 418-429, doi:10.1111/gbb.12127.
- 13 Kutsche, K. *et al.* (2000). Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat Genet* 26, 247-250, doi:10.1038/80002.
- 14 Ravindran, E. *et al.* (2017). Homozygous ARHGEF2 mutation causes intellectual disability and midbrain-hindbrain malformation. *PLoS Genet* 13, e1006746, doi:10.1371/journal.pgen.1006746.
- 15 De Ligt, J. *et al.* (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 367, 1921-1929, doi:10.1056/NEJMoa1206524.
- 16 Zhou, H. *et al.* (2018). ARHGEF39 promotes tumor progression via activation of Rac1/P38 MAPK/ATF2 signaling and predicts poor prognosis in non-small cell lung cancer patients. *Lab Invest* 98, 670-681, doi:10.1038/s41374-018-0022-y.
- 17 Xu, B., Lv, W., Li, X., Zhang, L. & Lin, J. (2019). Prognostic genes of hepatocellular carcinoma based on gene coexpression network analysis. *J Cell Biochem* 120, 11616-11623, doi:10.1002/jcb.28441.
- 18 Gao, J. & Jia, W. D. (2019). Expression of Rho Guanine Nucleotide Exchange Factor 39 (ARHGEF39) and Its Prognostic Significance in Hepatocellular Carcinoma. *Med Sci Monit* 25, 7826-7835, doi:10.12659/MSM.918270.
- 19 Cooke, M. *et al.* (2021). FARP1, ARHGEF39, and TIAM2 are essential receptor tyrosine kinase effectors for Rac1-dependent cell motility in human lung adenocarcinoma. *Cell Rep* 37, 109905, doi:10.1016/j.celrep.2021.109905.
- 20 Wang, H., Li, Y., Wang, Y., Han, Z. G. & Cai, B. (2012). C9orf100, a new member of the Dbl-family guanine nucleotide exchange factors, promotes cell proliferation and migration in hepatocellular carcinoma. *Mol Med Rep* 5, 1169-1174, doi:10.3892/mmr.2012.783.
- 21 Wang, H. *et al.* (2018). ARHGEF39 promotes gastric cancer cell proliferation and migration via Akt signaling pathway. *Mol Cell Biochem* 440, 33-42, doi:10.1007/s11010-017-3153-3.
- 22 Fritz, R. D. *et al.* (2013). A versatile toolkit to produce sensitive FRET biosensors to visualize signaling in time and space. *Sci Signal* 6, rs12, doi:10.1126/scisignal.2004135.
- 23 Fritz, R. D. *et al.* (2015). SrGAP2-Dependent Integration of Membrane Geometry and Slit-Robo-Repulsive Cues Regulates Fibroblast Contact Inhibition of Locomotion. *Dev Cell* 35, 78-92, doi:10.1016/j.devcel.2015.09.002.
- 24 Martin, K. *et al.* (2016). Spatio-temporal co-ordination of RhoA, Rac1 and Cdc42 activation during prototypical edge protrusion and retraction dynamics. *Sci Rep* 6, 21901, doi:10.1038/srep21901.
- 25 Loo, L. *et al.* (2019). Single-cell transcriptomic analysis of mouse neocortical development. *Nat Commun* 10, 134, doi:10.1038/s41467-018-08079-9.
- 26 Polioudakis, D. *et al.* (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* 103, 785-801 e788, doi:10.1016/j.neuron.2019.06.011.
- 27 Fan, X. *et al.* (2020). Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci Adv* 6, eaaz2978, doi:10.1126/sciadv.aaz2978.
- 28 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36, 411-420, doi:10.1038/nbt.4096.

- 29 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140, doi:10.1093/bioinformatics/btp616.
- 30 Zhou, Y. *et al.* (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10, 1523, doi:10.1038/s41467-019-09234-6.
- 31 Ridley, A. J. & Hall, A. (1992). The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* 70, 389-399, doi:10.1016/0092-8674(92)90163-7.
- 32 Kiselev, V. Y., Andrews, T. S. & Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat Rev Genet* 20, 273-282, doi:10.1038/s41576-018-0088-9.
- 33 Kang, H. J. *et al.* (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483-489, doi:10.1038/nature10523.
- 34 Li, M. *et al.* (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362, doi:10.1126/science.aat7615.
- 35 Muller, P. M. *et al.* (2020). Systems analysis of RhoGEF and RhoGAP regulatory proteins reveals spatially organized RAC1 signalling from integrin adhesions. *Nat Cell Biol* 22, 498-511, doi:10.1038/s41556-020-0488-x.
- 36 Machacek, M. *et al.* (2009). Coordination of Rho GTPase activities during cell protrusion. *Nature* 461, 99-103, doi:10.1038/nature08242.
- 37 Chircop, M. (2014). Rho GTPases as regulators of mitosis and cytokinesis in mammalian cells. *Small GTPases* 5, doi:10.4161/sgtp.29770.
- 38 Maddox, A. S. & Burridge, K. (2003). RhoA is required for cortical retraction and rigidity during mitotic cell rounding. *J Cell Biol* 160, 255-265, doi:10.1083/jcb.200207130.
- 39 Tatsumoto, T., Xie, X., Blumenthal, R., Okamoto, I. & Miki, T. (1999). Human ECT2 is an exchange factor for Rho GTPases, phosphorylated in G2/M phases, and involved in cytokinesis. *J Cell Biol* 147, 921-928, doi:10.1083/jcb.147.5.921.
- 40 Yuce, O., Piekny, A. & Glotzer, M. (2005). An ECT2-centralspindlin complex regulates the localization and function of RhoA. *J Cell Biol* 170, 571-582, doi:10.1083/jcb.200501097.
- 41 Niiya, F., Tatsumoto, T., Lee, K. S. & Miki, T. (2006). Phosphorylation of the cytokinesis regulator ECT2 at G2/M phase stimulates association of the mitotic kinase Plk1 and accumulation of GTP-bound RhoA. *Oncogene* 25, 827-837, doi:10.1038/sj.onc.1209124.
- 42 Matthews, H. K. *et al.* (2012). Changes in Ect2 localization couple actomyosin-dependent cell shape changes to mitotic progression. *Dev Cell* 23, 371-383, doi:10.1016/j.devcel.2012.06.003.
- 43 Minoshima, Y. *et al.* (2003). Phosphorylation by aurora B converts MgcRacGAP to a RhoGAP during cytokinesis. *Dev Cell* 4, 549-560, doi:10.1016/s1534-5807(03)00089-3.
- 44 Bastos, R. N., Penate, X., Bates, M., Hammond, D. & Barr, F. A. (2012). CYK4 inhibits Rac1-dependent PAK1 and ARHGEF7 effector pathways during cytokinesis. *J Cell Biol* 198, 865-880, doi:10.1083/jcb.201204107.
- 45 Zhang, D. & Glotzer, M. (2015). The RhoGAP activity of CYK-4/MgcRacGAP functions non-canonically by promoting RhoA activation during cytokinesis. *Elife* 4, doi:10.7554/eLife.08898.
- 46 Marchesi, S. *et al.* (2014). DEPDC1B coordinates de-adhesion events and cell-cycle progression at mitosis. *Dev Cell* 31, 420-433, doi:10.1016/j.devcel.2014.09.009.
- 47 Zamboni, V. *et al.* (2018). Rho GTPases in Intellectual Disability: From Genetics to Therapeutic Opportunities. *Int J Mol Sci* 19, doi:10.3390/ijms19061821.
- 48 Mei, C. *et al.* (2018). Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion. *Eur J Hum Genet* 26, 676-686, doi:10.1038/s41431-018-0102-x.
- 49 Chung, W. K., Roberts, T. P., Sherr, E. H., Snyder, L. G. & Spiro, J. E. (2021). 16p11.2 deletion syndrome. *Curr Opin Genet Dev* 68, 49-56, doi:10.1016/j.gde.2021.01.011.
- 50 Escamilla, C. O. *et al.* (2017). Kctd13 deletion reduces synaptic transmission via increased RhoA. *Nature* 551, 227-231, doi:10.1038/nature24470.
- 51 Martin Lorenzo, S., Nalesso, V., Chevalier, C., Birling, M. C. & Heralut, Y. (2021). Targeting the RHOA pathway improves learning and memory in adult Kctd13 and 16p11.2 deletion mouse models. *Mol Autism* 12, 1, doi:10.1186/s13229-020-00405-7.
- 52 Sundberg, M. *et al.* (2021). 16p11.2 deletion is associated with hyperactivation of human iPSC-derived dopaminergic neuron networks and is rescued by RHOA inhibition in vitro. *Nat Commun* 12, 2897, doi:10.1038/s41467-021-23113-z.
- 53 Urresti, J. *et al.* (2021). Cortical organoids model early brain development disrupted by 16p11.2 copy number variants in autism. *Mol Psychiatry* 26, 7560-7580, doi:10.1038/s41380-021-01243-6.
- 54 Pucilowska, J. *et al.* (2015). The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway. *J Neurosci* 35, 3190-

- 3200, doi:10.1523/JNEUROSCI.4864-13.2015.
- 55 Blackmon, K. *et al.* (2018). Focal Cortical Anomalies and Language Impairment in 16p11.2 Deletion and Duplication Syndrome. *Cereb Cortex* 28, 2422-2430, doi:10.1093/cercor/bhx143.
- 56 Kalebic, N. & Huttner, W. B. (2020). Basal Progenitor Morphology and Neocortex Evolution. *Trends Neurosci* 43, 843-853, doi:10.1016/j.tins.2020.07.009.
- 57 Florio, M. *et al.* (2015). Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* 347, 1465-1470, doi:10.1126/science.aaa1975.

Supplementary material

Figure S1. Titration of ARHGDIA for effective amount to increase dynamic range of FRET-based RhoGTPase biosensor assay.

Figure S2. Developmental mRNA expression of ARHGEF39 in post-mortem samples from BrainSpan Developmental Transcriptome and PsychENCODE Human Brain Development.

Figure S3. Overlap between lists of DEGs.

Figure S4. Log₂ fold changes for differentially expressed RhoGEFs and RhoGAPs.

Table S1. Counts and percentages of ARHGEF39 positive cells for all clusters in collected datasets.

Table S2. Lists of DEGs per cluster and cluster marker genes for Loo_RG2 and Polioudakis_PgG2M.

Table S3. Core set of ARHGEF39-related genes with DE logFoldChanges and assigned PsychEncode developmental WGCNA modules Li, et al. (2018)³⁴

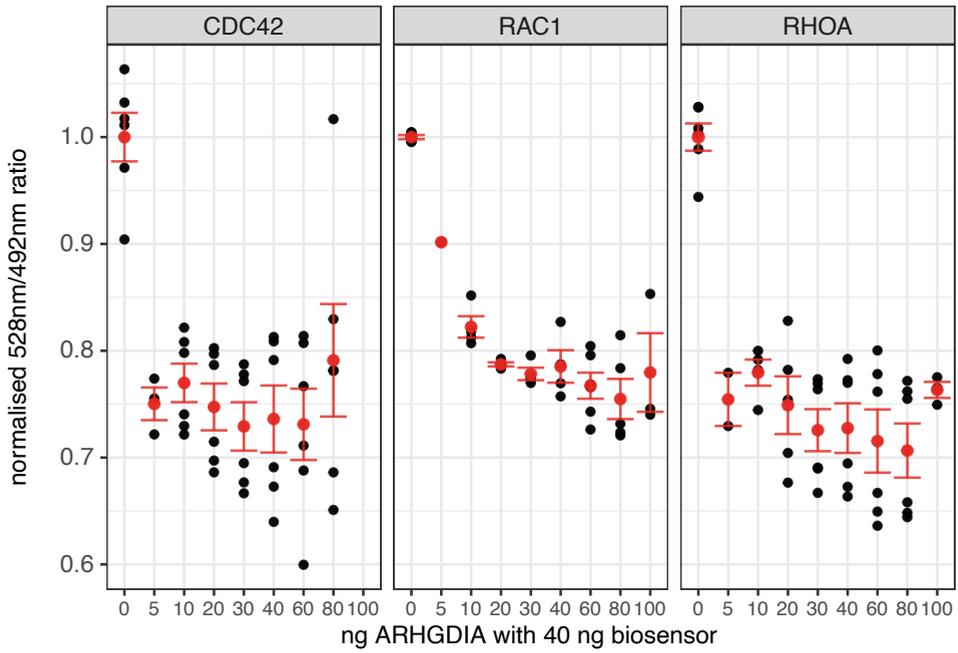


Figure S1 Titration of ARHGDI A for effective amount to increase dynamic range of FRET-based RhoGTPase biosensor assay

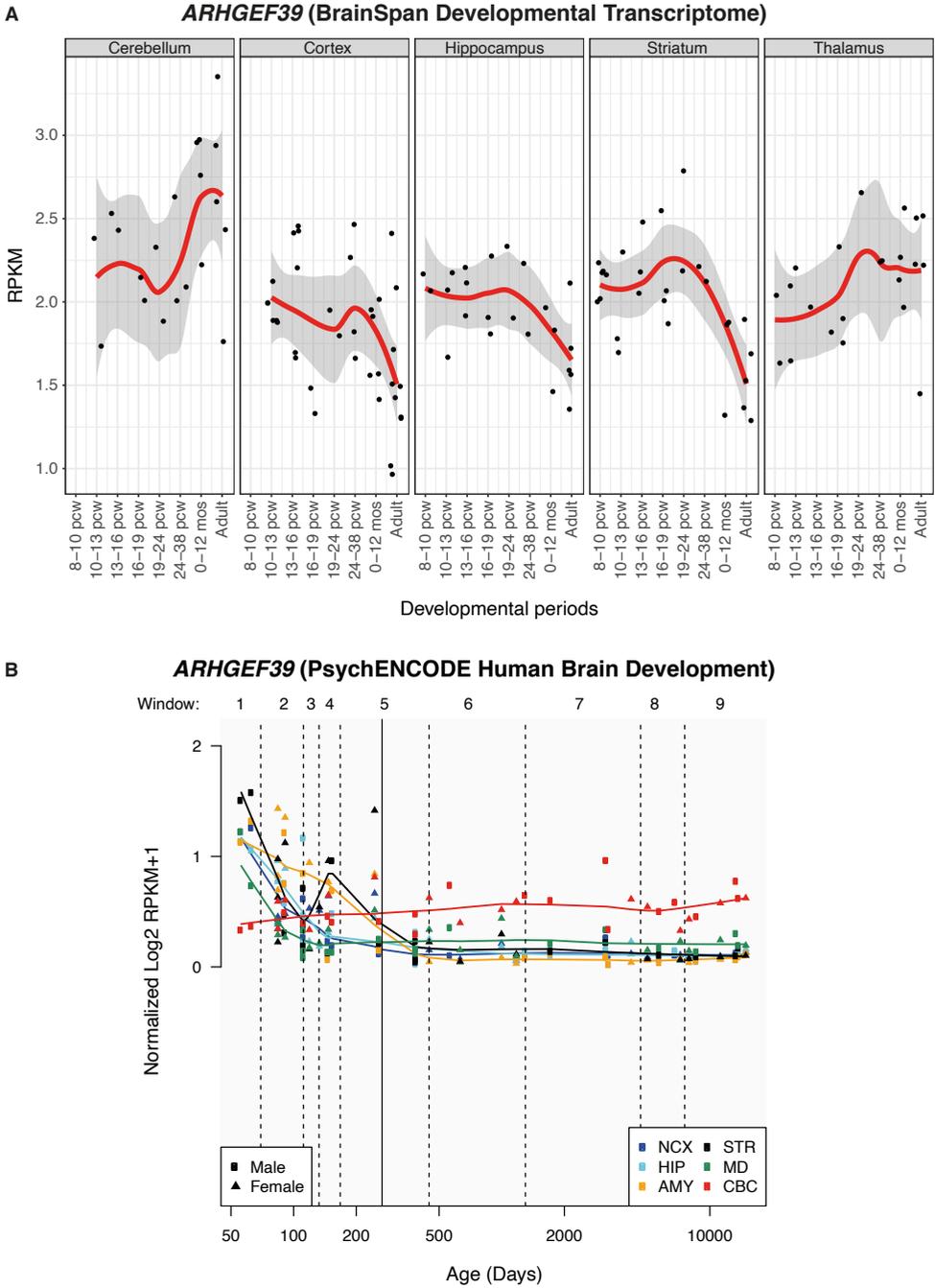


Figure S2 Developmental mRNA expression of *ARHGEF39* in post-mortem samples from (A) BrainSpan Developmental Transcriptome and (B) PsychENCODE Human Brain Development.

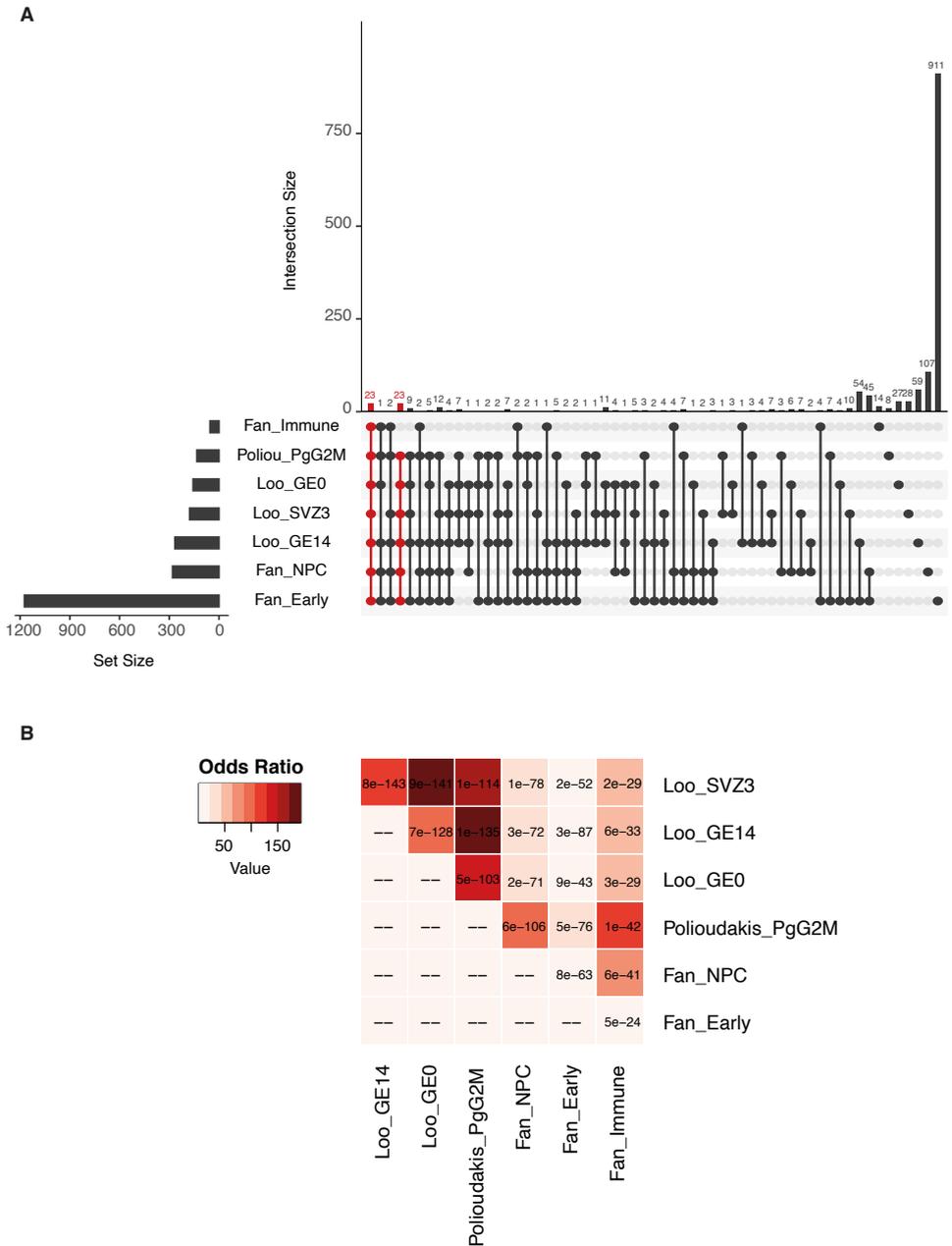


Figure S3 Overlap between lists of DEGs. (A) Upset plot of all intersections in ARHGEF39-associated gene lists. Core gene set is indicated in red. These are genes shared by Polioudakis_PgG2M, Loo_GE0, Loo_GE14, Loo_SVZ3, Fan_NPC and Fan_Early. (B) Gene overlap matrix. Odds ratios in shades of red indicate the strength of association between two gene lists. Fisher's exact test p-values are superimposed on the grid.

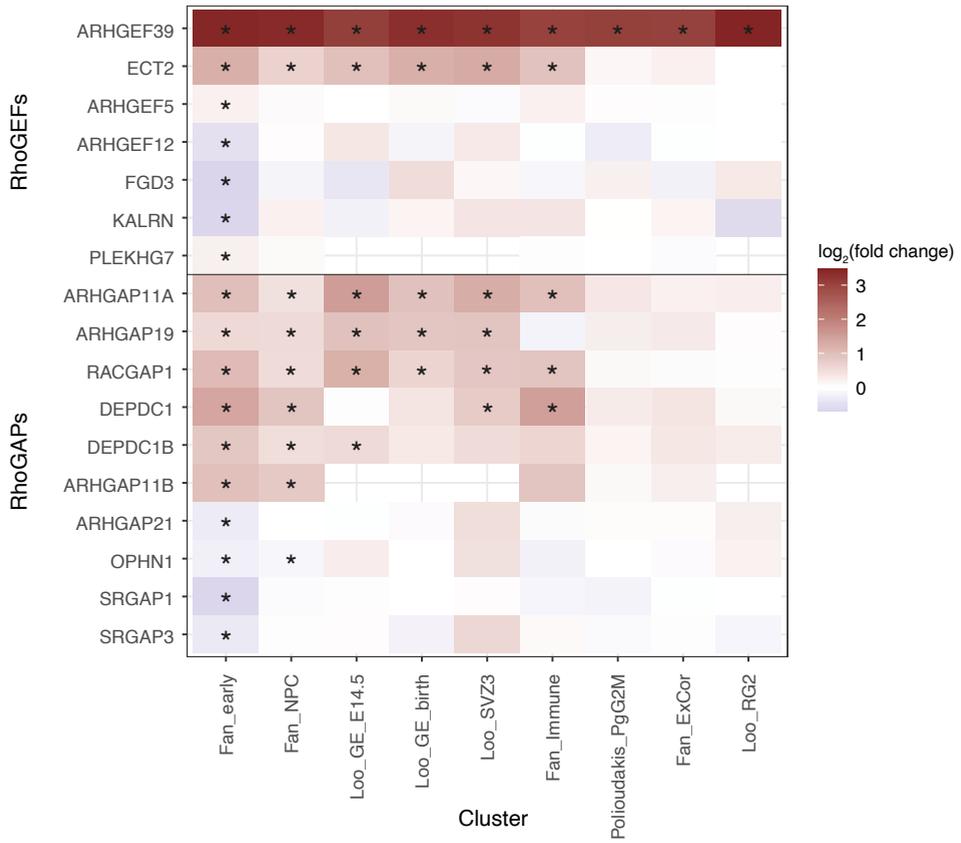


Figure S4 \log_2 fold changes for differentially expressed RhoGEFs and RhoGAPs.

* indicates FDR-corrected p-value < 0.01 in DEG analysis

Table S1 Counts and percentages of ARHGFE39 positive cells for all clusters in collected datasets.

Dataset	Cluster name	Number of cells	Number of cells (ARHGFE39 > 0)	Fraction ARHGFE39 > 0	Average expression (ARHGFE39 > 0)
Polioudakis2019	Cycling progenitors (G2/M phase)	695	141	0.20	1.45
Polioudakis2019	Cycling progenitors (S phase)	1232	64	0.05	1.44
Polioudakis2019	Oligodendrocyte progenitor cells (OPC)	306	10	0.03	1.43
Polioudakis2019	Ventricular radial glia (vRG)	984	10	0.01	1.43
Polioudakis2019	Outer radial glia (oRG)	1293	10	0.01	1.41
Polioudakis2019	Intermediate progenitors (IP)	2150	7	0.00	2.12
Polioudakis2019	Interneuron MGE (InMGE)	1705	4	0.00	2.05
Polioudakis2019	Maturing excitatory (ExM)	9822	20	0.00	1.84
Polioudakis2019	Excitatory deep layer 1 (ExDp1)	2039	4	0.00	1.34
Polioudakis2019	Migrating excitatory (ExN)	9995	15	0.00	2.12
Polioudakis2019	Interneuron CGE (InCGE)	1434	2	0.00	1.40
Polioudakis2019	Maturing excitatory upper enriched (ExM-U)	1756	1	0.00	1.68
Polioudakis2019	Exitatory deep layer 2 (ExDp2)	166	0	0.00	NA
Polioudakis2019	Microglia (Mic)	48	0	0.00	NA
Polioudakis2019	Pericyte (Per)	114	0	0.00	NA
Polioudakis2019	Endothelial (End)	237	0	0.00	NA
Loo2019-E14	RG2 [14-E]	334	152	0.46	1.31
Loo2019-E14	SVZ3 (proliferating) [15-E]	315	70	0.22	1.23
Loo2019-P0	Ganglionic eminences [9-P]	421	72	0.17	1.41
Loo2019-E14	Ganglionic eminences [6-E]	762	79	0.10	1.35
Loo2019-E14	RG3 (cortical hem) [21-E]	42	4	0.10	1.33
Loo2019-P0	Endothelial2 [21-P]	71	4	0.06	1.43
Loo2019-E14	RG1 [8-E]	614	34	0.06	0.97
Loo2019-P0	Choroid plexus [20-P]	85	4	0.05	1.16
Loo2019-E14	Choroid plexus [22-E]	27	1	0.04	0.90
Loo2019-E14	Endothelial [18-E]	87	3	0.03	0.94
Loo2019-E14	Microglia [20-E]	31	1	0.03	2.30
Loo2019-E14	RG4 [10-E]	604	12	0.02	1.42
Loo2019-P0	Microglia [22-P]	71	1	0.01	1.67
Loo2019-E14	Layer1 [17-E]	149	2	0.01	0.96

Dataset	Cluster name	Number of cells	Number of cells (ARHGEF39 > 0)	Fraction ARHGEF39 > 0	Average expression (ARHGEF39 > 0)
Loo2019-P0	Astrocytes (immature) 2 [13-P]	260	3	0.01	1.45
Loo2019-E14	Striatal inh1 [16-E]	263	3	0.01	0.84
Loo2019-P0	LayerI [19-P]	93	1	0.01	1.23
Loo2019-P0	LayerV-VI [12-P]	226	2	0.01	1.76
Loo2019-E14	LayerV-VI [3-E]	1068	8	0.01	1.49
Loo2019-E14	LayerV-VI [13-E]	510	3	0.01	0.89
Loo2019-E14	LayerV-VI [5-E]	804	4	0.00	0.95
Loo2019-P0	Astrocytes (immature) 1 [10-P]	417	2	0.00	1.19
Loo2019-P0	SVZ2 (migrating) [8-P]	436	2	0.00	1.56
Loo2019-P0	Oligodendrocytes [16-P]	219	1	0.00	1.39
Loo2019-P0	Int1 [5-P]	482	2	0.00	1.93
Loo2019-P0	Int2 [14-P]	244	1	0.00	1.87
Loo2019-E14	SVZ1 (migrating) [4-E]	847	3	0.00	0.98
Loo2019-P0	SVZ1 (migrating) [2-P]	577	2	0.00	1.38
Loo2019-E14	SVZ2 (VZ-SVZ) [11-E]	595	2	0.00	1.17
Loo2019-P0	LayerII-IV [1-P]	903	3	0.00	1.32
Loo2019-E14	Int1 [1-E]	1022	3	0.00	1.48
Loo2019-P0	Int4 [6-P]	406	1	0.00	1.95
Loo2019-P0	Striatal inh2 [7-P]	417	1	0.00	1.90
Loo2019-E14	Int2 [12-E]	474	1	0.00	0.59
Loo2019-P0	LayerII-IV [4-P]	571	1	0.00	1.40
Loo2019-E14	Striatal inh2 [9-E]	615	1	0.00	1.59
Loo2019-E14	LayerV-VI [7-E]	661	1	0.00	1.24
Loo2019-E14	LayerV-VI [2-E]	773	1	0.00	2.39
Loo2019-E14	Thalamic [19-E]	145	0	0.00	NA
Loo2019-P0	Endothelial1 [17-P]	121	0	0.00	NA
Loo2019-P0	Int3 [11-P]	287	0	0.00	NA
Loo2019-P0	Layer II-IV [15-P]	113	0	0.00	NA
Loo2019-P0	LayerV-VI [18-P]	178	0	0.00	NA
Loo2019-P0	Striatal inh1 [3-P]	513	0	0.00	NA
Fan2020	NPC	1333	404	0.30	1.32

Dataset	Cluster name	Number of cells	Number of cells (ARHGEF39 > 0)	Fraction ARHGEF39 > 0	Average expression (ARHGEF39 > 0)
Fan2020	Early	1392	306	0.22	1.36
Fan2020	Immune	511	78	0.15	1.05
Fan2020	EX_cor	2065	214	0.10	0.95
Fan2020	PONS_neu	1163	99	0.09	0.84
Fan2020	Blood	116	9	0.08	1.06
Fan2020	Astro	104	8	0.08	1.04
Fan2020	Oligo	66	5	0.08	1.39
Fan2020	Endo	57	4	0.07	0.85
Fan2020	IN_cor	72	5	0.07	1.73
Fan2020	CR	49	2	0.04	1.37

Table S2 Lists of DEGs per cluster and cluster marker genes for Loo_RG2 and Polioudakis_PgG2M.

Dataset	DEGs / cluster marker genes
Loo_RG2	ARHGEF39 CDC25C ASPM NEK2 PLK1 NDE1 CDKN3 GAS2L3
Loo_SVZ3	ARHGEF39 CDC25C ASPM NEK2 PLK1 NDE1 INCENP KIF11 HMGB2 CCNB1 SMC2 FAM83D CKS1B PRC1 MCM2 TOP2A NUF2 NME1 KNSTRN MCM3 SGO2 KIF20A BIRC5 KIF20B NUCKS1 CENPA CENPE SMC4 CDK1 TUBB4B CKAP5 CDT1 CDC20 AURKA PIMREG SGO1 CALM2 HSP90AB1 RAD21 RPA2 HMMR CDKN3 HJURP NCAPG CENPF CDCA8 NOTCH1 RACGAP1 AKIRIN2 KIF23 BRD8 SPC25 MIIP CKAP2L ARHGAP11A H2AX TACC3 MIS18BP1 UNG KIFC1 MKI67 RANGAP1 KIF22 GAS2L3 ECT2 NUSAP1 PSRC1 CEP55 KNL1 KIF2C TRIM59 BUB3 TUBA1C TPX2 CCNB2 UBB CENPL CCNA2 CKAP2 MCM6 ARL6IP1 CDCA3 HELLS UBE2C CHAF1B BUB1 MXD3 HYL51 ANLN RPS26 SIVA1 PIF1 BUB1B SLC25A5 PAICS HSP90B1 KIF18A SAPCD2 SPAG5 PCNA CDC25B RPS5 CCP110 MCM4 KIF4A FZR1 POC5 G2E3 CKS2 RTKN2 CENPC NPM1 DLGAP5 TTK SKA2 LMNB1 BORA NUP37 VBP1 UBALD2 DBF4 SLFN13 CIP2A CDC6 ODF2 MAD2L1 CLIC1 CLSPN RPS2 PABPC1 SDC1 ANK3 RANBP1 ILF3 RPA1 LIG1 CHAF1A TUBB GPD2 CCDC112 FAM111A CCND2 PRDX1 HBB CASP8AP2 TYMS SLC17A6 HES1 ZC3H7A NEDD4L CCND1 PNN MDK PIK3R3 DBI RPLP2 RPS24 CDCA7 RPS3 CCNE2 LH4 KLF6 SORBS2 DUT RPLP1 GMMN GATD3A SPDL1 UMPS KIF18B LPAR1 DHFR ATAD2 NICN1 SRSF10 TIPIN TNIK
Loo_GE14	ARHGEF39 CDC25C ASPM NEK2 PLK1 NDE1 INCENP KIF11 HMGB2 CCNB1 SMC2 FAM83D CKS1B PRC1 MCM2 TOP2A NUF2 NME1 KNSTRN MCM3 SGO2 KIF20A BIRC5 KIF20B NUCKS1 CENPA CENPE SMC4 CDK1 TUBB4B CKAP5 CDT1 CDC20 AURKA PIMREG SGO1 CALM2 HSP90AB1 RAD21 RPA2 HMMR CDKN3 HJURP NCAPG CENPF CDCA8 NOTCH1 RACGAP1 AKIRIN2 KIF23 BRD8 SPC25 MIIP CKAP2L ARHGAP11A H2AX TACC3 MIS18BP1 UNG KIFC1 MKI67 RANGAP1 KIF22 GAS2L3 ECT2 NUSAP1 PSRC1 CEP55 KNL1 KIF2C TRIM59 BUB3 TUBA1C TPX2 CCNB2 UBB CENPL CCNA2 CKAP2 MCM6 ARL6IP1 CDCA3 HELLS UBE2C CHAF1B BUB1 MXD3 HYL51 ANLN PIF1 KIF14 BUB1B TROAP PBK PAICS MCM5 KIF18A H110 SAPCD2 SPAG5 CDC25B WNK1 CCP110 MCM4 KIF4A FZR1 G2E3 NCAPD2 RTKN2 CDCA2 DLGAP5 TTK TERF1 SKA2 LMNB1 SKA3 GPSM2 BORA AURKB NUP37 VBP1 MIAT NDC80 UBALD2 DBF4 CIP2A CDC6 CDKN2D ANP32E ODF2 MAD2L1 GINS2 SKA1 TUBA1B SYT4 PCDHA2 RAB2A MPPED1 MRPL51 PNRC2 THRA PKM PTPRD H2AZ2 CCNG2 P4HA3 HSPH1 NR2E1 TMEM132A SPC24 TUBB3 REEP5 RNASEH2C RUNX1T1 TMPO NDRG4 MNS1 MIS18A NREP H3C3 USP1 MLX NOVA2 KIF15 NSG1 NPDC1 SAE1 MT3 RBFOX2 SRRM4 VIM SERF1A SERF1A SERF1B MAPT ERI2 RAB3C SCG5 KPNA2 LIPG PRDX6 TEAD2 OIP5 NCAPH SOX9 MAP2 PHGDH CCDC18 SOX4 FIGN SNAP25 NDUFAF3 RTN1 KIF5A TNC RGS20 MAP1LC3A KIF5C PPP2R5C SERPINI1 XPR1 PSAT1 SPRED1 SH3BP5 STMN2 PEA15 RUFY3 TMSB10 SCG3 OLFM1 STXBP1 MSH2 MEG3 MELK DTL MGST3 SLC25A4 CDKN1C SLC1A3 PODXL2 POU3F2 PTPN5 ANKLE1 TTYH1 PLPP3 S1PR1 NNAT TUBA1A NNAT PAFAH1B3 CDK5RAP2 RNF14 MAP1B LM03 DYNLL1 MASTL SCN3A NUDCD2 STAG2 TRAP PRDX2 PRR11 SDC3 SRSF3 PAK1 ARHGAP19 LBR NSG2 TAGLN3 PCF11 KLF7 CCDC77 NAA50 SHCBP1 SFRP1 PARP6 UBE2T LY6H DDX39A TMOD2 RAB3A CEP295 MAP9 L1CAM STMN3 TTC3 MYT1L MLT11
Loo_GEO	ARHGEF39 CDC25C ASPM NEK2 PLK1 NDE1 INCENP KIF11 HMGB2 CCNB1 SMC2 FAM83D CKS1B PRC1 MCM2 TOP2A NUF2 NME1 KNSTRN MCM3 SGO2 KIF20A BIRC5 KIF20B NUCKS1 CENPA CENPE SMC4 CDK1 TUBB4B CKAP5 CDT1 CDC20 AURKA PIMREG SGO1 CALM2 HSP90AB1 RAD21 RPA2 HMMR HJURP NCAPG CENPF CDCA8 NOTCH1 RACGAP1 AKIRIN2 KIF23 BRD8 SPC25 MIIP CKAP2L ARHGAP11A H2AX TACC3 MIS18BP1 UNG KIFC1 MKI67 RANGAP1 KIF22 ECT2 NUSAP1 PSRC1 CEP55 KNL1 KIF2C TRIM59 BUB3 TUBA1C TPX2 CCNB2 UBB CENPL CCNA2 CKAP2 MCM6 ARL6IP1 CDCA3 HELLS UBE2C CHAF1B BUB1 MXD3 RPS26 SIVA1 KIF14 SLC25A5 TROAP PBK HSP90B1 MCM5 H110 PCNA WNK1 RPS5 POC5 NCAPD2 CKS2 CENPC NPM1 CDCA2 TERF1 SKA3 GPSM2 AURKB MIAT NDC80 SLFN13 CDKN2D ANP32E CLIC1 GINS2 CLSPN SKA1 N1010 CCDC88A ZNF367 CHEK1 MCM7 IGF2 CIT RPLP0 NOP58 RPL4 E2F1 REEP4 TIMP1 TMEM100 UHRF1 EXO1 GSX1 DDHD1 IGFBP2 PARPBP SRM CHCHD1 SMTN RPS14 FKBP2 SNX8 ID3 GNL3 PPP1R14B RPL32 APRT GADD45G BCR1 GASE5 SULT1A1 COL22A1 TIMELESS TIGD2 HSP E1 BTG2
Polioudakis_PgG2M	ACBD7 ANLN ANP32E ARHGAP11A ARHGAP11B ARHGEF39 ARL6IP1 ASPM AURKA AURKB BIRC5 BORA BUB1 BUB1B C21orf58 CCDC18 CCNA1 CCNA2 CCNB1 CCNB2 CDC20 CDC25B CDC25C CDCA2 CDCA3 CDCA8 CDK1 CDKN2C CDKN3 CENPA CENPE CENPF CENPN CENPU CEP55 CEP70 CIT CKAP2 CKAP2L CKS1B CKS2 CRNDE DBF4 DEPDC1 DEPDC1B DIAPH3 DLGAP5 ECT2 ENSG00000280206 FAM111A PIMREG FAM83D FBXO5 FOXM1 GAS2L3 GTSE1 H2AZ2 H2AX HJURP HMGB2 HMGN2 HMMR INCENP CIP2A KIF11 KIF14 KIF15 KIF18A KIF20A KIF20B KIF22 KIF23 KIF2C KIF4A KIFC1 KNL1 KNSTRN KPNA2 MAD2L1 MELK MIS18BP1 MKI67 MXD3 MYBL1 NCAPD2 NCAPG NCAPH NDC80 NDE1 NEIL3 NEK2 NMU NUF2 NUSAP1 OIP5 PARPBP PBK PIF1 PLK1 PRC1 PRR11 PSRC1 PTTG1 PTTG3P RACGAP1 RTKN2 SAPCD2 SGO1 SGO2 SHCBP1 SKA3 SMC2 SMC4 SPA17 SPAG5 SPC24 SPC25 SPDL1 STIL TACC3 TMPO TOP2A TPX2 TRIM59 TROAP TTK TUBA1B TUBB4B TUBB6 UBE2C UBE2S UBE2T ZWINT
Fan_NPC	ARHGEF39 SPC25 MALAT1 PRC1 KIF15 KIF20A CDC25C CCNB2 CENPA DEPDC1 CDCA3 KNL1 MXD3 KIF23 CDK1 TMSB4X MKI67 HJURP ASPM NCAPG NCAPH FTL CCNB1 AURKB PLK1 BIRC5 FTH1 CCNA2 BUB1 ANLN CDCA8 KIF4A CDC20 NEIL3 MAP1B AURKA KIF11 MAGED1 RPL7 SAPCD2 NDC80 NDUFA4 GAS2L3 COX7C SGO2 SUMO2 SGO1 PARPBP ESPL1 FABP7 RACGAP1 LUC7L3 RNF26 EEF1A1 RTKN2 HMMR CEP70 SRP14 GINS2 KPNA2 MYL6 ACTG1 RPL18 HSP90AB1 KMT5A PBK EIF1 TBC1D31 RPL27A KIF20B KIF14 RPL6 G2E3 HYL51 SPDL1 SHCBP1 RPL5 RANG AP1 MZT1 TACC3 FOXM1 RPL37A RPS13 ARHGAP11A CALM1 CENPF ARHGAP11B

Dataset	DEGs / cluster marker genes
Fan_NPC <i>(continued)</i>	H33B MDK PTMA CDKN3 CIT APOLD1 NDE1 NNAT TMSB10 EEF1G SKP1 CIP2A SPAG5 NEK2 CDCA2 CCNF SMTN CENPE RPL10 RPL15 DBF4 KIF2C LOC100288637 SON BEX3 KIF18A NMU NCAPD2 PSRC1 SNRNP CFL1 CKAP2L DBF4B KNSTRN RPS3A NUF2 CCDC18 GTSE1 KIF18B CDC25B CKAP5 SOX4 HSBP1 PRR11 CKAP2 CCNA1 CEP55 DEPDC1B SKA1 FAM83D RPS27 CENPL BUB1B ECT2 PIMREG DLGAP5 ACTB RPL34 PIF1 HSP90AA1 OIP5 GABARAP SERF2 PKM GNG5 MIR91HG PEBP1 RPL30 SNRPE HSPB1 RPL35A RPS27A WDR5BDT LRRC37A4P RPS15 RPL3 RPS9 RPS19 HMG2N RANBP1 PRDX5 RPL24 FABP5 NPM1 EID1 HNRNPC FAM160A2 CKB RPL27 RPL9 TMA7 NLGN2 ELOB RPL7A FAU RPS6 RPL14 RPS7 RPS4X RACK1 ENO1 RPL12 TMSB15A RPSA RPL39 RPL23A MCM5 RPS23 RPS24 RPL38 RPS14 RPS3 CDKN2C RPL18A NASP SEPTIN11 ALDOA NHP2 HSPA8 RPS8 NME1 GSTP1 ATP5PO RPS5 RPS15A CHTF18 COX6B1 PGAM1 POC5 SOD1 CNBP RPS29 RPL31 MCM7 RPS2 RPL36 RPL41 RPS12 CHCHD2 DNMT1 COMMD6 HSPD1 RPL4 CRNDE PPA1 RPL10A RPL13 PPIA RPS20 RPL37 NCL DBI NACA RPLP2 EDF1 RPL29 EIF4EBP1 RPL23 RPL32 SUB1 RPLP1 FKBP3 SNRPD1 HINT1 PFDN5 BTF3 RPS11 PLEKHA8P1 RPL35 GAPDH RPS16 RPS25 RPS18 SLC4A8 PFI1 HMGB1 LDHB RPL19 PSMAT7 RPLP0 RPL26 HSD17B11 ATP5F1A SELENOH CDK4 IGSF9 ATP5MC2 PRDX2 RPL11 RPL8 CUTA RPS26 HSPE1
Fan_Early	ARHGEF39 SPC25 MALAT1 PRC1 KIF15 KIF20A CDC25C CCNB2 CENPA DEPDC1 CDCA3 KNL1 MXD3 KIF23 CDK1 TMSB4X MKI67 HJURP ASPM NCAPG NCAPH FTL CCNB1 AURKB PLK1 BIRC5 FTH1 CCNA2 BUB1 ANLN CDCA8 KIF4A CDC20 NEIL3 MAP1B AURKA KIF11 MAGED1 RPL7 SAPCD2 NDC80 NDUFA4 GAS2L3 COX7C SGO2 SUMO2 SGO1 PARPBP ESPL1 FABP7 RACGAP1 LUC7L3 RNF26 EEF1A1 RTKN2 HMHR CEP70 SRP14 GINS2 KPN2 MYL6 ACTG1 RPL18 HSP90A1 KMT5A PBK EIF1 TBC1D31 RPL27A KIF20B KIF14 RPL6 CENPO G2E3 HBG2 HYL51 SPDL1 SHCBP1 RPL5 RANGAP1 MZT1 TACC3 FOXM1 RPL37A RPS13 ARHGAP11A CALM1 CENPF ARHGAP11B H33B MDK PTMA CDKN3 CIT APOLD1 NDE1 NNAT TMSB10 EEF1G SKP1 CIP2A SPAG5 NEK2 CDCA2 CCNF SMTN CENPE RPL10 RPL15 DBF4 KIF2C LOC100288637 SON BEX3 KIF18A NMU NCAPD2 PSRC1 SNRNP CFL1 CKAP2L DBF4B KNSTRN RPS3A NUF2 CCDC18 GTSE1 KIF18B CDC25B CKAP5 SOX4 HSBP1 PRR11 CKAP2 CCNA1 CEP55 DEPDC1B SKA1 FAM83D RPS27 CENPL BUB1B ECT2 PIMREG DLGAP5 ACTB RPL34 PLK4 PIF1 SPP1 HSP90AA1 OIP5 GABARAP CEP83 DR1 PMAIP1 PDP1 DDX5 PTPRD ATP6V0D1 LINC02043 BMPR2 AKR1C1 SEC11C PPP2R1A PPIF RPL39L H15 ATPAF2 SPAG9 SVOP FAME118A ATP6V0A1 SV2A H2AZ2 ANP32E ETV4 BEX4 IGFBP1 PRKAR1A SNGC RUNX1T1 MTERF3 NGRN JADE1 RAD51B SOCS1 ITFG1 SVBP KLHL35 DDX24 ATP5F1B CRIM1 TICRR CACNA2D1 GRM2 SMIM19 RAB6A NRSN1 RUVBL2 GLO1 HERC1 LSAMP RAD51 FGF17 ST7 NR2E1 SDHD DLGAP4 RFWD3 MAPT EAVL4 MRPS15 SCG5 PAK3 IFT74 PXMP2 POLD1 KIF22 MTRF2 HES4 TERF2IP AFDN PRDX6 E2F8 SH3BGR13 KIF3A TMEM138 CCDC112 LINC02367 LINC01873 MECP2 SCAMP1 FBXO5 ESCO2 KLF7 CCDC77 TMTSF2 PBX3 MARCHF6 MRPL3 MNMAT2 CEP128 SHISA2 NXT2 PGAM2 NAPB BEX5 SNRNP25 DDX17 TERT SNAPC1 SARAF DMRTA2 NTRK2 NIF3L1 TK1 KDM6B HMG1A FRZB PLIN3 FANCI CCDC12 CDK5R2 TEDC1 KIF1A ADD3 CHRNA4 CD200 SGCB MRPL48 E7F8 STMN4 NRXN3 LBH THOC6 PTBP2 RALGPS1 C1orf109 FAT1 BASP1 CENPU SLC4A10 DPYSL2 CHGA TMEM237 H2AC20 TCEAL3 ANK2 REST TIMM23 BCL11A GAD2 TMEM59 DCC SNRPA1 OPNH1 ATP2A2 PCNA ANXA5 ARHGAP19 GEM ARHGEF12 TMEM141 SLC2A12 MRPL13 MT2A RPL22L1 GEMIN2 EFTYK PNMA2 POLR1C LBR TFDP1 PEG3 C11orf95 RAD51AP1 ANKRD12 ATF1 OGT H2BC21 GJC1 EPB41L4A GNG3 ARHGAP21 CENPV RIMBP2 MGST1 FZD2 TEAD2 NUDCD2 SBK1 CA14 EFNB1 MBIP DUT TFDP2 CALM2 RYK PCLAF FANCD2 FEN1 SYT5 MAD2L1 EC1 BAZ2B GNASAS1 CACNG2 GPC3 CPNE7 CRABP2 GRIA2 CYLD DM2L3 CHD5 RFC3 CALM3 RRAGA PTTG1 HBA1 CHL1 MRTO4 TEX30 JAKMIP2 ATP6V1A MYCBP2 CCDC88A MAP6 ATP6V0B PWP2 PHYHIPL DPYSL3 HBG1 MFSND4AAS1 DNAJC12 DCTN3 RSP01 BRCA2 SLC25A36 REEP4 ASF1B ALP1 PROK2 ANKRD36B RARRS2 RRAGD PPFIA2 SHMT2 GRIA1 ATP6V1G1 B4GALNT1 RPA1 NECAP2 PPP2R2B SCN2A MAP4 JAZF1 SFRP2 REEP1 EIF4A2 SKA3 ATP6V1H SEPHS1 GPM6A CIRBP PDZRN4 GDI1 SPOUT1 RUSC1 EFNB3 CKLF HPCAL4 CENPS RUNCDC3B BCL2L11 SEZ6L2 LRP4A51 SIVA1 AK1 KLLN CEP135 TMEFF2 PPL11 SARS1 CORO7 GGCT GABRB2 LINC00504 SBN01 GIHCG CD6 GOLGA8B CTPS1 CCL3 COQ2 CASP6 H11 COMMD10 CENPM STAU2 PPM1K NAA38 RTN3 SUV39H2 CCDC34 SH3GLB1 H3C3 NREP MIS18A EXOSC3 NFIX CLDN10 ACTN1 CHD6 SLC1A5 CENPI MNS1 PLEKHA6 NAP1L5 KIF21A SEPTIN10 NOB1 HECTD4 RGS7 OLFM2 DKC1 FOXP2 KMT2E RPA3 MORF4L1 ANP32B FEZF1A51 MIS18BP1 SHISA9 H2AX SAE1 CCDC184 GCDH KIDINS220 SRRM4 HAUS8 CCDC14 SMPD3 CADPS TAGLN2 CEP85 C4orf48 SMC2 CKS1B RAB3C RFESD NR2F1A51 MTSS1 C3 SYT1 AAMD CDATAD5 RAD54B SOX11 MAP2 CCL5 NIP7 CDC59 DUSP26 HIRIP3 HNRNPDL SCN3B DNA2 SUPT5H EME1 STX12 TMEM98 PTPRO RPAIN KIAA1217 SAPW70 BTG3 ATCAY SECISBP2 DNMT2 KALRN BLCAP C21orf58 MRPL22 CXADR CSRN3 NUP37 MIAT MFGE8 DYNLRB1 SHISA7 HLAFA51 GALM NRP1 TMEM123 INSR MXRA7 NHLH1 SGIP1 AKR1C2 RNPC3 CCNG1 LYAR ACA2A RBL1 POU3F2 NAP1L3 CENPH MATR3 SMIM30 RPS4Y1 GP5M2 PGM2L1 IGFBP5 ST8SIA2 C5orf34 FOXK1 TMEM169 ATRX POLE2 SDF2L1 STMN1 PAICS NUDT1 ERCC6L DUS3L JPH4 MYT1L SPRY1 GINS1 RRM1 DNPH1 L1CAM L2HGHDH PGAM4 CBS CELF4 LINC01551 CDC14A MTHFD2 SIRT2 FAM155A EMX2 MGME1 EAVL3 BRIP1 SCRT2 CDR1 FXD7 ROBO2 PNPLA8 LHX1 RAB13 BNIP3L BRCA1 ROBO1 SERINC1 KIF21B CADM2 DCL1 MAP1LC3A KIF5C EVPLL PPP2R5C SLC7A14 MYH11 AKAP5 RBM4 NUDT6 TMEM126B ATN1 RTL8C RASGEF1B LMNB2 NANOS1 LOC440896 CORO1C PTN LY6E RAD54L MCM10 PNPLA4 PYGO2 PRSS23 MAGED2 BAX ALP2 CEP170 PGAP2 NPDC1 ALKBH1 HTR2C NCAN BRAT1 APIP CAMLG C21orf62 PLPPP1 DNAJC9 SUCLG2 ODF2 NSF2 RBFOX2 G3BP2 HSD11 HBB GABARAP2 GATA3 ENO2 CYRIA GABRA2 AKAP9 CDC42 PEG10 DAAM1 OSUN7 KLC1 MPPE1 CLASP2 EMC9 CFAP298 NCAM2 CCDC150 NUDT5 TFF2 MOB3B RALYA51 SPATA5 SYT4 CEP152 DBN1 SENP7 NUSAP1 RGM5 HMG5 CYR1 BBS5 ADCYAP1 PHF19 SDC1 FEZF1 GALNT14 HMG2A NEDD1 CENPW

Dataset	DEGs / cluster marker genes
Fan_Early (continued)	MRE11 PSD2 PSAP MAGEH1 ANKS1A DSTN H2AC4 DNAJC19 C19orf48 SLC25A4 EZR ACAD9 HELLS ORC6 CCDC60 NHLH2 HDGF GGH RGS17 LSM4 HS6ST2 KRT8 GNAO1 CNR1 BRD3 CNN2 DTL MRPL39 GNG8 RFC5 MRPL11 SLC35C2 IGFBP4 AFAP1 ALG1 C20orf27 OLFM1 EIF1B CITED2 FAM162A GAL SPIN1 NCAM1 AAK1 PRRX1 AP2M1 PGM5P3AS1 SLC17A6 PLD3 SYBU ORC5 TAF8 HMGA2AS1 RAE1 LRR1 TAOK3 JUP ARL1 RALYL DYNCLL12 ANKRD36 RLN2 SENP5 MRPL37 FBLN1 BUB3 CBX2 ATP6AP1 EFCAB11 SPC24 OGA PLIN2 TMEM60 ADAM22 NELL1 MRFAP1 TCEAL2 HES1 ILVBL LGALS8 TESMIN CD24 CDC45 CENPC SEPTIN3 GNG2 SMIM20 DPPA4 PAFAH2 DTD1 ING1 BLM ATXN7L3B PJA2 NTHL1 FEZ1 GPM6B AASDHPT SLIT1 OPTN NOVA1 EIF4G2 MTAP ATP6V0E2 LINC02158 HAT1 TPTEP2 PPIH GDAP1L1 TGDS DIRAS3 PLEKHG7 RAB31 SCG3 POU2F2 FBXL22 SPTAN1 STMN2 MEG3 RBPJ CTSC PAK5 ATP6V1G2 FUCA2 RRAGB DSN1 CCDC15 MTG1 DACT1 MELK PER1 GLRX ALKBH7 SLC45A1 DIAPH3 PSPC1 FEV CTNNA2 PSME2 LMNB1 GAD1 NEK6 H2BC20P CDK5R1 CDH9 AGTRAP S100A10 MYL12A ATP6V0C SCN3A H2AJ EBPL AGPAT2 BTBD8 LOC642846 PRELID2 POMT2 LRRTM4 TM6SF2 RASSF1 FOXG1 BORA DNER SMIM18 NRXN1 H2BC9 AUTS2 GEN1 NKX21 CACNA1E CISD3 BEX1 BRK1 EXOSC8 STK17B RBM39 SLITRK1 PDE4DIP RTN4 SCN9A PTBP1 ALG8 SFRP1 ANKRD20A9P NME4 HINT2 ARL6IP4 LY6H DSCC1 HIP1R RFC4 FAM72B LCOR NKX23 SRGAP1 LSM5 H4C2 PKMYT1 PRMT2 TGIF1 SPIN3 GHRH FOXO3 CELF1 DYNC1I2 MAP1LC3B ARL6IP5 EXO1 SLC16A1 DDB2 SOBP OTOF CENPK NEFL KLHL24 THSD7A CACNB3 ATAD1 SMIM4 AKAP6 SLC2A1 PCDH10 SMPD4 TMEM39A EPB41L1 SMC4 SERPINB6 CACNB1 SYT13 CHMP2A TNFAIP8L1 CHGB MND1 SEMA6D MPHOSPH8 LGR5 SRGAP3 NXPH4 PPBP NTM SOX3 RAB1A GNAS DDIAS SPOCK1 CSE1L CMC1 KIFAP3 CRMP1 CLTC CALCOCO1 DLK1 FAM241B C1orf112 PTX3 RAB2A METTL4 SKP2 CALD1 SHMT1 BEX2 NSG1 INA DMAC1 RPA2 MTHFD1 DCX BCL2L12 GYPC POLD3 SCLT1 SLC8A1 MACF1 ING4 MAB21L3 MSMO1 SASS6 AK2 IDS ARHGEF5 TMEM106C AMER2 CCNI STXBP1 SCG2 ACOT13 NEFM PLK2 TMPOAS1 GSTA4 ELOVL4 MSL3P1 FKBPL RBFOX1 HBA2 CLSPN IQGAP3 PBX1 RUFY2 SNAP25 FAIM2 RND3 ATAT1 LACTB2 PHF5A MICAL1 NUDT15 TENT5B SLC6A1 NFIB MYL12B LIX1 COQ3 SORBS3 CHAC2 MYBL2 MFAP2 SYCE2 RHEBL1 AIF1L NOL4L MAPRE3 CDCA4 TAC1 AP1S2 CAMK2N1 BRWD1 PUS1 ATP9A LRRN3 TDO2 COL11A1 NRP2 DST CHEK2 MRPL23 PHACTR3 ITGB3BP MCUB CKS2 SERINC3 NDRG4 DYNC1H1 COMMD4 ACTL6B ADGLR3 AKR7A2 POC1A H3C8 MAGOHB AP3B2 PSAT1 NES SEMA6A HMG2 HAUS4 FGD3 ST18 TMEM59L GFRA1 NR2F1 CENPN TAGLN3 NOL4 SH3GL2 H14 ATP6AP2 COX7A2L TMEM35A NSMCE4A TBCB NAA50 DARS2 TIMELESS RAB30DT KCNQ10T1 RAB14 KIFC1 GNG11 RRM2 AHYC ADGRG1 TDP1 RAB18 LINC00618 POLA2 LOC100129434 PARP15 RTN1 MIR1241HG RUFY3 RERE GAP43 IFITM3 CNPY3 PCBPA CCDC28B BOLA3 H4C11 H4C13 CPE SPA17 HMGCS1 LOC102606465 DPP6 FGF18 ATP6V1F ATP1A3 MARCKS HROB H2AC17 CEP43 DPYSL5 CSGALNACT1 PHGDH RMRP IFT122 ANK3 CLIC1 GPX8 CYB5D2 GMNN RNASEH2A ATP1B1 APC TET2 IGF2BP1 PCSK1N EBF1 GRIA4 COX7A2 CEP78 CRIP2 FNBP1L EPHA2 ADGRB3DT POLQ OCIAD2 MEAF6 ANKRD44 CCND1 PABIR2 LEF1 KLHL7 CELF2 KIF1B DHX32 DDX12P LINC00632 RUSC2 LOC100419583 PRNP NUP35 NDFIP1 NAPA PPP1R9B MMACHC CMC2 GPC2 MZT2A NOTCH2NLA PPAN FXYD6 RAB3A ABCC9 DDX39A BRD8 NR2F2 RBM14 CDCA5 NSG2 MCM2 C1orf198 FBXO43 DDR1 RER1 ATRAID CLN6 PRDX4 PYCR1 PLCL1 SIX5 NOTCH1 CSTF21 RHNO1 CHST1 MLLT11
Fan_Excior	ARHGEF39 CDC25C
Fan_Immune	ARHGEF39 SPC25 MALAT1 PRC1 KIF15 KIF20A CCNB2 CENPA DEPDC1 CDCA3 KNL1 MXD3 KIF23 CDK1 TMSB4X MKI67 HJURP ASPM NCAPG NCAPH FTL CCNB1 AURKB PLK1 BIRC5 FTH1 CCNA2 BUB1 ANLN CDCA8 KIF4A CDC20 NEIL3 MAP1B AURKA CENPO HBG2 PLK4 SPP1 CD37 ASAH1 ITM2B B2M SLC1A3 IFI16 FCGR1A FOLR2 LAPTM5 CTSB GPR34 CD53 CD68 IL6ST P2RY12

Table S3 Core set of ARHGEF39-related genes with DE logFoldChanges and assigned PsychEncode developmental WGCNA modules from Li et al. (2018)

Gene name	Loo_ RG2	Loo_ SVZ3	Loo_ GE14	Loo_ GE0	Poliou_ PgG2M	Fan_ NPC	Fan_ Early	Fan_ Excor	Fan_ Immune	WGCNA module
ARHGEF39	3.49	3.27	3.11	3.32	3.07	3.41	3.45	3.06	3.04	ME5
CDC25C	-0.11	1.44	1.32	1.08	0.49	0.60	1.24	0.80	1.05	ME5
ASPM	0.11	1.32	1.44	1.21	-0.01	0.47	1.02	0.36	1.29	ME5
NEK2	0.00	1.33	0.95	1.08	0.23	0.73	1.22	0.34	0.93	ME5
PLK1	0.14	1.41	1.39	1.36	0.33	0.92	1.41	0.39	1.41	ME5
NDE1	0.20	1.07	0.82	0.83	0.08	0.64	0.91	0.12	0.34	ME13
KIF11	0.08	0.71	1.22	1.02	0.18	0.43	0.96	0.20	0.74	ME5
CCNB1	0.24	1.67	1.67	1.39	0.11	0.50	0.78	0.17	0.94	ME5
FAM83D	-0.20	1.12	1.49	1.29	0.41	0.92	1.26	0.53	1.24	ME5
PRC1	0.12	1.61	1.65	1.33	0.06	0.31	0.88	0.21	1.34	ME5
NUF2	0.29	1.05	0.96	0.87	0.18	0.52	1.04	0.44	0.82	ME5
KNSTRN	0.07	1.07	1.34	0.72	0.22	0.56	0.88	0.34	0.85	ME5
SGOL2	0.18	1.48	1.51	1.31	0.04	0.62	1.15	0.49	1.11	ME5
KIF20A	0.35	1.09	1.15	0.81	0.39	0.81	1.32	0.82	1.31	ME5
BIRC5	0.04	0.96	1.25	0.85	0.14	0.37	0.79	0.30	0.96	ME5
KIF20B	0.07	1.15	1.20	0.79	0.11	0.46	0.88	0.39	0.24	ME5
CENPA	0.03	1.15	1.49	1.06	0.01	1.01	1.49	0.24	1.33	ME5
CENPE	0.10	1.40	1.30	1.15	0.06	0.58	1.09	0.55	0.99	ME5
CDK1	0.09	1.13	1.45	0.92	0.11	0.36	0.96	0.39	1.34	ME5
CDC20	-0.22	1.74	1.79	1.13	0.36	0.59	1.00	0.41	1.20	ME5
AURKA	0.24	1.60	1.38	0.95	0.10	0.89	1.06	0.28	1.22	ME5
FAM64A	-0.01	1.46	1.37	1.13	0.18	0.37	1.02	0.31	0.84	ME5
SGOL1	0.13	1.08	1.13	0.74	0.27	0.58	1.11	0.54	1.08	ME5
HMMR	0.35	1.73	1.34	1.00	0.12	0.80	1.21	0.75	1.11	ME5
HJURP	0.06	0.68	0.65	0.64	0.06	0.65	1.19	0.36	1.35	ME5
NCAPG	-0.03	0.74	0.71	0.90	0.10	0.40	1.03	0.12	1.36	ME5
CENPF	0.05	1.40	1.73	1.23	0.11	0.22	0.65	0.23	0.89	ME5
CDCA8	-0.01	0.85	1.16	0.80	0.17	0.71	1.21	0.57	1.43	ME5
RACGAP1	-0.04	0.84	1.22	0.67	0.12	0.53	1.08	0.07	0.89	ME5
KIF23	0.26	1.33	1.39	0.98	0.21	0.76	1.25	0.30	1.51	ME5
SPC25	0.15	1.03	1.17	1.07	0.06	0.33	0.81	0.25	1.12	ME5
CKAP2L	0.10	1.14	1.28	0.86	0.08	0.59	1.27	0.39	1.21	ME5

Gene name	Loo_ RG2	Loo_ SVZ3	Loo_ GE14	Loo_ GE0	Poliou_ PgG2M	Fan_ NPC	Fan_ Early	Fan_ Excor	Fan_ Immune	WGCNA module
ARHGAP11A	0.25	1.28	1.52	0.95	0.35	0.44	0.99	0.22	0.96	ME5
TACC3	0.10	1.14	1.23	0.92	0.12	0.59	1.12	0.61	1.02	ME5
MKI67	0.04	0.85	0.86	0.80	0.22	0.35	0.83	0.20	1.21	ME5
ECT2	-0.02	1.30	0.97	1.21	0.14	0.67	1.21	0.21	0.93	ME5
PSRC1	0.35	1.40	1.60	1.10	0.25	0.57	1.06	0.53	0.45	ME62
CEP55	0.36	1.44	1.28	0.98	0.39	0.63	1.12	0.46	1.14	ME5
CASC5	-0.06	0.76	1.10	0.76	0.06	0.58	1.37	0.46	1.71	ME5
KIF2C	0.23	1.47	1.26	0.97	0.42	0.61	1.15	0.43	0.91	ME5
CCNB2	0.00	1.49	1.25	1.04	0.28	0.50	0.90	0.28	1.09	ME5
CCNA2	0.12	0.96	0.92	1.02	0.27	0.63	1.06	0.40	1.29	ME5
CKAP2	0.05	1.36	1.35	0.80	0.07	0.34	0.69	0.17	0.69	ME5
CDCA3	0.13	1.04	1.23	0.95	0.11	0.73	1.20	0.35	1.35	ME5
BUB1	0.30	0.77	0.97	0.93	0.19	0.65	1.31	0.41	1.23	ME5
MXD3	0.15	0.81	1.01	0.75	0.12	0.73	0.94	0.40	1.05	ME5

Information about WGCNA modules

WGCNA module	Spatiotemporal interpretation	Enriched cell types	#ARHGEF39 genes
ME5	High prenatally in all regions	NEP/RGC/OPC	44
ME13	High prenatally in all regions	NEP/RGC/OPC	1
ME62	Low in postnatal CBC	NEP/RGC	1

CBC = cerebellar cortex, NEP/RGC = neural epithelial progenitor/radial glial lineage, OPC = oligodendrocyte progenitor cells

6





**General
discussion**

CHAPTER 6

General discussion

Neurodevelopmental disorders that affect speech and language development have complex molecular aetiologies. At the genetic level, these disorders are likely to involve many risk loci. At the phenotypic level, there can be remarkable clinical variation between patients carrying pathogenic mutations in the same gene. This variation can be the result of genetic background effects, varying impacts of different mutations on gene function, or individual variation in the network of molecular interactions that is needed for a gene function. The aim of this thesis was to understand how language-related genes and gene networks influence human neuronal features that contribute to the development of language-related traits.

In this thesis, I have focused on two language-related genes: *CNTNAP2* and *ARHGEF39*. *CASPR2* deficiency disorder (CDD) is a severe condition, in which homozygous mutations in *CNTNAP2* give a wide range of varying symptoms. This variability indicates that *CASPR2* (the *CNTNAP2* protein product) has multiple functions and, together with molecular studies, this has suggested that this gene acts as part of an extended molecular network. Illustrating the phenotypic and genetic complexity of CDD, I presented the case of a patient with compound heterozygous mutations in *CNTNAP2* and hyperkinetic stereotyped movements - a novel neurological manifestation of CDD (Chapter 2). I used a human neuronal network model to investigate how mutations in *CASPR2* isoforms may differentially affect the development of disorder-related neuronal phenotypes (Chapter 3). I extended the molecular network of *CNTNAP2* by investigating the expression patterns of microRNAs (miRNAs), regulatory molecules that play essential roles in neurodevelopment, in response to loss of *CNTNAP2* in human neuronal networks (Chapter 4). *ARHGEF39* was more recently linked to language disorder phenotypes via a functional polymorphism in a microRNA binding site – pointing to a potential impact of miRNA networks on language development. I characterized the molecular function and explored the neurodevelopmental co-expression network of *ARHGEF39* using cellular assays and publicly available single cell RNA-sequencing datasets (Chapter 5).

In this final chapter, I will summarize the findings of each experimental chapter (section 6.1). From these results, I will discuss how these findings advanced our understanding of the role of *CNTNAP2* in disorder (section 6.2) and the molecular contributions to understanding language-related disorders more broadly (section 6.3). Furthermore, I will discuss future perspectives for gaining deeper insight in the relationships between gene networks, human neuronal features, and language-related computation in neurons by combining human cell models with other experimental models (section 6.4).

6.1 Summary of the findings presented in this thesis

6.1.1 Expanding the molecular and phenotypic spectrum of CDD (Chapter 2)

CDD is a rare syndromic neurodevelopmental disorder that is characterized by refractory epilepsy, intellectual disability, language impairment, and autistic features as core symptoms. However, there is some variability in these and other symptoms among CDD

cases (Rodenas-Cuadrado et al. 2016; Smogavec et al. 2016). In chapter 2, I presented the case of a boy with the common features of CDD and hyperkinetic stereotyped movements who carried a paternally inherited missense mutation and a maternally inherited duplication involving intron 1, exon 2 and intron 2 of *CNTNAP2*. The affected region in intron 1 contains a binding site for the transcription factor FOXP2, potentially leading to abnormal *CNTNAP2* expression. Hyperkinetic stereotyped movements are a novel addition to the phenotypic spectrum of CDD. It remains to be tested how these mutations affect neurodevelopment, and if this specific combination of coding and non-coding mutations contributed together to this novel motor phenotype. The boy's parents were unaffected, which supports the incomplete penetrance of heterozygous *CNTNAP2* mutations. It is, thus, recommended to include *CNTNAP2* regulatory regions in the genetic screening of patients with a CDD-suggestive phenotype, especially if only a single heterozygous *CNTNAP2* variant has been identified or if atypical neurological phenotypes are also present.

6.1.2 Defining isoform-specific functions of *CNTNAP2* in human neurons (Chapter 3)

Given the clinical heterogeneity of patients with *CNTNAP2* mutations, it is important to functionally characterize different types of mutations in *CNTNAP2* to understand potential mechanisms underlying this heterogeneity. I hypothesized that functional differences between isoforms may partially explain variation in *CNTNAP2* disorder phenotypes. To explore this, in chapter 3 I characterized three *CNTNAP2* mutant conditions in human neuronal network cultures: (1) heterozygous and (2) homozygous loss of the long isoform, and (3) homozygous loss of both the long and short isoforms. This is the first study that looks at multiple types of *CNTNAP2* mutations in an isogenic human background, allowing for a precise comparison of the effects of these mutations in a species-specific molecular context. This study shows that homozygous loss of *CASPR2* in human neuronal network cultures affects the differentiation of neural precursor cells, the number of excitatory synapses and gene expression for several biological processes, whereas heterozygous loss only produces a subset of the gene expression changes. The limited phenotypes observed in the heterozygous *CASPR2* condition are consistent with the lower risk of disorder in individuals carrying heterozygous *CNTNAP2* mutations. In contrast, mutations affecting the short isoform seem to be more disruptive than those affecting the rest of the gene. The cause of this is likely two-fold since these mutations are likely to affect both the long and short isoforms, but also because the short isoform is expressed much earlier in development. I showed that when both long and short isoform expression is lost, cilia assembly genes are specifically disrupted. This effect of *CNTNAP2* loss has not previously been described, possibly because mutant mouse models of *CNTNAP2* disruption targeted exons in the long isoform, leaving short isoform expression intact. I showed that short isoform expression peaks much earlier in development than the long isoform, and thus by disrupting both isoforms I was able to identify a potential relationship between cilia and *CNTNAP2* function that may be crucial to early brain development. Further investigation of a role for *CASPR2* in cilia during early brain development could help to understand patient phenotypes relating to neuronal migration and cortical lamination, and provide an explanation for clinical variation between cases with mutations that affect different isoforms.

6.1.3 MiRNA networks respond to loss of *CNTNAP2* in human neurons (Chapter 4)

MiRNAs are key regulators of several neurodevelopmental processes. In chapter 4, I established that the functions and development of miRNA networks in cortical development can be studied using human neuronal network cultures. In chapter 3, I showed that the human neuron model most closely resembled *in vivo* cortical development profiles based on gene expression patterns. In chapter 4, I showed that these cultures also have a distinctively cortical miRNA expression pattern when compared to samples of post-mortem human brain tissue, suggesting that they provide a good model for studying non-coding regulatory networks to understand their roles in neurodevelopment. I then used this model to demonstrate the effects of *CNTNAP2* loss on miRNA expression. The major perturbation to miRNA expression observed could partially be explained by changes in cell type composition. Still, by integrating motif-based analysis of miRNA targets with miRNA expression, I identified six miRNAs that seem to counteract previously established deficits in neurite outgrowth of *CNTNAP2* knockout neurons. Genes that inhibit neurite outgrowth were under stronger repression by some of these active miRNAs, whereas genes that promote neurite outgrowth were targeted by downregulated miRNAs. These miRNAs and their targets can be considered as extensions of the molecular network related to *CNTNAP2* function during cortical development and could point to places where compound mutations could have consequences for differences in clinical phenotypes.

6.1.4 Investigation of molecular and neurodevelopmental functions of *ARHGEF39* (Chapter 5)

Just as understanding gene function can provide windows into disorders and phenotypes, understanding miRNA networks and their regulatory targets can give insight into the complex molecular networks underlying neurodevelopmental disorders ¹. In chapter 5, I explored the molecular and neurodevelopmental functions of *ARHGEF39*. Misregulation of this gene was previously implicated in developmental language disorder (DLD) *via* a functional polymorphism that can disrupt post-transcriptional regulation by miRNAs ². Little was known about properties of *ARHGEF39*, or how its (dys)function might contribute to neurodevelopment or related disorders. *ARHGEF39* was known to be a RhoGEF family member, and as such it would be capable of activating one or more RhoGTPases. RhoGTPases are involved in every cellular process that requires cytoskeletal reorganization and thus are fundamental enzymes for cellular function. I showed that *ARHGEF39* specifically activates one RhoGTPase -RHOA – rather than RAC1 or CDC42. RHOA is involved in the assembly of cell-matrix interactions *via* focal adhesions ^{3,4}. I further showed that *ARHGEF39* expression in cell cultures leads to cell detachment. Together these findings suggest that *ARHGEF39* may regulate cell adhesion. Given this, and the prior work implicating *ARHGEF39* in language-related phenotypes, I went on to explore the role of *ARHGEF39* in neurodevelopment. I used published single cell RNA-seq datasets to determine the expression of *ARHGEF39* in the developing brain and uncover neurodevelopmental processes implicated in its function. I demonstrated that *ARHGEF39* is a marker gene for proliferating neural progenitor cells, and that it is co-expressed with genes involved in cell division. During cell division, precise regulation of RHOA activity is necessary for cell rounding and the controlled disassembly of focal adhesions. This led me to the novel suggestion that *ARHGEF39* could play a role in neurogenesis in the

developing brain. Understanding the GTPase substrate, co-expression network, and processes downstream of ARHGEF39 may shed light on the mechanisms by which altered expression levels of ARHGEF39 contribute to neurodevelopment and associated disorders.

6.2 Understanding the role of *CNTNAP2* in disorder

A complete understanding of the genetics of language-related disorders and their relationship to complex phenotypes must be addressed with a coherent and interdisciplinary strategy – a simplified version of which is found in Figure 1. These approaches can be thought of as cyclical since they are all interconnected, and research can move in both directions between most of the approaches. An entry point in this research cycle can be to identify new candidate genes by studying families and cohorts with language disorders. This often leads to exploration of the molecular networks of the identified candidate genes, sometimes resulting in the identification of more candidate genes. The functions of these genes can then be studied further in various experimental models to identify the molecular, neuronal (and sometimes behavioural) phenotypes affected by these genes, and providing clues about the mechanisms that lead from mutation to disorder. Together these findings shed light on the biological encoding of language and provide insight into the genetic aetiology of language-related disorders. In this thesis, I made contributions to understanding the role of *CNTNAP2* in disorder at multiple steps in this research cycle.

6.2.1 Clinical and mutational spectrum of CDD

CNTNAP2 was first identified as language-related gene due to its membership of the molecular network of FOXP2, together with genetic association to non-word repetition (NWR) performance in children with DLD⁵. However, mutations in *CNTNAP2* are not limited to language impairment, but are associated with a broad spectrum of other clinical

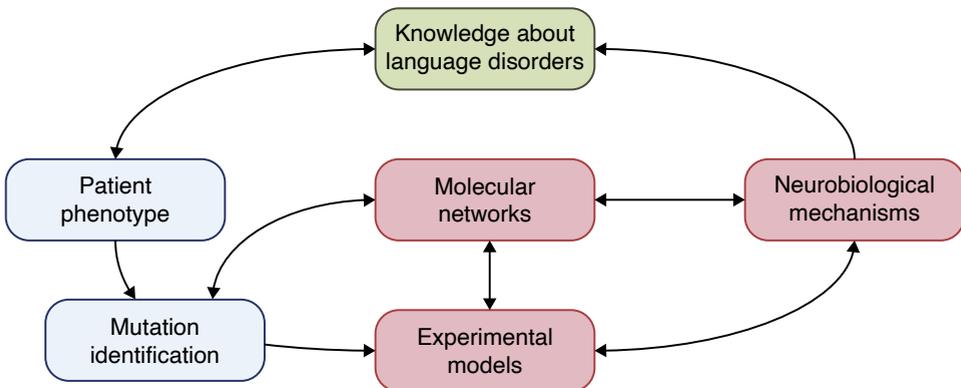


Figure 1 Research cycle for studying genetics of language-related disorders. In blue is the identification of candidate language-related genes by studying patients with language-related disorders and exploring the molecular networks of other candidate language-related genes. In red is the investigation of the functions of candidate language-related genes in experimental models by characterizing the molecular networks and affected neurobiological mechanisms. In green is the integration of neurobiological mechanisms and patient phenotypes to advance the knowledge about the genetic aetiology of language disorders.

phenotypes including intellectual disability, autistic characteristics, and seizures⁶⁻⁸. The broad spectrum and variability of *CNTNAP2*-related phenotypes suggests that *CNTNAP2* has multiple functions during neurodevelopment and several genetic interactions. In chapter 2, I expanded the mutational and clinical spectrum associated with bi-allelic mutations in *CNTNAP2* with the case of a patient that had inherited mutations from two unaffected parents, including a maternal duplication affecting the FOXP2 target site. This patient showed hyperkinetic stereotyped movements, representing a novel clinical feature. It remains to be tested if this specific combination of coding and non-coding mutations contributed together to this novel motor phenotype. In any case, this report emphasizes the importance of extending patient screening to non-coding regions of *CNTNAP2*, especially if only a single heterozygous *CNTNAP2* variant has been identified or if atypical neurological phenotypes are present. MiRNAs and their target sites present an additional way of identifying compound or non-coding mutations that can affect *CNTNAP2* function. In chapter 4, I showed how miRNA function seems to counterbalance reductions in neurite outgrowth upon loss of *CNTNAP2*. Mutations in these miRNAs or their targets may therefore exacerbate *CNTNAP2* phenotypes. The functional impact of *CNTNAP2* mutations may further depend on which protein-coding isoforms are affected. In chapter 3, I showed that short isoform mutations lead to an earlier and more severe disruption of neurodevelopment than long isoform variants, which may explain some of the variability in cortical lamination phenotypes observed in patients with bi-allelic mutations in *CNTNAP2*^{7,8}. Transcription factor sites, miRNA networks, and different isoform thus each present additional dimensions to the mutational complexity of *CNTNAP2* and may be used to better understand the various roles of this gene in neurodevelopment and related disorders.

6.2.2 Human neuronal networks as an experimental model to study *CNTNAP2* function

In this thesis, I applied human neuronal network cultures as an experimental model to study the functions of *CNTNAP2*. *In vitro* neural differentiation from stem cells offers a tractable model that reproduces (species-specific) features of *in vivo* brain development⁹. I specifically adopted a protocol that produces electrophysiologically mature neuronal networks¹⁰, as *CNTNAP2* has been reported to be important for synaptic function and functional network connectivity¹¹⁻¹³. Since the transcriptomic profile of a cell or tissue reflects its identity, this readout was used to establish the molecular identity of the human neuronal model system. Cross-correlation of samples from the human neuronal network with *post mortem* human brain samples showed the highest similarity in mRNA expression with mid-gestation samples (Chapter 3) and the highest similarity in miRNA expression with cortex samples (Chapter 4). In addition, developmental processes important for network development, such as astrocyte differentiation and neuronal maturation, could be traced via cell type deconvolution and expression patterns of miRNAs in time series data from differentiating cultures. CRISPR/Cas9 allowed the specific introduction of mutations in *CNTNAP2* in human embryonic stem cells, creating isogenic lines to study isoform-specific functions of *CNTNAP2*. Whole genome sequencing found an even distribution of background variation and lack of clear pathogenic variants in the generated lines, confirming that these lines can be used to specifically model the effects of loss of the long isoform and both long and short isoforms of *CNTNAP2* on cortical development.

The work in this thesis also addressed a technical challenge with this model and identified properties that need to be considered for future experiments. First, differences in growth rate between NPC lines can create variability in network density. I showed that adjusting the seeding density for neural network differentiation can compensate for this issue. However, this led to technical variability between experiments that probably affected the astrocyte maturation and baseline numbers of synapses, making it difficult to quantitatively compare conditions across experiments. Future experiments should take care to measure the growth rate of each NPC line after fluorescence activated cell sorting and before seeding. Second, the human neuronal networks produced mostly excitatory neurons, facilitating studies on effects on excitatory synapses and observations of changes in neural activity across conditions. However, it limited the possibility to study inhibitory synapses and the complexity of changes in neuronal network activity that could be observed had both inhibitory and excitatory synapses been present. In future, effects of *CNTNAP2* on inhibitory neurons could be addressed with protocols that co-culture or specifically generate GABAergic neurons^{14,15} or three dimensional brain organoid or assembloid cultures that display larger neuronal cell type diversity^{16,17}. The description herein regarding NPC growth and the high ratio of excitatory neurons will allow precise application of these human neuronal network cultures in the future.

6.2.3 Neurobiological mechanisms that are affected by the *CNTNAP2* molecular network

Mature neuronal networks were compared for differences in cell type composition, synapse counts, neuronal activity, and gene expression. These comparisons confirmed known *CNTNAP2*-related phenotypes, such as a decrease in the number of mature synapses in *Cntnap2* knockout neurons^{12,13,18-21} and dysregulation of (genes related to) neurite outgrowth^{12,18,22-24}. This thesis further contributed two neurobiological mechanisms that may contribute to *CNTNAP2*-related disorders. Network cultures with dual isoform *CNTNAP2* knockout were most severely affected, including a specific disruption of cilia assembly pathways. Ciliogenesis may therefore be a specific function in which the short isoform is involved. This hypothesis is supported by the observed downregulation of a CASPR2 interactor, *TCTN1*, and by the early expression of cilia assembly genes that coincide with the early expression of the short isoform. Ciliogenesis is important for correct neuronal migration²⁵⁻²⁷ and my findings suggest that this pathway could play a role in *CNTNAP2*-related disorders, such as the cortical dysplasia observed in the Old Order Amish children that, unlike most described patients, have a mutation in the C-terminal region of the protein that affects both isoforms²⁸.

In addition, analysis of miRNA expression in chapter 4 revealed that the activity of miRNAs that promote neurite outgrowth is upregulated in *CNTNAP2* knockout cultures. The findings suggest a counterbalancing mechanism for the decrease in neurite outgrowth that results from reduced levels of *CNTNAP2* and may indicate that miRNAs function to reduce the effects of this neurodevelopmental imbalance. In the future, the influence of these neurobiological mechanisms can be studied further in these human neuronal network cultures and other models. Additional mechanisms involving *CNTNAP2* will likely be uncovered in this process.

6.3 Unfolding complexities of language-related disorder

6.3.1 Genetic and phenotypic complexity

The genetics of language and language-related disorders is often described as complex; however, this complexity takes many forms. There is complexity at the genetic level, since multiple genetic factors contribute to the risk for language disorders and this risk is mediated by numerous genetic interactions²⁹. Therefore, it is important to investigate the molecular networks that are formed by these interactions to identify the processes that contribute to the risk for language-related disorders. At the phenotypic level, complexity is represented by the different neurobiological mechanisms that can contribute to similar clinical outcomes. Understanding these mechanisms may clarify the distinctions and overlaps between separate cases of genetic language-related disorder. In this thesis, I have made efforts to address these forms of complexity.

6.3.2 Addressing genetic complexity by characterizing non-coding sequences

The non-coding regions of the genome are sites of regulatory interactions that can provide new insights into the complex molecular architecture of language-related disorders. In this thesis, I highlighted the functional importance of non-coding sequences containing binding sites for transcription factors and miRNAs. In chapter 2, I highlighted that a patient mutation in intron 1 of *CNTNAP2* duplicated a binding site for *FOXP2* and may contribute to atypical neurological manifestations of CDD. In chapter 4, I described the roles of miRNAs in the development of neuronal networks and how these pathways may modulate the effects of homozygous mutations in *CNTNAP2* via interactions with non-coding UTRs. These conclusions provide reasons to further investigate the non-coding sequences around known candidate language-related genes. A previous study showed that polymorphisms in the 5' promoter of *CNTNAP2* were associated to autism spectrum disorder (ASD) and language development and that these variants altered transcription factor binding and transcriptional regulation of *CNTNAP2*³⁰. Such studies could be extended to other regions, such as the introns and 3'UTR of *CNTNAP2* to identify new patient variation and gain a better understanding of the molecular network around this gene. The importance of non-coding variants in disorder is also supported by the identification of *ARHGEF39*, which I studied in chapter 5, providing new insight into molecular pathways, cellular pathways and cell types that are relevant for language-related neurodevelopment. Investigating the functional impacts of non-coding mutations helps to understand the function of language-related genes and may identify new risk factors.

6.3.3 Non-word repetition as a gateway into phenotypic complexity

Polymorphisms in *CNTNAP2* and *ARHGEF39* have both been associated to DLD^{2,5}. DLD is a complex disorder, as it is multifactorial and encompasses a wide range of language problems³¹. Functional links between DLD-related genes may help to identify molecular pathways and neurobiological mechanisms that are relevant for the aetiology of this disorder. For example, the molecular link between *FOXP2* and *CNTNAP2* provided a potential mechanism for how mutations in *FOXP2* may affect neuronal networks that are important for language development via the regulation of *CNTNAP2* expression⁵. The

question of a functional link between *CNTNAP2* and *ARHGEF39* is of special interest, as polymorphisms in both genes were associated to performance on the NWR task ^{2,5}. NWR involves phonological working memory and is linked to several language skills such as vocabulary acquisition, reading, and language comprehension ^{32,33}. A recent genome-wide analysis confirmed the heritability of NWR task performance and its shared genetic architecture with other reading and language measures that is largely independent from performance IQ ³⁴. This task is therefore likely to be able to reveal genetic associations that are relevant for speech and language development.

However, the genetic architecture of NWR is likely as complex as its phenotypic architecture, because it likely captures variation in several different neurodevelopmental mechanisms. This is illustrated by the lack of a clear functional overlap between *ARHGEF39* and *CNTNAP2* when we look at these two genes implicated in the NWR phenotype. Thus far, the only neurodevelopmental phenotype that *ARHGEF39* has been implicated in is proliferation of NPCs ³⁵. By contrast, *CNTNAP2* function has been widely implicated in neuronal functions including synapse development, dendritic arborization, axonal excitability and network synchrony ^{11,12,19-21,23,36,37}. At a molecular level, the proteins encoded by these genes are not known to share interaction partners, nor do they seem to participate in the same pathways. However, in this thesis, I found that both genes may affect the numbers of NPCs during cortical development. In chapter 3, I described that human neuronal networks with homozygous mutations in *CNTNAP2* had a reduced number of NPCs, which coincides with an increase in neuronal differentiation. In chapter 5, I showed *ARHGEF39* is a marker for proliferating NPCs and *ARHGEF39* expression correlates with genes that regulate the cell cycle. Recent studies show that the diversity of NPC types is larger than previously thought, especially in the human cortex ³⁸. It remains to be investigated whether *ARHGEF39* and *CNTNAP2* affect the same or different types of progenitor cells during human brain development. However, this thesis provides intriguing clues that NPCs may provide a common developmental point at which *ARHGEF39* and *CNTNAP2* ultimately contribute to variation in NWR task performance.

6.4 Future perspectives

The genetics of language aims to identify molecular underpinnings of human speech and language abilities. Part of this enterprise is to understand how candidate language-related genes and gene networks influence neuronal features that contribute to the development of language-related traits. In this thesis, I demonstrated the value of human stem cell-derived neuronal networks as an experimental model to study gene function. I identified molecular networks around language-related genes via transcriptomic data. I generated new hypotheses about the neurobiological mechanisms that these genes are involved in by combining network data and data from cellular assays. The next steps to further our understanding of the influence of candidate language-related genes on neurodevelopment and cognition will require the combination of human neuronal models with other types of experimental models. Human cell models and animal models can complement each other in multiple ways for finding the neural features that contribute to language. Additionally, improved human cell models and the insights from computational models will make it possible to investigate the effects of genetic mutations on advanced neurobiological functions, such as information processing *in vitro*.

6.4.1 Human cell models and animal models complement each other as approaches to uncover neuronal features that contribute to language

Human neuronal models can capture human-specific aspects of neurodevelopment, genetics, and disease mechanisms³⁹. At the same time, cell models have their limits in terms of functional complexity, as they lack the architecture and circuitry of a full brain and are incapable of sentient behaviour. Animal models will, therefore remain essential to investigate gene-function relationship for complex behavioural phenotypes. Findings from human models can lead to functional hypotheses about the neurodevelopmental perturbations resulting from mutation of a gene of interest. For example, functional studies in human neurons from this thesis have led to the hypothesis that the short isoform of *CNTNAP2* may play a role in cilia assembly to affect neuronal migration and cortical lamination (Chapter 3). Animal models can be useful for *in vivo* validation of such a hypothesis and may be able to extend the understanding of neurobiological mechanisms by adding information about effects on the development of circuitry and behaviour. However, for these efforts to be successful, it will be necessary to account for potential species differences that affect translation between model systems. New insights into developmental trajectories can help to facilitate translation between *in vivo* humans, *in vivo* animal models, and *in vitro* human cell models. A comparative approach involving multiple animal species can further help to answer more ultimate questions about the functions of identified neurobiological mechanisms.

It is not easy to identify suitable animal models to study the functions of language-related genes given that language is a human specific trait. Human speech and language require multiple processes, and any animal model only allows investigation of a subset of similar features⁴⁰. The ability to reproduce features of a trait in a model is called face validity. In addition, human and non-human species respond differently to gene mutations. For example, humans, mice, and rats with mutations in *Cntnap2* have displayed dramatic variations in behaviour and seizure phenotypes⁴¹. This indicates that there are important factors in the model background that affect the outcome and suggests differences in construct validity between model systems. These problems require a detailed approach. It is too crude to label models as valid and invalid, but rather it is important to be as specific as possible about the process that is being modelled, and the known and unknown characteristics of the model⁴². Recently, high-resolution analysis of brain tissue with single cell and spatial omics has enhanced understanding of brain development in such a way that developmental trajectories can be compared between species and between *in vivo* and *in vitro* models and that risk factors can be mapped to time and location⁴³. This makes it easier to select the cell types and developmental stages that are relevant to model for a certain risk factor and to match *in vitro* and *in vivo* models based on these features. The comparison of human cell models and animal models is important to translate the findings from these models to human language. Convergent findings validate the involvement of a language-related gene in a specific neurobiological mechanism, whereas divergent findings can point to relevant species-specific differences that may underlie human-specific aspects of language.

Behavioural and neurobiological differences between animal species should thus not be seen only as a limitation, but also as an opportunity. Embracing the diversity of the animal kingdom offers the opportunity to investigate the roles of genes in the evolution

of neuronal features that affect language-relevant traits. Vocal production learning is an essential process in the acquisition of speech and this trait is shared by humans and some other animals⁴⁴. Convergent evolution between vocal learning species can point to neurobiological mechanisms that are essential for this trait. For example, human and songbird brains show convergence in the expression of genes related to circuit development⁴⁵, which could regulate the formation of a direct projection from the forebrain to the brain stem that is found in vocal learners⁴⁶. Songbirds have been used most extensively as model system for vocal learning, but bats receive increasing attention⁴⁷. Bats offer a widely diverse order of animals including a broad spectrum of vocal learning abilities⁴⁸. Further, the mammalian brain structure of bats is an advantage for comparisons with the human brain, and progress is being made in the neurobiological and molecular characterization of bat brains⁴⁹. It will be valuable to apply the insights from human cell models in a comparative approach. For example, a future question may include whether functional variation or expression of *CNTNAP2* or *ARHGEF39* relates to aspects of neocortical development and abilities for vocal learning. The answers to such questions would help to address not only how, but also why these genes are important for complex vocal learning behaviours.

6.4.2 Investigating computation to bridge the gap between human neuronal models and language

The capacity for human language can be decomposed into basic building blocks and core operations that require the involvement of multiple brain networks^{50,51}. The computational contribution of these networks and involved areas to cognition is a big question, as it requires a detailed understanding of both the neurobiological infrastructure and computational requirements of language components. Traditional cognitive neuroscience has mostly focused on the computational power of macroscopic neural circuitry, but increasing emphasis has been put on computations at the level of single neurons for learning and plasticity⁵². Single neurons can process information by transforming an input spike train into an appropriate output spike train using sophisticated computations that are supported by their individual cellular properties, such as dendritic tree shape, membrane ion channel expression, and synaptic plasticity^{53,54}. On top of that, the properties of individual neurons contribute to emergent computations in ensembles of neurons⁵⁵. The biological features of single neurons can thus contribute to abilities that are necessary for language. For example, computational modelling shows that neuronal spike rate adaptation via membrane excitability can support the rapid storage of information that is needed for sentence processing in a simple network⁵⁶. Neuronal features that are relevant for neuron computation are affected by language-related genes and gene networks. For example, chapter 3 showed that loss of *CNTNAP2* affected properties such as spike rate and the number of synapses. However, it remains an open question how mutations in a gene, such as *CNTNAP2*, directly affects the computational properties of a network of neurons. This could be addressed *in silico* by adjusting model parameters based on the measured effects in a gene knockout model. Recent advances in human cell models and electrophysiological assays also may allow investigation of this question *in vitro* by performing computational experiments in physical biological neuronal networks.

Human brain organoids offer a window into the complex neuronal activity in self-organised networks that mirror aspects of the developing human brain, such as tissue architecture, cell diversity, and neuronal maturation⁵⁷. The architecture of three-dimensional neuronal networks allows for more natural intercellular interactions and resulting network activity than two-dimensional cultures, but the activity in three-dimensional cultures is more difficult to measure than in two-dimensional cultures⁵⁸. Recent technical improvements for recording activity in three-dimensional cultures have enabled researchers to show that mature organoids are capable of sophisticated neuronal network dynamics, including rhythmic oscillatory activity⁵⁹⁻⁶¹. These oscillations depend on the tuned interaction of excitatory and inhibitory neurons⁶⁰, and resemble features of electroencephalography in human neonates of 25 to 38 post-conceptual weeks⁵⁹. Oscillations could be observed at multiple frequency bands and detailed analysis validated the presence of theta oscillations in organoids of 4-6 months⁶¹. Cortical oscillations have been proposed to be foundational to speech processing⁶². In this framework, theta oscillations track slow temporal information in speech, which may be important for speech intelligibility. Human brain organoids could thus provide an experimental system to study the contribution of language-related genes to the development of complex neuronal activity, while taking into account the contributions of human-specific cell types, genes, and molecular networks. Organoids with mutations in *MECP2* display the epileptiform-like activity that mimics seizures that are commonly seen in patients with Rett syndrome⁶⁰. This provides a system to study the precise cell-type-specific perturbations that underlie the pathological changes in neural network function. The aetiology of epilepsy in CDD patients could be studied in a similar approach by generating brain organoids with *CNTNAP2* mutations. Patient-derived organoids could help to explain interindividual variability in epilepsy symptoms in CCD patients. Pharmacological tests in patient-derived organoids may also help to tailor therapy better, as patients with bi-allelic mutations in *CNTNAP2* respond inconsistently to anti-epileptic drugs⁷.

Thus far, complex human neuronal models focus on intrinsic patterns of neuronal activity that are present in the absence of a task. If it is possible to control input and output of an *in vitro* neuronal network, new assays can be developed to test the computational performance of neuronal networks. Recent work in monolayer cultures of human and mouse neurons showed that these cultures can learn how to play the classical game Pong in a closed-loop system⁶³. The neurons received sensory input about ball position and the recorded network activity produced a motor output that let the paddle move. If the ball was hit, this would result in a predictable stimulus as the position of the ball followed the previous trajectory. If the ball was missed, the game would reset and the ball would be in an unpredictable position. The cultures displayed the ability to learn how to move the paddle based on sensory input. It may be possible to adapt such a closed-loop system for language-related tasks. For example, a computational model that tests sentence processing and labels the semantic roles of words⁵⁶ may then be implemented in a biological neuronal network model. The combination of these computational tasks with increasingly complex and sophisticated human neuronal models will eventually make it possible to dissect the neuronal features and relevant underlying molecular networks that correlate with language.

6.4.3 Conclusion

In closing, there are several steps in explaining how genes and the neuronal features they influence contribute to the development of language-related traits. Candidate language-related genes need to be identified and they need to be situated in dynamic molecular networks that change depending on cellular and developmental contexts. Experiments are necessary to characterize these molecular networks and test the contribution of individual genes to the neuronal features that are supported by these molecular networks. In this thesis, I contributed new insights into the molecular networks around *CNTNAP2* and *ARHGEF39* by using human cell models and molecular data, resulting in hypotheses about relevant neurobiological mechanisms for (variation in) pathology. These results enable follow-up studies that, with the addition of other types of experimental models, can investigate the mechanisms by which *CNTNAP2* and *ARHGEF39* affect brain structure, complex network dynamics, behaviour and computation. This promises valuable steps to further our understanding of the influence of candidate language-related genes on neurodevelopment and cognition.

References

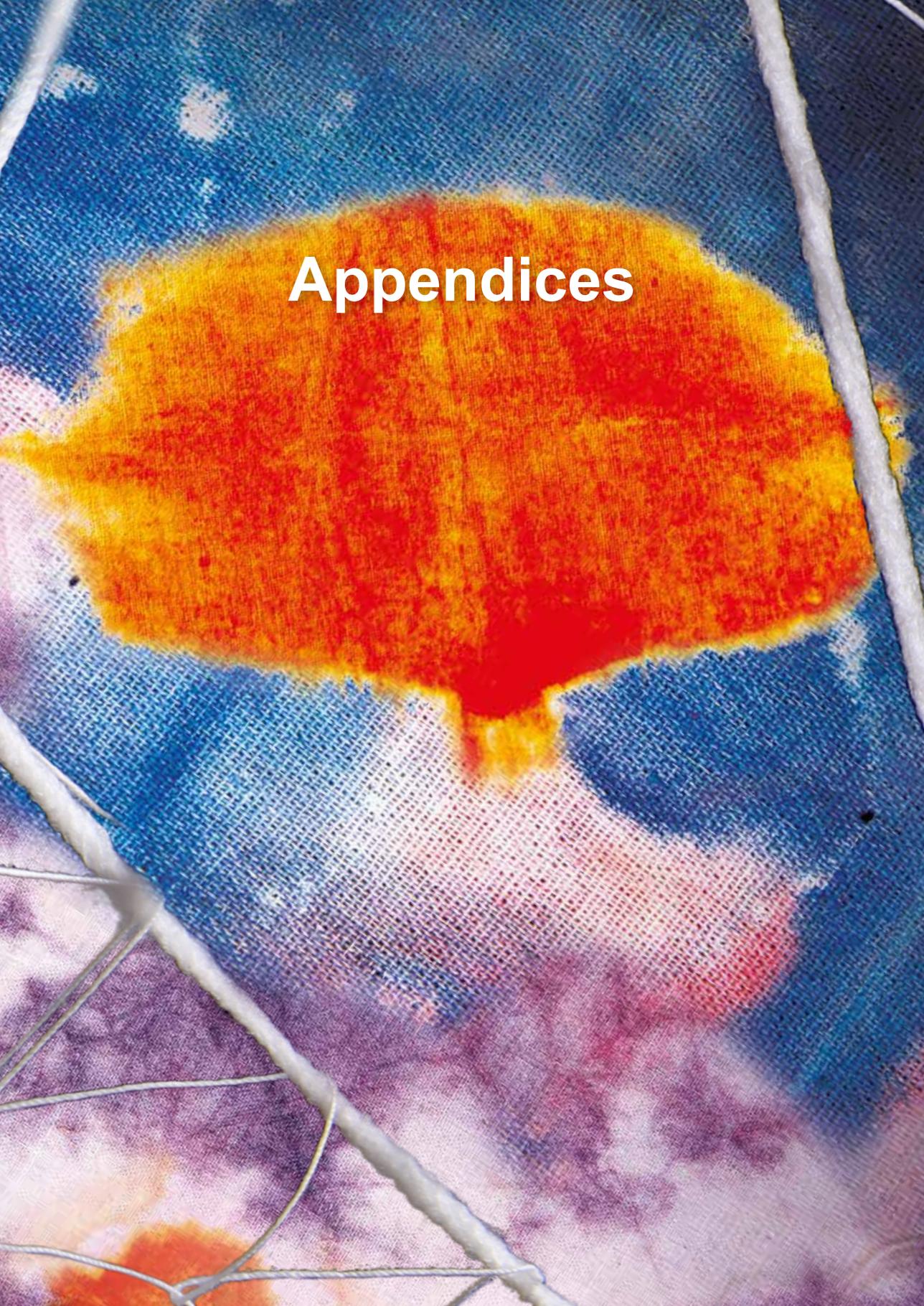
- 1 Wanke, K. A., Devanna, P. & Vernes, S. C. (2018). Understanding Neurodevelopmental Disorders: The Promise of Regulatory Variation in the 3'UTRome. *Biol Psychiatry* 83, 548-557, doi:10.1016/j.biopsych.2017.11.006.
- 2 Devanna, P. *et al.* (2018). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol Psychiatry* 23, 1375-1384, doi:10.1038/mp.2017.30.
- 3 Nobes, C. D. & Hall, A. (1995). Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* 81, 53-62, doi:10.1016/0092-8674(95)90370-4.
- 4 Ridley, A. J. & Hall, A. (1992). The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* 70, 389-399, doi:10.1016/0092-8674(92)90163-7.
- 5 Vernes, S. C. *et al.* (2008). A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359, 2337-2345, doi:10.1056/NEJMoa0802828.
- 6 Rodenas-Cuadrado, P., Ho, J. & Vernes, S. C. (2014). Shining a light on CNTNAP2: complex functions to complex disorders. *Eur J Hum Genet* 22, 171-178, doi:10.1038/ejhg.2013.100.
- 7 Rodenas-Cuadrado, P. *et al.* (2016). Characterisation of CASPR2 deficiency disorder--a syndrome involving autism, epilepsy and language impairment. *BMC Med Genet* 17, 8, doi:10.1186/s12881-016-0272-8.
- 8 Smogavec, M. *et al.* (2016). Eight further individuals with intellectual disability and epilepsy carrying bi-allelic CNTNAP2 aberrations allow delineation of the mutational and phenotypic spectrum. *J Med Genet* 53, 820-827, doi:10.1136/jmedgenet-2016-103880.
- 9 Kelava, I. & Lancaster, M. A. (2016). Stem Cell Models of Human Brain Development. *Cell Stem Cell* 18, 736-748, doi:10.1016/j.stem.2016.05.022.
- 10 Gunhanlar, N. *et al.* (2018). A simplified protocol for differentiation of electrophysiologically mature neuronal networks from human induced pluripotent stem cells. *Mol Psychiatry* 23, 1336-1344, doi:10.1038/mp.2017.56.
- 11 Penagarikano, O. *et al.* (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147, 235-246, doi:10.1016/j.cell.2011.08.040.
- 12 Lazaro, M. T. *et al.* (2019). Reduced Prefrontal Synaptic Connectivity and Disturbed Oscillatory Population Dynamics in the CNTNAP2 Model of Autism. *Cell Rep* 27, 2567-2578 e2566, doi:10.1016/j.celrep.2019.05.006.
- 13 Antoine, M. W., Langberg, T., Schnepel, P. & Feldman, D. E. (2019). Increased Excitation-Inhibition Ratio Stabilizes Synapse and Circuit Excitability in Four Autism Mouse Models. *Neuron* 101, 648-661 e644, doi:10.1016/j.neuron.2018.12.026.
- 14 Mossink, B. *et al.* (2022). Cadherin-13 is a critical regulator of GABAergic modulation in human stem-cell-derived neuronal networks. *Mol Psychiatry* 27, 1-18, doi:10.1038/s41380-021-01117-x.
- 15 Sun, A. X. *et al.* (2016). Direct Induction and Functional Maturation of Forebrain GABAergic Neurons from Human Pluripotent Stem Cells. *Cell Rep* 16, 1942-1953, doi:10.1016/j.celrep.2016.07.035.
- 16 Quadrato, G. *et al.* (2017). Cell diversity and network dynamics in photosensitive human brain organoids.

- Nature* 545, 48-53, doi:10.1038/nature22047.
- 17 Birey, F. *et al.* (2017). Assembly of functionally integrated human forebrain spheroids. *Nature* 545, 54-59, doi:10.1038/nature22330.
 - 18 Anderson, G. R. *et al.* (2012). Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proc Natl Acad Sci U S A* 109, 18120-18125, doi:10.1073/pnas.1216398109.
 - 19 Varea, O. *et al.* (2015). Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. *Proc Natl Acad Sci U S A* 112, 6176-6181, doi:10.1073/pnas.1423205112.
 - 20 Gdalyahu, A. *et al.* (2015). The Autism Related Protein Contactin-Associated Protein-Like 2 (CNTNAP2) Stabilizes New Spines: An *In Vivo* Mouse Study. *PLoS One* 10, e0125633, doi:10.1371/journal.pone.0125633.
 - 21 Jurgensen, S. & Castillo, P. E. (2015). Selective Dysregulation of Hippocampal Inhibition in the Mouse Lacking Autism Candidate Gene CNTNAP2. *J Neurosci* 35, 14681-14687, doi:10.1523/JNEUROSCI.1666-15.2015.
 - 22 Canali, G. *et al.* (2018). Genetic variants in autism-related CNTNAP2 impair axonal growth of cortical neurons. *Hum Mol Genet* 27, 1941-1954, doi:10.1093/hmg/ddy102.
 - 23 Gao, R. *et al.* (2018). CNTNAP2 stabilizes interneuron dendritic arbors through CASK. *Mol Psychiatry* 23, 1832-1850, doi:10.1038/s41380-018-0027-3.
 - 24 Argent, L. *et al.* (2020). Caspr2 interacts with type 1 inositol 1,4,5-trisphosphate receptor in the developing cerebellum and regulates Purkinje cell morphology. *J Biol Chem* 295, 12716-12726, doi:10.1074/jbc.RA120.012655.
 - 25 Higginbotham, H. *et al.* (2012). Arl13b in primary cilia regulates the migration and placement of interneurons in the developing cerebral cortex. *Dev Cell* 23, 925-938, doi:10.1016/j.devcel.2012.09.019.
 - 26 Higginbotham, H. *et al.* (2013). Arl13b-regulated cilia activities are essential for polarized radial glial scaffold formation. *Nat Neurosci* 16, 1000-1007, doi:10.1038/nn.3451.
 - 27 Park, S. M. *et al.* (2018). Brain Somatic Mutations in MTOR Disrupt Neuronal Ciliogenesis, Leading to Focal Cortical Dyslamination. *Neuron* 99, 83-97 e87, doi:10.1016/j.neuron.2018.05.039.
 - 28 Strauss, K. A. *et al.* (2006). Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med* 354, 1370-1377, doi:10.1056/NEJMoa052773.
 - 29 Fisher, S. E. (2006). Tangled webs: tracing the connections between genes and cognition. *Cognition* 101, 270-297, doi:10.1016/j.cognition.2006.04.004.
 - 30 Chiocchetti, A. G. *et al.* (2015). Variants of the CNTNAP2 5' promoter as risk factors for autism spectrum disorders: a genetic and functional approach. *Mol Psychiatry* 20, 839-849, doi:10.1038/mp.2014.103.
 - 31 Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T. & the, C.-c. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J Child Psychol Psychiatry* 58, 1068-1080, doi:10.1111/jcpp.12721.
 - 32 Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. (1994). The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory* 2, 103-127, doi:10.1080/09658219408258940.
 - 33 Coady, J. A. & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *Int J Lang Commun Disord* 43, 1-40, doi:10.1080/13682820601116485.
 - 34 Eising, E. *et al.* (2022). Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proc Natl Acad Sci U S A* 119, e2202764119, doi:10.1073/pnas.2202764119.
 - 35 Anijs, M., Devanna, P. & Vernes, S. C. (2022). ARHGEF39, a Gene Implicated in Developmental Language Disorder, Activates RHOA and Is Involved in Cell De-Adhesion and Neural Progenitor Cell Proliferation. *Front Mol Neurosci* 15, 941494, doi:10.3389/fnmol.2022.941494.
 - 36 Poliak, S. *et al.* (2003). Juxtaparanodal clustering of Shaker-like K⁺ channels in myelinated axons depends on Caspr2 and TAG-1. *J Cell Biol* 162, 1149-1160, doi:10.1083/jcb.200305018.
 - 37 Scott, R. *et al.* (2019). Loss of Cntnap2 Causes Axonal Excitability Deficits, Developmental Delay in Cortical Myelination, and Abnormal Stereotyped Motor Behavior. *Cereb Cortex* 29, 586-597, doi:10.1093/cercor/bhx341.
 - 38 Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J. & Kriegstein, A. R. (2021). Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat Neurosci* 24, 584-594, doi:10.1038/s41593-020-00794-1.
 - 39 Zhao, X. & Bhattacharyya, A. (2018). Human Models Are Needed for Studying Human Neurodevelopmental Disorders. *Am J Hum Genet* 103, 829-857, doi:10.1016/j.ajhg.2018.10.009.
 - 40 Fisher, S. E. & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends Genet* 25, 166-177, doi:10.1016/j.tig.2009.03.002.

- 41 Thomas, A. M., Schwartz, M. D., Saxe, M. D. & Kilduff, T. S. (2017). Cntnap2 Knockout Rats and Mice Exhibit Epileptiform Activity and Abnormal Sleep-Wake Physiology. *Sleep* 40, zsw026-zsw026, doi:10.1093/sleep/zsw026.
- 42 Hyman, S. E. (2021). Use of mouse models to investigate the contributions of CNVs associated with schizophrenia and autism to disease mechanisms. *Curr Opin Genet Dev* 68, 99-105, doi:10.1016/j.gde.2021.03.004.
- 43 Khodosevich, K. & Sellgren, C. M. (2023). Neurodevelopmental disorders-high-resolution rethinking of disease modeling. *Mol Psychiatry* 28, 34-43, doi:10.1038/s41380-022-01876-1.
- 44 Lattenkamp, E. Z. & Vernes, S. C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences* 21, 209-215, doi:10.1016/j.cobeha.2018.04.007.
- 45 Pfenning, A. R. et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346, 1256846, doi:10.1126/science.1256846.
- 46 Wang, R. et al. (2015). Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. *J Comp Neurol* 523, 892-906, doi:10.1002/cne.23719.
- 47 Vernes, S. C. (2017). What bats have to say about speech and language. *Psychon Bull Rev* 24, 111-117, doi:10.3758/s13423-016-1060-3.
- 48 Vernes, S. C. & Wilkinson, G. S. (2020). Behaviour, biology and evolution of vocal learning in bats. *Philos Trans R Soc Lond B Biol Sci* 375, 20190061, doi:10.1098/rstb.2019.0061.
- 49 Vernes, S. C. et al. (2022). The pale spear-nosed bat: A neuromolecular and transgenic model for vocal learning. *Ann NY Acad Sci* 1517, 125-142, doi:10.1111/nyas.14884.
- 50 Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Curr Opin Neurobiol* 28, 142-149, doi:10.1016/j.conb.2014.07.005.
- 51 Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science* 366, 55-58, doi:10.1126/science.aax0289.
- 52 Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys Life Rev* 11, 329-364, doi:10.1016/j.pprev.2014.04.005.
- 53 Koch, C. & Segev, I. (2000). The role of single neurons in information processing. *Nat Neurosci* 3 Suppl, 1171-1177, doi:10.1038/81444.
- 54 Brunel, N., Hakim, V. & Richardson, M. J. (2014). Single neuron dynamics and computation. *Curr Opin Neurobiol* 25, 149-155, doi:10.1016/j.conb.2014.01.005.
- 55 Yuste, R. (2015). From the neuron doctrine to neural networks. *Nat Rev Neurosci* 16, 487-497, doi:10.1038/nrn3962.
- 56 Fitz, H. et al. (2020). Neuronal spike-rate adaptation supports working memory in language processing. *Proc Natl Acad Sci U S A* 117, 20881-20889, doi:10.1073/pnas.2000222117.
- 57 Chiaradia, I. & Lancaster, M. A. (2020). Brain organoids for the study of human neurobiology at the interface of in vitro and in vivo. *Nat Neurosci* 23, 1496-1508, doi:10.1038/s41593-020-00730-3.
- 58 Bourke, J. L. et al. (2018). Three-dimensional neural cultures produce networks that mimic native brain activity. *J Tissue Eng Regen Med* 12, 490-493, doi:10.1002/term.2508.
- 59 Trujillo, C. A. et al. (2019). Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell* 25, 558-569 e557, doi:10.1016/j.stem.2019.08.002.
- 60 Samarasinghe, R. A. et al. (2021). Identification of neural oscillations and epileptiform changes in human brain organoids. *Nat Neurosci* 24, 1488-1500, doi:10.1038/s41593-021-00906-5.
- 61 Sharf, T. et al. (2022). Functional neuronal circuitry and oscillatory dynamics in human brain organoids. *Nat Commun* 13, 4403, doi:10.1038/s41467-022-32115-4.
- 62 Giraud, A. L. & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15, 511-517, doi:10.1038/nn.3063.
- 63 Kagan, B. J. et al. (2022). In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron* 110, 3952-3969 e3958, doi:10.1016/j.neuron.2022.09.001.



Appendices



English Summary

Most children acquire speech and language skills in the first few years of life in a way that seems effortless and without the need for formal instruction. However, developmental disorders that involve speech or language impairments are common, affecting up to 25% of school age children. The molecular architecture of speech and language disorders is complex. Most cases involve contributions from multiple genes or genetic risk factors. Sometimes mutations in the same gene contribute to the susceptibility for multiple disorders, suggesting functional links between these disorders. Studying the molecular and neurobiological roles of candidate language-related genes helps to understand the complex molecular architecture and spectrum of language disorder phenotypes.

In **Chapter 1** we provide an overview of the candidate language-related genes that have been identified so far, and we explain how genes form entry points to study gene networks and neuronal features that contribute to the ability of human language. In order to understand the function of a gene, it is necessary to consider genes and their products in the context of molecular networks that are formed through multiple types of molecular interactions. Different types of model systems, such as cell lines (used in chapter 5), animal models and human stem cell-derived neurons (used in chapter 3 and 4), each have their own set of advantages and limitations to study the functional properties of language-related genes.

The aim of this thesis was to understand how language-related genes and gene networks influence human neuronal features that contribute to the development of language-related traits. This was done by exploring genetic variation in patient populations (**Chapter 2**) and by using relevant *in vitro* model systems (**Chapter 3-5**) to understand the molecular mechanisms and neurodevelopmental processes that contribute to the development of language-related phenotypes.

CASPR2 deficiency disorder (CDD) is a rare syndrome that is caused by homozygous mutations in the gene *CNTNAP2*. The disorder is generally characterized by severe intellectual disability, early-onset drug-resistant epilepsy, and reduced or absent language. This broad range of phenotypic consequences suggest that CASPR2 (the protein product of *CNTNAP2*) has multiple functions and that this gene may have several interactions that could modify the effects of mutations in this gene. **Chapter 2** delves deeper into the clinical diversity of CDD by reporting the case of a patient with hyperkinetic stereotyped movements that are unlike other dyskinesias and have not been reported in the case of CDD. This individual inherited a missense mutation and partial duplication in *CNTNAP2* from his unaffected father and mother, respectively. These mutations affect the protein itself, but the maternal duplication also affects a target site for transcriptional regulation by FOXP2. It remains to be tested if this specific combination of coding and non-coding mutations contributed together to this novel motor phenotype. In any case, this report emphasizes the importance of extending patient screening to non-coding regions of *CNTNAP2*, especially if only a single heterozygous *CNTNAP2* variant has been identified or if atypical neurological phenotypes are present.

In **chapter 3**, a human stem cell-derived neuronal network model was used to investigate how mutations in CASPR2 isoforms may differentially affect the development of disorder-related neuronal phenotypes. Not every *CNTNAP2* mutation affects both the long and short isoform, which may be a cause for variation in CDD. To test this hypothesis, multiple mutations in *CNTNAP2* were engineered in the genome of human embryonic stem cells. Neuronal networks derived from these cells were compared for the consequences of each of these mutations on cell type composition, number of synapses, and global gene expression. Homozygous loss of the long isoform of CASPR2 affects the proportion of neural precursor cells, the number of excitatory synapses, and gene expression for several biological processes, whereas heterozygous loss of the long isoform only produces a subset of the gene expression changes. Gene expression data further revealed that short isoform of *CNTNAP2* is expressed early in development at high levels and that networks that lack both the long and short isoform of CASPR2 show disrupted expression of cilia assembly genes. Further investigation of a role for *CNTNAP2* in cilia function during early brain development could help clarify its roles in neuronal migration and cortical lamination, and provide a neurobiological explanation for clinical variation between cases with mutations that affect different isoforms.

In **Chapter 4**, the gene expression network around *CNTNAP2* is further explored by looking at the expression pattern of microRNAs (miRNAs). These regulatory molecules play key roles in coordinating several neurodevelopmental processes by silencing the expression of target genes. We first showed that the miRNA expression patterns in *in vitro* developing human neuronal networks resemble those of the developing human cortex. Then we showed that subsets of miRNAs change in expression activity upon loss of *CNTNAP2* in human neuronal networks. A majority of changes could be explained by changes in cell type composition. By integrating miRNA expression data with target gene expression data, we further identified six miRNAs that change expression in such a way that they seem to counterbalance a known neurite outgrowth deficit in neurons with reduced *CNTNAP2* expression. Genes that inhibit neurite outgrowth were under stronger repression by some of these active miRNAs, whereas genes that promote neurite outgrowth were targeted by downregulated miRNAs. These miRNAs and their targets can be considered as extensions of the molecular network related to *CNTNAP2* function during cortical development and point to places in the genome where additional mutations could have modifying effects that may account for differences in clinical phenotypes.

Chapter 5 investigates the function of *ARHGEF39*. This gene is an example of how studying miRNA regulation can provide an entry point to study the molecular underpinnings of language-related disorders, as *ARHGEF39* was linked to language disorder phenotypes via a polymorphism that disrupts post-transcriptional regulation by miRNAs. Using cellular assays, we showed that *ARHGEF39* activates RHOA, a member of the Rho family of GTPases that regulate intracellular actin dynamics and cell shape, and that overexpression of *ARHGEF39* leads to cell de-adhesion. Analysis of available single cell RNA-sequencing datasets from developing mouse and human brains show that *ARHGEF39* expression is a marker for proliferating progenitor cells, and that its expression is associated with cell division and other RHOA regulating genes. Further investigation of the effect of *ARHGEF39* expression

levels on neural progenitor cell division and attachment could help to understand its role in neocortical development and may provide an explanation for the contribution of altered *ARHGEF39* expression levels to the risk for language impairments.

Finally, in **chapter 6**, we summarize the research of the thesis, and discuss how results from these studies have expanded our knowledge about human neurobiological processes that are affected by *CNTNAP2*, *ARHGEF39*, and microRNA networks. This advances our understanding of the role of *CNTNAP2* in disorder and offers approaches to unfolding the genetic and phenotypic complexities of language-related disorders. Further, we discuss how the combination of human cell models with animal models and with computation models may offer deeper insight in the mechanisms by which genes such as *CNTNAP2* and *ARHGEF39* affect brain structure, complex network dynamics, behaviour and computation. This promises valuable steps to further our understanding of the influence of candidate language-related genes on neurodevelopment and cognition.

Nederlandse samenvatting

De meeste kinderen verwerven spraak- en taalvaardigheden in de eerste paar jaar van hun leven op ogenschijnlijk moeiteloze wijze en zonder behoefte aan formele instructies. Ontwikkelingsstoornissen met beperkingen in taal en spraak zijn echter ook veelvoorkomend: in tot wel 25% van de kinderen in de schoolgaande leeftijd. De onderliggende moleculaire oorzaken zijn complex. De meeste gevallen zijn belast met de bijdragen van meerdere genen of genetische risicofactoren. Soms kunnen mutaties in hetzelfde gen ook bijdragen aan risico's voor meerdere stoornissen. Het bestuderen van de moleculaire en neurobiologische functies van kandidaten voor taalgerelateerde genen kan ons begrip geven van de complexiteit van de onderliggende moleculaire oorzaken en het spectrum aan verschijningsvormen van taalstoornissen.

Hoofdstuk 1 biedt een overzicht van de kandidaten voor taalgerelateerde genen die tot dusver zijn geïdentificeerd en we gaan in op hoe genen een ingang vormen voor het bestuderen van gennetwerken en neuronale eigenschappen die bijdragen aan menselijke taalvaardigheid. Om de functie van een gen te kunnen begrijpen is het nodig om genen en hun producten te beschouwen in de context van moleculaire netwerken die worden gevormd door meerdere typen van moleculaire interacties. Verschillende soorten modelsystemen, zoals cellijnen (gebruikt in hoofdstuk 5), diersystemen en neuronen gekweekt uit menselijke stamcellen (gebruikt in hoofdstuk 3 en 4), hebben ieder hun eigen set van voordelen en beperkingen om de functionele eigenschappen van taalgerelateerde genen te bestuderen.

Het doel van deze thesis was om te begrijpen hoe taalgerelateerde genen en gennetwerken de eigenschappen van menselijke neuronen beïnvloeden die bijdragen aan de ontwikkeling van taalgerelateerde kenmerken. Dit werd gedaan door genetische variatie in patiënten te bestuderen (**Hoofdstuk 2**) en met behulp van relevante *in vitro* modelsystemen (**Hoofdstuk 3-5**) om inzicht te krijgen in de moleculaire mechanismen en processen van de hersenontwikkeling die bijdragen aan de ontwikkeling van taalgerelateerde fenotypes.

CASPR2 deficiëntie stoornis (CDD) is een zeldzaam syndroom dat wordt veroorzaakt door homozygote mutaties in het gen *CNTNAP2*. De stoornis wordt over het algemeen gekenmerkt door ernstige verstandelijke beperking, vroegtijdige medicijn-resistente epilepsie en verminderd of afwezig taalvermogen. Dit brede palet aan fenotypische gevolgen suggereert dat CASPR2 (het eiwit afkomstig van *CNTNAP2*) meerdere functies heeft en dat dit gen verscheidene interacties heeft die de effecten van mutaties in dit gen kunnen beïnvloeden. **Hoofdstuk 2** gaat dieper in op de klinische diversiteit van CDD aan de hand van de casus van een patiënt met een hyperkinetische en stereotiepe-bewegingsstoornis die anders is dan andere dyskinesieën en niet eerder is gerapporteerd in een geval van CDD. Dit individu heeft een *missense*-mutatie en gedeeltelijke duplicatie in *CNTNAP2* van respectievelijk zijn gezonde vader en moeder. De mutaties veranderen de structuur van het eiwit, maar de van de moeder afkomstige duplicatie raakt ook een bindingsplaats voor transcriptieregulatie door FOXP2. Het zal nog moeten blijken of deze specifieke combinatie van coderende en niet-coderende mutaties bijdraagt aan

dit nieuwe bewegingsfenotype. Deze rapportage benadrukt in ieder geval het belang om ook niet-coderende gebieden van *CNTNAP2* mee te nemen in de screening van patiënten, met name als slechts een enkele heterozygote variant in *CNTNAP2* is gevonden of als atypische neurologische kenmerken zich voordoen.

Voor **hoofdstuk 3** werd gebruik gemaakt van netwerken van neuronen die zijn gekweekt uit menselijke stamcellen om te bestuderen hoe mutaties in verschillende CASPR2 isoformen de ontwikkeling van ziektegerelateerde neuronale fenotypes verschillend kunnen beïnvloeden. Niet iedere mutatie in *CNTNAP2* beïnvloedt zowel de lange als de korte isoform, wat een reden kan zijn voor variatie in de presentatie van CDD. Om deze hypothese te testen werden meerdere knock-out mutaties aangebracht in *CNTNAP2* in menselijke embryonale stamcellen. Neuronnetwerken gekweekt uit deze cellen werden vergeleken voor de consequenties van ieder van deze mutaties op de aanwezige proporties van verschillende celtypen, het aantal synapsen en algehele genexpressie. Homozygote knock-out van de lange isoform van *CNTNAP2* beïnvloedt de proportie van neurale precursorcellen, het aantal excitatoire synapsen en genexpressie voor meerdere biologische processen, terwijl heterozygote knock-out alleen een deel van de veranderingen in genexpressie laat zien. Genexpressiedata toonde verder dat de korte isoform al vroeg in de ontwikkeling in hoge mate tot expressie komt en dat neuronnetwerken die zowel de lange als de korte isoform van CASPR2 missen een verstoorde expressie van genen betrokken in de opbouw van cilia laten zien. Verder onderzoek van de rol van *CNTNAP2* in ciliafunctie tijdens de vroege hersenontwikkeling zou kunnen helpen om de rol van dit gen in de migratie van neuronen en corticale laagvorming en het zou een neurobiologische verklaring kunnen bieden voor klinische variatie tussen gevallen waarin mutaties in *CNTNAP2* verschillende isoformen beïnvloeden.

In **hoofdstuk 4** wordt het genexpressienetwerk rondom *CNTNAP2* verder verkend door te kijken naar de expressiepatronen van microRNAs (miRNAs). Deze regulerende moleculen spelen sleutelrollen in de coördinatie van verscheidene processen in de hersenontwikkeling door de expressie van doelgenen te stillen. We laten als eerste zien dat de miRNA expressiepatronen van *in vitro* ontwikkelende menselijke neuronnetwerken gelijkenis vertonen met die van de ontwikkelende menselijke cortex. Daarna tonen we dat subsets van miRNAs veranderen in expressie na verlies van *CNTNAP2* in menselijke neuronnetwerken. De meeste van deze veranderingen worden verklaard door veranderingen in de compositie van celtypen. Door miRNA expressiedata te integreren met expressiedata over doelgenen konden we zes miRNAs identificeren die zodanig van expressie veranderden dat ze tegenwicht lijken te bieden aan een eerder onderzocht tekort in de groei van neurieten vanuit neuronen met verminderde expressie van *CNTNAP2*. Genen die de uitgroei van neurieten remmen werden sterker onderdrukt door miRNAs, terwijl genen die deze uitgroei stimuleren juist werden gereguleerd door miRNAs die minder tot expressie kwamen. Deze miRNAs en hun doelgenen kunnen worden beschouwd als uitbreidingen van het moleculaire netwerk waarin *CNTNAP2* functioneert tijdens de ontwikkeling van de hersenschors en zij kunnen locaties in het DNA aanwijzen waar extra mutaties mogelijk modificerende effecten kunnen hebben die verschillen tussen klinische fenotypes zouden kunnen verklaren.

Hoofdstuk 5 onderzoekt de functie van *ARHGEF39*. Dit gen is een voorbeeld van hoe het bestuderen van regulatie door miRNAs een ingang kan bieden voor het onderzoeken van de moleculaire oorzaken van taalstoornissen, aangezien *ARHGEF39* in verband werd gebracht met taalontwikkelingsstoornis via een polymorfisme dat posttranscriptionele regulatie door miRNAs verstoort. Met behulp van proeven in cellijnen hebben we laten zien dat *ARHGEF39* RHOA activeert, een lid van de Rho-familie van GTPases die de dynamiek van intracellulair actine en daarmee de vorm van de cel reguleren, en dat overexpressie van *ARHGEF39* leidt tot de loslating van cellen. Analyse van gepubliceerde single cell RNA-sequencing datasets van ontwikkelende muizen- en mensenhersenen liet zien dat expressie van *ARHGEF39* specifiek kenmerk is voor prolifererende neurale progenitorcellen en dat expressie van *ARHGEF39* is geassocieerd met de expressie van genen die betrokken zijn in celdeling en andere genen die RHOA reguleren. Verder onderzoek naar het effect van *ARHGEF39* expressieniveaus op celdeling en aanhechting van neurale progenitorcellen kan inzicht geven in de rol van *ARHGEF39* in de ontwikkeling van de hersenschors en kan helpen verklaren hoe veranderde expressieniveaus van *ARHGEF39* bijdragen aan het risico voor taalontwikkelingsstoornis.

Tot slot wordt in **hoofdstuk 6** het onderzoek in deze thesis samengevat en wordt besproken hoe resultaten uit deze studies hebben bijgedragen aan onze kennis over de menselijke neurobiologische processen die worden beïnvloed door *CNTNAP2*, *ARHGEF39* en microRNA-netwerken. Dit brengt ons begrip over de rol van *CNTNAP2* in ziekte vooruit en biedt nieuwe manieren om de genetische en fenotypische complexiteiten van taal-gerelateerde stoornissen te onderzoeken. Verder bespreken we hoe het combineren van menselijke celmodellen met diermodellen en met computermodellen een dieper inzicht kan verschaffen in de mechanismes waarin genen zoals *CNTNAP2* en *ARHGEF39* invloed hebben op hersenstructuur, hersenactiviteit, gedrag en informatieverwerking. Dit belooft waardevolle stappen in de uitbreiding van onze kennis over de invloed van taalgerelateerde genen op hersenontwikkeling en cognitie.

Acknowledgements

Genes, proteins, microRNAs, and neurons are all part of complex networks that affect their function. Similarly, the work in this PhD thesis has been made possible by many interactions. Here, I want to extend my gratitude for this support from many people, including my supervisors, collaborators, colleagues, friends and family.

First, I want to thank my supervisor **Sonja Vernes**. We have worked together for so many years. Our journey together started in the years before this PhD, when I did a Master's internship in your research group. Next, you supported me to go on an adventure in the United States to learn the art of growing human stem cells and cortical organoids. It was a unique opportunity to be able to apply those skills in this project and learn many new ones. I am grateful for the independence that you gave me. Your determination, eye for detail, and persuasion skills will keep inspiring me.

I also want to thank my other supervisor **Simon Fisher** for the environment that he created. The Language & Genetics department is a fascinating environment for all the diversity of research projects that it brings together. It also has the fanciest and best equipped laboratory that I have ever worked in. I am grateful for all the input and support that I received.

My thesis was evaluated by the members of the Manuscript Committee, **Erik Storkebaum**, **Esther Becker**, and **Dirk Schubert**. I want to thank them for their time and consideration.

The work in this thesis was made possible by both external and internal collaborations. I want to thank **Pasquale Striano** and **Marcello Scala** for bringing their clinical expertise to the collaboration that led to chapter 2. I am deeply grateful to **Steven Kushner** and **Femke de Vrij** at the Erasmus University Medical Center for sharing their optimized neuronal network differentiation protocol and their comments on chapter 3. I also thank them and **Bas Lendemeijer** for the practical demonstration of this protocol and answering additional practical questions. Further, I want to thank **Rob Woestenenk**, head of the Flow Cytometry facility of Radboud University, for his help with sorting the neural precursor cells.

My colleague and friend **Ksenia Lavrichenko** was essential in the design and execution of the whole genome analysis of the embryonic stem cell lines in chapter 3. I am grateful for everything that you taught me about bioinformatics and all the other wisdom that you shared with me. **Paolo Devanna's** unmatched love for microRNAs fed me during chapter 4 and his work on non-coding variation in neurodevelopmental disorders laid the foundation for chapter 5. Paolo, you have also been an anchor in the lab for me since my first internship at the MPI. **Ine Alvarez-van Tussenbroek** was the best neighbour at the desk that I could have wished. Ine, the co-writing sessions that you organized got me back on track after the pandemic and helped me finish my chapters. Ine and Paolo, I am also thankful for the countless coffee breaks that we had and I am extremely happy to have you as paranympths and friends.

During my PhD, I also had the pleasure to supervise a few students. I want to thank **Yarick Winmai** for his hard work to make CRISPR-mediated homology directed repair work as part of his bachelor thesis project with the for Life Sciences at

Hogeschool Arnhem Nijmegen. **Nikki Kolsters** helped me out as a student-assistant when the neuronal network differentiation work became really busy.

I have also been fortunate to have collaborated on two extra projects during my PhD. I thank **Dick van den Broek** and **Marvin Uhlmann** for introducing me into computational neuroscience and explaining what a neural network means to them. It has been really exciting to develop our ideas into an interdisciplinary research project and win the first Levelt innovation award for cross-disciplinary language research together. I also want to thank **Susanne Schmid** for the opportunity to do an exchange project comparing human and rat *CNTNAP2* knockout neurons at Western University of Ontario. She and her lab members made me feel very welcome in London in the beginning of 2020, despite the circumstances of Canadian winter and later the COVID-19 lockdown. I especially thank **Rajkamal Mann** for showing me the intricacies of whole-cell patch clamp electrophysiology and playing basketball and squash tennis together.

I also would like to thank many more colleagues at the Max Planck Institute for Psycholinguistics. **Pedro, Kai, Martin, Jon-Ruben, Moritz, Janine, Laura Nienke, Ella, Katharina, Lisa, Meret, Paul, Elpida**, for so much fun inside and outside of the weekly NVC meetings. **Arianna, Sara, Elliot, Joery, Maggie, Lot, Else, Fabian**, for sharing the lab space, asking questions and everyday help. From sharing meals and coffee breaks to going on trips together and from making musical discoveries together to celebrating international pancake day, there are too many good memories to name. I also thank **Jasper** and **Jurgen** for their support as lab management and **Martina** for her help in organizing almost everything. **Kevin** has helped me to find my way in the many procedures of completing a PhD, as coordinator of the International Max Planck Research School.

Further, I would like to thank a number of friends for their moral support. **Alice** and **Giuliana**, for staying in touch and reminding me about our wonderful time at Yale University together. **Justin**, among other things for our trip through Bulgaria in 2017. **Ike**, for regular check-ins and conversations about literature and world affairs. **Ksenija**, for animated and passionate stories. **Tom**, for your groundedness that always reminds me of what is important. **Toon** and **Else**, for your advice that came at the right moment for me. **Barbara**, for your sharp observations and love for drama. **Charlotte**, for your creativity and absurd sense of humor that let me see with new eyes. **Ida**, for building for a new peace movement together.

Finally, I want to thank my family and partner. **Albert** en **Femke**, papa en mama, bedankt voor alles wat jullie me hebben gegeven, voor het rijkelijk voeden van mijn nieuwsgierigheid, voor jullie luisterend oor en voor de moed om altijd jezelf uit te dagen. **Damiët**, lieve zus, dankjewel voor de mooie kaft en ik bewonder jouw bezieling als maker. Ook over meerdere grote afstanden heb ik veel steun van jou ontvangen. Lieve **Maartje**, ik ben heel gelukkig dat ik jou ook mag bedanken. Jij brengt me enorm veel liefde, vreugde en optimisme voor de toekomst.

Curriculum Vitae

Midas Anijs was born on the 10th of July 1992 in Nijmegen, the Netherlands. He completed his secondary education at Stedelijk Gymnasium Nijmegen in 2010, and started the bachelor's programme Biomedical Sciences at Radboud University in Nijmegen. During this period, he also participated in the Faculty of Medical Sciences Honours Programme. This provided him the opportunity to do a research internship abroad in the UK, in the laboratory of prof. dr. Robert Lightowlers and prof. dr. Zofia Chrzanowska-Lightowlers in the Wellcome Centre for Mitochondrial Research at Newcastle University. Here, he wrote his bachelor thesis on the potential role of the protein neugrin in mitochondrial translation. He obtained his bachelor's degree in 2013.

In the same year, Midas continued with the research master's programme Molecular Mechanisms of Disease at the Radboud University, which he completed in 2015 with the distinction cum laude. As part of this degree, he did a research internship at the Max Planck Institute for Psycholinguistics in Nijmegen in 2014 under supervision of dr. Sonja Vernes and Paolo Devanna, studying the functional properties of evolutionary variants of the transcription factor FOXP2. In February 2015, he moved to New Haven, Connecticut in the United States, to work on his second internship project and Master's thesis in the laboratory of Prof. dr. Flora Vaccarino at Yale University. With direct supervision from dr. Jessica Mariani, he studied the pathogenesis of FOXG1 syndrome in cortical organoids that were generated from patient-derived induced pluripotent stem cells. His thesis was awarded with a prize by the Dutch Rett Syndrome Society.

In October 2015, Midas returned to Nijmegen, being awarded a four-year International Max Planck Research School (IMPRS) for Language Sciences PhD Fellowship. There, he started his doctoral research in the Neurogenetics of Vocal Communication research group led by dr. Sonja Vernes, with the ambition to set up a new cellular model to investigate the roles of candidate language-related genes in neurodevelopment. His project was co-supervised by Prof. dr. Simon Fisher, head of the Language & Genetics Department. Work included in this thesis was presented at international conferences, such as the Society for Neuroscience Meeting in 2018 and the advanced course on RNA Transcriptomics at the Wellcome Genome Campus in Cambridge, UK in 2018. Next to his main projects, Midas also created new multidisciplinary collaborations. In 2017, Midas Anijs, Dick van den Broek and Marvin Uhlmann were awarded the first Levelt Innovation Award for cross-disciplinary language research for a project that combined genetics and computational neuroscience. They fed data from *CNTNAP2* mutant neuronal networks to a computational neuronal network model of sentence processing to build bridges between their research models and fields. In the beginning of 2020, he spent three months in the laboratory of Prof. dr. Susanne Schmid at Western University of Ontario in London, Canada with a project grant from the Donders Institute and Brain and Mind Institute exchange programme. This project aimed to facilitate interspecies translation between two disease models by comparing the electro physiological and morphological properties of human neurons and rat neurons with knockout mutations in *CNTNAP2*.

Since 2021, Midas has worked as a lecturer at Radboud University's Faculty of Social Sciences.

Publications

- Scala, M.*, [Anijs, M.*](#), Battini, R., Madia, F., Capra, V., Scudieri, P., Verrotti, A., Zara, F., Minetti, C., Vernes, SC., and Striano, P. (2021). Hyperkinetic stereotyped movements in a boy with biallelic CNTNAP2 variants. *Ital J Pediatr.* 12;47(1):208, doi: 10.1186/s13052-021-01162-w
- [Anijs, M.](#), Devanna, P., and Vernes, SC. (2022) *ARHGEF39*, a Gene Implicated in Developmental Language Disorder, Activates RHOA and Is Involved in Cell De-Adhesion and Neural Progenitor Cell Proliferation. *Front Mol Neurosci.* 15, 941494, doi: 10.3389/fnmol.2022.941494

