Discovery and Neurobiological Modeling of Rare Genetic Variant Risk for Mental Illness

Job Obe Zeno de Jong

Lay-out and printing by Optima Grafische Communicatie (www.ogc.nl)

Ontdekking en neurobiologisch modeleren van zeldzame genetische risicofactoren voor psychiatrische aandoeningen

Discovery and Neurobiological Modeling of Rare Genetic Variant Risk for Mental Illness

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. F.A. van der Duijn Schouten

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
28 september om uur 15.30 uur

door

Job Obe Zeno de Jong
geboren te Amsterdam.

Erasmus University Rotterdam

Erasmus

PROMOTIECOMMISSIE:

Promotor:	Prof. Dr. S.A. Kushner
Overige leden:	Dr. F. de Vrij Prof. Dr. Y. Elgersma Dr. H. Bruining Dr. L. De Witte Dr. B. Xu Prof. Dr. H. Wichterle
Copromotor:	Dr. S Markx

TABLE OF CONTENTS

Chapter 1

Introduction

INTRODUCTION

Mental illness is a major contributor to global disease burden (1). The prevention and treatment of mental illness pose a great challenge that will require integrated efforts in a multitude of disciplines ranging from politics to public health strategies to the development of new efficacious treatments. In order to realize the latter, we need a more thorough understanding of the biology of the functioning and dysfunctioning of the human brain. The focus of this work lies here, with an emphasis on the role of genomics in understanding the biology of mental illness, with the ultimate goal of working towards novel treatment strategies to improve public mental health. The scope of this work ranges from the discovery of a rare genetic variant predisposing for depression in a genetically isolated population to the modeling of rare, disease-causing genetic lesions for autism spectrum disorder and schizophrenia in stem cell-derived neuronal models.

Psychiatric disorders have been an enigmatic field of study, in part due to the complex nature of the causal determinants that lead to disease. The field of modern psychiatry has transitioned through multiple scientific paradigms in terms of thinking about causality –and partially as a consequence- the treatment of mental illness. The early 20th century was characterized by a strong emphasis on early life events that were thought to be causative to psychopathology, which were explored by psychoanalysts in talk-therapy sessions in an attempt to cure the mental disorder the patient was suffering from. Over the course of the century there was a gradual shift towards a more *biological* perspective on mental illness and talk-therapies were augmented or replaced by various pharmacotherapies, many of which are still used to this day (2).

Most pharmacotherapies used in psychiatry were introduced to the clinic in the second half of the 20th century. For example, lithium was discovered as a psychotropic medication with mood stabilizing effects for the first time in 1949, and in the following decade the major classes of other pharmacotherapies, including antipsychotics, benzodiazepines and antidepressants were introduced (3). These medications have in common that –like many medications used in medicine- none of them were designed to interfere with a known disease mechanism and thereby reverse biological processes and alleviate symptoms or cure disease. Rather, discoveries took place in a serendipitous manner. For example, the finding that MAO-inhibitors had a positive effect on mood symptoms and D2-receptor antagonists on psychotic symptoms offered an entry point for studying the neural substrate of these symptoms. Through this route, the involvement of the serotonin system in mood regulation and depression (4), and the involvement of the dopamine system in psychosis were uncovered (5).

This mode of studying biology can be seen as more of a *top-down* approach, where clues from scientific (or clinical) data that are possibly 'downstream' are followed up to study their po-

tential causes and correlates. Although this approach has led to important scientific insights, it has also led to persisting biased views on the nature and mechanisms of mental illness (5). Studying the human genome in an unbiased fashion in an attempt to study disease biology is more of a 'bottom-up' approach, where the entry-point of study, the genome, can be seen as a closer to root-causal phenomenon –that is, if the study design is chosen in a particular way. The involvement of the genome in mental illness was hypothesized after these appeared to be highly heritable.

The heritability of psychiatric disorders

An important lead about the etiology of psychiatric disorders came from the observation that mental illness tends to aggregate in families, suggesting a degree of heritability. This observation was confirmed in the second half of the twentieth century in a large number of family, twin- and adoption studies that aimed to quantitatively distinguish environmental from non-environmental (i.e. genetic) disease risk (6). These studies revealed a relatively high but varying degree of heritability for the major psychiatric illnesses. Heritability (h^2) is defined as a ratio of variances, in which the variance of genotypic values is divided by the variance in phenotypic (e.g. disease status) values ($h^2 = \sigma^2_{genotype}/\sigma^2_{phenotype}$)(7), leading to h^2 values ranging between 0 and 1. For example, major depressive disorder (MDD) has an estimated twin heritability of 0.37, autism spectrum disorder (ASD) 0.75 and schizophrenia (SCZ) 0.81 (8). So a high heritability –such as heritability of 0.75 as is seen in ASD- means that a lot of the variance of a phenotype in a given population can be explained by the variance in the genotypes in that same population. In the case of a lower heritability such as the heritability of 0.37 that is observed in MDD, a lower proportion of the phenotypic variance is explained by genotypic variance in a given population. From this logically follows that for a condition such as MDD there is a higher proportion of phenotypic variability that can be attributed to non-genotypic –i.e. environmental, variables. It is important to point out that heritability by itself does not provide any information about the genetic architecture of a phenotype. A phenotype that is highly heritable could have either one locus causal to disease status (e.g. the monogenic Huntington's disease), or many genetic loci contributing to disease (e.g. ASD).

The genomics revolution and its impact on psychiatric genetics

As with most scientific progress, the elucidation of the genetic architecture of neuropsychiatric disorders was preceded by the development of methods and technologies that enabled researchers to find answers to their questions. Genetic architecture refers to all genomic variations, including single base-pair substitutions (e.g. SNPs, missense, nonsense mutations), insertions and deletions (indels) and structural variation (CNVs) that confer risk for illness. Examples of early successes in elucidating the genetic architecture of brain disorders are the identification of the genomic region harboring the gene causative to Huntington's chorea using linkage analysis in 1983 in a genetically isolated population (9), and the association of common genetic

variation in the *APOE* gene with Alzheimer's disease using a candidate gene approach (10). After the completion of the human genome project (11), and with decreasing costs of whole exome (WES) and whole genome sequencing (WGS), Miller syndrome was the first Mendelian neurologic disorder for which the causative gene was identified using WES (12). Following this, many causative genes for Mendelian disorders were identified using similar approaches, with 3600+ disease gene-phenotype pairs described today (13).

The aforementioned examples are disorders for which variation in single genes lead to a disease phenotype. Some of these are relatively more common (e.g. Huntington's), but most are extremely rare on a population-wide level (e.g. Miller syndrome). For most common psychiatric disorders encountered in a clinical setting it became clear that the genetic architecture was quite different. A large number of gene mapping efforts revealed no reproducibly identifiable single genes (or even several genes) responsible for, or associated with common psychiatric disorders (14). Both linkage studies assuming a Mendelian inheritance pattern (even though inheritance patterns for mental illness in families typically do not appear to be classically Mendelian), and the hypothesis-driven candidate gene approach that extensively looked at the synthetic, degradative or receptor components of neurotransmitter systems, failed to produce any robustly replicable results (5, 15).

Common variation in psychiatric disorders

Genome-wide association (GWA) studies were a very promising design to overcome some of the obstacles encountered in linkage- and candidate gene designs. SNP-microarrays became inexpensive quickly after their introduction and thus enabled researchers to investigate relatively large cohorts of cases and controls. Furthermore, with SNP markers being distributed evenly across all chromosomes, the entire genome could be investigated in an unbiased fashion. Despite these advantages, the first years of GWA studies in psychiatry were again characterized by failures to identify any reproducible genomic loci that reached the threshold for genome-wide statistical significance. For example, nine GWA studies on MDD published between 2010 and 2013 with sample sizes ranging between 1000-9500 cases, did not reveal any genome-wide significant loci (15).

Successes using the GWA-approach began to emerge as sample sizes increased (16-18), and phenotypes were more narrowly defined (19). This development was facilitated by the formation of global research consortia such as the psychiatric genomics consortium (PGC) that allowed for the analysis of very large sample sizes and harmonized big data analyses with consistent quality control. To date, GWA studies have wielded 100+ loci for SCZ (20-23), 102 for MDD (16, 24-26) and 12 for ASD (27). Note that the sample size for e.g. the meta-analysis of MDD GWA-studies (24) exceeds a sample size of 800 000 cases and controls.

The GWA study design is suited to discover SNPs associated with disease that are commonly present in the population (minor allele frequency (MAF) > 1%). Individually, these SNPs typically confer only a small risk to disease status, with Odds Ratios (ORs) ranging between 1.05-1.15. In recent years, an increasing number of publications have investigated multiple SNPs that are differentially present in cases versus controls, leading to the generation of polygenic risk scores (PRS) for psychiatric disorders (28, 29). For SCZ, for example, when sorted on effect size (OR), inheriting the top 10% versus the bottom 10% of SNP-risk variants, increases one's chance for developing SCZ by 10-fold (30). Although informative on the role of SNPs on the genetic architecture and biology of SCZ, the positive- and negative predictive values of these PRSs are currently too low to make a clinically meaningful contribution to a psychiatric diagnosis (30).

To conclude, numerous publications with negative findings from linkage and candidate studies and eventually successful GWA studies have taught us that the genetic architecture of psychiatric disorders has a highly *polygenic* nature. This means that an estimated hundred to thousands (15) of genomic loci may contribute to disease risk in a given population. In this regard, psychiatric disorders are etiologically comparable to a complex trait such as height, where a very large number of alleles, in combination with environmental variables influence the outcome of the trait.

In parallel to the discoveries on the role of common SNPs was the elucidation of the role of rare genetic variation in psychiatric disorders. During this era, there were two prevailing hypotheses about the genetic architecture of mental illness. The common-disease-common-variant (CDCV) hypothesis postulated that the largest contribution of to disease risk arises from common variants to disease risk, whereas the common-disease-rare-variant (CDRV) hypothesis proposed the largest contribution from rare variation (31). Since then, it has become increasingly clear that the reality is a combination of these two hypotheses, and both types of variation play an important role in disease risk. This can be explained well from an evolutionary perspective. As new alleles constantly arise at a rate of approximately 175 per diploid genome per generation (32), variants with large effects on disease status will be under substantial evolutionarily pressure due to decreased fitness, and thus will remain rare (and typically relatively recent in their emergence) at the population level. Alleles conferring modest disease risk are subject to progressively decreasing natural selection pressure and therefore often present at higher frequencies with more evolutionarily distant emergence (33).

Rare variation in psychiatric disorders

The first findings on the role of rare genetic variation on psychiatric disease risk were CNVs that were found in patients with ASD and SCZ (34-36); disrupting the function of one or multiple genes. Following these early discoveries, WES and WGS studies have continued to reveal novel,

rare and highly penetrant variants implicated in psychiatric disease. Rare variant discovery has thus far been most successful for ASD, for which to date over 100 *high-confidence genes* have been described (37). These are usually *de novo*, protein-truncating variants (PTVs), that are individually rare but are estimated to jointly account for around 15% of cases of ASD (38). For SCZ, there have so far only been ten genes described in the published literature (39-41) in sample sizes that revealed numerous genes for ASD. For MDD, very little is known about the contribution of rare variants to disease, with to date only a single study showing segregation of a missense variant with MDD (42). For rare variants, the evidence of their involvement in disease has been strongest for PTVs. However, missense variants have also been implicated in ASD, although reported effect sizes are typically smaller than effect sizes for PTVs (43, 44).

Whereas common variation has been successfully investigated mainly using the GWA case-control study design, rare genetic variation has been studied using a variety of approaches, including case-control-, trio- and family-study designs. One approach that has been shown to be beneficial for rare variant discovery for complex traits is the use of genetically isolated populations A brief background on these populations will follow, as we employ their use in two chapters of this work.

The use of population isolates in psychiatric genomics research

The use of genetically isolated populations for disease gene discovery is advantageous for several reasons. First, there is a presumed reduced heterogeneity for causative alleles due to the bottleneck effect. Such an effect takes place when a small number of founding individuals establish a population that expands without genetic admixture from the outside, so there will be a large population with a relatively homogenous gene pool and relatively few disease-causing alleles (45). Second, some alleles may be present in relatively high frequencies due to genetic drift, a process by which allele frequencies within a finite population (such as a population isolate) shift as a result of chance events. Genetic drift may cause slightly deleterious alleles to be brought into fixation in a finite population, just as it may cause beneficial alleles to be lost. Third, genetic isolates often harbor higher frequencies of recessive disorders as a result of inbreeding (45).

Examples of successful gene discovery efforts in genetic isolates include the locus for Huntington's disease in a Venezuelan population (9), Glutaric aciduria and Ellis–van Creveld syndrome in the Amish (46) and Retinoschisis (47) in the Finnish population. In more recent years, the GWAS-approach was successfully applied to identify associations in several genetic isolates, such as for asthma in the Hutterites (48), as well as intracranial aneurysms (49) and bipolar disorder and schizophrenia in Finland (50).

Following up genomic discoveries with functional studies to elucidate disease mechanisms

Understanding the heritability and genetic architecture of psychiatric disorders is only the first step in understanding biology and working towards treatments that are designed with a mechanistic rationale. Even though the contribution of common variation exceeds the contribution of rare variation on a population-wide level for conditions such as SCZ and ASD (29, 51), thus far rare variation has led to more insight in the functional consequences of gene disruption on the brain. Rodent models with targeted modifications in disease-associated genes have facilitated the elucidation of disease phenotypes and mechanisms for both SCZ by modeling the 22q11.2 microdeletion (52), and ASD by knocking out disease-associated genes such as *CNTNAP2* (53) and *SHANK3* (54). For common genetic variation, it has been more challenging to make sense of the biological meaning of disease-associated SNPs on brain biology, although recently progress has been made in this field as well (55). For this reason, the studies included in this work will focus on rare variants of large effect on psychiatric disease risk.

Functional studies on psychiatric disorders using genetic animal models

Much has been learned from rodent models with modifications in disease-associated genes. For ASD, gene knock-outs include *NRXN1A* (56), *TSC1* (57), *SCN1A* (36) and *CNTNAP2 (53)*, all of which are causal to syndromic or non-syndromic ASD in humans (58-61). These rodent models all have partially overlapping behavioral phenotypes reminiscent of human ASD-related behavior, including impaired social interactions and vocal communication, repetitive grooming and seizures. Commonalities in these models on the molecular, cellular and circuit levels include neuronal morphology deficits on the level of the spines, dendrites and axons, as well as altered cellular signaling (PI3K, mTOR) and excitiation/inhibition imbalance (62). Rodent models for ASD have been useful as disease modeling organisms because of their relatively high degree of genetic homology to humans, their practical advantages as experimental animals, the availability of phenotypic batteries for core behavioral phenotypes, and mostly, the possibility for experimental inquiry at multiple levels of organization, from molecule to circuit to behavior.

For SCZ, the number of valid and therefore useful genetic rodent models has been significantly lower than for ASD (63). One reason for this is that it has been taking longer to uncover *high-confidence* risk genes/CNVs for SCZ. Therefore, the validity of certain rodent models, such as for example the *DISC1* KO (64) is questionable as *DISC1's* true contribution to SCZ-risk is uncertain (65). The two genetic risk factors that confer the highest risk for SCZ –the 22q11.2 microdeletion and *SETD1A* (66), have therefore wielded genetic rodent models that had a higher chance for face- and construct validity *a priori.* Behavioral phenotypes for these models include deficits in working memory and pre-pulse inhibition, hypothesized to be reminiscent of cognitive and negative schizophrenia symptoms.

Modeling psychiatric disorders in rodents comes with obvious shortcomings. The first and foremost being the complex definition of mental disorders. Psychiatric disorders typically affect higher order brain processes including cognition, emotion and perception and rely completely on the reporting of subjective experiences reported by the patient. Despite high hopes and expectations, the revolutions in both the neuroscience- and genomics fields have not yielded objectively measurable biomarkers with sufficiently high positive- and negative predictive values to aid a reliable diagnosis for the majority of common psychiatric disorders. An exception to this are the aforementioned rare, syndromic forms of both ASD and SCZ, for which genotyping can now confirm a diagnosis at the molecular level. Nonetheless, animal models incapable of human language will not be able to report symptoms involving higher order thought processes and therefore critical questions will remain open regarding the face- and predictive validity of these models, as we can only rely on our own observations of animal behavior and speculate on the level of analogy between ours and theirs. Another shortcoming of using rodent models to study psychiatric illness are the biological differences between rodents and humans. Although a reasonable level of orthology exists at the level of the genome nucleotide sequence (67), and many aspects of brain development appear to be conserved across different mammalian species (68), mice and humans are still separated by an estimated 85 million years of evolution (69). Species-specific differences in neurobiology appear to be especially apparent in the cortex. For example, the human neocortex has an approximate 1000-fold increase in both neuronal numbers and surface area compared to mice (70). At the tissue-organizational level, superficial layer neurons appear to be expanded in numbers in humans (71), as is the case for outer radial glial cells (oRGs) in the outer radial glial layer, that are only very scarcely present in mice (72-74). Recently, single-nucleus RNA-sequencing corroborated some of these and revealed other differences at the single-cell transcriptional level (75).

The importance of the shortcomings of rodent models becomes especially apparent when looking at the number of promising preclinical findings that translate successfully to the clinic. For brain disorders, the FDA-approval rate for new drugs has remained disappointingly low over the last several decades, with virtually no genuinely novel, revolutionary medications being approved. Regulatory agency approvals have almost entirely been based on minor chemical modifications of pre-existing drugs, sometimes referred to as 'me too' drugs. This has been quite different in other fields in medicine, such as oncology, in which first-in-class medications with dramatically improved efficacy and side effect profiles are being approved regularly by the FDA and EMA (75). These shortcomings underscore the need for a psychiatric disease modeling system with higher chances of successful translation from the lab to human patients.

The use of induced pluripotent stem cells to model psychiatric disorders in a dish

The generation of induced pluripotent stem cells (iPSCs) from somatic cells (76) allowed researchers access to human stem cells without the need for embryonic tissue. Embryonic stem cells (ESCs) had been initially deployed successfully for the generation of neuronal tissue such as retinal- and pituitary cells by pioneer Yosiki Sasai and colleagues in 2012 (77). The subsequent development of protocols to generating brain organoids (78, 79), combined with the use of the induced stem cells derived from patients with well-defined brain disorders, paved the way to study human-specific developmental aspects of these disorders that were not possible before. The addition of CRISPR-gene editing technology to the toolkit of researchers made the modeling system even more promising (80).

The first successes in brain disorder disease modeling were for microcephaly (79), Zika virus infection-related brain pathology (81), ASD (82) and Miller-Dieke Syndrome (83). These studies revealed disease-related phenotypes and provided insight into human-specific disease mechanisms and therefore hold promise for the investigation of treatment strategies or drug screening in the future. The greatest insights that have been revealed using brain organoid modeling are early developmental cell-biological processes including neural progenitor proliferation, differentiation, cellular signaling and neurogenesis. At this point, very little is known about properties related to neuronal activity in brain organoids, with only a few studies with early data published to date (81, 84). Since these early successes, the brain organoid disease-modeling field has been slowed down by considerable shortcomings of the system as well.

Even though brain organoids recapitulate the cellular diversity and micro-architectural features of the early embryonic human brain with an impressive accuracy, as has been shown using immunostaining (78) as well as (sc)RNA sequencing (85, 86),; anterior-posterior, dorso- ventral, and medio-lateral patterning processes appears highly unorganized. The result is that brain organoid region-anatomy appears random from sample to sample (87), making it difficult to conduct studies on developmental aspects that require accurate anatomical (and the related functional) features. This caveat can be overcome to some extent by choosing the appropriate organoid generation method. The amount of tissue heterogeneity appears to be related to the amount of external patterning that is applied experimentally by adding small molecule inhibitors to modulate the relevant signaling pathways (86). Therefore, whole-brain organoids that involve only self-patterning and produce most brain regions display the highest level of heterogeneity (79), whereas organoids receiving external patterning cues show lower levels of heterogeneity, and thus are more highly reproducible (86). Making a well-reasoned decision on which organoid generation method to use is therefore of crucial importance to answer relevant questions for a disease-modeling project. For example, dorsalizing forebrain organoids can be achieved through inhibiton of Wnt-signaling (78), which is also a pathway that is known to

play an important role in the etiology of ASD (88). If critical disease processes are predicted to already occur during the time-window of external patterning (based on for example knowledge on gene expression levels of a disease gene of interest), the patterning itself may interfere with – or mask, relevant phenotypes. Recently, there has been progress in increasing reproducibility of forebrain organoids by introducing an organizational center that generates a Sonic Hedgehoc gradient that orders self-organization (87). These fundamental studies on the brain organoid system itself are now of critical importance to bring the field of disease-modeling ahead.

Choosing the right model system to study disease

It should be clear that none of the described modeling systems will be capable of faithfully reproducing the biology of a human being, interacting with its external world and reporting its experiences. Just like animals, lab-grown brain organoids are incapable of reporting their experiences, and similar questions regarding the face validity of the system remain. In addition, brain organoids at this point do not have multisensory input, nor do they have naturalistic interactions with other organ systems such as immune-, gastrointestinal- and cardiovascular system. Furthermore, another important issue –and related to the face validity of the model, is the age-of-onset for different psychiatric disorders. As brain organoids have thus far been shown to resemble most closely the embryonic human brain (85, 86), it remains challenging to determine the relevance of uncovered embryonic disease processes for an early adulthood-onset disorder like for example SCZ. Therefore, choosing the appropriate model wisely to obtain relevant answers to scientific questions is critical, and combining different model systems for different levels of investigation to create a picture as complete as possible will be necessary to achieve a comprehensive understanding of brain function and disease.

Here, I will briefly introduce the three chapters included in this work, that span gene variant discovery, disease modeling using stem cell-derived brain organoids, and the initial steps of setting up a drug-screening platform utilizing a stem cell-derived neuronal culture. I will briefly outline the scientific rationale for each guided by the background described above.

Chapter 2

The aim of this study is to identify a rare genetic risk factor of large effect size for MDD. As GWA-studies have thus far been only able to identify common SNPs associated with depression that confer a small risk to illness, the identification of a variant of larger effect could facilitate the identification of relevant and actionable disease processes through subsequent disease modeling. In order to achieve this, we performed extensive psychiatric phenotyping and exome sequencing in an extended pedigree from the Old Order Mennonite population.

Chapter 3

The aim of this study is to develop a human-specific modeling system to investigate neurodevelopmental pathophysiological aspects of ASD, with the ultimate goal of identifying a targetable disease mechanism with potential for translation to the clinic. To this end, we generated forebrain organoids grown from induced pluripotent stem cells from patients carrying a homozygous protein-truncating mutation in *CNTNAP2,* that is responsible for a rare, syndromal form of ASD with comorbid intellectual impairment and epilepsy.

Chapter 4

The aim of this study is to develop a drug-screening platform using ESC-derived neuronal cultures from *Df(16)A*$^{+/-}$ mice carrying a genomic deletion syntenic to the 22q11.2 microdeletion in humans. *Df(16)A*$^{+/-}$ mice have served as a valuable mouse model for 22q11.2- associated SCZ, as the mice display behavioral and neurophysiological phenotypes that are reminiscent of SCZ phenotypes in human patients. We observed robust alterations in dendritic arborization as a neuronal cellular phenotype, which we intend to use as a readout for high-throughput drug screening in the future.

REFERENCES

1. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1211-59.
2. Ban TA. Pharmacotherapy of mental illness--a historical analysis. Prog Neuropsychopharmacol Biol Psychiatry. 2001;25(4):709-27.
3. Hyman SE. Revolution stalled. Sci Transl Med. 2012;4(155):155cm11.
4. Hirschfeld RM. History and evolution of the monoamine hypothesis of depression. J Clin Psychiatry. 2000;61 Suppl 6:4-6.
5. Kendler KS. The Dopamine Hypothesis of Schizophrenia: An Historical and Philosophical Analysis. Philosophy, Psychiatry, & Psychology. 2011;Volume 18,:pp. 41-63
6. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015;47(7):702-9.
7. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet. 2008;9(4):255-66.
8. Sullivan PF, Geschwind DH. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. Cell. 2019;177(1):162-83.
9. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature. 1983;306(5940):234-8.
10. Rebeck GW, Reiter JS, Strickland DK, Hyman BT. Apolipoprotein E in sporadic Alzheimer's disease: allelic variation and receptor interactions. Neuron. 1993;11(4):575-80.
11. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
12. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010;42(1):30-5.
13. Posey JE, O'Donnell-Luria AH, Chong JX, Harel T, Jhangiani SN, Coban Akdemir ZH, et al. Insights into genetics, human biology and disease gleaned from family based genomic studies. Genet Med. 2019;21(4):798-812.
14. Kendler KS. What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. Mol Psychiatry. 2013;18(10):1058-66.
15. Flint J, Kendler KS. The Genetics of Major Depression. Neuron. 2014;81(5):1214.
16. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668-81.
17. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016;48(9):1031-6.
18. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry. 2013;18(4):497-511.
19. consortium C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 2015;523(7562):588-91.
20. Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50(3):381-9.

21. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet. 2009;373(9659):234-9.
22. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 2003;60(12):1187-92.
23. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. PLoS Med. 2005;2(5):e141.
24. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. 2019;22(3):343-52.
25. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000;157(10):1552-62.
26. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA. 2003;289(23):3095-105.
27. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431-44.
28. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748-52.
29. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506(7487):185-90.
30. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421-7.
31. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009;19(3):212-9.
32. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. Genetics. 2000;156(1):297-304.
33. McClellan J, King MC. Genetic heterogeneity in human disease. Cell. 2010;141(2):210-7.
34. Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, Borrow J, et al. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc Natl Acad Sci U S A. 1995;92(17):7612-6.
35. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science. 2007;316(5823):445-9.
36. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. 2012;74(2):285-99.
37. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nat Neurosci. 2019;22(12):1961-5.
38. Sacco R, Gabriele S, Persico AM. Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. Psychiatry Res. 2015;234(2):239-51.
39. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci. 2016;19(4):571-7.
40. Steinberg S, Gudmundsdottir S, Sveinbjornsson G, Suvisaari J, Paunio T, Torniainen-Holm M, et al. Truncating mutations in RBM12 are associated with psychosis. Nat Genet. 2017;49(8):1251-4.

41. Tarjinder Singh BMN, Mark J. Daly. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. medRxiv 2020091820192815;.
42. Amin N, de Vrij FMS, Baghdadi M, Brouwer RWW, van Rooij JGJ, Jovanova O, et al. A rare missense variant in RCL1 segregates with depression in extended families. Mol Psychiatry. 2018;23(5):1120-6.
43. Ruzzo EK, Perez-Cano L, Jung JY, Wang LK, Kashef-Haghighi D, Hartl C, et al. Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. Cell. 2019;178(4):850-66 e26.
44. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015;87(6):1215-33.
45. Sheffield VC, Stone EM, Carmi R. Use of isolated inbred human populations for identification of disease genes. Trends Genet. 1998;14(10):391-6.
46. Biery BJ, Stein DE, Morton DH, Goodman SI. Gene structure and mutations of glutaryl-coenzyme A dehydrogenase: impaired association of enzyme subunits that is due to an A421V substitution causes glutaric acidemia type I in the Amish. Am J Hum Genet. 1996;59(5):1006-11.
47. Sauer CG, Gehrig A, Warneke-Wittstock R, Marquardt A, Ewing CC, Gibson A, et al. Positional cloning of the gene associated with X-linked juvenile retinoschisis. Nat Genet. 1997;17(2):164-70.
48. Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, et al. Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. N Engl J Med. 2008;358(16):1682-91.
49. Kurki MI, Gaal EI, Kettunen J, Lappalainen T, Menelaou A, Anttila V, et al. High risk population isolate reveals low frequency variants predisposing to intracranial aneurysms. PLoS Genet. 2014;10(1):e1004134.
50. Palo OM, Antila M, Silander K, Hennah W, Kilpinen H, Soronen P, et al. Association of distinct allelic haplotypes of DISC1 with psychotic and bipolar spectrum disorders and with underlying cognitive impairments. Hum Mol Genet. 2007;16(20):2517-28.
51. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46(8):881-5.
52. Stark KL, Xu B, Bagchi A, Lai WS, Liu H, Hsu R, et al. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nat Genet. 2008;40(6):751-60.
53. Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell. 2011;147(1):235-46.
54. Bozdagi O, Sakurai T, Papapetrou D, Wang X, Dickstein DL, Takahashi N, et al. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. Mol Autism. 2010;1(1):15.
55. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177-86.
56. Etherton MR, Blaiss CA, Powell CM, Sudhof TC. Mouse neurexin-1alpha deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. Proc Natl Acad Sci U S A. 2009;106(42):17998-8003.
57. Goorden SM, van Woerden GM, van der Weerd L, Cheadle JP, Elgersma Y. Cognitive deficits in Tsc1+/- mice in the absence of cerebral lesions and seizures. Ann Neurol. 2007;62(6):648-55.
58. Bolton PF. Neuroepileptic correlates of autistic symptomatology in tuberous sclerosis. Ment Retard Dev Disabil Res Rev. 2004;10(2):126-31.
59. Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, et al. 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. 2009;85(5):655-66.

60. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, et al. Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. N Engl J Med. 2006;354(13):1370-7.
61. Weiss LA, Escayg A, Kearney JA, Trudeau M, MacDonald BT, Mori M, et al. Sodium channels SCN1A, SCN2A and SCN3A in familial autism. Mol Psychiatry. 2003;8(2):186-94.
62. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nat Med. 2016;22(4):345-61.
63. Nestler EJ, Hyman SE. Animal models of neuropsychiatric disorders. Nat Neurosci. 2010;13(10):1161-9.
64. Jaaro-Peled H. Gene models of schizophrenia: DISC1 mouse models. Prog Brain Res. 2009;179:75-86.
65. Chubb JE, Bradshaw NJ, Soares DC, Porteous DJ, Millar JK. The DISC locus in psychiatric illness. Mol Psychiatry. 2008;13(1):36-64.
66. Takata A, Xu B, Ionita-Laza I, Roos JL, Gogos JA, Karayiorgou M. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. Neuron. 2014;82(4):773-80.
67. Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420(6915):520-62.
68. Defelipe J. The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. Front Neuroanat. 2011;5:29.
69. Springer MS, Murphy WJ. Mammalian evolution and biomedicine: new views from phylogeny. Biol Rev Camb Philos Soc. 2007;82(3):375-92.
70. Herculano-Houzel S, Mota B, Lent R. Cellular scaling rules for rodent brains. Proc Natl Acad Sci U S A. 2006;103(32):12138-43.
71. Hill RS, Walsh CA. Molecular insights into human brain evolution. Nature. 2005;437(7055):64-7.
72. Wang X, Tsai JW, LaMonica B, Kriegstein AR. A new subtype of progenitor cell in the mouse embryonic neocortex. Nat Neurosci. 2011;14(5):555-61.
73. Shitamukai A, Konno D, Matsuzaki F. Oblique radial glial divisions in the developing mouse neocortex induce self-renewing progenitors outside the germinal zone that resemble primate outer subventricular zone progenitors. J Neurosci. 2011;31(10):3683-95.
74. Hansen DV, Lui JH, Parker PR, Kriegstein AR. Neurogenic radial glia in the outer subventricular zone of human neocortex. Nature. 2010;464(7288):554-61.
75. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. Conserved cell types with divergent features in human versus mouse cortex. Nature. 2019;573(7772):61-8.
76. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76.
77. Eiraku M, Sasai Y. Self-formation of layered neural structures in three-dimensional culture of ES cells. Curr Opin Neurobiol. 2012;22(5):768-77.
78. Kadoshima T, Sakaguchi H, Nakano T, Soen M, Ando S, Eiraku M, et al. 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. 2013;110(50):20284-9.
79. Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, et al. Cerebral organoids model human brain development and microcephaly. Nature. 2013;501(7467):373-9.
80. Hockemeyer D, Jaenisch R. Induced Pluripotent Stem Cells Meet Genome Editing. Cell Stem Cell. 2016;18(5):573-86.
81. Quadrato G, Nguyen T, Macosko EZ, Sherwood JL, Min Yang S, Berger DR, et al. Cell diversity and network dynamics in photosensitive human brain organoids. Nature. 2017;545(7652):48-53.

82. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L, et al. FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. Cell. 2015;162(2):375-90.
83. Bershteyn M, Nowakowski TJ, Pollen AA, Di Lullo E, Nene A, Wynshaw-Boris A, et al. Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia. Cell Stem Cell. 2017;20(4):435-49 e4.
84. Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, et al. Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. Cell Stem Cell. 2019;25(4):558-69 e7.
85. Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Brauninger M, et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. Proc Natl Acad Sci U S A. 2015;112(51):15672-7.
86. Velasco S, Kedaigle AJ, Simmons SK, Nash A, Rocha M, Quadrato G, et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. Nature. 2019.
87. Cederquist GY, Asciolla JJ, Tchieu J, Walsh RM, Cornacchia D, Resh MD, et al. Specification of positional identity in forebrain organoids. Nat Biotechnol. 2019;37(4):436-44.
88. Bae SM, Hong JY. The Wnt Signaling Pathway and Related Therapeutic Drugs in Autism Spectrum Disorder. Clin Psychopharmacol Neurosci. 2018;16(2):129-35.

Chapter 2

A Rare Missense Variant of the Orphan G Protein-Coupled Receptor Gene *GPR156* is Associated with Major Depressive Disorder in Old Order Mennonites

Bradley R. Miller[1], Claudia Gonzaga-Jauregui[1], Karlla W. Brigatti[1], Job de Jong[2], Robert Breese[2], Susannah Brydges, David Carey, Andrew J. Dwork, Aris N. Economides, Kevin Elmwood, Jean Endicott, Michael B. First, Hilledna J. Gregoire, René Hen, Shane McCarthy, John D. Overton, Evangelos Pefanis, Erik G. Puffenberger, Jose Rojas, Gorazd Rosoklija, Alan R. Shuldiner, Jeffrey Staples, Cristopher Van Hout, Millie Young, Nicole Alessandri-Haber[3], Sander Markx[3*], Kevin A. Strauss[3*]

Manuscript submitted

A Rare Missense Variant of the Orphan G Protein-Coupled Receptor Gene *GPR156* is Associated with Major Depressive Disorder in Old Order Mennonites

1. Co-first authors
2. Co-second authors
3. Co-senior authors
*Corresponding authors

Affiliations:
Department of Psychiatry, Columbia University, New York, NY, 10032 (BRM, JD, HJG, MBF, JE, RH, AJD, GR, SM)
Division of Molecular Therapeutics, New York State Psychiatric Institute, New York, New York, 10032. (JD, SM)
Division of Systems Neuroscience, New York State Psychiatric Institute, New York, NY, (BRM, HJG, RH).
Clinic for Special Children, Strasburg, Pennsylvania, 17579. (EGP, KWB, KAS)
Department of Pediatrics, Penn Medicine-Lancaster General Hospital, Lancaster, Pennsylvania, 17602 (KAS)
Departments of Pediatrics and Molecular, Cell & Cancer Biology, University of Massachusetts School of Medicine, Worcester, Massachusetts, 01655 (KAS)
Regeneron Genetics Center, Tarrytown, New York, 10599. (CG-J, RB, SB, ANE, SM, JDO, ELP, JR, ARS, JS, CVH, NA-H)
Regeneron Pharmaceuticals Inc. Tarrytown, New York, 10599. (CG-J, RB, SB, ANE, SM, JDO, ELP, JR, ARS, JS, CVH, NA-H)
Geisinger, Danville, Pennsylvania 17822. (DC, KE)

ABSTRACT

Major depressive disorder (MDD) is the leading cause of disability worldwide. Risk for MDD is heritable, and the genetic structure of 'founder' populations enables investigation of rare susceptibility alleles with large effect. In an extended Old Order Mennonite pedigree (n=78), we identified a rare missense variant of *GPR156* (c.1599G>T, p.Glu533Asp) that increases the relative risk of MDD nearly three-fold. GPR156 expression is localized to the medial habenula, a region implicated in mood regulation, and insertion of the c.1599G>T change into the murine *Gpr156* locus induces a form of maladaptive behavior under stress that is rescued by antidepressants. This work, spanning humans and mice, reveals a human variant with a large effect size in depression, points to a novel receptor, and introduces a new animal model for the study of depression pathophysiology and treatment.

Keywords: major depressive disorder, depression, GPR156, G-protein coupled receptor, medial habenula

INTRODUCTION

Major depressive disorder (MDD) is the leading cause of disability worldwide (1), with an overall 12-month prevalence of 6% and lifetime risk of 15–18% (2, 3). It commonly strikes in early adulthood (4, 5); the median onset of MDD is 25 years and nearly 40% of those affected experience a first depressive episode before age 20 (6, 7). Major depression is characterized by subjective experiences of sadness and anhedonia, often colored by thoughts of worthlessness, guilt, and suicide (8). Subjective despair is at the core of MDD, and it often coincides with objective disturbances of sleep, appetite, attention, decision making, and motor activity (8, 9).

Major depression tends to be episodic but is best considered a lifelong burden, with phases of mental and physical debilitation that intrude upon periods of relative psychological stability (8). Almost 80% of individuals who experience a major depressive episode suffer at least one recurrence and, over the long term, only 17–30% fully recover (10, 11). Lifestyle modification, psychotherapy, and medication are the mainstays of treatment (9, 12). However, lifestyle changes are hard to implement, trained psychotherapists are in short supply (13, 14), and commonly prescribed antidepressants often have limited efficacy or are poorly tolerated (15, 16). Thus, there is a pressing need for more effective antidepressant therapies. The most commonly prescribed antidepressant therapies are built on serendipitous discoveries several decades ago that targeti the monoamine system and can alleviate depression (17). Ideally, the next generation of treatments would target the underlying pathophysiology of depression. To develop these, we need a better understanding of the biological basis of depression (4, 18).

Genetic research can elucidate fundamental mechanisms of MDD and reveal new therapeutic targets (3, 4). Twin and adoption data indicate that heritable susceptibility to MDD is between 29% (men) and 42% (women) (19). Genome-wide association studies (GWAS) identified dozens of common polymorphisms associated with depressive illness (20, 21). These GWAS are built on the common disease – common variant model of depression and they identified numerous common variants associated with depression with small effect sizes. Co-inheritance of multiple common DNA variants, each of small effect, may have additive or synergistic effects on disease susceptibility (22, 23).

We hypothesized that rare variants with large effect sizes also contribute to the genetic risk of major depression (common disease – rare variant model). This is in line with other common diseases, including psychiatric illnesses such as autism and schizophrenia, that have genetic architectures spanning common variants with small effects and rare variants with large effects (24). Rare variants with large effect sizes may provide great insight into the pathophysiology of depression.

Here, as an alternative to GWAS, we employed a family-based approach to search for rare, penetrant alleles that might underlie a high prevalence of depressive illness in an extended Old Order Mennonite family. Contemporary Mennonites represent a classic founder population; they descended from just a few hundred Swiss-German Anabaptists who migrated to North America during the 18th century and today face a distinctive constellation of heritable disease risks shaped by generations of reproductive isolation and genetic drift (25-28).

In an extended three-generation family spread across three states, we identified a rare variant of *GPR156* c.1599G>T (p.Glu533Asp) that increases the risk for MDD approximately three-fold. Expression of GPR156 localizes to the habenula and other brain regions linked to stress-related behavior (29, 30), and insertion of the human c.1599G>T variant into the murine *Gpr156* locus induces a maladaptive form of behavior under adverse conditions that can be partially rescued by antidepressant therapy (31, 32). To our knowledge, *GPR156* c.1599G>T is the strongest risk allele for MDD reported to date. This finding challenges basic assumptions about the genetic landscape of MDD (22, 33), reinforces theories about the habenula's role in emotional behavior (34, 35), and opens new territory for therapeutic development (18).

RESULTS

A rare GPR156 c.1599G>T variant confers strong relative risk for major depressive disorder in an extended Mennonite pedigree

A 40-year-old Mennonite woman presented to our community health center with a long-standing history of recurrent, severe depression beginning at age 17 years. She reported a high prevalence of depressive symptoms across three generations of her family (**Fig. 1**). We collected psychiatric phenotype information from her and 14 first- and second-degree relatives using the Structured Clinical Interview for DSM-5 (SCID-5) (36) corroborated by information from household members, clinical providers, and available medical records. These data were reviewed by a team of experienced research psychiatrists to arrive at a final consensus diagnosis for each individual (see **Online Methods**)(37). All study participants consented to genomic research and provided DNA samples for exome sequencing, as previously described (27). Laboratory scientists and field psychiatrists collected data independently, and genotype data were unblinded only after all DSM-5 diagnoses were 'locked in' by the diagnostic team.

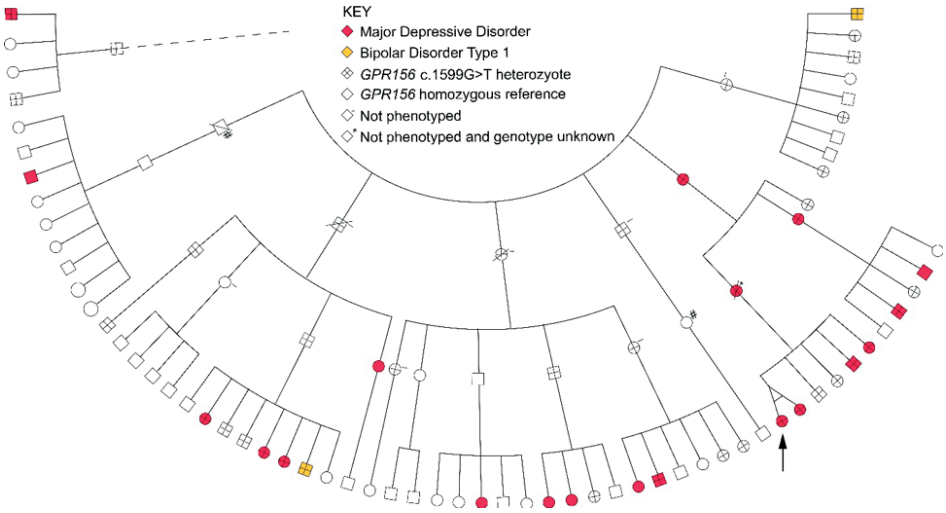


Figure 1. Extended Old Order Mennonite pedigree segregating the rare *GPR156* c.1599G>T (p.Glu533Asp) variant.

Subjects with major depressive disorder (MDD) and bipolar disorder type 1 (BP1) are shown with red and yellow shading, respectively. The proband is indicated with a vertical arrow. Cross-hatching denotes a *GPR156* c.1599G>T heterozygous genotype. Genotyping and psychiatric phenotyping were conducted in a parallel, mutually blinded fashion. All included subjects belong to a sibship in which each member has a 50% chance of inheriting the *GPR156* c.1599G>T allele.

A missense variant of *GPR156* (c.1599G>T, p.Glu533Asp; minor allele frequency, MAF<0.02; hg38.chr3:g.120167878[C>A]) was the only rare allele that segregated with affective illness in the proband and 14 family members. *GPR156* encodes a seven-transmembrane protein without a known ligand and belongs to a large class of orphan G protein-coupled receptors expressed in brain tissue (https://www.proteinatlas.org) (38). The glutamate to aspartate change in amino acid residue 533 of GPR156 is within the intracellular C-terminal tail of the protein and predicted *in silico* to be damaging (**Supplementary Fig. 1**).

To further investigate the link between *GPR156* c.1599G>T and mood disorders, we expanded the pedigree to encompass 78 individuals related in sibships across three generations from settlements in Pennsylvania, New York, and Ohio (**Fig. 1**). Within this larger cohort, *GPR156* c.1599G>T heterozygotes had nearly three-fold higher relative risk (RR) for MDD or bipolar disorder type 1(RR 2.9, CI 1.4-6.3, p=0.0041) (**Table 1, Fig. 2A**). Of the 15 *GPR156* c.1599G>T heterozygotes with a major affective disorder, 13 had MDD and two had bipolar disorder type 1; exclusion of these latter two individuals did not appreciably alter the results: *GPR156* c.1599G>T remained a significant risk for MDD (RR 2.7, CI 1.3-5.9, p=0.011).

Table 1. Statistical Associations Between *GPR156* Genotype and Psychiatric Phenotype (n=78)

Psychiatric Phenotype	Relative Risk	CI	Fisher's P	Percent Affected (%)	
				M/+	+/+
Major depressive disorder *or* Bipolar disorder 1	2.9	1.4-6.3	*0.0041*	45.5	15.6
Major depressive disorder	2.7	1.3-5.9	*0.011*	41.9	15.6
Transdiagnostic (any mood or anxiety disorder)	1.7	1.0-2.9	*0.029*	57.6	33.3
Anxiety disorders	1.4	0.60-3.3	0.31	25.0	17.8

Abbreviations: CI, 95% confidence interval; M/+, GPR156 c.1599G>T heterozygous; +/+, GPR156 reference allele

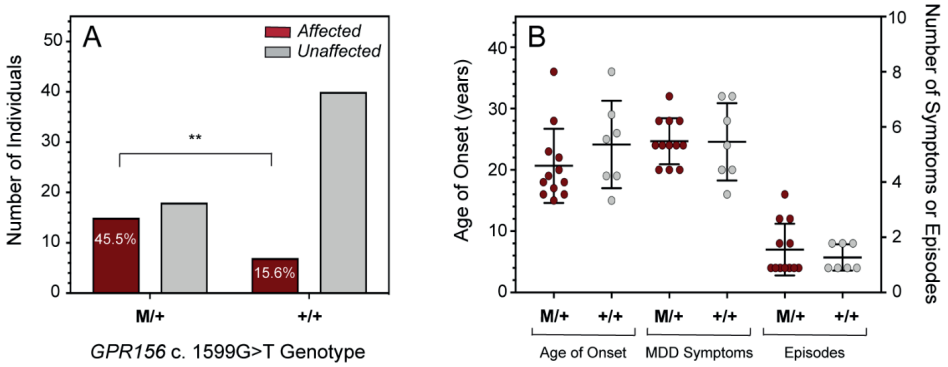


Figure 2. Association of *GPR156* c.1599G>T with affective illness.
(A) *GPR156* c.1599G>T heterozygotes (M/+) had a 2.9-fold (CI 1.4-6.3; p=0.0041) higher relative risk of being affected (red bars) with major mood disorders as compared to subjects homozygous for the *GPR156* reference allele (+/+). (B) The age of onset (left axis), number of depressive symptoms (right axis), and frequency of episodes (right axis) did not differ between individuals with MDD who harbored the *GPR156* c.1599G>T variant (red circles, n=13) and affected individuals who did not (gray diamonds, n=7).

Using the comprehensive SCID-5 assessment, we determined whether the GPR156 variant is associated with a transdiagnostic range of mood and anxiety disorders, or whether it is more restricted. Notably, common variants associated with depression in GWAS have a substantial overlap with those associated with anxiety disorders (39). When we evaluated a transdiagnostic range encompassing MDD, bipolar disorder, the full range of more mild mood disorders, and anxiety disorders, the association of *GPR156* c.1599G>T with mental illness was substantially weakened (**Table 1).** Furthermore, we found no association between *GPR156* c.1599G>T and anxiety disorders, suggesting specificity between *GPR156* genotype and major affective illness (**Table 1**). There was nothing distinctive about the expression of MDD in *GPR156* c.1599G>T heterozygotes: the onset, character, severity, duration, response rate, and recurrence of depressive symptoms were similar among those who harbored the *GPR156* c.1599G>T variant as compared to affected individuals who did not (**Fig. 2B**).

Increased prevalence of depression among *GPR156* c.1599G>T heterozygotes within a large unascertained clinical cohort

After demonstrating the clinical significance of*GPR156* c.1599G>T in a structured family study, we set out to determine its association with MDD in a more naturalistic setting. *GPR156* c.1599G>T (rs774090358) was exceedingly rare in outbred population databases, with one heterozygote each in gnomAD (n=250,562, MAF=0.000003) and TopMed (n=125,568, MAF=0.000008) and no homozygotes reported to date. However, we identified nine *GPR156* c.1599G>T heterozygotes among 144,204 individuals in the Geisinger-Regeneron DiscovEHR collaboration cohort (MAF=0.0000624)(40), which links genomic sequence data to deidentified electronic health record (EHR) information as part of a regional health initiative. Geisinger serves rural communities of Pennsylvania, including Mennonites; we thus expected to find *GPR156* c.1599G>T enriched in the DiscovEHR database. Deidentified clinical records were available for six of the DiscovEHR subjects.

Based on genomic data, the *GPR156* c.1599G>T heterozygotes in DiscovEHR did not overlap or have closer than a third-degree relation with any member of our founder pedigree, but all shared the same haplotype around *GPR156* on chromosome 3 suggesting shared ancestry. Remarkably, all six subjects (100%) with medical records had depressive symptoms and were prescribed at least one antidepressant medication (**Supplementary Table 1**). The background prevalence of depressive illness in the aggregate Geisinger clinical population (n=2,060,968) and DiscovEHR cohort are 22.8% and 25.7%, respectively, indicating that *GPR156* c.1599G>T confers an approximately four-fold relative risk for depressive illness in this sample (RR=3.9, CI 2.4–15.7, p=0.0003). This demonstrates that the GPR156 c.1599G>T (rs774090358) variant is associated with depressive symptoms that reach the threshold for antidepressant treatment in a community setting.

GPR156 is highly expressed in the medial habenula

To better understand the potential role of GPR156 in the pathophysiology of MDD, we generated three mouse models: a constitutive knock-out with an exon 3–5 deletion of murine *Gpr156* (*Gpr156*$^{-/-}$), a humanized model harboring the wild-type (WT) human gene in the murine locus (*GPR156*$^{hum/hum}$), and a humanized knock-in model with the exact variant change (c.1599G>T; p.Glu533Asp) in the human gene (*GPR156*$^{E533D/E533D}$)(**Fig. 3, Supplementary Fig. 2**). Mice from all three lines were viable and fertile and showed no growth derangements or anatomical deformities. *Gpr156*$^{-/-}$, *GPR156*$^{hum/hum}$ and *GPR156*$^{E533D/E533D}$ mice behaved similar to genetically unaltered WT mice on a battery of tests for motor coordination, balance, activity, and locomotion (**Supplementary Table 2**).

Gpr156$^{-/-}$ mice had an in-frame LacZ expression cassette in place of deleted exons 3–5 (**Fig. 3B**), which allowed us to interrogate expression through reporter LacZ staining. We found LacZ expression throughout the mouse brain, especially within the habenula and hippocampus, two areas implicated in the pathophysiology of depression (**Fig. 4A**) (41-43). Using RNAscope in situ hybridization, we confirmed that endogenous GPR156 is expressed in these areas, and found that it is most highly expressed in the medial habenula (**Fig. 4B-C**). Within the medial habenula, we found that it is expressed in VGLUT-2 expressing neurons, which receive input from the limbic system, project to brainstem nuclei, and mediate behavioral responses to aversive stimuli (**Fig. 4D-E, Supplementary Fig. 2**) (29, 30). *GPR156* mRNA was undetectable by RNAscope in the habenula of *Gpr156-/-* mice (**Supplementary Fig. 2**). We then performed RNAscope for GPR156 in postmortem human brain and found that GPR156 is also expressed in the medial habenula of the human brain (**Supplementary Fig. 2**).

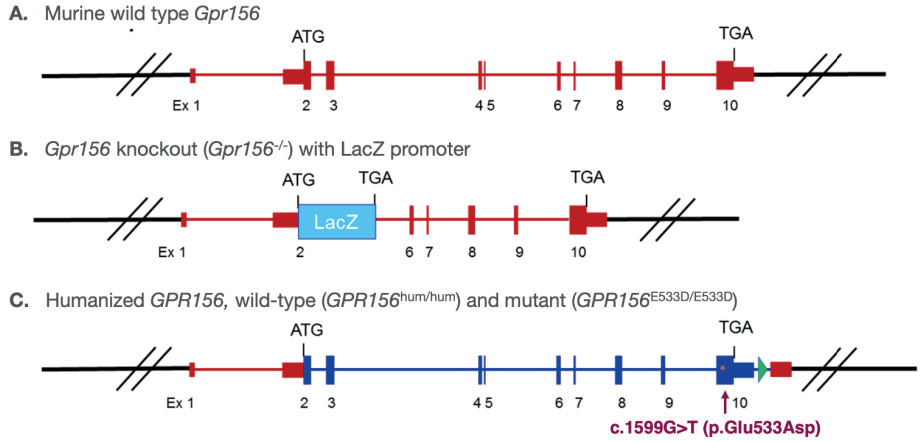


Figure 3. GPR156 genetically modified mouse models.
Schematic of the three genetically modified mouse lines produced for this study. We generated a constitutive knock-out mouse with a deletion of exons 3 through 5 of the endogenous murine *Gpr156* gene (*Gpr156*$^{-/-}$), a mouse with the wild-type human gene inserted in the murine locus (*GPR156*$^{hum/hum}$), and a humanized knock-in (*GPR156*$^{E533D/E533D}$) model with the variant change c.1599G>T (p.Glu533Asp) in the human gene.

GPR156 c.1599G>T (p.Glu533Asp) variant causes maladaptive stress-induced behavior in rodents that is rescued by antidepressant medications

To determine the functional impact of the *GPR156* c.1599G>T variant *in vivo,* we used the Porsolt forced swim test (FST) of perseverance under adverse conditions, originally developed in 1977 to screen for antidepressant efficacy in rodents (31, 44). When forced to swim in a cylinder from which they cannot escape, normal rodents swim vigorously for a period of time but then adopt an immobile posture. The time to immobility can be modified (lengthened) by exposure to certain antidepressants, which may signal their potential to alleviate depressive symptoms in humans (32, 45, 46).

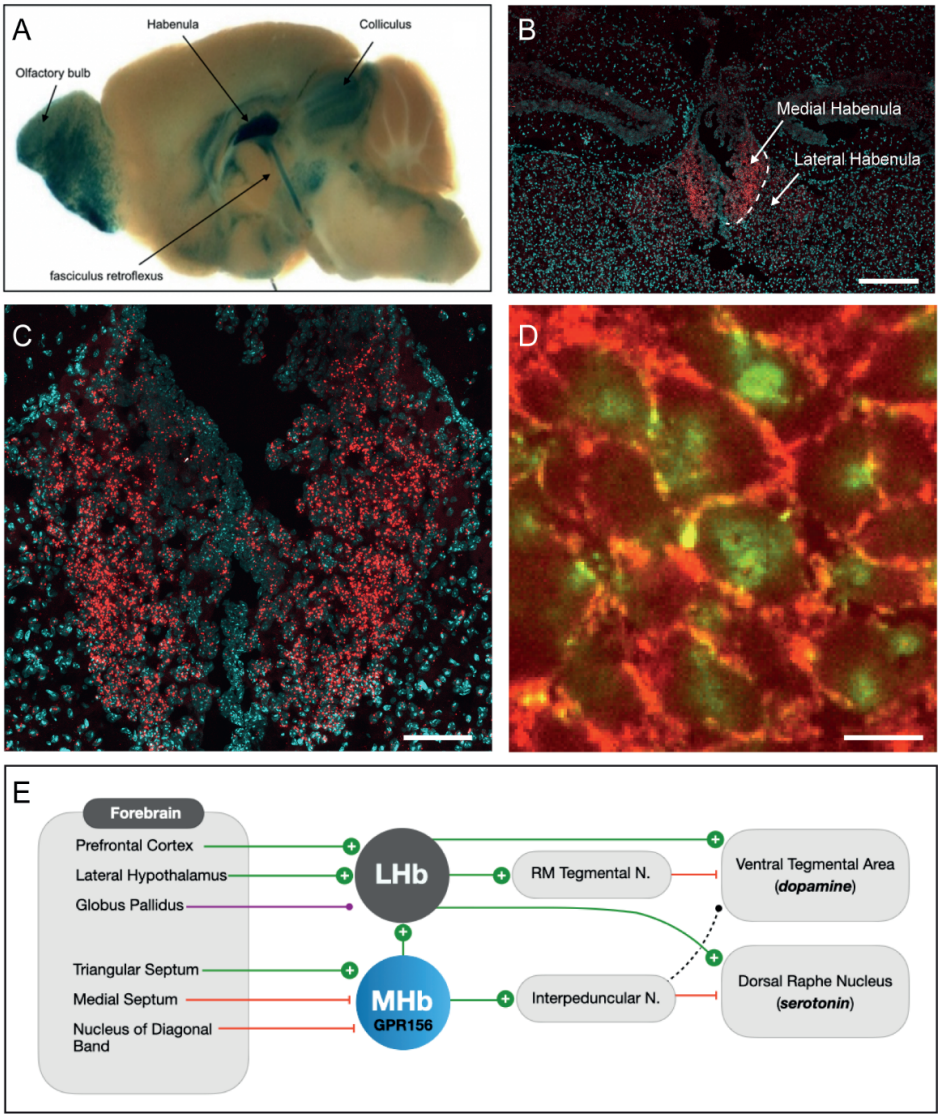


Figure 4. Expression of GPR156 in the medial habenula.

(A) Replacement of the endogenous mouse *Gpr156* locus with a LacZ reporter cassette shows the pattern of GPR156 expression in adult mouse brain. A sagittal brain section reveals highest expression in the habenula as indicated by the intensely dark staining. Areas of less strong expression include the olfactory bulb, fasciculus retroflexus, interpeduncular nucleus, colliculus, and dorsal brainstem nuclei. (B) *In situ* mRNA hybridization (RNAscope) of wildtype mouse brain shows abundant *GPR156* mRNA (red) in the medial habenula; no such expression was detected in *Gpr156*$^{-/-}$ animals (See Fig. S1). Scale bar = 400 µm. (C) Higher magnification image of medial habenula demonstrating high level of GPR156 expression. Scale bar = 100 µm. (D) Within the medial habenula, *GPR156* mRNA (green, perinuclear) localizes VGLUT2 expressing neurons, designated by td-tomato (red, cytoplasmic) expression. Scale bar = 10 µm. (E) The lateral (LHb) and medial (MHb) habenula receive diverse excitatory inputs (green lines), inhibitory inputs (red lines), and combined excitatory and inhibitory inputs (purple lines) from limbic and basal ganglia structures of the forebrain(34, 42, 76-79). They project via the rostromedial (RM) tegmental and interpeduncular nuclei to brainstem nuclei with broad dopaminergic and serotonergic projections. These circuits are believed to play a critical role in mediating adaptive (and maladaptive) responses to rewarding and aversive stimuli. Black dotted line represents unknown neurotransmitter.

The FST exposed a striking difference between *Gpr156*$^{-/-}$ and *GPR156*$^{E533D/E533D}$ mice as compared to WT and *GPR156*$^{hum/hum}$ animals. As expected, unmodified WT (n=59) and ***GPR156***$^{\mathbf{hum/hum}}$ (n=42) mice swam vigorously around the circumference of the reservoir with muzzle above water searching for an escape; 98% of animals from both groups swam for at least six minutes without assistance (Fig. 5A). In contrast, ***Gpr156***$^{\mathbf{-/-}}$(n=49) and ***GPR156***$^{\mathbf{E533D/E533D}}$ (n=80) mice exhibited hypermobility during 20–210 seconds of vigorous swimming before diving erratically beneath the water's surface in a disorganized and maladaptive manner (Fig. 5B, Supplemental Video 1). Once submerged, they had to be rescued from drowning.

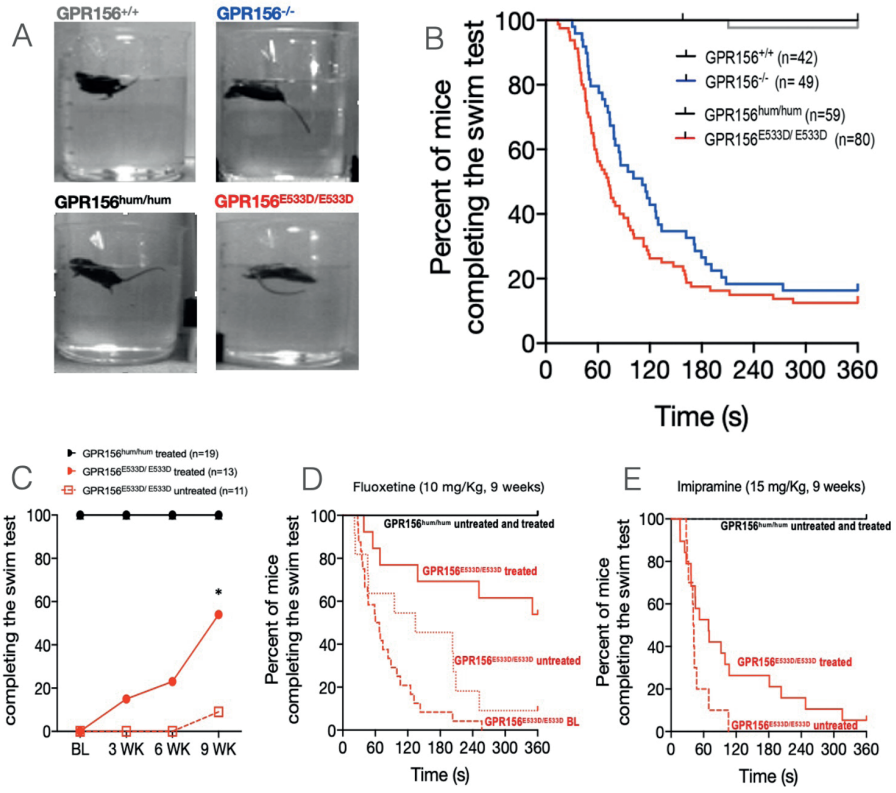


Figure 5. *Gpr156*$^{-/-}$ and *GPR156*$^{E533D/E533D}$ mice exhibit aberrant behavior in the forced swim test (FST) that can be rescued with antidepressant medications.

(A) On the six-minute FST, *Gpr156*$^{+/+}$ and *GPR156*$^{hum/hum}$ mice **swam vigorously around the reservoir with muzzle above water searching for an escape. In contrast, most** *Gpr156*$^{-/-}$ and *GPR156*$^{E533D/E533D}$ mice began to swim in an erratic and disorganized manner within four minutes, diving repeatedly beneath the water's surface until needing to be rescued from drowning. (B) Only 14% and 12.5% of *GPR156*$^{-/-}$ and *GPR156*$^{E533D/E533D}$ mice, respectively, completed the six-minute FST without nearly drowning. In contrast, 98% of *GPR156*$^{+/+}$ and *GPR156*$^{hum/hum}$ mice completed the test unassisted. (C) *GPR156*$^{hum/hum}$ and *GPR156*$^{E533D/E533D}$ mice received intraperitoneal (i.p.) fluoxetine (10 mg/kg•day) six days weekly for nine weeks and were subjected to the FST at weeks 3, 6, and 9 of treatment. After nine weeks of fluoxetine therapy, 53% of *GPR156*$^{E533D/E533D}$ completed the six-minute FST as compared to 9% of the untreated group. The probability of completing the FST for six minutes improved significantly in *GPR156*$^{E533D/E533D}$ mice treated with either fluoxetine (panel D) or imipramine (panel E)(logrank and Gehan-Breslow-Wilcoxon tests, p<0.0001).

Accordingly, we streamlined the FST to a standardized six-minute protocol (47). The impact of *GPR156* modification on the FST was profound; only 14% and 12.5% of ***Gpr156***$^{-/-}$ and ***GPR156***$^{\mathbf{E533D/E533D}}$ mice, respectively, could swim six minutes without nearly drowning (Fig. 5B). To our knowledge, no such phenotype has been previously described.

We next treated mice for nine weeks via daily intraperitoneal injection with one of two different antidepressants, fluoxetine (10 mg/kg•day) or imipramine (15 mg/kg•day), and assessed FST performance at three-week intervals. Subjecting untreated *GPR156*$^{E533D/E533D}$ mice to repetitive testing did not improve their swimming performance. After nine weeks of fluoxetine therapy, 54% of *GPR156*$^{E533D/E533D}$ mice achieved a normal six-minute FST as compared to only 8% of mice in the untreated group (**Fig. 5C and D**; logrank and Gehan-Breslow-Wilcoxon tests, $p<0.0001$). *GPR156*$^{E533D/E533D}$ mice treated with imipramine showed less robust but highly significant improvement in swimming performance (**Fig. 5E**; logrank and Gehan-Breslow-Wilcoxon tests, $p<0.0001$).

DISCUSSION

A family-based strategy using sibships (48) allowed us to discover an association between *GPR156* c.1599G>T and MDD. The strong relative risk conferred by this rare missense variant might have been obscured by a GWAS paradigm comparing much larger numbers of unrelated individuals (21). We confirmed this association in a naturalistic setting of a large healthcare system and established the functional impact of the variant by modeling it in mice.

Our discovery of a rare variant associated with depression with large effect sizes has implications for the genetic architecture of depression. It suggests that the common disease-common variant model of depression, supported by recent GWAS, is incomplete. Most genetic studies of depressive illness rely on a multifactorial and polygenic model of risk, in which MDD arises from combined effects of multiple common DNA variants, each of small effect (49). This common disease-common variant paradigm is popular for framing prevalent disorders with complex etiology (22), and it depends on stratification of each variant's relative 'burden', modeling of additive or synergistic effects, and a premise that all humans carry a number of genetic changes that constitute a continuous measure of disease susceptibility (49, 50). Based on the common disease-common variant model, aggregate GWAS data from more than one million subjects implicate dozens of common gene variants associated with depression (3,4) As expected, each has small effect, and together they account for ~25% of heritable risk (4).

As an alternative approach, we used a family-based strategy within a founder population (25) to identify a rare variant of large effect for a well-defined psychiatric phenotype. Our discovery

of a rare variant in GPR156 that has a large effect size association with depression challenges key assumptions the common disease-common variant model of depression (22, 51). It shows that in addition to common low penetrance variants often detected in GWAS, there are also more highly penetrant rare variants. This is supported by additional recent studies of familial depression (52). While it may be more tractable to discover these rare variants within founder populations (53), within 'outbred' populations highly penetrant alleles unique to individuals or otherwise rare (perhaps arising *de novo*) may also contribute genetic susceptibility to MDD but escape detection by GWAS (54). This hybrid architecture, comprising common variants with small effect sizes and rare variants with large effects sizes, is found in a wide variety of diseases, including cancer, cardiovascular disease, Alzheimer's disease, and schizophrenia. The relative contributions of common and rare variants to the risk of major depression remains to be determined.

Our study is also a departure from 'transdiagnostic' studies of mental illness, which tolerate more fluid boundaries between psychiatric diagnoses (55-57). The concept that *general* susceptibility to mental illness can manifest along a number of DSM-5 diagnoses is plausible but lacks predictive power. There is a great deal of overlap in psychiatric diagnoses associated with common variants found in GWAS (39). In contrast, *GPR156* c.1599G>T has more specific and potentially actionable implications: individuals heterozygous for this variant have a greater than 40% risk for MDD that will manifest during their second or third decade (**Table 1**).

A significant proportion of *GPR156* c.1599G>T heterozygotes had been prescribed at least one medication for depression. Commonly prescribed antidepressant medications target monoaminergic neurotransmission and they have several limitations. Many patients see only marginal benefit or experience side effects that limit tolerability (58). Furthermore, approximately one third of depressed individuals are refractory to antidepressant medications (59). Thus, we need safer, more effective antidepressant therapies, and GPR156 or one of its effectors might prove a suitable target.

GPR156 encodes a G protein-coupled receptor (GPCR) that belongs to a superfamily of more than 800 metabotropic signaling proteins with a common 7-transmembrane structure and varied binding pockets. GPCRs are extensively expressed throughout the nervous system and they are involved in a diversity of functions including primary sensory signaling, cognition, and mood (60, 61). About 34% of all approved drugs in the U.S. and Europe target GPCRs or their proximate effectors, making them the most highly exploited family of pharmacological targets in the human genome (62, 63).

Many GPCRs (including GPR156) have no known ligand and are therefore called 'orphan' receptors. Data from experimental animals, GWAS, and postmortem human transcriptomes

implicate more than a dozen orphan GPCRs in the modulation of emotional behavior (38, 64). GPR156 is one of the orphan class C receptors, first described in 2003 by Calver et al., who employed a bioinformatic strategy using sequence homology with γ-aminobutyric acid type B receptors ($GABA_{B1}$ and $GABA_{B2}$) to identify an uncharacterized protein in brain tissue they termed $GABA_{BL}$ (B-like) (65), which is now called GPR156. Although physiologic ligands and downstream effectors of GPR156 remain unknown, it is highly localized to the habenula (**Fig. 4**), which provides strong evidence for its role in emotional reactivity and stress adaptation. Our behavioral results in mice show that the mutation has a functional impact. Elucidating how GPR156 may regulate the habenula, and identifying its natural ligand, could lead to additional insights into the biology of depression and to the development of more targeted and effective treatments.

Persistent dysfunction or overactivity of neurons in the habenula may contribute to dysphoria (66). The habenula is a tiny structure (30-36 mm^3) just above the posterior thalamus close to the midline, and it connects limbic and basal ganglia structures of the forebrain to nuclei of the ventral midbrain and hindbrain, including the dorsal raphe nucleus, a major source of serotonin, and the ventral tegmental area (VTA), a major source of dopamine (**Fig. 4E**) (34, 67). The habenula is composed of distinct lateral (LHb) and medial (MHb) subdivisions, both of which influence behavior in response to rewarding and aversive stimuli (34, 67-69).

The highest concentration of *GPR156* mRNA is in neurons of the MHb, which receive excitatory inputs from limbic and basal ganglia regions, project to the LHb and interpeduncular nucleus (IPN), and thereby influence activity of the VTA and dorsal raphe nuclei via a dorsal diencephalic circuit (**Fig. 4E**). MHb-IPN firing is hyperactive in rodents subjected to uncontrollable stress (69, 70) or bred for helplessness (71). In Holtzman rats, a genetic strain susceptible to stress-evoked immobility on the FST, the protective actions of fluoxetine correlate with reduced metabolic activity in key structures of the dorsal diencephalic circuit—including the nucleus accumbens, habenula, IPN, and dorsal raphe nuclei (**Fig. 4E**) (45). These same structures are targets of experimental deep brain stimulation to treat intractable depression in humans (72, 73).

Direct implication of MHb-IPN activity in the Porsolt FST is of special relevance to our study (74). Historically, immobility is considered a behavioral correlate of despair in rodents subjected to the FST. *Gpr156*$^{-/-}$ and *GPR156*$^{E533D/E533D}$ show a related but distinctive response; hyperactive and disorganized behavior provoked by the FST might represent a maladaptation to emotional distress that, without rescue, would prove fatal. Patterns of neural activity that underlie this behavior may involve the MHb-IPN circuit (46, 75), and fluoxetine and imipramine might rescue the response by modulating downstream monoaminergic activity. Research to better understand these mechanisms is ongoing.

In conclusion, we demonstrate that a single rare gene variant increases the risk of MDD nearly three-fold. Our findings have implications for the genetic architecture of mental illness generally, and major depression in particular, and introduce a promising biochemical pathway for therapeutic development. This work also introduces a new animal model for study of depression pathophysiology and treatment grounded in human genetics. Furthermore, high expression of GPR156 in the habenula aligns with emerging data about the role of this structure in mediating behavioral adaptations to stress and motivates further research into the medial habenula.

Given their rare nature, it is inherently difficult to establish functional links between rare variants and common diseases. This is unfortunate because they may be highly informative. Here we show the power of using a multimodal approach to address this challenge by harnessing a combination of family studies, community clinical databases, and animal models.

Acknowledgements

B.R.M. is supported by NIMH (K08MH116368) and Hope for Depression Research Foundation (HDRF). R.H. is supported by NIMH: R37 MH068542, NIA: R01 AG043688, NIMH: R01 MH083862, NYSTEM: NYSTEM-C029157, S10 OD018464, and HDRF: RGA-13-003.

ONLINE METHODS

Study participants

The study was approved by the Lancaster General Hospital Institutional Review Board (protocol number 2008-095-CSC) and all subjects consented in writing to participate. The DiscovEHR participants represent a subset of the Geisinger MyCode® Community Health Initiative, which is comprised of individuals who consented to collection, storage, and analysis of biological samples for the purpose of correlation with de-identified electronic health record (EHR) information. Participants were consented in accordance with the Geisinger Institutional Review Board approved protocol number 2006-0258.

Psychiatric assessments

A total of 78 individuals from the Old Order Mennonite pedigree were extensively evaluated using the Structured Clinical Interview for DSM-5 (SCID-5) Research Version instrument. (36) Gathered information was corroborated through household members, clinical providers, and available medical records. The overall data were reviewed by a team of experts to arrive at a final consensus diagnosis for each individual, as previously described.(37)

Exome sequencing

Sample preparation, whole exome sequencing, and sequence data production were performed at the Regeneron Genetics Center (RGC) as previously described.(27)In brief, 1ug of high-quality genomic DNA was used for exome capture using the NimbleGen VCRome 2.1 or the IDT XGen target enrichment reagent. Captured libraries were sequenced on the Illumina HiSeq 2500 platform with v4 chemistry using paired-end 75 bp reads. Exome sequencing was performed such that >85% of the bases were covered at 20x or greater. Raw sequence reads were mapped and aligned to the GRCh38/hg38 human genome reference assembly using BWA-mem. Single nucleotide and indel variants and genotypes were called using GATK's HaplotypeCaller.

Genetic analyses

For downstream genetic analyses, variants were annotated and analyzed using an in-house implemented annotation and analysis pipeline and additional customized Perl bioinformatics scripts for data processing.

IBD segment detection

We used KING 2.1.5 [Manichaikul et al, PMID:20926424) with the --ibdseg command to identify all pairwise IBD segments between carriers of the hg38 chr3:g.120167878(C>A) variant using Illumina GSA genotyping array data for each sample. IBD segments on chromosome 3 were extracted from the KING results file and were visually tiled across the chromosome to illustrate the pileup of shared IBD that spanned the region containing the variant of interest. The shared IBD haplotype was defined as the minimum region of overlap across all variant carriers that has not been broken by recombination events.

Mouse generation

All animal experiments were conducted in accordance with the guidelines of the Institutional Animal Care and Use Committee (IACUC) of Regeneron Pharmaceuticals, Inc., Columbia University, and the New York Stare Psychiatric Institute. Mice were housed on 12 hours light/dark cycle with free access to water and food. Experiments were performed in wild type C57Bl/6NTac (*Gpr156*$^{+/+}$), which express wildtype murine GPR156; humanized mice (*GPR156*$^{hum/hum}$), which have the human *GPR156* gene in place of murine *Gpr156;* and GPR156 humanized mutant mice (*GPR156* $^{E533D/E533D}$) which express the human GPR156 p.Glu533Asp protein variant.

Gpr156 targeting constructs were designed as follows. For knockout of *Gpr156,* a bacterial artificial chromosome (BAC) containing *Gpr156* genomic sequence was modified such that a floxed lacZ reporter cassette containing a neomycin resistance gene under the control of the human *UBC* (ubiquitin) promoter replaced *Gpr156* coding exon 1 just after the start ATG to the end of coding exon 4. The cassette was cloned such that lacZ coding sequence was in frame with the start ATG. This construct was electroporated into100% C57Bl/6NTac embryonic stem cells. To "humanize" the *Gpr156* locus, the mouse ATG to stop genomic sequence was replaced with human *GPR156* ATG to 100 bp beyond the end of the 3' untranslated region. A self-deleting neomycin resistance cassette was placed just after the humanization, at the 3' end of the *Gpr156* locus. This targeting construct was then further modified by adding a single A to T base pair substitution in human *GPR156* coding exon 9 to generate the E533D human mutation and a puromycin resistance cassette replaced the neomycin version. These constructs were electroporated into a 50% C57Bl/6NTac/50% 129SvEvTac embryonic stem cell line. Successfully targeted clones from all electroporations were identified by TaqMan analysis. *Gpr156*$^{-/+}$, *Gpr156*$^{hum/+}$, and *Gpr156*$^{hum\ E533D/+}$ mice were generated using the VelociGene©

method (Valenzuela et al., 2003; Poueymirou et al., 2007) and backcrossed to C57Bl/6NTac as needed.

Animal phenotyping and behavioral assessments

Genetically modified mice were subjected to a battery of behavioral assessments to identify potential gross coordination, balance, locomotion, motor or vestibular deficits, as indicated in **Supplementary Table 2.**

Forced swim test

Mice were subjected to a swimming for six minutes in a cylinder from which they could not escape.(31) Briefly, each mouse was placed individually into a glass cylinder (5 liter capacity, Height 25 cm, Diameter 18 cm) filled with 4 liters of water at 23-25ºC. Four animals were recorded simultaneously using an ANY-maze™ video-tracking system (Stoelting Co.). Flasks were separated by white plastic partitions (40 cm high x 25 cm wide) to prevent mice from seeing one another. Video recording started automatically by motion detection and stopped as soon as a mouse was taken out of the water. An experimenter rescued mice that were unable to complete the six-minute swim test and recorded the time elapsed until rescue.

Drug administration

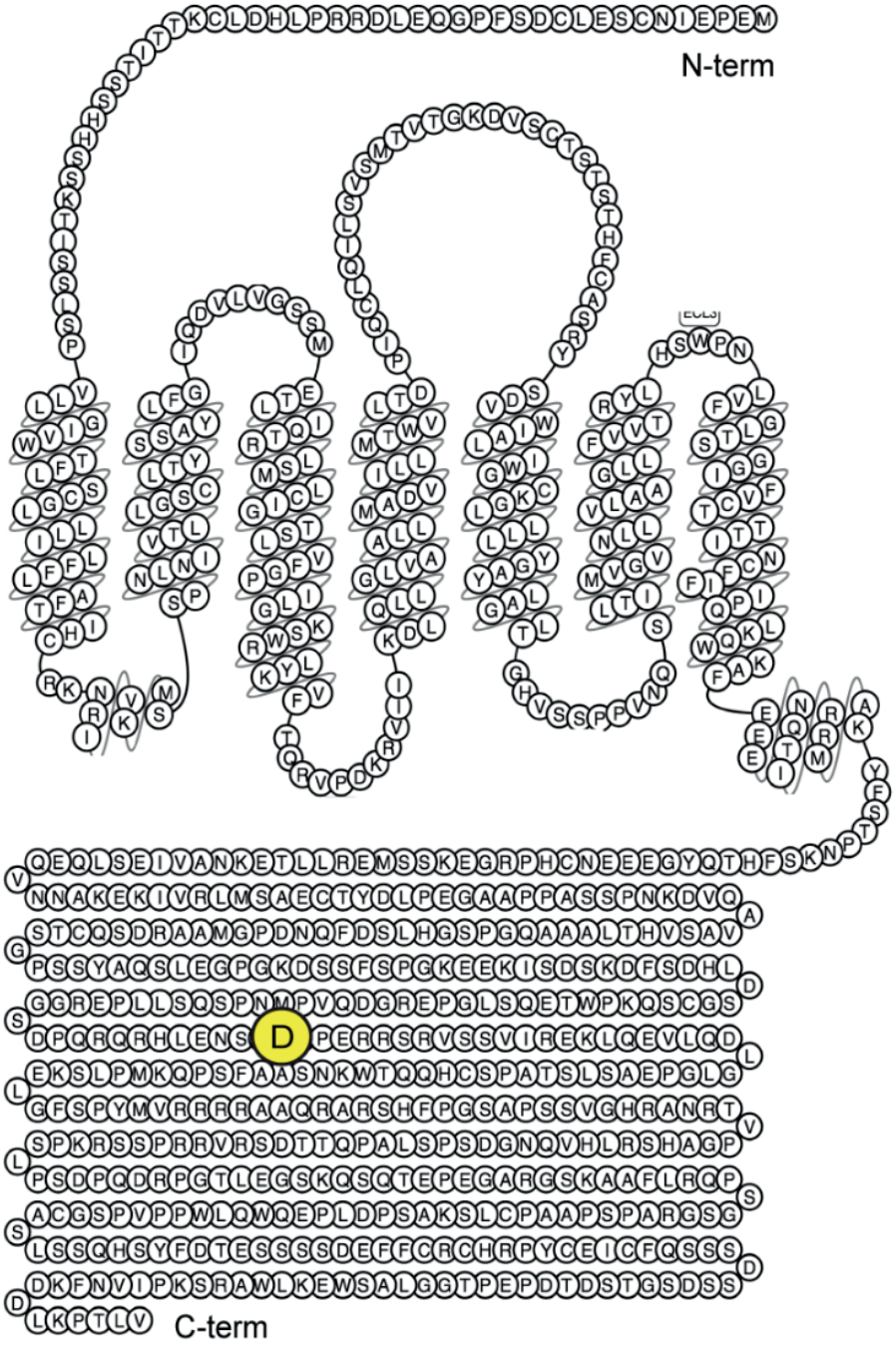
Fluoxetine hydrochloride (Sigma Cat # F132) was diluted in sterile injectable saline and injected i.p. at 10mg/kg•dose, 6 days/week (Monday to Saturday) for nine weeks. Imipramine hydrochloride (Sigma Cat # 10899) was diluted in sterile injectable saline and injected i.p. at 15mg/kg•dose, 6 days/week (Monday to Saturday) for nine weeks.

RNA in situ hybridization of mouse and human GPR156

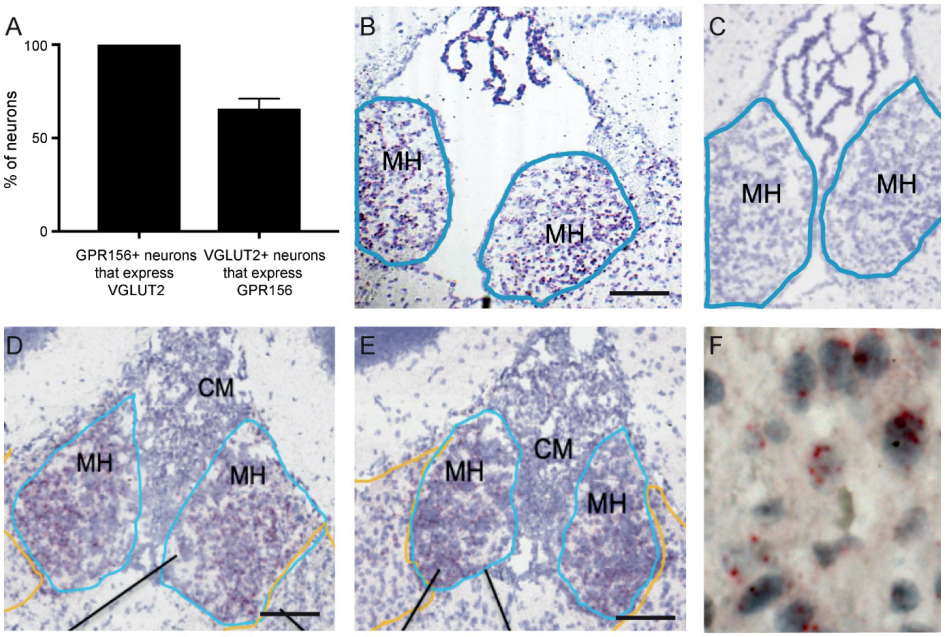
In situ RNA hybridization was performed in mouse brain tissue using the RNAscope 2.5 HD system RED (Advanced Cell Diagnostics; Newark, CA) following the manufacturer's protocol for fresh-frozen tissue samples. RNAscope probes specific for mouse and human GPR156 were designed; the 20 ZZ probe specific to mouse GPR156 targets nucleotides 1494 to 2601 of NM_153394.2 (Mm-Gpr156-No-XHs, cat # 535121, ACDbio) and the 20 ZZ probe specific to human GPR156 targets nucleotides 1313 – 2466 of NM_153002.2 (Hs-GPR156 NO-XMm, Cat #535131). PPIB probe (cat# 313911, ACDbio) and DAPB probe (cat#310043, ACDbio) were used as a positive and a negative control, respectively. For Supplemental Figure 2D-G, we used in a single-plex manual RNAscope assay, and colorimetric signal from the RNAscope probes was captured with a 40X objective using an Aperio slide scanner (Leica Microsystems; Buffalo Grove, IL). For Figure 2C-D, RNAScope Multiplex Fluorescent v2 Assay was performed in accordance with ACD's instructions (Doc. #:3233110-USM). GPR156 transcript was visualized with the red Opal 609 flourophore (Perkin Elmer) and nuclei were counterstained with DAPI (ACDbio). Images were taking using standard confocal microscopy.

For the VGLUT-2 and GPR156 colocalization experiments (Supplementary Figure 2B-C), a heterozygous Vglut2-ires-Cre mouse (Jackson Laboratory, stock # 016963) was crossed with a homozygous Rosa-CAG-LSL-tdTomato-WPRE (Jackson Laboratory, stock # 007909), a Cre-dependent fluorescent reporter. Mice from the F1 generation carrying the Vglut2-ires-Cre allele (Vglut2-tdTomato mice) express the red fluorescent protein tdTomato in all cells expressing vesicular glutamate transporter 2. GPR156 transcript was visualized as above using the RNAScope Multiplex Fluorescent v2 Assay with the green fluorescein fluorophore (PerkinElmer, NEL741001KT).

To detect GPR156 mRNA transcripts in the medial habenula in human tissue, RNAscope® 2.5 HD Detection Reagent-RED assay was performed according to ACD's user manual (Doc. #: 322360-USM). The postmortem interval was less than 26 hours, 2 cm-thick coronal blocks were made, placed for 5 days in 10% formalin solution; paraffin embedded; then cut into 8um thick sections. The medial habenula was identified using anatomical landmarks(80). 8 µm sections were cut from FFPE blocks, mounted onto Superfrost Slides, and tissue was prepped and pretreated as instructed in ACD's instructions (Doc. #: 322452-USM). Tissue sections were baked in a dry oven at 60°C for an hour, deparaffinized using xylene and 100% ethanol. After a 10 min incubation with hydrogen peroxide, Manual Target Retrieval was performed by submerging tissue slides into a beaker of boiling 1X RNAscope® Target Retrieval Reagents. Tissue was pretreated with Protease Plus from ACD kit, GPR156-C1 probe was hybridized 40°C for 2hr, followed by serial hybridization of Amps 1-6, and signal detection with 120µL of Fast RED solution. Cell nuclei were counterstained using and 50% Gil's Hematoxylin #2 followed by 0.02% Ammonia water, and cover-slipped using EcoMount Mounting Medium.



Supplementary Figure 1. Schematic of GPR156 protein with E533D mutation highlighted. Mutated residue is in the long intracellular c-terminal tail.



Supplementary Figure 2. GPR156 expression in the medial habenula.
(A) Percentages of GPR156 and VGLUT2 co-expression in the medial habenula. (B) Expression of GPR156 in medial habenula of wildtype mouse (dark purple) by RNAscope. (C) Loss of GPR156 expression in medial habenula of *GPR156*$^{-/-}$ mice by RNAscope. (D) Expression of human wildtype GPR156 mRNA in medial habenula of *GPR156*$^{hum/hum}$ mice by RNAscope. (E) Expression of human mutant GPR156 mRNA in medial habenula of *GPR156*$^{E533D/E533D}$ mice by RNAscope. (F) *In situ* mRNA hybridization (RNAscope) of postmortem human medial habenula shows *GPR156* mRNA (dark red dots). Scale bars = 100 µm (B-E) and 10 µm(F)

Subject	Sex	Electronic Medical Record Diagnoses	Antidepressant Treatment
Subject 1	Female	Depressive disorder, not elsewhere classified	Fluoxetine
Subject 2	Female	Depressive disorder, not elsewhere classified	Duloxetine
Subject 3	Male	Depressive disorder, not elsewhere classified	Citalopram
Subject 4	Male	Depressive disorder, not elsewhere classified	Mirtazapine
Subject 5	Male	Major depressive disorder	Mirtazapine, sertraline
Subject 6	Male	Sleep disturbance, unspecified	Sertraline

Supplementary Table 1. Association of *GPR156* c.1599G>T with depressive illness and anti-depressant treatment among patients in the regional Giesinger healthcare system. All patients with the variant and an electronic medical record had depressive symptoms that reached the threshold for antidepressant treatment in a community setting.

Phenotype	Behavioral assay	GPR156$^{-/-}$	GPR156$^{hum/hum}$	GPR156$^{E533D/E533D}$
Coordination, balance and motor deficits	Rotarod	No significant difference	No significant difference	No significant difference
	Catwalk	No significant difference	No significant difference	No significant difference
	Balance beam	No significant difference	No significant difference	No significant difference
Activity, general locomotion	Open field	No significant difference	No significant difference	No significant difference
Vestibular dysfunction	Righting reflex Tail hanging reflex Wire hanging	No significant difference	No significant difference	No significant difference

Supplementary Table 2. Summary of motor and vestibular behavioral assessments in GPR156 transgenic lines. There was no significant difference between the transgenic lines and wildtype mice in any of the behaviors.

REFERENCES

1. WHO. Depression and other common mental disorders: global health estimates. World Health Organization. 2017; https://apps.who.int/iris/handle/10665/254610.
2. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 2011;9:90.
3. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health. 2013;34:119-38.
4. Insel TR. Disruptive insights in psychiatry: transforming a clinical discipline. The Journal of clinical investigation. 2009;119(4):700-5.
5. Malhi GS, Outhred T, Morris G, Hamilton A, Das P, Mannie Z. Primary Prevention of Mood Disorders: A Primary Concern That Requires Urgent Action. Journal of the American Academy of Child and Adolescent Psychiatry. 2018;57(9):629-31.
6. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Archives of general psychiatry. 2005;62(6):593-602.
7. Hirschfeld RM. The epidemiology of depression and the evolution of treatment. The Journal of clinical psychiatry. 2012;73 Suppl 1:5-9.
8. Malhi GS, Mann JJ. Depression. Lancet. 2018;392(10161):2299-312.
9. Malhi GS, Outhred T, Hamilton A, Boyce PM, Bryant R, Fitzgerald PB, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders: major depression summary. Med J Aust. 2018;208(4):175-80.
10. Verduijn J, Verhoeven JE, Milaneschi Y, Schoevers RA, van Hemert AM, Beekman ATF, et al. Reconsidering the prognosis of major depressive disorder across diagnostic boundaries: full recovery is the exception rather than the rule. BMC Med. 2017;15(1):215.
11. Spijker J, de Graaf R, Bijl RV, Beekman AT, Ormel J, Nolen WA. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry. 2002;181:208-13.
12. Belmaker RH, Agam G. Major depressive disorder. The New England journal of medicine. 2008;358(1):55-68.
13. Bartram M, Stewart JM. Income-based inequities in access to psychotherapy and other mental health services in Canada and Australia. Health Policy. 2019;123(1):45-50.
14. Sudak DM. Advancing the accessibility of psychotherapy: learning from our international colleagues. J Psychiatr Pract. 2015;21(2):150-3.
15. Jakobsen JC, Gluud C, Kirsch I. Should antidepressants be used for major depressive disorder? BMJ Evid Based Med. 2020;25(4):130.
16. Sinyor M, Cheung CP, Abraha HY, Lanctot KL, Saleem M, Liu CS, et al. Antidepressant-placebo differences for specific adverse events in major depressive disorder: A systematic review. Journal of affective disorders. 2020;267:185-90.
17. Miller BR. A Mechanistic Study of p11 Reveals a Promising New Rapid-Action Antidepressant Target. Biological psychiatry. 2020;88(5):e23-e4.
18. Villas Boas GR, Boerngen de Lacerda R, Paes MM, Gubert P, Almeida W, Rescia VC, et al. Molecular aspects of depression: A review from neurobiology to treatment. Eur J Pharmacol. 2019;851:99-121.
19. Kendler KS, Gatz M, Gardner CO, Pedersen NL. A Swedish national twin study of lifetime major depression. The American journal of psychiatry. 2006;163(1):109-14.

20. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry. 2013;18(4):497-511.
21. Yu C, Arcos-Burgos M, Baune BT, Arolt V, Dannlowski U, Wong ML, et al. Low-frequency and rare variants may contribute to elucidate the genetics of major depressive disorder. Transl Psychiatry. 2018;8(1):70.
22. Mitchell KJ, Porteous DJ. GWAS for psychiatric disease: is the framework built on a solid foundation? Molecular psychiatry. 2009;14(8):740-1.
23. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature genetics. 2018;50(9):1219-24.
24. Bergen SE, Ploner A, Howrigan D, Group CNVA, the Schizophrenia Working Group of the Psychiatric Genomics C, O'Donovan MC, et al. Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. The American journal of psychiatry. 2019;176(1):29-35.
25. Puffenberger EG. Genetic heritage of the Old Order Mennonites of southeastern Pennsylvania. American journal of medical genetics Part C, Seminars in medical genetics. 2003;121C(1):18-31.
26. Strauss KA, Puffenberger EG. Genetics, medicine, and the Plain people. Annual review of genomics and human genetics. 2009;10:513-36.
27. Strauss KA, Gonzaga-Jauregui C, Brigatti KW, Williams KB, King AK, Van Hout C, et al. Genomic diagnostics within a medically underserved population: efficacy and implications. Genet Med. 2018;20(1):31-41.
28. Crowgey EL, Washburn MC, Kolb EA, Puffenberger EG. Development of a Novel Next-Generation Sequencing Assay for Carrier Screening in Old Order Amish and Mennonite Populations of Pennsylvania. J Mol Diagn. 2019;21(4):687-94.
29. Lazaridis I, Tzortzi O, Weglage M, Martin A, Xuan Y, Parent M, et al. A hypothalamus-habenula circuit controls aversion. Molecular psychiatry. 2019;24(9):1351-68.
30. Liu B, Cao Y, Wang J, Dong J. Excitatory transmission from ventral pallidum to lateral habenula mediates depression. World J Biol Psychiatry. 2020:1-7.
31. Porsolt RD, Bertin A, Jalfre M. Behavioral despair in mice: a primary screening test for antidepressants. Arch Int Pharmacodyn Ther. 1977;229(2):327-36.
32. Bogdanova OV, Kanekar S, D'Anci KE, Renshaw PF. Factors influencing behavior in the forced swim test. Physiol Behav. 2013;118:227-39.
33. Keller MC, Miller G. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? Behav Brain Sci. 2006;29(4):385-404; discussion 5-52.
34. Boulos LJ, Darcq E, Kieffer BL. Translating the Habenula-From Rodents to Humans. Biological psychiatry. 2017;81(4):296-305.
35. Li B, Piriz J, Mirrione M, Chung C, Proulx CD, Schulz D, et al. Synaptic potentiation onto habenula neurons in the learned helplessness model of depression. Nature. 2011;470(7335):535-9.
36. First MB, Williams JBW, Karg R, Spitzer RL. Structured Clinical Interview for DSM-5, Research Version, Non-patient Edition. Arlington, VA: American Psychiatric Association; 2015.
37. Strauss KA, Markx S, Georgi B, Paul SM, Jinks RN, Hoshi T, et al. A population-based study of KCNH7 p.Arg394His and bipolar spectrum disorder. Hum Mol Genet. 2014;23(23):6395-406.
38. Watkins LR, Orlandi C. Orphan G Protein Coupled Receptors in Affective Disorders. Genes (Basel). 2020;11(6).
39. Brainstorm C, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360(6395).

40. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science. 2016;354(6319).
41. Campbell S, Macqueen G. The role of the hippocampus in the pathophysiology of major depression. Journal of psychiatry & neuroscience : JPN. 2004;29(6):417-26.
42. Yang Y, Wang H, Hu J, Hu H. Lateral habenula in the pathophysiology of depression. Curr Opin Neurobiol. 2018;48:90-6.
43. Aizawa H, Zhu M. Toward an understanding of the habenula's various roles in human depression. Psychiatry and clinical neurosciences. 2019;73(10):607-12.
44. Porsolt RD, Anton G, Blavet N, Jalfre M. Behavioural despair in rats: a new model sensitive to antidepressant treatments. Eur J Pharmacol. 1978;47(4):379-91.
45. Padilla E, Shumake J, Barrett DW, Sheridan EC, Gonzalez-Lima F. Mesolimbic effects of the antidepressant fluoxetine in Holtzman rats, a genetic strain with increased vulnerability to stress. Brain research. 2011;1387:71-84.
46. Nair SG, Strand NS, Neumaier JF. DREADDing the lateral habenula: a review of methodological approaches for studying lateral habenula function. Brain research. 2013;1511:93-101.
47. Can A, Dao DT, Arad M, Terrillion CE, Piantadosi SC, Gould TD. The mouse forced swim test. J Vis Exp. 2012(59):e3638.
48. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). American journal of human genetics. 1993;52(3):506-16.
49. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668-81.
50. Psychiatric GCSC. A framework for interpreting genome-wide association studies of psychiatric disorders. Molecular psychiatry. 2009;14(1):10-7.
51. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008;456(7218):18-21.
52. Amin N, de Vrij FMS, Baghdadi M, Brouwer RWW, van Rooij JGJ, Jovanova O, et al. A rare missense variant in RCL1 segregates with depression in extended families. Mol Psychiatry. 2018;23(5):1120-6.
53. Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. Genome Biol. 2008;9(8):109.
54. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K, et al. A unified genetic theory for sporadic and inherited autism. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(31):12831-6.
55. Taquet M, Smith SM, Prohl AK, Peters JM, Warfield SK, Scherrer B, et al. A structural brain network of genetic vulnerability to psychiatric illness. Molecular psychiatry. 2020.
56. Huang J, Perlis RH, Lee PH, Rush AJ, Fava M, Sachs GS, et al. Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. The American journal of psychiatry. 2010;167(10):1254-63.
57. Cross-Disorder Group of the Psychiatric Genomics C, Genetic Risk Outcome of Psychosis C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381(9875):1371-9.
58. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet. 2018;391(10128):1357-66.

59. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. The American journal of psychiatry. 2006;163(11):1905-17.
60. Grammatopoulos DK. Regulation of G-protein coupled receptor signalling underpinning neurobiology of mood disorders and depression. Mol Cell Endocrinol. 2017;449:82-9.
61. Basith S, Cui M, Macalino SJY, Park J, Clavio NAB, Kang S, et al. Exploring G Protein-Coupled Receptors (GPCRs) Ligand Space via Cheminformatics Approaches: Impact on Rational Drug Design. Front Pharmacol. 2018;9:128.
62. Sriram K, Insel PA. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? Molecular pharmacology. 2018;93(4):251-8.
63. Hauser AS, Attwood MM, Rask-Andersen M, Schioth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov. 2017;16(12):829-42.
64. Alavi MS, Shamsizadeh A, Azhdari-Zarmehri H, Roohbakhsh A. Orphan G protein-coupled receptors: The role in CNS disorders. Biomed Pharmacother. 2018;98:222-32.
65. Calver AR, Michalovich D, Testa TT, Robbins MJ, Jaillard C, Hill J, et al. Molecular cloning and characterisation of a novel GABAB-related G-protein coupled receptor. Brain Res Mol Brain Res. 2003;110(2):305-17.
66. Kaye A, Ross DA. The Habenula: Darkness, Disappointment, and Depression. Biological psychiatry. 2017;81(4):e27-e8.
67. Hikosaka O, Sesack SR, Lecourtier L, Shepard PD. Habenula: crossroad between the basal ganglia and the limbic system. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2008;28(46):11825-9.
68. Metzger M, Souza R, Lima LB, Bueno D, Goncalves L, Sego C, et al. Habenular connections with the dopaminergic and serotonergic system and their role in stress-related psychiatric disorders. Eur J Neurosci. 2019.
69. McLaughlin I, Dani JA, De Biasi M. The medial habenula and interpeduncular nucleus circuitry is critical in addiction, anxiety, and mood regulation. Journal of neurochemistry. 2017;142 Suppl 2:130-43.
70. Mirrione MM, Schulz D, Lapidus KA, Zhang S, Goodman W, Henn FA. Increased metabolic activity in the septum and habenula during stress is linked to subsequent expression of learned helplessness behavior. Front Hum Neurosci. 2014;8:29.
71. Shumake J, Edwards E, Gonzalez-Lima F. Opposite metabolic changes in the habenula and ventral tegmental area of a genetic model of helpless behavior. Brain research. 2003;963(1-2):274-81.
72. Dandekar MP, Fenoy AJ, Carvalho AF, Soares JC, Quevedo J. Deep brain stimulation for treatment-resistant depression: an integrative review of preclinical and clinical findings and translational implications. Molecular psychiatry. 2018;23(5):1094-112.
73. Sartorius A, Kiening KL, Kirsch P, von Gall CC, Haberkorn U, Unterberg AW, et al. Remission of major depression under deep brain stimulation of the lateral habenula in a therapy-refractory patient. Biological psychiatry. 2010;67(2):e9-e11.
74. Thornton EW, Bradbury GE, Davies C. Increased immobility in an automated forced swimming test following lesion of the habenula in rats: absence of evidence for a contribution from motor impairment. Behav Neurosci. 1990;104(1):37-43.
75. Wang Z, Wang L, Yamamoto R, Sugai T, Kato N. Role of the lateral habenula in shaping context-dependent locomotor activity during cognitive tasks. Neuroreport. 2013;24(6):276-80.
76. Qin C, Luo M. Neurochemical phenotypes of the afferent and efferent projections of the mouse medial habenula. Neuroscience. 2009;161(3):827-37.

77. Antolin-Fontes B, Ables JL, Gorlich A, Ibanez-Tallon I. The habenulo-interpeduncular pathway in nicotine aversion and withdrawal. Neuropharmacology. 2015;96(Pt B):213-22.
78. Paolini M, De Biasi M. Mechanistic insights into nicotine withdrawal. Biochem Pharmacol. 2011;82(8):996-1007.
79. Graziane NM, Neumann PA, Dong Y. A Focus on Reward Prediction and the Lateral Habenula: Functional Alterations and the Behavioral Outcomes Induced by Drugs of Abuse. Front Synaptic Neurosci. 2018;10:12.
80. Kim JW, Naidich TP, Ely BA, Yacoub E, De Martino F, Fowkes ME, et al. Human habenula segmentation using myelin content. Neuroimage. 2016;130:145-56.

Chapter 3

Cortical Overgrowth in a Preclinical Forebrain Organoid Model of *CNTNAP2*-Associated Autism Spectrum Disorder.

Job O. de Jong[1,2], Ceyda Llapashtica[1,2], Matthieu Genestine[1,2], Kevin Strauss[3], Frank Provenzano[4], Yan Sun[1],Huixiang Zhu[1,2], Giuseppe P. Cortese[1,2], Francesco Brundu[5], Karlla W. Brigatti[3], Barbara Corneo[6], Bianca Migliori[7], Raju Tomer[8], Steven A. Kushner[1,9], Christoph Kellendonk[1,2,10], Jonathan A. Javitch[1,2,10], Bin Xu*[1,2] & Sander Markx*[1,2]

*Corresponding author

Affiliations
1. Department of Psychiatry, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA.
2. Division of Molecular Therapeutics, New York State Psychiatric Institute, New York, NY, USA.
3. Clinic for Special Children, Strasburg, Pennsylvania, USA.
4. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, NY, USA.
5. Department of Systems Biology, Columbia University, New York, NY, USA
6. Stem Cell Core Facility, Columbia University, New York, NY, USA.
7. Tech4Health and Neuroscience Institutes, NYU Langone Health, New York, NY, USA.
8. Department of Biological Sciences, Columbia University, New York, NY, USA.
9. Department of Psychiatry, Erasmus MC University Medical Center, Rotterdam, The Netherlands
10. Department of Pharmacology, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA.
Manuscript accepted
Contact information
Sander Markx, M.D.: sm2643@cumc.columbia.edu
Bin Xu, Ph.D.: bx2105@cumc.columbia.edu

ABSTRACT

We utilized forebrain organoids generated from induced pluripotent stem cells of patients with a syndromic form of Autism Spectrum Disorder (ASD) with a homozygous protein-truncating mutation in *CNTNAP2,* to study its effects on embryonic cortical development. Patients with this mutation present with clinical characteristics of brain overgrowth. Patient-derived forebrain organoids displayed an increase in volume and total cell number that is driven by increased neural progenitor proliferation. Single-cell RNA sequencing revealed PFC-excitatory neurons to be the key cell types expressing *CNTNAP2*. Gene ontology analysis of differentially expressed genes (DEgenes) corroborates aberrant cellular proliferation. Moreover, the DEgenes are enriched for ASD-associated genes. The cell type specific signature genes of the *CNTNAP2*-expressing neurons are associated with clinical phenotypes previously described in patients. The organoid overgrowth phenotypes were largely rescued after correction of the mutation using CRISPR-Cas9. This *CNTNAP2*-organoid model provides opportunity for further mechanistic inquiry and development of new therapeutic strategies for ASD.

INTRODUCTION

Autism spectrum disorder (ASD) and related neurodevelopmental disorders represent a major public health burden (1). To better understand these conditions, mechanistic insight into their development is needed. Rodent models with modified genes associated with ASD have facilitated this effort, but translating promising preclinical findings to the clinic has remained challenging (2). This is at least partly due to differences in the biology between species (3-5) and underscores the need for elucidating human-specific disease mechanisms for psychiatric disorders. Forebrain organoids generated from human induced pluripotent stem cells (hiPSCs) derived from subjects with well-characterized brain disorders are a promising method to study human-specific neurobiology. Indeed, brain organoid modeling of ASD and related neurodevelopmental disorders may help reveal critical disease phenotypes and associated disease mechanisms that may not be recapitulated in animal models (6-9).

A homozygous loss-of-function (LoF) mutation, c.3709DelG, in contactin-associated-protein-like 2 *(CNTNAP2)* causes a rare and severe neurodevelopmental syndrome characterized by ASD, intellectual disability, early-onset epilepsy and an increase in head circumference – collectively termed Cortical Dysplasia Focal Epilepsy (CDFE) syndrome (10). The inherited mutation and associated clinical syndrome were first described in a family from the genetically isolated Old-Older Amish population (10). Additional cases carrying different homozygous LoF mutations were subsequently discovered in subjects outside the Amish founder population, all presenting with a similar constellation of symptoms (11, 12). Thus far, no homozygous LoF mutations have been described in healthy control individuals, indicating that these are associated with a severe and a highly penetrant phenotype.

CNTNAP2 encodes a cell adhesion molecule that is structurally related to the neurexins, but has shown to be functionally distinct, fulfilling roles in both peripheral and central nervous system development and physiology (12). Known functions of *CNTNAP2* have been delineated in a number of *Cntnap2* knock-out and knock-down rodent models, and include critical roles in potassium channel clustering in myelinated axons (13, 14) as well as dendritic arborization (14), synaptic development and function (14, 15) and neuronal migration (16). Interestingly, *Cntnap2* null mice display several behavioral phenotypes reminiscent of the neurodevelopmental syndrome associated with homozygous LoF mutations, including deficits in social interaction and cognition, an increase in repetitive-stereotyped behaviors, and a lowered seizure threshold (16).

Little is known about the function of *CNTNAP2* in the human brain. Human expression data shows high abundance in different cortical areas during embryonic development (17). Histopathology and MRI findings from patients from the Old Order Amish population diagnosed

with CDFE-syndrome show cortical dysplasia, cortical grey matter thickening, ectopic neurons in the white matter tracts and increased hippocampal cell number (10). Together, these findings imply an early embryonic origin of pathophysiology. As embryonic development is faithfully recapitulated in brain organoid modeling systems as shown by transcriptome analyses (18, 19), we generated forebrain organoids to study how disruption of *CNTNAP2* affects cortical embryonic development.

To this end, we generated hiPSC lines from subjects from the Old Order Amish founder population, all of whom had been diagnosed with CDFE-syndrome and were genetically confirmed to carry the homozygous c.3709DelG mutation in *CNTNAP2.* Using these lines, we generated forebrain organoids and compared these to organoids generated from healthy controls from the same population to examine the effects of the mutation on cortical development. We show an increase in volume of patient-derived organoids compared to controls. This enlargement of organoid volume results from increased proliferation of neural progenitor cells (NPCs) and other proliferating cells, which leads to an increase in total cell number. Single-cell RNA sequencing (scRNAseq) revealed PFC-excitatory neurons to be the key cell types expressing *CNTNAP2*. Differential gene expression analyses at both the bulk- and single cell level corroborate changes in biological processes including cell proliferation and neuronal differentiation. Differentially expressed genes were enriched for ASD-associated genes and Weighted Gene Co-Expression Networks (WGCNAs). The cell type-specific signature genes in *CNTNAP2*-expressing neural cell types are associated with multiple clinical disease phenotypes that have previously described in patients with *CNTNAP2* LoF mutations. Together, our work provides a critical step towards understanding the role of *CNTNAP2* in human brain development. These findings provide possible mechanisms underlying the CDFE syndrome, including the increased head circumference and cortical gray matter thickening, epilepsy and impairment in language development (10). Future studies should focus on further elucidating these disease mechanisms which can, in turn, lead to the development of novel treatments of *CNTNAP2*-associated ASD and other genetic subtypes of ASD.

METHODS

Ascertainment of Subjects and Clinical Phenotypes

The study was approved by the Lancaster General Hospital (LGH) Institutional Review Board (IRB) (LGH IRB protocol number 2008-095). All investigators complied with all relevant ethical regulations for work with human participants. Parents of all subjects provided written informed consent to participate on behalf of their children. Neuropsychiatric diagnoses and head circumferences measurements were done during routine clinical visits for pediatric care at the Clinic for Special Children in Strasburg, Pennsylvania. The generation of hiPS cell lines

and derived forebrain organoids was approved by New York State Psychiatric Institute (NYSPI) IRB (NYSPI IRB protocol number 7500) and Columbia University's Human Embryo and Embryonic Stem Cell Research Committee.

MRI analysis

Gray matter volume

We quantified total gray matter volume relative to the total brain volume of 6 whole-brain MRIs from patients carrying the c.3709DelG mutation in *CNTNAP2*, and compared these to 4 template MRIs that were generated from whole-brain MRIs of age-matched healthy children (20, 21). We used Statistical Parametric Mapping for image segmentation and summed gray matter and white matter masks to calculate total brain volume. For each image, the sum value of voxels for each tissue class was used to generate a ratio between gray matter and white matter.

Corpus callosum volume

Clinical MRI scans were spatially normalized into the corresponding age-matched template space T1-weighted MRI scan. Since all scans were of varying thicknesses, with axial acquisitions preventing accurate fine quantification of the corpus callosum, a side-by-side qualitative evaluation was made for each subject and template scan. For each participant MRI scan in corresponding template space, a number was assigned to indicate "corpus callosum size deviation from normal" with scores of 0, 1 and 2 (no change, subtle change, marked change) and the sign being direction (+ was template larger than control). This was done three times for each participant blindly and the median score recorded. A one-sample t-test of the values and the template values (which are 0) was performed.

hiPSC generation and characterization

hiPSC lines were generated from three patients with *CNTNAP2*-associated ASD carrying the homozygous c.3709DelG mutation and three healthy controls from the same Amish population who did not carry the mutation, by a non-integrating Sendai virus-based reprogramming method as previously described (22). Karyotyping was performed on twenty G-banded metaphase cells at 450–500 band resolution as previously described (23). All hiPSC lines used in this study were between passage 10 and 25. The difference in passage number for each pair of patient- and control-derived line never exceeded 3 passages.

Forebrain organoid culture

Forebrain organoids were generated from patient- and control hiPSC-lines using a previously published organoid differentiation protocol (24), with minor modifications. Briefly, hiPSCs were dissociated with 0.5 mM EDTA in PBS and triturated to generate a single cell suspension. A total of 10,000 cells were then plated into each well of an ultra-low-attachment 96-well

plate (Nunc) to form embryoid bodies (EBs) in medium containing mTesR1, 1 µg/ml heparin and Penicillin-Streptomycin antibiotics. EBs were exposed to medium containing the ROCK inhibitor Y27632 (50 µM) for the first 24h, followed by 5 days without interference in 96-well plates. On day 5, the medium was switched to medium containing DMEM/F12 (1:1) (Gibco; 11330), N-2 supplement (Gibco; 17502-048), MEM-NEAA (Invitrogen), Glutamax (Invitrogen 35050-061), Penicillin-Streptomycin and 0.1 mM ®-Mercaptoethanol, Dorsomorphin (2 µM), SB431542 (10 µM) and IWR1e (3 µM). On day 25, 0.2% Chemically Defined Lipid Concentrate was added to this medium until day 39. On day 40, EBs were transferred to ultra-low attachment 24 well plates with 1 ml of medium per organoid/well. Between day 40 and 80, 1 ml of medium containing DMEM/F12 (1:1) (N-2 supplement MEM-NEAA, Glutamax, Penicillin-Streptomycin and 0.1 mM ®-Mercaptoethanol, 10% fetal bovine serum (FBS) and 1% Matrigel (Corning) was replaced weekly. After day 80 medium was replaced once per week. Matrigel was added bi-weekly to prevent excess Matrigel deposition on the outer shell of the organoid.

Immunohistochemistry

See table S1 for all antibodies and respective concentrations used. Organoids were fixed for 4 h in 10% formalin at room temperature, cryo-protected in 15% sucrose, dissolved in PBS for 4 h, and then placed in 30% sucrose in PBS overnight. The next day organoids were embedded in optimal cutting temperature (O.C.T.) solution and cut into 14 µm sections on a Leica Cryostat. After heat-induced epitope retrieval in 40 mM Sodium Citrate/ 0.05% Tween-20 (pH6), tissue sections were blocked for 1h in 10% horse serum in PBS containing 0.1% TritionX-100 (PBSTx0.1). Primary antibodies were incubated overnight at 4 ºC in PBSTx0.1. Secondary antibodies were incubated for 1h at room temperature and nuclei were counterstained with 300 nM DAPI solution. Sections were mounted on glass slides using Prolong antifade solution and imaged on a LEICA SP8 confocal microscope.

Western blotting

Proteins were extracted in lysis buffer containing protein inhibitor (cOmplet Mini, Roche, Ref: 11836170001) in tissue protein extraction reagent (Thermo Scientific , Ref: 78510) according to manufacturer's protocol. Briefly, organoids were washed twice with DPBS, then incubated in lysis buffer for 10 minutes on ice and vortexing vigorously. Samples were spun down at 10,000g and was stored at -80°C for downstream analysis. Extracted protein was separated on a NuPAGE 4 to 12 % gradient gel (Invitrogen, Ref: NP0322). Primary antibodies (see table S1) were bound by HRP-conjugated secondary antibodies and visualized using the Azure c600 imaging system.

Light-sheet microscopy

Whole organoids were fixed and stained in 300 nM DAPI in PBS for 30 mins to ensure staining of the nuclei of the outer shell of the organoid. Light-sheet microscopy images were acquired on a Zeiss Z-1 Light-sheet microscope from 6 angles in 45º degree increments starting at 0º.

CRISPR-Cas9 Rescue-line generation

We randomly selected one of the three patient hiPSC lines for CRISPR genome editing.

gRNA and donor template design. Guide-RNAs were designed using *CT-finder* software (25). gRNAs were designed around the c.3709DelG mutation site in the *CNTNAP2* locus on chromosome 7q35. gRNA target A on the minus strand is 5'-GGGTCGGTGGCGGAC-GACATGGG-3' and gRNA target B on the plus strand is 5' - GCACCTGGATCACCT-GATTCAGG-3'. The donor template was a single stranded 181-bp oligonucleotide (see Table S2). This sequence is complementary to the antisense DNA strand and contains the wildtype DNA sequence and a silent blocking mutation near the 5'-end of gRNA target A. The donor template flanks the mutation site with 90bp on each side (Fig. S2A).

hiPSC gene-editing using CRISPR Cas9. We followed a CRISPR Cas9 gene editing protocol previously described (26) with minor modifications. Briefly, gRNA oligonucleotides were cloned into the plasmid pSpCas9n(BB)-2A-GFP encoding Cas9-nickase. pSpCas9n(BB)-2A-GFP (PX461) was a gift from Feng Zhang (Addgene plasmid # 48140 ; http://n2t.net/addgene:48140 ; RRID:Addgene_48140). hiPSCs were maintained in MTesR1 medium. Cells were passaged using 0.05 mM EDTA in PBS or Accutase. Prior to nucleofection, cells were treated with ROCK-inhibitor for 24 hours. Plasmids and donor template were introduced into the hiPSCs through nucleofection using Amaxa nucleofector II. 72 hours post nucleofection, GFP + cells were selected using FACS and grown as single cell-derived colonies onto irradiated MEF coated 35 mm dishes. Colonies with undifferentiated hiPSC morphology were manually picked and further grown on 96-well plates. Clones derived from single cells were genotyped using Sanger sequencing on PCR amplicons (see table S2 for primers).

Viral transduction

pAAV.hSyn.eGFP.WPRE.bGH was a gift from James M. Wilson (Addgene plasmid # 105539; http://n2t.net/addgene:105539; RRID:Addgene_105539). The virus was added to the organoid growth media and incubated for 1 week at a concentration of 10^7 virus particles per ml. Transduction efficiency was evaluated 10 days post-transduction using an epi-fluorescence microscope.

Isotropic fractionation

We adapted a previously developed protocol to accurately calculate total cell number and relative cell type fractions based on nuclear immunostaining (27). Briefly, organoids were fixed for 4h in 10% formalin at room temperature. After blocking for 4h in 10% horse serum in PBS containing 0.2% Triton X-100 (PBSTx0.2), primary antibodies were incubated PBSTx0.2 on whole organoids for 20h on a shaker at room temperature to ensure complete tissue penetration. Organoids were washed for 4h in PBSTx0.2. Secondary antibodies were incubated for 20h in in PBS PBSTx0.2, followed by a 4h wash in PBSTx0.2. Organoids were fixed for a second time for 20h, and then homogenized using a glass tissue homogenizer in 10 mM sodium citrate buffer containing 1% Triton X-100. Total cell counts were performed by counting DAPI-stained nuclei in a hemocytometer on an upright Zeiss epifluorescence microscope. Relative cell type fractions were calculated from images acquired on a Leica SP8 confocal microscope.

BrdU labeling

Organoids were incubated using 10 µM BrdU dissolved in growth medium for 2h at 37 ºC, then washed twice with PBS and subsequently fixed for 4h in 10% formalin at room temperature. We subsequently followed our immunostaining protocol to stain for BrdU after heat denaturing the DNA in a 40 mM sodium citrate buffer containing 0.05% Tween-20. For the BrdU/Ki67 quantification analysis, serial sections of 14 µm were made every fifth section, with 3-4 sections for each organoid.

Total RNA isolation and bulk RNA sequencing

Ten forebrain organoids per line (i.e., 3 control-derived lines, 3 patient-derived lines) were collected at 8 weeks *in vitro*. Total RNA was isolated from hiPSC-derived cortical neurons using miRNeasy kit (Qiagen, USA) according to instructions of the manufacturer. RNA was suspended in RNase-free water. The concentration and purity of each sample was determined by spectrophotometer (ND1000; Nanodrop) and confirmed by Microchip Gel Electrophoresis (Agilent) using Agilent 2100 Bioanalyzer Chip according to the instructions of the manufactures. A poly-A pull-down step was performed to enrich mRNAs from total RNA samples (200 ng to 1 µg per sample, RIN > 8 required) and proceeded on library preparation by using Illumina TruSeq RNA prep kit. Libraries were then sequenced using Illumina HiSeq2000 at the Columbia Genome Center. Multiplex samples with unique barcodes were mixed in each lane, which yields targeted number of single-end 100 bp reads for each sample, as a fraction of 180 million reads for the whole lane. RTA software (Illumina) was used for base calling, and bcl2fastq (version 1.8.4) for converting BCL to fastq format, coupled with adaptor trimming. The reads were mapped to a reference human genome (hg19) using Tophat2.0 (version 2.0.4) with four mismatches (--read-mismatches = 4) and 10 maximum multiple hits (--max-multihits = 10). We have uploaded the RNAseq data used in this paper in NCBI's Gene Expression Omnibus under accession code GSE174569.

Transcriptome correlation analysis

Pearson's r correlations were calculated between the top 1000 most highly expressed genes from the bulk RNA sequencing data from the 3 control-derived organoid lines and the same genes from all available samples from the Brainspan transcriptome atlas (28).

Differential expression analysis

DESeq2 (Version 1.28.1), an R package based on a negative binomial distribution that models the number of reads from RNA-seq experiments and tests for differential expression(29), was used to determine differentially expressed genes (DEGs) between mutants and control samples. The list of significantly DEGs was defined at false discovery rate (FDR) adjusted p-value (padj) < 0.05.

Gene ontology analysis

To determine a common functional relationship among the top DEGs, the enrichment of biological processes was tested using Metascape (30) with default settings. Gene lists of upregulated and downregulated DEGs were analyzed separately.

Permutation tests

Permutation tests were performed by generating 10,000 lists of gene-sets with equal size to the differentially expressed genes. We then compared these randomly generated genesets to the gene lists of interest (WGCNA networks, ASD-SFARI genes (SFARI-Gene_genes_10-31-2019release_10-05-2020export) as well as control gene lists derived from other common diseases, and tested whether the overlap between DEGs and permuted gene lists differed in a statistically significant manner from the randomly generated gene lists. Two-sided p-values were derived from z-scores using formula $z = (x-\mu)/\sigma$.

Image analyses

Confocal- and light-sheet microscopy images were processed using ImageJ software (Version 2.0.0-rc-69/1.52p) (https://imagej.nih.gov/ij/). BrdU/Ki67 quantifications and nuclear fraction quantifications were carried out using Cell Profiler software (Version 4.1.3) (31).

Statistical analyses

All statistical and bioinformatic analyses were performed using *R* programming language (version 4.0.0) (32). The graphs in the manuscript display boxplots showing median, first-, and third quartiles. Whiskers show the largest and smallest values no further than 1.5 times the interquartile range from the first and third quartile, respectively, with all data points plotted individually. Comparisons between patient- and control-derived organoids were done using a linear mixed model from software package *Lme4* (33) with genotype status designated as a fixed effect and cell line as a random effect. Relationships between two continuous variables

were tested using linear regression. Statistical significance from the mixed model was calculated using Likelihood Ratio Tests (LRT). For comparisons of the CRISPR rescue-line to its original patient line, and the gray matter volume analysis, we used two-sided student's t-tests. P-values < 0.05 were considered statistically significant. Outliers were never removed from any of the analyses.

Single-cell RNAseq dissociation and sequencing

We collected 3 organoids per sample and dissociated using the papain dissociation protocol (Worthington Biochemicals) according to the manufacturer's protocol. Briefly, organoids were incubated for 30 minutes at 37°C in Earle's Balanced Salt Solution (EBSS) / Albumin-ovomucoid inhibitor/ papain with gentle shaking and mechanical dissociation. The pellet was resuspended and incubated with (EBSS) / Albumin-ovomucoid inhibitor / DNAse for 2 minutes followed by incubation with ovomucoid for 2 minutes. The cells were washed and resuspended with DMEM/F12. After determination of the cell density and viability, cells were submitted to single-cell RNA sequencing (10X Genomics, Chromium Single cell 3' v3) to recover about 5,000 sequenced cells per sample with more than 50,000 reads per cell. We have uploaded the RNAseq data used in this paper in NCBI's Gene Expression Omnibus under accession code GSE174569.

Single-cell RNAseq data analysis

The Cell Ranger 4.0.0 pipeline (10x Genomics) was used to align and aggregate all reads from scRNA-seq to the GRCh38 human reference genome with default settings and produce a balanced "filtered feature bc matrix". The data was imported into the SCANPY software (Version 1.7.1) (34) where genotype information was added. Quality control was conducted to ensure all samples analyzed contained highly consistent quality. Cells expressing a minimum of 200 genes were kept, and counts were normalized for each cell by the total expression, multiplied by 10^6 and log-transformed. LIGER (Version 2.0.1) software (35) was used to integrate the different samples and variation in the cells' transcriptional profile was visualized by uniform manifold approximation and projection for dimension reduction (UMAP) function (36) in SCANPY. To annotate the cell types in brain organoids, we used two independent approaches. First, we downloaded the gene-cell count matrix and cell type annotation files of the single cell RNAseq analysis of human embryonic prefrontal cortex tissue from the cell browser (37). We then utilized this dataset to train the SingleR software(Version 1.0.6) (38), and to infer the cell types in our dataset. In addition, we used the "Leiden clustering method" implemented in the SCANPY to conduct unsupervised cell population clustering (39). We used the Rank "rank genes groups analysis" of SCANPY to determine the top 100 cell type differentiating genes of each cell type or each cell cluster annotated by singleR and Leiden clusters. For each cell type or cluster, differentially expressed (DE) genes were determined using model-based analysis of single cell transcriptomics (MAST, Version 1.12.0) algorithm(40).

RESULTS

Clinical description of patients carrying the c.3709DelG mutation in *CNTNAP2*.

Patients carrying the c.3709DelG mutation in *CNTNAP2* were reported to have increased head circumference, epileptic seizures, language regression, intellectual disability and ASD (10). Because the sample size of the initial cohort in which the increased head circumference was described was small (10), we expanded this assessment by measuring the head circumferences of an additional 17 male and 20 female pediatric patients from the Old-Order Amish founder population all carrying the inherited homozygous c.3709DelG mutation in *CNTNAP2* and all presenting with the same syndromic ASD. We compared these data to head circumference reference data published by the WHO (41). This analysis revealed an increased head circumference in both male and female patients compared to the WHO mean (weighted z-score = 2.93, $p < 0.003$ (Fig. S1A). To expand upon the initial description of increased gray matter thickness (10), we subsequently analyzed brain MRIs from 6 patients carrying the c.3709DelG mutation in *CNTNAP2* and compared these to age-matched clinical templates. We found an increase in total gray matter (GM) volume relative to total brain volume (TBV) in patients compared to controls (Patients: GM/TBV= 0.73; Controls: GM/TBV= 0.54; t-test, $p = 0.004$, n= 6 patients, 4 control templates (Fig. S1B and Table S3).

hiPSC and Forebrain Organoid Generation and Characterization.

It has been previously shown that C*NTNAP2* expression is highest in frontal- and anterior cortical regions in the human embryonic brain (17). We therefore focused on investigating the role of *CNTNAP2* on forebrain development by generating forebrain organoids. We generated hiPSC lines from monocytes of three human female patients (ages 10, 12 and 17 years) carrying the homozygous c.3709DelG mutation in *CNTNAP2* (Fig. 1A) and three female healthy control subjects from the same population. All 3 patients were assessed by a pediatrician (KS) and were diagnosed with epilepsy and a syndromic form of ASD, characterized by cognitive-, behavioral-, and language regression and delayed developmental milestones starting around age 1.

To confirm stemness in hiPSC lines used for organoid generation we performed RT-PCR for markers *NANOG* and *OCT4* (Fig. S1C); we performed G-band karyotype analysis to ensure the absence of chromosomal abnormalities in all patient- and control-derived cell lines (Fig. S1D). We confirmed the genotypes of patient- and control-derived hiPSCs using Sanger sequencing (Fig. S1E).

We deployed three approaches to determine whether our forebrain organoid model recapitulates key characteristics of early telencephalic brain development. First, after 4 weeks in

culture, immunostaining of both patient- and control-derived organoids showed ventricle-like structures composed of NPCs expressing forebrain-specific transcription factor PAX6, (Fig. 1B). Dorsal forebrain identity was further confirmed by immunostaining for transcription factor FOXG1 (Fig. S1F). Organoids from both genotype groups demonstrate a clear pattern of MAP2-expressing neurons positioned around the ventricle-like structures and preferentially located in the outer layers of the organoids, indicative of initial NPC differentiation and subsequent migration towards the outer cortex-like region (Fig. 1B). At 13 weeks *in vitro,* organoids displayed expression of early born cortical layer marker TBR1 and CTIP2 as shown using immunostaining (Fig. S1G).

Second, at the transcriptional level, we performed bulk-RNA sequencing to reveal that at 8 weeks in culture, the organoids display the highest correlations with multiple forebrain structures from 9 week-old human fetal brain tissue, when compared to transcriptional profiles from samples from all brain structures and timepoints from the Brainspan transcriptome database (28) (Fig. 1C). Third, we evaluated whether the cell type composition of the organoids resembles that of early human fetal prefrontal brain, by performing scRNAseq on 13 week-old organoids. Unsupervised Leiden clustering revealed 27 distinct clusters that were consistently present in all case- and control-derived organoids (Fig. S1H and table S4). We annotated these cell populations using PFC cell-type gene expression signatures previously identified by scRNAseq of human embryonic PFC tissue (average age: 16.3 post conceptional weeks (pcw)) (37). These analyses demonstrate that the forebrain organoids contain a range embryonic prefrontal cortical cell types (Fig. 1D; Fig. S1I; Fig. S1J and table S5).

CNTNAP2 expression in forebrain organoids.

To determine the *CNTNAP2* expression pattern in control-derived organoids, we performed Western blotting, which revealed that protein expression starts around 4 weeks in culture and then increases over time (Fig. 1E). Western blotting for CNTNAP2 utilizing an antibody (Antibody A, intracellular, Fig. 1A) directed against the c-terminus that is deleted in the c.3079delG mutation showed that full-length CNTNAP2 protein was not detected in patient-derived organoids while it was indeed detected in control-derived organoids (Fig. 1F). We then used an antibody directed against the laminin G domain of the protein located upstream of the transmembrane domain (Antibody B Fig. 1A), which is predicted to be present in the truncated CNTNAP2 protein. Notably, while immunoblotting with antibody B revealed CNTNAP2 expression in the control samples, no protein was detected in patient-derived organoids suggesting that the truncated *CNTNAP2* mRNA and/or protein are unstable (Fig. S1K). Analyzing the mRNA sequence reads for *CNTNAP2* scRNAseq confirms that the mRNA is indeed present in lower numbers, with case organoids showing a lower read-coverage at the 3' end than controls (Fig. S1L). To compare these expression data in organoids to expression in the intact brain, we analyzed data from the publicly available gene-expression database Brain-

Span, which characterizes the embryonic spatiotemporal expression pattern of the developing human brain. These data show the highest *CNTNAP2* expression levels in the first trimester of embryonic development in multiple cortical areas, including orbitofrontal, prefrontal and temporal cortices (Fig. S1M). Other areas of high gene expression during embryonic development include the striatum and cerebellar cortex (Fig. S1M). The BrainSpan data are in line with previous expression studies of *CNTNAP2* in the human embryonic brain (17, 42). Analysis of scRNAseq data indicates that at 13 weeks *in vitro,* in control organoids, *CNTNAP2* is most highly expressed in different types of PFC-excitatory neurons as annotated previously (37): 'Early Born Deep Layer/subplate Excitatory Neuron PFC' (EN-PFC), 'Early Born Deep Layer/subplate Excitatory Neuron V1' (EN-V1) and 'New born excitatory neurons' (nEN) (hereafter referred to as '*CNTNAP2*-expressing cell types'). In patient derived-organoids, *CNTNAP2*-expression is markedly lower in these cell types (Fig 1G). This is in line with cell type expression pattern in human PFC cell types, with *CNTNAP2* being expressed most highly in excitatory and inhibitory neuronal cell types (Fig. S1N).

These results indicate that the *CNTNAP2* expression pattern in control organoids are in accordance with *in vivo* gene expression and thus indicate that forebrain organoids serve as a well suited preclinical neuronal model system to study the effects of this homozygous LoF-mutation on early human cortical development.

Patient-derived forebrain organoids display an increase in volume.

To investigate whether the brain overgrowth characteristic of these patients could be recapitulated in our model system, we set out to assess the size of patient- and control-derived organoids over time by measuring the projected surface area of the organoid using 2D bright field microscopy images (Fig. 2A). After 4 weeks in culture, case- and control organoids that were generated from an equal number of hiPSCs displayed no difference in surface area. At 9 weeks, case-derived organoids demonstrates an increase in surface area relative to control-derived organoids, and at 13 weeks *in vitro*, patient-derived organoids showed a 1.5-fold increase in 2D projected surface area compared to control-derived organoids (LRT, $p = 0.002$, n= 6-7 organoids/line) (Fig. 2B).

As the organoids are not spherical but irregularly shaped, we used light-sheet microscopy to more precisely quantify organoid volume (Fig. 2C). We found a 1.5-fold volumetric increase for patient-derived organoids compared to controls (LRT, $p = 0.005$) (Fig. 2D). We then utilized light-sheet imaging to investigate whether the total surface area of the middle z-section could be used as an reliable estimate for organoid volume using a linear model, which was indeed the case (Linear model, $R^2 = 0.606$, $p < 0.001$) (Fig. 2E). We therefore proceeded using 2D brightfield microscopy images for quantifying organoid volume in all following analyses.

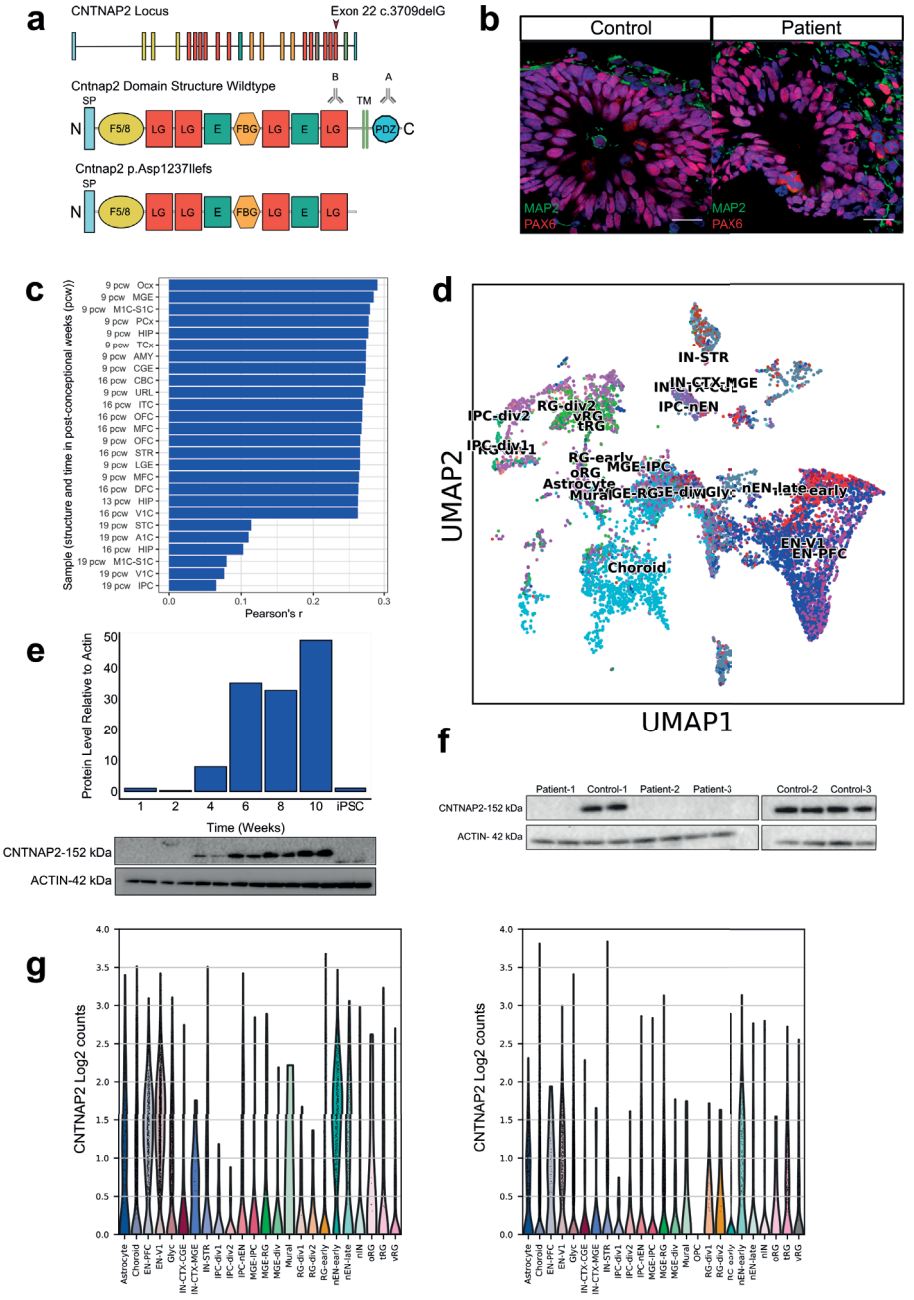

a
CNTNAP2 Locus
Exon 22 c.3709delG
Cntnap2 Domain Structure Wildtype
SP
B
A
TM
N
F5/8
LG
E
FBG
PDZ
C
Cntnap2 p.Asp1237Ilefs
b
Control
Patient
MAP2
PAX6
c
Sample (structure and time in post-conceptional weeks (pcw))
9 pcw Ocx
9 pcw MGE
9 pcw M1C-S1C
9 pcw PCx
9 pcw HIP
9 pcw TCx
9 pcw AMY
9 pcw CGE
16 pcw CBC
9 pcw URL
16 pcw ITC
16 pcw OFC
16 pcw MFC
9 pcw OFC
16 pcw STR
9 pcw LGE
9 pcw MFC
16 pcw DFC
13 pcw HIP
16 pcw V1C
19 pcw STC
19 pcw A1C
16 pcw HIP
19 pcw M1C-S1C
19 pcw V1C
19 pcw IPC
0.0
0.1
0.2
0.3
Pearson's r
d
IN-STR
IN-CTX-MGE
IN-CTX-CGE
IPC-nEN
IPC-div2
vRG
tRG
IPC-div1
RG-div2
RG-early
oRG
MGE-IPC
Astrocyte
Mural
GE-RG
E-div
Glyc
nEN-late
nEN-early
EN-V1
EN-PFC
Choroid
UMAP2
UMAP1
e
Protein Level Relative to Actin
50
40
30
20
10
0
1
2
4
6
8
10
iPSC
Time (Weeks)
CNTNAP2-152 kDa
ACTIN-42 kDa
f
Patient-1
Control-1
Patient-2
Patient-3
Control-2
Control-3
CNTNAP2-152 kDa
ACTIN- 42 kDa
g
CNTNAP2 Log2 counts
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Astrocyte
Choroid
EN-PFC
EN-V1
Glyc
IN-CTX-CGE
IN-CTX-MGE
IN-STR
IPC-div1
IPC-div2
IPC-nEN
MGE-IPC
MGE-RG
MGE-div
Mural
OPC
RG-div1
RG-div2
RG-early
nEN-early
nEN-late
oRG
tRG
vRG

Figure 1. Forebrain organoid characterization and *CNTNAP2* expression.

A. *CNTNAP2* locus and protein domain structures of wildtype and mutant protein. The 24 exons of *CNTNAP2* are color-coded based on domain structures of the full-length protein. Deletion of a guanine base at position 3709 in exon 22 (indicated with red arrow) of the gene results in shift in open reading frame and a stop codon at amino acid position 1253. The truncated variant of the protein (p.Asp1237Ilefs) is terminated right before the C-terminal transmembrane domain. Antibody A (intracellular) is a polyclonal antibody directed against the C-terminal end of the protein. Antibody B (extracellular) is a monoclonal antibody directed against the Laminin-G domain upstream of the transmembrane domain (also see table S1 for antibodies). SP= signal peptide; F5/8= discoidin/neuropilin homology domain; LG= Laminin G domain; E= EGF-Like Domain; FBG= Fibrinogen-Like Region; PDZ = PDZ-interaction domain.

B. Both control-derived and patient-derived organoids show PAX6-expressing dorsal forebrain neural progenitor cells in ventricular zone-like regions after 4 weeks in culture. Neuronal dendrosomatic marker MAP2-expressing cells are preferentially located in the outer zones of the cortical organoids. All control- and patient-derived organoids displayed similar expression patterns for these markers. Scale bar represents 50 µm.

C. Bar plot displaying Pearson's r correlations of the bulk RNAseq expression profile of 8 week-old wildtype organoids compared to all publicly available expression profiles of multiple brain structures across multiple time points from BrainSpan. Displayed are the top-20 highest correlating structures and the 6 lowest. The highest correlations are for multiple forebrain structures at 9 post-conceptional weeks, indicating that the organoids have the highest resemblance to these structures at the transcriptional level. (OCx = occipital neocortex, MGE = medial ganglionic eminence, M1C-S1C = primary motor-sensory cortex, PCx = parietal neocortex, HIP = hippocampus, TCx = temporal neocortex, AMY = amygdaloid complex, CGE = caudal ganglionic eminence, CBC = cerebellar cortex, URL = upper (rostral) rhombic lip, ITC = inferolateral temporal cortex (area TEv, area 20), OFC = orbital frontal cortex, MFC = anterior (rostral) cingulate (medial prefrontal) cortex, STR = striatum, LGE = lateral ganglionic eminence, MFC = anterior (rostral) cingulate (medial prefrontal) cortex, DFC = dorsolateral prefrontal cortex, V1C = primary visual cortex (striate cortex, area V1/17), STC = posterior (caudal) superior temporal cortex (area TAc), A1C primary auditory cortex (core), IPC = posteroventral (inferior) parietal cortex.

D. UMAP plot of combined control samples (n=2 samples; 11,328 cells) of 13 week-old brain organoids displaying the distribution of various cell lineages annotated according to Nowakowski et al. (2017) (ref 37).

E. Western blot analysis showing increasing CNTNAP2 protein levels over time relative to actin in a control-derived organoid line, roughly corresponding to the developmental pattern of CNTNAP2 expression pattern in human embryonic brain. Each sample contains two lanes with technical replicates; the bar graph represents mean quantification from these two lanes. Source data are provided as a Source Data file.

F. Western blot analysis showing presence of full-length CNTNAP2 protein levels in control-derived- and absence in patient-derived forebrain organoids using an antibody directed against the C-terminal end of the protein (Antibody A in panel A). Each sample contains two lanes with technical replicates. Source data are provided as a Source Data file.

G. Violin plot displaying *CNTNAP2* mRNA log2 counts for cell types in control organoids (left panel) and case organoids (right panel). *CNTNAP2* is most highly expressed in EN-PFC, EN-V1, nEN-early and nEN-late in control organoids (Abbreviations: EN-PFC = Early Born Deep Layer/subplate Excitatory Neuron PFC, EN-V1 = Early Born Deep Layer/subplate Excitatory Neuron V1, Glyc = Glycolysis, IN-CTX-CGE1 = CGE/LGE-derived inhibitory neurons, IN-STR = striatal interneurons, IPC-div1 = Dividing Intermediate Progenitor Cells RG-like, IPC-div2 = Intermediate Progenitor Cells RG-like, IPC-nEN = Intermediate Progenitor Cells EN-like, MGE-IPC = MGE Progenitors, MGE-RG = MGE Radial Glia 1, MGE-div = dividing MGE Progenitors, RG-div1 = Dividing Radial Glia (G2/M-phase), RG-div2 = Dividing Radial Glia (S-phase), RG-Early = early radial glia, U1 = unknown1, nEN-early = Newborn Excitatory Neuron - early born, nEN-late = Newborn Excitatory Neuron - late born, nIN = MGE newborn neurons, oRG = outer radial gla, tRG= truncated radial glia, vRG = ventricular radial glia) annotated according to Nowakowski et al. (2017) (ref 37).

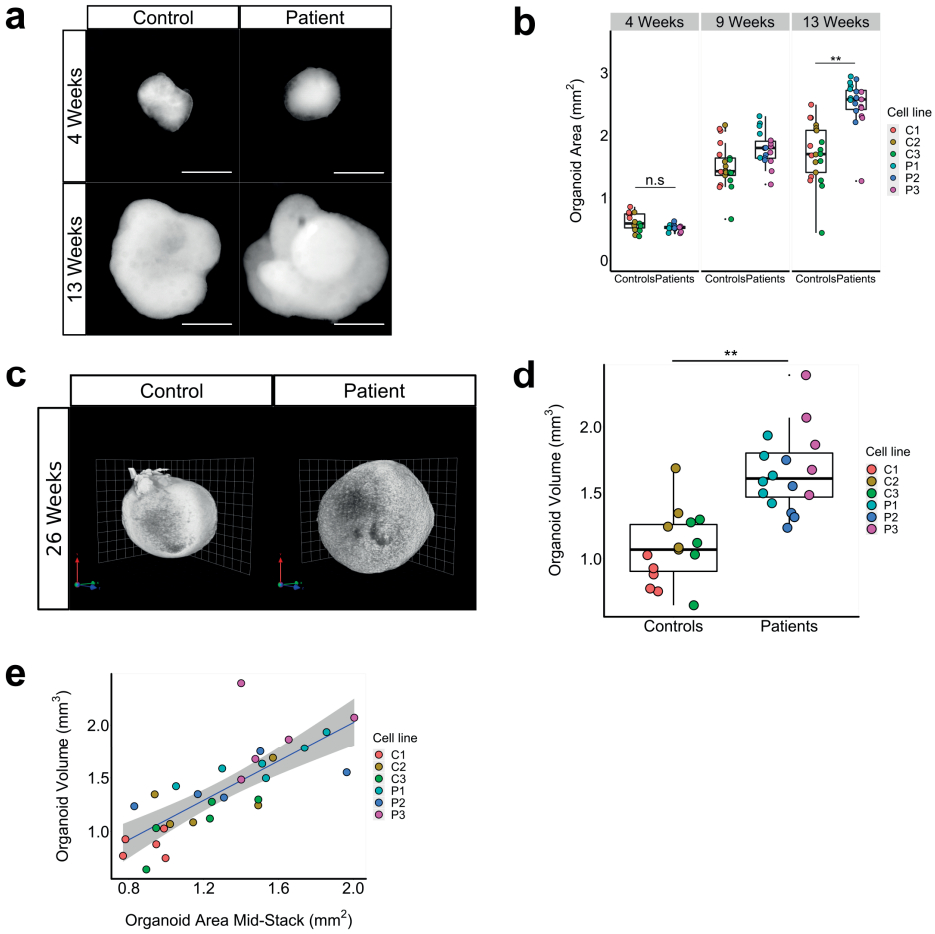


Figure 2. Patient-derived forebrain organoids display an increase in volume.

(A and B). Representative image of bright field images of patient- and control-derived organoids over time (A) and quantifications based on 2D bright field image measurements (B), showing an increase in surface area at 9 weeks and a 1.5-fold increase in area of patient-derived- versus control-derived organoids at week 13 (Likelihood Ratio Test (LRT), p = 0.002, $\chi 2(1)$= 9.42, n= 3 cell-lines/genotype, C1:n= 7 organoids (orgs), C2: n= 7 orgs, C3: n= 7 orgs, P1: n=6 orgs, P2: n= 6 orgs; P3: n= 6 orgs), while no difference in size was observed at week 4. Scale bar represents 500 µm.

(C and D). Representative image of light-sheet microscopy images of control- and patient-derived cortical organoids at 26 weeks *in vitro* (C) and volumetric quantification of light-sheet images showing a 1.5-fold volumetric increase for patient-derived cortical organoids compared to controls (D); (LRT, p = 0.005, $\chi 2(1)$= 7.85, n= 3 cell-lines/genotype, C1,C2,C3,P2,P3: n= 5 orgs, P1:n = 6 orgs).

E. Scatterplot showing the relationship between the area of a middle z-stack of a light-sheet image of a forebrain organoid and the volume at 26 weeks as measured by the 3D rendering of the light-sheet image. We constructed a linear model of volume as a function of the area of the middle z-stack (blue line) (F(1,29)= 44.59, R^2 = 0.606, p = 2.529e-07). The shaded grey area represents the constructed 95% confidence interval around the mean. This result indicates that 2D bright field microscope images can be used to reliably estimate organoid volume.

Boxplots display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = p < 0.05; ** = p < 0.01; *** = p < 0.001

Increased organoid volume is driven by an increased total cell number caused by increased proliferation of dividing cells

To establish the factors that contribute to the volumetric increase of the patient-derived organoids, we utilized isotropic fractionation (27) to estimate the total cell number in the organoids. We found a 2.5-fold increase in total cell number after 6 weeks in culture for patient-derived organoids compared to controls (LRT, p <0.001, n =4 organoids/line) (Fig. 3A). As expected, organoid surface area and total cell number correlated positively with each other at both 6- and 13 weeks *in vitro* (Fig. 3B and Fig. S2A). To determine whether cell size also contributes to the volumetric increase of the organoids, we measured the neuronal soma of AAV-Synapsin-GFP transduced organoids, and found no difference in neuronal soma volume between genotype groups (LRT, p= 0.69) (Fig. S2B). We then estimated the cell cycle length by co-labeling with BrdU and Ki67, and found that proliferating cells have a shorter cell cycle (Average cell cycle time (Tc) = 54h) for patient-derived organoids compared to controls (Tc= 83h) (LRT, p =0.02, 1.54-fold change = 1.54, n = 3 organoids/line) (Fig. 3C and D). Co-labeling of BrdU with PAX6 confirmed that most proliferation took place in the PAX6+ NPCs (Fig. S2C) in organoids from both genotype groups (LRT, p=0.46, n= 3 organoids/genotype). Next, we quantified the size of the progenitor pool by calculating the absolute numbers of PAX6+ cells at 6 weeks *in vitro*. We found a 2.6-fold increase in absolute numbers of NPCs in the patient-derived organoids compared to controls (LRT, p <0.001, n=4 organoids/line) (Fig. 3E-G). These data show that increased proliferation activity of NPCs and other proliferating cells leads to an increase in the progenitor pool, which is a major contributor to the volumetric increase observed in patient-derived organoids. At 13 weeks *in vitro*, the scRNAseq data reveal a decrease in the unsupervised clusters that represent *CNTNAP2*-expressing neuronal cell types (EN-PFC, EN-V1, nEN) in patient-derived organoids (Fig. 3H, Fig. S2D and E). A possible explanation could be a deficit in neuronal differentiation towards these cellular fates; alternatively, the increase in glycolysis cells ("Glyc") in patient-organoids could imply an increase in cellular stress in these cell types (Fig. S2E and Fig. 1G). This could be cell type specific due to the mutation in *CNTNAP2* and/or a property intrinsic to the brain organoid culture as described previously (43).

Gene-ontology analysis based on bulk RNAseq and scRNAseq data corroborates abnormal cellular proliferation and neurogenesis processes.

To evaluate the transcriptome changes associated with the phenotypes described above, we conducted bulk RNA sequencing (RNAseq) on 8-week old organoids. Principal component analysis (PCA) of the RNAseq data shows a clear separation in the transcriptional pattern between the two genotype groups, with 47% of the variance explained by principal component 1 (Fig. 4A). Differential gene expression analysis identified a total of 339 differentially expressed genes (false discovery rate adjusted p-value < 0.05), including 89 upregulated- and

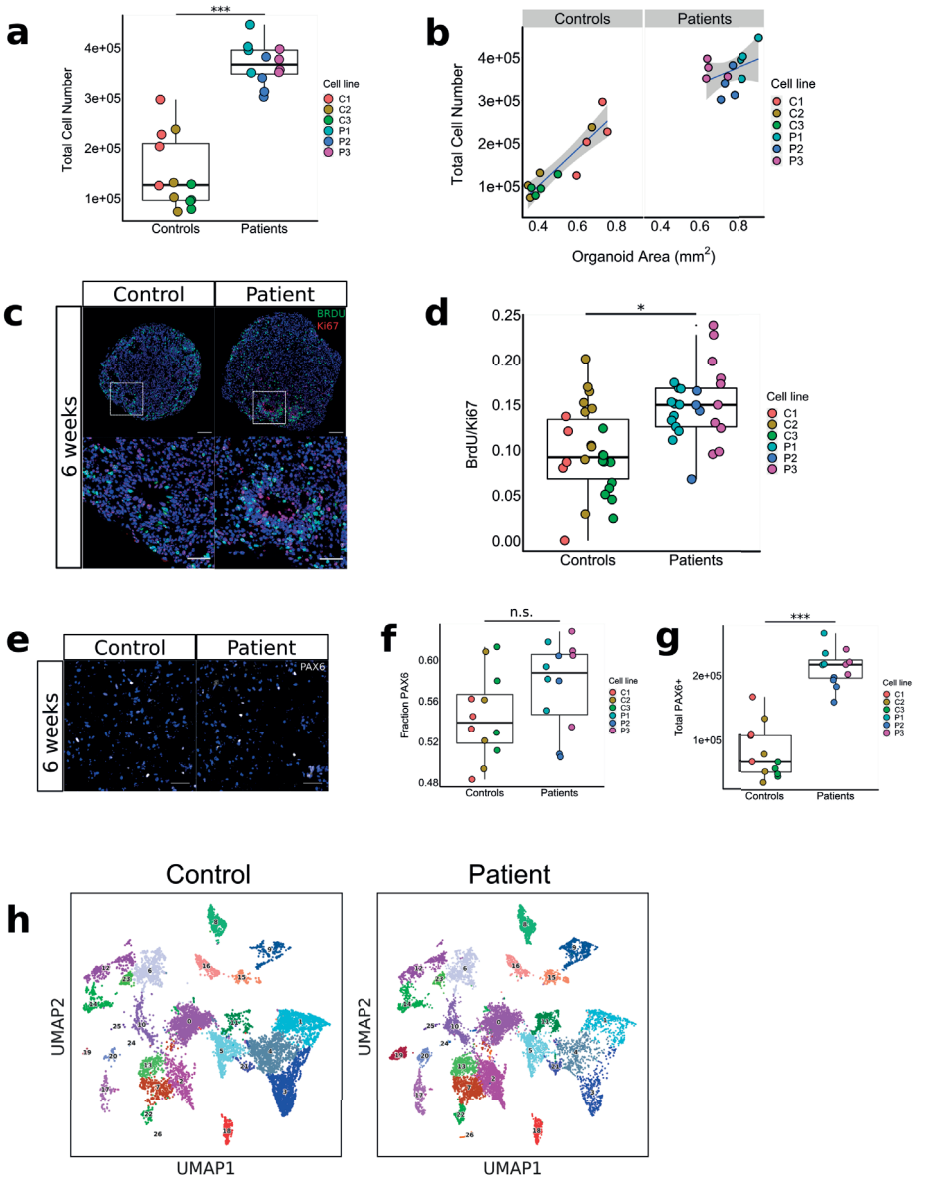


Figure 3. Volumetric increase is driven by increased proliferation, which leads to expansion of progenitor pool and increased total cell number.

(A and B). Total cell number quantification of patient- and control-derived organoids after 6 weeks in culture, showing a 2.5-fold increase in total cell number for patient-derived organoids (LRT, p = 0.00043, $\chi 2(1)$= 13.25, n= 3 cell lines/genotype, 4 orgs/cell line) (A), and the linear relationship between 2D organoid surface area and total cell number for patient- and control-derived organoids after 6 weeks in culture. (Controls*:* F(1,10) = 50.22, R^2= 0.83 ,p= 3.348e-05; patients: F(1,10) = 1.82, R^2= 0.15, p = 0.21, n= 3 cell-lines/genotype, 4 orgs/cell line). Boxplots in panel A display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = p < 0.05; ** = p < 0.01; *** = p < 0.001.

(C and D). Representative image of organoid sections stained for BrdU and Ki67 after 6 weeks in culture (C) and quantification (D) of ratio BrdU/Ki67 between patient- and control-derived organoids, showing a 1.5-fold increase in BrdU/Ki67 ratio for patient-derived organoids (LRT, p = 0.02, $\chi 2(1)= 5.55$, n= 3 cell-lines/genotype, 3 orgs/cell line, 3 sections/org). Using the formula Tc = Ts/(BrdU+/Ki67+), where Tc = cell cycle length, and Ts = S-phase length, estimated mean cell cycle lengths for patient- and control-derived organoids are ~54 and ~83 hours, respectively. Scale bars in C represent 100 µm in top panels, and 50 µm in lower panels. Boxplots in panel D display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = p < 0.05; ** = p < 0.01; *** = p < 0.001.
(E, F and G). Representative image of confocal microscopy images of nuclei from homogenized control and patient-derived organoids cultured for 6 weeks and stained for NPC-marker PAX6 (E), quantification of fractions of PAX6 + cells at 6 weeks *in vitro* (F) and total number of PAX6+ cells at 6 weeks *in vitro* (G), showing a 2.6-fold increase in total PAX6+ cells at 6 weeks (LRT, p = 0.0003, $\chi 2(1)= 13.35$, n= 3 cell-lines/genotype, 4 orgs/cell-line), with an equal fraction of PAX6+ cells at 6 weeks. Boxplots in panel F and G display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = p < 0.05; ** = p < 0.01; *** = p < 0.001.
H. UMAP plots displaying the 26 clusters of cell population determined by the unsupervised Leiden clustering method for all control-derived organoids (n=2 samples; 11,328 cells) (left panel) and all patient-derived organoids (n=3 samples; 16,780 cells) (right panel). There is a lower number of cells in clusters 1, 3 and 4 in patient-derived samples, corresponding to cell types EN-V1, EN-PFC and nEN-early annotated in Figure 1 D.

250 downregulated genes in patient-derived organoids compared to controls (Fig. 4B) (Table S6). To investigate whether these differentially expressed genes have known roles in biological processes related to the observed *CNTNAP2*-associated disease phenotypes, we performed Gene Ontology (GO) analysis (30) and found a statistically significant enrichment for genes involved in various biological processes, including cell proliferation and neurogenesis (Fig. 4B, S3A and S3B and Table S7 and S8). We then performed GO analysis on the cell type-specific signature genes for each of the 3 *CNTNAP2*-expressing cell types (EN-PFC, EN-V1 and nEN-early) in the control-organoids from the scRNAseq data at 13 weeks. These genes were enriched for involvement in biological processes including neuronal differentiation (Fig. S3C-E).

Differentially expressed genes show enrichment for ASD-associated genes and ASD Weighted Gene Co-expression Networks

As ASD has been described as one of the core features of the neurodevelopmental syndrome associated with the homozygous c.3709DelG *CNTNAP2* mutation in the Old-Order Amish patients (10), we set out to investigate whether affected genes are enriched for genes with an established association with ASD. An enrichment could suggest that different genetic predispositions could lead to autism through effects on overlapping sets of genes. To do so, we compared the differentially expressed genes from the bulk sequencing experiment to the ASD-associated genes curated by SFARI Gene (44). Of the 980 genes in SFARI categories 1-4 and the category titled 'syndromic', 19 were differentially expressed between patient- and control-derived organoids (Fig. 4C and table S9). To test whether this enrichment reached the threshold for statistical significance, we performed permutation analyses (n = 10,000). In each iteration, we generated a random gene list of equal length to the differentially expressed genes and found that this was indeed the case (z = 2.42, p=0.016) (Fig. 4D). We then performed pseudo-bulk differential expression analysis on the scRNAseq data, calculating differential gene expression in the three *CNTNAP2*-expressing cell types (EN-PFC, EN-V1, nEN-early) (Table

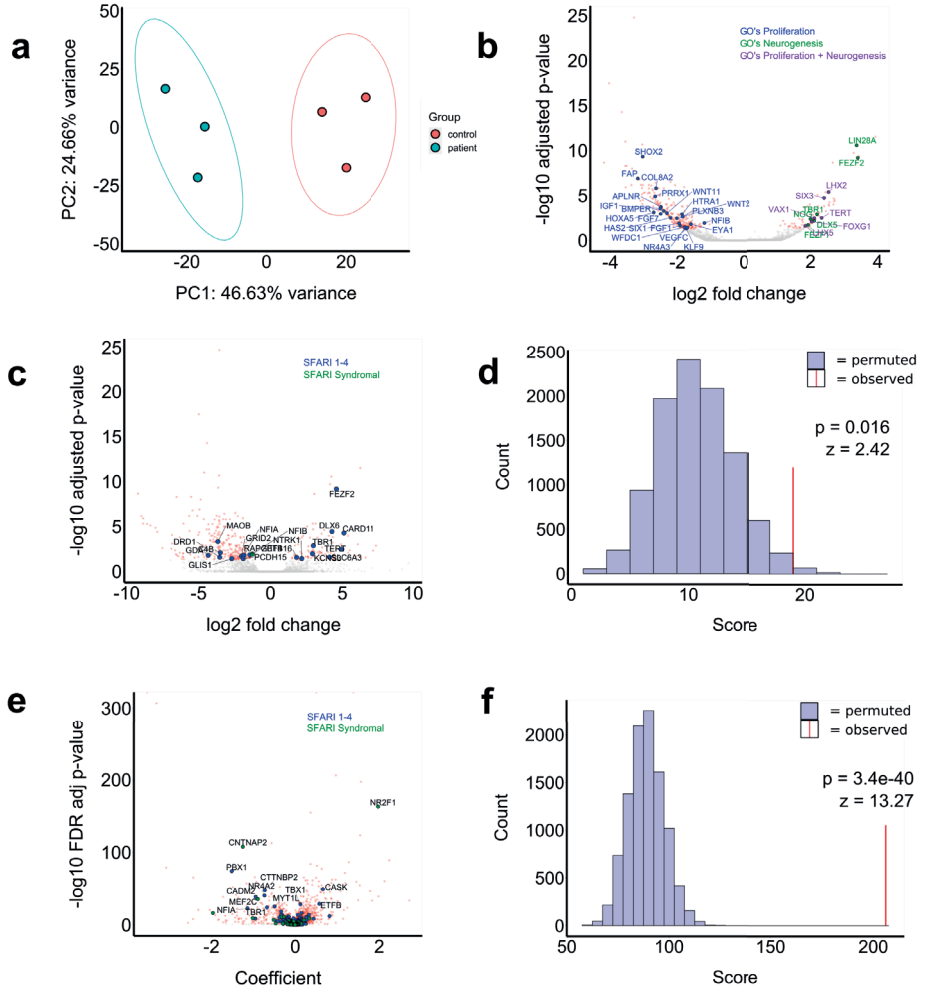


Figure 4. Differentially expressed genes are enriched for GOs related to cell proliferation and neurogenesis, as well as ASD-associated genes.

A. Principal component analysis shows a clear distinction between the expression profiles of patient- compared to control-derived organoids at 8 weeks *in vitro*.

B. Volcano plot showing differentially expressed genes between patient- and control-derived organoids at 8 weeks *in vitro*. Genes reaching statistical significance are colored in red, and genes below statistical significance threshold are colored in grey. A total of 250 genes are downregulated while 89 are upregulated. Genes are highlighted according to their Gene Ontology (GO). GOs related to cell proliferation (blue) (GO:0050673, GO:0050678, GO:0010463, GO:0010464, GO:0050679), neurogenesis (green) and both proliferation and neurogenesis (purple) (GO:2000177, GO:0072091, GO:0061351, GO:1902692, GO:0072089, GO:2000179, GO:0007405, GO:0050768, GO:005076).

(**C and D**) The same volcano plot shown in panel B, but now with ASD-associated genes curated by SFARI enriched among the DE genes highlighted. Blue dots are genes in SFARI categories 1-4, green dots are in category 'syndromic' (C). Statistical significance was tested using a permutation test (n=10,000) showing an enrichment higher than enrichment expected from random events (z= 3.28, p =0.001).

(**E and F**) Volcano plot showing genes that are differentially expressed in three *CNTNAP2*-expressing cell types (EN-V1, EN-PFC and nEN). Genes reaching statistical significance are colored in red, and genes below statistical significance threshold are colored in grey. Blue dots are genes in SFARI categories 1-4, green dots are in category 'syndromic. Of 2,800 differentially expressed genes in these cell types, 215 are present in the SFARI database (E). Enrichment for SFARI genes at the single cell level was more pronounced than in the bulk RNA sequencing analysis as indicated by the bar plot displaying the permutation analysis (n=100,000, z= 13.93, p= 4.1e-44).

S10). The enrichment for the same SFARI genes was more pronounced among differentially expressed genes in these cell types, where 204 out of 2,800 differentially expressed genes were genes present in the SFARI database (z= 13.27, p = 3.42e-40) (Fig. 3E- F and table S11) implying that transcriptional dysregulation of ASD-associated genes occurs in *CNTNAP2*-expressing cell types.

We subsequently used the same approach to test whether the differentially expressed genes discovered by bulk-sequencing were also enriched in three independently created Weighted Gene Co-Expression Networks (WGCNAs) for ASD (45-47). These WGCNAs were constructed based on the correlation of gene-expression levels between genes, and led to the generation of gene co-expression modules that are associated with specific biological or disease processes (48). We found a statistically significant enrichment for all three WGCNAs (Fig. S3F-H), indicating an overlap in transcriptional signature between the genes differentially expressed between patient and control-derived organoids and other forms of ASD. We then compared the differentially expressed genes to gene regulatory network data publicly available for other neuropsychiatric conditions, such as Alzheimer's disease (AD) (49) and Bipolar Disorder (BD) (50), and found no statistically significant enrichment for these WGCNAs (Fig. S3I-J). This suggests that the enrichment of differentially expressed genes in WGCNAs may be specific to ASD and related neurodevelopmental disorders. The enrichment for ASD WGCNAs was more pronounced as well for the DE genes in *CNTNAP2*-expressing cell types (Fig. S3K-M). Disease phenotype GO analysis (51) on the cell type-specific signature genes for each of the three *CNTNAP2*-expressing cell types are associated with ASD, epileptic encephalopathy, Pitt-Hopkins syndrome and hypoplasia of the corpus callosum (Fig. S3N-P). The first three disease phenotypes have all been described in patients with *CNTNAP2* LoF mutations (10, 52) while corpus callosum volumetric changes had not been investigated until now. In order to investigate the latter, we carried out a qualitative comparison of corpus callosum volumetric differences on MRI between the *CNTNAP2* mutation carriers and matched healthy controls without the mutation and found a decrease in corpus callosum volume in patients utilizing a one-sample t-test was performed between the size deviation for the templates (which are 0), and found that mutation carriers indeed demonstrate a decrease in corpus callosum volume (t-test; p=0.0127) (Table S12).

Site-specific repair of c.3709DelG mutation using CRISPR-Cas9 rescues cortical overgrowth phenotypes and transcriptional changes.

To confirm the causal relationship between the homozygous *CNTNAP2* c.3709DelG mutation and the cortical overgrowth disease phenotypes, we used clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 (53) to generate an isogenic 'rescue' line by reintroducing the guanine base at position 3709 in one of the patient-derived hiPSC lines (Fig. S4A). Sanger sequencing of a CRISPR edited single-cell derived clone confirmed the homozygous wildtype

DNA sequence (Fig. 5A). Western blotting confirmed the presence of the CNTNAP2 protein in the organoids derived from the CRISPR-line, at levels similar to those of control-derived organoids (Fig. 5B). We subsequently repeated the experiments by following the same organoid differentiation protocol described above. Organoids generated from the CRISPR-rescue hiPSC line had a 1.5-fold reduction in 2D projected surface area compared to the organoids derived from the unedited patient line (Fig. 5C-D) (t-test, n= 9/line p = 0.0014) – which corresponds with the observed baseline difference in 2D surface area between patient- and control-derived organoids. In addition, the other overgrowth phenotypes, including increased total cell number (t-test, n= 4/line p <0.001) (Fig. 5E) and increased PAX6+ progenitor pool at 6 weeks (t-test,n=4/line, p=0.002) (Fig. 5F), were rescued to comparable numbers observed in the organoids generated from healthy controls. At the bulk-transcriptional level, 576 genes were differentially expressed between organoids derived from the CRISPR-rescue line when compared to its parental patient-derived line. Of the 339 genes that were differentially expressed in patient-organoids, 67 of these were expressed in the opposite direction in the CRISPR-rescue line, indicating a partial rescue (Fig. S4B-C) (permutation analysis, n= 10000, z = 18.6, p = 2.3e-77). These 67 genes are enriched for biological processes related to neurodevelopment including cellular proliferation (Fig. S4D). Regardless of whether reaching the threshold for statistical significance, the gene expression directionality of the 339 DEgenes was generally reversed between patients- versus control lines and CRISPR-rescue versus patients line (Fig. 5G). PCA revealed a large batch effect in sample clustering represented by PC1 and a genotype-rescue effect represented by PC2 (Patient vs. CRISPR-rescue) (Fig. S4E and S4F). Of the 100 genes with the highest contribution to PC1 (86% of the variance) only 3 genes are present among the 339 DEgenes from the initial patient-control batch (permutation analysis; n = 10,000; z =0.99; p=0.32), whereas the top-100 contributing genes for PC2 (6% of the variance) include 41 of these 339 DEgenes (permutation analysis; n = 10,000; z =29.7; p=8.86e-195), indicating the batch effect manifests through genes different from the 339 DEgenes from the initial patient-control batch (Fig. S4G).

DISCUSSION

In this study, we utilized forebrain organoids generated from hiPSCs derived from patients carrying the homozygous c.3709DelG mutation in *CNTNAP2* and healthy controls to investigate the effects of this mutation on cortical embryonic development. We show that *CNTNAP2* is most highly expressed in several types of PFC excitatory neurons. We discovered increased proliferation in PAX6-positive NPCs and other dividing cells, leading to an increase in the generation of total cell number. This increase in number of cortical cells is responsible for a corresponding increase in overall

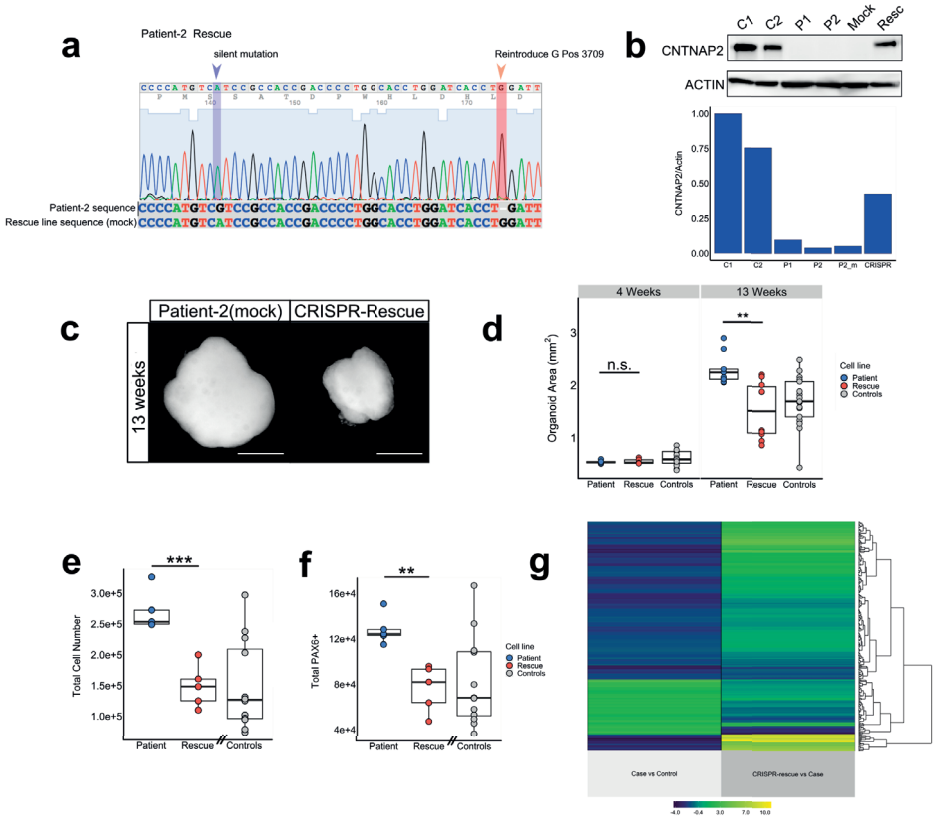


Figure 5. CRISPR-mediated repair of pathogenic c.3709DelG mutation rescues overgrowth phenotypes.

(A and B) Sanger sequencing trace of hiPSC-line derived from patient-2 after CRISPR-Cas9 genome editing. The rescued line shows a guanine base introduced at cDNA position 3709 and a silent blocking mutation at position 3675. Sequence is aligned to original patient-2 DNA sequence (A). Western blotting confirms the presence of CNTNAP2 protein in organoids generated from the CRISPR-rescued line at a similar level as control lines, and absence of the protein in the parental patient-2 cell line-derived organoids (B). Source data are provided as a Source Data file.

(C and D). Representative image of bright field microscopy images of organoids derived from patient-2 and CRISPR-rescued line at 13 weeks *in vitro* (C), and corresponding quantification showing a 1.5-fold decrease in area for the CRISPR rescue line compared to the unedited patient line (D) (two-sided t-test, t(14.1) = 3.93, p = 0.0014, n=9 orgs/ cell-line). Grey dots represent pooled control data points also displayed in figure 2A. Scale bar represents 500 µm. Boxplots in panel D display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

(E and F). Quantifications of total cell number at 6 weeks (t-test, t(7,98)= 5.68, $p < 0.001$, n= 5 orgs/cell-line) (E), total number of PAX6+ cells at 6 weeks(two-sided t-test, t(6,87)= 4.68, p= 0.002, n= 5 orgs/cell-line) (F). Total cell number, and total number of PAX6 + cells are decreased in the CRISPR rescue-line compared to the parental patient line, confirming the causal effect of the single mutation on the observed phenotypes. Grey dots represent pooled control data points also displayed in figure 3A. Boxplots display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

G. Heatmap displaying the log2 fold-change of the 339 genes differentially expressed between control- and patient-organoids (left column) and between the CRISPR-rescue line versus its parental patient-derived control (right column). The inversion of the expression direction for the majority of the 339 genes highlights that the DE gene expression-level is largely rescued in the CRISPR line derived organoids.

organoid volume in patient-derived organoids. GO analysis from both bulk RNAseq- and scRNAseq analysis corroborates abnormal cellular proliferation and differentiation. We further show that there is an enrichment of these differentially expressed genes for ASD-associated genes and WGCNAs associated with ASD and other neurodevelopmental disorders. The cell type-specific signature genes expressed in the affected neuronal types are associated with clinical disease phenotypes that have been described in patients with *CNTNAP2* LoF mutations. Finally, by repairing the pathogenic mutation using CRISPR-Cas9, we were able to rescue these cortical overgrowth phenotypes, thereby confirming a causative effect of the homozygous LoF mutation in the context of an identical genetic background.

The main findings from this study confirm that the homozygous c.3709DelG mutation in *CNTNAP2* leads to abnormal brain development in processes analogous to those occurring during the first trimester of embryonic development. Thus, the findings support the prevailing hypothesis of an early embryonic origin of the pathophysiology of ASD (54). This increase in total cell number could be related to the increased cortical thickening seen in ASD, as was also observed on MRI of patients with *CNTNAP2*-associated ASD. More recently, other studies deploying hiPSC-derived neuronal culture techniques have also reported increased NPC proliferation and aberrant neurogenesis in patients with ASD. This includes both monogenic forms of ASD (55) and idiopathic ASD (8, 56) – in both 2D and 3D neuronal cultures. These findings could be causally related to two clinical phenomena observed in patients with ASD that may have at least a partial overlapping etiology: early brain overgrowth and an increase in head circumference or macrocephaly (defined as head circumference > 97th percentile). A recent systematic review and meta-analysis that analyzed 27 studies reported macrocephaly in 15.7% of ASD cases 3% in controls (57). The same study also reported an increase in total brain volume (defined as brain volume 2 standard deviations above the mean) in 9% of patients with ASD.

Of note, the overgrowth phenotype described in this study has not been recapitulated in *Cntnap2* null mice(16, 58). One study quantified the volume of both medial prefrontal and somatosensory cortices, and found no difference between null- and wildtype mice (58). Thus, although mouse models can capture some of the clinical features of a human condition such as ASD, important aspects of such a disease may be missed since these may exclusively be present in humans. Given the fact that one of the defining features of the human brain compared to other mammals is its expanded gyrencephalic neocortex (59) that is thought to give rise to our abilities of higher order brain processes, such as cognition and language, it is not entirely surprising that a truncation of a protein with a critical role in cortical development could have different phenotypic outcomes in different mammalian species. Species-specific differences in the role that *CNTNAP2* plays in brain development is reflected in its differential expression pattern as well. This was shown using *in situ hybridization*, revealing that gene expression is

more diffusely distributed throughout the entire brain in rodents while there is a more focal expression in humans with a preferential localization in prefrontal cortex (42).

The findings from this study lead to interesting questions regarding the development of novel treatments for neurodevelopmental disorders, such as *CNTNAP2*-associated ASD. It is critical to determine the extent to which pathophysiological processes that occur as early as in the first trimester of pregnancy remain reversible after birth. Are the cortical overgrowth phenotypes critical to cognitive- and behavioral deficits seen in these patients, and if so, during which time window should treatments aim to target these early developmental processes? In this study, we show that the field of brain organoid modeling holds great promise for working towards a preclinical modeling system with face validity, with the ultimate goal of creating targeted treatments that may eventually improve the lives of patients with severe neurodevelopmental disorders such as *CNTNAP2*-associated ASD.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Yuanjia Wang and Chen Chen for their advice on the statistical analyses using the linear mixed model. We thank Dr. Maura Boldrini for her advice on the BrdU labeling experiments. We thank Drs. Alexander Sosunov, Andrew Dwork and Gorazd Rosoklija for their advice on histology experiments. We thank Joseph Sall from NYU Langone's Microscopy Laboratory (supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center) for technical assistance with the light-sheet microscopy image acquisition. We thank Luis Aparicio de Santiago for his contribution to the scRNAseq analysis. We thank Julie Shabto, Qanetha Ahmed, Yelizaveta Gribkova and Huber Rodriguez-Tejada for their contribution to Western Blotting and imaging experiments. We thank Dr. Theresa Swayne and Dr. Laura Munteanu for their advice on image processing and analysis at the Confocal and Specialized Microscopy Shared Resource of the Herbert Irving Comprehensive Cancer Center at Columbia University, supported by NIH grant #P30 CA013696 (National Cancer Institute). This research was partially supported by Irving Institute/CTO Pilot Award (UR007953) to B.X.; B.X. and S.M. are supported by NCAT UG3 NS115598.

AUTHOR CONTRIBUTIONS

S.M. and B.X. conceived the research project, J.O.J.,M.G., H.Z., C.L., Y.S., K.S. and G.C. performed experiments and collected data, J.O.J., B.X., B.M.,M.G., F.P. and G.C. analyzed data, B.C. reprogrammed patients' samples, S.M., K.S. and K.B. were involved in the recruit-

ment of the patients. S.M., B.X., C.K., J.A.J., S.K., J.O.J contributed to design of the research project and discussion of the data. J.O.J., S.M., B.X., J.A.J. and C.K. wrote the manuscript with input from the other authors.

DATA AVAILABILITY

All data are available from the authors upon request.

Bulk RNA sequencing transcriptome data are available under accession code:
GSE174569.

Publicly available data used in this paper includes:

- Brainspan transcriptome database (Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014;508(7495):199-206.) www.brainspan.org
- human embryonic PFC single cell transcriptome (Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science. 2017;358(6368):1318-23.) www.sciencemag.org/content/358/6368/1318/suppl/DC1

Figures associated with available raw data in the source data file:
Figure 1E, 1F, 4A, 4B, 4C, 4D, 4E, 4F, 5B, 5G, S1J, S3A-P, S4B-G

In the other figures, all data points have been plotted individually.

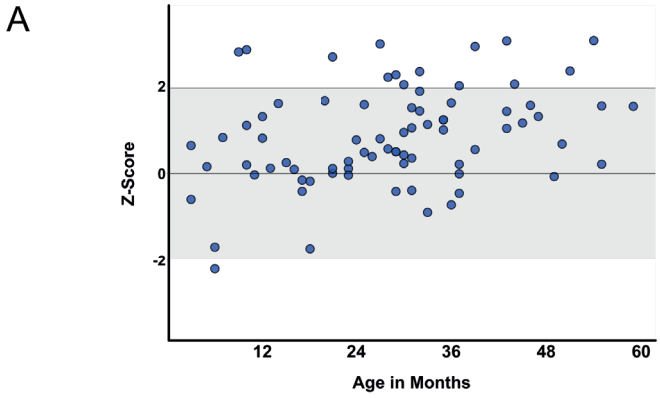
CODE AVAILABILITY

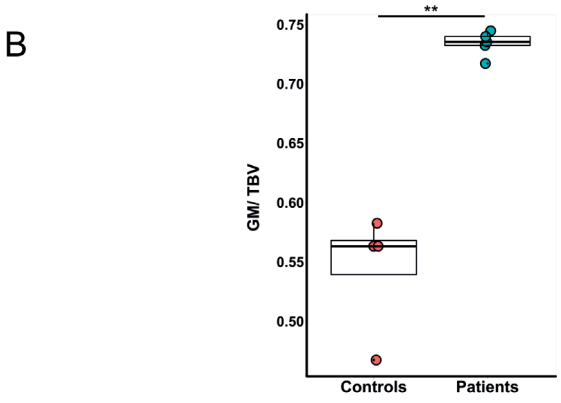
All code used is available from the authors upon request.

COMPETING INTERESTS

The authors declare no competing interests.

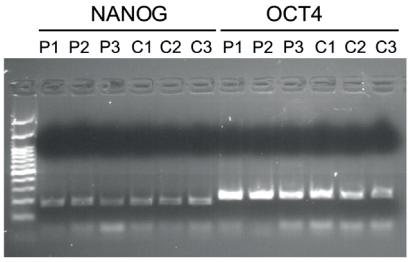
Supplementary figure 1.



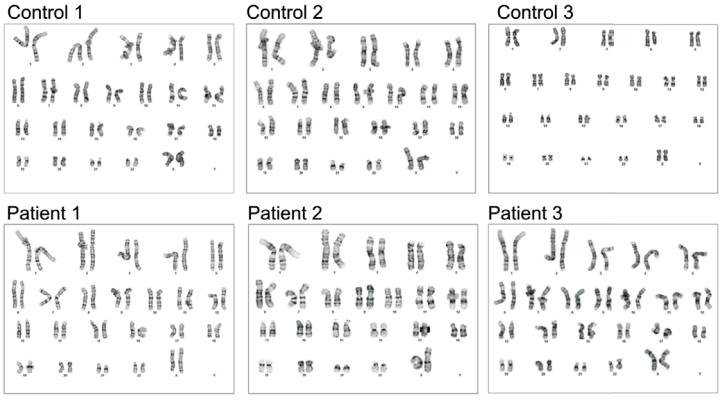


Supplementary figure 1. (continued)

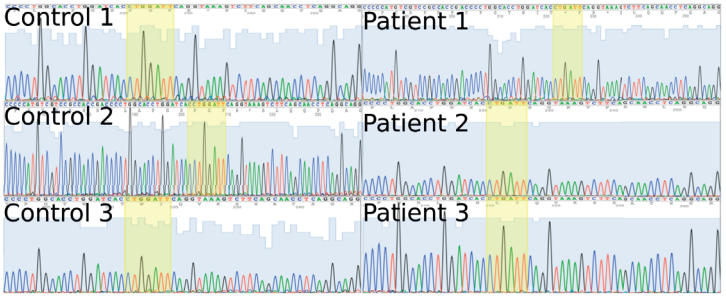
C



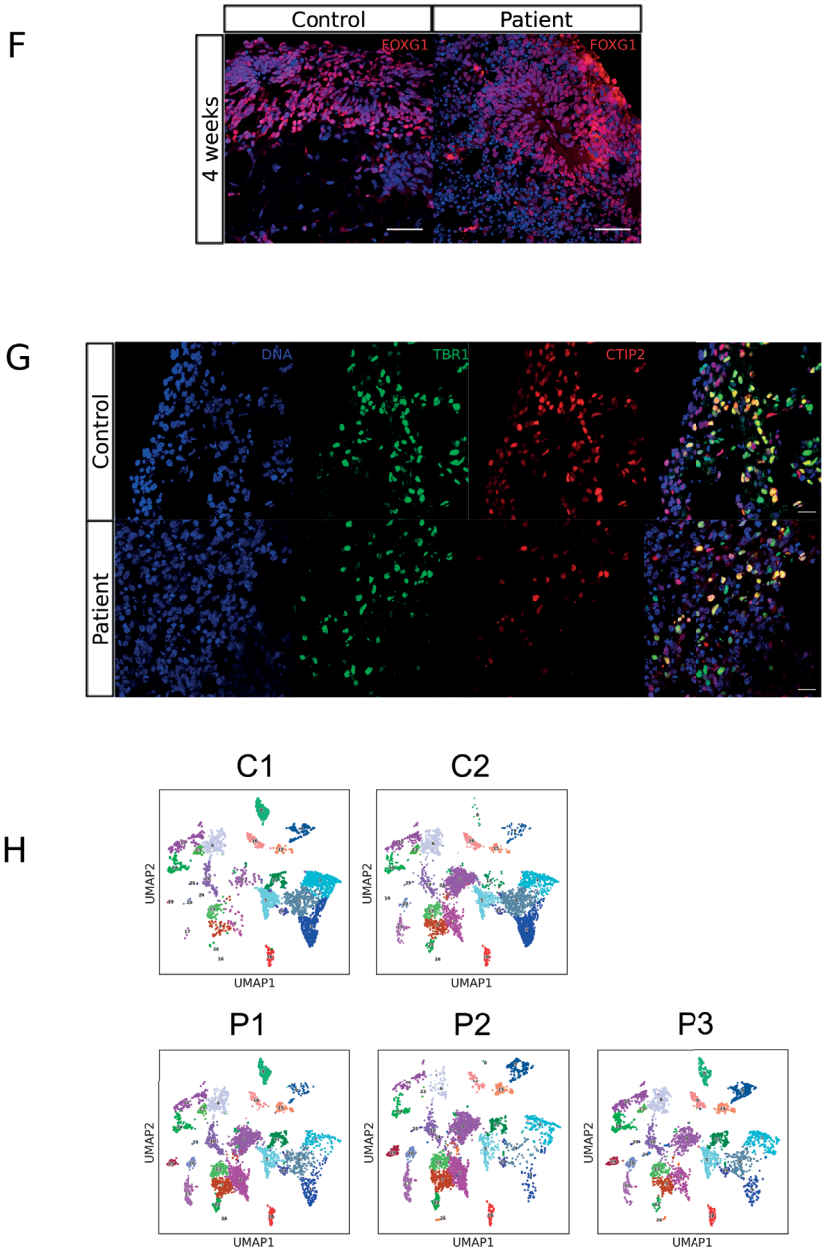
D



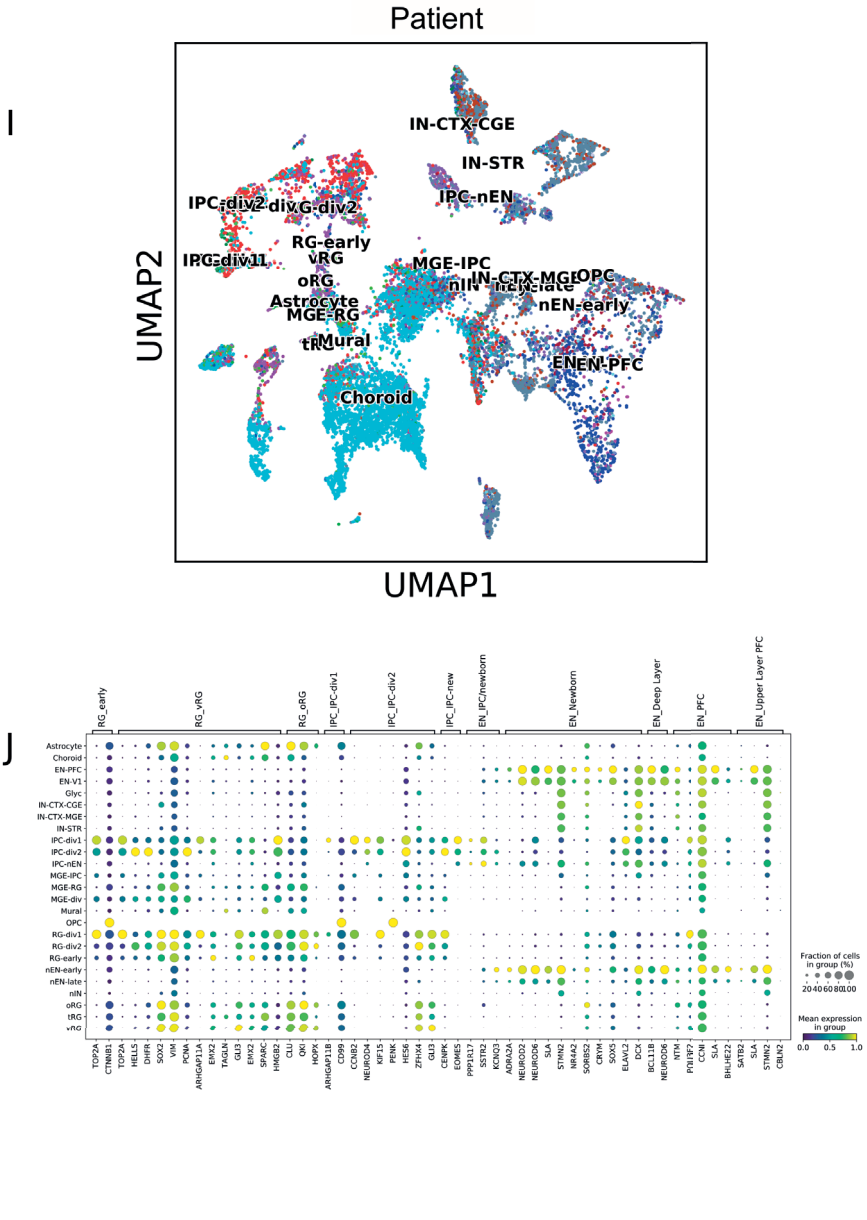
E



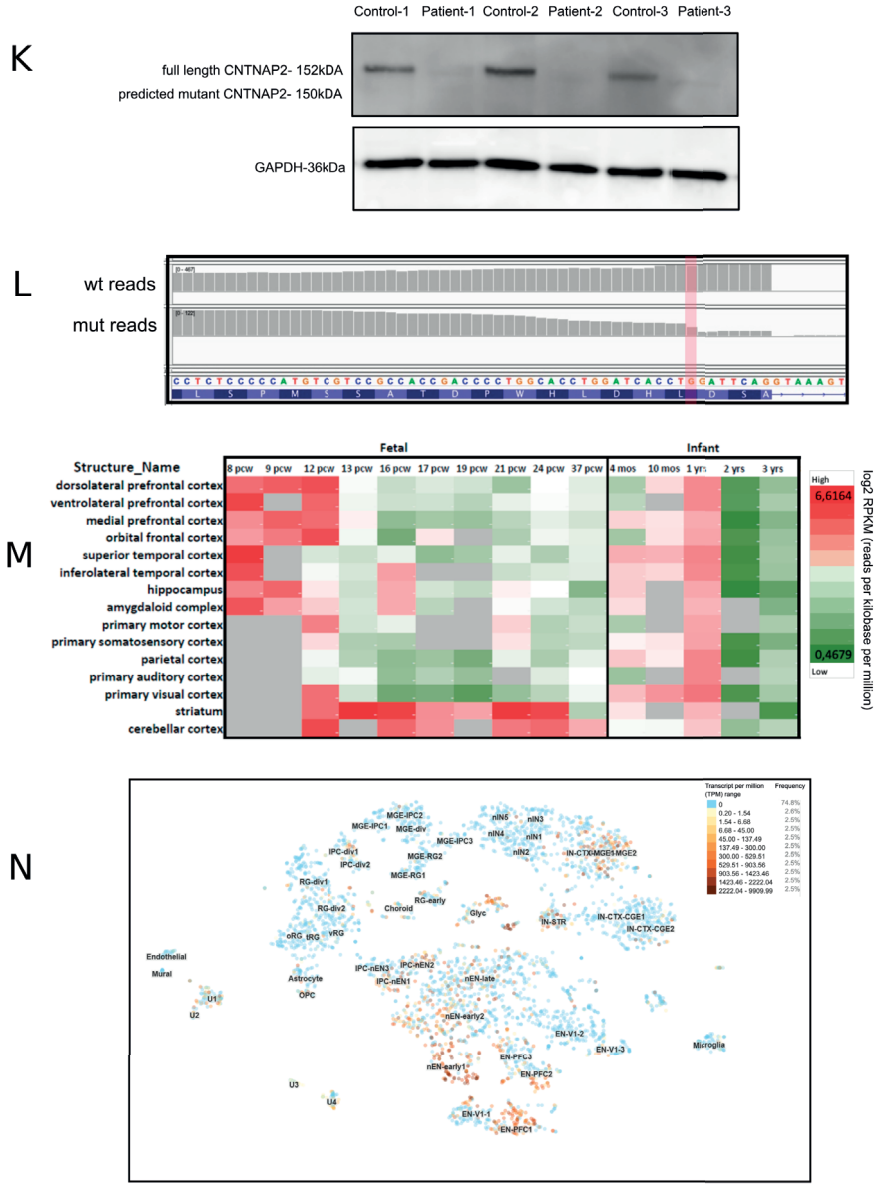
Supplementary figure 1. (continued)



Supplementary figure 1. (continued)



Supplementary figure 1. (continued)



Supplementary figure 1. (continued)

A. Longitudinal head circumference measurements were obtained from 20 unique female and 17 unique male pediatric patients who were between birth and 5 years old, all from the Old-Order Amish community, diagnosed with ASD and carrying the c.3709DelG mutation in *CNTNAP2*. Z-scores were calculated using average head circumference measurements per month and gender and mean and SD from WHO-reference data (weighted z-score = 2.93, p <0.003).

B. Quantification of total gray matter (GM) relative to total brain volume (TBV) of 6 patients carrying the homozygous c.3709DelG mutation in *CNTNAP2* compared to 4 age-matched control templates, each of which is based on a composite of MRIs from healthy individuals for that given age. (Patients: GM/TBV= 0.73; Controls: GM/TBV= 0.54;two-sided t-test, p = 0.004, n= 6 patients, 4 control templates) Boxplots display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

C. qRT-PCR shows the presence of stemness markers NANOG and OCT4 in all iPSC lines. This experiment was not repeated independently.

D. Karyotyping analysis of all patient- and control-derived hiPSC lines indicates that the iPSCs derived from each subject have normal karyotype.

E. Sanger sequencing analysis of the genomic fragment surrounding the mutation site confirms all iPSC lines have the correct *CNTNAP2* genotypes. All three patient-derived iPSCs show a single guanine deletion at cDNA position 3709, the three control lines show 2 guanines at the same position. Three surrounding base pairs upstream and downstream of mutation site marked in yellow.

F. Examples of immunohistochemistry images of 4-week-old cortical organoid sections showing expression of dorsal forebrain specific transcription factor *FOXG1* in both control- and patient-derived organoids. All control- and patient-derived organoids displayed similar expression patterns for this marker. Scale bar represents 50 µm.

G. Representative examples of immunohistochemistry images of 13 week-old cortical organoid sections showing expression of early born neuron layer markers TBR1 and CTIP2. All control- and patient-derived organoids displayed similar expression patterns for these markers. Scale bar represents 20 µm.

H. UMAP plots displaying the cell clusters determined by the unsupervised Leiden clustering method (see methods for details) for all individual control- and patient-derived organoids, showing that all 26 clusters are consistently present in all individual samples.

I. UMAP plot of cell lineage distribution annotated according to Nowakowski et al. (2017)(ref. 37) in patient-derived brain organoids (n=3 samples, 16,780 cells).

J. Dot plot displaying expression levels of signature markers expressed annotated cell types, present in the control-derived forebrain organoids at 13 weeks *in vitro*. Color of dots represents mean gene expression level, size of dots represent fraction of cells in cell type expressing the gene.

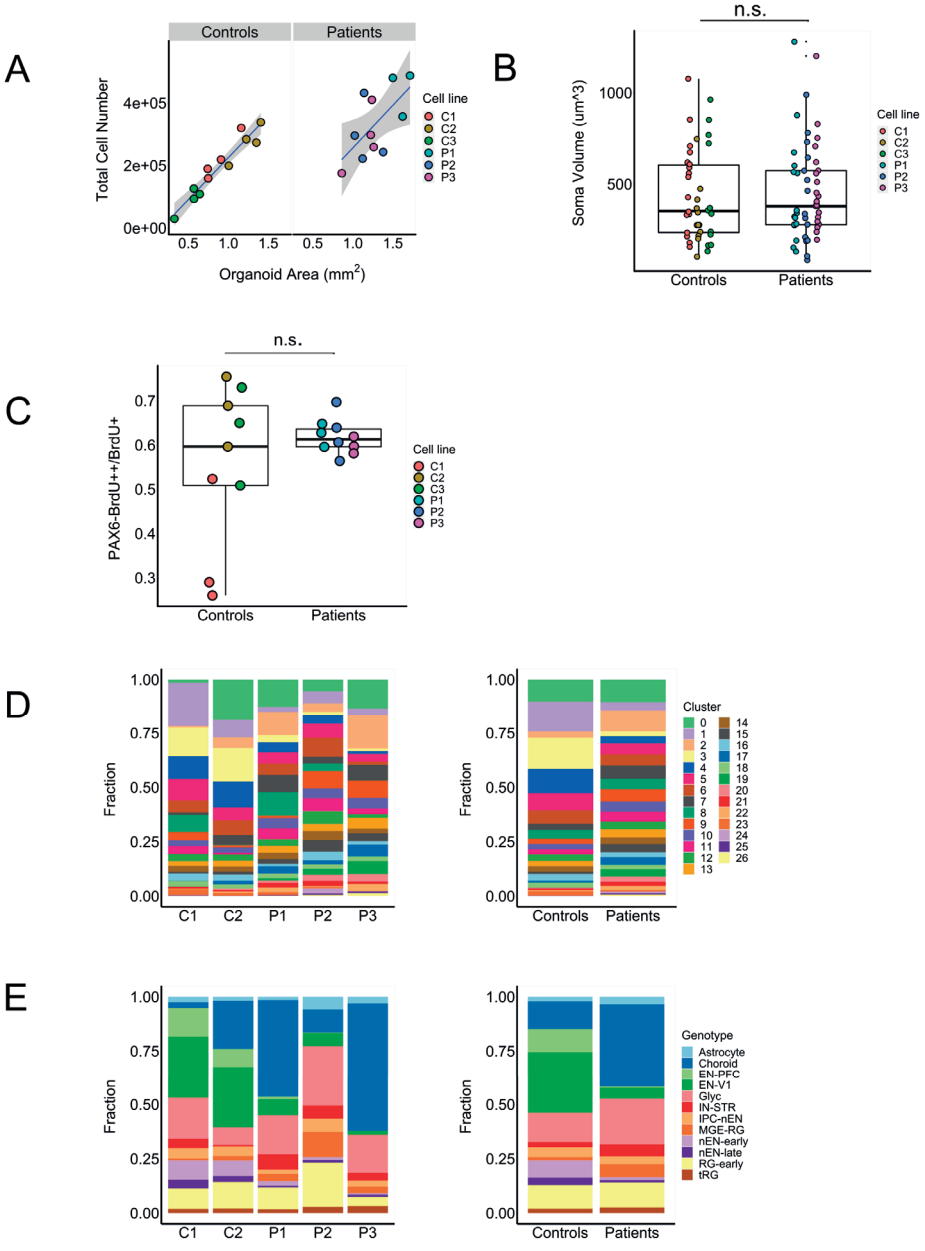
K. Western blot analysis showing presence of full length CNTNAP2 protein levels in all control-derived organoids and absence in all patient-derived cortical organoids using an antibody directed against the last Laminin G domain positioned before the transmembrane domain (Antibody B in Fig. 1A), indicating there is no truncated protein (predicted size ~150kDa) present in patient-derived organoids. This experiment was repeated independently and generated similar results.

L. Aligned mRNA sequence reads from scRNAseq of control- and patient derived organoids, showing lower read coverage for the patient-derived organoid at the 3'-end of the mRNA, suggesting that a nonsense-mediated decay occurs in *CNTNAP2* mRNA in patient organoids.

M. Heat map displaying *CNTNAP2* expression levels from human embryonic RNA sequencing data that are publicly available at www.brainspan.com (ref 28). Expression is highest in multiple cortical regions, amygdala and hippocampus in the first 12 weeks post conception, and in striatum and cerebellum between 12- and 37 weeks post conception.

N. t-SNE plot obtained from https://cells.ucsc.edu/?ds=cortex-dev showing the *CNTNAP2* mRNA levels in all cell lineages in developing human telencephalon, as annotated by Nowakowski *et al* (2017) (ref. 37). *CNTNAP2* is highly expressed in different types of excitatory and inhibitory neurons, which is in line with the observation in our brain organoids.

Supplementary figure 2.



Supplementary figure 2. (continued)

A. Plots of linear relationships between organoid projected surface area and total organoid cell number for both control- and patient-derived organoids at 13 weeks *in vitro*. Shaded grey area is 95% confidence interval estimate around the mean. (Linear model, Controls: $F(1,10)= 10.64$, $R^2= 0.52$, $p = 0.0085$. Patients: $F(1,9)=2.31$, $R^2= 0.34$, $p =0.16$, n= 3 cell lines, C1-C2-C3-P2-P3: n=4 orgs/cell line P1:n=3 orgs/cell_line). The shaded grey area represents the constructed 95% confidence interval around the mean.

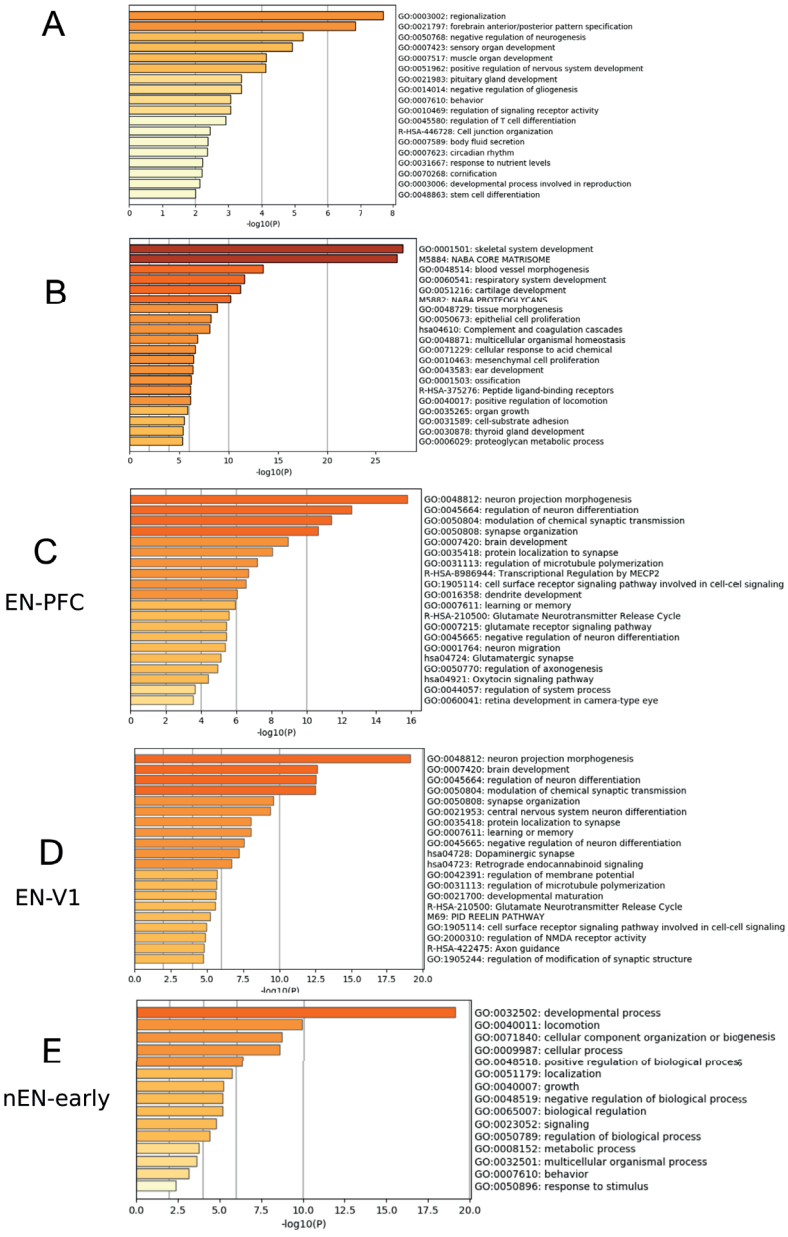
B. Quantification of neuronal soma from organoids infected with AAV-Synapsin-GFP at 40 weeks showing no difference in neuronal somatic volume between control- and patient-derived cortical organoids (LRT, $\chi^2(1)= 0.16$, $p = 0.69$, n= 3 cell lines, 3 orgs/cell line, 8 neuronal soma/org). Boxplots display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

C. Quantification of BrdU + cells that stained positive for PAX6 as well at 6 weeks *in vitro*, showing no difference between genotype groups (LRT, $\chi^2(1)= 0.54$, $p = 0.46$). Boxplots display median, first and third quartiles, and whiskers showing the largest and smallest values no further than 1.5 times the inter quartile range from first and third quartile, respectively. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

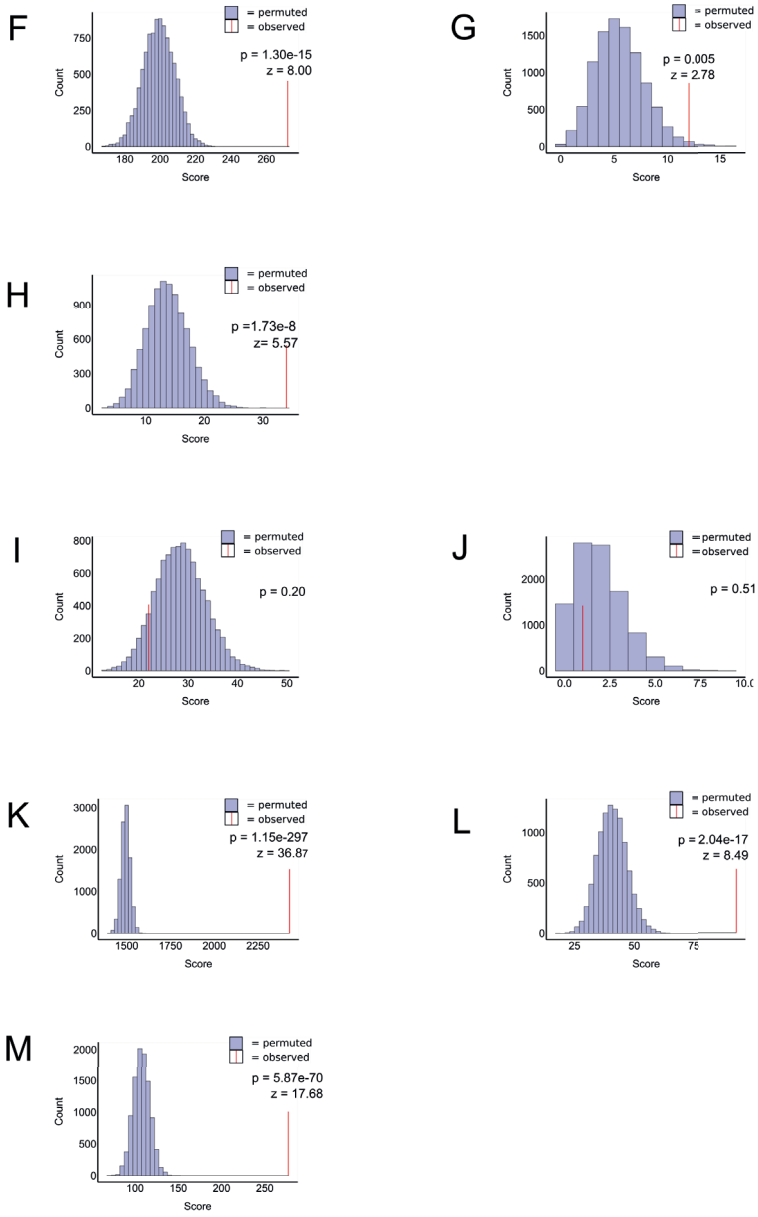
D. Stacked bar plot displaying relative cell fractions in each unsupervised Leiden cluster in patient (P1-3) and control (C1-C2) organoids as determined using scRNAseq for individual cell lines (left panel) and pooled per genotype (right panel).

E. Stacked bar plot displaying relative cell type fractions in patient (P1-3) and control (C1-C2) organoids as determined using scRNAseq and annotated using the cell types identified by Nowakowski et al. (ref 37) for individual cell lines (left panel) and pooled per genotype (right panel).

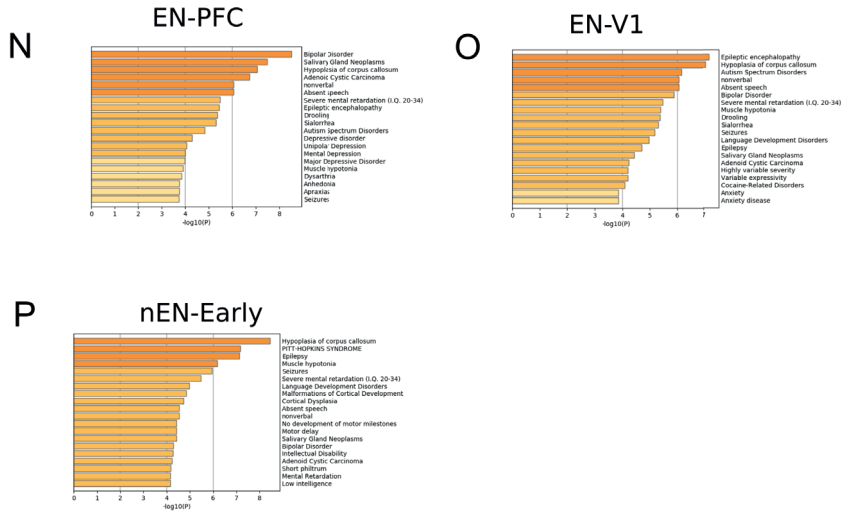
Supplementary figure 3.



Supplementary figure 3. (continued)



Supplementary figure 3. (continued)



A and B. Gene Ontology (GO) analysis of differentially expressed (DE) genes in patient organoids determined by bulk RNA sequencing. The results show the biological process ontologies sorted by the smallest FDR p-value. Panel A shows GO analysis of differentially upregulated genes, panel B shows GO analysis downregulated genes.

C - E. Gene Ontology (GO) analysis of the top-100 most highly expressed genes in cell types with high CNTNAP2 expression in the control organoids: EN-PFC (C), EN-V1 (D), nEN-early (E) as annotated by single cell RNA sequencing.

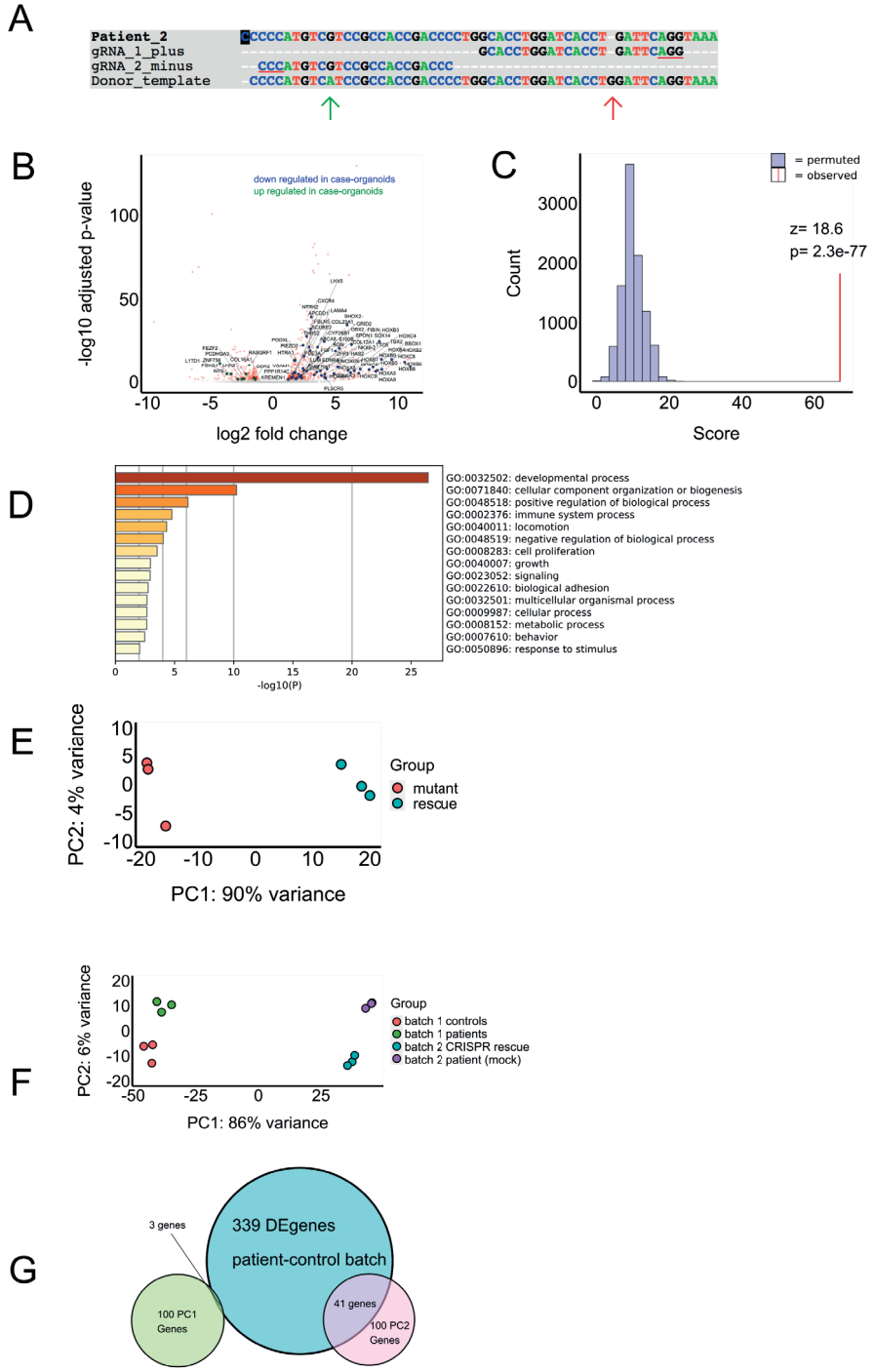
F - H. Histograms showing permutation analyses of 10000 iterations of randomly generated gene lists of equal size to gene list of DEgenes from bulk RNAseq, and ASD-related WGCNA from Parikshak, 2013 (ref. 46) (F), Willsey, 2013 (ref. 45) (G), Duda, 2018 (ref. 47) (H).Displayed are z-scores and two-sided p-values of the permutation tests.

I and J. Histograms showing permutation analyses of 10000 iterations of randomly generated gene lists of equal size to gene list of differentially expressed genes from the bulk RNAseq, compared to WGCNAs for AD (ref. 49) (I) and BD (ref.50) (J). Displayed are z-scores and two-sided p-values of the permutation tests.

K - M. Histograms showing permutation analyses of 10000 iterations of randomly generated gene lists of equal size to a list of DE genes from single-cell RNAseq, and ASD-related WGCNA from Parikshak, 2013 (ref. 46) (K), Willsey, 2013 (ref. 45) (L), Duda, 2018 (ref. 47) (M). Enrichment for these DEgenes at the single-cell level is more pronounced than at the bulk RNAsequencing level shown in panel F-H. Displayed are z-scores and two-sided p-values of the permutation tests.

N - P. Enrichment of the 100 cell type differentiating genes from each *CNTNAP2*-expressing cell types for disease phenotypes as curated by DisGeNET (ref. 49). The results are sorted by the smallest FDR p-value.

Supplementary figure 4.



Supplementary figure 4. (continued)

A. CRISPR design showing the genomic location of the mutant DNA sequence of patient 2 and the targets of designed gRNAs. The Protospacer adjacent motif (PAM) sequence is underlined in red. The donor template has a G base inserted at cDNA position 3709 (red arrow) and a silent blocking mutation 6 base-pairs downstream from the gRNA2 PAM sequence (green arrow).

B. Volcano plot displaying the differentially expressed genes between CRISPR-rescue organoids and its parental patient-derived organoid line at 8 weeks *in vitro.* DE genes reaching statistical significance are colored in shaded red, and genes below statistical significance threshold are colored in shaded grey. 386 genes were upregulated while 190 genes were downregulated. Genes that were downregulated in case organoids but upregulated in the CRISPR rescue line (57 genes) are indicated as blue dots, genes that were upregulated in case organoids but are downregulated in the CRISPR-rescue line (10 genes) are indicated as green dots, indicating a partial rescue of the transcriptional profile in the CRISPR rescue line.

C. Permutation analysis (n=10,000) indicating that the partial rescue at the transcriptional level is many times more as would be expected from random events.

D. Gene Ontology (GO) analysis on the 67 genes that are highlighted in panel B of this figure. The results show the biological process ontologies sorted by the smallest FDR p-value, and implicate multiple processes related to neurodevelopment including cellular proliferation.

E. Principal Component Analysis (PCA) shows a clear distinction between the expression profiles of CRISPR-rescue compared to it parental patient-derived organoids at 8 weeks *in vitro*.

F. PCA analysis showing that genotypes (Patient vs control and CRISPR-rescue vs patient) separate in a similar manner on PC2, which explains 6% of the variance. This analysis shows a large batch effect, with the two batches being clearly separated on the axis representing PC1 that explains 86% of the variance.

G. Circle diagram displaying that of the 339 genes differential expressed between patient- and control-organoids (Fig. 4A), 41 are present in the top-100 contributing genes to PC2 of panel F of this figure, where only 3 of the 339 DE genes are present in the top-100 contributing genes to PC1. This underscores that the separation on the horizontal axis is a batch effect that involves different genes than the 339 DE genes from the control-patient batch.

REFERENCES

1. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. Psychol Med. 2015;45(3):601-13.
2. Kaiser T, Feng G. Modeling psychiatric disorders for developing effective treatments. Nat Med. 2015;21(9):979-88.
3. Herculano-Houzel S, Mota B, Lent R. Cellular scaling rules for rodent brains. Proc Natl Acad Sci U S A. 2006;103(32):12138-43.
4. Hill RS, Walsh CA. Molecular insights into human brain evolution. Nature. 2005;437(7055):64-7.
5. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. Conserved cell types with divergent features in human versus mouse cortex. Nature. 2019;573(7772):61-8.
6. Bershteyn M, Nowakowski TJ, Pollen AA, Di Lullo E, Nene A, Wynshaw-Boris A, et al. Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia. Cell Stem Cell. 2017;20(4):435-49 e4.
7. Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, et al. Cerebral organoids model human brain development and microcephaly. Nature. 2013;501(7467):373-9.
8. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L, et al. FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. Cell. 2015;162(2):375-90.
9. Iefremova V, Manikakis G, Krefft O, Jabali A, Weynans K, Wilkens R, et al. An Organoid-Based Model of Cortical Development Identifies Non-Cell-Autonomous Defects in Wnt Signaling Contributing to Miller-Dieker Syndrome. Cell Rep. 2017;19(1):50-9.
10. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, et al. Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. N Engl J Med. 2006;354(13):1370-7.
11. Riccardi F, Urquhart J, McCullagh G, Lawrence P, Douzgou S. A patient with a novel CNTNAP2 homozygous variant: further delineation of the CASPR2 deficiency syndrome and review of the literature. Clin Dysmorphol. 2019;28(2):66-70.
12. Saint-Martin M, Joubert B, Pellier-Monnin V, Pascual O, Noraz N, Honnorat J. Contactin-associated protein-like 2, a protein of the neurexin family involved in several human diseases. Eur J Neurosci. 2018;48(3):1906-23.
13. Horresh I, Bar V, Kissil JL, Peles E. Organization of myelinated axons by Caspr and Caspr2 requires the cytoskeletal adapter protein 4.1B. J Neurosci. 2010;30(7):2480-9.
14. Anderson GR, Galfin T, Xu W, Aoto J, Malenka RC, Sudhof TC. 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. 2012;109(44):18120-5.
15. Varea O, Martin-de-Saavedra MD, Kopeikina KJ, Schurmann B, Fleming HJ, Fawcett-Patel JM, et al. Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. Proc Natl Acad Sci U S A. 2015;112(19):6176-81.
16. Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell. 2011;147(1):235-46.
17. Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet. 2008;82(1):150-9.

18. Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Brauninger M, et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. Proc Natl Acad Sci U S A. 2015;112(51):15672-7.
19. Velasco S, Kedaigle AJ, Simmons SK, Nash A, Rocha M, Quadrato G, et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. Nature. 2019.
20. Nooner KB, Colcombe SJ, Tobe RH, Mennes M, Benedict MM, Moreno AL, et al. The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. Front Neurosci. 2012;6:152.
21. Shi F, Yap PT, Wu G, Jia H, Gilmore JH, Lin W, et al. Infant brain atlases from neonates to 1- and 2-year-olds. PLoS One. 2011;6(4):e18746.
22. Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad Ser B Phys Biol Sci. 2009;85(8):348-62.
23. Riera M, Patel A, Bures-Jelstrup A, Corcostegui B, Chang S, Pomares E, et al. Generation of two iPS cell lines (FRIMOi003-A and FRIMOi004-A) derived from Stargardt patients carrying ABCA4 compound heterozygous mutations. Stem Cell Res. 2019;36:101389.
24. Kadoshima T, Sakaguchi H, Nakano T, Soen M, Ando S, Eiraku M, et al. 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. 2013;110(50):20284-9.
25. Zhu H, Misel L, Graham M, Robinson ML, Liang C. CT-Finder: A Web Service for CRISPR Optimal Target Prediction and Visualization. Sci Rep. 2016;6:25516.
26. Byrne SM, Church GM. Crispr-mediated Gene Targeting of Human Induced Pluripotent Stem Cells. Curr Protoc Stem Cell Biol. 2015;35:5A 8 1-22.
27. Herculano-Houzel S, Lent R. Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. J Neurosci. 2005;25(10):2518-21.
28. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014;508(7495):199-206.
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
30. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.
31. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006;7(10):R100.
32. (2020). RCT. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
33. Bates D MM, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015(67(1)):1–48.
34. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15.
35. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. Cell. 2019;177(7):1873-87 e17.
36. McInnes LH, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxivorg/abs/180203426 2018.
37. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, et al. Spatio-temporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science. 2017;358(6368):1318-23.

38. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163-72.
39. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9(1):5233.
40. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015;16:278.
41. Group WMGRS. Growth Standards: Growth velocity based on weight, length and head circumference: Methods and development. Geneva: World Health Organization 2009.
42. Abrahams BS, Tentler D, Perederiy JV, Oldham MC, Coppola G, Geschwind DH. Genome-wide analyses of human perisylvian cerebral cortical patterning. Proc Natl Acad Sci U S A. 2007;104(45):17849-54.
43. Bhaduri A, Andrews MG, Mancia Leon W, Jung D, Shin D, Allen D, et al. Cell stress in cortical organoids impairs molecular subtype specification. Nature. 2020;578(7793):142-8.
44. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. Nucleic Acids Res. 2009;37(Database issue):D832-6.
45. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013;155(5):997-1007.
46. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013;155(5):1008-21.
47. Duda M, Zhang H, Li HD, Wall DP, Burmeister M, Guan Y. Brain-specific functional relationship networks inform autism spectrum disorder gene prediction. Transl Psychiatry. 2018;8(1):56.
48. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
49. Miller JA, Oldham MC, Geschwind DH. A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. J Neurosci. 2008;28(6):1410-20.
50. Chen H, Wang N, Zhao X, Ross CA, O'Shea KS, McInnis MG. Gene expression alterations in bipolar disorder postmortem brains. Bipolar Disord. 2013;15(2):177-87.
51. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017;45(D1):D833-D9.
52. Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, et al. 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. 2009;85(5):655-66.
53. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154(6):1380-9.
54. Courchesne E, Pramparo T, Gazestani VH, Lombardo MV, Pierce K, Lewis NE. The ASD Living Biology: from cell proliferation to clinical phenotype. Mol Psychiatry. 2019;24(1):88-107.
55. Li Y, Muffat J, Omer A, Bosch I, Lancaster MA, Sur M, et al. Induction of Expansion and Folding in Human Cerebral Organoids. Cell Stem Cell. 2017;20(3):385-96 e3.
56. Marchetto MC, Belinson H, Tian Y, Freitas BC, Fu C, Vadodaria K, et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. Mol Psychiatry. 2017;22(6):820-35.
57. Sacco R, Gabriele S, Persico AM. Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. Psychiatry Res. 2015;234(2):239-51.

58. Lauber E, Filice F, Schwaller B. Dysregulation of Parvalbumin Expression in the Cntnap2-/- Mouse Model of Autism Spectrum Disorder. Front Mol Neurosci. 2018;11:262.
59. Molnar Z, Clowry G. Cerebral cortical development in rodents and primates. Prog Brain Res. 2012;195:45-70.

Chapter 4

Generating a high-throughput drug screening platform for 22q11.2 deletion syndrome using mouse embryonic stem cell-derived neuronal cultures from *Df(16)A*$^{+/-}$ mice.

Giuseppe Cortese[1], Francesco Brundu[2], Job de Jong[1], Joseph Gogos[3], Sander Markx[1]

Affiliations
1. Department of Psychiatry, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA.
2. Department of Systems Biology, Columbia University, New York, NY, USA
3. Departments of Neuroscience and Physiology, Columbia University, New York, NY, USA

Manuscript in preparation

ABSTRACT

Pharmacological treatment for schizophrenia (SCZ) has remained largely unchanged since the second half of the last century, is only partially efficacious and causes many unwanted side-effects. Therefore, there is a pressing need for the development of novel medications that can target the full range of symptoms that accompany a diagnosis of schizophrenia.

In this study, we used *Df(16)A*$^{+/-}$ mice to develop a mouse embryonic stem cell (mESC)-derived neuronal model with the goal of generating a high-throughput assessment system that can be used to screen for novel compounds for 22q11.2DS-associated SCZ. mESC-derived neurons from both genotype groups display characteristics of electrophysiological maturity. Single cell RNA sequencing (scRNAseq) revealed neuronal cultures consisting of multiple mouse CNS cell types including excitatory neurons, inhibitory neurons and neuroblasts. *Df(16)A*$^{+/-}$ mESC-derived cortical neurons have significant reductions in the number of dendrite branchpoints and total dendrite length, recapitulating the phenotype seen in the *Df(16)A*$^{+/-}$ mice, providing a useful cell-morphological read-out to use for drug screening.

INTRODUCTION

Pharmacological treatment for schizophrenia (SCZ) has remained largely unchanged since the second half of the last century, with first and second generation antipsychotics remaining the mainstay pharmacotherapy. Unfortunately, these medications mainly target the positive symptoms of the disorder and cause unwanted side effects for many patients (1). The negative symptoms of SCZ are more difficult to treat and are more functionally impairing compared to positive symptoms (2). Therefore, there is a pressing need for the development of novel medications that can target the full range of symptoms that accompany a diagnosis of schizophrenia.

About 2% of SCZ cases are due to a chromosomal microdeletion at the q11.2 locus on human chromosome 22 (3). Around 30% of microdeletion carriers display psychiatric symptoms that meet the DSM-V diagnostic criteria for schizophrenia, which is about 30 times higher than in the general population (4). The 22q11.2 locus contains a set of genes that are highly expressed during early fetal brain development (5). Therefore, the 22q11.2 microdeletion provides unique research opportunities to model early-stage neurodevelopmental abnormalities important for SCZ pathogenesis.

In recent years, the advent of using embryonic stem cell- and patient-derived induced pluripotent stem cell technology and the possibility to differentiate these into 2D and 3D neuronal cultures, has created new opportunities to study neurodevelopmental aspects of SCZ pathophysiology (6). Moreover, these cultures can be used for drug screening purposes (7).

Extensive work by our group and others have established that the 22q11.2 mouse model, *Df(16)A*$^{+/-}$ mice, resemble core SCZ cognitive phenotypes, including deficits in attention and working memory (8, 9). Assessments of neuronal morphology have revealed that neurons from primary cultures carrying the microdeletion have significantly reduced axonal and dendritic branching (10, 11)

In this study, we used *Df(16)A*$^{+/-}$ mice to develop a mouse embryonic stem cell (mESC)-derived neuronal model with the goal of generating a high-throughput assessment system that can be used to screen for novel compounds for 22q11.2DS-associated SCZ. mESC-derived neurons from both genotype groups display characteristics of electrophysiological maturity. Single cell RNA sequencing (scRNAseq) revealed neuronal cultures consisting of multiple mouse CNS cell types including excitatory neurons, inhibitory neurons and neuroblasts. *Df(16)A*$^{+/-}$ mESC-derived cortical neurons have significant reductions in the number of neurite branchpoints and total neurite length, recapitulating the phenotype seen in the *Df(16)A*$^{+/-}$ mice, providing a useful cell-morphological read-out to use for drug screening.

METHODS

ESC Culture and Differentiation

ES lines from the *Df(16)A*$^{+/-}$ mouse and wild-type littermates (mESC) were cultured as described in (12) and differentiated into neuroepithelial tissues by culturing as floating aggregates in a serum-free medium (serum-free floating culture of embryoid body–like aggregates with quick reaggregation: SFEBq culture). For SFEBq culture, mESCs were dissociated to single cells in 0.25% trypsin-EDTA (Invitrogen) and quickly reaggregated in ES growth medium (3000 cells/150 ml/well) using 96-well low cell-adhesion plates containing cortical differentiation medium (GMEM, non-essential amino acids, Na pyruvate, 2-mercaptoethanol, and KOSR). mESC cell aggregates were termed embryonic bodies (EBs). On day 7, EBs were transferred to a 10 cm bacterial-grade dish containing cortical maturation media (DMEM/F12 supplemented with N2 and penicillin streptomycin). On day 10, EBs were dissociated and the mESC-derived neurons were plated onto poly-ornithine (Sigma) and laminin (Life Technologies) coated coverslips (12mm: BD Biosciences), 6-well dishes (Falcon), or 96-well plates (Falcon) in plating media (N2/B27 + UfDU; uridine and fluorodeoxyuriding: Sigma). After three days media was changed to BrainPhys growth media (Stem Cell Technologies) for the duration of neuronal growth/maturation. Additionally, neuronal maturation was assessed at day 15 (5 days post-plating), and day 21 (11 days post-plating). We determined that by day 30 (20 days post-plating) neurons displayed appropriate characteristics of maturity.

Immunohistochemistry

mESC-derived neurons were plated onto dual-coated glass coverslips and fixed in 4% paraformaldehyde for 20 minutes on ice at Days 15, 21, and 30; then permeabilized with 0.25% Triton X-100 for 15 minutes, blocked with horse serum for 1 hour, and incubated overnight at 4°C with primary antibodies diluted in 3% BSA. The following day, the cells were incubated with secondary antibodies and mounted with Vectashield containing DAPI. Primary antibody dilutions were used as follows: anti-Tbr1 (Abcam, *ab46020,* 1:500), and anti-MAP2 (1:1000). Secondary antibodies conjugated to Alexa Fluor 488, 555, and 647 were used at a dilution of 1:1000 (Life Technologies). Images were acquired on an epifluorescence microscope under 10x and 40x magnifications.

Electrophysiology

Whole-cell electrophysiology was performed as previously described (13) on mESC-derived neurons at days 15, 21, and 30 post-plating onto glass coverslips (12mm; BD Biosciences). Electrophysiological signals were acquired using Multiclamp 700B amplifier, 1440A DigiData and pClamp10 software (Lowpass filter frequency: 10,000 Hz) (Molecular Devices). Whole-cell patch-clamp recordings were performed using borosilicate glass pipettes (initial resistance 5.0–8.0 MΩ). An external solution was used that contained (in mM): NaCl 145, KCl 5,

HEPES 10, Glucose 10, $CaCl_2$ 2, $MgCl_2$ 2, pH 7.3 with NaOH, adjusted to 325 mOsm with sucrose. An internal solution was used that contained (in mM): $CsMeSO_4$ 130, $NaMeSO_4$ 10, EGTA 10, $CaCl_2$ 1, HEPES 10, MgATP 5, Na_2GTP 0.5, pH 7.2 with KOH, adjusted to 290 mOsm with sucrose. Intrinsic properties were measured Input resistance was measured in the voltage-clamp mode at –70 mV from the current response to a 5 mV hyperpolarizing voltage step and calculated by Ohm's law from the current change. For assessing neuronal excitability, action potential (AP) firing was recorded in current-clamp mode in response to incremental, depolarizing current injections of 500 ms duration (10pA increment of 25 steps). The number of AP firings was plotted as an average of the maximum number of firings exceeding a threshold of 0mV during a given train. Waveform analysis of initial AP firing was performed using standard Clampfit routines as follows: *Duration*: the time from AP rise to AP decay from threshold; *Rheobase*: measured by a series of short (5 ms duration) current injections (20 pA increment of 80 steps) and calculated as the minimal current amplitude to elicit the initial AP. In current-clamp mode, the resting membrane potential (RMP) of all cells were reported as baseline measurements, and were then adjusted to –70 mV by injection of a small standing current before experiments began. Threshold was determined by the membrane potential of the cell when the first AP was elicited. Statistical analyses were performed using Igor Pro 8 and Prism 8 (GraphPad). Data are presented as means ± SEM.

High-Throughput Imaging and Analysis of Neurite Outgrowth and Branching

mESC-derived neurons were plated onto 96-well dual-coated plates (Falcon) at a density of 5,500 cells/well and allowed to mature for 20 days. On Day 30 cells were fixed in 4% PFA and stained with dendrite marker MAP2. Images were acquired with the IN Cell Analyzer 2000 (GE Healthcare) integrated with Cytomat microplate managing stations via a robotic manipulator (Mitsubishi) to accommodate a high volume of plates. High-content image acquisition and analysis were performed using IN Cell Investigator and MetaMorph (Molecular Devices) containing modules specific for assessment and quantification of dendrite outgrowth and branching. Multitarget analysis was used to analyze the fluorescence-based images, and the Top-hat segmentation method was used to analyze nuclei, with nuclei defined as having a minimum area of 20 μm^2 with software sensitivity set at 50. Hoechst staining of nuclei was used to acquire maintenance measures of nuclear area to confirm no difference in cell size between groups existed, and cell counts for normalization. Quantification of dendrite branching and outgrowth was performed using imageJ. For each plate, measures were normalized to cell number (Hoechst positive) and dendrite outgrowth and branching were normalized to the mean of DMSO control treated wells.

Single-cell RNAseq analysis

Single-cell count matrices were aggregated with CellRanger (cellranger aggr function) and loaded into a single scanpy AnnData object (scanpy v1.4.6). QC metrics were analyzed (total number of counts, number of genes detected, percentage of mitochondrial RNA) to identify outliers in the distribution of cells and remove them (PMID: 26887813). Afterwards, cell counts were normalized using pool-derived size factors (PMID: 27122128) and log-transformed. The LIGER algorithm with default parameters was used to integrate the different batches (Welch, Joshua D., et al. "Single-cell multi-omic integration compares and contrasts features of brain cell identity." Cell 177.7 (2019): 1873-1887.). We used UMAP (McInnes, L. & Healy, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426 (2018)) to visualize the distribution of cells. To identify cell sub-populations, we used the Leiden algorithm for community detection (Traag, Vincent A., Ludo Waltman, and Nees Jan van Eck. "From Louvain to Leiden: guaranteeing well-connected communities." Scientific reports 9.1 (2019): 1-12.) with different values for the resolution parameter, analyzing the average Silhouette score across all cells for each clustering (Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics 20 (1987): 53-65). We computed differentially expressed genes using the Wilcoxon Rank-Sum test as implemented in scanpy, and we labelled the different populations based on the genes differentially up-regulated in each population.

RESULTS

mESC-derived neurons from *Df(16)A*$^{+/-}$ and wildtype mice develop into electrophysiologically mature neuronal cultures.

In order to establish a high-throughput platform that could be used for drug screening purposes in the next phase of this study, we deployed the protocol developed by Sasai and colleagues (12) to generate neuronal cultures derived from mESCs from 3 *Df(16)A*$^{+/-}$ and 3 wildtype mice (Fig 1A). Both wildtype and *Df(16)A*$^{+/-}$ neurons displayed expression of the dendritic marker *MAP2* and deep layer early-born neuronal marker *TBR1*, whose expression was first confirmed at 15 days *in vitro* and remained at 21 and 30 days. As expected, dendritic complexity increased with time (Fig. 1B). Consistent with neuronal maturation, whole cell patch clamp recordings from neurons from both genotype groups display a shortening duration of action potentials over development (Fig. 1C). Several other membrane properties were measured, including resting membrane potential, input resistance, action potential threshold, none of which showed a statistically significant difference between genotype (Fig. 1C and Table 1). Together, these results indicate that the neurons generated in both genotype groups are electrophysiologically mature.

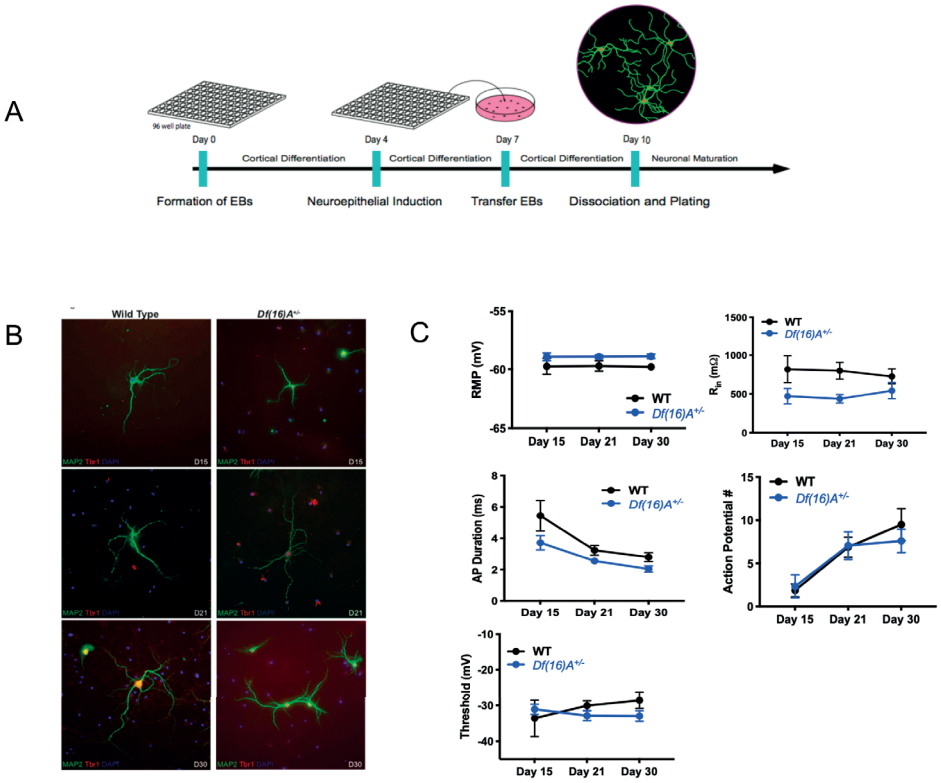


Figure 1. Generation and basic characterization of mESC-derived neuronal culture system.

A. Flow chart displaying major steps of the differentiation protocol used in this study.

B. Examples of representative immunohistochemistry images for wildtype and Df(16)A$^{+/-}$ neurons stained for dendritic marker MAP2 and early born neuronal marker TBR1. Expression of these marker is first seen at DIV15 and remains present until DIV30. For both genotypes, dendritic complexity increases as time progresses. Scale bar represents 20 µm.

C. Plots displaying electrophysiological properties indicating electrophysiological maturity for neurons from both genotype groups. Top-left panel: resting membrane potential, top-right panel: input resistance, mid-left panel: action potential duration, mid-right panel: number of spontaneous action potentials fired, bottom-left panel: action potential threshold.

Table 1. All data are presented as mean ± SEM. An unpaired t-test was used to compare cell types.

Properties	WT	*Df(16)A*$^{+/-}$
Resting Vm (mV)	-64.3 ± 0.3	-64.6 ± 0.5
Input Resistance (mΩ)	619 ± 153	671 ± 79
Spike Voltage Threshold (mV)	-30.8 ± 3	-31.1 ± 2.5

Single-cell RNA sequencing reveals mouse CNS cell type identities generated in neuronal cultures.

To characterize the neuronal culture in closer detail, we used scRNAseq to uncover the cell type composition of the neuronal culture system. Unsupervised Leiden clustering (14) of 16429 cells from 3 wildtype and 3 *Df(16)A*$^{+/-}$ neuronal cultures at 30 days *in vitro* revealed 11 distinct

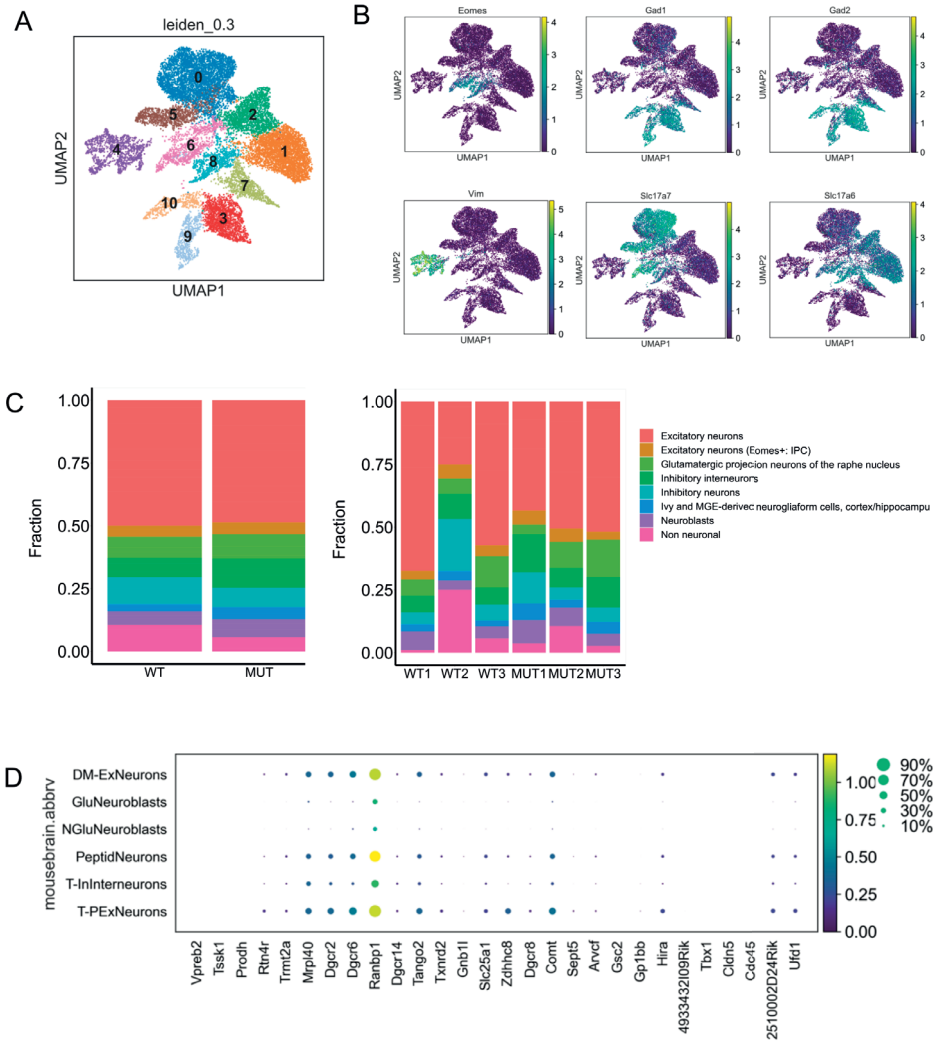


Figure 2. scRNAseq characterization of mESC-derived neuronal culture system.

A. Scatter plot showing distribution of cells based on UMAP dimensionality reduced single cell transcriptomes. Different colors represent 11 distinct clusters as determined using unsupervised Leiden clustering.

B. Scatter plot showing same distribution of UMAP dimensionality reduced cells as in panel A, but here the color gradient represents the mRNA normalized expression levels for a single marker in the different clusters. Eomes displays high expression in intermediate progenitor cells (top-left), Gad1 and Gad2 display high expression in inhibitory neurons (top-middle and top-right), Vim (bottom-left) and Slc17a7 and Slc17a6 (bottom-middle and bottom-right) display high expression in excitatory neurons.

C. Stacked bar plots displaying distribution of different cell types between different genotypes (left panel) and different cell lines (right panel). These plots show that neuronal cultures from both genotype groups contain the same cell types, with some variability between different cell lines but no marked difference between genotypes.

D. Dot plot displaying expression levels of genes deleted heterozygously in *Df(16)A*$^{+/-}$ mice, compared to expression levels for the same genes in wildtype neuronal cultures. In mutant neuronal cultures, the genes show an approximate 2-fold reduction in expression levels, which is expected considering the heterozygous microdeletion. Color of dots represents expression level, size of dots represent fraction of cells expressing the gene.

clusters (Fig. 3A). These clusters were annotated using the census cell types from the mouse central nervous system as published by Zeisel et al. (15). Cell types present in the culture system include several types of telencephalic excitatory neurons, glutamatergic projection neurons of the raphe nucleus, inhibitory interneurons and neuroblasts (Fig 2A and table 1). Composition of cell types was consistent between cell lines and genotypes (Table 1). Excitatory neurons show high expression of markers Slc17a7, Slc17a6 and Vim (Fig 2B). Given the fact that excitatory neurons make up the largest fraction of cell types for both genotypes, and these numbers are quite consistent between cell lines and genotypes, the cultures were considered appropriate for high-throughput morphologic assessment. Next, we analyzed expression levels of the genes present in the microdeletion region in multiple cell types, and discovered an approximate 50% reduction in mRNA expression for these genes, consistent with a heterozygous deletion of these genes (Fig 2D).

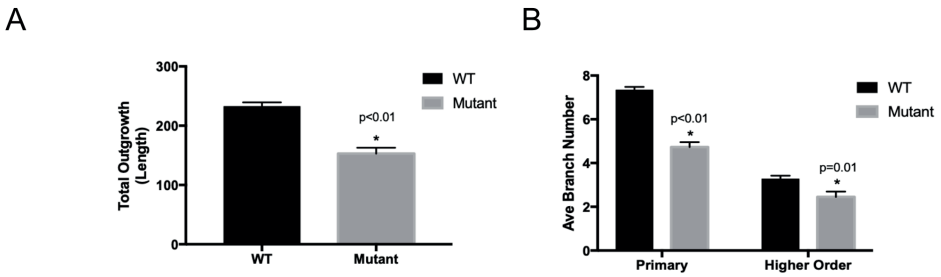


Figure 3. mESC-derived neuronal cultures recapitulate neurite-outgrowth phenotype observed in *Df(16)A*$^{+/-}$ mice.

A. Quantifications of total neurite outgrowth of of MAP2-stained neurons showing a robust neurite outgrowth deficit.

B. Quantification of number of primary and higher order dendrite branches between wildetype and mutant neurons, showing reductions in both primary and higher orders dendrite branches in mutant neurons.

mESC-derived neurons from *Df(16)A*$^{+/-}$ recapitulate impaired dendritic arborization phenotype observed in *Df(16)A*$^{+/-}$ mice.

Previous studies by our group have shown impaired dendritic and axonal morphology (neurite outgrowth) in dissociated hippocampal neurons (10, 11). Dendritic arborization deficits include a decrease in the number of primary dendrites, the number of branchpoints and total length of dendrites. We hypothesized the dendrite outgrowth phenotype would be a well-suited read-out to use for high throughput drug screening for several reasons. First, the phenotype is hypothesized to be present in cell types that can be generated using *in vitro* cultures. Second, the phenotype has a large effect size in mice. Therefore, a full rescue of the phenotype could be easily distinguished from biological variability, thereby reducing the need for a large sample size of neurons measured. Third, the phenotype can be easily measured in an automated, *high-throughput* manner, allowing a high number of compounds to be tested with high cost efficiency.

We analyzed the images of *Df(16)A*$^{+/-}$ mESC-derived neurons stained for MAP2, and discovered dendrite outgrowth deficits compared to wildtype neurons. *Df(16)A*$^{+/-}$ neurons displayed a decreased number of primary dendrites, a decreased number of branchpoints and a decreased total length of dendrites (Fig. 3A and B). The effect of this cellular phenotype is of large enough size to detect a pharmacologic rescue.

Table 2. Table showing fractions of annotated cell types for all independent cell types from both wildtype and mutant genotype groups.

Cluster N Cells	1 (wt)	2 (wt)	3 (wt)	4 (mut)	5 (mut)	6 (mut)	WT tot	*Df(16)A*$^{+/-}$	Annotation (Zeisel et al.)
0 4071 (24.8%)	41.1%	10.7%	29.7%	24.6%	22.5%	21.7%	27.1%	22.7%	Excitatory neurons
1 3133 (19.1%)	16.8%	12.3%	18.7%	11.8%	20.2%	27.1%	16.9%	21.1%	Excitatory neurons
2 1705 (10.4%)	6.6%	6.1%	12.5%	3.8%	10.5%	14.9%	10.1%	10.7%	Glutamatergic projection neurons of the raphe nucleus
3 1559 (9.5%)	6.5%	10.0%	6.8%	15.2%	7.7%	12.1%	7.5%	11.3%	Inhibitory interneurons
4 1232 (7.5%)	1.1%	25.1%	5.7%	3.8%	10.7%	2.7%	9.5%	5.7%	Non neuronal
5 1073 (6.5%)	9.5%	2.1%	8.9%	7.0%	7.9%	3.1%	7.4%	5.7%	Excitatory neurons
6 967 (5.9%)	7.4%	3.6%	4.9%	9.2%	7.3%	4.8%	5.0%	6.7%	Neuroblasts
7 873 (5.3%)	1.9%	13.3%	5.0%	6.2%	3.8%	3.5%	6.5%	4.3%	Inhibitory neurons
8 728 (4.4%)	3.3%	5.6%	4.3%	5.6%	5.2%	3.1%	4.4%	4.4%	Excitatory neurons (Eomes+: IPC)
9 609 (3.7%)	2.8%	3.8%	2.2%	6.6%	3.1%	4.8%	2.7%	4.7%	Ivy and MGE-derived neurogliaform cells, cortex/ hippocampus
10 479 (2.9%)	2.9%	7.4%	1.4%	6.2%	1.1%	2.2%	3.1%	2.8%	Inhibitory neurons

DISCUSSION

With this work, we have generated and characterized a mESC-derived neuronal culture system from *Df(16)A*$^{+/-}$ mice for future use as a drug screening platform to find novel compounds to treat 22q11.2DS-associated schizophrenia. The culture system produces electrophysiologically mature neurons and consistently produces the neuronal cell types suitable for phenotypic assessment. The dendrite outgrowth deficits present in *Df(16)A*$^{+/-}$ mice are recapitulated in the mECS-derived *Df(16)A*$^{+/-}$ neurons and are a well suited read-out that can be used for high throughput automated morphological assessment.

For this drug screening project to proceed, there are various steps that need to be accomplished. First, the system needs to be scaled up so that high number of compounds can be screened. Second, a treatment schedule needs to be established for time points and dosing of compounds after which an initial screening can be initiated. In later stages, in case there is an identification

of compounds that rescue the dendrite outgrowth phenotype, these compounds first need to be tested in live *Df(16)A*$^{+/-}$ mice to investigate whether the cellular and the behavioral phenotypes are rescued as well.

In parallel to what is outlined above, the *Df(16)A*$^{+/-}$ neurons in the culture system can be investigated further to elucidate possible mechanisms underlying the developmental dendrite outgrowth phenotype. The scRNAseq data will be analyzed in closer detail to uncover differential gene expression in mutant neurons to develop hypotheses as to which biological processes are altered in these neurons. Differential gene expression can subsequently be used as a secondary outcome measure in neurons that have displayed a dendrite outgrowth phenotype rescue.

For obvious reasons, the following steps after completing the drug screening in mESC-derived cultured neurons and validation in *Df(16)A*$^{+/-}$ mice, is to utilize iPSCs derived from patients carrying the 22q11.2 microdeletion to generate neuronal cultures, either 2D or 3D, to test whether identified compounds in the mouse system have similar effects in the human system. These will be the first steps towards testing potential novel compounds in human patients with this condition.

REFERENCES

1. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med. 2005;353(12):1209-23.
2. Webber MA, Marder SR. Better pharmacotherapy for schizophrenia: what does the future hold? Curr Psychiatry Rep. 2008;10(4):352-8.
3. Weinberger DR. Future of Days Past: Neurodevelopment and Schizophrenia. Schizophr Bull. 2017;43(6):1164-8.
4. Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry. 1999;56(10):940-5.
5. Xu B, Hsu PK, Stark KL, Karayiorgou M, Gogos JA. Derepression of a neuronal inhibitor due to miRNA dysregulation in a schizophrenia-related microdeletion. Cell. 2013;152(1-2):262-75.
6. Khan TA, Revah O, Gordon A, Yoon SJ, Krawisz AK, Goold C, et al. Neuronal defects in a human cellular model of 22q11.2 deletion syndrome. Nat Med. 2020.
7. Bershteyn M, Nowakowski TJ, Pollen AA, Di Lullo E, Nene A, Wynshaw-Boris A, et al. Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia. Cell Stem Cell. 2017;20(4):435-49 e4.
8. Stark KL, Xu B, Bagchi A, Lai WS, Liu H, Hsu R, et al. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nat Genet. 2008;40(6):751-60.
9. Emamian ES, Hall D, Birnbaum MJ, Karayiorgou M, Gogos JA. Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia. Nat Genet. 2004;36(2):131-7.
10. Mukai J, Dhilla A, Drew LJ, Stark KL, Cao L, MacDermott AB, et al. Palmitoylation-dependent neurodevelopmental deficits in a mouse model of 22q11 microdeletion. Nat Neurosci. 2008;11(11):1302-10.
11. Mukai J, Tamura M, Fenelon K, Rosen AM, Spellman TJ, Kang R, et al. Molecular substrates of altered axonal growth and brain connectivity in a mouse model of schizophrenia. Neuron. 2015;86(3):680-95.
12. Eiraku M, Sasai Y. Mouse embryonic stem cell culture for generation of three-dimensional retinal and cortical tissues. Nat Protoc. 2011;7(1):69-79.
13. Cortese GP, Zhu M, Williams D, Heath S, Waites CL. Parkin Deficiency Reduces Hippocampal Glutamatergic Neurotransmission by Impairing AMPA Receptor Endocytosis. J Neurosci. 2016;36(48):12243-58.
14. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9(1):5233.
15. Zeisel A, Hochgerner H, Lonnerberg P, Johnsson A, Memic F, van der Zwan J, et al. Molecular Architecture of the Mouse Nervous System. Cell. 2018;174(4):999-1014 e22.

Chapter 5

General discussion

GENERAL DISCUSSION

The levels of investigation of the three studies included in this work range from detailed psychiatric phenotyping in human research participants to generating a drug screening platform using embryonic stem cells derived from mice that were genetically engineered to model a human disease. The psychiatric disorders as classified by the DSM-V that are investigated in these studies include MDD, ASD and SCZ. The scientific community's depth of understanding of the genetic architecture and biological underpinnings of these three conditions differ substantially. The experimental designs of the three studies in this thesis are reflective of this. For MDD, the genetic architecture has remained most elusive, with mainly common variants of small effect size and of unclear biological function known to date (1-4). Therefore, disease modeling using genetically engineered mice has been the least advanced for MDD, and rodent-modeling for MDD relies heavily on environmental stress-induction paradigms such as maternal separation or social defeat (5). For SCZ, a higher number of both common SNPs (6-9) and highly penetrant high-confidence genes have been implicated (10), and extensive studies in mainly two genetic mouse models have shed light on possible mechanisms underlying SCZ-related mouse behavior (11-15). For ASD, the elucidation of its genetic architecture is advanced, with associated common SNPs (16) and several hundred high-confidence genes identified to date (17, 18). Numerous studies of ASD using genetically modified rodents or patient-derived iPSCs have uncovered common themes in underlying biology (19), thereby creating opportunities to search for novel therapeutics that interfere with these processes. As a result, preclinical treatment studies have shown promise for ASD (20-22), although successful translations to the clinic are yet to come.

There are several contrasting features of these disorders that give rise to the differences in their genetic architecture and challenges in identifying risk genes of large effect. These are complex issues that could be discussed at great length, but I will only elaborate briefly on a small selection here. First, are the differing heritabilities of the disorders. The lower heritability that is seen in MDD, versus the higher heritability in ASD and SCZ imply that for MDD there is a larger role for environmental factors to contribute to disease status. This explains well why larger cohorts were needed to uncover common the genetic variation that contributes to disease risk. Second is the way in which the diagnostic constructs are defined and how easily they are distinguished from 'normal' human behavior. One could argue that virtually every human being experiences feelings of sadness reminiscent of MDD in their lifetime, whereas for example fewer people experience psychotic symptoms that do not reach threshold for a psychiatric diagnosis (23). For MDD, this allows for more diagnostic uncertainty and therefore more noise in the genotype-phenotype correlation, especially in large scale studies that use less accurate phenotyping methods. The importance of this point is corroborated by the fact that looking at the extreme end of the psychiatric phenotypic spectrum allows utilizing smaller

sample sizes in order to reveal statistically significant genomic associations (24, 25). The third point is age-of-onset of disease, which has been especially relevant to the progress that has been made in elucidating (human-specific) disease mechanisms. In contrast to MDD and SCZ, ASD has its onset very early in life, and there is an increasing body of evidence for early embryonic origins of the disorder (26). This has led to early successes in ASD modeling studies using iPSC-derived neuronal systems, which are validated to model embryonic development most faithfully (27, 28). To conclude, our current understanding of the genetic architecture and disease mechanisms of ASD and SCZ is ahead of that of MDD, and perhaps for this reason disease mechanism-informed novel treatments for these conditions will be developed sooner than for MDD. In the following sections, I will discuss the advantages, limitations and future directions of the studies in this thesis and their respective fields.

Advantages, limitations and future directions of chapter 2.

Advantages of study

One of the strengths of this study is that the missense variant identified in *GPR156* confers a relatively large risk for mood disorders, with carriers displaying a three-fold increased risk for major depressive disorder or bipolar 1 disorder than non-carriers. This is in stark contrast with other (common) established genetic risk factors for depression, with most of the common variants for MDD having odds ratios between 1.05 and 1.15 (1). With this finding, we show that rare variants of large effect also play a role in the genetic architecture of depression, and that family studies using whole exome or -genome sequencing are valuable methods to uncover this genetic architecture alongside large scale GWA-studies. Although family studies for rare variant detection can be performed successfully in outbred populations (29), it is likely there are several aspects of doing this study in the Old-Order Mennonite population that allowed us to successfully identify this genetic variant, as discussed in the introduction of this work.

Another advantage of this study is the functional consequence of the c.1599G>T mutation on the protein, causing a remarkable behavioral phenotype in *knock-in* mice carrying the human mutation that has to our knowledge not been described before. This implies that the mutation is not only associated with MDD –as is the case with most common SNPs of which biological function are in most cases unclear (30), but could have a causal consequence on the depression phenotype. Together, these two advantages make the follow-up studies of this gene potentially highly informative for studying the consequences of the mutation on GPR156 at the molecular, cellular, circuit and behavioral level.

Limitations of this study

One of the limitations of this study is the fact that the variant we discovered is extremely rare in outbred populations, raising the question of whether the gene is relevant to the biology of depression in the general population. This issue may be reflected as well in the finding that

the *GPR156* locus did not come up in any of the GWA studies on depression so far, nor was it a result of putatively constructed gene sets using computational approach MAGMA (1). Of note, this does not exclude the possibility of involvement of *GPR156* in the biology of depression in the general population *per se.* For example, for ASD, PTVs in *CNTNAP2* are responsible for a syndromic form of the disorder (31), whereas the *CNTNAP2* locus was not in proximity to the hits from a large recent GWA-study on ASD (16). The implications of this potential shortcoming are especially relevant for possible downstream functional- and drug discovery studies on *GPR156*. In case a targetable disease mechanism would be uncovered, having this only be applicable in individuals with mutations in *GPR156* would make the pursuit of drug-discovery highly laborious and cost-inefficient. This is one of the inherent challenges of gene-driven therapeutics discovery, especially for phenotypes where hundreds to thousands of risk variants may be causally contributing to the development of the phenotype.

Another limitation that is also related to the rarity of the variant, is the limited sample size for the validation step of the study. Ideally, after the initial discovery of a variant in a discovery cohort, the variant is validated in an independent validation cohort of sufficient size to do reliable statistics. This approach has been widely applied in GWA-studies (32). In linkage and candidate designs, it was often the validation step that failed to confirm a positive finding (32, 33). In this study, we did identify the same variant in a healthcare database. Together, the validation cases confirm genome-wide statistically significant positive associations with depression. However, the sample size of the 'validation cohort' is very small, making the validation less robust than desirable. This challenge may be inherent to the discovery of (extremely) rare variants, and is especially challenging for a phenotype that is complex in etiology and common, like that of depression. Given the size of the genome and all of its potential variations, it is possible that a certain variant will only exist in one single family (34), making validation in replication cohorts or families very challenging.

Future directions

Solutions that could help overcome this validation problem, at least partially, is to further extend the multimodal approach that we employed in this study. Favorable methods include interrogation of other variants in the same gene in databases (burden analysis), functional studies including gene KO and knock-down in animal- or human models, and expression studies in human tissues. Combining multiple lines of evidence on different organizational levels will strengthen the overall evidence of the involvement of a gene in a complex phenotype. In addition to this, it will be of critical importance to have increasing numbers of publicly available databases with both genotypic and phenotypic data on large cohorts of individuals from various ethnical and geographical backgrounds, to allow researchers to mine for as much genotype-phenotype evidence as possible.

Generating a *knock-in* mouse line of the human mutation and discovering the presence of a behavioral phenotype was the first step that confirmed its functional importance and corroborated the involvement of *GPR156* in depression. There are several important future experiments to perform in order to uncover the functions of *GPR156*. At the receptor level, many questions remain unanswered. For example, what is the natural ligand of this receptor? This question could be answered using the traditional reverse pharmacology approach where a large library of possible ligands can be tested on the GPCR in a cell culturing system (35). However, this approach is not always successful and newer strategies using GPCR crystallography could be considered as well (36). Another interesting avenue would be to investigate the downstream signal transduction of this GPCR. In further stages, high throughput screening combined with structure based drug design could be used to investigate potential pharmacologic ligands to rescue receptor function. Free fatty acid 1 receptor (FFA1; previously known as GPR40), is an example of a GPCR involved in metabolic disorders that went from orphan to drug target using a similar approach (36). At the behavioral level, further phenotyping experiments could investigate additional behavioral phenotypes reminiscent of depression in mice, by for example inducing social defeat- or maternal separation stress, to investigate the loss of sucrose preference, which is thought to be analogous to anhedonia in humans. In addition, 2-photon calcium imaging could be utilized to investigate neural activity in the habenula that may be correlated with these behavioral phenotypes.

Advantages, limitations and future directions of chapter 3.

Advantages of study

A strength of this study is the finding of a robust phenotype that resembles one of the phenotypes also present in the human patients carrying the homozygous c.3709DelG mutation in *CNTNAP2*: the increased head circumference and the increase in grey matter thickness. Finding such a clinically-related phenotype in the modeling system is notable as it suggests that the disease process that is being modeled may be related to biological processes occurring during embryonic brain development in these patients. It is important to acknowledge, however, that to achieve certainty that this finding is indeed related to the clinical phenotype, would still require post-mortem examination of brains of pediatric patients carrying the homozygous c.3709DelG mutation in *CNTNAP2* to show that increased cortical thickness and cell numbers are indeed correlated -and increased in these patients.

Limitations of study

One of the limitations of this study is the issue of the extent to which the features seen in this reductionist *ectoderm-lineage-only* model system are a faithful recapitulation of the biology occurring in the developing human embryo. For example, it is well possible that the cellular behavior of proliferating and differentiating NPCs in the ventricular zone would be different in the presence of mesoderm-lineage-derived vasculature providing the cells with oxygen, nu-

trients and molecular signals from different developing organ systems. Therefore, we cannot be certain that the uncovered phenotypes are analogous to processes in the embryonic developing brains of individuals with PTVs in *CNTNAP2*. The validation of this model system compared to *in vivo* conditions in human embryos is therefore critical, but will remain challenging given the limited access to this tissue.

Another shortcoming of this study – and of the current brain organoid field overall, is the substantial amount of variability that exists between organoids generated from the same line. For example, this variability can be seen in the boxplots of the fraction of PAX6 positive cells in our organoid system (Fig. 3F). This variation possibly arises from the complex balance between tissue self-patterning properties and the use of external molecular cues to direct the tissue into desired lineage, and an absence of reproducible morphogen gradients. This has made it necessary to hone in on phenotypes of large effect (such as the organoid volume phenotype), and has made it challenging to detect more subtle phenotypes that could nonetheless be relevant to disease biology. We have made an effort to overcome this issue at least partly by selecting patient samples from a genetically isolated population. Thereby we have aimed to reduce a level of variability that could be coming from heterogeneity in genetic background that would exist among different individuals from the outbred population.

Future directions: possible mechanism of overgrowth phenotype in CNTNAP2 organoid model

One of the critical and interesting questions that remain after this study is; by which mechanism does CNTNAP2 truncation lead to cortical overgrowth? One of the first steps to answer this question would be to determine whether the NPC proliferation phenotype is a direct result of protein truncation –and thus cell-autonomous- or whether the phenotype is a non-cell-autonomous, secondary effect. Given our prior knowledge of CNTNAP2 being a cell adhesion molecule with similar protein domain structure to the neurexins and confirmed functions in synapse biology (37), it is tempting to speculate that the more likely possibility would be that the proliferation phenotype is a secondary effect. Even more so, because our scRNAseq data has shown that *CNTNAP2* expression is highest in cortical excitatory neurons, and relatively low (but not absent) in NPCs. A possible explanatory mechanism could be that disruption of biological processes related to cell adhesion (such as for example neuronal migration) triggers activation of certain signaling transduction cascades that regulate cell proliferation. Possible examples of these could be abnormal mTOR/PI3K signaling or WNT signaling, both of which have been implicated in ASDs and NPC proliferation/neurogenesis (38-41).

In case the NPC proliferation phenotype is a direct result of *CNTNAP2* loss-of-function, it would be possible that *CNTNAP2* fulfills a direct role in the process of cell division. The fact that both NPC proliferation and differentiation are impaired in the *CNTNAP2*-deficient

organoids, implies there could be a disruption of the mitotic process, in which both division speed and division plane would be altered. There is evidence for multiple genes in which disruption causes both abnormalities in proliferation and differentiation (42). If this hypothesis were to be true, *CNTNAP2* could be considered to have tumor suppressor-*like* functions. Interestingly, there have been several studies that have found evidence for the involvement of *CNTNAP2* in brain tumors (43, 44), and one study has found CNTNAP2 to have tumor-suppressive function in HEK cells (43). This latter possibility would fit in with the biology of a wide range of synaptic proteins that fulfill additional roles in different parts of the cell, such as for example transcriptional regulation (45).

Future directions: from abnormal early developmental processes to behavior

In this study we have demonstrated aberrant early developmental cell biological processes such as NPC proliferation and differentiation that lead to the generation of a higher total cell number in case-derived brain organoids when compared to controls. Whether neurons are produced in excess numbers in the cortices of these patients still has to be further determined, but this is certainly a possibility. An interesting outstanding question is how these early cellular processes lead to abnormalities at the circuit- and behavioral level. A possibility is that an increased total number of excitatory neurons may lead to an excitation/inhibition imbalance in favor of excitation. The relationship between the excess production of excitatory neurons and E/I imbalance and ASD-related behavior has previously been shown by Fang and colleagues in mice (46). The underpinnings at the level of neuronal circuits have remained unclear, however. In the case of *CNTNAP2*-related pathology, insight at the circuit level has been uncovered in a recent study that utilized *Cntnap2* null mice to reveal abnormal neuronal population activity in the domains of categorization of sensory stimuli and experience dependent refinement of social stimuli in the prefrontal cortex (47). Interestingly, however, increased numbers of excitatory neurons in the cortical areas of *Cntnap2* null mice have thus far not been demonstrated, and one study reported no difference in cortical volumes between case and control mice (48). Therefore, it remains unclear whether the population activity phenotype demonstrated by Levy et al. is related to the number of cells in the cortices of these mice. It would be of importance to investigate whether the findings from our human modeling system are also reproduced in *Cntnap2* null mice to then make the link to circuit level abnormalities underlying ASD-related behavior.

Future directions: identifying a pharmacologic rescue

Rescuing the observed phenotypes by repairing the homozygous protein truncating mutation using CRISPR-Cas9 in the brain organoid system confirmed the causative effect of the mutation. Although repairing this mutation in a fertilized oocyte would be the most efficacious way of treating this disorder in human patients, this is not possible with current gene editing

techniques and ethical regulations. Therefore, setting out to find a pharmacologic compound that rescues the observed phenotypes in the organoid modeling system is an important future experiment for finding a treatment for *CNTNAP2*-associated pathology in patients. In addition, as our bulk – and scRNAseq analyses have shown overlapping transcriptional signatures to other ASDs, finding a pharmacologic compound that rescues *CNTNAP2*-associated pathology could be of interest for other forms of ASD as well.

One route to search for a therapeutic compound would be to search for evidence of abnormal signaling pathway activity based on preexisting knowledge on their involvement in ASD and brain overgrowth. As noted above, some candidate pathways include PI3K-AKT-mTOR, Notch, and Wnt signaling. For some pathways, demonstrating their involvement in a phenotype can be achieved in straightforward ways by measuring read-out proteins using for example western blots or immunohistochemistry. This is for example the case for Pi3K-AKT-mTOR signaling, for which Zhang and colleagues rescued a brain overgrowth phenotype using AKT inhibitor in brain organoids containing a mutation in *RAB39b (49)*. For other pathways, such as Notch signaling, its involvement may have to be investigated using different methods because reliable read-out proteins are unknown. The advantage of this approach is that a limited number of pathways can be assessed for and therefore the number of candidate compounds is limited as well.

Another approach to find a compound that rescues the overgrowth phenotype would be to conduct a drug screening. Using organoid size as a read out would be a well-suited approach because the phenotype has a relatively large effect size and can be easily measured using regular light microscopy. High-throughput pipelines for drug screens have already been developed for brain organoids (50, 51). An advantage of this approach is that a broad range of compounds could be tested with no prior assumptions regarding biology relevant to the phenotype. Therefore, compounds targeting unexpected proteins and biological processes could be identified. A disadvantage of this approach is that it may be more costly and labor intensive, especially given the early stage of the organoid drug screening field.

Future directions of the brain organoid field as a whole

As mentioned before, one of the major challenges of the brain organoid field is the substantial amount of heterogeneity between organoids, which occurs especially at the tissue-organizational level. There are several ways in which these issues could be addressed in the future. First, the starting material from which organoids are grown, pluripotent stem cells, show considerable amounts of variability related to cell line, passage number and even location within the colony (52-54). Understanding and controlling these parameters well before growing organoids could reduce this level of variability. Second, organoid generation protocols should innovate in terms of methods by which reproducibility is optimized between organoids grown from the same

cell line. One way by which this has been achieved is by introducing an organizer center that generates an intrinsic morphogen gradient that ensures higher reproducibility on the tissue-organizational level (55).

Another limiting factor in growing brain organoids is the diffusion of oxygen and nutrients, so improving methods for supplying these is critical. Some advances on this have already been made by transplanting organoids into animals and using the vasculature of the host (56). Future methods could use microfluidic systems or biomaterials to generate artificial vascular networks to supply the necessary oxygen and nutrients.

Importantly, in further improving brain organoid models, it remains important to validate organoid tissue to its corresponding primary tissue (eg. fetal neocortex). So far, this has mainly been done using transcriptome datasets (27, 28). In the future, validation should happen in a broader range of functions such as (sub)cellular morphology, tissue architecture and biochemical measures (57), to ensure the tissue generated recapitulates *in vivo* biology in a faithful manner.

Advantages, limitations and future directions of chapter 4.

Advantages of study

One of the advantages of this study is that the impaired dendritic arborization phenotype is robust and consistent across all three cell lines tested. This will facilitate the process of testing large numbers of compounds by assessing whether the phenotype will be rescued. Finding such a phenotype in a stem cell-derived neuronal system is not trivial, as phenotypes often show a high level of variability between and within cell lines. Another advantage of this study is that the *Df(16)A*$^{+/-}$ mouse model is the most thoroughly studied mouse model of schizophrenia, with disease phenotypes described at both the cellular and behavioral level (13-15). Using these mice to screen for potential therapeutic compounds will help to answer the question of whether a rescue at the cellular-morphological level also has effects on abnormal mouse behavior. This question is critical as behavior, alongside cognitions and emotions, is one of the essential read-outs for a successful treatment of a psychiatric condition.

Limitations of study

One of the limitations of this study is the fact that a potential successful translation from the neuronal cell-culture system to a living mouse would still be far away from a successful translation to human patients with 22q11.2 DS-associated SCZ. Whether human patients display this neurodevelopmental phenotype would have to be further evaluated in 2D or 3D neuronal cultures grown from human iPSCs derived from patients with 22q11.2 DS- associated SCZ. Nonetheless, a successful translation from cells to living mice would make this study a promising proof-of-concept study that shows that cellular systems can be used for subsequent testing

in living organisms. Another potential limitation is that this phenotype has its origin early in embryonic development, whereas SCZ associated with 22q11.2DS has its onset later during adolescence (58), raising questions about the relationship between this cellular phenotype and disease onset. Possibly, the cellular phenotype could be related as well to other, more early-onset neurodevelopmental phenotypes that are part of the phenotypic spectrum of 22q11.2DS, such as intellectual disability, ASD and ADHD (59).

Future directions

As outlined in the discussion section of Chapter 4, in case a compound that rescues the cellular phenotype is discovered, it will be subsequently tested in *Df(16)A*$^{+/-}$ mice to investigate whether a behavioral phenotypes are rescued as well. The next challenge will be to make the translation to a human patient-derived neuronal system. There are several important issues to take into consideration to make this translational step. First, the presence of the impaired dendritic arborization phenotype has to be established and quantified in the human system. Second, the rescuing compounds from the drug screen could be tested first on the human neuronal system. This step will provide insight into whether human and mouse disease biology are in fact overlapping and can be targeted in similar ways. If this is not the case, the entire drug screening procedure should be repeated in the human system in order to identify compounds that can rescue the phenotype there. Third would be the translation of a candidate compound to human patients according to FDA and EMA regulatory procedures. Ideally, a rescuing compound in the mouse cellular system will also have rescuing properties in *Df(16)A*$^{+/-}$ mice as well as the human iPSC-derived neuronal system. Hypothetically, this would create the largest a-priori chance of successfully targeting psychiatric symptoms in human patients. However, in case disease biology does substantially differ between humans and mice, different routes to successful translation to the clinic are a possibility as well.

CONCLUDING REMARKS

Understanding and treating mental illness continues to be an enormous challenge that has occupied large numbers of scientists across the world for a long time. Despite the initial high expectations in the late 20th century, the advances in both the genomics and neuroscience fields have thus far not led to improved interventions that positively impact the lives of patients. Although perhaps disappointing to some in the field of psychiatry, the amount of progress that is being made in both fields has increased our knowledge on the nature of these disorders to such an extent that there is reason to be hopeful that patients will eventually benefit from this investment being made. Currently, our understanding is greatest of rare, genetically driven neurodevelopmental disorders, and perhaps patients suffering from these will be the first to receive novel treatments. Improved treatments for the majority of patients encountered in

the clinic that do not have a monogenetic disease etiology may require more patience. In the meantime, investing enough resources in strategies to improve public mental health remains highly important and should occur in parallel as much as possible.

REFERENCES

1. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. 2019;22(3):343-52.
2. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668-81.
3. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000;157(10):1552-62.
4. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA. 2003;289(23):3095-105.
5. Gururajan A, Reif A, Cryan JF, Slattery DA. The future of rodent models in depression research. Nat Rev Neurosci. 2019;20(11):686-701.
6. Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50(3):381-9.
7. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet. 2009;373(9659):234-9.
8. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 2003;60(12):1187-92.
9. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. PLoS Med. 2005;2(5):e141.
10. Tarjinder Singh BMN, Mark J. Daly. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. medRxiv 2020091820192815;.
11. Mukai J, Cannavo E, Crabtree GW, Sun Z, Diamantopoulou A, Thakur P, et al. Recapitulation and Reversal of Schizophrenia-Related Phenotypes in Setd1a-Deficient Mice. Neuron. 2019;104(3):471-87 e12.
12. Emamian ES, Hall D, Birnbaum MJ, Karayiorgou M, Gogos JA. Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia. Nat Genet. 2004;36(2):131-7.
13. Mukai J, Dhilla A, Drew LJ, Stark KL, Cao L, MacDermott AB, et al. Palmitoylation-dependent neurodevelopmental deficits in a mouse model of 22q11 microdeletion. Nat Neurosci. 2008;11(11):1302-10.
14. Mukai J, Tamura M, Fenelon K, Rosen AM, Spellman TJ, Kang R, et al. Molecular substrates of altered axonal growth and brain connectivity in a mouse model of schizophrenia. Neuron. 2015;86(3):680-95.
15. Xu B, Hsu PK, Stark KL, Karayiorgou M, Gogos JA. Derepression of a neuronal inhibitor due to miRNA dysregulation in a schizophrenia-related microdeletion. Cell. 2013;152(1-2):262-75.
16. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431-44.
17. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. Nucleic Acids Res. 2009;37(Database issue):D832-6.
18. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nat Neurosci. 2019;22(12):1961-5.

19. Sestan N, State MW. Lost in Translation: Traversing the Complex Path from Genomics to Therapeutics in Autism Spectrum Disorder. Neuron. 2018;100(2):406-23.
20. Hornberg H, Perez-Garci E, Schreiner D, Hatstatt-Burkle L, Magara F, Baudouin S, et al. Rescue of oxytocin response and social behaviour in a mouse model of autism. Nature. 2020;584(7820):252-6.
21. Sato A, Kasai S, Kobayashi T, Takamatsu Y, Hino O, Ikeda K, et al. Rapamycin reverses impaired social interaction in mouse models of tuberous sclerosis complex. Nat Commun. 2012;3:1292.
22. Penagarikano O, Lazaro MT, Lu XH, Gordon A, Dong H, Lam HA, et al. Exogenous and evoked oxytocin restores social behavior in the Cntnap2 mouse model of autism. Sci Transl Med. 2015;7(271):271ra8.
23. van Os J, Hanssen M, Bijl RV, Ravelli A. Strauss (1969) revisited: a psychosis continuum in the general population? Schizophr Res. 2000;45(1-2):11-20.
24. consortium C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 2015;523(7562):588-91.
25. Barnett IJ, Lee S, Lin X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. Genet Epidemiol. 2013;37(2):142-51.
26. Courchesne E, Pramparo T, Gazestani VH, Lombardo MV, Pierce K, Lewis NE. The ASD Living Biology: from cell proliferation to clinical phenotype. Mol Psychiatry. 2019;24(1):88-107.
27. Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Brauninger M, et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. Proc Natl Acad Sci U S A. 2015;112(51):15672-7.
28. Velasco S, Kedaigle AJ, Simmons SK, Nash A, Rocha M, Quadrato G, et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. Nature. 2019.
29. Peng G, Fan Y, Palculict TB, Shen P, Ruteshouser EC, Chi AK, et al. Rare variant detection using family-based sequencing analysis. Proc Natl Acad Sci U S A. 2013;110(10):3985-90.
30. Robert F, Pelletier J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. Front Genet. 2018;9:507.
31. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, et al. Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. N Engl J Med. 2006;354(13):1370-7.
32. Flint J, Kendler KS. The Genetics of Major Depression. Neuron. 2014;81(5):1214.
33. Kendler KS. What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. Mol Psychiatry. 2013;18(10):1058-66.
34. McClellan J, King MC. Genetic heterogeneity in human disease. Cell. 2010;141(2):210-7.
35. Civelli O, Reinscheid RK, Zhang Y, Wang Z, Fredriksson R, Schioth HB. G protein-coupled receptor deorphanizations. Annu Rev Pharmacol Toxicol. 2013;53:127-46.
36. Ngo T, Kufareva I, Coleman J, Graham RM, Abagyan R, Smith NJ. Identifying ligands at orphan GPCRs: current status using structure-based approaches. Br J Pharmacol. 2016;173(20):2934-51.
37. Anderson GR, Galfin T, Xu W, Aoto J, Malenka RC, Sudhof TC. 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. 2012;109(44):18120-5.
38. Lie DC, Colamarino SA, Song HJ, Desire L, Mira H, Consiglio A, et al. Wnt signalling regulates adult hippocampal neurogenesis. Nature. 2005;437(7063):1370-5.
39. Ryskalin L, Lazzeri G, Flaibani M, Biagioni F, Gambardella S, Frati A, et al. mTOR-Dependent Cell Proliferation in the Brain. Biomed Res Int. 2017;2017:7082696.
40. Bae SM, Hong JY. The Wnt Signaling Pathway and Related Therapeutic Drugs in Autism Spectrum Disorder. Clin Psychopharmacol Neurosci. 2018;16(2):129-35.

41. Winden KD, Ebrahimi-Fakhari D, Sahin M. Abnormal mTOR Activation in Autism. Annu Rev Neurosci. 2018;41:1-23.
42. Dewey EB, Taylor DT, Johnston CA. Cell Fate Decision Making through Oriented Cell Division. J Dev Biol. 2015;3(4):129-57.
43. Bralten LB, Gravendeel AM, Kloosterhof NK, Sacchetti A, Vrijenhoek T, Veltman JA, et al. The CASPR2 cell adhesion molecule functions as a tumor suppressor gene in glioma. Oncogene. 2010;29(46):6138-48.
44. Takita J, Chen Y, Kato M, Ohki K, Sato Y, Ohta S, et al. Genome-wide approach to identify second gene targets for malignant rhabdoid tumors using high-density oligonucleotide microarrays. Cancer Sci. 2014;105(3):258-64.
45. Jordan BA, Kreutz MR. Nucleocytoplasmic protein shuttling: the direct route in synapse-to-nucleus signaling. Trends Neurosci. 2009;32(7):392-401.
46. Fang WQ, Chen WW, Jiang L, Liu K, Yung WH, Fu AKY, et al. Overproduction of upper-layer neurons in the neocortex leads to autism-like features in mice. Cell Rep. 2014;9(5):1635-43.
47. Levy DR, Tamir T, Kaufman M, Parabucki A, Weissbrod A, Schneidman E, et al. Author Correction: Dynamics of social representation in the mouse prefrontal cortex. Nat Neurosci. 2020;23(4):594.
48. Lauber E, Filice F, Schwaller B. Dysregulation of Parvalbumin Expression in the Cntnap2-/- Mouse Model of Autism Spectrum Disorder. Front Mol Neurosci. 2018;11:262.
49. Zhang W, Ma L, Yang M, Shao Q, Xu J, Lu Z, et al. Cerebral organoid and mouse models reveal a RAB39b-PI3K-mTOR pathway-dependent dysregulation of cortical development leading to macrocephaly/autism phenotypes. Genes Dev. 2020;34(7-8):580-97.
50. Zhou T, Tan L, Cederquist GY, Fan Y, Hartley BJ, Mukherjee S, et al. High-Content Screening in hPSC-Neural Progenitors Identifies Drug Candidates that Inhibit Zika Virus Infection in Fetal-like Organoids and Adult Brain. Cell Stem Cell. 2017;21(2):274-83 e5.
51. Durens M, Nestor J, Williams M, Herold K, Niescier RF, Lunden JW, et al. High-throughput screening of human induced pluripotent stem cell-derived brain organoids. J Neurosci Methods. 2020;335:108627.
52. Rugg-Gunn PJ, Ferguson-Smith AC, Pedersen RA. Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. Hum Mol Genet. 2007;16 Spec No. 2:R243-51.
53. Li XY, Jia Q, Di KQ, Gao SM, Wen XH, Zhou RY, et al. Passage number affects the pluripotency of mouse embryonic stem cells as judged by tetraploid embryo aggregation. Cell Tissue Res. 2007;327(3):607-14.
54. Bauwens CL, Peerani R, Niebruegge S, Woodhouse KA, Kumacheva E, Husain M, et al. Control of human embryonic stem cell colony and aggregate size heterogeneity influences differentiation trajectories. Stem Cells. 2008;26(9):2300-10.
55. Cederquist GY, Asciolla JJ, Tchieu J, Walsh RM, Cornacchia D, Resh MD, et al. Specification of positional identity in forebrain organoids. Nat Biotechnol. 2019;37(4):436-44.
56. Mansour AA, Goncalves JT, Bloyd CW, Li H, Fernandes S, Quang D, et al. An in vivo model of functional and vascularized human brain organoids. Nat Biotechnol. 2018;36(5):432-41.
57. Bhaduri A, Andrews MG, Kriegstein AR, Nowakowski TJ. Are Organoids Ready for Prime Time? Cell Stem Cell. 2020;27(3):361-5.
58. Bassett AS, Chow EW, AbdelMalik P, Gheorghiu M, Husted J, Weksberg R. The schizophrenia phenotype in 22q11 deletion syndrome. Am J Psychiatry. 2003;160(9):1580-6.
59. Fiksinski AM, Schneider M, Murphy CM, Armando M, Vicari S, Canyelles JM, et al. Understanding the pediatric psychiatric phenotype of 22q11.2 deletion syndrome. Am J Med Genet A. 2018;176(10):2182-91.

Summary

Mental illness is a major contributor to global disease burden. The prevention and treatment of mental illness poses a great challenge that will require a more thorough understanding of the functioning and dysfunctioning human brain to ultimately enable the development of more efficacious treatments. The human genome is a well-suited starting point for understanding the neurobiology of mental illness, as the major psychiatric disorders show a relatively high degree of heritability. Genetic advances in the past few decades have revealed a genetic architecture that is highly polygenic, with common variants of small effect and rare variants of larger effect sizes each contributing significantly. Elucidating the genetic architecture of mental illness is merely the first step in working towards understanding their neurophysiological underpinnings. Functional studies using rodent models with genetic modifications in mental illness-associated genes have been highly useful for understanding disease biology, but have thus far not yield novel treatments. The more recent development of using induced pluripotent stem cells from human patients hold promise for elucidating critical human-specific disease processes that can be further investigated for drug discovery purposes.

Chapter 1 of this thesis is a general introduction to the three main studies included in this work. The introduction spans from a background on the genetic architecture of mental illness to a background on various study methods used to investigate disease-associated genes functionally.

Chapter 2 is a gene discovery study in which an extended pedigree from the Old Order Mennonite population (a genetic isolate) with a high prevalence of major depressive disorder is investigated to test the hypothesis of whether a rare genetic variant with large effect size contributes to this phenotype. Whole exome sequencing was used to identify a rare missense variant in *GPR156* that increases the risk for major depressive disorder by approximately three-fold. *GPR156* is an orphan g-protein coupled receptor that is highly expressed in the habenula, a brain region shown to play a critical role in mood regulation. Insertion of the human mutation in *GPR156* in mice results in abnormal emotional behavior that is rescued with conventional antidepressant medications.

Chapter 3 is a disease modeling study that deploys the use of induced pluripotent stem cells from patients with a rare syndromic form of autism spectrum disorder caused by a homozygous loss-of-function mutation in *CNTNAP2.* The goal of this study is to uncover a disease phenotype that can be investigated further to identify a drug-targetable disease mechanism. Patients with this this mutation display an increase in head circumference and gray matter volume. Forebrain organoids generated from the stem cell lines from these patients displayed an overgrowth phenotype characterized by increased neural progenitor proliferation, resulting in an increase in total cell number and an increased total organoid volume. Differentially expressed genes uncovered by both bulk- and single-cell RNA sequencing corroborate aberran-

cies in biological processes including cellular proliferation and differentiation. Furthermore, differentially expressed genes are enriched for autism-associated genes and autism-associated gene expression correlation networks. Finally, the overgrowth phenotypes and associated gene expression signature was rescued after repair of the of the mutation in *CNTNAP2* using CRISPR-Cas9.

Chapter 4 is a study that sets the initial steps of generating a cell-based platform that can be used for *high-throughput* drug screening. The ultimate goal of this study is to identify a compound that can be used to treat patients with 22q11.2 deletion syndrome. Previous work has shown that mice lacking the genomic region syntenic to the 22q11.2 locus, *D(f)16*$^{+/-}$ mice, display deficit in dendritic and axonal outgrowth (i.e. neurite outgrowth). Neuronal cultures generated from embryonic stem cells from *D(f)16*$^{+/-}$ mice recapitulate this robust phenotype. Therefore, the neurite outgrowth phenotype will be used as a read-out to use for screening compounds in the following steps of this study.

Chapter 5 is a general discussion about the strengths, shortcomings and future directions of these three studies and their respective fields.

Samenvatting

Psychiatrische aandoeningen leveren een grote bijdrage aan de wereldwijde ziekte last. Daarom zijn de behandeling en preventie van psychiatrische aandoeningen wereldwijd een belangrijke uitdaging. Om deze uitdaging aan te gaan is er groter begrip van de werkende en niet-werkende menselijke hersenen nodig, zodat er uiteindelijk effectievere behandelingen ontworpen kunnen worden. Het menselijk genoom is een geschikt beginpunt om de biologie van psychiatrische aandoeningen te ontcijferen, omdat deze aandoeningen in hoge mate erfelijk zijn. Vooruitgangen in de genetica de afgelopen paar decennia hebben laten zien dat de genetische architectuur van psychiatrische aandoeningen een hoge mate van polygeniciteit vertoont, waarbij zowel veelvoorkomende genetische varianten met kleine effectgroottes en zeldzame genetische varianten met grotere effectgroottes een rol spelen. Het in kaart brengen van deze genetische architectuur is slechts de eerste stap in het ontcijferen van de onderliggende biologie van de aandoeningen die ze veroorzaken. Studies die gebruik maken van muizen of ratten met modificaties in ziekte-geassocieerde genen hebben veel kennis opgeleverd over ziekte-biologie, maar hebben tot noch toe niet geleid tot nieuwe werkzame behandelingen. De recentere ontwikkeling van het gebruik van geïnduceerde pluripotente stamcellen die verkregen worden van patienten met psychiatrische aandoeningen houden de belofte om ziekteprocessen die mens-specifiek zijn op te helderen. Deze kunnen in de toekomst verder onderzocht worden in de zoektocht naar nieuwe effectievere behandelingen.

Hoofdstuk 1 van dit proefschrift is een algemene introductie over de drie onderzoeksprojecten in dit werk. In de introductie wordt achtergrond informatie gegeven over zowel de genetische architectuur van psychiatrische aandoeningen als verschillende methoden om ziekte-geassocieerde genen te bestuderen om hun functie en geassocieerde ziekte-mechanisme te verklaren.

Hoofdstuk 2 is een genetische familie-studie waarin een familie behorende tot de traditionele Mennotienen (een genetisch geïsoleerde populatie) met een hoge prevalentie van depressie wordt bestudeerd om te toetsen of er mogelijk een zeldzame gen-variant met hoge effect-grootte bijdraagt aan dit fenotype. *Whole exome sequencing* werd gebruikt om een zeldzame *missense* variant in het gen *GPR156* aan te tonen. De aanwezigheid van deze gen variant zorgt voor een drievoudige verhoging van het risico op het hebben van depressie. *GPR156* codeert voor een nog onbekende g-eiwit-gekoppelde receptor die in hoge mate tot expressie wordt gebracht in de habenula, een hersengebied met een kritieke rol in stemmingsregulatie. Het inbrengen van de menselijke mutatie in een muizenlijn zorgt voor afwijkend emotioneel gedrag in deze muizen, wat normaliseert als de muizen behandeld worden met conventionele antidepressiva.

Hoofdstuk 3 is een *disease-modeling* studie waarin gebruik wordt gemaakt van geïnduceerde pluripotente stamcellen die verkregen worden van patiënten met een zeldzame syndromale vorm van autisme spectrum stoornis die veroorzaakt wordt door een homozygote *loss-of-function* mutatie in het gen *CNTNAP2*. Het doel van deze studie is het identificeren van een

pathofysiologisch fenotype dat in toekomst verder mechanistisch opgehelderd kan worden zodat er vervolgens farmacologisch op geïntervenieerd kan worden. Patiënten met deze mutatie vertonen een vergrote hoofdomtrek en een toegenomen volume van grijze stof in de hersenen. Corticale organoïden die gegenereerd werden van deze patiënten vertonen tekenen van overgroei dat gekarakteriseerd wordt door toegenomen proliferatiesnelheid van neuronale voorlopercellen, wat leidt tot een toegenomen hoeveelheid cellen en toegenomen volume van de organoïden. *Single-cell-* en *bulk-RNA sequencing* werden gebruikt om genen te identificeren die in verschillende mate tot expressie komen in patient-verkregen organoïden. Een deel van deze genen zijn betrokken in biologische processen zoals cel proliferatie en differentiatie en zijn daarmee in lijn met het overgroei-fenotype. Een subset van dezelfde differentieel geëxpresseerde genen vertonen overlap met zowel autisme spectrum stoornis geassocieerde genen als met genen die deel uitmaken van autisme-geassocieerde gen expressie netwerken. Het repareren van de pathogene mutatie met CRISPR-Cas9 resulteerde in een normalisatie van het overgroei fenotype en het daarmee gepaard gaande transcriptie regulatie patroon, waardoor de causaliteit tussen genotype en fenotype bevestigd werd.

Hoofdstuk 4 is een studie waarin de eerste stappen worden gezet tot het creëren van een platform dat gebruikt maakt van neuronale celculturen om te screenen voor werkzame medicijnen voor het 22q11 deletiesydroom op een *high-throughput* manier. Eerder werk toonde aan dat het muismodel van het 22q11 deletiesyndroom, de *D(f)16*$^{+/-}$ muis, onder andere defecten vertoont in zowel dendritische als axonale uitgroei (gezamenlijk neuriet-uitgroei). Dit fenotype werd gereproduceerd in de neuronale culturen die gegenereerd werden van de embryonische stamcellen van *D(f)16*$^{+/-}$ muizen. Omdat dit een robuust fenotype is op cellulair niveau is het een geschikte uitkomstmaat om te gebruiken voor de medicijnscreening in de volgende stappen van deze studie.

Hoofdstuk 5 is een algemene discussie over de sterke- en zwakke punten en de toekomstperspectieven van deze drie studies en hun respectievelijke onderzoeksvelden.

Curriculum Vitae

Job de Jong was born in Amsterdam in 1988. He received his bachelor's degree in biology at the Vrije Universiteit in Amsterdam in 2011. He completed his education to become a physician-clinical researcher at Maastricht University in 2016. The same year he enrolled in the PhD program at the department of Psychiatry at Erasmus MC under the supervision of Prof. Dr. Steven Kushner, focusing on the role of rare genetic variation in psychiatric disorders. He conducted his doctoral research in the lab of Dr. Sander Markx and Dr. Bin Xu at Columbia University Medical Center in New York City. In 2020, he moved back to Amsterdam to work as a physician in psychiatry.

PhD Portfolio

Name PhD student: J. de Jong
Erasmus MC Department: Psychiatry
PhD period: September 2016 – January 2020

Promotor: Prof. dr. S.A. Kushner
Co-promotor : Dr. S. Markx

No course work was done at Erasmus MC as all research was carried out abroad.

Other research related activities

New York Stem Cell Symposium. Invited speaker, talk on 'Cortical Overgrowth in a Preclinical Forebrain Organoid Model of CNTNAP2-Associated Autism Spectrum Disorder'. September 2019

Attended Columbia Stem Cell Initiative (CSCI) Summit in May 2018.

Attended NYSCF conference – New York Stem Cell Foundation in New York City in September 2018

Attended NYSCF conference – New York Stem Cell Foundation in New York City in September 2017

Translational Therapeutics Drug Development Bootcamp, Columbia University, New York City. Project title: Using antisense oligonucleotides as a new therapeutic modality for psychiatric and cognitive disorders. Principal investigator: Dr. Joseph Gogos. February 2017

Publications

de Jong JO, Llapashtica C, Genestine M, Strauss K, Provenzano F, Sun Y, Zhu H, Cortese GP, Brundu F, Brigatti KW, Corneo B, Migliori B, Tomer R, Kushner SA, Kellendonk C, Javitch JA, Xu B, Markx S. Cortical Overgrowth in a Preclinical Forebrain Organoid Model of CNTNAP2-Associated Autism Spectrum Disorder. Accepted for publication at *Nature Communications*. Preprint available on bioRxiv. doi.org/10.1101/739391

Miller BR[*], Gonzaga-Jauregui C[*], Brigatti KW[*], **de Jong JO[#],** Breese R[#], Brydges S, Carey D, Dwork AJ, Economides AN, Elmwood D, Endicott J, First MB, Gregoire HJ, Hen R, McCarthy S, Overton JD, Pefanis E, Puffenberger EG, Rojas J, Rosoklija G, Shuldiner AR, Staples J, Van Hout C, Young M, Alessandri-Haber N, Markx S, Strauss KA. A Rare Missense Variant of the Orphan G Protein-Coupled Receptor Gene GPR156 is Associated with Major Depressive Disorder in Old Order Mennonites. Manuscript submitted.

Cortese G, Brundu F, de **Jong JO** , Gogos J, Markx S. Generating a high-throughput drug screening platform for 22q11.2 deletion syndrome using mouse embryonic stem cell-derived neuronal cultures from Df(16)A+/- mice. Manuscript in preparation.

Giniatullina A, Maroteaux G, Geerts CJ, Koopmans B, Loos M, Klaassen R, Chen N, van der Schors RC, van Nierop P, Li KW, **de Jong JO**, Altrock WD, Cornelisse LN, Toonen RF, van der Sluis S, Sullivan PF, Stiedl O, Posthuma D, Smit AB, Groffen AJ, Verhage M. Functional characterization of the PCLO p.Ser4814Ala variant associated with major depressive disorder reveals cellular but not behavioral differences. Neuroscience. 2015 Aug 6;300:518-38.

de Jong JO, Arts B, Boks MP, Sienaert P, van den Hove DL, Kenis G, van Os J, Rutten BP. Epigenetic effects of electroconvulsive seizures. The journal of ECT. 2014;30(2):152-9.

* Shared first-authorship
# Shared second-authorship

Acknowledgements

I'd like to thank my promotor Steven Kushner for supervising me through these exciting scientific projects. Although we were on different continents for most times during my PhD-trajectory, your long-distance scientific input was always incredibly helpful and provided me with critical insights that helped the projects move forward.

I'd like to thank Sander Markx and Bin Xu for being great PIs and mentors in the lab where most of the research was conducted. From both of you I received great mentorship in a broad range of subjects.

Sander, thank you for introducing me to translational science in psychiatry and getting me excited about these projects. As an aspiring MD-PhD, it was inspiring to work on scientific projects that involved interactions with human research participants while also working on patient-derived lab-grown brain tissue in order to understand embryonic disease mechanisms of a genetic condition. I've learned a great deal from you about psychiatric diagnostics, genetics and disease modeling. Your trust in my abilities made me grow as a researcher.

Bin, you have been a great mentor in the lab, supervising me in learning to do CRISPR gene-editing, stem cell culture, growing brain organoids and doing -omics bioinformatics analyses. The learning curve was quite steep coming from medical school and with limited experience in the lab. Your patience, kindness and humor allowed me to learn all these things at my own pace. We've had a lot of interesting discussions about brain development, genetics and designing experiments. Working with you opened up methods of thinking that were unknown to me before.

I'd like to give a special thanks to all the people in the Plain Community that participated in this research. It has been a privilege to have been able to work with the Amish and Mennonite families. I've admired the people from this community for their kindness, openness and welcoming attitude towards participating in these tedious and sometimes difficult studies. I had a lot of fun learning about the people through their personal stories and thereby gaining insight into the transcultural aspects of psychiatric diagnostics in this interesting community.

The work in this thesis had not been possible without the very kind people from the Clinic for Special Children in Strasbourg, PA. In the early phases of the genetic study I spent some time at the clinic to work on psychiatric assessments. Seeing how the people in the clinic worked both for and with the people from the Plain Community was refreshing to me. Participating in family-meetings whose purpose was to learn from, educate and share acquired scientific knowledge from these studies with the plain community was a great experience. Kevin Strauss, thank you for your co-mentorship on these projects; I learned a great deal from your doings

as a clinician-scientist and director of the clinic. Thank you Eric, Karlla, Matt, Adam, Donna, Christine and Millie; it was great working with you all.

Brad Miller, thanks for all the trips we made together to Pennsylvania, upstate New York and Ohio. It was great times being on the road with you, staying in weird motels and visiting families all around the country. I learned a lot from you about diagnostic interviewing in psychiatry.

Michael First and Jean Endicott, thank you for all the times spent during the diagnostic consensus meetings. These meetings were very helpful to me in learning how to conduct structured psychiatric diagnostics in a research study context. Michael, your thoroughness and attention to details still helps me currently when I'm doing diagnostic interviewing in a clinical setting. Jean, your knowledge on the transcultural aspects of diagnostic interviewing in the Plain Community was critical for making a valid diagnosis and taught me how to incorporate these aspects into the diagnostic process.

Thank you to everyone who I worked with in the lab of Sander and Bin. Ceyda, Yan, Vivian, Kevin, Yanru, Giuseppe, Liza and Anthony, it was so good working with you.

I'd like to thank Jonathan Javitch and Christoph Kellendonk for their co-mentorship during my research projects. Your input was always interesting and valuable and it was useful to receive mentorship from several scientific angles. I also enjoyed learning from attending your lab meetings as they are in such different fields than the work in this thesis.

Thanks to all the great people that worked in the molecular therapeutics division at NYSPI during my time there. Times were always fun and I learned a lot of things through you all. Caline, Audrey, Signe, Julia, Josh, Wes, Marie,Ying, Maria, Mike, Zack, Trang, Pattama, Troels, Pedro, Sara, Kelly, Brenna, Cory, Guowei, Laura, Lillian, Mark, Matthias, Melissa, Sandra, Sarah, Tisha, Xin, Yuchao. I miss hanging out in the kitchen with all of you and going to bars. Sasha, Doug and Jason: thanks for all your help getting the research done and for being great friends.

Mariëlle Caspers, thank you for all your help during the final phases of the trajectory.

Most importantly, I'd like to thank the people in the US and the Netherlands that I love and that were there for me both when times were good and bad. My dear family; Wilma, Willem, Joop, Ria, Sam, Bess, Freya, Rufus and Willemijn. A big thanks to my lovely friends Pia, Jo, Sean, Simone, Marlous, the people from 545, Simon, Boas, Giulia, Wester, Thomas, Felicia, Daan, Kay, Fred, Pepijn, Coco, Abel, Steven, Romy, Eddy, Emma, Sem, Frans & Brent, without you, no science.