Cellular Phenotypes Associated with *NRXN1* Mutations in Autism; an iPSC Study

Aicha Massrali

Department of Psychiatry

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Wolfson College

August 2020

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgements, the Preface, and specified in the text. It is not substantially the same as any that I have submitted or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. This dissertation does not exceed the prescribed word limit of 60,000 words set by the Degree Committee for Clinical Medicine and Clinical Veterinary Medicine.

Aicha Massrali

August 2020

Acknowledgements

First and foremost, I would wholeheartedly like to express my sincere gratitude towards my supervisor Prof. Simon Baron-Cohen for giving me the opportunity to undertake this project and being a constant source of support, motivation and inspiration throughout the course of my degree. Many thanks also to Dr. Mark Kotter who welcomed me in his team and provided me with all the technical facilities and support I needed to complete my experiments.

I am truly indebted to Dr. Deepak Srivastava King's College, London, for his guidance and feedback regarding experiments, analysis and thesis writing and most importantly for his insight that have inspired and taught me so much.

I am very thankful for Dwaipayan Adhya for being extremely generous with his time, revision and advice throughout the writing of this thesis. I feel so fortunate for having the best lab mate anyone can ask for, Arkoprovo Paul, his kindness and support helped get through the toughest times of this PhD.

My sincere gratitude for the Youssef Jameel Scholarship and the Autism Research Trust for funding my studentship and bearing the costs of my project. I am also thankful to Wolfson college for providing a warm, cosy and an intellectually stimulating environment throughout my stay there.

I was very lucky to have amazing colleagues across three labs, the autism research centre, lab of regenerative medicine at Cambridge and Srivastava's lab a at King's college London. Special thanks to Nataly Martynyuk, Koby Baranes, Joana Tavares, Varun Warrier, Helena Brunel, Nick Gatford and Lucia Dutan.

I was very lucky to have had a brilliant student joining this project for a short but a very critical period; Rugile Matulevičiūtė was very quick to learn, very keen to work and very pleasant to be around.

I would like to especially thank the Karim's (Muhammed-Kaiser and Zahabiya), the golden couple that has given me so much love, care and support.

Genuine heartfelt gratitude towards Matt Wayland and the department of Zoology at the University of Cambridge who has not only hosted me during my long live imaging experiments but also helped me with data analysis and interpretation.

I cannot express my appreciation to my 'squad', my lifelong friends who have always been my support system and comfort zone (even from overseas) since I was ten: Nour Shouhayeb, Kamila Bolbol, Sarah Wasfy and Salma Yaken. Without the friends I made during my PhD, my experience would not have been the same: Tatiana Vilsbol, Nadya Pohran, Ibrahim Khalil, Faraj Alnasser, Mie Mnt and Sol Aki.

A special thanks to my special someone, Kouceila Bouali whose love, patience and support helped me not only survive but also cherish the hardest times of this journey.

My deepest gratitude goes to my mom who is my hero, my inspiration and my wonder woman. My dad for being the most loving and caring father in the world and my brother who has always got my back.

All my work and ambition would mostly be dedicated to Samy, my cousin who has taught me that love needs no words.

Table of Contents

List of Figures

List of Tables

Abstract

Autism spectrum conditions are neurodevelopmental conditions that entail social-communication difficulties, unusually narrow interests and difficulties adjusting to unexpected change. While the role of *NRXN1* has been established in regulating normal synaptic function and physiology, its contribution to the aetiology of autism is only now emerging. However, to fully understand how *NRXN1* influences neuronal phenotypes, synaptic and functional changes a reliable and relevant human cellular model is needed. In this thesis I have used induced neurons derived from induced pluripotent stem cells (iPSC) using the OPTi-OX system to study the impact of *NRXN1* mutations on excitatory glutamatergic neurons. I specifically study the molecular, morphological and functional changes caused by 1) induced mutations in CRISPR-edited cell lines; and 2) iPSCs derived from autistic individuals with NRXN1 mutations. The hypothesis is that intronic mutations in the patient lines and exonic mutations in the mutant lines have distinct effects on molecular, morphological and functional activity of glutamatergic neurons. NRXN1-patient neurons displayed distinct patterns of *NRXN1* isoform expression. Patient line 211_NXM had a significantly higher levels of *NRXN1* expression compared to patient line 092_NXF. 211_NXM neurons also displayed significantly higher firing rates as determined by MEA recordings. Conversely, patient line 092-NXF had *NRXN1* expression and neurons with aberrant neurite outgrowth. I also generated four different NRXN1-mtant lines. Each NRXN1-mutant line exhibited distinct expression levels of *NRXN1* and its isoforms. Furthermore, each line displayed disrupted neurite outgrowth to varying degrees. Finally, RNA-Seq analysis revealed notable upregulation for genes enriched for cell adhesion biological functions in the patient lines and chemical synaptic transmission pathways in the mutant lines. Of particular interest were enrichment of differentially regulated genes in collagen and cadherin family genes, consistent with alterations in adhesion properties in neurons with NRXN1-mutations. The strength of the thesis lie in the technological advancements of forward programming and iPSC

technologies, however further experiments with a bigger sample size might be required to validate and support the findings described. Thus, these data contribute to our overall understanding of how NRXN1 may contribute to the aetiology of autism and demonstrates the complexity of its effects. In the second part of this thesis, I investigate the association between DNA methylation at birth (cord blood), and scores on the Social and Communication Disorders Checklist (SCDC), a measure of autistic traits, in 701 8-year-olds, by conducting a methylome-wide association study (MWAS). Using methylation data for autism in peripheral tissues, we did not identify a significant concordance in effect direction of CpGs with p value < 10–4 in the SCDC MWAS (binomial sign test, p value > 0.5). In contrast, using methylation data for autism from post-mortem brain tissues, a significant concordance in effect direction of CpGs with a p value < 10–4 in the SCDC MWAS (binomial sign test, p value = 0.004) was reported. Supporting this, I observe an enrichment for genes that are dysregulated in the post-mortem autism brain (one-sided Wilcoxon rank-sum test, p value=6.22×10–5). Finally, integrating genome-wide association study (GWAS) data for autism (n = 46,350) with mQTL maps from cord-blood (n = 771), we demonstrate that mQTLs of CpGs associated with SCDC scores at p value thresholds of 0.01 and 0.005 are significantly shifted toward lower p values in the GWAS for autism ($p < 5 \times 10–3$). Addition support for this using a GWAS of SCDC demonstrate a lack of enrichment in a GWAS of Alzheimer's disease. These results highlight the shared cross-tissue methylation architecture of autism and autistic traits and demonstrate that mQTLs associated with differences in DNA methylation associated with childhood autistic traits are enriched for common genetic variants associated with autism and autistic traits.

"Not all the features of atypical human operating systems are bugs."

By Steven Silberman[1]

[1] NeuroTribes: The Legacy of Autism and the Future of Neurodiversity

Chapter 1

Introduction

1 Introduction

1.1 Autism History

In 1943 Leo Kanner wrote an article entitled “Autistic disturbances of affect contact” in *The Nervous Child* in which he described eight boys and three girls (2-8 years old) who were struggling with communication and exhibiting repetitive and obsessive behaviour [1]. Seeing that difficulties in communication and in language were also seen in other neurological conditions, Kanner decided to exclude children with known conditions, brain injuries or intellectual disabilities. This led to the classification of autism as a rare disorder seen in no more than 0.04% of the population [1]–[3].

Before the use of the term “infantile autism”, this syndrome was categorised as a form of childhood schizophrenia (Infantile Psychosis) until 1980, when autism was listed as a subgroup within “Pervasive Developmental Disorders” (PDD), a term that was introduced in the Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III) [3], [4]. This rendered the diagnostic criteria ambiguous as there were no clarity in identifying different PDDs [4]. In the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), which is the latest edition of DSM, published in 2013, Asperger syndrome, and pervasive developmental disorder not otherwise specified (PDD-NOS) all fall under the umbrella of Autism Spectrum Disorders (ASD) [5].

The use of the term Autism spectrum conditions (ASC) instead of ASD is to avoid the pejorative connotation of the word 'disorder' given that, whilst some characteristics of autism are disabling, others (such as excellent attention to and memory for detail) may even confer talent [6]. ASC are now defined as early-onset neurodevelopmental syndromes and entail social-communication disability alongside unusually narrow interests and difficulties adjusting to unexpected change. These behavioural characteristics are often, but not always, accompanied by atypical language development and/or learning difficulties, as well as anxiety disorders [6]. To bypass the debate about whether to use ASD or ASC, many authors are now simply using the term 'autism' as a succinct, neutral umbrella term to encompass those on the autism spectrum. Importantly, since 2013, DSM now also includes atypical sensory processing, including sensory hyper-sensitivity, as a diagnostic feature of autism.

1.2 Autism Prevalence

Worldwide prevalence of autism is currently 1-2%, thus much more prevalent than previously recognised [7]–[10]. The number of diagnosed autism cases has been steadily increasing since the term was coined (70 years ago) with a substantial rise in the past two decades. This high and sudden rise can be attributable to better diagnostic criteria and greater public awareness and not to an actual increase in prevalence [8]. General population studies reported no significant change in the prevalence of autism in childhood and adulthood which supports the notion that the increase of reported cases doesn't reflect an actual increase in prevalence [4].

1.3 Autism Etiology

Autism is a complex condition in which the symptoms vary from mild to severe and phenotypes are extremely diverse. Autism can be defined as a multifactorial condition that results from an interaction between the genetic background and environmental factors, leading to disruption in typical neural development and connectivity [11].

1.3.1 Heritability

The earliest twin studies recognized the heritability in autism. In 1977, Folstein and Rutter's study of monozygotic (MZ) twins showed higher concordance than dizygotic (DZ) twins. In the late 1990s estimates for heritability was 91-93% [12]. More recently Taniai et al. (2008) and Rosenberg et al. (2009) reported the highest MZ concordances (88–95%) and DZ concordances of 31% [14]. The high heritability rate of autism shown by twin studies encouraged scientists to explore the molecular genetics involvement in autism aetiology.

Autism aetiology has not yet been fully understood and is predominantly "idiopathic" (unknown cause), however, an estimate of 20% of autism cases have a causative gene/mutation identified and are known as "monogenic syndrome".

- **Monogenic Syndromes** have been associated (co-occur) in 5-10% of autism cases. The most common co-occurring condition is Fragile X syndrome (FXS),

reported in more than 2% of autism cases [15]. FXS is cause by mutations in the FMR1 (Fragile X Mental Retardation 1), predominantly expressed in the brain and known for its regulatory role during synaptic plasticity [16]. Tuberous sclerosis complex (TSC) is another example of monogenic syndromes associated with autism. It is estimated that 1% of individuals with autism are also diagnosed with TSC. The mammalian target of the rapamycin signalling pathway (mTOR) plays a major role in local translation in the nervous system and is disrupted in in TSC individuals resulting from mutations in TSC1 and TSC2 (mTOR inhibitors) [17].

Angelman syndrome and Prader-Willi syndrome result from loss of maternal and paternal alleles, respectively of 15q11.2-q13.3. In some rare cases of Angelman syndrome, methylation of the imprinting control region is un-methylated and hemi-methylated in other rare case of Prader-Willi syndrome. While this locus is deleted in both syndromes, it is duplicated in autism [18], [19]. Rett Syndrome is another disorder that is closely related to autism (1% of autism females) and it is associated with a mutation in MeCP2 [20]. This methyl-CpG binding protein plays a crucial role in methylation-dependent gene silencing and its mutated form has been linked to autism cases along with other MBD family members such as MBD5 [21] and MBD3 & 4 [20]. Most recently MeCP2 showed increased binding to GAD1 and RELN. This binding is mediated by an enrichment of 5-hmC in the cerebellum in autism [22].

1.3.2 Autism Genetics

Many approaches have been used to identify genetic variants and chromosomal abnormalities, yielding hundreds of candidate genes, none of which account for more than 1% of autism cases [23]. Previously, biologists mainly relied on common Sanger and Maxam-Gilbert sequencing technology for genomic analysis. However, since the 2008, new sequencing technologies known as Next Generation Sequencing (NGS) emerged allowing massive parallel analysis with a high-throughput from multiple sample with a significantly reduced cost [24]. These approaches allow sequencing of the whole genome (whole genome sequencing) or parts of it (whole exome sequencing) or targeted panels at a much higher speed and much lower costs making them ideal for clinical applications [25].

Due to the lack of conclusive molecular diagnosis yielded by conventional molecular testing of patients with genetic conditions, several studies tested the feasibility of applying WES and WGS in the clinic. Increasing evidence shows the benefits of these methods in the clinic to improve diagnosis but the scientific community have also relied on these methods for the discovery of individual genetic variants that influence metabolite levels to provide insight into the molecular processes influencing disease pathogenesis [26]. Genome wide association studies relies on WGS to test genetic variants across the genomes of specific population samples (e.g., patient vs control) to identify genotype-phenotype association.

Post sequencing, association tests are used to identify regions specific for the tested phenotype at genome-wide significance, meta-analysis is then done to increase the statistical power of these associations. Functional characterization experiments are then done to confirm causal links (genotype-phenotype), especially in the non-coding genome. Risk variants identified by GWAS are classified into two main groups, 1) rare variants with small effect sizes (difficult to identify using GWAS), and 2) common variants with large effects (uncommon for complex conditions) [27]. The identification of rare variants is exceptionally challenging with genomic studies due to statistical power limitations. The exome, the protein coding region of the genome, is the best characterized and most conserved of the genome, and the source of most risk and causal variants for mendelian disorders. Which makes WES less expensive and more favourable for the identification of rare variants of large effect on disease than the WGS [28]–[30].

- **Chromosomal abnormalities** were reported in approximately 2-5% of autism cases [17], [31]. Structural abnormalities are not restricted to specific chromosomes, and they include deletions, duplications, inversions and translocations. Although most of those chromosomal anomalies are rare, some are recurrent and one important example is the 15q11q13 duplication (maternally derived) that has been reported in 1-3% of individuals with autism [32]. This region is of particular importance in brain development and function as it hosts important genes including *GABRA5* and *GABRB3* (GABA receptors), *UBE3A* and *HERC2* (components of the proteasome complex) and

SNRPN (ribonucleoprotein peptide N) as well as *CYFIP1* (the FMRP interacting protein) [15].

- **Copy number variations (CNV)** increase the likelihood of having autism in 5-10% of individuals. Some of these CNVs are inherited and some are de-novo (not detected in either parent or unaffected sibling). The CNVs associated with autism and found in the control population emphasise the variable and incomplete penetrance of those mutations [33]. Moreover, some of these CNVs have also been associated with other neuropsychiatric conditions including developmental delay, epilepsy, schizophrenia and ADHD (Attention Deficit and Hyperactivity Disorder), which supports that the notion that these conditions share a similar pathophysiological pathways and that the final phenotypes depend on the presence of additional variants within the individual's genome [34]. Conversely, de novo CNVs are estimated to occur in 4-7% of autism cases compared to less than 2% occurrence in unaffected siblings and controls [35]–[37]. In autism cases, de novo CNVs were identified in close proximity to genes associated with synaptic functions resulting [37].

- **Genome wide analysis studies (GWAS)** have identified a number of single nucleotide polymorphisms, in total those SNPs account for 20% of autism likelihood [38], [39]. De-novo mutations are found in 10-20% of cases and account for less than 5% of overall autism liability [40]. More interestingly, a recent GWAS implicated prenatal corticogenesis related genes in strong association with idiopathic autism [41]. The expression of the identified genes in this study showed peaks during

proliferation, neurogenesis, migration, and cortical layer formation beginning from the late first until the third trimesters but not strongly expressed postnatally [41].

- **Autism Genome Project (AGP)** is a large-scale sequencing project designed to unravel autism-associated genes and indeed it identified hundreds of rare mutations and genomic imbalances in autism cases [42]. Many of these genes had already been correlated with intellectual disability, this confirmed the suggested association between autism and development. The AGP also concluded that autism is "behavioural manifestation of tens or perhaps hundreds of genetic and genomic disorders" and is not a monogenic condition [42]. Further studies then investigated the association these genes and autism. The genes that were previously identified as causal genes for autism, turned out to be causal for other monogenic syndromes associated with autism (as described in the previous section "Monogenic syndrome") [42]. Moreover, most of the genes associated with autism also known as autism candidate genes were affecting neurodevelopment or neural pruning [43].

- **Mosaicism in Autism** refers to mosaic mutations: de novo mutations that occur in only a subset of the cells in the body as they occur after fertilisation. They have been implicated in cancer and a number of neurodevelopmental conditions [44]. Due to the complexity of the autism phenotypic outcome that is of behavioral nature, it is very challenging to confer the functional consequences of somatic mutations. A fair number of studies have shown a modest but consistent contribution of somatic

variants in autism. Freed et al (2016) and Lim et al (2017) suggested that up to 7.5% of de novo mutations in autism may in fact be postzygotic mosaic mutations [45], [46]. Another study in post-mortem autism brain identified mosaic mutations as potential contributors to autism [47]. Another study on the largest set of autism brain derived whole genome sequencing data identified risk-modifying somatic mutations, including CACNA1A and brain active enhancers which suggests that somatic mutations may contribute to risk through disruptions in the regulation of gene expression in addition to coding variation [48].

1.3.3 Autism Epigenetics

A potential contributor to the development of the autism spectrum lies in epigenetic mechanisms [23]. Epigenetics refers to the processes involved in potentially heritable changes in gene expression without altering the genetic code. The epigenome is the catalogue of the chemical modifications that have been added to the individual's genome. Those modifications arise from heritable modifications (in response to a maternal stress, for example), environmental factors, and replication errors [49]. There are two main molecular epigenetic mechanisms involved in gene expression: covalent modification of DNA through methylation and chromatin structure regulation via histone post-translational modification [50].

1.3.3.1 DNA methylation

DNA modification (DNA methylation and de-methylation) act at the interface of genetic and environmental factors and may contribute to autism's complex aetiology [51]. Previous studies identified perinatal environmental exposures to various environmental factors and their link to impaired methylation in children with autism [52]'[53]'[54], [55]. DNA methylation is catalysed by *DNA methyltransferases* (DNMTs), a class of enzymes that binds the methyl-group specifically to the 5'-position of the pyrimidine ring of cytosine in a DNA sequence. Another important, but less characterised process is demethylation, which is mediated by the *ten eleven translocation* (TET) enzymes responsible for 5-methylcytosines (5mCs) oxidization [56]. DNMTs and TETs protein families continue to evolve in brain regions throughout brain development in mammals. DNA (de)methylation profiles have a crucial role in neuronal plasticity and plasticity-related behaviours such as reward learning, addiction and stress response. Also, they are involved in neurobiological processes such as learning and memory [50], [57].

Recently, DNA methylation profiling of five CpG islands (CGI-1 to CGI-5) in the *SHANK3* gene in autism post-mortem brain tissue showed a significant increase overall DNA methylation [58]. Gregory et al. found evidence for differential methylation pattern of the oxytocin receptor gene in autism when compared to control [59]. Disruption of the methylation status of autism candidate genes such as glutamic acid decarboxylase 65 (*GAD65*), ubiquitin protein ligase E3A (*UBE3A*) and reelin (*RELN*) have been reported in several studies [19], [22], [60]. For instance, a

study investigating methylation signatures of autism candidate genes in human fetal membrane from preterm birth, reported hypermethylation of *OXTR* (Oxytocin receptor gene) promoter [60].

Moreover, global methylation analysis of LCLs from individuals with autism reveals epigenetic differential regulation in *RORA* expression [61]. One of *RORA*'s transcriptional targets is *CYP19A1* (aromatase), the enzyme responsible for testosterone-estradiol conversion. Mutations in *CYP19A1* has been reported in the middle frontal gyrus of the brain in autism [62].

Previous studies have investigated autism associated methylation signatures in both peripheral tissues [63]–[65] ($50 < N < 2917$) and in the post-mortem brain [66]–[69] ($31 < N < 81$). While post-mortem brain is pertinent for a neurodevelopmental condition like autism, it is not readily accessible, and will be confounded by post-mortem effects on DNA methylation. Studies of methylation signatures in post-mortem brains in autism have replicable identified differential methylation [66]–[69]. Further, they have demonstrated an enrichment for differentially methylated signatures in the immune system, synaptic signalling, and neuronal regulation [66], [67], [69]. In contrast, recent large-scale analysis of three different peripheral tissue datasets have not identified significantly differentially methylated CpG sites in autistic individuals compared to typically developing individuals [63], [65].

The lack of significant results in peripheral tissues may be attributable to small effect sizes, and significant heterogeneity in both CpG methylation and autism.

While a few studies have investigated DNA methylation underlying autism, to our knowledge, there has been no study investigating DNA methylation under- lying autistic traits in the general population, which are subthreshold manifestations of the autism phenotype. One measure of autistic traits is the Social and Communication Disorders Checklist (SCDC) [70]. Scores on the SCDC are associated (Cohen's D = 2.8, p value < 0.001) [70] and genetically correlated with autism (r_g ~ 0.3) [38], [70], [71]. The SCDC has a modest SNP heritability (h^2_{SNP} = 0.24, s.e. = 0.07) [72] , and polygenic scores for autism are associated with SCDC scores in the general population (max R^2 = 0.13%) [71]. An advantage of using a continuous measure of autistic traits is that it captures the underlying variance better and minimizes heterogeneity attributable to different diagnostic criteria and practices used to diagnose autism.

DNA methylation is partly heritable ($0.05 < h^2_{twin} < 0.19$, defined as the proportion of variance in methylation that is attributable to genetics) [73]–[75]. A few studies have integrated genetics and methylation to identify convergent signatures in autism. Andrews and colleagues demonstrated that autism associated GWAS loci are enriched for methylation QTLs (mQTLs) in fetal brain and blood, suggesting that at least some of the gen- etic loci associated with autism may contribute to the condition through

differential methylation [76]. In line with this, Hannon and colleagues demonstrated that polygenic risk for autism is associated with differential methylation at birth [63]. While these studies have demonstrated a role for common genetic variants associated with autism and influencing methylation, to our knowledge no study has investigated if methylation of CpGs associated with autistic traits are enriched for common genetic variants associated with autism or autistic traits. One way to test this hypothesis is using mQTLs.

In Part II/Chapter 7 of this thesis I hypothesized that mQTLs of significant CpGs in a methylome-wide association study (MWAS) of autistic traits will be enriched for lower p values in a GWAS of autism or autistic traits.

1.4 Autism Prenatal Origin

1.4.1 Human early brain development

Human early brain development is a complex and dynamic process orchestrated by a sequence of genetic, environmental, biochemical, and physical events. It begins as pseudostratified epithelial sheet (neuroepithelium), and then it starts to bend and fold to eventually close and generate the neural tube along the length of the developing embryo (Price 2011). The neural tube closes during the fifth week of gestation (embryonic day30), initiating a rise in intraventricular fluid pressure indicating the start of rapid brain enlargement [77]. Around week 4 of gestation, a layer of meningeal cells become visible and

form a major signalling centre between the cortex and the skull to control the proliferation and migration of neural progenitors and neurons [78]. Aberrant development or disruption of the meningeal formation and function result in complex brain disorders such as lissencephaly and polymicrogyria [79]. Week four of gestation also marks the start of an oscillatory process (interkinetic nuclear migration) during which the neuroepethelial cells divide symmetrically at the margin of the ventricle. This process consists of four phases: 1) cell nuclei positioning at the basal, abventricular locations, 2) migrate toward the apical ventricular surface, 3) divide symmetrically into 2 progenitor cells and 4) return to their initial basal position. Early development of the human brain is thus characterised through this early proliferation resulting in an increased surface area and thickening of the ventricular zone [77].

Neurogenesis then begins at gestational week 5; the progenitor cells in the ventricular zone (aka the radial glial cells) switch for symmetric to asymmetric proliferation. The radial glia are the neural stem cells, and they differentiate into neurons, intermediate progenitors, and basal radial glia [80]. Asymmetric division indicates that while one of the two daughter cells remains as radial glial cell at the ventricular zone, the other one differentiates into intermediate transient amplifying population, leading to the generation of neurons and contributing significantly to cortical growth and folding. Intermediate populations include intermedial progenitors (IPs) and basal radial glial cells (bRGs) [81]. The radial glial cells have a long basal process connecting it to the outer surface, those cells that remain in the VZ divide at the apical surface while the IPs and the bRGs are located at the subventricular zone (SVZ).

These intermediate populations are prolonged through the inner and outer subventricular zone (ISVZ and OSVZ), which makes them crucial for brain expansion. Neurons migrate through the intermediate zone and through the cortical plate through the basal processes, the development of the cortical plate begins at the 7^{th} gestational week and continues to differentiate to form its six-layered structure until gestation week 18. This is marked by the radial expansion of the cortical plate and subplate and the attenuation of the proliferative zone [77] and the cortical continue to differentiate throughout the gestational period.

During human brain development three main phases can account for differential growth; 1) Neuronal division (spanning the first half of gestation) and as described above, is characterized by neuron proliferation and migration to the outer brain surface. This phase is associated with cortical thickness and expansion, but it is not until the second phase that cortical starts folding [80]. The second phase spans from mid-gestation and until 24 months postnatally, during which the neuron connections start forming and inducing expansion of the outer cortex and thus folding occurs due to the increased cortical stress. During this phase, myelination also takes place inducing white matter growth. The third phase is associated with mild synaptogenesis and neuronal connectivity (synapting pruning and elimination) and this spans the entire lifetime. During this time, the cortex plasticity allows it to adapt to cortical stress and undergoes secondary and tertiary folding[77].

The disruption of cellular and developmental pathways during early brain development, including neurite outgrowth, synapse formation and synaptic transmission have been described in various neurodevelopmental conditions including autism [40], [82], [83].

1.4.2 Autism: a neurodevelopmental condition

Autism is a polygenic complex condition that results from an interplay between the genetic background and epigenetic regulation by internal and external environmental factors, as concluded by the previous sections (Autism Genetics and Epigenetics). Moreover, post-mortem studies of idiopathic autism children as young as 2 years provided evidence that autism is also a multistage condition of 'prenatal' development as maldevelopment was shown at first, second and third prenatal trimester [84]–[88]. Autism related phenotypes reported in children include excess neurons in the cerebral cortex (30% increase in autism versus control) [89]. Cortical neurons proliferate exponentially in the first and second trimester and do not proliferate postnatally supporting the notion the prenatal pathological origin of autism [90], [91]. Moreover, brain overgrowth has been reported in utero and in autism toddlers which was associated with dysregulation of cell cycle gene expression [92], [93]. Other studies in the prefrontal cortex of autism individuals reported dysregulation of cell cycle, differentiation, and DNA damage detection genes and pathways [94]. Furthermore, studies demonstrated the cortical layer disorganization and dysregulation of neuron migration in the prefrontal and temporal cortices during the second and third trimester [86], [95]. Other studies reported reduced neuron and dendritic arbours'

size in autism post-mortem tissues which reflects aberrant neuronal growth and maturation [85], [96]–[98]. Autism-derived iPSC studies have further confirmed the developmental aspect of autism pathogenesis. A study comprising 8 autism-derived iPSCs and 5 healthy controls demonstrated disruption in a number of prenatal neural developmental processes including neuronal proliferation, maturation and synaptogenesis in telencephalic organoids [99]. Concordant with post-mortem findings, this model showed excess proliferation, moreover abnormal neurite and synapse development were reported. The telencephalic organoids used in this study recapitulate mid-fetal human cortical development, and were derived from autism subjects with macrocephaly and thus the excess proliferation could be directly related to the expanded head size occurring during early cortical development and not directly associated with autism [99]. Other studies using autism-derived iPSC (and no macrocephaly) reported dysregulation of genes enriched neuron differentiation, axon guidance, regional patterning and neuronal maturation but not neuronal proliferation [100], [101].

1.5 Autism Synaptic Origins

Many of the identified autism candidate genes are involved in neuronal processes ranging from neurogenesis to synapse formation, function and transmssion [102]. In 2014, a large whole genome exome study comprising more than fifteen thousand DNA samples revealed de novo loss of function mutation in more than on hundred genes in 5% of autistic subjects. More importantly many of these genes encoded for protein involved in synapse-related functions [40]. Additionally, integrative functional

genomic analyses during early developmental periods reported the autism associated genes involved in synaptogenesis [103].

The synapse is defined as the specialised site at which one neuron communicate with another through a process called synaptic transmission. There are two types of synaptic transmission, electric and chemical and this depends on the structure of the apposition of the neurons involved. For example, in an electric synapse the pre and post-synaptic terminals are in very close apposition that they form gap junctions. These junctions allow the generation of fast electric currents through the gap junction channels that physically connects their cytoplasm. While at the chemical synapses, there are no channels connected the pre- and post-synapse, instead the action potential at the presynaptic cleft release neurotransmitters which are then diffused across the synaptic cleft to the post-synaptic membrane through ionotropic or metabotropic receptors. This results in the excitation or the inhibition of an action potential firing [104].

Most brain functions such as learning and memory heavily rely on changes in synapse structure and function, and disruption of synapse formation and development results is the pathological origin of a myriad of neuropsychiatric conditions including autism, schizophrenia and epilepsy [105]. Trans-synaptic cell-adhesion molecules modulate synapse formation and/or structure by conferring synaptic partners' binding specificity. These synaptic molecules are also involved in synaptic function and maturation [106]. Many studies

support the hypothesis that ASCs are mainly caused by synaptic defects during early brain development based on functional and transcriptomic analysis studies as many genes associated with autism play fundamental roles in synaptic function and development [107]. For a detailed review of synaptic proteins, receptors associated with autism spectrum disorders Chen et al 2014 review [107], the following section will focus on neurexin and neuroligins only.

1.5.1 Neurexins

Neurexins are a family of adhesion proteins consisting of three neurexin genes (*NRXN*1, *NRXN*2 and *NRXN*3), located on 2p16.3, 11q13, and 14q31, respectively. Each gene has two major isoforms (α and β) expressed from two independent promoters [108].

Both isoforms have an identical C-terminal transmembrane region, and different N-terminal extracellular sequence. The α neurexins contain laminin/neurexin/sex-hormones binding globulin domain also known as LNS domain, and three EGF-like repeats while the β neurexins contain one LNS domain and a short specific N-terminal motif [106], [109]. Another notable family of proteins are the CASPRs; contactin-associated proteins. This family of proteins share a similar structure to neurexins and belong to a larger neurexin family of cell adhesion molecule, but they are known to be more involved in neuron-glia interaction. Like neurexins, CASPR contain EGF-like

repeats and homologous sequences to the G domain of laminin A, CASPRs extracellular region also includes discoidin and a fibrinogen-like domain [110].

As noted above, neurexins are cell adhesion molecules, which are enriched at the presynaptic membrane [108]. Neurexins bind to post-synaptic protein forming a trans-synaptic assembly mediate synaptic function and maturation. These binding partners include neuroligins [111], leucine-rich repeat transmembrane proteins [112], neurexophilins [113], dystroglycan [114], GABAA receptors [115], and GluD2-cerebellin-1 [106].

Neurexins are extensively and alternatively spliced at six canonical sites, generating thousands of isoforms with differential expression patterns through brain regions and developmental stages [105], [116]. Many factors have been implicated in the alternative splicing of neurexins. STAR is a family of RNA-binding proteins involved in alternative splicing events in neuronal and non-neuronal cells. *Sam68*, *SLM1* and *SLM2* are members of the STAR family, and it has been suggested that together they have a role in the differential alternative splicing of neurexins. *SLM2* is implicated in splice site 4 variant in *NRXN*2 while *SLM2* and *Sam68* are involved in *NRXN*1 and *NRXN*3 splice site 4 variants [117]. Other studies implicated H3K9me3 at the splice site 4 of *NRXN*1 [118].

One theory suggests that the extensive alternative splicing, let neurexins act as surface recognition molecules to identify synapses, since it is regulated and altered by the neuronal activity [119]. There is increasing evidence to support this theory. A study by Ullrich et al in 1995 reported the generation of more than 1000 isoforms resulting from independent alternative splicing and expressed in distinct subsets of neurons showing a high degree of regional regulation [120]. Ding et al investigated epigenetic regulation of *NRXN*1 splicing and memory preservation and interestingly they reported the regulation of *NRXN*1 SS4 by neuronal activity [121].

Another study demonstrated how alternative splicing of neurexins regulates NMDA- and AMPA-receptor mediated responses at hippocampal synapses. For instance, SS4 of *NRXN*1 enhances only postsynaptic NMDA-receptor mediated responses while SS4 of *NRXN*3 selectively regulates AMPA-receptor mediate responses [122]. Another study showed that depolarization of cortical cultured neurons regulates *NRXN*2 splicing, particularly at SS3, moreover they confirm exon 11 exclusion (at SS3) is calcium-dependent [123]. Furthermore, they investigate heterogeneous nuclear ribonucleoprotein (hnRNP) K and hnRNP L binding to this exon and they report that hnRNP K is implicated in *NRXN*2α expression and more importantly hnRNP L is involved in activity dependent splicing of neurexins, potentially governing trans-synaptic signalling [124].

Extensive alternative splicing of neurexin mRNA contributes to their molecular diversity, producing hundreds to thousands neurexin isoforms. Many attempts to unravel this molecular diversity using conventional sequencing and PCR-based methods [117], [125], [126], but failed to assess combinatorial alternative splicing events in single mRNA molecules due to the lack of a suitable workflow that can provide long reads in single cell while providing enough sequencing depth. Due to this predicament, the exact number of NRXN isoforms in the mammalian brain is not currently known but in recent years great advances has been made to answer this particular question.

When faced with a similar challenge to interrogate combinatorial altenrative splicing effects in DSCAMs (Down syndrome cell adhesion molecule; a highly diversified receptor in the Drosophila nervous system), scientists used a circularization-assisted multisegment sequence method where bar-coded circularized fragments of DSCAM could be subjected to Illumina sequencing [127]. This might have been a perfect solution for DSCAM but not for NRXNs due to the size of their mRNA transcripts and thus an alternative solution is required. In 2014, Treutlin et al., used a single-molecule long-read sequencing approach for profiling neurexin isoform diversity. In their study they investigated more that 25K full neurexin mRNA reads within single mRNA molecules and observed an enormous diversity of α-neurexin isoforms in which no single splice variant dominates, while the diversity is much more limited for the β-neurexins as most variants can be explained by two splice sites (SS#4 and SS#5). They

report a minimum of 1159 isoforms for NRXN1α and a total of 152 isoforms for β-neurexins in the prefrontal cortex of adult mice [116].

Previously, the estimated number of neurexin variants resulting from alternative splicing was between two and three thousand variants, but this study suggests that the true number of neurexin variants is probably higher, as the numbers reported in this study included one brain region and one developmental stage [108], [119], [128]. Flaherty et al integrated targeted long and short-read sequencing data, to catalogue NRXN1α isoforms in human post-mortem brain tissues and hiPSC-derived neurons. In their result they report 123 human NRXN1α isoform, predicted to be translated across two hiPSC neuronal subtypes, three adult and three fetal prefrontal cortex samples [129].

Several studies implicated CNVs and mutation of *NRXN* genes in autism [35], [37], [130]–[132]. For instance, a truncation mutation in *NRXN*2 was found in autism individuals, derived from a father with severe language delay and a history of schizophrenia in the family and another study reported *NRXN*3 deletions in four autism individuals [133], [134]. Other member of the *NRXN* superfamily have also been associated with autism, for example 13 rare variant of Caspr2 (also known as CNTNAP2) have been identified uniquely in autism individuals [135].

1.5.2 Neuroligins

Neuroligins are also type 1 membrane proteins located on the post-synaptic membrane with a simple domain structure, homologous to acetylcholinesterase. The neuroligins family comprises 5 *NLGN* genes; *NLGN*1 located at 3q26, *NLGN*2 at 17p13, *NLGN*3 and *NLGN*4 at the X chromosome (Xq13 and Xp22.3) and finally *NLGN*4Y localized at the Y chromosome (Yq11.2). While *NLGN*1, *NLGN*3 and *NLGN*4 are mainly expressed in excitatory glutamatergic synapses, *NLGN*2 is exclusively expressed inhibitory GABAergic synapses [106]. *NLGN*s function at the synapse are diverse, *NLGN*1 potentiates synaptic NMDA/AMPA receptor ratios at excitatory synapses, while *NLGN*2 induces inhibitory GABAergic synaptic responses [136]. Moreover, *NLGN*1 has been implicated in the maturation of synapse formation as well as neurite outgrowth [136]–[138].

Animal studies confirmed a role for *NLGN*3 in spatial learning and memory, mice that recapitulate an autism-related mutation (R451C) in *NLGN*3 show behavioural and synaptic deficits including an increased in amplitude and frequency of IPSC (inhibitory postsynaptic current) and mIPSC (miniature-IPSC pyramidal neurons) [139], [140]. The same phenotype was not pronounced in *NLGN*3 knock-out mice indicating that this mutation results in more than loss of function of *NLGN*3. This phenotype is consistent with studies using cellular models where the mutations results in an abolished synaptogenic activity [141], [142]. In another study, *NLGN*3

R451C mice reported an enhanced AMPA and NMDA-mediated synaptic responses in the hippocampal Ca1 region [143]. Recently, Nguyen et al showed that *NLGN*4X and *NLGN*4Y are differentially regulated despite their high homology. In this study they demonstrated that unlike *NLGN*4X, *NLGN*4Y has a trafficking deficit that impedes its ability to induce synapses due to a single amino acid difference between the two genes. Correcting this single residue allows *NLGN*4Y to induce synapses to levels similar to that of *NLGN*4X. Moreover they report that autism-related variants in *NLGN*4X show a similar phenotype to that demonstrated by *NLGN*4Y [144]. This is suggested to hold an explanation to the male bias representation in autism [144], [145].

The five genes in this family are alternatively spliced a single splice site (SS#A), except for neuroligin1, which has an addition canonical splice site (SS#B) [106]. The affinity of their binding depends on different *NRXN*/*NLGN* pairing and their alternative splicing [146]. For example, SS#B in *NLGN*1 plays a crucial role in *NRXN*s binding. *NLGN*1 with an insert in SS#B binds exclusively to β-*NRXN*s lacking an insert in SS#4, while *NLGN*1 lacking an insert in SS#B binds freely to both *NRXN* isoforms regardless the presence of absence of an insert at SS#4 [146].

GWAS studies implicated *NLGN*1 CNVs in autism in European ancestry cohorts: Autism Genetic Research Exchange (AGRE) and Autism Case-Control (ACC) [35].

Moreover, Leblond et al identified a maternally inherited duplication in *NLGN*1 found in autism and intellectual disability cases, and Tejada 2019 reported a novel nonsense homozygous mutation in the NLGN1 gene in monozygotic autism twins with intellectual disability [147], [148]. Another study sought to systematically investigate functional effects of five missense mutations in NLGN1 that were identified in autism individuals. They report changes in cellular localization, protein degradation and impaired spine formation implicating NLGN1's synaptic functions in the aetiology of autism [149].

*NLGN*2 has also been implicated with a number of neuropsychiatric conditions including autism and schizophrenia [150], [151]. A study in 2016 identified a nonsense mutation in *NLGN*2 resulting in a loss of function in a young male diagnosed with autism, macrocephaly, anxiety, obsessive-compulsive behaviours, developmental delay, and obesity [152]. Moreover, a study using exome sequencing to analyse rare coding variants in autism and ancestry-matched controls implicated variants of *NLGN*2, *NRXN*1 and SHANK3 in autism [40].

Many studies reported mutations in *NLGN*3 and *NLGN*4 in individuals with autism [153]–[155]. In one multiplex autism family, three siblings were reported having the same frameshift mutation in *NLGN*4 inherited from their unaffected mother [156]. The same study reported a point mutation in *NLGN*3 in two siblings, inherited from their unaffected mother [156]. Moreover, in one large family in France, 10 males had

X-linked intellectual disability, 2 had autism and one was diagnosed with PDD. They all shared the same mutation (2 bp deletion) in *NLGN*3 genes, it resulted in a premature stop codon [157]. A GWAS of rare CNVs comprising more than one thousand autism families reported an exonic deletion of NLGN3 on chromosome X in a male proband[36]. Other studies found no links between rare and common variants in *NLGN*s and autism in their cohort [153], [158]–[162].

1.5.3 Neurexin/Neuroligins complex

Neuroligins are neurexins post-synaptic ligands, together neurexins and neuroligins form a trans-synaptic complex that governs synapse organization, neurite outgrowth and synaptic transmission [163].

Based on the crystal structure of the *NRXN*1/*NLGN*1 binding, we learned that the *NRXN* LNS domain binds to the esterase homology domain of the *NLGN* in the synaptic cleft [164]. This complex is coated on both sides by PDZ-domain containing proteins, and a heterophilic, Ca2+-dependent complex at synapses. The trans-synaptic interaction between the pre-and post-synaptic counterparts coordinates and promotes specialisation of GABAergic and glutamatergic synaptogenesis [163]. The direct binding between neurexin and neuroligin is facilitated by other molecules too; the *NRXN*1α/*NLGN*1ß are bridged by a synaptogenic protein (Hevin) secreted by astrocytes at thalamocortical synapses in mice [165]

The *NRXN* cytoplasmic sequence has a C terminal binding site for class II PDZ domain that binds to CASK and other related proteins. Whilst *NLGN*s, also on their C terminal end binds to class I PDZ domain such as PSD-95, a post synaptic protein involved in the recruitment of glutamate receptors. CASK and PSD-95 are membrane associated guanylate-kinase proteins also known as MAGUK proteins. This family of proteins (MAGUK) act as scaffolding molecules for cell adhesion and signalling molecules, at the pre-synaptic (e.g., CASK) and at the post-synaptic (=e.g., PSD95) membrane. Their main role during synapse development is to establish connection with other scaffolding proteins (e.g., Shank) and the actin skeleton.

1.6 NRXN1

*NRXN*1 is one of the largest (spans 1.12 Mb) and most complex genes in the human genome. The *NRXN*1 gene encodes two major isoforms transcribed under independent promoters: at the 5' end lies the *NRXN*1α and the *NRXN*1β is downstream exon17 [163]. *NRXN*1 and its two major isoforms (α and β) are highly expressed in the brain during fetal cortex development, both isoforms show a marked and constant increase with gestational age [166]. Jenkins et al also reported that *NRXN*1 expression peaked at birth, and until 3 years, after that the mRNA levels dramatically decreased until they reach a lower level of expression that remains constant across the lifespan [166]. This is consistent with other studies confirming a marked high expression of *NRXN*1 in the

prefrontal cortex during a critical time frame (late embryonic to early postnatal periods) for synapse development, formation and function [167], [168].

Moreover, *NRXN*1 role in synaptogenesis during early brain development was previously suggested by Huttenlocher where they report a correlation between *NRXN*1 and cortical synaptic density in the human cortex [168], [169].

The *NRXN1* gene contains six canonical splice sites, generating thousands of isoforms [105]. This gene undergoes extensive alternative splicing resulting in more than a thousand different isoforms and only a few have been documented [116], [129]. More than 200 different *NRXN1α* isoforms were reported from Single- molecule long-read RNA sequencing (RNA-Seq) of mouse prefrontal cortex (PFC) [116]. Moreover, novel mutations generate novel isoforms which makes it challenging to predict the outcome of a given mutation, or getting an accurate quantification of *NRXN1* expression without having the complete repertoire and better methods of detection [129].

Using short and long read sequencing authors surveyed and compared *NRXN1α* isoform across fetal and adult post-mortem tissue as well as in iPSC-derived neurons. They went on to catalogue those isoforms found in different types of iPSC-derived using dual SMAD inhibition protocol and generating a heterogenous mixed neuronal culture, NGN2- forced expression generating excitatory neurons and using forced expression of ASCL1 and DLX2 for

the generation of inhibitory neurons. They were able to validate that the iPSC-derived neurons largely overlap with the *NRXN*1 isoform profile of that of human post-mortem brains [129].

To investigate the effect of patient mutations on the *NRXN*1 isoform repertoire; four iPSC lines were generated from four psychosis individuals carrying the two 5'*NRXN1* and two 3'*NRXN1* deletions (5'*NRXN1*$^{+/-}$ and 3'*NRXN1*$^{+/-}$). They used short and long-read sequencing to quantify the *NRXN1*α isoform repertoire from induced neurons derived from the four aforementioned patient lines (*NRXN1*$^{+/-}$) as well as four control lines. The *NRXN1*α isoform catalogued from the *NRXN1*$^{+/-}$ neurons were different when compared to the control group. Whole transcriptome RNA-sequencing of the neural progenitor cells and induced neurons showed no significant reduction in *NRXN1* total gene expression in *NRXN1*$^{+/-}$ patient lines when compared to control. However, they reported a shift in *NRXN1* isoform expression between the 5'*NRXN1*$^{+/-}$, 3'*NRXN1*$^{+/-}$ and the control group. This highlights how *NRXN1* expression levels can be masked by alternative isoform expression, as well as the diversity of the effect of *NRXN1* mutation on isoform expression [129].

Earlier studies suggest that the highly dynamic process of *NRXN*1 splicing serves as a molecular code to determine the type of synapse being formed in the brain [119]. For example Graf 2004 suggested the involvement of *NRXN*1α in GABAergic synapses while the *NRXN*1β isoform is predominantly found in glutamatergic synapses [170].

To validate the role of neurexins in synaptogenic activity non-neuronal cells were co-cultured with neuronal cells, and *NRXN*s expression in the non-neuronal cells resulted in the induction of postsynaptic assembly in the neuronal cells. Furthermore, the transfection of the non-neuronal cells with *NRXN*1α, motivated the clustering of GABAergic postsynaptic proteins (gephyrin and neuroligin2) in the connecting neuron but not of the glutamatergic postsynaptic proteins (PSD95 or neuroligins 1,3 and 4) [171]. Other studies reported the clustering of PSD95 in connecting neurons resulting from *NRXN*1β transfection in PC12 and/or COS cells, consistent with studies using primary rat hippocampal neurons [170], [172]. Together these studies confirm the complexity of alternative splicing and how it is regulated by the types of synapse to be established during development.

1.6.1 *NRXN*1 mutations and neuropsychiatric conditions

Functional aberrations of NRXN1 contribute to a myriad of neuropsychiatric conditions including autism, schizophrenia and intellectual disability [133], [163]. Despite the growing interest in investigating *NRXN*1 synaptic function, many questions remain unanswered regarding their impact on the pathophysiology of these conditions. Large-scale genetic studies confirmed the converging of common pathological mechanisms for autism and schizophrenia conferred in some cases by mutations in *NRXN*1 [163].

1.6.1.1 NRXN1 mutations associated with Autism

Human genetic studies reported that mutations in *NRXN*1 are strongly associated with autism. Most autism associated deletions in *NRXN*1 affect *NRXN*1α and rarely affect the β isoform. For instance, a 2006 Friedman was the first to discover de novo deletion affecting the first 5 exons of the *NRXN*1α isoform in an autistic child [173]. Moreover, the autism genome project consortium reported a large deletion in two female siblings with autism[174]. Inherited deletions were reported in a genomic screening of close to 900 individuals with autism and confirmed in Marshall et al. [35], [130].

Another study reviewed clinical and control cases for comparative genomics and revealed 12 individuals with exonic deletions in *NRXN*1, interestingly these individuals were diagnosed with autism, intellectual disability, language delay and hypotonia [175]. Zweier et al reported a compound heterozygous mutation in *NRXN*1; a large deletion spanning exons 1-4 and a stop codon mutation on the second allele, resulting in a loss of function of *NRXN*1α but not *NRXN*1β isoform. The mutations were associated with autism and intellectual disability [176]. A rare heterozygous missense mutation was reported in four autism individuals and not in 1201 controls [133]. A study in 2012 performed a thorough clinical and molecular screening of 24 individuals with mutations in *NRXN*1, they reveal that 17 of these deletions were exonic while only 7 had intronic deletions[177].

More importantly, most deletions occurred in the *NRXN*1α isoforms. Those mutations were associated with intellectual disability (93%) and autism (56%). The mutations affecting the C-terminal (*NRXN*1β isoform) had a higher frequency of seizures when compared to mutations affecting *NRXN*α [177]. Similarly, Dabell et al showed that *NRXN*1α carries a higher burden of deletions (exons 1-4) and more likely to be associated with autism and/or epilepsy [178].

Most of the mutations reported in autism subjects were in *NRXN*1α, however efforts by Camacho-Garcia et al discovered four novel point mutations in Exon 18 (specific to *NRXN*1)β and Exon 24 (affecting both isoforms) [179].

More deletions have been reported in patients with autism, for instance study in 2013 reported multiple deletions spanning the first 5 exons with one deletion encompassing nearly exons 1 to 19 affecting both isoforms (Figure 1-1). Similarly, Szatmani 2007 reported an individual with autism with a deletion from exon8 to exon 18 but as shown in the figure below and as reported in most studies there is a much higher burden on the alpha isoform of NRXN1 as most deletions seem to lie in the first few exons of the gene [180]–[186].

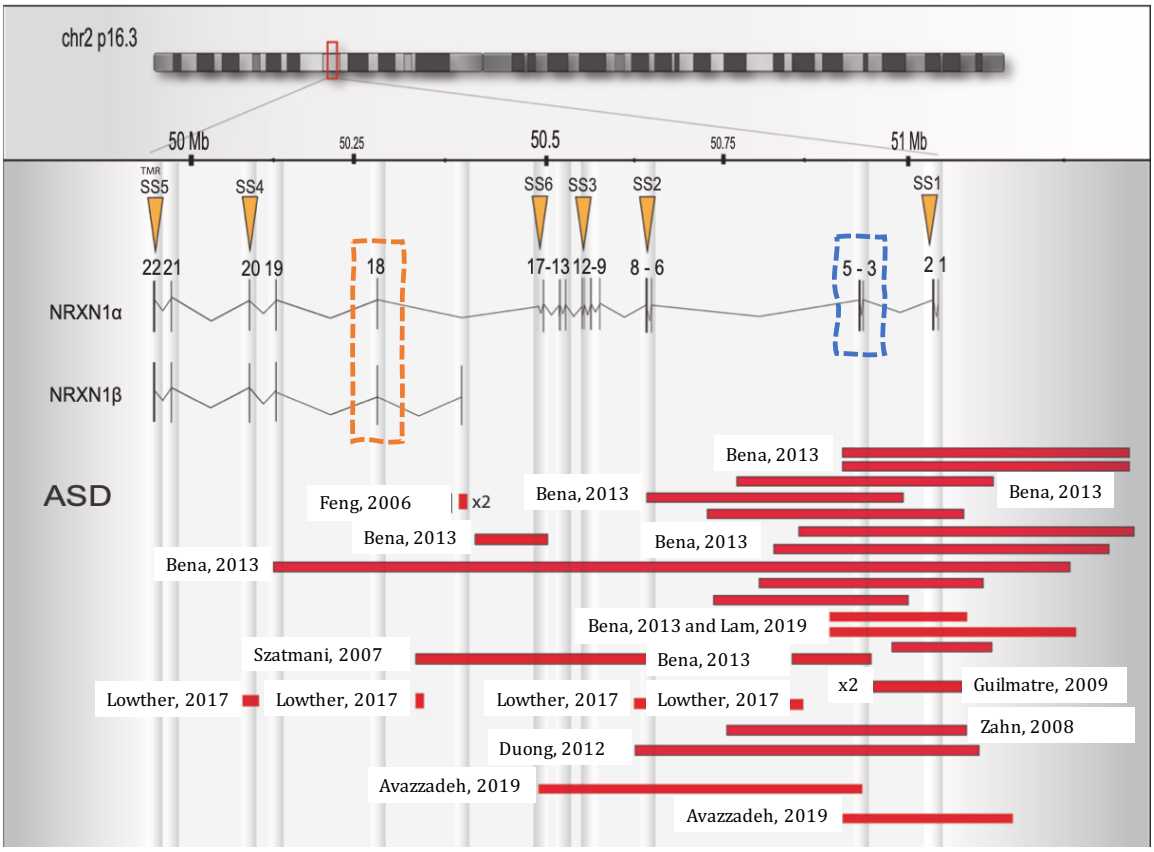


Figure 1-1: NRXN1 gene map showing the locations of the mutations associated with autism. The red bars represent the locations of the deletions reported spanning the exons 1 to 22 on NRXN1 isoforms, the yellow triangles show the location of the splice sites and the blue and orange dotted boxes highlight the locations of the mutations that will be studies in this thesis (Chapter 2, and Chapter 3). This figure is adapted and modified from Tromp et al 2020 [187].

1.6.2 *NRXN1* role in Neurite Outgrowth

Neuron's morphology is characterised by three distinct components: soma, axon and dendrite. The soma contains the nucleus and other organelles, the axon is a long-extended process responsible for transmitting information and the dendrite arbour is to connect and receive information from other neurons [170]. Neuritogenesis is the mechanism of generating new neurites from the soma, this is followed by axonal pathfinding and formation of synapse [102].

Disruption in this process result in aberrant network functions and is implicated in autism and other neurodevelopmental conditions [188]. Moreover, 88% of high-risk autism genes are involved in early neurodevelopmental functions and 80% of those genes are implicated in Neuritogenesis [102].

The synaptotropic hypothesis suggests that axonal growth and dendritic branching are controlled and stabilised by nascent synapses, guiding the growth of neurite in synaptic partners rich areas [189]. Previous work in Xenopus aimed revealed that β neurexins and *NLGN*1 contribute to dendritic arbour development and disruptions in these molecules result in destabilising filopodia, reducing synaptic density leading to a reduction in arbour complexity as the neurons mature [190]. Moreover, homozygous mutant *NRXN* and *NLGN*1 null mutation in drosophila showed significant defects in arbour length and morphology [189].

Collectively, these result support and expand on the synaptotropic hypothesis and incorporate the role of cell adhesion molecules in this process and even propose calling these molecules neuritic adhesion complex (NAC) to highlight their role in neurite outgrowth prior to synapse formation[189], [190]. The potential contribution of neurexins in neurite outgrowth has been further explored in rodents. Transfecting *NRXN*1α siRNA in rat hippocampal neurons resulted in a marked increase in neurite length [188]. The *NRXN*1 ligand (*NLGN*1) has also been implicated in the regulation of neurite outgrowth through its interaction with *NRXN*1β and

on activation of fibroblast growth factor receptor-1 in rat hippocampal neurons [137]. The role of *NLGN*1 in neurite outgrowth has been confirmed in a neuroblastoma cell line through protein-protein interaction involving the AChE-like domain and PDZ motif. Since *NLGN*1 is not localised at axonal growth cones, and cannot promote neurite outgrowth in the absence of AChE-like and PDZ domain, it is suggested that this novel role is mediated through its binding to other cell adhesion molecules such as neurexins [191].

1.6.3 *NRXN*1 role in Synaptic function

Cell adhesion molecules are deeply involved in synapse function, especially neurexins and neuroligins, impacting the ratio between excitatory and inhibitory synapses within the brain [102]. Several studies implicated α and β neurexins in synaptic function and synaptic transmission at excitatory and inhibitory terminals [192]. Animal studies showed that homozygous mutations in *NRXN*1 result in alterations in synaptic transmission and behavioral deficits such as anxiety and social interaction, while minimal effects where displayed in heterozygous mice [193].

Triple knockout of α-neurexins in mice resulted in a decrease in calcium channel functioning leading to an impaired neurotransmitter release [82]. Loss of function of the *NRXN*1α isoform in mice was associated with a marked decrease in excitatory (but not inhibitory) synaptic strength [82]. Knockout of ß-neurexins in mice resulted in the suppression of presynaptic release probability that is caused by the dysregulation of the post-synaptic endocannabinoid

synthesis [194]. In another mice model, deletion of all ß-neurexins resulted in impaired synaptic transmission in two different types of synapses through different mechanisms. For instance, in parvalbumin-positive interneurons of the prefrontal cortex, *NRXN* deletions resulted in loss of synapses with no changes in action potential triggered calcium influx while the same deletion caused a marked decrease in action potential signals and did not report any loss in synapses [194].

Recently, induced pluripotent stem cells allowed the investigation of those mutations in human neurons. Autism-derived iPSC with heterozygous *NRXN*1α mutations reported increased amplitude, frequency and duration of calcium transients, this was complemented with upregulation of genes implicated in voltage-gated calcium channels and ion transport pathways [181], [195].

Conversely, another study reported a significant decrease in calcium signalling activity in neurons derived from an autism iPSC carrying a bi-allelic mutation in *NRXN*1α [181]. Pak et al, generated two iPSC lines with mutations in exon 19 and 24 (shared by both *NRXN*1 isoforms) and report reduced frequency of spontaneous miniature excitatory postsynaptic currents (mEPSCs) compared to the control [180]. Flaherty et al (2019) investigated neuronal activity in 4 patient-derived iPSCs carrying 5' and 3' *NRXN*1 mutations and reported reduced number of spikes in induced neurons 3 weeks post induction. Furthermore, they show that the overexpression of 4 wild type

*NRXN*1α isoforms rescued the phenotype in the iN with 5'*NRXN*1 but not the ones with the 3'*NRXN*1 mutation. The overexpression of the 3'*NRXN*1 mutant isoforms in the control neurons resulted in a reduced neuronal activity, indicating that the phenotype seen in the 3'*NRXN*1 iNs is not directly caused by the low abundance of *NRXN*1α expression but rather a dominant double negative effect caused by the mutant isoforms [129].

1.7 Model systems

Understanding brain typical and atypical development required good model system, and historically animal models have provided us with a detailed mechanistic understanding of many human diseases including neurodevelopmental conditions including autism and schizophrenia [196]–[199].

Benefiting from the shared biological mechanisms that have been conserved throughout evolution, a handful of biological model systems. Ideally, the best model should by robust, fast-growing species, generating a large offspring, cost and time effective [200]. The most commonly used model systems include yeast (*Saccharomyces cerevisiae*), fruit flies (*Drosophila melanogaster*), worm (*Caenorhabditis elegans) and* mouse (*Mus musculus*). The most preferred model system is the mouse due to its high similarity with the human genome, ease of maintenance and its ability to recapitulate most of human physiological mechanisms

[200]. The introduction of novel genetic editing tools made mouse the most useful organisms for studying complex diseases, including heart disease, obesity, high blood pressure and stress and genetic disorders such as Huntington's and Fragile X syndrome [201].

Each model organism has its advantages and disadvantages, for example fairly recently (1970s), the worm (*Caenorhabditis elegans)* has been extensively used in research for their advantageous reproducible cell lineage that allows unparalleled analysis of cell fate decision. Moreover, more than 70% of C. elegans genes have human orthologues and more than 40% of human disease genes have C. elegans orthologues. Although the use of this model organism has led to the discovery of various key cell death pathways, other aspects of its physiology are very different from human physiology [202], [203].

Another model recently added to the list is the zebrafish, Danio rerio, the first vertebrate system in which saturation loss-of-function mutagenesis screening was performed. An important advantage the zebrafish has over most model organisms available is the transparent embryo which makes early development accessible and added to this is the potential to reproduce in huge numbers with a manageable expense as opposed to the expensive animal facilities required for the maintenance of mice models. Although, the zebrafish model has allowed scientists to unravel key

principles of early embryonic development, it shares only a limited degree of biological similarity with humans[204]–[207]. The diversity of those animal models has hugely served and continues to serve the scientific community in expanding our knowledge about typical and atypical human development, but it also revealed the crucial need for human cell-based models that can recapitulate biological pathways that are specific to humans and cannot be reproduced in animal models.

Investigating the pathophysiological mechanisms implicated in autism relied on different animal models induced with environmental triggers (e.g Valproic-acid induced rat autism model) that exhibits certain behavioural aspects similar to those exhibited in autism, or animal models that have been genetically modified to replicate specific mutations found in autism [196], [197], [208]–[210]. Post-mortem human brains have also served as valuable sources for examining pathological changes in patients. However, these tissues only represent the disease endpoint, whereas the alterations that lead to the development of the diseases often occur early in development. Post-mortem brain studies are thus limited in their ability to reveal dynamic cellular changes that are often important in the disease mechanisms.

While these models contributed significantly to the understanding of autism and other neuropsychiatric conditions, they were limited in revealing dynamic neuronal changes in the context of a complex genetic background. The development of patient-

derived induced pluripotent stem cells has given a myriad of opportunities from unravelling autism's complex aetiology to new therapeutic discoveries.

1.7.1 Modelling Autism using iPSC

Recently, a paradigm-shift in the research of psychiatric disorders (including ASC) has emerged, based on the burgeoning field of cellular reprogramming and induced pluripotency. The ground-breaking technology of induced pluripotency awarded Yamanaka and Gurdon the Nobel Prize in Physiology and Medicine in 2012. Yamanaka developed a method to reprogram adult fibroblast cells to a state of pluripotency similar to that of embryonic stem cells using four transcription factors (OCT3/4, SOX2, c-MYC and KLF4). The derived induced pluripotent stem cells (iPSCs) can then be differentiated into any cell type, giving scientists the opportunity to model neurological conditions by using cells derived from the patient [211].

With recent advances in the field, patient-derived iPSCs have been successfully differentiated into various neuronal subtypes in both monolayer (2D) and as organoids (3D) to mimic the cellular and functional circuitry exhibited in autism [212], [213]. Initially the starting material according to Takahashi and Yamanaka in 2006 was fibroblasts. Although this cell type is accessible, but it is invasive and unsuitable for children especially those diagnosed with ASC. A promising alternative cell type is keratinocytes derived from the root shaft of scalp hair [214]. A variety of

subtype-specific neural differentiation protocols have been developed recently for disease modelling.

1.7.2 Combined use of iPSC and CRISPR-Cas9

Genome editing technologies have proven to be very useful for experimenting with animal and cellular models. Programmable engineered nucleases enable the rapid generation of loss/gain of function mutation in a single gene thus allowing scientists to study the effect of variants on biological processes [215].

In recent years, the emergence of the CRISPR-Cas9 system has further improved ease of use and design, markedly higher specificity and efficiency when compared to other popular genome editing nucleases (ZFN: Zinc-finger, and TALEN: transcription activator–like effector nucleases)[215], [216]. The CRISPR systems were initially identified in *Escherichia coli* as an immune response to foreign genetic elements.

CRISPR, also known as clustered regularly interspaced short palindromic repeats; are repeats separated by short variable sequences. CRISPR systems are divided into two classes. The first class contains, type I and type III systems, found in Archaea and the second class contains type II, IV, V and VI CRISPR systems. Type II CRISPR-Cas9 system, derived from *Streptococcus pyogenes* is the most commonly used for genome editing [216].

The CRISPR sequences were present in 40% of bacteria and 90% Archaea, also they are usually adjacent to Cas (CRISPR-associated) genes. Most importantly, foreign genetic elements including viruses were identified within the CRISPR sequence. The proximity of the CRISPR sequences to the Cas genes and the inclusion of foreign genetic elements led to the big discovery of the CRISPR's defensive mechanism of action[215], [217]. The system relies on three elements: the mature CRISPR RNA (crRNA) and the transactivating crRNA (tracrRNA) and the Cas enzyme. The crRNA is made up of 20 nucleotide and it binds to the tracrRNA and serve as a guiding sequence (guide RNA). They next form a complex with the Cas9, which will be guided to bind and cleave upstream the target region in the genome called protospacer-adjacent motifs (PAM). To adopt this mechanism for genome editing, the guide RNA is designed to be complementary to the target region downstream to the PAM sequence specific to the Cas9 nuclease [215].

Like other endonucleases, Cas9 cleaves the targeted DNA creating a double strand break (DSB) which is followed by one of the DNA damage repair mechanisms: non-homologous end-joining (NHEJ) or homology directed repair (HDR) in mammalian cells. NHEJ is the main DNA-repair pathway in humans, it mediates the relegation of the DSB without requiring a homologous template DNA. This also means that NHEJ is not restricted to a specific phase of the cell cycle. In genome editing experiments, following the activity of Cas9 resulting in a DSB in the target sequence, the system relied on the host's own repair mechanism and repair by NHEJ results in frameshift

and premature stop codon mutations. HDR on the other hand is used to generate specific mutations at the target region and requires the presence of a repair template. It is more complicated and less frequent than NHEJ, but it has been widely used to induce SNP at a certain locus as well as knock-in experiment. It is important to note that HDR is only active in dividing cells, and efficiency depends on many variables including, cell type, genomic locus and the size of the repair template[216], [218].

The CRISPR-Cas9 system relies on the same mechanisms and strategies as other nucleases previously used in genome editing, however, it offers major advantages experimentally. For instance, unlike ZFN the system is easily customisable, and the same components can be used to target multiple regions just by changing the guide RNA sequence. Another major advantage is the specificity of the cleaving site, Cas9 cleaves at the bp 5' of the PAM sequence making a blunt cut between the 17^{th} and the 18^{th} nucleotide in the target sequence [215], [216], [218]. The ease of design, the specificity and the high efficiency offered by this system (CRISPR-Cas9) has made it an invaluable tool for modelling autism and other neuropsychiatric conditions in animals as well as cellular models. Patient-derived or CRISPR-edited iPSCs can be differentiated toward glutamatergic, GABAergic, dopaminergic, and motor neurons, as well as astrocytes and oligodendrocytes according to the condition one wants to model [219].

1.7.2.1 Forward Programming (OPTi-OX system)

Initially, the generation of neuronal lineage from embryonic or induced pluripotent cells relied on embryoid body formation which presented a major drawback due to its heterogeneity. Other common protocols previously used are neural-inducing stroma PA6 or MS5 cells and selective survival of neural progeny which were associated with protracted differentiation and a poor yield, respectively. SMAD proteins are crucial players in neural induction through interacting with TGFβ and BMP signalling pathways [220], [221].

Recent studies showed that dual inhibition of SMAD signalling using Noggin and SB431542,result in destabilizing the TGFb/activin- and Nanog-mediated pluripotency network, as well as suppressing differentiation towards trophoblast and mes-/endodermal fates and promoting neuralization of primitive ectoderm by BMP inhibition [220]. It was a great improvement on older methods that required the use of stromal feeders or embryoid bodies, and it has proven to efficiently convert more than 80% of pluripotent cells into neurons, it a slow process that has been shown to generate high variability from experiment to experiment [220], [222]. Dual-SMAD inhibition protocols produces a mixed neuronal population and minimum of three weeks to get to the neural precursor stage [222].

Neurogenesis is the production of a diverse array of neuronal cell types from neural stem cells. The molecular mechanisms governing this cellular complexity rely on the

combinatorial activity of transcription factors with basic helix-loop-helix (bHLH) motifs. These bHLH transcription factors are also known as “proneural factors” and they are necessary for the intiation of neurogenesis, and crucial for the specification of neuronal subtype identity [223]. Direct conversion of somatic cells into neurons using forced expression of proneural transcriptional factors have been proposed as an alternative for rapid generation of neurons. After careful investigation of a series of transcriptional factors, *NGN2* overexpression alone showed rapid conversion into neurons [222].

Integrative genomic analysis of NGN2 reprogramming of human iPSCs have shown the efficiency of the programming protocol; a dramatic transcriptional change is marjed by rapid downregulation of pluripotency genes and an onser of pro-neural transcriptional network. As previously shown in a 2017 study REST inhibition combined with neuronal reprogramming factors drive the conversion of fibroblasts into neurons [224]. REST is mainly known for its role in inhibiting neurogenesis by binding to NEUROD4, and in this protocol NGN2 competes with REST for the promoter region of NEUROD4 and therefore neurogenesis is induced. Unpublished scRNAseq data from our lab shows the dramatic downregulation of REST and the upregulation of NEUROD4 [225]. In this study they report an increase in NEUROD4 expression at Day1, and at Day2 they report the upregulation of CDKN1A (cell cycle inhibitor), and as early as Day4 post induction they show evidence of mass cell cycle exit. This is cue for the pro-differentiation program as shown by the expression of

NEUROD1, SOX4 and pan-neuronal genes such as TUBB3 and SYT1. This transition from a cycling state to a neuronal differentiation state is corresponding to the transition of NSCs to differentiating neurons during development or in conventional differentiation protocols [116], [225].

The neuronal differentiation induced by NGN2 generated cholinergic neurons and cholinergic-glutamatergic hybrids. Lee et al showed that NGN2 in combination with ISL1 and LHX3 lead to the generation of motor neurons [226]. NGN2 reprogramming have also been associated with the generation of cholinergic neurons in the presence of dorsomorphin and forskolin, but it is not yet clear the specific positional identity or subtype of the cholinergic neurons [227], [228].

Some studies rely on the use of lentiviral vectors for the delivery of the TET-ON components. Notwithstanding, their efficiency and scalability, lentiviral-based protocols bear the risk of random transgenes integration in the genome [229].

To circumvent this, the TET-ON elements can be incorporated in a safe locus within the genome that doesn't disrupt the expression of adjacent genes, (genomic safe harbour (GSH) loci) to generate inducible cell lines. This method is known as the OPTi-OX system (Optimised inducible overexpression) and is a platform that allows the forward programming of iPSCs into specific cell types [230].

The platform relies on the dual GSH targeting of the Tet-ON system components for homogenous and controllable expression of inducible transgenes. The Tet-ON system consists of two components: 1) a reverse tetracycline transactivator (*rtTA*) which is responsive to doxycycline (DOX) and driven by a CAG promoter and 2) a transgene (*NGN2*) driven by an *rtTA* inducible promoter: Tet-responsive element (TRE). The components are then targeted into two GSH loci to ensure that the overexpression of the transgene will not be affected by promoter interference and reduces the size of the cassette allowing the insertion of larger transgenes if needs be [229].

Unpublished single cell RNA analysis from our lab revealed that the NGN2-derived neurons generate 3 populations of neurons. The major population is glutamatergic neurons, a small subpopulation shows higher expression of cholinergic markers and the third is a small population of neurons expressing both glutamatergic and cholinergic markers [225].

A 2020 study compared NGN2-derived neurons to embryoid body-derived neurons (a lengthy and more native protocol), their findings highlighted the major limitations of NGN2-derived neurons [231]. For instance, EB-derived neurons are able to reach a more mature stage and develop oscillatory activity resembling that of the rodent's developing brain. EB-based protocols are lengthy, cumbersome and highly heterogenous, while NGN2-derived neurons provide a simple and fast method to

obtain a more homogenous culture. Moreover, co-culturing NGN2-derived neurons with mixed glia has been commonly adopted as a way for promoting its maturation and electrophysiological function [225]. Following the NGN2 protocol, neurons (co-cultured with astrocytes) start expressing synaptic markers and start firing at 2-3 weeks, resembling approximately activity of the second trimester, and reached the maturity levels of the third trimester around the sixth or seventh week of differentiation [231].

A recent study demonstrated the successful generation of human excitatory neurons from stem cells by combining two strong neuralizing factors (transcription factor programming and small molecule patterning). They further undertake single-cell and reporter gene approaches to select highly differentiated neurons with increased functionality, augmenting their utility in the modelling of nervous system disorders. In this study they found that the induced neurons generated was most strongly correlated with the prenatal brain, behaving mostly like layer 3 pyramidal neurons. Moreover, they show that coculturing the induced neurons resulted in increased firing and synchronicity showing a more mature phenotype which made it an attractive model for studying aspects of neurodevelopmental conditions. Unlike rat and mouse models it is still unknown how the timeline of these neurons match human foetal brain development[232].

Bulk and scRNA-Seq analysis of NGN2-derived neurons compared with NGN2-derived neurons co-cultured with glia demonstrated the crucial role glia plays in the maturation of the induced neurons. For instance, SYN1, SYP and SNAP25 genes (pre-synaptic markers) seem to be similarly expressed in both conditions, while activity-dependent genes in the postsynaptic compartment (e.g., HOMER1, CAMKIIN, ARC and NPTX2) seem to show an upregulation in the cocultured neurons, with most of them showing an increasing trend two weeks post induction [225].

1.7.3 Challenges to iPSC-based autism modelling

While patient and genetically modified iPSCs provide a valuable experimental system to uncover pathological mechanisms in neuropsychiatric conditions, many challenges remain to be overcome. Firstly, iPSC have decreased potential of developing PAX6+ neural progenitors as compared to ESCs [233] which is an important step in differentiating into many neuronal subtypes. This might be due the epigenetic differences from the donor somatic cell, that has undergone an incomplete reprogramming [234].

Many tests have been developed to validate the reprogrammed iPSC before their use in disease modelling which can reduce the effects of reprogramming on differentiation variability, however, heterogeneity of the neuronal population generated remains a hurdle. Moreover, modelling neurodevelopmental and

neuropsychiatric conditions to recapitulate a functional phenotype requires a more mature stage than the one that can be reached in-vitro [235]. Major efforts and advances have been made towards the development of novel differentiation protocols to generate pure, mature and specific neuronal populations [99], [229], [230], [236].

Forward programming techniques have been developed to derive distinct neuronal subtypes, however the sensitivity of the approach to subtle alterations in TF combinations and co-administered growth and patterning factors remains a challenging factor.

In an attempt to look for the *NRXN*s expression in the NGN2 derived neurons, a temporal sc-RNA analysis was performed from iPSC stage until Day21 neurons with and without astrocytes (Figure 1-2) and demonstrated the high expression of *NRXN*s 2 and 3 weeks post induction with a slight increase in the presence of glia [225]. Although the EB-protocol is able to replicate a wider developmental range that cannot be covered by the NGN2-protocol such as cytogenesis, and cell migration, the NGN2 protocol provide an invaluable tool to investigate neurotic outgrowth, cell growth, electrical activity and synaptogenesis [231].

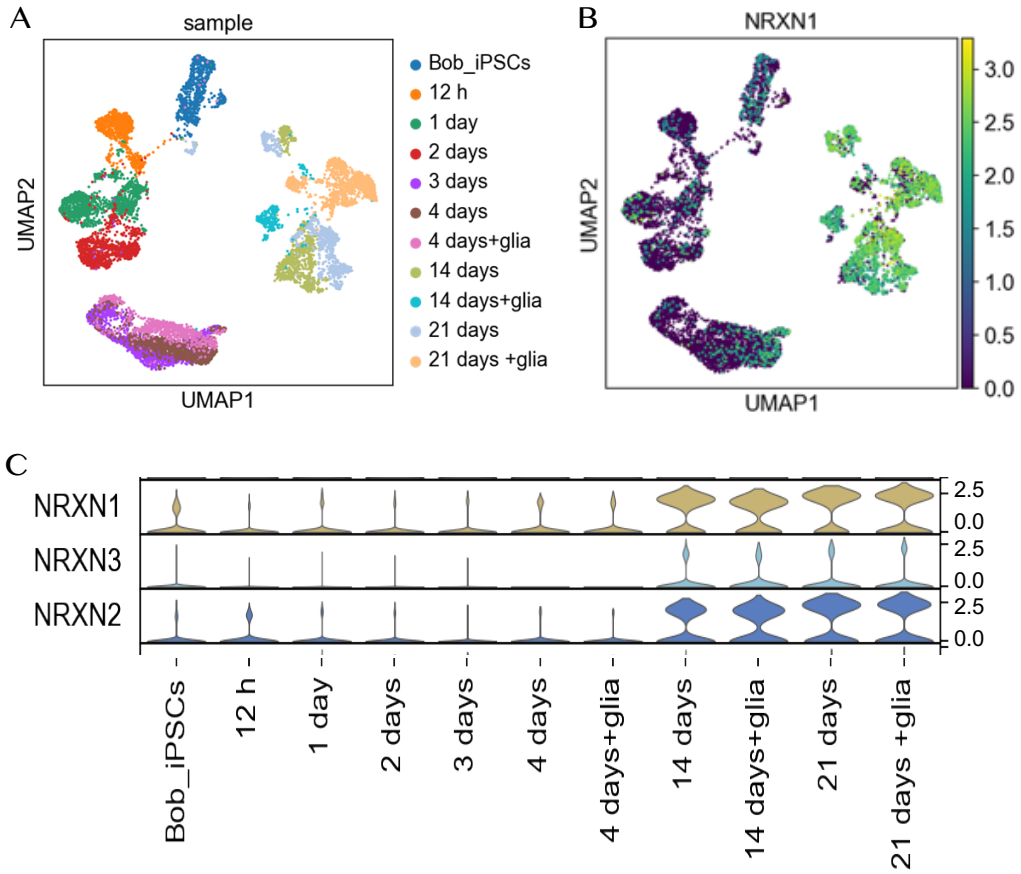


Figure 1-2: Single cell RNA Seq temporal analysis of NRXNs, A) A representative graph showing the position of the samples at each time point, B) NRXN1 expression across each timepoint - the colour scale indicated the level of expression, the lighter the green the higher the expression. C) Volcano plots of NRXNs (1-2-3) expression in neurons (with and without glia) across time

1.8 Hypotheses

This thesis is divided into two parts, the first and main part is an iPSC study focusing on the effect of NRXN1 mutations in iPSC-derived neurons, and the second part is an association study between DNA methylation at birth (cord blood), and autistic traits by conducting a methylome-wide association study (MWAS).

Part I: **Cellular Phenotypes Associated with NRXN1 Mutations in Autism; an iPSC Study (Chapter 3,4,5,6)**

NRXN1 has been implicated in autism but many questions remain unanswered regarding its impact on the condition's pathophysiology. Emerging evidence indicates that NRXN1 and its isoforms are highly expressed during early neuronal development, a period now recognised as critical period for the emergence of phenotypes associated with autism. However, our understanding of NRXN1 function during this period come from animal studies, making it more difficult to understand their implications in human neurons.

Development of stem cell technologies and gene editing techniques now allow us to generate autism-derived iPSCs carrying mutations in a particular gene, as well as for the generation of isogenic mutant lines to probe the role of gene in question. Coupled with advances in forward programming approaches (OPTi-OX), offer a reliable cellular model to investigate the contribution of specific genes in synaptic function in human neurons within a reasonable time frame. In this thesis, I relied on two

cellular models: 1) autism-derived iPSCs (patient) with NRXN1 mutations and 2) isogenic (mutant) iPSCs with induced NRXN1 mutations. By incorporating the OPTi-OX system, induced neurons were generated from patient and mutant iPSCs to reveal novel functions for *NRXN1* in development of synaptic function and contribution to autism related cellular phenotypes (Figure 1-3).

To this end, this thesis aims to address the following objectives:

1- Generating a stem cell model by introducing an inducible NGN2 factor to rapidly drive the differentiation of patient-derived iPSC into neurons (Chapter 3)
2- Generating 4 mutant lines by introducing indel mutation using CRISPR-Cas9 technology in healthy control iPSC lines (Chapter 3)
3- Study the effect of patient and CRISPR-induced mutations on the transcriptome of induced neurons (Chapter 4)
4- Utilise the system to study the effect of *NRXN1* mutations on temporal expression of Cell adhesion molecules and effect on neuritogenesis (Chapter 5)
5- Study the effect of *NRXN1* mutations on functional properties of the induced neurons (Chapter 6)

Upon completion of this project, we will have a clearer understanding of the contribution of *NRXN1* in the condition's aetiology and the underlying molecular and neurobiological pathways by dissecting the physiological effect of NRXN1 variants and their link to specific endophenotypes.

Part II: **Integrated genetic and methylomic analyses identify shared biology between autism and autistic traits (Chapter 7)**

We hypothesized that mQTLs of significant CpGs in a methylome-wide association study (MWAS) of autistic traits will be enriched for lower p values in a GWAS of autism or autistic traits.

To address these questions, we investigated the association of CpG methylation in cord blood using scores on the SCDC at age 8. The use of cord blood CpGs minimizes (though, does not eliminate) reverse causation (where the phenotype influences DNA methylation), as the methylation of CpG sites is measured very early in life. To investigate how comparable an MWAS of an autistic trait is to other MWAS of autism and related phenotypes conducted across different tissues, we investigated the overlap between the MWAS of SCDC and other MWAS of autism and communication-related traits in peripheral and post-mortem brain tissues.

We further investigated if genes that are transcriptionally dysregulated in the post- mortem autism brain are enriched for methylation CpGs associated with SCDC. Finally, integrating GWAS data for autism from 46,350 individuals, we investigated if mQTLs of CpGs associated with SCDC scores at various p value thresholds are significantly shifted toward lower p values in the autism GWAS. We validated these results using a smaller GWAS for SCDC.

In summary, this study had three specific aims: (1) to investigate if an MWAS for autistic traits identifies significant CpG methylation and (2) if it is comparable to MWAS of autism; and (2 and 3) to investigate if mQTLs of CpGs associated with autistic traits at various p value thresholds are enriched in GWAS of autism and autistic traits (Figure 1-3).

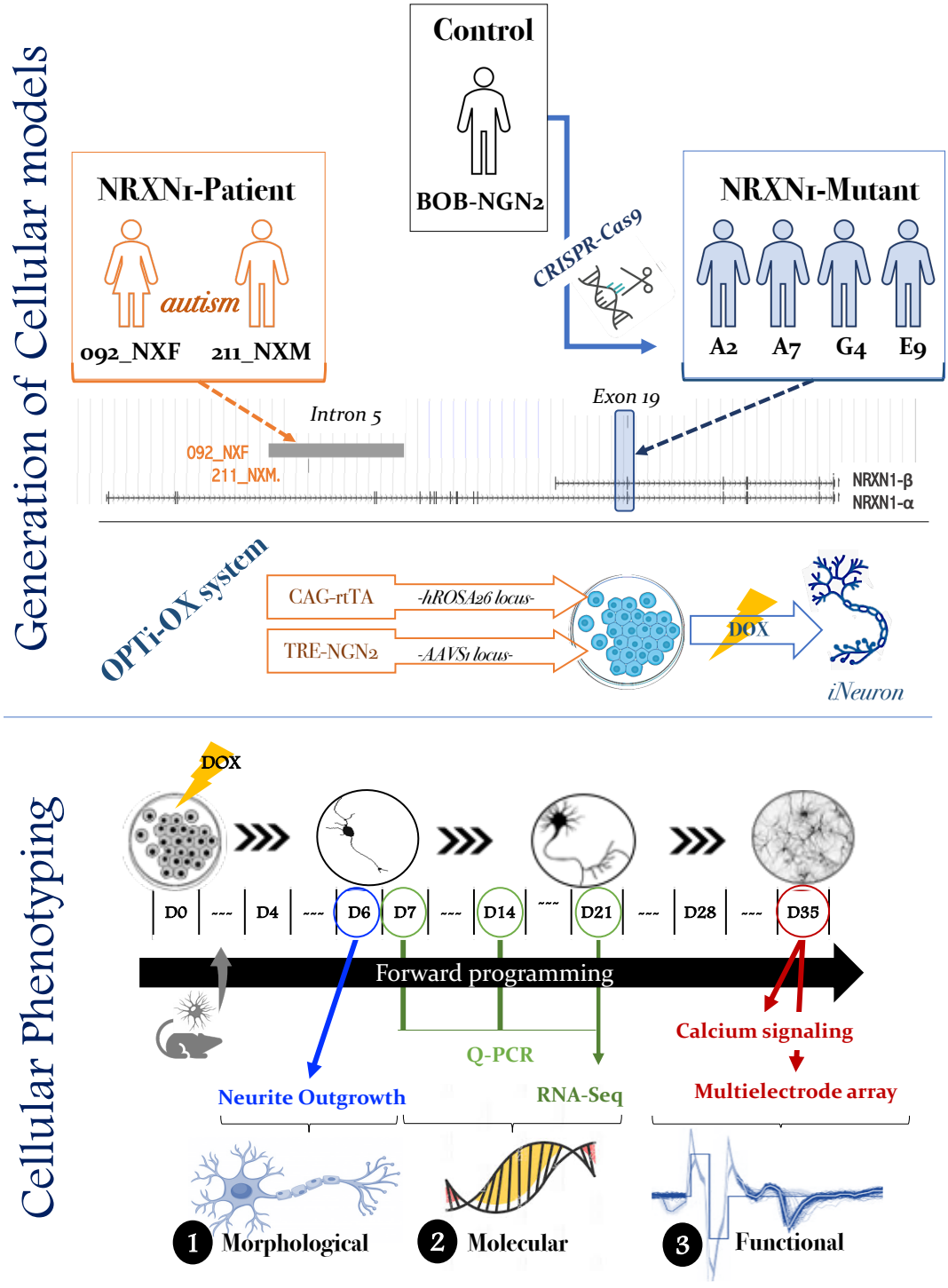
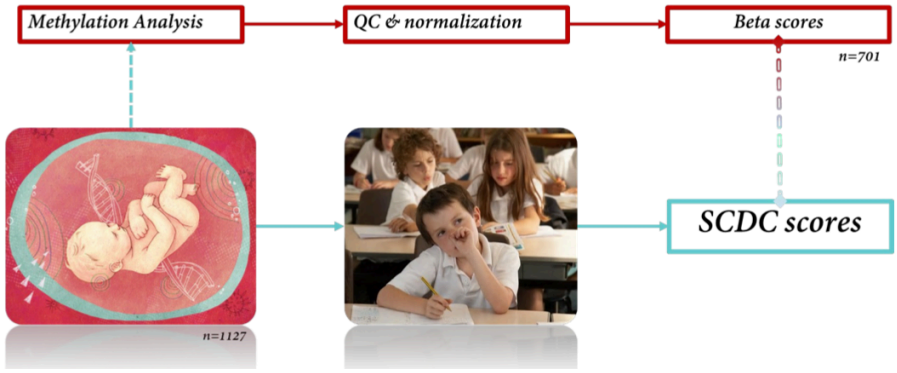
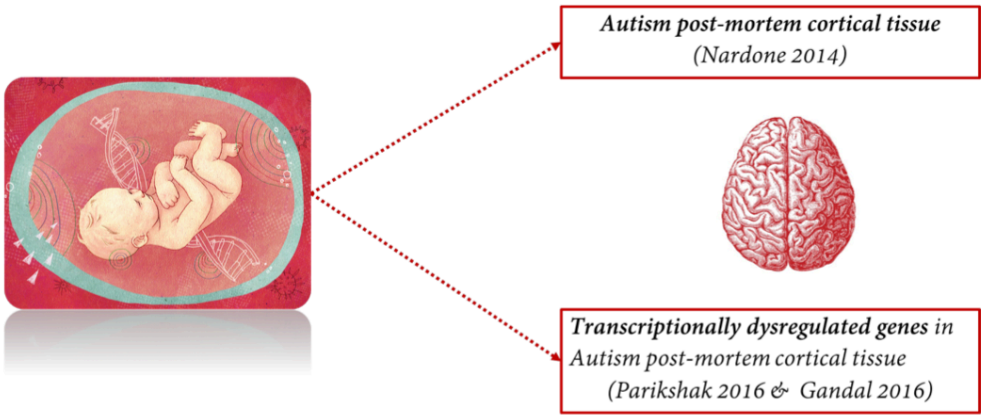


Figure 1-3: Schematic diagram of the study design, the first part shows the generation of two cellular models (NRXN1-patient) and (NRXN1-mutant) lines showing the loci of their mutations or target regions. The OPTi-OX components have been introduced in the generated cell lines to allow forward programing of iPSCs into induced neurons. The second part how I investigate the morphological, molecular and functional profiling associated with NRXN1 mutations during neural induction

1. Identify significant differentially methylated CpGs that are associated with scores on the SCDC scores



2. Investigate the extent of shared methylation across autistic traits (SCDC) and autism



3. Investigate the enrichment of mQTLs associated with SCDC CpGs for Autism GWAS loci

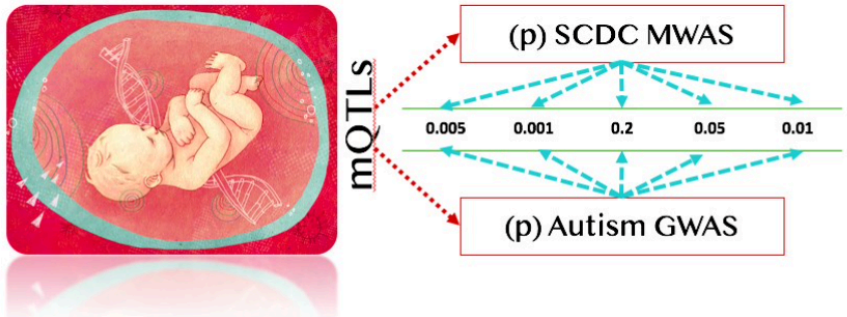


Figure 1-4: Schematic diagram of the aims of the Part II study

Chapter 2

Materials & Methods

2 Material and Methods

Part I: **Cellular Phenotypes Associated with NRXN1 Mutations in Autism; an iPSC Study**

2.1 Cell lines used in study

In this thesis, a total of three induced pluripotent stem cell lines were used (Table 2-1). Two patient iPSC lines were provided by Prof Jack Price at King's College London. The iPSC lines were derived from autism patients with NRXN1 mutations. The cell line 092_NXF was derived from a female diagnosed with developmental delay (high functioning autism). This patient carrying a *De novo* deletion of ~200kb in short arm of chromosome 2 – 2p16.3 (50,806,991 – 51,013,685), and a paternally inherited duplication in long arms of chromosome 1 – likely to be benign – 1q21.1 (144,679,874 – 145,747,269) x3, as identified by a clinical CGH array screen as part of the Brain and Body Genetic Recourse Exchange (BBGRE) program, King's College Hospital [237], [238].

The second cell line, 211_NXM was derived from a 50-year old male recruited from the Social Communication Disorders Clinic at Great Ormond Street Institute of Child Health (GOS-ICH). This individual was diagnosed with high-functioning autism, he was identified to carry a single nucleotide variant resulting in a missense mutation (G→C) on chromosome 2 (50724586). Additionally, some CNV VOUS (copy number

variant - variant of unknown significance) were identified for this patient: Xq28 dup-X: 152,900,100- 152,984,573 (84,472bp, SLC6A8, DUSP9, BCAP31, PNCK).

The third line used is a gene-corrected line from an individual with α1-antitrypsin deficiency, a well-established line used in our facility as a healthy control and is referred to as BOB-NGN2 [239]. The BOB-NGN2 has the pR26_CAG-rtTA and pAAVS1_TRE-NGN2 constructs previously inserted and rely on the OPTi-OX system for forward programming [229]. This line was provided by Dr Mark Kotter at the Laboratory of Regenerative Medicine. Summary of all the lines used in this thesis is listed in Table 2-1. Although it is not ideal to use a single control cell line to compare against the patient and mutant cell lines, due to the lack of available already established and characterised control lines in which the OPTi-OX system was already incorporated, I was forced to rely on only one control cell line.

Table 2-1: Summary of iPSC lines used, and gene edited in this study

		NGN2	***NRXN1*mut**	**Action**	**Product**
Control	Bob-NGN2	+	-	→ indel mutation	- A2 - A7 - G4 - E9

Patient	092_NXF	-	+	→ OPTi-OX	- 092_NXF-NGN2(6) - 092_NXF-NGN2(9) - 092_NXF-NGN2(12)
	211_NXM	-	+	→ OPTi-OX	- 211_NXM-NGN2(8) - 211_NXM-NGN2(13) - 211_NXM-NGN2(16)

2.1.1 Gene Editing

Gene editing tools were used to generate: **1) *NRXN1*-Isogenic (mutant) lines**: CRISPR-Cas9 was used to introduce indel mutations in the *NRXN1* gene in the control line; and **2) OPTi-NGN2 - *NRXN1*-Patient lines:** by inserting the two components of the OPTi-OX system into genomic safe harbour (GSH) loci in two patient lines with *NRXN1* mutations (092_NXF and 211_NXM).

2.1.1.1 Generation of NRXN1-Isogenic (mutated) lines

NRXN1 gene encodes 2 major isoforms, a long isoform: *NRXN1α*, and a short isoform *NRXN1ß*. Each isoform is controlled independently by two different promoters on separate positions on the gene: located upstream of exon 1 and downstream of exon 17, respectively [195]. BOB-NGN2 is a control iPSC line with an optimised

overexpressed inducible NGN2 (OPTi-OX NGN2 system), that has been well established and characterised in our lab [230]. CRISPR-editing tool was used to introduce indel mutation in this line via non-homologous end joining, to generate *NRXN1*-mutant lines. The gRNA was designed to target a shared exon between both the α and β isoform of the *NRXN1* gene as shown in Figure 2.1a. The gRNA synthesis and CRISPR-Cas9 nucleofection was done by AELIAN Biotechnology in Vienna, Austria using the method described by [218]. The gRNA sequence used is AGAAGCCGTATTGGTGCGAG; the target is highlighted in the *NRXN1* gene, Figure 2-1b.

A

B

NRXN1_human

>NM_004801.5:1478-5911 Homo sapiens neurexin 1 (NRXN1), transcript variant alpha1, mRNA

ATGGGGACGGCGCTGCTCCAGCGCGGGGGCTGTTTTCTTCTGTGCCTCTCGCTGCTGCTCCTGGGCTGCTGGGCGGAGCTGGG
CAGCGGGCTGGAGTTTCCGGGCGCCGAGGGCCAATGGACGCGCTTCCCCAAGTGGAATGCCTGCTGCGAGAGCGAGATGAGCT
TCCAGCTCAAGACTCGCAGCGCCCGCGGCCTCGTGCTCTACTTCGACGACGAGGGCTTCTGCGACTTCCTGGAGCTGATTCTG
ACGCGCGGCGGCCGCCTGCAGCTCAGCTTCTCCATCTTCTGCGCTGAGCCTGCGACGCTCCTGGCCGACACGCCGGTTAACGA
CGGCGCCTGGCACAGCGTGCGCATCCGCCGCCAGTTCCGCAACACCACGCTCTTCATCGACCAGGTGGAGGCCAAGTGGGTGG
AGGTCAAGTCCAAGCGCAGGGACATGACGGTGTTCAGCGGCCTTTTCGTCGGGGGGCTGCCCCCGGAACTGCGCGCCGCGGCG
CTCAAGCTCACCCTGGCCTCGGTGAGGGAGCGGGAGCCCTTCAAGGGGTGGATTCGTGACGTGAGGGTCAACTCCTCGCAGGT
CCTGCCCGTGGACAGCGGCGAGGTGAAGCTGGACGATGAGCCGCCCAACAGCGGCGGGGGAAGCCCGTGCGAGGCGGGCGAGG
AGGGCGAGGGCGGGGTGTGCCTCAACGGAGGTGTGTGCTCCGTGGTGGACGACCAGGCCGTGTGCGACTGCTCGCGAACCGGC
TTCCGCGGCAAGGACTGCAGCCAAGAAGACAACAATGTGGAAGGTCTGGCGCACCTGATGATGGGCGACCAAGGTAAAAGTAA
AGGAAAAGAAGAATATATTGCCACGTTCAAAGGATCTGAATACTTCTGCTACGACTTGTCTCAAAACCCCATTCAAAGCAGCA
GTGATGAAATAACTCTGTCATTTAAAACCCTTCAGAGGAATGGACTGATGCTTCACACTGGGAAATCGGCTGATTATGTCAAT
CTTGCCCTGAAAAATGGAGCTGTCTCTCTGGTCATTAATTTGGGATCAGGGGCCTTTGAAGCACTAGTGGAGCCTGTGAATGG
AAAGTTTAATGATAATGCCTGGCATGATGTGAAAGTCACCAGGAATCTGCGTCAGCACTCAGGCATTGGACACGCTATGGTGA
CAATATCAGTGGATGGGATTCTTACCACAACGGGCTACACGCAAGAAGATTATACCATGCTGGGGTCTGATGACTTTTTCTAT
GTTGGAGGCAGTCCCAGCACAGCCGACCTTCCAGGGTCACCAGTCAGTAACAACTTTATGGGCTGTCTCAAAGAGGTTGTATA
TAAAAATAATGATGTGAGGCTGGAATTATCTCGACTTGCCAAGCAAGGAGATCCTAAGATGAAGATCCATGGAGTGGTGGCAT
TTAAATGTGAGAATGTTGCAACTTTAGACCCAATCACCTTTGAAACCCCAGAGTCTTTCATCTCTTTGCCTAAATGGAATGCA
AAGAAAACTGGCTCCATATCATTTGATTTCCGTACAACAGAGCCAAATGGCCTCATCTTATTTAGCCATGGCAAGCCAAGACA
TCAGAAAGATGCCAAGCACCCACAGATGATAAAGGTGGACTTCTTTGCTATTGAGATGCTAGATGGCCACCTCTACCTCCTCC
TGGACATGGGGTCAGGTACTATAAAAATAAAAGCCCTGTTGAAGAAAGTGAATGATGGAGAATGGTATCATGTGGACTTCCAG
AGAGACGGACGGTCAGGTACCATTTCTGTCAACACGTTGCGTACTCCCTACACTGCTCCTGGTGAGAGTGAGATTCTGGACCT
GGATGATGAGTTGTACCTGGGGGGGCTGCCAGAAAATAAAGCTGGCCTTGTCTTCCCCACCGAGGTGTGGACTGCTCTGCTCA
ACTATGGCTACGTGGGCTGCATCAGGGATTTGTTCATCGATGGCCAAAGCAAAGATATCCGGCAAATGGCTGAAGTTCAAAGT
ACTGCTGGAGTGAAGCCTTCCTGCTCAAAGGAAACAGCAAAACCGTGCCTTAGCAACCCTTGCAAAAACAATGGCATGTGCAG
GGATGGGTGGAACAGATATGTCTGTGATTGTTCCGGAACAGGCTATCTTGGCAGGTCCTGTGAGAGAGAGGCAACGGTTTTGA
GCTATGATGGGAGCATGTTTATGAAAATTCAGCTCCCCGTAGTCATGCATACGGAGGCTGAGGATGTTTCCTTACGGTTCCGA
TCCCAGCGTGCATATGGCATTCTGATGGCAACCACTTCTAGAGACTCTGCTGACACCCTCCGCCTGGAGCTAGACGCAGGACG
TGTGAAACTGACGGTCAATCTAGATTGTATCAGGATTAACTGTAATTCCAGCAAAGGTCCCGAGACTCTTTTTGCTGGCTATA
ACCTCAATGATAACGAGTGGCACACAGTGCGTGTAGTTCGGCGTGGAAAAAGTTTAAAGTTAACAGTGGATGACCAACAGGCC
ATGACAGGTCAAATGGCAGGTGATCATACTAGGCTGGAGTTCCATAACATAGAGACTGGCATCATCACAGAACGACGGTATCT
TTCTTCTGTCCCCTCCAACTTCATTGGACACCTGCAGAGCTTGACATTTAATGGAATGGCATACATTGACCTGTGTAAAAATG
GCGACATAGATTACTGTGAGCTTAATGCCAGATTTGGCTTCAGGAACATCATAGCAGATCCTGTCACCTTCAAGACCAAATCG
AGCTATGTTGCCTTAGCTACCTTGCAAGCCTACACTTCTATGCATCTTTTTTTCCAGTTCAAGACAACATCCCTAGATGGATT
AATTCTATATAACAGTGGGGATGGAAATGACTTTATTGTGGTTGAATTAGTTAAAGGGTACTTACATTACGTGTTTGATTTGG
GAAATGGTGCTAACCTCATCAAAGGAAGCTCAAATAAACCTCTCAATGACAATCAGTGGCACAACGTGATGATATCAAGGGAC
ACCAGCAACCTCCACACTGTAAAGATTGACACAAAAATCACAACGCAAATCACCGCCGGAGCCAGGAACTTAGACCTCAAGAG
TGACTTATATATAGGAGGAGTAGCTAAAGAAACATACAAATCCTTACCAAAACTTGTACATGCCAAAGAAGGCTTTCAAGGCT
GCCTGGCATCAGTTGATTTAAATGGACGGCTTCCGGACCTCATCTCCGATGCTCTTTTCTGCAACGGACAGATCGAGAGAGGA
TGTGAAGGGCCCAGCACAACCTGCCAAGAGGACTCATGTTCCAATCAAGGTGTGTGCTTGCAACAATGGGATGGCTTCAGCTG
TGACTGTAGTATGACTTCCTTCAGTGGACCACTCTGCAATGACCCTGGGACGACATATATCTTTAGCAAAGGTGGTGGACAAA
TCACGTATAAGTGGCCTCCTAATGACCGACCCAGTACACGAGCAGACAGACTGGCCATAGGTTTTAGCACTGTTCAGAAAGAA
GCCGTATTGGTGCGAGTGGACAGTTCTTCAGGCTTGGGTGACTACCTAGAACTGCATATACACCAGGGAAAAATTGGAGTTAA
GTTTAATGTTGGGACAGATGACATCGCCATTGAAGAATCCAATGCAATCATTAATGATGGGAAATACCATGTAGTTCGTTTCA
CGAGGAGTGGTGGCAATGCCACGTTGCAGGTGGACAGCTGGCCAGTGATCGAGCGCTACCCTGCAGGGCGTCAGCTCACAATC
TTCAATAGCCAAGCAACCATAATAATTGGCGGGAAAGAGCAGGGCCAGCCCTTCCAGGGCCAGCTCTCTGGGCTGTACTACAA
TGGCTTGAAAGTTCTGAATATGGCAGCCGAAAACGATGCCAACATCGCCATAGTGGGAAATGTGAGACTGGTTGGTGAAGTGC
CTTCCTCTATGACAACTGAGTCAACAGCCACTGCCATGCAATCAGAGATGTCCACATCAATTATGGAGACTACCACGACCCTG
GCTACTAGCACAGCCAGAAGAGGAAAGCCCCCGACAAAAGAACCCATTAGCCAGACCACAGATGACATCCTTGTGGCCTCAGC
AGAGTGTCCCAGCGATGATGAGGACATTGACCCCTGTGAGCCGAGCTCAGGTGGGTTAGCCAACCCAACCCGAGCAGGCGGCA
GAGAGCCGTATCCAGGCTCAGCAGAAGTGATCCGGGAGTCCAGCAGCACCACGGGTATGGTCGTTGGGATAGTAGCCGCTGCC
GCCCTGTGCATCCTTATCCTCCTCTATGCCATGTACAAGTACAGAAACCGGGATGAAGGCTCATACCATGTGGACGAGAGTCG
AAACTACATCAGTAACTCAGCACAGTCCAATGGGGCTGTTGTAAAGGAGAAACAACCCAGCAGTGCGAAAAGCTCCAACAAAA
ATAAGAAAAACAAGGATAAAGAGTATTATGTCTGA

Figure 2-1: NRXN1 structure and gRNA design strategy, A)Schematic representation of NRXN1 isoforms and the first shared exon (Exon19) is highlighted as it represents the gRNA target region. B) NRXN1 human gene Fasta format showing the designed gRNA complimentary sequence (highlighted in yellow)

The PCR primers in Table 2-2 were used for genotyping; the PCR products were run on a 1% agarose gel.

Table 2-2: PCR primers used for NRXN1 genotyping

Pf	ACTTGGTCCACTTTCTTGAGCA
Pr	ACTGGTTTCTGGGGGAAGGA

2.1.2 Insertion of OPTi-OX system in *NRXN1*-patient lines

2.1.2.1 Strategy for GSH loci targeting with the OPTi-OX system

The pR26_CAG-rtTA was constructed by introducing the rtTA coding sequence into the BamHI/MluI sites of pR26_CAG-EGFP [230]. The pAAV_TRE-NGN2 was constructed by cloning the NGN2 coding sequence into the SpeI/EcoRI sites of pAAV_TRE-EGFP as previously described [229], [240]

2.1.2.2 Gene targeting

<u>The targeting of the GSH loci with the TET-ON components:</u>

Targeting of the hROSA26 locus: the 211_NXM and 092_NXF patient iPSC lines were seeded on 6 well geltrex-coated (Fisher Scientific, A1413302) plates and maintained in Stemflex media (Thermofisher, A3349401) until they reached 80-100% confluency. Twelve to twenty-four hours prior to nucleofection, the media was changed by Pen/Strep (Gibco,11548876) free Stemflex media.

On the day of nucleofection, the nucleofection master mix was prepared by adding 4 µg from both the gRNA/Cas9n plasmid and the pR26_CAG-rtTA vector to the P3 solution and P3 supplement supplied in the P3 Primary Cell 4D-Nucleofector X Kit (LONZA, V4XP-3032). The master mix was then kept on ice while the cells were prepared.

Before preparing the cells, 2x p100 plates were coated with geltrex, and Pen/Strep-free fresh media with CloneR (STEMCELL, 5889) (1:10) was added and left at the incubator. The cells were treated by Stempro accutase (Thermofisher, A1110501) for 5 min to dissociate them into single cells. The cells were then collected in a falcon tube containing Stemflex and centrifuged at 1000rpm for 3 min. The pellet was then resuspended in fresh media containing (1:10). Using the automated cell counter, the cells were counted and $2x10^6$ cells were transferred to a new tube and centrifuged. The cell pellet was then resuspended in 100 µL of the prepared master mix and transferred to the cuvette.

The cells in the cuvette were then pulsed using the CA-137 cycle in the Lonza 4D-Nucleofector System. The nucleofected cells were then supplemented with 500 µL Pen/Strep free media with CloneR(1:10) and incubated for 5 min. The cells were then divided into the 2x geltrex-coated p100 plates.

Forty-eight hours post nucleofection the media was changed with Stemflex with CloneR(1:10) and the following day the media was changed with Stemflex + CloneR(1:40). Three to four days post nucleofection, neomycin-resistant cells were selected by adding 100 µg/ml G418 (Sigma-Aldrich) for 4 days. The surviving colonies were then picked and expanded on geltrex-coated 24-well plate. The DNA was extracted using wizard genomic DNA purification kit (Promega, A1120). To validate the insertion of the pR26_CAG-rtTA construct in the selected colonies, their DNA was amplified using specific primer pairs as listed in Table 2-3. The PCR products were visualized on 1% agarose gel run for 30 min at 60V. After genotyping, the clones containing the pR26_CAG-rtTA plasmid were expanded and passaged to be used for downstream applications.

Table 2-3: pR26_CAG-rtTA PCR primers and program used for genotyping

A	B	C	D	E
	WT	Left Backbone	Left junction	Right junction
Pf	GCCAATCAGCGG AAGCCG	CGTTGTAAAACG ACGGCCAG	GCCAATCAGCGG AAGCCG	ATGCCCTGCCAA TCGAGATG
Pr	CGCGGGTTAGAC AACTTGCTG	GTGCCCAGTCAT AGCCGAAT	TCAAGGAAACCC TGGACTACTG	CGCGGGTTAGAC AACTTGCTG
Expected size	~2.2kb	~1.1kb	~1.2kb	~1.8kb
PCR Program				
	WT	Left side	Right side	

Initial Denaturation		98°Cx5'	98°Cx5'	98°Cx5'
X35 cycles	Denaturation	96°Cx30"	98°Cx30"	94°Cx30"
	Annealing	63°Cx30"	60°Cx30"	59°Cx30"
	Extension	70°Cx2'	70°Cx2'	70°Cx2'
Final Extension			70°Cx5'	70°Cx5'

Targeting of the AAVS1 locus: the 211_NXM-rtTA and 092_NXF-rtTA iPSC lines were seeded and maintained as described above. The nucleofection master mix was prepared by adding 4 µg from both the AAVS1 ZFN plasmids and the pAAV_TRE-NGN2 to the P3 solution and P3 supplement supplied in the Lonza P3 Primary Cell 4D-Nucleofector X Kit). The master mix was then kept on ice while the cells were being prepared. The cells were prepared and nucleofection was done as described above. Five days post nucleofection, the iPSC resistant cells were selected by adding 1µg/ml of puromycin (Sigma-Aldrich) for 3-5 days.

Colonies were picked and expanded. And the primers in Table 2-4 were used for genotyping to verify site- specific transgene integration, to determine the number of targeted alleles, and to exclude off-target integrations. The PCR products were visualized on a 1% Agarose gel run for 30 min at 60V.

Table 2-4: pAAV_TRE-NGN2 PCR primers and program used for genotyping

A		B	C	D	E
		WT	Left Backbone	Left junction	Right junction
Pf		CTGTTTCCCCTTCCCAGGCAG	CGCTATTACGCCAGCTGGC	CTGTTTCCCCTTCCCAGGCAG	TGCAGGGGAACGGGGCTCA
Pr		TGCAGGGGAACGGGGCTCA	TGAGGAAGAGTTCTTGCAGCTC	TCGTCGCGGGTGGCGAGGCGCACCG	TGCAGGGGAACGGGGCTCA
Expected size		~1.7kb	~1.2kb	~1.1kb	~1.3kb
PCR Program					
		WT	Left side		Right side
Initial Denaturation		94°Cx5'	94°Cx5'		94°Cx5'
X35 cycles	Denaturation	96°Cx30"	98°Cx30"		94°Cx30"
	Annealing	63°Cx30"	63°Cx30"		63°Cx30"
	Extension	65°Cx2'	65°Cx1'30"		65°Cx1'30"
Final Extension		65°Cx10'	65°Cx10'		65°Cx10'

After genotyping the selected clones and verifying the proper integration of both constructs in the GSH loci, three clones per line were selected for expansion and further passaging. To validate the regulation of the transgene by DOX, the OPTi-NGN2 PSCs were treated with neural induction media with and without DOX and RNA was collected from the cells at Day 4 and Day 14 as indicated in the neuronal timeline in Figure 2-2. Quantitative PCR was used to quantify the mRNA level of known pluripotent and neuronal genes in the cells treated with DOX vs the non-treated ones.

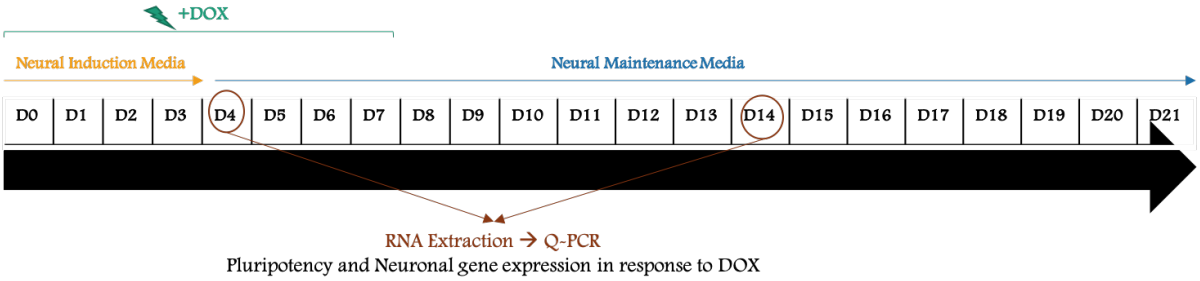


Figure 2-2: Neural induction timeline; cells were treated with neural induction media until day 3 and then moved to neural maintenance media in the presence or absence of DOX. Cell were collected at Day 4 and Day 14 to check for pluripotent and neural markers in response to DOX

2.2 Cell culture

2.2.1 Thawing

The iPSC lines are kept in liquid nitrogen for long term storage. Prior to thawing the iPSCs, two wells (per line) of 6-well plate were coated with 1% geltrex in DMEM: F12 (Gibco, 31330-038) for 4 hours at 37 °C. The wells were then washed with D-PBS (Gibco, 14190-094) and left in the incubator until the iPSCs were ready for plating.

The cryovials containing the iPSCs were briefly placed in a the water-bath, and then gently suspended in the Stemflex + 10 µM ROCK inhibitor (ROCKi) (TOCRIS, 1254/10) and centrifuged at 1000rpm for 3 min. The pellet was then resuspended in the same prewarmed mix and plates in the geltrex-coated 6 well plate. The media was changed with Stemflex media without ROCKi) every 48 hours.

2.2.2 Maintenance and expansion

For regular maintenance and expansion, iPSCs were passaged every week whenever the confluency reached 80% of the well. The cells were washed once with 1X D-PBS and then 500 µl of Relesr (STEMCELL, 5872) was added per one well of a 6-well plate, aspirated and then left at room temperature for 2 min. Cells were then detached with a cell lifter and suspended in a 15 ml falcon tube containing Stemflex media. Once the cells were settled at the bottom of the falcon tube, the medium sized colonies were carefully seeded onto 6 wells of the 6well plate. The media was regularly changed every 48 hours.

2.2.3 Freezing

To freeze the iPSCs, Relesr was used as described above, pelleted and the pellet was resuspended in 10% DMSO (Sigma, D2650) in 1 ml of Stemflex media and stored in a Mr. Frosty at 80ºC for 24hours before being transferred to the liquid nitrogen.

2.3 Neuronal Differentiation

2.3.1 Primary culture

Primary mixed glial cultures were derived from P0-P2 neonatal Spraque Dawley rats and were generated along the previous guidelines [241], with minor modifications [241], [242]. The pups were euthanized following Schedule 1 rules and regulations from the Home Office Animal Procedures Committee UK (APC).

To maintain aseptic conditions, all procedures were performed in a laminar flow hood. A horizontal flow hood was used to perform dissections and a vertical flow hood for tissue culture. In brief, the meninges, midbrain and olfactory bulbs were removed, and dissociated rat neonatal cortices cut thoroughly and incubated at 37°C for 30 min in Minimum Essential Medium Eagle (MEM) (Sigma, M7278) containing 4% Papain (Sigma,P3125), 1% of 4 mg/ml DNase I Type IV (Sigma, D5205) and 1% of 24 mg/ml Lcysteine. After the digestion step the dissociated cells were plated into PDL-coated (Poly-D-Lysine) (Sigma, P6407) cell culture flasks at a density of 2 brains/T75 flask. These mixed glia cultures were cultured for 10 days in DMEM high glucose (Sigma, D6546) supplemented with 10% FBS (Fetal Bovine Serum) (Gibco, 10500064) and 1%Pen/Strep and kept under a humidified atmosphere at 37°C and 7% CO_2.

2.3.2 Neural induction

The OPTi-NGN2 iPSCs were maintained in colonies in 6 well plate as described above. For neural induction cells were dissociated into single cells by using 1 ml of Stempro accutase per well and left in the incubator for 5min. The cells were then collected in a DMEM: F12 containing flacon tube. The cells were centrifuged at 1000rpm for 3 min and the resulting pellet is resuspended in 10µM ROCKi in 1 ml of Stemflex media. Using the automated cell counter, the cells were quantified and plated in a 6wp at a density of 500K per well.

After 24 hours (Day0), the neural induction was started by adding neural induction media+ DOX (Sigma, D9891). The same was repeated for Day 1 and 2. The iN were then co-cultured with Rat glia (as demonstrated in Figure 2-3) on Day 3 and transferred to Neural Maintenance media+DOX until Day7. After the first week, half media change is required with neural maintenance media, without DOX. The recipes for neural induction and neural maintenance media are listed in Table 2-5.

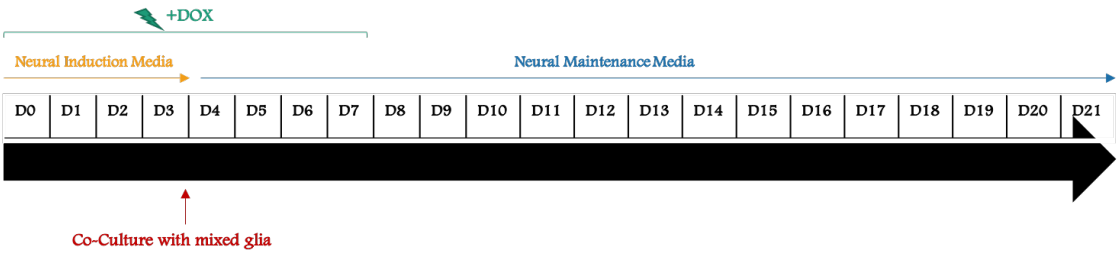


Figure 2-3: General Neural induction timeline; The green arrow show the time window for the DOX treatment (D0-D7), the yellow arrow is the time window for the Neural induction media (D0-D3), the red arrow is the timepoint for the addition of the rat mixed glia to the iN and the blue arrow is the time window where the media was switched to the neural maintenance media until the end of the neuronal differentiation

Table 2-5: Recipes for neural induction and maintenance media

Induction media	Maintenance media
DMEM:F12 (Gibco, 31330-038)	Neurobasal (Gibco, 21103-049)
N2 supplement- 100x (Gibco, 17502048)	B27 supplement - 100x (Gibco, 17504044)
Glutamax- 100x (Gibco, 35050-038)	
Non-essential amino acids - 100x (Gibco, 11140-050)	
2-Mercaptoethanol - 50 µm, Gibco, 31350010)	
Penicillin-Streptomycin - 100x (11548876)	

2.3.3 Rat Glia dissociation

The mixed-glia culture was treated with 5 ml of Trypsin-EDTA (Gibco, 25200056) and left at the incubator for 5 min. Once the cells appear to have dissociated from the flask, astrocyte media was quickly added to the T75 flask to deactivate the trypsin. The cells were then collected in a 50 ml flacon tube and centrifuged at 1200rpm for 5 min. The pellet was then collected and resuspend in Neural maintenance media and passed through a 33 mm diameter sterile syringe filter. Using the cell counter, the mixed glia was quantified and ready for co-culture. The dissociation is performed primarily to favor/enrich for astrocytes only, however, unpublished RNA-Seq data

from our lab revealed that the dissociated cells are a mixture of astrocytes, oligodendrocytes and microglia [225].

2.3.4 Co-culture

The iNs need to be co-cultured with the dissociated glial cells to reach a more matured state. Based on morphology and location, the glial cells are divided into two main subsets, microglia and macroglia (astrocytes and oligodendrocytes). Microglia are the first glial cells observed in the developing brain, they are derived from yolk-sac derived macrophages, while astrocytes are derived from a neural lineage, produced at the final stages of neurogenesis. Glial cells have distinct and diverse functions that direct brain development in several ways, and they work together with neurons to obtain normal brain functions. Microglia are involved in neuronal survival, programmed neural cell death, axonal and synaptic pruning.

They are also implicated in astroglial proliferation and differentiation. Similarly, studies have shown that astrocytes play a crucial role in neuronal survival, axonal guidance, neuro/gliogenesis as well as postnatal synaptogenesis. In-vitro the addition of astrocytes improved viability and maturity of NGN2-induced neurons, and allowed the formation of functional synapses [231].

On day 3 the iNs are dissociated using Stempro accutase as described before. The pellet was resuspended in neural maintenance media and counted using the cell

counter. The iNs and the glial mixed cultured were plated on PDL/Laminin-coated (Sigma, P6407/L2020) plates at a 1:1 ratio at the appropriate density according to vessel used and the downstream application as shown in the Table 2-6.

Table 2-6: Cell seeding density

Vessel	Density (per cell type)	Application
6 well plate (6-wp)	$5x10^5$	RNA extraction
12 well plate (12-wp)	$4x10^5$	RNA extraction
24 well plate (24-wp)	$2.5x10^5$	ICC
35mm µ-Dish	$3.5x10^5$	Calcium imaging
96 well plate (96-wp)	$8x10^3$	Neurite outgrowth assay
6 well multichannel array	$5x10^4$	MEA

2.4 RNA Extraction

As described above the iPSCs were induced until Day3 and then cocultured with mixed glial cells (1:1 ratio) at a density of $5x10^5$ /well (6-wp) for RNA-Seq and $4x10^5$/well (12-wp) for Q-PCR (Table 2-6). Depending on each experiment, the cells are harvested on their respective time points, as indicated in Figure 2-4 for the downstream application.

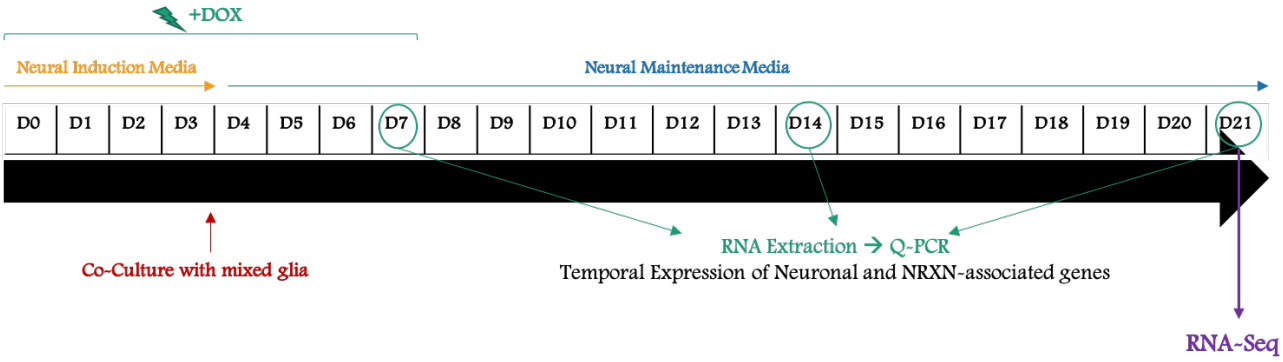


Figure 2-4: Neural induction timeline and collection timepoints per experiment; iN will by collected at Day 7, 14 and 21 for RNA extraction for measuring expression of neuronal and NRXN-associated genes (as indicated by the green circles and arrows) and at day 21 for RNA-Seq (purple arrow). The yellow arrow indicates the days the cells are treated with neural induction media+ DOX (green) and neural maintenance media (blue), the red arrow shows the day the iNs are co-cultures with the rat mixed glia (day3).

The RNA was extracted from the iPSC-induced neurons using Qiagen RNeasy mini kit (74106). The neural maintenance media was removed by aspiration, and cells washed once with 1x D-PBS before addition of RLT buffer (provided in the kit) to the well to form a cell lysate. Cell lysate was then collected in a microcentrifuge tube, an equal amount of 70% ethanol was then added to the lysate and was mixed well by pipetting. The mixture was then transferred to a RNeasy spin column placed in a collection tube and centrifuged for 15 sec at 10,000 rpm and the flow-through was discarded. The RW1 buffer (350 µl) provided in the kit was then added to the cells and centrifuged at 10,000rpm for 15 sec and the flow-through was discarded. In a separate tube 10µl DNase I stock solution to 70 µl Buffer RDD and were mixed by inversion. The DNase mix (80 µl) was then directly added to the column and placed at RT for 15 min. The RW1 buffer (350 µl) was again added to the column and centrifuged again at 10,000rpm for 15 sec. The RPE buffer (500 µl) was then added to the column, this was followed

by centrifugation at 10,000rpm for 15 sec and the flow-through was discarded. This step was repeated but this time the centrifugation was for 2 min to dry the spin column, ensuring that no ethanol is carried over during RNA elution. Finally, the spin column was placed in a new 1.5 ml collection tube, 50 µl RNase-free water was added directly to the spin column membrane and centrifuged at 10,000 rpm to elute the RNA. The RNA was quantified with a NanoDrop 1000 Spectrophotometer (Thermo scientific). And is ready for downstream applications. For neuronal and *NRXN1*-associated genes' gene expression the iN were collected at 3 timepoints, Day7, 14 and 21. For RNA-Seq, cells were collected at Day21.

2.5 Q-PCR

2.5.1 cDNA synthesis

The Maxima® First Strand cDNA Synthesis Kit (Thermofisher, K1642) was used to synthesise complimentary DNA from 1 µg of RNA following the manufacturer's protocol. The master mix was prepared by adding 5x Reaction mix, maxima enzyme mix and the appropriate amount of template RNA (100 ng/µl) in nuclease free water in a total volume of 20µL. The reaction was then incubated for 10 min at 25°C followed by 15 min at 50°C. And finally, the reaction was terminated by heating for 85°C for 5 min. The product was then directly used in qPCR or stored at -20°C for up to 1 week.

2.5.2 Taqman® Assay

The two major isoforms of *NRXN1*, α and β were measured using TaqMan™ Gene Expression Assay; (FisherScientific; Hs00985123_m1, Hs00985129_m1, Hs00373346_ml and Hs99999901_s1) for the detection of *NRXN1* (pan), *NRXN1α*, *NRXN1ß* and 18S rRNA (housekeeping gene). These assays were performed in a duplex real-time PCR reaction, using a FAM™ dye-labelled assay for the target gene in combination with an endogenous control labelled with, VIC™ dye.

2.5.3 SYBR™ green

For the temporal expression of neuronal and *NRXN1*-associated genes during neural induction, gene sequences were extracted from the NCBI GenBank and University of California and Santa Cruz (UCSC) genome and bioinformatics browser and primers were designed using Primer3Plus (v. 0.4.0) software. For the Q-PCR reaction, 0.4µl of the primer pair (0.2 µM), 3 µl of cDNA and 4.1 nucelase-free water were mixed with 2.5µl of the PowerUp™ SYBR™ Green Master Mix (ThemoFisher Scientific, A25780). The samples were run on Applied Biosystems 7500 fast PCR machine using the program described in Table 2-7 and using gene-specific primers as listed in Table 2-8. Samples were run in technical duplicates and normalised to 18S rRNA as the house keeping gene.

Table 2-7: The Q-PCR program used for TaqMan and SYBR green assays

Initial Denaturation	**95°Cx15 min**
X35 cycles Denaturation	95°Cx30 sec
Annealing	60°Cx30
Extension	72°Cx2'

Table 2-8: Primer pairs for Q-PCR assay

	Gene	**Pf**	**Pr**
Neuronal	VGLUT2	GGGAGACAATCGAGCTGACG	CAGCGGATACCGAAGGAGATG
	GRIA4	GGCCAGGGAATTGACATGGA	AACCAACCTTTCTAGGTCCTGTG
	MAP2	AGACTGCAGCTCTGCCTTTAG	AGGCTGTAAGTAAATCTTCCTCC
Pluripotent	NANOG	AGCAGATGCAAGAACTCTCCAA	TGAGGCCTTCTGCGTCACAC
	OCT4	GTGGAGGAAGCTGACAACAA	ATTCTCCAGGTTGCCTCTCA
Transgene	NGN2	TGTTCGTCAAATCCGAGACCT	CGATCCGAGCAGCACTAACA
RNA-Seq validation	OAF	GAAGGGGCAGAGTCAGTTCC	GTACTTGAGCATGCAGGGGT
	FN1	TGACAAGCAGACCAGCTCAG	GGCATTTGGATTGAGTCCCG

	FBLN1	CATCTCGCTGCCTACCTTCC	GGGGGTGACACTTAGCCTTC
	RPS21	TCCGCTAGCAATCGCATCAT	GCCCCGCAGATAGCATAAGT
	MMP14	GGCGAGTATGCCACATACGA	GTACTCGCTATCCACTGCCC
	GRM3	CTCCAACATCCGCAAGTCCT	ACACGTTGTATCGCCCCATT

The ΔΔCt method was used to measure gene expression levels of each sample at different time points relative to iPSC (before neural induction). Every experiment was replicated 3 times for each cell line, each replicate had a technical duplicate for the Q-PCR assay. The statistical analysis was performed using the Prism package of GraphPad software; the samples means were compared by a two-way ANOVA test with Tukey's multiple comparisons test.

2.6 RNA-Seq

2.6.1 Library Preparation

The RNA samples were handled by the NGS library facility at the Wellcome – MRC Cambridge Stem Cell Institute. The RNA samples underwent PolyA treatment using the NEBNext Poly (A) mRNA Magnetic Isolation Module (E7490S/E7490L) which isolates intact Poly(A)+ RNA from total RNA using paramagnetic beads coupled with Oligo d(T)25 to directly bind the desired fragments. The isolated mRNA was then processed using the NEBNext Ultra II directional RNA Library Prep Kit for Illumina®

(E7760S/E7760L) which produces strand specific libraries through use of the dUTP method to preserve directionality. Quality of final libraries was assessed using the Agilent 4200 Tapestation. The RNA integrity number (RIN) were obtained and shown in the electropherograms in Figure 2-5.

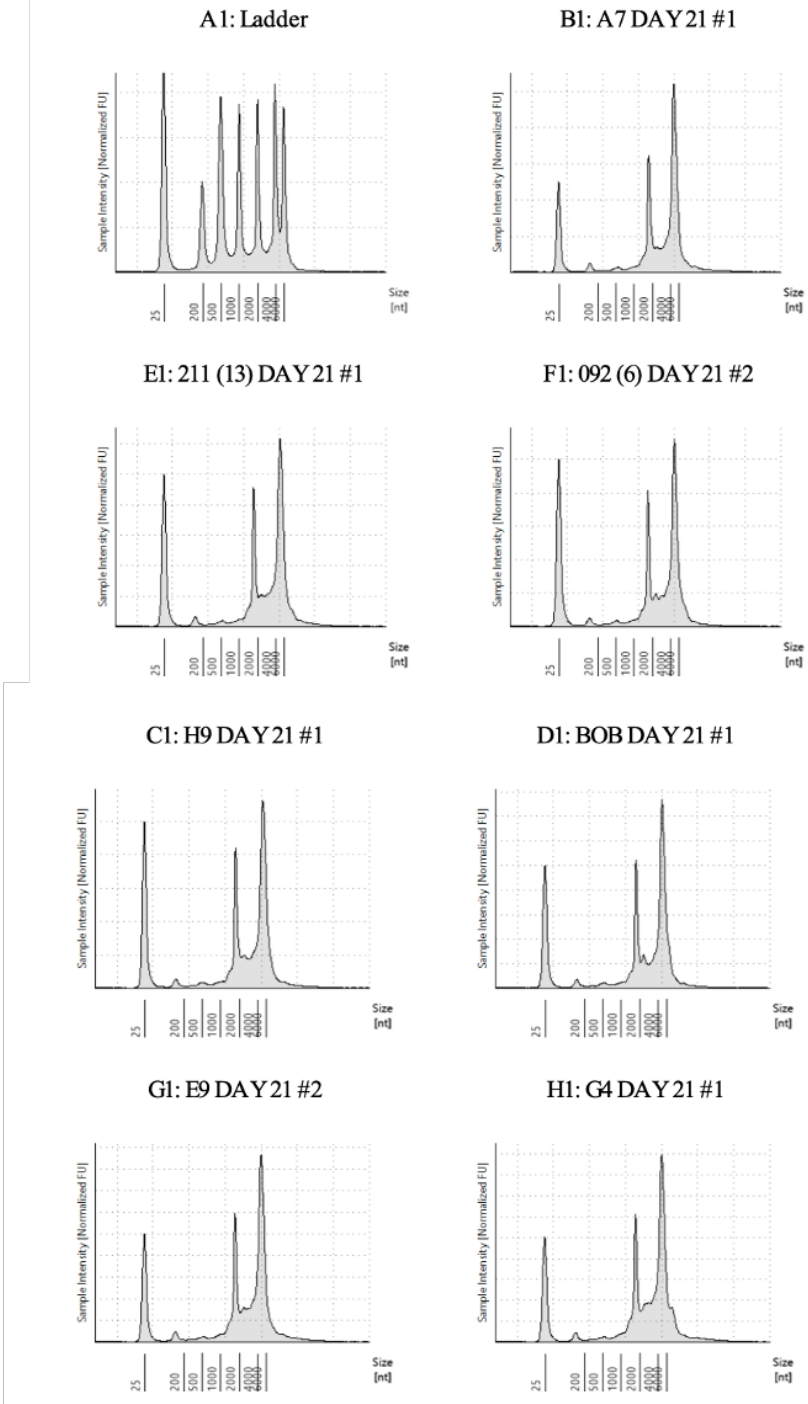


Figure 2-5: Electropherogram of the RNA extraction from the samples. The main peaks shown on the graphs represent the 18S and 28S ribosomal subunits (respectively). The small peaks on the left represent 5S ribosomal and tRNA. These peaks indicate the quality of the sample and show that no degradation has occurred.

2.6.2 Sequencing

The samples were sent for sequencing by the genomics unit at Cancer Research UK. The libraries were sequenced in a SLX_NOvaSeq6000 S1_Std to a depth of 200-250 million reads. FASTQ files were aligned to a reference genome by our collaborator Grant Belgard at the Bioinformatics CRO (Florida, USA).

2.7 RNA sequencing analysis

2.7.1 Raw data processing

Sargasso v2.0.1 was used to separate *Rattus norvegicus* (rat) and *Homo sapiens* (human) reads via alignment to the Rnor 6.0 and GRCh38.p12 / hg38 reference genomes [243]. Read count abundances were then generated from the separated raw data FASTQ files by pseudo-aligning these to reference transcriptomes using Kallisto v0.46.0, while adjusting for GC content bias and bootstrapping 50x. For human, abundances were counted over the GENCODE 'comprehensive' reference transcriptome (GRCh38.p12 / release 31), while, for rat, they were counted over Ensembl's Rnor 6.0.97 'cdna' and 'ncrna' combined [244]. Bootstrapped transcript isoform-level counts were then imported to R Programming Language v 3.6.1 (R) (R Core Team, 2016) and summarized to gene level counts (adjusting for gene-length) using the tximport v1.13.12 package in R.

Protein coding genes (total = 19703 for human; 22139 for rat) were then isolated from the raw counts based on GENCODE biotype keyword 'protein_coding'. A gene with zero counts across all samples was removed (total = 1479 for human; 3755 for rat). The raw counts of the remaining genes (total = 18224 for human; 18384) (as a tximport object) were then converted to a DESeq2 v1.25.10 (Love et al., 2014) object for normalization with βPrior set to FALSE. For downstream analyses, the negative binomial-distributed normalized counts were converted to regularized log (rlog) counts via the rlog function of DESeq2 in R, with blind set to FALSE. Variance-stabilized counts were also generated.

Tables of normalized and normalized + transformed:

- [human|rat]/protein_coding/counts/NormCounts.tsv (normalized RNA-seq counts)
- [human|rat]/protein_coding/counts /RlogCounts.tsv (normalized + regularized log transformed RNA-seq counts)
- [human|rat]/protein_coding/counts /VSTcounts.tsv (normalized + variance stabilized RNA-seq counts)
- [human|rat]/protein_coding/counts /RLogZcounts.tsv (normalized + regularized log transformed RNA-seq count Z-scores)

2.7.2 RNA-seq differential expression analysis

Differential expression analysis was conducted on the negative binomial-distributed normalized counts with FDR set at 5%. Following differential expression, log (base 2) fold changes (log_2FC) were shrunk via the lfcshrink function of DESeq2. A gene was defined as differentially expressed if it passed Benjamini-Hochberg Q≤0.05 and absolute $log_2FC \geq 2$ for human, and Benjamini-Hochberg Q≤0.05 and absolute $log_2FC \geq 1$ for rat. MA plots were generated using base R functions by plotting log_2FC (y-axis) versus natural log normalized counts (x-axis). Prior to logging, normalized counts were shifted by +1 to avoid negative log values. Genes passing Benjamini-Hochberg Q≤0.05 and absolute $log_2FC \geq 2$ were highlighted red. Volcano plots were generated using the EnhancedVolcano package v1.3.1 [245]. Plots were then plotted together using grid.arrange from the gridExtra v2.3 package. For labelling, cut-offs were set to unadjusted P≤0.0001 and absolute $log_2FC \geq 2$ for human, and unadjusted P≤0.0001 and absolute $log_2FC \geq 1$ for rat.

2.7.3 RNA-seq normalization QC

The dispersion of normalized counts was checked by plotting the maximum likelihood estimate of dispersion (overlaid with the final dispersion estimates) versus the mean of normalized counts. Unsupervised sample clustering was also performed using the rlog counts via the dist and hclust functions from the stats R package (using default parameters). These were then plotted as a symmetrical samples heatmap with the heatmap.2 function of gplots v3.0.1.1 package [246]. The distribution of rlog counts

across samples was additionally checked via a box-and-whisker plot via the boxplot function from graphics R package. Pairwise scatter plots were generated from rlog normalized counts via the scatterplotMatrix function from the car v3.0-3 package.

2.7.4 Highly expressed genes

The top 50 most highly expressed genes were determined by first averaging the unlogged, normalized counts across each sample group and then ordering by decreasing order (high to low counts). To generate the heatmap, the heatmap.2 function of the gplots package was used with both row and column dendrograms and clustering disabled.

2.7.5 PCA

Principal component analysis was performed using the PCAtools R package [245]. Regularized log counts were used as input. A bi plot comparing PC1 and PC2 was generated, as was a pairs plot comparing PCs 1-5 on a pairwise basis. The main genes responsible for variation along each PC was generated via a loadings plot. Finally, correlation between PCs and metadata was generated via an eigencorplot.

2.7.6 Supervised clustering and heatmap

Supervised (filtered) clustering was performed using the Heatmap function of the ComplexHeatmap v2.1.0 package. Regularized log counts were converted to the Z scale (scaled by row/gene) and then clustered via 1 minus Pearson correlation distance and Ward's linkage ('ward.D2'). A samples box-and-whisker plot of Z-scores was added to the heatmap bottom. A colour bar indicating the different sample groups was added at the heatmap top. To supervise clustering, genes from each differential expression analysis at Benjamini-Hochberg Q≤0.05 and absolute $\log_2$FC≥2 were included for human, and Benjamini-Hochberg Q≤0.05 and absolute $\log_2$FC≥1 for rat.

2.8 Immunocytochemistry

2.8.1 For Neurite outgrowth assay

Cells were seeded at $8x10^3$ cells/well on a PDL/Laminin-coated 96-well plate. The cells were induced and were not co-cultured with rat glial mixed culture for this particular assay. On Day 6 as indicated in Figure 2-6, cells were gently rinsed with 1x D-PBS then treated with 4%Paraformaldehyde (PFA) (Alfa Aesar, J61899) and left at RT for 20 min. Then the PFA is aspirated and the cells are rinsed twice with 1x D-PBS. Fixed cells were then treated with a permeabilization/blocking solution (2% normal donkey serum (NDS) (Sigma, D9663) and 0.1% triton X-100) for 1 hour at RT. The cells are treated with the chicken polyclonal MAP2 antibody (Abcam, ab5392) diluted in 2%NDS blocking solution and left overnight at 4°C. Following incubation, the cells were washed twice in 1x D-PBS followed by 3 washes (10-15 min each wash) in D-PBS.

The wells were then incubated for 1hour in the dark in 2%NDS blocking solution containing Donkey anti-Chicken secondary antibody (Thermofisher, SA1-72002) and DAPI (Theromfisher, b2883).

The cells were then washed again as mentioned above and left in D-PBS. The plates were wrapped in parafilm (to avoid evaporation) and in foil and stored at 4°C until imaging.

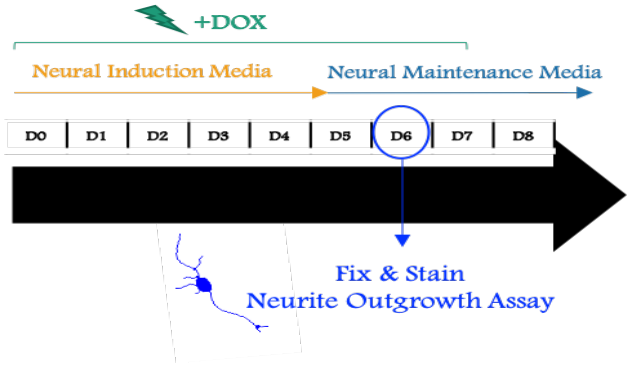


Figure 2-6: Figure 2 6: Neural induction for neurite outgrowth assay, the blue arrow indicated the timepoint chosen to fix and stain the cells (Day6) for the neurite outgrowth assay

The epi-fluorescent high-content screening (epi-HCS) platform at the Wohl Cellular Imaging Centre in King's college London was used for this experiment. The epi-HCS has four channels; two channels, the 405nm for DAPI and the 488nm channel for the morphological marker (MAP2) were used for this experiment. The images were acquired and processed using The HCS Studio software, this allows the automated detection of neurites based on an inclusion/exclusion criterion to accurately detect the neurites to be analysed.

The neurite detection criterion is set up based on 3-5 fields across 5 different wells. The software identified the cell body of the neurons based on the MAP2 (neuron morphological marker) staining and then traced the neurites. The inclusion/exclusion of the traced neurites depended on many parameters, such as minimum neurite length, minimum neurite thickness, minimum neurite signal intensity, and min/max neurite directionality. Once these parameters were optimised for a comprehensive morphological examination, the whole plate was scanned and the total neurite count, neurite length and total branch point count per neuron were quantified.

Statistical analysis: The ROUT method was used for determining the presence of any outliers, this method relies on the false discovery rate, in this analysis the maximum false discovery rate was set to 1%, aiming for no more than 1% of the identified outliers to be false. The Kruskal-Wallis test was used to find significant differences between the cell lines. The experiment was repeated 3 times and each line had 5 technical replicates.

2.9 Functional activity

The functional activity of the neurons was assessed using two methods, multi-electrode array was used for synchronised activity and calcium imaging for spontaneous activity. The iNs were induced as described in Figure 2-7 and mixed and

the MEA was used to record their electrophysiological activity at 4 timepoints (Day 14,21,28 and 35). The calcium imaging videos were taken only one at Day35.

Figure 2-7: Neural induction timeline for Calcium imaging and MEA recordings: Day35 (blue arrow).

2.9.1 Multielectrode Array (MEA)

The extracellular electrophysiological activity of live-mature iNeurons was recorded using the 60-6 well MEA200/30iR-Ti chips (Multi Channel Systems MCS GmbH). These chips contain 6 independent wells, each well containing nine electrodes with an internal reference electrode, so a total of 60 electrodes (Figure 2-8). The wells are separated by a macrolong ring and two bars extending from the centre, allowing to run 6 different samples, condition and/or treatment simultaneously. The MEAs have to be treated for their hydrophilic surface and coated prior to cell plating. Cells are cultured directly on the coated-MEA and the electrodes can be used to record their activity or even for stimulation. At the time of recording, the MEA is mounted on an amplifier in a specific orientation. The signals received from the cells in contact with the electrodes are recorded, amplified and sent to the data acquisition computer using the ME2100-system (Multi Channel Systems MCS GmbH).

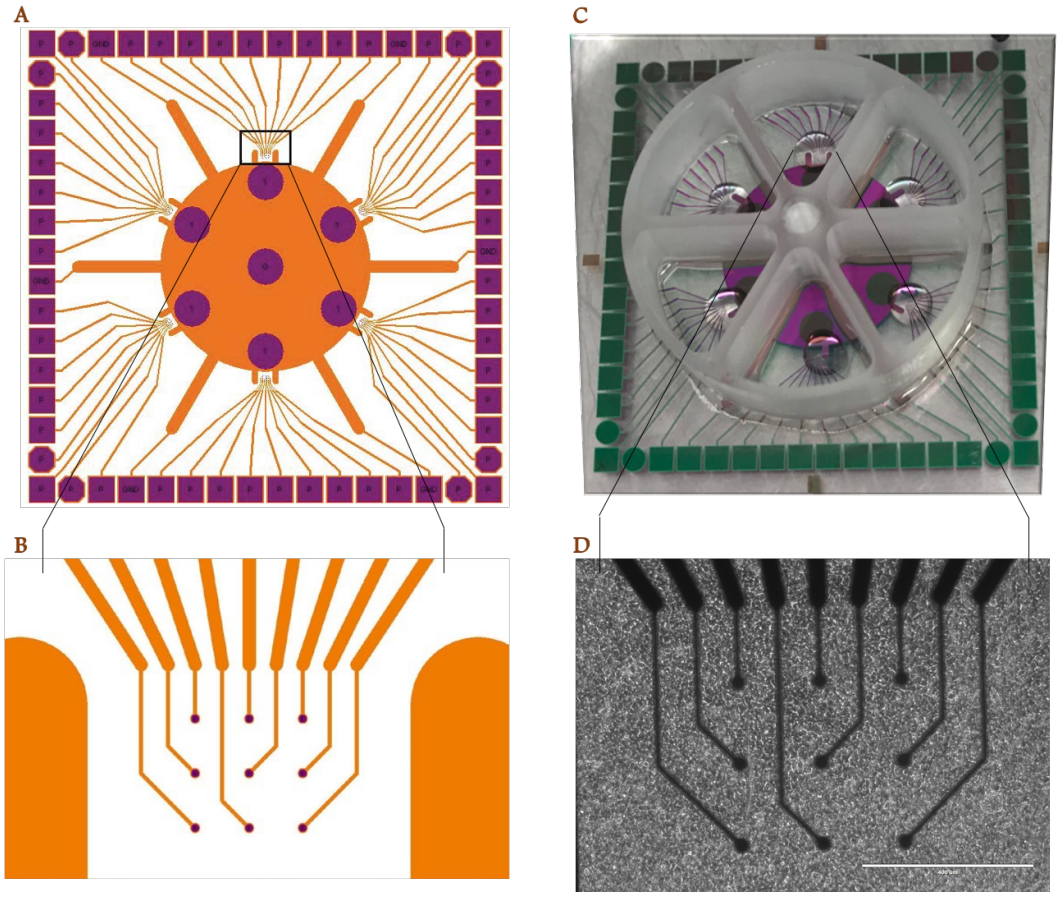


Figure 2-8: Multielectrode array, A) 60-6well MEA schematic representation, each well/chamber is separated by two bars, the purple circles represent the electrodes in each field, B) A zoom-in view on one well of the 60-6well MEA showing the nine electrodes, C) The 60-6well MEA with a 10 µl drop of laminin coating prior to plating, D) A zoom-in view on one well of the 60-6well MEA showing the plating of the iNeurons on top of the electrodes.

2.9.1.1 Hydrophilic treatment and sterilisation

To treat the hydrophobic surface of the new MEAs, 1 ml of sterilised bovine serum albumin (BSA) (Thermofisher, 23209) was added to cover the surface and was left in the incubator overnight. For sterilisation, 1%Tergazyme (in ddH_2O) was added to the MEA wells and left overnight. The MEA wells were then washed 3x with ddH_2O. The

wells were then gently rinsed with 70% Ethanol and washed again 3x with ddH_2O. The MEAs were then left to dry overnight under the UV in the laminar flow hood.

2.9.1.2 Plating

Prior to seeding, the MEA are coated with 50µg/ml PDL in ddH_2O then washed 3x with ddH_2O and left to dry under the UV. And whilst dissociating the iN Day3 (as described above) and glial mixed cells, a 10 µL drop of Laminin (Sigma, L2020) (20 µg/ml) was added the centre of each well as shown in Figure 2-8C, so that it only covers the electrode surface and left at 37°C for 30 min (maximum). Once the iN and glial cells were resuspended in the appropriate amount of Neural maintenance media + 10µM ROCKi to bring their concentration to $10x10^6$ cells/ml and are mixed together at 1:1 ratio. The Laminin was then carefully aspirated and replaced by a 10 µl drop of the cells mix ($5x10^4$) to cover the electrodes as presented in Figure 2-8D. The cells are left to attach for 1 hour and then 300 µl of neural maintenance media + DOX was added to cover the well. The co-cultured are then are maintained as described above and electrophysiological activity is recorded on Day35 for 15 minutes

2.9.1.3 Recording & Analysis

Before recording the contact, pads were gently wiped with a soft tissue moistened with alcohol before placing in the MEA amplifier. The raw voltages were subjected to a Bessel filter (a high pass cut off 200 Hz). Spikes were detected using a threshold of

5 times the standard deviation of noise signal calculated over a 10 ms moving window. Electrodes with less than 5 spikes per minute were considered inactive and were excluded from analysis. Single electrode bursts were identified if a minimum of 5 spikes were detected in a burst with a minimum 100 ms interval between bursts. Network bursts were defined as bursts firing simultaneously in 30% of active electrodes.

2.9.2 Live Imaging

The OPTi-NGN2 iPSCs are grown and maintained in colonies. For neural induction cells have to be dissociated into single cells and seeded at around 80% confluency. And as described above, the cells are induced the next day with Neural induction medium, then cocultured with glial mixtures on Day3 at a density of 3.5 $x10^5$ of each cell type in 35mm µ-Dish.

From day 7 onwards, the media was half changed every other day with neural maintenance media (without DOX). On Day 35, the cells are ready to be stained for live imaging. The Pluronic^{TM} F-127 (Invitrogen, P3000MP) has to be incubated at 37°C to be soluble, and 10mM D-Glucose was added to the Live imaging solution (Thermofisher, A14291DJ). Then the dye loading solution was prepared by diluting 10µM of Calbryte520 and 0.04% The Pluronic^{TM} F-127 in the live imaging solution. The staining solution was then properly mixed by vortexing and incubated in the dark for 30 min at 4°C.

Once the staining solution was ready, it was added to the cells and left to incubate in the dark at 37°C for 45 min. The cells were then washed with the live imaging solution and ready for imaging.

2.9.2.1 Data Processing

Data analysis was performed in R [247] (R Core Team, 2019) using the EBImage package [248] for image processing.

2.9.2.2 Drift Correction

Lateral drift was observed in some time series. Drift was measured using a template matching approach: The first frame in the time series was cropped to remove a margin 25 pixels wide. This cropped frame became the template. The template was slid over the second frame in the time series and its similarity to the patch of the second frame it covered was quantified using the sum of absolute differences (SAD).

Once the region of the second frame that was the best match to the template was identified, the lateral offset (in pixels) between the two frames could be calculated. The second frame was translated by the lateral offset, so that it was correctly aligned with the first frame. The procedures applied to frame 2 in steps 2-4 were repeated for all remaining frames.

For more details of the measurement and correction of drift, see the following scripts:

https://gist.github.com/WaylandM/279e2085a09ae19f6f3ed477cf066bf0

https://gist.github.com/WaylandM/8a32d468d8bd73e4b324498fff57efc8

2.9.2.3 Segmentation

Each time series was processed as follows: A maximum projection of the time series was computed. Adaptive thresholding was applied to the maximum projection to classify pixels as foreground or background using the *thresh* function in EBImage. A rectangular window measuring 20x20 pixels was moved across the image and the thresholding offset (offset from the averaged value in the window) was set to 0.002. Morphological noise was removed using an opening operation (https://en.wikipedia.org/wiki/Opening_(morphology)). The structuring element (brush) was diamond shaped and had a width of 7 pixels. Holes in objects were filled using the *fillHull* function in EBImage. The EBImage function *bwlabel* was used to label objects (connected sets of pixels) in the foreground image. Labelled objects touching the edge of the image were excluded from analysis. The labelled objects were also filtered on size. Objects with fewer than 30 pixels were removed. The remaining labelled objects were assumed to represent cell bodies and used as masks for measuring signal intensity in each frame of the time series. Cell body signal intensities were background corrected. The signal intensity of each cell body across all frames in the time series was scaled to the range 0:1 and smoothed using LOESS (locally estimated scatterplot smoothing) with span of 0.05 and degree of 2.

2.9.2.4 Data Analysis

Each time series was processed as follows: The signal intensity of each cell body was plotted against time so that patterns in activity could be observed. Peaks in signal intensity were detected using the following algorithm: https://gist.github.com/dgromer/ea5929435b8b8c728193 Delta (local threshold for peak detection) was set to 0.2.

For each time series the following summary statistics were computed:

1. Synchronicity of activity was measured by computing the correlation in signal intensity between all pairs of cell bodies using Pearson's correlation coefficient. The pairwise correlation coefficients were summarized for each time series by computing the mean, median and standard deviation.

2. Another measure of synchronicity was obtained by measuring the proportion of cell bodies that were showing simultaneous peaks in activity. A moving window with a width of 3 timepoints (1,170 ms) was slid through the time series and the number of cell bodies with a peak in each interval was recorded. The interval with the highest number of coincident peaks was identified. The number of cell bodies peaking in this interval was expressed as a proportion of the total number of cell bodies to provide a summary statistic.

3. Activity was measured as the number of peaks in activity per cell body, summarized by mean, median and standard deviation.

2.10 Statistical Analysis

All cell lines were treated the same way and run in parallel, a minimum of three biological replicates per experiments and two to five technical replicates were used per experiments. All datasets were tested for normality, datasets found to be normally distributed were analysed using parametric statistical tests while datasets found to be abnormally distributed were analysed using their non-parametric equivalent. Two-way one-way analysis of variance (ANOVA) was used for temporal endogenous expression RNA analyses with an alpha level of 0.05. Ordinary one-way ANOVA with Tukey's test for multiple comparisons or Kruskal-Wallis test with Dunn's post-hoc correction was used for calcium imaging, total neurite outgrowth analyses and multiple electrode analysis with an alpha level of 0.05 (MRD, Mean Rank Difference). All data visualisations were generated in GraphPad Prism 7.0 (GraphPad Software, La Jolla California USA, http://www.graphpad.com/scientific-software/prism/). All data are shown as mean ± standard error of the mean (SEM) to two decimal places where necessary and all error bars represent SEM.

Part II: **Integrated genetic and methylomic analyses identify shared biology between autism and autistic traits**

2.11 Participants

Participants were children from the Accessible Resource for Integrated Epigenetic Studies (ARIES, www.ariesepi- genomics.org.uk) [249], a subset of the Avon Longitudinal Study of Children and Parents (ALSPAC) [250]. Methylation data was only available for participants in the ARIES substudy. ALSPAC is a longitudinal cohort in which the participants were pregnant women in the Avon region in the UK. The initial cohort consists of 14, 541 initial pregnancies and 13,988 children who were alive at the age of 1. In addition, children were enrolled in further phases. Details of the data available can be found on the online data dictionary here: http://www. bristol.ac.uk/alspac/researchers/access/. Written informed consent was obtained from the parent or the guardian of the child and assent was obtained from the child where possible. The study was approved by the ALSPAC Ethics and Law committee, and the Cambridge Human Biology Research Ethics Committee.

The participants of the primary MWAS of SCDC were 701 children who completed the SCDC at age 8, and for whom epigenetic data was available (341 males and 360 females). Of the participants included in the primary MWAS (SCDC), only five participants had an autism diagnosis based on a parental questionnaire at 9.5 years of age. We conducted a secondary MWAS of pragmatic communication in 666 children. Pragmatic communication was measured using the Children's Communication Checklist [251] (CCC) at age 9 (323 males and

340 fe- males). In addition, we conducted a GWAS of SCDC scores in a sample of 5,628 8-year-olds from ALSPAC, details of which are provided below. This sample included participants who were included in the two MWAS (SCDC and CCC).

2.12 Phenotypic measures

The SCDC is a 12-item questionnaire that measures difficulties in verbal and nonverbal communication and social interaction including reciprocal interaction [70] . Scores range from 0 to 24 with high scores reflecting difficulties in social interaction and communication. The SCDC has good psychometric properties—internal consistency of 0.93 and test-retest reliability of 0.81 [70]. We used mother reported SCDC scores on children aged 8. The mean of SCDC scores in our sample was 14.65 (standard deviation = 3.44).

Previous research has demonstrated that the SCDC is stable over time and scores at different ages are genetically correlated [71], [252]. SNP heritability is highest for SCDC scores in childhood (at the age of 91 months) and in later adolescence (17 years) [71], [252]. We focused on SCDC scores at 91 months as the sample size was the largest, has highest genetic correlation with autism [71], and the exposure to environmental factors is limited at 91 months com- pared to other time points.

A second measure that we used in this study is the 53- item parent-completed CCC which measures pragmatic communication [251]. The CCC and subscales have moderate to high twin heritability [253], and moderate SNP heritability (h^2_{SNP} = 0.18) [72]. There is a negative correlation between the CCC and the SCDC [254]. The mean of the CCC in the sample of 666 children was 151.83 (standard deviation = 6.77), with scores ranging from 111 to 162. To make the analysis comparable with the SCDC (which measures difficulties rather than ability), we reverse scored the CCC so that higher scores measure difficulties in pragmatic communication.

The histograms of both the phenotypes in the samples used in the study are provided in Additional file 1: Figure S1. We calculated the phenotypic correlation be- tween the CCC and the SCDC in the samples used in this study using Pearson's correlation.

2.13 Cord blood DNA methylation, quality control and normalization

Array-based cord blood methylation quantification was conducted by ARIES [249]. Briefly, DNA was extracted from cord blood drawn from the umbilical cord upon de- livery. Following extraction, DNA was bisulfite-converted using the Zymo EZ-DNA MethylationTM kit (Zymo, Ir- vine, CA). Then, methylation of over 485,000 CpG sites was measured using the Illumina HumanMethylation450 BeadChip array according to the standard protocol. The arrays were

scanned using an Illumina iScan and initial quality review was assessed using GenomeStudio (version 2011.1).

Methylation assays utilize a pair of probes to detect methylation of cytosine at CpG sites. One is used to detect methylated loci (M) and the other is used to detect unmethylated CpG islands (U). The level of methylation at a locus is then estimated based on the ratio of signals from M to U, called "beta" value. Beta values range from 0 (no cytosine methylation) to 1 (complete cytosine methylation). Sample information and participant demo- graphics are provided in Additional file 1: Table S1.

2.14 QC and normalization

In total, there were 1,127 cord blood samples including technical replicates (i.e., samples that were of poor qual- ity with low detection score and were thus repeated). Of these, 241 were from blood spots and 886 were from white cells. Blood spots were obtained from cord blood and not taken from heel prick. The provided data was quality controlled by ARIES team. The QC procedure employed by the ARIES team includes removing partici- pants who did not pass mother-child genotype-based re- latedness control, participants who were outliers for genetic heterozygosity, genetic ethnicity outliers, samples with low bead numbers, and detection p value >0.05 (probability that the target sequence signal was distin- guishable from the background). This resulted in a total of 914 participants. None of these participants had a sex mismatch, where the genetic sex was different from re- ported sex. We further removed

nine duplicate samples, resulting in 905 participants. Further, 782 of these participants had phenotypic data on the SCDC. Finally, we removed 81 related individuals, resulting in a final sample of 701 participants who had both methylation and phenotypic data.

The data was normalized using functional normalization implemented in the R package meffil (https://github.com/ perishky/meffil) [255]. Functional normalization is a between-array normalization method for the Illumina Infi- nium HumanMethylation450 platform and an extension of quantile normalization. It removes unwanted technical variation. The normalization procedure was performed to the methylated and unmethylated signal intensities, and type I and type II probes separately. For X and Y chromo- somes, males and females were normalized separately using the sex at birth information.

We removed CpG sites whose probe or single-base ex- tension overlaps with a SNP with MAF > 0.01. We further removed cross-reactive probes identified in Chen et al. 2013 [256] as implemented in meffil. In total, 372,662 CpG sites remained after quality control. Cell proportions for CD4 T lymphocytes, CD8 T lymphocytes, B lymphocytes, natural killer cells, monocytes, and granulocytes were esti- mated using the minfi package [257]. These cell types were estimated using post hoc procedures as cell type informa- tion was not collected prior to DNA extraction, further details of which are provided elsewhere [249].

2.15 Methylome-wide association

A methylome-wide association study was run using a two-step regression model (model 1). In the first regression, normalized epigenetic probe betas were regressed against technical covariates (slide, sample type, i.e., white blood cells vs blood spots, and plate and cell counts), using the following model.

$$\beta\text{eta}_{\text{meth}} = \beta_0 + \beta_1 Slide + \beta_2 Sample + \beta_3 Plate + \beta_{4\ldots n}\, Cellcount + \varepsilon$$

The residuals from this regression were further used as corrected methylation values. In the second regression, SCDC (or CCC) scores were regressed against corrected methylation values with sex and the first two genetic principal components as covariates, as provided below:

$$\text{SCDC} = \beta_0 + \beta_1 \varepsilon + \beta_2 Sex + \beta_3 PC1 + \beta_4 PC2 + \zeta$$

We did not correct for age as methylation was measured at birth, and SCDC was measured at 8 years of age for all participants. Here, we were specifically testing if methylation status measured in cord blood was associated with autistic traits or pragmatic language measured at a later age. Given the highly skewed distribution of the SCDC scores, we used a negative binomial regression, using the MASS package in R, which involves by default applying a chi-square test to validate the model (goodness of fit test). We used a Bonferroni-corrected epigenome-wide significant threshold of 1×10^{-7} to identify significant associations. All analyses were conducted in R version 3.2. A Pearson correlation test between both regression coefficients and Z-score (regression Beta divided by the stand- ard error) from both SCDC and CCC MWAS models was performed to assess the epigenetic correlation between both traits.

To evaluate that the results are robust to methodo- logical differences, we used a second model to conduct the methylome-wide association (model 2). This too was conducted using a two-step regression. Here, in the first regression, we used the M value [258] rather than the methylation beta value and regressed it against cell counts generated using "GSE68456" [259] which includes nucleated red blood cells in cord blood. Thus, our first regression in model 2 is of the form:

$$\text{M–value}_{\text{meth}} = \beta_0 + \beta_1 Slide + \beta_2 Sample + \beta_3 Plate + \beta_{4\ldots n}\, Cellcount + \varepsilon$$

Where the cell counts include CD4 T lymphocytes, CD8 T lymphocytes, B lymphocytes, natural killer cells, monocytes, granulocyte, and nucleated RBCs. The resid- uals from this regression were regressed against SCDC scores using Eq. 2 outlined above. We then evaluated if the Z-scores varied substantially between the two models using a Pearson correlation test. Throughout the manu- script, we report the results from the first model because (1) beta values are easier to interpret than M-values, and (2) only 241 participants included in the MWAS had nu- cleated RBC count different from 0.

To identify gene sets and networks that were differentially methylated in the SCDC MWAS, we used mGSEA [260] and used Gene Ontology-based gene sets.

In order to interpret results from the MWAS, we de- signed a multi-step enrichment strategy including (1) a same-sample, same-tissue overlap and correlation ana- lyses between the SCDC and the CCC; (2) a cross-tissue overlap analysis between the SCDC MWAS and MWAS of autism in peripheral blood and post-mortem brain tissue; (3) enrichment for autism transcriptionally dysregulated genes; and (4) enrichment of CpG-associated mQTLs in autism and SCDC GWAS. A summary of the study design is provided in Fig. 1.

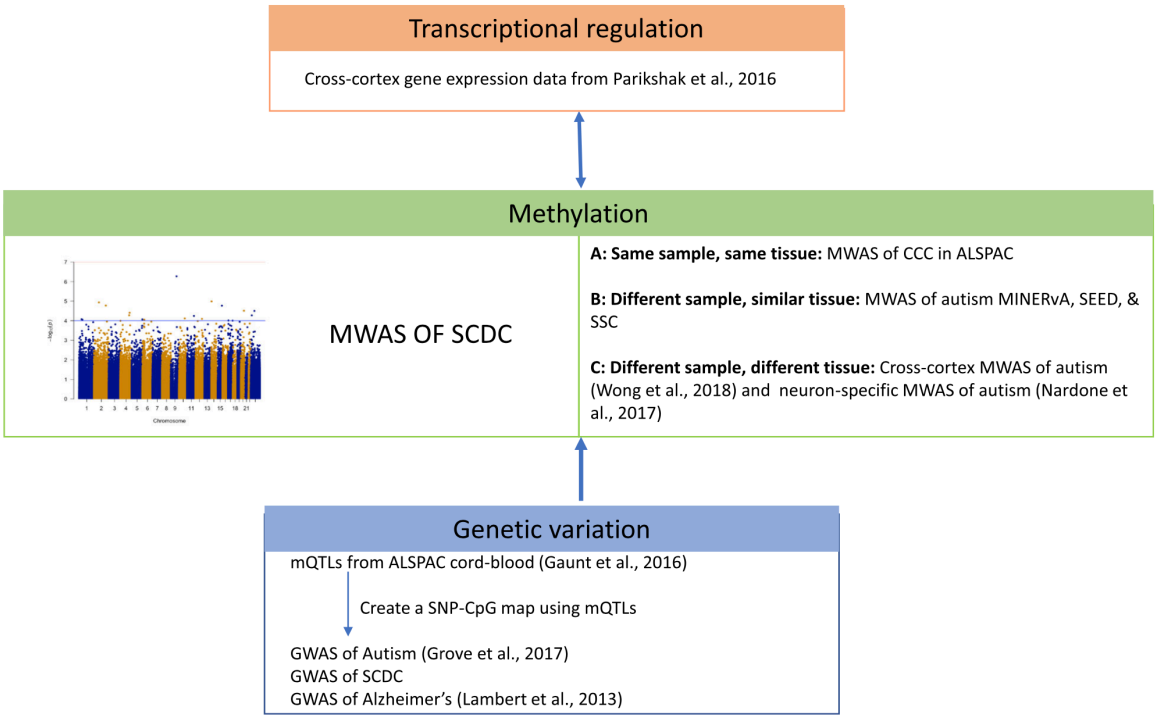


Figure 2-9: Schematic diagram of study design

2.16 Peripheral tissue (blood and bloodspot) overlap analysis

We had access to summary MWAS statistics from three peripheral tissue datasets described in detail elsewhere (SEED [65], Simons Simplex Collection (SSC) [65], and MINERvA [63]). For all overlap analyses, we conducted two statistical tests. In the first, we tested if all

nominally significant CpGs (p value < 0.01) in the three-peripheral tissue MWAS (SEED, SSC, and MINERvA) have a shift toward lower p values in the SCDC MWAS (one-sided Wilcoxon rank-sum test). This tests a larger number of CpGs and is consistent with the idea that each individual CpG contributes minimally to the phenotype suggesting a polymethylomic (or, polyepigenetic) architecture simi- lar to a polygenic architecture of complex traits. In addition, this does not test effect direction as effect direction may vary based on number of factors including tissue source. In the second analysis, we investigated effect direction concordance for CpGs with p value < 1×10^{-4} in either of the two MWAS being tested, thus, conducting 12 binomial sign tests tests in total. This restricts the analyses to a relatively small number of CpGs.

2.17 Post-mortem brain tissue overlap analysis

We had access to summary statistics from two post- mortem brain tissue datasets to investigate overlap with the SCDC MWAS. MWAS in both these datasets were conducted using the Illumina HumanMethylation450 BeadChip making the MWAS comparable to the SCDC MWAS. We used a recent MWAS conducted using tis- sue from 38 idiopathic autistic individuals and 38 controls [69]. Further details are provided elsewhere [69]. To investigate if there was an overlap between the SCDC MWAS and a neuron-specific MWAS in post-mortem autism brains, we used summary MWAS data from FACS-sorted neurons in brain samples from 15 autistic individuals and 16 typical controls. Further details are provided elsewhere

[66]. We did not have access to social communication data in the two post-mortem samples, making this comparison impossible.

For both datasets, our analysis was similar to the ana- lysis of peripheral tissue MWAS. We investigated effect direction concordance between the two post-mortem brain autism MWAS and the SCDC MWAS for all CpGs with p value $< 1 \times 10^{-4}$ in the post-mortem brain MWAS (binomial sign test). Additionally, we investigated if CpGs with p value < 0.01 in either of the two post-mortem brain MWAS had a significant shift toward lower p values in the SCDC MWAS (one-sided Wil-coxon rank-sum test).

2.18 Enrichment with autism-associated transcriptionally dysregulated genes

For enrichment analyses with transcriptionally dysregu- lated gene expression data, we used an RNA-sequencing study of 167 post-mortem cortical samples with n = 85 with a diagnosis of autism and n = 82 from nonpsychiat- ric controls. Samples were from BA9 (prefrontal cortex), or BA41/42 (temporal cortex) [261]. Significantly dysregulated genes had a Benjamini-Hochberg adjusted FDR < 0.05. We conducted enrichment analyses using a one-sided Wilcoxon rank-sum test. We first mapped the CpGs to genes using the CpG to gene annotation for the Illumina 450k methylation array using the IlluminaHu- manMethylation450k.db package in R (http://www.biocon- ductor.org/packages/release/data/annotation/html/Illumi- naHumanMethylation450kprobe.html). We restricted our analysis only to CpGs that

were mapped onto the genes tested for differential expression in the post-mortem brain dataset [261]. We then compared the distribution of the CDC p values for CpGs mapped to significantly differen- tially dysregulated genes vs the other genes.

2.19 Enrichment of CpG-associated mQTLs in GWAS of autism and SCDC

We investigated if mQTLs of CpGs below four p value thresholds in the SCDC MWAS (P_{SCDC}) had lower p values compared to other mQTLs in the GWAS (P_{GWAS}) of (1) autism, (2) SCDC, and (3) as a negative control, Alzheimer's. We hypothesized that the mQTLs of CpGs below P_{SCDC} will have significantly lower P_{GWAS} in comparison with remaining mQTLS. To map CpGs to mQTLS, we used mQTL maps generated by the ARIES cohort in cord blood (n = 771), restricting our analysis to only significant mQTLS identified after FDR correction (p value <0.05 after FDR correction) [262]. This cohort overlapped with the sample in which the MWAS was con- ducted. All mQTLs had a minor allele frequency > 1%. For each CpG-mQTL pair, we restricted our analysis to only those CpG-mQTL pairs investigated in both the SCDC MWAS and the GWAS of interest. In other words, the CpGs must have been investigated in the SCDC MWAS and the paired mQTL of the CpG must have been investi- gated in the GWAS of interest. As none of the CpGs meet the strict p value threshold, we had to use several thresh- olds at different levels of stringency. To control the signal- to-noise ratio in the context of an MWAS, we considered four empirical P_{SCDC} thresholds: 0.05, 0.01, 0.005, and 0.001.

Enrichment was conducted using permutation test- ing, where we defined 10,000 null sets. We identified three potential factors that may influence this analysis: (1) the linkage disequilibrium (LD) structure of mQTLs, (2) the number of mQTLs mapped onto a CpG, and (3) the num- ber of CpGs a single mQTL is mapped onto. To address LD, first, we clumped the list of mQTLs using an r^2 of 0.6 and distance of 1000 kb, to ensure that linkage disequilibrium among these mQTLs does not confound the out- come. In this clumped list of mQTLs, the majority were mapped to only one mQTL. Second, to account for the number of mQTLs mapped onto CpGs, we binned the CpGs into six groups based on the number of SNPs they mapped onto (1–5, 6–10, 11–15, 15–20, 20–25, and above 25), and conducted enrichment analysis so that every mQTL in the null set matched the original mQTL based on CpG bins. Third, one single mQTL may map onto multiple CpGs, resulting in non-unique CpG-mQTL pairs with P_{SCDC} < threshold, and P_{SCDC} > threshold. We retained unique CpG-mQTL pairs in each list before conducting permutation-based enrichment analysis. Finally, to account for multiple testing, as we tested across four non-independent p value thresholds, the empirical p values were corrected for the four tests using Benjamini-Hochberg FDR correction. Empirical p values were significant at FDR < 0.05.

We examined the results identified in the Autism GWAS using a GWAS of log-transformed SCDC scores in ALSPAC (details below). As a negative control, we used GWAS data for Alzheimer's (phase I), downloaded from IGAP (http://web.pasteur-lille.fr/en/recherche/u744/igap/ igap_download.php) [263], and tested for enrichment using an identical procedure as mentioned above. The Alzhei- mer's GWAS (phase I, for which

genome-wide summary data is available) consists of 17,008 cases and 37,154 con- trols, and identified 14 significant GWAS loci. While both autism and Alzheimer's are neuropsychiatric conditions, the genetic correlation between the two conditions is non- significant (r_g = 0.04 ± 0.10; p value = 0.102), suggesting minimal shared genetics. The number of cases and controls used in the two studies (phase 1 for the Alzheimer's GWAS) are comparable, providing approximately similar statistical power (mean chi-square: Alzheimer's = 1.114, autism = 1.2). Further, they are distinct in that autism is a neurodevelopmental condition diagnosable at childhood, while Alzheimer's is largely diagnosed in individuals who are 65 or older.

2.20 GWAS of SCDC scores

We conducted a log transformed GWAS of SCDC scores at age 8 in the ALSPAC data. Note that log- transformed phenotype models are computationally more efficient for high-dimensional GWAS data than negative binomial models used in the MWAS. Further, we identified a high correlation between the log- transformed SCDC MWAS and the negative binomial SCDC MWAS (r_{Beta} = 0.98, P 2.2×10^{-16}; $r_{Zscores}$ = 0.99, P 2.2×10^{-16}), suggesting that the results are almost identical between the two statistical models. Participants were genotyped using the Illumina® HumanHap550 quad chip by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. We restricted our analysis only to individuals of Euro- pean descent. This was identified using multidimensional scaling analysis and compared with Hapmap II (release 22) [264]. We excluded individuals with sex mis- matches, high

missingness (> 3%), and disproportionate heterozygosity, and if cryptic relatedness, identified using identity by descent, was greater than 0.1. We removed SNPs with greater than 5% missingness, those that violated Hardy-Weinberg equilibrium (p value $< 1 \times 10^{-6}$), and those with a minor-allele frequency less than 1%. This resulted in a total of 526,688 genotyped SNPs. Haplotypes were estimated using data from mothers and children using ShapeIT (v2.r644) [265]. Imputation was performed using Impute2 V2.2.2 against the 1000 genomes reference panel (Phase 1, Version 3) [266]. Imputed SNPs were excluded from all further analyses if they had a minor allele frequency < 1% and info < 0.8.

After quality control, there were 8,282,911 genotyped and imputed SNPs that were included in subsequent analyses. GWAS analysis was conducted for mother- reported SCDC scores at age 8 that was log-transformed given the highly skewed distribution. Linear regression was conducted in Plink v1.9 [266] that converted allele dosages into hard calls. We included the first two ancestry principal components and sex as covariates in the regression model. The first two ancestry principal com- ponents were calculated using Plink 1.9 in unrelated individuals, using SNPs with MAF > 5% that were pruned for LD ($r^2 < 0.1$).

As reported previously [38], [71], [72], the SNP heritability as quantified using LDSC [267], [268] was $h^2 = 0.12 \pm 0.05$. The LDSR intercept (0.99) suggested that there was no inflation in GWAS estimates due to population stratifi- cation. The λ_{GC} was 1.013. We replicated the previously identified genetic correlation (constrained intercept) [38] with autism using our

SCDC GWAS (PGC-autism: r_g = 0.46 ± 0.20, p value = 0.019; iPSYCH-autism: r_g = 0.45 ± 0.18, p value = 0.01).

2.21 Data, software, and script availability:

MWAS summary statistics:

- The summary statistics for the MWAS (SCDC and CCC) can be downloaded from here:

  https://www.dropbox.com/sh/8za5xspmbjydpst/AAA_ZGmMLOE8Ql7egi5Mcu8Ha?dl=0. These summary statistics are also provided as RData files with this manuscript.

- Summary statistics for the SEED and the SSC MWAS can be obtained from here: https://molecularautism.biomedcentral.com/articles/10. 1186/s13229-018-0224-6.
- Summary statistics for the MINERvA cohort can be obtained by contacting Jonas Bybjerg-Grauholm.

GWAS summary statistics:

- The summary statistics for the autism GWAS (iPSYCH) can be downloaded from http://www. med.unc.edu/pgc/results-and-downloads (iPSYCH-PGC GWAS-2017).
- The Alzheimer's GWAS can be downloaded from http://web.pasteur-lille.fr/en/recherche/ u744/igap/igap_download.php.

- The summary statistics for the SCDC GWAS can be obtained from https://www.dropbox.com/sh/8za5xspmbjydpst/AAA_ZGmMLOE8Ql7egi5Mcu8Ha?dl=0.
- Scripts for running the two regression models for the MWAS and running the enrichment analyses with the mQTL data are available here: https://github.com/autism-research-centre/MWAS_ autistictraits
- mQTL data used in this (cord blood) is a part of the ARIES cohort, and can be downloaded here: http://www.mqtldb.org/
- We used the following software/packages: Plink (http://zzz.bwh.harvard.edu/plink/); IlluminaHumanMethylation450k.db(http://www.bioconductor.org/packages/release/data/annotation/ html/IlluminaHumanMethylation450kprobe.html); MASS (https://cran.r-project.org/web/packages/ MASS/index.html); LDSC (https://github.com/bulik/ ldsc/wiki/Heritability-and-Genetic-Correlation).

Chapter 3

Generation of Cellular Models using Gene Editing tools

Summary

Heterozygous CNVs in *NRXN1* gene have been implicated in autism, but such mutations show incomplete penetrance. This chapter describes the generation of the cell lines included in this thesis, it will be divided in two-part, isogenic iPSC lines and iPSC lines from autistic individuals. The first part demonstrates the strategy used to generate the isogenic lines using CRISPR-Cas9 to target the first shared exon of the two *NRXN*1 isoforms. And the second part illustrates the targeting of the patient lines with OPTi-OX system to allow the forward programming of the iPSC into induced neurons. To ensure the validity of both gene editing methods in the generation of the cell lines, genotyping and neural induction and neuronal characterization was carried out.

3 Generation of Cellular Models using Gene Editing tools

3.1 Introduction

The neurexin family is a well-studied group of synaptic proteins; they play an important role in synaptic development as cell adhesion molecules and receptors localised at the pre-synaptic terminal [105]. Neurexins are type-1 membrane proteins produces by three genes, namely *NRXN*1, *NRXN*2 and *NRXN*3. All of them localise at the presynaptic terminals [105]. *NRXN*1 is 1.12 Mb long and contains 24 exons, making it one of the largest gene within the human genome [269]. It gives rise to two major isoforms; transcribed from two distinct promoters: (1) *NRXN*1α, the largest isoform and is usually more frequently affected by mutations, (2) *NRXN*1ß, transcribed by a promoter located downstream of exon17. Exon 19 is the first shared exon by both isoforms and deletions in this exon can potentially affect both isoforms [166], [195]. To add to its complexity *NRXN*1 is extensively spliced; 5 sites within the *NRXN*1α and 2 sites are shared between both isoforms (α and ß)[105]

*NRXN*1 has been implicated in autism spectrum conditions and is listed in Simon's Foundation for Autism Research Initiative (SFARI), a database that have curated genes have been associated with autism, with variable evidence [270].

Many studies reported heterozygous CNVs in *NRXN*1 that resulted in the deletion or loss of function of the gene. However, many autism related mutations show incomplete penetrance, this is intriguing because it makes it harder to dissect the role of the mutation in the aetiology of the condition from the patient's own genetic background [195].

To investigate this, we introduced indel mutations in a well-established iPSC line (BOB-*NGN*2) using CRISPR-Cas9 [230]. And we also complemented that by using two iPSC lines derived from autistic individuals with *NRXN*1 mutations.

3.1.1 *NRXN*1-Isogenic (mutant) lines

The CRISPR_Cas9 system is a genome editing tool that relies on a nuclease (Cas9) to target a genomic locus with high specificity and efficiency to stimulate a double strand break (DSB). This is followed by one of two major pathways for DNA damage repair: non-homologous end-joining (NHEJ) or the homology-directed repair (HDR). The latter way resulting in insertion/deletion (indel) mutations. If the resulting indel mutations lead to frameshift mutations and premature stop codons, this generates gene knockouts. HDR on the other hand, is used to generate specific modifications at the target region and hence is used for the introduction of single nucleotide mutations. The Cas9 nuclease is guided by the synthetic guide RNA, that is usually designed to complement the target locus where the DSB is supposed to happen [215]. In this study, the guide RNA was designed to target the first shared exon between the

two NRXN1 major isoforms (Figure 3.1) to result in a frameshift mutation leading to a loss of function.

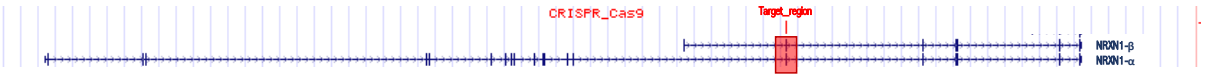


Figure 3-1: NRXN1 structure, showing the CRISPR-Cas9 target region on the first shared exon of NRXN1.

3.1.2 Patient lines

In this thesis, two patient lines were used: 092_NXF; a female iPSC line derived from an autistic individual and 211_NXM a male iPSC line derived from another autistic individual. Both patients have *NRXN*1 mutations, 092_NXF has a *De novo* deletion of ~200kb; 2p16.3 (50,806,991 – 51,013,685) and 211_NXM has a missense mutation (G→C) on chromosome 2 (50724586) as shown in Figure 3-2. These lines were provided by Prof Jack Price at King's College, London. Gene correction could potentially be done to produce an ideal control, but due to the highly complex regulation of NRXN1 and the large size of the deletions associated with it, creating an isogenic control through gene correction is technically very challenging.

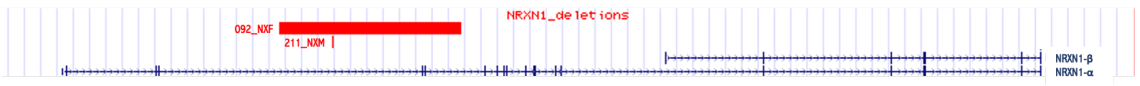


Figure 3-2: NRXN1 structure, showing the positions of NRXN1 deletions and mutations in the Patient iPSC lines; the 092_NXF patient line 200kb deletion and 211_NXM point mutation on the NRXN1 α isoform

The OPTi-OX system (Optimised inducible overexpression) is a platform that allows the forward programming of iPSCs into specific cell types. The platform relies on the dual GSH targeting of the Tet-ON system components for homogenous and controllable expression of inducible transgenes. The Tet-ON system consists of two components: 1) a reverse tetracycline transactivator (*rtTA*) which is responsive to doxycycline (DOX) and driven by a CAG promoter and 2) a transgene (*NGN2*) driven by an *rtTA* inducible promoter: Tet-responsive element (TRE). The components are then targeted into two GSH loci to ensure that the overexpression of the transgene will not be affected by promoter interference and reduces the size of the cassette allowing the insertion of larger transgenes if needs be [229], [230].

We relied on the OPTi-OX system to allow forward programming of 092_NXF and 211_NXM into mature induced neurons. The pR26_CAG-*rtTA* was constructed by introducing the rtTA coding sequence into the BamHI/MluI sites of pR26_CAG-EGFP [230]. The pAAV_TRE-*NGN2* was constructed by cloning the *NGN2* coding sequence into the SpeI/EcoRI sites of pAAV_TRE-EGFP [229], [240]. The *rtTA* is constitutively expressed in the presence of DOX, and results in the activation of the TRE promoter and driving the expression of *NGN2*. *NGN2* is a *bHLH* transcription factor, used for the induction of neuronal differentiation from iPSCs.

3.2 Results

3.2.1 CRISPR-Cas9-induced mutations in Exon19 produced 4 different in-del mutations

3.2.1.1 The generation of 4 clones with 4 different indel mutations were validated by Sanger sequencing

The gRNA (AEG6) was used to produce indel mutations in *NRXN*1 Exon19, it was designed to target (Chr2: 50,236,786-50,236,973). The resulting clones were genotyped using the PCR primers in Table 2-2 in Chapter 2 and sent for Sanger sequencing (Figure-3). Sanger sequencing is first generation sequence technique, typically used in clinical genomics to validate the presence of mutations in a specific region of interest [271]. Technically, this method can be used to cover any desired region, however, due to its hight cost it is usually a preferred method of screening specific regions and/or genes [272]. Although, Sanger sequencing is a first genetation DNA sequencing method, it has some advantages over next-generation sequencing (NGS) techniques, for instance, some regions can be poorly captured by targeted NGS due to high GC content. In this case, Sanger sequencing can be used to patch the poorly covered regions by NGS [272]. The generated CRISPR-edited clones included 2 heterozygous*:* A7 (wild type / 1bp insertion), E9 (wild type / 115 bp deletion) and 2 homozygous: G4 (homozygous 1b insertion) and A2 (2 bp deletion / 4bp deletion) lines (Figure 3-3B).

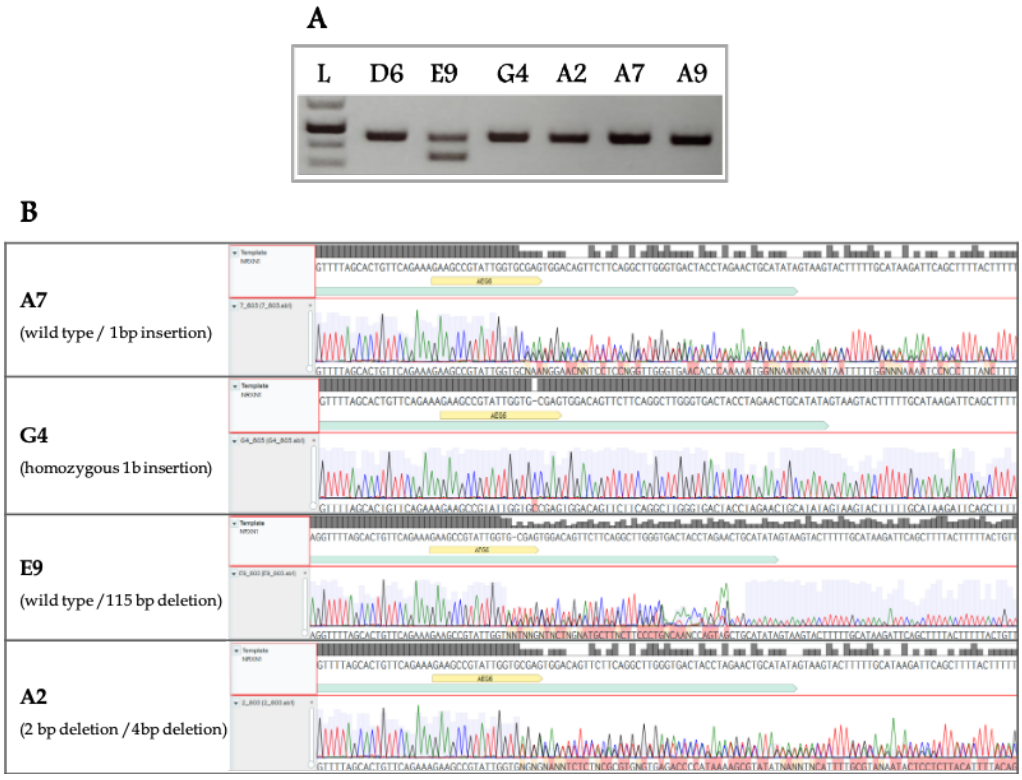


Figure 3-3: Genotyping results for the CRISPR-Cas9 induced NRXN1 mutations. A; Gel electrophoresis of the PCR products, B; Sanger sequencing chromatograms of the 4 clones (A7, G4, E9 and A2).

3.2.2 Characterization of *NRXN*1-isogenic (mutant) iNeurons

3.2.2.1 Induction of NRXN1-Isogenic lines into neurons and time-course expression demonstrated the initiation of pan and glutamatergic neuronal expression and dysregulation of pluripotency markers.

The iNeurons (iNs) produced by the clones generated by the gene editing of our well-established *NGN2* OPTi-OX control iPSC line (BOB-NGN2) was characterised to confirm that the gene editing has not affected reproducibility between the lines. The *NRXN*1-Isogenic lines were induced by treating them with DOX in chemically defined neural induction medium. Quantification of gene expression was done using Q-PCR,

gene expression was normalised against 18S housekeeping gene, and the log fold change was normalised to BOB Dayo. Gene expression of *NGN2* was quantified across 4 time points, dayo before induction and days 7, 14 and 21 post induction. As expected, the successful induction of *NGN2* expression peaked at day 7 (Figure 3-4).

The induction of *NGN2* showed a rapid drop in pluripotency markers such as *OCT4* and *NANOG*, 1-week post induction (Figure 3-4E, F and G). The iNs also showed high expression of a pan-neuronal marker, namely MAP2 and glutamatergic neuronal genes such as *GRIA4* and *vGLUT1 (Figure 3-4B, C and D)*. This demonstrated that the induction of *NRXN1*-Isogenic lines results in the generation of excitatory cortical neurons. The concordance in expression patterns between the mutant and the wildtype iNs confirms the validity of the system to study the effect of the induced mutations in the *NRXN1* genes.

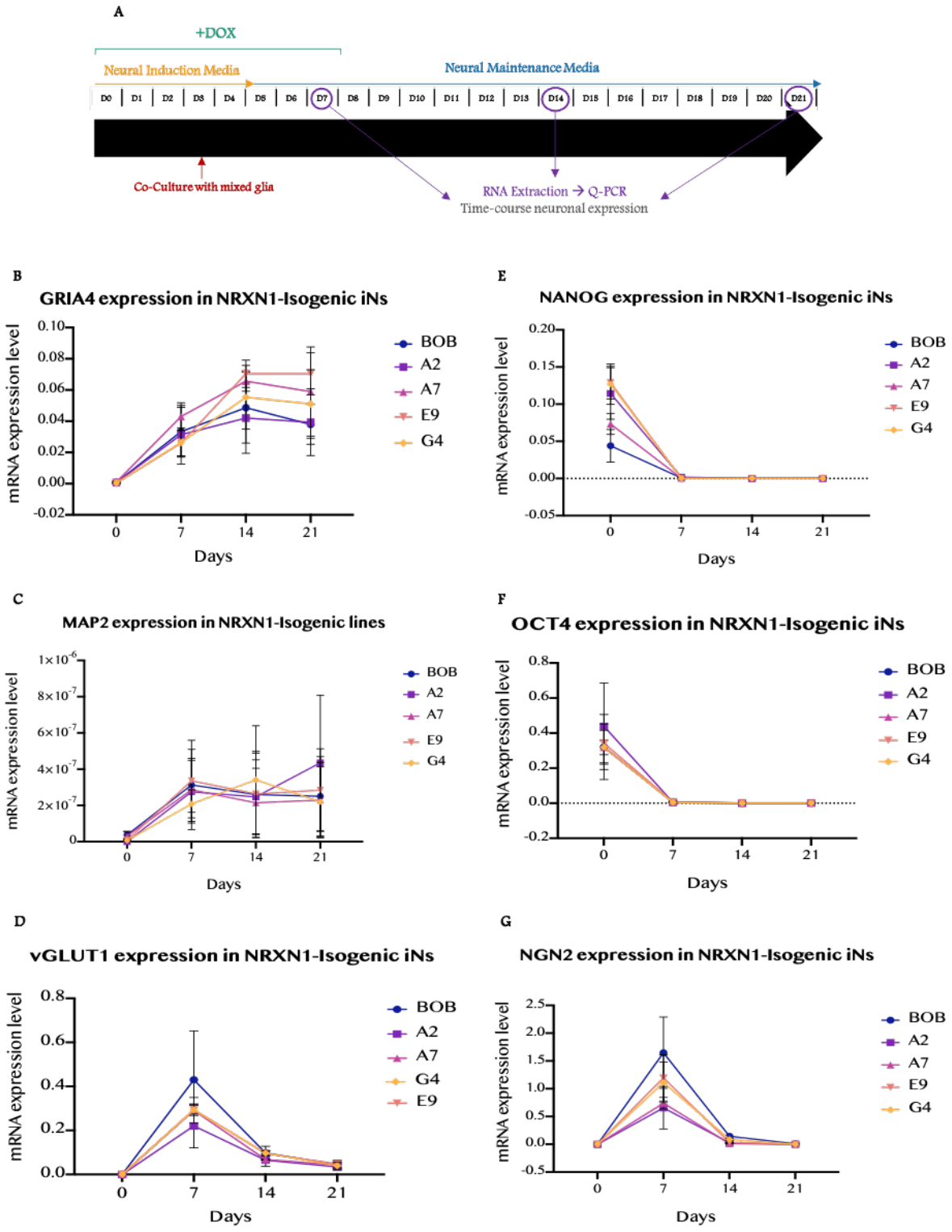


Figure 3-4: Time-course expression pattern of neuronal and pluripotency genes across 4 timepoints in isogenic lines, A) Schematic representation of our time-course neural induction, Time-course quantification of mRNA levels of GRIA4 (B), MAP2 (C), vGLUT1 (D), NANOG (E), OCT4 (F) and NGN2 (G)

3.2.3 Introducing the OPTi-OX system in Autism patient-derived iPSC

3.2.3.1 The optimised inducible NGN2 overexpression was targeted in dual GSH loci; ROSA26 and AAVS1 in two patient iPSC lines. Three clones were generated from each line and validated using PCR.

To validate the successful targeting of hROSA26 and AAVS1 with the OPTi-OX components, PCR primers were designed to genotype the selected colonies (Figure 3-5A). The primers for the ROSA26_CAG-*rtTA* and AAVS1_TRE-*NGN2* are listed in Table 2-3 and Table 2-4, respectively (Chapter 2). To identify the presence of the wild type allele, PCR primers were designed to bind to the genomic DNA outside the target region corresponding to the homology arms, therefore the successful integration of the gene of interest results in a loss of this PCR product. Another two primer pairs were designed to flank the 5' of the transgene and the 3' of the transgene (*rtTA* and *NGN2*).

Successful integration of the transgene would result in the presence of a PCR product. The last primer pair spans part of the homology arm and the plasmid's backbone junction; presence of a product indicates that the integration of the vector was non-specific (Figure 3-5A). From each line, clones were selected for genotyping, and the PCR product were visualised on a 1% agarose gel; the results are shown in Figure 3-5 B and C.

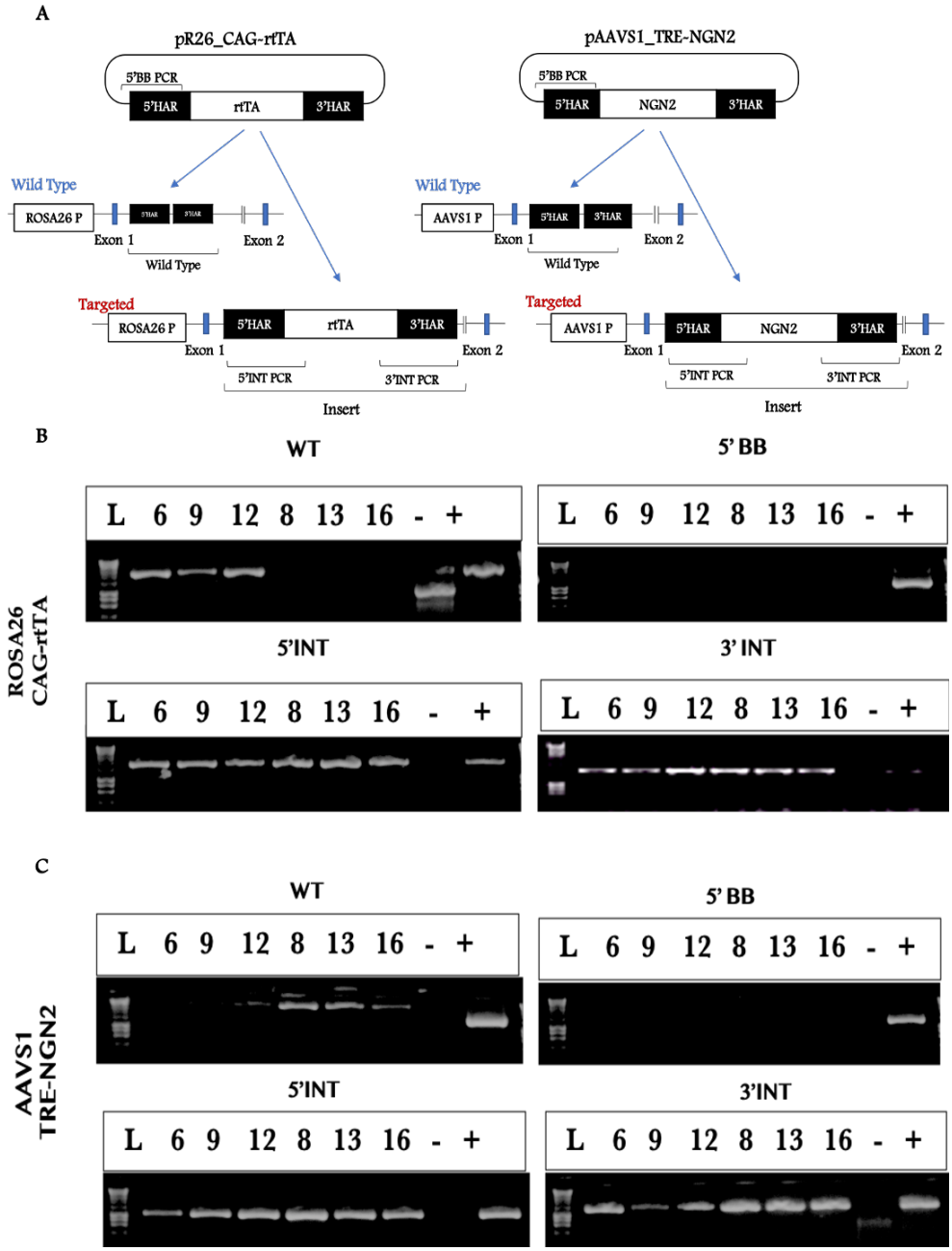


Figure 3-5: Inducible NGN2 expression, A; Genotyping strategy by PCR to vaildate the taregting of the pR26_CAG-rtTA and pAAVS1 in the patient lines, PCR primer pairs are designed to target the wild type by spanning the 5' and 3' end of the homology arms of the GSH (ROSA26 and AAVS1, respectively), the presence of the PCR product in this case indicates the presence of the Wild type allele, another primer pair was designed to spanning the bakcbone of the targeting vector and the 5'homology arm; the presence of a band at the correct size indicated the non-specific integration in the genome. Two other pairs of primers (5'INT and 3'INT) were designed to amplify the 5'HAR-transgene and the 3'HAR-transgene, respecively. B; Electrophoresis results for the genotyping of ROSA26 CAG-rtTA, WT: wild type, 5'BB: 5' Backbone, 5'INT and 3'INT are the 5' and 3' ends of the trangene (rtTA).C; Electrophoresis results for the genotyping of AAVS1 TRE-NGN2, WT: wild type, 5'BB: 5' Backbone, 5'INT and 3'INT are the 5' and 3' ends

of the trangene (NGN2). L: Ladder, 6-9-12: 092_NXF clones, 8-13-16: 211_NXM clones, - : negative control and +: positive control.

The pR26_CAG-*rtTA* and pAAVS1_TRE-*NGN2* vector were successfully integrated in the ROSA26 and the AAVS1 GSH loci by the presence of the bands at the correct size in both 5'INT and 3'INT. The presence of the WT allele in clones 092_NXF- 6, 9 and 12 indicates that these clones are heterozygous for the CAG-*rtTA* but homozygous for the TRE-*NGN2* (Figure 3-5). The 211_NXM 8, 13 and 16 clones are homozygous for both transgenes and all the clones showed no product for the 5'BB which is sign of no off-target effects. Over all the clones chosen all contain both constructs and are ready for forward programming (Table 3-1).

Table 3-1: Summary of genotyping results

			092NXF			211NXM		
		Clone	6	9	12	8	13	16
ROSA26	**CAG-rtTA**	Homozygous	-	-	-	x	x	x
		Heterozygous	x	x	x	-	-	-
AAVS1	**TRE-NGN2**	Homozygous	x	x	x	x	x	x
		Heterozygous	-	-	-	-	-	-

3.2.4 Forward programming of OPTi-OX iPSCs into iNs

3.2.4.1 Induction of NGN2 expression using DOX resulted in the differentiation of iPSC cells into iNs, showing a decrease in pluripotent markers and an increase in neuronal marker within 1 week of induction, as well as clear morphological changes using bright field images.

Clones 6-9 and 12 of 092_NXF and 8-13 and 16 of 211_NXM were seeded for neural induction. To confirm that the iPSCs are responsive to DOX-induced forward programming, the cells were grown in the presence and absence of DOX until D15. Bright field images and RNA were extracted from iNs 4 days post induction to assess expression of *NGN2* expression as well as pluripotency markers. RNA was also collected at day 14 in order to assess expression of neuronal markers, using Q-PCR. The 092NXF-NGN2 and 211_NXM-NGN2 lines showed neuronal projection in the presence of DOX at Day 4 while a heterogenous culture of multiple cell types presented in the absence of DOX (Figure 3-6B and Figure 3-7B). This strongly indicated the successful integration and response of the OPTi-OX system in the presence of DOX. To further demonstrate the efficiency of this system in developing mature neurons, the expression of pluripotency and neuronal markers was quantified at days 4 and 14 in DOX-treated vs non-treated cells by Q-PCR. In Day4 samples, the expression of the pluripotency markers, *OCT4* and *NANOG*, were significantly higher in the non-DOX treated cells while the expression of *NGN2* was significantly lower compared to DOX-treated cells. Moreover, the expression of vGLUT2, MAP2 and *GRIA4*, 3 pan neuronal genes was significantly higher in DOX-treated iPSC compared to the non-treated cells at Day 14 (Figure 3-6, and Figure 2-7).

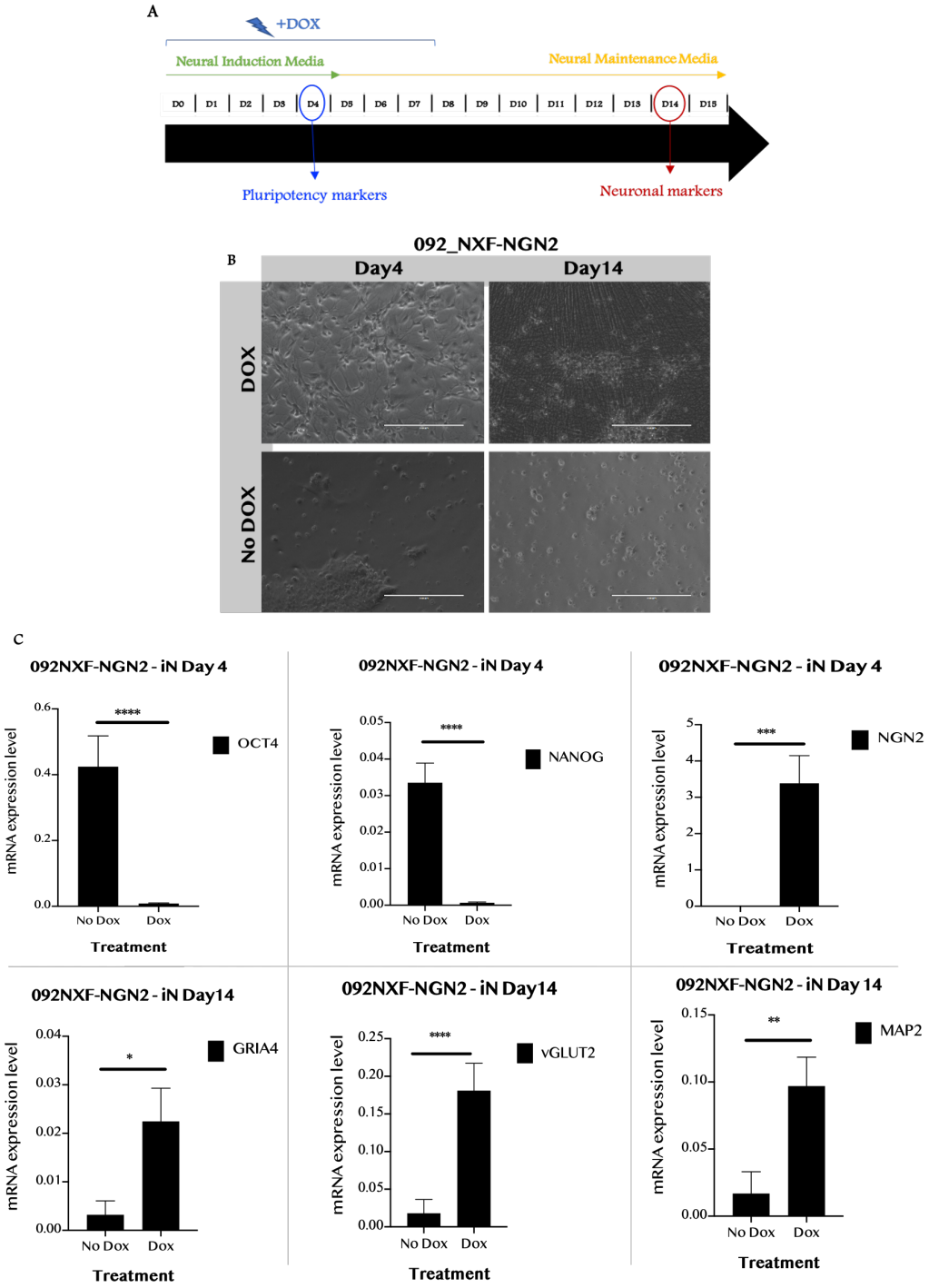


Figure 3-6: Experimental approach for validation of the iNs derived from OPTi-OC iPSCs in the presence of DOX in the 092_NXF-NGN2 patient line: A) Schematic representation of the neural differentiation timeline, B) Brightfield imaged demonstrating the morphological changes across time and treatment in the 092_NXF-NGN2. C) Quantification of NGN2 expression and pluripotency markers (OCT4 and NANOG) at Day4, and pan-neuronal markers (vGLUT2, MAP2, GRIA4) at Day15 in the presence and absence of DOX at Day14, absolute value of mRNA expression is presented on the y axis and Bonferroni

*correction was used for multiple comparisons, 95%confidence interval, stars indicate levels of statistical significance, adjusted P value: ****<0.0001, *** 0.0001, **0.002, *0.02.*

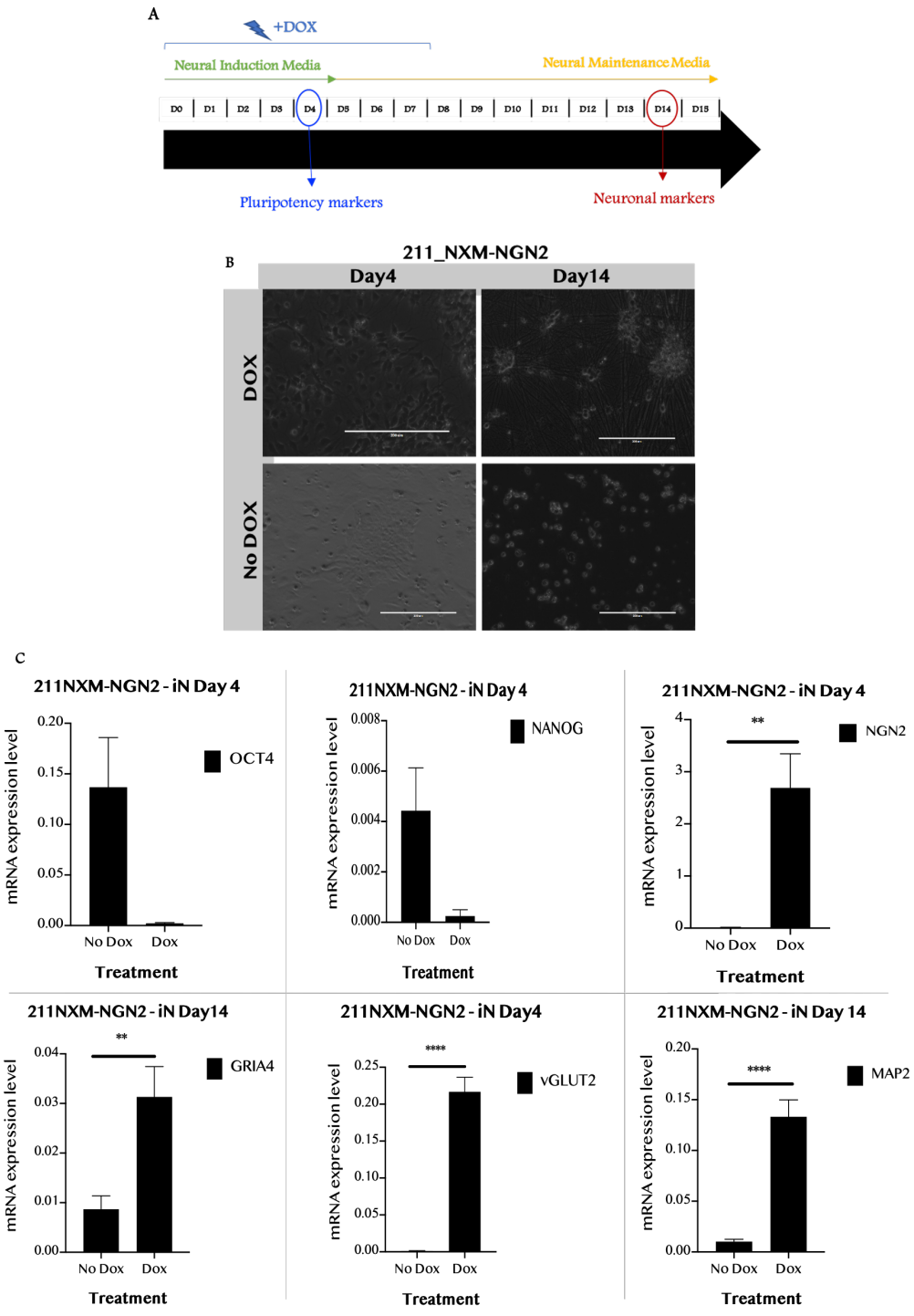


Figure 3-7: Experimental approach for validation of the iNs derived from OPTi-OC iPSCs in the presence of DOX in 211_NXM-NGN2 patient line: A) Schematic representation of the neural differentiation timeline, B) Brightfield imaged demonstrating the morphological changes across time and treatment in the 092_NXF-NGN2. C) Quantification of NGN2 expression and pluripotency markers (OCT4 and

*NANOG) at Day4, and pan-neuronal markers (vGLUT2, MAP2, GRIA4) at Day15 in in 211_NXM-NGN2 in the presence and absence of DOX at Day14, absolute value of mRNA expression is presented on the y axis and Bonferroni correction was used for multiple comparisons, 95%confidence interval, stars indicate levels of statistical significance, adjusted P value: ****<0.0001, *** 0.0001,**0.002, *0.02.*

3.2.5 Characterization of OPTi-OX patient-derived iNeurons

3.2.5.1 Induction of patient lines into neurons and time-course expression demonstrated the initiation of pan and glutamatergic neuronal expression and dysregulation of pluripotency markers.

To test the efficiency of the *NGN2* OPTi-OX system in the homogenous and controlled expression of *NGN2* in the targeted patient cell lines, iPSC cells were cultured with neuronal induction medium containing DOX and the expression of *NGN2* measured before induction (Day 0) and at days 7, 14 and 21 post induction. In addition, a well-established *NGN2* OPTi-OX iPSC line (BOB-*NGN2*) was used as a reference line. The iNs showed a peak in *NGN2* expression 7 days post induction, as expected. The iNs were also tested for pluripotency markers (*OCT4* and *NANOG*); a sudden drop in their expression was observed between day 0 and 4. The neuronal genes on the other hand showed strong expression starting day 7 (Figure 3-8). The neuronal markers used were MAP2, *vGLUT1* and *GRIA4* which demonstrate that the iNs generated have an excitatory cortical identity as shown in Figure 3-8.

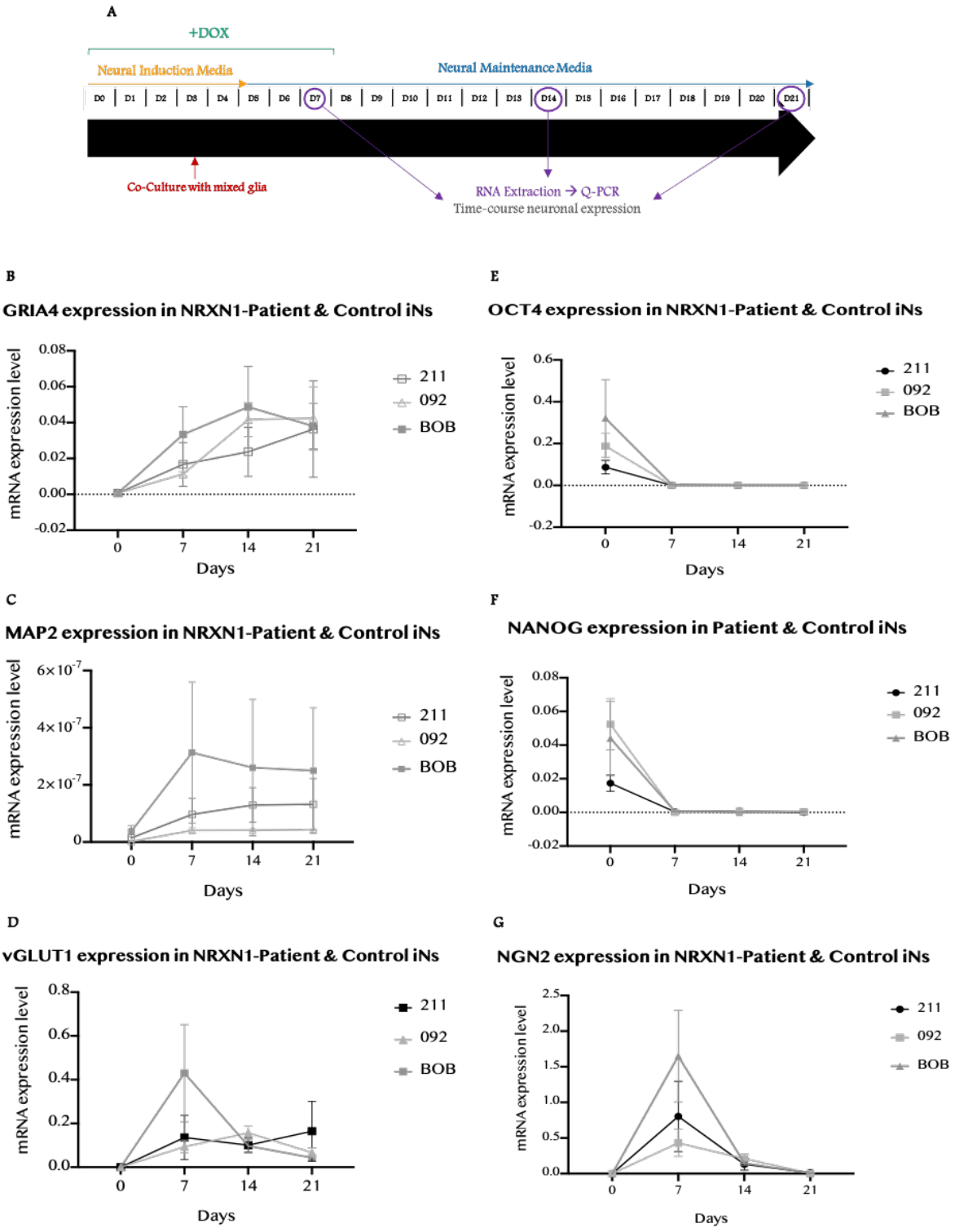


Figure 3-8: Time-course expression pattern of neuronal and pluripotency genes across 4 timepoints in patient and control lines, A) Schematic representation of our time-course neural induction, B, C and D)Time-course expression of glutamatergic neuronal markers (vG vGLUT2 and GRIA4), pan neuronal marker (MAP2) and E and F) pluripotency genes (OCT4 and NANOG), and NGN2 (G) shown in a time-course graph.

3.3 Discussion

NRXN1 heterozygous CNVs reported implicated in autism and other neuropsychiatric condition show variable penetrance and the mechanism underlying their clinical representation remain ambiguous. This suggest that the patient's genetic background is contributing the clinical outcome of those mutations. In this chapter, gene editing tools were used to generate two cellular models to examine functional significance of *NRXN1* mutations. Firstly, CRISPR-Cas9 NHEJ was used to introduce indel mutations in the *NRXN1* gene within a well characterised iPSC line already containing a *NGN2*-inducible cassette, and secondly, a *NGN2*-inducible cellular model was created using patient-derived iPSCs.

The first part of the chapter, the BOB-*NGN2* cell line, a well characterised control cell line that contains a TET-ON NGN2 cassette in a GSH site was used to generate an isogenic line with *NRXN*1 mutations [230]. Given the complexity of the gene, careful consideration was taken into account in the design of the gRNA. Despite the fact that most reported clinical *NRXN1* mutations affect only the *NRXN1*α isoform, some cases report having both isoforms affected [133]. Studies in mice showed that knock-out of the α neurexins have a lethal effect due to severe impairments in synaptic transmission, while perturbation in endocannabinoid signalling were reported in the β neurexins knockouts [82], [194].

Pak et al (2015), generated two heterozygous mutations, affecting both isoforms, one by creating a frameshift mutation and another by introducing a stop codon causing a truncation in *NRXN1*. This was the first reported study showing the functional effect of *NRXN1* mutation in human induced neurons and resulted in disruption in neurotransmitter release. In order to affect both *NRXN1* isoforms, the gRNA was designed to target the first shared exon, causing indel mutations. Theoretically, this should result in a frame shift mutation affecting the transcriptional levels of both *NRXN1* isoforms [195]. CRISPR-induced mutations in this region resulted in four different indel mutations, two heterozygous and two homozygous mutations. The four mutated lines were validated using Sanger sequencing and PCR.

This was followed by neural induction of the CRISPR-edited cell lines to ensure that the genetic modification did not tamper with the ability of the mutant lines to differentiate into neurons, the iPSCs (containing an *NGN2* inducible cassette) were induced using chemically defined neuronal media supplemented with DOX. This results in the rapid generation of excitatory cortical neurons as reported by Bertero et al (2016)[230].

In the second part of this chapter, the patient lines containing *NRXN1* mutations were targeted with the OPTi-OX system to allow the rapid and deterministic forward programming of the patient lines into a pure neuronal culture.

The TET-ON system was used for the inducible overexpression of NGN2, the TET-ON elements were inserted in two different GSH loci using CRISPR-Cas9 to target the human ROSA26 locus with the CAG-*rtTA* cassette, and zinc-finger nuclease (ZFN) to target the AAVS1 locus with the TRE-NGN2 cassette [229], [230]. The pRO26_CAG-*rtTA* containing a neomycin resistance gene was inserted in the NRXN1-Patient lines and the colonies surviving neomycin selection were further genotyped using four primer pairs spanning, the wild type, the 5'BB, the 3'BB and the transgene. After genotyping, the clones homozygous or heterozygous for *rtTA* were selected. These clones were then targeted with the pAAVS1_TRE-*NGN2* containing puromycin resistance gene, the clones surviving puromycin were then genotyped using a similar primer designing strategy to ensure the successful integration of the *NGN2*.

To validate their ability to generate cortical neurons, the OPTi-*NGN2* patient cell lines were then treated with chemically defined media plus or minus DOX. As early as four days post induction, the DOX-treated cells showed neural like morphology and formed neural network at day 14 as shown in the bright field images in Figure 3-6B and Figure 3-7B. The cells treated with DOX also showed a statistically significant

increase in *NGN2* expression and decrease in *OCT4* and *NANOG* expression at day 4 when compared to the non-treated cells, confirming that DOX induces *NGN2* expression leading the cells to differentiate, hence the decrease in pluripotency markers. The DOX-treated cells have also shown a higher expression of pan-neuronal (*MAP2*) and glutamatergic genes (*vGLUT1* and *GRIA4*) after two weeks of induction, indicating the generation of induced excitatory cortical neurons. For further confirmation of neuronal identity, forebrain markers such as BRN2 and FOXG1 will be useful.

During neural induction, *NGN2* levels increase driving the cells to a neuronal fate. Quantifying the expression levels of *NGN2*, showed it is dox-inducible, the *OCT4*, *NANOG* pluripotency markers were only expressed at the pluripotent stages and no expression was reported starting day7 post-induction, confirming that the cells are differentiating. *MAP2* (pan-neuronal gene), *vGLUT1* (presynaptic excitatory gene) and *GRIA4* mRNA expression levels in induced neurons 7days post-induction and until day21 confirm the efficient conversion of the *NRXN1*-Isogenic and *NRXN1*-Patient iPSCs into a homogenous cortical excitatory neuronal population.

The forward programming of these iPSC lines with different NRXN1 mutation provide an invaluable tool to study the mutation effects on molecular, morphological and electrophysiological properties of the neurons.

Chapter 4

Transcriptome analysis of *NRXN1*-Patient and *NRXN1*-Mutant lines

Summary

Autism is a polygenic multifactorial condition, caused by an interplay between genetics and the environments. Risk genes like *NRXN1* are only one player contributing to a wider network of genes and *NRXN1* is likely involved in a cascade of regulatory events affecting its associated gene networks [273]. This might explain its incomplete penetrance and the variability in its clinical representation. To investigate the pathways associated with *NRXN1* sequence variants, bulk RNA-Seq was performed on induced neurons (21 days post induction) from two autism patients (*NRXN1*-patient), four CRISPR-edited (*NRXN1*-mutant) and a wild-type control (BOB-NGN2) line.

4 Transcriptome analysis of *NRXN1*-Patient and *NRXN1*-mutant lines

4.1 Introduction

Recent transcriptomic analyses reveal that most of differentially expressed genes found in autism cases are predominantly expressed in cortical neuronal lineages [83]. Other studies in cortical post-mortem tissues reported that the upper-layer excitatory cortical neurons and microglia are the most affected in autism samples. These findings have encouraged the use of iPSC models to investigate molecular and functional aspects of autism for their ability to differentiate into specific cellular types of the brain in early developmental stages using forward programming [229], [230], [232]. Conversely the use of CRISPR genome editing tools allowed the development of iPSC models with specific autism variants [195], [274].

NRXN1 heterozygous mutations have been linked to autism and schizophrenia using genome wide association studies [130], [131], [273], [275]. Recent studies relied on the use of autism patient-derived iPSC with *NRXN*1 mutations or introduced specific mutations to investigate transcriptomic changes associated with *NRXN*1 variants [181]. Avazzadeh et al looked at transcriptomic and functional profiles of cultured neurons derived from autism patients carrying NRXN1α+/– iPSC and compared them to control. They report upregulated voltage gated calcium channel genes consistent with increased calcium signalling transients [181].

Conversely, another study used autism patient-derived iPSC with bi-allelic NRXN1α-/- deletion and reported a significant depression in calcium signalling activity. Moreover, single cell-RNA Sequencing reported increased expression of radial glia-like cells, consistent with results from previous iPSC studies [100], [180], [276] and autism cerebral cortex [277], [278]. Avazzadeh and Lam et al both used patient-derived iPSC carrying heterozygous and homozygous mutations, respectively, and reported conflicting results regarding calcium signalling. Given that both studies relied on patient-derived iPSC and not isogenic lines, it is unclear if the variability in the phenotypes reported are directly related to the type of mutation presented or due to the patient's own genetic background or a combination of the two. Therefore, the use of isogenic lines is ideal to dissect the impact of specific mutation on molecular or functional phenotypes. Pak et al introduce two conditional NRXN1 mutations in iPSC cells to investigate functional impacts of autism and schizophrenia related NRXN1 mutation on human neurons. This generated one cell line with a mutation in exon 19 and another cell line with a truncation mutation in exon 24. In their study they report that both mutations impair synaptic transmission [195].

The mutations reported in NRXN1 are mostly non-recurrent and with variable penetrance [129]. Evaluating the impact of specific mutations on human neurons is challenging without having a clearer view on their impact on the molecular pathways underlying the observed phenotype.

In this chapter, the aim is to explore the effects of NRXN1 mutations on the transcriptome of human neurons using patient-derived and CRISPR-edited neurons 3 weeks post induction.

4.2 Results

The iPSC lines used in this experiment were divided into three groups. The first group was the *NRXN1*-Patient lines, this group comprised two individuals with mutations in the *NRXN1α* isoform, a female with a 200kb deletion (092_NXF), and a male (211_NXM) with a point mutation. The second group is the NRXN1-isogenic lines (*NRXN1*-mutant lines), consisting of four CRISPR-edited lines, with homozygous and heterozygous indel mutations in the exon19 of the *NRXN1* gene, namely (A2, A7, E9 and G4). The last one is the control/wild-type group; a male iPSC line (BOB-NGN2) with no mutations in the *NRXN1* gene. The iPSC-derived neurons were collected at Day 21-post induction for bulk RNA sequencing, as the neurons shows high expression of pan-neuronal markers at this stage as shown in our Q-PCR data and according to recent literature (Figure 4-1) [229], [230].

Figure 4-1: Schematic representation of experimental design showing the neural induction timeline; iN will by collected at 21 for RNA extraction and RNA-Seq (orange arrow). The arrow indicates the days the cells are treated with neural induction media (green) and neural maintenance media (yellow), the red arrow shows the day the iNs are co-cultured with the rat mixed glial culture (day3).

4.2.1 Similarity between samples within each group

To look at the direct or indirect effect of *NRXN1* mutations on the transcriptome of Day 21 neurons and try to understand how much the genetic background the lines derived from autism patients contributes to changes in the transcriptome, three groups were used in this study: control with no mutations, another with induced mutations in *NRXN1* and the third autism patient derived with *NRXN1* mutations (Figure 4-2).

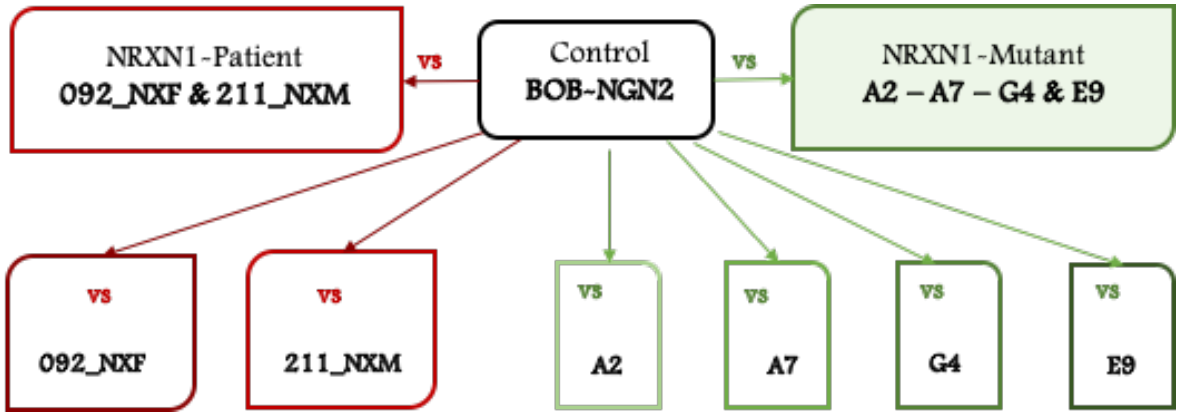


Figure 4-2: Group and Individual comparisons in the RNA-Seq data: Control (BOB-NGN2) was compared to the NRXN1-patient and the NRXN1-mutant groups and then individual comparisons were made between the BOB-NGN2 and each of the patient and the mutant lines.

To determine the main sources of variability and see how similar the samples of each group are; principal component analyses (PCA) was performed as shown in Figure 4-3A and a scree plot showing the components' contribution to the variability in all the samples (Figure 4-3B). The results of the PCA shows the clustering of patient line together at the top right quadrant. The control line is right at the lower left quadrant, clustered close to three *NRXN1*-mutant lines (A7, G4 and E9), except for the A2 line which lies at the bottom right quadrant. The largest source of variation revealed

across the samples accounted for 55.6% of the variance and the second component identified explained 18.03% of the observed variance across the dataset.

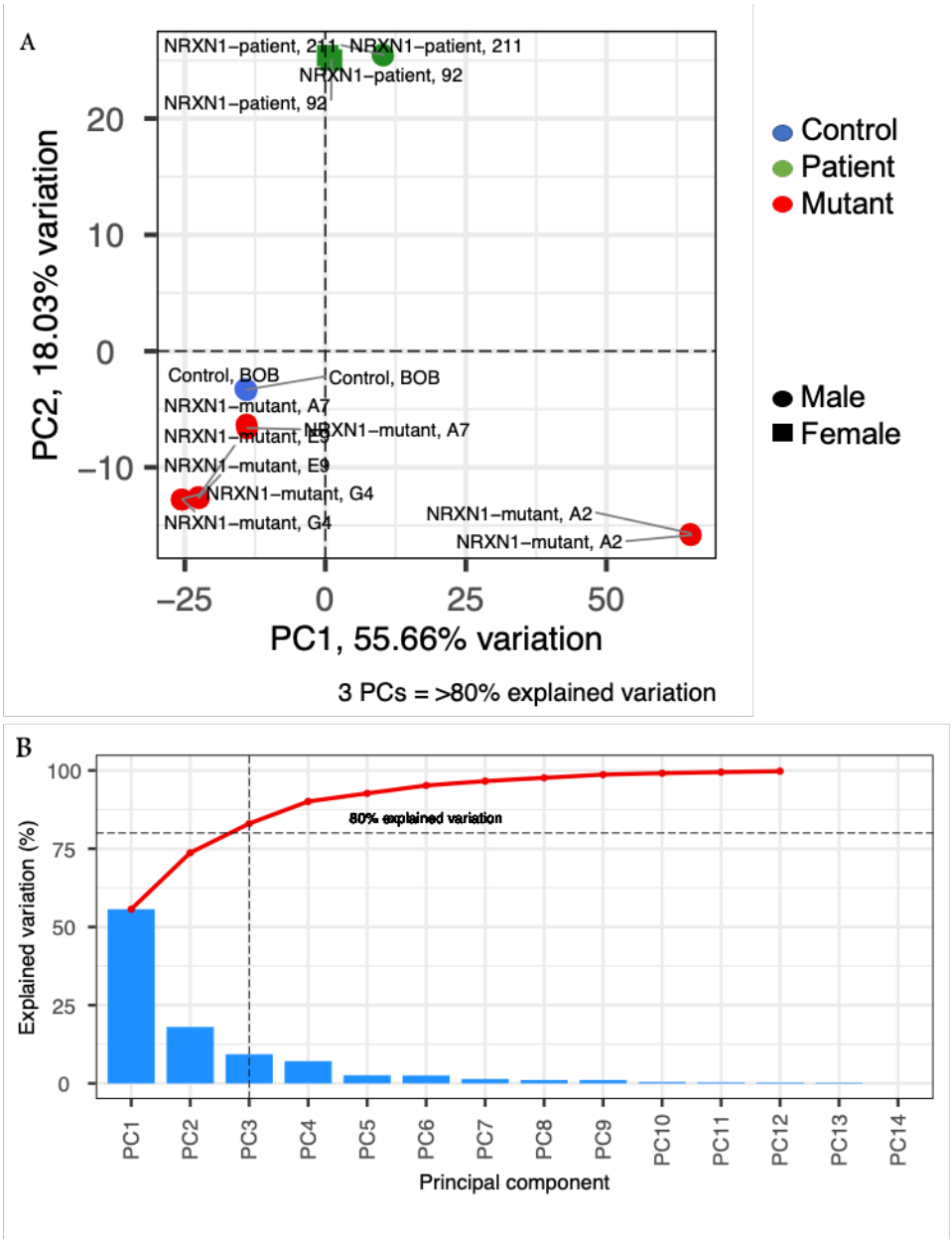


Figure 4-3: Principal component analyses of all samples. A) Biplot PCA, PCA1 is plotted at the X axis (55.6%variation) and PC2 on the Y axis (18.03% variation),the colours of the dots represent the group each sample belongs to and the symbol represent the gender B) Scree plot: all principal components plotted against the percentage of variation they account for, PC1-3 = 80% explained variation.

The distance between samples was calculated to further demonstrate the similarity and closeness between the samples and within groups used in this study. The distance is visualised in a histogram using a colour gradient that goes from red to blue. Red indicates the closeness of the samples and blue indicates the opposite (Figure 4-4). The *NRXN1*-Isogenic lines, the wild type (BOB-NGN2) and three mutant lines (A7, G4 and E9) were clustered together. E9 and G4 are closer to each other than BOB-NGN2 and A7. The *NRXN1*-patient lines (092_NXF and 211_NXM) were tightly grouped together. The A2 *NRXN1*-mutant line is not clustered or showing resemblance to any of the other lines included in this study.

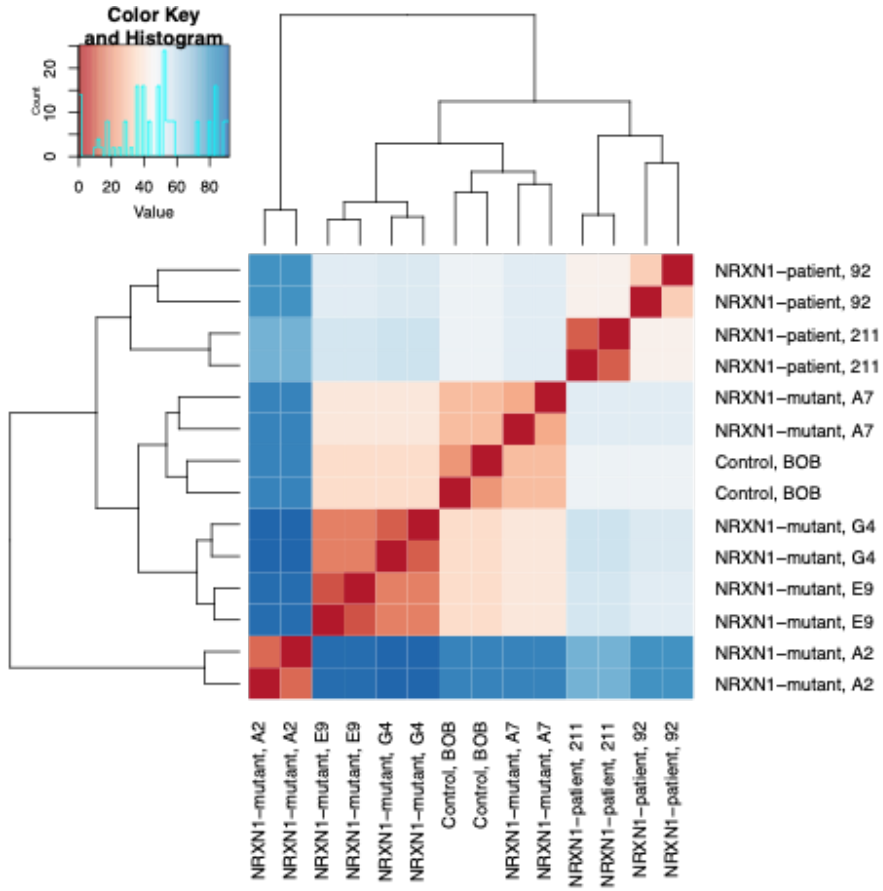


Figure 4-4: Heatmap of sample-to-sample distance. Samples are listed from bottom to top and from left to right in the same order. The colour key on the top left of the figure indicates the distance between samples, by representing the distance value by a colour gradient that goes from Red (0) to Blue (100).

4.2.2 Differentially expressed genes (DEGs) resulting from comparing the control group to the NRXN1-Patient and NRXN1-mutant groups

The DEGs were defined as those with the log2 fold change ≥2 or ≤-2 between two groups with an adjusted p-value and FDR < 0.05, corrected for, using the Benjamini-Hochberg procedure. All the lines were plotted in a supervised heatmap with a colour gradient showing the resemblance of the genes based on the Z-score assigned (Figure 4-5). Initially, two main comparisons were carried out: *NRXN1*-mutant vs Control and

NRXN1-patient vs Control (Figure 4-2). Then individual lines were each compared to the control group separately.

4.2.2.1 *Downregulation of neural development related genes in NRXN1-mutant group when compared to the control group*

The mutations in *NRXN1* in the *NRXN1*-mutant lines resulted in the downregulation of 7 genes, 3 of which are related to neural development such as *BRINP3*, NKX6-2 and *OTP (Figure 4-6)*. Those genes are involved in the regulation of neurogenesis. *BRINP3*, also known as Bone Morphogenetic Protein/Retinoic Acid Inducible Neural-Specific 3 is a protein coding gene associated with pituitary tumours and cancers. The main role of *BRINP3* is the inhibition of neuronal cell proliferation [279]. The transcription factor *NKX6-2*, downregulated in *NRXN1*-mutant lines, is known for its role in the regulation of axon-glial interactions [280]. And lastly *OTP*, a member of the homeodomain family (HD), playing a key role in cell fate specification during brain development [281]. Due to the small number of genes revealed in this group, no significant enrichment was reported for any gene ontology pathways (FDR<0.05).

4.2.2.2 *Upregulation of CNTN6 in the NRXN1-patient group when compared to the control*

In the *NRXN1*-patient lines, 19 genes were upregulated, and 8 genes were downregulated (Figure 4-6). The upregulated genes were not significantly enriched

in GO analysis, but the list included some interesting genes such as *CNTN6, PSMD13, CHL1* and *RPS21*, genes that have been previously linked to "axon guidance" (GO:0007411). Moreover *CNTN6 (*Contactin 6*)*, a paralog of *CNTN4*, is a strong autism candidate gene (SFARI gene database) and is also associated with Tourette syndrome and Attention Deficiency and Hyperactivity Disorder. *CHRNB4* a neuronal acetyl choline receptor, previously linked to nicotine addiction was downregulated in the *NRXN1*-patient lines [282].

CAT (catalase), the gene responsible for metabolising hydrogen peroxide and mitigating its toxic effects was upregulated in *NRXN1*-patient lines and downregulated in the *NRXN1*-mutant lines. This is the only gene that was differentially expressed in both groups. Known and reported effects of mutations in this gene are known to cause acatalasemia. Amongst the GO terms related to this gene are "antioxidant activity" (GO:0016209) and "signalling receptor binding" (GO:0005102) [283], [284].Surprisingly, *NRXN1* was not reported to be differentially regulated either in the *NRXN1*-patient, or the *NRXN1*-mutant neurons when compared to the control.

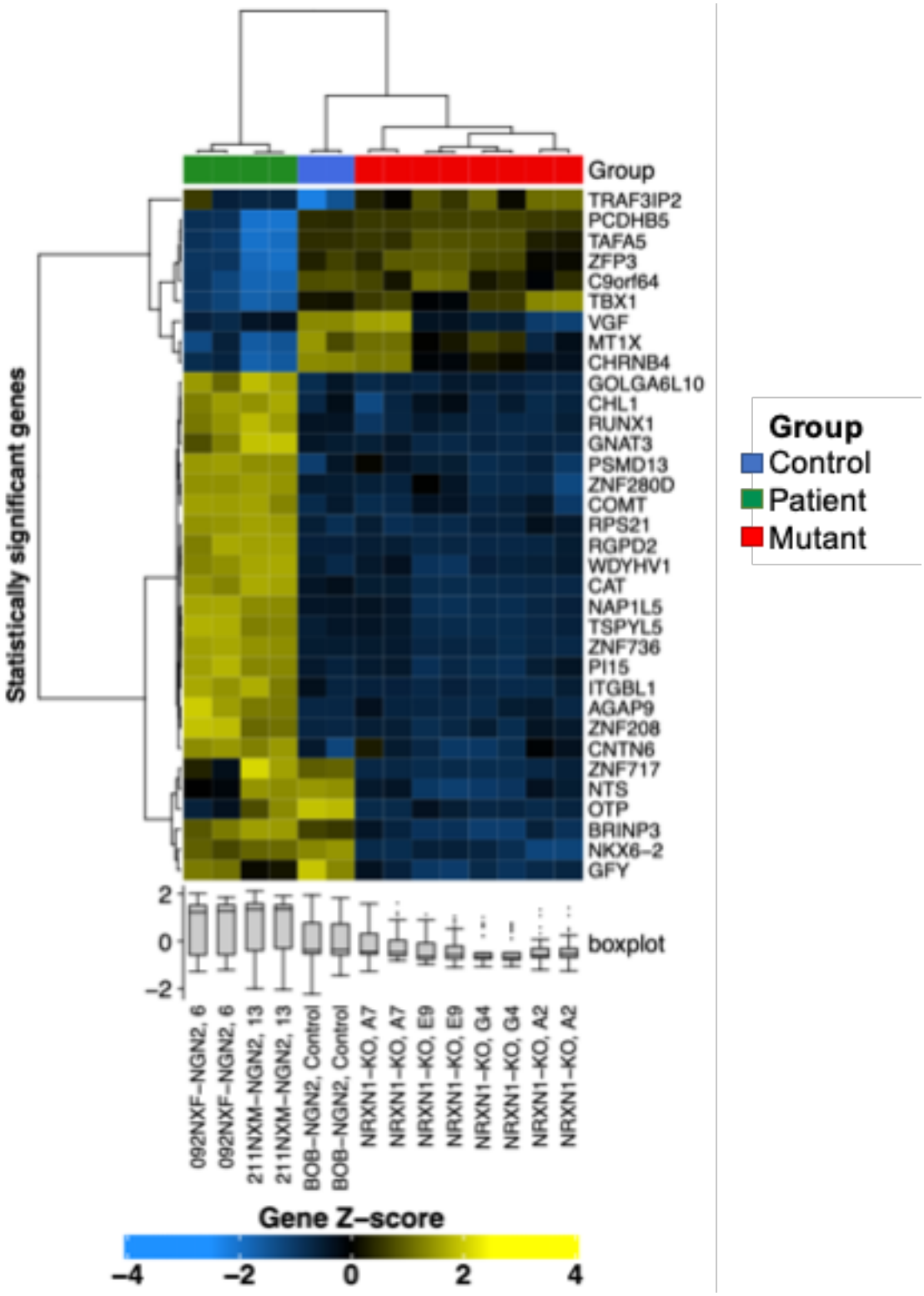


Figure 4-5: Heatmap of differentially expressed genes in all group, a the top of the graph the square colours represent the group: blue= control, green=patient and red=mutant. The colour key indicates the Gene Z-score, the blue indicates a lower Gene Z-score and yellow represents a higher score.

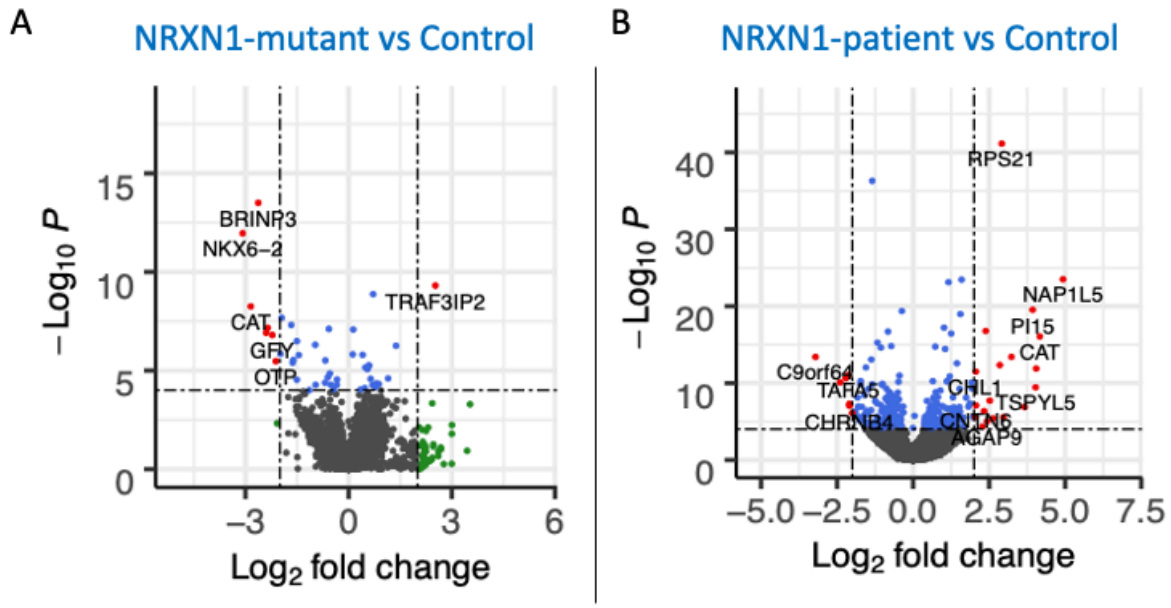


Figure 4-6: Volcano plots showing differential gene expression of NRXN1-Patient lines vs control (A), and NRXN1-mutant lines vs Control. The blue dots represent the expressed genes that are not significantly differentially regulated, while the red ones are significant upregulated (Log2 fold change >2 - upper right quadrant) or downregulated (Log2 fold change< -2 - upper left quadrant) in NRXN1-mutant (A) and NRXN1-patient (B)lines

4.2.2.3 NRXN1 coverage revealed differential expression pattern on individual cell line basis

To have a deeper look at *NRXN1* expression in the lines included in this study, a pile-up of RNA-Seq reads covering region chr2:49918505:51032561 (*NRXN1* sequence) was produced using bamCoverage from deepTools; smooth length 40 and bin size 25 by individual cell lines and not by group and visualised on bedGraph (Figure 4-7) in the UCSC Genome Browser [285]. This is a qualitative representation of *NRXN1* exonal presentation in each line by the presence of the peaks, the higher the peak is the higher the number of reads in this region. The *NRXN1*-patient lines show different expressions as can be seen in Figure 4-7, while 211_NXM shows larger peaks than that

of the BOB-NGN2, while the 092_NXF patient line do not show any peaks. This means that each mutation in the patient lines affected *NRXN1* gene differently, the 200Kb deletion in 092_NXF decreased its expression to almost null, while the point mutation in the 211_NXM resulted in an increased expression of the gene. Similarly, the mutations in the *NRXN1*-isogenic lines seem to affect the *NRXN1* gene differently. The A2 showed close to no expression of *NRXN1* exonal regions. The A7 shows little to no peaks when compared to the wild-type (BOB-NGN2) while E9 mutant line shows bigger peaks and seem to have higher expression than the wild type. The G4 doesn't show notable changes.

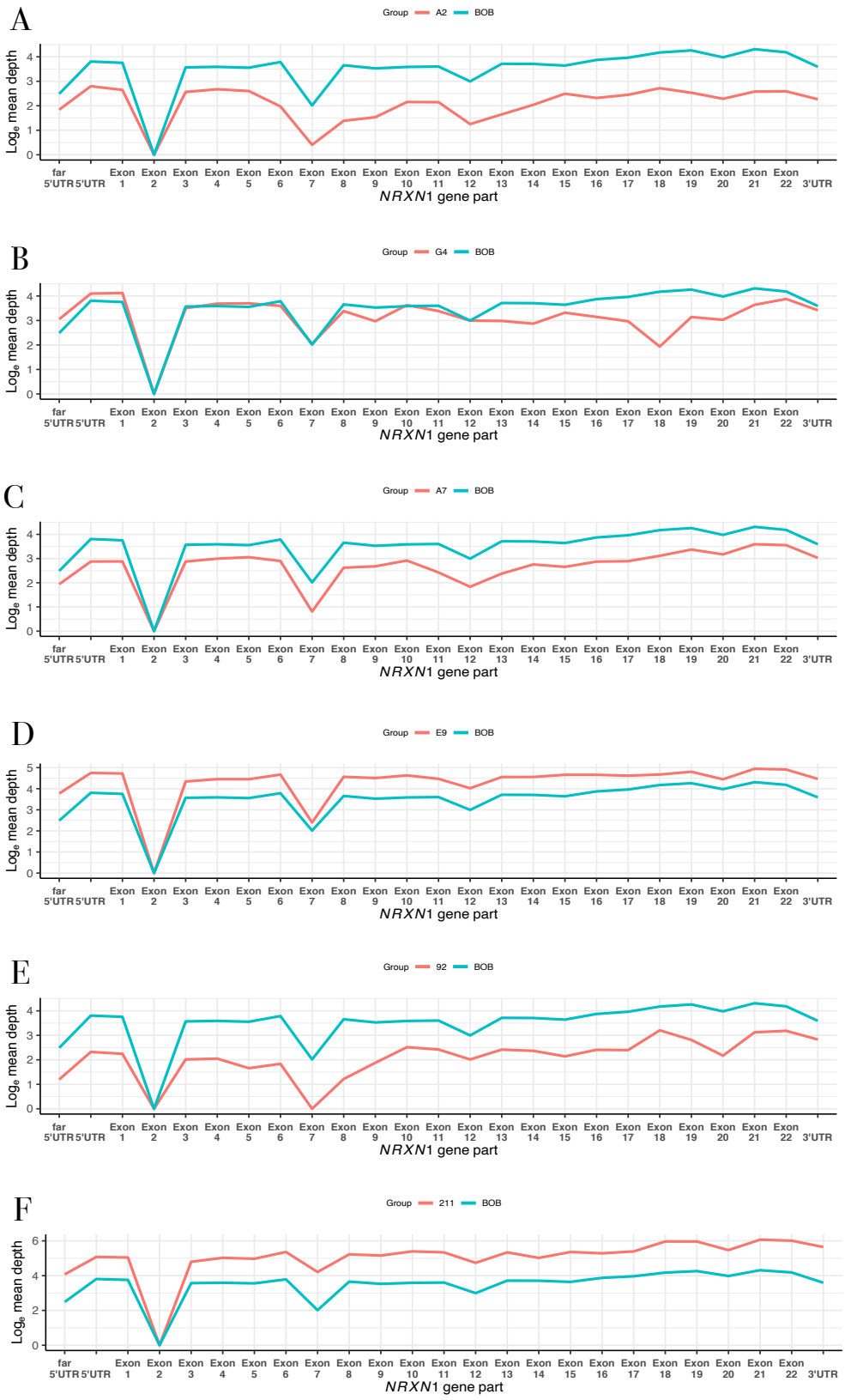


Figure 4-7: NRXN1 gene exonic profile for, A) A2, B) G4, C) A7, D) E9, E)092NXF and F) 211NXM (orange) each compared to the control BOB_NGN2 cell lines (blue line).

The different exonic profiles of NRXN1 exhibited by the samples within the same group warranted further investigations to compare individual cell lines within each group to the control group. Therefore, individual cell lines within each group were compared to the control cell line (BOB-NGN2). A total of six comparisons :(092NXF vs Control), (211NXM vs Control), (A2 vs Control), (A7 vs Control), (G4 vs Control) and (E9 vs Control) (Figure 4-2). The differentially expressed genes from each comparison were identified as previously described and an overrepresentation analysis (ORA) was done using Webgestalt. The genes were annotated and mapped using a reference gene list of all the expressed genes between the groups against three functional databases (KEGG, Panther and Reactome) to check for enrichment analysis. The top 15 statistically significant categories were identified where FDR<0.05 and corrected for using the Benjamini-Hochberg procedure.

4.2.3 DEGs resulting from individual NRXN1-patient lines comparisons (092_NXF vs BOB-NGN2) and (211_NXM vs BOB-NGN2)

The *NRXN1*-patient lines were compared individually to the control line (BOB-NGN2. In the 092NXF-NGN2 line; 273 genes were differentially regulated, amongst which 69 genes were downregulated (Figure6-8A) while 233 genes were upregulated, and 120 genes were downregulated in the 211NXM-NGN2 lines when compared to the control (Figure 4-8B). There were 111 upregulated genes overlapping between 092NXF and 211NXM and 25 genes were downregulated in both groups. Interestingly, two of the

upregulated genes in 092NXF and 211NXM have been listed as candidate genes linked to autism; FOXP2 and CNTN6.

FOXP2 a member of the forkhead/winged-helix (FOX) family of transcription factors expressed in fetal and adult brain, is strongly associated with autism [286], [287] and ranks as a high confidence gene in the SFARI database. *CNTN6* (Contactin 6) has also been listed as a strong candidate gene in the SFARI database, and many studies reported the association of this cell adhesion molecule with autism and schizophrenia [288], [289]. Amongst the downregulated genes in both groups, we found that Neuronal acetylcholine receptor subunit α-3 and 4 also known as *CHRNA3* and *CHRNB4, are* both enriched in "acetylcholine receptor activity", "acetylcholine-gated cation-selective channel activity" and "acetylcholine binding" functional pathways (GO:0015464, 0022848 and 0042166, respectively). Those two genes CHRNA3 and CHRNB4 were also enriched in biological processes such as "synaptic transmission involved in micturition" (GO:006008), "synaptic transmission, cholinergic" (GO:0007271), "regulation of neurotransmitter secretion" (GO:0046928) and "excitatory postsynaptic potential" (GO:0060079).

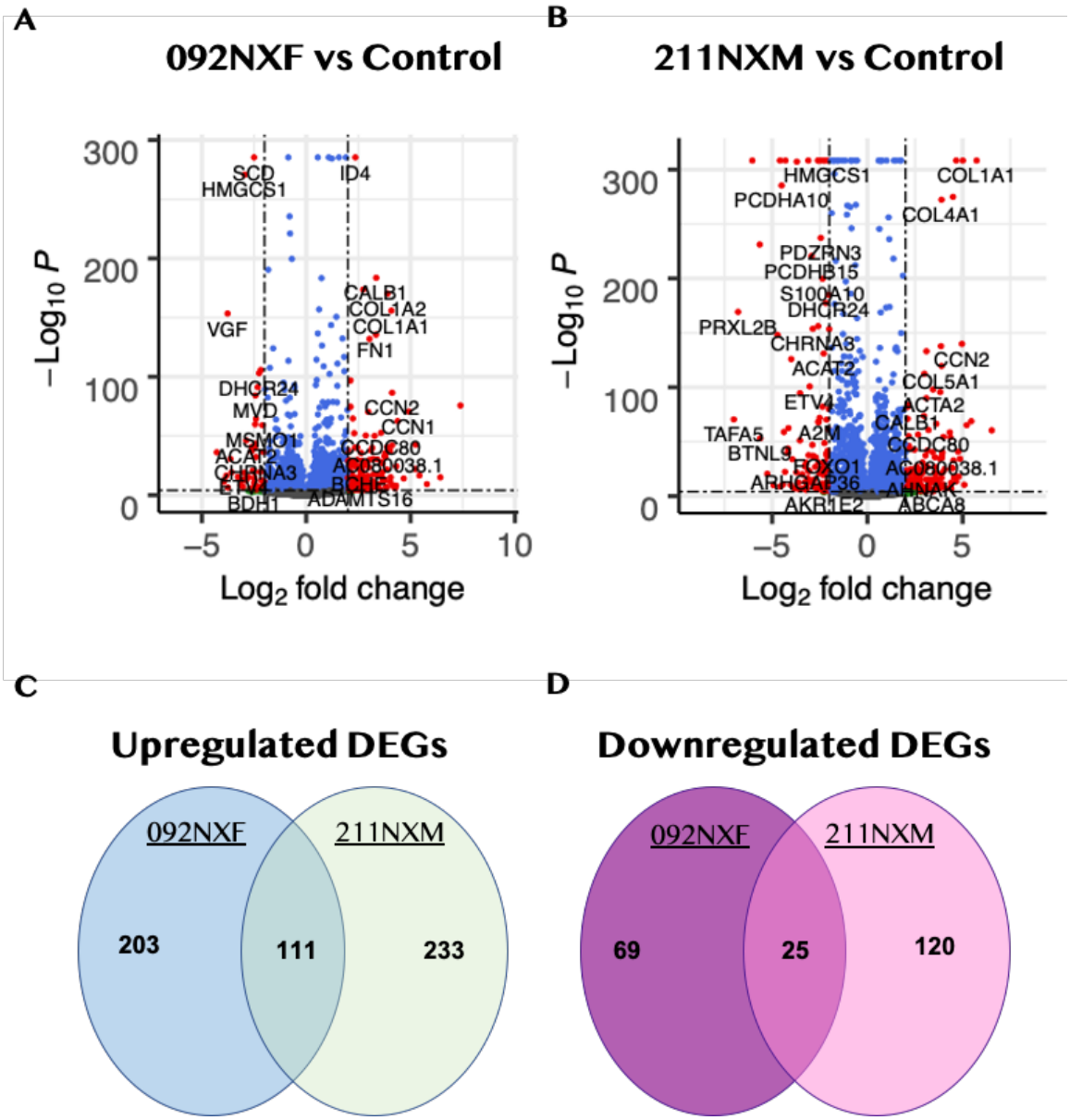


Figure 4-8: Differentially expressed genes in NRXN1-patients, Volcano plots displaying A)092NXF vs Control and B)211NXM vs Control and Venn diagram showing the overlap between C) Upregulated DEGs and D) Downregulated DEGs between 092NXF and 211NXM

4.2.3.1 *Significant enrichment of genes upregulated in 092NXF and 211NXM in Extracellular matrix and cell adhesion GO terms*

More than 30 of the upregulated genes in 092NXF were highly enriched for the gene ontology (GO) term "Extracellular matrix organization" (GO:0030198, FDR= 0.00) and 56 genes were enriched in for cell adhesion (GO:0007155, FDR= 0.00) (Figure 4-9A). To visualize the connection between genes, the same set of genes was analysed using STRING; the genes involved in extracellular matrix in purple, cell adhesion in green and "signalling receptor binding" (GO;0005102) in green (Figure 4-9B).

Consistent with the results shown in the 092NXF patient line, "Extracellular matrix organization" (46 genes, GO:0030198, FDR= 0.00) and "cell adhesion" (65 genes, GO:0007155, FDR= 0.00) were two of the most highly significant GO terms enriched for the genes upregulated in 211NXM when compared to control (Figure 4-10A). String analyses revealed that like the 092NXF line, the genes were related to "signalling receptor binding" (GO:0005102) (Figure 4-10B).

4.2.3.2 *Significant enrichment of genes downregulated in 092NXF in steroid metabolic process, cholesterol biosynthetic process*

The 092NXF line's downregulated genes were highly enriched in the "steroid metabolic process"; more than 10 genes overlapped with the genes in this category (GO:0008202, FDR=2.09 x 10^{-8}) (Figure 4-11A). ORA results also listed the "Cholesterol

biosynthetic process" (GO:0006695, FDR=1.98 x 10^{-8}) in the top significantly enriched pathways. Moreover, *EMX2, EOMES* and *PAX6* involved in the "cerebral cortex regionalization" (GO: 0021796) were also downregulated in this line as revealed by string analyses but not webgestalt.

String analysis also revealed an enrichment in the "highly calcium permeable nicotinic acetylcholine receptors pathway" (R-HAS-629597, FDR=2.853 x 10^{-3}) as 3 out of 9 genes were downregulated in 092NXF (*CHRNA3, CHRNA6* and *CHRNB4*) (Figure 4-11), which have been linked to *NRXN1* variants and nicotine addiction [282].

4.2.3.3 A significant enrichment of 211NXM downregulated genes in chemical synaptic transmission and cholesterol biosynthetic process

An overrepresentation analysis of the downregulated genes in 211NXM revealed a significant enrichment in "chemical synaptic transmission" (GO:0007268, FDR= 8.38 x 10^{-3}), "cholesterol biosynthetic process" (GO:0006695, FDR=1.03 × 10^{-3}) and "cadherin signalling pathway" (P00012, FDR=1.33 x 10^{-3}). (Figure 4-12). The downregulated genes included 9 genes of Cadherin-like superfamily (IPRO15919).

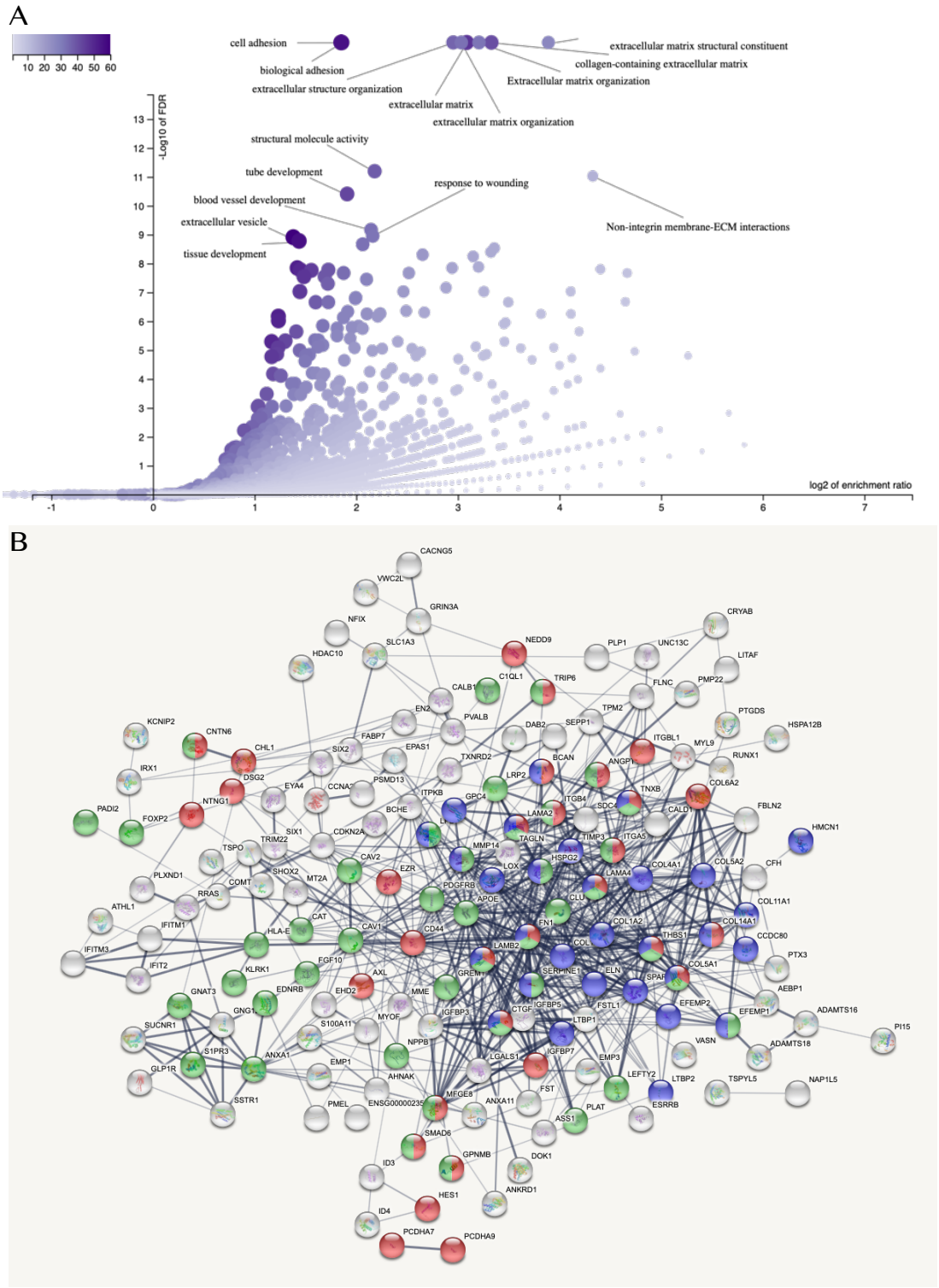


Figure 4-9: Overrepresentation analysis of upregulated genes in 092NXF patient line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for Extracellular matrix organization (purple), Cell adhesion (red) and signalling receptor binding (green), line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

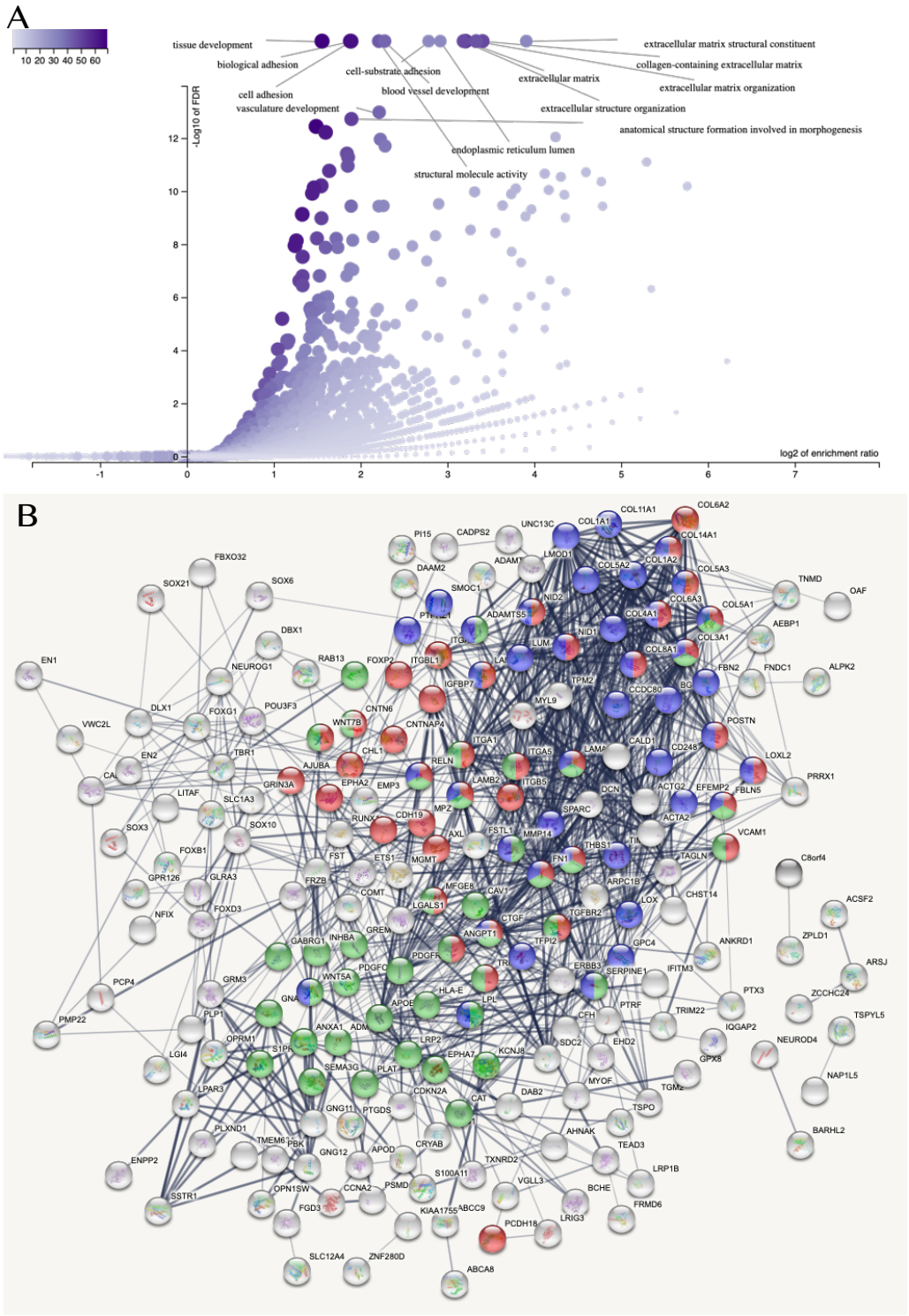


Figure 4-10: Overrepresentation analysis of upregulated genes in 211NXM patient line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for Extracellular matrix organization (purple), Cell adhesion (red) and signaling receptor binding (green). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

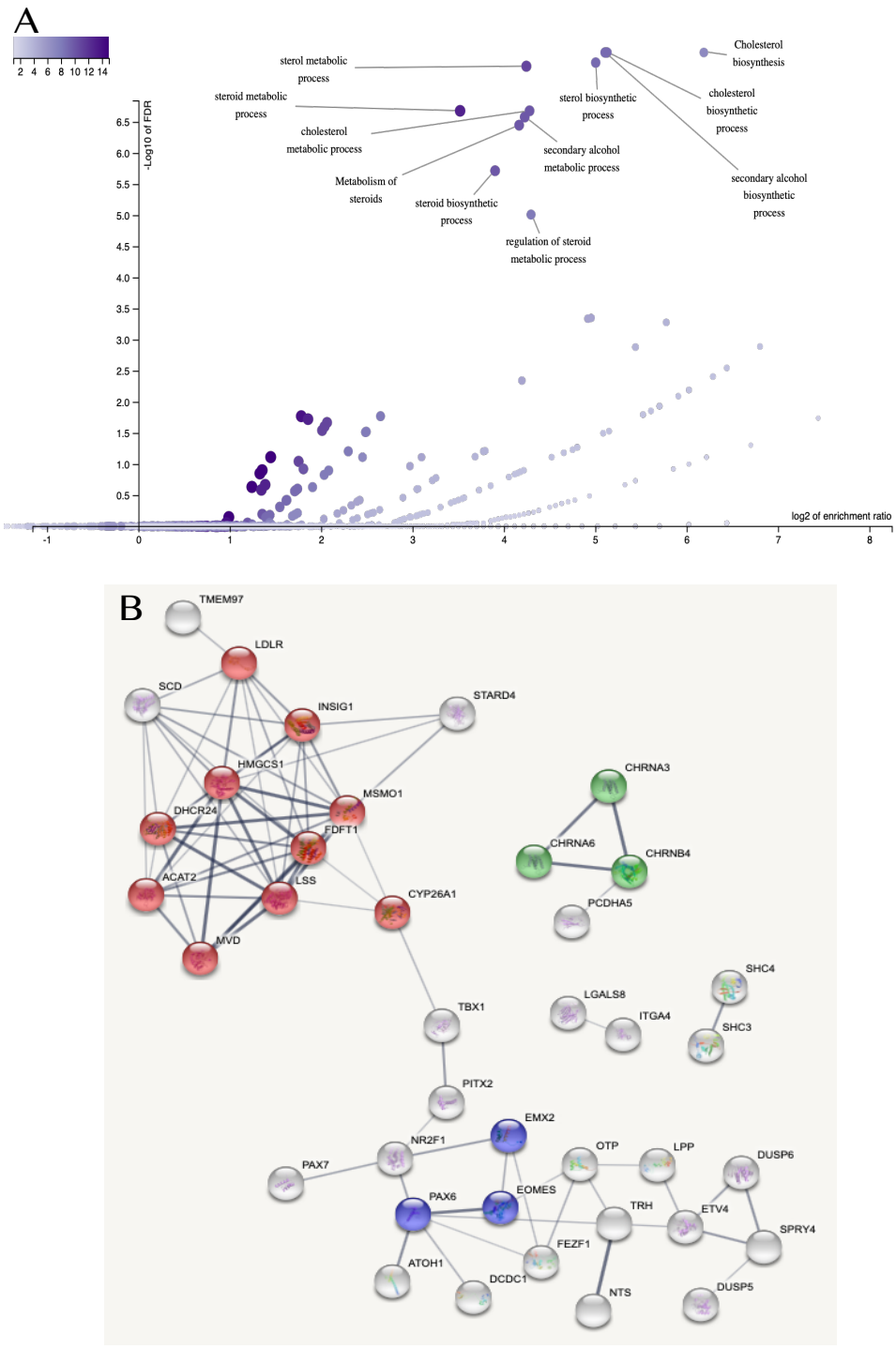


Figure 4-11: Overrepresentation analysis of downregulated genes in 092NXF patient line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for Steroid metabolic process (red), Cerebral cortex regionalization (purple) and Highly calcium permeable nicotinic acetylcholine receptors (green). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

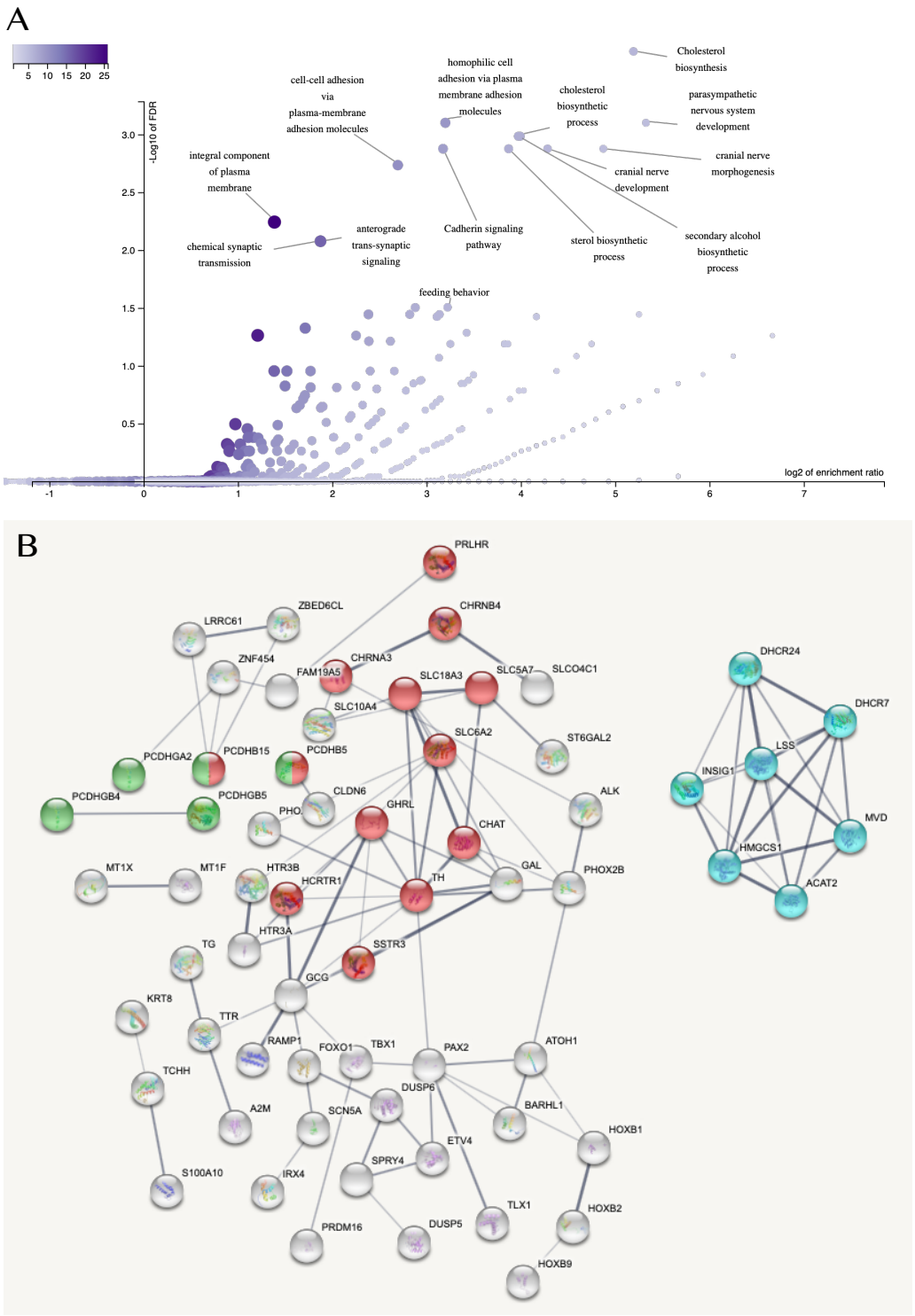


Figure 4-12: Overrepresentation analysis of downregulated genes in 211NXM patient line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for Cholesterol biosynthetic process (turquoise), Chemical synaptic transmission (red) and cadherin-like superfamily (green). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

4.2.4 DEGs resulting from individual NRXN1-mutant lines comparisons (A2 vs BOB-NGN2), (A7 vs BOB-NGN2), (G4 vs BOB-NGN2) and (E9 vs BOB-NGN2)

Each of the four mutant line was compared to the wild type individually (Figure 4-13). The *NRXN1*-mutant line A2 showed the highest number of differentially expressed genes (1364 genes) while only 35 genes were identified in the A7. The E9 and G4 lines revealed 178 and 151 DEGs, respectively. As shown in the Venn diagram in Figure 4-11E, there were a number of genes overlapping between the four lines. For example, 23 genes overlapped with the A2 DEGs, which was more than half of the DEGs identified in A7. Additionally, 114 genes were in common in E9 and G4. Lastly, only 11 genes overlapped across the four mutant lines. The mutations induced in the *NRXN1* gene had diverse effect on the transcriptome as shown by the different number of DEGs reported by each individual comparison. More importantly, *NRXN1* was only significantly downregulated in A2 but not in any of the other mutant lines.

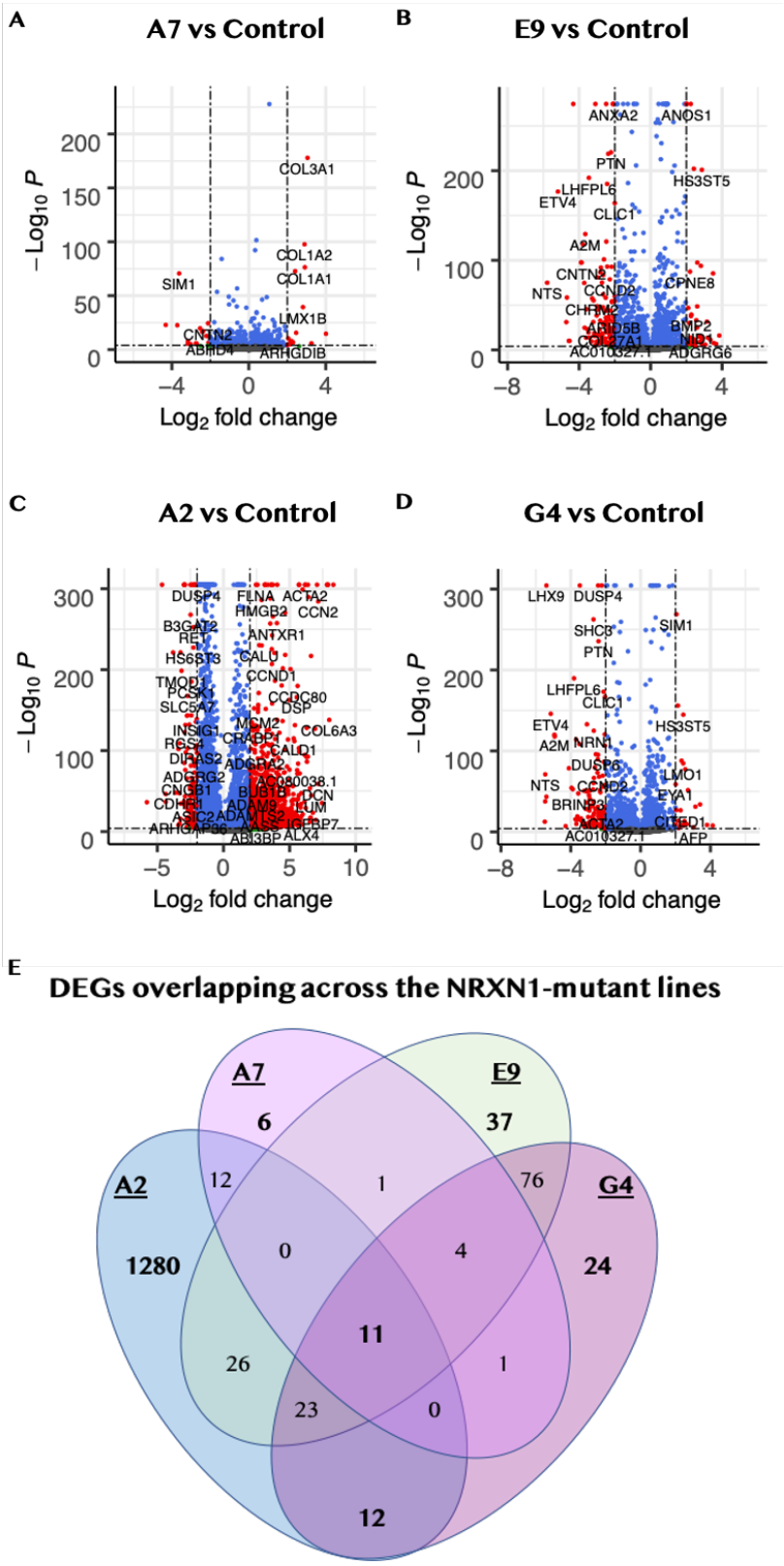


Figure 4-13: Differentially Expressed genes in NRXN1 mutant lines, Volcano plots of A) A7 vs control, B) E9 vs Control, C) A2 vs Control, D) G4 vs Control and E) Venn diagram of the overlapping DEGs across the four mutant lines

4.2.4.1 *The genes upregulated in the NRXN1-mutant line **A7** revealed a significant enrichment for Extracellular matrix organization and blood vessel development*

The A7 mutant line revealed 16 DEGs when compared to the wildtype. Almost half of these genes showed marked enrichment for "Extracellular matrix organization" (GO: 0030198, FDR=2.77 x 10^{-5}), 8 of which are also related to "Blood vessel development" (GO: 0001568, FDR= 5.12 x 10^{-5}) (Figure 4-14). The genes involved in these categories are part of the "Collagen triple helix repeats" family (IPRO8160): (*COL1A1, COL1A2, COL3A1* and *COL6A3*) (Figure 4-14B).

4.2.4.2 *The downregulated genes in the NRXN1-mutant A7 were significantly enriched for generation of neurons and cerebral cortex regionalization*

The mutation in A7 resulted in the downregulation of 19 genes enriched for "Cerebral cortex regionalization" (GO:0021796, FDR=2.86 x 10^{-2}) and "Generation of neurons" (GO: 0048699, FDR=3.86 x 10^{-3}). The top categories showing significant enrichment were related to other neuron development categories (neuron differentiation, central nervous system development, regulation of neurogenesis and forebrain development) (Figure 4-15).

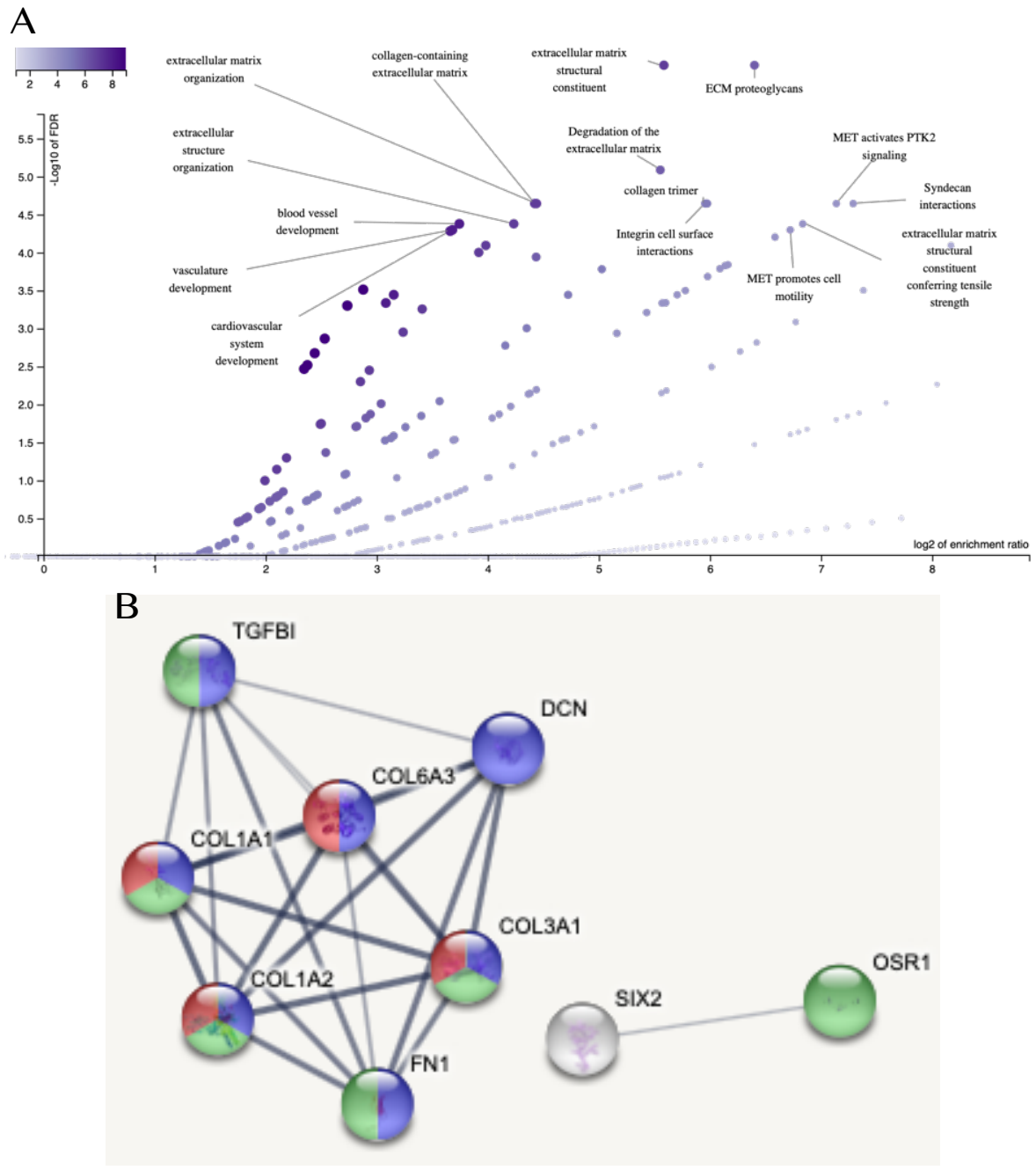


Figure 4-14: Overrepresentation of upregulated genes in A7 mutant line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for collagen triple helix repeats (red), Extracellular matrix organization (purple) and blood vessel development (green). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

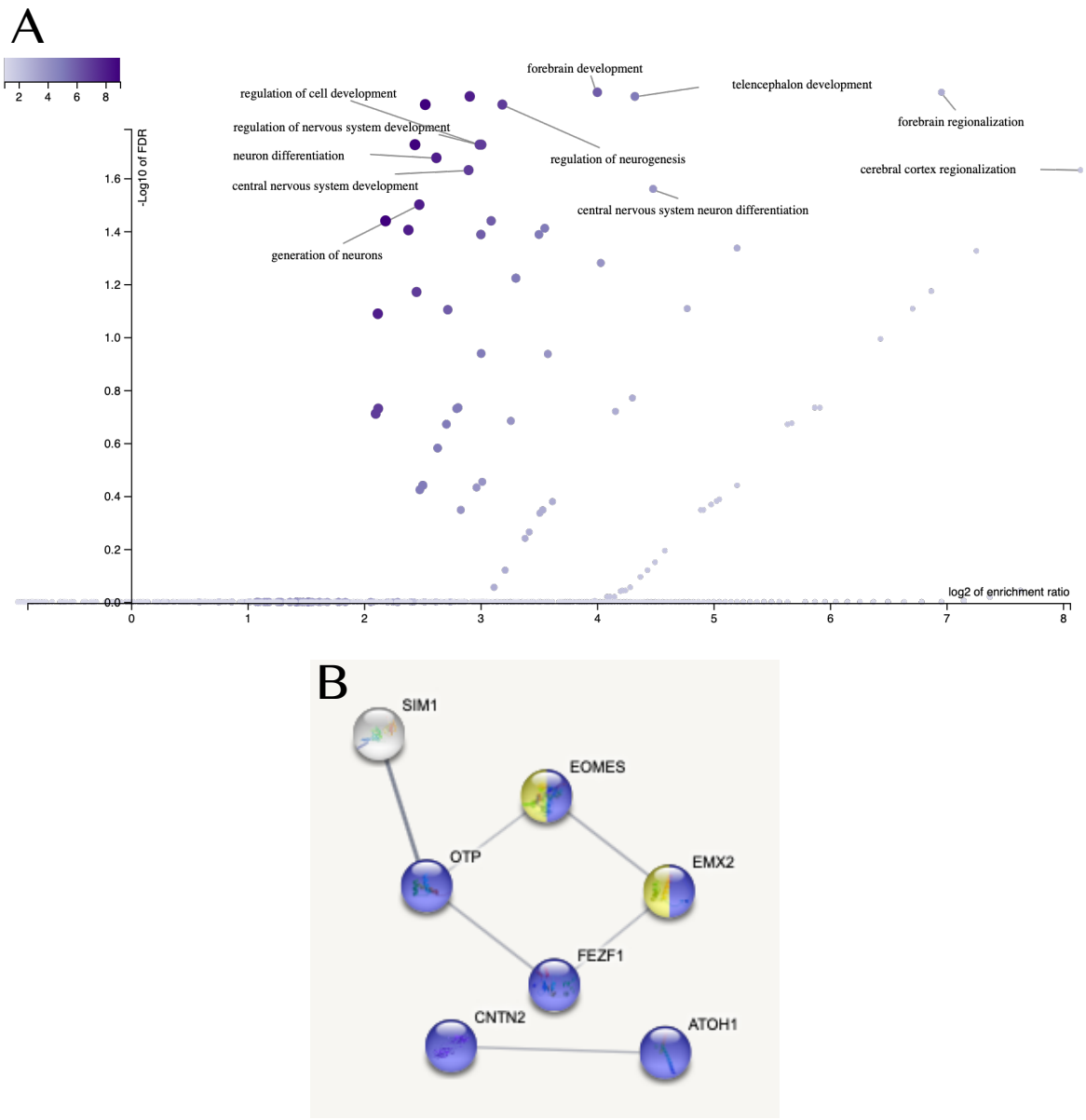


Figure 4-15: Overrepresentation of downregulated genes in A7 mutant line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for "Generation of neurons" (purple) and "Cerebral cortex regionalization" (yellow). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

4.2.4.3 *A significant enrichment of upregulated genes (A2) in Extracellular matrix organization*

More than a thousand genes were reported as significantly upregulated in A2. ORA revealed that 135 genes were significantly enriched in "Extracellular matrix organization" (GO: 0030198, FDR=0.00) which reported the highest enrichment ratio of the GO terms reported (Figure 4-16). Other GO terms included "blood vessel development" (GO: 0003197, FDR=0.00).

4.2.4.4 *The downregulated genes in A2 were significantly enriched in chemical synaptic transmission*

Comparing the A2 to the wildtype revealed 135 downregulated genes, more than 30 of those genes showed a significant enrichment for "Chemical synaptic transmission" (GO:0007268, FDR= 0.00), Cell-cell signalling (GO:0007268, FDR= 1.78 x 10^{-13}) and "Neuronal system" pathway (R-HSA112316, FDR=2.35 x 10^{-13}) (Figure 4-17). Moreover, this is the only mutant line that reported the significant downregulation of *NRXN1*.

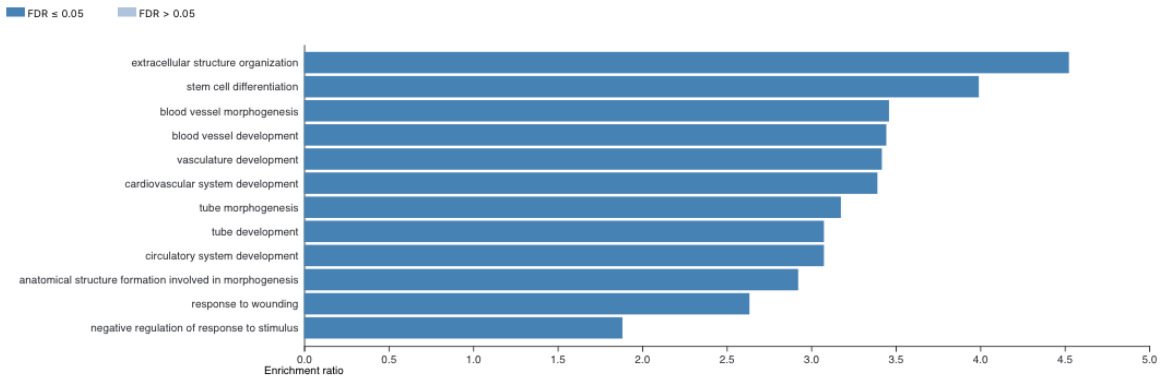


Figure 4-16: Overrepresentation analysis of A2 Upregulated DEGs, the bar graph shows the top significantly enriched GO terms listed according to their enrichment ratio

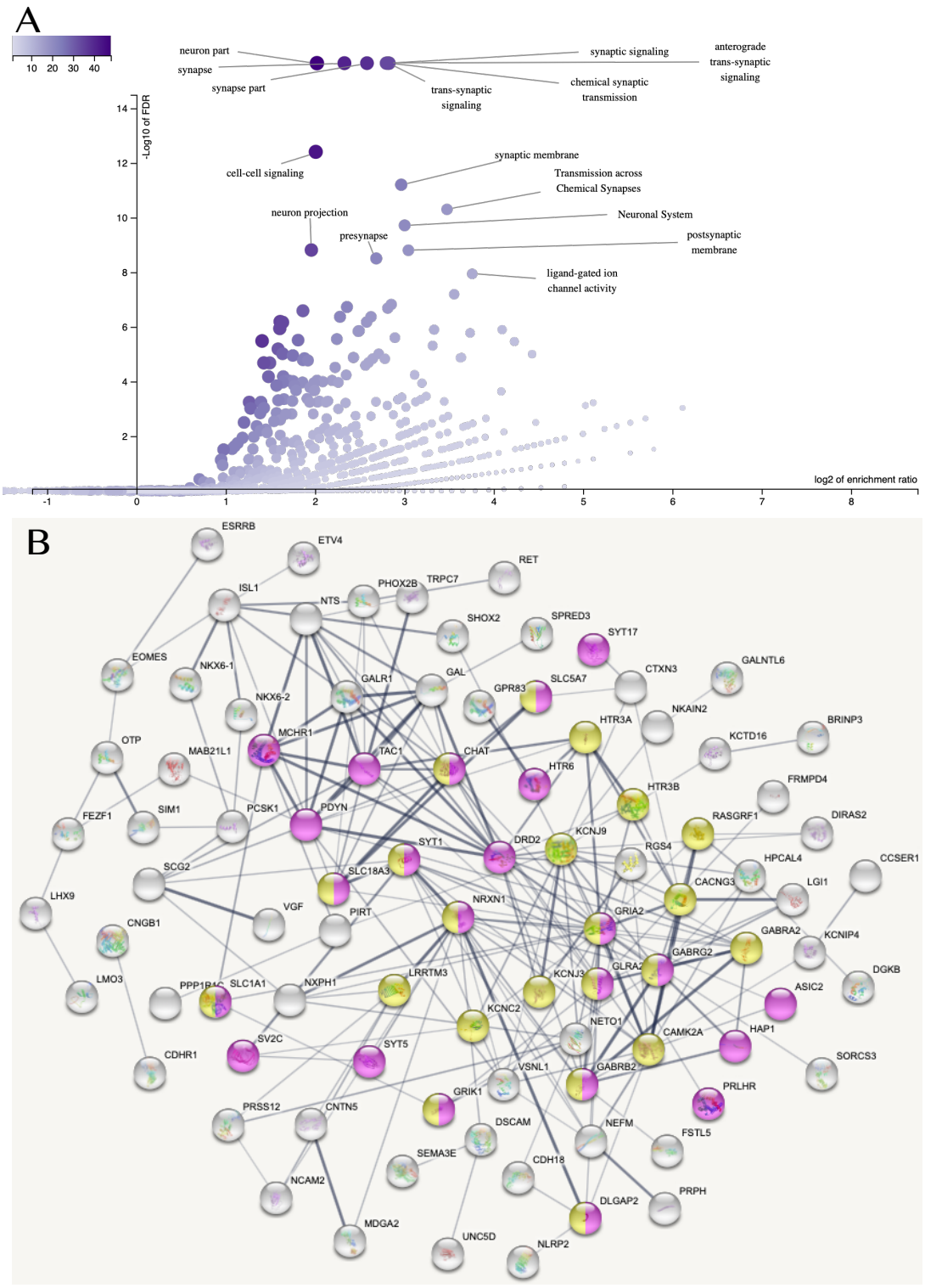


Figure 4-17: Overrepresentation of downregulated genes in A2 mutant line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for "Chemical synaptic transmission" (purple) and "Neuronal system" pathway (yellow). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

4.2.4.5 *E9 upregulated genes show a significant enrichment in gliogenesis and endocardial cushion development*

The DEGs reported in E9 included 45 upregulated genes. These genes were enriched for "Endocardial cushion development" (GO: 003197, FDR= 2.61 x 10^{-2}), "Gliogenesis" (GO: 0042063, FDR= 3.64 x 10^{-3}) and "Glial cell differentiation" (GO:0010001, FDR= 3.90 × 10^{-3}) (Figure 4-18).

4.2.4.6 *Enrichment analysis revealed that the downregulated genes in E9 and G4 are involved in the generation of neurons and Cell-cell signalling*

The G4 mutant line reported 128 downregulated genes while E9 reported 134 downregulated genes and 96 DEGs overlapped between the two groups. The great similarity between the two lines was displayed in the ORA which revealed that in both groups the top significantly enriched GO terms are "Cell-cell signalling" (GO:0007267, FDR= 2.56 x 10^{-5} [E9] and FDR= 3.52 x 10^{-5} [G4]) and "Generation of neurons" (GO:0048699, FDR= 1.72 x 10^{-6} [E9] and 2.84 x 10^{-6} [G4]) (Figure 4-19 and Figure 4-20, respectively).

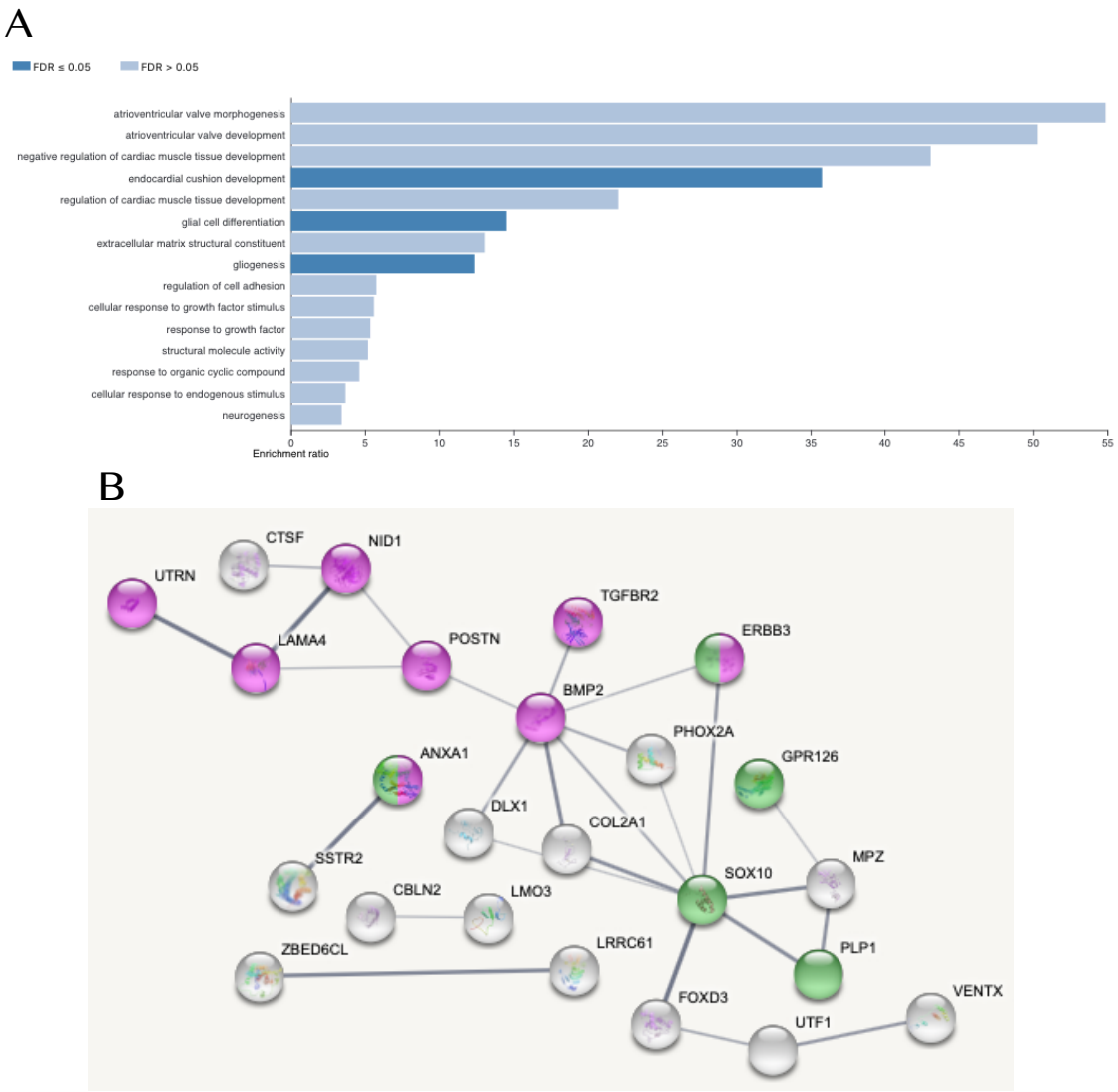


Figure 4-18: Overrepresentation of upregulated genes in E9 mutant line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for "Gliogenesis" (purple) and "Endocardial cushion development" pathway (green). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

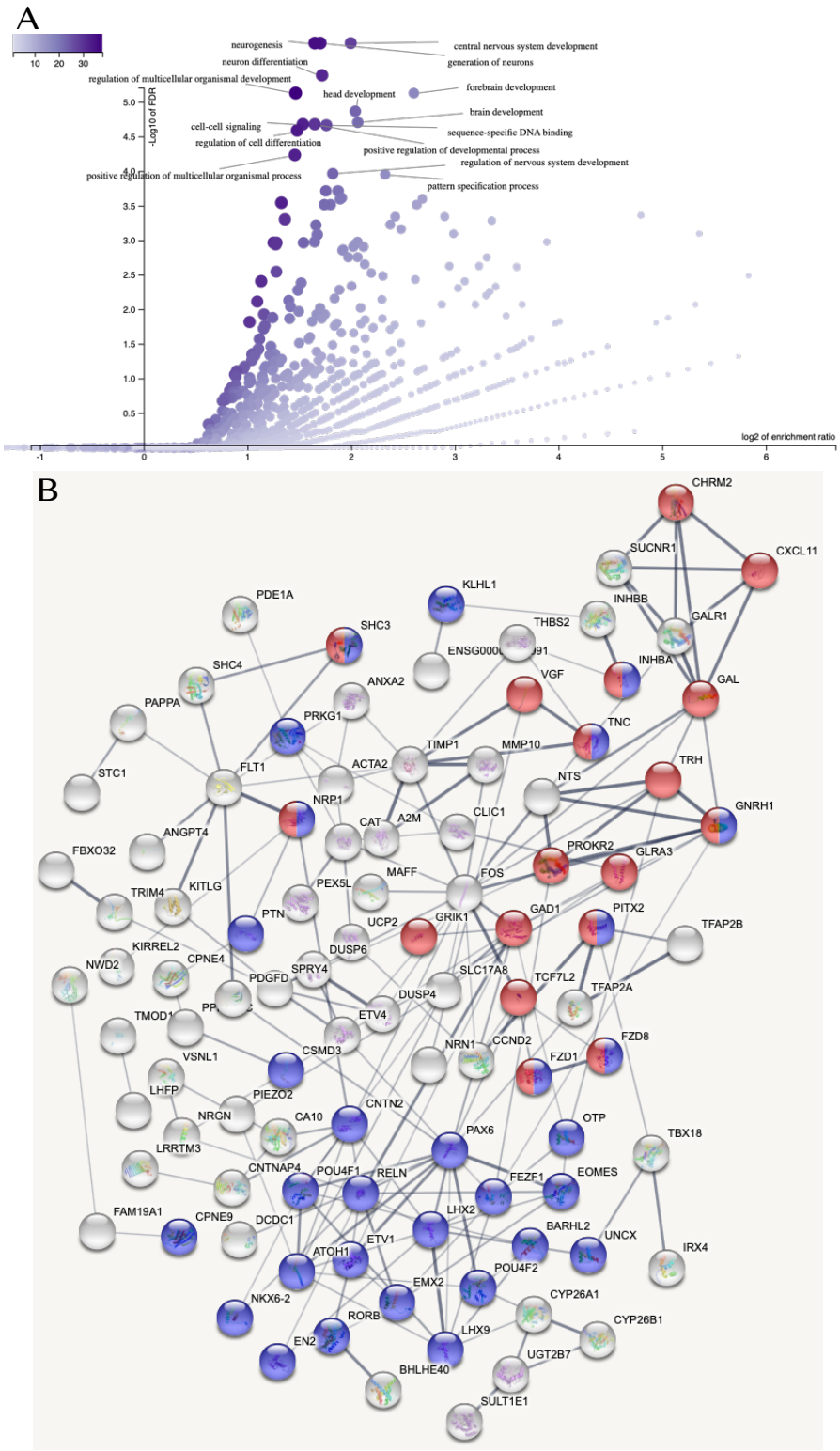


Figure 4-19: Overrepresentation of downregulated genes in E9 mutant line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for "Generation of neurons" (purple) and "Cell-cell signalling" pathway (red). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

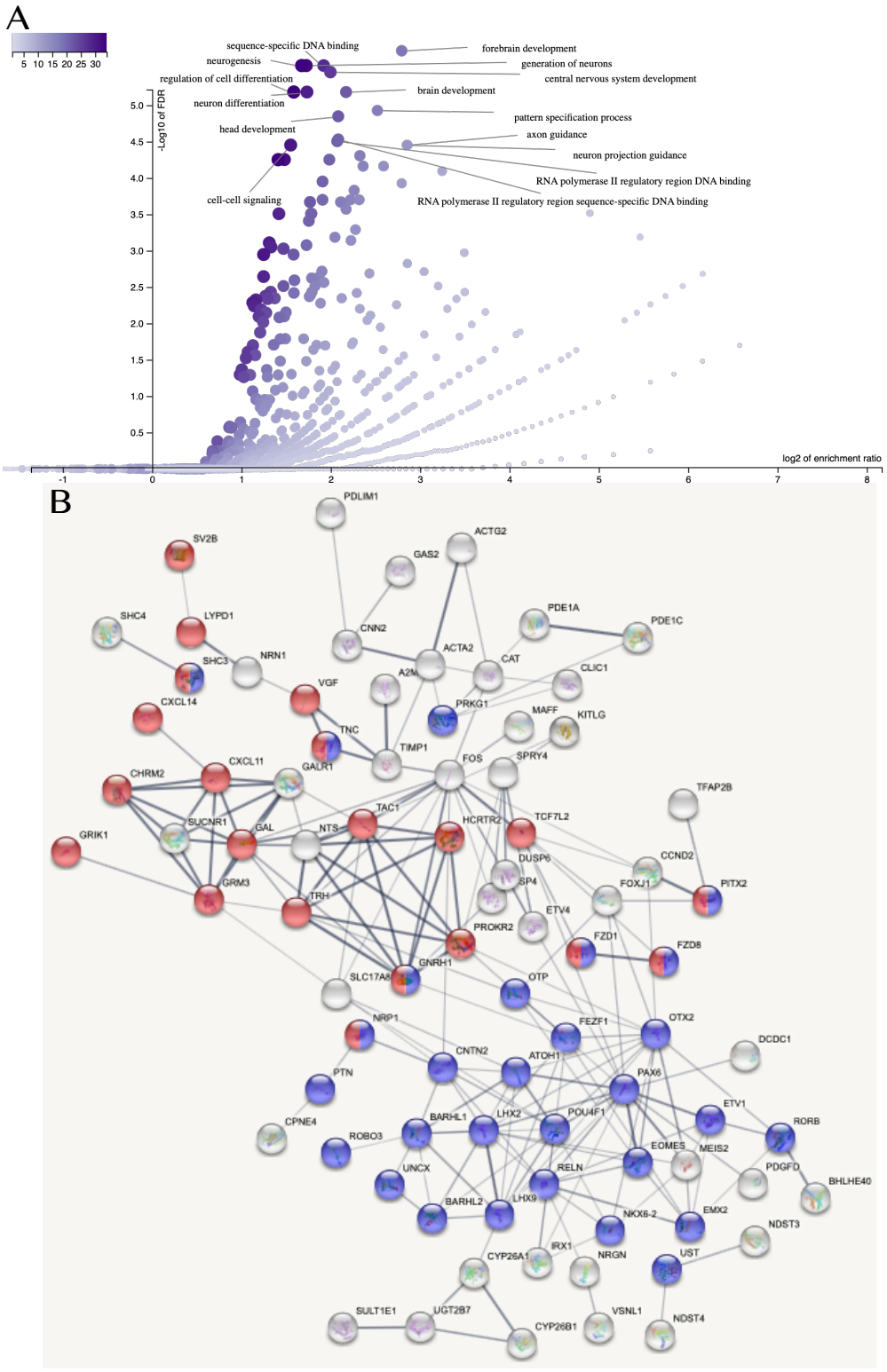


Figure 4-20: Overrepresentation of downregulated genes in G4 mutant line, A) Volcano plot showing the top significantly enriched GO terms (FDR<0.05 - the colour gradient indicates the enrichment ratio as a log2 function. B) STRING analyses showing genes enriched for "Generation of neurons" (purple) and "Cell-cell signaling" pathway (red). line thickness indicated the strength of data support and the disconnected nodes are hidden for better visualization

4.2.5 Summary of ORA resulting from individual comparisons

After carefully screening the DEGs resulting from each individual comparison, it was interesting to note the overlapping DEGs across groups and the large number of collagen-encoding genes; the cardinal components of extracellular matrices in the developing brain. Subsequently, GO analysis reported a significant enrichment for "Extracellular matrix organization" in 092NXF, 211NXM, A2 and A7 upregulated genes (Table 4-1). Interestingly, 7 and 8 upregulated genes in E9 were highly enriched for gliogenesis and glial cell differentiation while no significant enrichment was reported for upregulated genes in the G4 mutant line.

Moreover, the 092NXF, 211NXM and A2 reported 8, 13 and 6 downregulated cadherins. GO terms related neuron development came as top categories for the downregulated genes in the investigated lines (Table 4-2). A significant enrichment was reported for "Generation of neurons" in three out of four NRXN1-mutant lines. Additionally, downregulated genes in 211NXM and A2 were highly enriched for Chemical synaptic transmission. The two patient lines were associated with cholesterol biosynthetic process, and the only female line in this study (092NXF), reported a significant enrichment of downregulated genes for the steroid metabolic process.

Table 4-1: Summary of GO enrichment results for upregulated DEGs

	Top Categories		FDR	#	ER
092	- Extracellular matrix organization	- GO:0030198	▪ 0.00	▪33	- 8.1774
	- Cell adhesion	- GO:0007155	▪ 0.00	▪56	- 3.6205
211	- Extracellular matrix organization	- GO:0030198	▪ 0.00	▪46	- 10.043
	- Cell adhesion	- GO:0007155	▪ 0.00	▪65	- 3.7025
A2	- Extracellular matrix organization	- GO:0001568	▪ 0.00	▪127	- 5.3250
	- Blood vessel development	- GO:0030198	▪ 0.00	▪160	- 3.6254
A7	- Extracellular matrix organization	- GO:0001568	▪ 2.77×10^{-5}	▪7	- 21.683
	- Blood vessel development	- GO:0030198	▪ 5.12×10^{-5}	▪8	- 13.391
E9	- Gliogenesis	- GO:0042063	▪ 3.64×10^{-3}	▪8	- 12.381
	- Glial cell differentiation	- GO:0010001	▪ 3.90×10^{-3}	▪7	- 14.519
G4	No categories showed significant enrichment				

#: number of genes overlapping
ER: Enrichment Ratio

Table 4-2: Summary of GO enrichment results for downregulated DEGs

	Top Categories		FDR	#	ER
092	- Steroid metabolic process	- GO:0008202	▪ $2.09\text{x}10^{-8}$	▪13	- 11.475
	- Cholesterol biosynthetic process	- GO:0006695	▪ $1.98\text{x}10^{-8}$	▪9	- 34.979
211	- Cholesterol biosynthesis process	- GO:0006695	▪ 1.03×10^{-3}	▪7	- 15.937
	- Cadherin signalling pathway	- P00012	▪ 1.33×10^{-3}	▪9	- 9.0288
	- Chemical synaptic transmission	- GO:0007268	▪ $8.38\text{x}10^{-3}$	▪16	- 3.6688
A2	- Chemical synaptic transmission	- GO:0007268	▪ 0.00	▪34	- 7.0931
	- Cell-cell signalling	- GO:0007268	▪ $4.78\text{x}10^{-13}$	▪44	- 4.0205
	- Neuronal system	- R-HAS-112316	▪ $2.35\text{x}10^{-10}$	▪21	- 8.0138

A7	- Cerebral cortex regionalization - Generation of neurons	- GO:0021796 - GO:0048699	▪ 2.86×10^{-2} ▪ 3.86×10^{-3}	▪2 ▪8	- 284.03 - 5.5576
E9	- Cell-cell signalling - Generation of neurons	- GO:0007267 - GO:0048699	▪ 2.56×10^{-5} ▪ 1.72×10^{-6}	▪33 ▪35	- 2.8967 - 3.2547
G4	- Cell-cell signalling - Generation of neurons	- GO:0007267 - GO:0048699	▪ 52×10^{-5} ▪ 2.84×10^{-6}	▪31 ▪33	- 2.9186 - 3.3028

#: number of genes overlapping
ER: Enrichment Ratio

4.2.6 Overlap between DEGs in the present study dataset and SFARI gene database

The Simon's Foundation for Autism Research Initiative (SFARI) have curated close to a thousand genes associated with autism, with variable evidence. It is based on a regularly updated scoring system to assess the evidence level for individual genes giving them a score from 1 (strongest) to 4 (weakest) based on peer-reviewed research papers [270]. In a quest to find whether mutations in *NRXN1* might have an effect on any of these genes, a screening of the high confidence genes (score= 1) listed in SFARI gene database (192 genes) with the DEGs from each of the aforementioned comparisons was done.

The results are listed in TABLE 4-3. A total of 12 SFARI high confidence genes have been reported to be differentially regulated in the 6 cell lines included in this study. In the 092NXF, *FOXP2* was the only gene overlapping with the SFARI high confidence

list, while 211NXM shared this gene and four more (*DHCR7, FOXG1, RELN and TBR1*). The A2 mutant line reported 6 high confidence genes to be differentially regulated including *NRXN1* while A7 reported none. The E9 and G4 shared 3 DEGs that were also listed in the SFARI list.

Table 4-3: Overlap between the high confidence SFARI genes and differentially expressed genes

	092	211	A2	G4	E9	A7
1	*FOXP2*	*FOXP2*	*DSCAM*	*RELN*	*RELN*	
2		*DHCR7*	*GABRB2*	*TCF7L2*	*TCF7L2*	
3		*FOXG1*	*GRIA2*	*RORB*	*RORB*	
4		*RELN*	*MEIS2*	*MEIS2*		
5		*TBR1*	*NRXN1*			
6			*TEK*			

4.3 Discussion

The aim of this chapter was to analyse the effect of the NRXN1 mutations in 2 different groups: NRXN1-Patient and NRXN1-mutant day21 induced neurons co-cultured with rat glial mixed cells to induce maturation. Following RNA-Seq of the induced neurons, PCA revealed an interesting behaviour of one of the NRXN1-mutant lines (A2), which did not seem to cluster or resemble any of the other mutant lines not the control line despite sharing the same genetic background of this line. Differential expression analysis reported 27 genes in the NRXN1

patient lines and only 8 genes in the NRXN1-mutant lines. Due to the small number of DEG in both groups, no significant enrichment in GO terms was reported.

Moreover, NRXN1 was not differentially expressed in either group so to query its expression, a read pile up was produced using bam-coverage, which revealed that the mutations in the patient and the mutant lines exhibited distinct patterns within each group. This called for an individual comparison of each cell line against the control and the results were very interesting.

Although, the DEGs identified from the 092NXF and 211NXM were different, 50% of the upregulated genes overlapped and comprised a significant number of genes from the collagen superfamily (092NXF: 8 genes) and (211NXM: 12 genes). Collagens are cardinal components of the extracellular matrix of the developing brain. They have been implicated in neurite outgrowth, synaptogenesis and the establishment of the architecture of the brain, making them fundamental elements in neural development [290]. Out of 29 collagens in the collagen superfamily, 27 were upregulated in the NRXN1-mutant line A2, , 12 in 211NXM and 8 in 092NXF, this was associated with an altered extracellular matrix and abnormal neural development and synaptic transmission, which is not surprising given the important role played by collagens in these processes. Upregulation of collagens have been reported in the prefrontal cortex of a valproic rat model of autism [291]. This results in aberrant neuronal migration and defective layering of the cortical areas as reported in some autism cases [292], [293]. Moreover, 56 and 65 downregulated genes in 092NXF and 211NXM, respectively, were highly enriched for cell adhesion, the biological process that describes the attachment of a

cell to extracellular matrix molecules. This was previously reported in NRXN1-knockdown neurons and autism prefrontal cortex [276], [294].

Interestingly, an upregulation of genes involved in gliogenesis and glial cell differentiation was reported in the mutant line with the largest deletion in NRXN1 (E9). This is consistent with previous studies that reported astroglial upregulation in ASD cerebral cortex [277] and bi-allelic deletion of NRXN1 was associated with higher expression of radial glia-like genes[180].

Furthermore, the downregulation of protocadherin genes was noteworthy in the patient lines. The cadherin superfamily of genes has been primarily isolated as Calcium dependent cell adhesion molecules [295]. They are also known to be predominantly expressed in the CNS, and previous studies have implicated them in neuronal circuitry, synapse junction and synaptic plasticity [295], [296]. Often referred to as 'synaptic molecules', cadherins are involved in intracellular signalling and have been associated with many neuropsychiatric conditions such as autism and schizophrenia [297]. Dysregulation of cadherins have been associated with altered calcium ion binding, cell-cell signalling and synaptic transmission, pathways that are often downregulated in autism prefrontal cortex [277], [294]. GO analysis of the downregulated genes of the mutant cell lines was associated with neural development and synaptic function terms (Chemical synaptic transmission, cell-cell signalling and generation of neurons), which complements other transcriptomic analysis of other studies using NRXN1-mutated cellular models [100], [129], [180], [195].

The DEGs in 092NXF patient line included nicotinic acetylcholine receptors (*CHRNA3, CHRNA6* and *CHRNB4*), members of a superfamily of ligand-gated ion channels, mediating fast signal transmission at the synapse. These genes are related to the Highly calcium permeable nicotinic acetylcholine receptors pathway. The *CHRNA3* and *CHRNB4* bind to acetylcholine, this results in a conformation change in the *Ach-R* and affects all subunits and opens the ion-conduction channel across the plasma membrane. These genes have been associated with increased susceptibility to nicotine addiction and lung cancer [298]–[300].

This was not the first time *NRXN1* was linked with Nicotine addiction and polymorphisms in *CHRNA3* and *CHRNB4* [301]. A study in 2008 reported significant association with nicotine dependence in two independent groups [300]. *NGN2*-induced neurons reported a glutamatergic excitatory neuronal fate, with a robust *AMPA*-R expression and a limited expression of *NMDAR*s [232]; as the *NMDAR*s are expressed at more mature neuronal stages.

Single-cell RNA-Seq investigation of induced BOB-NGN2 neurons (Day21) in our lab (unpublished) reported the generation of three population of neurons: a glutamatergic neuronal population, another neuronal subpopulation identified is a *bonafide* cholinergic population with strong expression of *PHOX2B, NKX6-1* and *NKX6-2* cholinergic markers enriched in motor and spinal cord neurons. And a third neuronal population identified was a hybrid population expressing both glutamatergic and cholinergic markers [225]. The paired-like homeobox 2B (*PHOXB2*) gene is involved in the development of major noradrenergic neuron populations, and among its related pathways are neural crest differentiation [302], [303]. Alterations in this gene are associated with Congenital central hypoventilation

syndrome and neuroblastoma. CCHS is often associated with neurocognitive impairments possibly due to repeated hypoxemia and hypercarbia [304].

Additionally, the genes involved in the cholesterol biosynthetic process (*DHCR7*, *DHCR24, LSS, MVD, ACAT2, HMGCS1 and INSIG1*) were downregulated in the 092NXF and 211NXM patient lines. Deficiencies in *DHCR7*; the enzyme responsible for converting 7-dehydrocholesterol to cholesterol, result in Smith-Lemli-Opitz syndrome (SLOS). Close to 75% of children with SLOS were also diagnosed with autism as reported in a 2006 study suggesting a link between cholesterol metabolism and autism [305]. In addition to SLOS, other disorders of sterol metabolism or homeostasis associated with autism, for instance a study investigated the type and incidence of other sterol disorders in blood samples from a cohort of subjects with autism from multiplex families found that up to 20% of children have substantial hypocholesterolaemia[306].

In an attempt to dissect autism clinical complexity and heterogeneity, a study in 2020 integrated large data sets of familial whole-exome sequences combined and neurodevelopmental expression patterns with electronic health records (EHRs) and healthcare claims. In their study, functional enrichment analysis revealed lipid regulation, a previously unrecognized molecular convergence. Moreover, they assessed autism and dyslipidemia comorbidity using health claims and found a significant enrichment of dyslipidemia in individuals with autism [307]. Cholesterol, a steroid hormone precursor, is essential for neuronal development and dysregulation of cholesterol related genes have been

implicated in synaptic dysfunction leading to neurological conditions, including autism [308], [309].

The same set of genes involved in the cholesterol biosynthetic process were also enriched for the steroid metabolic process in the 092NXF patient line in addition to the downregulation of the *CYP26A1* gene; a member of the cytochrome *P450* superfamily involved in the synthesis of cholesterol, steroids and other lipids, and the *MSMO1* gene, a *Sterol-C4-mehtyl oxidase-like* protein and associated with microcephaly and developmental delay [310], [311].

The high confidence gene from the SFARI gene database were selected and queried for an overlap with the present study's dataset. Twelve high confidence autism candidate genes were identified in the investigated data set, 50% of those genes were identified in A2 mutant line alone including *GRIA2, GABRB2, DSCAM, MEIS2, NRXN1* and *TEK* while none was identified in A7. *FOXP2* and *RELN* have been previously differentially regulated in *NRXN1α* mutated neurons[129]. This is just an overlap and not an enrichment as more analysis are needed to identify an enrichment. The overlap may help to point towards what biological pathways to test in the future, that may have relevance for autism (hypothesis generating), but it does not tell identify mutations/deletion in NRXN1 that drive an 'autistic' phenotype. Therefore, NRXN1 is unlikely to be regulating the function/expression of genes associated with risk for autism - but might have effects on other biological pathways.

The perturbations observed in the *NRXN1*-mutant lines reflect a direct effect downstream the mutations induced in the *NRXN1* gene and highlight the diverse effect of variants of this gene.

This, however, cannot be exclusively suggested regarding the patient lines, as the perturbations reported can also result from independent contribution of autism risk variants.

In conclusion, transcriptomic analysis of individual cell lines revealed that although the mutations have affected the transcription of a different set of genes, their roles and involvement converged into shared pathways which might result in them exhibiting similar phenotypes in functional and morphological assays.

Chapter 5

Cell adhesion proteins and Neuritogenesis

Summary

*NRXN*s and *NLGN*s are cell adhesion molecules involved in synapse formation and neurite outgrowth. In this chapter the focus is on understanding 1) whether the expression of nrnx1 isoforms and nrnxn1 binding partners are altered in the isogenic *NRXN*1-mutant and/or the *NRXN*1-patient lines describes in Chapter 3; and 2) whether the initial development of neuronal morphology differs from control lines. Two major comparisons were carried throughout this chapter: The first comparison is between *NRXN*1-mutant versus isogenic control lines and the second between patient and control lines.

5 Cell adhesion proteins and Neuritogenesis

5.1 Introduction

*NRXN*1 is one of the largest and most complex genes in the human genome as it generates thousands of isoforms [105]. Many *NRXN*1 mutations have been implicated in neuropsychiatric conditions such as schizophrenia and autism [312]. Recent studies investigated the role of *NRXN*1 in synapse function. These studies have broadly taken two different approaches – the introduction of mutations by gene editing approaches or the use of iPSC generated from patients with mutations in *NRXN*1 and with a diagnosis of neuropsychiatric disorders. These studies have focused primarily on the impact of *NRXN*1 mutations on neuronal functions. Pak et al showed that conditional *NRXN*1 mutations impair neurotransmitter release in human neurons[195]. Similarly, *NRXN*1 knockdown resulted in alteration in synaptic adhesion and neuron differentiation pathways in neural stem cell models [276].

Interestingly, Flaherty et al revealed that large exonic mutations in the 3' end of *NRXN*1α did not just alter the expression of *NRXN*1, but also altered the profile of *NRXN*1 isoform profile; a 50% loss of wild type isoform and the expression of 31 novel mutant isoforms. The study also reported impaired neuronal activity and delayed maturation in human neurons as a result of those mutations [129]. However, a number of questions arise from these studies, firstly, whether single point mutations and or deletions that encompass the *NRXN*1α also impact the profile of *NRXN*1-pan and

NRXN1ß isoforms is unclear. Additionally, it is not known how and whether changes in *NRXN*1 expression levels occur in combination with alterations in the expression of *NRXN*1 binding partners.

Neurexins function by binding to other cell-adhesion molecules in the post-synapse and forming ligands with post-synaptic cell adhesion molecules such as neuroligins [106], [211]. There is no documented evidence on the effect of mutations in *NRXN*1 on other members of its family or its ligands. Pak et al 2015 introduced mutations in the *NRXN*1 gene in human induced neurons. They found no significant differences in mRNA levels of synaptic genes and *NRXN* ligands including *CASK*. They however reported a dramatic increase in *CASK* protein levels in the mutated lines when compared to controls. This is very intriguing, because the mRNA levels of *CASK* were not significantly higher in the mutated lines indicating that *CASK* protein synthesis is affected by *NRXN1* expression levels. It has been suggested that the regulation of *CASK* is normally regulated by *NRXN1* to control neurotransmitter release [195].

While previous studies have focused on the role of *NRXN*1 in synaptic function, it is unclear if *NRXN*1 plays a role in shaping the overall morphology of neurons, by regulating neurite outgrowth. One of the earliest studies suggested the synaptotropic hypothesis, which states that axonal growth and dendritic branching are controlled and stabilised by nascent synapses, guiding the growth of neurite in synaptic partners rich areas [189]. Previous work in Xenopus

aimed revealed that β neurexins and *NLGN*1 contribute to dendritic arbour development and disruptions in these molecules result in destabilising filopodia, reducing synaptic density leading to a reduction in arbor complexity as the neurons mature [190]. Moreover, homozygous mutant *NRXN* and *NLGN*1 null mutation in drosophila showed significant defects in arbour length and morphology[189]. Collectively, these result support and expand on the synaptotropic hypothesis which incorporate the role of cell adhesion molecules in this process and even propose calling these molecules neuritic adhesion complex (NAC) to highlight their role in neurite outgrowth prior to synapse formation[189], [190].

The role of these complexes in neuritogenesis has also been suggested in rodents. A significant increase in neurite length was reported as a result of transfecting *NRXN*1α siRNA to rat hippocampal neurons [188]. Another study implicated *NLGN*1 in the regulation of neurite outgrowth through its interaction with *NRXN*1β and on activation of fibroblast growth factor receptor-1 in rat hippocampal neurons [137]. Recently studies in neuroblastoma cell line confirmed the role of *NLGN*1 in neurite outgrowth through protein-protein interaction involving he AChE-like domain and PDZ motif. Since *NLGN*1 is not localised at axonal growth cones, and cannot promote neurite outgrowth in the absence of AChE-like and PDZ domain, it is suggested that this novel role is mediated through binding to other cell adhesion molecules such as neurexins [191].

Disruption of early neurodevelopmental process such as Neuritogenesis have been reported in autism in various studies [102]. Together with evidence on the potential involvement of *NRXN-NLGN* complex in neurite outgrowth we sought to investigate the effect of the mutations presented in the *NRXN*1-mutant lines as well as the *NRXN*1-patient lines on this process.

In this chapter, a time-course gene expression characterization is performed to study the effect of *NRXN*1 mutations in *NRXN*1-mutant and patient lines on *NRXN*1-isoforms, *NRXN*s and their main binding partners. Finally, neurite outgrowth analysis on early developing neurons was performed to investigate the potential involvement of *NRXN*1 mutations on neuronal morphology.

5.2 Results

5.2.1 Investigating *NRXN*1 isoform expression levels in *NRXN*1-mutant and *NRXN*1-patient lines across time

An emerging theme from recent studies suggests that mutations in *NRXN*1 impacts the profile of NRXN1 isoforms [129]. To understand whether *NRXN*1 and its predominant isoforms were altered in our CRISPR-edited and patient-derived line, the *NRXN*1-Taqman gene expression assay (Applied Biosystems, CA, USA) was used. Expression of pan-*NRXN*1, *NRXN*1α and *NRXN*1β isoforms were assessed at 0, 7-, 14 and 21-days post neuronal induction.

A 2way ANOVA and Tukey's test for multiple comparisons was used to compare *NRXN*1 isoform expression between either *NRXN*1-mutant or *NRXN*1-patient lines specifically to *NRXN*1-isoform expression at Day 21 (last timepoint) in a control line. Additional analysis was performed to compare the expression levels between isoforms within each line separately.

5.2.1.1 A significant reduction in overall expression of NRXN1 α in NRXN1-mutant lines

The guide RNA was designed to target the first shared exon between *NRXN*1α and *NRXN*1β isoforms, as previously described in Chapter 3, resulting in the generation of four mutant lines with different mutations in Exon 19. The mRNA levels of pan-*NRXN*1 were detected in the wild-type and mutant lines at 1, 2- and 3-weeks post-induction although the wildtype was consistently expressed at higher levels than the mutant lines, but this only reached statistical significance for G4 (p=0.0361) (Figure 5-1A). *NRXN*1α expression showed a gradual increase throughout differentiation as reported in previous studies using cellular based models [276]. The *NRXN*1α overall expression of the wild-type was significantly higher when compared to the *NRXN*1-mutant lines; A2 (p=0.0275), A7 (p=0.0051), G4 (p=0.0054) and E9 (p=0.0287) (Figure 5-1B).

*NRXN*1β isoform showed a marked lower expression in the *NRXN*1-mutant lines when compared to the WT; only A2 and G4 mutant lines were reported to be statistically significant (2way ANOVA-Tukey's test, p=0.0292 and 0.0055, respectively). This

isoform shares a similar expression pattern to that of *NRXN*1, where the mRNA levels peak at Day 7 followed by a subtle decrease in the following weeks. Looking at a single timepoint (Day 21), the expression of *NRXN*1α was higher than all the mutant lines, but 2way ANOVA (Tukey's post hoc test, $p<0.05$) reported statistical significance for A7 only ($p=0.0474$). No statistically significant differences were reported at Day 21 for *NRXN*1 nor *NRXN*1β. Additional analysis showed that *NRXN*1β was significantly higher than the *NRXN*1α isoform in the wild-type (BOB-NGN2) (Paired T-test $p=0.0370$) (Figure 5-2D).

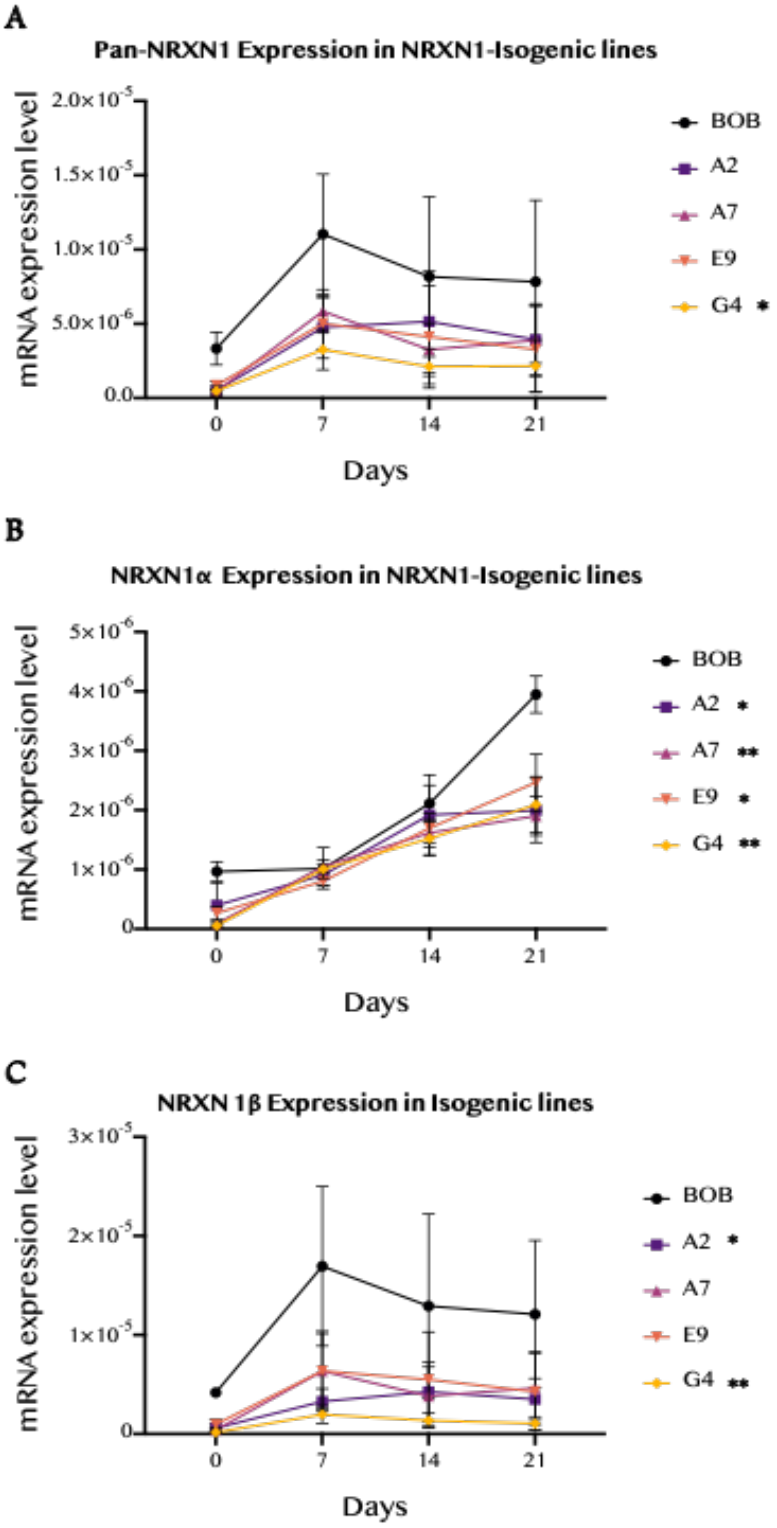


Figure 5-1: Temporal Expression of NRXN1 isoforms in NRXN1-isogenic iNs; A))NRXN1- B)NRXN1a and C)NRXN1β gene expression at Days 0, 7, 14 and 21. The graphs represent the mRNA absolute values normalised to an internal 18S reference gene (Y axis) and days post induction on the X axis. The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

5.2.1.2 *Varied expression patterns of NRXN1 and NRXN1β within the NRXN1-Patient lines*

Both patient-derived iPSC lines have mutations in the α isoform of the *NRXN*1 gene as described in Chapter 3; a point mutation (211_NXM-NGN2) and a 200kb deletion (092_NXF-NGN2 line). The patient lines showed a marked difference in expression; while the 200kb deletion in (*092_NXF*) reduced the overall expression, the iPSC with the point mutation in *211_NXM* showed a marked increase in expression. While the control line shows a peak in expression at Day 7 and then a decrease in expression was reported at Day 14 and 21 for both pan-*NRXN*1 and *NRXN*1β, the patient NXM_211, shows a peak in pan-*NRXN*1 gene expression at Day 14 and a gradual increase in *NRXN*1β until Day 21 (Figure4-2A and C, respectively). Moreover, *NRXN*1-Patient (092_NXF-NGN2) neurons shows a significant overall reduction in pan-*NRXN*1 and *NRXN*1β expression levels when compared to the 211_NXM-NGN2 patient line (2way ANOVA and Tukey's multiple test: p=0.0426 and p=0.0304, respectively). Moreover, no significant changes were reported in *NRXN*1α expression levels between the three cell lines.

Additional analysis was carried out to investigate the difference in isoform expression levels; overall *NRXN*1α was significantly lower than *NRXN*1β mRNA levels in BOB-NGN2, 092_NXF-NGN2 and 211_NXM-NGN2; (Paired T-test: p=0.0370, p=0.0220 and p=0.0350, respectively) as shown in Figure 5-2(D, E and F).

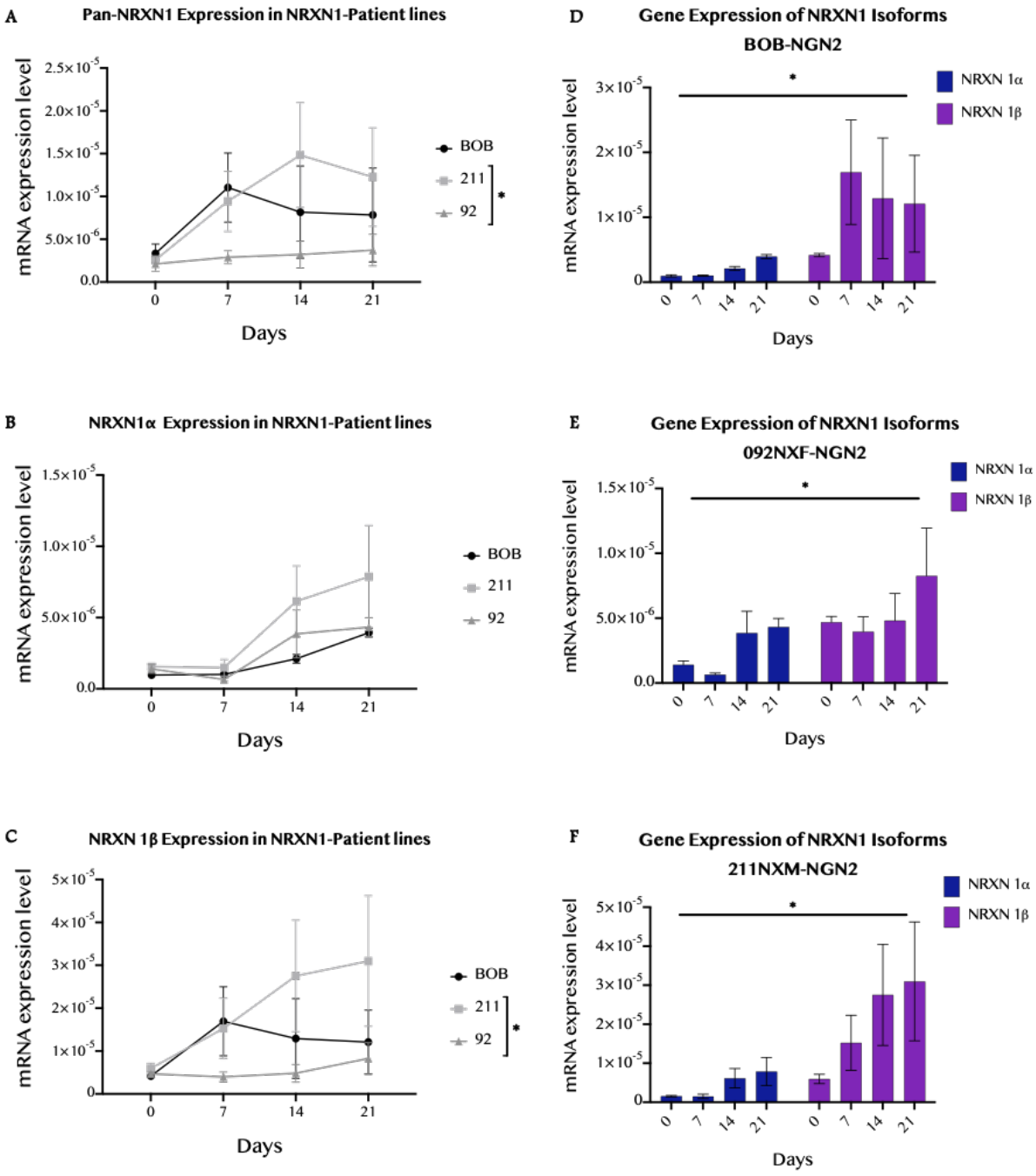


Figure 5-2:Temporal Expression in NRXN1-Patient and control lines expression at Days 0, 7, 14 and 21 for A)NRXN1- B)NRXN1α and C)NRXN1β gene. The plots display the mean with s.e.m, p<0.05 by two-way ANOVA, and Tukey's test for multiple comparisons. Gene expression levels of NRXN1 Isoforms of D) BOB-NGN2, E)092_NXF-NGN2 and F)211_NXM- NGN2, The plots display the mean with s.e.m, p<0.05, Paired T-test. All graphs represent the mRNA absolute values normalised to an internal 18S reference gene (Y axis) and days post induction on the X axis, n=3.

5.2.2 Investigating NRXN1-associated genes' expression in NRXN1-mutant and NRXN1-patient lines across time

*NRXN*s and *NLGN*s act as ligands and bind to other pre-and post-synaptic MAGUK proteins such as *CASK* to form trans-synaptic complexes. Mutations in *NRXN*1 might have effects on their associated proteins. In this experiment, *NRXN*-associated genes' expression across four time points was measured using gene-specific primers and SYBR-green assay to compare expression levels of *NRXN*1-mutant and *NRXN*-patient lines to control. Moreover, statistical analysis was performed to show the different expression levels of genes within the same family in each line separately.

5.2.2.1 NRXN1 mutant line (A7) shows significantly higher overall CASK gene expression levels than the wildtype and other NRXN1-mutant lines

Looking into the effect of the *NRXN*1 mutation in the isogenic lines on other *NRXN*s (*NRXN*2 & 3) and *NRXN*-ligands (*NLGN*1, 2 and 3); very subtle changes were observed. *NRXN*2 was consistently highly expressed with gradual increase across time, while *NRXN*3 remained lowly expressed with a slight increase in weeks 2 and 3. No differences in expression was reported across the wild-type and the mutant lines, nor within the mutant line. *CASK*, which binds directly to *NRXN*1 and connects it to the exocytotic machinery showed similar expression pattern across time, in all the lines, the highest expression was reported at Day 7 post induction with a slight reduction at the following weeks.

When comparing the lines, it was notable that the overall expression of *CASK* was significantly higher in A7 when compared to the wildtype (BOB-NGN2, p=0.0146) and other mutant lines (A2-p=0.0223, E9-p=0.0466 and G4-p=0.0315). This marked difference of *CASK* expression in the A7 mutant line still reported statistical significance when compared to (BOB-p=0.0487, E9-p=0.0468 and G4-p=0.0428) when comparing mRNA levels at a single timepoint (Day 21).

In summary, mutations in *NRXN*1 were not associated with differences in overall expression of other *NRXN*s or *NLGN*s when comparing the *NRXN*1-mutant lines to the wild-type, which was also shown in other studies [195], [276]. *CASK* is the only gene investigated that reported a significant increase in mRNA levels in one of the mutant lines (A7); a heterozygous line with one bp insertion. This is the first time a difference of *CASK* expression level was reported as a result of a mutation in *NRXN*1 although some studies suggested a regulatory effect of *NRXN*1 on *CASK* protein levels [195]. The mutations show no effect on the genes' expression pattern across time.

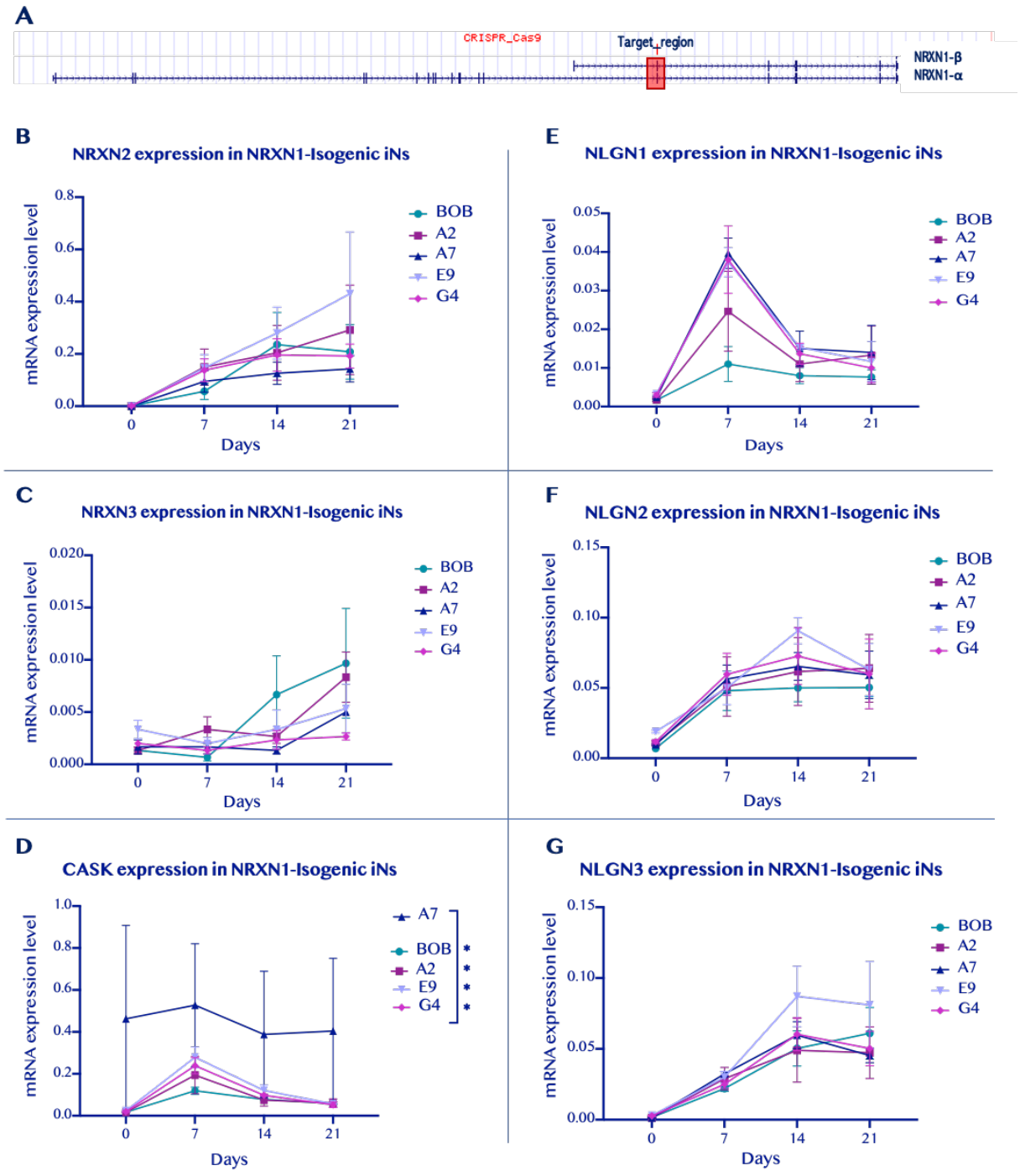


Figure 5-3 Temporal gene expression results of NRXNs, NLGNs and CASK;A)NRXN1 gene structure showing gRNA target region. Gene expression Q-PCR results at 0, 7-, 14- and 21-days post-neural induction for: B)NRXN2, C)NRXN3, D)CASK, E)NLGN1, F)NLGN2 and G)NLGN3 in Control and NRXN1-mutant lines The graphs represent the mRNA absolute values normalised to an internal 18S reference gene (Y axis) and days post induction on the X axis. The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

5.2.2.2 *Gene expression of NRXN-associated genes showed no significant differences in NRXN1-Patient lines when compared to the BOB-NGN2 control line*

NRXN2 mRNA levels increase steadily from Dayo to Day21 in both of the *NRXN1*-Patient lines (092_NXF and 211_NXM). This is not the same pattern seen in the BOB-NGN2, where the mRNA levels reach a peak at Day 14 and then decrease at Day21 as shown in Figure 5-4B. The *NRXN3* gene expression in the 092_NXF and the 211_NXM Patient lines shows higher expression at Day21 when compared to that of the control line (Figure 5-4B). The *NRXN1*-Patient line 092_NXF consistently shows higher *CASK* mRNA levels than 211_NXM and BOB-NGN2 across all time point (Figure 5-4D). *NLGN1* gene expression peaks one week post neuronal induction in the control and one of the *NRXN1*-Patient lines (092_NXF). This is however not the case for the other patient line (211_NXM), which shows no peak and continuous increase in mRNA levels until Day21 as seen in Figure 5-4E.

While monitoring the expression of *NLGN2* in the iNs derived from the patient and the control lines, three distinct patterns were observed (Figure 5-4F). While BOB-NGN2 peaks at Day 7, 092_NXF peaks at Day 14 and 211_NXM shows a continuous increase in expression across the weeks period of induction. *NLGN3* gene expression shows no differences across the cell lines included in this experiment and doesn't seem to be affected by the mutations in the *NRXN1*-patient lines (Figure 5-4G).

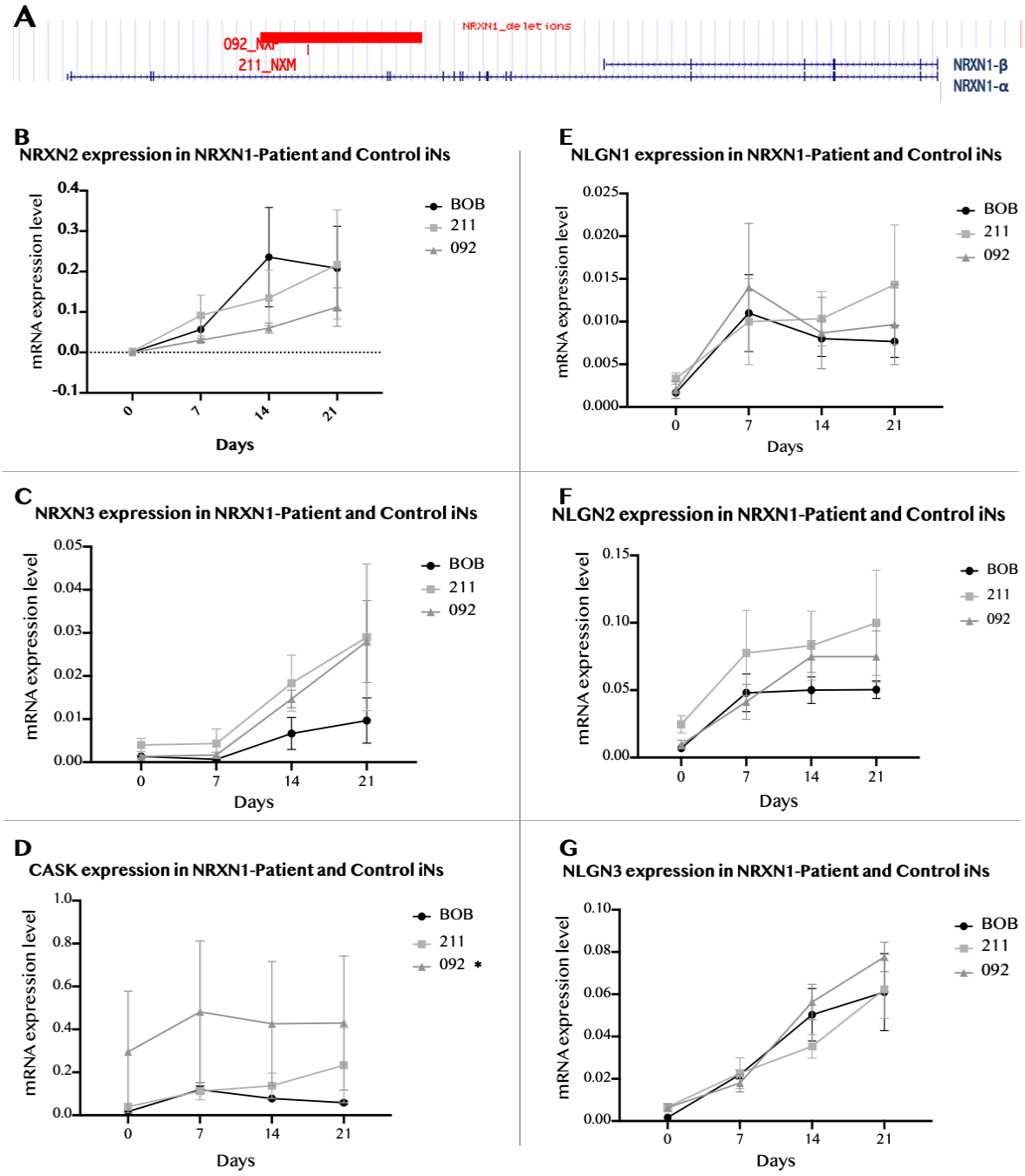


Figure 5-4: Temporal gene expression results of NRXNs, NLGNs and CASK; Gene expression Q-PCR results at 0, 7-, 14- and 21-days post-neural induction for synaptic genes: A)NRXN2, B)NRXN3, C)CASK, D)NLGN1, E)NLGN2 and F)NLGN3 in Control and Patient lines. The graphs represent the mRNA absolute values normalised to an internal 18S reference gene (Y axis) and days post induction on the X axis. The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

5.2.2.3 *NRXN2 showed a consistently higher expression that NRXN3 across time in all the cell lines*

The overall expression of *NRXN2* was significantly higher than that of *NRXN2* across the *NRXN*-Isogenic lines (BOB-NGN2 p=0.0096, A2 p=0.0003, A7 p=0.0003, G4 p=0.0009 and E9 p=0.0052) and the *NRXN1*-patient lines (*092_NXF-NGN2* p=0.0066 and *211_NXM-NGN2* p=0.0272) (Figure 5-5). These results are in agreement with what was found by Harkin et al, which report that *NRXN2* is consistently highly expressed in post-mortem tissue (PMT) of fetal brain at 8-12 post conceptional weeks (pcw) across four different brain regions while *NRXN3* was lowly expressed and shows moderate expression at later stages (11-12 pcw).

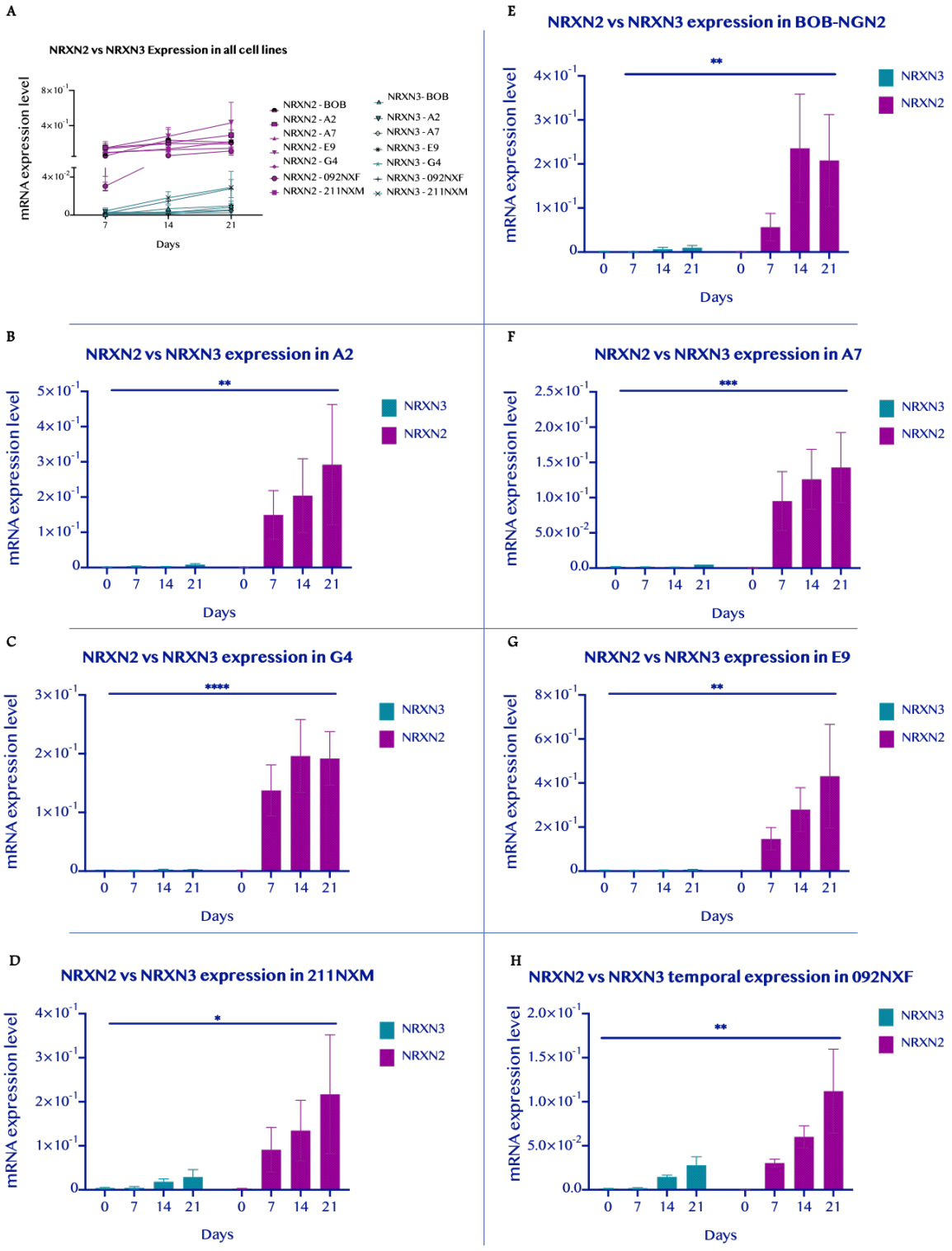


Figure 5-5 NRXN2 and NRXN3 expression levels in NRXN1-mutant and NRXN1-patient line across time. A) Collective representation of NRXN2and 3 expression including all cell lines, B) A2, C)G4, D) 211_NXM, E)BOB-NGN2, F) A7, G) E9 amd H)092_NXF. The graphs represent the mRNA absolute values normalised to an internal 18S reference gene (Y axis) and days post induction on the X axis. The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

5.2.2.4 *NLGN1 showed the lowest overall expression levels in the NLGN family across time in all cell lines*

The expression of *NLGN*1 was generally low and did not increase across time, instead the mRNA level peaks at 7 days post induction followed by a slight decrease in expression while *NLGN*2 and *NLGN*3 highest expression levels were reported at Day14 post induction. Moreover, the overall expression of *NLGN*1 reported a significantly lower levels when compared to *NLGN*2(BOB-p<0.001, A2-p=0.0098, A7-p p<0.0001, G4-p=0.0009, E9-p=0.0009) and *NLGN*3 (BOB-p=0.0005, A7-p=0.042 and E9-p=0.0037) in the isogenic lines (Figure 5-6).

Similar results were found when comparing the three *NLGN* genes in the *092_NXF*-NGN2 line, *NLGN*1 was significantly lower than *NLGN*2 (p<0.0001) and *NLGN*3 (p=0.0001) (Figure 5-6D). This is in agreement with other studies, which report that *NLGN*1 is consistently lowly expressed in fetal PMT (8-12 pcw) [313] and neural stem cells (week 0-4)[276]. Interestingly the mRNA levels of *NLGN*2 were significantly higher than *NLGN*1 (p=0.0001) and *NLGN*3 (p=0.0094) in *211_NXM*-NGN2 patient line (Figure 5-6G).

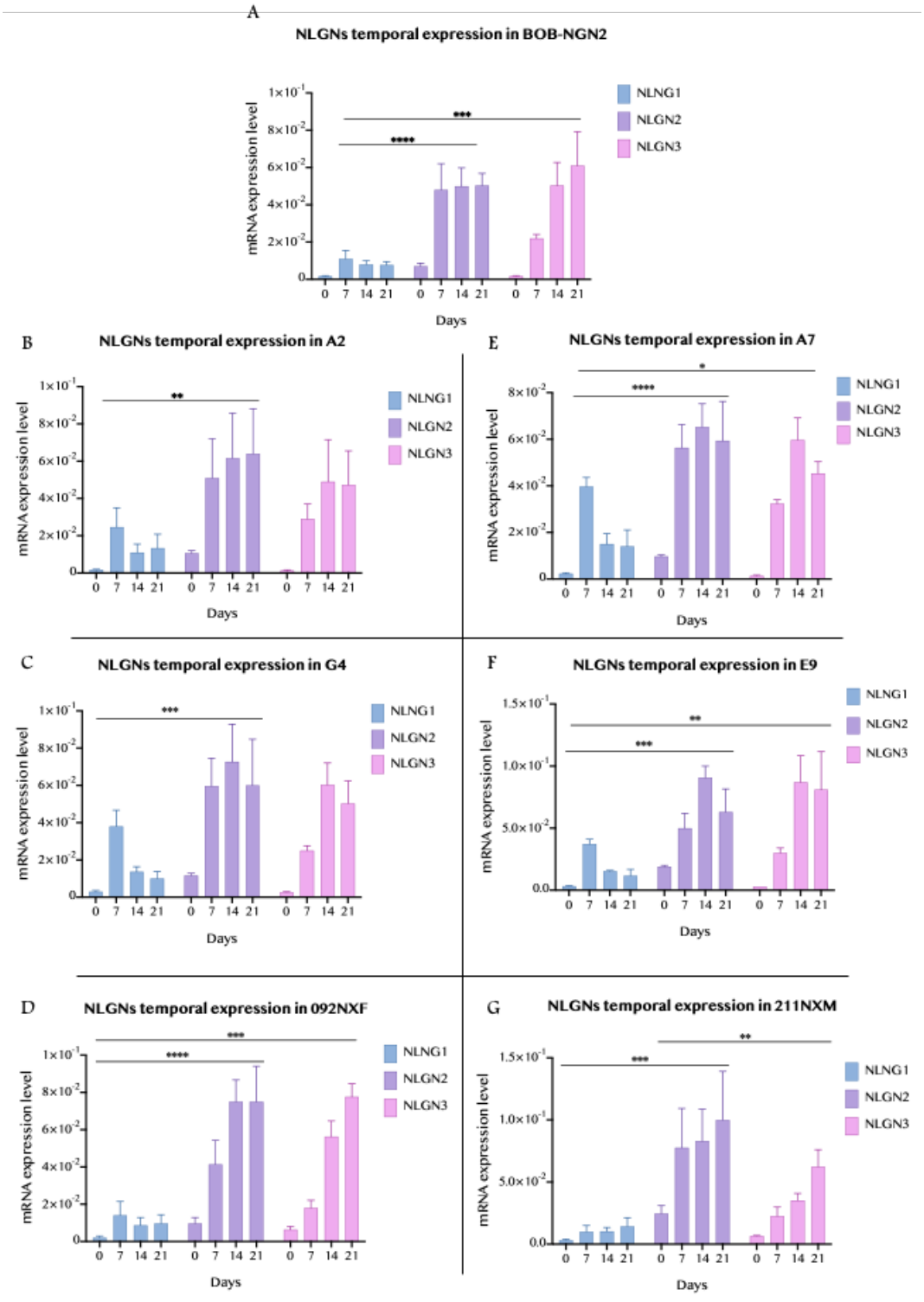


Figure 5-6: NLGN1, NLGN2 and NLGN3 expression levels in NRXN1-mutant and NRXN1-patient line across time. A) BOB-NGN2, B)A2, C)G4, D) 092_NXF, E)A7, F)E9 and G)092_NXF. The graphs represent the mRNA absolute values normalised to an internal 18S reference gene (Y axis) and days post induction on the X axis. The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

To summarize, the difference in expression across genes was highly significant, *NRXN2* reported a consistently higher expression than the other genes involved in this experiment. *CASK*, *NLGN2* and *NLGN3* genes were also highly expressed in all the cell lines and showed a marked higher expression than *NLGN1* and *NRXN3*.

5.2.3 Exploring the effect of *NRXN1* mutations on neurite outgrowth in *NRXN1*-mutant and *NRXN1*-patient neurons.

In this experiment, the NRXN1-mutant lines as well as the NRXN1-Patient lines were used as models to examine the possible effect of these mutations on neurite outgrowth in comparison to control.

The neurite outgrowth assay was used to identify the neurites and compute three major parameters: the neurite length, the neurite count and branch point count per neuron. Neurons were fixed at Day 6 and stained with *DAPI* and *MAP2* as a morphological marker to tag neurites (Figure 5-7). In Chapter 3, a molecular a temporal expression of neurons derived from Control, NRXN1-patient and NRXN1-mutant lines was conducted at day0, 7, 14 and 21 post induction, showing a comparable levels of pan neuronal markers including MAP2 and GRIA4, indicating similar levels of differentiation state. A QQ plot for the aforementioned parameters was created to confirm the normal distribution of the values plotted as shown in Figure 5-7B, C and D. Five technical replicates and three biological replicates from each cell line was used. To remove outliers, ROUT (Q=1%) analysis was used, and the

statistical significance was calculated using Kruskal-wallis and Dunn's multiple comparisons test ($p<0.05$).

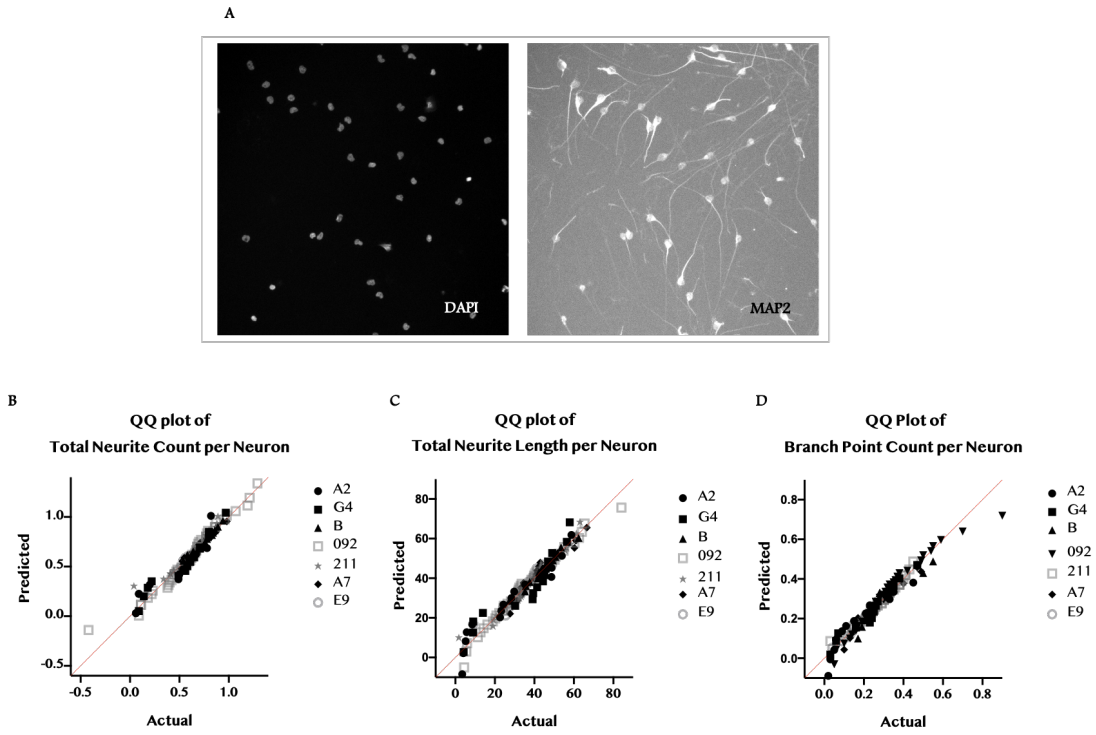


Figure 5-7: Neurite outgrowth assay A) Immunofluorescent staining of iN with MAP2 (morphological markers) and DAPI (nucleus), B) QQ plots of Total Neurite count/neuron, C) QQ plot of total neurite length/neuron and D) QQ plots of Branch point count/neuron of all cell lines included in this experiment

5.2.3.1 NRXN1-mutant lines show aberrant neurite outgrowth

Three out of four mutant lines were significantly affected in at least one of the three parameters tested (Figure 5-8). The total neurite length per neuron was significantly reduced in A2 ($p=0.0057$) and E9 (0.0067) (Figure 5-9B). These two lines also showed a marked decrease in branch point count per neuron (A2, $p=0.0236$) and (E9, $p=0.003$) when compared to the wild-type (Figure 5-9C).

Moreover, the total neurite count per neuron in A2 (p=0.0029), E9 (p=0.0038) and G4 (p=0.0387) was significantly lower than that of the control, while A7 the fourth mutant line was not significantly affected. Given the potential role of NRXN1 and NLGN1 in neurite outgrowth, a deeper look at their expression levels at an early timepoint was essential. We found that while the NRXN1 mRNA levels in the NRXN1 mutant lines was lower than the wild type at Da7, the NLGN1 expression level was higher (Figure4-8E). These differences did not reach statistical significance but show a potential association between reduced NRXN1 levels and elevated NLGN1 levels in neurogenesis.

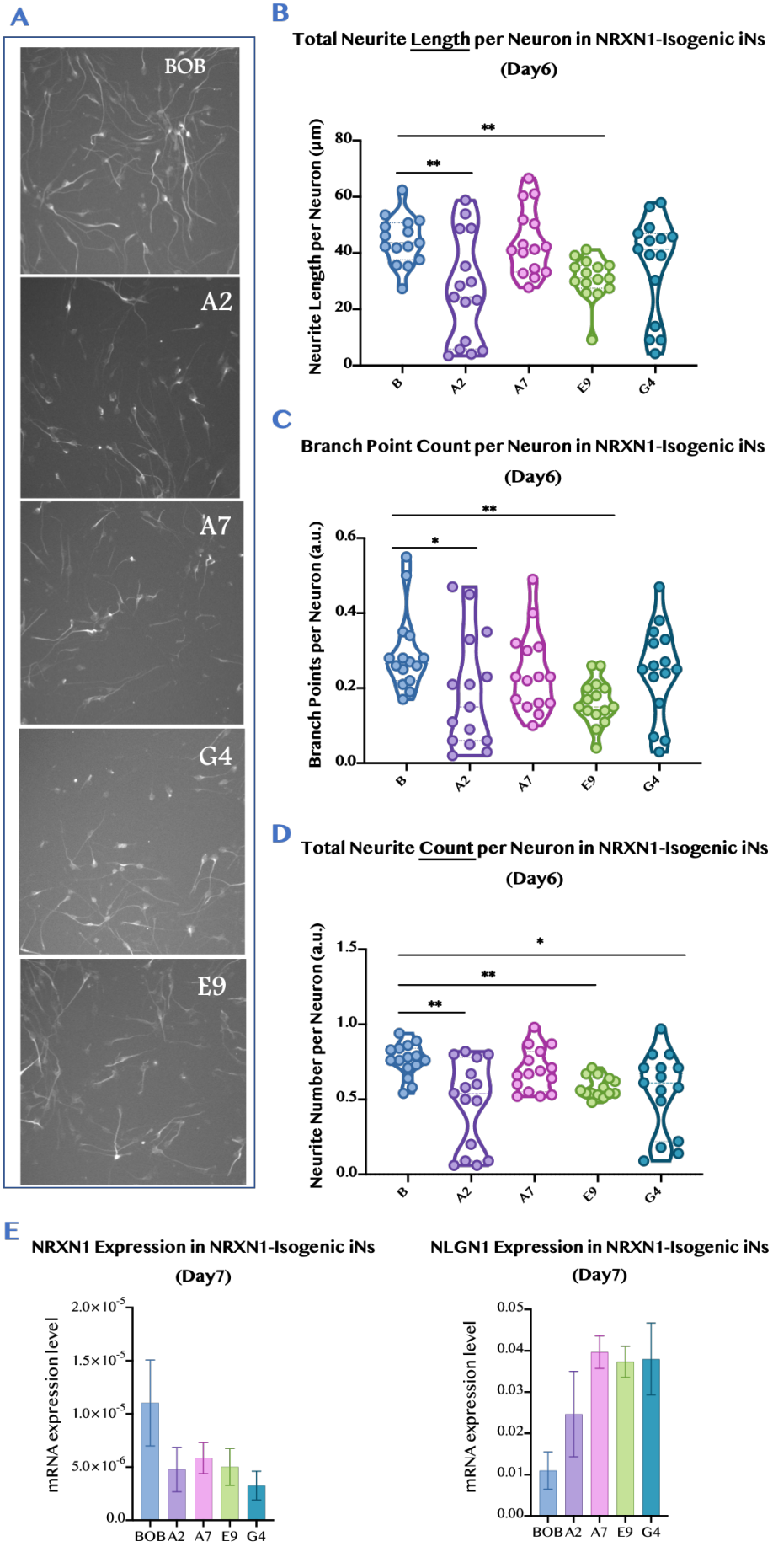
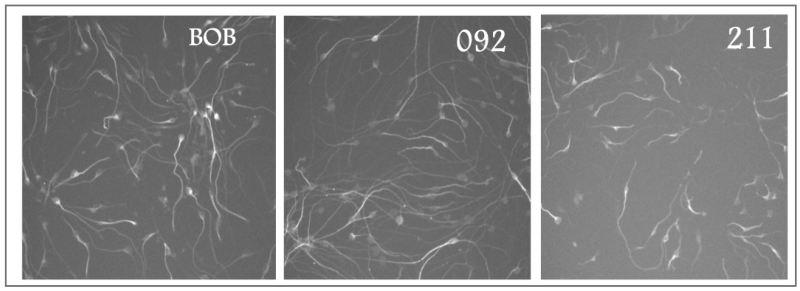

A
BOB
A2
A7
G4
E9
B
Total Neurite Length per Neuron in NRXN1-Isogenic iNs
(Day6)
Neurite Length per Neuron (µm)
0
20
40
60
80
**
**
B
A2
A7
E9
G4
C
Branch Point Count per Neuron in NRXN1-Isogenic iNs
(Day6)
Branch Points per Neuron (a.u.)
0.0
0.2
0.4
0.6
**
*
B
A2
A7
E9
G4
D
Total Neurite Count per Neuron in NRXN1-Isogenic iNs
(Day6)
Neurite Number per Neuron (a.u.)
0.0
0.5
1.0
1.5
*
**
**
B
A2
A7
E9
G4
E
NRXN1 Expression in NRXN1-Isogenic iNs
(Day7)
mRNA expression level
0.0
5.0×10-6
1.0×10-5
1.5×10-5
2.0×10-5
BOB A2 A7 E9 G4
NLGN1 Expression in NRXN1-Isogenic iNs
(Day7)
mRNA expression level
0.00
0.01
0.02
0.03
0.04
0.05
BOB A2 A7 E9 G4

Figure 5-8: Neurite outgrowth assay for iN-derived from Control and NRXN1-isogenic lines, A) Bright-field images of iNs at Day6, B)Violin plots of Total Neurite length/neuron , C) Violin plots of Branch point count/neuron and D) Violin plot of total neurite count/neuron in Control and NRXN1-mutant lines. The plots display the mean with s.e.m, P<0.05 by Kruskal-Wallis and Dunn's multiple comparisons test, E) mRNA expression levels of NRXN1 (left) and NLGN1 (right) in induced neurons Day7 using, Taq-man and SYBR green assays, respectively. The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

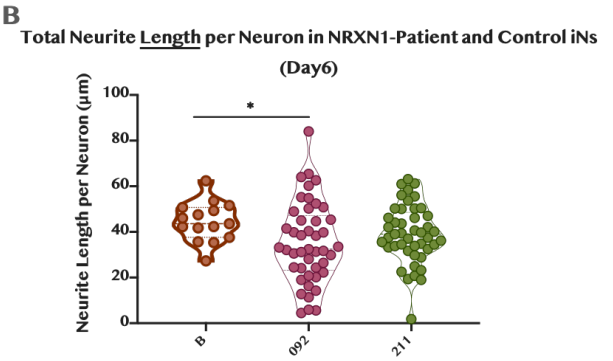
5.2.3.2 092_NXF-NGN2 patient line show lower total neurite count per neuron when compared to control

The NRXN1-patient lines do not seem to share the same neural development properties (Figure 5-9). While the 092_NXF patient line reported a significant decrease in neurite length (p=0.0273) when compared to the neurotypical control, the 211_NXM showed no differences (Figure 5-9B). Similarly, the total neurite count was significantly reduced in 092_NXF (p=0.0083), but not in 211_NXM, when compared to the control BOB-NGN2 line (Figure 5-9C). It is also important to note that similar to the NRXN1-mutant lines, the 092_NXF show reduced NRXN1 and increased NLGN1 expression at Day 7 (Figure 5-9E). A difference although not significant, was not observed in the 211_NXM line. The two patient lines showed no significant differences in the branch point count per neuron when compared to the control (Figure 5-9D).

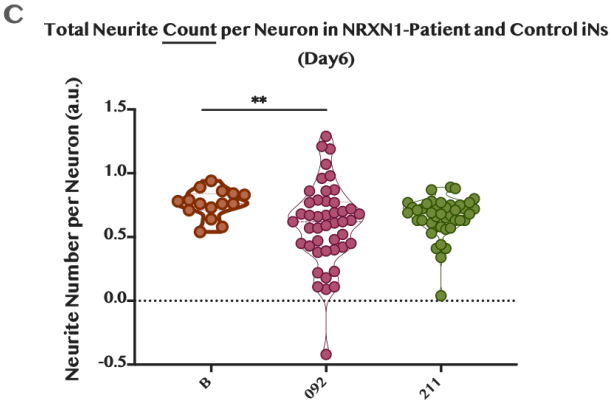
A


BOB
092
211

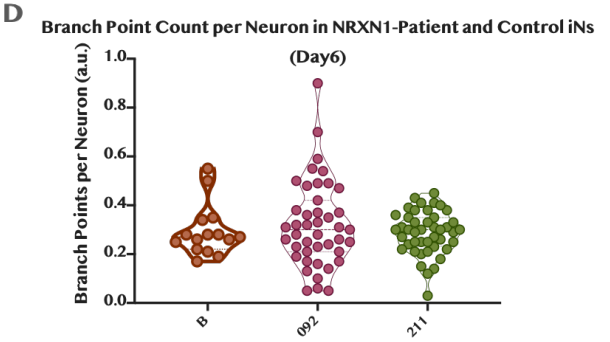
B


Total Neurite Length per Neuron in NRXN1-Patient and Control iNs
(Day6)
Neurite Length per Neuron (µm)
100
80
60
40
20
0
*
B
092
211

C


Total Neurite Count per Neuron in NRXN1-Patient and Control iNs
(Day6)
Neurite Number per Neuron (a.u.)
1.5
1.0
0.5
0.0
-0.5
**
B
092
211

D


Branch Point Count per Neuron in NRXN1-Patient and Control iNs
(Day6)
Branch Points per Neuron (a.u.)
1.0
0.8
0.6
0.4
0.2
0.0
B
092
211

E

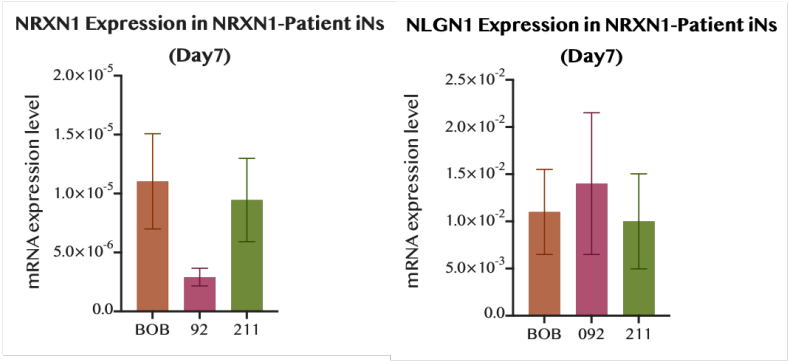

NRXN1 Expression in NRXN1-Patient iNs
(Day7)
mRNA expression level
2.0×10-5
1.5×10-5
1.0×10-5
5.0×10-6
0.0
BOB
92
211
NLGN1 Expression in NRXN1-Patient iNs
(Day7)
mRNA expression level
2.5×10-2
2.0×10-2
1.5×10-2
1.0×10-2
5.0×10-3
0.0
BOB
092
211

Figure 5-9: Neurite outgrowth assay for iN-derived from Control and NRXN1-Patient lines, A) Immunofluorescent staining of MAP2 (morphological markers) of iNs at Day6, B)Violin plots of Total Neurite length/neuron , C) Violin plots of Branch point count/neuron and D) Violin plot of total neurite count/neuron in Control and NRXN1-mutant lines. The plots display the mean with s.e.m, P<0.05 by Kruskal-Wallis and Dunn's multiple comparisons test, E) mRNA expression levels of NRXN1 (left) and NLGN1 (right) in induced neurons Day7 using, Taq-man and SYBR green assays, respectively. The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

5.3 Discussion

Alternative splicing of *NRXN*1 is it thought to be crucial for the recruitment of post-synaptic binding partners and hence impact synaptic function in human neurons [129]. It is therefore critical to understand the effect how mutations in *NRXN*1 alter the *NRXN*1 isoform repertoire in *NRXN*1-mutant lines as well as the patient lines. The *NRXN*-1 isogenic lines include 2 heterozygous and 2 homozygous mutations, all indel mutations were introduced in Exon 19 to target the first shared exon by both isoforms. *NRXN*1α expression increased in the induced neurons overtime, consistent with the gene's expression pattern reported in human brain and neural stem cell development [166], [276].

The four *NRXN*1-mutant lines showed a marked decrease in *NRXN*1α when compared to the wild type, while only two mutations (A2 and G4) altered the β isoform significantly. The *NRXN*1-mutant line G4 also showed a significant reduction in *NRXN*1 mRNA levels. This confirms that the CRISPR-induced mutations had different effects on the *NRXN*1 transcription machinery.

Moreover, no significant difference in expression in *NRXN*1 or its isoforms were reported in the *NRXN*1-patient lines when compared to the control. However, the two patient lines showed a very distinct patterns and levels of expression.

The *092_NXF NRXN*1 and *NRXN*1β mRNA were significantly lower than that of the *211_NXM* patient line. It is also worth noting that the *092_NXF* expression for *NRXN*1 and *NRXN*1β increased gradually with time, the BOB-NGN2 and *211_NXM* showed a peak in expression levels 1 week and 2 weeks post-induction, respectively. The mutations of the two patients' line are in the same region but give rise to opposite effects on *NRXN*1 expression. This might be due to the regulatory elements found in this region that might be controlling *NRXN*1 expression directly or indirectly. For example, the 092 deletion overlaps with the loci of an important regulatory element known as miR8485 or miR-NID1.

This *NRXN*1 intron- derived microRNA (miR-NID1) regulates *NRXN*1 expression through binding to Tar DNA-binding protein-43 [314]. TDP43 facilitates microRNA processing and has been implicated in neurodegeneration. Computation prediction models and functional enrichment analysis reported an overrepresentation of TDP-43 target genes in neural activity and neurological conditions[314].

It is therefore a possibility that the mutation harbouring miR-8485 in *092_NXF* to be responsible for the differential regulation of *NRXN*1.This however cannot be confirmed as this was not addressed nor investigated in this experiment. This is merely an example of how different mutations have different effects on *NRXN*1 gene expression, and regulatory elements in this region might hold key to this variability.

Another possible explanation to the variable outcomes seen in the mutant and patient lines that these mutations gave rise to novel and/or mutant isoforms that are not detectable using traditional methods (such as Taqman assay). The abundance of *NRXN*1 isoforms makes it challenging to answer this question without more novel in-depth analysis. Even using whole-transcriptome RNA-Seq Flaherty et al, reported no significant reduction in *NRXN*1 expression in *heterozygous NRXN1-mutant* hiPSC-NPCs or hiPSC-neurons compared to controls, this was then proven to be due to the shifting in isoform expression and the expression of novel isoforms by the mutant lines. They found a 50% reduction in wild type isoform expression and the expression of 30 novel isoforms [129].

The repertoire of *NRXN*1 isoform is very limited and do not include all potential variants and mutations result in the expression of novel isoform that have not been annotated. This limits the quantification of *NRXN*1 isoforms, and the development of novel targeted tools is essential to fully examine to effect of mutations on isoform expression. Recently, Flaherty et al developed a targeted hybrid sequencing approach

to overcome those limitations and to thoroughly investigate the effect of deletion mutations on *NRXN*1 isoform expression [129].

*NRXN*1β was significantly higher than that of *NRXN*1α in the control and the patient lines. The shift in *NRXN*1 isoform expression may also be due to synapse function, as it has been previously suggested that *NLGN*1 binding to α and β neurexins not only mediated trans-synaptic cell adhesion but serves as a signalling code to specify the type of synapse formed[111], [170]. It is however important to be cautious and not to draw any conclusions from these observations, as the difference in isoform expression levels might be due to primer pair efficiency and not necessarily reflect the different level of mRNA expression. Comparing expression levels of different genes/isoforms is better done using RNA-Seq where primer-pair efficiency will not be a contributing factor.

This is regulated by alternative splicing of *NLGN*1 and *NRXN*1, for example, *NRXN*1 has six canonical splice sites while *NLGN*1 has 2 splice sites (A and B). *NLGN*1 with an insert in SS#B binds exclusively to β-*NRXN*s lacking an insert in SS#4, while *NLGN*1 lacking an insert in SS#B binds freely to both *NRXN* isoforms regardless the presence of absence of an insert at SS#4 [106]. Many studies suggested that *NLGN*1 induce GABAergic but not glutamatergic synapse formation by binding to *NRXN*α and a subset of *NRXN*1β (with an insert in SS#4), while *NLGN*1 binding to *NRXN*1β (lacking

an insert in SS#4) restricts its function to glutamatergic neurons [315], [316]. This might explain the increased expression of *NRXN*1β in our induced-NGN2 glutamatergic neurons. As discussed in Chapter 3, the induced neurons are excitatory glutamatergic neurons and thus we expected to see a higher *NRXN*1β expression.

*NRXN*s are highly expressed in the early developing cerebral cortex, and each gene shows a distinct pattern as reported by sequencing fetal brain samples between 8 and 12 weeks post-conception [313]. This confirms that they and their ligands are highly implicated in synapse formation and synaptic transmission. In this chapter, we wanted to investigate the effect of *NRXN*1 mutations on the expression of *NRXN*-associated genes.

The effect of the CRISPR-edited mutations as well as those found in the patient lines on other *NRXN*s in the neurexin family and their post-synaptic ligands was investigated and as reported in previous studies, *NRXN*1 mutations have not significantly affected other members in the *NRXN* or the *NLGN* family of genes [195], [276]. This was not the case for *CASK*; a significant increase of mRNA levels was reported in the *NRXN*1-mutant line A7, when compared to the control and the other *NRXN*1-mutant lines. Similarly, the 092_NXF Patient line showed higher overall *CASK* expression when compared to the control and the other *NRXN*1-patient line (*211_NXM*). Pak et al 2015 have also reported higher *CASK* protein (but not mRNA)

levels in the heterozygous *NRXN*1$^{+/-}$ mutant line [195]. This indicates a possible regulatory role in *CASK* expression.

The NGN2-induced neurons showed distinct expression of the *NRXN* and *NLGN* genes in the investigated lines. *NRXN*2 is the smallest gene and the most highly expressed during early brain development [106]. Unlike *NRXN*3, *NRXN*2's small size allows its transcription to be completed within a cell cycle in dividing cells [128], [313]. NGN2-induced neurons in the *NRXN*1-isogenic and patient lines expressed *NRXN*3 at significantly lower levels than *NRXN*2. This is consistent with findings of *NRXN*s expression pattern across different brain regions in post-mortem tissue (8-12 pcw) by both PCR and RNA-Seq [313]. The same study reported that both *NLGN*2 and 3 are highly expressed while *NLGN*2 is lowly expressed and increased with age. These results were successfully replicated in our NGN2-induced neuron; *NLGN*2and3 were significantly higher than *NLGN*1 in the *NRXN*1-isogenic and patient lines. The data presented here confirm that *NRXN* and their binding partners show high expression levels during neural differentiation and are generally not affected by the *NRXN*1 mutation. Their expression at this critical time window provides further evidence to their role in neurogenesis and synapse formation in NGN2-induced neurons.

The interaction between cell adhesion molecules is critical for the neural developmental processes; neural differentiation, neurite outgrowth, axonal guidance

and synaptogenesis [137]. The growth of neurite is controlled by the interaction between growth cone receptors and their ligands. Neurexins are concentrated in growth cones and are thought to be involved in the neurite outgrowth by binding to *NLGN*1 [191], [317]. Some studies reported that *NRXN*1 and *NLGN*1 induce neurite outgrowth in rat hippocampal neurons [137], [188].

Since studies associated autism with aberrant neurogenesis and neurite outgrowth and several studies implicated *NRXN*1 in these processes, I hypothesized that the mutations found in the autism patient lines as the CRISPR-edited lines will exhibit atypical growth of neurites. We tested based our analysis on three main parameters: total neurite count, neurite length and branch point count per neuron. The results showed that the 211_NXM Patient line showed no difference in neurite length, neurite count or branch point count per neuron, while the 092_NXF Patient line showed a lower neurite count per neurons when compared to control. As previously reported, this line also showed a significantly lower *NRXN*1 expression level than the *211_NXM* line, without functional analysis, we cannot claim that this is a direct effect of this mutation or if this was caused by the patient's genetic background or other underlying conditions. Morphological and temporal analysis of early-born cortical neurons derived from idiopathic autism patients showed marked increase in neurite length and accelerated growth, which is inconsistent with our results [318].

It is important to note that the idiopathic lines included in their studies were also diagnosed with macrocephaly, a condition identified in 15% of autism adolescents and one of the main causes of macrocephaly or increased brain volume is the increased size and number of dendrites [102], [318]. Microcephaly on the other hand, has been reported in 20% of autism cases and reported in 3% of the general population[318]. The increase number and size of the dendrites is therefore specific to macrocephaly and not autism which explains why an opposite effect was observed in one of the *NRXN*1-patient lines in this study. To query the effect of the CRISPR-induced mutations on neurite outgrowth, we tested the same parameters iNeurons at Day6 on the NRXN1-mutant lines.

Three out of four *NRXN*1-mutant lines, and one of the *NRXN*1-patient lines, showed a significant reduction in at least one of the three tested parameters. The E9 and A2 of the *NRXN*1-mutant lines, had a diminished neurite count, length and branch point count per neuron. Also, the *092_NXF* patient line, showed a reduced number of neurites and total neurite length per neuron.

To query if this is associated with differences in *NRXN*1-*NLGN*1 expression level, we looked at iNs at Day7 in these cell lines and compared them to the control. Interestingly, the aforementioned cell lines showed a reduced *NRXN*1 and increased *NLGN*1 mRNA levels when compared to the control. Although this did not reach

statistical significance, it warrants some further studies and experiments to explore the possible link between these molecules and the morphological changes reported in our lines. These results suggest that disruption in *NRXN-NLGN* complex lead to aberrant neuritogenesis, in concordance with several studies that successfully showed that these complexes are involved in neurite outgrowth induction in Xenopus, Drosophila, rat hippocampal neurons and neural stem cells [137], [189]–[191].

The aim of this chapter was to thoroughly characterize important molecular and morphological properties of the *NRXN*1-mutant and patient lines. Further characterization of NRXN1 mutation on protein levels would be valuable for a more physiological read out, however due to the lack of working antibodies that specifically detect neurexin-1 protein levels, and the technical and time limitations to create to introduce an epitope tag for protein detection, molecular characterization in this thesis relied only on RNA-Seq data and Taq-man assay. The *NRXN*1-patient lines showed distinct patterns of expression for *NRXN*1 isoform and the *NRXN*-associated genes. They also showed different morphological phenotypes. The effects of induced mutation in the mutant lines showed to be affected in the same direction but to variant extent. While all the lines showed a marked reduction in *NRXN*1 α expression, only two lines showed a reduction in β expression. Similarly, with morphological analysis, different mutation affected different parameters of neurite outgrowth at varying magnitude.

The results presented confirm that the mutations in *NRXN*1 line was associated with molecular and morphological effects on the induced neurons. Furthermore, the autism patient lines also behaved differently, it is unclear if this is due to the difference in their mutations or other elements of their genetic background that might driving these differences. The variations in the phenotypes presented in this chapter highlights the diversity of *NRXN*1 mutation and showcases the heterogeneity of autism.

Chapter 6

Functional characterization of NRXN1-mutant and NRXN1-patient lines

Summary

NRXN1 mutations have been repeatedly associated with autism and other neuropsychiatric conditions, presenting variable penetrance and diverse clinical presentation. The impact of NRXN1 mutations alone on human neurons remains ambiguous. In this chapter, the aim is to analyse the effect of the NRXN1 mutations in the NRXN1-mutant lines and the NRXN1-patient lines on calcium signalling and electrophysiological activity of human neurons five weeks post induction by comparing them to control.

6 Functional characterization of NRXN1-mutant and NRXN1-patient lines

6.1 Introduction

The synapse is defined as the specialised site at which one neuron communicates with another through a process called synaptic transmission. Synaptic molecules have been extensively implicated in synapse formation maturation and function; especially neurexins and neuroligins that are known to mediate the excitation and inhibition (E/I) balance in neurons [192]. Several studies implicated α and β neurexins in synaptic function and synaptic transmission at excitatory and inhibitory terminals [82], [170], [319], [320].

Few studies have used iPSCs to describe changes in synaptic transmission associated with NRXN1 mutations [181], [195]. IPSCs derived from three autistic individuals with heterozygous mutations in three different regions within the NRXN1α reported increased amplitude, frequency and duration of calcium transients. This was complemented with upregulation of genes implicated in voltage-gated calcium channels and ion transport pathways [181]. Conversely, another study reported a significant decrease in calcium signalling activity in neurons derived from an autism iPSC carrying a bi-allelic mutation in NRXN1α [180].

Pak et al, generated two iPSC lines with mutations in exon 19 and 24 (shared by both NRXN1 isoforms) and report reduced frequency of spontaneous miniature excitatory postsynaptic currents (mEPSCs) compared to the control [195]. Although iPSC provide a useful tool to reveal the role of neurexins in synaptic transmission, most of what is known was revealed using animal models [82], [193], [319], [320].

Triple knockout of α-neurexins in mice resulted in a decrease in calcium channel functioning leading to an impaired neurotransmitter release [82]. Loss of function of the NRXN1α isoform in mice was associated with a marked decrease in excitatory (but not inhibitory) synaptic strength [193]. To investigate the role of ß-neurexins, Anderson et al generated knockout mice of all ß-neurexins and reported the suppression of presynaptic release probability that is caused by the dysregulation of the post-synaptic endocannabinoid synthesis [194].

In another mice model, deletion of all neurexins resulted in impaired synaptic transmission in two different types of synapses through different mechanisms. For instance, in parvalbumin-positive interneurons of the prefrontal cortex, NRXN deletions resulted in loss of synapses with no changes in action potential triggered calcium influx while the same deletion resulted in no loss in synapse number, however, it caused an impaired action potential-induced calcium influx as reported by the decrease of amplitude and increased decay times [211]. Taken together these results highlight that the functional effects of neurexin

deletions may vary dramatically depending on experimental design, the model used and/or the subtype of neurons being investigated [82], [194], [211].

Monitoring neurons activity in a neuronal network relies on detecting membrane voltage changes by measuring the action potential activity extracellularly (multielectrode array) and/or optical recording of ionic changes in the neurons (Intracellular Calcium ions) [321].

Multielectrode arrays have the ability to monitor the electrical activity of a global neuronal network in-vitro [321]. Using this technique allows the simultaneous recording of action potentials with an adequate temporal resolution. These arrays can only detect extracellular activity and therefore cannot identify the activity of individual cells within the neural network unlike optical imaging approaches [321], [322]. The combined use of advanced live imaging techniques and calcium indicators allows the monitoring of intracellular calcium signals of individual neurons [322], [323].

Calcium indicators include a wide range of classes, like bioluminescent proteins, chemical calcium chelators, FRET-based and single-fluorophore genetically encoded calcium indicator [322]. Calbryte 520 AM (ab171868) is a new synthetic calcium-sensitive dye, upon crossing the membrane and upon binding to calcium ions produced a bright measurable signal that is indicative of cellular calcium in the neurons. Since action potentials cause measurable

Calcium changes in all parts of the neuronal cell, these recordings are used as an indicator for the neuron electric activity and the optical nature of the experiment allow the identification of the individual cellular activity [323].

Driven by RNA-Seq data of the distinct transcriptomic profiles resulting from the patient and mutant line and the prominent downregulation of cadherins, I propose that the NGN2-dericed neurons exhibiting a functional phenotype. The aim of this chapter is to investigate neuronal activity and function those neurons intracellularly using calcium imaging and extracellularly using MEA, 5 weeks post-induction.

6.2 Results

6.2.1 Intracellular recording of neuronal activity (Calcium imaging)

NRXN1-patient and isogenic mutant-iPSCs were induced and co-cultured with rat mixed-glia and stained with a fluorogenic calcium-sensitive dye (Calbryte 520, AM) to monitor spontaneous calcium signalling in iNs at Day 35 for 3 minutes (Figure 6-1A). There was no difference in the percentage of active cells between different lines. Each time series was processed, the cell bodies were automatically segmented and numbered (using a well-defined criterion as described in chapter 2). The signal intensity was scaled for each cell body from 0-1 and the proportion of cell bodies was calculated (number of active cell bodies/total number

of cell bodies (Figure 6-1B). No difference was reported in the proportion of active cell bodies between the investigated cell lines (Figure 6-1C).

Calcium signals were analysed for calculation of peak duration and amplitude. The duration was computed by measuring the full width in the middle of the peak (full-width mid-height) and the amplitude was reported by measuring ΔF/F (>2%) (Figure 6-D and E).

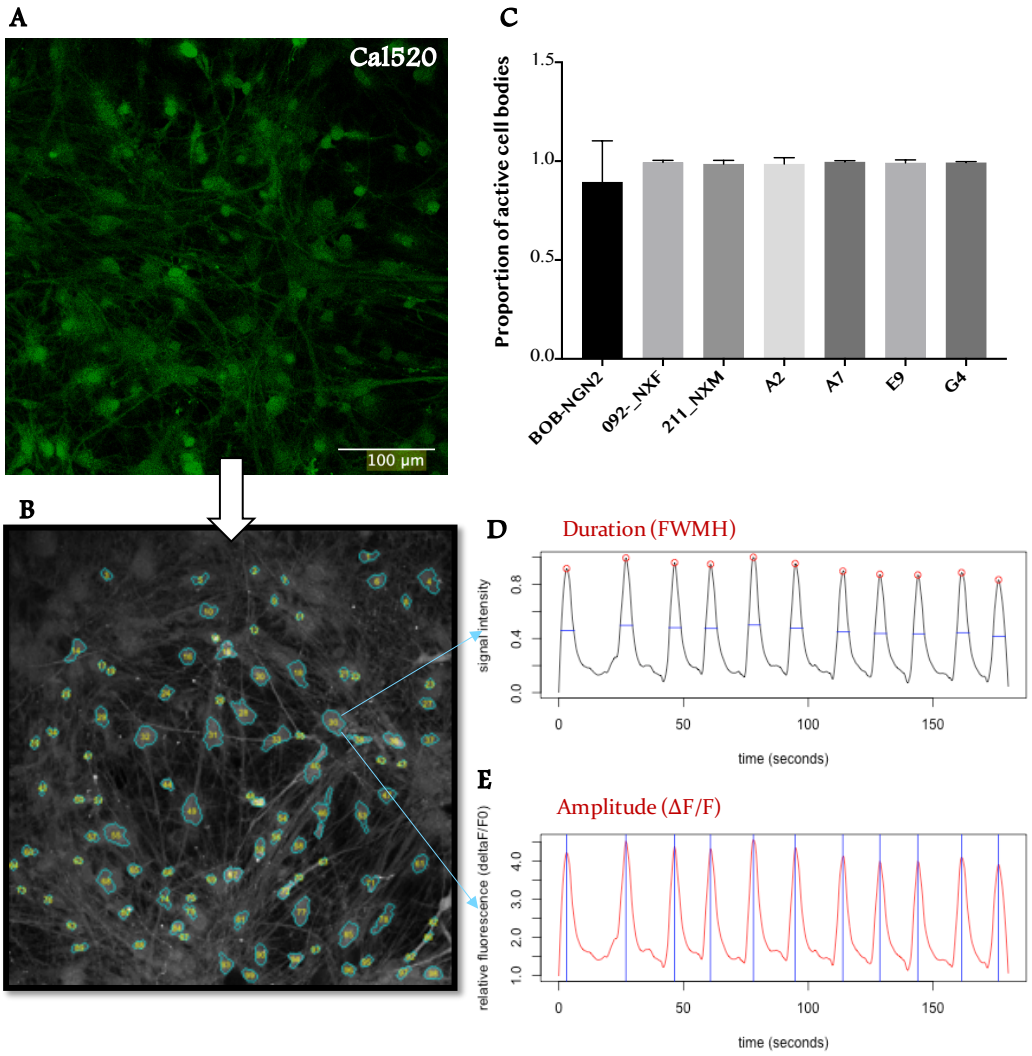


Figure 6-1: Calcium signalling analysis workflow. A) Day35 iN stained with Cal520 AM, B) Automated segmentation of cell bodies, C) Bar graph for the proportion of active cell bodies, D) Duration of peaks is calculated by measuring the peaks' FWMH (vertical blue line), E) Peak amplitude (ΔF/F), the horizontal blue line indicate the identification of a peak (ΔF/F>2%).

Given the suggested contribution of NRXN1 mutations in hyper synchronicity and hyperexcitability [324], the calcium signals were analysed to look at the frequency of spikes per minute, the frequency of spikes per cell body and the proportion of active cell body firing at the same timepoint (synchronicity).

6.2.1.1 The NRXN1 patient lines showed no significant changes in peak amplitude or duration when compared to control

The calcium signal amplitude in this experiment was reported by measuring ΔF/F > 2% (Figure 6-1). Since calbryte520 is a non-ratiometric dye, the calcium indicator dye does not report the values of free and bound calcium intracellular separately which is not ideal for measuring amplitude.

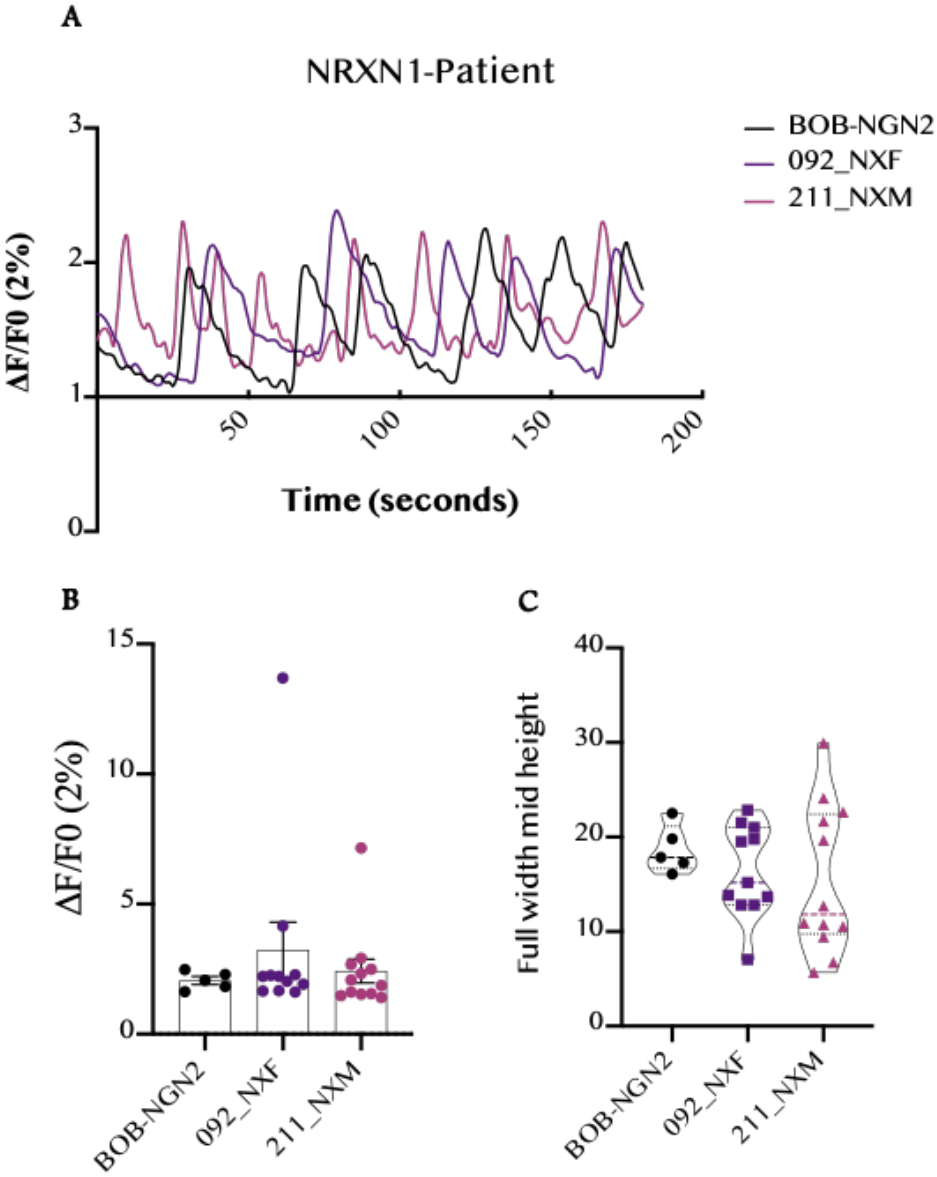


Figure 6-2: Measurement of Calcium signals in NRXN1-Patient line, A) Representative time graph of calcium signals (ΔF/F) over time (180 seconds), B) Amplitude ΔF/F, C) Duration (FWMH) The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, [BOB-NGN2 (n=5), 092_NXF (n= 11), 211_NXM (n= 12)]. The plots display the mean with s.e.m, P<0.05 by one-way ANOVA, and Tukey's post hoc test for correction.

6.2.1.2 *The NRXN1-mutant lines showed no significant changes in peak amplitude or duration when compared to control*

The amplitude peak and duration were not affected in the NRXN1-mutant lines when compared to controls as shown in Figure 6-3.

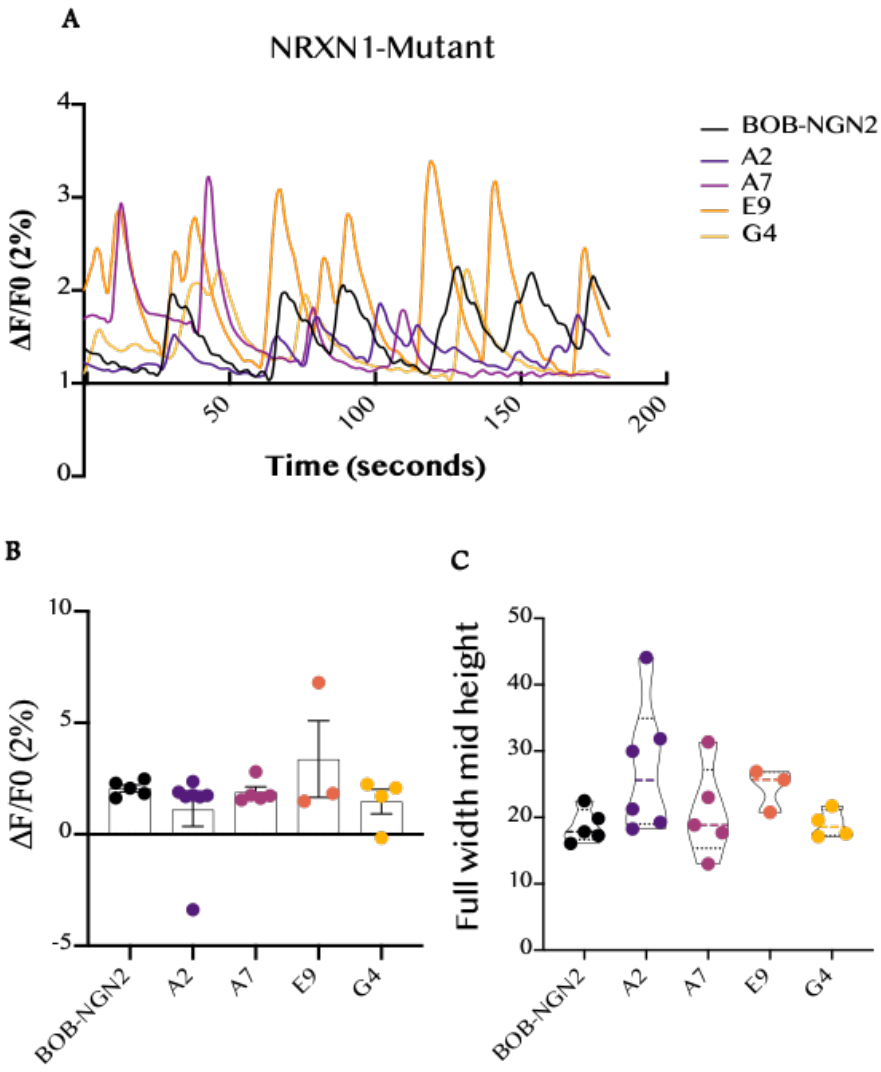


Figure 6-3: Measurement of calcium signals in NRXN1-Mutant line, A) Representative time graph of calcium signals (ΔF/F) over time (180 seconds), B) Amplitude ΔF/F, C) Duration (FWMH) The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction,[BOB-NGN2 (n=5), A2(n= 7), A7 (n=5), E9 (n=3), G4 (n=4)] . The plots display the mean with s.e.m, P<0.05 by two-way ANOVA, and Tukey's post hoc test for correction, n=3.

6.2.1.3 *The 092NXF-derived neurons show more synchronised firing when compared to the control line*

The synchronicity of the NRXN1-patient and control lines firing is represented by the waterfall and raster plot in Figure 6-4 and summarised in the synchronization graph showing the proportion of cell firing at every single time point. A marked difference was observed between the patient and control lines, a difference that only reached statistical significance for the 092NXF patient line. No difference was reported in the number of peaks per cell body or per minute between the control and the patient lines (Figure 6-4E and F).

6.2.1.4 *The firing synchronicity in the NRXN1-mutant line (A7) is significantly higher than the control and to other NRXN1 mutant lines (A2 and E9)*

The A7 mutant line showed a pattern of firing distinct form the other mutant lines and the control. This was clearly shown when plotting the proportion of cell bodies at each time point (Figure 6-5B), the size of the peaks on the graph represents the higher number of synchronised cell bodies. This difference was significantly higher than the control, A2 and E9 but not the G4 induced neurons. The frequency of peaks per minutes and per cell bodies showed a high variability within the A2 and the A7 neurons. No significant differences were reported between the control and the mutant lines (Figure 6-5F and G).

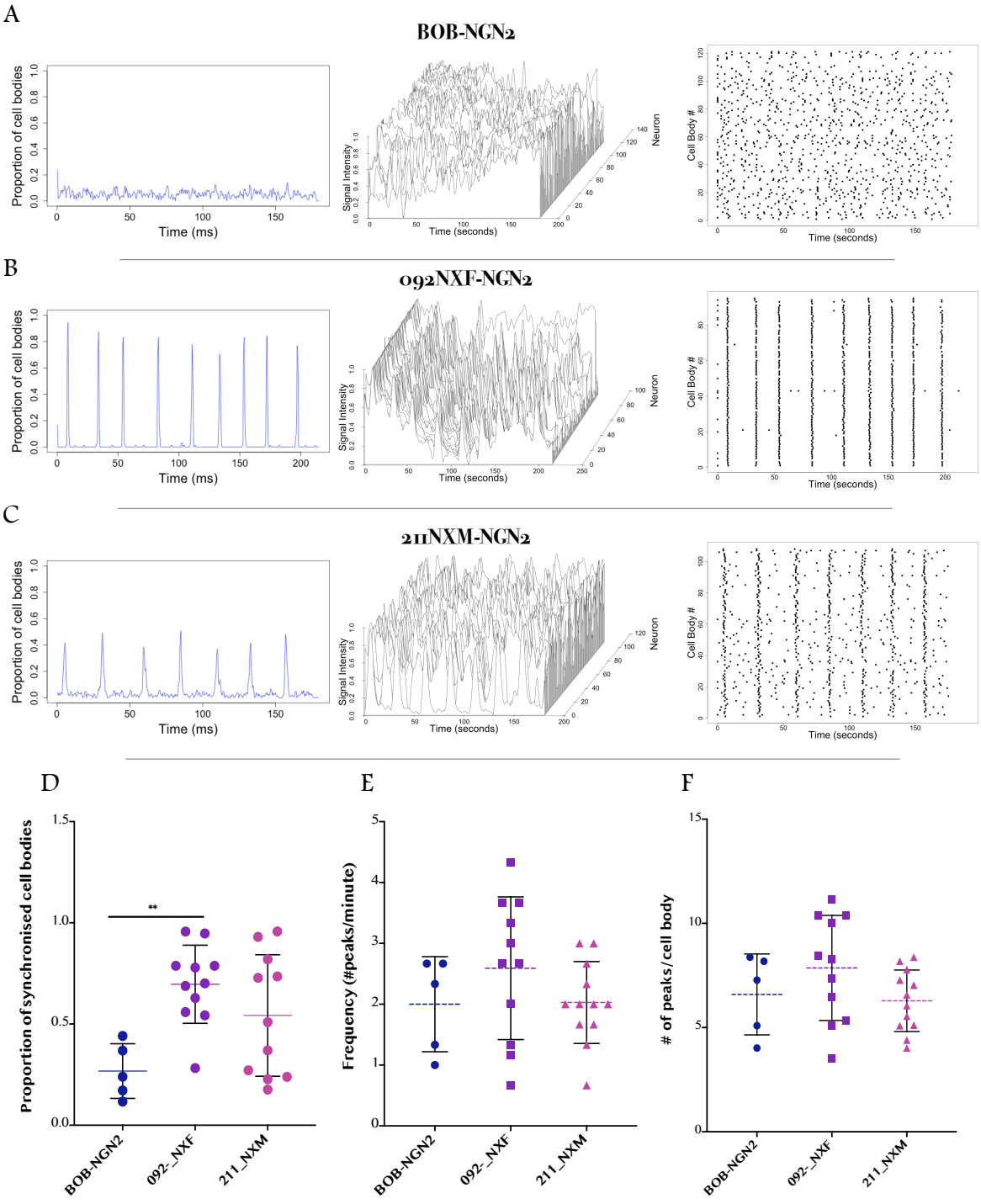


Figure 6-4: Measuring firing synchronicity and frequency in the NRXN1-patient lines, A) The far left graphs represent the proportion of cell bodies (Y axis) that are firing at the same time point (X axis), and the waterfall plot shows a 3D plot of the number of cell bodies (Z axis) firing at the same time (X axis) and their signal intensity (Y axis) and on the right a raster plot showing the cell bodies (Y axis) firing at each time point (Y axis) for A)BOB-NGN2, B) 092_NXF and C) 211_NXM. The mean values of each experiment were then plotted in D) Proportion of synchronised cell bodies, E) Frequency of peaks (per

minute), F) Frequency of peaks (per cell body). The plots display the mean with s.e.m, P<0.05 by one-way ANOVA, and Tukey's post hoc test for correction.

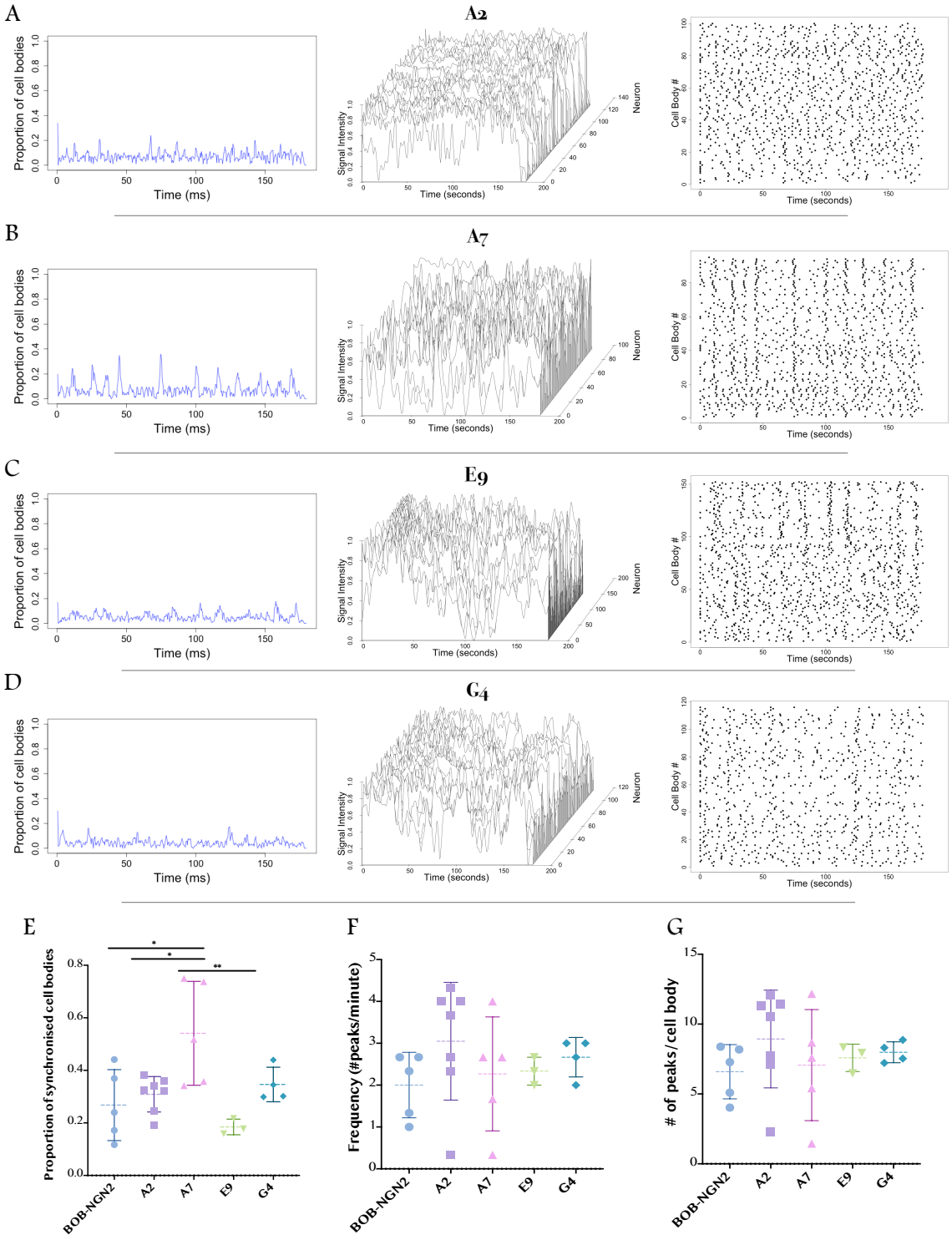


Figure 6-5: Measuring firing synchronicity and frequency in the NRXN1-Mutant lines. The far left graphs represent the proportion of cell bodies (Y axis) that are firing at the same time point (X axis), and the

waterfall plot shows a 3D plot of the number of cell bodies (Z axis) firing at the same time (X axis) and their signal intensity (Y axis) and on the right a raster plot showing the cell bodies (Y axis) firing at each time point (Y axis) for A) A2, B) A7, C) E9 and D) G4. The mean values were plotted in E) Proportion of synchronized cell bodies, F) Frequency of peaks (per minute), G) Frequency of peaks (per cell body). The plots display the mean with s.e.m, P<0.05 by one-way ANOVA, and Tukey's post hoc test for correction.

Finally, the neurons were stimulated with 10µM ionomycin, a calcium ionophore as a positive control resulting in a rapid and sustained calcium influx reflected in the sudden increase in signal intensity (Figure 6-6).

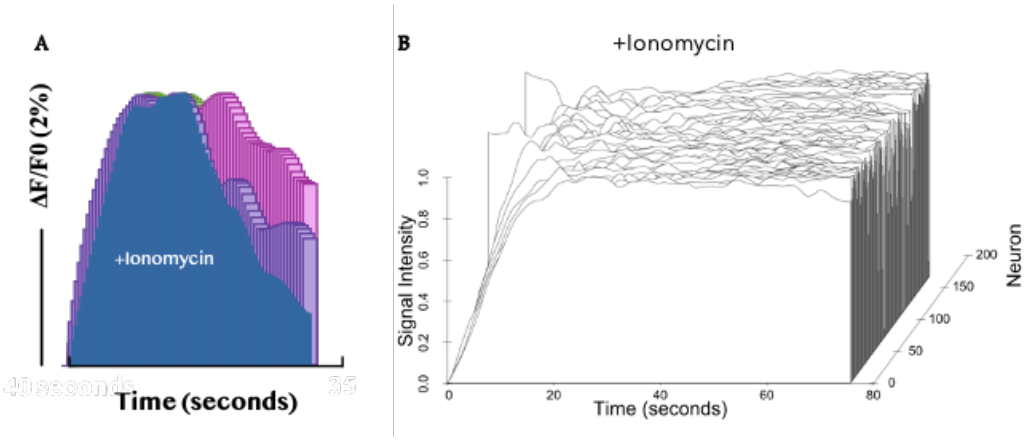


Figure 6-6: Stimulation of iN with 10µM ionomycin: representative waterfall plots showing the sudden rise of calcium signal using ΔF/F (A) or signal intensity (B) on the Y axis across time X axis resulting from ionomycin treatment.

6.2.2 Extracellular recording of neuron activity (MEA)

The multichannel 6 well MEA was used to record the activity of different neuronal populations (Control, Patient and Mutant) simultaneously in an unbiased manner [325]. The induced neurons were co-cultured with rat glia to induce their maturation and cultured on the 6 well MEA, 9 electrodes per well (Figure 2-8).

The real-time spontaneous activity of the cells was recorded on Day35 for 15 minutes. The raw voltages were subjected to a Bessel filter (a high pass cut off 200 Hz) (Figure 5-7). Spikes were detected using a threshold of 5 times the standard deviation of noise signal calculated over a 10 ms moving window. Electrodes with less than 5 spikes per minute were considered inactive and were excluded from analysis. Single electrode bursts were identified if a minimum of 5 spikes were detected in a burst with a minimum 100ms interval between bursts. Network bursts were defined as bursts firing simultaneously in 30% of active electrodes (Figure 6-9). In this experiment, only the control (BOB-NGN2), the NRXN1-patient lines (092_NXF and 211_NXM) and the NRXN1-mutant lines (A2 and A7) were investigated. The other two NRXN1-mutant lines E9 and G4 were not included for technical reasons.

6.2.2.1 A significantly higher mean firing rate in the 211_NXM patient line when compared to control

The extracellular activity of iNs at Day 35 was filtered and analysed using the multichannel analyser. The filtered spikes were counted per electrode and divided by 900 seconds to calculate the mean firing rate (Figure 6-7). The weighted mean firing rate of NRXN1-patient lines was plotted against the control and the 211_NXM patient line showed a marked increase in number of spikes per seconds when compared to control ($p=0.004$) and to the 092_NXF patient line ($p=0.002$) as reported by One-way ANOVA and tucker post hoc analysis (Figure 6-7C).

6.2.2.2 *Mutations in A2 and A7 had no effect on the weighted mean firing rate of cultured neurons*

The weighted mean firing rate in the NRXN1-mutant line (A2 and A7) showed some variability but no marked difference when compared to the control lines (Figure 6-7D).

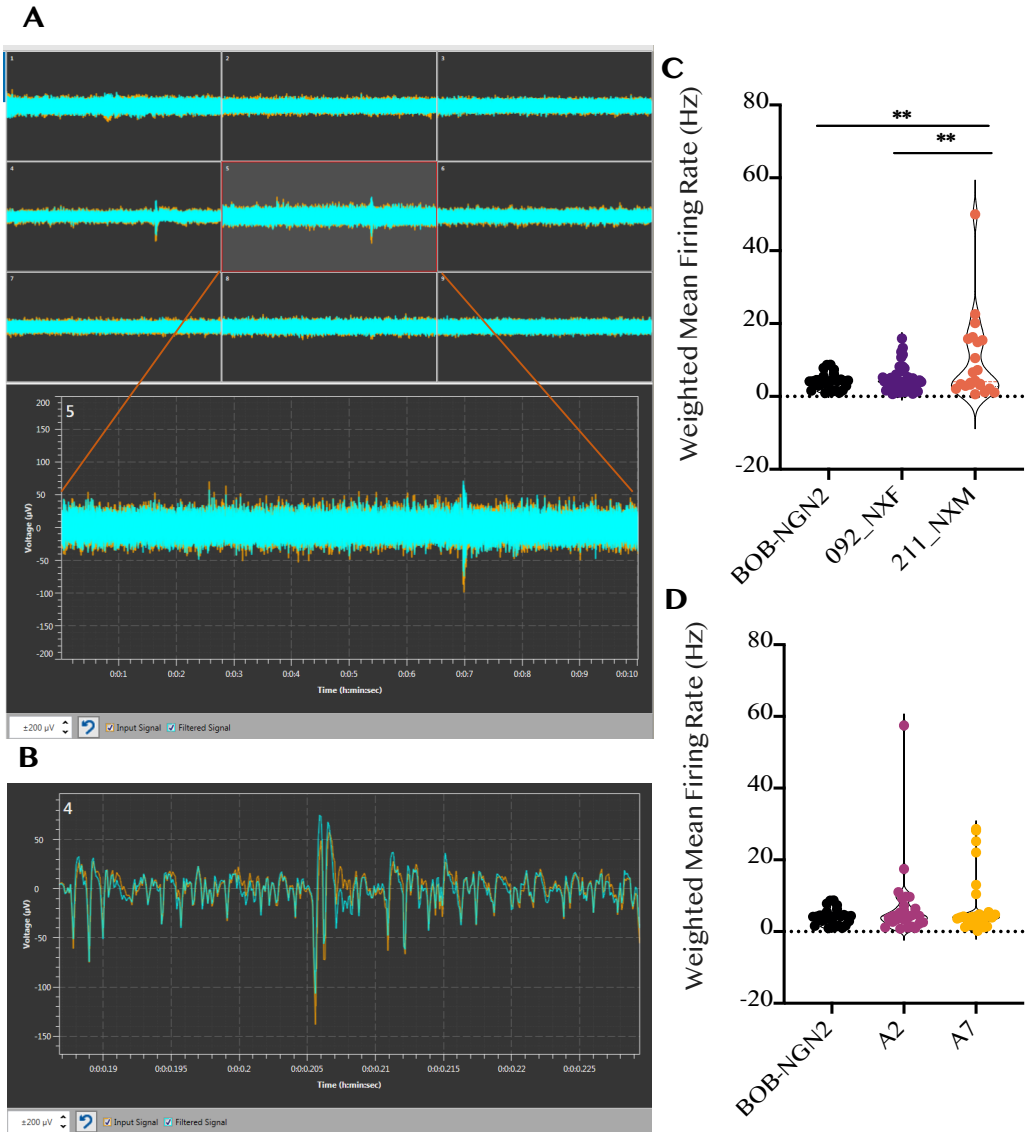


Figure 6-7 Spike analysis, A) Visual representation of input(orange) and filtered (blue) signals using a Bessel high pass filter (200Hz), the above panel shows the signal in the 9 electrodes of the MEA, and below is a single electrode view (Electrode 5). B) Filtered signal view for spike analysis showing a filtered

spike at +/-200μV resolution during 10ms. The mean firing rate (per single electrode) for C) NRXN1-patient and D) NRXN1-mutant lines. All the values were normalized to the control.

6.2.2.3 Burst analysis reported no significant difference between the control and the NRXN1-patient not the NRXN1-mutant lines

To monitor the synchronised activity of the neurons within the same electrode, spikes were counted across time and in the event of at least five spikes firing within 100 ms a burst was counted. The detected bursts were then analysed to indicate the number of bursts per minute, as well the burst duration and the duration between successive bursts (Figure 6-8). Surprisingly, no marked difference was reported between the patient lines and the control which was unexpected as it was hypothesized that the synchronised firing reported in the calcium imaging can results in higher bursting frequency in the MEA recording. Moreover, burst frequency, mean duration and inter-burst interval were not affected in the NRXN1-mutant lines A2 and A7, indicating that the NRXN1 mutations did not affect the bursting activity in this experiment.

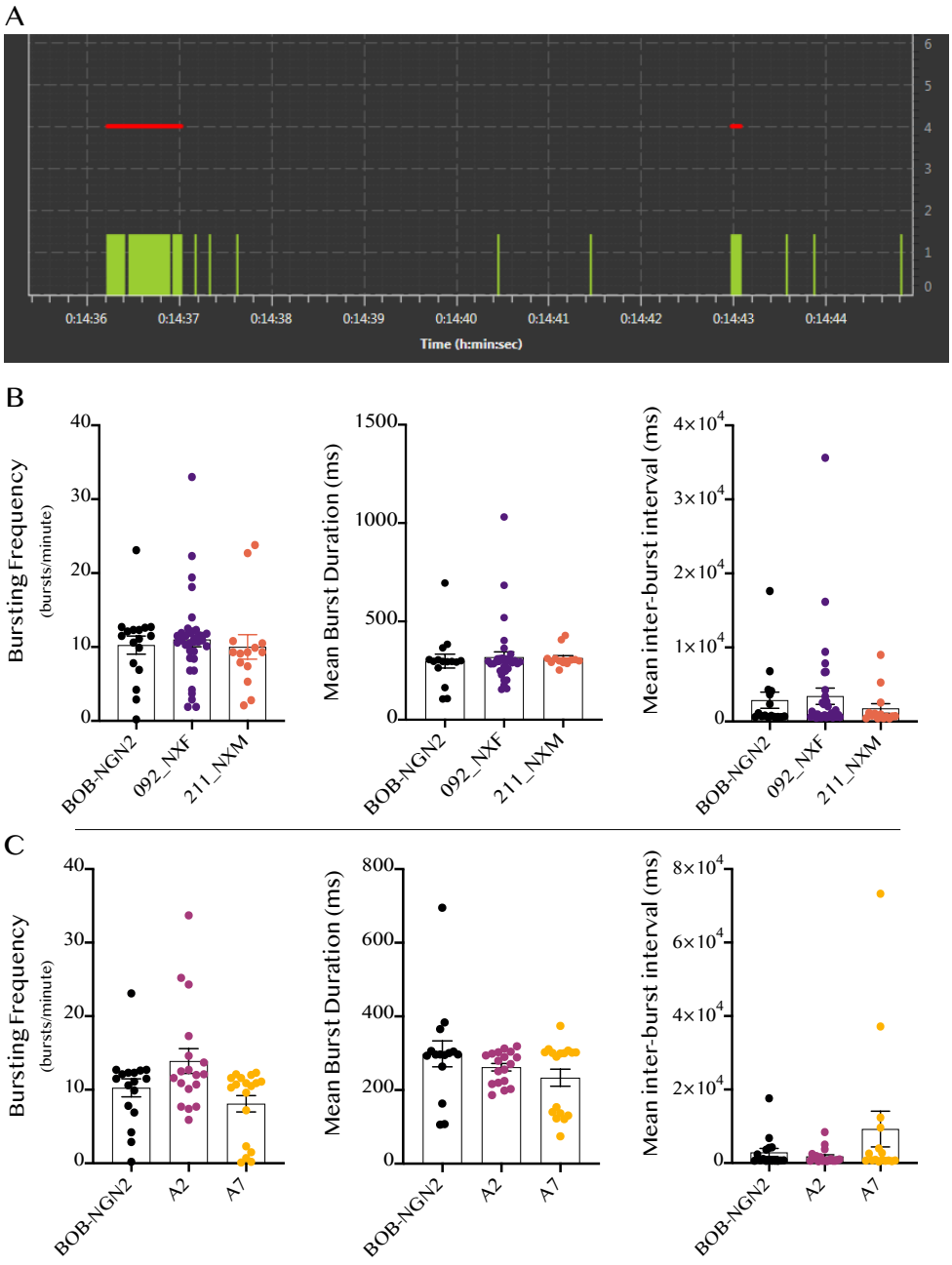


Figure 6-8: Burst analysis, A) Visual representation of bursts detected in 10seconds, the green bars indicate spikes detected and the red bars indicate the bursts. The left panel shows the bursting frequency (number of bursts/minute), the middle one shows the mean burst duration (in millisecond) and the right panel the mean inter-burst interval (in millisecond) for B) NRXN1-Patient lines and C) NRXN1-mutant lines. All the values in the graph have been normalized to the control line.

6.2.2.4 NRXN1 mutations in the patient and mutant lines did not seem to affect aggregated network bursting when compared to the control

The network bursts were defined as simultaneous bursting across at least 30% of active electrodes using the same burst detection parameters discussed above. The NRXN1-patient lines did not seem to be affected and showed no different in network bursts detected nor did the A2 and A7 mutant lines when compared to the control BOB-NGN2 (Figure 6-9B and C). It is noteworthy to highlight the variation of number of network bursts reported per independent experiment within the same line. Which is clearly shown in the volcano plots showing the 2-3 populations per cell lines (Figure 6-9).

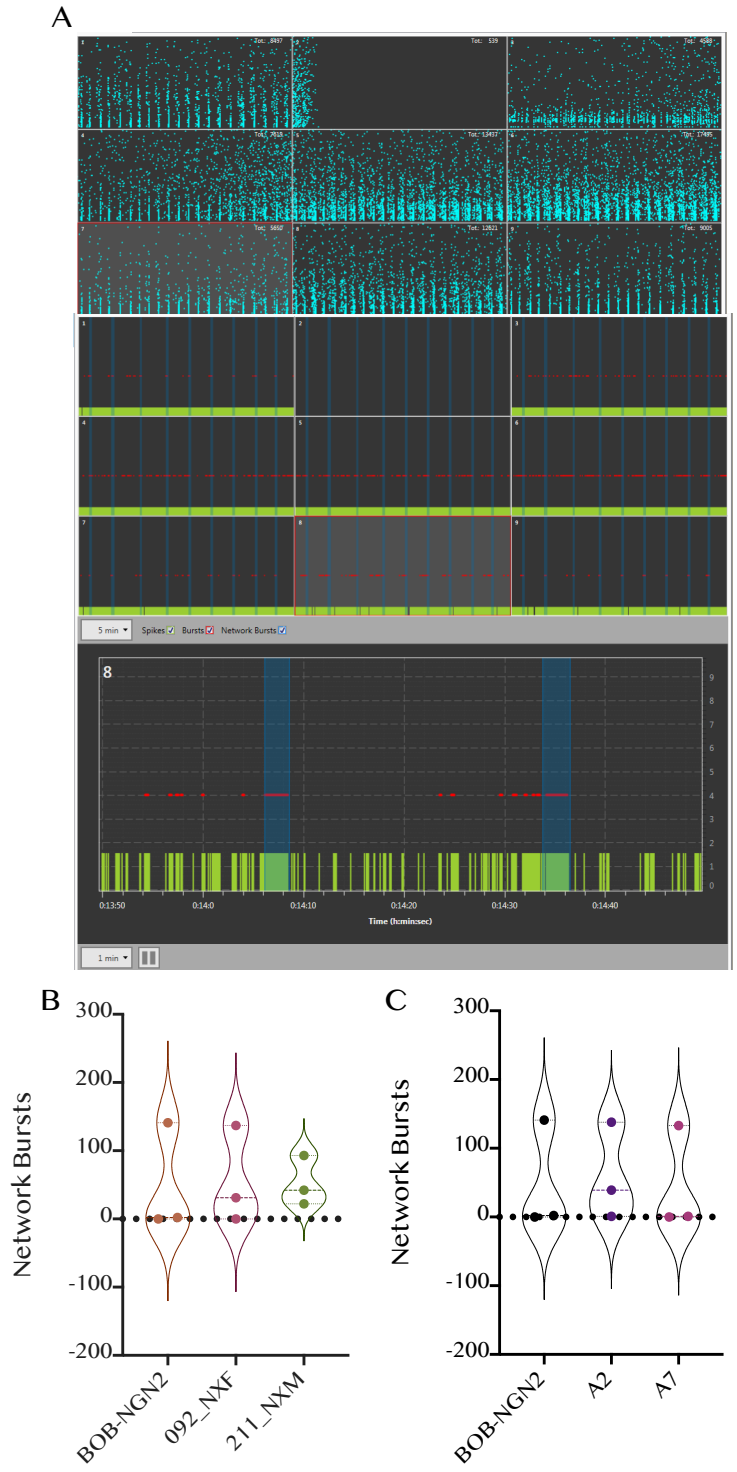


Figure 6-9: Network bursts detection, A) Raster plot of synchronized firing of cell bodies across time across electrodes (top panel), network busts across 5 minutes across electrodes (middle panel), single view network bursts detection (lower panel): the green bars represent the spikes, the red bard indicate the bursts, and the blue highlighted regions show network bursts

6.3 Discussion

Neurexins, given their complex interaction networks at the synapse, mediate various regulatory functions. In this chapter the aim was to characterise the functional properties associated with NRXN1 mutations in patient and mutant lines. The iPSCs were induced to generate excitatory glutamatergic neurons and co-cultured with rat mixed glia to induce maturation. Unpublished single cell RNA analysis from our lab revealed that the NGN2-derived neurons generate 3 populations of neurons. The major population is glutamatergic neurons, a small subpopulation shows higher expression of cholinergic markers and the third is a small population of neurons expressing both glutamatergic and cholinergic markers [225]. Neuronal activity was measured intracellularly using calcium imaging and extracellularly using MEA at Day 35.

Functional calcium signalling revealed that mutations in the NRXN1-patient and isogenic NRXN1-mutant lines did not alter the amplitude nor the duration of spontaneous calcium transients. This, is in disagreement with other iPSC studies that report altered calcium signalling in iPSC-derived neurons with NRXN1α heterozygous or homozygous mutations [180], [181], [195]. Avazzadeh generated iPSC lines derived from autism patient with heterozygous mutations in the NRXN1α isoform and differentiated them into neurons using the Dual-Smad (DS) inhibition protocol. The calcium transients reported in the $NRXN1^{+/-}$ lines had significantly higher amplitude, frequency and duration [181].

Another, study reported decreased calcium signalling in response to membrane depolarization in Day 28 neurons using the same method of differentiation (DS) [180]. The three different phenotypes presented here might be due to the effect of the mutation on isoform expression (α vs ß), method of neuron differentiation (DS vs NGN2) and maturity of the investigated neurons as previously demonstrated by forward programming of iPSC to iNs do not follow a typical path of neural development and do not reach the same grade of neuronal maturity compared to traditional methods [231].

Additionally, the use of patient cell lines adds a layer of complexity to the data as the patient's genetic background is another variable to the equation. In 2015, Pak et al investigated two different isogenic mutations, a deletion in Exon 19 (found in some autism cases) and a truncation mutation in Exon 24, and induced neuronal differentiation using NGN2-forced expression. The neurons derived from both cellular models exhibited a similar phenotype where neurotransmitter release was impaired and spontaneous mEPSC was diminished when compared to the isogenic control [195]. Spontaneous activity was measured in neurons derived from an autism a schizophrenia and a control group. The peak frequency and duration were significantly altered in the schizophrenia, but not the autism group when compared to the control. Interestingly, the amplitude of the peaks was significantly lower in the autism but not the schizophrenia group [326].

The synchronised responses of neuronal events are crucial for normal brain functions and have been associated with specific cognitive functions such as attention, memory formation and recall [327]. Synchronicity of neurons is described as the coherent and collective firing of neurons wired together in a network at the same timepoint. This time-locked activity requires inter-neuronal coherence that allows this synchronous activity to occur within a millisecond time window and results in an amplified field potential. Disruption of neural synchrony has been involved in several neurological pathologies including autism, schizophrenia and epilepsy [327], [328]. The proportion of cell bodies firing was calculated and plotted by every time point to look at the synchronicity of firing in the investigated cell lines (Figure 6-1C).

The NRXN1-patient lines showed higher synchronicity when compared to control however only the 092_NXF patient line reached statistical significance. Similarly, one of the NRXN1-mutant lines (A7) showed a significantly higher synchronised firing when compared to the control and the other mutant lines. The increased synchronicity of neurons firing in A7 and 092_NXF is difficult to interpret, as synchronicity of neural network is a sign of neuronal maturation and mutations in NRXN1 has been linked to delayed maturation of neurons in other studies [180].

The forced expression of NGN2 in our neurons might not show the delayed maturation reported in other studies using DS differentiation methods. Moreover, NGN2

induction generates excitatory neurons of glutamatergic and cholinergic identity as confirmed by single cell analysis of Day 21 neurons in our lab [225].

Other possible interpretation for this phenotype can be explained by taking a look at the co-occurrence of autism and epilepsy which warrants an investigation into their shared molecular and pathophysiological mechanisms [329]. Epilepsy is described as an excessive neuronal synchronization and hyperexcitability [327]. Various studies reported that epileptic hypersynchrony is connected with the increased excitability of neuronal events in epileptic experimental models. Thus, hypersynchronous neural events seen in epilepsy rely on neuronal hyperexcitability resulting from calcium release from internal calcium stores [330]. Based on their role in synaptic transmission and their involvement in brain pathologies, some studies suggested that neurexins and neuroligins may participate in the modulation of neuron hyperexcitability and implicate them in epilepsy [324], [331]. This is just a plausible explanation to the phenotype reported here but confirming or denying it require further investigation specifically designed to answer this question.

The simple non-invasive and unbiased high throughput recordings obtained by MEAs, rendered them a reliable tool for modelling autism and other neuropsychiatric conditions [332]. In this chapter, the control line was cultured alongside the two NRXN1-patient lines (092_NXF and 211_NXM) and two out of four NRXN1-mutant

lines (A2 and A7) in the same 6-well MEA plate and neuronal activity was recorded for 15 minutes on Day 35. The number of spikes and mean firing rate in the NRXN1-patient 211_NXM line was significantly higher than that of the control. This contradicts other studies that report lower number of spikes in neurons with NRXN1 mutations. For instance, a study in 2019 investigated neuronal activity in 4 patient-derived iPSCs carrying 5' and 3' NRXN1 mutations and reported reduced number of spikes in induced neurons 3 weeks post induction.

Furthermore, they show that the overexpression of 4 wild type NRXN1α isoforms rescued the phenotype in the iN with 5'NRXN1 but not the ones with the 3'NRXN1 mutation. The overexpression of the 3'NRXN1 mutant isoforms in the control neurons resulted in a reduced neuronal activity, indicating that the phenotype seen in the 3'NRXN1 iNs is not directly caused by the low abundance of NRXN1α expression but rather a dominant double negative effect cause by the mutant isoforms [129]. Other studies looking into ASD variant effect on electrophysiological activity reported no change in number of spikes in neurons with NRXN1 mutations [325].

Synchronous neuronal activity emerges as the neurons mature and form neural networks. Culturing iPSC-derived neurons on MEAs offer the tools to discover bursting patterns and network dynamic in autism-derived cultures. Following on the hypersynchronous firing resulting from the calcium signalling in the 092_NXF and

the A7 mutant line, I sought to investigate bursting parameters (frequency, duration and inter-burst intervals) in the two NRXN1-patient lines and the A2 and A7 NRXN1-mutant lines. Surprisingly, no differences were reported in any of those parameters. Additionally, network burst activity was computed whenever it was covered on at least 30% of the active electrodes. No differences were reported between the NRXN1-patient nor the NRXN1-mutant lines against the control.

6.4 Limitations

In the calcium imaging experiment, the NRXN1-patient lines showed a marked hypersynchronous phenotype, this was however only significant in the 092_NXF line and this novel phenotype should be addressed using more robust methods of calcium signal quantification. For instance, in this investigation, the non-ratiometric dye (Cal520) dye was used, and although it is very widely used for qualitative imaging it is very complicated to compute an absolute value for the calcium concentration, instead a relative fluorescence value is obtained [322], [333].

Another challenging factor is the presence of external technical factors that might contribute to the fluorescence intensity and skew the data. For these reasons, extra caution was drawn to the probe concentration, acquisition conditions and use of as many biological and technical replicates as possible.

Moreover, the lack of a suitable control to account for the variables mentioned above makes it challenging to draw a conclusion, while a deletion of the whole gene or creating a null mutation could potentially conclude the participatory role of NRXN1 towards the phenotype, it is technically challenging to delete a large and complex gene like NRXN1. Alternatively, creating an identical mutation to the one seen in the neurons (A7 and 092) exhibiting the phenotype but on a different control line could potentially lead to a more conclusive result.

Other concerns should be considered for results interpretation is that the forced expression of NGN2 generate excitatory neurons of glutamatergic identity. The neurons are then co-cultured with mixed rat-glia, the network signals can be mediated by the astrocytes in the co-culture, or the neurons and it is impossible to differentiate that. However, whether it is astrocyte or neurocentral, it is probably induced by the mutation present in NRXN1 and/or the patient's genetic background in the case of NRXN1-patient lines. To further confirm the validity of the hypersynchronous phenotype, replicating these findings in a different model will be necessary. Future model used should have at least two distinct neuronal cell types (as opposed to only excitatory neurons in this one), synapsed in a defined orientation, must have oligodendrocytes, astrocytes and microglia to recapitulate critical aspects of neuronal activity, inflammation and synaptic pruning.

Another strategy is to transplant the neurons into mouse/rat models and record the neurons' activity whilst monitoring the behaviour. Ultimately, it is crucial to have larger cohorts, isogenic controls, improved patterning and maturation of a variety of human neural cell types, cultured either as artificial circuits or transplanted into mice.

MEAs provide an attractive platform for electrophysiological activity which made it widely adopted especially in the field of neuroscience, yet little information was provided for rigorous assay development for the experimental application of these arrays [334]. In this experiment, many factors were considered to overcome the broad range of firing activity observed across arrays. For instance, all the lines to be compared were plated on wells of the same MEA and were normalised to the control well of each plate. The duration of the recording was another important factor to consider, in this experiment and as reported by other studies, 15 minutes seemed sufficiently long to get accurate and reproducible recordings of the neurons' firing [129]. However, a recent study recommended 30 minutes to adequately capture spike activity [334].

In the analysis of these recording, strict spike detection and event filtering metric were applied to confidently report the spike and burst activity and assess differences between the lines. The main challenge in the setup of this experiment is the use of a small number of electrodes (9 electrodes per well) to detect burst activity. In most

experiment, 3 replicates are acceptable for statistical power, however the small number of electrodes of the MEA used in this study required a higher number of replicates to more confidently report bursting activity in the investigated neuronal cultures [334].

Chapter 7

7 Integrated genetic and methylomic analyses identify shared biology between autism and autistic traits

In this chapter I investigated the association between DNA methylation at birth (cord blood), and scores on the Social and Communication Disorders Checklist (SCDC), by conducting a methylome-wide association study (MWAS). No significant CpGs were associated with SCDC except for cg14379490, on chromosome 9. Using methylation data for autism from post-mortem brain tissues, we identify a significant concordance in effect direction of CpGs with a p value < 10^{-4} in the SCDC MWAS (binomial sign test, p value = 0.004). Supporting this, we observe an enrichment for genes that are dysregulated in the post-mortem autism brain (one-sided Wilcoxon rank-sum test, p value=6.22×10^{-5}). Finally, integrating genome-wide association study (GWAS) data for autism with mQTL maps from cord-blood demonstrate that mQTLs of CpGs associated with SCDC scores at p value thresholds of 0.01 and 0.005 are significantly shifted toward lower p values in the GWAS for autism. Our results highlight the shared cross-tissue methylation architecture of autism and autistic traits and demonstrate that mQTLs associated with differences in DNA methylation associated with childhood autistic traits are enriched for common genetic variants associated with autism and autistic traits.

7.1 Introduction

Autism is a neurodevelopmental condition characterized by social-communication difficulties, unusually re- strictive, repetitive behavior and narrow interests, and sensory difficulties [6], [335]. The condition can be thought as a continuum, with autistic traits being normally dis- tributed in the general population, and autism at the extreme end of the continuum [38], [336], [337] . Both autism and autistic traits are highly heritable [338]–[341] , with variation across the allelic spectrum associated with the condition [21], [39], [41]. Despite a significant SNP heritability (Autism: h^2_{SNP} – 0.49 [39] – 0.12 [41]), recent studies have demonstrated that the variance explained per SNP is small, suggesting a highly polygenic architecture [41], [342]. None of the significant SNPs associated with autism alter protein coding, suggesting that gene expression is regulated through other mechanisms [41], [63]. For instance, a recent genome-wide association study (GWAS) of autism has identified an enrichment of GWAS signals in H3K4me1 histone marks, particularly in brain and neural cell lines [41], [342].

Previous studies have investigated autism associated methylation signatures in both peripheral tissues [63]–[65] ($50 < N < 2917$) and in the post-mortem brain [66]–[69] ($31 < N < 81$). While post-mortem brain is pertinent for a neurodevelopmental condition like autism, it is not readily accessible, and will be confounded by post- mortem effects on DNA methylation. Studies of methylation signatures in post-mortem brains in autism have replicable identified differential methylation [66]–[69]. Further, they have demonstrated an enrichment for

differentially methylated signatures in the immune system, synaptic signalling, and neuronal regulation [66], [67], [69]. In contrast, recent large-scale analysis of three different peripheral tissue datasets have not identified significantly differentially methylated CpG sites in autistic individuals compared to typically developing individuals [63], [65]. The lack of significant results in peripheral tissues may be attributable to small effect sizes, and significant heterogeneity in both CpG methylation and autism.

While a few studies have investigated DNA methylation underlying autism, to our knowledge, there has been no study investigating DNA methylation under- lying autistic traits in the general population, which are subthreshold manifestations of the autism phenotype. One measure of autistic traits is the Social and Communication Disorders Checklist (SCDC) [70]. Scores on the SCDC are associated (Cohen's D = 2.8, p value < 0.001) [70] and genetically correlated with autism (r_g ~ 0.3) [38], [70], [71]. The SCDC has a modest SNP heritability (h^2_{SNP} = 0.24, s.e. = 0.07) [72] , and polygenic scores for autism are associated with SCDC scores in the general population (max R^2 = 0.13%) [71]. An advantage of using a continuous measure of autistic traits is that it captures the underlying variance better and minimizes heterogeneity attributable to different diagnostic criteria and practices used to diagnose autism.

One potential mechanism through which common genetic variants can regulate gene expression is through DNA methylation. DNA methylation is partly heritable ($0.05 < h^2_{twin} < 0.19$, defined as the proportion of variance in methylation that is attributable to genetics) [73]–[75]. A few studies have integrated genetics and methylation to identify convergent signatures in autism. Andrews and colleagues demonstrated that autism associated GWAS loci are enriched for methylation QTLs (mQTLs) in fetal brain and blood, suggesting that at least some of the gen- etic loci associated with autism may contribute to the condition through differential methylation [76]. In line with this, Hannon and colleagues demonstrated that polygenic risk for autism is associated with differential methylation at birth [63]. While these studies have demonstrated a role for common genetic variants associated with autism and influencing methylation, to our knowledge no study has investigated if methylation of CpGs associated with autistic traits are enriched for common genetic variants associated with autism or autistic traits. One way to test this hypothesis is using mQTLs. We hypothesized that mQTLs of significant CpGs in a methylome-wide association study (MWAS) of autistic traits will be enriched for lower p values in a GWAS of autism or autistic traits.

To address these questions, we investigated the association of CpG methylation in cord blood using scores on the SCDC at age 8. The use of cord blood CpGs minimizes (though, does not eliminate) reverse causation (where the phenotype influences DNA methylation), as the methylation of CpG sites is measured very early in life. To investigate how comparable an MWAS of an autistic trait is to other MWAS of autism and related phenotypes conducted across different tissues, we investigated the overlap between the MWAS of SCDC and other

MWAS of autism and communication-related traits in peripheral and post-mortem brain tissues. We further investigated if genes that are transcriptionally dysregulated in the post-mortem autism brain are enriched for methylation CpGs associated with SCDC. Finally, integrating GWAS data for autism from 46,350 individuals, we investigated if mQTLs of CpGs associated with SCDC scores at various p value thresholds are significantly shifted toward lower p values in the autism GWAS. We validated these results using a smaller GWAS for SCDC.

In summary, this study had three specific aims: (1) to investigate if an MWAS for autistic traits identifies significant CpG methylation and (2) if it is comparable to MWAS of autism; and (3) to investigate if mQTLs of CpGs associated with autistic traits at various p value thresholds are enriched in GWAS of autism and autistic traits.

7.2 Results

7.2.1 Methylome-wide association study of the SCDC scores

Methylome-wide association analysis (Methods, model 1) did not identify any significant loci after Bonferroni correction (p value $< 1 \times 10^{-7}$). The top CpG site was cg14379490, on chromosome 9 (MWAS beta = – 1.78 ± 0.35, p value = 5.34×10^{-7}), which is equivalent to a 0.51 standard deviation unit decrease in SCDC scores. This CpG site is an "Open Sea" CpG site, whose closest gene is FAM120A, which encodes a scaffold protein that is expressed in a wide

number of human tissues. We identified 19 CpG sites with suggestive p values (p value < 1×10^{-4}) (Additional file 1: Table S2). The QQplot and the Manhattan plot are provided in Figure 7-1.

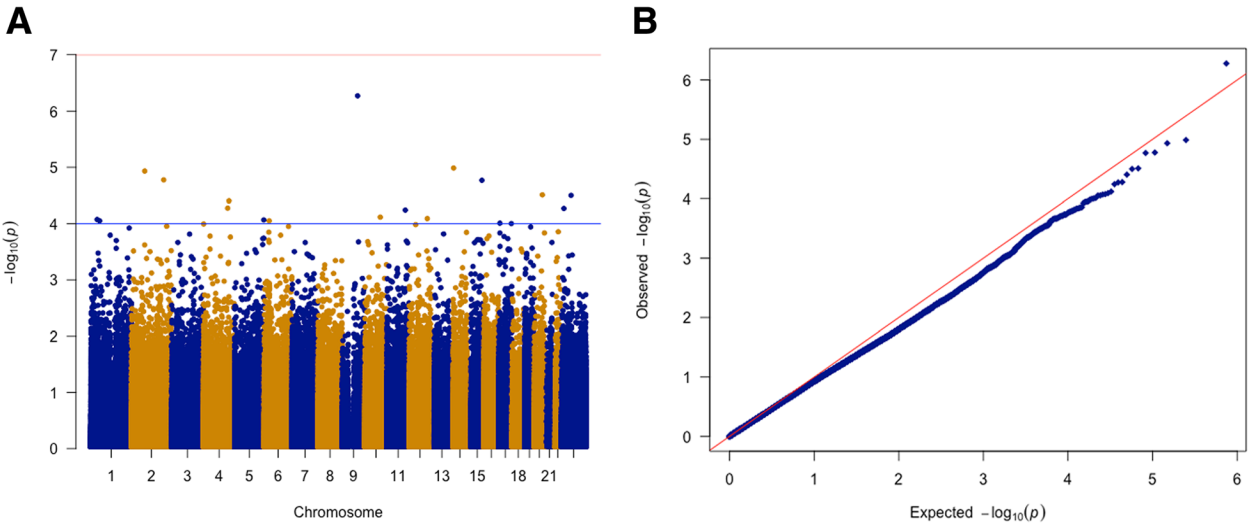


Figure 7-1: Manhattan plot and QQ plot for the SCDC MWAS. A) Manhattan plot of the SCDC MWAS. The blue line indicates significance (p value < 1 x 10-4), and the red line indicates the threshold of statistical significance after multiple testing correction B) Quantile-Quantile plot of the SCDC MWAS

We did not find any evidence for inflation in p values (λ = 0.88), possibly because of the relatively small sample size and the regression model used. Further, gene-set analysis also did not identify significant association after correcting for multiple testing (Additional file 1: Table S3). To confirm that the results are robust to methodological differences, we re-ran the MWAS by using M values instead of beta values and using a different cell count estimate which included nucleated RBCs (Methods, model 2). There was a high correlation in Z-scores (r = 0.92, 95% CI = 0.92–0.92, p value < 2.2×10^{-16}) between the two models. Subsequent analyses were conducted using model 1 MWAS results as these are easier to interpret.

To provide confidence to our primary analyses, we con- ducted an MWAS of scores on the CCC, which was re- versed scored to identify difficulties in communication (Methods). The most significant CpG was cg13711424 (MWAS beta = – 3.73 ± 0.71, p value = 1.79×10^{-7}), equivalent to a 0.55 standard deviation unit decrease in CCC scores. The Manhattan plot and QQ plot are included in Additional file 1: Figure S2. Of the 19 SCDC-associated CpGs of suggestive significance (p value < 1×10^{-4}), the effect direction was concordant for 18 of them in the CCC MWAS (p value = 7.62×10^{-5}, binomial sign test). Similarly, of the 32 CpGs with p value <1×10^{-4} in the CCC MWAS, 28 had concordant effect direction in the SCDC MWAS (p value = 1.93×10^{-5}, binomial sign test). Scores on the CCC and the SCDC were phenotypically correlated (r = 0.39, 95% CI = 0.32–0.45, p value < 2.2×10^{-16}) in the participants who were included in the MWAS (n = 666), and both questionnaires measure difficulties in pragmatic communication. Given that we were testing correlated phenotypes in the same cohort using CpG methylation measured in the same tissue, we hypothesized that the MWAS for the two phenotypes will be positively correlated. The Z-scores for the MWAS of the two phenotypes were significantly correlated (r=0.157, 95% CI= 0.153–0.160, p value < 2.2×10^{-16}), which increased if we considered only CpGs with p value < 0.01 in either one of the phenotypes (P_{SCDC} < 0.01: r = 0.40, 95% CI = 0.36–0.43, p value <2.2×10^{-16}, P_{CCC} <0.01: r=0.40, 95% CI=0.37– 0.42, p value < 2.2×10^{-16}).

7.2.2 Enrichment analyses with peripheral blood methylation signatures

To investigate if there is an overlap between the SCDC MWAS and MWAS of autism conducted in peripheral tissues, we conducted effect direction concordance ana- lysis with three autism MWAS datasets (MINERvA, SEED, and SSC, Methods). For all of them, we first investi- gated concordance in effect direction of all CpG sites with p value $< 1 \times 10^{-4}$. In contrast to the findings with the CCC MWAS, we did not identify a significant concordance in effect direction between the SCDC MWAS and any of the other three autism MWAS datasets. Comparing the three MWAS datasets to each other, we did not identify a significant concordance in effect direction for the suggestive CpGs in each of the comparisons (Table 7-1).

Table 7-1: Sign concordance of the SCDC MWAS and the three peripheral tissues MWAS at top loci (p value < 1 x 10^{-4}

		Testing dataset				Number of CpGs in the discovery dataset
		ALSPAC	MINERvA	SEED	SSC	
Discovery dataset	*ALSPAC*	*NA*	*10*	*11*	*4*	*17*
	MINERvA	*14*	*NA*	*17*	*19*	*29*
	SEED	*19*	*21*	*NA*	*18*	*37*
	SSC	*19*	*21*	*23*	*NA*	*47*

The table above provides the results of the tests for effect direction concordance. CpGs with p value $< 1 \times 10^{-4}$ in the discovery dataset were tested for concordance in effect direction in the testing dataset. The numbers in the cells (in italic) provides the total number of CpGs with concordant effect direction in the testing dataset. The number of CpGs in the discovery dataset provides the total number of CpGs in the discovery dataset with p value $< 1 \times 10^{-4}$. None of the results were significant (binomial sign test) after correcting for the multiple tests conducted

Given that there was limited evidence for concordance in effect direction between the datasets, we next tested if nominally significant CpGs (p value < 0.01) in the three autism MWAS have a shift toward lower p values in the SCDC MWAS using a one-sided Wilcoxon-rank sum test. This tests more CpGs than an effect direction concordance test and is agnostic to effect direction which may be discordant in different peripheral tissues measured at different developmental stages. After Bonferroni correction (alpha = 0.016), we did not identify a significant shift toward lower p values for the nominally significant CpGs from any of the three datasets (SEED: p value = 0.02; SSC: p value = 0.48; MINERvA: p value = 0.91), though we note a nominally significant shift in the SEED dataset. This lack of overlap may be due to the low statistical power of the MWAS of SSC and the three autism MWAS, none of which have identified significantly differentially methylated CpGs.

7.2.3 Enrichment analyses with autism post-mortem brain methylation signatures

Methylation signatures in post-mortem brain tissues are more relevant to neurodevelopmental phenotypes than methylation signatures in peripheral tissue, and, for autism, are statistically better powered than MWAS in peripheral tissues, as these have identified more differentially methylated loci compared to peripheral tissue analyses [66], [69]. Considering this, we investigated if there is an overlap between the SCDC MWAS and MWAS of the post-mortem autism brain. Using data from the latest post-mortem brain study [69], we investigated concordance in effect direction between all CpG probes with p value < 1

$\times 10^{-4}$ from the cross-cortex analysis in the SCDC MWAS. Further, 171 out of 293 CpGs had a concordant effect direction in the two datasets (p value= 0.004). At a more stringent p value threshold of p value $< 1 \times 10^{-5}$, 88 of the 133 probes had concordant effect directions in the two datasets (p value = 2.4×10^{-4}, binomial sign test).

In contrast, Wilcoxon rank-sum test of all CpGs with p value < 0.01 in the post-mortem MWAS did not identify a significant shift toward lower p values (p value = 0.99, one-tailed Wilcoxon rank-sum test). We next tested if the results were supported in a different dataset. A previous study investigated differential methylation in post- mortem neurons from the frontal lobe (identified using FACS sorting) in autism [66]. First, testing effect direction concordance, 44 of the 87 CpGs with p value $< 1 \times 10^{-4}$ had concordant effect direction in the two datasets (p value = 1, binomial sign test). However, we identified a significant shift toward lower p values (p value=9.3×10^{-3}, one-tailed Wilcoxon rank-sum test) of all CpGs with p value < 0.01 in the SCDC MWAS.

7.2.4 Enrichment with autism dysregulated genes

A few studies have identified consistent sets of dysregulated genes in autism, and co-expression modules enriched for these dysregulated genes [249]–[252]. Previous studies have identified a significant enrichment for differentially methylated autism CpGs in genes that are transcriptionally dysregulated in the post-mortem cortex in autism [63]. We investigated if

CpGs mapped to transcriptionally dysregulated genes in the autism post- mortem cortex [261] and associated co-expression mod- ules had a shift toward lower p values in the SCDC MWAS when compared to the other genes. We identified a significant shift toward lower p values for the transcriptionally dysregulated genes (one-sided Wilcoxon rank-sum test, p value = 6.22×10^{-5}), but did not identify a significant enrichment for any of the modules (M4: p value = 0.58, M9: p value = 0.59, M16: p value = 0.042, M10: p value = 0.31, M20: p value = 0.42, M19: p value = 0.105).

7.2.5 Genetic influences in SCDC methylation patterns

We next investigated if CpGs associated with SCDC scores are enriched for GWAS signals for autism. DNA methylation is under cis and, to a smaller extent, trans genetic control. We identified mQTLS associated with SCDC CpGs below four p value thresholds (P_{SCDC}, Methods), and compared the distribution of p value of these mQTLS in the autism GWAS against the p value distributions of mQTLs above the P_{SCDC} (Methods). After multiple testing correction, mQTLS of CpGs with P_{SCDC} = 0.01, and 0.005 have significantly lower p values in the autism GWAS (P_{SCDC} 0.01: FDR-corrected p value = 5×10^{-4}, P_{SCDC} 0.005, FDR-corrected p value = 4.75×10^{-3}) (Table 2, Fig. 2).

We provide additional support for this enrichment in a GWAS of SCDC, which is genetically correlated with autism. We identified an en- richment at P_{SCDC} 0.005 (FDR-corrected p value

= 0.046) and at P_{SCDC} 0.001 (FDR-corrected p value = 0.046). In contrast, we did not identify an enrichment for mQTLs in the Alzheimer's GWAS (Table 7-2, Figure 7-2).

Table 7-2: Results of the enrichment analysis of the top CpGs

CpG value threshold	GWAS	*p* value	Mean difference	FDR-corrected *p* value
.05:	Autism	0.54	0.004	0.54
0.01:	Autism	1.0×10^{-4}	0.039	5×10^{-4}
0.005	Autism	1.9×10^{-3}	0.031	4.75×10^{-3}
0.001:	Autism	0.040	0.037	0.063
0.05:	SCDC	0.735	0.007	0.73
0.01:	SCDC	0.301	0.015	0.40
0.005	SCDC	0.022	0.038	0.046
0.001:	SCDC	0.023	0.065	0.046
0.05:	Alzheimer's	0.343	0.008	0.853
0.01:	Alzheimer's	0.710	0.003	0.853
0.005	Alzheimer's	0.853	-0.003	0.853
0.001:	Alzheimer's	0.793	-0.009	0.853

The table provides the results of the enrichment analyses for the top loci. We calculated the difference between the average p values for all the mQTLs mapped to CpGs below a selected threshold in the MWAS (CpG p value threshold) and the mQTLs mapped to CPGs above the threshold in the SCDC MWAS. This value is referred to as "mean difference" in the table. A positive difference suggests and enrichment. We then permuted the results after correcting for various factors and computed a permuted p value (p value). We then corrected it for multiple testing using FDR correction (FDR-corrected p value). This was done using a GWAS for autism, SCDC, and Alzheimer's

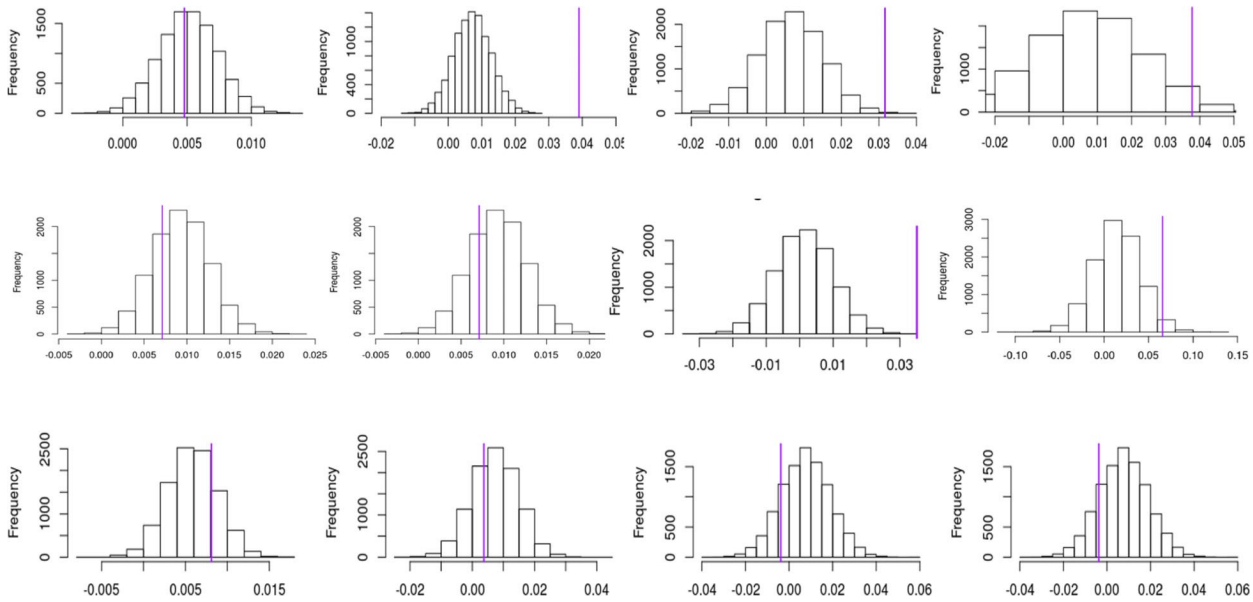


Figure 7-2: Permutation histogram of SNP-enrichment in top CpGs for three GWAS. The graphs present the results of the permutation analysis of the SNP enrichment. a Results of the autism GWAS. b Results of the SCDC GWAS. c Results of the Alzheimer's GWAS. p value thresholds of the CpGs for enrichment are provided at the bottom of each column. Y-axis of each plot represents the frequency of the difference in mean p value of the mQTLs of CpGs below the threshold from the mean p value of the mQTLs of the CpGs above the threshold. X-axis represents the differences in the mean p value of the mQTLs of CpGs below the threshold from the mean p value of the mQTLs of the CpGs above the threshold. A higher difference in the means indicates a greater enrichment. Purple lines indicate the difference in mean of the non-permuted data point, i.e., the actual difference in mean

7.3 Discussion

This study investigated the shared biology of autism and autistic traits by integrating genetic, methylation, and post-mortem gene expression data. We first investigated the validity of considering autistic traits for methylation studies. Considering autistic traits over a case-control design is useful in that (1) it captures greater variance across the underlying liability spectrum, (2) it can be used to increase sample sizes by phenotyping individuals for whom methylation data is available, and (3) it can be used to link methylation signatures from tissues col- lected in early life to the phenotype, as this can be more difficult for autism.

We conducted a prospective MWAS of autistic traits (SCDC) by measuring methylation signatures in the cord blood and linking it to autistic traits measured later in life. While we did not identify a significant CpG association with autistic traits after multiple testing correction, we were able to confirm that this analysis produced biologically several differences between the SCDC MWAS analysis and the three MWAS for autism. Of primary importance is the statistical model used in the analysis. While we were interested in investigating if methylation signatures from cord blood were associated with SCDC scores measured in later life, all three-peripheral tissue MWAS investigated if autism diagnosis was associated with differential methylation.

Thus, in our analysis, methylation was an independent variable, whereas in the three MWAS for autism, methylation was a dependent variable. Second, there are remarkable differences in age at which methylation was measured, and confounding variables included in the analyses; for instance, we included genetic principal components as covariates. Third, there are differences in tissue source as well. While the MINERvA cohort primarily used blood spots, the SEED and the SSC cohorts used whole blood for the respective MWAS. In comparison, our MWAS was conducted using cord blood. Fourth, while the SCDC MWAS was con- ducted in individuals of European ancestry, the SEED and SSC cohorts also included individuals of non-European ancestries.

Fifth, while both the SCDC MWAS and the MINERvA MWAS had largely balanced sex ratio, the SEED and SSC cohorts had more male autistic individuals than females, though sex was included as covariate in these cohorts. None of the three autism MWAS demonstrated a significant overlap with each other as investigated using a sign-concordance test of the most significant CpGs. It is critical to investigate this observed lack of concordance, though it may be driven by the low statistical power of each individual MWAS, similar to the early GWAS studies which were underpowered.

In contrast to the results from the peripheral tissues, we observed some degree of overlap between MWAS conducted in post-mortem brain tissues [66], [69] and the SCDC MWAS. First, we found a significant sign concordance in CpGs identified in the largest cross- cortex MWAS of autism using post-mortem tissue samples. However, we did not identify an enrichment using a Wilcoxon rank-sum test of p values. In contrast, using a neuron-specific MWAS generated using a different post-mortem tissue dataset, we identified a significant overlap using a Wilcoxon-rank sum test of p values but not a significant sign-concordance. Additionally, using an RNA sequencing dataset of autism and neurotypical post-mortem brains [261], we identified a significant enrichment for transcriptionally dysregulated genes using a Wilcoxon rank-sum test. Overall, we are unable to strongly suggest that there is a significant overlap be- tween the SCDC MWAS and the MWAS of autism in either post-mortem or peripheral blood tissues. This is likely due to multiple factors as outlined earlier.

In addition, measuring methylation in peripheral tissue, which is not necessarily a relevant tissue for a neurodevelopmental condition like autism, is likely to attenuate the signal-to-noise ratio. Indeed, the post- mortem brain MWAS study [69] has identified significant CpGs with fewer samples compared with any of the three peripheral tissue MWAS [63], [65]. Thus, due to both the increased statistical power and the use of a relevant tissue, the top CpGs in the post-mortem brain MWAS are more likely to be true positives than the top CpGs in the peripheral tissue MWAS.

Given the highly polygenic nature of autism [41], it is likely that GWAS loci that are not statistically significant in the current GWAS studies may still influence methylation. Thus, the second aim of this study was to investigate if the top CpG sites in the SCDC MWAS were enriched for GWAS signals for autism and autistic traits. Our results demonstrate an enrichment for mQTLs for CpGs associated with SCDC scores in the GWAS for autism. We were able to provide additional support for the results in a much smaller GWAS of SCDC scores but failed to identify an enrichment in a GWAS of Alzheimer's [263], which is of comparable statistical power to the GWAS of autism. This enrichment is observed at more stringent p value thresholds providing confidence in our results. We did not test this in other peripheral tissue MWAS for which we had access to summary statistics given the lack of overlap between these and the SCDC MWAS.

7.3.1.1 Limitations

Our study does not investigate causality. While methods such as Mendelian randomization can investigate causality [63], [343], this is typically restricted to a few number of loci based on current results of GWAS studies. In addition, we are restricted from using Mendelian randomization due to the low statistical power of both the MWAS and the GWAS datasets, resulting in the identification of a limited number of statistically significant loci. Two mechanisms may explain the overlap observed in the current dataset. The first is causal wherein genetic loci are likely to influence autism or autistic traits by influencing methylation of CpG sites, altering gene expression levels. The second is horizontal pleiotropy, where genetic loci are associated with autism or autistic traits, and separately, also influence methylation levels of CpG sites. This study cannot tease apart these two mechanisms.

A few caveats must be borne in mind while interpreting the results of this analysis. First, the current array-based method interrogates only a small proportion of all CpG sites in the genome. Thus, significant loci associated with autistic traits may lie outside of the regions interrogated. Second, due to the nature of the assay, the methylation values may also capture hydroxymethylation. We cannot exclude the possibility of signal attenuation due to assaying both hydroxymethylation and methylation in the current study, and the correlation between hydroxymethylation between blood and brain is low [344] . Third, while there is a modest but significant genetic and phenotypic correlation between autism and scores on the SCDC the SCDC only measures social aspects of autism and is not correlated with the non-social aspects of autism. Finally, age of gestation was not available to include as a covariate, and thus the current study does not account for it.

Chapter 8

General Discussion

8 General Discussion

8. 1 Part I: Cellular Phenotypes Associated with NRXN1 Mutations in Autism; an iPSC Study

The role of *NRXN1* gene has been established in regulating normal synaptic function and physiology, and functional changes in this gene have been associated with neuropsychiatric conditions including autism and schizophrenia. What remains poorly understood is its contribution to their aetiology. Since a reliable cellular model is crucial to dissect the effects of *NRXN*1 alterations on autism related phenotypes, two iPSC models were generated: 1) induced mutations in CRISPR-edited cell lines, 2) Autism-derived iPSC with *NRXN*1 mutations.

In this thesis I hypothesised that mutations in the patient lines and the mutant lines have diverse effect on neurite outgrowth, ligands expression, neuronal activity and affect the transcriptome of glutamatergic induced neurons.

Most reported clinical *NRXN1* mutations affect only the *NRXN1α* isoform, however, some cases report having both isoforms (α and ß) affected [133]. In this study, a single gRNA used was designed to target the first shared exon by both isoforms (Exon 19). This generated random mutations via non-homologous end joining generating 4 mutant lines: *NRXN1*-mutant G4 had

1 bp insertion in both alleles and the A2 mutant lines had different mutations on both alleles, a 2bp on one allele and 4 bp on the other one. The A7 mutant line had a 1bp insertion in one allele while the second was a wild type. Finally, the E9 mutant line had the biggest deletion (115 bp) on one allele but was wild type on the other allele.

The *NRXN1*-patient lines on the other hand had mutations on the *NRXN1α* isoform. The 092NXF-NGN2 has a 200 bp deletion in intronic region 5 – it spans more than just intron 5 – it also deletes several exons as well. This deletes a regulatory element miR8485 (a miR-NID1: *NRXN1* intron derived miRNA), located within intron 5. MiR8485 is known to bind to Tar DNA-binding proteins, absence of which has been shown to negatively regulate *NRXN1* expression [314]. MiR8485 specifically binds to TDP-43 to regulate *NRXN1* expression. TDP-43 binding RNAs are associated with neuron differentiation and development in humans [314]. It is unknown if variants in miR-NID1 can have an indirect effect on TDP-43 regulatory functions, as it is known to bind to miRNAs processed from multiple neuronal genes such as *NRXN1*, DSCAM and RIMS2 [314]. The 211_NXM was diagnosed with high-functioning autism, who was identified to carry a single nucleotide variant resulting in a missense mutation (G→C) on chromosome 2 (50724586) falling within exon14 in the NRXN1α isoform.

The patient lines containing *NRXN*1 mutations were targeted with the OPTi-OX system to allow the rapid and deterministic forward programming of the patient lines

into a homogenous neuronal culture. The TET-ON system was used for the inducible overexpression of NGN2. To circumvent the random transgene integration in the genome by lentiviral-based protocol; the TET-ON elements were inserted in two different GSH loci using CRISPR-Cas9 to target the human ROSA26 locus with the CAG-*rtTA* cassette, and zinc-finger nuclease (ZFN) to target the AAVS1 locus with the TRE-NGN2 cassette [229], [230]. The forward programming of these iPSC lines with different NRXN1 mutations provide an invaluable tool to study effect of different NRXN1 deletions on molecular, morphological and electrophysiological properties of the neurons.

8.1.1 General findings

The first aim was to look for differential isoform expression across time in each cell line generated. The two cellular models included in this thesis have mutations in different regions within the NRXN1 gene. *NRXN*1β was expressed significantly higher than that of *NRXN*1α in the induced neurons in all cell lines. It has been previously suggested that *NLGN*1 binding to α and β neurexins serves as a signalling code to specify the type of synapse formed [111], [170]. Many studies suggested that *NLGN*1 binding to *NRXN*1β (lacking an insert in SS#4) restricts its function to glutamatergic (and not GABAergic) neurons [315], [316]. It is therefore expected to observe a higher expression of *NRXN*1β in an induced-NGN2 glutamatergic neuronal population. It was also important to look at the temporal gene expression profile of *NRXNs* and their

binding partners (*NLGNs* and *CASK*). Consistent with findings in post-mortem tissue across different brain regions in post-mortem tissue (8-12 pcw), the NGN2-induced neurons in the *NRXN*1-isogenic and patient lines expressed *NRXN*3 at significantly lower levels than *NRXN*2 [313]. Moreover, *NLGN*2 and 3 were predominantly expressed in the investigated lines while *NLGN*2 is lowly expressed and increased with age, consistent with Harkin's finding in post-mortem tissue [313]. This highlights and supports the notion that *NRXN* and their binding partners expression is critical during neurogenesis and synapse formation, and that it is not affected by NRXN1 mutations.

The second aim of the study was to look at morphological alterations in neurite outgrowth. Since several studies implicated *NRXN*1 in Neuritogenesis, I hypothesized that the mutations found in the autism patient lines as the CRISPR-edited lines will exhibit atypical growth of neurites. Three *NRXN*1-mutant lines (A2, E9 and G4), and one of the *NRXN*1-patient lines (092_NXF), showed atypical neurite outgrowth at Day 6. This was associated with a reduced *NRXN*1 and increased *NLGN*1 mRNA levels when compared to the control at Day 7. Although this did not reach statistical significance, it warrants some further studies and experiments to explore the possible link between these molecules and the morphological changes reported in our lines. It is possible that disruption in the *NRXN-NLGN* complex lead to alteration in neurite outgrowth. Several studies that successfully showed that these complexes are involved in neurite outgrowth induction in Xenopus, Drosophila, rat hippocampal neurons and neural stem cells [137], [189]–[191], [317].

Findings from RNA-Seq data revealed that a large number of genes from the collagen superfamily was dysregulated in the patient and mutant lines when compared to control. Members of this family have been extensively involved in neural development, including neurite outgrowth and synaptogenesis [290].

The third aim was to investigate the effect of the NRXN1 mutations on neuronal activity. Multiple studies reported altered calcium signalling in iPSC-derived neurons with *NRXN1α* heterozygous or homozygous mutations [180], [181], [195]. Avazzadeh et al used iPSCs derived from autistic individuals with heterozygous NRXN1α mutations, and they reported increased amplitude, frequency and duration, this was concordant with their RNA-Seq data which revealed dysregulation of voltage-gated calcium channel genes [181]. This is in disagreement with another study that reported a marked reduction in calcium transients in an iPSC line with a bi-allelic NRXN1 α mutation [180]. Both of these studies relied on patient-derived iPSCs with mutations in the NRXN1α isoform.

In this study, the NRXN1-patient lines also had mutations within the NRXN1α isoform, however, they showed no effect on calcium transients' amplitude nor duration. They however presented an interesting phenotype: they both showed a hypersynchronous firing but only the 092_NXF reached statistical significance when compared to the

control iNs. 092_NXF also showed atypical neurite outgrowth at Day 6 and bulk RNA-Seq and ORA revealed that cell adhesion and extracellular matrix organization pathways are dysregulated at Day 35. The variability of the phenotypes presented reflects the complexity and incomplete penetrance of the NRXN1 genes and added to this is the involvement of the patient's genetic background in the development of the presented phenotypes. Isogenic lines allow scientists to dissect the role of the mutation by comparing it to the wildtype, having an identical genetic background except for the introduced mutation.

Pak et al investigated two different mutation, a deletion in the first shared exon between the two major NRXN1 isoforms (found in some autism cases) and a truncation mutation in Exon 24 and induced neuronal differentiation using NGN2-forced expression. In their study they used adeno-associated viruses to introduce the mutations in human embryonic stem cells, they flanked the target exon sequence with loxP sites to facilitate homologous direct repair that results in the deletion of the whole exon, disrupting the reading frame and causing a loss of function of NRXN1 [195]. The neurons derived from both cellular models exhibited a similar phenotype where neurotransmitter release was impaired and spontaneous mEPSC was diminished when compared to the isogenic control [195].

Similar to this study, the same exon was targeted to induce a frameshift mutation, but unlike Pak et al, the mutations I introduced were via non-homologous end joining and thus creating random indel mutations in the target sequence, the whole exon was not deleted rather a disruption in the reading frame was intended to cause a loss of function of NRXN1. Contrary to the findings of the 2015 study, the NRXN1-mutant lines generated in this thesis did not show differences in calcium transients. It is however important to point out that the mutation in their study was a different mutation (although in the same region) and that they used whole-cell patch-clamp to record individual neuron activity while this study relied on calcium imaging for intracellular recording of network activity [195], [345].

Moreover, the 092_NXF line, the A7 mutant line presented hypersynchronous bursts that were significantly higher than the control and the other mutant lines. This mutant line (A7) has a marked reduction in NRXN1β mRNA levels but reported no alteration in neurite growth. Although most studies do not report the hypersynchronous phenotype in iPSC-derived neurons, alterations in neural synchrony have been reported in autism and other neurological conditions such as schizophrenia and epilepsy-[327], [328].

Extracellular activity was recorded using MEA. Interestingly, the 211_NXM patient line showed increased mean firing rate when compared to control and to the 092_NXF

line. The 211_NXM line is the only line in this study that showed higher expression of NRXN1 as revealed by RNA-Seq data. Other studies looking at mutations affecting NRXN1 isoforms and reported reduced number of spikes in induced neurons 3 weeks post induction [129] while Deneault et al report no effect on firing patterns in neurons with NRXN1 mutations [325]. Consistent with Deneault's findings, the A2 and A7 mutant lines showed no differences in mean firing rate when compared to control, this might be due to the different type of mutations in the other studies that might result to a different phenotype, or might be due to a technical artefact in the experiment design as in the duration of the recording (15 minutes) is crucial for identifying difference between cell lines; recent studies now suggest extending the duration to 30 minutes for a more accurate depiction of neuronal activity [334].

Finally, we took a look at the effect of these mutations on the transcriptome and the pathways involved. First, we looked for overlapping differentially expressed genes, and found a 50% overlap between the 092_NXF and the 211_NXM. The overlapping genes comprised a large number of collagens. Similarly, 27 out of 29 members in the collagen family were upregulated in the NRXN1-mutant line A2. This reflected the important effects of NRXN1 on an important family of genes, that has been a cardinal component in neural development [290]. This suggests that the alterations observed in neurite outgrowth and functional activity of the neurons might be an indirect effect of the NRXN1 mutations mediated by the dysregulation of collagens.

In autism animal model studies such as in the valproic acid rat model, authors have reported upregulation of collagens in the prefrontal cortex [291]. This resulted in atypical neuronal migration and layering of the cortical areas in some autism cases [292], [293]. Moreover, 56 and 65 downregulated genes in 092NXF and 211NXM respectively were highly enriched for cell adhesion, the biological process that describes the attachment of a cell to extracellular matrix molecules. This was previously reported in NRXN1-knockdown neurons and autism prefrontal cortex [276], [294]. Other RNA-Seq studies in human neurons reported the dysregulation of calcium binding, ion transport, transporter activity, and voltage- gated channel complexes in autism-derived NRXN1$\alpha^{+/-}$ neurons [181].

Another family of genes that was severely affected in the patient lines was the cadherin superfamily. Cadherins are synaptic molecules that primarily act as calcium dependent cell adhesion molecules and are abundantly expressed in the brain (Yagi and Takeichi 2000). They have been implicated in intracellular signalling and have been associated with many neuropsychiatric conditions such as autism and schizophrenia [297]. In autism prefrontal cortex, alteration in calcium ion binding, cell-cell signalling and synaptic transmission pathways, the same pathways are often associated with dysregulation of cadherins [277], [294]. The downregulated genes in the NRXN1-mutant lines were highly enriched in chemical synaptic transmission, cell-cell signalling and generation of neurons pathways, consistent with previous studies using NRXN1-mutated cellular models [100], [129], [180], [195].

8.1.2 Individual cell lines' findings

Amongst the NRXN1-mutant line, the mutation in the A2 mutant line seemed to have the most prominent effect on NRXN1 expression, as revealed by RNA-Seq analysis. Similarly, TaqMan analysis showed that both isoforms were significantly lower than that of the control. It also impacted neurite length, count and branch point count. Moreover, transcriptome analysis reported downregulation in genes involved in chemical synaptic transmission (GO:0007267), cell-cell signalling (GO:0007268) and neuronal system (R-HAS-112316). Functional analysis, however, reported no difference in calcium imaging nor multielectrode array recordings for the same line (A2). It is noteworthy to mention that the RNA-Seq was performed on Day 21 iN while the electrophysiological activity was recorded at Day 35, due to the method used for neural induction and the mutation in NRXN1, a delay in neuronal maturation has been often reported, this might explain why in some cases the results reported by RNA-Seq at Day21 iN are not complemented by the functional analysis done at Day35 [129], [231], [276]. The mutation in the mutant line A7 significantly reduced the expression levels of NRXN1-α but had no major effects on neurite outgrowth. Interestingly, it showed a hypersynchronous firing as revealed by calcium imaging on Day35 neurons, this is a novel phenotype that was only replicated in one of the patient lines (092_NXF).

No significant changes in *NRXN1* expression were reported in the E9 mutant line although it had the largest deletion of all the mutant lines and showed a marked reduction in NRXN1 α isoform. Morphological analysis revealed significant disruption in neurite outgrowth as a result of the NRXN1 deletion. Interestingly, Day 21 neurons showed a significant dysregulation of genes related to generation of neurons which include genes important for neuronal morphology and neurite outgrowth. The downregulated genes in the E9 mutant were also enriched for the cell-cell signalling pathway while no significant difference was reported in calcium signalling recorded on Day 35.

No significant reduction in NRXN1 expression levels at Day 21 was reported in the G4 mutant line, however a Taqman assay showed that the overall expression of NRXN1, NRXN1α and NRXN1β across time was significantly reduced. This was consistent with a pile-up of exonal reads of NRXN1 revealed by the RNA-Seq results. Interestingly, apart from a reduced neurite count, no significant alterations appeared on the neurons' morphology nor disruption in calcium signalling was reported although downregulation of cell-cell signalling, and generation of neurons pathways was reported in Day 21 neurons. Based on the properties investigated in this thesis, this mutant line seems to be the least affected by the NRXN1 mutation and further investigation are warranted to explore how the mutation could have affected this particular line.

The patient lines carry mutations within the α isoform of NRXN1, however they showed very distinct pattern of expression. The point mutation in the 211_NXM line resulted in an increased expression of NRXN1 while the large deletion in 092_NXF resulted in the reduction of NRXN1 exonal reads. The 092_NXF also showed a hypersynchronous firing and altered neurite outgrowth, while the 211_NXM showed no morphological differences but reported an increased firing rate at Day 35. Interestingly, the upregulated genes in 092_NXF and 211_NXM were highly enriched for cell adhesion pathway (GO:0007155) despite having very distinct expression levels of NRXN1 and different genetic backgrounds. It seems like the dysregulation of NRXN1 expression (up or downregulation) results in the upregulation of other cell adhesion related genes as a compensatory mechanism.

8.1.3 Conclusion

In this thesis I used the OPTi-OX system to generate induced neurons from CRISPR-edited and autism-derived iPSC as model system, to investigate the role of NRXN1 in autism related phenotypes, and to test this hypothesis, the cellular models were generated and investigated for molecular, morphological and electrophysiological functions. The results support my hypothesis and confirm that this model system is useful for this type of investigation and furthermore it demonstrated variable and diverse effects of NRXN1 on neuronal profiles.

First, I have established that NRXN1 mutations may have diverse effects on NRXN1 expression levels. While most of the lines show a reduction in NRXN1 expression, only A2 reached statistical significance, and the mutation in 211_NXM line resulted in higher expression of NRXN1, as revealed by TaqMan and RNA-Seq analysis. Moreover, NRXN1 mutation had diverse effects on neuronal morphology and activity. Most importantly, transcriptomic analysis showed that although the mutations had affected the transcription of a different set of genes, their roles and involvement converged into shared pathways which might result in them exhibiting similar phenotypes in functional and morphological assays. It is unclear whether NRXN1 mutations has a direct effect on the cellular phenotypes observed in this study or are these effects are governed by transcriptional effects or other genetic variants associated with NRXN1 dysregulation. In the mutant lines, the observed phenotype reflected direct effect downstream the mutations induced in the NRXN1 gene. Conversely, the same cannot be concluded about the patient lines without correcting the mutations, as the perturbations reported can also result from independent contribution of autism risk variants.

8.1.4 Limitations

8.1.4.1 Novel and/or mutant isoforms that are not detectable using traditional methods.

The abundance and diversity of NRXN1 isoforms made it challenging to quantify using Taq-Man probes. The repertoire of *NRXN*1 isoforms that can be detected, was very limited and did not include all potential variants and, it is known that mutations result in the expression of novel isoform that have not been annotated. This limited

the quantification of *NRXN*1 isoforms, and the development of novel targeted tools is essential to fully examine to effect of mutations on isoform expression. A targeted hybrid sequencing approach was developed by Flaherty et al to overcome those limitations and to thoroughly investigate the effect of mutations on *NRXN*1 isoform expression [129].

8.1.4.2 NRXN1 isoforms are synapse specific

*NLGN*1 binding to α and β neurexins mediate trans-synaptic cell adhesion and are used as a molecular code to specify the type of synapse formed [111], [170]. Recently, it has been suggested that *NLGN*1 induce GABAergic but not glutamatergic synapse formation by binding to *NRXN*α and a subset of *NRXN*1β (with an insert in SS#4), while *NLGN*1 binding to *NRXN*1β (lacking an insert in SS#4) restricts its function to glutamatergic neurons [315], [316]. The induced neurons are excitatory glutamatergic neurons and thus it is expected to have a higher *NRXN*1β expression but not NRXN1α. It is therefore challenging to realistically assess NRXN1α expression using this model.

8.1.4.3 Replicating the Patient mutations to dissect genetic background from effect of mutation

Although recent literature supports the theory linking *NRXN1* mutations in the patient lines to the impaired synaptic transmission reported both functionally and transcriptionally. It is naïve to claim that this is directly caused by the *NRXN1*

mutations for a number of reasons. The induced mutations in the wild type gave rise to various phenotypes. This is probably due to the structural complexity of *NRXN1* and its variable penetrance. It would have been probably more conclusive to replicate the same mutations found in the patient lines in the wild type instead of targeting a different region. The goal from targeting a different region was to produce a loss of function of NRXN1 by targeting a shared exon for both isoforms instead of targeting the α isoform. Moreover, introducing a large deletion like the 092_NXF is technically very challenging, especially in a complex gene like NRXN1.

8.1.4.4 Gene correction

To confirm the genotype-phenotype correlation, it would have been beneficial to correct the mutations in the patient lines. It was not considered in the course of this project because of the technical challenges presented by the *NRXN1* gene size, deletion size and alternative splicing. An alternative would be rescuing the phenotype by regulating *NRXN1* isoform expression or introducing the downregulated isoform.

8.1.4.5 Identity of the neuronal culture

Recently, Rosa et al showed that NGN2-derived neurons do not reach the same level of maturity when compared to other traditional methods (Embryoid bodies) [231]. Moreover, the limited expression of NMDA-R in NGN2-iNs present a hurdle when investigating synaptic transmission pathways as variants in the glutamate ionotropic receptor NMDA type subunit

2A and 2B have been associated with autism [21]. However, the addition of primary rat glia or astrocytes helps improve NGN2-derived neuronal maturity and allows the formation of functional synapses with larger capacitance and smaller input resistance [231].

8.1.5 Future Directions

8.1.5.1 Detect isoform expression levels using targeted hybrid RNA-Seq approach

Due to the extensive alternative splicing of NRXN1, mutations in NRXN1 can result in the generation of hundreds of novel mutant isoforms. This restricts the availability of well documented and annotated isoform directory for NRXN1 and quantifying isoform expression by short-read RNA-Seq is limited. A study in 2019 developed a targeted hybrid (long and short read) sequencing approach to better investigate the effect of NRXN1 mutations on isoform expression [129]. The plan is to undertake this approach to dissect the molecular basis of the detected phenotypes.

8.1.5.2 Adding inhibitory neurons to better recapitulate neuron network activity

The NGN2-derived neurons provide an invaluable and fast tool to generate glutamatergic functional neurons, however, the population is mainly excitatory glutamatergic neurons and a small population of cholinergic neurons. The addition of rat mixed glia or astrocytes aids in the neurons' maturation and synaptic function. In future work, I aim to include inhibitory neurons to the culture to better recapitulate neural network activity especially that one of the phenotypes observed

in this study is hyper-synchronicity which might be masked in the presence of inhibitory neurons.

8.1.5.3 Rescue morphological and functional phenotypes with overexpression of α or β isoforms

It is impossible to conclude the direct effect of NRXN1 mutations on the morphological, molecular or functional effect demonstrated in this study unless the phenotype can be rescued by a correction or an overexpression of the wildtype isoform. Since gene correction for a complex and large gene like NRXN1 is technically challenging, overexpression of NRXN1 isoforms can be used as an alternative to confirm the impact of those mutation on the phenotype in question. If the overexpression rescues the phenotype, then NRXN1 is directly involved in the mechanism leading to the investigated phenotype. This should be a priority for future research.

8.1.5.4 Alternative models to test functional phenotypes of neurons

To validate/explore autism neuron functional phenotype, an iPSC-based programme of study that combine the latest iPSC- derived mature cerebral organoid techniques with cutting edge tools to trace early pre-natal transcriptional events and post-natal neuronal functions. Cerebral organoids are known for their remarkable fidelity in modelling early brain development; however, a big challenging factor is reaching

neuronal maturation, and this is mostly due to insufficient oxygen and nutrient availability. To overcome this challenge, growth factors can be added to the media ot transplanting neural organoids into rodents to allow vascularization. The classic approach of organotypic slice culture at the air liquid interface can be applied to organoids to improve long-term survival to improve oxygen supply. This method has been found to improve overall morphology as well as increased cortical neuron populations, these neurons are functional mature and have the ability to generate networks within the organoid that have the potential to investigate functional phenotypes that may be specific to autism.

8.2 Part II: Integrated genetic and methylomic analyses identify shared biology between autism and autistic traits

Previous studies have identified differences in DNA methylation in autistic individuals compared to neurotypical individuals. Yet, it is unclear if this extends to autistic traits—subclinical manifestation of autism features in the general population. In this study, the aim was to investigate the shared biology of autism and autistic traits by integrating genetic, methylation, and post-mortem gene expression data. We first investigated the validity of considering autistic traits for methylation studies. Considering autistic traits over a case-control design is useful in that (1) it captures greater variance across the underlying liability spectrum, (2) it can be used to increase sample sizes by phenotyping individuals for whom methylation data is available, and (3) it can be used to link methylation signatures from tissues collected in early life to the phenotype, as this can be more difficult for autism.

Firstly the association between DNA methylation at birth (cord blood), and scores on the Social and Communication Disorders Checklist (SCDC), a measure of autistic traits, in 701 8-year-olds, by conducting a methylome-wide association study (MWAS). We did not identify significant CpGs associated with SCDC. The most significant CpG site was cg14379490, on chromosome 9 (MWAS beta=–1.78±0.35, p value=5.34×10–7). Using methylation data for autism in peripheral tissues, we did not identify a significant concordance in effect direction of CpGs with p value < 10–4 in the SCDC MWAS (binomial sign test, p value > 0.5). In contrast, using methylation data for autism from post-mortem brain tissues, we identify a significant

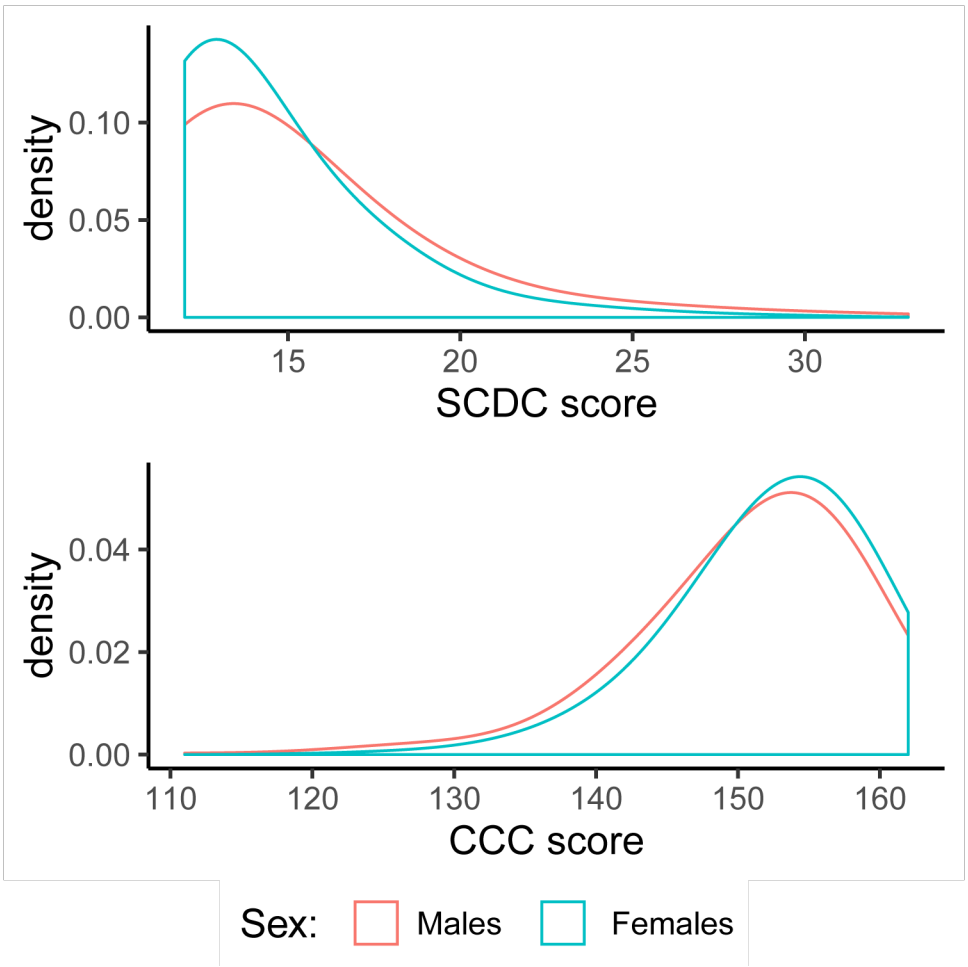
concordance in effect direction of CpGs with a p value < 10–4 in the SCDC MWAS (binomial sign test, p value = 0.004). Supporting this, we observe an enrichment for genes that are dysregulated in the post-mortem autism brain (one-sided Wilcoxon rank-sum test, p value=6.22×10–5). Finally, integrating genome-wide association study (GWAS) data for autism (n = 46,350) with mQTL maps from cord-blood (n = 771), we demonstrate that mQTLs of CpGs associated with SCDC scores at p value thresholds of 0.01 and 0.005 are significantly shifted toward lower p values in the GWAS for autism ($p < 5 \times 10–3$). We provide additional support for this using a GWAS of SCDC and demonstrate a lack of enrichment in a GWAS of Alzheimer's disease.

8.2.1 Conclusion

This study demonstrates a degree of methylomic overlap between autism and autistic traits, but we are limited in making further conclusions. Two factors—sample size and heterogeneity between the various samples—limit our understanding of methylation in autism. Future meta-analyses of both autism and autistic traits may help improve the statistical power of both the MWAS and aid in better understanding the shared etiology between the two phenotypes. We identified an enrichment for autism and autistic traits GWAS signals in the top CpG loci for autistic traits, but these must be replicated in in- dependent MWAS of autistic traits.

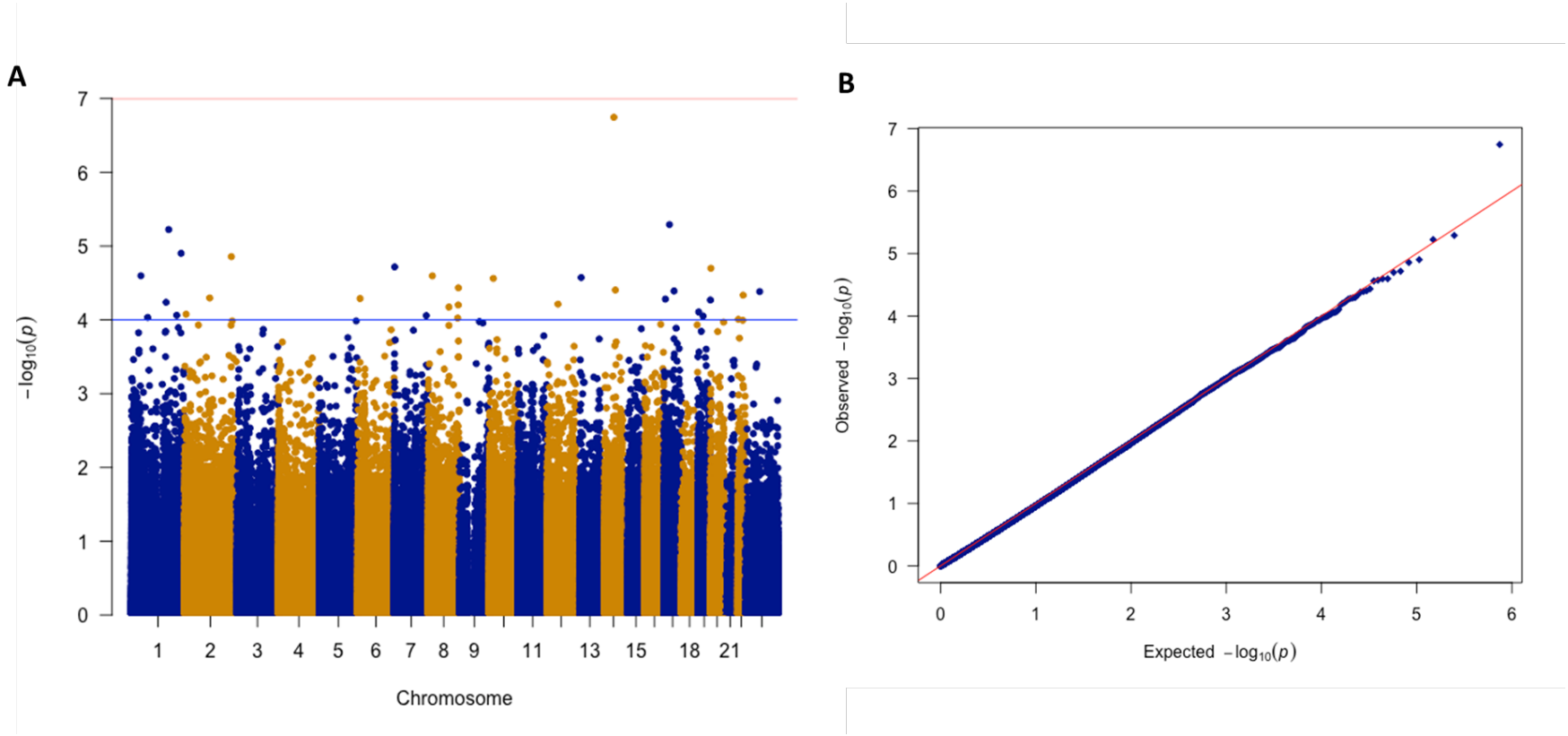
9 Supplementary information

Supplementary Figure 1: Histograms of SCDC and CCC scores in the samples used in the MWAS



Kernel density plots of SCDC scores (n = 701) and CCC scores (n = 666) in the sample used for the MWAS analyses.

Supplementary Figure 2: Manhattan plot and qq-plot of the MWAS for the CCC scores



A: Manhattan plot of the CCC MWAS. B: Quantile-Quantile plot of the CCC MWAS

Supplementary Table 1: Sample information and participant demographics

	SCDC	*CCC*
Mean	14.65 (3.44)	151.83 (6.77)
Sample size	701	666
Males	341	323
Whitecells	575	543
Bcell	0.16 (0.03)	0.16 (0.03)
CD3T	0.019 (0.05)	0.019 (0.05)
CD4T	0.16 (0.04)	0.16 (0.05)
CD8T	0.06 (0.03)	0.06 (0.04)
Gran	0.46 (0.08)	0.46 (0.09)
Mono	0.10 (0.03)	0.10 (0.03)
NK	0.14 (0.03)	0.14 (0.03)

This table provides sample information and participant demographics. We provide the mean values and , in paranthesis, standard deviation for the mean scores, and cell proportions for participants included in the SCDC and the CCC MWAS.

Supplementary Table 2: List of CpGs with P < 0.0001 in the SCDC MWAS

Name	Chr	Position	Gene symbol	Gene Region	BETA	SE	P
cg00156802	chrX	19358457	chrX:19361726-19362479		-0.81768	0.2024	5.35E-05
cg03098447	chr17	7210075	EIF5A;EIF5A;EIF5A	TSS1500;TSS1500;TSS1500	2.499287	0.641473	9.77E-05
cg03202738	chr4	1.56E+08	NPY2R	TSS200	3.25216	0.791181	3.95E-05
cg05125693	chr10	94051830	CPEB3;MARCH5	TSS1500;Body	2.865308	0.724334	7.63E-05
cg05877109	chr14	23775794	BCL2L2	TSS1500	2.574715	0.583753	1.03E-05
cg07640800	chr2	80529315	LRRTM1;CTNNA2;CTNNA2	3'UTR;Body;Body	1.673823	0.381845	1.17E-05
cg10894566	chr15	89905901	chr15:89904822-89906050		5.931252	1.379222	1.70E-05
cg11228785	chr5	1.79E+08	ADAMTS2;ADAMTS2	Body;Body	1.080653	0.275089	8.55E-05
cg11416605	chrX	63267928	chrX:63263904-63264129		1.233394	0.296407	3.17E-05

cg11490681	chr17	77460916	HRNBP3	5'UTR	0.591558	0.151996	9.94E-05
cg13448605	chr1	38100467	RSPO1;RSPO1	1stExon;5'UTR	11.17919	2.843069	8.42E-05
cg14379490	chr9	96221055	FAM120A	Body	-1.78917	0.35686	5.34E-07
cg15925695	chr11	1.12E+08	DIXDC1;DIXDC1	TSS200;Body	2.998837	0.745251	5.72E-05
cg17185953	chr1	53387522	ECHDC2	TSS200	3.780007	0.964301	8.86E-05
cg19478343	chr20	49620679	KCNG1	Body	-1.6772	0.402575	3.10E-05
cg19984781	chr6	30710898	FLOT1	TSS1500	0.932476	0.237908	8.87E-05
cg25165908	chr12	1.11E+08	CUX2	Body	-1.17914	0.299162	8.10E-05
cg25377985	chr4	1.48E+08	TTC29	TSS1500	1.876305	0.464107	5.28E-05
cg27314761	chr2	1.97E+08	SLC39A10;SLC39A10	5'UTR;5'UTR	-1.07153	0.248946	1.68E-05

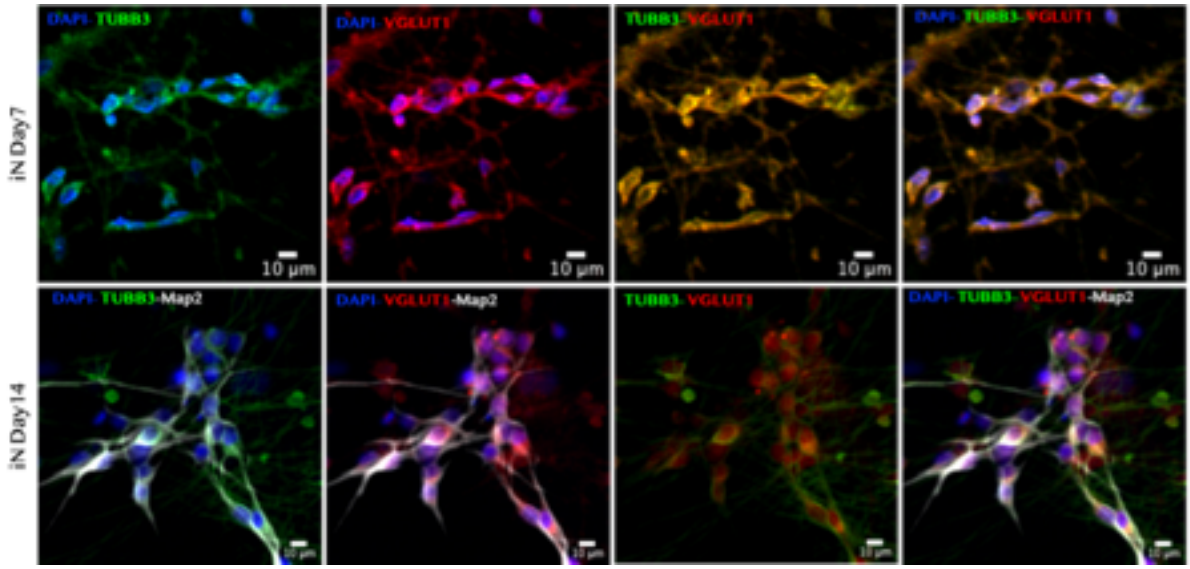
Supplementary Table 3: Results of the gene-set analyses for the SCDC MWAS (top 50 gene sets)

ID	Description	Size	p-value	padj
GO:0022626	cytosolic ribosome	261	9.12E-05	0.17
GO:0002181	cytoplasmic translation	119	0.000442	0.40
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	155	0.000639	0.40
GO:0071826	ribonucleoprotein complex subunit organization	308	0.001124	0.51
GO:0003735	structural constituent of ribosome	256	0.001618	0.51
GO:0022625	cytosolic large ribosomal subunit	140	0.002153	0.51
GO:0030183	B cell differentiation	128	0.002483	0.51
GO:0042100	B cell proliferation	105	0.002731	0.51
GO:0022618	ribonucleoprotein complex assembly	293	0.002934	0.51
GO:0010927	cellular component assembly involved in morphogenesis	125	0.003083	0.51
GO:0032279	asymmetric synapse	268	0.004005	0.51
GO:0006805	xenobiotic metabolic process	146	0.004308	0.51
GO:0098984	neuron to neuron synapse	271	0.004424	0.51
GO:0019210	kinase inhibitor activity	115	0.004464	0.51
GO:0099572	postsynaptic specialization	263	0.004754	0.51
GO:0004860	protein kinase inhibitor activity	110	0.004758	0.51
GO:0005840	ribosome	484	0.005194	0.51

GO:0034728	nucleosome organization	278	0.005454	0.51
GO:0042113	B cell activation	318	0.005769	0.51
GO:0006898	receptor-mediated endocytosis	381	0.006104	0.51
GO:0022627	cytosolic small ribosomal subunit	112	0.006365	0.51
GO:0050670	regulation of lymphocyte proliferation	238	0.006579	0.51
GO:0045047	protein targeting to ER	113	0.006649	0.51
GO:0072599	establishment of protein localization to endoplasmic reticulum	117	0.006713	0.51
GO:0014069	postsynaptic density	261	0.006767	0.51
GO:0048675	axon extension	128	0.008435	0.61
GO:0032944	regulation of mononuclear cell proliferation	241	0.009089	0.64
GO:0046651	lymphocyte proliferation	303	0.010073	0.64
GO:0031503	protein-containing complex localization	348	0.010266	0.64
GO:0070972	protein localization to endoplasmic reticulum	144	0.010555	0.64
GO:0070663	regulation of leukocyte proliferation	253	0.0109	0.64
GO:0031901	early endosome membrane	146	0.011318	0.64
GO:0050671	positive regulation of lymphocyte proliferation	153	0.011383	0.64
GO:0009897	external side of plasma membrane	303	0.012647	0.64
GO:0015935	small ribosomal subunit	170	0.012804	0.64

GO:0070665	positive regulation of leukocyte proliferation	163	0.013243	0.64
GO:1902936	phosphatidylinositol bisphosphate binding	111	0.01341	0.64
GO:0051291	protein heterooligomerization	124	0.01344	0.64
GO:0000786	nucleosome	154	0.013598	0.64
GO:0044445	cytosolic part	441	0.014362	0.64
GO:0032943	mononuclear cell proliferation	307	0.014697	0.64
GO:0044391	ribosomal subunit	393	0.014809	0.64
GO:1905897	regulation of response to endoplasmic reticulum stress	102	0.014985	0.64
GO:0044815	DNA packaging complex	166	0.015002	0.64
GO:2001237	negative regulation of extrinsic apoptotic signaling pathway	114	0.015159	0.64
GO:0031623	receptor internalization	108	0.016557	0.66
GO:0032946	positive regulation of mononuclear cell proliferation	156	0.016599	0.66
GO:0050864	regulation of B cell activation	176	0.01669	0.66
GO:0034614	cellular response to reactive oxygen species	177	0.01728	0.66
GO:0035690	cellular response to drug	363	0.017447	0.66

Supplementary Figure 3: Confocal images of BOB-NGN2 iN Day 7 and 14 stained with TUBB3 (green), vGLUT1 (red), MAP2 (grey) and DAPI (blue)



10 References

[1] L. Kanner, “Autistic Disturbances of affective contact,” 1943.

[2] J. Harris, “Leo Kanner and autism: a 75-year perspective.,” *Int. Rev. Psychiatry*, vol. 30, no. 1, pp. 3–17, Feb. 2018, doi: 10.1080/09540261.2018.1455646.

[3] J. C. Harris, “The origin and natural history of autism spectrum disorders.,” *Nat. Neurosci.*, vol. 19, no. 11, pp. 1390–1391, Oct. 2016, doi: 10.1038/nn.4427.

[4] C. Lord *et al.*, “Autism spectrum disorder.,” *Nat. Rev. Dis. Prim.*, vol. 6, no. 1, p. 5, Jan. 2020, doi: 10.1038/s41572-019-0138-4.

[5] S. M. Schaafsma and D. W. Pfaff, “Etiologies underlying sex differences in Autism Spectrum Disorders,” *Front. Neuroendocrinol.*, vol. 35, pp. 255–271, 2014.

[6] M.-C. Lai, M. V Lombardo, and S. Baron-Cohen, “Autism.,” *Lancet*, vol. 383, no. 9920, pp. 896–910, 2014, doi: 10.1016/S0140-6736(13)61539-1.

[7] G. Russell, L. R. Rodgers, O. C. Ukoumunne, and T. Ford, “Prevalence of parent-reported ASD and ADHD in the UK: Findings from the millennium cohort study,” *J. Autism Dev. Disord.*, vol. 44, no. 1, pp. 31–40, 2014, doi: 10.1007/s10803-013-1849-0.

[8] B. L. Leventhal *et al.*, “Prevalence of Autism Spectrum Disorders in a Total Population Sample,” *Am. J. Psychiatry*, vol. 168, no. 1, pp. 904–912, 2011, doi: 10.1176/appi.ajp.2011.10101532.

[9] E. Saemundsen, P. Magnússon, I. Georgsdóttir, E. Egilsson, and V. Rafnsson, “Prevalence of autism spectrum disorders in an Icelandic birth cohort.,” *BMJ Open*, vol. 3, no. 6, pp. 1–7, 2013, doi: 10.1136/bmjopen-2013-002748.

[10] Y. Wan *et al.*, “Prevalence of ASD among children in China,” *Shaghai*

Arch. Psychiatry, vol. 25, no. 2, pp. 70–80, 2013.

[11] L. R. Schaevitz and J. E. Berger-Sweeney, "Gene-environment interactions and epigenetic pathways in autism: the importance of one-carbon metabolism.," *ILAR J.*, vol. 53, no. 3–4, pp. 322–340, 2012, doi: 10.1093/ilar.53.3-4.322.

[12] A. Ronald and R. A. Hoekstra, "Autism spectrum disorders and autistic traits: A decade of new twin studies," *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, vol. 156, no. 3. pp. 255–274, 2011, doi: 10.1002/ajmg.b.31159.

[13] H. Taniai, T. Nishiyama, T. Miyachi, M. Imaeda, and S. Sumi, "Genetic influences on the broad spectrum of autism: Study of proband-ascertained twins," *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, vol. 147, no. 6, pp. 844–849, 2008, doi: 10.1002/ajmg.b.30740.

[14] S. S. Jeste and D. H. Geschwind, "Disentangling the heterogeneity of autism spectrum disorder through genetic findings.," *Nat. Rev. Neurol.*, vol. 10, no. 2, pp. 74–81, 2014, doi: 10.1038/nrneurol.2013.278.

[15] B. Wiśniowiecka-Kowalnik and B. A. Nowakowska, "Genetics and epigenetics of autism spectrum disorder—current evidence in the field," *J. Appl. Genet.*, vol. 60, no. 1, pp. 37–47, 2019, doi: 10.1007/s13353-018-00480-w.

[16] M. J. Ascano *et al.*, "FMRP targets distinct mRNA sequence elements to regulate protein expression.," *Nature*, vol. 492, no. 7429, pp. 382–386, Dec. 2012, doi: 10.1038/nature11737.

[17] X. Liu and T. Takumi, "Genomic and genetic aspects of autism spectrum disorder.," *Biochem. Biophys. Res. Commun.*, vol. 452, no. 2, pp. 244–253, Sep. 2014, doi: 10.1016/j.bbrc.2014.08.108.

[18] K. N. Leung *et al.*, "Neuronal chromatin dynamics of imprinting in development and disease," *J. Cell. Biochem.*, vol. 112, no. 2, pp. 365–373, 2011, doi: 10.1002/jcb.22958.Neuronal.

[19] J. M. LaSalle, L. T. Reiter, and S. J. Chamberlain, “Epigenetic Regulation of UBE3A and Roles in Human Neurodevelopmental Disorders,” *Epigenomics*, vol. 7, no. 7, pp. 1213–1228, 2015, doi: 10.2217/epi.15.70.Epigenetic.

[20] A. V. Ciernia and J. Lasalle, “The landscape of DNA methylation amid a perfect storm of autism aetiologies,” *Nat. Neurosci.*, vol. 41, 2016, doi: 10.1038/nrn.2016.41.

[21] S. J. Sanders *et al.*, “Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci Article Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci,” *Neuron*, vol. 87, no. 6, pp. 1215–1233, 2015, doi: 10.1016/j.neuron.2015.09.016.

[22] a Zhubi, Y. Chen, E. Dong, E. H. Cook, a Guidotti, and D. R. Grayson, “Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum.,” *Transl. Psychiatry*, vol. 4, no. 1, p. e349, 2014, doi: 10.1038/tp.2013.123.

[23] V. W. Hu, “The expanding genomic landscape of autism: discovering the ‘forest’ beyond the ‘trees,’” *Future Neurol.*, vol. 8, no. 1, pp. 29–42, 2013, doi: 10.1016/j.surg.2006.10.010.Use.

[24] M. Kchouk, J. F. Gibrat, and M. Elloumi, “Generations of Sequencing Technologies: From First to Next Generation,” *Biol. Med.*, vol. 09, no. 03, 2017, doi: 10.4172/0974-8369.1000395.

[25] K. Schwarze, J. Buchanan, J. C. Taylor, and S. Wordsworth, “Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature,” *Genet. Med.*, vol. 20, no. 10, pp. 1122–1130, 2018, doi: 10.1038/gim.2017.247.

[26] T. Long *et al.*, “Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites,” *Nat. Genet.*, vol. 49,

no. 4, pp. 568–578, 2017, doi: 10.1038/ng.3809.

[27] V. Tam, N. Patel, M. Turcotte, Y. Bossé, G. Paré, and D. Meyre, "Benefits and limitations of genome-wide association studies," *Nat. Rev. Genet.*, vol. 20, no. 8, pp. 467–484, 2019, doi: 10.1038/s41576-019-0127-1.

[28] J. C. Bis *et al.*, "Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation," *Mol. Psychiatry*, vol. 25, no. 8, pp. 1859–1875, 2020, doi: 10.1038/s41380-018-0112-7.

[29] H. Dan, X. Huang, Y. Xing, and Y. Shen, "Application of targeted panel sequencing and whole exome sequencing for 76 Chinese families with retinitis pigmentosa," *Mol. Genet. Genomic Med.*, vol. 8, no. 3, pp. 1–14, 2020, doi: 10.1002/mgg3.1131.

[30] J. Chahoud *et al.*, "Whole Exome Sequencing in Penile Squamous Cell Carcinoma Uncovers Novel Prognostic Categorization and Drug Targets Similar to Head and Neck Squamous Cell Carcinoma," *Clin. Cancer Res.*, p. clincanres.4004.2020, 2021, doi: 10.1158/1078-0432.ccr-20-4004.

[31] B. Devlin and S. W. Scherer, "Genetic architecture in autism spectrum disorder.," *Curr. Opin. Genet. Dev.*, vol. 22, no. 3, pp. 229–237, Jun. 2012, doi: 10.1016/j.gde.2012.03.002.

[32] A. Hogart, D. Wu, J. M. LaSalle, and N. C. Schanen, "The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13.," *Neurobiol. Dis.*, vol. 38, no. 2, pp. 181–191, May 2010, doi: 10.1016/j.nbd.2008.08.011.

[33] A. Huguet *et al.*, "A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression," *PLoS One*, vol. 11, no. 5, pp. 1–19, 2016, doi: 10.1371/journal.pone.0154248.

[34] S. Girirajan and E. E. Eichler, "Phenotypic variability and genetic susceptibility to genomic disorders.," *Hum. Mol. Genet.*, vol. 19, no. R2, pp. R176-87, Oct. 2010, doi: 10.1093/hmg/ddq366.

[35] J. T. Glessner *et al.*, "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.," *Nature*, vol. 459, no. 7246, pp. 569–573, May 2009, doi: 10.1038/nature07953.

[36] S. J. Sanders *et al.*, "Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism.," *Neuron*, vol. 70, no. 5, pp. 863–885, Jun. 2011, doi: 10.1016/j.neuron.2011.05.002.

[37] D. Pinto *et al.*, "Convergence of genes and cellular pathways dysregulated in autism spectrum disorders.," *Am. J. Hum. Genet.*, vol. 94, no. 5, pp. 677–694, May 2014, doi: 10.1016/j.ajhg.2014.03.018.

[38] E. B. B. Robinson *et al.*, "Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population," *Nat. Genet.*, vol. 48, no. 5, pp. 552–5, Mar. 2016, doi: 10.1101/027771.

[39] T. Gaugler *et al.*, "Most genetic risk for autism resides with common variation," *Nat. Genet.*, vol. 46, no. 8, pp. 881–885, 2014, doi: 10.1038/ng.3039.

[40] S. De Rubeis *et al.*, "Synaptic, transcriptional and chromatin genes disrupted in autism," *Nature*, vol. 515, no. 7526, pp. 209–15, Nov. 2014, doi: 10.1038/nature13772.

[41] J. Grove *et al.*, "Identification of common genetic risk variants for autism spectrum disorder," *Nat. Genet.*, vol. 51, no. 3, pp. 431–444, 2019, doi: 10.1038/s41588-019-0344-8.

[42] H. L. Johnson, "The promise of autism genetics: Still waiting," *J. Am. Assoc. Nurse Pract.*, vol. 31, no. 12, pp. 687–689, 2019, doi: 10.1097/JXX.0000000000000325.

[43] J. A. S. Vorstman, J. R. Parr, D. Moreno-De-Luca, R. J. L. Anney, J. I. J. Nurnberger, and J. F. Hallmayer, "Autism genetics: opportunities and challenges for clinical translation.," *Nat. Rev. Genet.*, vol. 18, no. 6, pp. 362–376, Jun. 2017, doi: 10.1038/nrg.2017.4.

[44] C. M. Dias and C. A. Walsh, “Recent Advances in Understanding the Genetic Architecture of Autism,” *Annu. Rev. Genomics Hum. Genet.*, vol. 21, pp. 289–304, 2020, doi: 10.1146/annurev-genom-121219-082309.

[45] D. Freed and J. Pevsner, “The Contribution of Mosaic Variants to Autism Spectrum Disorder.,” *PLoS Genet.*, vol. 12, no. 9, p. e1006245, Sep. 2016, doi: 10.1371/journal.pgen.1006245.

[46] E. T. Lim *et al.*, “Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder.,” *Nat. Neurosci.*, vol. 20, no. 9, pp. 1217–1224, Sep. 2017, doi: 10.1038/nn.4598.

[47] A. M. D’Gama *et al.*, “Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms,” *Neuron*, vol. 88, no. 5, pp. 910–917, 2015, doi: 10.1016/j.neuron.2015.11.009.

[48] R. Rodin *et al.*, “The Landscape of Mutational Mosaicism in Autistic and Normal Human Cerebral Cortex,” pp. 1–34, 2020, doi: 10.1101/2020.02.11.944413.

[49] T. Mbadiwe and R. M. Millis, “Epigenetics and Autism,” *Autism Res. Treat.*, vol. 2013, pp. 1–9, 2013, doi: 10.1155/2013/826156.

[50] J. D. Sweatt, M. J. Meaney, E. J. Nestler, and S. Akbarian, *Epigenetic Regulation in the Nervous System*. Elsevier, 2013.

[51] J. M. Lasalle, “Epigenomic strategies at the interface of genetic and environmental risk factors for autism,” *J. Hum. Genet.*, vol. 58, no. 7, pp. 396–401, 2013, doi: 10.1038/jhg.2013.49.Epigenomic.

[52] S. J. James *et al.*, “Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism 1 , 2,” *Am. J. Clin. Nutr.*, vol. 80, no. 1, pp. 1611–1617, 2004.

[53] J. A. Rusiecki, A. Baccarelli, V. Bollati, L. Tarantini, L. E. Moore, and E. C. Bonefeld-jorgensen, “Global DNA Hypomethylation Is Associated with High Serum-Persistent Organic Pollutants in Greenlandic Inuit,” *Environ.*

Health Perspect., vol. 116, no. 11, pp. 1547–1552, 2008, doi: 10.1289/ehp.11338.

[54] D. C. Dolinoy, J. R. Weidman, and R. L. Jirtle, "Epigenetic gene regulation : Linking early developmental environment to adult disease," *Reprod. Toxicol.*, vol. 23, no. 1, pp. 297–307, 2007, doi: 10.1016/j.reprotox.2006.08.012.

[55] R. L. Jirtle and M. K. Skinner, "Environmental epigenomics and disease susceptibility," *Nat. Genet.*, vol. 8, no. 1, pp. 253–262, 2007, doi: 10.1038/nrg2045.

[56] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, and S. Balasubramanian, "5-Hydroxymethylcytosine is a predominantly stable DNA modification.," *Nat. Chem.*, vol. 6, no. 12, pp. 1049–55, 2014, doi: 10.1038/nchem.2064.

[57] I. A. Qureshi and M. F. Mehler, "Epigenetic mechanisms governing the process of neurodegeneration," *Mol. Aspects Med.*, vol. 34, no. 4, pp. 875–882, 2013, doi: 10.1016/j.mam.2012.06.011.

[58] L. Zhu *et al.*, "Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders," *Hum. Mol. Genet.*, vol. 23, no. 6, pp. 1563–1578, 2014, doi: 10.1093/hmg/ddt547.

[59] S. G. Gregory *et al.*, "Genomic and epigenetic evidence for oxytocin receptor deficiency in autism.," *BMC Med.*, vol. 7, p. 62, 2009, doi: 10.1186/1741-7015-7-62.

[60] F. Behnia *et al.*, "Fetal DNA methylation of autism spectrum disorders candidate genes: association with spontaneous preterm birth," *Am. J. Obstet. Gynecol.*, vol. 212, no. 4, pp. 533.e1-533.e9, 2015, doi: 10.1016/j.ajog.2015.02.011.

[61] V. W. Hu *et al.*, "Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: Altered pathways in neuronal development and steroid biosynthesis," *PLoS One*, vol. 4, no. 6, 2009, doi:

10.1371/journal.pone.0005775.

[62] A. Crider, R. Thakkar, A. O. Ahmed, and A. Pillai, "Dysregulation of estrogen receptor beta (ERβ), aromatase (CYP19A1), and ER co-activators in the middle frontal gyrus of autism spectrum disorder subjects.," *Mol. Autism*, vol. 5, no. 1, p. 46, 2014, doi: 10.1186/2040-2392-5-46.

[63] E. Hannon *et al.*, "Elevated polygenic burden for autism is associated with differential DNA methylation at birth," *Genome Med.*, vol. 10, no. 1, p. 19, Dec. 2018, doi: 10.1186/s13073-018-0527-4.

[64] C. C. Y. Wong *et al.*, "Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits.," *Mol. Psychiatry*, vol. 19, no. 4, pp. 495–503, 2014, doi: 10.1038/mp.2013.41.

[65] S. V. Andrews *et al.*, "Case-control meta-analysis of blood DNA methylation and autism spectrum disorder," *Mol. Autism*, vol. 9, no. 1, p. 40, Dec. 2018, doi: 10.1186/s13229-018-0224-6.

[66] S. Nardone, D. S. Sams, A. Zito, E. Reuveni, and E. Elliott, "Dysregulation of Cortical Neuron DNA Methylation Profile in Autism Spectrum Disorder," *Cereb. Cortex*, vol. 27, no. 12, pp. 5739–5754, Dec. 2017, doi: 10.1093/cercor/bhx250.

[67] S. Nardone *et al.*, "DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways," *Transl. Psychiatry*, vol. 4, no. 9, pp. 1–9, 2014, doi: 10.1038/tp.2014.70.

[68] C. Ladd-Acosta *et al.*, "Common DNA methylation alterations in multiple brain regions in autism," *Mol. Psychiatry*, vol. 19, no. 8, pp. 862–871, 2014, doi: 10.1038/mp.2013.114.Common.

[69] C. Wong *et al.*, "Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic forms of autism in postmortem human brain tissue.," *bioRxiv*, p. 394387, Aug. 2018, doi: 10.1101/394387.

[70] D. H. Skuse, W. P. L. Mandy, and J. Scourfield, “Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist,” *Br. J. Psychiatry*, vol. 187, no. 6, pp. 568–572, Dec. 2005, doi: 10.1192/bjp.187.6.568.

[71] B. St Pourcain *et al.*, “ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties,” *Mol. Psychiatry*, Jan. 2017, doi: 10.1038/mp.2016.198.

[72] B. St Pourcain *et al.*, “Common variation contributes to the genetic architecture of social communication traits,” *Mol. Autism*, vol. 4, no. 1, p. 34, 2013, doi: 10.1186/2040-2392-4-34.

[73] E. Hannon *et al.*, “Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins,” *PLOS Genet.*, vol. 14, no. 8, p. e1007544, Aug. 2018, doi: 10.1371/journal.pgen.1007544.

[74] L. Gordon *et al.*, “Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence,” *Genome Res.*, vol. 22, no. 8, pp. 1395–1406, Aug. 2012, doi: 10.1101/gr.136598.111.

[75] J. van Dongen *et al.*, “Genetic and environmental influences interact with age and sex in shaping the human methylome,” *Nat. Commun.*, vol. 7, p. 11115, Apr. 2016, doi: 10.1038/ncomms11115.

[76] S. V. Andrews *et al.*, “Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder,” *Nat. Commun.*, vol. 8, no. 1, p. 1011, Dec. 2017, doi: 10.1038/s41467-017-00868-y.

[77] S. Budday, P. Steinmann, and E. Kuhl, “Physical biology of human brain development,” *Front. Cell. Neurosci.*, vol. 9, no. JULY, pp. 1–17, 2015, doi: 10.3389/fncel.2015.00257.

[78] R. O’Rahilly and F. Müller, “The meninges in human development.,” *J.*

Neuropathol. Exp. Neurol., vol. 45, no. 5, pp. 588–608, Sep. 1986.

[79] I. Kelava and M. A. Lancaster, "Stem Cell Models of Human Brain Development," *Cell Stem Cell*, vol. 18, no. 6, pp. 736–748, 2016, doi: 10.1016/j.stem.2016.05.022.

[80] T. Sun and R. F. Hevner, "Growth and folding of the mammalian cerebral cortex: from molecules to malformations.," *Nat. Rev. Neurosci.*, vol. 15, no. 4, pp. 217–232, Apr. 2014, doi: 10.1038/nrn3707.

[81] S. C. Noctor, V. Martínez-Cerdeño, L. Ivic, and A. R. Kriegstein, "Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases.," *Nat. Neurosci.*, vol. 7, no. 2, pp. 136–144, Feb. 2004, doi: 10.1038/nn1172.

[82] M. Missler *et al.*, "α-neurexins couple Ca 2+ channels to synaptic vesicle exocytosis," *Nature*, vol. 423, no. 6943, pp. 939–948, 2003, doi: 10.1038/nature01755.

[83] F. K. Satterstrom *et al.*, "Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism," *Cell*, vol. 180, no. 3, pp. 568-584.e23, 2020, doi: 10.1016/j.cell.2019.12.036.

[84] E. Courchesne, T. Pramparo, V. H. Gazestani, M. V Lombardo, K. Pierce, and N. E. Lewis, "The ASD Living Biology: from cell proliferation to clinical phenotype," *Mol. Psychiatry*, vol. 24, no. 1, pp. 88–107, 2019, doi: 10.1038/s41380-018-0056-y.

[85] E. Courchesne *et al.*, "Neuron Number and Size in Prefrontal Cortex of Children With Autism," *JAMA*, vol. 306, no. 18, pp. 2001–2010, Nov. 2011, doi: 10.1001/jama.2011.1638.

[86] R. Stoner *et al.*, "Patches of Disorganization in the Neocortex of Children with Autism," *N. Engl. J. Med.*, vol. 370, no. 13, pp. 1209–1219, 2014, doi: 10.1056/NEJMoa1307491.

[87] A. Packer, “Neocortical neurogenesis and the etiology of autism spectrum disorder,” *Neurosci. Biobehav. Rev.*, vol. 64, pp. 185–195, 2016, doi: https://doi.org/10.1016/j.neubiorev.2016.03.002.

[88] A. P. A. Donovan and M. A. Basson, “The neuroanatomy of autism – a developmental perspective,” *J. Anat.*, vol. 230, no. 1, pp. 4–15, 2017, doi: 10.1111/joa.12542.

[89] M. Santos *et al.*, “von Economo neurons in autism: A stereologic study of the frontoinsular cortex in children,” *Brain Res.*, vol. 1380, pp. 206–217, 2011, doi: https://doi.org/10.1016/j.brainres.2010.08.067.

[90] T. Rabinowicz, G. M. de Courten-Myers, J. M.-C. Petetot, G. Xi, and E. de los Reyes, “Human Cortex Development: Estimates of Neuronal Numbers Indicate Major Loss Late During Gestation,” *J. Neuropathol. Exp. Neurol.*, vol. 55, no. 3, pp. 320–328, Mar. 1996, doi: 10.1097/00005072-199603000-00007.

[91] J. M. Gohlke, W. C. Griffith, and E. M. Faustman, “Computational Models of Neocortical Neuronogenesis and Programmed Cell Death in the Developing Mouse, Monkey, and Human,” *Cereb. Cortex*, vol. 17, no. 10, pp. 2433–2442, Jan. 2007, doi: 10.1093/cercor/bhl151.

[92] F. Bonnet-Brilhault *et al.*, “Autism is a prenatal disorder: Evidence from late gestation brain overgrowth,” *Autism Res.*, vol. 11, no. 12, pp. 1635–1642, Dec. 2018, doi: 10.1002/aur.2036.

[93] A. F. Pardiñas *et al.*, “Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection,” *Nat. Genet.*, vol. 50, no. 3, pp. 381–389, 2018, doi: 10.1038/s41588-018-0059-2.

[94] M. L. Chow *et al.*, “Age-Dependent Brain Gene Expression and Copy Number Anomalies in Autism Suggest Distinct Pathological Processes at Young Versus Mature Ages,” *PLOS Genet.*, vol. 8, no. 3, p. e1002592, Mar. 2012, [Online]. Available: https://doi.org/10.1371/journal.pgen.1002592.

[95] J. Wegiel *et al.*, "The neuropathology of autism: defects of neurogenesis and neuronal migration, and dysplastic changes," *Acta Neuropathol.*, vol. 119, no. 6, pp. 755–770, 2010, doi: 10.1007/s00401-010-0655-4.

[96] M. Varghese *et al.*, "Autism spectrum disorder: neuropathology and animal models," *Acta Neuropathol.*, vol. 134, no. 4, pp. 537–566, 2017, doi: 10.1007/s00401-017-1736-4.

[97] I. A. J. van Kooten *et al.*, "Neurons in the fusiform gyrus are fewer and smaller in autism," *Brain*, vol. 131, no. 4, pp. 987–999, Mar. 2008, doi: 10.1093/brain/awn033.

[98] S. Jacot-Descombes *et al.*, "Decreased pyramidal neuron size in Brodmann areas 44 and 45 in patients with autism," *Acta Neuropathol.*, vol. 124, no. 1, pp. 67–79, 2012, doi: 10.1007/s00401-012-0976-6.

[99] J. Mariani *et al.*, "FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders," *Cell*, vol. 162, no. 2, pp. 375–390, 2015, doi: 10.1016/j.cell.2015.06.034.

[100] B. A. Derosa *et al.*, "Convergent Pathways in Idiopathic Autism Revealed by Time Course Transcriptomic Analysis of Patient-Derived Neurons," *Sci. Rep.*, vol. 8, no. 1, pp. 1–15, 2018, doi: 10.1038/s41598-018-26495-1.

[101] D. Adhya *et al.*, "Atypical neurogenesis in induced pluripotent stem cell (iPSC) from autistic individuals," *Biol. Psychiatry*, 2020, doi: 10.1016/j.biopsych.2020.06.014.

[102] J. Gilbert and H. Y. Man, "Fundamental elements in autism: From neurogenesis and neurite growth to synaptic plasticity," *Front. Cell. Neurosci.*, vol. 11, no. November, pp. 1–25, 2017, doi: 10.3389/fncel.2017.00359.

[103] N. N. Parikshak *et al.*, "Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism," *Cell*, vol. 155, no. 5, p. 1008, 2013, doi: 10.1016/j.cell.2013.10.031.

[104] S. A. Siegelbaum, E. R. Kandel, and R. Yuste, “Synaptic Transmission,” in *Principles of Neural Science*, Fifth edit., E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and H. A. J., Eds. McGraw-Hill - Medical, 2013, pp. 226–234.

[105] T. C. Südhof, “Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits,” *Cell*, vol. 171, no. 4, pp. 745–769, 2017, doi: 10.1016/j.cell.2017.10.024.

[106] T. C. Südhof, “Neuroligins and neurexins link synaptic function to cognitive disease,” *Nature*, vol. 455, no. 7215, pp. 903–911, 2008, doi: 10.1038/nature07456.

[107] J. Chen, S. Yu, Y. Fu, and X. Li, “Synaptic proteins and receptors defects in autism spectrum disorders,” *Front. Cell. Neurosci.*, vol. 8, no. September, pp. 1–13, 2014, doi: 10.3389/fncel.2014.00276.

[108] K. Tabuchi and T. C. Südhof, “Structure and Evolution of Neurexin Genes: Insight into the Mechanism of Alternative Splicing,” *Genomics*, vol. 79, no. 6, pp. 849–859, 2002, doi: https://doi.org/10.1006/geno.2002.6780.

[109] M. Missler and T. C. Südhof, “Neurexins: three genes and 1001 products.,” *Trends Genet.*, vol. 14, no. 1, pp. 20–26, Jan. 1998, doi: 10.1016/S0168-9525(97)01324-3.

[110] E. Peles, “Identification of a novel contactin-associated transmembrane receptor with multiple domains implicated in protein-protein interactions,” *EMBO journal.*, vol. 16, no. 5, pp. 978–988, 1997.

[111] A. A. Boucard, A. A. Chubykin, D. Comoletti, P. Taylor, and T. C. Südhof, “A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to α- and β-neurexins,” *Neuron*, vol. 48, no. 2, pp. 229–236, 2005, doi: 10.1016/j.neuron.2005.08.026.

[112] J. Ko, M. V Fuccillo, R. C. Malenka, and T. C. Südhof, “LRRTM2 Functions as a Neurexin Ligand in Promoting Excitatory Synapse Formation,”

Neuron, vol. 64, no. 6, pp. 791–798, 2009, doi: https://doi.org/10.1016/j.neuron.2009.12.012.

[113] M. Missler, R. E. Hammer, and T. C. Südhof, "Neurexophilin Binding to α-Neurexins," *J. Biol. Chem.*, vol. 273, no. 52, pp. 34716–34723, 1998, doi: 10.1074/jbc.273.52.34716.

[114] S. Sugita, F. Saito, J. Tang, J. Satz, K. Campbell, and T. C. Südhof, "A stoichiometric complex of neurexins and dystroglycan in brain ," *J. Cell Biol.*, vol. 154, no. 2, pp. 435–446, Jul. 2001, doi: 10.1083/jcb.200105003.

[115] C. Zhang *et al.*, "Neurexins physically and functionally interact with GABA(A) receptors.," *Neuron*, vol. 66, no. 3, pp. 403–416, May 2010, doi: 10.1016/j.neuron.2010.04.008.

[116] B. Treutlein, O. Gokce, S. R. Quake, and T. C. Südhof, "Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 13, pp. 1291–1299, 2014, doi: 10.1073/pnas.1403244111.

[117] T. Iijima *et al.*, "SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1," *Cell*, vol. 147, no. 7, pp. 1601–1614, 2011, doi: 10.1016/j.cell.2011.11.028.

[118] Y. Ding, L. Howard, L. Gallagher, and S. Shen, "Regulation and postsynaptic binding of neurexins – drug targets for neurodevelopmental and neuropsychiatric disorders," *Front. Biol. (Beijing).*, vol. 10, no. 3, pp. 239–251, 2015.

[119] Y. A. Ushkaryov, A. G. Petrenko, M. Geppert, and T. C. Südhof, "Neurexins: Synaptic cell surface proteins related to the α-latrotoxin receptor and laminin," *Science (80-.).*, vol. 257, no. 5066, pp. 50–56, 1992, doi: 10.1126/science.1621094.

[120] B. Ullrich, Y. A. Ushkaryov, and T. C. Südhof, "Cartography of neurexins: More than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons," *Neuron*, vol. 14, no. 3, pp. 497–507, 1995,

doi: 10.1016/0896-6273(95)90306-2.

[121] X. Ding *et al.*, "Activity-induced histone modifications govern Neurexin-1 mRNA splicing and memory preservation.," *Nat. Neurosci.*, vol. 20, no. 5, pp. 690–699, May 2017, doi: 10.1038/nn.4536.

[122] J. Dai, J. Aoto, and T. C. Südhof, "Alternative Splicing of Presynaptic Neurexins Differentially Controls Postsynaptic NMDA and AMPA Receptor Responses.," *Neuron*, vol. 102, no. 5, pp. 993-1008.e5, Jun. 2019, doi: 10.1016/j.neuron.2019.03.032.

[123] G. Rozic-Kotliroff and N. Zisapel, "Ca2+-dependent splicing of neurexin IIα," *Biochem. Biophys. Res. Commun.*, vol. 352, no. 1, pp. 226–230, 2007, doi: 10.1016/j.bbrc.2006.11.008.

[124] G. Rozic, Z. Lupowitz, and N. Zisapel, "Exonal elements and factors involved in the depolarization-induced alternative splicing of neurexin 2," *J. Mol. Neurosci.*, vol. 50, no. 1, pp. 221–233, 2013, doi: 10.1007/s12031-012-9919-x.

[125] T. Iijima, Y. Iijima, H. Witte, and P. Scheiffele, "Neuronal cell type-specific alternative splicing is regulated by the KH domain protein SLM1.," *J. Cell Biol.*, vol. 204, no. 3, pp. 331–342, Feb. 2014, doi: 10.1083/jcb.201310136.

[126] I. Ehrmann *et al.*, "The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain.," *PLoS Genet.*, vol. 9, no. 4, p. e1003474, Apr. 2013, doi: 10.1371/journal.pgen.1003474.

[127] W. Sun *et al.*, "Ultra-deep profiling of alternatively spliced Drosophila Dscam isoforms by circularization-assisted multi-segment sequencing.," *EMBO J.*, vol. 32, no. 14, pp. 2029–2038, Jul. 2013, doi: 10.1038/emboj.2013.144.

[128] L. Rowen *et al.*, "Analysis of the human neurexin genes: Alternative splicing and the generation of protein diversity," *Genomics*, vol. 79, no. 4,

pp. 587–597, 2002, doi: 10.1006/geno.2002.6734.

[129] E. Flaherty *et al.*, *Neuronal impact of patient-specific aberrant NRXN1α splicing*, vol. 51, no. 12. Springer US, 2019.

[130] C. R. Marshall *et al.*, "Structural Variation of Chromosomes in Autism Spectrum Disorder," *Am. J. Hum. Genet.*, vol. 82, no. 2, pp. 477–488, 2008, doi: 10.1016/j.ajhg.2007.12.009.

[131] M. Bucan *et al.*, "Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes," *PLoS Genet.*, vol. 5, no. 6, 2009, doi: 10.1371/journal.pgen.1000536.

[132] M. J. Gandal *et al.*, "Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder," *Science (80-.).*, vol. 362, no. 6420, 2018, doi: 10.1126/science.aat8127.

[133] J. Gauthier *et al.*, "Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia," *Hum. Genet.*, vol. 130, no. 4, pp. 563–573, 2011, doi: 10.1007/s00439-011-0975-z.

[134] A. K. Vaags *et al.*, "Rare deletions at the neurexin 3 locus in autism spectrum disorder.," *Am. J. Hum. Genet.*, vol. 90, no. 1, pp. 133–141, Jan. 2012, doi: 10.1016/j.ajhg.2011.11.025.

[135] B. Bakkaloglu *et al.*, "Molecular Cytogenetic Analysis and Resequencing of Contactin Associated Protein-Like 2 in Autism Spectrum Disorders," *Am. J. Hum. Genet.*, vol. 82, no. 1, pp. 165–173, 2008, doi: https://doi.org/10.1016/j.ajhg.2007.09.017.

[136] A. A. Chubykin *et al.*, "Activity-Dependent Validation of Excitatory versus Inhibitory Synapses by Neuroligin-1 versus Neuroligin-2," *Neuron*, vol. 54, no. 6, pp. 919–931, Jun. 2007, doi: 10.1016/j.neuron.2007.05.029.

[137] M. D. Gjørlund *et al.*, "Neuroligin-1 induces neurite outgrowth through interaction with neurexin-1β and activation of fibroblast growth factor receptor-1," *FASEB J.*, vol. 26, no. 10, pp. 4174–4186, 2012, doi: 10.1096/fj.11-

202242.

[138] E. Schnell, A. L. Bensen, E. K. Washburn, and G. L. Westbrook, "Neuroligin-1 overexpression in newborn granule cells in vivo.," *PLoS One*, vol. 7, no. 10, p. e48045, 2012, doi: 10.1371/journal.pone.0048045.

[139] X. Cao and K. Tabuchi, "Functions of synapse adhesion molecules neurexin/neuroligins and neurodevelopmental disorders," *Neurosci. Res.*, vol. 116, no. June, pp. 3–9, 2017, doi: 10.1016/j.neures.2016.09.005.

[140] K. Tabuchi *et al.*, "A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice.," *Science*, vol. 318, no. 5847, pp. 71–76, Oct. 2007, doi: 10.1126/science.1146221.

[141] B. Chih, S. K. Afridi, L. Clark, and P. Scheiffele, "Disorder-associated mutations lead to functional inactivation of neuroligins.," *Hum. Mol. Genet.*, vol. 13, no. 14, pp. 1471–1477, Jul. 2004, doi: 10.1093/hmg/ddh158.

[142] D. Comoletti *et al.*, "The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing.," *J. Neurosci.*, vol. 24, no. 20, pp. 4889–4893, May 2004, doi: 10.1523/JNEUROSCI.0468-04.2004.

[143] M. Etherton *et al.*, "Autism-linked neuroligin-3 R451C mutation differentially alters hippocampal and cortical synaptic function," *Proc. Natl. Acad. Sci.*, vol. 108, no. 33, pp. 13764–13769, 2011, doi: 10.1073/pnas.1111093108.

[144] T. A. Nguyen *et al.*, "A Cluster of Autism-Associated Variants on X-Linked NLGN4X Functionally Resemble NLGN4Y," *Neuron*, vol. 106, no. 5, pp. 759-768.e7, 2020, doi: https://doi.org/10.1016/j.neuron.2020.03.008.

[145] M.-C. Lai, M. V Lombardo, B. Auyeung, B. Chakrabarti, and S. Baron-Cohen, "Sex/Gender Differences and Autism: Setting the Scene for Future Research," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 54, no. 1, pp. 11–24, 2015, doi: 10.1016/j.jaac.2014.10.003.

[146] J. Koehnke *et al.*, "Crystal structures of beta-neurexin 1 and beta-neurexin

2 ectodomains and dynamics of splice insertion sequence 4.," *Structure*, vol. 16, no. 3, pp. 410–421, Mar. 2008, doi: 10.1016/j.str.2007.12.024.

[147] C. S. Leblond *et al.*, "Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders," *PLoS Genet.*, vol. 8, no. 2, 2012, doi: 10.1371/journal.pgen.1002521.

[148] M.-I. Tejada, X. Elcoroaristizabal, N. Ibarluzea, M.-P. Botella, A.-B. de la Hoz, and I. Ocio, "A novel nonsense homozygous variant in the NLGN1 gene found in a pair of monozygotic twin brothers with intellectual disability and autism.," *Clinical genetics*, vol. 95, no. 2. Denmark, pp. 339–340, Feb. 2019, doi: 10.1111/cge.13466.

[149] M. Nakanishi *et al.*, "Functional significance of rare neuroligin 1 variants found in autism.," *PLoS Genet.*, vol. 13, no. 8, p. e1006940, Aug. 2017, doi: 10.1371/journal.pgen.1006940.

[150] C. Sun *et al.*, "Identification and functional characterization of rare mutations of the neuroligin-2 gene (NLGN2) associated with schizophrenia.," *Hum. Mol. Genet.*, vol. 20, no. 15, pp. 3042–3051, Aug. 2011, doi: 10.1093/hmg/ddr208.

[151] N. Krumm *et al.*, "Excess of rare, inherited truncating mutations in autism.," *Nat. Genet.*, vol. 47, no. 6, pp. 582–588, Jun. 2015, doi: 10.1038/ng.3303.

[152] D. J. Parente *et al.*, "Neuroligin 2 nonsense variant associated with anxiety, autism, intellectual disability, hyperphagia, and obesity.," *Am. J. Med. Genet. A*, vol. 173, no. 1, pp. 213–216, Jan. 2017, doi: 10.1002/ajmg.a.37977.

[153] F. Blasi *et al.*, "Absence of coding mutations in the X-linked genes neuroligin 3 and neuroligin 4 in individuals with autism from the IMGSAC collection," *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, vol. 141B, no. 3, pp. 220–221, 2006, doi: 10.1002/ajmg.b.30287.

[154] A. Lawson-Yuen, J. S. Saldivar, S. Sommer, and J. Picker, "Familial

deletion within NLGN4 associated with autism and Tourette syndrome," *Eur. J. Hum. Genet.*, vol. 16, no. 5, pp. 614–618, 2008, doi: 10.1038/sj.ejhg.5202006.

[155] T. Ylisaukko-oja *et al.*, "Analysis of four neuroligin genes as candidates for autism," *Eur. J. Hum. Genet.*, vol. 13, no. 12, pp. 1285–1292, 2005, doi: 10.1038/sj.ejhg.5201474.

[156] S. Jamain *et al.*, "Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism.," *Nat. Genet.*, vol. 34, no. 1, pp. 27–29, May 2003, doi: 10.1038/ng1136.

[157] F. Laumonnier *et al.*, "X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family.," *Am. J. Hum. Genet.*, vol. 74, no. 3, pp. 552–557, Mar. 2004, doi: 10.1086/382137.

[158] Z. Talebizadeh, D. C. Bittel, O. J. Veatch, M. G. Butler, T. N. Takahashi, and J. H. Miles, "Do known mutations in neuroligin genes (NLGN3 and NLGN4) cause autism?," *Journal of autism and developmental disorders*, vol. 34, no. 6. pp. 735–736, Dec. 2004, doi: 10.1007/s10803-004-5295-x.

[159] J. B. Vincent, D. Kolozsvari, W. S. Roberts, P. F. Bolton, H. M. D. Gurling, and S. W. Scherer, "Mutation screening of X-chromosomal neuroligin genes: no mutations in 196 autism probands.," *Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.*, vol. 129B, no. 1, pp. 82–84, Aug. 2004, doi: 10.1002/ajmg.b.30069.

[160] J. Gauthier *et al.*, "NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population," *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, vol. 132B, no. 1, pp. 74–75, 2005, doi: 10.1002/ajmg.b.30066.

[161] A.-K. Wermter, I. Kamp-Becker, K. Strauch, G. Schulte-Körne, and H. Remschmidt, "No evidence for involvement of genetic variants in the X-linked neuroligin genes NLGN3 and NLGN4X in probands with autism

spectrum disorder on high functioning level.," *Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.*, vol. 147B, no. 4, pp. 535–537, Jun. 2008, doi: 10.1002/ajmg.b.30618.

[162] Y. Liu *et al.*, "Lack of association between NLGN3, NLGN4, SHANK2 and SHANK3 gene variants and autism spectrum disorder in a Chinese population.," *PLoS One*, vol. 8, no. 2, p. e56639, 2013, doi: 10.1371/journal.pone.0056639.

[163] Z. Hu, X. Xiao, Z. Zhang, and M. Li, "Genetic insights and neurobiological implications from NRXN1 in neuropsychiatric disorders," *Mol. Psychiatry*, vol. 24, no. 10, pp. 1400–1414, 2019, doi: 10.1038/s41380-019-0438-9.

[164] K. Mukherjee *et al.*, "CASK Functions as a Mg2+-independent neurexin kinase Konark," *Cell*, vol. 133, no. 2, pp. 328–339, 2008, doi: 10.1016/j.cell.2008.02.036.

[165] S. K. Singh *et al.*, "Astrocytes Assemble Thalamocortical Synapses by Bridging NRX1α and NL1 via Hevin.," *Cell*, vol. 164, no. 1–2, pp. 183–196, Jan. 2016, doi: 10.1016/j.cell.2015.11.034.

[166] A. K. Jenkins, C. Paterson, Y. Wang, T. M. Hyde, J. E. Kleinman, and A. J. and Law, "Neurexin 1 (NRXN1) Splice Isoform Expression During Human Neocortical Development and Aging," *Mol. Psychiatry*, vol. 21, no. 5, pp. 701–706, 2016, doi: 10.1001/jamainternmed.2014.5466.Association.

[167] S. L. Andersen, "Trajectories of brain development: point of vulnerability or window of opportunity?file:///Users/aicha/Downloads/19794405.nbib," *Neurosci. Biobehav. Rev.*, vol. 27, no. 1–2, pp. 3–18, 2003, doi: 10.1016/s0149-7634(03)00005-8.

[168] G. Z. Tau and B. S. Peterson, "Normal development of brain circuits.," *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.*, vol. 35, no. 1, pp. 147–168, Jan. 2010, doi: 10.1038/npp.2009.115.

[169] P. R. Huttenlocher, "Synaptic density in human frontal cortex -

developmental changes and effects of aging.," *Brain Res.*, vol. 163, no. 2, pp. 195–205, Mar. 1979, doi: 10.1016/0006-8993(79)90349-4.

[170] E. R. Graf, X. Zhang, S.-X. Jin, M. W. Linhoff, and A. M. Craig, "Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins," *Cell*, vol. 119, no. 7, pp. 1013–1026, 2004, doi: 10.1016/j.cell.2004.11.035.

[171] Y. Kang, X. Zhang, F. Dobie, H. Wu, and A. M. Craig, "Induction of GABAergic postsynaptic differentiation by alpha-neurexins.," *J. Biol. Chem.*, vol. 283, no. 4, pp. 2323–2334, Jan. 2008, doi: 10.1038/jid.2014.371.

[172] G. Giannone *et al.*, "Neurexin-1β binding to neuroligin-1 triggers the preferential recruitment of PSD-95 versus gephyrin through tyrosine phosphorylation of neuroligin-1.," *Cell Rep.*, vol. 3, no. 6, pp. 1996–2007, Jun. 2013, doi: 10.1016/j.celrep.2013.05.013.

[173] J. M. Friedman *et al.*, "Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation.," *Am. J. Hum. Genet.*, vol. 79, no. 3, pp. 500–513, Sep. 2006, doi: 10.1086/507471.

[174] Sa. Geno *et al.*, "Mapping autism risk loci using genetic linkage and chromosomal rearrangements.," *Nat. Genet.*, vol. 39, no. 3, pp. 319–328, Mar. 2007, doi: 10.1038/ng1985.

[175] M. S. L. Ching *et al.*, "Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders," *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, vol. 153, no. 4, pp. 937–947, 2010, doi: 10.1002/ajmg.b.31063.

[176] C. Zweier *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.*, vol. 85, no. 5, pp. 655–666, Nov. 2009, doi: 10.1016/j.ajhg.2009.10.004.

[177] C. P. Schaaf *et al.*, "Phenotypic spectrum and genotype-phenotype correlations of NRXN1 exon deletions.," *Eur. J. Hum. Genet.*, vol. 20, no.

12, pp. 1240–1247, Dec. 2012, doi: 10.1038/ejhg.2012.95.

[178] M. P. Dabell *et al.*, "Investigation of NRXN1 deletions: clinical and molecular characterization.," *Am. J. Med. Genet. A*, vol. 161A, no. 4, pp. 717–731, Apr. 2013, doi: 10.1002/ajmg.a.35780.

[179] R. J. Camacho-Garcia *et al.*, "Mutations affecting synaptic levels of neurexin-1β in autism and mental retardation.," *Neurobiol. Dis.*, vol. 47, no. 1, pp. 135–143, Jul. 2012, doi: 10.1016/j.nbd.2012.03.031.

[180] M. Lam *et al.*, "Single cell analysis of autism patient with bi-allelic NRXN1-alpha deletion reveals skewed fate choice in neural progenitors and impaired neuronal functionality," *Exp. Cell Res.*, vol. 383, no. 1, p. 111469, 2019, doi: 10.1016/j.yexcr.2019.06.014.

[181] S. Avazzadeh *et al.*, "Increased Ca2+ signaling in NRXN1α +/- neurons derived from ASD induced pluripotent stem cells," *Mol. Autism*, vol. 10, no. 1, pp. 1–16, 2019, doi: 10.1186/s13229-019-0303-3.

[182] C. Lowther *et al.*, "Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression.," *Genet. Med.*, vol. 19, no. 1, pp. 53–61, Jan. 2017, doi: 10.1038/gim.2016.54.

[183] F. Béna *et al.*, "Molecular and clinical characterization of 25 individuals with exonic deletions of NRXN1 and comprehensive review of the literature.," *Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.*, vol. 162B, no. 4, pp. 388–403, Jun. 2013, doi: 10.1002/ajmg.b.32148.

[184] A. Guilmatre *et al.*, "Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation.," *Arch. Gen. Psychiatry*, vol. 66, no. 9, pp. 947–956, Sep. 2009, doi: 10.1001/archgenpsychiatry.2009.80.

[185] L. Duong *et al.*, "Mutations in NRXN1 in a family multiply affected with

brain disorders: NRXN1 mutations and brain disorders.," *Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.*, vol. 159B, no. 3, pp. 354–358, Apr. 2012, doi: 10.1002/ajmg.b.32036.

[186] P. Szatmari *et al.*, "Mapping autism risk loci using genetic linkage and chromosomal rearrangements.," *Nat. Genet.*, vol. 39, no. 3, pp. 319–328, Mar. 2007, doi: 10.1038/ng1985.

[187] A. Tromp, B. Mowry, and J. Giacomotto, "Neurexins in autism and schizophrenia—a review of patient mutations, mouse models and potential future directions," *Mol. Psychiatry*, 2020, doi: 10.1038/s41380-020-00944-8.

[188] A. Wang, Y. Xiang, B. B. Yang, and W. Lu, "Neurexin-1α regulates neurite growth of rat hippocampal neurons," vol. 11, no. 4, pp. 115–125, 2019.

[189] W. D. Constance *et al.*, "Neurexin and neuroligin-based adhesion complexes drive axonal arborisation growth independent of synaptic activity," *Elife*, vol. 7, pp. 1–33, 2018, doi: 10.7554/eLife.31659.

[190] S. X. Chen, P. K. Tari, K. She, and K. Haas, "Neurexin-neuroligin cell adhesion complexes contribute to synaptotropic dendritogenesis via growth stabilization mechanisms in vivo," *Neuron*, vol. 67, no. 6, pp. 967–983, 2010, doi: 10.1016/j.neuron.2010.08.016.

[191] D. A. Murchison, "a Novel Function for Neurexin and Neuroligin Family," McGill University, 2009.

[192] T. Patriarchi *et al.*, "Imbalance of excitatory/inhibitory synaptic protein expression in iPSC-derived neurons from FOXG1(+/-) patients and in foxg1(+/-) mice," *Eur. J. Hum. Genet.*, vol. 24, no. 6, pp. 871–880, 2016, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26443267.

[193] M. R. Etherton, C. A. Blaiss, C. M. Powell, and T. C. Südhof, "Mouse neurexin-1α deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 42, pp. 17998–18003, 2009, doi:

10.1073/pnas.0910297106.

[194] G. R. Anderson *et al.*, "β-Neurexins Control Neural Circuits by Regulating Synaptic Endocannabinoid Signaling," *Cell*, vol. 162, no. 3, pp. 593–606, 2015, doi: 10.1016/j.cell.2015.06.056.

[195] C. H. Pak *et al.*, "Human Neuropsychiatric Disease Modeling using Conditional Deletion Reveals Synaptic Transmission Defects Caused by Heterozygous Mutations in NRXN1," *Cell Stem Cell*, vol. 17, no. 3, pp. 316–328, 2015, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26279266.

[196] C. Nicolini and M. Fahnestock, "The valproic acid-induced rodent model of autism.," *Exp. Neurol.*, vol. 299, no. Pt A, pp. 217–227, Jan. 2018, doi: 10.1016/j.expneurol.2017.04.017.

[197] T. M. Kazdoba, P. T. Leach, M. Yang, J. L. Silverman, M. Solomon, and J. N. Crawley, "Translational Mouse Models of Autism: Advancing Toward Pharmacological Therapeutics.," *Curr. Top. Behav. Neurosci.*, vol. 28, pp. 1–52, 2016, doi: 10.1007/7854_2015_5003.

[198] R. A. Martinez *et al.*, "Genome engineering of isogenic human ES cells to model autism disorders," *Nucleic Acids Res*, vol. 43, no. 10, p. e65, 2015, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25765640.

[199] V. Martínez-Cerdeño, "Dendrite and spine modifications in autism and related neurodevelopmental disorders in patients and animal models," *Dev. Neurobiol.*, vol. 77, no. 4, pp. 393–404, 2017, doi: 10.1002/dneu.22417.

[200] J. Kim, B. K. Koo, and J. A. Knoblich, "Human organoids: model systems for human biology and medicine," *Nat. Rev. Mol. Cell Biol.*, vol. 21, no. 10, pp. 571–584, 2020, doi: 10.1038/s41580-020-0259-3.

[201] M. Huch and B.-K. Koo, "Modeling mouse and human development using organoid cultures.," *Development*, vol. 142, no. 18, pp. 3113–3125, Sep. 2015, doi: 10.1242/dev.118570.

[202] R. Baumeister and L. Ge, “The worm in us - Caenorhabditis elegans as a model of human disease.,” *Trends in biotechnology*, vol. 20, no. 4. England, pp. 147–148, Apr. 2002, doi: 10.1016/s0167-7799(01)01925-4.

[203] J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson, “The embryonic cell lineage of the nematode Caenorhabditis elegans.,” *Dev. Biol.*, vol. 100, no. 1, pp. 64–119, Nov. 1983, doi: 10.1016/0012-1606(83)90201-4.

[204] S. S.S., S. R.J., and B. E.V., “Model systems for studying cellular mechanisms of SCN1A-related epilepsy,” *Journal of Neurophysiology*, vol. 115, no. 4. American Physiological Society (E-mail: subscrip@the-aps.org), D.K. O’Dowd, Univ. of California, 4221 McGaugh Hall, Irvine, CA 92697, United States. E-mail: dkodowd@uci.edu, pp. 1755–1766, 2016, [Online]. Available: http://jn.physiology.org/content/115/4/1767.full.pdf.

[205] C. Nüsslein-Volhard, “The zebrafish issue of Development.,” *Development*, vol. 139, no. 22, pp. 4099–4103, Nov. 2012, doi: 10.1242/dev.085217.

[206] P. Haffter and C. Nüsslein-Volhard, “Large scale genetics in a small vertebrate, the zebrafish.,” *Int. J. Dev. Biol.*, vol. 40, no. 1, pp. 221–227, Feb. 1996.

[207] M. C. Mullins, M. Hammerschmidt, P. Haffter, and C. Nüsslein-Volhard, “Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate.,” *Curr. Biol.*, vol. 4, no. 3, pp. 189–202, Mar. 1994, doi: 10.1016/s0960-9822(00)00048-8.

[208] J.-L. Bossu and S. Roux, “[The valproate model of autism].,” *Med. Sci. (Paris).*, vol. 35, no. 3, pp. 236–243, Mar. 2019, doi: 10.1051/medsci/2019036.

[209] C. Leung, Z. Jia, and L. C., “Mouse Genetic Models of Human Brain Disorders,” *Front. Genet.*, vol. 7, no. MAR, p. no pagination, 2016, [Online]. Available:

http://journal.frontiersin.org/article/10.3389/fgene.2016.00040/full.

[210] M. J.M., J. M. McCammon, and H. Sive, “Addressing the Genetics of Human Mental Health Disorders in Model Organisms,” *Annu. Rev. Genomics Hum. Genet.*, vol. 16, pp. 173–197, 2015, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26002061.

[211] L. Chen, M. Jiang, B. Zhang, O. and Gokce, and T. C. Südhof, “Conditional Deletion of All Neurexins Defines Diversity of Essential Synaptic Organizer Functions for Neurexins,” *Neuron*, vol. 94, no. 3, pp. 611-625.e4, 2017, doi: 10.1016/j.neuron.2017.04.011.

[212] M. W. Nestor *et al.*, “Human Inducible Pluripotent Stem Cells and Autism Spectrum Disorder: Emerging Technologies,” *Autism Res.*, vol. 9, no. 5, pp. 513–535, 2016, doi: 10.1002/aur.1570.

[213] J. K. Ichida and E. Kiskinis, “Probing disorders of the nervous system using reprogramming approaches.,” *EMBO J.*, vol. 34, no. 11, pp. 1456–1477, 2015, doi: 10.15252/embj.201591267.

[214] J. P. Thomson *et al.*, “Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques,” *Nucleic Acids Res.*, vol. 41, no. 22, pp. 1–15, 2013, doi: 10.1093/nar/gkt1080.

[215] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, “Genome engineering using the CRISPR-Cas9 system,” *Nat. Protoc.*, vol. 8, no. 11, pp. 2281–2308, 2013, doi: 10.1038/nprot.2013.143.

[216] M. Adli, “The CRISPR tool kit for genome editing and beyond,” *Nat. Commun.*, vol. 9, no. 1, 2018, doi: 10.1038/s41467-018-04252-2.

[217] R. R. Shah *et al.*, “Efficient and versatile CRISPR engineering of human neurons in culture to model neurological disorders,” *Wellcome Open Res*, vol. 1, p. 13, 2016, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/27976757.

[218] C. J. Giuliano, A. Lin, V. Girish, and J. M. Sheltzer, “Generating Single

Cell–Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9," *Curr. Protoc. Mol. Biol.*, vol. 128, no. 1, pp. 1–25, 2019, doi: 10.1002/cpmb.100.

[219] M. Mueller, T. Seufferlein, A. Illing, and J. Homann, "Modelling human channelopathies using induced pluripotent stem cells: a comprehensive review," *Stem Cells Int.*, 2013, doi: 10.1155/2013/496501.

[220] S. M. Chambers, C. A. Fasano, E. P. Papapetrou, M. Tomishima, M. Sadelain, and L. Studer, "Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling," *Nat. Biotechnol.*, vol. 27, no. 3, pp. 275–280, 2009, doi: 10.1038/nbt.1529.

[221] T. G. Fernandes *et al.*, "Neural commitment of human pluripotent stem cells under defined conditions recapitulates neural development and generates patient-specific neural cells," *Biotechnol. J.*, vol. 10, no. 10, pp. 1578–1588, 2015, [Online]. Available: http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1860-7314.

[222] Y. Zhang *et al.*, "Rapid single-step induction of functional neurons from human pluripotent stem cells," *Neuron*, vol. 78, no. 5, pp. 785–798, 2013, doi: 10.1016/j.neuron.2013.05.029.

[223] B. Aydin *et al.*, "Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes," *Nat Neurosci*, vol. 22, no. 6, pp. 897–908, 2019, doi: 10.1038/s41593-019-0399-y.Proneural.

[224] J. Drouin-Ouellet *et al.*, "REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways.," *EMBO Mol. Med.*, vol. 9, no. 8, pp. 1117–1131, Aug. 2017, doi: 10.15252/emmm.201607471.

[225] M. K. Abdul Karim, "Integrative genomic analysis of Neurogenin2 reprogramming of human iPSCs," University of Cambridge, 2020.

[226] S.-K. Lee, B. Lee, E. C. Ruiz, and S. L. Pfaff, "Olig2 and Ngn2 function in

opposition to modulate gene expression in motor neuron progenitor cells.," *Genes Dev.*, vol. 19, no. 2, pp. 282–294, Jan. 2005, doi: 10.1101/gad.1257105.

[227] M.-L. Liu *et al.*, "Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons," *Nat. Commun.*, vol. 4, no. 1, p. 2183, 2013, doi: 10.1038/ncomms3183.

[228] D. K. Smith, J. Yang, M. Liu, and C. Zhang, "Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming," *Stem Cell Reports*, vol. 7, no. 5, pp. 955–969, 2016, doi: 10.1016/j.stemcr.2016.09.013.

[229] M. Pawlowski *et al.*, "Inducible and Deterministic Forward Programming of Human Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes," *Stem Cell Reports*, vol. 8, no. 4, pp. 803–812, 2017, doi: 10.1016/j.stemcr.2017.02.016.

[230] A. Bertero *et al.*, "Optimized inducible shRNA and CRISPR/Cas9 platforms for in vitro studies of human development using hPSCs," *Dev.*, vol. 143, no. 23, pp. 4405–4418, 2016, doi: 10.1242/dev.138081.

[231] F. Rosa *et al.*, "In Vitro Differentiated Human Stem Cell-Derived Neurons Reproduce Synaptic Synchronicity Arising during Neurodevelopment.," *Stem cell reports*, vol. 15, 2020, doi: 10.1016/j.stemcr.2020.05.015.

[232] R. Nehme *et al.*, "Combining NGN2 Programming with Developmental Patterning Generates Human Excitatory Neurons with NMDAR-Mediated Synaptic Transmission," *Cell Rep.*, vol. 23, no. 8, pp. 2509–2523, 2018, doi: 10.1097/CCM.0b013e31823da96d.Hydrogen.

[233] B. Y. Hu *et al.*, "Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 9, pp. 4335–4340, 2010, doi: 10.1073/pnas.0910012107.

[234] D. X., X. Du, J. M. Parent, and D. X., "Using Patient-Derived Induced

Pluripotent Stem Cells to Model and Treat Epilepsies," *Curr. Neurol. Neurosci. Rep.*, vol. 15, no. 10, p. no pagination, 2015, [Online]. Available: http://www.springer.com.

[235] A. Gordon *et al.*, "Long-term maturation of human cortical organoids matches key early postnatal transitions," *Nat. Neurosci.*, 2021, doi: 10.1038/s41593-021-00802-y.

[236] K. I., I. Kelava, and M. A. Lancaster, "Dishing out mini-brains: Current progress and future prospects in brain organoid research," *Dev. Biol.*, vol. 420, no. 2, pp. 199–209, 2016, doi: 10.1016/j.ydbio.2016.06.037.

[237] S. Curran, J. W. Ahn, H. Grayton, D. A. Collier, and C. M. Ogilvie, "NRXN1 deletions identified by array comparative genome hybridisation in a clinical case series – further understanding of the relevance of NRXN1 to neurodevelopmental disorders," *J. Mol. Psychiatry*, vol. 1, no. 1, p. 4, 2013, doi: 10.1186/2049-9256-1-4.

[238] J. W. Ahn *et al.*, "BBGRE: Brain and body genetic resource exchange," *Database*, vol. 2013, pp. 1–4, 2013, doi: 10.1093/database/bat067.

[239] K. Yusa *et al.*, "Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells," *Nature*, vol. 478, no. 7369, pp. 391–394, 2011, doi: 10.1038/nature10424.

[240] J. Ladewig *et al.*, "Small molecules enable highly efficient neuronal conversion of human fibroblasts," *Nat. Methods*, vol. 9, no. 6, pp. 575–578, 2012, doi: 10.1038/nmeth.1972.

[241] K. D. Mccarthy and J. De Vellis, "Preparation of separate astroglial and oligodendroglial cell cultures from rat cereberal tissue," *J Cell Biol.*, vol. 85, no. June, pp. 890–902, 1980.

[242] Z. A. Syed, T. Härd, A. Uv, and I. F. van Dijk-Härd, "A potential role for Drosophila mucins in development and physiology," *PLoS One*, vol. 3, no. 8, 2008, doi: 10.1371/journal.pone.0003041.

[243] Y. Qiu, M. RD, W. IM, S. Seethapathy, B. RC, and K. IJ, “Enhanced Isotopic Ratio Outlier Analysis (IROA) Peak Detection and Identification with Ultra-High Resolution GC-Orbitrap/MS: Potential Application for Investigation of Model Organism Metabolomes.,” *Rev. Derecho Genoma Hum.*, vol. 8, no. 1, Jan. 2018.

[244] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, “Near-optimal probabilistic RNA-seq quantification,” *Nat. Biotechnol.*, vol. 34, no. 5, pp. 4–8, 2016, doi: 10.1038/nbt.3519.

[245] K. Blighe and M. Lewis, “Package ‘ PCAtools ,’” 2019.

[246] A. G. R. Warnes *et al.*, “Package ‘ gplots ,’” no. 1, 2016.

[247] R Core Team, “R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.,” 2018.

[248] G. Pau, F. Fuchs, O. Sklyar, M. Boutros, and W. Huber, “EBImage-an R package for image processing with applications to cellular phenotypes,” *Bioinformatics*, vol. 26, no. 7, pp. 979–981, 2010, doi: 10.1093/bioinformatics/btq046.

[249] C. L. Relton *et al.*, “Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES),” *Int. J. Epidemiol.*, vol. 44, no. 4, pp. 1181–1190, Aug. 2015, doi: 10.1093/ije/dyv072.

[250] A. Boyd *et al.*, “Cohort Profile: The ‘Children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children,” *Int. J. Epidemiol.*, vol. 42, no. 1, pp. 111–127, Feb. 2013, doi: 10.1093/ije/dys064.

[251] D. Bishop, “Development of the Children’s Communication Checklist (CCC): a method for assessing qualitative aspects of communicative impairment in children.,” *J. Child Psychol. Psychiatry.*, vol. 39, no. 6, pp. 879–91, Sep. 1998.

[252] B. St Pourcain *et al.*, “Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and

adolescence.," *Mol. Autism*, vol. 5, no. 1, p. 18, Feb. 2014, doi: 10.1186/2040-2392-5-18.

[253] D. V. M. Bishop, G. Laws, C. Adams, and C. F. Norbury, "High Heritability of Speech and Language Impairments in 6-year-old Twins Demonstrated Using Parent and Teacher Report," *Behav. Genet.*, vol. 36, no. 2, pp. 173–184, Mar. 2006, doi: 10.1007/s10519-005-9020-0.

[254] D. H. Skuse *et al.*, "Social Communication Competence and Functional Adaptation in a General Population of Children: Preliminary Evidence for Sex-by-Verbal IQ Differential Risk," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 48, no. 2, pp. 128–137, Feb. 2009, doi: 10.1097/CHI.0b013e31819176b8.

[255] J. Min, G. Hemani, G. D. Smith, C. L. Relton, and M. Suderman, "Meffil: efficient normalisation and analysis of very large DNA methylation samples," *bioRxiv*, p. 125963, Apr. 2017, doi: 10.1101/125963.

[256] Y. Chen *et al.*, "Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray," *Epigenetics*, vol. 8, no. 2, pp. 203–209, Feb. 2013, doi: 10.4161/epi.23470.

[257] M. J. Aryee *et al.*, "Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.," *Bioinformatics*, vol. 30, no. 10, pp. 1363–9, May 2014, doi: 10.1093/bioinformatics/btu049.

[258] P. Du *et al.*, "Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis," *BMC Bioinformatics*, vol. 11, 2010, doi: 10.1186/1471-2105-11-587.

[259] O. M. de Goede *et al.*, "Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells.," *Clin. Epigenetics*, vol. 7, no. 1, p. 95, 2015, doi: 10.1186/s13148-015-0129-6.

[260] K.-L. Tiong and C.-H. Yeang, "MGSEA - a multivariate Gene set enrichment analysis.," *BMC Bioinformatics*, vol. 20, no. 1, p. 145, Mar.

2019, doi: 10.1186/s12859-019-2716-6.

[261] N. N. Parikshak *et al.*, "Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism," *Nature*, vol. 540, no. 7633, pp. 423–427, Dec. 2016, doi: 10.1038/nature20612.

[262] T. R. Gaunt *et al.*, "Systematic identification of genetic influences on methylation across the human life course.," *Genome Biol.*, vol. 17, p. 61, Mar. 2016, doi: 10.1186/s13059-016-0926-z.

[263] J.-C. Lambert *et al.*, "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease," *Nat. Genet.*, vol. 45, no. 12, pp. 1452–1458, Oct. 2013, doi: 10.1038/ng.2802.

[264] R. A. Gibbs *et al.*, "The International HapMap Project," *Nature*, vol. 426, no. 6968, pp. 789–796, Dec. 2003, doi: 10.1038/nature02168.

[265] O. Delaneau, J. Marchini, and J.-F. Zagury, "A linear complexity phasing method for thousands of genomes," *Nat. Methods*, vol. 9, no. 2, pp. 179–181, Dec. 2011, doi: 10.1038/nmeth.1785.

[266] S. Purcell *et al.*, "PLINK: a tool set for whole-genome association and population-based linkage analyses.," *Am. J. Hum. Genet.*, vol. 81, no. 3, pp. 559–75, Sep. 2007, doi: 10.1086/519795.

[267] B. K. Bulik-Sullivan *et al.*, "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.," *Nat. Genet.*, vol. 47, no. 3, pp. 291–295, Feb. 2015, doi: 10.1038/ng.3211.

[268] B. K. Bulik-Sullivan *et al.*, "An atlas of genetic correlations across human diseases and traits.," *Nat. Genet.*, vol. 47, no. 11, pp. 1236–41, Sep. 2015, doi: 10.1038/ng.3406.

[269] E. Kasem, T. Kurihara, and K. Tabuchi, "Neurexins and neuropsychiatric disorders," *Neurosci. Res.*, vol. 127, pp. 53–60, 2018, doi: 10.1016/j.neures.2017.10.012.

[270] T. Williams, "Genomics offers new possibilities for global health through

international collaboration," *DMM Dis. Model. Mech.*, vol. 3, no. 3–4, pp. 131–133, 2010, doi: 10.1242/dmm.005215.

[271] A. Gomes and B. Korf, "Chapter 5 - Genetic Testing Techniques," N. H. Robin and M. B. B. T.-P. C. G. Farmer, Eds. Elsevier, 2018, pp. 47–64.

[272] I. S. Hagemann, "Chapter 1 - Overview of Technical Aspects and Chemistries of Next-Generation Sequencing," S. Kulkarni and J. B. T.-C. G. Pfeifer, Eds. Boston: Academic Press, 2015, pp. 3–19.

[273] K. Mozhui *et al.*, "Genetic regulation of Nrnx1 expression: An integrative cross-species analysis of schizophrenia candidate genes," *Transl. Psychiatry*, vol. 1, no. April, pp. 1–11, 2011, doi: 10.1038/tp.2011.24.

[274] P. Wang *et al.*, "CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in neurodevelopment," *Mol. Autism*, vol. 6, no. 1, p. 55, 2015, doi: 10.1186/s13229-015-0048-6.

[275] J. L. Stone *et al.*, "Rare chromosomal deletions and duplications increase risk of schizophrenia," *Nature*, vol. 455, no. 11, pp. 237–241, 2008, doi: 10.1038/nature07239.

[276] L. Zeng, P. Zhang, L. Shi, V. Yamamoto, W. Lu, and K. Wang, "Functional Impacts of NRXN1 Knockdown on Neurodevelopment in Stem Cell Models," *PLoS One*, vol. 8, no. 3, 2013, doi: 10.1371/journal.pone.0059685.

[277] M. J. Gandal *et al.*, "Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap," *Science (80-.).*, vol. 359, no. 6376, pp. 693–697, 2018, doi: 10.1126/science.aad6469.

[278] D. Velmeshev *et al.*, "Single-cell genomics identifies cell type–specific molecular changes in autism," *Science (80-.).*, vol. 364, no. 6441, pp. 685–689, 2019, doi: 10.1126/science.aav8130.

[279] L. Shorts-Cary *et al.*, "Bone morphogenetic protein and retinoic acid-inducible neural specific protein-3 is expressed in gonadotrope cell

pituitary adenomas and induces proliferation, migration, and invasion," *Endocrinology*, vol. 148, no. 3, pp. 967–975, 2007, doi: 10.1210/en.2006-0905.

[280] C. Baldi *et al.*, "Expanding the clinical and genetic spectra of NKX6-2-related disorder," *Clin. Genet.*, vol. 93, no. 5, pp. 1087–1092, 2018, doi: 10.1111/cge.13221.

[281] X. Lin, M. W. State, F. M. Vaccarino, J. Greally, M. Hass, and J. F. Leckman, "Identification, chromosomal assignment, and expression analysis of the human homeodomain-containing gene Orthopedia (OTP)," *Genomics*, vol. 60, no. 1, pp. 96–104, 1999, doi: 10.1006/geno.1999.5882.

[282] G. Pérez-Rubio, M. E. Pérez-Rodríguez, A. Ramírez-Venegas, R. Sansores, A. Camarena, and R. Falfán-Valencia, "Genetic variants in NRXN1 , CHRNA3 , CHRNA5 and CHRNB4 are associated with smoking and increased consumption of cigarettes per day in a Mexican mestizo population," no. September 2015, p. PA1224, 2015, doi: 10.1183/13993003.congress-2015.pa1224.

[283] M. O. Freitas *et al.*, "PEX5 protein binds monomeric catalase blocking its tetramerization and releases it upon binding the N-terminal domain of PEX14," *J. Biol. Chem.*, vol. 286, no. 47, pp. 40509–40519, 2011, doi: 10.1074/jbc.M111.287201.

[284] H. Yamashita *et al.*, "Characterization of human and murine PMP20 peroxisomal proteins that exhibit antioxidant activity in vitro," *J. Biol. Chem.*, vol. 274, no. 42, pp. 29897–29904, 1999, doi: 10.1074/jbc.274.42.29897.

[285] F. Ramírez *et al.*, "deepTools2: a next generation web server for deep-sequencing data analysis," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W160–W165, 2016, doi: 10.1093/nar/gkw257.

[286] Y. L. Chien *et al.*, "The central nervous system patterning gene variants

associated with clinical symptom severity of autism spectrum disorders," *J. Formos. Med. Assoc.*, vol. 116, no. 10, pp. 755–764, 2017, doi: 10.1016/j.jfma.2016.11.015.

[287] H. Li, T. Yamagata, M. Mori, and M. Y. Momoi, "Absence of causative mutations and presence of autism-related allele in FOXP2 in Japanese autistic patients," *Brain Dev.*, vol. 27, no. 3 SPEC. ISS., pp. 207–210, 2005, doi: 10.1016/j.braindev.2004.06.002.

[288] O. Mercati *et al.*, "CNTN6 mutations are risk factors for abnormal auditory sensory perception in autism spectrum disorders," *Mol. Psychiatry*, vol. 22, no. 4, pp. 625–633, 2017, doi: 10.1038/mp.2016.61.

[289] M. Poot, "A candidate gene association study further corroborates involvement of contactin genes in autism," *Mol. Syndromol.*, vol. 5, no. 5, pp. 229–235, 2014, doi: 10.1159/000362891.

[290] T. Hubert, S. Grimal, P. Carroll, and A. Fichard-Carroll, "Collagens in the developing and diseased nervous system," *Cell. Mol. Life Sci.*, vol. 66, no. 7, pp. 1223–1238, 2009, doi: 10.1007/s00018-008-8561-9.

[291] N. F. M. Olde Loohuis, G. J. M. Martens, H. van Bokhoven, B. B. Kaplan, J. R. Homberg, and A. Aschrafi, "Altered expression of circadian rhythm and extracellular matrix genes in the medial prefrontal cortex of a valproic acid rat model of autism," *Prog. Neuro-Psychopharmacology Biol. Psychiatry*, vol. 77, no. December 2016, pp. 128–132, 2017, doi: 10.1016/j.pnpbp.2017.04.009.

[292] S. L. Valk, A. Di Martino, M. P. Milham, and B. C. Bernhardt, "Multicenter mapping of structural network alterations in autism," *Hum. Brain Mapp.*, vol. 36, no. 6, pp. 2364–2373, Jun. 2015, doi: 10.1002/hbm.22776.

[293] H. C. Hazlett *et al.*, "Early Brain Overgrowth in Autism Associated With an Increase in Cortical Surface Area Before Age 2 Years," *Arch. Gen. Psychiatry*, vol. 68, no. 5, pp. 467–476, May 2011, doi:

10.1001/archgenpsychiatry.2011.39.

[294] I. Voineagu *et al.*, "Transcriptomic analysis of autistic brain reveals convergentmolecular convergent molecular pathology," *Nature*, vol. 474, no. 7351, pp. 380–384, 2011, doi: 10.1038/nature10110.

[295] T. Yagi and M. Takeichi, "Cadherin superfamily genes: Functions, genomic organization, and neurologic diversity," *Genes Dev.*, vol. 14, no. 10, pp. 1169–1180, 2000, doi: 10.1101/gad.14.10.1169.

[296] M. Yamagata, X. Duan, and J. R. Sanes, "Cadherins interact with synaptic organizers to promote synaptic differentiation," *Front. Mol. Neurosci.*, vol. 11, no. April, pp. 1–15, 2018, doi: 10.3389/fnmol.2018.00142.

[297] C. Redies, N. Hertel, and C. A. Hübner, "Cadherins and neuropsychiatric disorders," *Brain Res.*, vol. 1470, pp. 130–144, 2012, doi: 10.1016/j.brainres.2012.06.020.

[298] G. Haller, P. Li, C. Esch, S. Hsu, A. M. Goate, and J. H. Steinbach, "Functional characterization improves associations between rare non-synonymous variants in CHRNB4 and smoking behavior," *PLoS One*, vol. 9, no. 5, 2014, doi: 10.1371/journal.pone.0096753.

[299] M. E. Pérez-rodríguez, A. Ramírez-venegas, and J. García-colunga, "SNPs in NRXN1 and CHRNA5 are associated to smoking and regulation of GABAergic and glutamatergic pathways," *Pharmacogenomics*, vol. 17, no. 10, pp. 1145–1158, 2016.

[300] J. Nussbaum *et al.*, "Significant association of the neurexin-1 gene (NRXN1) with nicotine dependence in European- and African-American smokers," *Hum. Mol. Genet.*, vol. 17, no. 11, pp. 1569–1577, 2008, doi: 10.1093/hmg/ddn044.

[301] L. J. Bierut *et al.*, "Novel Genes Identified in a High Density Genome Wide Association study for nicotine dependence," *Hum. Mol. Genet.*, vol. 16, no. 1, pp. 24–35, 2007, doi: 10.1093/hmg/ddl441.

[302] A. Pattyn, X. Morin, H. Cremer, C. Goridis, and J. F. Brunet, "The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives.," *Nature*, vol. 399, no. 6734, pp. 366–370, May 1999, doi: 10.1038/20700.

[303] J. Mong *et al.*, "Transcription Factor-Induced Lineage Programming of Noradrenaline and Motor Neurons from Embryonic Stem Cells," *Stem Cells*, vol. 32, no. 3, pp. 609–622, Mar. 2014, doi: 10.1002/stem.1585.

[304] K. T. MacDonald, R. A. Mosquera, A. Yadav, M. C. Caldas-Vasquez, H. Emanuel, and K. Rennie, "Neurocognitive functioning in individuals with congenital central hypoventilation syndrome," *BMC Pediatr.*, vol. 20, no. 1, pp. 1–8, 2020, doi: 10.1186/s12887-020-2006-5.

[305] D. M. Sikora, K. Pettit-Kekel, J. Penfield, L. S. Merkens, and R. D. Steiner, "The near universal presence of autism spectrum disorders in children with Smith–Lemli–Opitz syndrome," *Am. J. Med. Genet. Part A*, vol. 140A, no. 14, pp. 1511–1518, Jul. 2006, doi: 10.1002/ajmg.a.31294.

[306] E. Tierney *et al.*, "Abnormalities of Cholesterol Metabolism in Autism Spectrum Disorders Elaine," *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, vol. 141B, no. 6, pp. 666–668, 2006, doi: 10.1038/jid.2014.371.

[307] Y. Luo *et al.*, "A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia," *Nat. Med.*, vol. 26, no. 9, pp. 1375–1379, 2020, doi: 10.1038/s41591-020-1007-0.

[308] M. Segatto, L. Leboffe, and L. T. and V. Pallottini, "Cholesterol Homeostasis Failure in the Brain: Implications for Synaptic Dysfunction and Cognitive Decline," *Current Medicinal Chemistry*, vol. 21, no. 24. pp. 2788–2802, 2014, doi: http://dx.doi.org/10.2174/09298673216661403031429 02.

[309] A. M. Petrov, M. R. Kasimov, and A. L. Zefirov, "Cholesterol in the Pathogenesis of Alzheimer's, Parkinson's Diseases and Autism: Link to Synaptic Dysfunction," *Acta Naturae*, vol. 9, no. 1, pp. 26–37, 2017,

[Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28461971.

[310] L. Fagerberg *et al.*, "Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics," *Mol. Cell. Proteomics*, vol. 13, no. 2, pp. 397–406, 2014, doi: 10.1074/mcp.M113.035600.

[311] M. He *et al.*, "Mutations in the human SC4MOL gene encoding a methyl sterol oxidase cause psoriasiform dermatitis, microcephaly, and developmental delay," *J. Clin. Invest.*, vol. 121, no. 3, pp. 976–984, 2011, doi: 10.1172/JCI42650.

[312] N. Schrode *et al.*, "Synergistic effects of common schizophrenia risk variants," *Nat. Genet.*, vol. 51, no. 10, pp. 1475–1485, 2019, doi: 10.1038/s41588-019-0497-5.

[313] L. F. Harkin *et al.*, "Neurexins 1-3 each have a distinct pattern of expression in the early developing human cerebral cortex," *Cereb. Cortex*, vol. 27, no. 1, pp. 216–232, 2017, doi: 10.1093/cercor/bhw394.

[314] Z. Fan, X. Chen, and R. Chen, "Transcriptome-wide analysis of TDP-43 binding small RNAs identifies miR-NID1 (miR-8485), a novel miRNA that represses NRXN1 expression," *Genomics*, vol. 103, no. 1, pp. 76–82, 2014, doi: 10.1016/j.ygeno.2013.06.006.

[315] B. Chih, L. Gollan, and P. Scheiffele, "Alternative Splicing Controls Selective Trans-Synaptic Interactions of the Neuroligin-Neurexin Complex," *Neuron*, vol. 51, no. 2, pp. 171–178, 2006, doi: 10.1016/j.neuron.2006.06.005.

[316] E. R. Graf, Y. Kang, A. M. Hauner, and A. M. Craig, "Structure Function and Splice Site Analysis ofthe Synaptogenic Activity ofthe Neurexin-1ᴺᴵLNS Domain Ethan," *J. Neurosci.*, vol. 26, no. 16, pp. 4256–4265, 2006, doi: 10.1523/JNEUROSCI.1253-05.2006.

[317] N. J. F. Gatford, M. Deans, R. R. R. Duarte, G. Chennell, P. Raval, and D. P. Srivastava, "Nanoscopic Clustering of Neuroligin-3 and Neuroligin-4X

Regulates Growth Cone Organization and Size," *bioRxiv*, p. 546499, 2019, doi: 10.1101/546499.

[318] S. T. Schafer *et al.*, "Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons," *Nat. Neurosci.*, vol. 22, no. 2, pp. 243–255, 2019, doi: 10.1038/s41593-018-0295-x.

[319] M. S. Sons *et al.*, "A-Neurexins Are Required for Efficient Transmitter Release and Synaptic Homeostasis At the Mouse Neuromuscular Junction," *Neuroscience*, vol. 138, no. 2, pp. 433–446, 2006, doi: 10.1016/j.neuroscience.2005.11.040.

[320] W. Zhang *et al.*, "Extracellular domains of α-neurexins participate in regulating synaptic transmission by selectively affecting N- and P/Q-type Ca2+ channels," *J. Neurosci.*, vol. 25, no. 17, pp. 4330–4342, 2005, doi: 10.1523/JNEUROSCI.0497-05.2005.

[321] M. E. Spira and A. Hai, "Multi-electrode array technologies for neuroscience and cardiology," *Nat. Nanotechnol.*, vol. 8, no. 2, pp. 83–94, 2013, doi: 10.1038/nnano.2012.265.

[322] C. Grienberger and A. Konnerth, "Imaging Calcium in Neurons," *Neuron*, vol. 73, no. 5, pp. 862–885, 2012, doi: 10.1016/j.neuron.2012.02.011.

[323] K. Miyazaki, J. E. Lisman, and W. N. Ross, "Improvements in simultaneous sodium and calcium imaging," *Front. Cell. Neurosci.*, vol. 12, no. January, pp. 1–10, 2019, doi: 10.3389/fncel.2018.00514.

[324] V. Harrison, L. Connell, J. Hayesmoore, J. McParland, M. G. Pike, and E. Blair, "Compound heterozygous deletion of NRXN1 causing severe developmental delay with early onset epilepsy in two sisters," *Am. J. Med. Genet. Part A*, vol. 155, no. 11, pp. 2826–2831, 2011, doi: 10.1002/ajmg.a.34255.

[325] E. Deneault *et al.*, "CNTN5 -/+ or EHMT2 -/+ human iPSC-derived neurons from individuals with autism develop hyperactive neuronal networks," *Elife*, vol. 8, pp. 1–26, 2019, doi: 10.7554/eLife.40092.

[326] L. M. Grunwald *et al.*, “Comparative characterization of human induced pluripotent stem cells (hiPSC) derived from patients with schizophrenia and autism,” *Transl. Psychiatry*, vol. 9, no. 1, 2019, doi: 10.1038/s41398-019-0517-3.

[327] D. G. Margineanu, “Epileptic hypersynchrony revisited,” *Neuroreport*, vol. 21, no. 15, pp. 963–967, 2010, doi: 10.1097/WNR.0b013e32833ed111.

[328] P. J. Uhlhaas and W. Singer, “Neural Synchrony in Brain Disorders: Relevance for Cognitive Dysfunctions and Pathophysiology,” *Neuron*, vol. 52, no. 1, pp. 155–168, 2006, doi: 10.1016/j.neuron.2006.09.020.

[329] A. Brooks-Kayal, “Epilepsy and autism spectrum disorders: Are there common developmental mechanisms?,” *Brain Dev.*, vol. 32, no. 9, pp. 731–738, 2010, doi: 10.1016/j.braindev.2010.04.010.

[330] W. J. Gao and P. S. Goldman-Rakic, “NMDA receptor-mediated epileptiform persistent activity requires calcium release from intracellular stores in prefrontal neurons,” *Exp. Neurol.*, vol. 197, no. 2, pp. 495–504, 2006, doi: 10.1016/j.expneurol.2005.05.018.

[331] M. Fang *et al.*, “Neuroligin-1 Knockdown Suppresses Seizure Activity by Regulating Neuronal Hyperexcitability,” *Mol. Neurobiol.*, vol. 53, no. 1, pp. 270–284, 2016, doi: 10.1007/s12035-014-8999-8.

[332] P. J. Ross *et al.*, “Modeling neuronal consequences of autism-associated gene regulatory variants with human induced pluripotent stem cells,” *Mol. Autism*, vol. 11, no. 1, p. 33, 2020, doi: 10.1186/s13229-020-00333-6.

[333] J. P. Y. Kao, G. Li, and D. A. Auston, *Practical aspects of measuring intracellular calcium signals with fluorescent indicators*, vol. 99. Elsevier Inc., 2010.

[334] J. Negri, V. Menon, and T. L. Young-Pearse, “Assessment of spontaneous neuronal activity In vitro using multi-well multi-electrode arrays: Implications for assay development,” *eNeuro*, vol. 7, no. 1, 2020, doi: 10.1523/ENEURO.0080-19.2019.

[335] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders (5th ed.)*. 2013.

[336] S. Baron-Cohen, S. J. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians.," *J. Autism Dev. Disord.*, vol. 31, no. 1, pp. 5–17, Feb. 2001.

[337] E. B. Robinson *et al.*, "Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%).," *Arch. Gen. Psychiatry*, vol. 68, no. 11, pp. 1113–21, Nov. 2011, doi: 10.1001/archgenpsychiatry.2011.119.

[338] E. Colvert *et al.*, "Heritability of autism spectrum disorder in a UK population-based twin sample," *JAMA Psychiatry*, vol. 72, no. 5, pp. 415–23, Mar. 2015, doi: 10.1001/jamapsychiatry.2014.3028.

[339] B. Tick, P. F. Bolton, F. Happé, M. Rutter, and F. Rijsdijk, "Heritability of autism spectrum disorders: A meta-analysis of twin studies," *J. Child Psychol. Psychiatry Allied Discip.*, vol. 57, no. 5, pp. 585–595, 2016, doi: 10.1111/jcpp.12499.

[340] S. Sandin, P. Lichtenstein, R. Kuja-Halkola, C. Hultman, H. Larsson, and A. Reichenberg, "The Heritability of Autism Spectrum Disorder," *JAMA*, vol. 318, no. 12, p. 1182, Sep. 2017, doi: 10.1001/jama.2017.12141.

[341] R. A. Hoekstra, M. Bartels, C. J. H. Verweij, D. I. Boomsma, S. PE, and B. DI, "Heritability of autistic traits in the general population.," *Arch. Pediatr. Adolesc. Med.*, vol. 161, no. 4, pp. 372–7, Apr. 2007, doi: 10.1001/archpedi.161.4.372.

[342] The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, "Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia," *Mol. Autism*, vol. 8, no. 1, p. 21, Dec. 2017, doi: 10.1186/s13229-017-0137-9.

[343] E. Hannon, M. Weedon, N. Bray, M. O'Donovan, and J. Mill, "Pleiotropic Effects of Trait-Associated Genetic Variation on DNA Methylation: Utility for Refining GWAS Loci," *Am. J. Hum. Genet.*, vol. 100, no. 6, pp. 954–959, Jun. 2017, doi: 10.1016/J.AJHG.2017.04.013.

[344] K. Lunnon *et al.*, "Variation in 5-hydroxymethylcytosine across human cortex and cerebellum," *Genome Biol.*, vol. 17, no. 1, p. 27, Dec. 2016, doi: 10.1186/s13059-016-0871-x.

[345] T. Kleindienst and C. Lohmann, "Simultaneous patch-clamping and calcium imaging in developing dendrites.," *Cold Spring Harb. Protoc.*, vol. 2014, no. 3, pp. 324–328, Mar. 2014, doi: 10.1101/pdb.prot080390.