ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

Biologia Cellulare e Molecolare

Ciclo XXXVI

Settore Concorsuale: 05/E2

Settore Scientifico Disciplinare: BIO/11

Decoding Agc1 Deficiency: Bioinformatics Approaches to Unveil the Functional Implications of Slc25a12 on the Transcriptome and Epigenome

Presentata da:

Nicola Balboni

Supervisore:

Federico M. Giorgi

Coordinatore dottorato:

Vincenzo Scarlato

Co-supervisore:

Barbara Monti

Esame finale anno 2024

1 Contents

2 Abstract 3

3 Scientific CV 5

4 Introduction 6

4.1 Agc1 deficiency: cellular, genetic and metabolic aspects of the disease 6

4.2 RNA-Seq: an outlook on transcriptome analysis 14

4.3 Master regulator analysis for transcriptomic networks 18

4.4 ATAC-Seq: Assay for Transposase-accessible chromatin 22

5 Methods 26

5.1 RNA-Seq pipeline 26

5.1.1 Salmon 26

5.1.2 DESeq2 28

5.1.3 GSEA 31

5.1.4 Majiq 32

5.1.5 Corto 34

5.1.6 Shiny and Docker for Interactive Web Application Frameworks 34

5.2 ATAC-Seq pipeline 37

5.2.1 MACS 37

5.2.2 diffTF 38

6 Results 41

6.1 RNA-Seq OliNeu cells 41

6.1.1 Experimental setup 41

6.1.2 Differential Gene Expression and GSEA 43

6.1.3 Agc1 Explorer web-app 48

6.1.4 Splicing analysis 50

6.1.5 Master Regulator Analysis 53

6.2 ATAC-Seq OliNeu 57
6.2.1 Experimental setup 57
6.2.2 Differential Accessibility Analysis 57
6.2.3 Comparison with RNA-Seq results 63
6.3 RNA-Seq iPSCs 67
6.3.1 Experimental setup 67
6.3.2 Exploratory Data analysis and Differential Gene Expression 68
6.3.3 GSEA results 71
7 Discussion 74
8 References 82

2 Abstract

In the brain, mutations in SLC25A12 gene encoding AGC1 cause an ultra-rare genetic disease reported as a developmental and epileptic encephalopathy associated with global cerebral hypomyelination. Symptoms of the disease include diffused hypomyelination, arrested psychomotor development, severe hypotonia, seizures and are common to other neurological and developmental disorders. Amongst the biological components believed to be most affected by AGC1 deficiency are oligodendrocytes, glial cells responsible for myelination. Recent studies (Poeta et al, 2022) have also shown how altered levels of transcription factors and epigenetic modifications greatly affect proliferation and differentiation in oligodendrocyte precursor cells (OPCs).

In this study we explore the transcriptomic landscape of Agc1 in two different system models: OPCs silenced for Agc1 and iPSCs from human patients differentiated to neural progenitors. Analyses range from differential expression analysis, alternative splicing, master regulator analysis. ATAC-seq results on OPCs were integrated with results from RNA-Seq to assess the activity of a TF based on the accessibility data from its putative targets, which allows to integrate RNA-Seq data to infer their role as either activators or repressors.

All the findings for this model were also integrated with early data from iPSCs RNA-seq results, looking for possible commonalities between the two different system models, among which we find a downregulation in genes encoding for SREBP, a transcription factor regulating fatty acids biosynthesis, a key process for myelination which could explain the hypomyelinated state of patients. We also find that in both systems cells tend to form more neurites, likely losing their ability to differentiate, considering their progenitor state. We also report several alterations in the chromatin state of cells lacking Agc1, which confirms the hypothesis for which Agc1 is not a disease restricted only to metabolic alterations in the cells, but there is a profound shift of the regulatory state of these cells.

The work reported in this thesis is solely restricted to projects funded directly by Telethon, which exclusively focus on the study of Agc1 deficiency. Several other projects were accomplished during the time of the PhD, even though they are not reported here, they are listed in Chapter 3 “Scientific CV”.

Abbreviations

ACSS1 (Acetyl-CoA Synthetase 1), AD (Alzheimer’s Disease), ADP (Adenosine Diphosphate), AGC1 (Aspartate-Glutamate Carrier 1), AGC2 (Aspartate-Glutamate Carrier 2), ASPA (Aspartoacylase), ATPAF2 (ATP synthase mitochondrial F1 complex Assembly Factor 2), BGN (ByGlycan), CDF (Cumulative Distribution Function), CHODL (Chondrolectin), CMT1A (Charcot-Marie-Tooth type-1A), CNS (Central Nervous System), CORO2A (Coronin 2A), DNA (Deoxyribonucleic Acid), EFCAB1 (EF-hand calcium-binding domain-containing protein 1), FASN (Fatty Acids Synthase N), FAT3 (FAT Atypical Cadherin 3), GSEA (Gene Set Enrichment Analysis), HDAC (Histone Deacetylase), iPSC (induce Pluripotent Stem Cell), ITGA5 (Integrin Alpha 5), KCNT2 (Potassium channel subfamily T member 2), LFC (Log2(FoldChange)), LSV (Local Splicing Variations), MAS (Malate-Aspartate Shuttle), MR (Master Regulator), MRPL55 (Mitochondrial Ribosomal Protein L55), MSigDb (Molecular Signature Database), MYT1 (Myelin Transcription Factor 1), NAA (N-Acetyl-Aspartate), NAT8L (N-Acetyltransferase 8 Like), NES (Normalized Enrichment Score), NPC (Neural Precursor Cell), OL (Oligodendrocyte), OMG (Oligodendrocyte Myelin Glycoprotein), OPC (Oligodendrocyte Precursor Cell), PCA (Principal Component Analysis), PDGFRα (Platelet Derived Growth Factor α), PMP22 (Peripheral Myelin Protein 22), PROX1 (Prospero Homeobox 1), PSI (Percent Selected Index), PSMB9 (Proteasome subunit beta type-9), PTN (Pleiotrophin), PTPRZ1 (Protein Tyrosine Phosphatase Receptor Type Z1), RA (Retinoic Acid), RAI1 (Retinoic Acid Induced 1), RARb (Retinoic Acid Receptor Beta), RARG (Retinoic Acid Receptor Gamma), RNA (Ribonucleic Acid), SMARCC2 (SWI/SNF Related Matrix Associated Actin Dependent Regulator Of Chromatin Subfamily C Member 2), SNAP47 (Synaptosome Associated Protein 47), SREBP (Sterol Regulatory Binding Protein), SVZ (Sub-Ventricular Zone), TAP1 (Transporter 1, ATP-binding cassette), TF (Transcription Factor), TPM (Transcripts Per Million), TRIM11 (Tripartite Motif Containing 11), VLCFA (Very Long-Chain Fatty Acids).

3 Scientific CV

Removed per thesis upload guidelines

4 Introduction

4.1 Agc1 deficiency: cellular, genetic and metabolic aspects of the disease

Manifesting as a neurological and hypomyelinating disorder, AGC1 deficiency (DEE39; OMIM: 612949; ICD-10: G31.8; ORPHA: ORPHA353217) develops from mutations within the SLC25A12 gene, encoding the mitochondrial aspartate-glutamate carrier isoform 1 (AGC1), that impair its transporter activity; during the initial months of life infantile development proceeds normally, only subsequently showing multiple and severe symptoms, such as neuromuscular delay, hypotonia, epilepsy, and psychomotor retardation. Magnetic resonance imaging (MRI) revealed a diminution in cerebral volume and age-correlated hypomyelination, along a reduction in N-acetyl aspartate (NAA) and an elevation in blood lactate, indicative of mitochondrial dysfunction in affected individuals (Wibom et al. 2009; Falk et al. 2014; Pfeiffer et al. 2020).

Two AGC isoforms, AGC1/SLC25A12 (aralar1) and AGC2/SLC25A13 (citrin), exhibit 77.8% sequence homology, with the former predominantly expressed in excitable tissues—heart, brain, and skeletal muscles—and the latter in the liver and non-excitable tissues (Del Arco, Agudo, and Satrústegui 2000); both isoforms facilitate a Ca^{2+}-stimulated, unidirectional exchange between intra-mitochondrial aspartate and cytosolic glutamate plus a proton, and constitute components of the malate-aspartate NADH shuttle (MAS, Fig 4.1A), pivotal for the mitochondrial entry of glycolysis-derived NADH reducing equivalents, key to correct pyruvate oxidation (Llorente-Folch et al. 2013; E. B. Taylor 2017).

A dysfunctional MAS caused by AGC1 failure produces a block in lactate utilization by neurons, which prevents the protective role of lactate on neuronal excitotoxicity (the pathological process by which nerve cells are damaged and killed by excessive stimulation by neurotransmitters such as glutamate and similar substances). This has been demonstrated both in AGC1-KO neurons in vitro and in $Agc1^{+/-}$ mice in vivo. During epilepsy, brain glutamate and lactate increase, and lack of Agc1-MAS activity limits glutamate-induced stimulation of neuronal mitochondrial respiration, enhancing cytosolic ATP/ADP ratio drop upon glutamate exposure. $Agc1^{+/-}$ mice are more susceptible to excitotoxicity, probably because a failure to use

lactate is detrimental, and this hypersensitizing effect to excitotoxicity could also occur in Aralar-MAS deficient patients (Pardo et al. 2022).

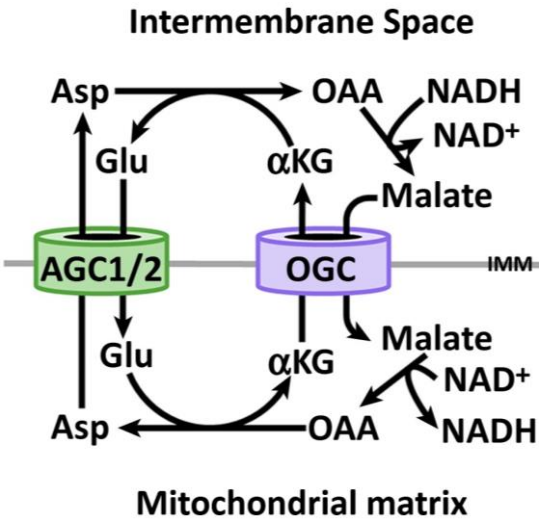

Figure 4.1A – Schematic representation of Malate-Aspartate Shuttle (MAS). Adapted from Taylor et al., Trend in Cell Biology 2017

As previously mentioned, the AGC1's role is to exchange glutamate and aspartate across the inner mitochondrial membrane, the latter being transported outside the mitochondrial matrix in the form of N-Acetylaspartate (NAA). Neurons are the primary NAA producers in the brain, which are then moved outside the mitochondria, to successively be delivered to oligodendrocytes (OLs), brain cells responsible for CNS myelination. NAA thereby provides an acetate source for acetyl-CoA synthesis and subsequent myelin-associated lipid production, establishing a continuous cross-talk and cooperation between neurons and OLs to support myelination (Moffett et al. 2013, Fig. 4.1B).

NAA also covers a role in modulating protein aggregation, which is a common feature in various neurodegenerative disorders (Warepam et al. 2021). Alterations in NAA levels is associated with various neurological disorders involving neuronal loss or dysfunction, such as amyotrophic lateral sclerosis, Alzheimer's, multiple sclerosis, epilepsy, schizophrenia, cerebral ischemia, and glioblastoma (Kantarci and Jack 2003; Criste and Trapp 2006; Abbott and Bustillo 2006). Evidence indicates that NAA is an important precursor of the neuropeptide N-acetylaspartylglutamate (NAAG),

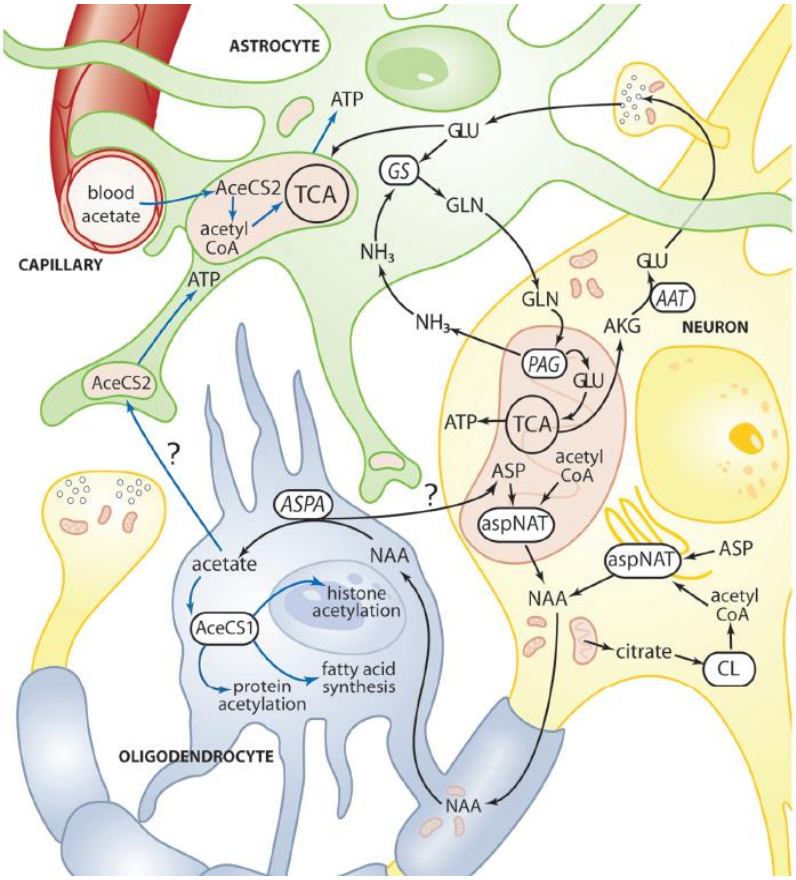

Figure 4.1B – Schematic representation of the metabolic traffic in neurons, oligodendrocytes, and astrocytes (Moffet et al. 2013).

and may be involved in osmoregulation of neuronal cells, axonglial signaling, maintenance of nitrogen balance in the brain, providing a basic source of acetate required for myelin lipid biosynthesis in oligodendrocytes, and facilitating energy metabolism in neuronal mitochondria (Moffett et al. 2007; D. L. Taylor et al. 1995).

Once NAA reaches OLs, it is converted to aspartate and acetate in a reaction catalyzed by aspartoacylase (ASPA); the acetate group is then used by acetyl-CoA synthetase to form acetyl-CoA which is fed to Fatty Acid Synthase (FASN). FASN is a crucial enzyme for fatty acid synthesis, which catalyzes the synthesis of medium and long-chain fatty acids, with its expression being regulated by sterol regulatory element-binding proteins (SREBP, encoded by SREBF1 and SREBF2 genes) and SREBP cleavage-activating protein (SCAP). Notably, SREBP activation is linked to cholesterol levels, establishing a homeostatic connection that activates SREBP when cholesterol levels diminish (Saher and Simons 2010). The elongation of long-chain fatty acids into very long-chain fatty acids (VLCFA) in the endoplasmic reticulum is also underscored for its significance in maintaining myelin, with both

deficiency and excess of VLCFAs being associated with myelin defects (Chrast et al. 2011). Additionally, the enzyme Pyruvate Dehydrogenase, pivotal for synthesizing acetyl-CoA from glycolytic sources, has been linked to myelin defects when deficient, although some studies present contrasting findings (Pliss et al. 2013).

The fatty acid synthesis pathway is required not only for the correct axon myelination, but also in ensuring a timely remyelination after myelin damage, a process which demands an exceptionally rapid synthesis of myelin proteins and lipids within a restricted time window, a task achieved through the coordinated gene expression of myelin components (Targett et al. 1996).

OLs are the brain cells responsible for the myelination process in the central nervous system. Axonal myelination is a fundamental process to grant neurons the capability for an efficient saltatory signal conduction, along with providing structural and trophic support to neurons. Even though myelination occurs during the initial phases of development, this process does not stop in the final phases of development but continues throughout the adult life to grant a continuous support to neurons (Yeung et al. 2014).

Mature OLs are formed in a process known as oligodendrogenesis, which is a progressive process consisting of 3 stages: oligodendrocyte precursor cells (OPCs, whose precursors are neural stem cells) which are able to proliferate and migrate to different areas of the CNS, but are not capable of myelination, pre-myelinating oligodendrocytes, which are already settled on a site and start forming connections with surrounding neurons, but do not form myelin sheaths yet, and finally mature OLs, which are fully differentiated and have effectively myelinated neighboring axons (Hughes and Stockton 2021). The differentiation process is determined by a plethora of timed expression of differentiation and proliferation markers, which are illustrated in Figure 4.1C (“A Guide to Oligodendrocyte Markers”):

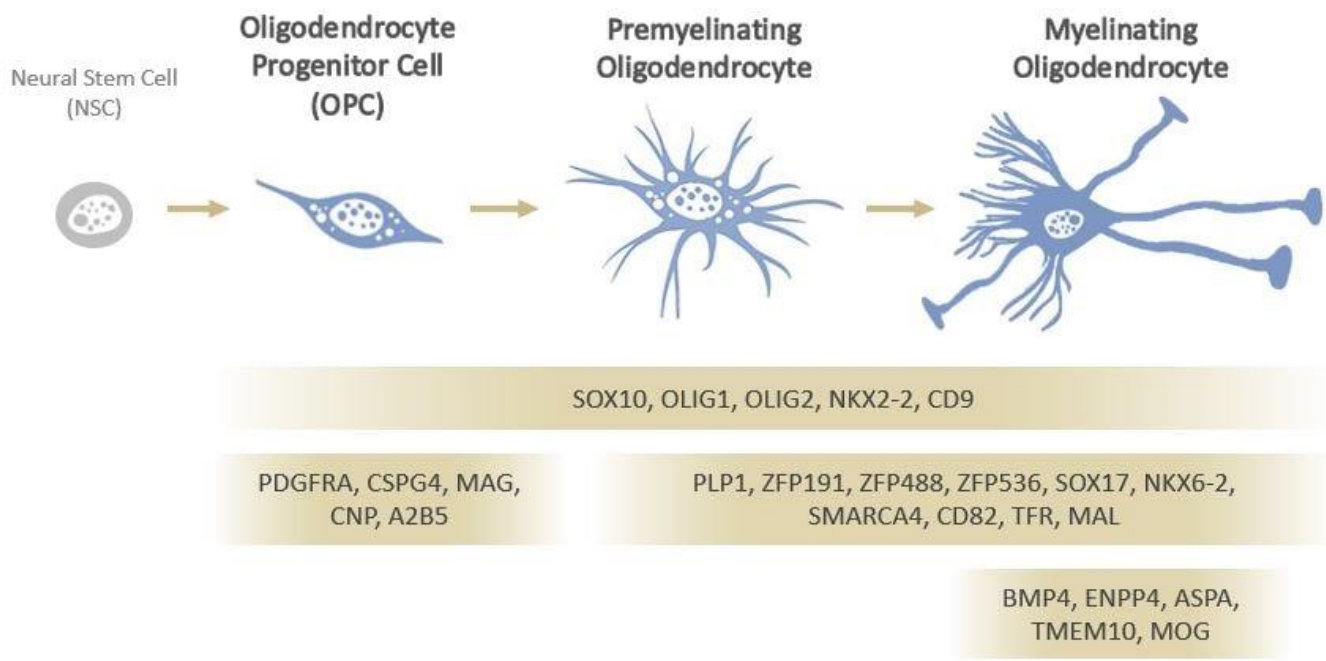

Figure 4.1C – Schematic representation of oligodendrogenesis

While we will not dive into the specifics of each single marker, it is important to remark the importance of:

- Sox10 (along with expression of Sox8 and Sox9, all together composing the SoxE family of transcription factors), is a transcription factor that influences neural precursor cells to become OPCs and is required throughout the whole process of oligodendrogenesis to reach maturity (Pozniak et al. 2010)
- Olig1 and Olig2 are transcription factors expressed to restricted areas during embryonic stages in the human central nervous system, with both of them preceding expression of OPCs markers (PDGFRA, NG2), and their distribution in brain cells correlates well with the distribution of OLs (Jakovcevski and Zecevic 2005)

One of the key regulatory processes determining cell fate is the interplay between Transcription Factors (TFs) and the chromatin state of the cells, which encompasses different types of modifications, including post-translational modifications of histones including acetylation/methylation which all concur to make gene promoters more or less accessible to modulate expression and consequently

cell fate (J. K. Kim, Samaranayake, and Pradhan 2009; Copray et al. 2009). The key feature that determines epigenetic regulation impact on cell fate and development is that epigenetic marks are inherited throughout the offspring of the cell, thus establishing a lineage-specific gene expression which controls what specific phenotype the cell will actually differentiate into (D'Urso and Brickner 2014).

Chromatin remodeling is therefore associated with a progressive restriction of the pool of potential cell lineages a cell can differentiate to even for cells of the central nervous system: multipotent neural stem cells are routed towards OPC differentiation via the suppression of neuronal and astrocytic genes by the activity of histone deacetylases (HDACs) (A. Liu et al. 2007). A well-established model for transition from OPCs to myelinating OLs theorized a progressive transition mediated by a combination downregulation of transcription inhibitors expression followed by an upregulation of myelin-associated genes activators (J. Liu and Casaccia 2010) (Figure 4.1D). Because of the crucial role chromatin remodeling has when studying cells that are not fully differentiated, especially OPCs, it is therefore important not only to consider marker levels at the protein levels, but also the state of the

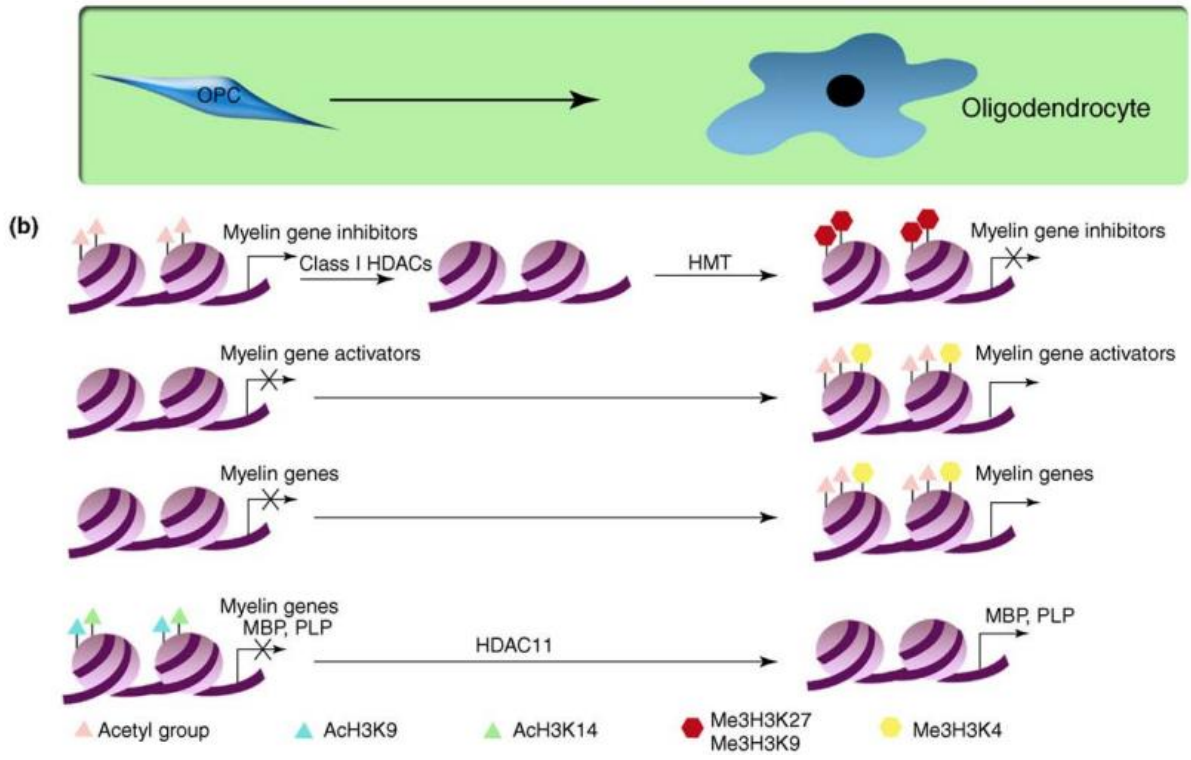

Figure 4.1D – Schematic illustration of chromatin and transcription regulation in OPC to OL transition (from J. Liu and Casaccia, 2010)

chromatin at the promoter level and transcript levels, in order to determine which cell fate is characterizing a specific cell line.

Historically, research has predominantly targeted mature neurons in the murine AGC1 deficiency model, uncovering significant neuronal metabolic disturbances (Llorente-Folch et al. 2013; Ramos et al. 2011; Juaristi et al. 2017). However, recent findings demonstrate that AGC1 down-regulation impedes proliferation in both neuronal precursor cells (NPCs) and oligodendrocyte precursor cells, precipitating spontaneous and premature differentiation in the latter. This proliferation impediment in NPCs correlates with diminished mitochondrial respiration and resultant energetic deficiency, while in OPCs, it primarily associates with a dysregulation in the expression of trophic factors and receptors implicated in proliferation/differentiation (Profilo, Peña-Altamira, and Corricelli 2017; Petralla et al. 2019), attributable to transcriptional alterations stemming from epigenetic modifications in either the OPCs themselves or in the Neural Stem Cells (NSCs) from which they originate (Poeta et al. 2022). Epigenetic modulation of gene expression via histone acetylation plays a pivotal role in both NSCs and OPCs proliferation/differentiation (Juliandi, Abematsu, and Nakashima 2010; Emery and Lu 2015; Hernandez and Casaccia 2015), and NAA can serve as an acetate source for histone acetylation (Long et al. 2013; Bogner-Strauss 2017), thereby suggesting a potential nexus between the perturbed metabolic/mitochondrial state and the epigenetic/transcriptional alterations, which may influence OPC pool maintenance, their differentiation into OLs, and consequently, the myelination/remyelination processes.

SLC25A12 mutations identified in AGC1 deficiency invariably precipitate a diminution or total cessation of carrier activity, concomitant with the emergence of pathological features ubiquitously observed in other genetically inherited white matter diseases (Wibom et al. 2009; Del Arco, Agudo, and Satrústegui 2000). These diseases are difficult to discern from one another, due to shared clinical phenotypes and common biomarkers that, even with the application of diverse screening assays, do not always provide a clear diagnostic pathway, necessitating a multifaceted approach to understanding and managing mitochondrial diseases effectively. This complexity in diagnosis and management also underscores the intricate nature of

mitochondrial diseases and the need for further research to enhance diagnostic and therapeutic strategies (Falk et al. 2014).

In the present study, data derived from bulk sequencing (RNA-seq and ATAC-seq) of an in vitro model of immortalized murine OPCs (Oli-Neu), partially silenced for AGC1 expression, were utilized and studied at different levels (chromatin accessibility and gene expression) to assess in a more comprehensive way which state cells affected by AGC1 deficiency were in and to better understand the molecular mechanisms underlying this disease. Data from AGC1 deficiency patients were also obtained in the form of dermal cells reprogrammed into pluripotent cells (iPSCs) subsequently differentiated to neural progenitor cells committed to neural differentiation, to cross the results between the murine model and a more representative human model.

The use of both mouse cell lines and iPSCs is required for the specific nature of the rare disease of the study. Diseases with a low incidence are difficult to model due to the lack and difficulty of collecting data, a con that carries along itself the difficulty of determining common symptoms and phenotypes that characterize the disease. In these terms, mouse cell lines offer immediate availability, genetic stability, and the feasibility of genetic manipulations for functional studies. Nonetheless, they also present significant drawbacks such as physiological differences from human cells and the potential risk of immune reactions in preclinical models. Conversely, iPSC lines provide a model closer to human conditions, allowing for the replication of specific pathological states and the analysis of individual genetic variations. However, their use poses substantial technical challenges, including variability in cell differentiation and the risk of genetic mutations during in vitro manipulation. Setting up a system that uses both models, allows to balance the advantages and disadvantages of each system to attain accurate and clinically relevant outcomes.

The results from this work, taken together, highlight new aspects of AGC1 deficiency that could improve the disease classification and therapies in the future.

4.2 RNA-Seq: an outlook on transcriptome analysis

RNA sequencing (RNA-seq) has emerged as a transformative tool in molecular biology, profoundly influencing our comprehension of genomic processes. Developed over a decade ago, RNA-seq primarily facilitates the analysis of differential gene expression (DGE).

These are the general steps followed when performing an RNA-seq experiment (illustrated in Figure 4.2A):

1. Isolation: RNA molecules are isolated from a cell (single-cell RNA-seq) or from a population of cells (bulk RNA-seq)
2. Fragmentation: RNA molecules are fragmented (on the order of a few hundred bases long).
3. Reverse transcription: RNA is reverse-transcribed into DNA using RNA-dependent DNA polymerases, also known as reverse transcriptases.
4. Amplification: DNA is amplified using polymerase chain reaction, creating amplicons, a set of double-stranded DNA fragments all similar to each other.
5. Sequencing: The amplified DNA amplicons undergo next-generation sequencing. During this process, the sequencer stochastically selects fragments and captures short terminal sequences, either from a single end or both ends, typically spanning approximately a hundred nucleotides. These acquired sequences are termed "sequencing reads": a single sequencing run yields millions of such reads, which are subsequently archived in a digital format.
6. Alignment: Align the sequencing reads to the reference genome: this task can be accomplished by using different computational methods that specifically establish a nucleotide-by-nucleotide correspondence between each read and a specific genomic segment. Given the extensive size of genomes, often encompassing billions of nucleotides, and the millions of reads generated in a typical RNA-seq experiment, this step poses a significant computational challenge. However, it's worth noting that contemporary algorithms, such as kallisto (Bray et al. 2016) and Salmon (Patro et al. 2017), bypass the computationally intensive step of direct nucleotide alignment, opting instead

for approximation methods termed "pseudoalignment" and "quasi-alignment" respectively. Both concepts bear striking similarities.

7. Quantification: For each gene, count/quantify the number of reads that aligned to that specific gene. It's relevant to note that due to potential sequencing inaccuracies and the occurrence of reads that may align to multiple genes, a statistical analysis is employed to infer the originating gene for each read. Thus, the read "quantities" attributed to each gene are inferred values.

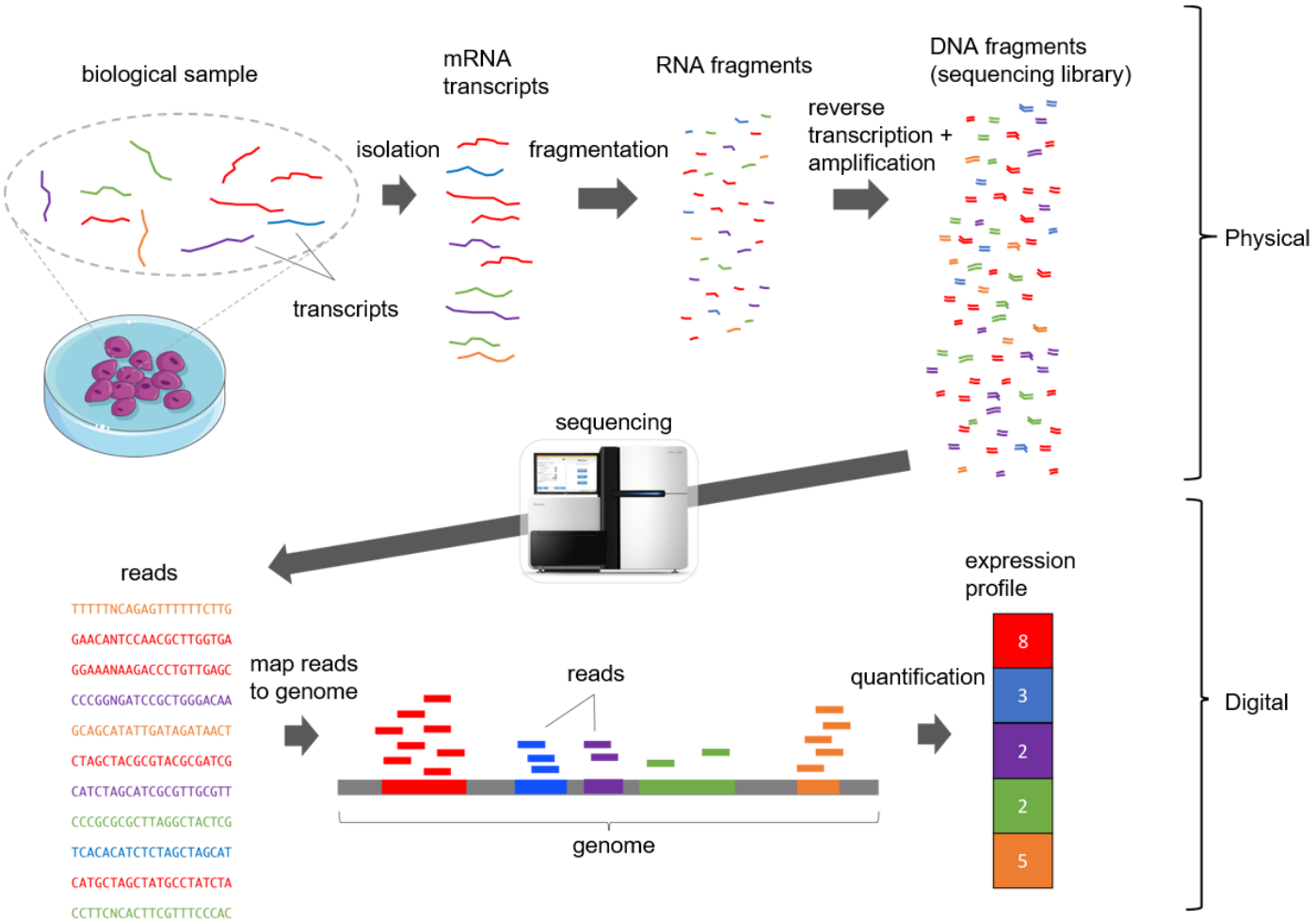

Fig 4.2A – Schematic representation of RNA-seq steps (adapted from https://mbernste.github.io/posts/rna_seq_basics)

The broader applications of RNA-seq have been instrumental in unveiling various aspects of biology. For instance, this technology has elucidated the extent of mRNA splicing (Pan et al. 2008), the regulatory roles of non-coding RNAs (Guttman

et al. 2009), and the functions of enhancer RNAs (Sartorelli and Lauberth 2020). The evolution and adaptation of RNA-seq have been propelled by both wet-lab and computational technological advancements that have facilitated a more comprehensive and unbiased view of RNA biology and the transcriptome, superseding the capabilities of previous microarray-based methodologies.

RNA-seq has proven as an indispensable tool for modern molecular biology studies, especially for brain-affecting diseases, the results from transcriptomics experiment can give new key insights to treat these sort of diseases for which older techniques are failing to deliver new findings and treatments (“Overcoming Gaps in the Treatment of Neurodegenerative Disease” 2020): RNA-seq can be useful to study often overlooked aspects of diseases such as alternative splicing, as a recent work from Heberle et al. identified 1,917 clinically relevant genes expressing multiple isoforms in the human frontal cortex; of these, 1,018 had multiple isoforms with different protein coding sequences. This underscores the need to delve deeper into how individual isoforms from a "single" gene influence human health and disease. Notably, 98 of these genes are implicated in brain-related diseases, including Alzheimer's disease-related genes, of which 99 RNA isoforms were differentially expressed between Alzheimer's cases and controls, even when the genes themselves exhibited no differential expression.

One of the other main advantages of RNA-seq data is that it can be made publicly available, thus improving research efficiency by allowing virtually no costs for reproducibility and the possibility to let researchers worldwide study the produced data. The advent of of general and specialized repositories of bulk and single cell RNA-seq data, such as GTEx (Aguet et al. 2017), CCLE (Barretina et al. 2012), Tabula Sapiens (Tabula Sapiens Consortium* et al. 2022), Tabula Muris (Schaum et al. 2018), GEO (Clough and Barrett 2016) and others, has revolutionized biomedical research. These repositories address a critical need in the scientific community by providing comprehensive, centralized, and easily accessible resources for understanding the molecular underpinnings of various diseases. By offering a wealth of genomic and transcriptomic data, these databases enable researchers to explore gene expression patterns across different tissues, conditions, or disease states. This, in turn, facilitates the identification of potential biomarkers, therapeutic targets, and insights into disease mechanisms. Moreover, the standardized data collection and

processing methods employed by these platforms ensure consistency and reproducibility, allowing for more robust cross-study comparisons. In essence, such databases accelerate scientific discoveries, foster collaboration, and pave the way for precision medicine approaches tailored to individual patient profiles.

4.3 Master regulator analysis for transcriptomic networks

The use of transcriptomic networks to study cellular phenotypes comes from the necessity to represent gene expression not only as effects of individual genes, but in a manner that should resemble the real regulatory mechanisms happening inside cells in a more comprehensive way. Cellular phenotypes are rarely the result of a single gene expression, but rather the concerted effect networks of co-regulated genes: this is a good starting point to explain genes working in networks but fails to separate effects caused by direct interactions from those caused by indirect interactions, such as signal cascades.

One of the successful approaches to tackle this challenge has been to represent gene networks as graphs, where genes are nodes and direct regulatory interactions are the edges of the network. An algorithm adopting this strategy is ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks, Margolin et al. 2006), whose focus is to reconstruct the statistical dependencies of a two-way network, where each gene statistical dependences are based on the gene itself and its first-order statistical interactors (genes that are not physically direct interactors but are considered so from a perspective of statistical dependence). The assumption of a two-way network was chosen due to limitations posed by the amount of data required to validate a network (the more the statistical interactions to estimate, the more data is required, which is a sensible parameter since retrieving gene expression data has high costs) and computational feasibility at the time.

Candidate interactions are estimated by computing the mutual information (MI, a measure of statistical dependency) of pairs of gene expression profiles, such that this value is zero if the expression profiles are deemed independent. Mutual information is defined as $I(x,y) = S(x) + S(y) - S(x,y)$ where $S(x)$ is the entropy of gene x and is computed as $S(x) = -\sum_i p(x_i) log p(x_i)$ where $p(x_i)$ is the probability of each discrete value of the gene. Further support to consider only first-order interactors is given by the fact that MI values tend to decrease in terms of both absolute value and significance, the further we move in terms of interactions from our candidate gene.

Then, to remove gene pairs that are statistically independent but have a positive MI value, a null-distribution is computed from a permutation of the gene

expression profiles and, based on the total number of gene pairs, a threshold for MI based on its associated *p-value* to be considered reliable is set (for 10^5 gene pairs, MI values are considered reliable up to a *p-value* of 10^{-4}).

Data Processing Inequality (DPI) is then used to remove the largest part of indirect gene interactions: DPI is based on the principle that if gene1 and gene3 interact only through a third gene, gene2, then $I(g_1, g_3) \leq min[I(g_1, g_2); I(g_2, g_3)]$: ARACNe scans each gene triplet and removes the edge with the smallest mutual information value, in a way that makes it independent from the order in which the triplets are scanned.

ARACNe has also been demonstrated to outperform competing algorithms in regulon reconstruction even in cases where the competing algorithms had access to genome-wide chromatin immunoprecipitation data. A regulon is the ensemble of all the transcriptional targets of any protein that has a regulatory function, namely TFs and coTFs.

Therefore, novel algorithms have employed ARACNe's ability to reconstruct regulons' networks from transcriptomics data to their advantage, as is the case of VIPER (Virtual Inference of Protein activity by Enriched Regulon analysis, Alvarez et al. 2016) an algorithm that aims at inferring the activity of a protein from gene expression data.

The motivation for VIPER is that the main drivers of cellular phenotypes are mediated in the end by protein activity, a feature difficultly measurable due to several cellular processes that include cascades of events, degradation, post-translational modifications, localization and more, and techniques which aim at quantifying protein levels show only a fraction of the picture, along with being time and cost expensive. VIPER tries to solve this by inferring protein activity based on ARACNe-inferred regulons, which accounts for the aforementioned regulatory processes since they are reflected in the gene expression profiles used to infer the network.

VIPER works by testing regulon enrichment in gene expression signatures, which are represented by a statistic (fold change, Student's t, Wilcoxon rank-sum, etc.) describing the comparison of two groups of samples. Based on Spearman correlation score between regulator and target expression, a set of positively regulated targets and one of negatively regulated are defined. aREA is then used to

compute an enrichment score for each regulon, where the score depends on both the absolute value of each gene expression and whether it is a negatively or positively regulated target. Each target gene score is also weighted by the confidence of the regulator-target interaction, defined as

$$IC = \frac{CDF_1}{CDF_1 + CDF_2}$$

Where CDF_1 is the cumulative distribution function of the MI values of the ARACNe-inferred regulon for a regulator, and CDF_2 is the CDF of a null regulon where genes are picked at random. A representation of VIPER's statistical workflow and how regulon's enrichment scores are interpreted is represented below in Fig. 4.3A.

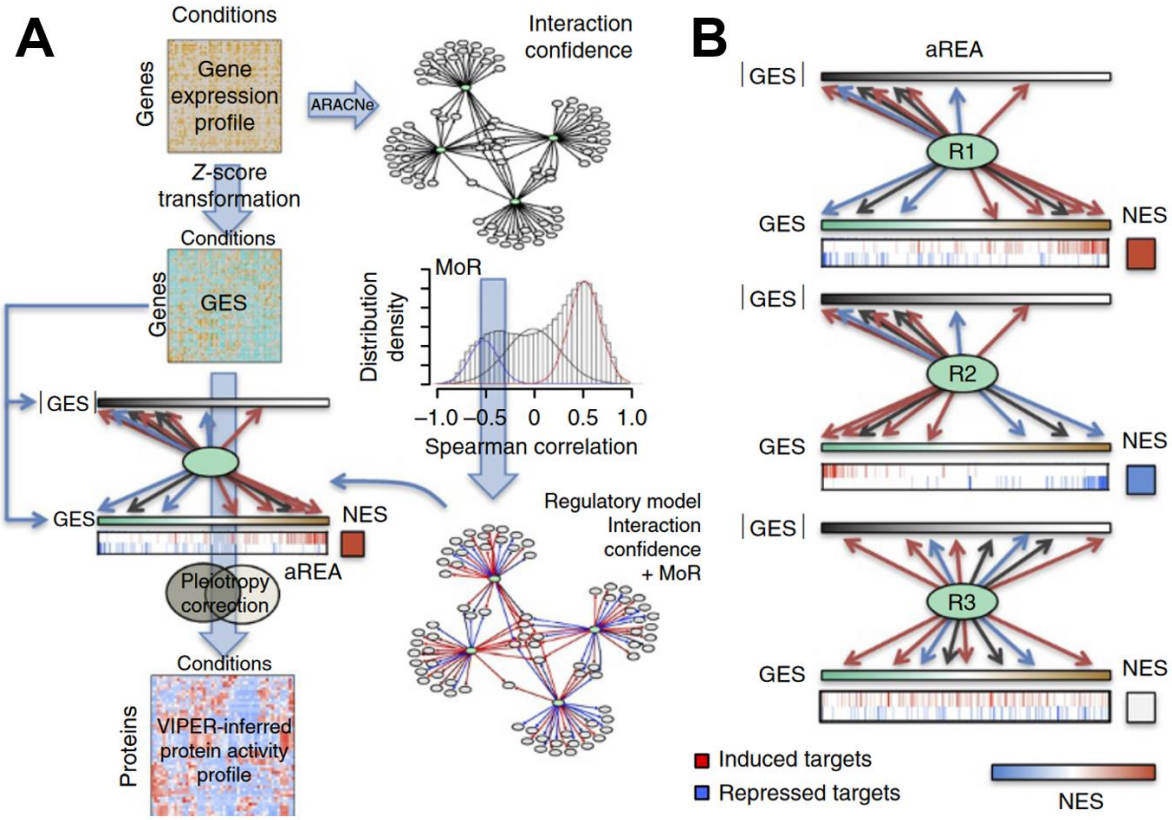

Figure 4.3A – A. VIPER statistical workflow which outputs protein activity values for each tested regulator. B. Interpretation of regulon enrichment: bars represent genes ranked by the statistic of choice from left (bottom ranking) to right (top ranking), if the enrichment is positive for top-ranking genes and negative for bottom-ranking genes the regulon will be positively enriched. The opposite case represents a negatively enriched regulon.

The use of VIPER and ARACNe to infer protein activity data has proven successful in reconstructing altered gene expression networks in cancer (Broyde et al. 2021; Obradovic et al. 2022). Its use has been further extended to study the impact of pharmacological therapies on transcriptional networks to identify changes that could effectively counteract the effect caused by oncoproteins in cancer transcriptomic networks (Alvarez et al. 2018), thus demonstrating the validity of this approach in overcoming the restrictions posed by classical approaches in cancer treatment: due to the positive results in the cancer field, this approach has also been tested in part of our data to test its effectiveness in a context of an ultra-rare disease.

4.4 ATAC-Seq: Assay for Transposase-accessible chromatin

Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) is a method developed to map open chromatin regions across the genome. These regions are enriched in regulatory DNA elements, such as promoters, enhancers, silencers and insulators, which play crucial roles in transcription factor binding and gene regulation (Brasset and Vaury 2005). The core principle of ATAC-seq is the utilization of the Tn5 transposase, a DNA-cutting enzyme that has the dual function of fragmenting DNA and simultaneously inserting sequencing adapters at the cleavage sites, which can then be used for DNA amplification. When introduced to nuclei, the Tn5 transposase tends to target open chromatin regions, facilitating their subsequent sequencing and identification (Luo, Gribskov, and Wang 2022).

The ATAC-seq process begins with the extraction of nuclei from cells, with protocols that maximize the preservation of the DNA's native chromatin state. The isolated nuclei are then subjected to the Tn5 transposase, which inserts sequencing adapters into accessible chromatin regions, which can be located in either open regions or the terminal regions nucleosome-bound DNA (Fig. 4.4A).

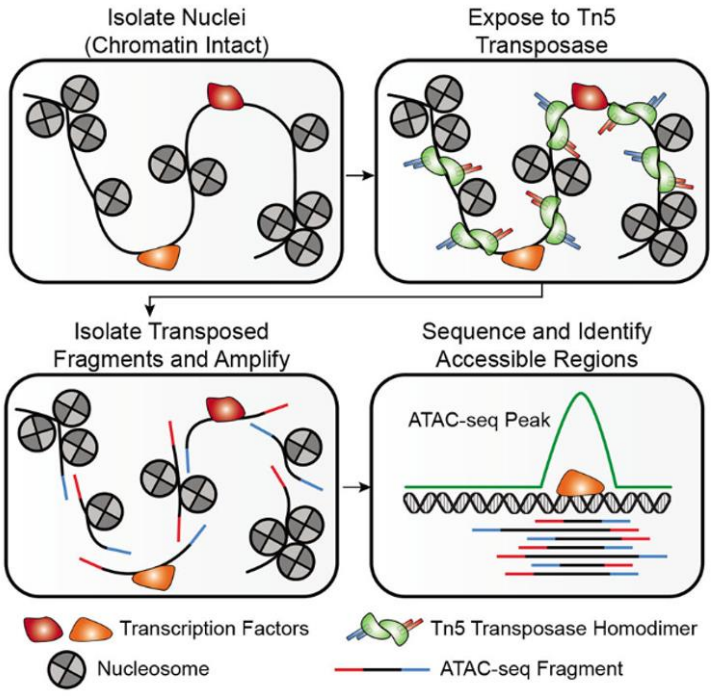

Figure 4.4A – Schematic representation of ATAC-Seq protocol

Following this *tagmentation* step, the DNA is purified and amplified to generate a sequencing library. High-throughput sequencing of this library produces reads that represent open chromatin regions and nucleosome-bound regions (Buenrostro et al. 2015; Grandi et al. 2022).

ATAC-seq offers several advantages over traditional methods of assessing chromatin accessibility, such as DNase-seq, MNase-seq, and FAIRE-seq. Firstly, ATAC-seq requires fewer cells, making it particularly suitable for studies with limited sample availability. Additionally, it offers a more straightforward and rapid protocol, eliminating the need for cross-linking or extensive purification steps. While DNase-seq and MNase-seq rely on enzymes to digest unprotected DNA, ATAC-seq utilizes the Tn5 transposase to directly tag accessible chromatin regions, providing a more direct measure of accessibility (Fig 4.4B)(Tsompana and Buck 2014). Furthermore, the simultaneous insertion of sequencing adapters during the transposition step streamlines the library preparation process.

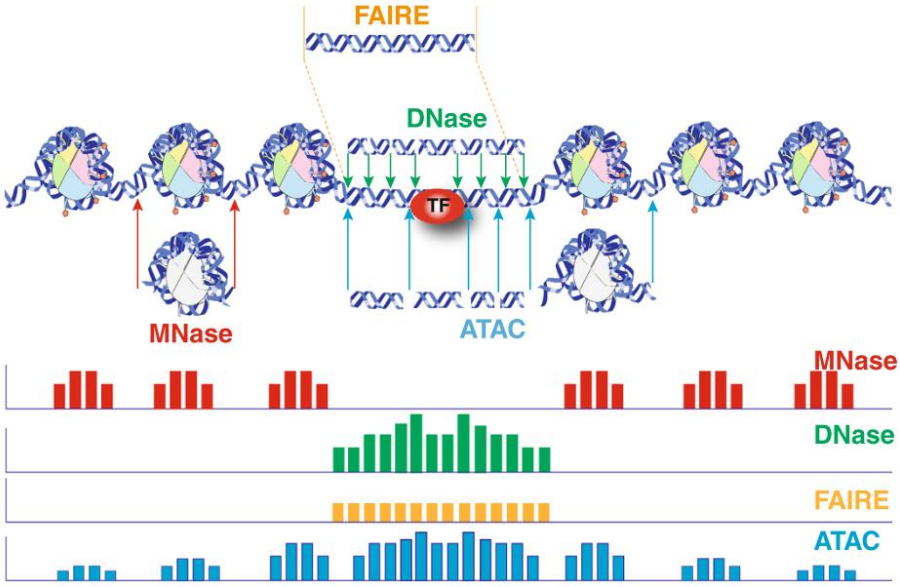

Figure 4.4B – ATAC-Seq schematic representation and comparison to other common techniques used in the study of chromatin organization (from Tsompana and Buck, 2014)

One of the distinguishing features of ATAC-seq is the ability to capture DNA from both open chromatin regions and nucleosome-bound DNA. This is significant as nucleosomes, the basic units of chromatin, can influence gene regulation by modulating the accessibility of DNA to transcription factors and other regulatory

proteins. By capturing nucleosome-bound DNA, ATAC-seq provides a more comprehensive view of the chromatin landscape, offering insights into not only regions that are free of nucleosomes but also those where nucleosomes are positioned.

The resulting data from ATAC-seq provides a genome-wide snapshot of chromatin accessibility, which provides invaluable for understanding the regulatory elements that influence gene expression and the potential roles they play in various biological processes. Furthermore, by integrating ATAC-seq data with other genomic datasets, researchers can gain deeper insights into the complex interplay between chromatin structure, transcription factor binding, and gene expression.

As reported in paragraph 4.1, chromatin remodelling is a key step in the differentiation process of oligodendrocytes. As we saw chromatin state regulates gene expression, first promoting pluripotency genes in embryonic stem cells, which are then turned off in favor of lineage-specific genes (SoxE factors for oligodendrocytes) which lose their repressive methylation marks and become activated (J. Liu and Casaccia 2010). Conservation of epigenetic marks through cell generations affects not only the differentiation process, but also the response to environmental stimuli: plasticity changes (i.e. structural changes of synaptic connectivity) is regulated by a transcription modulation which regulates the brain-derived neurotrophic factor expression, which is regulated by methylation of CpG islands (Martinowich et al. 2003). Long-term modifications in synaptic plasticity are mechanisms associated with learning and memory capabilities and epigenetic marks are used to regulate memory-related gene expression, for example via the widely known CREB1 transcription factor, which recruits CBP to activate downstream genes like C/EBP, required for long-term synaptic modification and to improve synaptic strength (Guan et al. 2002).

In conclusion, ATAC-seq is a powerful tool for studying chromatin accessibility at a genome-wide scale. Its ability to capture both open chromatin and nucleosome-bound DNA provides a comprehensive view of the regulatory landscape, making it an essential technique for researchers studying gene regulation and chromatin dynamics. We employed this technique in our work since it would prove a useful addition to RNA-seq data and, to our knowledge, no other work ever performed

ATAC-seq on an AGC1 deficiency model. Our objective was also to use the results from ATAC-seq and cross with RNA-seq results to gather a deeper understanding of which regulatory mechanisms are altered in AGC1 deficiency.

5 Methods

The code used for the analyses is publicly available on GitHub, in three different repositories:

1. OliNeu RNA-seq analysis: https://github.com/Nbalb/agc1_OliNeu
2. OliNeu ATAC-seq analysis: https://github.com/Nbalb/Agc1_ATAC
3. iPSCs RNA-seq analysis: https://github.com/Nbalb/AGC1_iPSCs

5.1 RNA-Seq pipeline

5.1.1 Salmon

Lightweight alignment tools have emerged as the most precise instruments for quantification of gene expression, encompassing tools like Kallisto, Sailfish, and Salmon, each exhibiting distinct operational peculiarities. In this work we utilized Salmon, the evolution of Sailfish which also has been demonstrated to outperform kallisto in terms of accuracy and inter-replicate concordance, two parameters used to assess mapping performance (Patro et al. 2017).

A shared characteristic among these tools is their ability to bypass the base-to-base alignment of reads, a step that older splice-aware alignment tools, such as STAR and HISAT2, found time-intensive. These modern lightweight aligners deliver quantification estimates at a pace significantly faster than their predecessors (often surpassing them by over 20-fold) and with enhanced accuracy. The output from these tools, commonly termed as 'pseudocounts' or 'abundance estimates', can either be consolidated at the gene level for differential gene expression analysis tools like DESeq2 or can be directly employed for isoform-level differential expression tools.

Salmon operates by taking the reference transcriptome (presented in FASTA format) and the raw sequencing reads (in FASTQ format) to concurrently map and quantify the reads. Salmon's "quasi-mapping" technique necessitates a reference index to check the best mapping location and orientation of the fragments before proceeding to quantification: this reference index essentially reformats the transcriptome into a structure that allows for to quickly identify transcriptomic

positions from which reads originated. Given that the search is conducted against the transcriptome, Salmon is not suited for discovering novel genes or isoforms, intron retention occurrences, or other techniques that require annotations missing from the transcriptome.

The first step to perform Salmon quantification is to create a transcriptome index, which consists of a suffix array (SA) of the reference transcriptome and a hash table to map transcripts in the reference transcriptome to the location in the SA.

The quasi-mapping procedure (Srivastava et al. 2016) is performed by (Fig. 5.1A):

1. An initial scan of the read from its beginning to its end, searching for a k-mer that matches an entry in the hash table.
2. Upon locating the k-mer in the hash table, the SA intervals are fetched, revealing all suffixes encompassing that specific k-mer.
3. Maximal matching prefix (MMP) is defined by identifying the longest read sequence that perfectly matches with the reference suffixes.
4. Advance by one k-mer to the next informative position (NIP).
5. This methodology is reiterated until the read's end.
6. The final mappings are derived by pinpointing the transcript set present in all the MMPs for a specific read. The output reports transcripts, orientation and transcript location.

Instead of counting the number of reads mapping to each transcript, Salmon exploits diverse modeling techniques, such as Expectation Maximization (EM), to estimate transcript abundances and to correct sample specific biases, such as: GC content bias, biases related to positional coverage, sequence discrepancies at both 5' and 3' fragment termini, distribution of fragment lengths, strand-specific techniques.

Overlooking these biases substantially increases false positive rates in downstream differential expression analyses.

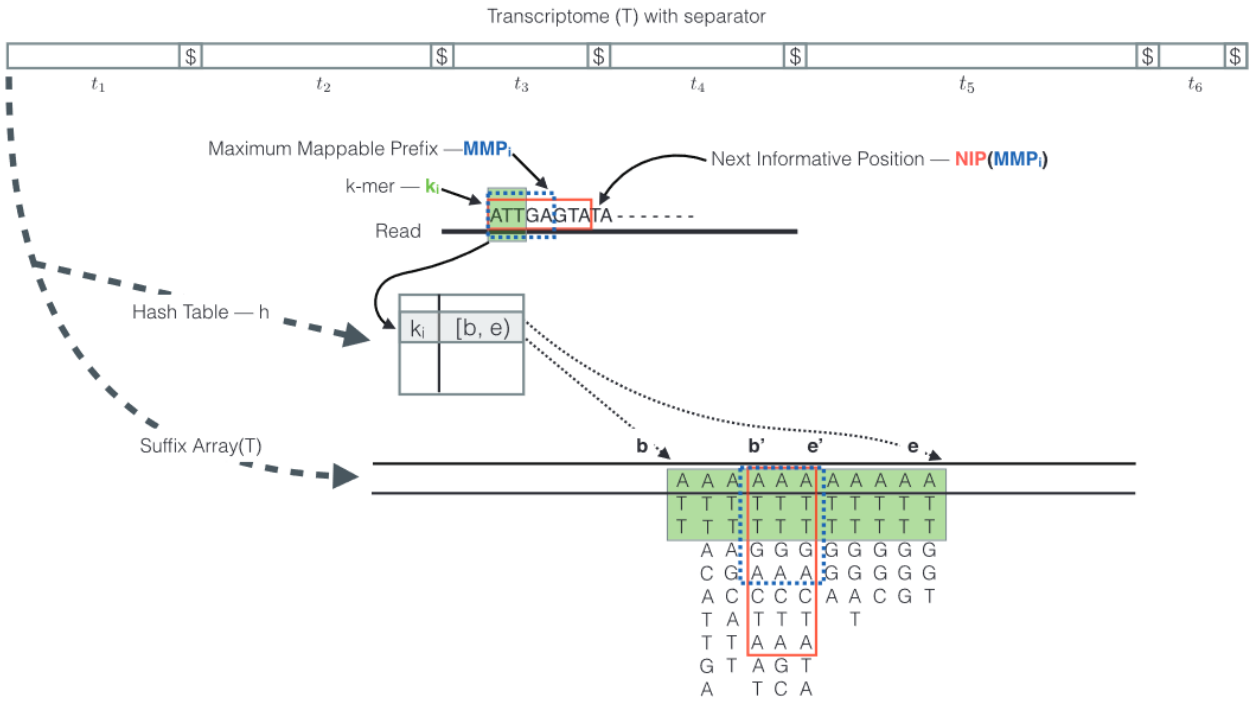

Figure 5.1A – Quasi-mapping procedure schematic representation (from Srivastava 2016)

5.1.2 DESeq2

DESeq2 (Love, Huber, and Anders 2014) is an R package that provides methods to test for differential gene expression by using negative binomial generalized linear models. It's tailored for high-throughput sequencing data, especially RNA-seq, but can also be applied to other count data and to abundance estimates, like Salmon's output.

The first step in DESeq2 is count normalization, to account for sequencing depth (total number of reads in a sample), gene length and RNA composition (A few highly differentially expressed genes between samples can “hog” most of the resources in the amplification step during sequencing, resulting in other genes that are erroneously underrepresented). Compared to other normalization methods (Counts/Transcripts per Million or reads/fragments per kilobase of exon per million reads/fragments) DESeq2 uses median of ratios which consists of:

1. Compute the geometric mean of raw counts across all samples to obtain a pseudo-reference and remove 0 counts (i.e. not expressed) genes
2. Compute the ratio between the raw counts in each sample and the pseudo-reference for each sample. The normalization factor for gene i in sample j is s_{ij} and represents the sample-specific median value of ratios
3. Raw counts are then normalized for each respective sample’s normalization factor

The second requires estimating within-group variability, which is represented by the dispersion parameter α used to describe the variance of counts as

$$Var = \mu + \alpha^2\mu$$

which returns reliable results when groups have high sample counts, which is not always the case. Therefore, information across all samples is shared for each gene separately and gene-wise dispersion estimates are computed using maximum likelihood estimation and a smooth curve is fit to the data to provide an estimate for the dispersion values. Gene-wise dispersion estimates are then shrunk towards the curve to obtain final dispersion values in an amount that depends on sample size, unless the final dispersion estimate is more than 2 standard deviations from the curve, in which case it does not get shrunk: this helps in reducing potential false positives while preserving dispersion of genes with high dispersion levels that do not follow model assumptions (Fig 5.1B).

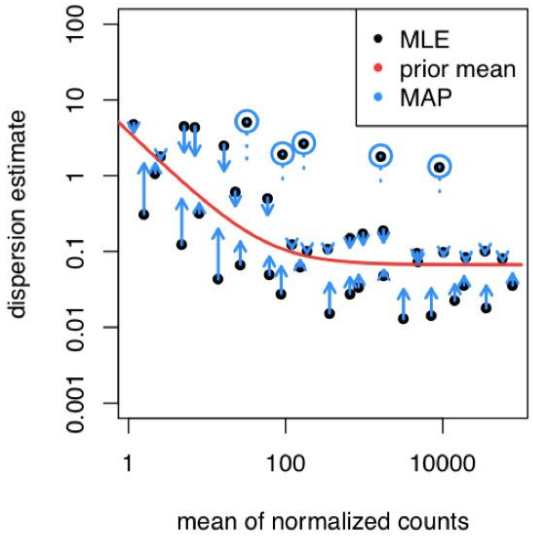

Figure 5.1B – DESeq2 dispersion estimation: points in black are the results of MLE of gene dispersion, and in red is the fitted curve. Estimates are then shrunk toward the fitted curve, unless they are considered outliers (> 2sd, blue circles). Adapted from Love, Huber and Anders 2014.

Log fold changes (LFCs) are then estimated by modelling normalized raw count data with a negative binomial model, chosen over other models since it allows to better represent distribution of RNA-seq count data that usually present variance > mean. Counts are modelled as

$$K_{ij} \sim NB(\mu_{ij}, \alpha_i)$$

$$\mu_{ij} = s_{ij} q_{ij}$$

Where K_{ij} is the raw counts for gene i of sample j, s_{ij} is the normalization factor previously computed, q_{ij} is a quantity proportional to the concentration of fragments from gene i of sample j and α_i is the dispersion estimate previously estimated. After model fit and coefficients estimation for each group, we can estimate LFCs with

$$log q_{ij} = \sum x_j \beta_i$$

The coefficients β_i give the LFCs for gene i for each column of the model matrix X. These estimates do not account for large dispersion when in presence of a

low read count. LFCs are therefore shrunken and in our specific case we used *apeglm* (Zhu, Ibrahim, and Love 2019), which uses a Bayesian shrinkage estimator with a heavy-tailed prior distribution on the effect sizes, where the shape of the prior distribution is fixed and the scale is adapted to the distribution of observed MLE of effect size for all genes.

Hypothesis test is then performed to determine significance associated with an estimated LFC using Wald test, which compares β_i divided by its standard error to a standard normal distribution, which outputs a *p-value* that is then adjusted for multiple hypotheses testing using Benjamini-Hochberg procedure.

5.1.3 GSEA

Gene Set Enrichment Analysis (GSEA) is a methodological approach developed to interpret genome-wide RNA expression data. Traditional methods often focus on individual genes, especially those with the most significant differential expression. However, such an approach can be limiting due to:

1. The potential absence of statistically significant genes after adjusting for multiple hypothesis testing.
2. The generation of a long list of statistically significant genes without a coherent biological theme.
3. The possibility of overlooking effects on pathways, as cellular processes often involve sets of genes acting in concert.
4. The inconsistency in the list of statistically significant genes when different groups study the same biological system.

To address these challenges, GSEA evaluates expression data at the level of gene sets, which are defined based on prior biological knowledge, such as biochemical pathways or coexpression from previous experiments. The primary objective of GSEA is to determine if members of a gene set tend to appear towards the top or bottom of a ranked list, indicating a correlation with the phenotypic class

distinction. An enrichment score (ES) is calculated for each tested gene set, as the maximum deviation of $P_{hit} - P_{miss}$

$$P_{\text{hit}}(S, i) = \sum_{\substack{g_j \in S \\ j \le i}} \frac{|r_j|^p}{N_R}, \quad \text{where } N_R = \sum_{g_j \in S} |r_j|^p$$

$$P_{\text{miss}}(S, i) = \sum_{\substack{g_j \notin S \\ j \le i}} \frac{1}{(N - N_H)}.$$

Genes are ranked on the basis of a score r_j (differential expression, correlation with a condition, …) in a list $L = \{g_1, \dots, g_N\}$, P_{hit} is the fraction of genes of L in gene set S and P_{miss} is the fraction of genes of L not in gene set S. p is a weighting factor that when set to 0 reduces the ES to the Kolmogorov-Smirnov test statistic, while when set to 1 the genes in S are weighted by the r_j score they were assigned normalized by the sum of the correlations over all the genes in S (which is the common use for GSEA and what we use throughout our work). Significance is estimated by permutation 10000 times the r_j scores obtaining a distribution of ES_{NULL} used to compute a p-value associated to S. NES is then obtained by correcting ES for gene set size and to account for False Discovery Rate using a fixed level of significance.

5.1.4 Majiq

Modeling Alternative Junction Inclusion Quantification (MAJIQ) (Vaquero-Garcia et al. 2016; 2021) is a computational tool designed to analyze variations in RNA splicing. Traditional transcriptome variations have been studied at the level of full gene isoforms or through the specification of alternative splicing 'events'. These events have been categorized into several common types, such as intron retention, alternative 3'/5' splice sites, or cassette exons. However, these predefined AS types often fail to capture the full complexity of spliceosome decisions, which can involve more than two alternative junctions. MAJIQ introduces the concept of local splicing

variations (LSVs) to address this complexity. LSVs capture previously defined types of alternative splicing as well as more intricate transcript variations. The primary objective of MAJIQ is to determine if members of a gene set tend to appear towards the top or bottom of a ranked list, indicating a correlation with the phenotypic class distinction. The tool employs a unique approach that focuses on LSVs, enabling the capture of both traditional AS events and more complex transcript variations.

MAJIQ quantifies the fraction each LSC junction is selected, determining the Percent Selected Index (*PSI* or Ψ) or the changes between each junction's *PSI* in different conditions (*ΔPSI* or $\Delta\Psi$), estimated as posterior distributions over possible fractions in range [0, 1] for *PSI* and [-1, 1] for *ΔPSI*. There are then … steps to estimate *PSI*:

1. Estimate the read rate per position in each junction which includes a global parametric model (a zero truncated negative binomial model) for read count variability, a procedure for stack removal (a read stack is intended as a position i in junction j with an observed $x_{i,j}$ which is unexpected given the dispersion parameter of the negative binomial model and the average read rate in other positions of the junction) and a bootstrapping for positions across each junction to estimate read rate to normalize for local dispersions

2. MAJIQ uses a Beta distribution to model the prior:

$$P_0(\Psi_{\mathrm{e,j}}) \sim Beta\left(\alpha = \frac{1}{J}\eta, \beta = \frac{J-1}{J}\eta\right)$$

   While for *ΔPSI* uses

$$P_0(\Delta\Psi) = \sum_{K=1}^{K} P(k) Beta\left(\Delta\Psi \middle| \alpha(k), \beta(k)\right)$$

   With k set to 3

3. The posterior is then computed given the estimated number of reads per junction in each sample

5.1.5 Corto

Corto (Mercatelli, Lopez-Garcia, and Giorgi 2020) is an R package that allows for a lightweight implementation of ARACNe and VIPER that can run in reasonable times on local machines, overcoming the issues of necessitating HPC infrastructure with large amount of RAM to execute to pipeline.

Corto was built by improving by 50% speed, efficiency and accuracy of ARACNe-AP (Lachmann et al. 2016), which itself is a more efficient implementation of the original ARACNe algorithm able also to cover uneven probability density spaces, built around the use of adaptive partitioning for MI estimates. ARACNe-AP results in being 200x faster, 2x more memory-efficient compared to the original ARACNe implementation and allows for multithreading. Networks inferred by ARACNe-AP are also virtually identical to those inferred by ARACNe, in terms of both estimated MI and network topology. Unlike ARACNe-AP, which bases gene-gene relationships on the inference of Mutual Information, corto uses Pearson's correlation.

Copy number variations (CNV) are changes in the number of copies of one or more sections of the DNA that can encompass both deletions (losses) and duplications (gains) of DNA regions. These variations can be as small as a single nucleotide or as large as an entire chromosome and can have strong regulatory effect when a regulator is affected. Therefore, corto introduces CNV correction via a linear model to offset the impact of CNVs in network reconstruction.

For the current analysis, the signature used to evaluate MRs activation in our model was obtained by the formula:

$$signature = log2FoldChange * -log_{10}padj$$

After computing the networks, this signature is used to determine the model-specific activation levels, used to determine which MR is activated based on each target signature score.

5.1.6 Shiny and Docker for Interactive Web Application Frameworks

Shiny, a product from RStudio, is a groundbreaking framework tailored for R, a language renowned for statistical computing and graphics. It enables R developers

to transform static data visualizations and analyses into interactive web applications. Without necessitating proficiency in web development languages like HTML, CSS, or JavaScript, Shiny allows data scientists to craft user-centric applications, making complex data analyses accessible and understandable to a broader audience.

Central to Shiny is its integration of reactive programming with R's analytical capabilities. This allows for automatic updates of outputs based on user input, eliminating the need for page reloads. A Shiny application comprises two main components: the user interface (UI) and the server function. The UI, crafted using R, defines the application's layout, encompassing elements like sliders and plots. Conversely, the server function manages the interplay between inputs and outputs, epitomizing the reactive model.

Docker, on the other hand, changed the realm of software deployment: at its core, Docker is a platform designed for containerization—a technology that packages an application and its dependencies into a 'container'. This ensures that the software runs consistently across various computing environments, be it a developer's local machine or a production server in the cloud. By abstracting away, the underlying system specifics, Docker alleviates the common software leitmotiv of "it works on my machine", ensuring uniformity and reliability in software deployment.

Shiny's strength lies in its ability to transform intricate R scripts into user-friendly, interactive web interfaces. However, the deployment of Shiny applications can be fraught with challenges. Differing library versions, system dependencies, or OS-specific quirks can hinder consistent deployment across varied environments. This is where Docker's utility becomes evident. By encapsulating a Shiny application within a Docker container, one ensures that the application, along with its dependencies and environment specifics, is packaged into a singular, cohesive unit. This container can then be executed consistently across any platform supporting Docker, be it a local machine, a cloud server, or a complex cluster.

The process begins with defining a Dockerfile, a script detailing the environment setup, including the base OS, R version, required libraries, and the Shiny application itself. Once crafted, this Dockerfile is used to build a Docker image, a static snapshot of the application and its environment. This image serves as a blueprint, from which Docker containers—live, running instances of the application—

can be instantiated. One of Docker's salient features is its layering mechanism. Each step in the Dockerfile creates a layer, and Docker caches these layers. This means that frequent updates to a Shiny application, perhaps to incorporate new data or refine visualizations, won't necessitate a complete rebuild of the Docker image. Only modified layers are rebuilt, ensuring efficient and rapid deployment cycles.

5.2 ATAC-Seq pipeline

5.2.1 MACS

In the domain of ChIP-Seq data analysis, understanding the precise protein-DNA interaction sites is fundamental. MACS, or Model-based Analysis of ChIP-Seq, can effectively tackle this challenge, particularly in modeling the shift size and detecting peaks. Due to the similarities in the data produced by ChIP-seq and ATAC-seq experiment, MACS is commonly used also for ATAC-seq data even if in the original paper the focus was solely on ChIP-seq (Zhang et al. 2008).

Counts data for MACS were obtained using *featureCounts* (Liao, Smyth, and Shi 2014), part of the *Subread* package which allows for fast read counting using an alignment file and a genome annotation file; *featureCounts* also returns useful information other than counts, namely gene ids and genomic positions.

One of the first issues MACS addresses is to achieve a good tag to site distance estimate, since ChIP-Seq tags represent the ends of the ChIP-DNA fragments, not the exact protein-DNA binding sites but they are often shifted towards the 3’ end of the strand. Estimating the shift size helps in approximating the center of the DNA fragments, which is closer to the actual binding sites of proteins on the DNA: by shifting the tags towards the center, one can achieve a higher resolution in pinpointing these binding sites.

To determine the shift size, MACS employs a dynamic strategy: given a sonication size (*bandwidth*) and a high-confidence fold-enrichment (*mfold*), MACS scans the genome with a window of 2 times the *bandwidth*, identifying tagged regions that are > *mfold* enriched relative to a random tag genome distribution. From 1,000 randomly sampled high-quality peaks, MACS separates the Watson and Crick tags (Watson strand tags are enriched upstream of binding site, Crick’s downstream) and aligns them by the midpoint between their Watson and Crick tag centers. The distance (d), between the modes of the Watson and Crick peaks is computed and all tags are shifted by $\frac{d}{2}$ towards the 3' ends, therefore determining the most probable protein-DNA interaction sites.

For peak detection, especially in experiments with a control, MACS scales the total control tag count linearly to match the total ChIP tag count. Recognizing that the

same tag can sometimes be sequenced repeatedly due to biases during ChIP-DNA amplification and sequencing library preparation, MACS removes duplicate tags that exceed what's warranted by a binomial model of sequencing depth. With the current genome coverage of most ChIP-Seq experiments, the tag distribution along the genome can be modeled by a Poisson distribution: this model's advantage is its ability to capture both the mean and variance of the distribution using a single parameter (λ_{BG}). After shifting every tag by $\frac{d}{2}$, MACS slides $2d$ windows across the genome, identifying candidate peaks with significant tag enrichment. Overlapping enriched peaks are merged, and each tag position is extended by d bases from its center. To account for local fluctuations in tag distribution and biases, λ is defined for each peak as

$$\lambda_{local} = max(\lambda_{BG}, [\lambda_{1K},]\lambda_{5K}, \lambda_{10K})$$

Where the difference between the parameters is determined by the kilobases (1, 5, 10) of the window centered at the peak position in the control sample. This parameter is then used to compute a *p-value* for each peak, determining which are called (meaning they pass a user defined threshold), based also on an empirical estimation of FDR established as number of control peaks/number of ChIP peaks.

5.2.2 diffTF

Traditional analysis of ATAC-Seq data, reliant on differential peak calling, often falls short in discerning subtle changes in TF binding which may result in missing key details in understanding transcriptional regulation. Moreover, the noise associated with TF binding site predictions often leads to an increased number of false-positive predictions when compared with ChIP-seq experiments, necessitating a more refined approach. diffTF is a computational method that overcomes these issues by estimating TF activity based on the chromatin state of TF's targets captured via ATAC-seq (Berest et al. 2019).

TF activity is defined as the effect of a transcription factor on the state of chromatin or active chromatin marks, meaning that TF activity has a causal role in determining chromatin accessibility, therefore changes in chromatin accessibility in TF binding sites (TFBS) can be used to infer TF capability to affect chromatin state.

This brings to the assumption that an increase in activator TF's activity will increase its binding site accessibility, while a repressor TF will decrease accessibility on its binding sites. ATAC-seq data is used to compute differential TF activity based on the differential accessibility of the target sites in 2 separate conditions, while RNA-seq data is used to estimate whether a TF is an activator or a repressor.

diffTF first generates a consensus peak set across samples (i.e. a collection of genomic regions that are consistent across samples) from aligned peaks data. Then a set of TFBSs is generated based on the TFs of interest by using data user-provided or publicly available (we used HOCOMOCO database (Kulakovskiy et al. 2018), which is also default in diffTF). Counts for the consensus peaks set are obtained used *featureCounts* and *DESeq2* for count normalization using a cyclic loess approach which also considers the user-supplied design formulas. TFBS are then filtered on the overlap with the consensus peak set (multiple TFBS per peak are allowed), then *limma* (Ritchie et al. 2015) functions *lmFit and eBayes* are used independently for each TF, using the previously provided design formula. GC-content for the local environment of each TF is computed via the *bedtools* package, creating 10 bins based on total GC-content; differential TF activity is then computed by comparing the distribution of TF-specific log2FoldChanges vs the log2FoldChanges of all other TFBS from TFs in the same GC bin.

TFs are then classified into activators and repressors by computing the Pearson correlation coefficient between the expression level of each TF and the ATAC-seq signal of each TFBS. Correlation medians are then computed across foreground (peaks in TFBS of TF-specific targets) and background (all other peaks). Background correlation medians are used to estimate the noise level for median correlations. A *p-value* for this step is also computed using a Wilcoxon rank sum test. Significant positive correlations classified TFs into activators, while negative correlations classify them into repressors.

6 Results

6.1 RNA-Seq OliNeu cells

6.1.1 Experimental setup

A detailed study on AGC1 deficiency was conducted through transcriptomic analysis, starting from pseudo-alignment performed using Salmon (results reported in table 6.1A).

Sample	Total fragments observed	Mapping rate (%)
Control rep 1	39361509	83.3841
Control rep 2	41912812	82.6381
Control rep 3	30297972	81.5560
siAgc1 rep 1	43464546	83.8745
siAgc1 rep 2	32785369	82.3057
siAgc1 rep 3	41443682	83.8530

Table 6.1A – Observed fragments and mapping rate from Salmon pseudo-alignment procedure

Differential gene expression was observed by comparing three sets of Oli-Neu cells with partial AGC1 silencing via short-interfering RNA to three sets of control Oli-Neu cells. Western Blot tests indicated a reduction in AGC1 protein levels in the modified Oli-Neu cells compared to the control cells. (Fig. 6.1A).

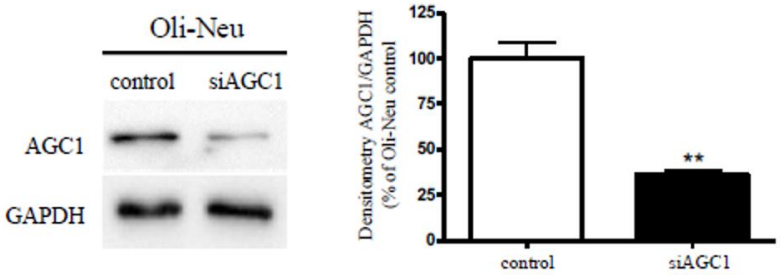

Figure 6.1A – Protein levels of Agc1 in control and siAgc1 samples

The effect of AGC1 silencing is clearly visible also through dimensionality reduction of the dataset via Principal Component Analysis (Fig. 6.1B).

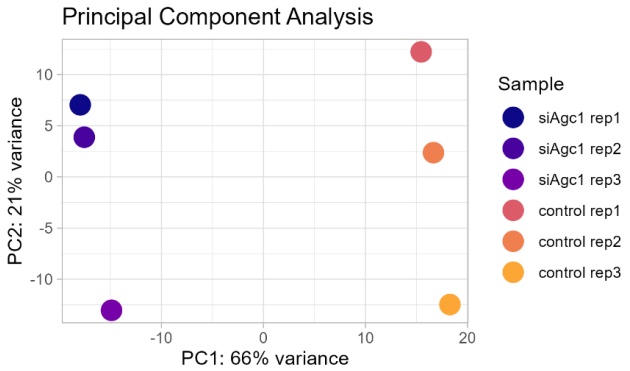

Figure 6.1B – Control and siAgc1 samples PCA

A clear separation between control and siAGC1 samples is observed on the first principal component after dimensionality reduction, indicating that the most relevant feature that separates samples is their genotype. It's harder to state what is driving the differences along the second PC, but large part of the effect might be explained by the technical replicate number.

The distance matrix of the samples also indicates that samples with the same genotype tend to have lower distances compared to samples with the other genotype, which brings them to cluster together when using hierarchical clustering (Fig. 6.1C).

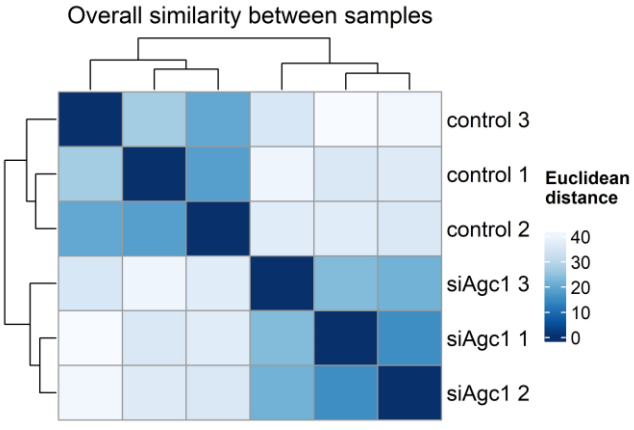

Figure 6.1C – Distance matrix of control and siAgc1 samples

The data reported here certify an overall good quality of the samples which are key to reliable results from the subsequent analyses, which cover several aspects of transcriptomics.

6.1.2 Differential Gene Expression and GSEA

To assess the degree of inhibition of Agc1 at the RNA level, we compared the quantified levels of both Agc1 and Agc2 transcripts (transcripts per million - TPMs) in siAGC1 and control samples and observed a similar degree of downregulation for both genes: mean TPMs for Agc1 were 26.9 for control and 20.3 for siAGC1, with a significant Log2FoldChange (LFC) of -0.39 (adjusted P = 0.039). The mean TPMs for Agc2 were 1.06 for control and 0.429 for siAGC1, with a significant LFC of -0.97 (adjusted P = 0.027) (Fig. 6.1D).

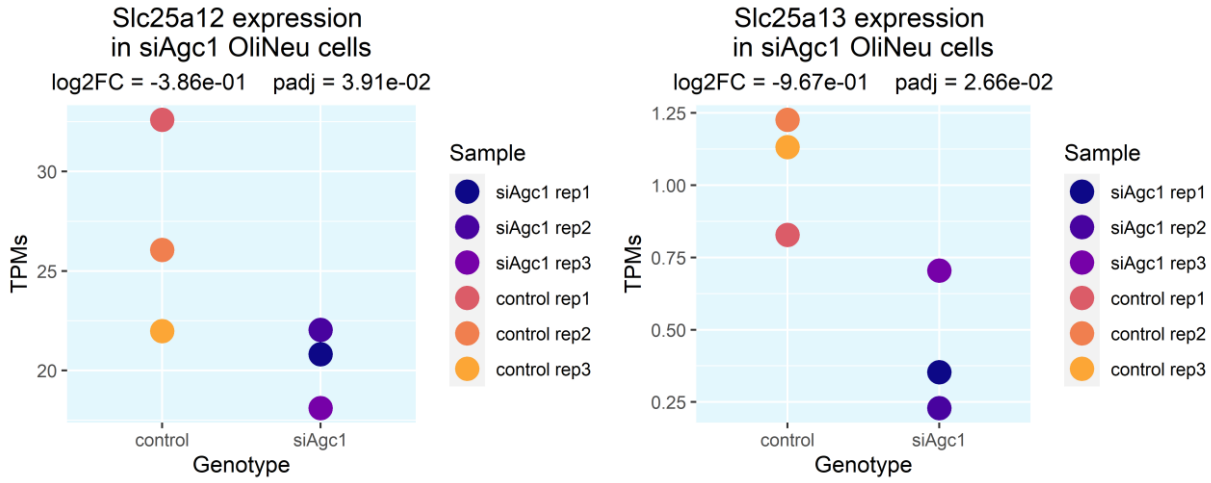

Figure 6.1D – TPM values for Slc25a12 and Slc25a13 in control and siAgc1 OliNeu cells

Nonetheless, the Western blot examination of AGC2 protein levels yielded a contrasting outcome. In silenced cells, there was an observed increase in the expression of the second isoform of the carrier when compared to control cells.

This observation hints at a potential compensatory mechanism triggered in response to AGC1 silencing, which may provide insight into why mitochondrial

activity remains unaffected in siAgc1 Oli-Neu cells (Petralla et al. 2019). In the context of estimating Log-Fold Changes (LFCs), it is well-recognized that genes with limited sample counts or high coefficients of variation can pose challenges, particularly in experimental scenarios with a limited number of replicates. This often results in LFC values that do not accurately mirror the true expression levels within the biological model being investigated (Love, Huber, and Anders 2014). Various strategies have been proposed to alleviate this issue, and we opted to implement the apeglm method (Zhu, Ibrahim, and Love 2019) due to its numerous advantages. These include the absence of the need for fixed thresholds to filter genes with low counts, the incorporation of pseudocounts, and its straightforward integration into the analysis pipeline. By visualizing the LFC against each gene's normalized counts (MA plot), it is evident how the use of this shrinkage procedure significantly lowers the LFC of genes with lower counts, while no impact is recorded for genes with a higher gene count (Fig. 6.1E).

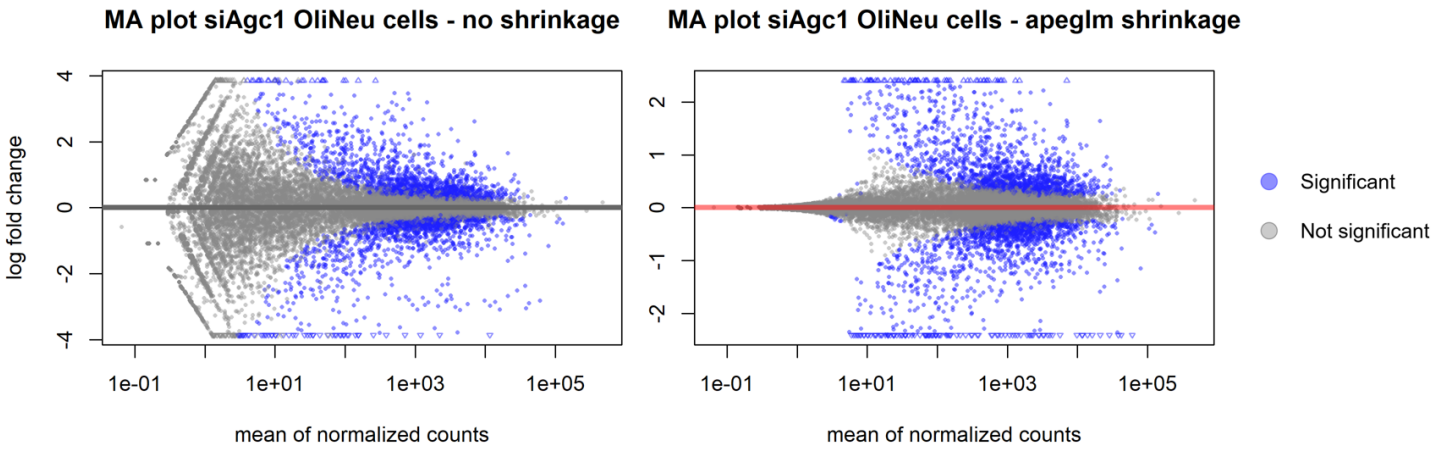

Figure 6.1E – MA plot before (left) and after (right) applying apeglm shrinkage to modelled counts.

Statistical analysis of differential expression allowed us to identify 769 genes exhibiting significant changes in siAgc1, categorized as 356 downregulated genes and 413 upregulated genes. Notably, among the most significant alterations, there is a preponderance of downregulated genes. For instance, Chodl, which encodes the transmembrane protein Chondrolectin (a finding corroborated via real-time PCR), and Mrpl55, responsible for a mitochondrial ribosomal protein, are among the

prominently downregulated genes. Conversely, among the genes displaying the most pronounced upregulation induced by siAgc1, several encode cell adhesion molecules, including Coro2a, Itga5, Bgn, and Fat3. This observation suggests potential alterations in both cell-cell adhesion and interactions with the extracellular matrix (Fig. 6.1F).

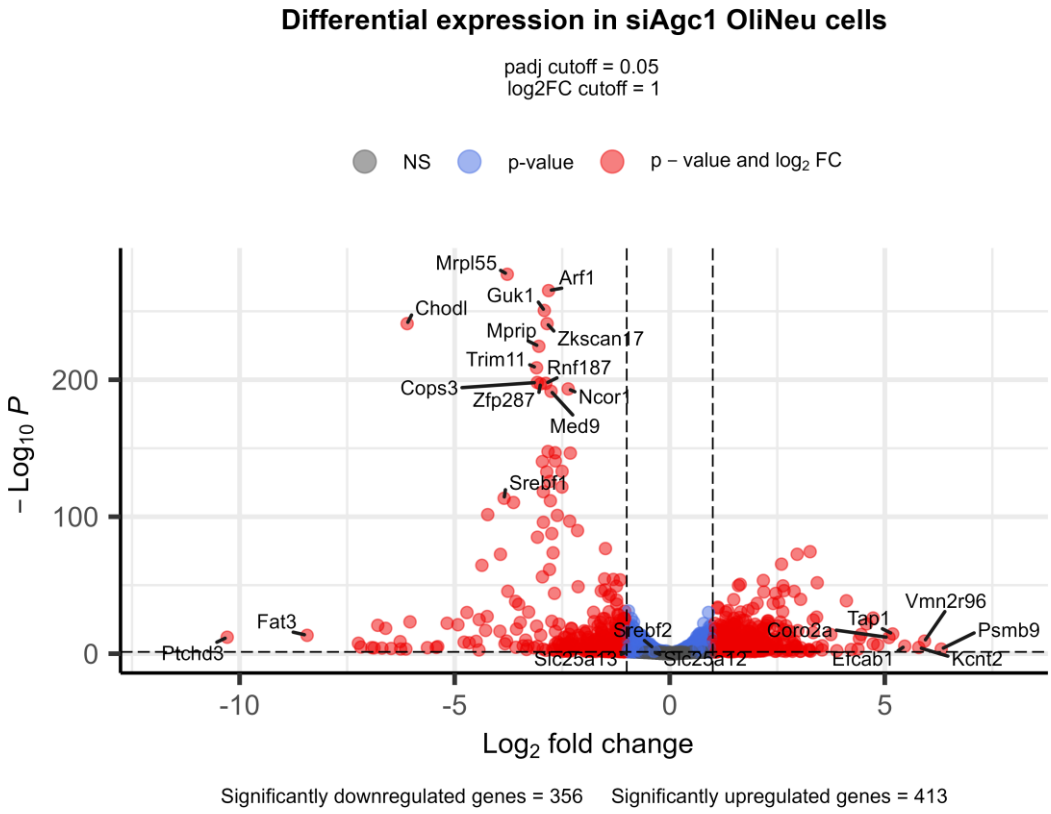

Figure 6.1F – Volcano plot of siAgc1 vs control OliNeu

To more succinctly discern the significantly altered cellular pathways, we carried out a pathway enrichment analysis using the Molecular Signature Database (MSigDb). This analysis specifically focused on pathways within the Hallmark, Curated, and Ontology gene sets. Subsequently, we refined the results by merging similar pathways to address redundancy, particularly within the Curated and Ontology gene sets. The resulting NES (Normalized Enrichment Score) for pathways found to be significantly enriched are reported in Fig. 6.1G. NES quantifies the enrichment of a predefined gene set in a ranked list of genes, with normalization ensuring comparability across datasets by standardizing NES values to a mean of

zero and a standard deviation of one across permutations of the gene expression data.

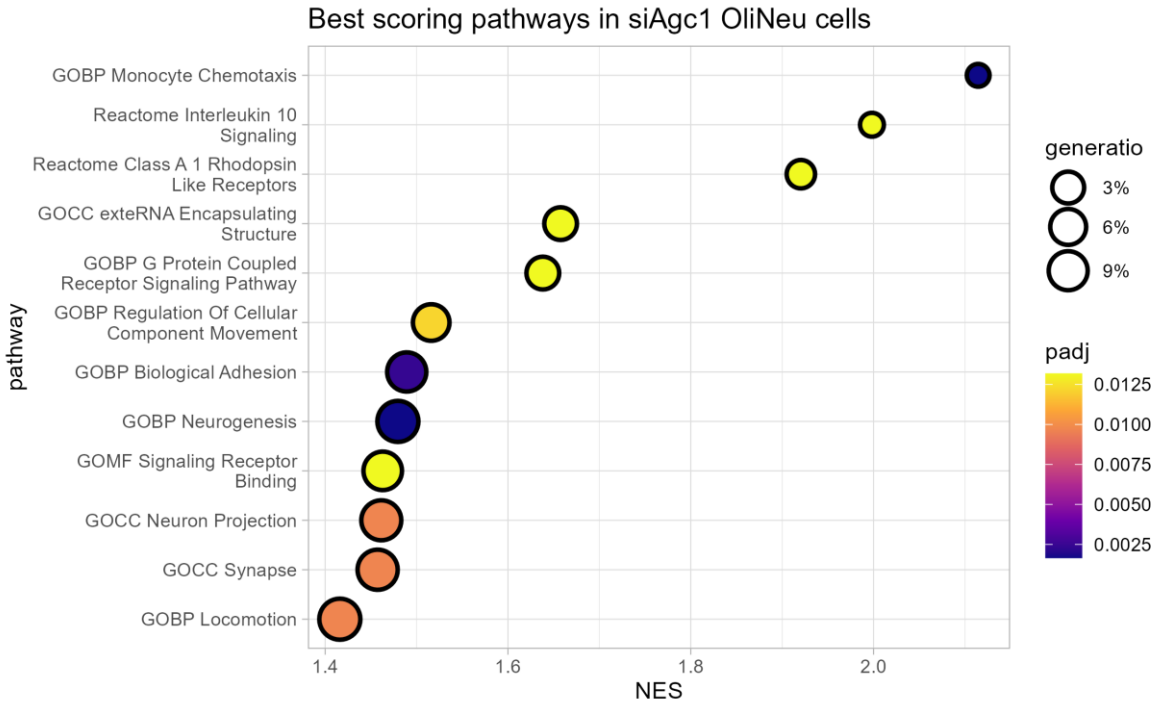

Figure 6.1G – NES values for GSEA on siAgc1 vs control OliNeu cells

This analysis underscores the profound impact of AGC1 downregulation on the cell's capacity to interact with other cells and the extracellular matrix (see Fig. 6.1G). Within siAGC1 samples, several pathways, including those related to biological adhesion and cellular motion (locomotion), exhibit notable upregulation. Moreover, we observe an upregulation in the neurogenesis pathway within siAGC1 samples, aligning with recent investigations conducted on neurospheres derived from heterozygous mice (Petralla et al. 2019). These studies reveal a decrease in the OPCs (oligodendrocyte progenitor cells) population, accompanied by an increase in mature OLs (oligodendrocytes), neuronal progenitors, and astrocytes.

To observe changes in genes associated with myelination between silenced and control samples, we focused on genes within the Gene Ontology myelin-related pathway that showed significance in differential expression analysis (Fig. 6.1H).

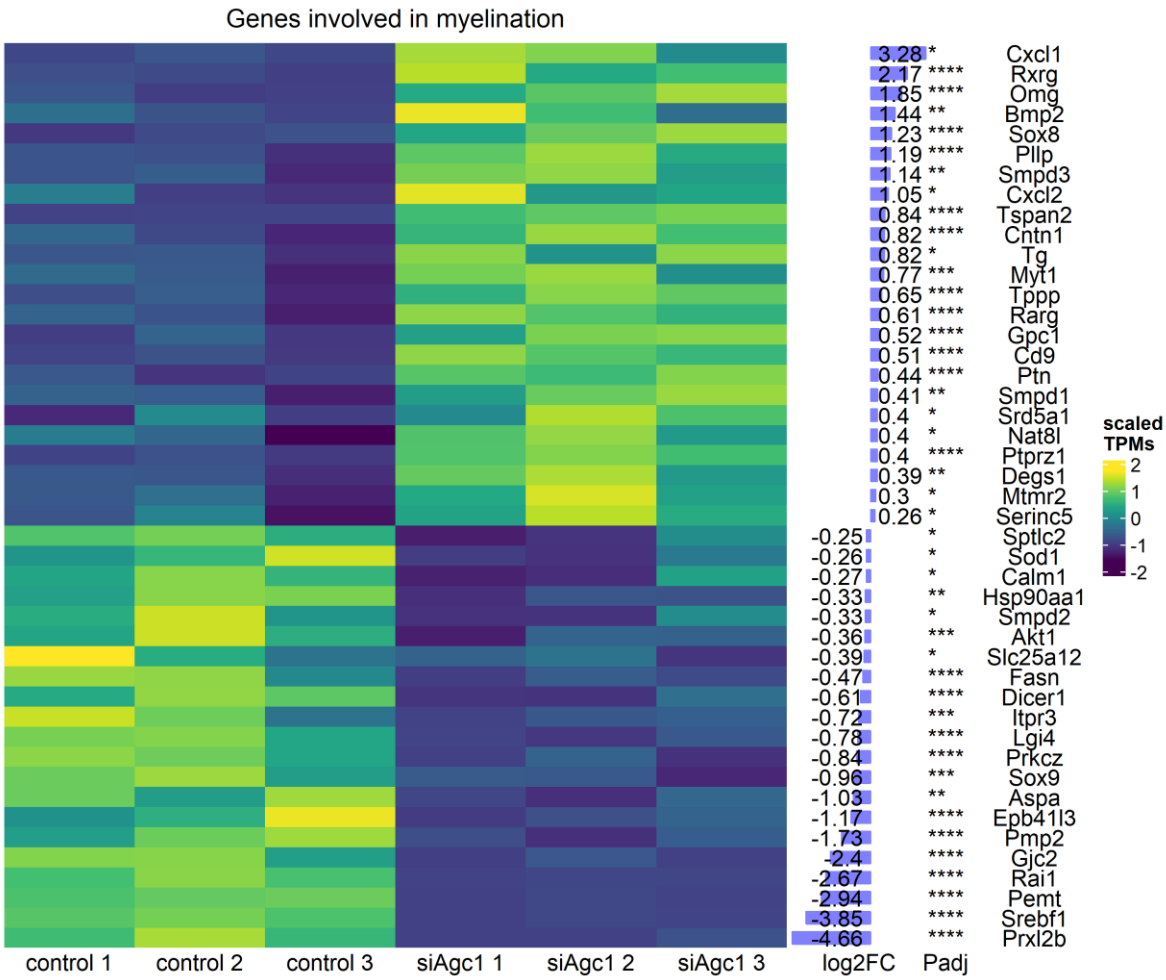

Figure 6.1H – TPMs heatmap for genes involved in myelination.

In this analysis, we found that Fasn (Fatty Acids Synthase N) is downregulated. Fasn controls fatty acid synthesis by producing palmitate from acetyl-CoA and malonyl-CoA. Since it also influences myelination and re-myelination processes, its depletion can lead to aberrant myelin production and the instability of myelinated axons (Dimas, Montani, and Pereira 2019). This finding confirms the hypomyelination observed in AGC1 deficiency patients. Moreover, the downregulation of Srebp1 (Sterol Regulatory Element Binding Protein 1), a master regulator of fatty acid synthesis, further contributes to reduced fatty acid synthesis and altered differentiation processes (Monnerie et al. 2017). This downregulation

might explain the aberrant myelination observed in AGC1 deficiency patients (Wibom et al., 2009) and the previously observed premature differentiation of siAGC1 Oli-neu cells (Petralla et al. 2019). Notably, Akt1 is also downregulated, hinting at its potential role in SREBP1 altered maturation and downregulation, as it controls the translation of many genes involved in lipogenesis (Porstmann et al. 2005).

In contrast, Ptn and Ptprz1 show upregulation in siAGC1 samples. Given their roles in regulating OPCs differentiation (Kuboyama et al. 2016), these alterations could account for the premature differentiation of silenced cells, which subsequently impedes proper myelin production. This premature differentiation appears to be a distinctive characteristic of these silenced cells, as indicated by the upregulation of markers for mature OLs in RNA-Seq analysis, such as Myt1 (Myelin transcription factor 1) and Omg (Oligodendrocyte myelin glycoprotein). Both of these markers influence the precise temporal progression of OPCs proliferation and differentiation (Nielsen et al. 2004; Lee et al. 2011).

Furthermore, there is notable upregulation in the gene expression of Retinoic Acid (RA) receptors (RARG and RARB) in silenced samples (Fig 6.1H). RA is known to be involved in the transition from neural stem cells (NSCs) to OPCs (S. young Kim et al. 2017). However, there are conflicting opinions regarding whether it inhibits or stimulates myelination (Latasa et al. 2010).

6.1.3 Agc1 Explorer web-app

To make results completely explorable in an intuitive and immediate fashion, we have developed an interactive web-app (Agc1 Transcriptome Explorer, https://giorgilab.unibo.it/Agc1, Fig. 6.1I). Users can explore the results of the differential expression analysis, presented as a table that allows for gene filtering to selectively display data in volcano plots and boxplots featuring count values. Search bars have been incorporated to facilitate dataset navigation based on specific genes, gene groups within GOBP, KEGG, and REACTOME pathways, or according to log2FC and adjusted p-values.

The web-app was developed using a Docker container to host a Shiny server with all the data needed for the web-app to display analyses results and to make it possible to have the web-app running across different machines, times and systems

with a seamless transition if needed. The *renv* R package was used to keep track of all the packages versions needed for the web-app to run and to maintain a separate R environment from the other parts of the analysis, to keep the container as lightweight as possible.

To make plots interactive, the *plotly* library was used (supported in R via the *ggplotly* package, which enables interactivity in *ggplot2* - a popular data visualization library for R - outputs), allowing to automatically represent subset of points selected from either tables or search bars in the related plots and enabling high-resolution downloads of the displayed plot.

Based on the relevance for interactivity, we decided to introduce the Volcano plot to allow exploration of LFC and *adjusted p-value* along with the related name, a feature that wouldn't have been graphically feasible without an interactive plot. Along with the volcano plots we also introduced the boxplots, which are useful representations often used by wet-lab scientists, allowing for multiple selection to compare subsets of genes between each other.

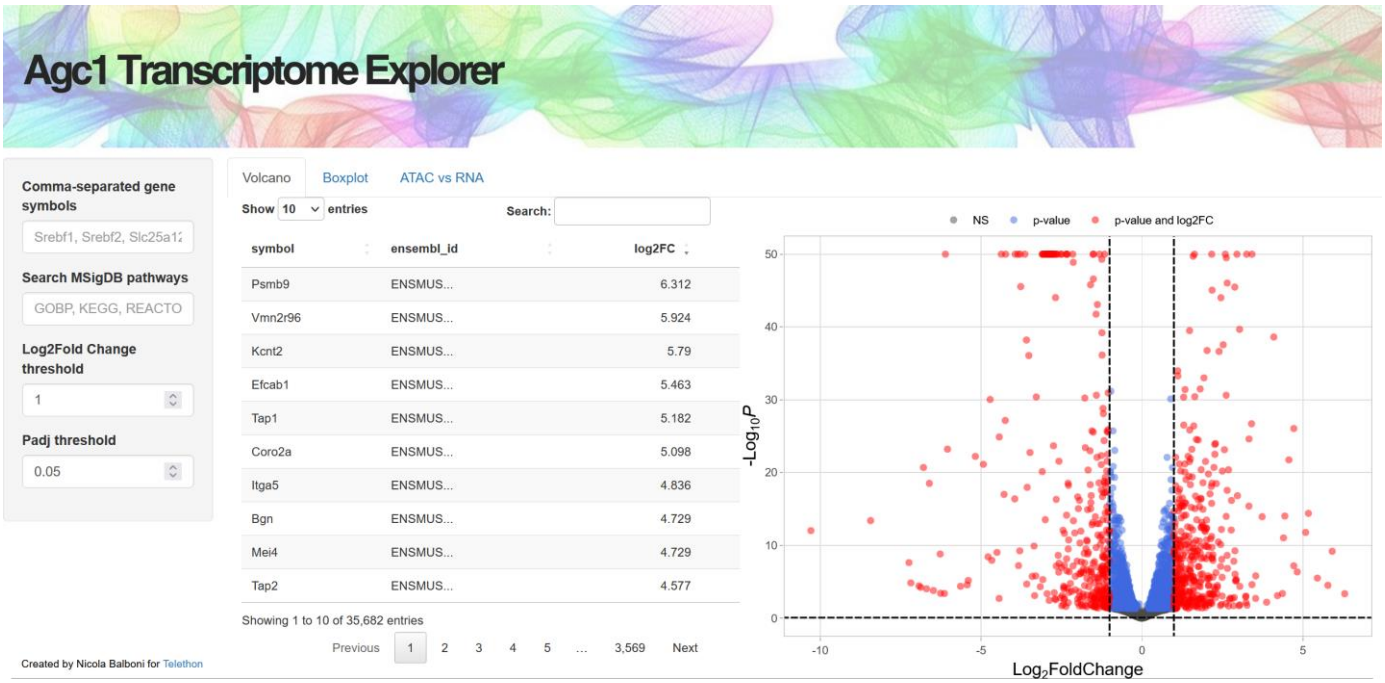

Figure 6.1I – Initial page of giorglab.unibo.it/Agc1

6.1.4 Splicing analysis

Transcript splicing plays a pivotal role in various aspects that impact oligodendrocytic cells, particularly concerning myelination. Many proteins found in myelin sheaths have been reported to undergo mRNA splicing prior to translation and splicing isoform expression can vary depending on the stage of the myelination process (Campagnoni 1988; Delarasse et al. 2006; Naef and Suter 1998).

To investigate the differential expression of splicing isoforms between our siAGC1 and control samples, we utilized MAJIQ (Vaquero-Garcia et al. 2021), a computational framework capable of identifying novel transcript isoforms and associated classical splicing variations (e.g., exon skipping and exon inclusion) as well as non-classical splicing variations (such as intron retention, alternative donor sites, and alternative acceptor sites). The results are presented as a difference in the percent spliced in (ΔPSI), with PSI representing the relative ratio of isoforms that include a specific splicing junction or retained intron. ΔPSI quantifies the change in the ratios between the two conditions for each splicing event, with values ranging from -1 to 1.

The alternative splicing analysis unveiled significant effects of AGC1 silencing on the relative abundance of isoforms of a myelin-related protein, Pmp22.

Pmp22 is known for its predominant expression in myelinating cells of the peripheral nervous system, although it has also been detected in the central nervous system. Pmp22 plays a crucial role in the proper myelination of peripheral nerves and the maintenance of axon myelination (Li et al. 2013). Genetic alterations, such as gene duplication or mutations, in Pmp22 are responsible for various inherited peripheral neuropathies, including Charcot-Marie-Tooth type-1A (CMT1A), hereditary neuropathy with liability to pressure palsies (HNPP), and a subtype of Dejerine-Sottas Syndrome (DSS) (Wang et al. 2017). In siAGC1 samples, we observe a specific local splicing variation (depicted in blue, Fig. 6.1J) involving exon 5 splicing upstream with exon 3, occurring 50% more frequently compared to control samples. In contrast, control samples do not exhibit such significant changes in local splicing variations compared to siAGC1 samples. Notably, a case involving two siblings with CMT1 highlighted a deletion in Pmp22 (Peripheral Myelin Protein 22) exon 4, which results in the mislocalization of Pmp22 to the endoplasmic reticulum instead of the

plasma membrane. This phenomenon contributes to the pathogenesis of CMT1 (see Fig. 6.1J).

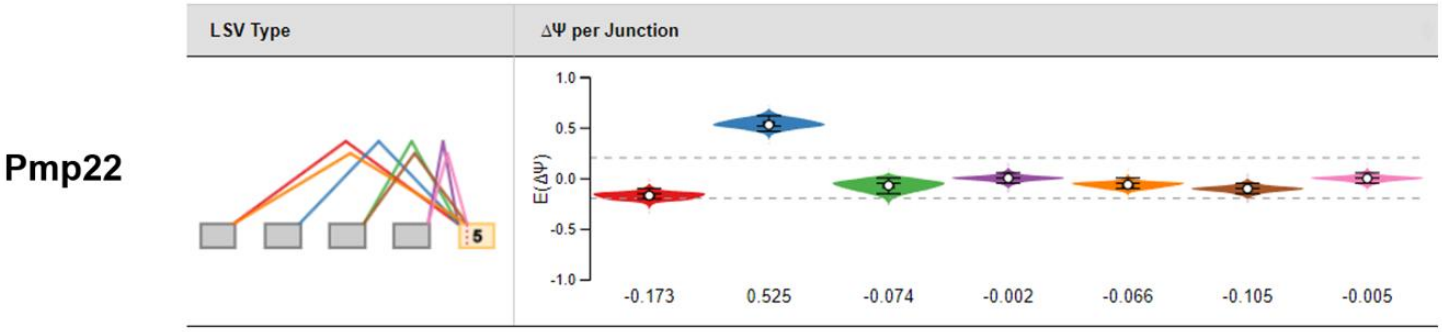

Figure 6.1J – Splicing events related to Pmp22's exon 5. Left panel shows the registered LSVs for exon 5 with different colors for each event, while on the right are violin plots reporting the estimated changes of inclusion for the specific event between siAgc1 and control

Furthermore, Retinoic acid receptor gamma (Rarg) also exhibits a significant difference in isoform levels, with silenced samples showing an exon-skipping event involving five exons, occurring 35% more frequently compared to control reads. Although the functionality of the proteins translated from the two main isoforms expressed in siAGC1 and control samples remains unknown, it is worth noting that different isoforms of Rarg are expressed at various stages of mouse embryo development (Noll and Miller 1994). Additionally, since Rarg negatively regulates OPCs differentiation in the spinal cord (Joubert et al. 2010) and in the Oli-neu model (Dahlin et al. 2015), a potential disruption in splicing variants could potentially impede this process (see Fig. 6.1K).

Rarg

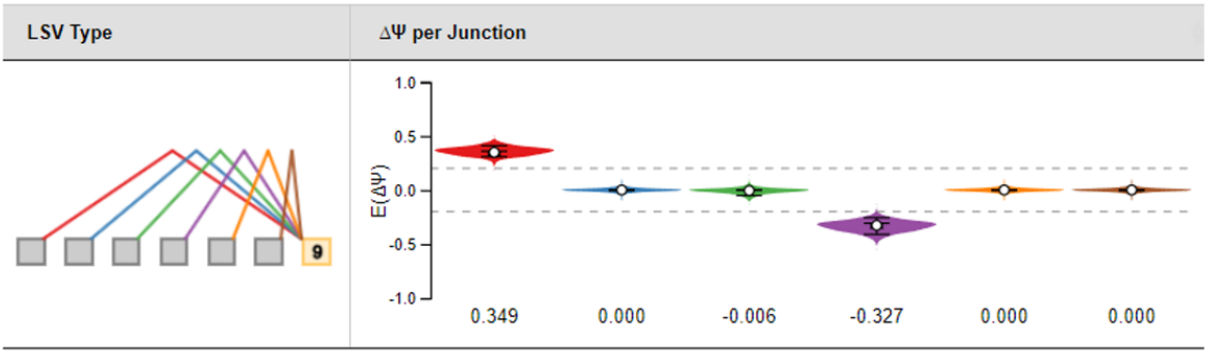

Figure 6.1K – Splicing events related to Rarg's exon 9.

Along with Majiq results, we tested another known package for differential isoform analysis, DEXSeq (Anders, Reyes, and Huber 2012), which is an exon-based method (differently form MAJIQ that is an event-based method) that uses a generalized linear model to model the differential usage of exons in different sample groups. The assumptions are similar to those of DESeq2: basically, the counts in the exons follow a negative binomial distribution and it estimates the biological variability for each exon. Results show an overall downregulation in expression of both genes (represented in the first row of figure 6.1L) but also significant alterations in exon usage (changes in the expression of individual exons that are not simply the consequence of overall up- or downregulation of the gene, second row of figure 6.1L) which consist of an upregulation in exon usage of first exon for Pmp22 and a downregulation on the second exon, while Rarg show upregulation on exon usage for exon 14 and 15, and a strong downregulation in exon usage of exon 17 to 21.

We decided to test these two different methods due to the different approaches they utilize in analyzing splicing events/exon usage, but both report as significant results these two proteins that are deeply involved cellular processes related to AGC1 deficiency.

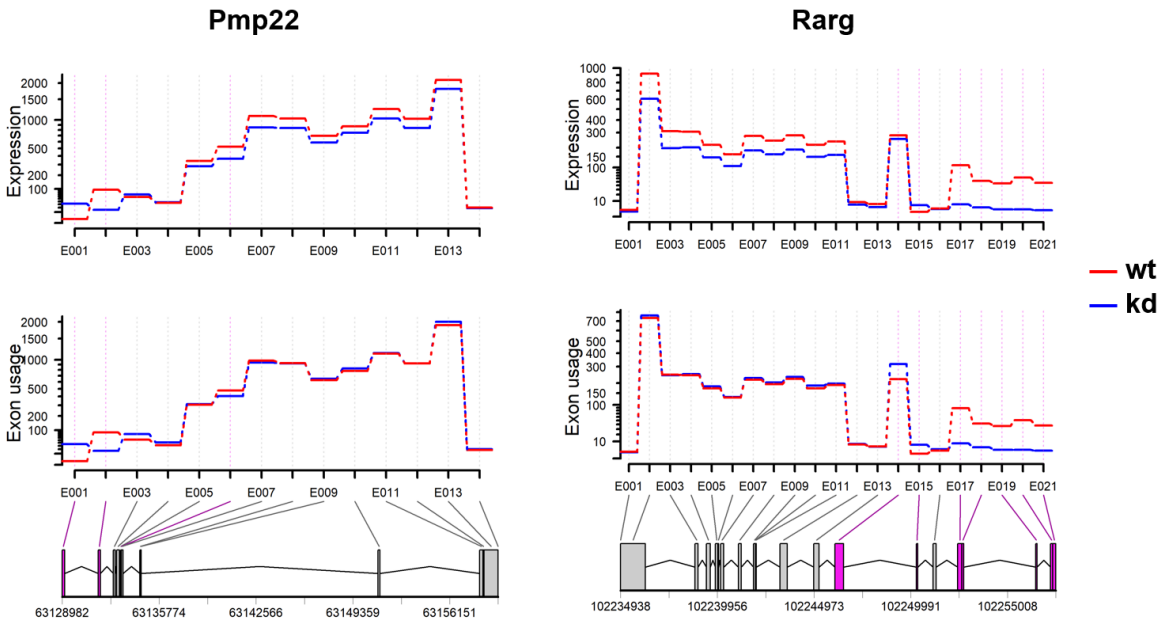

Figure 6.1L – DEXSeq results showing in the first row single expression levels exon-by-exon in Pmp 22 and Rarg and exon-by-exon usage in second row.

6.1.5 Master Regulator Analysis

To comprehensively grasp the molecular implications of AGC1 deficiency, a deeper exploration of the transcriptional data is required. This entails investigating both differential splicing and leveraging Master Regulator Analysis, a methodology designed to identify the most likely transcriptional regulators responsible for the observed transcriptional change.

Master Regulator Analysis was conducted using the *corto* R package (Mercatelli, Lopez-Garcia, and Giorgi 2020), which combines established algorithms for gene network generation, namely ARACNe-AP (Lachmann et al. 2016) and VIPER (Ding et al. 2018). This analysis revolves around estimating siAGC1-specific activation levels, represented by enrichment scores, within precise and contextually relevant gene coexpression networks centered around transcriptional regulators. To achieve this, a network of gene interactions is constructed using a sizable (>100 samples) tissue-specific RNA-seq dataset, establishing regulons that comprise groups of target genes regulated by a common master regulator. The activation level of a master regulator is subsequently inferred based on the expression patterns of its target genes within a specific experimental context (e.g., if all the target genes of

Myc, considered a master regulator, are positively upregulated in a treated vs. control experiment, then Myc's activity is considered positive and highly elevated. The degree of activation varies based on differential expression levels derived from RNA-seq analysis results).

One of the primary challenges for such algorithms lies in identifying an RNA-seq dataset closely resembling the tissue or cell type under study: this requirement necessitates a dataset with a sufficient number of samples to accurately reconstruct gene networks. To address this, we examined two different networks obtained from GTEx tissues: the frontal cortex and hippocampus (Aguet et al. 2017), which were converted to mouse orthologs via the DIOPT consensus phylogenetic approach.

After selecting the gene expression data for network construction and creating the network, results from the differential expression analysis were used as input to estimate the activation levels of each master regulator for the contrast siAGC1 vs control samples.

This analysis unveiled two master regulators, Prox1 and Smarcc2, which we hypothesize play pivotal roles in the underlying transcriptional alterations discussed earlier. Notably, one of the most prominent master regulators with consistently activated regulons in both networks is PROX1 (Fig 6.1M). PROX1 is a homeobox transcription regulator known to influence cell-fate decisions between neurogenesis and oligodendrogenesis. It acts as a robust repressor of neuronal lineage commitment in Neural Stem Cells (NSCs) derived from the sub-ventricular zone (Chang and Teng 2018). Importantly, PROX1 is highly expressed in both OPCs and OLs, with its expression increasing as OPCs undergo differentiation (Kato et al. 2015). Furthermore, it restricts OPC proliferation by inhibiting NG2 expression (Narayanan et al. 2015). These findings align closely with our in vitro AGC1 deficiency model, which exhibits reduced OPC proliferation, decreased NG2 expression, and premature differentiation (Petralla et al. 2019; Poeta et al. 2022).

Smarcc2 emerges as another significant master regulator, with an enriched regulon displaying a negative association (Fig. 6.1M). Human SMARCC2 encodes BAF170, a core subunit of the ATP-dependent BAF complexes (analogous to the SWI/SNF complex in mammals). These complexes are known for their role in regulating chromatin remodeling and gene expression during embryogenesis and

are vital for neurodevelopment. Studies on Smarcc2 knockout mice have demonstrated that its absence leads to BAF complex degradation, impairing forebrain development (Abbas et al. 2021). Specifically, BAF170, along with BAF155, plays a critical role in oligodendrogenesis: deletion of BAF170 impairs proliferation, depleting the number of PDGFRα-positive (a marker of differentiating OPCs) cells in the mouse forebrain, and hinders differentiation from OPCs to immature OLs (Berry, Wang, and Lu 2020). These findings align with prior observations concerning defects in the AGC1 deficiency model related to proliferation and differentiation processes, underscoring the strong connection between epigenetic dysregulation and OL maturation, further reinforcing the established link between epigenetic mechanisms and OL maturation (de Ferra et al. 1985).

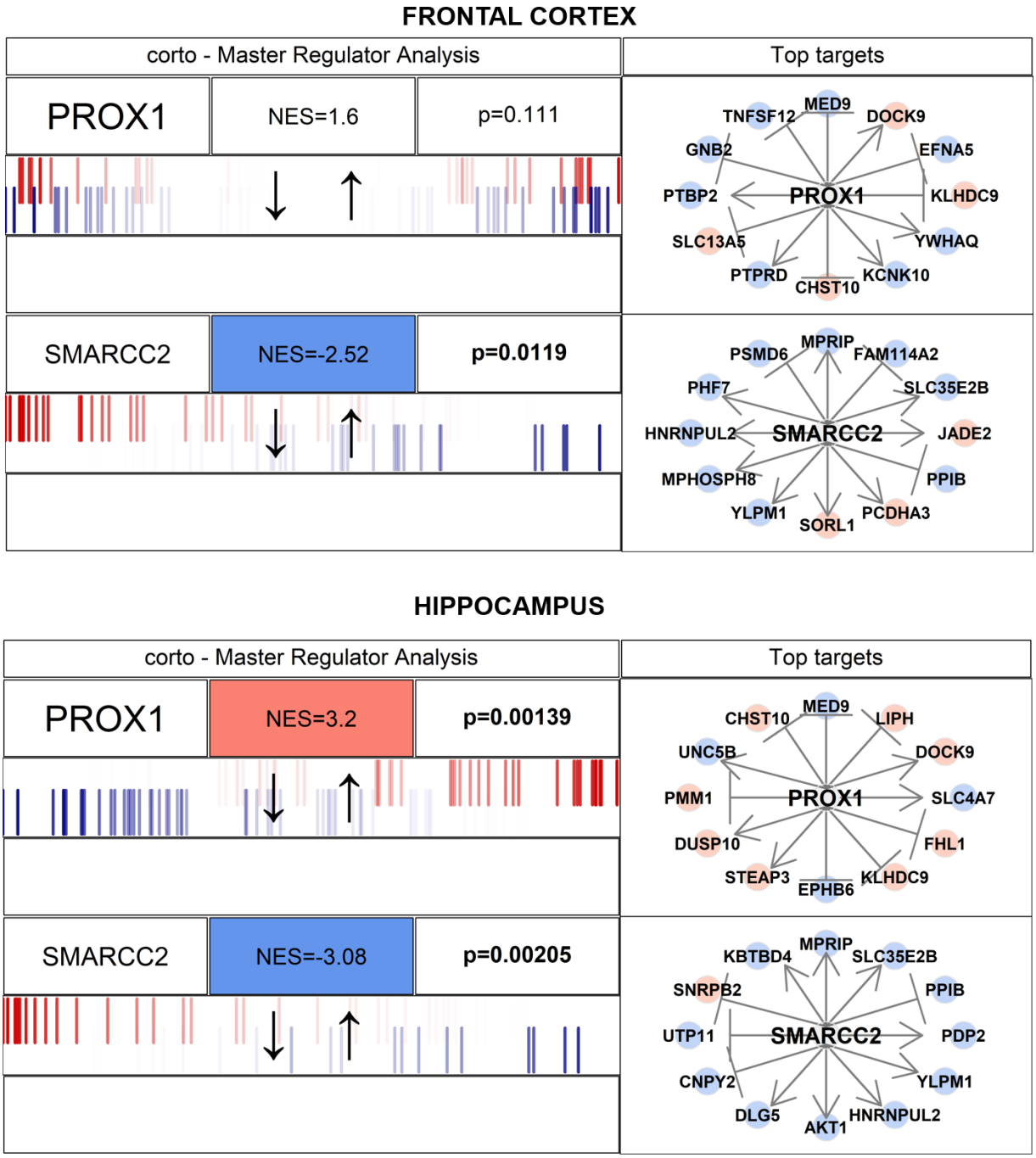

Figure 6.1M – Master Regulator Analysis results

6.2 ATAC-Seq OliNeu

6.2.1 Experimental setup

Driven by previous results from Poeta et al. (2022), which found epigenetic alterations significantly affecting proliferation and differentiation capabilities in OliNeu cells, we decided to investigate the epigenetic landscape of siAgc1 OliNeu cells to gather a deeper knowledge of regulatory processes involving chromatin remodelling. This analysis, coupled with transcriptomic data, would also allow us to get a better understanding of regulatory processes as a whole in Agc1 deficiency.

For the analysis, the same siAgc1 OliNeu (carrying a partial silencing of Agc1 via a short-interfering RNA) and control OliNeu cells were used, with 4 different replicates per condition. ATAC-Seq (Assay for Transposase Accessible Chromatin) was performed on the samples. ATAC-Seq allows for the detection of both open regions and histon-bound DNA, with a sensitivity for open regions that is usually sufficient for the detection of TF's footprinting, thus enabling the study of TF's binding from a quantitative standpoint, which paves the way for the study of the impact of a TF's binding levels on its' targets expression levels, obtained by cross-referencing chromatin accessibility data with RNA-Seq data.

6.2.2 Differential Accessibility Analysis

ATAC-Seq data were obtained in the form of FASTQ data that were first aligned to mouse genome GRCm39 using Subread. The results from alignment are reported below in table 6.2B.

Sample	Mapped reads	Unmapped reads	Percent mapped reads (%)
Control 1	63659068	7502158	88.2
Control 2	67209831	8226687	87.8
Control 3	92286175	13291061	85.6
Control 4	53560959	6654149	87.6
siAgc1 1	35031727	5494849	84.3
siAgc1 2	108518069	15911575	85.3
siAgc1 3	46305165	6691777	85.5

Peak calling, a procedure to identify regions of the genome where there is an enrichment of sequenced reads compared to the background, was done using MACS2. These regions are referred to as "peaks" and represent various biological phenomena, such as transcription factor binding sites, histone modifications, open chromatin regions, or other genomic features. Encode-suggested blacklisted regions (Amemiya, Kundaje, and Boyle 2019) were removed prior to continuing with downstream analysis. These blacklisted regions are removed due to presence of repetitive elements, high signal-to-noise ratio or the fact that they are more prone to artifacts. Peaks were then annotated to genes using a UCSC-derived mm39 annotation database and, after counting, a peak count matrix was used as input for DESeq2 to model differential chromatin accessibility.

Results from the alignment procedure show the typical ATAC-Seq profile when considering read counts with respect to fragment size: we can see, even if faintly, an increase in the number of read corresponding to fragments coming from nucleosome-free regions (NFR, < 100bp), mono-nucleosomic regions (180 - 247bp) and di-nucleosomic regions (315 - 437bp) in both siAgc1 and control samples (Fig 6.2A). The fact that the increase in read counts while in proximity of these regions largely depends on how the experimental procedure was carried out, but the fact that these regions are visible is still considered a good indicator of reliable data.

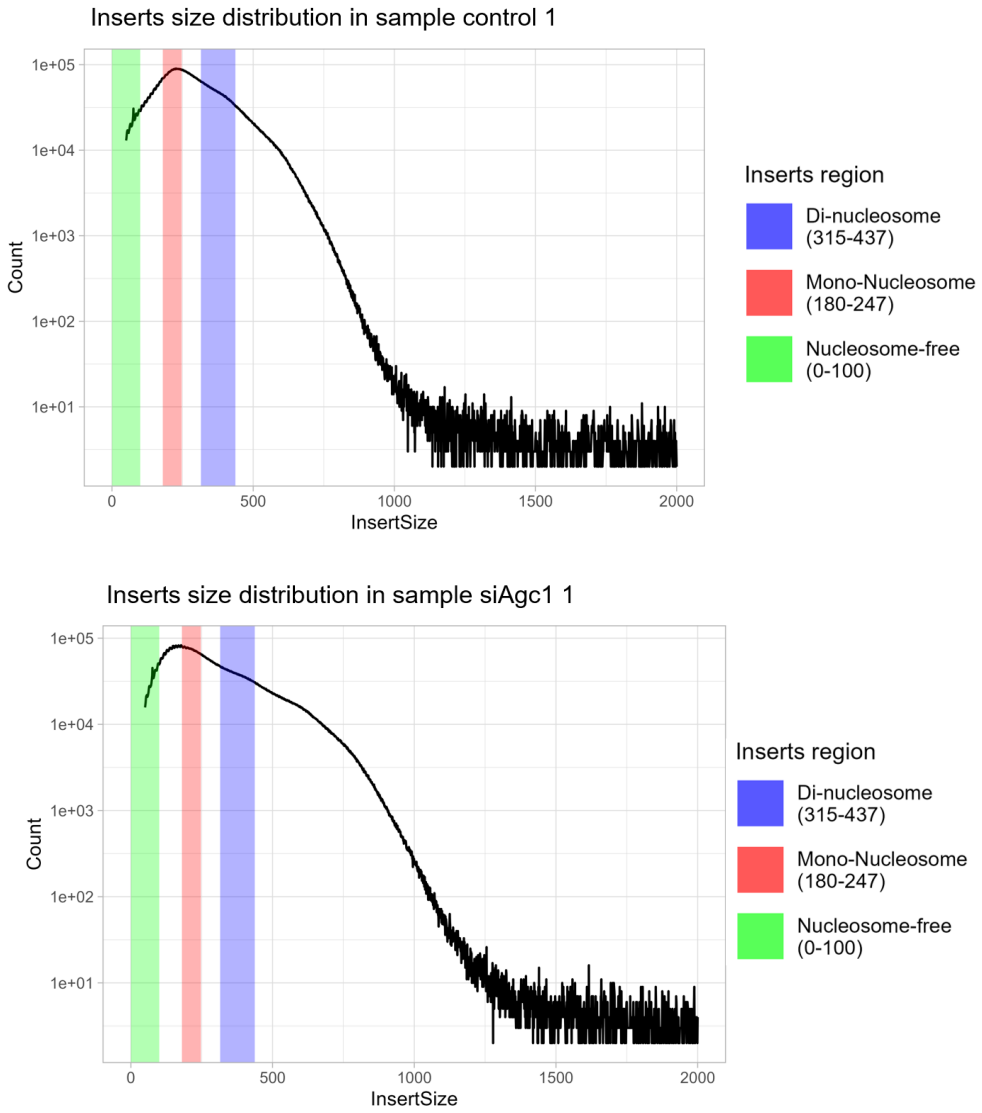

Figure 6.2A – Insert size distribution counts in samples control 1 and siAgc1 1.

At this point the focus shifted solely on peaks coming from NFRs, since these are the regions that would be accessible for TFs binding and whose accessibility would be most impactful for transcription regulation. First, to check consistency of the data, we assessed whether called peaks mapped to the same regions across samples and conditions or they diverged in some way. A large part of the peaks falls, as expected, in the promoter regions, with almost the entirety of these peaks coming from the proximal upstream promoter region (between 0 and 1000 bp) with only a minimal part mapping to distal promoter regions (above 1000 bp). The rest of the features to which the read count map to are mostly intergenic and intronic regions, and just a residual part mapping to exons. A key takeaway, which also supports the

reliability of the dataset, is that the feature percentages are comparable across samples and conditions (Fig. 6.2B).

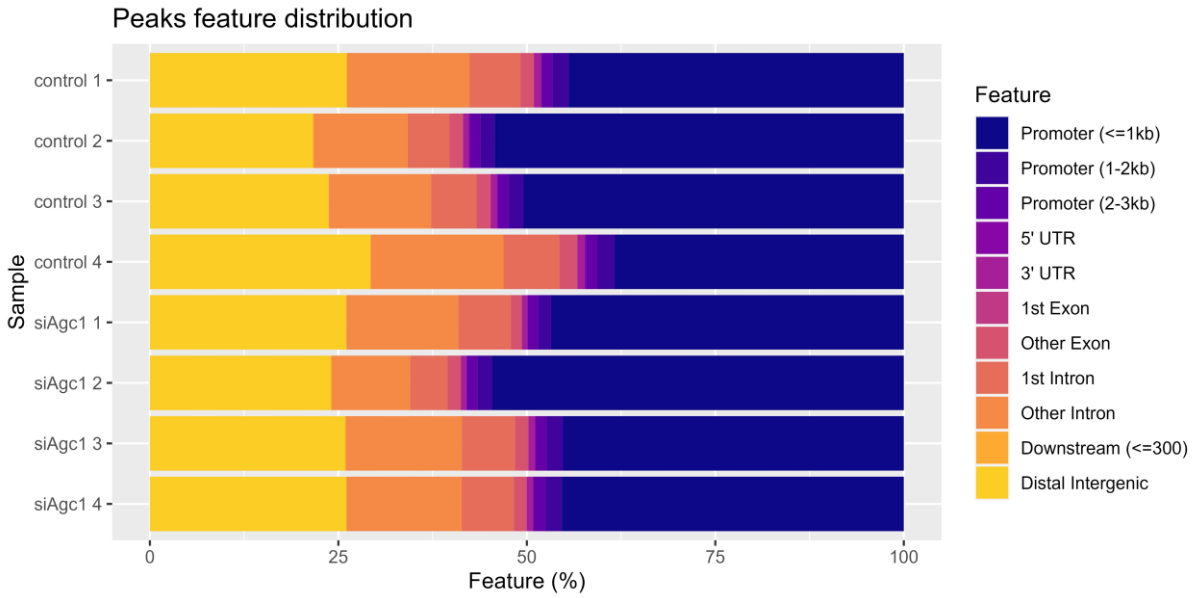

Figure 6.2B – Percentage features distribution of mapped peaks in control and siAgc1 samples

For the subsequent differential accessibility analysis we collapsed redundant peaks to unique peaks and selected those peaks that were present in at least 2 samples, to be able to compute significant fold changes. As shown in figure 6.2C, for each condition there are samples that carry unique peaks that are not shared with other samples. For control samples 60,88% of peaks are present in at least 2 samples, while in siAgc1 samples this percentage amounts to 16,35%. For the subset of non-redundant peaks recorded in at least 2 samples that we established to use, 58,08% of the peaks are recorded at least once in both control and siAgc1, giving a relevant number of peaks to compute a differential accessibility score.

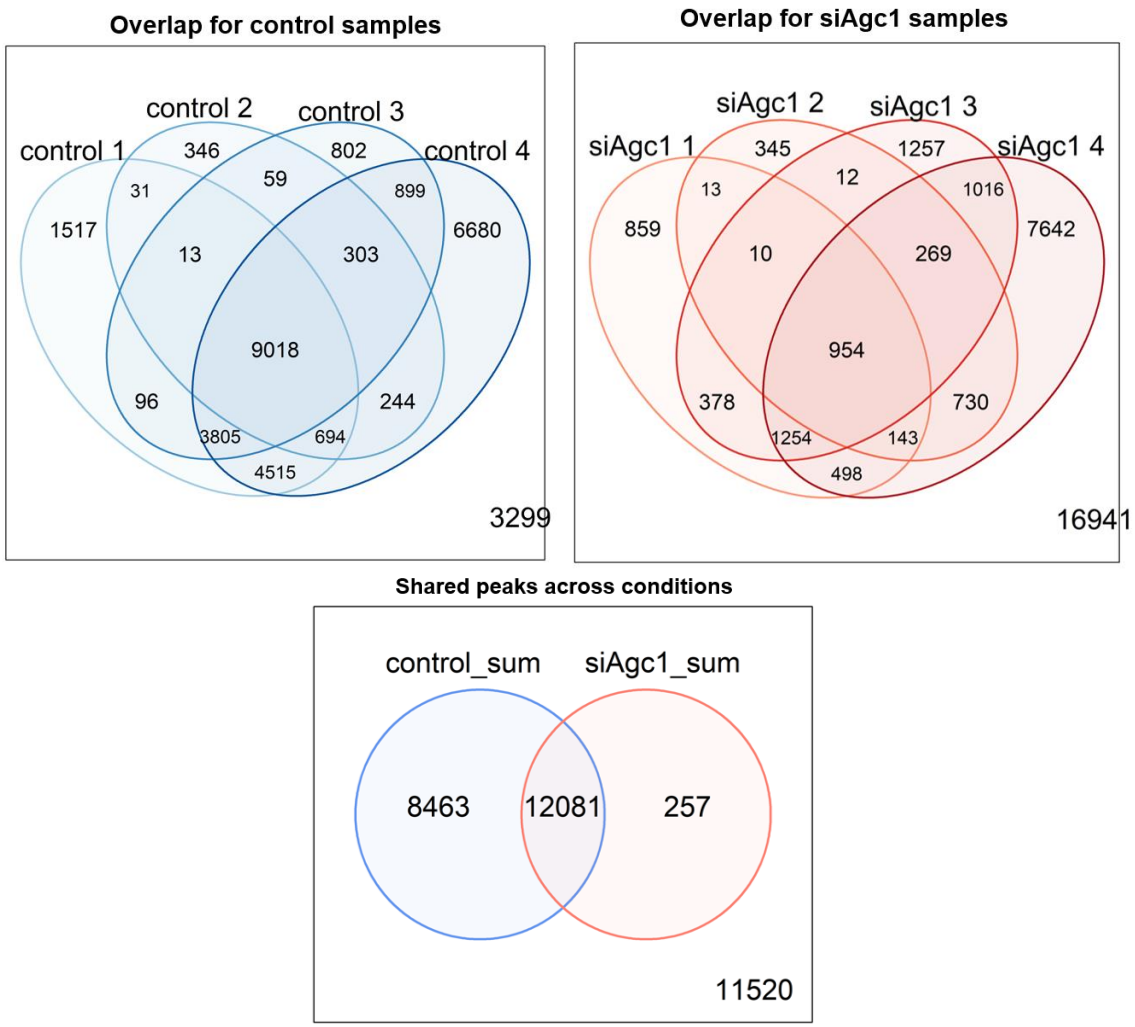

Figure 6.2C – Venn Diagram plots of shared peaks across controls (blue), shared peaks across siAgc1 samples (red), and shared peaks with at least 2 recorded peaks across all samples between control and siAgc1 samples

After obtaining a count matrix for the established peak set, we observed a clear separation between control and siAgc1 samples when checking their dimensionality reduction in PCA: control samples are markedly separated from siAgc1 samples based on their genotype, while another variable that could contribute to separation along the second principal component might be the cell passage, which could show a stronger effect on siAgc1 (Fig. 6.2D)

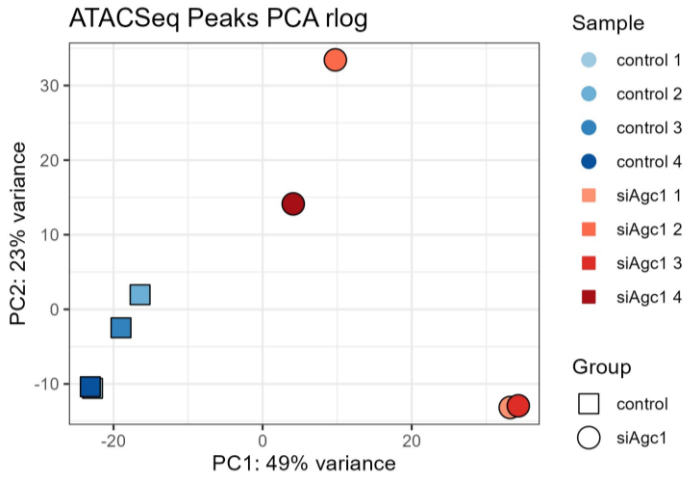

Figure 6.2D – PCA of all analyzed samples

At this point, we selected only peaks that were part of known promoter regions, ranging from -1000bp to +500 bp with respect to transcription start-site (TSS) to estimate LFCs. We report a total of 217 peaks upregulated in siAgc1 and 387 peaks downregulated in siAgc1 (Fig 6.2E). In this case an upregulation in terms of counts coming from a specific peak, represents a region in which the chromatin

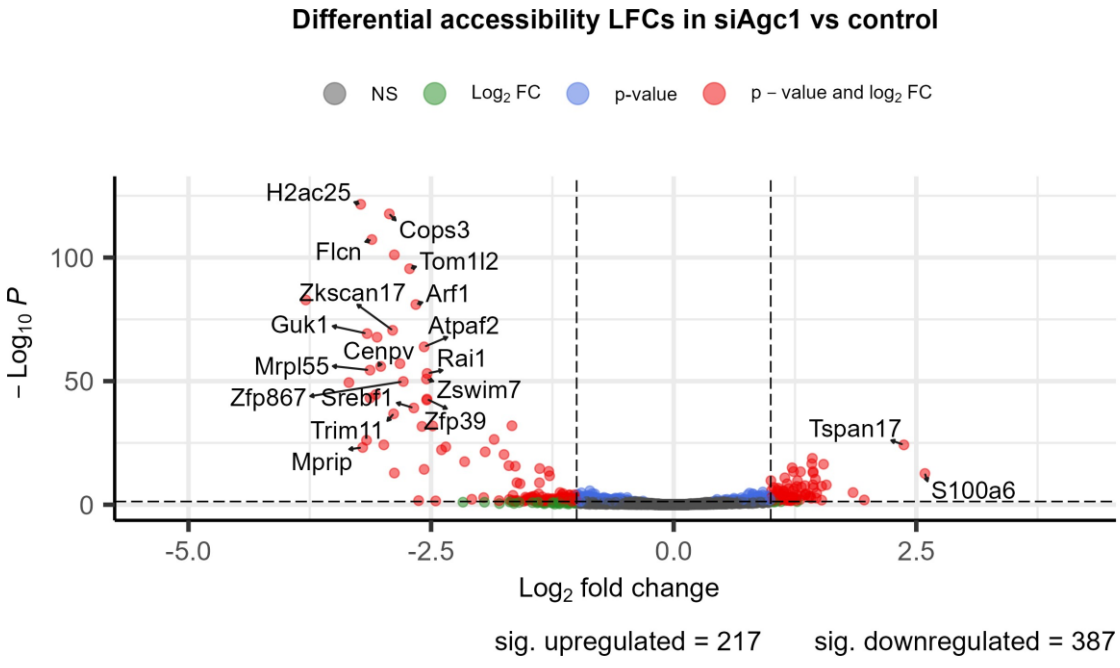

Figure 6.2E – Volcano plot of peaks counts in siAgc1 vs control OliNeu cells.

resulted in being more “open” in siAgc1 samples, compared to the same region in control samples.

Among the genes that we find to be significantly altered in the ATAC-seq experiment, we identified some of those genes that were found to be differentially expressed also from RNA-seq results. Part of these genes (Cops3, Tom1l2, Guk1, Zkscan17, Rai1, Atpaf2, Cenpv, Trim11, Mprip and other genes) are reported having a less open chromatin on their site, while also showing a downregulation in terms of RNA levels, hinting at an alteration of a regulatory process in siAgc1 cells that causes chromatin remodelling changes that bring to a closed chromatin state on the promoters of these genes. Part of these genes have already been explored for their role in the brain in the RNA-seq results chapter. Here, we highlight the fact that the downregulation in these genes is caused by an epigenetic mechanism.

6.2.3 Comparison with RNA-Seq results

After finding concordant results with RNA-seq results from the final part of the previous analysis, we focused on further investigating the relationship between chromatin and transcription in OliNeu cells. The first tested evaluation was based on a correlation analysis: our expectation was to find a positive correlation between LFCs score from chromatin and transcriptomic data, suggested by both the direction of regulation (chromatin accessibility → transcription) and from previous results that showed most downregulated genes in terms of expression also having a heavily impaired accessibility on their promoters.

As expected, Spearman’s r is positive (0.22) and statistically significant (p-value = 1.35e-80) for 7268 tested genes, representing the fact that chromatin accessibility and transcription tend to behave in concert, but, nonetheless, there are exceptions to this rule that make up a relevant number of cases (Fig. 6.2F).

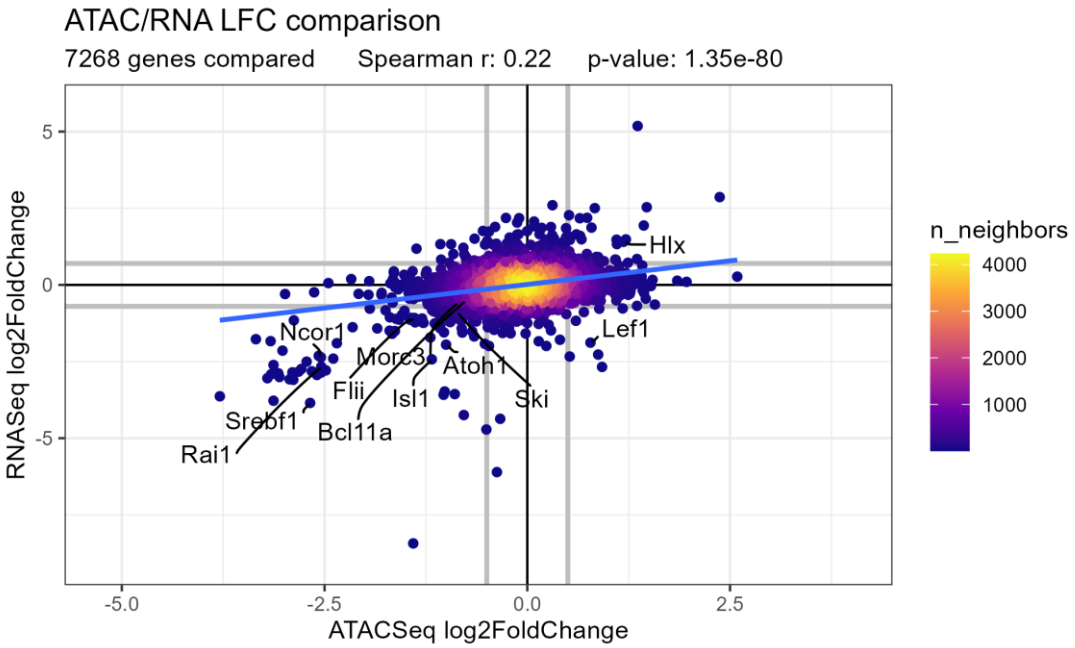

Figure 6.2F – Correlation analysis of transcriptomics LFCs over promoters' accessibility LFCs. Grey lines indicate thresholds for which we would consider interesting TFs for further analyses. Points with labels are selected TFs.

The first and third quadrant encompass points that "behave" in the expected way based on the relationship between chromatin accessibility and transcription, the more open the chromatin the higher the transcription levels, with closer chromatin. Second and fourth quadrant, on the contrary, behave in the opposite way: in the second quadrant we have genes showing a decrease in their promoter's chromatin accessibility, but their transcript levels rise, while on fourth quadrant an increase in accessibility corresponds to a decrease in transcription levels. At this point we were intrigued by the idea of if and how TFs behave differently when they have a "canonical" response to promoter occupancy compared to a "reverse" response.

diffTF (Berest et al. 2019) is a method designed to assess differential TFs activity and classify TFs as either activators or repressors. This classification is achieved by integrating various types of genome-wide chromatin accessibility data with RNA-Seq data and *in-silico* predicted TF binding sites. The method can determine if a specific TF acts as a global transcriptional repressor or activator

based on the expression of putative targets identified via the predicted binding sites. Results are reported as significance scores over weighted mean difference (the difference in TF activity between siAgc1 cells and control), which, if we consider weighted mean difference as a form of effect size, can be reconducted to a typical volcano plot (Fig. 6.2G).

It is important to notice that the score reported along the x axis of Fig 6.2H is

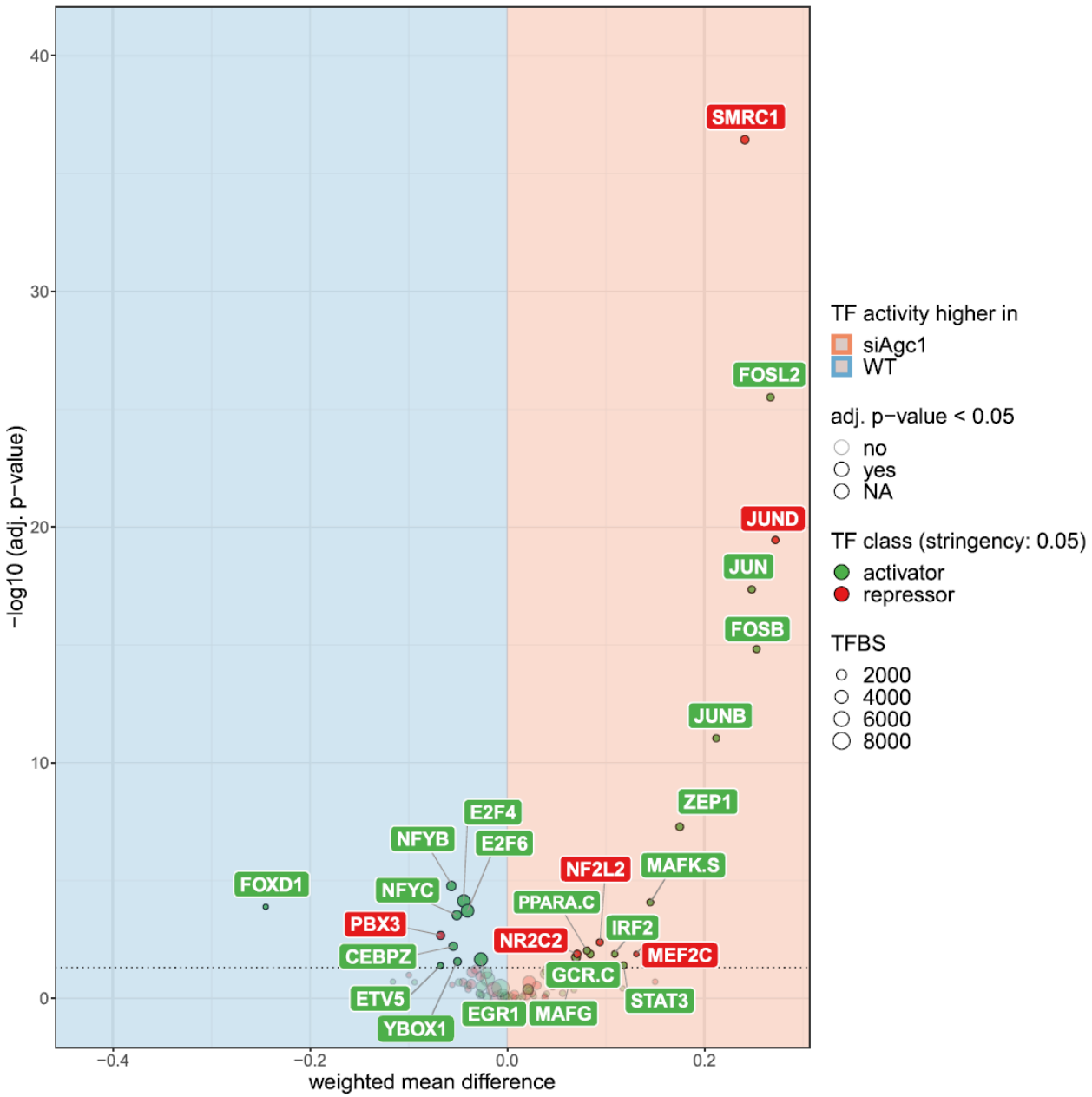

Figure 6.2G – diffTF results. On the x axis the score of the differential accessibility between the two conditions weighted across all the targets of each respective represented transcription factor, on the y axis the significance associated to the score.

not a score related to a TF promoter accessibility as we showed in previous analyses, but it's rather a weighted score based on the accessibility of its target genes, so that the score represents the weighted mean of how much their targets (genes that have a binding site for that specific TF in their promoter) are accessible.

Results from diffTF identify 20 activators showing a statistically significant weighted mean (at these thresholds of significance: 0.05 for adjusted p-value and 0.05 for TF class stringency) even though the largest p-value is associated with a repressor (5 significant in total), SMRC1, the protein encoded by Smarcc1 gene, a paralog of the best known Smarcc2: both take part in the formation of npBAF (neural – progenitor BAF) complex regulating different aspects of neural development as we reported in chapter 6.1.5. Between the most upregulated factors we also register different paralogs of the subunits that bind to form the AP-1 complex: Jun and Fos. It has been shown that levels of Jun increase when in presence of nerve damage and is key for neuronal regeneration (Raivich et al. 2004), which in our case could be a system's response to the lack of well-myelinated axons. JunB and FosB, similarly, increase their expression in response to an increase of stimuli the evoke seizures in order to repress the activity of MMP-9, a metalloproteinase associated with neurodegeneration (Rylski et al. 2009). So, even in the case of JunB and FosB, this seems to be a response mechanism to the degenerative condition that is verified in Agc1 deficiency patients. Foxd1, on the contrary, is downregulated, and since this TF is known to sustain the expression of specific genes expressed in anterior-hypothalamic mature neurons (Newman et al. 2018), this could be a further signal of impaired differentiation in siAgc1 OliNeu cells.

6.3 RNA-Seq iPSCs

6.3.1 Experimental setup

For this part of the project we analyzed 6 different human iPSCs (induced Pluripotent Stem Cells) obtained by dermal fibroblasts of different patients: 2 healthy donors were used as controls (C1, C2) and 3 samples from patients affected by Agc1 deficiency, each sample was analysed in 3 technical replicates, summing up to 6 control replicates and 9 patients' replicates. The samples used to model Agc1 deficiency were obtained by 2 different patients: patient 1 (P1) previously described in (Falk et al. 2014) carries a homozygous missense mutation (c.1058G>A; p.R353Q) in SLC25A12 encoding a mutant AGC1 with low residual activity (~15% of WT AGC1), and patient 2 (P2, from which two clones P2A and P2B were obtained) carrying two compound heterozygous variants (c.225del; p.(Glu76Serfs*17) and c.1747C>A; p.(=)), leading to a complete loss of AGC1 protein expression (Bölsterli et al. 2022). After collecting the dermal fibroblasts from patients, these were differentiated into neuronal progenitors committed to immature proliferating neurons, that were tested for their phenotype via Western Blot analysis and immunohistochemistry. These samples were then analyzed via RNA-Seq, alignment results are shown in the table 6.3A below:

Sample	Total reads number	Mapping Genome Ratio (%)
C1 rep 1	90061564	96.15
C1 rep 2	90039870	96.33
C1 rep 3	90126082	96.15
C2 rep 1	92444880	95.54
C2 rep 2	92431888	95.91
C2 rep 3	92371084	95.99
P1 rep 1	92178698	94.23
P1 rep 2	92562372	94.00
P1 rep 3	92371084	94.08
P2A rep 1	92359998	94.56
P2A rep 2	90063754	94.78
P2A rep 3	90113510	94.88
P2B rep 1	89940376	95.13
P2B rep 2	90043606	95.19
P2B rep 3	90127796	95.11

Table 6.3A – Alignment statistics for control and mutated Agc1 iPSCs

6.3.2 Exploratory Data analysis and Differential Gene Expression

Following the rationale of the previous analysis, we checked the coherence between the experimental design and the data obtained from RNA-Seq. The distance matrix reports as expected, smaller intra-genotype distances than inter-genotype distances, while also showing replicates of the same sample having smaller distances in each sample. We also report a smaller distance between P2A and P2B samples, coming from the fact that they are clones of the same cell line. Interestingly enough, control samples have higher distances between each other compared to patients' samples, which could hint at the fact that even though patients are characterized by different mutations leading to different levels of AGC1 protein in the cells, the overall condition from a transcriptomic point of view seems to be quite similar (Fig. 6.3A).

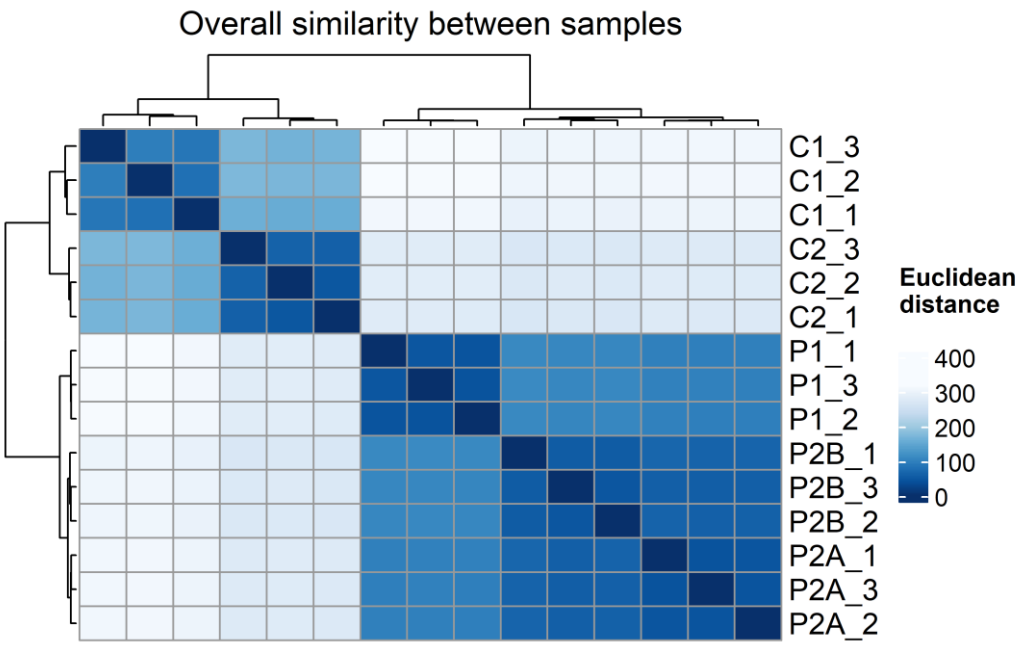

Figure 6.3A – Distance matrix of iPSCs Agc1 deficiency model replicates on transcriptomic data

These results are confirmed also from dimensionality reduction (PCA), where the first principal component, explaining 79% of the total variance of the data, separates the samples based on their genotype (Fig 6.3B). The second principal component seems to be driven mostly by the differences between the controls, an effect recorded also from the distance matrix.

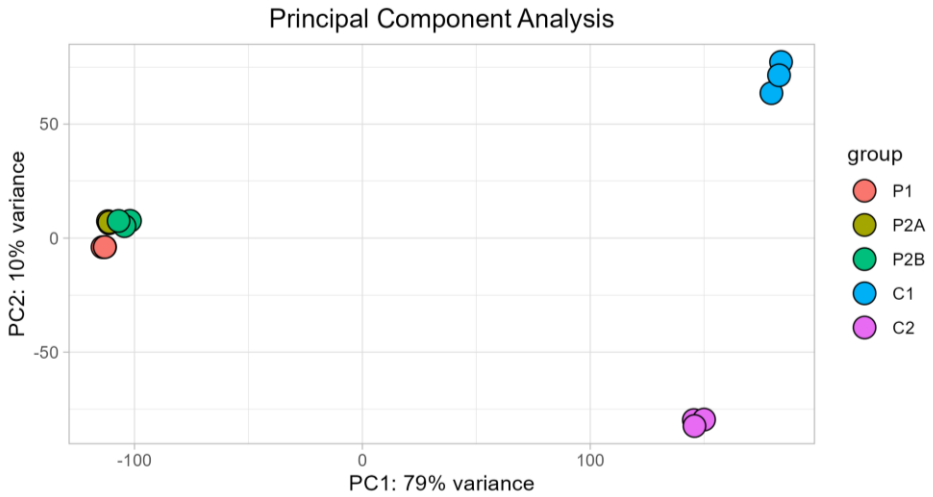

Figure 6.3B – Control and patients iPSCs PCA from transcriptomic data

In terms of differential expression, we report a total of 3634 significantly downregulated genes and 4695 upregulated genes, with SLC25A12 reported as downregulated here, while SLC25A13, which encodes AGC2, isoform of AGC1 with the same metabolic function, is upregulated (Fig. 6.3C). The increase in amount of significantly altered genes can be partially explained by the larger number of replicates that has been used for this experiment, which greatly increases the power of the statistical analysis conducted here. Another possible explanation for this increase in number of significant genes can be due to the reprogramming step that the cells had to undergo to become pluripotent and then neural precursors, a different type of precursor with a different set of expressed genes compared to OPCs like the OliNeu cells.

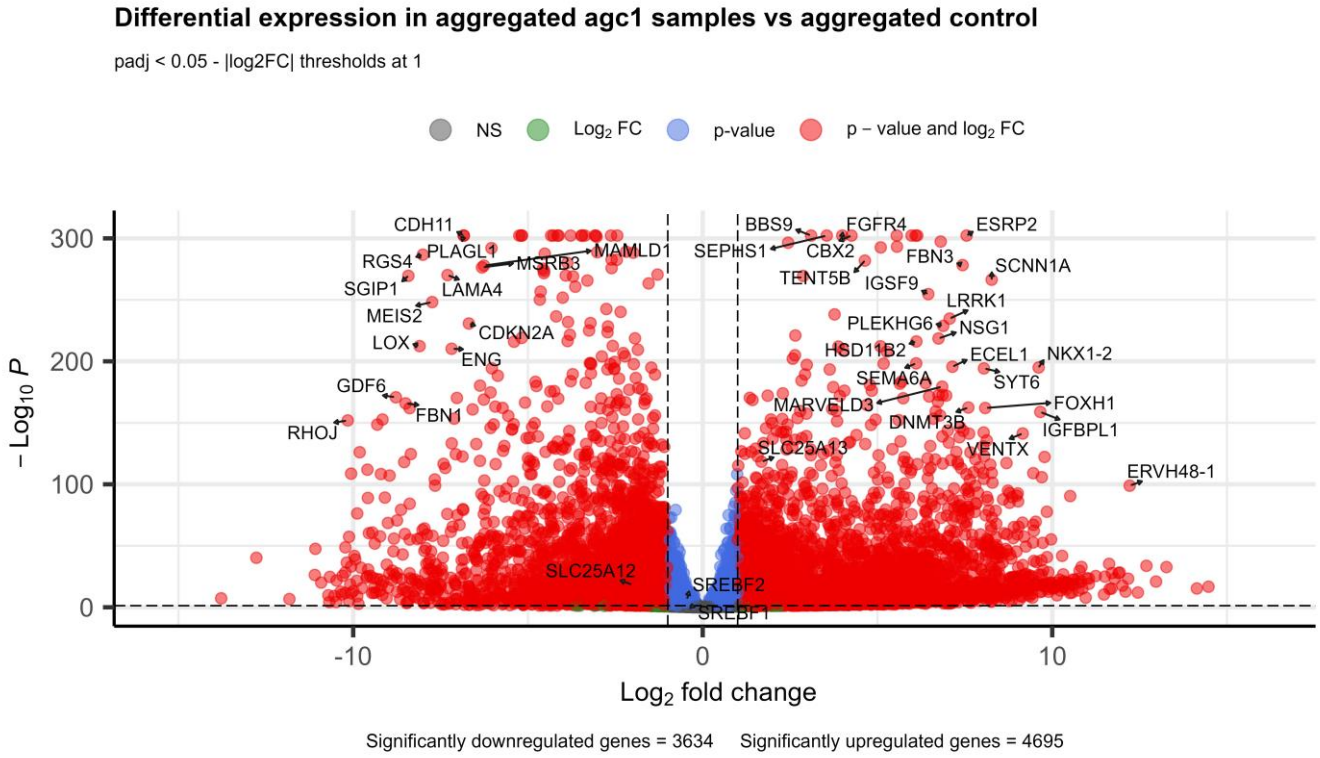

Figure 6.3C – Volcano plot of AGC1 deficiency patients iPSCs vs control iPSCs

It's also relevant to point out how the transcript levels of SLC25A13 (isoform of SLC25A12) are higher in patients' cells than control cells, probably because of a compensatory mechanism which enables to sustain the mitochondrial bioenergetics of patients' cells. Even in this model we report a significant, even if small in terms of effect size, downregulation of both SREBF1 and SREBF2, the two key fatty-acid synthesis regulators we reported downregulated in OliNeu cells. This result hints at

the fact that the underactivation of the biosynthetic pathway of fatty acids is also embedded in patients' cells and supports the idea that this metabolic and regulatory alteration is shared across different models of AGC1 deficiency.

6.3.3 GSEA results

GSEA (Gene Set Enrichment Analysis) was restricted the search to pathways included in the Hallmark gene sets and Gene Ontology: Biological Processes gene sets. Results from this analysis on iPSCs show a greater number of significantly altered pathways, with some similarities with the results obtained from GSEA on OliNeu cells: two pathways related to neurite outgrowth (Regulation of Cell Projection Size and Dendrite Extension) significantly upregulated, which, as previously mentioned, align with a more mature cellular profile. This result goes along with the downregulation of the epithelial-to-mesenchymal transition (EMT), a biological process wherein epithelial cells, characterized by their polarity and tight cell-cell junctions, undergo a series of molecular and morphological changes to adopt a mesenchymal phenotype resulting in a more migratory behavior compared to their epithelial counterparts (Fig. 6.3D).

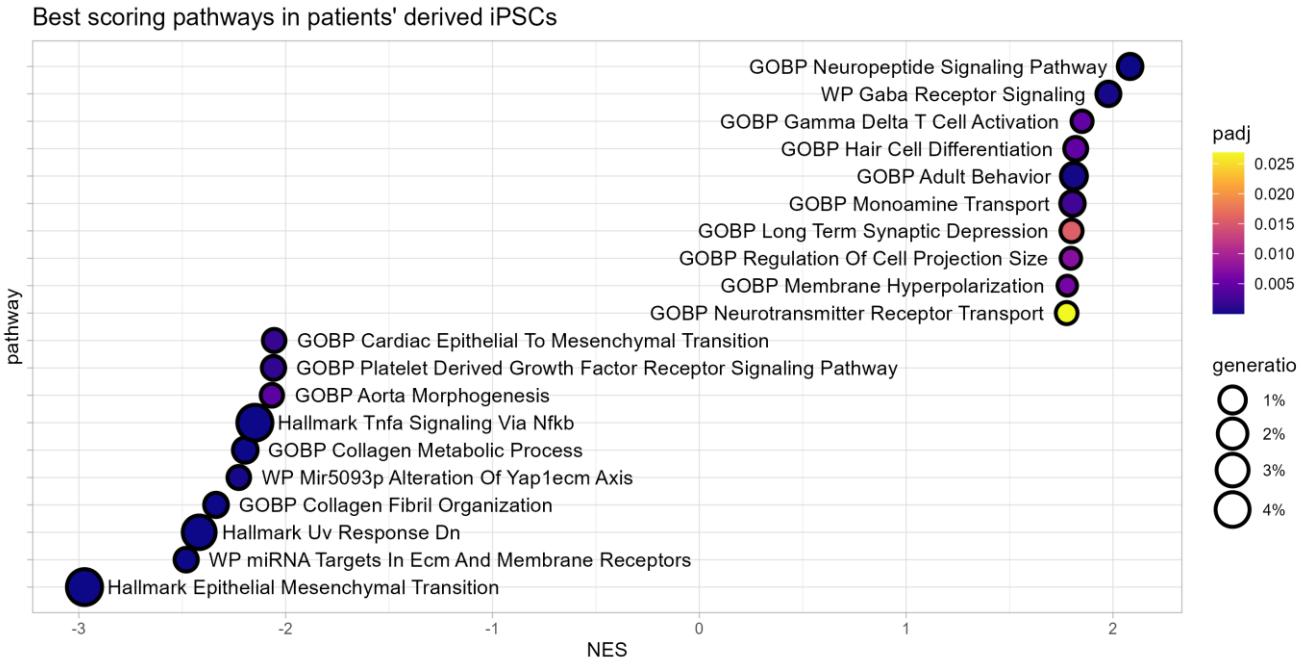

Figure 6.3D – Top and bottom 10 pathways by enrichment score for iPSCs vs control.

We also report an upregulation in two key signaling pathways: neuropeptides and monoamine. Neuropeptides signaling is a generic term due to the vast number of processes the neurotransmitters regulate in the brain, but coupled with the upregulation of the neurotransmitter receptor transport pathway it becomes clearer that the function in this specific case that these molecules are actively supporting neurotransmission in AGC1 deficiency. Monoamines are neurotransmitters such as dopamine, serotonin, and norepinephrine, which play a pivotal role in regulating higher brain functions. These neurotransmitters modulate various cognitive, emotional, and motor processes in the brain. Alterations in monoaminergic systems have been implicated in several brain diseases. For instance in a rat model of AD, there was a significant decrease in the levels of norepinephrine, dopamine, and serotonin in various brain regions (Ganguly and Guha 2008) and even in Parkinson's disease a dopamine deficit is reported to be characteristic (Masato et al. 2019). The upregulation in the monoamine transport pathway could therefore be related to an increased request from brain cells for these neurotransmitters in Agc1 deficiency.

7 Discussion

In this thesis we present the work related to different Telethon projects with the aim of elucidating the molecular mechanism that play a key role in the development and upkeep of Agc1 deficiency, an ultrarare disease that affects newborn babies starting a few months after birth, impairing their development and causing severe symptoms including seizures, muscle hypotonia and global cerebral hypomyelination.

Two cellular models were established to study the disease:

1. A targeted genetic downregulation of mouse oligodendrocyte precursor cells (OPCs) by partial silencing of Slc25a12 (the gene encoding AGC1 protein) in OliNeu mouse cells (silenced cells named siAgc1). This led to a reduction in AGC1 abundance to approximately 40% of the physiological levels for both mRNA and protein, mirroring observations in human patients (Petralla et al. 2019).
2. Generation of neural progenitors from the reprogramming of human dermal fibroblasts of two male Agc1 deficiency patients into iPSCs, leading to 3 different lines: one from a patient (P1) which showed 15% of AGC1 residual activity and two (P2A, P2B) from a patient which showed no residual activity of AGC1.

In OliNeu cells, silencing AGC1 resulted in a notable transcriptome shift in the OPCs' model compared to control cells, affecting the expression profiles of over 700 genes. Notably, AGC1's paralog, AGC2 (encoded by the Slc25a13 gene), was downregulated at the mRNA level but upregulated at the protein level, suggesting a potential compensatory mechanism.

Among the upregulated genes in response to Agc1 knockdown, several are associated with cell adhesion, such as Coro2a (encoding the Coronin 2A protein), Itga5 (integrin subunit alpha 5), Bgn (Biglycan), and Fat3 (Fat cadherin 3). These findings suggest structural remodeling of Oli-Neu cells reported also from optical microscopy analysis (not reported in this work), which revealed longer branches and a more elongated and branched cellular morphology compared to control cells.

Notably, several genes upregulated in siAGC1 cells are also upregulated in Alzheimer's disease (AD), including Itga5, Bgn, Fat3, Tap1 (Transporter 1, ATP-binding cassette), Kcnt2 (Potassium channel subfamily T member 2), Efcab1 (EF-hand calcium-binding domain-containing protein 1), and Psmb9 (Proteasome subunit beta type-9) (Nasrabady et al. 2018; Sanna et al. 2023). This similarity suggests shared dysregulation of OPCs/OLs proliferation and differentiation processes between AGC1 deficiency and AD, potentially contributing to abnormal white matter formation and cognitive dysfunction (Lorenzini et al. 2020; Pu, Stephenson, and Yong 2018). More genes overlapping with those affected in AD, include the nuclear receptor corepressor Ncor1 (Castelo-Branco et al. 2014). Ncor1 plays a pivotal role in controlling the fate of neural stem cells (NSCs), influencing OL-specific gene expression, oligodendrogenesis, and OL maturation (Castelo-Branco et al. 2014). This suggests a potential role of Ncor1 in white matter maintenance and its relevance to AGC1 deficiency.

Furthermore, the observed gene expression alterations, particularly the deregulation of Rai1, link AGC1 deficiency to other neurodevelopmental disorders, including Smith-Magenis and Potocky-Lupski syndromes. These disorders share features such as motor and cognitive dysfunction, seizures, and delayed neurodevelopment (Rinaldi et al. 2022), highlighting AGC1's potential influence on various biological processes beyond amino acid transport and reducing equivalents balance in brain cells.

AGC1 deficiency in OliNeu cells also led to the repression of several hundred genes, including the Srebp transcription factors encoded by Srebf1 and Srebf2 genes. These factors are central regulators of sterol biosynthesis, and their downregulation may impair OL maturation, affecting growth, myelin protein expression, and cholesterol synthesis, ultimately leading to hypomyelination (Monnerie et al. 2017). Additionally, genes encoding fatty acid synthase (Fasn) and aspartoacylase (Aspa) were downregulated, along with Akt1, a key player in OL myelination (Jaegle et al. 2003). The coordinated downregulation of these genes aligns with the observed perturbation in the fatty acid synthesis and myelination pathway, consistent with hypomyelination observed in patients (Wibom et al. 2009).

Furthermore, AGC1 downregulation led to alterations in the expression of transcription factors Sox8 and Sox9, members of the SoxE family known for regulating late stages of myelination and maintenance. While Sox8 exhibited significant upregulation, Sox9 experienced significant downregulation, in line with compensatory mechanisms described in prior studies (Turness et al. 2018). These findings support the disruption of proliferation and differentiation observed in *in vitro* models of AGC1 deficiency (Poeta et al. 2022).

Moreover, significant alterations in splicing patterns and isoform prevalence were observed for hundreds of genes, with significant differences in terms of PSI (Percent Selected Index) for Rarg (Retinoic Acid Receptor Gamma), which shows an increased number of splicing events between exon 5 and 2 for siAgc1 samples, and Pmp22 (Peripheral Myelin protein 22) which shows an significantly altered splicing events arising from exon 9.

8 samples from the OliNeu model (4 siAgc1 and 4 controls) were also tested via ATAC-Seq for chromatin accessibility. This was done to further explore the role of epigenetic regulation in Agc1 deficiency, which has been demonstrated to affect both differentiation and proliferation in the affected cells (Poeta et al. 2022).

Here, we report the results on the analysis of differential chromatin accessibility analysis on nucleosome-free regions with a specific focus on promoter's accessibility, which make up ~50% of the total collected reads. Samples result clearly separated based on their genotype in dimensionality reductions, allowing us to conclude that there is a relevant impact on the chromatin state of OPCs when the levels of Agc1 drop.

We compared the results from differential gene expression with chromatin accessibility, verifying a significant positive correlation as expected, but still reporting a number of cases where chromatin accessibility was not positively associated with increased transcription. This event is to be expected, due to the fact that chromatin accessibility and TF binding have a complex relationship, especially in specific contexts like mitosis, where chromatin accessibility does not always imply TF binding (Coux, Owens, and Navarro, n.d.).

To more accurately determine the impact of chromatin accessibility in the systems, we focused on the activity of TFs using diffTF, a framework that uses ATAC-

Seq data and RNA-Seq data in combination to assess in a more targeted and precise way the increase (or decrease) of TFs' binding activity based on the accessibility level of their targets; RNA-Seq data of both the TF and the targets is used to establish whether the TF is acting as an activator or a repressor. Results from this analysis report many more activators altered (both up and down) compared to the repressors even though the most significant result is represented by Smarcc1, which is part of the npBAF (neural-progenitor BAF) complex that is a key regulator of differentiation, cell fate and proliferation in brain development and function, regulating gene expression crucial for CNS development and function (Phelan et al. 1999). The complex, characterized by subunit variability, plays diverse roles in various cellular processes and mutations in BAF subunits are linked to numerous neurodevelopmental disorders (Ho and Crabtree 2010; Rive Le Gouard et al. 2023). Smarcc2, a paralog of Smarcc1, was also reported to have a significant and strong downregulation by master regulator analysis, which bases the selection of TF's targets on coexpression data from large transcriptomic databases. The fact that the two paralogs result in being significant from two different analysis that base their score on the targets' activity let use conclude that most reliably a large part of the targets of these (which due to their nature of being part of the same complex are shared) must be altered and further experiments could clarify which of the two TFs has the most impact on this effect.

Several isoforms of cJun and Fos are also reported to be highly upregulated from diffTF analysis: these two genes dimerize to form the AP1 TF complex which is normally expressed during neurogenesis and in the adult brain. c-Jun, a major component of the AP-1 transcription factor complex, is expressed during neurogenesis and in the adult brain. Its expression is highly induced in response to neuronal injury (Herdegen and Leah 1998). Neuronal injury triggers an axonal response which involves dramatic changes in the transcriptional program of the injured neuron, some of which are mediated by the induction of immediate-early transcription factors of the AP-1 family (Raivich et al. 2004). This response aims to induce and support axon elongation and achieve a reconstruction of effective synapses, a pathway that we report to be upregulated from our GSEA, bringing us to conclude that in Agc1 deficiency there is a broad response to the hypomyelination condition by the AP1 complex.

Along with these results we report a set of genes that found a strong reduction in promoter accessibility that are also reported to be significantly and strongly downregulated at the RNA levels: we identified 100 genes having LFCs < -0.5 and p-adjusted < 0.05 in both differential chromatin accessibility and differential gene expression, for which we also tested a GSEA on the most utilized curated pathway databases (KEGG, WikiPathways, Reactome and Hallmarks of Cancer). This analysis returned 0 significant results, as well as a Fisher test for the same pathways, which also returned 0 significant results, suggesting that this set of 100 genes is not enriched for any known pathway that is listed among the 4 databases.

We compared this set of genes with the results obtained from differential gene expression analysis of iPSCs cells and we found that only 40 out of 78 significantly detected genes in iPSCs were downregulated. This result becomes clearer in the light of the fact that there is a very low correlation between LFCs from the two different models (Fig. 7.1A).

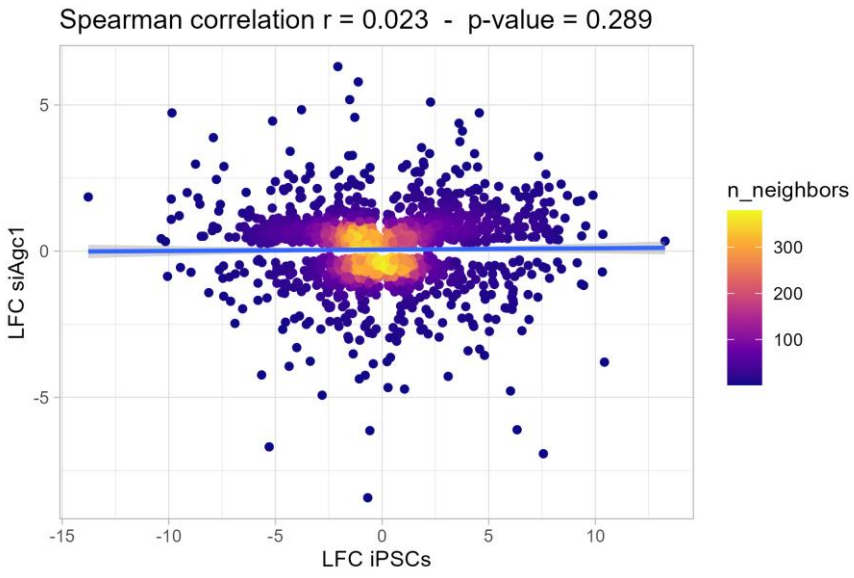

Figure 6.1A – Correlation between gene expression LFCs of iPSCs model and OliNeu model

This is of course a reflection of the very different biological nature of the two models, one coming from the genetic reprogramming and induction of human patients' dermal fibroblasts into neural progenitors, the other coming from RNA interference in mouse OPCs and, since our idea was to find a signature for Agc1 disease, more analysis at both gene expression and protein level could be useful to

identify genes that are conserved across the models that would allow us to identify specific markers for Agc1 deficiency. Having a large number of genes, or a signature, as a disease's marker would be useful for classification since it has recently been demonstrated that various psychiatric and neurodegenerative diseases exhibit shared epidemiologic risk with correlations between psychiatric and neurodegenerative diseases based on Linkage Disequilibrium regression on GWAS data, probably caused by a set of proteins and transcripts that show strong pleiotropic effect (Wingo et al. 2022).

A result that has been established by both models and has a relevant impact on the development of Agc1 deficiency is the loss of expression of Srebf1 and Srebf2, two genes encoding for SREBP, a TF promoting the expression of genes involved in the fatty-acid biosynthesis pathway, which are the building blocks of myelin sheaths. The downregulation of these TFs exacerbates a situation where the lack of Agc1 is already under-supplying OLs with enough acetyl groups. This result further supports why the current treatment for Agc1 deficiency is ketogenic diet, which is very high in fatty acids, which likely works as a supply of acetyl groups for the myelination process (Dahlin et al. 2015)

GSEA on iPSCs data also reports the enrichment of pathways related to neurite outgrowth, which goes well in accordance with the activation of AP1 targets genes, which gets expressed in order to promote neuron repairing mechanisms as we mentioned earlier. Another key result verified by iPSCs, is the downregulation in the EMT process, which supports the idea that cells are less migratory and are starting to become more differentiated, once again supporting the idea that the differentiation process in Agc1 deficiency patients has been disrupted.

To answer the reviewers' doubts about the possibility of using ATAC-seq data to determine candidate target genes for each tested regulator in master regulator analysis, it would be possible to use ATAC-seq data to better refine the process of building a gene regulatory network, but this would imply:

1. Having a sufficiently large ATAC-seq datasets in order to properly infer associations between regulators and targets
2. Taking into consideration that restricting the pool of a regulator's target genes to solely genes from an association map obtained from ATAC-seq

data could lead to miss the effect that second order targets have on the activity of the regulator.

What would be advisable is to leverage ATAC-seq data to determine a subset of targets whose expression data could be weighed more compared to genes that only show an association in expression data, without removing them.

Concluding, with this ensemble of works on Agc1's deficiency we gathered several information that we summarise below:

1. The global hypomyelination proposed by patient's diagnosis in several works is not only driven by the lack of Agc1, but also by the downregulation of two key regulators of fatty-acid biosynthesis, Srebf1 and Srebf2

2. Agc1 deficiency is a much complex disease from a regulatory point of view: several alterations in terms of chromatin accessibility are reported supporting recent studies (Poeta et al. 2022) that highlight how epigenetic alterations impact proliferation and differentiation, supported by the concordant results on activation of Smarcc1/Smarcc2, part of the npBAF chromatin remodelling complex.

Taken together, these results expand the actual knowledge of Agc1 as a disease mostly related to the metabolics of the cells, putting at the center the idea that the alteration is not just an unbalance in NAA levels and its consequences, but a rearrangement of regulatory networks inside affected cells, which would benefit of more studies that combine different methodologies to scout all the different faces of this disease.

8 References

“A Guide to Oligodendrocyte Markers.” n.d. Accessed October 14, 2023. https://www.biocompare.com/Editorial-Articles/590587-A-Guide-to-Oligodendrocyte-Markers/.

Abbas, Eman, Mohamed A. Hassan, Godwin Sokpor, Kamila Kiszka, Linh Pham, Cemil Kerimoglu, Andre Fischer, Huu Phuc Nguyen, Jochen F. Staiger, and Tran Tuoc. 2021. “Conditional Loss of BAF (mSWI/SNF) Scaffolding Subunits Affects Specification and Proliferation of Oligodendrocyte Precursors in Developing Mouse Forebrain.” *Frontiers in Cell and Developmental Biology* 9. https://www.frontiersin.org/articles/10.3389/fcell.2021.619538.

Abbott, Christopher, and Juan Bustillo. 2006. “What Have We Learned from Proton Magnetic Resonance Spectroscopy about Schizophrenia? A Critical Update.” *Current Opinion in Psychiatry* 19 (2): 135. https://doi.org/10.1097/01.yco.0000214337.29378.cd.

Aguet, François, Andrew A. Brown, Stephane E. Castel, Joe R. Davis, Yuan He, Brian Jo, Pejman Mohammadi, et al. 2017. “Genetic Effects on Gene Expression across Human Tissues.” *Nature* 550 (7675): 204–13. https://doi.org/10.1038/nature24277.

Alvarez, Mariano J., Yao Shen, Federico M. Giorgi, Alexander Lachmann, B. Belinda Ding, B. Hilda Ye, and Andrea Califano. 2016. “Functional Characterization of Somatic Mutations in Cancer Using Network-Based Inference of Protein Activity.” *Nature Genetics* 48 (8): 838–47. https://doi.org/10.1038/ng.3593.

Alvarez, Mariano J., Prem S. Subramaniam, Laura H. Tang, Adina Grunn, Mahalaxmi Aburi, Gabrielle Rieckhof, Elena V. Komissarova, et al. 2018. “A Precision Oncology Approach to the Pharmacological Targeting of Mechanistic Dependencies in Neuroendocrine Tumors.” *Nature Genetics* 50 (7): 979–89. https://doi.org/10.1038/s41588-018-0138-4.

Amemiya, Haley M., Anshul Kundaje, and Alan P. Boyle. 2019. “The ENCODE Blacklist: Identification of Problematic Regions of the Genome.” *Scientific Reports* 9 (1): 9354. https://doi.org/10.1038/s41598-019-45839-z.

Anders, Simon, Alejandro Reyes, and Wolfgang Huber. 2012. “Detecting Differential Usage of Exons from RNA-Seq Data.” *Genome Research* 22 (10): 2008–17. https://doi.org/10.1101/gr.133744.111.

Barretina, Jordi, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A. Margolin, Sungjoon Kim, Christopher J. Wilson, et al. 2012. “The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity.” *Nature* 483 (7391): 603–7. https://doi.org/10.1038/nature11003.

Berest, Ivan, Christian Arnold, Armando Reyes-Palomares, Giovanni Palla, Kasper Dindler Rasmussen, Holly Giles, Peter-Martin Bruch, et al. 2019. “Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF.” *Cell Reports* 29 (10): 3147-3159.e12. https://doi.org/10.1016/j.celrep.2019.10.106.

Berry, Kalen, Jiajia Wang, and Q. Richard Lu. 2020. “Epigenetic Regulation of Oligodendrocyte Myelination in Developmental Disorders and Neurodegenerative Diseases.” *F1000Research* 9 (February): F1000 Faculty Rev-105. https://doi.org/10.12688/f1000research.20904.1.

Bogner-Strauss, J.G. 2017. “N-Acetylaspartate Metabolism Outside the Brain: Lipogenesis, Histone Acetylation, and Cancer. Front Endocrinol (Lausanne.” https://doi.org/10.3389/fendo.2017.00240.

Bölsterli, Bigna K., Eugen Boltshauser, Luigi Palmieri, Johannes Spenger, Michaela Brunner-Krainz, Felix Distelmaier, Peter Freisinger, et al. 2022. “Ketogenic Diet Treatment of Defects in the Mitochondrial Malate Aspartate Shuttle and Pyruvate Carrier.” *Nutrients* 14 (17): 3605. https://doi.org/10.3390/nu14173605.

Brasset, E., and C. Vaury. 2005. “Insulators Are Fundamental Components of the Eukaryotic Genomes.” *Heredity* 94 (6): 571–76. https://doi.org/10.1038/sj.hdy.6800669.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. “Near-Optimal Probabilistic RNA-Seq Quantification.” *Nature Biotechnology* 34 (5): 525–27. https://doi.org/10.1038/nbt.3519.

Broyde, Joshua, David R. Simpson, Diana Murray, Evan O. Paull, Brennan W. Chu, Somnath Tagore, Sunny J. Jones, et al. 2021. “Oncoprotein-Specific Molecular Interaction Maps (SigMaps) for Cancer Network Analyses.” *Nature Biotechnology* 39 (2): 215–24. https://doi.org/10.1038/s41587-020-0652-7.

Buenrostro, Jason, Beijing Wu, Howard Chang, and William Greenleaf. 2015. “ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide.” *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* 109 (January): 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109.

Campagnoni, A. T. 1988. “Molecular Biology of Myelin Proteins from the Central Nervous System.” *Journal of Neurochemistry* 51 (1): 1–14. https://doi.org/10.1111/j.1471-4159.1988.tb04827.x.

Castelo-Branco, Gonçalo, Tobias Lilja, Karolina Wallenborg, Ana M. Falcão, Sueli C. Marques, Aileen Gracias, Derek Solum, et al. 2014. “Neural Stem Cell Differentiation Is Dictated by Distinct Actions of Nuclear Receptor Corepressors and Histone Deacetylases.” *Stem Cell Reports* 3 (3): 502–15. https://doi.org/10.1016/j.stemcr.2014.07.008.

Chang, Wenguang, and Junfang Teng. 2018. “Prox1 Is Essential for Oligodendrocyte Survival and Regulates Oligodendrocyte Apoptosis via the Regulation of NOXA.” *Acta Biochimica et Biophysica Sinica* 50 (7): 709–17. https://doi.org/10.1093/abbs/gmy061.

Chrast, Roman, Gesine Saher, Klaus-Armin Nave, and Mark H. G. Verheijen. 2011. “Lipid Metabolism in Myelinating Glial Cells: Lessons from Human Inherited Disorders and Mouse Models.” *Journal of Lipid Research* 52 (3): 419–34. https://doi.org/10.1194/jlr.R009761.

Clough, Emily, and Tanya Barrett. 2016. “The Gene Expression Omnibus Database.” *Methods in Molecular Biology (Clifton, N.J.)* 1418: 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.

Copray, Sjef, Jimmy Long Huynh, Falak Sher, Patrizia Casaccia-Bonnefil, and Erik Boddeke. 2009. “Epigenetic Mechanisms Facilitating Oligodendrocyte Development, Maturation, and Aging.” *Glia* 57 (15): 1579–87. https://doi.org/10.1002/glia.20881.

Coux, Rémi-Xavier, Nick D.L. Owens, and Pablo Navarro. n.d. “Chromatin Accessibility and Transcription Factor Binding through the Perspective of Mitosis.” *Transcription* 11 (5): 236–40. https://doi.org/10.1080/21541264.2020.1825907.

Criste, Gerson A., and Bruce D. Trapp. 2006. “N-Acetyl-L-Aspartate in Multiple Sclerosis.” In *N-Acetylaspartate*, edited by John R. Moffett, Suzannah B. Tieman, Daniel R. Weinberger, Joseph T. Coyle, and Aryan M. A. Namboodiri, 199–214. Advances in Experimental Medicine and Biology. Boston, MA: Springer US. https://doi.org/10.1007/0-387-30172-0_14.

Dahlin, Maria, Daniel A. Martin, Zandra Hedlund, Monica Jonsson, Ulrika von Döbeln, and Anna Wedell. 2015. “The Ketogenic Diet Compensates for AGC1 Deficiency and Improves Myelination.” *Epilepsia* 56 (11): e176–81. https://doi.org/10.1111/epi.13193.

Del Arco, A, M Agudo, and J Satrústegui. 2000. “Characterization of a Second Member of the Subfamily of Calcium-Binding Mitochondrial Carriers Expressed in Human Non-Excitable Tissues.” *Biochemical Journal* 345 (Pt 3): 725–32.

Delarasse, Cécile, Bruno Della Gaspera, Chuan Wei Lu, François Lachapelle, Antoinette Gelot, Diana Rodriguez, André Dautigny, Claude Genain, and Danielle Pham-Dinh. 2006. “Complex Alternative Splicing of the Myelin Oligodendrocyte Glycoprotein Gene Is Unique to Human and Non-Human Primates.” *Journal of Neurochemistry* 98 (6): 1707–17. https://doi.org/10.1111/j.1471-4159.2006.04053.x.

Dimas, P., L. Montani, and J.A. Pereira. 2019. “CNS Myelination and Remyelination Depend on Fatty Acid Synthesis by Oligodendrocytes.” *Elife* 8:e44702. https://doi.org/10.7554/eLife.44702.

Ding, Hongxu, Eugene F. Douglass, Adam M. Sonabend, Angeliki Mela, Sayantan Bose, Christian Gonzalez, Peter D. Canoll, Peter A. Sims, Mariano J. Alvarez, and Andrea Califano. 2018. “Quantitative Assessment of Protein Activity in Orphan Tissues and Single Cells Using the metaVIPER Algorithm.” *Nature Communications* 9 (1): 1471. https://doi.org/10.1038/s41467-018-03843-3.

D’Urso, Agustina, and Jason H. Brickner. 2014. “Mechanisms of Epigenetic Memory.” *Trends in Genetics : TIG* 30 (6): 230–36. https://doi.org/10.1016/j.tig.2014.04.004.

Emery, B., and Q.R. Lu. 2015. *Transcriptional and Epigenetic Regulation of Oligodendrocyte Development and Myelination in the Central Nervous System*. Cold Spring Harb Perspect Biol. https://doi.org/10.1101/cshperspect.a020461.

Falk, Marni J., Dong Li, Xiaowu Gai, Elizabeth McCormick, Emily Place, Francesco M. Lasorsa, Frederick G. Otieno, et al. 2014. “AGC1 Deficiency Causes Infantile Epilepsy, Abnormal Myelination, and Reduced N-Acetylaspartate.” In *JIMD Reports, Volume 14*, edited by Johannes Zschocke, K. Michael Gibson, Garry Brown, Eva Morava, and Verena Peters, 77–85. JIMD Reports. Berlin, Heidelberg: Springer. https://doi.org/10.1007/8904_2013_287.

Ferra, Francesca de, Helen Engh, Lynn Hudson, John Kamholz, Carmie Puckett, Susan Molineaux, and Robert A. Lazzarini. 1985. “Alternative Splicing Accounts for the Four Forms of Myelin Basic Protein.” *Cell* 43 (3): 721–27. https://doi.org/10.1016/0092-8674(85)90245-4.

Ganguly, R., and D. Guha. 2008. “Alteration of Brain Monoamines & EEG Wave Pattern in Rat Model of Alzheimer’s Disease & Protection by Moringa Oleifera.” *The Indian Journal of Medical Research* 128 (6): 744–51.

Grandi, Fiorella C., Hailey Modi, Lucas Kampman, and M. Ryan Corces. 2022. “Chromatin Accessibility Profiling by ATAC-Seq.” *Nature Protocols* 17 (6): 1518–52. https://doi.org/10.1038/s41596-022-00692-9.

Guan, Zhonghui, Maurizio Giustetto, Stavros Lomvardas, Joung-Hun Kim, Maria Concetta Miniaci, James H. Schwartz, Dimitris Thanos, and Eric R. Kandel. 2002. “Integration of Long-Term-Memory-Related Synaptic Plasticity Involves Bidirectional Regulation of Gene Expression and Chromatin Structure.” *Cell* 111 (4): 483–93. https://doi.org/10.1016/S0092-8674(02)01074-7.

Guttman, Mitchell, Ido Amit, Manuel Garber, Courtney French, Michael F. Lin, David Feldser, Maite Huarte, et al. 2009. “Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals.” *Nature* 458 (7235): 223–27. https://doi.org/10.1038/nature07672.

Heberle, Bernardo Aguzzoli, J. Anthony Brandon, Madeline L. Page, Kayla A. Nations, Ketsile I. Dikobe, Brendan J. White, Lacey A. Gordon, et al. 2023. “Using Deep Long-Read RNAseq in Alzheimer’s Disease Brain to Assess Clinical

Relevance of RNA Isoform Diversity." bioRxiv. https://doi.org/10.1101/2023.08.06.552162.

Herdegen, T., and J. D. Leah. 1998. "Inducible and Constitutive Transcription Factors in the Mammalian Nervous System: Control of Gene Expression by Jun, Fos and Krox, and CREB/ATF Proteins." *Brain Research. Brain Research Reviews* 28 (3): 370–490. https://doi.org/10.1016/s0165-0173(98)00018-6.

Hernandez, M., and P. Casaccia. 2015. "Interplay between Transcriptional Control and Chromatin Regulation in the Oligodendrocyte Lineage." *Glia* Aug;63(8):1357-75. https://doi.org/10.1002/glia.22818.

Ho, Lena, and Gerald R. Crabtree. 2010. "Chromatin Remodelling during Development." *Nature* 463 (7280): 474–84. https://doi.org/10.1038/nature08911.

Hughes, Ethan G., and Michael E. Stockton. 2021. "Premyelinating Oligodendrocytes: Mechanisms Underlying Cell Survival and Integration." *Frontiers in Cell and Developmental Biology* 9. https://www.frontiersin.org/articles/10.3389/fcell.2021.714169.

Jaegle, Martine, Mehrnaz Ghazvini, Wim Mandemakers, Marko Piirsoo, Siska Driegen, Françoise Levavasseur, Smiriti Raghoenath, Frank Grosveld, and Dies Meijer. 2003. "The POU Proteins Brn-2 and Oct-6 Share Important Functions in Schwann Cell Development." *Genes & Development* 17 (11): 1380–91. https://doi.org/10.1101/gad.258203.

Jakovcevski, Igor, and Nada Zecevic. 2005. "Olig Transcription Factors Are Expressed in Oligodendrocyte and Neuronal Cells in Human Fetal CNS." *The Journal of Neuroscience* 25 (44): 10064–73. https://doi.org/10.1523/JNEUROSCI.2324-05.2005.

Joubert, Lara, Isabelle Foucault, Yves Sagot, Lilia Bernasconi, François Duval, Chantal Alliod, Marie-José Frossard, et al. 2010. "Chemical Inducers and Transcriptional Markers of Oligodendrocyte Differentiation." *Journal of Neuroscience Research* 88 (12): 2546–57. https://doi.org/10.1002/jnr.22434.

Juaristi, Inés, María L. García-Martín, Tiago B. Rodrigues, Jorgina Satrústegui, Irene Llorente-Folch, and Beatriz Pardo. 2017. "ARALAR/AGC1

Deficiency, a Neurodevelopmental Disorder with Severe Impairment of Neuronal Mitochondrial Respiration, Does Not Produce a Primary Increase in Brain Lactate." *Journal of Neurochemistry* 142 (1): 132–39. https://doi.org/10.1111/jnc.14047.

Juliandi, B., M. Abematsu, and K. Nakashima. 2010. "Epigenetic regulation in neural stem cell differentiation." *Development, growth & differentiation*. https://doi.org/10.1111/j.1440-169X.2010.01175.x.

Kantarci, Kejal, and Clifford R. Jack. 2003. "Neuroimaging in Alzheimer Disease: An Evidence-Based Review." *Neuroimaging Clinics of North America* 13 (2): 197–209. https://doi.org/10.1016/s1052-5149(03)00025-x.

Kato, Kentaro, Daijiro Konno, Martin Berry, Fumio Matsuzaki, Ann Logan, and Alicia Hidalgo. 2015. "Prox1 Inhibits Proliferation and Is Required for Differentiation of the Oligodendrocyte Cell Lineage in the Mouse." *PLOS ONE* 10 (12): e0145334. https://doi.org/10.1371/journal.pone.0145334.

Kim, J. K., M. Samaranayake, and S. Pradhan. 2009. "Epigenetic Mechanisms in Mammals." *Cellular and Molecular Life Sciences* 66 (4): 596–612. https://doi.org/10.1007/s00018-008-8432-4.

Kim, Sun young, Eve E. Kelland, Ji hong Kim, Brett T. Lund, Xiao Chang, Kai Wang, and Leslie P. Weiner. 2017. "The Influence of Retinoic Acid on the Human Oligodendrocyte Precursor Cells by RNA-Sequencing." *Biochemistry and Biophysics Reports* 9 (March): 166–72. https://doi.org/10.1016/j.bbrep.2016.12.004.

Kuboyama, Kazuya, Akihiro Fujikawa, Ryoko Suzuki, Naomi Tanga, and Masaharu Noda. 2016. "Role of Chondroitin Sulfate (CS) Modification in the Regulation of Protein-Tyrosine Phosphatase Receptor Type Z (PTPRZ) Activity." *The Journal of Biological Chemistry* 291 (35): 18117–28. https://doi.org/10.1074/jbc.M116.742536.

Kulakovskiy, Ivan V, Ilya E Vorontsov, Ivan S Yevshin, Ruslan N Sharipov, Alla D Fedorova, Eugene I Rumynskiy, Yulia A Medvedeva, et al. 2018. "HOCOMOCO: Towards a Complete Collection of Transcription Factor Binding Models for Human and Mouse via Large-Scale ChIP-Seq Analysis." *Nucleic Acids Research* 46 (Database issue): D252–59. https://doi.org/10.1093/nar/gkx1106.

Lachmann, Alexander, Federico M. Giorgi, Gonzalo Lopez, and Andrea Califano. 2016. “ARACNe-AP: Gene Network Reverse Engineering through Adaptive Partitioning Inference of Mutual Information.” *Bioinformatics (Oxford, England)* 32 (14): 2233–35. https://doi.org/10.1093/bioinformatics/btw216.

Latasa, Maria-Jesus, Mario Ituero, Alejandra Moran-Gonzalez, Ana Aranda, and Jose Miguel Cosgaya. 2010. “Retinoic Acid Regulates Myelin Formation in the Peripheral Nervous System.” *Glia* 58 (12): 1451–64. https://doi.org/10.1002/glia.21020.

Lee, Xinhua, Yinghui Hu, Yiping Zhang, Zhongshu Yang, Zhaohui Shao, Mengsheng Qiu, Blake Pepinsky, Robert H. Miller, and Sha Mi. 2011. “Oligodendrocyte Differentiation and Myelination Defects in OMgp Null Mice.” *Molecular and Cellular Neuroscience* 46 (4): 752–61. https://doi.org/10.1016/j.mcn.2011.02.008.

Li, Jun, Brett Parker, Colin Martyn, Chandramohan Natarajan, and Jiasong Guo. 2013. “The PMP22 Gene and Its Related Diseases.” *Molecular Neurobiology* 47 (2): 673–98. https://doi.org/10.1007/s12035-012-8370-x.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Liu, Aixiao, Yu R. Han, Jiadong Li, Dongming Sun, Ming Ouyang, Mark R. Plummer, and Patrizia Casaccia-Bonnefil. 2007. “The Glial or Neuronal Fate Choice of Oligodendrocyte Progenitors Is Modulated by Their Ability to Acquire an Epigenetic Memory.” *Journal of Neuroscience* 27 (27): 7339–43. https://doi.org/10.1523/JNEUROSCI.1226-07.2007.

Liu, Jia, and Patrizia Casaccia. 2010. “Epigenetic Regulation of Oligodendrocyte Identity.” *Trends in Neurosciences* 33 (4): 193–201. https://doi.org/10.1016/j.tins.2010.01.007.

Llorente-Folch, Irene, Ignasi Sahún, Laura Contreras, María José Casarejos, Josep María Grau, Takeyori Saheki, María Angeles Mena, Jorgina Satrústegui, Mara Dierssen, and Beatriz Pardo. 2013. “AGC1-Malate Aspartate Shuttle Activity Is

Critical for Dopamine Handling in the Nigrostriatal Pathway.” *Journal of Neurochemistry* 124 (3): 347–62. https://doi.org/10.1111/jnc.12096.

Long, P.M., Namboodiri Moffett, JR, Viapiano AMA, Lawler MS, Jaworski SE, and D.M. 2013. “N-Acetylaspartate (NAA) and N-Acetylaspartylglutamate (NAAG) Promote Growth and Inhibit Differentiation of Glioma Stem-like Cells.” *The Journal of Biological Chemistry*, September. https://doi.org/10.1074/jbc.M113.487553.

Lorenzini, Luca, Mercedes Fernandez, Vito Antonio Baldassarro, Andrea Bighinati, Alessandro Giuliani, Laura Calzà, and Luciana Giardino. 2020. “White Matter and Neuroprotection in Alzheimer’s Dementia.” *Molecules* 25 (3): 503. https://doi.org/10.3390/molecules25030503.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Luo, Liheng, Michael Gribskov, and Sufang Wang. 2022. “Bibliometric Review of ATAC-Seq and Its Application in Gene Expression.” *Briefings in Bioinformatics* 23 (3): bbac061. https://doi.org/10.1093/bib/bbac061.

Margolin, Adam A, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. 2006. “ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context.” *BMC Bioinformatics* 7 (S1): S7. https://doi.org/10.1186/1471-2105-7-S1-S7.

Martinowich, Keri, Daisuke Hattori, Hao Wu, Shaun Fouse, Fei He, Yan Hu, Guoping Fan, and Yi E. Sun. 2003. “DNA Methylation-Related Chromatin Remodeling in Activity-Dependent Bdnf Gene Regulation.” *Science* 302 (5646): 890–93. https://doi.org/10.1126/science.1090842.

Masato, Anna, Nicoletta Plotegher, Daniela Boassa, and Luigi Bubacco. 2019. “Impaired Dopamine Metabolism in Parkinson’s Disease Pathogenesis.” *Molecular Neurodegeneration* 14 (1): 35. https://doi.org/10.1186/s13024-019-0332-6.

Mercatelli, Daniele, Gonzalo Lopez-Garcia, and Federico M. Giorgi. 2020. “Corto: A Lightweight R Package for Gene Network Inference and Master Regulator

Analysis.” *bioRxiv*, February, 2020.02.10.942623. https://doi.org/10.1101/2020.02.10.942623.

Moffett, John R., Peethambaran Arun, Prasanth S. Ariyannur, and Aryan M. A. Namboodiri. 2013. “N-Acetylaspartate Reductions in Brain Injury: Impact on Post-Injury Neuroenergetics, Lipid Synthesis, and Protein Acetylation.” *Frontiers in Neuroenergetics* 5 (December). https://doi.org/10.3389/fnene.2013.00011.

Moffett, John R., Brian Ross, Peethambaran Arun, Chikkathur N. Madhavarao, and M. A. A. Namboodiri. 2007. “N-Acetylaspartate in the CNS: From Neurodiagnostics to Neurobiology.” *Progress in Neurobiology* 81 (2): 89–131. https://doi.org/10.1016/j.pneurobio.2006.12.003.

Monnerie, Hubert, Micah Romer, Brigid K. Jensen, John S. Millar, Kelly L. Jordan-Sciutto, Sangwon F. Kim, and Judith B. Grinspan. 2017. “Reduced Sterol Regulatory Element-Binding Protein (SREBP) Processing Through Site-1 Protease (S1P) Inhibition Alters Oligodendrocyte Differentiation in Vitro.” *Journal of Neurochemistry* 140 (1): 53–67. https://doi.org/10.1111/jnc.13721.

Naef, Roland, and Ueli Suter. 1998. “Many Facets of the Peripheral Myelin Protein PMP22 in Myelination and Disease.” *Microscopy Research and Technique* 41 (5): 359–71. https://doi.org/10.1002/(SICI)1097-0029(19980601)41:5<359::AID-JEMT3>3.0.CO;2-L.

Narayanan, Ramanathan, Mehdi Pirouz, Cemil Kerimoglu, Linh Pham, Robin J. Wagener, Kamila A. Kiszka, Joachim Rosenbusch, et al. 2015. “Loss of BAF (mSWI/SNF) Complexes Causes Global Transcriptional and Chromatin State Changes in Forebrain Development.” *Cell Reports* 13 (9): 1842–54. https://doi.org/10.1016/j.celrep.2015.10.046.

Nasrabady, Sara E., Batool Rizvi, James E. Goldman, and Adam M. Brickman. 2018. “White Matter Changes in Alzheimer’s Disease: A Focus on Myelin and Oligodendrocytes.” *Acta Neuropathologica Communications* 6 (1): 22. https://doi.org/10.1186/s40478-018-0515-3.

Newman, Elizabeth A., Dong Won Kim, Jun Wan, Jie Wang, Jiang Qian, and Seth Blackshaw. 2018. “Foxd1 Is Required for Terminal Differentiation of Anterior

Hypothalamic Neuronal Subtypes." *Developmental Biology* 439 (2): 102–11. https://doi.org/10.1016/j.ydbio.2018.04.012.

Nielsen, Joseph A., Jo Ann Berndt, Lynn D. Hudson, and Regina C. Armstrong. 2004. "Myelin Transcription Factor 1 (Myt1) Modulates the Proliferation and Differentiation of Oligodendrocyte Lineage Cells." *Molecular and Cellular Neurosciences* 25 (1): 111–23. https://doi.org/10.1016/j.mcn.2003.10.001.

Noll, Elizabeth, and Robert H. Miller. 1994. "Regulation of Oligodendrocyte Differentiation: A Role for Retinoic Acid in the Spinal Cord." *Development* 120 (3): 649–60. https://doi.org/10.1242/dev.120.3.649.

Obradovic, Aleksandar, Lorenzo Tomassoni, Daoqi Yu, Kristina Guillan, Katie Souto, Elise Fraser, Susan Bates, et al. 2022. "Case Study of Single-Cell Protein Activity Based Drug Prediction for Precision Treatment of Cholangiocarcinoma." Preprint. Cancer Biology. https://doi.org/10.1101/2022.02.28.482410.

"Overcoming Gaps in the Treatment of Neurodegenerative Disease." 2020. *eBioMedicine* 60 (October). https://doi.org/10.1016/j.ebiom.2020.103088.

Pan, Qun, Ofer Shai, Leo J. Lee, Brendan J. Frey, and Benjamin J. Blencowe. 2008. "Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing." *Nature Genetics* 40 (12): 1413–15. https://doi.org/10.1038/ng.259.

Pardo, Beatriz, Eduardo Herrada-Soler, Jorgina Satrústegui, Laura Contreras, and Araceli del Arco. 2022. "AGC1 Deficiency: Pathology and Molecular and Cellular Mechanisms of the Disease." *International Journal of Molecular Sciences* 23 (1): 528. https://doi.org/10.3390/ijms23010528.

Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford. 2017. "Salmon: Fast and Bias-Aware Quantification of Transcript Expression Using Dual-Phase Inference." *Nature Methods* 14 (4): 417–19. https://doi.org/10.1038/nmeth.4197.

Petralla, Sabrina, Luis Emiliano Peña-Altamira, Eleonora Poeta, Francesca Massenzio, Marco Virgili, Simona Nicole Barile, Luigi Sbano, et al. 2019. "Deficiency of Mitochondrial Aspartate-Glutamate Carrier 1 Leads to Oligodendrocyte Precursor

Cell Proliferation Defects Both In Vitro and In Vivo." *International Journal of Molecular Sciences* 20 (18): 4486. https://doi.org/10.3390/ijms20184486.

Pfeiffer, Brian, Kuntal Sen, Shagun Kaur, and Kara Pappas. 2020. "Expanding Phenotypic Spectrum of Cerebral Aspartate–Glutamate Carrier Isoform 1 (AGC1) Deficiency." *Neuropediatrics* 51 (02): 160–63. https://doi.org/10.1055/s-0039-3400976.

Phelan, M. L., S. Sif, G. J. Narlikar, and R. E. Kingston. 1999. "Reconstitution of a Core Chromatin Remodeling Complex from SWI/SNF Subunits." *Molecular Cell* 3 (2): 247–53. https://doi.org/10.1016/s1097-2765(00)80315-9.

Pliss, Lioudmila, Kathryn A. Hausknecht, Michal K. Stachowiak, Cynthia A. Dlugos, Jerry B. Richards, and Mulchand S. Patel. 2013. "Cerebral Developmental Abnormalities in a Mouse with Systemic Pyruvate Dehydrogenase Deficiency." *PLOS ONE* 8 (6): e67473. https://doi.org/10.1371/journal.pone.0067473.

Poeta, Eleonora, Sabrina Petralla, Giorgia Babini, Brunaldo Renzi, Luigi Celauro, Maria Chiara Magnifico, Simona Nicole Barile, et al. 2022. "Histone Acetylation Defects in Brain Precursor Cells: A Potential Pathogenic Mechanism Causing Proliferation and Differentiation Dysfunctions in Mitochondrial Aspartate-Glutamate Carrier Isoform 1 Deficiency." *Frontiers in Cellular Neuroscience* 15 (January): 773709. https://doi.org/10.3389/fncel.2021.773709.

Porstmann, Thomas, Beatrice Griffiths, Yuen-Li Chung, Oona Delpuech, John R. Griffiths, Julian Downward, and Almut Schulze. 2005. "PKB/Akt Induces Transcription of Enzymes Involved in Cholesterol and Fatty Acid Biosynthesis via Activation of SREBP." *Oncogene* 24 (43): 6465–81. https://doi.org/10.1038/sj.onc.1208802.

Pozniak, Christine D., Abraham J. Langseth, Gerrit J. P. Dijkgraaf, Youngshik Choe, Zena Werb, and Samuel J. Pleasure. 2010. "Sox10 Directs Neural Stem Cells toward the Oligodendrocyte Lineage by Decreasing Suppressor of Fused Expression." *Proceedings of the National Academy of Sciences of the United States of America* 107 (50): 21795–800. https://doi.org/10.1073/pnas.1016485107.

Profilo, E., L.E. Peña-Altamira, and M. Corricelli. 2017. "Down-Regulation of the Mitochondrial Aspartate-Glutamate Carrier Isoform 1 AGC1 Inhibits Proliferation

and N-Acetylaspartate Synthesis in Neuro2A Cells.” *Biochimica et Biophysica Acta. Molecular Basis of Disease*, June. https://doi.org/10.1016/j.bbadis.2017.02.022.

Pu, Annie, Erin L. Stephenson, and V. Wee Yong. 2018. “The Extracellular Matrix: Focus on Oligodendrocyte Biology and Targeting CSPGs for Remyelination Therapies.” *Glia* 66 (9): 1809–25. https://doi.org/10.1002/glia.23333.

Raivich, Gennadij, Marion Bohatschek, Clive Da Costa, Osuke Iwata, Matthias Galiano, Maria Hristova, Abdolrahman S. Nateri, et al. 2004. “The AP-1 Transcription Factor c-Jun Is Required for Efficient Axonal Regeneration.” *Neuron* 43 (1): 57–67. https://doi.org/10.1016/j.neuron.2004.06.005.

Ramos, Milagros, Beatriz Pardo, Irene Llorente-Folch, Takeyori Saheki, Araceli del Arco, and Jorgina Satrústegui. 2011. “Deficiency of the Mitochondrial Transporter of Aspartate/Glutamate Aralar/AGC1 Causes Hypomyelination and Neuronal Defects Unrelated to Myelin Deficits in Mouse Brain.” *Journal of Neuroscience Research* 89 (12): 2008–17. https://doi.org/10.1002/jnr.22639.

Rinaldi, Berardo, Roberta Villa, Alessandra Sironi, Livia Garavelli, Palma Finelli, and Maria Francesca Bedeschi. 2022. “Smith-Magenis Syndrome—Clinical Review, Biological Background and Related Disorders.” *Genes* 13 (2): 335. https://doi.org/10.3390/genes13020335.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. “Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47. https://doi.org/10.1093/nar/gkv007.

Rive Le Gouard, Nicolas, Romain Nicolle, Mathilde Lefebvre, Antoinette Gelot, Solveig Heide, Anna Gerasimenko, Romulus Grigorescu, et al. 2023. “First Reports of Fetal SMARCC1 Related Hydrocephalus.” *European Journal of Medical Genetics* 66 (8): 104797. https://doi.org/10.1016/j.ejmg.2023.104797.

Rylski, Marcin, Renata Amborska, Katarzyna Zybura, Piotr Michaluk, Beata Bielinska, Filip A. Konopacki, Grzegorz M. Wilczynski, and Leszek Kaczmarek. 2009. “JunB Is a Repressor of MMP-9 Transcription in Depolarized Rat Brain Neurons.” *Molecular and Cellular Neuroscience* 40 (1): 98–110. https://doi.org/10.1016/j.mcn.2008.09.005.

Saher, Gesine, and Mikael Simons. 2010. “Cholesterol and Myelin Biogenesis.” In *Cholesterol Binding and Cholesterol Transport Proteins: Structure and Function in Health and Disease*, edited by J. Robin Harris, 489–508. Subcellular Biochemistry. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-8622-8_18.

Sanna, Pietro Paolo, Chiara Cabrelle, Tomoya Kawamura, Daniele Mercatelli, Nathan O’Connor, Amanda J. Roberts, Vez Repunte-Canonigo, and Federico M. Giorgi. 2023. “A History of Repeated Alcohol Intoxication Promotes Cognitive Impairment and Gene Expression Signatures of Disease Progression in the 3xTg Mouse Model of Alzheimer’s Disease.” *eNeuro* 10 (7): ENEURO.0456-22.2023. https://doi.org/10.1523/ENEURO.0456-22.2023.

Sartorelli, Vittorio, and Shannon M. Lauberth. 2020. “Enhancer RNAs Are an Important Regulatory Layer of the Epigenome.” *Nature Structural & Molecular Biology* 27 (6): 521–28. https://doi.org/10.1038/s41594-020-0446-0.

Schaum, Nicholas, Jim Karkanias, Norma F. Neff, Andrew P. May, Stephen R. Quake, Tony Wyss-Coray, Spyros Darmanis, et al. 2018. “Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris.” *Nature* 562 (7727): 367–72. https://doi.org/10.1038/s41586-018-0590-4.

Srivastava, Avi, Hirak Sarkar, Nitish Gupta, and Rob Patro. 2016. “RapMap: A Rapid, Sensitive and Accurate Tool for Mapping RNA-Seq Reads to Transcriptomes.” *Bioinformatics* 32 (12): i192–200. https://doi.org/10.1093/bioinformatics/btw277.

Tabula Sapiens Consortium*, Robert C. Jones, Jim Karkanias, Mark A. Krasnow, Angela Oliveira Pisco, Stephen R. Quake, Julia Salzman, et al. 2022. “The Tabula Sapiens: A Multiple-Organ, Single-Cell Transcriptomic Atlas of Humans.” *Science (New York, N.Y.)* 376 (6594): eabl4896. https://doi.org/10.1126/science.abl4896.

Targett, M.P., J. Sussman, N. Scolding, M.T. O’Leary, D.A. Compston, and W.F. Blakemore. 1996. “Failure to Achieve Remyelination of Demyelinated Rat Axons Following Transplantation of Glial Cells Obtained from the Adult Human Brain.” *Neuropathology and Applied Neurobiology* Jun;22(3):199-206. PMID: 8804021.

Taylor, D. L., S. E. Davies, T. P. Obrenovitch, M. H. Doheny, P. N. Patsalos, J. B. Clark, and L. Symon. 1995. “Investigation into the Role of N-Acetylaspartate in Cerebral Osmoregulation.” *Journal of Neurochemistry* 65 (1): 275–81. https://doi.org/10.1046/j.1471-4159.1995.65010275.x.

Taylor, Eric B. 2017. “Functional Properties of the Mitochondrial Carrier System.” *Trends in Cell Biology* 27 (9): 633–44. https://doi.org/10.1016/j.tcb.2017.04.004.

Tsompana, Maria, and Michael J. Buck. 2014. “Chromatin Accessibility: A Window into the Genome.” *Epigenetics & Chromatin* 7 (1): 33. https://doi.org/10.1186/1756-8935-7-33.

Turnescu, Tanja, Juliane Arter, Simone Reiprich, Ernst R. Tamm, Ari Waisman, and Michael Wegner. 2018. “Sox8 and Sox10 Jointly Maintain Myelin Gene Expression in Oligodendrocytes.” *Glia* 66 (2): 279–94. https://doi.org/10.1002/glia.23242.

Vaquero-Garcia, Jorge, Joseph K Aicher, Paul Jewell, Matthew R Gazzara, Caleb Matthew Radens, Anupama Jha, Christopher J Green, et al. 2021. “RNA Splicing Analysis Using Heterogeneous and Large RNA-Seq Datasets.” Preprint. Bioinformatics. https://doi.org/10.1101/2021.11.03.467086.

Vaquero-Garcia, Jorge, Alejandro Barrera, Matthew R Gazzara, Juan González-Vallinas, Nicholas F Lahens, John B Hogenesch, Kristen W Lynch, and Yoseph Barash. 2016. “A New View of Transcriptome Complexity and Regulation through the Lens of Local Splicing Variations.” Edited by Juan Valcárcel. *eLife* 5 (February): e11752. https://doi.org/10.7554/eLife.11752.

Wang, David S., Xingyao Wu, Yunhong Bai, Craig Zaidman, Tiffany Grider, John Kamholz, James R. Lupski, Anne M. Connolly, and Michael E. Shy. 2017. “PMP22 Exon 4 Deletion Causes ER Retention of PMP22 and a Gain-of-Function Allele in CMT1E.” *Annals of Clinical and Translational Neurology* 4 (4): 236–45. https://doi.org/10.1002/acn3.395.

Warepam, Marina, Awdhesh Kumar Mishra, Gurumayum Suraj Sharma, Kritika Kumari, Snigdha Krishna, Mohd Sajjad Ahmad Khan, Hamidur Rahman, and Laishram Rajendrakumar Singh. 2021. “Brain Metabolite, N-Acetylaspartate Is a

Potent Protein Aggregation Inhibitor." *Frontiers in Cellular Neuroscience* 15. https://www.frontiersin.org/articles/10.3389/fncel.2021.617308.

Wibom, Rolf, Francesco M. Lasorsa, Virpi Töhönen, Michela Barbaro, Fredrik H. Sterky, Thomas Kucinski, Karin Naess, et al. 2009. "AGC1 Deficiency Associated with Global Cerebral Hypomyelination." *New England Journal of Medicine* 361 (5): 489–95. https://doi.org/10.1056/NEJMoa0900591.

Wingo, Thomas S., Yue Liu, Ekaterina S. Gerasimov, Selina M. Vattathil, Meghan E. Wynne, Jiaqi Liu, Adriana Lori, et al. 2022. "Shared Mechanisms across the Major Psychiatric and Neurodegenerative Diseases." *Nature Communications* 13 (1): 4314. https://doi.org/10.1038/s41467-022-31873-5.

Yeung, Maggie S. Y., Sofia Zdunek, Olaf Bergmann, Samuel Bernard, Mehran Salehpour, Kanar Alkass, Shira Perl, et al. 2014. "Dynamics of Oligodendrocyte Generation and Myelination in the Human Brain." *Cell* 159 (4): 766–74. https://doi.org/10.1016/j.cell.2014.10.011.

Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology* 9 (9): R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhu, Anqi, Joseph G Ibrahim, and Michael I Love. 2019. "Heavy-Tailed Prior Distributions for Sequence Count Data: Removing the Noise and Preserving Large Differences." *Bioinformatics* 35 (12): 2084–92. https://doi.org/10.1093/bioinformatics/bty895.