Modern human-specific changes in regulatory regions implicated in cortical development

Juan Moriano[1,2] and Cedric Boeckx[1,2,3,*]

[1]Universitat de Barcelona

[2]Universitat de Barcelona Institute of Complex Systems (UBICS)

[3]Catalan Institution for Research and Advanced Studies (ICREA)

[*]Corresponding author: cedric.boeckx@ub.edu

July 24, 2019

Abstract

Recent paleogenomic studies have highlighted a very small set of proteins carrying modern human-specific missense changes in comparison to our closest extinct relatives. Despite being frequently alluded to as highly relevant, species-specific differences in regulatory regions remain understudied. Here, we integrate data from paleogenomics, chromatin modification and physical interaction, and single-cell gene expression of neural progenitor cells to report a set of genes whose enhancers and/or promoters harbor modern human-specific single-nucleotide changes and do not contain Neanderthal/Denisovan-specific changes. These regulatory regions exert their regulatory functions at early stages of cortical development and some of the genes controlled by them figure prominently in the neurodevelopmental/neuropsychiatric disease literature. In addition, we identified a substantial proportion of genes related to chromatin regulation and specifically an enrichment for the SETD1A/HMT complex, known to regulate the generation and proliferation of intermediate progenitor cells.

1 Introduction

Progress in the field of paleogenomics has allowed researchers to interrogate the genetic basis of modern human-specific traits in comparison to our closest extinct relatives, the Neanderthals and Denisovans [1]. One such trait concerns the period of growth and maturation of the brain, which is a major factor underlying the characteristic 'globular' head shape of modern humans [2]. Comparative genomic analyses using high-quality Neanderthal/Denisovan genomes [3–5] have revealed missense changes in the modern human lineage affecting proteins involved in the division of neural progenitor cells, key for the proper generation of neurons in an orderly spatiotemporal manner [4, 6]. But the total number of fixed missense changes amounts to less than one hundred proteins [1, 6]. This suggests that changes falling outside protein-coding regions may be equally relevant to understand the genetic basis of modern human-specific traits, as proposed more than four decades ago [7]. In this context it is noteworthy that human positively-selected genomic regions were found to be enriched in regulatory regions [8], and that signals of negative selection against Neanderthal DNA introgression were reported in promoters and conserved genomic regions [9]. Likewise, a fair amount of archaic-introgressed material falls within regulatory regions and is known to have an impact on gene expression in a tissue-specific way [10, 11].

In this work, we report a set of genes under the control of regulatory regions that harbor modern human-specific changes and are active at early stages of cortical development (Figure 1). We integrated data of chromatin immunoprecipitation and open chromatin regions identifying enhancers and promoters active during human cortical development, and the genes regulated by them as revealed by data of chromatin physical interaction, with paleogenomic data of single-nucleotide changes distinguishing modern humans and Neanderthal/Denisovan lineages. This allowed us to uncover those enhancer and promoters that harbor modern human-specific changes (at fixed or nearly fixed frequency in present-day human populations) and that, in addition, are depleted of Neanderthal/Denisovan-specific changes (*Methods* section 4.1). Next, we analysed single-cell gene expression data and performed co-expression network analysis to identify the genes under putative human-specific regulation within genetic networks in neural progenitor cells (*Methods* section 4.2

& 4.3). A substantial proportion of genes controlled by regulatory regions satisfying the aforementioned criteria are involved in chromatin regulation, and prominently among these, the SETD1A histone methyltransferase complex (SETD1A/HMT). This complex appears to have been targeted in modern human evolution and specifically regulates the indirect mode of neurogenesis through the control of WNT/β-CATENIN signaling (Figure 3).

2 Results

212 genes were found associated to regulatory regions active in the human developing cortex (from 5 to 20 post-conception weeks) that harbor human-specific single-nucleotide changes (hSNCs) and do not contain Neanderthal/Denisovan-specific changes (Suppl. Mat. Table S1 & S2). Among these, some well-studied disease-relevant genes stand out: *HTT* (Huntington disease) [12], *FOXP2* (language impairment) [13], *CHD8* and *CPEB4* (autism spectrum disorder) [14, 15], *TCF4* (Pitt-Hopkins syndrome and schizophrenia) [16, 17], *GLI3* (macrocephaly and Greig cephalopolysyndactyly syndrome) [18], *PHC1* (also known as primary, autosomal recessive, microcephaly-11) [19], *RCAN1* (Down syndrome) [20], and *DYNC1H1* (cortical malformations and microcephaly) [21].

Twelve out of the 212 genes contains fixed hSNCs in enhancers, with *LRRC23* having three such changes, and *GRIN2B, DDX12P* and *TFE3*, two each. Fourteen genes have fixed hSNCs in their promoters. Only one gene, *LRRC23*, exhibits fixed changes in both its enhancer and promoter regions. Applying a stricter filtering to the 212 genes dataset in order to select only those genes that are not regulated by any other enhancers with Neanderthal/Denisovan-specific changes, we obtained a reduced list of 118 genes (Suppl. Mat. Table S3), which includes seven genes harboring fixed hSNCs in enhancers (*NEUROD6, GRIN2B, LRRC23, RNF44, KCNA3, TCF25, TMLH3*), and nine genes, fixed hSNCs in promoters (*SETD1A, LRRC23, FOXJ2, LIMCH1, ZFAT, SPOP, DLGAP4, HS6ST2, UBE2A*).

No significant enrichment was found for the 212 gene list in Gene Ontology categories or KEGG pathways. However, the CORUM protein complexes database returns a significant enrichment for the SETD1A histone methyltransferase complex (hypergeometric test; FDR < 0.01). Our dataset

contains three members of that complex: SETD1A (promoter), ASH2L (enhancer) and WDR82 (enhancer). SETD1A, with a fixed hSNC, associates to the core of an H3K4 methyltransferase complex (ASH2L, WDR5, RBBP5, DPY30) and to WDR82, which recruits RNA polymerase II, to promote transcription of target genes through histone modification H3K4me3 [22]. Furthermore, the *SETD1A* promoter and the *WDR82* enhancer containing the relevant changes fall within positively-selected regions in the modern human lineage [8].

An enrichment was found for enhancers (permutation test; p-value < 0.05) and promoters (permutation test; p-value 10-4) overlapping with modern human positively-selected regions [8]. In addition, we found a significant over-representation of enhancers (permutation test; p-value < 0.05; while for promoter regions p-value < 0.08) for genetic loci associated to schizophrenia [23]. No significant overlap was found for enhancers/promoters and autism spectrum disorder risk variants ([24], retrieved from [25]) (Suppl. Fig. S1).

From 5 to 20 post-conception weeks, different types of cells populate the germinal zones of the human developing cortex (Figure 2). We analyzed gene expression data at single-cell resolution from a total of 762 cells (*Methods* section 4.2). We particularly focus on two progenitor cell-types—radial glial and intermediate progenitor cells (RGCs and IPCs, respectively)—since these are two of the main types of progenitor cells that give rise, in a orderly manner, to the neurons present in the adult brain (Figure 2). Two sub-populations of RGCs were identified (*PAX6*+/*EOMES*- cells), whereas three sub-populations of IPCs were detected (*EOMES*-expressing cells, with cells retaining *PAX6* expression and some expressing differentiation marker *TUJ1*), largely replicating what has been reported in the original publication for this dataset (Suppl. Fig. 2). We also performed a weighted gene co-expression analysis to assess the relevance of our 212 genes as part of genetic networks in the different sub-populations of progenitor cells (except for IPC sub-population 3, which was excluded due to the low number of cells).

SOX2 is a classical neural stem cell marker and we found that one of its related enhancers contains a high-frequency hSNC. *SOX2* is expressed at very early stages of neurodevelopment and is a key transcription factor for stem cell pluripotency [26]. SOX2 promotes neural stem cell maintenance by controlling long-range chromatin interactions [27]. Together with *SOX2* we find in

the same modules *SOCS3* (RGC-1 skyblue module), *JUN* and *JUNB* (IPC-2 red module), which are among the most highly downregulated genes in SOX2–deficient neural stem cells [27].

CHD8 (high-frequency hSNC in enhancer), a high risk candidate for autism spectrum disorder [14], is an ATP-dependent chromatin remodeler with highest expression between 9 and 16 post-conception weeks [14]. CHD8 controls neural progenitor cell proliferation through WNT-signaling related genes [28, 29]. CHD8 preferential binds to promoters of a conserved set of target genes [30]. In our dataset, *CHD8* is present in the RGC-2 turquoise module alongside with a considerable amount of known CHD8 targets in the human developing cortex [31]. Among these targets we find 10 genes in our 212 gene list (*CPEB4, KDM6A, FUT8, APOPT1, BCAR1, SPG7, ARID3A, ERF, RNF44, PROSC*; Suppl. Mat. Table S4 for the full list).

The RGC-2 turquoise module stands out because of the high number of genes related to neurodevelomental disorders that it contains. We found an enrichment (hypergeometric test; FDR < 0.01) for the following human phenotype ontology terms: Neurodevelopmental abnormality, abnormality of nervous system physiology and morphology, and neurodevelopmental delay (Suppl. Mat. Table S5). Along with *CHD8* and *CPEB4*, other autism-related genes (syndromic genes from the SFARI database) are found in this module, such as *CNTNAP2* and *FOXP1*. Another Forkhead box factor, FOXP2, is also in this module and harbors an almost fixed (>99%) hSNC in its promoter. FOXP2 is a highly conserved protein involved in language-related disorders whose evolutionary changes are particularly relevant for understanding human cognitive traits [32]. This hSNC (7:113727420) in the *FOXP2* promoter adds new evidence for a putative modern human-specific regulation of FOXP2 together with the nearly fixed intronic SNC that affects a transcription factor-binding site [32]. In the context of the RGC-2 turquoise module, it is worth noting the presence of *GLI3*, a gene that falls within a modern human positively-selected region [8] and whose promoter harbors a hSNC. GLI3 is a transcription factor primarily expressed in RGCs where it controls cell-cycle length through CDK6 regulation [33].

SETD1A (fixed hSNC in promoter), implicated in schizophrenia and developmental language impairment [34, 35], is expressed in RGC modules together with other genes implicated in proliferation and differentiation of progenitor cells such as lysine demethylase *KDM7* and transcription factors

SOX11 and *TEAD2* [36–38]. SETD1A regulates cell proliferation and differentiation in different cell-types, impacting H3K4 methylation levels and long-range chromatin interactions [39–41]. It has been shown that SETD1A promotes RGC proliferation through H3K4 promoter trimethylation via a mechanism dependent on the nucleosome assembly factor NAP1L1: suppression of NAP1L1 early on in embryonic cortex development decreases RGCs and causes premature differentiation resulting in more IPCs and more neurons [42] . In the IPC-2 midnightblue module, *SETDIA* is co-expressed with other SET-domain containing methyltransferases (*SETDB2, SETD2, SETD7*), core subunit of COMPASS complex *DPY30*, *β-CATENIN*, WNT-effector *TCF3*, and differentiation marker *NEUROD1*.

ASH2L (high-frequency hSNC in enhancer) is known to act mostly at advanced stages of neurogenesis and specifically controls the proliferation capacity of IPCs [43], a later stage of progenitor differentiation. In our dataset, ASH2L is found in the IPC-2 green module alongside neuronal differentiation marker *NEUROD6* (harboring a fixed and three other changes at $> 99\%$ frequency in an enhancer) and *GPM6A* (two high-frequency hSNC in an enhancer). The latter falls within a modern human selective sweep [8] and is involved in filopodia and neurite formation [44, 45]. Also found in this module is another gene associated with signatures of positive selection [8]: *DYNC1H1*, a cytoplasmic dynein involved in retrogade axonal transport and neuronal migration [46], and, as already mentioned above, implicated in malformations of cortical development and microcephaly [21]. The *DYNC1H1* promoter contains the highest number (five) of high-frequency hSNCs. Finally, *RCAN1* (two hSNCs in an enhancer), part of the Down Syndrome critical region [20], is also present in the IPC-2 green module. RCAN1 participates in progenitor cell proliferation and neuronal radial migration, and its mutation causes periventricular heterotopia [47].

3 Discussion

By integrating data from paleogenomics and chromatin interaction and modification, we identified a set of regulatory regions, active at early stages of cortical development, that harbor modern human-specific single-nucleotide changes and are depleted of Neanderthal/Denisovan-specific changes. Gene

expression data at single-cell resolution allowed us to identify genes regulated by these regions in genetic networks relevant for processes such as cell proliferation, differentiation or migration. These modern human-specific changes are new candidates to investigate species-specific differences during cortical development and complement previous studies that focused on protein-coding changes [4, 6].

A considerable amount of these changes affects chromatin regulators, most noteworthy the SETD1A/HMT complex. Regulators of chromatin dynamics are known to play key roles during cell-fate decisions through the control of specific transcriptional programs [48–50]. Both SETD1A and ASH2L, core components of the HMT complex, regulate WNT/β-CATENIN signaling [43, 51–53], which influences cell-fate decisions by promoting either self-maintenance or differentiation depending on the stage of progenitor differentiation (Figure 3).

SETD1A has been found to act in collaboration with a histone chaperone to promote proliferation of neural progenitor cells through H3K4 trimethylation at the promoter of β-CATENIN, while its knockdown causes reduction in proliferative neural progenitor cells and an increase in cells at the cortical plate [51]. In addition, one of SETD1A direct targets is the WNT-effector TCF4 [54], whose promoter also harbors a hSNC.

ASH2L has a more pronounced effect on the generation of upper-layer neurons during late neurogenesis [43]. Conditional knock-out of *ASH2L* early in neurodevelopment does not affect proliferation of progenitor cells, but significantly compromises the proliferative capacity of RGCs and IPCs at later stages, with these progenitor cells showing a marked reduction in H3K4me3 levels and subsequent downregulation of WNT/β-CATENIN signaling-related genes (defects that can be rescued by over-expression of β-CATENIN) [43].

Given that depletion of components of the SETD1A/HMT complex impairs the proliferative capacity of progenitor cells, and primarily at late stages of neurogenesis in the case of ASH2L, mutations in regulatory regions of these genes are expected to most likely affect the indirect mode of neurogenesis (RGC-IPC-neuron transition) and the generation of upper-layer cortical neurons.

Other chromatin regulators with human-specific changes in regulatory regions are also relevant for early progenitor cell-fate decisions. Another histone modifier, histone demethylase KDM6A, associates to the H3K4 methyltransferase complex [22] and harbors a hSNC in its promoter.

KDM6A regulates the *PTEN* promoter to enhance its expression, controlling in this way neural stem cells proliferation and differentiation (Lei and Jiao 2018). *PTEN* is essential for normal brain growth [55] and possesses an excess of hSNCs in modern human genomes in comparison to Neanderthal/Denisovan genomes [6].

ATP-dependent chromatin remodeler CHD2 has a hSNC in an enhancer and promotes RGC self-maintenance at the expense of IPC generation [56]. In addition, CHD2 has been related to epilepsy and other neurodevelopmental disorders [57, 58]. Intriguingly, the *SETD1A* promoter region with a fixed hSNC presents a binding motif for CHD2 (16:30969654 (hg19); UCSC Genome Browser). From the same family of proteins, CHD8, discussed above, has been found to associate with the HMT complex in a carcinoma cell-line [59]. Among the reported targets of CHD8 in the human developing brain [31], we found AKAP8, which harbors a high-frequency missense hSNC [6] and is involved in facilitation of protein kinase A-binding to its substrates, including the core subunit of the HMT complex DPY30 [60].

The dysfunction of chromatin regulators is among the most salient features behind causative mutations in neurodevelopmental disorders [61]. The modern human-specific changes reported here highlight chromatin modifiers and remodelers in the context of regulatory regions active at early stages of cortical development and further experimental elucidation will be required to assess their relevance in modern human evolution.

4 Methods

4.1 Data processing

Integration and processing of data from different sources was performed using IPython v5.7.0. We used publicly available data from [6] of single-nucleotide changes in modern humans (at fixed or above 90% frequency in present-day human populations, as defined by [6]) and Neanderthal/Denisovan-specific changes. For enhancer–promoter linkages, we used publicly available data from [62], based on transposase-accessible chromatin coupled to sequencing and chromatin capture via Hi-C, from 15 to 17 post-conception weeks of the human developing cortex. We additionally retrieved enhancer-

gene data of chromatin interaction via Hi-C of the adult prefrontal cortex from [63] (PsychENCODE resource portal: `http://resource.psychencode.org/`), filtered by signals of active enhancers (H3K27ac) that do not overlap with promoter signal (H3K4me3) from 7 to 12 post-conception weeks of the human developing cortex retrieved from [64]. Human positively-selected regions coordinates were retrieved from [8].

The coordinates (hg19 version) of the modern human-specific regulatory regions harboring hSNCs are available in the Supplementary Material Tables S1 & S2.

4.2 Single-cell RNA-seq analysis

The single-cell transcriptomic analysis was performed using the Seurat package v2.4 [65] in RStudio v1.1.463 (server mode).

Single-cell gene expression data was retrieved from [64] from PsychENCODE portal (`http://development.psychencode.org/#`). We used raw gene counts thresholding for cells with a minimum of 500 genes detected and for genes present at least in 10% of the total cells (n=762). Data was normalized using "LogNormalize" method with a scale factor of 1000000. We regressed out cell-to-cell variation due to mitochondrial and cell-cycle genes (*ScaleData* function). For the latter, we used a list of genes ([66]) that assigns genes to either G1/S or G2/M phase (function *CellCycleScoring*). We further filtered cells (*FilterCells* function) setting a low threshold of 2000 and a high threshold of 9000 gene counts per cell, and a high threshold of 5% of the total gene counts for mitochondrial genes.

We assigned the label 'highly variable' to genes whose average expression value was between 0.5 and 8, and variance-to-mean expression level ratio between 0.5 and 5 (*FindVariableGenes* function). We obtained a total of 4261 genes for this category. Next, we performed a principal component (PC) analysis on highly variable genes and determined significance by a jackStraw analysis (*JackStraw* function). We used the first most significant principal components (n = 13) for clustering analysis (*FindClusters* function; resolution = 3). Data was represented in two dimensions using t-distributed stochastic neighbor embedding (*RunTSNE* function). The resulting twelve clusters were plotted using *tSNEplot* function. Cell-type assignment was based on the metadata from the

original publication [64].

4.3 Weighted gene co-expression network analysis

For the gene co-expression network analysis we used the WGCNA R package [67, 68]. For each population of progenitor cells (RGC-1, 34 cells (15017 genes); RGC-2, 30 cells (14747 genes); IPC-1, 52 cells (15790 genes); IPC-2, 41 cells (15721 genes); IPC3 population was excluded due to low number of cells), log-transformed values of gene expression data were used as input for weighted gene co-expression network analysis. A soft threshold power was chosen (12, 12, 14, 12 for RGC-1, RGC-2, IPC-1, IPC-2 populations, with R2: 0.962, 0.817, 0.961, 0.918, respectively) and a bi-weight mid-correlation applied to compute a signed weighted adjacency matrix, transformed later into a topological overlap matrix. Module detection (minimum size 200 genes) was performed using function *cutreeDynamic* (method = 'hybrid', deepSplit = 2), getting a total of 32, 26, 9, 23 modules for RGC-1, RGC-2, IPC-1, IPC-2, respectively (Suppl. Fig. S3 & S4).

4.4 Enrichment analysis

The g:Profiler R package [69] was used to perform enrichment analyses (hypergeometric test) for gene/phenotype ontology categories, biological pathways (KEGG, Reactome) and protein databases (CORUM, Human Protein Atlas) for the gene lists generated in this study. Permutation tests (10,000 permutations) were performed to evaluate enrichment of enhancers/promoters regions in different genomic regions datasets using the R package regioneR [70].

Author contributions

Conceptualization: C.B. & J.M.; Data Curation: J.M.; Formal Analysis: J.M.; Funding Acquisition: C.B.; Investigation: C.B. & J.M.; Methodology: C.B. & J.M.; Software: J.M.; Supervision: C.B.; Visualization: C.B. & J.M.; Writing – Original Draft Preparation: C.B. & J.M.; Writing – Review & Editing: C.B. & J.M.

Funding

C.B. acknowledges research funds from the Spanish Ministry of Economy and Competitiveness/FEDER (grant FFI2016-78034-C2-1-P), Marie Curie International Reintegration Grant from the European Union (PIRG-GA-2009-256413), research funds from the Fundació Bosch i Gimpera, MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas 4903 (Evolinguistics: JP17H06379), and Generalitat de Catalunya (Government of Catalonia) – 2017-SGR-341.

Data availability

The data not present in the manuscript or in the supplementary material will be available in an online open access repository, as well as the code used to perform the analysis reported in this study.

Competing interests

Authors declare NO competing financial or non-financial interest.

References

[1] S. Pääbo, "The Human Condition—A Molecular Approach," *Cell*, vol. 157, pp. 216–226, Mar. 2014.

[2] J.-J. Hublin, *et al.*, "Brain ontogeny and life history in Pleistocene hominins," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, pp. 20140062–20140062, Jan. 2015.

[3] M. Meyer, *et al.*, "A High-Coverage Genome Sequence from an Archaic Denisovan Individual," *Science*, vol. 338, pp. 222–226, Oct. 2012.

[4] K. Prüfer, *et al.*, "The complete genome sequence of a Neanderthal from the Altai Mountains," *Nature*, vol. 505, pp. 43–49, Jan. 2014.

[5] K. Prüfer, *et al.*, "A high-coverage Neandertal genome from Vindija Cave in Croatia," *Science*, vol. 358, pp. 655–658, Nov. 2017.

[6] M. Kuhlwilm *et al.*, "A catalog of single nucleotide changes distinguishing modern humans from archaic hominins," *Scientific Reports*, vol. 9, Dec. 2019.

[7] M. King *et al.*, "Evolution at two levels in humans and chimpanzees," *Science*, vol. 188, pp. 107–116, Apr. 1975.

[8] S. Peyrégne, *et al.*, "Detecting ancient positive selection in humans using extended lineage sorting," *Genome Research*, vol. 27, pp. 1563–1572, Sept. 2017.

[9] M. Petr, *et al.*, "Limits of long-term selection against Neandertal introgression," *Proceedings of the National Academy of Sciences*, vol. 116, pp. 1639–1644, Jan. 2019.

[10] R. C. McCoy, *et al.*, "Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression," *Cell*, vol. 168, pp. 916–927.e12, Feb. 2017.

[11] M. Silvert, *et al.*, "Impact and Evolutionary Determinants of Neanderthal Introgression on Transcriptional and Post-Transcriptional Regulation," *The American Journal of Human Genetics*, vol. 104, pp. 1241–1250, June 2019.

[12] F. Saudou *et al.*, "The Biology of Huntingtin," *Neuron*, vol. 89, pp. 910–926, Mar. 2016.

[13] C. S. L. Lai, *et al.*, "A forkhead-domain gene is mutated in a severe speech and language disorder," *Nature*, vol. 413, pp. 519–523, Oct. 2001.

[14] R. Bernier, *et al.*, "Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development," *Cell*, vol. 158, pp. 263–276, July 2014.

[15] A. Parras, *et al.*, "Autism-like phenotype and risk gene mRNA deadenylation by CPEB4 missplicing," *Nature*, vol. 560, pp. 441–446, Aug. 2018.

[16] C. Zweier, *et al.*, "Haploinsufficiency of TCF4 Causes Syndromal Mental Retardation with Intermittent Hyperventilation (Pitt-Hopkins Syndrome)," *The American Journal of Human Genetics*, vol. 80, pp. 994–1001, May 2007.

[17] M. P. Forrest, *et al.*, "The Psychiatric Risk Gene Transcription Factor 4 (TCF4) Regulates Neurodevelopmental Pathways Associated With Schizophrenia, Autism, and Intellectual Disability," *Schizophrenia Bulletin*, vol. 44, pp. 1100–1110, Aug. 2018.

[18] M. Kalff-Suske, *et al.*, "Point Mutations Throughout the GLI3 Gene Cause Greig Cephalopolysyndactyly Syndrome," *Human Molecular Genetics*, vol. 8, pp. 1769–1777, Sept. 1999.

[19] S. Awad, *et al.*, "Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis," *Human Molecular Genetics*, vol. 22, pp. 2200–2213, June 2013.

[20] J.-J. Fuentes, *et al.*, "A new human gene from the Down syndrome critical region encodes a proline-rich protein highly expressed in fetal brain and heart," *Human Molecular Genetics*, vol. 4, no. 10, pp. 1935–1944, 1995.

[21] K. Poirier, *et al.*, "Mutations in TUBG1, DYNC1h1, KIF5c and KIF2a cause malformations of cortical development and microcephaly," *Nature Genetics*, vol. 45, pp. 639–647, June 2013.

[22] A. Piunti *et al.*, "Epigenetic balance of gene expression by Polycomb and COMPASS families," *Science*, vol. 352, p. aad9780, June 2016.

[23] Schizophrenia Working Group of the Psychiatric Genomics Consortium, "Biological insights from 108 schizophrenia-associated genetic loci," *Nature*, vol. 511, pp. 421–427, July 2014.

[24] Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, *et al.*, "Identification of common genetic risk variants for autism spectrum disorder," *Nature Genetics*, vol. 51, pp. 431–444, Mar. 2019.

[25] A. Buniello, *et al.*, "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019," *Nucleic Acids Research*, vol. 47, pp. D1005–D1012, Nov. 2018.

[26] S. Masui, *et al.*, "Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells," *Nature Cell Biology*, vol. 9, pp. 625–635, June 2007.

[27] J. A. Bertolini, *et al.*, "Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance," *Cell Stem Cell*, vol. 24, pp. 462–476.e6, Mar. 2019.

[28] A. Sugathan, *et al.*, "regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors," *Proceedings of the National Academy of Sciences*, vol. 111, pp. E4468–E4477, Oct. 2014.

[29] O. Durak, *et al.*, "Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling," *Nature Neuroscience*, vol. 19, pp. 1477–1488, Nov. 2016.

[30] A. A. Wade, *et al.*, "Common CHD8 Genomic Targets Contrast With Model-Specific Transcriptional Impacts of CHD8 Haploinsufficiency," *Frontiers in Molecular Neuroscience*, vol. 11, Jan. 2019.

[31] J. Cotney, *et al.*, "The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment," *Nature Communications*, vol. 6, May 2015.

[32] S. E. Fisher, "Human Genetics: The Evolving Story of FOXP2," *Current Biology*, vol. 29, pp. R65–R67, Jan. 2019.

[33] K. Hasenpusch-Theil, *et al.*, "Gli3 controls the onset of cortical neurogenesis by regulating the radial glial cell cycle through *Cdk6* expression," *Development*, vol. 145, p. dev163147, Sept. 2018.

[34] A. Takata, *et al.*, "Loss-of-Function Variants in Schizophrenia Risk and SETD1a as a Candidate Susceptibility Gene," *Neuron*, vol. 82, pp. 773–780, May 2014.

[35] E. Eising, *et al.*, "A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development," *Molecular Psychiatry*, vol. 24, pp. 1065–1078, July 2019.

[36] C. Huang, *et al.*, "Dual-specificity histone demethylase KIAA1718 (KDM7a) regulates neural differentiation through FGF4," *Cell Research*, vol. 20, pp. 154–165, Feb. 2010.

[37] C. Chen, *et al.*, "Orchestration of Neuronal Differentiation and Progenitor Pool Expansion in the Developing Cortex by SoxC Genes," *Journal of Neuroscience*, vol. 35, pp. 10629–10642, July 2015.

[38] K. Saito, *et al.*, "Neural Progenitor Cells Undergoing Yap/Tead-Mediated Enhanced Self-Renewal Form Heterotopias More Easily in the Diencephalon than in the Telencephalon," *Neurochemical Research*, vol. 43, pp. 180–189, Jan. 2018.

[39] B. K. Tusi, *et al.*, "*Setd1a* regulates progenitor B-cell-to-precursor B-cell development through histone H3 lysine 4 trimethylation and *Ig heavy-chain* rearrangement," *The FASEB Journal*, vol. 29, pp. 1505–1515, Apr. 2015.

[40] Y. Li, *et al.*, "Setd1a and NURF mediate chromatin dynamics and gene regulation during erythroid lineage commitment and differentiation," *Nucleic Acids Research*, pp. gkw327–, May 2016.

[41] A. S. Bledau, *et al.*, "The H3k4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation," *Development*, vol. 141, pp. 1022–1035, Mar. 2014.

[42] H. Qiao, *et al.*, "Nap1l1 Controls Embryonic Neural Progenitor Cell Proliferation and Differentiation in the Developing Brain," *Cell Reports*, vol. 22, pp. 2279–2293, Feb. 2018.

[43] L. Li, *et al.*, "The COMPASS Family Protein ASH2l Mediates Corticogenesis via Transcriptional Regulation of Wnt Signaling," *Cell Reports*, vol. 28, pp. 698–711.e5, July 2019.

[44] N. M. Rosas, *et al.*, "Alanine Scanning Mutagenesis of the C-Terminal Cytosolic End of Gpm6a Identifies Key Residues Essential for the Formation of Filopodia," *Frontiers in Molecular Neuroscience*, vol. 11, Sept. 2018.

[45] A. Honda, *et al.*, "Extracellular Signals Induce Glycoprotein M6a Clustering of Lipid Rafts and Associated Signaling Molecules," *The Journal of Neuroscience*, vol. 37, pp. 4046–4064, Apr. 2017.

[46] S. Sasaki, *et al.*, "A LIS1/NUDEL/Cytoplasmic Dynein Heavy Chain Complex in the Developing and Adult Nervous System," *Neuron*, vol. 28, pp. 681–696, Dec. 2000.

[47] Y. Li, *et al.*, "Rcan1 Deficiency Impairs Neuronal Migration and Causes Periventricular Heterotopia," *The Journal of Neuroscience*, vol. 35, pp. 610–620, Jan. 2015.

[48] T. Chen *et al.*, "Chromatin modifiers and remodellers: regulators of cellular differentiation," *Nature Reviews Genetics*, vol. 15, pp. 93–106, Feb. 2014.

[49] Y. Hirabayashi *et al.*, "Epigenetic control of neural precursor cell fate during development," *Nature Reviews Neuroscience*, vol. 11, pp. 377–388, June 2010.

[50] T. C. Tuoc, *et al.*, "Control of cerebral size and thickness," *Cellular and Molecular Life Sciences*, vol. 71, pp. 3199–3218, Sept. 2014.

[51] Y. Li *et al.*, "Histone chaperone HIRA regulates neural progenitor cell proliferation and neurogenesis via beta-catenin," *The Journal of Cell Biology*, vol. 216, pp. 1975–1992, July 2017.

[52] K. Hoffmeyer, *et al.*, "Wnt/ -Catenin Signaling Regulates Telomerase in Stem Cells and Cancer Cells," *Science*, vol. 336, pp. 1549–1554, June 2012.

[53] T. Salz, *et al.*, "hSETD1a Regulates Wnt Target Genes and Controls Tumor Growth of Colorectal Cancer Cells," *Cancer Res*, vol. 74, pp. 775–786, Feb. 2014.

[54] J. Mukai, "Recapitulation and reversal of schizophrenia-related phenotypes in Setd1a-deficient mice," *bioRxiv*, p. 120, 2019.

[55] Y. Chen, *et al.*, "Pten Mutations Alter Brain Growth Trajectory and Allocation of Cell Types through Elevated -Catenin Signaling," *Journal of Neuroscience*, vol. 35, pp. 10252–10267, July 2015.

[56] T. Shen, *et al.*, "CHD2 is Required for Embryonic Neurogenesis in the Developing Cerebral Cortex," *STEM CELLS*, vol. 33, no. 6, pp. 1794–1806, 2015.

[57] G. L. Carvill, *et al.*, "Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1," *Nature Genetics*, vol. 45, pp. 825–830, July 2013.

[58] S. Chénier, *et al.*, "CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioural problems," *Journal of Neurodevelopmental Disorders*, vol. 6, no. 1, p. 9, 2014.

[59] J. A. Yates, *et al.*, "Regulation of HOXA2 gene expression by the ATP-dependent chromatin remodeling enzyme CHD8," *FEBS Letters*, vol. 584, no. 4, pp. 689–693, 2010.

[60] A. Bieluszewska, *et al.*, "PKA-binding domain of AKAP8 is essential for direct interaction with DPY30 protein," *The FEBS Journal*, vol. 285, pp. 947–964, Mar. 2018.

[61] M. Gabriele, *et al.*, "The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 84, pp. 306 – 327, 2018.

[62] L. de la Torre-Ubieta, *et al.*, "The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis," *Cell*, vol. 172, pp. 289–304.e18, Jan. 2018.

[63] D. Wang, *et al.*, "Comprehensive functional genomic resource and integrative model for the human brain," *Science*, vol. 362, no. 6420, 2018.

[64] M. Li, *et al.*, "Integrative functional genomic analysis of human brain development and neuropsychiatric risks," *Science*, vol. 362, p. eaat7615, Dec. 2018.

[65] A. Butler, *et al.*, "Integrating single-cell transcriptomic data across different conditions, technologies, and species," *Nature Biotechnology*, vol. 36, p. 411, Apr. 2018.

[66] I. Tirosh, *et al.*, "Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq," *Science*, vol. 352, pp. 189–196, Apr. 2016.

[67] P. Langfelder *et al.*, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, p. 559, Dec. 2008.

[68] P. Langfelder *et al.*, "Fast *R* Functions for Robust Correlations and Hierarchical Clustering," *Journal of Statistical Software*, vol. 46, no. 11, 2012.

[69] J. Reimand, *et al.*, "g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments," *Nucleic Acids Research*, vol. 35, pp. W193–W200, July 2007.

[70] B. Gel, *et al.*, "regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests," *Bioinformatics*, p. btv562, Sept. 2015.

[71] E. Calo *et al.*, "Modification of Enhancer Chromatin: What, How, and Why?," *Molecular Cell*, vol. 49, pp. 825–837, Mar. 2013.

[72] M. Florio *et al.*, "Neural progenitors, neurogenesis and the evolution of the neocortex," *Development*, vol. 141, pp. 2182–2194, June 2014.

[73] A. Chenn *et al.*, "Regulation of Cerebral Cortical Size by Control of Cell Cycle Exit in Neural Precursors," *Science*, vol. 297, no. 5580, pp. 365–369, 2002.

[74] R. N. Munji, *et al.*, "Wnt Signaling Regulates Neuronal Differentiation of Cortical Intermediate Progenitors," *Journal of Neuroscience*, vol. 31, pp. 1676–1687, Feb. 2011.

[75] K. Draganova, *et al.*, "Wnt/beta-Catenin Signaling Regulates Sequential Fate Decisions of Murine Cortical Precursor Cells: Beta-Catenin Signaling Regulates Sequential Neural Fate," *STEM CELLS*, vol. 33, pp. 170–182, Jan. 2015.

[76] C. A. Mutch, *et al.*, "Beta-Catenin Signaling Negatively Regulates Intermediate Progenitor Population Numbers in the Developing Cortex," *PLoS ONE*, vol. 5, p. e12376, Aug. 2010.

Figures

Figure 1: **Regulatory regions characterized in this study**. Active enhancers are typically located in regions of open chromatin and nucleosomes in their vicinity are marked by histone modifications H3K27 acetylation and H3K4 mono-methylation. By contrast, H3K4 trimethylation defines active promoters [71]. We considered signals of active enhancers and promoters in the human developing brain (from 5 to 20 post-conception weeks) that harbor modern human-specific single-nucleotide changes (SNC) filtering out those regulatory regions that also contain Neanderthal/Denisovan-specific changes. Chromosome conformation capture (Hi-C) data revealed the genes controlled by these regulatory regions.

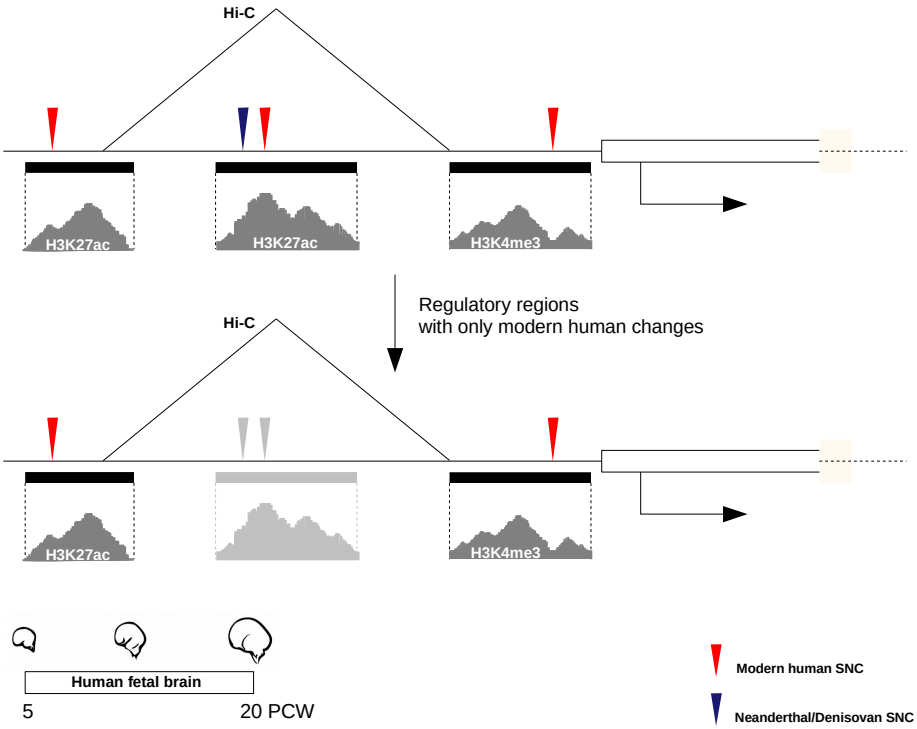


Figure 2: **Cell-type populations at early stages of cortical development**. (a) Apical radial glial cells (RGCs) populate the ventricular zone and prolong one process apically to the ventricular surface and another one to the basal lamina, which serves as a scaffold for neuronal migration. RGCs also proliferate and differentiate to give rise to another RGC, a basal intermediate progenitor (indirect neurogenesis), or a neuron (direct neurogenesis) [72]. Intermediate progenitor cells (IPCs) are basal progenitors lacking of apical-basal cell polarity. IPCs migrate to the subventricular zone and, after a couple of self-renewal divisions, differentiate to give rise to two neurons [72]. (b) The tSNE plot shows twelve clusters detected analyzing a total of 762 cells. (c) The violin plots show expression of two markers (*PAX6, EOMES*) across the different clusters, distinguishing between RGCs and IPCs. (d) The miniature tSNE plots show the distribution across the sub-populations of a selection of genes discussed in the main text.
IPC: Intermediate progenitor cells; NascentN: Nascent neurons; ExN: Excitatory neurons; Astro: Astrocytes; RGC: Radial glial cells; InN: Interneurons; Oligodendrocyte progenitor cells: OPC; Oligo: Oligodendrocytes; Cl_12: Cluster 12 (unidentified cells).

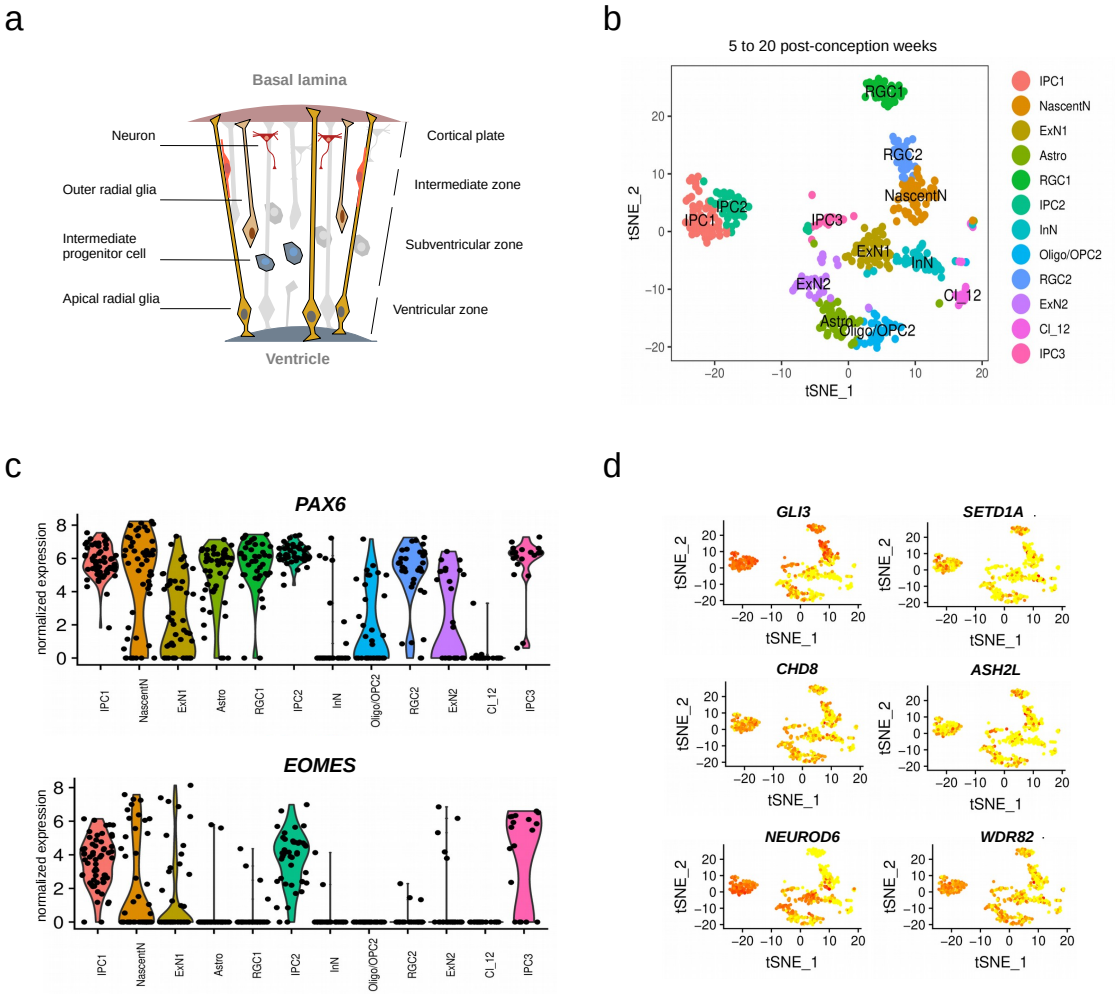
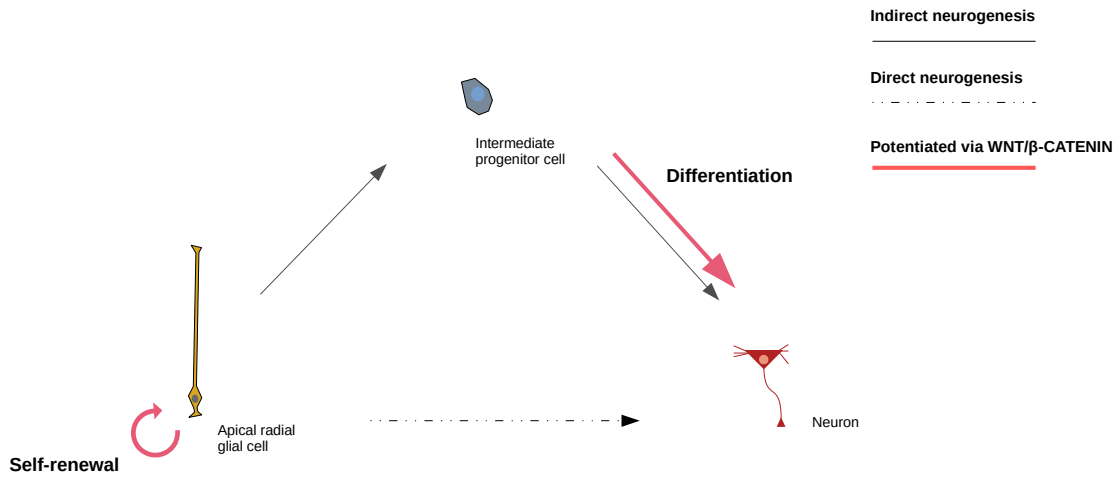


Figure 3: **Progenitor cell-fate decisions shaped by WNT/β-CATENIN signaling**. Studies in mice show that early during neurodevelopment, WNT/β-CATENIN signaling promotes neural stem and progenitor cell self-renewal whereas its depletion causes premature neuronal differentiation [73–75]; later on, its down-regulation is required for generation of intermediate progenitor cells from radial glial cells [75, 76]. Lastly, WNT/β-CATENIN signaling promotes differentiation of intermediate progenitors to produce neurons [74].



Supplementary Figures

Figure S1: **Enrichment analyses**. Permutation tests were applied to determine enrichment of enhancers/promoters within human positively-selected regions (n=314, from [8]), schizophrenia genomic loci (n=108, [23]), and autism risk variants (n=58, [24], retrieved from [25])

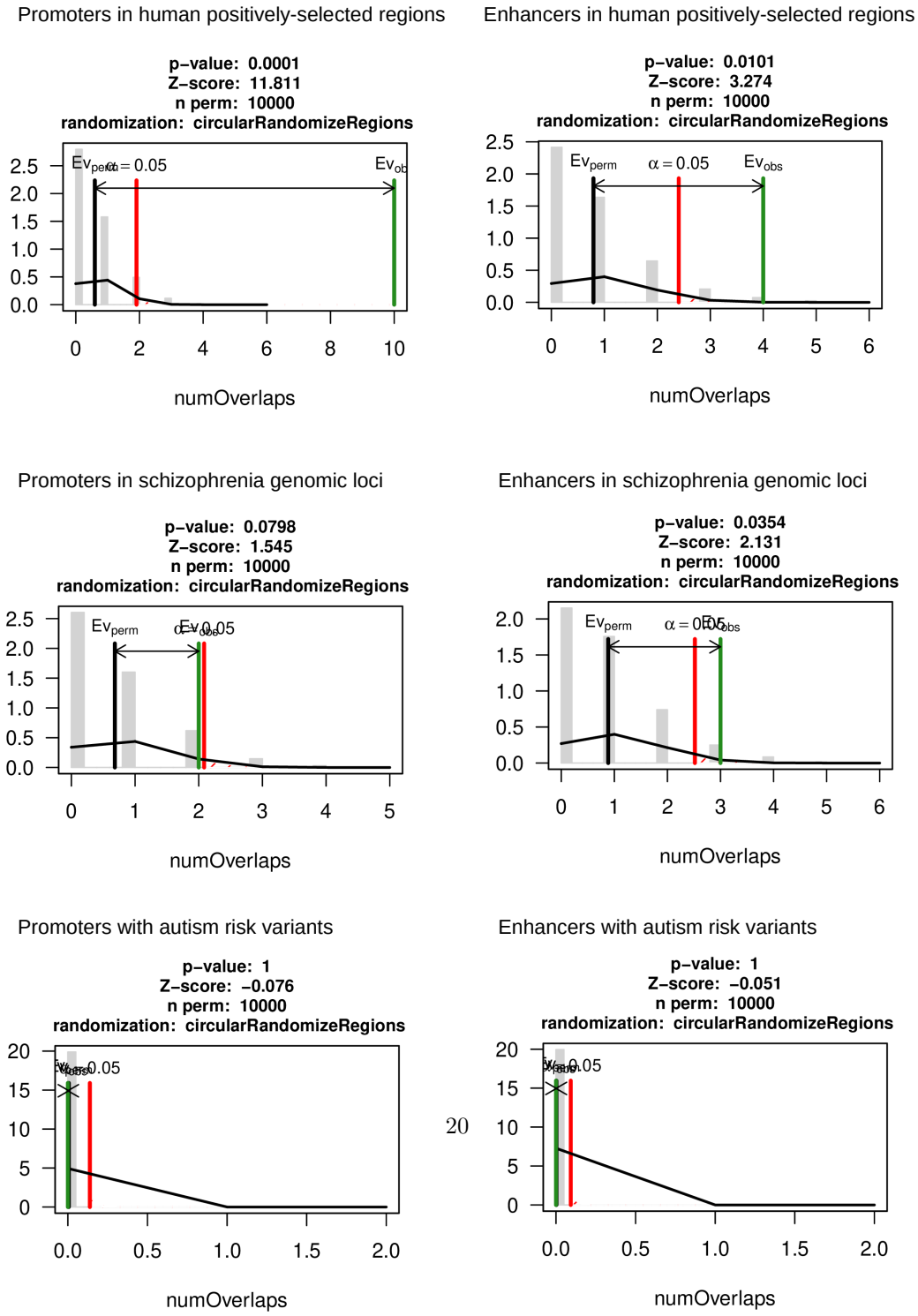


Figure S2: **Clustering results obtained in the original study**. [64] generated and analysed the single-cell transcriptomic data used in this study (5 to 20 post-conception weeks human prenatal brain). They reported two clustering results (A and B) after using different methodologies. Note in B the presence of the three sub-populations for intermediate progenitor cells and the two clusters of radial glial cells, as in our analysis. Images are reproduced with copyright permission from *Science*.

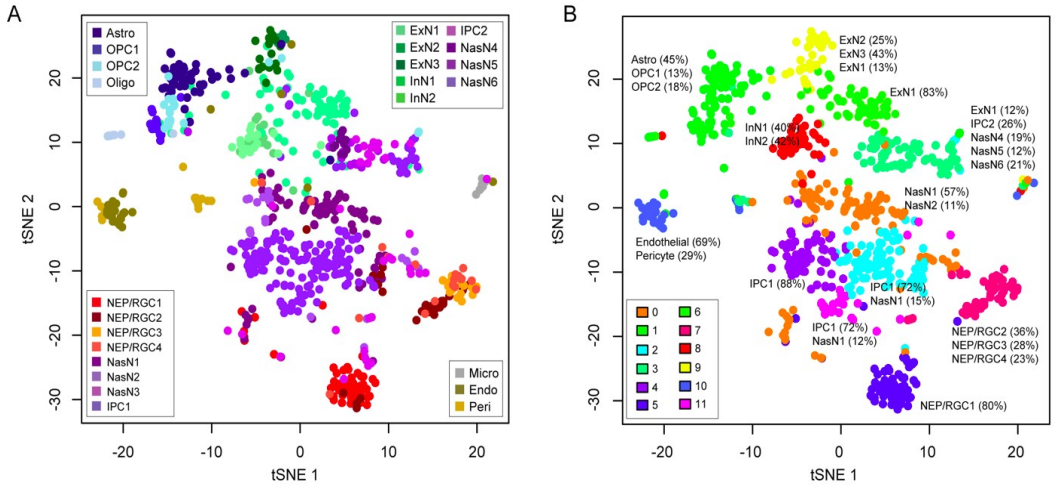
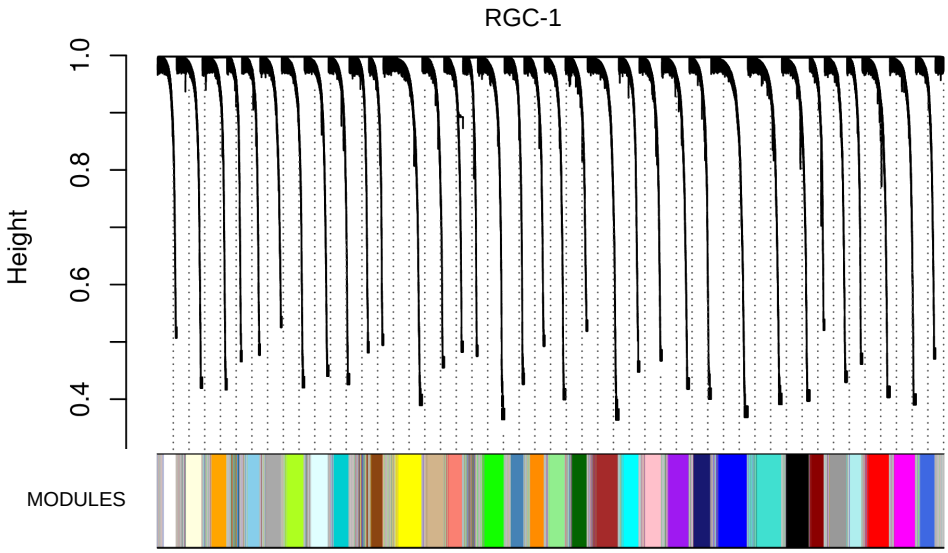


Figure S3: **WGCNA - Radial glial cells gene modules**. A total of 32 modules were detected for RGC-1 sub-population and 26 modules for RGC-2 sub-population.



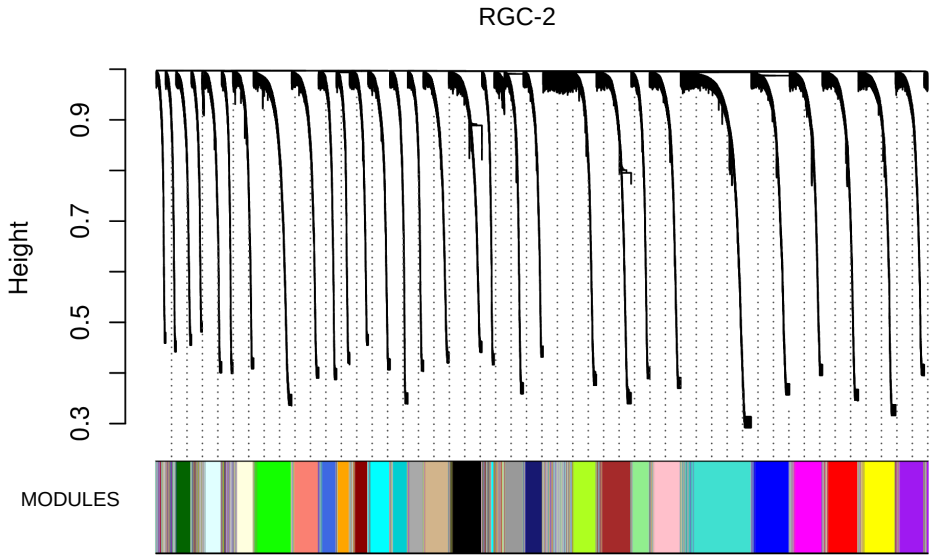
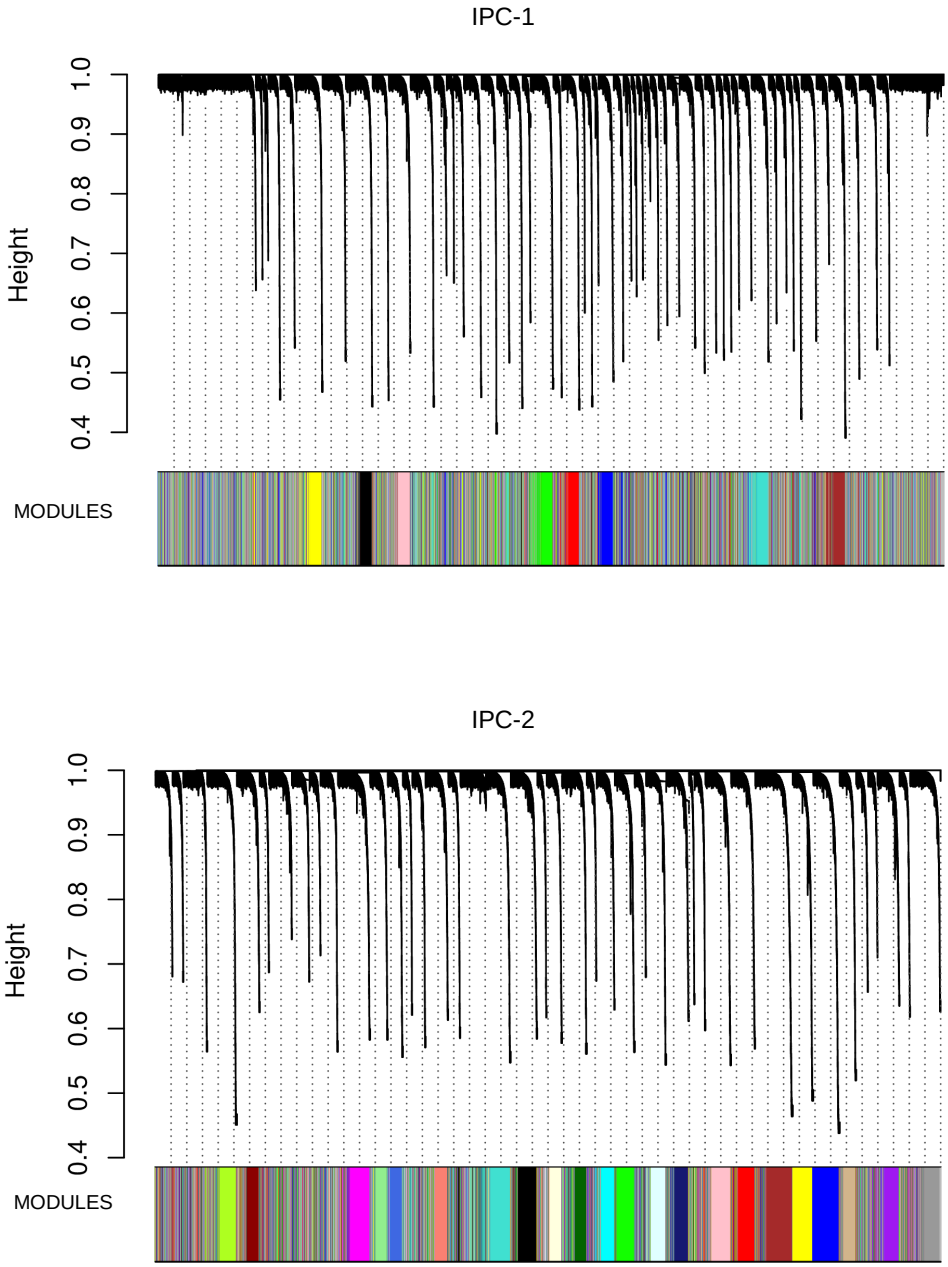



Figure S4: **WGCNA - Intermediate progenitor cells gene modules**. A total of 9 modules were detected for IPC-1 sub-population and 23 modules for IPC-2 sub-population.