

RESEARCH ARTICLE

Open Access

Modern human changes in regulatory regions implicated in cortical development



Juan Moriano^{1,2*}  and Cedric Boeckx^{1,2,3*}

Abstract

Background: 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 identify derived regulatory changes in the modern human lineage in comparison to Neanderthals/Denisovans. We report a set of genes whose enhancers and/or promoters harbor modern human single nucleotide changes and are active at early stages of cortical development.

Results: We identified 212 genes controlled by regulatory regions harboring modern human changes where Neanderthals/Denisovans carry the ancestral allele. These regulatory regions significantly overlap with putative modern human positively-selected regions and schizophrenia-related genetic loci. Among the 212 genes, we identified a substantial proportion of genes related to transcriptional regulation and, specifically, an enrichment for the SETD1A histone methyltransferase complex, known to regulate WNT signaling for the generation and proliferation of intermediate progenitor cells.

Conclusions: This study complements previous research focused on protein-coding changes distinguishing our species from Neanderthals/Denisovans and highlights chromatin regulation as a functional category so far overlooked in modern human evolution studies. We present a set of candidates that will help to illuminate the investigation of modern human-specific ontogenetic trajectories.

Keywords: Modern humans, Neanderthals/Denisovans, Paleogenomics, Regulatory regions, Chromatin regulation, SETD1A/histone methyltransferase complex

Background

Progress in the field of paleogenomics has allowed researchers to study 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 in the modern human lineage 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

* Correspondence: jmoriano@ub.edu; cedric.boeckx@ub.edu

¹Universitat de Barcelona, Gran Via de les Corts Catalanes, Barcelona, Spain
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

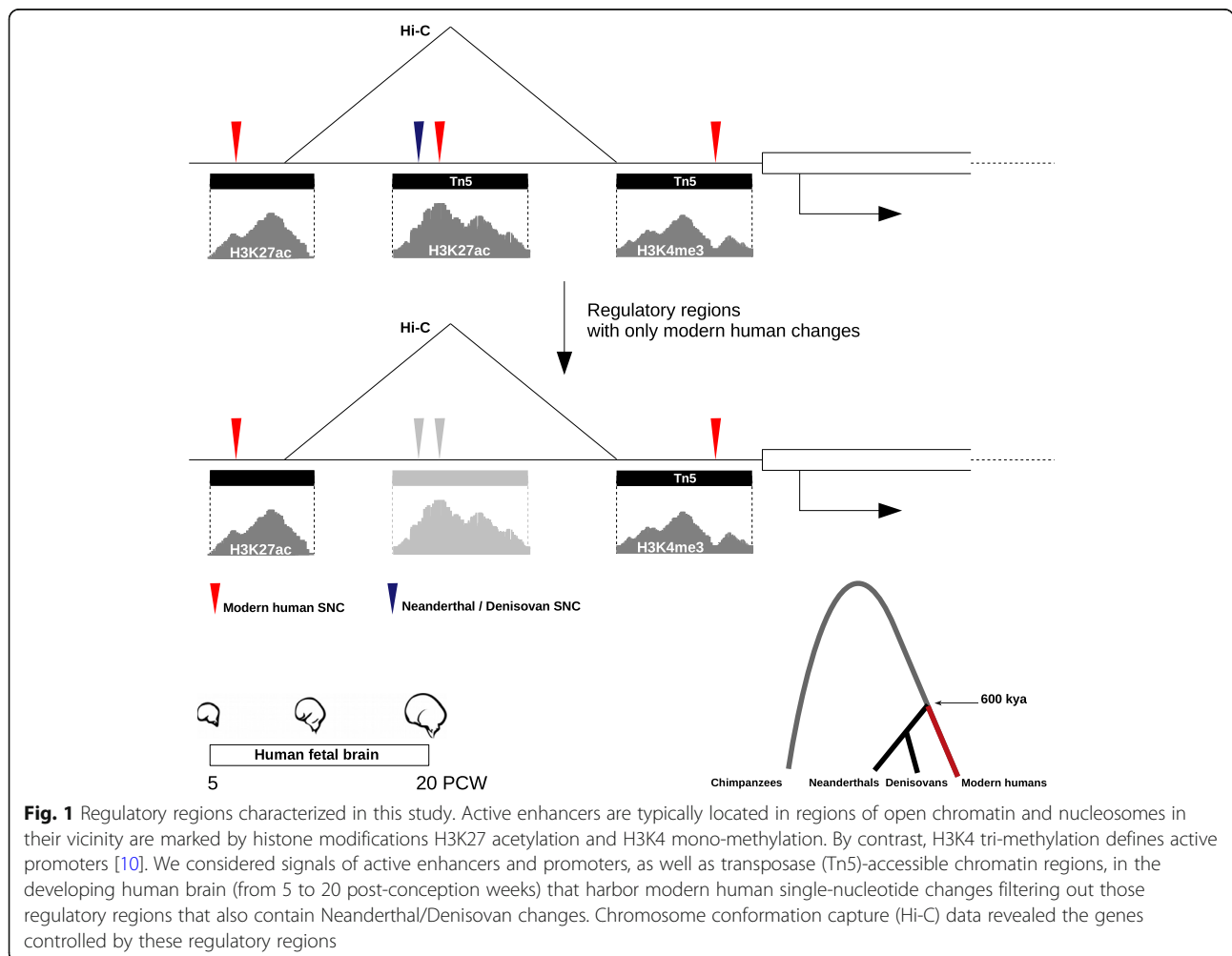
DNA introgression were reported in promoters and conserved genomic regions [9].

Here, we report a set of genes under the control of regulatory regions that harbor modern human-lineage genetic changes and are active at early stages of cortical development (Fig. 1). We integrated data on chromatin immunoprecipitation and open chromatin regions identifying enhancers and promoters active during human cortical development, and the genes regulated by them as revealed by chromatin physical interaction data, together with paleogenomic data of single-nucleotide changes (SNC) distinguishing modern humans and Neanderthal/Denisovan lineages. This allowed us to uncover those enhancer and promoters that harbor modern human SNC (thereafter, mSNC) at fixed or nearly fixed frequency (as defined by [6]) in present-day human populations and where the Neanderthals/Denisovans carry the ancestral allele (Methods section). Next, we analysed single-cell gene expression data and performed co-expression network analysis to identify the genes plausibly under human-specific regulation within genetic

networks in neural progenitor cells (Methods section). Many of the genes controlled by regulatory regions satisfying the aforementioned criteria are involved in chromatin regulation, and prominently among these, the SETD1A histone methyltransferase (SETD1A/HMT) complex. This complex, which has not figured prominently in the modern human evolution literature until now, 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.

Results

Two hundred and twelve genes were found associated to regulatory regions active in the developing human cortex (from 5 to 20 post-conception weeks) that harbor mSNCs and do not contain Neanderthal/Denisovan changes (Suppl. Mat. Tables S1 & S2). Among these, some well-studied disease-relevant genes are found: *HTT* (Huntington disease) [11], *FOXP2* (language impairment) [12], *CHD8* and *CPEB4* (autism spectrum



disorder) [13, 14], *TCF4* (Pitt-Hopkins syndrome and schizophrenia) [15, 16], *GLI3* (macrocephaly and Greig cephalopolysyndactyly syndrome) [17], *PHC1* (primary, autosomal recessive, microcephaly-11) [18], *RCANI* (Down syndrome) [19], and *DYNC1H1* (cortical malformations and microcephaly) [20].

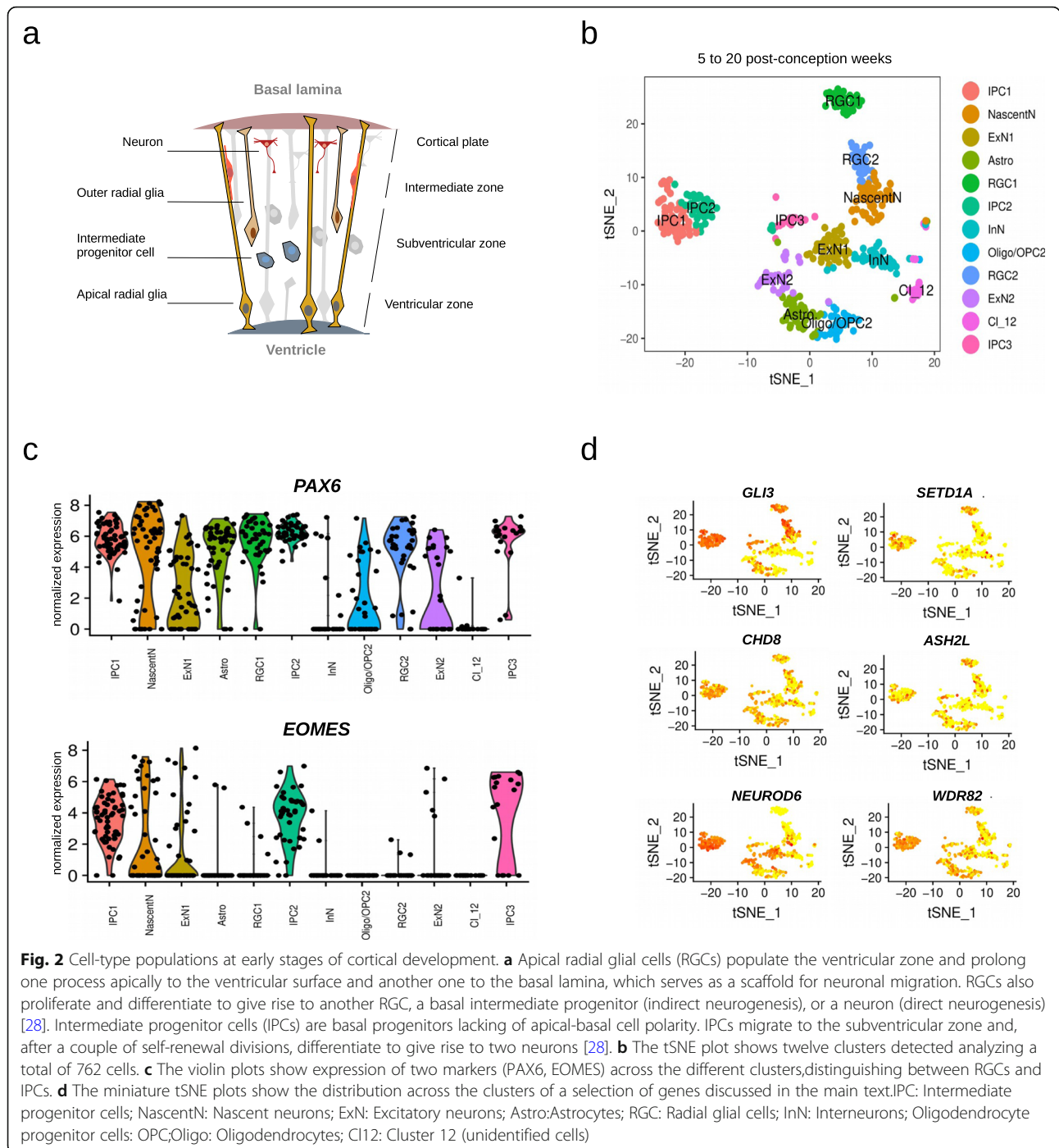
Twelve out of the 212 genes contain fixed mSNCs in enhancers (*NEUROD6*, *GRIN2B*, *LRRC23*, *RNF44*, *KCNA3*, *TCF25*, *TMLHE*, *GLI4*, *DDX12P*, *PLP2*, *TFE3*, *SPG7*), with *LRRC23* having three such changes, and *GRIN2B*, *DDX12P* and *TFE3*, two each. Fourteen genes have fixed mSNCs in their promoters (*LRRC23*, *SETD1A*, *FOXJ2*, *LIMCH1*, *ZFAT*, *SPOP*, *DLGAP4*, *HS6ST2*, *UBE2A*, *FKBP1A*, *RPL6*, *LINC01159*, *RBM4B*, *NFIB*). Only one gene, *LRRC23*, exhibits fixed changes in both its enhancer and promoter regions. To identify putatively mSNC-enriched regions, we ranked regulatory regions by mutation density (Methods section). Top candidates enhancers (top 5% in hits-per-region length distribution) were associated with potassium channel *KCNQ5*, actin-binding protein *FSCN1*, and neuronal marker *NEUROD6*. Top candidate promoters were linked to cytoplasmic dynein *DYNC1H1*, nuclear factor *NFIB*, PHD and RING finger domains-containing *PHRF1*, and kinesin light *KLC1* (Suppl. Mat. Table S3 & S4). Interestingly, most of these are known to be involved in later stages of neurogenesis (differentiation and migration steps).

Previous work has shown an enrichment of enhancer and promoter regions within modern human putative positively-selected regions [8]. For those regulatory regions containing mSNC, a significant over-representation was found for enhancers (permutation test; p -value 0.01) and promoters (permutation test; p -value 10^{-4}) overlapping with modern human candidate sweep regions [8] (Suppl. Fig. 1; Suppl. Mat. Table S5). In addition, we found a significant enrichment for enhancers (permutation test; p -value 0.04; while for promoter regions p -value 0.08) overlapping with genetic loci associated to schizophrenia [21]. By contrast, no significant overlap was found for enhancers/promoters and autism spectrum disorder risk variants ([22], retrieved from [23]) (Suppl. Fig. 1). Single-nucleotide variants can have an impact on epigenetic signals and transcription factor binding affinity in regulatory regions, and thus can alter gene expression levels [24–26]. We also performed motif enrichment analysis for our enhancer/promoter region datasets (Methods section). We found a motif enrichment in enhancer regions for transcriptional regulators *IRF8*, *PU.1*, *CTCF* (Benjamini q -value 0.01) and *OCT4* (Benjamini q -value 0.02); while for promoter regions a motif enrichment was detected for the zinc finger-containing (and *WNT* signaling regulator) *ZBTB33* (Benjamini q -value 0.03).

Next, we evaluated relevant gene ontology and biological categories in our 212 gene list (Methods section). We identified a substantial proportion of genes related to β -catenin binding (GO:0008013; hypergeometric test (h.t.): adj p -value 0.11) and transcriptional regulation (GO:0044212; h.t.: adj p -value 0.17), and detected a significant enrichment from the CORUM protein complexes database for the SETD1A/HMT complex (CORUM:2731; h.t.: adj p -value 0.01). Indeed, three members of the SETD1A/HMT complex are present in our 212 gene list: SETD1A (fixed mSNC in promoter), ASH2L (mSNC in enhancer) and WDR82 (mSNC in enhancer). SETD1A 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 [27]. Furthermore, the *SETD1A* promoter and the *WDR82* enhancer containing the relevant changes fall within putative positively-selected regions in the modern human lineage [8] (Suppl. Mat. Table S5).

The abundance of transcriptional regulators and the specific enrichment for the SETD1A/HMT complex led us to examine the gene expression programs likely under their influence in neural progenitor cells. From 5 to 20 post-conception weeks, different types of cells populate the germinal zones of the developing cortex (Fig. 2). We re-analyzed gene expression data at single-cell resolution from a total of 762 cells from the developing human cortex, controlling for cell-cycle heterogeneity as a confounding factor in the analysis of progenitor populations (Methods section). We focused on two progenitor cell-types—radial glial and intermediate progenitor cells (RGCs and IPCs, respectively)—two of the main types of progenitor cells that give rise, in an orderly manner, to the neurons present in the adult brain (Fig. 2). Two clusters of RGCs were identified (*PAX6+* and *EOMES-* cells), and three clusters 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 next identified genetic networks (based on highly-correlated gene expression levels) in the different cluster of progenitor cells (except for IPC cluster 3, which was excluded due to the low number of cells) (Methods section; Suppl. Figs. 3 & 4). This allowed us to identify genes present in the 212 gene list within modules of co-expressed genes whose biological relevance was assessed through a functional enrichment analysis using *g:Profiler2* R package [29] (hypergeometric test; see Methods).

An over-representation of genes related to the human phenotype ontology term ‘Neurodevelopmental abnormality’ was detected in the RGC-cluster 2 turquoise



module (HP:0012759; h.t.: adj *p*-value 0.03, [Suppl. Mat. Table S6](#)). Indeed, a considerable amount of genes were found to be associated to phenotype terms ‘Neurodevelopmental delay’ and ‘Skull size’ (HP:0000240, respectively; h.t.: adj *p*-value 0.07 and 0.13, respectively; [Suppl. Mat. Table S7](#)). These terms have appeared prominently in the human evolution literature in the context of neoteny and delay brain maturation, brain growth and the craniofacial phenotype in between

species comparisons [6, 30–32]. Two chromatin regulators with mSNC in regulatory regions are present in these two ontology terms and are associated to neurodevelopmental disorders: KDM6A (mSNC in promoter), which associates to the H3K4 methyltransferase complex [27], and is mutated in patients with Kabuki syndrome [33]; and PHC1 (mSNC in promoter), a component of the repressive complex PRC1 [27], found in patients with primary microcephaly-11 [18]. Among the total

genes related to the ‘Skull size’ term ($n = 109$), we found an over-representation of genes (*CDON*, *GLI3*, *KIF7*, *GAS1*) related to the hedgehog signaling pathway (KEGG:04340; h.t.: adj p -value 0.05). Of these, *GLI3* (mSNC in promoter) is perhaps the most salient member, highlighted in previous work as a gene harbouring an excess of mSNC [6]. *GLI3* is a gene linked to macrocephaly and the craniofacial phenotype [17, 34] and under putative modern human positive selection [8]. Considering that hedgehog signaling plays a critical role in basal progenitor expansion [35], we note the presence in this turquoise module of the outer radial glia-specific genes *IL6ST* and *STAT3* [36]. The forkhead-box transcription factor *FOXP2* is also present in RGC-cluster 2 turquoise module and associated to the ‘Neurodevelopmental delay’ ontology term. Its promoter harbors an almost fixed (>99%) mSNC. *FOXP2* is a highly conserved protein involved in language-related disorders whose evolutionary changes are particularly relevant for understanding human cognitive traits [37]. This mSNC (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 [37].

While we did not detect a specific enrichment in the modules containing SETD1A/HMT complex components *ASH2L* or *WDR82* genes, the IPC-cluster 2 midnightblue module, which contains SETD1A, shows an enrichment for a β -CATENIN-containing complex (SETD7-YAP-AXIN1- β -CATENIN complex; CORUM:6343; h.t.: adj p -value 0.05; Suppl. Mat. Table S8) and indeed contains WNT-effector TCF3, which harbors nearly fixed missense mutations in modern humans [6]. SETD1A is known to interact with β -CATENIN [38, 39] and increase its expression to promote neural progenitor proliferation [40].

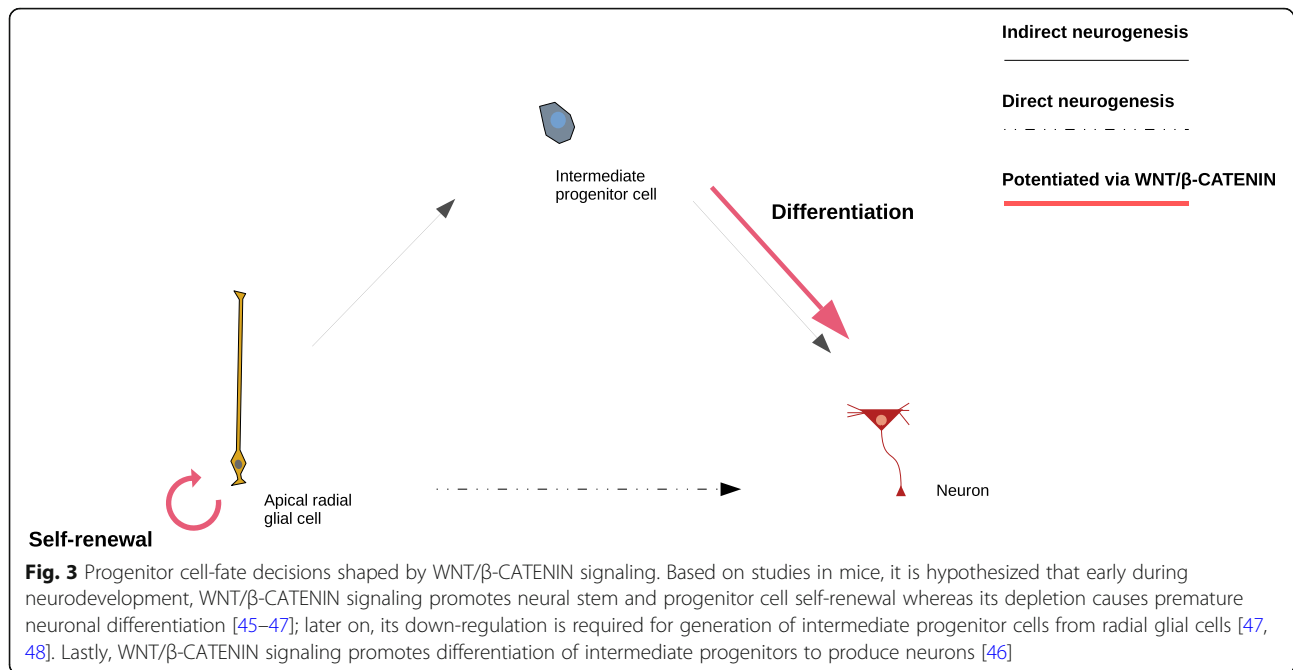
Discussion

By integrating data from paleogenomics and chromatin interaction and modification, we identified a set of genes controlled by regulatory regions that are active during early cortical development and contain single nucleotide changes that appeared in the modern human lineage after the split from the Neanderthal/Denisovan lineage. The regulatory regions reported here significantly overlap with putative modern human positively-selected regions and schizophrenia genomic loci, and control a set of genes among which we find a high number related to chromatin regulation, and most specifically 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 [41–43]. Both SETD1A and *ASH2L*, core components of the HMT complex, regulate WNT/ β -

CATENIN signaling [38–40, 44], which influences cell-fate decisions by promoting either self-maintenance or differentiation depending on the stage of progenitor differentiation (Fig. 3).

SETD1A (fixed mSNC in promoter), implicated in schizophrenia and developmental language impairment [49, 50], acts 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 [40]. In addition, one of SETD1A direct targets is the WNT-effector TCF4 [51], whose promoter also harbors a mSNC. Similarly, *ASH2L* specifically regulates WNT signaling: Conditional knockout of *ASH2L* significantly compromises the proliferative capacity of RGCs and IPCs by the time of generation of upper-layer neurons, with these progenitor cells showing a marked reduction in H3K4me3 levels and downregulation of WNT/ β -CATENIN signaling-related genes (defects that can be rescued by over-expression of β -CATENIN) [44]. Taken together, depletion of components of the SETD1A/HMT complex impairs the proliferative capacity of progenitor cells, altering the indirect mode of neurogenesis, with a specific regulation of the conserved WNT signaling. Interestingly, in addition to the aforementioned properties of the regulatory regions of the SETD1A complex components found in modern humans (overlap with modern human positively-selected regions and containing mSNCs), a recent work studying species-specific differences in chromatin accessibility using brain organoids reported that regulatory regions associated to *SETD1A* and *WDR82* were found in differentially-accessible regions in human organoids in comparison to chimpanzee organoids, with the *SETD1A* region overlapping with a human-gained histone modification signal when compared to macaques [52].

The dysfunction of chromatin regulators is among the most salient features behind causative mutations in neurodevelopmental disorders [53]. Our data highlights chromatin modifiers and remodelers that play prominent roles in neurodevelopmental disorders affecting brain growth and facial features. Along with the aforementioned chromatin regulators *PHC1* (microcephaly) and *KDM6A* (Kabuki syndrome), another paradigmatic example is the ATP-dependent chromatin remodeler *CHD8* (mSNC in enhancer), which controls neural progenitor cell proliferation through WNT-signaling related genes [54, 55]. *CHD8* is a high-risk factor for autism spectrum disorder and patients with *CHD8* mutations characteristically present macrocephaly and distinctive facial features [13]. Intriguingly, another ATP-dependent chromatin remodeler, *CHD2* (mSNC in enhancer),



presents a motif in the SETD1A promoter region containing the fixed mSNC (16:30969654; UCSC Genome Browser).

We have focused on the early stages of cortical development. While single-cell gene expression data of neural progenitor cells still remains limited, future integration of these data with other datasets covering different neo-cortical regions [56] will shed further light on modern human changes and cortical areas-specific progenitor cells. We acknowledge, in addition, that the genetic changes distinguishing modern humans and Neanderthals/Denisovans may be relevant at other stages of neurodevelopment, including the adult human brain. Progress in single-cell multi-omic technologies applied to brain organoid research will be critical to assess the impact of such changes in the diverse neural and non-neural cell-types through different developmental stages. Moreover, we excluded the examination of regulatory regions harboring Neanderthal/Denisovan changes due to the low number of high-quality genomes from Neanderthal/Denisovan individuals, which makes the determination of allele frequency in these species unreliable. We hope that the availability of a higher number of high-quality genomes for these species in the future will make such examination feasible.

Conclusions

This study complements previous research focused on protein-coding changes [4, 6] and helps extend the investigation of species-specific differences in cortical development that has so far relied on detailed comparisons between humans and non-human primates [52, 57–60].

We provide a list of new candidate genes for the study of human species-specific differences during the early stages of cortical development. The study of modern human evolutionary changes affecting chromatin regulators integrated with the examination of neurodevelopmental disorders could be a valuable entry point to understand modern human-specific brain ontogenetic trajectories.

Methods

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 SNC in the modern human lineage (at fixed or above 90% frequency in present-day human populations) and Neanderthal/Denisovan changes. [6] analyzed high-coverage genotypes from one Denisovan and two Neanderthal individuals to report a catalog of SNC that appeared in the modern human lineage after their split from Neanderthals/Denisovans. Similarly, [6] also reported a list of SNC present in the Neanderthal/Denisovan lineages where modern humans carry the inferred ancestral allele.

For enhancer–promoter linkages, we used publicly available data from [61], based on transposase-accessible chromatin coupled to sequencing and integrated with chromatin capture via Hi-C data, from 15 to 17 post-conception weeks of the developing human cortex. A total of 92 promoters and 113 enhancers were selected as harboring mSNC and being depleted of Neanderthal/Denisovan SNC (from a total of 2574 enhancers and 1553 promoters present in the original dataset). Additionally, we completed the previous dataset filtering

annotated enhancer-gene linkages via Hi-C from the adult prefrontal cortex [62] (PsychENCODE resource portal: <http://resource.psychencode.org/>). In this case, enhancers ($n = 32,803$) were selected for further analyses if their coordinates completely overlapped with signals of active enhancers (H3K27ac) (that do not overlap with promoter signals (H3K4me3)) from the developing human cortex between 7 to 12 post-conception weeks [63]. A total of 43 enhancers, containing mSNC but free of Neanderthal/Denisovan SNC, passed this filtering. As a whole, the final integrated dataset covered regulatory regions active at early stages of human prenatal cortical development and linked to 212 genes. The coordinates (hg19 version) of the regulatory regions containing mSNC are available in the [Supplementary Material Tables S1 & S2](#).

Human positively-selected regions coordinates were retrieved from [8].

Single-cell RNA-seq analysis

The single-cell transcriptomic analysis was performed using the Seurat package v2.4 [64] in RStudio v1.1.463 (server mode).

Single-cell gene expression data was retrieved from [63] 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 1,000,000. 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 [65] that assigns scores genes to either G1/S or G2/M phase (function *CellCycleScoring*), allowing us to reduce heterogeneity due to differences in cell-cycle phases. 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 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 [63].

Weighted gene co-expression network analysis

For the gene co-expression network analysis we used the WGCNA R package [66, 67]. For each cluster of progenitor cells (RGC-1, 34 cells (15,017 genes); RGC-2, 30 cells (14,747 genes); IPC-1, 52 cells (15,790 genes); IPC-2, 41 cells (15,721 genes); IPC-3 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 clusters, with R^2 : 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. Figs. 3 & 4).

Enrichment analysis

We ranked regulatory regions by mutation density calculating number of single nucleotide changes per regulatory region length (for those regions spanning at least 1000 base pairs). Top candidates were those ranking in the distribution within the 5% out of the total number of enhancers or promoters (Suppl. Mat Tables S3 & S4). The g:Profiler2 R package [29] was used to perform enrichment analyses (hypergeometric test; correction method ‘gSCS’; background genes: ‘only annotated genes’, *Homo sapiens*) 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 *regionR* [68]. The Hypergeometric Optimization of Motif EnRichment (HOMER) software v4.10 [69] was employed for motif discovery analysis, selecting best matches (Benjamini q -value < 0.05) of known motifs ($n = 428$; ChIP-seq-based) in our promoter and enhancer datasets.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-6706-x>.

Additional file 1. Supplementary Material.

Additional file 2. Supplementary Figures.

Abbreviations

SNC: Single-nucleotide change; mSNC: Modern human single-nucleotide change; HMT: Histone methyltransferase complex; h.t.: Hypergeometric test; RGC: Radial glial cell; IPC: Intermediate progenitor cell

Acknowledgements

We thank members of the 'Cognitive Biology of Language' group for helpful discussions.

Author's 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. The authors read and approved this manuscript.

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 (Evolvinguistics: JP17H06379), and Generalitat de Catalunya (Government of Catalonia) – 2017-SGR-341. The funding bodies played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets supporting the conclusions of this article (tables and code) are available in the Figshare repository, under <https://doi.org/10.6084/m9.figshare.11603478.v1> and <https://doi.org/10.6084/m9.figshare.11608074.v1>. We also made use of web-based data resources from [62] and [63] through the PsychENCODE portal (<http://development.psychencode.org/>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Universitat de Barcelona, Gran Via de les Corts Catalanes, Barcelona, Spain.

²Universitat de Barcelona Institute of Complex Systems, Martí Franquès, Barcelona, Spain. ³Catalan Institution for Research and Advanced Studies, Passeig Lluís Companys, Barcelona, Spain.

Received: 8 November 2019 Accepted: 30 March 2020

Published online: 16 April 2020

References

- Pääbo S. The human condition—a molecular approach. *Cell*. 2014;157(1):216–26 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S009286741301605X>.
- Hublin J-J, Neubauer S, Gunz P. Brain ontogeny and life history in Pleistocene hominins. *Philos Transact Royal Soc B: Biol Sci*. 2015;370(1663):20140062 Available from: <http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2014.0062>.
- Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science*. 2012;338(6104):222–6 Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.1224344>.
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 505(7481):43–9 Available from: <http://www.nature.com/articles/nature12886>.
- Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija cave in Croatia. *Science*. 2017;358(6363):655–8 Available from: <http://www.sciencemag.org/lookup/doi/10.1126/science.aao1887>.
- Kuhlwil M, Boeckx C. A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Sci Rep*. 2019;9(1) Available from: <http://www.nature.com/articles/s41598-019-44877-x>.
- King M, Wilson A. Evolution at two levels in humans and chimpanzees. *Science*. 1975;188(4184):107–16 Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.1090005>.
- Peyrégne S, Boyle MJ, Dannemann M, Prüfer K. Detecting ancient positive selection in humans using extended lineage sorting. *Genome Res*. 2017;27(9):1563–72 Available from: <http://genome.cshlp.org/lookup/doi/10.1101/gr.219493.116>.
- Petr M, Pääbo S, Kelso J, Vernot B. Limits of long-term selection against Neandertal introgression. *Proc Natl Acad Sci*. 2019;116(5):1639–44 Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1814338116>.
- Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell*. 2013;49(5):825–37 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1097276513001020>.
- Saudou F, Humbert S. The biology of huntingtin. *Neuron*. 2016;89(5):910–26 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0896627316000969>.
- Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*. 2001;413(6855):519–23 Available from: <http://www.nature.com/articles/35097076>.
- Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, et al. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell*. 2014;158(2):263–76 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867414007491>.
- Parras A, Anta H, Santos-Galindo M, Swarup V, Elorza A, Nieto-González JL, et al. Autism-like phenotype and risk gene mRNA deadenylation by CPEB4 mis-splicing. *Nature*. 2018;560(7719):441–6 Available from: <http://www.nature.com/articles/s41586-018-0423-5>.
- Zweier C, Peippo MM, Hoyer J, Sousa S, Bottani A, Clayton-Smith J, et al. Haploinsufficiency of TCF4 causes Syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am J Hum Genet*. 2007;80(5):994–1001 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S000292707609562>.
- Forrest MP, Hill MJ, Kavanagh DH, Tansey KE, Waite AJ, Blake DJ. The psychiatric risk gene transcription factor 4 (TCF4) regulates neurodevelopmental pathways associated with schizophrenia, autism, and intellectual disability. *Schizophr Bull*. 2018;44(5):1100–10 Available from: <https://academic.oup.com/schizophreniabulletin/article/44/5/1100/4700989>.
- Kalff-Suske M, Wild A, Topp J, Wessling M, Jacobsen E-M, Bornholdt D, et al. Point mutations throughout the GLI3 gene cause Greig Cephalopolysyndactyly syndrome. *Hum Mol Genet*. 1999;8(9):1769–77 Available from: <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/8.9.1769>.
- Awad S, Al-Dosari MS, Al-Yacoub N, Colak D, Salih MA, Alkuraya FS, et al. Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis. *Hum Mol Genet*. 2013;22(11):2200–13 Available from: <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddt072>.
- Fuentes J-J, Pritchard MA, Planas AM, Bosch A, Ferrer I, Estivill X. A new human gene from the Down syndrome critical region encodes a proline-rich protein highly expressed in fetal brain and heart. *Hum Mol Genet*. 1995;4(10):1935–44 Available from: <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/4.10.1935>.
- Poirier K, Lebrun N, Broix L, Tian G, Saillour Y, Boscheron C, et al. Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. *Nat Genet*. 2013;45(6):639–47 Available from: <http://www.nature.com/articles/ng.2613>.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–7 Available from: <http://www.nature.com/articles/nature13595>.
- Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team, Grove J, Ripke S, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431–44 Available from: <http://www.nature.com/articles/s41588-019-0344-8>.
- Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2018;47(D1):D1005–12. <https://doi.org/10.1093/nar/gky1120>.

24. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, et al. Extensive variation in chromatin states across humans. *Science*. 2013;342(6159):750–2 Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.1242510>.
25. Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science*. 2013;342(6159):744–7 Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.1242463>.
26. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. *Science*. 2013;342(6159):747–9 Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.1242429>.
27. Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science*. 2016;352(6290):aad9780 Available from: <http://www.sciencemag.org/lookup/doi/10.1126/science.aad9780>.
28. Florio M, Huttner WB. Neural progenitors, neurogenesis and the evolution of the neocortex. *Development*. 2014;141(11):2182–94 Available from: <http://dev.biologists.org/cgi/doi/10.1242/dev.090571>.
29. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. G:profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. 2007;35 Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkm226>.
30. Somel M, Franz H, Yan Z, Lorenc A, Guo S, Giger T, et al. Transcriptional neoteny in the human brain. *Proc Natl Acad Sci*. 2009;106(14):5743–8 Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.0900544106>.
31. Liu X, Somel M, Tang L, Yan Z, Jiang X, Guo S, et al. Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res*. 2012;22(4):611–22 Available from: <http://genome.cshlp.org/cgi/doi/10.1101/gr.127324.111>.
32. Lesciotta KM, Richtsmeier JT. Craniofacial skeletal response to encephalization: how do we know what we think we know? *Am J Phys Anthropol*. 2019;168(S67):27–46 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.23766>.
33. Miyake N, Koshimizu E, Okamoto N, Mizuno S, Ogata T, Nagai T, et al. *MLL2* and *KDM6A* mutations in patients with kabuki syndrome. *Am J Med Genet*. 2013;161(9):2234–43. <https://doi.org/10.1002/ajmg.a.36072>.
34. Adhikari K, Fuentes-Guajardo M, Quinto-Sánchez M, Mendoza-Revilla J, Camilo Chacón-Duque J, Acuña-Alonzo V, et al. A genome-wide association scan implicates *DCHS2*, *RUNX2*, *GLI3*, *PAX1* and *EDAR* in human facial variation. *Nat Commun*. 2016;7(1):11616 Available from: <http://www.nature.com/articles/ncomms11616>.
35. Wang L, Hou S, Han Y-G. Hedgehog signaling promotes basal progenitor expansion and the growth and folding of the neocortex. *Nat Neurosci*. 2016;19(7):888–96 Available from: <http://www.nature.com/articles/nn.4307>.
36. Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, et al. Molecular identity of human outer radial glia during cortical development. *Cell*. 2015;163(1):55–67 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867415011241>.
37. Fisher SE. Human genetics: the evolving story of *FOXP2*. *Curr Biol*. 2019;29(2):R65–7 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S096098221831546X>.
38. Hoffmeyer K, Raggioli A, Rudloff S, Anton R, Hierholzer A, Del Valle I, et al. Wnt/beta-catenin signaling regulates telomerase in stem cells and cancer cells. *Science*. 2012;336(6088):1549–54 Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.1218370>.
39. Salz T, Li G, Kaye F, Zhou L, Qiu Y, Huang S. *hSETD1A* regulates Wnt target genes and controls tumor growth of colorectal cancer cells. *Cancer Res*. 2014;74(3):775–86 Available from: <http://cancerres.aacrjournals.org/content/74/3/775>.
40. Li Y, Jiao J. Histone chaperone HIRA regulates neural progenitor cell proliferation and neurogenesis via beta-catenin. *J Cell Biol*. 2017;216(7):1975–92 Available from: <http://www.jcb.org/lookup/doi/10.1083/jcb.201610014>.
41. Chen T, Dent SYR. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*. 2014;15(2):93–106 Available from: <http://www.nature.com/articles/nrg3607>.
42. Hirabayashi Y, Gotoh Y. Epigenetic control of neural precursor cell fate during development. *Nat Rev Neurosci*. 2010;11(6):377–88 Available from: <http://www.nature.com/articles/nrn2810>.
43. Tuoc TC, Pavlakis E, Tylkowski MA, Stoykova A. Control of cerebral size and thickness. *Cell Mol Life Sci*. 2014;71(17):3199–218 Available from: <http://link.springer.com/10.1007/s00018-014-1590-7>.
44. Li L, Ruan X, Wen C, Chen P, Liu W, Zhu L, et al. The COMPASS family protein *ASH2L* mediates Corticogenesis via transcriptional regulation of Wnt signaling. *Cell Rep*. 2019;28(3):698–711.e5 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2211124719308289>.
45. Chenn A, Walsh CA. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science*. 2002;297(5580):365–9 Available from: <https://science.sciencemag.org/content/297/5580/365>.
46. Munji RN, Choe Y, Li G, Siegenthaler JA, Pleasure SJ. Wnt signaling regulates neuronal differentiation of cortical intermediate progenitors. *J Neurosci*. 2011;31(5):1676–87 Available from: <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5404-10.2011>.
47. Draganova K, Zemke M, Zurkirchen L, Valenta T, Cantú C, Okoniewski M, et al. Wnt/Beta-catenin Signaling546Regulates sequential fate decisions of murine cortical precursor cells: Beta-catenin signaling Regulates547Sequential neural fate. *Stem Cells*. 2015;33(1):170–82 Available from: <http://doi.wiley.com/10.1002/stem.1820>.
48. Mutch CA, Schulte JD, Olson E, Chenn A. Beta-catenin signaling negatively regulates intermediate progenitor population numbers in the developing cortex. *PLoS One*. 2010;5(8):e12376 Available from: <https://dx.plos.org/10.1371/journal.pone.0012376>.
49. Takata A, Xu B, Ionita-Laza I, Roos J, Gogos JA, Karayiorgou M. Loss-of-function variants in schizophrenia risk and *SETD1A* as a candidate susceptibility gene. *Neuron*. 2014;82(4):773–80 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0896627314003584>.
50. Eising E, Carrion-Castillo A, Vino A, Strand EA, Jakielski KJ, Scerri TS, et al. A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol Psychiatry*. 2019;24(7):1065–78 Available from: <http://www.nature.com/articles/s41380-018-0020-x>.
51. Mukai J, Cannavò E, Crabtree GW, Sun Z, Diamantopoulou A, Thakur P, et al. Recapitulation and reversal of schizophrenia-related phenotypes in *Setd1a*-deficient mice. *Neuron*. 2019; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0896627319307871>.
52. Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchís-Calleja F, et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature*. 2019;574(7778):418–22 Available from: <https://www.nature.com/articles/s41586-019-1654-9>.
53. Gabriele M, Tobon AL, D'Agostino G, Testa G. The chromatin basis of neurodevelopmental disorders: rethinking dysfunction along the molecular and temporal axes. *Prog Neuro-Psychopharmacol Biol Psychiatry*. 2018;84:306–27 Available from: <http://www.sciencedirect.com/science/article/pii/S027854617305389>.
54. Sugathan A, Biagioli M, Golzio C, Erdin S, Blumenthal I, Manavalan P, et al. *CHD8* regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci*. 2014;111(42):E4468–77 Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1405266111>.
55. Durak O, Gao F, Kaeser-Woo YJ, Rueda R, Martorell AJ, Nott A, et al. *Chd8* mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. *Nat Neurosci*. 2016;19(11):1477–88 Available from: <http://www.nature.com/articles/nn.4400>.
56. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*. 2017;358(6368):1318–23 Available from: <http://www.sciencemag.org/lookup/doi/10.1126/science.aap8809>.
57. Mora-Bermúdez F, Badsha F, Kanton S, Camp JG, Vernot B, Köhler K, et al. Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. Musacchio a, editor. *eLife*. 2016 Sep;5:e18683. <https://doi.org/10.7554/eLife.18683>.
58. Otani T, Marchetto MC, Gage FH, Simons BD, Livesey FJ. 2D and 3D stem cell models of primate cortical development identify species-specific differences in progenitor behavior contributing to brain size. *Cell Stem Cell*. 2016;18(4):467–80 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1934590916001090>.
59. Marchetto MC, Hrvoj-Mihic B, Kerman BE, Yu DX, Vadodaria KC, Linker SB, et al. Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. Zoghbi HY, Arlotta P, editors. *eLife*. 2019;8:e37527. <https://doi.org/10.7554/eLife.37527>.
60. Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, et al. Establishing cerebral organoids as models of human-specific brain evolution. *Cell*. 2019;176(4):743–756.e17 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867419300509>.

61. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, et al. The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell*. 2018;172(1–2):289–304.e18 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867417314940>.
62. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. Ashley-Koch AE, Crawford GE, Garrett ME, Song L, Safi a, Johnson GD, et al., editors. *Science*. 2018;362(6420) Available from: <https://science.sciencemag.org/content/362/6420/eaat8464>.
63. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. 2018;362(6420):eaat7615 Available from: <http://www.sciencemag.org/lookup/doi/10.1126/science.aat7615>.
64. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36:411. <https://doi.org/10.1038/nbt.4096>.
65. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352(6282):189–96 Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.aad0501>.
66. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. <https://doi.org/10.1186/1471-2105-9-559>.
67. Langfelder P, Horvath S. Fast *r* functions for robust correlations and hierarchical clustering. *J Stat Softw*. 2012;46(11) Available from: <http://www.jstatsoft.org/v46/i11/>.
68. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*. 2015;btv562 Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv562>.
69. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1097276510003667>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

