

PhD degree in Systems Medicine (Curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

Settore disciplinare: MED/04

**Brain organoid modelling of human corticogenesis:
the paradigm of Weaver syndrome**

Sebastiano Trattaro

European Institute of Oncology

University of Milan

Matricola n. R11759

Tutor: Prof. Giuseppe Testa

European Institute of Oncology, Milan

University of Milan, Milan

Human Technopole, Milan

PhD Coordinator: Prof. Saverio Minucci

Anno accademico 2019-2020

TABLE OF CONTENTS

TABLE OF CONTENTS.....	i
LIST OF ABBREVIATIONS.....	iv
FIGURE INDEX.....	viii
TABLE INDEX.....	xii
1. ABSTRACT.....	1
2. INTRODUCTION.....	3
2.1 Developmental basis of human neocortical development.....	3
2.1.1 Molecular features of mammalian cortical development.....	4
2.1.2 Neurogenesis in the developing mammalian neocortex.....	5
2.1.3 Outer subventricular zone.....	6
2.2 Investigating human brain development.....	8
2.2.1 Primary and post mortem samples.....	8
2.2.2 Induced Pluripotent Stem Cells.....	10
2.2.2.1 Cell reprogramming.....	10
2.2.2.2 Advantages and disadvantages of iPSC use.....	11
2.2.2.3 Neural differentiation using pluripotent stem cells.....	12
2.3 Brain organoids to model human brain development.....	13
2.3.1 Cross-comparison of different brain organoid protocols.....	14
2.3.2 Brain organoids: reproducibility and culture-dependent stress.....	15
2.3.3 Brain organoids and disease modelling.....	16
2.4 Regulation of gene expression in human brain development.....	18
2.5 Polycomb group proteins.....	21
2.5.1 Polycomb Repressive Complexes: a focus on PRC2.....	21
2.5.1.1 Enhancer of Zeste 2.....	23
2.5.2 PRC2 function in pluripotency and cortical development.....	24
2.6 Overgrowth and intellectual disability: the example of Weaver Syndrome.....	26
2.6.1 Clinical features associated with Weaver Syndrome.....	26
2.6.1.1 Intellectual disability.....	27
2.6.1.2 Neural crest stem cells.....	27
2.6.2 Molecular basis of Weaver Syndrome.....	28
2.7 Genome editing approaches for disease-modelling.....	29
2.7.1 Clustered Regularly Interspaced Short Palindromic Repeats.....	29
2.7.2 Genome engineering applied to disease modelling.....	31
2.7.3 Genome engineering applied to neurodevelopment.....	32
3. AIMS OF THE STUDY.....	34

4. RESULTS	35
4.1 Analysis of the BrainSpan Atlas revealed transcriptional programmes of the developing human fetal cortex	35
4.2 Cortical Brain Organoids globally resembled human fetal cortex and evolved in a two-step fashion	45
4.3 CBO reproducibly modulated co-expression patterns of cortical development.....	58
4.4 Benchmarking of brain organoids against pre-natal human fetal cortical development revealed heterochronicity of differentiation across protocols	66
4.5 Gene signatures specific for brain populations and functions unveiled dynamics of brain organoid differentiation	74
4.6 Transcriptional footprints of pre-natal cortical development and cortical organoidogenesis strongly overlapped and were partially recapitulated by different brain organoid paradigms...	85
4.7 Definition of the WS hiPSC cohort for disease modelling in relevant cellular lineages	91
4.8 The WS mutational spectrum differentially affected CBO growth revealing patient-specific phenotypes	95
4.9 WS CBO upregulated neuronal maturity genes in early phases and displayed metabolic alterations at later developmental time-points	101
4.10 Integration of transcriptomic and epigenomic WS datasets uncovered a set of genes directly linked to EZH2 LOF relevant for the WS-associated intellectual disability.....	110
4.11 2D-induced neurons did not reveal differences between patients and controls.....	117
4.12 Patient-derived neural crest stem cells show downregulation of migration	119
4.13 Isogenic line engineering for validation of WS molecular phenotypes.....	120
5. DISCUSSION	123
6. MATERIALS AND METHODS.....	135
6.1 Experimental procedures	135
6.1.1 Cell reprogramming	135
6.1.2 hiPSC maintenance	135
6.1.3 CBO generation and maintenance.....	136
6.1.4 NSCS differentiation	137
6.1.5 iN differentiation	137
6.1.6 Culture conditions for fetal cortical cell	138
6.1.7 CBO and hiPSC immunofluorescence	138
6.1.8 Growth curves.....	140
6.1.9 Taqman assays for glycolytic genes	140
6.1.10 RNA extraction and library preparation for RNA-seq.....	140
6.1.11 ChIP-seq	140
6.1.12 DNA methylation profiling.....	142
6.1.13 CRISPR/Cas9 engineering.....	142
6.1.13.1 Donor construct design for isogenic lines generation.....	142
6.1.13.2 sgRNA design	142

6.1.13.3 sgRNAs synthesis.....	143
6.1.13.4 CAS9 purified protein.....	144
6.1.13.5 CRISPR/Cas9 delivery into hiPSC.....	144
6.1.13.6 Hygromycin selection and clones isolation.....	145
6.1.13.7 Genomic DNA extraction and screening.....	145
6.2 Bioinformatic procedures.....	147
6.2.1 BrainSpan.....	147
6.2.2 Dimensionality reduction of BrainSpan pre-natal cortical samples.....	147
6.2.3 Correlation pre- and post-natal cortical samples BrainSpan.....	147
6.2.4 BrainSpan WGCNA.....	147
6.2.5 RNA-seq alignments for CBO, NCSC, iN, internal fetal cortical dataset.....	148
6.2.6 Dimensionality reduction CBO dataset and internal fetal samples.....	148
6.2.7 Dimensionality reduction CBO dataset.....	148
6.2.8 DEA on CBO dataset and functional annotation of DEGs.....	149
6.2.9 Bulk deconvolution CBO dataset.....	149
6.2.10 Cortex-specific genes determination and visualisation of their behaviour in CBO.....	150
6.2.11 CBO WGCNA.....	150
6.2.12 Download and alignment other brain organoid datasets.....	151
6.2.13 Dimensionality reduction for CO, FO and TA datasets.....	151
6.2.14 Bulk deconvolution of CBO and other brain organoid datasets.....	151
6.2.15 Correlation CBO versus other brain organoid datasets.....	151
6.2.16 Correlation organoid datasets versus BrainSpan.....	151
6.2.17 DEA on CO, FO and TA datasets.....	152
6.2.18 Literature-curated gene signatures visualisation in BrainSpan and Brain organoids..	152
6.2.19 Module overlap of BrainSpan and CBO WGCNA.....	152
6.2.20 Analysis of BrainSpan modules in brain organoids.....	152
6.2.21 Analysis of CBO modules in BrainSpan and other brain organoids.....	153
6.2.22 DEAs Weaver CBO and DEGs characterisation.....	153
6.2.23 DEGs overlap and visualisation.....	154
6.2.24 H3K27me3 CHIP-seq analysis.....	154
6.2.25 DNA methylation analysis.....	154
7. LIST OF PUBLICATIONS.....	156
7.1 Published papers.....	156
7.2 Book chapters.....	156
7.3 Awarded grants.....	157
7.4 Submitted papers.....	157
8. ACKNOWLEDGEMENTS.....	158
9. BIBLIOGRAPHY.....	160

LIST OF ABBREVIATIONS

2D: Two-dimensional

ALS: Amyotrophic Lateral Sclerosis

AMPA: Alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-type

ApicalProg: aRG signature

aRG: Apical Radial Glia

Astrocyte: Astrocytic marker signature

Bmp: Bone morphogenetic protein

CAS: CRISPR associated genes

CBO: Cortical Brain Organoids

ChIP: Chromatin Immuno-precipitation

ChIP-seq: ChIP-sequencing

CNS: Central Nervous System

CO: Cerebral Organoid

CP: Cortical Plate

CPM: Counts per million reads

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

crRNA: CRISPR-RNA

cPRC1: canonical PRC1

Cycling: Cell cycle signature

DEA: Differential Expression Analysis

DEGs: Differentially Expressed Genes

DNMTs: DNA methyltransferases

Drosophila: Drosophila Melanogaster

EGF: Epidermal growth factor

ESCs: Embryonic Stem Cells

ExcN: excitatory glutamatergic neuron signature

ExN: excitatory neurons

EZH2-LOF: CRISPR/Cas9 line carrying homozygous frameshift in the SET domain of EZH2

FACS: Fluorescence Activated Cell Sorting

FC: Fold Change

FDR: False Discovery Rate

Fgf: Fibroblast growth factor

FO: Forebrain Organoids

FPKM: Fragments Per Kilobase of transcript per Million mapped reads

GO: Gene Ontology

GOF: Gain of function

GSEA: Gene Set Enrichment Analysis

GW: Gestational Week

H2AK119Ub: ubiquitination of Lysine 119 on histone H2A

H3K4: lysine 4 of histone H3

H3K4me3: Trimethylation of lysine 4 of the histone H3

H3K27ac: Acetylation of the lysine 27 of histone H3

H3K27me1/2/3: methylation of the Lysine 27 on histone 3

H3K36: lysine 36 of histone H3

HAR: Human Accelerated Regions

HAs: Homology arms

HDR: Homologous Directed Recombination

HGE: Human Gained Enhancers

hiPSC: human induced pluripotent stem cells

HOX: Homeobox

HRE: Human Lost Enhancers

ID: Intellectual disability

IKNM: Interkinetic Nuclear Migration

In: inhibitory neurons

iN: Induced neurons

InhN: GABAergic inhibitory neuron signature

IntermediateProg: IP signature

IP: intermediate progenitors

iPSC: Induced Pluripotent stem cells

ISVZ: Inner Subventricular Zone

KD: Knock down

KO: Knock out

LOF: Loss of function

LowLayerN: signature of lower layer neurons

ME: Module Eigengene

NCSC: neural crest stem cells

NHEJ: Non Homologous End Joining

NMDA: N-methyl-D-aspartate receptor

NPCs: Neural Progenitor Cells

NSC: Neural stem cell signature

ON: Overnight

oRG: Outer radial glia

OSKM: Oct3/4, Sox2, Klf4 and c-Myc

OSVZ: Outer Subventricular Zone

PanNeuron: signature for general markers of post-mitotic neurons

PAM: protospacer adjacent motif

PC: Principal Component

PCA: Principal Component Analysis

PcG: Polycomb group of proteins

PCW: Post Conceptional Week

Pg: Cycling progenitors signature

PostSynaptic: signature of genes specifically found in the post-synaptic density

PRC1: Polycomb repressive complex 1

PRC2: Polycomb repressive complex 2

PREs: Polycomb Responsive Elements

PSC: Pluripotent stem cells

RG: radial glia

RNAseq: RNA sequencing

RNase III: RNA-specific ribonuclease

RPKM: Reads Per Kilobase of transcript per Million mapped reads

scRNAseq: single-cell RNAseq

SFEB: Serum-free embryoid bodies

SFEBq: Serum-free embryoid bodies quick

sgRNA: single-guide RNA

Shh: Sonic hedge-hog

Sox: SRY-related HMG-box

SV: Surrogate Variable

SVA: Surrogate Variable Analysis

SVZ: Subventricular Zone

TA: Telencephalic Aggregates

TALENs: Transcription Activator-Like Effector Nucleases (TALENs)

TFs: Transcription Factors

tracrRNA: trans-activating crRNA

TrxG: Tritorax group proteins

UpLayerN: signature of upper layer neurons

UPR: Unfolded Protein Response

vRG: Ventricular radial glia signature

VZ: Ventricular Zone

WBD: WD-40 binding domain

WGCNA: Weighted Gene Co-expression Network Analysis

WS: Weaver syndrome

ZFNs: Zinc-Finger Nucleases (ZFNs)

FIGURE INDEX

Fig. 1: Developmental basis of human brain development.	3
Fig. 2: Schematic representation of mammalian cortical development.	6
Fig. 3: Outer subventricular zone.	8
Fig. 4: Homogeneity vs complexity in brain organoids.	16
Fig. 5: Architecture of the PRC2 complex.	23
Fig. 6: Ezh2 expression along corticogenesis.	25
Fig. 7: Type II CRISPR-Cas system in <i>S. pyogenes</i>	31
Fig. 8: Genome engineering using CRISPR/Cas9.	32
Fig. 9: Cohort of fetal cortical samples from the BrainSpan Atlas dataset.	35
Fig. 10: Correlation analysis of BrainSpan pre-natal and post-natal samples.	36
Fig. 11: Dimensionality reduction of pre-natal cortical samples from BrainSpan.	36
Fig. 12: Characterization of top genes driving sample distribution in PCA.	37
Fig. 13: Gene dendrogram for BrainSpan WGNA module selection.	38
Fig. 14: Correlation of BrainSpan WGCNA modules with stage.	38
Fig. 15: Characterisation of the BrainSpan turquoise module.	40
Fig. 16: Characterisation of the BrainSpan pink module.	40
Fig.17: Characterisation of the BrainSpan midnight-blue module.	41
Fig. 18: Characterisation of the BrainSpan grey60 module.	41
Fig. 19: Characterisation of the BrainSpan magenta module.	42
Fig. 20: Characterisation of the BrainSpan yellow module.	42
Fig. 21: Characterisation of the BrainSpan black module.	43
Fig. 22: Characterisation of the BrainSpan red module.	44
Fig. 23: Characterisation of the BrainSpan blue module.	44
Fig. 24: Graphical representation of the in-house generated CBO and fetal brain datasets.	46
Fig. 25: Dimensionality reduction of CBO, fetal brain tissues and 2D cultured fetal cortical cells.	47
Fig. 26: Dimensionality reduction of the CBO dataset.	48
Fig. 27: Characterization of top genes driving sample distribution in the PCA of the CBO dataset.	48
Fig. 28: Differential expression analysis of the CBO dataset.	49
Fig. 29: Functional analysis of CBO DEAs until day 100.	50
Fig. 30: Functional analysis of CBO DEAs from day 100 until day200.	50
Fig. 31: FC scatter plot CBO day50vs25 against day100vs50.	51
Fig. 32: FC scatter plot CBO day100vs50 against day150vs100.	52
Fig. 33. FC scatter plot CBO day150vs100 against day200vs150.	52
Fig. 34: Bulk deconvolution of the CBO dataset.	53

Fig. 35: Visualisation of cortex-specific upregulated DEGs in the CBO dataset.	54
Fig. 36: Visualisation of cortex-specific downregulated DEGs in the CBO dataset.	54
Fig. 37: Functional analysis of cortex-specific DEGs.	55
Fig. 38: Clustering of cortex-specific DEGs based on their expression trend during CBO differentiation.	56
Fig. 39: Gene dendrogram for CBO WGCNA module selection.	58
Fig. 40: Correlation of CBO WGCNA modules with stage.	59
Fig. 41: Characterisation of the CBO turquoise module.	59
Fig. 42: Characterisation of the CBO black module.	60
Fig. 43: Characterisation of the CBO blue module.	61
Fig. 44: Characterisation of the CBO brown module.	61
Fig. 45: Functional analysis of main CBO WGCNA modules.	62
Fig. 46: Characterisation of the CBO green module.	63
Fig. 47: Characterisation of the CBO red module.	63
Fig. 48: Functional analysis of CBO green and red modules.	64
Fig. 49: Characterisation of the CBO magenta module.	64
Fig. 50: Schematic representation of the brain organoid datasets analysed in this work.	66
Fig. 51. Dimensionality reduction of the CO dataset.	67
Fig. 52: Dimensionality reduction of the FO dataset.	67
Fig. 53: Dimensionality of the TA dataset.	68
Fig. 54: Bulk deconvolution of the CBO dataset.	69
Fig. 55: Bulk deconvolution of the CO dataset.	69
Fig. 56: Bulk deconvolution of the FO dataset.	69
Fig. 57: Bulk deconvolution of the TA dataset.	70
Fig. 58: Differential expression analysis of CO, FO and TA datasets.	70
Fig. 59: Correlation analysis of CBO against the other brain organoid datasets analysed.	71
Fig. 60: Correlation analysis of CBO against the BrainSpan pre-natal cortical samples.	71
Fig. 61: Correlation analysis of CO against the BrainSpan pre-natal cortical samples.	72
Fig. 62: Correlation analysis of FO against the BrainSpan pre-natal cortical samples.	72
Fig. 63: Correlation analysis of TA against the BrainSpan pre-natal cortical samples.	73
Fig. 64: Literature curated signatures for cortical populations.	78
Fig. 65: Literature curated signatures for brain areas, off-target tissues and stress.	81
Fig. 66: Literature curated signatures for neuronal transmission systems.	84
Fig. 67: Gene overlap between BrainSpan and CBO WGCNAs.	85
Fig. 68: BrainSpan turquoise and pink module behaviour across BrainSpan and brain organoid datasets.	86
Fig. 69: BrainSpan grey60, midnight-blue and magenta module behaviour across BrainSpan and brain organoid datasets.	87

Fig. 70: CBO turquoise and black module behaviour across BrainSpan and brain organoid datasets.	87
Fig. 71: BrainSpan yellow and black module behaviour across BrainSpan and brain organoid datasets.	88
Fig. 72: CBO brown and magenta module behaviour across BrainSpan and brain organoid datasets.	88
Fig. 73: BrainSpan red and blue module behaviour across BrainSpan and brain organoid datasets.	89
Fig. 74: CBO green and red module behaviour across BrainSpan and brain organoid datasets.	89
Fig. 75: CBO blue module behaviour across BrainSpan and brain organoid datasets.	90
Fig 76: Collection of WS samples.	91
Fig. 77: Weaver disease-causing mutations.	91
Fig. 78: Representative immunostainings for pluripotency markers.	92
Fig. 79: Graphical representation of the cohort, experimental models and tools used.	92
Fig. 80: EZH2-LOF engineering.	93
Fig. 81: EZH2 levels in EZH2-LOF samples.	94
Fig. 82: CBO growth WS vs controls.	96
Fig. 83: CBO growth WS vs controls divided by batch.	97
Fig. 84: CBO growth WVS02A vs controls.	98
Fig. 85: CBO growth WVS04A and WVS03B vs controls.	99
Fig. 86: CBO growth EZH2-LOF vs controls.	100
Fig. 87: Stage-wise DEA WS vs controls.	101
Fig. 88: Transcriptomic analysis WS vs controls at day 25.	102
Fig. 89: Transcriptomic analysis WS vs controls at day 50.	102
Fig. 90: Transcriptomic analysis WS vs controls at day 100.	103
Fig. 91: PCA and DEA subdividing WS patients in groups.	104
Fig. 92: Functional analysis of DEGs subdividing WS patients in groups.	105
Fig. 93: Transcriptomic analysis EZH-LOF vs controls at day 25.	105
Fig. 94: Representative CBO immunostainings.	107
Fig. 95: Controls, WS and EZH2-LOF immunostainings quantification.	108
Fig. 96: Intersection DEA WS vs controls and EZH2-LOF vs isogenic control.	110
Fig. 97: Overlap with roadmap epigenomic H3K27me3 datasets.	112
Fig. 98: Distribution of H3K27me3 peaks at regulatory regions.	113
Fig. 99: Intersection DEAs and ChIPseq peaks lost in WS.	114
Fig. 100: PCA of DNA methylation profiles of the WS cohort in hiPSC, CBO, NCSC and iN.	115
Fig. 101: DNA methylation profiles at the promoter of relevant targets.	116
Fig. 102: PiggyBac strategy used to overexpress Ngn2 for iN induction.	117
Fig. 103: Heatmap of WS vs control iN DEGs.	117

Fig. 104: Electrophysiological characterisation of iN.	118
Fig. 105: Validation of NCSC differentiation.....	119
Fig. 106: GO analysis for biological processes of NCSC DEGs WS vs controls.	119
Fig. 107: Schematic representation of the donor construct for WS isogenic line generation.....	120
Fig. 108: Screening of CRISPR/Cas9 clones.	121
Fig. 109: Validation and efficiency of isogenic lines engineering.	122
Fig. 110: Schematic representation of the DNA template generation for sgRNA synthesis.	144

TABLE INDEX

Table 1: Genes involved in extracellular matrix organisation.....	57
Table 2: List of hiPSC used in this work.....	135
Table 3: List of antibodies used for immunofluorescence analyses.....	139
Table 4: List of sgRNAs.....	143
Table 5: sgRNA synthesis oligos.	143
Table 6: Primers screening EZH2-LOF line.....	145
Table 7: Screening WVS02A isogenic 1.....	146
Table 8: Screening WVS02A isogenic 2.....	146
Table 9: Screening WVS02A isogenic 3.....	146
Table 10: Screening WVS02A isogenic 4.....	146

1. ABSTRACT

Brain organoid technology is transforming neurobiology and neuropsychiatry by enabling the investigation of human neurodevelopment *in vitro* and thereby making its spatiotemporal dynamics for the first time accessible and experimentally tractable. While an increasing range of protocols, spanning undirected to patterning-driven differentiation, have achieved remarkable degrees of recapitulation of brain development, what is acutely lacking is a comprehensive analysis providing a resource for the field in terms of gene expression patterns specific of pre-natal cortical development and the quantification of their recapitulation in brain organoids. Here we complement the current knowledge with a comprehensive benchmarking of cortical brain organoids *vis a vis in vivo* corticogenesis, integrating: i) our in-house cohort of cortical brain organoids (CBO) from several individuals in multiple differentiation rounds and densely profiled over 200 days, thereby covering the equivalent of early to mid-late gestation while quantifying inter-individual and technical variability; ii) our in-house cohort of brain tissue primary samples, including cortex; iii) meta-analysis of publicly available transcriptomic data from brain organoid protocols entailing different degree of patterning and iv) meta-analysis of the pre-natal cortical samples of the BrainSpan Atlas. The transcriptomic analysis of developing CBO revealed a two-step development composed of a fast-evolving phase, corresponding to robust increase of the glutamatergic neuron differentiation signature, followed by a slow-evolving one, characterised by milder transcriptional changes mainly related to astrocyte development. We found that the timing of this transcriptional evolution was closely recapitulating the physiological one, in terms of global transcriptome, appearance of key cell populations, and preservation of gene expression patterns, while we observed some degree of heterochronicity when considering other protocols, with a more compressed differentiation when analysing comparable early time-points of *in vitro* development. Cross-visualisation of the trends followed by main monotonic gene co-expression patterns of BrainSpan in CBO and *vice versa* revealed concordance in their behaviour, a further confirmation of the fidelity of CBO in recapitulating the fundamental transcriptional dynamics of the human developing cortex. Importantly, this was also true for the other brain organoid paradigms analysed, however when we focused on modules not following a strictly monotonic trend, we observed that CBO more closely resembled pre-natal cortex. These non-monotonic patterns were mainly related to extracellular matrix organisation, which we confirmed to be preserved in CBO by multiple independent analyses. Upon characterization of CBO differentiation dynamics, we used them to investigate, as case study, Weaver syndrome (WS), a rare congenital disorder characterized by overgrowth, craniofacial abnormalities and intellectual disability associated to malfunction of the Polycomb Repressive Complex 2 (PRC2), a well-known repressor of gene expression in vertebrates. While much is known

about the function of its catalytic subunit EZH2 in orchestrating developmental programs, there is no understanding of how the heterozygous mutations found in WS bring about disease phenotypes. The reason is twofold: i) first, as most functional studies were performed in Ezh2 null mice and cells, we do not know how one mutant allele impacts the global landscape of H3K27 methylation; ii) second, virtually nothing is known about the possible mechanisms through which PRC2 can cause intellectual disability, a major feature of WS that poses a significant burden for patients and families. By means of patient-derived CBO profiled up to 100 days *in vitro* at a transcriptomic and epigenomic level, we aimed of uncovering the molecular circuitries underpinning disease-relevant phenotypes. With this approach, we found that, unlike 2D differentiated neurons, CBO were able to highlight the molecular features of the disease. Indeed, patient-derived CBO upregulated genes involved in neuronal maturation and migration as well as displayed alteration of glucose metabolism at later stages of development. Intersection of the differentially expressed genes between WS and control CBO with dysregulated genes in a CRISPR/Cas9 engineered model of severe EZH2 loss of function, H3K27me3 ChIP-seq peaks, and DNA methylation profiles revealed a set of bona-fide PRC2 targets possibly responsible for the WS intellectual disability phenotype. Lastly, we found that patient-derived neural crest stem cells (NCSC) carry transcriptional dysregulation of genes involved in neural crest migration, grounding the hypothesis that the craniofacial defects characteristic of WS have their molecular roots at the NCSC level.

Together, our results contribute to the definition of transcriptional footprints and dynamics specific of pre-natal cortical development, representing a collection of prioritised known and novel hubs that we categorised in well-defined functional domains. Our approach describes the extent of *in vivo/in vitro* alignment of developmentally relevant processes and temporality, highlighting commonalities and diversities of different brain organoid paradigms and providing a resource available for consultation when modelling physiological or pathological human cortical development. We exploited CBO to get insights into the role of PRC2 in the onset of WS both in terms of general pathways dysregulated upon its mutation as well as in terms of the identification of druggable targets for which we brought evidence of their involvement in the WS-related intellectual disability phenotype, providing the framework for translational studies.

2. INTRODUCTION

2.1 Developmental basis of human neocortical development

Human brain development is a very complex process that takes place pre- and post-natally until adulthood. In humans, the embryonic period ends after post conceptional week 8 (PCW8) with the development of the central nervous system (CNS) backbone, constituted by transient structures that eventually mature into final ones. After gastrulation, the embryo is constituted by cells organised in multiple layers referred to as endo, meso and ectoderm, which will give rise to all tissues of the foetus. The nervous system develops from neuroectoderm, a peculiar portion of the ectoderm that undergoes neuralisation from PCW3, when neural progenitor cells (NPCs) start to appear. Neuralisation starts with the formation of the neural tube, where NPCs organise in apico-basal orientation leaving the apical portion facing towards the cavity, which in turn will give rise to the ventricular system. Between PCW3 and 8, three protuberances along the antero-posterior axis appear (prosencephalon, mesencephalon and rhombencephalon) and later divide into five the secondary structures constituting the primordium of the brain^{1,2,3,4}, as outlined in **Fig. 1**. During the fetal period, the development of cortical and subcortical areas takes place and completion of developmental neurogenesis is largely achieved by mid-gestation^{5,6,7}. New-born cortical neurons migrate in a sequential fashion to form the six layers of the neocortex and, once reached their final position, they differentiate by extending dendritic and axonal processes. By the end of the fetal period, most of the major pathways including cortico-cortical and thalamo-cortical connections are established^{5,8}).

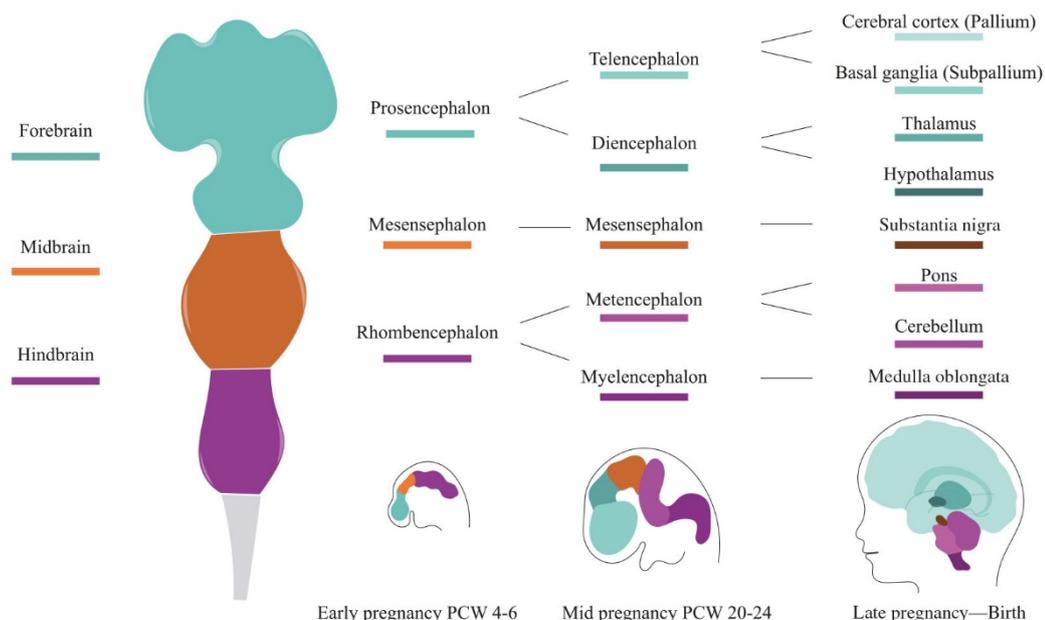


Fig. 1: Developmental basis of human brain development. Adapted from⁹.

2.1.1 Molecular features of mammalian cortical development

The cerebral cortex is defined as a bilateral structure located in the superficial portion of the cerebral hemispheres¹⁰. The largest region of the cerebral cortex is the neocortex, whose neurons are organised in six layers. The interplay between intrinsic genetic programmes and extrinsic morphogenetic patterning results in the expansion, layering, folding and functional organisation of the cortex¹¹. Initial patterning of cortex into antero-posterior and dorso-ventral axes results from diverse molecular signals derived from different regions of the neocortical proliferative zone, which are encoded by intrinsic genetic programmes. Seminal work using mice lacking thalamic afferents showed a near complete preservation of the molecular and anatomical organisation of the cortex¹², highlighting the robustness of the genetically encoded regionalisation. In addition to genetically encoded programmes, different patterning centres secrete positional cues resulting in gradient expression of transcription factors (TFs) important for the establishment of anterior-posterior and dorso-ventral patterning¹³.

Patterning centres along the anterior and posterior midline express opposing gradients. The anterior neural ridge (which becomes the commissural plate) expresses different types of Fgfs (mainly Fgf8 and Fgf17), while the cortical hem (located posteriorly) secretes Bmps and Wnts. Bmps signaling represses Fgf8 expression, which in turn represses Bmp4 and Wnt8b. A ventral patterning centre, located in the ventral telencephalon, expresses Sonic hedge-hog (Shh), implicated in regional patterning of the forebrain¹⁴. Finally, the antihem, located close to the boundary between dorsal and ventral telencephalon, produces Tgfa, Neuregulin1, Neuregulin3, Fgf7 and Sfrp2¹⁵. Cortical hem and antihem cooperate with the anterior neural ridge to establish anterior-posterior and lateral-medial axes. These opposite gradients impact on the expression of several master transcription factors, including COUP-TFI, Emx2, Pax6, and Sp8, whose controlled expression across the embryonic cortical axes determines size and position of cortical areas by specifying or repressing areal identities¹⁶. On top of this, neuronal determinants regulate cell fate in the developing cortex. For example, the balance between dorsal excitatory and ventral inhibitory neurons is mainly regulated by a transcriptional axis linked to the expression levels of FOXG1 and NKX2.1, respectively, and by the cascade of TFs signaling deriving from their activity^{17,18}. Within the excitatory glutamatergic neurons forming the six layers of the neocortex, laminar organisation is controlled by layer-specific transcription factors. Sox5 and TBR1 are genes specifically expressed by early-born neurons that will occupy layers 6 and 5. On the other hand, late-born neurons differentiation is characterised by the expression of SATB2 or POU3F2 and POU3F3, highly enriched in neurons of layers 2-5¹⁹. Finally, the positioning of neurons and the establishment of specific connections largely depend on cytoskeleton regulation by modulators such as doublecortin (DCX) and tubulin subunits (TUBA1A, TUBB2B and TUBB3), among others^{20,21}.

The exclusive expression of specific genes is not sufficient for explaining the enormous cell types diversity of the cortex, and in general of the CNS. Indeed, almost 90% of the protein coding genes and a plethora of non-coding genes contribute to the specific cell fate determination of all CNS cell types²².

2.1.2 Neurogenesis in the developing mammalian neocortex

The cerebral cortex develops from the dorsal part of telencephalic vesicles, referred to as pallium¹⁰. In all mammalian species, the neocortex is composed by six layers of pyramidal (or projection) glutamatergic neurons, with GABAergic interneurons originating in the ventral telencephalon tangentially migrating into the neocortex and organising in horizontal layers²³. Pyramidal neurons derive from the migration of new-born neurons originated in the proliferative zones: i) ventricular zone (VZ) and ii) subventricular zone (SVZ)^{24,23,25,26}.

During development, the neuroepithelium organises forming the VZ, where neural stem cells, called apical radial glia (aRG), organise radially and coordinate neurogenesis and neuronal migration towards the cortical plate (CP). Thus, aRG directly derives from the neuroepithelium and owns two fundamental characteristics: i) radial organisation, with cell bodies situated in the VZ contacting both the ventricular and the pial surface; ii) expression of glial markers such as the glial fibrillary acid protein (GFAP) and the presence of glycogen granules²⁷. The contact with the ventricular surface is mediated by a single primary cilium, while the radial glia cell body emits a long extension, called radial glia fibre, towards the pial surface²⁸. Imaging studies demonstrated that intermediate progenitors (IP) in the SVZ derive directly from aRG²⁸. Neurons can be generated both directly from aRG, or indirectly through the generation of IP. Indirect neurogenesis led to the amplification of the final number of neurons of the cerebral neocortex in mammals^{29,30,31,32}. During their cell cycle, aRG exhibit a stereotyped behaviour referred to as interkinetic nuclear migration (IKNM)^{33,34}. The cell body shuttles up and down in the VZ during different phases of the cell cycle. During G2 phase, the nucleus descends from the most basal portion of the VZ to undergo mitosis at the ventricular surface. The cleavage plane of aRG cells is controlled so that the self-renewed cell maintains both apical and basal structures, while the committed cell loses apico-basal polarity and delaminates becoming an IP³⁵.

The formation of the neocortex proceeds with an inside-out pattern, meaning that deeper layers are generated before than superficial layers³⁶. Progenitors located in the VZ or SVZ divide according several modes. Indeed, they can divide: i) symmetrically, increasing the pool of neural progenitors; (2) asymmetrically, maintaining the pool of progenitors and at the same time generating one committed cell; or (3) symmetrically, depleting the pool of progenitors by originating two committed cells. Almost all IP divide symmetrically, with about 90% of these divisions generating two neurons and the remaining 10% generating two IP²⁹. Neurons then migrate along the radial

fibres to reach their final position within the CP, assuming a characteristic bipolar morphology. This process is called radial migration or locomotion migration and it was the only migratory model accepted for years. Later, other studies shed light on the so called multipolar migration, with neurons extending and retracting multiple processes searching for positional cues while reaching slowly the cortical plate³⁷. Before reaching their final destination, multi-polar neurons come back to the radial migration. Somal migration is another type of migration observed in the developing neocortex that consists in new-born neurons extending a long ascending process towards the pial surface, which is then retracted to elevate the cell soma^{37,38,26}. Following neurogenesis, astrocytes are generated from the very same progenitors, which undergo a switch towards the gliogenic differentiation programme through activation of Notch1 and Jak/STAT signaling pathways as well as expression of specific transcription factors (TFs) such as SOX9 and NF1A^{39,40,28}). A summary of the mammalian cortical development is outlined in **Fig. 2**.

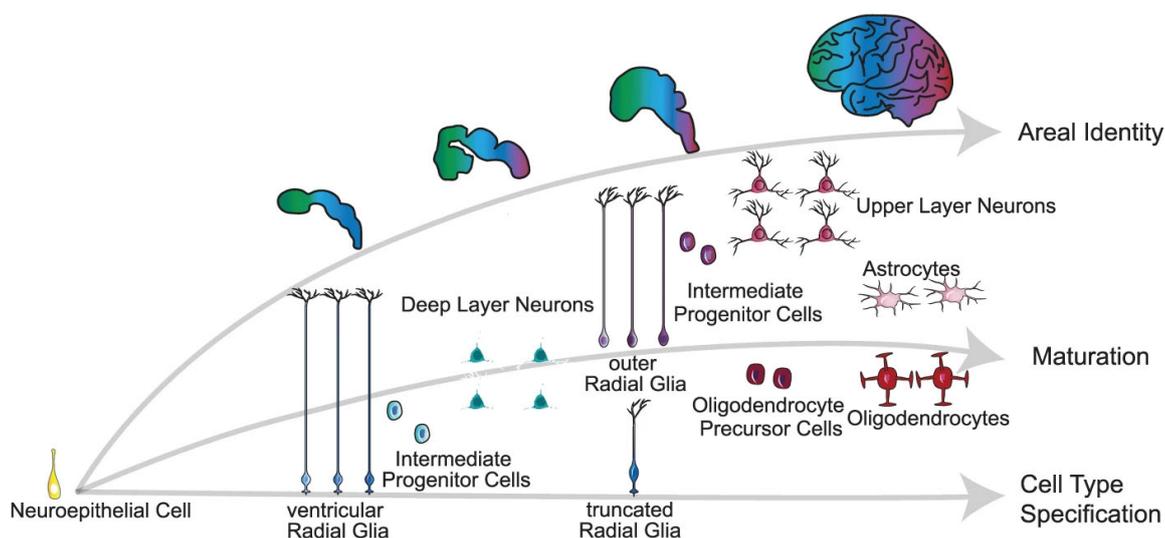


Fig. 2: Schematic representation of mammalian cortical development. Adapted from⁴¹.

2.1.3 Outer subventricular zone

The organisation of the SVZ differs between lissencephalic and gyrencephalic mammals. Indeed, while the SVZ in gyrencephalic mammals, particularly in primates, is very expanded and subdivided into discrete cytoarchitectural regions called inner SVZ (ISVZ) and outer SVZ (OSVZ), the SVZ of rats and mice is much thinner and less organised^{42,36,34} (**Fig.3**). Studies of thymidine labelling in primates demonstrated that the proliferation within the OSVZ coincides with the major wave of cortical neurogenesis, suggesting that the OSVZ contributes to neuron production^{35,31}. The outer subventricular zone contains a class of neural stem cells called outer, or basal, radial glia (oRG), which express RG markers such as nestin and GFAP^{43,44}. This cellular type divides mainly asymmetrically to maintain the pool of progenitors and to generate committed neurons. Morphologically, oRG cells are unipolar, with a unique process that extends basally towards the pial

surface. Interestingly, mTOR signaling has been shown to regulate the cytoskeletal organisation of oRG cells for the maintenance of this peculiar morphology⁴⁵. Molecularly, oRG cells preferentially express genes related to extracellular matrix formation, migration, and stemness, including TNC, PTPRZ1, FAM107A, HOPX, and LIFR⁴⁶.

In humans, cortical neuron production begins by gestational week (GW) 6 and the OSVZ does not arise until GW11. Over the following 6 weeks, the OSVZ expands dramatically to become the predominant germinal region of the neocortex. Cells constituting the OSVZ lineage can be subdivided into three population: (1) oRG cells, (2) intermediate progenitor cells that do not have oRG cell morphology, but express markers of the undifferentiated state, and (3) immature neurons that account for about 45% of the OSVZ cell population³¹. During evolution, brain size have increased exponentially. In mammals this can be explained by the two-step neurogenesis involving intermediate progenitor⁴⁷. Further increase of neuronal output in the human cortex was achieved by: i) the presence of oRG cells, ii) multiple rounds of oRG asymmetric divisions to generate other oRG cells, and (3) increased proliferative capacity of IP cells³¹.

Despite being initially characterised as an almost exclusively neurogenic population, oRG has been shown to have gliogenic potential in macaque. Indeed, a transition from neurogenesis to oligodendrogenesis between E70 and E92 of macaque embryonic development was observed. Further studies are needed to determine the astrogenic role of oRG, as at the time when GFAP+ astrocyte differentiation begins in macaques, researchers have not been able to determine if cortical astrocytes predominantly originate from VZ or OSVZ progenitors⁴⁸.

Most of our current knowledge about cortical development comes from animal models. Even though many step of neurogenesis have been addressed in those models, they obviously cannot fully recapitulate human neocortex development. Therefore, the need of generating human models has become prominent along years.

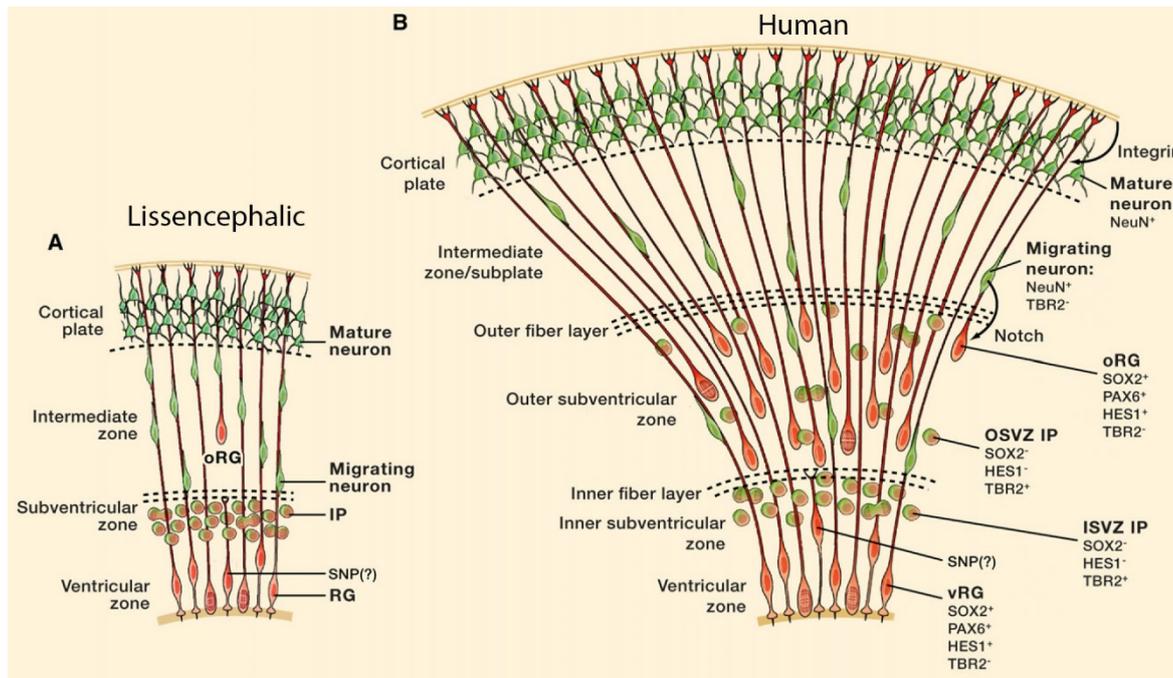


Fig. 3: Outer subventricular zone. (A) Lissencephalic corticogenesis. (B) Human corticogenesis. Adapted from³¹.

2.2 Investigating human brain development

Being human brain development one of the most complicated developmental processes entailing the interplay between genetics and environment, which shape the complex series of dynamic processes promoting the emergence of diverse set of cell types, structures and functions, the need of performing specie-specific studies increased over time. A large body of work coming from animal models has provided immense knowledge about these processes, but it is obvious that human-specific features cannot be fully understood without using human-specific models. A clear example is the rate of failure of drug development for neurological disorders, with 1 every 99 compounds coming from animal studies successfully tested on humans for neuropsychiatric disorder treatment⁴⁹.

2.2.1 Primary and post mortem samples

For decades, the study of cellular and molecular architectures of the developing human CNS has relied on the availability of fetal primary tissue, with several limitations hampering the progress of the field: i) limited access to this kind of specimen, particularly at early developmental stages, ii) difficulty in *in vitro* maintenance, with complete maturation impossible to achieve or sustain for long periods, iii) ethical constrains. Successful culture of primary dissociated fetal neurons was achieved with serum-free conditions. These two-dimensional cultures contained morphologically differentiated neurons with complex dendritic arbours and cell-cell contacts as well as synapses⁵⁰, yet this complexity has not been reliably reproduced. For this reason, researchers focused on the harvest of primary cells without passaging in culture to classify human brain cells types, including neuronal, glial and vascular lineages, and to perform molecular studies^{51,52}, which however are

devoid of any morphological information. Despite these difficulties, there are examples of studies performed on very long cultures of primary neurons. Single cell RNA sequencing of neurosurgically resected neurons from adult individuals, cultured in two dimensions for almost 90 days, revealed that neuronal-subtype-specific genes identified from animal models are often lowly expressed or even absent in adult human neurons, while more complex signatures including non-coding RNAs are more informative of subtype specification⁵³. Epigenomic characterisation of human primary fetal brain tissue revealed human-specific features such as increased presence of super enhancer sites with an average distance from target genes higher than in mice, as well as increased frequency of bivalent domains⁵⁴. Moreover, mapping of the trimethylation of lysine 4 of the histone H3 (H3K4me3) across primates identified human-specific enrichment or depletion⁵⁵ at around 500 *loci*, which interestingly do not define specific functions, but rather identify genes with multifaceted roles⁵⁶.

Despite providing paramount insights into human brain development, primary and post-mortem tissues entail limitations such as the lack of resolution given by the intrinsic cellular heterogeneity of the brain. Most studies indeed rely on the analysis of bulk homogenates resulting in measurements of average effects rather than cell type specific information^{57,58,59}. One alternative has been the characterisation of dissected sub-areas or the use of fluorescence activated cell sorting (FACS) for the isolation of more homogeneous cell populations^{60,61}. Nonetheless, these approaches do not allow to overcome the issue of post mortem tissue decay due to tissue autolysis and processing, which has been shown to be particularly relevant when studying some types of histone modifications and higher-order chromatin structures^{62,63,64}.

A significant breakthrough came with the development of methods that allowed the isolation and propagation of progenitors from the central nervous system in defined media conditions. Expansion in culture of the proliferating populations of the striatum SVZ in aggregates defined the cellular model known as primary neurospheres⁶⁵. Key modifications in the culture conditions that allowed neuronal cells to grow as neurospheres were the development of a well-defined medium (i.e., serum-free and high in EGF) and the culture in ultralow attachment conditions. These innovations allowed the formation of clonal spheres that can be serially passaged, increasing the total number of cells generated over time. Neurospheres became a promising platform for drug testing⁶⁶ and later gained particular prominence as an efficient method to isolate and expand tumor cells. Indeed, following the hypothesis that a subpopulation of tumour cells is responsible for the origin, maintenance, and resistance of solid tissue malignancies and given the potential similarity between somatic NPCs and cancer-like stem cells, Ignatova et al. pioneered the use of neurospheres to isolate and expand cells from adult human brain tumors^{67,68}.

2.2.2 Induced Pluripotent Stem Cells

Pluripotency is defined as the capacity to generate all three embryonic germ layers (endo-, meso- and ecto-derm) and to self-renew indefinitely⁶⁹. In mammalian cells, this feature is observed only in the inner cell mass (ICM) of the blastocyst, during a very short time-window of early embryonic development⁷⁰. Induced pluripotent stem cells (iPSC) are reprogrammed from adult somatic cells by overexpressing fundamental factors required to maintain pluripotency. iPSC are a transformative tool for the derivation and study of disease-relevant cell types.

2.2.2.1 Cell reprogramming

The isolation of mouse embryonic stem cells (ESCs) took place in the early 1980s⁷¹ and led to the optimisation of culture conditions to maintain their capacity to differentiate into all the embryonic germ layers⁷⁰. The theory developed by Waddington according to which cell differentiation entails irreversible steps that make impossible to revert cell fate commitment was considered true until the advent of reprogramming techniques^{70,72}. In 1952 the first successful nuclear transfer was performed in *R. Pipiens* injecting blastula cells into enucleated eggs. Interestingly, the efficiency of the technique decreased with cells derived from more mature embryos, leading to the conclusion that when committed to specific fate during development, nuclei of differentiated cells lose part of the information to generate the entire embryo, thus making the process irreversible⁷³. In contrast, in 1962, John Gurdon proved that somatic cell nuclear transfer in enucleated eggs was possible, thus confirming that epigenetic reprogramming is achievable by exposing somatic cells to a totipotent environment^{74,75}. These data were confirmed by multiple evidence culminating with the cloning of a mammal for the first time in 1997 and with the finding that ESCs contain factors able to induce reprogramming of somatic cells in 2001^{70,76}. All these reprogramming methods are defined as direct.

Other approaches, defined as indirect reprogramming, require the introduction of factors leading to de-differentiation of somatic cells. The first example was given by Harold Weintraub in 1987, who showed that fibroblasts can be converted into muscle lineage by transduction with the MyoD gene⁷⁷. In 2006, studies by Takahashi and Yamanaka pioneered the field of iPSC, achieving reprogramming of mouse embryonic fibroblasts by overexpressing only four TFs. They surveyed the overexpression of twenty-four candidate genes and, through withdrawal of single factors, narrowed down the list to four pivotal genes essential for maintaining pluripotency. In particular, they generated a reporter system in which a gene specifically expressed in pluripotency, Fbx15, was fused to neomycin resistance. With this system, Oct3/4, Sox2, Klf4 and c-Myc (OSKM) were identified as the key factors whose expression is required to reprogram somatic cells into iPSC⁷². POU5F1 (Oct3/4) is a POU family transcription factor expressed in ESCs, early embryos and germ cells. Oct3/4-null embryos die in utero at peri-implantation stages of development and *in vitro*

culture of the inner cell mass of homozygous mutant blastocyst produces only trophoblast lineages. Sox2 is a Sox (SRY-related HMG-box) family transcription factor expressed in ESCs, early embryos, germ-cells and neural stem cells. Sox2-null embryos die because of failure in the primitive ectoderm (epiblast) development. c-Myc is a helix-loop-helix/leucine zipper transcription factor playing important roles in self-renewal and maintenance of pluripotency in mouse ESCs. c-Myc has a large number of binding sites in the genome and is thought to modify chromatin structure and activate expression of some miRNAs. Klf4 is associated with both tumour suppression and oncogenesis. Klf4 is involved in self-renewal of mouse ESCs, and it is known to cooperate with Oct3/4 and Sox2 to activate the Lefty1 core promoter. Other cocktails such as, OCT4 and SOX2, NANOG and LIN28, can be used to achieve the generation of iPSC^{78,79}.

As of today, iPSC reprogramming is performed with non-integrating methods, but the field developed along the years starting from integrating strategies requiring retroviral or lentiviral vectors carrying the OSKM reprogramming factors, with the disadvantage of possibly disrupting gene expression networks due to their random integration⁸⁰. For this reason, non-integrating methods took place. Three non-integrating methods are currently widely used in the field: i) Sendai-viral reprogramming, in which Sendai viral particles are used to deliver OSKM factors; ii), Epi reprogramming, in which expression of reprogramming factors is achieved by Epstein-Barr virus-derived sequences that facilitate episomal plasmid DNA replication in dividing cells and iii) mRNA-mediated reprogramming, which relies on the delivery of *in-vitro* transcribed mRNAs encoding OSKM factors. Each method has advantages and disadvantages and the choice depends on each laboratory requirement in terms of cell types to be reprogrammed, input material available, timing of reprogramming needed⁸¹. Pluripotency of established iPSC lines by assessing the expression of pluripotency-specific genes and by assessing their differentiation capacity into the three-germ layers *in vitro* or *in vivo*⁸².

2.2.2.2 Advantages and disadvantages of iPSC use

In the seminal work of Takahashi and Yamanaka, they assessed whether iPSC were identical to ESCs. Global gene expression and DNA methylation status showed that this was not the case⁷². Accordingly, many reports showed that iPSC, despite being pluripotent, retain some epigenetic features of somatic origin. Other studies, instead, did not show any difference between these two types of pluripotency. In support of the latter claim, it has been reported that when iPSC colonies emerge, reprogramming is usually not complete. Indeed, in early passages iPSC show differences of gene expression and epigenetic status compared to ESCs maintained for a longer period in culture. Continuous passaging of iPSC allows to get rid of transcriptional, epigenetic and functional differences between iPSC and ESCs^{83,70}.

iPSC have shown to be a transformative tool in research for multiple reasons. First, they do not face the ethical issue of manipulating an embryo for their derivation, contrarily to ESCs. Second, the possibility of generating pluripotent stem cell lines directly from patients allows the differentiation and study of disease-relevant human cell types directly in the diseased genetic background and to generate isogenic controls where disease-causing genetic lesions are reverted through genome editing^{84,85,70,86}. Other big advantages of iPSC are their application in regenerative medicine and in drug discovery^{87,70}.

Despite the valuable advantages, iPSC show some disadvantages. First, reprogramming procedures generate genetic and epigenetic abnormalities in culture⁸¹. Moreover, clinical applications of iPSC result in high medical costs and in highly time-consuming procedures for personalised patient treatment⁸⁷.

2.2.2.3 Neural differentiation using pluripotent stem cells

Researchers spent years refining PSCs differentiation into neurons. PSCs can be grown as embryoid bodies (EBs) that, upon prolonged culture, contain ectodermal, mesodermal and endodermal derivatives. In 1994, the group of David I. Gottlieb for the first time showed that EBs exposure to retinoic acid gave rise to neural cells⁸⁸ and the group of Austin Smith purified neural precursors from retinoic acid exposed EBs in 1998⁸⁹. Although this method produced a good proportion of neural cells, EBs contained different kind cell types, including mesodermal and endodermal lineages⁹⁰. For this reason, researchers developed another technique entailing the co-culture of ESCs with stromal cells favouring differentiation; this strategy allowed the derivation of dopaminergic neurons⁹⁰ and of neural rosettes further differentiated in motoneurons⁹¹, but still required stromal cells releasing not chemically-defined factors. In the latter study and many others, scientists interfered with the SMAD pathway to improve neuronal differentiation^{92,93,92}, but only in 2009 full neuroectoderm conversion was achieved by dual SMAD inhibition⁹⁴, representing the foundation of the most widely used methods for neuronal induction. Despite these seminal achievements, differentiating cultures take usually months to reach neuronal maturity, limiting large-scale studies. For this reason, researchers developed TFs over-expression based methods and small molecules combinations to obtain mature neurons in less than one month^{95,96,97}.

Although the vast majority of the scientific community focused in optimising neuronal differentiation, many labs have invested in developing protocols to derive monolayer cultures of astrocytes, oligodendrocytes and microglia from PSCs^{98,99,100}, opening the possibility to study non-neuronal phenotypes in neurodevelopmental disorders.

Lastly, in recent years, PSCs-derived three-dimensional cultures referred to as organoids have been shown to integrate distinct neural cell types in dynamic cultures resembling the physiology of brain development.

2.3 Brain organoids to model human brain development

Organoids are defined as three-dimensional structures that recapitulate the spatio-temporal development of the particular organ they are intended to resemble. Indeed, organoids are generated from renewable cellular sources that, by differentiating in most of the cell types composing a given organ, organise spatially in functional units similar to the physiological ones and with the same temporal sequence found *in vivo*. These experimental systems offer an unprecedented level of resolution in the study of human organ development physiology and they have shown particular relevance for the brain. With the advent of iPSC technology, human cell types previously inaccessible have become experimentally tractable. In the last 15 years, our ability to push iPSC differentiation into neural cell types has greatly improved and we are now able to direct their differentiation with precise temporal patterning, also taking advantage of cell-autonomous differentiation programmes. The advent of brain organoids has been a long journey starting from our ability to differentiate neuronal precursors organised in neural-rosettes resembling the neural tube structure using FGF2 as induction factor¹⁰¹. In 2005, the lab of Yoshiki Sasai demonstrated that ESCs aggregates cultured in suspension and serum-free conditions (SFEB culture) with the addition of a basic neuronal differentiation medium underwent differentiation into forebrain neural precursors⁹². An improvement of this protocol with controlled and quick aggregation (SFEBq culture) was published later with more reproducible results¹⁰². Later, the same lab further optimised SFEBq cultures by adding TGF β and Wnt inhibitors for the first period of culture to enhance telencephalic differentiation, obtaining also oRG cells and therefore paving the way for human-specific studies using 3D cultures¹⁰³. Contextually with this work, Lancaster et al., developed the so-called cerebral organoids. The novelties of this culture system were the use of matrigel as scaffold to embed differentiating 3D structures and the improvement of culture conditions using spinning bioreactors to enhance nutrient assimilation. Most importantly, this method did not rely on specific patterning molecules, but rather on basic neuronal induction medium, thus allowing cells to unleash their cell-autonomous differentiation potential. Indeed, cerebral organoids resemble many regional identities of the brain, including forebrain, midbrain and hindbrain, and include oRG cells. These characteristics lead to define cerebral organoids also as mini-brains or whole-brain organoids. In 2015, Pasca et al. published a protocol for the generation of so-called cortical brain organoids. The peculiarity of this method is the absence of extracellular scaffolding as well as heavy patterning towards dorsal telencephalon through sequential exposure to dual SMAD inhibitors for neuroepithelial specification, EGF-FGF2 for neural progenitor expansion and the

neurotrophins BDNF and NT3 for cortical glutamatergic neurons differentiation. Mature cortical brain organoids are composed of neurons from deep and superficial cortical layers, astrocytes as well as oRG cells¹⁰⁴. Another brain organoid system published the same year as a modification of the SFEBq has peculiar characteristics given by a two-step differentiation where neural rosettes are differentiated in 2D and then picked and transferred in suspension for terminal patterning towards forebrain, the so called telencephalic aggregates¹⁰⁵. Lastly, in 2016, forebrain organoids were generated in 3D-printed multi-well spinning bioreactors, with significant production scale-up. This differentiation protocol includes dual SMAD inhibition, matrigel embedding and patterning towards forebrain, as well as inhibition of GS3K to promote cell survival¹⁰⁶. These seminal papers paved the way for the explosion of the brain organoids field, with set-up of many different protocols aimed at improving culture conditions and reproducibility as well as at capturing different features of human brain development. Examples of these developments are: i) fusion of brain organoids patterned towards different brain regions^{107,108,109,110}; ii) polarised organoids, which thanks to a fixed patterning center overexpressing SHH get certain degrees of self-regionalisation¹¹¹; iii) vascularised organoids, where endothelial cells are co-cultured with brain organoids allowing the formation of vascular structures within them¹¹²; iii) air-liquid interface organoids, where organoids are sliced and let grow at the air-liquid interface for better nutrient and oxygen diffusion¹¹³; iv) organoids on a chip, thanks to which researchers performed studies on cerebral cortex folding¹¹⁴; v) cerebellar organoids¹¹⁵; iv) midbrain organoids¹¹⁶.

The diversification of brain organoid protocols allows to perform more tailored disease-modelling studies by choosing the experimental set-up that best suits the biological question under examination. Nonetheless, this wealth of possibilities generates the need of comprehensive benchmarking against the reference tissue and systematisation of each model characteristics.

2.3.1 Cross-comparison of different brain organoid protocols

The determination of the degree of organoid approximation to the actual human fetal brain development has been crucial for scoring the trustworthiness of these models for studying brain disorders such as autism and intellectual disability. Several studies contrasted the organoid differentiation method used against the human fetal brain demonstrating its suitability to study neuronal differentiation, with a particular focus on neocortical development. Transcriptionally, brain organoids include the populations found in the human developing cortex¹¹⁷ and different methods have been shown to mimic early embryonic to early fetal stages, with plateau in the emergence of relevant cellular types after three months of differentiation^{118,119}.

Cross-comparison among protocols uncovered peculiarities of brain organoids *vis a vis* human fetal cortex, as for example the overexpression of ECM-related genes in 2D-cultured neurons and in brain

organoids that require a first steps of differentiation in 2D¹²⁰. Other integrative analyses started to shed light on the neuronal differentiation dynamics occurring in brain organoids, pointing at the existence of different bypasses through which neurons arise, not excluding the possibility that culture conditions may determine the propensity towards one rather than others¹²¹. Moreover, the vulnerability of organoids cultures in terms of cellular stress was reported for several brain organoid differentiation methods^{122,41}. Despite this body of research has lately grown exponentially, we still lack a comprehensive understanding of the transcriptional hubs mediating pivotal developmental transitions during brain organoidogenesis, let alone a precise quantitation of brain organoids developmental timings with respect to primary human cortical tissue. The latter point has enormous relevance for disease-modelling using brain organoids, given the emergent convergence of molecular phenotypes of delay or acceleration in neuronal differentiation characterising different clinical conditions¹²³.

2.3.2 Brain organoids: reproducibility and culture-dependent stress

Given the complexity of brain organoids culture in term of cell types composition and spatio-temporal organisation to be recapitulated, reproducibility of brain organoids is a pre-requisite for studying them, thus representing a long-standing topic in the field. As previously mentioned, brain organoids can be classified in two groups based on the use of patterning molecules directing differentiation through specific cell fates. Not surprisingly, patterning increases homogeneity of brain organoids in terms of cells types and their proportions at expenses of complexity and *vice versa*¹²⁴ (**Fig. 4**). The same principle applies to the concept of reproducibility, which is more controllable in contexts exogenously restricting organoids' differentiation potential rather than almost completely relying on cell-autonomous processes. Studies on the reproducibility of cortical brain organoids, which rely on a strict patterning in the first phases of differentiation, across lines and batches of differentiation revealed a very high correlation among organoids¹²⁵. Instead, single cell RNA seq analysis of 31 whole-brain organoids derived from a healthy human iPSC (hiPSC) lines revealed that some clusters of cells were represented only in a subset of organoids analysed, while the majority were consistently represented, suggesting the presence of cell types more reproducibly generated than others. Importantly, authors found that organoids deriving from the very same bioreactor where showing more similarity among themselves, pointing at the growth environment as a key element in controlling differentiation¹²⁶. Patterning towards dorsal forebrain applied to the same protocol increased dramatically the reproducibility across batches and hiPSC lines¹¹⁹. Importantly not-patterned organoids are more likely to produce off-target lineages, such as mesodermal cells¹²⁶.

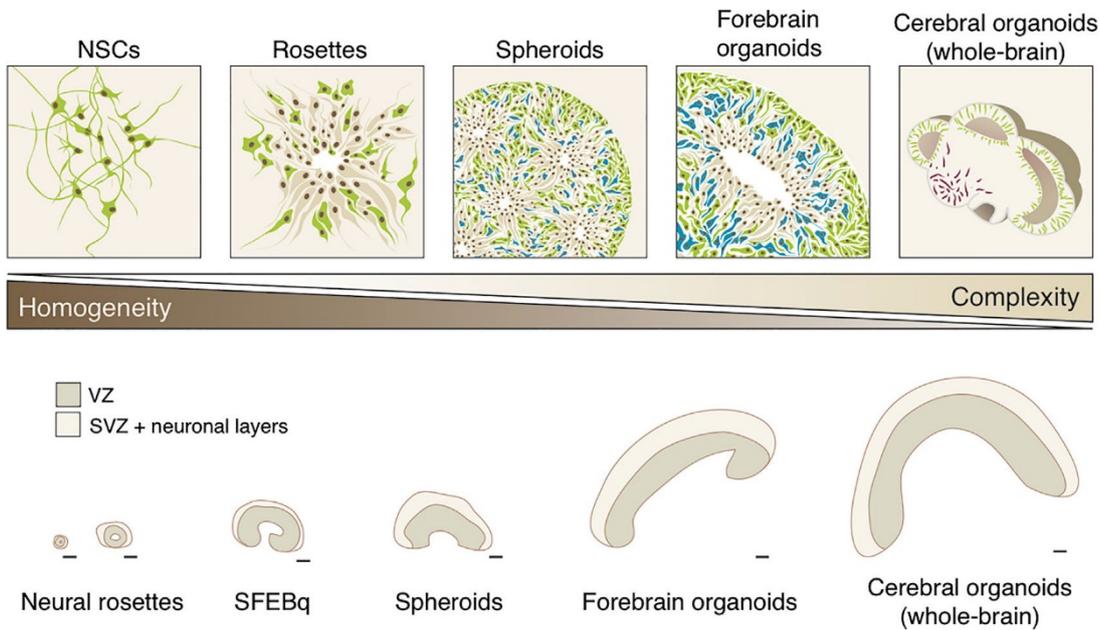


Fig. 4: Homogeneity vs complexity in brain organoids. Adapted from¹²⁴.

Another long-standing matter of discussion in the field is culture conditions, which have been reported to induce stress response within brain organoids¹²². Indeed, comparison between brain organoids with different degree of patterning against the human fetal brain revealed increased expression of glycolytic markers and ER stress genes, suggesting that culture conditions may impact on organoids' metabolism and differentiation dynamics¹²². It has been demonstrated that proportion of cell types within brain organoids compared to the ones of human fetal brain differs at comparable developmental weeks. Organoids show less intermediate progenitors, outer radial glia and post-mitotic neurons (in particular upper layer neurons), while being enriched in apical neuronal progenitors⁴¹. Interestingly, primary fetal cortical cells grown as aggregates in organoid culture conditions acquired expression of the stress-related signature and, *vice versa*, organoid cells transplanted into mice came back to a condition of reduced stress⁴¹. These experiments demonstrated that culture conditions negatively affect the metabolism of cortical cells, thus increasing the need for their further optimisation.

2.3.3 Brain organoids and disease modelling

Brain organoid development paved the way for a new level of understanding about neurological diseases, especially neurodevelopmental disorders, with the possibility of studying relationships between cell types composing the human brain in an experimental model with a close-to-physiological cytoarchitecture. Indeed, since the very first study, whole brain organoids have been used for modelling diseases, in particular primary microcephaly caused by mutations in the centrosomal protein CDK5RAP2. Patient-derived organoids showed reduced neuroepithelial regions with reduced amount of radial glia and increased neurons. This phenotype was compatible with a premature neural differentiation and was attributed to a defect in the orientation of the

mitotic spindle of aRG cells, crucial for their symmetric expansion¹²⁷. Similarly, primary microcephaly caused by mutation in the centrosomal protein WDR62, of which ablation causes retarded cilium assembly, long cilium and delayed cell cycle progression, is characterised by decreased proliferation and premature differentiation of NPCs, including oRG cells¹²⁸. Contextually with the development of the telencephalic aggregates protocol, important advancement in the understanding of idiopathic ASD was made. Indeed, organoids from idiopathic cases showed differences in NPCs proliferation, neuronal differentiation and synaptic assembly as well as an overproduction of GABAergic inhibitory neurons, chiefly mediated by dysregulation of FOXP1, corroborating the hypothesis of glutamatergic/GABAergic imbalance in ASD¹⁰⁵. Studies regarding Miller-Dieker Syndrome using forebrain organoids revealed a switch from symmetric to asymmetric cell division in diseased organoids due to an impairment of N-cadherin/ β -catenin signaling axis in the ventricular zone¹²⁹. Defects in oRG cells proliferation were also found in Miller-Dieker Syndrome using brain organoids¹³⁰. Another example of brain organoids use to study neurodevelopmental disorders is Leigh syndrome. Organoids revealed that mutation in the mitochondrial complex IV assembly factor SURF1 causes impaired neuronal differentiation by disrupting the neurogenic capacity of neural progenitors¹³¹. Migration studies in organoids revealed that disruption of the cadherin receptor-ligand pair DCHS1 and FAT4, responsible for periventricular heterotopia, causes defective migration dynamics of subsets of neurons due to changes in the morphology of neuronal progenitors, while the vast majority correctly migrate towards the cortical plate¹³². Functional studies on brain organoids revealed that UBE3A suppresses neuronal hyperexcitability via ubiquitin-mediated degradation of specific potassium channels in the context of Angelman Syndrome¹³³. Organoids have been also used for unravelling the molecular aetiology of genetic ASD. Mutation of different ASD causing genes caused acceleration in the development of cortical neurons, despite affecting different populations, thus pointing at a convergence of phenotype. Indeed, SUV420H1, PTEN and CHD8 mutations caused acceleration of excitatory deep layer neurons, callosal neurons and inhibitory interneurons production, respectively¹³⁴. The acceleration of inhibitory neurons has been demonstrated for CHD8 haploinsufficient brain organoids. This phenotype was accompanied by a delayed production of excitatory neurons along with a protracted neuronal progenitor proliferation phase¹³⁵.

Despite recapitulating early embryonic to early fetal stages of human brain development, organoids have been extensively harnessed for the study of neurodegenerative disorders. Brain organoids derived from patients affected by familiar Alzheimer disease displayed relevant characteristics such as amyloid aggregation, hyperphosphorylated tau protein and endosome abnormality. Moreover, treatment with β - and γ -secretase inhibitors reduced amyloid and tau accumulation, making this model suitable pre-clinical drug testing¹³⁶. Studies on Parkinson disease have been carried out using midbrain-specific organoids. Patient-derived organoids show a decrease in the number and

complexity of dopaminergic neurons¹³⁷. Additionally, Huntington disease has been modelled through brain organoids. Large huntingtin CAG repeats show complete failure of neuroectodermal acquisition, while shorter repeats show abnormalities in neural rosette formation and disrupted cytoarchitecture of cortical organoids. Huntington organoids show a more immature gene expression profile compared to control organoids, possibly connecting abnormal huntingtin to abnormal neural development in Huntington disease¹³⁸. Interestingly, brain organoids are also a suitable model to study glioblastoma cell invasion dynamics in the surrounding healthy tissue^{139,140}. Being brain organoids suitable to recapitulate human-specific features, researchers have used them to highlight differences between humans and other primates. Leveraging primary and brain organoids single cell RNAseq data, researchers were able to find that human-specific gene candidates, single or clustered in specific gene co-expression modules. These genes were enriched in genes affecting transcription during G1/S transition and neuronal apoptosis, processes that may influence the number of progenitors and neurons. Most importantly, it emerged that the PI3K/AKT/mTOR pathway was particularly activated in the oRG cells of humans, compared to other primates, which may promote stemness and long-term self-renewal, compatibly with an increased neuronal output in humans¹²². Another field of applications of brain organoids is the modelling of CNS infection effects. Studies on the impact of Zika virus infection on the DNA methylation pattern of organoids revealed that the infection alters the methylome of neural progenitors, astrocytes and differentiated neurons at genes implicated in brain disorders such as mental retardation and schizophrenia¹⁴¹. Zika virus also caused reduced proliferation and premature differentiation of RG cells as well as aberrant positioning of new-born neurons, partially through interference with innate immunity, thus potentially explaining the microcephaly phenotype that this infection causes^{142,143,144}. Brain organoids have also been infected with the Japanese encephalitis virus, which preferentially targeted astrocytes and oRG cells, revealing differential antiviral response, active in more mature compared to young organoids¹⁴⁵. The recent Sars-CoV-2 virus has also been shown to have effects at the level of the CNS. Indeed, infection of brain organoids showed breakdown of the CSF-brain barrier¹⁴⁶. Lastly, given the significance of brain organoids in the disease-modeling field, researchers set up high-throughput screening platforms integrating different technologies, such as high-content imaging and multi-electrode arrays, to score the functional activity of brain organoids upon exposure to libraries of drugs¹⁴⁷.

2.4 Regulation of gene expression in human brain development

Among the mechanisms controlling gene expression during human brain development, DNA methylation is the most studied. Promoters activity through different developmental stages is influenced by the balance between DNA methylation and chromatin modifications^{148,149}. Genes regulating these developmental transition have been found to be related to low-methylation

genomic clusters referred to as DNA methylations valleys (DMVs), which localise at their promoters and become hypermethylated in committed neuronal cells^{149,150}. Cytosine methylation is usually associated with the presence of CpG dinucleotides, although methylation in non-CpG context was found to be more abundant in human ESCs and brain, with perinatal accumulation that continues until adolescence. Methylation at non-CpG cytosines becomes predominant in human neurons, in correspondence of CAC sequences^{151,152,153}. Importantly, the result of this epigenetic modification is anyway transcriptional repression. Indeed, MeCP2 is able to recognise non-CpG methylated sites and to recruit in their correspondence the NCoR/SMRT co-repressor complex, impacting on neuronal survival and connectivity through the modulation of neurotrophic factors, such as BDNF^{154,155,156}. Given the importance of DNA methylation in the course of neurodevelopment, going from neuronal plasticity and memory formation to the regulation of cognitive functions^{157,158,159}, the relationship with neurodevelopmental disorders is expected. Methylome analysis of post-mortem brains from autistic patients revealed global hypomethylation in young autistic subjects and that differentially methylated *loci* are associated with genes involved in neuron proliferation, differentiation and migration. Moreover, methylation at cytosines was reduced at putative aberrantly alternatively spliced exons in favour of their hydroxymethylation in autistic subjects. The same study also hypothesised that, in ASD cases compared to controls, developmentally-relevant DNA methylated *loci* have a higher correlation with early time points of fetal brain development, suggesting an epigenetic delay in the acquisition of DNA methylation states in ASD¹⁶⁰.

Enhancer regions represent another important layer of gene expression regulation. Nowadays, it is known that changes in the DNA sequence or epigenetic status of enhancers led to the modification of their activity during evolution, thus contributing to the evolution human brain. Comparative analysis of different vertebrate genomes revealed evolutionarily conserved non-coding sequences that undergone human-specific changes at an accelerated rate than expected by chance. These human-specific variations have been attributed to segmental duplication or single nucleotide changes, although in both cases the rate of mutation and the selective pressure point to a sort of “acceleration” in their acquisition^{161,162,163}. Interestingly, these human accelerated regions (HAR) have been associated with developmental enhancers¹⁶⁴. Moreover, a subset of evolutionarily conserved enhancers did not undergo DNA sequence alteration, but rather epigenetic activation or repression due to variation of acetylation of the lysine 27 of histone H3 (H3K27ac), thus defining human gained or lost enhancers (HGE or HRE)^{165,166}. Deep comprehension of the role of evolutionarily-relevant enhancer regions can only be achieved by integration with chromatin interaction data, linking their activity to target genes. Such analysis revealed that HARs are related to genes involved in brain development, patterning and regionalisation, neuronal progenitor proliferation and cortical lamination, suggesting their role in the human-specific regulation of these processes. These genes include crucial developmental regulators mostly related to forebrain

development, such as EMX2, PAX6, HES1, SOX2, TBR1, CUX1. HGEs and HLEs instead regulate genes involved in GTPase regulators, GPCR signaling pathways, mTOR signaling, etc. Common patterns among the genes regulated by HAR, HGE and HLE revealed their importance in primate cortical expansion, particularly in relation to oRG cells and their progeny. Indeed, overlap between genes associated with human evolution elements and single cell RNAseq of the developing prefrontal cortex revealed that the majority of these genes were enriched in oRG cells. Interestingly, HAR appear to regulate preferentially ASD-relevant genes, while HGE genes are more related to developmental delay and intellectual disability¹⁶⁷. Another study performed functional studies on these two classes of enhancer regions, revealing a role of these elements in neuronal progenitors differentiation by CRISPR/Cas9 perturbation¹⁶⁸. Enhancers have also fundamental roles in cortical regionalisation and patterning. Indeed, several enhancers define protodomains of the pallial neuroepithelium crucial for cortical arealization. The control of the activity of those enhancers has been linked to the TFs regulating cortical patterning, such as PAX6¹⁶⁹.

Histone post-translational modifications are responsible for the dynamic regulation of gene expression, thus they are essential in brain development for the regulation of genes, and relative regulatory regions, activity. Our knowledge of human-specific histone post-translational modification regulation during brain development is still limited, although studies in this direction are starting to emerge. H3K4 methylation profiles of the human prefrontal cortex provided important evolutionary differences. In particular, a group of transcription start sites (TSS) enriched in *loci* associated with neuropsychiatric susceptibility genes resulted differentially H3K4 methylated in humans compared to primates⁵⁵. There is also evidence that H3K4 methylation in the human cerebral cortex is dynamically regulated during pre- and post-natal brain development and that it is associated with autism and schizophrenia^{170,171}. Interestingly, H3K4me3, besides being enriched at active promoters, was also found at 5'UTR of a subset of genes in neuronal cells, suggesting a link with METTL3-mediated N6-methyladenosine deposition on RNAs, a process that involves about half of expressed mRNA in neurons and impairment of which causes neuronal malfunctioning¹⁷². H3K4me3 and H3K27ac, together with RNAseq, have also been profiled in brain organoids defining gene regulatory regions at three different developmental time points. Interestingly, the transition from neuronal progenitors to early neurons was the most enriched in differentially expressed genes and differential active enhancers, pointing at this step as the most vulnerable in human brain development. Notably, HGE were found to be mostly active during this transition, linking them to neuronal precursors proliferation. The same study also classified enhancer regions in active or repressing regulators, highlighting a so far underestimated layer of trans-repression during early human brain development¹¹⁸.

Lastly, given the importance of the epigenetic control in all developmental processes, it is not surprising that alteration of chromatin regulators, such as ADNP, KMT2B, EZH2 among many others, leads to neurodevelopmental defects, pointing at their crucial role in brain development. Later in this chapter, a particular focus will be on the role of Polycomb Repressive Complex 2 (PRC2) in cortical development and its involvement in human neurodevelopmental disorders.

2.5 Polycomb group proteins

Polycomb group of proteins (PcG) are encoded by a set of genes, referred to as Polycomb genes, firstly identified in *Drosophila Melanogaster* (*Drosophila*) as regulators of Homeobox (HOX) genes. HOX genes define the specification of *Drosophila* body segments thanks to their spatial and temporal expression pattern. Indeed, they encode subclasses of transcriptional regulators that elicit distinct developmental programmes along the head-to-tail axis of animals¹⁷³. The name PcG derives from the particular phenotype that their mutation causes. Indeed, the sex comb characterising male adults of *Drosophila* is a specialised row of bristles present only on the first pair of thoracic legs in wild-type animals, while mutants of the Pc gene present extra combs on the second and third pair of legs^{174,175}. The molecular phenotype of homozygous Pc mutant larvae is the loss of repression of HOX genes and, subsequently, other mutations causing the appearance of extra-combs were identified in other protein coding genes, thus leading to the collective name of Polycomb group proteins (PcG)¹⁷⁶. Mutation of PcG causes the transformation of *Drosophila* body segments into others, defining the so-called homeotic transformation^{176,177,178}. HOX genes are not exclusively regulated by PcG, indeed Tritorax group proteins (TrxG) antagonise PcG activity at these loci, complementing the spatio-temporal control of their expression¹⁷⁹. Animals carrying mutation in Trithorax have three abnormal thoracic segments and, later, all other positive regulators of HOX genes expression were collectively defined as the TrxG. Mutations in PcG cause anterior embryonic segments to become more posterior, and *vice versa* for TrxG.

PcG and TrxG were relatively conserved during evolution and their function has been linked to differentiation, cell fate determination, proliferation, senescence and cancer. Indeed, they can regulate a plethora of cellular processes, including X chromosome inactivation, genomic imprinting, cell cycle control, stem cell biology, and cancer cell maintenance. This functional diversity is achieved by the assembly of the PcG and TrxG proteins in a cell- and developmental-stage-specific manner to control gene expression via histone-modifying or chromatin-remodelling activities¹⁸⁰.

2.5.1 Polycomb Repressive Complexes: a focus on PRC2

PcG are organised in two families of protein complexes: Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2), with distinct functional regulation of chromatin. PRC1

mediates the ubiquitination of Lysine 119 on histone H2A (H2AK119Ub), whereas PRC2 mediates the methylation of the Lysine 27 on histone 3 (H3K27me1/2/3)¹⁸¹.

All PRC1 complexes share a conserved protein core in animals and plants. In mammals, canonical PRC1 are composed by: i) the catalytic subunit, which is one of the RING proteins, ii) one chromobox protein (CBX), iii) PCGF1 and iv) HPH. The two catalytical subunits RING1 and RING2 can both mutually exclusively assemble in the complex, although RING2 is more prevalent¹⁸². CBX proteins (CBX2, CBX4, CBX6, CBX7, CBX8) are able to bind H3K27me3 deposited by PRC2 mediating the recruitment of PRC1 to PRC2 targets. The presence of this protein characterises the canonical PRC1 (cPRC1)^{183,184}. A number of non-canonical PRC1 complexes exist and they are characterised by their ability to target chromatin independently of H3K27me3.

PRC2 is composed of one of the SET-domain-containing histone methyltransferases Enhancer of Zeste (EZH1 or EZH2), embryonic ectoderm development (EED), suppressor of Zeste (SUZ12) and the CAF1 histone-binding proteins RBBP4 and RBBP7. EZH1 and EZH2 are the two mutually exclusive catalytic subunits, which mediate H3K27 methylation^{185,186}. EED is responsible for the recognition of methylated lysine residues through an aromatic cage present at its C-terminal, allowing the propagation of H3K27me3 repressive marker on target *loci*. Mutations in the aromatic cage of EED in *Drosophila* reduces PRC2 recruitment to the Polycomb Responsive Elements (PREs)¹⁸⁷. SUZ12 is another core subunit of PRC2 necessary for its enzymatic activity. It mediates the interaction between EZH2 and nucleosomes through a C2H2 zinc finger. Knockdown (KD) of SUZ12 results in a genome-wide decrease of H3K27me3¹⁸⁸. RBBP4 and 7 are histone-binding subunits present in several complexes which regulate chromatin metabolism¹⁸⁹. During evolution, accessory subunits of PRC2 emerged. Through a proteomic approach in human cells, it was possible to identify two fundamental alternative assemblies¹⁹⁰: i) PRC2.1, characterised by mutually exclusive binding of one of the three Polycomb-like homologs (PCLs) PHF1, PHF19, or MTF2, with PHF1 stimulating efficient trimethylation activity of EZH2 on the H3K27me2 substrate^{191,192}; ii) PRC2.2, defined by the presence of the zinc-finger proteins AEBP2 and JARID2, which enhance enzymatic activity and regulate the recruitment of PRC2 to chromatin. Finally, the two mammalian-specific proteins, C10ORF12/LCOR and C17ORF96/EPOP, also co-purify with PRC2 complexes^{193,194,180}.

PRCs regulation of gene expression is significantly different in mammals compared to *Drosophila*. Indeed, indeed, in *Drosophila*, PRCs are recruited to responsive elements on the DNA in a cooperative way with specific transcription factors, while in mammals their binding correlates with GC rich regions¹⁹⁵. In this case, the recruitment model is “chromatin sampling”, where PcG weakly interact with GC rich regions and their binding is secured by the interactions with transcription factors, non-coding RNAs and histone marks¹⁸⁰. The PRC2 subunits AEBP2 and EZH2 have affinity for unmodified CpG-containing chromatin, which is instead reduced for methylated cytosines¹⁹⁶.

2.5.1.1 Enhancer of Zeste 2

EZH1 and EZH2 mutually exclusively bind PRC2 and have slightly different catalytic potential. While EZH2 KD affects global H3K27me3 levels, EZH1 performs this function weakly, not affecting global H3K27me3¹⁸⁵. EZH2 is a histone-lysine N-methyltransferase encoded by the homonym gene on chromosome 7. It contains a C-terminal SET domain necessary for the methyltransferase activity. The SET domain is characteristic of lysine methyltransferases and it was originally identified in different *Drosophila*'s proteins, from which the acronym derives (Su(var)3-9, Enhancer-of-zeste and Trithorax). Crystallography studies characterised the structure of EZH2, demonstrating that its catalytic domain conserves the overall structure of canonical SET domain. Nonetheless, it presents some atypical features, as for example the presence of a CXC domain absent in other SET proteins. Moreover, the post-SET domain appears to diverge from the canonical structure because it folds in a way that it cannot contribute to the formation of the cofactor binding site. This could be explained by the fact that the CXC domain occupies the SAM binding site and the region in which the post-SET domain normally resides. Moreover, the EZH2 I-SET domain forms a secondary pocket juxtaposed to the SAM binding site. These characteristics may be responsible for the inactivity of EZH2 in absence of EED and SUZ12 within PRC2, although more studies are needed to exclude potential artifacts¹⁹⁷. EZH2 is composed by other structural domains that allow its function in the context of PRC2: i) the WD-40 binding domain (WBD) is necessary to bind EED and RBBP4/7; ii) two SANT domains, necessary to interact with histones; iii) the domain 1, necessary to interact with PHF1; iv) the domain 2, which mediates the interaction with SUZ12¹⁹⁸. A graphical representation of PRC2 composition in *Drosophila* and in mammals and of the protein structure of PRC2 subunits is outlined in **Fig. 5**.

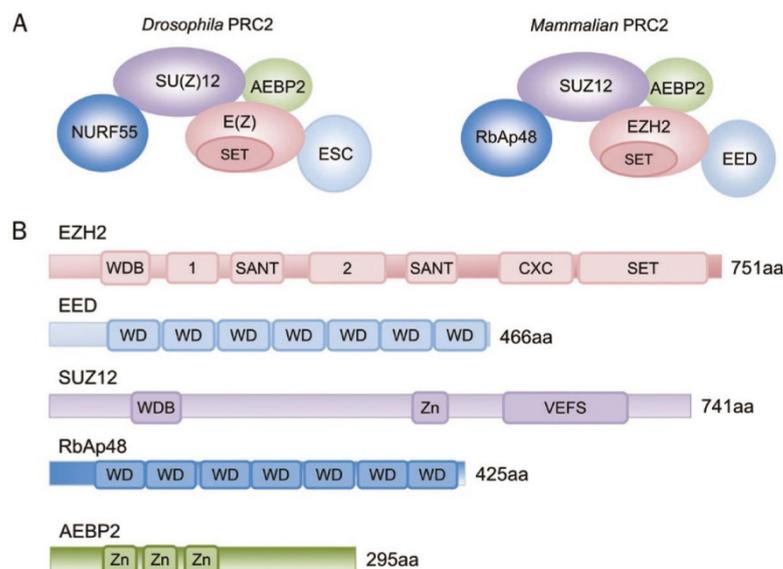


Fig. 5: Architecture of the PRC2 complex. (A) *Drosophila* and mammalian PRC2 composition. (B) Domain structure PRC2 core subunits. Adapted from¹⁹⁸.

In addition to its histone methyltransferase activity, EZH2 has been found to interact with DNA methyltransferases (DNMTs) and is associated to their activity *in vivo*. The direct connection between these two modalities of gene repression is demonstrated by the fact that DNMTs binding to several PRC2 repressed genes depends on EZH2¹⁹⁹. EZH2 expression is associated with proliferative tissues¹⁸⁵ and its levels are particularly high in ESCs and in adult stem cells, where it contributes to maintain their multipotent identity²⁰⁰. Moreover, EZH2 has been recently found to have PRC2 independent and non-catalytic activity, promoting nuclear excision repair by complexing with DDB2 and enhancing its stability²⁰¹. Another example of methyltransferase-independent activity is its function as activator of the androgen receptor gene transcription in prostate cancer, demonstrated by the direct occupancy of its promoter in presence of H3K27ac rather than H3K27me3. In this context, EZH2 activity was demonstrated to be also PRC2-independent, since KD of SUZ12 did not reproduce the downregulation of the androgen receptor transcription observed with EZH2 KD²⁰².

In conclusion, EZH2 primarily acts as part of PRC2, although several reports started to elucidate its Polycomb-independent functions.

2.5.2 PRC2 function in pluripotency and cortical development

Ezh2 has been extensively studied during the reprogramming of mouse embryonic fibroblasts into iPSC, where, despite its inactivation, H3K27me3 is deposited on a selected group of PRC2 targets enriched in developmental regulators controlling the expression of lineage specific genes²⁰³. Indeed, several groups demonstrated that while PRC2 is dispensable for ESCs maintenance, it is instead essential to preserve their plasticity during embryonic development^{204,205,206,203,186}. Importantly, many promoters in ESCs carry a bivalent chromatinic signature consisting of large H3K27me3 regions harbouring smaller H3K4me3 peaks. Bivalent promoters tend to disappear upon ESCs differentiation, acquiring either one or the other marker, and are usually associated with developmental TFs genes expressed at low levels. Given its properties, this bivalent signature defines the initial poised epigenetic state of a particular locus at the pluripotent stage, which resolves upon differentiation in transcriptional activation or repression (H3K4me3-only or H3K27me3-only, respectively) according to lineage-specific gene expression programmes²⁰⁷. A similar principle applies to poised enhancers, initially defined in ESCs and marked with a bivalent signature consisting of H3K27me3 and H3K4me1, while being devoid of H3K27ac²⁰⁸.

Deletion of Ezh2 in ESCs results in loss of neurogenic capacity, pointing at a crucial role during neurodevelopment²⁰⁹. Indeed, it is known that different sets of genes are differentially regulated by Polycomb during neurodevelopment, resulting in the fine-tuned control of the timing of neural maturation²¹⁰. In particular, only a subset of PRC2 targets is specified in ESCs, while novel lineage-specific genes become transiently H3K27me3 in NPCs upon cell fate commitment. The repression

of non-neuronal factors mediated by PRC2 in adult neurons is actively needed at the level of bivalent genes and can change due to cell intrinsic or extrinsic stressors^{211,212,213}. Ezh2 and Ezh1 expression varies during cortical neurons differentiation, with decreasing levels of Ezh2 during neuronal specification and increasing levels of Ezh1 accordingly, regulating not only the timing of neuronal maturation but also the neurogenic to astrogenic transition²¹³ (**Fig. 6**).

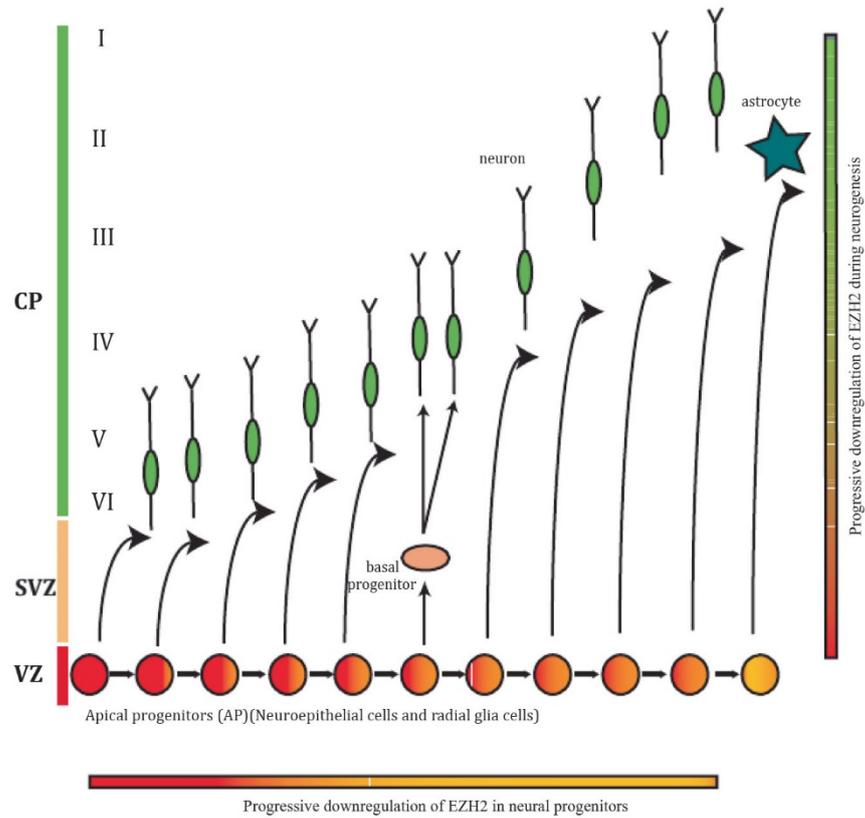


Fig. 6: Ezh2 expression along corticogenesis. Adapted from²¹⁴.

The evidence that Ezh2 acts as a molecular switch during cortical development is corroborated by different studies that addressed part of the Ezh2-mediated regulation by abolishing its function in different phases of corticogenesis. Ezh2 KD in neural tube from chicken embryos induced defects in the apico-basal polarity of neuroblasts leading to impaired neural tube organisation caused by expression deregulation of the cell cycle controller p21WAF1/CIP1²¹⁵. Instead, loss of function of Ezh2 in mouse NPCs before the onset of neurogenesis changed the balance between differentiation and self-renewal towards differentiation²¹⁰. Indeed, despite a broadly conserved temporal order of maturation, the neurogenic period was shorter, with a reduced neuronal output and unbalanced representation of deep and upper layer neurons towards the former²¹⁰. Conversely, ablation of Ezh2 at the neurogenic to astrogenic switch period led to the late onset of astrogenesis due to a prolonged neurogenic phase caused by increased levels of Ngn1, a direct PRC2 target²¹⁶. PRC2 also acts in concert with DNA methylation for regulating gene expression during cortical development.

In particular H3K27me3 promoters are more prone to acquire DNA methylation during neuronal differentiation, defining restriction and potential of NPCs²¹⁷.

Although the role of Ezh2 during mouse corticogenesis has started to be elucidated, our knowledge is limited to few targets and to a general understanding of the PRC2 function in the regulation of the timing of neuronal differentiation, though without a deep comprehension of the dysregulation propagated in neurons during their fate specification. Interestingly, mutation of EZH2, and in other subunits of PRC2, have been identified as the molecular cause of Weaver syndrome (WS), increasing the need of clarifying the PRC2 role in human corticogenesis.

2.6 Overgrowth and intellectual disability: the example of Weaver Syndrome

Overgrowth is a clinical condition that may depend on genetic, endocrine or metabolic factors. When metabolic imbalance takes place, overgrowth is usually the only clinical condition, while dysmorphic features and cognitive impairment are often present if the cause of overgrowth is genetic²¹⁸.

Syndromes characterised by genetic alterations causing generalised overgrowth usually account for very few cases worldwide, often leading to the grouping of patients with atypical manifestations into the same clinical domain. For example, Sotos syndrome is mainly caused by loss of function of NSD1, responsible for H3K36 methylation, although mutations in other epigenetic regulators (SETD2, DNMT3A or APC2) lead to Sotos-like phenotypes similar enough to fall in the same clinical classification^{219,220,221}. This is not the case for mutations in NFIX, which lead to Sotos-like manifestations, although being classified as the cause of a completely new syndrome, Malan Syndrome²²².

WS is another typical example of a rare overgrowth disease accounting for very few cases worldwide (less than 100)²²³, with heterogeneous clinical phenotypes anyway classified as WS. Although it was first described in 1974²²⁴, the molecular roots of WS remained unknown until 2011, when EZH2 heterozygous mutations were designated as the molecular basis of the disease^{223,225}.

2.6.1 Clinical features associated with Weaver Syndrome

WS is characterized by pre and post-natal overgrowth, macrocephaly, peculiar facial dysmorphisms and intellectual disability. The growth profile of Weaver patients, in the cohort analysed by Tatton-Brown²²³, 38% of Weaver patients had a birth weight greater than 2 standard deviations (SD) above the mean. 91% of patients showed a height at least 2 SD above the mean. Compared to other diseases characterised by macrocephaly, occipitofrontal circumference measurements in 40 Weaver patients ranged from -0,9 to +5,5 SDs, with a median value of +1,8 SDs. Facial features that characterize WS individuals include: a high hairline, ocular hypertelorism, broad forehead, almond shaped palpebral fissures, a pointed “stuck-on” chin with horizontal skin crease and, in early

childhood, large fleshy ears and retrognathia²²⁶. Within WS, intellectual disability affects 82% of the patients with mild to severe impact²²⁷, making this aspect of the disease the major cause of distress. A severe level of intellectual disability is quite rare. The remaining 18% of patients show cognitive abilities within the normal range. Brain imaging studies were also performed on nine WS patients and results showed: i) four patients with isolated ventriculomegaly; ii) one individual with ventriculomegaly and periventricular leukomalacia; and iii) four individuals each with either periventricular leukomalacia, pachygyria and polymicrogyria, a small infarcted area in the cerebellum, or a persistent cavum septum pellucidum²²³. Additional features of WS are advanced osseous maturation; poor coordination; soft, doughy skin; camptodactyly, hoarse, low pitched cry and umbilical hernia^{227,228}.

Given EZH2 role in cancer maintenance and progression²²⁹, it would not be surprising if WS had an increased probability to develop neoplasms, although the small number of patients makes it difficult to precisely determine the risk. Until now, only five WS patients have been diagnosed cancer, three of them with neuroblastoma, one with malignant ovarian endodermal sinus tumor and one with sacrococcygeal teratoma²³⁰. Although relevant, all these cases have been identified before linking PRC2 to Weaver syndrome, thus leaving the possibility that the patients affected by cancer may not be true Weavers.

2.6.1.1 Intellectual disability

Intellectual disability (ID) is defined as a clinical conditions characterised by deficits in intellectual functions and in adaptive behaviours affecting 1-3% of the world-wide population, thus representing the most common developmental disorder^{231,232}. Other common clinical features of subjects affected by ID are delayed speech, seizures and hypotonia²³³. The causes of intellectual disability are multiple and can be grouped in genetic and environmental. Part of the environmental causes are alcoholic syndrome, infections (such as measles or meningitis) and severe head injuries. Genetic causes are instead more common and can be chromosomal aberrations, single gene alterations or epigenetic defects leading to abnormal gene expression²³⁴. Based on their inheritance pattern, ID caused by single gene mutation can be divided in autosomal recessive, autosomal dominant or X-linked. Interestingly, genes involved in non-syndromic autosomal recessive ID, where ID is the only clinical phenotype, are being studied to understand normal variation of intelligence. Indeed, differences in the intelligence quotient are linked with genes that, if mutated, cause variation in intellectual ability²³⁵. Weaver syndrome is instead an autosomal dominant syndromic condition characterised by ID.

2.6.1.2 Neural crest stem cells

The neural crest is a transient multipotent cell population that originates during vertebrates gastrulation, giving rise to many cell types, including adipocytes, osteocytes and chondrocytes

composing craniofacial structures. Neural crest induction starts at the neural tube border and continues until neural tube closure under spatio-temporal patterning mediated by Wnt, Fgf, BMPs. These molecules induce expression of master regulators of neural plate border, which later differentiate in neural crest stem cells (NCSC). NCSC represent the multipotent intermediate population between neural plate border cells and terminally differentiated cell types. Indeed, neural crest specifiers promote epithelial to mesenchymal transition and motility by regulating the expression of specific adhesion molecules^{236,237,238,239,240,241}). Different subtypes of NCSC exist based on their axial origin and the cell types they generate. In particular cranial NCSC, originating from the regions surrounding the rudiments of encephalic vesicles, give rise to the main facial structures. These include: i) the frontonasal process; ii) the mesenchyme of the first pharyngeal arch that will eventually give rise to jawbones and ear bones malleus and incus; ii) the mesenchyme of the second pharyngeal arch from which facial cartilage and the stapes bone of the ear are derived; iii) the mesenchyme of the third, fourth and sixth pharyngeal arches that produce muscles and bones of the neck²⁴². The fifth pharyngeal arch degenerates in humans²⁴². Given the frequent occurrence of facial dysmorphisms in neurodevelopmental disorders, cranial NCSC became a valuable tool to study their underpinning molecular features. Interestingly, there is evidence that mild neural crest deficit underlie domestication in humans²⁴³. Indeed, overall facial reduction and retraction, more gracile profile, reduced prognathism, reduced brow ridges and nasal projections are characteristic of anatomically modern humans²⁴⁴. These changes correlate with prosociality, thus leading to the so called self-domestication hypothesis^{245,246} and to a link between neurodevelopment and facial dysmorphisms. Given these premises, cranial NCSC, and their progeny, represent a disease-relevant model for WS.

2.6.2 Molecular basis of Weaver Syndrome

The main mutational mechanism of WS are missense mutations in EZH2, with only 10% of cases presenting truncating mutations within the terminal exon. However, the link between genetic alteration and clinical manifestations remain largely unknown. As previously mentioned, EZH2 is the main catalytic subunit of PRC2 and it mediates H3K27 methylation. Additionally, EZH2 directly controls DNA methylation by recruiting DNA methyltransferases via its amino-terminal region to target genes leading to transcriptional repression¹⁹⁹.

Mutations in other PRC2 subunits have been identified in WS patients. Mutations in EED, necessary to stabilise the complex and required for the methyltransferase activity of EZH2, also cause an overgrowth phenotype falling into the WS clinical picture^{247,248}. More recently, mutations in SUZ12 have been identified in patients showing Weaver-like phenotypes, including overgrowth²⁴⁹.

Functional studies demonstrated that most of WS-associated EZH2 variants present in the SET domain reduce the histone methyltransferase function of EZH2, indicating a possible loss of

function mechanism. Partial loss of function was also proved in mice engineered to carry the same mutation found in one WS patient. Indeed, homozygous and heterozygous mice embryos showed a decreased H3K27 methylation²⁴⁷. Despite this, evidence of gain of function mutations of EZH2 were identified in a patient with a Weaver-like phenotype affected by growth restriction rather than overgrowth. Interestingly, researchers found that this variant is responsible for a unique DNA methylation signature different from the one found in patients with EZH2 loss of function²⁵⁰.

The finding that mutations in different subunits of PRC2, either gain or loss of function, cause a Weaver-like condition with convergent and divergent phenotypes leads to the hypothesis that this syndrome may be classified as a spectrum of clinical conditions, which may broadly be defined as Polycombopathies.

2.7 Genome editing approaches for disease-modelling

In the last decade, genome-editing techniques have been coupled to iPSC technology to demonstrate causal relationship between genetic mutations and disease phenotypes^{75,251}. Indeed, genetic manipulation of disease-causing mutations has allowed the understanding of the contribution of particular genomic alterations to disease phenotypes regardless of the specific genetic background. Different techniques, such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), have been developed in order to enhance homologous-directed recombination (HDR) in cells, but, so far, the CRISPR/Cas9 system is the most efficient and the most widely used²⁵².

2.7.1 Clustered Regularly Interspaced Short Palindromic Repeats

Prokaryotic genomes contain a family of repeated DNA sequences that consists of non-contiguous 24 to 40 nucleotides recurrent motifs regularly spaced by unique sequences of similar sizes, referred to as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). These repeats were first identified in 1987²⁵³ and later were found as the most widely distributed family of repeats among prokaryotic genomes^{251,254}. Within CRISPR arrays, spacers derive from foreign sequences found in bacteriophages and conjugative plasmids, suggesting the implication of these arrays in the genetic memory against infective agents, conferring a sort of “prokaryotic immunity”^{255,256}. To test this hypothesis, alteration of *S. Thermophilus* CRISPR *loci* by addition or deletion of spacers were performed. Spacers deletion led to sensitivity to the phage from which the spacer derived, while new spacers addition, previously absent in the bacterial chromosome, conferred resistance to the corresponding phage^{255,256}.

In close relation with the CRISPR array, there are genes named CAS (CRISPR associated genes), suggesting a functional cooperation. Spacers can have either of the two orientations within a CRISPR locus, and anti-sense RNA are transcribed from them^{253,257}. Northern blot analysis using a

probe containing single-stranded spacer sequences, showed transcription of the CRISPR region in the same direction of the CAS genes. This pointed to the conclusion that the system uses small RNAs to target foreign DNAs^{255,256}. The RNA transcribed from CRISPR *loci* is a precursor, 120-180 nucleotides long, of the so called CRISPR-RNA (crRNA)²⁵⁸.

There are three types of CRISPR/Cas systems. Type I and III systems share some features such as specialised Cas endonucleases that process the pre-crRNAs and the fact that mature crRNA assembles into a large multi-Cas protein complex, the Cascade complex, capable of recognising and cleaving nucleic acids complementary to the crRNA. Conversely, in the type II system, pre-crRNAs processing is mediated by a trans-activating crRNA (tracrRNA) that triggers maturation by the double-stranded RNA-specific ribonuclease (RNase III) in the presence of the Cas9 protein, which is the only CAS protein responsible for exogenous DNA interference²⁵⁷. The pre-tracrRNA is a transcript of 210 nucleotides encoded upstream and in the opposite strand of the leader-repeat-spacer array. This transcript contains a stretch of about 25 nucleotides complementary to the pre-crRNA, suggesting a co-maturation process for crRNA and tracrRNA^{258,259}.

The tracrRNA is required for target DNA recognition and may play a role in the orientation of the crRNA for interaction with the complementary strand of target DNA. Their interaction creates a structure in which the 5'-terminal 20 nucleotides of the crRNA, which define target specificity, are available for target DNA binding²⁵⁷. For target recognition, the Cas9 protein requires a sequence called protospacer adjacent motif (PAM) located within the target DNA downstream the sequence recognised by the protospacer. The PAM sequence varies depending of the CRISPR system. For instance, for *S. Thermophilus* CRISPR is an NGGNG sequence, while for *S. Pyogenes* system requires a NGG. Point mutations introduced at the 5' end of the protospacer do not interfere with the cleavage activity of the Cas9/tracrRNA:crRNA complex, but mutations close to the PAM sequence are not tolerated, defining a seed region at the 3' end of the protospacer sequence^{257,260}.

Since dsDNA lacking a PAM is not bound, this sequence may be essential for strand separation and stabilisation of the R-loop structure that allows the cleavage by the Cas9. In this loop structure, the negative strand of the DNA is displaced and the positive strand is bound to the crRNA unique sequence²⁶¹. This configuration allows dsDNA cleavage, which occur three nucleotides upstream the PAM sequence in presence of Mg²⁺ ions and the consequent generation of a blunted end dsDNA²⁵⁷.

A schematic representation of type II CRISPR/Cas system is depicted in **Fig. 7**.

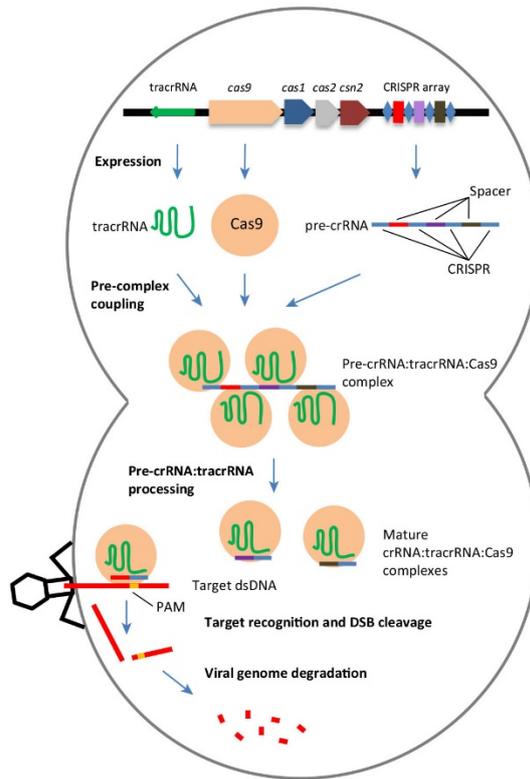


Fig. 7: Type II CRISPR-Cas system in *S. pyogenes*. Adapted from²⁶².

2.7.2 Genome engineering applied to disease modelling

In 2013, F Ann Ran and colleagues for the first time were able to use the CRISPR/Cas9 system, derived from *S. Pyogenes*, to perform genome editing of cultured mammalian cells. By designing a single-guide RNA (sgRNA) as a fusion of the scaffold portion of the tracrRNA (80bp) and the target-specific portion of the crRNA, they were able to target a specific *locus* of the genome²⁶³. The delivery of the sgRNA, Cas9 and repairing template containing homology arms against the sequence of interest triggers homologous recombination at the specific target-site. Conversely, non-homologous end joining (NHEJ) is activated if the repairing template is not provided (**Fig. 8**). Recombinant clones can be screened by different methods, such as *locus*-specific PCRs followed by Sanger sequencing.

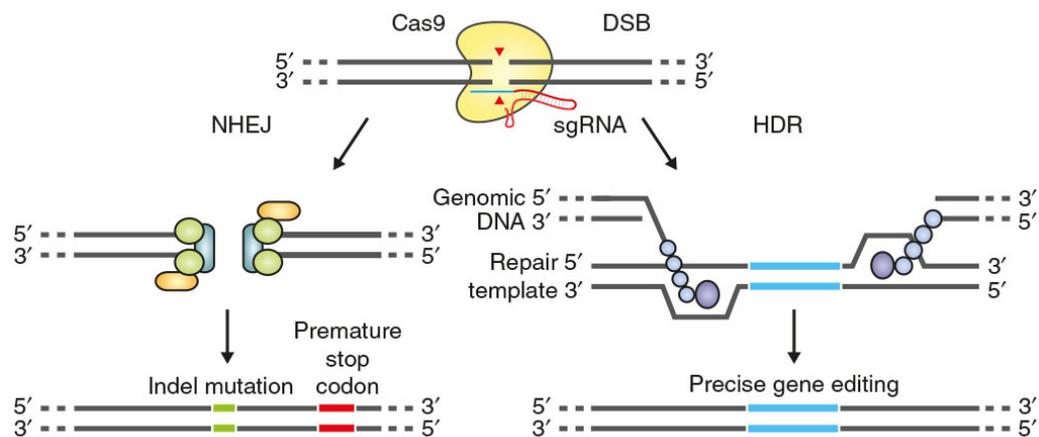


Fig. 8: Genome engineering using CRISPR/Cas9. Adapted from²⁶³.

The combination of CRISPR/cas9 system and iPSC technology opened the possibility to perform experiments that demonstrate causal relationships between genotypes and phenotypes in human genetic disorders⁷⁵. In particular, the possibility of applying CRISPR/Cas9 genome editing to engineer isogenic cell lines carrying only specific mutations of interests in the same genetic background of the parental line is particularly relevant. The advantage of this method is the study of pathogenic mutations regardless of confounding factors due to the genetic background²⁶⁴.

Isogenic lines can be either patient-derived cell lines in which the mutation of interest is corrected or control lines engineered to carry a specific mutation, although the information obtained with the two set-ups is slightly different. Corrected patient-lines allow to study the genetic rescue effect in a diseased genetic background, while if the mutation is engineered in a control line, the information obtained is exclusively related to its molecular effects. To generate true isogenic cell lines, the repair template for homologous recombination should carry only the mutation of interest flanked by HAs. However, post-genome editing screening for the detection of correctly edited clones is complicated by the absence of a selection method. For this reason, methods relying on the engineering of a removable antibiotic resistance cassettes could be used.

Other applications of genome editing include the generation of full knock out (KO) lines for specific genes, as well as the generation of fusion proteins useful to study the cellular and molecular function of particular genes of interest²⁶⁵. Moreover, it is possible to use catalytically inactive versions of the Cas9 fused to activatory or inhibitory domains to perform CRISPR interference for the epigenetic control of transcription. Indeed, CRISPR/Cas9 has been coupled to single-cell RNAseq (scRNAseq) to perform large-scale screenings where hub disease-relevant genes have been simultaneously perturbed to deconvolve directional interdependencies²⁶⁶.

2.7.3 Genome engineering applied to neurodevelopment

Isogenic cell lines, coupled to iPSC technology and their differentiation into neuronal cell types, have been used to deconvolve the gene-phenotype relationships in mono- (or oligo-) genic neurodevelopmental disorders. This kind of application has been used by Ross and colleagues²⁶⁷ to study the autism spectrum disorder (ASD) caused by genetic alterations in the Xp22.11 locus, which encompasses three genes: PTCHD1, the long noncoding RNA PTCHD1-AS, and DDX53. PTCHD1-AS is frequently disrupted in male subjects with ASD. Isogenic cell lines carrying two polyadenylation sequences disrupting the exon 3 this transcript, resembling the genetic configuration found in patients, have been generated and differentiated in neurons. Using this approach, researchers found a decrease in the frequency of AMPAR-mEPSCs in engineered neurons, confirming the results observed in patients. Interestingly, the amplitude of AMPAR-mEPSCs was decreased in CRISPR-edited neurons, but not in neurons derived by patient iPSC carrying the same type of genetic

alteration, probably revealing a phenotype hidden in the non-isogenic setting. Kleefstra syndrome is another neurodevelopmental disorder for which an isogenic setting was crucial to unravel disease-relevant phenotypes. Kleefstra syndrome is caused by mutations in the histone methyltransferase EHMT1 and is associated with intellectual disability and autism. An isogenic line carrying a disease-causing mutation in a control genetic background established the causal role for EHMT1 in the determination of the neuronal alterations at the synaptic network level in Kleefstra syndrome²⁶⁸. More recently, CRISPR/Cas9 coupled to scRNAseq has been used to find convergence of phenotypes in neurodevelopmental disorders by interfering with multiple autism-causing genes in multiplex screenings. With this approach, researchers were able to classify genes in two modules of mechanistic convergence, namely genes that delay neuronal differentiation and genes that accelerate it¹²³. CRISPR/Cas9 scarring coupled with scRNAseq is another powerful tool that allows to perform lineage tracing experiments, particularly useful to understand the dynamic of neuronal progenitor proliferation and lineage development in the brain. Indeed, this approach was used to study zebrafish brain lineages, with the main result being that most individual embryonic progenitors generate multiple cell types, although these tend to remain in relatively restricted spatial domains^{269,270}. A similar approach was applied to human brain organoids, leading to the conclusion that iPSC clones used to initiate organoids tend to accumulate in distinct brain regions. Spatial transcriptomics coupled with 4-D light sheet microscopy confirmed local proliferation of clones with limited migration during neuroepithelial formation, linking clone enrichment in brain organoid regions to positions in the neuroectoderm.

3. AIMS OF THE STUDY

This work bridges benchmarking of cortical brain organoids against the human fetal brain with the application of this experimental system for disease modelling of a paradigmatic neurodevelopmental disorder caused by mutations in the epigenetic regulator PRC2, Weaver Syndrome, with overarching aims that unfold as follows:

- The long-standing challenge of modelling human cortical development derives from the complexity of genetically encoded as well as extrinsic events that need to take place with a precise spatio-temporal sequence during development. In the past two decades, with the advent of hiPSC and later of organoids, we have started to recapitulate some of these events *in vitro*. However, concerns such as the quantification of the reliability and reproducibility of organoid differentiation arose in the field and, as of today, they are only partially solved. We aimed at expanding the current knowledge on this matter by defining the transcriptional dynamics of cortical brain organoids with the objective of providing a resource of gene expression patterns specific of pre-natal cortical development and a quantitation of the extent to which organoids reliably recapitulate them.
- The knowledge about PRC2 function during cortical development derives almost exclusively from studies based on mouse models, where its role has been dissected by knock out (KO) of key subunits, starting from pluripotency throughout the astrogenic phase. Less is known about PRC2 activity during human neurodevelopment, where its alteration causes Weaver syndrome. This disorder remains a clinical unmet need due to the experimental challenge of studying disease-relevant cell types in a human and patient-specific setting. We aimed at dissecting transcriptional and epigenetic alterations due to PRC2 malfunction in human corticogenesis by means of i) a patient cohort including a spectrum of Weaver-syndrome-causing mutations and of ii) CRISPR/Cas9 engineered isogenic lines, in disease-relevant *in vitro* models.

4. RESULTS

4.1 Analysis of the BrainSpan Atlas revealed transcriptional programmes of the developing human fetal cortex

To identify relevant transcriptional circuitries of human cortical development, we took advantage of the most comprehensive and time-resolved human fetal brain transcriptional dataset, the BrainSpan Atlas ⁽²⁷¹⁾ and white paper: [https://help.brain-map.org/display/devhumanbrain/Documentation Developmental Transcriptome](https://help.brain-map.org/display/devhumanbrain/Documentation+Developmental+Transcriptome)). We focused our attention on the pre-natal cortical development by selecting all available samples covering from the end of the human embryonic period (PCW8) to late-gestation (PCW37), for a total of 162 samples distributed as outlined in **Fig. 9**.

Age	n° samples	Area
PCW 8	9	Cortex
PCW 9	7	Cortex
PCW 12	31	Cortex
PCW 13	32	Cortex
PCW 16	28	Cortex
PCW 17	9	Cortex
PCW 19	8	Cortex
PCW 21	11	Cortex
PCW 24	11	Cortex
PCW 25	1	Cortex
PCW 26	3	Cortex
PCW 35	1	Cortex
PCW 37	11	Cortex



Fig. 9: Cohort of fetal cortical samples from the BrainSpan Atlas dataset ([https://help.brain-map.org/display/devhumanbrain/Documentation Developmental Transcriptome](https://help.brain-map.org/display/devhumanbrain/Documentation+Developmental+Transcriptome)). The number of specimens from every PCW is reported for a total of 162 bulk-RNAseq samples

Correlation analysis between the different PCW revealed two main clusters, one formed by PCW8-9 and one by PCW12-24. Interestingly, this subdivision resembled the timing of human neurogenesis, with early neurogenic stages (PCW8-9) clustering apart from time-points in which the drastic expansion of the OSVZ takes place together with the biggest wave of neuronal output (PCW9-onwards)³¹ (**Fig. 10A**). Samples from PCW25 onwards showed a generally lower correlation with other stages, possibly due to the very low number of samples for PCW25, 26 and 35. Remarkably, correlation of post-natal cortical samples showed very high concordance between all stages, indicating that most of the transcriptional changes in the cortex take place during pre-natal development (**Fig. 10B**).

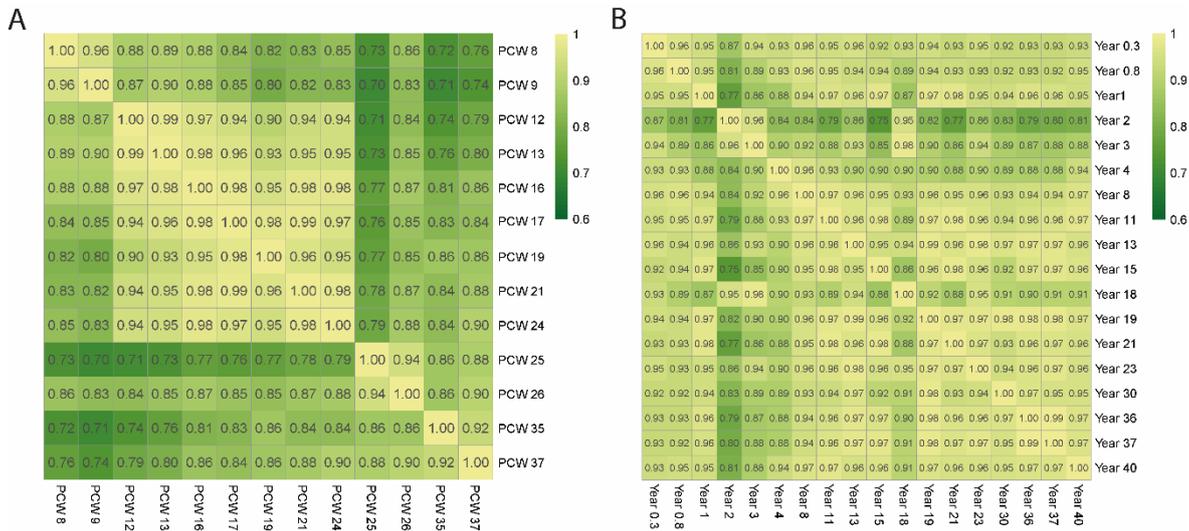


Fig. 10: Correlation analysis of BrainSpan pre-natal and post-natal samples. (A) Heatmap showing Spearman correlation coefficient for every developmental time-point. The mean expression for all samples at a given PCW was considered for correlation analysis. **(B)** Heatmap showing Spearman correlation coefficient for every post-natal time-point. The mean expression for all samples at a given age was considered for correlation analysis.

After this first exploration, we performed dimensionality reduction through principal component analysis (PCA) and identified developmental stage as the main factor driving samples differences (**Fig. 11**), with a peculiar U-shaped distribution when considering Principal Component (PC) 1-2. Indeed, PCW8-9 were separated from later time-points mainly by PC2, while a progression by stage in PC1 was observed from PCW12-13 to PCW37, which in turn were localised between very early and later stages for PC2 values.

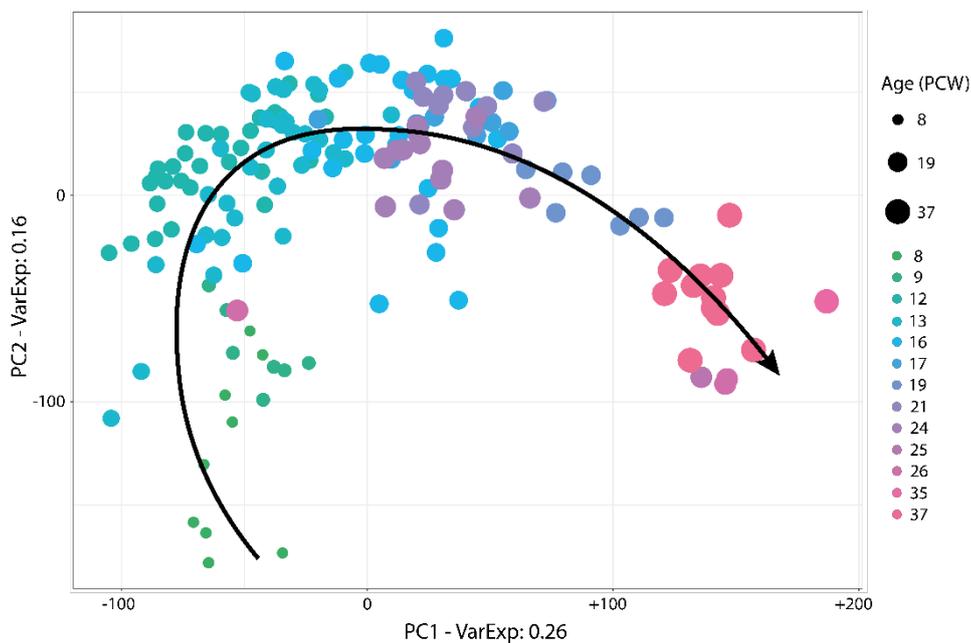


Fig. 11: Dimensionality reduction of pre-natal cortical samples from BrainSpan. Principal component analysis on pre-natal cortical samples from BrainSpan was performed. PC1 and 2 were visualised. Dot colour and size indicate proceeding developmental time-points. The arrow highlighted the U-shape distribution of samples throughout stage-progression.

We thus investigated the genes mainly driving this distribution (**Fig. 12A and B**). Gene ontology (GO) was performed selecting the top 300 genes having the highest positive and negative loading value

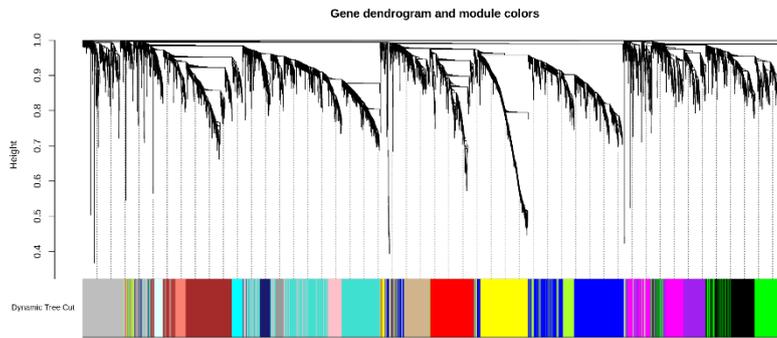


Fig. 13: Gene dendrogram for BrainSpan WGNA module selection. Gene dendrogram generated using as input for average-linkage hierarchical clustering the dissimilarity matrix relative to the topological overlap measure calculated from the adjacency matrix. Network modules were detected as branches of the dendrogram by using the DynamicTree Cut algorithm (deepSplit=1; minimum cluster size= 50; PAM stage TRUE; cutHeight 0.998). For details on the parameters used see²⁷². Colours represent WGCNA gene modules.

Using this strategy, we found 17 gene modules, ranging in size from 105 to 782 genes, and summarised the expression profile of each of them by its first principal component (module eigengene, ME) to relate it with stage either as a continuous variable or series of categorical variables, thus identifying modules positively or negatively associated with development as well as modules highly correlating with particular time-points (**Fig. 14**).

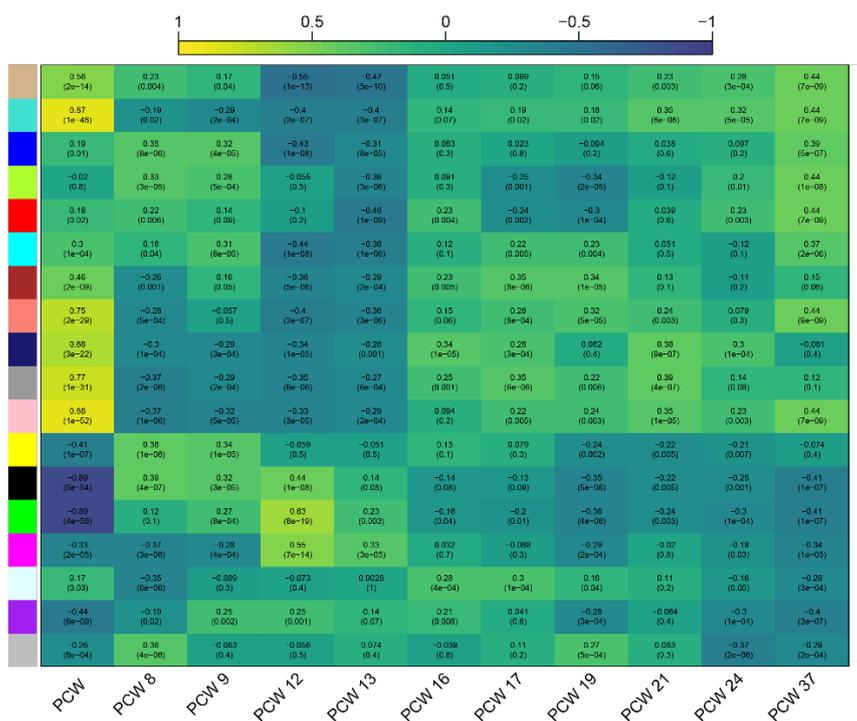


Fig. 14: Correlation of BrainSpan WGCNA modules with stage. Heatmap showing the correlation between ME (represented by the colours on the left) and developmental stage considered as a continuous variable (PCW) or as a series of categorical variables (PCW8-37). Coefficients of correlation were calculated using Spearman correlation and define the heatmap colours. PValue for every correlation, represented in brackets, was calculated and considered significant when < 0.01.

For interesting modules, we visualised the trend over developmental time of their ME, performed functional analysis of the genes composing the module, and reconstructed weighted networks of the top 75 genes according to their intramodular connectivity and with a threshold of 0.2 as

minimum edge weight. Intramolecular connectivity represents a measure of the connectivity of each gene within a single module to all other genes within the same module, while the edge weight refers to the strength between two nodes of the module. We then used CytoNCA²⁷³ to calculate the following centrality scores to identify the most important nodes: i) degree, calculated as a sum of the weight of all connections of a given node, ii) closeness, an indicator of the distance of a given node to all other nodes, iii) betweenness, the quantification of the number of times a node is a bridge along the shortest path between other two nodes, iv) eigenvector, a measure of the influence of a node in the network. Degree, closeness, betweenness and eigenvector were represented by node label transparency, node colour darkness, node border width, and node size/node label font size, respectively.

Among the modules with a strong positive correlation with stage, genes of the turquoise and pink ones resulted to be the most consistently upregulated along cortical development (**Fig. 15A and 16A**). The turquoise module was enriched in genes contributing to the glutamatergic transmission, synaptic development and neuron morphogenesis (**Fig. 15B**). Genes with high centrality indices in this network were indeed mainly related to pivotal neuronal functions. CAMK2N2 encodes for inhibitor of calcium/calmodulin-dependent protein kinase and its rat orthologue has been linked to neuronal synaptic plasticity by phosphorylation of the alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-type (AMPA) glutamate receptors²⁷⁴. Interestingly, this gene has been related to amyotrophic lateral sclerosis (ALS), as C9orf72, another turquoise gene with high betweenness²⁷⁵. Other central turquoise genes such as PITPNM3, CPNE7 PIP5KL1 have known roles in the regulation of neuronal function through calcium signaling^{276,277,278}. Remarkably, genes such as TRF2, GDA, C9orf72, and GRIN3A showed the highest betweenness in the network, despite not having high degree, closeness and eigenvector. TRF2 is implicated in the regulation of neural stem cell proliferation and neuronal differentiation²⁷⁹, GDA is involved in the regulation of synaptic function through its interaction with PSD-95²⁸⁰, GRIN3A is a subunit of the N-methyl-D-aspartate receptor (NMDA) glutamate receptors with non-conventional signalling properties related to the delay of synapse maturation until the arrival of sensory experience and to synapse pruning²⁸¹ (**Fig15C**).

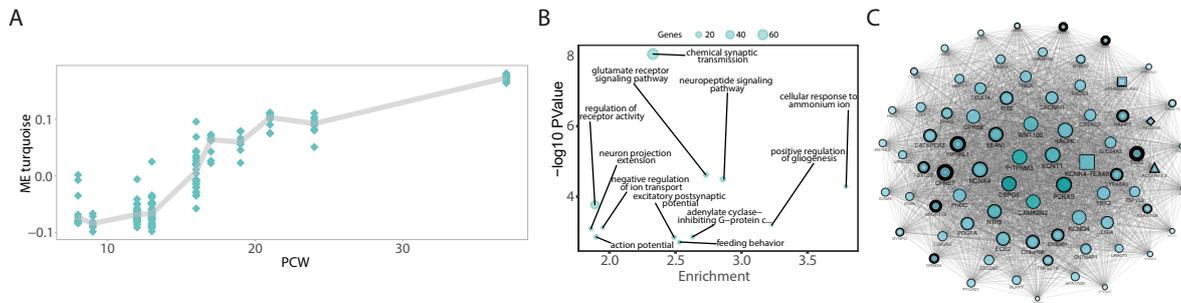


Fig. 15: Characterisation of the BrainSpan turquoise module. (A) Ribbon chart representing the turquoise ME (Y-axis) trend over developmental time (X-axis). (B) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of turquoise genes enriched for the visualised category. The enrichment, defined as the ratio between the number of turquoise genes belonging to a given category and the number of turquoise genes expected to be found by chance for that category, is represented on the X-axis. (C) Network reconstruction for the top 75 turquoise genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

The pink module instead included mostly genes involved in membrane potential control and ion channels (Fig. 16B). Central genes of the pink modules were indeed voltage-gated potassium channels, such as *KCNC3* and 4. Mutations in *KCNC3* cause developmental defects of the nervous system²⁸². Moreover, genes such as *CASKIN1*, *CAMK2B* are related to calcium intracellular signaling regulating dendritic morphology and neuronal homeostasis^{283,284}. *KIF2C* is instead related to intracellular trafficking towards dendrites²⁸⁵. *AATK* was found to promote neuronal differentiation²⁸⁶. Genes such as *ABLIM2*, *SEMA4A*, and *CARMIL2*, showed particularly high betweenness, compared to the other centrality indices, and are related to axonal growth and axon guidance^{287,288,289} (Fig. 16C).

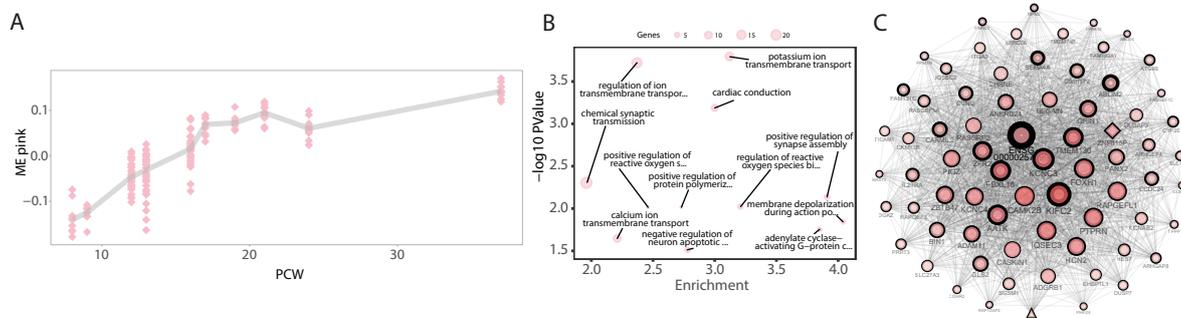


Fig. 16: Characterisation of the BrainSpan pink module. (A) Ribbon chart representing the pink ME (Y-axis) trend over developmental time (X-axis). (B) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of pink genes enriched for the visualised category and the number of pink genes expected to be found by chance for that category, is represented on the X-axis. (C) Network reconstruction for the top 75 pink genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Interestingly, also the midnight-blue module was positively correlated with stage, although its ME did not show a constant increase in expression as turquoise and pink, but rather a drop at very late developmental stages (Fig. 17A). This module was enriched in genes involved in general telencephalic development, with upper layer neuron determinants such as *CUX2* and *SATB2* being

among the top nodes defining this module (**Fig. 17B and C**). Among the most central midnight-blue genes we found FRMPD4, which regulates dendritic spine morphogenesis²⁹⁰, CHN2, mediating semaphorin-dependent synapse pruning²⁹¹, and ITPR1, of which missense mutations cause spinocerebellar ataxia²⁹² (**Fig. 17C**).

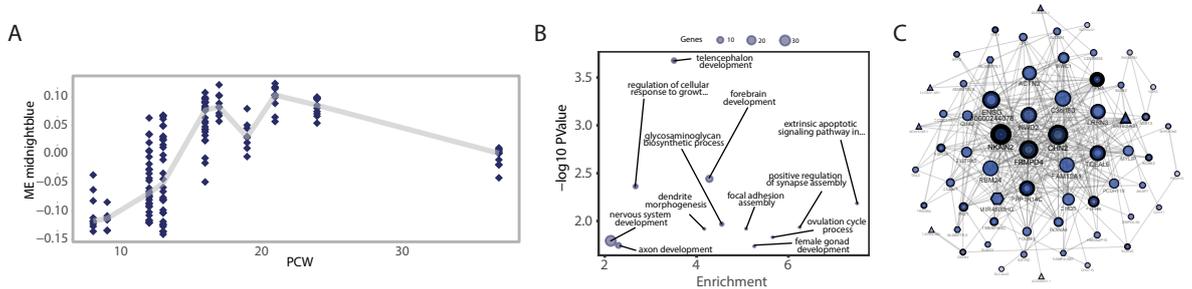


Fig.17: Characterisation of the BrainSpan midnight-blue module. (A) Ribbon chart representing the midnight-blue ME (Y-axis) trend over developmental time (X-axis). **(B)** Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of midnight-blue genes enriched for the visualised category. The enrichment, defined as the ratio between the number of midnight-blue genes belonging to a given category and the number of midnight-blue genes expected to be found by chance for that category, is represented on the X-axis. **(C)** Network reconstruction for the top 75 midnight-blue genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Other modules with a positive correlation along development were the grey60 and the magenta one (**Fig. 18A and 19A**). Grey 60 module was enriched mostly in genes involved in action potential and membrane channels control. Indeed, highly central grey60 genes were: i) SHANK1, which mutations have been linked to ASD and which function is to regulate excitatory synaptic strength and number^{293,294,295}; ii) CACNAC1, a voltage-gated calcium channel subunit implicated in neuropsychiatric syndromes such as schizophrenia and depression as well as in neurodevelopmental disorders^{296,296}; iii) PITPNM2, which acts as regulator of phosphoinositide-dependent pathways and has been implicated in neurite outgrowth²⁹⁷; iv) JAG2, a mediator of the Notch pathway and regulator motor-neuron progenitor differentiation²⁹⁸ (**Fig. 18B and C**)

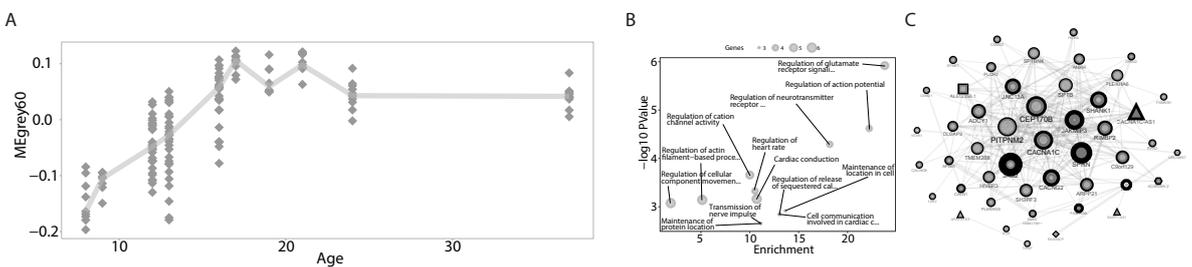


Fig. 18: Characterisation of the BrainSpan grey60 module. (A) Ribbon chart representing the grey60 ME (Y-axis) trend over developmental time (X-axis). **(B)** Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of grey60 genes enriched for the visualised category. The enrichment, defined as the ratio between the number of grey60 genes belonging to a given category and the number of grey60 genes expected to be found by chance for that category, is represented on the X-axis. **(C)** Network reconstruction for the top 75 grey60 genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

The magenta module had stage-specific positive correlation, particularly at PCW12-13. Magenta genes were instead mostly enriched in important synaptic regulators such as CDKL5, which regulates neurite morphogenesis and synapse formation²⁹⁹, DOCK4, which regulate the expression of AMPA and NMDA receptor subunits³⁰⁰, SYT6, belonging family of synaptotagmins³⁰¹, and NYAP2, involved in neurite elongation³⁰² (**Fig. 19B and C**).

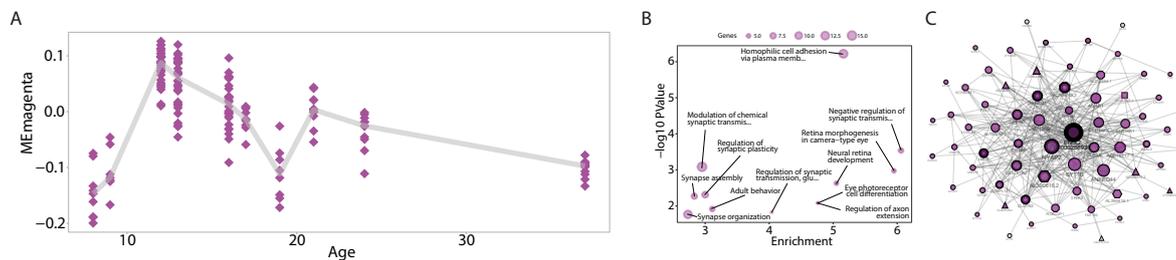


Fig. 19: Characterisation of the BrainSpan magenta module. (A) Ribbon chart representing the magenta ME (Y-axis) trend over developmental time (X-axis). (B) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of magenta genes enriched for the visualised category. The enrichment, defined as the ratio between the number of magenta genes belonging to a given category and the number of magenta genes expected to be found by chance for that category, is represented on the X-axis. (C) Network reconstruction for the top 75 magenta genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Conversely, yellow and black modules showed a negative correlation with stage (**Fig. 20A and 21A**). The yellow module was enriched mostly in cell cycle regulators (**Fig. 20B**). Of note, the genes from this module drastically reduced their expression between PCW9 and PCW12, again in line with the increase in neurogenesis at this time-points. Central genes in the yellow module comprised several kinesin genes involved in microtubules dynamics during cell cycle. KIF20A and KIF11 have reported roles in aRG proliferation. Loss of function of KIF20A causes early cell cycle exit and precocious neuronal differentiation³⁰³, while KIF11 was reported to be essential for proper neurogenesis in zebrafish³⁰⁴. AURKA and B were also found to be central yellow genes and they have important roles in regulating cell cycle, and in particular asymmetric cell division, polarity and migration³⁰⁴. CDC20, CCNB2 and BUB1B are other highly connected genes in the yellow module and they were reported to regulate cell cycle and self-renewal in neuronal progenitors^{305,306,307} (**Fig. 20C**).

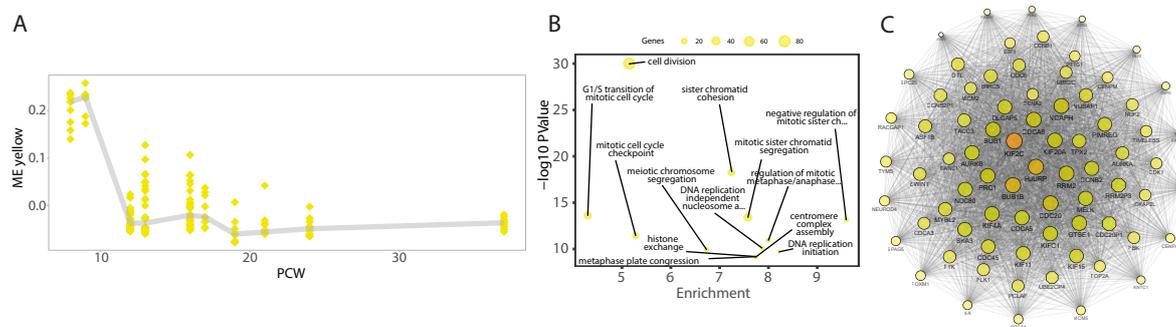


Fig. 20: Characterisation of the BrainSpan yellow module. (A) Ribbon chart representing the yellow ME (Y-axis) trend over developmental time (X-axis). (B) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of yellow genes enriched

for the visualised category. The enrichment, defined as the ratio between the number of yellow genes belonging to a given category and the number of yellow genes expected to be found by chance for that category, is represented on the X-axis. (C) Network reconstruction for the top 75 yellow genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Central genes of the black module were instead mainly related to DNA metabolism, in particular DNA replication (Fig. 21B). WDHD1 plays important roles in DNA replication upon STAT3 targeting³⁰⁸. LIN9 inhibits DNA synthesis and oncogenic transformation by interacting with retinoblastoma 1 protein and its loss causes genomic instability³⁰⁹. NOC3L is a homologue of NOC3, which plays important roles in regulating DNA replication at the beginning of S-phase³¹⁰. PPM1D mediates the relief of p53-mediated cell cycle arrest. HAUS6 is instead part of the augmin complex, which plays important roles in microtubule attachment to the kinetochore for mitotic spindle formation and disruption of which causes p53-dependent apoptosis and aborts brain development (pre-print, doi: 10.1101/2020.11.18.388694). Interestingly, the HAUS6 pseudogene, HAUS6P1, showed high betweenness in the black network (Fig. 21C).

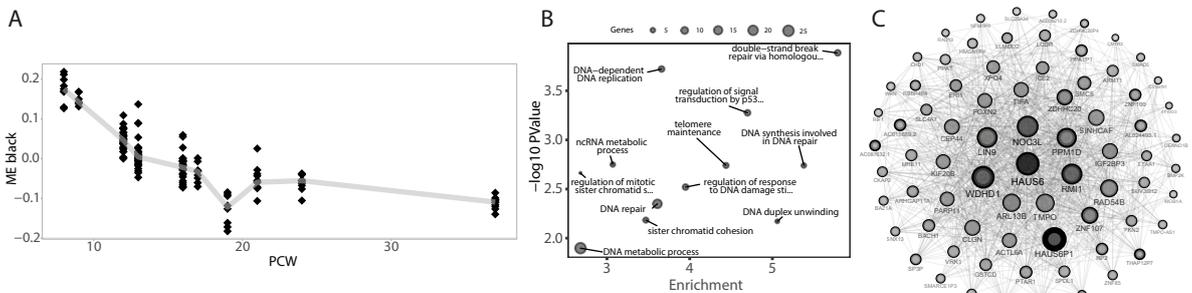


Fig. 21: Characterisation of the BrainSpan black module. (A) Ribbon chart representing the black ME (Y-axis) trend over developmental time (X-axis). (B) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of black genes enriched for the visualised category. The enrichment, defined as the ratio between the number of black genes belonging to a given category and the number of black genes expected to be found by chance for that category, is represented on the X-axis. (C) Network reconstruction for the top 75 black genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Finally, we focused on modules with peculiar behaviour along cortical development. Indeed, the red and blue module followed an U-shaped trend along PCW, with a peak of positive correlation at early stages, followed by a decrease at intermediate time points, and again an increase at the end of gestation (Fig. 22A and 23A). Surprisingly, both modules were related to extracellular matrix, particularly the red one. Red central genes were indeed collagen subunits such as COL5A1, COL1A2 and COL8A1. Interestingly, COL8A1 has been involved in Tourette Syndrome³¹¹ and, in general, collagens have been linked to behavioural phenotypes in mice³¹². In this module, we also found as central genes different laminin subunits such as LAMC3, LAMA4 and LAMA2. LAMC3 has been linked to gyrfication abnormalities in humans³¹³. Mutation in the extracellular-matrix-related gene NID1, another central red gene and interactor of LAMC1, was instead linked to occipital cephalocele³¹⁴ (Fig. 22B and C).

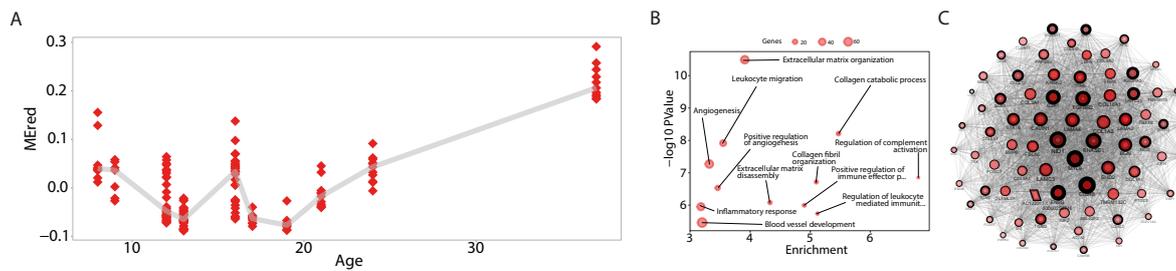


Fig. 22: Characterisation of the BrainSpan red module. (A) Ribbon chart representing the red ME (Y-axis) trend over developmental time (X-axis). (B) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of red genes enriched for the visualised category. The enrichment, defined as the ratio between the number of red genes belonging to a given category and the number of red genes expected to be found by chance for that category, is represented on the X-axis. (C) Network reconstruction for the top 75 red genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Also the blue module included central genes related to extracellular matrix regulation such as PLPP3, involved in cell migration and adhesion through dephosphorylation of extracellular sphingosine-1-phosphate³¹⁵, LFGN, which if mutated leads to extracellular matrix remodeling³¹⁶, and EMILIN3, a protein located at the interface of elastin microfibril³¹⁷. Importantly, one of the top categories enriched for the blue module was linked to astrocyte differentiation. Two of the main central genes in this module were LGALS3BP and AXL. LGALS3BP has been found to be part upregulated signature of astrocyte-secreted extracellular matrix upon injury³¹⁸, while AXL is a tyrosine kinase receptor that transduces signals from the extracellular matrix into the cytoplasm and it is highly expressed in aRG and astrocytes³¹⁹ (Fig. 23B and C).

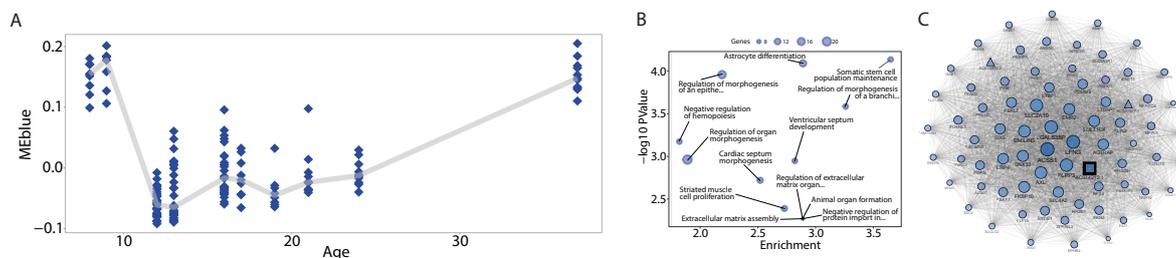


Fig. 23: Characterisation of the BrainSpan blue module. (A) Ribbon chart representing the blue ME (Y-axis) trend over developmental time (X-axis). (B) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of blue genes enriched for the visualised category. The enrichment, defined as the ratio between the number of blue genes belonging to a given category and the number of blue genes expected to be found by chance for that category, is represented on the X-axis. (C) Network reconstruction for the top 75 blue genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

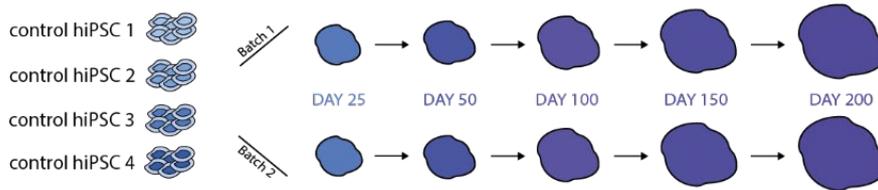
In summary, we characterised the gene expression landscape of human pre-natal cortex, identifying developmental transitions defined by precise transcriptional footprints that we described exploiting WGCNA. Indeed, we were able to extract networks of co-expressed genes with characteristic behaviours during the pre-natal period, thus prioritising well-known as well as novel hubs relevant for human cortical development.

4.2 Cortical Brain Organoids globally resembled human fetal cortex and evolved in a two-step fashion

After having defined the transcriptional changes characteristic of human fetal corticogenesis, we profiled an in-house cohort of cortical brain organoids (CBO), differentiated using an adaptation of the protocol published by Pasca et al in 2015¹⁰⁴, in order to benchmark cortical organoidogenesis against its *in vivo* counterpart. Briefly, the dorsal telencephalon fate was achieved by neuroepithelial specification followed by neural progenitors expansion and differentiation towards glutamatergic neurons thanks to heavy sequential patterning using small molecules and growth factors. Our CBO transcriptomic cohort included 43 samples obtained by two independent differentiation rounds of four hiPSCs lines reprogrammed from different individuals and densely profiled over 200 days (time-points: hiPSC, 25, 50, 100, 150 and 200), thereby covering the equivalent in time of early to mid-late gestation. Importantly, every CBO sample consisted of a single organoid, thus providing the possibility to quantify inter-individual and technical variability accurately. Additionally, we profiled a cohort of primary fetal central nervous system tissues and 2D cultures. We leveraged these two datasets in order to identify cortex-specific determinants and analyse their behaviour in organoids. A summary of samples available is outlined in **Fig. 24A**, together with representative images and immunostainings of CBO at different time-points (**Fig. 24B**).

A

CBO dataset



Human fetal CNS tissue dataset



Area	n° samples	Age
Cortex	2	PCW 13 & 15
MGE	2	PCW 13 & 15
LGE	1	PCW 13
Striatum	1	PCW 15
Diencephalon	2	PCW 13 & 15
MidBrain	2	PCW 13 & 15
Pons	2	PCW 13 & 15
Cerebellum	2	PCW 13 & 15
Medulla	1	PCW 15
Spinal Cord	1	PCW 13

Human fetal 2D-cultured cells



Area	n° samples	Age
Cortex	6	PCW 11
Cortex	5	PCW 19

B

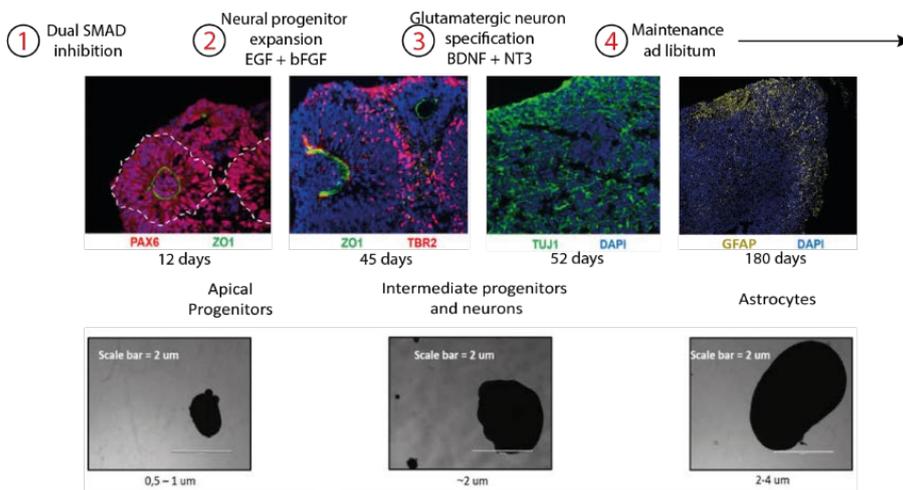


Fig. 24: Graphical representation of the in-house generated CBO and fetal brain datasets. (A) Graphical representation of the in-house generated CBO and fetal brain datasets detailing number of individuals, batches and time-points for CBO and number of samples, brain area and PCW for fetal samples. **(B)** Representative images of organoid differentiation with immunostaining for key population markers (PAX6 = apical progenitors, ZO-1 = neuroepithelial cells, TBR2 = intermediate progenitors, TUJ1 = post-mitotic neurons, GFAP = astrocytes).

Dimensionality reduction through PCA showed a clear distribution of organoid samples by stage when considering PC1-2, with day 25 and day 50 organoids forming individual clusters separated from day 100-200 organoids, which instead resulted less obviously stage-wise distributed and thus suggesting a slow-down of transcriptional maturation from day 100 onwards. Mature organoids positioned themselves in PC1-2 coordinates close to human fetal brain tissues. Strikingly, cortical tissue samples were interspersed among mature CBO, indicating that > 100 days organoids closely resemble the transcriptome of fetal cortex (PCW13 and 15). Interestingly, also the lateral ganglionic eminence tissue clustered with cortical samples and mature CBOs. 2D cultured fetal cortical cells (primary neural stem cells) instead clustered apart from all brain tissues and organoids, also from CBO at very early stages (**Fig. 25**).

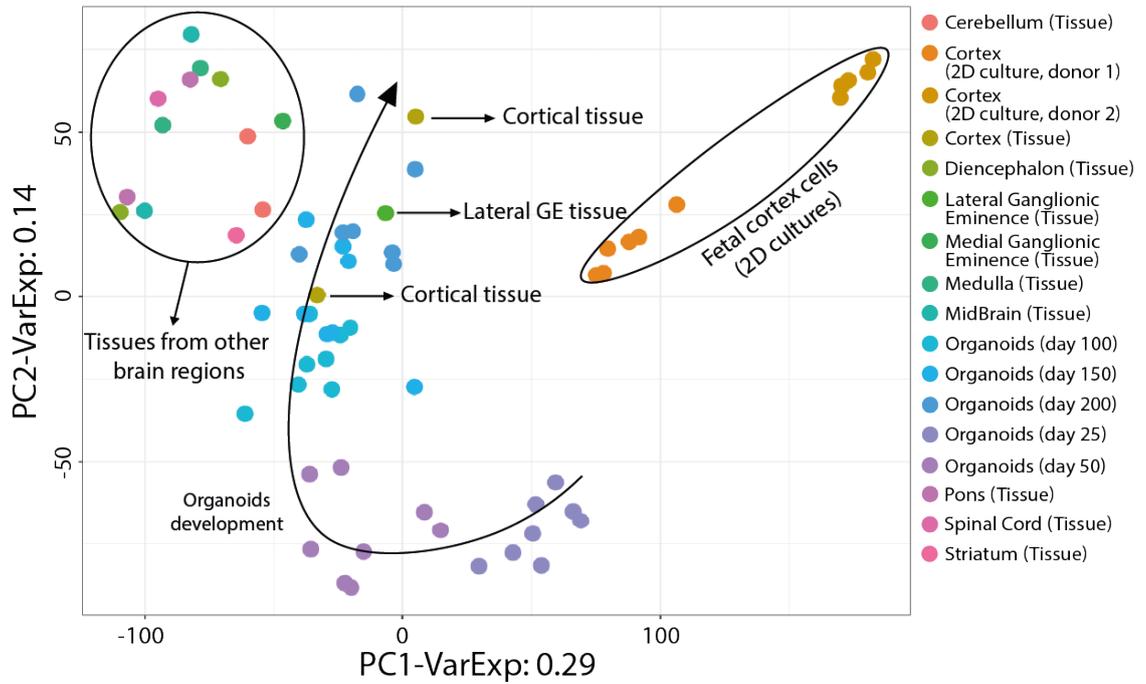


Fig. 25: Dimensionality reduction of CBO, fetal brain tissues and 2D cultured fetal cortical cells. Principal component analysis on internal CBO, fetal brain tissues and 2D cultured fetal cortical cells was performed. PC1 and 2 were visualised. Dot colour-code is depicted in the figure legend. The arrow highlighted the distribution of CBO samples throughout stage-progression. Fetal samples were also highlighted differently as an exemplification of the figure.

Focusing on CBO development, we performed PCA and visualisation of PC1-2 on our complete organoid dataset, including parental hiPSCs transcriptomes, to better characterise stage-wise differences. We observed a clear separation between hiPSCs and all CBO samples on PC2, while stage was clearly driving PC1, again with day 25 and day 50 CBO forming individual and well defined clusters and day 100-200 CBO forming an unique one. This result corroborated the hypothesis that global transcriptomic differences between day 100, 150 and 200 CBO were milder than the ones observed at early stages (**Fig. 26**).

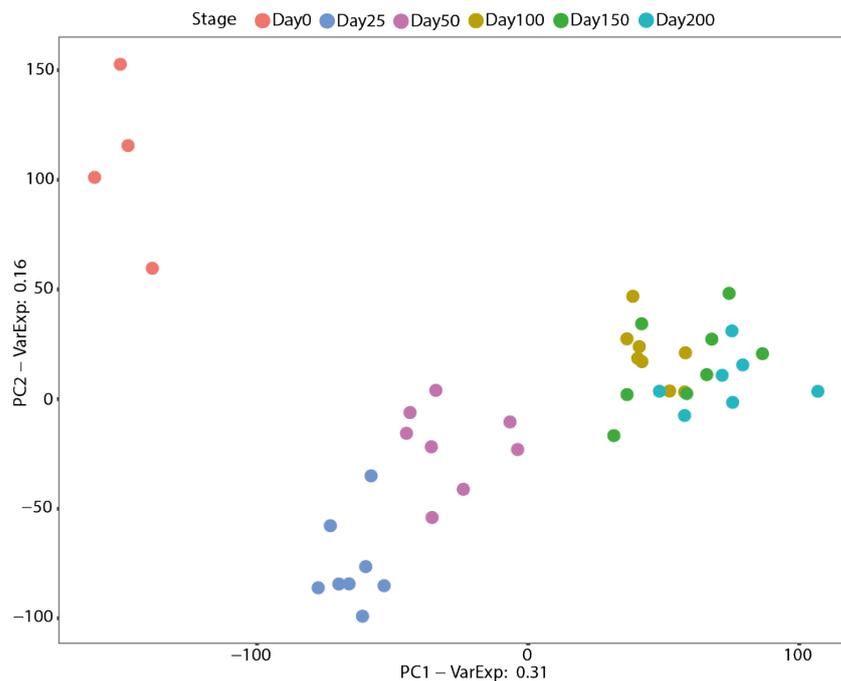


Fig. 26: Dimensionality reduction of the CBO dataset. Principal component analysis performed on the CBO dataset including the iPSC stage. PC1 and 2 were visualised. Dot colour-code is depicted in the figure legend.

We then characterised the top 150 genes driving PC1, highlighting that positively correlated genes were involved in neuronal maturation and synapse formation, while negatively correlated ones played a role in cell cycle, in line with the decrease of cell proliferation along CBO development (**Fig. 27A**). The top 150 PC2 genes were instead related to more general categories, many of them being not clearly ascribable to relevant neuronal functions (**Fig. 27B**).

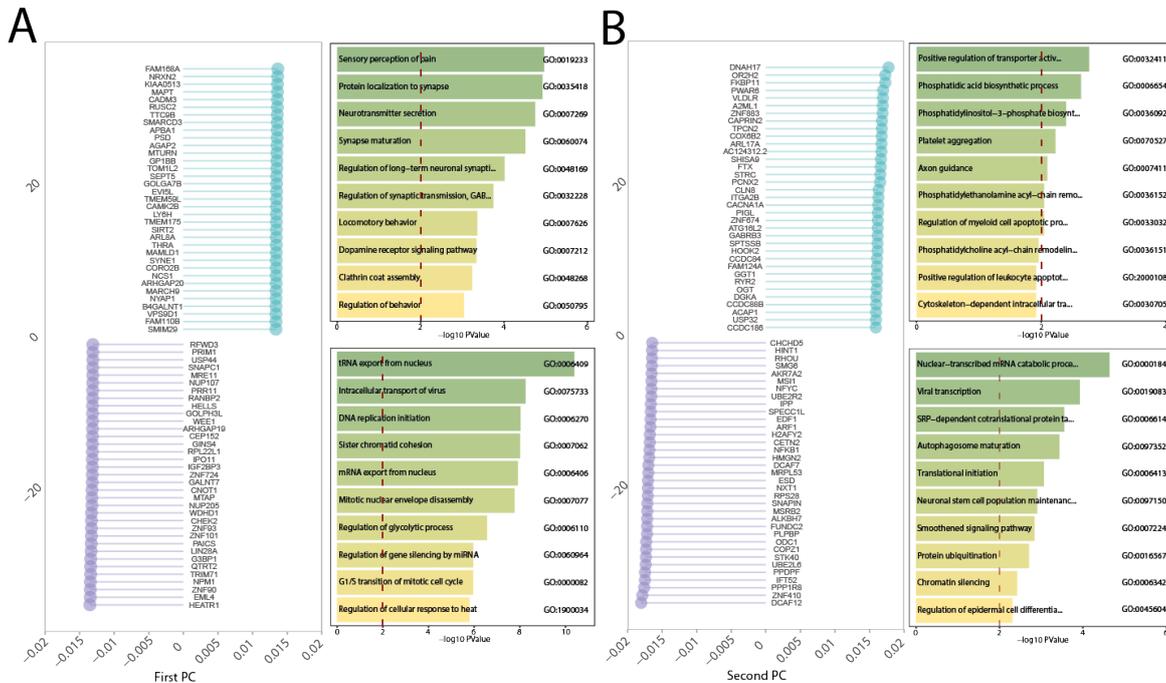


Fig. 27: Characterization of top genes driving sample distribution in the PCA of the CBO dataset. (A) Top 35 genes positively associated with PC1 and top 35 genes negatively associated with PC1 were represented as lollipop charts. X-values represented the PC1 loadings for the visualised genes. Y-values represented the order of genes by PC1 loadings. Gene ontology analysis for biological processes was performed on the top 150 PC1 with positive or negative loadings. X-values represent $-\log_{10}P$ Value, which was considered significant when above 2 (P Value < 0.01). (B) Top 35 genes positively associated with PC2 and top 35 genes negatively associated with PC2 were represented as lollipop charts. X-values represented the PC2 loadings for the visualised genes. Y-values represented the order of genes by PC2 loadings. Gene ontology analysis for biological processes was performed on the top 150 PC2 with positive or negative loadings. X-values represent $-\log_{10}P$ Value, which was considered significant when above 2 (P Value < 0.01).

We next characterised stage-wise differences in gene expression by differential expression analysis (DEA). We performed DEA comparing each stage against the previous one for a total of 5 comparisons. Considering as differentially expressed genes (DEGs) genes with false discovery rate (FDR) < 0.05 and absolute fold change (FC) greater than 2, we obtained: i) 6524 DEGs (3531 downregulated and 2993 upregulated) for the comparison CBO day25 vs iPSCs, ii) 3625 DEGs (1276 downregulated and 2349 upregulated) for CBO day 50 vs 25, iii) 2825 DEGs (1283 downregulated and 1542 upregulated) for CBO day 100 vs 50; iv) 242 DEGs (33 downregulated and 209 upregulated) for CBO day 150 vs 100 and, v) 482 DEGs (182 downregulated and 300 upregulated) for CBO day 200 vs 150 (**Fig. 28**). This approach revealed drastic reduction of the number of DEGs after day 100, identifying a two-step evolution of CBO along time with fast transcriptomic changes

until day 100 and the achievement of a plateau from day 100 onwards, thus confirming what we observed in PCA.

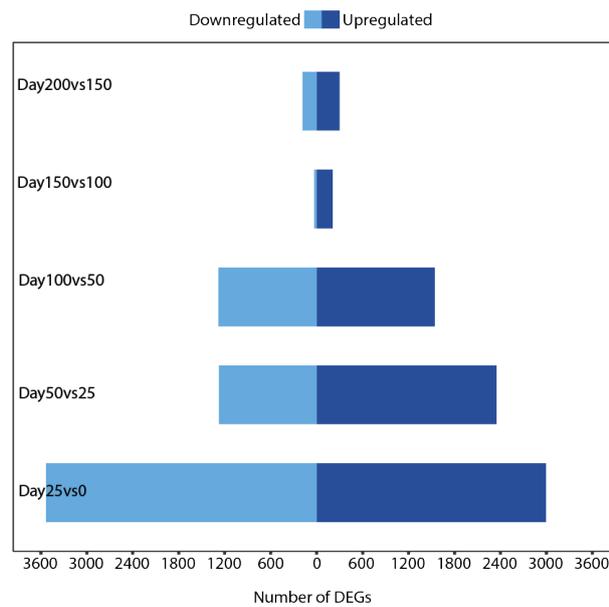


Fig. 28: Differential expression analysis of the CBO dataset. (A) Bar plot showing the number of upregulated and downregulated DEGs with absolute FC > 2 and FDR < 0.05 of pair-wise DEAs for the CBO dataset considering each stage against the previous one in development. The X-axis indicated the number of DEGs, while the Y-axis showed the comparison.

These results led us to the characterisation of the fast evolution phase by gene ontology of the DEGs for every comparison until day 100. Upregulated genes were enriched in categories related to neuronal function, axon and synaptic development, ion transport. Downregulated categories were instead related to very diverse categories, with several of them associated with cell cycle (DNA-replication) and to extracellular matrix (**Fig. 29A and B**).

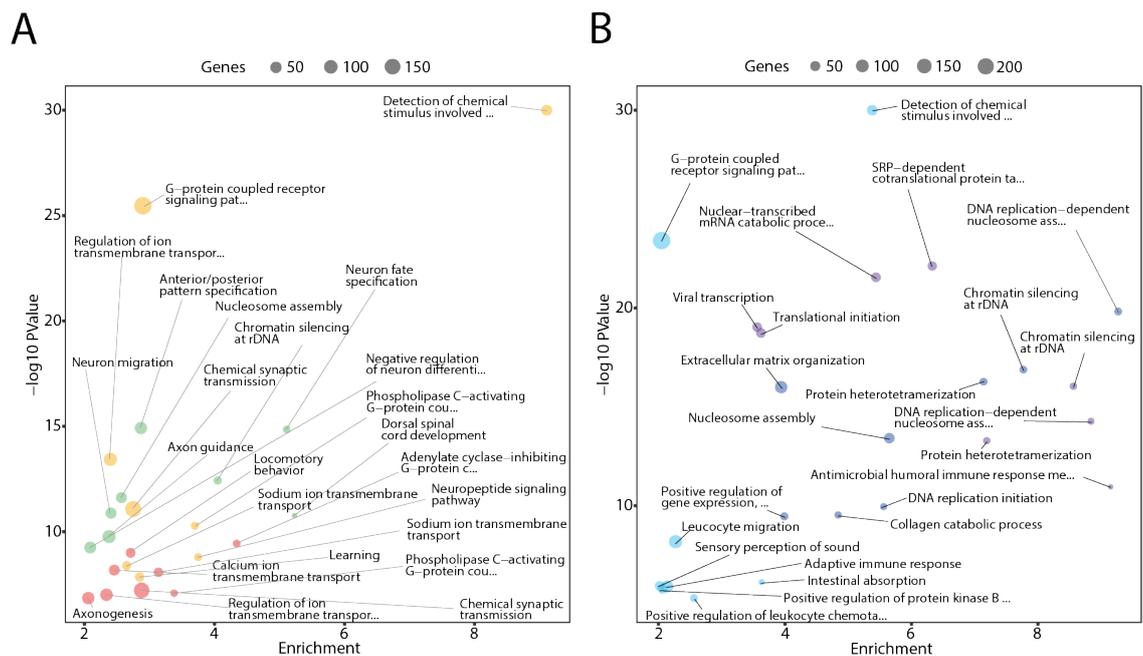


Fig. 29: Functional analysis of CBO DEAs until day 100. (A) Bubble plot summarising the top 8 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (PValue < 0.01) (Y-axis) and enrichment greater than 2 (X-axis) for upregulated DEGs in the comparisons Day25 vs Day0 (green), Day50 vs Day100 (yellow) and Day100 vs Day50 (red). (B) Bubble plot summarising the top 8 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (PValue < 0.01) (Y-axis) and enrichment greater than 2 (X-axis) for downregulated DEGs in the comparisons Day25 vs Day0 (sky-blue), Day50 vs Day100 (blue) and Day100 vs Day50 (violet).

Surprisingly, when analysing categories enriched for the comparison day 150 vs 100 and day 200 vs 150, characterising the slow-evolving phase of CBO, we found upregulation of genes involved in muscle function and fibroblast proliferation and, again, downregulation of cell division (Fig. 30A and B). The unexpected upregulation of genes involved in off-target cell types specification led us to look at single genes composing these categories. We found that many of them - such as TMOD1, DES, TNNC11, MYBPC1, MYH2, A2M - are expressed also in differentiating astrocytes, suggesting that the astrogenic phase begun starting from day 100 onwards in CBO^{320,321,322,323,324,325,326} (Fig. 30A and B).

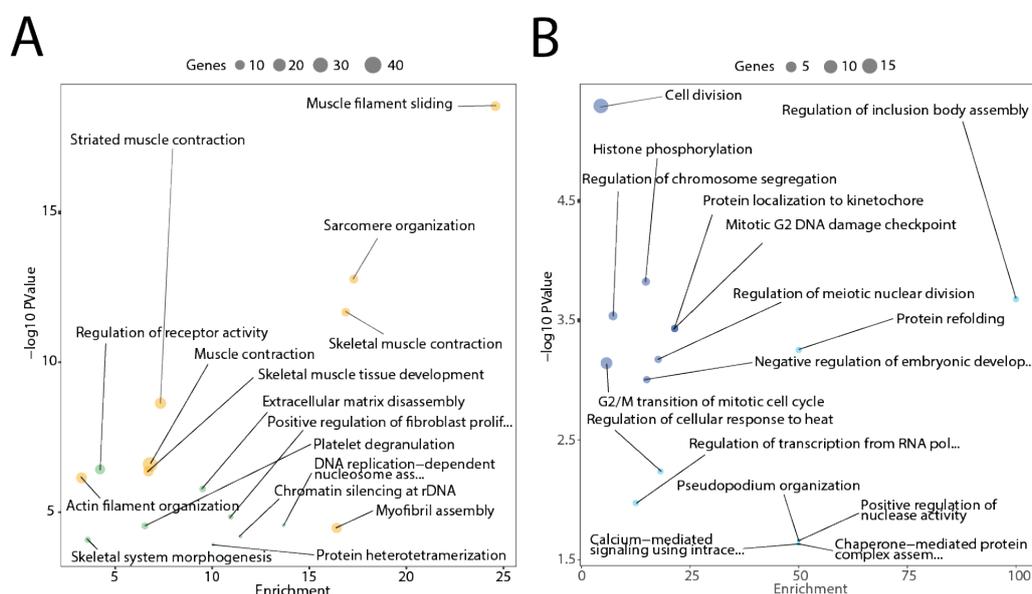


Fig. 30: Functional analysis of CBO DEAs from day 100 until day 200. (A) Bubble plot summarising the top 8 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (PValue < 0.01) (Y-axis) and enrichment greater than 2 (X-axis) for upregulated DEGs in the comparisons Day150 vs Day100 (green) and Day200 vs Day150 (yellow). (B) Bubble plot summarising the top 8 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (PValue < 0.01) (Y-axis) and enrichment greater than 2 (X-axis) for downregulated DEGs in the comparisons Day150 vs Day100 (sky-blue) and Day200 vs Day150 (blue).

We then analysed the dynamics of differential expression along CBO development by looking at the directionality of DEGs for all CBO comparisons. To do this, we plotted the fold-change of consecutive comparisons (e.g. day 50vs 25 against 100 vs 50, etc), therefore visualizing genes that were modulated in a consistent or different direction across stages. This approach allowed the identification of four classes of genes for every pair of comparisons: i) genes upregulated in both pairs; ii) genes downregulated in both pairs; and iii) iv) genes upregulated in one pair and downregulated in the other and *vice versa*.

The comparison of day 50 vs 25 and 100 vs 50 revealed similar distribution of common DEGs in the four defined classes. Interestingly, among the top genes downregulated at day 50 compared to day 25 and upregulated at day 100 vs 50 we found key genes for neuronal differentiation. Indeed, EOMES is an intermediate progenitor marker³²⁷, LHX2 is a dorsal telencephalon determinant³²⁸, and FEZF2 regulates CTIP2 mediating the extension of axons towards subcortical targets³²⁹. These results indicated that several key cell fate determinants follow a dynamic expression during cortical organoidogenesis (**Fig. 31**).

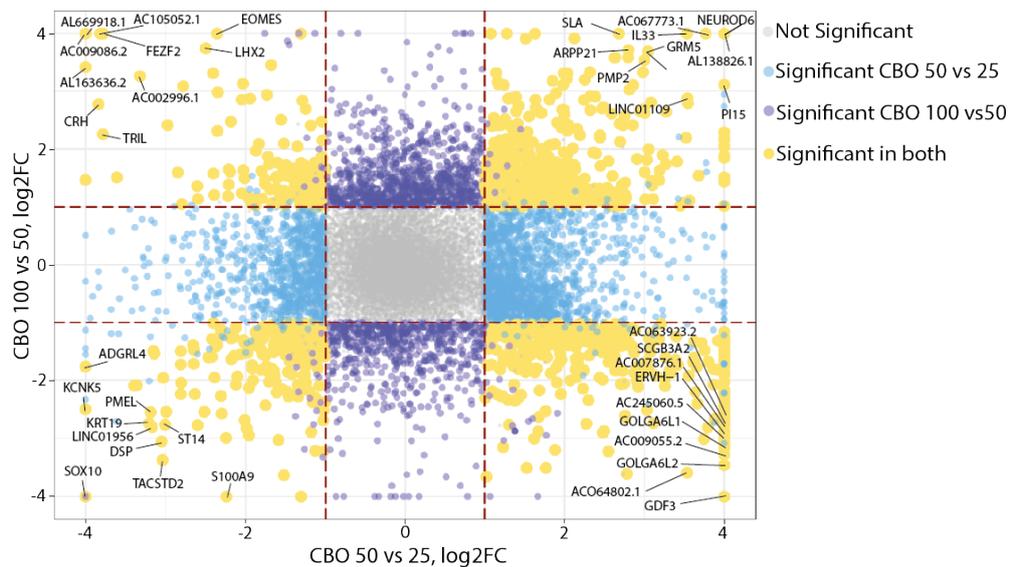


Fig. 31: FC scatter plot CBO day50vs25 against day100vs50. Scatter plot showing the log₂FC of DEGs for the comparisons Day50 vs Day25 (X-axis) and Day100 vs 50 (Y-axis). DEGs specific for one of the two comparisons or common to both were highlighted according to the colour-code depicted in the figure legend.

For the pair of comparisons day 100 vs 50 and day 150 vs 100 we observed consistent upregulation of astrocytic markers (HEPACAM, AQP4 and AGT^{330,331,332}). This result was in line with the gene ontology data obtained for the comparisons day 150 vs 100 and day 200 vs 150, where we observed upregulation of genes mainly express in astrocytes within the brain. Importantly, we also measured consistent upregulation of SCGN and CALB2, two markers of GABAergic interneurons^{333,334}, which, from a theoretical point of view, should not be found in mature CBO given the heavy patterning towards dorsal telencephalon. Despite this, other organoid systems entailing similar degrees of patterning showed expression of GABAergic genes³³⁵ and our lateral ganglionic eminence sample clustered with mature CBO in PCA (**Fig. 25**), in line with the upregulation of these inhibitory interneurons markers in CBO (**Fig. 32**)

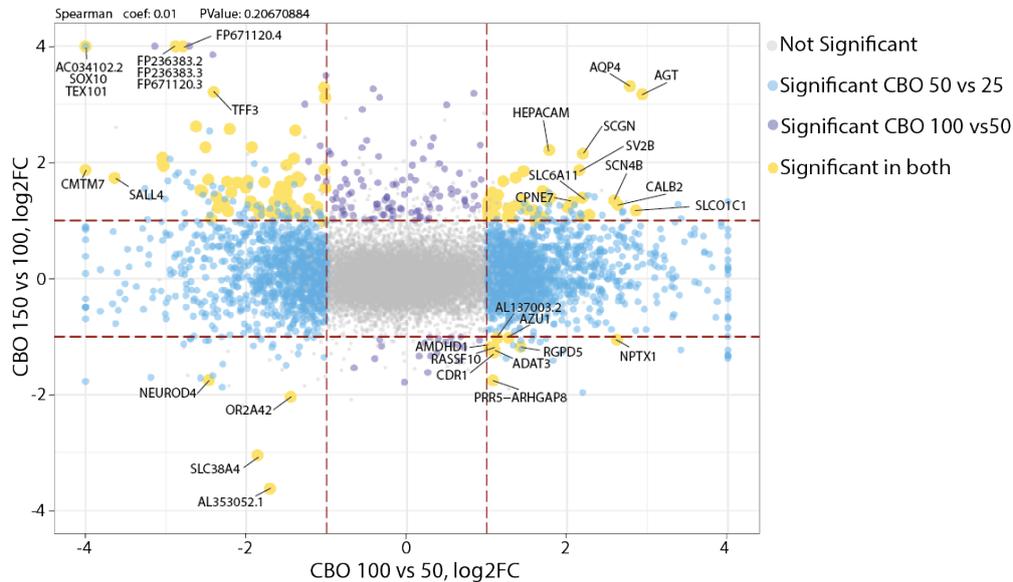


Fig. 32: FC scatter plot CBO day100vs50 against day150vs100. Scatter plot showing the log₂FC of DEGs for the comparisons Day100 vs Day50 (X-axis) and Day150 vs 100 (Y-axis). DEGs specific for one of the two comparisons or common to both were highlighted according to the colour-code depicted in the figure legend.

The analysis for the pair of comparisons day 150 vs 100 and 200 vs 150 showed similar results, with the astrocytic markers HEPACAM and AQP4 as well as the GABAergic interneuron marker SCGN consistently upregulated (**Fig. 33**). Interestingly, APOE was also among the upregulated genes in day 150 and 200 CBO and a pre-print study found it associated to astrocyte identity in the developing human cortex (doi:10.1101/2020.12.29.424636).

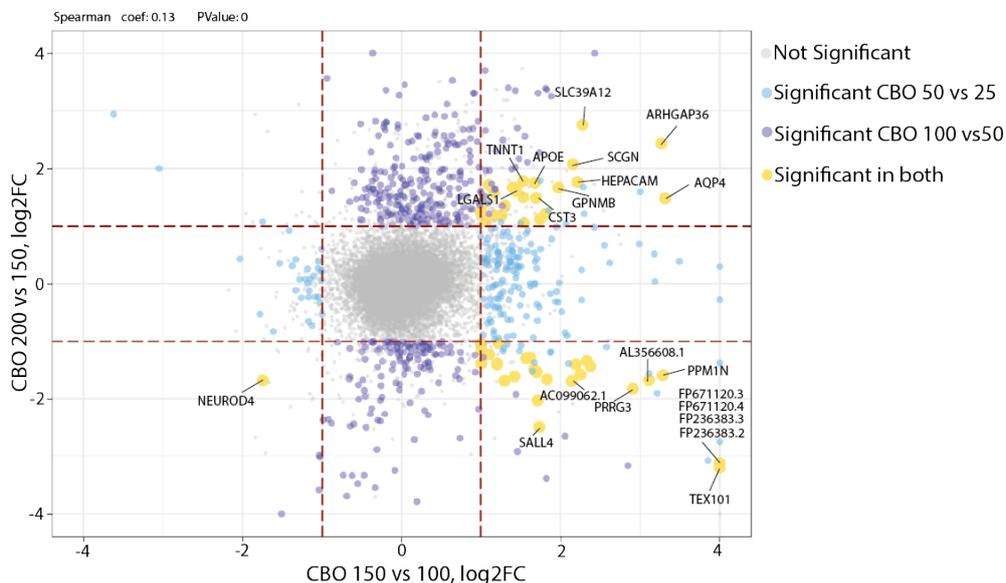


Fig. 33: FC scatter plot CBO day150vs100 against day200vs150. Scatter plot showing the log₂FC of DEGs for the comparisons Day150 vs Day100 (X-axis) and Day200 vs 150 (Y-axis). DEGs specific for one of the two comparisons or common to both were highlighted according to the colour-code depicted in the figure legend.

After characterising the transcriptomic evolution of CBO along differentiation, we sought to understand the cell composition within organoids at all time-points analysed. To do so, we estimated cell type proportions in our bulk transcriptomic dataset using the SCDC deconvolution

algorithm³³⁶ and a scRNAseq atlas of the developing human cortex as reference³³⁷. We considered 6 populations from this atlas: i) excitatory neurons (ExN), ii) inhibitory neurons (In), iii) intermediate progenitors (IP), iv) outer radial glia (oRG), v) cycling progenitors (Pg), and vi) ventricular radial glia (vRG). We observed a clear increase of excitatory neurons along development, with a rapid increase until day 100 mirrored by a drop in ventricular radial glia, cycling progenitors and intermediate progenitors. Interneurons showed the most variable and less clear pattern, probably reflecting the difficulty in discriminating them from excitatory neurons. Interestingly, outer radial glia increased from day 100 onwards (**Fig. 34**). Being the signatures defining these categories identified on the basis of a single cell RNAseq dataset of which we maintained the annotation, the nomenclature “cycling progenitors” defines early progenitors with a prevalent component of cell cycle genes. The fact that oRG is increasing after day 100 while Pg are virtually absent at that time point does not exclude that oRG expresses cell cycle genes, but rather reflects the fact that the signature through which oRG is defined differs significantly from the one of Pg and that cell cycle genes weigh less in oRG definition. We ourselves evaluated the presence of *bona fide* oRG in CBO by staining for HOPX, a oRG specific marker³³⁸.

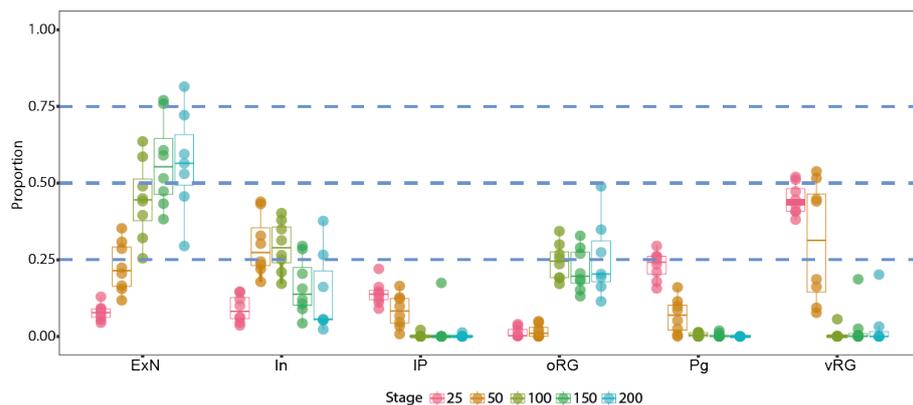


Fig. 34: Bulk deconvolution of the CBO dataset. Box plot showing the estimated proportion of excitatory neurons (ExN), inhibitory neurons (In), intermediate progenitors (IP), outer radial glia (oRG), cycling progenitors (Pg), and ventricular radial glia (vRG) (Y-axis) based on the results given by the SCDC deconvolution algorithm. Every stage of CBO differentiation was displayed with a different colour according to the figure legend.

Upon characterisation of the cell types composing CBO along differentiation, we proceeded with their benchmark against the *in vivo* counterpart. We identified human fetal cortex specific genes by comparing the cortical samples in our internal fetal brain tissue dataset (PCW13 and 15) against hiPSCs and subsequently excluding genes commonly DEGs when performing the same comparisons starting from the other brain areas of our cohort. We considered significant genes with absolute $\log_{2}FC > 3$ and $FDR < 0.05$. We then checked the expression of these genes in the CBO cohort dividing them in upregulated (**Fig. 35**) and downregulated ones (**Fig. 36**).

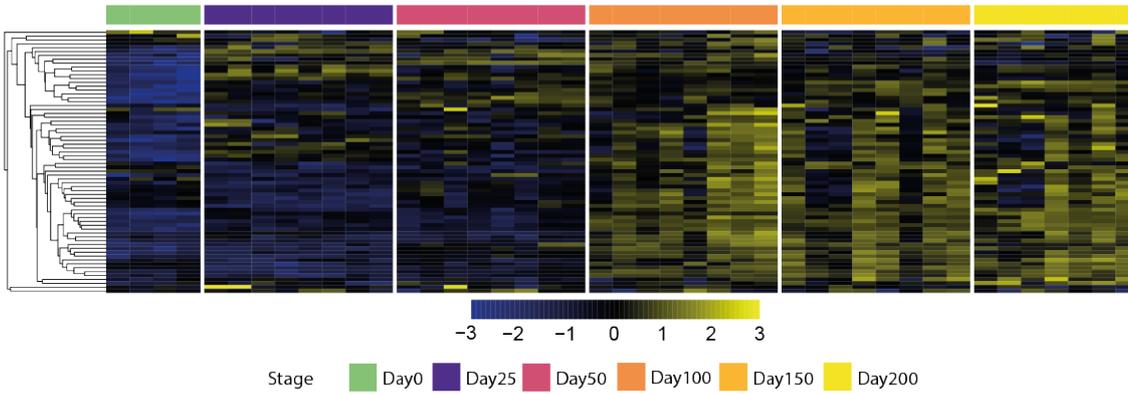


Fig. 35: Visualisation of cortex-specific upregulated DEGs in the CBO dataset. Heatmap showing the behaviour of cortex-specific upregulated DEGs in the CBO dataset. Values represented the Z-score for every gene (row) in all samples (column). The colour-code for stage is depicted in the figure legend.

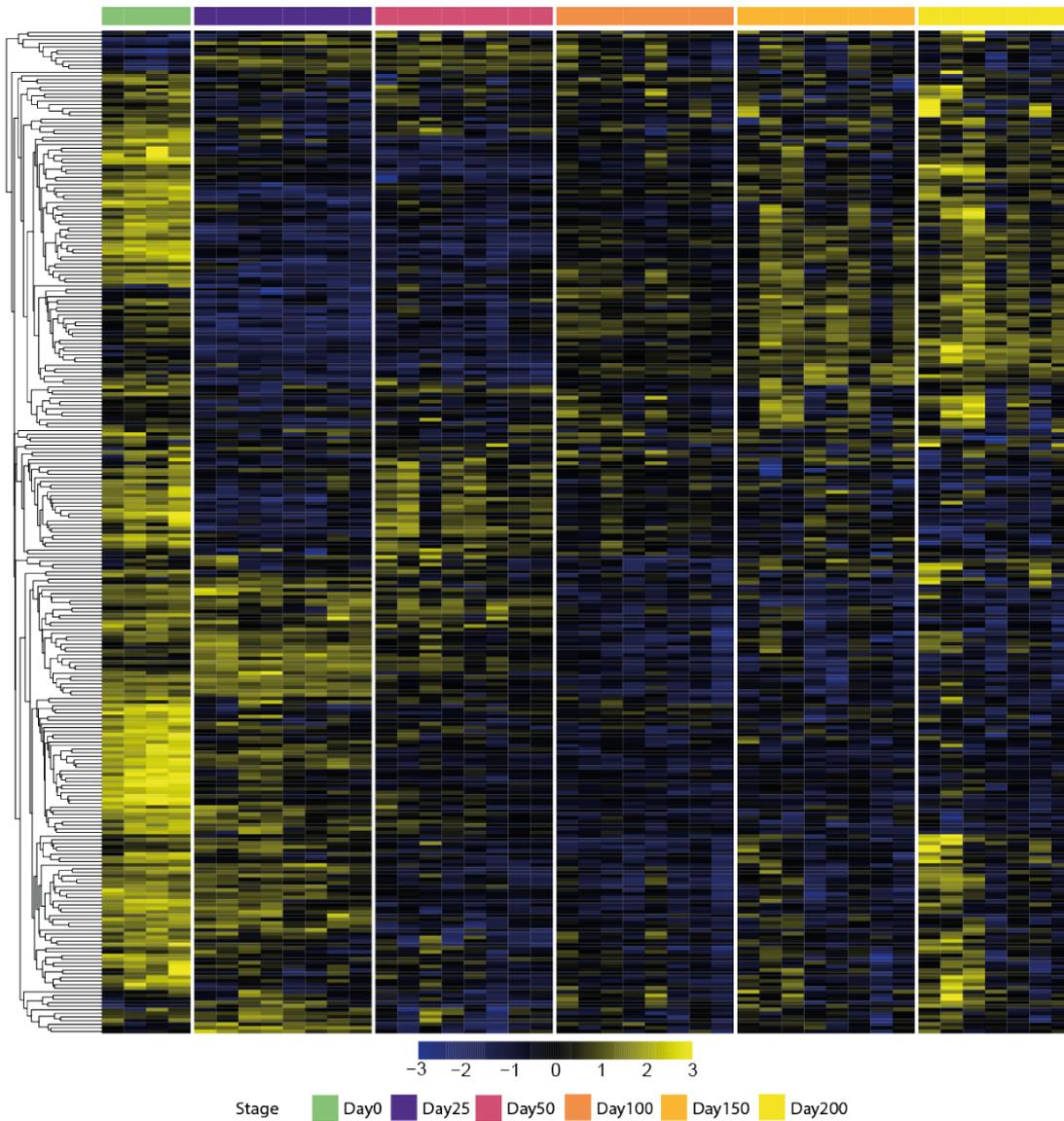


Fig. 36: Visualisation of cortex-specific downregulated DEGs in the CBO dataset. Heatmap showing the behaviour of cortex-specific downregulated DEGs in the CBO dataset. Values represented the Z-score for every gene (row) in all samples (column). The colour-code for stage is depicted in the figure legend.

We looked at the biological pathways in which these genes were involved and we found that upregulated genes were enriched in categories related to glutamatergic neuron function, while

downregulated ones were involved in extracellular matrix functions, among other pathways (**Fig. 37A and B**).

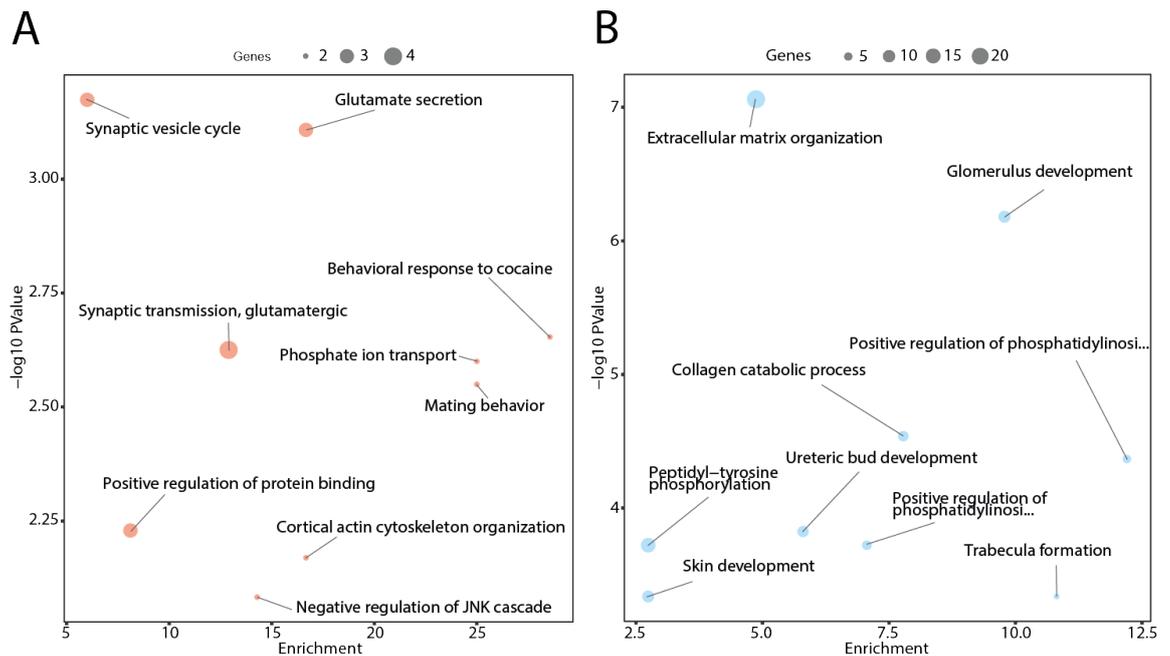


Fig. 37: Functional analysis of cortex-specific DEGs. (A) Bubble plot summarising the top 10 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis) and enrichment greater than 2 (X-axis) for cortex-specific upregulated DEGs. **(B)** Bubble plot summarising the top 10 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis) and enrichment greater than 2 (X-axis) for cortex-specific downregulated DEGs.

Looking at the expression dynamics of cortex specific genes in CBO, we were able to identify several patterns. We classified upregulated and downregulated genes as follows based on the average-linkage hierarchical clustering computed for heatmaps generation: i) cortex-specific upregulated genes stably expressed in CBO; ii) cortex-specific upregulated genes increasingly upregulated along CBO differentiation; iii) cortex-specific downregulated genes increasingly repressed along CBO differentiation; iv) cortex-specific downregulated DEGs increasingly upregulated along CBO differentiation; v) cortex-specific downregulated genes transiently upregulated along CBO differentiation (**Fig. 38**).

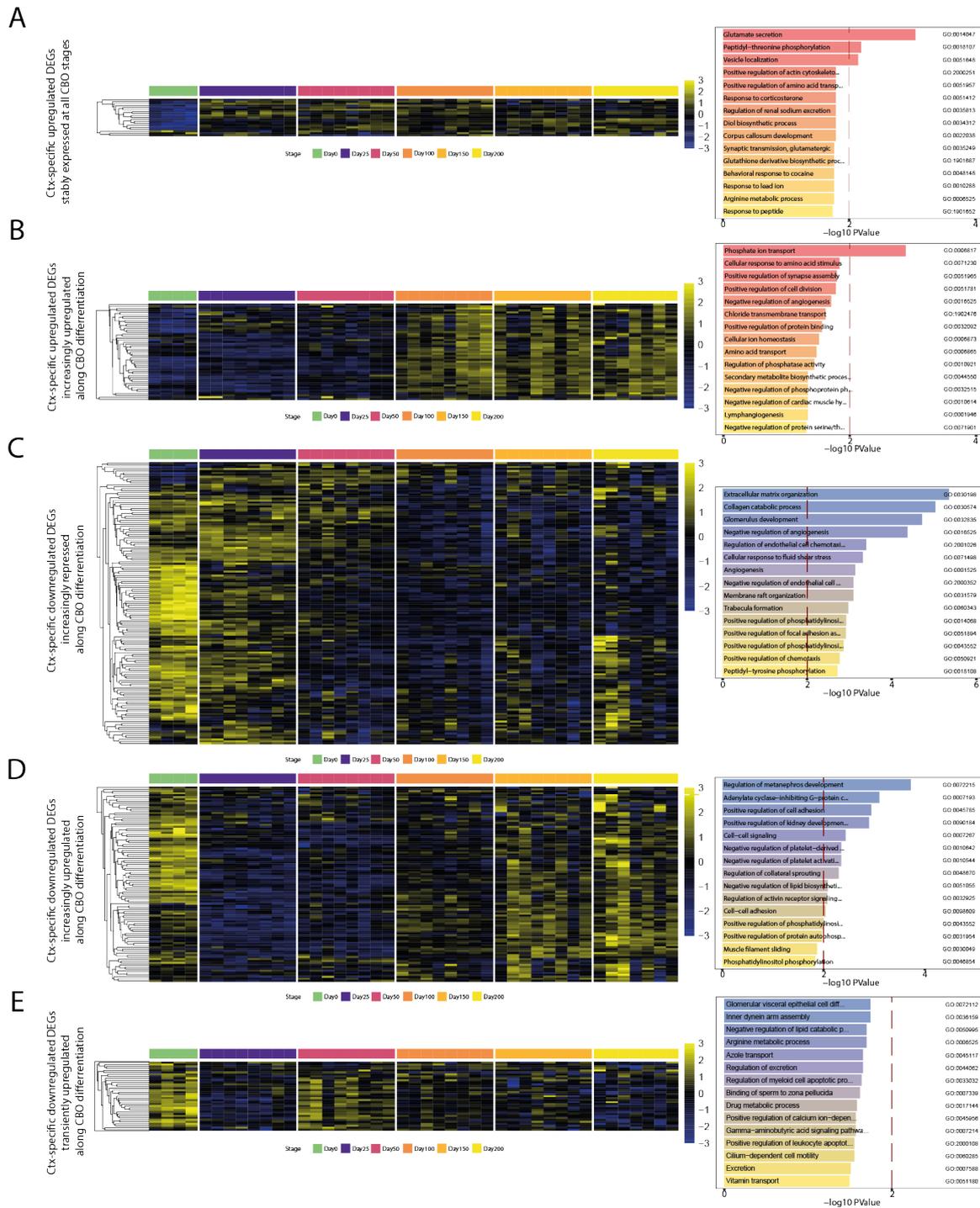


Fig. 38: Clustering of cortex-specific DEGs based on their expression trend during CBO differentiation. Average-linkage hierarchical clustering was used to identify clusters of cortex-specific upregulated or downregulated genes with specific trends during CBO development. (A) Cortex-specific upregulated DEGs stably expressed at all CBO stages and relative GO for biological processes; (B) Cortex-specific upregulated DEGs increasingly upregulated along CBO differentiation and relative GO for biological processes; (C) Cortex-specific downregulated DEGs increasingly repressed along CBO differentiation and relative GO for biological processes; (D) Cortex-specific downregulated DEGs increasingly upregulated along CBO differentiation and relative GO for biological processes; (E) Cortex-specific downregulated DEGs transiently upregulated along CBO differentiation and relative GO for biological processes.

We characterised the genes of every category by gene ontology. Interestingly, we found that genes increasingly repressed along CBO development were mostly related to extracellular matrix, partially in contrast with previously published results showing overexpression of extracellular matrix genes

in brain organoids compared to the human fetal brain¹²⁰ (**Fig. 38**). A list of these genes can be found in **Table 1**. The other gene clusters were not particularly enriched in meaningful categories (**Fig. 38**).

Extracellular matrix organisation	
ACAN	HAS2
ADAMTS2	KDR
ADAMTS2	LAMC2
ANXA2	MMP2
COL4A1	SERPINE1
COL4A3	SMOC2
COL4A4	SULF1
CREB3L1	TCF15
FGF2	TGFBI

Table 1: Genes involved in extracellular matrix organisation. List of genes belonging to the extracellular matrix organisation category of the GO for biological processes for the cluster of cortex-specific downregulated genes increasingly repressed along CBO differentiation.

Overall, we determined the transcriptional dynamics of CBO differentiation identifying two phases of evolution, one fast-evolving until day 100 and one slow-evolving after day 100. We characterised them highlighting that changes in the fast-evolving phase were mainly related to neuronal differentiation, while the slow-evolving one was defined by upregulation of genes expressed in astrocytes. These results suggested that astrocytes specification started between day 100 and 150 of CBO differentiation, while the neuronal signature was transcriptomically stable at those stages. Moreover, we found that mature CBO globally resembled the transcriptome of human fetal cortex and we characterised the dynamics of cortex-specific modulated genes in organoidogenesis, pointing at extracellular matrix as an important pathway regulated during corticogenesis and defining its dynamics along CBO differentiation.

4.3 CBO reproducibly modulated co-expression patterns of cortical development

To identify the transcriptional networks driving cortical organoidogenesis, we applied WGCNA on our CBO cohort. Using the same strategy applied for the WGCNA on cortical samples from BrainSpan, we used hierarchical gene clustering on the Topological Overlap Dissimilarity Measure to identify networks that were then detected as branches of the dendrogram with the DynamicTree Cut algorithm (Fig. 39).

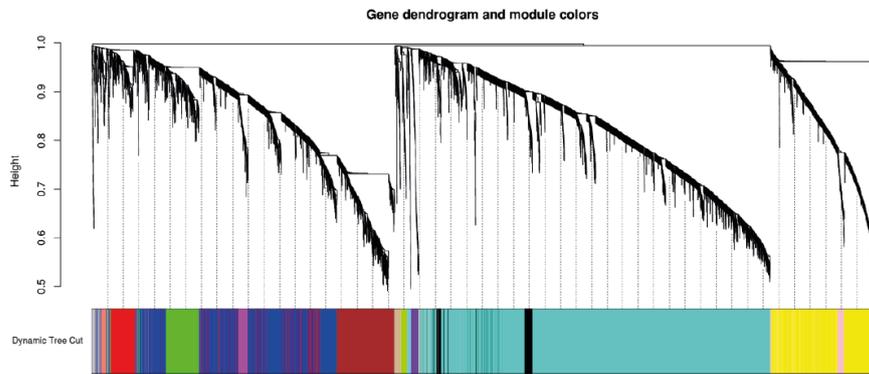


Fig. 39: Gene dendrogram for CBO WGCNA module selection. Gene dendrogram generated using as input for average-linkage hierarchical clustering the dissimilarity matrix relative to the topological overlap measure calculated from the adjacency matrix. Network modules were detected as branches of the dendrogram by using the DynamicTree Cut algorithm (deepSplit=1; minimum cluster size= 30; PAM stage TRUE; cutHeight 0.999). Colours represent WGCNA gene modules.

We found 14 gene modules, ranging in size from 47 to 3279 genes, and summarized the expression profile of each one by its ME to relate it with stage either as a continuous variable or series of categorical variables, thus identifying modules positively or negatively associated with development as well as modules highly correlating with particular time-points (Fig. 40).

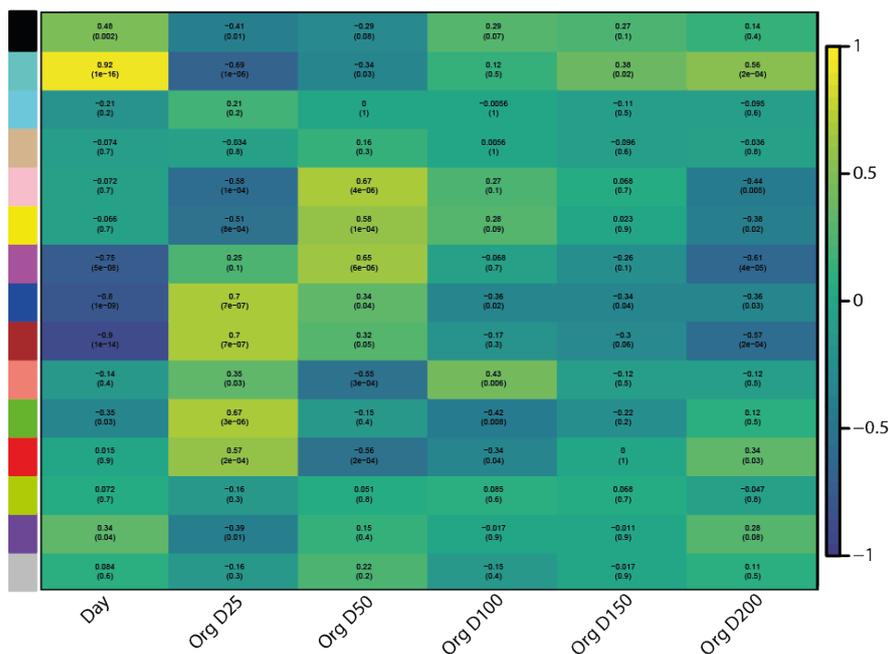


Fig. 40: Correlation of CBO WGCNA modules with stage. Heatmap showing the correlation between ME (represented by the colours on the left) and developmental stage considered as a continuous variable (Day) or as a series of categorical variables (OrgD25-200). Coefficients of correlation were calculated using Spearman correlation and define the heatmap colours. PValue for every correlation, represented in brackets, was calculated and considered significant when < 0.01 .

For interesting modules, we visualized the trend over developmental time of their ME, performed functional analysis of the genes composing the module, and reconstructed weighted networks of the top 75 genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. We then used CytoNCA to identify the most important nodes by computing degree, closeness, betweenness, and eigenvector as centrality indices. Degree, closeness, betweenness and eigenvector were represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

We identified two modules with the highest positive correlation with stage and two with the highest negative correlation with stage. We visualised the ME along time-points for all samples in order to check for consistency among lines and replicates.

The turquoise module contained the highest number of genes that were consistently upregulated along development among lines and replicates (**Fig. 41A**). Turquoise genes with the highest centrality scores were involved in neuronal functions. Indeed, we found again CAMK2B, which was one of the most connected genes in the turquoise module of the BrainSpan WGCNA (**Fig. 15**), related to calcium intracellular signaling regulating dendritic morphology and neuronal homeostasis²⁸⁴. KIF2C was also found in the pink module of the BrainSpan WGCNA and it is related to intracellular trafficking towards dendrites²⁸⁵. Among the most central genes for the turquoise module, there were also: i) AGAP2, a GAP/GEF protein localizing at the post-synaptic density³³⁹; ii) MAPT, encoding for the microtubule associated protein tau implicated in the onset of Alzheimer disease³⁴⁰; iii) MMP24, which regulates axonal outgrowth³⁴¹; iv) SLC4AE, a Cl-/HCO₃⁻ exchanger involved in the regulation of response to GABA and glycine signaling also implicated in epilepsy³⁴²; v) NRXN2, which encodes for neurexin 2 and plays important roles in the excitatory cortical synapse³⁴³; vi) subunits of neurotransmitter receptors such as GABBR1 and GRM5 (**Fig. 41B**).

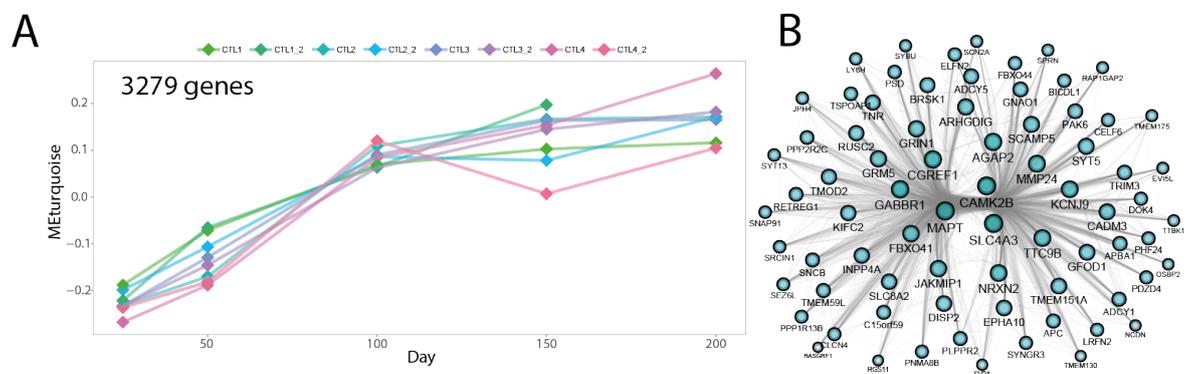


Fig. 41: Characterisation of the CBO turquoise module. (A) Ribbon chart representing the turquoise ME (Y-axis) trend over developmental time (X-axis) for every line and replicate of the CBO dataset. (B) Network reconstruction for the top 75 turquoise genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight.

Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Another module with high positive correlation with stage was the black one (**Fig. 42A**), which was composed by a much smaller number of genes and showed a particularly steep increase in ME between Day50 and Day100. Compared to most modules, the black one showed also a higher degree of variability among lines and replicates, pointing to a small set of genes less robustly induced in CBO. The most central black genes were NEUROD6, a transcription factor with prominent expression during embryonic development in neuronal precursors committed to the pyramidal lineages and demonstrated roles in the regulation of axonal fasciculation³⁴⁴, and SLA, also known as SLAP-1 and selectively expressed in deep layer neurons in the developing cortex³⁴⁵. Other central black genes were ARPP21, which controls dendritic branching³⁴⁶, CELF2, found to be enriched in excitatory neurons (10.1101/244905), and BCL11B, a determinant of deep layer neurons identity³⁴⁷. Interestingly, among the genes with high betweenness and lower values for the other centrality scores measured we found RORB, which specifies layer 4 neurons³⁴⁸. Other black genes with high betweenness were ZBTB18, involved in the regulation of radial migration of new-born neurons within the fetal brain³⁴⁹, and MPPED1, which if absent causes reduced upper layer neuron output³⁵⁰ (**Fig. 42B**).

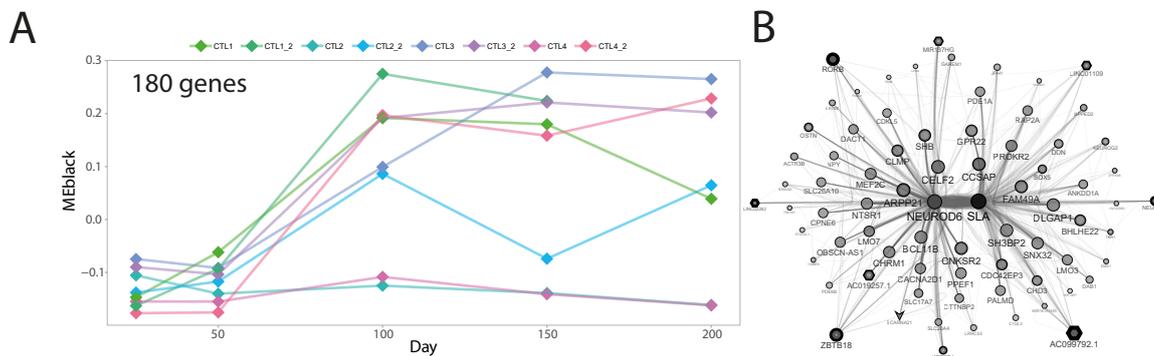


Fig. 42: Characterisation of the CBO black module. (A) Ribbon chart representing the black ME (Y-axis) trend over developmental time (X-axis) for every line and replicate of the CBO dataset. **(B)** Network reconstruction for the top 75 black genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

The blue module instead showed high negative correlation with stage, and indeed it was consistently downregulated in all lines and replicates with a drop until day 100 and then a stabilization of the genes composing it until day 200 (**Fig. 43A**). Central blue genes were involved in diverse general functions, although anyway relevant for human cortical development, such as transcriptional and translational regulation. CDC14A regulates cell migration and adhesion³⁵¹, TSPAN6 was specifically found in neuronal progenitors during the G0 phase of their cell cycle in a preprint study (doi: 10.1101/446344). Interestingly YAP1, a crucial mediator of the HIPPO pathway, was a central blue gene and it has been linked to the regulation of neuronal progenitor proliferation

together with TEAD4³⁵². NUAK2, a target of YAP1, was also found in the blue network and there is evidence that its expression is restricted to the proliferative zone of the developing cortex³⁵³. NPM1, instead, regulates neuronal viability³⁵⁴. Importantly, MRPL45, the gene with the highest betweenness in the blue network, is involved in mitochondrial function regulation and it has been reported to be a downstream target of ERR γ in regulating long-term potentiation, learning and memory in hippocampal neurons³⁵⁵ (**Fig. 43B**).

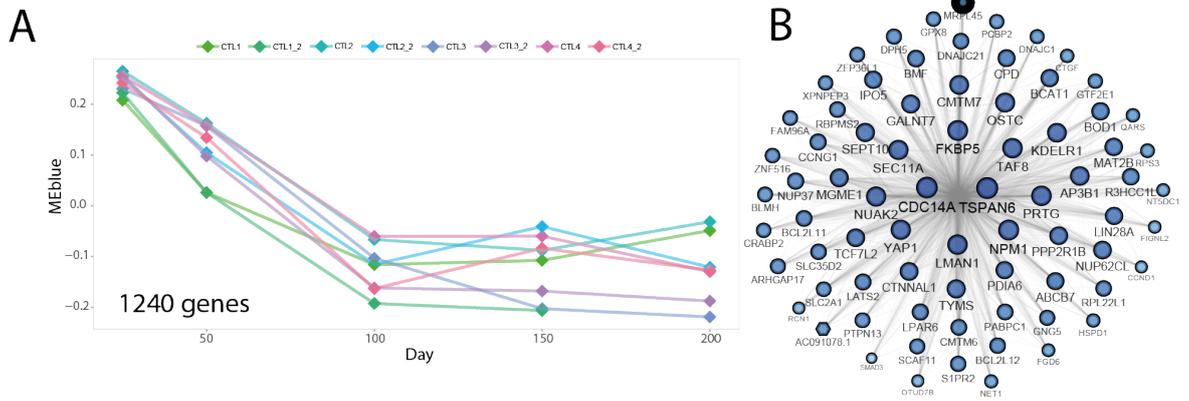


Fig. 43: Characterisation of the CBO blue module. (A) Ribbon chart representing the blue ME (Y-axis) trend over developmental time (X-axis) for every line and replicate of the CBO dataset. (B) Network reconstruction for the top 75 blue genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

The other module with high negative correlation along CBO development was the brown one (**Fig. 44A**). Brown central genes were clearly related to cell cycle. Example of brown genes were the cyclins CCNB1 and B2, AURKA and B, also found in the yellow module in the BrainSpan WGCNA, and DLGAP5, a target of AURKA signaling³⁵⁶. Moreover, the brown module included genes important for the control of mitosis, such as KIF18A, KIF4A, MELK and DEPDC1B^{357,356,358,359}. Genes required for the control of DNA repair during the cell cycle, such as FANCI and FANCD2, were also central in this module³⁶⁰ (**Fig. 44B**).

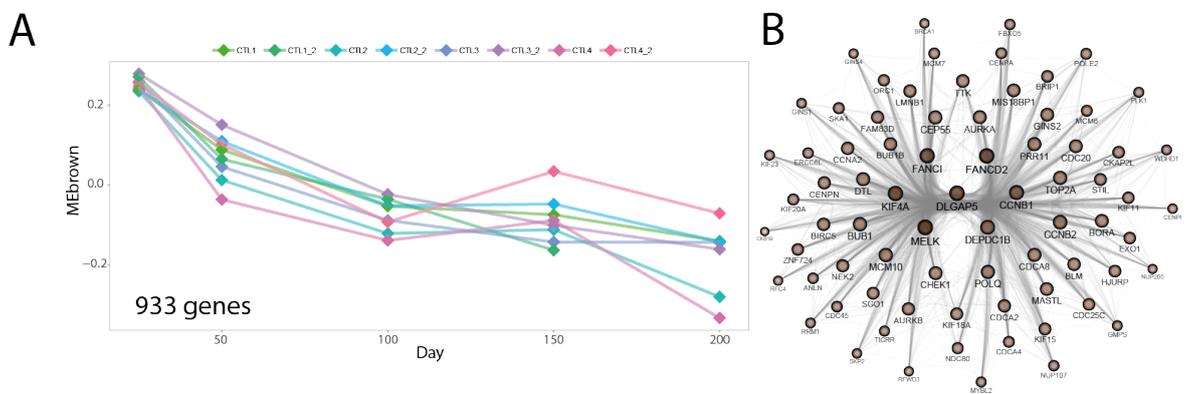


Fig. 44: Characterisation of the CBO brown module. (A) Ribbon chart representing the brown ME (Y-axis) trend over developmental time (X-axis) for every line and replicate of the CBO dataset. (B) Network reconstruction for the top 75 brown genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and

represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

We then characterised through gene ontology the most enriched categories for the genes of these 4 modules confirming that: i) the turquoise and black ones were mostly related to cortical neuron differentiation; ii) the brown module was enriched in cell cycle genes; and iii) the blue module was instead composed of genes generally important for neuronal progenitor and neuronal functions (Fig. 45).

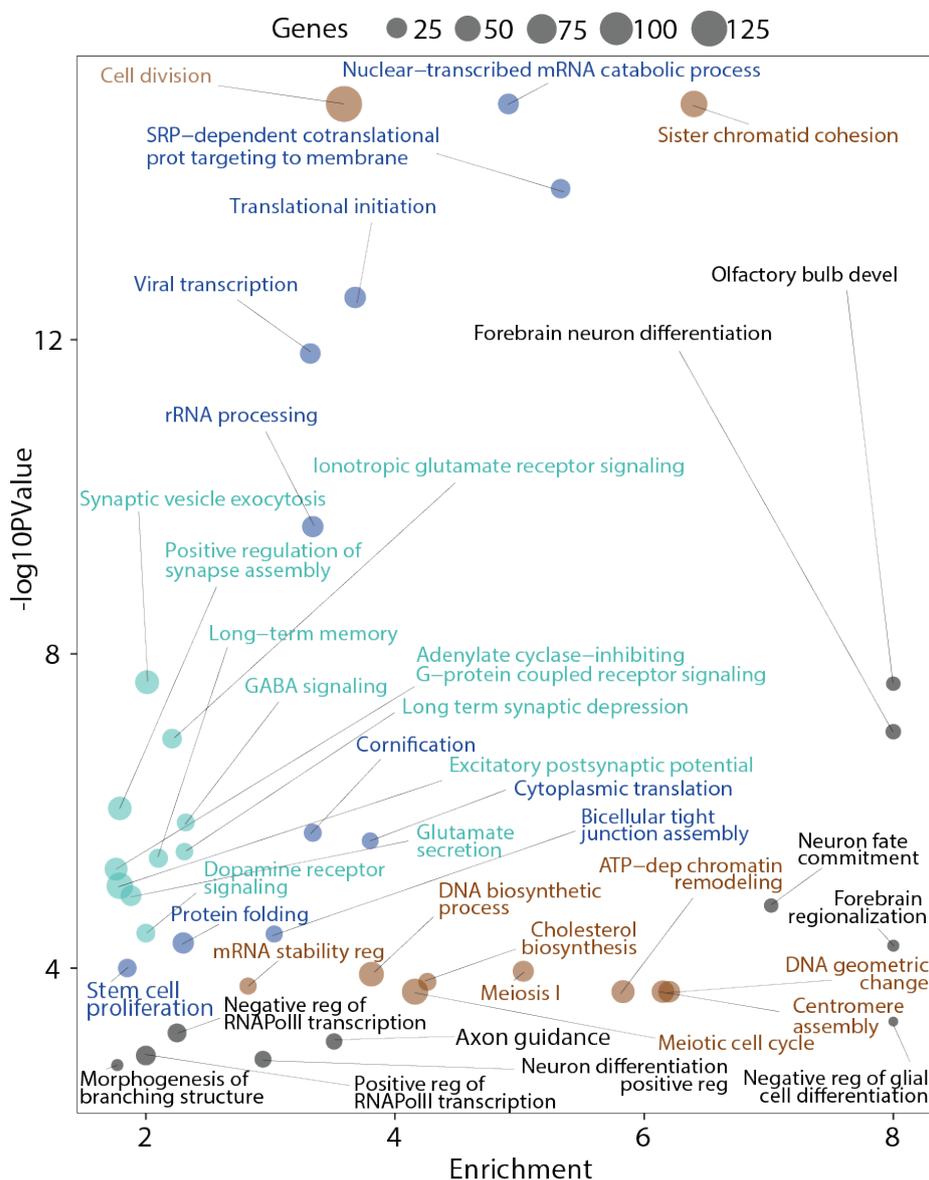


Fig. 45: Functional analysis of main CBO WGCNA modules. Bubble plot summarising the top 10 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of genes enriched for the visualised category. The enrichment, defined as the ratio between the number of genes belonging to a given category and the number of genes expected to be found by chance for that category, is represented on the X-axis. Dot colours represent the different modules.

WGCNA on CBO also revealed modules that were mostly correlated to specific stages rather than showing high positive or negative correlation with CBO development as a continuous variable, thus resulting in particular behaviour of these genes during differentiation. For instance, the green and

red modules had a peak of positive correlation with day 25 that dropped at subsequent stages to then increase again at later time-points, thus resulting in a U-shaped dynamic of expression along CBO development. Importantly, the expression of genes belonging to these modules was less consistent among cell lines and replicates at late stage of CBO development (**Fig. 46A and 47A**).

When looking at the function of central genes for the green and red module, some of them had reported function during neurodevelopment. For instance, the green module gene *ESYT1* promotes neurotransmission and synaptic growth³⁶¹, *IL4R* is involved in neural stem cell plasticity³⁶², and *CYFIP1* regulates brain connectivity³⁶³. Examples of central red module genes involved in neuronal functions were *TNFRSF1A*, which participate in the regulation of neuronal differentiation and regional pattern development³⁶⁴, and *LAMB2*, which regulates cortical lamination³⁶⁵ (**Fig. 46B and 47B**).

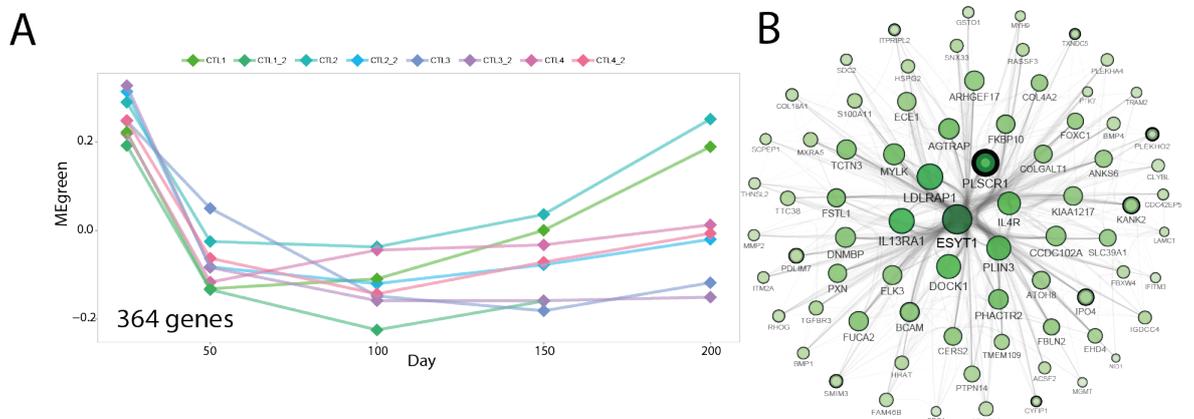


Fig. 46: Characterisation of the CBO green module. (A) Ribbon chart representing the green ME (Y-axis) trend over developmental time (X-axis) for every line and replicate of the CBO dataset. (B) Network reconstruction for the top 75 green genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

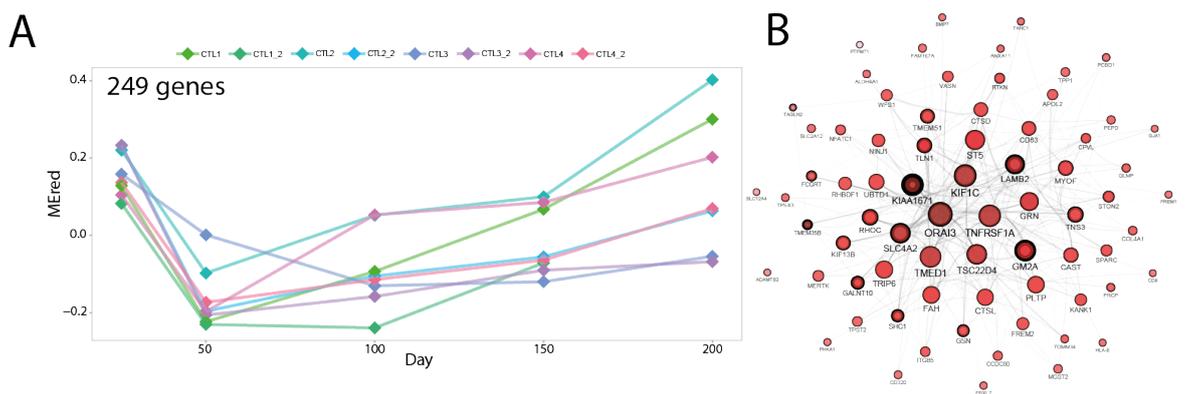


Fig. 47: Characterisation of the CBO red module. (A) Ribbon chart representing the red ME (Y-axis) trend over developmental time (X-axis) for every line and replicate of the CBO dataset. (B) Network reconstruction for the top 75 red genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

Interestingly, when looking at the function of genes belonging to the green and red modules using gene ontology we found enrichment for categories related to extracellular matrix and cell adhesion. The behaviour of those genes was in line with what we observed for cortex-specific DEGs that were increasingly repressed along CBO-development (**Fig. 48**), adding a further layer of information regarding the expression and reproducibility dynamics of extracellular-matrix-related genes in CBO over time.

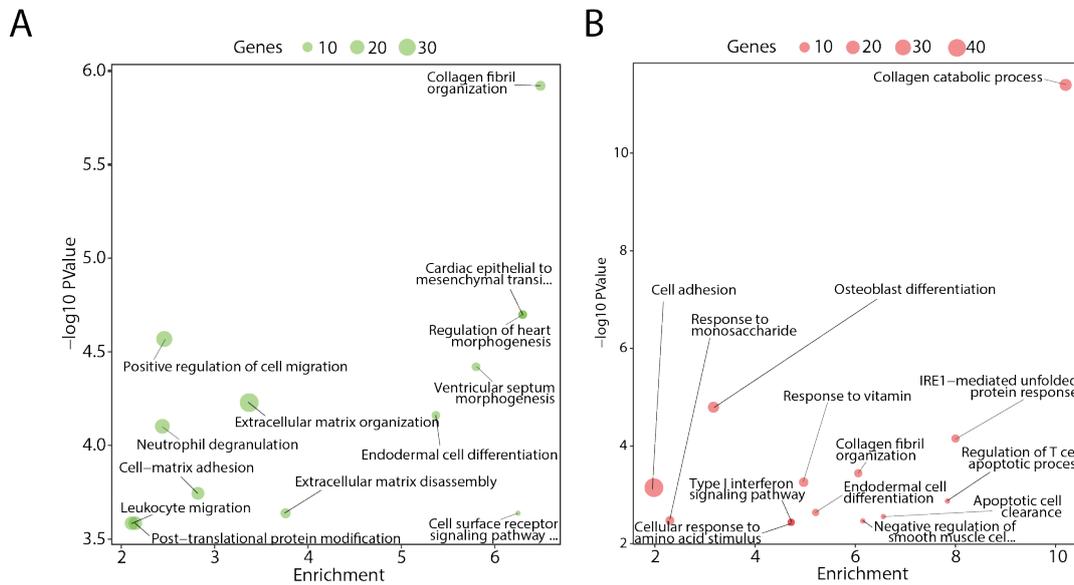


Fig. 48: Functional analysis of CBO green and red modules. Bubble plot summarising the top 13 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis) for the green (**A**) and red (**B**) CBO WGCNA modules. The size of the dots represents the number of genes enriched for the visualised category. The enrichment, defined as the ratio between the number of green or red genes belonging to a given category and the number of green or red genes expected to be found by chance for that category, is represented on the X-axis.

Another interesting module with stage-specific behaviour, with a peak of positive correlation at day 50, was the magenta one. Genes of these module were consistently downregulated after day 50 and were related to transcriptional and epigenetic regulation, among other cellular processes (**Fig. 49**).

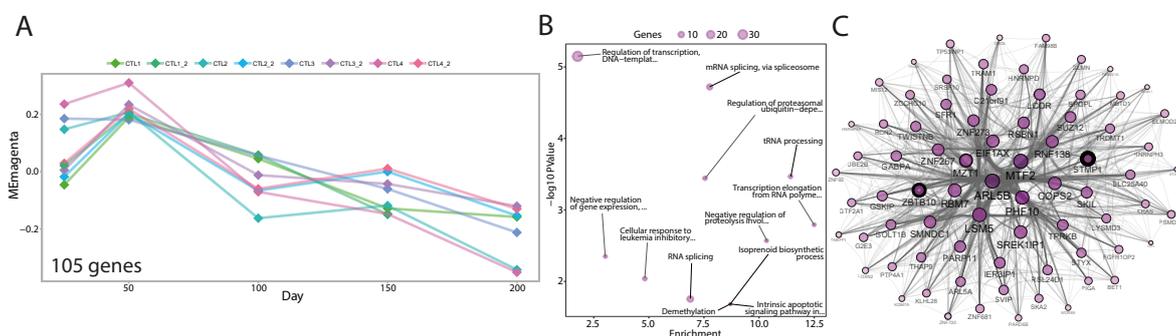


Fig. 49: Characterisation of the CBO magenta module. (**A**) Ribbon chart representing the magenta ME gene (Y-axis) trend over developmental time (X-axis) for every line and replicate of the CBO dataset. (**B**) Bubble plot displaying the top 12 GO categories for biological processes with $-\log_{10}P$ Value greater than 2 (P Value < 0.01) (Y-axis). The size of the dots represents the number of red genes enriched for the visualised category. The enrichment, defined as the ratio between the number of magenta genes belonging to a given category and the number of magenta genes expected to be found by chance for that category, is represented on the X-axis. (**C**) Network reconstruction for the top 75 magenta genes according to their intramodular connectivity and with a threshold of 0.2 as minimum edge weight. Degree centrality, closeness

centrality, betweenness centrality and eigenvector centrality were computed using CytoNCA and represented by node label transparency, node colour darkness, node border width, and node size/ node label font size, respectively.

In conclusion, we were able to time-resolve the transcriptional evolution of CBO by identifying key co-expression patterns driving their maturation and hubs of cortical organoidogenesis. Overall, we found consistency among lines deriving from independent individuals and among technical replicates for the majority of transcriptional changes along development. Finally, we highlighted genes with a stage-wise peculiar behaviour and we related them mainly to extracellular matrix regulation.

4.4 Benchmarking of brain organoids against pre-natal human fetal cortical development revealed heterochronicity of differentiation across protocols

To evaluate the fidelity of transcriptional programmes in developing human brain organoids, we benchmarked our internal CBO dataset, together with other three external transcriptional datasets of brain organoids generated using different methods, against cortical samples from the BrainSpan atlas. We chose the organoid paradigms to analyse based on their degree of patterning, in order to have a full-scale overview of cortical organoidogenesis in different culture conditions. Indeed, we selected: i) the unpatterned cerebral organoid (CO) dataset from Luo et al¹²⁰; ii) the dataset of forebrain organoids (FO) from Qian et al¹⁰⁶, which develop cell types specific of dorsal and ventral telencephalon; iii) the telencephalic aggregates (TA) dataset from Mariani et al¹⁰⁵, with organoids that are generated upon a first step of 2D differentiation and have similar patterning of forebrain organoids; and iv) our internal dataset of CBO specifically patterned towards dorsal telencephalon. A summary of the developmental time points for every dataset is outlined in **Fig. 50**.

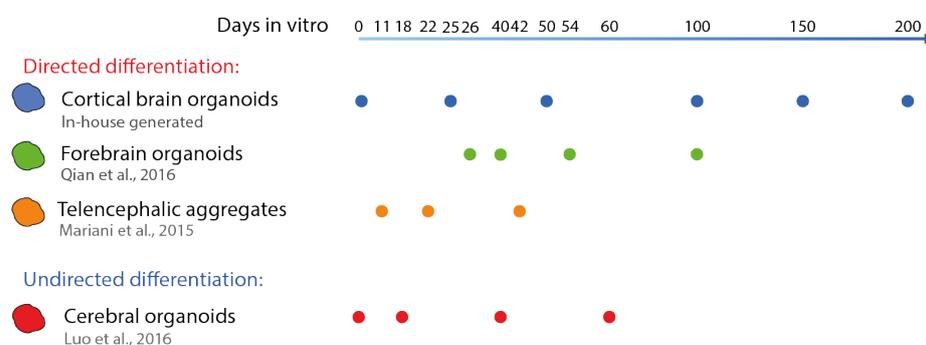


Fig. 50: Schematic representation of the brain organoid datasets analysed in this work. Summary of the developmental time-points for which bulk RNAseq was available for the brain organoid datasets analysed: i) CBO, differentiated and profiled in-house; ii) FO, from Qian et al; iii) TA from Mariani et al; iv) CO, from Luo et al.

Besides CBO, the CO dataset was the only including the pluripotent stage, which showed a clear separation from all organoid samples in PCA. CO were clustering by stage, with day 40 and 60 samples globally more similar within each other and different from day 18. This dataset included CO from one embryonic stem cell line (**Fig. 51**).

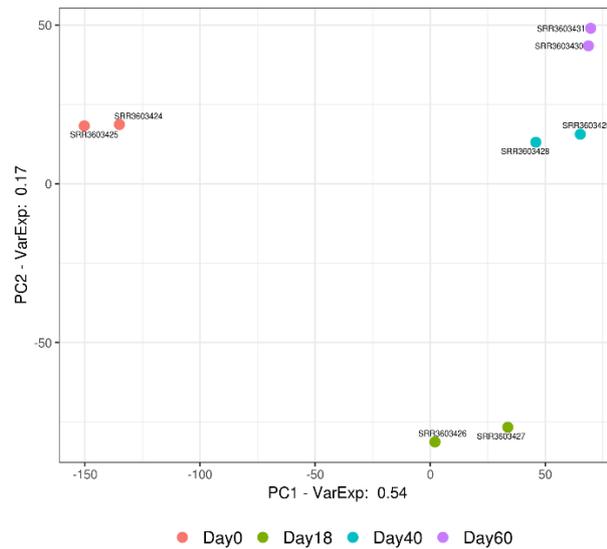


Fig. 51. Dimensionality reduction of the CO dataset. PCA showing PC1 and PC2 for the CO dataset. Every dot represented a samples. Colour-code is depicted the figure legend.

Also for the FO dataset, organoids were clustering by stage, with a peculiar behaviour of day 100 organoids. We attributed such big difference between day 100 FO and the rest of the dataset to the fact that single-end sequencing was performed only for this time-point, while all other samples were sequenced in paired-end. For this reason, we removed this stage for all subsequent analyses for which we could not differentiate trustworthy results from batch effect. (Fig. 52). Also in this case, only organoids from a single hiPSC lines were included.

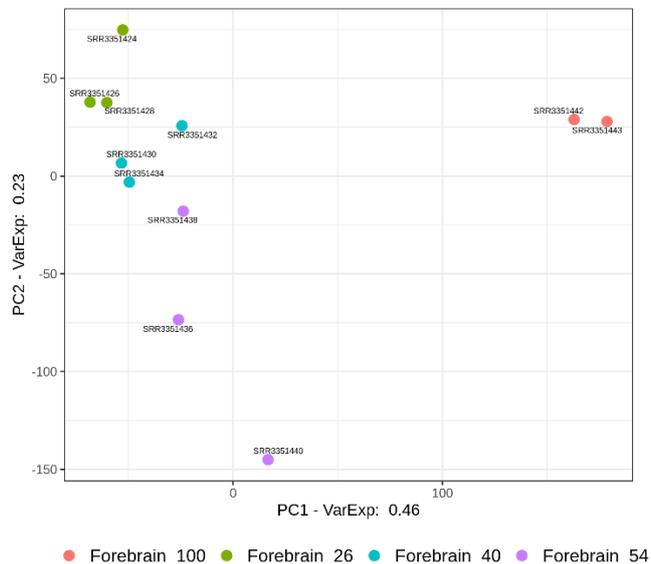


Fig. 52: Dimensionality reduction of the FO dataset. PCA showing PC1 and PC2 for the FO dataset. Every dot represented a samples. Colour-code is depicted the figure legend.

The TA dataset included a total of 10 lines deriving from different control individuals. In this case, not all individuals were sequenced at the three time-points analysed and for several individuals two differentiation batches were performed. Organoids clustered by developmental stage (Fig. 53).

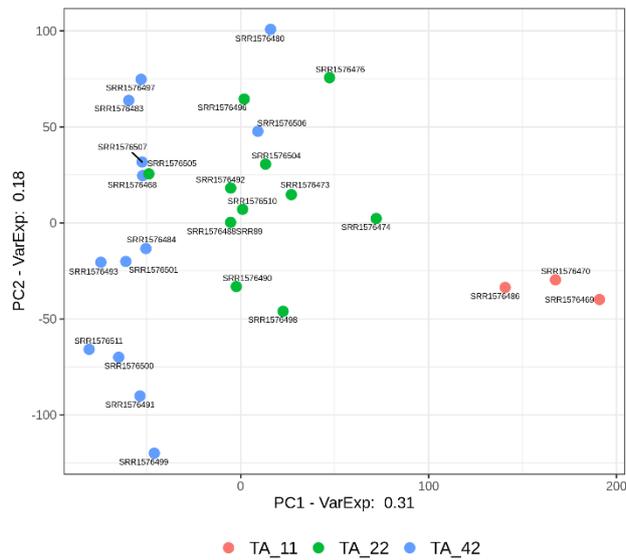


Fig. 53: Dimensionality of the TA dataset. PCA showing PC1 and PC2 for the TA dataset. Every dot represented a samples. Colour-code is depicted the figure legend.

We then unravelled the differentiation dynamics of the organoid systems analysed by using the same bulk transcriptome deconvolution approach used for the CBO dataset, this time considering only two broad populations: neural progenitors, including cycling progenitors and ventricular radial glia, and neurons, including excitatory and inhibitory neurons.

CBO showed a preponderance of progenitor cells at the first day 25, similar levels of the two populations at day 50, and stably high proportions of neurons from day 100 onwards. CO, FO and TA showed similar dynamics of decrease of neuronal progenitors and increase in neurons, although at a similar *in vitro* age of around 20 days all these three organoid systems showed much more similar proportions of the two populations than CBO. This result indicated heterochronicity of differentiation between different systems, with CO, FO and TA showing a faster progression of neuronal progenitors towards neurons (**Fig. 54-57**).

Importantly, the only datasets including multiple lines from different individuals were the CBO and the TA ones. Overall, CBO displayed high reproducibility across lines and differentiation batches, with day 50 representing the most variable stage, while for TA reproducibility was poorly preserved at day 22 and 44 (**Fig. 54-57**).

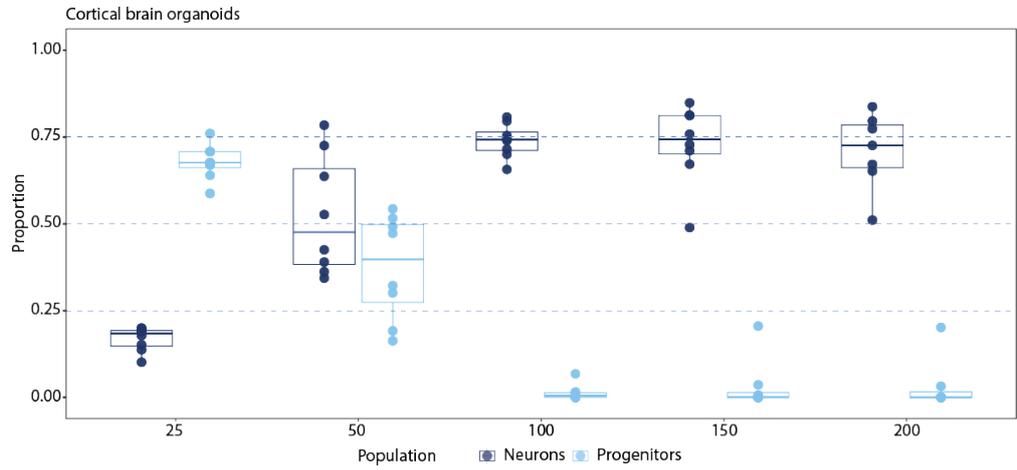


Fig. 54: Bulk deconvolution of the CBO dataset. Box plot showing the estimated proportion of progenitors and neurons (Y-axis) at each stage of CBO differentiation (X-axis) and for every CBO sample based on the results given by the SCDC deconvolution algorithm.

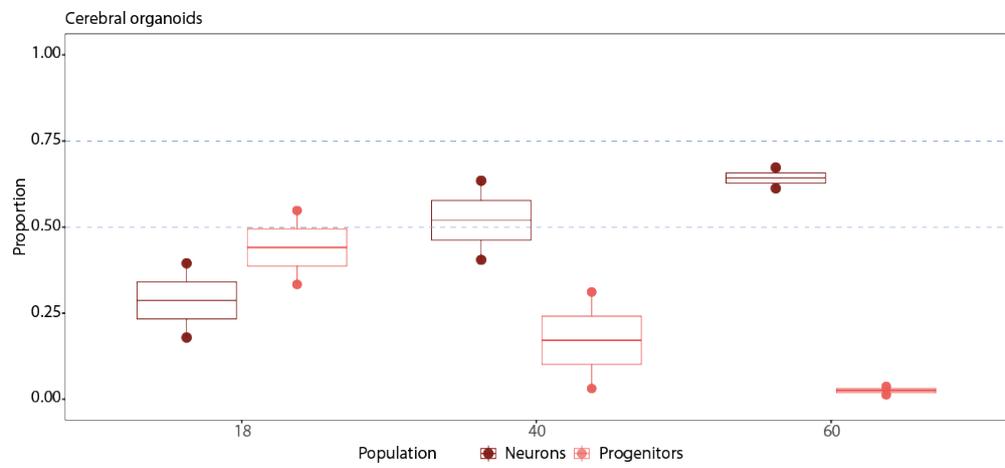


Fig. 55: Bulk deconvolution of the CO dataset. Box plot showing the estimated proportion of progenitors and neurons (Y-axis) at each stage of CO differentiation (X-axis) and for every CO sample based on the results given by the SCDC deconvolution algorithm.

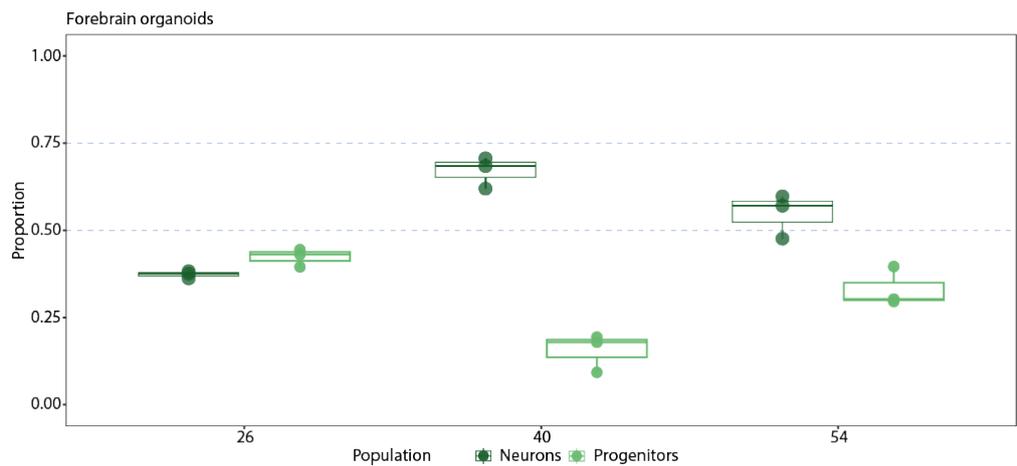


Fig. 56: Bulk deconvolution of the FO dataset. Box plot showing the estimated proportion of progenitors and neurons (Y-axis) at each stage of FO differentiation (X-axis) and for every FO sample based on the results given by the SCDC deconvolution algorithm.

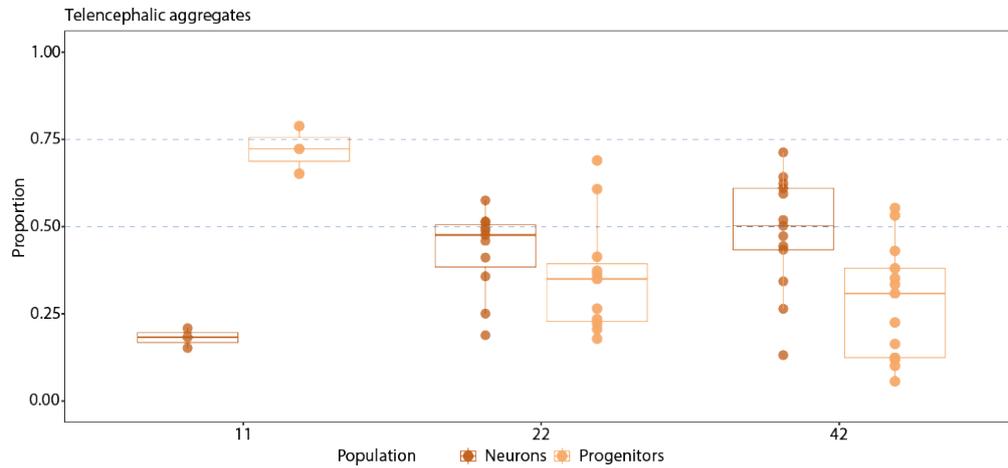


Fig. 57: Bulk deconvolution of the TA dataset. Box plot showing the estimated proportion of progenitors and neurons (Y-axis) at each stage of TA differentiation (X-axis) and for every TA sample based on the results given by the SCDC deconvolution algorithm.

We thus performed DEA for each external dataset using the same approach applied for CBO, namely performing pair-wise comparisons of each stage against the previous one (**Fig. 58**). Similarly to CBO, this analysis showed for all protocols a reduction of the number of DEGs over time, although a comparable DEGs contraction observed for day 150 CBO (**Fig. 28**) was reached much earlier for all external datasets (around 40-60 days). This result corroborated the hypothesis of heterochronicity of differentiation, with CBO reaching a plateau in transcriptional dynamics later than CO, FO and TA.

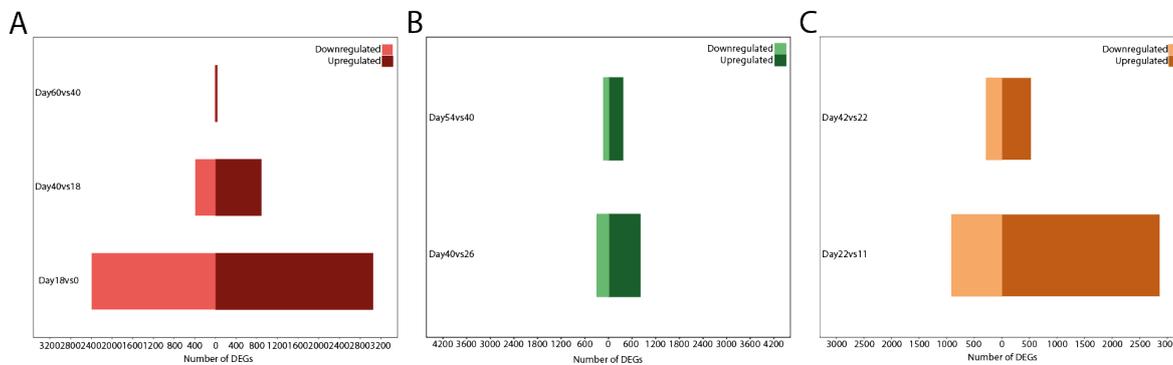


Fig. 58: Differential expression analysis of CO, FO and TA datasets. (A) Bar plot showing the number of upregulated and downregulated DEGs with absolute FC > 2 and FDR < 0.05 of pair-wise DEAs for the CO dataset considering each stage against the previous one in development. The X-axis indicated the number of DEGs, while the Y-axis showed the comparison. (B) Bar plot showing the number of upregulated and downregulated DEGs with absolute FC > 2 and FDR < 0.05 of pair-wise DEAs for the FO dataset considering each stage against the previous one in development. The X-axis indicated the number of DEGs, while the Y-axis showed the comparison. (C) Bar plot showing the number of upregulated and downregulated DEGs with absolute FC > 2 and FDR < 0.05 of pair-wise DEAs for the TA dataset considering each stage against the previous one in development. The X-axis indicated the number of DEGs, while the Y-axis showed the comparison.

To confirm this result, we computed Spearman correlation for CO, FO and TA against CBO. We observed that CO and FO from day 40 onwards highly correlated with > 100 day CBO, again supporting the hypothesis of their global faster differentiation. Interestingly, for TA we observed high inter-individual variability especially at day 22 and 42, confirming the bulk deconvolution results (**Fig. 59**).

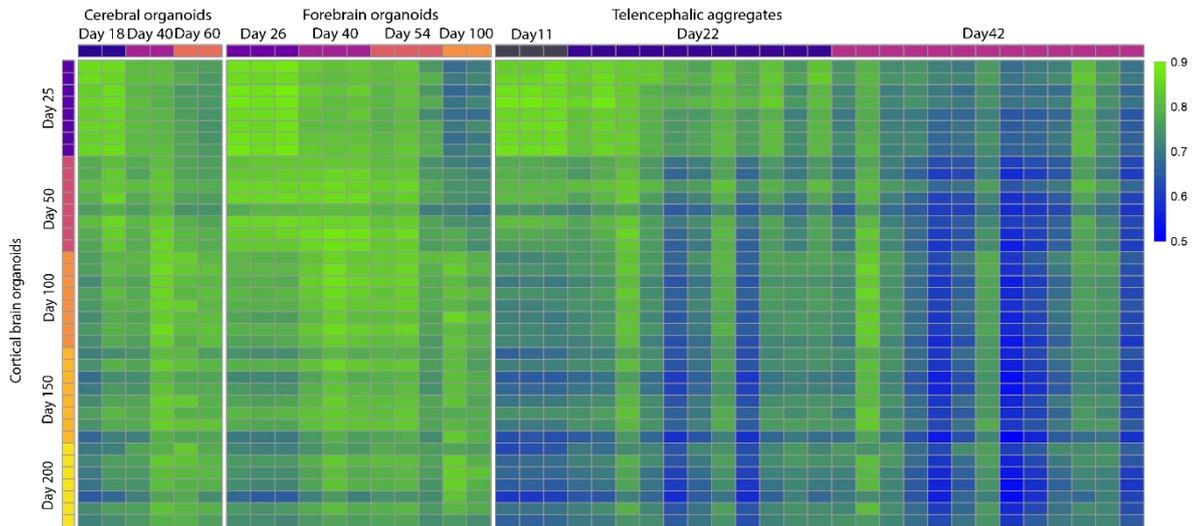


Fig. 59: Correlation analysis of CBO against the other brain organoid datasets analysed. Heatmap showing the result of a Spearman correlation analysis of global transcriptomes (protein-coding genes) considering CO versus CBO, FO versus CBO, and TA versus CBO. Rows and columns represented CBO and CO/FO/TA single samples, respectively. Colour-scale represented the Spearman correlation coefficient.

We thus wondered how these transcriptional differences among brain organoid paradigms related to the actual human cortical development. To answer this question, we correlated CBO, CO, FO and TA to pre-natal cortical samples of the BrainSpan atlas (**Fig. 60-63**).

For CBO, we observed a gradual increase in correlation over time with cortical samples from late PCW (**Fig. 60**). Interestingly, we again observed a fast-evolution phase in CBO until day 100 followed by a slow-evolution one from day 100 to 200. Moreover, this analysis showed high reproducibility between lines and replicates, as a confirmation of what we already observed for WGCNA modules. Another interesting finding for this analysis was that slow-evolving stages correlated similarly well with late and early PCW, possibly due to the presence of a progenitor signature (**Fig. 34**) in CBO that could derive from developing oRG. This result was partially in line also with the finding that PCW12-24 globally highly correlate with each other (**Fig. 10A**).

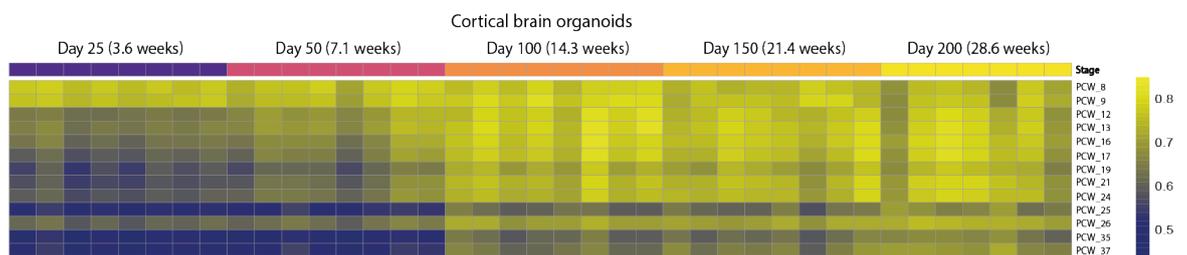


Fig. 60: Correlation analysis of CBO against the BrainSpan pre-natal cortical samples. Heatmap showing the result of a Spearman correlation analysis of global transcriptomes (protein-coding genes) of CBO versus BrainSpan pre-natal cortical samples. Rows and columns represented the mean expression level of all BrainSpan cortical samples at a given PCW and CBO single samples, respectively. Colour-scale represented the Spearman correlation coefficient.

The same analysis performed for CO showed a more compressed evolution, with day 40-60 time-points correlating with PCW for which similar levels of correlation for CBO were reached only at day

100. Also in this case, we observed high correlation with all stages of fetal cortical development for the latest CO time-point (Fig. 61).

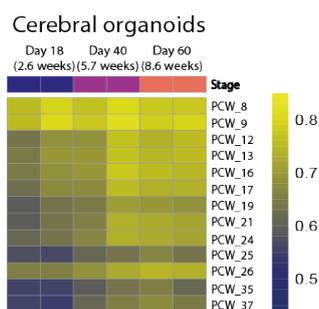


Fig. 61: Correlation analysis of CO against the BrainSpan pre-natal cortical samples. Heatmap showing the result of a Spearman correlation analysis of global transcriptomes (protein-coding genes) of CO versus BrainSpan pre-natal cortical samples. Rows and columns represented the mean expression level of all BrainSpan cortical samples at a given PCW and CO single samples, respectively. Colour-scale represented the Spearman correlation coefficient.

FO instead showed a slightly more compressed development than CBO, with day 40-54 correlating higher with later PCW than CBO at similar developmental age, although in this case this difference was less clear-cut. Interestingly, day 100 FO showed extremely high correlation with all fetal cortex samples compared to all other protocols and developmental time-points (Fig. 62).

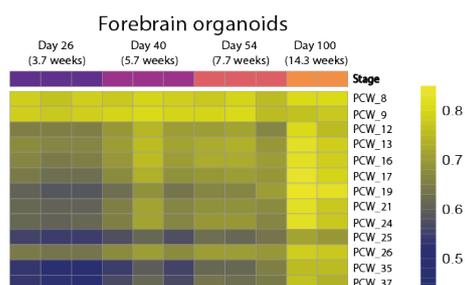


Fig. 62: Correlation analysis of FO against the BrainSpan pre-natal cortical samples. Heatmap showing the result of a Spearman correlation analysis of global transcriptomes (protein-coding genes) of FO versus BrainSpan pre-natal cortical samples. Rows and columns represented the mean expression level of all BrainSpan cortical samples at a given PCW and FO single samples, respectively. Colour-scale represented the Spearman correlation coefficient.

TA differentiation showed to be less tolerant to interindividual variability than CBO, with day 11 organoids homogeneously correlating with the early PCWs and day 22 and 42 being less more variable, supporting the bulk deconvolution analysis (Fig. 55). Day 22 and 44 samples that appeared to follow a correct cortical differentiation were showing high correlation with all PCWs (Fig. 63).

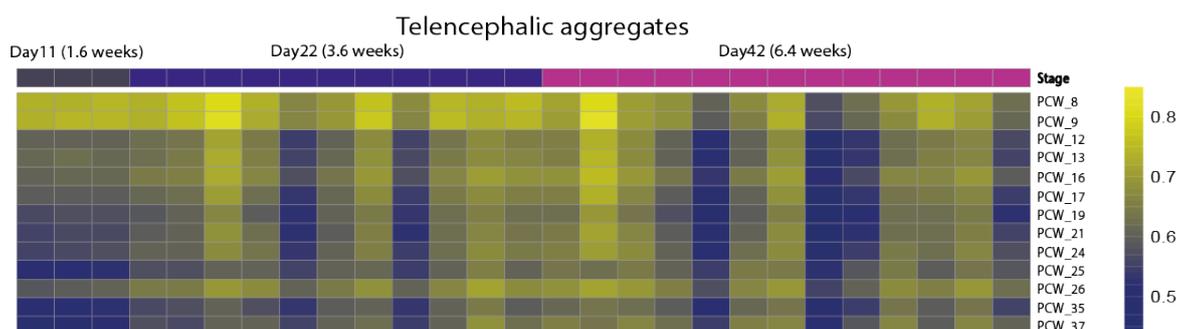


Fig. 63: Correlation analysis of TA against the BrainSpan pre-natal cortical samples. Heatmap showing the result of a Spearman correlation analysis of global transcriptomes (protein-coding genes) of TA versus BrainSpan pre-natal cortical samples. Rows and columns represented the mean expression level of all BrainSpan cortical samples at a given PCW and TA single samples, respectively. Colour-scale represented the Spearman correlation coefficient.

Overall, we demonstrated that CBO showed a rise in correlation with increasing fetal-cortex age along organoidogenesis, confirming the two-step evolution observed with DEA. We also interrogated external organoid datasets with different degree of patterning and peculiar culture conditions to understand whether the transcriptional evolution of organoidogenesis was similar within distinct paradigms. We found heterochronicity in the transcriptomic recapitulation of cortical development, with CO, FO and TA showing a more compressed differentiation than CBO. Indeed, CBO appeared to better recapitulate the timing of transcriptomic evolution in cortical development.

4.5 Gene signatures specific for brain populations and functions unveiled dynamics of brain organoid differentiation

Upon characterisation of global differences between organoid paradigms in terms of general populations and transcriptome dynamics, we sought to investigate how brain organoids regulate different sets of genes that we used as hallmarks of specific brain populations and functions, with a particular focus on cortex. We curated signatures from literature, taking also advantage of previously published studies summarising relevant genes for populations and functions in the cortex^{122,118}.

We first looked at the expression of cell populations markers in cortical samples from Brainspan and in brain organoids dividing genes as following: i) relevant cell cycle genes (*cycling*); ii) neural stem cell genes (*NSC*); iii) aRG genes (*ApicalProg*); iv) IP genes (*IntermediateProg*); v) *oRG* genes; vi) general markers of post-mitotic neurons (*PanNeuron*); vii) genes specific for excitatory glutamatergic neurons (*ExcN*); viii) genes specific for GABAergic inhibitory neurons (*InhN*); ix) markers of lower layer neurons (*LowLayerN*); x) markers of upper layer neurons (*UpLayerN*); xi) genes specifically found in the post-synaptic density (*PostSynaptic*); xii) astrocytic markers (*Astrocyte*).

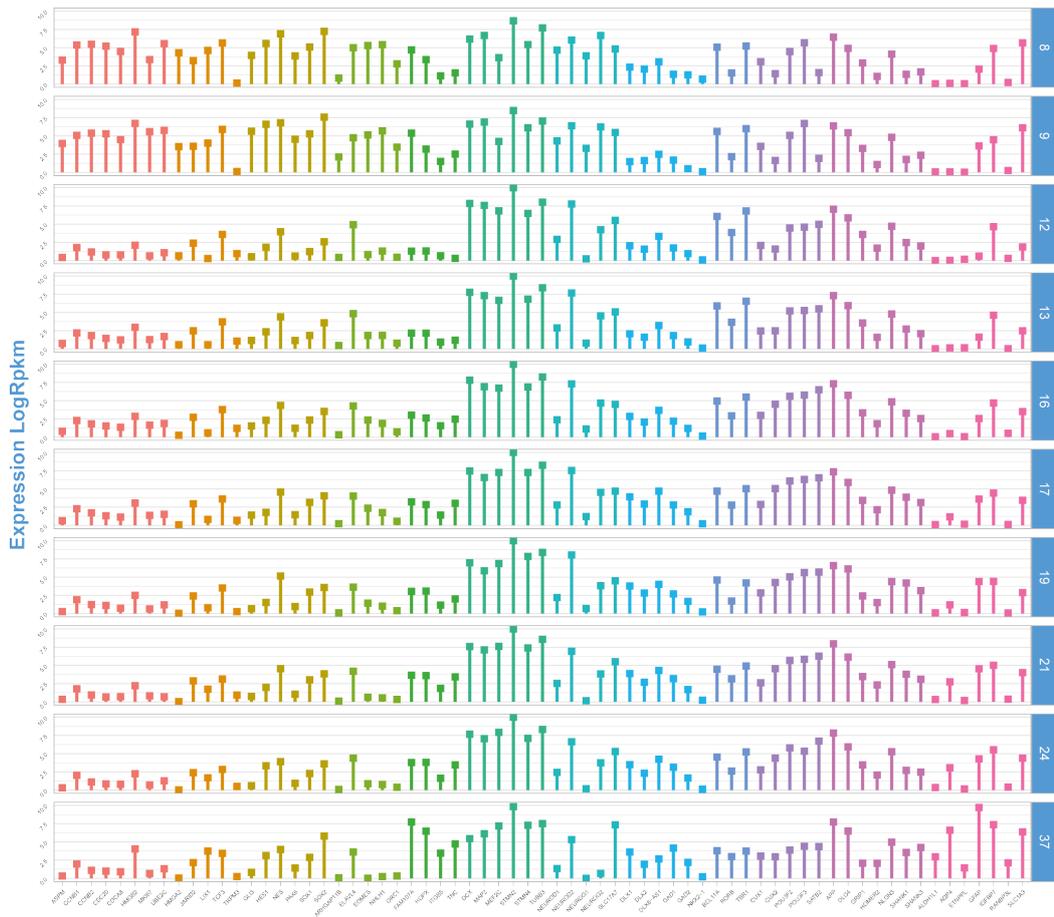
We first looked at the expression of markers for these classes in cortical samples from BrainSpan. The first noticeable feature was that important genes for the cell cycle drastically reduced between PCW9 and 12, in line with our WGCNA results. The same feature applied to markers of neural stem cells, apical progenitors and intermediate progenitors. Comparing the *cycling* signature across brain organoid protocols, we observed that its reduction was less abrupt. The same observation was made for the *ApicalProg* signature, which in organoids never reduced as much as in cortical samples, possibly explaining the correlation of organoids with early post-conceptual weeks. Regarding the *IntermediateProg* signature, in organoids it did not reach the same stability of expression observed in BrainSpan at PCW8 and 9, although CO at day 18 and FO at day 26 showed consistent upregulation of the complete signature, which was present at all subsequent time points less coherently. Interestingly, CBO showed upregulation of the signature especially at day 100, with EOMES, a known determinant of intermediate progenitors, only present at this time-point. TA showed the least expression of the intermediate progenitor signature. The *oRG* signature was well expressed in BrainSpan at PCW8 and 9 to then decrease at PCW12 and increase again until PCW37. CBO showed constant upregulation of the whole signature starting from day 100, while for FO this was observed already at day 40. CO showed its partial upregulation, although HOPX, the most widely recognised specific marker of *oRG*, was never detected. The same pattern was found for TA, where also FAM107A was not detected at any stage.

The *PanNeuro* signature was constantly upregulated in BrainSpan, with higher expression at PCW12-37 than PCW8-9. In organoids, this signature was constantly upregulated in all protocols analysed. Interestingly, the *ExcN* was particularly expressed at PCW8-9, while it was more variable at later stages. CBO largely resembled this variability at later stages, while showing global upregulation of this group of genes at day 25 and 50. The other organoid paradigms showed a constant upregulation of the signature through time, with less variability. Inhibitory neuron genes were constantly increasing in BrainSpan while showing different patterns in organoids. CO and TA showed very mild upregulation of the signature at all stages, FO stably expressed it until day 40 with a drop at later time points, and CBO showed increase in these genes from day 100 onwards. This behaviour resulted particularly interesting because the patterning towards dorsal telencephalon of CBO should not lead to the production of inhibitory neurons, although this result was in line with what we observed by DEA and comparable protocols showed similar results³³⁵. Regarding the layer identity of neurons, lower and upper layer neuron makers were upregulated along development in BrainSpan. CBO showed upregulation of the *LowLayerN* signature starting from day 100, while the *UpLayerN* signature was constantly expressed, although the expression of the upper layer neuron determinant SATB2 was not observed until day 100. CO started to upregulate the *LowLayerN* signature and SATB2 at day 40, while FO stably expressed lower layer neuron markers at day 40 and did not express SATB2 at any stage. TA did not express the *LowLayerN* signature at any stage, while very low levels of SATB2 were measured at day 11 and 22, but not at day 42. The *PostSynaptic* signature was stably expressed at all stages in BrainSpan and organoids.

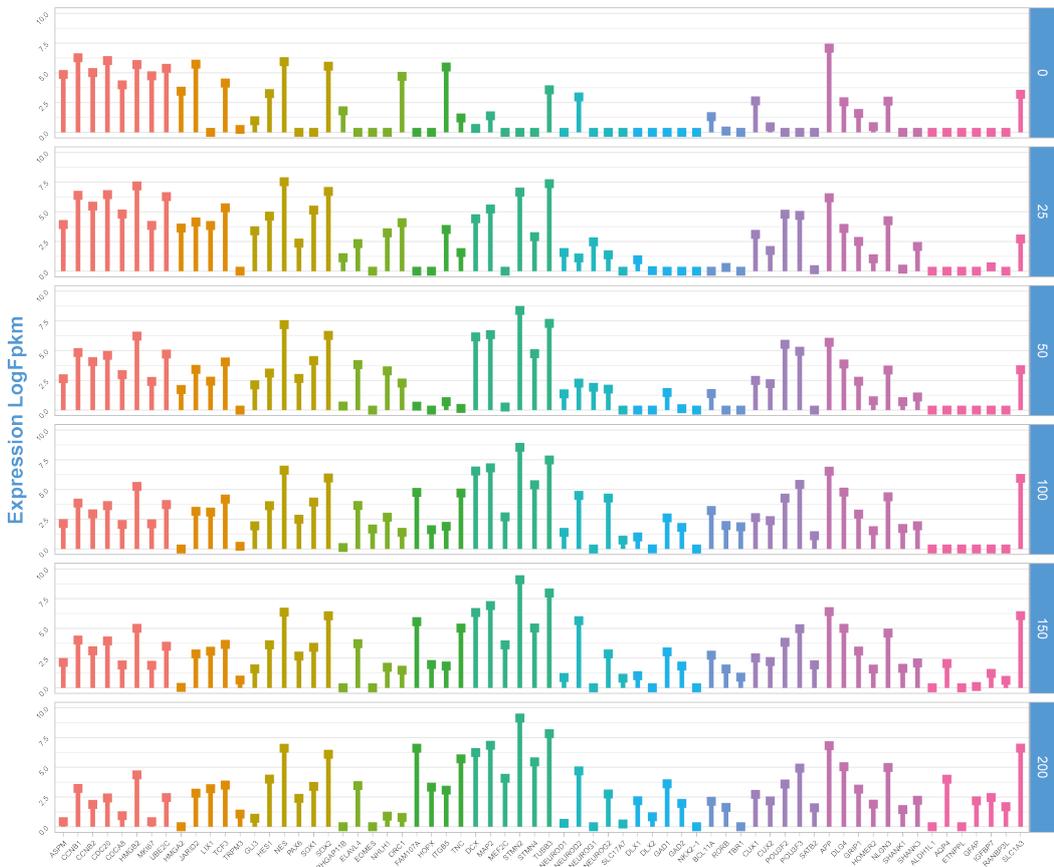
Regarding the *Astrocyte* signature, it was stably expressed in BrainSpan at PCW37, although it was expressed also earlier but less stably. Importantly AQP4, a water channel strictly specific of astrocytes³⁶⁶, started to be expressed in BrainSpan at PCW17. In CBO, AQP4 was expressed at day 150 and the signature was more stably upregulated at day 200, in line with our DEA results. The other organoid paradigms did not stably express this signature at any stage, probably due to the time-points analysed.

The levels of expression for all described signatures in BrainSpan and in all organoid systems are reported in **Fig. 64**.

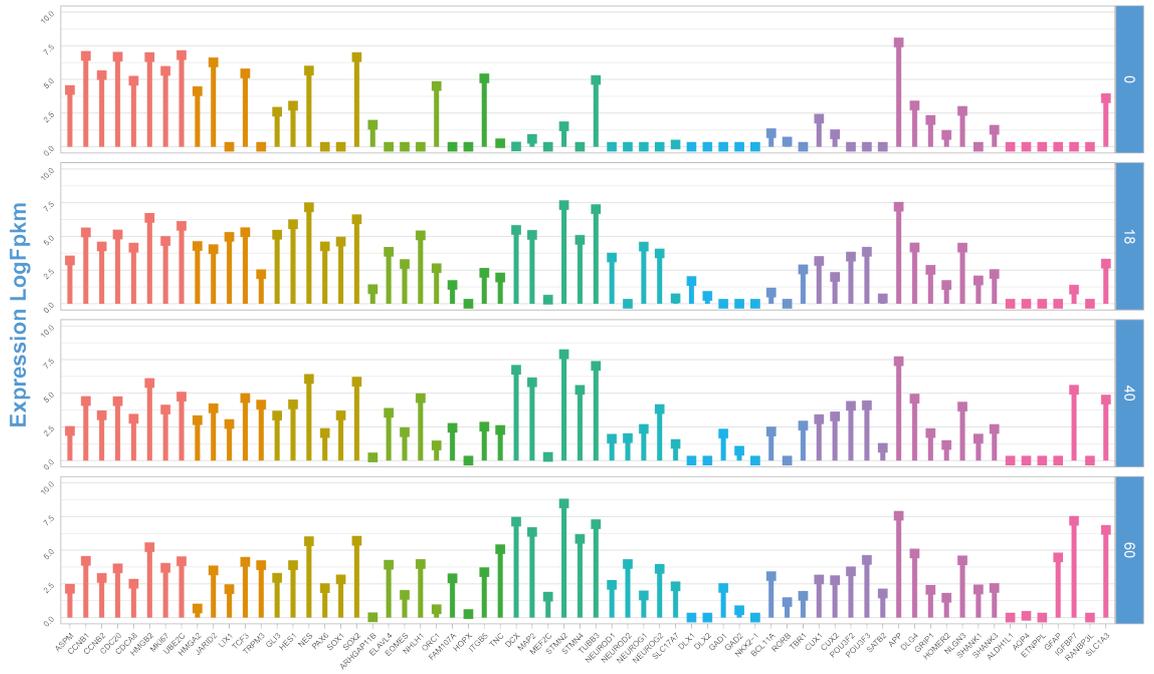
A



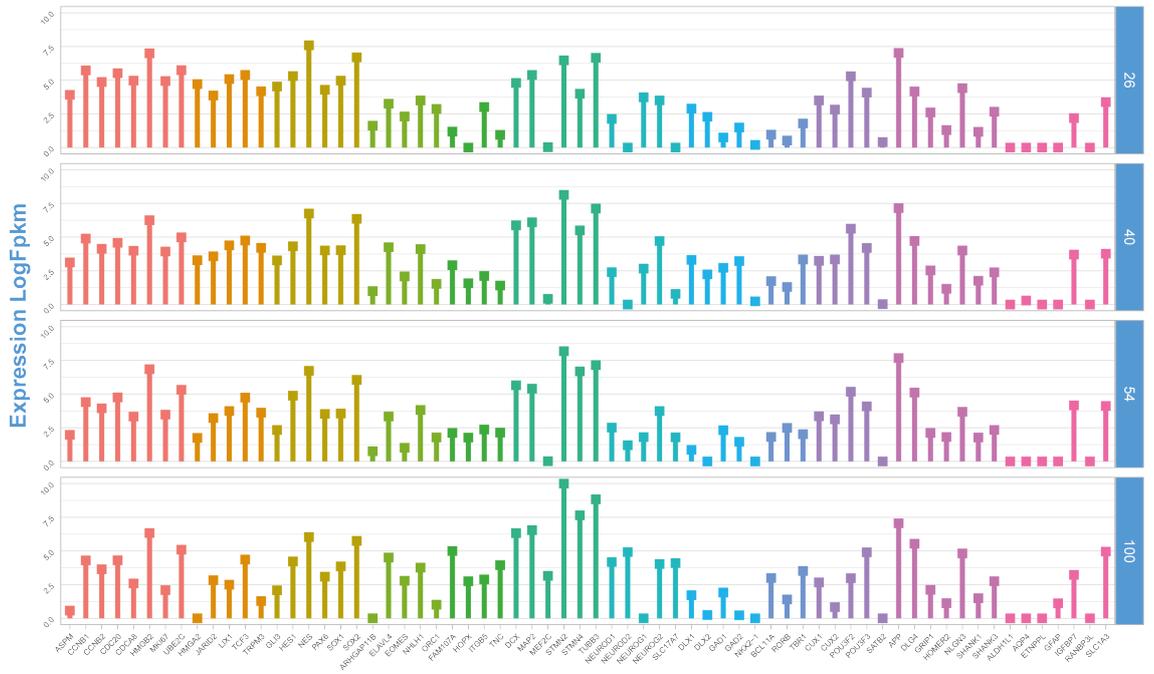
B



C



D



E

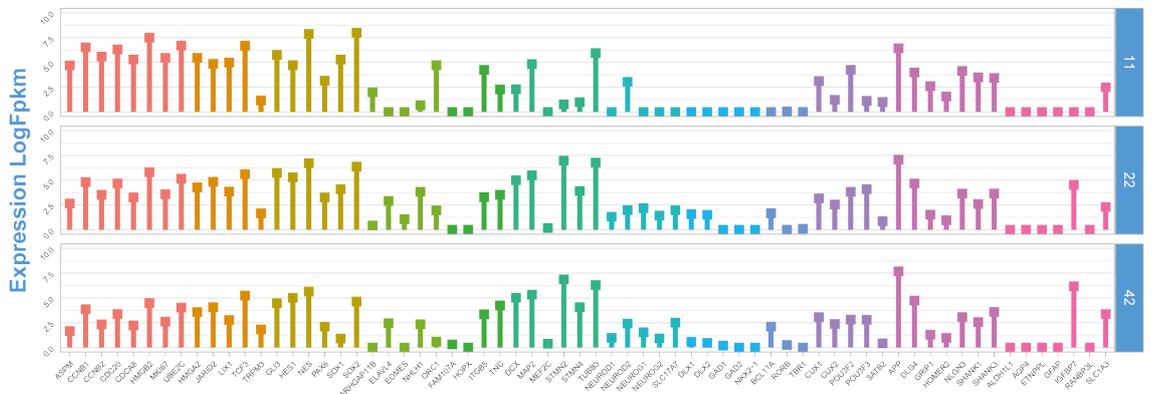
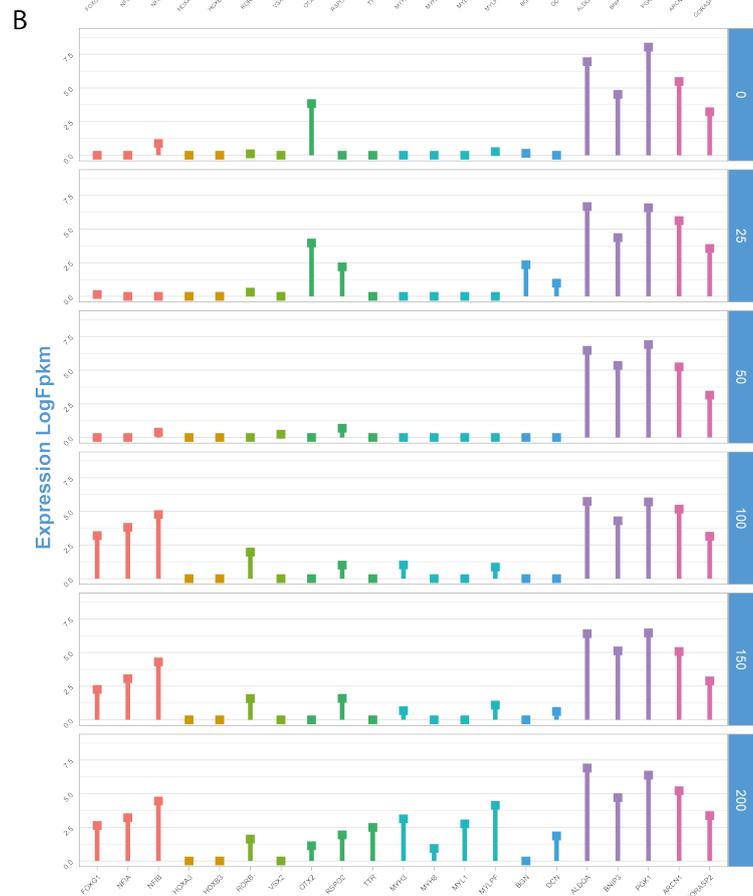
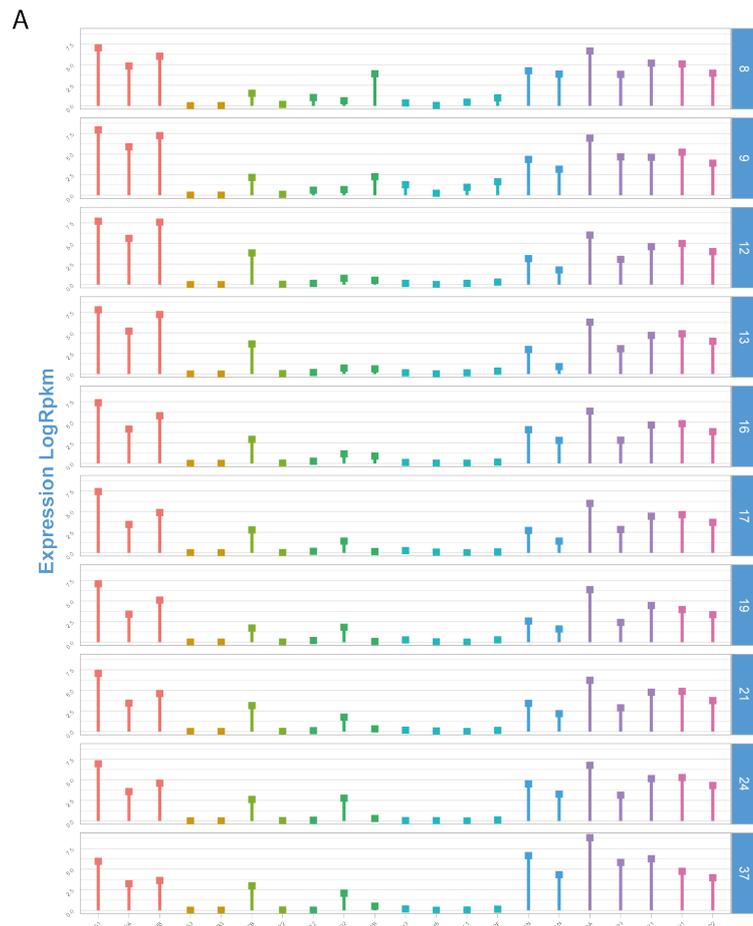


Fig. 64: Literature curated signatures for cortical populations. Lollipop charts representing the expression levels calculated as log₁₀ FPKM for genes specific of the main cortical populations in (A) BrainSpan pre-natal cortical samples, (B) CBO, (C) CO, (D) FO, and (E) TA. Each bar colour corresponds to a signature as depicted in the figure legend.

We then considered brain area specific genes, markers of off-target tissues and stress-related genes to understand to which extent brain organoids develop on-target cell types and whether culture conditions produce stress responses. When looking at telencephalon-specific determinants, we found their stable expression in BrainSpan and their upregulation along development for all organoid protocols. As expected, we did not observe any expression of hindbrain markers in BrainSpan. The same result was found in CBO, while CO and TA showed upregulation of these genes along development. Of note, FO expressed the hindbrain signature until day 54 and not at day 100. Regarding the retina signature, RORB was expressed throughout cortical development and it was found also in organoids, although this represents also a marker of low layer neurons. VSX2 instead was not expressed in BrainSpan, CBO and TA, while it was upregulated in FO until day 54 and not at day 100. Epithelial choroid plexus genes were very lowly expressed in BrainSpan at all stages. In CBO this signature was selectively expressed at day 25 and 200. The other organoids system highly expressed these genes at all stages. Surprisingly, mesodermal gene expression was found only in CBO at day 200, while they were not detected in other organoids and in cortical samples. Conversely, for mesenchymal genes we found the opposite behaviour, with low expression in CBO and FO at every stage and stable expression in the cortex, CO and TA.

Lastly, since several studies reported elevated metabolic stress in brain organoids compared to fetal tissue^{122;41}, we looked at the expression of glycolytic and ER-stress genes. Unexpectedly, we found stable expression of those genes throughout cortical development, mirrored in all brain organoid paradigms, thus ruling out a scenario where prolonged culture of organoids causes increase in cell stress.

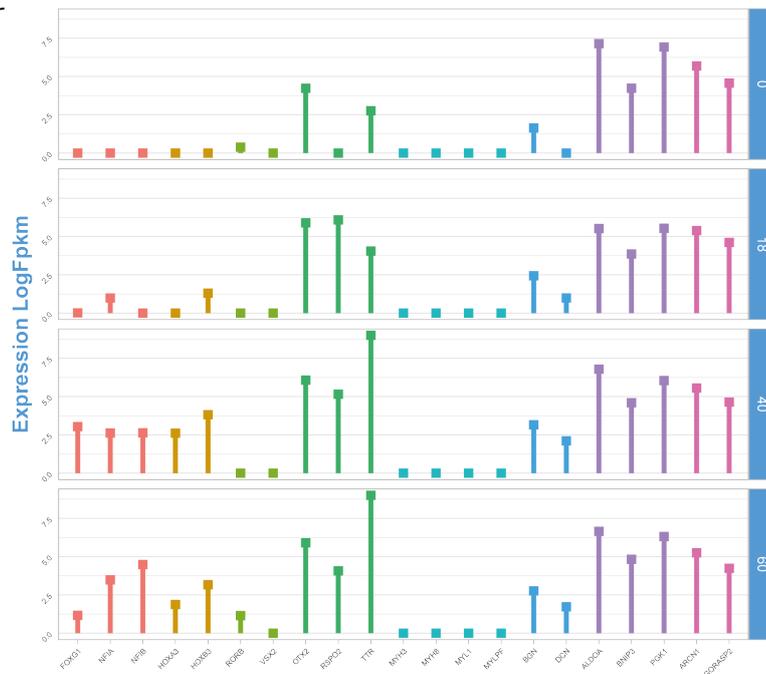
The levels of expression for all gene signatures analysed in BrainSpan and all organoid systems are reported in **Fig. 65**.



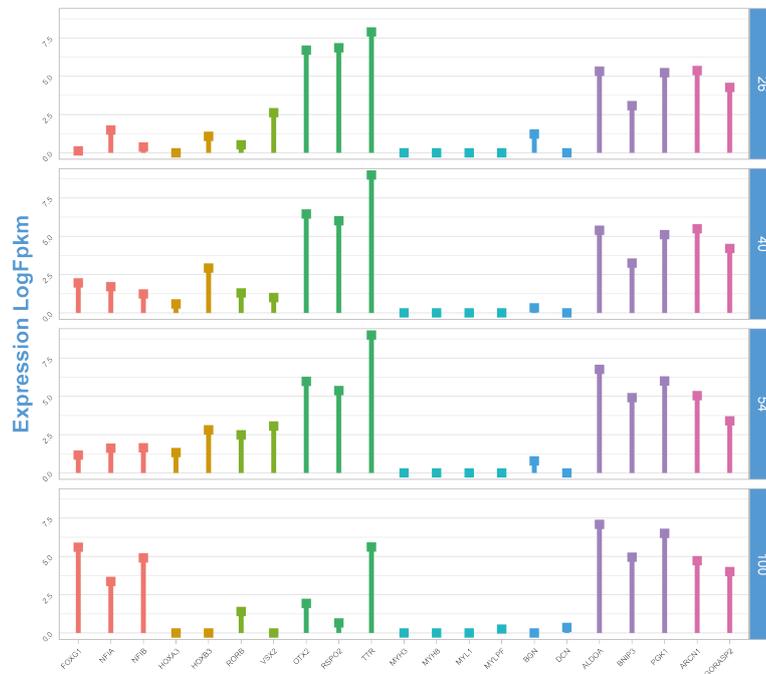
Signature

- Telencephalon
- Hindbrain
- Retina
- Choroid
- Mesoderm
- Mesenchyme
- Glycolysis
- ERStress

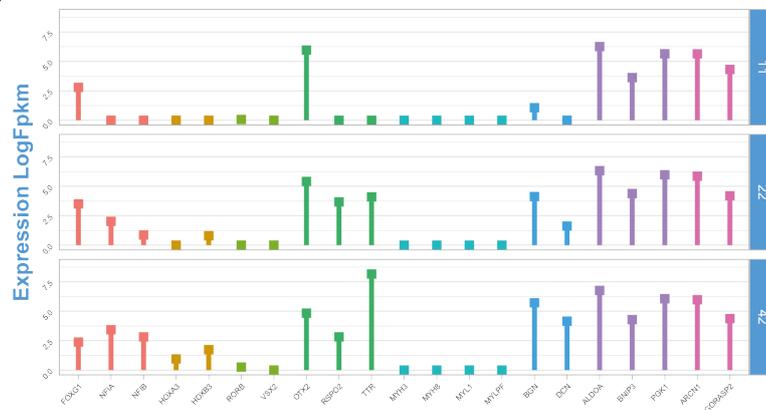
C



D



E



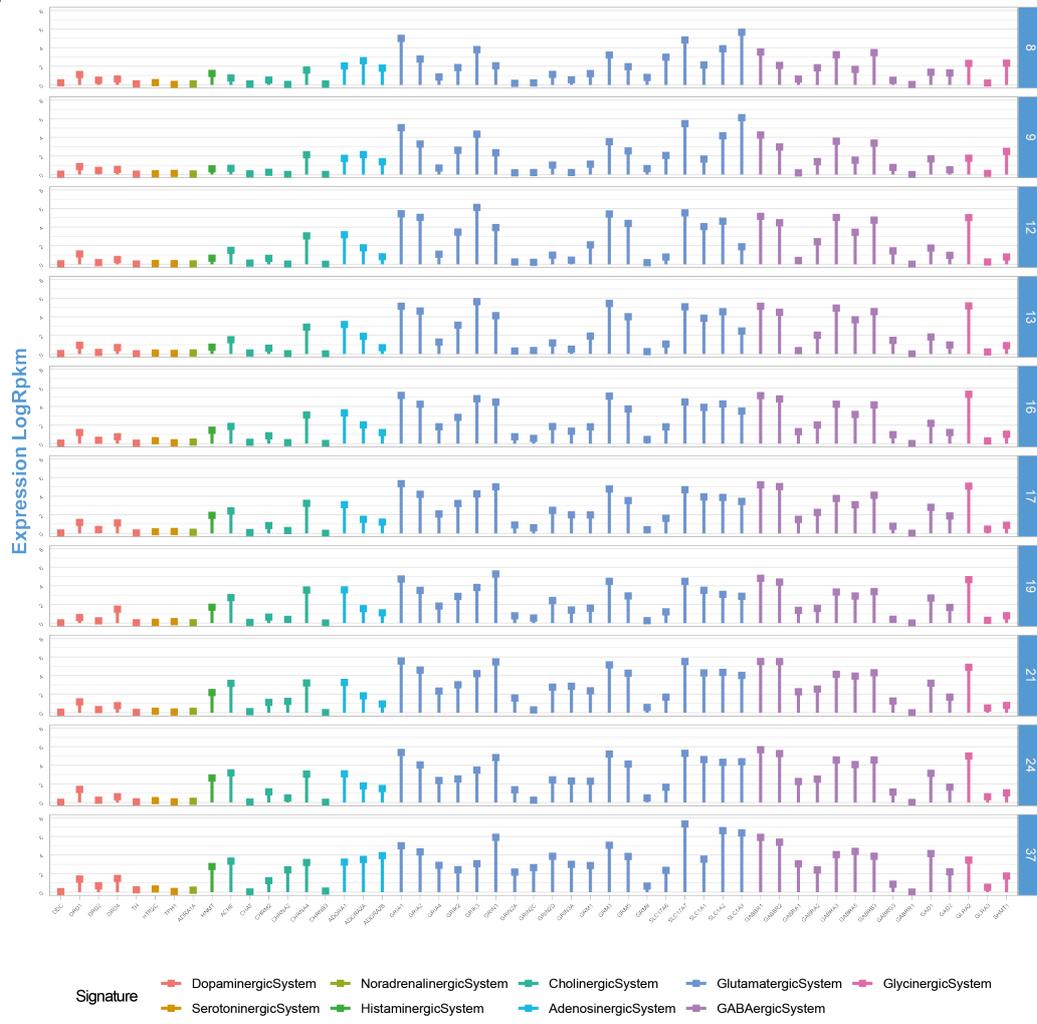
Signature

- Telencephalon
- Hindbrain
- Retina
- Choroid
- Mesoderm
- Mesenchyme
- Glycolysis
- ERStress

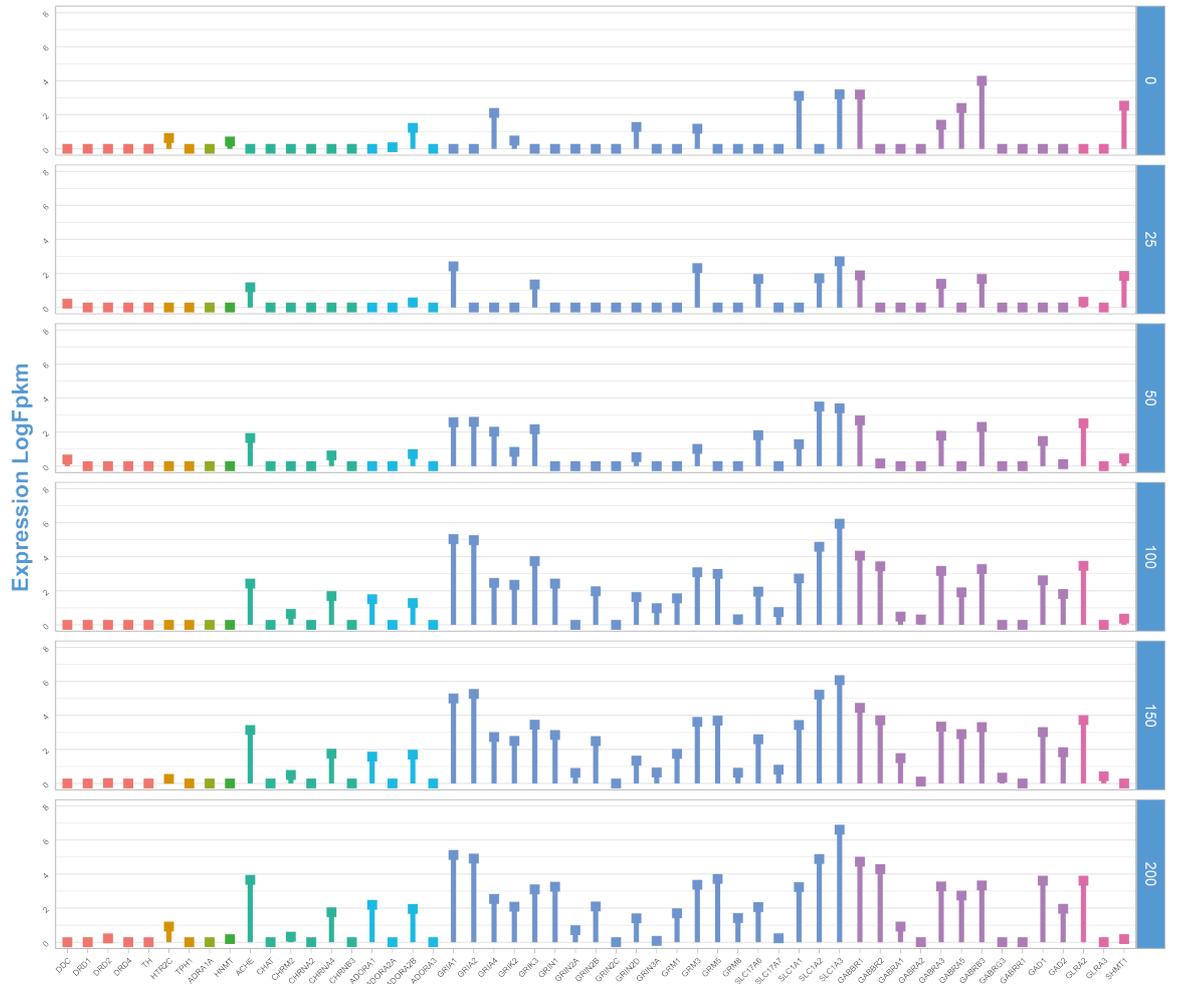
Fig. 65: Literature curated signatures for brain areas, off-target tissues and stress. Lollipop charts representing the expression levels calculated as log₁₀ FPKM for genes specific of the main cortical populations in (A) BrainSpan pre-natal cortical samples, (B) CBO, (C) CO, (D) FO, and (E) TA. Each bar colour corresponds to a signature as depicted in the figure legend.

Finally, we analysed markers for specific neurotransmission systems. As expected, the glutamatergic and GABAergic system signatures increased along development in the fetal cortex. CBO mirrored this behaviour, especially after day 100. FO behaved similarly to CBO. CO upregulated the glutamatergic system signature already at day 60 together with the GABAergic system signature. TA showed mild upregulation of the glutamatergic system signature, while not expressing the GABAergic one. BrainSpan cortical samples also showed expression of the adenosinergic, cholinergic and histaminergic system signatures, which were not stably expressed in brain organoids. Regarding the glycinergic system signature, BrainSpan cortical samples showed consistent expression only of the GLRA2 receptor, which was also expressed in CBO from day 50, CO and FO from day 40, while in TA it was not detected. The dopaminergic, serotonergic and noradrenalinergic system signatures were not detected in BrainSpan, which were also not found in CBO. CO, FO and TA expressed the serotonin receptor HTR2C. CO and FO expressed some dopaminergic receptors. TA expressed also enzymes important for dopamine metabolism, TH and DDC (**Fig. 66**)

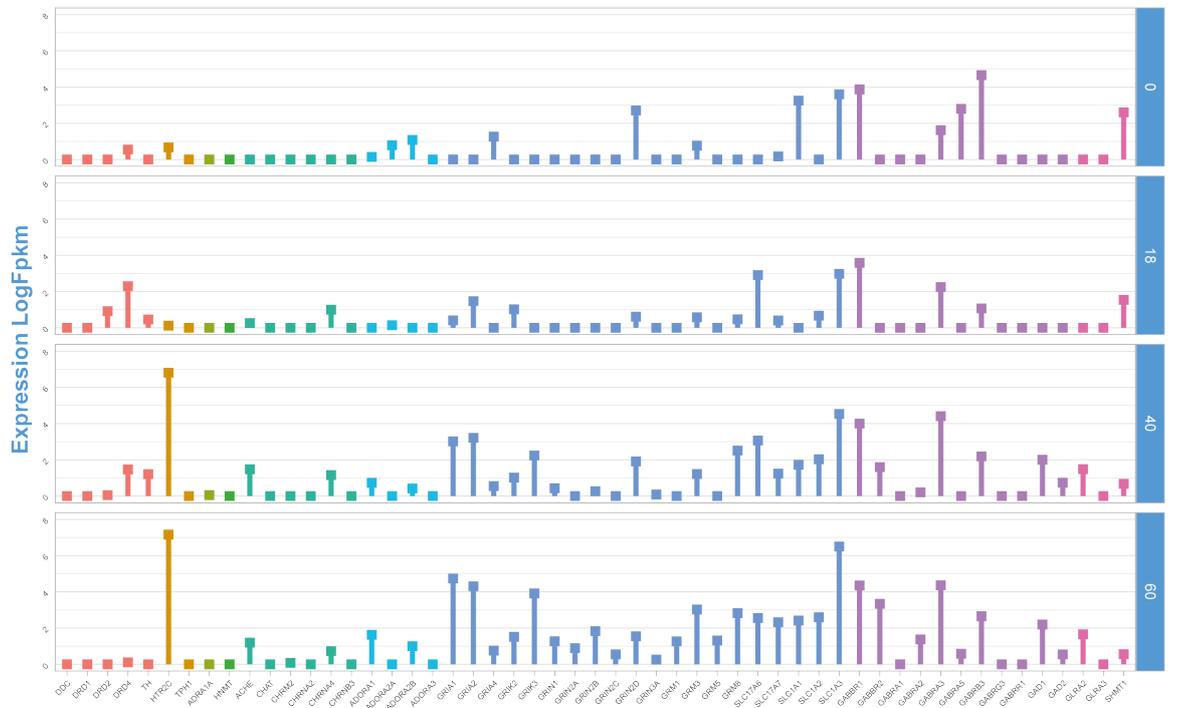
A



B



C



Signature

- DopaminergicSystem
- NoradrenalinergicSystem
- CholinergicSystem
- GlutamatergicSystem
- GlycinergicSystem
- SerotonergicSystem
- HistaminergicSystem
- AdenosinergicSystem
- GABAergicSystem

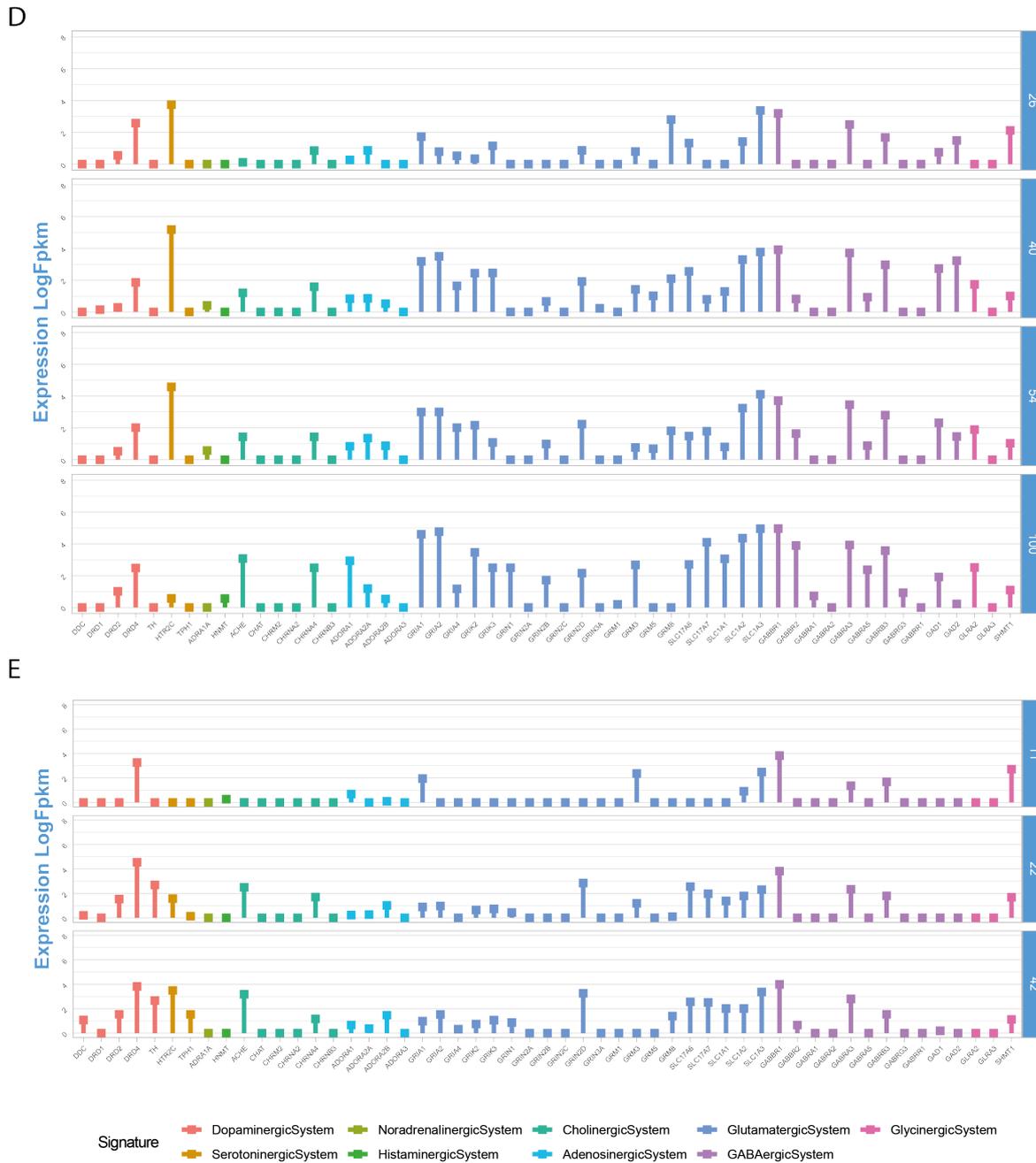


Fig. 66: Literature curated signatures for neuronal transmission systems. Lollipop charts representing the expression levels calculated as log₁₀ FPKM for genes specific for different neuronal transmission systems in (A) BrainSpan pre-natal cortical samples, (B) CBO, (C) CO, (D) FO, and (E) TA. Each bar colour corresponds to a signature as depicted in the figure legend.

In conclusion, by interrogating pre-natal cortex and different brain organoid datasets with literature-curated gene signatures we were able to pinpoint similarities and peculiarities of brain organoid systems in recapitulating populations, functions and timings of the human fetal cortex. Noteworthy, we found that CBO expressed genes specific to GABAergic interneurons, despite being patterned towards dorsal telencephalon. Moreover, we found stable expression of genes relative to glycolytic and ER stress in all organoid paradigms at all stages analysed and, most importantly, these gene were also stably expressed in the developing fetal cortex.

4.6 Transcriptional footprints of pre-natal cortical development and cortical organoidogenesis strongly overlapped and were partially recapitulated by different brain organoid paradigms

Lastly, we determined whether the WGCNA-derived co-expression patterns in fetal cortical samples and CBO were shared and how they were modulated across different brain organoid paradigms.

To do so, we measured the overlap of genes deriving from the BrainSpan and CBO WGCNA modules and calculated the odds ratio and PValue across all modules (**Fig. 67**)

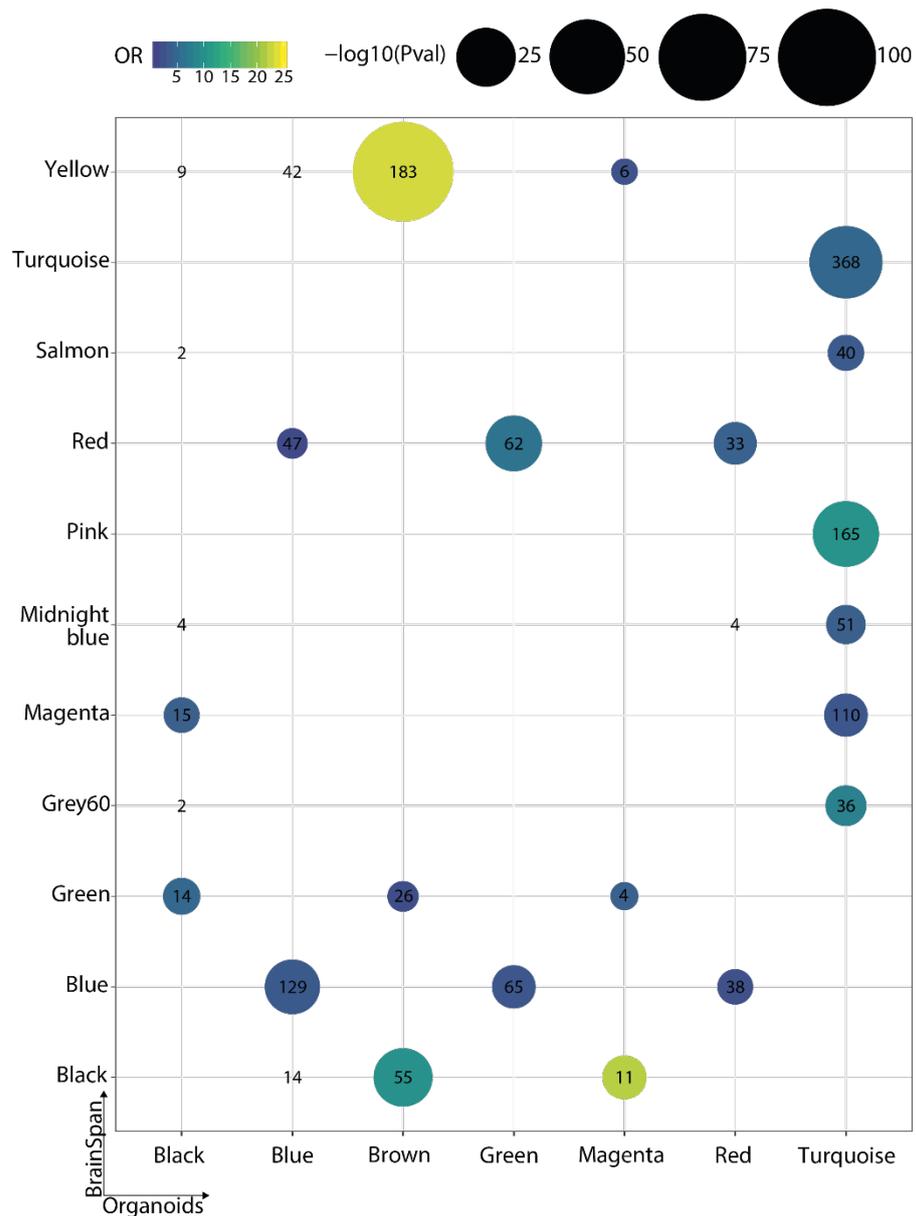


Fig. 67: Gene overlap between BrainSpan and CBO WGCNAs. Dot plot representing the overlap between BrainSpan (Y-axis) and CBO (X-axis) WGCNA modules selected as meaningful in the two analyses. Odds ratio (OR) and PValue were calculated for each overlap using as universe genes shared across the two analyses. Numbers represented shared genes and were shown only for OR>1, while dots were shown for those having also PValue < 0.05. Dot colour was assigned according to OR values and dot size according to PValue.

We first focused on modules that had positive correlation with stage. Interestingly, we found that many genes that in CBO belonged to the turquoise and black modules were included in the modules

turquoise, salmon, pink, midnight-blue, magenta and grey60 in BrainSpan (**Fig. 67**). Indeed, comparing the functions in which central genes of the CBO turquoise and black module were involved, we found consistency with the overlapping BrainSpan modules. More in details, while CBO turquoise and black genes were implicated in glutamatergic transmission, synaptic maturation, action potential (turquoise) and were generally important for forebrain development (black) (**Fig. 45**), these diverse processes were subdivided in the overlapping BrainSpan modules (**Fig. 15-19**). This finding suggested more clear-cut co-expression patterns along development for specific subfunctions in fetal cortex than in CBO. We then looked at the behaviour of this group of modules in all brain organoid systems analysed by calculating the eigengene of each gene set for every dataset and visualising their variation along differentiation. This analysis showed an overall recapitulation of their trends in all paradigms (**Fig. 68 and 69**). Indeed, the turquoise and pink module, which had the highest correlation with developmental stage, showed similar trends in all protocols (**Fig. 68A and B**). The turquoise module was more related to glutamatergic neuron differentiation and morphogenesis, while the pink genes were involved in membrane potential regulation (**Fig. 15 and 16**).

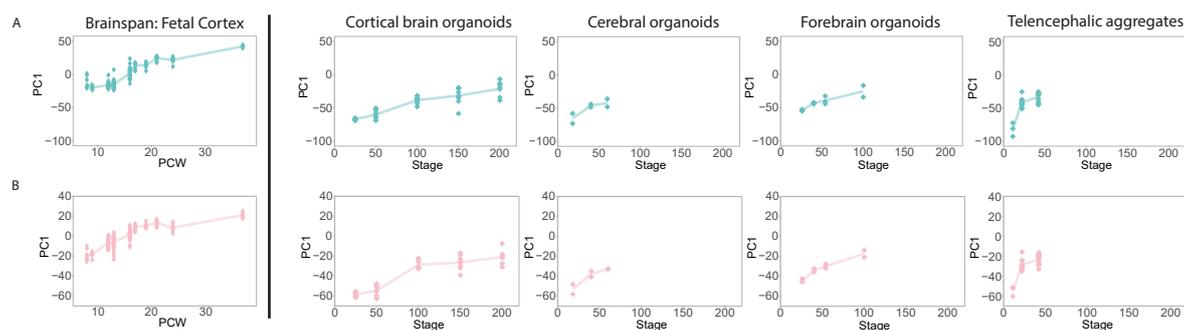


Fig. 68: BrainSpan turquoise and pink module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the BrainSpan turquoise ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(B)** Ribbon charts representing the BrainSpan pink ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

The grey60 module, also involved in regulation of action potential and ion channels (**Fig. 18**), showed stabilisation along development. This was also the case in CBO, while for CO, FO and TA we probably did not have enough time-resolution to appreciate the achievement of a plateau (**Fig. 69A**). Interestingly, CBO also recapitulated the midnight-blue and magenta modules, which for FO showed little changes along development. For CO and TA we probably did not have enough time-resolution to thoroughly analyse this behaviour (**Fig. 69B and C**). The midnight-blue module was related to general forebrain differentiation, while the magenta was more related to synapsis (**Fig. 17 and 19**). The salmon module did not show particularly interesting functions and therefore it was not taken into consideration for this analysis.

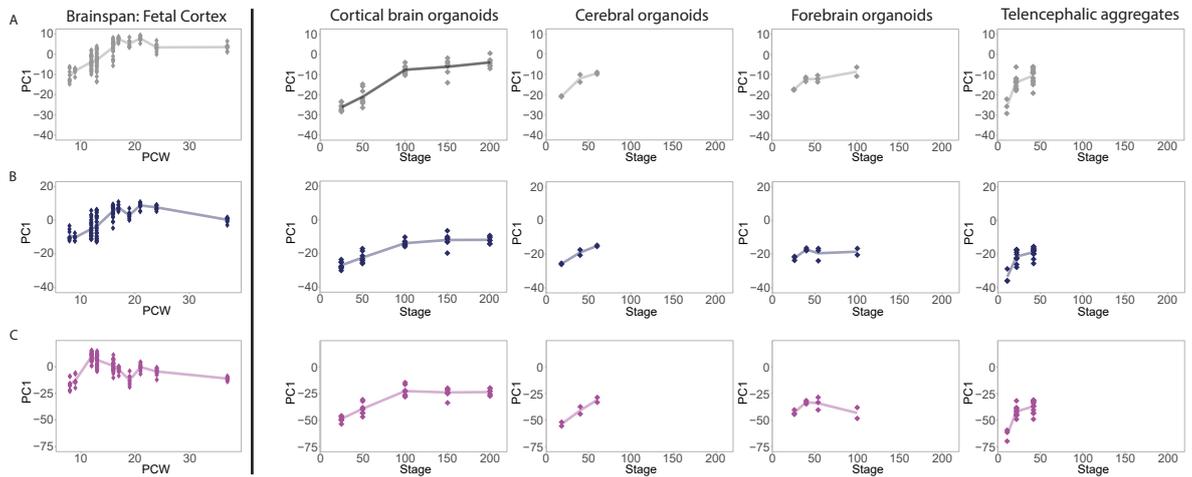


Fig. 69: BrainSpan grey60, midnight-blue and magenta module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the BrainSpan grey60 ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(B)** Ribbon charts representing the BrainSpan midnight-blue ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(C)** Ribbon charts representing the BrainSpan magenta ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

When considering the turquoise module from the CBO WGCA, we found consistent results. Indeed all protocols showed increase in the eigengene for this module (**Fig. 70A**). The CBO black module, also highly correlating with stage, was instead generally variable among samples for CBO starting from day 100 and CO at day 60, while it was less variable for FO and TA (**Fig. 70B**). The turquoise module of the CBO WGCNA was related to glutamatergic neurons differentiation and synapse, while the black module included more categories related to general forebrain development (**Fig. 45**).

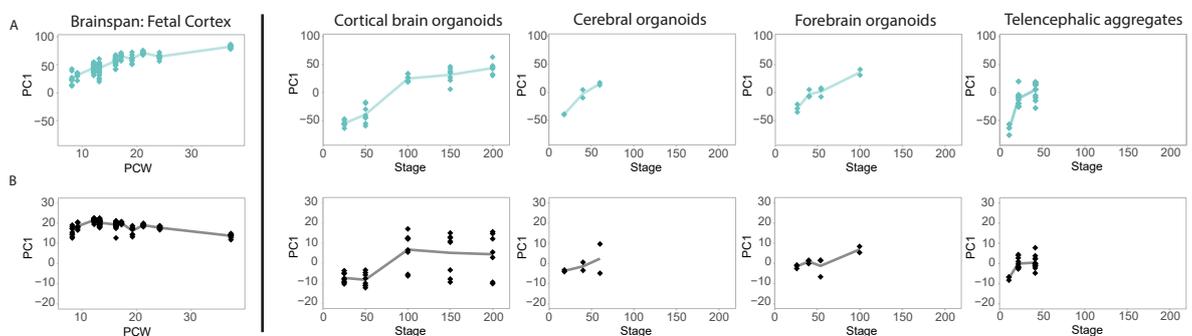


Fig. 70: CBO turquoise and black module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the CBO turquoise ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(B)** Ribbon charts representing the CBO black ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

We then focused on modules with negative correlation with stage. BrainSpan yellow and black modules overlapped with CBO modules brown and magenta (**Fig. 67**). CBO brown genes were indeed related to cell cycle and DNA replication as well as BrainSpan yellow and black ones. CBO magenta genes were related mainly to transcriptional regulation.

Looking at the behaviour of yellow module, we observed a less sudden reduction of the expression of yellow genes compared to BrainSpan for all brain organoid systems, but the overall trend of

reduction was recapitulated (**Fig. 71A**). The pattern found for the black module in BrainSpan was instead more preserved across protocols (**Fig. 71B**).

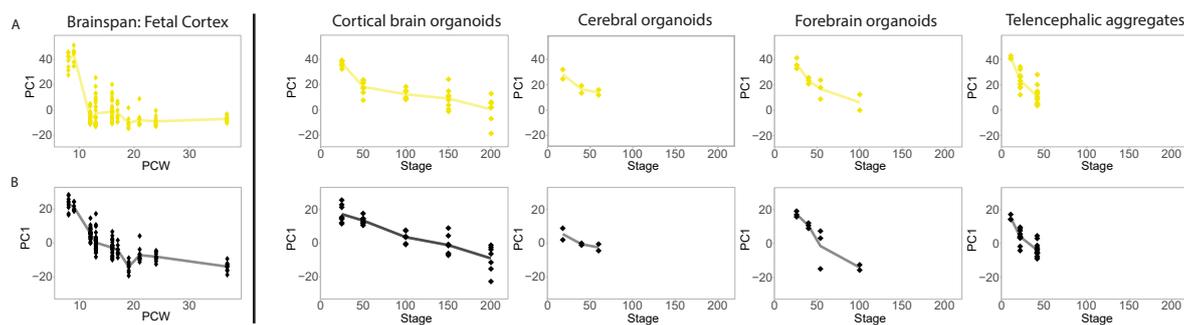


Fig. 71: BrainSpan yellow and black module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the BrainSpan yellow ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(B)** Ribbon charts representing the BrainSpan black ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

When looking at the CBO brown module, which was mostly related to cell cycle, we found a very similar behaviour in BrainSpan compared to the yellow one. In organoids, the decrease in expression was smoother in all protocols, but overall preserved (**Fig. 72A**). The CBO magenta module, related to transcriptional regulation, showed reduction along development in BrainSpan. Its behaviour in CBO was characterised by a peak of eigengene at day 50, but then the tendency of BrainSpan was preserved, as in the other protocols (**Fig. 72B**).

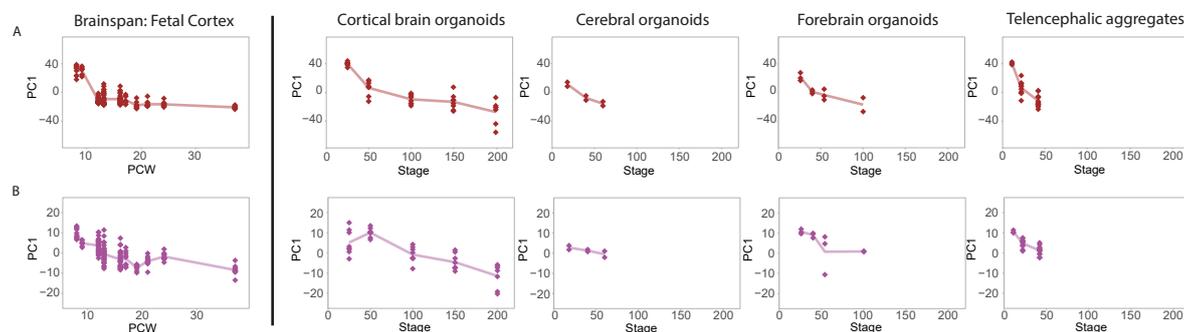


Fig. 72: CBO brown and magenta module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the CBO brown ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(B)** Ribbon charts representing the CBO magenta ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

We lastly focused on the BrainSpan blue and red module, which significantly overlapped with the CBO blue, green and red ones (**Fig. 67**). Noteworthy, these modules showed a particular U-shaped behaviour along pre-natal development with high expression at early stages, a drop in intermediate ones and then an increase towards the end of gestation (**Fig. 22A and 23A**). Importantly, a similar pattern was observed for the green and red CBO modules (**Fig. 46A and 47A**). Interestingly, the visualisation of the trend of these gene sets in brain organoids revealed different recapitulation in the systems analysed. While CBO mostly recapitulated the U-shape found in BrainSpan, although with a milder drop at intermediate stages than in cortical samples, CO, FO and TA did not. CO showed very little variation of these module along differentiation, FO showed a drop at day 100

and TA showed an opposite trend for the red module and no changes for the blue one (**Fig. 73A and B**).

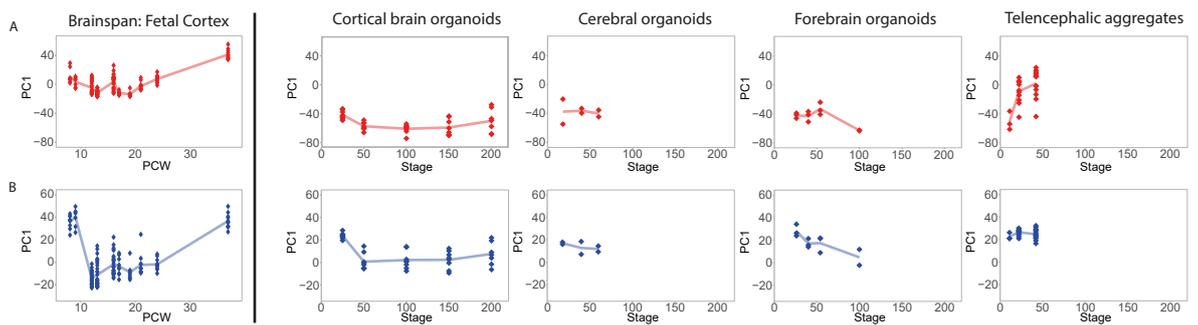


Fig. 73: BrainSpan red and blue module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the BrainSpan red ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(B)** Ribbon charts representing the BrainSpan blue ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

Genes of the BrainSpan red module were mainly related to extracellular matrix regulation, which was also a function enriched in the blue one. Of note, the blue module included genes related to astrocyte differentiation.

As already anticipated, the CBO green and red modules showed a very similar behaviour to the BrainSpan red and green ones, and indeed they were also enriched in genes involved mainly in extracellular matrix regulation (**Fig. 48**). The behaviour of those modules in CBO was extremely similar to the one in BrainSpan, while CO and TA showed little variation. FO showed only a reduction for the green module and little variation for the red one (**Fig. 74 A and B**).

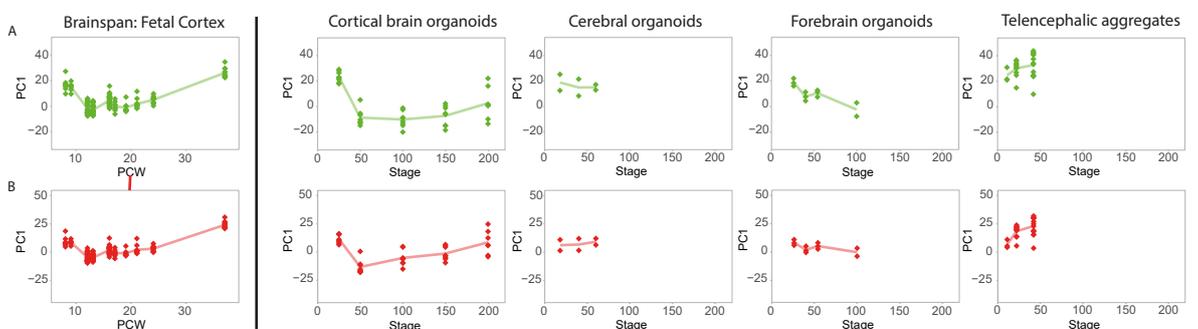


Fig. 74: CBO green and red module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the CBO green ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA. **(B)** Ribbon charts representing the CBO red ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

The BrainSpan red and blue module also significantly overlapped with the CBO blue module (**Fig. 67**). This module was in general regulated to transcriptional and translational regulation. Interestingly, it showed a reduction in CBO until day 100 to then stabilise. Instead, in BrainSpan an U-shaped behaviour was observed. As for CBO, CO and FO showed a reduction of the blue module along differentiation, while TA showed little variation (**Fig. 75**).

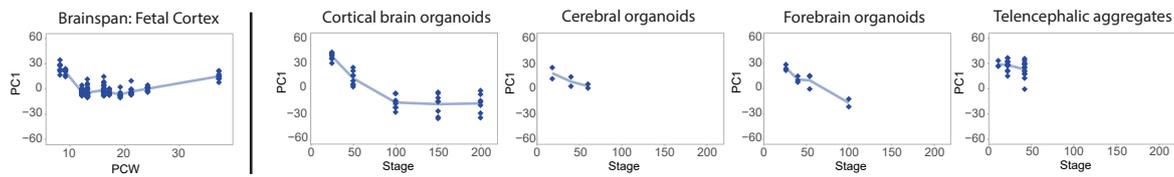


Fig. 75: CBO blue module behaviour across BrainSpan and brain organoid datasets. (A) Ribbon charts representing the CBO blue ME (Y-axis) trend over developmental time (X-axis) in BrainSpan cortical pre-natal samples and CBO, CO, FO and TA.

In summary, we demonstrated that co-expression patterns of the developing cortex were overall recapitulated in brain organoids by two independent analyses. Indeed, WGCNA on BrainSpan and CBO revealed strong overlap between modules and relative functions. The projection of the BrainSpan and CBO modules highly positively or negatively correlating with developmental stage in different organoid paradigms revealed similar trends for all protocols. In parallel, these analyses also revealed transcriptomic signatures particularly related to extracellular matrix that followed peculiar behaviours along pre-natal development and that were fully recapitulated by CBO but not by the other organoid systems analysed.

4.7 Definition of the WS hiPSC cohort for disease modelling in relevant cellular lineages

Following the detailed definition of CBO transcriptional dynamics and their benchmarking against the human pre-natal cortex as well as comparison against other brain organoid systems, we used CBO as a model for the study of a paradigmatic syndrome caused by mutation in epigenetic regulators and characterised by intellectual disability, WS. Indeed, this disease is caused by mutation in multiple components of the PRC2 complex, mainly in the catalytic subunit EZH2, and it is characterised by two main dysfunctional domains: i) intellectual disability and ii) overgrowth with particularly severe facial dysmorphism²²³. In order to study the molecular deregulation due to PRC2 malfunction in WS-disease-relevant cell types, we reprogrammed skin fibroblasts from 4 patients along with 1 half-matched control (**Fig. 76**).

Sample	Sex	Mutation	Main clinical signs
WVS02A	M	EZH2 exon16: c.1876G>A; p.Val626Met	Tall stature, advanced bone age, moderate delayed motor, speech and cognitive development, macrocephaly, large ears, mild ocular hypertelorism, relative retrognathia, large hands
WVS01H	F	EZH2: c.394 C>T; p.Pro132Ser	Excessive growth, severe accelerated osseous maturation, hypotonia, mild intellectual disability, poor fine motor coordination and balance, prominent macrocephaly, large bifrontal diameter
WVS04A	M	Other PRC2 subunit: EED pHis258Tyr	Macrosomia, advanced bone, generalized overgrowth, macrocephaly, large ears, retrognathia, kyphosis, large hands
WVS03B	F	EZH2 exon20: c.2212G>A; p.Ala738Thr	"The phenotype of the patient seems quite far from typical Weaver syndrome although some facial features may resemble Weaver patients"
CTL09A	M	Control (Father of WVS04A)	N/a

Fig 76: Collection of WS samples. Collection of WS samples of the cohort analysed complete of the information regarding genetic lesions and clinical features.

Noteworthy, 3 patients had mutations in EZH2 (Val626Met, Pro132Ser, Ala738Thr), while the fourth carried a mutation that affected EED (His258Tyr) (**Fig. 77**).

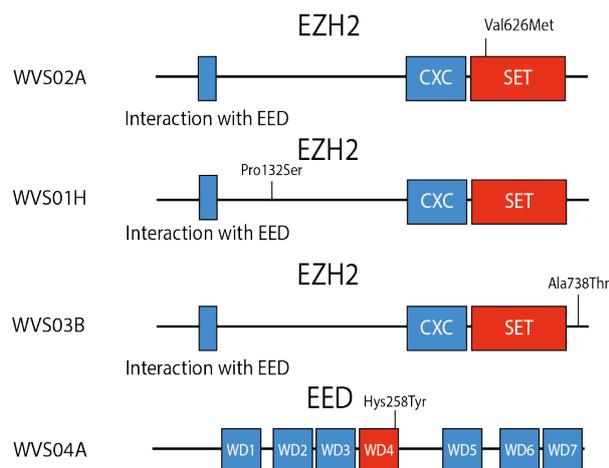


Fig. 77: Weaver disease-causing mutations. Schematic representation of the Weaver causing mutations in the cohort analysed .

For cell reprogramming, we used a non-integrating self-replicating mRNAs³⁶⁷. The obtained iPSC clones were characterised by immunofluorescence for pluripotency markers NANOG, and OCT-4 (Fig. 78).

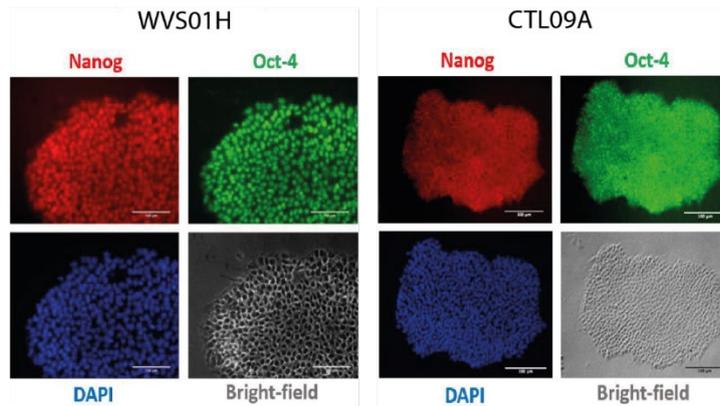


Fig. 78: Representative immunostainings for pluripotency markers. Representative immunostainings for pluripotency markers of hiPSC lines obtained by reprogramming of WS patient-derived fibroblasts: in red Nanog, in green Oct-4, in blue DAPI and in grey the bright-field image.

To complete our cohort, we included 3 unrelated control hiPSC lines for a total of 4 individuals per group, along with a CRISPR/Cas9 engineered line described below. We differentiated these hiPSC lines into 3 different *in vitro* cellular models and profiled them transcriptomically and epigenomically : i) CBO, to study the WS-related intellectual disability in a developing system recapitulating human cortical development where all the relevant cell types appear in a chronologically ordered fashion; ii) induced neurons (iN), to compare homogeneous populations of mature glutamatergic neurons for the identification of alterations in their function possibly contributing to the clinical phenotype of patients; and iii) NCSC, as they represent the precursors of many skeletal structures, including the craniofacial ones (experimental scheme in Fig. 79).

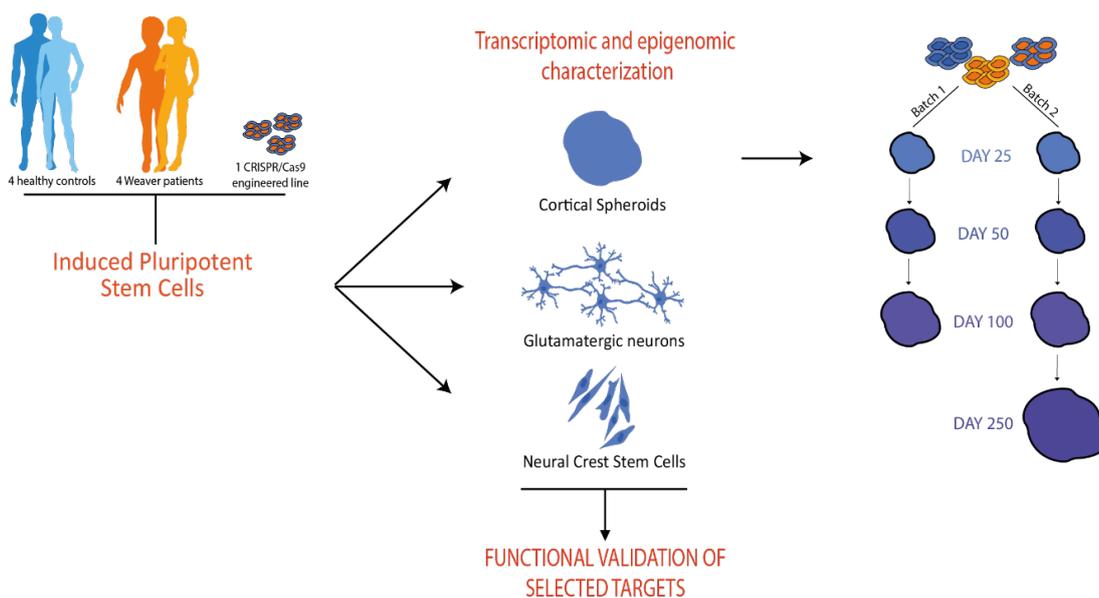


Fig. 79: Graphical representation of the cohort, experimental models and tools used.

A recent study demonstrated that mutations in EZH2 leading to WS resulted in a decrease of its enzymatic activity²⁴⁸. To study the effect of a severe EZH2 loss of function (LOF) and understand its reverberation on CBO differentiation capability, we generated a homozygous frameshift in the SET domain of EZH2 in a control hiPSC line (EZH2-LOF) through CRISPR/Cas9 (**Fig. 80**).

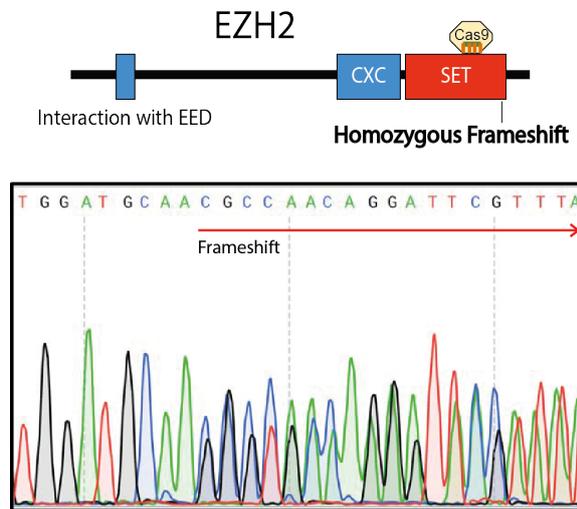


Fig. 80: EZH2-LOF engineering. Sanger sequencing confirmation of the correct CRISPR/Cas9-mediated editing of the EZH2-LOF line carrying an homozygous frameshift in the SET domain of EZH2 (EXON18).

We analysed the differentiation potential of this isogenic mutant line into the previously described lineages. Outstandingly, the specification of EZH2-LOF into NCSC failed completely (data not shown), confirming that EZH2 is pivotal for the differentiation of this lineage. Instead, iN derivation did not show any obvious differentiation defect, suggesting that bypassing the progenitor stage avoided PRC2-related deficiencies (data not shown). As expected, we found that mRNA levels of EZH2 were drastically reduced in CBO from the EZH2-LOF line at different stages of differentiation (**Fig. 81**) and that EZH2-LOF CBO showed reduced growth (see paragraph below). This isogenic control enabled us to infer the molecular dysregulation directly dependent on PRC2 LOF.

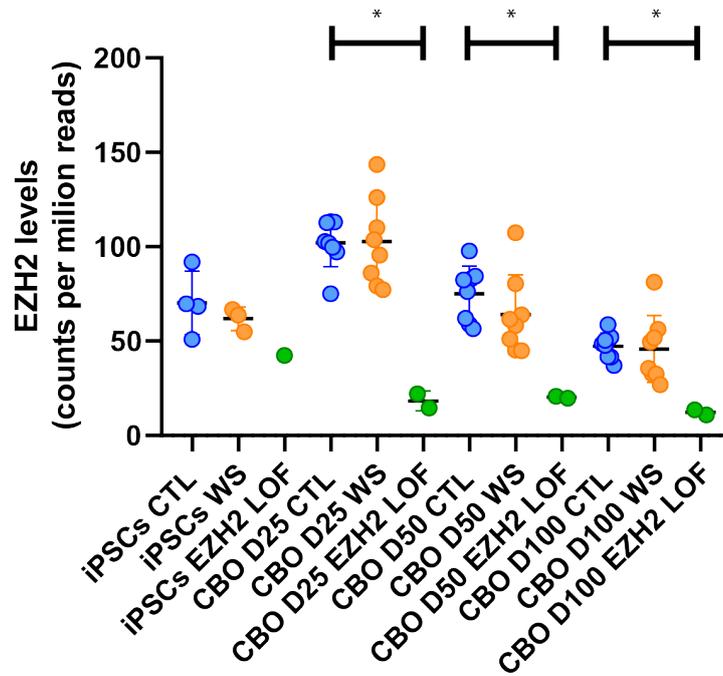


Fig. 81: EZH2 levels in EZH2-LOF samples. EZH2 levels in the transcriptome of control, patient and frameshift CBO expressed in library-size-normalised counts per million reads. Statistical test: Mann-Whitney; * PValue < 0.05. iPSCs: n= 4 controls, 4 WS, 1 EZH2 LOF; CBO day 25: n= 8 controls, 8 WS, 2 EZH2-LOF; CBO day 50: n= 8 controls, 8 WS, 2 EZH2-LOF; CBO day 100: n= 8 controls, 8 WS, 2 EZH2-LOF.

4.8 The WS mutational spectrum differentially affected CBO growth revealing patient-specific phenotypes

Since macrocephaly represents one of the features of WS patients, we wondered whether CBO were able to recapitulate this phenotype by measuring their growth over time. To ensure reproducibility in the quantification of organoids size, we grew them in 96 well plates up to 50 days of differentiation instead of transferring them in 10 cm plates at day 12 (see methods), thus preventing their fusion that could confound our results and allowing us to generate growth curves following single organoids over time. Importantly, after day 50 we could not proceed with the analysis since organoids started to disintegrate as they become too big for the culture format used. By assuming the spherical shape of organoids, we measured the area of 12 organoids per line derived in a total of 4 independent batches of differentiation to tackle inter-line and technical variability. All measurements were normalised on the size of the respective EB at day 5, the end of the neuroepithelium specification phase of organoidogenesis. The time-points analysed were day 6, 15, 25 and 50.

We first evaluated results by pooling together control- and patient-organoids from all differentiation rounds. With this approach, we observed that WS organoids were significantly smaller at day 6 and at day 15, whereas at days 25 and 50 differences were not statistically significant (**Fig. 82A**). Representative images of control and WS CBO used for the growth curves are reported in **Fig. 82B**.

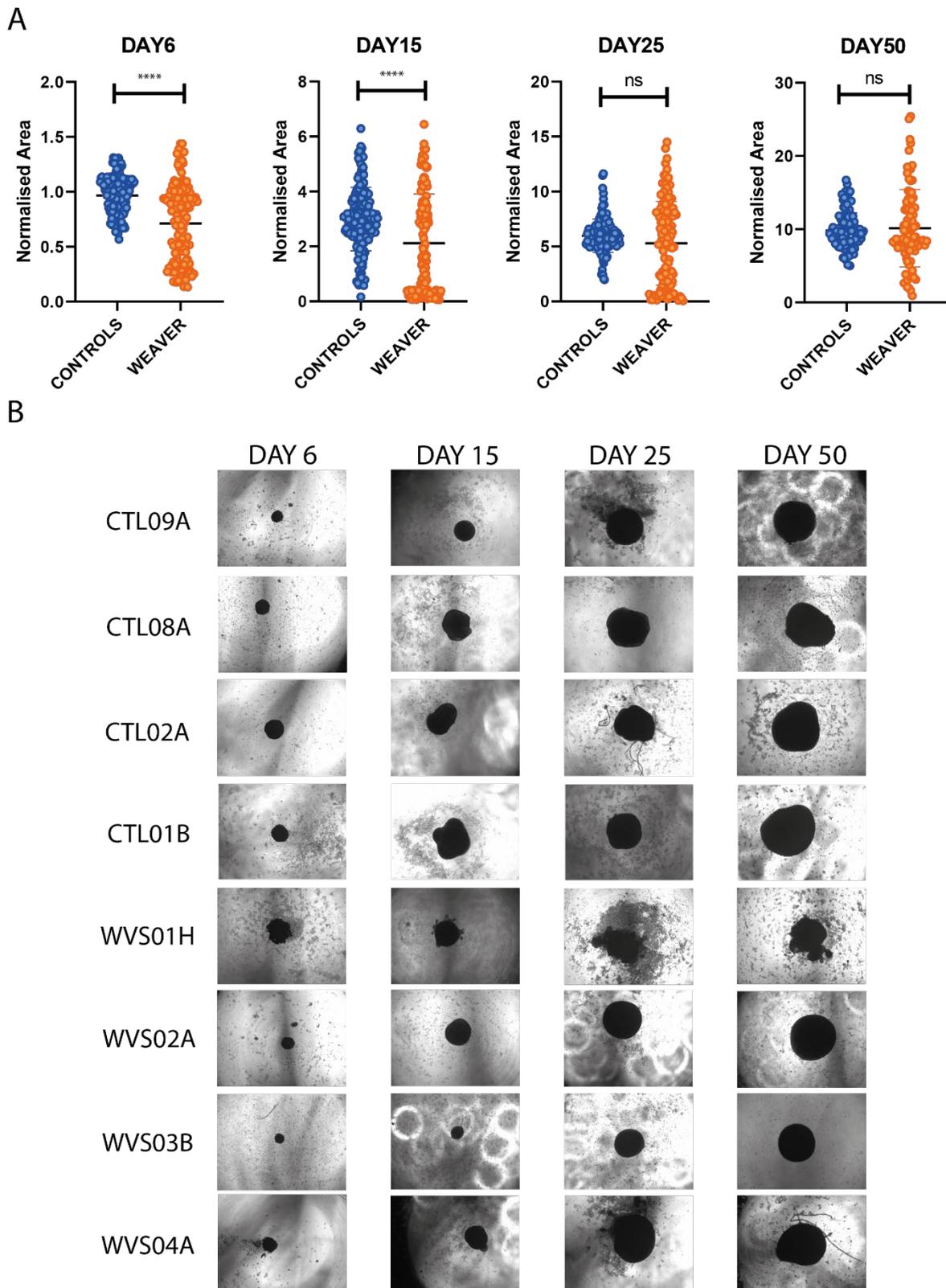


Fig. 82: CBO growth WS vs controls. (A) Overall measurements of the area of control and WS CBO at day 6, 15, 25 and 50 of differentiation grouping all batches of differentiation together. Statistical test: Mann-Whitney; **** PValue < 0.0001 . Day 6: Controls = 156; WS = 129. Day 15: Controls = 155; WS = 143. Day 25: Controls = 149; WS = 117. Day 50: Controls = 109; WS = 88. Differentiation batches = 4. **(B)** Representative images of the CBO used for the growth curves for all the patents and control lines at the time-points analysed.

We then considered all batches as single experiments to understand whether there was variability of CBO differentiation. At day 6, all the analyses showed a sustained lower growth of WS CBO. At

day 15 and 25, the same results were confirmed for 2 out of 4 batches of differentiation, while we did not observe any significant difference for batch n° 2 and an increase in size for batch n° 1. Day 50 WS CBO showed increased size than control in the first round of differentiation, while in the second and third they were smaller and in the fourth the difference was not significant (**Fig. 83**).

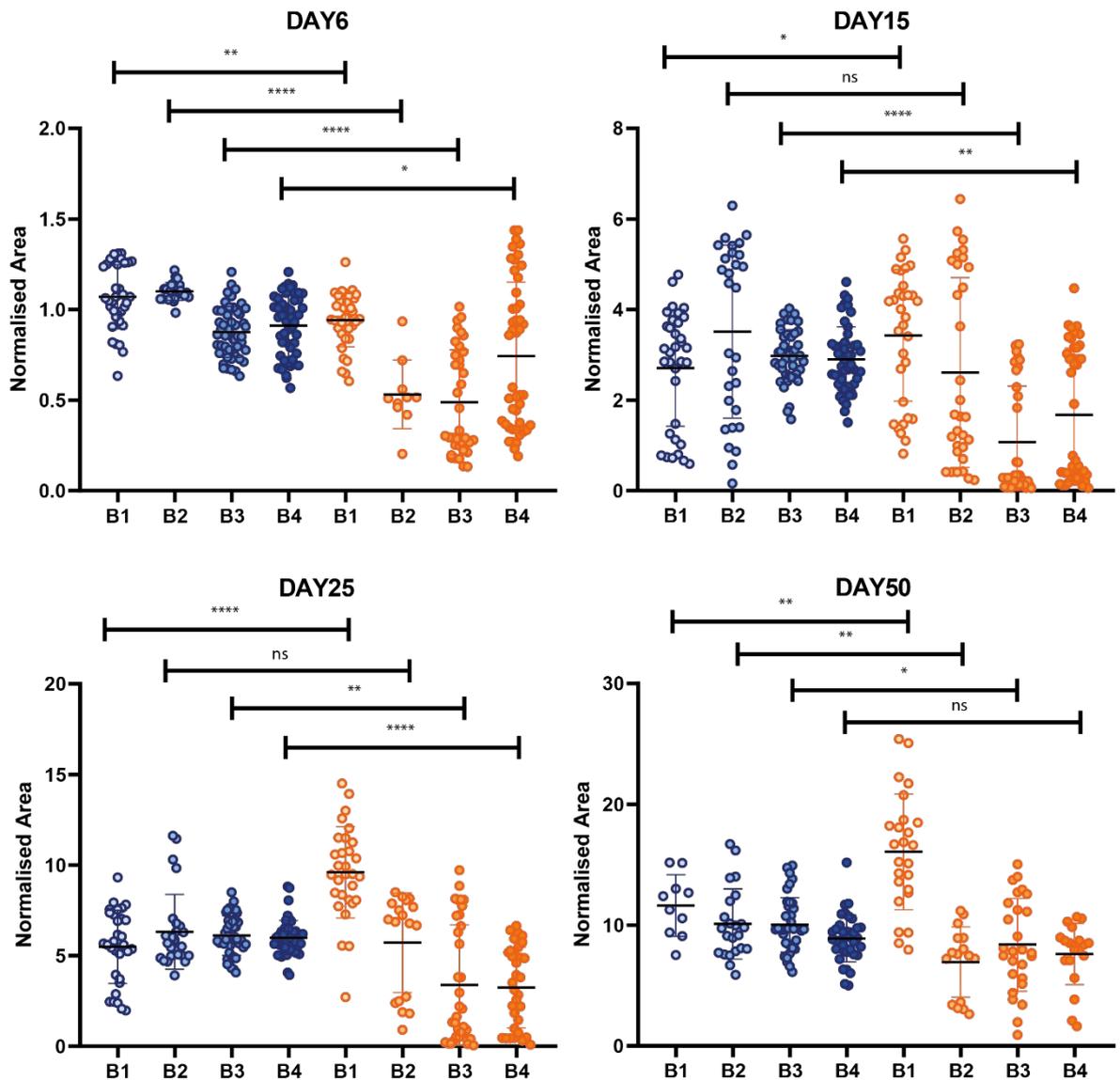


Fig. 83: CBO growth WS vs controls divided by batch. Measurements of the area of control and WS CBO at day 6, 15, 25 and 50 of differentiation considering each batch as a single experiment. Statistical test: Mann-Whitney; * PValue = 0.05; ** PValue = 0.01; *** PValue = 0.001; **** PValue = 0,0001. Day 6: Controls = 156; WS = 129. Day 15: Controls = 155; WS = 143. Day 25: Controls = 149; WS = 117. Day 50: Controls = 109; WS = 88. Differentiation batches = 4.

These discordant results led us to the analysis of each line independently to understand whether they behaved differently, given also the intrinsic genetic and phenotypic differences of WS patients in our cohort. We pooled together all WS CBO deriving from specific patient lines from the four batches of differentiation and we analysed their differential growth against the totality of control CBO.

This analysis showed that the WVS02A formed consistently bigger organoids than controls, in accordance with the clinical features of the patient, who displayed a macrocephaly²⁴⁸ (Fig. 84).

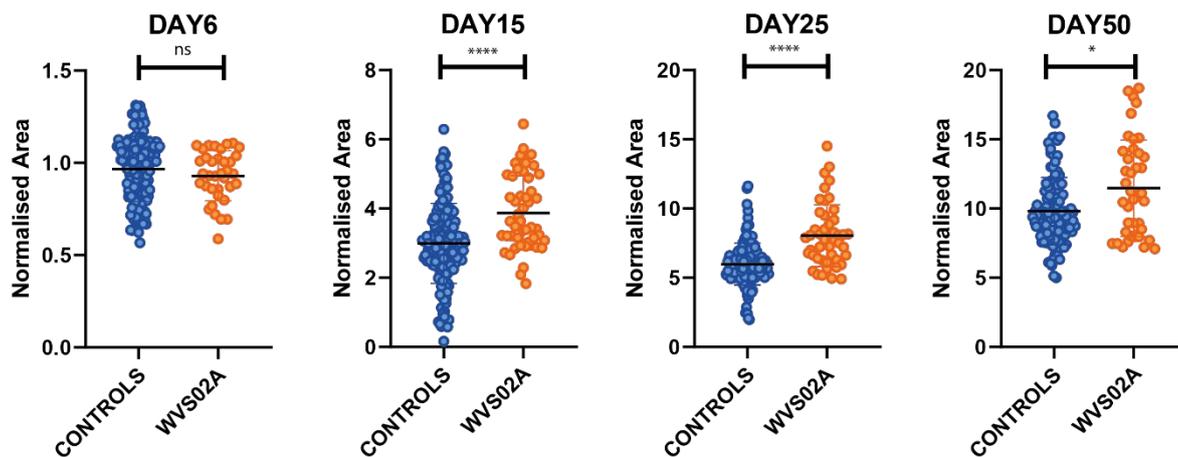


Fig. 84: CBO growth WVS02A vs controls. Measurements of the area of control and WVS02A CBO at day 6, 15, 25 and 50 of differentiation. Day 6: Controls = 156; WVS02A = 36. Day 15: Controls = 155; WVS02A = 48. Day 25: Controls = 149; WVS02A = 47. Day 50: Controls = 109; WVS02A = 42. Statistical test: Mann-Whitney test; * p value = 0.05; ** p value = 0.01; *** p value = 0.001; **** p value = 0,0001. Differentiation batches = 4.

On the other hand, WVS03B and WVS04A CBO showed an overall reduced growth compared to controls (Fig. 85). Interestingly, for WVS03B this phenotype was observed until day 25. This patient showed a clinical phenotype quite far from a typical Weaver patient, although the characteristic craniofacial defects resembled WS. The results observed for WVS04A were instead in contrast with the diagnosed macrocephaly of this patient. For WVS01H these specific analyses were not performed because of the low number of measurements due to the difficulties of this hiPSC line to differentiate in brain organoids.

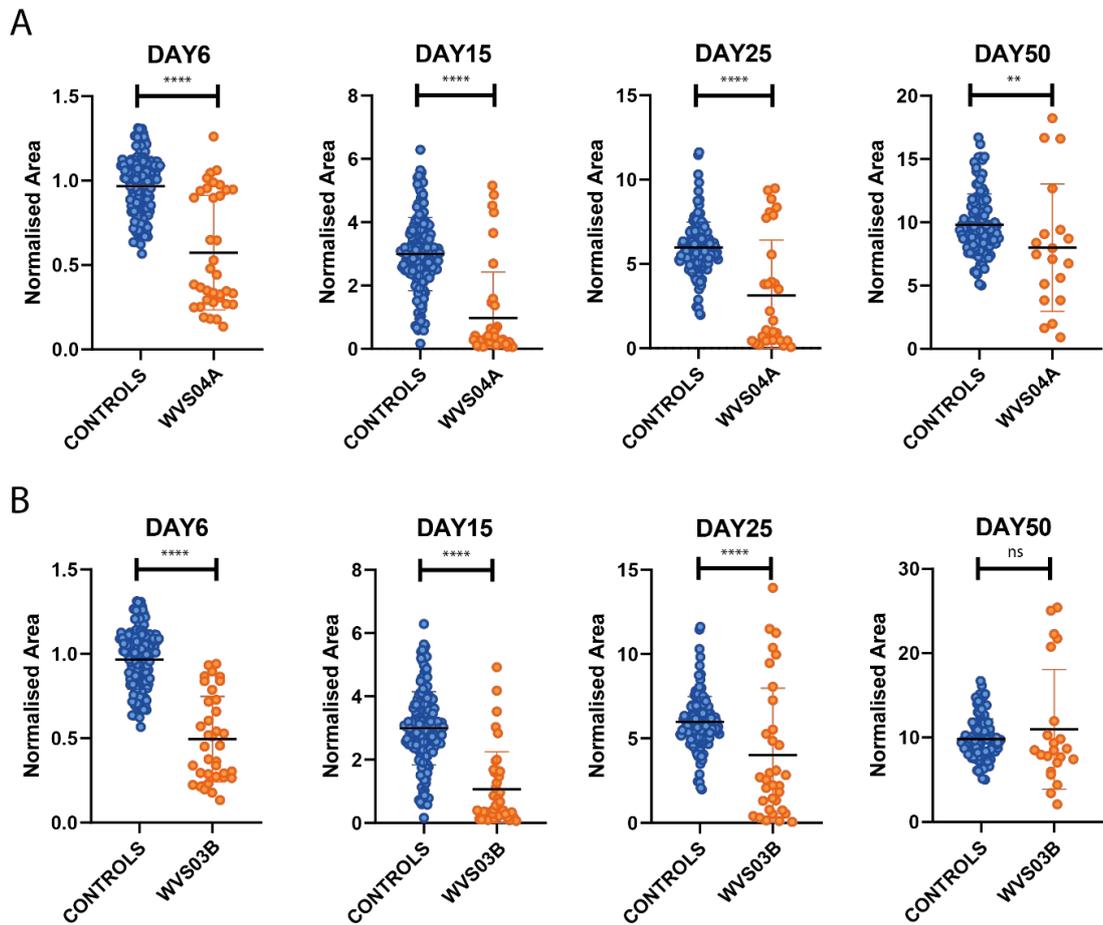


Fig. 85: CBO growth WVS04A and WVS03B vs controls. (A) Measurements of the area of control and WVS04A CBO at day 6, 15, 25 and 50 of differentiation. Day 6: Controls = 156; WVS04A = 36. Day 15: Controls = 155; WVS04A = 40. Day 25: Controls = 149; WVS04A = 31. Day 50: Controls = 109; WVS04A = 19. Statistical test: Mann-Whitney; * p value = 0.05; ** p value = 0.01; *** p value = 0.001; **** p value = 0,0001. Differentiation batches = 4. **(B)** Measurements of the area of control and WVS03B CBO at day 6, 15, 25 and 50 of differentiation. Day 6: Controls = 156; WVS03B = 36. Day 15: Controls = 155; WVS03B = 40. Day 25: Controls = 149; WVS03B = 33. Day 50: Controls = 109; WVS03B = 22. Statistical test: Mann-Whitney; * p value = 0.05; ** p value = 0.01; *** p value = 0.001; **** p value = 0,0001. Differentiation batches = 4.

The same analyses were performed also comparing the EZH2-LOF line against its isogenic control. Remarkably, EZH2-LOF CBO displayed a consistent and sustained lower growth compared to control, indicating that severe EZH2 LOF impaired CBO differentiation (**Fig. 86A**). Representative images of EZH2-LOF CBO used for the growth curves at the stages analysed are depicted in **Fig. 86B**.

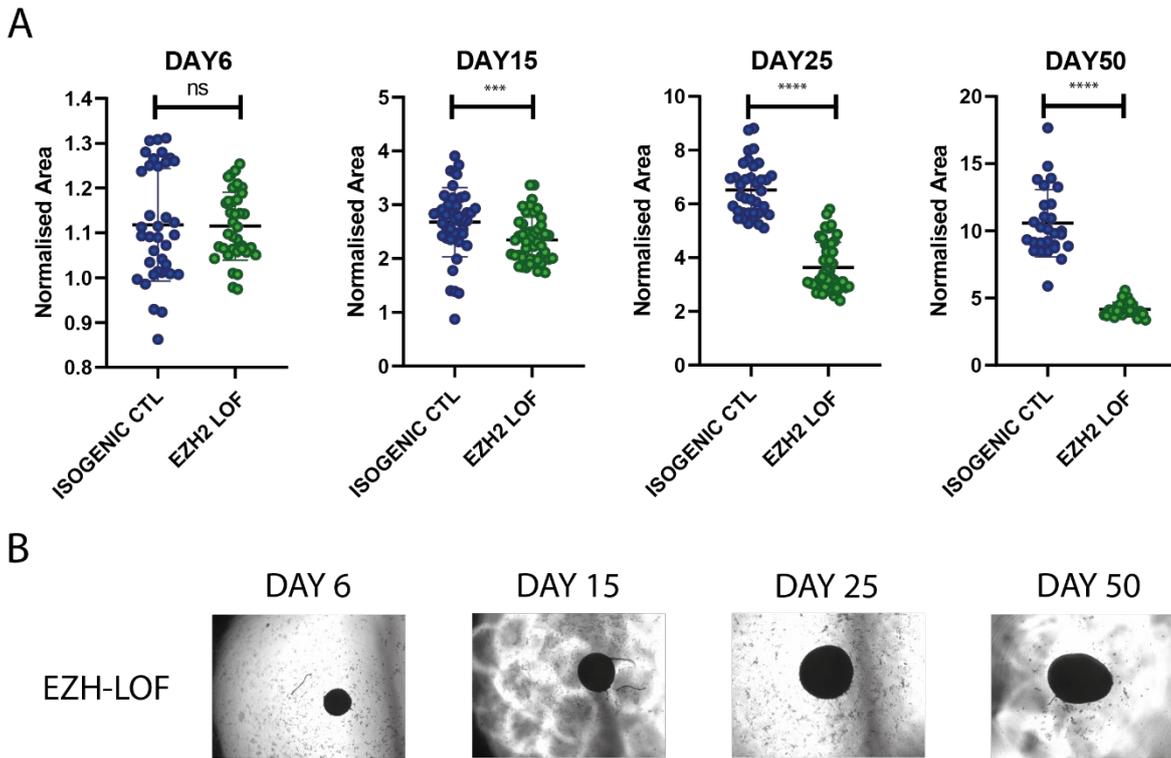


Fig. 86: CBO growth EZH2-LOF vs controls. (A) Overall measurements of the area of control and EZH2 LOF CBO at day 6, 15, 25 and 50 of differentiation grouping all batches of differentiation together. Statistical test: Mann-Whitney; * p value = 0.05; ** p value = 0.01; *** p value = 0.001; **** p value = 0,0001. Day 6: Isogenic Control = 36; EZH2 LOF = 36. Day 15: Isogenic Control = 43; EZH2 LOF = 47. Day 25: Isogenic Control = 36; EZH2 LOF = 44. Day 50: Isogenic Control = 28; EZH2 LOF = 25. Differentiation batches = 4. **(B)** Representative images of the CBO used for the growth curves for the EZH2-LOF line at the time-points analysed.

In summary, these results demonstrated that WS CBO were characterised by differential growth compared to controls, with patient-specific behaviour. Moreover, we demonstrated that severe EZH2-LOF caused reduced CBO growth, in line with the essential role of EZH2 during neurodevelopment^{203, 209} and recapitulating the WS03B and WVS04A CBO growth phenotype.

4.9 WS CBO upregulated neuronal maturity genes in early phases and displayed metabolic alterations at later developmental time-points

Next, we focused on the characterisation of the global differences in gene expression in developing WS CBO. To this end, we differentiated CBO from the cohort of 4 controls, 4 patients and the isogenic EZH2-LOF line in 2 independent rounds of differentiation and analysed their bulk transcriptome at 3 critical developmental stages (day 25, 50 and 100), since we demonstrated that the differences observed in bulk transcriptomics were more subtle after day 100 (Fig. 28). The characterisation of a total of 54 individual-organoid-samples, encompassing 9 lines at 3 stages, all in duplicate, allowed us to control for technical variability of organoidogenesis, leading to an increased robustness of the analysis. Stage-wise differential expression analysis of control *versus* WS organoids revealed a group of 514 genes that showed similar levels of expression at day 25 and that are abnormally silenced in WS organoids compared to controls. Specifically, we were able to score an higher global fold change of these genes in WS over time, indicating that they were less or not silenced in this condition (Fig. 87A). Intersection with publicly available EZH2 ChIP-seq data from ESCs highlighted 27 *bona fide* targets of EZH2 generally associated with development (regionalisation and pattern specification process), pointing at a broad dysregulation of gene silencing in WS that could underlie the WS characteristic intellectual disability phenotype (Fig. 87B and C).

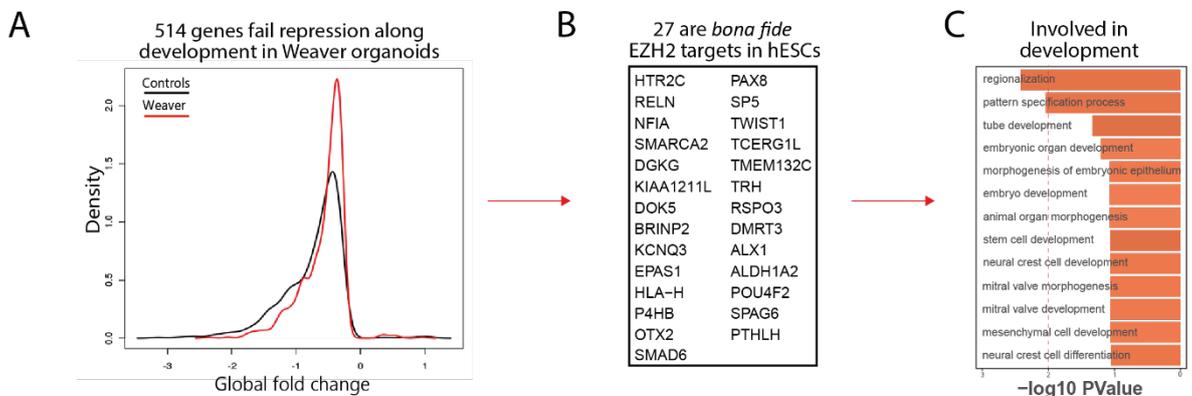


Fig. 87: Stage-wise DEA WS vs controls. (A) Stagewise differential expression analysis of control and WS CBO showing a group of genes (514) expressed at the same levels at day 25 in both genotypes that fail correct silencing at later stages in WS CBO. The graph displays the global fold change of this group of genes in both control and WS CBO. Global fold change refers to the fold changes across stages of differentiation and not between genotypes (the analysis was performed stage-wise in controls and Weavers separately). (B) List of the 27 EZH2 targets in ESCs among the 514 genes in 87A. (C) GO for biological processes of the 27 genes in 87B showing the involvement of those genes in general development. The bar plot displays $-\log_{10}P$ Value on the X-axis and $-\log_{10}P$ Value greater than 2 (P Value < 0.01) was considered significant (blue-scale = categories enriched for downregulated genes; red-scale = categories enriched for upregulated genes).

To dissect these results, we performed differential expression analysis at each CBO developmental stage independently, with the aim of revealing stage-specific transcriptomic dysregulation of cortical neuron differentiation in WS. GO analysis for biological processes at day 25 showed upregulation of genes involved in neuronal maturation, in particular of axonal fasciculation, and

downregulation of the unfolded protein response, as corroborated also by Gene Set Enrichment Analysis (GSEA) (Fig. 88A and B).

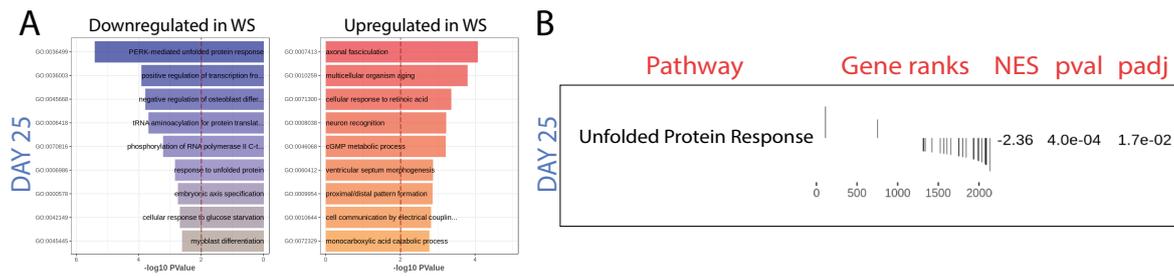


Fig. 88: Transcriptomic analysis WS vs controls at day 25. (A) GO for biological processes of the DEGs with FDR < 0.05 for the comparison WS and Control CBO at day 25. The bar plot displays $-\log_{10}P$ Value on the X-axis and $-\log_{10}P$ Value greater than 2 (PValue < 0.01) was considered significant (blue-scale = categories enriched for downregulated genes; red-scale = categories enriched for upregulated genes). **(B)** Results of the Gene Set Enrichment analysis performed on DEGs with PValue < 0.05 showing: pathway, gene ranks, enrichment score (NES), PValue and PValue adjusted (significant when PValue adjusted < 0.05).

Interestingly, the upregulation in WS of genes involved in cortical neuron maturation remained conserved also after 50 days of differentiation, pointing at a developmental transcriptional acceleration in WS CBO that may be partially dependent on the dysregulation of neuronal migration and cell-cell adhesion that we also observed at this stage (Fig. 89A). GSEA at day 50 revealed downregulation in WS of pathways related to proliferation (G2M checkpoint, MTORC1 signaling and E2F targets) and migration (MTORC1 signaling) (Fig. 89B). Importantly, RELN, an important mediator of the migration of maturing cortical neurons through the RG that was also linked to polymicrogyria (one of the key features of WS³⁶⁸, resulted upregulated WS CBO at day 50 of differentiation.

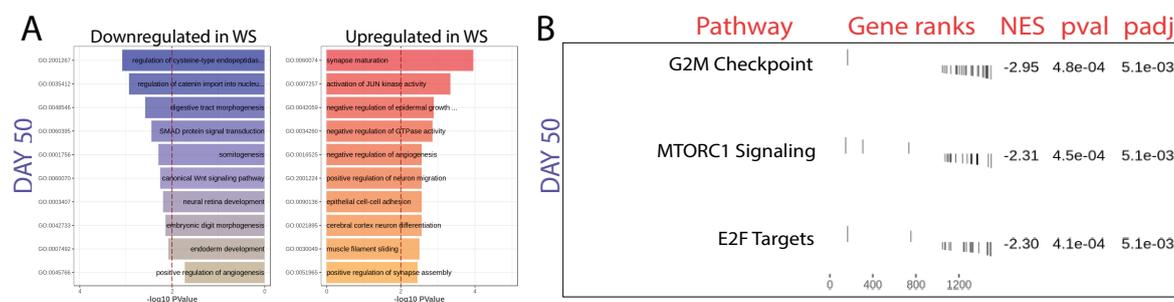


Fig. 89: Transcriptomic analysis WS vs controls at day 50. (A) GO for biological processes of the DEGs with FDR < 0.1 for the comparison WS and Control CBO at day 50. The bar plot displays $-\log_{10}P$ Value on the X-axis and $-\log_{10}P$ Value greater than 2 (PValue < 0.01) was considered significant (blue-scale = categories enriched for downregulated genes; red-scale = categories enriched for upregulated genes). **(B)** Results of the Gene Set Enrichment analysis performed on DEGs with PValue < 0.05 showing: pathway, gene ranks, enrichment score (NES), PValue and PValue adjusted (significant when PValue adjusted < 0.05).

Characterisation of day 100 CBO transcriptomes revealed mild upregulation in WS CBO of 6 genes (GAPDH, ALDOA, ALDOC, PGK1, TPI and PKM) coding for enzymes that catalyse irreversible reaction of the glycolytic pathway, indicating an aberrant increase in glucose metabolism (Fig. 90A, C and D). Further confirmation of this observation was achieved by measuring mRNA levels of GAPDH and

WS and EZH2LOF clustering apart from all other samples (**Fig. 91A**). WVS03B instead showed the less reliable behaviour among replicates, with one of them clustering with controls and the other completely apart from the other samples. We performed DEA at day 50, the last stage where we saw differences in neuronal maturation signatures, and we selected DEGs for the comparison of classical WS *versus* controls. When considering the DEGs with FDR < 0.01 and absolute FC > 2, we observed that DEGs behaved similarly in classical WS and EZH2-LOF, while they were more similar between controls and WVS03A (**Fig. 91B**)

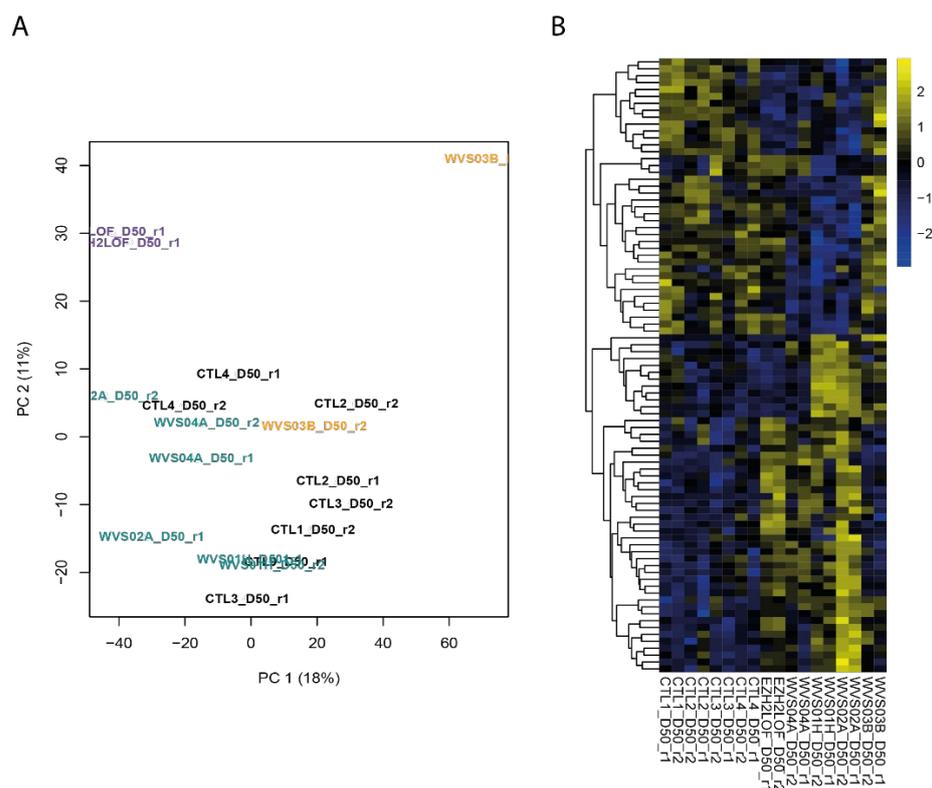


Fig. 91: PCA and DEA subdividing WS patients in groups. (A) PCA showing the PC1 and PC2 of the WS CBO cohort at day 50 of differentiation. Labels indicate the code of the individual. Green = classic WS patients; Orange = putative gain of function; Violet = EZH2-LOF. **(B)** Heatmap showing the behaviour of DEGs with FDR < 0.01 and absolute FC > 2 in all day 50 samples. Values represented the Z-score for every gene (row) in all samples (column). The colour-code for stage is depicted in the figure legend.

GO analysis for BP for the comparison classic WS vs controls revealed again upregulation of genes related to neuronal differentiation and morphogenesis, thus confirming the already observed transcriptional phenotype also with this approach (**Fig. 92A**). Reassuringly, master regulator analysis of DEGs showed significant enrichment for the PRC2 subunits EZH2 and SUZ12 (**Fig. 92B**)

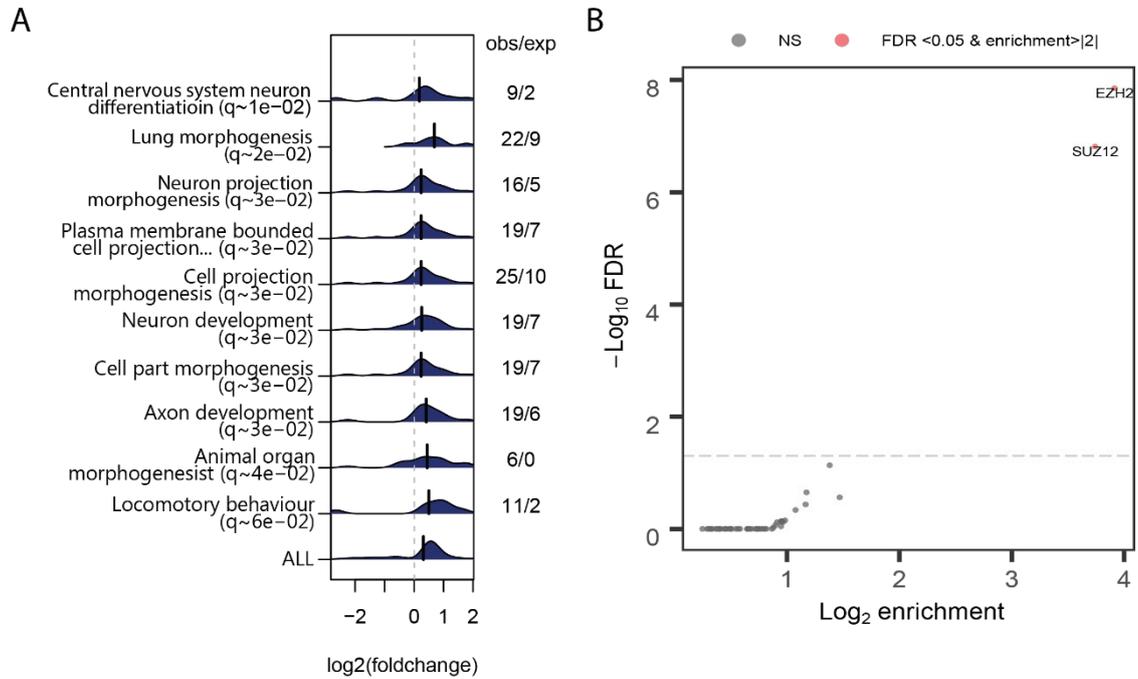


Fig. 92: Functional analysis of DEGs subdividing WS patients in groups. (A) GO analysis performed on DEGs with FDR < 0.01 and absolute FC > 2 showing the top 10 categories with FDR < 0.05 and enrichment > 2 coupled with the distribution of classic WS fold changes for the DEGs of a given category. ALL represents the fold change distribution of the totality of DEGs. The enrichment was defined as the ration between the number of DEGs belonging to a given category and the number of genes expected to be found by chance (obs/exp). **(B)** Master regulator analysis performed on DEGs with FDR < 0.01 and absolute FC > 2. Significant master regulators were considered the ones with FDR < 0.05 ($-\log_{10}\text{FDR} = 1.3$) and enrichment greater than 2. The enrichment was defined as the ration between the number of target of a given master regulator among DEGs and the number of its target expected to be found by chance among DEGs.

Finally, we performed transcriptomic analysis of EZH2-LOF CBO against its isogenic control. This analysis revealed already at day 25 a robust downregulation of general signatures of forebrain regionalisation and cerebral cortex maturation, confirming a general detrimental effect of severe EZH2 LOF in CBO differentiation and in line also with the marked reduced CBO growth derived from this line (**Fig. 93**).

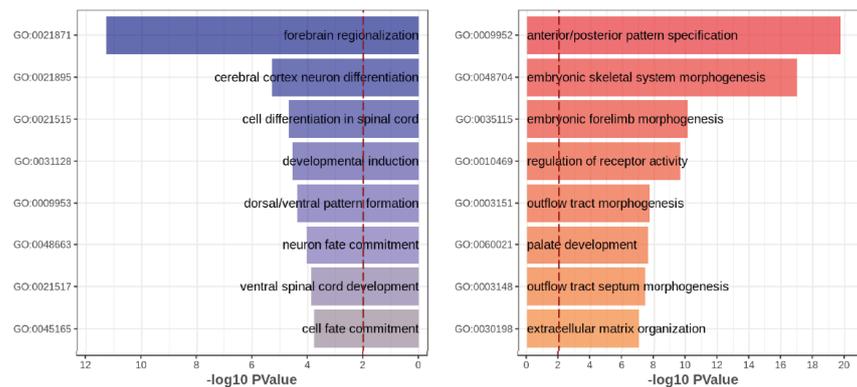
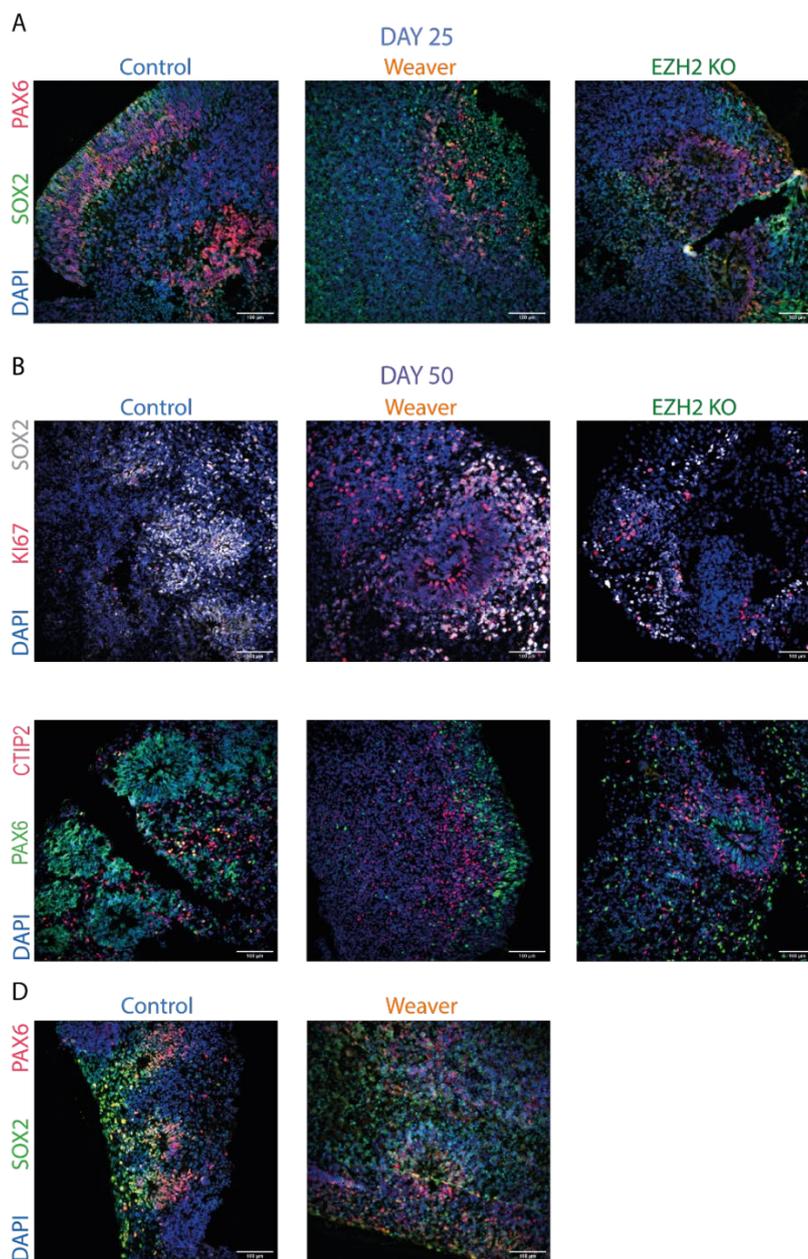


Fig. 93: Transcriptomic analysis EZH-LOF vs controls at day 25. GO for biological processes of day 25 DEGs with absolute FC > 2 and FDR < 0.05 for the comparison EZH2-LOF vs isogenic control CBO. The bar plot displays $-\log_{10}\text{PValue}$ on the X-axis and $-\log_{10}\text{PValue}$ greater than 2 (PValue < 0.01) was considered significant (blue-scale = categories enriched for downregulated genes; red-scale = categories enriched for upregulated genes).

Immunofluorescence analyses of organoids produced according to the protocol previously described, reflected the organisation of the developing cortex with the classical structure of the

ventricles from which radially oriented neuronal progenitors organise, as previously observed in our laboratory. Interestingly, ventricular-like structures were also observed in EZH2-LOF CBO (**Fig. 94**). Given the transcriptomic results pointing at a possible precocious neuronal differentiation in WS and the reported evidence that EZH2 ablation in mouse cortical progenitors causes an unbalance between self-renewal and differentiation towards differentiation²¹⁰, we performed immunofluorescence analyses for markers of apical progenitors and neuronal maturation in WS, control and EZH2-LOF CBO (**Fig. 94 and 95**).



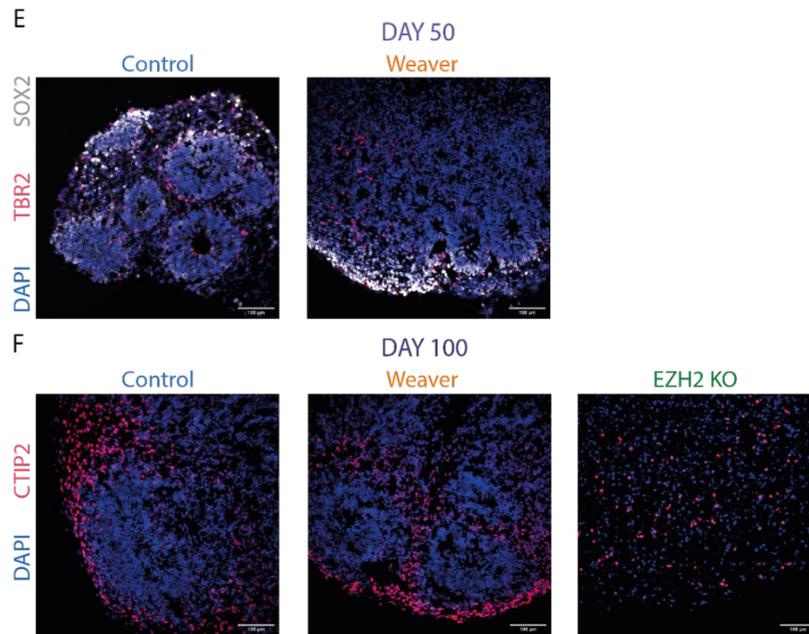


Fig. 94: Representative CBO immunostainings. (A) Representative widefield images showing the expression of PAX6 and SOX2 in day 25 control, patient and EZH2 LOF CBO. (B) Representative widefield images showing the expression of KI67 and SOX2 in day 50 control, patient and EZH2 LOF CBO. (C) Representative widefield images showing the expression of PAX6 and CTIP2 in day 50 control, patient and EZH2 LOF CBO. (D) Representative widefield images showing the expression of PAX6 and SOX2 in day 50 control and patient CBO. (E) Representative widefield images showing the expression of TRB2 and SOX2 in day 50 control and patient CBO. (F) Representative widefield images showing the expression of CTIP2 in day 100 control, patient and EZH2 LOF CBO. All images were acquired on a widefield microscope (Leica DMI6 B), equipped with an Andor Zyla (VSC-04470 sCMOS), using 20X/0.75 dry objective.

By measuring PAX6 levels at the three stages analysed, we found significant higher proportions of aRG in WS and EZH2-LOF CBO. Interestingly, when looking at the proportion of aRG in more advanced time-points, we found that its levels at day 50 were unchanged between genotypes and then significantly increased again in day 100 WS CBO (Fig. 95A). The aberrantly dynamic regulation of this marker along WS CBO development led us to investigate whether the neuronal output was also changed. We then measured the levels of CTIP2 (day 50 and 100) and SATB2 (day 100) as markers of lower and upper layer neurons, respectively (Fig. 95B and C). We did not find statistically significant differences between groups, although there was a clear trend of increased proportion of CTIP2 positive cells at day 50 for both WS and EZH2-LOF CBO to be confirmed with further experiments.

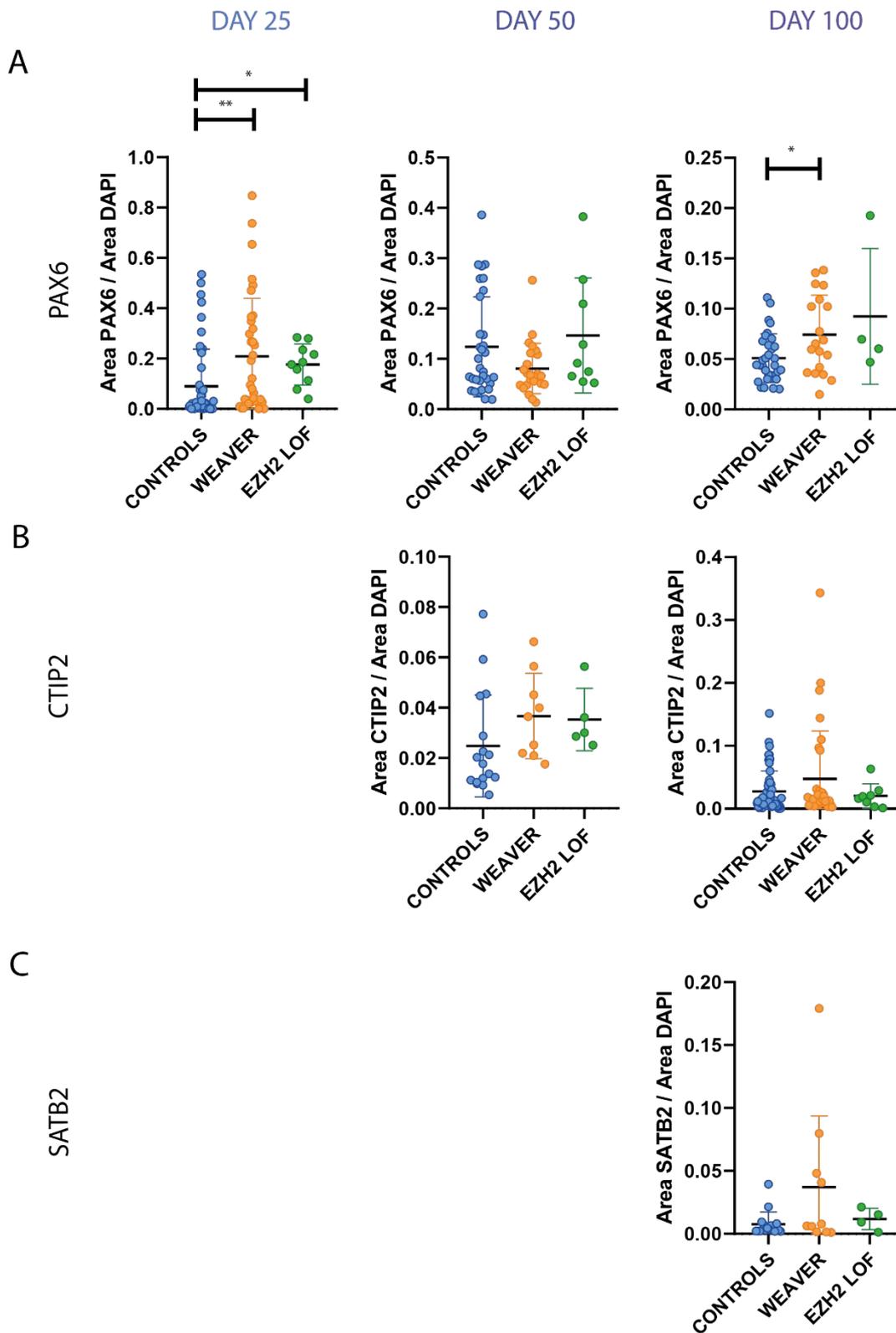


Fig. 95: Controls, WS and EZH2-LOF immunostainings quantification. (A) Quantification of the fluorescent signal for the apical progenitors marker PAX6 calculated as ratio between the area occupied by the PAX6 signal and the area occupied by DAPI at the three time-points analysed in patient and EZH2 LOF CBO. **(B)** Quantification of the fluorescent signal for the lower layer neuron marker CTIP2 calculated as ratio between the area occupied by the CTIP2 signal and the area occupied by DAPI in day 50 and 100 control, patient and EZH2 LOF CBO. **(C)** Quantification of the fluorescent signal for the upper layer neuron marker SATB2 calculated as ratio between the area occupied by the SATB2 signal and the area occupied by DAPI in day 100 control, patient and EZH2 LOF CBO. Statistical test: Mann-Whitney; * p value = 0.05; ** p value = 0.01; *** p value = 0.001; **** p value = 0.0001.

Overall, WS CBO showed upregulation of genes related to neuronal maturation at day 25 and day 50 of differentiation, with altered expression of neuronal migration related genes at day 50. Conversely, day 100 CBO showed upregulation of a glycolytic stress signature probably independent from differential CBO size between patients and control. Transcriptomic analysis of EZH2-LOF CBO showed that severe EZH2 LOF caused downregulation of gene signatures related to forebrain development, confirming the EZH2 essential role in controlling gene expression during human corticogenesis. Finally, by quantifying the cell type population proportions in CBO, we found dynamic regulation of aRG cell proportions between WS and control CBO, with an increase in the patient group at day 25 and 100. This regulation was mirrored in CBO with severe EZH2 LOF at day 25 but not at day 100. This dynamic regulation did not reflect on the neuronal output, although a clear trend of CTIP2 upregulation was observed at day 50 for WS and EZH2-LOF CBO.

4.10 Integration of transcriptomic and epigenomic WS datasets uncovered a set of genes directly linked to EZH2 LOF relevant for the WS-associated intellectual disability

To identify transcriptomic deregulations directly related to the EZH2 LOF in WS, we intersected the lists of DEGs of the WS (as a unique group and excluding EZH2-LOF) *versus* controls comparison at the three developmental stages analysed with the respective lists of DEGs of EZH2-LOF *versus* isogenic control (**Fig. 96**). By plotting the fold changes of the genes included in the two lists for every stage, we were able to find a subset of them that followed the same expression pattern in both frameshift and WS CBO as well as a subset that showed opposite expression, with a global overlap of 17.3%, 28.3% and 25.5% in day 25, 50 and 100, respectively (**Fig. 96**).

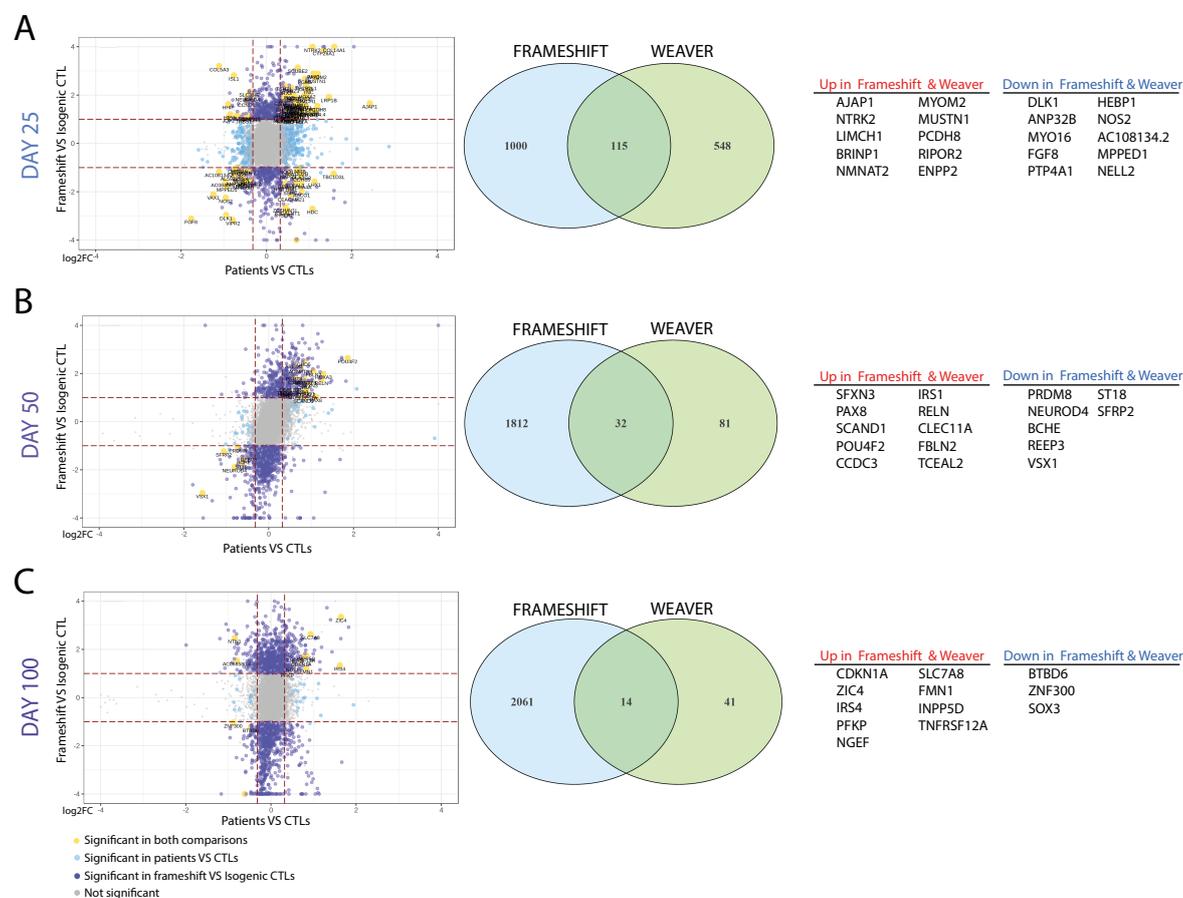


Fig. 96: Intersection DEA WS vs controls and EZH2-LOF vs isogenic control. (A) Scatter plot showing the log₂FC of DEGs for the comparisons WS vs controls (X-axis) against and EZH2-LOF vs isogenic control (Y-axis) at day 25. DEGs specific for one of the two comparisons or common to both were highlighted according to the colour-code depicted in the figure legend. Venn diagram representing the size of the overlap between the two DEA at day 25. Top significant upregulated or downregulated DEGs in WS CBO found differentially expressed and following the same direction of dysregulation also in EZH2-LOF CBO at day 25. Frameshift = EZH2-LOF. **(B)** Scatter plot showing the log₂FC of DEGs for the comparisons WS vs controls (X-axis) against and EZH2-LOF vs isogenic control (Y-axis) at day 50. DEGs specific for one of the two comparisons or common to both were highlighted according to the colour-code depicted in the figure legend. Venn diagram representing the size of the overlap between the two DEA at day 50. Top significant upregulated or downregulated DEGs in WS CBO found differentially expressed and following the same direction of dysregulation also in EZH2-LOF CBO at day 50. Frameshift = EZH2-LOF. **(C)** Scatter plot showing the log₂FC of DEGs for the comparisons WS vs controls (X-axis) against and EZH2-LOF vs isogenic control (Y-axis) at day 100. DEGs specific for one of the two comparisons or common to both were highlighted according to the colour-code depicted in the figure legend. Venn diagram representing the size of the overlap between the two DEA at day 100. Top significant upregulated or downregulated DEGs

in WS CBO found differentially expressed and following the same direction of dysregulation also in EZH2-LOF CBO at day 100. Frameshift = EZH2-LOF.

With a close look at the top 10 genes upregulated or downregulated in both condition at all the stages, we found 3 *bona fide* EZH2 target that failed downregulation in WS CBO in the stage-wise analysis (**Fig. 87**) to be upregulated at day 50 (namely RELN, PAX8 and POU4F2), confirming that they were directly regulated by PRC2. Of note, copy number variation of the gene AJAP1³⁶⁹, one of the most significantly and strongly upregulated genes in both conditions at day 25 involved in adherens junctions formation and cell migration, is associated with intellectual disability and craniofacial abnormalities³⁷⁰. Our findings demonstrated that severe EZH2 LOF in CBO led to impaired human corticogenesis and revealed partial overlap with the transcriptional dysregulation observed in WS, with EZH2 target related to neuronal differentiation and migration upregulated.

To get further insights into the direct EZH2-dependent transcriptomic dysregulation in the context WS corticogenesis we performed H3K27me3 ChIP-seq at the three stages of differentiation. We considered the H3K27me3 histone mark lost or gained in WS when a certain peak was present in at least 3 out of 4 control and in at the most 1 Weaver out of 4 and *vice versa*, respectively. This analysis allowed us to score major differences in H3K27me3 between WS and controls, highlighting genes that lost or gained the histone mark at their promoter or at associated enhancers (publicly available data for hippocampal enhancers were retrieved from the 4D nucleome project). Before proceeding with the analysis of WS-specific differences, we decided to validate our experiments by intersecting significant peaks at every stage for control and WS with publicly available histone mark ChIP-seq datasets from hESC, hESC-derived neuronal progenitors and neurons, and hippocampal neurons. As expected, we found the greatest overlap with H3K27me3 peaks for all cell types analysed and we also observed that our H3K27me3 peaks were overlapping more with neuronal progenitor and neuron profiles than hESC, thus confirming the fidelity of our dataset (**Fig. 97**).

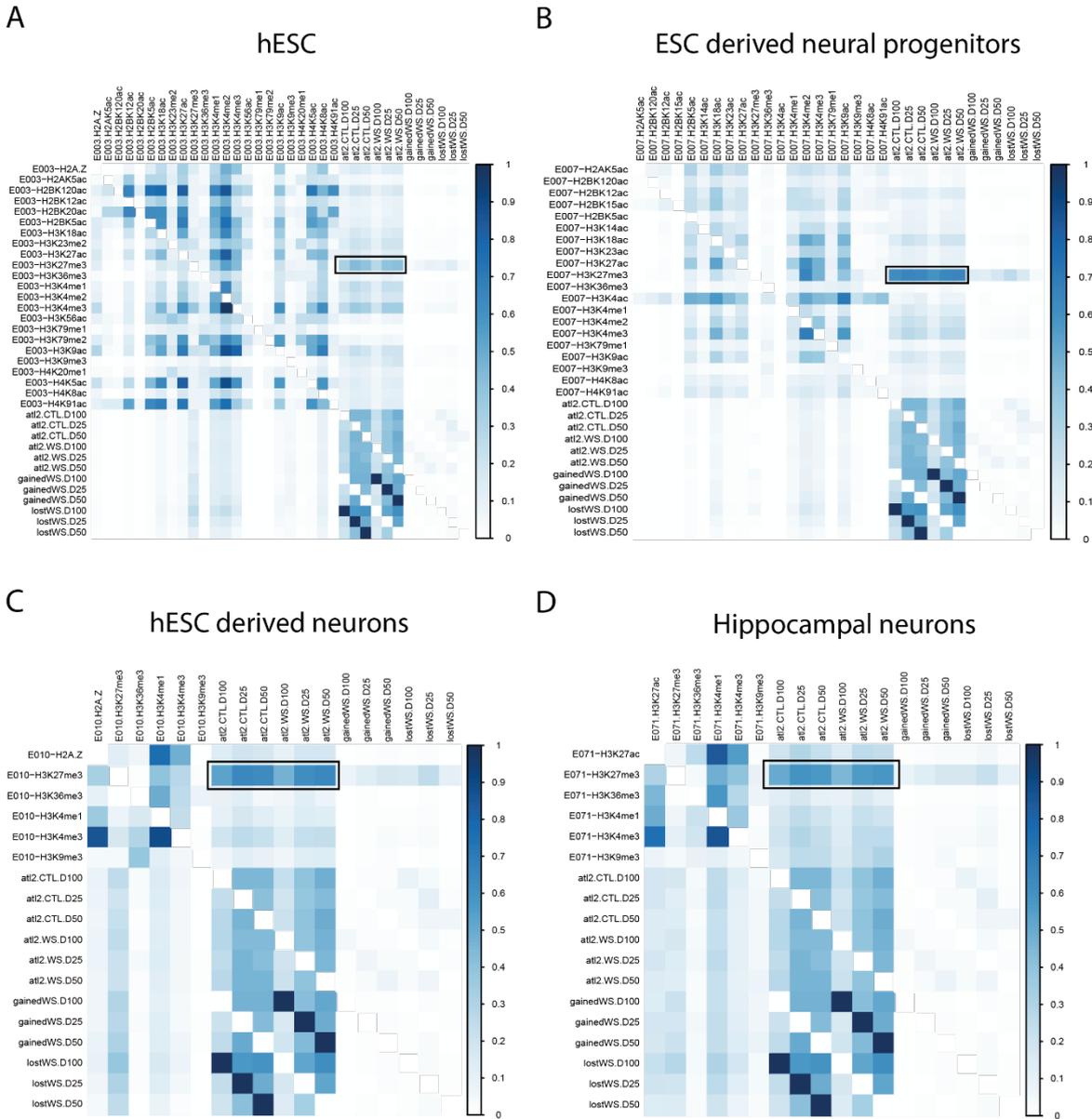


Fig. 97: Overlap with roadmap epigenomic H3K27me3 datasets. Overlaps of peaks for main regulatory histone marks with FDR < 0.1 retrieved from Roadmap Epigenomic ChIP-seq datasets generated from hESC (A), hESC derived neuronal progenitors (B) and neurons (C) or hippocampal neurons (D) vis a vis H3K27me3 CBO peaks in controls and Weaver at day 25, 50 and 100 (FDR < 0.1). The colour-code indicates the fraction of the overlap. At12 = peaks present in at least two samples, gained = peaks specifically gained in WS samples, lost = peaks specifically lost in WS samples.

We analysed the distribution of H3K27me3 at regulatory elements to understand whether WS peaks showed a peculiar behaviour compared to controls. Interestingly, both WS and control samples had similar distribution of H3K27me3, with no difference in the proportion of peaks falling in different regulatory regions. However, when considering peaks specifically lost or gained in WS, we observed a significant enrichment at enhancers, suggesting that WS-specific epigenetic dysregulations were mainly localised at these regions (Fig. 98).

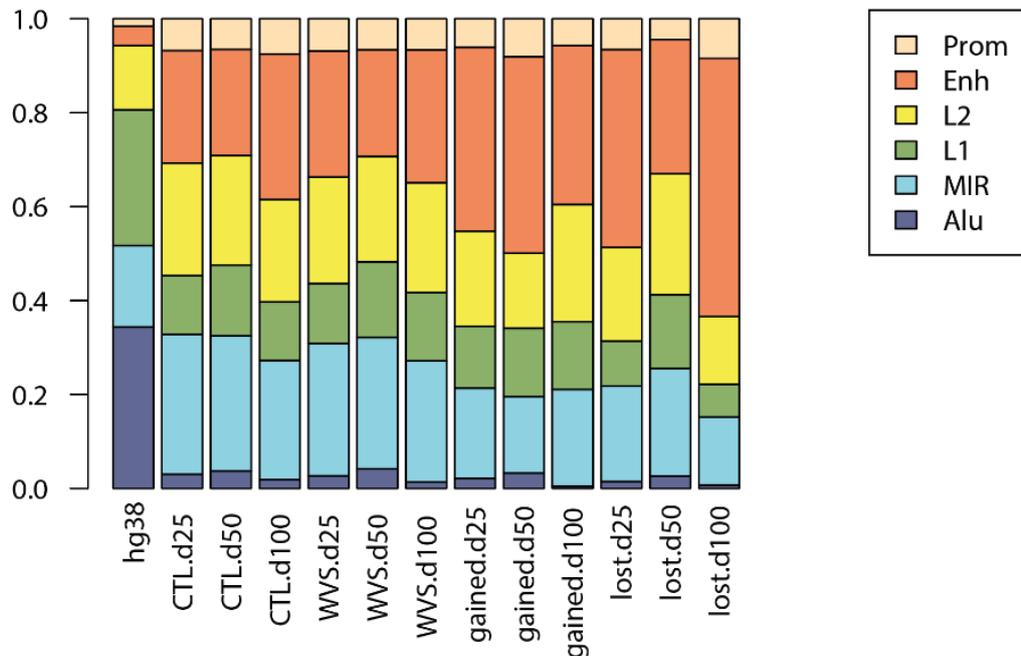


Fig. 98: Distribution of H3K27me3 peaks at regulatory regions. Distribution of H3K27me3 peaks at regulatory regions. hg38 represents the proportion of the analysed regions in the human genome. “CTL” and “WS” bars represent the distribution of the totality of significant peaks (FDR < 0.1) at a given stage. “Gained” bars represent the distribution of significant gained peaks (FDR < 0.1) in WS at a given stage. “Lost” bars represent the distribution of significant gained peaks (FDR < 0.1) in WS at a given stage. Significance was tested using the non-parametric Kolmogorov Smirnov Test. PValue of gained day 25 vs controls day 25 < 0.05; PValue of gained day 50 vs controls day 50 < 0.01; PValue of gained day 100 vs controls day 100 < 0.01; PValue of lost day 25 controls day 25 < 0.01; PValue of lost day 50 controls day 50 < 0.01; PValue of lost day 100 controls day 100 < 0.01.

We then intersected DEGs from the analysis considering WS as a unique group and excluding EZH-LOF with genes that lost H3K27me3 at promoters, enhancers or both, again to highlight WS-specific features in line with a EZH2 LOF and because we did not have sufficient sample size to perform the qualitative analysis of H3K27me3 distribution considering classical WS, the atypical WS and EZH2-LOF as different groups. Overall, the intersections ranged from 3.7% to 50% with a general reduced number of total overlapping genes along CBO maturation, probably due to a general reduction of the number of DEGs throughout differentiation (**Fig. 99**). Moreover, the highest overlaps were observed at enhancers, in line with our results on the distribution of lost and gained peaks in WS (**Fig. 98**).

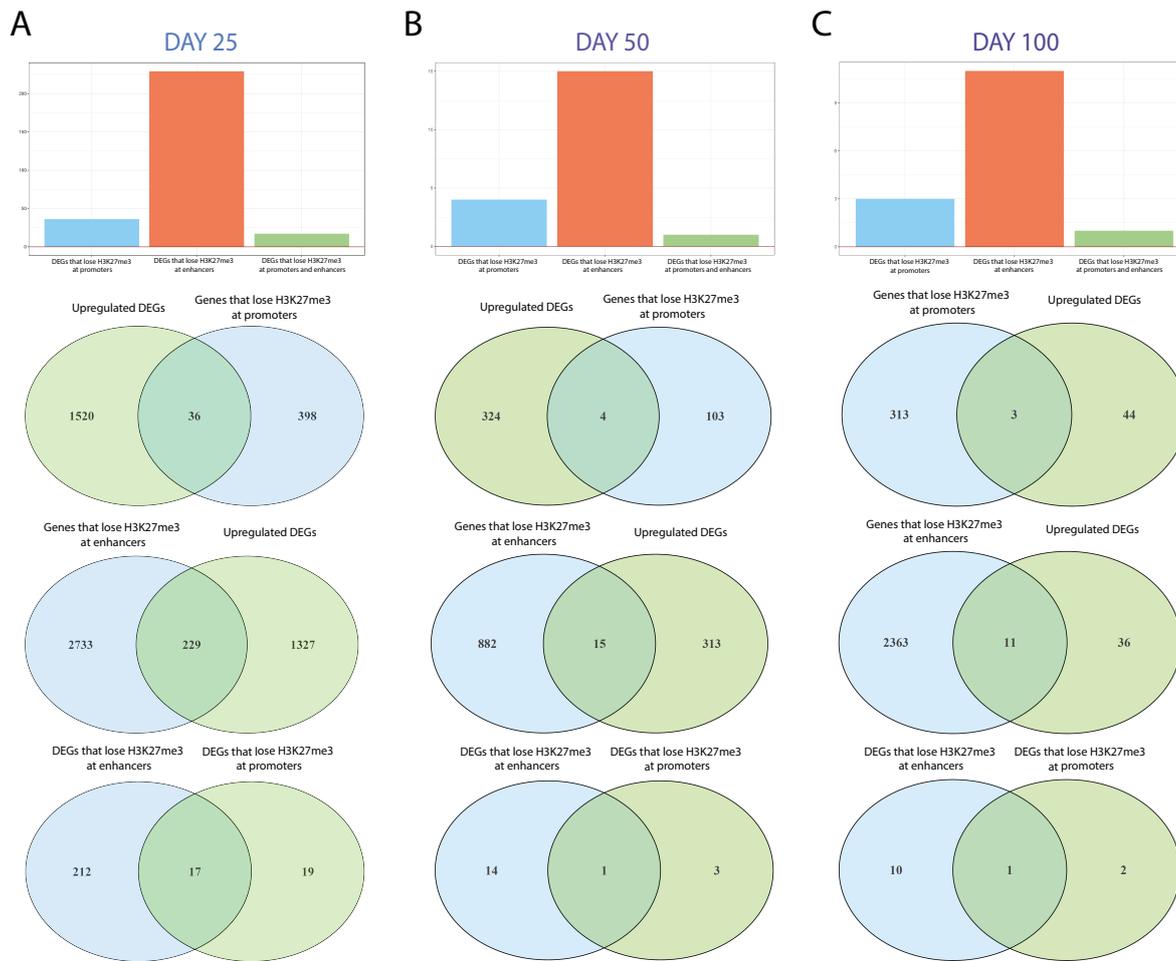


Fig. 99: Intersection DEAs and ChIPseq peaks lost in WS. (A) Barplot representing the number of DEGs with FDR 0.25 that lost H3K27me3 at promoters, enhancers or both at day 25. Venn diagrams representing the same intersections showing the size of the overlaps at day 25. **(B)** Barplot representing the number of DEGs with FDR 0.25 that lost H3K27me3 at promoters, enhancers or both at day 50. Venn diagrams representing the same intersections showing the size of the overlaps at day 50. **(C)** Barplot representing the number of DEGs with FDR 0.25 that lost H3K27me3 at promoters, enhancers or both at day 100. Venn diagrams representing the same intersections showing the size of the overlaps at day 100.

Despite this, we found concordance with the top upregulated genes in both WS and EZH2-LOF CBO at the three stages of differentiation (**Fig. 96**) and the genes that lost H3K27me3 at promoters or enhancers. AJAP1 and LIMCH1 lost H3K27me3 at enhancers at day 25 and resulted upregulated at the same stage in WS and frameshift CBO. Other genes, SFNXN3 and PAX8, lost this histone mark at enhancer and promoter and at the promoter, respectively, at day 25, while being upregulated only at day 50. KCNQ3 lost H3K27me3 at enhancers and was upregulated at day 50; this gene, together with PAX8, was also part of the 27 genes that failed downregulation in WS stage-wise (**Fig. 87**). CDKN1A, a known target of EZH2 which dysregulation leads to impaired neuroblasts differentiation in chicken embryos²¹⁵, INPP5D, FMN1 and NGEF lost H3K27me3 at enhancers at day 100 and were found upregulated at the same stage of differentiation in both WS and frameshift. These results demonstrated that the dysregulation in gene expression observed for this set of genes directly depended from EZH2 LOF in the context of WS corticogenesis.

Given the known interplay between PcG proteins and DNA methylation^{180,199}, we then profiled the methylome of WS and CBO organoids up to day 200 (until day 100 for WS CBO and up to day 200 for control CBO). Principal component analysis revealed different global methylation profiles of CBO at the different stages analysed, with more variance observed between day 25-50 and later stages of differentiation (**Fig. 100**), as opposed to the transcriptome in which less variability among stages is observed after 100 days (**Fig. 26**). In this analysis we also included NCSC and iN. As expected, iPSCs and NCSC clustered apart from iN and CBO. Of note, iN positioned in between iPSCs and day 25 CBO in PC1, possibly indicating that the developmental changes in DNA methylation occurring in corticogenesis were not acquired in this model probably due to the compressed time-frame of differentiation (**Fig. 100**)

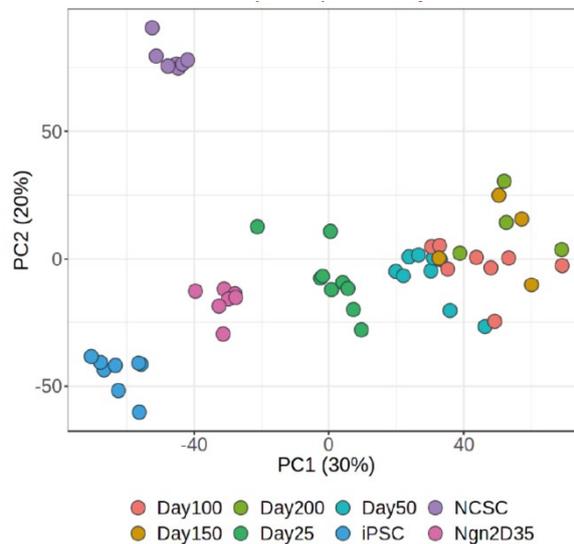


Fig. 100: PCA of DNA methylation profiles of the WS cohort in hiPSC, CBO, NCSC and iN. PCA including the DNA methylation profiles of WS and control iPSCs, NCSCs, iN, CBO at day 25, 50, 100 and control CBO at day 150 and 200. PC1 was driven by progression of CBO stage, with iN (Ngn2) that positioned between hiPSC and CBO at day 25. PC2 was mainly driven by the differences between NCSC and the rest of the samples. Colour-code shown in the figure legend.

We then looked at the methylation profile of the set of genes upregulated in frameshift and patients that lost H3K27me3 at promoter or enhancer and observed, at day 25, that AJAP1 showed a strong reduction in DNA methylation at the promoter in WS, compatibly with its increased RNA expression (**Fig. 101**). The same observation, with less strong reduction in DNA methylation, applies to PAX8 at day 25 and to CDKN1A at day 100 (**Fig. 101**).

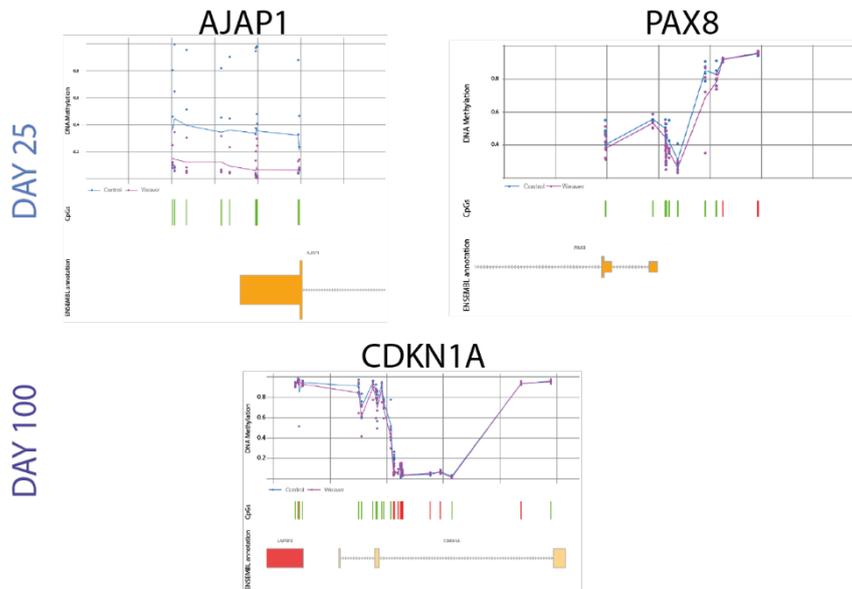


Fig. 101: DNA methylation profiles at the promoter of relevant targets. DNA methylation levels at the promoters of relevant genes emerged from the transcriptomic and H3K27me3 ChIP-seq analyses. Normalised methylation levels range from 0 to 1, FDR < 0.05. Green bars represent differentially methylated CpG while red bars represent not differentially methylated CpG.

In conclusion, the integration of transcriptomic and epigenomic data from patient-derived CBO revealed high-confidence molecular intermediates between genetic lesions in PRC2 and the neurodevelopmental phenotype characteristic of WS.

4.11 2D-induced neurons did not reveal differences between patients and controls

In parallel to CBO differentiation, we aimed at infer disease-relevant alterations in mature cortical neurons. For this reason, we used a PiggyBac-based overexpression system of the master regulator of glutamatergic neurons differentiation Ngn2 to differentiate iN in 35 days of differentiation (**Fig. 102**).

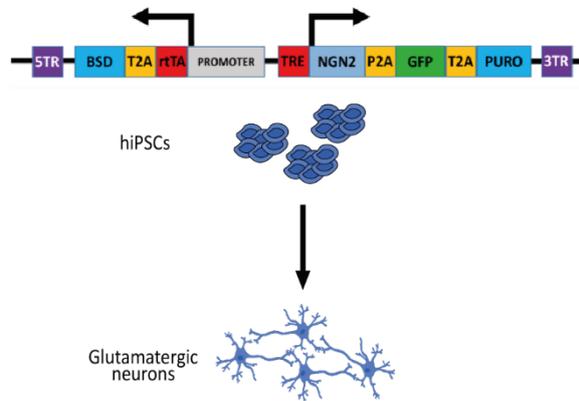


Fig. 102: PiggyBac strategy used to overexpress Ngn2 for iN induction.

We profiled the transcriptome of iN from both Weaver and controls at day 35. We found extremely few DEGs with $FDR < 0.05$, highlighting very subtle differences between the two groups. Despite this, when we looked at DEGs with $PValue < 0.01$ we found that this group of gene distinguished well patients from controls (**Fig. 103**).

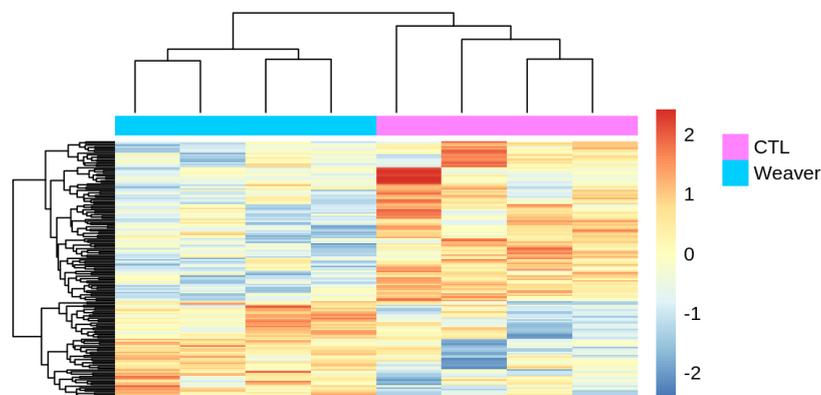


Fig. 103: Heatmap of WS vs control iN DEGs. Heatmap showing the behaviour of DEGs with $PValue < 0.01$ in day 35 WS vs control iN. Values represented the Z-score for every gene (row) in all samples (column).

Finally, measurement of the intrinsic electrical properties of iN at the same stage did not reveal differences in the action potential number as well as in iN membrane properties (**Fig. 104**).

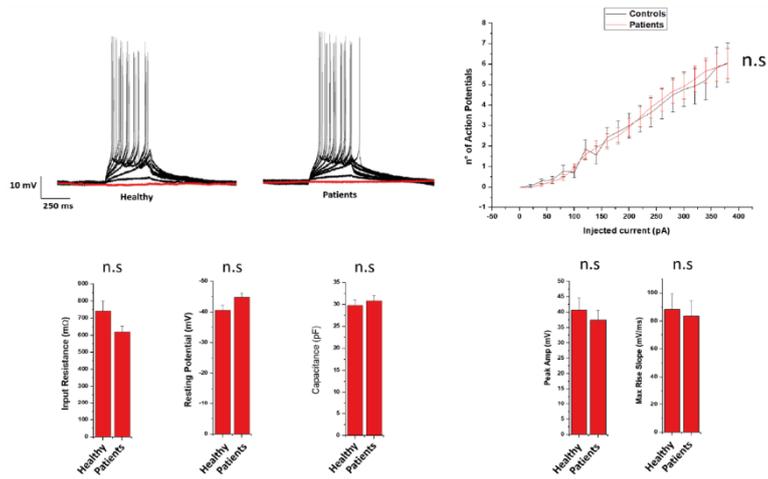


Fig. 104: Electrophysiological characterisation of iN. WS neurons do not show differences in their intrinsic membrane properties when compared to control neurons.

In conclusion, these results suggested that accelerated differentiation of neurons from hiPSC could mask important disease-relevant molecular phenotypes that were instead preserved in CBO, where the timing of the different phases of corticogenesis was maintained.

4.12 Patient-derived neural crest stem cells show downregulation of migration

The second main disease phenotype associated to WS is represented by severe craniofacial abnormalities²²³. In order to investigate the molecular mechanisms underlying these WS clinical features, we reproduced a well-established NCSC differentiation protocol³⁷¹. We confirmed the NCSC identity by immunostaining followed by FACS for anti-NGFR and HNK1 (**Fig. 105A**), as well as by the expression of cardinal cranial NCSC markers in their bulk transcriptomes (**Fig. 105B**), confirming the suitability of this model for the study of the craniofacial phenotype characteristic of WS.

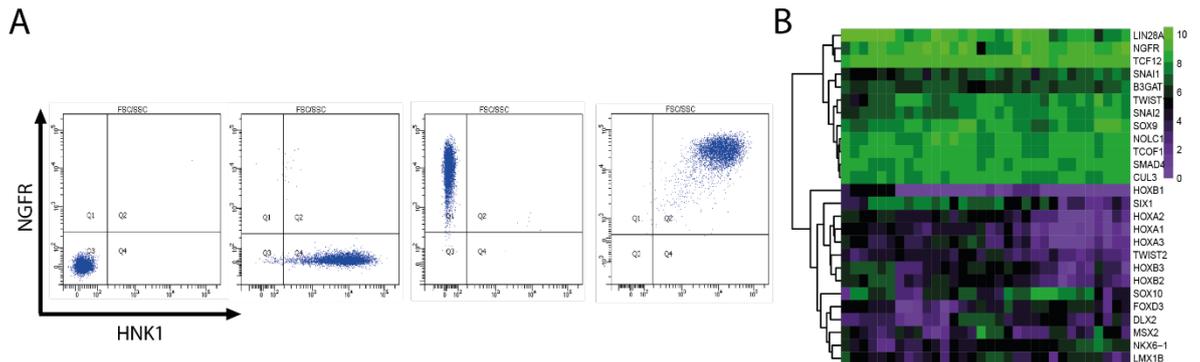


Fig. 105: Validation of NCSC differentiation. (A) Representative FACS analyses showing the expression of NGFR and HNK1, two cardinal markers of NCSCs differentiation. (B) Heatmap showing the expression of cranial NCSCs markers expressed in normalised log₁₀-transformed read counts.

Transcriptomic analysis of NCSC revealed a strong downregulation of genes involved in neural crest migration (**Fig. 106**), one of the key pathway at the basis of the facial abnormalities characteristic of neurodevelopmental disorders^{372,373}.

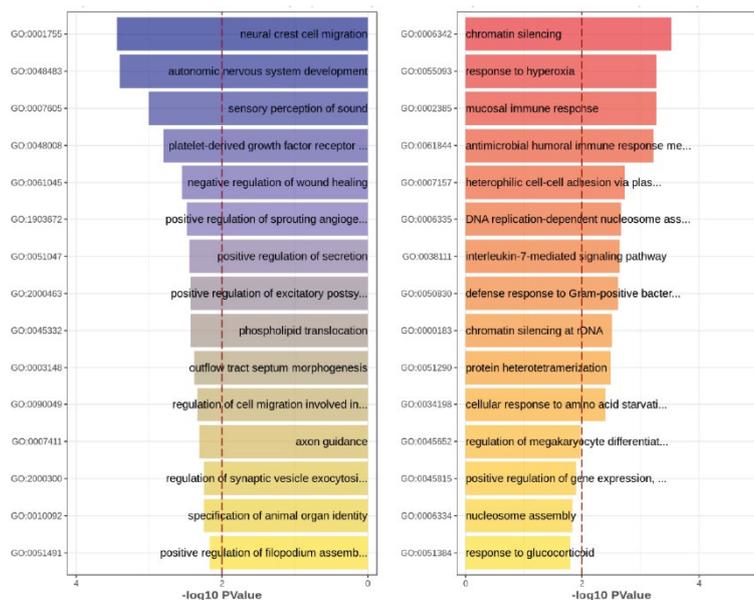


Fig. 106: GO analysis for biological processes of NCSC DEGs WS vs controls. GO for biological processes of DEGs with absolute FC > 1.5 and FDR < 0.05 for the comparison WS vs control NCSC. The bar plot displays -log₁₀PValue on the X-axis and -log₁₀PValue greater than 2 (PValue < 0.01) was considered significant (blue-scale = categories enriched for downregulated genes; red-scale = categories enriched for upregulated genes).

4.13 Isogenic line engineering for validation of WS molecular phenotypes

To validate the phenotypes observed using patient-derived CBO and NCSC, we designed a CRISPR/Cas9 engineering strategy to insert disease-relevant mutations in a control genomic background and, *vice versa*, to correct the disease-causing mutation in patient-lines. These tool had already shown to be essential for the scoring of disease-relevant phenotypes net of the confounding effects due to different genetic backgrounds³⁷⁴. In this work, a proof of principle of the editing strategy efficiency is reported.

We designed a donor plasmid to use as a repair template for the HDR. The construct carried a resistance against hygromycin under the control of hPGK promoter. This region was flanked by two LoxP sites recognisable by the Cre-recombinase to cleave the sequence comprised between them. In this way, once selected the clones edited in the correct way, the resistance cassette could be removed. To mediate HDR, the cassette was flanked by two 100 bp long homology arms (HAs) binding the genomic DNA flanking the region of interest. Finally, we inserted restriction enzyme recognition sites at each end of the HAs to make this construct as more versatile as possible for its use with multiple genomic regions.

Since WVS02A was the only patient-line recapitulating the macrocephaly phenotype in CBO, we decided to engineer the EZH2 mutation of this patient in a control genetic background. HAs were designed in order to direct the insertion of hygromycin into the intron upstream the exon carrying the mutation and the 3' HA carried the disease-causing alteration along with a synonymous change in the PAM sequence to avoid the Cas9-mediated re-editing (**Fig. 107**).

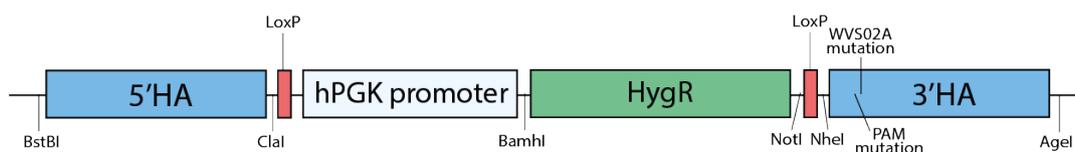


Fig. 107: Schematic representation of the donor construct for WS isogenic line generation.

Resistant clones were then subjected to a series of PCR directed to the edited locus in order to confirm the presence of the cassette and the heterozygosity. First we amplified the resistance cassette (**Fig. 108A**) and then we checked for the integration in the correct genomic locus by using construct-external and -internal primer combinations spanning the whole edited region (**Fig. 108B and C**). Positive clones to both the latter PCRs were then subjected to the heterozygosity test, performed by using primers designed at the extremes of the two HAs (**Fig. 108D**).

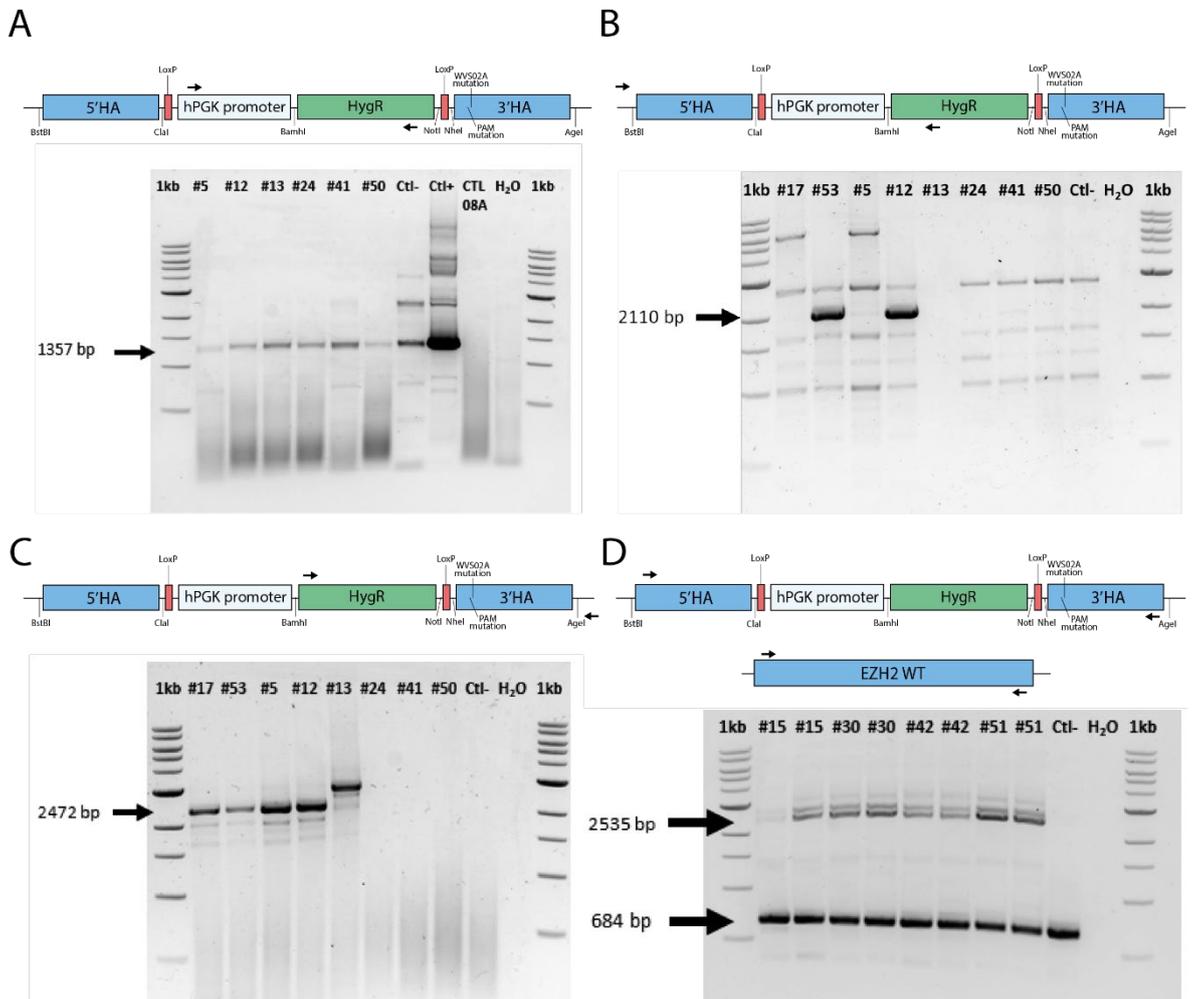


Fig. 108: Screening of CRISPR/Cas9 clones. (A) Scheme of the PCR reaction and a representative gel electrophoresis for the detection of the resistance cassette in the genome. (B) Scheme of the PCR reaction and a representative gel electrophoresis to demonstrate the correct integration at the 5' insertion site. (C) Scheme of the PCR reaction and a representative gel electrophoresis to demonstrate the correct integration at the 3' insertion site. (D) Scheme of the PCR amplification in the mutated and WT alleles and representative gel electrophoresis of clones that integrated the construct in the right position and in heterozygosis.

Finally, we performed Sanger sequencing on the edited and the WT alleles of 5 clones, confirming for all of them the presence and absence of the mutation, respectively (Fig. 109A and B). We then calculated the efficiency of correct genome editing finding that, out of 43 screened clones, 23% carried the mutation in the locus of interest, of which half was confirmed to be correctly heterozygous, as observed in patients (Fig. 109C)

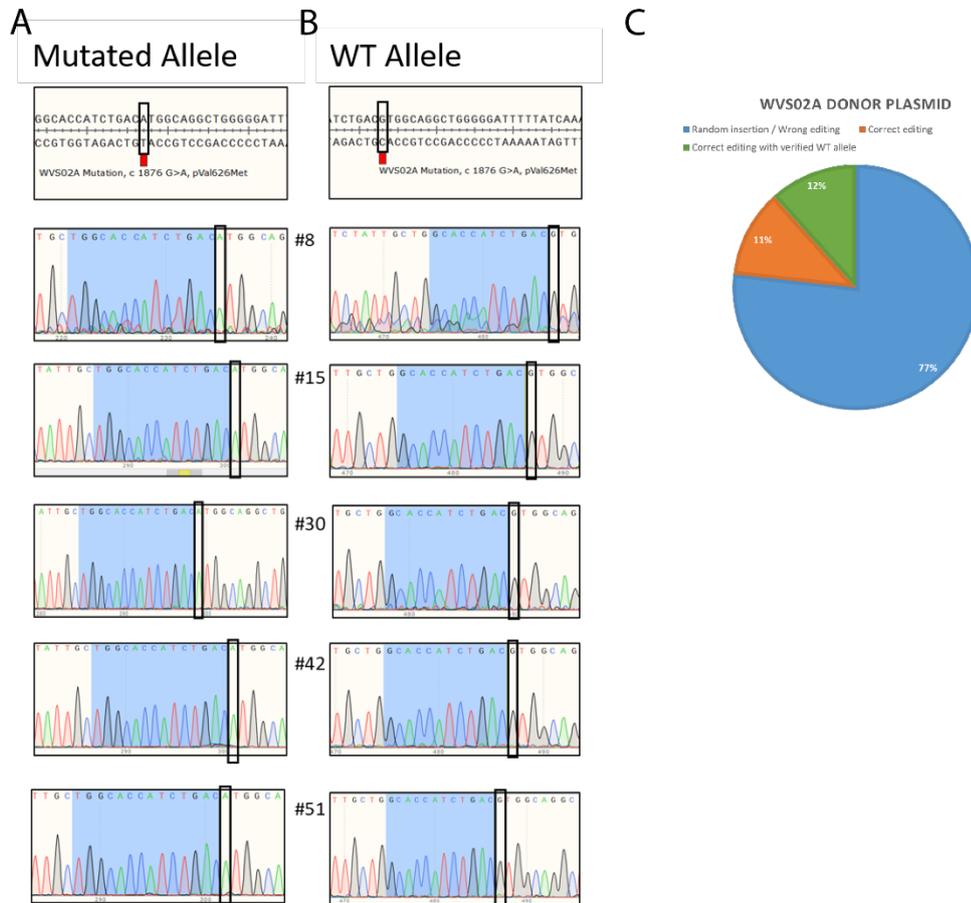


Fig. 109: Validation and efficiency of isogenic lines engineering. Sanger sequencing of edited (**A**) and WT (**B**) alleles of 5 clones from a control line that were correctly engineered with the WVS02A mutation in heterozygosis. (**C**) Pie chart representing the percentage of incorrect and correct editing using the donor plasmid.

This proof of principle demonstrated that we were able to engineer isogenic lines with a highly efficient and streamlined genome editing strategy.

5. DISCUSSION

My objective in this work was to provide a comprehensive transcriptomic characterisation of human cortical brain organoid differentiation dynamics by means of a highly time-resolved integration of datasets, including a large one generated in house. Leveraging transcriptomic data from human fetal cortex and from different brain organoid paradigms, as well as diverse computational tools, we were able to identify key gene networks regulating human cortical development. These transcriptional patterns were largely albeit differentially recapitulated in brain organoids, with protocol-specific peculiarities that can guide the choice of a specific system depending on the experimental question. I then harnessed this high resolution map to investigate, as case study, the transcriptomic and epigenomic profiling underlying the pathogenic alterations caused by mutations of PRC2 in a prototypical example of neurodevelopmental disorder, Weaver Syndrome, ushering in the identification of high-confidence, PRC2 dosage-dependent targets whose dysregulation paves the way to translational studies.

The long-standing challenge of modelling human cortical development derives from the complexity of genetically encoded as well as extrinsic events that need to take place with a precise spatio-temporal sequence during development. In the past two decades, with the advent of hiPSC and later of organoids, we have started to recapitulate some of these events *in vitro*. However, concerns such as the quantification of the reliability and reproducibility of organoid differentiation arose in the field and, as of today, they are only partially solved^{117,121,122,41}. We designed our CBO cohort and analyses to tackle precisely these issues. Indeed, our dataset included CBO from four different individuals (three males and one female), starting from hiPSCs reprogrammed with two different reprogramming methods (three with self-replicating RNAs and one with Sendai virus), differentiated in duplicate and profiled at a single-organoid level. Our results demonstrated that CBO reproducibly recapitulated the transcriptional landscape of pre-natal cortical development and that they modulated signatures of cortical cell types with a developmentally relevant timing. These results are in line with other published reports assessing CBO reproducibility^{125,375}. Additional studies provided insights on the variability of brain organoids, with considerable efforts in terms of standardisation of culture conditions and with the emerging evidence that patterning plays a key role in reducing variability^{126,119}. We were able to evaluate the reproducibility of brain organoids from different individuals in terms of recapitulation of fetal cortex signatures for CBO and TA, while for CO and FO only organoids derived from one line were available. Our analyses revealed that TA have variable behaviour in terms of global transcriptome correlation with pre-natal cortex as well as in terms of modulation of important cortical signatures and of cell composition.

Organoids are very complex cultures entailing the presence of multiple cell populations developing over time. Therefore, scRNAseq approaches have been proven to be powerful to dissect

developmental trajectories in organoids and in the developing brain^{117,122,376}. However, the level of resolution in terms of genes per cell detected is still not sufficient to have a very broad overview of the transcriptional changes in the system analysed, especially when considering non-highly expressed genes and the dropout noise, issues that the field is trying to tackle^{377,378}. Instead, our bulk-RNAseq-centred experimental design allowed us to identify resolved patterns of transcriptional modulations in the developing cortex and in CBO over time, leading to the identification of pathways and hubs relevant for the human pre-natal cortical development and to the understanding of how different brain organoid paradigms resemble those patterns, objectives that could have been only partially achieved relying exclusively on a scRNAseq approach due to its still limited sensitivity. We leveraged publicly available scRNAseq profiling of the human pre-natal cortical development to complement our dataset with the information regarding cell type composition in CBO overtime, finding appearance of relevant populations in line with the expected developmental timings. Thus, in our case bulk- and sc-RNAseq resulted to be complementary and synergic, as demonstrated also by others^{379,375}.

A very recent report assessed the capability of CBO to recapitulate early post-natal development, with a thorough analysis of very late differentiation stages (> 200 days)³⁷⁵. Our work represents the first analysis of CBO with such a time-resolved and detailed reconstruction of the transcriptional footprints of early-mid (day 25 to 200) cortical organoidogenesis phases vis a vis *in vivo* pre-natal cortical development.

The transcriptomic analysis of developing CBO revealed a fast-evolving phase followed by a slow evolving one, as confirmed by DEA, correlation with BrainSpan pre-natal cortical samples and bulk deconvolution. Interestingly, while the fast-evolving phase was characterised by increase in neuronal output and the appearance at day 100 of the oRG signature, the slow-evolving one was characterised by a drastic reduction of the magnitude of transcriptional changes compared to early phases, which were characterised by very large number of DEGs. This is compatible with a more subtle transcriptional evolution accompanied by the onset of other processes related to neuronal maturation such as neuron migration and synaptic establishment. Among the functional categories emerging when analysing the slow-evolving phase, we observed a further decrease in cell cycle, and an upregulation of ECM-related genes. Surprisingly, we noticed upregulation of markers of muscular development and function. With a closer look at the genes leading to this observation, we found that the vast majority of them was expressed also in astrocytes, in line with previous reports analysing CBO-derived astrocytes³³⁰. The developmental timing of their appearance in CBO (day 100 to 150 CBO, corresponding to 14.3 to 21.4 weeks *in vitro*) was aligned with human pre-natal cortex, where we observed the appearance of astrocyte-specific marker AQP4 around PCW 16-17. Interestingly, the blue gene module of the BrainSpan WGCNA included genes related to astrocytes

differentiation as well as categories related to skeletal muscle differentiation. This module showed a U-shaped behaviour, largely recapitulated by CBO, with a high expression at PCW8-9 followed by a drop in expression between PCW12-13 and a raise starting from PCW 16 onwards, again in line with what reported for glial genes during cortical development and with the developmental timing of astrocytes appearance in humans^{330,375}. However, we found that day 200 CBO upregulated genes specific for mesoderm, especially myosin subunits, which were not observed in pre-natal cortex. This finding may be ascribed to an increase in astrocyte reactivity at very late stages, since remodelling of the cytoskeleton, involving also myosins, is a hallmark of astrocyte response to brain injury³⁸⁰ and that it is difficult to maintain long-term astrocyte cultures in a non-reactive state³⁸¹, although previous reports studying CBO-derived astrocytes did not report significant astrogliosis³³⁰. Noteworthy, all other brain organoid systems did not robustly upregulate astrocyte signatures, probably due to the unavailability of late time-points for the external datasets analysed.

Despite the limited time-frame available for external organoid systems, we observed a certain degree of heterochronicity among different protocols in terms of global transcriptomic correlation with BrainSpan, appearance of key cell populations, and subtle transcriptomic differences between consecutive stages observed at early *in vitro* age, resulting in a more compressed transcriptional evolution compared to CBO. To our knowledge, this is the first description of such temporal dynamics in brain organoids. Other reports comparing different paradigms focused on more heterogenous time-points and other aspects of recapitulation^{122,41}. A comparison of single-cell brain organoid datasets against the human fetal brain revealed that organoid age can be assigned on the basis of “age signatures” retrieved from fetal cortex, however authors did not comment on the correspondence between *in vitro* and *in vivo* timings¹²¹. We compared similar *in vitro* ages for all protocols analysed, with the CBO dataset including also very late stages, and showed that CO, FO and TA reach the slow-evolving phase earlier than CBO, with a greater proportion of neurons at earlier time points. Additionally, CO and FO globally correlated earlier with late stages of cortical and CBO development. A possible explanation for this finding could relate to the use of EGF/FGF in the in-house CBO, which would maintain a more immature state for longer time. While CO and FO do not require EGF/FGF for differentiation, TA do, thus going against the hypothesis of faster maturation due to the absence EGF/FGF. However, only early time points were available for TA and the lack of reproducibility among replicates made difficult to draw conclusion on this issue. Another hypothesis could be the timing of use of those patterning factors, which for CBO was extended, possibly leading to a more protracted differentiation matching better the *in vivo* counterpart.

The analysis of BrainSpan pre-natal cortical samples revealed a clear subdivision in two different clusters in terms of transcriptional landscapes, subdividing developmental timings in PCW8-9 and PCW12-24. Interestingly, this distinction coincided with the appearance of the OSVZ from PCW9

and expanding for the following 6 weeks³¹. More specifically, we attributed these broad transcriptional changes to the downregulation of genes involved in cell cycle, drastic reduction of which was observed between PCW9 and 12. Indeed, while the proliferation of progenitors is still ongoing and even increased at these time-points, the neuronal output dramatically raises, thus leading to the observed shift in transcriptomic signatures (yellow-black vs turquoise-pink-midnightblue-gray60-magenta BrainSpan WGCNA modules). In CBO the upregulation of neuronal signatures resulted in an increase over time of the estimated proportion of excitatory glutamatergic neurons, which raised from around 25% to 50% of the whole-organoid composition between day 50 and 100 days (7.1 to 14.3 weeks). This event coincided with the upregulation of the oRG signature reaching about 25% of the CBO composition at day 100, again in line with the developmental timings of human cortical development in terms of OSVZ development and increase of the neuronal output in the developing brain³¹.

Temporal dynamics recapitulation has enormous relevance for disease-modelling, given the emergent convergence of molecular phenotypes of delay or acceleration in neuronal differentiation characterising different clinical conditions¹²³ and the eventuality of masking disease-relevant phenotypes if they are not correctly preserved³⁸². Our analysis started to elucidate these dynamics in brain organoidogenesis, suggesting that CBO *in vitro* development aligns with *in vivo* cortical temporality more closely than the other systems analysed. Depending on the experimental question behind the use of brain organoids, having a model that recapitulates not only the main neurodevelopmental processes but also their timing is extremely important, as in the case of neurodevelopmental disorders³⁸². For neurodegenerative diseases, instead, it may be more convenient to accelerate the organoid maturation process to study mature cell types in a shorter time-frame.

Surprisingly, during CBO development we also noticed upregulation of genes specific for GABAergic interneurons (such as DLX1-2, GAD1-2, and CALB2), which corresponded to the classification as interneurons with our bulk deconvolution approach. Despite resulting unexpected, this result is compliant with previous studies reporting the presence of GABAergic interneuron signatures in dorsalisating brain organoid protocols^{121,335,119}, although we could not exclude an experimental artifact or their origin as interneurons of the olfactory bulb, considering that we did not observe at any stage the expression of the GABAergic neuron progenitor marker NKX2-1^{119,383}. Another peculiar finding was the observation of high correlation between late organoid differentiation time points and early PCW when correlating brain organoids to pre-natal cortex, a result in line also with other reports³⁷⁵. This could be partially explained with the fact that also considering only BrainSpan samples we observed high correlation from PCW12 to 24, although leaving open the question about the reason behind elevated correlation of late-stage organoids with PCW8-9. We tend to exclude

that culture conditions of organoids could lead to this result due to our bulk deconvolution analysis showing very low proportion of early neuronal progenitors at late developmental stages, although other reports found that organoids retain higher progenitor signatures at very late stages compared to the actual fetal cortex⁴¹. Indeed, in this study based on scRNAseq researchers found upregulation of glycolytic genes and endoplasmic reticulum stress signatures in several brain organoid paradigms compared to the actual fetal brain. This aberrant upregulation was linked to impairment in cell type specification with altered proportion of cell types within organoids compared to fetal cortex⁴¹. However, we detected a stable expression of stress-related signatures throughout differentiation in CBO and in all other brain organoid datasets, a pattern observed also in BrainSpan pre-natal cortical samples, although at slightly lower levels. This finding is supported by a very recent report assessing the expression of these signatures in BrainSpan and CBO³⁷⁵ and corroborates the hypothesis of a homeostatic metabolic state in organoids rather than a deleterious increase over-time. This opposite finding could be due to different culture conditions of the BO analysed, to the time-points available in the different studies or to the different resolution of bulk- and sc-RNAseq in terms of sensibility and discrimination of cell types, respectively.

Our comprehensive analysis of the BrainSpan dataset exclusively focused to pre-natal cortex revealed a series of gene modules defining developmental patterns of co-expression. We explored them at an unprecedented level of resolution, providing a resource to the field about the genes mainly driving pre-natal developmental transitions in the cortex and placing key hubs, known or novel, inside relevant functional domains. By applying a specular approach on our CBO dataset, we could observe preservation of modules accounting for fundamental neuronal functions and following a monotonic behaviour along development, with significant overlaps with the corresponding BrainSpan ones. Cross-visualisation of the trends followed by these BrainSpan modules in CBO and *vice versa* revealed concordance in their behaviour, a further confirmation of the fidelity of CBO in recapitulating the principal transcriptional dynamics of the human developing cortex. Importantly, also in the other brain organoid datasets analysed these general expression patterns were preserved. Conversely, when we focused on modules not strictly following a monotonic or close-to monotonic behaviour, we observed that CBO closely resembled BrainSpan trends, while the other brain organoid paradigms did not. Indeed, BrainSpan blue and red modules presented an U-shape behaviour with higher expression at early stages, followed by a drop and again a raise towards the end of gestation. As already discussed, the blue module was composed of genes related to astrocyte development, although it included other genes related to extracellular matrix (ECM) organisation, which was also the main function enriched in the red one. CBO modules green and red showed the highest overlap with BrainSpan blue and red genes and had exactly the same U-shape behaviour, while showing little or opposite variation in other brain organoid systems.

In addition to the BrainSpan dataset, we also compared our CBO to primary fetal tissues sequenced in house. Although this dataset did not cover the same time-frame of BrainSpan, it resulted to be complementary due to the added value of a completely comparable wet and analytical processing, affording the possibility of a direct comparison with CBO. We found that the best strategy to determine cortex-specific patterns in our organoid dataset was to compare all brain tissues available against hiPSC and to then subtract common changes between cortex and different brain areas. The visualisation of these patterns along CBO development highlighted the upregulation of glutamatergic neuron maturation genes, as expected, but also the gradual downregulation of ECM-related genes with a slight re-increase towards day 200, again confirming the relevance of this functional domain in cortical development and the reliability of CBO in recapitulating it. Indeed, many studies have explored the role of ECM during cortical development, reporting its dynamic expression during human cortical development as well as its role in regulating cortical folding and neuronal progenitor proliferation and migration^{60,384,385}. In humans, germinal zones show higher expression of ECM genes compared to the cortical plate⁶⁰, in line with our observation along differentiation. Moreover, ECM is also produced by maturing and mature astrocytes, playing important roles in their own development and in maintaining neuronal homeostasis in health and disease³⁸⁶, again confirming the reliability of the behaviour observed for ECM genes in BrainSpan and CBO. Finally, ECM has implications in the evolution of the human brain, with progenitor cells expressing ECM components at higher levels than in mice³⁸⁷, making its recapitulation in *in vitro* models essential for studying physiological and pathological cortical development. Indeed, ECM expression has already been investigated in brain organoids, with the conclusion that models entailing a 2D step in culture (TA and 2D neuronal cultures) were overexpressing ECM-related genes compared to fetal cortex, while CO and FO showed downregulation of the same genes¹²⁰. Our analyses provide a detailed overview of ECM trends in the developing cortex and in CBO, resulting in a complete *in vivo/in vitro* alignment not observed for CO, FO and TA, which showed little or opposite changes.

Together, our results contribute to the definition of transcriptional footprints and dynamics specific of pre-natal cortical development, representing a collection of prioritised known and novel hubs that we categorised in well-defined functional domains. Our approach describes the extent of *in vivo/in vitro* alignment of developmentally relevant processes and temporality, highlighting commonalities and diversities of different brain organoid paradigms and providing a resource available for consultation when modelling physiological or pathological human cortical development.

Upon characterisation of the transcriptional dynamics governing CBO differentiation, we used this model to study Weaver Syndrome. As of today, WS remains a clinical unmet need due to the

experimental challenge of studying disease-relevant cell types in a human and patient-specific setting. Moreover, what is known about PRC2 function in cortical development derives exclusively from studies based on mouse models, where its role has been dissected by KO of key subunits, starting from pluripotency^{388,203} throughout the astrogenic phase²¹⁶. WS is instead caused by heterozygous mutations in different PRC2 components with confirmed LOF only for a subset of them²⁴⁸, while others have been classified as putative gain of function (GOF)²⁵⁰. In this work, we analysed a spectrum of mutations falling in the PRC2 subunits EZH2 and EED to score differences between WS and controls in the two main WS clinical domains, namely intellectual disability and craniofacial abnormalities, by means of CBO and NCSC, respectively. The choice of using CBO as a model for the WS-related intellectual disability, besides our comprehensive characterisation of their reliability in modelling cortical developmental processes and timings, ensured the specification of glutamatergic cortical neurons, which we hypothesised to be the cell type most probably responsible for patient phenotypes. This hypothesis grounded on clinical data from WS patients, whose MRI profiles show pachygyria and perisylvian polymicrogyria³⁶⁸, malformations due to alteration of the cortical development in late stages of neuronal migration, where the deeper layers of the cortex form multiple small gyri with derangement of the normal lamination and sulcation³⁸⁹.

One of the typical hallmarks of WS is overgrowth, which can be or not accompanied by macrocephaly^{228,225,227}. In our cohort two patients had this clinical feature and therefore we tested whether their increased head circumference was dependent on abnormal neuronal differentiation, thus leading to difference in CBO size³⁹⁰. Our results show divergent behaviour for different patients, with only WVS02A forming organoids bigger than control, in line with the macrocephaly phenotype of the patient. The other line deriving from a macrocephalic WS individual actually showed an opposite behaviour, with CBO resulting smaller than controls. This was true also for a third patient within the cohort, while for the fourth we could not perform this analysis due to low efficiency of organoid generation. Interestingly, EZH2-LOF CBO also showed marked reduction in organoid growth, suggesting that LOF of EZH2 is responsible for reduced CBO size, which could be explained with an acceleration of differentiation, a phenotype already observed in a model of *Ezh2* ablation at the beginning of neurogenesis²¹⁰. Nonetheless, this hypothesis is partially in contrast with a report suggesting that one of the patients of our cohort, showing reduced CBO growth (WVS03B), could actually carry an EZH2 GOF, although this was not demonstrated biochemically²⁵⁰. While EZH2 complete loss in pluripotency leads to failure in neuronal differentiation²⁰³, our settings are different because patients carry a functional EZH2 allele and EZH2-LOF CBO retain the expression of EZH2 to some extent, probably being able to fulfil structural roles in chromatin regulation, which have been shown to be essential for correct neuronal maturation³⁹¹. Indeed, transcriptomic analysis of EZH2-LOF CBO shows downregulation of pathways related to general

forebrain development, however we detect markers of neuronal progenitors and neurons by immunostainings.

Overactivation of the PI3K-MTOR pathway leads to megalencephaly, a phenotype already explored using brain organoids^{392,390}. Transcriptomic analysis of WS CBO at day 50 revealed downregulation of signatures related to MTORC1 signaling and cell proliferation, in line with the phenotype of reduced CBO growth observed. EZH2 has been reported to positively regulate MTOR-dependent signaling in the hippocampus³⁹³, therefore its haploinsufficiency could cause the inhibition of this pathway. Our results suggest that EZH2-LOF causes overall reduced growth of CBO at early stages of differentiation. However, this phenotype must be compensated later in development since microcephaly is not among the clinical features of WS. This set of experiments was performed in 96-well format, making impossible the analysis beyond day 50 due to the organoids' size. To analyze more advanced stages of differentiation another culture format could be considered.

Transcriptomic analyses of patient- and control-derived CBO highlighted different characteristic of WS at the developmental stages analysed. The stage-wise analysis including all time-points revealed a set of genes that selectively failed downregulation along maturation in WS, including some important developmental regulators known to be *bona fide* targets of EZH2. This result confirmed at least a partial loss of function of PRC2 in WS organoids. The stage-by-stage analysis revealed instead specific characteristic of WS. Day 25 WS CBO were characterised by a general downregulation of the PERK-mediated unfolded protein response (UPR). Our results are in contrast with the ones from Yao-Yu and colleagues, who demonstrated that EZH2 inhibitors upregulate the PERK-mediated arm of the UPR in colon rectal cancer³⁹⁴. This may be due to different reasons, including that in corticogenesis the regulation of this pathway is different from the one of colon rectal cancer or that in the context of WS it is plausible that PRC2 mutations are not merely LOF but may selectively lead to upregulation of some targets while reinforcing/not influencing the repression of others, especially considering the heterozygosity of the genetic lesions. At the same developmental stage, we also observed an upregulation of genes involved in synapse maturation and axon development, again supporting the hypothesis that timing of WS corticogenesis is impaired and tends to an acceleration. This result was still maintained at day 50 of differentiation where again WS CBO upregulated genes related to neuronal maturation while downregulating cell cycle. In parallel to this, day 50 WS CBO also upregulated genes involved in neuronal migration, primarily RELN, a well-known target of EZH2 and mediator of the correct migration of maturing neurons through the RG layer³⁹⁵. Interestingly, RELN has been linked to polymicrogyria³⁹⁶, reinforcing the evidence that neuronal migration is molecularly dysregulated during WS corticogenesis. Intriguingly, day 100 WS CBOs showed an impaired glucose metabolism, which was not apparently due to their differential size compared to controls. Ablation of EZH2 promotes

tumorigenesis and malignant progression of glioblastoma in part by activating glycolysis through an EAF2-HIF1 α signaling axis³⁹⁷, in line with a possible metabolic alteration that could in turn lead to impaired neuronal migration in WS³⁹⁸.

Due to the heterogeneity of patients in our cohort and to the fact that one of the mutations has been classified as a putative GOF²⁵⁰, we performed an analysis that took into account this different mutational landscape. Indeed, we subdivided WS lines in classical WS (WVS02A, WVS01H, WVS04A), putative GOF (WVS03B) and severe LOF (EZH2-LOF). This analysis confirmed that classical WS, as well as EZH2-LOF, were again characterised by upregulation of genes related to neuronal differentiation. Interestingly, the putative GOF line did not show the same trends for top DEGs, despite displaying the same CBO growth phenotype of EZH2-LOF and that genes related to neuronal differentiation resulted anyway upregulated when considering it as a classical WS.

These results led us to investigate the cell composition within our organoids. Interestingly, we observed significant upregulation of the aRG marker PAX6 in both WS and EZH2-LOF at day 25, a difference that was not significant anymore at day 50. When measuring the lower layer neuronal marker CTIP2, we observed a trend of increased expression in both WS and EZH2-LOF. Despite not being statistically significant, probably due to the low sample size, this result coupled with the upregulation of PAX6 at day 25 is in line with the accelerated differentiation hypothesis in WS. Surprisingly, we found PAX6 re-upregulation at day 100 in WS, however further experiments are needed to interpret this result. The intersection of transcriptomic data from WS CBO with the transcriptomic analysis performed on EZH2-LOF CBO revealed that around 20-30 % of the DEGs observed in WS overlap with the ones of EZH2-LOF at all stages analysed, suggesting that part of the phenotype aligns with a loss of function, while another part can not be explained merely by a lower EZH2 activity.

Given the epigenetic function of PRC2¹⁸¹, we profiled H3K27me3 in CBO to relate differences in gene expression to actual changes at the chromatin level. CHIP-seq on brain organoids was already performed in bulk¹¹⁸. However, the heterogeneity in term of cell type composition and the often low input material, especially for late stages of differentiation, hampers the achievement of optimal experimental conditions. We profiled CBO at day 25, 50 and 100 of differentiation and we ensured reliability of observed peaks by intersecting our CHIP-seq with publicly available datasets from hESC, neuronal progenitors and neurons, confirming overlap especially with the neuronal lineage. Our epigenomic profiling indicated that peaks specifically lost or gained in WS are enriched in enhancer regions, while there is no difference in terms of distribution at promoters. This finding is particularly relevant given the demonstrated role of PRC2 during neuronal differentiation. Indeed, while loss of PRC2 at the pluripotent stage does not cause changes in the activation of poised enhancers, when impaired during neuronal lineage specification it hampers the induction of important anterior

neuronal genes regulated by poised enhancers due to disruption of a permissive topology³⁹¹. Another epigenetic modification tightly related to PRC2 activity is DNA methylation. The interplay between these two chromatinic marks has been demonstrated to be particularly relevant in neuronal specification. Indeed, the vast majority of promoters acquires DNA methylation during the differentiation from pluripotency to neuronal progenitors and H3K27me3 promoters result more prone to acquire DNA methylation during neuronal development²¹⁷. We profiled DNA methylation profiles of day 25, 50 and 100 CBO, together also with hiPSC, NCSC and iN.

After having identified the general molecular pathways dysregulated in WS corticogenesis, we performed integration of transcriptomic, H3K27me3 and DNA methylation CBO profiles in order to identify a core of genes reliably linked to the intellectual disability associated to WS. We focused on WS upregulated genes, compatible with a EZH2 LOF, because of the general evidence that WS genetic lesions cause impairment of PRC2 function²⁴⁸ and of the fact that EZH2 heterozygous deletions lead to intellectual disability³⁹⁹. Moreover, the focus on aberrantly upregulated genes leaves open the possibility of pharmacologically inhibit the identified targets to score for the rescue of the phenotypes observed.

The intersection of these three layers highlighted the robust dysregulation of 3 main targets, namely AJAP1, PAX8 and CDKN1A across all layers analysed. AJAP1 is an adherens junction associated protein that localises and interacts with β -catenin in the E-cadherin-catenin complex that was found dysregulated in different types of tumor, including glioblastoma^{369,400,401}. AJAP1 is located in the 1p36 genomic region along with CHD5, also found upregulated in both WS and frameshift CBO at day 50. Interestingly, copy number variation of this region was associated to intellectual disability³⁷⁰, making them interesting targets also in the context of WS. Although very few is known about PAX8 function in corticogenesis, it was reported to regulate the expression of the mouse *Engrailed* gene, essential for the development of midbrain and cerebellum. Moreover, it is a known target of EZH2 in the context of thyroid carcinomas⁴⁰². CDKN1A is also a well-known target of EZH2 and it was reported to be essential in neuroblasts migration in chicken embryos²¹⁵. Despite CDKN1A is an early target of EZH2 in corticogenesis²¹⁵, we found it dysregulated only at day 100 of differentiation. This may be due to the haploinsufficiency context of WS, where the dysregulation of important targets may be milder due to the presence of a fully functional allele of EZH2.

Summarising, our work provides new insights into the role of PRC2 in human corticogenesis in the context of WS both in terms of general pathway dysregulated upon its mutation as well as in terms of the identification of druggable *bona-fide* targets for which we brought evidence of their involvement in the WS-related intellectual disability phenotype. Moreover, we showed that CBO are able to recapitulate WS-relevant phenotypes, as opposite to iN where we did not score

significant differences against controls, again in line with recent research showing that accelerated neuronal differentiation masks ASD relevant phenotypes³⁸². Lastly, we found that WS NCSC carry transcriptional deregulation of genes involved in neural crest migration, grounding the hypothesis that the craniofacial defects characteristic of WS have their molecular roots at the NCSC level.

In conclusion, we contributed to the elucidation the molecular landscape of the two main domains of WS clinical manifestations, namely intellectual disability and craniofacial abnormalities, although further experiments are needed to crystalise these results. Indeed, sc-RNAseq of WS and control CBO has already been performed and is under analysis for day 50, 100 and 250 to further dissect the accelerated neuronal maturation phenotype observed, which will also need to be complemented with functional experiments on CBO such as the expansion of the sample size for markers already profiled and the dissection of dysregulated pathways found in bulk transcriptomics (e. g. MTROC1, neuronal migration, glycolysis). Moreover, further analysis on ChIP and DNA methylation profiles, complemented also with EZH2 tracks, will be pivotal to understand the effect of PRC2 alteration at the enhancer level. Importantly, functional validation of the common deregulated targets across layers will be carried out by means of isogenic lines and CRISPR activation/interference. Importantly, we already standardised a straightforward pipeline to engineer isogenic lines, as demonstrated in this work by the engineering of a WS-specific mutation in a control genetic background. The available isogenic pair will be used to perform CRISPR epi-engineering of the molecular targets already identified from RNAseq and ChIPseq, using as a functional readout migration assays and evaluation of organoids' composition at different time-points, given that migration/cell adhesion and accelerated differentiation were the two main phenotypes found by transcriptomics in the patient cohort.

As a model for the intellectual disability found in WS we hypothesise that the transcriptional alteration due to PRC2 malfunction is reverberating mainly on circuits related to neuronal migration and differentiation, with key hubs strongly deregulated that may mediate this phenotype and require further validation as described above. Growth curve results show a differential behaviour of WS CBO, which may be partially explained by the different mutations of our patient cohort. However, we believe that the impaired WS CBO growth may lead to the same phenotype of accelerated differentiation, since this is a transcriptional phenotype common to all patients. This hypothesis could be explained by a transient more pronounced expansion of apical progenitors in WS, which resolves in different moments depending on the patient-specific mutation, thus leading to different growth phenotypes in early stages but eventually to a common increase of neurons in WS compared to controls. This hypothesis could be answered by the sc-RNAseq analysis we are carrying out and by the more functional studies proposed above in the context of the validation of the molecular targets identified. Another phenotype observed in WS is the dysregulation of

H3K27me3 apposition at enhancer regions in WS CBO, which requires further investigation to be linked to the phenotypes observed.

Lastly, we will carry out on NCSC the same epigenomic profiling applied for CBO as well as the validation of the migration phenotype observed in transcriptomics. Further differentiation of NCSC in craniofacial-relevant cell types will also provide further insights on the PRC2-dependent alterations of this lineage in the context of WS.

6. MATERIALS AND METHODS

6.1 Experimental procedures

6.1.1 Cell reprogramming

Skin fibroblasts from WS patients were received from Prof. William Gibson and Prof. Rosanna Weksberg from BC Children's Hospital (Vancouver) and The Hospital for Sick Children (Toronto). Reprogramming was performed using non integrating self-replicating mRNAs as previously described³⁶⁷ (Stemgent, 00-0071) or Sendai virus (CytoTune-iPS 2.0 Sendai Reprogramming Kit; Thermo Fisher Scientific, A16517). One of the control individual (CTL09A) of the WS cohort was the father of WVS04A and was received from Prof. William Gibson. CTL01A and CTL02A were received from the Department of Biomedical Sciences at Antwerpen Universiteit (Belgium) and the genomic and Genetic Disorders Biobank at the IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Italy, affiliated to the Telethon Network of Genetic Biobanks), respectively. CTL08A was purchased from the Wellcome Trust Sanger Institute and was the only line reprogrammed using the Sendai virus. The information regarding hiPSC used in this work is summarised in **table 2**.

Internal code	Code used in this work	Derivation	Reprogramming method	Starting material
CTL09A	CTL1	BC Children's Hospital (Vancouver)	Self-replicating RNA	Fibroblasts
CTL08A	CTL2	Wellcome Trust Sanger Institute	Sendai virus	Fibroblasts
CTL02A	CTL3	University Of Sheffield	Self-replicating RNA	Fibroblasts
CTL01A	CTL4	Genomic and Genetic Disorders Biobank at the IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, affiliated to the Telethon Network of Genetic Biobanks	Self-replicating RNA	Fibroblasts
WVS01H	WVS01H	The Hospital for Sick Children (Toronto)	Self-replicating RNA	Fibroblasts
WVS02A	WVS02A	The Hospital for Sick Children (Toronto)	Self-replicating RNA	Fibroblasts
WVS03B	WVS03B	BC Children's Hospital (Vancouver)	Self-replicating RNA	Fibroblasts
WVS04A	WVS04A	BC Children's Hospital (Vancouver)	Self-replicating RNA	Fibroblasts

Table 2: List of hiPSC used in this work. The list includes all the cell lines used, their codified name, derivation, reprogramming method and the starting material.

6.1.2 hiPSC maintenance

hiPSC were cultured in TeSR-E8 medium (Stemcell technologies, 05990) supplemented with penicillin-streptomycin (P/S, 100 U/mL; Thermo Fisher Scientific, 15140-122), with daily media change, at 37 °C, 5 % CO₂ and 3 % O₂ in standard incubators. hiPSC were grow on matrigel-coated dishes prepared as follows: matrigel stock solution (Corning, 354248) was diluted 1:40 in DMEM /F-12 1:1 medium (Lonza, BE12-614F and Thermo Fisher Scientific, 11765054, respectively) supplemented with P/S 100 U/mL and used to coat dishes for 15-30 minutes at 37 °C. Passages 1:6-1:8 were performed using ReLeSR™ (Stemcell technologies, 05872) or Accutase solution (Sigma-Aldrich, A6964). ReLeSR was used to detach hiPSC in clumps for expansion and standard maintenance. Accutase solution was used for single cell passaging; in this case, ROCK inhibitor 5µM (Sigma, Y0503) was added to the culture overnight (ON) to enhance single hiPSC survival. Cryopreservation of hiPSC was performed by single cell dissociation and storage in complete TeSR-E8 medium plus 10% DMSO supplemented with ROCK inhibitor 5µM.

6.1.3 CBO generation and maintenance

CBO were generated using an adaptation of the previously described protocol published by Pasca et al in 2015¹⁰⁴. More in details, hiPSC from a confluent 6 cm plate were detached with dispase solution (Stem Cell Technologies, 07923), partially dissociated mechanically and plated on top of a layer of irradiated feeder (on a 10 cm matrigel-coated dish). hiPSC were grown on feeders for 3-4 days in a medium composed of 80% DMEM/F12 medium (1:1), 20% Knockout serum (Thermo Fisher Scientific, 10828028), 1% Non-essential amino acids (NEAA, Lonza BE13-114E), 0.1 mM cell culture grade 2-mercaptoethanol solution (Thermo Fisher Scientific, 31350010), 2 mM L-Glutamine (Thermo Fisher Scientific, 25030081), P/S 100 U/mL, and FGF2 at a final concentration of 20 ng/mL (Thermo Fisher Scientific, PHG0021). Daily media change was performed. Embryoid bodies (EB) were generated by detaching hiPSC with dispase for 40 minutes and plating on ultralow attachment 10 cm plates (Corning, 3262) in the first differentiating medium composed of 80% DMEM/F12 medium (1:1), 20% Knockout serum, 1% NEAA, 0.1 mM cell culture grade 2-mercaptoethanol solution, 2 mM L-Glutamine, P/S, 100 U/mL, 7,5 μ M Dorsomorphin (MedChem express, HY-13418A), 10 μ M TGF β inhibitor SB431542 (MedChem express, HY-10431), and ROCK inhibitor 5 μ M. EB were grown in normal oxygen incubators. EB were left undisturbed for 1 day and at 48h media change was performed with differentiation medium 1 without ROCK inhibitor. Dorsomorphin and TGF β inhibitor are used to perform Dual-SMAD inhibition, pushing neuroectoderm specification. Dual-SMAD inhibition was performed for a total of 5 days, with daily media change. On day 6 the second differentiation medium was added until day 25 with daily media change for the first 12 days, and then every other day. The second differentiation medium was composed of neurobasal medium (Thermo Fisher Scientific, 12348017) supplemented with 1X B-27 supplement without vitamin A (Thermo Fisher Scientific 12587001), 2 mM L-Glutamine, P/S, 100 U/mL, 20 ng/mL FGF2 and 20 ng/mL EGF (Thermo Fisher Scientific, PHG0313). Human FGF2 and EGF were used to amplify the pool of neural progenitors. On day 12, CBO were moved to ultra-low attachment 10 cm dishes and grown on shakers to enhance oxygen and nutrient supply. On day 26, FGF2 and EGF were replaced with 20 ng/mL brain-derived neurotrophic factor (BDNF, Peprotech 450-02) and 20 ng/mL neurotrophin-3 (NT3, Peprotech 450-03) to promote differentiation of neural progenitors towards the glutamatergic fate. From day 43 onwards, BDNF and NT3 were removed and from day 50 the medium was supplemented with Amphotericin β to prevent mould formation.

Alternatively to EB formation starting from hiPSC grown on irradiated feeders, EB were also generated by single-cell aggregation in 96-well ultralow-attachment plates (S-bio Duotech, MS-9096UZ). In this case, hiPSC were expanded on matrigel-coated 10 cm plates and dissociated at 60% confluency with Accutase solution for 3 minutes. Cells were centrifuged to remove the enzymatic suspension (160g for 3 minutes). After resuspension in TeSR-E8 medium supplemented with 5 μ M ROCK inhibitor cells were counted with a TC20 automatic cell counter (Biorad) and seeded into 96

ultra-low attachment well plates at a final concentration of 2×10^4 cells in each well. Plates were then centrifuged at 160g for 3 minutes to enhance EB formation. The following day medium was not changed, leaving EB undisturbed. From this point, dual-SMAD inhibition was performed and then the same differentiation pipeline described above was applied. The time-points day 25, 50, 100, 150 and 200 were used in this work.

6.1.4 NSCS differentiation

hiPSC were differentiated into NCSC as previously described⁴⁰³. NCSC differentiation required 15-20 days and was carried out as follows: 90% confluent hiPSC were detached with Accutase solution and plated on matrigel coated dishes in TeSR-E8 medium supplemented with 5 μ M ROCK inhibitor at a density of $\sim 9.2 \times 10^4$ cells per cm^2 . The day after, NCSC differentiation medium was added and changed every day for 15-20 days. NCSC medium was composed of DMEM-F-12 1:1, 10% probumin Life Science Grade from the 20% stock solution (20% m/v in DMEM F-12 1:1, stock solution; Millipore, 821001), P/S 100 U/mL, 2 mM L-Glutamine, 1% NEAA, 0.1% 1000X trace elements complex (CA055-010, Gentaur Italy Srl), 0.2% 50 mM b-mercaptoethanol, 10 μ g/ml Transferrin, bovine (Holo form) (Life Technologies, 11107-018), 50 μ g/ml (+)-Sodium L-ascorbate (Sigma, A4034), 10 ng/ml Heregulin-1 (Peprotech, 100-03), 200 ng/ml LONGÒR3 IGF-I (Sigma, 85580C), 8 ng/ml FGF2, 3 μ M GSK3 inhibitor IX (BIO) (MedChem express, HY-10580) and 20 μ m SB431542 (MedChem express, HY-10431). Cells were passaged every 4-5 days and plated at high concentration (1:1 the first time and 1:2 the following ones) on matrigel coated dishes for the entire duration of the differentiation. Upon differentiation, NCSC were stocked as stable lines and cultured in the medium and splitting ratios. FACS following staining for HNK1 (Sigma, c6680) and NGFR (Advanced Targeting System, AB-N07) was performed to assess NCSC identity.

6.1.5 iN differentiation

iN were differentiated with an adaptation of the protocol described in⁹⁵. Briefly, the lentiviral construct was substituted with a PiggyBac-based construct that was electroporated into hiPSC using the Neon Transfection System. 4×10^5 cells per line were transfected with 2.25 μ g of PiggyBac construct and 250 ng of helper construct carrying the transposase. Upon selection with blasticidin (Merck, 3513-03-9), the lines were differentiated as previously described⁹⁵. Briefly, 1×10^6 hiPSC were plated on matrigel-coated 10 cm plates in complete TeSR-E8 medium supplemented with ROCK inhibitor 5 μ M. On the following day, media 1 supplemented with 2 μ g/mL doxycycline (Sigma-Aldrich, 24390-14-5) was applied for one day to induce transgenic Ngn2 expression. On day 2 media 1 supplemented with doxycycline 2 μ g/mL and puromycin 1 μ g/mL (Sigma-Aldrich, 58-58-2) to select cells that correctly induced Ngn2 expression. On day 3 media 2 was applied. On day 6-8 of differentiation, iN were transferred on poly-D-lysine (Sigma-Aldrich 27964-99-4) coated plates alone (transcriptomic) or together with astrocytes (1:1) (electrophysiology). Medium was replaced

once per week until 35 days of differentiation. Media were composed as follows: i) Medium 1: DMEM/F-12 1:1, P/S 100 U/mL, 1% NEAA, 1X N2 supplement (Thermo Fisher Scientific 17502001), BDNF (10ng/mL), NT3 (10ng/mL), 1 µg/mL Laminin (R&D systems (Space) 3400-010-01); ii) medium 2: Neurobasal-A (ThermoFisher)/F-12 1:1, (100 U/mL) and streptomycin (100 µg/mL), N2 supplement (1:200), Non-essential amino acids (1 mM), Laminin (1 µg/mL), BDNF (10 ng/mL), NT3 (10 ng/mL), B-27 supplement without vitamin A (1:100), Doxycycline (2 µg/mL). Electrophysiology was performed at Humanitas University by Davide Pozzi from Prof. Michela Matteoli's group. Patch-clamp recordings were performed in an extracellular solution with the following composition (in mM): 130 NaCl, 5 KCl, 1.2 KH₂PO₄, 1.2 MgSO₄, 2 CaCl₂, 25 HEPES, and 6 Glucose, pH 7.4, in the presence of synaptic transmission blockers, CNQX 10 mM and APV 20 mM. Borosilicate glass pipettes of 4-6 MΩ were filled with the following internal solution (in mM): 135 K-gluconate, 5 KCl, MgCl₂, 10 HEPES, 1 EGTA, 2 ATP, 0.5 GTP, pH 7.4. Electrical signals were amplified by a Multiclamp 200 B (Axon instruments), filtered at 5 kHz, digitised at 20 kHz with a digidata 1440 and stored with pClamp 10 (Axon instruments). Passive properties including capacitance and membrane resistance were calculated in voltage clamp mode using a pulse test of 10 mV. Only neurons with a stable (max deviation 10%) access resistance <15 mW and with a holding current < 100 pA were considered. Intrinsic neuronal excitability was calculated as the total number of action potentials for each current step, whereas the current threshold density was calculated as the minimum depolarising current needed to elicit at least one action potential. The parameters describing the action potential shape (AP peak and Max Rising Slope) were analysed using the pClamp software (Molecular Devices) and data were analysed using Prism software (GraphPad).

6.1.6 Culture conditions for fetal cortical cell

Human fetal primary neural stem cells were provided by Dr. Steve Pollard's laboratory. They were derived from the cortex of post-conception week 11 and 19, male embryos. They were cultured in the following medium DMEM/F12 medium (1:1), P/S (100 U/mL), 0.1 mM cell culture grade 2-mercaptoethanol solution, 1% NEAA, 0.5% N2 supplement, 1X B27 Supplement 100X (Thermo Fisher Scientific, 17504-044), 0,012% Bovine Albumin Fraction V (Thermo Fisher Scientific, 15260-037), 1,5 g/L glucose (Sigma-Aldrich, G8644). Washes for this type of cells were performed with a medium composed of DMEM/F12 medium (1:1), P/S (100 U/mL) and 0,015% Bovine Albumin Fraction V.

6.1.7 CBO and hiPSC immunofluorescence

CBO were harvested on day 25, 50, 100 and fixed 2 hours in paraformaldehyde 4% (ChemCruz, sc-281692) on orbital shaking. Then, sucrose 30% was added for 12-16 hours after a wash in PBS 1X. CBO were then embedded in cryostat embedding medium (Bio Optica 05-9801). Cryosection were obtained with a standard protocol using Leica CM 1900 instrument with 10 µm thickness. Sodium

Citrate Buffer (10 mM Sodium Citrate, 0,05% Tween, in ddH₂O, pH 6.0) was used for antigen retrieval. Slides were immersed in Sodium citrate buffer in the water-bath at 95°C for 45 minutes and then left at room temperature for 45 minutes. Slides were then treated with 0,1 M Glycine pH 7.4 for 10 minutes to reduce autofluorescence. Subsequently, blocking solution was added (5% serum, 1% triton X-100 in PBS 1X) and the slides were incubated 30 minutes at room temperature. Primary antibodies in antibody buffer (2% South America serum, Euroclone ECS0182L, in PBS 1X) were added and slides were incubated ON at 4°C. The day after, 5 consecutive washes in PBS were performed and then the slides were incubated for 1 hour at room temperature with the secondary antibodies in antibody buffer. DAPI (Sigma, D9542) was added for 5 minutes at room temperature after 3 washes in PBS 1X. Slides were washed with PBS 1X and then water, dried and mounted using VectaMount mounting media (Vector Laboratories, H-5501).

hiPSC were plated on matrigel-coated glass coverslips were fixed with paraformaldehyde 4% for 10 minutes on ice and then washed with PBS 1X. Subsequently, blocking solution was added and then the same protocol used for CBO was applied.

The following primary antibodies were used:

Antibody	Dilution	Species secondary	Brand
NANOG	1 in 1000	Goat	Everest Biotech,EB06860
OCT3/4	1 in 100	Rabbit	Santacruz, sc-9081
CTIP2	1 in 200	Rat	Abcam, ab18465
KI67	1 in 500	Rabbit	Abcam, ab15580
PAX6	1 in 250	Rabbit	Biolegend, 561664
SOX2	1 in 50	Goat	R&D system, AF2018
TBR2	1 in 200	Rabbit	Abcam, ab23345
SATB2	1 in 100	Mouse	Abcam, ab51502

Table 3: List of antibodies used for immunofluorescence analyses

CBO and hiPSC samples were acquired on a widefield microscope (Leica DMI6 B), equipped with an Andor Zyla (VSC-04470 sCMOS), using 20X/0.75 dry objective.

For CBO, single fields of view were automatically merged in order to reconstruct the whole slice. The steps of image processing and analysis were automatized through an in-house developed plug-in for ImageJ. Images were processed by applying a rolling ball background subtraction (radius = 50 px) and a gaussian blur as a noise filter (radius = 2 px). To overcome the impossibility to perform single cell segmentation, especially regarding the DAPI channel, we choose to use the fraction of thresholded area in the channel of the marker of interest over the thresholded area occupied by DAPI signal as a parameter of comparison between conditions. The algorithms used for the threshold were Li, Moments or Otsu, according to the best performance on the image analysed. Importantly, using this approach, quantifications were performed on whole-organoid. Numerical

results were analysed with PRISM (GraphPad, version 8) applying a non-parametric T-test (Mann-Whitney test).

6.1.8 Growth curves

CBO were grown as described above with the only difference that they were kept into 96 ultralow attachment well plates up to day 50. Photos of 12 organoids per cell line per batch of differentiation were captured at day 5, 6, 15, 25 and 50 with an Evos Cell imaging system XL (Thermo Fisher Scientific) with the following parameters: Phase 4, Objective 2x and Illumination 35-55%. The area of the CBO was measured with ImageJ and then the area of each organoid was normalised according to its area at day 5. The resulting ratios were used to create the growth curves. Data were then analysed at days 6, 15, 25 and 50 applying a non-parametric T-test (Mann-Whitney test) with PRISM (GraphPad, version 8).

6.1.9 Taqman assays for glycolytic genes

Total RNA was extracted from snap-frozen pellets CBO at day 100 using the RNeasy Mini Kit (Qiagen). Purified RNA was quantified using a NanoDrop spectrophotometer. Taqman assays to validate the overexpression of glycolytic genes in WS day 100 CBO were performed by Cogentech SRL according to standard procedures. Each sample was analysed in triplicate and normalised to GAPDH. Relative mRNA quantity was calculated by the comparative cycle threshold (Ct) method using the formula $2^{-\Delta\Delta Ct}$. GAPDH assay: GAPDH:Hs99999905_m1; PGK1:hs99999906_m1; TBP assay: hs99999910_m1.

6.1.10 RNA extraction and library preparation for RNA-seq

Total RNA was extracted from snap-frozen pellets fetal of cortical cells, NCSC, iN, CBO at day 25, 50, 100, 150, 200 using the RNeasy Mini Kit (Qiagen, 74104). Purified RNA was quantified using a NanoDrop spectrophotometer and RNA quality was checked with an Agilent 2100 Bioanalyzer using the RNA nano kit (Agilent, 5067-1512). Library preparation for RNA sequencing was performed according to TruSeq Total RNA sample preparation protocol (Illumina, RS-122-2202), starting from 250 ng - 1 µg of total RNA. cDNA library quality was assessed at on Agilent 2100 Bioanalyzer, using the high sensitivity DNA kit (Agilent 5067-4626). Libraries were sequenced with the Illumina Novaseq machine at a read length of 50 bp paired-end and a coverage of 35 million of reads per sample.

6.1.11 ChIP-seq

ChIP-seq for H3K27me3 was performed on CBO deriving for all the lines of the cohort (4 controls, 4 WS and EZH2-LOF) at day 25, 50 and 100. 2 frozen organoid pellets per line per stage were thawed on ice and then formaldehyde 1% in PBS 1X (Sigma, F8775) was added and organoids were mechanically disrupted with a pestle. Then, crosslinking was carried out for 15 minutes at RT in

rotation. Quenching was performed using glycine at a final concentration of 125 mM for 5 minutes at RT in rotation. Samples were centrifuged at 160g at RT for 5 minutes and then resuspended in nuclei lysis buffer supplemented with PIC (Roche, 11697498001). Lysis proceeded for 20 minutes in ice and sonication was performed using a S220 Focused-Ultrasonicator (Covaris) with the following parameters in 130 μ L vessels: 150 watt, 20% duty, 200 cycle per burst for 330 seconds. The two sonicated CBO samples for the same line and stage were pooled together and centrifuged at 4°C for 30 minutes at maximum speed. Supernatants were recovered and protein were quantified using the BCA protein quantification kit (Thermo Fisher Scientific, 23225). 40 μ g of chromatin per sample were taken for the H3K27me3 and IgG immunoprecipitations. Samples were diluted to 200 μ L with ChIP dilution buffer and 2 μ L for every samples were taken as input. Then, 2 μ g of H3K27me3 antibody (Cell signaling technology, C36B11) or IgG (Invitrogen, 02-6102) were added to samples and incubation was carried out ON. The day after, 25 μ L of Dyna Beads Protein G (Thermo Fisher Scientific 10003D) per sample were magnetised for 1 minutes and then washed with ChIP dilution buffer for 3 times. After the last wash, Dyna Beads were resuspended in the initial volume using ChIP dilution buffer and 25 μ L were added to each sample (beads had a concentration of 30 mg/mL with a binding capacity of 6,25 μ g of antibody every 25 μ L). Beads and samples were incubated for 4h in rotation at 4°C. Following incubation, samples were washed two times with RIPA buffer (5 minutes each wash), one time with RIPA buffer high salt (5 minutes), one time with LiCl buffer (5 minutes) and one time with TE buffer (5 minutes). After completion of washes cycles, elution buffer was added to every samples and decrosslinking was performed for 30 minutes at 37°C with RNAseA (10 μ g per sample) and then ON at 65°C with the addition of Proteinase K (10 μ g per sample) in agitation. The following day, beads were removed and samples purified with the Qiagen PCR purification kit (Qiagen, 28104). Quantification was performed using the Qubit dsDNA HS assay kit (Thermo Fisher Scientific, Q32851). DNA libraries were prepared by the sequencing facility at IFOM/IEO campus according to the protocol described by Blecher-Gonen and colleagues⁴⁰⁴ and sequenced on the Illumina Novaseq platform at a read length of 50 bp paired-end and a coverage of 30 million of reads per sample. Input samples of controls and WS at each stage were pooled together in the same amount before library preparation.

Buffers used for ChIP experiments: i) Nuclei lysis buffer, SDS 1%, EDTA 10 mM, Tris-HCl pH 8.1 50 mM, in ddH₂O; ii) ChIP dilution buffer, Tris HCl pH8 50 mM, NaCl 150 mM, Triton X-100 1%, EGTA 50 mM, EDTA 50 mM, Deoxycholic acid 0,1%, in ddH₂O; iii) RIPA buffer, Deoxycholic acid 0,1%, SDS 0.1%, Triton X-100 1%, Tris-HCl pH8 10 mM, in ddH₂O; iv) RIPA high salt buffer, Deoxycholic acid 0,1%, SDS 0.1%, Triton X-100 1%, Tris-HCl pH8 10 mM, EDTA 0,1 mM, NaCl 140 mM, in ddH₂O; v) LiCl buffer, LiCl 250 mM, NP40 0,5%, Deoxycholic acid 0,5%, EDTA 5 mM, NaCl 50 mM, SDS 1%, in ddH₂O; vi) TE buffer, Tris-HCl pH8, EDTA 1 mM in ddH₂O; vii) Elution buffer, Tris-HCl pH 7,5 20 mM, EDTA 5 mM, NaCl 50 mM, SDS 1%, in ddH₂O.

6.1.12 DNA methylation profiling

Genomic DNA was extracted (LeGene Biosciences, 7400-25) according to manufacturer instructions. Starting from gDNA, bisulfite conversion was carried out with the EZ DNA methylation kit (Zymo research, D5001) according to Illumina's recommended incubations. Converted DNA was subjected to a preamplification step, fragmentation of DNA and hybridisation on a Illumina Infinium EPIC BeadChip (Illumina), which contained allele specific probes recognising either an uracil (generated by the conversion of unmethylated cytosines) or a cytosine (not converted by bisulphite treatment in uracil due to the methyl group). Fluorescent nucleotides were inserted by single base extension and signals were imaged with a HiScan system (Illumina). Methylome profiles were generated by the lab of Prof. Rosanna Weksberg at The Hospital for Sick Children, Toronto and analysed at IEO.

6.1.13 CRISPR/Cas9 engineering

6.1.13.1 Donor construct design for isogenic lines generation

The donor plasmid (WVS02A CONSTRUCT) was designed to introduce the mutation of WVS02A in a control line. 4 restriction enzyme recognition sites flanking the homologous regions were added to allow HAs substitution and use of the plasmid for the engineering of all other mutations of our cohort of patients. 1 Kb HAs were chosen on both sides of the Cas9-mediated cut for every patient. Chosen restriction enzymes were unique in the plasmid and also in the homology arms of all the patients: BstBI, ClaI, NheI and AgeI. Sequences from the last release of the human genome were used (GRCh38.p7). This construct contained also a floxed hPKG promoter - Hygromycin cassette to be engineered in the closest intron in order to select hiPSC where the recombination occurred. The resistance cassette was flanked by two unique restriction enzymes (BamHI and NotI) for future substitution according to the need. The WVS02A-ready-to-use construct was synthesised by GeneArt™ (Thermo Fisher Scientific).

6.1.13.2 sgRNA design

sgRNAs design was performed using the CRISPR design tool of Benchling available at the following website: <https://www.benchling.com/>. For the design, the GRCh38 assembly of the human genome was used. sgRNAs were chosen according to the quality score and the position of the cut. Two scores were evaluated, the on-target and the off-target score defined based on algorithm by Doench et al and Hsu et al^{405,406}. These scores ranged from 0 to 100. The off-target score indicated the inverse probability of the Cas9/gRNA complex to bind sequences in the rest of the genome. The on-target score represented the cleavage efficiency of the Cas9. One gRNA was designed for the EZH2-LOF line generation and four were designed for the engineering of isogenic lines of each mutation of the WS cohort. All the sgRNAs for WS isogenic lines were suitable for the engineering of the mutation in a control line and for the correction of the mutation in patient lines. Moreover, all the

gRNAs showed a middle-high quality because of the presence of many predicted off-targets, although with very low identity (less than 5% with the target region and almost all in non-coding regions).

sgRNA	On-target score	Off-target score	Sequence
sgRNA_EZH2-LOF	35/100	65,9/100	TGTGGTGGATGCAACCTGCA
sgRNA_WVS02A	72,6/100	52,6/100	AGGAGGGTGAAAATGGACTG
sgRNA_WVS01H	64,4/100	69,1/100	CAGATATGCTTATTGGTGAG
sgRNA_WVS03B	50,1/100	79,6/100	ATGCACCCACTATCTTCAGC
sgRNA_WVS04A	59,4/100	59,6/100	GTCTTCTAAACTCATTGTTG

Table 4: List of sgRNAs

6.1.13.3 sgRNAs synthesis

sgRNA were synthesised using the kit GeneArt™ Precision gRNA Synthesis Kit (Thermo Fisher Scientific, A29377) by *in vitro* transcription on a DNA template containing the T7 promoter, the target sequence and the constant part of the tracrRNA. The DNA template was generated by PCR amplification using two overlapping primers: i) the forward, containing the target sequence and a 5' prime overhang containing the sequence of the T7 promoter; ii) the reverse primer, containing the reverse complement of the target sequence and a 5' overhang containing the complementary sequence of the last portion of the constant part of the tracrRNA. The constant portion of the tracrRNA was provided by the manufacturer (**Table 5**). Oligos were synthesised by Sigma-Aldrich.

sgRNA synthesis oligos

sgRNA synthesis oligos	Sequence
Target FW primer sgRNA_EZH2-LOF	TAATACGACTCACTATAGTGTGGTGGATGCAACCTGCA
Target REV primer sgRNA_EZH2-LOF	TTCTAGCTCTAAAACCTGCAGGTTGCATCCACCACA
Target FW primer sgRNA_WVS02A	TAATACGACTCACTATAGAGGAGGGTGAAAATGGACTG
Target REV primer sgRNA_WVS02A	TTCTAGCTCTAAAACCTGCAGGTTGCATCCACCACA
Target FW primer sgRNA_WVS01H	TAATACGACTCACTATAGCAGATATGCTTATTGGTGAG
Target REV primer sgRNA_WVS01H	TTCTAGCTCTAAAACCTGCAGGTTGCATCCACCACA
Target FW primer sgRNA_WVS03B	TAATACGACTCACTATAGATGCACCCACTATCTTCAGC
Target REV primer sgRNA_WVS03B	TTCTAGCTCTAAAACCTGCAGGTTGCATCCACCACA
Target FW primer sgRNA_WVS04A	TAATACGACTCACTATAGGTCTTCTAAACTCATTGTTG
Target REV primer sgRNA_WVS04A	TTCTAGCTCTAAAACCTGCAGGTTGCATCCACCACA

Table 5: sgRNA synthesis oligos. Oligos used as template for sgRNA synthesis

After assembling the PCR reactions according to the manufacturer instructions, the reaction proceeded as follows:

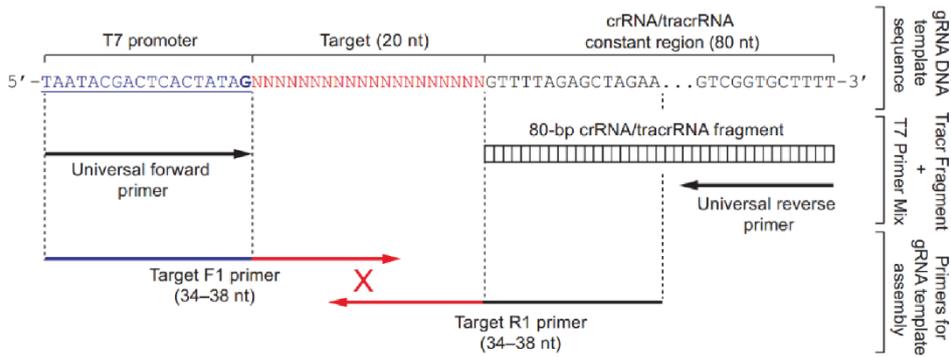


Fig. 110: Schematic representation of the DNA template generation for sgRNA synthesis. Taken from the GeneArt™ Precision gRNA Synthesis Kit manual

In vitro transcription was performed for 4 hours at 37°C using the T7 RNA polymerase and template DNA was removed using DNaseI (both provided with the kit). *In vitro* transcribed RNA was purified using the gRNA Clean Up reagents of the kit. sgRNAs were loaded on an denaturing gel as a control of the procedure. The concentration was determined using a NanoDrop spectrophotometer.

6.1.13.4 CAS9 purified protein

CRISPR/Cas9 conditions were standardised using a Cas9 purified by Cogentech biochemistry facility (Cogentech SRL). The latter was available solubilised in transduction buffer (5x Transduction buffer: 500 mM NaCl, 25 mM NaH₂PO₄, 250 mM NDSB-201, 150 mM glycerol, 75 mM glycine, 1.25 mM MgCl₂, 1 mM 2-mercaptoethanol at pH 8.0 in milliQ water⁴⁰⁷).

6.1.13.5 CRISPR/Cas9 delivery into hiPSC

The Neon Transfection System (Thermo Fisher Scientific) with the 100 µL kit (Thermo Fisher Scientific, MPK10096) was used to electroporate hiPSC. 70-80 % confluent cells were pre-treated for 2-4 hours with 10 µM ROCK inhibitor to enhance cell-survival after electroporation. At the moment of the electroporation, hiPSC were detached using Accutase solution (1 mL x 10 cm plate), centrifuged for 3 minutes at 160g with a standard centrifuge. Cells were then resuspended in PBS and counted with an automated cell counter. In the meantime, the ribonucleic complex Cas9/sgRNA (molar ratio: 1:2,58 or 1:4 during the last electroporation) was assembled and incubated 10 minutes at 37°C to enhance the complex formation. Once formed, the complex was stable for two hours. For each electroporation reaction, 4x10⁵ cells, 1.5 µg donor plasmid and corresponding amounts of Cas9/gRNA complex were used (10 µg of Cas9 each 4x10⁵ cells). Cells were resuspended in electroporation buffer T (provided by the kit) in order to have 4x10⁵ cells each 50 µL. The donor plasmid was added to the Cas9/gRNA complex for the HDR-mediated integration of the WVS02A mutation in a control line, while only the Cas9/gRNA complex was added for generating the EZH2-LOF line. The mixture was added to 50 µL of cells and 120 µL volume was reached adding buffer T. The Neon pipet was used to slowly introduce the electroporation mix in the 100 µL tip avoiding bubbles formation. The 20 µL excess was necessary to avoid bubble formation. The tip was

introduced into the electroporation station containing 3 mL of buffer E2 (provided with the kit). Electroporation was performed with the following conditions: amplitude of the pulse: 900 V, pulse width: 20 ms, n° of pulses: 3. Electroporated cells were directly plated on matrigel coated dishes in TeSR-E8 supplemented with 10 μ M ROCK inhibitor. A total $3,2 \times 10^6$ cells were electroporated in rounds of 4×10^5 and plated in 15 cm in a dish to allow clonal growth.

6.1.13.6 Hygromycin selection and clones isolation

Hygromycin B (Merck, 31282-04-9) selection was carried out after 48h from the electroporation for the WVS02A isogenic line using a final concentration of 50 μ g/mL for 8 days. This concentration was previously standardised in the lab as the optimal for hiPSC. Cells edited with EZH2-LOF sgRNA were not subjected to selection. After 10-15 days colonies were big enough for manual picking. Clones were transferred in Eppendorf tubes, centrifuged 160g for 10' and then the pellet was resuspended in TeSR-E8 supplemented with 5 μ M ROCK inhibitor for plating in 48 well plate for expansion dedicated to screening and freezing.

6.1.13.7 Genomic DNA extraction and screening

Cells were incubated ON at 60°C with Bradley Lysis Buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0,5% SDS, 10 mM NaCl in ddH₂O) containing Proteinase K (1 mg/ml). The day after, ice-cold EtOH/NaCl mix (EtOH 100%, NaCl 5M) was added to precipitate the DNA and the cells were incubated 30' at -80°C and then centrifuged 20' at 3000 rpm. The pellet was then resuspended twice in cold 70% EtOH and centrifuged 10' at 3000 rpm. Once EtOH was completely removed, DNA was eluted in 30 μ l of warm TE buffer (Tris-HCl pH8, EDTA 1 mM) pH 8.0. DNA was then incubated 10' at 56°C. Nanodrop spectrophotometer was used for quantification. 25 ng of genomic DNA were used for PCR amplification. The screening strategy for the EZH2-LOF line was performed by PCR and Sanger sequencing. For EZH2-LOF, primers around the mutation site were designed and are reported in **Table 6**.

	Sequence
FW	ACAATAGTGTGTTCTTCCAAATGTC
REV	CTTGTTACATAACAAACAACTATCC

Table 6: Primers screening EZH2-LOF line

The screening strategy for the WVS02A isogenic line was performed by sequential PCRs on clones to verify the integration of the cassette in the correct locus. First, clones were subjected to amplification with two primers annealing on the resistance cassette (amplicon length: 1357 bp; primers in **table 7**). Positive clones were then screened for the correct integration at the 5' (amplicon length: 2110 bp; **table 8**) and at the 3' (amplicon length: 2472 bp; **table 9**) using couples of primers annealing outside and inside the donor construct. Lastly, clones with the correct integration site were amplified with two primers designed in the HAs and spanning the whole

construct (amplicon length: 2535bp for the edited allele and 684 bp for the WT allele; primers in **table 10**) as test of heterozygosity. Amplicons obtained from the PCRs performed to understand whether the construct was engineered in the correct genomic locus were subjected to Sanger sequencing to confirm the presence of the mutation and the absence of unwanted rearrangements. Moreover, the amplicon of the WT allele was also subjected to Sanger sequencing.

All oligos were ordered by Sigma-Aldrich. All Sanger sequencing experiments were performed by Cogentech SRL.

	Sequence
FW	CATTCTTCACGTCCGTTTCGC
REV	ATTTGTGTACGCCCGACAGT

Table 7: Screening WVS02A isogenic 1. Primers amplifying the resistance cassette.

	Sequence
FW	CAGGCTCCCATTAACCTGCT
REV	ATAGGTCAGGCTCTCGCTGA

Table 8: Screening WVS02A isogenic 2. Primers amplifying the 5' integration site

	Sequence
FW	ACGTCCTGCGGGTAAATAGC
REV	AAGGGTCACTACAGCCAAGTT

Table 9: Screening WVS02A isogenic 3. Primers amplifying the 3' integration site

	Sequence
FW	CAGGCTCCCATTAACCTGCT
REV	AAGGGTCACTACAGCCAAGTT

Table 10: Screening WVS02A isogenic 4. Primers for heterozygosity test

6.2 Bioinformatic procedures

6.2.1 BrainSpan

Data from the BrainSpan Atlas were downloaded from here: <http://www.brainspan.org/static/download.html>. Dataset organisation was described in the following white paper: [https://help.brain-map.org/display/devhumanbrain/Documentation,Developmental Transcriptome](https://help.brain-map.org/display/devhumanbrain/Documentation,Developmental+Transcriptome). Data were only available already pre-processed (RPKM values).

6.2.2 Dimensionality reduction of BrainSpan pre-natal cortical samples

Raw data from BrainSpan were selected for pre-natal cortical specimens for a total of 162 samples. Genes with expression levels greater than 1 RPKM in at least 25% of the observations were considered for further analysis (16824 genes). Principal component analysis was calculated using the `prcomp` R function and PC1-PC2 were visualised. Gene loadings for PC1 and PC2 were retrieved from this analysis and the top 35 ones with positive and negative scores were visualised as lollipops. The top 300 genes with the highest positive loading and the top 300 with the highest negative loading were selected to perform GO analysis for biological processes using the TopGO package⁴⁰⁸. Fisher Test and Weight01 method were used to select significant categories to account for ontology hierarchy. PValue threshold for significance: 0.01. The totality of genes used to compute PCA (16824 genes) were used as universe. Results were visualised as bar plots (`ggplot2`).

6.2.3 Correlation pre- and post-natal cortical samples BrainSpan

Raw data from BrainSpan were selected for pre-natal (n = 162) or post-natal cortical samples (n = 200). Genes with RPKM lower than 1 in at least 75% of samples were discarded (resulting in 16824 genes). The mean expression for each gene at each stage was calculated considering protein-coding genes and taking into account all cortical sub-areas. Correlation between developmental stages was computed using Spearman metrics and visualised as heatmaps (`Pheatmap` package).

6.2.4 BrainSpan WGCNA

Pre-natal cortical samples were selected (162) and genes with RPKM lower than 1 in at least 75% of samples were discarded. Importantly, since PCW 25, 26 and 35 included a very low number of samples, they were excluded from this analysis, resulting in 157 samples. Coefficient of variation was calculated on log-transformed expression data and selection of the genes with the highest CV was applied (35% of the total, 5889 genes). The WGCNA R package was used to generate a signed co-expression network²⁷². Biweight mid-correlation was used to calculate the correlation matrix. Scale-free topology criterion was used to choose a soft correlation threshold β of 18 and the correlation matrix was elevated to the power of β to obtain the adjacency matrix, which was in turn used to calculate the topological overlap measure. The dissimilarity matrix relative to the topological overlap measure was used for average-linkage hierarchical clustering and dendrogram

generation. 17 gene modules were defined as branches of the dendrogram using the DynamicTree Cut algorithm (deepSplit=1; minimum cluster size= 50; PAM stage TRUE; cutHeight 0.998⁴⁰⁹). Every module was correlated with stage as phenotypic trait. To do so, module eigengenes, the first principal component of a given module, were calculated and correlated using Spearman metrics to stage as a continuous variable or as a series of categorical variables. ME of interesting modules were visualised through developmental time to analyse their trends. GO analysis for biological processes for all the genes of each module was performed using TopGO. Fisher Test and Weight01 method were used to select significant categories to account for ontology hierarchy. PValue 0.01 and enrichment of 1.75 were used as significance thresholds. Enrichment score was calculated as ratio between the number of genes of a given module associated to a particular category and the number of genes associated to that category expected to be found by chance in the module. As universe for GO, the number of selected genes for network generation was chosen (5889 genes). Results were visualised as bubble plots (ggplot2). Node and edges of the top-75 genes of each module, defined according to intramolecular connectivity with minimum edge weight of 0.2, were exported using the dedicated function of the WGCNA R package and imported in Cytoscape (version 3.8.2). Betweenness, closeness, eigenvector and degree were considered as centrality indices and were computed using CytoNCA²⁷³.

6.2.5 RNA-seq alignments for CBO, NCSC, iN, internal fetal cortical dataset

RNAseq FASTQ data were quantified using Salmon⁴¹⁰ (version 0.8.2) at the gene level. GRCh38 Genecode 27 was used as reference for quantification and annotation.

6.2.6 Dimensionality reduction CBO dataset and internal fetal samples

A total of 66 samples considering CBO and internal fetal samples were used to perform PCA. Gene filtering using a threshold of 2 counts per million reads (cpm) in at least 2 samples was used, resulting in 17759 analysed. PCA was computed using prcomp R function. PC1 and PC2 were visualised.

6.2.7 Dimensionality reduction CBO dataset

A total of 43 samples considering CBO dataset (including the hiPSC stage). Gene filtering using a threshold of 2 cpm in at least 4 samples was used, resulting 16901 genes used for the analysis. PCA was computed using prcomp R function. PC1 and PC2 were visualised. Gene loadings for PC1 and PC2 were retrieved from this analysis and the top 35 with positive and negative scores were visualised as lollipops. The top 150 with the highest positive loading and the top 150 with the highest negative loading were selected to perform GO analysis for biological processes using TopGO. Fisher Test and Weight01 method were used to select significant categories to account for ontology hierarchy. PValue threshold for significance: 0.01. The totality of genes used to compute PCA (16901 genes) were used as universe. Results were visualised as bar plots (ggplot2).

6.2.8 DEA on CBO dataset and functional annotation of DEGs

DEAs were performed between CBO at different stages of differentiation versus the previous time-point using edgeR. Genes with expression levels higher than 2 cpm in at least 4 samples were tested for differential expression (16901 genes). The information about lines was used as a covariate for DEGs determination. DEGs were determined using generalised linear regression methods. The option “robust” was used for the estimation of dispersion. DEGs with $FDR < 0.05$ and absolute $FC > 2$ were considered for further analysis. DEGs for every comparison were divided in up- and down-regulated and number of DEGs were visualised as bar plots (ggplot2) to visualise the entity of the difference between stages. GO analysis for biological processes was performed for every comparison dividing genes in up- and down-regulated. The analyses were performed using TopGO. Fisher Test and Weight01 method were used to select significant categories to account for ontology hierarchy. PValue 0.01 and enrichment of 2 were used as significance thresholds. As universe for GO, 16879 genes were used. The inconsistency between the number of genes used for DEA (16901) and the universe for GO was due to the presence of a little subset of genes with duplicated gene symbol, used by the TopGO function, which were eliminated. These genes had different Ensembl gene names, which was used for DEA. Results were visualised as bubble plots grouping the comparisons in before day 100 and after day 100 (ggplot2).

Visualisation of common DEGs between the different sequential CBO comparisons was performed using scatter plots (ggplot2). The \log_2FC of all tested genes was visualised. DEGs between sequential CBO comparisons (e. g. day 50vs25 against day 100vs50) were intersected and differentiated in genes not significant, significant in one of the comparisons or significant in both. In this way, the behaviour of genes in common between the two DEAs was analysed, thus finding genes upregulated in both, downregulated in both or upregulated in one and downregulated in the other. The top 10 genes in terms of absolute fold change for the four types of behaviours were labelled in the plot.

6.2.9 Bulk deconvolution CBO dataset

Proportions of cell populations in CBO at different time-points were estimated applying a deconvolution approach based on the SCDC algorithm³³⁶ using as reference a scRNAseq dataset of the developing human cortex³³⁷. Bulk-RNAseq data not filtered for lowly expressed genes were used as input for the SCDC algorithm. The single cell raw count expression matrix was filtered to discard low quality cells: cells with mitochondrial RNA content $< 5\%$; ribosomal protein RNA content $< 50\%$; detected genes > 450 and < 3000 ; UMI counts > 750 and < 10000 from the scRNAseq reference were considered for the deconvolution. Starting from the clustering annotation performed in the original work, pericytes, microglia, endocytes and oligodendrocyte progenitors were discarded as we did not expect to find these cell type in CBO. The remaining cell populations were then grouped in the

following categories: aRG, oRG, IP, cycling progenitors, interneurons, excitatory neurons. Cells with questionable cell-type assignments were removed (SCDC qc threshold = 0.65). Cell proportions were retrieved using the SCDC_prop function and were visualised as boxplots showing the proportion of every cell type per stage (ggplot2).

6.2.10 Cortex-specific genes determination and visualisation of their behaviour in CBO

DEAs of the different fetal tissues from our internal cohort versus hiPSC were performed for all tissues including at least two samples using edgeR. Genes with expression levels higher than 2 cpm in at least 2 samples were tested for differential expression (17275 genes). DEGs were determined using generalised linear regression methods. The option “robust” was used for the estimation of dispersion. DEGs with FDR < 0.05 and absolute log₂FC > 3 were considered for further analysis. This very stringent threshold on fold change was applied due to the profound difference between contrasted samples to consider only very robust DEGs. Cortex-specific DEGs were determined by subtracting to DEGs of the cortex vs hiPSC comparison DEGs found in the comparisons between other tissues and hiPSC. GO analysis for biological processes was performed for cortex-specific genes dividing them in up- and down-regulated. This analysis was performed using TopGO. Fisher Test and Weight01 method were used to select significant categories to account for ontology hierarchy. PValue 0.01 and enrichment of 2 were used as significance thresholds. As universe for GO, 17275 genes were used. Results were visualised as bubble plots (ggplot2).

The behaviour of cortex-specific DEGs in the CBO dataset, including hiPSC, was visualised by heatmap (Pheatmap package) using average-linkage hierarchical clustering for rows. Row cluster dendrogram for both heatmaps was retrieved and k = 7 or k = 6 for upregulated or downregulated genes was used, respectively, to define clusters with the CutTree function. Clusters of upregulated cortex specific DEGs were grouped according to their behaviour in the CBO dataset in genes stably expressed in CBO or increasingly upregulated along CBO differentiation. The same approach was applied for downregulated cortex specific DEGs defining the following three groups: i) cortex-specific downregulated genes increasingly repressed along CBO differentiation; ii) cortex-specific downregulated DEGs increasingly upregulated along CBO differentiation; iii) cortex-specific downregulated genes transiently upregulated along CBO differentiation. GO analysis for biological processes was performed for every group. This analysis was performed using TopGO. Fisher Test and Weight01 method were used to select significant categories to account for ontology hierarchy. PValue 0.01. As universe for GO, 17275 genes were used. Results were visualised as bar plots (ggplot2).

6.2.11 CBO WGCNA

The same pipeline used to perform WGCNA on BrainSpan was applied also to the CBO dataset. The analysis was performed on a total of 39 samples with *in vitro* age spanning from day 25 to day 200.

Filtering was applied by keeping genes with an expression of at least 2 cpm in at least 7 samples, which is the minimum number of samples per stage, in order not to discard genes that were time-point specific. Coefficient of variation was calculated on log-transformed data and was set to 0.5, resulting in a total of 7831 genes considered for the analysis. The soft threshold β was set to 15 using the scale-free topology criterion. The DynamicTree Cut algorithm parameters used for gene module identification were DeepSplit of 1; minClusterSize 30; PAM stage TRUE; cutHeight 0.999), for a total of 14 modules that were characterised using the same pipeline applied for the BrainSpan WGCNA.

6.2.12 Download and alignment other brain organoid datasets

Data for CO, FO and TA were downloaded from Gene Expression Omnibus using the relative article accession numbers (GSE82022, GSE80073, GSE61476, respectively) . RNAseq FASTQ data were quantified using Salmon (version 0.8.2) at the gene level. GRCh38 Genecode 27 was used as reference for quantification and annotation.

6.2.13 Dimensionality reduction for CO, FO and TA datasets

A total of 8, 11 and 27 samples for CO, FO and TA, respectively, were used to perform PCA, separately for each dataset. Gene filtering using a threshold of 2 cpm in at least 2 samples was used, resulting in 15339, 16585 and 16522 genes for CO, FO and TA, respectively, used for the analysis. PCA was computed using prcomp R function. PC1 and PC2 were visualised.

6.2.14 Bulk deconvolution of CBO and other brain organoid datasets

The same approach used for the deconvolution of CBO was applied for all brain organoid datasets with the difference that in this case only two broad categories were deconvolved. This allowed to observe general variations in broad populations (Progenitors: cycling progenitors + ventricular radial glia); Neurons (excitatory neurons + inhibitory neurons) between protocols.

6.2.15 Correlation CBO versus other brain organoid datasets

To perform correlation analysis between CO, FO or TA versus CBO, FPKM were calculated for all datasets and filtering was applied keeping genes with FPKM greater than 1. The analysis was performed only on protein coding genes. Correlation was computed for each dataset using Spearman metrics and visualised as separated heatmaps for each comparison (Pheatmap package).

6.2.16 Correlation organoid datasets versus BrainSpan

To perform correlation analysis between CBO, CO, FO or TA versus BrainSpan pre-natal cortex, FPKM for all brain organoid datasets were calculated due to the availability of BrainSpan data as RPKM. Filtering was applied to exclude genes mean expression level lower than 1 RPKM in at least 75% of samples (BrainSpan) and 1 FPKM (Brain Organoids). The analysis was performed only on protein coding genes. Mean expression per gene per stage was calculated for all BrainSpan pre-

cortical samples. Correlation was computed using Spearman metrics and visualised as separated heatmaps for each comparison (Pheatmap package).

6.2.17 DEA on CO, FO and TA datasets

DEAs were performed for CO, FO and TA independently. For every dataset every stage of differentiation was compared against the previous one using edgeR. Genes with expression levels higher than 2 cpm in at least 2 samples were tested for differential expression for all datasets (CO: 15339; FO: 16585; TA: 16522 genes). The information about lines was used as a covariate for DEGs determination for the TA dataset, while for CO and FO only one line was available. DEGs were determined using generalised linear regression methods. The option “robust” was used for the estimation of dispersion. DEGs with FDR < 0.05 and absolute FC > 2 were considered for further analysis. DEGs for every comparison were divided in up- and down-regulated and number of DEGs were visualised as bar plots to visualise the entity of the difference between stages (ggplot2 package).

6.2.18 Literature-curated gene signatures visualisation in BrainSpan and Brain organoids

To evaluate expression levels selected gene signatures, LogFPKM expression values were calculated using edgeR for all brain organoid datasets. Genes with duplicated gene names were excluded. RPKM of BrainSpan prenatal cortical samples were log₂ transformed. Values for the gene signatures of interest were retrieved and the mean expression was calculated for each stage of every dataset. Finally, mean expression levels for every gene at every stage were plotted as lollipops (ggplot2).

6.2.19 Module overlap of BrainSpan and CBO WGCNA

Module assignment for WGCNA and CBO were retrieved. Genes shared across the two networks were selected for the analysis (2643) and modules of interest were restricted to the genes represented in the universe. Number of shared genes, odds ratio and PValue across CO and BrainSpan modules were calculated by the GeneOverlap R package. Results were visualised as dot plot (ggplot2) where numbers (shared genes) were shown for odds ratio >1, dots were shown for those having also PValue < 0.05. Color-code was assigned according to OR, dot size varied according to PValue.

6.2.20 Analysis of BrainSpan modules in brain organoids

Module eigengene for each BrainSpan WGCNA module of interest was calculated in all organoid datasets as a prediction (R function predict) based on the module eigengene of BrainSpan itself. Results were visualised in ribbon plots (ggplot2) showing first principal component coefficients for each module along BrainSpan and organoid developmental time-points.

6.2.21 Analysis of CBO modules in BrainSpan and other brain organoids

Module eigengene for each CBO WGCNA module of interest was calculated in BrainSpan and in all other organoid datasets as a prediction (R function predict) based on the module eigengene of CBO itself. Results were visualised in ribbon plots (ggplot2) showing first principal component coefficients for each module along BrainSpan and organoid developmental time-points.

6.2.22 DEAs Weaver CBO and DEGs characterisation

DEAs were performed between control and Weaver CBO at the three time-points analysed using edgeR. Genes with expression levels higher than 2 cpm in at least 8 samples were tested for differential expression (Day 25: 13664, Day 50: 14493, Day 100: 14767). Batch correction was performed using the surrogate variable (SV) analysis (SVA) package (SV day 25: 5; SV day 50: 4; SV day 100: 5). DEGs were determined using generalised linear regression methods. The option “robust” was used for the estimation of dispersion. DEGs with FDR < 0.05 for day 25 and FDR < 0.1 for day 50 and day 100 were considered for further analysis. DEGs for every comparison were divided in up- and down-regulated and GO analysis for biological processes was performed using TopGO. Fisher Test and Weight01 method were used to select significant categories to account for ontology hierarchy. PValue 0.01. As universe for GO, 13664, 14493, 14767 genes were used for day 25, 50 and 100 comparisons, respectively. Results were visualised as bubble plots grouping the comparisons in before day 100 and after day 100 (ggplot2). GSEA was performed on genes with PValue < 0.05 ranked by fold change using the fgsea R package and visualising enriched H1 packages with PValue adjusted < 0.05 and enrichment score greater than 2. DEAs were also performed considering the EZH2-LOF line versus its isogenic control at the three time-points using the same approach applied for patients. Genes with expression levels higher than 2 cpm in at least 2 samples were tested for differential expression (Day 25: 13946, Day 50: 14459, Day 100: 15176). DEGs with FDR < 0.05 and absolute FC > 2 were considered for further analyses. GO and GSEA were performed as for patients.

For the stage-wise analysis considering controls and WS samples separately, it was performed considering stage as a continuous variable and including hiPSC. Genes resulting upregulated in both groups at day 25 that underwent silencing later in controls but not in WS were extracted using average-linkage hierarchical clustering and the distribution of their global fold change was visualised. This group of genes was then overlapped with H3K27me3 peaks in hESC. GO was performed on the overlapping genes using the same approach applied previously in this work.

For the analysis subdividing WS in groups based on the EZH2 functionality, we performed DEA using a series of 4 categorical variable indicating controls, classic WS, putative EZH2 gain of function and EZH2-LOF and compared classical WS vs controls using edgeR. PCA, GO and master regulator analysis was performed on DEGs with FDR < 0.01 and absolute FC > 2 using as universe the genes

filtered for low-expression. GO was performed with TopGO and master regulator analysis with the VIPER package⁴¹¹. SVA was applied with 3 SV.

6.2.23 DEGs overlap and visualisation

Visualisation of common DEGs between the WS vs control and EZH2-LOF vs isogenic control comparisons was performed using scatter plots (ggplot2). The log₂FC of all tested genes was visualised. DEGs between WS vs controls and EZH-LOF vs isogenic control for every stage (e. g. day 25 WS vs controls against day 25 EZH2-LOF vs isogenic control) were intersected and differentiated in genes not significant, significant in one comparison or significant in both. In this way, the behaviour of genes in common between the two comparisons was analysed finding genes upregulated in both, downregulated in both or upregulated in one and downregulated in the other comparison. The size of the overlap was visualised as venn diagrams (VennDiagram R package) and top 10 genes upregulated or downregulated in both comparisons were extracted.

6.2.24 H3K27me3 ChIP-seq analysis

ChIP-seq reads alignment was performed on the human hg38 genome using Bowtie 1.0 (-v 2 -m 1). Peak calling was performed with Homer, using broad parameters (histone). Logical analysis of the peaks was performed via BedTools 2.26 (multiIntersect function) to identify lost and gained peaks specifically absent or present in WS samples. Lost peaks were recognised as the ones present in n-1 control and absent in all WS or present in all controls and at the most in one WS at a given stage. Conversely, gained peaks were recognised as peaks present in n-1 WS and absent in all control or present in all WS and at the most in one control at a given stage. To independently validate the quality of our data we selected representative tissues and gathered all peaks passing FDR 0.1 from the Roadmap Epigenomics dataset, for H3K27me3 and major regulatory regions marks (H3K4me1, H3K4me1, H3K27ac, etc). Specifically, we used E003 (H1 hESC), E007 for NPC (H1 derived neural precursors), E071 for cortical neurons (H9 derived), and E071 for hippocampal neurons. We visualised the overlap between CBO H3K27me3 peaks (FDR < 0.1) and the external datasets. Peak-gene association was performed using the intersectBed function (Bedtools), taking as reference i) cell-type specific 4DGenome enhancer (hippocampal neurons and or neural progenitors depending on the most enriched tissue in each stage), or ii) promoters, defined as regions spanning from 500bp before and 250bp after the TSS of each gene.

6.2.25 DNA methylation analysis

Raw DNA methylation data were analysed by the minfi 1.30 R package⁴¹². M-Values, defined as log₂ ratio between the intensities of the methylated probe versus the unmethylated one for each CpG, were retrieved after raw data pre-processing through SWAN normalisation. Beta-values, defined ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities), were calculated starting from M-Values using the m2beta

function. PCA was performed using the `prcomp` R function. For differential methylation analysis and visualisation purpose, the `mCSEA 1.0.1` R package was used. Promoters with $FDR < 0.05$ were selected as differentially methylated.

7. LIST OF PUBLICATIONS

7.1 Published papers

1. Cavallo F*, Troglia F*, Fagà G, Fancelli D, Shyti R, **Trattaro S**, Zanella M, D'Agostino G, Hughes JM, Cera MR, Pasi M, Gabriele M, Lazzarin M, Mihailovich M, Kooy F, Rosa A, Mercurio C, Varasi M, Testa G. High-throughput screening identifies histone deacetylase inhibitors that modulate GTF2I expression in 7q11.23 microduplication autism spectrum disorder patient-derived cortical neurons. *Mol Autism*. 2020 Nov 19;11(1):88. doi: 10.1186/s13229-020-00387-6. PMID: 33208191; PMCID: PMC7677843.
2. López-Tobón A*, **Trattaro S***, Testa G. The sociability spectrum: evidence from reciprocal genetic copy number variations. *Mol Autism*. 2020 Jun 16;11(1):50. doi: 10.1186/s13229-020-00347-0. PMID: 32546261; PMCID: PMC7298749.
3. Drakulic D*, Djurovic S, Syed YA, **Trattaro S**, Caporale N, Falk A, Ofir R, Heine VM, Chawner SJRA, Rodriguez- Moreno A, van den Bree MBM, Testa G, Petrakis S, Harwood AJ. Copy number variants (CNVs): a powerful tool for iPSC-based modelling of ASD. *Mol Autism*. 2020 Jun 1;11(1):42. doi: 10.1186/s13229-020-00343-4. PMID: 32487215; PMCID: PMC7268297.
4. Zanella M*, Vitriolo A*, Andirko A, Martins PT, Sturm S, O'Rourke T, Laugsch M, Malerba N, Skaros A, **Trattaro S**, Germain PL, Mihailovic M, Merla G, Rada-Iglesias A, Boeckx C, Testa G. Dosage analysis of the 7q11.23 Williams region identifies BAZ1B as a major human gene patterning the modern human face and underlying self- domestication. *Sci Adv*. 2019 Dec 4;5(12):eaaw7908. doi: 10.1126/sciadv.aaw7908. PMID: 31840056; PMCID: PMC6892627.
5. López-Tobón A*, Villa CE*, Cheroni C*, **Trattaro S**, Caporale N, Conforti P, Iennaco R, Lachgar M, Rigoli MT, Marcó de la Cruz B, Lo Riso P, Tenderini E, Troglia F, De Simone M, Liste-Noya I, Macino G, Pagani M, Cattaneo E, Testa G. Human Cortical Organoids Expose a Differential Function of GSK3 on Cortical Neurogenesis. *Stem Cell Reports*. 2019 Nov 12;13(5):847-861. doi: 10.1016/j.stemcr.2019.09.005. Epub 2019 Oct 10. PMID: 31607568; PMCID: PMC6893153.

*First author

7.2 Book chapters

1. Alejandro Lopez-Tobon, Nicolò Caporale, **Sebastiano Trattaro**, and Giuseppe Testa; Book: *Stem Cell Epigenetics* (2020); Chapter 11: Three-dimensional models of human brain development (257-278); Elsevier, eBook ISBN: 9780128140864.

7.3 Awarded grants

1. Jan 2020-Dec 2022: 3 years FIRC/AIRC fellowship "Assunta Lombardelli Giuliano Mordini" on the project "Organoid-based dissection of the master oncogenic role of GTF2I in thymic neoplasms".

7.4 Submitted papers

1. Serena Buontempo*, Pasquale Laise*, James M. Hughes*, **Sebastiano Trattaro***, Vivek Das, Chantal Rencurel, Giuseppe Testa. EZH2-mediated H3K27me3 targets transcriptional circuits of neural differentiation; Under revision, Scientific Reports.
2. Nicolò Caporale*, Michelle Leemans*, Lina Birgersson*, Pierre-Luc Germain*, Cristina Cheroni*, Gábor Borbély, Elin Engdahl, Christian Lindh, Raul Bardini Bressan, Francesca Cavallo, Nadav Even Chorev, Giuseppe Alessandro D'Agostino, Vesna Munic Kos, Steven M Pollard, Marco Tullio Rigoli, Erika Tenderini, Alejandro Lopez Tobon, Sebastiano Trattaro, Flavia Troglio, Matteo Zanella, Åke Bergman, Pauliina Damdimopoulou, Maria Jönsson, Wieland Kiess, Efthymia Kitraki, Hannu Kiviranta, Eewa Nånberg, Mattias Öberg, Panu Rantakkoko, Christina Rudén, Olle Söder, Carl-Gustaf Bornehag, Barbara Demeneix, Jean-Baptiste Fini, Chris Gennings, Joëlle Rüegg, Joachim Sturve and Giuseppe Testa. From Cohorts to Molecules: Adverse Impacts of Endocrine Disrupting Mixtures. Under revision, Science.

*First author

8. ACKNOWLEDGEMENTS

Giuseppe Testa's laboratory is an interdisciplinary environment where scientific projects are carried out with a great collaborative spirit that was key for my professional growth as a scientist.

I wish to thank Giuseppe for guiding me through my PhD not only as a boss, but rather as a leader. Thanks to his mentorship and teachings I had the opportunity to undertake a path of all-round growth in terms of project design and execution as well as in terms of scientific vision. He allowed me to be involved in a multiplicity of scientific projects that he carries out in his lab and that led me to develop my practical skills about the most cutting-edge technologies as well as my critical thinking and approach to research, thus giving me the chance of appreciating in full the topics and the field of disease-modelling, also with world-wide collaborations.

I would like to thank Alejandro López-Tobón, who provided me my initial training in Giuseppe's laboratory and with whom we developed ideas and projects that were and are still being realised in practice. With our fruitful discussions I could shape my capabilities as researcher and with our friendship I could enjoy our research.

Together with Alejandro and Nicolò Caporale, we imported and standardised the differentiation of brain organoids in our lab. Our work as the founder "organoiders team" brought us to develop independent projects in a highly collaborative and friendly environment that facilitated troubleshooting and advancement.

I would like to thank Cristina Cheroni, who shared with me all her experience in bioinformatics giving me the opportunity of developing my hybrid skills. Cristina and I, together with Alejandro and Nicolò, carried out the organoid benchmarking and characterisation, which constitutes a big part of this work. I would like to thank her not only as a great colleague, but also as a sincere friend.

A special thank also to Alessandro Vitriolo, who greatly contributed to the development of my bioinformatic skills and with whom we are carrying out one to one the dissection of the molecular underpinnings of Weaver syndrome, the other project presented in this work.

I would like to thank Erika Tenderini who helped with library preparation and hiPSC handling as well as with all organisational activities. I would like to thank her also for our friendship and long discussions about life vision and for her constant support.

I wish to thank Sara Sebastiani, a highly talented master student that I had the opportunity to supervise and a good friend, who developed and carried out with me the engineering of isogenic lines and the phenotypic profiling of patient-derived brain organoids.

I wish to thank Pietro Lo Riso who carried out the DNA methylation analysis and shared with me his knowledge and pipelines. I would also like to thank him for his advices and for his friendship.

I wish to thank Elena Signaroldi, for her constant support in project management and for being always available to help and provide advices.

I wish to thank Flavia Troglio, for her every day support of lab activities and for sharing with me her wet experience.

I would like to thank Michele Gabriele, who performed the reprogramming of patient-derived fibroblasts.

I wish to thank Patricio Fuentes Bravo, who thought me how to perform ChIP.

I wish to thank Rosanna Weksberg and Saana Choufani for performing the DNA methylation profiling.

I wish to thank Davide Pozzi for carrying out electrophysiology of iN.

I wish to thank Raul Bressan and Steven Pollard for providing primary fetal brain tissues.

I would also like to thank all the members of the Testa lab for their everyday support and friendship, especially Martina Pezzali, who recently joined the lab as a PhD student and who will work on the Weaver project with me.

I would also like to thank Stefania Faletti, with whom I shared my experience within the SEMM PhD programme.

I wish to thank the SEMM PhD office for their administrative support.

I wish to thank Cecilia Toscani, Camilla Galli and Francesca Fiore from the ENABLE network, with whom I shared the experience of organising an international symposium.

I wish to thank Madeline Lancaster and Diego Pasini, who accepted to supervise me as external and internal supervisors of my PhD, respectively, and who provided very useful feedbacks along my PhD.

I wish to thank the examiners of this thesis, Martin Schaefer and Jürgen Knoblich, who accepted to evaluate my PhD work and to provide feedbacks.

Last but not least, a special thank goes my girlfriend Serena, who supports and understands me every day and to my parents, who support me constantly and closely from very far away.

9. BIBLIOGRAPHY

1. Desmond, M. E. & O’Rahilly, R. The growth of the human brain during the embryonic period proper. *Anat. Embryol. (Berl.)* (1981).
2. Stiles, J. The fundamentals of brain development: Integrating nature and nurture. (2008).
3. Copp, A. J. Greene ND, Murdoch JN. *The genetic basis of mammalian neurulation. Nat Rev ...* (2003).
4. Yamada, S. *et al.* Developmental atlas of the early first trimester human embryo. *Dev. Dyn.* **239**, 1585–1595 (2010).
5. Kostović, I. & Jovanov-Milosević, N. The development of cerebral connections during the first 20-45 weeks’ gestation. *Semin Fetal Neonatal Med* **11**, 415–422 (2006).
6. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016).
7. Markram, K. & Markram, H. The intense world theory - a unifying theory of the neurobiology of autism. *Front. Hum. Neurosci.* **4**, 224 (2010).
8. Lopez-Tobon, A., Caporale, N., Trattaro, S. & Testa, G. Three-dimensional models of human brain development. *Stem Cell Epigenetics* (2020).
9. Lopez-Tobon, A., Caporale, N., Trattaro, S. & Testa, G. Three-dimensional models of human brain development. *Stem Cell Epigenetics* (2020).
10. Rubenstein, J. L. R. Annual Research Review: Development of the cerebral cortex: implications for neurodevelopmental disorders. *J. Child Psychol. Psychiatry* **52**, 339–355 (2011).
11. Sur, M. & Rubenstein, J. L. R. Patterning and plasticity of the cerebral cortex. *Science* **310**, 805–810 (2005).
12. Miyashita-Lin, E. M., Hevner, R., Wassarman, K. M., Martinez, S. & Rubenstein, J. L. Early neocortical regionalization in the absence of thalamic innervation. *Science* **285**, 906–909 (1999).
13. O’Leary, D. D. M., Chou, S.-J. & Sahara, S. Area patterning of the mammalian cortex. *Neuron* **56**, 252–269 (2007).
14. Ohkubo, Y., Chiang, C. & Rubenstein, J. L. R. Coordinate regulation and synergistic actions of BMP4, SHH and FGF8 in the rostral prosencephalon regulate morphogenesis of the telencephalic and optic vesicles. *Neuroscience* **111**, 1–17 (2002).
15. Assimacopoulos, S., Grove, E. A. & Ragsdale, C. W. Identification of a Pax6-dependent epidermal growth factor family signaling source at the lateral edge of the embryonic cerebral cortex. *J. Neurosci.* **23**, 6399–6403 (2003).
16. Bishop, K. M. & Rubenstein, J. L. R. Distinct Actions of Emx1, Emx2, and Pax6 in Regulating the Specification of Areas in the Developing Neocortex. *Journal of ...* (2002).
17. Goulburn, A. L., Stanley, E. G., Elefanta, A. G. & Anderson, S. A. Generating GABAergic cerebral cortical interneurons from mouse and human embryonic stem cells. *Stem Cell Res.* **8**, 416–426 (2012).
18. Sussel, L., Marin, O., Kimura, S. & Rubenstein, J. L. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359–3370 (1999).
19. Kwan, K. Y., Sestan, N. & Anton, E. S. Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development* **139**, 1535–1546 (2012).
20. Manzini, M. C. & Walsh, C. A. What disorders of cortical development tell us about the cortex: one plus one does not always make two. *Curr. Opin. Genet. Dev.* **21**, 333–339 (2011).
21. Liu, J. S. Molecular genetics of neuronal migration disorders. *Curr Neurol Neurosci Rep* **11**, 171–178 (2011).

22. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
23. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
24. Olson, E. C. & Walsh, C. A. Smooth, rough and upside-down neocortical development. *Curr. Opin. Genet. Dev.* **12**, 320–327 (2002).
25. Pinto, L. & Götz, M. Radial glial cell heterogeneity--the source of diverse progeny in the CNS. *Prog. Neurobiol.* **83**, 2–23 (2007).
26. Tabata, H. & Nagata, K.-I. Decoding the molecular mechanisms of neuronal migration using in utero electroporation. *Med Mol Morphol* **49**, 63–75 (2016).
27. Campbell, K. & Götz, M. Radial glia: multi-purpose cells for vertebrate brain development. *Trends Neurosci.* **25**, 235–238 (2002).
28. Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).
29. Noctor, S. C., Martínez-Cerdeño, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat. Neurosci.* **7**, 136–144 (2004).
30. Aboitiz, F. & Zamorano, F. Neural progenitors, patterning and ecology in neocortical origins. *Front. Neuroanat.* **7**, 38 (2013).
31. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
32. Corbin, J. G. *et al.* Regulation of neural progenitor cell development in the nervous system. *J. Neurochem.* **106**, 2272–2287 (2008).
33. Striedter, G. F. & Charvet, C. J. Telencephalon enlargement by the convergent evolution of expanded subventricular zones. *Biol. Lett.* **5**, 134–137 (2009).
34. Fish, J. L., Dehay, C., Kennedy, H. & Huttner, W. B. Making bigger brains-the evolution of neural-progenitor-cell division. *J. Cell Sci.* **121**, 2783–2793 (2008).
35. LaMonica, B. E., Lui, J. H., Wang, X. & Kriegstein, A. R. OSVZ progenitors in the human cortex: an updated perspective on neurodevelopmental disease. *Curr. Opin. Neurobiol.* **22**, 747–753 (2012).
36. Molnár, Z. Evolution of cerebral cortical development. *Brain Behav. Evol.* **78**, 94–107 (2011).
37. Tabata, H. & Nakajima, K. Multipolar migration: the third mode of radial neuronal migration in the developing cerebral cortex. *J. Neurosci.* **23**, 9996–10001 (2003).
38. Nadarajah, B., Alifragis, P., Wong, R. O. L. & Parnavelas, J. G. Neuronal migration in the developing cerebral cortex: observations based on real-time imaging. *Cereb. Cortex* **13**, 607–611 (2003).
39. Farhy-Tselnicker, I. & Allen, N. J. Astrocytes, neurons, synapses: a tripartite view on cortical circuit development. *Neural Dev.* **13**, 7 (2018).
40. Miller, F. D. & Gauthier, A. S. Timing is everything: making neurons versus glia in the developing cortex. *Neuron* **54**, 357–369 (2007).
41. Bhaduri, A. *et al.* Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).
42. Martínez-Cerdeño, V. *et al.* Comparative analysis of the subventricular zone in rat, ferret and macaque: evidence for an outer subventricular zone in rodents. *PLoS One* **7**, e30178 (2012).
43. Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
44. Fietz, S. A. *et al.* OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat. Neurosci.* **13**, 690–699 (2010).
45. Andrews, M. G., Subramanian, L. & Kriegstein, A. R. mTOR signaling regulates the morphology and migration of outer radial glia in developing human cortex. *Elife* **9**, (2020).
46. Pollen, A. A. *et al.* Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).

47. Cheung, A. F. P., Pollen, A. A., Tavaré, A., DeProto, J. & Molnár, Z. Comparative aspects of cortical neurogenesis in vertebrates. *J. Anat.* **211**, 164–176 (2007).
48. Rash, B. G. *et al.* Gliogenesis in the outer subventricular zone promotes enlargement and gyrification of the primate cerebrum. *Proc. Natl. Acad. Sci. USA* **116**, 7089–7094 (2019).
49. Farid, S. S., Baron, M., Stamatis, C., Nie, W. & Coffman, J. Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D. *MAbs* **12**, 1754999 (2020).
50. Hammond, R. R. *et al.* A reliable primary human CNS culture protocol for morphological studies of dendritic and synaptic elements. *J. Neurosci. Methods* **118**, 189–198 (2002).
51. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. USA* **112**, 7285–7290 (2015).
52. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37–53 (2016).
53. Spaethling, J. M. *et al.* Primary cell culture of live neurosurgically resected aged adult human brain cells and single cell transcriptomics. *Cell Rep.* **18**, 791–803 (2017).
54. Yan, L. *et al.* Epigenomic landscape of human fetal brain, heart, and liver. *J. Biol. Chem.* **291**, 4386–4398 (2016).
55. Shulha, H. P. *et al.* Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol.* **10**, e1001427 (2012).
56. Lopez-Tobon, A., Caporale, N., Trattaro, S. & Testa, G. Three-dimensional models of human brain development. *Stem Cell Epigenetics* (2020).
57. Siegmund, K. D. *et al.* DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS One* **2**, e895 (2007).
58. Grayson, D. R. *et al.* Reelin promoter hypermethylation in schizophrenia. *Proc. Natl. Acad. Sci. USA* **102**, 9341–9346 (2005).
59. Ninkovic, J. *et al.* The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network. *Cell Stem Cell* **13**, 403–418 (2013).
60. Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc. Natl. Acad. Sci. USA* **109**, 11836–11841 (2012).
61. Mitchell, A., Roussos, P., Peter, C., Tsankova, N. & Akbarian, S. The future of neuroepigenetics in the human brain. *Prog Mol Biol Transl Sci* **128**, 199–228 (2014).
62. Mitchell, A. C. *et al.* The genome in three dimensions: a new frontier in human brain research. *Biol. Psychiatry* **75**, 961–969 (2014).
63. Huang, H.-S., Matevosian, A., Jiang, Y. & Akbarian, S. Chromatin immunoprecipitation in postmortem brain. *J. Neurosci. Methods* **156**, 284–292 (2006).
64. Lopez-Tobon, A., Caporale, N., Trattaro, S. & Testa, G. Three-dimensional models of human brain development. *Stem Cell Epigenetics* (2020).
65. Kilpatrick, T. J. & Bartlett, P. F. Cloning and growth of multipotential neural precursors: requirements for proliferation and differentiation. *Neuron* **10**, 255–265 (1993).
66. Moors, M. *et al.* Human neurospheres as three-dimensional cellular systems for developmental neurotoxicity testing. *Environ. Health Perspect.* **117**, 1131–1138 (2009).
67. Ignatova, T. N. *et al.* Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia* **39**, 193–206 (2002).
68. Lopez-Tobon, A., Caporale, N., Trattaro, S. & Testa, G. Three-dimensional models of human brain development. *Stem Cell Epigenetics* (2020).
69. Bouwman, B. A. M. & de Laat, W. Architectural hallmarks of the pluripotent genome. *FEBS Lett.* **589**, 2905–2913 (2015).
70. De Vos, J., Bouckenheimer, J., Sansac, C., Lemaître, J. M. & Assou, S. Human induced pluripotent stem cells: A disruptive innovation. *Curr. Res. Transl. Med.* **64**, 91–96 (2016).

71. Takahashi, K. & Yamanaka, S. A developmental framework for induced pluripotency. *Development* **142**, 3274–3285 (2015).
72. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
73. Briggs, R. & King, T. J. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc. Natl. Acad. Sci. USA* **38**, 455–463 (1952).
74. Gurdon, J. B. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol* **10**, 622–640 (1962).
75. Okano, H. & Yamanaka, S. iPS cell technologies: significance and applications to CNS regeneration and disease. *Mol. Brain* **7**, 22 (2014).
76. Tanabe, K., Takahashi, K. & Yamanaka, S. Induction of pluripotency by defined factors. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **90**, 83–96 (2014).
77. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
78. Yamanaka, S. Pluripotency and nuclear reprogramming. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **363**, 2079–2087 (2008).
79. Bayart, E. & Cohen-Haguener, O. Technological overview of iPS induction from human adult somatic cells. *Curr Gene Ther* **13**, 73–92 (2013).
80. *Stem cell and gene therapy for cardiovascular disease.* (Elsevier, 2016). doi:10.1016/C2014-0-00701-0
81. Schlaeger, T. M. *et al.* A comparison of non-integrating reprogramming methods. *Nat. Biotechnol.* **33**, 58–63 (2015).
82. Yoshida, Y. & Yamanaka, S. Recent stem cell advances: induced pluripotent stem cells for disease modeling and stem cell-based regeneration. *Circulation* **122**, 80–87 (2010).
83. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat. Biotechnol.* **28**, 848–855 (2010).
84. Merkert, S. & Martin, U. Site-Specific Genome Engineering in Human Pluripotent Stem Cells. *Int. J. Mol. Sci.* **17**, (2016).
85. Wu, M., Chen, G. & Hu, B. Induced pluripotency for translational research. *Genomics Proteomics Bioinformatics* **11**, 288–293 (2013).
86. Nagata, N. & Yamanaka, S. Perspectives for induced pluripotent stem cell technology: new insights into human physiology involved in somatic mosaicism. *Circ. Res.* **114**, 505–510 (2014).
87. Inoue, H., Nagata, N., Kurokawa, H. & Yamanaka, S. iPS cells: a game changer for future medicine. *EMBO J.* **33**, 409–417 (2014).
88. Bain, G., Kitchens, D., Yao, M., Huettner, J. E. & Gottlieb, D. I. Embryonic stem cells express neuronal properties in vitro. *Dev. Biol.* **168**, 342–357 (1995).
89. Li, M., Pevny, L., Lovell-Badge, R. & Smith, A. Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr. Biol.* **8**, 971–974 (1998).
90. Kawasaki, H. *et al.* Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. *Neuron* **28**, 31–40 (2000).
91. Lee, H. *et al.* Directed differentiation and transplantation of human embryonic stem cell-derived motoneurons. *Stem Cells* **25**, 1931–1939 (2007).
92. Watanabe, K. *et al.* Directed differentiation of telencephalic precursors from embryonic stem cells. *Nat. Neurosci.* **8**, 288–296 (2005).
93. Elkabetz, Y. *et al.* Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.* **22**, 152–165 (2008).
94. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
95. Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
96. Qi, Y. *et al.* Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. *Nat. Biotechnol.* **35**, 154–163 (2017).

97. Lopez-Tobon, A., Caporale, N., Trattaro, S. & Testa, G. Three-dimensional models of human brain development. *Stem Cell Epigenetics* (2020).
98. Tcw, J. *et al.* An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells. *Stem Cell Rep.* **9**, 600–614 (2017).
99. Ehrlich, M. *et al.* Rapid and efficient generation of oligodendrocytes from human induced pluripotent stem cells using transcription factors. *Proc. Natl. Acad. Sci. USA* **114**, E2243–E2252 (2017).
100. Abud, E. M. *et al.* iPSC-Derived Human Microglia-like Cells to Study Neurological Diseases. *Neuron* **94**, 278–293.e9 (2017).
101. Zhang, S. C., Wernig, M., Duncan, I. D., Brüstle, O. & Thomson, J. A. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat. Biotechnol.* **19**, 1129–1133 (2001).
102. Eiraku, M. *et al.* Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell* **3**, 519–532 (2008).
103. Kadoshima, T. *et al.* Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc. Natl. Acad. Sci. USA* **110**, 20284–20289 (2013).
104. Paşca, A. M. *et al.* Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* **12**, 671–678 (2015).
105. Mariani, J. *et al.* FOXP1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375–390 (2015).
106. Qian, X. *et al.* Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell* **165**, 1238–1254 (2016).
107. Sloan, S. A., Andersen, J., Paşca, A. M., Birey, F. & Paşca, S. P. Generation and assembly of human brain region-specific three-dimensional cultures. *Nat. Protoc.* **13**, 2062–2085 (2018).
108. Bagley, J. A., Reumann, D., Bian, S., Lévi-Strauss, J. & Knoblich, J. A. Fused cerebral organoids model interactions between brain regions. *Nat. Methods* **14**, 743–751 (2017).
109. Miura, Y. *et al.* Generation of human striatal organoids and cortico-striatal assembloids from human pluripotent stem cells. *Nat. Biotechnol.* **38**, 1421–1430 (2020).
110. Xiang, Y. *et al.* hESC-Derived Thalamic Organoids Form Reciprocal Projections When Fused with Cortical Organoids. *Cell Stem Cell* **24**, 487–497.e7 (2019).
111. Cederquist, G. Y. *et al.* Specification of positional identity in forebrain organoids. *Nat. Biotechnol.* **37**, 436–444 (2019).
112. Cakir, B. *et al.* Engineering of human brain organoids with a functional vascular-like system. *Nat. Methods* **16**, 1169–1175 (2019).
113. Giandomenico, S. L. *et al.* Cerebral organoids at the air-liquid interface generate diverse nerve tracts with functional output. *Nat. Neurosci.* **22**, 669–679 (2019).
114. Karzbrun, E., Kshirsagar, A., Cohen, S. R., Hanna, J. H. & Reiner, O. Human brain organoids on a chip reveal the physics of folding. *Nat. Phys.* **14**, 515–522 (2018).
115. Muguruma, K., Nishiyama, A., Kawakami, H., Hashimoto, K. & Sasai, Y. Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Rep.* **10**, 537–550 (2015).
116. Jo, J. *et al.* Midbrain-like Organoids from Human Pluripotent Stem Cells Contain Functional Dopaminergic and Neuromelanin-Producing Neurons. *Cell Stem Cell* **19**, 248–257 (2016).
117. Camp, J. G. *et al.* Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. USA* **112**, 15672–15677 (2015).
118. Amiri, A. *et al.* Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* **362**, (2018).
119. Velasco, S. *et al.* Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
120. Luo, C. *et al.* Cerebral organoids recapitulate epigenomic signatures of the human fetal brain. *Cell Rep.* **17**, 3369–3384 (2016).

121. Tanaka, Y., Cakir, B., Xiang, Y., Sullivan, G. J. & Park, I.-H. Synthetic Analyses of Single-Cell Transcriptomes from Multiple Brain Organoids and Fetal Brain. *Cell Rep.* **30**, 1682–1689.e3 (2020).
122. Pollen, A. A. *et al.* Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* **176**, 743–756.e17 (2019).
123. Lalli, M. A., Avey, D., Dougherty, J. D., Milbrandt, J. & Mitra, R. D. High-throughput single-cell functional elucidation of neurodevelopmental disease-associated genes reveals convergent mechanisms altering neuronal differentiation. *Genome Res.* **30**, 1317–1331 (2020).
124. Kelava, I. & Lancaster, M. A. Stem cell models of human brain development. *Cell Stem Cell* **18**, 736–748 (2016).
125. Yoon, S.-J. *et al.* Reliability of human cortical organoid generation. *Nat. Methods* **16**, 75–78 (2019).
126. Quadrato, G. *et al.* Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* **545**, 48–53 (2017).
127. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
128. Zhang, W. *et al.* Modeling microcephaly with cerebral organoids reveals a WDR62-CEP170-KIF2A pathway promoting cilium disassembly in neural progenitors. *Nat. Commun.* **10**, 2612 (2019).
129. Iefremova, V. *et al.* An Organoid-Based Model of Cortical Development Identifies Non-Cell-Autonomous Defects in Wnt Signaling Contributing to Miller-Dieker Syndrome. *Cell Rep.* **19**, 50–59 (2017).
130. Bershteyn, M. *et al.* Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia. *Cell Stem Cell* **20**, 435–449.e4 (2017).
131. Inak, G. *et al.* SURF1 mutations causative of Leigh syndrome impair human neurogenesis. *BioRxiv* (2019). doi:10.1101/551390
132. Klaus, J. *et al.* Altered neuronal migratory trajectories in human cerebral organoids derived from individuals with neuronal heterotopia. *Nat. Med.* **25**, 561–568 (2019).
133. Sun, A. X. *et al.* Potassium channel dysfunction in human neuronal models of Angelman syndrome. *Science* **366**, 1486–1492 (2019).
134. Paulsen, B. *et al.* Human brain organoids reveal accelerated development of cortical neuron classes as a shared feature of autism risk genes. *BioRxiv* (2020). doi:10.1101/2020.11.10.376509
135. Villa, C. E. *et al.* CHD8 haploinsufficiency alters the developmental trajectories of human excitatory and inhibitory neurons linking autism phenotypes with transient cellular defects. *BioRxiv* (2020). doi:10.1101/2020.11.26.399469
136. Raja, W. K. *et al.* Self-Organizing 3D Human Neural Tissue Derived from Induced Pluripotent Stem Cells Recapitulate Alzheimer’s Disease Phenotypes. *PLoS One* **11**, e0161969 (2016).
137. Smits, L. M. *et al.* Modeling Parkinson’s disease in midbrain-like organoids. *npj Parkinsons Disease* **5**, 5 (2019).
138. Conforti, P. *et al.* Faulty neuronal determination and cell polarization are reverted by modulating HD early phenotypes. *Proc. Natl. Acad. Sci. USA* **115**, E762–E771 (2018).
139. Ogawa, J., Pao, G. M., Shokhirev, M. N. & Verma, I. M. Glioblastoma model using human cerebral organoids. *Cell Rep.* **23**, 1220–1229 (2018).
140. da Silva, B., Mathew, R. K., Polson, E. S., Williams, J. & Wurdak, H. Spontaneous Glioblastoma Spheroid Infiltration of Early-Stage Cerebral Organoids Models Brain Tumor Invasion. *SLAS Discov.* **23**, 862–868 (2018).
141. Janssens, S. *et al.* Zika virus alters DNA methylation of neural genes in an organoid model of the developing human brain. *mSystems* **3**, (2018).
142. Yoon, K.-J. *et al.* Zika-Virus-Encoded NS2A Disrupts Mammalian Cortical Neurogenesis by Degrading Adherens Junction Proteins. *Cell Stem Cell* **21**, 349–358.e6 (2017).

143. Xu, Y.-P. *et al.* Zika virus infection induces RNAi-mediated antiviral immunity in human neural progenitors and brain organoids. *Cell Res.* **29**, 265–273 (2019).
144. Dang, J. *et al.* Zika Virus Depletes Neural Progenitors in Human Cerebral Organoids through Activation of the Innate Immune Receptor TLR3. *Cell Stem Cell* **19**, 258–265 (2016).
145. Zhang, B. *et al.* Differential antiviral immunity to Japanese encephalitis virus in developing cortical organoids. *Cell Death Dis.* **9**, 719 (2018).
146. Pellegrini, L. *et al.* SARS-CoV-2 Infects the Brain Choroid Plexus and Disrupts the Blood-CSF Barrier in Human Brain Organoids. *Cell Stem Cell* **27**, 951–961.e5 (2020).
147. Durens, M. *et al.* High-throughput screening of human induced pluripotent stem cell-derived brain organoids. *J. Neurosci. Methods* **335**, 108627 (2020).
148. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
149. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
150. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* **46**, 17–23 (2014).
151. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
152. Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
153. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17**, 215–222 (2014).
154. Chen, L. *et al.* MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci. USA* **112**, 5509–5514 (2015).
155. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013).
156. Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89–93 (2015).
157. Borrelli, E., Nestler, E. J., Allis, C. D. & Sassone-Corsi, P. Decoding the epigenetic language of neuronal plasticity. *Neuron* **60**, 961–974 (2008).
158. Guo, J. U. *et al.* Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat. Neurosci.* **14**, 1345–1351 (2011).
159. Day, J. J. & Sweatt, J. D. DNA methylation and memory formation. *Nat. Neurosci.* **13**, 1319–1323 (2010).
160. Corley, M. J. *et al.* Epigenetic delay in the neurodevelopmental trajectory of DNA methylation states in autism spectrum disorders. *Front. Genet.* **10**, 907 (2019).
161. Bird, C. P. *et al.* Fast-evolving noncoding sequences in the human genome. *Genome Biol.* **8**, R118 (2007).
162. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).
163. Pollard, K. S. *et al.* Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168 (2006).
164. Prabhakar, S. *et al.* Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008).
165. Reilly, S. K. *et al.* Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
166. Vermunt, M. W. & Creighton, M. P. Transcriptional Dynamics at Brain Enhancers: from Functional Specialization to Neurodegeneration. *Curr Neurol Neurosci Rep* **16**, 94 (2016).
167. Won, H., Huang, J., Opland, C. K., Hartl, C. L. & Geschwind, D. H. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat. Commun.* **10**, 2396 (2019).
168. Geller, E. *et al.* Massively parallel disruption of enhancers active during human corticogenesis. *BioRxiv* (2019). doi:10.1101/852673

169. Pattabiraman, K. *et al.* Transcriptional regulation of enhancers active in protodomains of the developing cerebral cortex. *Neuron* **82**, 989–1003 (2014).
170. Shulha, H. P. *et al.* Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. *Arch. Gen. Psychiatry* **69**, 314–324 (2012).
171. Santen, G. W. E. *et al.* Coffin-Siris syndrome and the BAF complex: genotype-phenotype study in 63 patients. *Hum. Mutat.* **34**, 1519–1528 (2013).
172. Koshi-Mano, K. *et al.* Neuron-specific analysis of histone modifications with post-mortem brains. *Sci. Rep.* **10**, 3767 (2020).
173. Mallo, M. & Alonso, C. R. The regulation of Hox gene expression during animal development. *Development* **140**, 3951–3963 (2013).
174. Slifer, E. H. A mutant stock of *Drosophila* with extra sex-combs. *J. Exp. Zool.* **90**, 31–40 (1942).
175. Lewis, E. B. & Mislove, R. F. New mutants report. *Drosoph. Inf. Serv.* (1947).
176. Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
177. Ingham, P. W. Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. *Nature* **306**, 591–593 (1983).
178. Ingham, P. W. trithorax and the regulation of homeotic gene expression in *Drosophila*: a historical perspective. *Int. J. Dev. Biol.* **42**, 423–429 (1998).
179. Kassis, J. A., Kennison, J. A. & Tamkun, J. W. Polycomb and trithorax group genes in *Drosophila*. *Genetics* **206**, 1699–1725 (2017).
180. Schuettengruber, B., Bourbon, H.-M., Di Croce, L. & Cavalli, G. Genome regulation by polycomb and trithorax: 70 years and counting. *Cell* **171**, 34–57 (2017).
181. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
182. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
183. Vandamme, J., Völkel, P., Rosnoblet, C., Le Faou, P. & Angrand, P.-O. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol. Cell Proteomics* **10**, M110.002642 (2011).
184. Gabriele, M., Lopez Tobon, A., D’Agostino, G. & Testa, G. The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **84**, 306–327 (2018).
185. Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).
186. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).
187. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762–767 (2009).
188. Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell* **15**, 57–67 (2004).
189. Zhang, Q., Vo, N. & Goodman, R. H. Histone binding protein RbAp48 interacts with a complex of CREB binding protein and phosphorylated CREB. *Mol. Cell. Biol.* **20**, 4970–4978 (2000).
190. Hauri, S. *et al.* A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep.* **17**, 583–595 (2016).
191. Nekrasov, M. *et al.* Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* **26**, 4078–4088 (2007).
192. Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol. Cell. Biol.* **28**, 2718–2731 (2008).
193. Beringer, M. *et al.* EPOP functionally links elongin and polycomb in pluripotent stem cells. *Mol. Cell* **64**, 645–658 (2016).
194. Kloet, S. L. *et al.* The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat. Struct. Mol. Biol.* **23**, 682–690 (2016).

195. Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl. Acad. Sci. USA* **104**, 5521–5526 (2007).
196. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* **1**, e1500737 (2015).
197. Wu, H. *et al.* Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. *PLoS One* **8**, e83737 (2013).
198. Tan, J., Yan, Y., Wang, X., Jiang, Y. & Xu, H. E. EZH2: biology, disease, and structure-based drug discovery. *Acta Pharmacol Sin* **35**, 161–174 (2014).
199. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
200. Chou, R.-H., Yu, Y.-L. & Hung, M.-C. The roles of EZH2 in cell lineage commitment. *Am J Transl Res* **3**, 243–250 (2011).
201. Koyen, A. E. *et al.* EZH2 has a non-catalytic and PRC2-independent role in stabilizing DDB2 to promote nucleotide excision repair. *Oncogene* **39**, 4798–4813 (2020).
202. Kim, J. *et al.* Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. *Cell Rep.* **25**, 2808–2820.e4 (2018).
203. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* **27**, 3769–3779 (2007).
204. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
205. Chamberlain, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**, 1496–1505 (2008).
206. O'Carroll, D. *et al.* The polycomb-group gene *Ezh2* is required for early mouse development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).
207. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
208. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
209. van der Stoep, P. *et al.* Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. *PLoS One* **3**, e2235 (2008).
210. Pereira, J. D. *et al.* *Ezh2*, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. *Proc. Natl. Acad. Sci. USA* **107**, 15957–15962 (2010).
211. Burgold, T. *et al.* The H3K27 demethylase JMJD3 is required for maintenance of the embryonic respiratory neuronal network, neonatal breathing, and survival. *Cell Rep.* **2**, 1244–1258 (2012).
212. Corley, M. & Kroll, K. L. The roles and regulation of Polycomb complexes in neural development. *Cell Tissue Res.* **359**, 65–85 (2015).
213. von Schimmelmann, M. *et al.* Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. *Nat. Neurosci.* **19**, 1321–1330 (2016).
214. Testa, G. The time of timing: how Polycomb proteins regulate neurogenesis. *Bioessays* **33**, 519–528 (2011).
215. Akizu, N. *et al.* EZH2 regulates neuroepithelium structure and neuroblast proliferation by repressing p21. *Open Biol* **6**, 150227 (2016).
216. Hirabayashi, Y. *et al.* Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition. *Neuron* **63**, 600–613 (2009).
217. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
218. Brioude, F. *et al.* Overgrowth syndromes - clinical and molecular aspects and tumour risk. *Nat. Rev. Endocrinol.* **15**, 299–311 (2019).

219. Luscan, A. *et al.* Mutations in SETD2 cause a novel overgrowth condition. *J. Med. Genet.* **51**, 512–517 (2014).
220. Tlemsani, C. *et al.* SETD2 and DNMT3A screen in the Sotos-like syndrome French cohort. *J. Med. Genet.* **53**, 743–751 (2016).
221. Almuriekh, M. *et al.* Loss-of-Function Mutation in APC2 Causes Sotos Syndrome Features. *Cell Rep.* (2015). doi:10.1016/j.celrep.2015.02.011
222. Martinez, F. *et al.* Novel mutations of NFIX gene causing Marshall-Smith syndrome or Sotos-like syndrome: one gene, two phenotypes. *Pediatr. Res.* **78**, (2015).
223. Tatton-Brown, K. *et al.* Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. *Oncotarget* **2**, 1127–1133 (2011).
224. Weaver, D. D., Graham, C. B., Thomas, I. T. & Smith, D. W. A new overgrowth syndrome with accelerated skeletal maturation, unusual facies, and camptodactyly. *J. Pediatr.* **84**, 547–552 (1974).
225. Gibson, W. T. *et al.* Mutations in EZH2 cause Weaver syndrome. *Am. J. Hum. Genet.* **90**, 110–118 (2012).
226. Tatton-Brown, K. *et al.* Weaver syndrome and EZH2 mutations: Clarifying the clinical phenotype. *Am. J. Med. Genet. A* **161A**, 2972–2980 (2013).
227. Edmondson, A. C. & Kalish, J. M. Overgrowth Syndromes. *J. Pediatr. Genet.* **4**, 136–143 (2015).
228. Tatton-Brown, K. & Rahman, N. The NSD1 and EZH2 overgrowth genes, similarities and differences. *Am. J. Med. Genet. C Semin. Med. Genet.* **163C**, 86–91 (2013).
229. Kim, K. H. & Roberts, C. W. M. Targeting EZH2 in cancer. *Nat. Med.* **22**, 128–134 (2016).
230. Coulter, D., Powell, C. M. & Gold, S. Weaver syndrome and neuroblastoma. *J. Pediatr. Hematol. Oncol.* **30**, 758–760 (2008).
231. Bertelli, M. O., Munir, K., Harris, J. & Salvador-Carulla, L. Intellectual developmental disorders“: reflections on the international consensus document for redefining “mental retardation-intellectual disability” in ICD-11. *Adv Mental Hlth Intell Disabil* **10**, 36–58 (2016).
232. Harripaul, R. *et al.* Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Mol. Psychiatry* **23**, 973–984 (2018).
233. Redin, C. *et al.* Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J. Med. Genet.* **51**, 724–736 (2014).
234. Ilyas, M., Mir, A., Efthymiou, S. & Houlden, H. The genetics of intellectual disability: advancing technology and gene editing. [version 1; peer review: 2 approved]. *F1000Res.* **9**, (2020).
235. Hill, W. D. *et al.* Examining non-syndromic autosomal recessive intellectual disability (NS-ARID) genes for an enriched association with intelligence differences. *Intelligence* **54**, 80–89 (2016).
236. Muñoz, W. A. & Trainor, P. A. Neural crest cell evolution: how and when did a neural crest cell become a neural crest cell. *Curr Top Dev Biol* **111**, 3–26 (2015).
237. Basch, M. L., Bronner-Fraser, M. & García-Castro, M. I. Specification of the neural crest occurs during gastrulation and requires Pax7. *Nature* **441**, 218–222 (2006).
238. Garnett, A. T., Square, T. A. & Medeiros, D. M. BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. *Development* **139**, 4220–4231 (2012).
239. Knight, R. D. *et al.* lockjaw encodes a zebrafish tfap2a required for early neural crest development. *Development* **130**, 5755–5768 (2003).
240. Monsoro-Burq, A.-H., Wang, E. & Harland, R. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during *Xenopus* neural crest induction. *Dev. Cell* **8**, 167–178 (2005).
241. Cheung, M. *et al.* The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev. Cell* **8**, 179–192 (2005).
242. Bhatt, S., Diaz, R. & Trainor, P. A. Signals and switches in Mammalian neural crest cell differentiation. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).

243. Wilkins, A. S., Wrangham, R. W. & Fitch, W. T. The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics* **197**, 795–808 (2014).
244. Theofanopoulou, C. *et al.* Self-domestication in Homo sapiens: Insights from comparative genomics. *PLoS One* **12**, e0185306 (2017).
245. Hare, B. Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annu. Rev. Psychol.* **68**, 155–186 (2017).
246. Sánchez-Villagra, M. R. & van Schaik, C. P. Evaluating the self-domestication hypothesis of human evolution. *Evol Anthropol* **28**, 133–143 (2019).
247. Cohen, A. S. A. *et al.* A novel mutation in EED associated with overgrowth. *J. Hum. Genet.* **60**, 339–342 (2015).
248. Lui, J. C. *et al.* Ezh2 mutations found in the weaver overgrowth syndrome cause a partial loss of H3K27 histone methyltransferase activity. *J. Clin. Endocrinol. Metab.* **103**, 1470–1478 (2018).
249. Imagawa, E. *et al.* Novel SUZ12 mutations in Weaver-like syndrome. *Clin. Genet.* **94**, 461–466 (2018).
250. Choufani, S. *et al.* DNA methylation signature for EZH2 functionally classifies sequence variants in three PRC2 complex genes. *Am. J. Hum. Genet.* **106**, 596–610 (2020).
251. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
252. Smith, C. *et al.* Efficient and allele-specific genome editing of disease loci in human iPSCs. *Mol. Ther.* **23**, 570–577 (2015).
253. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
254. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, Engl.)* **151**, 2551–2561 (2005).
255. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
256. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
257. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
258. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
259. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
260. Sapranaukas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
261. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* **109**, E2579–86 (2012).
262. Mojica, F. J. M. & Montoliu, L. On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals. *Trends Microbiol.* **24**, 811–820 (2016).
263. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
264. Ma, N. *et al.* Determining the Pathogenicity of a Genomic Variant of Uncertain Significance Using CRISPR/Cas9 and Human-Induced Pluripotent Stem Cells. *Circulation* **138**, 2666–2681 (2018).
265. Zanella, M. *et al.* Dosage analysis of the 7q11.23 Williams region identifies BAZ1B as a major human gene patterning the modern human face and underlying self-domestication. *Sci. Adv.* **5**, eaaw7908 (2019).

266. Boettcher, M. *et al.* Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* **36**, 170–178 (2018).
267. Ross, P. J. *et al.* Synaptic Dysfunction in Human Neurons With Autism-Associated Deletions in PTCHD1-AS. *Biol. Psychiatry* **87**, 139–149 (2020).
268. Frega, M. *et al.* Neuronal network dysfunction in a model for Kleefstra syndrome mediated by enhanced NMDAR signaling. *Nat. Commun.* **10**, 4928 (2019).
269. Briscoe, J. & Marín, O. Looking at neurodevelopment through a big data lens. *Science* doi:10.1126/science.aaz8627
270. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
271. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
272. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
273. Tang, Y., Li, M., Wang, J., Pan, Y. & Wu, F.-X. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* **127**, 67–72 (2015).
274. Convertini, P. *et al.* Genome wide array analysis indicates that an amyotrophic lateral sclerosis mutation of FUS causes an early increase of CAMK2N2 in vitro. *Biochim. Biophys. Acta* (2013). doi:10.1016/j.bbadis.2013.03.015
275. Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat. Rev. Neurol.* **14**, 544–558 (2018).
276. Köhn, L. *et al.* Mutation in the PYK2-binding domain of PITPNM3 causes autosomal dominant cone dystrophy (CORD5) in two Swedish families. *Eur. J. Hum. Genet.* **15**, 664–671 (2007).
277. Liu, C.-Y. *et al.* Copine-7 is required for REM sleep regulation following cage change or water immersion and restraint stress in mice. *Neurosci. Res.* (2020). doi:10.1016/j.neures.2020.04.002
278. Chang, J. D., Field, S. J., Rameh, L. E., Carpenter, C. L. & Cantley, L. C. Identification and characterization of a phosphoinositide phosphate kinase homolog. *J. Biol. Chem.* **279**, 11672–11679 (2004).
279. Grammatikakis, I., Zhang, P., Mattson, M. P. & Gorospe, M. The long and the short of TRF2 in neurogenesis. *Cell Cycle* **15**, 0 (2016).
280. Patel, M. V. *et al.* A Novel Short Isoform of Cytosolic PSD-95 Interactor (Cypin) Regulates Neuronal Development. *Mol. Neurobiol.* **55**, 6269–6281 (2018).
281. Pérez-Otaño, I., Larsen, R. S. & Wesseling, J. F. Emerging roles of GluN3-containing NMDA receptors in the CNS. *Nat. Rev. Neurosci.* **17**, 623–635 (2016).
282. Waters, M. F. *et al.* Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes. *Nat. Genet.* **38**, 447–451 (2006).
283. Bencsik, N. *et al.* Dendritic spine morphology and memory formation depend on postsynaptic Caskin proteins. *Sci. Rep.* **9**, 16843 (2019).
284. Küry, S. *et al.* De novo mutations in protein kinase genes CAMK2A and CAMK2B cause intellectual disability. *Am. J. Hum. Genet.* **101**, 768–788 (2017).
285. Saito, N. *et al.* KIFC2 is a novel neuron-specific C-terminal type kinesin superfamily motor for dendritic transport of multivesicular body-like organelles. *Neuron* **18**, 425–438 (1997).
286. Raghunath, M. *et al.* A novel kinase, AATYK induces and promotes neuronal differentiation in a human neuroblastoma (SH-SY5Y) cell line. *Brain Res. Mol. Brain Res.* **77**, 151–162 (2000).
287. Dent, E. W., Gupton, S. L. & Gertler, F. B. The growth cone cytoskeleton in axon outgrowth and guidance. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
288. Yukawa, K. *et al.* Semaphorin 4A induces growth cone collapse of hippocampal neurons in a Rho/Rho-kinase-dependent manner. *Int. J. Mol. Med.* **16**, 115–118 (2005).
289. Vanderzalm, P. J. *et al.* C. elegans CARMIL negatively regulates UNC-73/Trio function during neuronal development. *Development* **136**, 1201–1210 (2009).

290. Piard, J. *et al.* FRMPD4 mutations cause X-linked intellectual disability and disrupt dendritic spine morphogenesis. *Hum. Mol. Genet.* **27**, 589–600 (2018).
291. Riccomagno, M. M. *et al.* The RacGAP β 2-Chimaerin selectively mediates axonal pruning in the hippocampus. *Cell* **149**, 1594–1606 (2012).
292. Huang, L. *et al.* Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia. *Orphanet J Rare Dis* **7**, 67 (2012).
293. Sato, D. *et al.* SHANK1 Deletions in Males with Autism Spectrum Disorder. *Am. J. Hum. Genet.* **90**, 879–887 (2012).
294. Mao, W. *et al.* Shank1 regulates excitatory synaptic transmission in mouse hippocampal parvalbumin-expressing inhibitory interneurons. *Eur. J. Neurosci.* **41**, 1025–1035 (2015).
295. Shi, R. *et al.* Shank proteins differentially regulate synaptic transmission. *Eneuro* **4**, (2017).
296. Moon, A. L., Haan, N., Wilkinson, L. S., Thomas, K. L. & Hall, J. CACNA1C: association with psychiatric disorders, behavior, and neurogenesis. *Schizophr. Bull.* **44**, 958–965 (2018).
297. Cockcroft, S. The diverse functions of phosphatidylinositol transfer proteins. *Curr. Top. Microbiol. Immunol.* **362**, 185–208 (2012).
298. Rabadán, M. A. *et al.* Jagged2 controls the generation of motor neuron and oligodendrocyte progenitors in the ventral spinal cord. *Cell Death Differ.* **19**, 209–219 (2012).
299. Zhu, Y.-C. & Xiong, Z.-Q. Molecular and synaptic bases of CDKL5 disorder. *Dev. Neurobiol.* (2018). doi:10.1002/dneu.22639
300. Guo, D. *et al.* Autism-like social deficit generated by Dock4 deficiency is rescued by restoration of Rac1 activity and NMDA receptor function. *Mol. Psychiatry* (2019). doi:10.1038/s41380-019-0472-7
301. Südhof, T. C. Synaptotagmins: why so many? *J. Biol. Chem.* **277**, 7629–7632 (2002).
302. Yokoyama, K. *et al.* NYAP: a phosphoprotein family that links PI3K to WAVE1 signalling in neurons. *EMBO J.* **30**, 4739–4754 (2011).
303. Geng, A. *et al.* KIF20A/MKLP2 regulates the division modes of neural progenitor cells during cortical development. *Nat. Commun.* **9**, 2707 (2018).
304. Johnson, K. *et al.* Kif11 dependent cell cycle progression in radial glial cells is required for proper neurogenesis in the zebrafish neural tube. *Dev. Biol.* **387**, 73–92 (2014).
305. Delgado-Esteban, M., García-Higuera, I., Maestre, C., Moreno, S. & Almeida, A. APC/C-Cdh1 coordinates neurogenesis and cortical size during development. *Nat. Commun.* **4**, 2879 (2013).
306. Hagey, D. W. *et al.* CYCLIN-B1/2 and -D1 act in opposition to coordinate cortical progenitor self-renewal and lineage commitment. *Nat. Commun.* **11**, 2898 (2020).
307. Guo, J. *et al.* Developmental disruptions underlying brain abnormalities in ciliopathies. *Nat. Commun.* **6**, 7857 (2015).
308. Zhou, Y. & Chen, J. J. STAT3 plays an important role in DNA replication by turning on WDHD1. *Cell Biosci.* **11**, 10 (2021).
309. Hauser, S. *et al.* Loss of LIN9, a member of the DREAM complex, cooperates with SV40 large T antigen to induce genomic instability and anchorage-independent growth. *Oncogene* **31**, 1859–1868 (2012).
310. Cheung, M.-H. *et al.* Human NOC3 is essential for DNA replication licensing in human cells. *Cell Cycle* **18**, 605–620 (2019).
311. Nag, A. *et al.* CNV analysis in Tourette syndrome implicates large genomic rearrangements in COL8A1 and NRXN1. *PLoS One* **8**, e59061 (2013).
312. Smagin, D. A., Galyamina, A. G., Kovalenko, I. L., Babenko, V. N. & Kudryavtseva, N. N. Aberrant Expression of Collagen Gene Family in the Brain Regions of Male Mice with Behavioral Psychopathologies Induced by Chronic Agonistic Interactions. *Biomed Res. Int.* **2019**, 7276389 (2019).
313. Barak, T. *et al.* Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat. Genet.* **43**, 590–594 (2011).
314. Darbro, B. W. *et al.* Mutations in extracellular matrix genes NID1 and LAMC1 cause autosomal dominant Dandy-Walker malformation and occipital cephaloceles. *Hum. Mutat.* **34**, 1075–1079 (2013).

315. Touat-Hamici, Z. *et al.* Role of lipid phosphate phosphatase 3 in human aortic endothelial cell function. *Cardiovasc. Res.* **112**, 702–713 (2016).
316. Xu, K. *et al.* Lunatic fringe deficiency cooperates with the Met/Caveolin gene amplicon to induce basal-like breast cancer. *Cancer Cell* **21**, 626–641 (2012).
317. Schiavinato, A. *et al.* EMILIN-3, peculiar member of elastin microfibril interface-located protein (EMILIN) family, has distinct expression pattern, forms oligomeric assemblies, and serves as transforming growth factor β (TGF- β) antagonist. *J. Biol. Chem.* **287**, 11498–11515 (2012).
318. Das, S., Li, Z., Noori, A., Hyman, B. T. & Serrano-Pozo, A. Meta-analysis of mouse transcriptomic studies supports a context-dependent astrocyte reaction in acute CNS injury versus neurodegeneration. *J. Neuroinflammation* **17**, 227 (2020).
319. Chen, J. *et al.* AXL promotes Zika virus infection in astrocytes by antagonizing type I interferon signalling. *Nat. Microbiol.* **3**, 302–309 (2018).
320. Obayashi, S., Tabunoki, H., Kim, S. U. & Satoh, J. Gene expression profiling of human neural progenitor cells following the serum-induced astrocyte differentiation. *Cell Mol. Neurobiol.* **29**, 423–438 (2009).
321. Dahl, D. & Bignami, A. Immunohistological localization of desmin, the muscle-type 100 A filament protein, in rat astrocytes and Müller glia. *J. Histochem. Cytochem.* **30**, 207–213 (1982).
322. Fink, K. L., López-Giráldez, F., Kim, I.-J., Strittmatter, S. M. & Cafferty, W. B. J. Identification of Intrinsic Axon Growth Modulators for Intact CNS Neurons after Injury. *Cell Rep.* **18**, 2687–2701 (2017).
323. Borggrewe, M. *et al.* Regionally diverse astrocyte subtypes and their heterogeneous response to EAE. *Glia* (2020). doi:10.1002/glia.23954
324. Manning, T. J., Rosenfeld, S. S. & Sontheimer, H. Lysophosphatidic acid stimulates actomyosin contraction in astrocytes. *J. Neurosci. Res.* **53**, 343–352 (1998).
325. Ferrier, R., Had, L., Rabié, A. & Faivre-Sarrailh, C. Coordinated expression of five tropomyosin isoforms and beta-actin in astrocytes treated with dibutyryl cAMP and cytochalasin D. *Cell Motil. Cytoskeleton* **28**, 303–316 (1994).
326. Magistri, M. *et al.* A comparative transcriptomic analysis of astrocytes differentiation from human neural progenitor cells. *Eur. J. Neurosci.* **44**, 2858–2870 (2016).
327. Arnold, S. J. *et al.* The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. *Genes Dev.* **22**, 2479–2484 (2008).
328. Chou, S.-J. & Tole, S. Lhx2, an evolutionarily conserved, multifunctional regulator of forebrain development. *Brain Res.* **1705**, 1–14 (2019).
329. Chen, B. *et al.* The Fezf2-Ctip2 genetic pathway regulates the fate choice of subcortical projection neurons in the developing cerebral cortex. *Proc. Natl. Acad. Sci. USA* **105**, 11382–11387 (2008).
330. Sloan, S. A. *et al.* Human Astrocyte Maturation Captured in 3D Cerebral Cortical Spheroids Derived from Pluripotent Stem Cells. *Neuron* **95**, 779–790.e6 (2017).
331. Iliff, J. J. *et al.* A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β . *Sci. Transl. Med.* **4**, 147ra111 (2012).
332. Sherrod, M., Liu, X., Zhang, X. & Sigmund, C. D. Nuclear localization of angiotensinogen in astrocytes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R539–46 (2005).
333. Raju, C. S. *et al.* Secretagogin is expressed by developing neocortical gabaergic neurons in humans but not mice and increases neurite arbor size and complexity. *Cereb. Cortex* **28**, 1946–1958 (2018).
334. Sellers, K., Zyka, V., Lumsden, A. G. & Delogu, A. Transcriptional control of GABAergic neuronal subtype identity in the thalamus. *Neural Dev.* **9**, 14 (2014).
335. Trujillo, C. A. *et al.* Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell* **25**, 558–569.e7 (2019).
336. Dong, M. *et al.* SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief. Bioinformatics* (2020). doi:10.1093/bib/bbz166

337. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785–801.e8 (2019).
338. López-Tobón, A. *et al.* Human cortical organoids expose a differential function of GSK3 on cortical neurogenesis. *Stem Cell Rep.* **13**, 847–861 (2019).
339. Wilkinson, B., Li, J. & Coba, M. P. Synaptic GAP and GEF complexes cluster proteins essential for GTP signaling. *Sci. Rep.* **7**, 5272 (2017).
340. Balaji, V., Kaniyappan, S., Mandelkow, E., Wang, Y. & Mandelkow, E.-M. Pathological missorting of endogenous MAPT/Tau in neurons caused by failure of protein degradation systems. *Autophagy* **14**, 2139–2154 (2018).
341. Beroun, A. *et al.* MMPs in learning and memory and neuropsychiatric disorders. *Cell Mol. Life Sci.* **76**, 3207–3228 (2019).
342. Vilas, G. L., Johnson, D. E., Freund, P. & Casey, J. R. Characterization of an epilepsy-associated variant of the human Cl⁻/HCO₃⁻ exchanger AE3. *Am. J. Physiol. Cell Physiol.* **297**, C526–36 (2009).
343. Born, G. *et al.* Genetic targeting of NRXN2 in mice unveils role in excitatory cortical synapse function and social behaviors. *Front. Synaptic Neurosci.* **7**, 3 (2015).
344. Bormuth, I. *et al.* Neuronal basic helix-loop-helix proteins Neurod2/6 regulate cortical commissure formation before midline interactions. *J. Neurosci.* **33**, 641–651 (2013).
345. Alifragis, P., Molnár, Z. & Parnavelas, J. G. Restricted expression of Slap-1 in the rodent cerebral cortex. *Gene Expr Patterns* **3**, 437–440 (2003).
346. Rehfeld, F. *et al.* The RNA-binding protein ARPP21 controls dendritic branching by functionally opposing the miRNA it hosts. *Nat. Commun.* **9**, 1235 (2018).
347. Cánovas, J. *et al.* The Specification of Cortical Subcerebral Projection Neurons Depends on the Direct Repression of TBR1 by CTIP1/BCL11a. *J. Neurosci.* **35**, 7552–7564 (2015).
348. Oishi, K. *et al.* Identity of neocortical layer 4 neurons is specified through correct positioning into the cortex. *Elife* **5**, (2016).
349. Hemming, I. A. *et al.* Disease-associated missense variants in ZBTB18 disrupt DNA binding and impair the development of neurons within the embryonic cerebral cortex. *Hum. Mutat.* **40**, 1841–1855 (2019).
350. Chen, C.-M., Wang, H.-Y., You, L.-R., Shang, R.-L. & Liu, F.-C. Expression analysis of an evolutionarily conserved metallophosphodiesterase gene, Mpped1, in the normal and beta-catenin-deficient malformed dorsal telencephalon. *Dev. Dyn.* **239**, 1797–1806 (2010).
351. Chen, N.-P., Uddin, B., Voit, R. & Schiebel, E. Human phosphatase CDC14A is recruited to the cell leading edge to regulate cell migration and adhesion. *Proc. Natl. Acad. Sci. USA* **113**, 990–995 (2016).
352. Cao, X., Pfaff, S. L. & Gage, F. H. YAP regulates neural progenitor cell number via the TEA domain transcription factor. *Genes Dev.* **22**, 3320–3334 (2008).
353. Courchet, J. *et al.* Terminal axon branching is regulated by the LKB1-NUAK1 kinase pathway via presynaptic mitochondrial capture. *Cell* **153**, 1510–1525 (2013).
354. Pfister, J. A. & D’Mello, S. R. Insights into the regulation of neuronal viability by nucleophosmin/B23. *Exp. Biol. Med.* **240**, 774–786 (2015).
355. Pei, L. *et al.* Dependence of hippocampal function on ERR γ -regulated mitochondrial metabolism. *Cell Metab.* **21**, 628–636 (2015).
356. Hu, C.-K., Coughlin, M., Field, C. M. & Mitchison, T. J. KIF4 regulates midzone length during cytokinesis. *Curr. Biol.* **21**, 815–824 (2011).
357. Czechanski, A. *et al.* Kif18a is specifically required for mitotic progression during germ line development. *Dev. Biol.* **402**, 253–262 (2015).
358. Ganguly, R. *et al.* MELK-a conserved kinase: functions, signaling, cancer, and controversy. *Clin. Transl. Med.* **4**, 11 (2015).
359. Yu, C.-T. R. *et al.* Phosphorylation and stabilization of HURP by Aurora-A: implication of HURP as a transforming target of Aurora-A. *Mol. Cell. Biol.* **25**, 5789–5800 (2005).
360. Smogorzewska, A. *et al.* Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289–301 (2007).

361. Kikuma, K., Li, X., Kim, D., Sutter, D. & Dickman, D. K. Extended synaptotagmin localizes to presynaptic ER and promotes neurotransmission and synaptic growth in drosophila. *Genetics* **207**, 993–1006 (2017).
362. Papadimitriou, C. *et al.* 3D Culture Method for Alzheimer’s Disease Modeling Reveals Interleukin-4 Rescues A β 42-Induced Loss of Human Neural Stem Cell Plasticity. *Dev. Cell* **46**, 85–101.e8 (2018).
363. Domínguez-Iturza, N. *et al.* The autism- and schizophrenia-associated protein CYFIP1 regulates bilateral brain connectivity and behaviour. *Nat. Commun.* **10**, 3454 (2019).
364. Twohig, J. P., Cuff, S. M., Yong, A. A. & Wang, E. C. Y. The role of tumor necrosis factor receptor superfamily members in mammalian brain development, function and homeostasis. *Rev Neurosci* **22**, 509–533 (2011).
365. Radner, S. *et al.* β 2 and γ 3 laminins are critical cortical basement membrane components: ablation of Lamb2 and Lamc3 genes disrupts cortical lamination and produces dysplasia. *Dev. Neurobiol.* **73**, 209–229 (2013).
366. Szu, J. I. & Binder, D. K. The Role of Astrocytic Aquaporin-4 in Synaptic Plasticity and Learning and Memory. *Front Integr Neurosci* **10**, 8 (2016).
367. Yoshioka, N. *et al.* Efficient generation of human iPSCs by a synthetic self-replicative RNA. *Cell Stem Cell* **13**, 246–254 (2013).
368. Cohen, A. S. A. *et al.* Weaver Syndrome-Associated EZH2 Protein Variants Show Impaired Histone Methyltransferase Function In Vitro. *Hum. Mutat.* **37**, 301–307 (2016).
369. Gross, J. C., Schreiner, A., Engels, K. & Starzinski-Powitz, A. E-cadherin surface levels in epithelial growth factor-stimulated cells depend on adherens junction protein shrew-1. *Mol. Biol. Cell* **20**, 3598–3607 (2009).
370. Isidor, B. *et al.* Complex constitutional subtelomeric 1p36.3 deletion/duplication in a mentally retarded child with neonatal neuroblastoma. *Eur J Med Genet* **51**, 679–684 (2008).
371. Lee, G., Chambers, S. M., Tomishima, M. J. & Studer, L. Derivation of neural crest cells from human pluripotent stem cells. *Nat. Protoc.* **5**, 688–701 (2010).
372. Cordero, D. R. *et al.* Cranial neural crest cells on the move: their roles in craniofacial development. *Am. J. Med. Genet. A* **155A**, 270–279 (2011).
373. Snider, T. N. & Mishina, Y. Cranial neural crest cell contribution to craniofacial formation, pathology, and future directions in tissue engineering. *Birth Defects Res C Embryo Today* **102**, 324–332 (2014).
374. Drakulic, D. *et al.* Copy number variants (CNVs): a powerful tool for iPSC-based modelling of ASD. *Mol. Autism* **11**, 42 (2020).
375. Gordon, A. *et al.* Long-term maturation of human cortical organoids matches key early postnatal transitions. *Nat. Neurosci.* (2021). doi:10.1038/s41593-021-00802-y
376. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
377. Lähnemann, D. *et al.* Eleven grand challenges in single-cell data science. *Genome Biol.* **21**, 31 (2020).
378. Zheng, Y., Zhong, Y., Hu, J. & Shang, X. SCC: an accurate imputation method for scRNA-seq dropouts based on a mixture model. *BMC Bioinformatics* **22**, 5 (2021).
379. Kagohara, L. T. *et al.* Integrated single-cell and bulk gene expression and ATAC-seq reveals heterogeneity and early changes in pathways associated with resistance to cetuximab in HNSCC-sensitive cell lines. *Br. J. Cancer* **123**, 101–113 (2020).
380. Schiweck, J., Eickholt, B. J. & Murk, K. Important shapeshifter: mechanisms allowing astrocytes to respond to the changing nervous system during development, injury and disease. *Front. Cell Neurosci.* **12**, 261 (2018).
381. Krencik, R. & Ullian, E. M. A cellular star atlas: using astrocytes from human pluripotent stem cells for disease studies. *Front. Cell Neurosci.* **7**, 25 (2013).
382. Schafer, S. T. *et al.* Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat. Neurosci.* **22**, 243–255 (2019).
383. Kohwi, M. *et al.* A subpopulation of olfactory bulb GABAergic interneurons is derived from Emx1- and Dlx5/6-expressing progenitors. *J. Neurosci.* **27**, 6878–6891 (2007).

384. Long, K. R. *et al.* Extracellular Matrix Components HAPLN1, Lumican, and Collagen I Cause Hyaluronic Acid-Dependent Folding of the Developing Human Neocortex. *Neuron* **99**, 702–719.e6 (2018).
385. Franco, S. J., Martinez-Garay, I., Gil-Sanz, C., Harkins-Perry, S. R. & Müller, U. Reelin regulates cadherin function via Dab1/Rap1 to control neuronal migration and lamination in the neocortex. *Neuron* **69**, 482–497 (2011).
386. Wiese, S., Karus, M. & Faissner, A. Astrocytes as a source for extracellular matrix molecules and cytokines. *Front. Pharmacol.* **3**, 120 (2012).
387. Florio, M. *et al.* Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **347**, 1465–1470 (2015).
388. Fragola, G. *et al.* Cell reprogramming requires silencing of a core subset of polycomb targets. *PLoS Genet.* **9**, e1003292 (2013).
389. Pang, T., Atefy, R. & Sheen, V. Malformations of cortical development. *Neurologist* **14**, 181–191 (2008).
390. Pirozzi, F., Nelson, B. & Mirzaa, G. From microcephaly to megalencephaly: determinants of brain size. *Dialogues Clin Neurosci* **20**, 267–282 (2018).
391. Cruz-Molina, S. *et al.* PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20**, 689–705.e9 (2017).
392. Zhang, W. *et al.* Cerebral organoid and mouse models reveal a RAB39b-PI3K-mTOR pathway-dependent dysregulation of cortical development leading to macrocephaly/autism phenotypes. *Genes Dev.* **34**, 580–597 (2020).
393. Jarome, T. J., Perez, G. A., Hauser, R. M., Hatch, K. M. & Lubin, F. D. EZH2 Methyltransferase Activity Controls Pten Expression and mTOR Signaling during Fear Memory Reconsolidation. *J. Neurosci.* **38**, 7635–7648 (2018).
394. Hsieh, Y.-Y., Lo, H.-L. & Yang, P.-M. EZH2 inhibitors transcriptionally upregulate cytotoxic autophagy and cytoprotective unfolded protein response in human colorectal cancer cells. *Am. J. Cancer Res.* **6**, 1661–1680 (2016).
395. Zhao, L. *et al.* Ezh2 is involved in radial neuronal migration through regulating Reelin expression in cerebral cortex. *Sci. Rep.* **5**, 15484 (2015).
396. Eriksson, S. H. *et al.* Persistent reelin-expressing Cajal-Retzius cells in polymicrogyria. *Brain* **124**, 1350–1361 (2001).
397. Pang, B. *et al.* EZH2 promotes metabolic reprogramming in glioblastomas through epigenetic repression of EAF2-HIF1 α signaling. *Oncotarget* **7**, 45134–45143 (2016).
398. Kathagen-Buhmann, A. *et al.* Glycolysis and the pentose phosphate pathway are differentially associated with the dichotomous regulation of glioblastoma cell migration versus proliferation. *Neuro. Oncol.* **18**, 1219–1229 (2016).
399. Suri, T. & Dixit, A. The phenotype of EZH2 haploinsufficiency-1.2-Mb deletion at 7q36.1 in a child with tall stature and intellectual disability. *Am. J. Med. Genet. A* (2017). doi:10.1002/ajmg.a.38356
400. Jakob, V., Schreiner, A., Tikkanen, R. & Starzinski-Powitz, A. Targeting of transmembrane protein shrew-1 to adherens junctions is controlled by cytoplasmic sorting motifs. *Mol. Biol. Cell* **17**, 3397–3408 (2006).
401. Han, L. *et al.* AJAP1 is dysregulated at an early stage of gliomagenesis and suppresses invasion through cytoskeleton reorganization. *CNS Neurosci Ther* **20**, 429–437 (2014).
402. Borbone, E. *et al.* Enhancer of zeste homolog 2 overexpression has a role in the development of anaplastic thyroid carcinomas. *J. Clin. Endocrinol. Metab.* **96**, 1029–1038 (2011).
403. Menendez, L. *et al.* Directed differentiation of human pluripotent cells to neural crest stem cells. *Nat. Protoc.* **8**, 203–212 (2013).
404. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).

405. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
406. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
407. D’Astolfo, D. S. *et al.* Efficient intracellular delivery of native proteins. *Cell* **161**, 674–690 (2015).
408. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).
409. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
410. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
411. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
412. Fortin, J.-P., Triche, T. J. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558–560 (2017).