

ARTICLE OPEN



Benchmarking brain organoid recapitulation of fetal corticogenesis

Cristina Cheroni^{1,2,3,13}, Sebastiano Trattaro^{1,2,3,13}, Nicolò Caporale^{1,2,3}, Alejandro López-Tobón^{1,2,3,7}, Erika Tenderini^{1,8}, Sara Sebastiani^{1,9}, Flavia Troglio^{1,2,3}, Michele Gabriele^{1,2,10,11,12}, Raul Bardini Bressan⁴, Steven M. Pollard⁴, William T. Gibson^{5,6} and Giuseppe Testa^{1,2,3}✉

© The Author(s) 2022

Brain organoids are becoming increasingly relevant to dissect the molecular mechanisms underlying psychiatric and neurological conditions. The *in vitro* recapitulation of key features of human brain development affords the unique opportunity of investigating the developmental antecedents of neuropsychiatric conditions in the context of the actual patients' genetic backgrounds. Specifically, multiple strategies of brain organoid (BO) differentiation have enabled the investigation of human cerebral corticogenesis *in vitro* with increasing accuracy. However, the field lacks a systematic investigation of how closely the gene co-expression patterns seen in cultured BO from different protocols match those observed in fetal cortex, a paramount information for ensuring the sensitivity and accuracy of modeling disease trajectories. Here we benchmark BO against fetal corticogenesis by integrating transcriptomes from in-house differentiated cortical BO (CBO), other BO systems, human fetal brain samples processed in-house, and prenatal cortices from the BrainSpan Atlas. We identified co-expression patterns and prioritized hubs of human corticogenesis and CBO differentiation, highlighting both well-preserved and discordant trends across BO protocols. We evaluated the relevance of identified gene modules for neurodevelopmental disorders and psychiatric conditions finding significant enrichment of disease risk genes especially in modules related to neuronal maturation and synapsis development. The longitudinal transcriptomic analysis of CBO revealed a two-step differentiation composed of a fast-evolving phase, corresponding to the appearance of the main cell populations of the cortex, followed by a slow-evolving one characterized by milder transcriptional changes. Finally, we observed heterochronicity of differentiation across BO models compared to fetal cortex. Our approach provides a framework to directly compare the extent of *in vivo/in vitro* alignment of neurodevelopmentally relevant processes and their attending temporalities, structured as a resource to query for modeling human corticogenesis and the neuropsychiatric outcomes of its alterations.

Translational Psychiatry (2022)12:520; <https://doi.org/10.1038/s41398-022-02279-0>

INTRODUCTION

The introduction of cell reprogramming technologies and 3D brain organoids (BO) have made the spatial and temporal dynamics of human brain development experimentally accessible. BO are thus becoming central to the investigation of how genetic vulnerabilities or environmental perturbations can alter physiological neurodevelopment and seed the unfolding of psychiatric and neurological conditions [1, 2]. As we and others recently demonstrated, the exposure of specific windows of vulnerability is proving of particular value, highlighting how temporally defined or even transient alterations in neurodevelopmental trajectories can bring about major mental health outcomes, from language

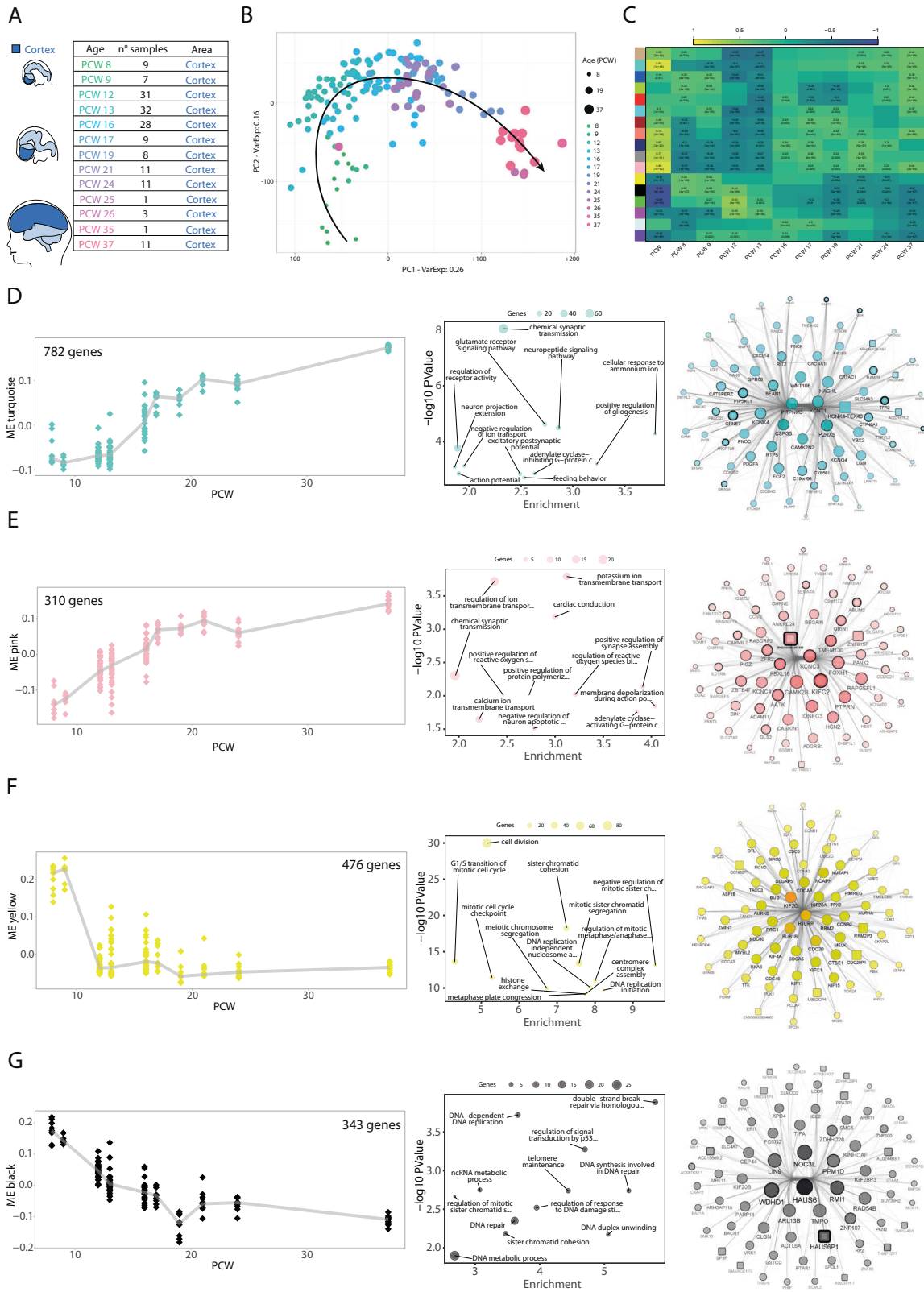
delay to autism spectrum disorders (ASD) [3–5]. Indeed, a growing body of literature testifies to the edge that BO are bringing to the modeling of complex neuropsychiatric disorders, from ASD to schizophrenia, bipolar disorder, and beyond [6–8]. Given the central role of the cortex for the higher-order functions primarily affected in neuropsychiatric conditions [9], determining how closely BO recapitulate human cerebral corticogenesis is thus crucial for harnessing their full potential as *in vitro* models of neurodevelopmental processes and their physiopathological outcomes. To this end, comparison between *ex vivo* human fetal cortex and BO gene expression has started to uncover the extent of this recapitulation, alongside the peculiarities of different

¹Department of Experimental Oncology, European Institute of Oncology IRCCS, Milan, Italy. ²Department of Oncology and Hemato-Oncology, University of Milan, Via Santa Sofia 9, 20122 Milan, Italy. ³Human Technopole, Viale Rita Levi-Montalcini 1, 20157 Milan, Italy. ⁴Centre for Regenerative Medicine, Institute for Regeneration and Repair, and Edinburgh Cancer Research UK Centre, University of Edinburgh, Edinburgh EH16 4UU, UK. ⁵BC Children's Hospital Research Institute, 950 West 28th Avenue, Vancouver, BC V5Z 4H4, Canada. ⁶Department of Medical Genetics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. ⁷Present address: Herophilus, Inc., San Francisco, CA 94107, USA. ⁸Present address: Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Via Campi Flegrei, 34 - 80078 Naples, Italy. ⁹Present address: Department of Molecular Neurobiology, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Nijmegen, The Netherlands. ¹⁰Present address: Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ¹¹Present address: The Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA. ¹²Present address: Koch Institute for Integrative Cancer Research, Cambridge, MA 02139, USA. ¹³These authors contributed equally: Cristina Cheroni, Sebastiano Trattaro.

✉email: giuseppe.testa@fht.org

Received: 17 November 2022 Revised: 28 November 2022 Accepted: 30 November 2022

Published online: 20 December 2022



methods. Specifically, the transcriptomic and epigenomic landscapes of BO were characterized and compared to isogenic fetal cortices, finding significant overlaps [10], as well as to mice data, highlighting human-specific genes involved in lineage establishment [11]. Very long-term BO cultures were meanwhile shown to capture early post-natal developmental transitions [12], while

spatial similarity maps of BO against reference were generated to assess organoid engineering protocols and to annotate cell fates [13]. Further detail emerged from integrative analyses showing an overexpression of extracellular matrix (ECM)-related genes in BO associated with the first steps of differentiation in 2D [14], a higher vulnerability to cellular stress in some organoid cultures compared

Fig. 1 Reconstruction of transcriptional programs of the developing fetal cortex. **A** Cohort of fetal cortical samples from the BrainSpan (BS) Atlas. The number of specimens for each post-conceptual week (PCW) is reported, for a total of 162 bulk RNASeq data points. **B** PCA on prenatal cortical samples from BS. Dot color and size is set according to developmental stage. **C** WGCNA pinpoints gene modules in the developing cortex. The heatmap shows the correlation between the first principal component of each module (module eigengene, ME) and the developmental stage, either as a continuous variable (PCW) or as a categorical variable for each stage. Coefficients of correlation were calculated using Spearman correlation; *P*-values are reported for significant correlations (*P*-value < 0.01). Each row represents a gene module and it is identified with a specific color. **D–G** Characterization of BS_Turquoise, BS_Pink, BS_Yellow, and BS_Black modules. For each module, ribbon plots visualize the behavior of the ME (Y-axis) through developmental stages; each dot represents a data point, while the line connects the median value for each PCW. The bubble plots show the *P*-value (Y-axis) and enrichment score (X-axis) for the top-12 GO categories (ranked according to *P*-value, Biological Process domain of the ontology); dot size represents the number of module genes belonging to the GO term (complete results are reported in Supplementary Table 3). Network reconstruction for the top-75 genes of each module, selected according to the intramodular connectivity. Degree, closeness, betweenness, and eigenvector are represented by node label transparency, node color darkness, node border width, and node size/node label font size, respectively. Node shape represents gene biotype, with protein-coding genes as circles and non-coding genes as squares.

to primary tissue [15, 16], and the existence of protocol-specific transcriptional bypasses in BO differentiation [17].

These efforts have provided a wealth of information, which remains, however, difficult to harmonize due to the lack of dedicated resources where transcriptional hubs of development/differentiation are categorized, ranked, and made available for consultation. Such tools are needed to help researchers select protocols and time-points based not only on the expression of genes of interest but also on the broader context of functional partitions and temporal dynamics of that expression. Building on these considerations, we benchmarked selected BO paradigms against human corticogenesis with the aim of templating such a resource through a framework that allows its adaptation and growth as the neural modeling field continues to mature. Specifically, we profiled in-house a cohort of cortical BO (CBO) [18, 19], derived from multiple individuals and differentiation rounds over 200 days, and integrated it with (i) our in-house cohort of primary samples, (ii) publicly available transcriptomic data from BO different for degree of guidance, and (iii) prenatal cortical samples of the BrainSpan Atlas (BS). The characterization of the gene expression landscape of BS and CBO led to the definition of co-expression patterns relevant for human prenatal corticogenesis and CBO differentiation, as well as to the ranking and categorization in functional domains of their transcriptional hubs. We then cross-compared primary samples, CBO and other BO systems [14, 20, 21]. These analyses revealed the overlap between co-expression patterns of prenatal cortex and CBO, allowed their visualization in the selected BO systems, and pointed at partial heterochronicity in the transcriptional recapitulation of corticogenesis by different BO methods.

In sum, our analyses represent a benchmark for the interrogation of the transcriptional networks characterizing human prenatal corticogenesis and CBO differentiation and uncover their modulation also in other BO protocols of widespread use for neuropsychiatric disease modeling. By comparing co-expression patterns of human corticogenesis and BO differentiation across protocols, our work represents a template for the benchmarking of BO and their applications in disease modeling.

RESULTS

Gene co-expression analysis highlights the transcriptional programs of the prenatal human corticogenesis

To categorize the transcriptional hubs of the developing human cortex, we took advantage of the BrainSpan Atlas (BS), the most comprehensive transcriptional characterization of the human brain encompassing both fetal and post-natal/adult stages.

In a landmark effort, Miller et al. profiled by microarray fetal brain at mid-gestation [22], establishing a first reference transcriptional atlas. More recently, a multi-modal characterization of human brain was performed [23]. We applied a similar approach at the transcriptome level, focusing our effort specifically on the prenatal cerebral cortex. This allowed us to reconstruct the transcriptional circuitries of the human fetal corticogenesis with increased

resolution. We selected a total of 162 data points (Fig. 1A) from cerebral cortex at post-conceptual weeks (PCW) 8–37.

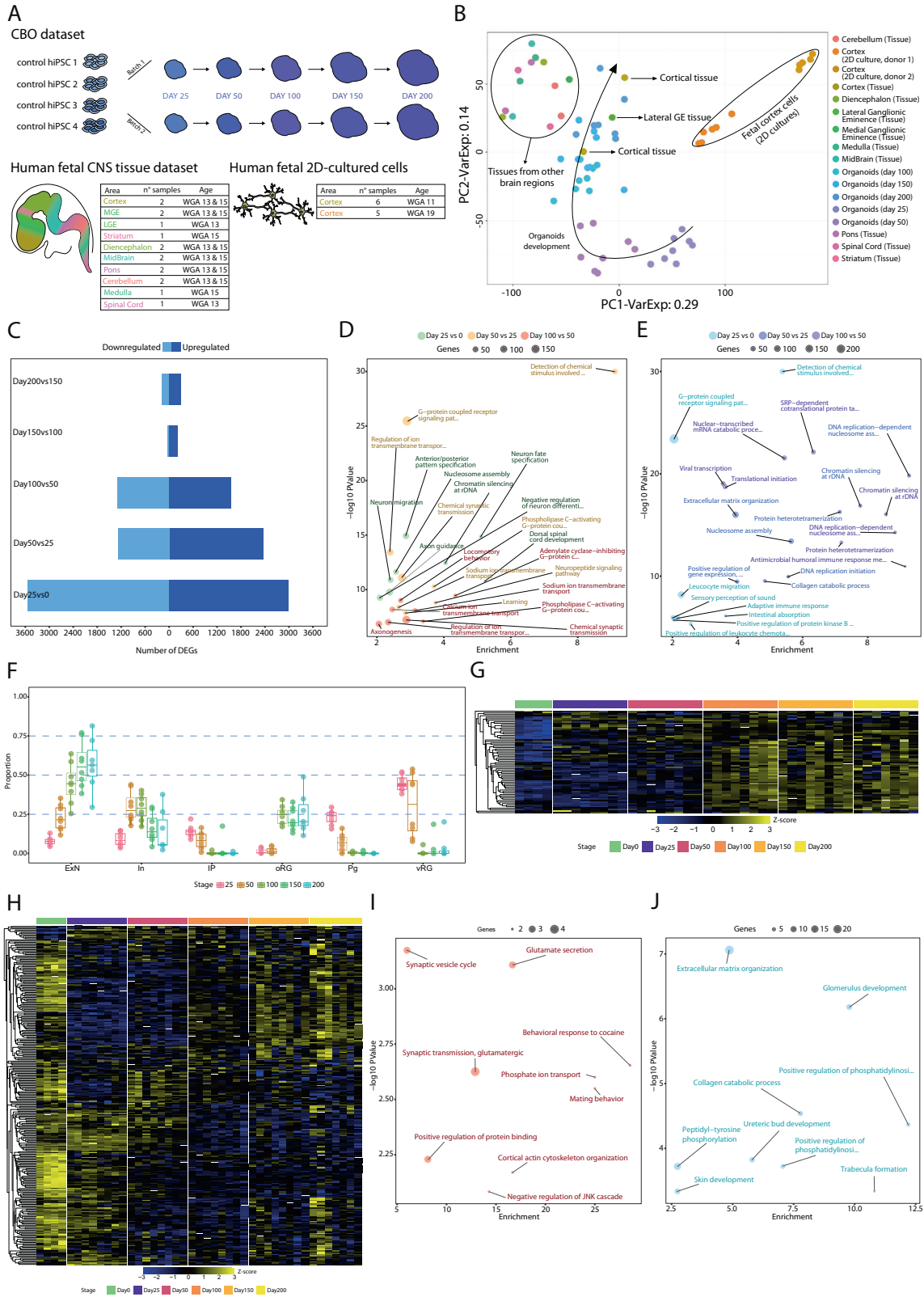
Principal component analysis (PCA) identified developmental stage as the main driver of sample differences, with PC2 separating very early stages (PCW 8–9) from early ones (PCW 12–13), followed by a progression of time-points in PC1 (Fig. 1B). Likewise, stage-wise correlation analysis detected a first shift in the transcriptional landscape from PCW 8–9 to 12–24 and a second, less clear-cut, characterizing PCW 25–37 (Supplementary Fig. 1A). The same analysis on BS post-natal cortical samples showed higher homogeneity compared to prenatal time-points (Supplementary Fig. 1B), indicating a more profound transcriptional evolution during the fetal phase, especially till PCW24.

We applied gene ontology enrichment analysis (GO) to the top 600 genes associated with PC1 or PC2 according to PCA loadings. We retrieved for PC1 categories associated with ion channels, lipid metabolism, and transcriptional regulation, while we found terms related to neuronal maturation and cell division for PC2 (Fig. 1B; Supplementary Fig. 1C, D).

To identify the transcriptional programs regulating corticogenesis in an unsupervised manner, we applied weighted gene co-expression network analysis (WGCNA), uncovering 17 gene co-expression modules (Fig. 1C, Supplementary Fig. 2A–C and Supplementary File 1). We summarized the co-expression profile of each module by its first PC (module eigengene, ME) and related it to developmental stage, highlighting several modules as positively or negatively correlated. BS_Turquoise, BS_Pink, BS_Grey60, and BS_Midnightblue showed positive correlation, while BS_Black and BS_Green were negatively associated with developmental progression (Fig. 1C). Other modules displayed changes in more restricted time windows (e.g. BS_Yellow, BS_Blue, BS_Red, Fig. 1C). The functional characterization of BS modules pointed to clear-cut biological domains for several of them, such as glutamatergic transmission and synapse for BS_Turquoise, ion channels for BS_Pink, DNA replication for BS_Black, and cell division for BS_Yellow (Fig. 1D–G). Finally, we reconstructed the co-expression network selecting the top-75 genes (according to intramodular connectivity) and then applied network analysis. Central nodes of BS_Turquoise and BS_Pink networks were related to neuronal functions and included synaptic proteins (CAMK2N2, CAMK2B, and GDA), receptor subunits (GRIN1, GRIN3A) and potassium channel subunits (KCNT1, KCNQ4, KCNC3, KCNC4) (Fig. 1D–G). BS_Midnightblue included the upper layer neuron markers CUX2 and SATB2 among its most central genes (Supplementary Fig. 2D). BS_Yellow and BS_Black hubs were strongly enriched in cell cycle genes.

The reconstructed networks also encompassed genes less studied in corticogenesis (e.g. for BS_Turquoise CATSPERZ, HAGHL, ADAMTS8, and KCN4-TEX40), thus allowing us to hypothesize their involvement by a guilt-by-association approach (Fig. 1D–G and Supplementary Fig. 2D).

Overall, we identified transcriptional circuitries related to developmental transitions and functional domains of human corticogenesis, reconstructing relevant gene co-expression networks.



Main modules, their behavior, and the relevance of every gene composing them are available in Supplementary File 1. To our knowledge, this collection represents the first resource of gene co-expression networks specifically focusing on prenatal corticogenesis that can be used to explore genes of interest in relation to physiological or pathological conditions.

Cortical brain organoids globally resemble the developing human fetal cortex and evolve in two steps

Upon characterization of the transcriptional dynamics defining human corticogenesis, we investigated the extent of their recapitulation in BO, with an experimental design that allowed us to measure interindividual and technical variability.

Fig. 2 Principal component and differential expression analyses uncovered CBO temporal dynamics. **A** Experimental design, detailing the number of individuals, batches and time-points for CBO and number of samples, brain area and WGA for fetal samples. **B** PCA on CBO, fetal brain tissues and 2D cultured fetal cortical progenitors, distinguished by dot color. The arrow highlights the distribution of CBO samples throughout stage progression. **C** Number of DEGs (FDR < 0.05, log₂FC > 1 as absolute value) from stagewise differential expression analysis in the CBO dataset. Barplots show the number of upregulated and downregulated genes for each comparison. **D, E** Bubble plots showing the functional characterization of upregulated (**D**) and downregulated (**E**) genes by functional enrichment analysis for the following comparisons: (I) Day25 vs Day0; (II) Day50 vs Day25; (III) Day100 vs Day50. The top-8 GO categories from the Biological Process domain of the GO are reported for each comparison (details are reported in Supplementary Table 3). **F** Boxplots displaying 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) from bulk deconvolution in each stage of CBO differentiation. **G, H** Heatmaps showing the behavior of cortex-specific upregulated (**G**) and downregulated (**H**) DEGs in the CBO dataset. Values are shown as Z-scores calculated on the expression values. **I, J** Bubble plots showing the functional characterization of the genes in (**G**) and (**H**), respectively.

CBO has been shown to recapitulate key events of human corticogenesis [5, 18, 19]. Indeed, immunohistochemical characterization confirmed the presence of key population markers for neural stem cells (SOX2), cell cycle (KI67), apical progenitors (PAX6), intermediate progenitors (TBR2), outer radial glia (HOPX), layer-specific neuronal markers (BCL11B, SATB2), and astrocytes (GFAP) (Supplementary Fig. 3 and 4 and [5, 18]).

We analyzed the transcriptional landscape of in-house differentiated CBO generated as previously described [18], for a total of 39 single-organoid and 4 hiPSCs samples from 4 control lines profiled over 200 days. We profiled single organoids to tackle technical variability of differentiation and we analyzed 2 independent organoid batches to measure reproducibility. In addition, we profiled a cohort of primary fetal CNS tissues (weeks of gestational age -WGA- 13 and 15) and 2D cultures (WGA 11, donor 1, and 19, donor 2) for a total of 4 individuals (described in [5], Fig. 2A). This dataset, having been processed as CBO in terms of sample preparation, library generation, sequencing platform, and computational pipelines, allowed a direct comparisons between the fetal tissue and the organoid model.

PCA showed temporal evolution of CBO with early stages forming individual clusters and later time-points resulting more intermingled. CBO evolved towards the fetal tissue, with clustering of mature organoids in proximity of fetal cortex. Conversely, 2D cultures clustered apart (Fig. 2B). Specific focus on CBO and relative functional analysis allowed us to associate PC1-driving genes to cell cycle and neuronal differentiation (Supplementary Fig. 5A–C).

To quantitatively characterize CBO evolution, we performed stage-wise differential expression analysis (DEA) comparing each stage against the previous. This approach highlighted a biphasic differentiation dynamic, with fast changes until day 100 followed by subtler modulations (Fig. 2C). The fast-evolving phase was characterized by a large number of differentially expressed genes (DEGs), related to neuronal fate commitment and maturation (upregulated) and cell cycle, transcription, and translation dynamics (downregulated) (Fig. 2D, E). The number of DEGs decreased considerably in the slow-evolving phase and related functional categories were generally less clear-cut (Supplementary Fig. 5D, E) [24–30].

We then analyzed dynamic changes among DEGs by visualizing the fold-change of consecutive comparisons (e.g. day 50vs25 against 100vs50, etc). Neuronal differentiation regulators, such as EOMES, LHX2, and FEZF2 [31–33], were modulated dynamically in the fast-evolving phase, while we detected upregulation of the astrocytic markers HEPACAM, AQP4, AGT, and APOE in the slow-evolving phase [34–37]. We also observed upregulation of GABAergic interneuron markers after day 100, in line with other studies [38, 39] (Supplementary Fig. 5F).

We then estimated CBO cell-type proportions with bulk deconvolution exploiting a scRNAseq atlas of the developing human cortex as reference [40]. Deconvolution methods have already been employed to estimate cell type proportions in the fetal brain [23] and have been shown to give reliable estimates when benchmarked against immunohistochemical quantification

of the main adult brain cell types, overcoming some of the selection biases affecting single-cell estimations [41]. From day 25 till day 100 CBO showed increase of excitatory neurons mirrored by a drop in ventricular radial glia, cycling progenitors, and intermediate progenitors. Outer radial glia increased from day 100 onwards (Fig. 2F).

Lastly, we identified cortex-enriched genes by comparing the cortical samples of our in-house fetal brain tissue dataset (WGA 13, 15) against hiPSCs and subsequently excluding common DEGs found with the same approach for other brain areas. We analyzed expression of cortex-enriched DEGs in CBO and found upregulation of genes related to glutamatergic neuron function alongside downregulation of ECM genes (Fig. 2G–J). In sum, we determined the transcriptional dynamics of CBO differentiation and we found that CBO resembled the transcriptome of human fetal cortex in terms of key cell populations and transcriptional programs, with a period of vast transcriptional changes followed by a plateau after day 100 in culture. This pattern of differentiation is in line with what has already been reported previously in BO [14] and reflects the shift from a first phase with predominance of progenitors to a second in which maturing neurons become the predominant population, which occurs within 100 days of differentiation for CBO.

WGCNA identifies cortical brain organoids' differentiation trajectories towards the glutamatergic fate

To identify transcriptional patterns in CBO differentiation, we applied WGCNA and identified 14 co-expression modules (Supplementary Fig. 6A–C, Supplementary File 2). Correlation with differentiation stage uncovered as strongly correlated modules the CBO_Turquoise and CBO_Black (positive correlation) and CBO_Brown and CBO_Blue (negative correlation) (Fig. 3A). We analyzed the behavior of CBO_Turquoise and CBO_Black ME in the 8 samples across time-points (Fig. 3B) detecting a steady increase over time. While CBO_Turquoise modulation was reproducible across replicates, the CBO_Black was more variable, pointing to a small set of genes (180 genes against the 3279 of the CBO_Turquoise) with a less robust behavior. The variability of this subset of genes was observed mainly in 2 samples, which, however, showed patterns in line with the other samples for other modules.

GO enrichment analysis for CBO_Turquoise and CBO_Black retrieved terms related to neuronal fate commitment and maturation. Network reconstruction confirmed CBO_Turquoise hub genes as related to neurotransmission and synaptic function (e.g. NRXN2, AGAP2, SLC4AE, GABBR1, and GRM5), while the CBO_Black network comprised several transcription factors related to excitatory neuron identity (NEUROD6, SLA, BCL11B, RORB) (Fig. 3B, C). CBO_Brown and CBO_Blue ME were instead decreasing along differentiation; the first was associated to DNA replication and cell cycle, while the second to more general functions such as transcriptional and translational regulation (Fig. 3B, C). WGCNA also detected modules characterized by non-monotonic trends through differentiation. Among them, CBO_Green and CBO_Red showed levels dropping from Day25 to Day50 and increasing

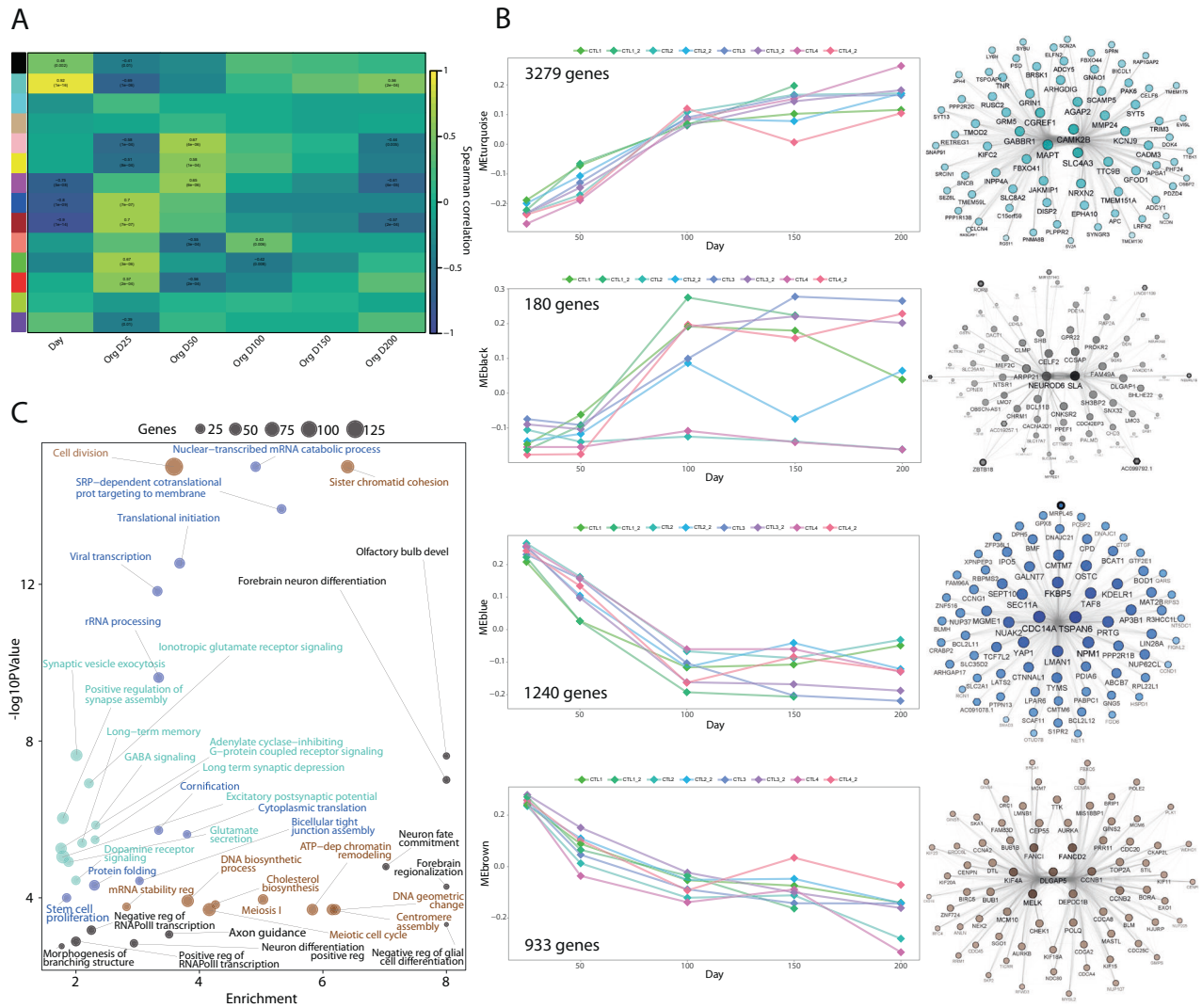


Fig. 3 Cortical brain organoids evolve towards the glutamatergic fate. **A** Heatmap showing the correlation between gene modules (summarized as ME) identified by WGCNA in cortical brain organoid (CBO) dataset and differentiation stage either as a continuous variable (Day) or a categorical variable for each time-point. Correlation coefficient and *P*-values are reported for significant correlations (Spearman *P*-value < 0.01). **B** Characterization of CBO_Turquoise, CBO_Black, CBO_Blue, and CBO_Brown gene modules. For each module, the ribbon plot visualizes the behavior of the module eigengene throughout differentiation time-points; each dot represents a data point, while the line connects time points of the same replicate (line and differentiation batch). Network reconstruction for the top-75 genes of each module, selected according to the intramodular connectivity value. Degree, closeness, betweenness, and eigenvector are represented by node label transparency, node color darkness, node border width, and node size/node label font size, respectively. Node shape represents gene biotype, with protein-coding genes as circles. **C** Bubble plot depicting the results of GO enrichment analysis for the four modules: *P*-values and enrichment scores are reported (details are reported in Supplementary Table 3).

again at late stages. Both modules were enriched in genes related to cell adhesion and ECM organization (Supplementary Fig. 6D, E).

In summary, we time-resolved the transcriptional evolution of CBO by identifying co-expression patterns and transcriptional hubs driving their differentiation. We generated a knowledge base that classifies genes in distinct modules of behavior along organoid differentiation and links them to specific functional domains. We found overall consistency among lines from independent individuals and replicates. Main modules, their behavior, and relevance of every gene composing them are available in Supplementary File 2.

Benchmarking of brain organoids against prenatal human corticogenesis reveals heterochronicity of differentiation across protocols

To compare CBO against other protocols and evaluate them versus the fetal cortex, we selected three external BO datasets for

which RNAseq was publicly available. These datasets of archetypal BO protocols, different for degree of guidance and culture conditions, included samples at comparable time-points (from 0 to 100 days). The selected BO datasets were: (i) minimally-guided neural organoids (MGO) [14]; (ii) forebrain organoids (FO) [20]; (iii) telencephalic aggregates (TA) [21], including a first step of 2D differentiation (Fig. 4A and Supplementary Fig. 7).

Stage-wise DEA revealed also for MGO, FO, and TA a decrease in the number of DEGs as differentiation progresses (Fig. 4B). However, the slow-evolving phase, which started between day 100–150 in CBO (Fig. 2C), was anticipated for all other protocols (day 40–60) (Fig. 4B). We then compared the DEGs identified for each stage transition in each external BO dataset to those retrieved for CBO (Supplementary Fig. 8A–C). The results highlighted a similarity in the transcriptional evolution of the models, with a high degree of overlap especially for MGO. In line with DEA, we found a more compressed time frame in the evolution of the

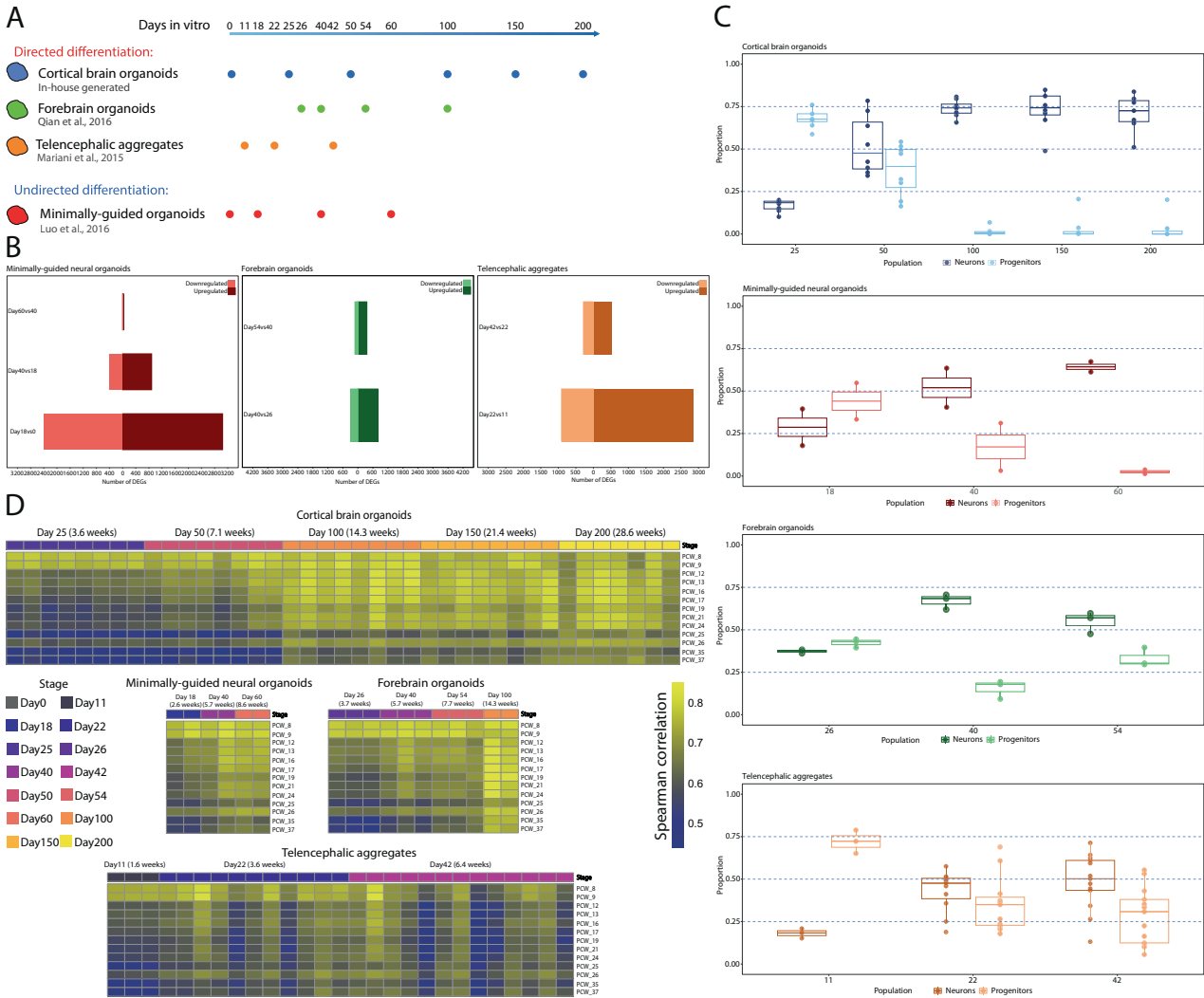


Fig. 4 Brain organoids differentially recapitulated the timing of corticogenesis. **A** Schematic representation of in-house and external BO datasets and relative differentiation time points (CBO cortical brain organoids, FO forebrain organoids, TA telencephalic aggregates, MGO minimally-guided neural organoids). **B** Stage-wise differential expression analysis in MGO, FO, and TA. Barplots reporting the number of DEGs (FDR < 0.05, log2FC > 1 as absolute value), shown as upregulated and downregulated genes for each comparison. **C** Boxplots reporting the fraction of Progenitors and Neurons estimated by bulk deconvolution in each organoid dataset. **D** Correlation analysis of organoid datasets against the BS fetal cortex. Heatmaps visualize the correlation coefficient (Spearman correlation) for each organoid data point and each BS PCW.

other models compared to CBO, given the time points for which the overlap was significant.

Bulk deconvolution analysis focusing on two broad cell populations (early progenitors and neurons) showed progenitors as predominant in all models at early stages (Fig. 4C). In CBO, proportions of the two populations were comparable at day 50, with neurons prevalent from day 100 onwards. MGO, FO, and TA showed a similar but accelerated dynamic compared to CBO, again indicating a more rapid transcriptional maturation.

We then tested the transcriptional similarity of each model towards BS fetal cortex (Fig. 4D). Whole-transcriptome correlation showed for CBO a gradual increase of similarity towards mid and late PCW over time. We observed a more time-compressed evolution among MGO and FO, which already by day 60 showed an extent of similarity with late PCW that CBO reached only by day 100. For CBO and TA, the dataset encompassed organoids generated from different control individuals. CBO demonstrated robust reproducibility across genetic backgrounds and batches of differentiation, while TA were less tolerant to interindividual variability. These observations were further confirmed by using CBO as reference (Supplementary Fig. 8D).

Overall, several lines of evidence pointed towards heterochrony across BO models in recapitulating the transcriptional modulations of corticogenesis. Although other experiments including more BO models and the same genetic backgrounds for all differentiation methods are needed to fully characterize this phenomenon, this is the first report of such differences in timings of differentiation of BO, which is an essential variable to control for pinpointing specific developmental transitions and disease phenotypes [42].

Gene signatures specific for brain cell subpopulations and functions unveil dynamics of brain organoid differentiation

We then examined how specific hallmarks of brain cell subpopulations and functions evolve during BO differentiation. To this end, we compiled a catalog of genes by manually curating new signatures from literature and by using signatures proposed in published studies [10, 15]. We observed their expression dynamics across BO and the developing cortex. Expression levels of signature genes are reported in Fig. 5 and Supplementary Figs. 9 and 10.

Fetal cortex showed drastic reduction of cell cycle, neural stem cell, apical progenitor, and intermediate progenitor gene levels

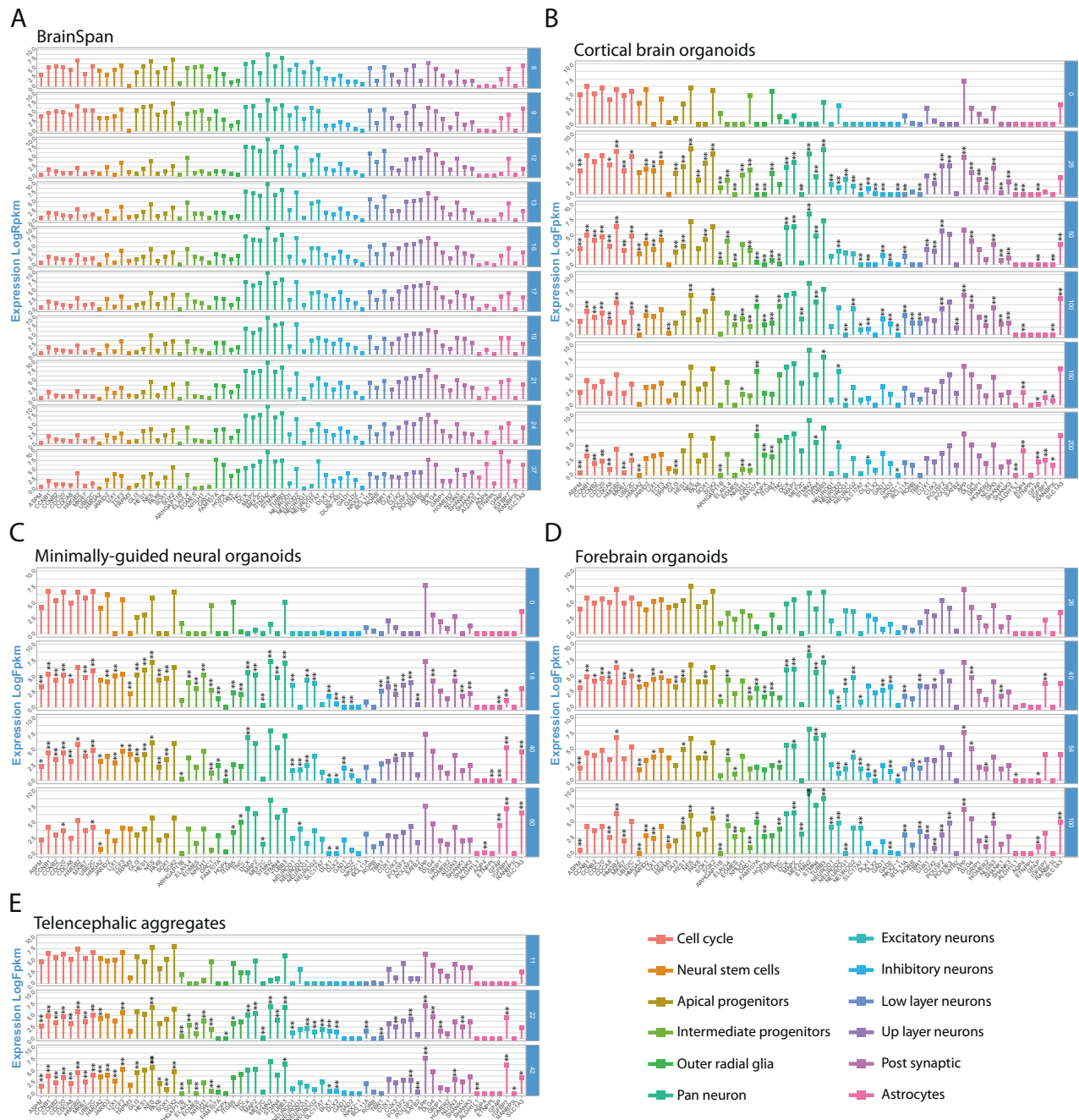


Fig. 5 Gene expression levels of signatures related to cell population identity and function in BO. Expression levels (Log₂Fpkpm or Log₂Rpm) in BS (A), CBO (B), MGO (C), FO (D), and TA (E) along time-points are reported in each dataset as the mean value across replicates. Each bar color corresponds to a specific signature, as reported in the plot legend. For brain organoids (B–E), asterisks identify genes that result significantly modulated in stage-wise differential expression analyses (comparison of each differentiation stage with the previous) reported in Fig. 2 for CBO and Fig. 4 for the other protocols. **: FDR < 0.05; *: Conventional *P*-value < 0.05.

between PCW 9–12, in line with WGCNA results. We found in BO models a less abrupt reduction of cell cycle and apical progenitor markers. The intermediate progenitor signature was found to peak at different time-points across models.

Excitatory neuron markers were particularly expressed at early PCW in the fetal cortex, a trend mirrored by CBO; MGO, FO, and TA showed a more persistent expression throughout differentiation.

oRG genes were well expressed in BS at PCW 8–9, decreased from PCW 12 and increased again at later stages, when astrocyte markers also appeared. CBO showed constant upregulation of the whole oRG signature starting from day 100, while for FO this was

observed already at day 40; MGO and TA showed a partial expression at the available time-points.

Considering that the examined BO protocols were based on different degrees of guidance and culturing conditions, we then looked at brain area-specific genes, markers of off-target tissues, markers of neurotransmission, and stress-related genes (detailed in Supplementary Fig. 9, Supplementary Fig. 10, Supplementary Table 2). As markers of metabolic stress, we looked at the expression of glycolytic and ER-stress genes [15]. We found fluctuations of expression of those genes throughout cortical development as well as in BO, however, without indications of a

sustained and strong increase at advanced time-points. In conclusion, by interrogating prenatal cortex and different BO datasets with literature-curated gene signatures, we highlighted biological domains for which each of these datasets manifested similarities to human fetal cortex development, as well as temporal peculiarities in modulating specific neural population markers.

Brain organoids differentially capture distinct transcriptional patterns of fetal corticogenesis

The knowledge base of gene co-expression modules we reconstructed for corticogenesis and CBO differentiation allowed us to directly compare modules identified in cortex/CBO and follow their behavior in the other BO models. First, we measured the overlap between the two sets of WGCNA modules (Supplementary Fig. 11). Fetal BS_Turquoise, BS_Pink, BS_Midnight-blue, and BS_Grey60 genes, steadily increasing their expression along corticogenesis, significantly overlapped with CBO_Turquoise, which showed the same trend. Similarly, BS_Yellow and BS_Black, functionally associated to cell cycle, showed steady decrease over time and overlapped with CBO_Brown, which had similar behavior and functional characterization (Supplementary Fig. 11).

We then sought to evaluate the relevance of the identified gene modules in the molecular circuitries that are affected in neurodevelopmental disorders and psychiatric conditions (Supplementary Fig. 12). We performed an overlap analysis to quantify the enrichment of genes from SFARI and Development Disorder Genotype - Phenotype Database (DD2P); moreover, we retrieved from the GWAS Catalog risk genes for four psychiatric disorders (Attention Deficit Hyperactive Disorder, Autism Spectrum Disorder, Schizophrenia, Unipolar Depression) and two non-psychiatric conditions (Diabetes Mellitus, Inflammatory Bowel Disease).

For the fetal cortex, we found a significant overlap found for the modules identified as related to neurogenesis and neural maturation (e.g. BS_Grey60, BS_Midnightblue, BS_Turquoise) and genes genetically associated to ASD from the SFARI repository, as well as risk genes for several psychiatric disorders (Supplementary Fig. 12A, C). Similarly, we identified the enrichment of genes associated to psychiatric conditions specifically in CBO gene modules associated to neuron commitment and maturation (CBO_Turquoise, CBO_Black, Supplementary Fig. 12B, D). We then analyzed the behavior of the same BS modules in each organoid dataset by following their ME across differentiation, revealing a close resemblance to fetal cortex trends (Fig. 6A, B). In a complementary approach, we applied the same analysis on the most relevant modules detected in CBO. The behavior of CBO_Turquoise and CBO_Brown in fetal cortex and other BO protocols was consistent with the one described in CBO. CBO_Black genes induction was variable across models and replicates (Supplementary Fig. 13A). CBO_Blue genes behaved similarly in all BO, and showed instead a non-monotonic behavior in BS, with peaks of expression at very early and very late PCW (Supplementary Fig. 13B).

Finally, we analyzed how BO recapitulated patterns of non-monotonic expression through time. BS_Red, BS_Blue, and CBO_Red, CBO_Green significantly overlapped between the two networks. The visualization of BS_Red and BS_Blue trends in BO revealed differences across protocols. CBO mostly displayed the same U-shape found in BS, although with a milder drop at intermediate stages. Conversely, MGO showed very little variation, FO showed a drop at day 100, and TA showed an opposite trend for the red module and no change for the blue one (Fig. 6C). Likewise, CBO and BS behaved similarly for the set of CBO_Red and CBO_Green genes, while these modules showed no, little or opposite variation in MGO, FO, and TA (Supplementary Fig. 13C). Among the functions of this set of modules, ECM and cell adhesion were prevalent.

Together, these results suggested that monotonic gene expression dynamics related to neuron specification and progenitor

proliferation are shared between fetal cortex and CBO and are largely recapitulated by the analyzed BO paradigms. Conversely, trends of non-monotonic expression are largely recapitulated in CBO and less robustly in other BO systems at the examined time-points. The overlap between co-expression patterns of development/differentiation with genes related to neurodevelopmental and psychiatric disorders further substantiated the ability of our WGCNA approach to pinpoint gene modules that are meaningfully related from a functional and pathogenic point of view. This confirms the relevance of the identified networks for exploring key genes for the physiology and pathology of corticogenesis. Importantly, this analysis also confirmed the preservation of transcriptional programs relevant for disease modeling in CBO.

DISCUSSION

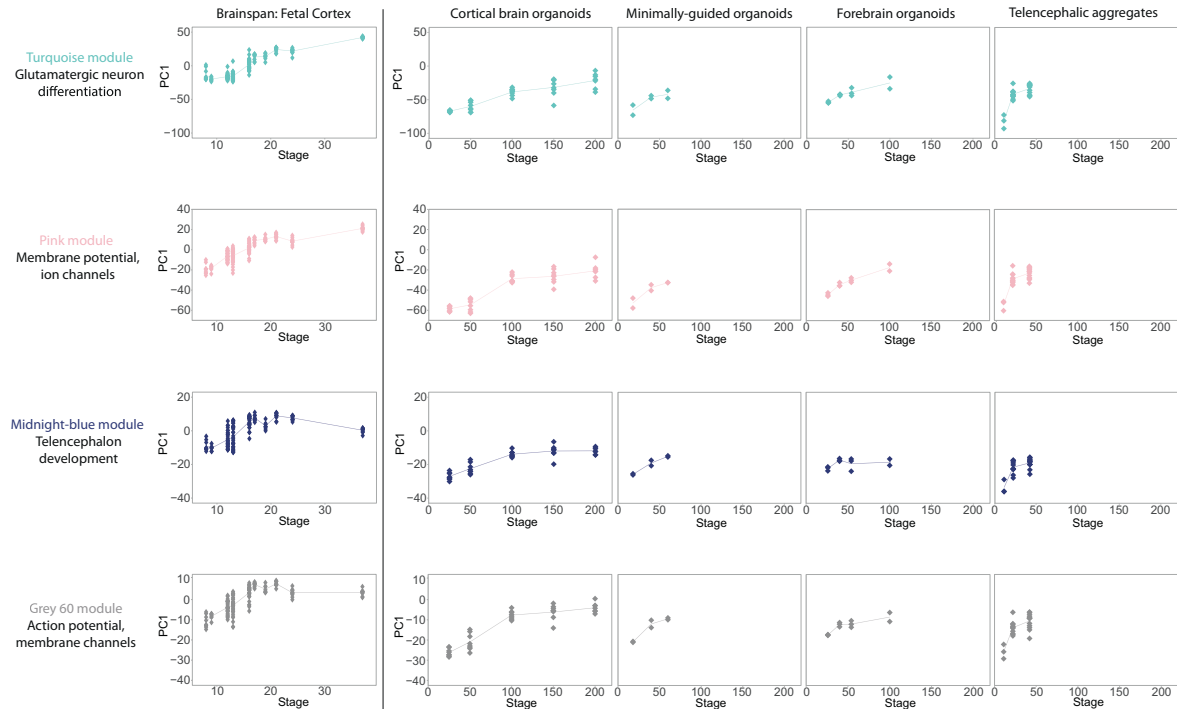
Despite the complexity of neuropsychiatric conditions, in terms of both phenotypic and time course heterogeneity, it is now well established that BO can allow the discovery of relevant developmental endophenotypes that are starting to illuminate disease pathogenesis with first inroads also in terms of patients' stratification and drug discovery [8, 43]. The ambition is thus rapidly progressing from the initial focus on monogenic syndromes of high penetrance to the more high-throughput and challenging interrogation of polygenic loadings for mental health vulnerability, including in terms of the developmental antecedents of its later onset manifestations. Yet, the more central neurodevelopment becomes to our understanding of mental health, the more we need to gain a full understanding of its experimental recapitulation *in vitro*, so as to define clear benchmarks across systems and thereby empirically guide the most appropriate disease modeling designs. Towards this goal, here we first defined co-expression patterns characterizing human corticogenesis *in vivo* and *in vitro* and then organized them in interactive tables where gene relevance can be visualized, together with the overall behavior of these patterns along fetal cortex development and CBO differentiation (Supplementary File 1 and 2). The generation of this knowledge base allowed us to perform cross-comparisons between BS, CBO, and other BO paradigms, quantifying the preservation of co-expression patterns of corticogenesis in these models. To our knowledge, this represents the first resource of such kind. Similar analyses were performed on BS and BO [12, 14], however, without focus on prenatal cortex, accessibility to gene modules composition, or prioritization of transcriptional hubs (both known and novel). Thus, this knowledge base templates an open framework for the benchmarking of modeling studies and serves as a platform to interrogate when choosing the experimental system that best suits specific modeling needs or research questions. Several studies provided insights on the variability of BO, with considerable efforts of standardization. Our results demonstrated CBO reproducibility in recapitulating the transcriptional landscape of prenatal corticogenesis, in line with other reports [12, 44], with a developmentally relevant timing.

We observed that TA were more variable compared to CBO in terms of cell composition and global transcriptome when considering hiPSCs from different individuals. The same evaluation could not be carried out for MGO and FO due to the availability of a single genetic background.

Several analytical approaches (DEA, whole-transcriptome correlation with fetal cortex, bulk deconvolution) revealed a fast-evolving phase followed by a slow-evolving one during CBO differentiation. Among genes showing modulation also in the slow-evolving phase, we found upregulation of GABAergic neuron markers (such as CALB2, SCGN, SLC6A11), in line with previous studies [17, 38, 39, 45]. Among the functional categories emerging when analyzing the slow-evolving phase, we observed an upregulation of ECM-related genes. We also observed upregulation of genes related to mesoderm, however, the majority of them

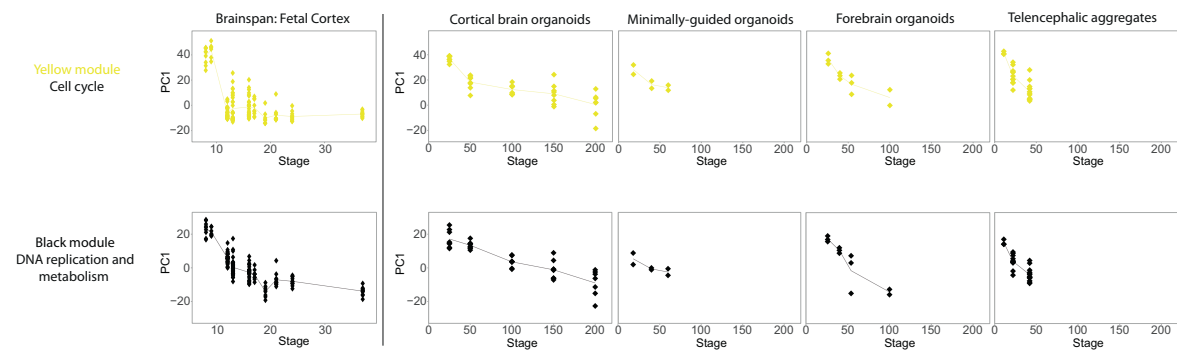
A

BrainSpan: monotonic behaviour with increased expression over developmental time



B

BrainSpan: monotonic behaviour with decreased expression over developmental time



C

BrainSpan: non-monotonic behaviour

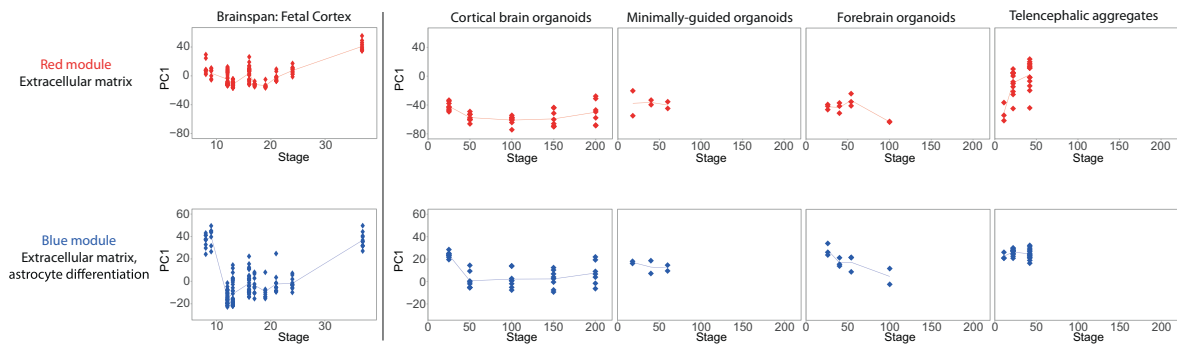


Fig. 6 Behavior of the module eigengene for BrainSpan gene co-expression patterns in BO. **A** Visualization of BS_Turquoise, BS_Pink, BS_Midnightblue, and BS_Grey60 in prenatal fetal cortex as well as in the BO datasets. PCA was performed to calculate the module eigengene on the BrainSpan and then applied to each BO dataset. Each dot represents a data point, while the line connects the median value for each post-conceptual week (BS) or differentiation day (BO). The same analysis and visualization was applied to BS modules with a decreased expression along development (**B**) or with a non-monotonic behavior (**C**).

is expressed also in astrocytes/reactive astrocytes [34, 46, 47]. The developmental timing of astrocytes appearance in CBO (day 100–150—week 14.3–21.4) aligned with the human prenatal cortex. The interplay between neuronal and astrocytic cells is increasingly recognized in neuropsychiatric disorders [48–50], therefore the characterization of their differentiation in BO compared to fetal cortex is important to guide disease modeling.

In the limited time-frame available for external BO datasets, we confirmed their biphasic evolution. However, we observed heterochronicity among protocols. We compared similar *in vitro* ages for all protocols analyzed, with the CBO dataset including also very late stages, and observed that MGO, FO and TA reached the slow-evolving phase earlier than CBO, with a greater proportion of neurons at earlier time-points. In addition, MGO and FO globally correlated earlier with late stages of corticogenesis and CBO differentiation.

A possible explanation for these findings could be related to the maintenance of a more immature stage for longer time due to the prolonged use of EGF/bFGF during CBO differentiation, which, however, ended up matching more faithfully the *in vivo* counterpart. It is also known that iPSCs can have different propensity to neuronal induction depending on the line itself and on culture conditions [51, 52]. However, both CBO and TA were generated from multiple hiPSCs lines reprogrammed from different individuals, thereby making improbable that differences across cell lines represent the main source of the observed heterochronicity among protocols. To our knowledge, this is the first description of such temporal dynamics in BO. Other reports comparing different paradigms focused on more heterogeneous time-points across protocols and other aspects of recapitulation [15–17]. Temporal dynamics recapitulation has remarkable relevance for disease-modeling, given the emergent convergence of molecular phenotypes of delay or acceleration in neuronal differentiation characterizing different clinical conditions [4, 53], and the eventuality of masking disease-relevant phenotypes if they are not correctly preserved [42]. Moreover, the tight link between neurodevelopment and risk/resilience of neuropsychiatric disorders positions the *in vitro* recapitulation of developmental timing at the core of the choice between BO systems, also with a view to balance the preservation of accurate developmental timing *vis a vis* the accelerated yield of more mature cell types involved in later onset neuropsychiatric disorders. Although further studies are needed to confirm and further dissect timing differences in BO, including multiple differentiation protocols from the same hiPSCs lines at the same time points, our analyses started to elucidate these dynamics, suggesting that CBO differentiation aligns with the *in vivo* temporality of corticogenesis.

Cellular stress in BO was previously reported by scRNAseq with the upregulation of glycolytic and endoplasmic reticulum stress signatures in several paradigms compared to primary samples [15, 16]. However, we did not observe a sustained increase of stress-related signatures along BO differentiation, corroborating the hypothesis of a homeostatic metabolic state in organoids rather than a deleterious stress increase over-time, in line with recent reports [11, 12].

Cross-visualization of monotonic BS co-expression patterns in CBO and vice versa revealed concordance in their behavior, a finding observed also for the other BO datasets. Overlap analysis with knowledge bases of genes associated to neurodevelopmental and psychiatric disorders indicated their selective enrichment in gene modules functionally associated to neuron fate commitment and maturation. Importantly, these modules were conserved from fetal brain to brain organoids, a relevant feature for disease modeling.

Modules with more complex temporal patterns were better recapitulated in CBO than in other protocols. Functional characterization of these modules pointed towards astrogenesis and ECM as predominant terms.

We also processed primary fetal tissues *in-house*, allowing direct comparisons against CBO because of the identical processing of the two datasets. The visualization of cortex-specific modulated genes in CBO highlighted again ECM. ECM genes have dynamic expression during corticogenesis, and have a role in regulating cortical folding, neuronal progenitor proliferation, and neuronal migration [54–56]. They have high expression in germinal zones and maturing/mature astrocytes [54, 57] and are important for the evolution of the human brain, with progenitor cells expressing ECM components at higher levels than in mice [58]. ECM also plays critical roles in synaptic plasticity and electrophysiological properties [59, 60], and its alteration has been linked to neuropsychiatric disorders [61, 62]. Our exposure of the differential regulation of ECM genes in BO compared to fetal cortex provides a relevant insight to consider when interpreting neuropsychiatric disease modeling datasets.

As already discussed, one limitation of this study, inherently due to the original design of the respective studies, is that the MGO and FO datasets derive from a single hiPSCs line. We therefore cannot exclude that the results we observed could be partially influenced also by the genetic background of the examined line for those BO models. Further studies are warranted to confirm the observations across different genetic backgrounds, building on the approaches we previously established for guiding optimal disease modeling designs by benchmarking transcriptional signal to noise detection *vis a vis* human genetic diversity [63]. Moreover, by inception this study did not aim to cover the complete landscape of BO protocols that have been developed so far. We focused our analyses on four selected protocols because of the range of epistemic objectives and attending culture conditions that they represent. While this could *prima facie* restrict the ordering reach of our work, we note that many if not most protocols currently used are characterized by minor adjustments and optimizations derived from these four archetypal protocols which, especially as far as CBO and MGO are concerned, represent the main forerunners and references of the field. Finally, the framework we have established in this work is easy to translate to other BO systems in follow-up studies and can thus be flexibly repurposed and harnessed, as the field progresses, as an analytical companion to the standardizing nomenclature effort we recently spearheaded [64].

Taken together, our results contribute to the definition of transcriptional footprints and dynamics specific to prenatal corticogenesis, representing a collection of prioritized known and novel hubs categorized in well-defined functional domains and made available as interactive tables (Supplementary File 1 and 2). Our approach describes the extent of *in vivo/in vitro* alignment of developmentally relevant processes and temporality, highlighting commonalities and differences across BO paradigms and providing a resource to be harnessed when modeling physiological or pathological human corticogenesis.

METHODS

hiPSCs reprogramming and maintenance

Skin fibroblasts from four healthy individuals were reprogrammed using non-integrating self-replicating mRNAs as previously described [65] (Stemgent, 00-0071) (CTL1, CTL3, and CTL4) or Sendai virus (CTL2) (CytoTune-iPS 2.0 Sendai Reprogramming Kit; Thermo Fisher Scientific, A16517). Fibroblasts of CTL1 (phenotypically normal male without intellectual disability or other physician-diagnosed neuropsychiatric diagnosis) were received from BC Children's Hospital, Vancouver. CTL2 (male) hiPSC line was received from the Wellcome Trust Sanger Institute. CTL3 (ShiPS-MIFF1 [66], male) hiPSC line was received from the university of Sheffield. Fibroblasts of CTL04 (female) were received from the GGDBbank, member of the Telethon Network of Genetic Biobanks (project no. GTB12001), funded by Telethon Italy, and of the EuroBiobank network. hiPSC were cultured in TeSR-E8 medium (Stemcell technologies, 05990), with daily media change, at 37°C, 5% CO₂, and 3% O₂ in standard

incubators. hiPSC were grown on matrigel-coated dishes (Corning, 354248) and passaged using ReLeSR™ (Stemcell technologies, 05872). Cell lines were routinely profiled through STR analysis to determine authenticity and to mycoplasma test. All lines were negative to mycoplasma. All subject signed an informed consent and the use of hiPSCs was approved by the ethical committee of the University of Milan.

CBO differentiation

CBO were generated using an adaptation of the previously described protocol published by Pasca et al. in 2015 [19], introducing orbital shaking on day 12 of differentiation as described in [18]. Briefly, 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 BE13-14E), 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 min and plating on ultra-low 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 μM Dorsomorphin (MedChem express, HY-13418A), 10 μM TGFβ inhibitor SB431542 (MedChem express, HY-10431), and ROCK inhibitor 5 μM. EB were grown in normal oxygen incubators. EB were left undisturbed for 1 day and at 48 h media change was performed with differentiation medium 1 without ROCK inhibitor. Dorsomorphin and TGFβ 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 Amphotericinβ to prevent mold formation. Before processing for transcriptomics, organoids were visually inspected and only organoids with round shape and with comparable size among lines were selected. Each line was differentiated and profiled at the following time points: iPSC stage (Day0), Day25, Day50, Day100, Day150, and Day200, in two different rounds of differentiation. Both differentiation rounds were profiled for each line, except for Day 200 for one of the lines, for which only one replicate was available. The cohort of sample is therefore composed by a total of 43 samples (39 organoids plus 4 iPSCs).

Immunostainings of CBO

CBO were harvested on day 25, 50, 100, and fixed 2 h in paraformaldehyde 4% (ChemCruz, sc-281692) on orbital shaking. Then, sucrose 30% was added for 12–16 h after a wash in PBS 1X. CBO were then embedded in cryostat embedding medium (Bio Optica 05-9801). Cryosection was 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 min and then left at room temperature for 45 min. Slides were then treated with 0.1 M Glycine pH 7.4 for 10 min to reduce autofluorescence. Subsequently, blocking solution was added (5% serum, 1% triton X-100 in PBS 1X) and the slides were incubated 30 min 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 h at room temperature with the secondary antibodies in antibody buffer. DAPI (Sigma, D9542) was added for 5 min 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). The following primary antibodies were used: BCL11B, 1 in 200, Rat, Abcam, ab18465; GFAP, 1 in 50, Rabbit, Sigma, G9269; KI67, 1 in 500, Rabbit, Abcam, ab15580; PAX6, 1 in 250, Rabbit, Biolegend, 561664; SOX2, 1 in 50, Goat, R&D system, AF2018; SATB2, 1 in 100, Mouse, Abcam, ab51502; TBR2, 1 in 200; Rabbit, Abcam, ab23345. 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.

Culture conditions for fetal cortical cell

Human fetal neural stem cell cultures were derived and maintained as previously described [5, 67]. They were derived from the cortex of WGA 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-βmercaptoethanol 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.

Download of external datasets

BrainSpan dataset: pre-processed Rpkv values for 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.

MGO, FO and TA datasets: bulk RNA sequencing data were downloaded from Gene Expression Omnibus using the relative article accession numbers (GSE82022, GSE80073, GSE61476, respectively). These protocols were selected because they represent a wide range of guidance and culture conditions. Moreover, they are references from which several derivations have been developed by several labs, with many of them being characterized by minor adjustments of these archetypal protocols which, especially for CBO and MGO, represent the main forerunners and references of the field.

RNA extraction and library preparation for RNA-seq

Total RNA was extracted from snap-frozen pellets of CBO at day 25, 50, 100, 150, 200, fetal cortical progenitors and fetal brain tissues 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 to 1 μg of total RNA. cDNA library quality was assessed with an 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.

RNAseq quantification for CBO, in-house fetal dataset, external brain organoids datasets

RNAseq FASTQ data were quantified at the gene level using Salmon (version 0.8.2 [68]). GRCh38 Genecode 27 was used as reference for quantification and annotation.

BrainSpan correlation analysis across developmental stages

The analysis was applied on BS specimens from prenatal and post-natal cortex. After selecting protein-coding genes, the mean expression was calculated for each stage (taking into account all the samples and sub-areas). Spearman correlation across samples was calculated and the correlation coefficient was visualized by heatmaps using pheatmap R package (<https://CRAN.R-project.org/package=pheatmap>).

Principal component analysis

BrainSpan: after selecting only samples from prenatal cortex, a filtering to discard not-expressed or low-expressed genes was applied by keeping genes with expression of at least 1 Rpkv in at least 1/4 of the samples (16,824 genes selected).

Internal fetal dataset and CBO: a total of 66 samples considering CBO and internal fetal samples was 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 17,759 analyzed. CBO dataset: a total of 39 samples considering CBO dataset (from day 25 to day 200). Gene filtering using a threshold of 2 cpm in at least 4 samples was used, resulting 16,901 genes used for the analysis.

For all datasets, PCA was computed using R `prcomp` function. For both BS and CBO PCA, gene loadings for PC1 and PC2 were retrieved from this analysis and the top 35 ones with positive and negative scores were visualized as lollipop graphs. The top 300 genes with the highest positive loading and the top 300 with the highest negative loading were selected to perform GO for biological processes using the TopGO package [69], relying on Fisher test and Weight01 method. P -value < 0.01 and enrichment > 1.75 were used as thresholds to select significantly enriched GO terms.

WGCNA

BrainSpan. *Weighted Gene Co-expression network generation and module identification:* After selecting only samples from prenatal cortex, not-expressed or low-expressed genes were discarded by keeping genes with expression of at least 1 Rpkm in at least 1/4 of the samples (16,824 genes selected). Samples from post-conceptual weeks 25, 26, and 35 were identified as outliers by the sample clustering, and therefore excluded from downstream analyses. Starting from this set of 16,824 genes measured in 157 specimens, a gene selection strategy was applied by calculating for each gene the coefficient of variation (CV) across the experimental conditions after log-transformation. A 65-percentile threshold was then imposed, thus selecting the 35% of genes showing the highest values of CV (5889 genes). A signed gene co-expression network was generated relying on WGCNA R package (version 1.64.1 [70]). The correlation matrix was calculated by applying a biweight mid-correlation and then transformed into an adjacency matrix by raising it to the power of $\beta = 18$. Topological Overlap Measure was calculated from the adjacency matrix and the relative dissimilarity matrix was used as input for average-linkage hierarchical clustering and gene dendrogram generation. 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` [71]).

Module-trait correlation: As phenotypic trait for module-trait correlation, module eigengenes were related to each sample developmental stage (PCW), considered either as a continuous quantitative variable or dichotomized as a series of categorical variable (dummy variables).

Module functional analysis: Gene ontology enrichment analysis for the Biological Process domain was performed on the genes belonging to each module of interest, using the list of 5889 genes selected for network generation as custom reference set. Analysis was performed by TopGO as described above. P -value < 0.01 and enrichment > 1.75 were used as thresholds to select significantly enriched GO terms.

Sub-network visualization and analysis in Cytoscape: Node and edge information for the selected modules were exported from the adjacency matrix, selecting for each module the top-75 according to the intramodular connectivity and imposing a threshold of 0.2 as minimum edge weight. Nodes and edges for each module were therefore imported in Cytoscape (version 3.8.2) for centrality analysis (CytoNCA [72]).

CBO. The same pipeline used to perform WGCNA on BS 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.

We excluded from differential expression analysis genes with biotypes that were not of interest (e.g. ribosomal RNA, antisense, snoRNA), and focused prevalently on protein_coding and long intergenic non-coding genes). Filtering on gene expression was applied by keeping genes with an expression of at least 2 cpm in at least 7 samples (15,663 genes selected). 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. 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 then characterized using the same approach described for BS.

For the generation of the bubble plot relative to CBO turquoise, black, blue and brown modules, REVIGO web tool was employed to summarize GO terms [73].

Differential expression analyses (DEA)

CBO: DEAs were performed for CBO in a stage-wise approach comparing each differentiation stage with the previous time-point using edgeR 3.20.9.

Genes with expression levels higher than 2 cpm in at least 4 samples were tested for differential expression (16,901 genes); after calculation of normalization factors, normalization and estimation of dispersion, differential expression was tested by glmFit approach. The information about line identity was used as a covariate in the statistical model. DEGs were selected imposing as thresholds FDR < 0.05 and absolute $\log_2FC > 1$. For each comparison, the number of DEGs, splitted in upregulated or downregulated, was represented by barplots. Functional enrichment analysis for the Biological Process domain of GO was performed by TopGO for every comparison dividing genes in up- and downregulated as setting the following parameters: 'weight01' as algorithm, 'fisher' as statistics and 15 as 'nodeSize' and 15 as 'minTerms'.

P -value < 0.01 and enrichment > 2 were used as thresholds to select significantly enriched GO terms.

Visualization of DEA results between the different sequential CBO comparisons was performed using scatter plots visualizing the \log_2FC of all tested genes; the color-code was set according to FDR values. Gene resulting differentially expressed in both sequential CBO comparisons (e.g. day 50vs25 against day 100vs50) were identified and divided in 4 quadrants. In this way, the behavior of genes in common between the two DEAs was analyzed, thus finding genes upregulated in both, downregulated in both or upregulated in one and downregulated in the other. The top 10 protein-coding genes in terms of absolute fold change for the four types of behaviors were labeled in the plot.

MGO, FO, and TA datasets: the same stage-wise DEA approach was applied for MGO, FO, and TA, analyzing each dataset independently. Genes with expression levels higher than 2 cpm in at least 2 samples were tested for differential expression with edgeR 3.20.9 (MGO: 15,339; FO: 16,585; TA: 16,522 genes). Line identity was used as a covariate for the TA dataset, while for MGO and FO only one line was available. DEGs were selected and visualized as described for CBO.

DEGs overlap: DEGs identified by stage-wise analysis in CBO were compared to the results of the same analysis on the other BO protocols. Overlap significance was tested by GeneOverlap R library (version 1.30.0, R version 4.1.1), considering the genes tested in CBO as universe. To retrieve transcriptional modulations robustly similar across protocols, we imposed stringent thresholds, selecting overlaps having P -value < 0.01 and odds ratio > 3. Results were visualized by heatmaps produced by the `drawHeatmap` function of the package. Only values with P -value < 0.01 and OR > 3 have been kept in the heatmap.

Cortex-specific genes determination and visualization of their behavior in CBO

DEAs of the different fetal tissues from our internal cohort versus hiPSC were performed for all areas including at least two samples using edgeR 3.20.9. Genes with expression levels higher than 2 cpm in at least 2 samples were tested for differential expression (17,275 genes). DEGs with FDR < 0.05 and absolute $\log_2FC > 3$ were considered for further analysis. Cortex-specific DEGs were determined by subtracting to DEGs of the cortex vs hiPSC comparison DEGs found in the analysis between other tissues and hiPSC. Functional enrichment analysis for GO biological processes was performed by TopGO for cortex-specific genes dividing them in up- and downregulated. P -value < 0.01 and enrichment > 2 were used as thresholds to select significantly enriched GO terms. The behavior of cortex-specific DEGs in the CBO dataset, including hiPSC, was visualized by heatmap using average-linkage hierarchical clustering for rows.

Bulk deconvolution

Proportions of cell populations were estimated applying a deconvolution approach based on the SCDC algorithm [74] using bulk-RNAseq counts as input and a scRNAseq dataset of the developing human cortex [40] as reference. The single cell raw count expression matrix was filtered to discard low-quality cells, by keeping cells compliant with the following threshold: mitochondrial RNA content < 5%; ribosomal protein RNA content < 50%; detected genes > 450 and < 3000; UMI counts > 750 and < 10,000; after filtering, 14,610 genes measured in 27,527 cells were selected for downstream steps. Starting from the clustering annotation performed in the original work, pericytes, microglia, endocytes and oligodendrocyte progenitors were not included, as we did not expect to find these cell type in CBO. The remaining cell populations were then grouped in the following 6 categories: ventricular Radial Glia (vRG); outer Radial Glia (oRG); Intermediate Progenitors (IP); cycling progenitors; interneurons: excitatory neurons. After deconvolution and for visualization purposes, the results were further grouped in two broader categories:

progenitors (cycling progenitors + vRG) and neurons (excitatory neurons + inhibitory neurons). Cells with uncertain cell-type assignments were removed (SCDC qc threshold = 0.65). Cell proportions were retrieved using the SCDC_prop function and were visualized as boxplots showing the proportion of every cell type per stage.

Analysis was performed in R, version 3.6.1.

Correlation of BO transcriptome versus BS or versus CBO

Transcriptome-wide correlation was calculated on Fpkm (Rpkms for BS) after selecting protein-coding genes and filtering out lowly-expressed ones (mean expression levels lower than 1 Rpkms in at least 75% of samples (BrainSpan) and 1 Fpkm (Brain Organoids)). Mean expression per gene per stage was calculated for all BS prenatal cortical samples. Correlation was computed in R using Spearman metrics and visualized as separated heatmaps for each comparison. The same approach was applied to perform correlation of external brain organoid datasets against CBO.

Literature-curated gene signatures visualization in BS and BO

Expression values (in log₂Fpkm or log₂Rpkms) for the gene signatures of interest were retrieved and the mean expression was calculated for each stage of every dataset, then visualized by lollipop plots.

Module overlap of BS and CBO WGCNA

Genes shared across the CBO and BrainSpan networks were selected for the analysis (2643); overlap across modules of interest was performed. Number of shared genes, odds ratio, and *P*-value across CBO and BrainSpan modules were calculated by the GeneOverlap R package. Results were visualized as dot plot where numbers (shared genes) were shown for odds ratio >1, dots were shown for those having also *P*-value <0.05. Color-code was assigned according to OR, dot size varied according to *P*-value.

Overlap of WGCNA gene modules and gene-disorder knowledge bases

The following gene-phenotype knowledge bases were considered for the overlap analysis: (I) Development Disorder Genotype-Phenotype Database (DD2P); (II) SFARI Genes; (III) the GWAS Catalog (EBI). SFARI and DD2P were downloaded respectively from <https://gene-archive.sfari.org/tools/> and <https://www.deciphergenomics.org/dd/ddgenes>. For DD2P, the overlap was tested with the complete database as well as on a subset related to the CNS, obtained by filtering the term 'brain' in the 'Organ' field.

GWAS catalog (<https://www.ebi.ac.uk/gwas/>) was interrogated to retrieve risk genes for 4 psychiatric disorders and two non-psychiatric conditions by searching for the indicated disease code: Attention Deficit Hyperactive disorder (EFO_0003888); Autism Spectrum Disorder (EFO_0003756); Schizophrenia (MONDO_0005090); Unipolar Depression (EFO_0003761); Diabetes Mellitus (EFO_0000400); Inflammatory Bowel Disease (EFO_0003767).

Overlap was performed for both Brainspan and CBO WGCNA gene modules, selecting the ones identified as the most interesting from correlation analysis: CBO_black, CBO_blue, CBO_brown, CBO_green, CBO_red, CBO_turquoise modules for cortical organoids and BS_black, BS_blue, BS_grey60, BS_midnightblue, BS_pink, BS_red, BS_turquoise, BS_yellow for the fetal cortex. Overlap significance was tested by GeneOverlap R library (version 1.30.0, R version 4.1.1). Overlaps were considered significant with an odds ratio (OR) higher than 1 and *P*-value lower than 0.01. Results were visualized as dot plot with numbers (shared genes) shown for OR > 1, dots shown for those having also *P*-value < 0.01. Color-code was assigned according to OR, dot size varied according to *P*-value.

Analysis of BS modules in BO and of CBO modules in BS and other brain BO

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. Likewise, module eigengene for each CBO WGCNA module of interest was calculated and predicted in BrainSpan and in external brain organoid datasets. Results were visualized as ribbon plots showing first principal component coefficients for each module along BrainSpan and organoid developmental time points.

Statistical analyses

Unless otherwise specified, all bioinformatic analyses were performed using R version 3.4.4 except for the bulk deconvolution analysis performed using R version 3.6.1. The statistical details of all analysis can be found in the relative figure legend and text of the "Results" section.

DATA AVAILABILITY

Bulk RNAseq data generated in this study have been deposited in ArrayExpress and will be made available upon publication at the accession number E-MTAB-11239. No new algorithms were developed in this work.

CODE AVAILABILITY

Code snippets for data processing and characterization/visualization of gene modules included as supplementary files.

REFERENCES

- Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental disorders: from genetics to functional pathways. *Trends Neurosci.* 2020;43:608–21.
- Cheroni C, Caporale N, Testa G. Autism spectrum disorder at the crossroad between genes and environment: contributions, convergences, and interactions in ASD developmental pathophysiology. *Mol Autism.* 2020;11:69.
- Villa CE, Cheroni C, Dotter CP, López-Tóbon A, Oliveira B, Sacco R, et al. CHD8 haploinsufficiency links autism to transient alterations in excitatory and inhibitory trajectories. *Cell Rep.* 2022;39:110615.
- Paulsen B, Velasco S, Kedaigle AJ, Pigoni M, Quadrato G, Deo AJ, et al. Autism genes converge on asynchronous development of shared neuron classes. *Nature.* 2022;602:268–73.
- Caporale N, Leemans M, Birgeron L, Germain P-L, Cheroni C, Borbély G, et al. From cohorts to molecules: adverse impacts of endocrine disrupting mixtures. *Science.* 2022;375:eabe8244.
- Amin ND, Paşca SP. Building models of brain disorders with three-dimensional organoids. *Neuron.* 2018;100:389–405.
- Shi Y, Wu Q, Wang X. Modeling brain development and diseases with human cerebral organoids. *Curr Opin Neurobiol.* 2020;66:103–15.
- Wang M, Zhang L, Gage FH. Modeling neuropsychiatric disorders using human induced pluripotent stem cells. *Protein Cell.* 2020;11:45–59.
- Teissier A, Pierani A. Wiring of higher-order cortical areas: spatiotemporal development of cortical hierarchy. *Semin Cell Dev Biol.* 2021;118:35–49.
- Amiri A, Coppola G, Scuderi S, Wu F, Roychowdhury T, Liu F, et al. Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* 2018;362. <https://doi.org/10.1126/science.aat6720>.
- Uzquiano A, Kedaigle AJ, Pigoni M, Paulsen B, Adiconis X, Kim K, et al. Proper acquisition of cell class identity in organoids allows definition of fate specification programs of the human cerebral cortex. *Cell.* 2022;185:3770–3788.e27.
- Gordon A, Yoon S-J, Tran SS, Makinson CD, Park JY, Andersen J, et al. Long-term maturation of human cortical organoids matches key early postnatal transitions. *Nat Neurosci.* 2021;24:331–42.
- Fleck JS, Sanchis-Calleja F, He Z, Santel M, Boyle MJ, Camp JG, et al. Resolving organoid brain region identities by mapping single-cell genomic data to reference atlases. *Cell Stem Cell.* 2021;28:1148–1159.e8.
- Luo C, Lancaster MA, Castanon R, Nery JR, Knoblich JA, Ecker JR. Cerebral organoids recapitulate epigenomic signatures of the human fetal brain. *Cell Rep.* 2016;17:3369–84.
- Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, et al. Establishing cerebral organoids as models of human-specific brain evolution. *Cell.* 2019;176:743.e17–756.e17.
- Bhaduri A, Andrews MG, Mancía Leon W, Jung D, Shin D, Allen D, et al. Cell stress in cortical organoids impairs molecular subtype specification. *Nature.* 2020;578:142–8.
- Tanaka Y, Kahir B, Xiang Y, Sullivan GJ, Park I-H. Synthetic analyses of single-cell transcriptomes from multiple brain organoids and fetal brain. *Cell Rep.* 2020;30:1682–1689.e3.
- López-Tóbon A, Villa CE, Cheroni C, Trattaro S, Caporale N, Conforti P, et al. Human cortical organoids expose a differential function of GSK3 on cortical neurogenesis. *Stem Cell Rep.* 2019;13:847–61.
- Paşca AM, Sloan SA, Clarke LE, Tian Y, Makinson CD, Huber N, et al. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat Methods.* 2015;12:671–8.
- Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C, et al. Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell.* 2016;165:1238–54.

21. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L, et al. FOXP1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell*. 2015;162:375–90.
22. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508:199–206.
23. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 2018;362. <https://doi.org/10.1126/science.aat7615>.
24. Obayashi S, Tabunoki H, Kim SU, Satoh J. Gene expression profiling of human neural progenitor cells following the serum-induced astrocyte differentiation. *Cell Mol Neurobiol*. 2009;29:423–38.
25. Dahl D, Bignami A. Immunohistological localization of desmin, the muscle-type 100 A filament protein, in rat astrocytes and Müller glia. *J Histochem Cytochem*. 1982;30:207–13.
26. Fink KL, López-Giráldez F, Kim I-J, Strittmatter SM, Cafferty WBJ. Identification of intrinsic axon growth modulators for intact CNS neurons after injury. *Cell Rep*. 2017;18:2687–701.
27. Borggrewe M, Grit C, Vainchtein ID, Brouwer N, Wesseling EM, Laman JD, et al. Regionally diverse astrocyte subtypes and their heterogeneous response to EAE. *Glia*. 2021;69:1140–54.
28. Manning TJ, Rosenfeld SS, Sontheimer H. Lysophosphatidic acid stimulates actomyosin contraction in astrocytes. *J Neurosci Res*. 1998;53:343–52.
29. 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*. 1994;28:303–16.
30. Magistri M, Khoury N, Mazza EMC, Velmeshev D, Lee JK, Bicciato S, et al. A comparative transcriptomic analysis of astrocytes differentiation from human neural progenitor cells. *Eur J Neurosci*. 2016;44:2858–70.
31. Arnold SJ, Huang G-J, Cheung AFP, Era T, Nishikawa S-I, Bikoff EK, et al. The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. *Genes Dev*. 2008;22:2479–84.
32. Chou S-J, Tole S. Lhx2, an evolutionarily conserved, multifunctional regulator of forebrain development. *Brain Res*. 2019;1705:1–14.
33. Chen B, Wang SS, Hattox AM, Rayburn H, Nelson SB, McConnell SK. The Fezf2-Ctip2 genetic pathway regulates the fate choice of subcortical projection neurons in the developing cerebral cortex. *Proc Natl Acad Sci USA*. 2008;105:11382–7.
34. Sloan SA, Darmanis S, Huber N, Khan TA, Birey F, Caneda C, et al. Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron*. 2017;95:779–790.e6.
35. Iliff JJ, Wang M, Liao Y, Plogg BA, Peng W, Gundersen GA, et al. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β . *Sci Transl Med*. 2012;4:147ra111.
36. Sherrad M, Liu X, Zhang X, Sigmund CD. Nuclear localization of angiotensinogen in astrocytes. *Am J Physiol Regul Integr Comp Physiol*. 2005;288:R539–46.
37. Trevino AE, Müller F, Andersen J, Sundaram L, Kathiria A, Shcherbina A, et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell*. 2021;184:5053–5069.e23.
38. Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, et al. Complex oscillatory waves emerging from cortical organoids model early human brain network development. *Cell Stem Cell*. 2019;25:558–569.e7.
39. Velasco S, Kedaigle AJ, Simmons SK, Nash A, Rocha M, Quadrato G, et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature*. 2019;570:523–7.
40. Polioudakis D, de la Torre-Ubieta L, Langerman J, Elkins AG, Shi X, Stein JL, et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron*. 2019;103:785–801.e8.
41. Park YP, He L, Davila-Velderrain J, Hou L, Mohammadi S, Mathys H, et al. Single-cell deconvolution of 3000 post-mortem brain samples for eQTL and GWAS dissection in mental disorders. *BioRxiv* 2021. <https://doi.org/10.1101/2021.01.21.426000>.
42. Schafer ST, Paquola ACM, Stern S, Gosselin D, Ku M, Pena M, et al. Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat Neurosci*. 2019;22:243–55.
43. Costamagna G, Comi GP, Corti S. Advancing drug discovery for neurological disorders using iPSC-derived neural organoids. *Int J Mol Sci*. 2021;22. <https://doi.org/10.3390/ijms22052659>.
44. Yoon S-J, Elahi LS, Paşca AM, Marton RM, Gordon A, Revah O, et al. Reliability of human cortical organoid generation. *Nat Methods*. 2019;16:75–8.
45. Kohwi M, Petryniak MA, Long JE, Ekker M, Obata K, Yanagawa Y, et al. A sub-population of olfactory bulb GABAergic interneurons is derived from Emx1- and Dlx5/6-expressing progenitors. *J Neurosci*. 2007;27:6878–91.
46. Schiweck J, Eickholt BJ, Murk K. Important shapeshifter: mechanisms allowing astrocytes to respond to the changing nervous system during development, injury and disease. *Front Cell Neurosci*. 2018;12:261.
47. Krecenk R, Ullian EM. A cellular star atlas: using astrocytes from human pluripotent stem cells for disease studies. *Front Cell Neurosci*. 2013;7:25.
48. Petrelli F, Pucci L, Bezzi P. Astrocytes and microglia and their potential link with autism spectrum disorders. *Front Cell Neurosci*. 2016;10:21.
49. Notter T. Astrocytes in schizophrenia. *Brain Neurosci Adv*. 2021;5:23982128211009148.
50. Maragakis NJ, Rothstein JD. Mechanisms of disease: astrocytes in neurodegenerative disease. *Nat Clin Pr Neurol*. 2006;2:679–89.
51. Hu B-Y, Weick JP, Yu J, Ma L-X, Zhang X-Q, Thomson JA, et al. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci USA*. 2010;107:4335–40.
52. Koehler KR, Tropel P, Theile JW, Kondo T, Cummins TR, Viville S, et al. Extended passaging increases the efficiency of neural differentiation from induced pluripotent stem cells. *BMC Neurosci*. 2011;12:82.
53. Lalli MA, Avey D, Dougherty JD, Milbrandt J, Mitra RD. High-throughput single-cell functional elucidation of neurodevelopmental disease-associated genes reveals convergent mechanisms altering neuronal differentiation. *Genome Res*. 2020;30:1317–31.
54. Fietz SA, Lachmann R, Brandl H, Kircher M, Samusik N, Schröder R, 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*. 2012;109:11836–41.
55. Long KR, Newland B, Florio M, Kalebic N, Langen B, Kolterer A, et al. Extracellular matrix components HAPLN1, lumican, and collagen I cause hyaluronic acid-dependent folding of the developing human neocortex. *Neuron*. 2018;99:702–719.e6.
56. Franco SJ, Martinez-Garay I, Gil-Sanz C, Harkins-Perry SR, Müller U. Reelin regulates cadherin function via Dab1/Rap1 to control neuronal migration and lamination in the neocortex. *Neuron*. 2011;69:482–97.
57. Wiese S, Karus M, Faissner A. Astrocytes as a source for extracellular matrix molecules and cytokines. *Front Pharm*. 2012;3:120.
58. Florio M, Albert M, Taverna E, Namba T, Brandl H, Lewitus E, et al. Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science*. 2015;347:1465–70.
59. Sugiyama S, Di Nardo AA, Aizawa S, Matsuo I, Volovitch M, Prochiantz A, et al. Experience-dependent transfer of Otx2 homeoprotein into the visual cortex activates postnatal plasticity. *Cell*. 2008;134:508–20.
60. Gogolla N, Caroni P, Lüthi A, Herry C. Perineuronal nets protect fear memories from erasure. *Science*. 2009;325:1258–61.
61. Gisabella B, Babu J, Valeri J, Rexrode L, Pantazopoulos H. Sleep and memory consolidation dysfunction in psychiatric disorders: evidence for the involvement of extracellular matrix molecules. *Front Neurosci*. 2021;15:646678.
62. Koskinen M-K, van Mourik Y, Smit AB, Riga D, Spijker S. From stress to depression: development of extracellular matrix-dependent cognitive impairment following social stress. *Sci Rep*. 2020;10:17308.
63. Germain P-L, Testa G. Taming human genetic variability: transcriptomic meta-analysis guides the experimental design and interpretation of iPSC-based disease modeling. *Stem Cell Rep*. 2017;8:1784–96.
64. Paşca SP, Arlotta P, Bateup HS, Camp JG, Cappello S, Gage FH, et al. A nomenclature consensus for nervous system organoids and assembloids. *Nature*. 2022;609:907–10.
65. Yoshioka N, Gros E, Li H-R, Kumar S, Deacon DC, Maron C, et al. Efficient generation of human iPSCs by a synthetic self-replicative RNA. *Cell Stem Cell*. 2013;13:246–54.
66. Desmarais JA, Unger C, Damjanov I, Meuth M, Andrews P. Apoptosis and failure of checkpoint kinase 1 activation in human induced pluripotent stem cells under replication stress. *Stem Cell Res Ther*. 2016;7:17.
67. Bressan RB, Southgate B, Ferguson KM, Blin C, Grant V, Alfazema N, et al. Regional identity of human neural stem cells determines oncogenic responses to histone H3.3 mutants. *Cell Stem Cell*. 2021. <https://doi.org/10.1016/j.stem.2021.01.016>.
68. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;4:417–9.
69. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006;22:1600–7.
70. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma*. 2008;9:559.
71. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*. 2008;24:719–20.
72. 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*. 2015;127:67–72.
73. Supek F, Božić njak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. 2011;6:e21800.
74. Dong M, Thennavan A, Urrutia E, Li Y, Perou CM, Zou F, et al. SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief Bioinformatics* 2020. <https://doi.org/10.1093/bib/bbz166>.

ACKNOWLEDGEMENTS

The authors wish to thank the Wellcome Trust Sanger Institute, its funders and clinical collaborators, and Life Science Technologies Corporation for providing CTL2 hiPSCs. The authors wish to thank the University of Sheffield for providing CTL3 hiPSCs. The authors wish to thank the GGDBbank, member of the Telethon Network of Genetic Biobanks (project no. GTB12001), funded by Telethon Italy, and of the EuroBiobank network, for providing CTL4 fibroblasts. This work received funding from: the European Research Council (ERC) (DISEASEAVATARS grant no. 616441 to GT); ENDpoiNTs, European Union's Horizon 2020 research and innovation program (grant no. 825759 to GT); Fondazione Cariplo (grant no. 2017-0886 to ALT and GT); Fondazione Telethon (grant no. GEP13105 to GT; grant no. GGP19295 to ALT). ST performed this work while being PhD student within the European School of Molecular Medicine (SEMM). WTG is supported by the British Columbia Children's Hospital Research Institute by an intramural IGAP research salary award. RBB was supported by a PhD fellowship from the Science Without Borders Program (CAPES, Governo Dilma Rousseff, Brazil). The authors express their gratitude to the research participants and their family members who donated biopsies.

AUTHOR CONTRIBUTIONS

CC designed the analytical strategy of transcriptomic data. ST designed and performed CBO differentiation. CC and ST performed bioinformatic analyses. NC cultured human fetal primary stem cells and contributed to bioinformatic analyses. ET, FT, and NC helped with cortical organoids maintenance and library preparation. ST, SS, and ALT performed immunofluorescence staining of organoids and image acquisition. ALT curated literature-derived gene signatures. WTG provided skin fibroblasts of CTL1, and provided editorial comments on the manuscript. MG reprogrammed hiPSCs from CTL1's fibroblasts. RBB and SP provided human fetal primary samples. CC, ST, ALT, and NC contributed to the study design, discussion of the results, and critical reading of the manuscript. CC, ST, and GT wrote the manuscript. GT conceived, designed, and supervised the study.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41398-022-02279-0>.

Correspondence and requests for materials should be addressed to Giuseppe Testa.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2022