

ARTICLE OPEN



YY1 mutations disrupt corticogenesis through a cell type specific rewiring of cell-autonomous and non-cell-autonomous transcriptional programs

Marlene F. Pereira^{1,2,3,11}, Veronica Finazzi^{3,11}, Ludovico Rizzuti^{1,2,3}, Davide Aprile³, Vittorio Aiello^{2,3}, Luca Mollica⁴, Matteo Riva⁴, Chiara Soriani¹, Francesco Dossena³, Reinald Shyti^{1,3}, Davide Castaldi^{2,3}, Erika Tenderini¹, Maria Teresa Carminho-Rodrigues⁵, Julien F. Bally⁶, Bert B. A. de Vries⁷, Michele Gabriele^{1,2,8,9,10,12}, Alessandro Vitriolo^{1,2,3,12} and Giuseppe Testa^{1,2,3,12}✉

© The Author(s) 2025

Germline mutations of YY1 cause Gabriele-de Vries syndrome (GADEVS), a neurodevelopmental disorder featuring intellectual disability and a wide range of systemic manifestations. To dissect the cellular and molecular mechanisms underlying GADEVS, we combined large-scale imaging, single-cell multiomics and gene regulatory network reconstruction in 2D and 3D patient-derived physiopathologically relevant cell lineages. YY1 haploinsufficiency causes a pervasive alteration of cell type specific transcriptional networks, disrupting corticogenesis at the level of neural progenitors and terminally differentiated neurons, including cytoarchitectural defects reminiscent of GADEVS clinical features. Transcriptional alterations in neurons propagated to neighboring astrocytes through a major non-cell autonomous pro-inflammatory effect that grounds the rationale for modulatory interventions. Together, neurodevelopmental trajectories, synaptic formation and neuronal-astrocyte cross talk emerged as salient domains of YY1 dosage-dependent vulnerability. Mechanistically, cell type resolved reconstruction of gene regulatory networks uncovered the regulatory interplay between YY1, NEUROG2 and ETV5 and its aberrant rewiring in GADEVS. Our findings underscore the reach of advanced in vitro models in capturing developmental antecedents of clinical features and exposing their underlying mechanisms to guide the search for targeted interventions.

Molecular Psychiatry (2025) 30:3413–3429; <https://doi.org/10.1038/s41380-025-02929-x>

INTRODUCTION

YY1 is a ubiquitous zinc-finger transcription factor (TF) whose haploinsufficiency causes Gabriele-de Vries syndrome (GADEVS, MIM: 600013), a predominantly neurodevelopmental disorder featuring mostly intellectual disability, gliosis, hypo- and dystonia with other additional congenital abnormalities of the eye, heart, kidney, genital and skeletal system in a subset of patients [1–7]. The role of YY1 in neurodevelopment was first shown in a *Yy1*^{+/-} mouse model [8], in which a significant fraction of embryos showed exencephaly, pseudo-ventricles, and asymmetry of the developing brain. YY1 exerts its function by binding to regulatory sequences and recruiting chromatin regulators such as the INO80 chromatin remodeling complex, the p300/CBP histone acetylase, polycomb group proteins, cohesin and condensin, eventually also regulating pre-mRNA splicing [9]. Moreover, we previously showed that YY1 haploinsufficiency caused a defective gene expression regulation associated with a loss of the histone post-translational modification associated with enhancer activation (H3K27ac) at YY1-bound

enhancers [1]. Shortly after, YY1 was showed to act as an enhancer-promoter (E-P) looping factor [10, 11]. For these characteristics, GADEVS can be identified as enhanceropathy [12, 13]. However, acute depletion of YY1 in mouse embryonic stem cells showed a minimal effect of E-P looping and transcription activation [14]. This recent evidence raised a fundamental question about the ability of YY1 in forming and maintaining E-P loops and gene expression. This led us to hypothesize that YY1 regulates the establishment of new E-P rather than their maintenance, in a highly tissue-specific manner. The inability to fully understand the role of YY1 and the reason why its haploinsufficiency causes GADEVS poses a limitation in improving affected individuals' care. To address these questions and to identify the pathological mechanisms underlying impairments in learning processes and brain abnormalities observed in GADEVS individuals, we derived induced pluripotent stem cells (iPSCs) from multiple individuals affected by YY1 haploinsufficiency and studied its impact in pluripotency as well as 2D and 3D neuronal models.

¹Department of Experimental Oncology, European Institute of Oncology IRCCS, Via Adamello 16, 20139 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. ⁴Department of Medical Biotechnology and Translational Medicine, University of Milan, Milan, Italy. ⁵Department of Genetics, University of Geneva & University Hospitals of Geneva, Geneva, Switzerland. ⁶Service of Neurology, Department of Clinical Neurosciences, Lausanne University Hospital & University of Lausanne, Lausanne, Switzerland. ⁷Radboud University UMC, Nijmegen, The Netherlands. ⁸Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ⁹The Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA. ¹⁰Koch Institute for Integrative Cancer Research, Cambridge, MA 02139, USA. ¹¹These authors contributed equally: Marlene F. Pereira, Veronica Finazzi. ¹²These authors jointly supervised this work: Michele Gabriele, Alessandro Vitriolo, Giuseppe Testa. ✉email: giuseppe.testa@fht.org

Received: 31 March 2024 Revised: 30 December 2024 Accepted: 10 February 2025
Published online: 22 February 2025

RESULTS

YY1 pathogenic mutations disrupt DNA binding and cause widespread transcriptional downregulation in iPSCs

GADEVs is caused by haploinsufficiency of *YY1* and results from truncating mutations or missense variants in the zinc-finger motifs [1]. To study the molecular dysregulations caused by *YY1* mutations in disease-relevant cell types, we focused on two truncating and one missense mutation in the Zinc-finger domain (GAD01: p.(Lys179*); GAD02: p.(Asp129Ilefs*127) [1], and GAD03: p.(Cys303Arg) [15] (Fig. 1a). We derived iPSCs from three affected individuals' along with healthy donors' lymphoblastoid cell lines or skin fibroblasts (Fig. 1b). All iPSC lines were extensively characterized by immunofluorescence and RNA-seq for pluripotency markers (Fig. S1a–c), and qRT-PCR confirmed that iPSCs were vector-free after approximately 10–12 passages in culture (Fig. S1d). Moreover, all lines had the capacity to differentiate into derivatives of all three embryonic germ layers (Fig. S2), even though different reprogramming methods were used (Fig. S1e). Next, we assessed *YY1* protein levels in iPSCs. As expected from truncating mutations, we were able to detect halved levels of *YY1* in GAD01 and GAD02 ($47.62 \pm 12.18\%$; $61.60 \pm 7.79\%$ vs. CTL, respectively). In GAD03 ($285.20 \pm 72.06\%$ vs. CTL), we observed instead increased expression of *YY1* transcript (Fig. S1g) and protein levels (Fig. 1c). Given the dosage imbalances in *YY1* expression in different mutants we took advantage of molecular dynamics simulations of pathogenic missense mutations of *YY1*, including the one harbored by GAD03, to evaluate the structural features of the DNA-binding subunits of *YY1* in mutants with respect to the healthy form, finding a reduction of protein stability and DNA-binding ability across all mutants (Fig. 1d). Despite increased levels of *YY1* in GAD03, all mutants seem to have a reduced affinity for DNA binding. As a fundamental regulator directly controlling gene expression programs, to functionally validate *YY1* DNA binding across control and mutant lines we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq). We profiled *YY1* binding genomic distribution across our cohort and found it enriched at promoters and distal intergenic regions, consistently with *YY1* known transcriptional regulation function (Fig. S3a). Moreover, our ChIP-seq showed coherent peaks distributions with respect to reference ENCODE tracks from H1 human embryonic stem cells (hESC), while *YY1* motifs seemed equally represented across genomic regions, with a higher abundance at introns and distal intergenic regions (Fig. S3b). Finally, the human *YY1* motif (MA0095.2 in JASPAR 2020) was by far the most enriched in iPSC *YY1* ChIP-seq (Fig. S3c), approximately occupying 28% of *YY1* peaks in CTLs ($P < 1 \times 10^{-3311}$, 2% of background), 49% of *YY1* peaks in GADEVs ($P < 1 \times 10^{-2293}$, 3% of background), and 39% of lost peaks, indicating that lost peaks were effectively bound by *YY1*, but also that *YY1* haploinsufficiency depletes co-factors bindings while retaining high affinity sites. Despite the increased protein levels in GAD03, *YY1* ChIP-seq showed a drop of binding signals in the three affected lines (Fig. 1f, S3d) and a general decrease in binding sites across replicates (Fig. 1e). As *YY1* was described to work as dimer [10, 16], this observation corroborates the prediction that in the mutant setting only $\frac{1}{4}$ of the dimers would be functional and hence with an equivalent impact of truncating and missense mutations such as p.(Cys303Arg) that emerged from our assays as a dominant negative (Fig. S1f). Indeed, $\frac{1}{4}$ of the average level observed in GAD03 is similar to those observed in GAD01 and GAD02, where only half of the protein level was observed. Coherently, among differentially expressed genes (DEGs) we observed a widespread downregulation (585 out of 764 with abs (FC) > 1.5 and FDR < 0.05) in the mutant iPSC lines (Fig. 1g).

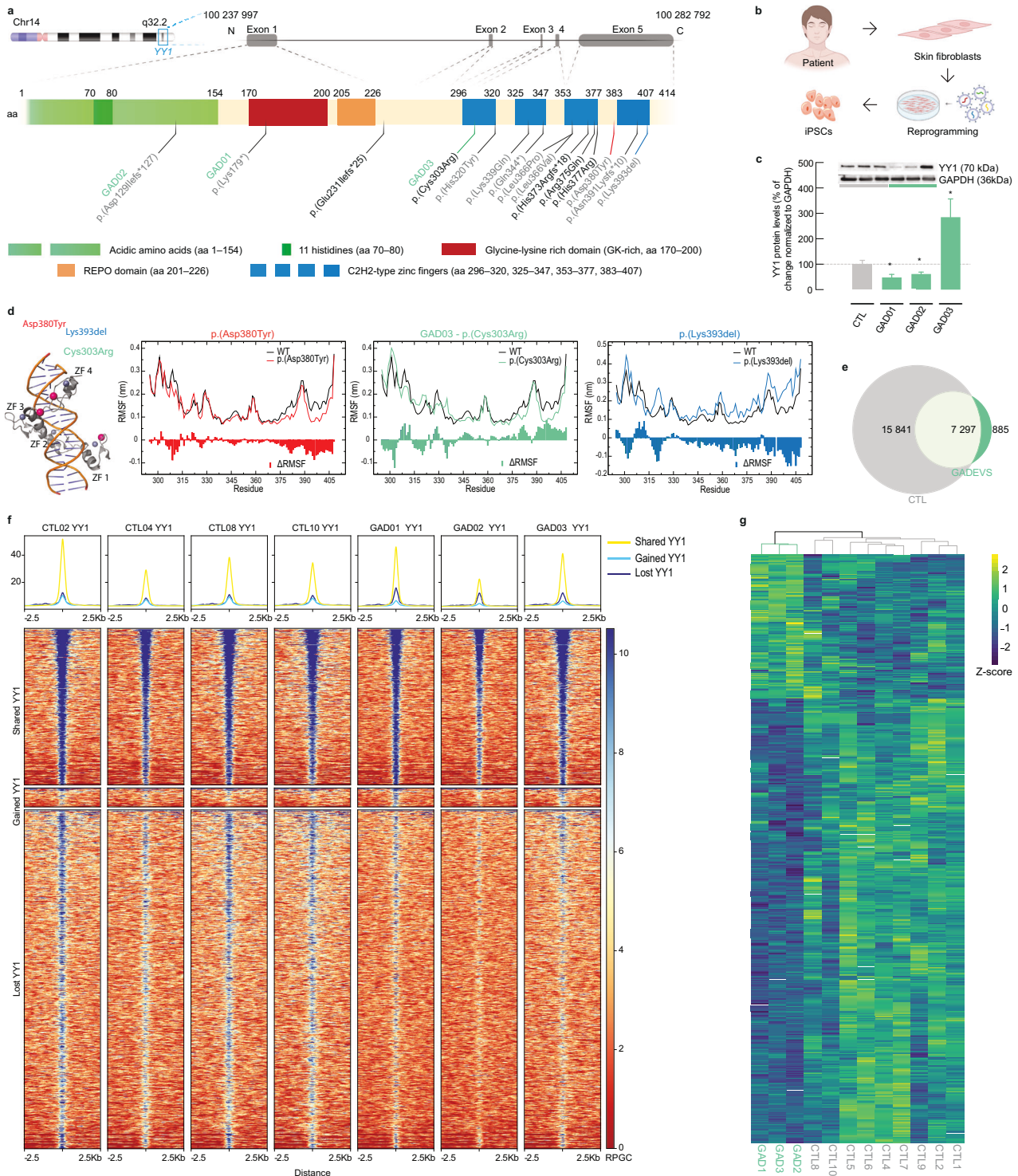
Downregulated *YY1* targets enriched for pathologically clinically relevant phenotypes

Given the prevalence of downregulated genes among DEGs and the enhancer-associated role of *YY1* as well, to understand the

impact of the widespread dysregulation observed in affected individuals' iPSC at the transcriptomic level, we characterized them via enrichment analyses and integration with *YY1* ChIP-seq (Fig. 2a) we previously discussed. Gene Ontology (GO) enrichments showed that biological processes relevant for neurological and organogenesis phenotypes associated with GADEVs symptoms were disrupted already at the pluripotent stage, as analogously observed in the case of other neurodevelopmental disorders [17, 18] (Fig. 2b). Then, we focused on genes directly regulated by *YY1* and observed that roughly 25% of downregulated genes were effectively bound by *YY1* (hypergeometric test, $P = 2.86e-179$) at regulatory regions (Fig. 2c). In fact, if we compare the genes bound by *YY1* (~5 thousand) and the DEGs (450 with FC > 2 and FDR < 0.05, out of ~19K genes expressed), the DEGs constitute an ~8.5 fold enrichment with respect to the percentage of genes differentially expressed. Notably, 60% of DEGs bound by *YY1* are downregulated and those bound both at enhancer and promoters by *YY1* are crucial markers of corticogenesis and, most importantly, genes implicated in a wide range of developmental or neurologic aberrations (Fig. 2d). These include genes such as *SOBP* (associated with mental retardation, maxillary protrusion and strabismus (MRAMS) syndrome [19]), *MLXIPL* (linked with Williams-Beuren syndrome and non-alcoholic fatty liver disease [20]), *KCNN3* (associated with Zimmermann-Laband syndrome [21]), *SEZ6* (occurrence of febrile seizures and dystonia [22, 23]), *MAP2* (associated with central neurocytoma within ventricles [24, 25] and neurodegenerative olivopontocerebellar atrophy [26]), *SLC1A1* (implicated in Schizophrenia 18 [27]), *OTX2* (syndromic microphthalmia [28–30]), *EEF1A2* (with mutations leading to developmental and epileptic encephalopathy 33, and autosomal dominant intellectual developmental disorder [31, 32]), among many others (Fig. 2d). Indeed, Human Phenotype Ontology (HPO) over-representation analysis of *YY1* bound downregulated genes corroborates the observations provided by GO enrichments by highlighting significant enrichments for pathologically relevant phenotypes reported in multiple affected individuals (Table S1). Among the significant HPOs, we found delayed speech and language development, intellectual disability, seizures, depressed nasal bridge, sensorineural hearing impairment, short stature, visual impairment, and cryptorchidism [5, 33] (Fig. 2e). The molecular and cellular origin of GADEVs clinical features remains unexplored. Given the urgency of identifying the molecular alterations caused by *YY1* haploinsufficiency to identify druggable pathways and improve affected individuals' care, we differentiated iPSCs derived from affected individuals towards clinically relevant cell types. The presence of anomalies in the central and peripheral nervous systems, including intellectual disability, craniofacial and skeletal dysmorphisms, as well as cardiovascular defects, have often been described in GADEVs individuals [5, 33]. Thus, exploring neural crest-derived tissues in *YY1*-mutated samples remains one of the best in vitro models to capture craniofacial dysmorphisms and defects in peripheral nervous systems precursors. Incidentally, while control lines successfully differentiated into neural crest stem cells (NCSC) (Fig. S3e, f), differentiation of iPSCs into NCSCs caused apoptotic blebbing events, similar to what has been already described for the complete loss of *Yy1* in mouse models [34], corroborating our observation in iPSC that heterozygous pathogenic mutations of *YY1* induce a loss of function-like phenotype.

YY1 mutations impair the formation of ventricle-like structures at early stages of cortical brain organoid generation

To model cortical development defects in GADEVs individuals, we resorted to organoids, a complex in vitro model representing spatiotemporal neuronal development and capturing the cellular



diversity required to understand neurological alterations. To better characterize the pathological relevance of YY1 mutations in early neuronal differentiation and capture the molecular dysfunction that could explain the mild to severe ID and other neuronal phenotypes such as ventricle atrophy and frontal gliosis observed in GADEVS individuals, iPSCs were differentiated into cortical organoids [35, 36] and cultured for short- (days 3, 7, 12, and 30) and long-term (day 90) assessment (Fig. 3a, S3a). First, we sought to assess general morphology properties at early stages of differentiation (Fig. S4a) and observed a significant reduction in

the cross-section area of spheroid bodies at day 3 (GADEVS $n = 4766$, $26,438 \mu\text{m}^2$, $427,715$ vs. CTL $n = 8197$, $31,493 \mu\text{m}^2$, $403,003$) (Fig. S4b–d), mirrored by similar reductions in their perimeter as well (GADEVS $n = 4766$, $584.1 \mu\text{m}$, 2028 vs. CTL $n = 8197$, $642.9 \mu\text{m}$, 1971) (Fig. S4e). These surface imbalances were compensated later, and no further area differences were observed until day 12 (Fig. S4f). However, cross-sectional perimeter was increased at day 7 (GADEVS $n = 2722$, $984.3 \mu\text{m}$, 2723 vs. CTL $n = 4578$, $952.3 \mu\text{m}$, 1868) and 12 (GADEVS $n = 3563$, $1405 \mu\text{m}$, 5137 vs. CTL $n = 8093$, $1352 \mu\text{m}$, 3881), suggesting that

Fig. 1 YY1 dosage imbalances are equally detrimental for DNA binding and transcriptional regulation at the pluripotency stage. **a** Map of YY1 gene structure and mutation distribution. From the top, YY1 location on chromosome 14 is reported above exon composition and genomic coordinates. Dashed lines report the contribution of each exon to the coded protein. A scheme of the protein is reported with amino acid residues numbering of each functional domain. Previously described mutations are reported in light grey, newly described mutations are reported in black. Samples used in this study are named as GAD01-03. On the bottom, a legend describing protein domains is reported. **b** Illustration shows that samples from healthy and affected individuals (GAD01-03) were reprogrammed into induced pluripotent stem cells (iPSCs). **c** YY1 protein levels (normalized to GAPDH) revealed opposite levels of YY1 in GAD03 compared to the remaining GADEVs cohort samples. * $P < 0.05$, $n = 6$, unpaired Student's t -test vs. CTL. Values are mean \pm SEM of n independent experiments. n refers to the number of protein extracts processed per line. **d** Molecular dynamics of YY1-DNA complex represented in terms of Root Mean Square Fluctuation (RMSF), reported on the y-axis (nm) per residue (x-axis). Flexibility changes are reported around zinc-binding sites. **e** Venn diagram reporting the intersection of YY1 peaks in CTL and GADEVs lines. Light grey reports CTL specific peaks (lost), light green refers to shared peaks, dark grey indicates GADEVs-specific (gained) peaks. **f** Heatmap of library-size normalized YY1 ChIP-seq coverage (RPGC) at YY1 binding sites. Each row represents a 5-kb window centered on peak summits. Reads distribution at Shared, Gained, and Lost peaks are respectively depicted in yellow, blue and cyan. **g** Heatmap showing z-scores of log(TMM) read counts for DEGs (764) derived from comparing YY1-mutant vs. control iPSCs (FDR ≤ 0.05 and FC ≥ 1.5). Yellow and dark blue colors respectively indicate expression levels of up- (179) and downregulated (585) genes.

YY1-mutated organoids acquired a more irregular shape along differentiation (Fig. S4g). To substantiate this hypothesis, we assessed the circularity of control and mutated organoids and indeed, they show a coherent reduction at days 7 (GADEVs $n = 2722$, 0.97, 0.29 vs. CTL $n = 4578$, 0.99 μm , 0.27) and 12 (GADEVs $n = 3563$, 0.93, 0.45 vs. CTL $n = 8093$, 0.99, 0.49) (Fig. S4h). Moreover, we calculated the number of organoids in both experimental groups and observed that YY1-mutated lines originated less organoids (Fig. S4i). Regardless of these perimeter and circularity differences, both control and affected cortical organoids showed no clear imbalances in the cellular composition of the progenitor pools (Fig. 3b) nor in actively proliferating cells (phospho-histone H3, pHH3-positive cells) within the organoids (Fig. S4j). However, following quantification of PAX6-positive ventricle-like structures (VLS), YY1-mutated organoids failed to recapitulate the complexity and typical cytoarchitectural organization observed in CTL samples, resulting in significant detection of 'irregular' VLS instead. (Fig. 3c).

YY1 haploinsufficiency affects intermediate cellular states in early organoid development and impacts differentiated lineages in later stages

To further characterize the cellular complexity and composition of YY1-mutated organoids at day 30 and day 90 regarding gene expression and DNA regulatory regions, we resorted to single cell multiomic profiling. Indeed, single-cell multiome captures the transcriptome and chromatin accessibility within the same cell nucleus, allowing the reconstruction of gene regulatory networks (GRN) along differentiation trajectories [37–39]. We profiled day 30 cortical organoids from 7 control and 3 GADEVs lines, and retained 31,214 high quality cells, globally expressing 27,701 genes. At day 90 cortical, retaining 32,677 high quality cells, globally expressing 33,876 genes (Fig. S5a, b). Given the presence of multiple cell types with different abundances, in both datasets, cells were kept if they expressed between 500 genes and 10,000 genes, retaining at least 5000 reads per cell.

To assign cell type identities to our datasets, we combined Leiden clustering with the identification of key known markers and differentiation trajectories (Fig. 3d, f, S5a, b see Methods). To further validate the cell type annotation, we ingested the two datasets onto a reference single-cell fetal cortex dataset [40] (Fig. S5c, d). The two separate ingestions of day 30 and day 90 organoids confirmed that the former are mostly composed of early lineages, while the latter contains a higher portion of neurons and differentiated lines. Then, to further corroborate our cell type annotation, we intersected the top expressed genes in our clusters with the top expressed in each cell type identified in the reference fetal data (Fig. S5e, f), supporting our original annotation. Already at day 30 we identified different types of progenitors, including SOX2⁺ apical radial glia (aRG)

like-cells, SLC1A3⁺ radial glia (RG), and EOMES⁺ intermediate progenitors (IP) (Fig. 3e). Moreover, we observed two main differentiation paths, mirroring direct (IP-independent) and indirect (IP-dependent) neurogenesis (Fig. S5h). We respectively identified (i) a cluster of maturing neurons (DCX⁺) derived from aRG cluster, and (ii) a cluster of TBR1⁺ cells derived from the IP, on two separated paths (Fig. 3d, Fig. S5c). Moreover, we observed the expression of DLX genes in IP at day 30, yet previous reports have already shown the expression of DLX5 in late cortical IP [41]. Despite detecting irregular VLS at day 12, we did not observe significant differences in cell type abundances in the RNA modality of our single cell multiomic datasets. However, we observed a severe transcriptional dysregulation in RG, IP and maturing excitatory neurons (Fig. 3e), which might indicate an impairment in the generation of mature neurons later in development. As expected, at day 90 we identified more cell types representative of a more mature stage of differentiation, comprising HOPX⁺ outer radial glia (oRG)-like cells, SATB2⁺ upper layer-like neurons, and a minor representation of AQP4⁺ cells, unmasking the presence of a rare population of astrocytes, and at the beginning of gliogenesis (Fig. S5d). Of note, at this later stage of development we saw a wide transcriptional dysregulation affecting the majority of identified cell types (Fig. 3g), with neurons derived from both IP and RG as the most disrupted, suggesting a cumulative/epigenetic effect.

YY1 haploinsufficiency triggers an extensive rewiring of cell type specific transcriptional programs

Next, we measured the direct effect of YY1 haploinsufficiency on the underlying GRNs in cortical organoids at day 30. To reconstruct the GRNs in the identified cell types, we applied Cell Oracle [42] because it leverages co-accessibility data from single cell ATAC and single-cell RNA-seq to build cell type specific GRNs. First, we used GRNs to further confirm the identity of IP by looking at their most central TFs (Fig. S6a). Among the first 30 TFs, we identified migration genes such as *NFIB* and *POU3F2*, and *RORB* in accordance with IP role in originating lower layer neurons. Moreover, we found *EOMES* centrality enriched in IP compared to RG (Fig. S6b, see methods), which showed basal expression of *EOMES* and constitute the closest cell type from a developmental perspective, hence reassuring us on the cell identity annotation. Once we validated IP identity, we summed raw counts into pseudobulks by grouping cells by individuals by cell type, to perform differential expression analysis (Fig. 3e, g, Table S1) by comparing GADEVs lines with controls in a cell type-specific manner (as advised in [43]). Then, to pinpoint the direct activity of YY1, we generated cell type-specific GRN with CellOracle and extracted the DEGs directly regulated by YY1 in each cell type. Direct targets of YY1 significantly enriched DEGs in RG and IP

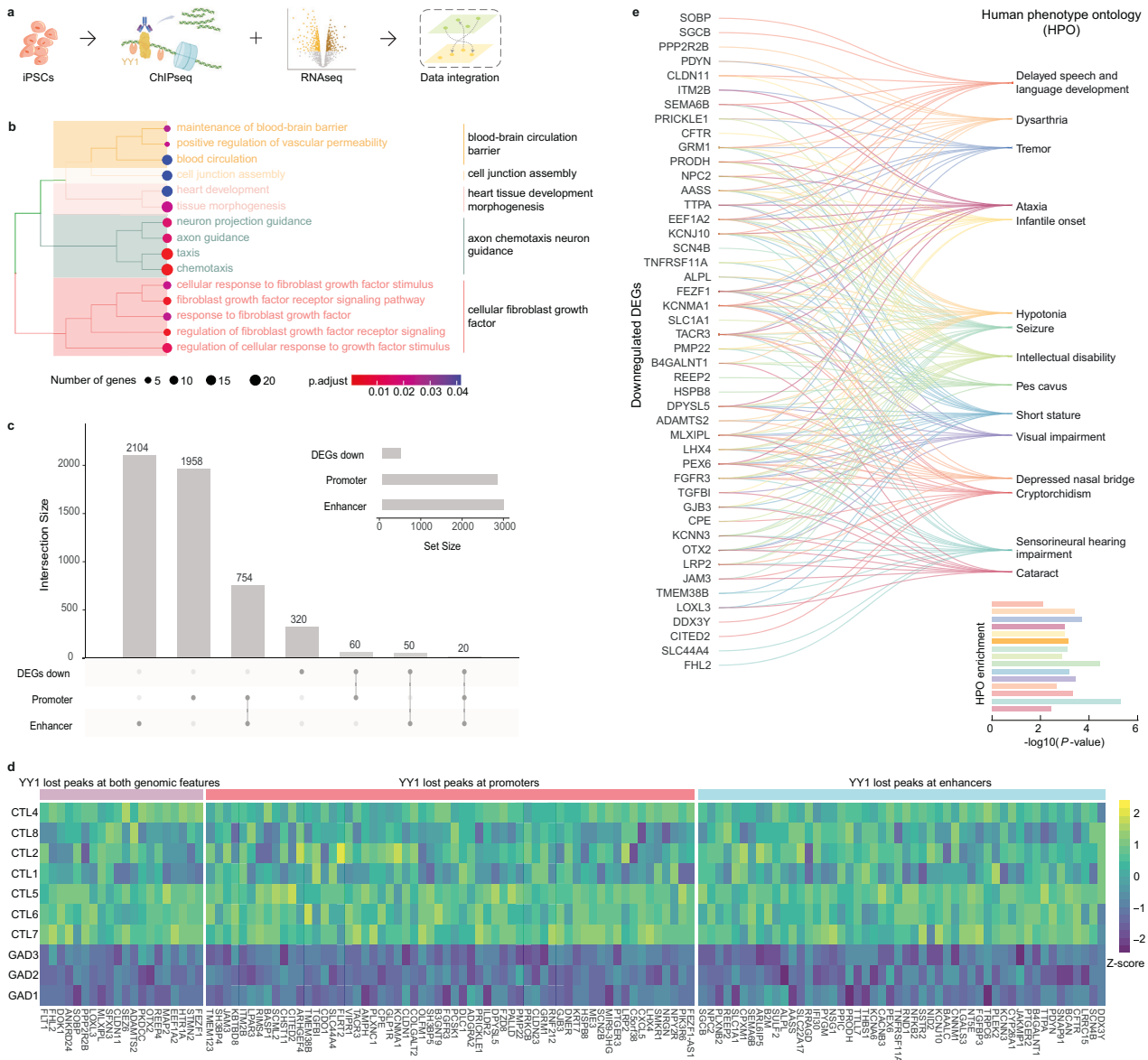


Fig. 2 YY1-controlled downregulated genes in iPSC point to pathologically relevant GO and HPO categories. **a** Diagram of the data integration framework applied in the figure. ChIP-seq and RNA-seq from iPSC have been combined to identify the role of differentially expressed genes directly targeted by YY1. **b** Hierarchical clustering representation of GO (biological processes) enrichments. Hierarchical clustering by Jaccard index similarity index is reported on the right. Number of genes defines node's size and node's color and it is defined by p -adjusted of each category enrichment. **c** UpSet plot reporting the intersection of genes whose promoter or enhancers (or both) are bound by YY1 in control lines, intersected with downregulated DEGs. On the top right, the set size (number of genes) is reported for DEGs, genes bound by YY1 at the promoter, and genes bound by YY1 at the enhancer. Genes found in more than one list are indicated on the bottom side as linked dots. **d** Heatmap reporting z-scores of logTMM read counts for downregulated DEGs, annotated following the intersection reported on panel b. On the y-axis, each row represents the expression of each sample; on the x-axis each DEG is annotated whether it lost YY1 binding at both promoter and enhancer or only at promoter or enhancer. **e** Bipartite plot representing gene-HPO associations for differentially expressed genes, and HPO categories significantly enriched by them. All genes reported are downregulated YY1 direct targets enriching at least one HPO category.

($P = 3.60e-65$ for RG and $P = 3.789e-95$ for IP). Enriched categories among DEGs directly regulated by YY1 in RG referred to *cell growth* and *positive regulation of developmental process* (Fig. 4a, Table S1). In IP, direct targets of YY1 enriched *tissue morphogenesis* and related categories such as *neural tube morphogenesis* (Fig. 4b, Table S1). Hence, GO categories enriched by DEGs directly regulated by YY1 both in RG and in IP point to defects in development, positioning YY1 upstream of the dysregulations observed in both cell types, and potentially pinpointing the pseudo-ventricle phenotype observed at earliest stages to a core of genes directly controlled by YY1.

The two mature neuronal subtypes were the two most disrupted cell types at day 90. TBR1⁺ neurons showed disruption in several “synaptic signalling” categories, as well as “long-term synaptic potentiation”, which are directly connected with neuronal activity and *cholesterol transport*, essential for neuronal metabolism [44, 45] (Table S1). Direct targets of YY1 identified as actively regulated in these neurons enrich for DNA binding and categories linked to transcription regulation (Table S1). When we focused on differentially expressed YY1 direct targets, we found enrichments for molecular functions related to potassium, calcium and glutamate channel activity, ultimately leading to putative direct

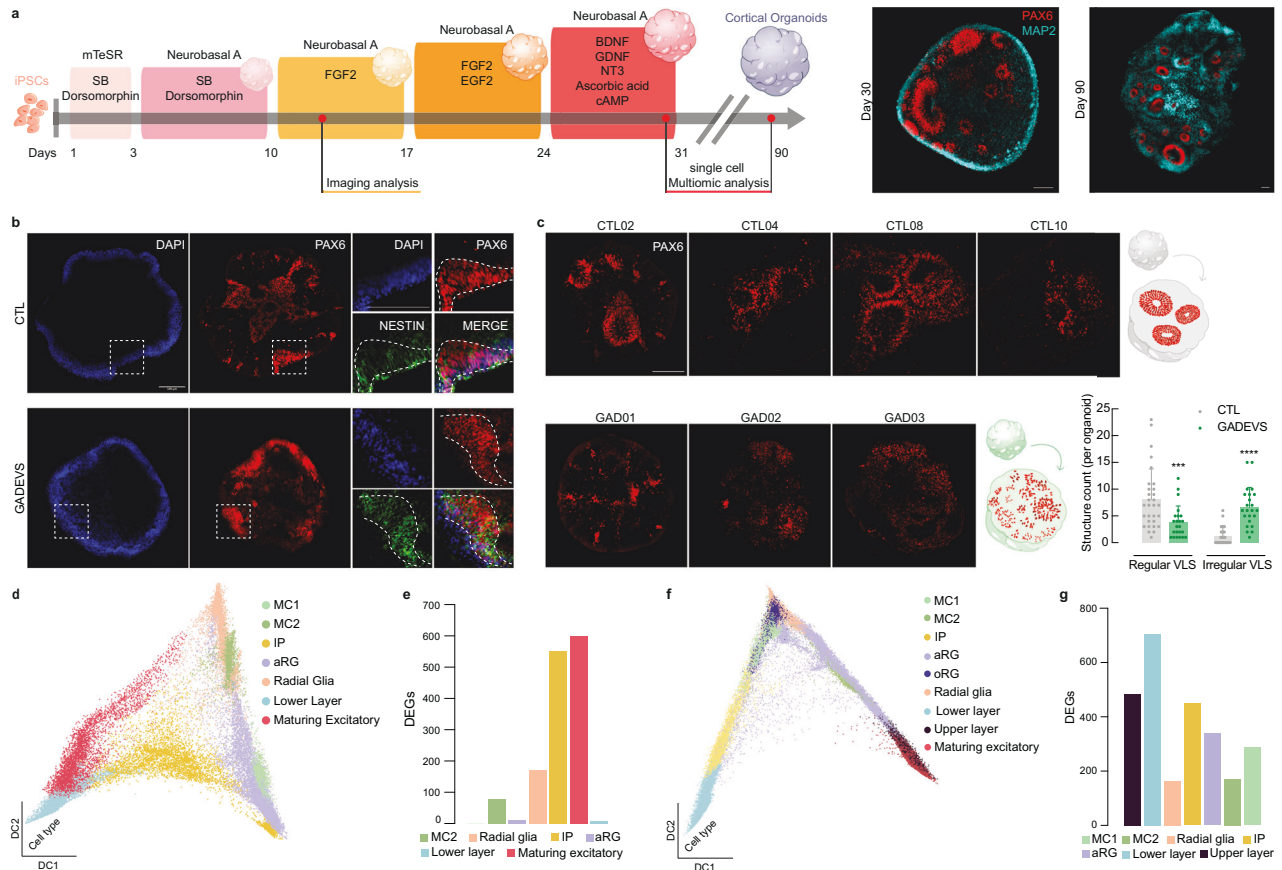


Fig. 3 GADEVS cortical brain organoids show impairments in ventricle-like structure (VLS) formation and stage-specific transcriptional disruptions. **a** Illustration of organoids generation protocol, representing telencephalon small molecule patterning and stage-specific characterization performed. Representative images of immunofluorescence (IF) performed using PAX6, and MAP2 at days 30 and 90, respectively. **b** Representative pictures of IF and cleared whole cortical organoids with PAX6, and NESTIN for the two different genotypes (CTL and GADEVS). **c** Representative images of PAX6-positive VLS counted per organoid and classified as regular and irregular depending on the shape and presence of ventricle-like lumen. (whole-organoids acquired in CTL, $n = 32$; and GADEVS, $n = 25$). Scale bar is 100 μm for all images. **d** Diffusion map dimensionality reduction reporting cell type distributions at day 30. **e** Barplot of differentially expressed genes number per cell type at day 30. **f** Diffusion map dimensionality reduction reporting cell type distributions at day 90. **g** Barplot of differentially expressed genes number per cell type at day 90. Scale bar of 100 μm for all images.

impairments on basal synaptic transmission. The same genes enriched cellular components for “ion channel complex” and “synaptic membrane” (Fig. 4c, Table S1). SATB2⁺ neurons showed dysregulations in “synaptic signaling”, “neuron development”, and “regulation of membrane depolarization”, which hinted at functional dysregulations (Table S1). Notably, “regulation of cell development”, “regulation of synaptic plasticity” and “regulation of the nervous system process” were enriched among YY1 targets expressed in the same neurons (Table S1), hinting at a direct link between YY1 activity and the dysregulation globally found in this cell type. Similarly to what we observed in TBR1⁺ neuronal populations, crossing YY1 targets with DEGs in SATB2⁺ neurons, we found enrichments for cellular compartment categories linked to ion channels, ultimately pinpointing YY1 contribution to their dysregulations in both neuronal lineages through two different gene modules (Fig. 4d, Table S1).

Since YY1 was not differentially expressed in IP at day 30 and in neurons at day 90, but we found (i) cell signaling differentially expressed in IP at day 30, (ii) upregulation of YY1 direct targets enriching GO categories related to *gene expression regulation*, and (iii) several cell types that showed broad transcriptional dysregulation at later stages, we speculated that compensatory mechanisms, including non-cell-autonomous effects, might contribute to organoids transcriptional phenotypes. Thus, to pinpoint transcriptional dysregulations that could be informative of non-cell autonomous

mechanisms, we investigated cell type abundances and cell-cell interactions. First, we applied miloR [46] to measure changes in cell type abundances, and observed an increase in the upper-layer neurons cluster in GADEVS only at day 90 (Fig. 4e). This could be a downstream effect of the vast dysregulation we observe in RG downstream of YY1, since they are precursors of upper-layer neurons. Hence, to better understand the relationships between RG and other cell types that could be triggering this increased differentiation of upper-layer neurons, we estimated [44] the contribution to gene expression of specific subgroups of receptor-ligand interactions (Fig. S6d), which could be informative of internal cell signalling and cell-cell communication defects. To model the expression patterns of receptors and ligands, we combined tools such as LIANA [47] and tensor-cell2cell [48] (see Methods), which are able to translate into cell-cell communication events the information of which receptors and ligands are expressed in each cell and in each cell type. We found a disruption in the communication between progenitors that was explained by a specific fraction of the transcriptome, called Factor 6 (Fig. S6e).

This factor highlights a disruption in cell-cell communication centered on RG and linking aRG, oRG, and cycling cells (MC1) clusters (Fig. 4f). Specifically, we observed a strong enrichment of EGFR and MAPK in Factor 6 (Fig. S6c), and targets of both pathways enriching YY1 targets ($p < 0.05$) in MC1 and the central RG clusters. However, we did not find a significant enrichment for EGFR or MAPK

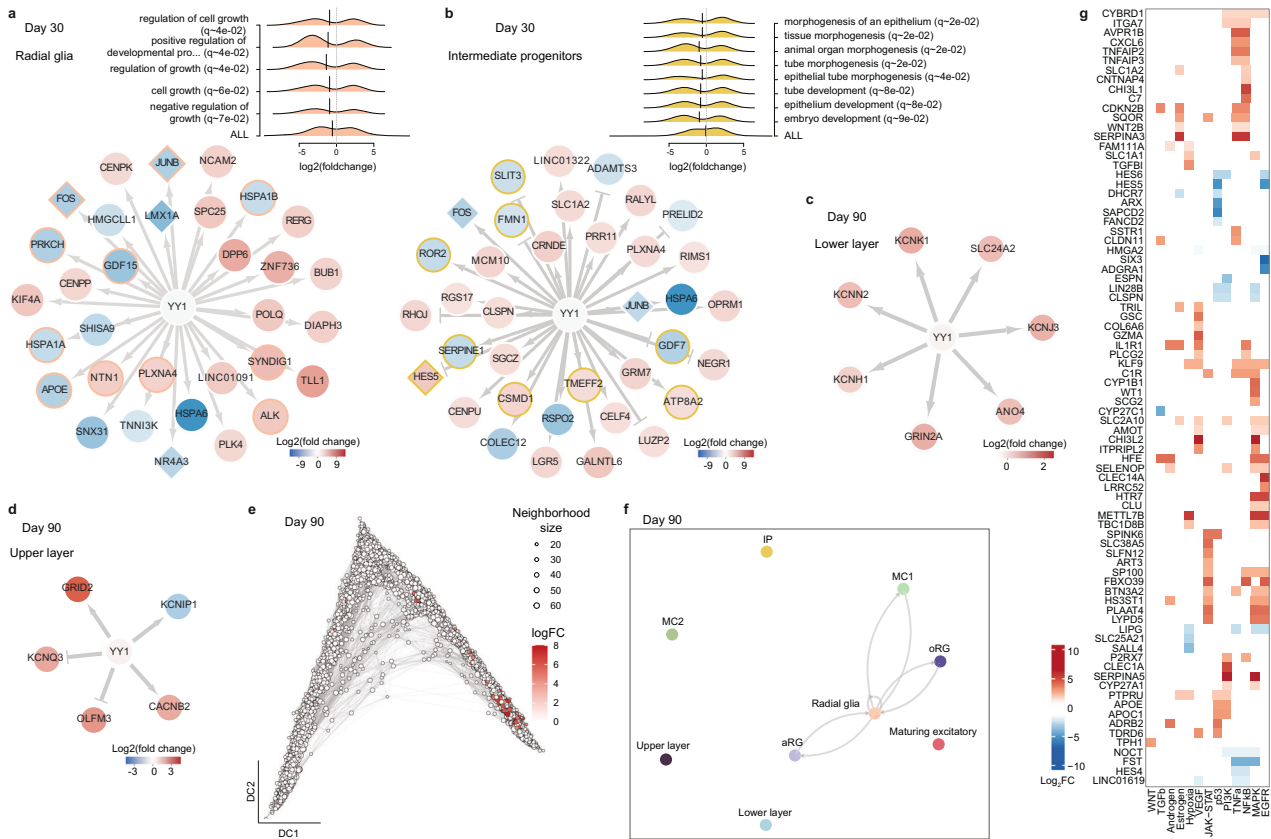


Fig. 4 Mechanistic dissection of cell type specific YY1-dependent dysregulations pinpoint YY1 regulatory activity to cell migration in IP and ion channels in mature neurons, as well as cell-cell communication defects among glia cells in late stages of GADEVs organoid corticogenesis. **a** Graph representation of the RG gene-regulatory-network comprising YY1 and its differentially expressed direct targets. Transcription factors are represented as diamond. Nodes are coloured by log₂FC. Purple borders represent association to “positive regulation of developmental process” GO category. Arrows represent activation by YY1 and T-shaped edges represent putative repressive activity by YY1. **b** Gene regulatory network representation of YY1 direct targets in IP. Arrows represent activation by YY1 and T-shaped edges represent putative repressive activity by YY1. Nodes color represent log₂(FC) measured by differential expression on IP pseudobulks, comparing controls with GADEVs lines. Green borders represent association to “morphogenesis” GO categories. **c** Gene regulatory network representation of YY1 direct targets enriching cell ion channel category in lower layer-like neurons. Edges and nodes are depicted as in panel b, using log₂(FC) from differential expression performed on lower-layer neurons. **d** Gene regulatory network representation of YY1 direct targets enriching cell ion channel category in upper layer-like neurons. Edges and nodes are depicted as in panel b, using log₂(FC) from differential expression performed on upper-layer neurons. **e** Diffusion map of day 90 organoids representing differential abundance scores measured with miloR (see methods). Neighbors of cells are represented as circles of proportional sizes with respect to the number of cells embedded in each group. Dots are colored in red proportionally with log₂FC measured as the relative abundance of GADEVs with respect to CTL in each neighbor ($p < 0.01$). **f** Differential cell-cell interactions calculated by Liana at day 90. Arrows depict interactions between cell types inferred by expression of receptor-ligand pairs. Only significantly affected interactions are depicted. **g** heatmap depicting log₂(FC) of genes involved in PROGENY cell-cell signaling pathways, derived from differential expression analysis on pseudobulk of RG at day 90.

targets among genes differentially expressed between GADEVs and CTL in any of the two cell types, which hinted at a more subtle but broad dysregulation. Thus, to make sense of which genes could explain the disrupted cell-cell interaction pathways found with LIANA and tensor cell2cell, we calculated the enrichment of genes differentially expressed between CTL and GADEVs in RG at day 90 against a curated database of signaling pathways called PROGENY, and identified the MAPK and EGFR pathways genes upregulated in GADEVs RG (Fig. 4g, see Methods). Given the non-cell autonomous type of dysregulation observed and the fact that we found cells expressing AQP4 – an astrocytes-specific water channel - within the central RG cluster (Fig. 3e), we speculated that astrocytes-mediated non-cell autonomous functions might also be impaired in GADEVs.

Abnormal neuronal organization and synaptic impairments in YY1-mutated glutamatergic neurons

Given the dysfunctions observed in late-stage organoids in lower- and upper-layer neurons, we further probed their mechanistic and functional impact in NGN2-derived glutamatergic neurons

(Fig. S7a). Given the disruption of non-cell-autonomous interactions in late-stage organoids, and the critical role of astrocytes in neuronal maturation, and response to environmental signals and synapse modulation, NGN2-derived neurons were co-cultured with fresh mouse astrocytes using transwell systems in the presence of healthy fresh mouse astrocytes (Fig. S7b). This approach enabled us to keep physically separated NGN2-derived neurons from astrocytes while preserving their interaction (Fig. 5a). First, bulk RNA-seq in glutamatergic neurons allowed us to quantify allelic abundances of YY1 (Fig. S8b) and observe that YY1 expression patterns were similar to the ones observed in iPSCs (Fig. S8a). Moreover, GAD03 increased expression was coupled to a higher detection of reads from the mutated allele (Fig. S8a), further corroborating the dominant negative model emerging from the YY1 loss of function described in iPSC (Fig. S1f).

To capture the diversity of neuronal fates generated by the comparatively homogeneous NGN2 neuronal differentiation paradigm [49] we applied single-cell approaches using gene expression and chromatin accessibility modalities, and we took

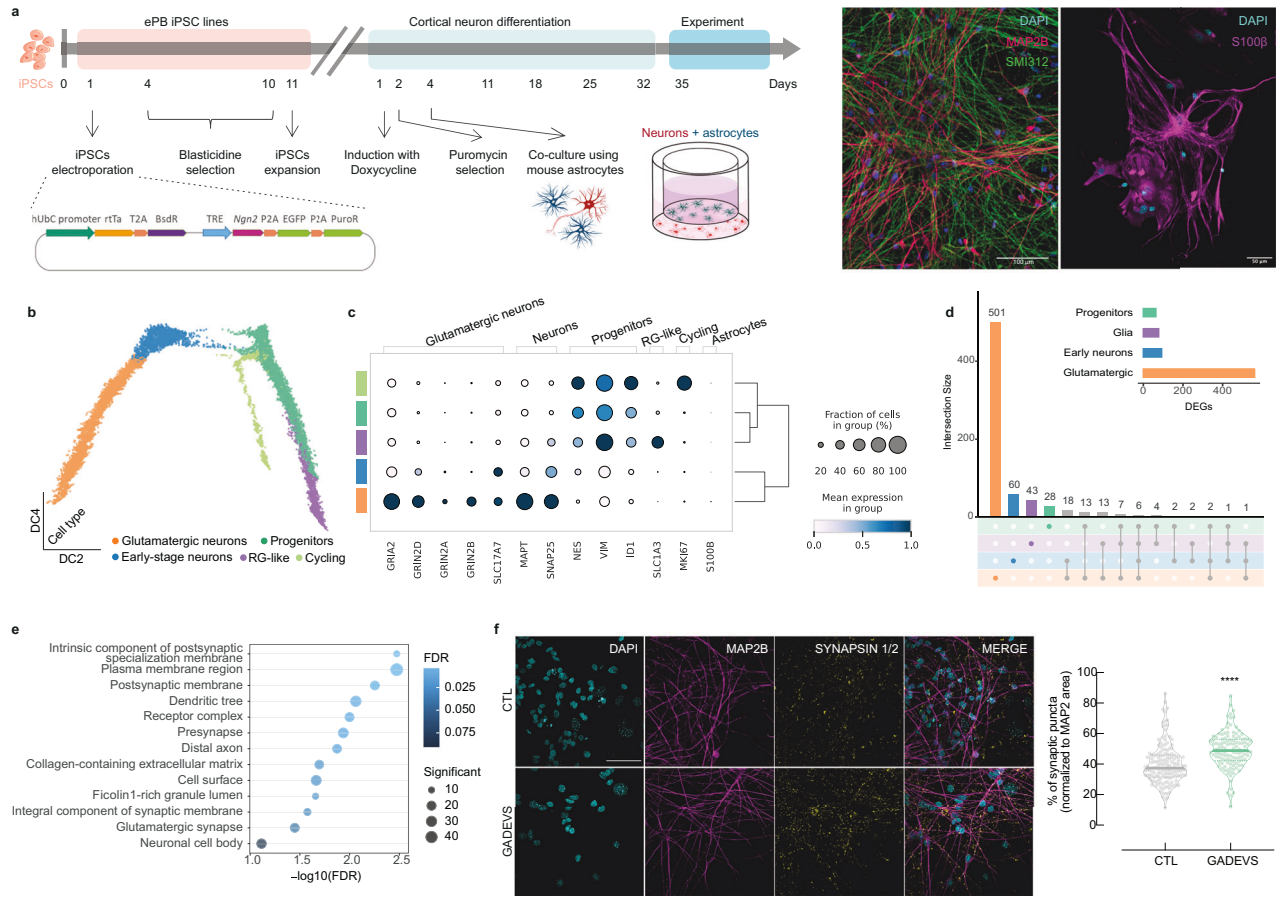


Fig. 5 GADEVS iPSC-derived glutamatergic neurons display transcriptional and morphological synaptic abnormalities. **a** Differentiation protocol to generate glutamatergic neurons starting from human derived iPSC. To induce NGN2 ectopic expression a ePiggyBac plasmid was transfected by electroporation. After selection, iPSC were co-cultured on transwell with fresh mouse astrocytes for 35 days. Immunostainings showed coherent expression of dendritic and pan-axonal markers MAP2B and SMI312 (scale bar 100 μm) and co-culture was achieved by addition of astrocytes (S100 β -positive, scale bar 50 μm). **b** Diffusion map of single-cell RNA modality from NGN2 iPSC-derived neurons show a heterogeneous set of cell types including early and mature neuronal lineages. A color legend depicting cell types is provided on the bottom of the panel. **c** Dotplot showing representative markers of glutamatergic neurons, pan markers of neuronal lineages, and markers of progenitors and glial cells. **d** Upset plot reporting the intersection of lists of differentially expressed genes (DEGs). Glutamatergic neurons (orange) show the largest cell type specific set of DEGs. **e** Dotplot of GO enrichments (biological process) for genes differentially expressed in the glutamatergic neurons cluster. **f** Representative figures of glutamatergic neurons immunostained with MAP2B and synapsins 1/2. Scale bar is 50 μm . Synaptic puncta were quantified by counting the number of synapsins 1/2 when co-localized with MAP2B de-noised signal for all fields of view acquired. Graph on the right shows the calculated percentage of synaptic puncta normalized to MAP2B (grouped analysis CTL $n = 156$ vs. GADEVS $n = 117$ fields of view across 4 independent seedings). Results are presented as median (min, max, and range) of percentages of synaptic puncta counted in n different fields of view independently acquired. **** $P < 0.0001$ vs. CTL using a two-tailed Mann-Whitney test.

advantage of the physical separation of the two cell types to also contextually profile neurons via single-cell multiomics (contextually capturing RNA- and ATAC from each single nuclei) and astrocytes via bulk RNA-seq. The single-cell transcriptome of neurons grown in co-culture with fresh astrocytes yielded 14,828 cells after quality control filtering (Fig. S8c). This data showed the presence of multiple progenitors (Fig. S8d, e), coherently with previous descriptions of heterogeneity in 2D Ngn2 induction experiments [49]. Diffusion maps effectively highlighted a population of cycling cells that differentiated into progenitors, which then generated two lineages (Fig. 5b), one of *NEUROG2*⁺ bona fide neurons, as expected, and surprisingly, a second one of *SLC1A3*⁺ cells (Fig. 5c, Fig. S8d, e), marking a cluster of radial glial-like cells that had not been annotated in previous studies. Additionally, we could identify two different states of neuronal maturation, which we designated as early-stage and glutamatergic neurons (Fig. 5c, Fig. S8f). After performing differential expression analysis (DEA) between GADEVS and CTL in cell type-specific pseudobulks, we observed that glutamatergic neurons

were the most dysregulated (Fig. 5d). Accordingly, GO analysis highlighted significant enrichments for cellular component categories related with glutamatergic pre- and postsynaptic compartments (Fig. 5e) and to biological processes related to neuronal functions, such as *anterograde trans-synaptic signaling* (Fig. S8g), whose disruption was detected also in organoids (Table S1). Thus, to account for disparities in synaptic assembly and organization we quantified a group of synapsins, which are important components of mature synapses in the cortex. As anticipated by GO enrichment analysis, *YY1*-mutated neurons exhibited a significant increase of dendritic synaptic puncta (GADEVS $n = 117$, 48.83%, 72.70 vs. CTL $n = 156$, 37.43%, 74.97) (Fig. 5f).

Astrocytes respond to environmental changes caused by *YY1*-deficient neurons and undergo cell activation

Based on the extensive alteration of the synaptic compartments, we reasoned that the consequences of *YY1* haploinsufficiency could extend beyond the affected neurons and adversely affects

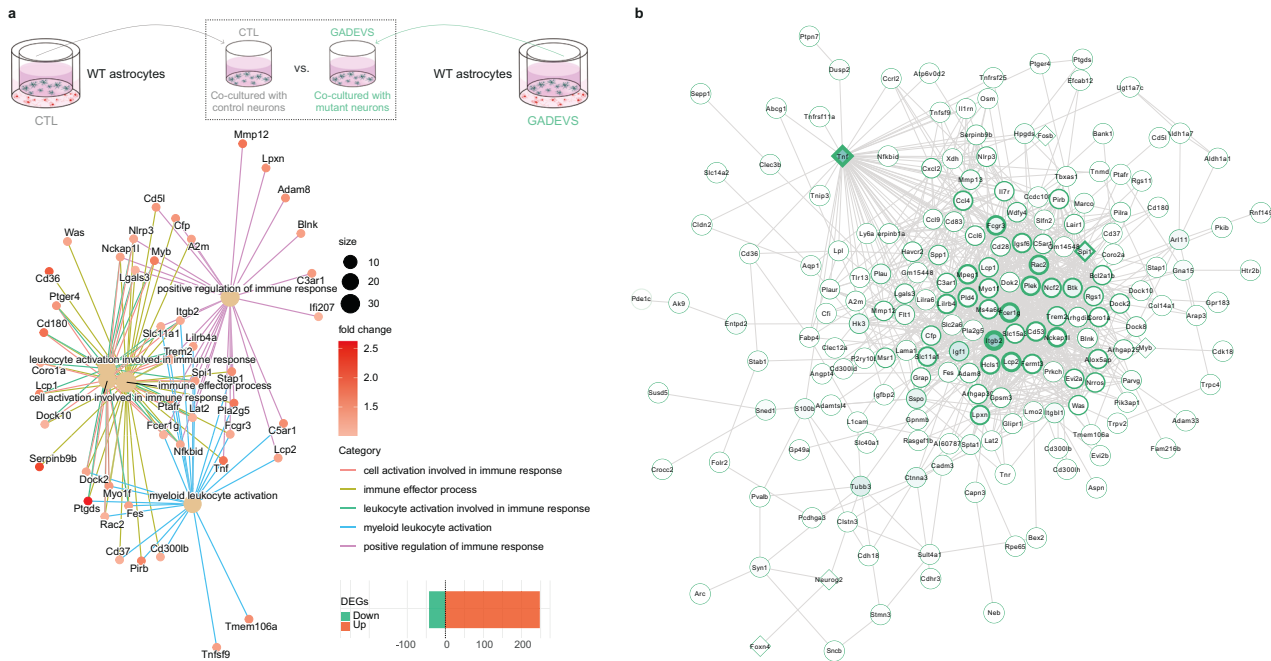


Fig. 6 Astrocytes co-cultured with affected neurons undergo cell activation. **a** Wild type mouse astrocytes co-cultured with both CTL and GADEVS neurons were harvested and profiled using bulk RNA-seq. DEA revealed predominant upregulation out of the total 291 DEGs (248 upregulated DEGs, $FDR \leq 0.05$ and $FC \geq |2|$). Gene-concept network plot describes the interactions across genes belonging to significant biological processes GO terms. **b** Main component of co-expression network generated from DEGs enriching *cell activation* GO categories depicted in panel a. Nodes of TFs are indicated as squares (*Spi1*, *Myb*, *Fosb*, *Foxn4*, *Neurog2*, and *Tnf*). Edges are denoted as light grey strokes. Node borders are proportional to its degree (highest degree scored 5, DEGs linked are *Itgb2*, *Tnf*, *Fcer1g*, *Lcp2*, and *Fcgr3*). The color of the node span from white (lowest) to green (highest). Nodes with highest centrality scores include *Itgb2*, *Igf1*, *Fcer1g*, *Tnf*, and *Sspo*. Centrality here is the topological index of each node (i.e., DEG). Degree centrality is proportional to the number of connections of a given node, while betweenness centrality correlates with the number of times a node is included in the shortest path between any other two nodes in the network.

the physiology of surrounding astrocytes. To quantify the possible impact of GADEVS NGN2-derived neurons on the adjacent co-cultured mouse astrocytes, we profiled mouse astrocytes using bulk RNA-seq, as they were co-cultured into distinct compartments separated by the transwells/cell culture inserts (Fig. 6a).

Specifically, to measure the non-cell autonomous effect of altered neurons on surrounding cells, we examined the transcriptional profile by bulk RNA-seq of healthy mouse astrocytes *vis à vis* healthy or GADEVS NGN2-derived neurons. With respect to astrocytes grown with healthy control NGN2-derived neurons, astrocytes co-cultured with GADEVS neurons showed 291 predominantly upregulated DEGs (Fig. 6a, 248 and 43 up and downregulated, respectively). These genes were enriched in GO processes involved in immune regulation and leukocyte reactive responses (Fig. 6a), including the more general “cell activation” GO category. Consistent with the notion that dysfunctional synaptic activity suffices to induce astrogliosis-like responses [50, 51], we observed that all genes enriching these GO categories were found to be exclusively upregulated (genes are depicted in each node in the gene-concept network plot, Fig. 6a). To verify the hypothesis that astrocytes are undergoing the activation process marked as astrogliosis, we performed a gene-regulatory network reconstruction using the genes associated with the “cell activation” GO category enriched in the DEGs. In the absence of chromatin accessibility data, to build the gene-regulatory network underlying astrocytes transcriptional rewiring in the presence of GADEVS neurons we inferred transcription factor activity with *viper* using the DoRothEA database [52] (see Methods). As a result, we identified 193 nodes forming one unique network in which *Spi1*, *Myb*, *Fosb*, *Foxn4*, *Neurog2*, and *Tnf* represent master regulators, controlling most of the nodes (Fig. 6b). Of particular interest, *Tnf* is one of the genes with highest degree score in the network, which

represents a measure of its direct connections with the rest of the network (genes). *Tnf* encodes a multifunctional proinflammatory cytokine belonging to the tumor necrosis factor (TNF) superfamily, which is one of the main pro-inflammatory cytokines involved in the regulation of astrocytes and microglia activation [53]. Directly linked with *Tnf*, other TNF ligands and receptors including *Tnfsf9*, *Tnfrsf11a*, and *Tnfrsf25* are among the upregulated DEGs linked with astrogliosis gene signatures. Likewise, macrophage-associated genes (*Fcer1g*, *Fcgr3*, *Mpeg1*, *Lcp2*, *Rac2*, *Trem2*, and *Nckap1l*), chemokines, interleukin receptors, channels, and serpin protease inhibitors (*Il1rn*, *Il7r*, *Ccl4*, *Ccl6*, *Ccl9*, *Serp1b1*, and *Serp1b9b*), potentiators and markers of astrogliosis (*S100b*, *Lgals3*, *Nfkbid*, *P2ry10b*, *P2ry12*, *Nfkbid*, *Spp1*, *Aif1*, *Cd83*, *Cd1*, *Pdgfb*, and *Vegfa*), as well as other pivotal genes associated with inflammatory responses (*A2m*, *Lilrb4*, *Adam8*, *Lpl*, *Sifn2*, *Plaur*, *Plau*, *Cxcl2*, *Adamts4*, *Hpgds*) are depicted in the DEGs network underscoring a primary composition of astrocyte activation genes (Fig. 6b).

Mechanistic dissection of YY1-mediated dysregulations in glutamatergic neurons revealed a broad feedback loop attempting at rescuing YY1 haploinsufficiency and a pervasive role of NEUROG2 in counteracting YY1 activity across biological processes

GADEVS glutamatergic neurons displayed transcriptional and synaptic dysregulations that were sufficient to trigger astrocytic activation, shown by upregulation of gliosis signatures. Thus, to causally pinpoint the disruptions in YY1 regulatory activity, and identify its regulatory partner we leveraged the contextual profiling of transcription and chromatin accessibility in neurons via multiomics described in the previous chapters. This allowed us to identify differential TF activity and generate cell type specific GRN to assess the YY1 dysfunctions in mature cortical excitatory neurons.

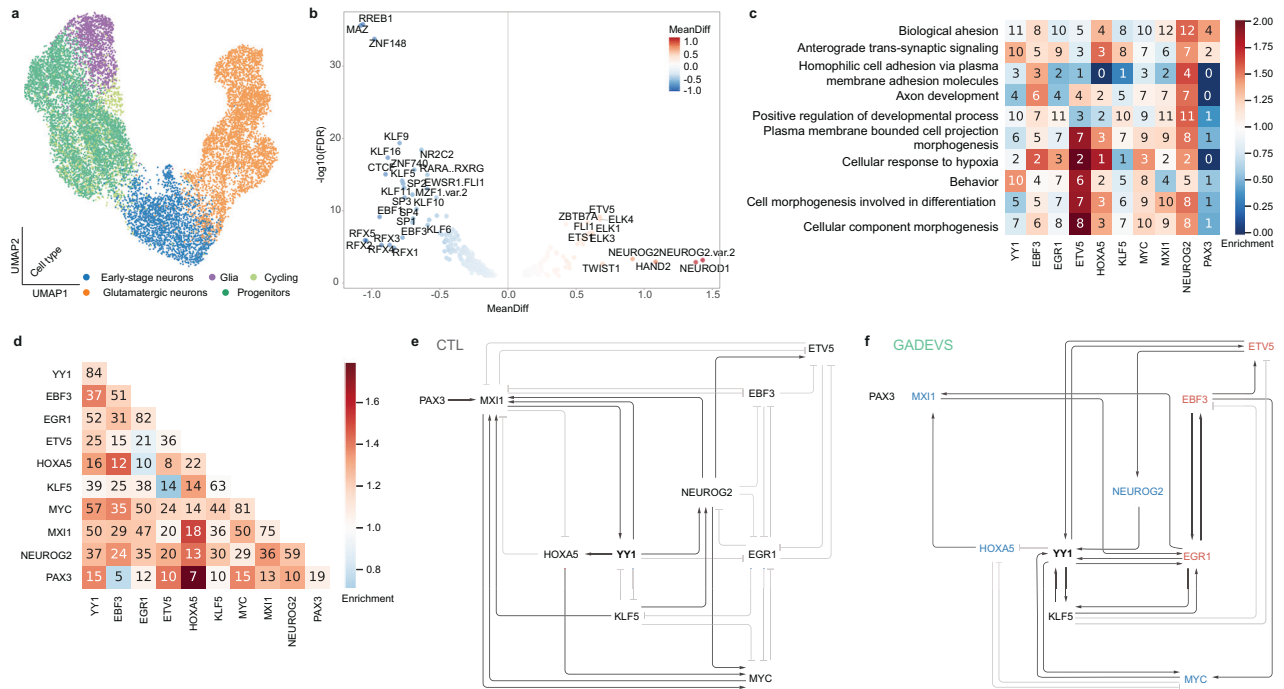


Fig. 7 Mechanistic dissection of molecular dysregulations in GADEVS iPSC-derived NGN2 neurons depicts a broad loss of TF activity, and a subnetwork of transcriptional regulation pinpointing YY1 haploinsufficiency to pathologically relevant processes. **a** UMAP representation of single-cell ATAC data. Cells are colored by transferring cell type annotation from each cell shared with the RNA-seq modality. Cell type specific colors are reported on the bottom of the panel. **b** Differential TF activity measured with ChromVar by inference from accessibility changes between control and GADEVS neurons at regulatory regions identified by co-accessibility with Cicero. MeanDiff represent the differential TF activity measured by ChromVar. **c** heatmap of GO (biological process) categories enriched by differentially expressed genes. Number of DEGs enriched in each category under the control of each TF (reported below) is indicated in each square. Each square is colored by enrichment measured by the representation of targets of each TF in each category. **d** Heatmap of shared targets between pairs of TFs among differentially expressed genes. Each square is coloured by the relative enrichment measured on the targets of the TF indicated on the rows with respect to the target indicated in the column (the number reported in each square of the heatmap is the number of shared DEGs between the factor on the x-axis and the factor on the y-axis). **e, f** Subnetwork of control (**e**) and GADEVS-specific (**f**) regulatory interactions among the TFs reported in panels c and d. Arrows and T-shaped edges respectively represent activation and repression (as inferred by CellOracle). Blue and red colored gene names respectively denote downregulation and upregulation of the TFs in GADEVS glutamatergic neurons.

First, we verified the consistency of the single-cell ATAC modality of NGN2-derived neurons with respect to the RNA modality, by transferring cell type annotations from the latter onto the UMAP of the former (Fig. 7a). Then, we quantified TF differential activity to identify the most disrupted regulators (Fig. 7b). We identified a general depletion of accessibility for most differentially accessible motifs, and a strong increase in accessibility in fewer TFs, including NEUROG2 and NEUROD1. To reconstruct the contribution of differentially active TFs to the differential expression previously identified in GADEVS mature neurons, we first performed a GO enrichment analysis on the differentially expressed genes, and then measured the enrichments within each enriched GO category for the direct targets of YY1 and those of TFs that simultaneously display significant differential activity and differential expression in neurons (Fig. 7c). By doing so, we found *ETV5* to be the most strongly enriched for a subset of rather distinguished categories, such as *hypoxia*, *morphogenesis* and *behavior*, and we found *NEUROG2* enriched in all categories, including *cell adhesion*, and developmental processes. Thus, to further pinpoint the molecular underpinnings of such dysregulations, we extended our gaze to the contribution of these differentially active and differentially expressed TFs to the regulation of all DEGs in mature neurons and plotted the number of shared targets between each TFs and YY1 (Fig. 7d). We found a large portion of DEGS to be target of MYC (the TFs with the highest number of target DEGs shared with YY1) and its two partners MXI1 and EBF3, immediately followed by NEUROG2. Thus,

to pinpoint the mechanistic co-dependency between these TFs, we separately derived networks of their reciprocal regulation for control and GADEVS samples and analyzed interactions specific to each genotype to observe potential disruptions in this crosstalk in the patients. In the controls (Fig. 7e), we found YY1 as a regulator of NEUROG2, together with KLF5, which appeared involved in a self-repressive interaction with YY1. The GADEVS-specific GRN showed a complete rewiring of this network. Apparently, YY1 haploinsufficiency triggered the contextual downregulation of *NEUROG2*, *MXI1*, and *HOXA5*, and the upregulation of *EGR1*, *EBF3* and *ETV5*. Concomitantly, the rewired network highlighted a coordinated attempt at upregulating YY1, which included a positive regulation of YY1 by NEUROG2 (Fig. 7f). Notably, this comparison is based on the independent analyses of GADEVS and CTL cells with CellOracle and the integration of RNA and ATAC modalities. Hence, it's grounded on the contextual profiling of accessibility and expression in each high-quality profiled cell, but it is dependent on the correct annotation of cell types and on the adherence between CellOracle motif database and the effective binding of TFs. We believe that the former (cell type annotation) is well substantiated in the previous chapters of this work while the latter (TF-targets association), which has been inferred from the accessibility of regulatory regions associated with each gene via co-accessibility (estimation of all known TFs binding) could not be systematically profiled experimentally. Hence, while the motif enrichment of accessible regions could be a suboptimal proxy of TF binding activity, the identification of regulatory regions by

co-accessibility performed on ATAC data - coupled at the single-cell level with the transcriptional profiles analyzed - constitutes the state-of-the art for these types of endeavors.

DISCUSSION

We previously showed that de novo mutations of *YY1* cause GADEVs through haploinsufficiency-dependent defects in the activation of *YY1*-bound enhancers [1]. Consistently, *YY1* was found shortly thereafter to act as a general enhancer-promoter looping factor [10, 11]. However, genome-wide studies on *YY1* following its acute depletion assayed its role in regulating the 3D genome architecture and transcription initiation levels debated its looping role [14]. Nevertheless, these studies were done in mouse embryonic stem cells, and did not account for the effects of *YY1* in differentiation [54]. Therefore, we hypothesized that *YY1* has a role in establishing cell type identity rather than maintaining pluripotency, altering specific differentiation axes. Consequently, to identify the physio-pathologically relevant cell type specific impact of *YY1* mutations, we modelled GADEVs neurodevelopment by leveraging patients derived iPSCs, 2D and 3D neuronal lineages. First, in iPSC we identified a broad transcriptional downregulation explained by a global loss of *YY1* across different pathogenic truncating or missense mutations. Notably, dysregulated genes in iPSC significantly relate to physio-pathological phenotypes, reinforcing the importance of early developmental settings in revealing the molecular priming of neurodevelopmental conditions [18].

To define the impact of GADEVs-causing *YY1* mutations on corticogenesis we harnessed patterned cortical brain organoids. Despite an overall similarity in cellular composition in terms of both neuroepithelial and downstream progenitor cells at early developmental stages, we found a striking and highly reproducible faulty organization of ventricle-like structures within GADEVs organoids. Studies conducted across various animal models [8, 33, 55–58] had also found evidence of dysfunctional VLS features, hence our observations underscored the fidelity of our model in capturing fundamental aspects of neurodevelopment with the first recapitulation of a pseudo-ventricle phenotype in human organoids which emerges as earliest correlate of the morphological brain phenotype detected in GADEVs individuals with MRI or CT scans^{1,15,33,58}. Subsequently, at 30 days of organoid differentiation when neurogenesis kickstarts following the expansion of neuroepithelial cells, with progenitor diversification and the emergence of neuronal lineages, we observed a very specific dysregulation affecting intermediate progenitors. Following on the course of differentiation, late stages (90 days) manifested further major differences in terms of neuronal subtypes among lower- and upper-layer neurons, with dysregulated *YY1* direct targets pointing to disruption of ion channels expression and dysregulation of cell-cell signaling in the EGFR, MAPK, WNT, TNF α and NF-KB (indicative of glial-derived inflammatory processes) pathways.

The cell-cell signaling and synaptic dysregulation were further probed in the 2D NGN2-driven neuronal differentiation paradigm. This revealed in neurons a profound transcriptional rewiring induced by *YY1* mutations, subverting the regulons of 6 main transcription factors (*NEUROG2*, *MXI1*, *HOXA5*, *EGR1*, *EBF3* and *ETV5*) and thereby unmasking a physiological feedback loop between *YY1* and NGN2, whose positive regulation by *NEUROG2* on *YY1* proved however ineffective in rescuing the haploinsufficient setting. Finally, *YY1*-defective neurons exerted a profound effect on neighboring astrocytes, triggering an inflammatory and astroglial response that represents, alongside altered pseudo-ventricles, yet another remarkable endophenotypic correlate of specific alterations observed in the frontal cortex of GADEVs affected individuals [5]. In sum, our work defined the activity of *YY1* in human neuronal lineages, across multiple stages of

development, highlighting both *YY1* regulatory activity and its downstream effects on neuronal development and mature neuronal functions, and extending the understanding of *YY1* haploinsufficiency to non-cell autonomous alterations. As with any condition of such complexity, this multiscale dissection of endophenotypes can now guide further modelling studies with the aim of identifying the most actionable targets that relay the molecular mechanisms exposed by single cell multiomics to the main cell biological defects at the level of pseudo-ventricles, synaptic puncta and reactive astrocytes. In this regard, given that even astrocyte malfunctions alone are sufficient to induce changes in basal transmission and plasticity processes [59], our findings of neuronal-mediated inflammation points to look at inflammation modulation as a potential path ahead for preclinical testing in GADEVs.

As an increasingly recognized condition [6, 33] GADEVs calls for a matched research and care strategy aimed at ever more accurate screening and detailed phenotyping of case reports to refine clinical description and align it to the underlying molecular mechanisms. Therefore, we established a website to collect detailed phenotypic data of individuals harboring *YY1* mutations not only to gain insight into the clinical spectrum that these mutations might cause but also to obtain a fundamental understanding of the pathogenic mechanisms underlying the Gabriele-de Vries syndrome (<https://humandiseasesgenes.nl/yy1>). With this first modelling of GADEVs, we provide a high-resolution definition of its key pathogenic mechanisms, uncovering multi-scale longitudinal phenotypes, cell-autonomous and non-cell autonomous alike, whose multiomic underpinnings and functional readouts orient the search for druggable targets and repurposed compounds in physio-pathologically relevant human neuronal lineages [59–61].

MATERIALS AND METHODS

Cell reprogramming

Human fibroblasts were isolated from skin biopsies and lymphoblastoid cell lines (LCLs) were generated by Epstein-Barr virus (EBV) transformation of the B-lymphocytes within the peripheral blood lymphocyte leading to proliferation and subsequent immortalization of these cells. Both non affected and affected individual's - samples (fibroblasts and LCLs) were reprogrammed using the non-integrating Sendai virus (SeV) method to safely and efficiently deliver and express key genetic factors (Yamanaka factors OCT3/4, SOX2, KLF4, and c-MYC). For all lines, a total of 3×10^5 cells were transduced per individual using a multiplicity of infection (MOI) of 5–5–3 (KOS comprising human Klf4, Oct3/4, and Sox2; c-MYC containing human c-Myc; Klf4 with human Klf4, respectively). Moreover, the amount of virus used was calculated according to each certificate of analysis respective of each SeV kit. The reprogramming procedure was performed according to the manufacturer's protocol using feeder-dependent protocols for both fibroblasts and PBMCs (feeder layer containing mouse embryonic fibroblasts). Cells were maintained every day and small colonies were subsequently picked and expanded approximately 1–1.5 months post-transduction. After reprogramming from LCL, EBV expression in iPSCs was evaluated, and samples completely turned off EBV expression compared to a positive control cDNA taken from LCL, showing irrelevant Ct, and with melting curves comparable with the negative (no cDNA). Reprogrammed fibroblast-derived iPSCs were also validated to be free of SeV transduced vectors by RT-qPCR. iPSC cells were pulled down and pelleted for RNA extraction and depletion was checked by RT-qPCR according to manufacturer's instructions. A reverse transcription was carried out using 2.5 μ g of RNA following the instructions provided with the kit (Cat N^o, 11766050, Invitrogen). For PCR reaction mix preparation, 12 ng of cDNA for each sample was used and performed in triplicates using the Fast SYBR Green Master Mix (visualize primers in Table S2). Positive control cells were set aside during the first stages of the reprogramming procedures (Fig. S1, Cat N^o A16517, ThermoFisher).

Cell culture

Fibroblast/LCL handling and iPSC culturing. Human fibroblasts were cultured in DMEM, FBS 10%, non-essential amino acids (NEAA) 1%,

penicillin-streptomycin (P/S) 1%, and β -mercaptoethanol 0.1%. LCLs were grown in RPMI 1640, FBS 15%, HEPES 1%, L-glutamine (Gln) 1%, and P/S 1%. Trypsin was used to passage fibroblasts whereas LCLs were cultivated in suspension and expanded by dilution. Human iPSCs were cultured in feeder-free conditions onto matrigel-coated dishes using TeSR-E8 medium supplemented with P/S 1% and passaged using either ReLeSR (Cat N° 100-0483 stemcell technologies, generates cell aggregates for expansion and standard maintenance) or Accutase solution (Cat N° A6964 Sigma, exclusively for experimental procedures) supplemented with ROCK inhibitor 5 μ M (Y-27632 dihydrochloride, #A3008 ApexBio) was added to the culture to enhance single cell survival. Cryopreservation of hiPSCs was performed in complete TeSR-E8 medium with 10% DMSO and supplemented with ROCK inhibitor 5 μ M. All samples were routinely tested for mycoplasma.

Directed differentiation of hiPSCs into ectoderm, mesoderm, and endoderm lineages. To functionally validate the potential of human iPSC lines to differentiate into all three embryonic germ layers, we took advantage of the STEMdiff™ trilineage differentiation kit and followed the manufacturer's instructions with minor modifications (Cat N° 05230, Stemcell technologies). hiPSCs were cultured onto Matrigel-coated glass coverslips and on day 0, 2.5×10^5 , 7.5×10^4 and 2.5×10^5 cells were plated for further differentiation into ecto-, meso-, and endoderm, respectively. At the end of differentiation, cells were fixed, and lineage-specific markers were used for immunofluorescence using the following established markers NESTIN and PAX6 for ectoderm; Brachyury and CXCR4 for mesoderm; and FOXA2 and SOX17 for endoderm (Fig. S2).

Differentiation of iPSCs into neural crest stem cells (NCSCs). hiPSCs were differentiated into NCSCs as previously described [62]. NCSC differentiation required 15–20 days and was carried out as follows: 90% confluent iPSCs were detached with Accutase solution and plated onto matrigel-coated dishes using TeSR-E8 medium supplemented with 5 μ M ROCK inhibitor at a density of approximately 9.2×10^4 cells per cm^2 . After 24 h, NCSC differentiation medium was added and changed every day. NCSC medium was composed of DMEM-F-12 1:1, 10% probumin (stock solution of 20% m/v in DMEM F-12 1:1), P/S 1%, L-Glutamine 1%, NEAA 1%, trace elements complex 0.1, 0.2% β -mercaptoethanol (50 mM), Transferrin (10 μ g/mL), sodium L-ascorbate (50 μ g/mL), Heregulin-1 (10 ng/mL), LONG®R3 IGF-1 (200 ng/mL), FGF2 (8 ng/mL), GSK3 inhibitor IX Bio (3 μ M) and SB431542 (20 μ M). Cells were passaged every 4–5 days and plated at high densities. Upon differentiation, NCSC were stocked as stable lines and cultured using NCSC medium with routine splitting ratios of 1:4–5 (Fig. S3e, f).

Preparation of primary mouse astrocytes. Primary astrocyte cultures were created from the cerebral cortices of embryonic day 18 (E18) mice embryos and maintained as previously described [63, 64]. In a nutshell, after embryo collection, cortices were dissected from each embryo's brain, and chemical dissociation with trypsin (2.5%, at 37 °C for 30 min) was performed. Afterwards, mechanical dissociation was performed by vigorous pipetting. Following digestion, cortex tissues were centrifuged (300 g for 5 min) and supernatant discarded. Finally, astrocyte medium (DMEM supplemented with DMEM medium supplemented with FBS 20%, P/S 1%) was added and $1\text{--}1.5 \times 10^7$ cells were seeded in each T75 culture flask. After reaching confluence (10–15 days), astrocytes were passaged with trypsin and expanded. This protocol allowed the generation of pure astrocytic cultures due to the gradual absence of viable neurons and other glial cells. Astrocytes were generated and maintained 3–4 weeks prior NGN2-derived neuron differentiation.

Differentiation of iPSCs into NGN2-derived cortical neurons. hiPSCs were engineered using a ePiggyBac (ePB) transposon previously electroporated using the Neon™ Transfection System (MPK5000, ThermoFisher). Electroporation parameters for each reaction of 4×10^5 cells were: 900 V, 20 ms, 2 pulses using 5 μ g of total DNA in a 1:10 ratio of helper:vector (0.5 μ g of a helper plasmid expressing a transposase and 4.5 μ g of donor plasmid with a transposable element). NGN2 ePB donor plasmid has the following genetic configuration: hUbc promoter - rTA - T2A - BsdR - TRE - Ngn2 - P2A - EGFP - T2A - PuroR and allows the selection with blasticidin (Fig. S7a). Following transfection, cells were allowed to recover for 48 h and selection was performed using blasticidin 5 μ g/mL until all cells were eliminated from the negative control performed without the addition of the transposase (usually 5–7 days of selection). Stable iPSCs containing the ePB transposon were stored until needed for neuronal differentiation.

Neuronal differentiation is induced by doxycycline and is followed by

NGN2 overexpression, as previously described [65]. Cortical neurons were maintained in Neurobasal medium fully supplemented (B27 with vit.A + P/S + GlutaMAX + doxycycline + puromycin + N2 + NEAA + human Laminin + NT3 + BDNF). Fresh mouse astrocytes were added to the culture by means of hanging cell culture inserts (MCRP06H48, Millipore, transwell for clarity) on day 4 after definitive plating of neurons in poly-D-lysine-coated 6 well plates or added into poly-L-ornithine-coated glass coverslips treated with nitric acid for IF experiments at a ratio 1:1 (human neuron to mouse astrocyte).

Differentiation of iPSCs into 3D human in vitro systems – cortical brain organoids. Cortical brain organoids were generated using an adaptation of a previously described protocol [36]. Prior to organoid culture, iPSCs were instead routinely cultured with mTeSR™1 basal medium (Cat N° 85850 stemcell technologies). hiPSC were expanded in 10 cm dishes and detached at 60–70% confluency with Accutase solution for obtaining single cell suspensions. After centrifugation (150 g, 3 min), cells were resuspended in mTeSR™1 basal medium supplemented with ROCK inhibitor (5 μ M), TGF- β inhibitor SB-431542 (10 μ M), and dorsomorphin (1 μ M) and counted. To form free-floating spheres, approximately 4×10^6 cells were added to a single 6-well plate and kept under rotation (95 rpm) inside the incubator (37 °C, high O₂ conditions) for 24 h. The medium was refreshed every day by removing half (2 mL) and adding half (2 mL) for the following 2 days. After 3 days (day 4 to 10), supplemented mTeSR™1 basal medium was replaced by Media 1, comprised of Neurobasal A medium supplemented with 1% GlutaMAX, 1% N2 NeuroPlex, 1% B27 with vitamin A, 1% NEAA, 1% P/S, SB (10 μ M) and Dorsomorphin (1 μ M). Subsequently, from days 11 to 17, cells were maintained in Media 2: Neurobasal A medium with GlutaMAX, 1% B27 + vitamin A, 1% NEAA and 1% P/S supplemented with FGF2 (20 ng/mL), followed by 7 additional days (day 18 to 24) in Media 2 supplemented with FGF2 (20 ng/mL) and EGF (20 ng/mL) to boost neural progenitor proliferation. Finally, organoid medium conditions shifted to Media 3, composed of Media 2 supplemented with BDNF (10 ng/mL), GDNF (10 ng/mL), NT-3 (10 ng/mL), L-ascorbic acid (200 μ M) and dibutyryl-cAMP (1 mM) to promote network maturation, gliogenesis and enable neuronal activity. After 7 days (31 days in vitro), cortical organoids were maintained in Media 3 for as long as needed, with media changes every 3–4 days. Catalog numbers for all reagents and small molecules used for the cell culture medium were previously described [36].

RNA extraction and library preparation for bulk RNA sequencing (RNA-seq)

Cultured iPSCs and NGN2-derived neurons were washed with phosphate-buffered saline (PSB) and treated with 600 μ L RTL buffer supplemented with 10% β -mercaptoethanol. Afterwards, total RNA was extracted using the RNeasy Mini Kit. Purified RNA was quantified using a NanoDrop spectrophotometer and RNA quality was further checked with an Agilent 2100 Bioanalyzer using the RNA nano kit (Cat N° 5067-1511, Agilent). Only samples with RIN > 9 were used for library preparation. Prior to library preparation, ERCC spike-in mixes (Cat N° 4456739, ThermoFisher) were added to all samples to facilitate data normalization and quantification. RNA sequencing libraries were prepared following manufacturer's protocols for Truseq-stranded Total RNA (RiboZero depletion) or using Illumina Stranded mRNA Prep, starting from 250–500 ng of total RNA. cDNA library quality was assessed on Agilent 2100 Bioanalyzer, using the Agilent High Sensitivity DNA Kit (Cat N° 5067-4626, Agilent). Libraries were then sequenced with the Illumina Novaseq 6000 instrument at a read length of 50 bp (paired-end) with a coverage of 35 million reads per sample.

Chromatin immunoprecipitation followed by sequencing (ChIP-Seq)

ChIP-seq for TF YY1 was performed in all iPSC lines (4 CTLs vs. 3 GADEVs lines) using approximately 1×10^7 cells per sample. First, cells were harvested and crosslinked with formaldehyde 1% in PBS for 8 min at RT under continuous rotation. To quench the reaction, glycine (125 μ M) was added to each mix and incubated for 5 min on ice. Cells were subsequently pelleted by centrifugation (500 g, 3 min at 4 °C) and washed in ice-cold PBS. Cells were lysed in 10 mL buffer A (50 mM HEPES pH 8.0, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) for 10 min on ice. After centrifugation, cell pellet was resuspended in 5 mL buffer B (10 mM Tris pH 8, 1 mM EDTA, 0.5 mM EGTA and 200 mM NaCl) and incubated for 5 min on ice. Next, nuclei were pelleted (500 g, 3 min at 4 °C), resuspended in 150 μ L buffer C (50 mM Tris pH 8, 5 mM EDTA, 1% SDS, 100 mM NaCl, 1x Roche complete mini protease inhibitors) and incubated

on ice for 10 min. Afterwards, 350 μ L ice-cold TE buffer was added and chromatin was sheared in 1.5 mL microtubes using a Bioruptor Pico sonication device (B01060010, Diagenode) for 3 \times 5 cycles (30 s ON/ 30 s OFF) at 4 $^{\circ}$ C. Upon addition of 55 μ L of 10 \times ChIP buffer (0.1% SDS, 10% Triton X-100, 12 mM EDTA, 167 mM Tris-HCl pH 8, 1.67 M NaCl), chromatin was centrifuged (13,000 g, 10 min at 4 $^{\circ}$ C). At this step, 1% of sheared chromatin was saved as input control, whereas the rest was transferred into fresh tubes. 10 μ g of antibody were added for YY1, and 5 μ g for normal rabbit IgG immunoprecipitations (antibody information on Table S3) were then incubated overnight on a rotating wheel at 4 $^{\circ}$ C. The following day, 40 μ L of protein G Dynabeads (mixed 1:1 and washed with 1 \times ChIP buffer) were added and incubation was continued for another 4 h. Each ChIP were washed: 2 \times LSB (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 140 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate); 3 \times HSB (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 360 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate); 1 \times LiSB (10 mM Tris-HCl pH 8.0, 1 mM EDTA, pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate); and 1 \times TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 50 mM NaCl). Beads were transferred into a new tube during the last wash, and wash buffer was completely removed before adding 100 μ L of elution buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 150 mM NaCl, 1% SDS). Mix was incubated for 30 min at 65 $^{\circ}$ C with constant shaking, and after 2 μ L RNaseA (20 μ g/ μ L) were added, samples were incubated for 1 h at 37 $^{\circ}$ C. Input samples were adjusted to 150 μ L total volume with elution buffer and processed similar to ChIP samples. Then 2 μ L Proteinase K (20 mg/mL) was added and samples were incubated 2 h at 55 $^{\circ}$ C and then O/N at 65 $^{\circ}$ C. After two volumes of AMPure XP beads and one volume of Isopropanol were added, samples were vigorously mixed and incubated for 10 min at RT. Next, beads were collected on a magnetic rack, washed twice with 80% EtOH, and DNA was finally eluted in 40 μ L 10 mM Tris pH8.0 for 5 min at 37 $^{\circ}$ C. DNA libraries were prepared and sequenced on the Illumina Novaseq 6000 instrument at 50 bp paired-end read length and a coverage of 35 million reads per sample.

Single cell multiomics (gene expression + chromatin accessibility)

NGN2-derived neurons were detached with Accutase, and single cell suspension was immediately processed. As opposed to neurons, cortical organoid dissociation required a more tailored and gentle procedure. 30- and 90-days cortical organoids were dissociated using a papain-based dissociation buffer composed of 30 units/mL papain (Cat N $^{\circ}$ 07466, stemcell technologies), dissolved in filter-sterilized activation solution comprised of 1.1 mM EDTA, 0.067 mM mercaptoethanol, 5.5 mM L-cysteine HCl) and 125 units/mL DNase I (07900, Stemcell technologies) in Hanks' Balanced Salt Solution (HBSS) with 10 mM HEPES, without phenol red (Cat N $^{\circ}$ 37150, stemcell technologies). Pools of homogeneously sized organoids (4–8 organoids depending on timepoint) were incubated on a rotating wheel (approximately 30 min at 37 $^{\circ}$ C) with 1 mL of activated dissociation buffer, followed by manually pipetting. Afterwards, cells were transferred into a new eppendorf leaving behind undissociated pieces, and further centrifuged (300 g, 5 min at 4 $^{\circ}$ C). Supernatant was removed and cells were resuspended in 1 mL of PBS-BSA 0.04% and filtered with 40 μ m Flowmi cell strainer to remove residual aggregates and debris. Single cell suspensions were counted manually (diluting 1:1 with trypan blue), and different lines were multiplexed together to obtain a final number of 1 million cells, equally mixed (e.g., 250k cells in case of 4 samples multiplexed), assembling a single reaction. This step allowed to pool together different genotypes which can later be demultiplexed based on their transcriptome previously profiled with bulk RNA-seq (at iPSC stage or NGN2-derived neurons) [66].

Nuclei isolation for single cell multiome (ATAC + Gene expression sequencing) was performed according to the manufacturer instructions in the demonstrated protocol from 10X Genomics (CG000365-RevB). Briefly, each reaction was washed twice with PBS-BSA 0.04%, and pellet was resuspended in 100 μ L of lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween-20, 0.1% Nonidet P40 Substitute, 0.01% digitonin, 1% BSA, 1 mM DTT, 1 U/ μ L RNase inhibitor, nuclease-free water), and incubated for 3 min on ice. Lysis buffer was washed away three times with 1 mL of wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 1% BSA, 0.1% Tween-20, 1 mM DTT, 1 U/ μ L RNase inhibitor, nuclease-free water). Assuming 50% nuclei loss during lysis, nuclei were resuspended in diluted nuclei buffer (1X Genomics Nuclei Buffer, 1 mM DTT, U/ μ L RNase inhibitor, nuclease-free water) according to the reference table provided by 10X protocol appendix. The volume of nuclei diluted

buffer is critical to fit the right range of concentration based on the number of targeted nuclei recovery, therefore avoiding overcrowding of the Chromium machine during tagmentation and GEM preparation steps. Resuspended nuclei were passed again through a Flowmi cell strainer and counted to fit the right concentration for the number of targeted nuclei (5 000 was the number of targeted nuclei recovery for each multiplexed sample in all the experiments). DNA libraries were prepared according to the manufacturer's recommendations (10X Genomics) and further sequenced on the Illumina Novaseq 6000 instrument at a coverage of 50,000 reads per nucleus.

Western Blot

Human iPSCs were grown as previously described, pelleted (300 g, 5 min), and then resuspended in 100 μ L RIPA buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5 mM EDTA, and 5% glycerol) supplemented with protease inhibitor cocktail (PIC) and 1 mM PMSF. Total protein lysate concentration was measured using the Biorad protein assay. For western blotting, 30 μ g of protein were resolved on polyacrylamide gels (12% separating gel), which were transferred on nitrocellulose membrane. Blocking was performed for 45 min in 5% non-fat dry milk in TBS plus 0.05% Tween 20, and membranes were incubated with primary (overnight at 4 $^{\circ}$ C) and secondary (2 h at RT) antibodies. Signal was detected with corresponding horseradish peroxidase (HRP)-conjugated secondary antibodies (antibody information on Table S3), imaged with ChemiDoc XRS+ System (Biorad), and quantified.

Immunofluorescence, image acquisition, and data analysis

Immunocytochemistry using 2D samples. Control and mutated cells (iPSCs, human-derived neurons, and mouse astrocytes) were cultured onto nitric acid-treated glass coverslips as previously described until wash (PBS) and fixed (4% methanol-free formaldehyde for 10 min at RT). For immunofluorescence (IF), fixed cells were then permeabilized with 0.2% Triton X-100 diluted in PBS for 10 min and blocked for 1 h in 5% serum matched with the species of the secondary antibody (donkey serum). Next, samples were incubated with the primary (overnight at 4 $^{\circ}$ C, antibodies described in Table S3) and secondary (1 h at RT) antibodies diluted in blocking buffer. Antibodies were removed by washing with PBS in between incubations. Afterwards, cells were stained with DAPI (1:5000) for 5 min, washed with PBS, and mounted onto glass slides with Mowiol-Dabco mounting medium. Samples were stored at –20 $^{\circ}$ C until visualized in the confocal microscope.

Acquisition and data analysis for counting synaptic puncta. Following fixation, staining, and mounting into glass slides, human-derived neurons were visualized using a Leica SP8 Confocal microscope equipped with acousto-optical beam splitter (AOBS), resonant scanner, motorized stage x-y-z as well as DM camera for widefield acquisition (Leica Microsystems, Germany). Previews of the whole coverslip were taken in widefield mode (10X/0.3 dry) using the DAPI channel to choose continuously the central area of the slide, that was further acquired at a higher resolution in confocal mode using the LASX software equipped with navigator (version 3.1.5.16308). Three-channel (MAP2, Synapsin and DAPI) z-stack images (z-step intervals of 1 μ m) were sequentially acquired using a 63X/1.4 oil objective and a DFC365 FX CCD Camera (Leica) with a x-y sampling of 72 nm. A total of 10 fields of view were acquired for each glass slide (with 4 slides being imaged, 40 fields of view per sample imaged and analyzed). After acquisition, each field of view was analyzed individually using an open-source software followed by custom design of a semi-automated macro (FIJI-ImageJ v2.1.0, USA, macro 1) by segmentation the C1 (MAP2B) and C2 (Synapsin1/2) channels for further analysis. First, all C1 channel fields of view were transformed into binary images and converted to masks, from which the area was measured to serve as normalizer as well as used to capture exclusively synapsin1/2-positive signal co-localized within MAP2B-positive structures. Then, synapsin1/2-positive particles were measured, ROIs saved, and counts normalized in relation to MAP2B area for each field of view. All data is presented as median of *n* different fields of view for all conditions (transformed as percentages).

Immunostaining and clearing of 3D structures - cortical brain organoids. Whole cortical organoids were immunostained and cleared using the MACS $^{\circ}$ Clearing Kit (Miltenyi Biotec, 130-126-719) and following entirely the manufacturer's instructions (Immunostaining and clearing of organoids and spheroids for 3D imaging analysis) comprising some minor modifications. Organoids were collected at days 12, 30, and 90 and were washed in

PBS followed by fixation using 4% methanol-free paraformaldehyde (20 min at RT). Following multiple washes to remove the fixative agent, organoids were permeabilized (6 h at RT) under continuous rotation. Incubation with primary antibodies was performed for 2 days at 37 °C under rotation (antibodies listed in Table S3). To remove unbound antibodies, antibody staining solution was discarded and washed 5 times for 30 min, with slow continuous rotation. Secondary antibodies were diluted according to manufacturer's recommendations and added to each well (incubated for 1 h at RT). If required, DAPI staining was performed at this step (10 min at RT). Afterwards, antibody staining solution was discarded and replaced with fresh ones (30 min incubations at RT). These steps were repeated 5 times to ensure full removal of unbound antibodies. Organoids were afterwards embedded in 1% agarose in ddH₂O and following solidification were cut into approximately 2 × 2–5 × 5 mm pieces (depending on organoid size), each containing one or more organoids. Smaller agarose cubes facilitate imaging conditions, and many orientations ensure visualization of the entire organoid. Dehydration solutions were prepared by diluting absolute ethanol in sterile water to obtain 50 and 70% ethanol solutions. Up to ten embedded organoids were dehydrated with a series of ethanol dilutions in 1.5 mL tubes at RT under slow continuous rotation: 50% ethanol was incubated for 2 h, followed by 70% ethanol for 2 h and finally 100% ethanol overnight. The next day, clearing solution was added into 1.5 mL tubes containing dehydrated organoids, and incubated at RT under slow continuous rotation for a total of 6 h. Clearing solution was discarded and substituted with imaging solution to proceed with imaging acquisition and long-term storage (protected from light).

Sample acquisition and data analysis. Whole brain organoids were further visualized in a Yokogawa Spinning Disk Field Scanning Confocal System (Yokogawa CSU-W1 25–50 µm pinhole dual disk, Nikon, Japan), equipped with motorized stage x-y-z, and a Prime BSI camera (Teledyne Photometrics, Arizona, USA), in confocal mode. Four channel (DAPI, NESTIN, PAX6, and SOX2) z-stack images of the whole organoid (z-step intervals of 5 µm) were acquired using the Nikon NIS Elements AR software (version 5.02.03) at a 10x/0.3 magnification (dry, no binning). For counting cells actively undergoing mitosis in the two conditions, images were processed in the Arivis Vision 4D software (version 3.5.0, Arivis, Germany), using a fully automated custom pipeline. Briefly, to facilitate the downstream 3D workflow, DAPI and pHH3 channels were processed separately, and advanced image enhancement filters were applied. For volume measurements, the DAPI channel in each organoid was later segmented using the "Li" thresholder, and objects smaller than 10 000 µm³ were filtered out. For counting the number of pHH3-positive cells, the "Blob Finder" analysis operator was used to segment cells and the following parameters were applied: (i) averaged diameter of 6 µm; (ii) 5% probability threshold; (iii) 60% split sensitivity. Finally, raw data was processed, and the number of particles was normalized to DAPI volume. All data is presented as median of *n* different organoids for all analysis. For analysis of cytoarchitectures and classification of ventricle-like structures (VLS) in CTL- and GADEV5-derived organoids, VLS measurements were classified as "regular" or "irregular". Given the empirical/arbitrary logic of the analysis, in order to ensure bias-free annotation of "regular" and "irregular" VLS, individual files containing a single organoid were anonymized and scrambled (named with 5 random characters) before manual annotation. Each structure was eligible for classification in "regular VLS" if the following criteria were met: (i) structure is PAX6-positive; (ii) structure has clear PAX6-lined ventricle lumens; (iii) structures were sufficiently separated from neighboring structures to ensure accurate counting; (iv) truncated structures where we could not discern between start and end of VLS were not counted (verified with 3D rendering of each organoid). For classification of structures in "irregular VLS", the following premises include: (i) structure is PAX6-positive; (ii) structure lack well-defined lumen; (iii) structures were adequately apart, otherwise were not counted.

Morphometrical characterization of organoids

Organoid morphometric properties were assessed for each individual cell line and bright-field images were acquired using the OLYMPUS IX81-ZDC inverted microscope, equipped with a Hamamatsu ORCA-ER B/W CCD camera at a magnification of 4X/0.16 (dry) without binning. Image acquisition was fully automated, and analysis was performed using Fiji software (Fiji-ImageJ v2.1.0, USA). Two complementary macros were developed to first stitch multiple fields of view (see Macro 2) and afterwards perform the morphological analysis (see Macro 3). Two distinct methods were used to detect objects according to organoid density for each dataset. When handling crowded acquisitions, the plugin 'StarDist'

was used to automatically select each organoid. Otherwise, samples were segmented and converted into masks and objects were marked for analysis. Regardless of the selection method used, objects touching the edges in each field of view were discarded and analysis followed the same principles in which: (i) objects smaller than 25,00 µm² were discarded from the pipeline; (ii) contrast was enhanced in all fields of view; (iii) images were segmented and binarized; (iv) area, perimeter, and circularity were analyzed using Fiji's 'Analyze Particle' plugin; (v) data was saved on the region of interest (ROI) manager. The measured parameters correspond to the largest cross-section of each organoid.

Computational analysis

Bulk RNA-seq analysis. All bulk RNA-seq data were quantified by means of Salmon v1.4 [67], using reference GENCODE genome hg38 v35 (CRCh38.p13) or GENCODE assembly GRCm39 (release M27), if handling human or mouse datasets, respectively. We derived HUGO Gene Nomenclature Committee (HGNC) symbol gene names from the corresponding gene transfer format (GTF). Each bulk dataset was filtered using a pre-normalization threshold: only genes with at least 20 raw read counts in at least two biological replicates were kept. Pseudo-bulks were generated via *adpbulk* (github.com/noamteyssier/adpbulk), aggregating raw reads by cell type and by individual, and genes with at least 5 raw read counts in at least two individuals were filtered.

Differential expression analysis was performed using previously published "edg2" edgeR wrapper [68] for bulk RNA-seq and "edg1" for pseudo-bulk, respectively using *estimateGLMRobustDisp* or *estimateDisp* for dispersion estimation. FDR ≤ 0.05 and FC ≥ 1.25 were used as thresholds in bulk, and FDR ≤ 0.01 and FC ≥ 2 were used in pseudo-bulks. Gene Ontology (GO) enrichments were performed with topGO [69] and represented with internal scripts on RStudio (R) 4.1.0. enrichGO [70] was used on genes passing false discovery rate (FDR) ≤ 0.05 to generate dotplots and GO semantic networks. Transcription factor enrichments (also "master regulatory analyses" elsewhere) were performed by means of internal R scripts, based on (i) recursive hypergeometric tests performed at gene level, by comparing known targets of TF that have been derived from motifs and ENCODE ChIP-seq data using TFBS database [71], with differentially expressed genes (DEGs), using as background universe the expressed genes in the filtered count matrix; (ii) enrichment was calculated as follows: observations are set as the intersection of DEGs and targets of each TF; expected observations are measured as the multiplication of lengths of the two genesets, divided by the length of the universe; (iii) FDR is measured by Benjamini & Hochberg multiple-test correction.

HPO term enrichments were calculated with a hypergeometric test, using all expressed genes as universe. The *ggraph* R package (v.2.0.6) was used for visualization of the results.

Astrocytes network deconvolution was performed as follows: starting from differential expression, we inferred transcription factor activity by means of *viper* using the DoRothEA database [52]. We performed network analysis in Cytoscape [72] to calculate topological indexes of each node (i.e., DEG), such as degree (calculated as a sum of the weight of all connections of a given node), and betweenness centrality (the quantification of the number of times a node is a bridge along the shortest path between other two nodes), which were used as centrality indices. Fosb, Foxn4, Myb, Neurog2, Spi1 and Tnf were all differentially expressed transcription factor with high betweenness and centrality scores. We further applied a heat diffusion algorithm to measure network propagation as previously done [73] to verify that the majority of the network was met by these TFs. They were thus predicted to be responsible for differential expression patterns (i.e., master regulators) identified in astrocytes grown with patient-specific neurons.

ChIP-seq analysis. Chromatin Immunoprecipitation sequencing (ChIP-seq) reads were trimmed for library specific adaptor contamination before being aligned to the hg38 genome with Bowtie 2 (removing multi-mapping reads with *samtools*). We performed peak calling via MACS 2.1 using narrow settings for YY1 (-q 0.05). Peaks overlaps were performed by means of *bedtools* v2.28, and bound genes were defined as intersecting promoters (from 500 bp upstream to 250 bp downstream TSS), or intersecting enhancers (cell type specific 4D Genome consortium peak sets). Presence of peaks in exclusion lists from ENCODE was verified. Motif enrichments were performed with Homer. To perform motif enrichment analyses on bulk TF ChIP-seq we started from summit peaks, extended them by 100 bp on both sides of the summit using *BedTools*, intersected all control replicates and kept peaks found in at least 2 of them. Hg38 was

used as reference genome and Homer parameters `--mask` and `-size 200` were applied.

Tracks and Heatmaps were generated with DeepTools v3.5 [74]. Reference peaks were defined as present in at least 2 control samples. We identified “lost peaks” as regions preferentially found in CTLs (i.e., regions in at least $n = 1$ controls and none of patients, plus regions found in all controls and at most 1 patient). Peaks were considered “gained” following the inverse logic, as regions preferentially found in patients. Motif enrichments were performed with Homer v.4.11 using default parameters. Coverage heatmaps were generated using deepTools plotHeatmap on bamCoverage (RPGC normalization) or bamCompare (normalization on Inputs) outputs, as stated in the text.

Multiomic data preprocessing. Multiomic sequencing data was preprocessed using 10X CellRanger ARC [75, 76] and computationally demultiplexed based on individual-specific single nucleotide polymorphisms using SCanSNP [66]. The scRNA-seq data was analyzed using the Scanpy package [77] (v.1.7.2). The Matplotlib [78] (v. 3.4.145), and Seaborn [79] (v.0.11.146) packages were used for visualization. Pandas [80] (v1.2.4) and Numpy [81] (v1.22.3) were used for data handling.

Differential expression analysis between GADEVs and control cells in each cluster was performed by pseudo-bulk. Cells were grouped by cell line and subsampled to have a similar number of cells in each group before summing the raw counts. EdgR [82] (v.3.32.1) was then used to perform differential expression between the aggregated counts of the controls and the disease patients: first we normalized by trimmed mean of M values (TMM) using the CalcNormFactors function, then we obtained negative binomial (NB) dispersion estimates using estimateGLMRobustDisp, we applied a GLM likelihood ratio test (glmFit) and found genes differentially expressed between conditions (glmLRT, topTags). Sex was included as a covariate in the design matrix.

The topGO R package (v.2.42.0) was then used for Gene Ontology [83] term enrichment analysis. We tested for Biological Process, Molecular Function and Cellular Component Terms, using as background all genes expressed in our dataset. We applied a Benjamini-Hochberg p-value adjustment procedure and set a significance threshold of 0.1 on the FDR.

scATAC-seq data analysis. The ArchR [84] (v.1.0.2) R toolkit was used for the main steps of scATAC-seq data analysis. ggplot2 [85] (v.3.3.4) was used for visualization.

Fragments were filtered by length and counted in genome-wide 500-bp bins. After generating Arrow files and filtering cells based on scATAC-seq-specific quality metrics (see provided scripts), we intersected the cell barcodes with those that passed scRNA-seq QC. Normalization and dimensionality reduction were performed using Latent Semantic Indexing. A UMAP projection was computed for visualization. After confirming that cell types annotated in scRNA-seq clustered coherently on the scATAC-seq UMAP, the annotation was given as grouping, combined with sequencing batches, to run ArchR's iterative peak calling using MACS2 [86] (v.2.2.7). Normalization and dimensionality reduction steps were then recalculated on the resulting peak count matrix. We downloaded motif position frequency matrices from the JASPAR database [87] (2020 version) and added motif information to the ArchR object. We used ChromVar [88] (v. 1.16.0) to compute per-cell motif activity scores. We then tested for differential TF activity scores between GADEVs and control samples by applying a Wilcoxon Rank Sum Test and computing the average difference in z-score between the two groups.

Promoter-enhancer association. We used Cicero [89] (v. 1.8.1) to find cell type- and condition-specific co-accessible peaks. We generated a binarized peak accessibility count matrix for cells of each cell type, divided by genotype and gave these as input to Cicero, specifying the UMAPs generated from the LSI reductions as coordinates. For the Ngn2 dataset, Glia cells were aggregated to Progenitor clusters and Early Neurons to Glutamatergic Neurons. Additionally, for each cell type, the GADEVs and control groups of cells were downscaled to be of the same size, as we have previously observed that the number of connections predicted by Cicero is negatively correlated with the number of input cells. Only peaks with positive accessibility correlations (Cicero score greater than zero) were included in the following analyses.

We used a custom script to find putative cis-regulatory regions for each gene, starting with co-accessibility scores. We first annotated the scATAC-seq peaks based on their overlap with promoters, exons and 5' untranslated regions (annotation downloaded from the Ensembl database (ensemblldb R package, v. 2.14.1)).

We then constructed promoter-enhancer networks using the igraph R package (v. 1.2.6) for each celltype (divided by genotype), selecting only links that passed a co-accessibility threshold (0.4 for Ngn2, 0.6 for organoids).

Gene regulatory networks. Gene regulatory networks (GRNs) were built using CellOracle [42] (v. 0.10.12). All peaks in the previously generated promoter-enhancer association lists were scanned for the presence of TF motifs from the gimmemotifs database (gimme.vertebrate.v5.0). Base GRNs were then generated for each combination of cell type and genotype, connecting each TF to the genes whose enhancers or promoters have the TF binding motif. Using CellOracle, we then filtered the GRNs based on the celltype-specific expression of the genes, derived from our scRNA-seq datasets. Edges in the resulting networks are drawn based on the concurrence of the following criteria: presence of the source TF's binding motif in the gene promoter or its enhancers, co-expression of the TF and the gene, and the discernible impact of TF expression on the transcript levels of the gene, inferred by a Bagging Ridge ML model.

The GRNs were then processed using custom scripts and visualized with Cytoscape [72] (v.3.10.0), selecting the yFiles Orthogonal layout algorithm.

DATA AVAILABILITY

Processed data can be found at <https://zenodo.org/records/14225858> and supplementary tables at https://github.com/GiuseppeTestaLab/2024_YY1_Organoids.

CODE AVAILABILITY

All the code is available upon request.

REFERENCES

- Gabriele M, van Silfhout ATV, Germain PL, Vitriolo A, Kumar R, Douglas E, et al. YY1 haploinsufficiency causes an intellectual disability syndrome featuring transcriptional and chromatin dysfunction. *Am J Hum Genet.* 2017;100:907–25.
- Vissers LELM, de Ligt J, Gilissen C, Janssen I, Stehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. *Nat Genet.* 2010;42:1109–12.
- Ferng A, Thulin P, Walsh E, Weissbrod PA, Friedman J. YY1: a new gene for childhood onset dystonia with prominent oromandibular-laryngeal involvement? *Mov Disord Off J Mov Disord Soc.* 2022;37:227–8.
- Zorzi G, Keller Sarmiento IJ, Danti FR, Bustos BI, Invernizzi F, Panteghini C, et al. YY1-related dystonia: clinical aspects and long-term response to deep brain stimulation. *Mov Disord Off J Mov Disord Soc.* 2021;36:1461–2.
- Nabais Sá MJ, Gabriele M, Testa G, de Vries BB. Gabriele-de Vries Syndrome. *GeneReviews®.* Seattle (WA): University of Washington, Seattle; 2019.
- Woo H, Kim WS, Kim JS. A rare epilepsy phenotype in Gabriele-de Vries syndrome: a new case and literature review. *Neurol Asia.* 2023;28:1063–7.
- Brooker SM, Mencacci NE. The expanding genetic landscape of myoclonus-dystonia syndrome: YY1 and ATP1A3 are added to the list. *Parkinsonism Relat Disord.* 2023;117:105929.
- Donohoe ME, Zhang X, McGinnis L, Biggers J, Li E, Shi Y. Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol Cell Biol.* 1999;19:7237–44.
- Verheul TCJ, van Hijfte L, Perenthaler E, Barakat TS. The why of YY1: mechanisms of transcriptional regulation by Yin Yang 1. *Front Cell Dev Biol.* 2020;8:592164.
- Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, et al. YY1 is a structural regulator of enhancer-promoter loops. *Cell.* 2017;171:1573–88.e28.
- Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, et al. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* 2017;27:1139–52.
- Vitriolo A, Gabriele M, Testa G. From enhanceropathies to the epigenetic manifold underlying human cognition. *Hum Mol Genet.* 2019;28:R226–R234.
- 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.* 2018;84:306–27.
- Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R. Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet.* 2022;54:1919–32.
- Carminho-Rodrigues MT, Steel D, Sousa SB, Brandt G, Guipponi M, Laurent S, et al. Complex movement disorder in a patient with heterozygous YY1 mutation (Gabriele-de Vries syndrome). *Am J Med Genet A.* 2020;182:2129–32.
- López-Perrote A, Alatwi HE, Torreira E, Ismail A, Ayora S, Downs JA, et al. Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. *J Biol Chem.* 2014;289:22614–29.

17. Gabriele M, Vitriolo A, Cuvertino S, Pereira MF, Franconi C, Germain PL, et al. KMT2D haploinsufficiency in Kabuki syndrome disrupts neuronal function through transcriptional and chromatin rewiring independent of H3K4-monomethylation. *bioRxiv*:2021.04.22.440945 [Preprint]; 2021. Available from: <https://www.biorxiv.org/content/10.1101/2021.04.22.440945v1>
18. Adamo A, Atashpaz S, Germain PL, Zanella M, D'Agostino G, Albertin V, et al. 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat Genet*. 2015;47:132–41.
19. Birk E, Har-Zahav A, Manzini CM, Pasmanik-Chor M, Kornreich L, Walsh CA, et al. SOBP is mutated in syndromic and nonsyndromic intellectual disability and is highly expressed in the brain limbic system. *Am J Hum Genet*. 2010;87:694.
20. de Luis O, Valero MC, Jurado LA. WBSR14, a putative transcription factor gene deleted in Williams-Beuren syndrome: complete characterisation of the human gene and the mouse ortholog. *Eur J Hum Genet EJHG*. 2000;8:215–22.
21. Bauer CK, Schneeberger PE, Kortüm F, Altmüller J, Santos-Simarro F, Baker L, et al. Gain-of-Function mutations in KCNN3 encoding the small-conductance Ca²⁺-activated K⁺ channel SK3 cause zimmermann-laband syndrome. *Am J Hum Genet*. 2019;104:1139–57.
22. Mulley JC, Iona X, Hodgson B, Heron SE, Berkovic SF, Scheffer IE, et al. The role of seizure-related SEZ6 as a susceptibility gene in febrile seizures. *Neurol Res Int*. 2011;2011:917565.
23. Gunnersen JM, Kim MH, Fuller SJ, De Silva M, Britto JM, Hammond VE, et al. Seiz-6 proteins affect dendritic arborization patterns and excitability of cortical pyramidal neurons. *Neuron*. 2007;56:621–39.
24. Goto S, Nagahiro S, Ushio Y, Kitaoka M, Nishio S, Fukui M. Immunocytochemical detection of calcineurin and microtubule-associated protein 2 in central neurocytoma. *J Neurooncol*. 1993;16:19–24.
25. Hassoun J, Söylemezoglu F, Gambarelli D, Figarella-Branger D, von Ammon K, Kleihues P. Central neurocytoma: a synopsis of clinical and histological features. *Brain Pathol*. 1993;3:297–306.
26. Kobayashi K, Miyazu K, Katsukawa K, Fukutani Y, Mukai M, Nakamura I, et al. Cytoskeletal protein abnormalities in patients with olivopontocerebellar atrophy—an immunocytochemical and Gallyas Silver Impregnation study. *Neuropathol Appl Neurobiol*. 1992;18:237–49.
27. Myles-Worsley M, Tiobech J, Browning SR, Korn J, Goodman S, Gentile K, et al. Deletion at the SLC1A1 glutamate transporter gene co-segregates with schizophrenia and bipolar schizoaffective disorder in a 5-generation family. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet*. 2013;162B:87–95.
28. Ragge NK, Brown AG, Poloschek CM, Lorenz B, Henderson RA, Clarke MP, et al. Heterozygous mutations of OTX2 cause severe ocular malformations. *Am J Hum Genet*. 2005;76:1008–22.
29. Patat O, van Ravenswaaij-Arts C, Tantau J, Corsten-Janssen N, van Tintelen J, Dijkhuizen T, et al. Otocephaly-Dysgnathia complex: description of four cases and confirmation of the role of OTX2. *Mol Syndromol*. 2013;4:302.
30. Chassaing N, Sorrentino S, Davis EE, Martin-Coignard D, Iacovelli A, Paznekas W, et al. OTX2 mutations contribute to the otocephaly-dysgnathia complex. *J Med Genet*. 2012;49:373–9.
31. Nakajima J, Okamoto N, Tohyama J, Kato M, Arai H, Funahashi O, et al. De novo EEF1A2 mutations in patients with characteristic facial features, intellectual disability, autistic behaviors and epilepsy. *Clin Genet*. 2015;87:356–61.
32. Lam WWK, Millichap JJ, Soares DC, Chin R, McLellan A, FitzPatrick DR, et al. Novel de novo EEF1A2 missense mutations causing epilepsy and intellectual disability. *Mol Genet Genomic Med*. 2016;4:465–74.
33. Yang J, Yu C, Lyn N, Liu L, Li D, Shang Q. Clinical analysis of Gabriele-de Vries caused by YY1 mutations and literature review. *Mol Genet Genomic Med*. 2024;12:e2281.
34. Varum S, Baggiolini A, Zurkirchen L, Atak ZK, Cantù C, Marzorati E, et al. Yin Yang 1 orchestrates a metabolic program required for both neural crest development and melanoma formation. *Cell Stem Cell*. 2019;24:637–53.e9.
35. Cheroni C, Trattaro S, Caporale N, López-Tobón A, Tenderini E, Sebastiani S, et al. Benchmarking brain organoid recapitulation of fetal corticogenesis. *Transl Psychiatry*. 2022;12:520.
36. 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;25:558–69.e7.
37. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol*. 2019;37:1452–7.
38. Zhu C, Yu M, Huang H, Juric I, Abnousi A, Hu R, et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat Struct Mol Biol*. 2019;26:1063–70.
39. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*. 2018;361:1380–5.
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. Pebworth MP, Ross J, Andrews M, Bhaduri A, Kriegstein AR. Human intermediate progenitor diversity during cortical development. *Proc Natl Acad Sci USA*. 2021;118:e2019415118.
42. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. *Nature*. 2023;614:742–51.
43. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. *Nat Commun*. 2021;12:5692.
44. Moriano J, Leonardi O, Vitriolo A, Testa G, Boeckx C. A multi-layered integrative analysis reveals a cholesterol metabolic program in outer radial glia with implications for human brain evolution. *Dev Camb Engl*. 2024;151:dev202390.
45. Cheon SY. Impaired cholesterol metabolism, neurons, and neuropsychiatric disorders. *Exp Neurobiol*. 2023;32:57–67.
46. Dann E, Henderson NC, Teichmann SA, Morgan MD, Marioni JC. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol*. 2022;40:245–53.
47. Dimitrov D, Schäfer PSL, Farr E, Rodriguez-Mier P, Lobentanz S, Badia-i-Mompel P, et al. LIANA+ provides an all-in-one framework for cell–cell communication inference. *Nat Cell Biol*. 2024;26:1613–22.
48. Armingol E, Baghdassarian HM, Martino C, Perez-Lopez A, Aamodt C, Knight R, et al. Context-aware deconvolution of cell–cell communication with TensorCell2. *Nat Commun*. 2022;13:3665.
49. Lin HC, He Z, Ebert S, Schörrig M, Santel M, Nikolova MT, et al. NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Rep*. 2021;16:2118–27.
50. Moulson AJ, Squair JW, Franklin RJM, Tetzlaff W, Assinck P. Diversity of reactive astrogliosis in CNS pathology: heterogeneity or plasticity? *Front Cell Neurosci*. 2021;15:703810.
51. Escartin C, Galea E, Lakatos A, O'Callaghan JP, Petzold GC, Serrano-Pozo A, et al. Reactive astrocyte nomenclature, definitions, and future directions. *Nat Neurosci*. 2021;24:312–25.
52. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*. 2019;29:1363–75.
53. Matusova Z, Hol EM, Pekny M, Kubista M, Valihrach L. Reactive astrogliosis in the era of single-cell transcriptomics. *Front Cell Neurosci*. 2023;17:1173200.
54. Pereira MF, Shyti R, Testa G. In and out: Benchmarking in vitro, in vivo, ex vivo, and xenografting approaches for an integrative brain disease modeling pipeline. *Stem Cell Rep*. 2024;19:767–95.
55. Girtan JR, Jeon SH. Novel embryonic and adult homeotic phenotypes are produced by pleiohomeotic mutations in *Drosophila*. *Dev Biol*. 1994;161:393–407.
56. Kwon HJ, Chung HM. Yin Yang 1, a vertebrate polycomb group gene, regulates anterior-posterior neural patterning. *Biochem Biophys Res Commun*. 2003;306:1008–13.
57. Morgan MJ, Woltering JM, In der Rieden PMJ, Durston AJ, Thierry JP. YY1 regulates the neural crest-associated slug gene in *Xenopus laevis*. *J Biol Chem*. 2004;279:46826–34.
58. Khamirani HJ, Zoghi S, Namdar ZM, Kamal N, Dianatpour M, Tabei SMB, et al. Clinical features of patients with Yin Yang 1 deficiency causing Gabriele-de Vries syndrome: a new case and review of the literature. *Ann Hum Genet*. 2022;86:52–62.
59. Coulter DA, Steinhäuser C. Role of astrocytes in epilepsy. *Cold Spring Harb Perspect Med*. 2015;5:a022434.
60. Çarçak N, Onat F, Sitnikova E. Astrocytes as a target for therapeutic strategies in epilepsy: current insights. *Front Mol Neurosci*. 2023;16:1183775.
61. Lawrence JM, Schardien K, Wigdahl B, Nonnemacher MR. Roles of neuropathology-associated reactive astrocytes: a systematic review. *Acta Neuropathol Commun*. 2023;11:42.
62. Menendez L, Kulik MJ, Page AT, Park SS, Lauderdale JD, Cunningham ML, et al. Directed differentiation of human pluripotent cells to neural crest stem cells. *Nat Protoc*. 2013;8:203–12.
63. Pereira MF, Amaral IM, Lopes C, Leitão C, Madeira D, Lopes JP, et al. I-alpha-aminoadipate causes astrocyte pathology with negative impact on mouse hippocampal synaptic plasticity and memory. *FASEB J*. 2021;35:e21726.
64. Schildge S, Bohrer C, Beck K, Schachtrup C. Isolation and culture of mouse cortical astrocytes. *J Vis Exp*. 2013; <https://doi.org/10.3791/50079>.
65. Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron*. 2013;78:785–98.
66. Caporale N, Castaldi D, Rigoli MT, Cheroni C, Valenti A, Stucchi S, et al. Multiplexing cortical brain organoids for the longitudinal dissection of developmental traits at single-cell resolution. *Nat Methods*. 2024;22:358–70.

67. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14:417–9.
68. Zanella M, Vitriolo A, Andirko A, Martins PT, Sturm S, O'Rourke T, 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*. 2019;5:eaaw7908.
69. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinforma Oxf Engl*. 2006;22:1600–7.
70. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*. 2021;2:100141.
71. Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinforma Oxf Engl*. 2016;32:1555–6.
72. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504.
73. Iannuccelli M, Vitriolo A, Licata L, Lo Surdo P, Contino S, Cheroni C, et al. Curation of causal interactions mediated by genes associated with autism accelerates the understanding of gene-phenotype relationships underlying neurodevelopmental disorders. *Mol Psychiatry*. 2024;29:186–96.
74. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42:W187–191.
75. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8:14049.
76. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol*. 2019;37:925–36.
77. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. 2018;19:15.
78. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng*. 2007;9:90–5.
79. Waskom ML. seaborn: statistical data visualization. *J Open Source Softw*. 2021;6:3021.
80. McKinney W. Data structures for statistical computing in python. *Proc 9th Python Sci Conf*. 2010;56–61.
81. Harris, Millman CR, van der Walt KJ, Gommers R SJ, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020;585:357–62.
82. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma Oxf Engl*. 2010;26:139–40.
83. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOid mine. *Nucleic Acids Res*. 2021;49:D325–34.
84. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet*. 2021 Mar;53:403–11.
85. Wickham H. ggplot2 [Internet]. Cham: Springer International Publishing; 2016. (Use R!).
86. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.
87. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2022;50:D165–73.
88. Schep AN, Wu B, Buenostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods*. 2017;14:975–8.
89. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell*. 2018;71:858–71.e8.

ACKNOWLEDGEMENTS

We thank all individuals affected with GADEVs and their parents for donating samples, Dr. Daniele Capocéfalo for feedbacks on network analysis, we thank all European Institute of Oncology and Human Technopole facilities.

AUTHOR CONTRIBUTIONS

MFP, LR, DA, ET, RS, and MG conducted experiments. AV, VF, VA, LM, MR, FD, and DC analyzed data. MFP, AV, MG, and CS planned experiments. BV, MTC-R, and JFB provided samples. MFP, VF, AV, MG, wrote the first draft of the manuscript. All authors contributed and approved the final manuscript. MG, AV, GT supervised and conceived the study.

FUNDING

This work was supported by Horizon 2020 Innovative Training Network EpiSyStem (GT), H2020 Marie Skłodowska-Curie Actions, Grant/ Award Number: 765966 (MFP), Associazione Italiana per la Ricerca sul Cancro (AIRC) “Fellowships for Italy” (MG), American-Italian Cancer Foundation (AICF) research fellowship (MG), NIH K99GM149815 (MG). E-Rare IMPACT consortium (GT), Telethon Research Grant GGP19226 (GT). Figures were created with Biorender.com.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All methods were performed in accordance with the relevant guidelines and regulations. This study was approved by the institutional review board of the University of Milan ethics committee. Written informed consents were obtained for all individuals. All animal experiments were done in accordance with the Italian Laws (D.L.vo 116/92 and following additions), which enforces EU 86/609 Directive (Council Directive 86/609/EEC of 24 November 1986), and were approved by institutional ethics committee (organismo preposto al benessere degli animali, OPBA committee) and the Italian Ministry of Health (Authorization 1073/16-PR).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41380-025-02929-x>.

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025