

RESEARCH

Open Access



Comprehensive multi-omics integration identifies differentially active enhancers during human brain development with clinical relevance

Soheil Yousefi¹, Ruizhi Deng^{1†}, Kristina Lanko^{1†}, Eva Medico Salsench^{1†}, Anita Nikoncuk^{1†}, Herma C. van der Linde¹, Elena Perenthaler¹, Tjakko J. van Ham¹, Eskeatnaf Mulugeta^{2*} and Tahsin Stefan Barakat^{1*} 

Abstract

Background: Non-coding regulatory elements (NCREs), such as enhancers, play a crucial role in gene regulation, and genetic aberrations in NCREs can lead to human disease, including brain disorders. The human brain is a complex organ that is susceptible to numerous disorders; many of these are caused by genetic changes, but a multitude remain currently unexplained. Understanding NCREs acting during brain development has the potential to shed light on previously unrecognized genetic causes of human brain disease. Despite immense community-wide efforts to understand the role of the non-coding genome and NCREs, annotating functional NCREs remains challenging.

Methods: Here we performed an integrative computational analysis of virtually all currently available epigenome data sets related to human fetal brain.

Results: Our in-depth analysis unravels 39,709 differentially active enhancers (DAEs) that show dynamic epigenomic rearrangement during early stages of human brain development, indicating likely biological function. Many of these DAEs are linked to clinically relevant genes, and functional validation of selected DAEs in cell models and zebrafish confirms their role in gene regulation. Compared to enhancers without dynamic epigenomic rearrangement, DAEs are subjected to higher sequence constraints in humans, have distinct sequence characteristics and are bound by a distinct transcription factor landscape. DAEs are enriched for GWAS loci for brain-related traits and for genetic variation found in individuals with neurodevelopmental disorders, including autism.

Conclusion: This compendium of high-confidence enhancers will assist in deciphering the mechanism behind developmental genetics of human brain and will be relevant to uncover missing heritability in human genetic brain disorders.

Keywords: Clinical genetics, Non-coding genome, Enhancer, Gene regulatory elements, Meta-analysis, Data integration, Epigenome, Computational analysis, Mendelian disorders, Human brain development

* Correspondence: e.mulugeta@erasmusmc.nl; t.barakat@erasmusmc.nl

[†]Ruizhi Deng, Kristina Lanko, Eva Medico Salsench and Anita Nikoncuk contributed equally to this work.

²Department of Cell Biology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

¹Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Non-coding regulatory elements (NCREs), such as enhancers, play a pivotal role in gene regulation [1, 2]. Enhancers ensure correct spatio-temporal gene expression, and it is increasingly recognized that genetic aberrations disturbing enhancer function can lead to human disease, including brain disorders [3–6]. Such non-coding genetic variants are expected to explain a considerable fraction of so-called missing heritability (e.g., the absence of a genetic diagnosis despite a high genetic clinical suspicion). These developments are pushing genetic diagnostic investigations to shift from whole-exome sequencing to whole genome sequencing, and the number of potentially pathogenic non-coding variants found in patients is expected to rise [4]. It is therefore of urgent clinical interest to understand where functionally relevant non-coding sequences are located in the human genome, as this will help to interpret the effects on health and disease.

Despite tremendous progress over the last decades, our understanding of the underlying mechanisms of enhancer biology remains limited due to challenges in annotating functional enhancers genome-wide. Large-scale community-driven efforts [7–11] and an uncountable plethora of individual studies have produced a vast amount of epigenome data sets, such as profiles of histone modifications, chromatin accessibility, and chromatin interactions for different human tissues and cell types, that can be used to predict putative enhancers at a large scale. More recently, new technologies such as massively parallel reporter assays and CRISPR-Cas9-based screens have entered the stage [12–14], providing novel means to directly test the functionality of non-coding regions. In addition, computational prediction algorithms [15, 16], trained on epigenome and experimental data, are improving the capability to predict functional sequences and the effects of variants in these regions.

One of the inherent problems with this increasing amount of data is the difficulty in keeping track of individual data sets and the ability to integrate data from various sources. Usually, individual studies focus on a limited number of cell types or tissues and compare their findings to a small number of previously published data sets. Although this is a logical step, it does not leverage the potential to fine-tune enhancer predictions which integrating all available enhancer data could have. This is illustrated by our previous findings that the overlap between individual enhancer predictions from several studies tends to be quite poor [4]. This is likely caused by heterogeneity of starting biological samples, limitations of current technologies, and differences in data analysis. Although the first two are difficult to change, analyzing

these data in a similar way could avoid some of the noise and difference generated by data analysis.

Here we undertook such an integrative effort, focusing on human brain development (Fig. 1A, Additional file 1: Fig. S1). We retrieved virtually all previously published putative enhancers for brain (from PubMed and enhancer databases, $n = \sim 1.6$ million putative enhancers) (Additional file 2: Table S1) [9, 11, 21–31], and performed an integrative analysis of relevant available epigenome data sets ($n = 494$) [9, 10, 19, 31–35] (Additional file 3: Table S2), after reanalyzing the data. Using this approach, we identify around 200 thousand putative critical regions (pCRs) in reported brain enhancers, of which around 40 thousand show dynamic epigenomic rearrangement during fetal brain development, indicating switching on and off of regulatory elements during development. We thus refer to these regions as differentially active enhancers (DAEs). Compared to their non-variable counterparts (nDAEs), DAEs have a higher level of sequence constraint, regulate genes that are expressed during fetal brain development and are associated with brain developmental processes. DAEs are enriched for binding sites of brain-relevant transcription factors, brain-related GWAS loci and are regulating disease-relevant *Online Mendelian Inheritance in Man* (OMIM) genes. We validate a selected number of DAEs using *in vitro* reporter assays and CRISPRi in cell lines, and reporter assays during zebrafish development. Together, this provides an easily accessible and comprehensive resource of NCREs that are likely functional during human brain development.

Methods

Data visualization

To visualize enhancers and epigenome data, we used the UCSC Genome Browser (<https://genome.ucsc.edu/>). To generate UCSC Genome Browser Tracks, aligned reads were converted to bedgraph using genomeCoverageBed, after which the bedGraphToBigWig tool from the UCSC Genome Browser was used to create a bigwig file [36, 37]. All enhancer regions, enhancer-gene interactions, and topologically associating domain (TAD) coordinates were uploaded directly as bed files. Other plots were drawn using R packages and Figs. 1, 2, 3, 4, 5, and 6 and Additional file 1: Figures S1–8 were assembled in Adobe Illustrator [47]. Additional file 2: Table S1, Additional file 3: Table S2, Additional file 4: Table S3, Additional file 5: Table S4, Additional file 6: Table S5, Additional file 7: Table S6, Additional file 8: Table S7, Additional file 9: Table S8, Additional file 10: Table S9, Additional file 11: Table S10, Additional file 12: Table S11, Additional file 13: Table S12, Additional file 14: Table S13 and Additional file 15: Table S14 were exported as text or Excel files.

Data collection and processing

Collection of putative brain enhancers

To generate a comprehensive set of putative brain enhancers active during fetal brain development, we scrutinized PubMed and various enhancer databases (last assessed: April 2019), including amongst others EnhancerAtlas, the FANTOM5 Project, and the Vista Enhancer database [9, 11, 21–31]. This resulted in 1,595,292 putative enhancers (Additional file 2: Table S1). Enhancers with identical coordinates were deduplicated and the unique regions were used to determine putative critical regions (pCRs), reasoning that overlapping parts of a putative enhancer obtained from different sources might point to functional relevant regions of that putative enhancer. If there is any overlap between coordinates of putative enhancers derived from two or more databases, the pCRs were defined as maximum overlapping regions present in those databases using the BEDtools suite (mergeBed, intersectBed, genomeCoverageBed, and groupBy sub-commands) (version 2.30.0) [36]. Putative enhancers that were only present in one of the input sources were also included in the pCRs (Fig. 1A, step 1), as it cannot be excluded that these putative enhancers are biologically relevant. pCRs with length less than 50 bp and more than 1000 bp were excluded. To avoid any overlap with gene promoters, enhancers located within 2 kb upstream or 1 kb downstream of a transcriptional start site (TSS) (Ensembl GRCh37.p13 Release 102) were excluded using intersectBed. Following this procedure, we identified a total of 202,462 pCRs which were used for downstream analyses. Next, we excluded 299 pCRs that were not covered by sufficient amounts of epigenome data (less than 10 reads in at least two samples (see section on defining DAEs)), resulting in a final number of 202,163 pCRs (Additional file 4: Table S3). GREAT web interface was used (version 4.0.4) (<http://great.stanford.edu/public/html/>) [17] to visualize enhancer-TSS distance (with *basal plus extension, proximal 5 kb upstream and 1 kb downstream, plus distal up to 100 kb, including curated regulatory domains, and whole genome (GRCh37/hg19) as background* parameters) (Additional file 1: Fig. S2B).

Epigenome data

Epigenome data were collected from the Roadmap Epigenomics Consortium, ENCODE, PsychENCODE, and other studies (Additional file 3: Table S2). Epigenome data sets used for integration included histone modifications (H3K27ac, H3K27me3, H3K4me1, H3K4me2, H3K4me3) and chromatin accessibility (ATAC-seq and DNase-seq) from different brain regions and different human developmental stages (Fig. 1A, step 1). To avoid any possible confounding biases because of the various pipelines used in different studies, we reanalyzed the raw

FASTQ files using our analysis pipeline (Additional file 1: Fig. S1). First, adaptor contamination was removed using Trim Galore (version 0.6.5) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and trimmed data were aligned to the GRCh37/hg19 human genome using Bowtie2 aligner (version 2.4.2) (with *--very-sensitive* parameter) [48]. Only properly paired and uniquely mapped reads, with mapping quality more than 30 ($\text{MAPQ} \geq 30$), were kept followed by removing any possible duplicated reads using Picard's MarkDuplicates (version 4.0.1.1) (<http://broadinstitute.github.io/picard/>). These reads were used to define differentially active enhancers (DAEs).

Defining differentially active enhancers (DAEs)

We assumed that pCRs with high variability in different epigenome data (dynamic epigenomic rearrangement) across different developmental stages are more likely to be functional than other pCRs. To determine this variability, the number of overlapping reads (for each epigenome mark) with pCRs was counted using the multiBamCov sub-command of BEDtools and a matrix was generated that included enhancers as rows and epigenome features as columns. Epigenome features were from different brain regions and developmental stages. In total, 299 pCRs with less than 10 reads were excluded, leaving 202,163 pCRs for this analysis. Subsequently, the raw read count matrix was normalized using TMM-normalization [49] and the normalized count matrix was used to define DAEs across different developmental stages and brain regions of a given epigenome data using edgeR (version 3.32.1) [50]. Since there were different developmental stages (time-point factor) and brain regions (brain part factor) in each epigenome data, a design matrix was generated for each factor separately. A limited number of samples without biological replicates were grouped together with other samples based on high correlation (Pearson correlation; $r > 0.89$). The DAEs were defined based on each design matrix using a generalized linear model and quasi-likelihood F-tests. In order to define the final DAE list, DAEs identified from at least two epigenome data-specific matrices were pooled. In total, this resulted in 39,709 DAEs (FDR adjusted p value < 0.05). The remaining 162,454 pCRs that did not show variability were considered as nDAEs (Additional file 4: Table S3).

Identifying chromatin interactions

Enhancer-gene interactions

In order to define enhancer-gene interaction, published HiC data from 3 human fetal brains, for cortical plate (CP) and germinal zone (GZ) at gestation weeks 17–18 were used [30]. This data provides 10 kb resolution bins for gene loop interactions and 40 kb resolution for

TADs. Pre-calculated significant interactions were intersected with pCRs (DAEs and nDAEs) using intersectBed to define gene-enhancer interaction for both CP and GZ separately. Out of the 202,163 pCRs, 41,041 pCRs engaged in 101,366 interactions in CP, and 41,085 pCRs had 100,521 interactions in GZ. Enhancer-gene interactions locating within the same TAD were considered for downstream analyses (almost 80% of all interactions were intra-TAD). We only included protein-coding and lincRNA genes in our analysis. To determine enhancer-enhancer interactions in Additional file 1: Fig. S7A, we also intersected HiC data with pCRs, focusing on interactions between DAEs and both DAEs and nDAEs.

In addition to HiC, we employed other enhancer-gene interaction predictions including JEME (<http://yiplab.cse.cuhk.edu.hk/jeme/>) [51], ENCODE (<https://ernstlab.biolchem.ucla.edu/roadmaplinking/>) [52], FOCS (<http://acgt.cs.tau.ac.il/focs/download.html>) [53], and GeneHancer (downloaded from UCSC table browser; hg19; updated 2019) [54]. These databases apply statistical models on different types of omics data to predict enhancer-gene interactions. We collected fetal brain enhancer-gene predictions from JEME and ENCODE and all brain-related enhancer-gene predictions from FOCS and GeneHancer, as the latter two resources do not specify fetal-specific interactions. In addition, we used H3K27ac HiChIP derived chromatin interactions from several postnatal brain regions [55] cell type-specific chromatin conformation capture data from PLAC-seq experiments in postnatal brain tissue [56] and enhancer-gene interaction predictions generated by the Activity-by-contact (ABC) model (<https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>) [57]. We performed the ABC model by fixing the length of pCRs to 500 bps from the center (250 bps from each side). The enhancer activity was then determined considering DNase, and H3K27ac samples, and gene expression data from fetal brain [10] using default settings of the “run.neighborhoods.py” function. The ABC score was calculated by integrating the fetal HiC data and enhancer activity defined using the default settings of the “predict.py” function and adjusting “--hic_type bedpe,” “--hic_resolution 10000” flags, and ignoring “--cellType” flag.

Intersections between the pCRs and each of these predictions were considered as enhancer-gene interaction (Additional file 6: Table S5). The coordinates of the HiChIP interactions were lifted over to hg19 before intersecting with pCRs.

Functional enrichment analysis

Enhancer sequence characteristics analysis

To determine whether different DNA sequence features distinguish different enhancer groups and whether there

is any association between these features and functional prediction, we considered the following features: (i) the non-coding essential regulation (ncER) score (https://github.com/TelentiLab/ncER_datasets/; updated 06-03-2019) [38]; (ii) GC content, as determined by the GCcontent R packages based on BSgenome. Hsapiens.UCSC.hg19 (version 1.4.3); (iii) conservation score for each enhancer, as derived from the gscores R packages based on phastCons100way.UCSC.hg19 (version 3.7.2) [39]; (iv) Orion scores [40]; (v) CADD scores [41]; (vi) Haploinsufficiency scores [43], and (vii) probability of loss-of-function intolerance (pLI) score [42]. The overlaps between DNA sequence features and enhancer coordinates were defined using intersectBed. As assessed enhancers (e.g., pCRs) varied in length between 50 and 1000 bp, and the abovementioned scores were given either at the nucleotide level or in certain bins (depending on the given scores from the individual resources), we calculated the mean value for each enhancer and used this in group comparisons. For gene-specific scores (e.g., pLI), we plotted the scores of the genes linked to the enhancers. Statistical significant differences between groups were determined using Wilcoxon signed rank test in R.

Gene expression correlation

To compare gene expression levels of enhancer target genes between different groups, various transcriptome data were collected. This included transcriptome data from different brain regions and developmental stages, and also various control data from other fetal tissues from the Roadmap Epigenomics Consortium, ENCODE project, Allen human brain atlas, and other studies (Additional file 7: Table S6) [7, 10, 18–20]. Raw data (FASTQ) was quality controlled and adaptors and other contaminants were removed using Trim Galore (version 0.6.5), reads were mapped to the GRCh37/hg19 human genome assembly using STAR aligner (version 2.7) [58], and gene counts were obtained using htseq-count (version 0.12.4) [59]. Gene expression levels were normalized based on fragments per kilobase of transcript per million mapped reads (FPKM). To correlate enhancers to gene expression, enhancer-gene interactions were derived from the HiC data or the alternative enhancer-gene predictions as described above. Gene expression levels were plotted and statistical comparison was performed, between expression levels of subgroups, using Wilcoxon signed rank test in R. We also compared genes linked to DAEs and nDAEs by HiC, to the three trajectory gene groups from BrainVar [60]. For this, we first found the overlap between genes interacting with DAEs/nDAEs using HiC-CP/GZ and each of the three trajectory groups (e.g., falling, rising, and constant genes). We then determined the odds ratio between

DAE and nDAE linked genes for each of the three gene trajectories, and used Fisher's exact test to determine significance.

Gene ontology analysis

For functional enrichment analysis, we used GREAT [17], Enrichr [61], and Metascape [62]. GREAT was used via the web interface (version 4.0.4) (<http://great.stanford.edu/public/html/>) using the following settings: basal plus extension, proximal 5 kb upstream and 1 kb downstream, plus distal up to 100 kb, including curated regulatory domains, and either whole genome or all pCRs as background, as indicated in the tabs of Additional file 5: Table S4. The $-\log_{10} p$ value was used to rank GREAT enrichment. Enrichr and Metascape were also used via the web interface (<https://maayanlab.cloud/Enrichr/>; <https://metascape.org/gp/index.html#/main/step1>), using the default settings and the whole genome set as background. All outputs of p value, adjusted p value (q value), and combined score (which is the estimation of significance based on the combination of Fisher's exact test p value and z score deviation from the expected rank) for Enrichr and of $\text{Log}P$, enrichment, z score, and $\log(q$ value) for Metascape are reported in Additional file 5: Table S4, Additional file 8: Table S7 and Additional file 10: Table S9.

Transcription factor binding enrichment

We used LOLA [63] using default settings to assess binding of known transcription factors to DAEs and nDAEs (Fig. 3). We used motifs from the JASPAR motif database (using reference genome GRCh37/hg19 and LOLAJaspar database core), to test the TF enrichment in DAEs and nDAEs, using all pCRs as background. The mean rank index (a combination of p value, odds ratio, from Fisher's exact test and the raw number of overlapping regions), was used to rank the known motifs. To display TF enrichment in Fig. 3C and F, we re-scaled the rank between 0 and 100 using the `rescale()` R function. To further identify motifs across the different relative DAE or nDAE bins and distinguish motifs in the central versus peripheral parts of the enhancers (Fig. 3G, H), we split the 100 relative bins into 20 groups of 5 consecutive bins and performed motif enrichment analysis using HOMER (version 4.11) [64], using function "findMotifs-Genome.pl" and all pCRs as background. A p value ≤ 0.01 was considered to select significantly enriched motifs.

Transposable element enrichment

The RepeatMask (GRCh37/hg19, updated 20-02-2020) was downloaded from the UCSC table browser and joined to the pCRs. To determine enrichment of transposable elements in brain enhancers, we followed a

strategy previously used when investigating active enhancers in human embryonic stem cells [65]. The number of overlaps of each type of repeat (n_{overlaps}) with all pCRs (n) was used to calculate the relative frequency ($f_{\text{all}} = n_{\text{overlaps}}/n$). Multiplication of the relative frequency with the number of regions (n_{test} , e.g., DAE, nDAE) in any tested group yields the expected frequency (E). This number was compared with the actual observed frequency in the subgroups ($f_{\text{test}} = (n_{\text{overlap, test}}/n_{\text{test}}) = O$) to calculate the observed versus expected ratio (O/E). We considered repeats with $O/E < 0.5$ as depleted, or $O/E > 1$ as enriched. For the subsequent data interpretation we only focused on transposable elements that were present multiple times ($n_{\text{overlap}} > 15$) in all pCRs (Additional file 9: Table S8).

Disease-relevance enrichment

The *Online Mendelian inheritance in Man* (OMIM) gene list (updated 28-09-2020) was downloaded using biomaRt R package [66] from Ensembl GRCh37.p13 Release 101. The GWAS catalog (GRCh37/hg19, updated 17-03-2021) was downloaded from the UCSC table browser. The GWAS catalog was manually filtered to keep brain-related studies and their variants with p value $\leq 9e-08$ (Additional file 12: Table S11). Stratified LD score regression analysis was performed by implementing the full baseline model to calculate enrichment (<https://github.com/bulik/ldsc/wiki> [67, 68]). Annotation and LD score files were created using the "make_annot.py" and "ldsc.py" functions, respectively. Partitioning heritability was performed using the "ldsc.py" script considering default parameters with "-- h2" flag. We obtained GWAS summary statistics for several brain-related traits including Alzheimer's disease [69], anorexia nervosa [70], anxiety [71], attention deficit hyperactivity disorder [72], autism spectrum disorder [73], bipolar disorder and schizophrenia [74], epilepsy [75], insomnia [76], intelligence [77], major depressive disorder [78], neuroticism [79, 80], obsessive compulsive disorder / Tourette syndrome [81], Parkinson's disease [82], and schizophrenia [83] (Additional file 12: Table S11). Z -scores were used to calculate the p values which were corrected for multiple hypothesis testing using the Benjamini-Hochberg method. For CNV analysis, we retrieved pre-processed published data from Brandler et al. (their supplemental Table 9: *de_novo_SVs* sheet, and their supplemental Table 7: *Primary CR Trans and Replication CR Trans* sheets) [44] and Monlong et al. (*cnvs-PopSV-Epilepsy-198affected-301controls-5 kb.tsv.gz* file in <https://figshare.com/s/20dfdedcc4718e465185>) [45]. For SNV analysis of the ASD simplex families, we collected *de novo* variants from supplemental Table 1 of Zhou et al [46]. Autism genes were collected from the SFARI Gene database (<http://gene.sfari.org/database/>

human-gene) [84]. The overlap between enhancer regions (DAE and nDAE) and each data set was determined using intersectBed. The odds ratio and *p* value between DAE and nDAE was calculated using fisher.test () R function. The Haldane–Anscombe correction was used to adjust the odds ratio.

Distribution of features across enhancer bins

To investigate the distribution of enrichment of different features (ncER score, GC content, phastCons score, and epigenome data) across enhancers, we divided the enhancer regions into 10 bp bins and calculated the relative scores (the median value for ncER score, GC content, phastCons score) and the number of reads (for epigenome data) for each bin. As the enhancers under investigation differed in size between 50 and 1000 bp, to make enrichments between enhancers comparable, we re-scaled each enhancer bin. To this end, we calculated a relative position between 1 and 100 for each bin of each enhancer, where 1 is the first bin, and 100 is the last bin of each individual enhancer. We then plotted the distribution of each feature across all these re-scaled enhancer bins.

DAE clustering analysis

The matrix of DAEs was used to determine the pattern of epigenome data through different developmental stages. To determine the optimal clustering algorithm, we used cValid R package which simultaneously compares multiple clustering algorithms (hierarchical, kmeans, model-based, pam, and clara). Based on this, the pam algorithm (which is similar to *k*-means but more robust to noise and outliers) was selected to cluster DAEs using the Spearman distance and ward.D² method. To define the optimal number of clusters, we used fviz_nbclust and NbClust R packages which compute different indices by bootstrapping (*n* = 1000). The predicted number of clusters was tested using the silhouette R package to examine whether the clustering performed correctly. This approach resulted in 2 clusters for DAEs and epigenome features at 8–12 PCW, 3 clusters for 13–18 PCW, and 2 clusters for > 18 PCW, for each of CP and GZ, respectively. For each cluster, we determined the gene expression of protein-coding genes interacting with the DAEs from each cluster, as obtained from published RNA-seq data sets. Significant differences in expression levels between different clusters were determined using the Wilcoxon signed rank test in R. Also, target genes linked to each cluster were used for functional enrichment analysis using Enrichr [61], as described under gene ontology analysis (Additional file 10: Table S9).

Enhancer cell type specificity and their dynamics in adult brain

To determine cell type specificity of enhancers, we compared DAEs and nDAEs to recently described cell type-specific regulatory elements from two studies on adult brain (obtained from Supplementary Data Set 4 (data lifted over to hg19) of Corces et al. [55] and Supplementary Table 5 of Nott et al [56]) and a study of fetal brain (obtained from Supplementary file 4 of Domcke et al., specificity scores for top 10000 regions [85]). We used bedtools to intersect DAEs or nDAEs and different cell type-specific regulatory elements. For all DAEs and nDAEs linked to target genes in CP and GZ by HiC, we compared dynamics of H3K27ac levels in both fetal and adult samples, using H3K27ac data from Li et al. [86]. Clustering analysis was performed as described under “DAE clustering analysis” above. Gene ontology analysis for each defined cluster was performed using Enrichr, as described above.

Experimental validation

Cell culture

HEK293 LTV cells (Cell Biolabs) were cultured in DMEM medium (Gibco), supplemented with 10% FBS at 37 °C, 5% CO₂. Human neural stem cells (NSCs) (Gibco) were cultured in NSC medium (KnockOut DMEM-F12 (Gibco), 2 mM L-glutamine (Gibco), 20 ng/ml bFGF (Peprotech), 20 ng/ml EGF (Peprotech), 2% StemPro Neural supplement (Gibco), 100 U/ml penicillin and 100 µg/ml streptomycin), as previously described [87].

Enhancer activity in STARR-seq reporter plasmids

For experimental validation in Fig. 1G, we randomly selected 22 DAEs that showed interaction with a target gene by HiC, and of which the target gene was expressed in neural stem cells, as indicated from our previously generated RNA-seq data (GSE137129 [87]); DAEs were amplified from genomic DNA and cloned into the STARR-seq plasmid (kind gift of A.Stark) [88] as previously described [65]. For the additional tested enhancer deletions (Additional file 1: Fig. S4), the obtained STARR-seq plasmids containing *IRF2BPL*, *CHD2*, and *MACF1* enhancers were modified by site-directed mutagenesis to remove regions with high or low ncER score. The following regions were deleted: *IRF2BPL* (chr14: 77422484-77422514); *CHD2* (ncER1 chr15: 93363603-93363640, ncER3 chr15: 93363780-93363790); *MACF1* (ncER1 chr1: 39598824-39598844, ncER2 chr1:39598744-39598754). The regions with low ncER score at the 5' and 3' ends (80–100 bp) of *IRF2BPL*, *CHD2*, and *MACF1* enhancers were excluded by Gibson assembly. Primer sequences are provided in Additional file 15: Table S14. HEK293 and NSC were transfected with STARR-seq plasmid containing enhancer regions using polyethylenimine (PEI, Sigma) or

Lipofectamine™ Stem Transfection Reagent (Thermo Scientific) respectively. Spike-in of a pmCherry-N1 plasmid (Clontech) was used as a transfection control. Twenty-four hours post transfection cells were collected, stained with Hoechst dye and the enhancer activity was measured by FACS analysis (20,000 cells per sample). GFP-positive cells within the mCherry-positive population were quantified to assess enhancer activity compared to an empty STARR-seq vector. Two independent transfection experiments were performed, each in duplicates. Statistical analysis was performed using a one-way ANOVA test followed by multiple comparison test (Fisher's LSD test). Calculations were conducted in GraphPad Prism (version 8).

dCas9-KRAB-MeCP2 silencing of active enhancers in NSC

We selected DAEs linked to *CHD2*, *CAD*, and *TRAK1* and designed for each DAE two targeting gRNAs (primer sequences are given in Additional file 15: Table S14). gRNAs were cloned into a pgRGFP plasmid (Addgene #82695, a kind gift of Allan Mullen) [89]. NSCs were co-transfected with dCas9-KRAB-MeCP2 (Addgene #110824, kind gift of Alejandro Chavez and George Church) [90] and the two gRNAs/DAE and collected for RNA isolation 48 h post transfection. Transfection efficiency was estimated by FACS analysis (78–92% GFP-positive cells detected). RNA was isolated using TRI reagent (Sigma) followed by cDNA preparation using iSCRIPT cDNA synthesis kit (Bio-Rad). Fold change in gene expression ($\Delta\Delta\text{ct}$ method) was evaluated by qPCR (iTaQ universal SYBR Green Supermix) (Sigma), performed in CFX96RTS thermal cycler (Bio-Rad), as previously described [87]. TBP expression was used as housekeeping normalization control. Statistical analysis was performed using a one-way ANOVA test followed by multiple comparison test (Fisher's LSD test). Calculations were conducted in GraphPad Prism (version 8).

Zebrafish studies

Zebrafish (*Danio rerio*) were raised and maintained under standard conditions [91]. Adult and larval fish were kept on a 14 h/10 h light–dark cycle at 28 °C. Larvae were kept in HEPES-buffered E3 medium. Media was refreshed daily, and at 24 hpf, 0.003% 1-phenyl 2-thiourea (PTU) was added to prevent pigmentation. All zebrafish experiments were performed in compliance with Dutch animal welfare legislation. Selected DAEs used in the in vitro experiments were transferred by Gibson assembly between the *AscI* and *PacI* site of a E1b-GFP-Tol2 enhancer assay plasmid (a kind gift of Ramon Birnbaum) [92] containing an E1b minimal promoter followed by GFP, using the following transfer primers: Transfer_fw: 5'-AGATGGGCCCTCGGGTAG

AGCATGCACCGG-3' and Transfer_rv: 5'-TCGAGA GATCTTAATGGCCGAATTCGTCTGA-3'. Constructs were injected into zebrafish embryos using standard procedures, together with Tol2 mRNA to facilitate genomic integration. At least 50 embryos were injected per construct in at least two different injection experiments. GFP expression was observed and annotated at 1, 2, and 3 dpf by a fluorescent Leica M165FC stereomicroscope (Additional file 14: Table S13). Images were analyzed using ImageJ (FIJI). An enhancer was considered active when at least 30% of the larvae showed consistent GFP expression.

Results

Integrative data analysis identifies differentially active regions during fetal brain development

We started our analysis by collecting relevant fetal brain epigenome data sets and previously published putative enhancers (Additional file 2: Table S1, Additional file 3: Table S2). Epigenome data sets included ChIP-seq for various histone modifications, DNase- and ATAC-seq data from various developmental time points and anatomical regions of human fetal brain, generated by several independent studies, including Roadmap, PsychENCODE, and other publications [9, 10, 19, 31–35, 86]. All primary data were reanalyzed using identical computational pipelines, and in total we processed 494 data sets. Scrutinizing through previously published literature on enhancers in brain and neuronal cell types, we collected 1,595,292 putative brain enhancers (Additional file 2: Table S1). These included enhancers retrieved from various enhancer databases, such as VISTA, FANTOM, and EnhancerAtlas, enhancer predictions from the PsychENCODE consortium, human accelerated regions, ultra-conserved regions, and others [9, 11, 21–31]. We first analyzed the overlap between the different putative enhancers and found only a small overlap between enhancer predictions from different studies (Additional file 2: Table S1). We reasoned that if different enhancer prediction methods used in the individual studies identified the same enhancers that only differ by the exact location or length, by merging the overlaps between different studies we could identify functional relevant parts of enhancers. We thus proceeded to determine putative critical regions (pCRs), by determining the unifying overlaps between all putative enhancers (Fig. 1A, step 1). In this analysis, we kept those putative enhancers that were only found in a single study, merged the overlaps between multiple studies and eliminated those regions that were located within 2 kb upstream and 1 kb downstream of a transcriptional start site (TSS) or which had < 10 reads in epigenome data (see “Methods”). This resulted in 202,163 pCRs, with a total length of 93 Mb, an average size of 460 bps, and

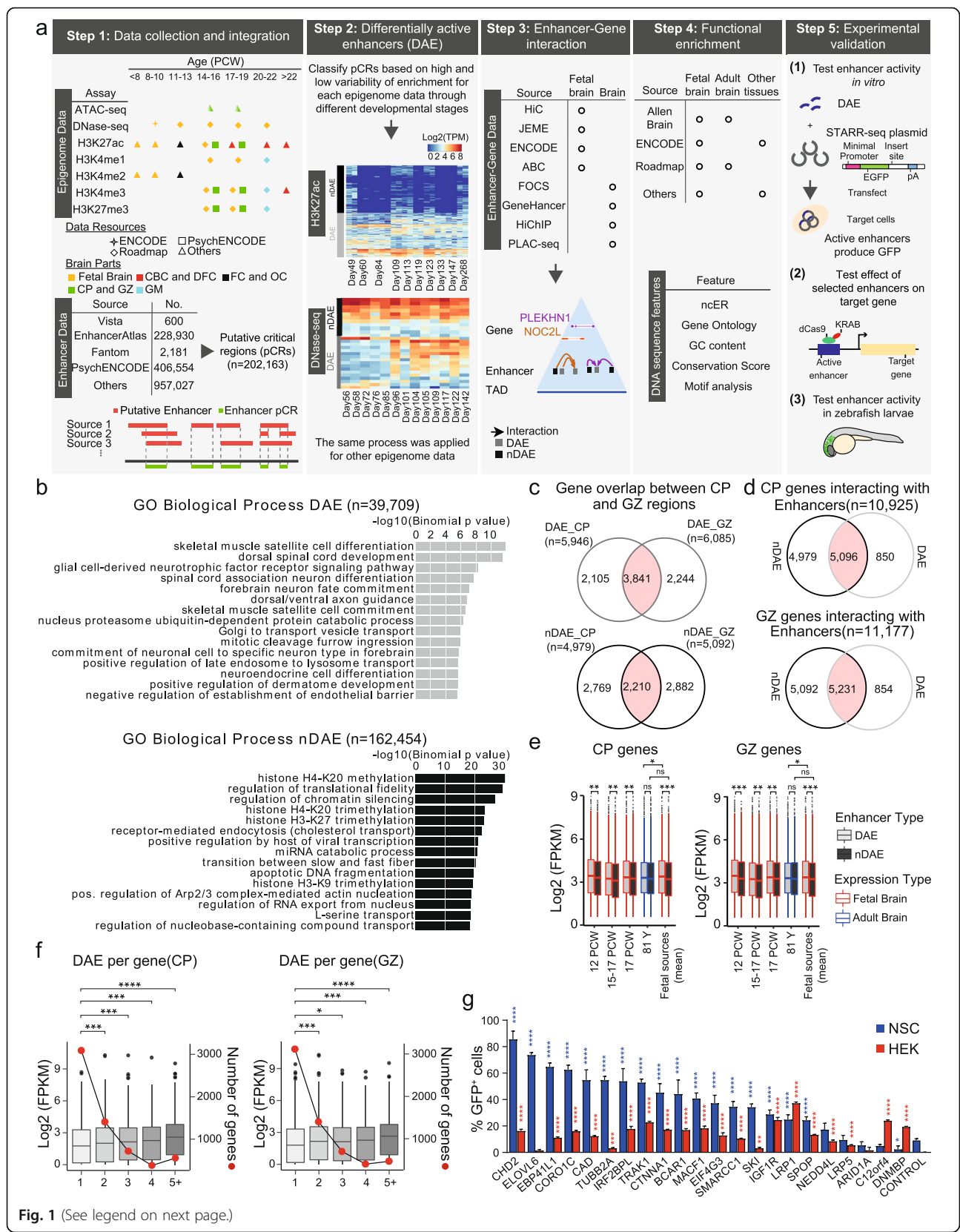


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Integrative analysis of brain enhancers during fetal development. **A** Various steps taken in the integrative analysis of this study. See text for details. **B** Functional enrichment analysis using GREAT [17], for DAEs (upper panel, $n = 39,709$) and nDAEs (lower panel, $n = 162,454$), determined using whole genome as a background. X-axis reports the $-\log_{10} p$ value as determined by GREAT. **C** Venn diagram showing the overlap between DAEs (upper panel) and nDAEs (lower panel) interacting with protein-coding and lincRNA genes in CP (left) and GZ (right). **D** Venn diagram showing the overlap between interactions of protein-coding and lincRNA genes with nDAEs (left) and DAEs (right), for protein-coding and lincRNA genes in CP (upper panel) and GZ (lower panel). **E** Box plots showing gene expression levels as determined by RNA-seq, for genes that interact by HiC with DAEs (light gray) or nDAEs (dark gray) in CP (left) and GZ (right), for fetal (red) or adult (blue) brain samples. Boxes are interquartile range (IQR); line is median; and whiskers extend to 1.5 the IQR. PCW, postconceptional week. FPKM, fragments per kilobase of transcript per million mapped reads. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ns, not significant (wilcox.test). Data obtained from: 12 PCW, Yan et al [18]; 15-17 PCW, De la Torre-Ubieta et al [19]; 17 PCW, Roadmap [10]; 81 years, Roadmap [10]; mean of fetal sources is the mean expression of the first three fetal samples. **F** Box plots showing RNA-seq gene expression for genes interacting with 1, 2, 3, 4, or 5 or more DAEs in CP (left) and GZ (right). Left y-axis shows gene expression (\log_2 FPKM), right y-axis, and line plot shows the number of genes per DAE group. * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$ (wilcox.test). RNA-seq data from Allen human brain atlas [20]. **G** Bar plot showing the percentage of GFP+ cells in NSCs (blue) and HEK cells (red), from cell transfection experiments with an enhancer reporter plasmid for 22 tested enhancers and an empty plasmid control. Plotted is the percentage of GFP+ in cells co-transfected with an mCherry expressing plasmid, to correct for transfection efficiency. Bars show the average from two independent experiments, with each enhancer tested each in duplicate. Error bars represent standard deviation. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$ (one-way ANOVA test followed by multiple comparison test (Fisher's LSD test))

most pCRs located between 5 and 50 kb away from the closest gene TSS (Additional file 1: Fig. S2A, B).

We assumed that enhancers that have functional relevant roles during brain development would show dynamic changes in the levels of histone modifications and chromatin accessibility correlating with their function. To investigate this, we next intersected all pCRs with all epigenome data sets from different time points of fetal brain development and calculated the read count for each pCR region. After TMM-normalization, we performed differential accessibility analysis (for ATAC-seq and DNase data) and generated differential histone modification profiles (for H3K27ac, H3K27me3, H3K4me1, H3K4me2, H3K4me3) using edgeR [50]. This resulted in 39,709 pCRs that showed a high variability for these features across developmental time points (Fig. 1B, Additional file 1: Fig. S2C, Additional file 4: Table S3, see “Methods”) which we refer to as differentially active enhancers (DAEs). In contrast, the remaining 162,454 pCRs showed a more constant epigenome pattern and we thus refer to them as not-differentially active enhancers, nDAEs (Fig. 1B, Additional file 1: Fig. S2C, Additional file 4: Table S3).

Gene ontology analysis using GREAT [17] showed that DAEs were significantly enriched for terms related to brain development, including processes such as forebrain neuron fate commitment, dorsal/ventral axon guidance, and spinal cord development (Fig. 1B, Additional file 5: Table S4). nDAEs appeared to be enriched for more general terms, including various chromatin modifications and receptor-mediated endocytosis (Fig. 1B, Additional file 5: Table S4).

To have a more specific view about the genes regulated by these pCRs, we next linked DAEs and nDAEs to their target genes, using different resources, which either link enhancer to gene promoters by direct chromatin

interaction as determined by chromatin conformation capture techniques (HiC [30], HiChIP [55], PLAC-seq [56]) or by predicting enhancer-gene interactions using statistical models and correlation between gene expression, omics data, and epigenome features (JEME, FOCS, GeneHancer, ENCODE, Activity-by-contact (ABC) method) [51–54, 57] (Additional file 6: Table S5). Since only a limited number of interactions between DAEs or nDAEs and target genes identified by these different methods were supported by >2 of the available resources (Additional file 1: Fig. S3A), and as most interactions and target genes were predicted by the HiC data (Additional file 1: Fig. S3B), we focused on these HiC predicted target gene interactions for the remainder of the analysis. These HiC data were generated from post conceptional week (PCW) 17–18 human brains [30] and were available for the germinal zone (GZ) (containing primarily mitotically active neural progenitors), and the cortical and subcortical plate (CP) (consisting primarily of post-mitotic and migrating neurons). Enhancer-promoter interactions derived from these HiC data do not exclude the fact that the identified DAEs and nDAEs interact with or regulate other genes at other developmental time points or in other cell types, for which at this moment no specific enhancer-promoter predictions are available.

Taking only those enhancer-promoter interactions that occurred in the same topological associated domain (TAD) into account, we found that from all DAEs, 6858 and 6883 for CP and GZ, respectively, interacted with promoters of protein-coding genes or lincRNAs, of which the majority of interactions occur with protein-coding genes. Similarly, 27,004 and 27,161 nDAEs interacted with target genes in CP and GZ, respectively, with a similar distribution between protein-coding and lincRNAs (Additional file 1: Fig. S3C, D).

In total, DAEs interacted with 5946 and 6085 protein-coding and lincRNA genes in CP and GZ, respectively, of which 3841 genes were shared between both CP and GZ (Fig. 1C). The majority of these genes (86%) also had interactions with nDAEs (Fig. 1D). We next integrated available gene expression data from fetal and adult brain (Additional file 7: Table S6) and found that genes that interacted with a DAE had a significantly higher gene expression compared to those genes not interacting with a DAE, at various regions and stages of fetal development but not in adult brain (12 PCW: CP genes p value = 0.002671, GZ genes p value = 5.111e-05; 15–17 PCW: CP genes p value = 0.003251, GZ genes p value = 0.003813; 17 PCW: CP genes p value = 0.002533, GZ genes p value = 0.001813; 81 years: CP genes p value = 0.1377, GZ genes p value = 0.2641; fetal sources mean: CP genes p value = 0.0002696, GZ genes p value = 0.00046; DAE fetal sources mean vs DAE 81 years: CP genes p value = 0.04744, GZ genes p value = 0.01525; nDAE fetal sources mean vs nDAE 81 years: CP genes p value = 0.781, GZ genes p value = 0.4904, wilcox.test) (Fig. 1E, Additional file 1: Fig. S3E). Similar observations were made when using the alternative, not HiC-based enhancer-promoter predictions (Additional file 1: Fig. S3F). In line with earlier findings [93], we find that the more enhancers a gene is interacting with, the higher the gene expression is, and this was also true for the DAEs (Fig. 1F). A recent study determined gene expression trajectories in the dorsolateral prefrontal cortex during pre- and postnatal development. This study identified constant, rising and falling genes, that showed respectively similar, increased or decreased gene expression levels upon development [60]. In line with the earlier gene expression findings, we found that the odds ratio between DAE and nDAE linked genes (as determined by HiC) was significantly higher (odds ratio = 1.183, p value = 0.0008 for GZ; odds ratio = 1.198, p value = 0.0004 for CP, Fisher's exact test) for falling genes, that showed higher gene expression levels in prenatal RNA-seq samples (Additional file 1: Fig. S3G,H).

Finally, to validate that DAEs can function as enhancers, we selected 22 DAEs linked to genes that are expressed in human neural stem cells (NSCs), cloned them in an enhancer reporter plasmid [88] and tested their enhancer activity in cell transfection experiments. Upon transfection in NSCs, 18 out of 22 tested sequences showed significantly increased percentage of GFP+ cells compared to control (normalized for transfection efficiency using an mCherry spiked-in control), confirming enhancer activity (Fig. 1G). Transfecting the same plasmids in non-neural HEK cells showed less pronounced activity. This indicates that 81.8% of the tested DAEs had a measurable enhancer activity using this assay in an in vitro neural cell type. Of note, this does

not exclude activity of those 4 DAEs that do not show enhancer activity in NSCs, in other cell types during fetal brain development.

We conclude that an integrative data analysis of virtually all previously reported brain enhancers identifies a set of DAEs which are associated with a brain developmental gene ontology, increased gene expression in fetal brain and display enhancer activity in vitro.

Multi-gene-interacting enhancers regulate genes implicated in multiple cellular processes and have distinguishing sequence characteristics

In order to understand the biological function of DAEs and nDAEs in more detail, we further characterized these two groups. When determining the number of genes that each DAE is interacting with, we found that the majority of DAEs interact with 1 or 2 genes; but, in addition, a considerable fraction of DAEs also interact with more than 2 genes (19.7 % for CP, 19.4% for GZ) (Fig. 2A), and the same was found for nDAEs (Fig. 2B). When comparing the enrichment of biological processes for the genes that interact by HiC with these multi-gene-interacting DAEs using Enrichr, we found that these genes were enriched for broader developmental and metabolic processes. However, genes that interact with DAEs that only regulate single genes were enriched for more specific brain-related terms, such as “neuron differentiation” and “neuron migration” (Additional file 8: Table S7). Similar results were obtained using GREAT and Metascape analysis, where multi-gene-interacting DAEs for example were enriched in mouse phenotypes associated with “early lethality,” whereas DAEs associated with only a single gene were enriched for “regulation of neural precursor cell proliferation” (Additional file 8: Table S7).

We next asked whether DAEs that regulate single or multiple genes could have distinguishing DNA sequence characteristics, which could support their presumed distinct functional roles. To answer this, we focused on scores that provide some weight based on the underlying sequences: non-coding essential regulation (ncER) score [38], GC content [39], and phastcons score [39]. The ncER scores were recently established using a machine learning model [38], taking functional, mutational, and structural features into account, including sequence constraint in the human population, and provides a score where 0 is non-essential and 1 is putative-essential. We observe that DAEs that interact with 3 or more genes have a significantly higher ncER percentile compared to DAEs that interact with only 1 gene (Fig. 2C). This might reflect their biological function regulating multiple genes, resulting in a higher tendency to be constraint. A similar trend was observed for GC content, where DAEs interacting with more than one gene had a significantly

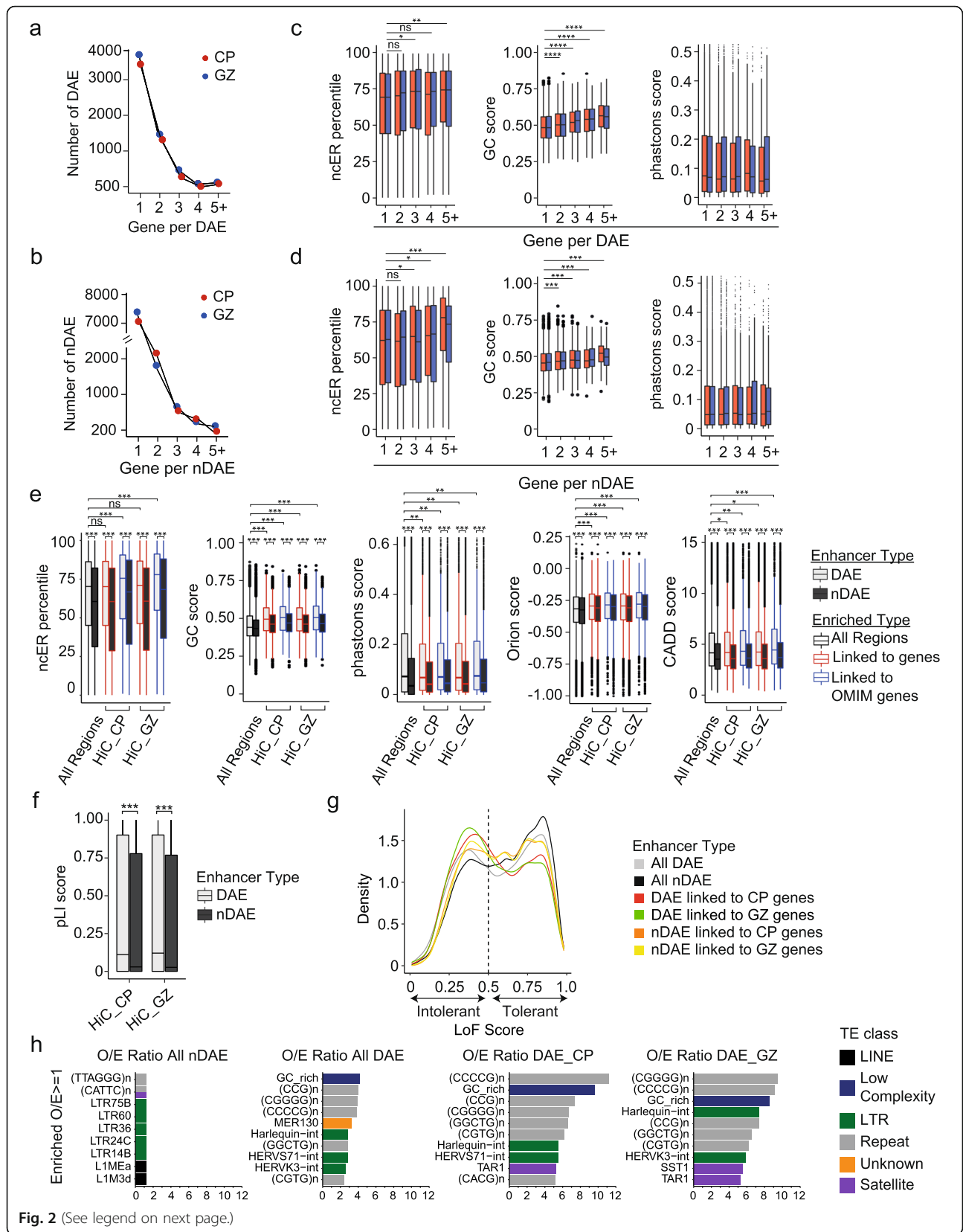


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Distinct sequence characteristics between DAEs and nDAEs. **A** Line graph showing the number of protein-coding and lincRNA genes (1, 2, 3, 4, or 5 or more) that each DAE is interacting with, and the number of DAEs per category, for CP (red) and GZ (blue). **B** As **A**, but here for nDAEs. **C** Box plots showing the median ncER percentile (left) [38], GC content score (middle) [39] or phastcons score (right) [39] for DAEs-CP (red) and DAEs-GZ (blue) that interact with 1, 2, 3, 4, or 5 or more protein-coding and lincRNA genes. Boxes are IQR; line is median; and whiskers extend to 1.5 the IQR. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; ns, not significant (wilcox.test). **D** As **C**, but here for nDAEs. **E** Box plots, showing from left to right ncER percentile [38], GC content score [39], phastcons score [39], Orion score [40], and CADD score [41], for all DAEs (light gray) and nDAEs (dark gray), or for those DAEs and nDAEs that are interacting in CP or GZ with protein-coding and lincRNA genes (red) or genes with a known OMIM phenotype (blue). Boxes are IQR; line is median; and whiskers extend to 1.5 the IQR. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ns, not significant (wilcox.test). **F** Box plot showing the pLI score [42] of genes interacting with DAEs (light gray) and nDAEs (dark gray) in CP or GZ. Boxes are IQR; line is median; and whiskers extend to 1.5 the IQR. *** $p < 0.001$; (wilcox.test). **G** Kernel density plot showing the distribution of loss-of-function tolerance scores for non-coding sequences [43] for all DAEs (light gray), all nDAEs (dark gray), DAEs linked to protein-coding and lincRNA genes in CP (red), DAEs linked to protein-coding and lincRNA genes in GZ (green), nDAEs linked to protein-coding and lincRNA genes in CP (orange), and nDAEs linked to protein-coding and lincRNA genes in GZ (yellow). **H** Bar chart showing the most enriched transposable elements (TEs) overlapping with from left to right all nDAEs, all DAEs, DAEs interacting with protein-coding and lincRNA genes in CP, and DAEs interacting with protein-coding genes in GZ. Plotted is a ratio between the observed (O) number of TEs over the expected (E). Different classes of TE are indicated with different colors as indicated

higher GC content, whereas for the phastcons score, an indicator of multi-species conservation, differences were not significant (Fig. 2C). Similar observations were made for nDAEs (Fig. 2D). Higher GC content has also been observed in more broadly active enhancers in the immune system [94] and might be explained by binding of broadly active transcription factors (TFs) to GC-rich motifs [95].

Sequence characteristics distinguish DAEs from nDAEs

Given the differences in gene ontology between DAE and nDAE linked genes (Fig. 1B) and the differences in ncER score and CG content between enhancers that regulate single versus multiple genes (Fig. 2C, D), we next asked whether there are differences between these scores in DAEs and nDAEs, and whether any potential difference would be influenced by gene interactions that these regulatory elements have. We observed a significantly higher ncER percentile, CG content, and phastcons score when comparing all DAEs to nDAEs (Fig. 2E). Interestingly, some of these scores further increased, when only considering those DAEs and nDAEs that interact with target genes (as determined by HiC). This increased even further when only considering interacting target genes that are associated with known *Online Mendelian Inheritance in Man* (OMIM) phenotypes. Similar observations were made when using the Orion [40] and CADD scores [41] (Fig. 2E) that similarly take depletion of variation in the human population and likelihood of deleteriousness of a given nucleotide based on integration of various annotations into account, respectively. Again, DAEs scored significantly higher for Orion and CADD scores than nDAEs, emphasizing the potentially biological important role of DAEs during brain development. Genes that are essential in humans are generally depleted of loss-of-function alleles, and this is reflected by a higher probability of loss-of-

function intolerance (pLI) score [42]. When we plotted the median pLI of genes linked to DAEs, or to nDAEs, genes linked to DAEs scored significantly higher (Fig. 2F). Finally, a recent study determined loss-of-function tolerance scores for non-coding sequences, by using machine learning and structural variants from whole genome sequencing, including homozygous enhancer deletions [43]. Using this analysis, we observed that DAEs were more likely to be intolerant to loss-of-function, whereas nDAEs were more often tolerant to loss-of-function (Fig. 2G). Again, when only considering those interactions linked to known target genes, scores further improved, in favor of DAEs.

We and others previously showed that functional enhancers can be enriched for transposable elements (TEs), some of which can be human specific [65, 96–98]. We thus asked whether DAEs and nDAEs showed a similar TE enrichment, and whether any TEs could distinguish both groups (Fig. 2H, Additional file 9: Table S8). nDAEs showed a small enrichment for various LTR-containing TEs (e.g., LTR75B, LTR60, LTR36). Compared to nDAEs, DAEs were mainly enriched for CG-rich repeat sequences, and a number of LTR repeats, such as Harlequin-int, HERVS71-int, and HERV K3-int. Enrichment of the latter LTR repeats was not seen when only considering gene-interacting DAEs. The MER130 repeat family was previously shown to be enriched near critical genes for the development of the mouse neocortex and suggested to be co-opted for developmental enhancers of these genes [99]. Interestingly, MER130 repeats were enriched in all DAEs, but this enrichment was lost when only assessing DAEs that interact with genes, which made it difficult to further investigate the role of MER130 in human brain regulation. Compared to our previous findings in human embryonic stem cells (ESCs) [65], the overall TE enrichment in enhancers in brain was markedly different,

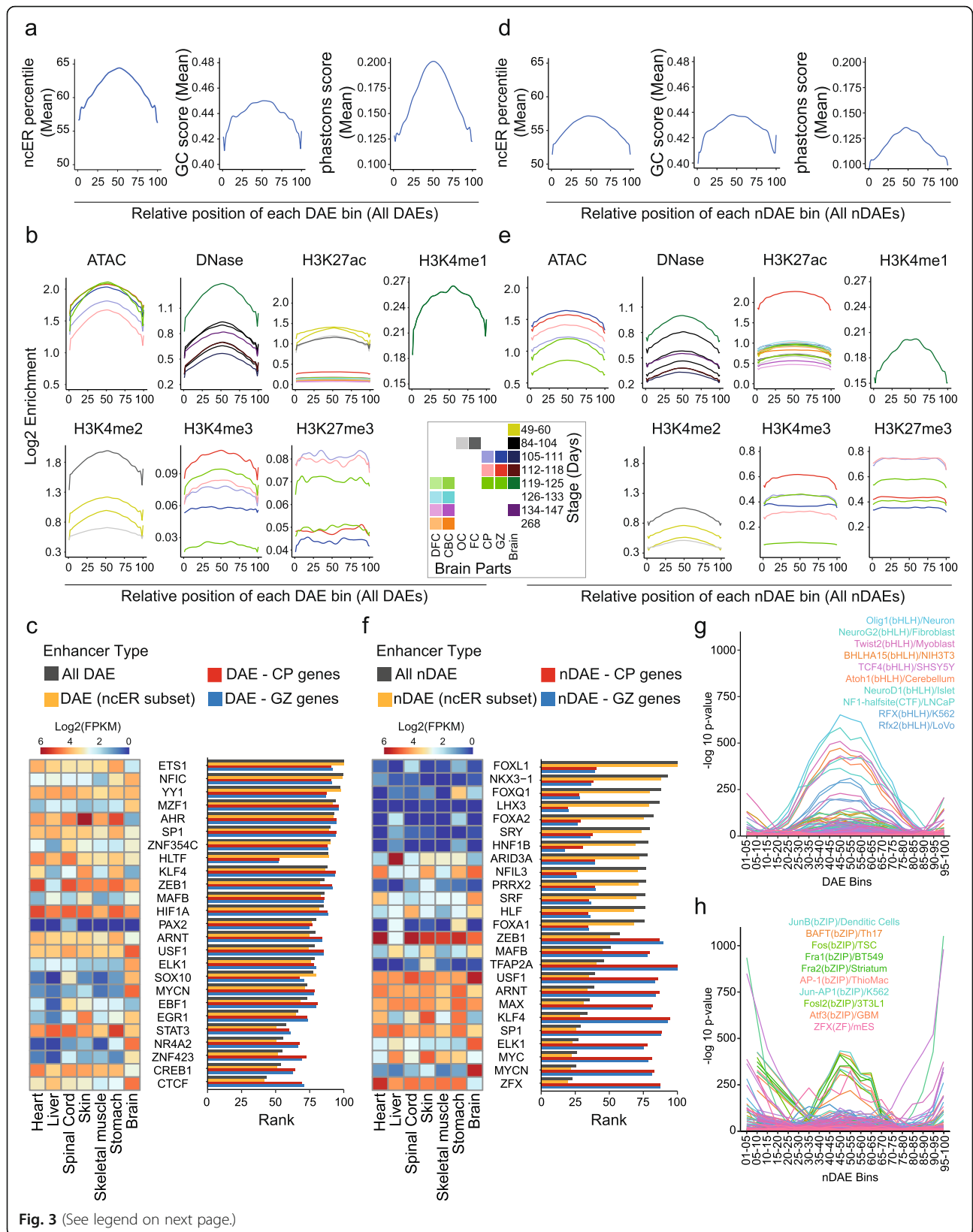


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 DAEs and nDAEs are enriched for distinct transcription factor binding sites. **A** Line plot showing the distribution of the mean ncER percentile (left) [38], GC content score (middle) [39], and phastcons score (right) [39] over the relative bin position for all DAEs. **B** Line plot showing the log₂ enrichment for various epigenome features as indicated, over the relative bin positions for all DAEs. Different colors indicate different time points of human brain development and different brain regions from which the data were obtained. DFC, dorsal frontal cortex; CBC, cerebellar cortex; OC, occipital cortex; FC, frontal cortex; CP, cortical plate; GZ, germinal zone; Brain, whole brain. Epigenome data used are summarized in Additional file 3: Table S2. **C** Bar chart showing the relative LOLA enrichment of TFs from JASPAR in all DAEs (light gray), in the central part of all DAEs (ncER subset, orange), in DAEs linked to genes in CP (red) and in DAEs linked to genes in GZ (blue). X-axis displays the rank score (a combination of *p* value, odds ratio from Fisher's exact test, and the raw number of overlapping regions) from LOLA. The rank was re-scaled between 0 and 100, so that DAEs with a larger TFs enrichment have a higher rank. Also shown is a heatmap showing the RNA-seq expression levels (Log₂ FPKM) of the same TFs across various human fetal tissues. RNA-seq data obtained from ENCODE project [7]. **D** As in **A**, but here for nDAEs. **E** As in **B**, but here for nDAEs. Note the difference in y-axis scale for H3K4me3 and H3K27me3 compared to panel **B** given the higher enrichment in nDAEs. **F** As in **C**, but now for nDAEs. **G** Line plot showing the distribution of enrichment ($-\log_{10} p$ value as determined by HOMER analysis) across the relative DAE bins, for the 251 TF motifs that were not equally enriched in all 20 bin groups. The most enriched TF motifs are indicated. **H** As **G**, but now for 218 TFs that were not equally enriched across the 20 bin groups of all nDAEs

with none of the TEs enriched in active enhancers in ESC showing enrichment at brain enhancers. This could indicate that different TEs co-opted into the regulatory landscape acquired different tissue-specific roles during evolution.

Together this indicates that by investigating unbiased variability in epigenome marks over putative brain enhancers across developmental time points, DAEs and nDAEs can be identified which are associated with different gene ontologies, show different enrichments, have different sequence characteristics, and are distinctively linked to disease-relevant genes.

DAEs and nDAEs are enriched for distinct transcription factor binding sites

As the merging of pCRs and subsequent variability calling identified DAEs with distinct sequence characteristics, we next wondered whether we could further zoom in into each of the DAEs, to identify functional relevant nucleotides. To this end, we again made use of the ncER, CG, and phastcons scores, assuming that the functional relevant nucleotides in each DAE might be those that have higher scores. As the identified DAEs varied in size between 50 and 1000 bps, we first split up each DAE into 10 bp bins and assigned the median ncER, CG, and phastcons scores to each bin. To be able to compare the score distribution within each bin between all DAEs, we re-scaled each DAE to a relative bin position from 1 to 100 (see "Methods" for details). Strikingly, the mean of ncER, CG, and phastcons scores were highest between bins 40 and 60 (Fig. 3A). To exclude that this was an artifact from the bin-rescaling, we plotted the mean distribution for the same scores also for DAEs that had an identical length and found similar results (Additional file 1: Fig. S4A). We next calculated the number of reads from all epigenome data sets and plotted the log₂ enrichment over the same relative DAE bin positions. We found that ATAC-seq, DNase-seq, H3K27ac, H3K4me1,

and H3K4me2 signals (all associated with enhancers) again were most enriched between bins 40 and 60, whereas signal for H3K4me3 (of which high levels are associated with promoters and lower levels are found at enhancers) and H3K27me3 (associated with repressed chromatin) showed a broader distribution (Fig. 3B), and this holds true for all developmental time points assessed. This suggests that on average, the center of the DAEs most likely contains the functional relevant sequences, and given the increased chromatin accessibility at those locations, this could indicate binding of functionally relevant TFs in these central regions.

To investigate this further, we first performed TF enrichment analysis using *Locus Overlap Analysis* (LOLA) [63], on both full length DAEs, as well as on only the central DAE parts between bins 40 and 60 (ncER subset). LOLA performs enrichment analysis based on genomic regions and tests the overlap of the query regions with a core reference database assembled from public data, including amongst others ChIP-seq data from CODEX [100]. We found a similar enrichment of TF binding sites between full length and central parts of DAEs (Fig. 3C, Additional file 9: Table S8), and between all DAEs and those interacting with target genes in CP and GZ. The most enriched TFs at DAEs according to LOLA included well-known TFs with essential roles for brain development. This includes amongst others ETS1, a widely studied TF with functions in different biological systems which was previously shown to be necessary for radial glia formation in vertebrates [101] and FGF-dependent patterning of anterior-posterior compartments in the central nervous system of *Ciona* (a marine invertebrate that is a well-suited model to study cell fate specification in chordates) [102]; YY1, a crucial TF which is involved in both gene activation and repression [103], mediating enhancer-promoter interactions [104] and of which mutations cause a neurodevelopmental disorder [105]; and CTCF, a master regulator of chromatin

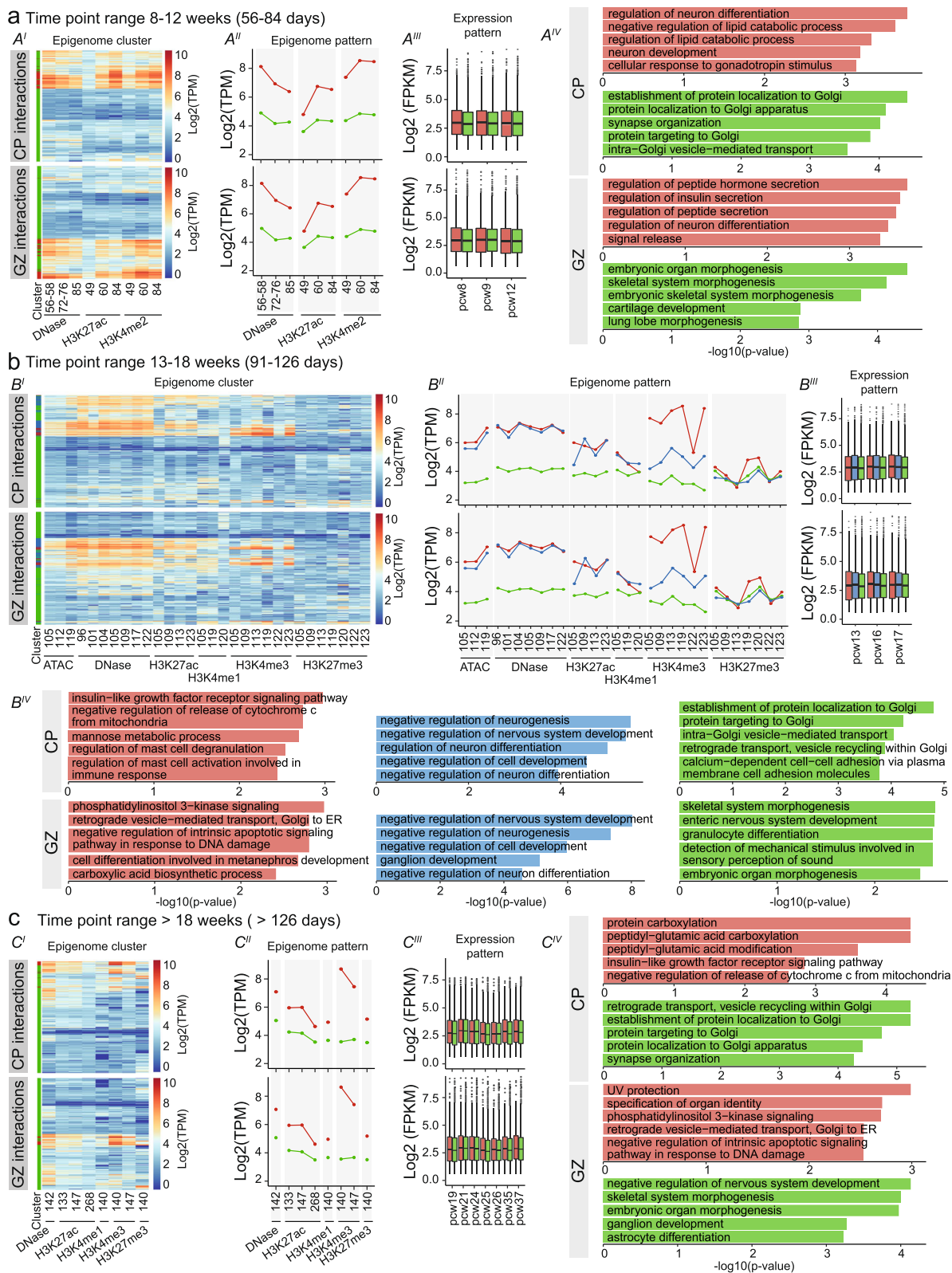


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Clustering of DAEs unravels temporal dynamics of brain gene regulation. **A** Heatmap displaying all available epigenome features for PCW 8–12, across all DAEs interacting with protein-coding genes in CP (upper heatmap) and GZ (lower heatmap) (A^I). *K*-means clustering analysis of epigenome features (A^II) identifies two clusters, cluster 1 (red) and cluster 2 (green). Level of enrichment is indicated on the y-axis in Log₂ TPM. Box plots (A^III) shows RNA-seq gene expression of protein-coding genes regulated by the DAEs from each cluster (Expression pattern), for available data from PCW 8, 9, and 12 [20]. Boxes are IQR; line is median; and whiskers extend to 1.5 the IQR. Gene enrichment analysis for the corresponding genes in each cluster (A^IV). X-axis shows the $-\log_{10}$ (p value) from Enrichr. **B** As for **A**, but now for PCW 13–18. **C** As for **A**, but now for PCW > 18

structure, of which de novo mutations cause intellectual disability [106]. We next repeated the same analysis for nDAEs (Fig. 3D–F, Additional file 9: Table S8). Similar to our observations for DAEs, nDAEs had higher ncERs, CG content, and conservation at the central part, with those regions being enriched for enhancer associated histone marks, but showed less variability over time. When performing TF enrichment analysis using LOLA, we observed a different and less specific set of TFs enriched at nDAEs compared to DAEs. Also, enrichment was lower at those nDAEs that were interacting with target genes. Again similar enrichment was found in the central part compared to the whole nDAEs. Enriched TFs for nDAEs included amongst others FOXL1, a transcriptional repressor that regulates central nervous system development [107]; the LIM homeodomain TF LHX3, that is essential for pituitary and nervous system development [108, 109]; and FOXA2, which plays a role in midbrain dopaminergic neurons [110, 111] (Fig. 3F). Shared TFs enriched both at DAEs and nDAEs included SP1, loss of which in astrocytes impacts on neurons in the cortex and hippocampus of mice [112]; MAFB, a basic leucine zipper TF that plays a role in hindbrain development [113–115] and postnatal brain development [116, 117]; and ZEB1 which is required for neuronal differentiation [118, 119].

As LOLA analysis considers a single shared base pair being sufficient for regions to count as overlapping, this analysis could not distinguish well between TFs specifically enriched at the central part of DAEs and nDAEs relative to the flanking regions. We therefore further investigated which TFs motifs were specifically enriched at the central parts versus other parts of DAEs and nDAEs, using motif enrichment analysis with HOMER [64], a motif discovery algorithm, which identifies regulatory elements that are specifically enriched in the query set relative to background. We first split the 100 relative bins into 20 groups of 5 consecutive bins each and determined the significantly enriched TF motifs ($p \leq 0.01$) for each of these 20 bin groups (Additional file 9: Table S8). Amongst the enriched motifs, we found back, amongst others, the motifs for the TFs enriched using the LOLA analysis, validating these findings (Additional file 9: Table S8). When plotting the number of

significant motifs ($p \leq 0.01$) per bin group and the number of target sequences with those motifs, we found that bins located in the central part of both DAEs and nDAEs had both the highest numbers of significant TF motifs and the highest number of target sequences (Additional file 1: Fig. S4B,C). As most enriched motifs were found in multiple bins, and there can be multiple TF bindings sites of the same TF within the same enhancer, we next focused on only those TF motifs which were not equally enriched in all 20 bin groups ($n = 251$ for DAEs and $n = 218$ for nDAEs). For both DAEs and nDAEs, we again found most motif enrichment in the central enhancer part, with DAEs being more enriched than nDAEs (Fig. 3G, H, Additional file 1: Fig. S4D,E). Amongst the most enriched TF motifs at the center of DAEs were motifs for the proneural basic helix-loop-helix transcription factors NEUROG2, ATOH1, and NEUROD1 that promote neurogenesis [120–122], OLIG1, a marker of oligodendrocytes [123] that also regulates the neuron-glia switch during earlier embryonic development [124, 125], TCF4, that is necessary for neuronal migration and the correct development of the cerebral cortex [126] and loss of which is associated with intellectual disability [127], and NF1 that regulates neuronal and glial differentiation and is causative of neurofibromatosis type 1 when mutant [128] (Fig. 3G). Enriched TF motifs at the central part of nDAEs are involved in more ubiquitous processes and include mainly activator protein 1 (AP-1), a heterodimer composed of members of the JUN (including JUNB), FOS (including FOSL2, FRA1, FRA2), ATF (including ATF3, BAFT), and MAF family that regulates a wide variety of cellular processes in response to a wide range of extracellular cues [129] (Fig. 3H).

Together this indicates that on average the central part of brain enhancers (both DAE and nDAEs) contains relevant but partially distinct TF binding sites and might be enriched for functional relevant sequences, which can be further fine-mapped using ncER scores and other sequence characteristics. To test this directly, we selected three DAEs, linked to *IRF2BPL*, *CHD2*, and *MACF1*, that showed activity in reporter assays in NSCs (Fig. 1G) and deleted 10–30 bp of those regions that had the highest ncER scores in those enhancers. Upon transfection of

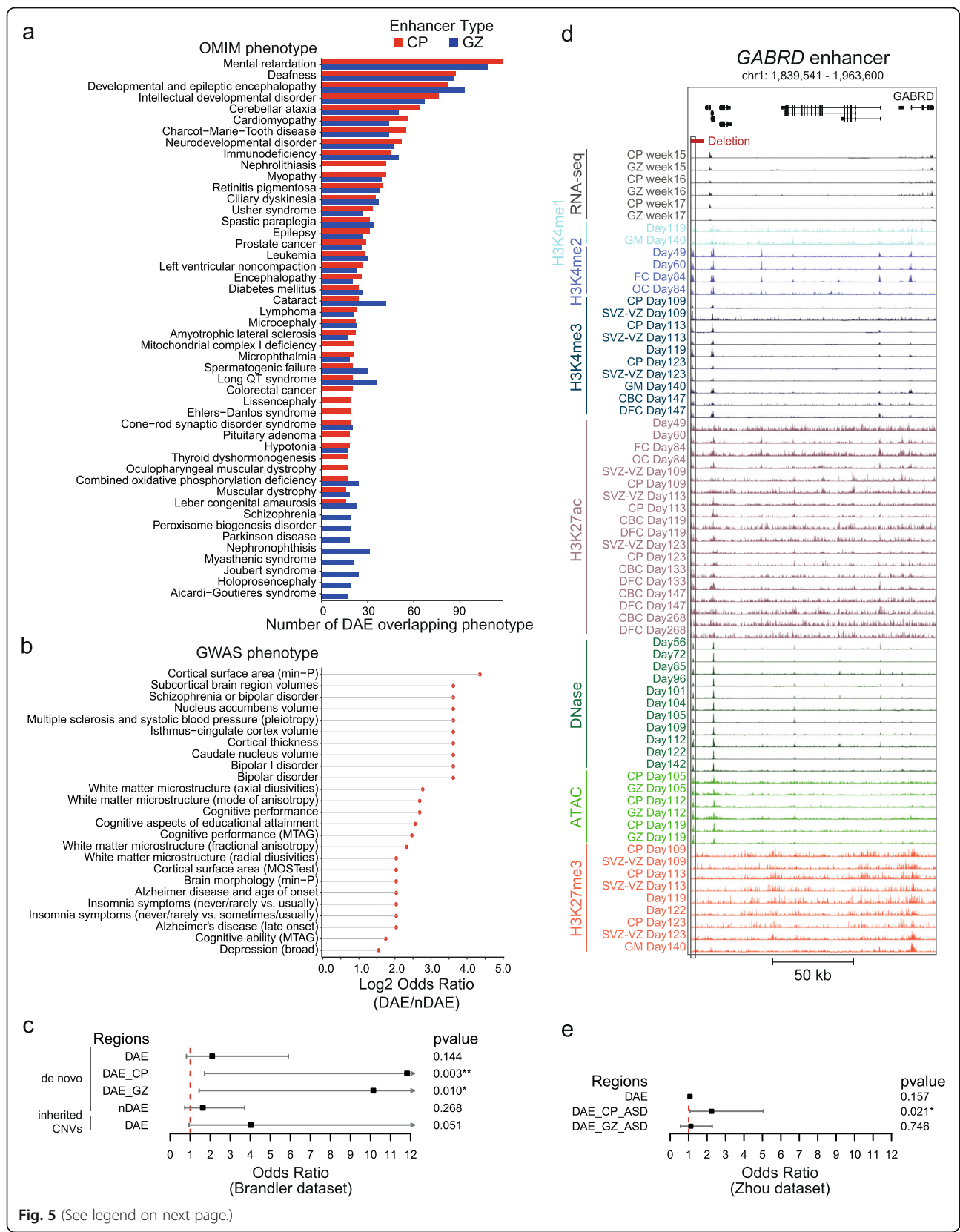


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Variants in DAEs and nDAEs are associated with human disease. **A** Bar graph showing the number of DAEs linked to their target genes in CP and GZ and their most enriched OMIM phenotypes. **B** Plot showing the top-25 GWAS phenotypes that are enriched in DAEs compared to nDAEs (log₂ odds ratio DAE/nDAE). **C** Line graph showing the odds ratio, confidence interval, and *p* value for enrichment of CNVs from an ASD cohort at DAEs and nDAEs. CNVs data obtained from Brandler et al [44]. * *p* < 0.05; ** *p* < 0.01 (Fisher's exact test). **D** Genome browser track showing the regulatory landscape of the *GABRD* gene. Indicated are a DAE (chr1: 1,840,449-1,840,835) that is interacting with the *GABRD* promoter, and a deletion (chr1: 1,840,001-1,845,000) that is found in an epilepsy patient (CNET0068) from Monlong et al. [45]. * *p* < 0.05 (Fisher's exact test). **E** Line graph showing the odds ratio, confidence interval, and *p* value for enrichment of SNV from an ASD cohort at DAEs and nDAEs. SNV data obtained from Zhou et al. [46]

these mutant DAEs, we observed a significantly reduced enhancer activity for *IRF2BPL* and *CHD2*, but not for *MACF1* (Additional file 1: Fig. S4F). Deleting regions with a lower ncER score did not affect enhancer activity. Together, this indicates that integrative analysis, variability analysis during development, and sequence characteristics can identify functional relevant nucleotides in brain enhancers.

DAEs show temporal epigenome dynamics during human brain development

To further understand the dynamics of enhancer regulation, we subdivided DAEs interacting with genes in GZ and CP by performing clustering analysis on all available epigenome data sets, at different developmental stages (between 8 and 12 PCW, 13–18 PCW, and > 18 PCW) (Fig. 4A, Additional file 10: Table S9). At 8–12 PCW, we found two clusters for both GZ and CP that showed relatively constant enrichments over time, with the first cluster (red) showing a higher enrichment for all epigenome features available for that developmental stage, compared to the second cluster (green). No statistically significant differences in gene expression levels between genes linked to both clusters were found. Genes associated with cluster 1 DAEs in CP were enriched for gene ontology terms related to neuronal differentiation, whereas cluster 2 was dominated by processes in the Golgi. Likewise, for GZ, genes associated with cluster 1 seemed to be associated with more specific biological functions, whereas processes associated with cluster 2 showed more broad involvements (Fig. 4A, Additional file 10: Table S9).

At 13–18 PCW, three clusters emerged in both GZ and CP (Fig. 4B, Additional file 10: Table S9). Whereas cluster 3 (green) showed relatively low levels of epigenome marks similar to cluster 2 at 8–12 PCW, cluster 1 (red) and cluster 2 (blue) showed higher epigenome enrichments. Both cluster 1 and 2 had similar levels of H3K27ac, but mainly diverged from each other on the levels of H3K4me3. Cluster 2 was strongly enriched for processes involved in neural system development both in CP and GZ. Gene ontology of genes associated with cluster 1 (red) which

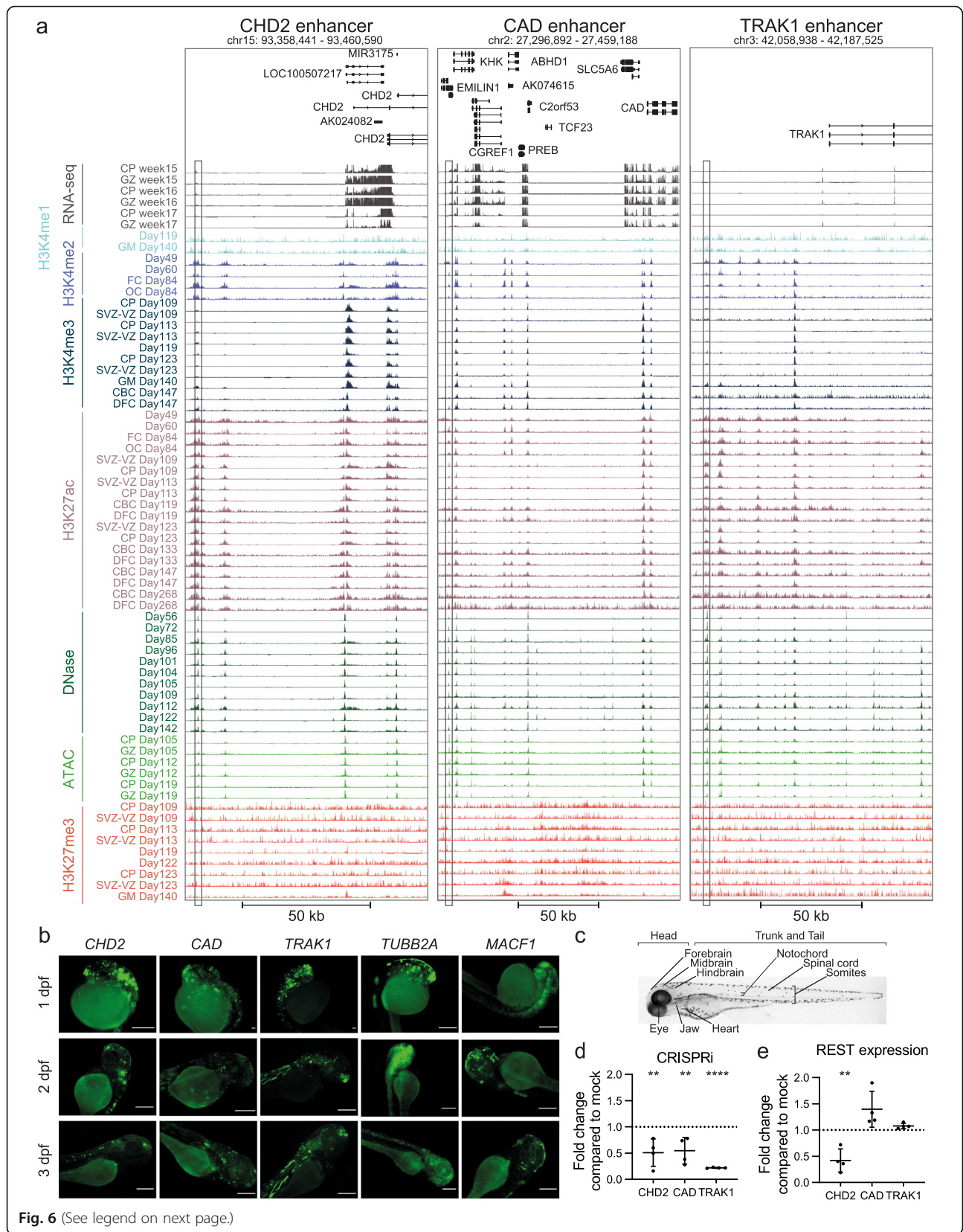
showed higher H3K4me3 levels, showed enrichment for insulin-like growth factor receptor signaling and immune cell-related processes in CP. Insulin-like growth factors are important for neuronal survival and neurogenesis [130]. As high levels of H3K4me3 have also been found at enhancers in blood cells [131], possibly stabilizing their transcription, it is tempting to speculate that part of this cluster reflects enhancers active in hematopoietic cells from the developing vasculature [132] and microglia (brain tissue macrophages) that are invading the brain at these developmental time points [133]. In GZ, cluster 1 was associated with phosphatidylinositol 3-kinase signaling, which is important for commitment of neural progenitor cells [134, 135].

Finally, at > 18 PCWs, we found two clusters of DAEs, of which cluster 1 (red) was marked by higher levels of epigenome marks (Fig. 4C, Additional file 10: Table S9). In CP, genes associated with this cluster were enriched for carboxylation processes and insulin-like growth factor receptor signaling. Genes associated with the second cluster (green) were again more enriched for broad developmental processes, including the Golgi system. In GZ, genes associated with cluster 1 (red) were amongst others involved in DNA damage repair. Indeed, alterations in this pathway can lead to reduced proliferation of neural progenitor cells leading to microcephaly [136, 137]. Cluster 2 (green) in GZ was associated with terms related to neurodevelopment and organ morphogenesis.

Together, this shows that temporal epigenomic rearrangement in DAEs is reflected in regulating the expression level of genes that are important in developmental and cell type-specific processes.

Cell type specificity of DAEs and nDAEs and their dynamics in adult brain

To further investigate cell type specificity of DAEs and nDAEs, we performed two additional analyses. First, we compared DAEs and nDAEs to recently identified cell type-specific regulatory elements. A recent study used scATAC-seq to generate a human cell atlas of fetal chromatin accessibility spanning 15 organs, including fetal brain [85]. When overlapping DAEs and nDAEs to the



(See figure on previous page.)

Fig. 6 CRISPRi and zebrafish experiments validate activity of DAEs regulating genes involved in neurogenetic disorders. **A** Genome browser tracks showing enhancers interacting with *CHD2* (left), *CAD* (middle), and *TRAK1* (right). Shown are RNA-seq expression profiles, various histone modifications, and ATAC-seq and DNase profiles for various time points during human fetal brain development, as indicated. The tested DAEs are indicated by the box. **B** Representative fluorescent images showing GFP expression of transgenic enhancer reporter assays in zebrafish larvae at 1, 2, and 3 dpf. Tested are the enhancers for *CHD2*, *CAD*, and *TRAK1* (shown in **A**), and two additional enhancers for *MACF1* and *TUBB2A*. The five tested enhancers induced GFP expression in the head of the larvae, amongst others in the forebrain in 61.1%, 81.8%, and 87.9% larvae for *CHD2*; 88.9%, 85.4%, and 85.7% for *CAD*; 87.1%, 70%, and 88.5% for *TRAK1*; 81.5%, 85.7%, and 76.2% for *MACF1*; and 87.5%, 100%, and 100% for *TUBB2A*, respectively at 1, 2, and 3 dpf. Also peripheral neuron-specific GFP expression was found, with 0%, 60.6%, and 21.2% for *CHD2*; 68.9%, 24.4%, and 51.4% for *CAD*; 83.6%, 65.5%, and 67.3% for *TRAK1*; 37%, 50%, and 33.3% for *MACF1*; and 50%, 83.3%, and 63.3% for *TUBB2A*, respectively at 1, 2, and 3 dpf. See also Additional file 14: Table S13. Scale bars represent 500 μm . **C** Bright-field image of a wild type zebrafish larvae at 3 dpf (lateral view), with the anatomical sites that were scored for GFP expression indicated. **D** qRT-PCR showing reduction of *CHD2*, *CAD*, and *TRAK1* expression in NSCs upon silencing of respective enhancer by dCas9-KRAB-MECP2. Data represent fold change of expression of respective genes compared to mock transfected cells (KRAB-MECP2 plasmid only, no gRNA plasmid). Two independent transfection experiments were performed, each in duplicate. All data points and standard deviation are shown. ** $p < 0.01$; **** $p < 0.0001$ (one-way ANOVA test followed by multiple comparison test (Fisher's LSD test)). **E** qRT-PCR showing reduction of *REST* expression in NSCs upon silencing of *CHD2*, *CAD*, or *TRAK1* enhancers by dCas9-KRAB-MECP2. Data represent fold change of REST expression compared to mock transfected cells (KRAB-MECP2 plasmid only, no gRNA plasmid). Two independent transfection experiments were performed, each in duplicate. All data points and standard deviation are shown. ** $p < 0.01$ (one-way ANOVA test followed by multiple comparison test (Fisher's LSD test))

most specific chromatin accessibility peaks per cell type, we found 7753 DAEs and 7946 nDAEs that overlapped with these cell type-specific chromatin accessibility peaks, including those found in several types of neurons and astrocytes (Additional file 1: Fig. S5A,B). This indicates that our bulk analysis re-identifies cell type-specific chromatin accessibility peaks which might therefore present cell type-specific enhancers, and it also shows that the bulk analysis identifies additional enhancers that are not captured by the single-cell chromatin accessibility profiles.

We next investigated how these cell type-specific enhancer might behave over time. Two recent studies determined cell type-specific regulatory elements from postnatal brain with a reasonable overlap between both studies (Additional file 1: Fig. S5C,D), by either isolating cell type-specific bulk populations from brain followed by ATAC-seq and ChIP-seq for H3K27ac and H3K4me3 [56], or by performing scATAC-seq [55]. Comparing the DAEs and nDAEs to these cell type-specific regulatory elements showed as expected that only a fraction of DAEs and nDAEs from the fetal brain analysis showed an overlap with the cell type-specific regulatory elements derived from postnatal samples. Amongst those, we found overlap with cell type-specific regulatory elements from neurons, oligodendrocytes, astrocytes, and microglia (Additional file 1: Fig. S5E,F). This indicates that despite determined from an integrative analysis of bulk samples derived during fetal brain development, a fraction of DAEs and nDAEs can be linked to cell type-specific regulatory elements which are likely to also have roles in postnatal brain. In contrast, other DAEs and nDAEs are likely having fetal-specific functions.

To further investigate the dynamics of DAEs and nDAEs in adult brain, in the second analysis, we compared H3K27ac levels obtained from both fetal and adult

samples derived from a single study [86] for all DAEs and nDAEs linked to target genes in GZ and CP by HiC, and performed clustering and gene ontology analysis (Additional file 1: Fig. S6, Additional file 10: Table S9). We found that DAEs that were mainly enriched for H3K27ac in fetal samples were as expected associated with gene ontology terms related to fetal brain development, including regulation of neuron differentiation. DAEs which also showed H3K27ac enrichment in adult samples were associated with more broad physiological processes.

Together, this shows that part of DAEs and nDAEs can be linked back to cell type-specific regulatory elements despite being identified from bulk tissue analysis and that some DAEs and nDAEs are likely to also function in postnatal brain.

DAEs regulate disease-relevant genes and are enriched for disease implicated variants

Given our findings that DAEs are associated with genes relevant for brain development, we further investigated which disease-relevant genes are regulated by DAEs. We first focused on known disease causing genes retrieved from OMIM. We found that 1556 OMIM genes are regulated by DAEs (of which 1165 and 1166 from the interactions found in GZ and CP, respectively) (Additional file 11: Table S10). Most DAEs are linked to genes involved in mental retardation, developmental and epileptic encephalopathy, and neurodevelopmental disorders (Fig. 5A). This included genes like *KMT2C*, involved in Kleefstra syndrome (OMIM #617768), and *GRIN2A* of which heterozygous mutations cause epilepsy and speech delay (OMIM #245570). Next to genes, enhancers can also interact with other additional enhancers. Interestingly, the more additional enhancers (DAE and/or nDAE) a DAE was interacting with, the more likely the

target gene of this DAE was an OMIM gene (Additional file 1: Fig. S7A). This supports recent findings that the number of enhancers linked to a gene reflects its disease pathogenicity [138] and confirms enhancer redundancy for disease-relevant genes [139].

We next leveraged published GWAS loci for brain-related traits and disorders (Additional file 12: Table S11). When comparing the odds ratio between DAEs and nDAEs, we found that DAEs were more often enriched for various significant GWAS loci, reflecting a broad variety of both brain developmental processes (e.g., volumes of different anatomical brain regions) and neurodevelopmental disorders (e.g., mental development, autism) (Fig. 5B). Similarly, using LD score regression analysis we found enrichment of heritability for variants within DAEs, nDAEs, and pCRs, including for the trait “intelligence” (Additional file 1: Fig. S7B).

Encouraged by these findings, we next asked whether copy number variants (CNVs) or single-nucleotide variants (SNVs) at DAEs could be involved in causing genetic disease. We first leveraged previously published disease implicated CNVs. Brandler et al. performed WGS in their discovery cohort of individuals affected by an autism spectrum disorder (ASD) and unaffected individuals and reported on 135 de novo CNVs (104 deletions, 29 duplications, and 2 inversions) [44]. Of these, 25 overlapped a DAE in cases, and 8 in controls (odds ratio = 2.10, p value = 0.144101). When only considering those CNVs overlapping DAEs linked to target genes, this became 17 in cases and 1 in control for DAEs linked to CP genes (odds ratio = 11.83, p = 0.003003) and 15 in cases and 1 in control for DAEs linked to GZ genes (odds ratio = 10.14, p value = 0.010423). For nDAEs, 36 CNVs were found in cases and 15 in controls (odds ratio = 1.63, p value = 0.267964). However, as not all these CNVs exclusively covered non-coding regions, it cannot be excluded that the observed association is due to disrupted coding genes, rather than involvement of DAEs. We therefore also assessed rare inherited deletions from the same study that did not overlap with coding exons (n = 213 in total, 175 in cases and 38 in controls). From these, 32 cases had a deletion covering a DAE, compared to two controls (odds ratio = 4.027972, p value = 0.05119). Although not significant, this might point to more deletions covering DAEs in ASD individuals but would require a larger sample size to be confirmed (Fig. 5C, Additional file 13: Table S12).

In another study, Monlong et al. [45] reported on CNVs in 198 epilepsy patients detected by WGS. They found an enrichment of rare non-coding CNVs near known epilepsy genes, with the *GABRD* gene showing the strongest and only nominally significant association with 4 non-coding deletions amongst the epilepsy patients. Interestingly, a 4999 bp deletion reported in that

study, overlapped with a 386 bp DAE which is located ~ 110 kb upstream of *GABRD* and which interacts with its promoter (Fig. 5D). Hence, it is possible that deletion of this DAE affects *GABRD* expression, which might be implicated in the phenotype of that individual.

Third, we made use of de novo SNVs found in WGS from 1790 ASD simplex families [46]. We found 932 de novo variants that overlapped all DAEs in ASD individuals compared to 829 variants overlapping all DAEs in unaffected individuals (odds ratio = 1.07, p value = 0.157). We next repeated the analysis with only those DAEs that are interacting with known autism genes from the SFARI Gene database (n = 1003 genes) [84]. We found 26 cases and 11 controls with de novo variants in DAEs that interact with autism genes in CP (odds ratio = 2.249703, p value = 0.021455), whereas for DAEs interacting with autism genes in GZ, this was 20 cases and 17 controls (odds ratio = 1.11955, p value = 0.745628) (Fig. 5E, Additional file 13: Table S12). Interestingly, for each of the genes *CIB2*, *FBRSL1*, *PACS2*, *KDM4B*, and *MYT1L*, we found 2 individuals with autism with de novo variants in DAEs interacting with these genes. These variants are either absent or extremely rare in a large control cohort of gnomAD [140], possibly pointing to a role in causing the phenotype, although this will require further validation.

Together this indicates that DAEs are linked to disease-relevant genes and are enriched for GWAS loci relevant for brain-related traits and for variants linked to genetic disorders.

CRISPRi and zebrafish experiments confirm enhancer activity of DAEs regulating genes involved in epileptic encephalopathy

To further substantiate our findings, we validated the biological role of selected enhancers, using in vivo zebrafish transgenic reporter assays and CRISPR inhibition in human NSCs by focusing on enhancers linked to disease-relevant genes.

CHD2 belongs to the chromodomain helicase DNA-binding families of chromatin remodeling proteins, and haploinsufficiency of this gene has been associated with a developmental and epileptic encephalopathy, presenting with early onset intractable seizures, cognitive regression, intellectual disability and ASD behaviors (OMIM #615369) [141]. Around 80 kb upstream of *CHD2*, we found a DAE that interacts with the *CHD2* promoter (Fig. 6A). In NSC reporter assays, this region showed strong enhancer activity, and this was less pronounced in non-neural HEK cells (Fig. 1G). To further study the biological relevance of this region, we first tested enhancer activity in vivo using zebrafish transgenesis. Out of the 36 analyzed zebrafish larvae, 61.1% showed GFP expression in the forebrain at 1 day post

fertilization (dpf), and this increased to 81.8% at 2dpf and 87.9% at 3dpf, indicating enhancer activity (Fig. 6B, C). Expression was also found in midbrain and hindbrain, at a slightly lower extent, in the eyes, in peripheral neurons, and in the spinal cord (Additional file 14: Table S13). GFP expression in the developing zebrafish brain correlated with in situ hybridizations of endogenous *chd2* [142]. To test whether epigenome silencing of this enhancer would affect *CHD2* expression, we performed CRISPR interference (CRISPRi) by targeting dCas9-KRAB-MeCP2 to the enhancer region by coexpression of gRNAs with a GFP fluorescent reporter. Transfection efficiency in these experiments, based on FACS for GFP, was 78–92%, and this resulted in around 50% reduction of *CHD2* expression compared to mock cells transfected solely with dCas9-KRAB-MeCP2 (Fig. 6D). Interestingly, it was previously shown that silencing of *CHD2* leads to reduced expression of *REST* [143]. In agreement with this, cells with reduced *CHD2* expression upon *CHD2* enhancer silencing showed reduced *REST* expression (Fig. 6E). This confirms that *CHD2* is under control of the investigated DAE.

Bi-allelic variants in *CAD* cause an early infantile epileptic encephalopathy (OMIM #616457) [144] that is characterized by global developmental delay, loss of skills, therapy refractory epilepsy, brain atrophy, and dyserythropoietic anemia. We found an enhancer located in the third intron of *EMILIN1*, around 135 kb upstream of *CAD* that interacts with the *CAD* promoter (Fig. 6A) and which showed strong enhancer reporter activity in NSCs and only limited activity in HEK cells (Fig. 1G). Targeting this region in NSCs by CRISPRi significantly diminished gene expression of *CAD* to around 50% compared to mock (Fig. 6D). Similar to *CHD2*, in vivo reporter assays in zebrafish recapitulated in situ hybridisation results for *cad* [145]. From the 45 analyzed larvae, GFP expression was found in the forebrain of 88.9% larvae at 1 dpf, which remained ~85% at 2 and 3 dpf. Again, GFP expression was observed also in midbrain, hindbrain, eyes, in peripheral neurons, notochord, and spinal cord (Fig. 6B, Additional file 14: Table S13).

We next focused on an enhancer interacting with *TRAK1*, located ~65 kb upstream of the TSS (Fig. 6A). *TRAK1* is involved in mitochondrial trafficking, and bi-allelic loss-of-function variants in *TRAK1* are associated with developmental and epileptic encephalopathy (OMIM #618201) [146, 147]. Similar to the *CHD2* enhancer results, the *TRAK1* enhancer showed higher reporter assay activity in NSCs than in HEK cells (Fig. 1G). Targeting of dCas9-KRAB-MeCP2 to the *TRAK1* enhancer reduced *TRAK1* expression to ~25% residual expression (Fig. 6D). Interestingly, in the VISTA enhancer browser, another enhancer linked to *TRAK1* (hs2359), ~18 kb upstream of the TSS, has been

reported which did not show enhancer reporter activity in E11.5 mouse embryos. When testing the *TRAK1* enhancer identified here in zebrafish (Fig. 6B), we found that from 55 larvae, 89.1% showed GFP expression in the forebrain, as well as in the midbrain (74.5%) and hindbrain (85.5%). The larvae showed decreasing GFP expression in neurons outside of the brain over the different time-point (83.6% at 1 dpf, 65.5% at 2 dpf, and 67.3% at 3 dpf) and increasing expression in both somites (89.1%) and heart (58.2%) at 3 dpf, compared to 32.7% and 1.8% at 1 dpf larvae, respectively. Moreover, this enhancer was active also in the eye, trunk and tail, notochord, and at 1 dpf, in the spinal cord (Additional file 14: Table S13).

Finally, next to these three enhancers, we validated 7 additional enhancers linked to the genes *LRP1*, *LRP5*, *TUBB2A*, *ELOVL6*, *MACF1*, *C12orf4*, and *EBP41L1* using zebrafish reporter assays and could confirm enhancer activity for all of them with >60% larvae expressing GFP (Fig. 6B, Additional file 1: Fig. S8, Additional file 14: Table S13). These included enhancers linked to the disease genes *MACF1* (OMIM #618325) and *TUBB2A* (OMIM #615763), of which coding pathogenic mutations cause brain malformations [148, 149], and *C12orf4* (OMIM #618221) of which bi-allelic variants cause intellectual disability [150]. Together, this shows that DAEs identified in this integrative analysis show enhancer activity in vitro and in vivo and regulate, amongst others, genes linked to Mendelian disorders.

Discussion

Understanding the role of NCREs in development and disease still needs a significant effort at multiple levels: starting from identifying and annotating NCREs to investigating their target gene(s) and function. In the past few years, the identification and annotation of NCREs have gained a lot of attention. However, despite these developments, due to their sheer number and complex function, more studies and concerted efforts are needed to understand the role of NCREs in development and disease. Here we performed an integrative analysis of virtually all previously described putative enhancers and epigenome datasets of relevance for human brain development.

Our analysis has allowed us to first identify the intersection between previous studies and identify a list of putative NCREs. This is an important step as the different regions that were identified by previous investigations often have slightly different coordinates, length, and quality. Our putative regions are thus the commonality between all the different studies that are conducted *hitherto*, but at the same time keep the originality in each of them. To further specify enhancers that might

have a biological relevance, mapping epigenomic data to these putative regions allowed us to identify around 40 thousand enhancers that display epigenomic rearrangement during human brain development. These DAEs have different sequence characteristics compared to non-variable enhancers, are bound by distinct sets of TFs, regulate disease-relevant genes, and can harbor non-coding variants that are associated with human disease. Furthermore, our integrative analysis identified a large number of enhancers linked to known disease genes and expands on the knowledge of regulation of these genes. For example, *CHD2* expression regulation has so far only been known to be influenced by a highly conserved long non-coding RNA (lncRNA) referred to as *CHD2* Adjacent Suppressive Regulatory RNA (CHAS-ERR), which is located in proximity to the *CHD2* TSS, and which represses *Chd2* gene expression in cis [151]. It has been hypothesized that targeting CHASERR could be used to increase expression of *CHD2* in haploinsufficient individuals [151], and it will be interesting to explore whether targeting the enhancer region of *CHD2* that we find and validate here could be exploited as an alternative target of such a strategy. Similarly, the regulation by enhancers of other disease implicated genes that we validate here adds to the list of potential targets to find disease causing non-coding variants that disturb this regulation.

An interesting finding of our study is that by starting with putative enhancers and variability of epigenome features over time during development, we recover DAEs and nDAEs that can be distinguished based on sequence characteristics, such as differences in GC content, the level of sequence constraint, tolerance to loss-of-function, and differential profiles of TF binding. Also, these DAEs and nDAEs seem to be associated with distinct developmental processes and result in differences in gene expression levels. It is tempting to speculate that the distinctive features between these two types of enhancers can be used to uncover key nucleotides responsible for those biological regulatory differences. It seems plausible that disturbing these functionally causative sequences could lead to altered physiology resulting in disease. Our analysis revealing GWAS loci enrichment and the link of DAEs supports this statement. We suggest that our results might help interpreting the effects of SNVs in non-coding sequences, which is at this stage not a trivial task. Our annotated database of DAE and nDAE will be instrumental to prioritize SNVs based on distinct sequence characteristics identified for these elements as well as to provide cues on potentially disturbed developmental processes based on differential temporal activity and regulatory targets of the enhancer in question. This in turn can instruct

functional validation and help deciphering pathogenicity of variants. With an increasing number of whole genome sequencing data available, it is expected that more, possibly disease implicated, non-coding variants will be identified, and the need to classify those sequences in benign or pathogenic will only further increase. With more computational pathogenicity prediction tools available, such as the ncER score and outcomes of integrative analyses such as performed here that pinpoint likely functional sequences, it might become possible to further decipher the impact of these SNVs.

Conclusions

In this study, by using an integrative computational analysis of virtually all previously described putative enhancers and epigenome datasets, we identified a comprehensive compendium of likely functional enhancers that are involved human brain development and disease. By applying CRISPRi-based silencing and zebrafish enhancer reporter assays, we show that these putative regions possess enhancer characteristics. We foresee that these enhancer sequences will be instrumental in identifying disease causing variants which might explain parts of the missing heritability in the field of clinical genetics.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-021-00980-1>.

Additional file 1: Fig. S1: Flow chart of integrative data analysis; **Fig. S2:** Derivation of pCRs and DAEs; **Fig. S3:** Enhancer-Gene predictions and target gene expression; **Fig. S4:** Features and motifs in DAEs and nDAEs; **Fig. S5:** Cell type specificity of DAEs and nDAEs; **Fig. S6:** Dynamics of DAEs and nDAEs in comparison to adult brain; **Fig. S7:** DAEs and nDAEs in human disease, related to Fig. 5; **Fig. S8:** Zebrafish enhancer reporter assay

Additional file 2: Table S1: List of all putative enhancers collected

Additional file 3: Table S2: Overview of all epigenome data processed in this study

Additional file 4: Table S3: List of all pCRs, with DAEs and nDAEs indicated

Additional file 5: Table S4: Functional enrichment analysis using GREAT for DAEs and nDAEs

Additional file 6: Table S5: enhancer-gene predictions

Additional file 7: Table S6: Overview of all RNA-seq data sets used in this study

Additional file 8: Table S7: Functional enrichment analysis using GREAT, Enrichr and Metascape for multigene interacting DAEs and nDAEs

Additional file 9: Table S8: TF and TE enrichment at DAEs and nDAEs

Additional file 10: Table S9: DAE clusters and associated functional enrichment using Enrichr

Additional file 11: Table S10: DAEs linked to OMIM genes

Additional file 12: Table S11: Significant GWAS loci used in this study

Additional file 13: Table S12: Calculations and *p*-values for gene variant associations

Additional file 14: Table S13: Zebrafish quantifications

Additional file 15: Table S14: Oligonucleotides used in this study

Acknowledgements

We would like to dedicate this paper to prof. Robert M.W. Hofstra, head of the department of Clinical Genetics at Erasmus MC, who sadly passed away during the preparation of this work. Ramon Birnbaum (Ben-Gurion University of the Negev, Israel) is acknowledged for sharing zebrafish reporter plasmids, and Raymond Poot (Erasmus MC, The Netherlands) for providing neural stem cells. We would like to thank the following third parties for providing approved access to the indicated data sets that were used to support the findings presented in this study:

-phs000755.v2.p1: "BrainSpan Atlas of the Human Brain". The datasets used for the analysis described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number phs000406.v2.p1. Submission of the data, phs000406.v2.p1, to dbGaP was provided by Dr. Nenad Sestan. Collection of the data and analysis was supported by grants from the National Institutes of Health (MH089929, MH081896, and MH090047). Additional support was provided by the Kavli Foundation, a James S. McDonnell Foundation Scholar Award, NARSAD, and the Foster-Davis Foundation.

-phs000791.v1.p1: "Roadmap Epigenomics Program - UCSF". Funding support for the NIH Roadmap Epigenomics Program was provided through the NIH Common Fund (Office of Strategic Coordination). Support for collection of datasets and samples was provided by a series of UO1 cooperative agreements with The Broad Institute [1U01ES017155-01], The Ludwig Institute for Cancer Research [1U01ES017166-01], The University of California San Francisco [1U01ES017154-01], and The University of Washington [1U01ES017156-01]. Data analysis and coordination were supported by an agreement with Baylor College of Medicine [1U01DA025956-01]. Assistance with data curation was supplied by GEO, and data access and visualization was supported by the NCBI.

-phs001226.v1.p1: "Regulatory Genomics of Human Embryonic Development". The datasets used for the analysis described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number phs001226.v1.p1.

-phs001438.v1.p1: "The dynamic landscape of open chromatin during human cortical neurogenesis". Datasets from this study used for the analyses described in this manuscript were generated in the Geschwind laboratory and supported by NIH grants to D.H.G. (5R01MH060233; 5R01MH100027; 3U01MH103339; 1R01MH110927; 1R01MH094714). Datasets were obtained from dbGaP found at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP study accession numbers phs001438.v1.p1.

-Data access from PsychENCODE was obtained for the following data sets: SynapseID: syn12033248 [9]; SynapseID: syn17092080 [86]. Data were generated as part of the PsychENCODE Consortium supported by: U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to: Shahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (USC), Mark Gerstein (Yale), Daniel Geschwind (UCLA), Thomas M. Hyde (LIBD), Andrew Jaffe (LIBD), James A. Knowles (USC), Chunyu Liu (UIC), Daliya Pinto (Icahn School of Medicine at Mount Sinai), Nenad Sestan (Yale), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (UCSF), Patrick Sullivan (UNC), Flora Vaccarino (Yale), Sherman Weissman (Yale), Kevin White (UChicago) and Peter Zandi (JHU).

Authors' contributions

SY performed primary computational analysis, with help of RD, KL, AN, and EP performed validation experiments in cells. EMS, HCvSL, and TvH performed zebrafish validation experiments. SY, EM, and TSB wrote the manuscript with input from all authors. All authors read and approved the final manuscript. EM and TSB conceived and jointly supervised the work.

Funding

RD is supported by a China Scholarship Council (CSC) PhD Fellowship (201906300026) for her PhD studies at the Erasmus Medical Center, Rotterdam, the Netherlands. EM is supported by Netherlands Organisation for Scientific Research (ZonMw Off Road grant). TSB is supported by the

Netherlands Organisation for Scientific Research (ZonMw Veni, grant 91617021), a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation, an Erasmus MC Fellowship 2017, and Erasmus MC Human Disease Model Award 2018. Funding bodies did not have any influence on study design, results, and data interpretation or final manuscript.

Availability of data and materials

All data generated in this study is included in this published article and its supplementary files. Tables Additional file 2: Table S1, Additional file 3: Table S2 and Additional file 7: Table S6 summarizes all primary data used for the meta-analysis. Some of the primary data that were used to support the findings of this study are available from dbGaP and PsychENCODE, but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available (third party data). The source codes and all processed data for all analysis performed in this study are available in the repositories <https://github.com/syousefi87/Differentially-Active-Enhancers> [152], and <https://figshare.com/projects/Differentially-Active-Enhancers/122965> [153].

Declarations

Ethics approval and consent to participate

For zebrafish studies, no larvae older than 5 days post fertilization were used. Zebrafish were kept according to guidelines of the Erasmus MC animal welfare office and all zebrafish experiments were performed in compliance with Dutch animal welfare legislation.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 28 April 2021 Accepted: 29 September 2021

Published online: 19 October 2021

References

- Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet.* 2012;13:613–26.
- Nord AS, West AE. Neurobiological functions of transcriptional enhancers. *Nat Neurosci.* 2020;23:5–14.
- D'Haene E, Vergult S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet Med.* 2021;23:34–46.
- Perenthaler E, Yousefi S, Niggli E, Barakat TS. Beyond the exome: the non-coding genome and enhancers in neurodevelopmental disorders and malformations of cortical development. *Front Cell Neurosci.* 2019;13:352.
- Carullo NVN, Day JJ. Genomic enhancers in brain health and disease. *Genes (Basel).* 2019;10(1):43.
- Chatterjee S, Ahituv N. Gene regulatory elements, major drivers of human disease. *Annu Rev Genomics Hum Genet.* 2017;18:45–63.
- Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489:57–74.
- Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518:317–30.
- Amiri A, Coppola G, Scuderi S, Wu F, Roychowdhury T, Liu F, et al. Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science.* 2018;362(6420):eaat6720.
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol.* 2010;28:1045–8.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507:455–61.
- Townsley KG, Brennand KJ, Huckins LM. Massively parallel techniques for cataloging the regulome of the human brain. *Nat Neurosci.* 2020;23:1509–21.
- Ryan GE, Farley EK. Functional genomic approaches to elucidate the role of enhancers during development. *Wiley Interdiscip Rev Syst Biol Med.* 2020; 12:e1467.

14. Montalbano A, Canver MC, Sanjana NE. High-throughput approaches to pinpoint function within the noncoding genome. *Mol Cell*. 2017;68:44–59.
15. Klefogiannis D, Kalnis P, Bajic VB. Progress and challenges in bioinformatics approaches for enhancer identification. *Brief Bioinform*. 2016;17:967–79.
16. Rojano E, Seoane P, Ranea JAG, Perkins JR. Regulatory variants: from detection to predicting impact. *Brief Bioinform*. 2019;20:1639–54.
17. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28:495–501.
18. Yan L, Guo H, Hu B, Li R, Yong J, Zhao Y, et al. Epigenomic landscape of human fetal brain, heart, and liver. *J Biol Chem*. 2016;291:4386–98.
19. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, et al. The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell*. 2018;172(289–304):e218.
20. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508:199–206.
21. Gao T, Qian J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res*. 2020;48:D58–64.
22. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464.
23. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res*. 2007;35:D88–92.
24. Vermunt MW, Reinink P, Korving J, de Bruijn E, Creyghton PM, Basak O, et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep*. 2014;9:767–79.
25. Valensisi C, Andrus C, Buckberry S, Doni Jayavelu N, Lund RJ, Lister R, et al. Epigenomic landscapes of hESC-derived neural rosettes: modeling neural tube formation and diseases. *Cell Rep*. 2017;20:1448–62.
26. Sun W, Poschmann J, Cruz-Herrera Del Rosario R, Parikshak NN, Hajan HS, Kumar V, et al. Histone acetylo-me-wide association study of autism spectrum disorder. *Cell*. 2016;167(1385–1397):e1311.
27. Emera D, Yin J, Reilly SK, Gockley J, Noonan JP. Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc Natl Acad Sci U S A*. 2016;113:E2617–26.
28. Vermunt MW, Tan SC, Castelijn B, Gieven G, Reinink P, de Bruijn E, et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat Neurosci*. 2016;19:494–503.
29. Yao P, Lin P, Gokoolparsadh A, Assareh A, Thang MW, Voineagu I. Coexpression networks identify brain region-specific enhancer RNAs in the human brain. *Nat Neurosci*. 2015;18:1168–74.
30. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016;538:523–7.
31. Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond Ser B Biol Sci*. 2013;368:20130025.
32. Reilly SK, Yin J, Ayoub AE, Emera D, Leng J, Cotney J, et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science*. 2015;347:1155–9.
33. Meuleman W, Muratov A, Rynes E, Halow J, Lee K, Bates D, et al. Index and biological spectrum of human DNase I hypersensitive sites. *Nature*. 2020; 584:244–51.
34. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, et al. Global reference mapping of human transcription factor footprints. *Nature*. 2020; 583:729–36.
35. Zhang J, Lee D, Dhiman V, Jiang P, Xu J, McGilivray P, et al. An integrative ENCODE resource for cancer genomics. *Nat Commun*. 2020;11:3696.
36. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*. 2014;47:11 12:11–34.
37. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*. 2010;26: 2204–7.
38. Wells A, Heckerman D, Torkamani A, Yin L, Sebat J, Ren B, et al. Ranking of non-coding pathogenic variants and putative essential regions of the human genome. *Nat Commun*. 2019;10:5241.
39. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15:1034–50.
40. Gussow AB, Copeland BR, Dhindsa RS, Wang Q, Petrovski S, Majoros WH, et al. Orion: detecting regions of the human non-coding genome that are intolerant to variation using population genetics. *PLoS One*. 2017;12: e0181604.
41. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310–5.
42. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536:285–91.
43. Xu D, Gokcumen O, Khurana E. Loss-of-function tolerance of enhancers in the human genome. *PLoS Genet*. 2020;16:e1008663.
44. Brandler WM, Antaki D, Gujral M, Kleiber ML, Whitney J, Maile MS, et al. Paternally inherited cis-regulatory structural variants are associated with autism. *Science*. 2018;360:327–31.
45. Monlong J, Girard SL, Meloche C, Cadieux-Dion M, Andrade DM, Lafreniere RG, et al. Global characterization of copy number variants in epilepsy patients from whole genome sequencing. *PLoS Genet*. 2018;14:e1007285.
46. Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, et al. Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nat Genet*. 2019;51:973–80.
47. Golding M. Adobe Illustrator CSS 5: for Web and Interactive Design; 2010.
48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
49. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11:R25.
50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
51. Cao Q, Anyansi C, Hu X, Xu L, Xiong L, Tang W, et al. Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nat Genet*. 2017;49:1428–36.
52. Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473:43–9.
53. Hait TA, Amar D, Shamir R, Elkon R. FOCS: a novel method for analyzing enhancer and gene activity patterns infers an extensive enhancer-promoter map. *Genome Biol*. 2018;19:56.
54. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*. 2017; bax028.
55. Corces MR, Shcherbina A, Kundu S, Gloudemans MJ, Fresard L, Granja JM, et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer’s and Parkinson’s diseases. *Nat Genet*. 2020; 52:1158–68.
56. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366:1134–9.
57. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. 2019;51:1664–9.
58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
59. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
60. Werling DM, Pochareddy S, Choi J, An JY, Sheppard B, Peng M, et al. Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell Rep*. 2020;31: 107489.
61. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
62. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523.
63. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*. 2016;32: 587–9.
64. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-

- regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
65. Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, et al. Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell*. 2018;23(276-288):e278.
 66. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. 2005;21:3439–40.
 67. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47:1228–35.
 68. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–5.
 69. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51:414–30.
 70. Duncan L, Yilmaz Z, Gaspar H, Walters R, Goldstein J, Anttila V, et al. Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am J Psychiatry*. 2017;174:850–8.
 71. Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry*. 2016;21:1391–9.
 72. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51:63–75.
 73. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51:431–44.
 74. Bipolar D. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address drve, Bipolar D, Schizophrenia Working Group of the Psychiatric Genomics C: Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*. 2018;173(1705-1715):e1716.
 75. International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address e-aeua: Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2014;13:893–903.
 76. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet*. 2019;51:394–403.
 77. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50:912–9.
 78. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50:668–81.
 79. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet*. 2018;50:920–7.
 80. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*. 2016;48:624–33.
 81. Yu D, Sul JH, Tsetsos F, Nawaz MS, Huang AY, Zelaya I, et al. Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. *Am J Psychiatry*. 2019;176:217–27.
 82. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18:1091–102.
 83. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. 2018;50:381–9.
 84. Banerjee-Basu S, Packer A. SFARI Gene: an evolving database for the autism research community. *Dis Model Mech*. 2010;3:133–5.
 85. Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, et al. A human cell atlas of fetal chromatin accessibility. *Science*. 2020;370(6518):eaba7612.
 86. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. 2018;362(6420):eaat7615.
 87. Perenthaler E, Nikoncuk A, Yousefi S, Berdowski WM, Alsagob M, Capo I, et al. Loss of UGP2 in brain leads to a severe epileptic encephalopathy, emphasizing that bi-allelic isoform-specific start-loss mutations of essential genes can cause genetic diseases. *Acta Neuropathol*. 2020;139:415–42.
 88. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013;339:1074–7.
 89. Daneshvar K, Pondick JV, Kim BM, Zhou C, York SR, Macklin JA, et al. DIGIT is a conserved long noncoding RNA that regulates GSC expression to control definitive endoderm differentiation of embryonic stem cells. *Cell Rep*. 2016;17:353–65.
 90. Yeo NC, Chavez A, Lance-Byrne A, Chan Y, Menn D, Milanova D, et al. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat Methods*. 2018;15:611–6.
 91. Westerfield M. *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio)*; 2000.
 92. D'Haene E, Bar-Yaacov R, Bariah I, Vantomme L, Van Loo S, Cobos FA, et al. A neuronal enhancer network upstream of MEF2C is compromised in patients with Rett-like characteristics. *Hum Mol Genet*. 2019;28:818–27.
 93. Berthelot C, Villar D, Horvath JE, Odom DT, Flicek P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat Ecol Evol*. 2018;2:152–63.
 94. Lecellier CH, Wasserman WW, Mathelier A. Human enhancers harboring specific sequence composition, activity, and genome organization are linked to the immune response. *Genetics*. 2018;209:1055–71.
 95. Colbran LL, Chen L, Capra JA. Short DNA sequence patterns accurately identify broadly active human enhancers. *BMC Genomics*. 2017;18:536.
 96. Kumarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet*. 2010;42:631–4.
 97. Jacques PE, Jeyakani J, Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet*. 2013;9:e1003504.
 98. Glinisky G, Barakat TS. The evolution of Great Apes has shaped the functional enhancers' landscape in human embryonic stem cells. *Stem Cell Res*. 2019;37:101456.
 99. Notwell JH, Chung T, Heavner W, Bejerano G. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat Commun*. 2015;6:6644.
 100. Sanchez-Castillo M, Ruau D, Wilkinson AC, Ng FS, Hannah R, Diamanti E, et al. CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res*. 2015;43:D1117–23.
 101. Kiyota T, Kato A, Kato Y. Ets-1 regulates radial glia formation during vertebrate embryogenesis. *Organogenesis*. 2007;3:93–101.
 102. Gainous TB, Wagner E, Levine M. Diverse ETS transcription factors mediate FGF signaling in the Ciona anterior neural plate. *Dev Biol*. 2015;399:218–25.
 103. 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.
 104. 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(1573-1588):e1528.
 105. Gabriele M, Vulto-van Silfhout AT, 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.
 106. Gregor A, Oti M, Kouwenhoven EN, Hoyer J, Sticht H, Ekici AB, et al. De novo mutations in the genome organizer CTCF cause intellectual disability. *Am J Hum Genet*. 2013;93:124–31.
 107. Nakada C, Satoh S, Tabata Y, Arai K, Watanabe S. Transcriptional repressor foxl1 regulates central nervous system development by suppressing shh expression in zebra fish. *Mol Cell Biol*. 2006;26:7246–57.

108. Mullen RD, Park S, Rhodes SJ. A distal modular enhancer complex acts to control pituitary- and nervous system-specific expression of the LHX3 regulatory gene. *Mol Endocrinol*. 2012;26:308–19.
109. Savage JJ, Hunter CS, Clark-Sturm SL, Jacob TM, Pfaffle RW, Rhodes SJ. Mutations in the LHX3 gene cause dysregulation of pituitary and neural target genes that reflect patient phenotypes. *Gene*. 2007;400:44–51.
110. Pristera A, Lin W, Kaufmann AK, Brimblecombe KR, Threlfell S, Dodson PD, et al. Transcription factors FOXA1 and FOXA2 maintain dopaminergic neuronal properties and control feeding behavior in adult mice. *Proc Natl Acad Sci U S A*. 2015;112:E4929–38.
111. Stott SR, Metzakopian E, Lin W, Kaestner KH, Hen R, Ang SL. Foxa1 and foxa2 are required for the maintenance of dopaminergic properties in ventral midbrain neurons at late embryonic stages. *J Neurosci*. 2013;33:8022–34.
112. Hung CY, Hsu TI, Chuang JY, Su TP, Chang WC, Hung JJ. Sp1 in astrocyte is important for neurite outgrowth and synaptogenesis. *Mol Neurobiol*. 2020; 57:261–77.
113. Manzanares M, Trainor PA, Nonchev S, Ariza-McNaughton L, Brodie J, Gould A, et al. The role of kreisler in segmentation during hindbrain development. *Dev Biol*. 1999;211:220–37.
114. Bianchi B, Kelly LM, Viemari JC, Lafon I, Burnet H, Bevengut M, et al. MafB deficiency causes defective respiratory rhythmogenesis and fatal central apnea at birth. *Nat Neurosci*. 2003;6:1091–100.
115. Koshida R, Oishi H, Hamada M, Takei Y, Takahashi S. MafB is required for development of the hindbrain choroid plexus. *Biochem Biophys Res Commun*. 2017;483:288–93.
116. Pai EL, Vogt D, Clemente-Perez A, McKinsey GL, Cho FS, Hu JS, et al. MafB and c-Maf Have prenatal compensatory and postnatal antagonistic roles in cortical interneuron fate and function. *Cell Rep*. 2019;26(1157–1173):e1155.
117. Maimaiti S, Koshida R, Ojima M, Kulathunga K, Oishi H, Takahashi S. Neuron-specific MafB knockout causes growth retardation accompanied by an impaired growth hormone/insulin-like growth factor I axis. *Exp Anim*. 2019;68:435–42.
118. Wang H, Xiao Z, Zheng J, Wu J, Hu XL, Yang X, et al. ZEB1 represses neural differentiation and cooperates with CTBP2 to dynamically regulate cell migration during neocortex development. *Cell Rep*. 2019;27(2335–2353):e2336.
119. Jiang Y, Yan L, Xia L, Lu X, Zhu W, Ding D, et al. Zinc finger E-box-binding homeobox 1 (ZEB1) is required for neural differentiation of human embryonic stem cells. *J Biol Chem*. 2018;293:19317–29.
120. Aslanpour S, Han S, Schuurmans C, Kurrasch DM. Neurog2 acts as a classical proneural gene in the ventromedial hypothalamus and is required for the early phase of neurogenesis. *J Neurosci*. 2020;40:3549–63.
121. Mulvaney J, Dabdoub A. Atoh1, an essential transcription factor in neurogenesis and intestinal and inner ear development: function, regulation, and context dependency. *J Assoc Res Otolaryngol*. 2012;13:281–93.
122. Pataskar A, Jung J, Smialowski P, Noack F, Calegari F, Straub T, et al. NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J*. 2016;35:24–45.
123. Xin M, Yue T, Ma Z, Wu FF, Gow A, Lu QR. Myelinogenesis and axonal recognition by oligodendrocytes in brain are uncoupled in Olig1-null mice. *J Neurosci*. 2005;25:1354–65.
124. Silbereis JC, Nobuta H, Tsai HH, Heine VM, McKinsey GL, Meijer DH, et al. Olig1 function is required to repress dlx1/2 and interneuron production in Mammalian brain. *Neuron*. 2014;81:574–87.
125. Jakovcevski I, Zecevic N. Olig transcription factors are expressed in oligodendrocyte and neuronal cells in human fetal CNS. *J Neurosci*. 2005;25: 10064–73.
126. Chen T, Wu Q, Zhang Y, Lu T, Yue W, Zhang D. Tcf4 controls neuronal migration of the cerebral cortex through regulation of Bmp7. *Front Mol Neurosci*. 2016;9:94.
127. Whalen S, Heron D, Gaillon T, Moldovan O, Rossi M, Devillard F, et al. Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: clinical score and further delineation of the TCF4 mutational spectrum. *Hum Mutat*. 2012; 33:64–72.
128. Hegeudus B, Dasgupta B, Shin JE, Emmett RJ, Hart-Mahon EK, Elghazi L, et al. Neurofibromatosis-1 regulates neuronal and glial cell differentiation from neuroglial progenitors in vivo by both cAMP- and Ras-dependent mechanisms. *Cell Stem Cell*. 2007;1:443–57.
129. Shaulian E, Karin M. AP-1 as a regulator of cell life and death. *Nat Cell Biol*. 2002;4:E131–6.
130. Werner H, LeRoith D. Insulin and insulin-like growth factor receptors in the brain: physiological and pathological aspects. *Eur Neuropsychopharmacol*. 2014;24:1947–53.
131. Russ BE, Olshansky M, Li J, Nguyen MLT, Gearing LJ, Nguyen THO, et al. Regulation of H3K4me3 at transcriptional enhancers characterizes acquisition of virus-specific CD8(+) T cell-lineage-specific function. *Cell Rep*. 2017;21:3624–36.
132. Paredes I, Himmels P, Ruiz de Almodovar C. Neurovascular communication during CNS development. *Dev Cell*. 2018;45:10–32.
133. Monier A, Evrard P, Gressens P, Verney C. Distribution and differentiation of microglia in the human encephalon during the first two trimesters of gestation. *J Comp Neurol*. 2006;499:565–82.
134. Zhang X, He X, Li Q, Kong X, Ou Z, Zhang L, et al. PI3K/AKT/mTOR signaling mediates valproic acid-induced neuronal differentiation of neural stem cells through epigenetic modifications. *Stem Cell Reports*. 2017;8:1256–69.
135. Sanchez-Alegria K, Flores-Leon M, Avila-Munoz E, Rodriguez-Corona N, Arias C. PI3K signaling in neurons: a central node for the control of multiple functions. *Int J Mol Sci*. 2018;19(12):3725.
136. Jayaraman D, Bae BI, Walsh CA. The genetics of primary microcephaly. *Annu Rev Genomics Hum Genet*. 2018;19:177–200.
137. Oegema R, Barakat TS, Wilke M, Stouffs K, Amrom D, Aronica E, Bahi-Buisson N, Conti V, Fry AE, Geis T, et al. International consensus recommendations on the diagnostic work-up for malformations of cortical development. *Nat Rev Neurol* 2020, 16:618–635.
138. Wang X, Goldstein DB. Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. *Am J Hum Genet*. 2020;106:215–33.
139. Kvon EZ, Waymack R, Elabd MG, Wunderlich Z. Enhancer redundancy in development and disease. *Nat Rev Genet*. 2021;22(5):324–36.
140. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.
141. Wilson MM, Henshall DC, Byrne SM, Brennan GP. CHD2-related CNS pathologies. *Int J Mol Sci*. 2021;22(2):588.
142. Thisse B, Thisse C. Fast release clones: a high throughput expression analysis. ZFIN Direct Data Submission (<http://zfin.org>). 2004. Available: <https://zfin.org/ZDB-PUB-040907-1>.
143. Shen T, Ji F, Yuan Z, Jiao J. CHD2 is required for embryonic neurogenesis in the developing cerebral cortex. *Stem Cells*. 2015;33:1794–806.
144. Rymen D, Lindhout M, Spanou M, Ashrafzadeh F, Benkel I, Betzler C, et al. Expanding the clinical and genetic spectrum of CAD deficiency: an epileptic encephalopathy treatable with uridine supplementation. *Genet Med*. 2020;22:1589–97.
145. Thisse B, Pflumio S, Fürthauer M, Loppin B, Heyer V, Degraeve A, et al. Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). ZFIN Direct Data Submission (<http://zfin.org>). 2001. Available: <https://zfin.org/ZDB-PUB-010810-1>.
146. Sagie S, Lerman-Sagie T, Maljevic S, Yosovich K, Detert K, Chung SK, et al. Expanding the phenotype of TRAK1 mutations: hyperekplexia and refractory status epilepticus. *Brain*. 2018;141:e55.
147. Barel O, Malicdan MCV, Ben-Zeev B, Kandel J, Pri-Chen H, Stephen J, et al. Deleterious variants in TRAK1 disrupt mitochondrial movement and cause fatal encephalopathy. *Brain*. 2017;140:568–81.
148. Dobyms WB, Aldinger KA, Ishak GE, Mirzaa GM, Timms AE, Grout ME, et al. MACF1 mutations encoding highly conserved zinc-binding residues of the GAR domain cause defects in neuronal migration and axon guidance. *Am J Hum Genet*. 2018;103:1009–21.
149. Brock S, Vanderhasselt T, Vermaning S, Keymolen K, Regal L, Romaniello R, et al. Defining the phenotypical spectrum associated with variants in TUBB2A. *J Med Genet*. 2021;58:33–40.
150. Philips AK, Pinelli M, de Bie CI, Mustonen A, Maatta T, Arts HH, et al. Identification of C12orf4 as a gene for autosomal recessive intellectual disability. *Clin Genet*. 2017;91:100–5.
151. Rom A, Melamed L, Gil N, Goldrich MJ, Kadir R, Golan M, et al. Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nat Commun*. 2019;10:5092.
152. Yousefi S, Deng R, Lanko K, Medico Salsench E, Nikoncuk A, Van der Linde HC, et al. Differentially-active-enhancers. GitHub. <https://github.com/yousefi87/Differentially-Active-Enhancers>. 2021.
153. Yousefi S, Deng R, Lanko K, Medico Salsench E, Nikoncuk A, Van der Linde HC, et al. Differentially-active-enhancers. figshare. <https://figshare.com/projects/Differentially-Active-Enhancers/122965>. 2021.

Publisher's Note

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