Supplemental Methods

Human DNase-seq and RNA-seq data

DNase-seq and RNA-seq data from human fetal tissues of week 10-20 generated within the Roadmap Epigenomics project (Bernstein et al., 2010) were downloaded from the NCBI's Sequence Read Archive (Dec. 15, 2014, summary table on GitHub).

cis-regulatory element (CRE) region determination and tissue-specificity scoring

DNase-seq reads were mapped to human genome version GRCh37 using NextGenMap (Sedlazeck et al., 2013, version 0.0.1). Aside from a few exceptions (dualstrand=1; min identity=0.9; min residues=0.5), the default parameters were used. PCR duplicates were removed using SAMtools rmdup (H Li and Durbin, 2009, version 1.1). We used JAMM (Ibrahim et al., 2015, version 1.0.7) to call peaks per tissue considering the biological replicates for the DNase-seq data using the recommended settings. To compare peaks across tissues, we merged overlapping peaks using the resulting union peaks as putative CRE, which are the basis of most further analyses. We removed peaks mapping to Y or MT chromosomes. Furthermore, we removed 26 CREs whose width exceeded 5000 bp ($< 0.0001\%$), resulting in a set of 465,281 CREs. We then used the number of overlapping peaks, i.e the number of tissues in which a CRE is accessible as a proxy for pleiotropy. This score ranges between 1 (tissue-specific) to 9 (ubiquitously open).

After the initial tissue assignment by simple presence-absence, we filtered elements that were not significantly DA between tissues where we assigned them as active vs. all others. For each of our CREs and each human tissue sample, we counted the reads falling within the coordinates of the CRE. Further, for each combination of tissues and PD class (2^9 - 2 = 510 in total), we performed DA analyses using DESeq2 (Love et al., 2014) between the tissues

where the CREs were detected as accessible versus the tissues where no peak was called. In addition, we also generated controls by shuffling the order of labels in each DA contrast (repeated 10×). For PD 1-6, the proportion of CREs that are significantly (BH-adjusted p-value$<$ 0.1) more accessible in the assigned tissues compared to other tissues is $>=$ 0.95, while in PD7 and PD8 it drops to 0.89 and 0.73, respectively. We excluded the ambiguously assigned CREs, resulting in a total of 443,188 (95%) remaining CREs.

Human and cynomolgus macaque iPSC differentiation into NPCs

iPSCs from human and cynomolgus macaque (Geuder et al., 2021) were differentiated to neural progenitor cells via dual-SMAD inhibition as three-dimensional aggregation culture (Chambers et al., 2009; Ohnuki et al., 2014). Briefly, iPSCs were dissociated and 9×10^3 iPSCs were seeded in a low attachment U-bottom 96-well-plate in 8GMK medium consisting of GMEM (Thermo Fisher), 8% KSR (Thermo Fisher), 5.5 ml 100× NEAA (Thermo Fisher), 100 mM Sodium Pyruvate (Thermo Fisher), 50 mM 2-Mercaptoethanol (Thermo Fisher) supplemented with 500 nM A-83–01 (Sigma Aldrich), 100 nM LDN 193189 (Sigma Aldrich) and 30 µM Y27632 (biozol). Culture medium of the spheres was changed every second day until they were harvested or plated for further culture. In order to obtain stable NPC lines, spheres were dissociated on day 7 of the differentiation process using Accumax (Sigma Aldrich) and plated onto Geltrex (Thermo Fisher) coated dishes. NPCs were subsequently cultured in NPC proliferation medium (DMEM F12 (Fisher Scientific) supplemented with 2 mM GlutaMAX-I (Fisher Scientific), 20 ng/mL bFGF (Peprotech), 20 ng/mL hEGF (Miltenyi Biotec), 2% B-27 supplement (50×) minus vitamin A (Gibco), 1% N2 supplement 100× (Gibco), 200 µM L-ascorbic acid 2-phosphate (Sigma), and 100U/ml 100µg/ml penicillin-streptomycin). All cell lines have been authenticated using RNA sequencing (RNA-seq) (Geuder et al., 2021), and the current study.

RNA-seq data generation and processing

Samples for RNA-seq were taken from 3 clones of 3 human individuals and 4 clones of 2 cynomolgus macaque individuals at the iPSC stage (time point 0) and after 1, 5, 7 and 9 days during the neural maturation process. Spheres were dissociated at each time point using Accumax (Sigma Aldrich) and live cells were sorted using the BD FACS Aria II.

cDNA libraries for samples from the different species and differentiation time points were generated using the prime-seq protocol (Janjic et al., 2022) and we obtained 100 bp cDNA, 10 bp UMI and 6 bp sample barcode reads from an Illumina HiSeq 1500. We used functions `bbduk` to filter out reads that have low sequence complexity (estimated entropy<0.5) and `repair` to pair the remaining reads from BBTools, BBMap v. 38.02 (Bushnell, 2014). Further processing is described in the Methods under "Cross-species gene expression and accessibility analysis".

ATAC-seq data generation and processing

iPSCs of 2 clones from 2 human individuals and 2 clones of 2 cynomolgus macaque individuals were differentiated using the protocol as described above. The NPC lines were cultured in NPC proliferation medium and passaged 2-4× until they were dissociated and subjected to ATAC-seq together with the respective iPSC clones.

ATAC-seq libraries were generated using the Omni-ATAC protocol (Corces et al., 2017) with minor modifications. In brief, cells were washed with PBS and dissociated using Accumax (Sigma Aldrich) for iPSCs or TrypleSelect (Thermo Fisher) for NPCs at 37°C for 5-10 min. After cells were counted, 100,000 cells were pelleted at 500 rcf for 5 min, washed with 1 ml PBS and pelleted at 500 rcf for 5 min at 4°C. The supernatant was removed completely and cells were resuspended in 100 µl chilled nuclei lysis buffer (10 mM Tris-HCl pH7.4, 10 mM NaCl, 3 mM MgCl2 in water, supplemented with 0.1% Tween-20, 0.1% NP-40, 0.01% Digitonin and 1% BSA) by pipetting up and down 3×, followed by incubation on ice for 3

min. After lysis, 1 ml of lysis wash buffer (10 mM Tris-HCl pH7.4, 10 mM NaCl, 3 mM MgCl2 in water, supplemented with 0.1% Tween-20 and 1% BSA) was added, and tubes were inverted 3×. After counting, 50,000 nuclei were pelleted at 500 rcf for 10 min at 4°C, the supernatant was removed and nuclei were resuspended in 50 µl transposition mix (25 µl 2x TD buffer, 2.5 µl TDE1, 16.5 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20 and 5 µl ddH2O) by pipetting 6×. Transposition reactions were incubated at 37°C for 1 h at 1000 rpm shaking, followed by a clean-up using the DNA Clean & Concentrator-5 kit (Zymo). For library generation, 20 µl of the transposed sample was mixed with 2.5 µl p5 custom primer (25 µM), 2.5 µl p7 custom primer (25 µM) (Buenrostro et al., 2013) and 25 µl NEBNext Ultra II Q5 2x Master Mix (NEB) and a PCR with 10 cycles was conducted as stated in the Omni-ATAC protocol. Libraries were purified using the DNA Clean & Concentrator-5 kit, run on a 2% E-Gel (Thermo Fisher) and gel excision of DNA between 150 bp and 1,500 bp was performed using the Monarch DNA Gel Excision Kit (NEB). Concentrations of the purified libraries were measured using PicoGreen (Thermo Fisher) and quality was assessed using a Bioanalyzer High-Sensitivity DNA Analysis Kit (Agilent). Libraries were pooled and sequenced on NovaSeq 6000 instrument with the following setup: R1: 151, i7: 8, R2: 151 cycles.

Sequenced human and cynomolgus macaque reads were mapped to GRCh38 and macFas6 genomes, respectively. For mapping, we used BWA-MEM2 (Vasimuddin et al., 2019, version 2.0pre2), using the following command: `bwa-mem2 mem -M -t 20 -I 250,150`. Furthermore, `samtools fixmate -m - -` and `samtools sort` commands were applied (Li and Durbin, 2009, version 1.11). Peak calling was performed using Genrich (`https://github.com/jsh58/Genrich`) on the 2 biological replicates per species per cell type. We applied the following parameter settings: `-j -y -r -q 0.05 -a 200 -e MT,Y -E $blacklist -s 20`, where as a `$blacklist` the ENCODE blacklist with GRCh38 coordinates (Amemiya et al., 2019) was supplied for human (910 regions), and a reciprocal liftOver version of it to

macFas6 (558 regions) was supplied for the peak-calling in macFas6 genome space.

liftOver file generation

Human genome liftOver files were downloaded from USCS
`https://hgdownload.soe.ucsc.edu/gbdb/hg19/liftOver/hg19ToHg38.over.chain.gz`
`https://hgdownload.soe.ucsc.edu/goldenPath/hg38/liftOver/hg38ToHg19.over.chain.gz`.
liftOver files hg38toMacFas6 and macFas6toHg38 were generated from blastz alignments (Schwartz et al., 2003; Kent et al., 2003) of the canonical chromosomes from both genomes, as reported here (`https://genomewiki.ucsc.edu/index.php?title=DoBlastzChainNet.pl`).
Further processing is described in the Methods under "Cross-species gene expression and accessibility analysis".

Data sources for estimation of CRE phylogenetic conservation and PD across mammals

Cross-species histone modification ChIP-seq data from Roller et al. (2021) were downloaded from `https://www.ebi.ac.uk/research/flicek/publications/FOG29/`. This data contains samples from rhesus macaque and marmoset, rabbit, rat, mouse, cat, dog, horse, pig and opossum, four tissues each.
All macaque CRE coordinates were collected from the 9 pairwise species files `Macaque_regRegions_allTissue_mainRegs_to_*_active.txt` where * indicates the other mammalian species.

Evolutionary sequence analysis of CREs

To be able to intepret evolution rates as a result of the genetic element's CRE activity, we excluded all CREs that overlapped CDSs (GENCODE v.19) in all sequence evolution analyses (6.6% of the gene-assigned CREs).

INSIGHT We ran the web tool INSIGHT (Gronau et al. (2013)) on the CRE or peak coordinates of each PD class in GRCh37 using the default settings. To re-calculate the evolutionary rates on various CRE subsets more efficiently, we downloaded the INSIGHT script `runINSIGHT-EM.sh` that applies expectation-maximization (EM) algorithm on the provided INSIGHT files (`.ins`) and the complementary flanking sequence INSIGHT files (`.flankPoly.forBetas.ins`). The scripts for subsetting the INSIGHT output files and re-calculating the evolutionary rates can be found on github.

phastCons and phyloP Pre-calculated 46-way GRCh37 phastCons and phyloP scores for the 10 primate subset were downloaded from
http://hgdownload.cse.ucsc.edu/goldenpath/hg19/phastCons46way/ and
http://hgdownload.cse.ucsc.edu/goldenpath/hg19/phyloP46way/
(versions from 2009-11-11) in a bigWig file format. For each CRE, the average conservation score was calculated for each conservation metric.

Parameter specification for quantification of transcription factor binding

To get position weight matrices, we used JASPAR 2020 collection, core vertebrate set (Fornes et al., 2020) using R packages `JASPAR2020` (version 0.99.10) and `TFBSTools` (Tan and Lenhard, 2016, version 1.36.0). These PWMs were provided to Cluster-Buster (Frith et al., 2003, downloaded on 2020-05-07). Cluster-Buster was ran with the following parameters: `-c0 -m0 -r10000 -b500 -f5`. The orthologous human and cynomolgus macaque CRE input sequences were extended by 500 bp in each direction, thereby providing an approximation of the background base composition (parameter `-b500`).

Gene set enrichment analysis

Gene-set enrichment analysis contrasting PD-overrepresented TF groups with the rest of the expressed TFs was conducted using the Bioconductor package `topGO` (Alexa and Rahnen-

fuhrer, n.d., version 2.50.0), setting the following parameters: `ontology="BP", nodeSize = 10, algorithm = "elim", statistic = "fisher"`. We ordered the enriched categories (Fisher's p-value$<$ 0.05) according to their Significant / Expected proportion.

Analysis of TF ChIP-seq data across mammals

Published data from Ballester et al. (2014) `https://doi.org/10.7554/eLife.02626.009` was downloaded, specifically, table elife-02626-fig2-data1-v1.xlsx. It contains four TF (CEBPA, FOXA1, HNF4A, ONECUT1) and *cis*-regulatory module (CRM) ChIP-seq peak coordinates in the human liver (n=75,589; hg18) and annotations on their binding conservation in the livers of 4 mammals. In the original publication, CRMs were defined as distinct ($>$ 1) TF binding co-occurence where the distance of the TF ChIP-seq summits are no more than 300 bp away.

To select PD9 CREs, CREs that are pleiotropic both in our study and across Roller et al. (2021) macaque tissues were used. To identify putative liver-specific CREs, we selected liver-specific CREs from Roller et al. (2021) and excluded all CREs that were accessible in the 9 tissues in our study. We overlapped the coordinates of PD1 (n=6,300) and PD9 (n=14,442) CREs in RheMac10 genome space with RLO TF ChIP-seq coordinates (n=68,417, 91%).

Alignment of human and macaque orthologous CREs

Orthologous human and macaque CRE sequences were pairwise aligned using MAFFT (Katoh and Standley, 2013, version 7) using the following parameters `--adjustdirection --maxiterate 1000 --auto`. We quantified alignment length (median 1273, 90% CI [1133, 1790]), fraction of mismatches in bp (median: 0.058, 90% CI [0.0357, 0.1432]), the fraction of indels in bp (median: 0.018, 90% CI [0.0031, 0.0916]) and the number of indels (median 6, 90% CI [2, 13]). We subsequently trimmed gaps in the remaining CRE alignments.