Supporting Online Material for:

A temporal *in-vivo* catalogue of chromatin accessibility and expression profiles in pineoblastoma reveals a prevalent role for repressor elements

Salam Idriss[1&], Mohammad Hallal[1,2,&], Abdullah El-Kurdi[1,3,&], Hasan Zalzali[4,5], Inaam El-Rassi[2], Erik A. Ehli[6], Christel M. Davis[6], Philip E. D. Chung[7,8], Deena M. A. Gendoo[9], Eldad Zacksenhaus[7,8,10], Raya Saab[4,5,11*], and Pierre Khoueiry[1,3,12*]

[1]Department of Biochemistry and Molecular Genetics, Faculty of Medicine, American University of Beirut, Beirut, Lebanon
[2] Biomedical Engineering Program, American University of Beirut, Lebanon
[3] Pillar Genomics Institute, Faculty of Medicine, American University of Beirut, Beirut, Lebanon
[4]Department of Pediatric and Adolescent Medicine, American University of Beirut, Beirut, Lebanon
[5]Department of Anatomy, Cell Biology, and Physiological Sciences, Faculty of Medicine, American University of Beirut, Beirut, Lebanon
[6]Avera Institute for Human Genetics, Sioux Falls, South Dakota, USA
[7]Toronto General Research Institute, University Health Network, 101 College Street, Toronto, ON M5G 1L7, Canada
[8]Department of Laboratory Medicine & Pathobiology, University of Toronto, Toronto, ON, Canada.
[9]Centre for Computational Biology, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK.
[10]Department of Medicine, University of Toronto, Toronto, ON, Canada
[11] Current address: Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA, USA
[12] Current address: Helmholtz Institute for RNA-based Infection Research (HIRI) - Josef-Schneider-Straße 2/D15, 97080 Würzburg
[&]These authors contributed equally
[*] Co-senior author

Supplemental Methods

H&E staining

For hematoxylin and eosin staining (H&E), tissues were fixed in 4% paraformaldehyde for 72 hours, then in 70% ethanol before being embedded in paraffin. Four (4) µm sections were deparaffinized and stained with hematoxylin and eosin. Representative images were taken at different magnification (×10, ×20 and ×40 oil) using a ZEISS Primo Vert Microscope. Stained slides were reviewed by a neuropathologist, and five slides from 3 different animals were examined from each time point.

RNA Extraction and Sequencing

Tissues designated for RNA extraction were "flash" frozen in liquid nitrogen and stored at -80°C. Total RNA was extracted from pineal glands collected on day 10, 49, and 90 from *Rbp3-Ccnd1/Trp53*$^{-/-}$ mice using the RNeasyPlus Mini Kit (Qiagen, 74134). Tissues were harvested in RLT Plus buffer and mechanically dissociated with needles of 18 G then 21 G and finally 25 G. Samples were then processed for RNA extraction according to the manufacturer's guidelines. RNA was quantified using NanoDrop and RNA integrity was verified with the 2100 Bioanalyzer prior to sequencing.

Total RNA was assessed for sample concentration and degradation on RNA 6000 Nano Chip ran on a 2100 Bioanalyzer (Agilent; Santa Clara, CA, 5067-1529) where the average RNA integrity (RIN) score for the samples set was 9.2. Sequencing libraries were prepared using the TruSeq Stranded Total RNA Library Prep Kit (Illumina, Inc; San Diego, CA, 20020598) with an input of 500 ng of total RNA following the low sample procedure. Briefly, ribosomal RNA (rRNA) was depleted from total RNA and the remaining RNA was purified, fragmented appropriately, and primed for cDNA synthesis. Blunt-ended cDNA was generated after first and second-strand synthesis. Adenylation of the 3' blunt-ends was followed by adapter

ligation prior to the enrichment of the cDNA fragments. Final library quality control was carried out by evaluating the fragment size on a DNA1000 chip ran on a 2100 Bioanalyzer (Agilent; Santa Clara, CA, 5067-1504). The concentration of each library was determined by quantitative PCR (qPCR) by the KAPA Library Quantification Kit for Next Generation Sequencing (KAPA Biosystems; Woburn, MA, KK4824) prior to sequencing.
Libraries were normalized to 2 nmol/L in 10 mM Tris-Cl, pH8.5 with 0.1% Tween-20 then pooled evenly. The pooled libraries were denatured with 0.1 N NaOH and diluted to 20 pmol/L. Cluster generation of the denatured libraries was performed according to the manufacturer's specifications (Illumina, Inc; San Diego, CA, FC-402-4021 & PE-402-4002) utilizing the HiSeq PE Cluster Kit v2 chemistry and flow cells. Libraries were clustered appropriately with a 1% PhiX spike-in. Sequencing-by-synthesis (SBS) was performed on a HiSeq 2500 utilizing v2 chemistry with paired-end 101-bp reads resulting in an average of 58.9 million paired-end reads per sample. Sequence read data were processed and converted to FASTQ format for downstream analysis by Illumina BaseSpace analysis software, FASTQ Generation v1.0.0.

Nuclei isolation for ATAC-seq

For ATAC-seq experiments, pineal gland tissues were cryopreserved in 200 µL CryoStor® cell cryopreservation media (Sigma-Aldrich C2874). Cryopreserved tissues were allowed to freeze in an isopropanol freezing chamber at -80°C. To isolate nuclei, pineal gland tissues were resuspended in 500 µL of 1× Homogenization buffer (30 mM $CaCl^2$, 18 mM Mg(Ac)2, 60 mM Tris pH 7.8, 320 mM Sucrose, 0.1 mM EDTA and 0.1% NP-40). Tissues were then homogenized followed by mechanical disruption with needles of 18 G then 21 G and finally 25 G. The homogenate was centrifuged at 500RCF for 7 min at 4°C. Supernatants were discarded and pellets containing the nuclei were resuspended in ice-cold PBS. Nuclei were

counted using Trypan blue staining and 50,000 nuclei were collected for ATAC-seq library preparation by centrifugation at 13000 rpm for 20 min, 4°C.

ATAC-seq and ChIP-seq processing and library preparation

Nuclei were subjected to transposition reaction using Nextera DNA Sample Preparation kit (Illumina, no. FC-121-1030). Briefly, 50,000 nuclei were pelleted after isolation, resuspended in 50µL transposition buffer (25 µL 2× TD buffer, 22.5 µL Nuclease-free H2O, 2.5 µL Illumina Tn5 transposase), and incubated at 37°C for 30 min as previously described (Buenrostro et al. 2015). Transposed DNA was purified with MinElute PCR Purification Kit (Qiagen, 28004), and eluted in 10 µL EB buffer (10 mM Tris buffer, pH 8).

ATAC-seq libraries were prepared by PCR amplification using Nextera Index Kit (Illumina, no. 15055289) and the NEBNext High-Fidelity 2× PCR Master Mix (NEB, M0541L) with 4–6 cycles. Libraries were quantified by PicoGreen DNA Quantification assay and sequenced on Illumina NovaSeq S2 using 2×50 paired-end reads in biological duplicates.

For chromatin isolation, tissues were disrupted as described above for RNA-seq and ATAC-seq. Tissues were cross-linked with 1% formaldehyde on an orbital rotator for 10 min at RT and subsequently quenched by 0.125 M glycine at RT for 5 min with mild rotation. Cross-linked tissues were lysed on ice with cell lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris, 1× Protease Inhibitor cocktail and PMSF) followed by nuclear lysis buffer (10 mM Tris, 1 mM EDTA, 1% SDS, and 1× Protease inhibitors and PMSF) to release nuclear chromatin. Chromatin was sonicated in the nuclear lysis buffer into 200bp -700bp fragments using Diagenode Bioruptor for 40 cycles with 30 sec. ON, 30 sec. OFF at high frequency at +4°C, yielding an average shearing size of 250–300 bp. Chromatin fragment sizes were checked using 1% conventional agarose gel.

For H3K27ac ChIP, 1 µg of antibody (Abcam ab4729) was incubated with 50 uL Dynabeads on an orbital rotator overnight. 25 µg of sheared chromatin was then added to the bead-antibody complex for incubation overnight. Beads were then washed using 4 washes of RIPA 500 buffer (500 mM NaCl, 1% Deoxycholate, 0.1% SDS, 1% Triton X-100, 1 mM EDTA and 10 mM Tris), 4 washes of LiCl buffer (500 mM LiCl, 1% Deoxycholate, 1% NP-40 and 100 mM Tris) and 1 wash with Tris-EDTA buffer (50 mM Tris pH 8 and 10 mM EDTA). All ChIP steps were performed in a cold room. Samples were then eluted in elution buffer (50 mM Tris, 10 mM EDTA and 1% SDS) followed by reverse cross-linking in parallel to 1% Input for each condition. DNA was extracted using Phenol:Chloroform and precipitated with ethanol. DNA recovered by IP and input samples were quantified using High Sensitivity Qubit Assay (Invitrogen, Q3285). The efficiency of immunoprecipitation (reflected by fold enrichment and recovery over input) was evaluated by RT-qPCR using positive and negative primers targeting mouse genomic regions that are potentially enriched or depleted for H3K27ac marks respectively. The primers used are: from *Gapdh* as a positive primer (gapdh_fwd AAGAAAGAAGCCCCGGACTG; gapdh_rev CTGCACCTCTGGTAACTCCG) and a gene desert region as a negative primer (Neg3_fwd AGTGCAAGGTTGTGGGTAAGA; Neg3_rev GCAATGCAGGATGGTGAAGT).
Prior to library preparation, ChIP DNA was quantified using PicoGreen DNA Quantification kit according to the manufacturer's instructions. ChIP-seq library preparation was performed using the Thruplex DNA-seq Kit. Sequencing was performed on the Illumina NovaSeq S1 (2×50 Paired-end). Two genomic input samples were sequenced for each condition. ATAC-seq and ChIP-seq libraries were sequenced at the University of Minnesota Genomics Center (UMGC).

RNA extraction and Quantitative real-time PCR

Total RNA was extracted using QIAzol reagent (Qiagen) according to the manufacturer's instructions. DNase-treated total RNA was used for cDNA synthesis with random hexamers using RevertAid 1st strand cDNA synthesis kit (Thermo Fisher Scientific). Real-time PCR was performed using the iTaq™ Universal SYBR® Green Supermix in a CFX96 system (Bio-Rad Laboratories). Products were amplified using the primers listed in Supplemental Table S17. PCR parameters consisted of denaturation at 95°C for 3 min followed by 40 cycles of 95°C for 15 sec, and 72°C for 1 min. Annealing temperature was 55°C for all transcripts. A final extension at 72°C for 10 min was performed followed by a melting curve, with temperature gradually increased (0.5°C) to 95°C. Standard curves were plotted using serially diluted cDNA. The geometric mean of housekeeping gene Gapdh was used as an internal normalization control and data analysis was performed using the ΔΔCT method. Primers used are listed in Supplemental Table S17.

RNA-seq, ATAC-seq, and ChIP-seq data analyses

Quality checks on raw FASTQ data were performed using FastQC v0.11.8 and MultiQC v1.9 (Ewels et al. 2016). For RNA-seq, sequence alignment to mm10 reference genome GRCm38.p5 from GENCODE (Frankish et al. 2019) was performed using HISAT2 (Kim et al. 2015) with parameters (--rna-strandedness RF). Alignment SAM files were converted into BAM format, sorted and indexed using SAMtools v1.4.1 (Li et al. 2009). Post alignment quality checks were done using RSeQC v2.6.4 (Wang et al. 2012) and deepTools v3.5.0 (Ramirez et al. 2014). Counts Per Million (CPM) normalized bigWig files were generated using deepTools with parameters (--effectiveGenomeSize 2652783500 --normalizeUsing "CPM" --binSize 20 --smoothLength 60). Raw read counts over genomic features was done using featureCounts v2.0.1 (Liao et al. 2014) and annotation GTF file from GENCODE v25 with parameters (-t gene -g gene_id -s 2 -p -B).

Genes whose sum of raw reads across all 9 replicates was greater than 10 reads were considered as expressed. Differential expression analysis was performed using DESeq2 v1.26.0 (Love et al. 2014) and genes with an absolute $\log_2$ fold change (LFC) > 1 and a p-adjusted < 0.05 were considered differentially expressed. Unsupervised *k*-means clustering of differentially expressed genes between time-points was done on *Z*-score transformed $\log_2$(TPM+1) mean expression values of the three biological replicates representing each time-point using Euclidean distance and Ward.D2 clustering method. Downstream Gene Ontology and over-representation analysis of the differentially expressed gene clusters was done using clusterProfiler v3.14.3 (Yu et al. 2012).

Copy-number variant calling from RNA-seq data was done using CaSpER v0.2.0 (Serin Harmanci et al. 2020) and amplification or deletion events predicted in all three replicates at each time-point were considered for further analysis. Gene fusion and tandem duplication events were predicted from RNA-seq data using Arriba v2.1.0 (Uhrig et al. 2021) and high confidence events common to all replicates within each time-point were considered for further analysis. Visualization of genomic tracks was done using the Integrative Genomics Viewer (IGV) v2.8.6 (Thorvaldsdottir et al. 2013) and plotting was done using pygenometracks v3.6 (Lopez-Delisle et al. 2021).

For ATAC-seq, adapter trimming was performed using Trimmomatic (Bolger et al. 2014) followed by sequence alignment using Bowtie 2 v2.3.4.1 (Langmead and Salzberg 2012) with parameters (-end-to-end -very-sensitive -maxins 2000). SAM files were converted into BAM format, sorted and indexed using SAMtools. Post alignment quality checks and correction for ATAC-seq transposase cut-site bias were performed using deepTools v3.5.0. Peak calling was done using MACS2 v2.2.7.1 (Zhang et al. 2008) with parameters (-f BAMPE -g mm -nomodel -q 0.05 --cutoff-analysis -call-summits) and blacklisted regions

retrieved from https://github.com/Boyle-Lab/Blacklist/tree/master/lists were filtered out. RPGC normalized bigWig files were generated using deepTools v3.5.0 with parameters (-effectiveGenomeSize 2652783500 -normalizeUsing "RPGC" -e -centerReads -binSize 20 -smoothLength 60). The best two representative samples for each time-point P10, P49, and P90 were selected for downstream analysis taking into consideration the pairwise Irreproducible Discovery Rate (IDR) values, Transcription Start Site (TSS) Enrichment as well as the Fraction of reads in peaks (FRIP) metric.

Consensus ATAC-seq peaks at each time-point were selected using the following approach: First, alignment BAM files for the selected two replicates representing each time-point were merged and peaks called using MACS2 with the same parameters as before. Next, peaks called using the merged BAM file and found in at least one of the two individual replicates were retained in the final consensus peak set for each time-point. For each time-point, excluded peaks using this approach showed constantly a decreased bigWig signal when compared to peaks common between replicates or included “rescued” peaks (Supplemental Fig. S1C). Consensus peaks selected for each time-point were merged using BEDTtools (Quinlan and Hall 2010) v2.30.0 to form a union set of peaks for downstream differential accessibility analysis. Raw read counts over the union set of peaks were generated using featureCounts (Liao et al. 2014) v2.0.1.

Replicate concordance was assessed as above and peaks with absolute LFC > 1 and p-adjusted < 0.05 were regarded as differentially accessible with peaks having LFC > 1 designated as “Gained-Open (GO)” and peaks with LFC < -1 designated as “Gained-Close (GC)”. Unsupervised *k*-means clustering of differentially accessible peaks between time-points was done on *Z*-score transformed $\log_2$(CPM+1) mean accessibility values of the two biological replicates from each time-point as above.

Transcription Factor (TF) motifs enrichment analysis was done using HOMER (Heinz et al. 2010) on the identified GO and GC differentially accessible region categories and on time-point specific peaks. TF motifs identified by HOMER with p-value < 0.01 and (percentage of Targets Sequences with Motif) / (percentage of Background Sequences with Motif) > 1.5 were retained for further analysis. Identification of enriched KEGG pathways of the selected TFs was done using Enrichr (Kuleshov et al. 2016). To assess whether the loci of enriched TF motifs were simultaneously shared between the different GO and GC open chromatin region (OCR) groups, we extracted the loci of each enriched TF motif using HOMER, then we determined the fraction of overlap between each enriched motif's loci and the different GO and GC OCR groups using BEDTools intersect. The closest gene as predicted by HOMER was used as target.

For ChIP-seq, quality checks on raw FASTQ data were performed as above and aligned using Bowtie 2 v2.3.4.1 with default parameters. Picard (Broad Institute, 2019) and SAMtools v1.4.1 were used to filter out duplicates, unmapped reads, secondary, and supplemental alignments. RPGC normalized bigWig files were generated using deepTools v3.5.0 with parameters (-effectiveGenomeSize 2652783500 -normalizeUsing "RPGC" -e -centerReads -binSize 20 -smoothLength 60).

The two highly concordant replicates with good insert size were kept. Peak calling was performed as for ATAC-seq above with the additional use of the merged input BAM file of the selected replicates as background control for MACS2. Summary of alignment results for RNA-seq, ATAC-seq, and ChIP-seq experiments is listed in Supplemental Table S1.

For quantifying the overlap in Fig. 3C, we normalized the counts of common genes by the min of absolute counts for both, as indicated on the figure. We also calculated the

Bonferroni-corrected p values of a hypergeometric test represented by the color scale to highlight significant cases.

Calling enhancers from H3K27ac ChIP-seq data: Typical enhancers (TEs) and super-enhancers (SEs) were called from the H3K27ac ChIP-seq data using Rose2 (Whyte et al. 2013) with parameters (-t 2000 -s 12500) after filtering out H3K4me3 peaks of all available brain tissues from ENCODE (Consortium 2012) to exclude promoter-specific peaks (Davis et al. 2018). The closest gene with a TSS within 10 kb or 50 kb upstream or downstream of a TE and SE respectively were designated as target gene.

Identifying pineal gland / PB-specific SEs: Pineal gland/pineoblastoma specific SEs were identified by filtering out all previously identified typical and super enhancers that intersected either gene bodies or distal and proximal *cis*-Regulatory Elements (CREs) from ENCODE (ENCODE references ENCSR695LYW and ENCSR770MVN respectively).

GSEA-PCA Analysis: Mouse and human gene expression profiles were collated from in-house experiments and the NCBI Gene Expression Omnibus (GEO) across various brain cancer tumors and subtypes. NCBI datasets were collected for Medulloblastoma (GSE37382), Glioblastoma (GSE36245), and Retinoblastoma (GSE29685, GSE24673, GSE59983). We also included mouse *Dicer1/Trp53*-deleted PB and *mouse Rb/Dicer1/Trp53*-deleted PB described previously (Chung et al. 2020), and 17 samples (including biological replicates and different time-points) for the mouse models of *Irpb-Ccnd1/Trp53*$^{-/-}$ PB and mouse *Irpb-Ccnd1* (Pinealoma) described in this study. Gene expression profiles were processed using the geoquery and affy packages from BrainArray, as described previously (Gendoo et al. 2015), and FPKM values were extracted for the *Irpb-Ccnd1/Trp53*$^{-/-}$ PB and

mouse *Irpb-Ccnd1* models as described above. Single-sample Gene Set Enrichment Analysis (ssGSEA) (Barbie et al. 2009) on Gene Ontology Biological Processes (GO BP) was conducted for all the human and mouse expression profiles as previously described (Gendoo et al. 2015), using the GSVA package (version 1.38.2) (Hanzelmann et al. 2013) in R (version 4.0.5) (R Core Team 2021). Genesets common between human and mouse datasets were then selected, from which enrichment scores (ES) across the datasets were combined to develop a ssGSEA-ranked matrix (Gendoo et al. 2015). Principal component analysis (PCA) was conducted on 549 ssGSEA-ranked genesets (rank matrix) across 609 human and mouse samples. Additionally, we included in our analysis data for 5 pineoblastoma samples, one pineocytoma sample, one Pineal Parenchymal Tumor of Intermediate Differentiation (PPTID) sample and one Pineal Tumor sample from the St. Jude portal (https://platform.stjude.cloud/data/diseases). We also included data for 6 PB samples from 5 patients downloaded from the kidsfirst portal (https://portal.kidsfirstdrc.org/)

Identifying PB active transcriptional regulators: Active PB enhancers were identified by intersecting intergenic and intronic P90 ATAC-seq peaks with P90 H3K27ac peaks, and HOMER was used to identify TF regulators acting on these enhancers similar to the TF motif enrichment analysis described earlier.

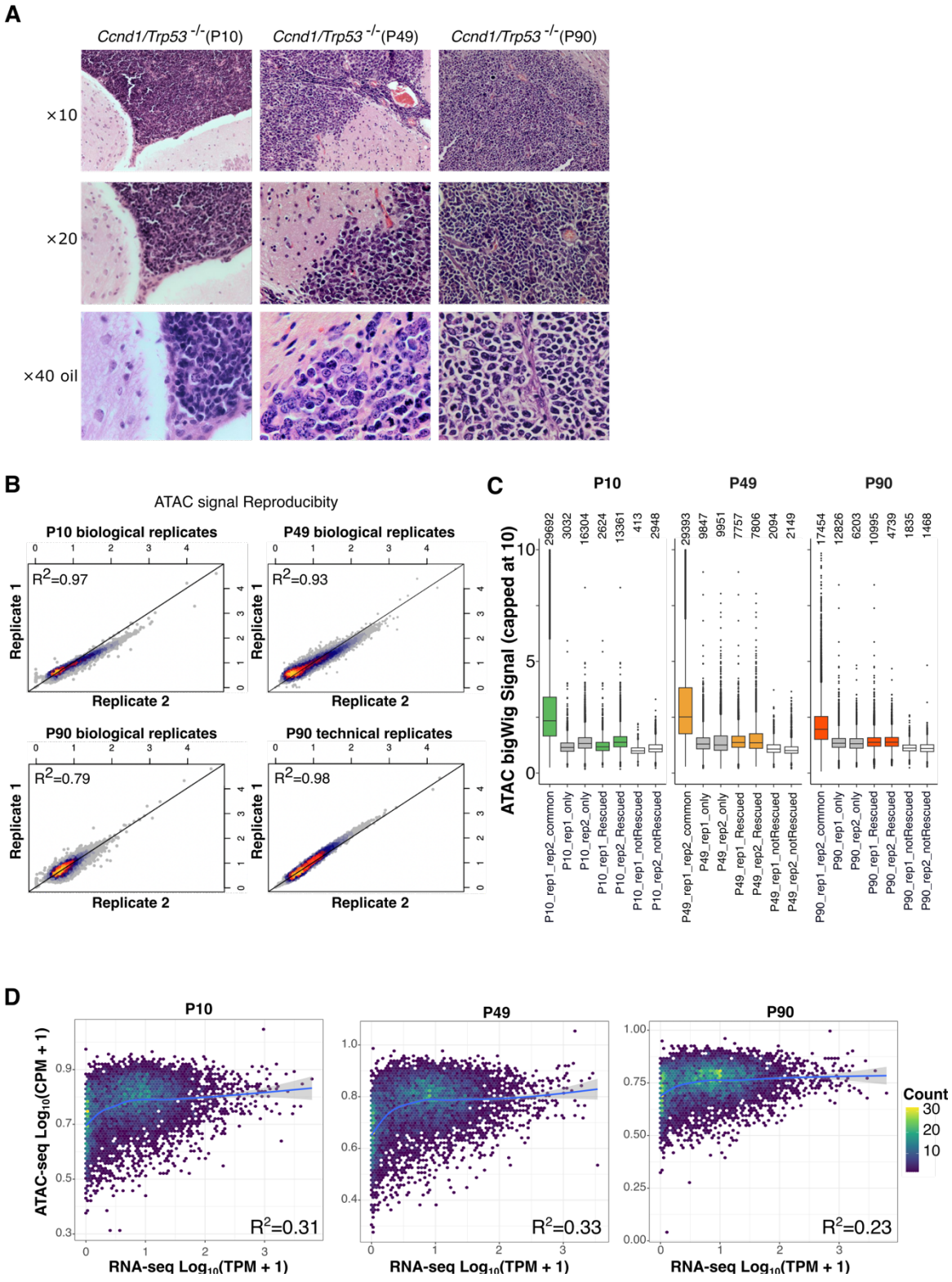
Annotating open chromatin regions with ChromHMM annotations: Chromatin states (ChromHMM 18-state model; http://compbio.mit.edu/ChromHMM/) mm10 annotations were downloaded from ENCODE for all available tissues (forebrain, midbrain, hindbrain, heart, lung, liver, kidney, stomach, intestine) at P0 (the available postnatal time-point in this dataset) (van der Velde et al. 2021). For each tissue, ChromHMM states were grouped into the five categories identified, thus excluding the "quiescent" states (i.e. states QuiesG, Quies,

Quies2, Quies3, and Quies4) as it covers ~80% of the genome. The remaining categories were “Active TSS” (Tss and TssFlnk), “Transcribed Genes”, (Tx and TxWk), “Enhancers” (EnhG, Enh, EnhLo, EnhPois, and EnhPr); “Bivalent Tss” (TssBiv), and “Repressive State” (ReprPC, ReprPCW, and Het). Next, the number of overlaps between the selected categories (either the 6 GO/GC categories or the P10, P49, P90, and common peaks between the three time-points), and the 5 ChromHMM categories for all 9 tissues were calculated using BEDTools intersect. The Jaccard index for each peak set and ChromHMM category was calculated to assess the similarity between our peaks and the ChromHMM annotations. Accession numbers for ENCODE files are ENCFF403VHJ, ENCFF643RYT, ENCFF758EGD, ENCFF762DAY, ENCFF554BKJ, ENCFF600PBT, ENCFF809HLK, ENCFF955TQG, ENCFF228XKW.

Additional data analysis and graphical representations: Mouse model and experimental design illustrations were made using Servier Medical Art (https://smart.servier.com). Bar plots and violin plots were generated using ggplot2 R package. Venn diagrams were generated using eulerr R package v6.1.0 (Micallef and Rodgers 2014). Heatmaps were generated using ComplexHeatmap v2.7.9.1007 (Gu et al. 2016). H3K27ac profile heatmap was generated using deepTools.

Supplemental Figures

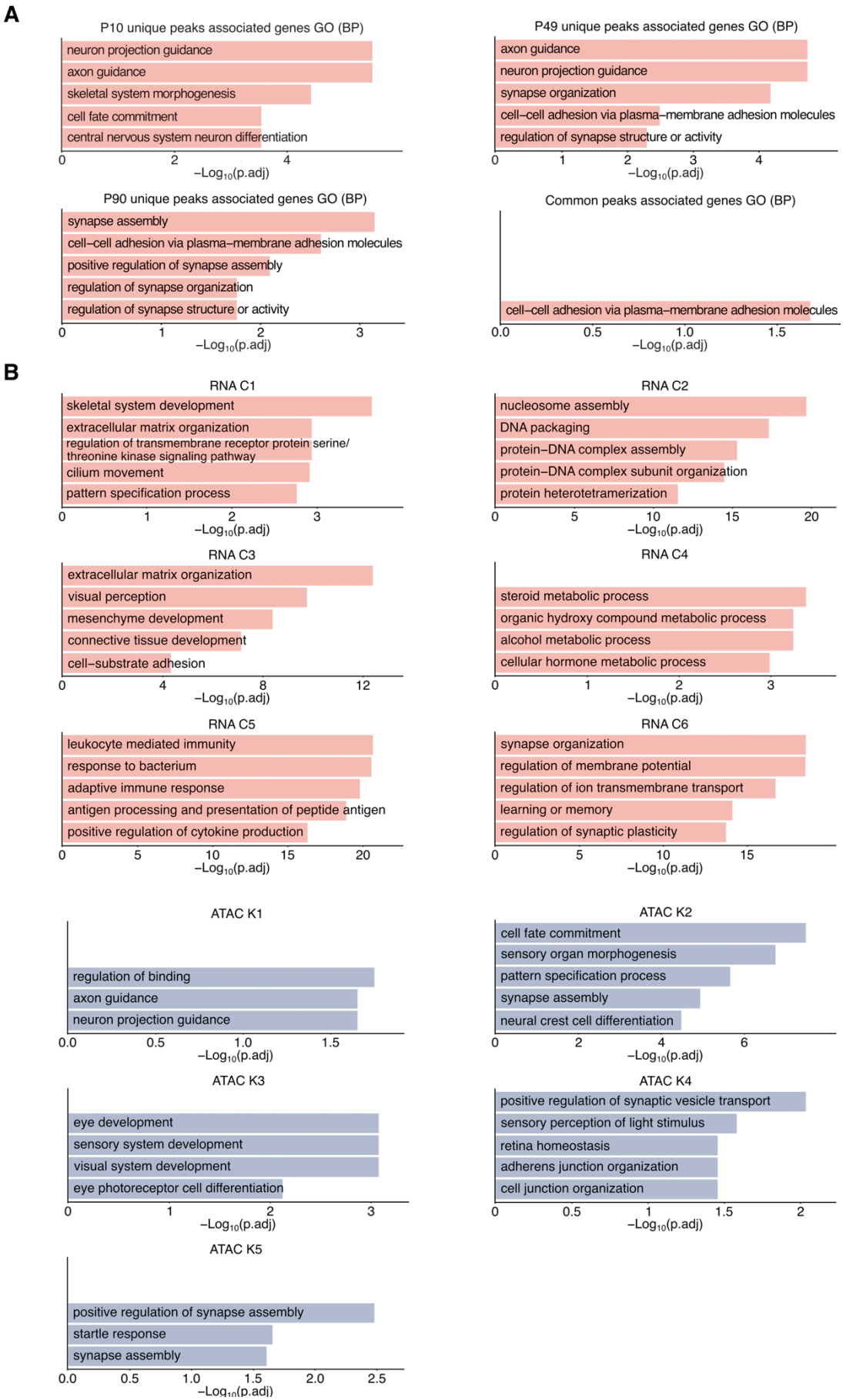
Supplemental Figure S1



Supplemental Figure S1: Quality checks on tissue and sequenced samples (A) Representative hematoxylin and eosin staining (H&E) in *Rpb3-Ccnd1*, *p53*$^{-/-}$ pineal glands at

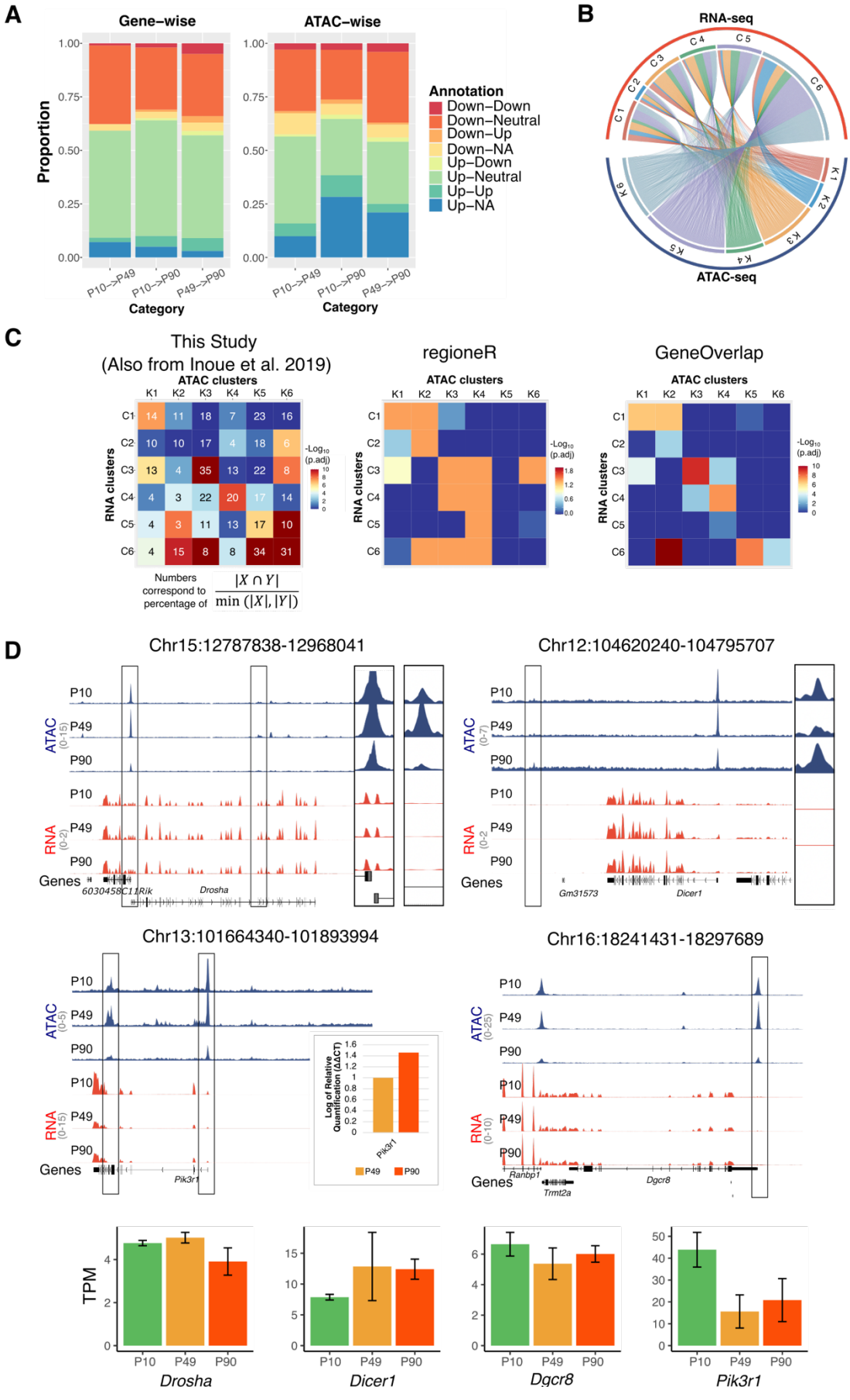
the indicated time points (P10, P49 and P90/ Tumor). (**B**) Correlation of ATAC-seq signal on peaks. Each dot is a peak and its signal in both replicates was extracted from BAM files in order to compare raw signals and plotted in log scale in. The bottom right plot corresponds to technical replicates from P90. (**C**) Boxplot of bigWig signal over called peaks. First (common) boxplot is for the common set of peaks between replicates. Second and third boxplots (grey in all three time points) are for replicate specific peaks (e.g peaks that do not intersect if we intersect peaks of both replicates). Fourth and fifth boxplots are for peaks that are specific to the first or the second replicate but were rescued using our approach and included in the final set of peaks. Sixth and seventh boxplots (white color) are for peaks that are specific to the first or the second replicate but were not rescued. The "notRescued" peak set have lower bigWig signal. Please not that Rescued and notRescued set of peaks are both subpopulations of the replicate only peak set (grey boxplots). The count of peaks in each category is shown above the corresponding boxplot. (**D**) Correlation analysis of ATAC-seq (y-axis) and RNA-seq (x-axis) signals over the three time points.

Supplemental Figure S2



Supplemental Figure S2: Top 5 gene ontology terms for different DEG and OCR groups. (A) Top 5 enriched gene ontology terms (Biological Processes) for P10, P49, and P90 unique and common peaks. **(B)** Top 5 enriched gene ontology terms (Biological Processes) for clusters of DEGs (C1 to C6) and DARs (K1 to K6) illustrated in Fig. 2b.

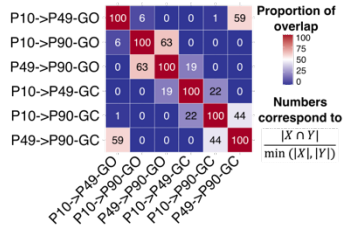
Supplemental Figure S3



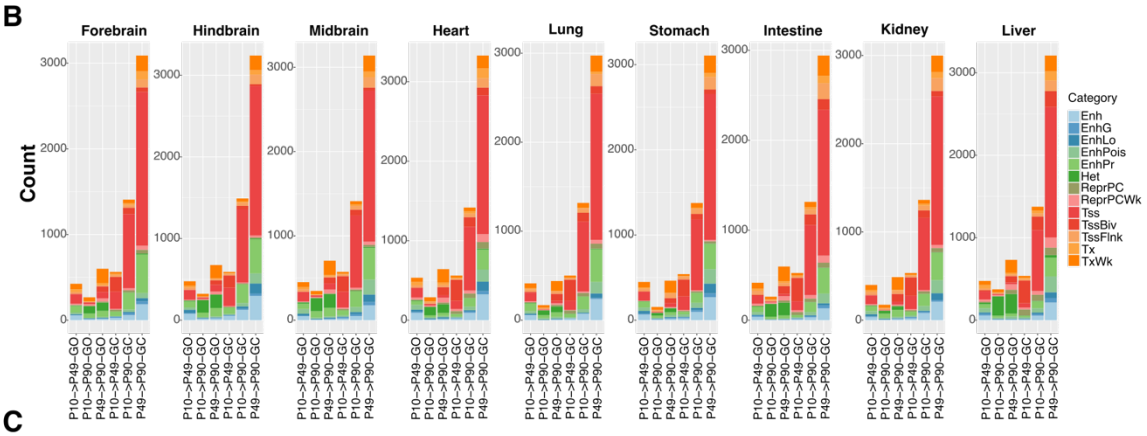
Supplemental Figure S3: (A) Proportions of genes-accessible region associations from Supplementary Table S9. Down = down regulated (for genes) and gained close (for peaks); Up = up regulated (for genes) and gained open (for peaks); Neutral = no change; NA = no information (for genes this means no peak is associated with it and for peaks this means that the statistics failed for this gene) (**B**) Circos plot linking the DARs in each cluster (Fig. 3B, K1 to K6) to their corresponding target genes in the DEGs clusters (Fig. 3B, C1 to C6). (**C**) Side by side comparison of association heatmaps for ATAC and RNA clusters using the gene sets intersections method as in Fig. 3C (left), features overlap as reported by regioneR (middle) (Gel et al. 2016), and gene set overlaps reported by GeneOverlap (Li Shen and Icahn School of Medicine at Mount Sinai 2021). (**D**) Selection of genomic tracks of normalized RNA-seq and ATAC-seq profiles showing expression and accessibility levels for 4 PB related genes (*Drosha*, *Dicer1*, *Pik3r1*, and *Dgcr8*). DAR associated with the genes are highlighted and a zoomed browser screenshot is shown to the right. The TPM values for the target gene are shown in the barplots below. The coordinates of the shown loci are indicated above. The y-axis scales for both assays are shown in parenthesis. Error bars represent standard deviations from the assessed biological replicates. For *Pik3r1*, we also show the RT-qPCR validation to the right.

Supplemental Figure S4

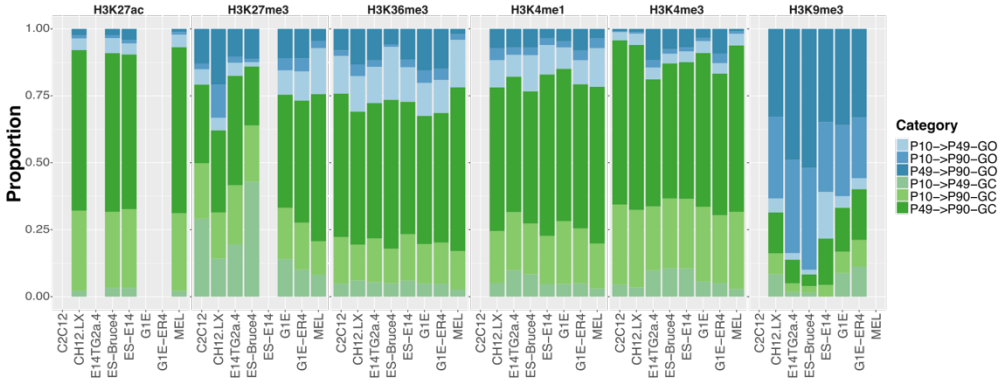
A



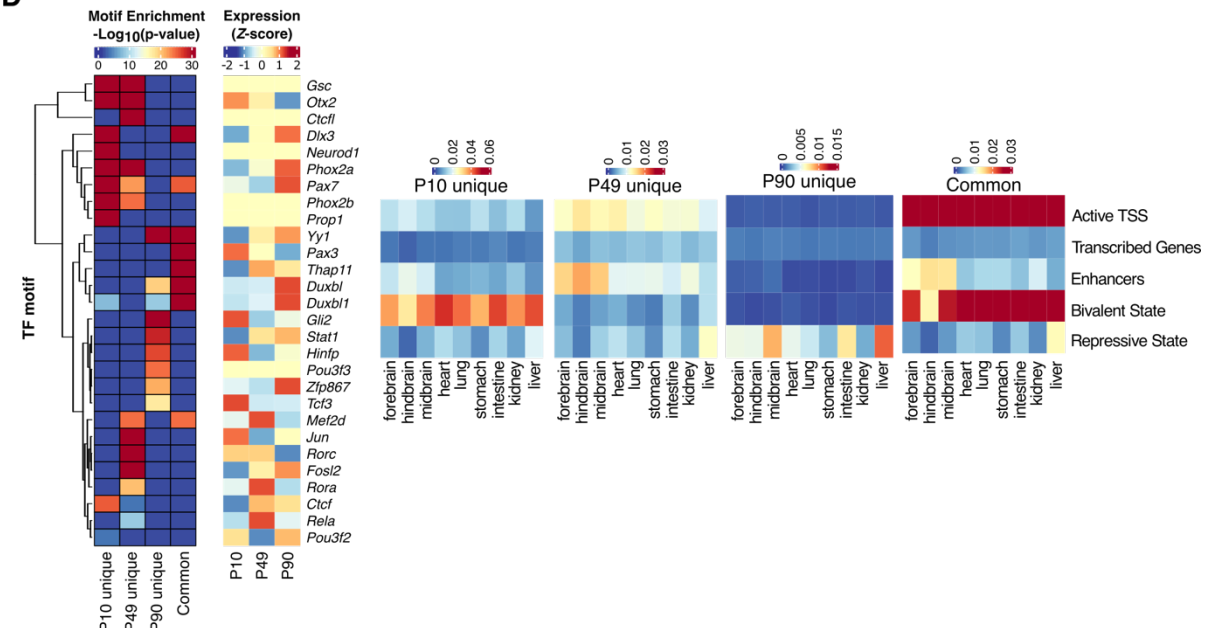
B



C

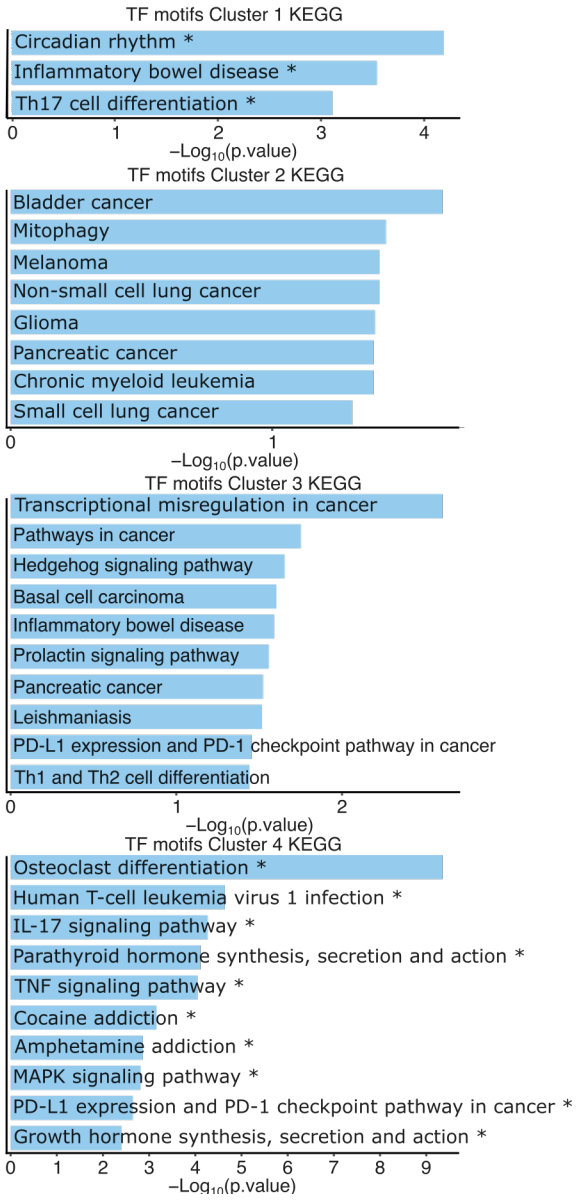


D



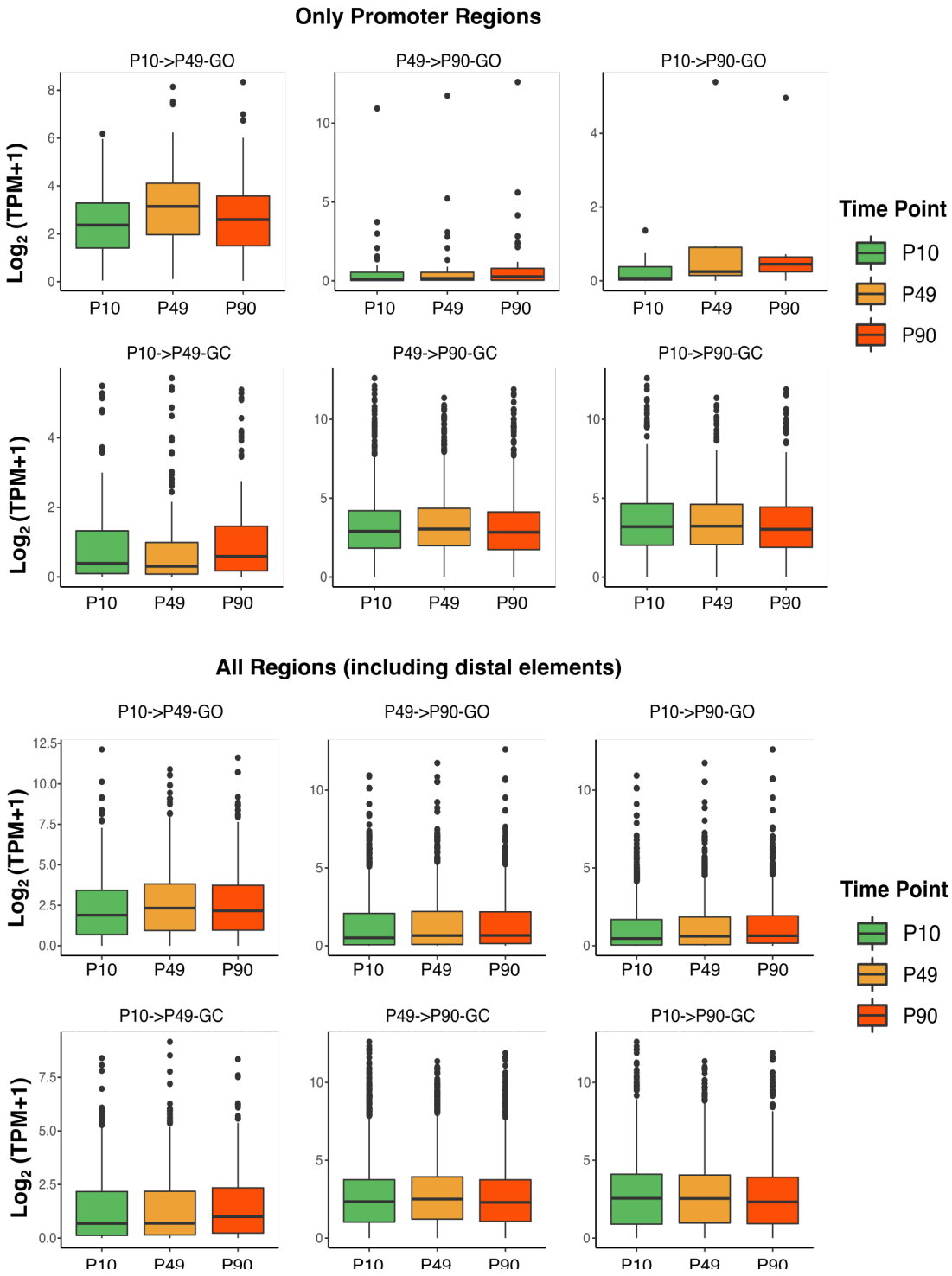
Supplemental Figure S4: (**A**) Tile plot indicating the pairwise GO/GC OCRs groups' percentage overlap calculated from the fraction of OCRs shared between GO/GC groups in each row/column divided by the minimum of the two groups. **(B)** Count of overlap between entries in each of the GO and GC OCR groups (x-axis) and the list of regions from each of the ChromHMM categories. (**C**) Proportion of overlap between each of the GO and GC OCR groups and different chromatin marks from different cell lines (Supplemental Table S18). Note the increase in overlap with the H3K9me3 repressive mark and the P10->P49-GO or the P49-P90-GO groups in all cell lines. (**D**) (Left) Heatmap of TF motif enrichment in P10, P49, and P90 unique and common OCRs with *Z*-score normalized mean expression values of the corresponding TFs to the right. (Right) Heatmaps of Jaccard indices for the 6 GO and GC OCR groups and P10, P49, and P90 unique and common OCR groups (respectively) in 9 ChromHMM annotated mouse tissues from ENCODE (Methods) with respect to 5 ChromHMM categories.

Supplemental Figure S5



Supplemental Figure S5: Enrichment for KEGG pathways in GO and GC cluster categories. For each cluster, the top enriched KEGG pathways are shown with p-value < 0.05. Stars indicate pathways with an adjusted p-value < 0.05.

Supplemental Figure S6



Supplemental Figure S6: Target gene and gained-open/gained-close association.
Boxplots of log_2 normalized gene expression for the different gained open and gained close categories at the three time points. The upper panel is for regions overlapping promoter elements and the bottom one for all regions.

References:

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C et al. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**: 108-112.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**: 21 29 21-21 29 29.

Chung PED, Gendoo DMA, Ghanbari-Azarnier R, Liu JC, Jiang Z, Tsui J, Wang DY, Xiao X, Li B, Dubuc A et al. 2020. Modeling germline mutations in pineoblastoma uncovers lysosome disruption-based therapy. *Nat Commun* **11**: 1825.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK et al. 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**: D794-D801.

Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047-3048.

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766-D773.

Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**: 289-291.

Gendoo DM, Smirnov P, Lupien M, Haibe-Kains B. 2015. Personalized diagnosis of medulloblastoma subtypes across patients and model systems. *Genomics* **106**: 96-106.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847-2849.

Hanzelmann S, Castelo R, Guinney J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**: 7.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357-360.

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**: W90-97.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li Shen and Icahn School of Medicine at Mount Sinai. 2021. GeneOverlap: Test and visualize gene overlaps.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923-930.

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Gruning B, Ramirez F, Manke T. 2021. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**: 422-423.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Micallef L, Rodgers P. 2014. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* **9**: e101717.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

R Core Team. 2021. R: A Language and Environment for Statistical Computing.

Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187-191.

Serin Harmanci A, Harmanci AO, Zhou X. 2020. CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. *Nat Commun* **11**: 89.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.

Uhrig S, Ellermann J, Walther T, Burkhardt P, Frohlich M, Hutter B, Toprak UH, Neumann O, Stenzinger A, Scholl C et al. 2021. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res* **31**: 448-460.

van der Velde A, Fan K, Tsuji J, Moore JE, Purcaro MJ, Pratt HE, Weng Z. 2021. Annotation of chromatin states in 66 complete mouse epigenomes during development. *Commun Biol* **4**: 239.

Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**: 2184-2185.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**: 307-319.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**: 284-287.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.