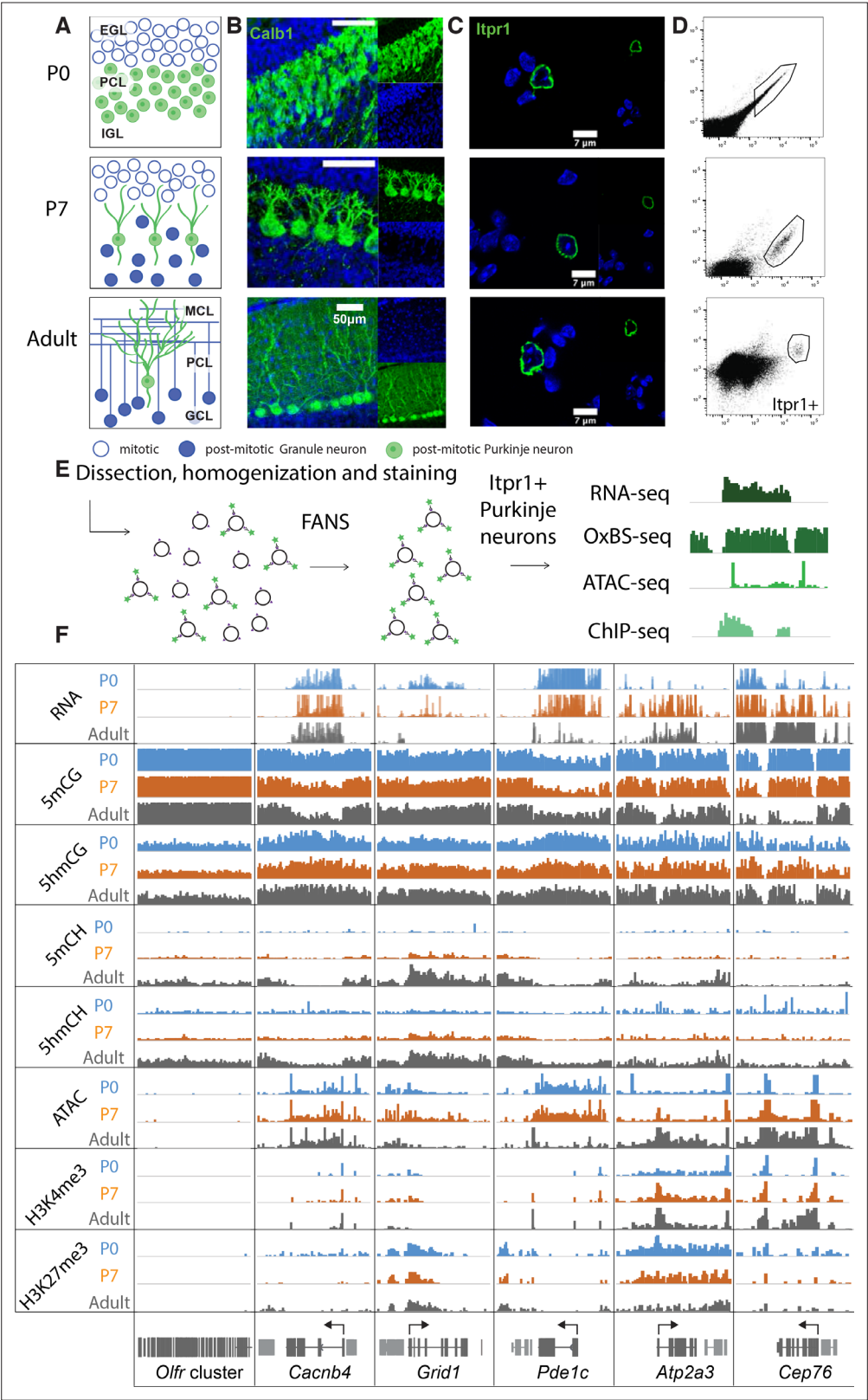


---

## Figures and figure supplements

5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function

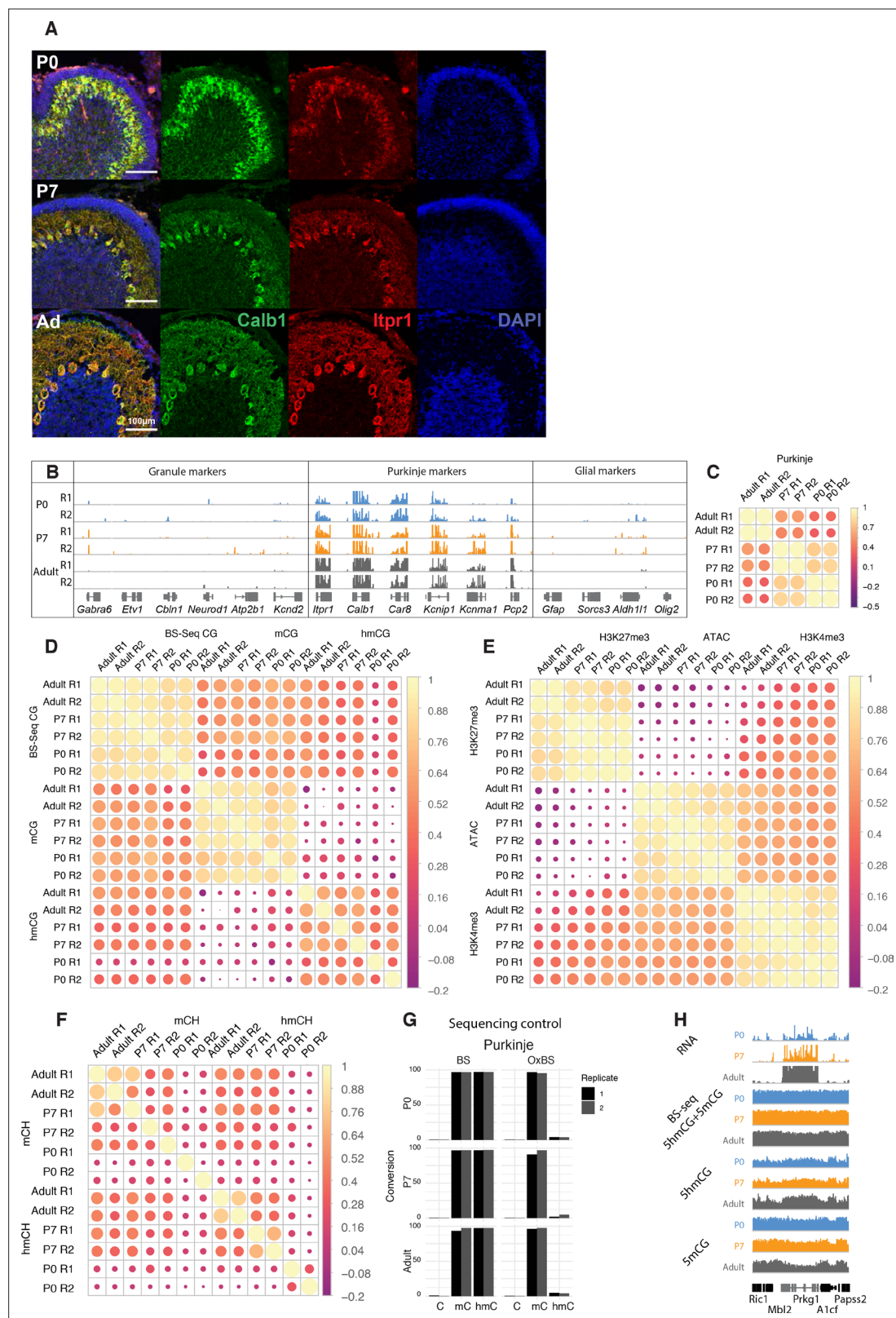
**Elitsa Stoyanova et al**



**Figure 1.** Chromatin landscape in differentiating Purkinje neurons. **(A)** Schematic of Purkinje cell (PC) differentiation and growth at P0, P7 – and adult (approximately 8-week-old) timepoints. EGL – external granule layer, PCL – Purkinje cell layer, IGL – internal granule layer, MCL – molecular cell layer, GCL – granule cell layer. **(B)** Immunofluorescence staining with Calb1 (green) of PCs in murine cerebella at P0, P7, and adult timepoints. **(C)** Figure 1 continued on next page

*Figure 1 continued*

Example of PC nuclei stained with Itpr1 (green) post-dissociation and pre-sorting, counterstained with DAPI. **(D)** Representative plots of fluorescence-activated nuclear sorting of PCs at P0, P7, and adult timepoints with Itpr1. **(E)** Workflow schematic of nuclei isolation, antibody staining with anti-Itpr1, fluorescence-activated sorting, and downstream sequencing applications. **(D)** Integrated genome viewer (IGV) representation of example regions of differentially regulated genes (*Olfr* cluster – always silent, *Cacnb4* – always expressed, *Grid1* and *Pde1c* – developmentally down-regulated, *Atp2a3* and *Cep76* – developmentally up-regulated). Top tracks show RNA expression in RPKM (reads per kilobase per million mapped reads), mCG tracks show methylation level in CG context from 0 to 0.8, hmCG tracks show hydroxymethylation level in CG context from 0 to 0.45, mCH and hmCH show methylation and hydroxymethylation level in CH context (H = A, C, or T) from 0 to 0.04. ATAC tracks show ATACSeq read density in RPKM from 0 to 10. H3K4me3 tracks show input normalized enrichment in RPKM from 0 to 5. H3K27me3 tracks show input normalized enrichment in RPKM from 0 to 3.



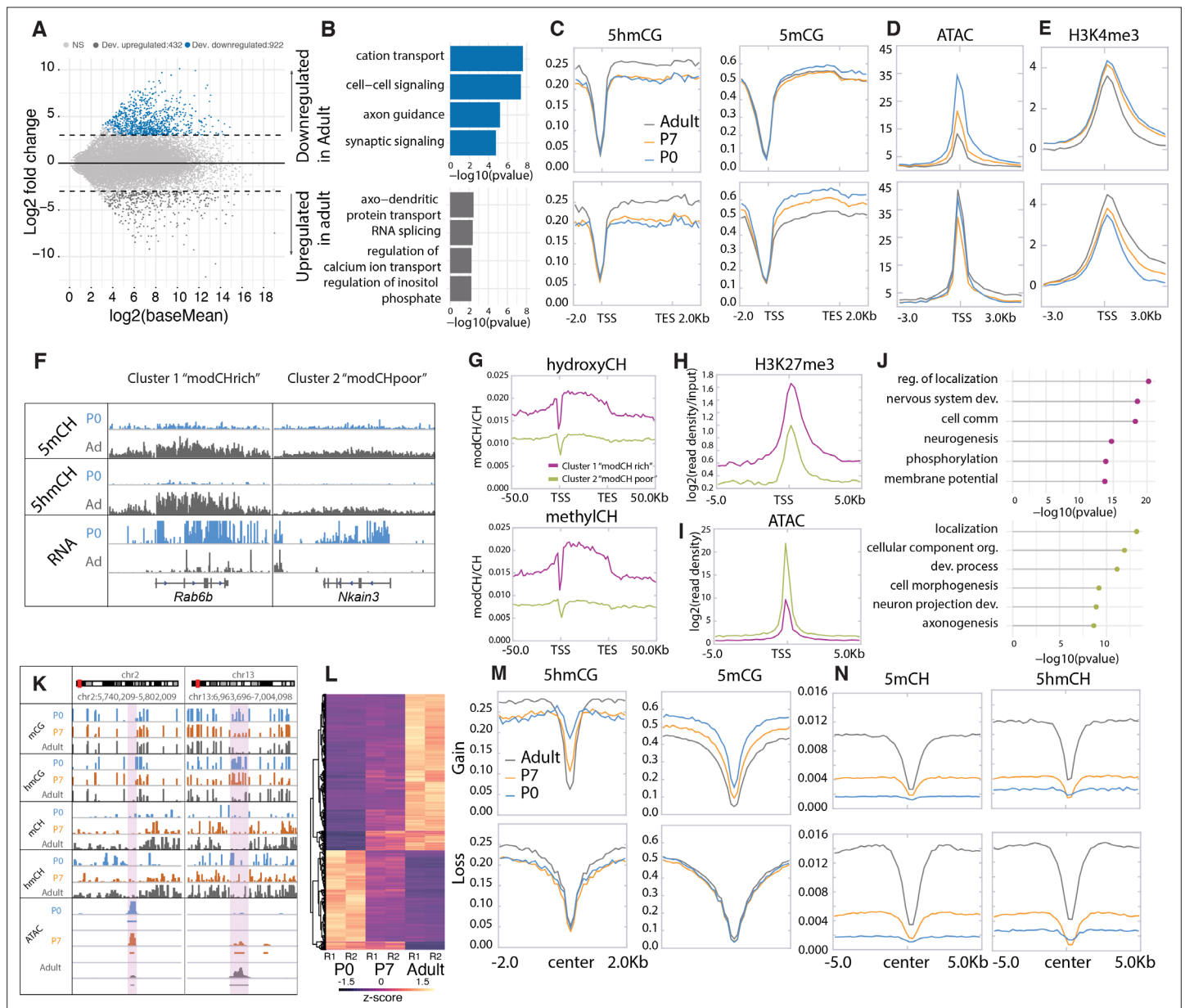
**Figure 1—figure supplement 1.** Quality control of sequencing datasets. **(A)** Immunofluorescence staining of Purkinje cells at P0, P7, and adult timepoints with Calb1 (green) and Itpr1 (red). **(B)** Integrated genome viewer (IGV) representation of Purkinje-specific markers (*Itpr1*, *Calb1*, *Car8*, *Kcnip1*, *Kcnma1*, *Pcp2*) enrichment and depletion of granule (*Gabra6*, *Etv1*, *Cbln1*, *Neurod1*, *Atp2b1*, *Kcnd2*) and glial (*Gfap*, *Sorcs3*, *Aldh1l1*, *Olig2*) markers. **(C)** Pearson correlation of RNASeq datasets. **(D)** Pearson correlation of gene body accumulation of 5hmCG + mCG (bisulfite sequencing)

Figure 1—figure supplement 1 continued on next page

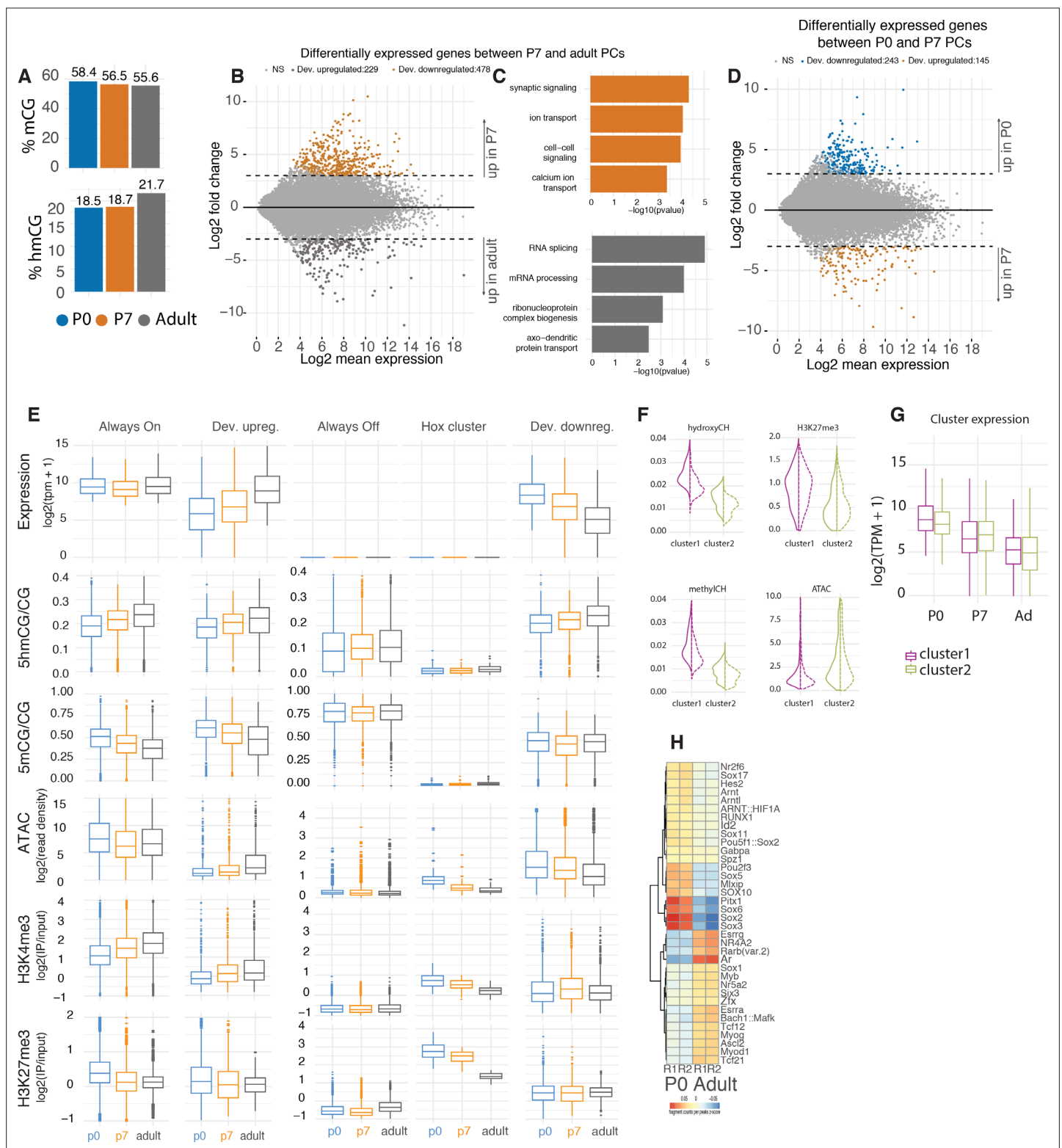


*Figure 1—figure supplement 1 continued*

[BSSeq]), 5hmCG, and 5mCG (derived from maximum likelihood estimator model of BSSeq and oxidative bisulfite sequencing [OxBSSeq]). **(E)** Pearson correlation of assay for transposase-accessible chromatin sequencing (ATACSeq), H3K4me3, and H3K27me3 chromatin immunoprecipitation sequencing (ChIPSeq) dataset enrichment  $\pm 5$  kb around the gene promoters. **(F)** Pearson correlation of gene body accumulation of 5hmCH and 5mCH (derived from maximum likelihood estimator model of BSSeq and OxBSSeq). **(G)** Bisulfite conversion and oxidation efficiency in BSSeq and OxBSSeq datasets. **(H)** IGV representation of BSSeq (showing 5hmCG + 5 mCG signal together) and OxBSSeq (showing 5hmCG + 5 mCG signal separately) at the three developmental timepoints.



**Figure 2.** Cytosine modification dynamics in relation to differential expression and accessibility. **(A)** MA plot representing statistically significant ( $p < 0.01$ ,  $\log_2(\text{fold change}) > 2$ ) differential gene expression between P0 (blue) and adult (dark gray). Light gray dots represent genes that are not statistically significant. **(B)** Gene ontology analysis of the developmentally up- and down-regulated genes. **(C–E)** Metagene plots representing the mean value of 5hmCG/CG, 5mCG/CG (**C**), assay for transposase-accessible chromatin sequencing (ATACSeq) read density (**D**) and H3K4me3 accumulation (**E**) over the gene bodies and promoters of developmentally up- and down-regulated genes over the three timepoints (P0 – blue, P7 – orange, and adult – dark gray). **(F)** Genome browser representation of the two classes of developmentally down-regulated genes with differential accumulation of CpH modifications. **(G–I)** Metagene plots representing the mean value of 5hmCH/CH, 5mCH/CH (**G**), ATACSeq read density (**H**), and H3K27me3 accumulation (**I**) over the gene bodies and promoters of developmentally down-regulated genes divided into two clusters by k-means clustering analysis (green for 'modCHpoor', purple for 'modCHrich'). **(J)** Gene ontology analysis of the two clusters. **(K)** Genome browser representation of two regions with differential accessibility. **(L)** Heatmap representing the differentially accessible regions between P0 and adult Purkinje cells (PCs) ( $p < 0.01$ ,  $\log_2(\text{fold change}) > 4$ ). **(M–N)** Metagene plots representing the mean values of 5hmCG/CG and 5mCG/CG (**M**) and 5hmCH/CH and 5mCH/CH (**N**) over the centers and flanking regions of peaks that gained or lost accessibility relative to adult PCs.

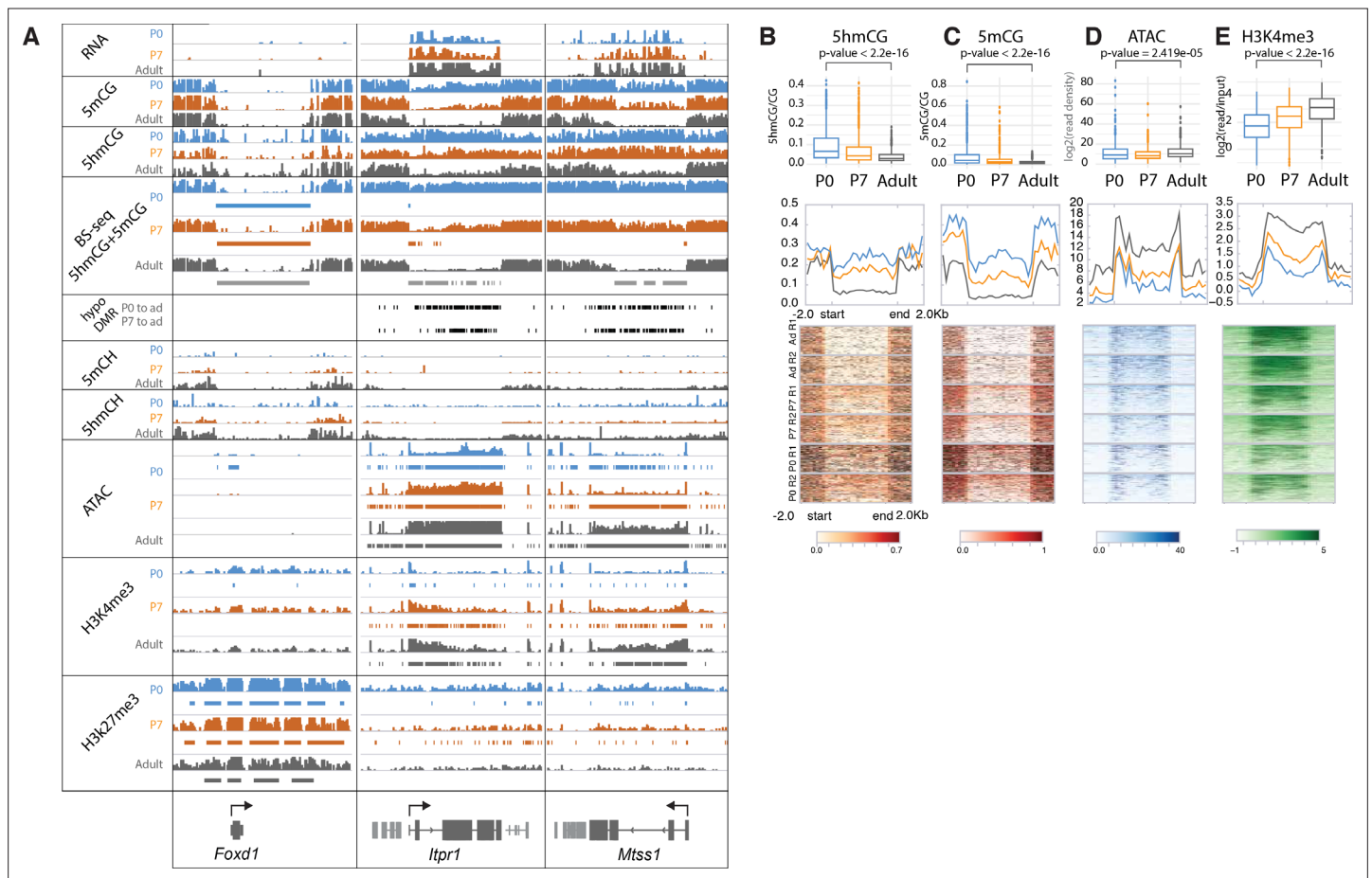


**Figure 2—figure supplement 1.** Gene expression and chromatin dynamics during Purkinje cell (PC) development. **(A).** Genome-wide quantification of the levels of 5hmCG and 5mCG at each timepoint. **(B)** MA plot representing differential gene expression between P7 and adult. Orange shaded dots represent genes enriched in P7 and gray dots represent genes enriched in adult, with absolute log2 fold change of 2 and p-adj value of 0.05. **(C)** Gene ontology (GO) analysis of differentially expressed genes between P7 and adult. **(D)** MA plot representing differential gene expression between P0 and P7 timepoints. Blue shaded dots represent genes enriched in P0 and orange dots represent genes enriched in P7, with absolute log2 fold change of 2 and p-adj value of 0.05.

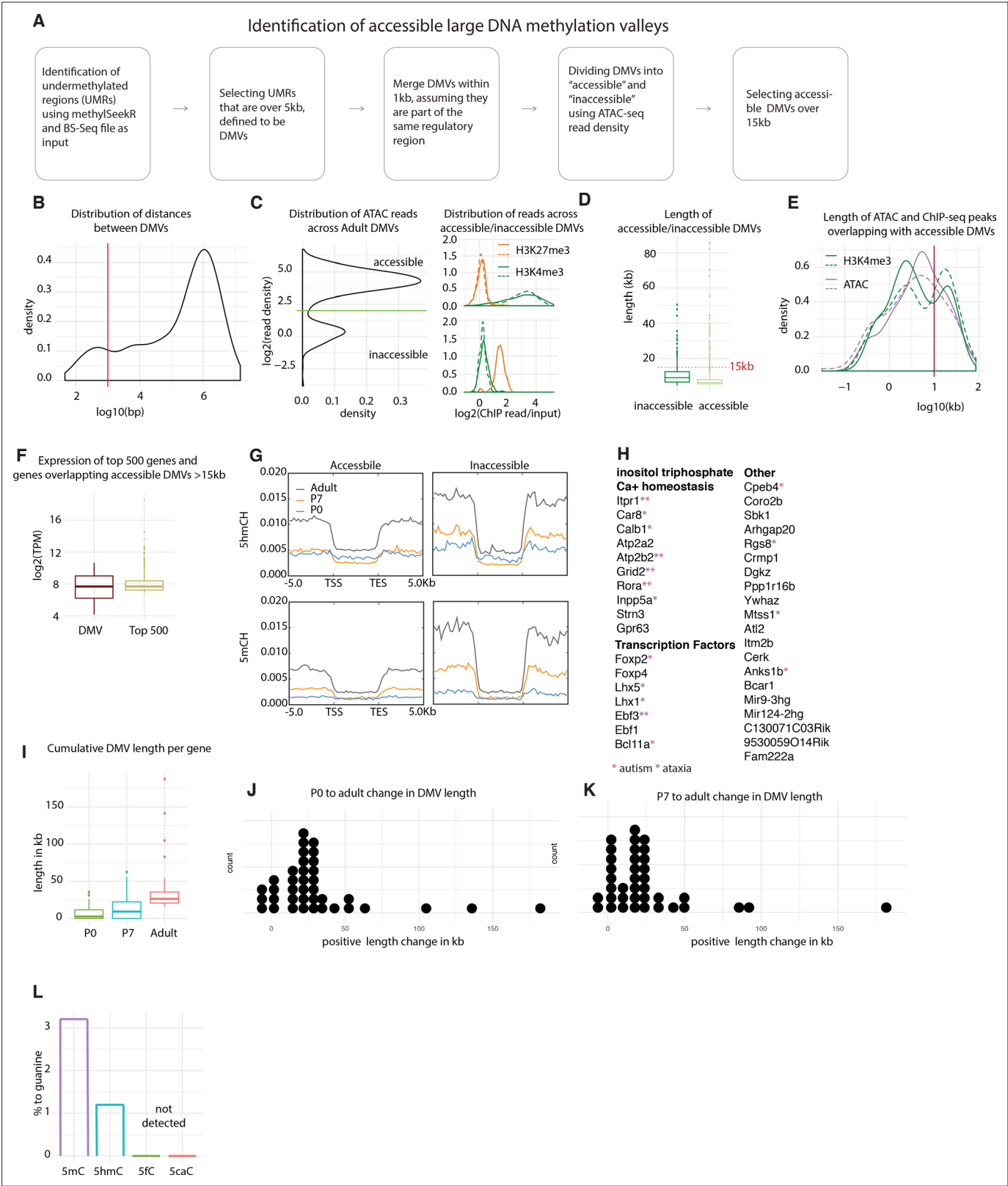
Figure 2—figure supplement 1 continued on next page

*Figure 2—figure supplement 1 continued*

2 and p-adj value of 0.05. **(E)** Quantification of expression, chromatin accessibility, 5mCG, 5hmCG, H3K4me3, and H3K27me3. **(F)** Violin plots of the quantifications of 5hmCH, 5mCH, ATAC, and H3K27me3 density over the two clusters of repressed genes. **(G)** Expression levels  $\log_2(\text{TPM} + 1)$  of the genes in each cluster. **(H)** Differentially abundant motifs of transcription factors in chromatin accessibility data between P0 and adult PCs.



**Figure 3.** Loss of 5mCG and 5hmCG in a novel class of epigenetically regulated highly expressed Purkinje-specific genes. **(A)** Integrated genome viewer (IGV) representation of example regions of inactive DNA methylation valley (DMV) (*Foxd1*) and active Purkinje cell (PC)-specific DMVs (*Itpr1*, *Mtss1*). Bisulfite sequencing (BS-seq) tracks shows combined levels of mCG and hmCG ranging from 0 to 1. Bars under BS-seq tracks are computationally identified DMVs, black bars denote hypo-DMRs (differentially methylated regions) in adult, compared to P0 or P7. Bars under assay for transposase-accessible chromatin sequencing (ATAC-seq), H3K4me3, and H3K27me3 tracks denote broad peaks of signal enrichment. **(B–E)** 5hmCG/CG (**B**), 5mCG/CG (**C**), ATAC-seq log<sub>2</sub>(read density) RPKM enrichment (**D**) and H3K4me3 log<sub>2</sub>(input normalized) RPKM enrichment (**E**) quantification over DMVs at each timepoint. Boxplots show mean value per DMVs, the test for significance is Wilcoxon. Metagene plots show mean value  $\pm$  2 kb around the DMV region, regardless of gene directionality.



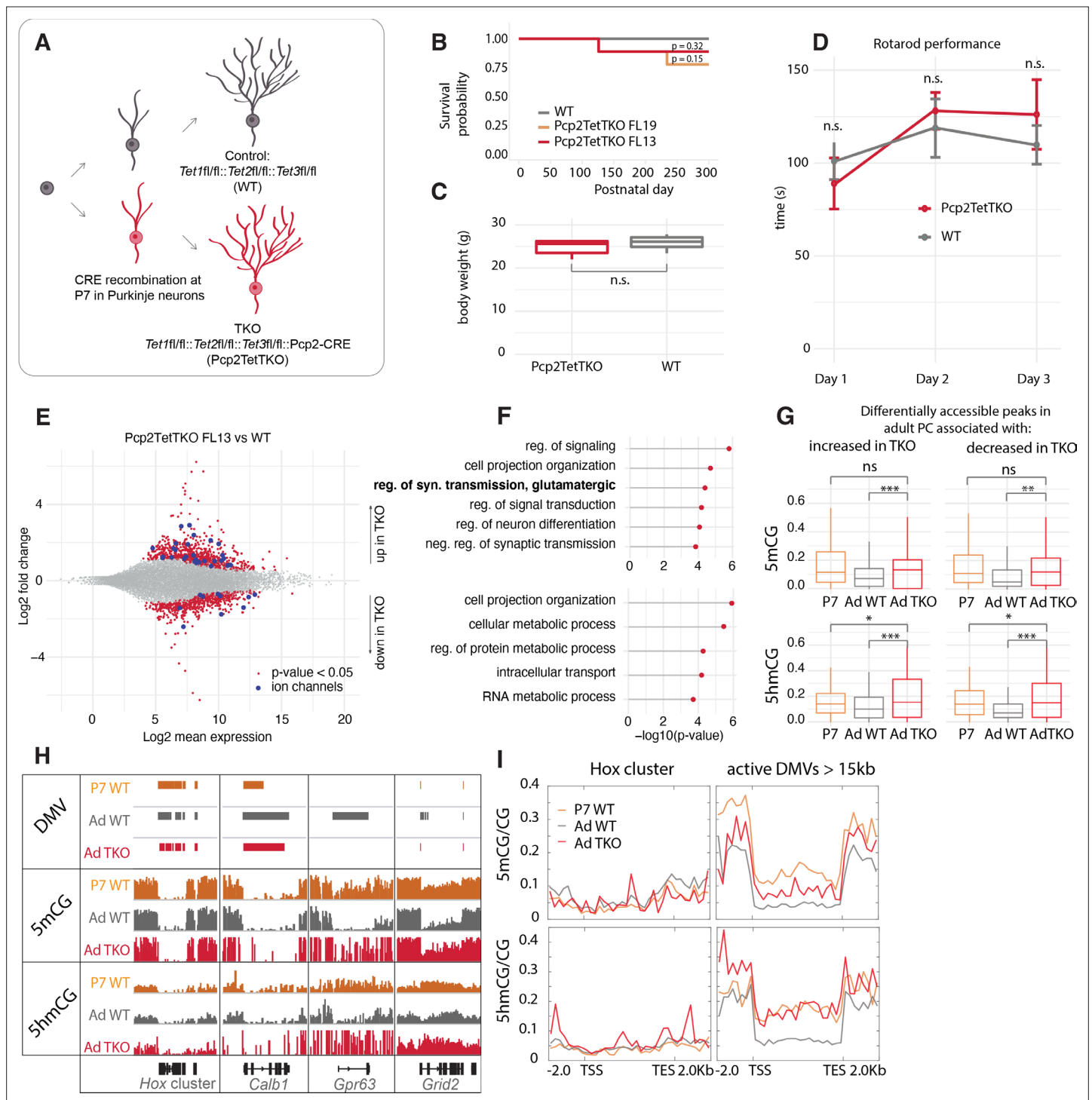
**Figure 3—figure supplement 1.** Characterization of Purkinje-specific DNA methylation valleys (DMVs). **(A)** Schematic of DMV identification and filtering. **(B)** Distribution of distances between DMVs. DMVs within 1 kb of each other are assumed to be part of the same regulatory region and therefore merged. **(C)** Assay for transposase-accessible chromatin sequencing (ATACSeq) read density over DMVs shows a bimodal distribution, segregating DMVs into accessible and inaccessible (left panel). Inaccessible DMVs show enrichment for H3K27me3 reads, while accessible DMVs show

Figure 3—figure supplement 1 continued on next page

*Figure 3—figure supplement 1 continued*

enrichment for H3K4me3 reads (right panel, dotted and straight lines designate replicates). **(D)** Distribution of lengths of accessible and inaccessible DMVs. **(E)** Length of broad peaks of ATACSeq and H3K4me3 chromatin immunoprecipitation sequencing (ChIPSeq) overlapping large accessible DMVs. **(F)** Expression of the top 500 most expressed genes in Purkinje cells (PCs) and the genes associated with DMVs. **(G)** DMVs do not accumulate any modifications in CH context. **(H)** Genes associated with DMVs, red star denotes role in autism, gray in ataxia **(I)**. Cumulative length of DMVs per gene. **(J–K)** Change of DMV length between P0 and adult **(J)** and P7 and adult **(K)**. **(L)** Mass spectrometry analysis of genomic DNA from adult PCs.



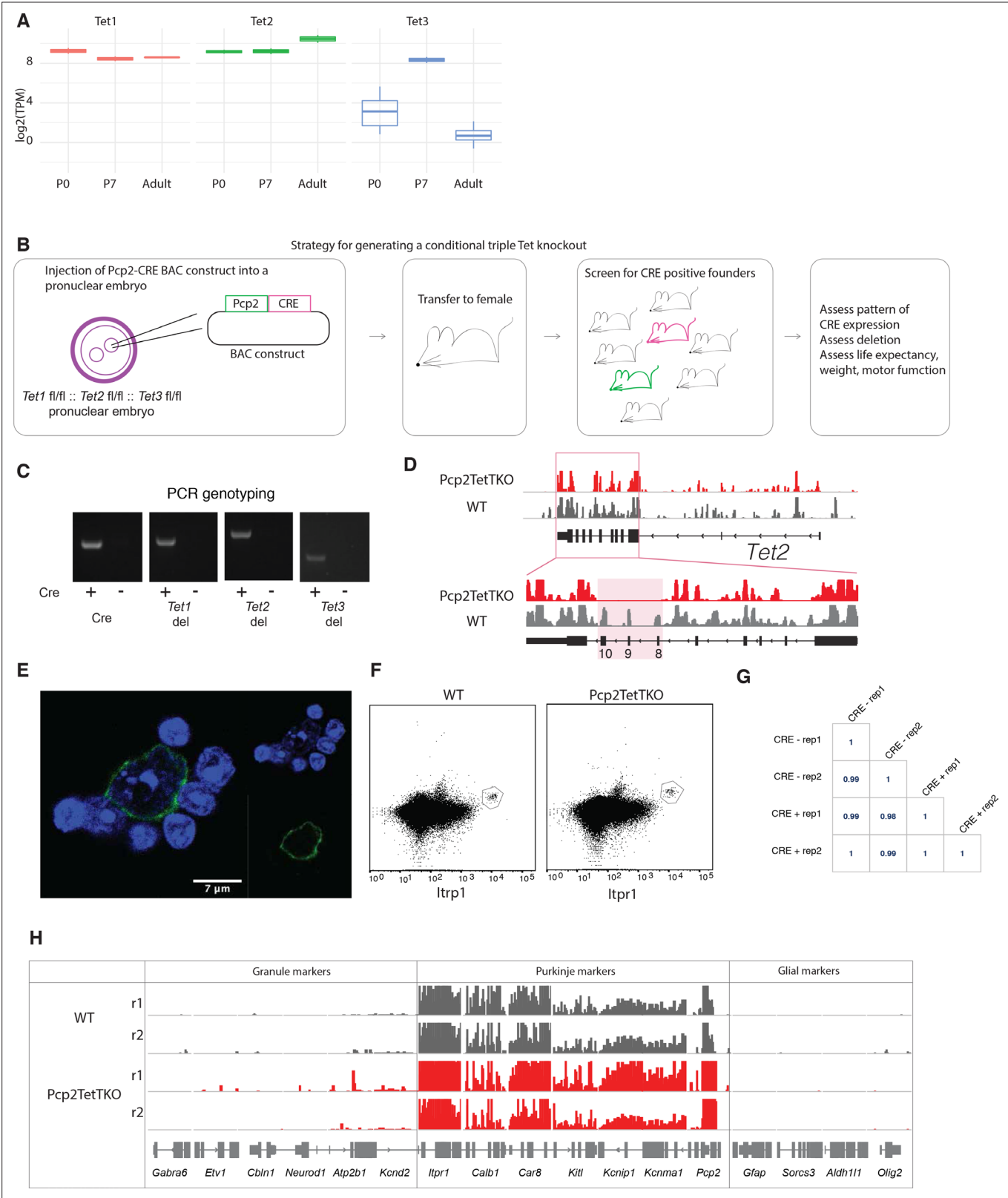


**Figure 4.** Loss of Tet activity leads to impaired gene expression regulation in Purkinje cells (PCs). **(A)** Schematic of experimental design to probe the triple Tet1, Tet2, Tet3 knockout effect in adult PCs. **(B)** Kaplan-Meier curve representing the survival fraction of wild-type (WT) (n = 8) and PC-specific triple knockout (Pcp2TetTKO) (Founder13 n = 8, Founder19 n = 8) mice since date of birth. **(C)** Body weight of WT and Pcp2TetTKO at 8 weeks. **(D)** Rotarod evaluation of motor skills in WT and Pcp2TetTKO at 8 weeks. **(E)** MA plot representing differential gene expression analysis between WT and Pcp2TetTKO. Navy dots represent genes with p-value < 0.05, red dots – ion channels. **(F)** Gene ontology analysis of statistically significant differentially expressed genes. Top panel shows categories of genes with increased expression in Pcp2TetTKO, bottom panel shows genes with decreased expression in Pcp2TetTKO. **(G)** Levels of 5hmCG/CG and 5mCG/CG over differentially accessible peaks only present in adult PCs (compared to P0). **(H)** Integrated genome viewer (IGV) representation of example DNA methylation valleys (DMVs) affected by Pcp2TetTKO. Hox cluster and Calb1 show minor changes in length as they are established before the P7 onset of Cre. Gpr63 and Grid2 show significant reduction in length as they are established.

Figure 4 continued on next page

Figure 4 continued

after P7. Solid bars represent DMVs identified at each condition. (I) Quantification of 5hmCG/CG and 5mCG/CG over the two classes of DMVs – Hox cluster and large active DMVs.

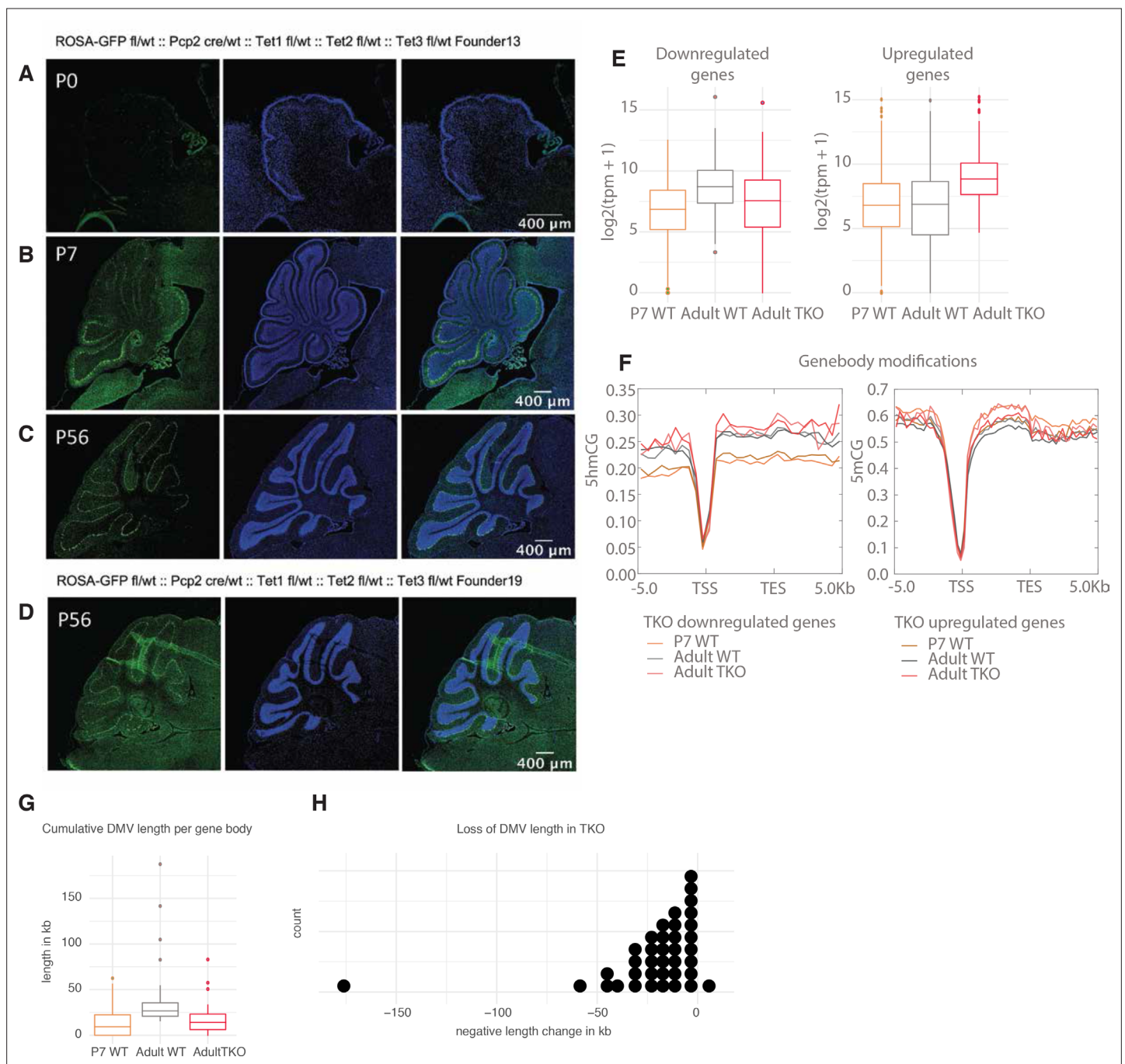


**Figure 4—figure supplement 1.** Generation of *Pcp2* cre/wt:: *Tet1* del/del:: *Tet2* del/del:: *Tet3* del/del (Purkinje cell-specific triple knockout [*Pcp2TetTKO*]) mouse founder lines. **(A)** Expression levels of *Tet1*, *Tet2*, and *Tet3* proteins at P0, P7, and adult developmental stages of Purkinje cells. **(B)** Schematic of *Pcp2TetTKO* generation via pronuclear injection. **(C)** PCR genotyping for CRE, *Tet1* deletion, *Tet2* deletion, and *Tet3* deletion in adult *Pcp2TetTKO* Purkinje genomic DNA. **(D)** Deletion of *Tet2* exons evident in nuclear RNASeq *Pcp2TetTKO*. **(E)** Example of *Pcp2TetTKO* nuclei stained with

Figure 4—figure supplement 1 continued on next page

*Figure 4—figure supplement 1 continued*

*Itpr1* post-dissociation and pre-sorting, counterstained with DAPI, a heterochromatin marker. (F) Example fluorescence-activated nuclear sorting (FANS) plot of adult *Pcp2TetTKO* nuclei. (G) Pearson correlation between *TKO* and control. (H) Integrated genome viewer (IGV) representation of Purkinje-specific markers (*Itpr1*, *Calb1*, *Car8*, *Kitl*, *Kcnp1*, *Kcnma1*, *Pcp2*) enrichment and depletion of granule (*Gabra6*, *Etv1*, *Cbln1*, *Neurod1*, *Atp2b1*, *Kcnd2*) and glial (*Gfap*, *Sorcs3*, *Aldh1l1*, *Olig2*) markers in *Pcp2TetTKO* and wild type (WT).



**Figure 4—figure supplement 2.** Cytosine dynamics in Purkinje cell-specific triple knockout (Pcp2TetTKO). (A–C) Pcp2 cre/wt:: Tet1 del/del:: Tet2 del/del:: Tet3 del/del mice were crossed to ROSA26-eGFP reporter line to evaluate CRE expression during development. Panels represent murine cerebella at P0 (A), P7 (B), and P56 (C) timepoints in FL13, stained with anti-GFP antibody. (D) Anti-GFP staining in P56 FL19. (E) Expression of genes with increased and decreased expression in Pcp2TetTKO (boxplots in left panel, heatmaps in right panel). (F) Metagenes plot of mean mCG and hmCG modifications over promoters and gene bodies of genes with increased or decreased expression in Pcp2TetTKO. (G) Cumulative DNA methylation valley (DMV) length per gene body between Pcp2TetTKO and wild type (WT). (H) Change in cumulative DMV length between adult Pcp2TetTKO and WT.