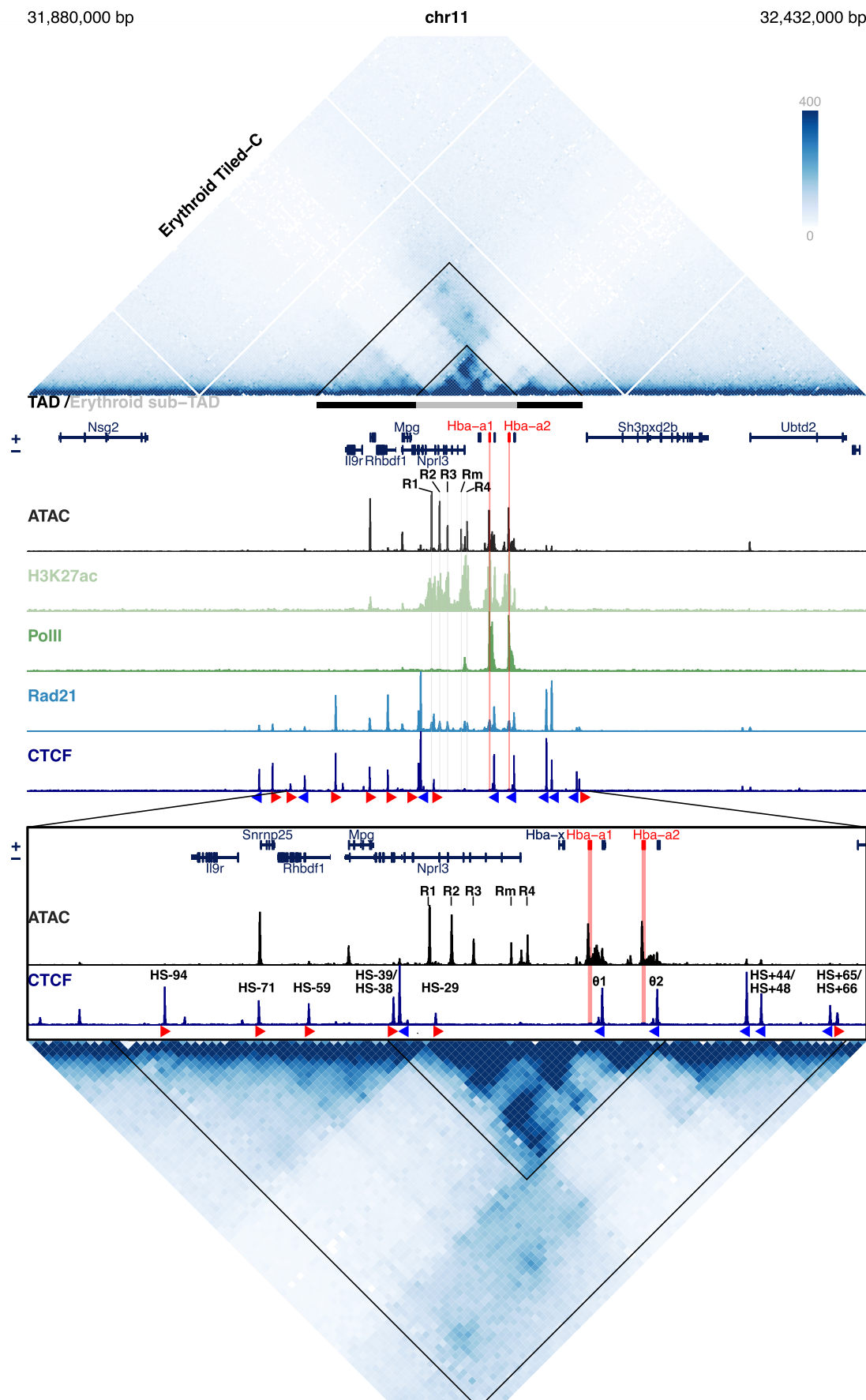


Expanded View Figures

Figure EV1. 3D structure and cis-regulatory elements of the α globin locus.

An expanded view of the α -globin locus in primary Ter119+ erythroid cells. Top panel shows Tiled-C interaction heatmap at 2 kb resolution adapted from (Oudelaar et al, 2020), the horizontal grey bars between the tracks represent the ~70 kb α globin sub-TAD (light grey, mm9 chr11:32,136,000–32,202,000) nested within a larger ~165 kb TAD (dark grey, mm9 chr11:32,080,000–32,245,000), these domains are further highlighted as triangles in the Tiled-C panel. The adult α globin genes are highlighted in red. The individual alpha globin superenhancer elements (R1, R2, R3, Rm & R4) are highlighted in grey. Below show chromatin characterisation of the region in primary definitive erythroid cells; ATAC-seq (black, this study), H3K27ac ChIP-seq (light green) (Kowalczyk et al, 2012), PolII (dark green) Rad21 (blue) and CTCF ChIP-seq (dark blue) (Hanssen et al, 2017). The orientation of CTCF motifs is shown under peaks by red (forward) and blue (reverse) arrows. Bottom panels show ATAC, CTCF ChIP-seq (labelled with CTCF site identifiers) and Tiled-C interactions zoomed in on the TAD region (mm9 chr11:32,050,000–32,250,000).



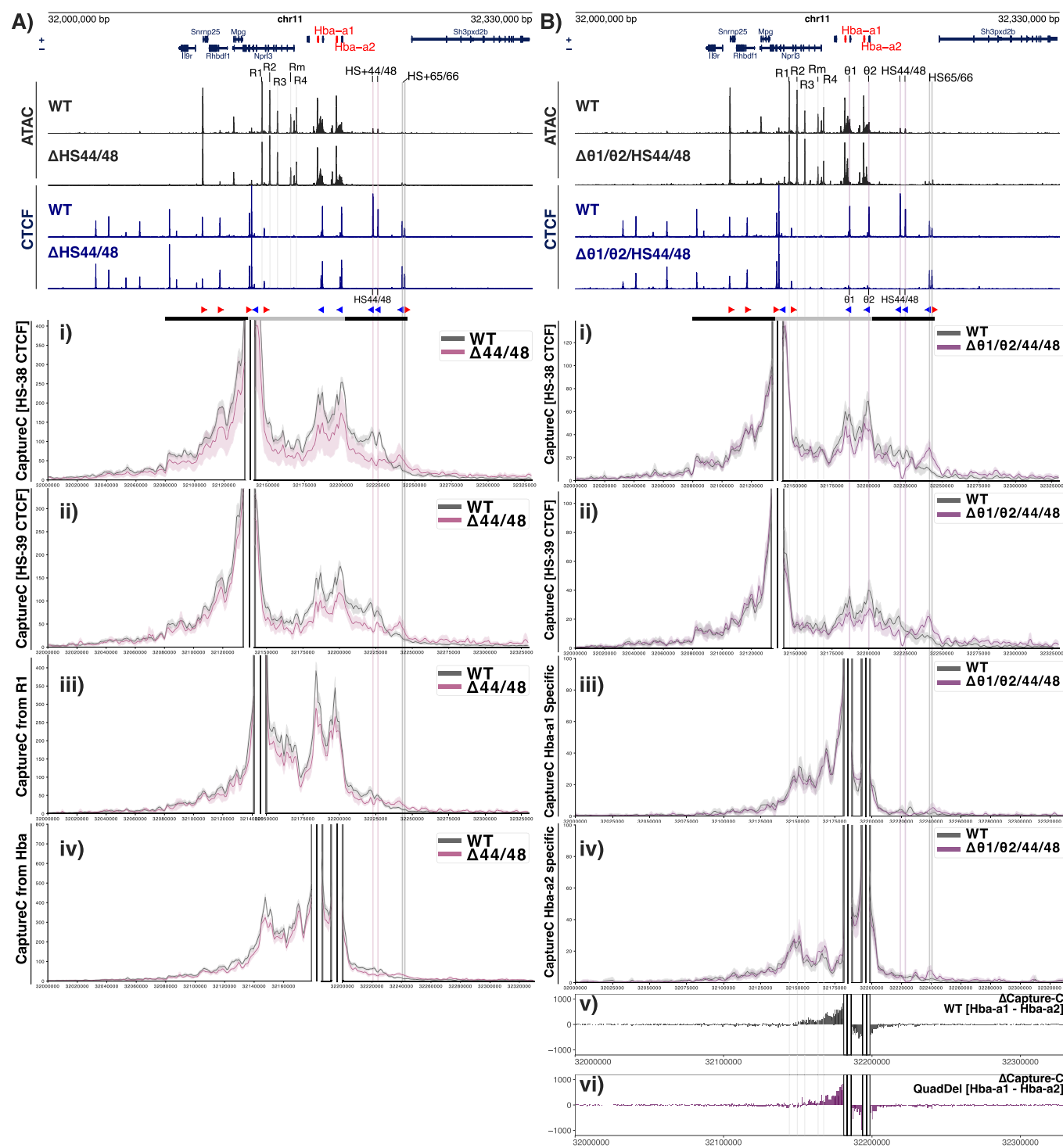




Figure EV2. Capture-C interaction profiles of the α globin locus from various viewpoints in $\Delta 44/48$ and $\Delta \theta 1/\theta 2/44/48$ erythroid cells.

(A) Top tracks show profiles for ATAC-seq (black) and CTCF ChIP-seq (navy) in primary erythroid cells (Ter119 +) isolated from WT and Δ HS44/48 mice. Profiles show normalised (RPKM) and averaged data from $n = 3$ biological replicates across the α -globin locus (mm9, chr11:32,000,000–32,330,000) with genes and genomic position, with positioning of genes above or below representing sense and antisense transcription, respectively. The adult α globin genes are highlighted in red. The individual α globin superenhancer elements (R1, R2, R3, Rm, and R4) are highlighted in grey. The horizontal grey bars between the tracks represent the ~70 kb α globin sub-TAD (light grey, chr11:32,136,000–32,202,000) nested within a larger ~165 kb TAD (dark grey, chr11:32,080,000–32,245,000). The orientation of CTCF motifs is shown under peaks by red (forward) and blue (reverse) arrows. NG Capture-C interaction profiles of the α -globin locus from WT (grey) and Δ HS44/HS48 (purple) Ter119+ primary erythroid cells, the following viewpoints: (i) HS-38 CTCF, (ii) HS-39 CTCF, (iii) R1 enhancer element and (iv) *Hba-a1/2* genes. The profiles represent normalised and averaged unique interactions from $n = 3$ biological replicates, smoothed with a 1D Gaussian filter. (B) As in (A) but showing profiles from primary APH-treated spleen cells isolated from WT and $\Delta \theta 1/\theta 2/HS + 44/HS + 48$ mice. NG Capture-C interaction profiles of the α -globin locus from WT (grey) and $\Delta \theta 1/\theta 2/HS + 44/HS + 48$ (purple) from the following viewpoints: (i) HS-38 CTCF, (ii) HS-39 CTCF, (iii) *Hba-a1* SNP-specific interactions and (iv) *Hba-a2* SNP-specific. The profiles represent normalised and averaged unique interactions from $n = 3$ biological replicates and halos representing \pm standard deviation, smoothed with a 1D Gaussian filter. Differential tracks (Δ CaptureC) show subtractions (v) WT [*Hba-a1* – *Hba-a2*] and (vi) $\Delta \theta 1/\theta 2/HS + 44/HS + 48$ [*Hba-a1* – *Hba-a2*]) of the mean number of unique interactions per restriction fragment, scaled to a total of 100,000 interactions in cis from SNP-specific counts.

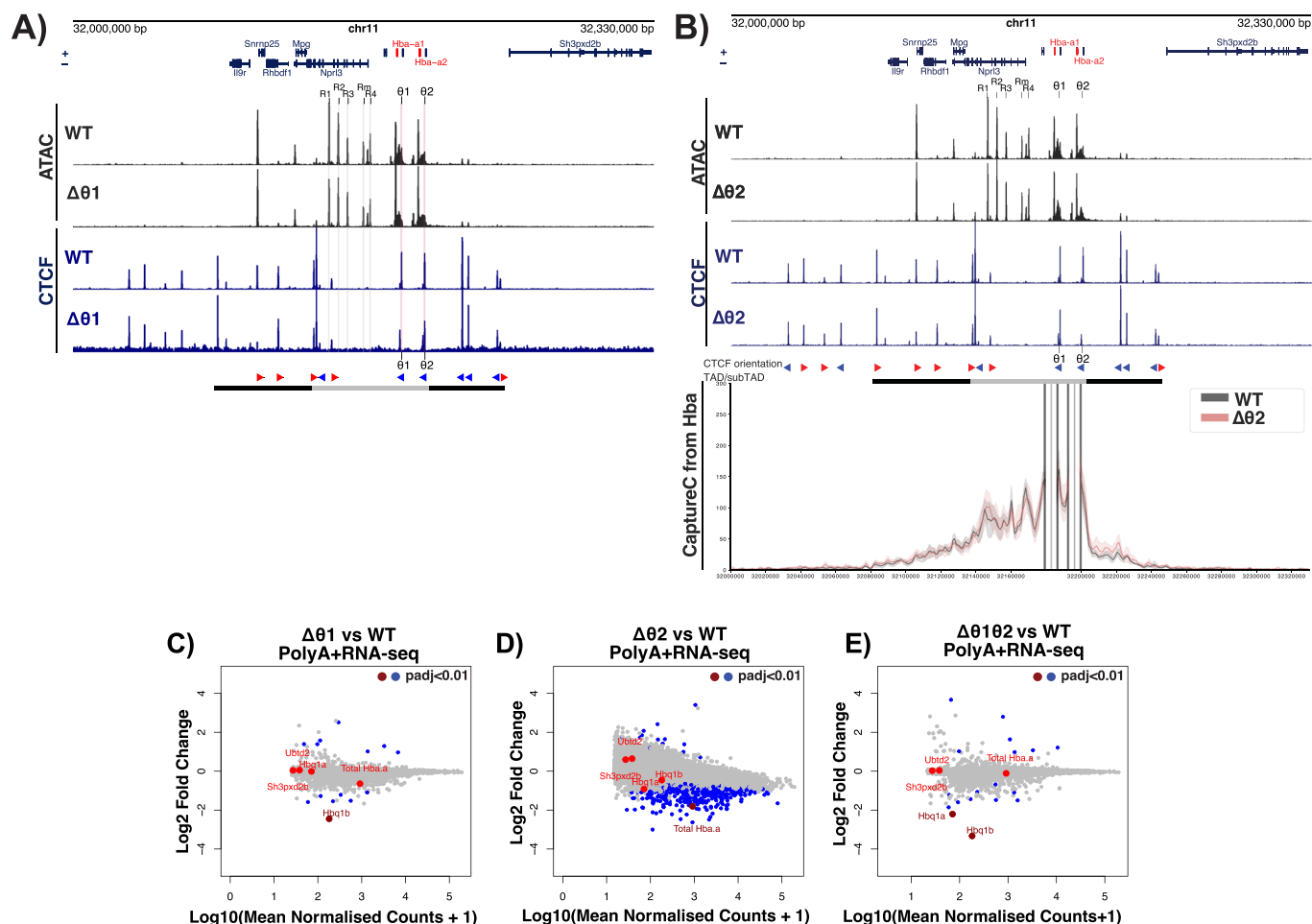
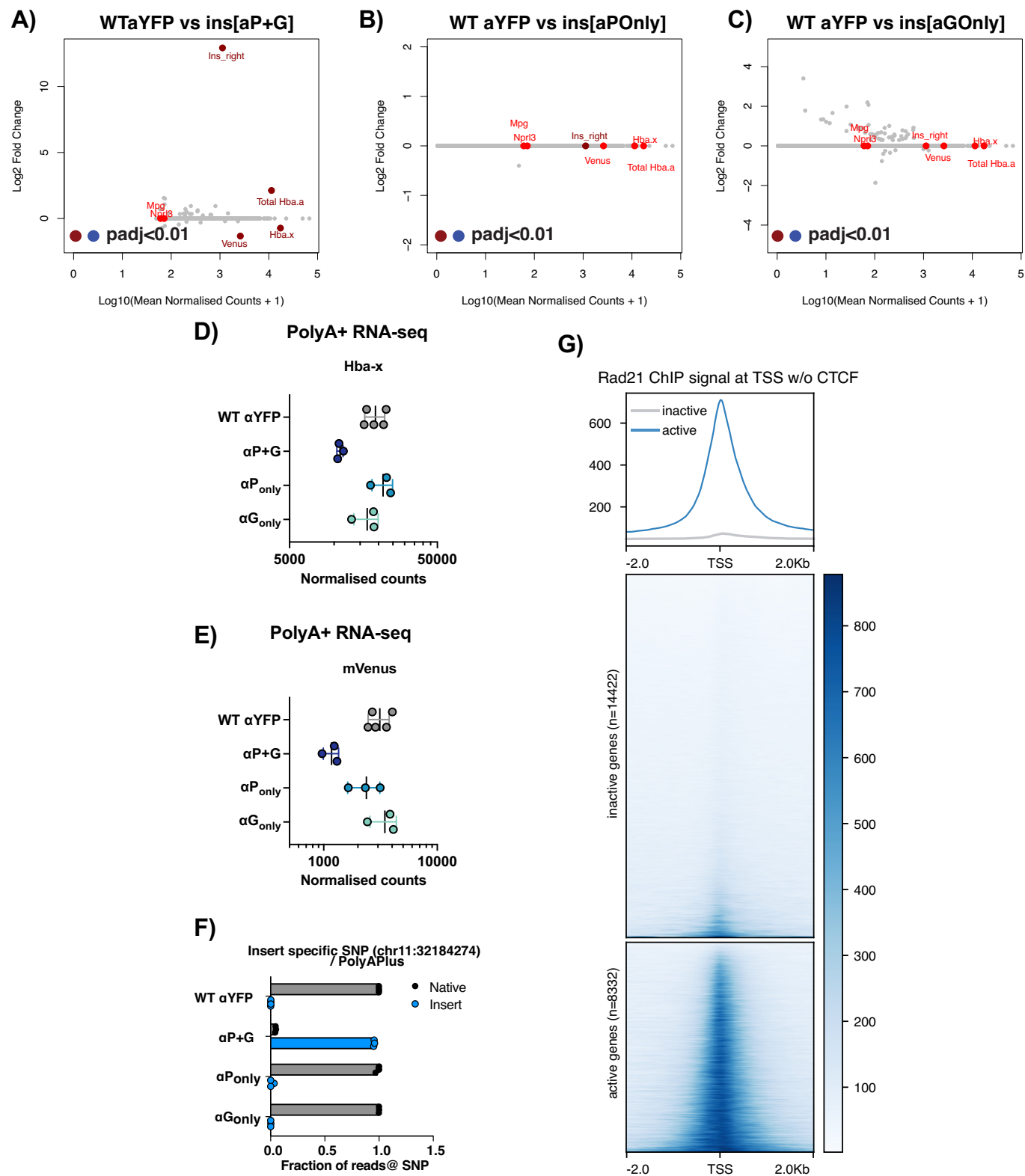


Figure EV3. Chromatin accessibility, CTCF binding and PolyA+RNA-seq in $\Delta\theta 1$ and $\Delta\theta 2$ erythroid cells.

(A) Characterisation of $\Delta\theta 1$ primary erythroid cells. Top tracks show profiles for ATAC-seq (black) and CTCF ChIP-seq (navy) in primary erythroid cells (Ter119+) isolated from WT and $\Delta\theta 1$ mice. Profiles show normalised (RPKM) and averaged data from $n = 3$ biological replicates across the α -globin locus (annotations are as in Fig. 1A). (B) Characterisation of $\Delta\theta 2$ primary erythroid cells. Top tracks show profiles for ATAC-seq (black) and CTCF ChIP-seq (navy) in primary erythroid cells (Ter119+) isolated from WT and $\Delta\theta 2$ mice. Profiles show normalised (RPKM) and averaged data from $n = 3$ biological replicates across the α -globin locus (annotations are as in Fig. 1A). NG Capture-C interaction profiles of the α -globin locus from the combined viewpoint of the *Hba* genes each with an exclusion zone, in WT (grey) and $\Delta\theta 1/\theta 2$ (pink) Ter119+ primary erythroid cells. The interaction profiles represent normalised and averaged unique interactions from $n = 2$ biological replicates and halos representing \pm standard deviation, smoothed with a 1D Gaussian filter. (C-E) Differential expression (PolyA+RNA-seq) in $\Delta\theta 1$, $\Delta\theta 2$ and $\Delta\theta 1/\theta 2$ primary erythroid cells. MAplot of Log2 Fold change versus Log10 of normalised counts in the models above vs WT; each dot represents a gene. Genes with an adjusted P value (p_{adj} , Benjamini-Hochberg corrected) < 0.01 are highlighted in blue. Genes of interest are highlighted in red (3' genes *Sh3pxd2b*, *Ubt2*, Total Hba (*Hba-a1/2*) and $\theta 1/\theta 2$ associated genes *Hbq1b/a* respectively) and those with a significant difference from WT highlighted with dark red. Results from $n = 3$ biological replicates from each genotype. There is an unexpectedly high number of differential genes in the $\Delta\theta 2$ model further *Hbq1a* appears unchanged upon $\Delta\theta 2$, which is incongruent with the result in $\Delta\theta 1/\theta 2$. As the $\Delta\theta 1/\theta 2$ does not show these differences and is a combinatorial deletion of both $\theta 1/\theta 2$ and we must assume these differences in $\Delta\theta 2$ expression are due to technical error.



◀ **Figure EV4. Extended characterisation of in vitro-derived CD71⁺ cells with inserts.**

(A–C) Differential expression between the WT aYFP reporter and the insertion models. MAplot of Log2 fold change in gene expression relative to WT against the Log10 of read counts; each dot represents a gene. Highlighted in red are *Hba-a* (representing total *Hba-a1/2* and aP+G in the P + G model), the region downstream of the insertion site (*ins_right*) and other genes in the locus. Significant changes plotted with dark red or blue (Wald test *P* value, Benjamini–Hochberg corrected: *Padj* < 0.01), non-significant changes are in grey and bright red. Data is *n* = 5 replicates of WT-aYFP and *n* = 3 replicates of each other genotype. As these libraries were not globin-depleted, the representation is skewed toward globin genes. (D) *Hba-x* expression as normalised counts in PolyA-Plus RNA-seq. Data is representative of *n* = 5 replicates of WT-aYFP and *n* = 3 replicates of each other genotype. Error bars represent ± standard deviation. (E) *mVenus* expression as normalised counts in PolyA-Plus RNA-seq. Data is representative of *n* = 5 replicates of WT-aYFP and *n* = 3 replicates of each other genotype. Error bars represent ± standard deviation. (F) Exonic SNP-specific count PolyA-Plus RNA-seq. The aP+G gene had its own unique SNP in exon 3 allowing counting of the proportion of transcripts between the inserted or native α-globin copies. SNPs specific counts of each variant of *Hba* were counted similarly to Fig. 1E in RNA-seq from the CD71⁺ cells. Error bars represent ± standard deviation. (G) Heatmap and summary profile displaying Rad21 ChIP-seq signal in WT in vitro-derived CD71⁺ erythroid cells at non-redundant transcription start sites (TSS) of inactive (14422) and active (8332) genes, which do not have a CTCF binding site within a 2 kb window around the TSS. Source data are available online for this figure.