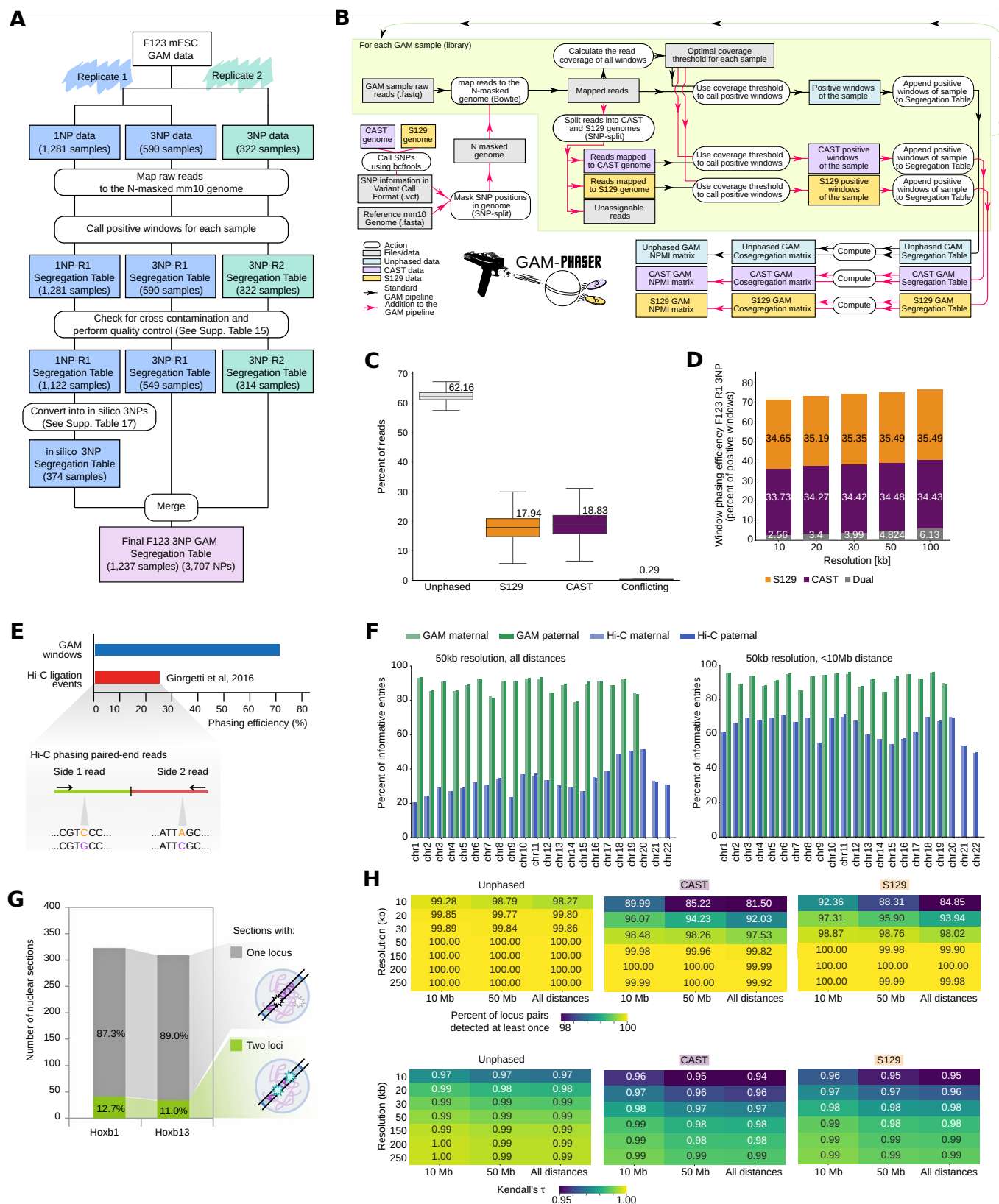
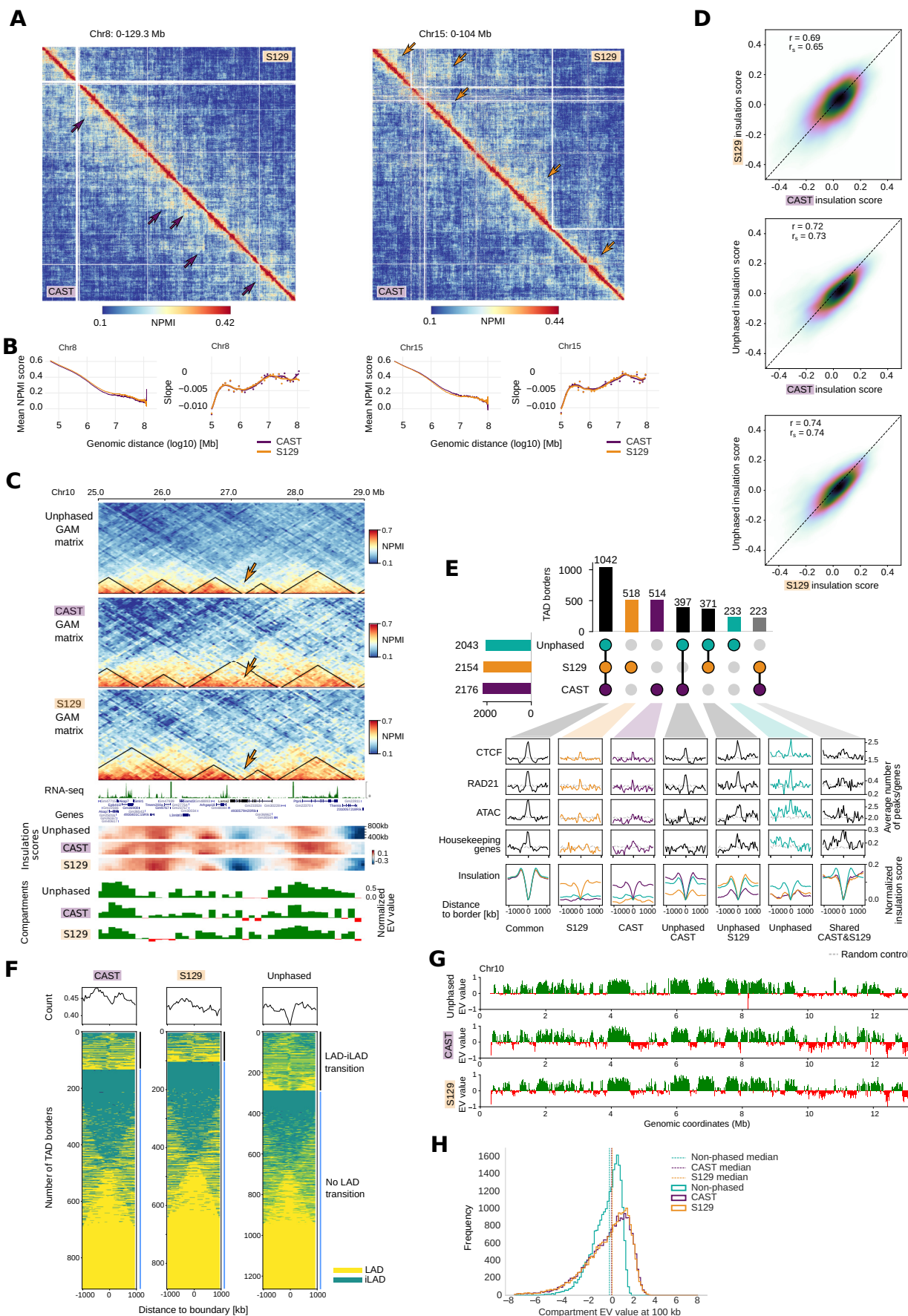


## Expanded View Figures

### Figure EV1. Strategy, methodology and evaluation of GAM-Phasing pipeline for allele-specific contact maps.

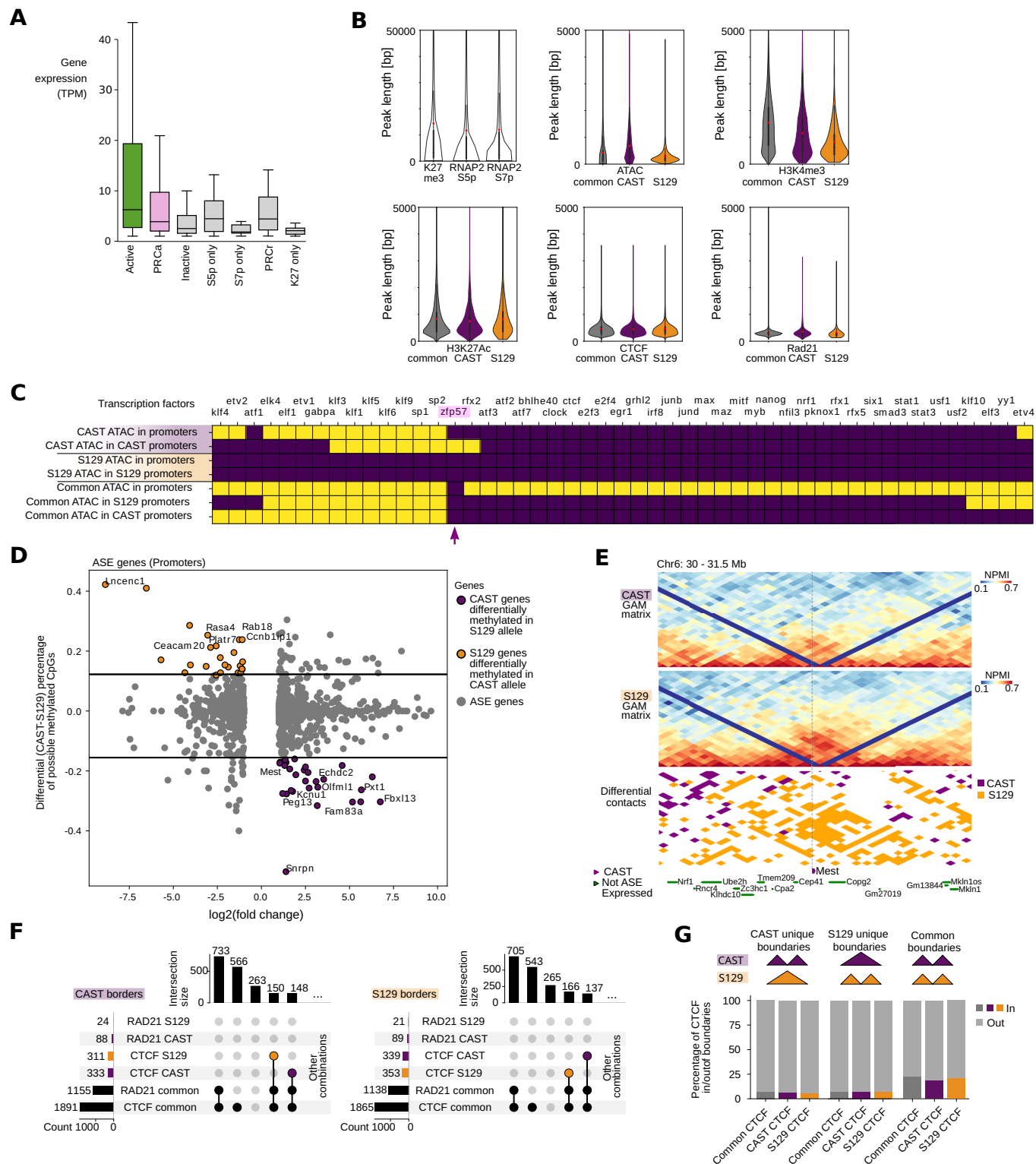
(A) Schematic overview of GAM data collection, quality control steps, and merging of replicates R1 and R2. (B) GAM-phaser pipeline. (C) Percentage of reads that were phased to each allele. Conflicting reads are reads containing SNPs from both alleles. (D) Percentage of phased positive windows in the entire segregation table for all F123 3NPs passed quality controls GAM samples. (E) Phasing efficiency between GAM and Hi-C. GAM efficiency is measured as phased windows divided by the total number of called windows, while Hi-C efficiency is calculated dividing phased ligation events to unique ligation events; reported phasing efficiency was obtained from (Giorgetti et al, 2016). Below, schematic of a phaseable Hi-C ligation event. (F) Number of informative contact entries in the phased F123 GAM dataset in comparison with phased Hi-C data collected for human GM12878 B-lymphoblastoid cells (Rao et al, 2014b), at all intrachromosomal distances and for distances up to 10 Mb. (G) Number of nuclear sections that are positive for the presence of two or one *Hoxb1* or *Hoxb13* locus detected by cryo-FISH using 40 kb fosmid probes ( $n = 341$  *Hoxb1* loci,  $n = 362$  *Hoxb13* loci,  $n = 2584$  nuclear sections imaged; data source from Barbieri et al, 2017). (H) Percentage of locus pairs detected at least once and Kendall's  $\tau$  coefficient values for different resolutions and different distances. These metrics were used to decide on optimal resolutions of the maps.





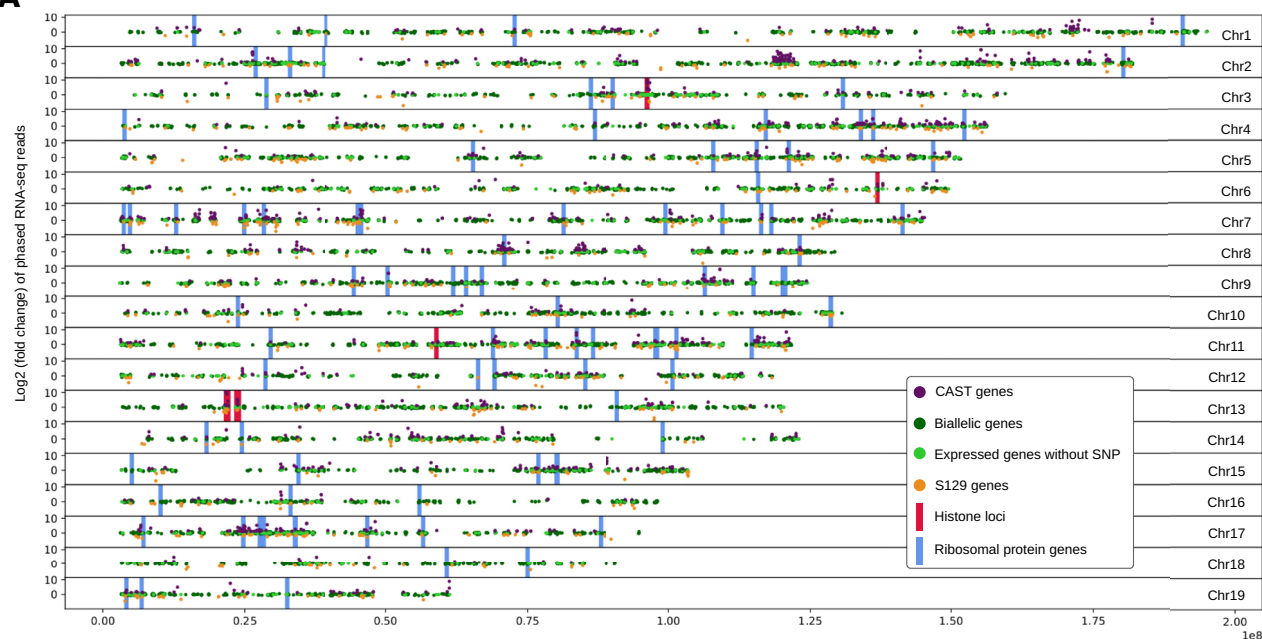
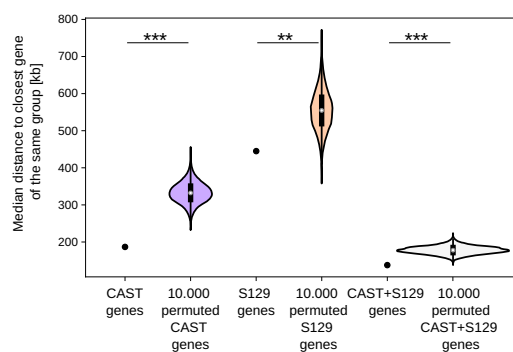
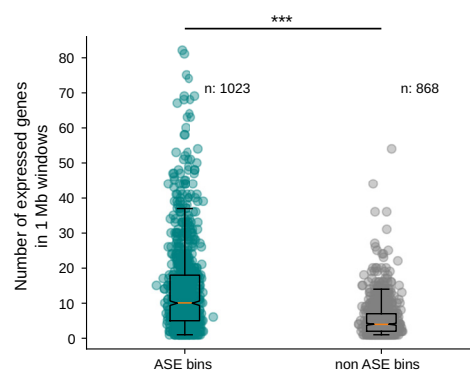
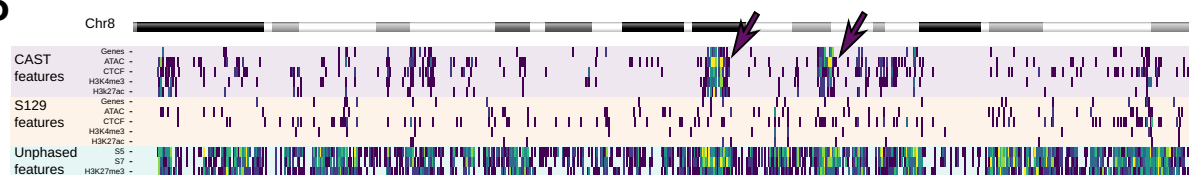
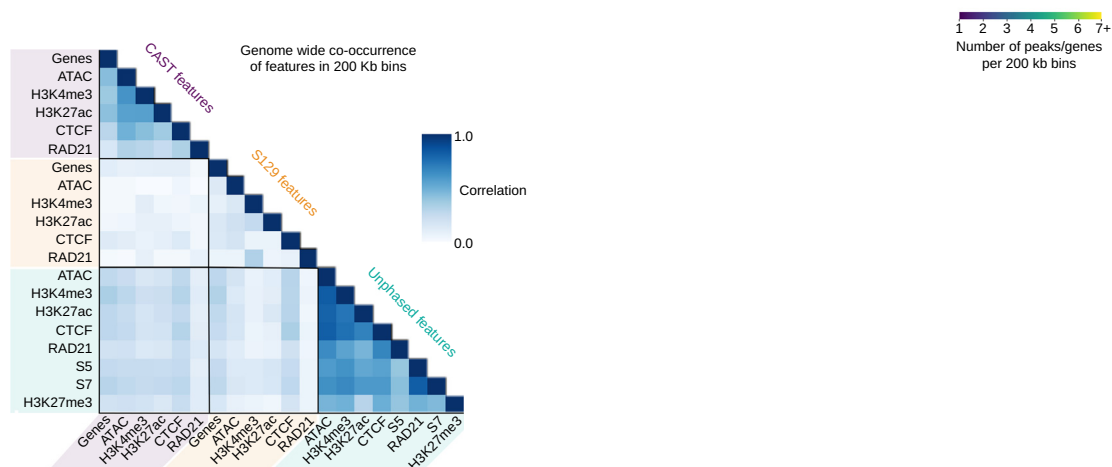
◀ **Figure EV2. Comparison of insulation, TADs and compartments between S129 and CAST haplotypes.**

(A) GAM matrices of chromosomes 8 and 15 showing both alleles at 50 kb resolution. Colored arrows show structural differences between alleles. (B) Distance decay curves and momentum curves for contact intensities across all distances in CAST and S129 chromosomes 8 and 15. (C) 4 Mb region in chromosome 10 showing an allele-specific TAD border in the S129 allele. Below, RNA-seq track, insulation scores and compartment tracks for all maps. (D) Pearson correlation coefficient ( $r$ ) between combinations of CAST, S129 and unphased insulation scores at 400 kb. (E) Upset plot of TAD border combinations between CAST, S129 and the unphased maps. Below, aggregate plots for CTCF, Rad21 and ATAC-seq peaks and housekeeping genes, centered at the TSS ( $\pm 1$  kb). Normalized Insulation score is also shown for each group. (F) Overlap of LADs and iLADs with  $\pm 1,000$  kb around CAST, S129 and common TAD borders, computed from 100 kb resolution GAM matrices to match LAD annotations. Each heatmap is clustered depending on whether the border overlaps with a LAD/iLAD transition or not. (G) Compartment tracks for CAST, S129 and the unphased maps for chromosome 10. (H) Compartment eigenvector values distribution for CAST, S129 and the unphased datasets. Discontinuous lines show the median for each dataset.



◀ **Figure EV3. Association of allele-specific epigenetic marks and transcription factors with promoters of ASE genes and 3D genome organization.**

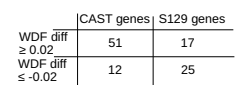
(A) Gene expression of the different groups in Fig. 2E. (B) Distribution of Pol2-S5p, Pol2-S7p and H3K27me3 peaks, and phased and unphased ATAC-seq, H3K4me3, CTCF, H3K27ac and RAD21 peak sizes. Red dots indicate the average size for each dataset. (C) Heatmap showing the enriched presence (cutoffs Q value  $\leq 0.05$  and P value of  $\leq 0.001$ ) of different transcription factors that overlap with the peaks of different ATAC-seq groups. ZFP57 is the only transcription factor enriched for an allele-specific group. (D) ASE gene promoters regarding their differential percentage of methylated CpGs. Colored are those genes with a significant amount of methylated CpGs in their promoter (top and bottom 5%) in the allele they are not expressed. (E) CAST and S129 GAM matrices for the *Mest* locus (Chr6: 30–31.5 Mb). Below, differential contacts and two tracks showing CAST genes and expressed genes. (F) Most borders contain common CTCF and RAD21 or only CTCF and each allele has a similar number of CTCF specific to either of the alleles in its borders. (G) Percentage of CTCF peaks that are inside or outside borders.

**A****B****C****D****E**

◀ **Figure EV4. Co-presence of ASE genes and chromatin features in the linear genome.**

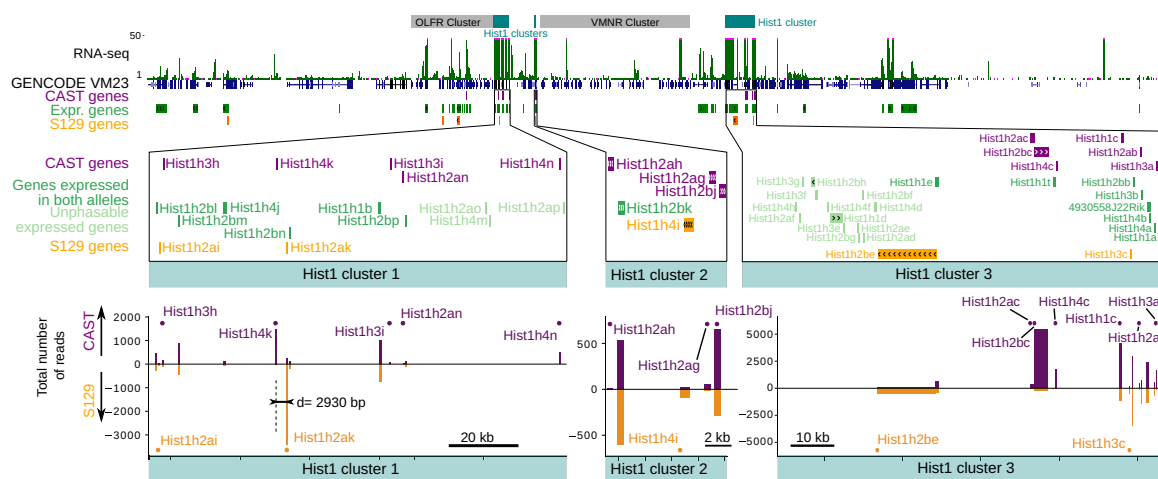
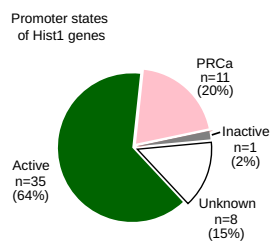
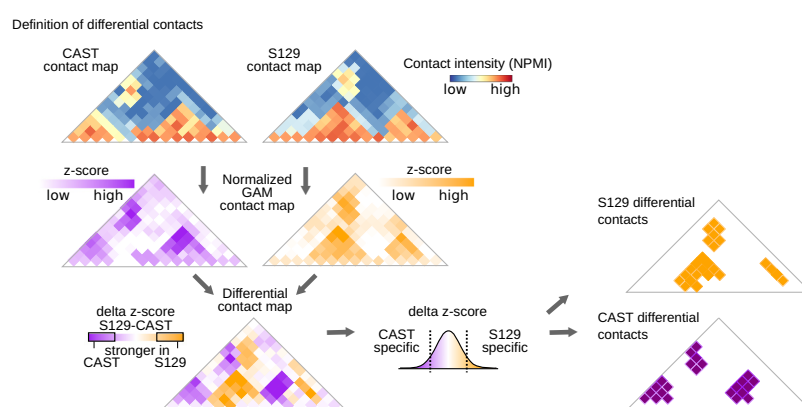
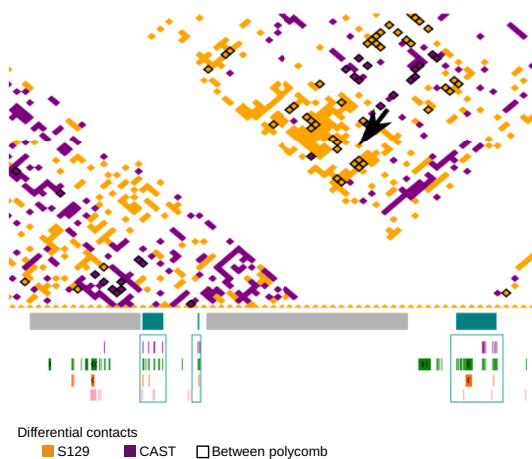
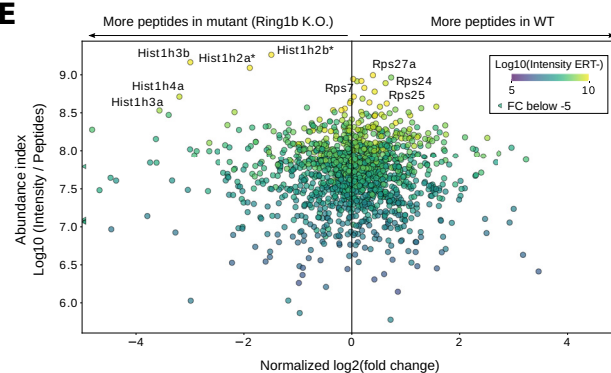
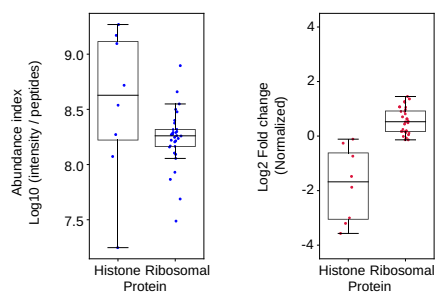
(A) Genomic position of all expressed genes: CAST genes, S129 genes, biallelic genes and genes without SNPs. Red and Blue bars indicate the position of Histone protein genes and Ribosomal protein genes. (B) Each of the 3 dots indicate the average distance of all genes of each type (CAST, S129 and CAST or S129) to the closest gene of that type. The violin plot shows the distribution of these averages if we permute the position of the genes 10,000 times. The permutation is carried out by randomly selecting the same number of CAST, S129 or CAST + S129 genes from all expressed genes.  $P$  values = 0.0001, 0.0145, 0.0001 for CAST, S129 and CAST + S129. (C) 1 Mb windows containing at least 1 ASE gene tend to contain more expressed genes than 1 Mb windows that do not contain ASE genes. T test:  $P$  value =  $2.7 \times 10^{-71}$ . Number of ASE windows, 1023. Number of non-ASE windows, 868. (D) Genomic location of allele-specific features (genes, ATAC-seq, CTCF, H3K4me3 and H3K27ac peaks) and unphased features (Pol2-S5p, Pol2-S7p and H3K27me3 peaks) and their density in bins of 200 kb. Arrows indicate two regions with an enrichment of CAST-specific features. (E) Genome-wide Pearson correlation ( $r$ ) of the co-occurrence of the features in (D). CAST features correlate well between each other while S129 features do not.





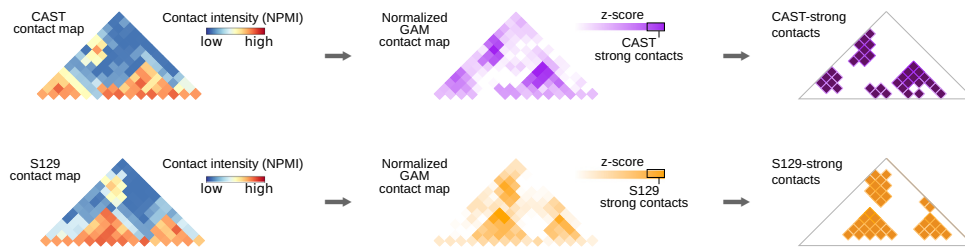
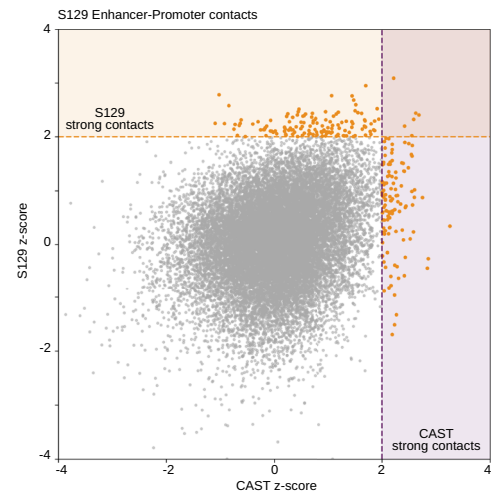
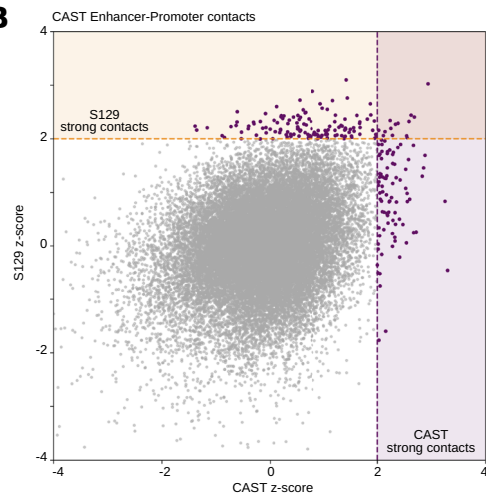
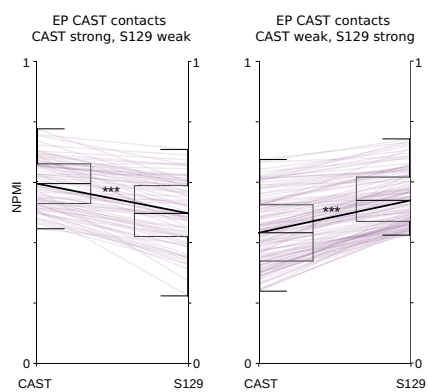
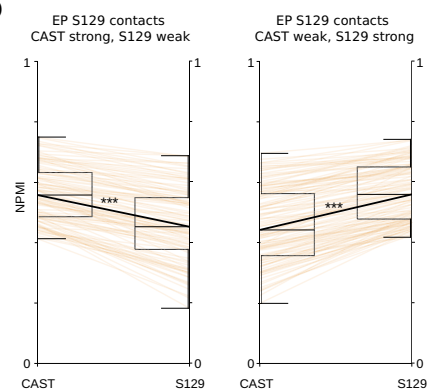
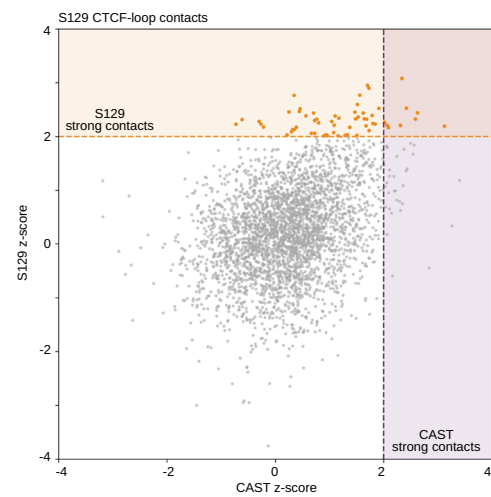
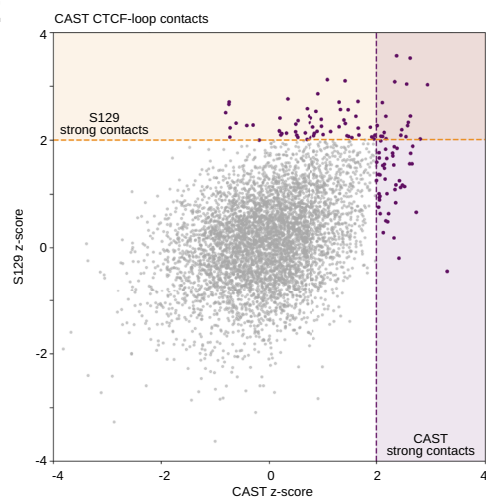
**Figure EV5. Comparative analysis of CAST and S129 alleles in genome compartmentalization and chromatin accessibility.**

(A) Normalized eigenvector (EV) values for the CAST and S129 allele for each 100 kb bin. Color coded are bins containing only not expressed genes, bins containing biallelic genes but not ASE genes, bins containing at least one CAST gene but not S129, bins containing at least one S129 gene but not CAST genes and bins containing at least one CAST gene and one S129 gene. (B) Percentage of ATAC-seq, CTCF, H3K4me3 or H3K27ac peaks in each compartment combination (A and A, B and B, A and B or B and A for CAST and S129 alleles, respectively). CAST-specific features show a tendency to overlap more in A/B (A specific compartment in the CAST allele), S129-specific features tend to overlap more in B/A (A specific compartment in the S129 allele.). (C) UpSet plots showing for the S129 allele, groups of TADs containing different sets of types of genes and their number. (D) Relation between the TAD length, the number of expressed genes in a TAD, and number of genes specific to that allele (dot size) for TADs in CAST and S129. Purple refers to TADs containing CAST genes, orange to TADs containing S129 genes, and gray to TADs containing CAST and S129 genes (for both CAST allele and S129 allele, respectively). (E) Violin plots showing the number of genes per TAD (observed, Obs.) compared to circular permutations of gene positions in the genome (permuted, Perm.). 10,000 permutations were done for each of the 4 examples in (D) and are compared to the number of genes per TAD in the original data (called *real*). All *P* values are  $\leq 0.0001$ . Numbers are 911, 889, 1265 and 1265, respectively. (F) Related to (C), number of H3K27me3 peaks normalized by TAD length (two-sided *t* test:  $*P \leq 0.05$ ,  $**P \leq 0.01$ ,  $***P \leq 0.001$ ; *P* values from top to bottom in S129 TADs:  $1.9 \times 10^{-14}$ ,  $3.2 \times 10^{-30}$ ,  $1.1 \times 10^{-10}$ , 0.00011,  $1.8 \times 10^{-5}$ , n.s.: 0.5092. Number of TADs with: expressed genes, 737; not expressed genes, 483; CAST genes, 285; S129 genes, 232; and with CAST and S129 genes, 262). (G) Loci with the same genomic length can have different volumes due to varying compaction. Decompacted loci with larger volumes are captured more frequently in the collection of GAM cryosections than more compacted chromatin. Window Detection Frequency (WDF) is a GAM-intrinsic measure of relative chromatin compaction, defined by the number of locus detection events in the collection of GAM nuclear slices (Beagrie et al, 2017). From phased window segregation tables, the WDF can be calculated separately for CAST and S129, as a measure of relative compaction between all loci in each haplotype. (H) Related to (C), for each group, the differential (CAST-S129) window detection frequency is represented. Positive values indicate decompaction in the CAST allele, while negative values indicate decompaction in the S129 allele (two-sided *t* test:  $*P \leq 0.05$ ,  $**P \leq 0.01$ ,  $***P \leq 0.001$ ; *P* values from top to bottom for S129 TADs: 0.031, 0.025). Number of TADs analyzed are the same as in (F). (I) Window detection frequency (WDF) values in the CAST and S129 allele for each bin containing genes. Fisher's exact test ( $P = 5.3 \times 10^{-5}$ ) shows the significant tendency of windows with high WDF in the CAST allele containing CAST genes and windows with high WDF in the S129 allele containing S129 genes compared to windows with lower WDF. Numbers for CAST and S129 genes with differential WDF  $\geq 0.02$  are 51 and 17, respectively. For CAST and S129 genes with differential WDF  $\leq -0.02$  are 12 and 25, respectively.

**A****B****C****D****E****F**


**Figure EV6. Differential gene regulation at the Hist1 gene cluster.**

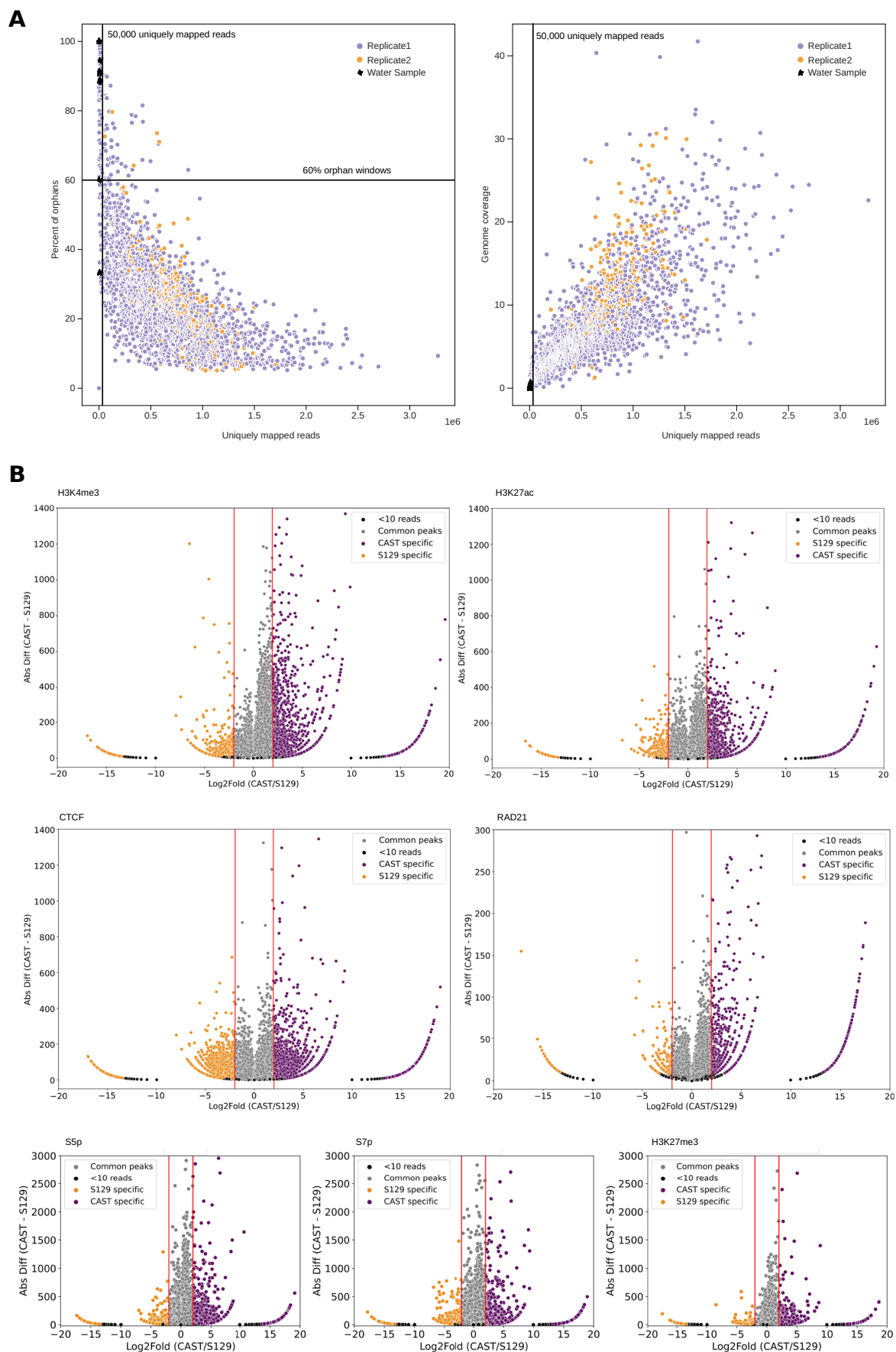
(A) Hist1 locus region from Fig. 4A with greater detail. Histone genes are depicted. Number of reads phased for the CAST or S129 allele are shown for all genes in the Hist1 clusters. (B) Proportion of Hist1 genes according to promoter state. (C) Schematic showing the pipeline used to compare CAST and S129 contact intensities and to extract CAST-specific and S129-specific contacts. (D) Allele-specific contacts at the Hist1 locus as shown in Fig. 4A. Black squares show the contacts where H3K27me3 peaks are present in both windows of the contact. (E) SILAC experiments were performed in the ESC-ERT2 cells in the presence and absence of tamoxifen to induce knockout of *Ring1b*, in three biological replicates. *Ring1b* knockout results in upregulation of histone proteins. Abundance was estimated by the ratio of intensity and number of peptides. Normalized log2 fold change was calculated applying the z-score normalization to the log2 of heavy/light (H/L) ratio of the untreated experiment divided by the H/L ratio of the conditional knockout. (F) Boxplots showing the abundance index and the log2 fold change for detected histone proteins and ribosomal proteins. Numbers of data points are 8, 32, 8, 32, respectively from left to right.

**A** Definition of strong contacts**B****C****D****E**



**Figure EV7. Strong allelic contacts bridge promoters, enhancers, and CTCF sites.**

(A) Schematic showing the strategy to identify strong allelic contacts. (B) All possible contacts involving conditions for CAST-specific enhancer-promoter contacts and S129-specific enhancer-promoter contacts. Lines mark cutoffs for strong and allele-specific contacts in each haplotype. (C) Differences in contact intensities observed in the CAST and S129 haplotypes for allele-specific enhancer-promoter (E-P) elements associated with CAST contacts that were found to be strong in CAST but weak in S129 (on the left), or strong in S129 but weak in CAST (on the right). Two sample *t* test: *P* values are 4.17e-10 and 8.78e-17, respectively. Numbers are 84 and 130 respectively. (E) All possible contacts involving conditions for CAST-specific CTCF loops and S129-specific CTCF loops. Lines mark cutoffs for strong and allele-specific contacts in each haplotype. (D) Differences in contact intensities observed in the CAST and S129 haplotypes for allele-specific enhancer-promoter (E-P) elements associated with S129 contacts that were found to be strong in CAST but weak in S129 (on the left), or strong in S129 but weak in CAST (on the right). Two sample *t* test: *p* values are 3.92e-12 and 4.42e-14, respectively. Numbers are 112 and 129, respectively.



**Figure EV8. Assessment of read-count thresholds in phasing.**

(A) Distribution of percentage of orphan windows, uniquely mapped reads and genome coverage in each GAM sample. Replicate 1, replicate 2 and water (ONP) samples are shown. Thresholds used to remove potentially low from high quality GAM samples are shown in vertical and horizontal black lines. (B) Distribution of phased H3K4me3, H3K27ac, CTCF, RAD21, Pol2-S5p, Pol2-S7p, and H3K27me3 peaks, showing their absolute difference in phased reads (CAST-S129) and their log2 fold change.