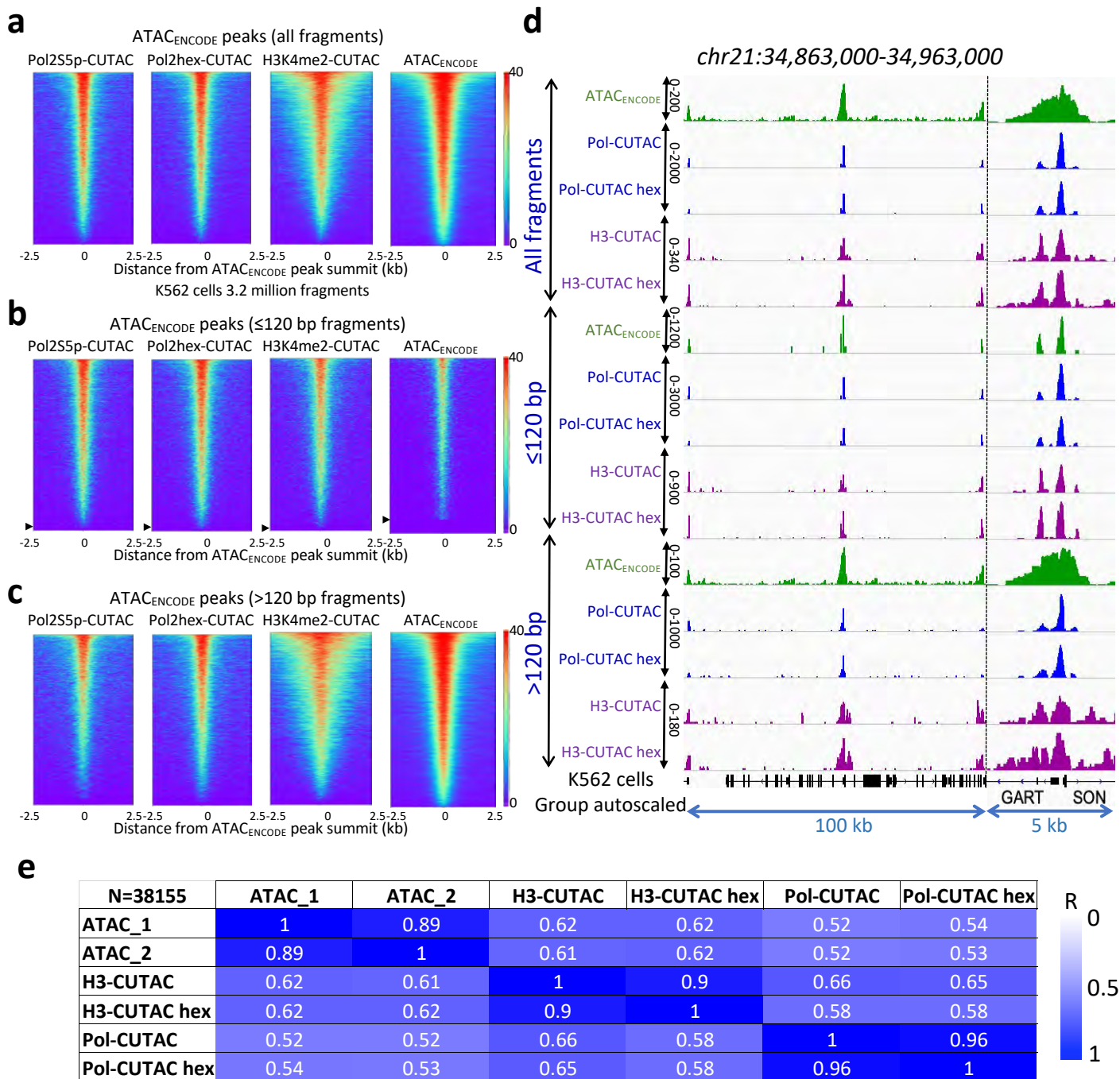
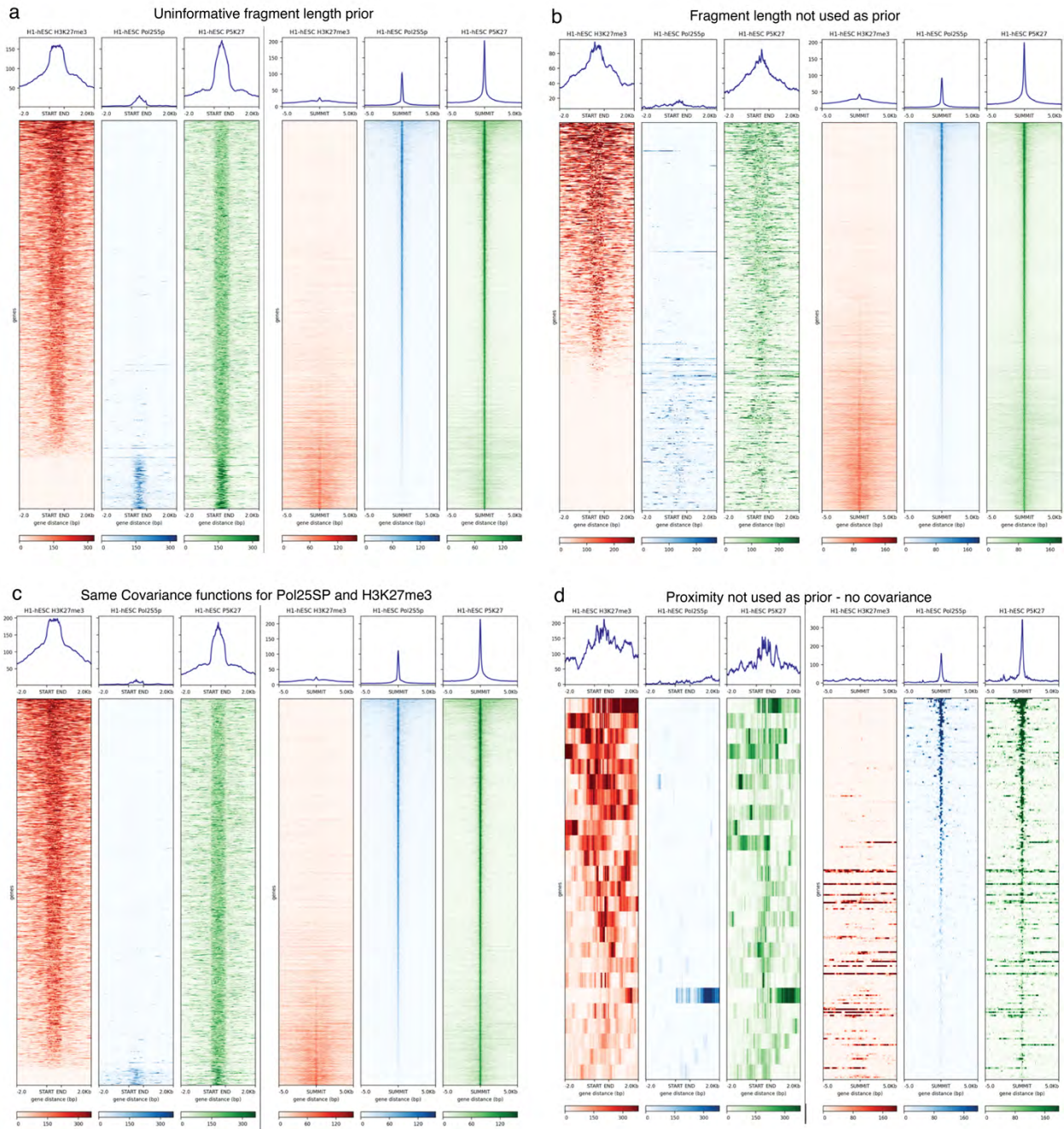


Additional file 1: Fig S1. Feature definition and fragment length separation under CUTAC conditions. (a)

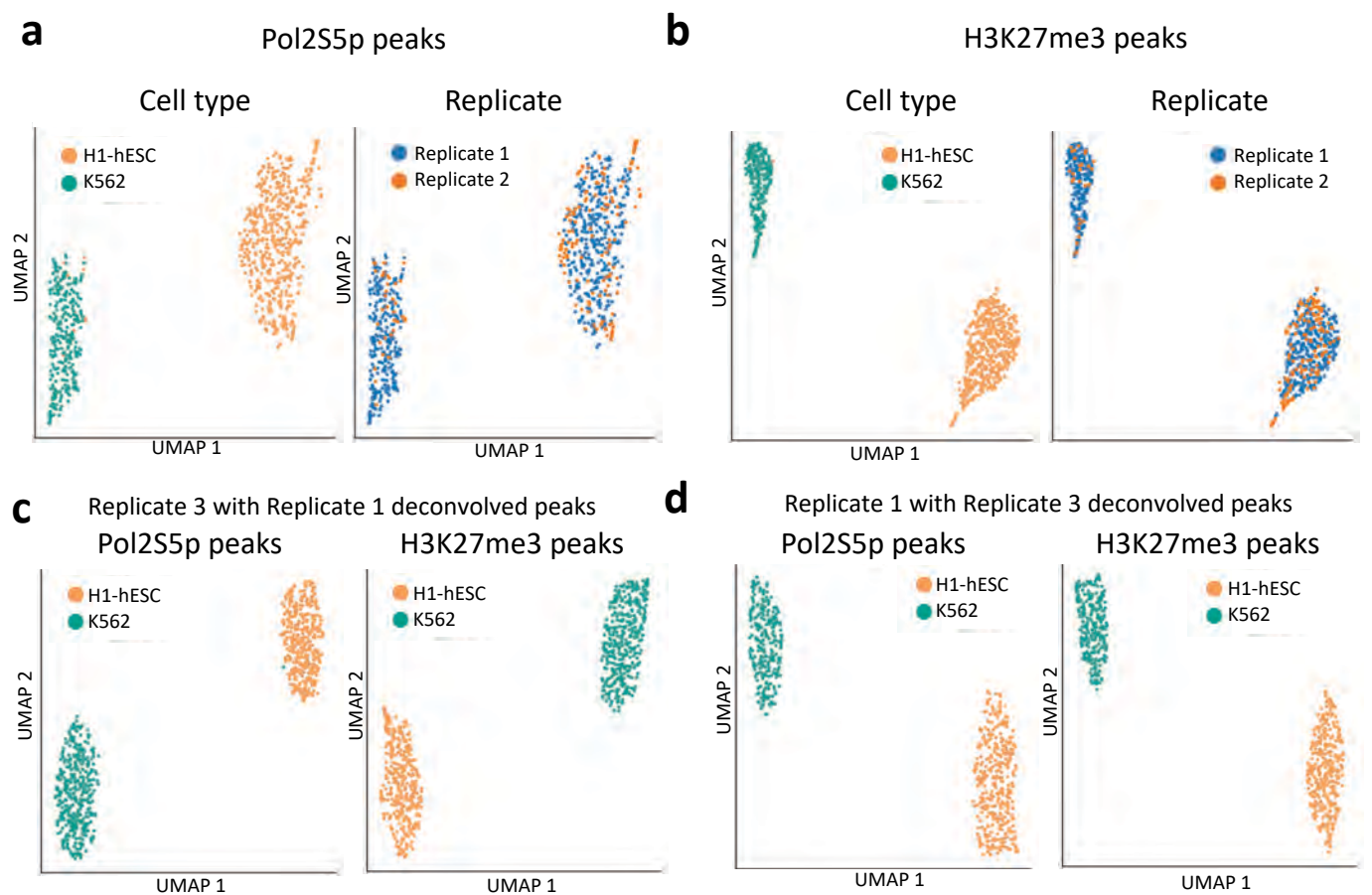
Fragments were mapped to hg19, and 3.2 million fragments were randomly sampled from each dataset and used to make bedgraph tracks. A representative region is shown. To compare peaks with very different signal-to-noise levels, samples were group-autoscaled with ranges indicated to the left of each set of tracks. Pol2S5p CUTAC of K562 cells with linear pre-amplification using only P5 primers for 12 cycles was followed by addition of P7 primers and PCR for various numbers of cycles. (b) Size distributions were not affected by differences in the number of PCR cycles following linear amplification. (c) Fragment length distributions for K562 cells (left) and H1 ES cells (right) are shown for linear pre-amplified fragments after tagmentations in 10 mM TAPS + 5 mM MgCl₂ ± 1,6-hexanediol or 300 mM NaCl + 10 mM MgCl₂ plotted as fractions of the total mapped fragments for linear pre-amplified datasets. Percentages of total fragments ≤120 bp are shown. Tagmentation in 1,6-hexanediol generally results in a smaller fragment distribution, especially conspicuous for H1 cell nucleosomes. The higher recovery of nucleosome-sized fragments from K562 cells than H1 ES cells reflects the much lower abundance of Polycomb domains in H1 cells.



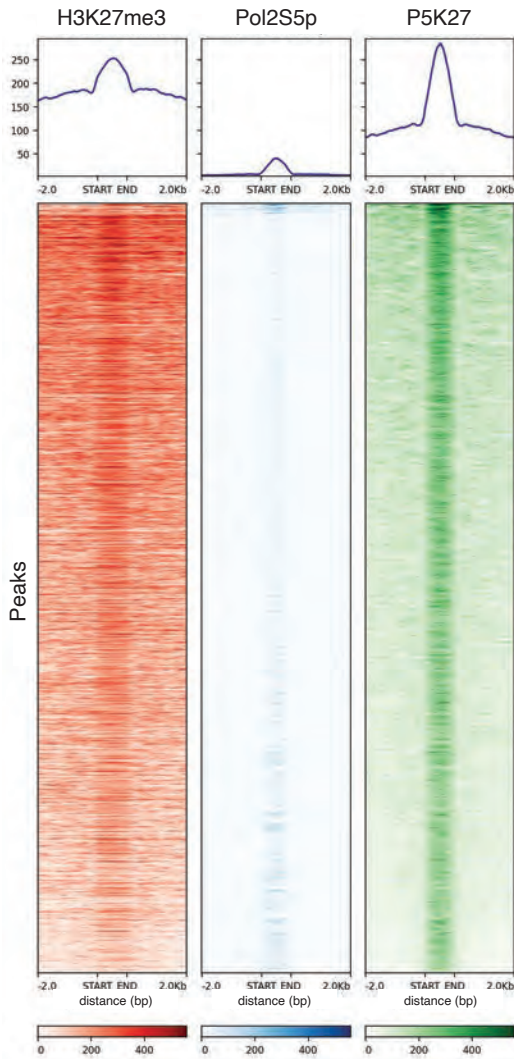
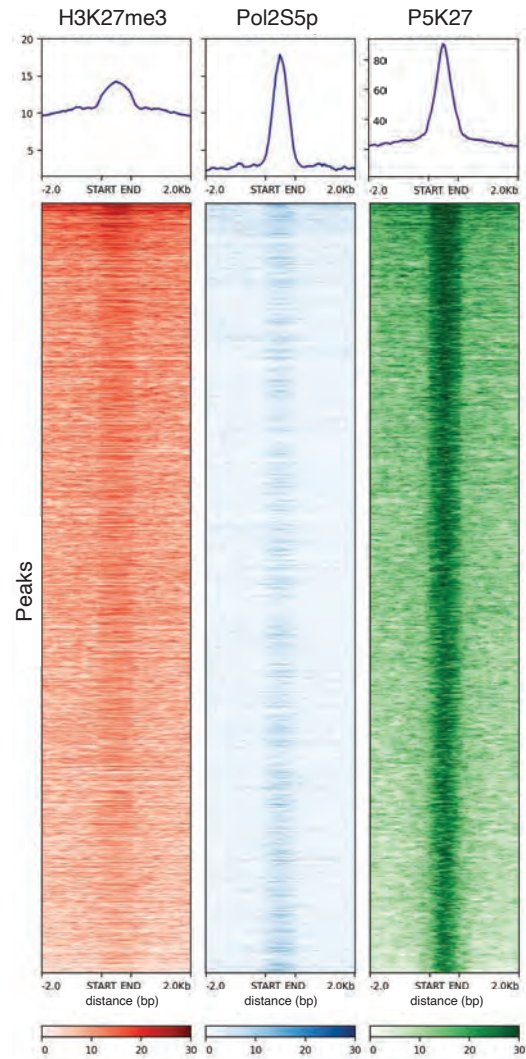
Additional file 1: Fig. S2. Close correspondences between CUTAC and ATAC-seq. (a) Heatmaps showing alignments of Pol2S5p-CUTAC $\pm 10\%$ 1,6-hexanediol during tagmentation, H3K4me2-CUTAC and ATAC_{ENCODE} data over peaks called on an ATAC_{ENCODE} replicated dataset. (b) Same as (a) showing the ≤ 120 -bp fragment subset, where arrowheads indicate the row where enrichment of signal over peaks ends: 98-99% for CUTAC and 93% for ATAC_{ENCODE}. (c) Same as (b) showing the > 120 bp subset. Fragments were mapped to hg19, and 3.2 million fragments were randomly sampled from each dataset and used to make bedgraph tracks. (d) Tracks comparing Pol2S5p and H3K4me2 CUTAC to ATAC-seq using K562 cell data generated by the ENCODE project. The region around the GART-SON bidirectional promoter is shown at two different scales. (e) Correlation matrix of CUTAC and ATAC-seq datasets used in this analysis. ATAC_1 is the ENCODE dataset used to call peaks and ATAC_2 is a replicate of ATAC_1.



Additional File 1: Fig. S3. Determination of critical features for deconvolution by 2for1separator. (a) Deconvolution of H1 CUT&Tag2for1 data using an uninformative dirichlet prior (1, 1, 1, 1) for fragment length distribution. Left heatmaps show signal for H3K27me3 and right heatmaps show signals for Pol25SP. Deconvolution is not dependent on prior knowledge about fragment size distributions. (b) Deconvolution without fragment length as prior leads to drop in performance. (c) Use of same covariance function for H3K27me3 and Pol25SP does not degrade performance whereas not using proximity information leads to failure of deconvolution (d).



Additional file 1: Fig. S4. Single Cell CUT&Tag2for1 and 2for1separator are consistent across replicates. (a) UMAP plot showing that single K562 cells run as independent replicates cluster together according to their deconvolved Pol2S5p signal and away from the two replicates of H1-hESCs. (b) Same as (a) showing cells from two independent replicates clustered according to their deconvolved H3K27me3 signal. (c) UMAP showing that single cells from replicate 3 cluster according to the K562 and H1-hESCs cell types based on either the Pol2S5p peaks (left) or H3K27me3 peaks (right) deconvolved from replicate 1. (d) Same as (c) but showing the cells from replicate 1 plotted on the deconvolved peaks from Replicate 3.

a K562**b** H1-hESC

Additional file 1: Fig. S5. Overlapping peaks. (a) Single antibody and 2for1 data at the overlapping peaks for K562 cells. **(b)** Same as (a), for H1-hESC cells.