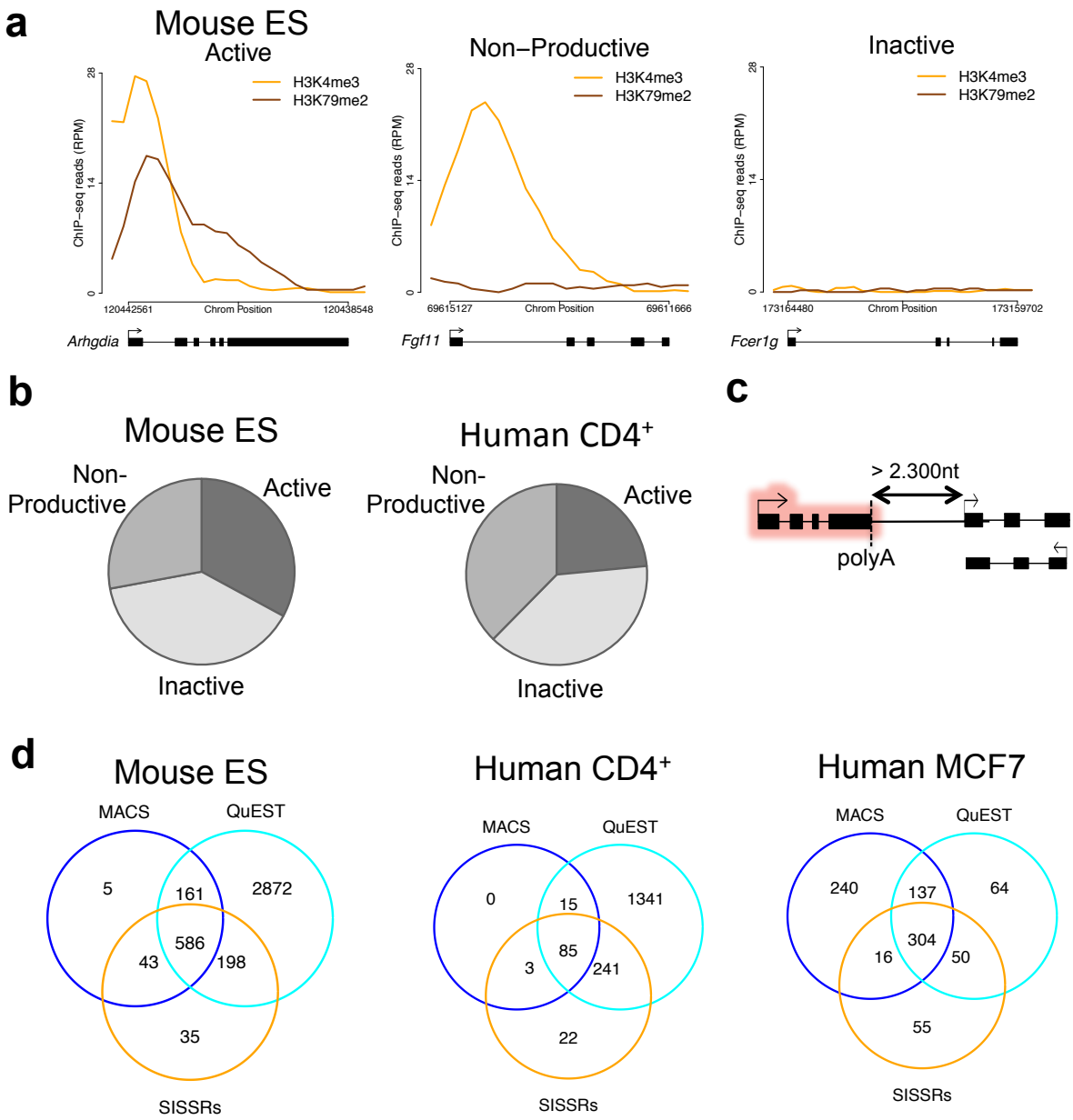
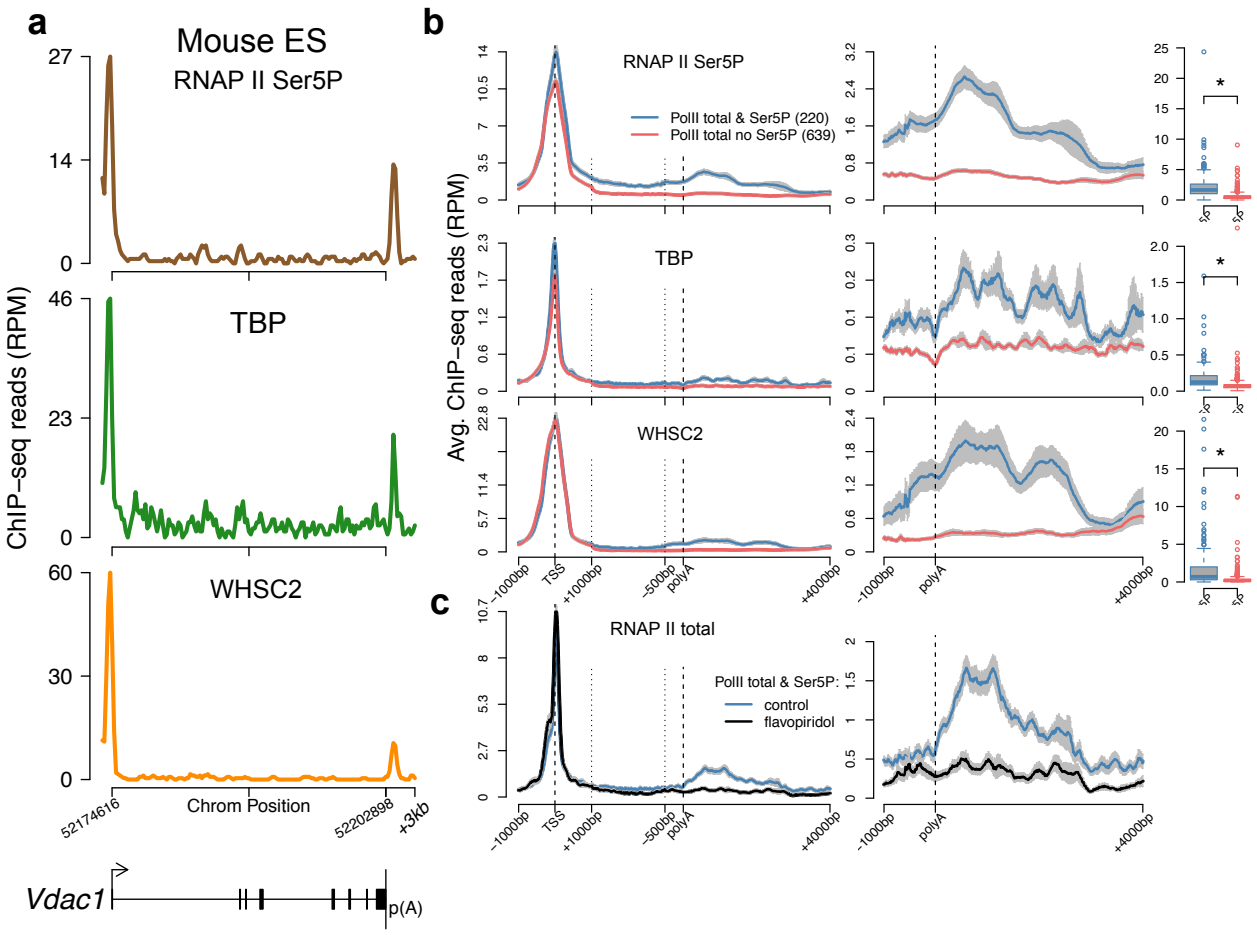
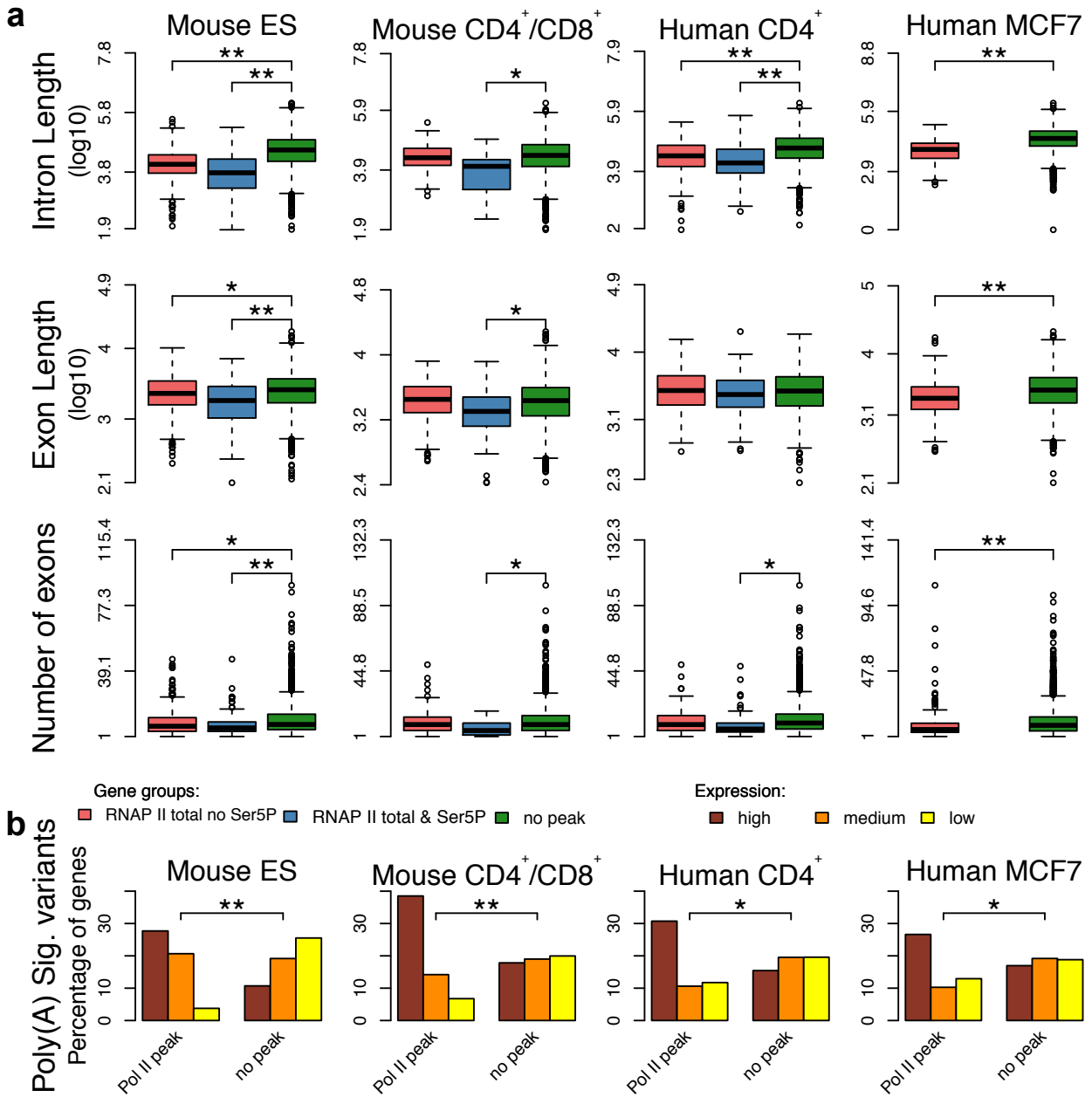


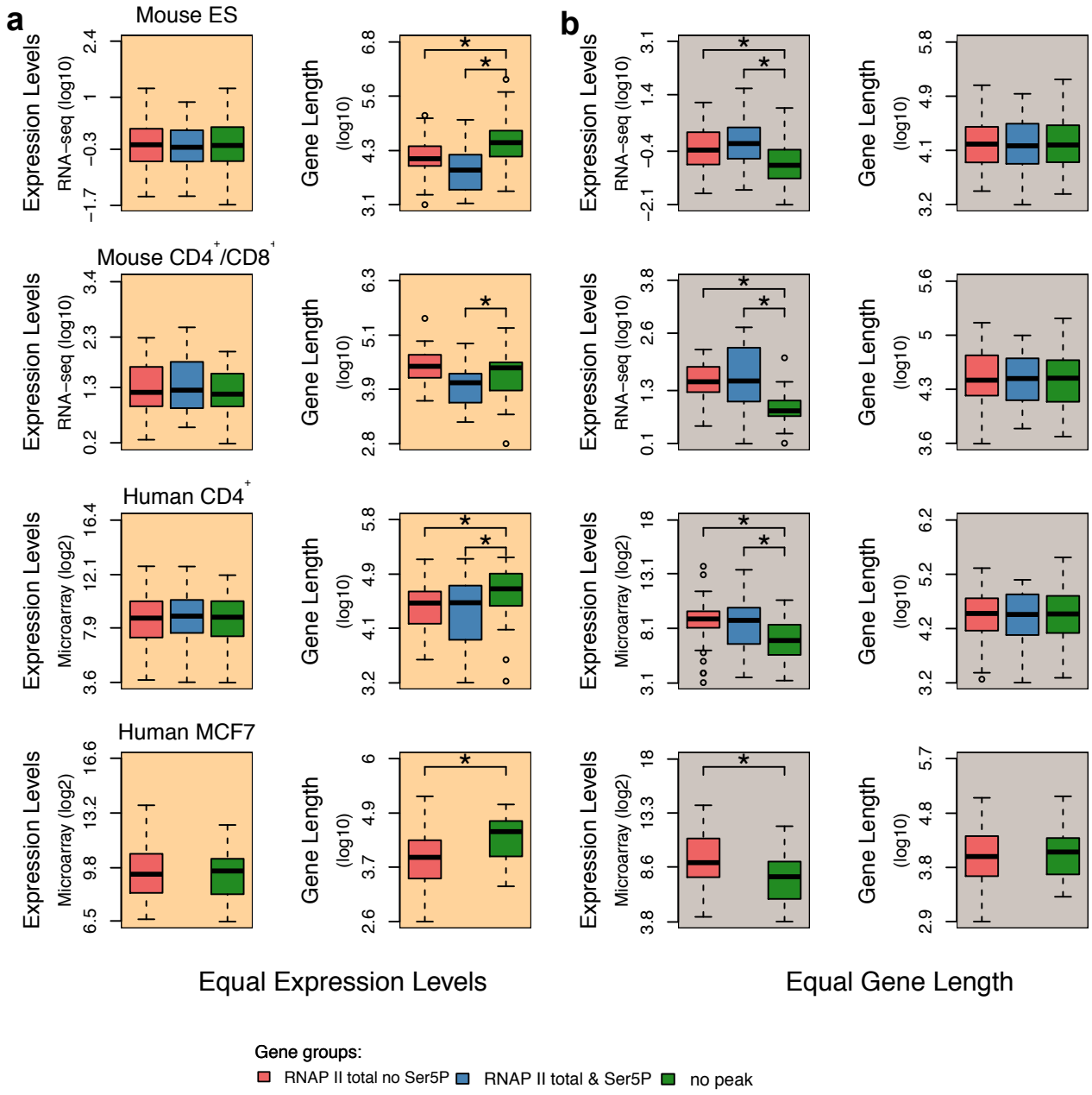
Supp. Figure 1. Global search for genes with 3' peaks of RNAPII. a) Actively transcribed genes were identified by increased 5' levels of both H3K4me3 and H3K79me2, which are nucleosome modifications associated with transcription initiation and elongation, respectively. Occupancy of H3K4me3 and H3K79me2 at a representative active gene (*Arhgdia*), non-productive gene (*Fgf11*) and inactive gene (*Fcerg1*), determined by ChIP-seq analysis of murine ES cells (Mikkelsen et al 2007; Marson et al 2008; Rahl et al 2010). Active genes are enriched in both H3K4me3- and H3K79me2-modified nucleosomes, whereas non-productive genes show evidence of initiation but not elongation and inactive genes show no evidence of initiation. **b)** Genome-wide proportion of active genes, non-productive genes and genes devoid of transcription initiation in murine ES cells and human CD4+ T cells. **c)** For each gene, a 3' flanking region was defined from the annotated 3'-end [poly(A) site] to 2.3Kb downstream. Genes were discarded if there is a neighbouring gene, in either strand, within less than 2.3Kb downstream of the poly(A) site. **d)** Each gene was independently interrogated using the following peak calling tools: MACS (Zhang et al 2008), QuEST (Valouev et al 2008) and SISSRs (Jothi et al 2008). The diagrams show the number of genes with 3' peak detected by each method in murine ES cells, and human CD4+ and MCF7 cells.



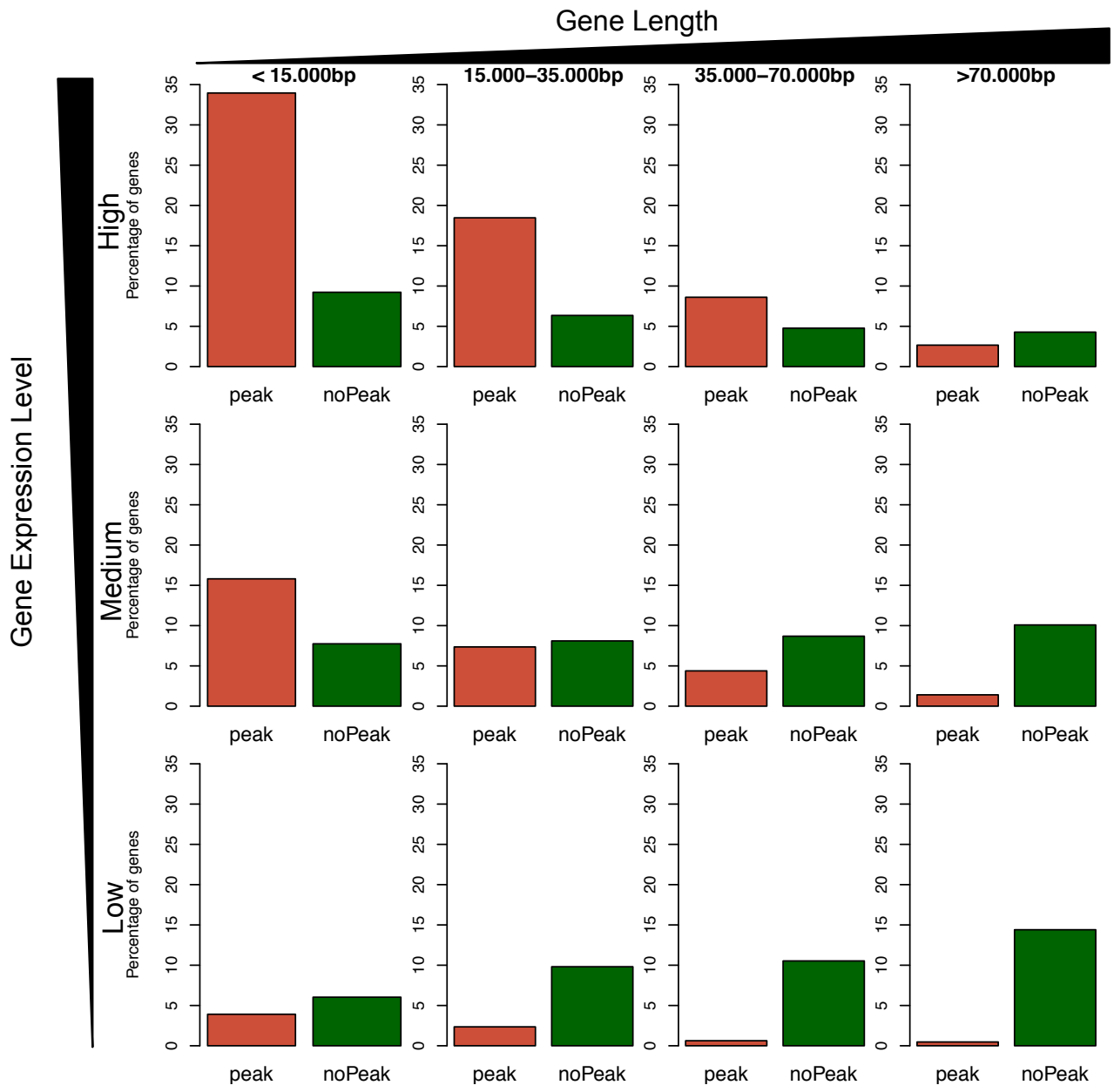
Supp. Figure 2. Genes with increased levels of Ser5P RNAPII in the termination region are likely to be in a looped conformation. a) Profiles of RNAPII (Pol II) Ser5P (brown), TBP (green), and WHSC2 (orange) in a representative gene (*Vdac1*). **b**) Average profiles of RNAPII Ser5P, TBP and WHSC2 in genes with a 3' peak of total RNAPII that either contain (blue) or do not contain (red) Ser5P enrichment. The boxplot represents the average ChIP-seq read signal for each gene in the termination region (polyA site to 2.3Kb downstream). **p-value* < 0.005 by two-sided Mann-Whitney test. **c)** Average profiles of RNAPII Ser5P in genes with a 3' peak of total and Ser5P RNAPII that were either non-treated (blue) or treated with flavopiridol (black). To show the average profiles across genes, a "metagene" profile was plotted for each group. The 5' end (1 kb upstream of the transcription start site (TSS) to 1 kb downstream) and the 3'-end (500 bp upstream of the polyA site to 4 kb downstream) are unscaled. The remainder of the gene is rescaled to 2k so that all genes appear to have the same length. Error bars (gray) represent standard error of the mean.



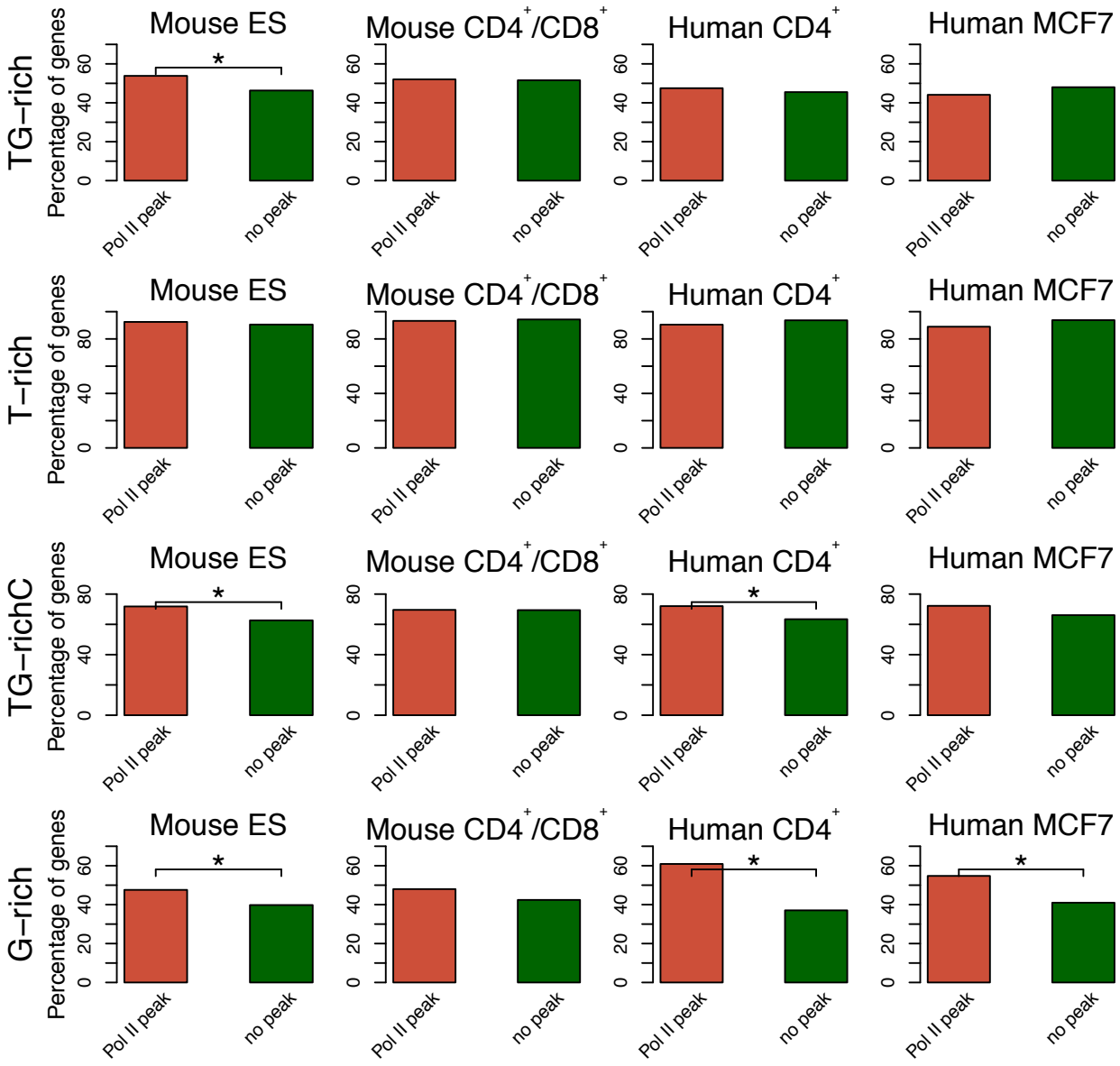
Supplementary Figure 3. Features that distinguish genes with and without 3' peaks of RNAPII. a) Comparison of intron length, exon length, and exon number in the three categories of genes: genes that are devoid of any peak of total RNAPII (Pol II) at the 3'-end (green); genes with a 3' peak of total RNAPII that is not phosphorylated on Ser5 (red); and genes with a 3' peak of total RNAPII that is phosphorylated on Ser5 (blue). ***p-value* < 0.0005 or **p-value* < 0.05 by two-sided Mann-Whitney test. **b)** Frequency of polyA signal variants (all single base substitutions variations of AATAAA) in genes that either contain a 3' peak of total RNAPII or are devoid of peak. Genes were split into three equally size groups according to expression level (high, medium and low). ***p-value* < 0.005 by Chi-squared test.



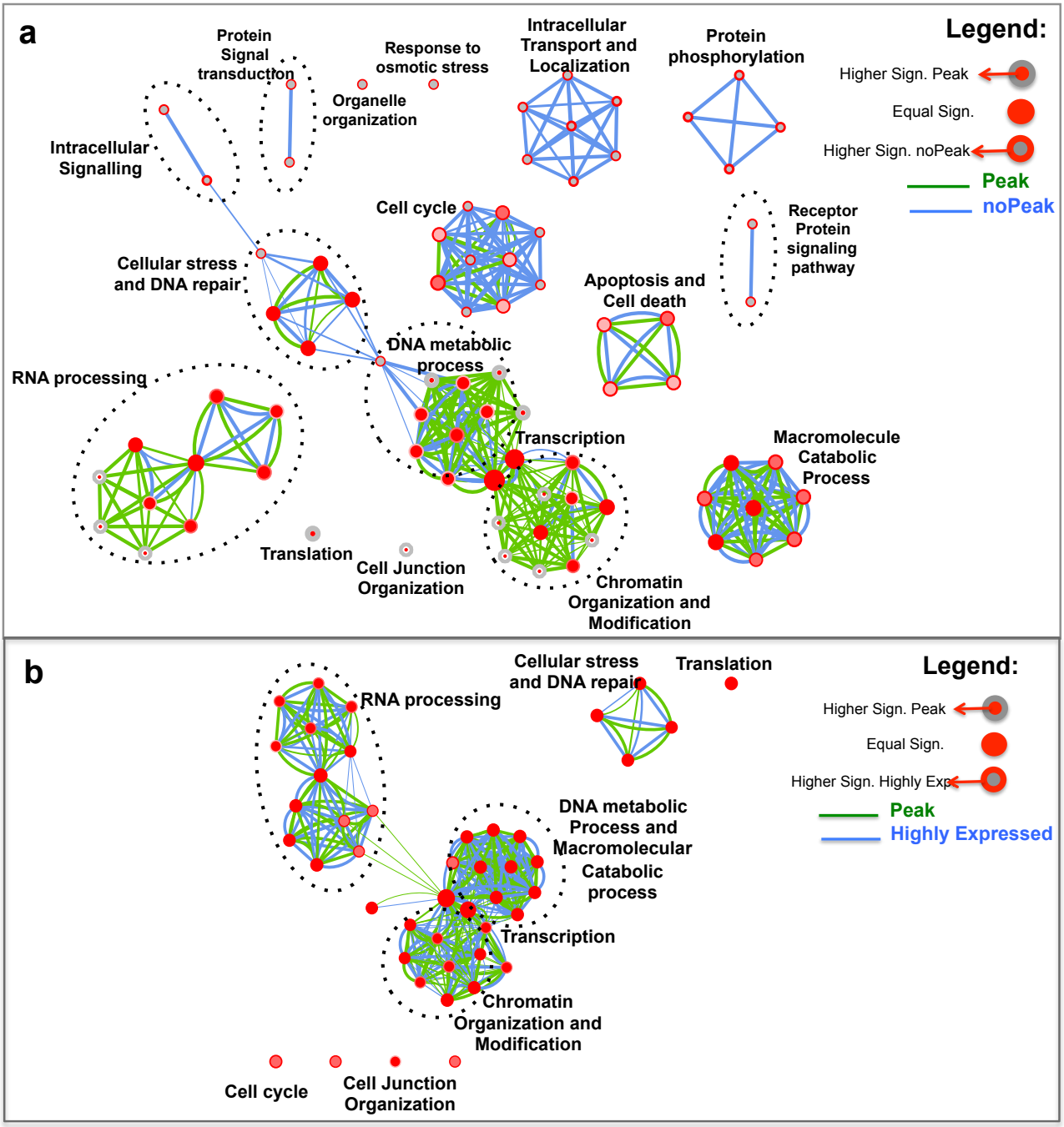
Supplementary Figure 4. Genes with 3' peaks of RNAPII are shorter and expressed at higher levels than genes with no peak. From each gene category, i.e., genes that are devoid of any peak of total RNAPII (Pol II) at the 3'-end (green); genes with a 3' peak of total RNAPII that is not phosphorylated on Ser5 (red); and genes with a 3' peak of total RNAPII that is phosphorylated on Ser5 (blue), we selected a sample with similar expression levels (**a)** and similar gene length **(b)**. For each sample, we then compared gene length and expression levels, respectively. Number of genes in each sample: 35 (mouse ES and human MCF7 cells); 20 (mouse $CD4^+/CD8^+$ and human $CD4^+$ cells). **p-value* < 0.005 by two-sided Mann-Whitney test.



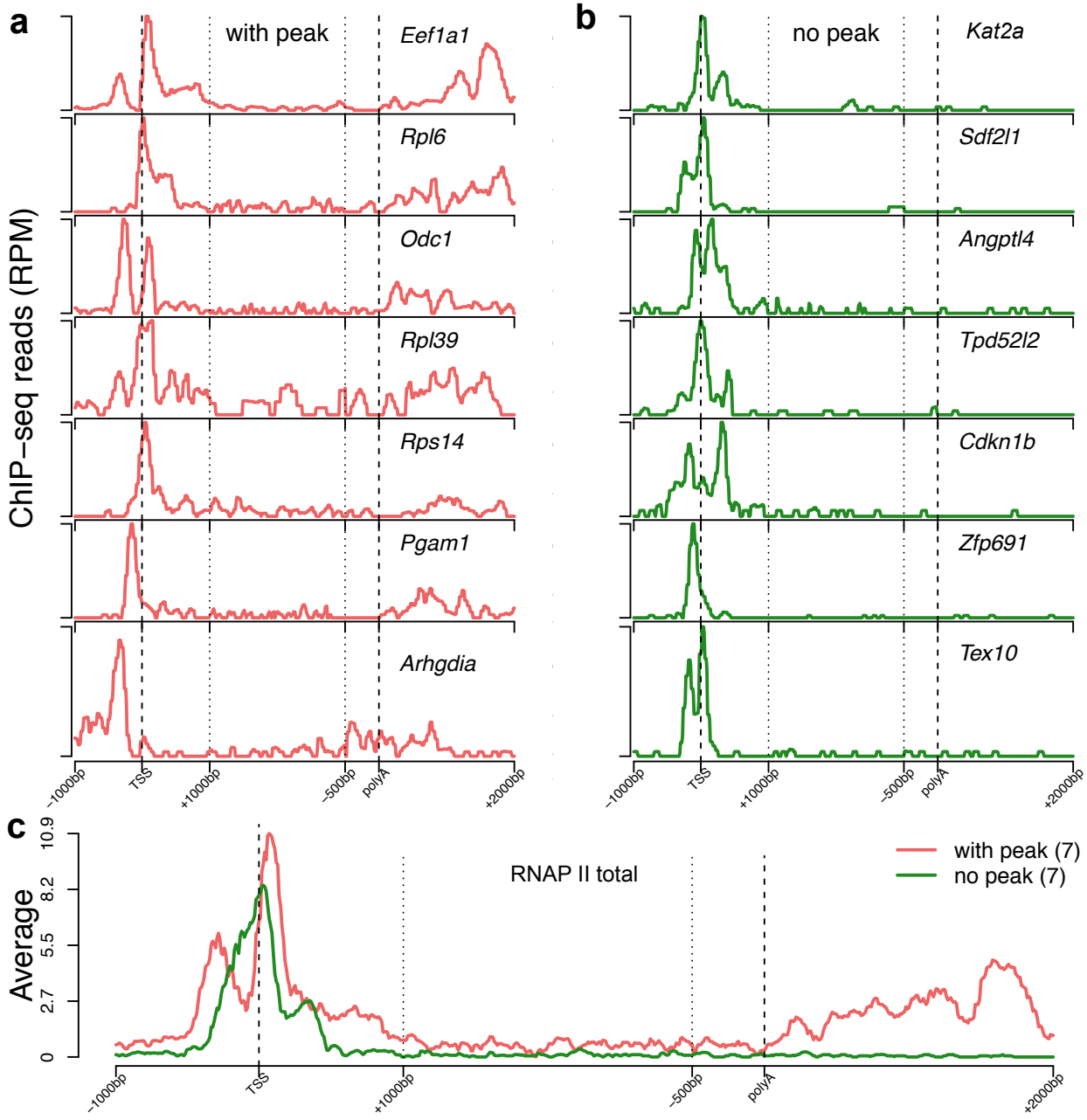
Supplementary Figure 5. The presence or absence of 3' peaks correlates with both level of gene expression and gene length. Genes that are devoid of any peak of total RNAPII at the 3'-end (green) and genes with a 3' peak of total RNAPII that is not phosphorylated on Ser5 (red) were distributed in groups according to gene expression level and gene length. Data from murine ES cells.



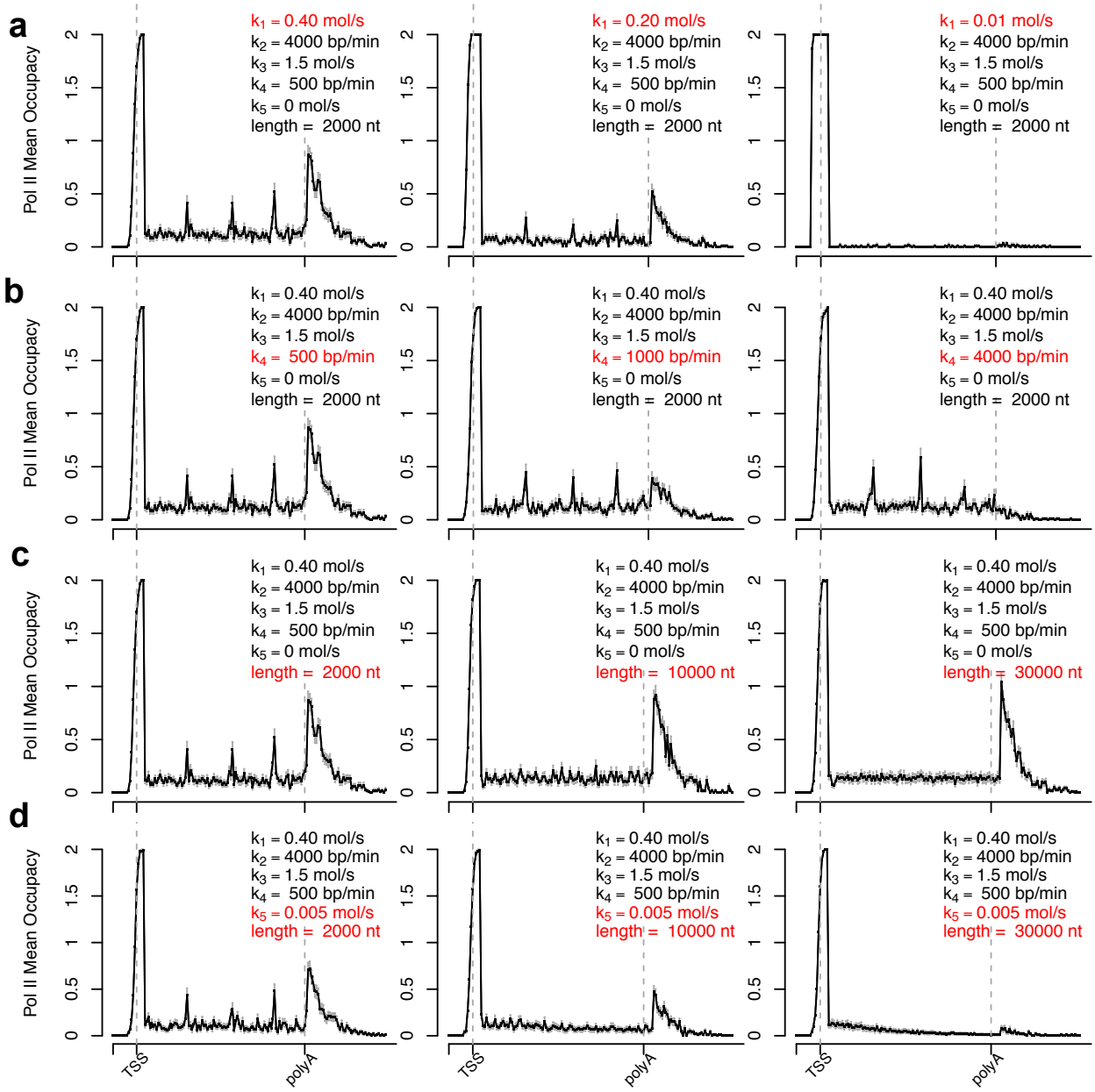
Supplementary Figure 6. Frequency of polyA site downstream elements in genes that either contain a 3' peak of total RNAPII or are devoid of peak. Several tetramers were assessed for the poly(A) site downstream signaling region (60nt after the annotated gene 3'-end) as described previously (Salisbury et al. 2006): TG-rich element (including TGTG and GTGT); T-rich element (including all single base substitutions of TTTT); alternative TG-rich element with G to C transversion (including TCTG, CTGT, TGTC, and GTCT) and G-rich element (including GGGG, GGGA, GGAG, GAGG, and AGGG). **p-value* < 0.05 by Chi-squared test.



Supplementary Figure 7. The biological functions enriched for genes with 3' peaks of RNAPII are distinct from those enriched for genes with no peak. a) Network of biological functions (gene ontology) enriched for genes that are devoid of any peak of total RNAPII at the 3'-end (blue) and genes with a 3' peak of total RNAPII that is not phosphorylated on Ser5 (green). **b)** Network of biological functions (gene ontology) enriched for genes with a 3' peak of total RNAPII that is not phosphorylated on Ser5 (green) and highly expressed genes (blue). The size of nodes is proportional to gene number (genes with 3' peak are represented as inner circles and genes with no peak or highly expressed genes are represented as outer circles); colour represents low (gray) to high (red) enrichment significance. Edges size represents number of genes common to nodes. Data from murine ES cells.



Supplementary Figure 8. Intragenic peaks of RNAPII density. RNAPII density profiles of seven representative genes that either contain (**a**, red) or are devoid (**b**, green) of RNAPII enrichment at the 3'-end. The corresponding average profiles are depicted in **c)**. The genes were rescaled to 2kb so that all genes appear to have the same length. Note that in the average profiles, the mean density signal is higher in genes with a 3' peak than in genes with no peak. This effect results from averaging multiple peaks scattered throughout each individual gene.



Supplementary Figure 9. Stochastic computational modeling of RNAPII distribution. The values of predicted RNAPII density along a hypothetical gene were estimated using the following rate constants: promoter-proximal clearance (k_1), elongation throughout the gene body (k_2), pausing at intragenic sites positioned at a regular distance of 500 nt along the gene body (k_3), elongation rate after the poly(A) site (k_4), and processivity rate (k_5). The positions of TSS and poly(A) site are indicated by dashed lines. Panels **a)** to **d)** depict a parameter sensitivity analysis: the changes of predicted RNAPII average density profile with variation of either rate of productive transcription **(a,** reflective of k_1**)**; speed of the polymerase elongating past the poly(A) site **(b,** reflective of k_4**)**; gene length **(c)**; and gene length with premature termination **(d,** reflective of k_5**).**