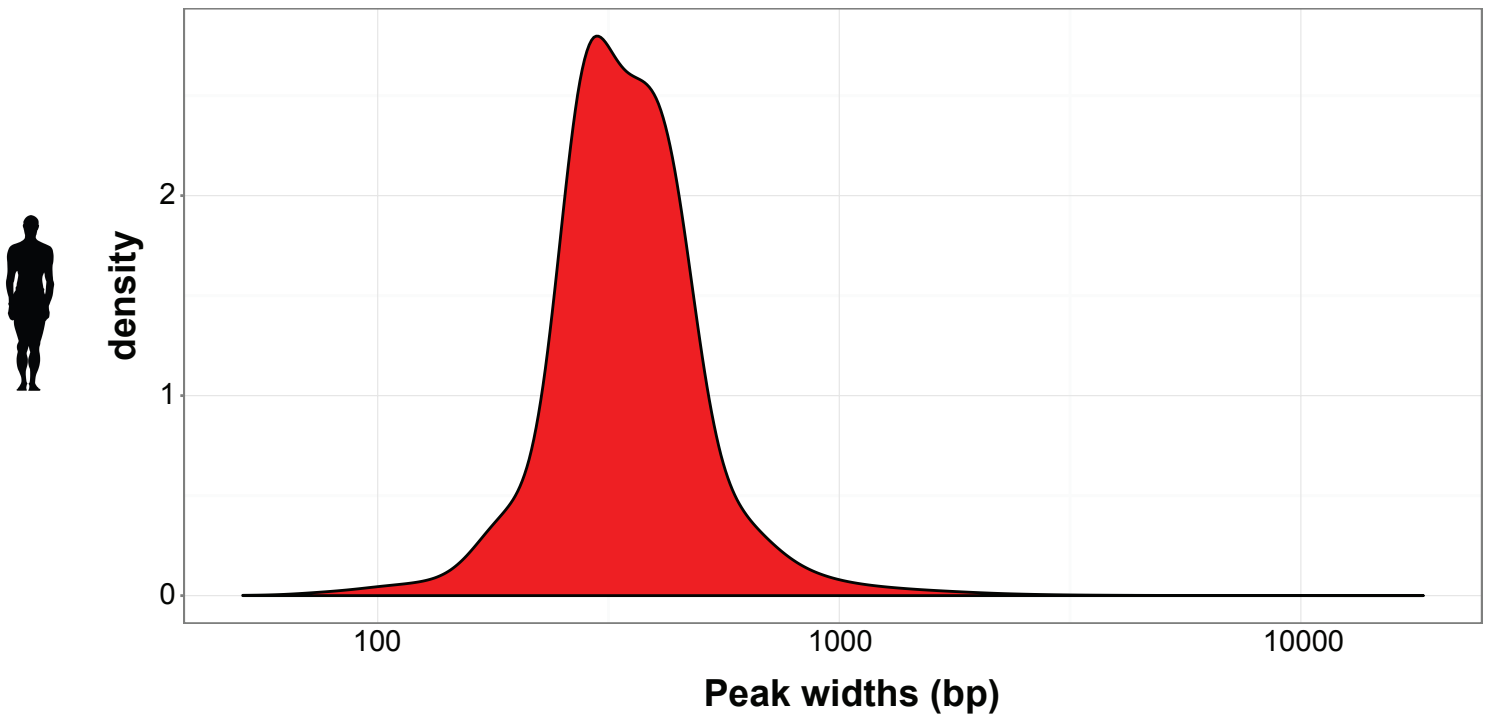
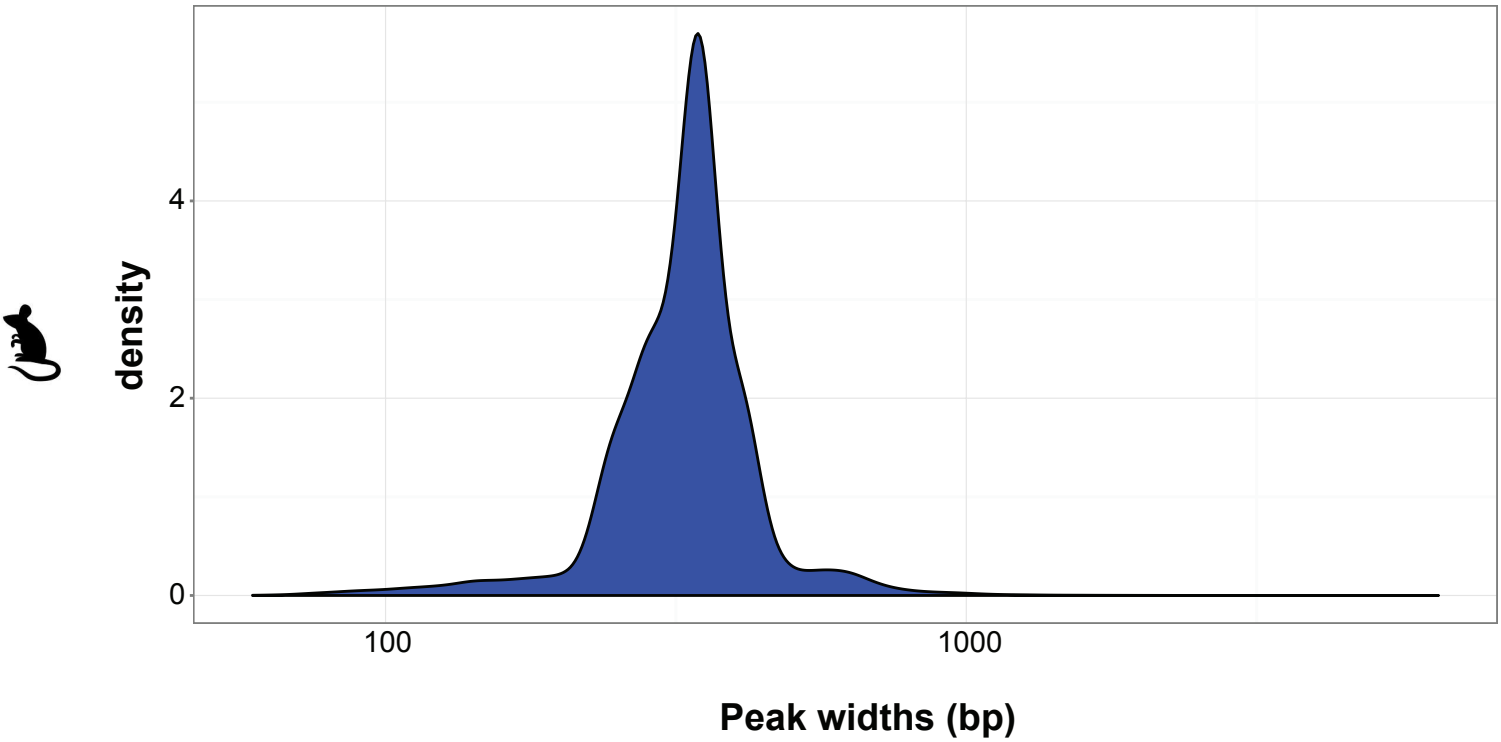
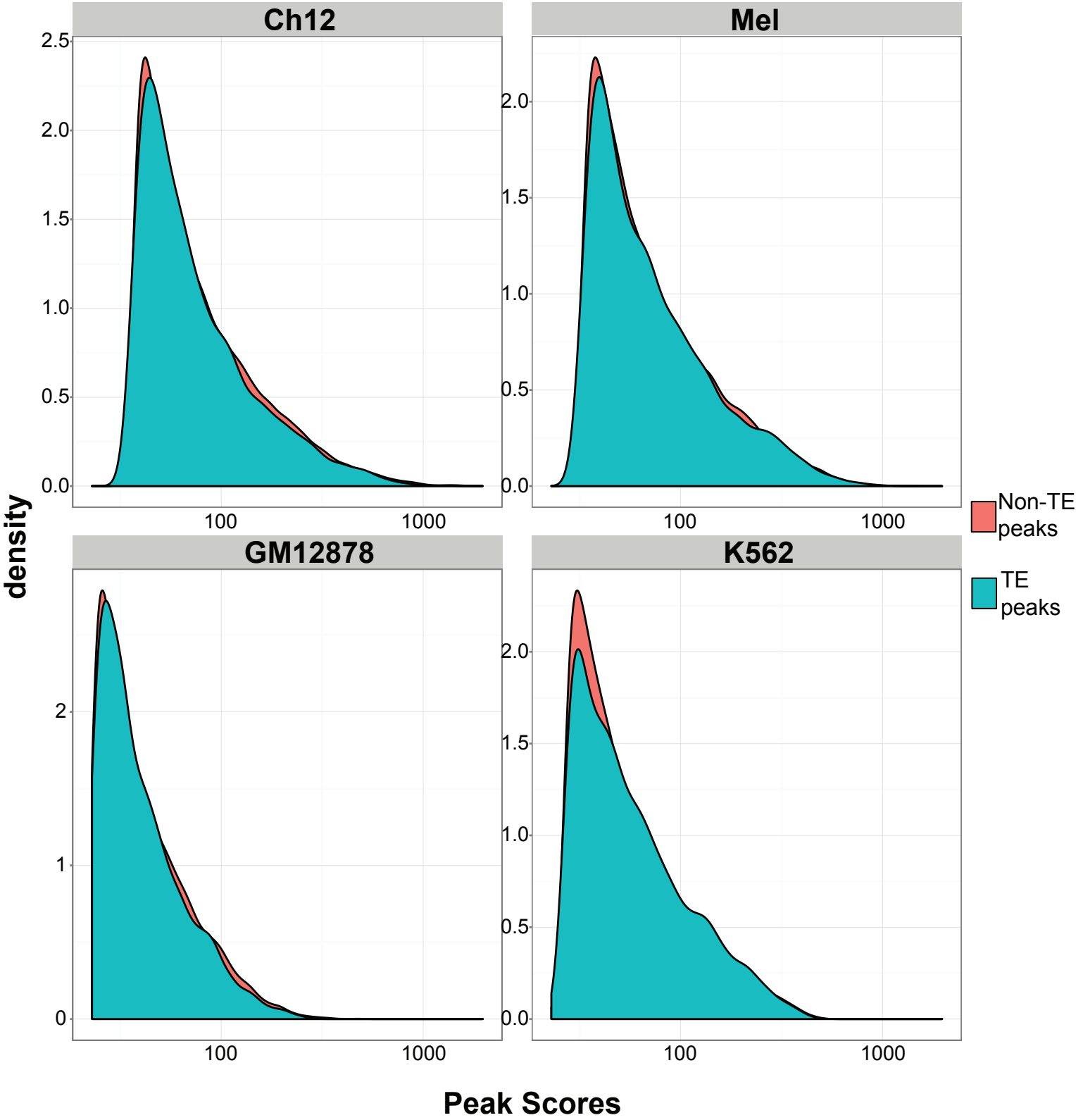
A



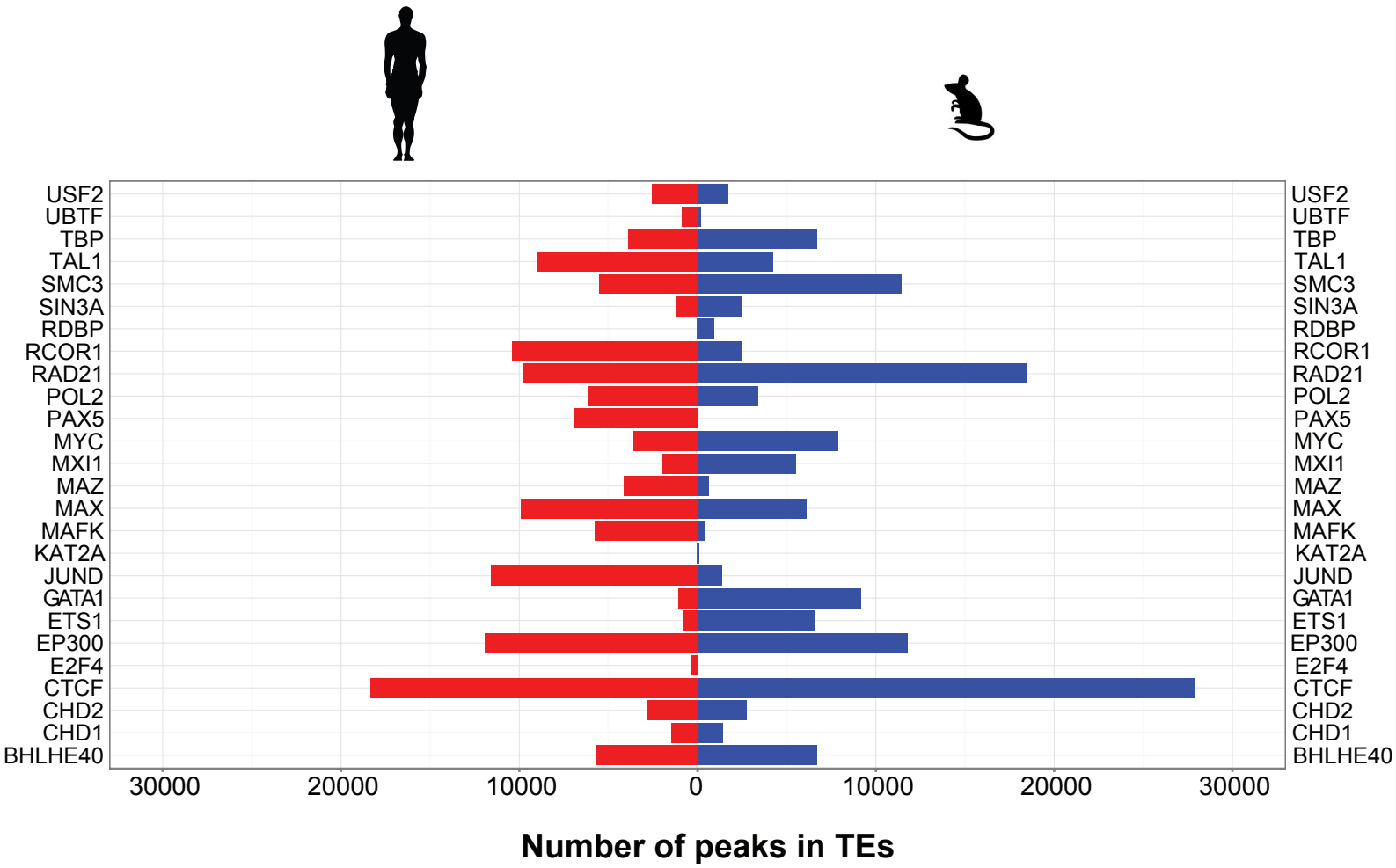
B



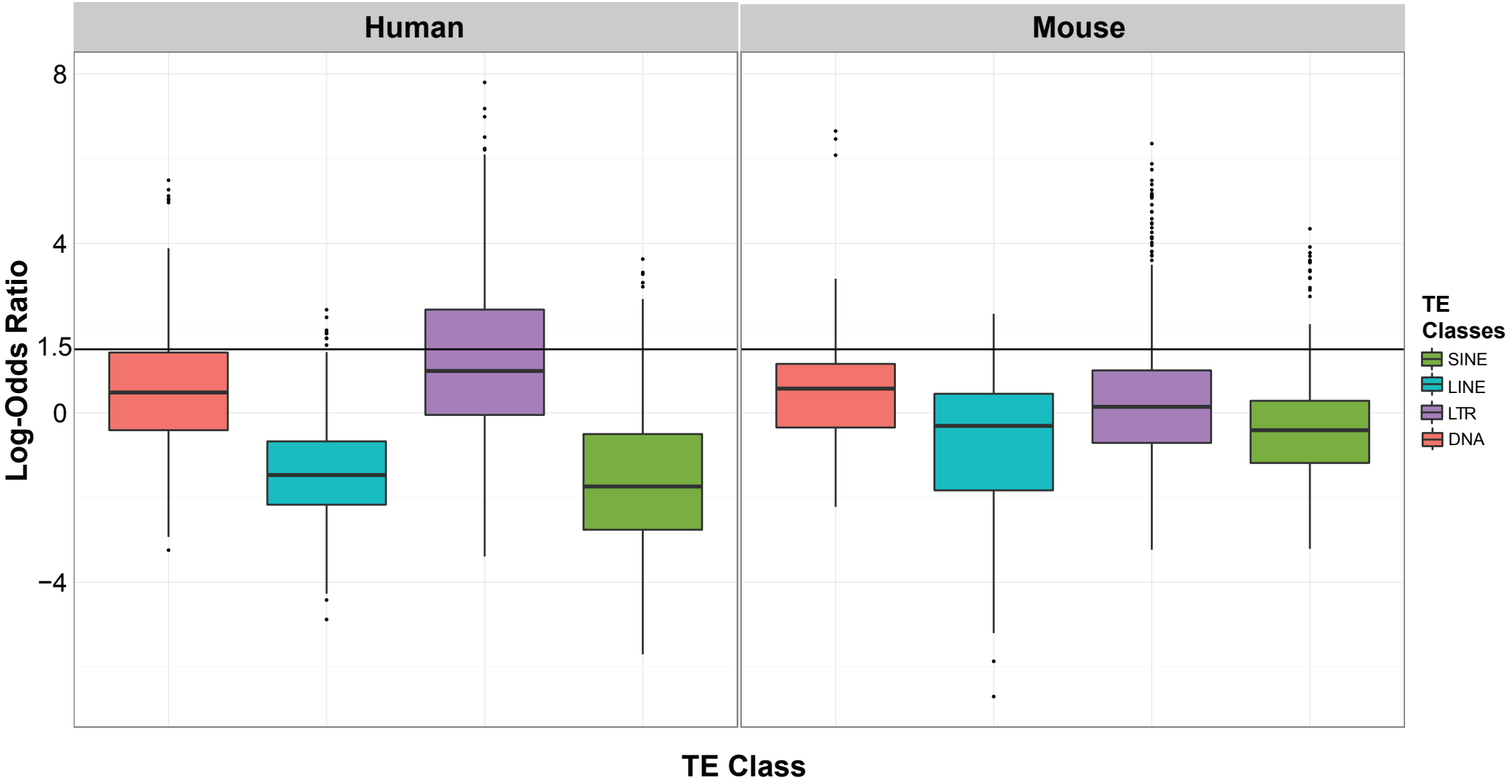
Supplementary Figure 1: Distribution of widths of ChIP-Seq TF binding peaks, in human (A) and mouse (B). The data for these plots was smoothened in ggplot, using the adjust parameter of geom_density() (human: adjust = 6, mouse: adjust=5).



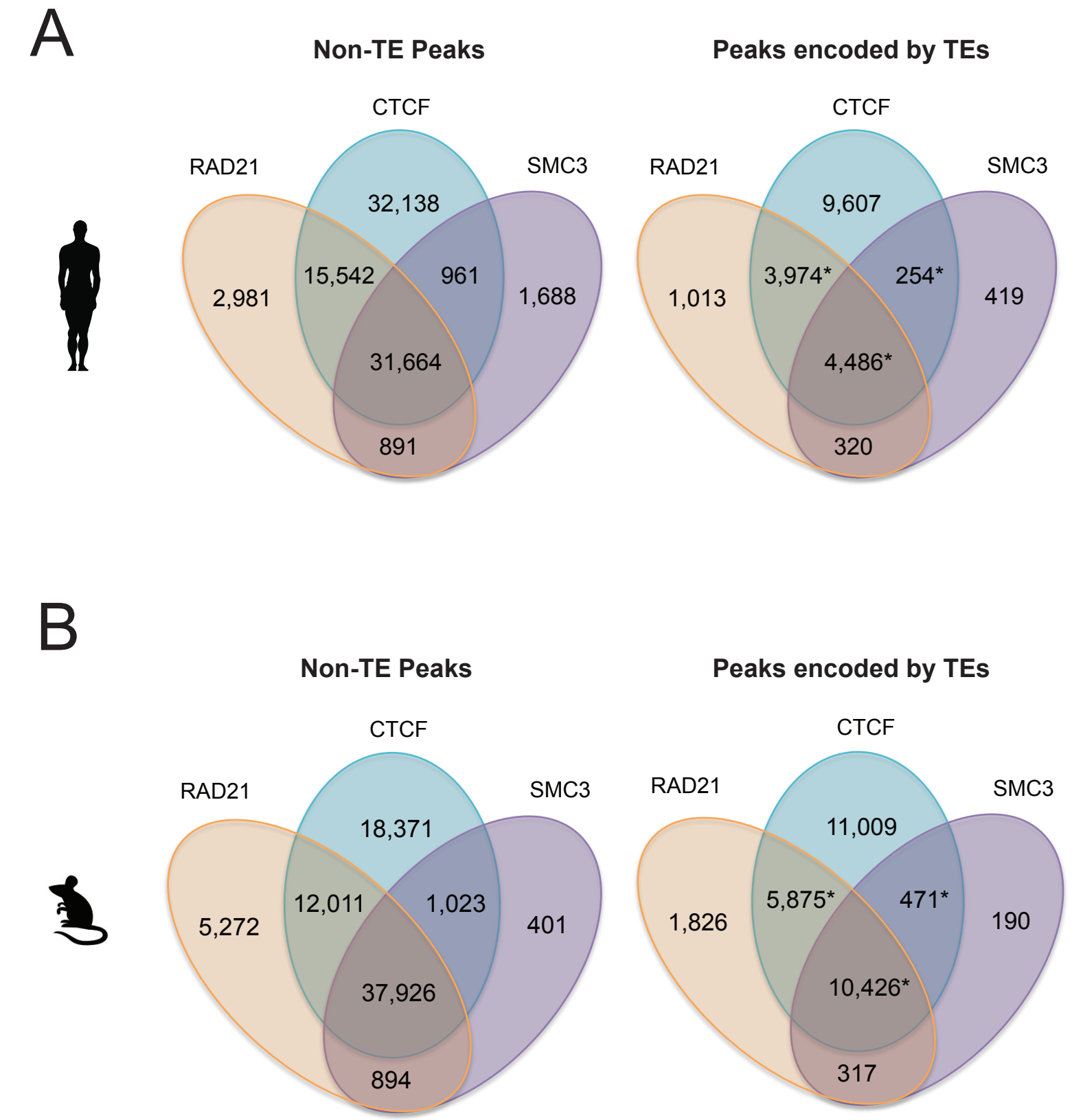
Supplementary Figure 2: Distribution of peak scores for EP300 binding peaks in lymphoblast and leukemia cell lines, in human and mouse. We chose the EP300 dataset as a representative dataset to evaluate the variability of peak scores in peaks in TE, and non-TE regions. We find that overall, the distribution of peak scores in TE and non-TE regions are highly similar.



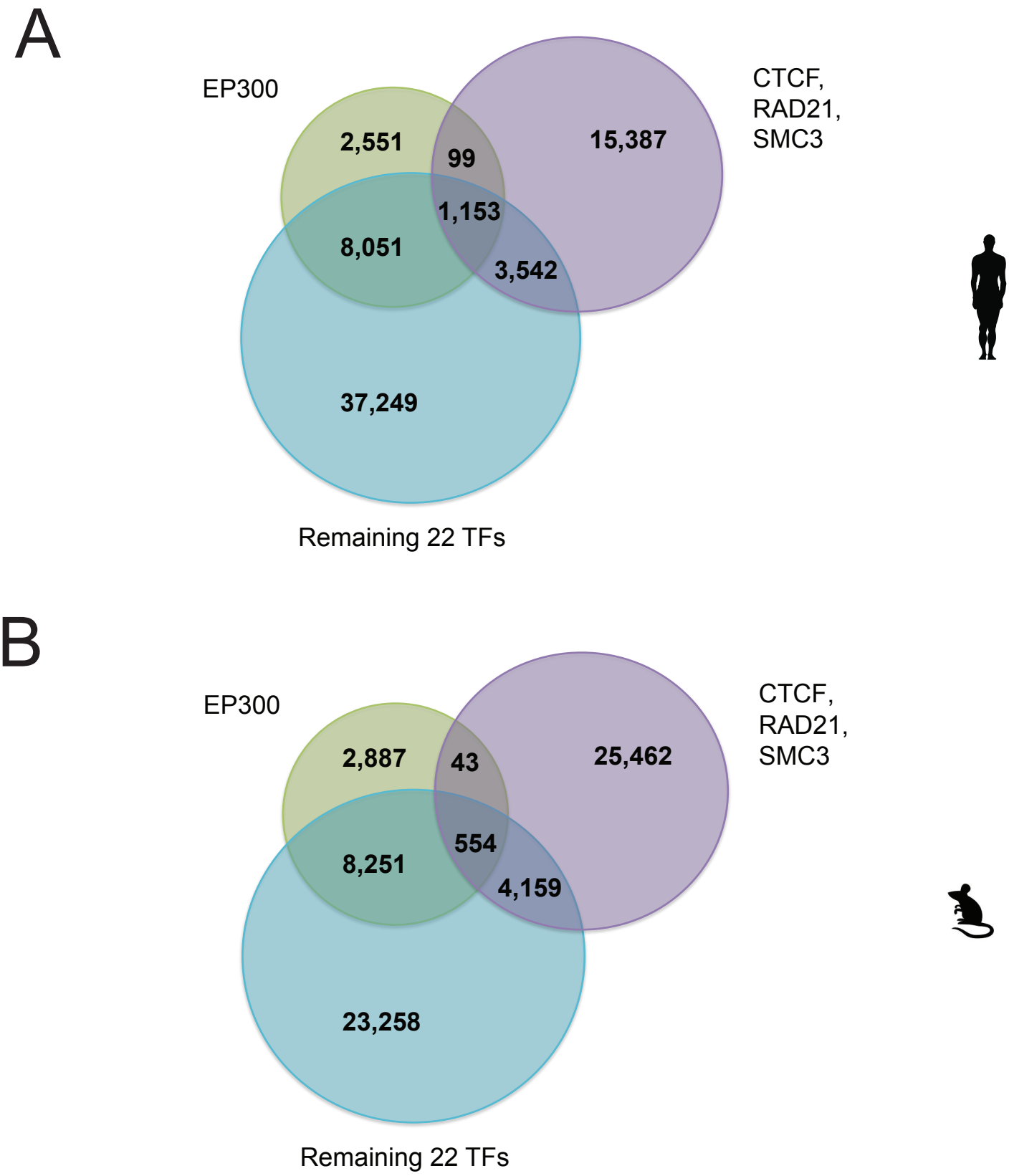
Supplementary Figure 3: Number of TE-derived peaks for each TF, in human (red) and mouse (blue). These numbers show TF-specific and species-specific differences, suggesting a role for TEs in mediating TF-specific and species-specific expansion of TF binding events. *(See Methods for definition of TE-derived peaks)*.



Supplementary Figure 4: Distribution of Log-Odds Ratio scores (i.e., enrichment scores, on the y-axis) for the various TF-TE relationships identified, categorized by TE class (on the x-axis). In both human and mouse, it is evident that the distribution of log-odds ratio scores shifted towards the positive for DNA and LTR elements, while they were negative for the LINE and SINE elements. Additionally there were more positive outliers in the LTR class compared with the other classes, in both human and mouse.

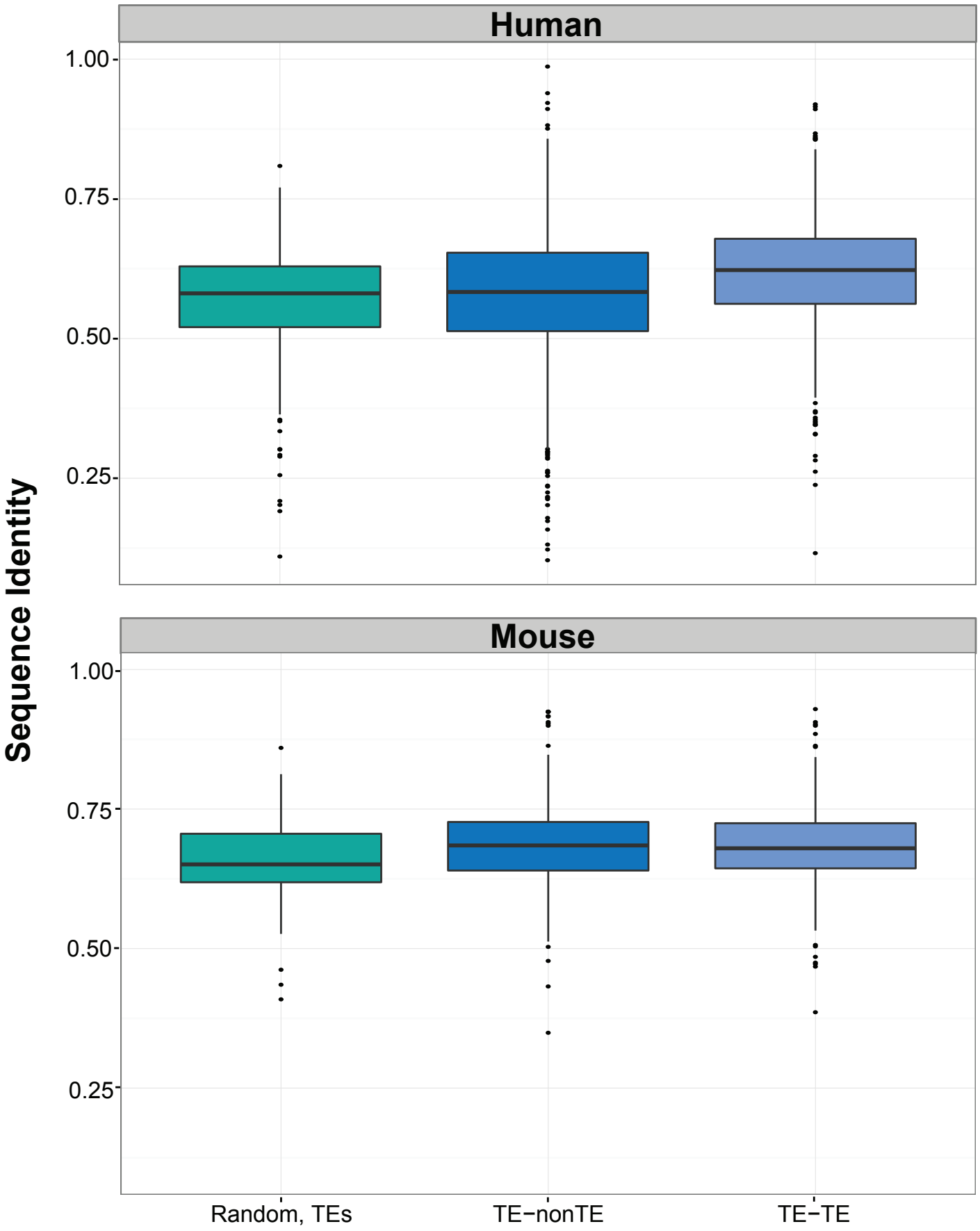


Supplementary Figure 5: Number of co-localized CTCF, and Cohesin complex factors (RAD21, and SMC3) binding peaks in TEs (right) and in non-TE genomic sequences (left). We observe that most of CTCF binding co-localizes with the binding of Cohesin complex proteins, as shown earlier (Nitzsche et al., 2011, Wendt et al., 2008). Additionally, we observe that the co-localized binding of CTCF and Cohesin also occurs in TEs, which is statistically significant (hypergeometric p-value < 10e-3).

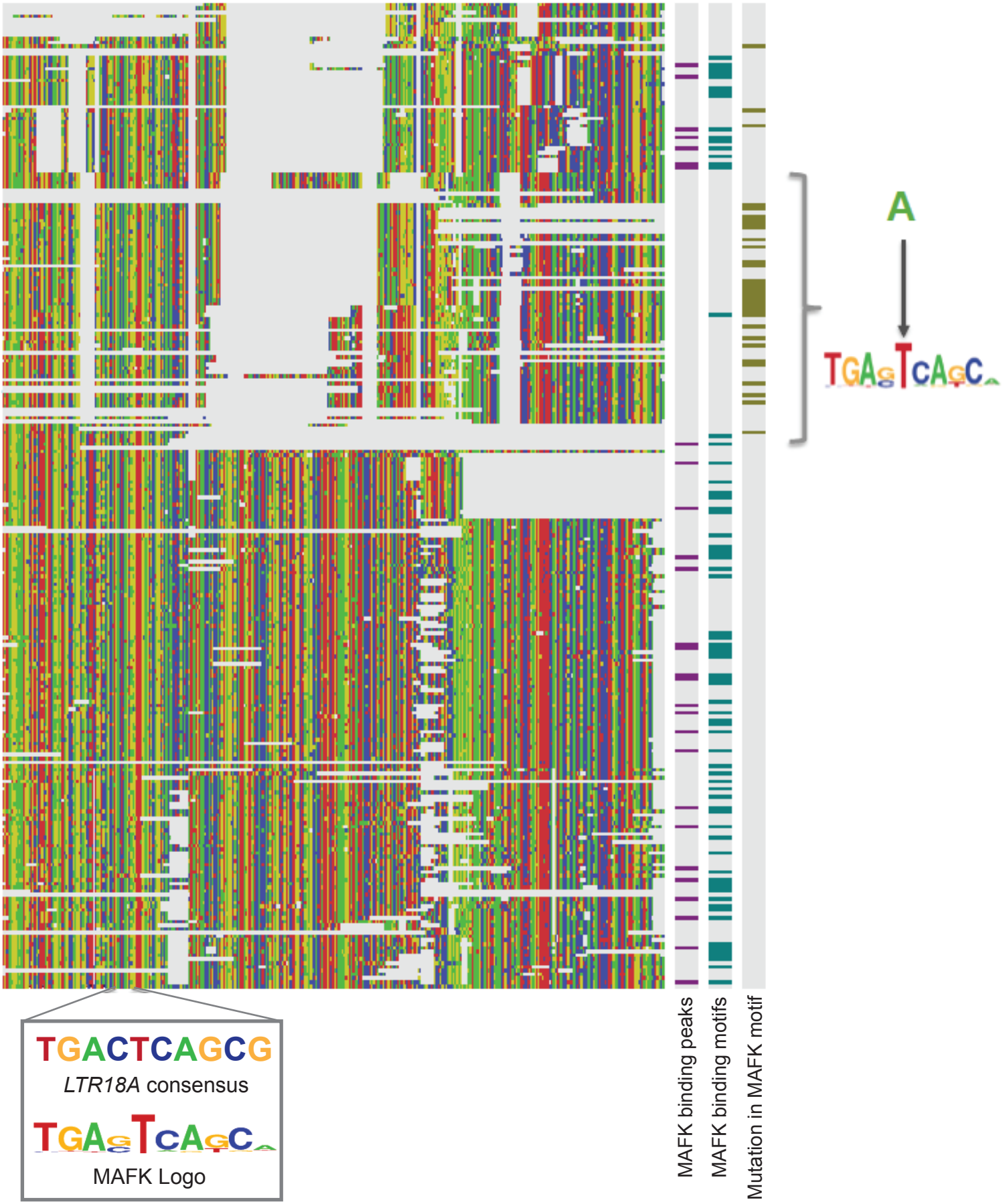


Supplementary Figure 6: EP300 and CTCF-associated factors have distinct connections with TE-derived TF binding peaks, in human (A) and mouse (B). In the Venn diagram, each circle represents the number of TE-derived peaks for one or more TFs. Most TE-derived EP300 binding peaks interact with other TF binding peaks. These interactions outnumber the interactions of CTCF-RAD21-SMC3 binding peaks with EP300 binding peaks and other TF binding peaks.

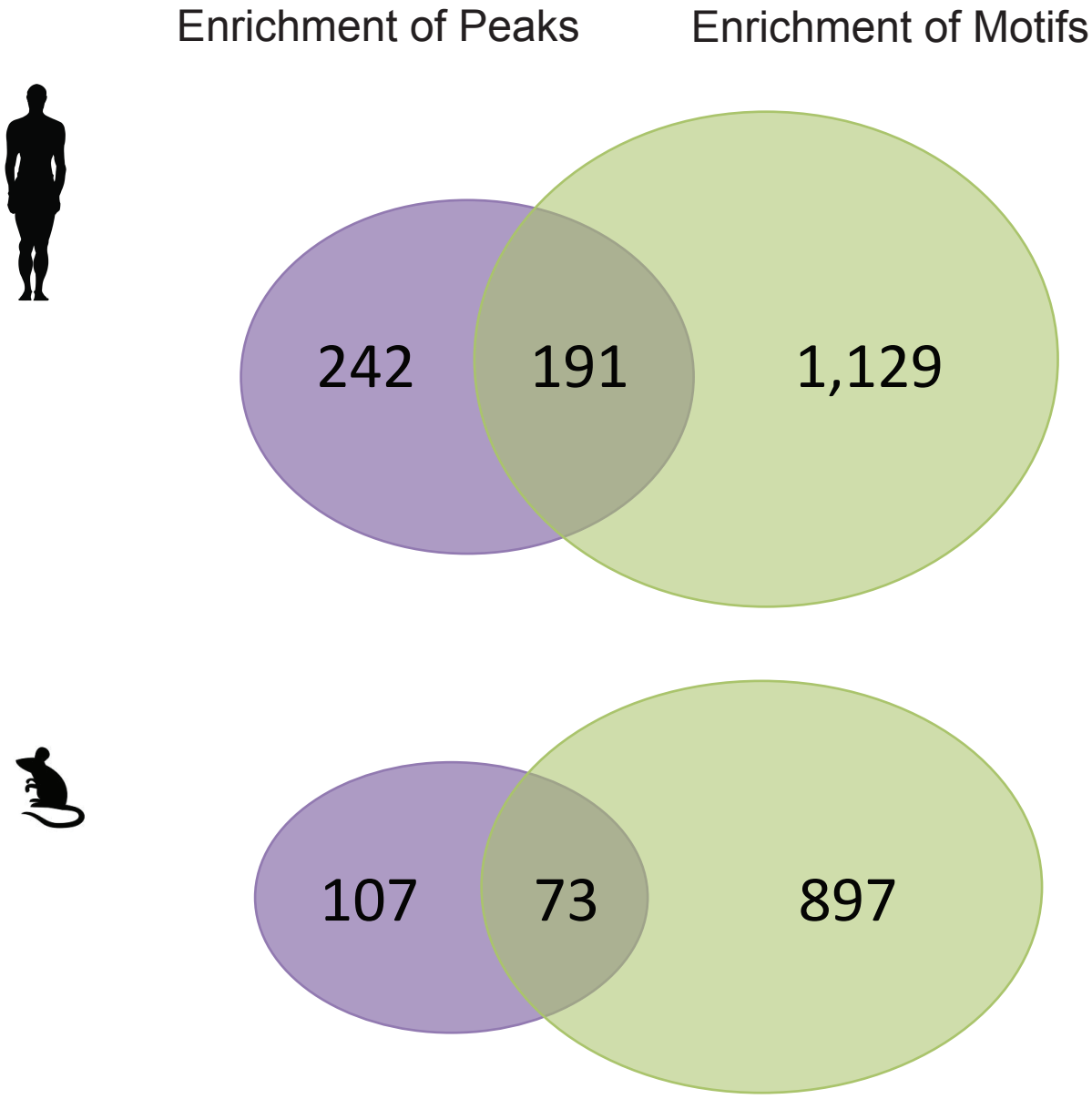
Supplementary Figure 7: Correlation between enrichment scores of TF-TE associations found by reads (x-axis), and peaks (y-axis), for human (left panel) and mouse (right panel). Each point on this plot represents one TF-TE relationship. Overall, the enrichment scores are positively correlated, as noted by the Pearson's correlation coefficient (r) in the plot.



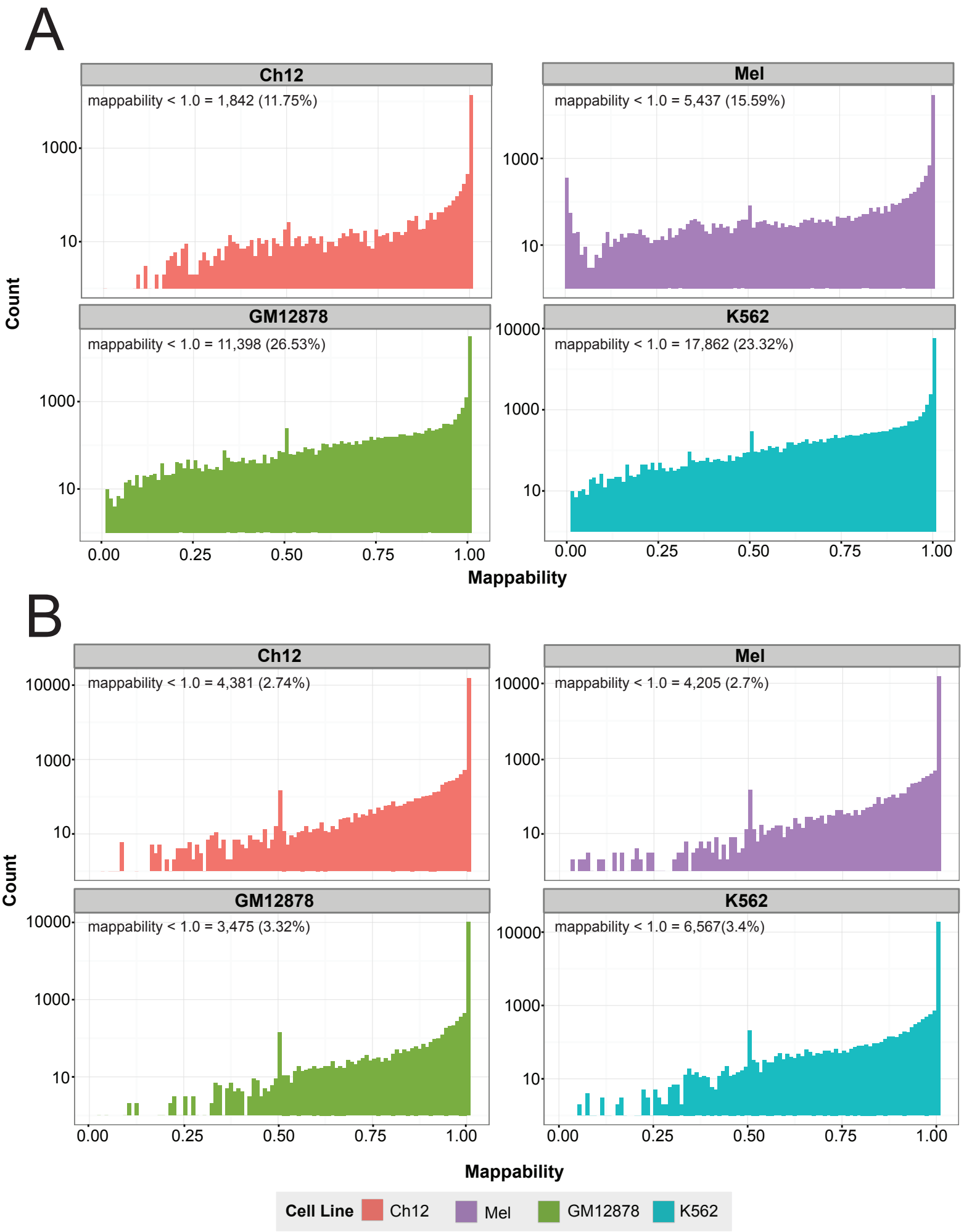
Supplementary Figure 8: Comparison of the sequence identity of pairs of sequences from human (upper panel) and mouse (lower panel), using the reciprocal-best chain files between hg19 and mm9. TEs that contributed conserved TF binding peaks and lacked the same TE subfamily in the orthologous region, showed high sequence identity with their counterparts in human and mouse ("TE-nonTE"), like conserved TE-derived TF binding peaks ("TE-TE"). For a comparison, we included random TEs ("Random, TEs").



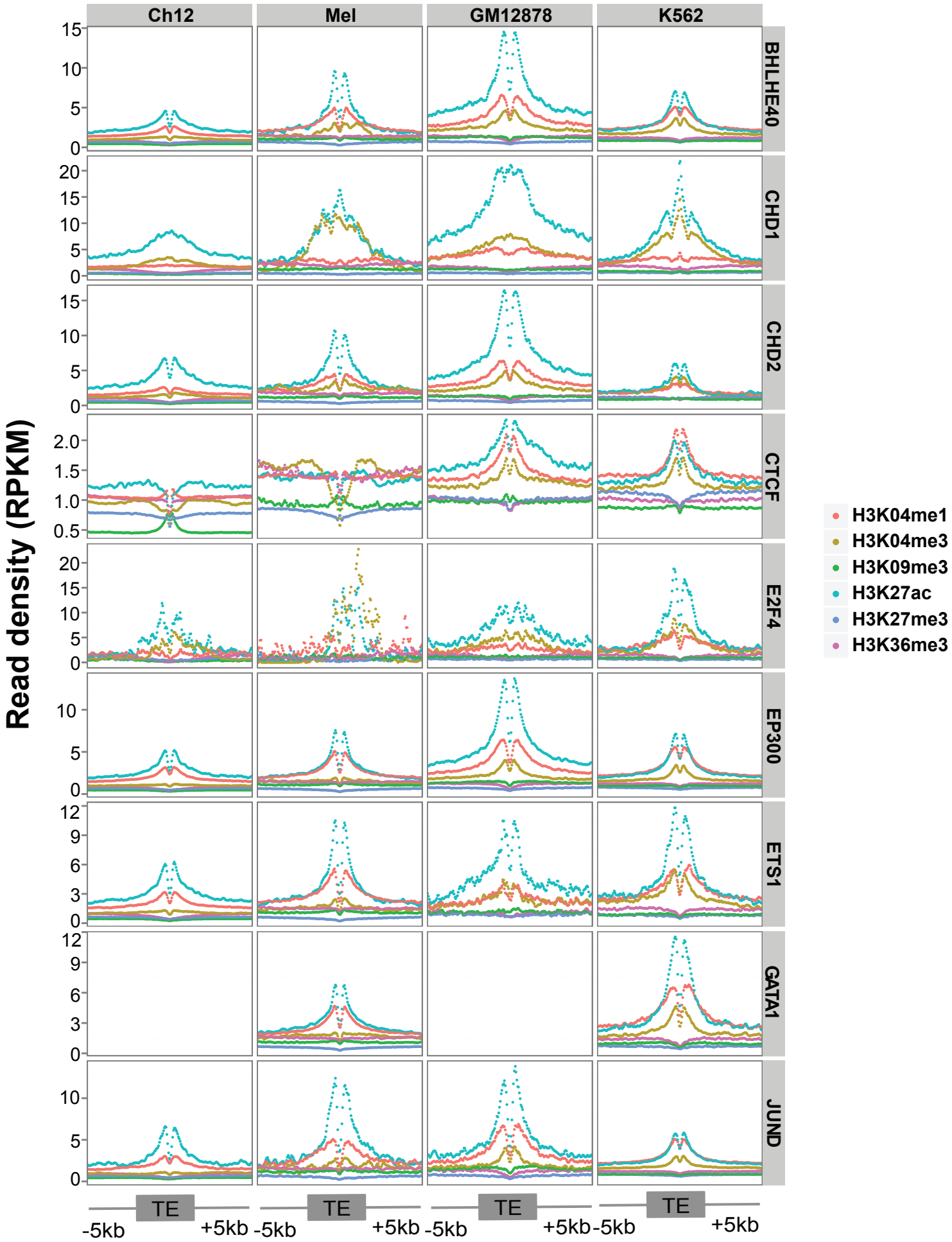
Supplementary Figure 9: Binding site analysis of *LTR18A* fragments in human that enrich for MAFK binding peaks and motifs. *LTR18A* has 259 genomic copies, 72 of which contain MAFK motifs and 28 are bound by MAFK (26 out of the 28 contain motifs). The multiple sequence alignment of the genomic copies of *LTR18A* to the consensus sequence (bottom row in the multiple sequence alignment) shows a cluster of sequences in the middle that lacks MAFK binding peaks and motifs. This cluster of *LTR18A* fragments lack the binding site due to a mutation (T -> A) at position 5 of the MAFK motif.



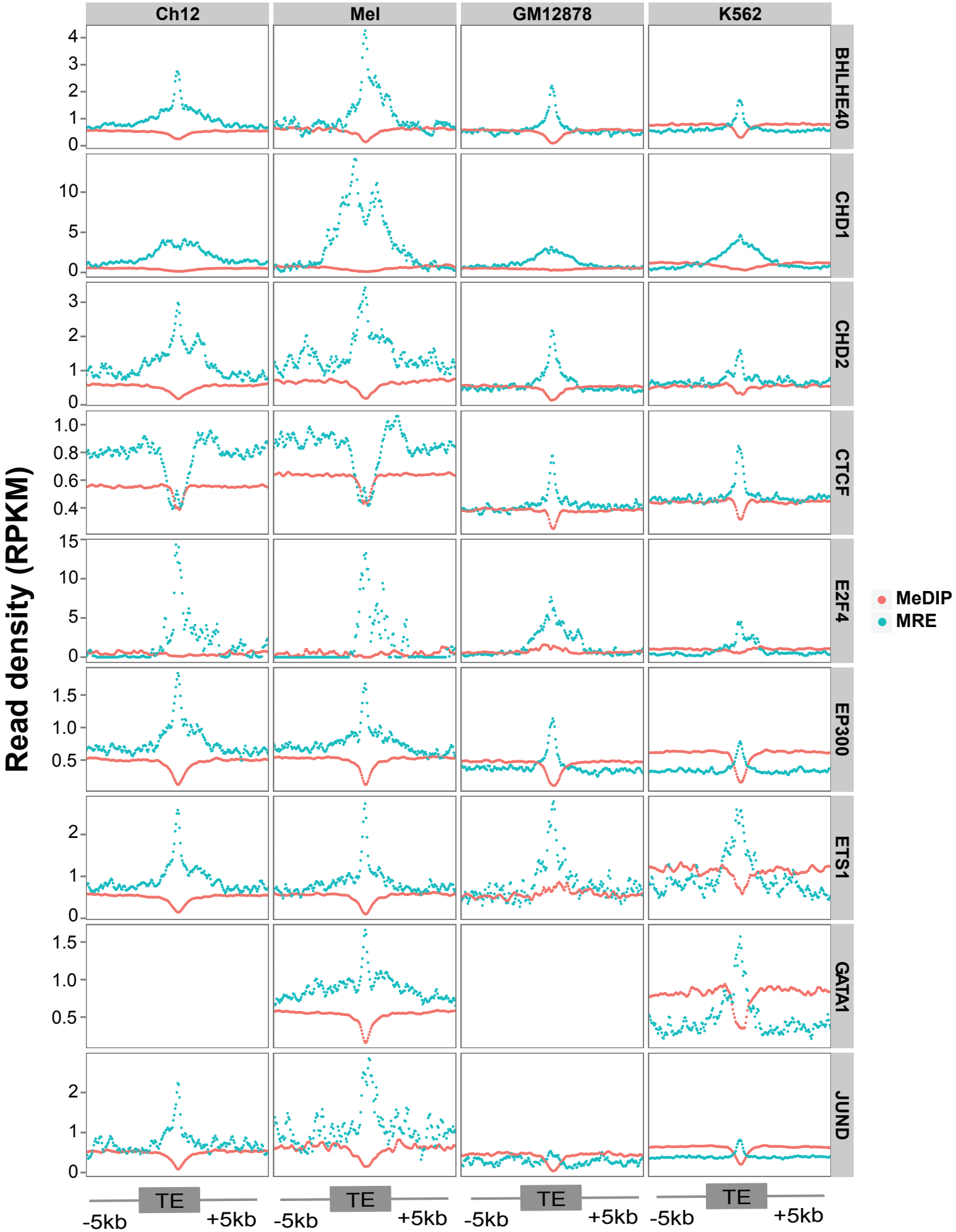
Supplementary Figure 10: Distribution of TE subfamilies that are enriched for peaks (purple), and motifs (green) in human (upper panel), and mouse (lower panel). Many TE subfamilies enrich for TF binding motifs and no TF binding peaks in both human and mouse. This can be explained by various other factors controlling binding, including chromatin. The relatively fewer TE subfamilies that enrich for TF binding peaks but not TF motifs, may represent non-sequence-specific binding, or binding motifs not captured by the motifs we predicted. (*See Methods for definition of TE enrichment of peaks, and* motifs).

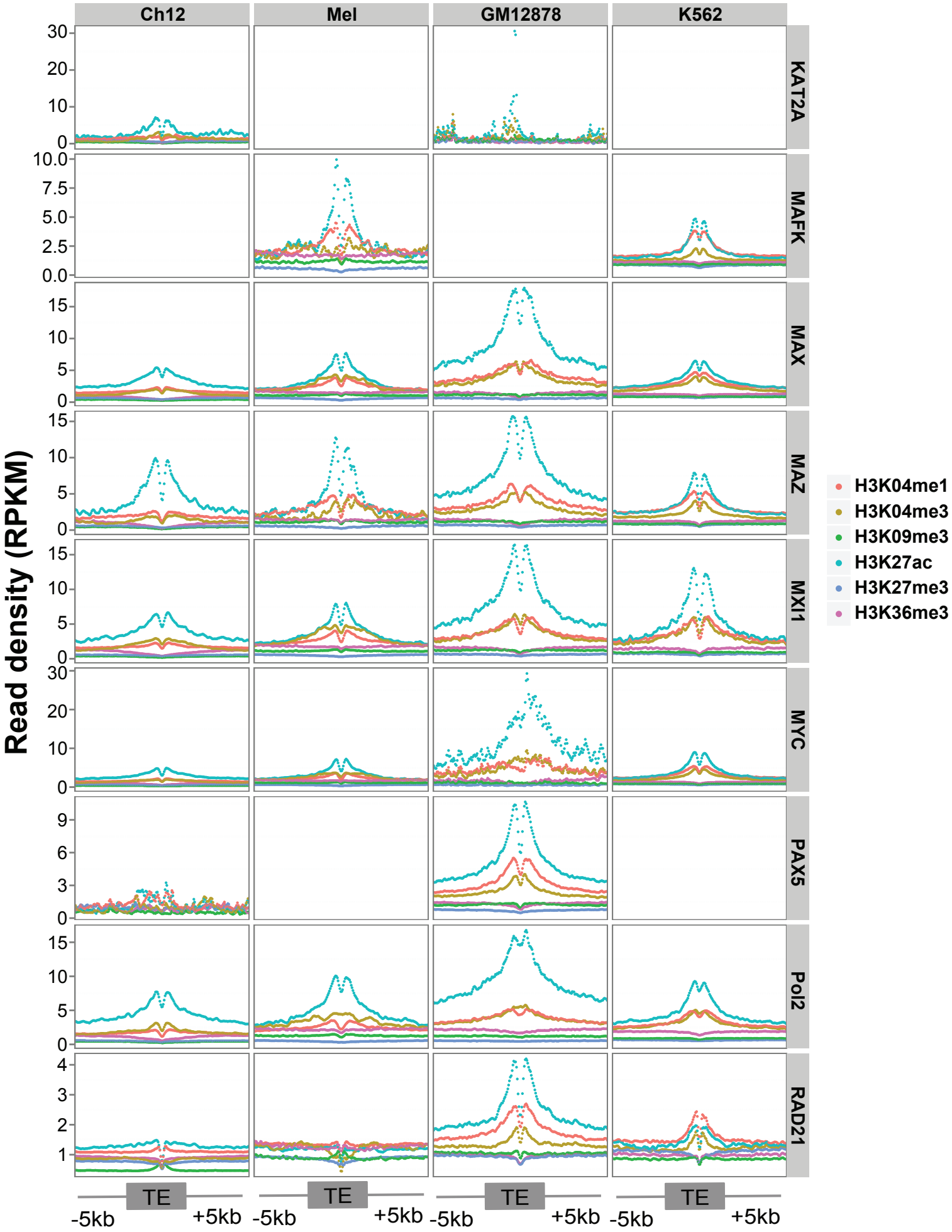


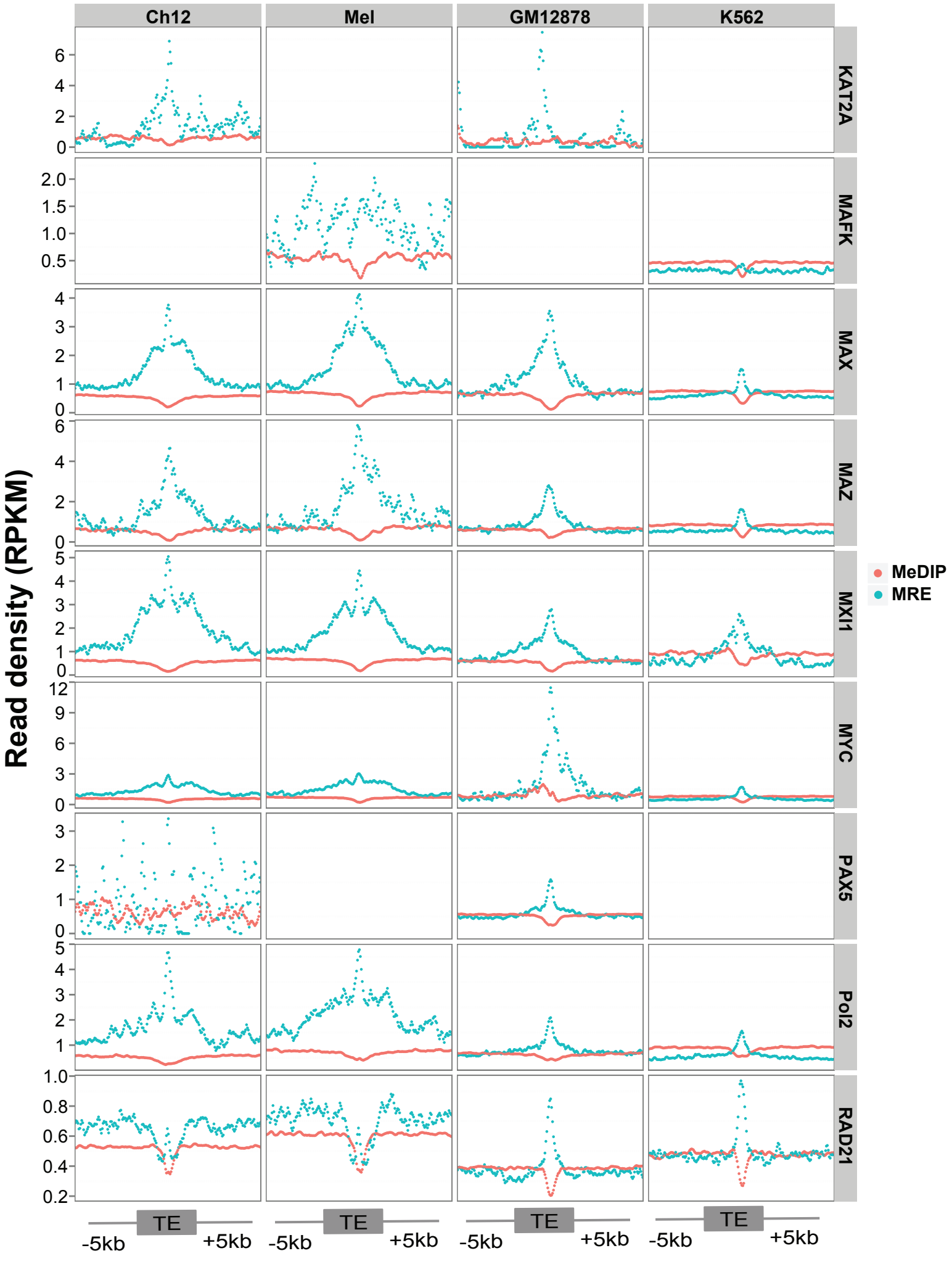
Supplementary Figure 11: Distribution of mappability scores of (A) TEs with TF binding peaks, and (B) TF binding peaks without TEs. Panels (A) and (B) here, correspond to the regions in Figure 6A and 6B. The percentage (%) of the regions in each sub-panel (corresponding to cell lines) that have a mappability score below 1.0 are listed in each corresponding sub-panel.

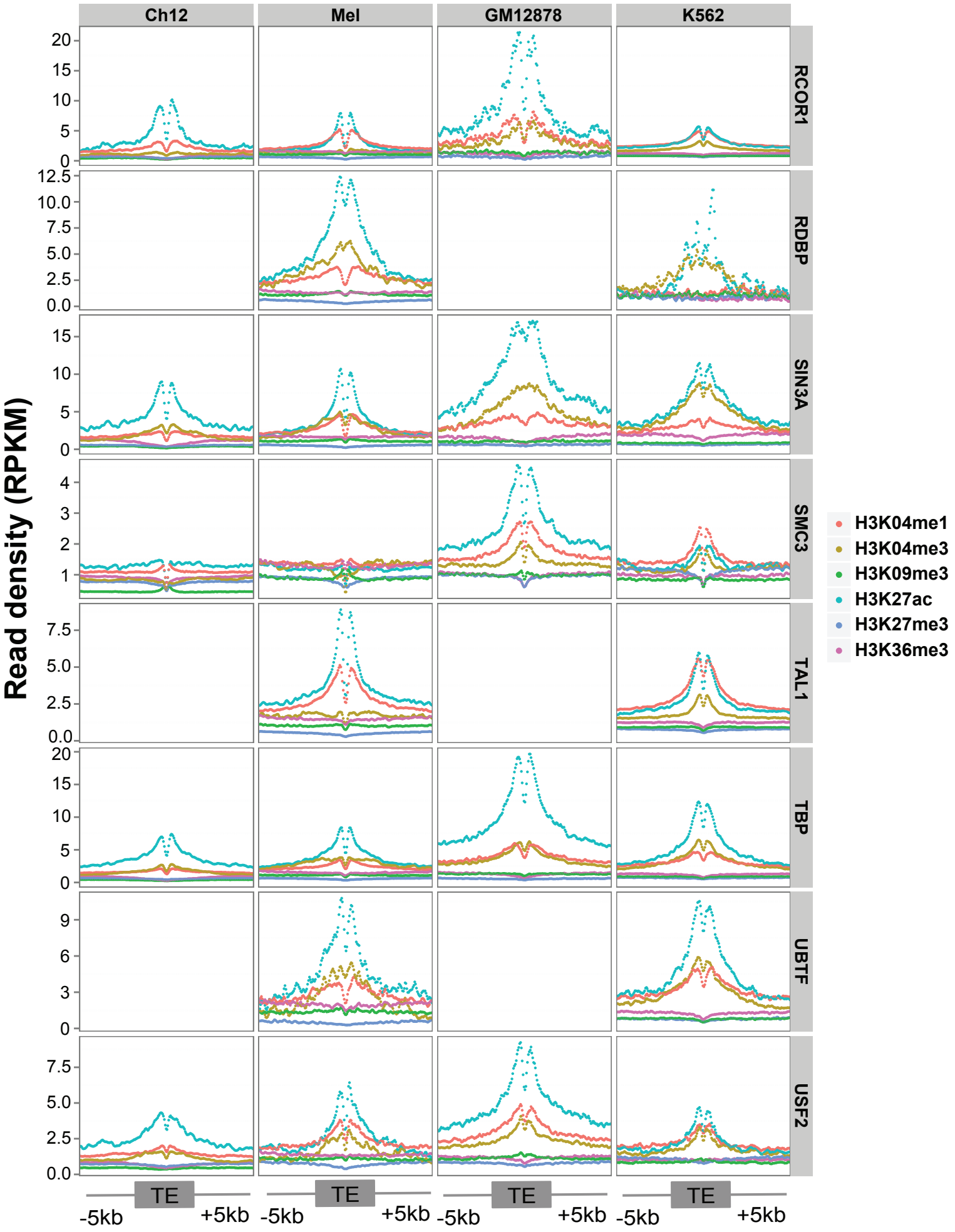


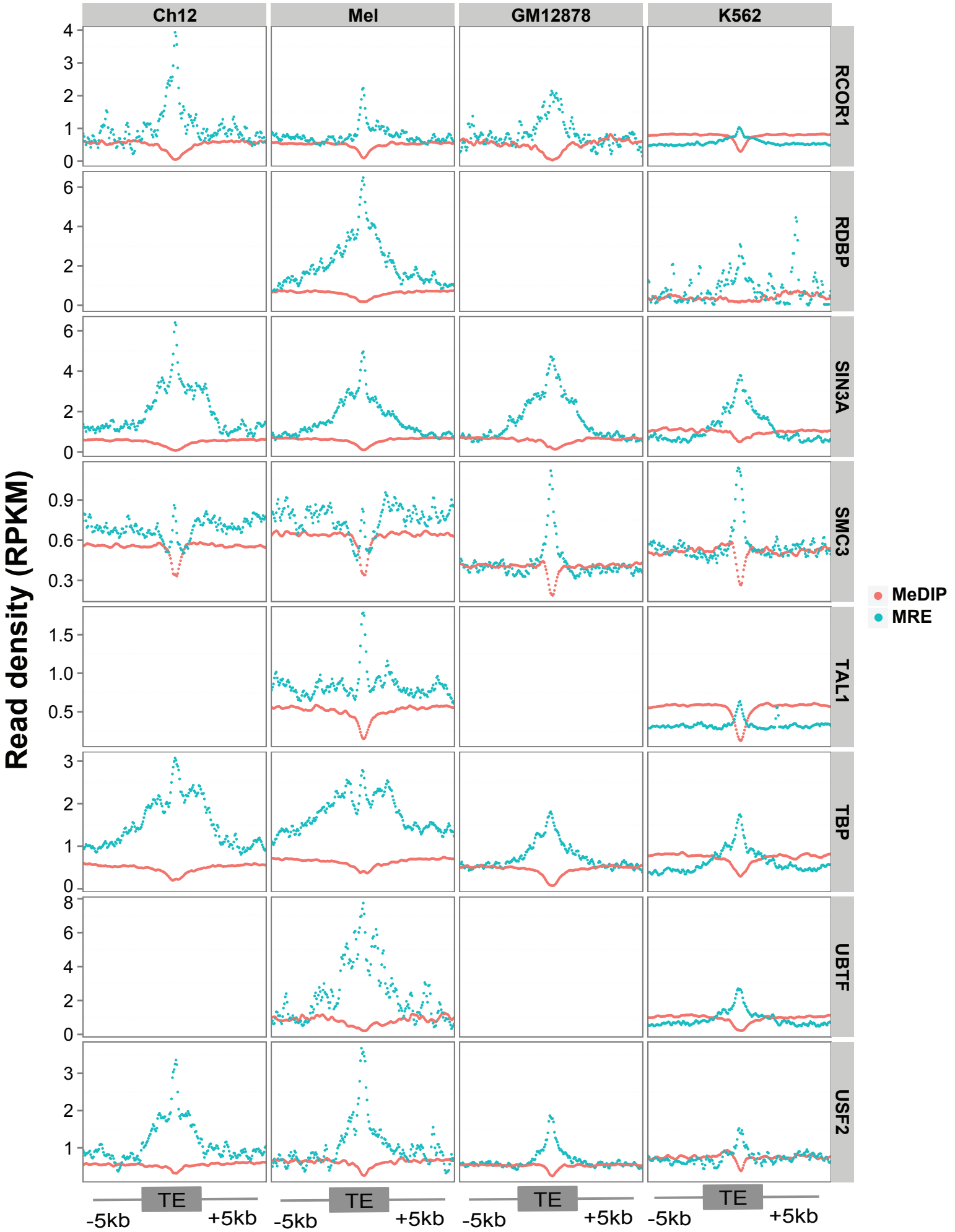
Supplementary Figure 12: Epigenetic profile TE-derived peaks for each of the 26 TFs. For each TF, in each cell line we obtained the TEs that contributed binding peaks and profiled their epigenetic signature for six histone modification marks and DNA methylation (subsequent pages).

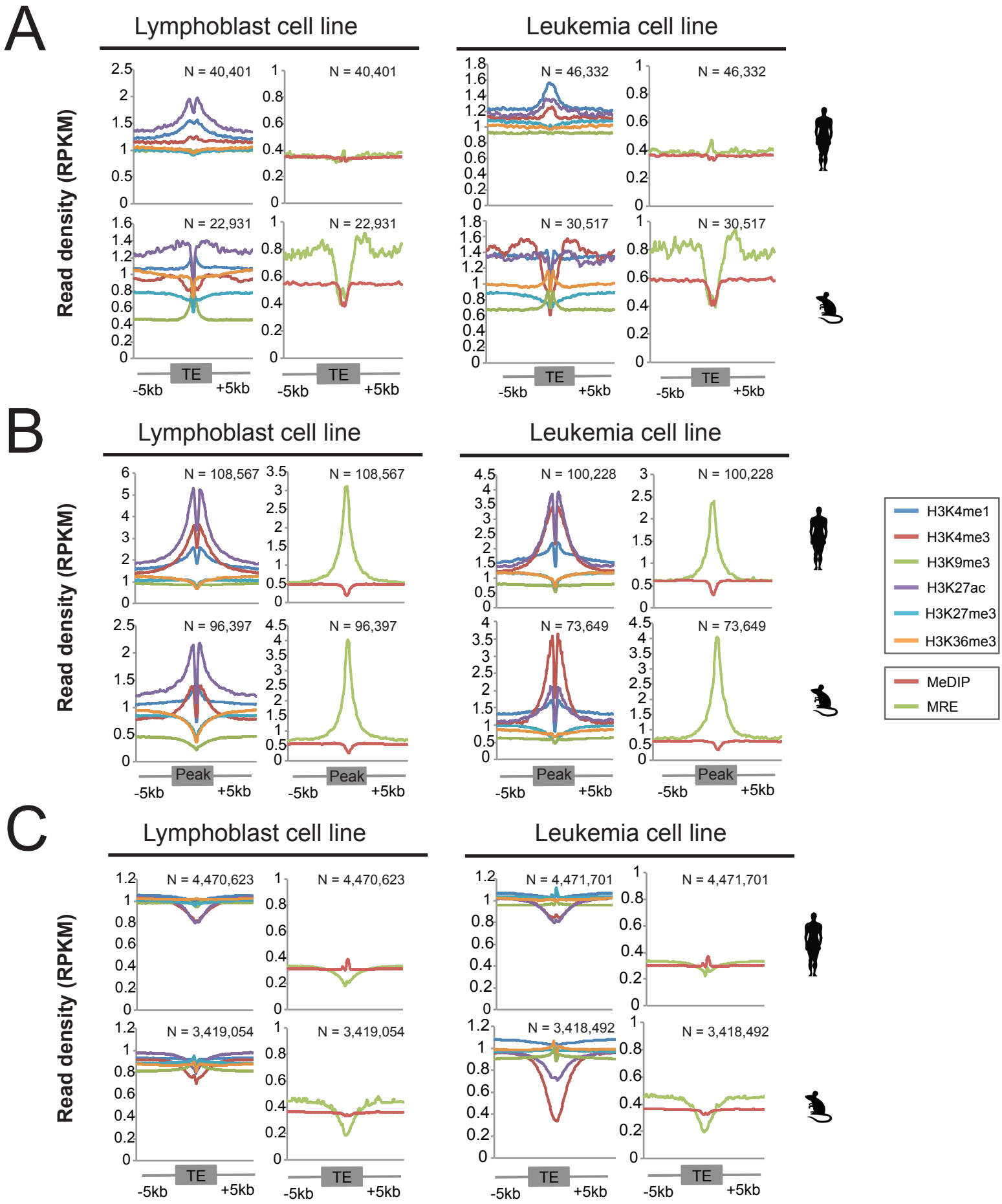

Ch12
Mel
GM12878
K562
BHLE40
CHD1
CHD2
CTCF
E2F4
EP300
ETS1
GATA1
JUND
Read density (RPKM)
MeDIP
MRE
-5kb
TE
+5kb


Ch12
Mel
GM12878
K562
KAT2A
MAFK
MAX
MAZ
MXI1
MYC
PAX5
Pol2
RAD21
Read density (RPKM)
H3K04me1
H3K04me3
H3K09me3
H3K27ac
H3K27me3
H3K36me3
-5kb
TE
+5kb

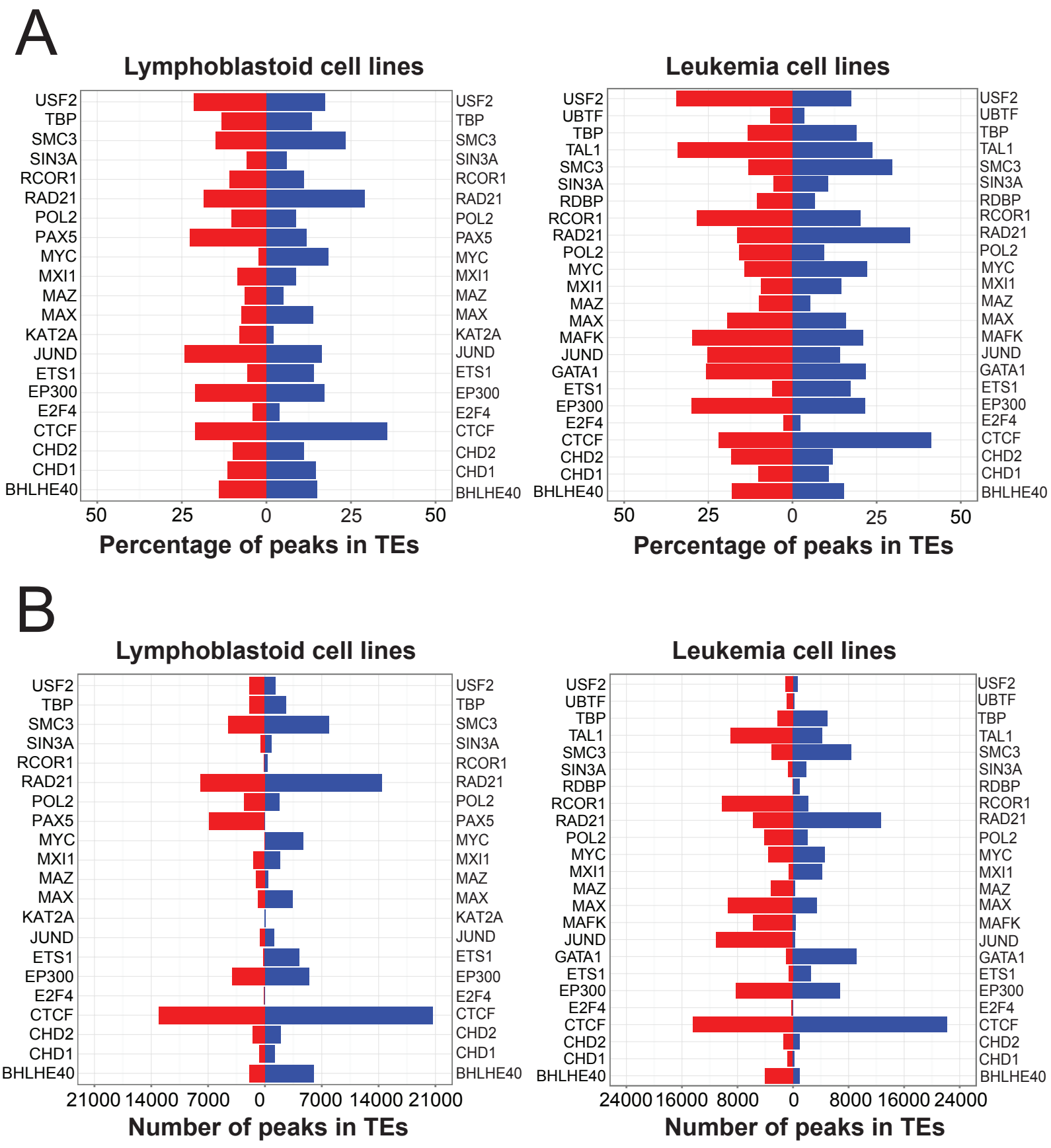

Ch12
Mel
GM12878
K562
KAT2A
MAFK
MAX
MAZ
MXI1
MYC
PAX5
Pol2
RAD21
Read density (RPKM)
MeDIP
MRE
-5kb
TE
+5kb


Ch12
Mel
GM12878
K562
RCOR1
RDBP
SIN3A
SMC3
TAL1
TBP
UBTF
USF2
Read density (RPKM)
H3K04me1
H3K04me3
H3K09me3
H3K27ac
H3K27me3
H3K36me3
-5kb
TE
+5kb


Ch12
Mel
GM12878
K562
RCOR1
RDBP
SIN3A
SMC3
TAL1
TBP
UBTF
USF2
Read density (RPKM)
MeDIP
MRE
-5kb
TE
+5kb



Supplementary Figure 13: Epigenetic profile of TE-derived peaks for CTCF-associated factors. All panels represent a 10kb regions centered on the region of interest (TEs, or peaks). We averaged the epignetic signal in 50bp bins, over the regions of interest for histone marks (left panel), and DNA methylation (right panel), in human (upper panels) and mouse (lower panels). (A) TEs that contributed to TF-binding peaks differ from the signature of non-TE peaks (B). (C) TE sequences that did not contributed TF binding peaks, also differ greatly.



Supplementary Figure 14: TE-encoded peaks in human (red) -mouse (blue) pairs of lymphoblast (left panels) and leukemia (right panels) cells. (A) Percentage of TF binding peaks in TEs in human-mouse pairs of lymphoblast (left), and leukemia (right) cells. (B) Number of TF binding peaks in TEs in human-mouse pairs of lymphoblast (left) and leukemia (right) cells. (*See Methods for description of TE-derived peaks*).