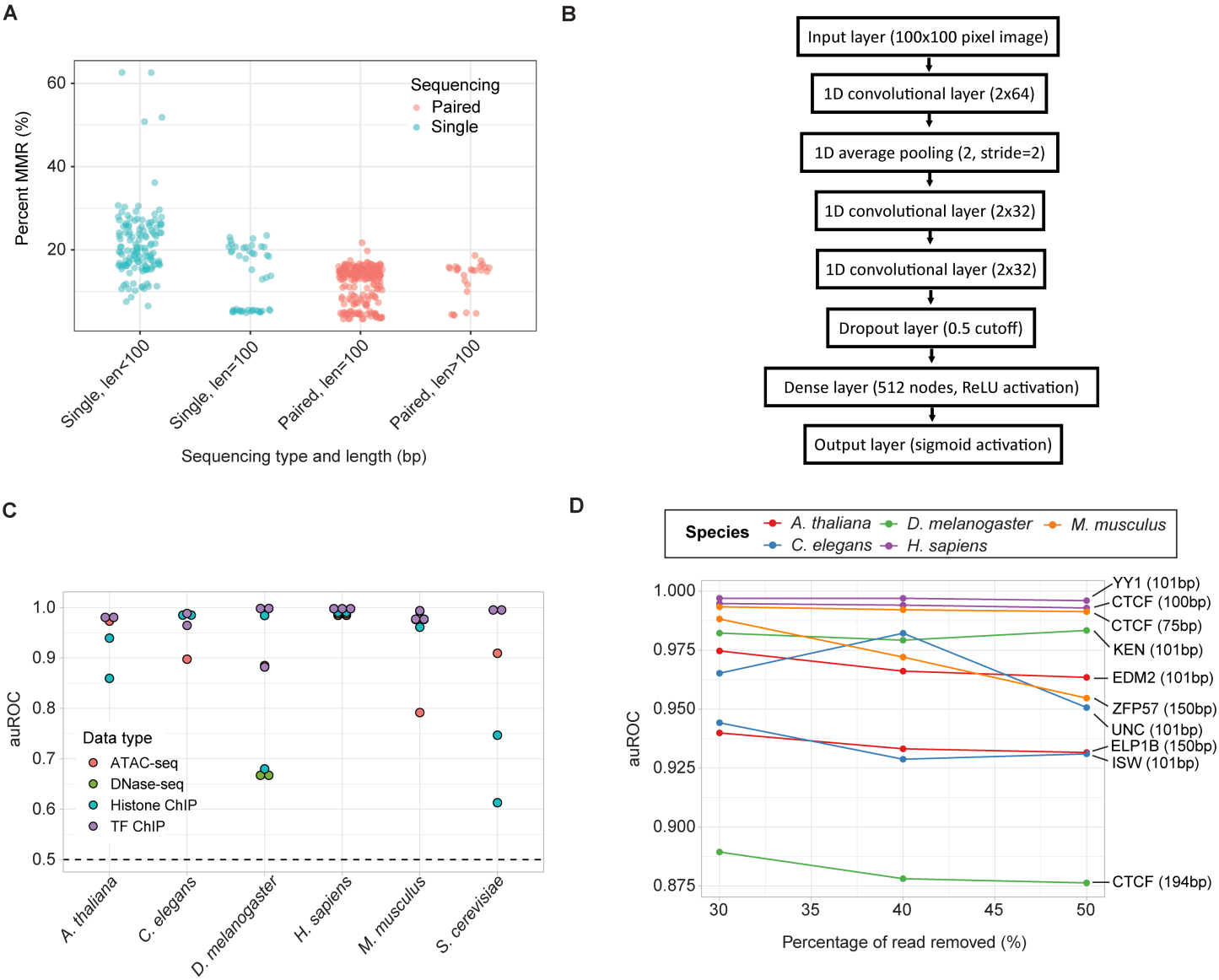
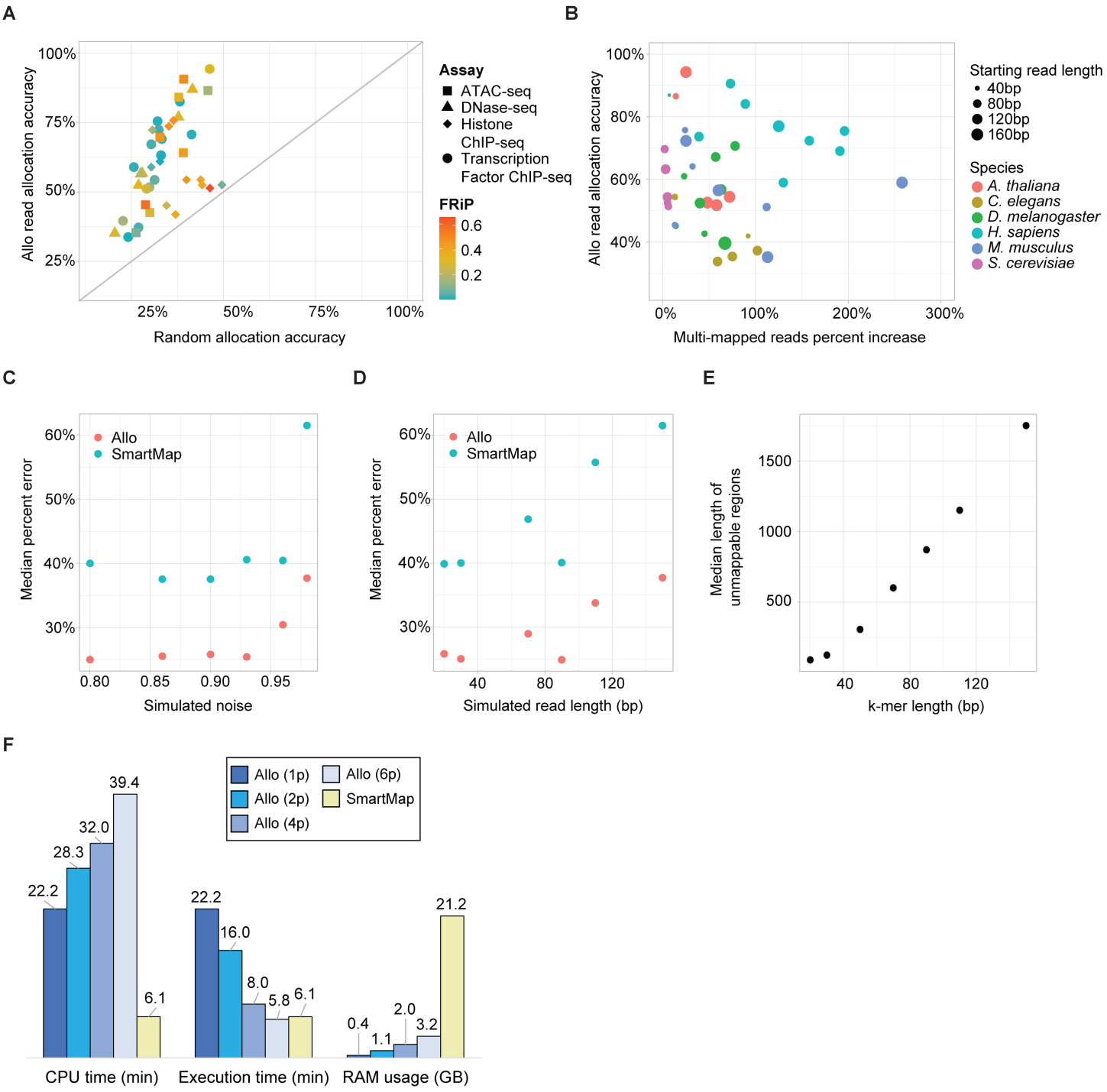
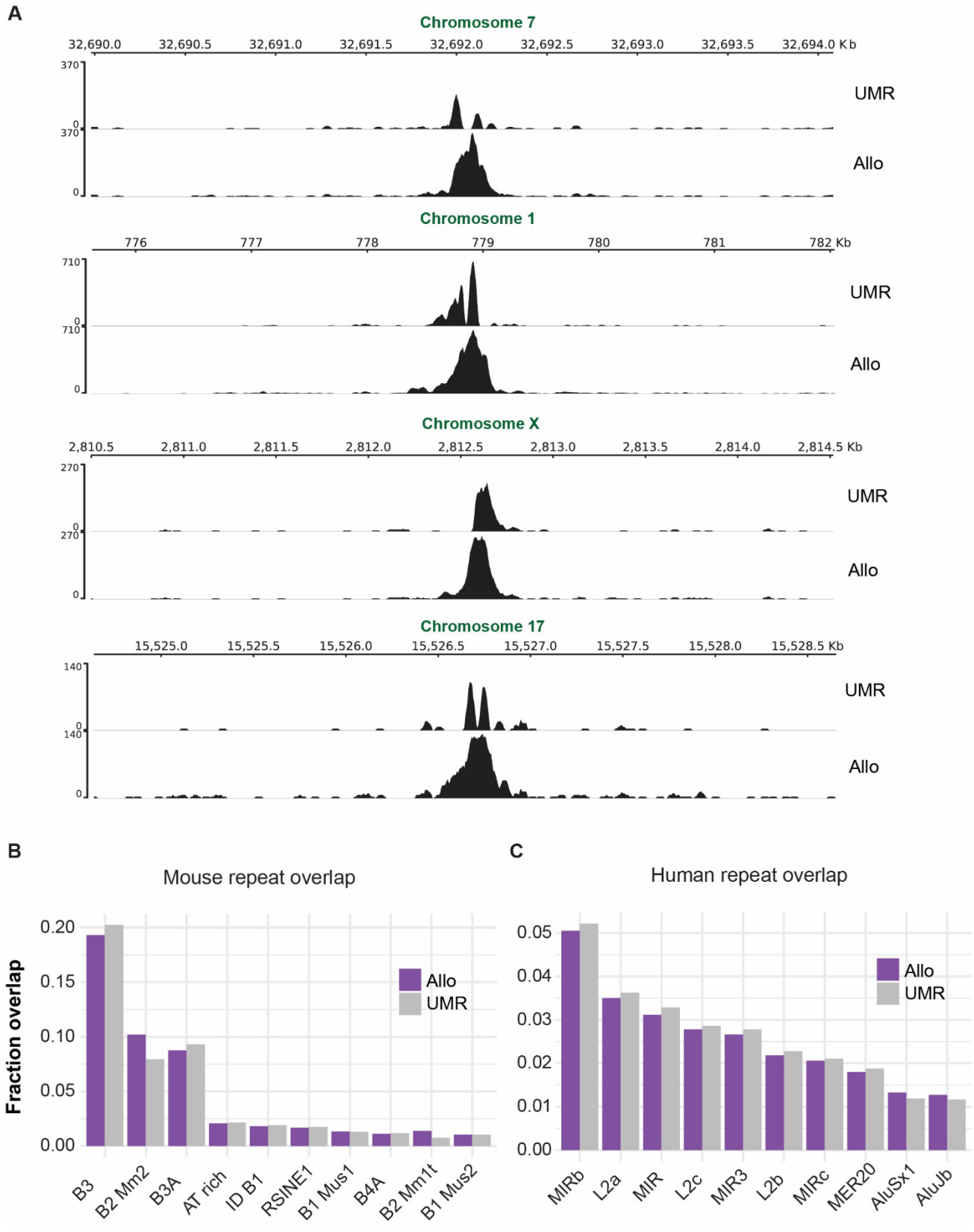
Supplemental Figures



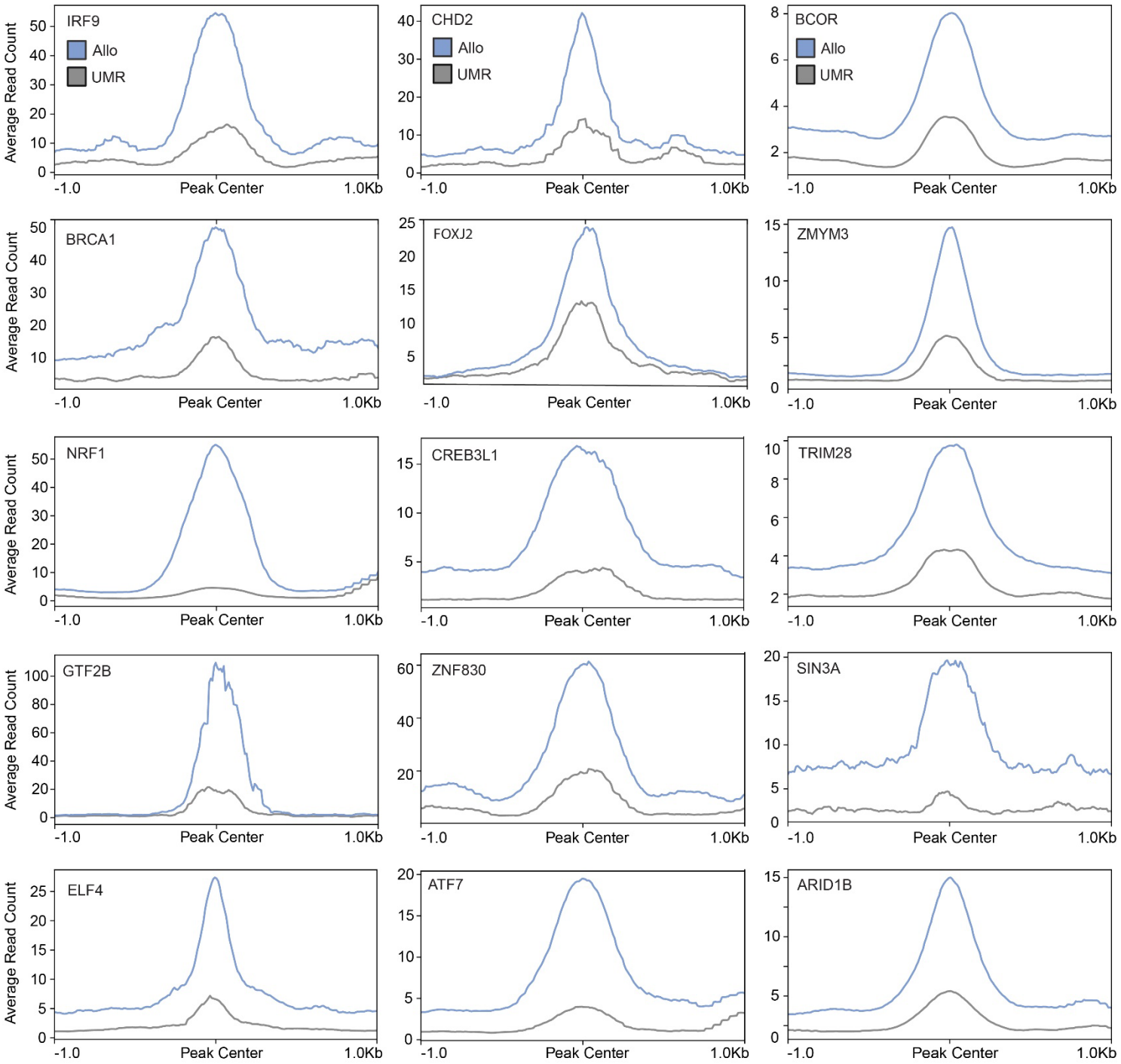
Supplemental Figure 1: A) Percentage of multi-mapped reads in 481 K562 ChIP-seq datasets, given read length and sequencing type (paired-end or single-end). Sequencing type is indicated by color. **B)** Architecture of Allo's convolutional neural network. The structure is the same for both the narrow and mixed peak CNNs. **C)** auROC values of Allo's neural networks across various species. **D)** auROC values from a selection of the testing datasets at different lengths after artificial trimming. There are two transcription factors represented per organism tested. The read length of the original dataset is listed in parentheses beside the name of the transcription factor.



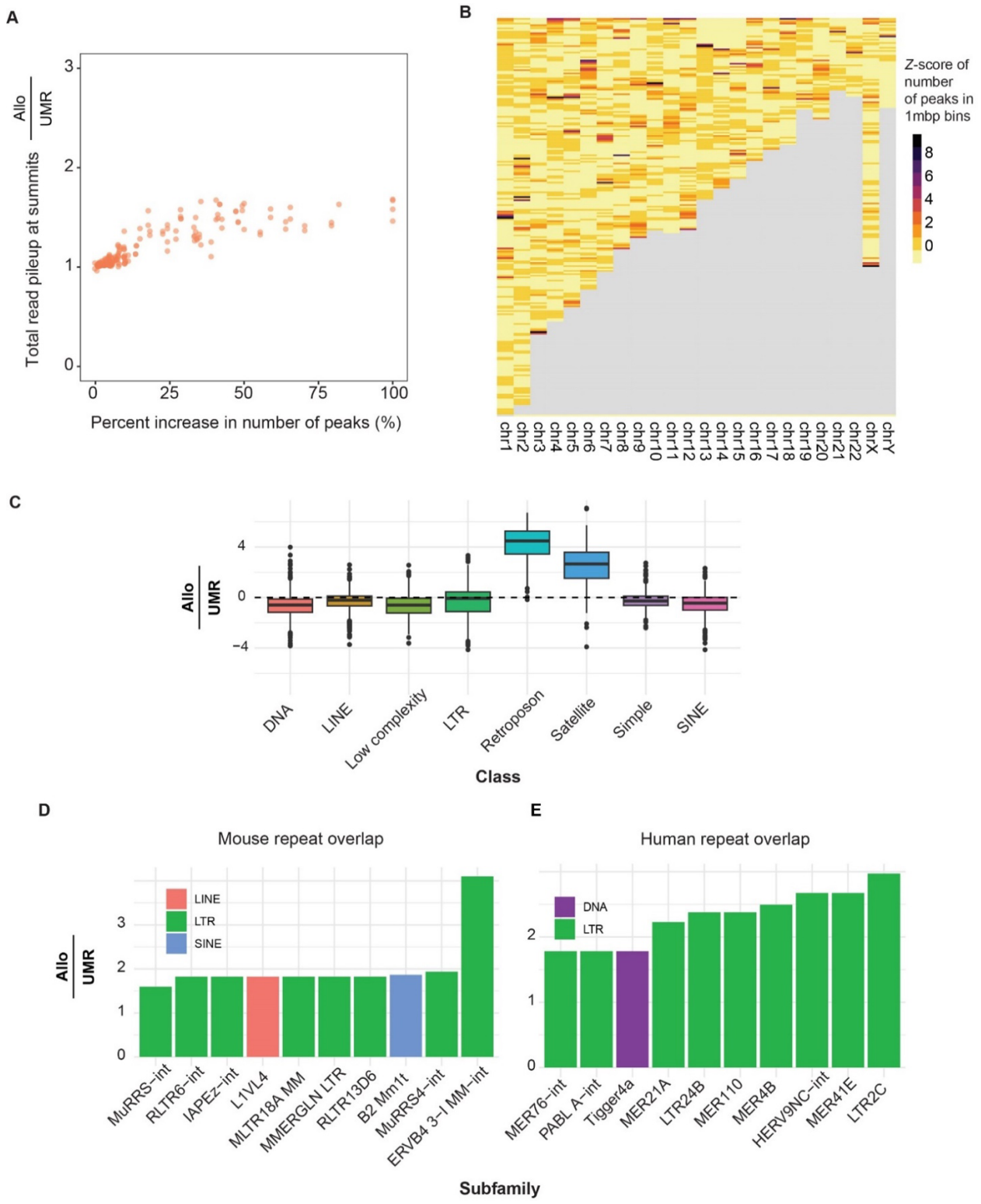
Supplemental Figure 2: A) Comparison between Allo's read allocation accuracy and random read allocation accuracy. Color indicates the FRiP score of the full-length dataset. Shapes indicate the assay examined in each sample. **B)** Comparison between Allo's read allocation accuracy and the proportional increase read counts due to multi-mapped reads (when datasets are trimmed to 30bp). The size of the dots indicate the starting read length of the datasets before trimming. The color indicates source species. **C)** Median percent error of Allo and SmartMap compared with the simulated noise level used during ChIP-seq data simulation. Color indicates software used for allocation. **D)** Median percent error of Allo and SmartMap compared with the simulated read length in basepairs. Color incudes software used for allocation. **E)** Median length of unmappable regions in hg38 at different lengths of k-mers. **F)** Computational performance comparisons between Allo using various numbers of processes and SmartMap. Values given are per 1 million reads allocated.



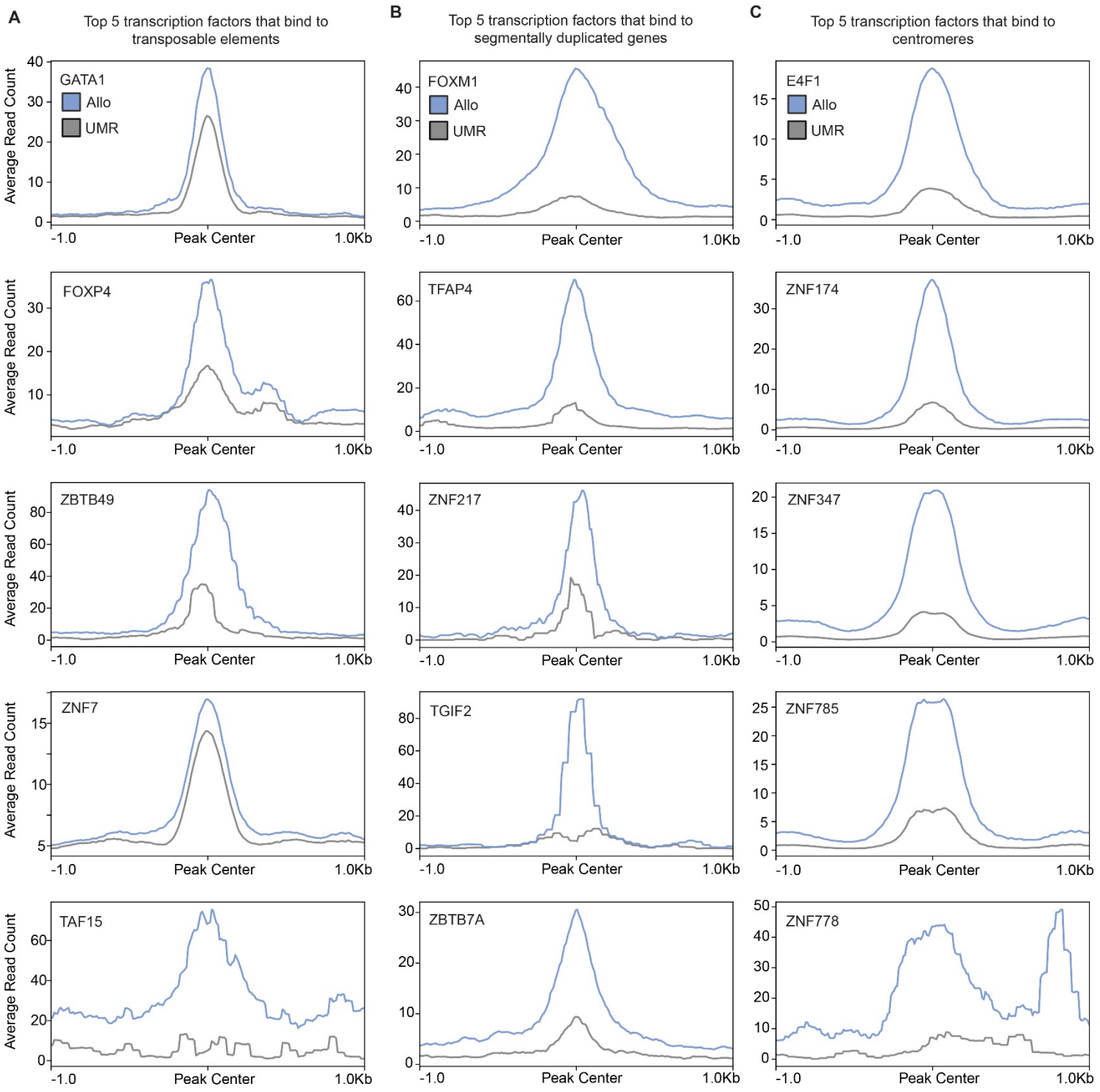
Supplemental Figure 3: A) Genome browser screenshots at locations of K562 CTCF UMR-derived peaks that increased in resolution after the inclusion of Allo-allocated multi-mapped reads. "UMR" tracks display only uniquely-mapped reads, while "Allo" tracks display both uniquely-mapped and Allo-allocated multi-mapped reads. **B,C)** Fraction of overlap with repetitive element subfamilies in UMR CTCF peaks and Allo-only CTCF peaks in mouse (**B**) and human (**C**) datasets. Top 10 subfamilies with the highest overlaps are shown. Color indicates Allo-only or UMR peak sets.



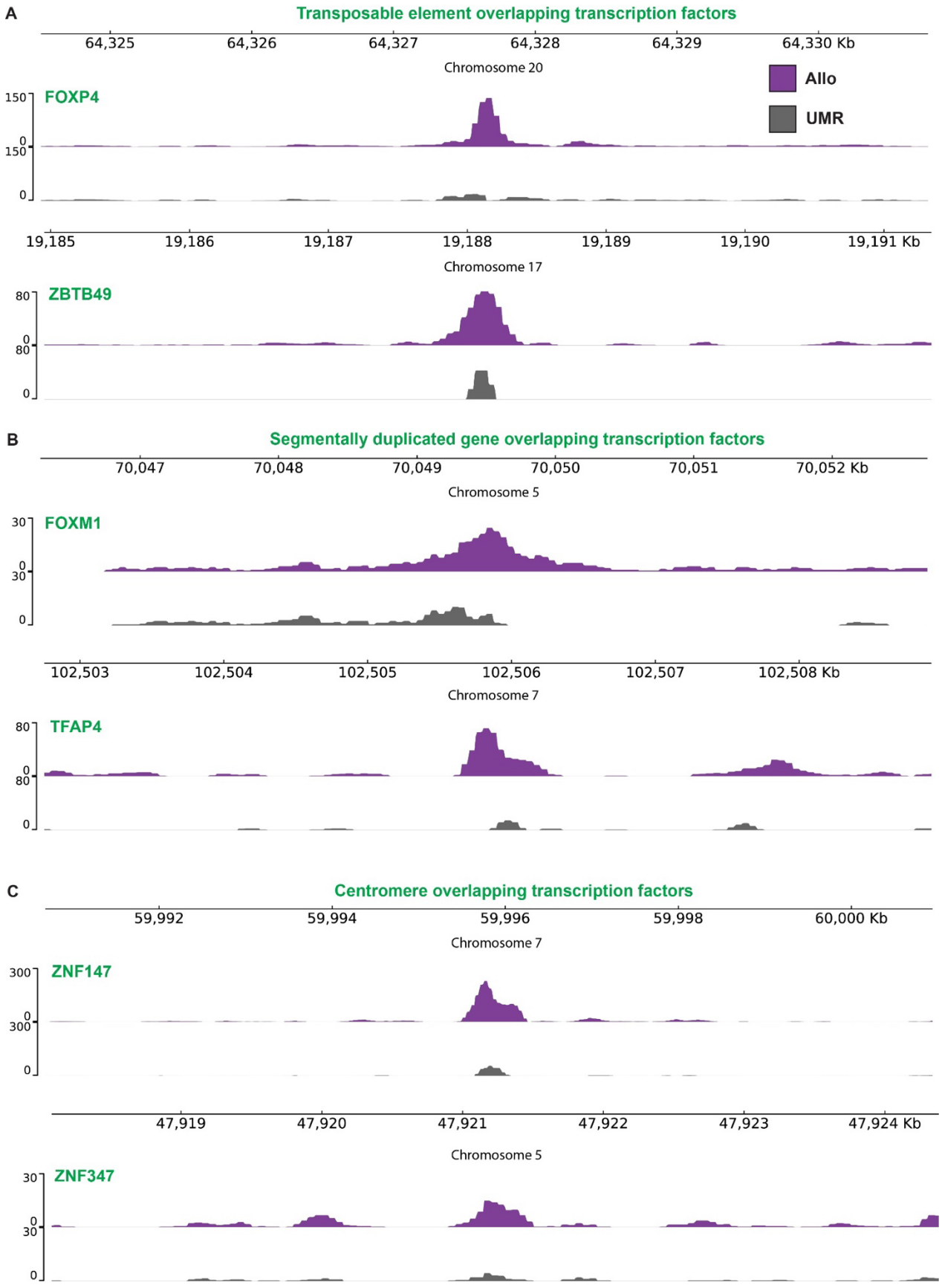
Supplemental Figure 4: Profile plots before and after the inclusion of Allo within Allo-only peaks in 15 randomly chosen K562 TF ChIP-seq datasets. Grey lines represent uniquely mapped read counts only and blue lines represent the inclusion of Allo-allocated MMRs.



Supplemental Figure 5: A) Comparison between the total read pileup count at peaks before and after the inclusion of Allo on 481 K562 ChIP-seq datasets. **B)** Locations of Allo-only peaks across 481 K562 ChIP-seq samples, normalized by *Z*-score. Each bin represents a 1 million base pair window. Color indicates the magnitude of the *Z*-score after normalization. **C)** The fraction of overlap of Allo-only peaks with specific repetitive element classes divided by the fraction of overlap of UMR-derived peaks at the same repetitive element classes. **D,C)** The fraction of overlap in Allo-only CTCF peaks versus UMR CTCF peaks in repetitive element subfamilies in mouse and human datasets. Top 10 subfamilies with the highest ratio are shown. Color indicates the repetitive element class.



Supplemental Figure 6: Profile plots before and after the inclusion of Allo within Allo-only peaks that overlap: **A)** transposable elements; **B)** centromeric satellite repeats; and **C)** segmentally duplicated genes. Grey lines represent uniquely mapped read counts only and blue lines represent the inclusion of Allo-allocated MMRs.



Supplemental Figure 7: Genome browser screenshots at randomly selected Allo-only peaks that overlapped various genomic regions including: **A)** transposable elements; **B)** segmentally duplicated genes; and **C)** centromeric satellite repeats. Grey tracks represent uniquely mapped read counts only and purple tracks represent the inclusion of Allo-allocated MMRs.