Supplementary Figure 1. Quality control and benchmarking data for GRO-seq samples. (A) Highly consistent results were obtained from technical replicates, as shown by a correlation plot depicting the GRO-seq signal from NHA cells. Spearman correlation value for the technical replicate pair was 0.99. (B) Highly consistent results were obtained from biological replicates, as shown by a correlation plot depicting the GRO-seq signal from HUVEC cells. Spearman correlation value for the biological replicate pair was 0.98. (C) Spearman correlation coefficients are plotted for all samples produced and submitted as GEO entry GSE92375. (D). The performance of different transcript discovery settings to capture the full transcribed area (as defined by combined result) is compared between cell types of different read depth (A549, low, 10.3 million reads; HUVEC, medium, 47.4 million reads; HCT116, high, 923 million reads). (E) The contribution of each setting (as percentage) to the TSS discovered is compared similarly as in D. Notice that the pooled setting contains some unique transcripts due to more robust transcript body assembly.

Supplementary Figure 2. Workflow for pri-miRNA analysis and benchmarking for TSS identification. (A) Workflow chart describing the approach used for transcript detection, evaluation and quantification. Main tools and parameter settings implemented as three consecutive pipelines are illustrated (refer to Material and Methods for details). In each cell type, the TSS detected with different settings are supported based on the CAGE-seq and ChIP-seq data. The initial transcript body assembly includes extension across gaps and inclusion of intragenic TSS. Subsequently, transcript data is evaluated and combined across cell types to define regions for quantification, called TSS elements. Exons and non-mappable regions are subtracted for quantification from these regions that extend from each TSS to the next, and the last element is refined to end where the signal drops significantly. As final step, the differential TSS activity at a given gene locus is calculated by subtracting the RPKM value of an upstream TSS element, resulting in a heatmap reflecting the TSS activities across cell types. (B) Venn diagram comparing the number of pre-miRNA with an overlapping transcript element identified in Chang et al., 2015 (total 1267) and in the approach presented here (total 1291). Common pre-miRNAs (1142) include also 68 pre-miRNA TSS identified by both approaches but with a lack of GRO-seq data signal for transcript elongation until the pre-miRNA coordinate. (C) Expression levels (RPKM) of pre-miRNAs identified by using only our approach (total 149), by both approaches (total 1142) and only in Chang et al., 2015 (total 125) are shown as boxplots.

Supplementary Figure 3. GRO-seq signal across cell types and individual samples reveals distinct TSS usage at the hsa-mir-29a~29b-1 cluster locus. (A) GRO-seq signal across 19/28 cell types in the pri-miRNA hsa-mir-29a~29b-1 cluster locus as in Figure 3 (chr7:130,376,783-131,016,782, similar cell types to those shown were omitted for clarity, refer to Figure 3A). (B) Signal from individual samples from LNCaP cells in response to TNFα and/or Dihydrotestosterone (DHT) or control (DMSO) treatment, in two replicates, is visualized from the same region. (C) TT-seq, PRO-seq, GRO-seq and CAGE-seq signal from K562 cells is visualized from the same region and from the start of the transcript (+/- 1000 bp around the TSS4).

Supplementary Figure 4. miRNA- and TF-expression level-derived stratification of samples. Sample location based t-SNE dimensionality reduction of pri-miRNA (A) or TF expression profiles (B) is shown on the map. Different symbols and colors indicate the cell origin, as specified in the legend.

Supplementary Figure 5. Short RNA-seq and RT-qPCR data for mature miRNA species. (A) Short RNA-seq data from (8), RT-qPCR data and ENCODE short RNA-seq based quantification of mature miRNA species produced from the hsa-mir-221~222 locus. (B) Short RNA-seq data from (8) based quantification of mature miRNA species produced from the hsa-mir-100~let-7a-2~mir-125b-1 locus. (C) Short RNA-seq data from (8) based quantification of mature miRNA species produced from the hsa-mir-29a~29b-1 locus. Relative fold changes to mature miRNA levels in the A549 cell are shown for each cluster.

Supplementary Figure 6. eRNA levels across all enhancers from SE regions in the vicinity of hsa-mir-221~222 locus. The barplots showing eRNA expression are organized by SE numbering (SE1-6). The coloring of cell types corresponds to that used in the GRO-seq signal tracks (in Figure 4A).

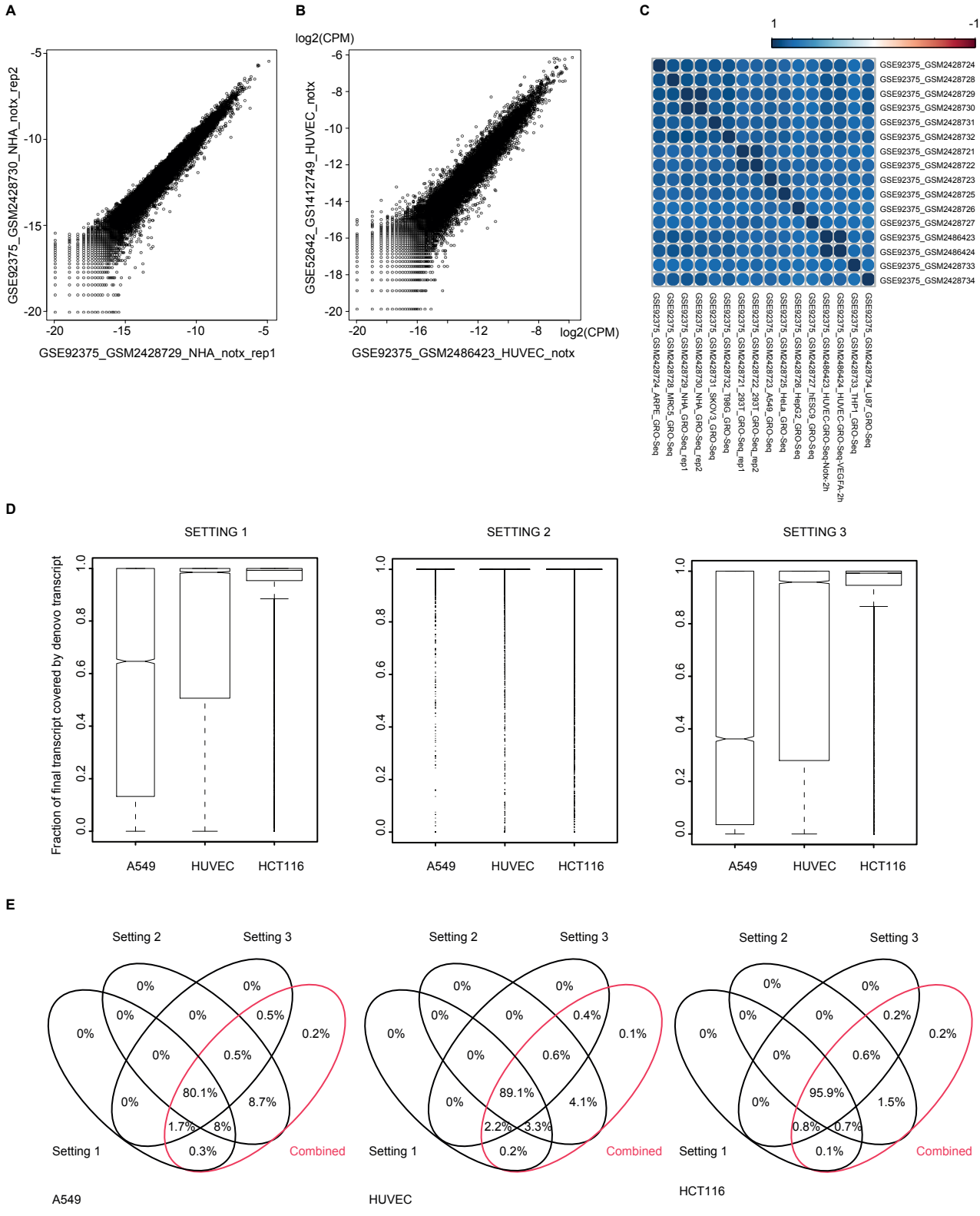
Supplementary Figure 7. Transcriptional changes within the hsa-mir-221~222 locus upon inflammatory activation. GRO-seq signal from the TNFα-stimulated LNCaP cells is shown across the whole hsa-miR-221~222 cluster locus (chrX:45,169,213-46,205,597). NF-κB binding as measured in the same condition (GM12878 cells) and layered H3K27ac are shown below the GRO-seq tracks.

Supplementary Table 1. List of all GRO-seq samples used in the analysis.

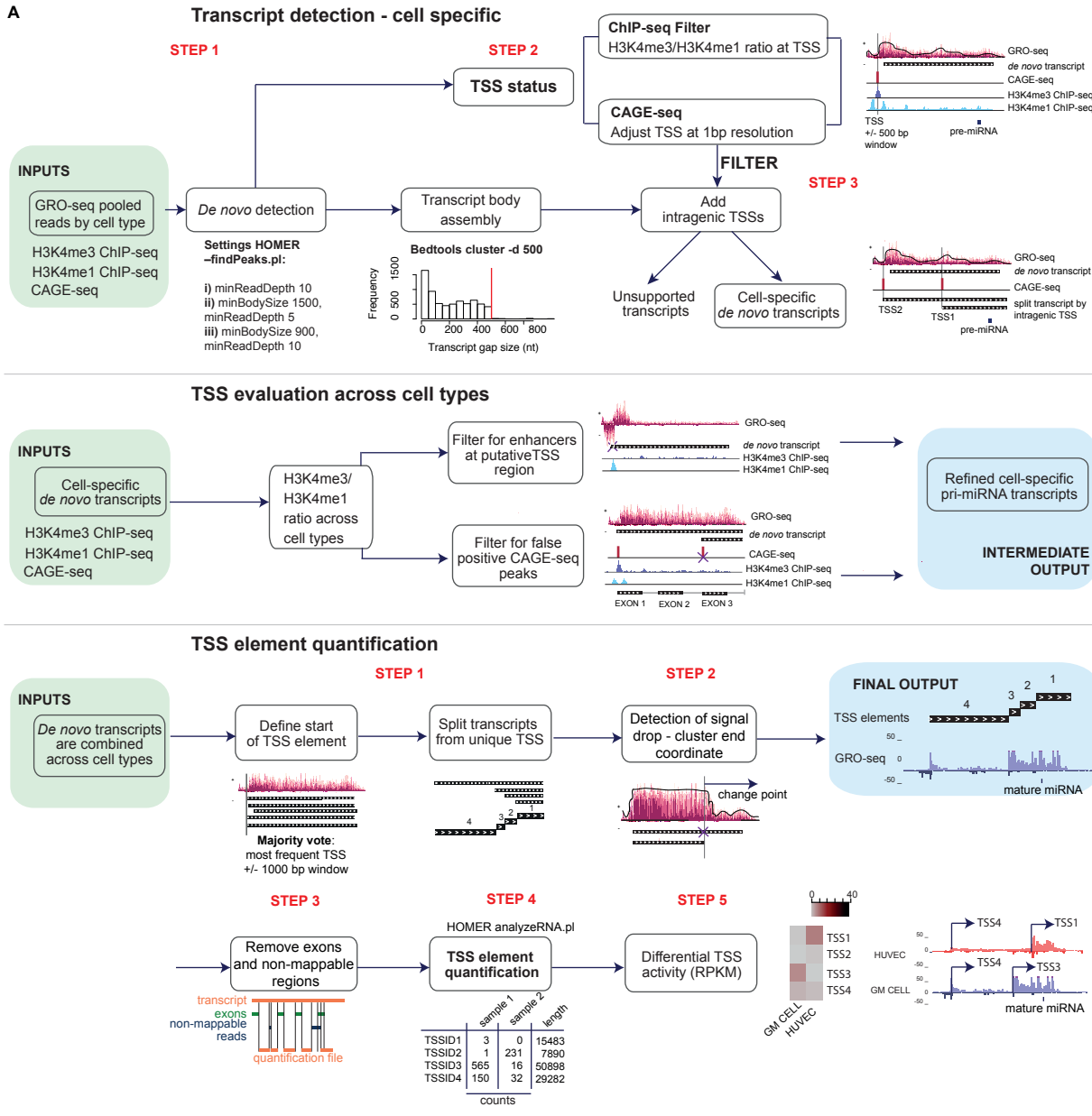
Supplementary Table 2. List of additional datasets - ChIP-seq, CAGE-seq and Hi-C data.

Supplementary Table 3. Summary of detected pri-miRNA TSS elements. The TSS identifier, chromosome, start, end, unique name and strand information is followed by list of cell types where each TSS was identified, CAGE-seq scores and ChIP-seq verification (H3K4me3 confirmed: H3K4me3 > H3K4me1 in at least one cell type; H3K4me1 found: H3K4me1 > H3K4me3 in at least one cell type). Additional datasheet lists TSS with pre-miRNA overlap assigned based on Chang 2015 assembly.

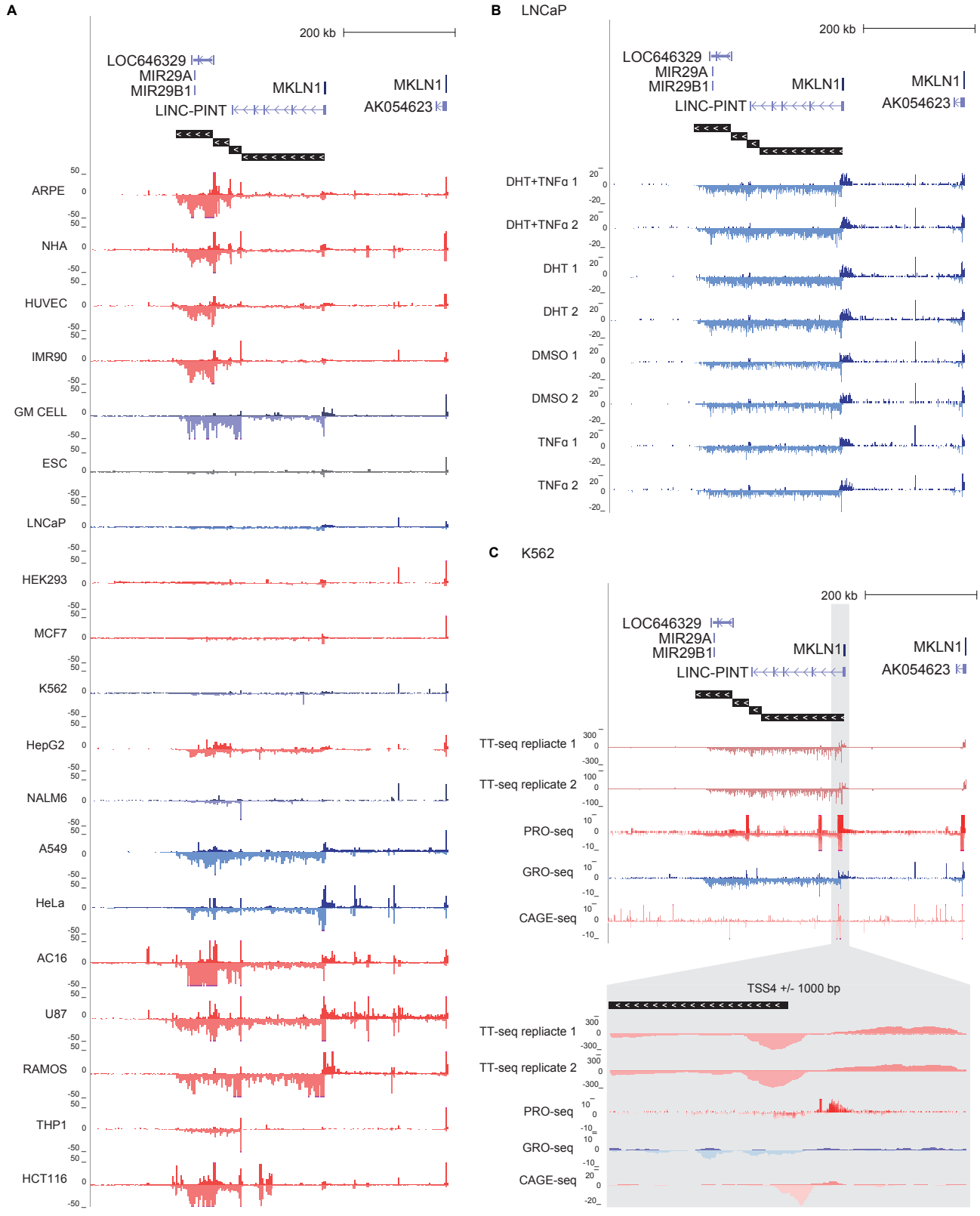
Supplementary Figure 1



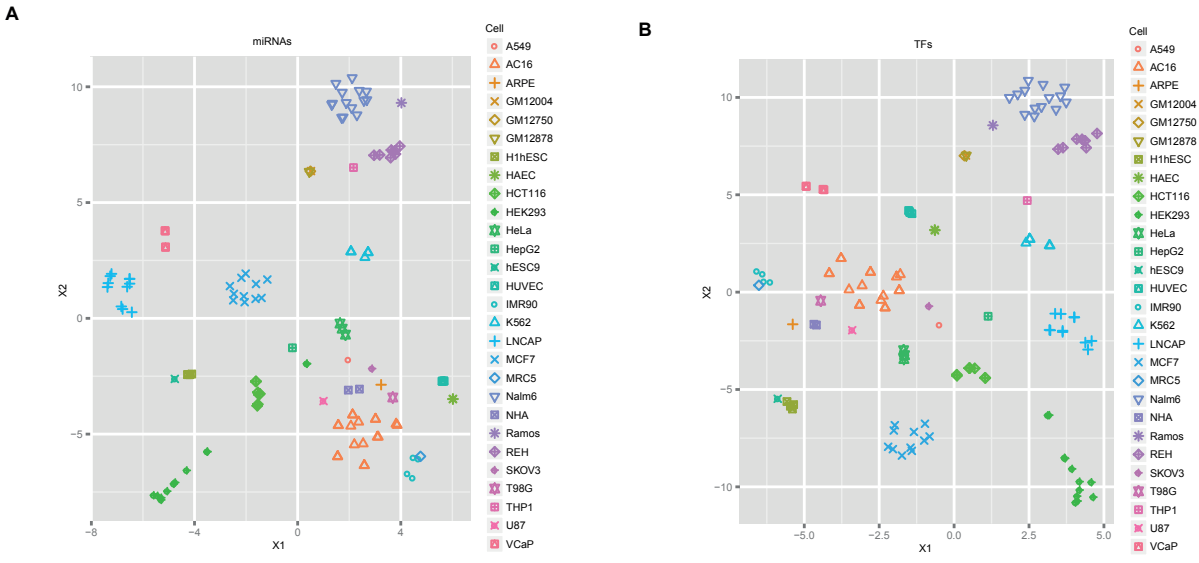
Supplementary Figure 2



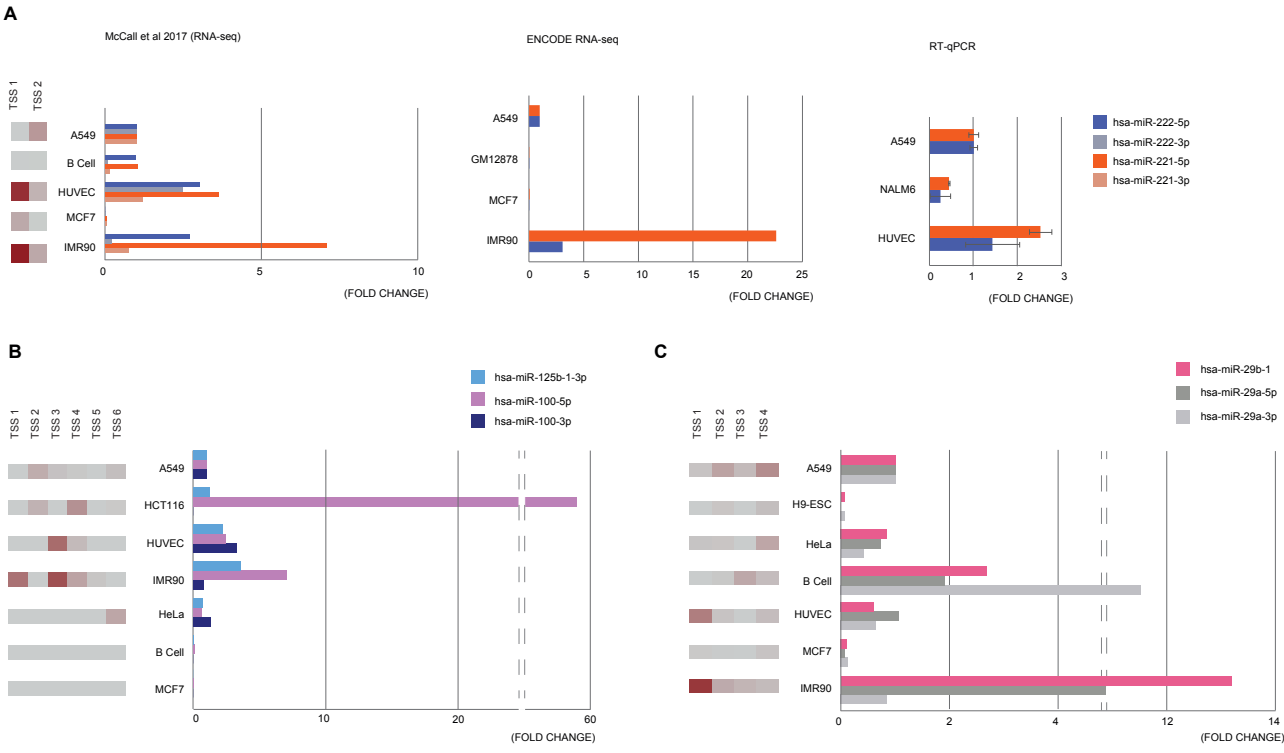
Supplementary Figure 3



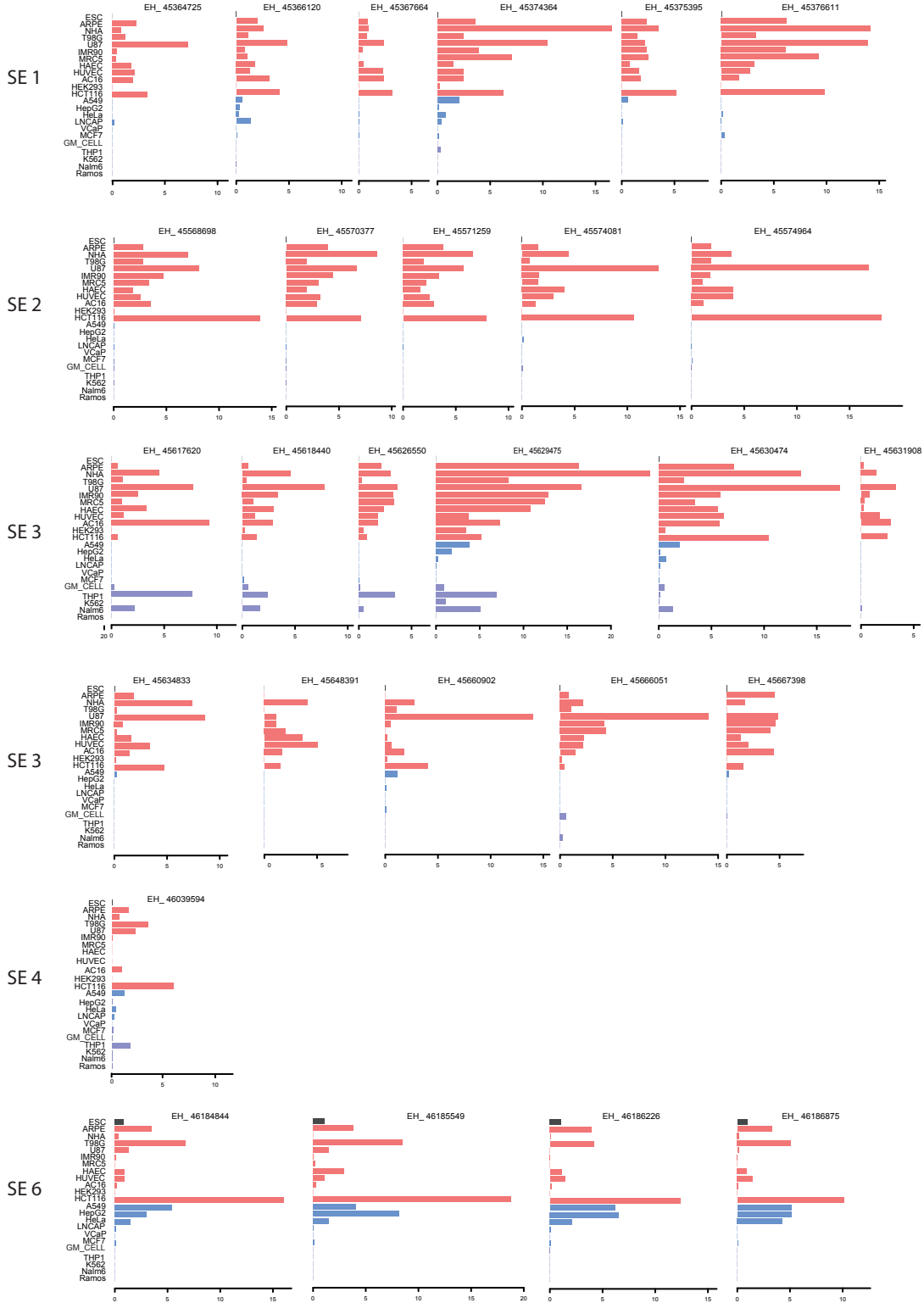
Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6



Supplementary Figure 7