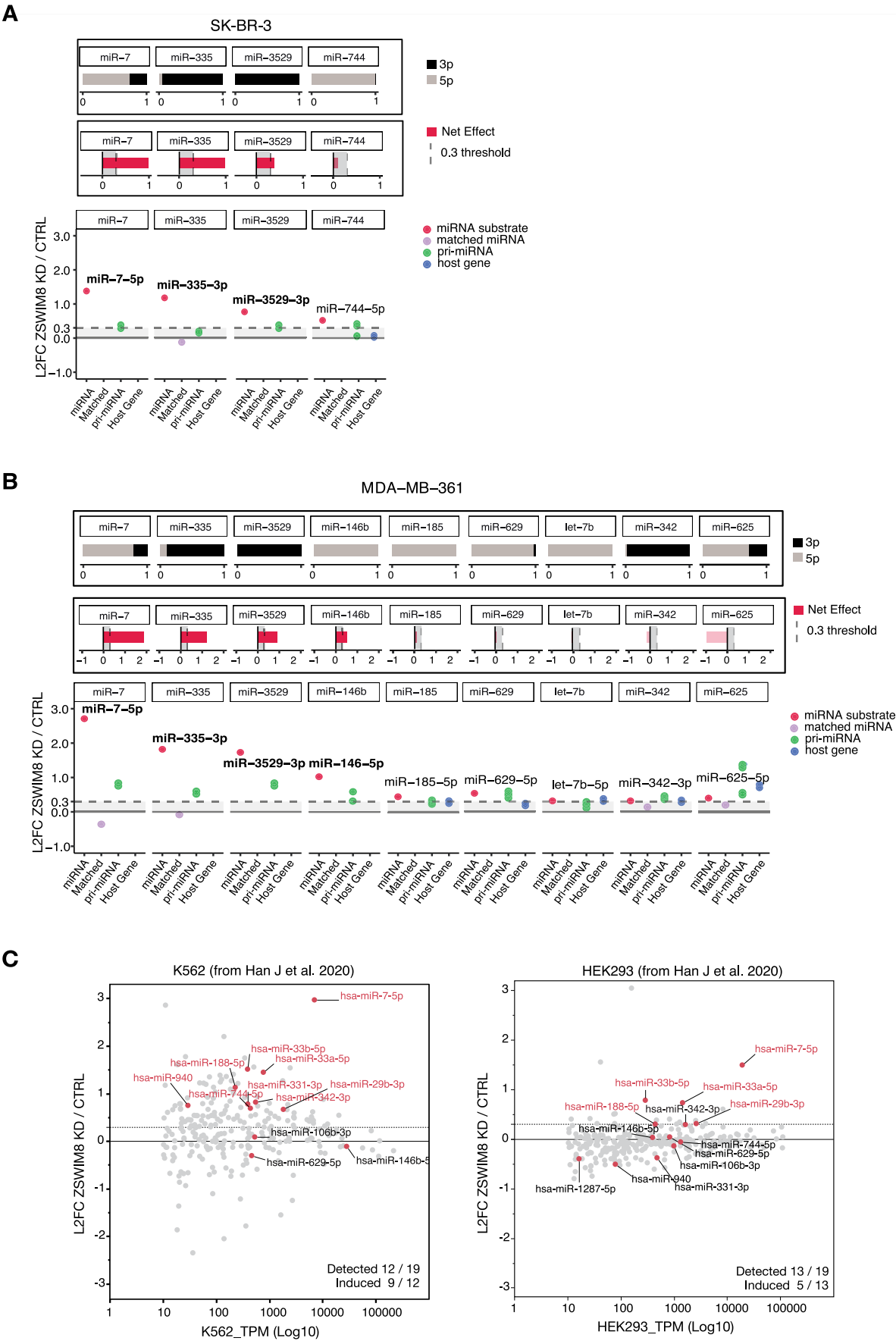
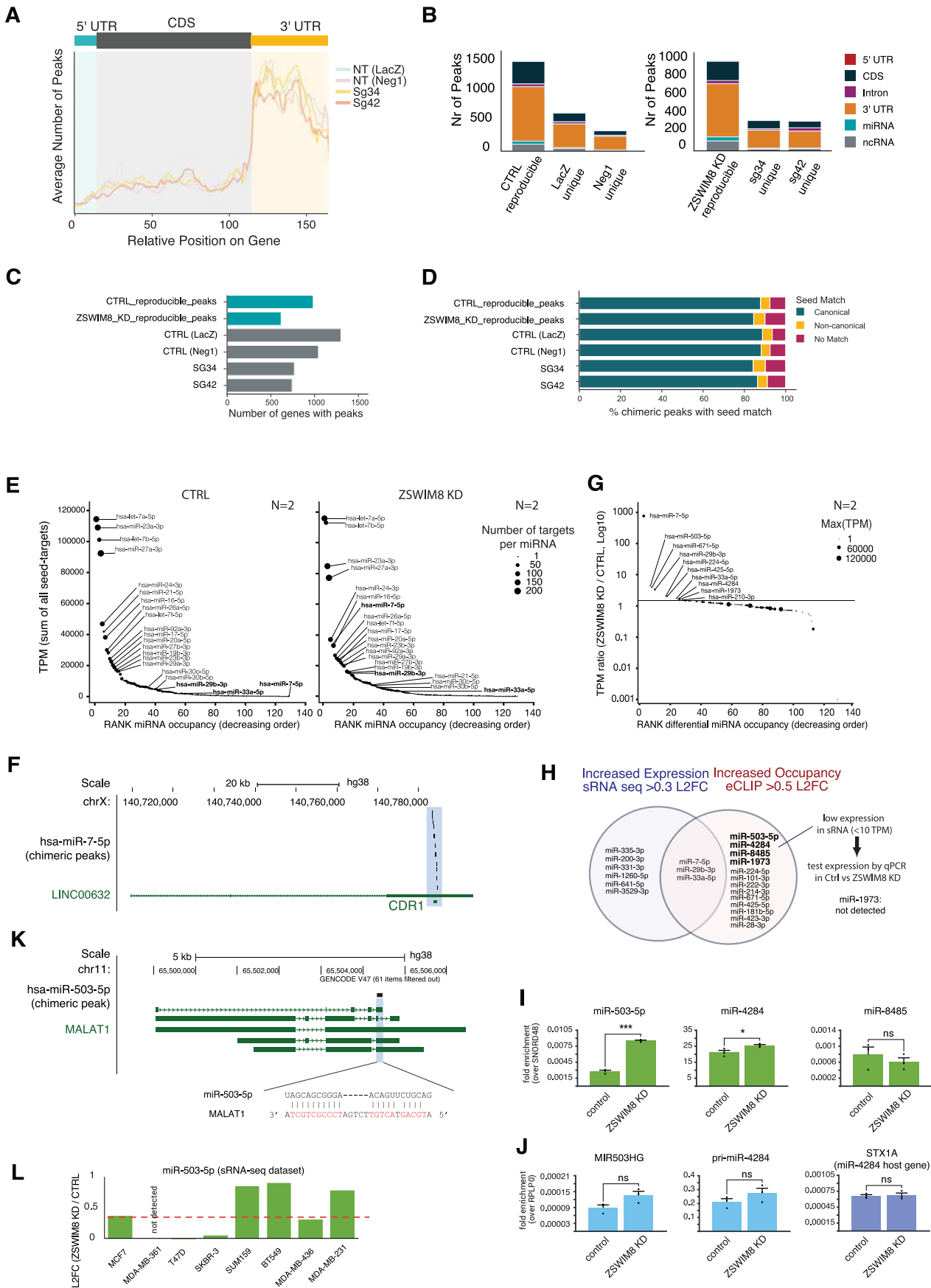


## Expanded View Figures

### Figure EV1. Identification of high-confidence TDMD miRNA substrates.

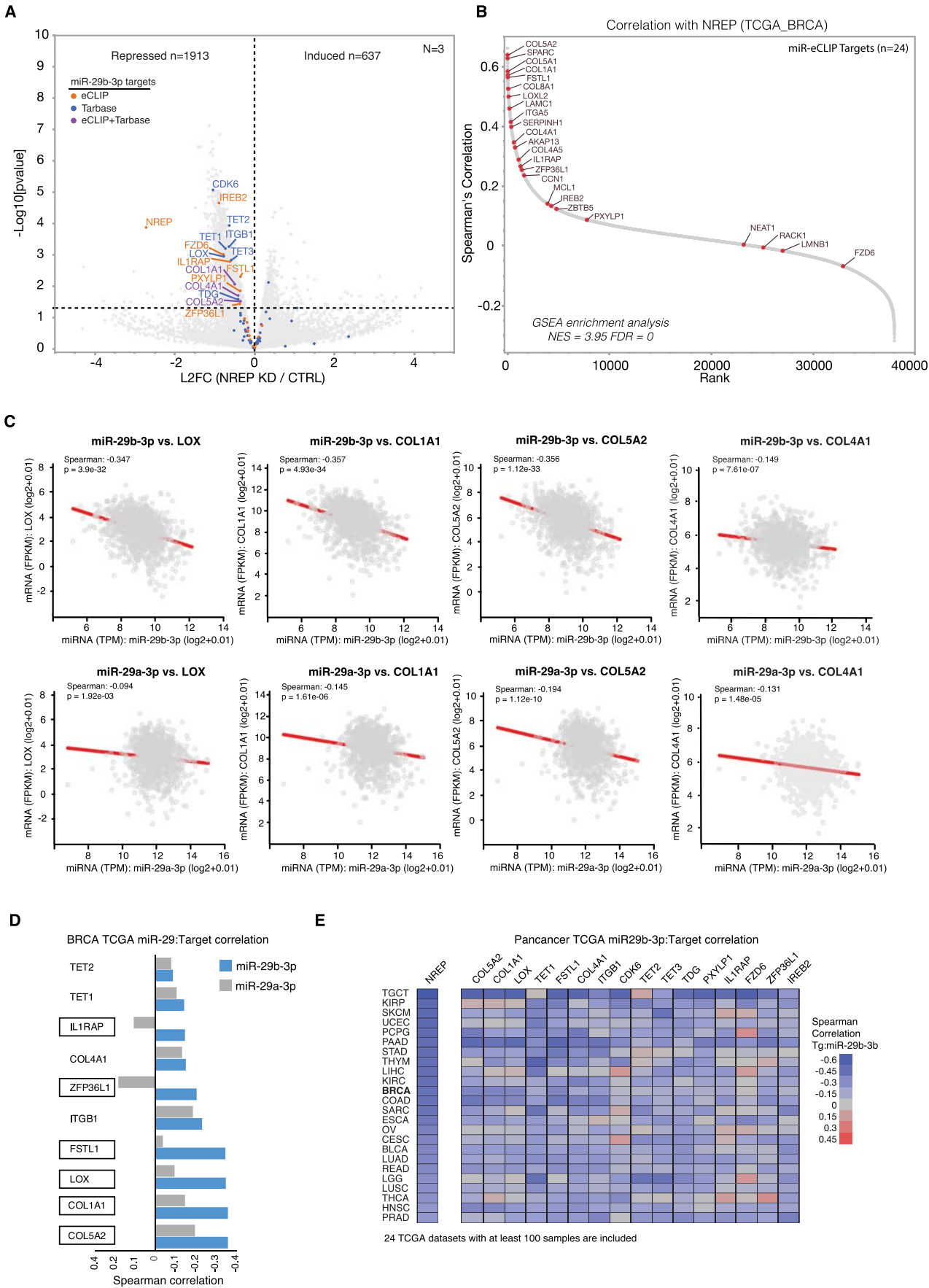
(A, B) For each miRNA found as upregulated upon ZSWIM8 KD (see Fig. 2E) is reported the distribution of reads from the 5p/3p strands, the TDMD net effect (red bar) and the values (L2FC values upon ZSWIM8-KD) of miRNA substrate (red dots) and matched miRNA (pink) obtained by sRNA-Seq analysis, together with their pri-miRNAs (green), and host genes (blue dots) obtained by RT-qPCR, normalized with either RPLP0 or GADPH (two or more dots are shown, considering the two normalizations and different pair of primers used). A dashed bar indicates the threshold for the "TDMD net effect" ( $\geq 0.3$  L2FC).  $N = 3$ . (C) Regulation of TDMD substrates identified in this study in other human cell lines (left K562; right HEK293; data from (Han et al, 2020)). The graphs show the fold change of miRNA levels assessed upon ZSWIM8 KD over the average expression of the controls. The red dots represent the breast cancer substrates that were detected ( $>10$  TPM) in the corresponding cell line; miRNAs showing an L2FC  $>0.3$  were considered as "induced" and are bolded.





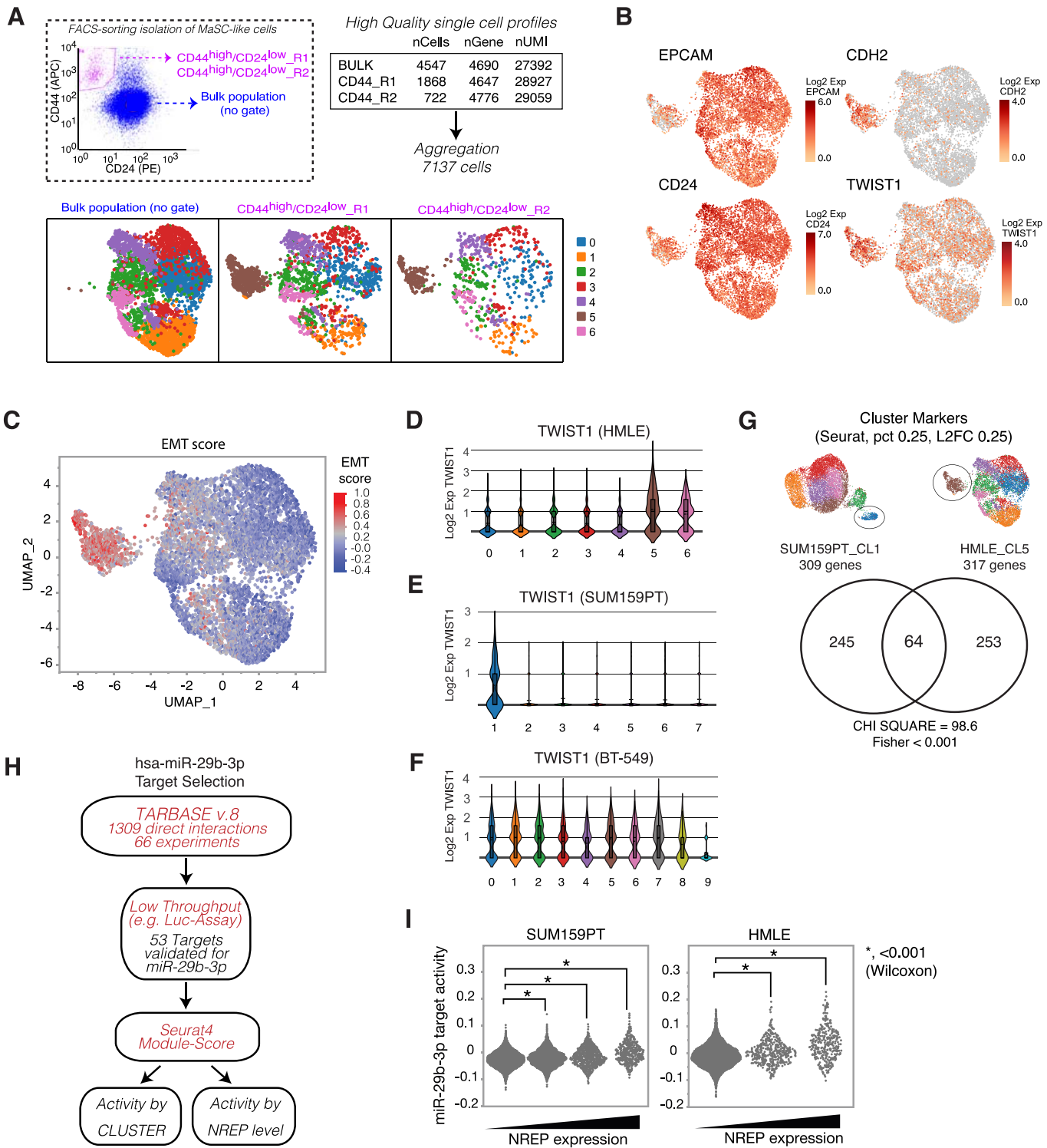
**Figure EV2. miR-eCLIP identifies endogenous miRNA:target interactions.**

(A) Metagene plot showing the average number of AGO2 peaks along the 5'UTR, CDS, and 3'UTR for peaks called in each sample (control sgRNA118 and 119; ZSWIM8 KD sgRNA34 and 42). In order to create the metagene, peaks were downsampled to match the peak set with the smallest number to account for differences in the number of significant peaks and gene lengths were normalized. (B) The AGO2 peak upset plot shows the feature distribution for reproducible peaks, as well as unique peaks for each replicate, in control (left) or ZSWIM8 KD cells (right). (C) Number of genes with chimeric peaks in each sample or reproducible across replicates (see Methods for the identification of reproducible peaks). (D) Proportion of seed match among chimeric peaks. Canonical sites are 6mer, 7mer-1A, 7mer-m8, and 8mer; noncanonical sites are seed-less: 3' compensatory or centered. (E) Bubble plot showing miRNA target occupancy in control (left) or ZSWIM8 KD (right) cells, as estimated from the cumulative count of all targets found in chimeric reads (expressed in transcript per million, TPM). MiRNAs are ranked from high to low occupancy. The dimension of the bubbles is proportional to the number of targets. (F) Chimeric peaks for hsa-miR-7-5p and LINC00632. All the peaks localize on the CDR1as region. (G) Bubble plot reporting the differential miRNA activity (ZSWIM8 KD vs Ctrl), calculated by a ratio of the values reported in panel E. The dimension of the bubble is proportional to the max miRNA target occupancy observed either in control or ZSWIM8 KD cells. (H-J) MiRNAs with increased target occupancy by eCLIP and low expression by sRNA-seq were re-analyzed by RT-qPCR in control and ZSWIM8 KD cells. MiR-1973 was not detected even by RT-qPCR. (I) Bar charts reporting the expression levels by RT-qPCR of mature miRNAs measured upon 0.5 µg/ml doxycycline (+) for seven cell doublings on cells infected with controls and ZSWIM8 targeting sgRNAs (average and SEM;  $N = 3$ ). Data were normalized over *SNORD48*. Statistical analysis with the Student's *t*-test; \* $p \leq 0.05$ , \*\*\* $p \leq 0.001$ . (J) Bar charts showing the expression levels of pri-miRNAs and host genes by RT-qPCR, as in (I). Data were normalized over *RPLP0* (average and SEM;  $N = 3$ ). Statistical analysis with the Student's *t*-test. (K) Chimeric peak for hsa-miR-503-5p and MALAT1. The peak contains a region of MALAT1 displaying an extended complementarity to miR-503-5p, as shown. (L) sRNA-seq data for miR-503-5p were re-analyzed in all BC samples (as in Fig. 1), including those with miR-503-5p expression <10 TPMs, and are reported as average L2FC between control and ZSWIM8 KD cells (see also Dataset EV1).



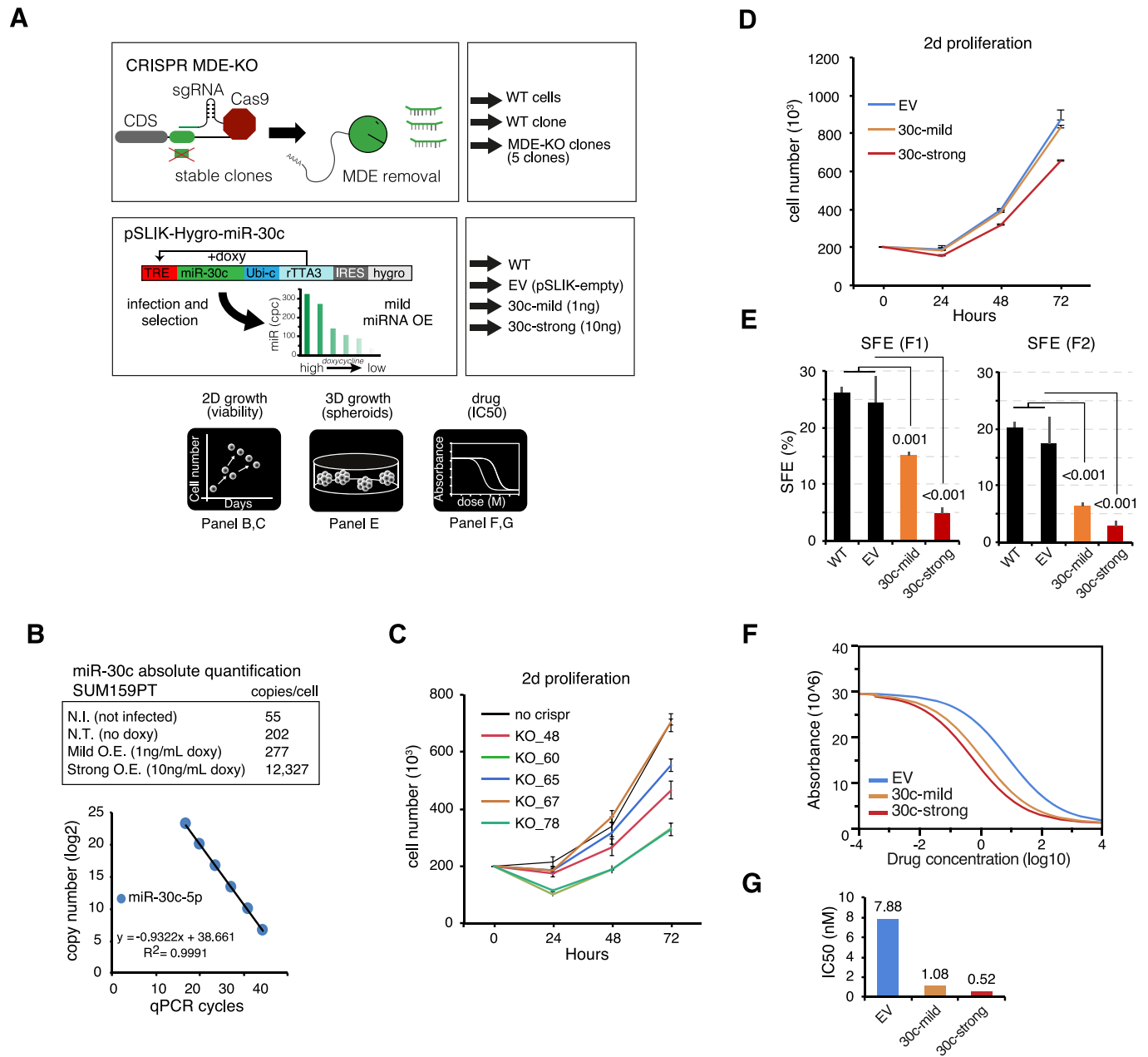
◀ **Figure EV3. NREP perturbation increases miR-29b-3p activity.**

(A) Volcano plot showing differentially regulated genes ( $p$  value  $< 0.05$ ; NREP KD vs CTRL,  $N = 3$ ; Deseq2) upon NREP perturbation in SUM159PT cells. Highlighted are miR-29b-3p targets identified by miR-eCLIP or from Tarbase. (B) Spearman pair-wise correlation analysis between the expression level of NREP and all other genes ( $n = 38,005$ ) was calculated in the TCGA BRCA dataset. Genes are ranked according to the correlation with NREP (high to low). MiR-29b-3p targets, defined from miR-eCLIP chimeric peaks, are highlighted in red. The targets were used as a gene set to run gene set enrichment analysis (GSEA). The observed normalized enrichment score (NES = 3.95) indicates a strong positive association. (C) Top: Spearman pair-wise correlation analysis between the expression level of miR-29b-3p (x-axis, data in TPM) and selected target genes (y-axis, data in FPKM) calculated using the TCGA BRCA dataset. Bottom: the same analysis was performed with miR-29a-3p. (D) Bar chart summarizing the Spearman coefficient between miR-29b-3p (blue) or miR-29a-3p (gray) and NREP-dependent miR-29b-3p targets (i.e., top 10 from Fig. 4G). A box indicates genes showing stronger correlation with miR-29b-3p as compared to miR-29a-3p. (E) Heatmap showing the Spearman correlation of miR-29a-3p with NREP-dependent miR-29b-3p targets across 24 TCGA datasets (Breast cancer, BRCA, highlighted in bold). NREP is shown as a reference.



◀ **Figure EV4. NREP controls miR-29b-3p levels and activity in CD44<sup>high</sup>/CD24<sup>low</sup> HMLE cells.**

(A) Single-cell analysis of stem-like cells (maSC-like) in HMLE. Top: schematic workflow, with representative FACS profile of CD44<sup>high</sup>/CD24<sup>low</sup> cells and summary of cells/features analyzed. Bottom: UMAP visualization of the aggregate dataset (7137 cells), grouped by samples and colored by clusters. Clusters were identified using standard methods (see Methods) based on transcriptional similarity within each sample; cluster numbers are arbitrary and not comparable across different cell lines. (B) UMAP plot of HMLE cells, colored by the expression of typical epithelial (EPCAM, CD24) or mesenchymal (CDH2 and TWIST1) markers. (C) UMAP plot colored by the EMT module score (see Methods). (D, F) Violin plots of TWIST1 expression across HMLE clusters (D), SUM159PT clusters (E), and BT-549 clusters (F). (G) Overlap between the mesenchymal clusters in SUM159PT and HMLE cell lines based on cluster markers. (H) Strategy for the definition of miR-29b-3p targets activity in single-cell gene expression profiles. (I) Activity of miR-29b-3p targets stratified according to different bins of NREP expression (bins are defined as mean, +1 SD, +2 SD, +3 SD; the first bin contains all cells with zero values). Statistical analysis by the Wilcoxon test, comparing each cluster versus all the others. \* $p < 0.001$ .



**Figure EV5. Phenotypes related to SERPINE1:miR-30c-5p interaction.**

(A) Schematic representation of the different methodological strategies used to investigate the phenotypes that depend on the SERPINE1:miR-30c-5p interaction. (B) Upper panel: a table summarizes miR-30c-5p levels by absolute quantification in cells infected with an inducible construct for miR-30c-5p overexpression. NI not infected cells, NT infected cells not treated with doxycycline. Lower panel: titration curve used for the absolute quantification of miR-30c-5p (see Methods). (C) Growth curve of SUM159PT clones bearing the MDE deletion in SERPINE1 3'UTR.  $N = 3$  (average cell number and SEM). (D) Growth curve of SUM159PT cells upon mild or strong miR-30c-5p overexpression.  $N = 3$  (average cell number and SEM). (E) Bar charts reporting the results of the sphere formation assay performed in SUM159PT. Shown is the sphere forming efficiency (average % and SEM from three replicate wells) measured at the first (F1) and at the second generation (F2).  $N = 2$ .  $P$  values by Student's  $T$ -test. (F, G) Cell viability was measured by ADP-Glo Max Assay in SUM159PT cells treated for 3 days with Paclitaxel at different concentrations (panel F) and used to calculate IC50 values (panel G). Results shown are from one of two independent experiments.