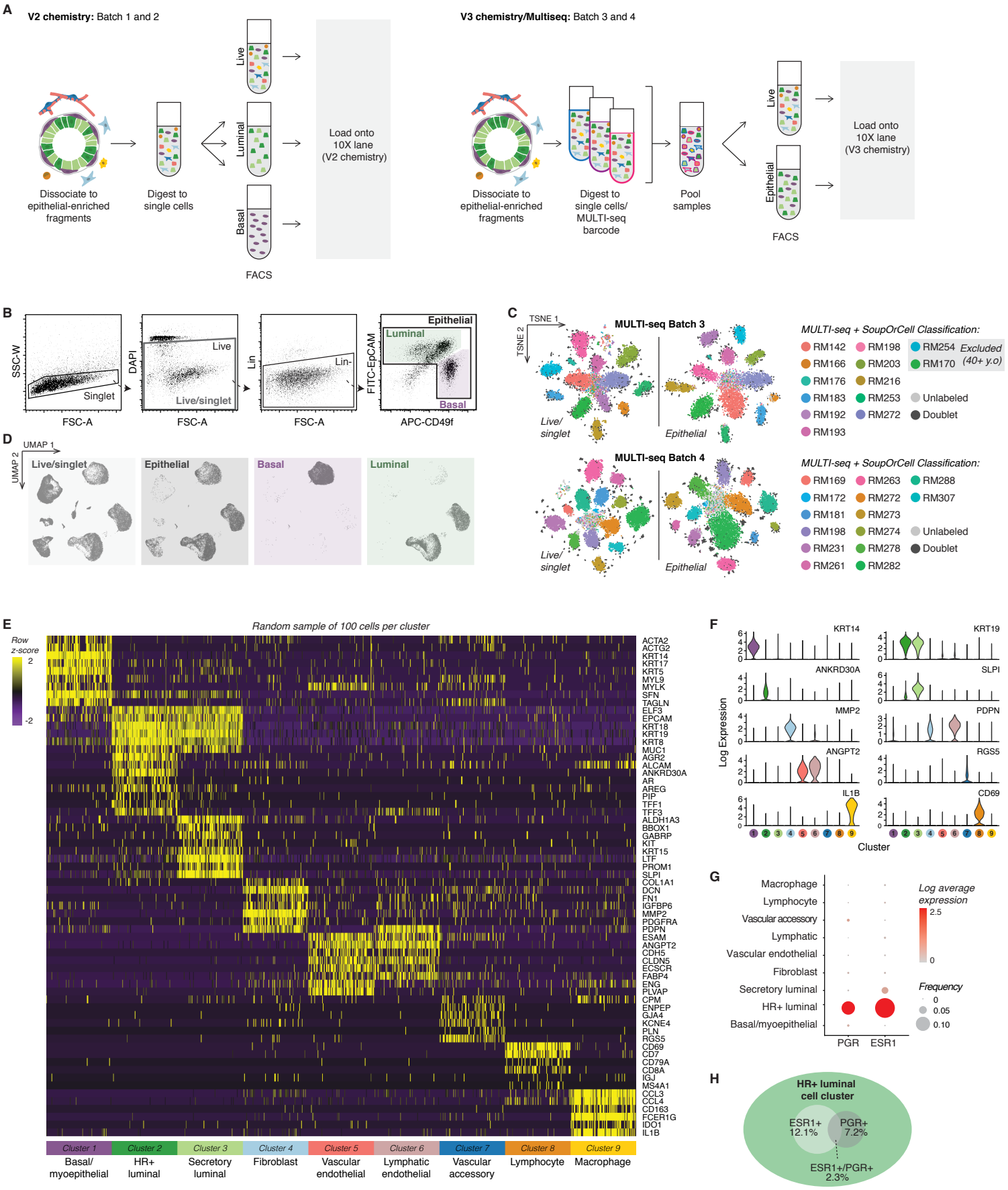

Figure S1 - Overview of scRNAseq workflow and basic clustering results

Figure S1. Overview of scRNA-seq workflow and basic clustering results; Related to Figure 1

(A) Overview of sample multiplexing and sorting strategies across batches (also see Table S2).

(B) FACS plots depicting sort gates used for sequencing.

(C) TSNE dimensionality reduction of the normalized barcode count matrices and final sample classification for MULTI-seq experiments (Batches 3 and 4, also see *STAR Methods*).

(D) UMAP dimensionality reduction of the combined data from twenty-eight reduction mammoplasty samples (GSE198732) for each sort population.

(E) Heatmap highlighting marker genes used to identify each cell type, colored by row z-score (linear scale). For visualization purposes, we randomly selected 100 cells from each cluster.

(F) Violin plot highlighting the log expression of selected marker genes in each cluster.

(G) Dot plot depicting the log normalized average and frequency of ESR1 and PGR expression across cell type clusters.

(H) Euler diagram highlighting the frequency of ESR1 and PGR expression and percent overlap in the HR+ luminal cell cluster.

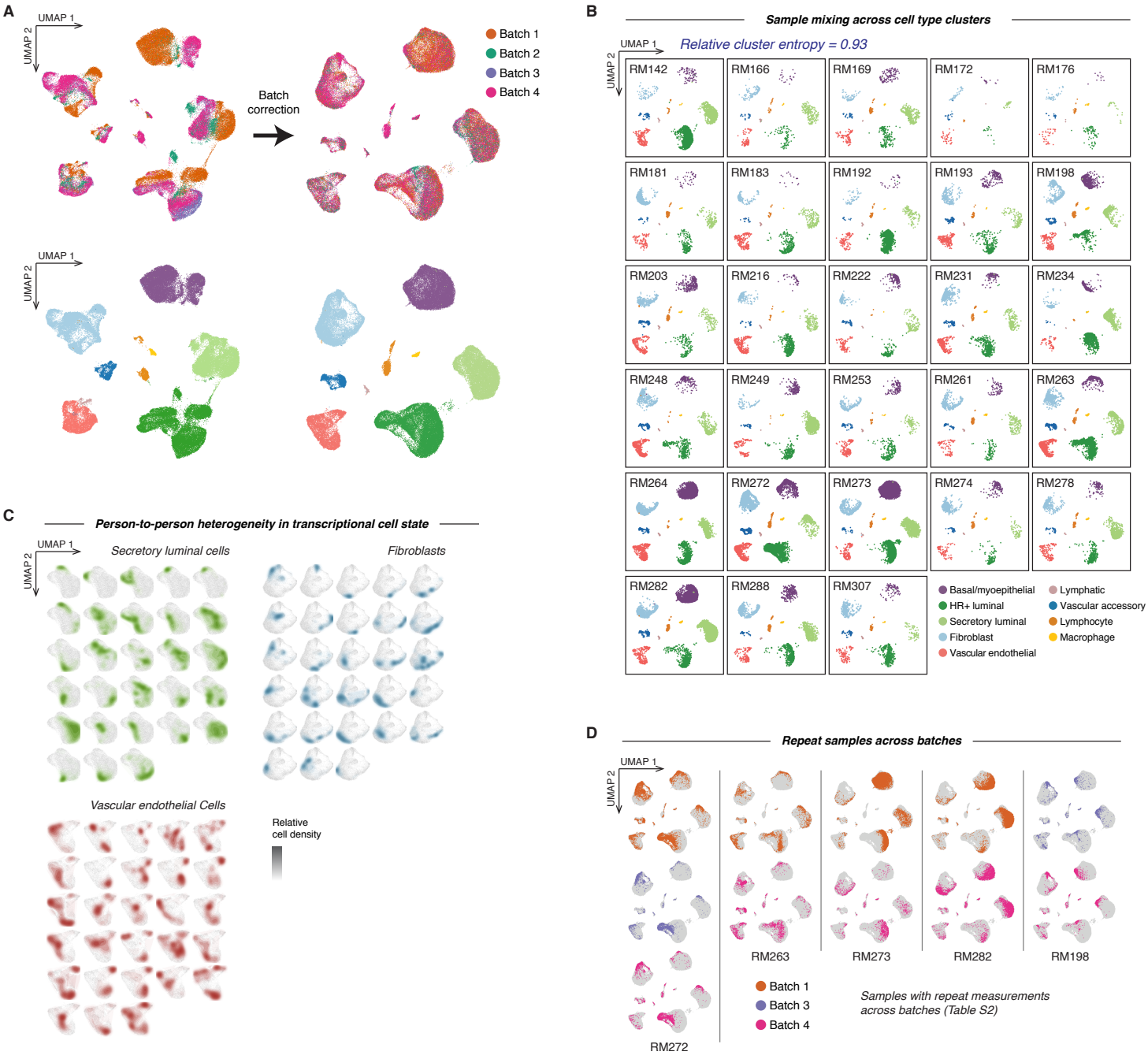

Figure S2 - Transcriptional variation between samples is independent of batch effects

Figure S2. Transcriptional variation between samples is independent of batch effect; Related to Figure 1

(A) UMAP dimensionality reduction of the combined data from twenty-eight reduction mammoplasty samples (GSE198732) for each sort population before (left) and after (right) batch correction.

(B) UMAP for each reduction mammoplasty sample highlighting cell types identified by unsupervised clustering. Cells from different individuals are well-mixed across all clusters (cluster entropy = 0.93, *STAR Methods*).

(C) Density plots (arbitrary units, linear scale) highlighting the transcriptional cell state of the indicated cell types from each sample.

(D) UMAP of reduction mammoplasty samples that were run as repeat measurements across multiple batches, highlighting cells from each batch. See Table S2 for sample and batch information.

(E) Quantification of the pairwise alignment—or “similarity score”—between cells from the same or different sample and within versus across batches for the indicated cell types. See Table S2 for sample and batch information. The dashed line represents the expected similarity score for random mixing.

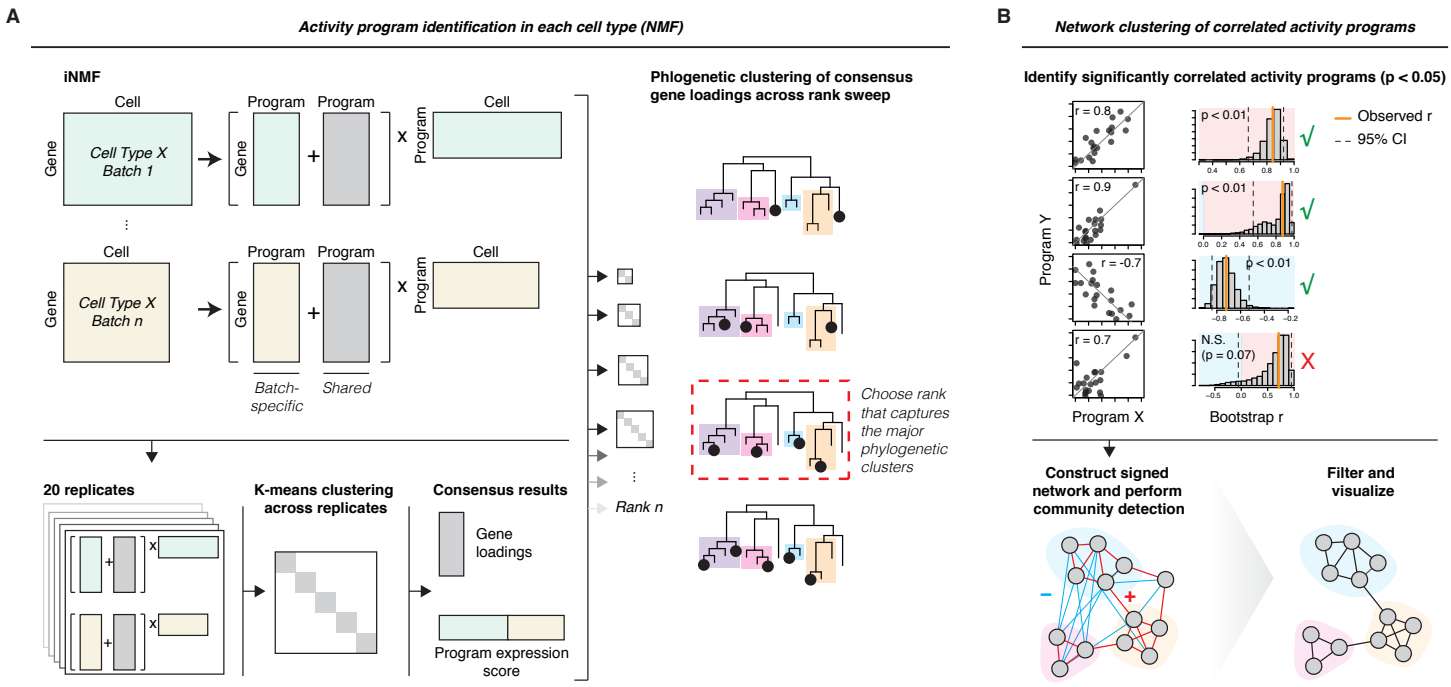

Figure S3 - Overview of DECIPHER-seq workflow

Figure S3. Overview of DECIPHER-seq workflow; Related to Figure 2

(A) To identify gene activity programs in the breast, we performed non-negative matrix factorization (NMF) on each of the major cell types in our dataset, using integrative NMF (iNMF) to account for batch differences. As NMF solutions are non-unique, we adapted a consensus matrix factorization approach to identify activity programs that were consistent across multiple replicates. To optimize rank *K*, we used phylogenetic clustering of consensus gene loadings across a range of *K* values and chose the point at which increasing granularity did not identify major new subtrees.

(B) To build a network map of cell-cell interactions, we quantified the average activity program expression for each sample and constructed a weighted network based on the pair-wise Pearson correlations *r*. The resulting correlation matrix was transformed into a signed weighted adjacency matrix by using bootstrap resampling to estimate confidence intervals associated with each correlation, and setting all correlations with p-values > 0.05 to zero. Finally, we identified modules of highly correlated gene expression programs using a community detection algorithm, and filtered out isolated links and modules using weighted topological overlap.

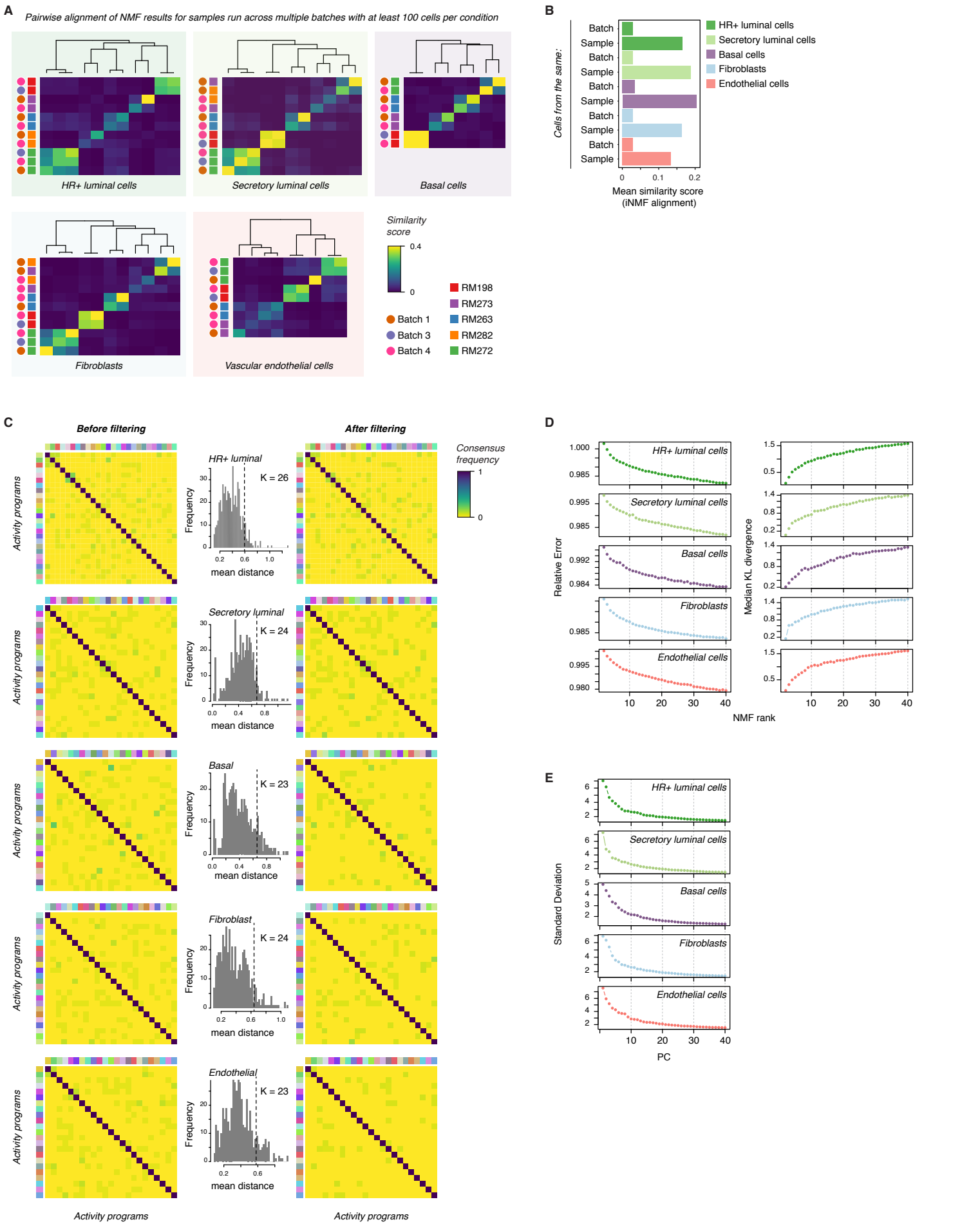

Figure S4 - NMF integration, consensus clustering of replicates, and metrics for choice of rank sweep

Figure S4. NMF integration, consensus clustering of replicates, and metrics for choice of rank sweep; Related to Figure 2

(A) Heatmaps showing the pairwise alignment ("similarity score"; fraction of nearest neighbors from the corresponding sample/batch, linear units) of NMF results for samples run across multiple batches. Hierarchical clustering (Ward D2 linkage) groups cells by sample rather than batch.

(B) Quantification of the mean pairwise alignment—or "similarity score"—of iNMF results for cells from the same sample but different batch ("Sample") or different sample but same batch ("Batch") for the indicated cell types. See Table S2 for sample and batch information.

(C) Consensus matrix (frequency, linear units) showing the clustered NMF components for the indicated values of K, combined across 20 replicates, before (left) and after filtering (right). The histogram shows the mean distance of each component to its 6 nearest neighbors with a dashed line showing the threshold for filtering outliers (see *STAR Methods*).

(D) Parameter selection for *K* sweep range. Plots depict the relative Frobenius reconstruction error and median KL divergence for consensus results at each value of *K*.

(E) Parameter selection for *K* sweep range. Plot depicting the standard deviation of the first forty principal components for each cell type.

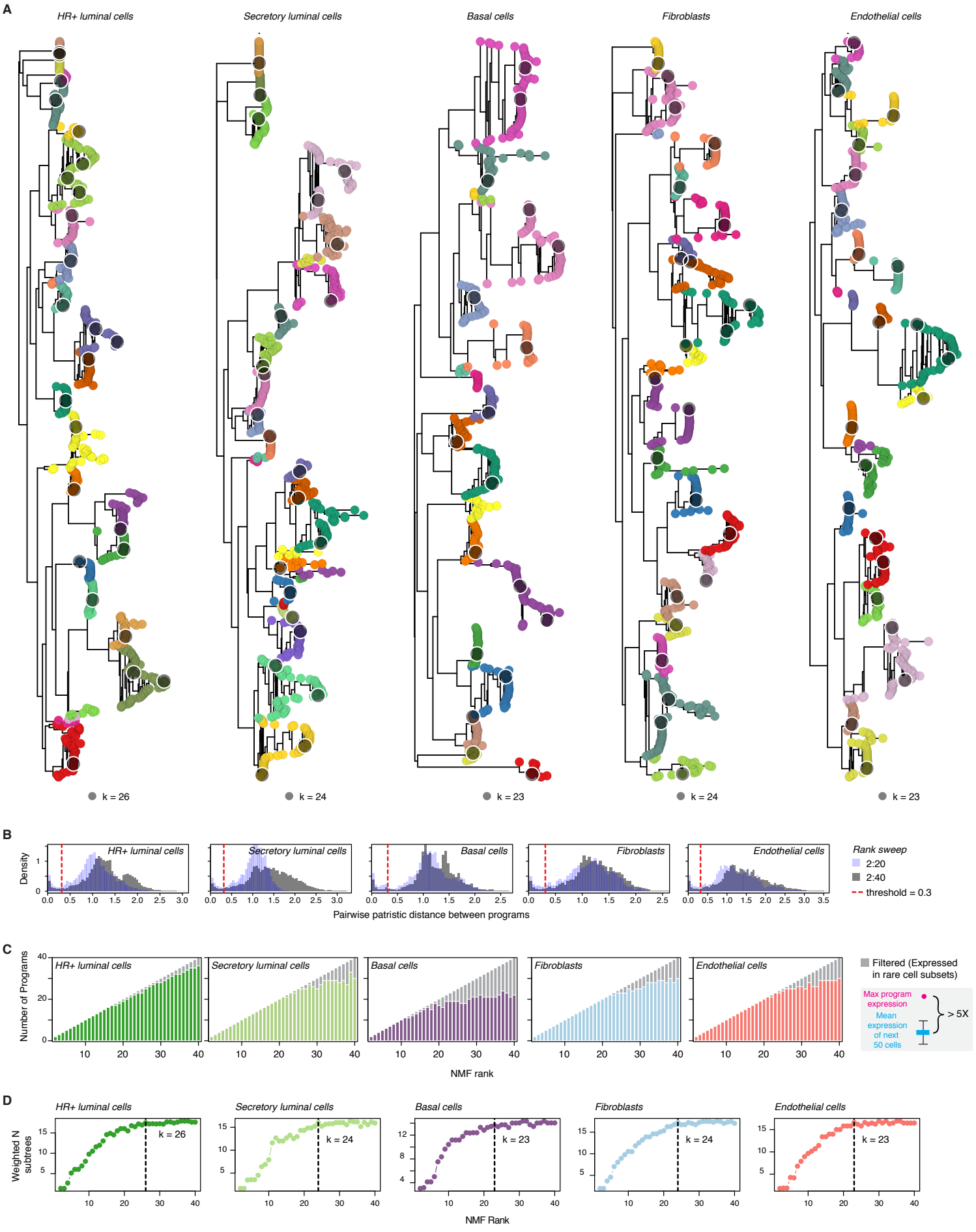

Figure S5 - Phylogenetic clustering of gene programs across rank sweep and final choice of rank

Figure S5. Phylogenetic clustering of gene programs across rank sweep and final choice of rank; Related to Figure 2

(A) Phylogenetic trees of consensus iNMF components ("activity programs") for *K* values ranging from 2 to 40, based on a balanced minimum evolution algorithm. Each tree was partitioned into distinct clusters using a depth-first search to identify sub-trees with median patristic distance below a threshold of 0.3 (see below).

(B) Histogram of each phylogenetic tree's patristic distance distribution, for phylogenetic trees constructed from *K* sweep values ranging from 2 to 40 (grey) or 2 to 20 (blue). The vertical red line shows the distance threshold used in this study of 0.3. We chose an absolute rather than relative (e.g. percentile) distance threshold as it was robust to different sweep ranges of *K* (e.g. the blue versus grey histograms).

(C) Plot depicting the number of "outlier" activity programs filtered at each value of *K*. Outliers were defined as activity programs where the maximum expression score of an activity program across all cells was more than 5 times greater than the mean expression score of that activity program in the next 50 highest-scoring cells.

(D) Plots depicting the number of subtrees identified at each *K*, weighted by the total number of programs in each subtree (see *STAR Methods)*. Dashed lines depict the chosen values for *K* (K_{opt}). We chose K_{opt} as the saturation point representing the point at which increasing the granularity of matrix factorization did not identify activity programs that comprise major new subtrees.

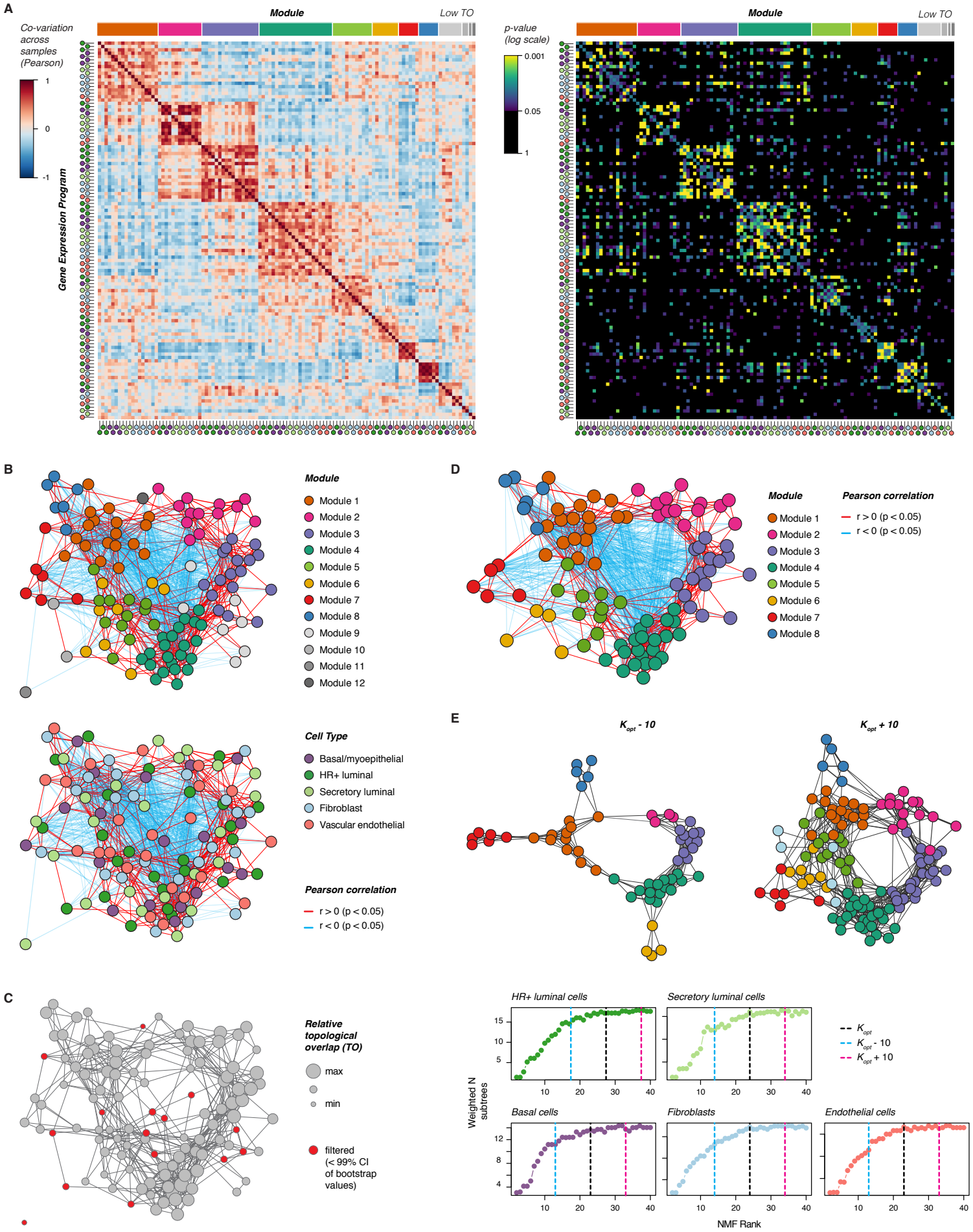

Figure S6 - Network clustering of correlated activity programs

Figure S6. Network clustering of correlated activity programs; Related to Figure 2

(A) *Left:* Heatmap depicting Pearson correlation coefficients (*r*) between activity programs in all identified modules, including filtered modules with low topological overlap. Modules were identified using a Constant Potts Model for community detection. *Right:* Heatmap depicting p-values for the Pearson correlation between activity programs in all identified modules (log scale). To estimate p-values, we used bootstrap resampling to determine confidence intervals for each correlation coefficient, and calculated p-values based on the null hypothesis that $r = 0$.

(B) Network graph of correlated activity programs in the human breast, including modules with low topological overlap, and depicting both positive and negative edges. Nodes represent distinct activity programs in the indicated cell types, and edges connect significantly correlated programs (red: Pearson correlation coefficient > 0, p-value < 0.05; blue: Pearson correlation coefficient < 0, p-value < 0.05). Modules were identified using a Constant Potts Model for community detection.

(C) Network graph of correlated activity programs in the human breast, with node sizes proportional to the relative weighted topological overlap (wTO) of each node with other nodes in the same module (arbitrary units, linear scale). Filtered activity programs with a wTO within the 99% confidence interval for nodes samples from “random” modules of the same size are highlighted in red.

(D) Network graph of correlated activity programs in the human breast, excluding modules with low topological overlap, and depicting both positive and negative edges. Nodes represent distinct activity programs in the indicated cell types, and edges connect significantly correlated programs (red: Pearson correlation coefficient > 0, p-value < 0.05; blue: Pearson correlation coefficient < 0, p-value < 0.05).

(E) *Top:* Network graph of correlated activity programs in the human breast, excluding modules with low topological overlap, for values of *K* at (K_{opt} - 10) and (K_{opt} + 10). Nodes represent distinct activity programs in the indicated cell types, and edges connect significantly correlated programs (Pearson correlation coefficient > 0, p-value < 0.05). Modules of correlated programs were identified using a Constant Potts Model for community detection. *Bottom:* Plots depicting the number of subtrees identified at each *K*, weighted by the total number of programs in each subtree (see *STAR Methods)*. Dashed lines depict the chosen values for *K* (K_{opt}), (K_{opt} - 10), and (K_{opt} + 10) for each cell type.

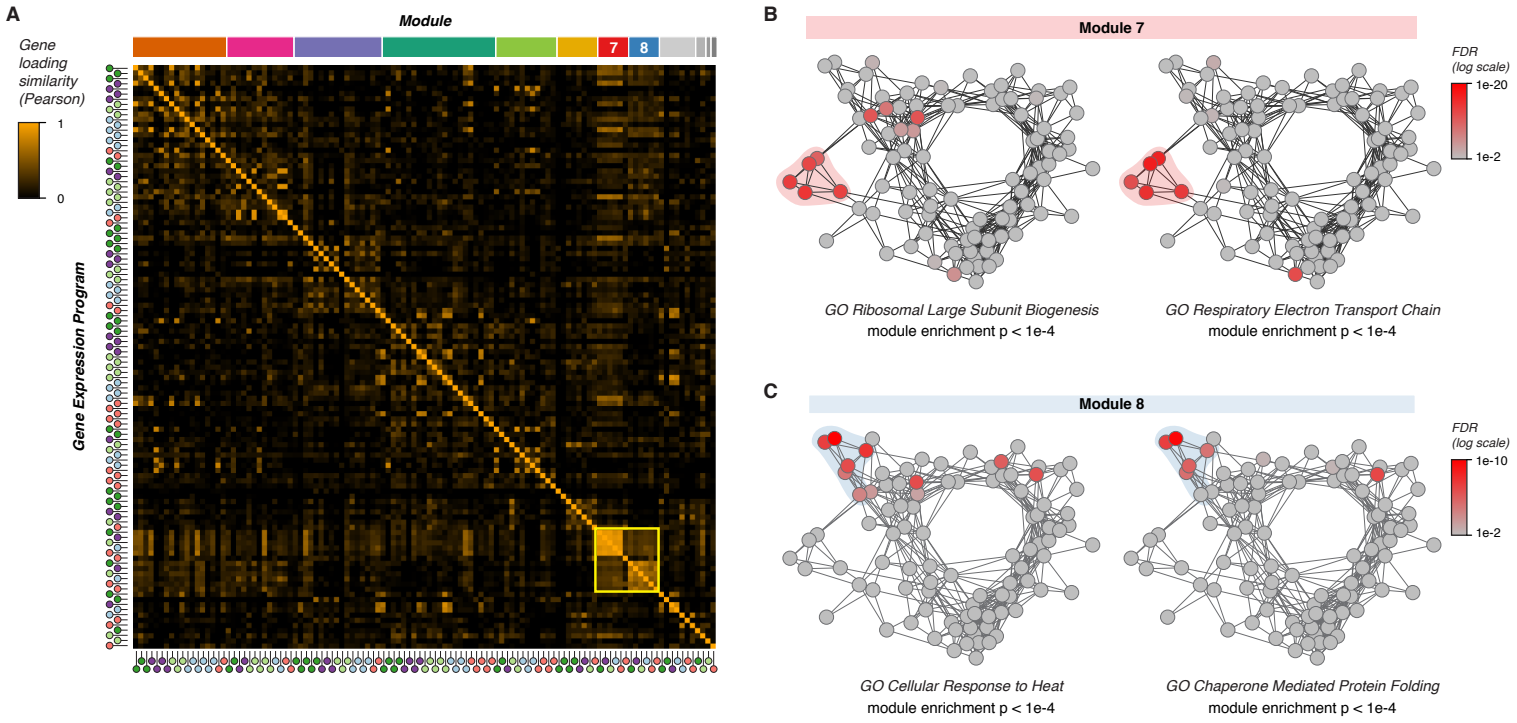

Figure S7 - Non-cell-type-specific shared transcriptional responses across cell types in the human breast

Figure S7. Non-cell-type-specific shared transcriptional responses across cell types in the human breast; Related to Figure 2

(A) Heatmap depicting Pearson correlation coefficients between gene loadings for all activity programs. Activity programs in modules 7 and 8 are highlighted by a yellow box.

(B) Network graph of activity programs, colored by the FDR for enrichment of the indicated gene sets in each activity program (log scale). Overall enrichment of gene sets within module 7 was determined by permutation analysis.

(C) Network graph of activity programs, colored by the FDR for enrichment of the indicated gene sets in each activity program (log scale). Overall enrichment of gene sets within module 8 was determined by permutation analysis.

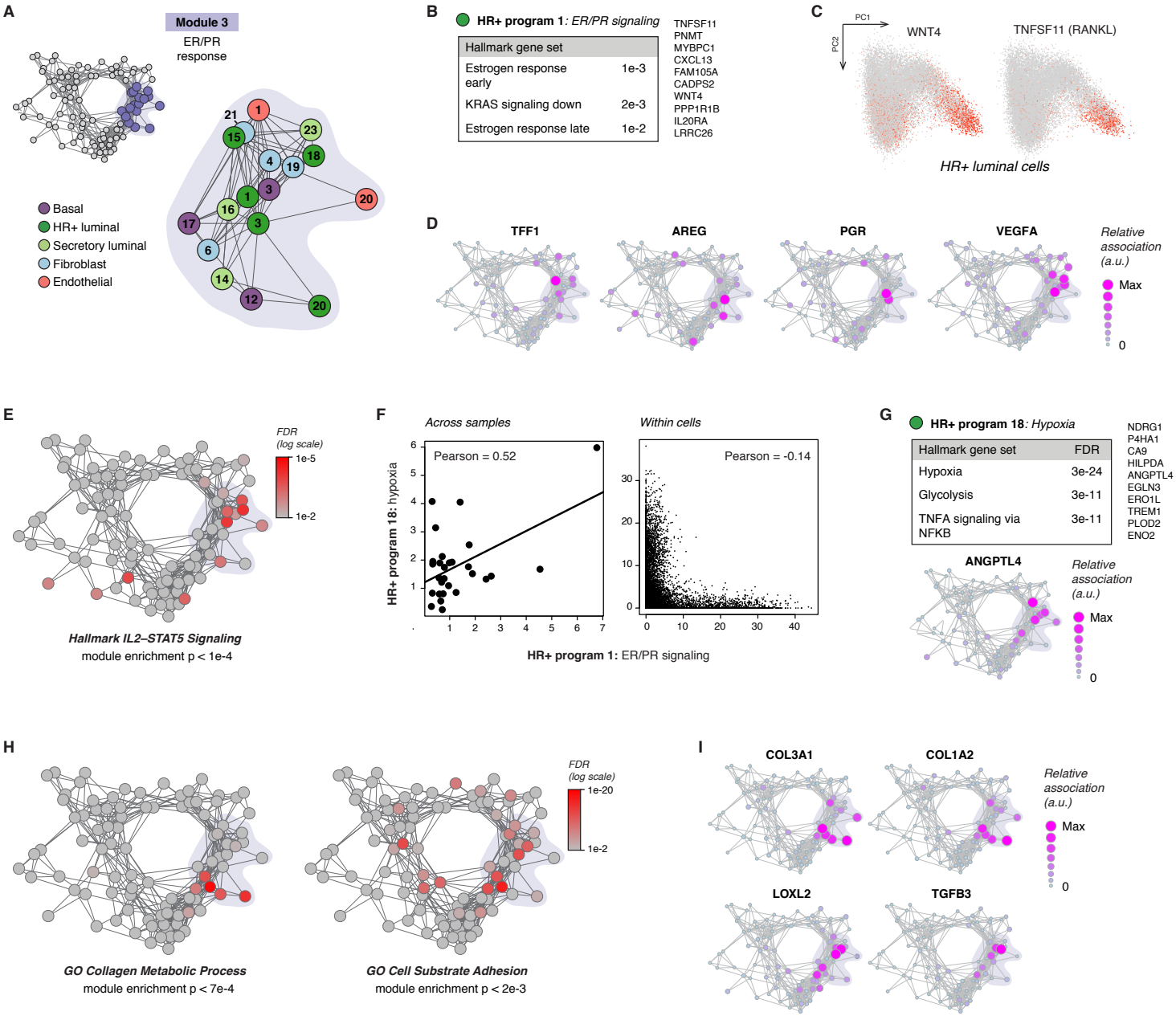

Figure S8 - The “ER/PR response” module

Figure S8. The "ER/PR response" module; Related to Figure 3

(A) Diagram highlighting activity programs in the "ER/PR response" module.

(B) Gene set enrichment analysis for HR+ activity program 1 ("ER/PR signaling"), showing the top pathways identified from the Molecular Signatures Database Hallmark gene sets, and the top 10 gene loadings in HR+ program 1.

(C) PCA plot of HR+ luminal cells depicting expression of WNT4 and TNFSF11 (RANKL) in log normalized counts.

(D) Network graph of activity programs, depicting the relative association of the indicated hormone-responsive genes with each activity program (arbitrary units, linear scale).

(E) Network graph of activity programs, colored by the FDR for enrichment of the "IL2-STAT5 signaling" Hallmark gene set in each activity program (log scale). Overall enrichment within module 3 was determined by permutation analysis.

(F) *Left*: Plot of the average expression score for HR+ activity program 1 ("ER/PR signaling") versus activity program 18 ("hypoxia") (Pearson $r = 0.52$). Dots represent the average expression score of each gene program within a sample. *Right:* Scatter plot of the per-cell average expression score for HR+ activity program 1 ("ER/PR signaling") versus activity program 18 ("hypoxia") (Pearson $r = -0.14$). Dots represent the expression score of each activity program within individual HR+ luminal cells.

(G) *Top:* Gene set enrichment analysis for HR+ activity program 18 ("hypoxia"), showing the top pathways identified from the Molecular Signatures Database Hallmark gene sets, and the top 10 gene loadings in HR+ program 18. *Bottom:* Network graph of activity programs, depicting the relative association of the hypoxia-related gene ANGPTL4 with each activity program (arbitrary units, linear scale).

(H) Network graph of activity programs, colored by the FDR for enrichment of the indicated Molecular Signatures Database gene sets in each activity program (log scale). Overall enrichment of gene sets within module 3 was determined by permutation analysis.

(I) Network graph of activity programs, depicting the relative association of the indicated marker genes with each activity program (arbitrary units, linear scale).

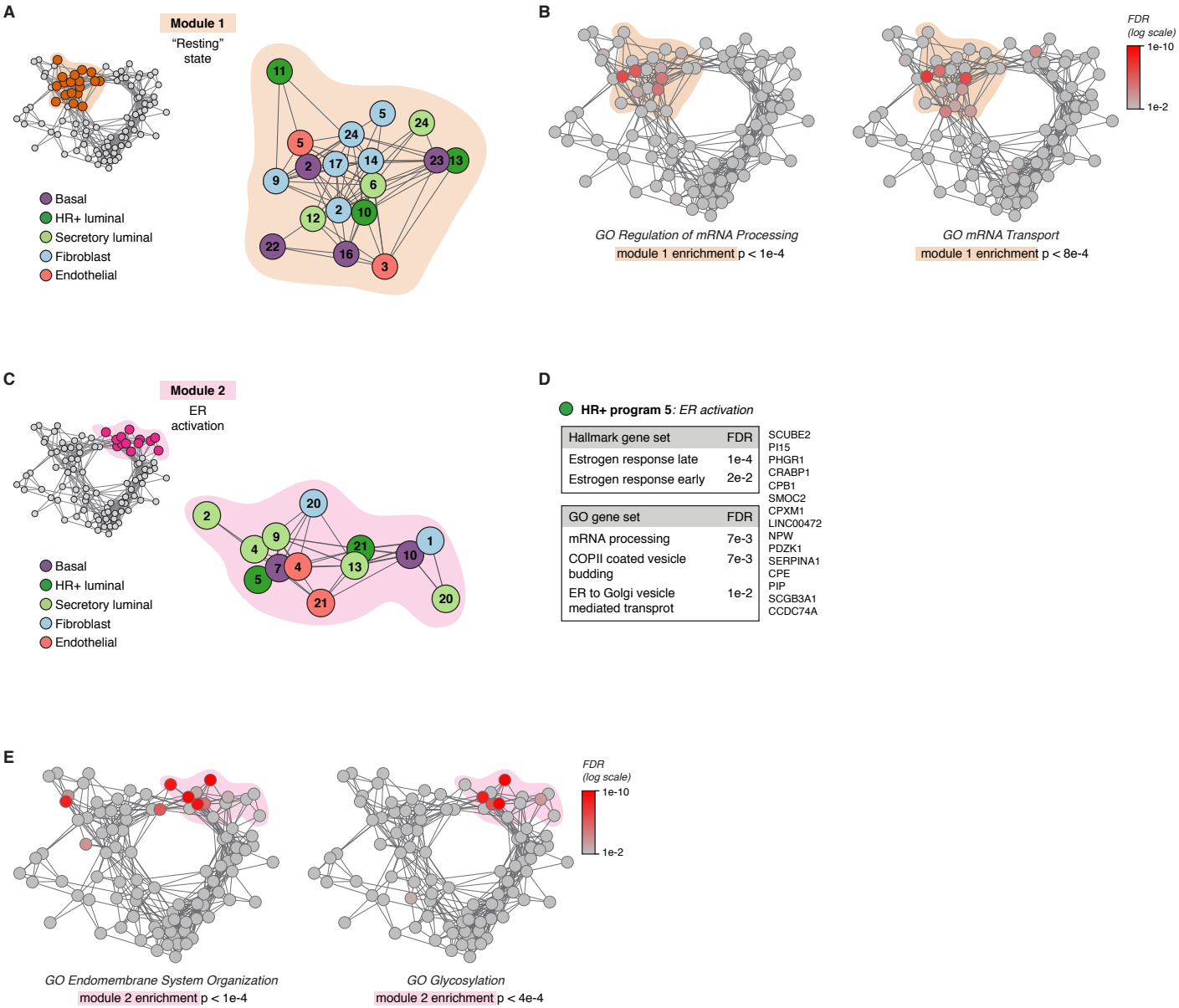

Figure S9 - The "Resting state" and "Estrogen receptor activation" modules

Figure S9. The “Resting state” and “Estrogen receptor (ER) activation” modules; Related to Figure 4

(A) Diagram highlighting activity programs in the “Resting state” module.

(B) Network graph of activity programs, colored by the FDR for enrichment of the indicated Molecular Signatures Database gene sets in each activity program (log scale). Overall enrichment of gene sets within module 1 was determined by permutation analysis.

(C) Diagram highlighting activity programs in the “ER activation” module.

(D) Gene set enrichment analysis for HR+ activity program 5 (“ER activation”), showing the top pathways identified from the Molecular Signatures Database Hallmark and GO Biological Process gene sets, and the top 15 gene loadings in HR+ program 5.

(E) Network graph of activity programs, colored by the FDR for enrichment of the indicated Molecular Signatures Database gene sets in each activity program (log scale). Overall enrichment of gene sets within module 2 was determined by permutation analysis.

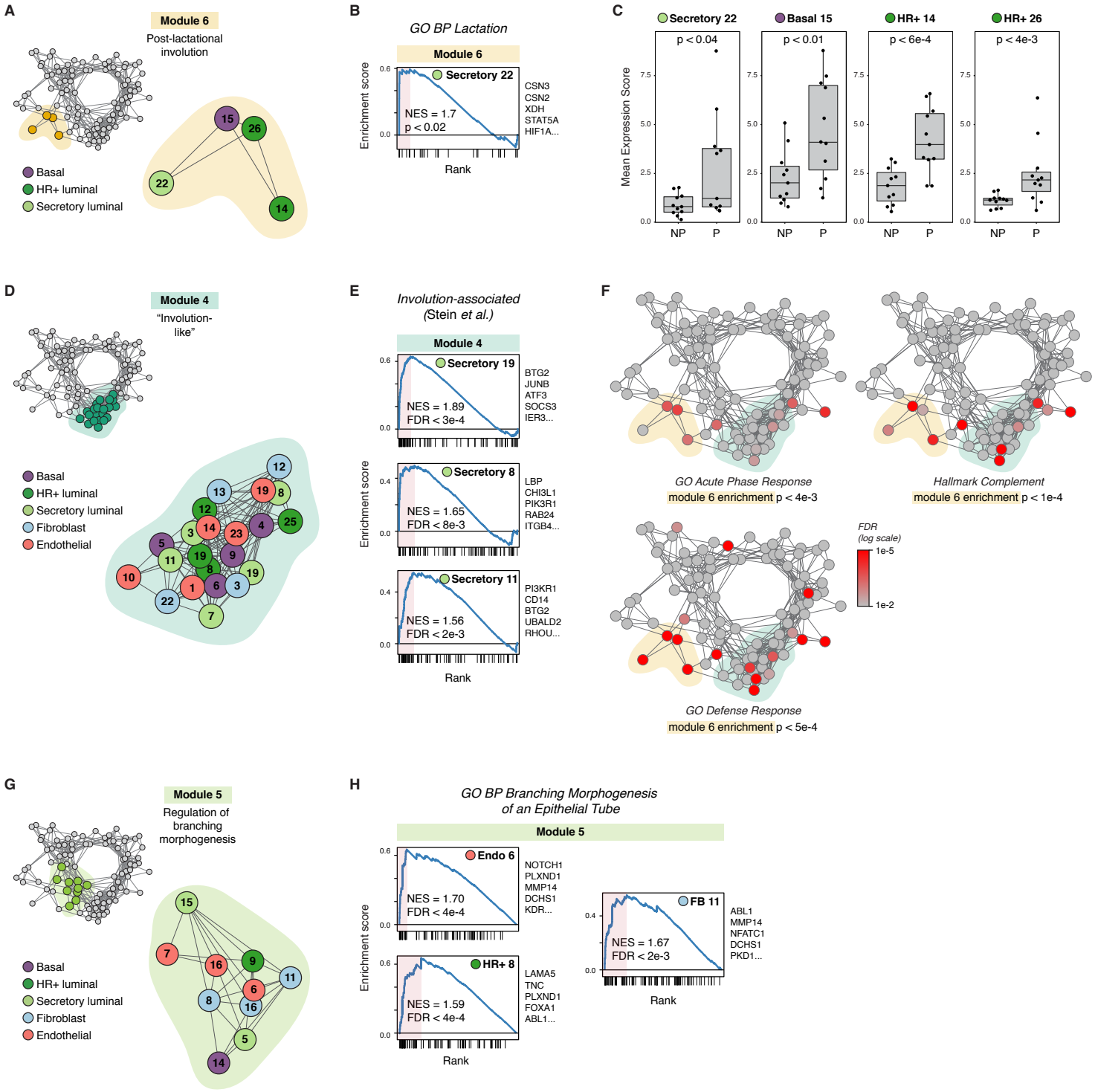

Figure S10 - The "Post-lactational involution", "Involution-like", and "Regulation of branching morphogenesis" modules

Figure S10. The "Post-lactational involution", "Involution-like", and "Regulation of branching morphogenesis" modules; Related to Figure 4

(A) Diagram highlighting activity programs in the "Post-lactational involution" module.

(B) Gene set enrichment analysis of Secretory cell activity program 22 in the "Post-lactational involution" module, showing enrichment of genes in the Molecular Signatures Database GO "Lactation" gene set. The top five leading edge genes are listed.

(C) Quantification of the mean expression score for the indicated activity programs for nulliparous (NP) versus parous (P) samples ($n = 22$ samples, $p < 0.05$, Mann-Whitney test).

(D) Diagram highlighting activity programs in the "Involution-like" module.

(E) Gene set enrichment analysis of the indicated activity programs in the "Involution-like" module, showing enrichment of genes previously shown to be upregulated during post-lactational involution in the mouse (Stein et al., 2004). The top five leading edge genes for each activity program are listed.

(F) Network graph of activity programs, colored by the FDR for enrichment of the indicated Molecular Signatures Database gene sets in each activity program (log scale). Overall enrichment of gene sets within module 6 was determined by permutation analysis.

(G) Diagram highlighting activity programs in the "Regulation of branching morphogenesis" module.

(H) Gene set enrichment analysis of the indicated activity programs in the "Regulation of branching morphogenesis" module, showing enrichment of the Molecular Signatures Database gene set "Branching Morphogenesis of an Epithelial Tube". The top five leading edge genes for each activity program are listed.

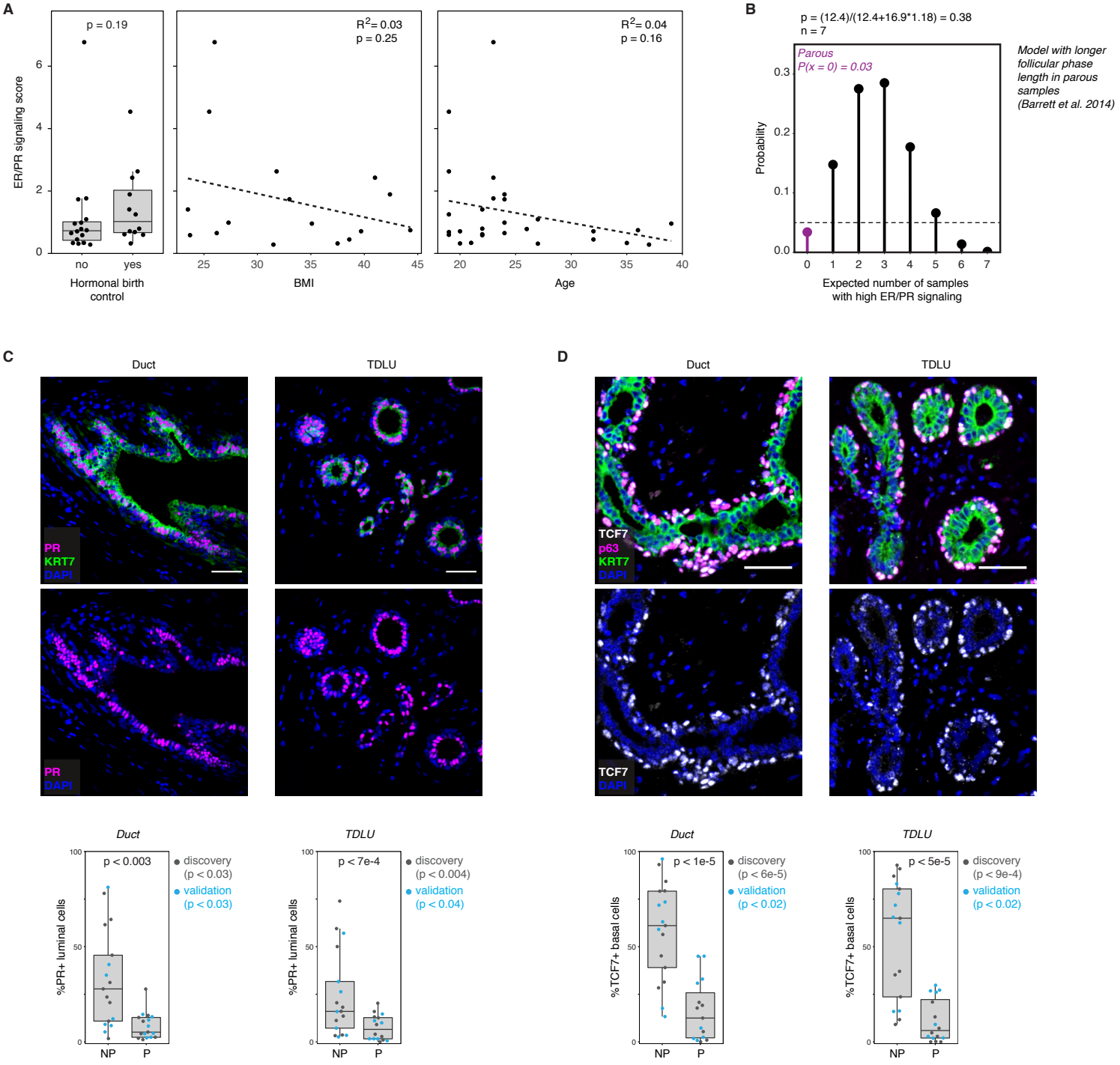

Figure S11 - Reduced ER/PR signaling in parous samples

Figure S11. Reduced ER/PR signaling in parous samples; Related to Figure 5

(A) Quantification of the average expression score of HR+ program 1 (ER/PR signaling) across the indicated biological variables (Hormonal contraceptive use: n = 28 samples, p = 0.19, Mann-Whitney test; BMI: n = 16, R^2 = 0.01, p = 0.25, Wald test; Age: n = 28 samples, R^2 = 0.04, p = 0.16, Wald test).

(B) Binomial probability distribution for the expected number of parous samples with high ER/PR signaling. The binomial probability of high ER/PR signaling was modeled as the average length of the luteal phase of the menstrual cycle in parous women, in days, divided by the average total length of the menstrual cycle in parous women (P = 0.03) (Barrett et al., 2014).

(C) Immunostaining for PR and KRT7 in ducts and terminal ductal lobular units (TDLUs), and quantification of the percentage of PR+ cells within the KRT7+ luminal compartment for nulliparous (NP) versus parous (P) samples (n = 34 samples, Mann-Whitney test). Results are shown for a subset of the original cohort of sequenced samples ("discovery set", n=19 samples) and a second independent cohort of samples ("validation" set, n = 15 samples). Scale bars 100 μm.

(D) Immunostaining for TCF7, p63, and KRT7 in ducts and terminal ductal lobular units (TDLUs), and quantification of the percentage of TCF7+ cells within the p63+ basal/myoepithelial cell compartment for nulliparous (NP) versus parous (P) samples (n = 33 samples, Mann-Whitney test). Results are shown for both the original cohort of sequenced samples ("discovery set", n=18 samples) and a second independent cohort of samples ("validation" set, n = 15 samples). Scale bars 50 μm.

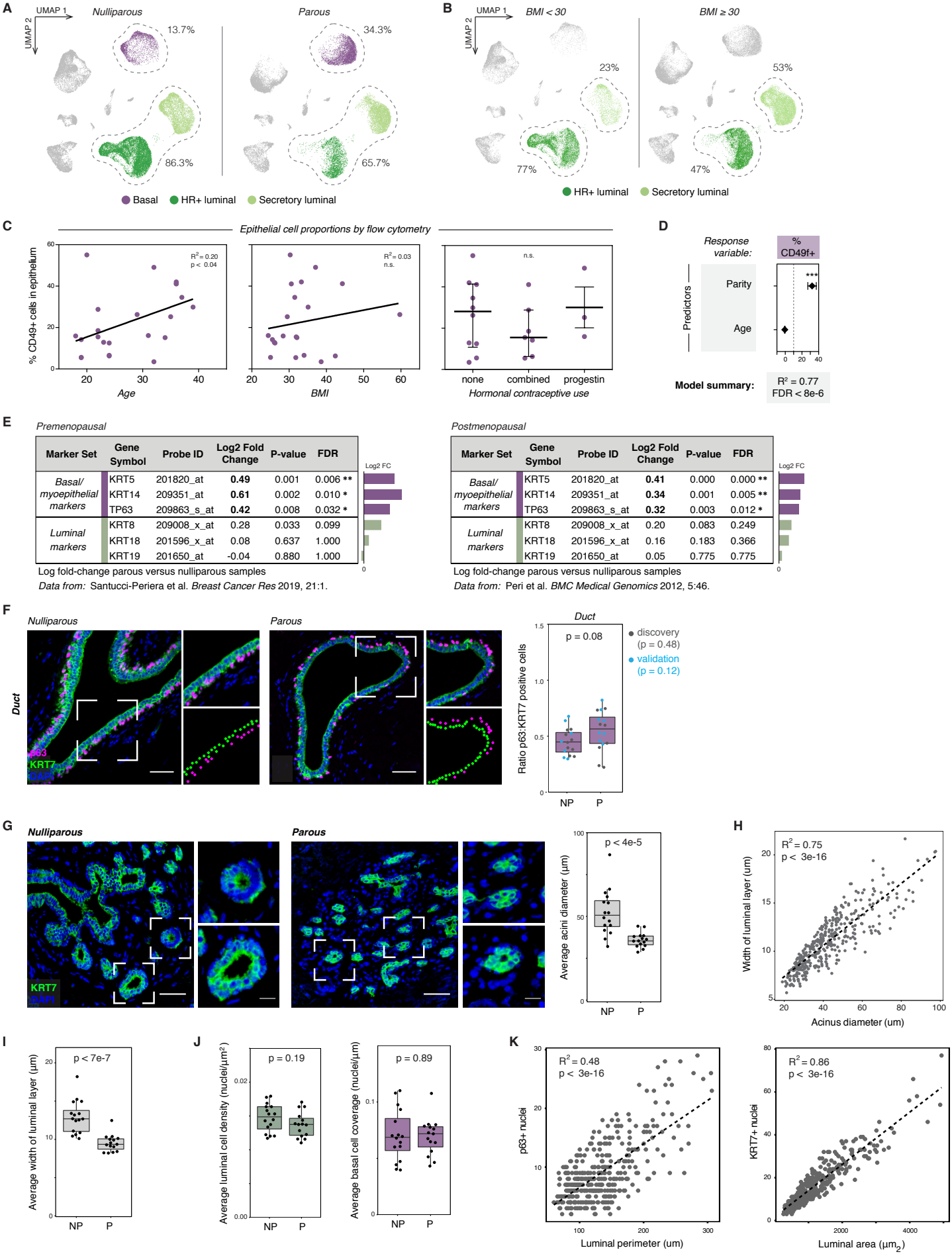

Figure S12 - Prior history of pregnancy is associated with an increased proportion of basal cells in the mammary epithelium

Figure S12. Prior history of pregnancy is associated with an increased proportion of basal cells in the mammary epithelium; Related to Figure 6

(A) UMAP plot of sorted live/singlet and epithelial cells from nulliparous and parous samples from the reduction mammoplasty dataset (GSE198732), with the percent of luminal and basal/myoepithelial cells highlighted.

(B) UMAP plot of sorted live/singlet and epithelial cells from non-obese (BMI < 30) and obese (BMI ≥ 30) samples from the reduction mammoplasty dataset (GSE198732), highlighting hormone-responsive (HR+) and secretory luminal cells.

(C) Quantification of the percentage of $EpCAM^-/CD49f^+$ basal cells identified by flow cytometry analysis versus age (n = 23; $R^2 = 0.20$; $p < 0.04$, Wald test), body mass index (n = 21; $R^2 = 0.03$; p = 0.44, Wald test), or hormonal contraceptive use (n = 23; p = 0.50, Kruskal-Wallis test).

(D) Results from multiple linear regression analysis, with prior pregnancy (parity) and age as predictors and the percentage of $EpCAM^-/CD49f^+$ basal cells in the epithelium as the response variable (Parity $p < 2e-5$; Age p = 0.17; overall $R^2 = 0.77$, $p < 8e-6$, Wald test).

(E) Microarray differential expression analysis for selected genes from Santucci-Periera *et al.* and Peri *et al.* (Peri et al., 2012; Santucci-Pereira et al., 2019).

(F) Immunostaining for the basal/myoepithelial marker p63 and pan-luminal marker KRT7 in ductal regions, and quantification of the ratio of p63+ basal cells to KRT7+ luminal cells in nulliparous (NP) versus parous (P) women (n = 32 samples; p = 0.08, Mann-Whitney test). Results are shown for a subset of the original cohort of sequenced samples ("discovery set", n=17 samples, p = 0.48) and a second independent cohort of samples ("validation" set, n = 15 samples, p = 0.12). Scale bars 50 μm.

(G) Immunostaining for the pan-luminal marker KRT7, and quantification of the average acinar diameter in TDLUs from nulliparous (NP) versus parous (P) samples (n = 31 samples, $p < 4e-5$, Mann-Whitney test). Scale bars 50 μm. Inset scale bars 15 μm.

(H) Linear regression analysis of the width of the luminal layer versus acinus diameter for individual acini (n = 391 acini from 37 samples; $R^2 = 0.75$, $p < 3e-16$, Wald test).

(I) Quantification of the average width of the luminal layer in acini from TDLUs in nulliparous (NP) versus parous (P) samples (n = 37 samples; $p < 7e-7$, Mann-Whitney test).

(J) Quantification of the average luminal cell density (nuclei per μm^2 of luminal area) and basal cell coverage (nuclei per μm of luminal perimeter) in acini from TDLUs in nulliparous (NP) versus parous (P) samples (n = 37 samples; luminal cell density: p = 0.19, Mann-Whitney test; basal cell coverage: p = 0.89, Mann-Whitney test).

(K) *Left:* Linear regression analysis of the perimeter of the luminal layer versus the number of p63+ basal cells for individual acini (n = 391 acini from 37 samples; $R^2 = 0.48$, $p < 3e-16$, Wald test). *Right:* Linear regression analysis of the area of the luminal layer versus the number of KRT7+ luminal cells for individual acini (n = 391 acini from 37 samples; $R^2 = 0.86$, $p < 3e-16$, Wald test).

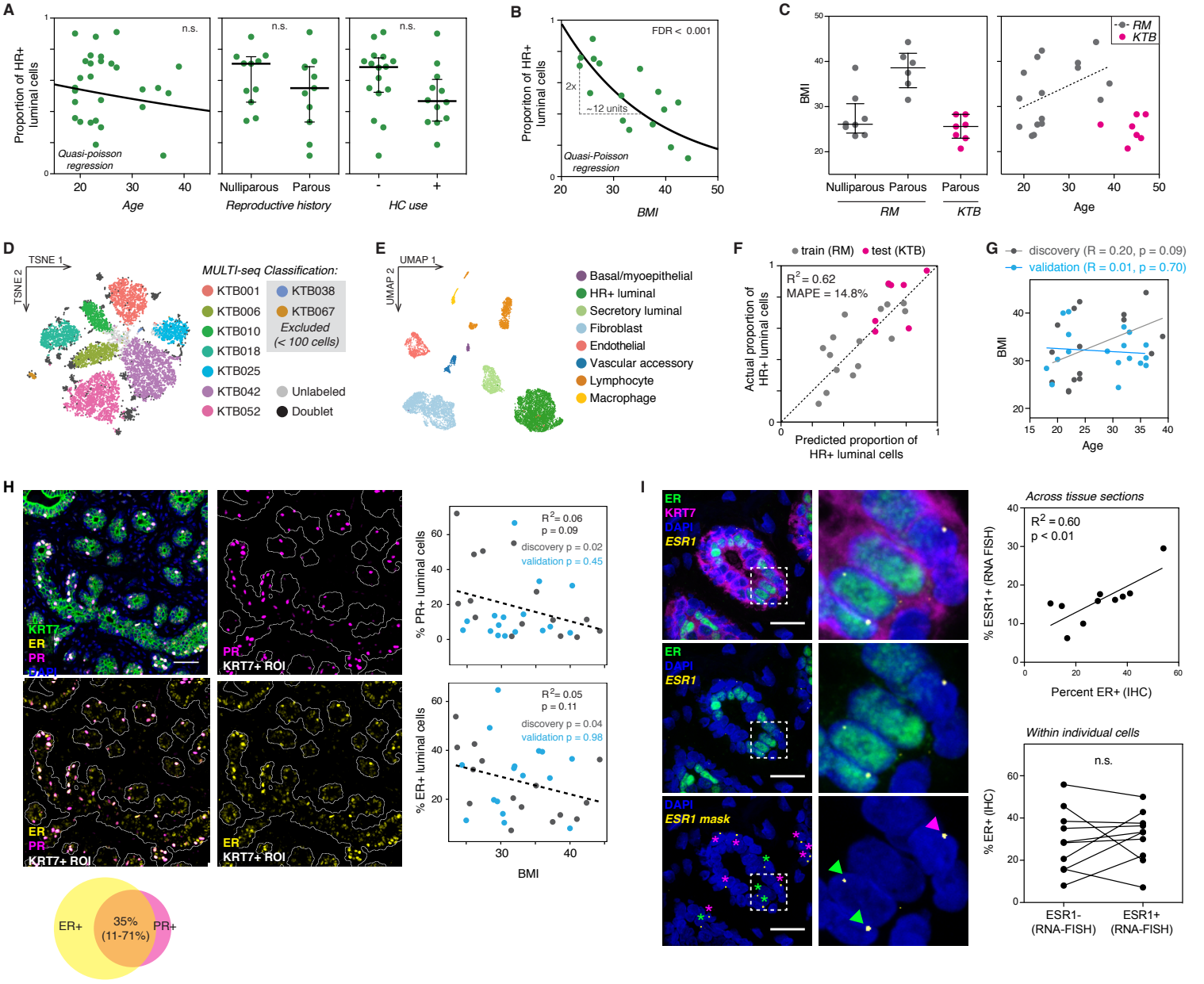

Figure S13 - The proportion of HR+ luminal cells is reduced in obese women, does not vary with other discriminating factors, and is underestimated by staining for ER/PR.

Figure S13. The proportion of HR+ luminal cells is reduced in obese women, does not vary with other discriminating factors, and is underestimated by staining for ER/PR; Related to Figure 6

(A) Proportion of HR+ luminal cells in each sample identified by scRNA-seq clustering, as a function of age, reproductive history, or hormonal contraceptive (HC) use (Wald test).

(B) Quasi-Poisson regression model of the proportion of HR+ cells in the luminal compartment as a function of BMI (FDR < 0.001, Wald test).

(C) Scatter plots highlighting differences in body mass index (BMI), reproductive history, and age between the Komen Tissue Bank (KTB) and reduction mammoplasty cohorts (see also Table S1). The trend line depicts the positive association of BMI with age in the reduction mammoplasty cohort, which is absent in the KTB cohort.

(D) TSNE dimensionality reduction of the normalized barcode count matrices and final sample classification for MULTI-seq barcoding of KTB samples.

(E) UMAP dimensionality reduction and unsupervised clustering of the combined data from seven KTB samples (GSE198732, Table S1) identifies the major epithelial and stromal cell types in the breast.

(F) A quasi-Poisson regression model accurately predicts the proportion of HR+ cells in the luminal compartment as a function of BMI in an independent cohort of core biopsy samples from the Komen Tissue Bank (KTB; mean absolute percentage error = 14.8%).

(G) Linear regression analysis of BMI versus age in the “discovery” and “validation” cohorts. The “validation” cohort is well-balanced across age and BMI.

(H) *Left:* Representative image of co-immunostaining of ER, PR, and KRT7. *Top right:* Linear regression analysis of the percentage of PR+ luminal cells versus BMI (n = 31 samples, R^2 =0.06, p = 0.09, Wald test). *Bottom right:* Linear regression analysis of the percentage of ER+ luminal cells versus BMI (n = 29 samples R^2 =0.05, p = 0.11, Wald test). Scale bars 50 μm. Venn diagram highlights the average percent overlap between ER and PR as measured by immunostaining, for samples where both ER and PR were co-immunostained (n = 5 samples, range = 11-71%).

(I) Multiplexed *in situ* hybridization of estrogen receptor transcript (ESR1) and immunostaining for estrogen receptor protein (ER) and KRT7. Scale bars 25 μm. *Right:* Quantification of the expression of ESR1 and ER across multiple tissue sections (R^2 = 0.6, p < 0.01, Wald test) or within individual cells (p = 0.63, Wilcoxon matched pairs signed-rank test).

(J) Table and bar plot depicting the sensitivity and specificity for ESR1 or PGR transcript expression in the HR+ luminal cell versus secretory luminal cell cluster based on scRNA-seq analysis.

(K) Representative images of co-immunostaining of PR, KRT23, and the pan-luminal marker KRT7. (See quantification of n = 41 samples in Figure 6F).

(L) Co-immunostaining of ER, KRT23, and the pan-luminal marker KRT7 and quantification of the percentage of ER+ cells within the KRT7+/KRT23- and KRT7+/KRT23+ luminal cell populations (n = 5 samples; p < 0.01 Mann-Whitney test). Scale bars = 50 μm.

(M) Results from multiple linear regression analysis, with body mass index (BMI), prior pregnancy (parity), and age as predictors and the percentage KRT23+ cells in the KRT7+ luminal compartment as the response variable (BMI p < 1e-4; Parity p = 0.29; Age p = 0.35; overall R^2 = 0.69, p < 7e-6, Wald test).

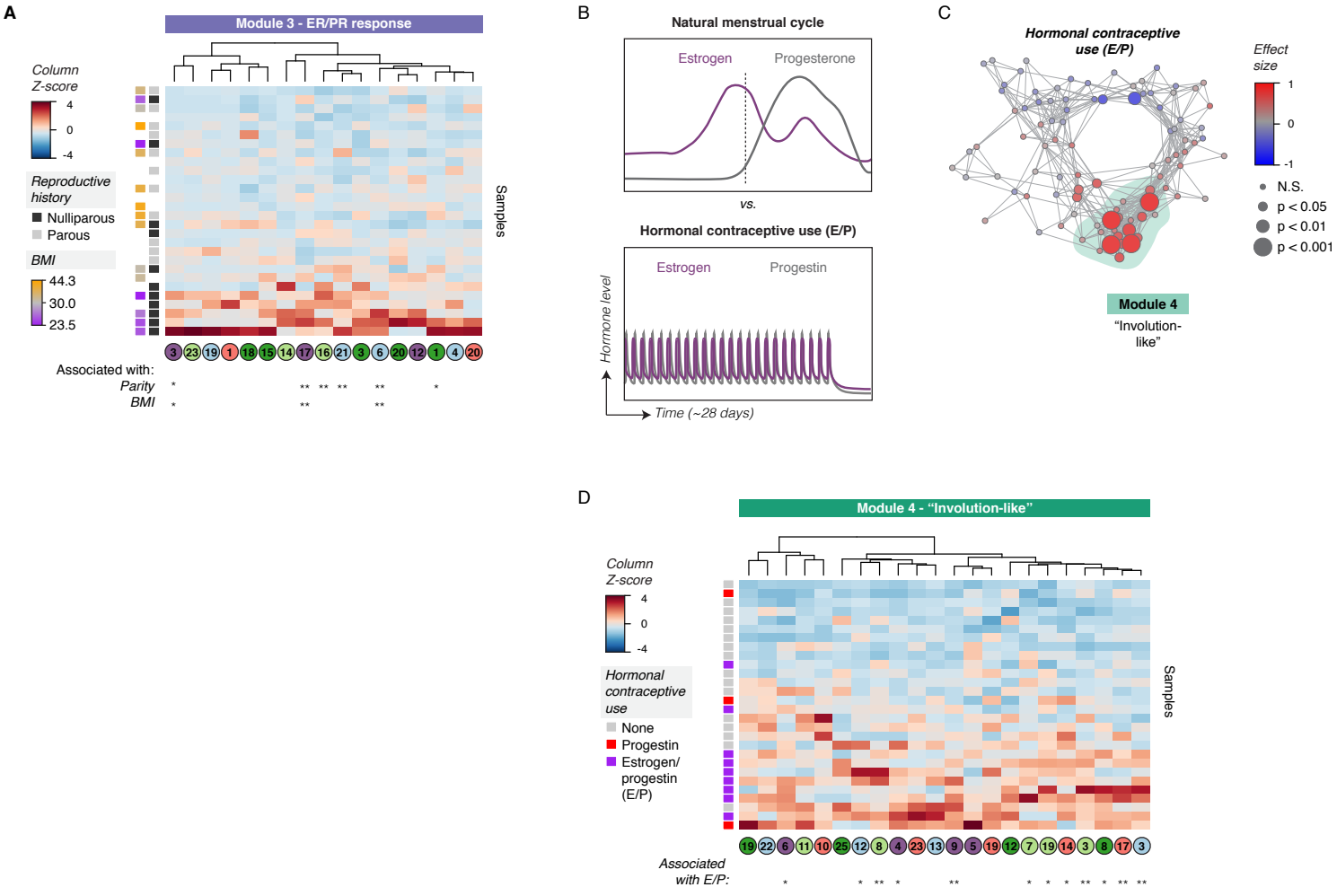

Figure S14 - Association of cell-cell interaction modules with reproductive history, body mass index, and hormonal contraceptive use

Figure S14. Association of cell-cell interaction modules with reproductive history, body mass index, and hormonal contraceptive use; Related to Figure 7

(A) Heatmap depicting the relative average expression score (column z-score, linear scale) of the indicated activity programs in each sample. Samples were ordered by the mean Z-score of all programs within the “ER/PR response” module. Activity programs were ordered by hierarchical clustering (complete linkage). Significant associations with parity or BMI are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$; parity: Mann-Whitney test; BMI: Wald test).

(B) Schematic depicting relative estrogen and progestin levels and dynamics across the natural menstrual cycle and in donors using combination (E/P) hormonal contraceptives. Samples from donors using hormonal contraceptives were used as a “virtual experiment” to test the effects of estrogen/progesterone treatment on downstream signaling pathways.

(C) Network graph of activity programs in the human breast, colored by the effect size of combined hormonal contraceptive use (Wilcoxon effect size, linear scale)) on each activity program. Significant positive and negative associations are represented by larger nodes (Mann-Whitney test).

(D) Heatmap depicting the relative average expression score (column z-score, linear scale) of the indicated activity programs in each sample. Samples were ordered by the mean Z-score of all programs within the “Involution-like” module. Activity programs were ordered by hierarchical clustering (complete linkage). Significant associations with combined hormonal contraceptive use are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$; Mann-Whitney test).