Supplemental Materials and Methods

Bioinformatics analysis:

IDC samples were extracted from the Ciriello *et. al.* study cohort based on the final pathology classification (1). Genes that were not detected in >50% of these samples were removed from the analysis. Sample purity can confound network analysis of gene expression data, therefore we applied linear regression to model the relationship between gene expression and tumor purity across samples. The residuals from this model were converted to z-scores (scaled by gene) and used for downstream analysis.

WGCNA was performed using the R package as described previously (2), with minor alterations. To select genes, principal components analysis (PCA) was performed, which revealed that the first 9 principal components accounted for ~95% of the variance within the data. Genes were ranked by their highest absolute correlation to one of these components and the top 50% were selected for use in WGCNA. A correlation matrix was calculated for these genes. This was converted to a 'signed hybrid', scale-free adjacency matrix by raising each positive correlation to a power of 7 (the lowest power for which the scale free topology fit $R_2 > 0.85$). This adjacency matrix was then used to calculate the topological overlap (TO) between genes. Finally, unsupervised hierarchical clustering was carried out using 1-TO as the distance measure to identify gene modules. 2 summary statistics were calculated for each module: the module eigengene (first principal component), used for correlation analysis; and the average expression of each gene within the module, used for comparing expression. GO Biological processes and KEGG Pathway enrichment analyses were then carried out on the genes within each module with a module membership score (correlation to the module eigengene) >0.7, using the enrichR package in R (3). For GSEA all genes were ranked by their correlation to the module eigengene and analyzed using the fgsea package in R (4).

Flow cytometry:

Epithelial and immune cell depletion efficiency was determined using flow cytometry (Attune NxT). Cell suspensions were stained with the Zombie Green™ Fixable Viability Kit (BioLegend), fixed for 20 mins in 4% PFA and stored at 4ºC overnight. Antibody staining was then carried out in PBS + 2% FCS + 1 mM EDTA using FITC conjugated anti-mouse TER-119 (Biolegend); PE/Cy5 conjugated anti-mouse CD45 (30-F11; Biolegend); and APC/Cy7 conjugated anti-mouse CD326 (G8.8; Biolegend). Flow cytometry data analysis was carried out using the FlowJo software package. Dead and erythroid cells were excluded using a FSC-A x FITC plot and doublets were excluded using a FSC-A x FSC-H plot. CD45 and EpCAM positive cells were then gated using fluorescence minus one control to set a quadrant gating strategy.

Tumor organoid and stromal cell culture:

3D organotypic cultures were setup in collagen gels as follows. A neutralized collagen solution was prepared at approximately 3 mg/ml (typically 87.5% rat tail collagen (SIGMA), 10% 10X DMEM (SIGMA) and 1.5% 1N NaOH (SIGMA); further NaOH was added if required to generate a neutral pH). To facilitate gel adherence to glass bottom 24 well plates, a portion of this solution was diluted 1:1 in DMEM and 15 µls were used to coat the center of each well with a thin layer of collagen. The plates were then incubated at 37ºC for approximately 30 minutes. During this time the stock neutralized collagen solution was incubated at 4ºC to stimulate nucleation of collagen fibers. Following this incubation period tumor organoids were embedded in 100 µl collagen gels at a density of 1 organoid/µl with or without 50,000 stromal cells. The gels were incubated at 37ºC for 1 hour to stimulate gelation and 1 ml of basal medium supplemented with 2.5nM FGF2 was added to each well. Inhibitors such as the γ-secretase inhibitor DAPT (D5942, Sigma), the TGF-β RI kinase inhibitor II TGFBRIi (616452,

Millipore) and the NOX1/4 inhibitor NOX1/4i (GKT831), or an equivalent volume of DMSO, were added at this point where indicated.

Stromal cells culture in 2D:
After isolation, some stromal cells were cultured in 2D in DMEM-F12 complemented with Gluta-MAX, insulin-transferrin-selenium (Gibco) and penicillin-streptomycin (Sigma). Cells were treated with TGF-β (T7039 Sigma), TGF-β RI kinase inhibitor II TGFBRIi (616452, Millipore) or an equivalent amount of DMSO for 72 hours before proceeding to protein extraction.

Western blot:
Stromal cells proteins were extracted in a lysis buffer (1X Complete mini tablets (Roche), 1X PhoSTOP (Roche), 5% glycerol (Sigma), 1X RIPA buffer (Millipore), 0,1% SDS) and kept on ice for 20 minutes. After centrifugation at 14,000 rpm for 10 minutes at 4°C, protein amount was quantified using a BCA assay kit (Thermo Fisher). Twenty ug of protein were then loaded with 1X Laemmli (Biorad) and β-mercaptoethanol (Sigma) in 4-15% Mini-PROTEAN pre-cast gels (BioRad). After migration, the proteins were transferred onto PVDF membranes (Millipore). Membranes were then blocked with Odyssey blocking buffer (Licor) for 1 hour at room temperature and probed with Nox4 (PA5-72816 Thermo Fisher; 1:1,000) and β-actin (A5441 Sigma; 1:1,000) primary antibodies diluted in blocking buffer and 0.1% TBS-Tween overnight at 4°C. Membrane were washed three times with 0.1% TBS-Tween and incubated with the following secondary antibodies: anti-rabbit IRDye 800CW (925-32211 Licor; 1:10,000) and anti-mouse IRDye 680CW (925-68070 Licor; 1:10,000) for 1 hour at room temperature. Membrane were then washed three times with 0.1% TBS-Tween. The membranes were scanned using the Odyssey CLx imaging system and band intensities were quantified using the ImageJ software.

Immunofluorescent staining:
Organotypic cultures were fixed in 4% PFA for 10 minutes at 37ºC and then incubated in blocking buffer (PBS+0.5%Triton+10%FCS+1%BSA) for 2 hours at room temperature. All antibody labelling steps were carried out in blocking buffer followed by 3 x 15 minute washes in blocking buffer. For anti-cytokeratin (wide spectrum,1:250; Abcam ab9377) staining the primary antibody was incubated for 2 hours at room temperature and over-night at 4°C; followed by a 4 hour incubation at room temperature with anti-rabbit AF647 (1:250; ThermoFisher); finally anti-KRT14 conjugated to FITC (Biolegend 19053; custom conjugation; 1:300) was then incubated over-night at 4°C.

Image acquisition:
Confocal micrographs were acquired as z-stacks with 5µm spacing across 45 µm using a custom spinning disc microscope with 10/20/40x objectives. Maximum intensity z-projections are presented in all figures. Time-lapse differential interference contrast (DIC) imaging of tumor organoids was conducted using a Zeiss Cell Observer system with a Zeiss AxioObserver Z1 and an AxioCam HR3 camera.

Image analysis:
All images were analyzed using Fiji (5) open source image analysis software with bespoke image processing macros. Organoids were segmented using the fluorescence signal from epithelial/organoid specific markers. This was achieved in two ways: where non-fluorescent stromal cells were used, the endogenous mTomato signal specifically emitted from the organoid was used for segmentation. When

organoids did not express the mTomato reporter, or both tumor and stromal cells expressed it, pan-cytokeratin immunofluorescence was used for segmentation.

Organoid segmentation was carried out on maximum intensity z-projections with the following image processing steps: a 3x3 pixel average (mean) filter was applied 3 times, followed by Triangle thresholding. 2 3x3 pixel maximal filters followed by a 3x3 pixel minimal filter was then applied to close any small holes in the detected particles. The organoid was then detected as the particle with centroid coordinates closest to the center of the image, a region of interest (ROI) outlining this particle was used to calculate shape descriptors, saved for quality control assessment and use in quantifying KRT14 expression. KRT14 staining fluorescence intensity was measured on maximum intensity projections from within the organoid ROI (detected as described above). The mean fluorescence intensity was measured for both KRT14 staining and the ubiquitously expressed endogenous reporter used for organoid segmentation. Normalized KRT14 expression was calculated as KRT14/ubiquitous reporter to account for fluorescence intensity variation due to differing depths of organoids within the gel.

Supplemental Figure 1

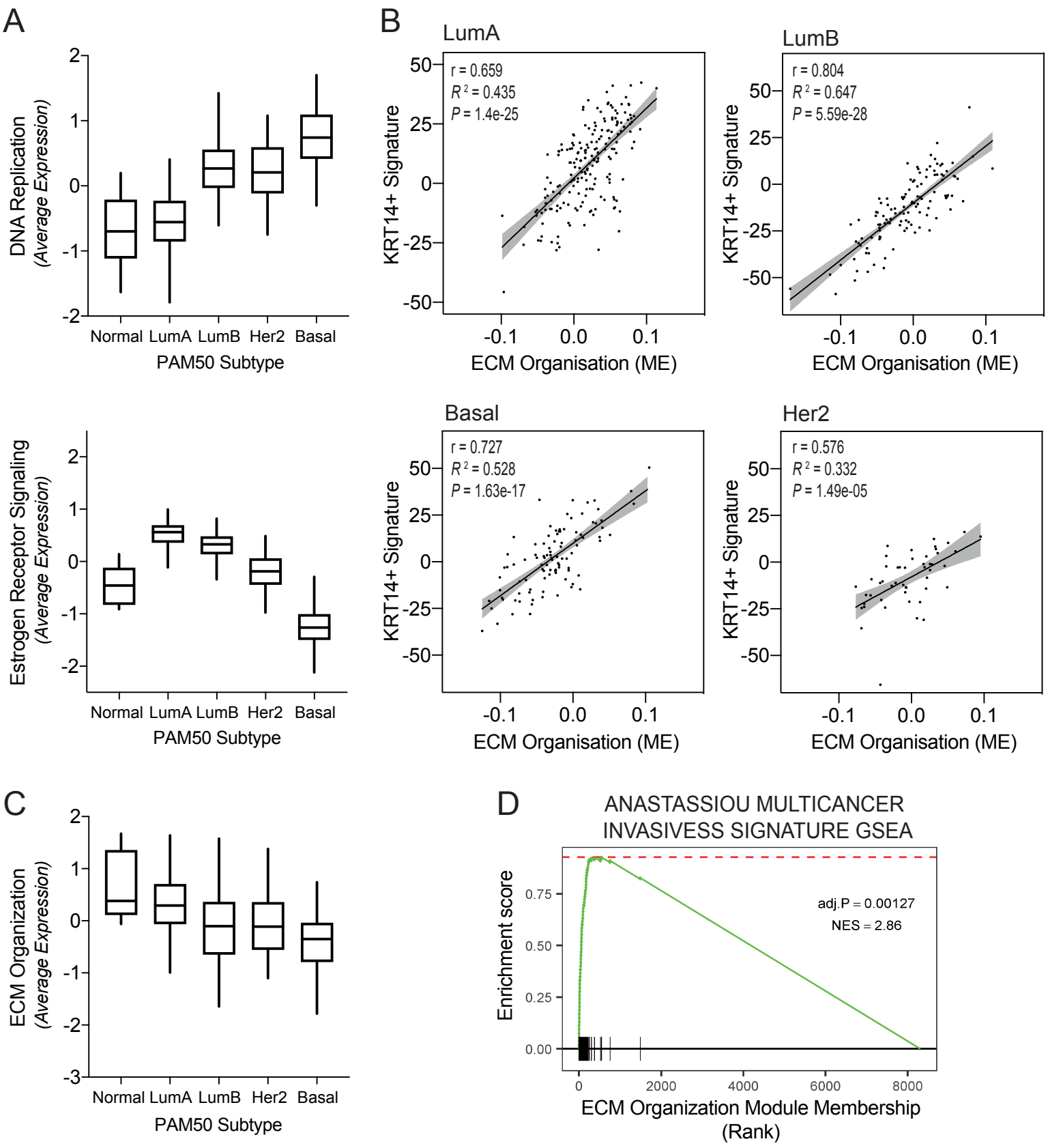

Figure S1. Weighted gene correlation network analysis of RNA-Sequencing from human invasive ductal breast cancer primary tumor samples. A-B) Boxplots showing average gene expression in the DNA replication, estrogen receptor signaling and ECM organization modules in samples grouped by PAM50 subtypes. **C**) Scatterplots showing the relationship between KRT14+ tumor cells and the ECM organization module eigengene (first principal component) in different PAM50 subtypes. Pearson's r (r) with corresponding Student's P value *(P*), and adjusted R_2 from linear regression analysis are displayed for each subtype. **D**) Gene set enrichment analysis plot showing significant enrichment of genes from the Anastassiou multicancer invasiveness signature in the ECM organization module.

Supplemental Figure 2

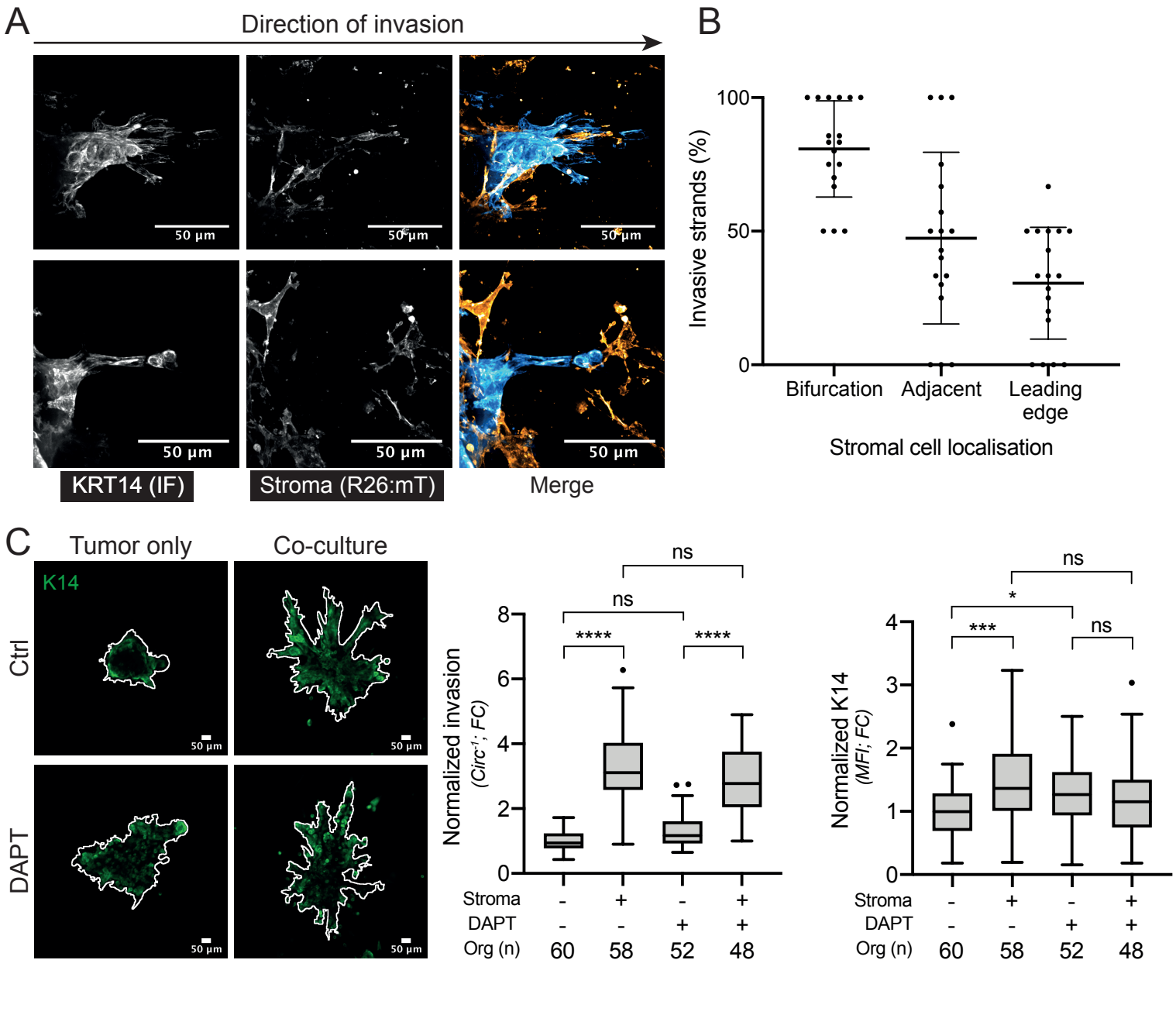

Figure S2: Physical interaction between stromal cells and invasive organoids. A) Confocal imaging of organoid invasive strands in co-culture with stromal cells. Stromal cells accumulate at the bifurcation point (e.g. white arrows), adjacent to the invasive strand (e.g. white arrowhead), and at the leading edge (e.g. red arrow). **B**) Quantification of stromal cell localisation at different positions on the invasive strand. Please also see Supplementary video 2. Scale bars = 50µm. **C**) MMTV-PyMT organoids, either in mono-culture or co-culture, were treated with 50 µM gamma secretase inhibitor (DAPT) or vehicle (DMSO). Left panel shows representative confocal images of organoid invasion (organoid outline in white) and KRT14 expression. Scale bar = 50 µm. Right panel shows quantifications of invasion and KRT14 intensity from 3 different tumors. ns = not-significant ($p>0.05$), *$p<0.0332$, ***$p<0.0002$, ****$p<0.0001$ (ANOVA test). $Circ^{-1}$ = inverse circularity; FC = Fold Change; MFI = Mean Fluorescence Intensity.

Supplemental Figure 3

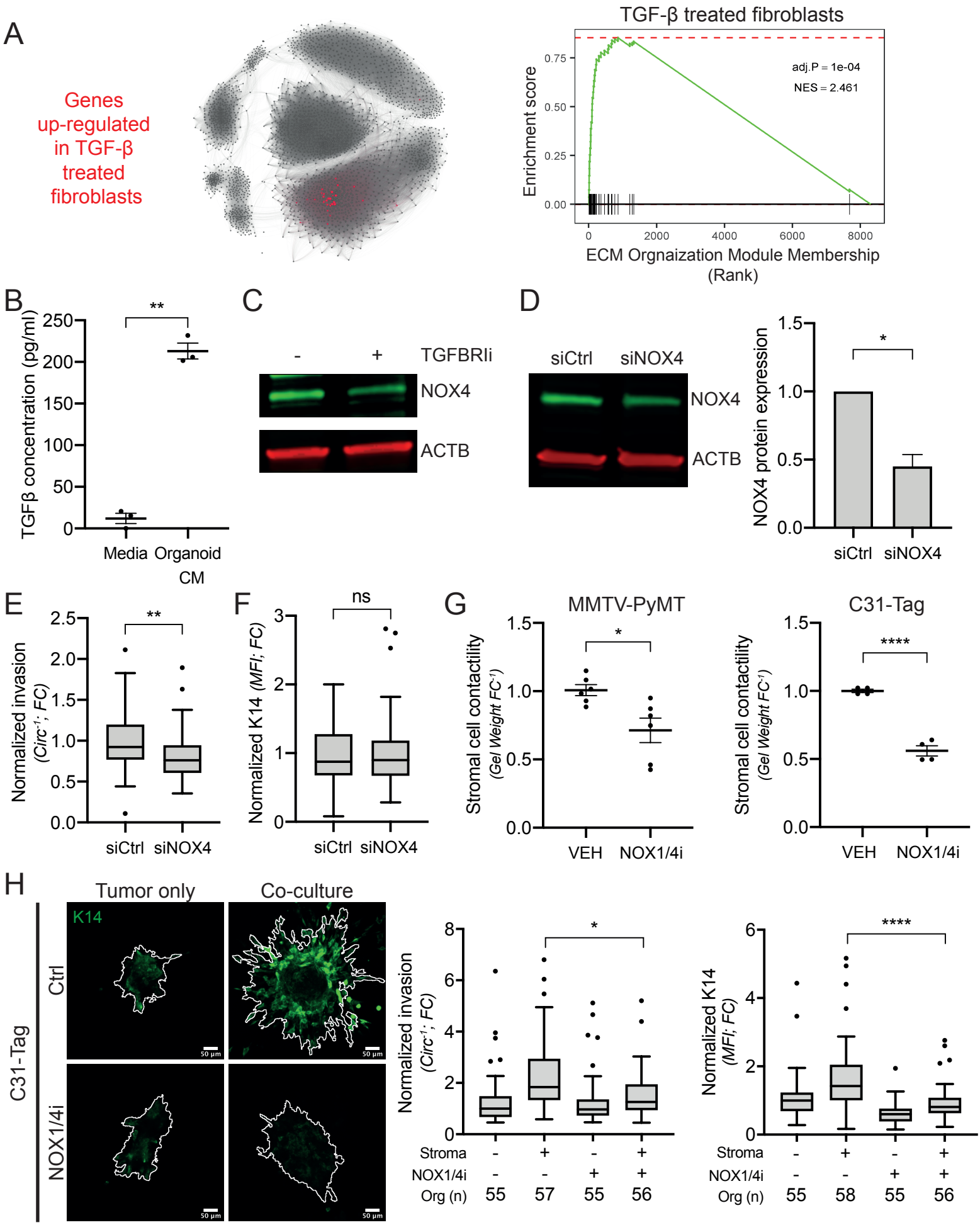

Figure S3: TGF-β-NOX4 regulate stromal cell activation and reciprocal induction of tumor cell invasion. A) Network plot highlighting TGF-β target genes in red and showing GSEA for a TGF-β treated fibroblast gene signature in the ECM organization WGCNA module. **B**) ELISA measuring MMTV-PyMT tumor organoids secretion of TGF-β without stromal cells from 3 different tumors. **C**) Western blot showing NOX4 expression in stromal cells treated with organoid conditioned media +/- TGFβRIi (1μM). **D**) Western blot showing effect of NOX4 knockdown in stromal cells by siRNA targeting, data is shown from three independent experiments, statistical comparison made using an unpaired T-test with Welch's correction to account for zero variance in the control. *p<0.0332. **E-F**) Quantification of tumor organoid invasion (E) and KRT14 expression (F) when co-cultured with stromal cells treated with control or NOX4 targeting siRNA from 3 different tumors. Statistical testing was performed using a Mann-Whitney test. **p<0.0021, ns= not-significant (p>0.05) **G**) Stromal cell collagen gel contraction assays, treated with vehicle or NOX1/4i (40μM). Points represent individual gels with the mean +/- SEM also shown; statistical testing was performed using an unpaired 2-tailed t-test comparing gels from 3 biological replicates (independent tumors). *p<0.0332, ****p<0.0001. **H**) C31-Tag basal breast cancer model organoids, either in mono-culture or co-culture, treated with vehicle (DMSO) or NOX1/4i (40μM). Left panel shows representative confocal images of C31-Tag organoids invasion (organoid outline in white) and KRT14 expression. Scale bar = 50 μm. Right panel shows quantification of invasion and KRT14 intensity from 3 different tumors. *p<0.0332, ****p<0.0001 (ANOVA test). $Circ^{-1}$ = inverse circularity; FC = Fold Change; MFI = Mean Fluorescence Intensity.

Supplemental Movie Legends

Supplemental Movie S1. Co-culture of tumor organoids with stromal cells increases tumor cells invasion in collagen I. Representative time-lapse DIC movie of tumor organoids in mono-culture (left panel) or in co-culture with stromal cells (right) panel. Frames were collected every 20 minutes for 8 hours post-plating (displayed at 6 fps).

Supplemental Movie S2. Stromal cells remodel collagen matrix. Representative time-lapse DIC movie of matrix remodeling by a stromal cell adjacent to a tumor organoid (bottom left corner). Frames were captured every 20 minutes form 8 hours post-plating (displayed at 6 fps).

Supplemental Table Legends

Supplementary Table 1. Summary statistics from WGCNA. Correlations between WGCNA module eigengenes and KRT14+ tumor cell signature/RPPA signature and Gene ontology analysis.

Supplementary Table 2. Module membership statistics for genes in the ECM organization module.