Meanwhile, @Nawrocki, Cole C. could you please send me one or two paragraphs for the description of the bioinformatic analyses for the RNA-seq data handling, cold/hot tumor DEGs & pathway, fibroblast Deconv & pathway? I will make sure to include them in the updated manuscript text.

**Processing NanoString GeoMx NGS Data**

NanoString GeoMx NGS data was processed using the same strategies as those used in the corresponding vignette[[1]](#footnote-1) released by NanoString. The GeomxTools[[2]](#footnote-2) R package was used to convert DCC files to counts data and to perform quality control (QC). Specifically, the following QC flags were passed to the setSegmentQCFlags function: minSegmentReads = 5000, percentTrimmed = 90, percentStitched = 90, percentAligned = 90, percentSaturation = 90, minNegativeCount = 0, maxNTCCount = 10000, minNuclei = 10, minArea = 500. Furthermore, any area of interest (AOI) that did not detect ≥10% of genes in the panel was excluded. Lastly, any gene that was not detected in at least 10% of AOIs was excluded. After QC, 155 of 190 AOIs and 14,833 of 18,676 genes remained for downstream analysis. Processing and analysis code is available in the following GitHub repository: ccnawrocki/ellisen-breast-geomx-project.

**Deconvolution and Abundance Analysis**

The processed counts data was third quantile (Q3) normalized, and this Q3 normalized data was used for deconvolution with the SpatialDecon[[3]](#footnote-3) R package. First, a custom reference profile was derived from breast cancer scRNA-seq data[[4]](#footnote-4), using the create\_profile\_matrix function. Differentially expressed marker genes had previously been defined for each fibroblast sub-type in the scRNA-seq data. From this list of marker genes, only the significant (FDR < 0.05) genes with the 50 largest log fold-change values for each sub-type were included for reference derivation. Next, the spatialdecon function was used to estimate the proportion of each fibroblast AOI that aligned with each fibroblast sub-type in the reference data. Using these results and the number of total cells present in each AOI, the number of cells corresponding with each sub-type was estimated. The cell counts were modeled for each sub-type, using infiltration identity (cold versus hot) as a predictor variable, in a Poisson regression via the glm function (R stats package). For clarity, the model formula used was as follows: count ~ 1+group+offset(log(ncells)), where group is a binary categorical variable defining an AOI’s infiltration identity, and ncells is the total number of cells in an AOI. The average cell count difference across levels of the group variable was then tested with a two-sided Local Wald Test, implemented with the multcomp[[5]](#footnote-5) R package. Significance was assessed after all p-values were corrected with Bonferroni adjustment.

**Differential Expression Analysis**

Differential expression analysis was performed using the limma[[6]](#footnote-6) R package. First, using the trimmed mean of M-values (TMM) method[[7]](#footnote-7), effective library sizes for each AOI were obtained with the calcNormFactors function from edgeR[[8]](#footnote-8). Using these library sizes, the counts data was normalized with limma’s voom function. Normalized expression of each gene was then modeled, using infiltration identity (cold versus hot) and patient identity as fixed predictor variables. Setting patient identity as a random effect in expression models was explored, although model fitting with this approach led to unreliable fits. Thus, the simpler modeling scheme was chosen for downstream analysis. Using results from the limma modeling pipeline, the following linear contrasts were tested for all genes: cold tumor AOIs versus hot tumor AOIs and cold fibroblast AOIs versus hot fibroblast AOIs. For each comparison, each gene’s differential expression significance was assessed after all p-values were corrected with Benjamini-Hochberg adjustment, the default for limma.

**Gene Set Enrichment Analysis**

Gene set enrichment analysis (GSEA) was performed using the clusterProfiler[[9]](#footnote-9) R package. For each of the two linear contrasts tested during differential expression analysis, the obtained results were pre-sorted by t statistic value. GSEA was then implemented with the fgsea[[10]](#footnote-10) algorithm on this pre-sorted gene list for the following gene set collections: Hallmark gene sets from MSigDB[[11]](#footnote-11) and Biological Processes gene sets from Gene Ontology[[12]](#footnote-12). Significance was assessed after all p-values were corrected with Benjamini-Hochberg adjustment.

1. https://bioconductor.org/packages/release/workflows/vignettes/GeoMxWorkflows/inst/doc/GeomxTools\_RNA-NGS\_Analysis.html [↑](#footnote-ref-1)
2. https://bioconductor.org/packages/release/bioc/html/GeomxTools.html [↑](#footnote-ref-2)
3. https://bioconductor.org/packages/release/bioc/html/SpatialDecon.html [↑](#footnote-ref-3)
4. https://www.nature.com/articles/s41467-023-39762-1 [↑](#footnote-ref-4)
5. https://cran.r-project.org/web/packages/multcomp/index.html [↑](#footnote-ref-5)
6. https://bioconductor.org/packages/release/bioc/html/limma.html [↑](#footnote-ref-6)
7. https://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-3-r25 [↑](#footnote-ref-7)
8. https://bioconductor.org/packages/release/bioc/html/edgeR.html [↑](#footnote-ref-8)
9. https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html [↑](#footnote-ref-9)
10. https://bioconductor.org/packages/release/bioc/html/fgsea.html [↑](#footnote-ref-10)
11. https://www.gsea-msigdb.org/gsea/msigdb [↑](#footnote-ref-11)
12. https://geneontology.org/docs/introduction-to-go [↑](#footnote-ref-12)