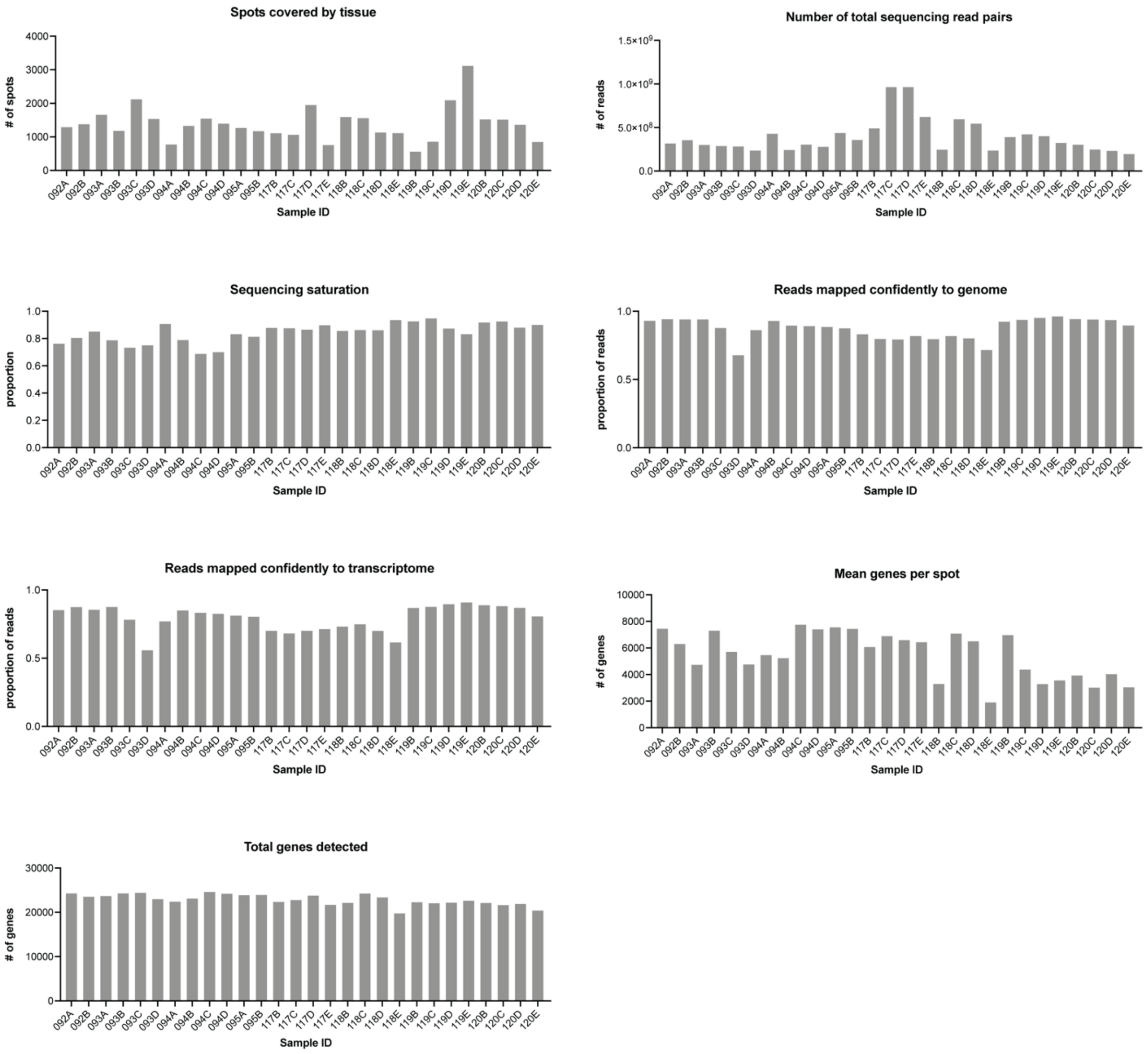
Supplemental figures corresponding to “Spatial transciptomic analysis of a diverse patient cohort reveals a conserved architecture in triple-negative breast cancer” by Bassiouni et al.

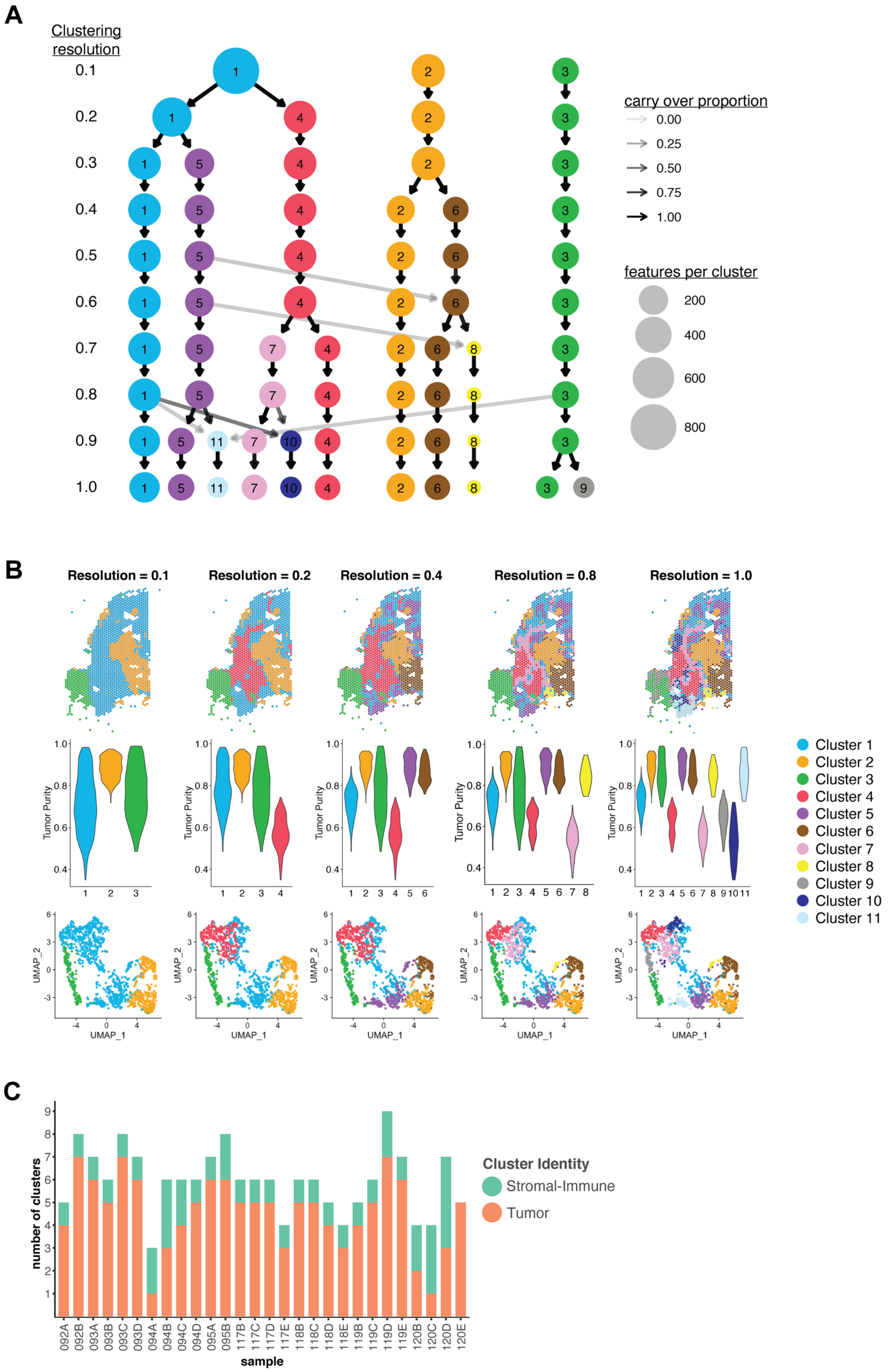
Supp Figure 1. Sequencing quality metrics



<u>Supplemental Figure 1: Sequencing quality metrics</u>

Sequencing quality metrics for 28 reference cohort samples.

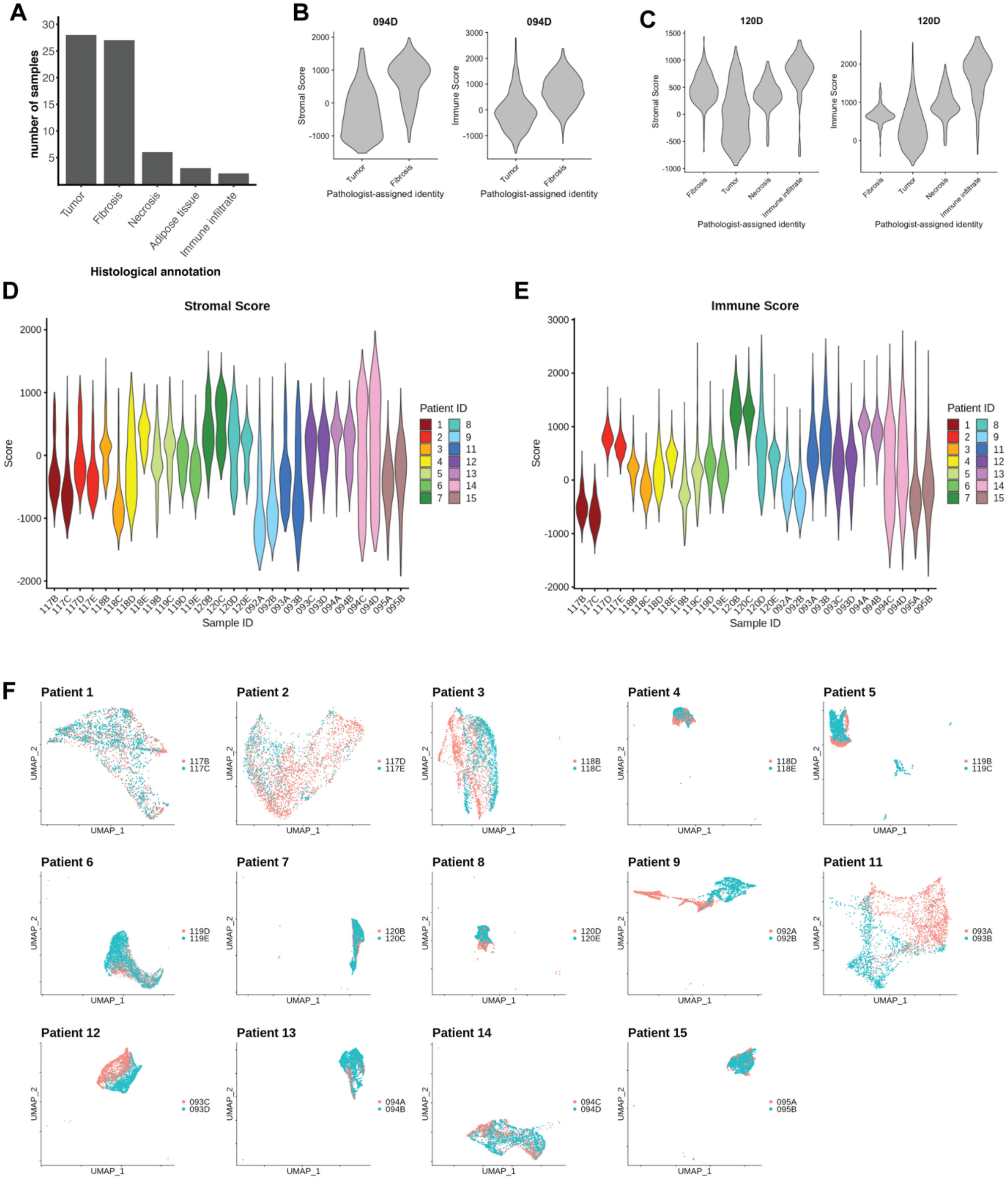
Supp Figure 2. Assessment of clustering resolutions



Supplemental Figure 2: Assessment of clustering resolutions

(A) Clustering tree of sample 094D, illustrating the change in total cluster number and cluster size at varying clustering resolution. Each row represents the clusters produced at the resolution indicated on the left. The size of each circle represents the number of features in that cluster. Arrows indicate the movement of spatial features as cluster resolution increases, and their opacity indicates the proportion of cells represented in each arrow. A new cluster that contains features from more than one input cluster may be an indication of overclustering (as seen at resolution 0.6, 0.7, and 0.9). Clustering was stable at 0.4, and this was selected as the resolution for further analysis. (B) Spatial maps, tumor purity profiles, and UMAPs of sample 094D at varying clustering resolutions illustrate the difference between under-clustering (at resolution 0.1) and over-clustering (at resolution 1.0). (C) Clustering analysis at resolution = 0.4 was applied to all samples. The bar graph summarizes the number of resulting clusters and their identity, as determined by ESTIMATE analysis and comparison with manual annotation.

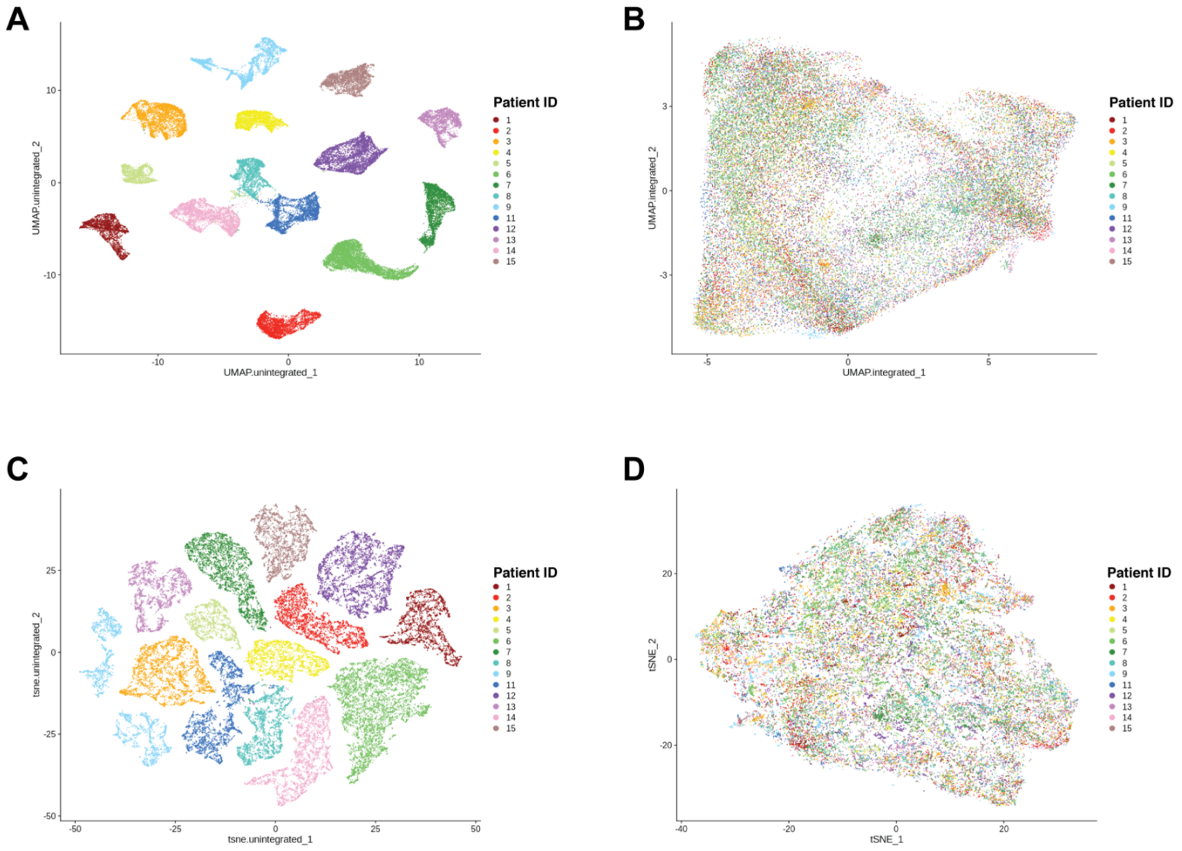
Supp Figure 3. Additional sample annotations



Supplemental Figure 3: Additional sample annotations

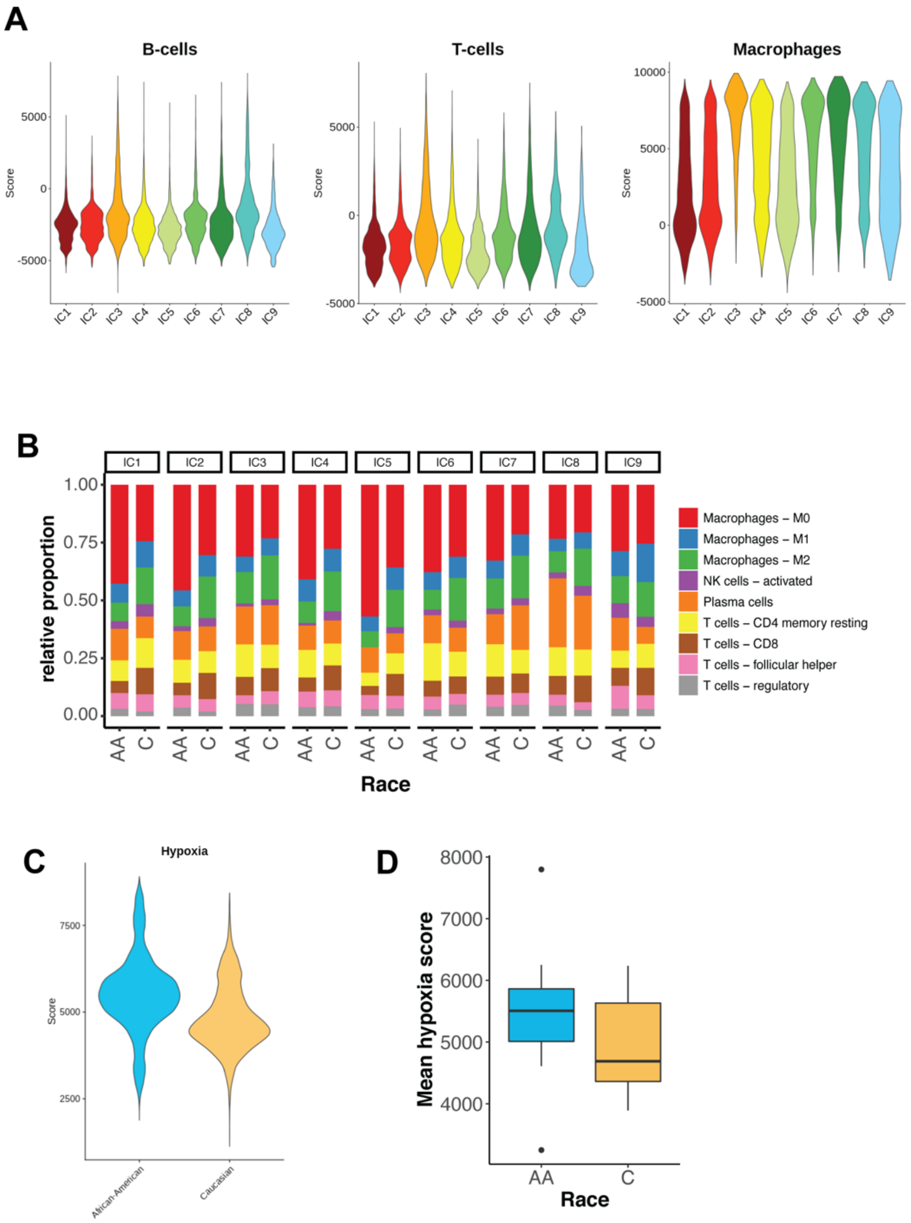
(A) Summary of tissue types identified by manual pathological annotation in 28 samples. The y-axis represents the number of samples in which the annotation was present. (B) Violin plots of stromal and immune scores for histologically distinct regions in sample 094D, as determined by a pathologist. (C) Violin plots of stromal and immune score compared to pathologist's annotations for sample 120D. (D-E) Cumulative stromal and immune scores for all 28 samples in the reference cohort. (F) UMAP embeddings of paired samples from each patient, reflecting the degree of heterogeneity in sampled tumors.

Supp Figure 4. Dimensionality reduction of integrated reference dataset



Supplemental Figure 4: Dimensionality reduction of the integrated reference dataset
(A-B) UMAP embeddings of 28 samples before (A) and after (B) integration, colored by patient ID. (C-D) t-SNE of 28 samples before (C) and after (D) integration, colored by patient ID.

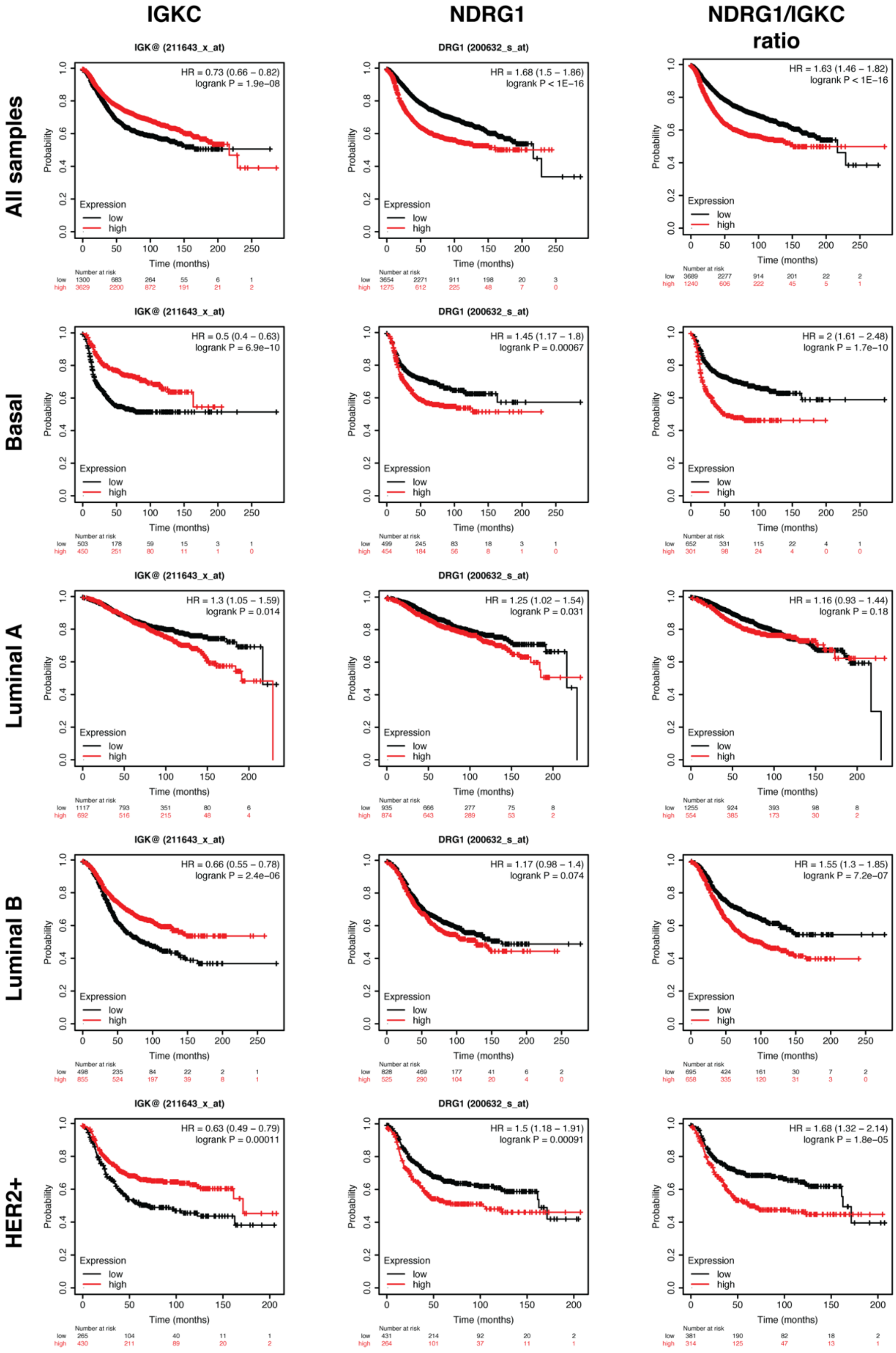
Supp Figure 5. Additional annotation of integrated clusters



Supplemental Figure 5: Additional annotation of integrated clusters

(A) Enrichment scores of lymphocytes across ICs, as determined by single sample gene set enrichment analysis (ssGSEA) using gene sets defined by Yu et al [25]. (B) CIBERSORTx analysis of ICs divided by race. Results were scaled to a cumulative maximum of 1, and relative proportions of individual cell types are indicated on the y-axis. AA = African-American; C = Caucasian. (C) Violin plot of feature hypoxia enrichment scores by race in the reference dataset, as determined by ssGSEA. (D) Boxplot representing average hypoxia score per sample divided by race. The mean of the distribution is represented by a bold horizontal line.

Supp Figure 6. Additional survival analysis



Supplemental Figure 6: Additional survival analysis

Kaplan-Meier curves depicting relationship of gene expression with relapse-free survival in publicly available breast cancer datasets, as described in Methods. Genes examined were *IGKC*, *NDRG1*, and the ratio of *NDRG1* to *IGKC*. The cohort was examined as a whole, or divided into PAM50 subtypes: basal, luminal A, luminal B, and HER2+. Hazard ratio (HR), p-value, and sample size are displayed.