

Supplementary Information

Characterization of four subtypes in morphologically normal tissue excised proximal and distal to breast cancer. Gadaleta *et al.* 2020

Supplementary Data Set 1 Clinical and pathological characteristics of the cohort.

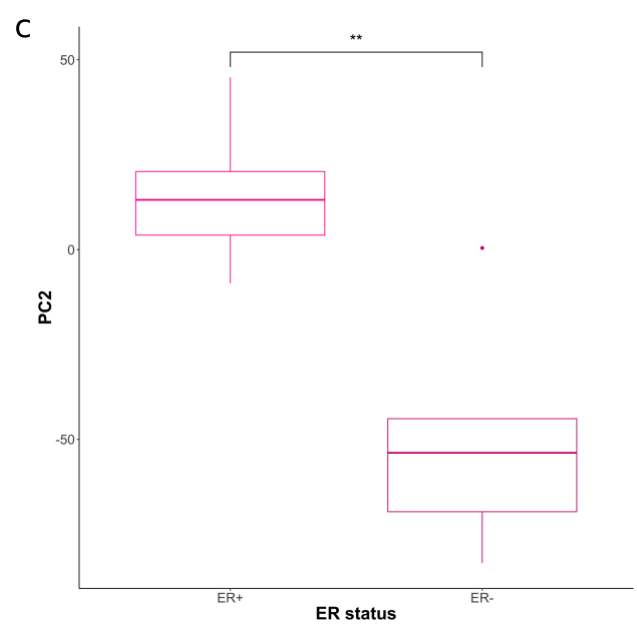
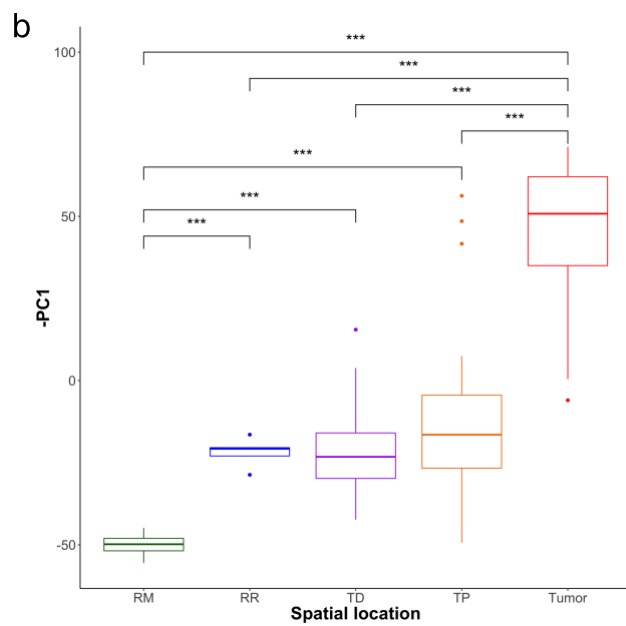
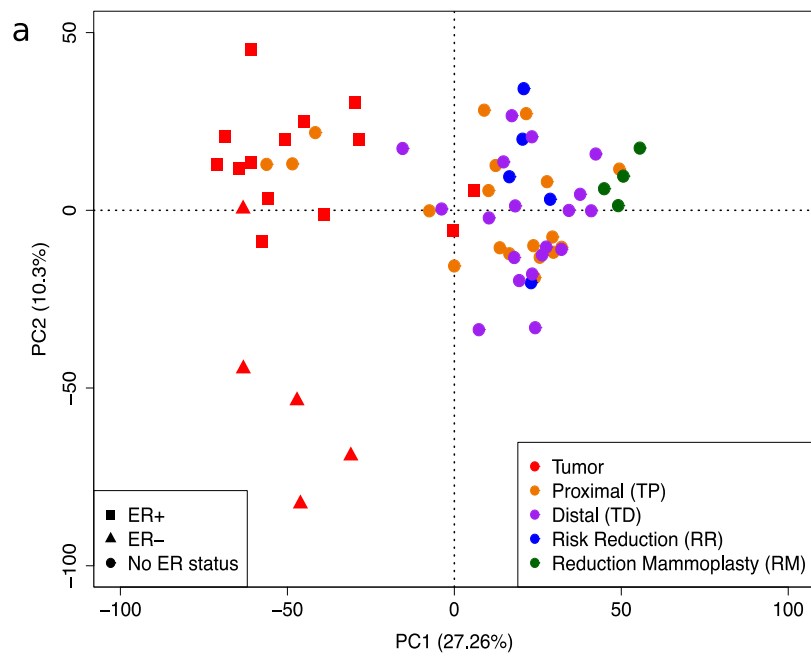
Supplementary Data Set 2 PAM50 probability scores.

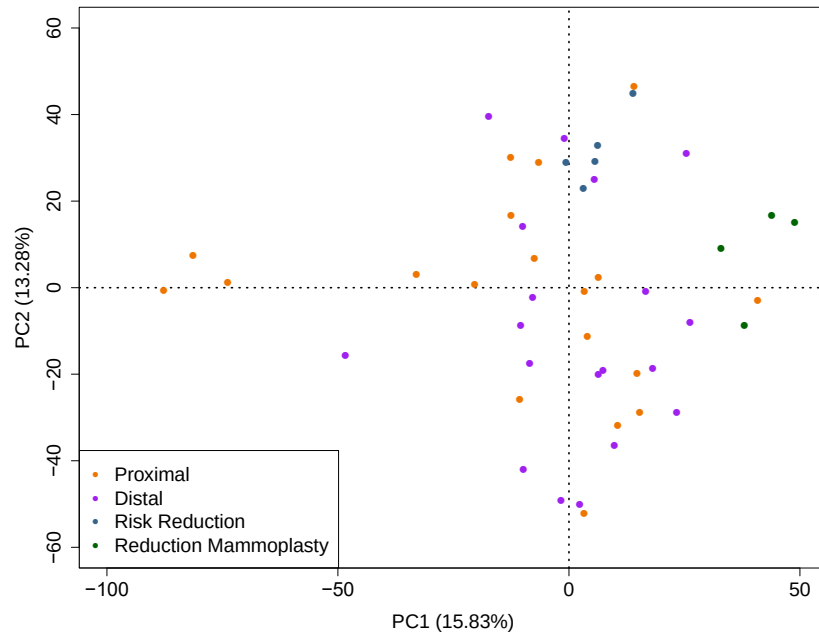
Supplementary Data Set 3 Gene classifier.

Supplementary Data Set 4 NCI pathways enriched and depleted in each subtype.

Supplementary Data Set 5 Non-coding RNAs defining the nc-enriched subtype.

Supplementary Data Set 6 Genes with prognostic potential in histologically normal tissues.

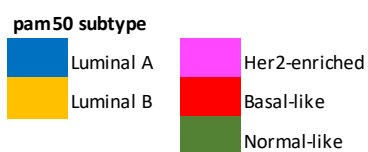




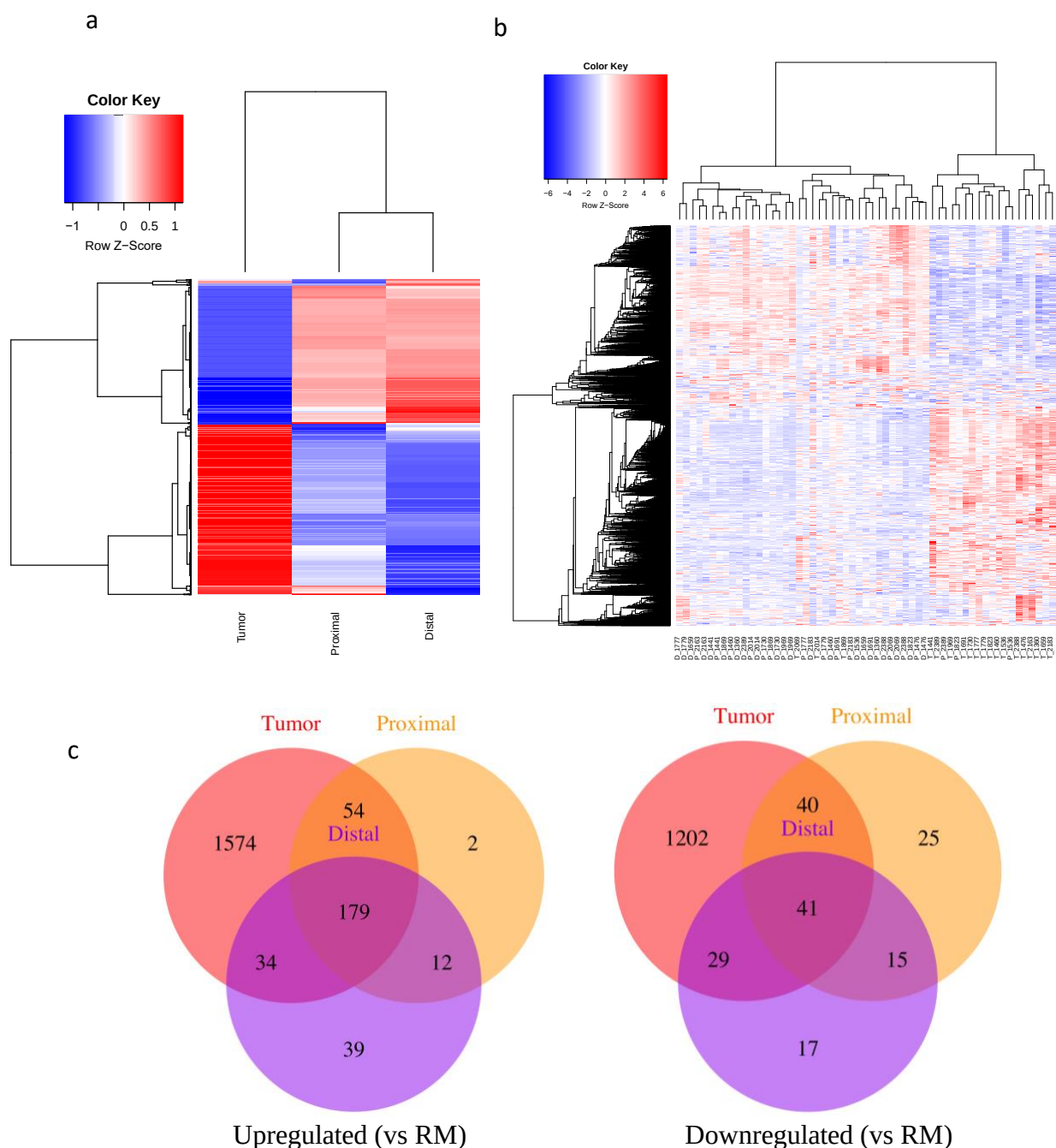
Supplementary Figure 1 PCA of the transcriptomic landscape of morphologically normal tissues. The transcriptomic landscape of breast cancer and morphologically normal tissues displays a wide dispersion of sample profiles across the main principal axes. **a**, Each point represents the orientation of a specimen projected into transcriptional space, with colors highlighting the relative location of the specimen. The shape of the tumor points represents ER status. Tumor profiles aggregate in distinct quadrants of transcriptional space, as do RR and RM. Profiles from HN tissues are interspersed and spread across all quadrants. **b**, There is a significant difference in the first PC between tumor and RM versus all other profiles. **c**, The spread of the data in the second PC is explained by ER status. **d**, There is a wide spread of proximal and distal profiles across the main principal axes. RR and RM specimens tend to exhibit homogeneity within their own groups, bar RM patient 2024.

	Luminal A													Luminal B		Triple Negative			
Patient	1441	1460	1476	1536	1659	1730	1777	1779	1823	1969	2014	2069	2389	1691	1869	1360	2163	2183	2388
Tumor																			
Proximal																			
Distal																			

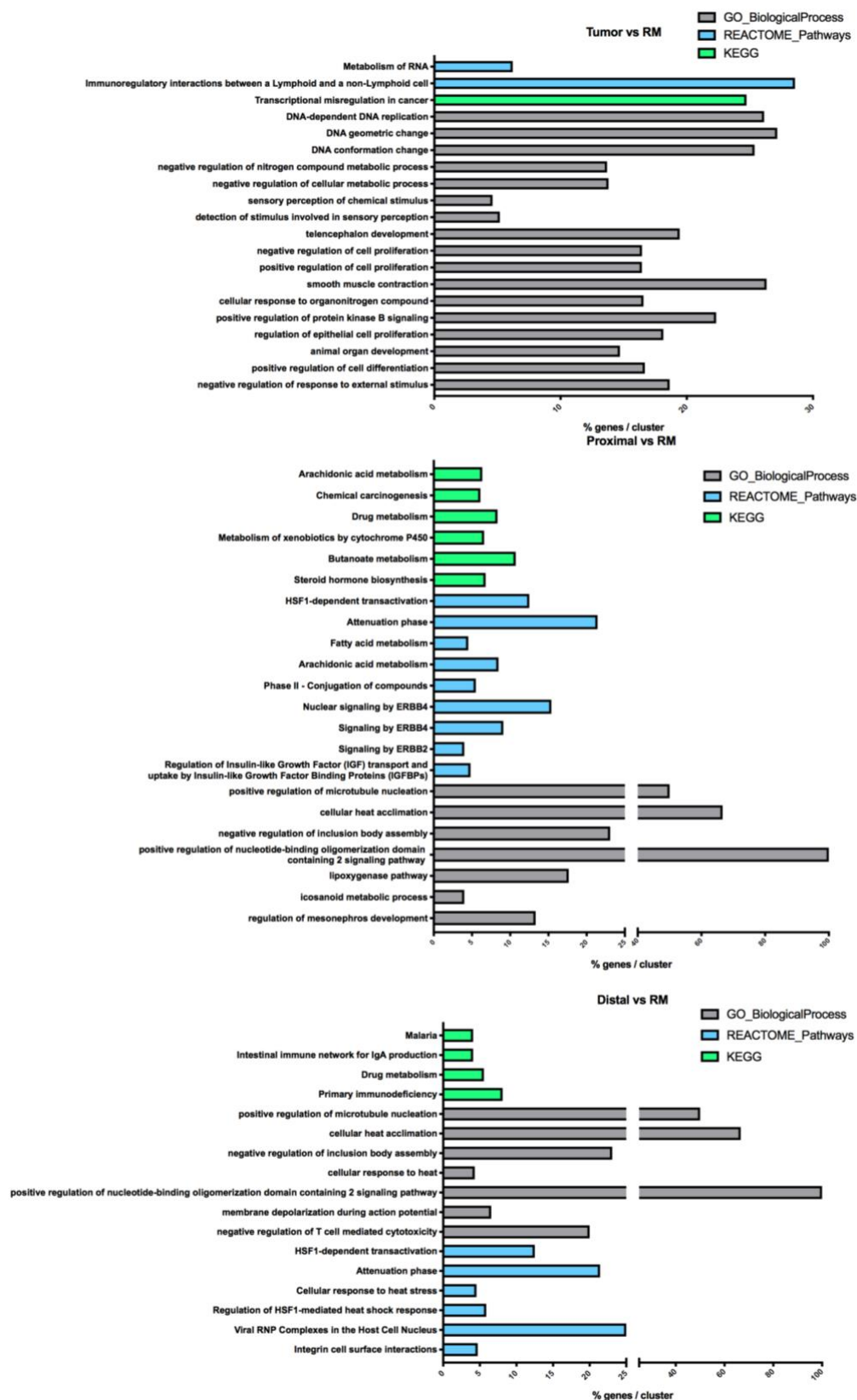
	Risk Reduction					Reduction Mammoplasty			
Patient	2101	2113	2204	1991	1795	1492	2182	1218	1989



Supplementary Figure 2 Molecular classification of proximal and distal samples, with patients sorted based on the clinical classification of their primary tumor.

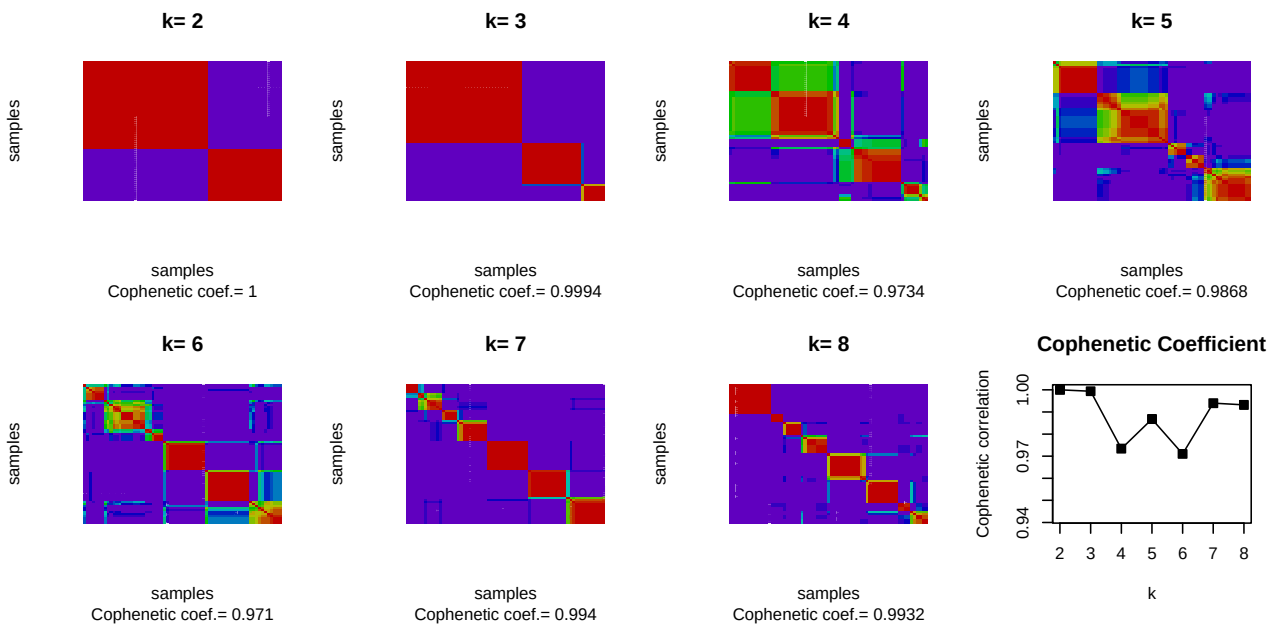


Supplementary Figure 3 Differential expression analysis of tumor, and matched normal tissues, using RM as baseline. **a**, Heatmap of the genes differentially expressed between each spatial group and RM shows dysregulation in tumor and matched tissues ($FDR \leq 0.05$, log fold change ≥ 1). There is overlap in the differential expression profiles for the TP and TD groups. **b**, The individual-level heatmap reflects observations from the PCA. While tumors tend to cluster together, extratumoral specimens do not cluster based on their spatial location or patient. **c**, The total number of genes differentially expressed decreases as distance from primary tumor increases.

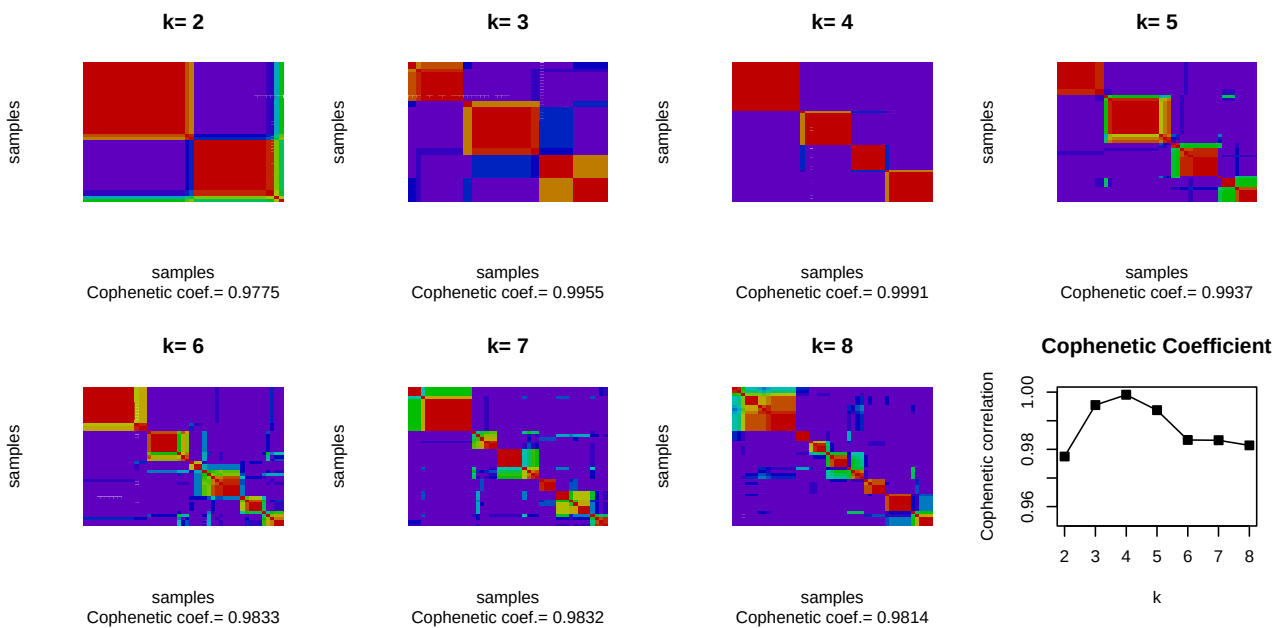


Supplementary Figure 4 GO terms and pathways enriched in each spatial group. The ClueGO Cytoscape plugin was employed to determine GO terms and pathways enriched in genes differentially expressed between each spatial group relative to RM ($p_{adj} < 0.05$). More than 600 cancer-associated pathways were reported enriched in the tumor group. For clarity, the top 20 most significant results pertaining to this group are presented.

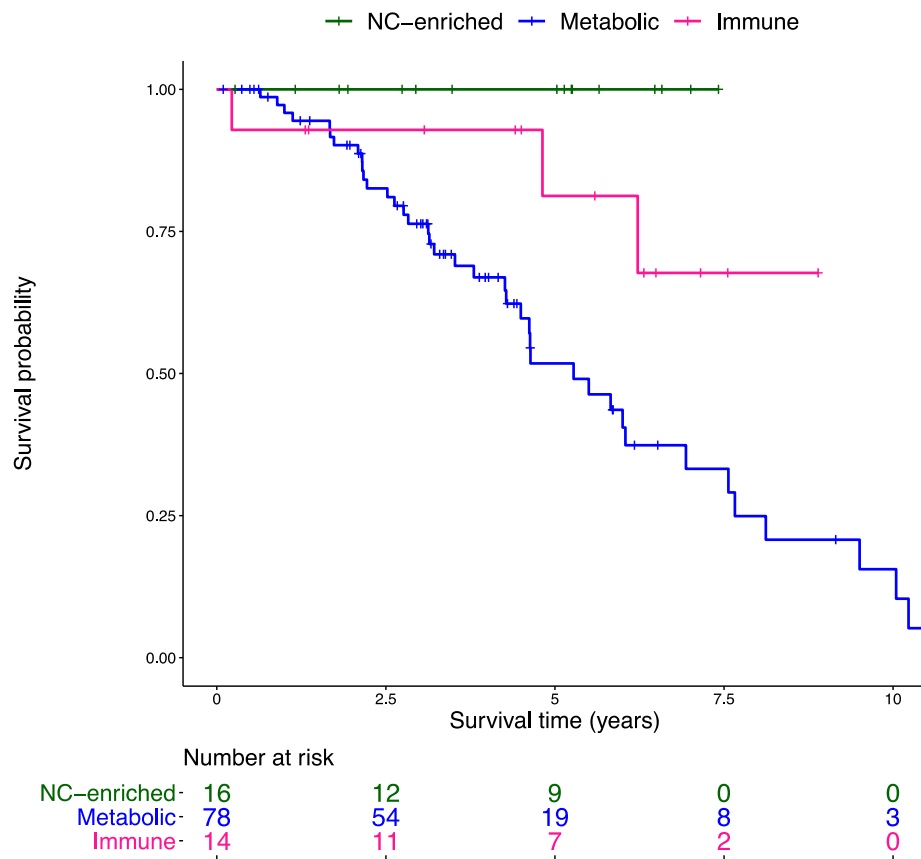
a



b



Supplementary Figure 5 NMF consensus clustering of the samples. **a**, Unsupervised NMF consensus clustering of tumor and HN samples for identify peak cophenetic coefficient for k=2 classes (cophenetic coefficient 1). **b**, NMF clustering of HN samples for solutions k=2 to k=8 identify four stable clusters (cophenetic coefficient 0.9991).



Supplementary Figure 6 Survival analysis indicates that the metabolic group has worse prognosis (log-rank $p < 0.001$, hazard ratio 6.1) relative to the immune and the nc-enriched subtype. No events are present in the nc-enriched group. However, the small sample size prevents any conclusive interpretations to be made.