Original Article

Spatial transcriptomics reveals the two cancer stem cell-like populations in triple-negative breast cancer

Jun Nakayama[1,2,3,*], Hiroko Matsunaga[4], Koji Arikawa[3,4], Takuya Yoda[2,4], Masahito Hosokawa[2,3,4], Haruko Takeyama[2,3,4], Yusuke Yamamoto[1], Kentaro Semba[2,5,*]

1 Laboratory of Integrative Oncology, National Cancer Center Research Institute, Tokyo, Japan.

2 Department of Life Science and Medical Bioscience, School of Advanced Science and Engineering, Waseda University, Tokyo, Japan.

3 Computational Bio-Big Data Open Innovation Lab. (CBBD-OIL), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan.

4 Research Organization for Nano & Life Innovation, Waseda University, Tokyo, Japan.

5 Translational Research Center, Fukushima Medical University, Fukushima, Japan.

*Correspondence to: Jun Nakayama (junakaya@ncc.go.jp, jnakayama.re@gmail.com) and Kentaro Semba (ksemba@waseda.jp)

Abstract

Gene expression analysis at the single-cell scale by next generation sequencing has revealed the existence of clonal dissemination in cancer metastasis. The current spatial

analysis technologies elucidate the heterogeneity of cell-cell interactions *in situ*; however, further analysis is needed to elucidate the nature of tumor heterogeneity. To reveal the expressional heterogeneity and cell-cell interactions in primary tumors and metastases, we performed transcriptomic analysis of microtissues dissected from a triple-negative breast cancer (TNBC) cell line MDA-MB-231 xenograft model by our automated tissue microdissection punching technology. This multiple-microtissue transcriptome analysis revealed that there were existed three cell-type clusters in the primary tumor and axillary lymph node metastasis, two of which were cancer stem cell-like clusters (CD44/MYC-high, HMGA1-high). The CD44/MYC-high cluster showed aggressive proliferation with MYC expression. The HMGA1-high cluster exhibited HIF1A activation and upregulation of ribosomal processes. Furthermore, we developed a cell-cell Interaction (CCI) analysis to investigate the ligand-receptor interactions (cancer cell to stroma and stroma to cancer cell) in each spot. The CCI analysis revealed the interaction dynamics generated by the combination of cancer cells and stromal cells in primary tumors and metastases. Two cancer stem cell-like populations were also detected by the scRNA-seq analysis of TNBC patients. In addition, the gene signature of the HMGA1-high cancer stem cell-like cluster has the potential to serve as a novel biomarker for diagnosis. The mixture of these multiple cancer stem cell-like populations may cause differential anticancer drug resistance, increasing the difficulty of curing this cancer.

Introduction

Breast cancer cells metastasize to multiple distant organs, such as the axillary lymph nodes, lungs, bone, liver, and brain (Nakayama *et al*, 2021; Obenauf & Massague, 2015). In particular, metastasis to axillary lymph nodes is an indicator of cancer grade in breast cancer patients (Giuliano *et al*, 2011). Most breast cancer tissues, including distant metastases, exhibit genetic heterogeneity (McGranahan & Swanton, 2017). Single-cell analyses have revealed that cancer cells evolve through the acquisition of genomic mutations in the primary tumor and metastases (Echeverria *et al*, 2018; Yates *et al*, 2017). Most previous analyses have been performed using isolated cancer cells and stromal cells from cancer tissues. Thus, the cell-cell interactions between cancer cells and stromal cells remains to be analyzed. In recent studies, the current single-cell analysis and spatial transcriptome technologies reveal the heterogeneity of cell-cell interactions between cancer cells and stromal cells *in situ* (Andersson *et al*, 2021; Wu *et al*, 2021); however, further analysis is needed to elucidate the nature of tumor heterogeneity.

Comprehensive gene expression analysis of metastases harvested from approximately 500 specimens in a various cancer types and metastatic organs (MET500 cohort) have suggested that metastatic tissues could be divided into several categories (e.g. proliferative or EMT-like/inflammatory) (Robinson *et al*, 2017). In particular, some samples were found to show signatures of more than one category, suggesting that these samples had micro intraheterogeneity. To clarify such heterogeneity, microtissue sectioning using the laser capture microdissection has often

been performed (Civita *et al*, 2019). This method has several disadvantages, however, including the laborious and time-consuming nature of sample handling and a high risk of RNA degradation. Thus, in previous work, we developed a system involving automated tissue microdissection punching followed by transcriptomic analysis of the tumor microtissue (Yoda *et al*, 2017). To analyze the expressional heterogeneity in microtissues from the primary tumor and axillary lymph node metastases, we performed analysis of the spatial microtissue transcriptome in the triple-negative breast cancer (TNBC) cell line MDA-MB-231 xenograft model. We focused on the expression profiles of known metastasis-promoting genes and cancer stem cell markers in dissected microtissues.

Results

Sampling microtissues from primary tumors and axillary lymph node metastases in MDA-MB-231 xenografts

Primary tumors and axillary lymph node metastases were harvested from NOD-SCID mice with MDA-MB-231-parent-*Venus* cell line xenografts. We subjected the sliced tissues to microtissue dissection by an automated tissue microdissection punching system (Figure 1A). RNA was successfully recovered from the microtissues collected at 93 spots in the primary lesion and 44 spots in axillary lymph node metastases using a microtissue automatic sampling device (Figure 1B). In samples of this size, although the number of cells present in the tumor tissue varies, it can be inferred that several to

approximately 10-30 cells are present in each spot (Yoda *et al.*, 2017). RNA-seq analysis was performed on the total RNA extracted from each spot. We checked the quality of the fastq files by FASTQC. Total RNA samples contained RNA from human cancer cell lines and RNA from mouse stromal cells in the tumor microenvironment. Therefore, the obtained sequences were mapped to both the human reference genome and the mouse reference genome by HISAT2 (Kim *et al*, 2019). Protein-coding genes (human: 19961 genes, mouse: 22050 genes) were extracted as transcripts per million (TPM) for spatial transcriptome analysis with Seurat (Butler *et al*, 2018; Stuart *et al*, 2019) (Figure 1C, Supplementary Figure S1A, S1B and S1C).

Analysis of microtissue transcriptomes

The clustering analysis and UMAP plots showed 3 clusters of cancer cells (transcripts mapped to the human reference genome) and 4 clusters of stromal cells (transcripts mapped to the mouse reference genome) in the microspots dissected from primary tumors and axillary lymph node metastases (Figure 2A and B). Next, we evaluated the expression of cancer stem cell markers to focus on cell–cell interactions in the metastatic stem cell niche (Oskarsson *et al*, 2014). We found that human cancer clusters showed specific gene expression patterns for high mobility group AT-Hook1 (HMGA1), CD44, and MYC (Figure 2C). Consequently, these human clusters were named HMGA1-high, CD44/MYC-high, and Marker-low. Mouse stromal clusters showed specific gene expression patterns for transthyretin (Ttr), Cd3d (T-cell marker), membrane spanning 4-domains a1 (Ms4a1, B-cell marker), and inhibin subunit beta A

(Inhba; a subunit of both activin and inhibin) (Figure 2D). Ttr and Inhba were highly expressed in their respective specific clusters (Supplementary Table S2). Therefore, these mouse clusters were named Ttr-high, Tcell-like, Inhba-high, and Bcell-like. CD44 was broadly expressed in all human clusters; however, HMGA1 was expressed in only HMGA1-high clusters (Figure 2E). CD44 and HMGA1 are well-known markers of cancer stem cells in breast cancer (Liu *et al*, 2010; Pegoraro *et al*, 2013). These results suggested that 2 types of cancer stem cell-like populations existed in MDA-MB-231 xenograft models.

The heterogeneity of each cluster was demonstrated by calculation of network topology using the normalized closeness centrality (Watanabe *et al*, 2020). The centralities showed the correlation of gene expression in each cluster. These results showed that the cancer cells in the HMGA1 clusters had a the expanded diverse of expressional heterogeneity compared with CD44/MYC clusters (Supplementary Figure S1D and S1E). Interestingly, although 3 human clusters were present in both the primary tumors and the lymph node metastases (Figure 2F), mouse stromal clusters showed a site-specific pattern. Most of the Ttr-high clusters were observed in the primary tumors. On the other hand, most Tcell-like clusters and B-cell-like clusters were found in the lymph node metastases (Figure 2G).

Next, we performed spot analysis with the spatial information to determine the spatial heterogeneity in the xenografts (Figure 3A, 3B, and 3C). Human CD44/MYC-high cancer cells tend to localize the outside of the primary tumor. However, lymph node metastases did not show such a tendency (Figure 3A). Cell cycle analysis

of cancer cells showed that cell proliferation occurred outside of the primary tumor and at sparse sites among the lymph node metastases (Figure 3B). Approximately 50 % of the cells in CD44/MYC-high clusters and 30% of the cells in HMGA1-high clusters were actively undergoing cell division (Supplementary Figure S2A, 3D and Supplementary Table S3). These results suggested that the cells divided from two cancer stem cell-like clusters, leading to cancer expressional heterogeneity. Mouse stromal cell localization showed that most Tcell-like clusters were present throughout the entire lymph node metastases; on the other hand, Tcell-like clusters also exited the outside of the primary tumor. Most Ttr-high clusters were present at the sparse primary tumor (Figure 3C). The mouse Tcell-like clusters and Inhba-high clusters were recruited into the cell cycle (Supplementary Figure S2B, S2C, S2D and Supplementary Table S3). Next, a comparative analysis of the cell cycle in the primary tumor and lymph node metastasis showed that the two cancer stem cell-like clusters (HMGA1-high and CD44/MYC-high) increased the cell division index in lymph node metastases (Figure 3E). In contrast, Marker-low clusters did not change the cell cycle index in either location. This result suggested that the cancer stem-like cells in metastatic tissues proliferated aggressively.

Enrichment analysis of microspots

The differentially expressed genes (DEGs) in human clusters and mouse clusters were extracted and visualized in heatmaps (Figure 4A, Supplementary Figure S51, S5B, Table S2 and S4). The human HMGA1-high cluster showed that high expression of TMSB10 (Zhang *et al*, 2017), CTSD (Ashraf *et al*, 2019) and LGALS1

(Balestrieri *et al*, 2021; Jung *et al*, 2007), which are correlated with poor prognosis in breast cancer. The human CD44/MYC-high cluster expressed SENPK, SENPN and PTK2 (focal adhesion kinase: FAK), which regulate the cell cycle and cell division. The human Marker-low clusters showed that low expression levels of these genes. The mouse Tcell-like clusters expressed cytokines and immune receptors. To determine the biological function of DEGs, we performed upstream analysis, GO enrichment analysis and pathway enrichment analysis using Metascape (Zhou *et al*, 2019). DEGs from each human cluster showed the upregulated DEGs in the HMGA1-high clusters and CD44/MYC-high clusters, and downregulated DEGs in the Marker-low clusters (Figure 4A, Supplementary Table S2 and S4). Upstream analysis of DEGs showed that HIF1A downstream genes were upregulated in the HMGA1-high cluster, and MYC downstream genes were upregulated in the CD44/MYC-high cluster. In the Marker-low clusters, E2F1 downstream genes were downregulated (Figure 4B). We performed pathway and GO enrichment analyses focused on the two cancer stem cell-like clusters that had upregulated DEGs. Amide metabolites, VEGFA-VEGFR signaling, and apoptosis regulation were enriched in both clusters. On the other hand, the DEGs in CD44/MYC-high clusters were enriched in many terms related to the cell cycle and cell division, namely, cell division, cell cycle, and telomeres. In HMGA1-high clusters, ribosome assembly and ribosome biogenesis were significantly enriched (Figure 4C, 4D and Supplementary Figure S3). The results of enrichment analysis using DEGs in the Marker-low clusters showed that metabolism of RNA, translation, and mitochondrial organization were enriched (Supplementary Figure S4).

176 In mouse clusters, enrichment analysis of the DEGs was performed by using
177 Metascape. Enrichment analysis of the DEGs in the Ttr-high clusters showed no
178 enriched terms. The Tcell-like clusters and Inhba-high clusters had two common
179 enriched terms; glycolysis and neutrophil degradation (Supplementary Figure S5).
180
181 **Cell-cell interaction (CCI) analysis of microspots**
182 To estimate the CCI score, we utilized a cell-cell interaction database
183 (https://baderlab.org/) and extracted a total of 115,900 interactions (3,209 ligands,
184 4,364 receptors and 433 extracellular matrix) (Watanabe *et al.*, 2020). Highly expressed
185 genes (expression level > 2) were selected for CCI analysis, and ligand-receptor
186 interactions (human to mouse and mouse to human) were estimated in each spot
187 (Figure 5A). We extracted 2,432 interactions and 7 clusters of “mouse (stromal) to
188 human (cancer)” CCI patterns (Figure 5B and Supplementary Table S7). In particular,
189 Cluster 2 was constituted by only lymph node metastasis spots, while Cluster 6 was
190 constituted by only primary tumor spots. We thus focused on these clusters: the CCI
191 heatmap showed lymph node metastasis-specific CCI (Figure 5C and 5D) and primary
192 tumor-specific CCI (Figure 5E and 5F). The “human (cancer) to mouse (stromal)” CCI
193 pattern showed 741 interactions (Supplementary Figure S6, Supplementary Table S7
194 and S8).
195 Next, we focused on cell type-specific CCI. Ten sets of cancer cell-stromal
196 cell interactions were observed (Figure 6A and 6B). The PT-1, PT-2 and PT-3 groups
197 existed in only the primary tumors; in contrast, the Mix-1~7 groups were present in both

198 the primary tumors and lymph node metastases (Figure 6C). As a result, 4,126
199 interactions “cancer to stromal” (Supplementary Table S8) and 4,165 interactions of
200 “mouse (stromal) to human (cancer)” (Supplementary Table S9) were estimated from
201 the cell type-specific CCI analysis. The ligand-receptor interaction with the highest CCI
202 score was shown for each group (Figure 6D-G). Among cancer cell-to-stromal
203 interactions, annexin A2 (ANXA2) and heat shock protein 90 alpha (HSP90AA1) had
204 high CCI scores in various groups. On the other hand, from stromal to cancer cell
205 interactions, B2m interactions were estimated in various groups. Alb interaction was
206 estimated in the Mix5 and Mix6 groups. The interactions of the Tcell-like cluster (Mix-1,
207 Mix-3 and Mix-5) changed depending on the type of cancer cell (Figure 6E and 6G).
208 Similarly, the Ttr-high (PT-1, PT-2 and Mix-4) cluster interactions also changed
209 depending on the type of cancer cell. On the other hand, in view of the stromal to cancer
210 interaction, the Mix5 and Mix6 clusters have specific interactions with proteins such as
211 integrin B1 (ITGB1) and cystatin C3 (CST3) (Figure 6G).
212
213 **TNBC patients showed two cancer stem cell-like populations**
214 To confirm our findings in clinical samples, we reanalyzed the public single-cell
215 RNA-seq (scRNA-seq) dataset (Karaayvaz *et al*, 2018) and the Molecular Taxonomy of
216 Breast Cancer International Consortium (METABRIC) (Pereira *et al*, 2016). First, we
217 analyzed the scRNA-seq dataset of 5 TNBC patients (Figure 7A). We extracted 546
218 cancer cells from the dataset with UMAP visualization (Figure 7B). Module analysis was
219 performed using the HMGA1 signatures and CD44/MYC signatures from DEG analysis

220 to detect the HMGA1-high cluster and CD44/MYC cluster in TNBC patients. TNBC
221 patients had the two cancer stem cell-like populations and a double-positive population
222 (Figure 7C, 7D, Supplementary Tables S10 and S11). Next, survival analysis was
223 performed using the HMGA1 signatures and CD44/MYC signatures with METABRIC
224 claudin-low subtype (TNBC) cohorts. High expression of HMGA1 signatures correlated
225 with poor prognosis in the claudin-low subtype (Figure 7E). In contrast, CD44/MYC
226 signatures did not correlate with prognosis in these cohorts (Supplementary Figure S7).
227
228
229 **Discussion**

230 Spatial transcriptomics technologies (Rodriques *et al*, 2019; Yoda *et al.*, 2017)
231 have enabled us to reveal the *in situ* expressional profiles and microheterogeneity of
232 cancer. In particular, in a xenograft model, both human-derived RNA and
233 mouse-derived RNA can be analyzed both simultaneously and individually by mapping
234 the sequence reads to a human genome reference or mouse genome reference
235 (Bradford *et al*, 2013; Callari *et al*, 2018). In this study, by combining microtissue
236 sampling and the isolation of human-mouse gene expression by mapping, we revealed
237 the expressional heterogeneity of cancer cells and stromal cells, and the heterogeneity
238 of cancer-stroma interactions in MDA-MB-231 primary tumors and the axillary lymph
239 node metastases.

240 Interestingly, we observed two types of cancer stem cell-like populations in
241 both the primary tumors and lymph node metastases. One of the cancer stem cell-like

242 populations expressed CD44 and MYC. The CD44 gene is a well-known cancer stem
243 cell marker in breast cancer (Liu *et al.*, 2010; Marotta *et al*, 2011; Sheridan *et al*, 2006).
244 Most MDA-MB-231 cells expressed high levels of CD44 (Burdick *et al*, 2012), although
245 the expression levels varied (Figure 2E). The cluster with the highest CD44 expression
246 was MYC, and these cells proliferated aggressively *in vivo*. The other cancer stem
247 cell-like population, the HMGA1-high cluster, was observed in both the primary tumors
248 and lymph node metastases. HMGA1 promotes tumor initiation, cancer stemness and
249 metastasis in TNBC (Huang *et al*, 2015; Pegoraro *et al.*, 2013; Shah *et al*, 2013). The
250 DEGs enrichment analysis indicated that HIF1A activation and ribosomal-related
251 bioprocesses were enriched in the HMGA1-high cancer stem cell-like population. HIF1A
252 is a well-known regulator of hypoxia that activates stemness, glycolysis, angiogenesis,
253 and invasion/metastasis (Choudhry & Harris, 2018; Petrova *et al*, 2018). In terms of
254 ribosomal-related processes, the upregulation of translation and ribosomal processes
255 may promote distant metastasis in breast cancer (Ebright *et al*, 2020). In the clinical
256 scRNA-seq analysis, both types of cancer stem cell-like populations were observed in
257 single-cell analysis of TNBC patients (Karaayvaz *et al.*, 2018). Our results showed that
258 the mixture of these multicancer stem cell-like populations makes curative treatment
259 difficult and causes the anticancer drug resistance in the clinic. In addition, the HMGA1
260 signatures has the potential to be a novel biomarker for diagnosis, and HMGA1-high
261 cancer stem cells may contribute to poor prognosis.

262 Our results showed Ttr-high and Inhba-high stromal populations in the
263 xenograft model. High expression of transthyretin (Ttr) enhances tumor proliferation and

264 growth (Lee *et al*, 2019). Inhba is a member of the TGF-beta superfamily (Bloise *et al*,
265 2019). Inhba is upregulated in breast tumors, and induces epithelial-mesenchymal
266 transition (EMT), tumor growth and distant metastasis (Bashir *et al*, 2015; Kalli *et al*,
267 2019). Most Inhba-high populations also existed in the primary tumor. Our results
268 suggested that stromal expression of Ttr and Inhba enhanced tumor growth in the
269 primary tumors of MDA-MB-231 xenografts.

270 CCI analysis of each the spot and group was performed to understand the
271 interaction dynamics of different combination of cancer cell and stromal cell types. This
272 analysis will be helpful for discovering the cancer stem cell niche and metastatic niche
273 (Oskarsson *et al.*, 2014). B2M is a gene that presents self-antigens on the plasma
274 membrane. Cancer cells present self-antigens to immune cells in the tumor tissue
275 (Popat *et al*, 2020). B2M has a different molecular regulatory mechanism in ER-positive
276 and ER-negative breast cancer, and it controls the proliferation of cancer cells (Chai *et*
277 *al*, 2019). Our results suggested that MDA-MB-231 cells interact with immune cells
278 through B2M antigen presentation, which controls cancer cell proliferation in the
279 xenograft model. In addition, HSP90AA1 and ANXA2 expressed by cancer cells had
280 high CCI scores in the spot CCI analysis. High expression of HSP90AA1 in TNBC or
281 HER2-/ER+ breast cancer patients is correlated with poor prognosis, and the
282 HSP90AA1 gene is often amplified (Cheng *et al*, 2012). HSP90AA1 is secreted in
283 extracellular vesicles under hypoxia and enhances the migration of cancer cells and
284 stromal cells in breast cancer (Santos *et al*, 2017). HSP90AA1 may contribute to the
285 involvement of extracellular proteins in the cell–cell interactions. High expression of

ANXA2 is correlated with poor prognosis in TNBC patients (Gibbs & Vishwanatha, 2018) and regulates drug resistance to EGFR1-targeted therapy (Fan *et al*, 2019; Zhang *et al*, 2018). ANXA2 controls angiogenesis in TNBC xenografts and has the potential to be a novel therapeutic target in TNBC (Sharma & Jain, 2020). ANXA2 is a regulator of endocytosis and exocytosis on the plasma membrane (Bharadwaj *et al*, 2013; Grindheim *et al*, 2017). This interaction has the potential to mediate cell-cell communication via exosomes, and thereby promotes the migration of cancer cells and immune cells.

Several spots did not contain enough RNA for analysis or exhibited bias toward either human RNA or mouse RNA (Supplementary Figure S1A). Thus, one limitation of this sampling method is that some spots have a biased cell type or no cells. Read counts of each cluster showed lower counts for the human Marker-low cluster, mouse Ttr-high cluster, mouse Inhba-high cluster, and mouse B-cell like cluster than for other clusters (Supplementary Figure S1B and S1C). The limitations of cell type bias and low RNA extraction efficiency caused these low transcript counts.

Our results showed that HMGA1 signatures correlated with poor prognosis in TNBC patients in the METABRIC cohort; on the other hand, CD44/MYC signatures did not correlate with progression (Supplementary Figure S7). Previous research reported that the diagnosis of invasive breast cancer on the basis of CD44 expression alone is difficult and that it is necessary to examine the expression of other genes (Mylona *et al*, 2008; Wang *et al*, 2017). Our study supports that CD44 signatures are not sufficient for diagnosis. In contrast, HMGA1 signatures is useful for the diagnosis of TNBC patients.

Our results showed that these spatial transcriptomics methods will be helpful for the diagnosis, further identification of biomarkers, and elucidation of the essential characteristics of cancer.

Materials and Methods

Cell culture

The MDA-MB-231-*luc2-Venus* cell line was cultured in RPMI-1640 (Fujifilm Wako, Osaka, Japan) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Fujifilm Wako), 100 µg/ml streptomycin (Meiji Seika Pharma Co. Ltd. Tokyo, Japan) and 100 U/ml penicillin (Meiji Seika Pharma) at 37°C with 5% CO_2.

Animal studies

A. breast cancer xenograft model was established in NOD.CB-17-Prkdc^{scid}/J mice (NOD-SCID; Charles River Laboratories Japan, Inc., Kanagawa, Japan) by orthotopic transplantation as previously described (Nakayama *et al*, 2017). A total of 1.0 x 10^6 cells were injected into the 4th fat pad of NOD-SCID mice. The primary tumor was removed 8 weeks after transplantation. An axillary lymph node metastasis was sampled 2 weeks after removing the primary tumor. The growth of the primary tumors and metastases were monitored by bioluminescence using an in vivo imaging system (IVIS-XRMS, PerkinElmer, MA, USA). For bioluminescence monitoring by IVIS, mice were anesthetized with 2.5% isoflurane (Fujifilm Wako) and intraperitoneally injected with

3 mg D-luciferin (Gold Biotechnology Inc., MO, USA) in 200 µl PBS as previously described (Han *et al*, 2020; Nakayama *et al*, 2020). The harvested organs were placed in ice-cold PBS (Fujifilm Wako) and embedded in SCEM (Super Cryoembedding Medium, SECTION-LAB, Japan) using liquid nitrogen and stored at -80 degree until sectioning.

Microtissue dissection and RNA-seq analysis

Microtissue sampling was performed by an automated tissue microdissection punching system as previously described (Yoda *et al.*, 2017). Frozen sections were sliced at a thickness of 20 µm and transferred on an LMD film II (SECTIOIN-LAB). Microspots were sampled with a 100 µm needle in the dissection instrument. RNA-seq was performed by Illumina HiSeq as previously described (Yoda *et al.*, 2017).

Mapping and quality check

Transcriptome analysis was performed with HISAT2 version 2.0.5 (Kim *et al.*, 2019) and RSEM version 1.3.0 (Li & Dewey, 2011). The gene expression of cancer cells was obtained by mapping RNA-sequence reads to the human reference genome or mouse reference genome. We subjected 'protein_coding' genes to spatial transcriptome analysis.

Clustering and UMAP visualization

Data mining analyses such as clustering, UMAP analysis, and DEG extraction were performed with the functions 'runPCA', 'FindNeigbors', 'FindClusters', and 'runUMAP' and 'FindAllMarkers' in 'Seurat' version 3.2. (Stuart *et al.*, 2019). Cell cycle detection was performed by the function 'CellCycleScoring'. Heatmap drawing was performed using 'ComplexHeatmap'(Gu *et al*, 2016). These packages and functions were run in R version 3.6.3.

Network analysis for expressional heterogeneity

Correlational network analysis for calculation of the expressional heterogeneity was performed by the 'igraph' package as previously described (Nakayama *et al.*, 2017; Watanabe *et al.*, 2020). We calculated Pearson's correlational coefficients between the spots classified into the same clusters. Next we calculated the normalized closeness centrality using the correlational network.

Enrichment analysis using DEGs

Pathway and GO enrichment analyses were performed by the Metascape (https://metascape.org/gp/index.html#/main/step1) (Zhou *et al.*, 2019). DEGs from each cluster were subjected to the Metascape interface. Differential enrichment terms were analyzed by multiple gene list mode. The results of enrichment analysis were visualized as heatmaps and networks.

Cell-Cell Interaction (CCI) analysis

Ligand-receptor interactions between human cancer cells and mouse stromal cells were performed using the interaction database of the Bader laboratory from Toronto University (https://baderlab.org/CellCellInteractions#Download_Data) in R software version 3.6.3. For spot CCI analysis, we extracted the genes whose expression value was greater than 2. We selected the combinations representing ligand-receptor interactions, in which both ligand genes and receptor genes were expressed in the same spot. Hierarchical clustering was performed by 'hclust' and it was visualized as a circular clustering plot by the 'circlize' package in R (Gu *et al*, 2014).

In the CCI analysis of the group with containing both human cancer cells and mouse stromal cells, we calculated the number of spots with expression values greater than 2. Only groups whose expression cell ratio exceeded 10% were extracted for CCI analysis, and the CCI score between each group was calculated as previously described (Watanabe *et al.*, 2020).

Analysis of the public single-cell RNA-seq dataset

To confirm the cancer stem cell signatures in MDA-MB-231 xenografts, we performed reanalysis of scRNA-seq of TNBC patients from the public cohort (Karaayvaz *et al.*, 2018). The normalized scRNA-seq dataset was downloaded the GSE138390 dataset from the Gene Expression Omnibus (GEO) and analyzed with the annotation metadata. The dataset was analyzed and visualized by the UMAP plot with Seurat in R. Module analysis was performed using the function 'AddModuleScore' with gene signatures (Supplementary Table S10) in 'Seurat'.

395

396 **Clinical dataset analysis**

397 Survival analysis of the Molecular Taxonomy of Breast Cancer International Consortium
398 (METABRIC) cohort (Ali *et al*, 2020; Curtis *et al*, 2012; Pereira *et al.*, 2016) was
399 performed by the Kaplan-Meier method using the 'ggplot2', 'survminer' and 'survival'
400 packages with R as previously described (Kuroiwa *et al*, 2020; Murakami *et al*, 2019;
401 Nishiyama *et al*, 2021)

402

403 **Code availability**

404 The source code of spatial transcriptome analysis is available on GitHub
405 (https://github.com/JunNakayama/Spatial-Transcriptomics-of-MDA-MB-231-xenografts)
406 .

407

408 **Data availability**

409 Spatial expression data (RNA-seq of each microspot) were deposited at GEO
410 accession number GSE184720.

411

412

Acknowledgements

We thank Prof. Semba's laboratory, Prof. Takeyama's laboratory and CBBD-OIL (AIST) for meaningful discussion and Mr. Kiyofumi Takahashi and Ms. Chikako Sakanashi for technical support.

Funding

This study was supported in part by JSPS KAKENHI (Grant Numbers: 18K16269, 20J01794, and 21K15562 to J.N.), and Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP19am0101004. The supercomputing resource was provided by the Human Genome Center (University of Tokyo).

Authors' Contributions

JN designed the study, performed the animal experiments and analysis of spatial transcriptomics, and wrote the manuscript. HM and TY performed the dissection of micro-tissues. KA performed analysis of NGS data. MH designed the experiment of micro-tissue dissection and interpreted the results. HT, YY and KS interpreted the results. All authors reviewed the manuscript.

Competing Interests

The authors declare that no conflict of interest exists.

435 **Reference**

436 Ali HR, Jackson HW, Zanotelli VRT, Danenberg E, Fischer JR, Bardwell H,

437 Provenzano E, Ali HR, Al Sa'd M, Alon S *et al* (2020) Imaging mass cytometry

438 and multiplatform genomics define the phenogenomic landscape of breast

439 cancer. *Nat Cancer* 1: 163-175

440 Andersson A, Larsson L, Stenbeck L, Salmén F, Ehinger A, Wu SZ, Al-Eryani G,

441 Roden D, Swarbrick A, Borg Å *et al* (2021) Spatial deconvolution of

442 HER2-positive breast cancer delineates tumor-associated cell type interactions.

443 *Nat Commun* 12: 6012

444 Ashraf Y, Mansouri H, Laurent-Matha V, Alcaraz LB, Roger P, Guiu S, Derocq D,

445 Robin G, Michaud HA, Delpech H *et al* (2019) Immunotherapy of triple-negative

446 breast cancer with cathepsin D-targeting antibodies. *J Immunother Cancer* 7: 29

447 Balestrieri K, Kew K, McDaniel M, Ramez M, Pittman HK, Murray G, Vohra NA,

448 Verbanac KM (2021) Proteomic identification of tumor- and

449 metastasis-associated galectin-1 in claudin-low breast cancer. *Biochim Biophys*

450 *Acta Gen Subj* 1865: 129784

451 Bashir M, Damineni S, Mukherjee G, Kondaiah P (2015) Activin-A signaling
452 promotes epithelial-mesenchymal transition, invasion, and metastatic growth of
453 breast cancer. *NPJ Breast Cancer* 1: 15007

454 Bharadwaj A, Bydoun M, Holloway R, Waisman D (2013) Annexin A2
455 heterotetramer: structure and function. *Int J Mol Sci* 14: 6259-6305

456 Bloise E, Ciarmela P, Dela Cruz C, Luisi S, Petraglia F, Reis FM (2019) Activin A
457 in Mammalian Physiology. *Physiol Rev* 99: 739-780

458 Bradford JR, Farren M, Powell SJ, Runswick S, Weston SL, Brown H, Delpuech
459 O, Wappett M, Smith NR, Carr TH *et al* (2013) RNA-Seq Differentiates Tumour
460 and Host mRNA Expression Changes Induced by Treatment of Human Tumour
461 Xenografts with the VEGFR Tyrosine Kinase Inhibitor Cediranib. *PLOS ONE* 8:
462 e66003

463 Burdick MM, Henson KA, Delgadillo LF, Choi YE, Goetz DJ, Tees DF, Benencia
464 F (2012) Expression of E-selectin ligands on circulating tumor cells:
465 cross-regulation with cancer stem cell regulatory pathways? *Front Oncol* 2: 103

466 Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating
467 single-cell transcriptomic data across different conditions, technologies, and
468 species. *Nat Biotechnol* 36: 411-420

469 Callari M, Batra AS, Batra RN, Sammut S-J, Greenwood W, Clifford H, Hercus C,
470 Chin S-F, Bruna A, Rueda OM *et al* (2018) Computational approach to
471 discriminate human and mouse sequences in patient-derived tumour xenografts.
472 *BMC Genomics* 19: 19-19

473 Chai D, Li K, Du H, Yang S, Yang R, Xu Y, Lian X (2019) β2-microglobulin has a
474 different regulatory molecular mechanism between ER(+) and ER(-) breast
475 cancer with HER2(). *BMC Cancer* 19: 223

476 Cheng Q, Chang JT, Geradts J, Neckers LM, Haystead T, Spector NL, Lyerly
477 HK (2012) Amplification and high-level expression of heat shock protein 90
478 marks aggressive phenotypes of human epidermal growth factor receptor 2
479 negative breast cancer. *Breast Cancer Res* 14: R62

480 Choudhry H, Harris AL (2018) Advances in Hypoxia-Inducible Factor Biology.
481 *Cell Metab* 27: 281-298

482 Civita P, Franceschi S, Aretini P, Ortenzi V, Menicagli M, Lessi F, Pasqualetti F,
483 Naccarato AG, Mazzanti CM (2019) Laser Capture Microdissection and
484 RNA-Seq Analysis: High Sensitivity Approaches to Explain Histopathological
485 Heterogeneity in Human Glioblastoma FFPE Archived Tissues. *Front Oncol* 9:
486 482

487 Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D,
488 Lynch AG, Samarajiwa S, Yuan Y *et al* (2012) The genomic and transcriptomic
489 architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486:
490 346-352

491 Ebright RY, Lee S, Wittner BS, Niederhoffer KL, Nicholson BT, Bardia A,
492 Truesdell S, Wiley DF, Wesley B, Li S *et al* (2020) Deregulation of ribosomal
493 protein expression and translation promotes breast cancer metastasis. *Science*
494 367: 1468-1473

495 Echeverria GV, Powell E, Seth S, Ge Z, Carugo A, Bristow C, Peoples M,
496 Robinson F, Qiu H, Shao J *et al* (2018) High-resolution clonal mapping of
497 multi-organ metastasis in triple negative breast cancer. *Nat Commun* 9: 5079

498 Fan Y, Si W, Ji W, Wang Z, Gao Z, Tian R, Song W, Zhang H, Niu R, Zhang F
499 (2019) Rack1 mediates tyrosine phosphorylation of Anxa2 by Src and promotes
500 invasion and metastasis in drug-resistant breast cancer cells. *Breast Cancer*
501 *Res* 21: 66

502 Gibbs LD, Vishwanatha JK (2018) Prognostic impact of AnxA1 and AnxA2 gene
503 expression in triple-negative breast cancer. *Oncotarget* 9: 2697-2704

504 Giuliano AE, Hunt KK, Ballman KV, Beitsch PD, Whitworth PW, Blumencranz
505 PW, Leitch AM, Saha S, McCall LM, Morrow M (2011) Axillary dissection vs no
506 axillary dissection in women with invasive breast cancer and sentinel node
507 metastasis: a randomized clinical trial. *Jama* 305: 569-575

508 Grindheim AK, Saraste J, Vedeler A (2017) Protein phosphorylation and its role
509 in the regulation of Annexin A2 function. *Biochim Biophys Acta, Gen Subj* 1861:
510 2515-2529

511 Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and
512 correlations in multidimensional genomic data. *Bioinformatics* 32: 2847-2849

513 Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize implements and
514 enhances circular visualization in R. *Bioinformatics* 30: 2811-2812

515 Han Y, Nakayama J, Hayashi Y, Jeong S, Futakuchi M, Ito E, Watanabe S,
516 Semba K (2020) Establishment and characterization of highly osteolytic luminal
517 breast cancer cell lines by intracaudal arterial injection. *Genes Cells* 25: 111-123

518 Huang R, Huang D, Dai W, Yang F (2015) Overexpression of HMGA1 correlates
519 with the malignant status and prognosis of breast cancer. *Mol Cell Biochem* 404:
520 251-257

521 Jung EJ, Moon HG, Cho BI, Jeong CY, Joo YT, Lee YJ, Hong SC, Choi SK, Ha
522 WS, Kim JW *et al* (2007) Galectin-1 expression in cancer-associated stromal
523 cells correlates tumor invasiveness and tumor progression in breast cancer. *Int J*
524 *Cancer* 120: 2331-2338

525 Kalli M, Mpekris F, Wong CK, Panagi M, Ozturk S, Thiagalingam S,
526 Stylianopoulos T, Papageorgis P (2019) Activin A Signaling Regulates IL13Rα2
527 Expression to Promote Breast Cancer Metastasis. *Front Oncol* 9: 32

528 Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC,
529 Specht MC, Bernstein BE, Michor F, Ellisen LW (2018) Unravelling subclonal
530 heterogeneity and aggressive disease states in TNBC through single-cell
531 RNA-seq. *Nat Commun* 9: 3588

532 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome
533 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37:
534 907-915

535 Kuroiwa Y, Nakayama J, Adachi C, Inoue T, Watanabe S, Semba K (2020)
536 Proliferative Classification of Intracranially Injected HER2-positive Breast
537 Cancer Cell Lines. *Cancers (Basel)* 12: 1811

538 Lee CC, Ding X, Zhao T, Wu L, Perkins S, Du H, Yan C (2019) Transthyretin
539 Stimulates Tumor Growth through Regulation of Tumor, Immune, and
540 Endothelial Cells. *J Immunol* 202: 991-1002

541 Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq
542 data with or without a reference genome. *BMC Bioinformatics* 12: 323

543 Liu H, Patel MR, Prescher JA, Patsialou A, Qian D, Lin J, Wen S, Chang YF,
544 Bachmann MH, Shimono Y *et al* (2010) Cancer stem cells from human breast
545 tumors are involved in spontaneous metastases in orthotopic mouse models.
546 *Proc Natl Acad Sci U S A* 107: 18115-18120

547 Marotta LL, Almendro V, Marusyk A, Shipitsin M, Schemme J, Walker SR,
548 Bloushtain-Qimron N, Kim JJ, Choudhury SA, Maruyama R *et al* (2011) The
549 JAK2/STAT3 signaling pathway is required for growth of $CD44^+CD24^-$ stem
550 cell-like breast cancer cells in human tumors. *J Clin Invest* 121: 2723-2735

551 McGranahan N, Swanton C (2017) Clonal Heterogeneity and Tumor Evolution:
552 Past, Present, and the Future. *Cell* 168: 613-628

553 Murakami A, Maekawa M, Kawai K, Nakayama J, Araki N, Semba K, Taguchi T,
554 Kamei Y, Takada Y, Higashiyama S (2019) Cullin-3/KCTD10 E3 complex is
555 essential for Rac1 activation through RhoB degradation in human epidermal
556 growth factor receptor 2-positive breast cancer cells. *Cancer Sci* 110: 650-661

557 Mylona E, Giannopoulou I, Fasomytakis E, Nomikos A, Magkou C, Bakarakos P,
558 Nakopoulou L (2008) The clinicopathologic and prognostic significance of

559 CD44+/CD24(-/low) and CD44-/CD24+ tumor cells in invasive breast
560 carcinomas. *Hum Pathol* 39: 1096-1102

561 Nakayama J, Han Y, Kuroiwa Y, Azuma K, Yamamoto Y, Semba K (2021) The
562 In Vivo Selection Method in Breast Cancer Metastasis. *Int J Mol Sci* 22

563 Nakayama J, Ito E, Fujimoto J, Watanabe S, Semba K (2017) Comparative
564 analysis of gene regulatory networks of highly metastatic breast cancer cells
565 established by orthotopic transplantation and intra-circulation injection. *Int J*
566 *Oncol* 50: 497-504

567 Nakayama J, Saito R, Hayashi Y, Kitada N, Tamaki S, Han Y, Semba K, Maki
568 SA (2020) High Sensitivity In Vivo Imaging of Cancer Metastasis Using a
569 Near-Infrared Luciferin Analogue seMpai. *Int J Mol Sci* 21

570 Nishiyama K, Maekawa M, Nakagita T, Nakayama J, Kiyoi T, Chosei M,
571 Murakami A, Kamei Y, Takeda H, Takada Y *et al* (2021) CNKSR1 serves as a
572 scaffold to activate an EGFR phosphatase via exclusive interaction with
573 RhoB-GTP. *Life Sci Alliance* 4

574 Obenauf AC, Massague J (2015) Surviving at a Distance: Organ-Specific
575 Metastasis. *Trends Cancer* 1: 76-91

576 Oskarsson T, Batlle E, Massagué J (2014) Metastatic Stem Cells: Sources,
577 Niches, and Vital Pathways. *Cell Stem Cell* 14: 306-321

578 Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del
579 Sal G, Manfioletti G (2013) HMGA1 promotes metastatic processes in basal-like
580 breast cancer regulating EMT and stemness. *Oncotarget* 4: 1293-1308

581 Pereira B, Chin S-F, Rueda OM, Vollan H-KM, Provenzano E, Bardwell HA,
582 Pugh M, Jones L, Russell R, Sammut S-J *et al* (2016) The somatic mutation
583 profiles of 2,433 breast cancers refine their genomic and transcriptomic
584 landscapes. *Nat Commun* 7: 11479

585 Petrova V, Annicchiarico-Petruzzelli M, Melino G, Amelio I (2018) The hypoxic
586 tumour microenvironment. *Oncogenesis* 7: 10

587 Popat S, Grohé C, Corral J, Reck M, Novello S, Gottfried M, Radonjic D, Kaiser
588 R (2020) Anti-angiogenic agents in the age of resistance to immune checkpoint

589 inhibitors: Do they have a role in non-oncogene-addicted non-small cell lung
590 cancer? *Lung Cancer* 144: 76-84

591 Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, Cao X, Rabban
592 E, Kumar-Sinha C, Raymond V *et al* (2017) Integrative clinical genomics of
593 metastatic cancer. *Nature* 548: 297-303

594 Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR,
595 Welch J, Chen LM, Chen F, Macosko EZ (2019) Slide-seq: A scalable
596 technology for measuring genome-wide expression at high spatial resolution.
597 *Science* 363: 1463-1467

598 Santos TG, Martins VR, Hajj GNM (2017) Unconventional Secretion of Heat
599 Shock Proteins in Cancer. *Int J Mol Sci* 18

600 Shah SN, Cope L, Poh W, Belton A, Roy S, Talbot CC, Jr., Sukumar S, Huso DL,
601 Resar LM (2013) HMGA1: a master regulator of tumor progression in
602 triple-negative breast cancer cells. *PLoS One* 8: e63419

603 Sharma MC, Jain D (2020) Important role of annexin A2 (ANXA2) in new blood
604 vessel development in vivo and human triple negative breast cancer (TNBC)
605 growth. *Exp Mol Pathol* 116: 104523

606 Sheridan C, Kishimoto H, Fuchs RK, Mehrotra S, Bhat-Nakshatri P, Turner CH,
607 Goulet R, Jr., Badve S, Nakshatri H (2006) CD44+/CD24- breast cancer cells
608 exhibit enhanced invasive properties: an early step necessary for metastasis.
609 *Breast Cancer Res* 8: R59

610 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao
611 Y, Stoeckius M, Smibert P, Satija R (2019) Comprehensive Integration of
612 Single-Cell Data. *Cell* 177: 1888-1902.e1821

613 Wang H, Wang L, Song Y, Wang S, Huang X, Xuan Q, Kang X, Zhang Q (2017)
614 CD44(+)/CD24(-) phenotype predicts a poor prognosis in triple-negative breast
615 cancer. *Oncol Lett* 14: 5890-5898

616 Watanabe N, Nakayama J, Fujita Y, Mori Y, Kadota T, Shimomura I, Ohtsuka T,
617 Okamoto K, Araya J, Kuwano K *et al* (2020) Single-cell Transcriptome Analysis
618 Reveals an Anomalous Epithelial Variation and Ectopic Inflammatory Response

619 in Chronic Obstructive Pulmonary Disease. *medRxiv*:
620 2020.2012.2003.20242412

621 Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A,
622 Thennavan A, Wang C, Torpy JR, Bartonicek N *et al* (2021) A single-cell and
623 spatially resolved atlas of human breast cancers. *Nat Genet* 53: 1334-1347

624 Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I,
625 Alexandrov LB, Van Loo P, Haugland HK, Lilleng PK *et al* (2017) Genomic
626 Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* 32: 169-184

627 Yoda T, Hosokawa M, Takahashi K, Sakanashi C, Takeyama H, Kambara H
628 (2017) Site-specific gene expression analysis using an automated tissue
629 micro-dissection punching system. *Sci Rep* 7: 4325

630 Zhang X, Ren D, Guo L, Wang L, Wu S, Lin C, Ye L, Zhu J, Li J, Song L *et al*
631 (2017) Thymosin beta 10 is a key regulator of tumorigenesis and metastasis and
632 a novel serum marker in breast cancer. *Breast Cancer Res* 19: 15

633 Zhang Y, Bi J, Zhu H, Shi M, Zeng X (2018) ANXA2 could act as a moderator of

634 EGFR-directed therapy resistance in triple negative breast cancer. *Biosci*

635 *Biotechnol Biochem* 82: 1733-1741

636 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner

637 C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the

638 analysis of systems-level datasets. *Nat Commun* 10: 1523

639

640

641 **Figure Legends**

642 **Figure 1. Microtissue sectioning from the primary tumors and the axillary lymph**
643 **node metastases in the TNBC xenograft model.**

644 (A) Experimental flowchart of spatial transcriptomics. The human TNBC cell line
645 MDA-MB-231-Parent-Venus was transplanted orthotopically into a female NOD-SCID
646 mouse. After 8 weeks, the primary tumor was harvested, and the center of the tumor
647 was sectioned. After 4 weeks, axially lymph node metastases were harvested and
648 sectioned from the same mouse. Sectioning was performed by an automated tissue
649 microdissection punching system with a 100 μm needle. (B) In total, 93 microspots were
650 sectioned from the primary tumor, and 43 microspots were sectioned from axillary
651 lymph node metastases. RNA was extracted from a total of 137 spots. (C) A flowchart of
652 the transcriptome analysis. Quality check was performed by FASTQC. The reads were
653 mapped to the human genome reference and the mouse genome reference by HISAT2.
654 Protein-coding genes were selected for analysis with Seurat.

655

656 **Figure 2. Transcriptome profiling of the microspots.**

657 (A) UMAP plot of human (cancer cell) spot clustering. (B) UMAP plot of mouse (stromal
658 cell) spot clustering. (C) Violin plot of cancer marker genes. (D) Violin plot of stromal
659 marker genes. (E) Heatmap of CD44, HMGA1, and MYC expression in each human
660 cluster. (F & G) Bar plot of spot counts in the primary tumor and lymph node
661 metastases: (F) human (G) mouse.

662

663 **Figure 3. Spatial transcriptome with cell cycle detection.**

664 (A) Spatial transcriptomics of human (cancer cell) clusters at the primary tumors and the
665 lymph node metastases. (B) Cell cycle phase of cancer cells at the primary tumor and
666 lymph-node metastases. (C) Spatial transcriptomics of mouse (stromal cell) clusters at
667 the primary tumors and lymph node metastases. (D) Sunburst plot of the cell cycle in
668 human cancer cell clusters. (E) Bar plot of cell cycle phases in the primary tumors and
669 lymph node metastases.

670

671 **Figure 4. Enrichment analysis of two cancer stem cell-like populations.**

672 (A) Heatmap of cluster marker genes in human cancer cell DEGs. (B) Heatmap of
673 enrichment scores in upstream analysis. (C) Heatmap of enrichment analysis in the two
674 cancer stem cell-like populations (HMGA1-high and CD44/MYC-high). (D) Enrichment
675 network in two cancer stem cell-like populations.

676

677 **Figure 5. Spot cell-cell interaction analysis in 'stromal to cancer interaction'.**

678 (A) A flowchart of the CCI analysis. (B) Circular clustering plot of hierarchical analysis of
679 spot CCI analysis (stromal cell to cancer cell). (C & D) Heatmap of CCI in the spots.

680

681 **Figure 6. Cell-cell interaction analysis with cell type combinations.**

682 (A) A table of the combinations of cell types. (B) UMAP plot of the combination group.
683 (C) Bar plot of counts of the group. (D) Circular bar plot of the top CCI
684 (cancer-to-stromal cell interaction) in each group. (E) Bubble chart of the

cancer-to-stromal CCI in each group. (F) Circular bar plot of the top CCI (stromal-to-cancer cell interaction) in each group. (G) Bubble chart of the stromal-to-cancer CCI in each group.

Figure 7. Reanalysis of clinical scRNA-seq and cohorts with cancer stem cell-like signatures.

(A) A flowchart of the reanalysis of a public scRNA-seq dataset. We downloaded GSE118389 (scRNA-seq data of 5 TNBC patients) and analyzed it with Seurat. Log normalization, scaling, PCA and UMAP visualization were performed following the basic protocol in Seurat. To extract the cancer cells, cells expressing EPCAM (epithelial marker) were filtered. (B) UMAP plot of cancer cell from 5 TNBC patients. (C&D) Module analysis of HMGA1-high signatures and CD44/MYC-high signatures with UMAP plots. The pie chart showed the ratio of cells that expressed the signatures. (E) Survival analysis of claudin-low (TNBC) patients in METABRIC cohorts by the Kaplan-Meier method. Survival analysis with the expression of the HMGA1 signatures.

703 **Supplementary Figure Legends**

704 **Supplementary Figure S1. Read counts and heterogeneity per spot.**

705 (A) Read counts of human and mouse transcripts in each spot. (B) Read counts of
706 human cancer clusters. (C) Read counts of mouse stromal clusters. (D) Normalized
707 closeness centrality in each human cancer cluster. (F) Normalized closeness centrality
708 in each mouse stromal cluster.

709

710 **Supplementary Figure S2. Cell cycle phase in mouse clusters.**

711 (A) The ratio of the cell cycle phase of human cancer clusters. (B) Cell cycle phase of
712 mouse stromal cells at the primary tumor and lymph-node metastasis. (C) Sunburst plot
713 of the cell cycle in mouse stromal cell clusters. (D) The ratios of the cell cycle phases of
714 mouse stromal clusters.

715

716 **Supplementary Figure S3. Enrichment network of two cancer stem cell-like**
717 **populations.**

718 (A) Circos plot of DEGs in cancer clusters. The edges showed the overlap of DEGs
719 between each cluster. (B) Network visualization of p values in enrichment analysis.

720

721 **Supplementary Figure S4. Enrichment analysis of Marker-low clusters.**

722 (A) Heatmap of enrichment analysis in Marker-low clusters. (B) Enrichment network in
723 Marker-low clusters.

724

Supplementary Figure S5. Enrichment analysis in mouse stromal clusters.

(A) Circos plot of DEGs. The edges showed the overlap of DEGs between each cluster. (B) Heatmap of DEGs in mouse stromal cell. (C) Heatmap of enrichment analysis in Ttr-high, Tcell-like, and Bcell-like clusters (D) Enrichment network in Ttr-high, Tcell-like, and Bcell-like clusters

Supplementary Figure S6. Spot Cell-Cell Interaction analysis in 'cancer to stromal interaction'

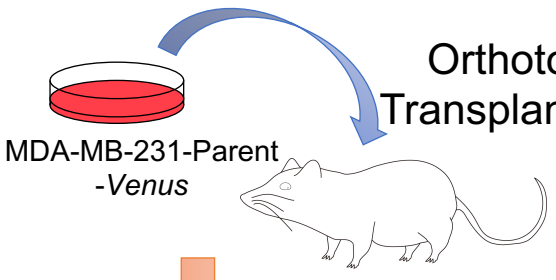
(A) Circular clustering plot of hierarchical analysis of spot CCI analysis (cancer-to-stromal cell). (B & C) Heatmap of CCI in the interactive spots.

Supplementary Figure S7. Survival analysis with the expression of the CD44/MYC signatures

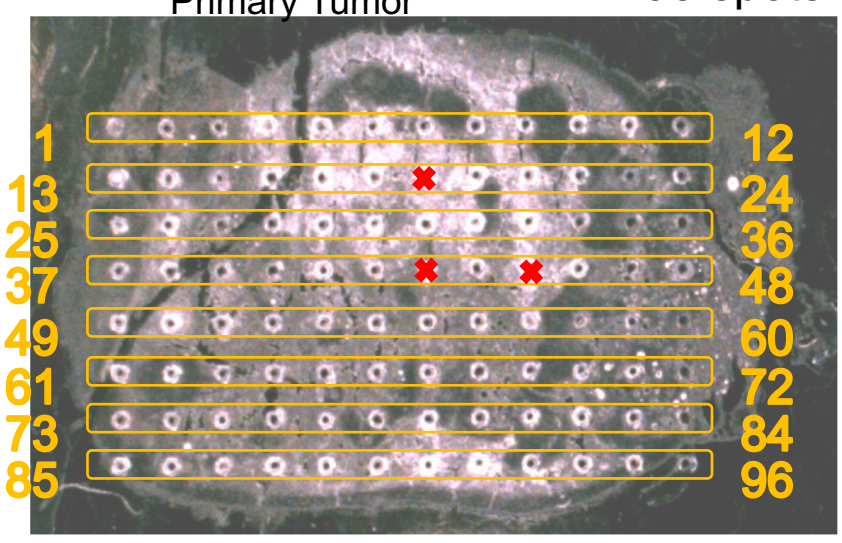
Survival analysis of claudin-low patients in METABRIC cohorts by Kaplan-Meier method. Survival analysis with the expression of the CD44/MYC signatures.

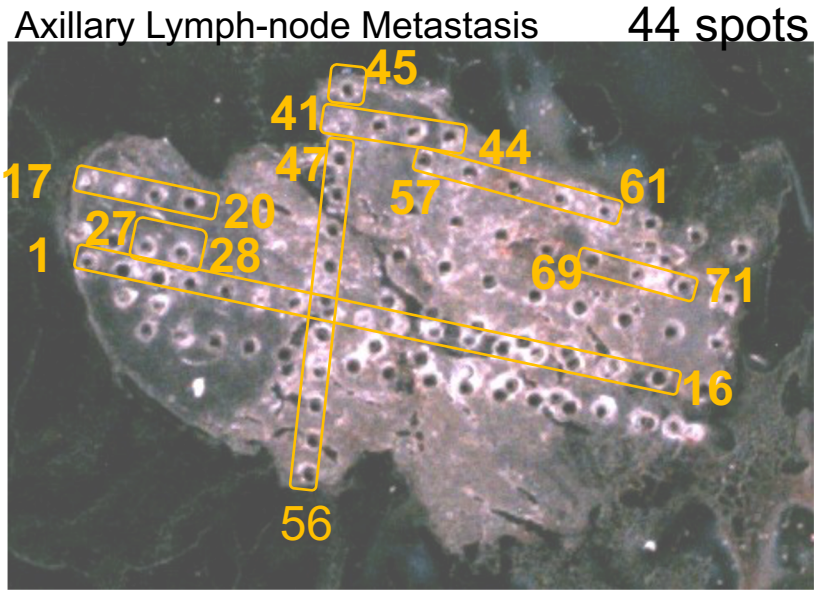
Figure 1

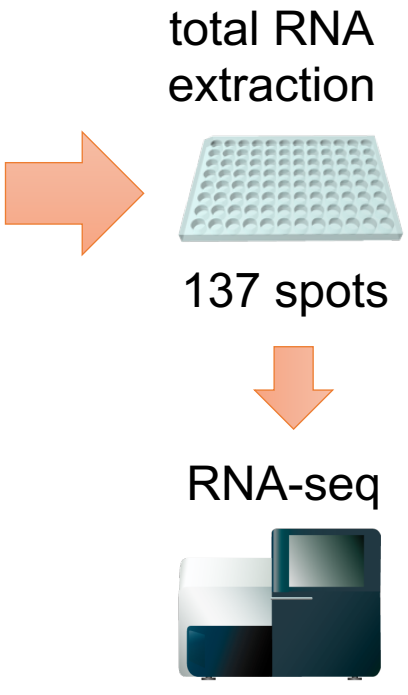
A


MDA-MB-231-Parent
-Venus
Orthotopic
Transplantation
Primary Tumor at
4th Fat Pad
Cut at center of primary tumor
followed dotted line
Automated tissue microdissection
punching system
Control
monitor
PC
Punching hollow
needle
96-well plate
Sample
table
Sample
LED Light
Video camera
Tissue
Scaffold
film
Plastic substrate
Y
X
Section at cut site

B


Primary Tumor
93 spots
1
12
13
24
25
36
37
48
49
60
61
72
73
84
85
96


Axillary Lymph-node Metastasis
44 spots
45
41
44
47
17
57
61
20
27
1
28
69
71
16
56


total RNA
extraction
137 spots
RNA-seq

C

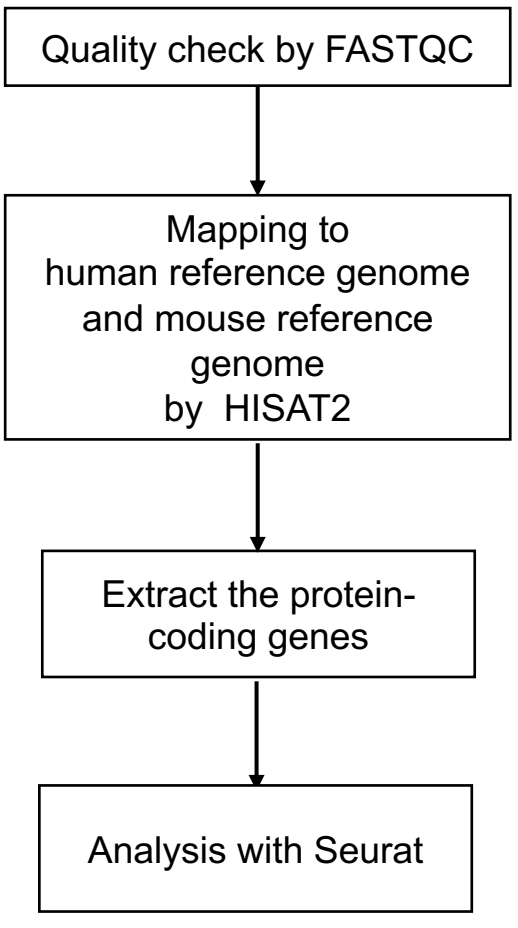
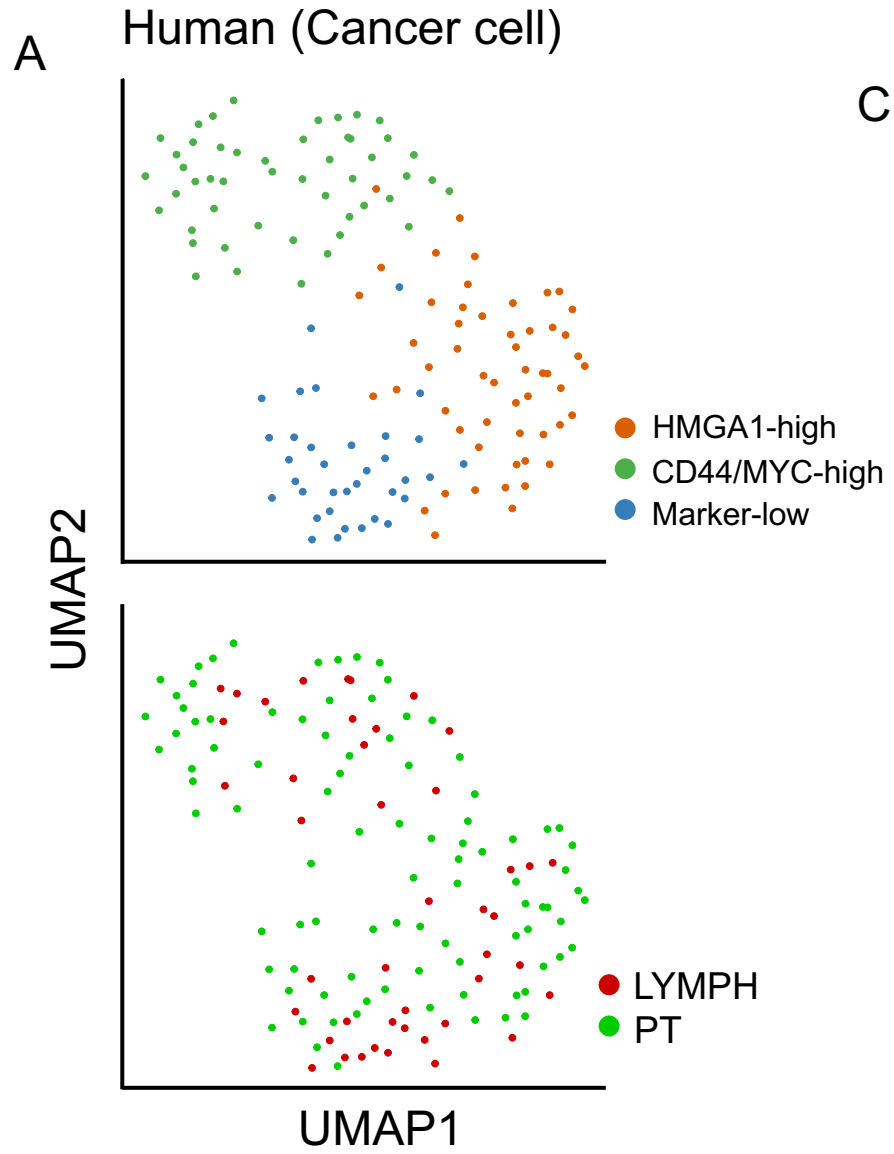
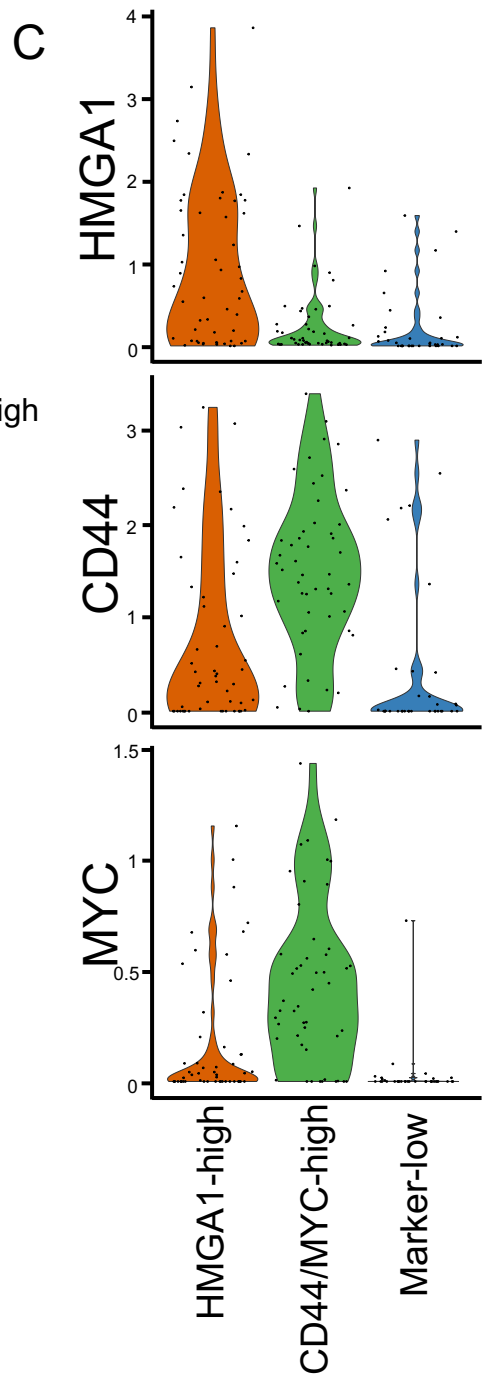
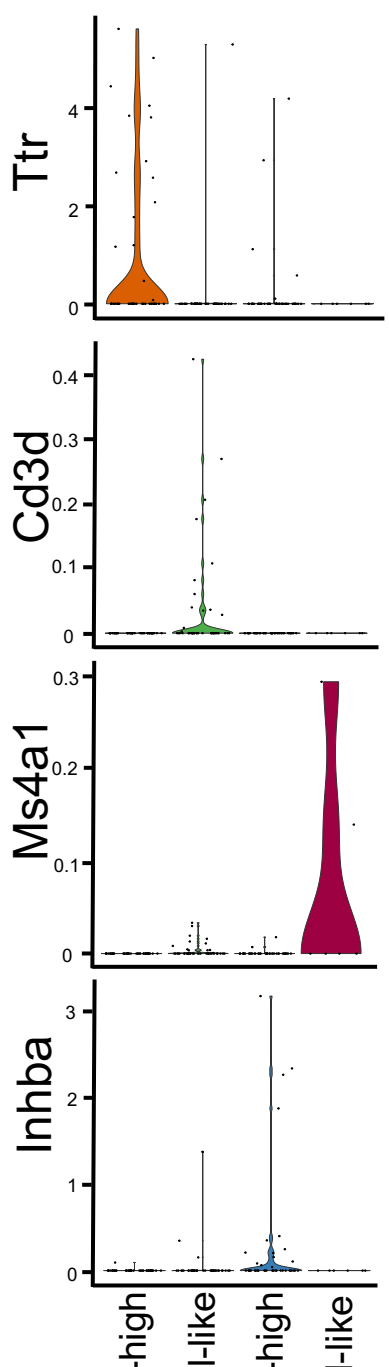
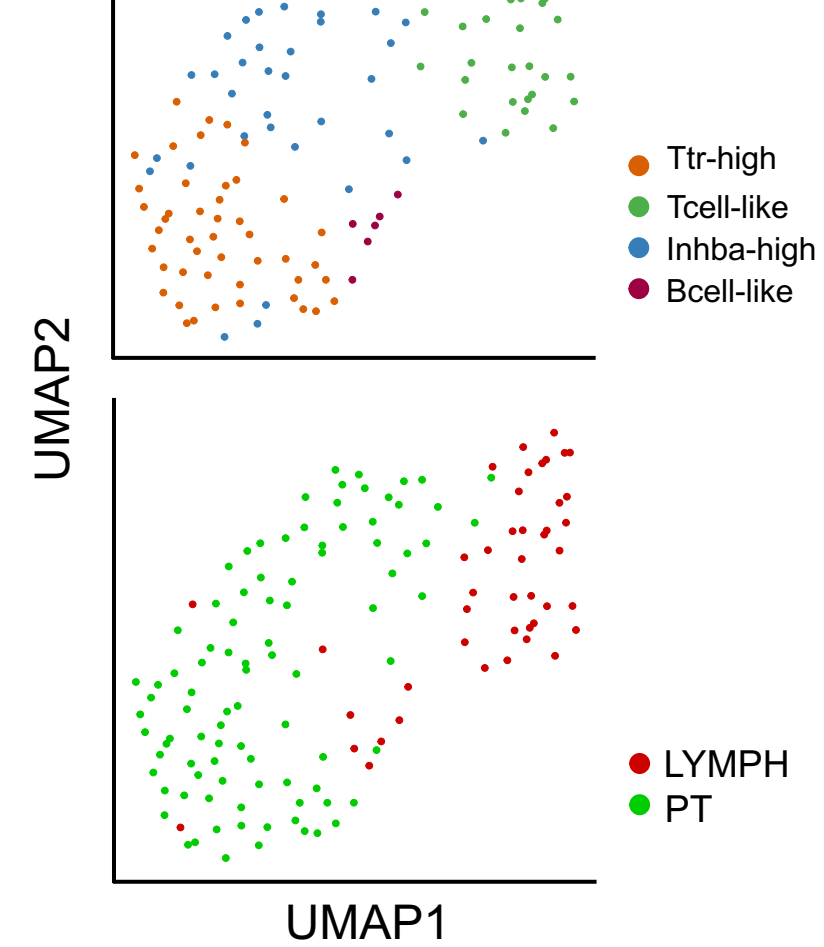

Quality check by FASTQC
Mapping to
human reference genome
and mouse reference
genome
by HISAT2
Extract the protein-
coding genes
Analysis with Seurat

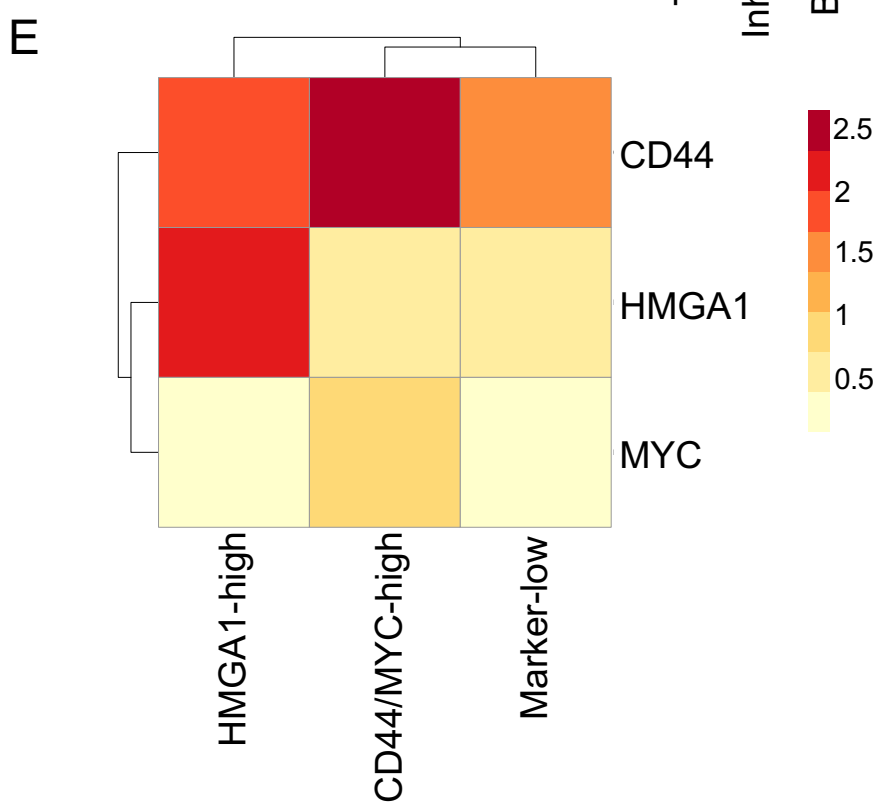
Figure 2

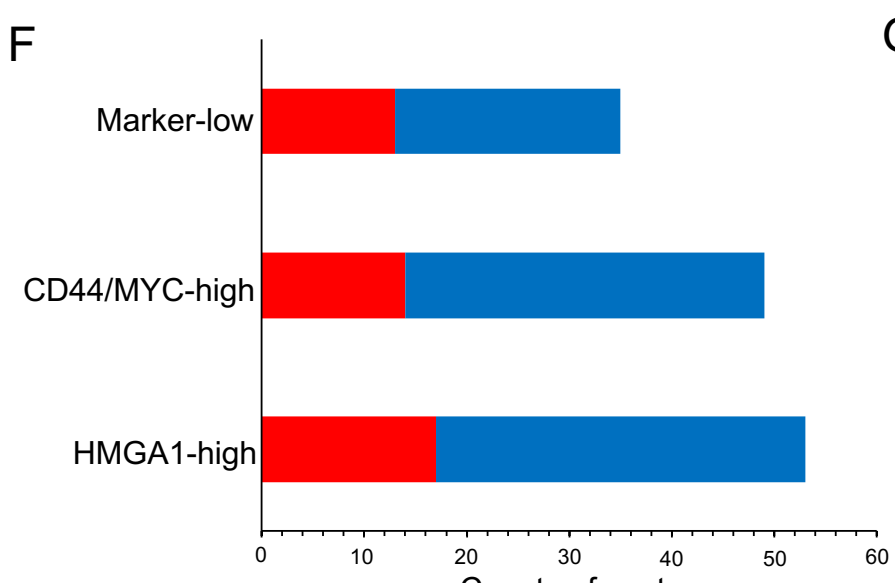

A
Human (Cancer cell)
UMAP2
UMAP1
HMGA1-high
CD44/MYC-high
Marker-low
LYMPH
PT


C
HMGA1
CD44
MYC
HMGA1-high
CD44/MYC-high
Marker-low


D
Ttr
Cd3d
Ms4a1
Inhba
Ttr-high
Tcell-like
Inhba-high
Bcell-like


B
Mouse (Stromal cell)
UMAP2
UMAP1
Ttr-high
Tcell-like
Inhba-high
Bcell-like
LYMPH
PT


E
CD44
HMGA1
MYC
HMGA1-high
CD44/MYC-high
Marker-low


F
Marker-low
CD44/MYC-high
HMGA1-high
Counts of spot

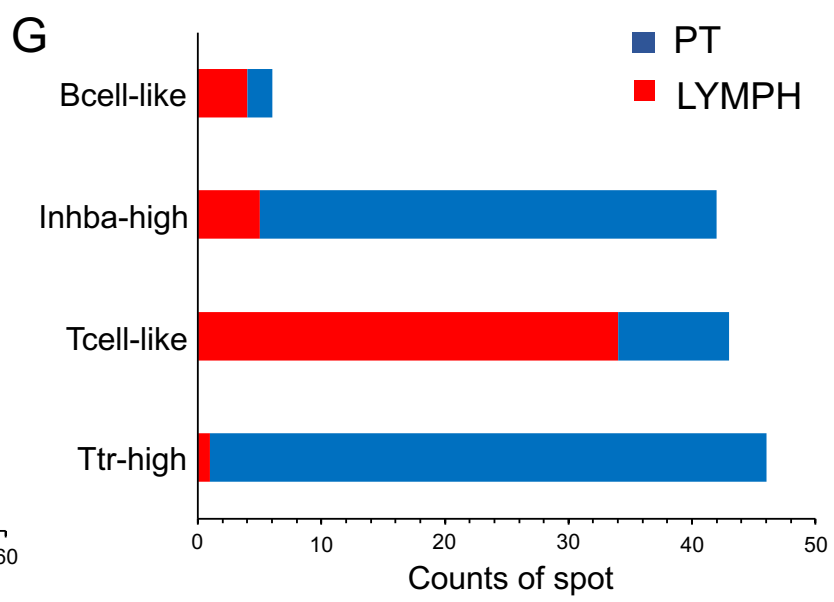

G
PT
LYMPH
Bcell-like
Inhba-high
Tcell-like
Ttr-high
Counts of spot

Figure 3

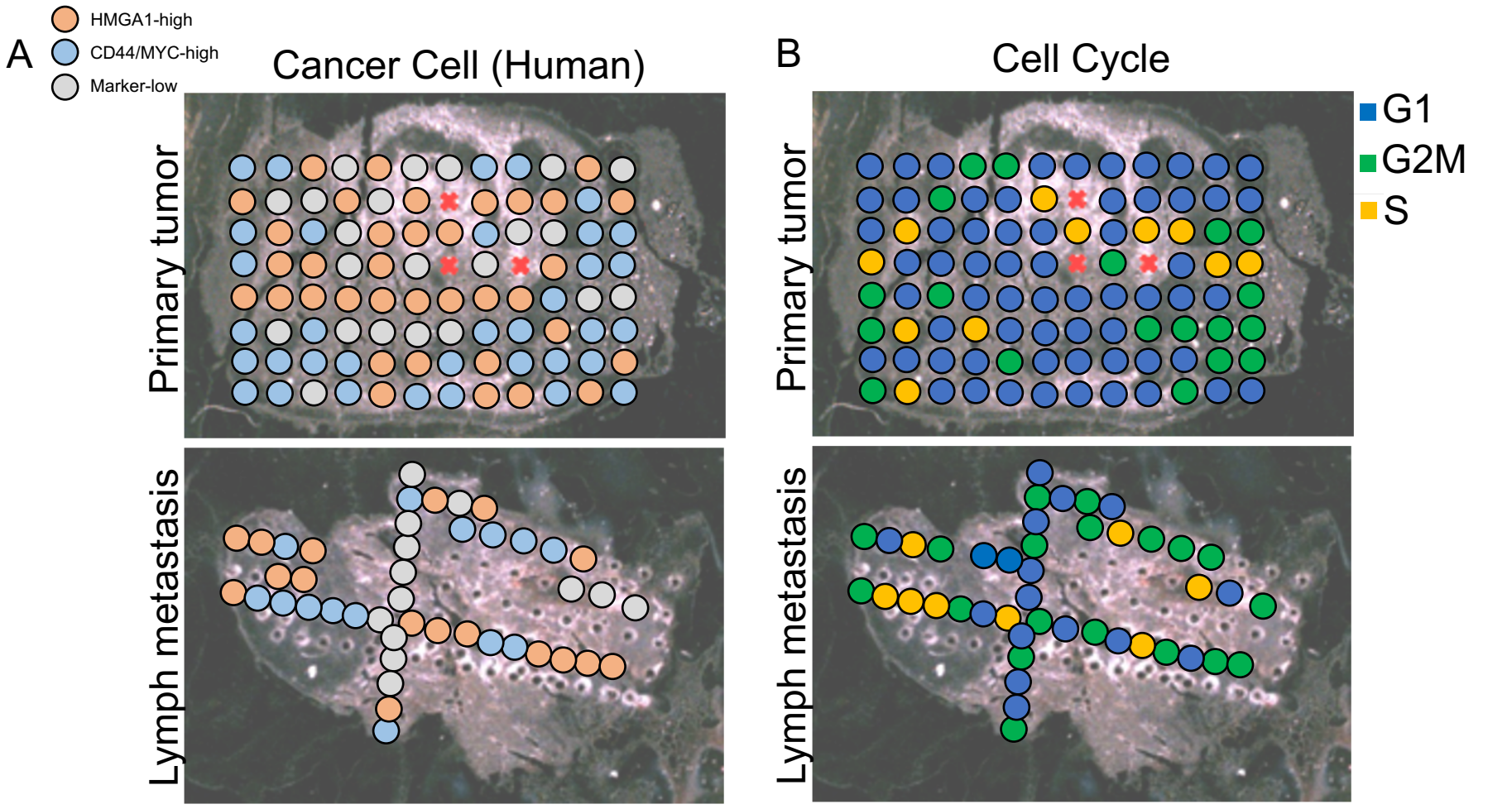
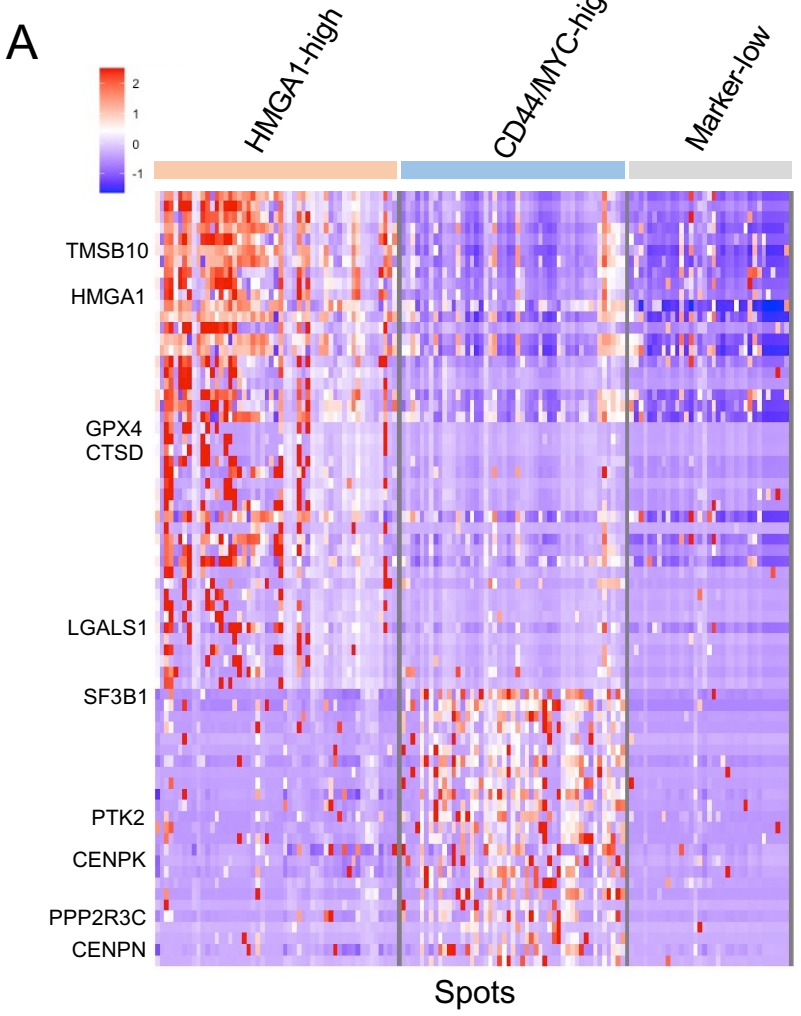
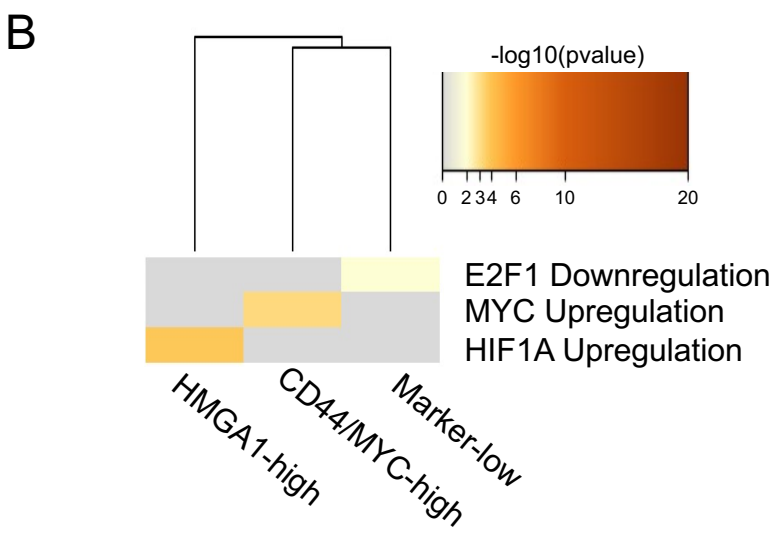
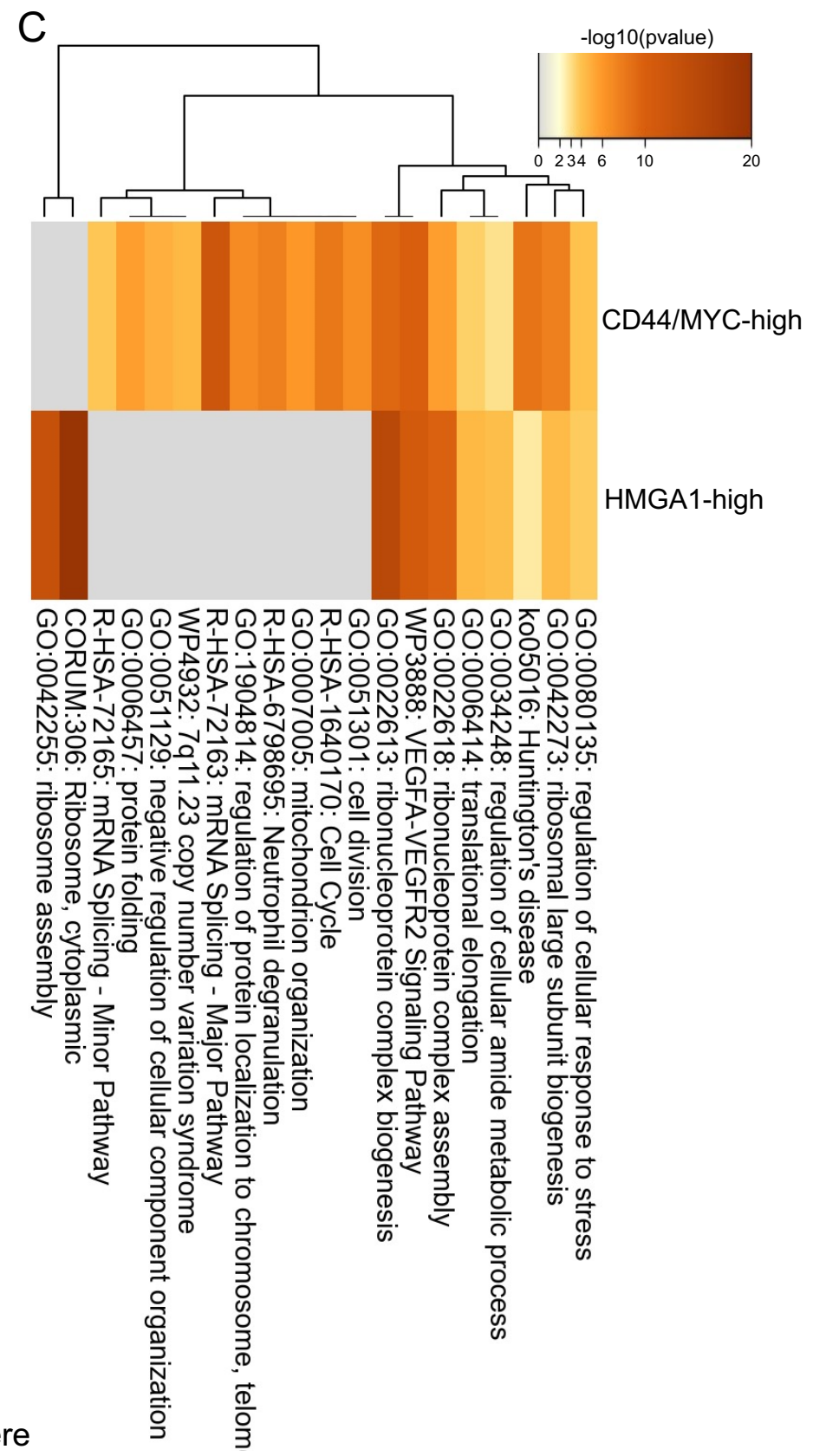

A
HMGA1-high
CD44/MYC-high
Marker-low
Cancer Cell (Human)
Primary tumor
Lymph metastasis
B
Cell Cycle
G1
G2M
S
C
Stromal Cell (Mouse)
Ttr-high
Tcell-like
Inhba-high
Becll-like
D
22.9%
17.1%
67.9%
60.0%
Maker-low
HMGA1
Cell Cycle
CD44/MYC
22.5%
28.3%
30.6%
46.9%
3.8%
E
G1
S/G2M
(%)
100
90
80
70
60
50
40
30
20
10
0
Cell Cycle Ratio
PT
LYMPH
HMGA1
CD44/MYC
Marker-low

Figure 4


A
HMGA1-high
CD44/MYC-high
Marker-low
TMSB10
HMGA1
GPX4
CTSD
LGALS1
SF3B1
PTK2
CENPK
PPP2R3C
CENPN
Spots


B
-log10(pvalue)
0 2 3 4 6 10 20
E2F1 Downregulation
MYC Upregulation
HIF1A Upregulation
HMGA1-high
CD44/MYC-high
Marker-low


C
-log10(pvalue)
0 2 3 4 6 10 20
CD44/MYC-high
HMGA1-high
GO:0080135: regulation of cellular response to stress
GO:0042273: ribosomal large subunit biogenesis
ko05016: Huntington's disease
GO:0034248: regulation of cellular amide metabolic process
GO:0006414: translational elongation
GO:0022618: ribonucleoprotein complex assembly
WP3888: VEGFA-VEGFR2 Signaling Pathway
GO:0022613: ribonucleoprotein complex biogenesis
GO:0051301: cell division
R-HSA-1640170: Cell Cycle
GO:0007005: mitochondrion organization
R-HSA-6798695: Neutrophil degranulation
GO:1904814: regulation of protein localization to chromosome,
R-HSA-72163: mRNA Splicing - Major Pathway
WP4932: 7q11.23 copy number variation syndrome
GO:0051129: negative regulation of cellular component organiz
GO:0006457: protein folding
R-HSA-72165: mRNA Splicing - Minor Pathway
CORUM:306: Ribosome, cytoplasmic
GO:0042255: ribosome assembly

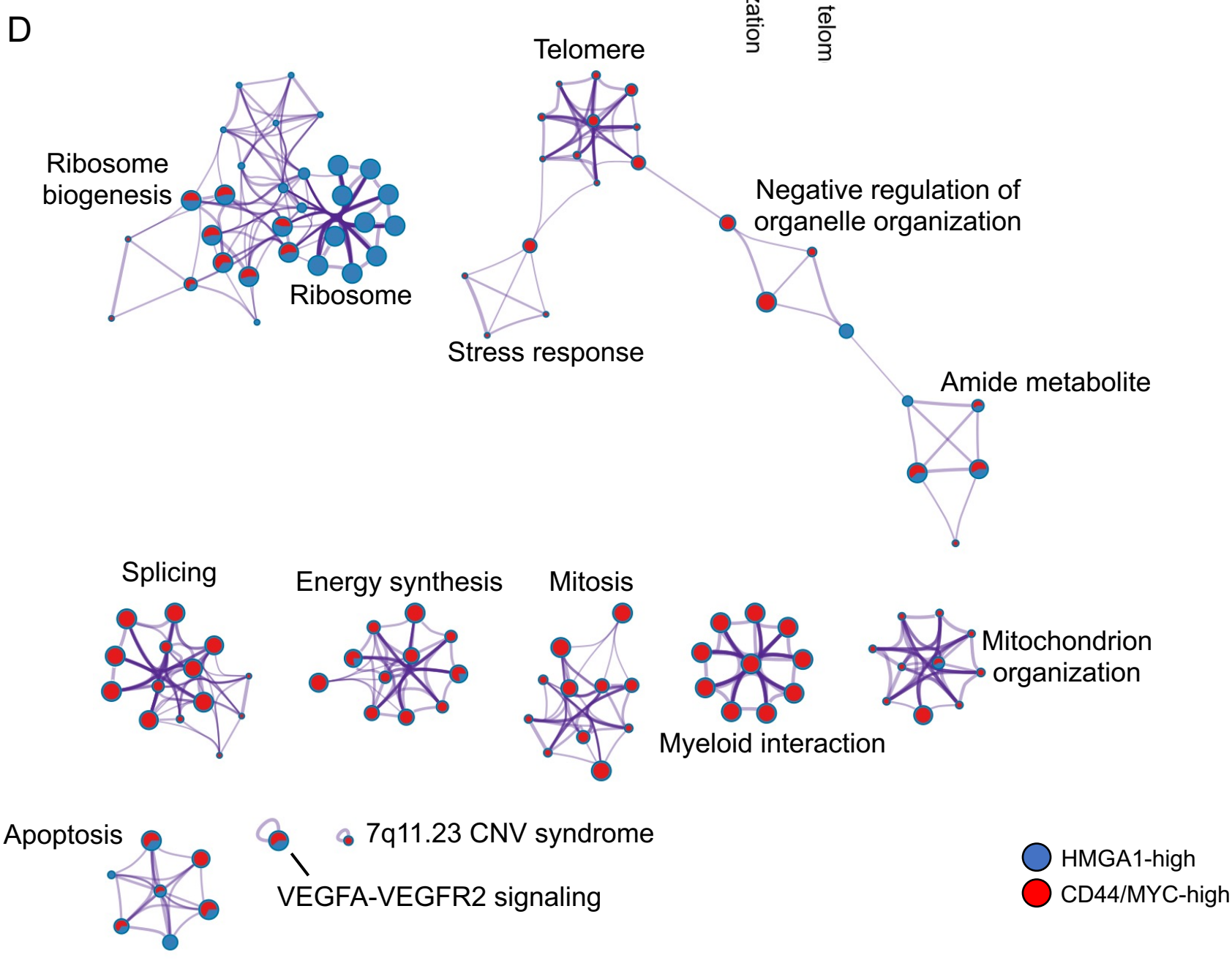

D
Telomere
Ribosome biogenesis
Ribosome
Negative regulation of organelle organization
Stress response
Amide metabolite
Splicing
Energy synthesis
Mitosis
Myeloid interaction
Mitochondrion organization
Apoptosis
7q11.23 CNV syndrome
VEGFA-VEGFR2 signaling
HMGA1-high
CD44/MYC-high

Figure 5

A

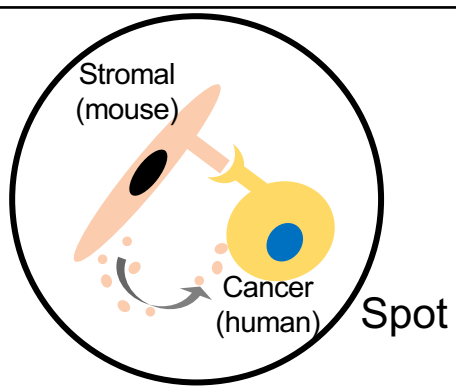
Cell-Cell Interaction Database

3,209 Ligands
4,364 Receptors
433 Extracellular Matrix

Total 115,900 interactions

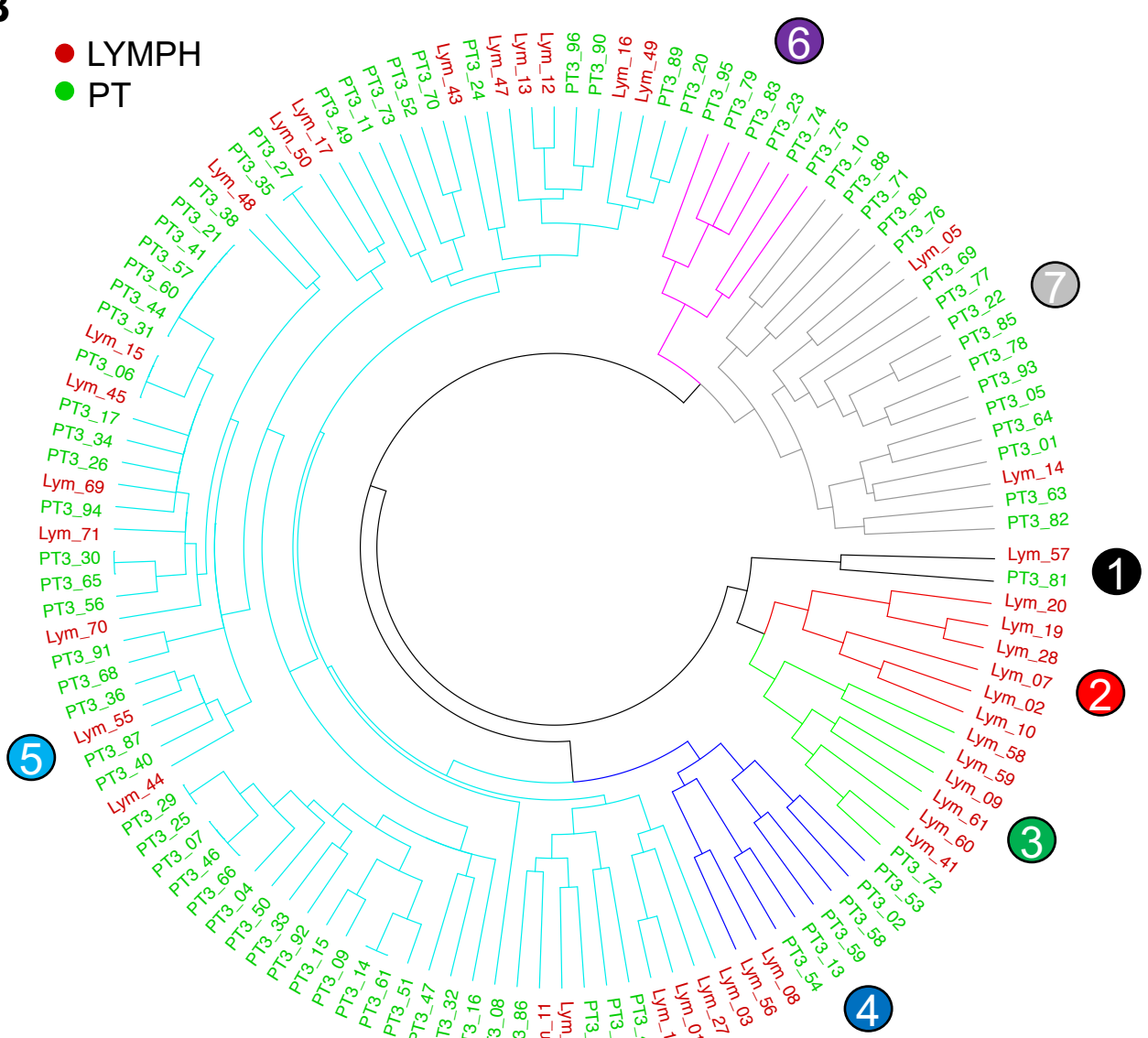
Choose expressed genes
Expression > 2

Estimation of
ligand-receptor interaction
in the same spot



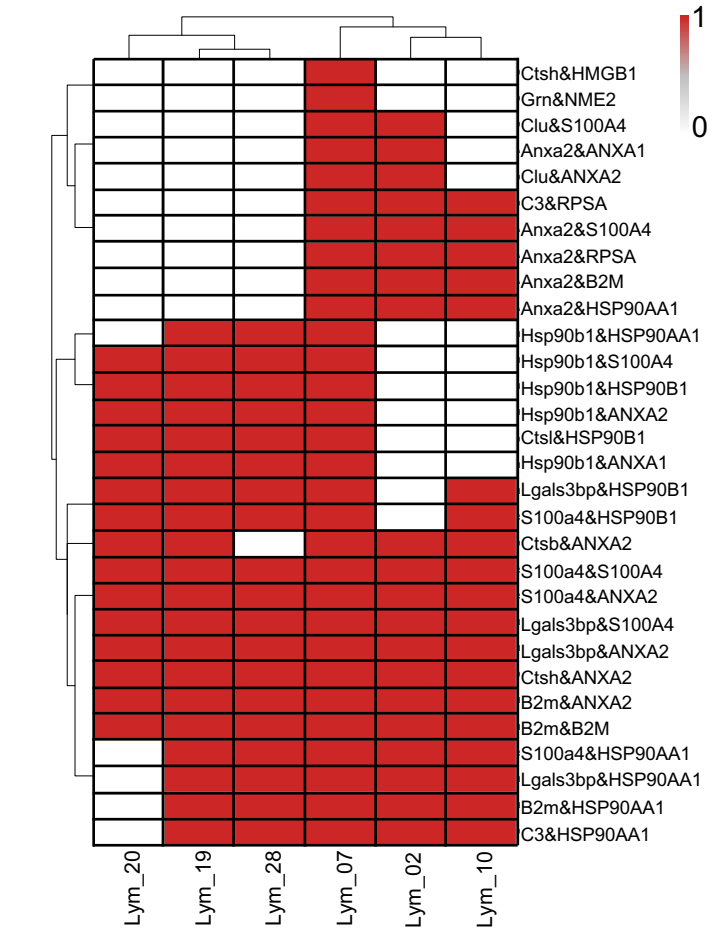
B

Stromal to Cancer Interaction

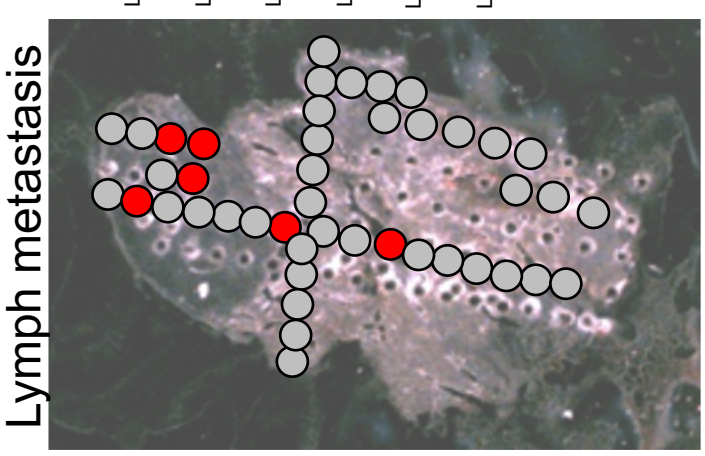


C

2 Lymph specific interaction

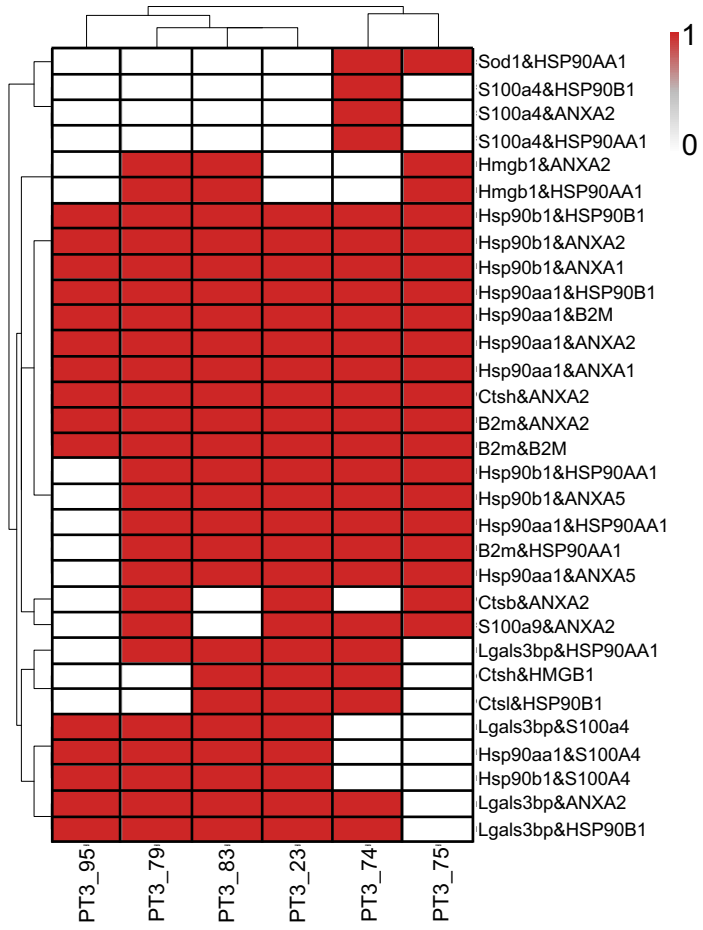


D



E

6 Primary tumor specific interaction



F

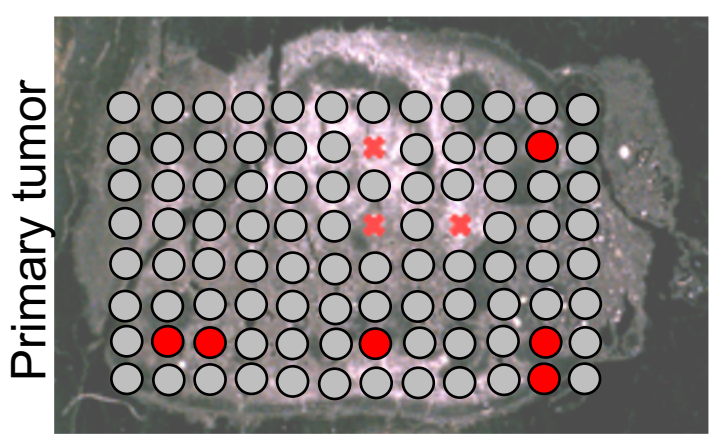
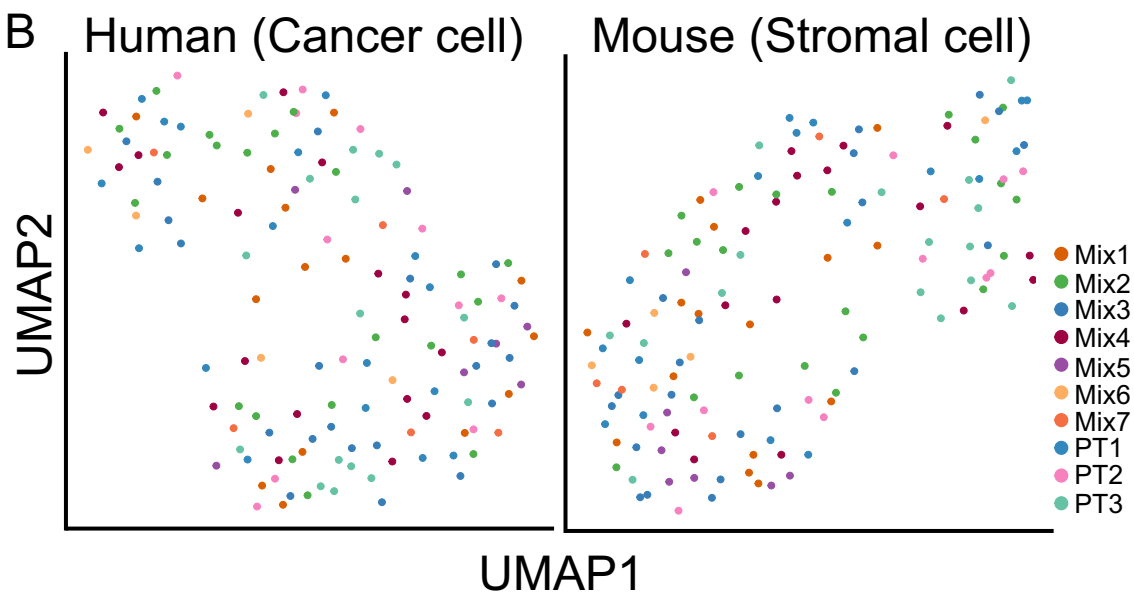
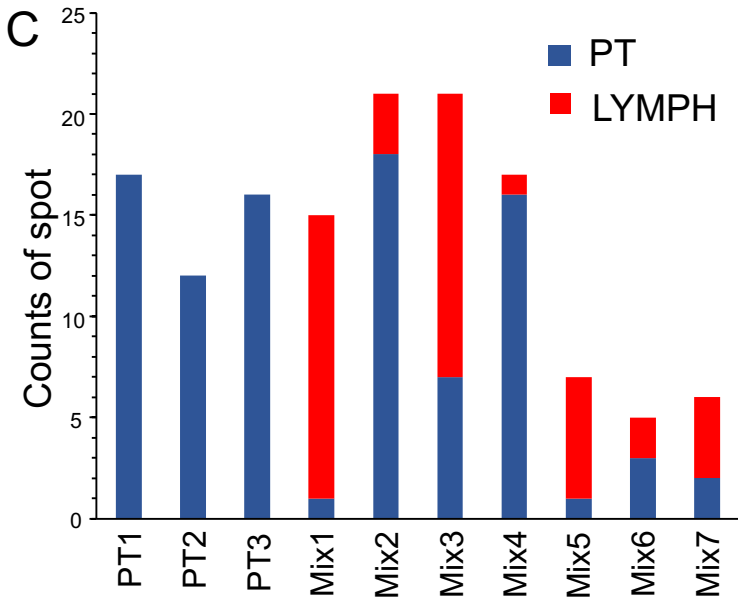


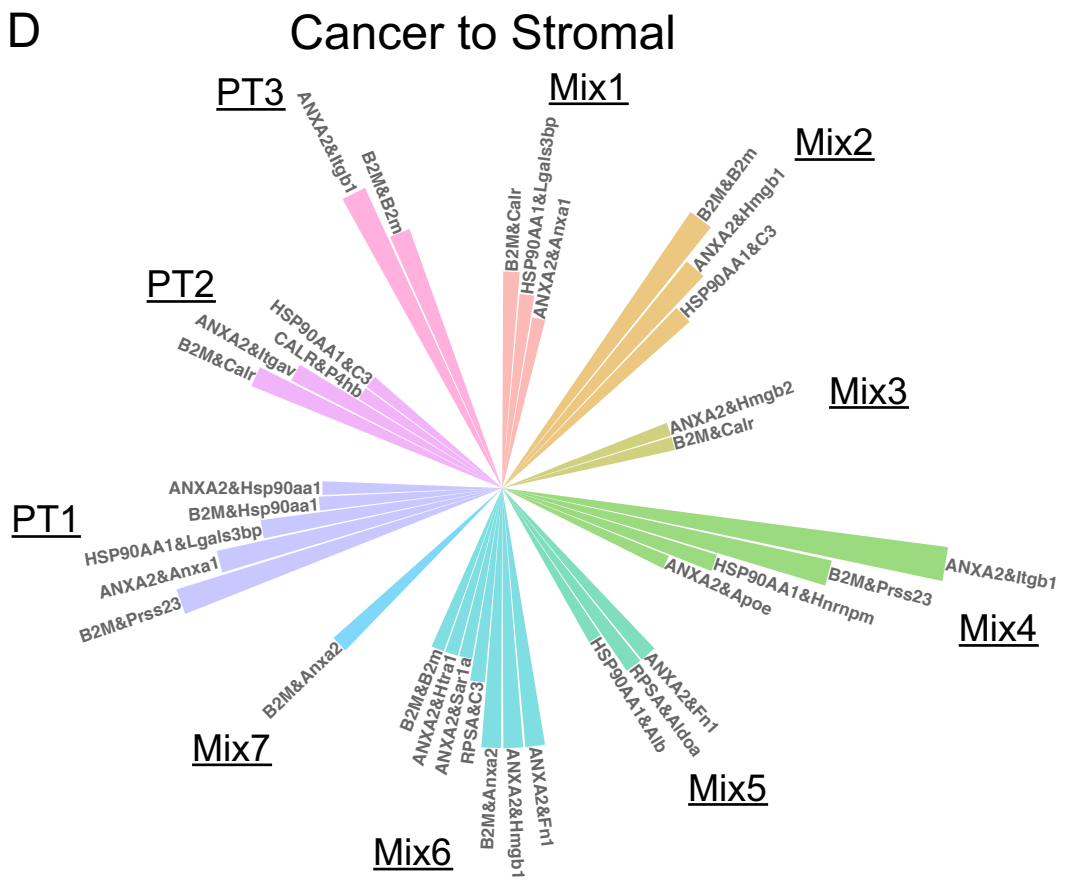
Figure 6

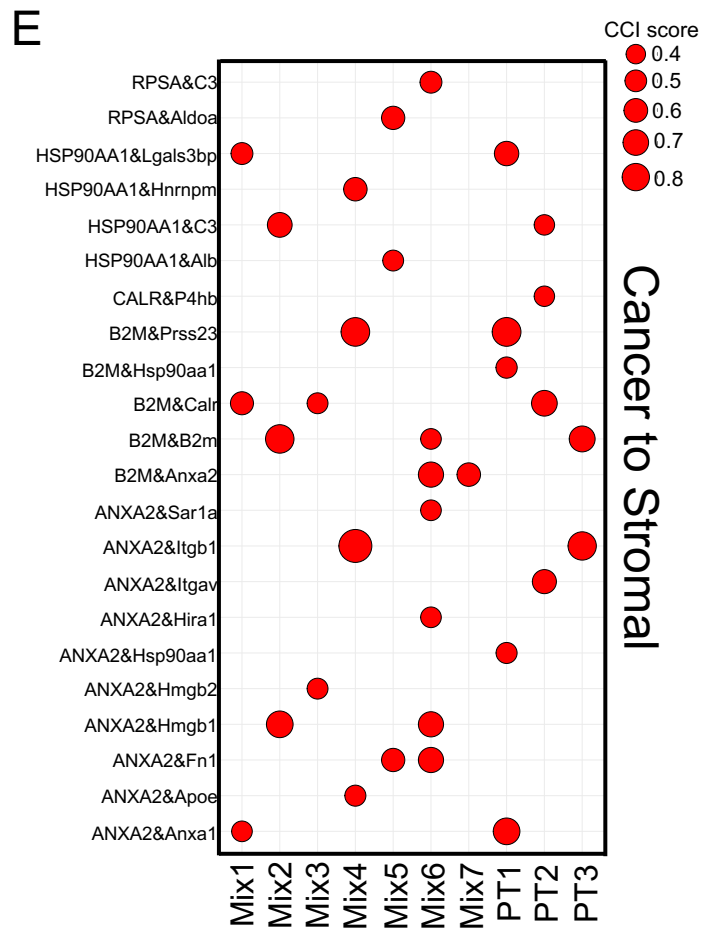
A

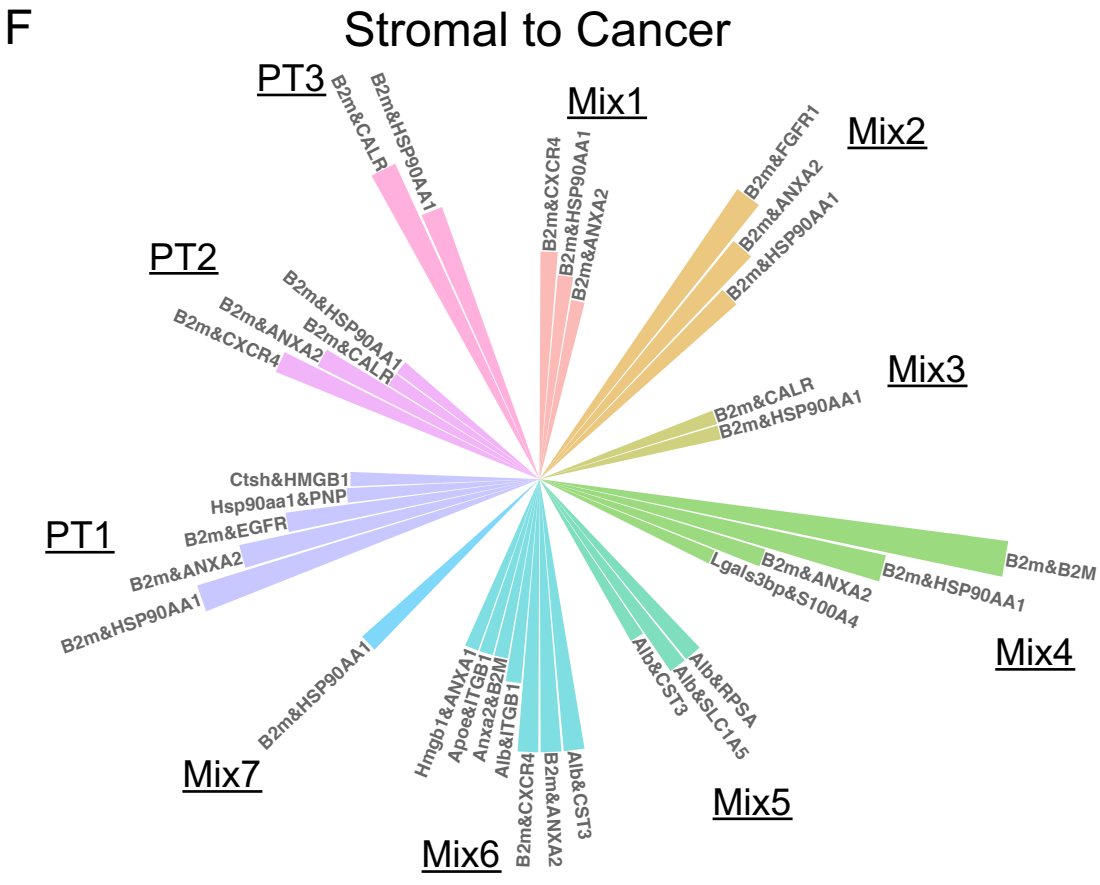
	PT-1	PT-2	PT-3	Mix-1	Mix-2	Mix-3	Mix-4	Mix-5	Mix-6	Mix-7
human	HMGA1-high	CD44/MYC-high	CD44/MYC-high	HMGA1-high	HMGA1-high	CD44/MYC-high	Marker-low	Marker-low	Marker-low	Marker-low
mouse	Ttr-high	Ttr-high	Inhba-high	Tcell-like	Inhba-high	Tcell-like	Ttr-high	Tcell-like	Inhba-high	Bcell-like











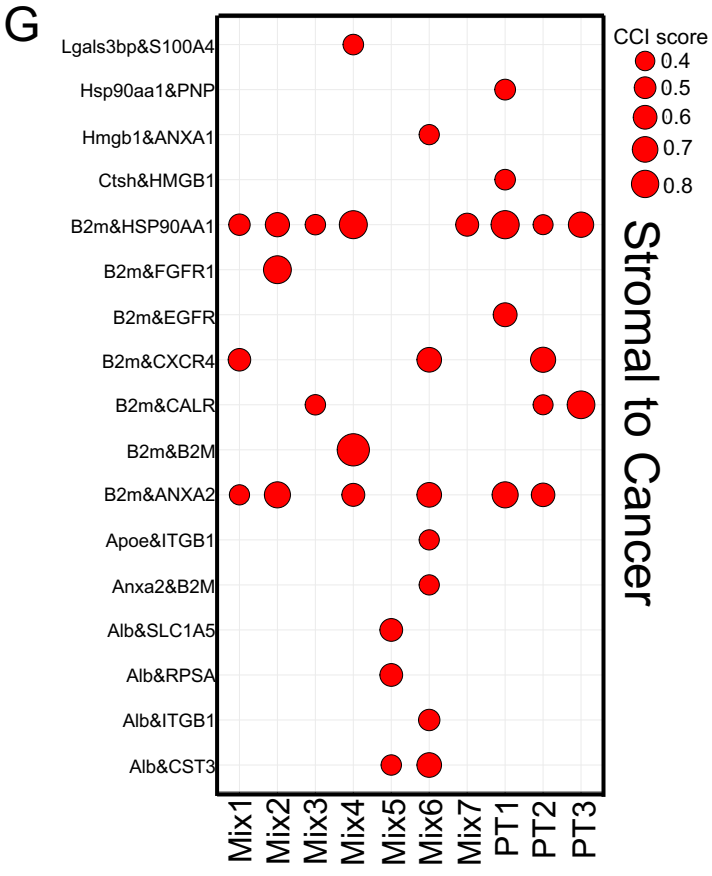
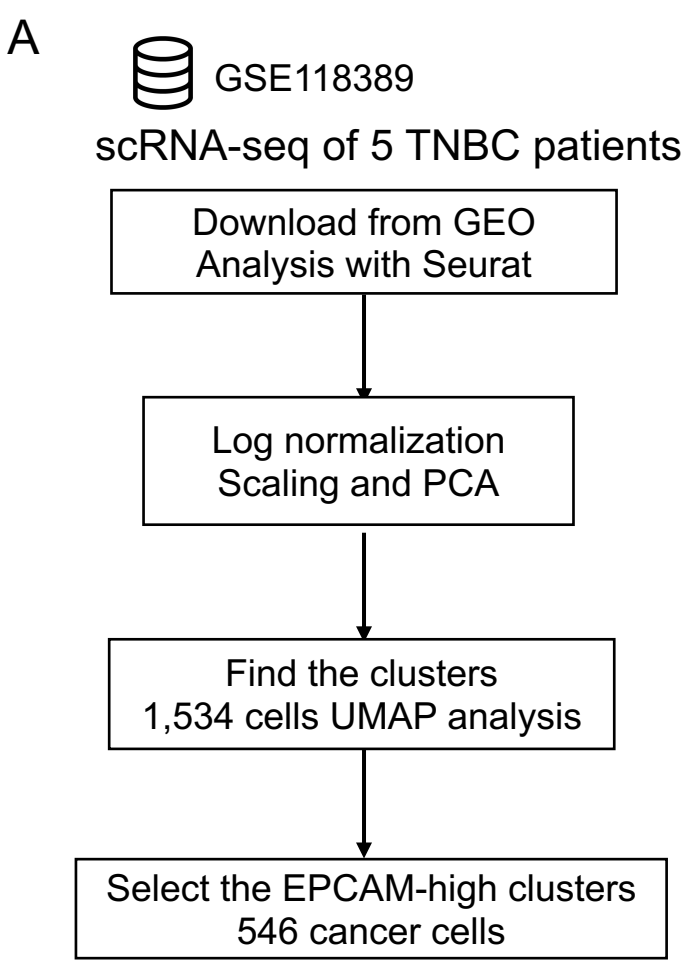
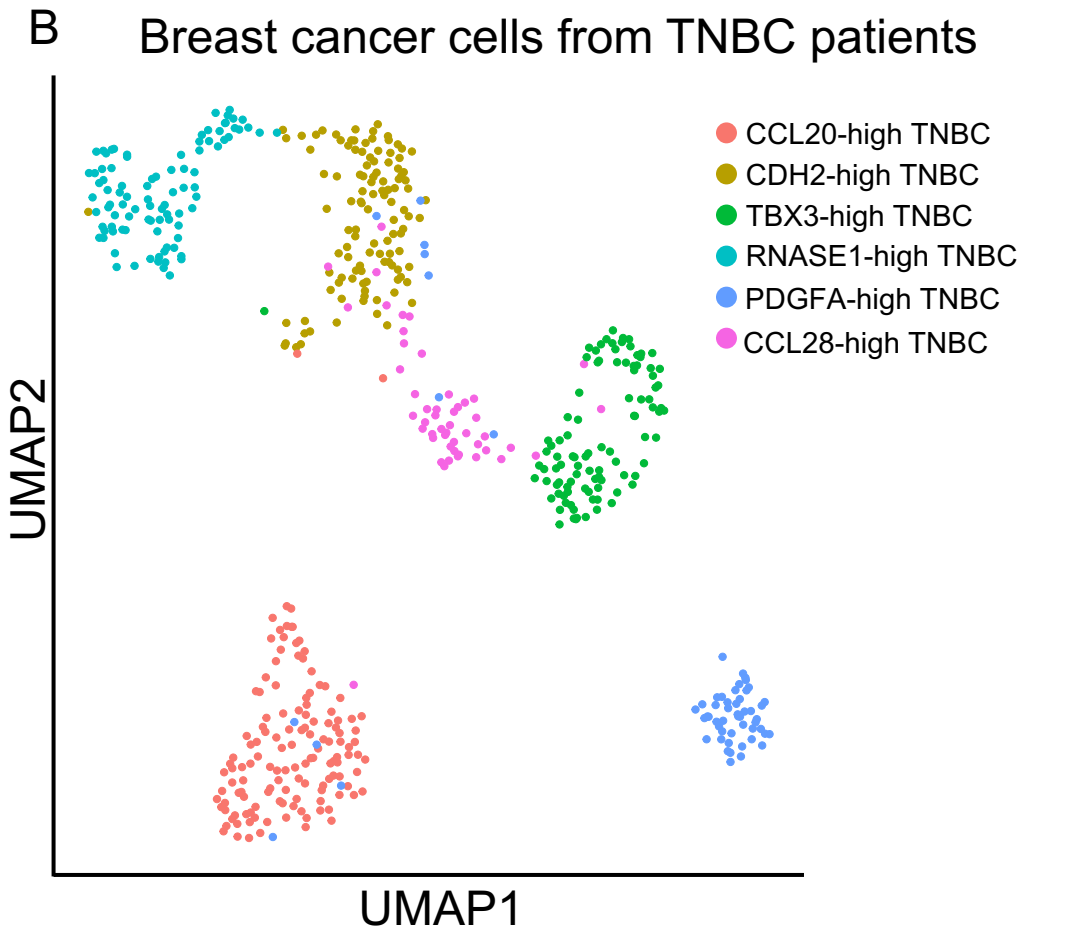
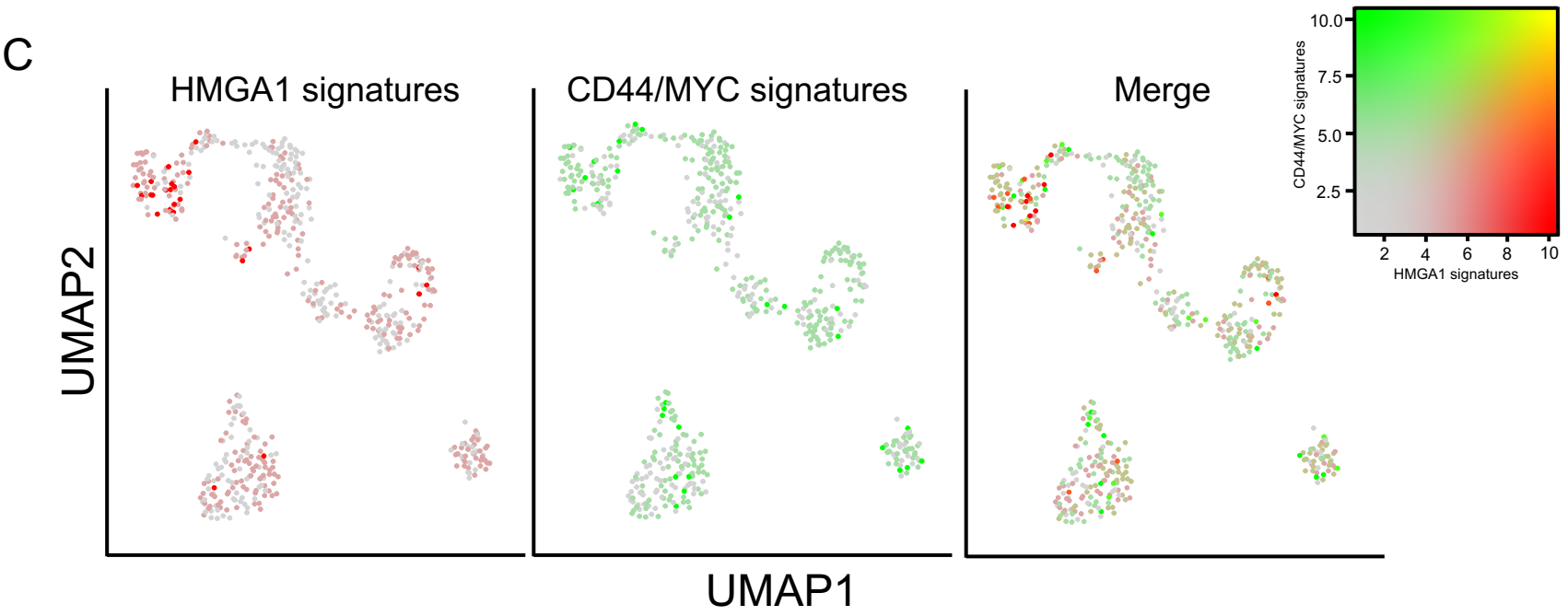
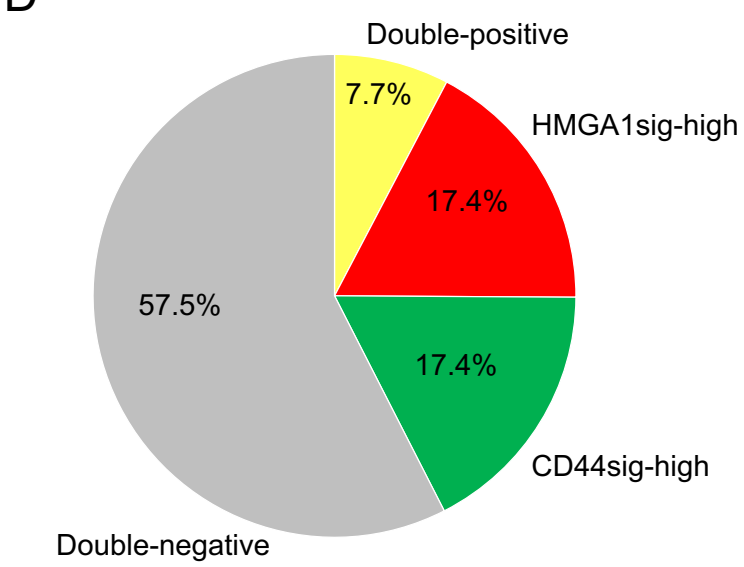


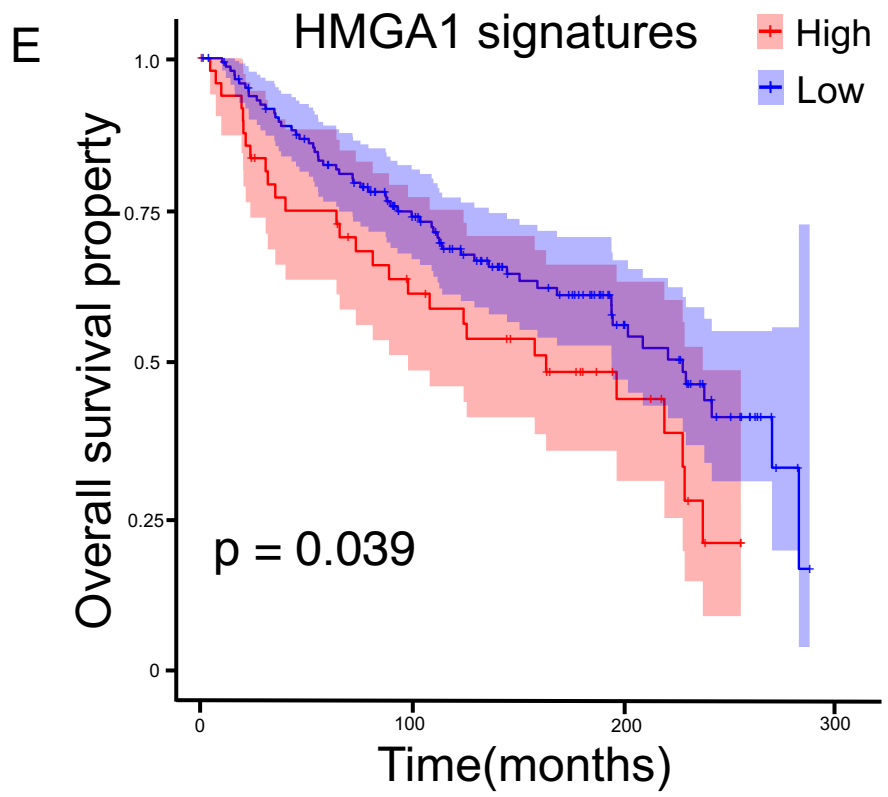
Figure 7


A
GSE118389
scRNA-seq of 5 TNBC patients
Download from GEO
Analysis with Seurat
Log normalization
Scaling and PCA
Find the clusters
1,534 cells UMAP analysis
Select the EPCAM-high clusters
546 cancer cells

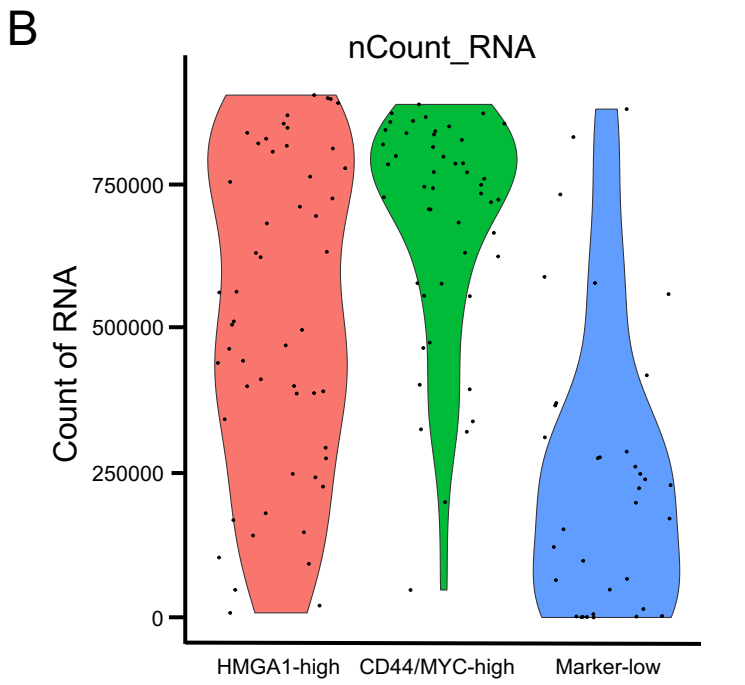

B
Breast cancer cells from TNBC patients
CCL20-high TNBC
CDH2-high TNBC
TBX3-high TNBC
RNASE1-high TNBC
PDGFA-high TNBC
CCL28-high TNBC
UMAP2
UMAP1

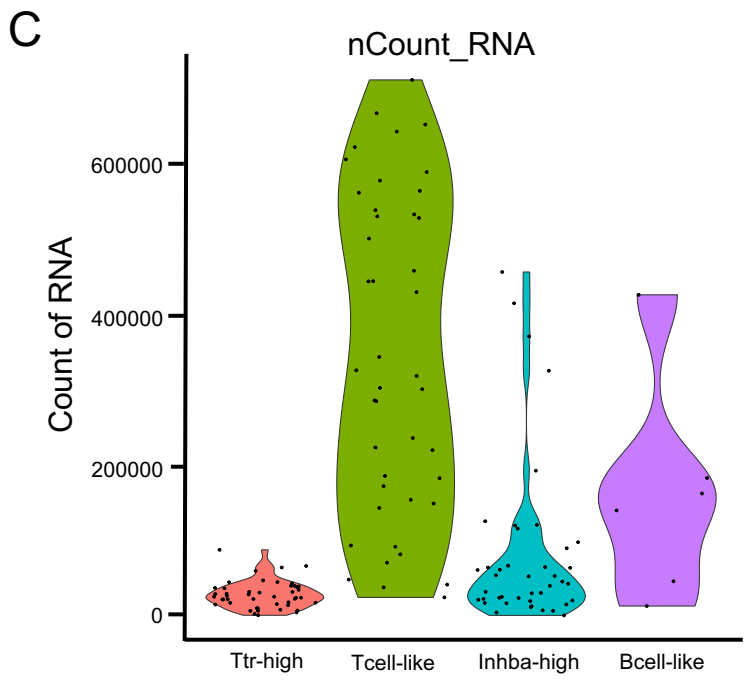

C
HMGA1 signatures
CD44/MYC signatures
Merge
UMAP2
UMAP1
CD44/MYC signatures
HMGA1 signatures
10.0
7.5
5.0
2.5
2
4
6
8
10

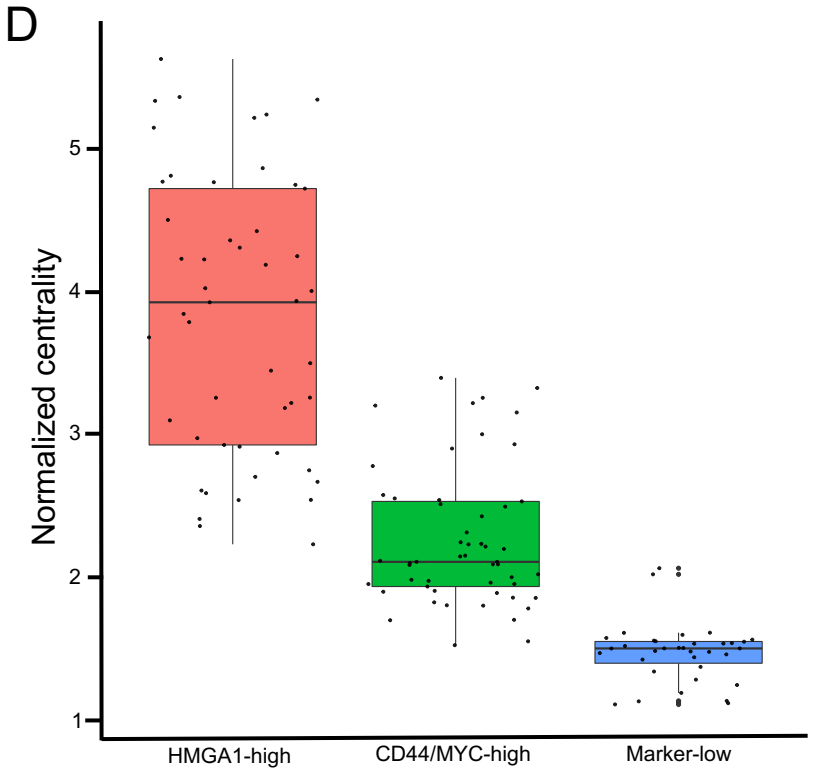

D
Double-positive
7.7%
HMGA1sig-high
17.4%
57.5%
17.4%
CD44sig-high
Double-negative

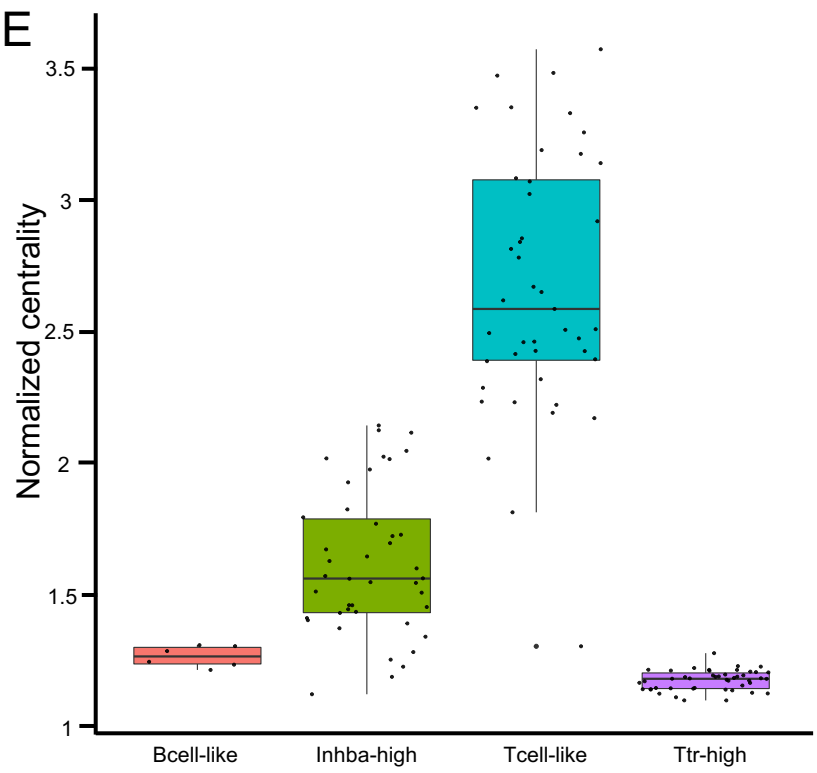

E
HMGA1 signatures
High
Low
Overall survival property
1.0
0.75
0.5
0.25
0
p = 0.039
0
100
200
300
Time(months)

Supplementary Figure S1

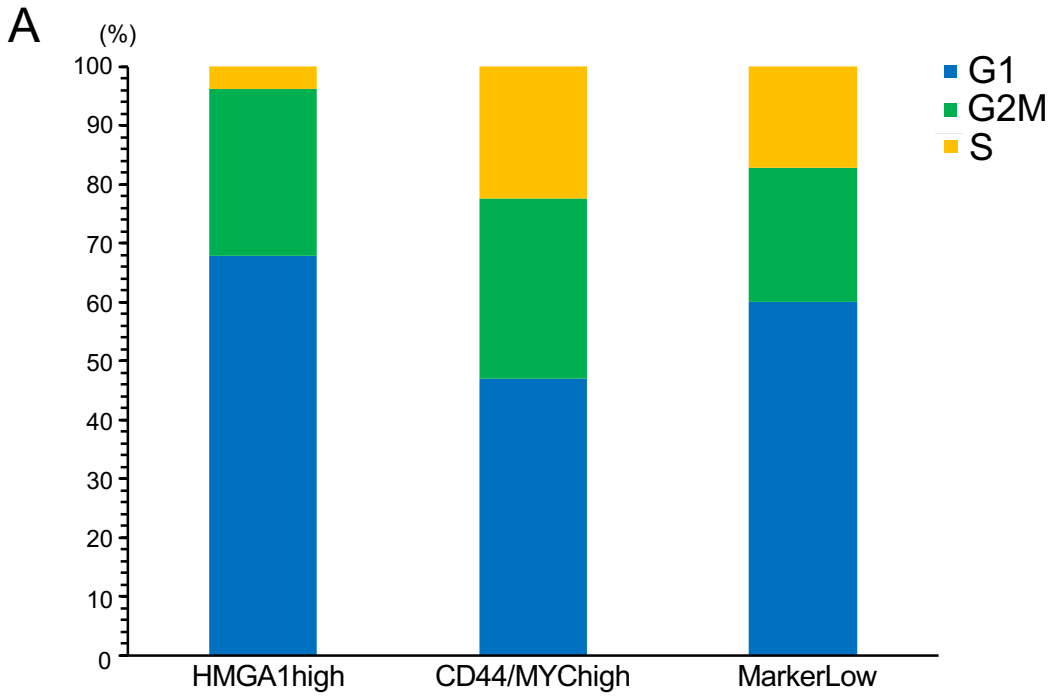


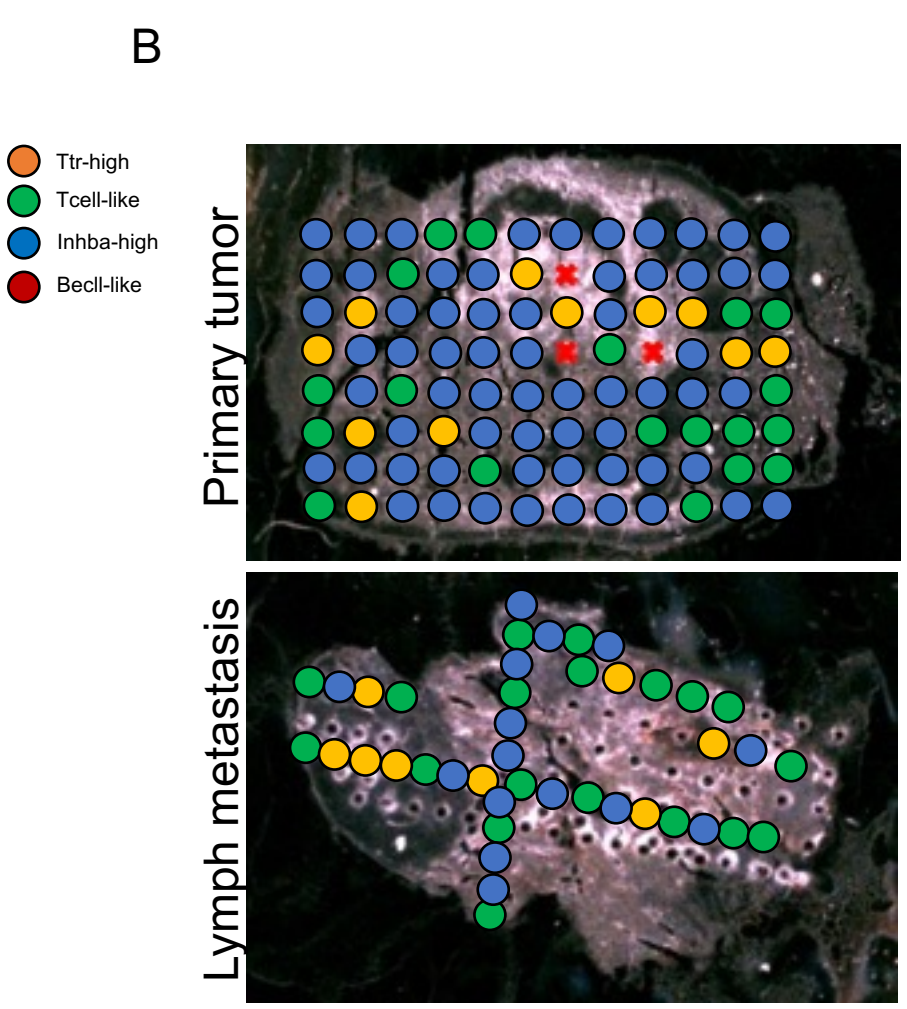


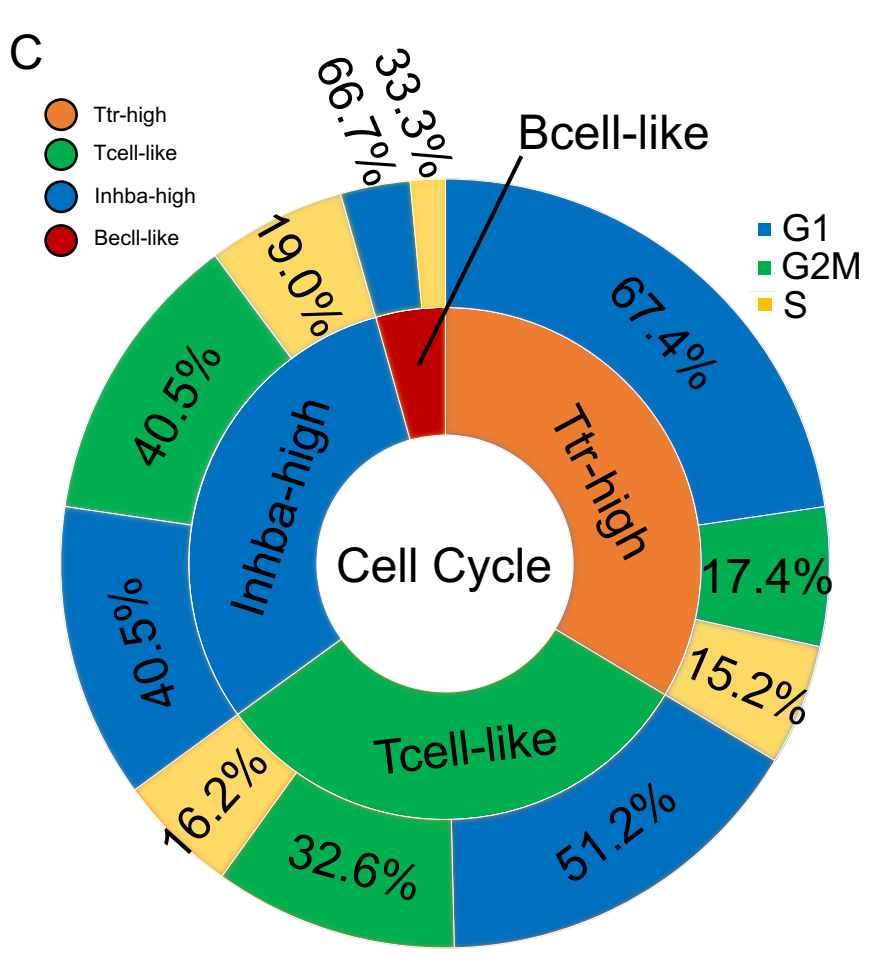


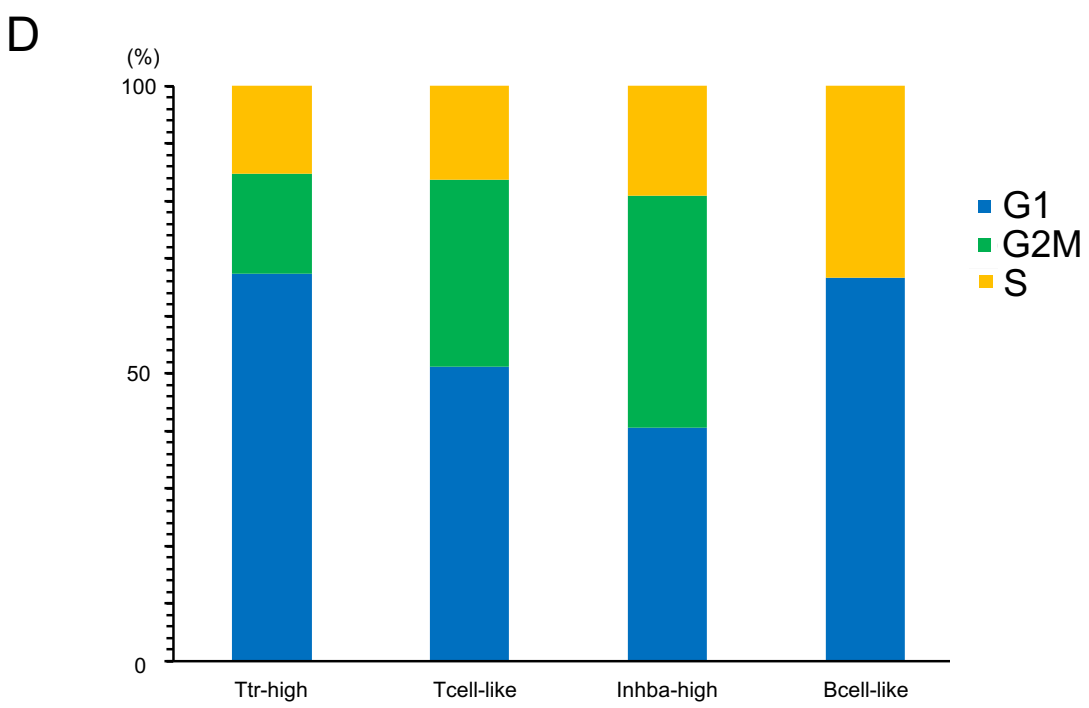


Supplementary Figure S2

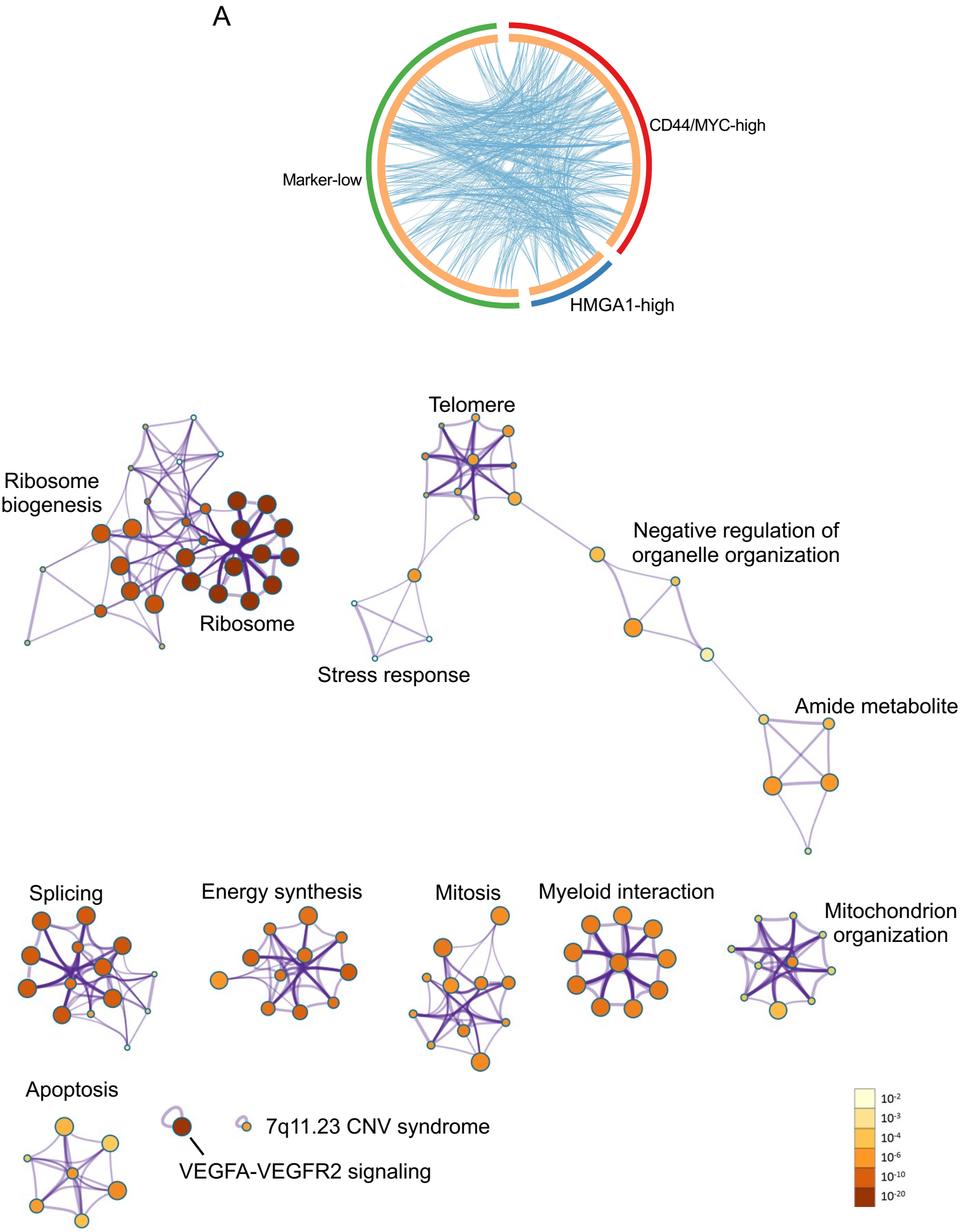






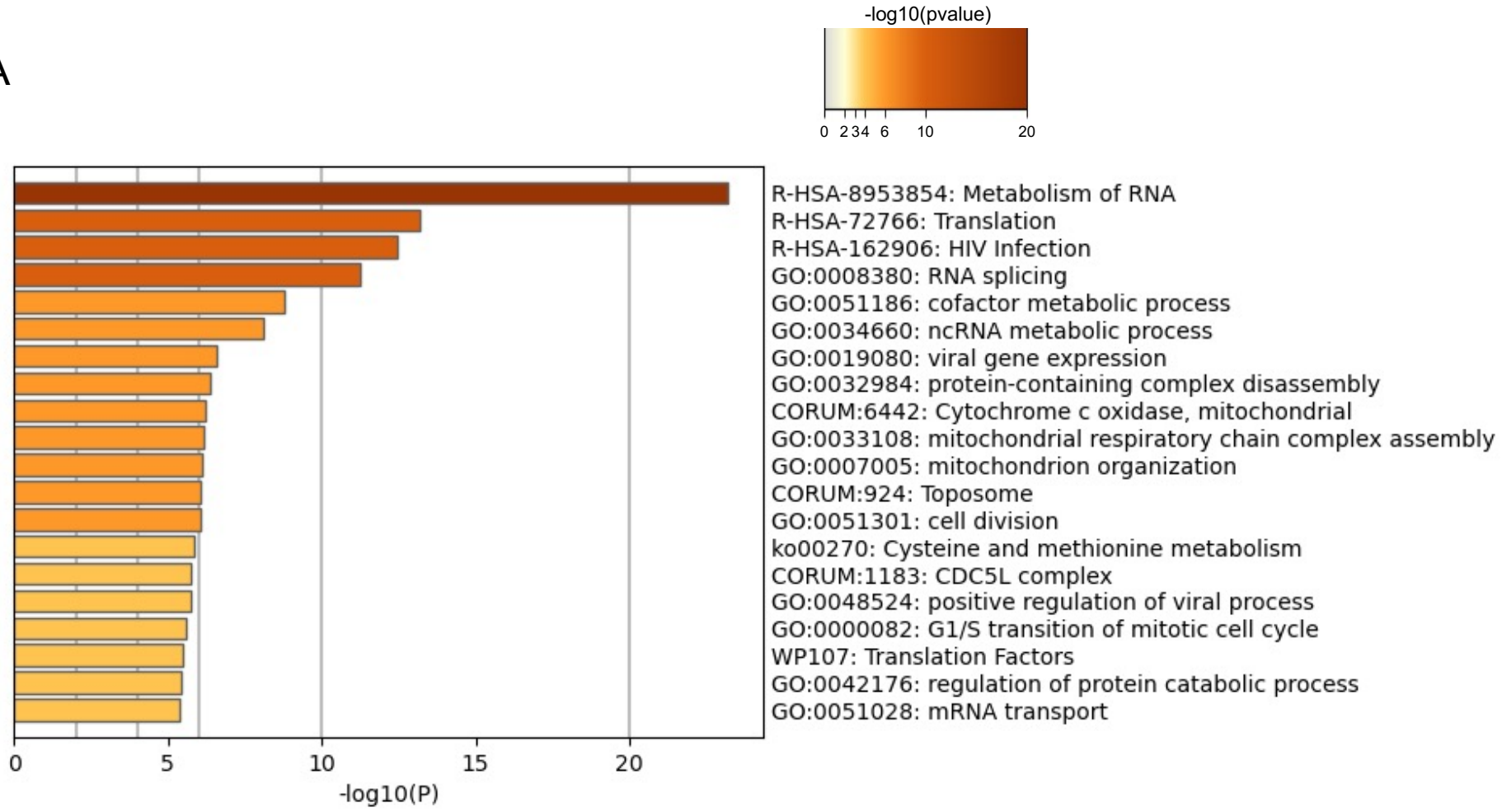


Supplementary Figure S3

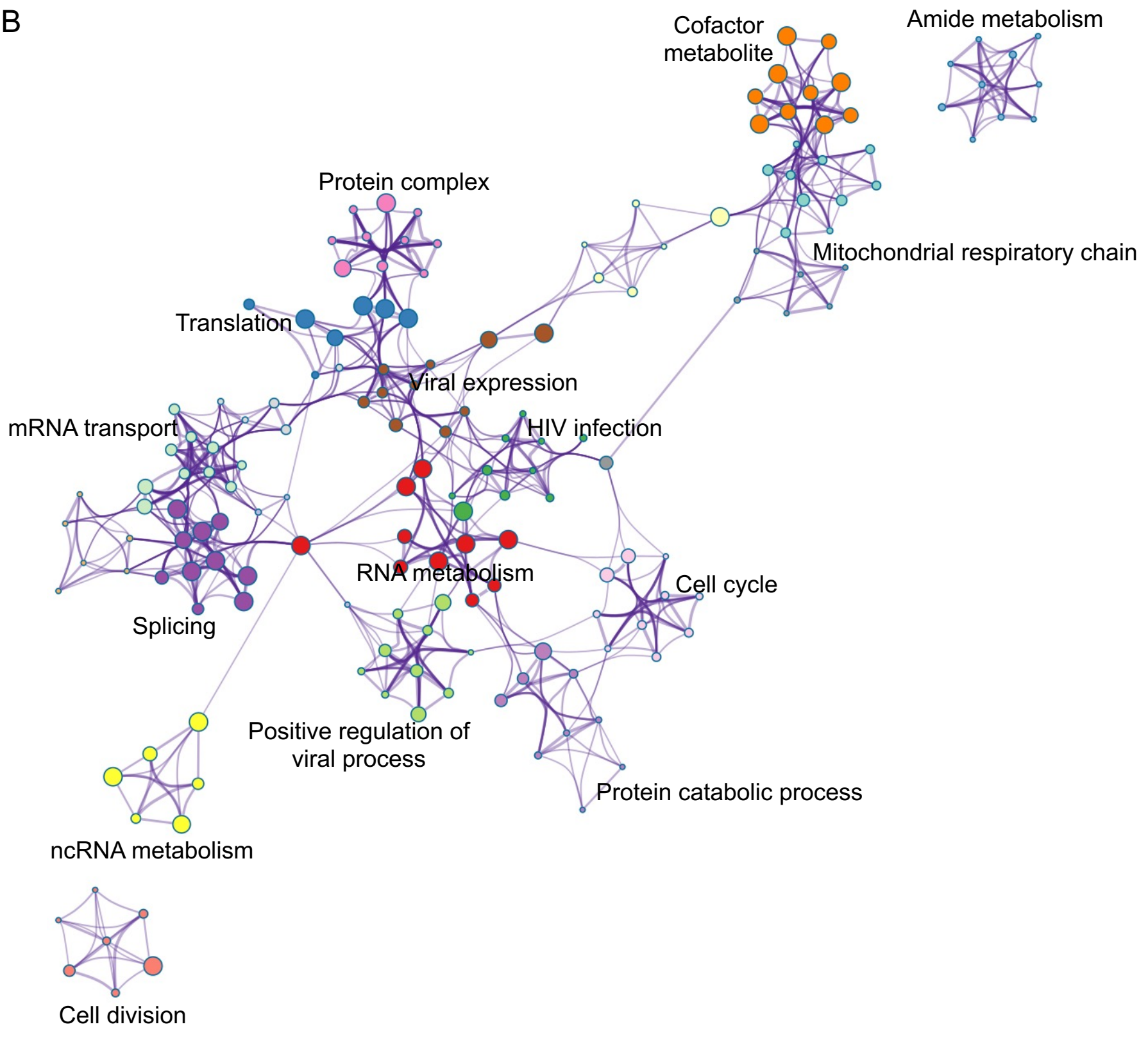


Supplementary Figure S4

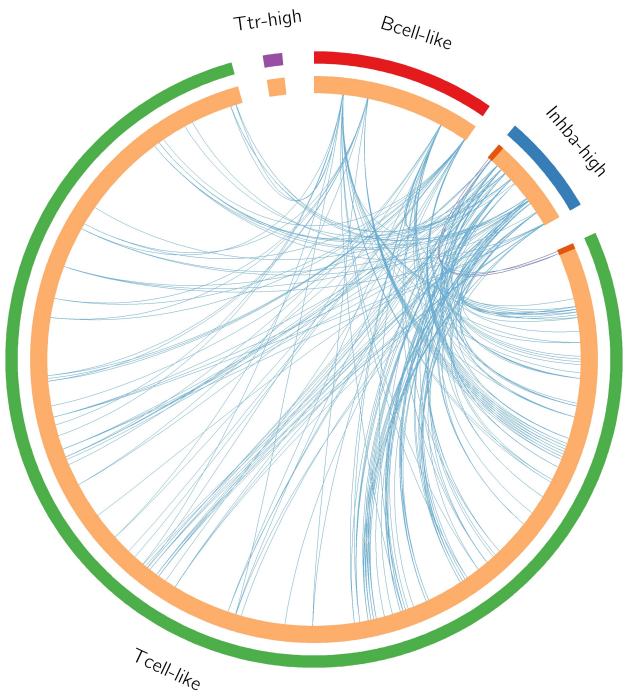
A

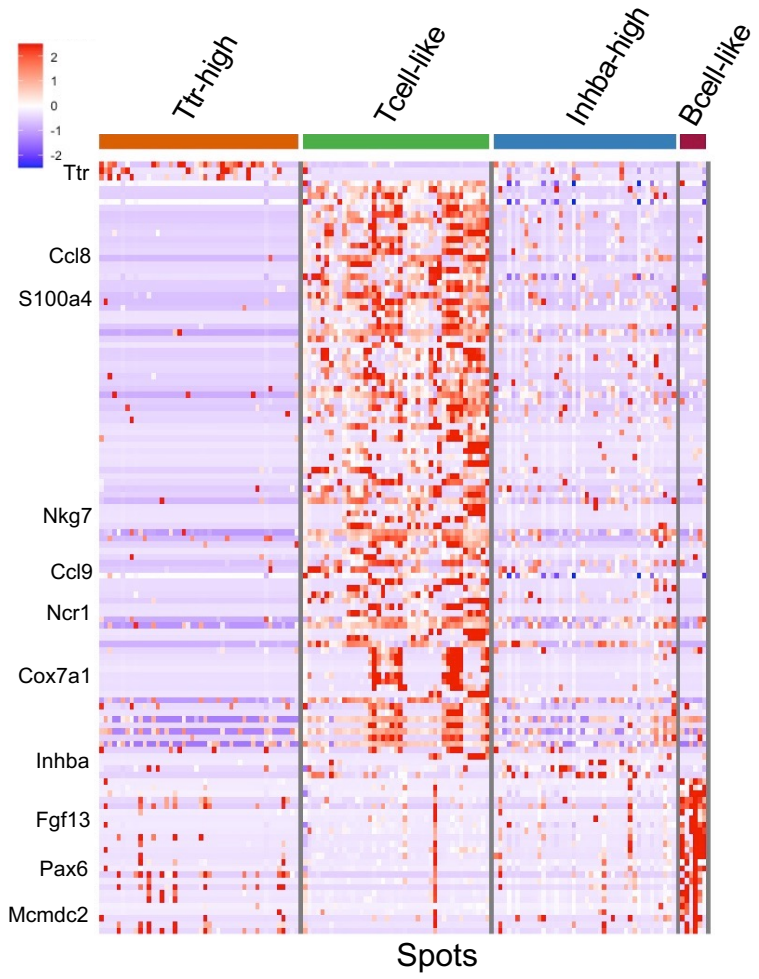


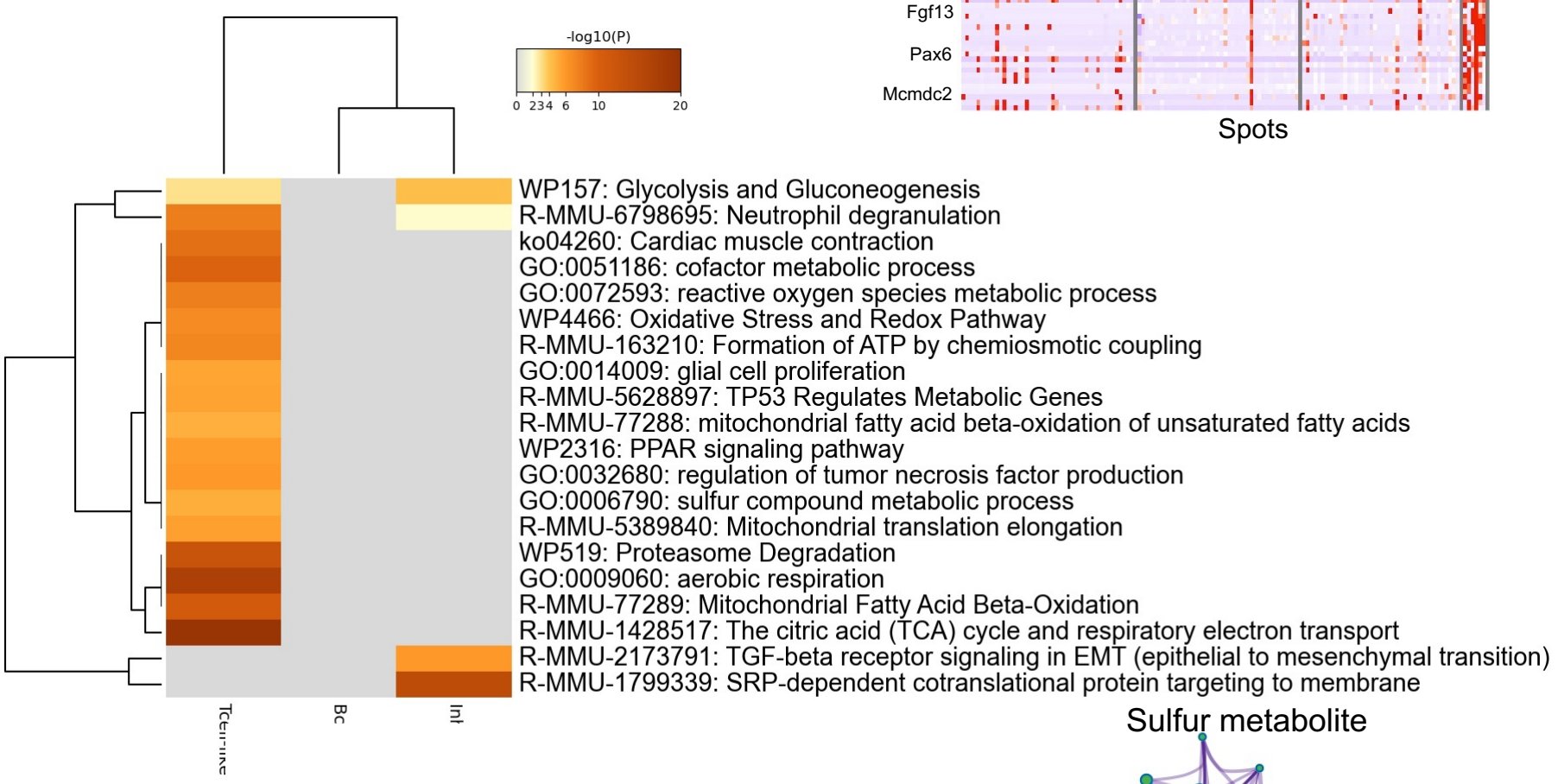
B

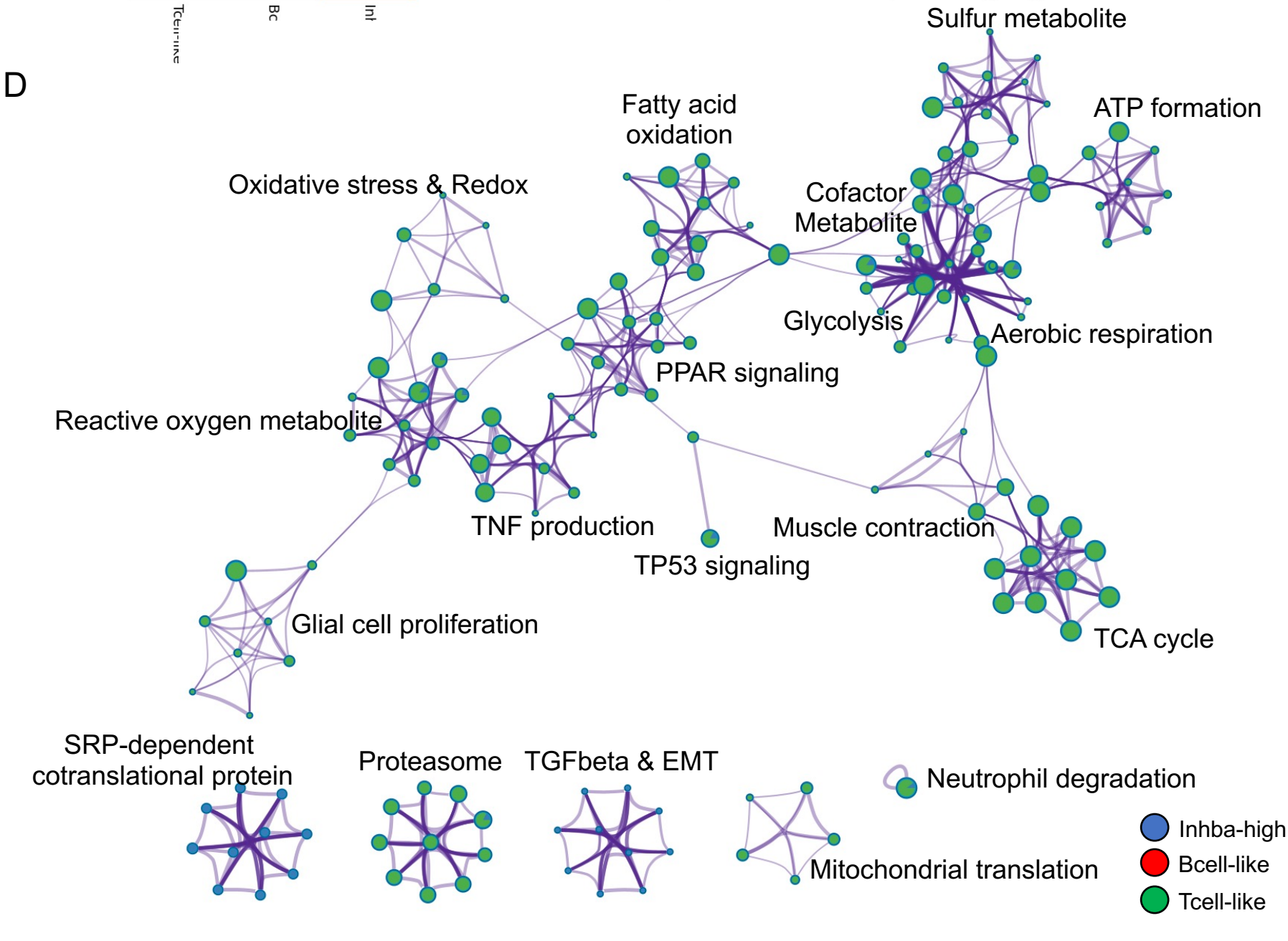


Supplementary Figure S5



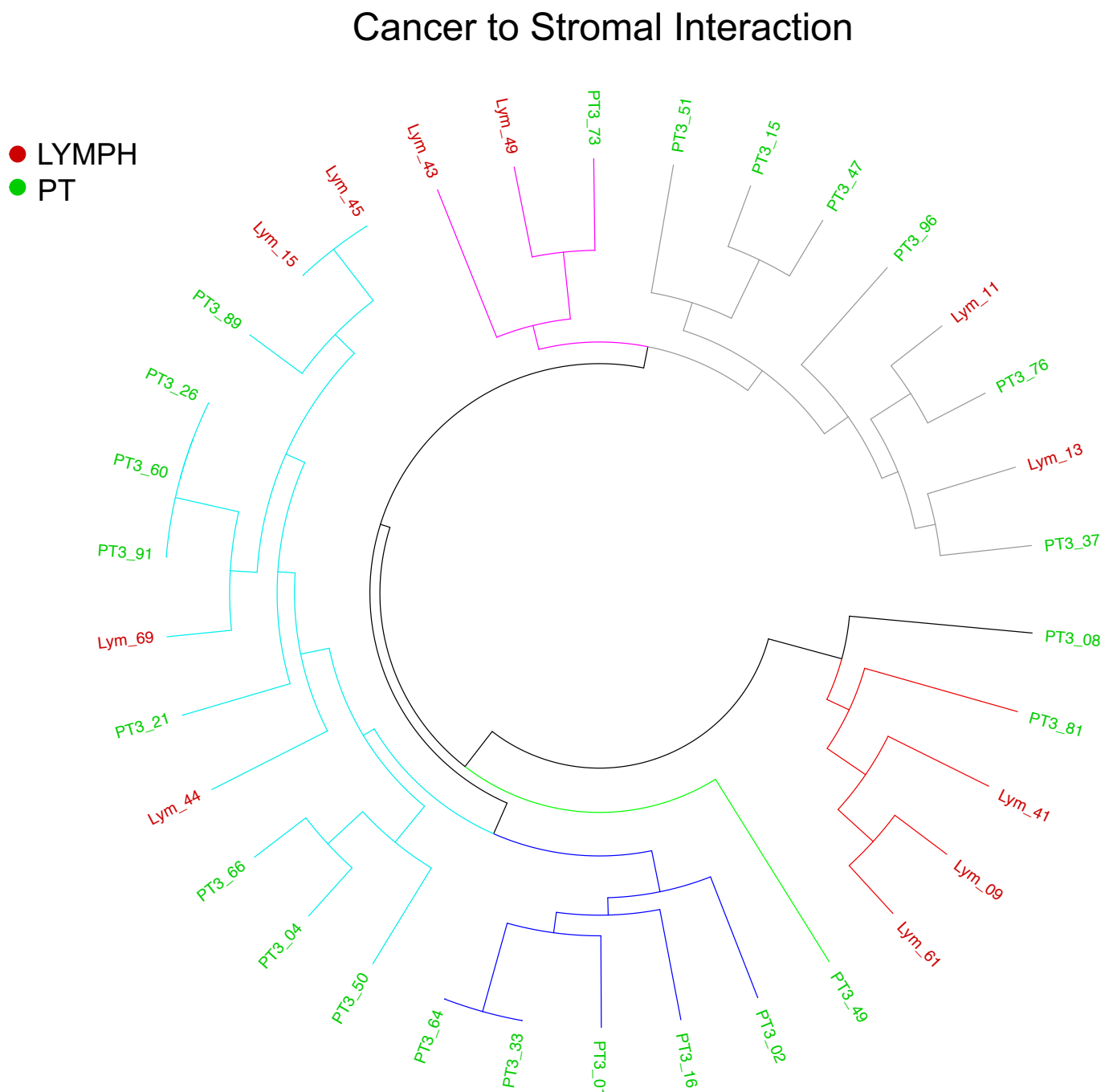




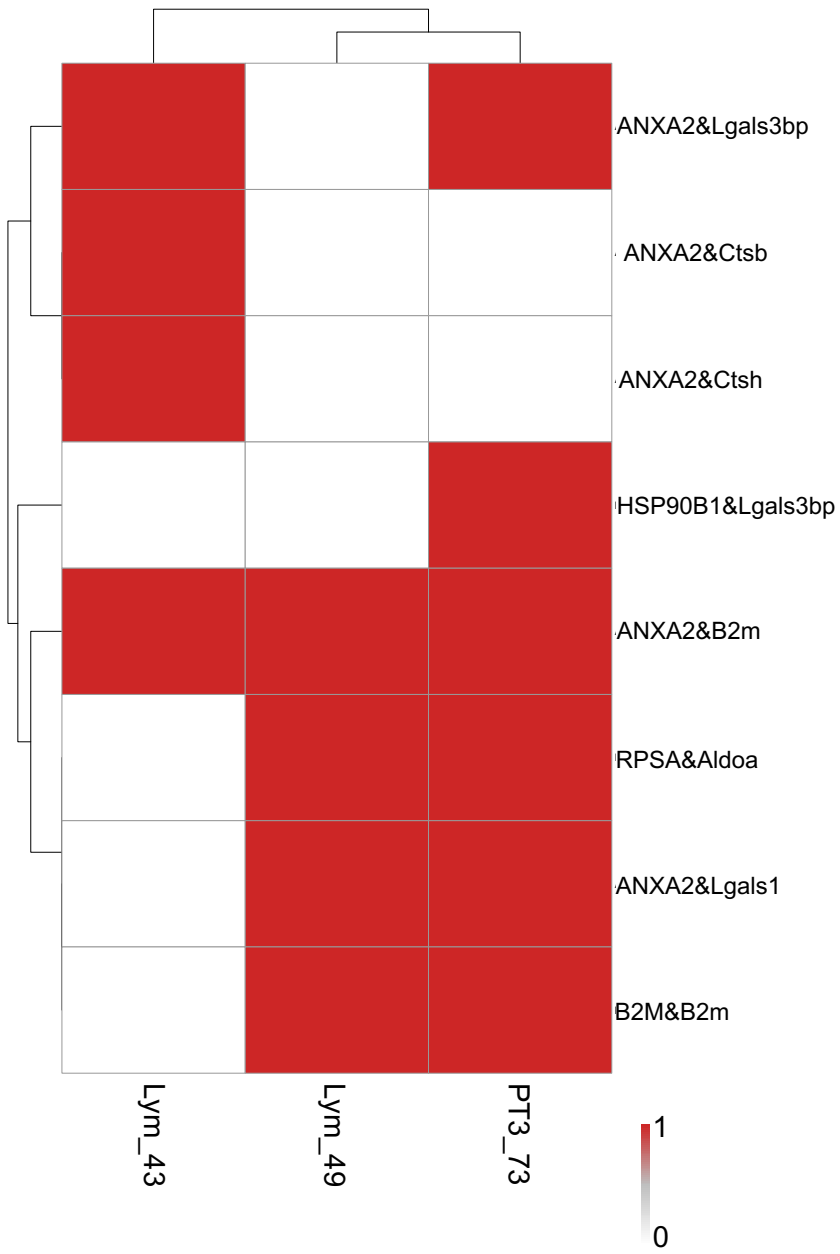


Supplementary Figure S6

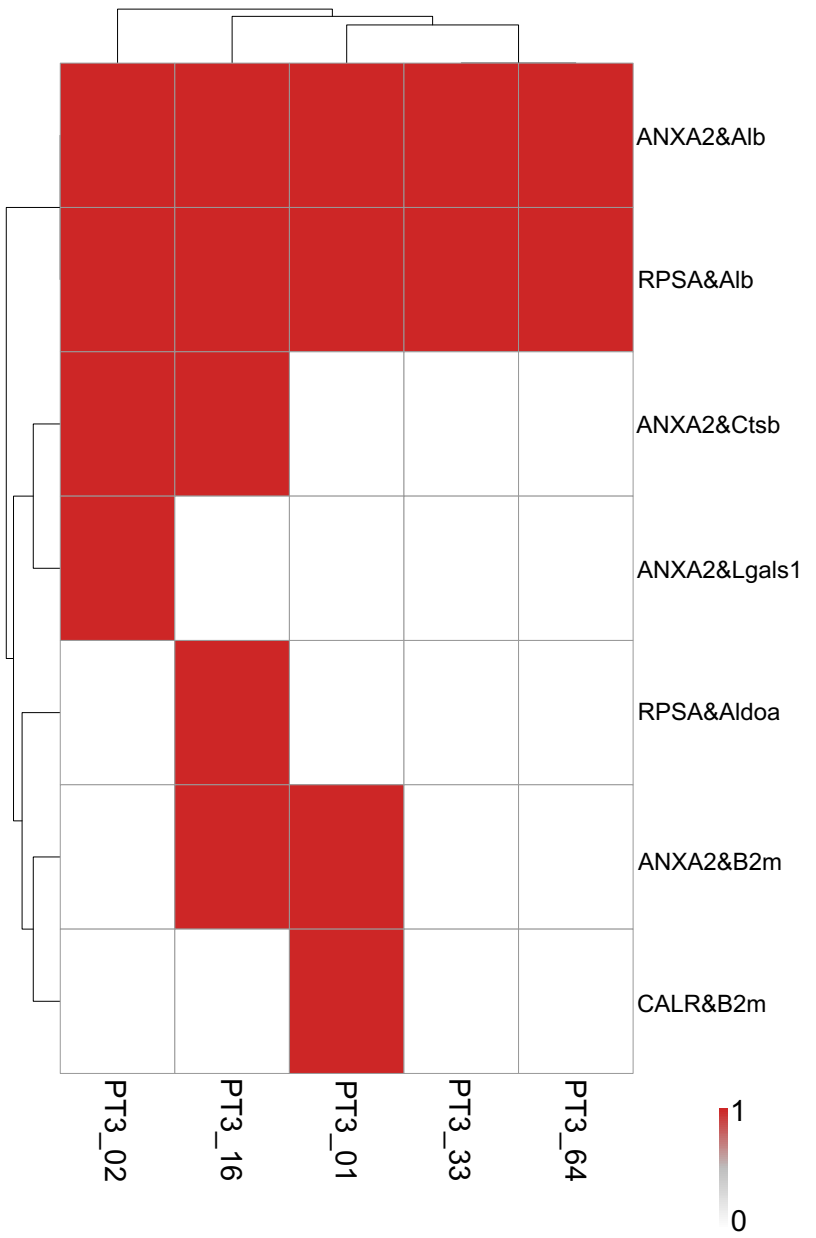
A



B 2 Lymph-PT interaction



C 5 PT specific interaction



Supplementary Figure S7