Cellular heterogeneity map of diverse immune and stromal phenotypes within breast tumor microenvironment

Yuan Li [Corresp., 1], **Zuhua Chen** [2], **Long Wu** [1], **Junjie Ye** [1], **Weiping Tao** [Corresp. 1]

[1] Department of Oncology, Renmin Hospital of Wuhan University, Wuhan, China

[2] Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Corresponding Authors: Yuan Li, Weiping Tao
Email address: liyuanpumc@gmail.com, taowpwp@sina.com

Background: Cellular heterogeneity within the tumor microenvironment is essential to tumorigenesis and tumor development. A high-resolution global view of the tumor-infiltrating immune and stromal cells in breast tumors is needed.

Methods: xCell was used to create a cellular heterogeneity map of 64 cell types in 1,092 breast tumor and adjacent normal tissues. xCell digitally dissects tissue cellular heterogeneity based on gene expression. Integrated statistical analyses were then performed.

Results: There were noticeable differences between the cell fractions in tumor tissues and normal tissues. Tumors displayed higher proportions of immune cells, including CD4+ Tem, CD8+ naïve T cells, and CD8+ Tcm compared with normal tissues. Immune inhibitory receptors (PD1, CTLA4, LAG3 and TIM3) were co-expressed on certain subtypes of T cells in breast tumors, and PD1 and CTLA4 were both positively correlated with CD8+ Tcm and CD8+ T cells. 28 cell types were significantly associated with overall survival in univariate analysis. CD4+ Tem, CD8+ Tcm, CD8+ T-cells, CD8+ naive T-cells, and B cells were positive prognostic factors but CD4+ naive T-cells were negative prognostic factors for breast cancer patients. TDRD6 and TTK are promising T cell and B cell targets for tumor vaccines. Endothelial cells and fibroblasts were significantly less prevalent in tumor tissues; astrocytes and mesangial cells were negatively correlated with the T stage. Mesangial cells and keratinocytes were found to be favorable prognostic factors and myocytes were negative prognostic factors. Five cell types were found to be independent prognostic factors and we used these to create a reliable prognostic model for breast cancer patients. Cellular heterogeneity was discovered among different breast cancer subtypes by Her2, ER, and PR status. Tri-negative patients had the highest fraction of immune cells while luminal type patients had the lowest. The various cells may have diverse or opposing roles in the prognosis of breast cancer patients.

Conclusions: We created a unique cellular map for the diverse heterogeneity of immune and stromal phenotypes within the breast tumor microenvironment. This map may lead to potential therapeutic targets and biomarkers with prognostic utility.

Cellular heterogeneity map of diverse immune and stromal phenotypes within breast tumor microenvironment

Yuan Li[1*], Zuhua Chen[2], Long Wu[1], Junjie Ye[1], Weiping Tao[1*]

1 Department of Oncology, Renmin Hospital of Wuhan University, Wuhan, China

2 Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei Province, China.

**Correspondence to*: Yuan Li and Weiping Tao, Department of Oncology, Renmin Hospital of Wuhan University, No.99, Zhangzhidong Road, Wuhan, 430060, China.

Tel: +86-27-88041911-82291.

Email: liyuanpumc@gmail.com; taowpwp@sina.com

Abstract

Background: Cellular heterogeneity within the tumor microenvironment is essential to tumorigenesis and tumor development. A high-resolution global view of the tumor-infiltrating immune and stromal cells in breast tumors is needed.

Methods: xCell was used to create a cellular heterogeneity map of 64 cell types in 1,092 breast tumor and adjacent normal tissues. xCell digitally dissects tissue cellular heterogeneity based on gene expression. Integrated statistical analyses were then performed.

Results: There were noticeable differences between the cell fractions in tumor tissues and normal tissues. Tumors displayed higher proportions of immune cells, including CD4+ Tem, CD8+ naïve T cells, and CD8+ Tcm compared with normal tissues. Immune inhibitory receptors (PD1, CTLA4, LAG3 and TIM3) were co-expressed on certain subtypes of T cells in breast tumors, and PD1 and CTLA4 were both positively correlated with CD8+ Tcm and CD8+ T cells. 28 cell types were significantly associated with overall survival in univariate analysis. CD4+ Tem, CD8+ Tcm, CD8+ T-cells, CD8+ naive T-cells, and B cells were positive prognostic factors but CD4+ naive T-cells were negative prognostic factors for breast cancer patients. TDRD6 and TTK are promising T cell and B cell targets for tumor vaccines. Endothelial cells and fibroblasts were significantly less prevalent in tumor tissues; astrocytes and mesangial cells were negatively correlated with the T stage. Mesangial cells and keratinocytes were found to be favorable prognostic factors and myocytes were negative prognostic factors. Five cell types were found to be

independent prognostic factors and we used these to create a reliable prognostic model for breast cancer patients. Cellular heterogeneity was discovered among different breast cancer subtypes by Her2, ER, and PR status. Tri-negative patients had the highest fraction of immune cells while luminal type patients had the lowest. The various cells may have diverse or opposing roles in the prognosis of breast cancer patients.

Conclusions: We created a unique cellular map for the diverse heterogeneity of immune and stromal phenotypes within the breast tumor microenvironment. This map may lead to potential therapeutic targets and biomarkers with prognostic utility.

Introduction

Breast cancer is a common cancer in women and drug resistance and distal metastasis remain major causes of mortality despite improvements in the early diagnosis and treatment of the disease (Cassetta & Pollard, 2017). Tumors have complex microenvironments composed of malignant cells, immune cells, and stromal infiltrate. Growing evidence suggests that this microenvironment plays a fundamental role in the development of malignancy and resistance to therapy (Noy & Pollard, 2014). Tumor-infiltrating cells can demonstrate either tumor-suppressing or tumor-promoting effects, depending on the cancer type. For instance, regulatory T cells (Tregs) and tumor associated macrophages (TAMs) are associated with pro-tumor functions (De Palma & Lewis, 2013; Nishikawa & Sakaguchi, 2014; Noy & Pollard, 2014), but CD8+ T cells are associated with improved clinical outcomes and better responses to immunotherapy (Tumeh et al., 2014). Research in cancer immunology has led to the development and approval of checkpoint blockers. These remarkably effective drugs augment T cell activity by blocking cytotoxic lymphocyte antigen-4 (CTLA4), programmed cell death protein 1 (PD1), and PD1 ligand (PDL1).

A better understanding of the cellular heterogeneity within the tumor microenvironment may reveal predictive biomarkers, improve existing treatments, and help to develop novel therapeutic strategies. Cellular heterogeneity is traditionally determined using flow cytometry and immunohistochemistry, however, these methods are extremely difficult to apply to solid tumors with limited throughput (Gentles et al.,

2015). Bioinformatics advancements have created novel methods that dissect cellular heterogeneity based on gene expression profiles (Abbas et al., 2009; Newman et al., 2015; Rooney et al., 2015; Shen-Orr & Gaujoux, 2013). For instance, CIBERSORT can estimate the abundances of 22 immune cell types (Newman et al., 2015). However, rare subsets of immune cells and stromal cells recognized to be important in the promotion or inhibition of tumor growth, invasion, and metastasis are ignored by CIBERSORT (Galon et al., 2006; Hanahan & Coussens, 2012). xCell can define 64 cell types within tissues, including immune and stroma cells (Aran et al., 2017a; Aran et al., 2017b). Recent analysis with xCell reveals that plasma cells and CD4+ Tcm in the tumor microenvironment may play a role in the progression of triple-negative breast cancer (Deng et al., 2019), although only immune cells were investigated. We used xCell to digitally depict the cellular heterogeneity map within the breast tumor microenvironment to reveal potential interactions and to uncover predictive biomarkers or therapeutic targets for breast cancer.

Methods and materials

Data curation and cohort characteristics

The RNA-seq data and clinical parameters from 1,092 patients with breast cancer were obtained from The Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov/). Of the 1,092 breast cancer patients included in this study, 112 paired normal tissues were identified and recurrent tumor tissues were excluded. The clinical characteristics of the cohort are listed in Table 1. 276 known Cancer/Testis (CT)

87 genes were downloaded from CTDatabase (http://www.cta.lncc.br/).

88 **Bioinformatics analysis**

89 xCell (http://xcell.ucsf.edu/) is a high-resolution gene-signature-based method for cell
90 type enrichment for up to 64 cell types, including immune and stroma cells. We used xCell
91 R package (Aran et al., 2017a) (Beta version) from GitHub in R (version 3.3.1) to
92 deconvolute the cellular heterogeneity within the breast tumor microenvironment from
93 RNA sequencing data. We determined the cellular heterogeneity of 1,092 breast tumor
94 tissues and 112 normal tissues using the xCell method. The 64 cell types were divided
95 into four groups, including 34 immune cells, 13 stromal cells, 9 stem cells, and 8 other
96 cells (**Table 2**). Over half of the 64 cell types were immune cells, providing a full view of
97 the innate and adaptive immune status with detailed cell subtypes, including CD4+ naive
98 cells, CD4+ T-cells, CD4+ Tcm, and CD4+ Tem. Stromal cells, including fibroblasts,
99 osteoblasts, and pericytes were also included. ImmuneScore and StromaScore were
100 generated by the xCell package using the sums of fractions of certain cell types (Aran et
101 al., 2017b). Tests for differences and correlations were performed. We used the t-
102 Distributed Stochastic Neighbor Embedding (t-SNE) method with tsne package (version
103 0.1-3) to perform cluster analysis based on cell fraction types.

104 **Survival analysis**

105 Univariate and multivariate COX regressions were performed using the survival
106 package (version 3.1-7) to search for survival-associated genes. The best cutoff value for
107 each factor was determined using the Survminer package (version 0.4.6). Significant

prognostic factors were displayed in a forest plot and the most significant factors were further evaluated using multivariate analysis. The final prognostic model was built with five independent factors including CD8+ T cells, mesangial cells, NKT, keratinocytes, and class-switched memory B cells. We used the survivalROC package (Version 1.0.3) in R, which uses a time-dependent ROC curve estimation with censored data (Heagerty et al., 2000) to compare the aptitude of the individual prognostic factors. The final prognostic model was used to generate the area under the curve (AUC) of the receiver-operator characteristic (ROC) curve for each parameter.

Statistical analysis

Differentially enriched cell types between groups were compared using the Student's t-test (two groups) or one-way ANOVA analysis (three groups). Correlation analyses were performed using the Spearman method. The survival curves were compared using the Kaplan-Meier method and log-rank test. All tests were two-sided and $p < 0.05$ was considered to be statistically significant unless otherwise noted. Data were analyzed using R (version 3.4.4).

Results

Breast tumor tissues had higher fractions of immune cells than normal tissues.

The median fractions for each cell type were calculated for normal and breast tumor tissues and the proportions of the 64 cells were found to differ between breast tumor and normal tissues (**Figure 1A, Table 2**). Breast tumor tissue had higher fraction of immune

cells with red to light blue markers whereas normal tissue had larger proportions of stem and stromal cells with blue to red markers (**Figure 1A**). Unsupervised cluster analysis revealed that breast tumor tissues and the adjacent normal tissues were clustered into different groups. Immune cells were also clustered into several subgroups (**Figure 1B**), indicating that the cellular heterogeneity in tumor vs. normal tissues was much greater than that in a single sample. Dimensionality reduction and visualization by t-Distributed Stochastic Neighbor Embedding (t-SNE) also suggested clear difference between the tumor and normal tissues (**Figure 1C**).

We compared the fractions of each cell type between breast tumor and normal tissues, revealing dramatic differences in the number of cell types between the tumor and normal tissues (**Figure 1D-W)**. Immune cells tended to be more diverse compared to stem or stromal cells. For innate immune cells, neutrophils were more prevalent in normal tissues whereas eosinophils were higher in tumor tissues (**Figure 1D-E**). There was not a significant difference in DC cells between normal and tumor tissues. However, iDC was significantly lower in tumor tissues but pDC and aDC were significantly higher (**Figure 1F-I**). This phenomenon was also seen in macrophages, in which macrophage M1 was higher in tumor tissues while macrophage M2 was lower (**Figure 1J-K**). CD4+ Tcm was found to be significantly lower in tumor tissues, while CD4+ Tem, CD8+ naïve T cells, and CD8+ Tcm were significantly higher (**Figure 1L-O**). Plasma cells, pro B cells, Tgd, Th1, Th2 cells, and Tregs were also found to be significantly higher in tumor tissues (**Figure 1P-U**). Representative stromal cells, such as endothelial cells and fibroblasts, were found

150 to be significantly lower in tumor tissues (**Figure 1V-W**). Differential analysis with paired
151 tumor and normal tissues showed similar patterns (Table 2).
152
153 **Inhibitory receptors were co-expressed on certain subtypes of T cells**
154 Inhibitory receptors, including PD1, CTLA4, LAG3, and TIM3, expressed on T cells
155 often led to T-cell exhaustion allowing tumors to evade the immune response (Huang et
156 al., 2017; Nirschl & Drake, 2013). The use of specific antibodies to inhibit CTLA4 or PD1
157 and overcome immune suppression and tumor regression is promising (Brahmer et al.,
158 2012; Callahan et al., 2010).
159 We investigated the correlations between these inhibitory receptors and CD4+/CD8+
160 T cells. Heatmaps suggested that the expression patterns of the inhibitory receptors
161 correlated with specific subsets of T cells with distinctions between tumor tissues and
162 normal tissues (**Figures 2A and 2B**). Correlation analyses also demonstrated that CD8+
163 T-cells, CD8+ Tcm, CD8+ naive T-cells, CD4+ memory T cells, and CD4+ naïve T cells
164 were all positively correlated with expressions of these inhibitory receptors in tumor
165 tissues, especially with PD1 and CTLA4 ($P<0.05$) (**Figures 2C and D**). CD8+ Tem, CD4+
166 Tcm, and CD4+ T-cells were not strongly correlated with the expressions of the inhibitory
167 receptors (**Figure 2C**). Only a few T cells were significantly correlated with these inhibitory
168 receptors in normal tissues; TIM3 expression was negatively correlated with CD4+ Tcm
169 (**Figure 2C and F**). We observed a significant correlation among the expression of
170 inhibitory receptors (**Figure 2E and G**).

171

172 **Cancer/testis genes TDRD6 and TTK show promise as breast cancer targets**

173 Cancer/Testis (CT) genes are a cluster of tumor-associated proteins normally
174 expressed in germ cells and different cancers. However, they are not typically seen in
175 normal somatic cells (Scanlan et al., 2002). The limited expression of CT genes makes
176 them ideal cancer and immunotherapy biomarkers.

177 We studied the antitumor immunity response to antigens generated by CT genes by
178 examining 276 known CT genes (obtained from the CTDatabase) for their association
179 with immune components. The significant associations between immune cells and CT
180 genes ($P<0.001$) are shown in **Figure 3A**. Most of the adaptive immune cells were
181 significantly correlated with CT genes. T cells, such as CD8+ T cells and aDC, which
182 belonged to adaptive and innate immune responses, were positively correlated with most
183 of the CT genes (**Figure 3A-C**). Moreover, two CT genes, TDRD6 and TTK, were
184 positively correlated with a number of immune cells, especially the CD4+/CD8+ T cells
185 (**Figure 3D and E**), implying strong host immune reactions to these two cancer antigens.

186

187 **Cellular heterogeneity correlated with clinical pathology of breast cancer**

188 Cellular heterogeneity is an important part of the tumor microenvironment and is
189 necessary for the growth and development of a tumor. We studied whether certain cell
190 types were significantly correlated with certain clinical parameters, including age, sex, T
191 stage, N stage, M stage, and TNM stage. A number of cell types were significantly

192 correlated with clinical parameters, especially T stage and M stage (**Figure 4A**).
193 Astrocytes, mesangial cells, and mast cells were negatively correlated with the T stage
194 and plasma cells were positively correlated with the T stage (**Figure 4B-E**). CD4+ Tcm,
195 CD4+ Tem, microvascular (mv) endothelial cells, NKT, and MSC were all significantly
196 higher at the M stage in patients with distal metastasis (**Figure 4F-J**). CLP was
197 significantly higher at the N stage in patients with lymph node metastasis (**Figure 4K**).
198 Th1 cells and MSC were both positively correlated with TNM stage (**Figure 4L-M**). The
199 12 male breast cancer patients studied tended to have higher proportion of CLP and NKT
200 compared with the female breast cancer patients (**Figure 4N-O**).
201
202 **Prognostic model with survival associated cell types**
203 Emerging evidence suggests that the number of tumor-infiltrating lymphocytes (TILs)
204 of primary tumors consistently predicts favorable outcomes for a number of tumor types,
205 including breast cancer. Therefore, survival analyses were performed to find survival-
206 associated cell types within the tumor microenvironment (**Figure 5A)**. Immune cells were
207 more strongly associated with overall survival, especially CD4+ and CD8+ T cells (**Figure**
208 **5A**). Most T cells, including CD8+ T cells, CD8+ Tcm, CD8+ naïve T cells, and CD4 Tem,
209 were favorable prognostic factors. However, high CD4+ naïve T cells were associated
210 with worse overall survival (**Figure 5A-M**). NKT, class switched memory B cells, NK cells,
211 cDC, and pDC were also significantly associated with overall survival (**Figure 5A-M**). A
212 number of stromal cells, including mesangial cells and keratinocytes, were favorable

213 prognostic factors but myocytes were adverse prognostic factors (**Figure 5A-M**).
214 Multivariate COX regression revealed that CD8+ T cells, mesangial cells, keratinocytes,
215 NKT, and class switched memory B cells were independent prognostic factors. We built
216 a prognosis predictor model with five independent prognostic factors. Our model more
217 reliably determined the survival of breast cancer patients with the highest AUC of ROC of
218 0.708, versus when the factors were analyzed separately (**Figure 5N, 5O**).
219
220 **Subtypes of breast cancer had diverse phenotypes of cellular heterogeneity**
221 Emerging evidence suggests that the breast cancer transcriptome has a wide range
222 of intratumoral heterogeneity, as well as genomic heterogeneity based on ER, PR, and
223 Her2, which are determined by the tumor cells and immune cells in the surrounding
224 microenvironment (Chung et al., 2017). We explored cellular heterogeneity among
225 different subtypes of breast cancer by Her2, ER, and PR status. 1,092 breast cancer
226 patients were classified into five groups according the clinicopathological parameters
227 provided by TCGA, including 30 Her2+_HR- patients, 59 Her2+_HR+ patients, 426
228 Luminal type (Her2-_HR+) patients, 97 triple negative (Tri-negative) patients, and 480
229 unknown patients (Table 1). The relative proportion of different cells varied widely among
230 these five subtypes (Figure 6). Tri-negative patients had the highest fraction of immune
231 cells while luminal type patients had the smallest fraction of immune cells, especially
232 CD4+ and CD8+ T cells (**Figure 6A, Supplementary Figure S1**). Cluster analyses from
233 ImmuneScore, StromaScore, and heatmap suggested the absence of certain cell types

used to distinguish these five subtypes (**Figure 6B**). Furthermore, t-SNE cluster analysis suggested a large difference in tumor-infiltrating cells among these five subtypes (**Figure 6C and D**). B cells, T cells, macrophages, Th cells and stromal cells, including keratinocytes, were significantly differentially enriched in these subtypes (**Figure 6E-P and Figure S2**). Tri-negative breast cancer tissues had the highest fractions of plasma cells, pro B cells, macrophages M1, Th1, and Th2 cells, but M2 cells had the lowest fraction of macrophages (**Figure 6E-L**). Keratinocytes, sebocytes, and pericytes were found frequently in Tri-negative breast cancer whereas MSC cells were found in low amounts (**Figure 6M-P**). Survival analysis revealed interesting differences between the five subtypes (**Figure 7**). Each subtype of breast cancer had a unique pattern of survival-associated tumor-infiltrating cells. Different cell types may have different functions in the prognosis of breast cancer patients. Keratinocytes had a favorable effect on the prognostic factors while neurons were associated with adverse prognosis factors in luminal type patients (**Figure 7A**). However, in Tri-negative patients, keratinocytes predicted a worse overall survival and neurons predicted a better overall survival (**Figure 7D**). Taken together, the diversity of cellular heterogeneity among the different subtypes of breast cancers suggested that tumor-infiltrating cells within the tumor microenvironment were essential in shaping the intratumor heterogeneity of breast cancer.

Discussion

255 We observed distinct tumor-infiltrating cell types within the tumor microenvironment.
256 The abundance and activation status of these cell types is of interest to researchers for
257 their novel bioinformatic techniques. Tumor-infiltrating cells are known to play important
258 roles in the regulation of tumor proliferation, metastasis, and invasion (Galon et al., 2006;
259 Hanahan & Coussens, 2012). The rapid accumulation of high-throughput data and the
260 evolution of bioinformatics algorithms allows us to digitally dissect the interactions
261 between tumors cells and tumor-infiltrating cells, including immune cells and stromal cells
262 (Aran et al., 2017a; Hackl et al., 2016). The high-throughput approach may help
263 understand the complexity of the tumor microenvironment and lead to innovations in
264 breast cancer treatment and prognosis. xCell analysis reveals that plasma cells and CD4+
265 Tcm in the tumor microenvironment may play a role in the progression of triple-negative
266 breast cancer (Deng et al., 2019), although only immune cells were investigated.
267 We used the digital deconvolution from xCell to determine the cellular heterogeneity
268 within breast tumor and normal tissues. A total of 64 cell types with more than 30 immune
269 cell types were characterized at high resolution. This was the most studied set of cell
270 types, especially for tumor-infiltrating lymphocytes (TILs). We focused on immune cell
271 types, especially the CD4+/CD8+ T cells, and discovered differences between breast
272 tumor tissues and adjacent normal tissues with polarized enrichment of certain cell types.
273 Our results demonstrated that the expression of inhibitory receptors (including PD1,
274 CTLA4, LAG3, and TIM3) were positively correlated and were associated with certain
275 types of T cells in tumor tissues, especially CD8+ Tcm and CD8+ T cells. CD4+ Tem,

CD8+ Tcm, CD8+ T-cells, CD8+ naive T-cells, and B cells were associated with better prognosis whereas CD4+ naive T-cells were negatively associated with prognosis for breast cancer patients. Innate and adaptive immune cells had active immune responses to tumor antigens, including T cells, B cells and DC. TDRD6 and TTK are promising targets for cancer vaccines that could activate a number of immune cells, especially T cells and B cells. Stromal cells were also widely involved in the development of breast cancer. Endothelial cells and fibroblasts were not observed as frequently in tumor tissues. Astrocytes and mesangial cells were negatively correlated with T stage. Mesangial cells and keratinocytes were favorable prognostic factors and myocytes were adverse prognostic factors. We built a prognosis predictor with survival-associated cell types to determine the overall survival of breast cancer patients. Cellular heterogeneity was also profiled in different subtypes of breast cancer based on Her2, ER, and PR status. Five subtypes of breast cancer demonstrated various phenotypes and the cell types may have had different or opposing roles in each subtype of breast cancer.

Immunotherapies, including immune checkpoint blockers, therapeutic vaccines, and engineered T cells are being intensively investigated (Schumacher & Schreiber, 2015) to determine how tumor cells interact with immune cells. The tumor-immune cell interaction poses considerable challenges since the development of cancer and immune surveillance by innate and adaptive immune cells with plasticity and memory are evolving ecosystems. The complex interplay between solid tumors and host immunity has been widely studied but is not well understood. Tumor infiltrating lymphocytes (TILs) have been associated

297 with clinical outcomes in many tumor types (Anagnostou & Brahmer, 2015; Schoenfeld,
298 2015). For example, CD8+ TILs are prognostically favorable in melanoma, colorectal,
299 ovarian, and non-small cell lung cancer. CD8+ TILs are able to kill tumor cells in specific
300 cancers (Yee et al., 2002). Immunity in breast cancer remains largely unstudied with only
301 a few preliminary evaluations on the prognostic value of CD4+/CD8+ T lymphocytes. The
302 presence of TILs is potentially predictive and prognostic in specific breast cancer
303 subtypes, especially in patients with human epidermal growth factor receptor 2 positive
304 and triple-negative breast cancer. Large adjuvant studies have shown that higher levels
305 of TILs in primary biopsies are associated with improved overall survival and fewer
306 recurrences, regardless of therapy (Adams et al., 2014; Dieci et al., 2015; Loi et al., 2013).
307 We provided detailed information about the immune cells in breast cancer with
308 numerous novel findings. Inhibitory receptors were expressed on certain types of T cells,
309 preferring CD8+ T cells and CD8+ Tcm. The co-expression of PD1, CTLA4, LAG3, and
310 TIM3 were more commonly observed in tumor tissues compared with normal tissues,
311 which may explain the limited effects of a single immune checkpoint inhibitor and the use
312 of combined strategies. The simultaneous inhibition of PD1 and CTLA4 (Wolchok et al.,
313 2013) or TIM3 (Fourcade et al., 2010) in advanced melanoma patients show promise in
314 clinical trials. CD8+ naive T cells versus CD4+ naive T cells were favorable prognostic
315 factors for the overall survival of breast cancer patients, suggesting that not all T cells
316 were protective. These results suggest that the upregulated co-expression of multiple
317 immune inhibitory receptors may contribute to immune suppression. More attention

should be paid to subtypes of T cells when using immune checkpoint blockers since immune cells are highly conditional and may have different or even opposing roles in response to tumor cells.

Growing evidence suggests that immune cells and tumor cell-extrinsic factors, including fibroblasts, endothelial cells, adipocytes within the tumor microenvironment have important roles in inhibiting apoptosis, enabling immune evasion, and promoting proliferation, angiogenesis, invasion, and metastasis (Whiteside, 2008). We found that endothelial cells were significantly higher in adjacent normal tissues (Figure 1J) and breast cancer patients with metastasis had a higher fraction of microvascular (mv) endothelial cells (Figure 4C). A high level of mv endothelial cells was significantly associated with worse overall survival (Figure 5A). Recent studies have shown that endothelial cells may promote triple-negative breast cancer cell metastasis via PAI-1 and CCL5 signaling (Zhang et al., 2018). The presence of endothelial cells significantly enhanced the angiogenic activity of breast cancer cells (Buchanan et al., 2012). These results support our analysis and further study of the clinical relevance of these cell types may provide novel insights into the initiation and progression of breast cancer.

We analyzed and described the potential roles of different tumor-infiltrating cells. Our study would benefit from additional analysis and experimental validations to further investigate the roles of the 64 types of cells profiled in this study.

Conclusions

We revealed the landscape of cellular heterogeneity at high resolution and provided

novel insights into cell interactions within the tumor microenvironment in breast cancer. Our results may assist in the development of future therapeutic and predictive strategies. Further study should focus on the subtypes of immune cells and stromal cells identified in this study.

Reference:

Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, and Clark HF. 2009. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One* 4:e6098. 10.1371/journal.pone.0006098

Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, Martino S, Wang M, Jones VE, Saphner TJ, Wolff AC, Wood WC, Davidson NE, Sledge GW, Sparano JA, and Badve SS. 2014. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *Journal of Clinical Oncology* 32:2959-2966. 10.1200/JCO.2013.55.0491

Anagnostou VK, and Brahmer JR. 2015. Cancer immunotherapy: a future paradigm shift in the treatment of non-small cell lung cancer. *Clin Cancer Res* 21:976-984. 10.1158/1078-0432.CCR-14-1187

Aran D, Hu Z, and Butte AJ. 2017a. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *bioRxiv*. 10.1101/114165

Aran D, Hu Z, and Butte AJ. 2017b. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology* 18:220. 10.1186/s13059-017-1349-1

Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, Drake CG, Camacho LH, Kauh J, Odunsi K, Pitot HC, Hamid O, Bhatia S, Martins R, Eaton K, Chen S, Salay TM, Alaparthy S, Grosso JF, Korman AJ, Parker SM, Agrawal S, Goldberg SM, Pardoll DM, Gupta A, and Wigginton JM. 2012. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* 366:2455-2465. 10.1056/NEJMoa1200694

Buchanan CF, Szot CS, Wilson TD, Akman S, Metheny-Barlow LJ, Robertson JL, Freeman JW, and Rylander MN. 2012. Cross-talk between endothelial and breast cancer cells regulates reciprocal expression of angiogenic factors in vitro. *Journal of Cellular Biochemistry* 113:1142-1151. 10.1002/jcb.23447

Callahan MK, Wolchok JD, and Allison JP. 2010. Anti-CTLA-4 antibody therapy: immune monitoring during clinical development of a novel immunotherapy. *Semin Oncol* 37:473-484. 10.1053/j.seminoncol.2010.09.001

Cassetta L, and Pollard JW. 2017. Repolarizing macrophages improves breast cancer therapy. *Cell Research* 27:963-964. 10.1038/cr.2017.63

Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, Ryu HS, Kim S, Lee JE, Park YH, Kan Z, Han W, and Park WY. 2017. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nature Communications* 8:15081. 10.1038/ncomms15081

De Palma M, and Lewis CE. 2013. Macrophage regulation of tumor responses to anticancer therapies. *Cancer Cell* 23:277-286. 10.1016/j.ccr.2013.02.013

Deng L, Lu D, Bai Y, Wang Y, Bu H, and Zheng H. 2019. Immune Profiles of Tumor Microenvironment and Clinical Prognosis among Women with Triple-Negative Breast Cancer. *Cancer Epidemiology, Biomarkers and Prevention* 28:1977-1985. 10.1158/1055-9965.EPI-19-0469

Dieci MV, Mathieu MC, Guarneri V, Conte P, Delaloge S, Andre F, and Goubar A. 2015. Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. *Annals of Oncology* 26:1698-1704. 10.1093/annonc/mdv239

Fourcade J, Sun Z, Benallaoua M, Guillaume P, Luescher IF, Sander C, Kirkwood JM, Kuchroo V, and Zarour HM. 2010. Upregulation of Tim-3 and PD-1 expression is associated with tumor antigen-specific CD8+ T cell dysfunction in melanoma patients. *J Exp Med* 207:2175-2186. 10.1084/jem.20100637

Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoue F, Bruneval P, Cugnenc PH, Trajanoski Z, Fridman WH, and Pages F. 2006. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 313:1960-1964. 10.1126/science.1129139

Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M, West RB, Plevritis SK, and Alizadeh AA. 2015. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 21:938-945. 10.1038/nm.3909

Hackl H, Charoentong P, Finotello F, and Trajanoski Z. 2016. Computational genomics tools for dissecting tumour-immune cell interactions. *Nat Rev Genet* 17:441-458. 10.1038/nrg.2016.67

Hanahan D, and Coussens LM. 2012. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 21:309-322. 10.1016/j.ccr.2012.02.022

Heagerty PJ, Lumley T, and Pepe MS. 2000. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56:337-344.

Huang RY, Francois A, McGray AR, Miliotto A, and Odunsi K. 2017. Compensatory upregulation of PD-1, LAG-3, and CTLA-4 limits the efficacy of single-agent checkpoint blockade in metastatic ovarian cancer. *Oncoimmunology* 6:e1249561. 10.1080/2162402X.2016.1249561

Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E, de Azambuja E, Quinaux E, Di Leo A, Michiels S, Piccart MJ, and Sotiriou C. 2013. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *Journal of Clinical Oncology* 31:860-867. 10.1200/JCO.2011.41.0902

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, and Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12:453-457. 10.1038/nmeth.3337

411 Nirschl CJ, and Drake CG. 2013. Molecular pathways: coexpression of immune checkpoint molecules: signaling
412 pathways and implications for cancer immunotherapy. *Clin Cancer Res* 19:4917-4924. 10.1158/1078-
413 0432.CCR-12-1972
414 Nishikawa H, and Sakaguchi S. 2014. Regulatory T cells in cancer immunotherapy. *Curr Opin Immunol* 27:1-7.
415 10.1016/j.coi.2013.12.005
416 Noy R, and Pollard JW. 2014. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* 41:49-61.
417 10.1016/j.immuni.2014.06.010
418 Rooney MS, Shukla SA, Wu CJ, Getz G, and Hacohen N. 2015. Molecular and genetic properties of tumors associated
419 with local immune cytolytic activity. *Cell* 160:48-61. 10.1016/j.cell.2014.12.033
420 Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, and Chen YT. 2002. Cancer/testis antigens: an expanding family of
421 targets for cancer immunotherapy. *Immunol Rev* 188:22-32.
422 Schoenfeld JD. 2015. Immunity in head and neck cancer. *Cancer Immunol Res* 3:12-17. 10.1158/2326-6066.CIR-14-
423 0205
424 Schumacher TN, and Schreiber RD. 2015. Neoantigens in cancer immunotherapy. *Science* 348:69-74.
425 10.1126/science.aaa4971
426 Shen-Orr SS, and Gaujoux R. 2013. Computational deconvolution: extracting cell type-specific information from
427 heterogeneous samples. *Curr Opin Immunol* 25:571-578. 10.1016/j.coi.2013.09.015
428 Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu
429 V, West AN, Carmona M, Kivork C, Seja E, Cherry G, Gutierrez AJ, Grogan TR, Mateus C, Tomasic G,
430 Glaspy JA, Emerson RO, Robins H, Pierce RH, Elashoff DA, Robert C, and Ribas A. 2014. PD-1 blockade
431 induces responses by inhibiting adaptive immune resistance. *Nature* 515:568-571. 10.1038/nature13954
432 Whiteside TL. 2008. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* 27:5904-5912.
433 10.1038/onc.2008.271
434 Wolchok JD, Kluger H, Callahan MK, Postow MA, Rizvi NA, Lesokhin AM, Segal NH, Ariyan CE, Gordon RA,
435 Reed K, Burke MM, Caldwell A, Kronenberg SA, Agunwamba BU, Zhang X, Lowy I, Inzunza HD, Feely
436 W, Horak CE, Hong Q, Korman AJ, Wigginton JM, Gupta A, and Sznol M. 2013. Nivolumab plus
437 ipilimumab in advanced melanoma. *N Engl J Med* 369:122-133. 10.1056/NEJMoa1302369
438 Yee C, Thompson JA, Byrd D, Riddell SR, Roche P, Celis E, and Greenberg PD. 2002. Adoptive T cell therapy using
439 antigen-specific CD8+ T cell clones for the treatment of patients with metastatic melanoma: in vivo
440 persistence, migration, and antitumor effect of transferred T cells. *Proc Natl Acad Sci U S A* 99:16168-16173.
441 10.1073/pnas.242600099
442 Zhang W, Xu J, Fang H, Tang L, Chen W, Sun Q, Zhang Q, Yang F, Sun Z, Cao L, Wang Y, and Guan X. 2018.
443 Endothelial cells promote triple-negative breast cancer cell metastasis via PAI-1 and CCL5 signaling. *FASEB*
444 *Journal* 32:276-288. 10.1096/fj.201700237RR

445

446

447

448 **Figure legends:**

449 **Figure 1. Differences of cellular heterogeneity between breast tumor tissue and**
450 **normal tissues.** A. Median fractions of 64 cell types in breast tumor and normal tissues.
451 64 cell types were grouped into four groups: immune, stem, stromal, and other cells. B.
452 Heatmap of fractions of 64 cell types in 1,092 breast tumor tissues and 112 adjacent
453 normal tissues. C. Dimensionality reduction and visualization by t-Distributed Stochastic
454 Neighbor Embedding (t-SNE) clustering based on cell fractions. D. to W. Dot plots of
455 fractions of certain cell types in breast tumor and normal tissues. Lines between dots
456 indicated paired tissues from the same breast cancer patient. *, $P<0.05$. **, $P<0.01$. ***,
457 $P<0.001$.

458

459 **Figure 2. Expression patterns of inhibitory receptors on CD4+/CD8+ T cells.** A and
460 B. Heatmaps of expression of inhibitory receptors including PD1, CTLA4, LAG3 and
461 TIM3, and fractions of CD4+/CD8+ T cells in tumor tissues (A) and normal tissues (B).
462 Data were transformed by rank and normalized. C. Clustered correlation matrixes among
463 inhibitory receptors and CD4+/CD8+ T cells in tumor tissues (up-right triangle) and normal
464 tissues (low-left triangle). D. Dot plot of correlations between PD1 expression and
465 fractions of CD8+ Tcm in tumor tissues. E. Dot plot of correlations between PD1
466 expression and CTLA4 expression in tumor tissues. F. Dot plot of correlations between
467 TIM3 expression and fractions of CD4+ Tcm in normal tissues. G. Dot plot of correlations
468 between PD1 expression and LAG3 in normal tissues.

469

470 **Figure 3. Correlations between cancer/testis genes and immune cells.** A. Significant
471 correlations between cancer/testis (CT) genes and immune cells. Scaled color dots
472 represented significant correlations between CT genes and immune cells (P<0.001) and
473 red dots represented positive correlations while blue dots represent negative correlations.
474 B. and C. CD8+ naïve T-cells and aDC were positively correlated with most of the CT
475 genes. D. and E. TDRD6 and TTK were positively correlated with a number of immune
476 cells.

477

478 **Figure 4. Involvement of cellular heterogeneity in clinic-pathology of breast cancer.**
479 A. A number of cell types were significantly correlated with clinical parameters. B. to O.
480 Examples of significant correlations between different cell types and clinical parameters,
481 y-axis represents the fractions of each cell type. *, P<0.05. **, P<0.01. ***, P<0.001.

482

483 **Figure 5. Survival associated tumor-infiltrating cells in breast cancer.** A. Forest plot
484 of hazard ratios of survival associated cell types. B. to M. Kaplan-Meier curves of survival
485 associated cell types. Red lines indicated high fraction while blue lines indicated low
486 fraction of each cell types, respectively. N. Kaplan-Meier curves of predictor built with five
487 independent prognostic factors (P<0.01). O. ROC curves of prognostic model and the five
488 independent prognostic factors.

489

Figure 6. Differences of cellular heterogeneity among different subtypes of breast cancer. A. Median fraction of 64 cell types in five subtypes of breast tumor. B. Cluster analysis by ImmuneScore and StromaScore, which were calculated by summing up the fractions of immune and stromal cells, respectively. C. Dimensionality reduction and visualization by t-Distributed Stochastic Neighbor Embedding (t-SNE) clustering with fractions of all the 64 cell types. D. Heatmap of fractions of 64 cell types in five subtypes of breast cancer. E. to P. Box plots with dots of fractions of certain cell types in five subtypes of breast cancer. *, P<0.05. **, P<0.01. ***, P<0.001, ****, P<0.0001.

Figure 7. Survival associated tumor-infiltrating cells in five subtypes of breast cancer. A. to E. Forest plots of hazard ratios of survival associated cell types in five subtypes of breast cancer.

Supplementary Figure S1. Median fractions of 64 types of cells in five subtypes of breast cancer.

Supplementary Figure S2. Diverse differences of 64 types of cells among the five subtypes of breast cancer.

Figure 1

Differences of cellular heterogeneity between breast tumor tissue and normal tissues.

A. Median fractions of 64 cell types in breast tumor and normal tissues. 64 cell types were grouped into four groups: immune, stem, stromal, and other cells. B. Heatmap of fractions of 64 cell types in 1,092 breast tumor tissues and 112 adjacent normal tissues. C. Dimensionality reduction and visualization by t-Distributed Stochastic Neighbor Embedding (t-SNE) clustering based on cell fractions. D. to W. Dot plots of fractions of certain cell types in breast tumor and normal tissues. Lines between dots indicated paired tissues from the same breast cancer patient. *, $P<0.05$. **, $P<0.01$. ***, $P<0.001$.

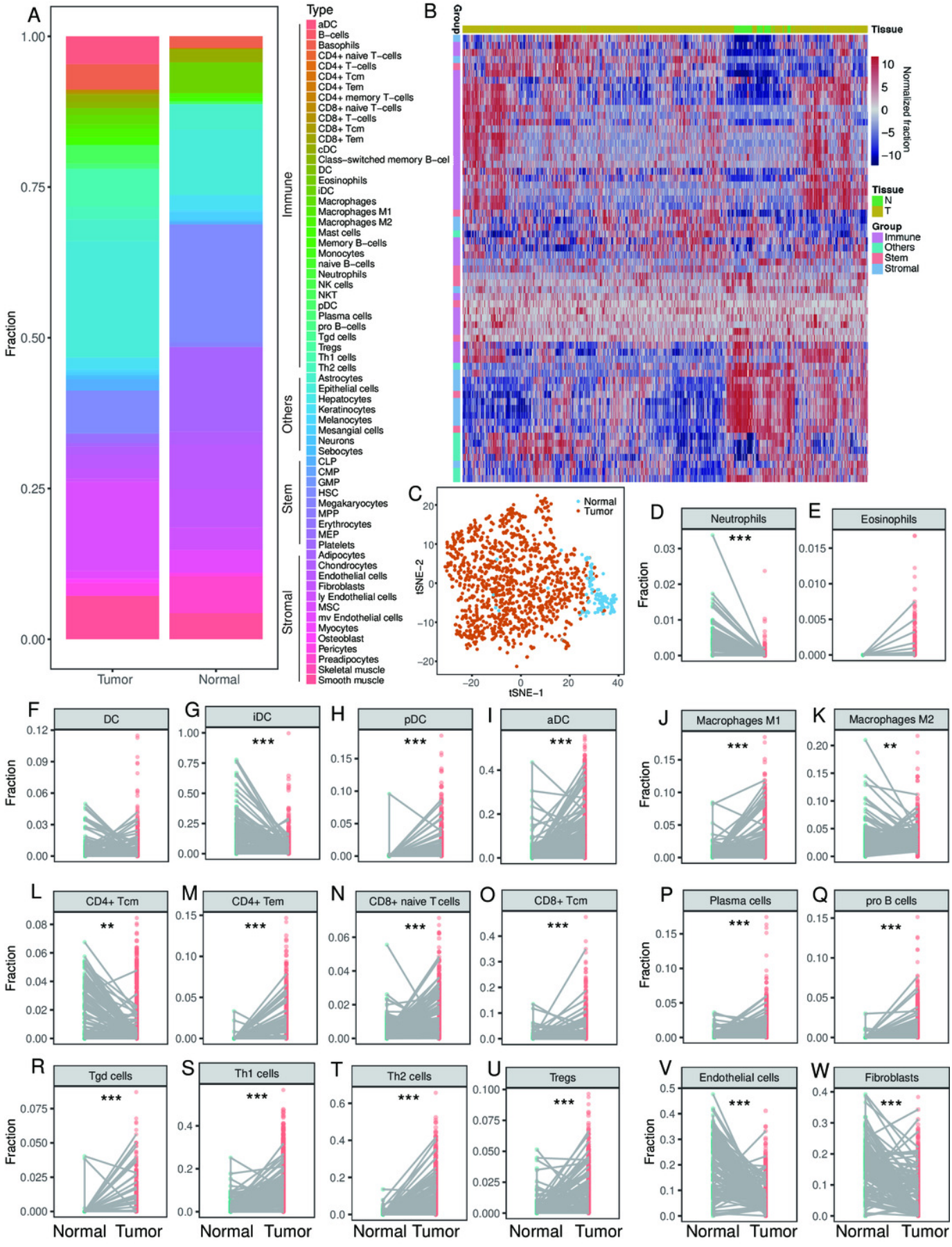
A
Fraction
1.00
0.75
0.50
0.25
0.00
Tumor
Normal
Type
aDC
B−cells
Basophils
CD4+ naive T−cells
CD4+ T−cells
CD4+ Tcm
CD4+ Tem
CD4+ memory T−cells
CD8+ naive T−cells
CD8+ T−cells
CD8+ Tcm
CD8+ Tem
cDC
Class−switched memory B−cel
DC
Eosinophils
iDC
Macrophages
Macrophages M1
Macrophages M2
Mast cells
Memory B−cells
Monocytes
naive B−cells
Neutrophils
NK cells
NKT
pDC
Plasma cells
pro B−cells
Tgd cells
Tregs
Th1 cells
Th2 cells
Astrocytes
Epithelial cells
Hepatocytes
Keratinocytes
Melanocytes
Mesangial cells
Neurons
Sebocytes
CLP
CMP
GMP
HSC
Megakaryocytes
MPP
Erythrocytes
MEP
Platelets
Adipocytes
Chondrocytes
Endothelial cells
Fibroblasts
ly Endothelial cells
MSC
mv Endothelial cells
Myocytes
Osteoblast
Pericytes
Preadipocytes
Skeletal muscle
Smooth muscle
Immune
Others
Stem
Stromal
B
Group
Tissue
10
5
0
−5
−10
Normalized fraction
Tissue
N
T
Group
Immune
Others
Stem
Stromal
C
Normal
Tumor
tSNE−2
tSNE−1
D
Neutrophils
E
Eosinophils
Fraction
F
DC
G
iDC
H
pDC
I
aDC
J
Macrophages M1
K
Macrophages M2
L
CD4+ Tcm
M
CD4+ Tem
N
CD8+ naive T cells
O
CD8+ Tcm
P
Plasma cells
Q
pro B cells
R
Tgd cells
S
Th1 cells
T
Th2 cells
U
Tregs
V
Endothelial cells
W
Fibroblasts
Normal Tumor

Figure 2

Expression patterns of inhibitory receptors on CD4+/CD8+ T cells.

A and B. Heatmaps of expression of inhibitory receptors including PD1, CTLA4, LAG3 and TIM3, and fractions of CD4+/CD8+ T cells in tumor tissues (A) and normal tissues (B). Data were transformed by rank and normalized. C. Clustered correlation matrixes among inhibitory receptors and CD4+/CD8+ T cells in tumor tissues (up-right triangle) and normal tissues (low-left triangle). D. Dot plot of correlations between PD1 expression and fractions of CD8+ Tcm in tumor tissues. E. Dot plot of correlations between PD1 expression and CTLA4 expression in tumor tissues. F. Dot plot of correlations between TIM3 expression and fractions of CD4+ Tcm in normal tissues. G. Dot plot of correlations between PD1 expression and LAG3 in normal tissues.

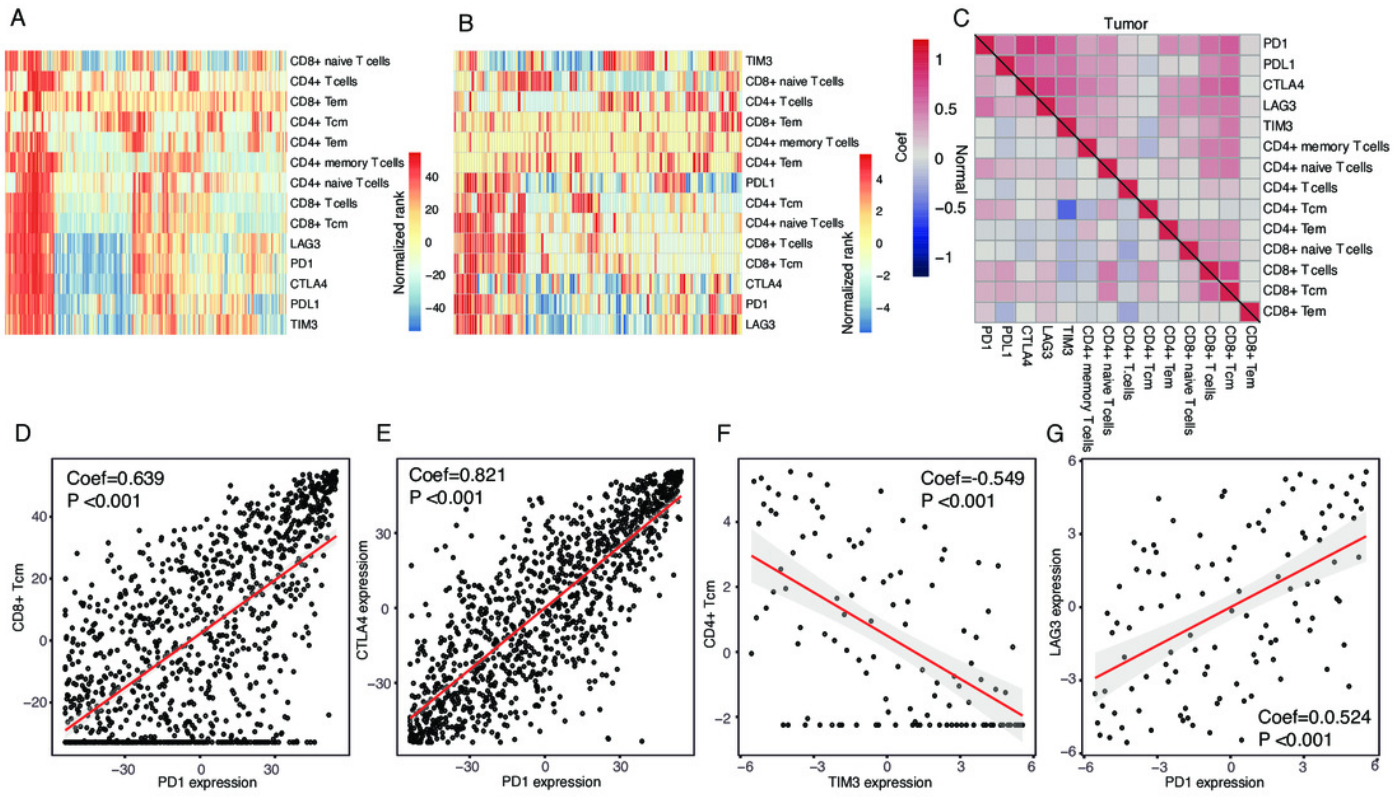

Figure 3

Correlations between cancer/testis genes and immune cells.

A. Significant correlations between cancer/testis (CT) genes and immune cells. Scaled color dots represented significant correlations between CT genes and immune cells ($P<0.001$) and red dots represented positive correlations while blue dots represent negative correlations. B. and C. CD8+ naïve T-cells and aDC were positively correlated with most of the CT genes. D. and E. TDRD6 and TTK were positively correlated with a number of immune cells.

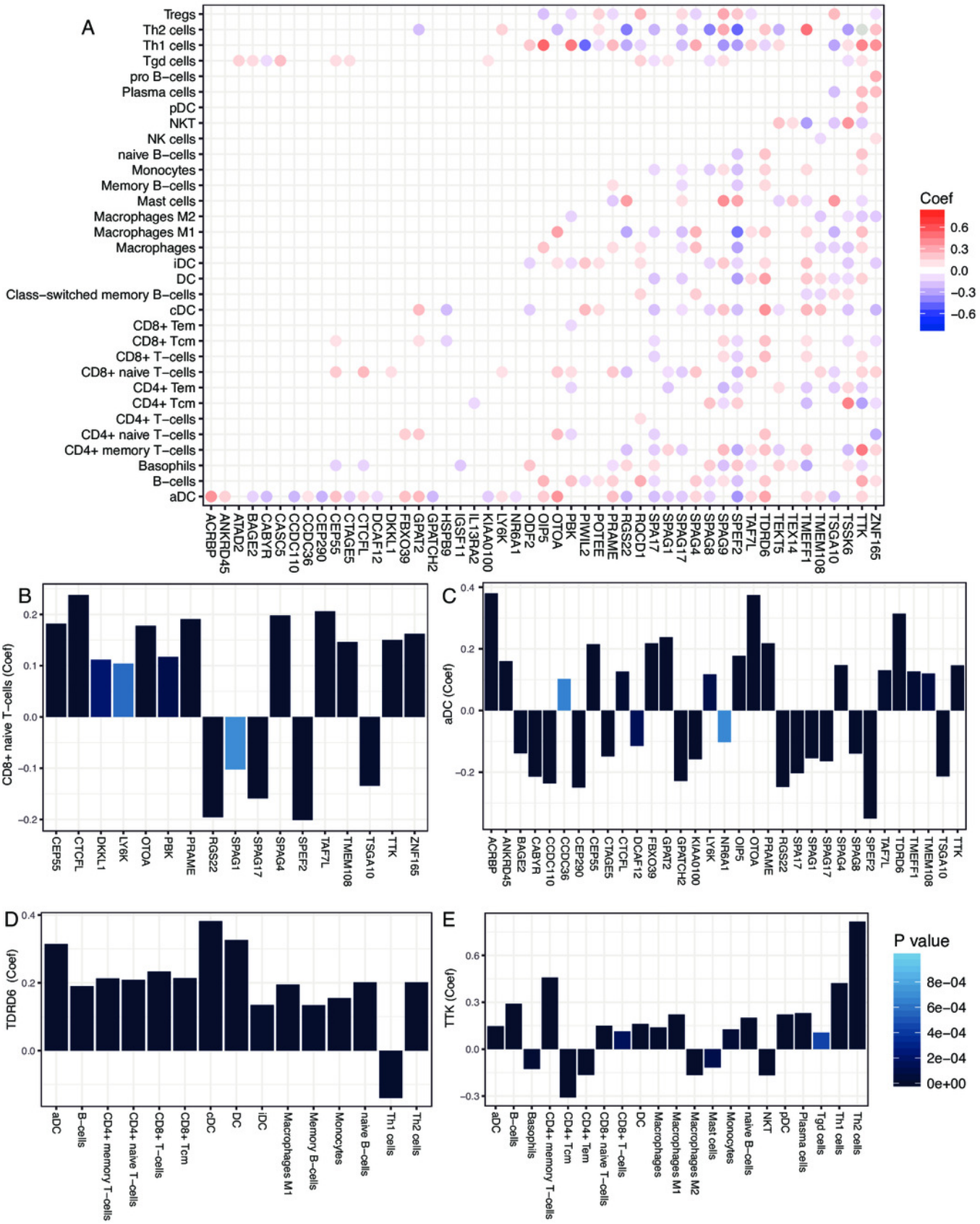
A
Tregs
Th2 cells
Th1 cells
Tgd cells
pro B−cells
Plasma cells
pDC
NKT
NK cells
naive B−cells
Monocytes
Memory B−cells
Mast cells
Macrophages M2
Macrophages M1
Macrophages
iDC
DC
Class−switched memory B−cells
cDC
CD8+ Tem
CD8+ Tcm
CD8+ T−cells
CD8+ naive T−cells
CD4+ Tem
CD4+ Tcm
CD4+ T−cells
CD4+ naive T−cells
CD4+ memory T−cells
Basophils
B−cells
aDC
ACRBP ANKRD45 ATAD2 BAGE2 CABYR CASC5 CCDC110 CCDC36 CEP290 CEP55 CTAGE5 CTCFL DCAF12 DKKL1 FBXO39 GPAT2 GPATCH2 HSPB9 IGSF11 IL13RA2 KIAA0100 LY6K NR6A1 ODF2 OIP5 OTOA PBK PIWIL2 POTEE PRAME RGS22 RQCD1 SPA17 SPAG1 SPAG17 SPAG4 SPAG8 SPAG9 SPEF2 TAF7L TDRD6 TEKT5 TEX14 TMEFF1 TMEM108 TSGA10 TSSK6 TTK ZNF165
Coef
0.6
0.3
0.0
−0.3
−0.6
B
CD8+ naive T−cells (Coef)
0.2
0.1
0.0
−0.1
−0.2
CEP55 CTCFL DKKL1 LY6K OTOA PBK PRAME RGS22 SPAG1 SPAG17 SPAG4 SPEF2 TAF7L TMEM108 TSGA10 TTK ZNF165
C
aDC (Coef)
0.4
0.2
0.0
−0.2
ACRBP ANKRD45 BAGE2 CABYR CCDC110 CCDC36 CEP290 CEP55 CTAGE5 CTCFL DCAF12 FBXO39 GPAT2 GPATCH2 KIAA0100 LY6K NR6A1 OIP5 OTOA PRAME RGS22 SPA17 SPAG1 SPAG17 SPAG4 SPAG8 SPEF2 TAF7L TDRD6 TMEFF1 TMEM108 TSGA10 TTK
D
TDRD6 (Coef)
0.4
0.2
0.0
aDC B−cells CD4+ memory T−cells CD4+ naive T−cells CD8+ T−cells CD8+ Tcm cDC DC iDC Macrophages M1 Memory B−cells Monocytes naive B−cells Th1 cells Th2 cells
E
TTK (Coef)
0.6
0.3
0.0
−0.3
aDC B−cells Basophils CD4+ memory T−cells CD4+ Tcm CD4+ Tem CD8+ naive T−cells CD8+ T−cells DC Macrophages Macrophages M1 Macrophages M2 Mast cells Monocytes naive B−cells NKT pDC Plasma cells Tgd cells Th1 cells Th2 cells
P value
8e−04
6e−04
4e−04
2e−04
0e+00

Figure 4

Involvement of cellular heterogeneity in clinic-pathology of breast cancer.

A. A number of cell types were significantly correlated with clinical parameters. B. to O. Examples of significant correlations between different cell types and clinical parameters, y-axis represents the fractions of each cell type. *, P<0.05. **, P<0.01. ***, P<0.001.

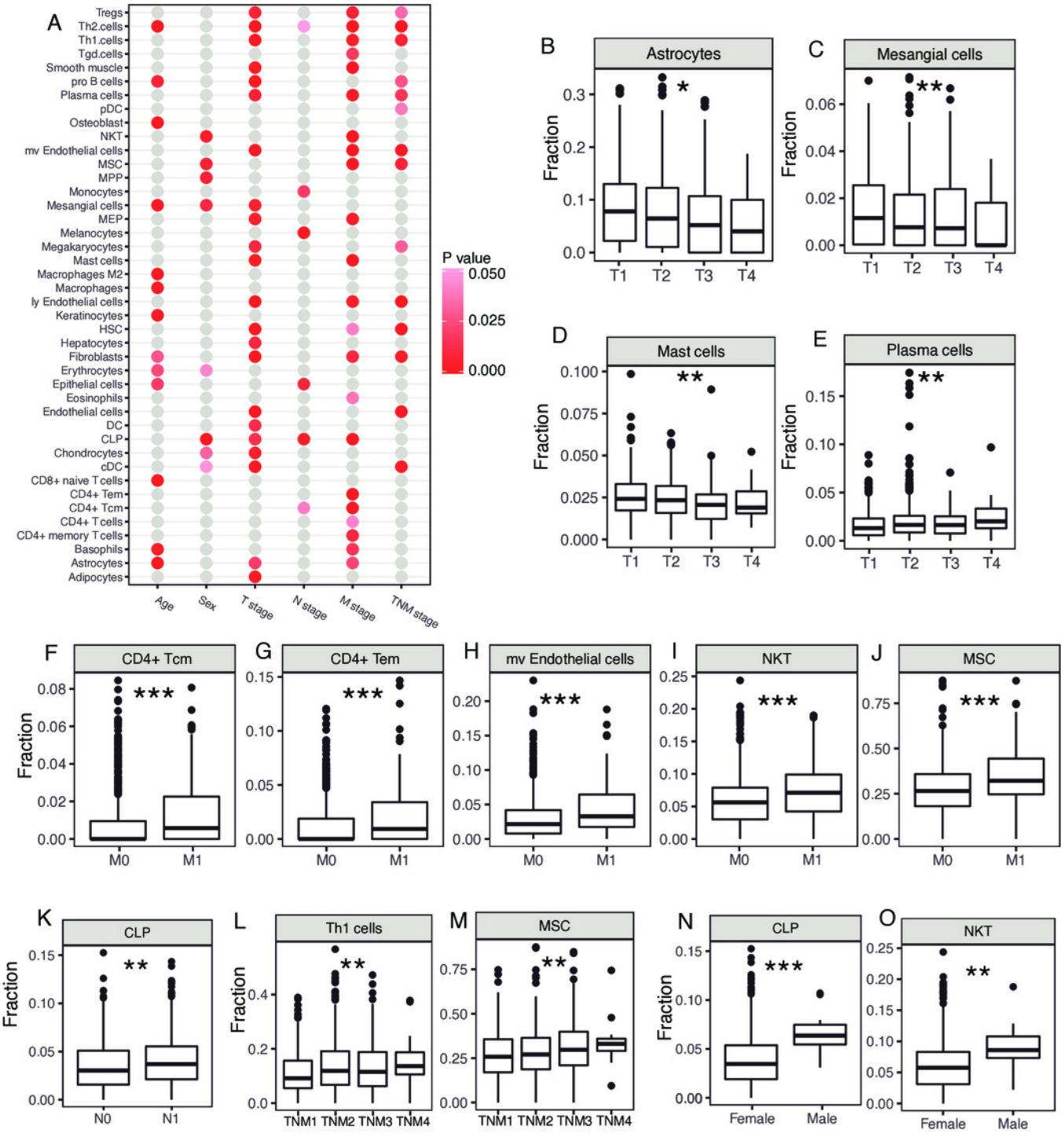
A
Tregs
Th2.cells
Th1.cells
Tgd.cells
Smooth muscle
pro B cells
Plasma cells
pDC
Osteoblast
NKT
mv Endothelial cells
MSC
MPP
Monocytes
Mesangial cells
MEP
Melanocytes
Megakaryocytes
Mast cells
Macrophages M2
Macrophages
ly Endothelial cells
Keratinocytes
HSC
Hepatocytes
Fibroblasts
Erythrocytes
Epithelial cells
Eosinophils
Endothelial cells
DC
CLP
Chondrocytes
cDC
CD8+ naive T cells
CD4+ Tem
CD4+ Tcm
CD4+ T cells
CD4+ memory T cells
Basophils
Astrocytes
Adipocytes
Age
Sex
T stage
N stage
M stage
TNM stage
P value
0.050
0.025
0.000
B
Astrocytes
C
Mesangial cells
D
Mast cells
E
Plasma cells
Fraction
T1
T2
T3
T4
F
CD4+ Tcm
G
CD4+ Tem
H
mv Endothelial cells
I
NKT
J
MSC
M0
M1
K
CLP
L
Th1 cells
M
MSC
N
CLP
O
NKT
N0
N1
TNM1
TNM2
TNM3
TNM4
Female
Male

Figure 5

Survival associated tumor-infiltrating cells in breast cancer.

Forest plot of hazard ratios of survival associated cell types. B. to M. Kaplan-Meier curves of survival associated cell types. Red lines indicated high fraction while blue lines indicated low fraction of each cell types, respectively. N. Kaplan-Meier curves of predictor built with five independent prognostic factors ($P<0.01$). O. ROC curves of prognostic model and the five independent prognostic factors.

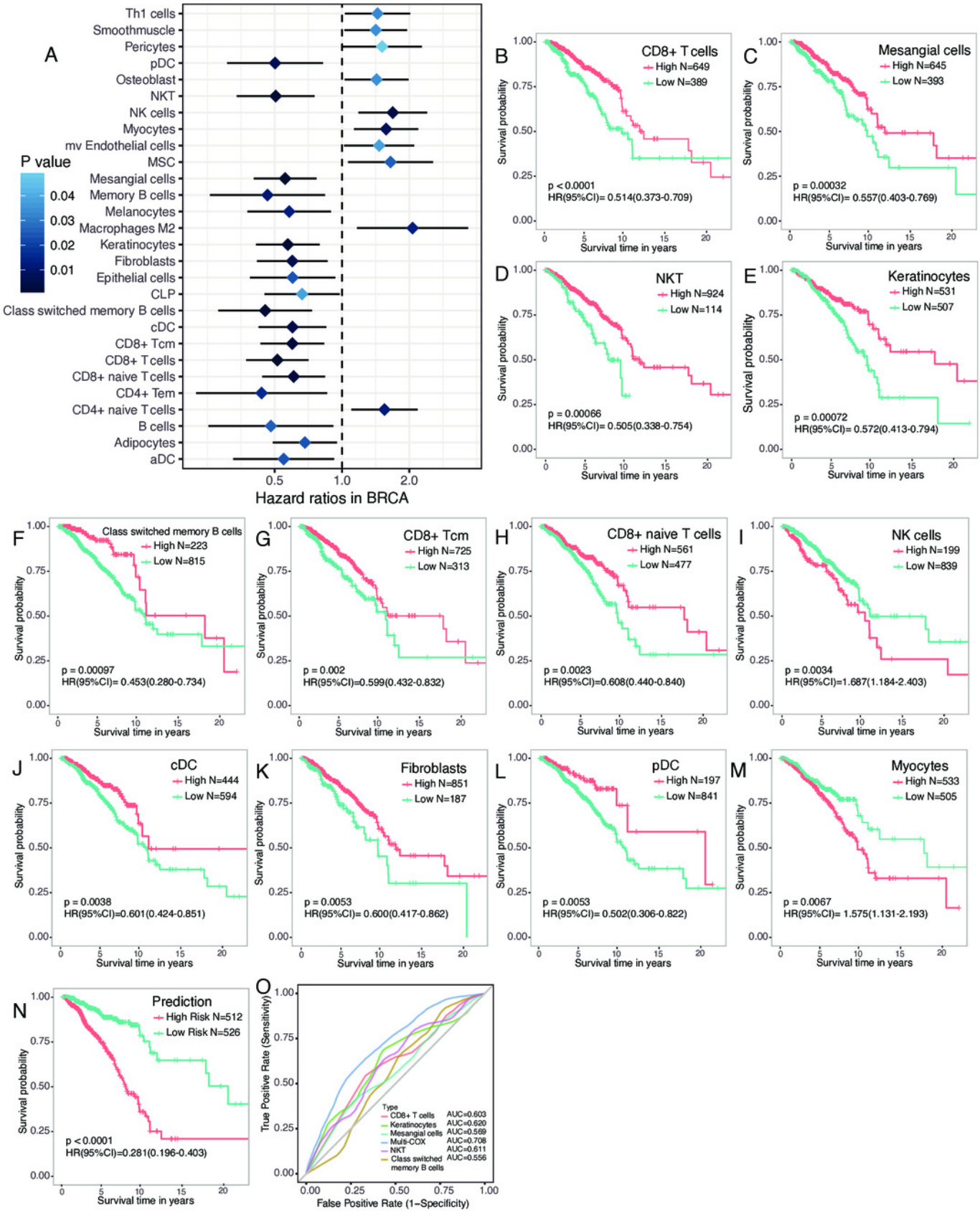
A
Th1 cells
Smoothmuscle
Pericytes
pDC
Osteoblast
NKT
NK cells
Myocytes
mv Endothelial cells
MSC
Mesangial cells
Memory B cells
Melanocytes
Macrophages M2
Keratinocytes
Fibroblasts
Epithelial cells
CLP
Class switched memory B cells
cDC
CD8+ Tcm
CD8+ T cells
CD8+ naive T cells
CD4+ Tem
CD4+ naive T cells
B cells
Adipocytes
aDC
P value
0.04
0.03
0.02
0.01
0.5
1.0
2.0
Hazard ratios in BRCA
B
CD8+ T cells
High N=649
Low N=389
p < 0.0001
HR(95%CI)= 0.514(0.373-0.709)
C
Mesangial cells
High N=645
Low N=393
p = 0.00032
HR(95%CI)= 0.557(0.403-0.769)
D
NKT
High N=924
Low N=114
p = 0.00066
HR(95%CI)= 0.505(0.338-0.754)
E
Keratinocytes
High N=531
Low N=507
p = 0.00072
HR(95%CI)= 0.572(0.413-0.794)
F
Class switched memory B cells
High N=223
Low N=815
p = 0.00097
HR(95%CI)= 0.453(0.280-0.734)
G
CD8+ Tcm
High N=725
Low N=313
p = 0.002
HR(95%CI)=0.599(0.432-0.832)
H
CD8+ naive T cells
High N=561
Low N=477
p = 0.0023
HR(95%CI)=0.608(0.440-0.840)
I
NK cells
High N=199
Low N=839
p = 0.0034
HR(95%CI)=1.687(1.184-2.403)
J
cDC
High N=444
Low N=594
p = 0.0038
HR(95%CI)=0.601(0.424-0.851)
K
Fibroblasts
High N=851
Low N=187
p = 0.0053
HR(95%CI)= 0.600(0.417-0.862)
L
pDC
High N=197
Low N=841
p = 0.0053
HR(95%CI)= 0.502(0.306-0.822)
M
Myocytes
High N=533
Low N=505
p = 0.0067
HR(95%CI)= 1.575(1.131-2.193)
N
Prediction
High Risk N=512
Low Risk N=526
p < 0.0001
HR(95%CI)=0.281(0.196-0.403)
O
True Positive Rate (Sensitivity)
False Positive Rate (1−Specificity)
Type
CD8+ T cells AUC=0.603
Keratinocytes AUC=0.620
Mesangial cells AUC=0.569
Multi-COX AUC=0.708
NKT AUC=0.611
Class switched memory B cells AUC=0.556
Survival probability
Survival time in years

Figure 6

Differences of cellular heterogeneity among different subtypes of breast cancer.

A. Median fraction of 64 cell types in five subtypes of breast tumor. B. Cluster analysis by ImmuneScore and StromaScore, which were calculated by summing up the fractions of immune and stromal cells, respectively. C. Dimensionality reduction and visualization by t-Distributed Stochastic Neighbor Embedding (t-SNE) clustering with fractions of all the 64 cell types. D. Heatmap of fractions of 64 cell types in five subtypes of breast cancer. E. to P. Box plots with dots of fractions of certain cell types in five subtypes of breast cancer. *, $P<0.05$. **, $P<0.01$. ***, $P<0.001$, ****, $P<0.0001$.

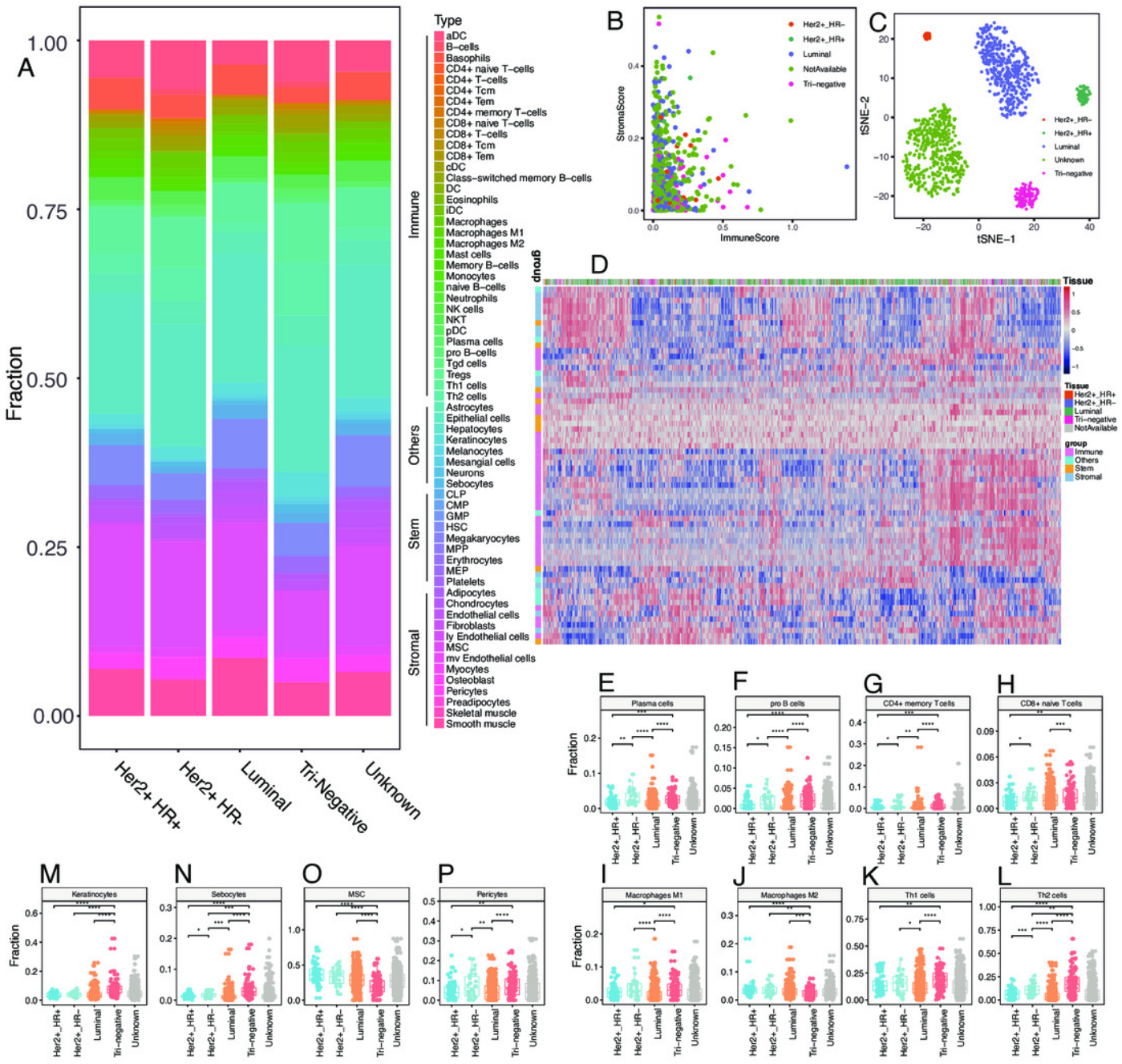
A
Fraction
1.00
0.75
0.50
0.25
0.00
Her2+ HR+
Her2+ HR-
Luminal
Tri-Negative
Unknown
Type
aDC
B-cells
Basophils
CD4+ naive T-cells
CD4+ T-cells
CD4+ Tcm
CD4+ Tem
CD4+ memory T-cells
CD8+ naive T-cells
CD8+ T-cells
CD8+ Tcm
CD8+ Tem
cDC
Class-switched memory B-cells
DC
Eosinophils
iDC
Macrophages
Macrophages M1
Macrophages M2
Mast cells
Memory B-cells
Monocytes
naive B-cells
Neutrophils
NK cells
NKT
pDC
Plasma cells
pro B-cells
Tgd cells
Tregs
Th1 cells
Th2 cells
Astrocytes
Epithelial cells
Hepatocytes
Keratinocytes
Melanocytes
Mesangial cells
Neurons
Sebocytes
CLP
CMP
GMP
HSC
Megakaryocytes
MPP
Erythrocytes
MEP
Platelets
Adipocytes
Chondrocytes
Endothelial cells
Fibroblasts
ly Endothelial cells
MSC
mv Endothelial cells
Myocytes
Osteoblast
Pericytes
Preadipocytes
Skeletal muscle
Smooth muscle
Immune
Others
Stem
Stromal
B
StromaScore
ImmuneScore
Her2+_HR-
Her2+_HR+
Luminal
NotAvailable
Tri-negative
C
tSNE-2
tSNE-1
Her2+_HR-
Her2+_HR+
Luminal
Unknown
Tri-negative
D
group
Tissue
Tissue
Her2+_HR+
Her2+_HR-
Luminal
Tri-negative
NotAvailable
group
Immune
Others
Stem
Stromal
E
Plasma cells
F
pro B cells
G
CD4+ memory T cells
H
CD8+ naive T cells
I
Macrophages M1
J
Macrophages M2
K
Th1 cells
L
Th2 cells
M
Keratinocytes
N
Sebocytes
O
MSC
P
Pericytes
Fraction
Her2+_HR+
Her2+_HR-
Luminal
Tri-negative
Unknown

Figure 7

Survival associated tumor-infiltrating cells in five subtypes of breast cancer.

A. to E. Forest plots of hazard ratios of survival associated cell types in five subtypes of breast cancer.

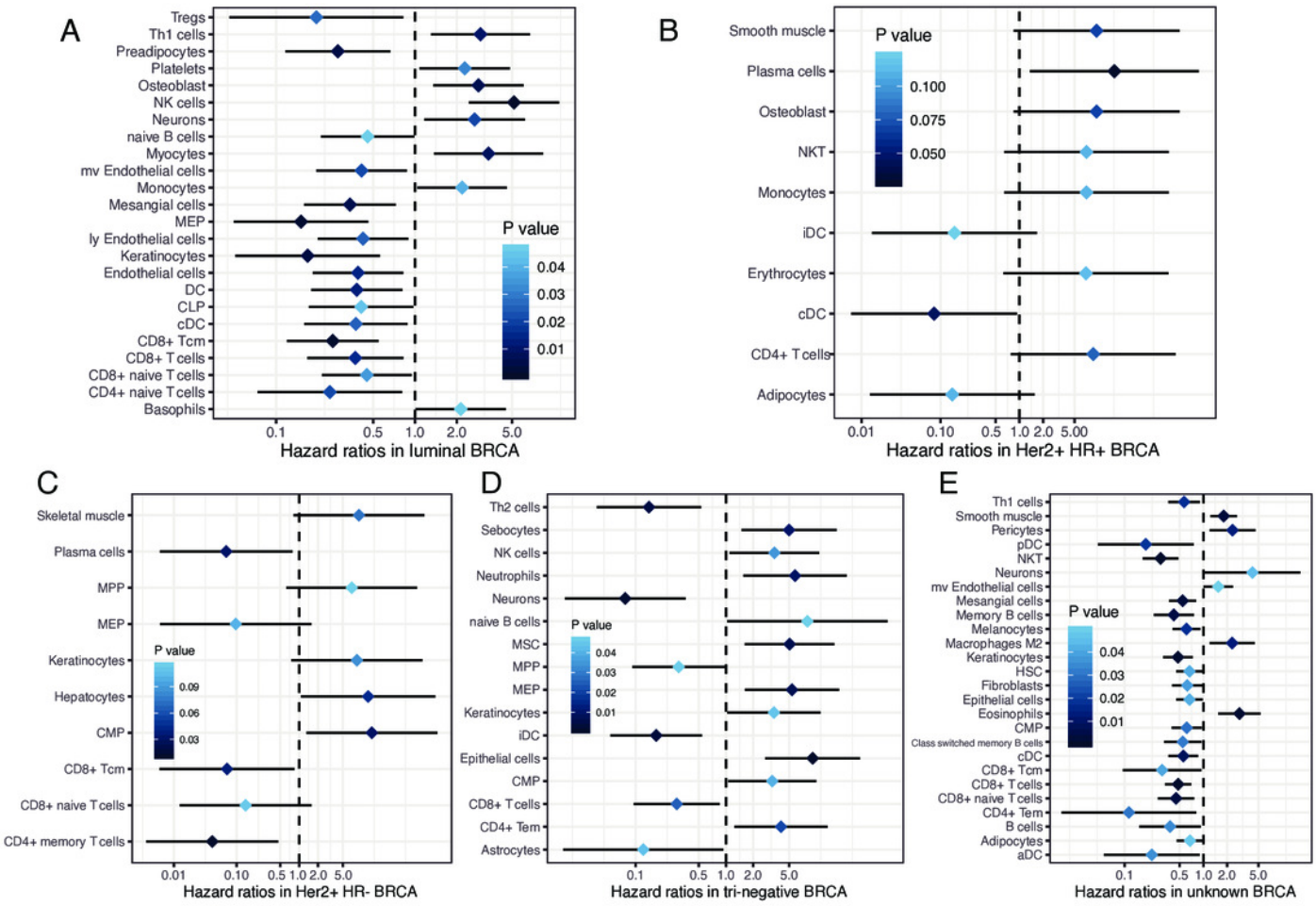

Table 1(on next page)

Clinical characteristics of the 1,092 breast cancer patients from TCGA.

Clinical characteristics of the 1,092 breast cancer patients from TCGA.

Table 1. Clinical characteristics of the 1,092 breast cancer patients from TCGA.

Characteristic	TCGA BRCA (N=1,092)
Age median (range)	59 (26-90)
Sex	
Female	1,080 (98.9%)
Male	12 (1.1%)
Tumor stage	
T1	279 (25.5%)
T2	635 (58.2%)
T3	138 (12.6%)
T4	40 (3.7%)
Lymph node stage	
N (-)	333 (30.5%)
N (+)	759 (69.5%)
Metastasis stage	
M (-)	903(82.7%)
M (+)	189(17.3%)
TNM stage	
TNM Ⅰ	181 (16.5%)
TNM Ⅱ	637 (58.3%)
TNM Ⅲ	254 (23.3%)
TNM Ⅳ	20 (18.3%)
Subtypes	
Luminal	426 (39.0%)
Her2+_HR+	59 (5.4%)
Her2+_HR-	30 (2.7%)
Tri-negative	97 (8.9%)
Unknown	480(44.0%)
Status	
Living	940 (86.1%)
Decreased	152 (13.9%)

Table 2(on next page)

Abbreviations and statistical summary of enriched fractions of the 64 cell types.

Abbreviations and statistical summary of enriched fractions of the 64 cell types.

Table 2. Abbreviations and statistical summary of enriched fractions of the 64 cell types.

Full name	Abbreviations	Group	Tumor N=1,092 (Mean ± SD)	Normal N=112 (Mean ± SD)	Paired Tumor N=112 (Mean ± SD)	P value	P value (paired)
Activated dendritic cells	aDC	Immune	0.116±0.110	0.016±0.045	0.103±0.104	<0.001	<0.001
B-cells	/	Immune	0.031±0.085	0.004±0.039	0.021±0.052	<0.001	<0.001
Basophils	/	Immune	0.078±0.056	0.058±0.040	0.071±0.054	<0.001	0.077
CD4+ naive T-cells	/	Immune	0.010±0.029	0.003±0.015	0.008±0.020	0.009	0.004
CD4+ T-cells	/	Immune	0.003±0.010	0.000±0.003	0.001±0.004	0.101	0.036
Central memory CD4+ T Cell	CD4+ Tcm	Immune	0.009±0.015	0.016±0.019	0.003±0.007	0.006	<0.001
Effector memory CD4+ T cell	CD4+ Tem	Immune	0.013±0.022	0.000±0.002	0.010±0.018	<0.001	<0.001
CD4+ memory T-cells		Immune	0.007±0.017	0.000±0.004	0.007±0.012	<0.001	<0.001
CD8+ naive T-cells	/	Immune	0.012±0.011	0.005±0.007	0.011±0.010	<0.001	<0.001
CD8+ T-cells	/	Immune	0.018±0.039	0.006±0.016	0.016±0.032	<0.001	0.001
Central memory CD8+ T Cell	CD8+ Tcm	Immune	0.026±0.049	0.005±0.015	0.027±0.046	<0.001	<0.001
Effector memory CD8+ T cell	CD8+ Tem	Immune	0.001±0.007	0.000±0.000	0.000±0.002	0.805	0.608
Conventional dendritic cells	cDC	Immune	0.037±0.046	0.072±0.057	0.040±0.043	<0.001	<0.001
Class-switched memory B-cells	/	Immune	0.026±0.031	0.002±0.014	0.020±0.023	<0.001	<0.001
Dendritic cells	DC	Immune	0.005±0.011	0.005±0.011	0.004±0.008	0.826	0.927
Eosinophils	/	Immune	0.000±0.001	0.000±0.000	0.000±0.001	0.737	0.763
Immature dendritic cells	iDC	Immune	0.042±0.073	0.167±0.171	0.047±0.053	<0.001	<0.001
Macrophages	/	Immune	0.038±0.036	0.008±0.023	0.038±0.034	<0.001	<0.001
Inflammatory (M1) macrophages	Macrophages M1	Immune	0.022±0.027	0.003±0.010	0.019±0.027	<0.001	<0.001
Reparative (M2) macrophages	Macrophages M2	Immune	0.031±0.021	0.030±0.034	0.031±0.017	0.001	0.003
Mast cells	/	Immune	0.024±0.012	0.015±0.008	0.025±0.010	<0.001	<0.001
Memory B-cells	/	Immune	0.006±0.028	0.001±0.011	0.003±0.014	<0.001	0.083
Monocytes	/	Immune	0.004±0.012	0.003±0.013	0.004±0.011	0.007	0.040

naive B-cells	/	Immune	0.004±0.016	0.001±0.009	0.003±0.009	<0.001	0.004
Neutrophils	/	Immune	0.000±0.001	0.003±0.005	0.000±0.000	<0.001	<0.001
Nature killer cells	NK cells	Immune	0.000±0.002	0.000±0.000	0.000±0.001	0.342	0.759
Natural killer T cells	NKT	Immune	0.061±0.039	0.028±0.036	0.043±0.028	<0.001	<0.001
Plasmacytoid dendritic cells	pDC	Immune	0.008±0.019	0.000±0.000	0.007±0.017	<0.001	0.004
Plasma cells	/	Immune	0.019±0.018	0.001±0.003	0.015±0.012	<0.001	<0.001
pro B-cells	/	Immune	0.009±0.017	0.000±0.000	0.006±0.013	<0.001	<0.001
Gamma delta T cells	Tgd cells	Immune	0.003±0.009	0.000±0.000	0.004±0.011	<0.001	0.003
Regulatory T cells	Tregs	Immune	0.012±0.016	0.004±0.009	0.013±0.016	<0.001	<0.001
Type 1 T helper (Th1) cells	Th1 cells	Immune	0.133±0.092	0.014±0.026	0.101±0.067	<0.001	<0.001
Type 2 T helper (Th2) cells	Th2 cells	Immune	0.075±0.096	0.002±0.007	0.082±0.091	<0.001	<0.001
Astrocytes	/	Others	0.076±0.069	0.110±0.065	0.081±0.064	<0.001	<0.001
Epithelial cells	/	Others	0.365±0.092	0.277±0.146	0.354±0.089	<0.001	<0.001
Hepatocytes	/	Others	0.001±0.002	0.005±0.003	0.002±0.002	<0.001	<0.001
Keratinocytes	/	Others	0.047±0.038	0.065±0.037	0.046±0.038	<0.001	<0.001
Melanocytes	/	Others	0.010±0.008	0.008±0.008	0.010±0.009	0.011	0.162
Mesangial cells	/	Others	0.014±0.015	0.034±0.013	0.015±0.015	<0.001	<0.001
Neurons	/	Others	0.004±0.008	0.004±0.002	0.004±0.006	<0.001	<0.001
Sebocytes	/	Others	0.016±0.019	0.009±0.009	0.017±0.024	<0.001	<0.001
Common lymphoid progenitor	CLP	Stem	0.039±0.026	0.013±0.015	0.041±0.026	<0.001	<0.001
Common myeloid progenitor	CMP	Stem	0.001±0.003	0.003±0.004	0.001±0.003	<0.001	0.008
Granulocyte-macrophage progenitor	GMP	Stem	0.002±0.006	0.002±0.006	0.002±0.008	0.001	0.025
Hematopoietic stem cells	HSC	Stem	0.150±0.110	0.492±0.191	0.174±0.125	<0.001	<0.001
Megakaryocytes	/	Stem	0.004±0.004	0.021±0.009	0.005±0.005	<0.001	<0.001
Multipotent progenitors	MPP	Stem	0.00±0.001	0.000±0.000	0.000±0.000	<0.001	0.027
Erythrocytes	/	Stem	0.000±0.000	0.000±0.000	0.000±0.000	0.471	0.306
Megakaryocyte-	MEP	Stem	0.035±0.030	0.011±0.017	0.027±0.022	<0.001	<0.001

erythroid progenitor							
Platelets	/	Stem	0.000±0.002	0.000±0.001	0.000±0.001	0.062	0.334
Adipocytes	/	Stromal	0.050±0.088	0.382±0.207	0.063±0.102	<0.001	<0.001
Chondrocytes	/	Stromal	0.035±0.037	0.053±0.022	0.040±0.038	<0.001	<0.001
Endothelial cells	/	Stromal	0.058±0.056	0.209±0.104	0.066±0.059	<0.001	<0.001
Fibroblasts	/	Stromal	0.058±0.068	0.172±0.078	0.063±0.074	<0.001	<0.001
Lymphatic endothelial cells	ly Endothelial cells	Stromal	0.020±0.027	0.110±0.071	0.021±0.025	<0.001	<0.001
Mesenchymal stem cells	MSC	Stromal	0.281±0.144	0.019±0.054	0.252±0.139	<0.001	<0.001
Microvascular endothelial cells	mv Endothelial cells	Stromal	0.032±0.032	0.104±0.062	0.029±0.028	<0.001	<0.001
Myocytes	/	Stromal	0.004±0.010	0.009±0.042	0.005±0.008	0.012	0.387
Osteoblast	/	Stromal	0.025±0.031	0.005±0.013	0.018±0.024	<0.001	<0.001
Pericytes	/	Stromal	0.053±0.052	0.027±0.035	0.057±0.055	<0.001	<0.001
Preadipocytes	/	Stromal	0.024±0.040	0.179±0.090	0.032±0.045	<0.001	<0.001
Skeletal muscle	/	Stromal	0.001±0.011	0.010±0.101	0.001±0.003	0.442	0.365
Smooth muscle	/	Stromal	0.133±0.091	0.125±0.067	0.175±0.088	0.345	<0.001
ImmuneScore	/	/	0.082±0.103	0.029±0.049	0.074±0.079	<0.001	<0.001
StromaScore	/	/	0.083±0.088	0.382±0.174	0.096±0.097	<0.001	<0.001

2

3