MetaTiME: Meta-components of the Tumor Immune Microenvironment

Yi Zhang[1,2], Guanjue Xiang[1,2], Alva Yijia Jiang[1], Allen Lynch[1,2], Zexian Zeng[1,2], Chenfei Wang[1,2], Wubing Zhang[1,2], Jingyu Fan[1,2], Jiajinlong Kang[1], Shengqing Stan Gu[4], Changxin Wan[1,2], Boning Zhang[1,2], X. Shirley Liu[1,2,3]*, Myles Brown[3,4*], Clifford A Meyer[1,2,3]*

1. Department of Data Science, Dana-Farber Cancer Institute, Boston, MA 02215, USA

2. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02215 USA.

3. Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA

4. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

*Corresponding author. Email: cliff_meyer@ds.dfci.harvard.edu

*Co-corresponding author. Email: myles_brown@dfci.harvard.edu

*Co-corresponding author. Email: xsliu.res@gmail.com

Abstract

Recent advances in single-cell RNA sequencing have revealed heterogeneous cell types and gene expression states in the non-cancerous cells in tumors. The integration of multiple scRNA-seq datasets across tumors can reveal common cell types and states in the tumor microenvironment (TME). We developed a data driven framework, MetaTiME, to overcome the limitations in resolution and consistency that result from manual labelling using known gene markers. Using millions of TME single cells, MetaTiME learns meta-components that encode independent components of gene expression observed across cancer types. The meta-components are biologically interpretable as cell types, cell states, and signaling activities. By projecting onto the MetaTiME space, we provide a tool to annotate cell states and signature continuums for TME scRNA-seq data. Leveraging epigenetics data, MetaTiME reveals critical transcriptional regulators for the cell states. Overall, MetaTiME learns data-driven meta-components that depict cellular states and gene regulators for tumor immunity and cancer immunotherapy.

Keywords

Tumor microenvironment, cancer immunity, single-cell RNA sequencing, transcriptional regulators, gene expression cell state, cell state representation.

Introduction

Recent advances in cancer research have revealed the integral role of the tumor microenvironment (TME) in tumor progression and therapy responses[1–6]. Understanding interactions between cancer cells and the non-cancer compartments, including immune cells, fibroblasts, and endothelial cells, has revealed potential targets for cancer immunotherapy. Specifically, single-cell RNA-sequencing (scRNA-Seq) applied on multiple patient tumors has enabled the high-resolution identification of TME constituents that interfere with the elimination of cancer cells. For example, exhausted tumor-infiltrating lymphocytes (TILs)[4,7,8], and certain tumor-associated macrophages subtypes[9–11], have been associated with tumor development. However, the definition of cell types and cell states in tumor scRNA analyses still relies on manual labeling by experts using known exclusive biomarkers following unsupervised clustering[12,13], which lacks consistency and varies between different cohorts.

As single-cell data accumulate, integrating a large collection of cells from multiple cohorts can help unify the definition of cell types and states to facilitate the automatic annotation of new scRNA-seq data[14,15]. One approach to cell annotation is to use predefined biomarker lists. However, these biomarkers might not cover domain-specific cellular states, for example, well-defined immune cell markers derived from blood immune cells may not fully cover the TME disease context[14]. Moreover, although cell type definitions in reference databases such as CIBERSORT, Azimuth, and Human Primary Cell Atlas[16–18] can be useful, the granularity of

these definitions varies between databases. Several efforts integrating pan-cancer scRNA data have revealed subtypes in the TME through the manual annotation of clusters using a shortlist of exclusive gene markers[13,19,20].

Another approach is to map a dataset containing unannotated cell states onto an annotated reference. Methods to obtain such representations include canonical correlation analysis (CCA)[21], adjusted principal components (Harmony)[22], or generative deep learning models using variational autoencoders (scVI)[23]. These methods use dimension reduction onto a common latent space to align cells with similar states between datasets, without ascribing meaning to the latent space representations. An alternative data driven approach is to identify low dimensional latent space representations in which a biological meaning can be ascribed to each latent dimension. Several matrix factorization algorithms have been developed to represent high dimensional data in a low-dimensional space with interpretable components, including non-negative matrix factorization (NMF)[24] and independent component analysis (ICA)[25,26].

In this study, we developed a computational framework for mapping millions of single cells from multiple cohorts onto a comprehensive and interpretable latent space, learnt from the data. The framework, MetaTiME (Meta-components of the Tumor immune MicroEnvironment), identifies reproducible low-dimensional meta-components that reflect independent components of gene expression variation across cohorts and cancer types. MetaTiME adopted ICA for dimensional reduction to maximize the mutual independence among components. We used MetaTiME to obtain meta-components (MeCs) from 1.7 million single cells across 79 tumor datasets. These MeCs represent the TME landscape along 86 data-driven transcriptional directions mirroring lineage-specific cell states and signaling activities. Furthermore, we developed a MetaTiME toolkit for using the MeCs to annotate cellular states and signature continuums in tumor scRNA datasets, and to reveal differential signatures across immunotherapy responses. Finally, by

incorporating transcription factor binding data, MetaTiME revealed and prioritized putative transcriptional regulators that may modulate tumor immunity.

Results

MetaTiME as a general framework to discover consensus transcriptomic programs

The MetaTiME framework consists of three stages: meta-component (MeC) discovery, interpretation of MeCs, and application of cell state annotations (**Fig.1a**). The MeC discovery stage detects repeatable sources of variation from multiple single-cell measurements sharing similar cellular properties. The MeC interpretation step involves a one-time curation effort using biomarker databases, pathway information and Cistrome DB chromatin profiling data[27]. In the third step, users map MeCs onto their new tumor scRNA-seq datasets using MetaTiME application tools, to obtain annotated cell states and signature continuums.

To train MeCs for the TME context, we collected and curated 2,157,387 cells from 76 studies ranging across 27 cancer types, using publicly available tumor scRNA-Seq data mostly from TISCH[28]. After removing the TISCH annotated malignant cells using MAESTRO[17], 102,703 stromal cells and 1,617,110 immune cells were retained for downstream training (**Fig.S1**). The 76 studies were further partitioned according to cancer type, resulting in 93 datasets, including 7 datasets with immune checkpoint blockade (ICB) treatment and 3 10x Genomics provided datasets representing peripheral blood mononuclear cells (PBMC) sampled from healthy donors.

In the MeC discovery stage, MetaTiME first decomposes the log-transformed expression matrix of each single dataset using Independent Component Analysis (ICA)[25]. We adopted ICA to maximize mutual independence among gene expression components. In simulations ICA performed slightly better than Non-negative Matrix Factorization (NMF) in simulated single-cell

105 data with pre-embedded transcriptional signatures (**Fig.S2**). The feature weight distribution of
106 each Independent Component (IC) also enables normalization of the gene contribution scores
107 for measuring similarity among components. MetaTiME then applies two transformations to the
108 IC vectors, z-weight normalization and skewness alignment, to ensure the scales of gene
109 representation scores are comparable among components (**Methods, Fig.S3a**). Next,
110 MetaTiME filters ICs to retain ones that are reproducible across multiple cohorts (the minimum
111 Pearson correlation with any other IC >=0.3). These are passed to the Louvain graph clustering
112 algorithm to merge IC groups into MeCs (**Methods**). Lastly, MetaTiME computes averaged
113 profiles of gene z-weights within each IC cluster, yielding 86 MeCs trained for the TME (**Fig.1a,**
114 **Fig.2a**). The number of MeCs was automatically determined by simultaneously optimizing
115 granularity and independence in IC clustering (**Fig.S3b**). Importantly, the MeC clustering does
116 not depend on cohort source (**Fig.S4a,** cohort source of component), and is robust during leave-
117 one-out testing in which the MeCs are learnt with a single cohort left out of the analysis
118 (**Fig.S5**). The leave-one-out MeCs are highly correlated with the MeCs learnt on the complete
119 data (mean maximum correlation=0.988), and the numbers of components are similar (mean
120 leave-one-out MeC number=84.64, full-set MeC number=86). This integration after
121 decomposition approach overcomes batch effects, which are often a challenge in single cell
122 RNA-seq data analysis. Moreover, we implemented the standard approach in which cluster-wise
123 signatures are obtained by batch effect removal, clustering, and differential expression analysis
124 (**Fig.S6a**). Limited by server computing memory, the maximum number of datasets we could
125 integrate in this way was 21, and the cluster-wise signatures display a lower level of specificity
126 in the test data (**Fig.S6b,c**). Therefore, the MetaTiME approach allows for the effective
127 integration of large numbers of single cell datasets, which will become increasingly important as
128 data accumulates.
129
130 **MetaTiME defines interpretable meta-components**

131 In principle, each MeC represents one independent source of transcriptional variation commonly
132 present in the TME. We investigated top ranked genes in MeCs and found MeCs are highly
133 interpretable, reflecting common biological processes in the TME. For instance, the MeC
134 derived from the largest IC cluster is highly enriched in interferon response genes, such as
135 *ISG15*, *IFI6*, *LY6E*, and *MX1*, indicating that the underlying interferon response is among the
136 most common source of transcriptional variation shared across tumor samples and cohorts
137 (**Fig.2a, b**). Intriguingly, top genes of each MeC are enriched in known biomarkers or regulators.
138 For example, several T cell-related MeCs identify different gene modules co-expressed in T
139 cells reflecting activation of different T-cell related processes (**Fig.2c, Fig.S7**). The "T cell co-
140 signaling" MeC-65 features T cell receptors in co-stimulatory and co-inhibitory pathways[29,30],
141 such as *TNFRSF4* (OX40), *TNFRSF18* (GITR) *TNFRSF9* (4-1BB), and *ICOS* (**Fig.2c**, left). The
142 "CXCL13+ exhausted CD8 T cell" MeC-40 features receptors characterizing the exhausted CD8
143 T cell state[8], including *HAVCR2* (TIM3), *LAG3*, *TIGIT*, and *PDCD1* (PD1) (**Fig.S7**), each being
144 potential ICB targets[31]. In addition, this MeC is characterized by a high level of *CXCL13*
145 (**Fig.2c**, second panel), a cytokine mediating immune cell trafficking to tertiary lymphoid
146 structures[32]. In contrast, a related MeC representing T cell co-signaling receptors in regulatory
147 CD4T cells (Treg) has a different ranking, including *TNFRSF18*, *TNFRSF4*, *TIGIT, TNFRSF1B*,
148 *CTLA4*, *CD27* among the top 20 genes, along with the regulatory T cell-specific marker *FOXP3*
149 (**Fig.2c**, right). Though ICB has been an extremely successful therapy for some patients, it has
150 not yet had an impact on the majority of patients[33]. Investigating the top members in the MeCs
151 involving T cell receptor pathways may help identify new ICB targets.
152

153 **MetaTiME depicts the functional landscape of transcriptomic variation and cell states in**
154 **the tumor microenvironment**

155 We provided functional annotations of all MeCs by examining top z-weight genes and compared
156 these with functional gene sets, such as immune cell type markers[15,18] and gene ontology

157 databases[34]. We found that 86 MeCs clearly mirror gene expression patterns corresponding to
158 cell types, cell states and signaling pathway activities, depicting a landscape of non-cancer cell
159 states in the TME (**Fig.2, Table S1**: MeC annotation). The top genes of the cell type MeCs
160 match well-known lineage-specific markers[15,18]. Examples include *CD74*, *CD79A*, *MS4A1* for B
161 cells ("B cell" MeC-18), *CD3D*, *CD8A*, *CD8B* for T cells ("CD3 - CD8 T cell" MeC-32), and *LYZ*,
162 *VCAN*, *S100A9* for CD14+ Monocytes ("CD14 monocyte" MeC-17) (**Fig.2b, Fig.2d**). The
163 majority of MeCs define high resolution lineage-specific cell states (**Fig.2b, Fig.S4b**). Taking the
164 B cell lineage as an example, multiple MeCs harbor genes specific to B cell developmental
165 stages[35], ranging from a progenitor B cell state (*CD69* and *PAX5* in the "PAX5 B cell" in MeC-
166 50), to a mature B cell state (*CD79A* in the "B cell" MeC-18), an antibody-secreting plasma cell
167 state (*XBP1* in the "B_Plasma" MeC-4, and *JCHAIN* in the "JCHAIN+ plasma B cell" MeC-77),
168 and immunoglobulin secretion states (IGK and IGH in the "immunoglobin kappa B cell" MeC-30,
169 IGL and IGH in the "immunoglobin lambda B cell" MeC-50) (**Fig.2b**). Lastly, like the interferon
170 responsive MeC mentioned above, we found a subset of MeCs that are more accurately
171 interpreted as signaling pathways because their top genes are more related to pathways or
172 molecular functions than to cell identities.
173
174 We organized the 86 annotated MeCs into six cell lineage-focused categories and one signaling
175 pathway-focused category (**Fig.2a,b** and **Table S1**: MeC annotation, MeC enrichment). Among
176 these, there are 6 B lineage-related MeCs for B cells; 20 T cell lineage MeCs covering CD8 T
177 cells, CD4 T cells, and natural killer (NK) cells; 4 dendritic cell (DC) lineage MeCs; 12 monocyte
178 and macrophage-related MeCs; 3 platelet, erythrocyte, and mast cell MeCs; 6 stromal cell-
179 related MeCs for fibroblasts, myofibroblasts and endothelial cells; and 35 MeCs in the signaling
180 category (**Fig.2b**). We demonstrated that the MeCs are of high specificity, visualizing the z-
181 weights of known cell subtype markers and pathway biomarkers (**Fig.2d**). Correlating MeCs
182 with the comprehensive immune cell type database Azimuth[15] validated the lineage-specificity of

183 several MeCs, while most MeCs reflect cell states that appear specific to the tumor context
184 (**Fig.S4b**).
185
186 **MetaTiME annotates cell states and signature continuums when applied to the tumor**
187 **microenvironment single-cell data**
188 As MetaTiME MeCs provides a highly interpretable basis for the TME in single cells, we
189 provided a toolkit to reveal MeC signature continuums and enriched cell states in scRNA-seq
190 TME data (code deposited in https://github.com/yi-zhang/MetaTiME). The MetaTiME annotation
191 toolkit takes as input the scRNA-seq expression matrix after depth normalization and log
192 transformation, maps each single cell onto the pre-trained MeC space, and annotates the most
193 highly enriched cell states for pre-defined cell clusters. The cell clusters are by default
194 calculated using graph clustering with high resolution after an optional batch effect correction
195 with Harmony[22]. We demonstrate the application of MetaTiME on basal cell carcinoma (BCC)
196 single-cells from *Yost et al.*[8]. These test cells were excluded from the MetaTiME training stage.
197 MetaTiME annotates the enriched cell states (**Fig.3a**) highlighting gradients of exhausted CD8 T
198 cells and follicular helper T cells (Tfh) (**Fig.3c**). The most enriched cell states consistently match
199 the manual labelling from the original study with improved resolution (**Fig.3b**) and highly
200 express corresponding markers (**Fig.3d**). In addition, compared to Seurat's[14] automated
201 CIBERSORT marker-based annotations (14 cell types, **Fig.S6a**), MetaTiME provides higher
202 resolution (38 cell states, **Fig.3b**). Other automatic annotation gene panels were also tested,
203 including the human primary cell atlas (HPCA) panel and Blueprint-ENCODE panel used in
204 SingleR[16], where macrophages and plasma cells appear to be mislabeled as subclusters within
205 T cell clusters (**Fig.S8b, c**). Interestingly, the MetaTiME annotation not only indicates the CD8 T
206 cell and CD4 T cell subtypes, but also splits cells further into cell states with polarized
207 expression in proliferation, cytotoxicity, exhaustion level, heat stress, co-signaling pathways,
208 and so on. (**Fig.3a, Fig.S8a**). The B cell group is further partitioned into distinct B cell

209 developmental states including a B cell cluster with cell cycle and MYC activities (**Fig.3a,b,**
210 **Fig.S8**), which possibly represent germinal center (GC) B cells undergoing active expansion
211 and maturation[36].
212
213 We thus re-annotated all tumor scRNA cohorts using MetaTiME and investigated the distribution
214 of cell state compositions across cancer cohorts. As shown in **Fig. 3e** and **Fig.S9**, tumors are
215 highly heterogenous and the TME cellular composition is only partially determined by cancer
216 type. For example, Cholangiocarcinoma (CHOL) is highly enriched in stromal cells including
217 collagen-secreting fibroblast, as expected[37], while other samples including ovarian cancer (OV),
218 pancreatic adenocarcinoma (PAAD), and multiple myeloma (MM) are also stromal-rich.
219 Furthermore, tumors with high infiltration of the "GZMK+ CCL5+ CD8 T cell" state include
220 multiple tumor types including colon cancer, breast cancer, and skin cancer, suggesting that
221 immune infiltration is sample-dependent and that cancer treatments should be personalized[38].
222
223 **Differential MetaTiME analysis detects alterations of transcriptional programs in**
224 **immunotherapy**
225 Single-cell data derived from ICB trials is invaluable for identifying cell types associated with ICB
226 treatment or response[8]. However, the detection of differential cell type abundances in ICB
227 cohorts has been challenging due to the heterogeneity of cell type proportions and to the limited
228 numbers of patients in each cohort[39]. We compared differences in MeC signatures instead of
229 cell count proportions, to understand immune responses during ICB. We analyzed two ICB
230 cohorts, a basal cell carcinoma (BCC) cohort with samples from pre- or post-ICB treatment[8],
231 and a bladder cancer (BLCA) cohort with samples from ICB responders and non-responders[40].
232 We applied MetaTiME for per-cluster cell state annotation and per-cell MeC signature
233 evaluation. For each cell state cluster, we tested all MeC signatures passing significance
234 (average z-weight>2) between conditions using the Wilcoxon rank-sum test. We plotted cluster-

235 wise signatures in the significance – effect size scatterplot to highlight the most significant
236 differential MeCs (**Methods**). In a comparison of pre- and post-ICB treatment, we observed
237 higher expression of cytotoxic T cell and B cell MeCs in the post-ICB samples. Moreover,
238 several monocyte and macrophage states are also suppressed after ICB treatment (**Fig.4a**).
239 Notably, the IL1B-positive macrophage signature is also found to be elevated in non-responders
240 compared to responders in the BLCA ICB cohort (**Fig.4b**). Since activation of the IL1B pathway
241 is a known regulator of inflammatory processes[41], we sought to investigate whether the IL1B-
242 positive macrophage signature is associated with tumor survival prognosis in bulk RNA-seq
243 data from The Cancer Genome Atlas Program (TCGA). We evaluated TCGA tumors using the
244 averaged expression of the top 20 genes from the "Macrophage IL1-NFkB" MeC (MeC-42)
245 which ranks first in elevated MeCs in non-responders (**Fig.4a**). We found that higher expression
246 of the *IL1B* signature is associated with lower survival rate in multiple cancer types, especially in
247 Low Grade Glioma (LGG) and in Kidney renal cell carcinoma (KIRC) (**Fig.S10**). This suggests
248 the macrophage state with *IL1B* pathway activation is associated with poor prognosis and lower
249 ICB efficacy.
250
251 **MetaTiME delineates myeloid cells in different metabolic states**
252 As specific myeloid cell states have been associated with cancer survival and treatment
253 response, we sought to systematically characterize MeCs related to monocytes and
254 macrophages. Although the canonical definition of M1 and M2 macrophages is derived from
255 cytokine polarized macrophages *in vitro*[42], MetaTiME's myeloid-related MeCs represent a more
256 complex framework for understanding tumor-infiltrating macrophages. MetaTiME's 12 monocyte
257 and macrophage related MeCs can be summarized into six central monocyte or macrophage
258 states for the TME, after merging similar states such as "Macrophage IL1-NFkB" and
259 "Macrophage IL1- JUN" due to similarity among top genes (**Fig.4c**). Monocytes are classified as
260 two categories, CD14+ and CD16+. For macrophages, four MeCs define common states of

261 intra-tumor macrophages: C1Q+, SPP1+, lipid-rich, and IL1B+ macrophages, and two MeCs,
262 representing interferon and MHC-II signaling pathways, are less frequently observed among
263 macrophages (**Fig.2b, Fig.4c**). In comparison, previous studies defined different tumor
264 associated macrophages in terms of manually selected representative genes after clustering
265 myeloid cells. For example, *Cheng et al.*[10] defined several tumor associated macrophage types
266 including *ISG15*+, *SPP1*+, *INHBA*+, *VCAN*+, *NLRP3*+, and *FN1*+ macrophages, while *Bi et al.*[43]
267 defined *CXCL10*-high, *GPNMB*-high, *FOLR2*-high, *VSIR*-high, and cycling macrophages for
268 advanced renal cell carcinoma (ccRCC). We find that the MetaTiME-defined macrophage MeCs
269 reflect co-expression relationships with the selected marker genes. For example, macrophage
270 markers from *Bi et al*. rank high in several macrophage-related MeCs (**Fig.S11**), and the
271 expression pattern of the marker genes picked by *Bi et al.* (*CXCL10*, *GPNMB*, *VSIR*, *FOLR2*,
272 Cycling marker *MKI67*) correspond to several MeCs ("interferon response" MeC-0, "PPARG+
273 lipid-rich" MeC-37, "MHCII-high" MeC-58, and "RNASE1+,C1Q+" MeC-49) (**Fig.S11**). However,
274 MetaTiME reveals additional distinct components such as the "SPP1+" and "C1Q+" MeCs,
275 which were detected as separate myeloid types in the *Cheng et al.* multi-cohort study (**Fig.S12**).
276 While the manual reconciliation of cell types from multi-cohort scRNA data shows many marker
277 genes to be consistent with the top genes in the MetaTiME MeCs, the myeloid cell population is
278 not neatly partitioned into cell clusters and might be better represented in terms of expression
279 signature continuums. For example, when mapping myeloid MeCs onto the kidney myeloid
280 cells, the IL1B+ MeC signature (MeC-42, MeC-45) is distributed across the "Macro_IL1B"
281 cluster as well as the CD14 monocyte cluster (**Fig.S12**).
282
283 To investigate functional differences among the different macrophage states, we applied gene
284 set enrichment (GSEA)[34] analysis using the top MeC genes (**Methods**). Interestingly, the
285 different macrophage states have different metabolic preferences (**Fig.4d**). Glucose metabolism
286 and the glycosylation pathway are highly active in SPP1+ macrophages, while lysosome and

287 phagosome activity are the most highly enriched in C1Q+ macrophages. Lysosome and
288 cholesterol metabolism, including *PPARG* signaling, are enriched in the lipid-rich state. The
289 inflammatory IL1B and NFkB pathways are highly active in IL1B+ macrophages. Several
290 macrophage states are related to cell signaling. *SPP1* for example, encodes Osteopontin, which
291 has been found to foster an environment that promotes cancer metastasis[44]. The C1Q+ MeC
292 features *C1QA*, *C1QB*, and *C1QC*, members of the family of complement molecules that could
293 play dual roles in chronic inflammation[45]. The IL1B+ meta-components features cytokines co-
294 expressed with *IL1B*, including *CXCL8*, *CXCL2*, and *CXCL3*, all of which can interact with other
295 cells in the TME by binding to cytokine receptors[46] (**Fig.2b**).
296
297 **Incorporation of epigenetic data prioritizes transcriptional regulators of tumor immunity**
298 We next investigated the transcription factors (TFs) that regulate the MeCs, hypothesizing that
299 the co-expression of genes in a subset of MeCs is determined through TF regulatory events.
300 Our group previously developed the Cistrome Data Browser and Lisa to predict transcriptional
301 regulators of gene sets based on chromatin immunoprecipitation with sequencing (ChIP-seq)
302 data[27,47]. Thus, we used Lisa to predict the TFs that regulate the top genes of each MeC, and
303 compared these Lisa regulatory prediction scores with the MeC z-weights across TFs. We found
304 that, for many MeCs, the same TFs were predicted to be both regulators of the MeC and were
305 highly expressed in the MeC itself, indicating an autoregulatory control scheme. Often, however,
306 TFs that were predicted by Lisa to be MeC regulators were not represented by high MeC z-
307 weights, and TFs with high MeC z-weight were not always found to have high Lisa scores
308 (**Fig.5, Table S1**: MeC regulators). TFs predicted by Lisa but not represented by high MeC z-
309 weight could be the result of TF activities being regulated through non-transcriptional
310 mechanisms[48] or multiple TFs in a family having similar binding patterns but only a subset being
311 the regulators[49]. TFs that have high MeC z-weights but low Lisa scores are most likely not well
312 represented in the relevant cell types in available ChIP-seq data. In the “interferon response”

313 MeC-0, *STAT1* is highly represented in the MeC z-weight and Lisa ranks *STAT1* as the top
314 regulator, consistent with *STAT1* being known as the master regulator of the interferon
315 response (**Fig.5a**). Several lineage-defining TFs display the autoregulatory pattern, including
316 *TCF4* in plasmacytoid dendritic cells (pDC) (**Fig.5b**) and *XBP1* in B plasma cells (**Table S1:**
317 MeC regulators**)**. The macrophage related MeCs are regulated by myeloid lineage TFs like
318 *CEBPB*, and TFs related to immune stimulus responses, including NFkB complex TFs. In the
319 "lipid-rich macrophage" MeC-37, although PPARG ranks among the top Lisa-predicted
320 regulators (**Fig.5c**), *PPARG* expression is not highly represented as a MeC z-weight. In this
321 MeC, the top co-expressed genes are indeed enriched in the PPARG signaling pathway
322 (**Fig.4d**); this result can be accounted for by PPARG being regulated through its ligands, which
323 include a variety of lipophilic acids[48].
324
325 We found glucocorticoid receptor (GR) signaling to be implicated in the regulation of the
326 “CXCL13+ Tfh” MeC-20, with GR being most highly ranked TF in both MeC and Lisa scores
327 (**Fig.5d**). Top genes in the “CXCL13+ Tfh” MeC include several direct target genes of GR,
328 including *SRGN* and *FKBP5* (**Fig.S13b**). We investigated whether the *CXCL13* cytokine itself
329 could be a direct target of GR in *CXCL13* secreting Tfh cells. Since GR ChIP-seq data is not
330 available for the exact Tfh cell state, we collected GR ChIP-seq data from several other cell
331 types. Direct binding of GR is observed at the *CXCL13* gene promoter and nearby the gene
332 locus at putative enhancers, which are conserved across multiple cell lines (**Fig.S13a**), including
333 the B cell line Nalm6, the monocyte cell line THP1, and cancer cell lines. Moreover, in another
334 CXCL13 secreting cell state, the “CXCL13+ exhausted CD8 T cell” MeC-40, GR is also highly
335 ranked in both Lisa and MeC scores (**Fig.5d**). Thus, we hypothesize that GR is likely to be a
336 transcriptional driver of the CXCL13-secreting cell states in exhausted CD8 T cells[50] as well as
337 in CD4 T follicular helper T cells. Thus, the GR pathway could be a candidate target in tumor
338 immunity modulation.

339

340 ## Discussion

341 We developed the MetaTiME (Meta-components of the tumor immune Microenvironment)
342 framework and performed a large-scale and pan-cancer integration of tumor single cell datasets
343 using ICA to optimize information independence among components[51]. We identified 86
344 interpretable meta-components (MeCs) that describe common aspects of TME gene expression
345 variation across multiple tumors.

346

347 The MetaTiME MeCs serve as comprehensive transcriptional signatures that depict a functional
348 landscape of TME transcriptional programs and cell states. For monocytes and macrophages,
349 the related MeCs revealed heterogeneity and plasticity of tumor-associated macrophages
350 (TAMs). We thus propose that TAMs, especially for solid tumors, should be classified based on
351 the major states with different metabolic preferences instead of the canonical M1 and M2
352 classification[42]. Similar states that do not fit well into the M1 versus M2 classification scheme
353 were also observed in previous studies analyzing myeloid cells in the TME, where single cells
354 were clustered and labeled using differential markers[10,12,43]. *Cheng et al.*[10] defined several TAM
355 types by clustering myeloid cells separately for each study and naming the TAMs with manually
356 selected top marker, chosen based on consistency across cohorts. Similarly, *Bi et al.*[43] defined
357 TAM types by harmonizing patients and naming the TAMs with top genes in each cluster for an
358 advanced renal cell carcinoma (ccRCC) cohort. Cell type definitions in the previous studies
359 were based on representative genes, which were chosen differently in the respective studies.
360 We propose that the MetaTiME derived monocyte and macrophage MeCs could be used to
361 define macrophage states and functional co-expressed gene modules more consistently for the
362 TME. Reexamining the marker genes from the previous studies: *NLRP3* is highly ranked in the
363 "IL1B+" MeC; in fact, the NLRP3 inflammasome mediates interleukin-1β production. *GPNMB* is

364 weighted among the top 20 genes in both the “C1Q+ macrophage” and “SPP1+ macrophage”
365 MeC-19 and MeC-26; it encodes a membrane glycoprotein which is typically highly expressed in
366 macrophages. *FOLR2* is ranked 29th in the “C1Q+ macrophage” MeC, indicating this
367 macrophage state also encodes a high folate-activated pathway. Finally, the “SPP1/C1Q
368 macrophage” MeC features an intermediate state with both SPP1 and C1QA, indicating the
369 plasticity and mixed nature of pathways activated in TAMs that could not be defined using
370 exclusive markers. Thus, the myeloid MeCs may provide a consistent definition of TAM states
371 corresponding to different metabolic processes.
372
373 MetaTiME provides a toolkit for analyzing independent TME scRNA-seq datasets by mapping
374 gene expression onto the MeC space. The outputs include signature continuums and the most
375 highly enriched cell states. Recent useful single-cell dataset integration algorithms such as
376 Harmony[22] and scArches[52] infer a joint low-dimensional representation among data. In these
377 approaches the shared space is re-computed every time a new dataset is incorporated. The
378 MetaTiME strategy builds upon previous approaches that transfer latent representations from
379 large datasets, but provides a stable and interpretable representation specialized for the TME.
380
381 By leveraging ChIP-seq data, MetaTiME reveals critical transcriptional regulators in tumor
382 immunity. In many cases, we found the joint consideration of MeC-specific co-expression
383 patterns and TF binding enrichments reveals the roles of TFs in defining cellular states and
384 gene expression programs. MetaTiME captured multiple known TFs critical to tumor immunity
385 and could serve as immune modulation targets; this includes *TOX* in the “CXCL13-secreting
386 exhausted CD8 T cell” MeC-40, a recently discovered regulator of T cell exhaustion[53] (**Table
387 S1**). The MeCs further implicated the glucocorticoid receptor pathway in the regulation of
388 several T cell states. Glucocorticoids are a class of steroid hormones essential to the
389 modulation of multiple biological processes, including immune related ones[54], although the role

390 of the GR pathway in different immune cell types is not fully understood. Since GR is broadly
391 expressed in many cell types, and is regulated through ligand binding, differential analysis of
392 GR expression is unlikely to fully capture GR regulation in single-cell data analysis. Though GR
393 ChIP-seq is not available in the contexts of the relevant T cell states, GR ChIP-seq in other cell
394 lines demonstrate robust binding nearby the top gene *CXCL13*. *CXCL13* is crucial to T follicular
395 helper cell communication with germinal center B cells, through interaction with its receptor
396 CXCR5[55,56].
397
398 Overall, MetaTiME depicts the functional landscape of transcriptomic variation and cell states in
399 the tumor microenvironment. It provides a computational framework to facilitate the elucidation
400 of the identity and function of cells in the TME in future studies and will facilitate the
401 identification of potential new therapeutic targets for immune modulation.
402
403
404 Methods
405
406 **Tumor single-cell RNA-seq data collection and processing**
407 For an extensive collection of single cells from tumor microenvironment, we utilized the public
408 tumor scRNA-seq collection, TISCH[28]. The TISCH collection uniformly processed each dataset
409 with MAESTRO[17] and isolated non-malignant environmental cells from malignant cells. Overall,
410 we collected 2,157,387 cells from 76 studies ranging 27 cancer types. The MAESTRO
411 annotation was labeled using CIBERSORT gene panels[18] followed by curation, enabling
412 selection of 1,719,813 tumor micro-environment cells, including 1,617,110 immune cells and
413 102,703 stromal cells, were retained for integrative analysis in this study. For studies with data
414 measured from multiple cancer types, cells different cancer types were split into independent

415 datasets, resulting in 93 datasets; it includes 3 PBMC datasets from healthy donors from 10X
416 Genomics as baseline and 7 datasets with ICB treatment.
417
418 For an unsupervised component analysis, each dataset was re-analyzed. For datasets with raw
419 count matrices available, gene expression was normalized towards per-cell read depth 10,000
420 followed by log transformation. For datasets with only TPM or FPKM values available, including
421 Smart-seq data or studies with only normalized matrix available, gene expression underwent log
422 transformation. Cells were filtered based on minimum library size 1000, gene number 500, and
423 maximum mitochondrial read proportion 5%.
424
425 **Decomposing individual studies and denoising low-dimensional components.**
426 We then decomposed the expression matrix of each scRNA-seq dataset using fastICA[51] into an
427 independent component (IC) vector matrix and a projection weight matrix. We tested different
428 values for the number of components (k) and chose k to be 100 uniformly for each dataset,
429 given it could cover more variations than the number of cell types in the TME, which is around
430 twenty. We applied two denoising approaches to handle sparsity and potentially noisy ICs. First,
431 we performed a z-score transformation of the gene loadings in the component, scaling all gene
432 loading values by the standard deviation of each IC. The gene loadings indicate the degree of
433 contribution to the component as a “metagene” from each gene, and we observed that most
434 genes contribute neutrally to the metagene. Thus, genes with significant contributions are
435 selected using the two-standard deviation threshold from either the positive or the negative side.
436 Second, we aligned the positive skewness of components (using scipy.stats.skewtest[57]) since
437 the sign of an independent component is randomly assigned in fastICA optimization. We
438 observed that asymmetrically extreme gene loading values highlight genes representative of the
439 component's function; thus, we computed each component's skewness statistics and flipped the
440 sign of component loadings if the skewness is negative. We excluded genes with a low

441 contribution (gene weight not passing two standard deviations) to any reproducible components
442 and kept 6623 genes with potentials in driving the reproducible components. The post-
443 decomposition steps ensured the attitude and sign of gene weights are comparable across
444 cohorts, depicting degree of contribution from each gene in the genome-wide background.
445

446 **Meta-component calling and functional annotation**

447 We then aim to discover reproducible patterns from all components from each dataset. We
448 evaluated similarity between pairs of components using cosine distance and retained a set of
449 1043 candidate reproducible components from 69 datasets, each with a minimum Pearson
450 correlation coefficient 0.3 with at least one different IC. We then clustered ICs using Louvain
451 clustering, a graph-based community detection algorithm where the resolution parameter
452 controls segmentation granularity. Clusters with at least five ICs were retained as reproducible
453 IC clusters. The number of clusters is determined by optimizing both Silhouette's score for
454 optimal within-cluster similarity compared to inter-cluster similarity and number of reproducible
455 clusters. The final resolution parameter was chosen to be 1.25 resulting in 86 clusters for meta-
456 component (MeC) calling. The consensus gene z-weights in each MeC were then calculated by
457 averaging ICs in each cluster. Genes of outlier z-weights passing two standard deviations were
458 highlighted as significant, and the ones with positive largest z-weights were considered
459 representative of the MeC.
460

461 MetaTiME MeCs were assigned curated annotation by matching top z-weighted genes to
462 functional biological information including cell type markers, pathway databases from GSEA,
463 cell types expressing top MeC genes, and high-rank transcription factors. In detail, GSEA
464 enrichment analyses utilized the top 100 highest z-weighted genes and TF database was
465 obtained from AnimalTFDB[49]. GSEA was also performed using the z-weight rank of all genes[58].

466 The 86 MeCs were first ordered by MeC cluster size, then organized into seven functional
467 categories with six lineage-related categories and one signaling category.
468

469 **Simulating multi-cohort single-cell RNA data with expression programs**

470 To benchmark dimensional reduction methods, we built upon previous effort from *Kotliar et al.*[59]
471 to use the scsim package to simulate multiple count matrices with built-in transcriptional
472 programs. In principle, the built-in gene expression programs (GEP) were sampled as random
473 scaling factors on a subset of genes mimicking overexpression or suppression of a pathway.
474 For testing whether a higher number of cohorts facilitate GEP recovery, we simulated 20 single-
475 cell datasets and tested usage of 5 cohorts, 10 cohorts, and 20 cohorts. Each dataset was
476 embedded with a subset of 14 pre-defined GEPs, since the real tumor scRNA data may not
477 cover every possible cell type or gene program in every dataset. The 14 GEPs contain 13 cell
478 type-specific programs with distinct cell type-specific genes, and one signaling gene expression
479 program that is randomly active in multiple cell types. Two low-dimensional reduction method
480 are benchmarked using simulated scRNA data: independent component analysis (ICA) and
481 non-negative matrix factorization (NMF). Decomposition was performed on each single cohort
482 separately, and meta-component calling was done as similar in MetaTiME: components are
483 filtered, clustered into meta-components, followed by averaging gene z-weights per cluster as
484 predicted gene expression programs (GEP). The predicted GEPs were compared with pre-
485 defined True GEPs using Pearson correlation. Overall, both ICA and NMF can recover GEPs,
486 while the ICA-based GEPs are more mutually independent and performs slightly better. Since
487 the GEP recovered in the 20 cohorts case matched true GEP better than 5 cohorts and 10
488 cohorts, the increased number of cohorts also improves GEP recovery. Thus, we chose ICA for
489 component integration and use all available datasets for GEP discovery for tumor
490 microenvironmental cells.

491

492 **MeC calling robustness with leave-one-out testing.**

493 MetaTiME's robustness was examined by re-calling MeCs after leaving one dataset out and
494 calculate the Pearson correlation between the left out dataset MeCs and the full-set MeCs. The
495 leave-one-out analysis was first performed by removing the SCC cells from the Yost et al.
496 dataset, which was shown as an example of MetaTiME annotation. Each dataset in the MeC
497 pool was left out to call new MeCs using the same parameters, resolution=1.25 and
498 min_cluster_size=5. We used two metrics to check the similarity of each leave-one-out
499 experiments compared to the full set. The first metric is the mean maximum correlation between
500 leave-one-out and the full set. This is calculated by taking the maximum correlation, row-wise
501 and column-wise in the correlation heatmap, and then taking the average. The second metric is
502 the number of MeCs.

503

504 **Comparing MeCs, signatures by post-embedding integration, to signatures by cluster-**
505 **wise differential expression.**

506 We compared MetaTiME MeCs to the standard approach that extracts signatures after
507 integrating cells across datasets. The standard approach to obtain cluster-wise signatures is to
508 perform dataset harmonization, cell clustering, and cluster-wise differential expression analysis.
509 When performing dataset integration, we first used CCA to include as many datasets as
510 possible given our maximum memory available in our computing resource (150GB). The
511 maximum number of datasets integrated by CCA is 10 when using datasets ranked by the
512 number of TME cells. We then used scanpy to achieve the successful integration of 21
513 datasets. The integrated cells are harmonized with Harmony[22] and clustered, followed by
514 differential expression analysis using the Wilcoxon test to extract cluster-wise signatures. The
515 signatures are in the format of gene vector of log fold change, and named as "Cluster DE
516 signatures". To control the confounding effect of signature numbers, the resolution of clustering

517 is chosen to reach 90 clusters, a similar number with the number of full-set MeCs which is 86.
518 Then, we compared MetaTiME MeCs to the Cluster DE signature projected onto a test dataset.
519 We used same BCC cells as the test cells that are never seen in the training step from either
520 method. To reflect how variable the projected scores are across cells, we grouped cells from the
521 test BCC data into 25 clusters with Leiden clustering. The heatmap of the cluster-wise signature
522 mapping is plotted to observe how specific each type of score is across cell groups. We further
523 used cross-cluster entropy to quantify the information content of scores relative to clusters in the
524 test dataset.
525
526 **The MetaTiME annotator for analyzing new tumor scRNA-seq data.**
527 MetaTiME provides an analytical toolkit for annotating cell states and signature activities for
528 tumor scRNA-seq data (https://github.com/yi-zhang/MetaTiME). The scRNA-seq data is first
529 processed following standard procedures, which includes cell depth normalization, log-
530 transformation, batch effect removal using Harmony[22], neighboring graph construction, graph
531 clustering, and UMAP embedding for visualization. Specifically, the clustering step uses an
532 over-clustering strategy, which sets a high-resolution parameter (default 8) that generates a
533 larger number of clusters and help reveal fine structures among the cells. Then, the MetaTiME
534 annotator tool takes as input a single log-transformed expression matrix for TME cells from the
535 dataset. The outputs include both per-cell MetaTiME MeC signature scores and per-cluster
536 enriched MeC state. For the per-cell score, MetaTiME projects each cell onto the MeC space by
537 calculating the dot product between the expression vector and the z-weight vector of each MeC,
538 using genes passing the significant z-weight criterion (z-weight >2). The projection matrix is
539 then scaled across all cells to ensure normally distributed scores within each MeC, outputting
540 the cell-by-MeC score matrix. Meanwhile, the UMAP view of the projection score shows the
541 signature gradient across the cells positioned by similarity. Lastly, the cluster-wise MeC
542 enrichment results are also generated. The per-cluster MeC enrichment score is calculated by

543 averaging profile of cells along each MeC; MeCs with mean score passing the significant cutoff
544 (2 in the z-weight scale) are called as the set of enriched MeCs. Each cluster may enrich
545 multiple number of MeCs, and the top enriched MeC with highest score is used in UMAP
546 visualization.
547
548 **Differential MeC signature analysis**
549 For tumor scRNA-seq data with different conditions, a differential signature analysis can be
550 carried out following MetaTiME annotation, which provides enriched MeCs for each cluster and
551 names each cluster with the top enriched MeC. Thus, for each cell cluster, the MeC signature
552 strength can be compared across conditions, for all enriched MeC in the current cluster. In
553 details, a simple Wilcoxon rank-sum test is adopted to compare MeC scores of cells in one
554 condition with another. The log-fold change of MeC scores were calculated by the ratio between
555 cell means from the two conditions in comparison. To plot the cluster-specific differential
556 signature plot, the signatures are marked using “EnrichedMeC@ClusterName”, where the
557 “ClusterName” is the top first enriched MeC used as cell state as current cluster. When
558 “EnrichedMeC” is the same as the “ClusterName”, only “ClusterName” is marked on the
559 Significance-Effect size plot.
560
561 **Incorporation of epigenetic data using Lisa**
562 Our group previously developed Lisa that predicts the influence of TFs on a set of genes. Lisa
563 models public chromatin accessibility and TF binding profiles to score TFs in gene regulation
564 from an epigenetic perspective. We developed Lisa2 that improves on running speed and
565 pipeline integration, which is applied on each MeC to score TFs in regulation potential on top
566 100 high z-weighted genes. The impact scores of TFs are thus from two sources: MeC z-
567 weights for expression representation, and Lisa scores for binding potential. The TFs are

grouped into three classes, TFs highly ranked based on both MeC gene weights and Lisa significance, TFs representative only in MeC, and TFs based on binding information only. In TableS1, we marked TFs from different classes in different colored columns. Significant TFs based on both MeC and Lisa (MeC z-weight >=2, Lisa score -log p-value >=2) are marked as orange color; furthermore, the TFs ranking among top 40 (aggregated rank of MeC and Lisa) compared to all genes are further marked red. TFs ranking among top 10 only in MeC z-weight are colored green, and TFs ranking among top 10 only in Lisa score are colored blue.

Grant information

NIH grants P01CA163222, U24CA237617, and R01HG011139.

Acknowledgement

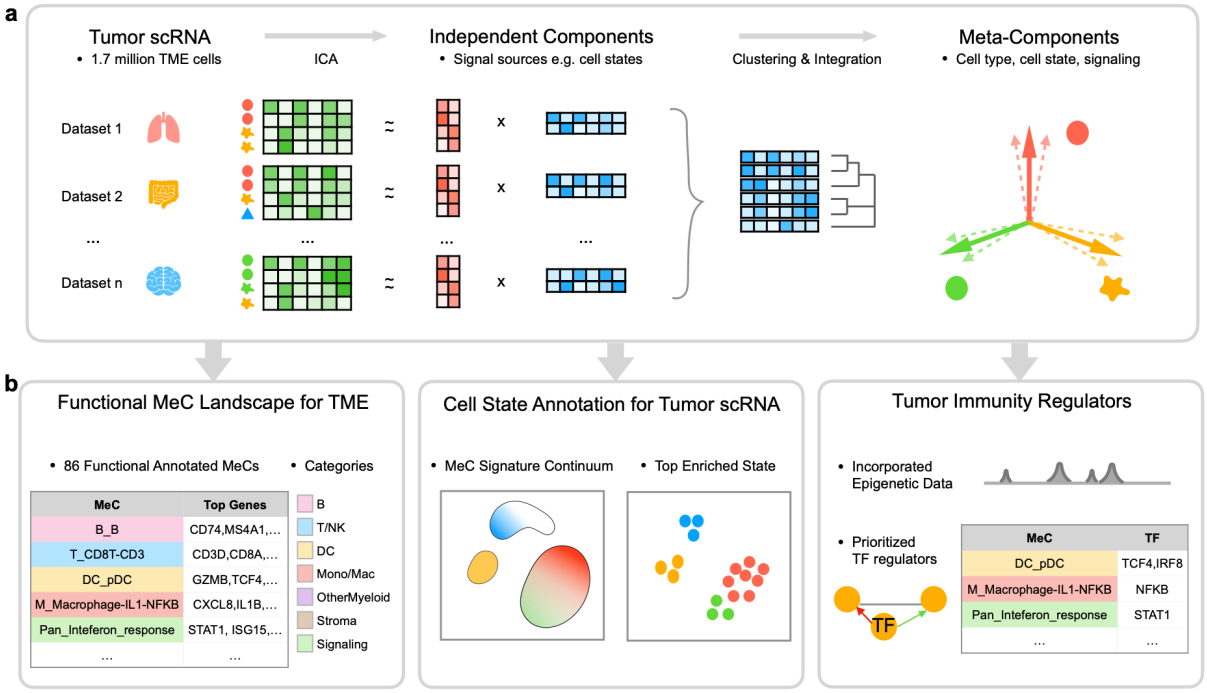
We acknowledge all authors of the original tumor single-cell studies that generated publicly available data. The results appearing here are in part based upon the data generated by the TCGA Research Network (http://cancergenome.nih.gov/).

584 Conflict of Interest

585 MB is a consultant to and receives sponsored research support from Novartis. MB serves on the
586 SAB of H3 Biomedicine, Kronos Bio, and GV20 Oncotherapy. X.S.L conducted the work while
587 being on the faculty at DFCI, and is currently a board member and CEO of GV20 Therapeutics.
588 The remaining authors declare no competing interests.
589

590 Figures

591 **Fig.1.**



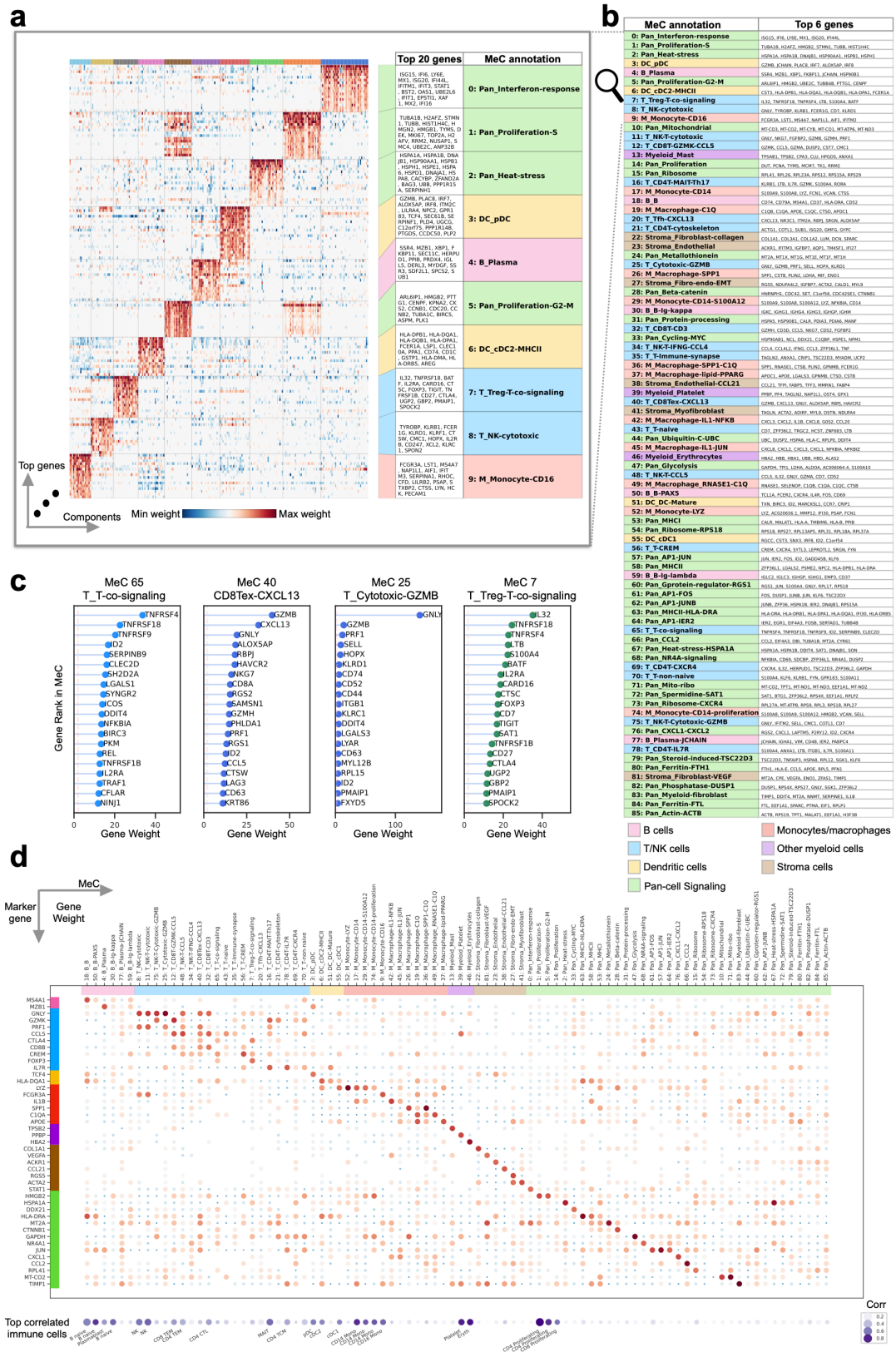
592

593

594 **Fig.1. Overview of MetaTiME.** MetaTiME integrates 1.7 million single cells to learn common
595 transcriptional programs in the tumor microenvironment (TME). **(a)** Steps for Meta-components
596 (MeCs) discovery. For each scRNA dataset, the expression matrix of TME cells is decomposed
597 into a loading matrix (red) and an independent component (IC) matrix through independent
598 component analysis (ICA). The ICs represent mutually independent sources of transcriptional
599 variation. ICs from each dataset are concatenated and clustered into groups of ICs with high
600 similarity, representing transcriptional programs shared across TME. MeCs are then calculated
601 as averaged profiles of ICs from each cluster. Each MeC is interpretable, representing gene
602 signatures of cell type, cell states, or signaling pathway activities. **(b)** Left: MetaTiME provides
603 86 functionally annotated MeCs that depict the TME transcriptional landscape. They are
604 grouped into six lineage-related categories and one category reflecting signaling activities.
605 Middle: the MetaTiME annotation tool facilitates automatic annotation of cell states for new

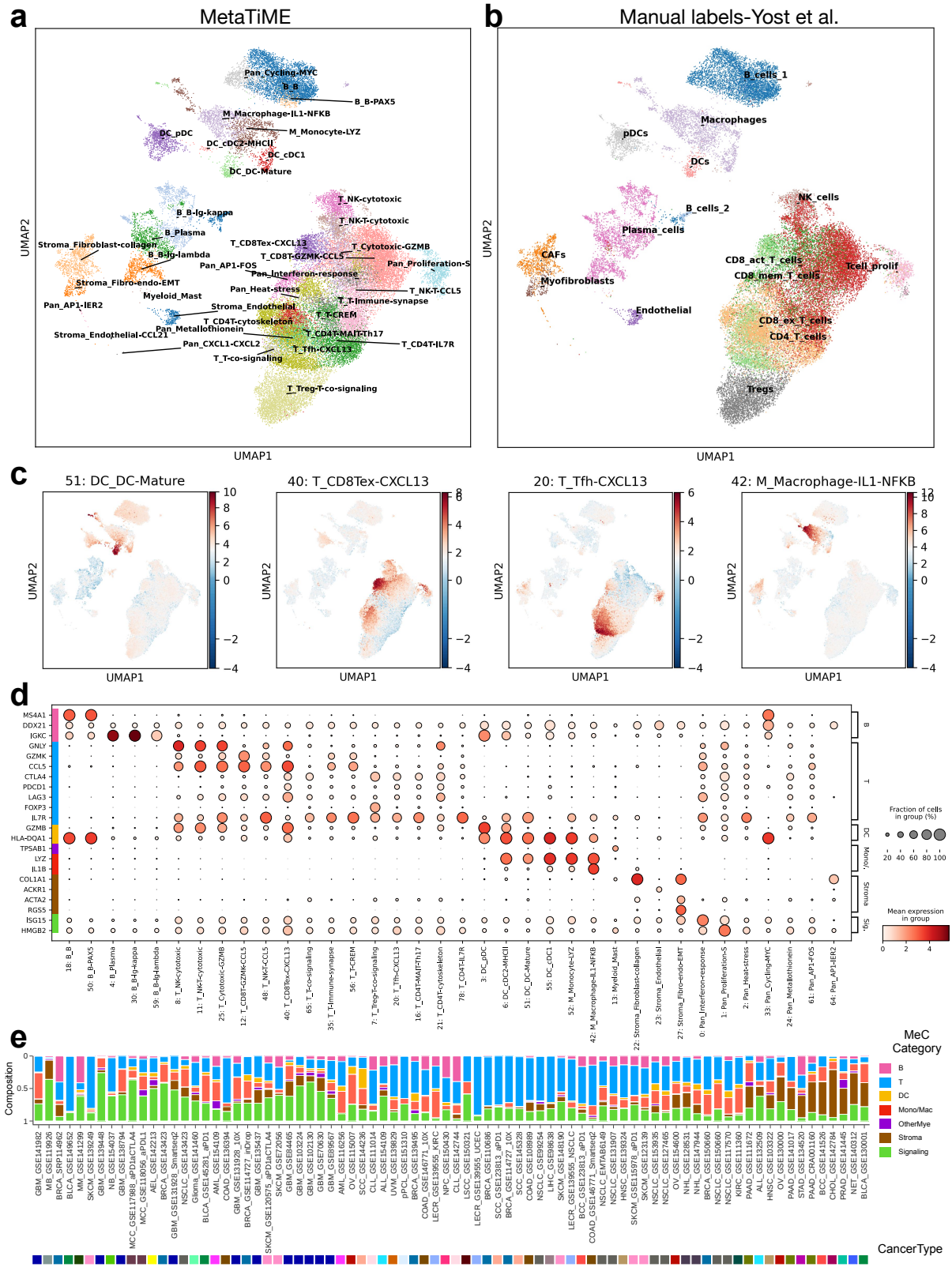
606 tumor scRNA data. Right: candidate regulators of each MeC are prioritized by combining MeC
607 gene weights with epigenetics data. MeC: meta-components, TME: tumor microenvironment,
608 ICA: independent component analysis, MeC: meta-component, TF: transcription factor.
609

Fig. 2.


a
Top genes
Components
Min weight
Max weight
Top 20 genes
MeC annotation
0: Pan_Interferon-response
1: Pan_Proliferation-S
2: Pan_Heat-stress
3: DC_pDC
4: B_Plasma
5: Pan_Proliferation-G2-M
6: DC_cDC2-MHCII
7: T_Treg-T-co-signaling
8: T_NK-cytotoxic
9: M_Monocyte-CD16
b
MeC annotation
Top 6 genes
B cells
T/NK cells
Dendritic cells
Pan-cell Signaling
Monocytes/macrophages
Other myeloid cells
Stroma cells
c
MeC 65
T_T-co-signaling
MeC 40
CD8Tex-CXCL13
MeC 25
T_Cytotoxic-GZMB
MeC 7
T_Treg-T-co-signaling
Gene Rank in MeC
Gene Weight
d
MeC
Marker gene
Gene Weight
Top correlated immune cells
Corr

612 **Fig.2. MetaTiME meta-components are biologically interpretable with top genes. (a)**
613 Heatmap of top ten most recurrent clusters of MeCs showing normalized gene weights. **(b)**
614 Biological characterization of each MeC with top genes. To facilitate biological interpretation,
615 MeCs are categorized into six lineage-associated classes (B cells, T cells for CD4T, CD8T and
616 NK cells, dendritic cells, monocyte and macrophages, other myeloid cell types, and stroma
617 cells) and one signaling pathway-associated class. **(c)** Examples of T cell related MeCs with top
618 20 genes with largest weights. **(d)** Gene contribution of known lineage-related biomarkers for
619 each MeC, and correlation with known immune markers from Azimuth. In the top dot plot, size
620 and color represents MeC z-weights of each gene in each MeC. In the bottom dot plot, size and
621 color represents the maximum correlation coefficient between MeC and Azimuth defined marker
622 genes per cell type. MeC: meta-component, DC: dendritic cell, Mono/Mac: monocytes and
623 macrophages.
624

Fig.3.


a
MetaTiME
Pan_Cycling-MYC
B_B
B_B-PAX5
M_Macrophage-IL1-NFKB
M_Monocyte-LYZ
DC_pDC
DC_cDC2-MHCII
DC_cDC1
DC_DC-Mature
T_NK-cytotoxic
B_B-Ig-kappa
T_NK-T-cytotoxic
B_Plasma
Stroma_Fibroblast-collagen
T_CD8Tex-CXCL13
T_Cytotoxic-GZMB
B_B-Ig-lambda
T_CD8T-GZMK-CCL5
Pan_AP1-FOS
Pan_Proliferation-S
Pan_Interferon-response
Stroma_Fibro-endo-EMT
Pan_Heat-stress
T_NK-T-CCL5
Myeloid_Mast
T_T-Immune-synapse
Pan_AP1-IER2
Stroma_Endothelial
T_T-CREM
T_CD4T-cytoskeleton
Pan_Metallothionein
T_CD4T-MAIT-Th17
Stroma_Endothelial-CCL21
Pan_CXCL1-CXCL2
T_CD4T-IL7R
T_Tfh-CXCL13
T_T-co-signaling
T_Treg-T-co-signaling
UMAP2
UMAP1
b
Manual labels-Yost et al.
B_cells_1
Macrophages
pDCs
DCs
NK_cells
B_cells_2
Plasma_cells
CAFs
CD8_act_T_cells
Tcell_prolif
CD8_mem_T_cells
Myofibroblasts
Endothelial
CD8_ex_T_cells
CD4_T_cells
Tregs
UMAP2
UMAP1
c
51: DC_DC-Mature
40: T_CD8Tex-CXCL13
20: T_Tfh-CXCL13
42: M_Macrophage-IL1-NFKB
UMAP2
UMAP1
d
MS4A1
DDX21
IGKC
GNLY
GZMK
CCL5
CTLA4
PDCD1
LAG3
FOXP3
IL7R
GZMB
HLA-DQA1
TPSAB1
LYZ
IL1B
COL1A1
ACKR1
ACTA2
RGS5
ISG15
HMGB2
B
T
DC
Mono/Mac
Stroma
Sig.
18: B_B
50: B_B-PAX5
4: B_Plasma
30: B_B-Ig-kappa
59: B_B-Ig-lambda
8: T_NK-cytotoxic
11: T_NK-T-cytotoxic
25: T_Cytotoxic-GZMB
12: T_CD8T-GZMK-CCL5
48: T_NK-T-CCL5
40: T_CD8Tex-CXCL13
65: T_T-co-signaling
35: T_T-Immune-synapse
56: T_T-CREM
7: T_Treg-T-co-signaling
20: T_Tfh-CXCL13
16: T_CD4T-MAIT-Th17
21: T_CD4T-cytoskeleton
78: T_CD4T-IL7R
3: DC_pDC
6: DC_cDC2-MHCII
51: DC_DC-Mature
55: DC_cDC1
52: M_Monocyte-LYZ
42: M_Macrophage-IL1-NFKB
13: Myeloid_Mast
22: Stroma_Fibroblast-collagen
23: Stroma_Endothelial
27: Stroma_Fibro-endo-EMT
0: Pan_Interferon-response
1: Pan_Proliferation-S
2: Pan_Heat-stress
33: Pan_Cycling-MYC
24: Pan_Metallothionein
61: Pan_AP1-FOS
64: Pan_AP1-IER2
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2 4
e
Composition
0
0.5
1
MeC Category
B
T
DC
Mono/Mac
OtherMye
Stroma
Signaling
GBM_GSE141982
MB_GSE119926
BRCA_SRP114962
BLCA_GSE149652
MM_GSE141299
SKCM_GSE139249
GBM_GSE139448
NB_GSE154237
GBM_GSE123814
MCC_GSE117988_aPD1aCTLA4
MCC_GSE118056_aPD1L1
AEL_GSE142213
BRCA_GSE143423
GBM_GSE131928_Smartseq2
NSCLC_GSE143423
Glioma_GSE141460
BLCA_GSE145281_aPD1
AML_GSE154109
COAD_GSE136394
GBM_GSE131928_10X
BRCA_GSE114727_inDrop
GBM_GSE135437
SKCM_GSE120575_aPD1aCTLA4
SKCM_GSE72056
GBM_GSE84465
GBM_GSE103224
GBM_GSE139324
GBM_GSE87630
GBM_GSE89567
AML_GSE116256
OV_GSE115007
SCC_GSE144236
CLL_GSE111014
ALL_GSE154109
UVM_GSE139829
pPCL_GSE151310
BRCA_GSE139495
COAD_GSE146771_10X
LECR_GSE139555_KIRC
NPC_GSE150430
CLL_GSE142744
LSCC_GSE150021
LECR_GSE139555_UCEC
BRCA_GSE110686
SCC_GSE123813_aPD1
BRCA_GSE114727_10X
SCC_GSE145328
COAD_GSE108989
NSCLC_GSE99254
LHC_GSE98638
SKCM_GSE148190
LECR_GSE139555_NSCLC
BCC_GSE123813_aPD1
COAD_GSE146771_Smartseq2
NSCLC_EMTAB6149
NSCLC_GSE131907
HNSC_GSE139324
SKCM_GSE115978_aPD1
SKCM_GSE123139
NSCLC_GSE153935
NSCLC_GSE127465
OV_GSE154600
NHL_GSE94820
NHL_GSE147944
BRCA_GSE150660
NSCLC_GSE150660
NSCLC_GSE117570
KIRC_GSE111360
PAAD_GSE111672
ALL_GSE132509
HNSC_GSE103322
OV_GSE130000
PAAD_GSE141017
STAD_GSE134520
PAAD_CRA001160
BCC_GSE141526
CHOL_GSE142784
PRAD_GSE141445
NET_GSE140312
BLCA_GSE130001
CancerType

627 **Fig.3. MetaTiME annotates cell states with high resolution on tumor microenvironment**
628 **single-cell data. (a)** MetaTiME cell state annotation of cell clusters in a basal cell carcinoma
629 scRNA dataset based on top enriched MeCs. **(b)** Manual annotation labels by experts from the
630 original study shown on the same UMAP space. **(c)** Signature continuums of four MeCs
631 representing the mature dendritic cell state, the CXCL13-secreting exhausted T cell state, the
632 CXCL13-secreting T follicular helper cell state, and the IL1B pathway-activated macrophage
633 state. **(d)** Marker gene expression for each annotated cell cluster as in (a)**. (e)** Bar plot showing
634 cell state composition of tumor microenvironment for tumor scRNA dataset cell states. The
635 proportion of cell states from the same MeC category are aggregated.
636
637
638
639

Fig.4

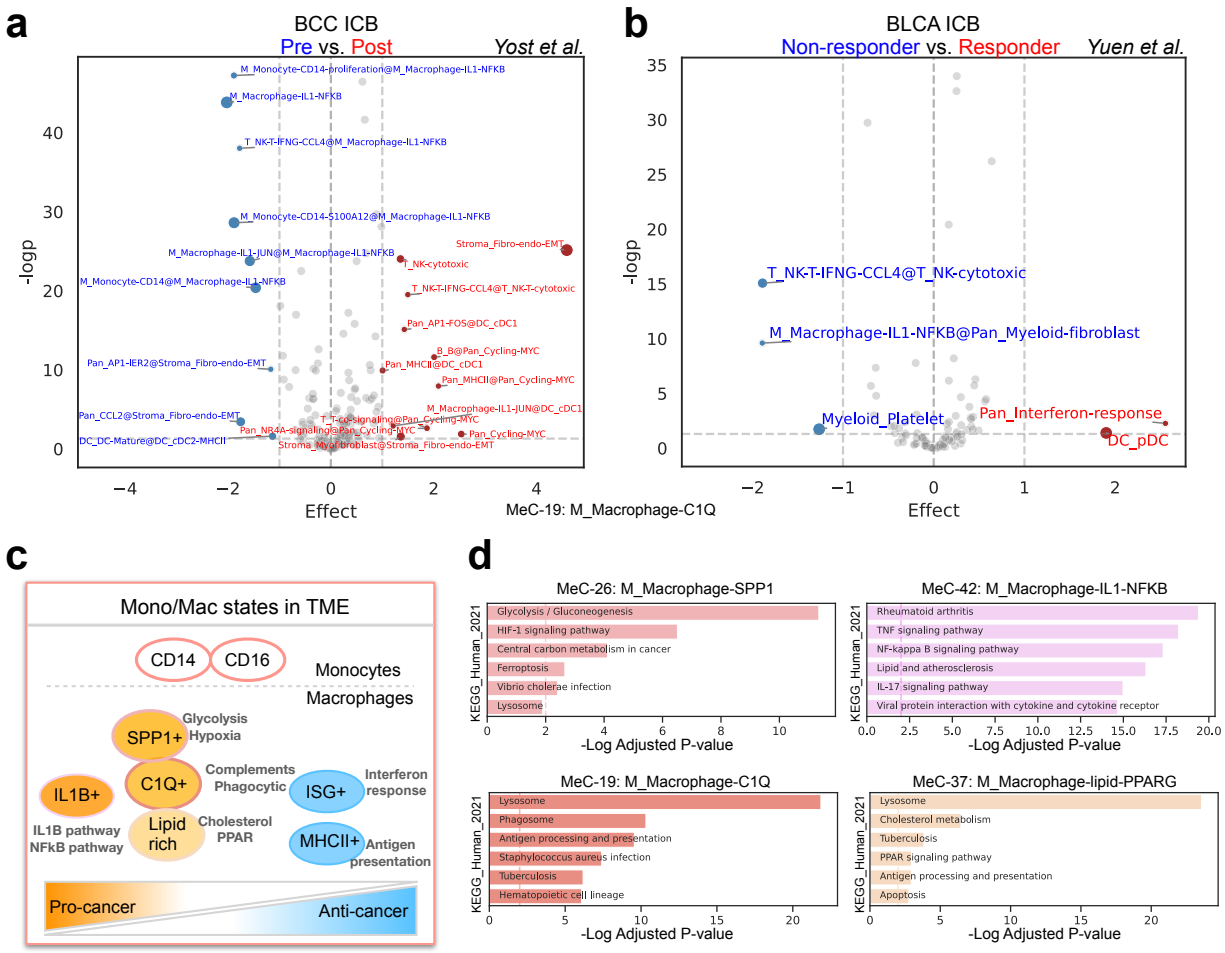


Fig.4. Differential signature analysis and delineated macrophage states in TME. (a) Differential MeC signature testing for enriched cell states comparing pre- and post-immunotherapy conditions in basal cell carcinoma (BCC). X-axis: Log odds ratio of mean signature scores between post- and pre-immunotherapy conditions. Y-axis: minus log p-value from Wilcoxon test. The significant cluster-wise differential signature is marked as "EnrichedMeC@ClusterName"; when the enriched MeC is the same as cluster name, the signature is marked "ClusterName". **(b)** Differential signature testing for enriched cell states comparing non-responders and responders from pre-treatment condition in bladder carcinoma (BLCA). **(c)** Model of monocytes and macrophage states in tumor and their metabolic

651 differences. **(d)** Top pathways enriched in different macrophage MeCs. BCC: basal cell

652 carcinoma, BLCA: bladder carcinoma.

Fig.5.

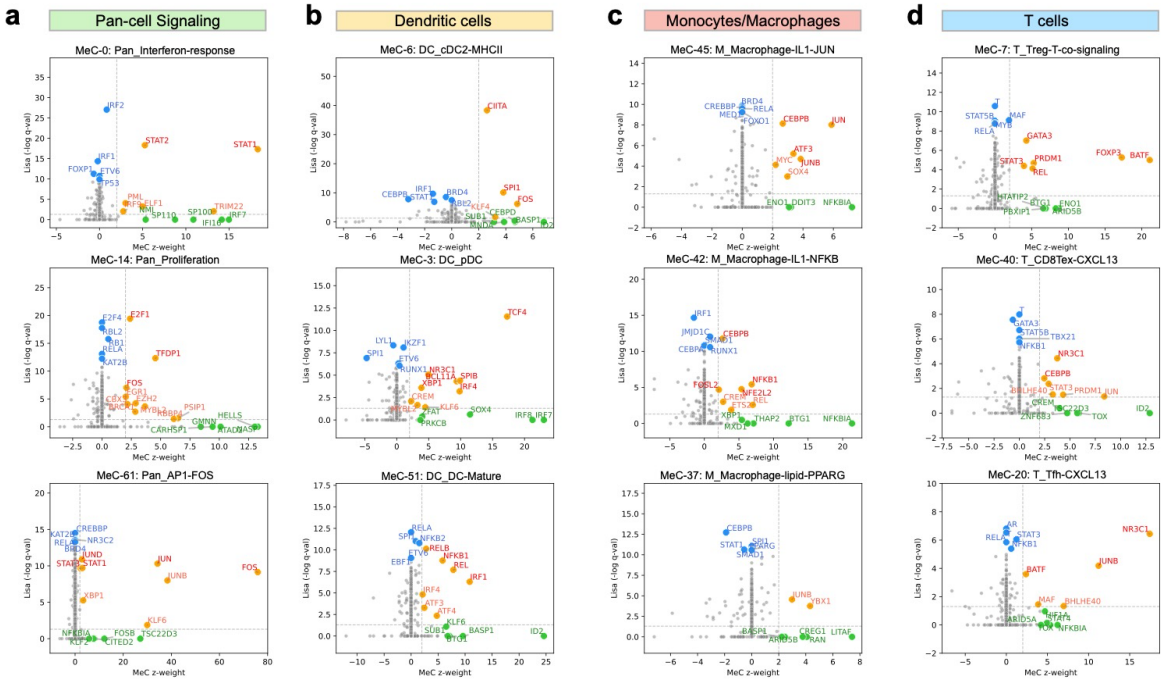


Fig.5. MetaTiME prioritizes tumor immunity transcriptional regulators. For selected MeCs, TFs are prioritized by their MeC expression representation and Lisa, ChIP-seq based, regulatory potentials. X-axis: gene z-weight of the TF for the current MeC. Y-axis: Lisa-based regulatory potential significance for top genes in the current MeC. Red and orange factors: MeC regulators prioritized based on both MeC gene weights and Lisa regulatory potential significance. Green factors: TFs highly weighted in MeCs and not in Lisa analysis. Blue factors: TFs with high Lisa regulatory potential and not highly weighted in MeCs. **(a)** TFs prioritized for three MeCs in the signaling category. **(b)** TFs prioritized for three MeCs in the dendritic cell category. **(c)** TFs prioritized for three MeCs representing different macrophage states. **(d)** TFs prioritized for three MeCs representing different T cell states. TF: transcription factor, MeC: meta-component.

669

670 Reference

671 1. Liu, X. S. & Mardis, E. R. Applications of Immunogenomics to Cancer. *Cell* **168**, 600–612
672 (2017).

673 2. Jin, M. Z. & Jin, W. L. The updated landscape of tumor microenvironment and drug
674 repurposing. *Signal Transduct. Target. Ther.* **5**, (2020).

675 3. Penter, L. *et al.* Local and Systemic Effects of Immune Checkpoint Blockade on Relapsed
676 Myeloid Malignancies Following Allogeneic Hematopoietic Stem Cell Transplantation.
677 *Blood* **136**, 34–35 (2020).

678 4. Zhang, Y. *et al.* The cancer-associated genetic variant rs3903072 modulates immune
679 cells in the tumor microenvironment. *Front. Genet.* (2019). doi:10.3389/fgene.2019.00754

680 5. Gu, S. S. *et al.* Therapeutically Increasing MHC-I Expression Potentiates Immune
681 Checkpoint Blockade. *Cancer Discov.* **11**, 1524–1541 (2021).

682 6. Zeng, Z. *et al.* TISMO: syngeneic mouse tumor database to model tumor immunity and
683 immunotherapy response. *Nucleic Acids Res.* **50**, D1391–D1397 (2022).

684 7. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer
685 immunotherapy response. *Nat. Med.* (2018). doi:10.1038/s41591-018-0136-1

686 8. Yost, K. E. *et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade.
687 *Nat. Med.* (2019). doi:10.1038/s41591-019-0522-3

688 9. Zhang, L. *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies
689 in Colon Cancer. *Cell* **181**, 442-459.e29 (2020).

690 10. Cheng, S. *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid
691 cells. *Cell* **184**, 792-809.e23 (2021).

692 11. Wang, X. *et al.* In vivo CRISPR screens identify the E3 ligase Cop1 as a modulator of
693 macrophage infiltration and cancer immunotherapy target. *Cell* **184**, 5357-5374.e22
694 (2021).

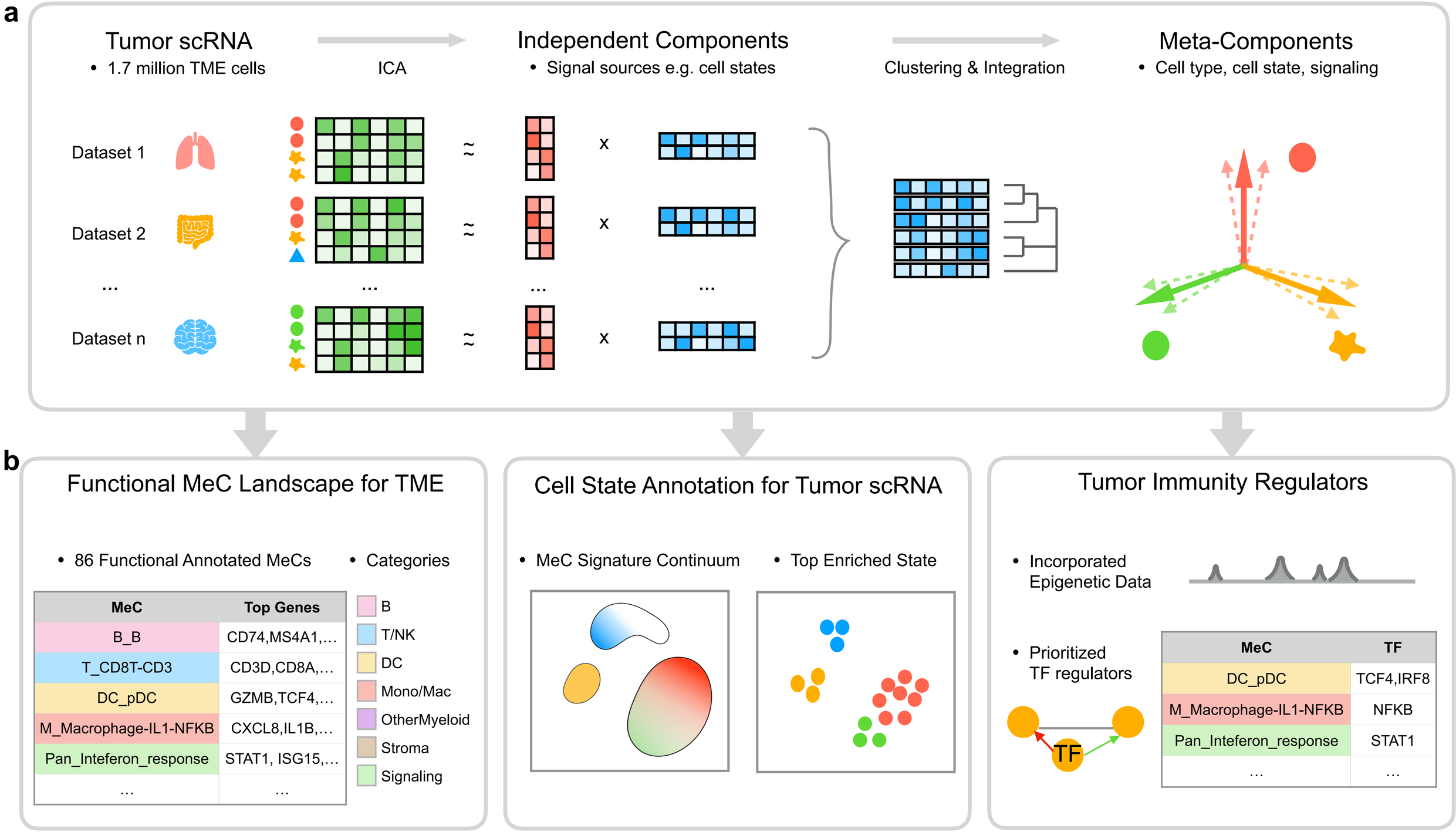
695 12. Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat.*
696 *Genet.* **53**, 1334–1347 (2021).
697 13. Nieto, P. *et al.* A single-cell tumor immune atlas for precision oncology. *Genome Res.* **31**,
698 1913–1926 (2021).
699 14. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21
700 (2019).
701 15. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29
702 (2021).
703 16. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a
704 transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
705 17. Wang, C. *et al.* Integrative analyses of single-cell transcriptome and regulome using
706 MAESTRO. *Genome Biol.* **21**, (2020).
707 18. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles.
708 *Nat. Methods 2015 125* **12**, 453–457 (2015).
709 19. Andreatta, M. *et al.* Interpretation of T cell states from single-cell transcriptomics data
710 using reference atlases. *Nat. Commun.* **12**, 1–19 (2021).
711 20. Zheng, A. L., Qin, S., Wang, A., Xing, B. & Gao, R. Pan-Cancer Single Cell Landscape of
712 Tumor-Infiltrating T Cells. **6474**, (2021).
713 21. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell
714 transcriptomic data across different conditions, technologies, and species. *Nat.*
715 *Biotechnol. 2018 365* **36**, 411–420 (2018).
716 22. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
717 Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
718 23. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for
719 single-cell transcriptomics. *Nat. Methods* (2018). doi:10.1038/s41592-018-0229-2
720 24. Pelka, K. *et al.* Spatially organized multicellular immune hubs in human colorectal cancer.

721 *Cell* **184**, 4734-4752.e20 (2021).
722 25. Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis.
723 *IEEE Trans. Neural Networks* (1999). doi:10.1109/72.761722
724 26. Wang, W. *et al.* Independent component analysis based gene co-expression network
725 inference (ICAnet) to decipher functional modules for better single-cell clustering and
726 batch integration. *Nucleic Acids Res.* **49**, E54 (2021).
727 27. Zheng, R. *et al.* Cistrome Data Browser: expanded datasets and new tools for gene
728 regulatory analysis. *Nucleic Acids Res.* **47**, D729–D735 (2019).
729 28. Sun, D. *et al.* TISCH: A comprehensive web resource enabling interactive single-cell
730 transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.* **49**, D1420–
731 D1430 (2021).
732 29. Wykes, M. N. & Lewin, S. R. Immune checkpoint blockade in infectious diseases. *Nat.*
733 *Rev. Immunol. 2017 182* **18**, 91–104 (2017).
734 30. Zhang, Q. & Vignali, D. A. A. Co-stimulatory and co-inhibitory pathways in autoimmunity.
735 *Immunity* **44**, 1034 (2016).
736 31. Qin, S. *et al.* Novel immune checkpoint targets: moving beyond PD-1 and CTLA-4. *Mol.*
737 *Cancer 2019 181* **18**, 1–14 (2019).
738 32. Thommen, D. S. *et al.* A transcriptionally and functionally distinct PD-1+ CD8+ T cell pool
739 with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat.*
740 *Med. 2018 247* **24**, 994–1004 (2018).
741 33. Chowell, D. *et al.* Improved prediction of immune checkpoint blockade efficacy across
742 multiple cancer types. *Nat. Biotechnol. 2021 404* **40**, 499–506 (2021).
743 34. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for
744 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–
745 15550 (2005).
746 35. Morgan, D. & Tergaonkar, V. Unraveling B cell trajectories at single cell resolution.

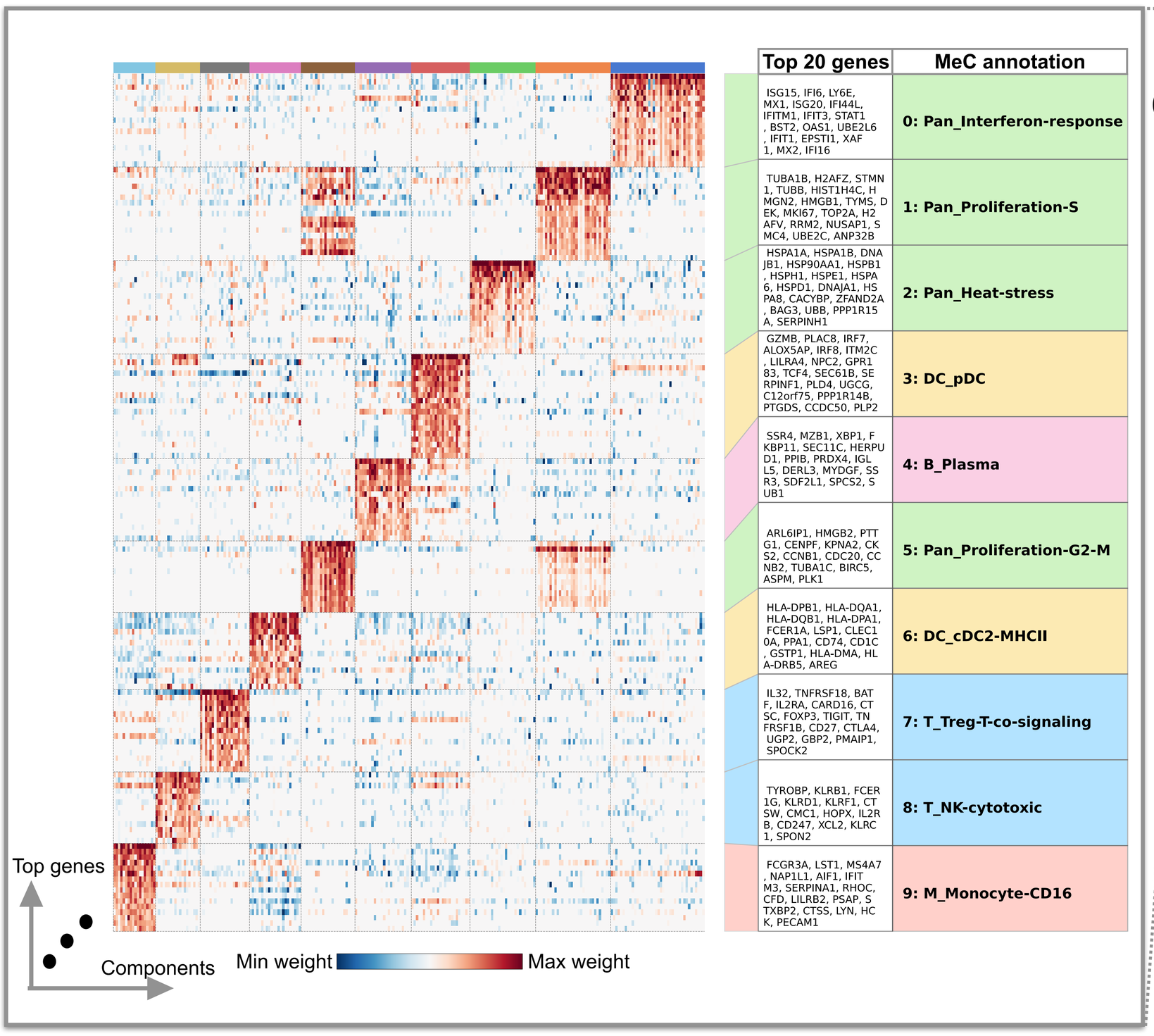
747 *Trends Immunol.* **43**, 210–229 (2022).
748 36. Finkin, S., Hartweger, H., Oliveira, T. Y., Kara, E. E. & Nussenzweig, M. C. Protein
749 Amounts of the MYC Transcription Factor Determine Germinal Center B Cell Division
750 Capacity. *Immunity* **51**, 324-336.e5 (2019).
751 37. Vaquero, J., Aoudjehane, L. & Fouassier, L. Cancer-associated fibroblasts in
752 cholangiocarcinoma. *Curr. Opin. Gastroenterol.* **36**, 63–69 (2020).
753 38. Luca, B. A. *et al.* Atlas of clinically distinct cell states and ecosystems across human solid
754 tumors. *Cell* **184**, 5482-5496.e28 (2021).
755 39. Zhao, J. *et al.* Detection of differentially abundant cell subpopulations in scrna-seq data.
756 *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
757 40. Yuen, K. C. *et al.* High systemic and tumor-associated IL-8 correlates with reduced
758 clinical benefit of PD-L1 blockade. *Nat. Med.* **26**, 693–698 (2020).
759 41. Kaplanov, I. *et al.* Blocking IL-1β reverses the immunosuppression in mouse breast
760 cancer and synergizes with anti-PD-1 for tumor abrogation. *Proc. Natl. Acad. Sci. U. S. A.*
761 **116**, 1361–1369 (2019).
762 42. Duan, Z. & Luo, Y. Targeting macrophages in cancer immunotherapy. *Signal Transduct.*
763 *Target. Ther. 2021 61* **6**, 1–21 (2021).
764 43. Bi, K. *et al.* Tumor and immune reprogramming during immunotherapy in advanced renal
765 cell carcinoma. *Cancer Cell* **39**, 649-661.e5 (2021).
766 44. Zhao, H. *et al.* The role of osteopontin in the progression of solid organ tumour. *Cell*
767 *Death Dis.* **9**, (2018).
768 45. Spivia, W., Magno, P. S., Le, P. & Fraser, D. A. Complement protein C1q promotes
769 macrophage anti-inflammatory M2-like polarization during the clearance of atherogenic
770 lipoproteins. *Inflamm. Res.* **63**, 885–893 (2014).
771 46. Liu, Q. *et al.* The CXCL8-CXCR1/2 pathways in cancer. *Cytokine Growth Factor Rev.* **31**,
772 61–71 (2016).

773 47. Qin, Q. *et al.* Lisa: inferring transcriptional regulators through integrative modeling of
774 public chromatin accessibility and ChIP-seq data. *Genome Biol.* (2020).
775 doi:10.1186/s13059-020-1934-6
776 48. Peters, J. M., Shah, Y. M. & Gonzalez, F. J. The role of peroxisome proliferator-activated
777 receptors in carcinogenesis and chemoprevention. *Nat. Rev. Cancer 2012 123* **12**, 181–
778 195 (2012).
779 49. Hu, H. *et al.* AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of
780 animal transcription factors. *Nucleic Acids Res.* **47**, D33–D38 (2019).
781 50. Li, H. *et al.* Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated
782 Compartment within Human Melanoma. *Cell* **176**, 775-789.e18 (2019).
783 51. Hyvärinen, A. & Oja, E. Independent component analysis: algorithms and applications.
784 *Neural Networks* **13**, 411–430 (2000).
785 52. Lotfollahi, M. *et al.* Mapping single-cell data to reference atlases by transfer learning. *Nat.*
786 *Biotechnol.* (2021). doi:10.1038/s41587-021-01001-7
787 53. Khan, O. *et al.* TOX transcriptionally and epigenetically programs CD8+ T cell
788 exhaustion. *Nat. 2019 5717764* **571**, 211–218 (2019).
789 54. Tang, Q. & Brown, M. A. Abstract A218: Modulating glucocorticoid receptor-mediated
790 signaling for enhancement of cancer immunotherapy. *Cancer Immunol. Res.* **7**, A218–
791 A218 (2019).
792 55. Ohandjo, A. Q. *et al.* Transcriptome Network Analysis Identifies CXCL13-CXCR5
793 Signaling Modules in the Prostate Tumor Immune Microenvironment. *Sci. Reports 2019*
794 *91* **9**, 1–13 (2019).
795 56. Havenar-Daughton, C. *et al.* CXCL13 is a plasma biomarker of germinal center activity.
796 *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2702–2707 (2016).
797 57. D'Agostino, R. B., Belanger, A. & D'Agostino, R. B. A Suggestion for Using Powerful and
798 Informative Tests of Normality. *Am. Stat.* **44**, 316 (1990).

799 58. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing
800 biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
801 59. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular
802 activity with single-cell RNA-Seq. *Elife* **8**, 1–26 (2019).
803


a
Tumor scRNA
• 1.7 million TME cells
ICA
Independent Components
• Signal sources e.g. cell states
Clustering & Integration
Meta-Components
• Cell type, cell state, signaling
Dataset 1
Dataset 2
...
Dataset n
b
Functional MeC Landscape for TME
• 86 Functional Annotated MeCs
• Categories
MeC | Top Genes
B_B | CD74,MS4A1,...
T_CD8T-CD3 | CD3D,CD8A,...
DC_pDC | GZMB,TCF4,...
M_Macrophage-IL1-NFKB | CXCL8,IL1B,...
Pan_Inteferon_response | STAT1, ISG15,...
B
T/NK
DC
Mono/Mac
OtherMyeloid
Stroma
Signaling
Cell State Annotation for Tumor scRNA
• MeC Signature Continuum
• Top Enriched State
Tumor Immunity Regulators
• Incorporated Epigenetic Data
• Prioritized TF regulators
TF
MeC | TF
DC_pDC | TCF4,IRF8
M_Macrophage-IL1-NFKB | NFKB
Pan_Inteferon_response | STAT1

a



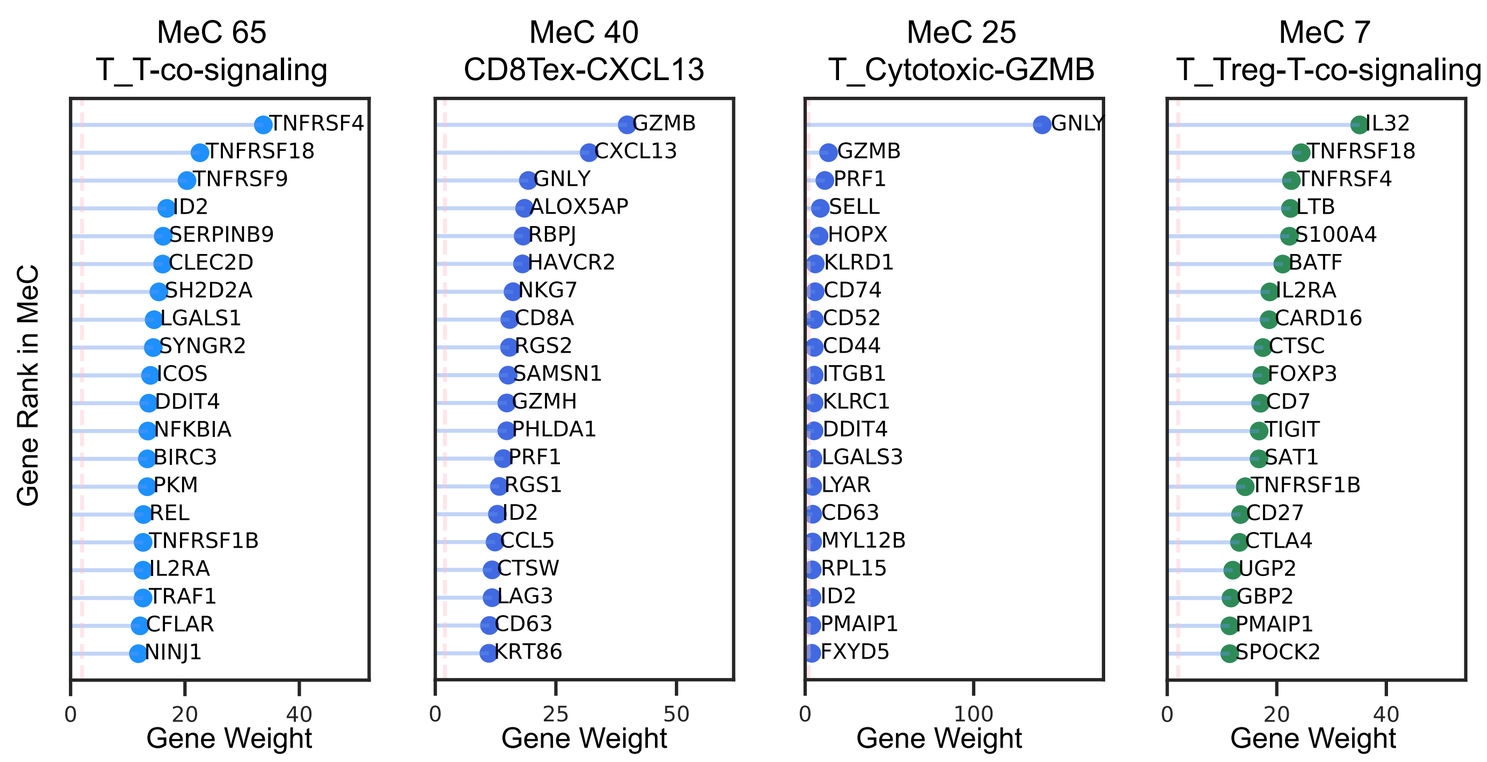
Top 20 genes	MeC annotation
ISG15, IFI6, LY6E, MX1, ISG20, IFI44L, IFITM1, IFIT3, STAT1, BST2, OAS1, UBE2L6, IFIT1, EPSTI1, XAF1, MX2, IFI16	0: Pan_Interferon-response
TUBA1B, H2AFZ, STMN1, TUBB, HIST1H4C, HMGN2, HMGB1, TYMS, DEK, MKI67, TOP2A, H2AFV, RRM2, NUSAP1, SMC4, UBE2C, ANP32B	1: Pan_Proliferation-S
HSPA1A, HSPA1B, DNAJB1, HSP90AA1, HSPB1, HSPH1, HSPE1, HSPA6, HSPD1, DNAJA1, HSPA8, CACYBP, ZFAND2A, BAG3, UBB, PPP1R15A, SERPINH1	2: Pan_Heat-stress
GZMB, PLAC8, IRF7, ALOX5AP, IRF8, ITM2C, LILRA4, NPC2, GPR183, TCF4, SEC61B, SERPINF1, PLD4, UGCG, C12orf75, PPP1R14B, PTGDS, CCDC50, PLP2	3: DC_pDC
SSR4, MZB1, XBP1, FKBP11, SEC11C, HERPUD1, PPIB, PRDX4, IGLL5, DERL3, MYDGF, SSR3, SDF2L1, SPCS2, SUB1	4: B_Plasma
ARL6IP1, HMGB2, PTTG1, CENPF, KPNA2, CKS2, CCNB1, CDC20, CCNB2, TUBA1C, BIRC5, ASPM, PLK1	5: Pan_Proliferation-G2-M
HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DPA1, FCER1A, LSP1, CLEC10A, PPA1, CD74, CD1C, GSTP1, HLA-DMA, HLA-DRB5, AREG	6: DC_cDC2-MHCII
IL32, TNFRSF18, BATF, IL2RA, CARD16, CTSC, FOXP3, TIGIT, TNFRSF1B, CD27, CTLA4, UGP2, GBP2, PMAIP1, SPOCK2	7: T_Treg-T-co-signaling
TYROBP, KLRB1, FCER1G, KLRD1, KLRF1, CTSW, CMC1, HOPX, IL2RB, CD247, XCL2, KLRC1, SPON2	8: T_NK-cytotoxic
FCGR3A, LST1, MS4A7, NAP1L1, AIF1, IFITM3, SERPINA1, RHOC, CFD, LILRB2, PSAP, STXBP2, CTSS, LYN, HCK, PECAM1	9: M_Monocyte-CD16

b

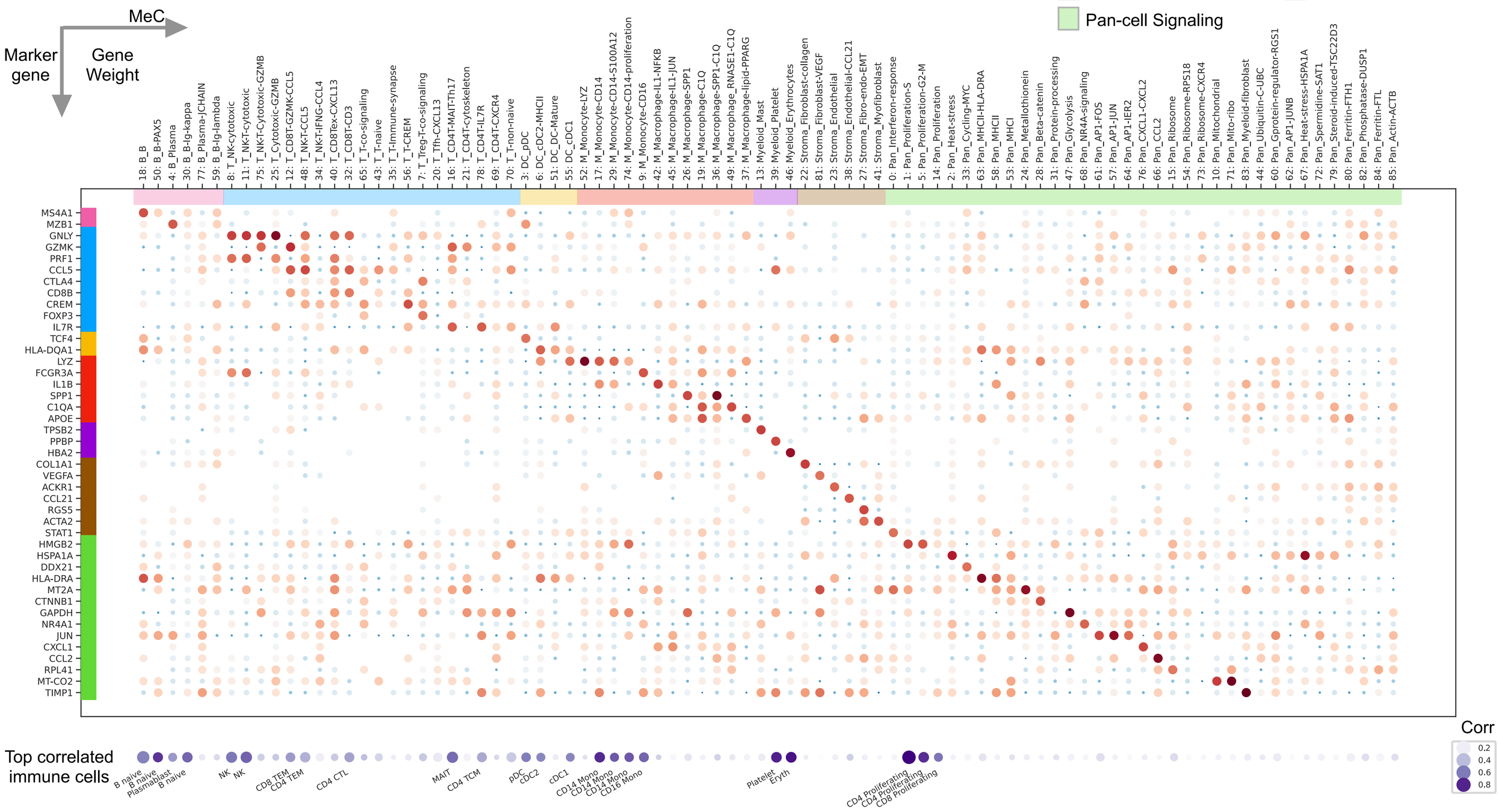
MeC annotation	Top 6 genes
0: Pan_Interferon-response	ISG15, IFI6, LY6E, MX1, ISG20, IFI44L
1: Pan_Proliferation-S	TUBA1B, H2AFZ, HMGB2, STMN1, TUBB, HIST1H4C
2: Pan_Heat-stress	HSPA1A, HSPA1B, DNAJB1, HSP90AA1, HSPB1, HSPH1
3: DC_pDC	GZMB, JCHAIN, PLAC8, IRF7, ALOX5AP, IRF8
4: B_Plasma	SSR4, MZB1, XBP1, FKBP11, JCHAIN, HSP90B1
5: Pan_Proliferation-G2-M	ARL6IP1, HMGB2, UBE2C, TUBB4B, PTTG1, CENPF
6: DC_cDC2-MHCII	CST3, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DPA1, FCER1A
7: T_Treg-T-co-signaling	IL32, TNFRSF18, TNFRSF4, LTB, S100A4, BATF
8: T_NK-cytotoxic	GNLY, TYROBP, KLRB1, FCER1G, CD7, KLRD1
9: M_Monocyte-CD16	FCGR3A, LST1, MS4A7, NAP1L1, AIF1, IFITM2
10: Pan_Mitochondrial	MT-CO3, MT-CO2, MT-CYB, MT-CO1, MT-ATP6, MT-ND3
11: T_NK-T-cytotoxic	GNLY, NKG7, FGFBP2, GZMB, GZMH, PRF1
12: T_CD8T-GZMK-CCL5	GZMK, CCL5, GZMA, DUSP2, CST7, CMC1
13: Myeloid_Mast	TPSAB1, TPSB2, CPA3, CLU, HPGDS, ANXA1
14: Pan_Proliferation	DUT, PCNA, TYMS, MCM7, TK1, RRM2
15: Pan_Ribosome	RPL41, RPL26, RPL23A, RPS12, RPS15A, RPS29
16: T_CD4T-MAIT-Th17	KLRB1, LTB, IL7R, GZMK, S100A4, RORA
17: M_Monocyte-CD14	S100A9, S100A8, LYZ, FCN1, VCAN, CTSS
18: B_B	CD74, CD79A, MS4A1, CD37, HLA-DRA, CD52
19: M_Macrophage-C1Q	C1QB, C1QA, APOE, C1QC, CTSD, APOC1
20: T_Tfh-CXCL13	CXCL13, NR3C1, ITM2A, RBPJ, SRGN, ALOX5AP
21: T_CD4T-cytoskeleton	ACTG1, COTL1, SUB1, ISG20, GMFG, GYPC
22: Stroma_Fibroblast-collagen	COL1A1, COL3A1, COL1A2, LUM, DCN, SPARC
23: Stroma_Endothelial	ACKR1, IFITM3, IGFBP7, AQP1, TM4SF1, IFI27
24: Pan_Metallothionein	MT2A, MT1X, MT1G, MT1E, MT1F, MT1H
25: T_Cytotoxic-GZMB	GNLY, GZMB, PRF1, SELL, HOPX, KLRD1
26: M_Macrophage-SPP1	SPP1, CSTB, PLIN2, LDHA, MIF, ENO1
27: Stroma_Fibro-endo-EMT	RGS5, NDUFA4L2, IGFBP7, ACTA2, CALD1, MYL9
28: Pan_Beta-catenin	HNRNPH1, CDC42, SET, C1orf56, CDC42SE1, CTNNB1
29: M_Monocyte-CD14-S100A12	S100A9, S100A8, S100A12, LYZ, NFKBIA, CD14
30: B_B-Ig-kappa	IGKC, IGHG1, IGHG4, IGHG3, IGHGP, IGHM
31: Pan_Protein-processing	HSPA5, HSP90B1, CALR, PDIA3, PDIA6, MANF
32: T_CD8T-CD3	GZMH, CD3D, CCL5, NKG7, CD52, FGFBP2
33: Pan_Cycling-MYC	HSP90AB1, NCL, DDX21, C1QBP, HSPE1, NPM1
34: T_NK-T-IFNG-CCL4	CCL4, CCL4L2, IFNG, CCL3, ZFP36L1, TNF
35: T_T-Immune-synapse	TAGLN2, ANXA1, CRIP1, TSC22D3, MYADM, UCP2
36: M_Macrophage-SPP1-C1Q	SPP1, RNASE1, CTSB, PLIN2, GPNMB, FCER1G
37: M_Macrophage-lipid-PPARG	APOC1, APOE, LGALS3, GPNMB, CTSD, CSTB
38: Stroma_Endothelial-CCL21	CCL21, TFPI, FABP5, TFF3, MMRN1, FABP4
39: Myeloid_Platelet	PPBP, PF4, TAGLN2, NAP1L1, OST4, GPX1
40: T_CD8Tex-CXCL13	GZMB, CXCL13, GNLY, ALOX5AP, RBPJ, HAVCR2
41: Stroma_Myofibroblast	TAGLN, ACTA2, ADIRF, MYL9, DSTN, NDUFA4
42: M_Macrophage-IL1-NFKB	CXCL3, CXCL2, IL1B, CXCL8, G0S2, CCL20
43: T_T-naive	CD7, ZFP36L2, TRGC2, HCST, ZNF683, LTB
44: Pan_Ubiquitin-C-UBC	UBC, DUSP2, HSPA6, HLA-C, RPLP0, DDIT4
45: M_Macrophage-IL1-JUN	CXCL8, CXCL2, CXCL3, CXCL1, NFKBIA, NFKBIZ
46: Myeloid_Erythrocytes	HBA2, HBB, HBA1, UBB, HBD, ALAS2
47: Pan_Glycolysis	GAPDH, TPI1, LDHA, ALDOA, AC006064.4, S100A10
48: T_NK-T-CCL5	CCL5, IL32, GNLY, GZMA, CD7, CD52
49: M_Macrophage_RNASE1-C1Q	RNASE1, SELENOP, C1QB, C1QA, C1QC, CTSB
50: B_B-PAX5	TCL1A, FCER2, CXCR4, IL4R, FOS, CD69
51: DC_DC-Mature	TXN, BIRC3, ID2, MARCKSL1, CCR7, CRIP1
52: M_Monocyte-LYZ	LYZ, AC020656.1, MMP12, IFI30, PSAP, FCN1
53: Pan_MHCI	CALR, MALAT1, HLA-A, TMBIM6, HLA-B, PPIB
54: Pan_Ribosome-RPS18	RPS18, RPS27, RPL13AP5, RPL31, RPL18A, RPL37A
55: DC_cDC1	RGCC, CST3, SNX3, IRF8, ID2, C1orf54
56: T_T-CREM	CREM, CXCR4, SYTL3, LEPROTL1, SRGN, FYN
57: Pan_AP1-JUN	JUN, IER2, FOS, ID2, GADD45B, KLF6
58: Pan_MHCII	ZFP36L1, LGALS2, PSME2, NPC2, HLA-DPB1, HLA-DRA
59: B_B-Ig-lambda	IGLC2, IGLC3, IGHGP, IGHG1, EMP3, CD37
60: Pan_Gprotein-regulator-RGS1	RGS1, JUN, S100A4, GNLY, RPL17, RPS18
61: Pan_AP1-FOS	FOS, DUSP1, JUNB, JUN, KLF6, TSC22D3
62: Pan_AP1-JUNB	JUNB, ZFP36, HSPA1B, IER2, DNAJB1, RPS15A
63: Pan_MHCII-HLA-DRA	HLA-DRA, HLA-DRB1, HLA-DPA1, HLA-DQA1, IFI30, HLA-DRB5
64: Pan_AP1-IER2	IER2, EGR1, EIF4A3, FOSB, SERTAD1, TUBB4B
65: T_T-co-signaling	TNFRSF4, TNFRSF18, TNFRSF9, ID2, SERPINB9, CLEC2D
66: Pan_CCL2	CCL2, EIF4A3, DBI, TUBA1B, MT2A, CYR61
67: Pan_Heat-stress-HSPA1A	HSPA1A, HSPA1B, DDIT4, SAT1, DNAJB1, SON
68: Pan_NR4A-signaling	NFKBIA, CD69, SDCBP, ZFP36L1, NR4A1, DUSP2
69: T_CD4T-CXCR4	CXCR4, IL32, HERPUD1, TSC22D3, ZFP36L2, GAPDH
70: T_T-non-naive	S100A4, KLF6, KLRB1, FYN, GPR183, S100A11
71: Pan_Mito-ribo	MT-CO2, TPT1, MT-ND1, MT-ND3, EEF1A1, MT-ND2
72: Pan_Spermidine-SAT1	SAT1, BTG1, ZFP36L2, RPS4X, EEF1A1, RPLP2
73: Pan_Ribosome-CXCR4	RPL27A, MT-ATP8, RPS9, RPL3, RPS18, RPL27
74: M_Monocyte-CD14-proliferation	S100A8, S100A9, S100A12, HMGB2, VCAN, SELL
75: T_NK-T-Cytotoxic-GZMB	GNLY, IFITM2, SELL, CMC1, COTL1, CD7
76: Pan_CXCL1-CXCL2	RGS2, CXCL1, LAPTM5, P2RY12, ID2, CXCR4
77: B_Plasma-JCHAIN	JCHAIN, IGHA1, VIM, CD48, IER2, PABPC4
78: T_CD4T-IL7R	S100A4, ANXA1, LTB, ITGB1, IL7R, S100A11
79: Pan_Steroid-induced-TSC22D3	TSC22D3, TNFAIP3, HSPA8, RPL12, SGK1, KLF6
80: Pan_Ferritin-FTH1	FTH1, HLA-E, CCL5, APOE, RPL5, PFN1
81: Stroma_Fibroblast-VEGF	MT2A, CPE, VEGFA, ENO1, ZFAS1, TIMP1
82: Pan_Phosphatase-DUSP1	DUSP1, RPS4X, RPS27, GNLY, SGK1, ZFP36L2
83: Pan_Myeloid-fibroblast	TIMP1, DDIT4, MT2A, NNMT, SERPINE1, IL1B
84: Pan_Ferritin-FTL	FTL, EEF1A1, SPARC, PTMA, EIF1, RPLP1
85: Pan_Actin-ACTB	ACTB, RPS19, TPT1, MALAT1, EEF1A1, H3F3B

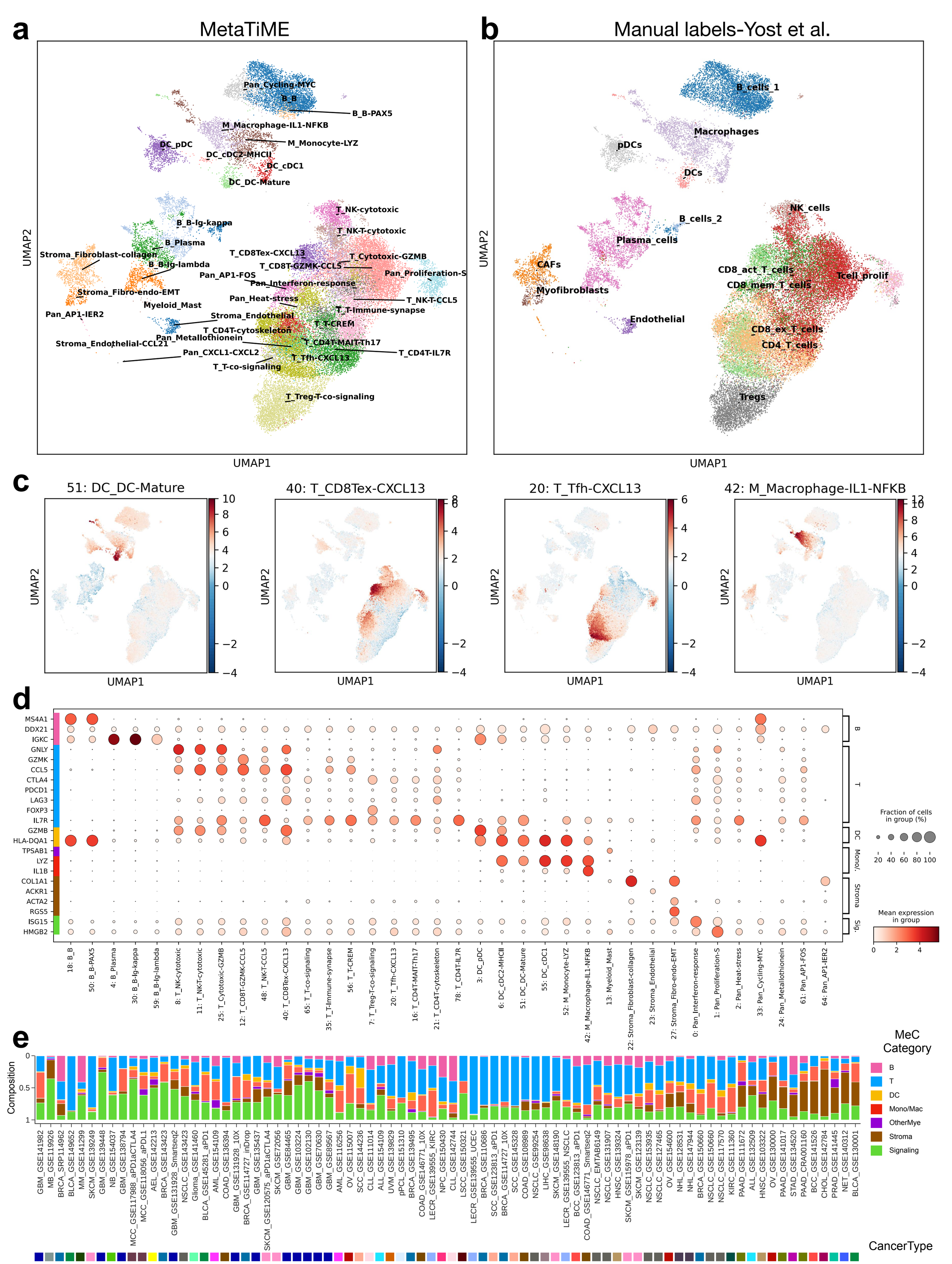
B cells
T/NK cells
Dendritic cells
Pan-cell Signaling
Monocytes/macrophages
Other myeloid cells
Stroma cells

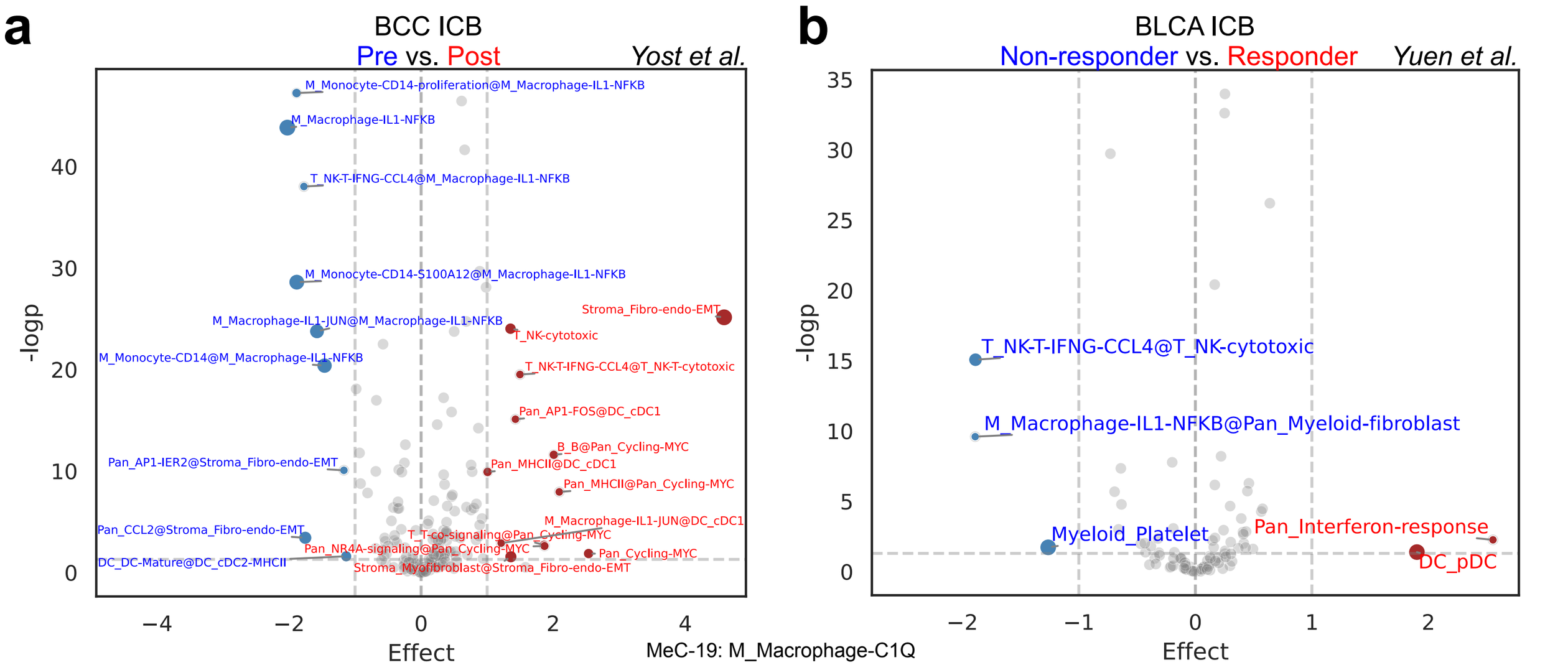
c

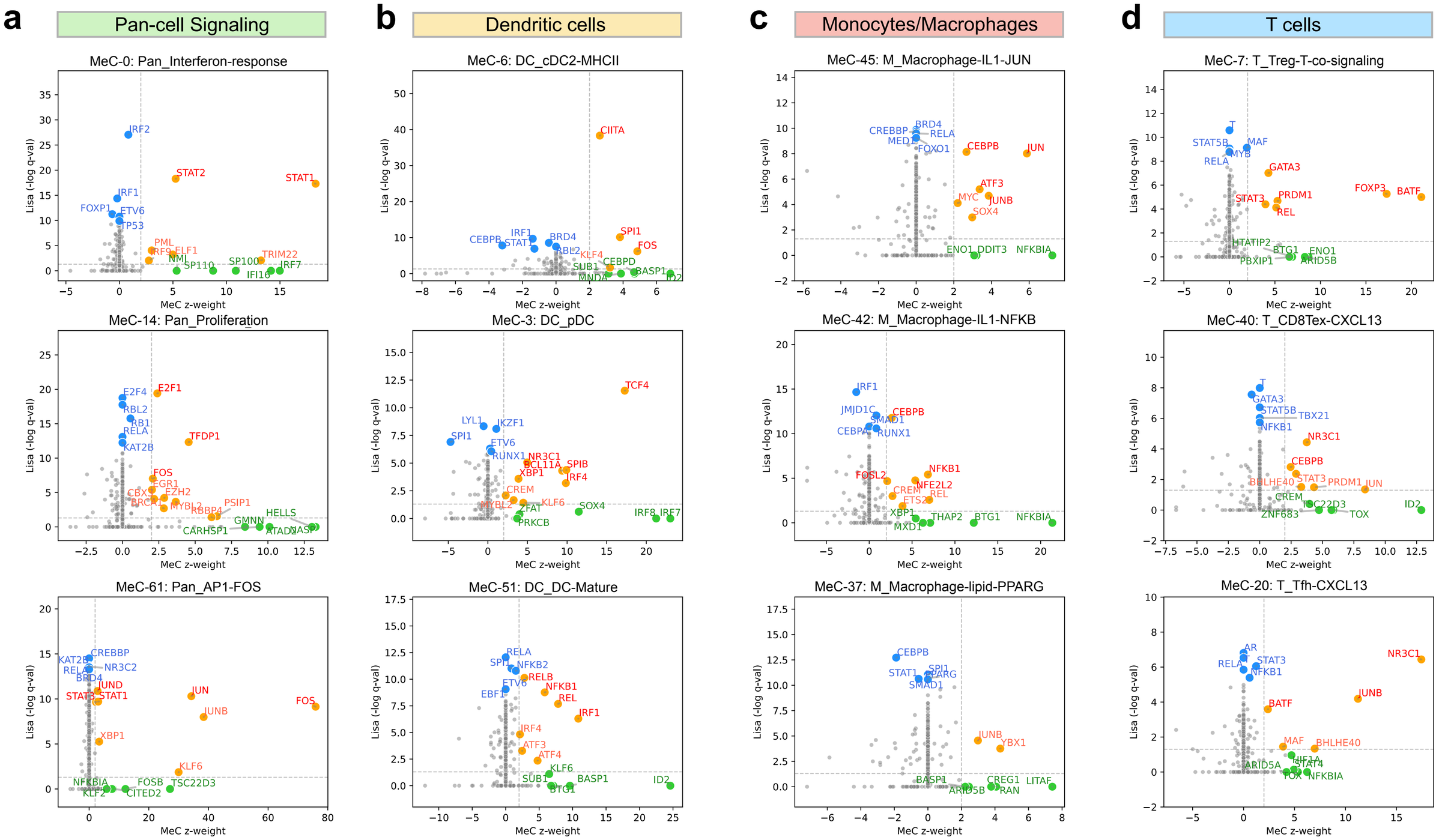


d




a
MetaTiME
Pan_Cycling-MYC
B_B
B_B-PAX5
M_Macrophage-IL1-NFKB
M_Monocyte-LYZ
DC_pDC
DC_cDC2-MHCII
DC_cDC1
DC_DC-Mature
B_B-Ig-kappa
B_Plasma
Stroma_Fibroblast-collagen
B_B-Ig-lambda
Stroma_Fibro-endo-EMT
Pan_AP1-IER2
Myeloid_Mast
T_NK-cytotoxic
T_NK-T-cytotoxic
T_CD8Tex-CXCL13
T_Cytotoxic-GZMB
T_CD8T-GZMK-CCL5
Pan_AP1-FOS
Pan_Proliferation-S
Pan_Interferon-response
Pan_Heat-stress
T_NK-T-CCL5
T_T-Immune-synapse
Stroma_Endothelial
T_T-CREM
T_CD4T-cytoskeleton
Pan_Metallothionein
Stroma_Endothelial-CCL21
T_CD4T-MAIT-Th17
Pan_CXCL1-CXCL2
T_CD4T-IL7R
T_Tfh-CXCL13
T_T-co-signaling
T_Treg-T-co-signaling
UMAP2
UMAP1
b
Manual labels-Yost et al.
B_cells_1
Macrophages
pDCs
DCs
B_cells_2
NK_cells
Plasma_cells
CAFs
CD8_act_T_cells
Tcell_prolif
CD8_mem_T_cells
Myofibroblasts
Endothelial
CD8_ex_T_cells
CD4_T_cells
Tregs
UMAP2
UMAP1
c
51: DC_DC-Mature
40: T_CD8Tex-CXCL13
20: T_Tfh-CXCL13
42: M_Macrophage-IL1-NFKB
UMAP2
UMAP1
d
MS4A1
DDX21
IGKC
GNLY
GZMK
CCL5
CTLA4
PDCD1
LAG3
FOXP3
IL7R
GZMB
HLA-DQA1
TPSAB1
LYZ
IL1B
COL1A1
ACKR1
ACTA2
RGS5
ISG15
HMGB2
B
T
DC
Mono/.
Stroma
Sig.
18: B_B
50: B_B-PAX5
4: B_Plasma
30: B_B-Ig-kappa
59: B_B-Ig-lambda
8: T_NK-cytotoxic
11: T_NK-T-cytotoxic
25: T_Cytotoxic-GZMB
12: T_CD8T-GZMK-CCL5
48: T_NK-T-CCL5
40: T_CD8Tex-CXCL13
65: T_T-co-signaling
35: T_T-Immune-synapse
56: T_T-CREM
7: T_Treg-T-co-signaling
20: T_Tfh-CXCL13
16: T_CD4T-MAIT-Th17
21: T_CD4T-cytoskeleton
78: T_CD4T-IL7R
3: DC_pDC
6: DC_cDC2-MHCII
51: DC_DC-Mature
55: DC_cDC1
52: M_Macrophage-LYZ
42: M_Macrophage-IL1-NFKB
13: Myeloid_Mast
22: Stroma_Fibroblast-collagen
23: Stroma_Endothelial
27: Stroma_Fibro-endo-EMT
0: Pan_Interferon-response
1: Pan_Proliferation-S
2: Pan_Heat-stress
33: Pan_Cycling-MYC
24: Pan_Metallothionein
61: Pan_AP1-FOS
64: Pan_AP1-IER2
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2 4
e
Composition
0
0.5
1
GBM_GSE141982
MB_GSE119926
BRCA_SRP114962
BLCA_GSE149652
MM_GSE141299
SKCM_GSE139249
GBM_GSE139448
NB_GSE154037
GBM_GSE138794
MCC_GSE117988_aPD1aCTLA4
MCC_GSE118056_aPDL1
AEL_GSE142213
BRCA_GSE143423
GBM_GSE131928_Smartseq2
NSCLC_GSE143423
Glioma_GSE141460
BLCA_GSE145281_aPD1
AML_GSE154109
COAD_GSE136394
GBM_GSE131928_10X
BRCA_GSE114727_inDrop
GBM_GSE135437
SKCM_GSE120575_aPD1aCTLA4
SKCM_GSE72056
GBM_GSE84465
GBM_GSE103224
GBM_GSE102130
GBM_GSE70630
GBM_GSE89567
AML_GSE116256
OV_GSE115007
SCC_GSE144236
CLL_GSE111014
ALL_GSE154109
UVM_GSE139829
pPCL_GSE151310
BRCA_GSE139495
COAD_GSE146771_10X
LECR_GSE139555_KIRC
NPC_GSE150430
CLL_GSE142744
LSCC_GSE150321
LECR_GSE139555_UCEC
BRCA_GSE110686
SCC_GSE123813_aPD1
BRCA_GSE114727_10X
SCC_GSE145328
COAD_GSE108989
NSCLC_GSE99254
LIHC_GSE98638
SKCM_GSE148190
LECR_GSE139555_NSCLC
BCC_GSE123813_aPD1
COAD_GSE146771_Smartseq2
NSCLC_EMTAB6149
NSCLC_GSE131907
HNSC_GSE139324
SKCM_GSE115978_aPD1
SKCM_GSE123139
NSCLC_GSE153935
NSCLC_GSE127465
OV_GSE154600
NHL_GSE128531
NHL_GSE147944
BRCA_GSE150660
NSCLC_GSE150660
NSCLC_GSE117570
KIRC_GSE111360
PAAD_GSE111672
ALL_GSE132509
HNSC_GSE103322
OV_GSE130000
PAAD_GSE141017
STAD_GSE134520
PAAD_CRA001160
BCC_GSE141526
CHOL_GSE142784
PRAD_GSE141445
NET_GSE140312
BLCA_GSE130001
MeC Category
B
T
DC
Mono/Mac
OtherMye
Stroma
Signaling
CancerType


a
BCC ICB
Pre vs. Post
Yost et al.
M_Monocyte-CD14-proliferation@M_Macrophage-IL1-NFKB
M_Macrophage-IL1-NFKB
T_NK-T-IFNG-CCL4@M_Macrophage-IL1-NFKB
M_Monocyte-CD14-S100A12@M_Macrophage-IL1-NFKB
Stroma_Fibro-endo-EMT
M_Macrophage-IL1-JUN@M_Macrophage-IL1-NFKB
T_NK-cytotoxic
M_Monocyte-CD14@M_Macrophage-IL1-NFKB
T_NK-T-IFNG-CCL4@T_NK-T-cytotoxic
Pan_AP1-FOS@DC_cDC1
B_B@Pan_Cycling-MYC
Pan_AP1-IER2@Stroma_Fibro-endo-EMT
Pan_MHCII@DC_cDC1
Pan_MHCII@Pan_Cycling-MYC
M_Macrophage-IL1-JUN@DC_cDC1
Pan_CCL2@Stroma_Fibro-endo-EMT
T_T-co-signaling@Pan_Cycling-MYC
Pan_NR4A-signaling@Pan_Cycling-MYC
Pan_Cycling-MYC
DC_DC-Mature@DC_cDC2-MHCII
Stroma_Myofibroblast@Stroma_Fibro-endo-EMT
-logp
Effect
b
BLCA ICB
Non-responder vs. Responder
Yuen et al.
T_NK-T-IFNG-CCL4@T_NK-cytotoxic
M_Macrophage-IL1-NFKB@Pan_Myeloid-fibroblast
Myeloid_Platelet
Pan_Interferon-response
DC_pDC
-logp
Effect
c
Mono/Mac states in TME
CD14
CD16
Monocytes
Macrophages
SPP1+
Glycolysis
Hypoxia
C1Q+
Complements
Phagocytic
IL1B+
IL1B pathway
NFkB pathway
Lipid rich
Cholesterol
PPAR
ISG+
Interferon response
MHCII+
Antigen presentation
Pro-cancer
Anti-cancer
d
MeC-26: M_Macrophage-SPP1
Glycolysis / Gluconeogenesis
HIF-1 signaling pathway
Central carbon metabolism in cancer
Ferroptosis
Vibrio cholerae infection
Lysosome
MeC-42: M_Macrophage-IL1-NFKB
Rheumatoid arthritis
TNF signaling pathway
NF-kappa B signaling pathway
Lipid and atherosclerosis
IL-17 signaling pathway
Viral protein interaction with cytokine and cytokine receptor
MeC-19: M_Macrophage-C1Q
Lysosome
Phagosome
Antigen processing and presentation
Staphylococcus aureus infection
Tuberculosis
Hematopoietic cell lineage
MeC-37: M_Macrophage-lipid-PPARG
Lysosome
Cholesterol metabolism
Tuberculosis
PPAR signaling pathway
Antigen processing and presentation
Apoptosis
KEGG_Human_2021
-Log Adjusted P-value


a
Pan-cell Signaling
MeC-0: Pan_Interferon-response
MeC-14: Pan_Proliferation
MeC-61: Pan_AP1-FOS
b
Dendritic cells
MeC-6: DC_cDC2-MHCII
MeC-3: DC_pDC
MeC-51: DC_DC-Mature
c
Monocytes/Macrophages
MeC-45: M_Macrophage-IL1-JUN
MeC-42: M_Macrophage-IL1-NFKB
MeC-37: M_Macrophage-lipid-PPARG
d
T cells
MeC-7: T_Treg-T-co-signaling
MeC-40: T_CD8Tex-CXCL13
MeC-20: T_Tfh-CXCL13
Lisa (-log q-val)
MeC z-weight