Title: A multi-tissue single-cell tumor microenvironment atlas reveals myeloid-derived cell states with significant impact on clinical outcome

Authors: Giovanna Resk Maklouf [1]†, Gabriela Rapozo Guimarães[1]†, Cristiane Esteves Teixeira[1], Marco Antônio Pretti[1], Leandro de Oliveira Santos[1], Nayara Gusmão Tessarollo[1], Nayara Evelin Toledo[1], Marcelo Falchetti[2], Mylla M. Dimas[2], Alessandra Freitas Serain[1], Nina Carrossini Bastos[3], Fabiane Carvalho de Macedo[3], Fabiana Resende Rodrigues[3], Jesse Lopes da Silva[4], Edroaldo Lummertz da Rocha[2], Cláudia Bessa Pereira Chaves[4,5], Andreia Cristina de Melo[4], Pedro Manoel Mendes Moraes-Vieira[6], Marcelo A. Mori[7,8], Mariana Boroni[1*]

Affiliations:
[1]Bioinformatics and Computational Biology Lab, Division of Experimental and Translational Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil.

[2]Department of Microbiology, Immunology, and Parasitology, Federal University of Santa Catarina, Santa Catarina, Brazil.

[3]Pathology Department, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil.

[4]Division of Clinical Research and Technological Development, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil.

[5]Gynecology Oncology Section, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil.

[6]Laboratory of Immunometabolism, Department of Genetics, Evolution, Microbiology, and Immunology, Institute of Biology, University of Campinas, Campinas, Brazil.

[7]Laboratory of Aging Biology, Department of Biochemistry and Tissue Biology, University of Campinas, Campinas, Brazil.

[8]Obesity and Comorbidities Research Center (OCRC), University of Campinas, Campinas, Brazil.

*Corresponding author: Mariana Boroni - mariana.boroni@inca.gov.br

†These authors contributed equally to this work.

One Sentence Summary: Myeloid-derived cell atlas exhibit heterogeneous phenotypes in the tumor microenvironment strongly linked with clinical outcomes in cancer.

Abstract: Immunotherapies provide long-lasting responses across a wide range of cancers; however, only a fraction of patients responds to them. Multiple reasons contribute to the failure, including the existence of myeloid-derived cells (MDCs) within tumors. Due to their high plasticity, these cells display numerous cell states and, as a result, distinct pro-tumorigenic behaviors, making them interesting targets. However, the creation of cutting-edge anticancer therapies is hampered by the lack of MDC-specific markers. To fully define the MDC landscape in solid tumors, we combined single-cell RNA-Seq from 13 public datasets, including samples from seven different cancers and normal samples, yielding the largest collection of MDC subpopulations within the tumor microenvironment. We identified five major lineages subdivided into one mast cell cluster, three neutrophils, eight dendritic cells, six monocytes, and

eleven macrophage states. Transcriptional profiles coupled with deconvolution estimates of cell populations in large cohorts revealed five MDC subpopulations as independent prognostic markers in different cancer types, including resident tissue interstitial macrophages and *FCGR3A*$^+$ monocytes associated with an unfavorable clinical outcome in ovarian and breast cancer patients, respectively. Our work reveals that *TREM2*$^+$ macrophages can be distinguished in different populations and associated with distinct prognoses. By analyzing a Brazilian cohort of ovarian cancer, we found that TREM2$^+$ macrophages are associated with a better prognosis, indicating that their role might be dependent on the tumor niche and co-expression of immunosuppressive markers. Collectively, this atlas reveals in high-resolution the heterogeneous MDC identity as well as new avenues for understanding and manipulating their fate in cancer.

Main Text:

INTRODUCTION

The tumor microenvironment (TME) is composed of recruited and resident cells as well as acellular components that interact with malignant cells to form a complex network. TME resident immune cells, which are derived from myeloid or lymphoid precursors, can either suppress or promote tumor growth by enabling and sustaining hallmarks related to cancer progression *(1)*. The comprehensive characterization of immune cells in the TME, particularly lymphocytes, has enabled the development of immunotherapies that elicit long-lasting responses across a wide range of cancers *(2)*. However, only a subset of individuals benefits from immune-checkpoint inhibitor therapies *(3)*. Immunosuppressive cells within tumors, including myeloid-derived cells (MDCs) like monocytes, macrophages, conventional dendritic cells (cDC), and polymorphonuclear granulocytes, have been recognized as one of the reasons why cancer therapies fail, among numerous others *(4–6)*.

MDCs are a highly diverse and plastic population that can exhibit many cell states and thus perform a broad array of functions in the niche in which they are found *(7)*. In terms of ontogeny, the majority of these cells are derived from bone marrow progenitor cells and recruited into the TME. Macrophages can be derived from two main sources: 1) from erythro-myeloid progenitors in the yolk sac and fetal liver before birth - these macrophages have enriched expression of the markers *FOLR2, PLTP,* and *LYVE1 (8, 9)*, and 2) from circulating monocytes derived from the bone marrow and recruited into the tumor tissue *(10)*. Recruited macrophages have a short lifespan and require constant replenishment from circulating monocytes. This cell population expresses monocyte markers such as *FN1*, *SELL*, and *VCAN* as a proxy for their ontogeny *(11)*.

Regardless of their origin, macrophages may manifest a constellation of phenotypes. Based primarily on *in vitro* experiments *(12)*, macrophages have been classified into two main polarization states: classically activated M1, which includes cells with a proinflammatory phenotype, and alternatively activated M2, which includes cells with tissue remodeling and/or anti-inflammatory properties *(10)*. In the TME context, M2 macrophages are commonly referred to as tumor-associated macrophages (TAMs), and their presence is associated with a poor prognosis in several tumor types *(13–15)*. However, not all TAMs exhibit a clear M2-like

phenotype, and they often express markers of both activation states *(16, 17)*, emphasizing the necessity of distinguishing TAM states beyond the M1/M2 dichotomy.

TAMs have been linked to a plethora of functions in the TME. These roles have been associated with tumor growth, epithelial-mesenchymal plasticity, extracellular matrix remodeling, cell invasion, and angiogenesis, which in turn influence tumor progression, metastasis, and therapy resistance. In addition, the crosstalk between TAMs and malignant cells can promote immune evasion via TME shaping *(18)*. Since TAMs and other MDCs are key in promoting tumor growth and metastasis, several therapeutic strategies to deplete TAMs or modulate their functional or phenotypic reprogramming, infiltration, or activation are emerging *(19)*. These therapies, however, are limited by the lack of specific markers to distinguish MDC subpopulations, which highlights the urgent need for an in-depth understanding of TME and determination of the subtypes of tumor-associated cells, which will allow the selective targeting of these abnormally expanded MDC populations *(19)*.

Although recent studies have used high-resolution technologies such as single-cell RNA sequencing (scRNA-Seq) and spatial transcriptomics to characterize the immune landscape of various cancer types, no pan-cancer integrated data allowed a systematic comparison across cancer types to reveal the similarities and differences of MDCs in the TME. To address this issue, we have built the most comprehensive pan-cancer catalog of TME subpopulations across eight different solid tumor types by combining publicly accessible scRNA-Seq data from three different techniques. The high-resolution characterization of tumor-associated cell types allowed the identification of MDC subpopulations associated with a poor prognosis in tumors of different origins, as well as the determination of molecules that allow the specific targeting of these tumor-associated cells. Our thorough identification and characterization of MDC landscapes in the TME will pave the path for the development of effective immunotherapy strategies targeting specific populations of these cells.

RESULTS

Thorough integration strategy recovers a comprehensive pan-cancer atlas of the tumor microenvironment

To define a comprehensive landscape of widespread and rare cell subtypes in the TME**,** we have integrated a large amount of scRNA-Seq data comprising 392,204 high-quality cells from 13 public datasets (Data file S1, Fig. S1A). The analyzed data was generated utilizing three different technologies (10x Genomics, InDrop, and Smart-Seq2, Fig. S1B) and samples from different anatomical sites of healthy donors and patients diagnosed with different cancer types, totaling 142 individuals (Data file S2). We developed a robust analytical pipeline with stringent quality control to integrate a vast and heterogeneous dataset, retaining crucial biological variation while removing typical contaminants like doublets and ambient RNA, as well as batch corrections (Fig. 1A). When compared to previous strategies that evaluated small and isolated datasets, our approach improved the ability to detect and define common and unusual cell states. The dataset encompassed seven tumors, including breast, colorectal, liver, lung, ovary, skin, and uveal melanoma, as well as their adjacent tissue counterparts; blood samples from patients with two different tumor types and one healthy donor; and adjacent lymph nodes from patients with breast tumors (Fig. 1B).

Through our approach, we identified highly homogeneous and robust cell groups (Fig. S1C-D), resulting in a precise map of cells distributed throughout clusters. The clusters were then annotated based on canonical gene markers (Data file S5) of major cell types and differentially expressed genes (DEGs), yielding 11 broadly recognized cell types (Fig. 1C). We also annotated low-quality cells as not identifiable (Fig. S1C). T and Natural Killer (NK) cells (n = 119,472), which expressed *CD3E*, *CD4* or *CD8A,* and *NKG7,* were the most abundant cell type, followed by MDC subpopulations, such as mononuclear phagocytes (n = 51,687), which were positive for *LYZ*, *AIF1,* and *CD68*; neutrophils (n = 12,033), marked by *CXCR2* and *FCGR3B*; mast cells (n = 4,539), identified by *TPSAB1* and *CPA3* expression; and megakaryocytes (n = 185), the smallest group, marked by *PPBP* expression. B cells (n = 22,102) were identified by the expression of *MS4A1* and *MZB1*; endothelial cells (n = 28,775) were marked by *VWF* and *CCL21*; fibroblasts (n = 48,159) by *COL1A1* and *COL1A2*, and plasmacytoid dendritic cells (pDC) (n = 1,365) by *LILRA4* and *IRF7* (Fig. 1D). *EPCAM* and *CDH1* expression allowed the identification of normal (n = 56,060) and malignant (n = 47,827) epithelial cells, which were further distinguished by the amount of copy number variation events observed in the malignant cells (Fig. S1E). When comparing tumor samples to normal samples, we found that tumors had a higher percentage of lymphocytes and mononuclear phagocytes. Except for lung metastasis samples, the proportion of mononuclear phagocytes was constant across tissue sources (Fig. 1E).

To characterize the subsets of MDCs, we analyzed each cluster independently. We performed unsupervised clustering on MDCs and identified five common major lineages (Fig. 1F): mast cells (n = 4,539), neutrophils (n = 12,033), cDCs (n = 6,212), monocytes (n = 10,321), and macrophages (n = 38,517) based on canonical gene markers (Figure 1G; Data file S5). MDCs were separated into 29 subpopulations (Fig. 1H).

Myeloid cells are heterogeneous and phenotypically diverse across the tumor samples

Mast cells, the less frequent MDC subpopulation, were identified as having a single state (Data file S3, Fig. S2A), equally distributed in tumor and normal samples (Fig. S2B). However, in contrast to normal tissues, tumor-associated mast cells displayed higher expression of genes related to PD-1 signaling (Fig. S2C). Neutrophils were subdivided into three functionally distinct subpopulations (Fig. S2D-E): the Neutrophil_TAGLN2 (n = 6,570) population expressed numerous genes involved in the interferon signaling pathway (Fig. S2F), reflecting the probable pro-inflammatory role of these cells; the Neutrophil_MMP9 (n = 4,024) population was enriched in pathways related to signaling processes by *VEGF* (Fig. S2F); and the Neutrophil_CXCL8 (n = 1,439) population was linked to anti-inflammatory pathways such as IL-4 and IL-13 signaling (Fig. S2F) - this population was found to be more prevalent in tumors (lung and breast - generated by InDrop technique) than in blood or normal tissues (Fig. S2G-H).

cDCs were the third most common subpopulation of mononuclear phagocytes, accounting for 12% of these cells. We identified eight cDC states (Fig. 2A-B): cDC type 1 (cDC1), expressing *CLEC9A* and *CADM1,* denoted cDC1_CLEC9A (n = 580); and six subpopulations of cDC type 2 (cDC2), marked by *CD1C* and *CLEC10A* expression. cDC2 subpopulation was separated into types A, resembling an anti-inflammatory profile, and B, a pro-inflammatory profile, named cDC2A_AREG (n = 2,044) and cDC2B_FCER1A (n = 726). These populations were previously described by Brown et al. (2020) (19) (Fig. S3A, top). In addition,

the cDC2A_AREG was enriched in tumors and lymph nodes compared to normal tissue, whereas the cDC2B_FCER1A was found mostly in the blood (Fig. 2C and S3B). Other subpopulations found predominantly in tumor samples included cDC2_CD14 (n = 765) and cDC2_FCGR3A (n = 314), both of which are likely monocyte-derived (Fig. 2C and S3B).

Another cDC2 subpopulation was identified by the expression of a set of chemokines (*CXCL1*, *CXCL8,* and *CXCL14*), named cDC2_CXCL8 (n = 993) (Fig. 2B). We also identified a subset of CD1C$^+$ cDC2 (cDC_CD207, n = 505) which expressed *CD1A* and *CD207*, the latter being a langerin protein encoder. Although langerin is primarily expressed by epidermal macrophages known as Langerhans cells *(20, 21)*, cDC_CD207 is a type 2 DC subpopulation distinct from the aforesaid cell (Fig. 2B and S3A, top). This subpopulation was more prevalent in tumors (lung, ovary, and lymph nodes) than in normal samples (Fig. S3B). The transcriptional profile of the Langerhans-like cells closely resembled that of cDC rather than macrophages. We also identified a mature cDC subpopulation named cDC_LAMP3 (n = 285), which was distinguished by the expression of *LAMP3* and *CCR7* genes (Fig. 2B and S3A, top). The expression of genes coding for PD-1 and PD-2 ligands was found to be increased in mature cDC_LAMP3, indicating that this population may have an immunosuppressive role (Fig. S3A, bottom). The distribution of dendritic cells across cancer types and samples was consistent, except for the lymph node, where an increase in the frequency of the cDC2A_AREG and cDC_LAMP3 subpopulations was observed (Fig. 2C). Functional enrichment analysis also confirmed these profiles (Fig. S3C).

We found that the populations cDC2_FCGR3A and cDC2A_AREG were enriched with pathways associated with an immunosuppressive role, such as PD1-signaling and interleukin-4 and interleukin-13 signaling. In contrast, the cDC_LAMP3, cDC1_CLEC9A, cDC2_CXCL8, and cDC_14 were enriched with pathways related to interferon-α/β and -γ signaling and antigen presentation consistent with a pro-inflammatory profile (Fig. S3C). The cDC_CD207 presented both anti- and pro-inflammatory pathways.

Monocytes are important precursors of tumor-infiltrating myeloid cells. A total of 10,321 monocytes were sub-clustered, and six major groups were identified (Fig. 2D). Mono_FCG3RA (n = 2,018) represents the non-classical phenotype that expressed *FCGR3A, MS4A7, ITGAL,* and *SIGLEC10* in a single cluster (Fig. 2E) and was absent in lymph nodes (Fig. 2F and S3D). The classical phenotype was characterized by the expression of common monocyte markers such as *CD14, SELL,* and *FCN1,* and was further distinguished by the expression of *FOS* into Mono_CD14_FOS$^-$ and Mono_CD14_FOS$^+$ subpopulations (Fig. 2E). When compared to other tissues, the Mono_CD14_FOS$^+$ (n = 1,497) population was enriched in lymph nodes and normal tissues (ovary and lung) (Fig. 2F and S3D). In contrast, Mono_CD14_FOS$^-$ (n = 2,150) was more prevalent in blood samples; the specific role of these subpopulations remains unknown (Fig. 2F and S3D). Additionally, a monocyte subpopulation with a pro-inflammatory profile was named Mono_IL1B (n = 1,739). This subpopulation expresses high levels of *IL1B*, *CXCL2,* and *NFKBIA* (Fig. S3E), as well as other inflammatory genes when compared to other monocytes (Fig. S3F). Pathway enrichment analysis showed this profile has increased levels of genes involved with Toll-like receptor cascades (Fig. S3G).

At this resolution, we were also able to identify two monocyte subpopulations displaying an intermediate phenotype, which we characterized as being between monocytes and antigen-presenting cells, expressing genes coding for CD16 and CD14. We called this population Mono_Inter, from which CXCL10 and CLEC10A expression distinguished subpopulations of

MonoInter_CXCL10 and MonoInter_CLEC10A cells. These subpopulations express several markers related to antigen presentation, such as *HLA-DPB1, HLA-DMA*, *HLA-DPA1,* and *HLA-DRA*, in addition to *CD163,* a marker of monocyte and macrophage activation, and *MARCO,* a scavenger receptor class A protein expressed on macrophage*s* (Fig. 2E). Functional enrichment analysis revealed the MHC class II antigen presentation pathway was enriched in this intermediate monocyte phenotype (Fig. S3G).

Macrophages are the largest subpopulation of mononuclear phagocytes found in this study, corresponding to 74.5% of all myeloid cells. By exploring the expression of markers regarding their ontogeny, macrophages were grouped into three main clusters: 1) recruited, which consists of cells expressing monocyte markers, such as *VCAN*, *S100A8*, *FCN1*, and *CD300E;* 2) resident-tissue macrophages (RTM), distinguished by the expression of *FOLR2*, *LYVE1*, and *PLTP;* and 3) RTM-like cells, which express markers of both types of macrophages (recruited and RTM) (Fig. 2G-H). RTM were enriched in tumors but also found in normal samples (Fig. S3H-I). RTM-like cells were found in all conditions, except for metastatic and primary uveal melanoma, skin melanomas, liver tumors, and normal breast tissue. Also, we highlight the RTM-like enrichment in ovarian tissue both in normal and tumor samples (Fig. S3J). In general, uveal and skin melanoma samples were enriched with resident cells, while in other tumors, recruited macrophages were predominantly found (Fig. S3J).

Besides classifying macrophages based on origin, we also identified 12 subpopulations based on gene programs (Fig. 2I). The diversity of the macrophage populations was higher in tumors compared to normal samples from the same tissue (Fig. 2J). The subpopulations were stratified according to their functional gene signatures, mostly following the suggestion of a consensus model for TAM diversity *(22)*. The alveolar-like Mac_Alv-like (n = 8,816), marked by the expression of *PPARG* and *MCEMP1,* was predominantly found in lung samples (Fig. 2K, S4A). The Mac_Angio subpopulation (n = 3,108) displayed high expression of genes associated with a pro-angiogenic signature, such as *VEGFA*, *VCAN*, and *EREG* (Fig. 2K, S4B); and Mac_Hypo (n = 1,802) presented an enrichment of genes associated with hypoxia (Fig. 2K, S4C), such as *SLC2A1* and *ERO1*. The Mac_Hypo subpopulation was majoritarily found in the tumor samples (lung and ovary) (Fig 2J, S3K). Two populations expressing high levels of interferon-primed genes (Fig. S4D), such as *CCL8, IDO1,* and *CXCL9,* were named RTM_IFN and Mac_IFN. While Mac_IFN was abundant in lung tumor samples, RTM_IFN was found in melanoma tumors (Fig. 2J-K and S3K). Two subpopulations revealing a lipid-associated (LA) metabolism transcriptional signature (Fig. S4E), including genes such as *FABP5*, *LPL*, and *LIPA,* named RTM_LA (n = 3,569) and Mac_LA (n = 1,230), were found enriched in breast tumors (Fig. S3K); We identified a large population of macrophages expressing high levels of *CCR2,* a receptor for monocyte chemoattractant protein-1, and *FCN1,* which is predominantly expressed in the peripheral blood leukocytes. Therefore, this population was named recruited macrophage Mac_Rec (n = 4,075) (Fig. 2K, S4F); One subpopulation expressing high levels of *MKI67* and *STMN1* was identified as Mac_Prolif (n = 2,234) and enriched in lung and ovary tumor samples (Fig. S3K).

We also identified one recruited subpopulation enriched for regulatory molecules (*CXR3C1*, *TREM2*) and antigen-presentation signatures (*HLA-A/C*, *HLA-DQA1/B1)* named Mac_Reg (n = 3,965) (Fig. 2K, S4G). In contrast, another population expressing high levels of resident markers such as *LYVE1, FOLR2,* and *PLTP,* resembled interstitial macrophages (IM) with specific and marked expression of genes such as *MAF* and *SEPP1*. This subpopulation was

named RTM_IM (n = 3,507) (Fig. 2K, S4H) and consisted of a well-described macrophage type related to vessels and connective tissue *(22, 23)*. The RTM-like subpopulation was characterized by a high expression of metallothioneins (Fig. S4I) and named RTM-like_MT (n = 1,209) (Fig. 2G-K).

To confirm the consistency and reproducibility of the mononuclear phagocyte populations described in this work, we correlated the gene expression profile of these populations with populations described by other groups *(24, 25)* (Fig. 2L-M, S5A-F). When evaluating cluster purity, our approach retrieved highly homogenous clusters, higher than the ones found for mononuclear phagocytes previously identified also using an integrative approach (Fig. S5G). Overall, our findings strongly suggest that, when compared to previous studies, our method for integrating data, clustering, and characterizing subpopulations were capable of significantly expanding the catalog of mononuclear phagocyte subpopulations in the TME, also highlighting rare populations poorly described before, such as the cDC2_CXCL8, unreported in previous works (Fig. 2L-M), the RTM distinct phenotypes (interstitial, lipid-associated metabolism and interferon-primed macrophages), and the monocyte-derived subpopulations, such as the Mac_Rec, Mac_Hypo and RTM-like_MT.

Macrophages exhibit diverse phenotypes and functional states in the TME

Since M1 and M2 polarization gene programs are well characterized in the literature *(26–28)* and they reflect important macrophage-related features, we attempted to assign these profiles to the 12 macrophage subpopulations we identified (Fig. 3A-C). Although the RTM-IFN and RTM_IM were, respectively, the subpopulations displaying the higher scores for M1 and M2 gene signatures (Data file S4), none of the subpopulations clearly presented overexpression of genes connected to either polarization profile. In fact, populations such as the RTM-IFN express high levels of pro-inflammatory cytokines such as *CXCL9*, *CXCL10,* and *CXCL11* (Fig. 3A), but also immunosuppressive molecules such as PDL-1 and PD-L2 (Fig. 3B). Although M1/M2 ratio scores (Fig. 3C) discriminate the subpopulations regarding their polarization states, this dichotomy oversimplifies the highly specialized, transcriptomically dynamic, and extremely heterogeneous nature of macrophages *in vivo*.

To further dissect the roles of macrophage subpopulations in the TME, we examined if each population was related to important cancer hallmark signatures (Fig. 3D-H). Mac_Angio and Mac_Hypo subpopulations exhibited high scores related to angiogenesis (Fig. 3D). Regarding the hypoxia signature, the Mac_Hypo, Mac_IFN, and Mac_Angio subpopulations exhibited higher scores (Fig. 3E). Interestingly, these Mac_Hypo and Mac_Angio subpopulations also exhibited an epithelial-mesenchymal transition score (Fig. 3F). Interestingly, the signature of the extracellular matrix on macrophages *(29)* is enriched in Mac-Alv-like and Mac_LA (Fig. 3G). The antigen presentation score was higher in RTM_LA, RTM_IFN, and Mac_Reg (Fig. 3H).

Given the possible role of macrophages as immunosuppressive cells, we also investigated the co-expression of checkpoint genes and immune regulators. We observed higher expression of *PDCD1* (the gene coding for PD1 protein) and *TREM2* (identified as a marker for TAMs *(30)*) in the Mac_LA, a subpopulation displaying a low M1/M2 score (high anti-inflammatory profile) (Fig. 3C, 3I, and S6A). *TREM2* expression was found to be significantly higher in both LA populations as well as the Mac_Reg and RTM_IFN (Fig 3I and S6A, top). Furthermore, RTM_IFN and Mac_IFN had higher pro-inflammatory scores (Fig. 3A) and higher expression of

CD274 and PDCD1LG2, the genes that encode for PD-L1 and PD-L2 proteins, respectively (Fig. 3I). Of note, *TREM1*, which is normally highly expressed on macrophages in inflamed tissues *(31, 32)*, was mainly expressed by the Mac_Alv-like, Mac_Hypo, and Mac_Angio subpopulations (Fig. 3I and S6A, bottom). Although we have investigated the myeloid-derived suppressor cell signature proposed by Alshetaiwi and colleagues (2020) *(33)* in all MDCs observed in our research, none of the subpopulations showed consistently elevated genes from that signature (Fig S6B). Interestingly, Mac_LA and RTM_LA showed higher suppressive profiles based on *CD84* expression, followed by RTM_IM, RTM_IFN, Mac_IFN, and mast cells (Fig. S6B).

We also investigated the pathways enriched in each macrophage subpopulation, and we discovered pathways compatible with the pro- and anti-inflammatory gene expression patterns observed in some phenotypes. Enriched pathways in the Mac_Rec, Mac_IFN, and RTM_IM phenotypes were associated with the interferon signaling pathway and Toll-like receptor cascades, for example, which is consistent with a pro-inflammatory profile. Interestingly, the Mac_Hypo and Mac_Angio exhibited an enrichment of the interleukin-10 pathway, which relates to anti-inflammatory processes such as angiogenesis, confirming the score presented in Fig 3D. The expression of genes implicated in pathways including IL-4 and IL-13 signaling in the RTM_IFN, RTM_IM, Mac_Rec, Mac_Hypo, and Mac_LA subpopulations indicates their probable anti-inflammatory function. Activation of matrix metalloproteinases, collagen formation, and degradation of the extracellular cellular matrix are just a few of the pathways displayed by macrophages engaged in matrix remodeling, such as Mac_LA and Mac_Hypo (Data file S3, Fig. 3K).

We also investigated the main metabolic pathways that were enriched in each macrophage phenotype. Macrophages can use different sources to fuel energy metabolism, shifting between glucose and lipid metabolism depending on the TME context and their inflammatory role (pro-inflammatory or anti-inflammatory) *(34)*. Thus, in our analysis, the macrophages that exhibited high levels of *TREM2*, such as Mac_LA, RTM_IFN, and Mac_Reg, were enriched in arginine and proline metabolism genes. In contrast, macrophages that presented high expression of *TREM1,* such as Mac_Alv-like, Mac_Hypo, and Mac_Angio, were enriched in glycolysis and gluconeogenesis metabolism genes (Data file S3, Fig 3K). NAD metabolism genes were highly enriched in the Mac_Rec subpopulation (Fig 3L). An overview of the main metabolic pathways is described in Fig. S6C.

In order to clarify the complex trajectory of differentiation in the monocyte-macrophage and monocyte-DC axes, we applied an unsupervised strategy of pseudo-time analysis. Using Mono_CD14_FOS$^-$ as the "root" for the trajectory inference, our analysis supported the notion that CD14^{hi} monocytes represent the precursor population of both Mono CD16^{hi} and monocyte-derived cells in tissues (Fig. 3M). Intermediate monocytes (MonoInter_CXCL10 and MonoInter_CLEC10A) can be sources for both cDC2 and Mac_Rec *(35)* (Fig. S6D-E). Interestingly, the Mac_Rec appears to be, ontologically, the first monocyte-derived macrophage subpopulation and a key precursor of other macrophage states, including RTM-like. Two different phenotypes, arising from two main branches, were shown as the most distant from the root: 1) Mac_Reg and 2) Mac_LA (Fig. S6D-E). Through the analysis of the genes that co-vary over pseudo-time, we were able to identify a set of monocyte-related genes (*VCAN, SELL,* and *FCN1*) at the beginning of the trajectory and both specific and general phenotype markers of the

macrophage lineages (*APOE*, *TREM2*, *SPP1*, *HLA-DRA*, *C1QA,* and *CD63*) in late branches (Fig. 3N).

Macrophages and monocytes are associated with distinct clinical outcomes depending on their niche

We next sought to investigate the clinical impact of the MDC subpopulations in different tumor types using larger cohorts. Forty-seven subpopulation signatures were used to estimate cell type proportions in bulk RNA-Seq samples from The Cancer Genome Atlas (TCGA) in nine tumor types through deconvolution analysis. We calculated the correlation between bulk and scRNA-Seq data to confirm whether our signature matrix could be used to estimate cell populations in the TCGA data and found a significant correlation in all TCGA cohorts (Pearson $R^2 > 0.7$, p-value ≤ 0.05) (Fig. S8).

As expected, all the tumor samples were enriched with malignant cells (~75%), except for lung adenocarcinoma (LUAD), which had very similar proportions of malignant cells (16%) and fibroblasts (19.9%) and a higher content of epithelial cells (43.4%) (Fig. 4A and Fig. S8B). Breast cancer (BRCA) presented a higher proportion of non-epithelial cells, and together with LUAD, these tumors were the most enriched in MDCs (Fig. 4A). Stratifying for mononuclear phagocyte subpopulations, macrophages were among the most abundant cell types (7.69% of total MDCs), and the RTM_IFN subpopulation was more commonly found in different tumor types, especially in skin melanoma (SKCM) (7.6% for metastatic and 5.6% for primary tumor) and uveal melanoma samples (1.9%), whereas RTM_IM was predominantly found in lung tumors (6.06% in LUAD and 3.9% in LUSC). The Mac_LA subpopulation was mainly found in LUAD and BRCA tumors, while Mac_Reg (2.6%) was mainly found in colorectal tumors (Fig. 4B).

By associating clinical parameters such as overall survival with TME subpopulation estimates, the Mac_Angio subpopulation was associated with a poorer prognosis in LUSC, LUAD, and breast cancer subtype Luminal B (BRCA_LumB), whereas the Mac_Reg and cDC2_FCGR3A subpopulations showed a better prognosis in the liver tumor (LIHC) and breast cancer subtype basal (basal BRCA). Interestingly, populations such as RTM_IM, RTM_INF, and cDC2_CXCL8 were associated with distinct prognoses depending on the tumor type (Fig. 4C). Of note, RTM_IM (Fig. 4D and 5C) and Mono_FCG3RA (Fig. 4D, 6C-D) showed as independent predictive biomarkers of poor prognosis in ovary cancer (OV) and basal BRCA cancer patients. Complete results of the univariate and multivariate Cox regression survival analysis are in Data file S7 and S8.

TREM2-expressing macrophages have been associated with a poor prognosis in cancer patients *(36)*. Even though this scenario was observed for RTM_LA and Mac_IFN in LIHC (Fig. 4C), the opposite was verified in the same tumor type for Mac_Reg, the second most highly *TREM2* expressing subpopulation (Fig. S6A). Univariate Cox analysis linked this cell type to a better LIHC outcome (Fig. 4C), most likely due to the absence of known immunosuppressive genes co-expressed in this subpopulation (Fig. 3J). Compared to RTM_LA, Mac_IFN had lower *TREM2* expression but higher *CD274* and *PDCD1LG2* expression (Fig. 3J).

The clustering of OV samples from TCGA based on the inferred TME profile revealed two major groups: one enriched in fibroblasts (14.2%) (Fig. 5A, yellow bars), primarily composed of tumors classified as mesenchymal subtype, which is known to present a poorer

prognosis subtype; and a second group with a more diverse profile. Furthermore, Kaplan-Meier analysis reveals that OV patients with the highest RTM_IM enrichment have poor overall survival (log-rank p-value = 0.017) but no significant difference regarding disease progression (log-rank p-value = 0.65) (Fig. 5B). This subpopulation was enriched in the mesenchymal molecular type (p-value = 1.6e-15) (Fig. 5D, first). Moreover, increased enrichment of RMT_IM was significantly associated with recurrence among patients with OV (Fig. 5D, third). Additionally, RTM_IM levels differed significantly (p-value = 0.04) between patients with complete (R0) and incomplete (R1) therapy responses (Fig 5D, second). While investigating the clinical impact of subpopulations with tumor suppression-related markers such as TREM2 *(37)*, Mac_LA was found to be significantly augmented in the mesenchymal molecular subtypes (p-value = 0.0005). Furthermore, it was increased in patients with an incomplete response (R1) to treatment (Fig. 5E). Additionally, through Pearson correlation analyses between cell subtypes in ovarian tumors, a significant positive moderate correlation (R^2 = 0.3, p-value = $6.69x10^{-9}$) was observed between RTM_IM and endothelial cells, MonoInter_CLEC10A (R^2 = 0.24, p-value = $6.84x10^{-6}$), and with TCD4_em (R^2 = 0.14, p-value = $9.94x10^{-3}$). For Mac_LA, no significant correlation was found (Fig. S11A).

When examining the inferred TME subpopulations in BRCA tumor types that have a significant clinical impact on survival, such as Luminal A and basal, it is noteworthy the high percentage of malignant cells (4.2% for LumA and 4.6% for basal), following of Mac_Rec (9.4% for LumA and 6.3% for basal), and Mac_LA (6.3% for LumA and 8.4% for basal). When it comes to Mono_FCG3RA, the basal subtype had a larger enrichment than the LumA subtype (3.9% and 1.2%, respectively). Intriguingly, Mono_FCG3RA subpopulations showed different prognoses between LumA and basal subtypes. Patients with a higher level of Mono_FCG3RA enrichment in the LumA had a significantly better prognosis (log-rank p-value = 2 x 10^4) (Fig. 6C and E). In contrast, patients with the basal subtype and higher monocyte FCG3RA enrichment had a worse prognosis (log-rank p-value = 0.0045) (Fig. 6D and E, second). Moreover, Mono_FCG3RA was significantly enriched in metastasis (p-value = 0.0023) and in patients with incomplete response to treatment (p-value = 0.0095) in the basal subtype (Fig 6E). Furthermore, for the LumA subtype, Mac_LA was significantly enriched in stage 3 (p-value = 0.05) (Fig. 6G). Additionally, through Pearson correlation analyses between cell subtypes in breast tumors, significant positive correlations were found for the LumA subtype, between Mono_FCG3RA and malignant cells (R^2 =0.18, p-value= $3.52x10^{-3}$) and TCD4_em cells (R^2=0.12, p-value = $1.49x10^{-5}$) (Fig. S11D). For the basal subtype, a moderately significant negative correlation was observed between Mono_FCG3RA and MonoInter_CXCL10 (R^2= -0.23, p-value = $1.21x10^{-3}$). Regarding Mac_LA, a significant moderate to strong negative correlation was found with TCD4_em (R^2 = -0.49, p-value = $2.43x10^{-2}$) (Fig. S11C). All proportions of subpopulation cells across tumor types are shown in Fig. S9, and survival analyses, including Cox regression multivariate analysis, are shown in Fig. S10A.

To confirm our findings, immunohistochemical staining on tissue microarrays was performed to evaluate the expression of specific macrophage subpopulations (CD68[+], TREM2[+]), T lymphocytes (CD8[+], PD-1+), and malignant cells (Ki67[+]) in high-grade serous ovarian cancer (HGSOC) primary tumors, the most lethal gynecologic malignancy. Overall, 69.8% (n = 44) of the patients were in stage 3 and 17.4% (n = 11) were in stage 4, both advanced stages, while the early stages corresponded to 6.3% in stage 1 (n = 4), and 6.3% in stage 2 (n = 4). In addition, 25% (n = 13) of patients underwent neoadjuvant chemotherapy, 91.1% (n = 41) of patients had adjuvant chemotherapy, and 6.5% (n = 11) of the patients underwent both approaches.

Furthermore, 82.5% died (mean 31 months), and 17.4% are still alive (Data file S9). The markers were classified as having high or low expression and categorized as peritumoral or intratumoral, depending on their localization. The presence of macrophages expressing both CD68[+] and TREM2[+] markers (high > 1 cell) in the peritumoral was significantly associated with a better prognosis for HGSOC patients (log-rank p = 0.015) (Fig. 4O). Representative images of tissue microarrays are shown in Fig. 4P and S10A. No significant differences in overall survival (OS) were found between the low- and high-PD-1, CD8, and TREM2-expression groups using the Kaplan-Meier analysis (Fig. S12B).

Crosstalk between macrophages, T, NK, and malignant cells in TME

To identify communication axes mediated by MDC subpopulations, ligand-receptor inference analysis was performed between MDC subpopulations and T lymphocytes, key players in the development of the immune response, as well as with malignant cells (divided per their tumor origin). To accomplish this, it was necessary to classify lymphocytes into subpopulations. In our analysis, we were able to differentiate three populations of NK cells: NK resting (NK_rest), NK cytotoxic (NK_cyto), and NKT. Both CD4[+] and CD8[+] subpopulations were divided into three groups according to their functional roles: naïve (TCD4_naive and TCD8_naive); effector memory (TCD4_em and TCD8_em); and the exhausted state (TCD4_ex and TCD8_ex). For CD4[+], we also identified the subpopulation of regulatory T cells. Additionally, a subpopulation that expresses both γ and δ T-cell receptors (named TGD) was identified (Fig. S7A). The clusters were annotated based on canonical gene markers and DEGs, yielding 11 broad cell types, as mentioned: NK_rest which expressed *XCL1* and *AREG;* NK_cyto positive for *GZMB, FGFBP2,* and *FCGR3A*; NKT which expressed *CD8A, CD8B,* and *FGFBP2*; TCD4_ex which expressed *CTLA4, PDCD1,* and *CD200;* TCD4_em positive for *IL7R, CD40LG* and *ANXA1;* TCD4_naive positive for *CCR7, LEF1, SELL* and *TCF7;* TCD4_reg enriched of FOXP3 and CTLA4; TCD8_em which expressed *GZMK, EMOS,* and *ITM2C;* TCD8_ex enriched of *LAG3* and *GZMB;* TCD8_naive *IL7R* and *ANXA1*; and the last subpopulation TGD positive for *TRDV1, TRDV2,* and *TRGV9* (Fig. S7B).

We obtained 3422 putative intercellular communications, as assessed by ligand-receptor pairings between macrophages and lymphocytes, of which 526 were with NKT, 484 with TCD4_ex, 587 with TCD4_reg, 645 with TCD8_em, 644 with TCD8_ex, and 536 with TGD. We also explored the axis of macrophage and malignant cell interaction, which were subdivided according to tissue origin. This interaction returned 4986 putative intercellular communications. Of these, 942 occur between macrophages and breast cancer, 543 with colorectal cancer, 1474 with liver cancer, 900 with lung cancer, 297 with melanoma, and 830 with ovarian cancer.

Considering the clinical impact of RTM_IM, we explore the specific communication axis between it, malignant cells, and T/NK cells. We detected 10 specific intercellular communications involving the "RTM_IM" cell type and malignant cells that were not detected between other types of macrophages, tumor cells, or lymphocytes. Of these, the APOE:SDC3 interaction was specific between RTIM_IM and malignant cells in ovarian tumors, IGF1:INSR in liver tumors, and TNF:TNFRSF1A in breast tumors (Fig. S7C). Among the 27 interactions involving RTM_IM and lymphocytes and NK cells, AIMP1:RRBP1, CALR:CD44, ITGB1:CD47, PECAM1:CXCR4 and THBS1:CD47 were specific with "TGD lymphocytes", CD28:CD4 with "TCD4_ex", CD320:TGFBR2 and TNF:RASA3 with "NKT", ITGB1:RACK1, TNF:FAS with "TCD4_reg" (Fig. S7D). By analyzing intracellular signaling led by the aforesaid

receptors, RACK1→NPM1 (Fig. S7E) is highlighted given its role as a transcription regulator of PD-L1 in tumors *(38)*.

DISCUSSION

Integration of single-cell data has proven to be a game changer for analyzing large and complex volumes of data. Indeed, this information is critical in the search to better understand the role of cells that compose the TME, which have been described as heterogeneous in several types of cancer and have a direct influence on tumor progression. The immune cells within the TME have either pro- or anti-tumor properties and have been widely researched for their immunotherapeutic potential. Here we have thoroughly integrated and characterized cell subpopulations in the TME, yielding a large and comprehensive dataset comprising different tumors, technologies, and sample types, generated across multiple conditions and obtained from multiple donors. Although high correlations were observed when comparing subpopulations identified in previous works with ours, showing the consistency of our approach, only a few subpopulations exclusively matched the populations we identified, showing our ability to identify novel, more well-defined subpopulations/states. Of note, TREM2 macrophages *(24)* highly correlate with different cells in our catalog, such as RTM_LA, Mac_LA, RTM_IFN, Mac_IFN, Mac_Angio, and Mac_Hypo, highlighting that there is no single TREM2$^+$ macrophage phenotype, as observed by our data. A rigorous methodology and extensive curation allowed us to identify more homogeneous clusters than previous approaches *(24, 25)*, unraveling high-quality clusters that represent a wide range of cells found in the TME of solid tumors, including rare states never previously characterized. These findings can be used to guide future research into obtaining novel biological insights and developing new immune checkpoint blockades and target therapies.

Our work has compiled and analyzed in great detail, at the single-cell level, the largest set of mononuclear phagocytes subpopulations within the TME. The cells we studied include monocytes, macrophages, and cDCs beyond other MDCs, such as neutrophils and mast cells, which are traditionally poorly characterized cells in the tumor context. While no integrative approach has previously described different mast cell phenotypes/states in the TME, our findings showed that they are linked to a better prognosis in LUAD, although they can display gene programs involved in immune response inhibition, such as PD-1 signaling. Because single-cell characterization of tumor-associated neutrophils has been limited due to technological limitations, our work has yielded the largest collection of these cells within the TME by integrating data from different technologies, including InDrop Seq, which does not carry bias in granulocyte retrieval *(39)*. In this study, all identified neutrophil subpopulations, particularly neutrophil_CXCL8, were associated with an anti-inflammatory profile involving interleukin-4 and -13 signaling. According to Xiong and colleagues (2022), CXCL8 mediates neutrophil extracellular traps, which have been associated with poor outcomes in colorectal cancer (19). The lack of neutrophils in normal tissues could be a possible bias for only observing an association with pro-tumorigenic activities in these cells.

Mononuclear phagocytes are key cells in organ development and homeostasis maintenance *(40)*. Their great plasticity enables them to take part in organogenesis, tissue repair, and protection against infections. In addition to producing essential growth and inflammatory factors that orchestrate lymphocyte responses, they are involved in extracellular matrix maintenance, angiogenesis, tissue enervation, and the clearance of apoptotic cells. Given that

tumor development resembles the development of a tissue or organ, it is not surprising that these functions have also been found in these cells within the TME.

Among the eight cDC states identified, cDC1 was the least representative subpopulation within the TME scRNA-Seq. Their function is well understood and is related to CD8$^+$ T cell activation, whereas cDC2 are more pleiotropic, with their main function being the presentation of exogenous antigens to CD4$^+$ T cells, thereby initiating T helper cell differentiation *(37)*. Interestingly, the clinical impact assessment of cCD2_CXCL8 cells, a newly described phenotype, in different cancer types using Cox multivariate analysis showed that these cells are associated with poor prognostics in lung cancer but good prognostics in liver tumors, similar to the results found by Li and colleagues *(41)* that identified *CXCL8* in a high-ImmuneScore population that contributed to better survival in colorectal cancer using bulk RNA-Seq, demonstrating that the TME context determines the function of each cell. The cDC_CD207 phenotype was previously demonstrated by Qian and colleagues (2020) who found a population called C5_CD207 that expressed *CD207* and *CD1A* but not the epithelial markers *CDH1* and *EPCAM (42)*, which are typically expressed in macrophages known as Langerhans cells. This is in agreement with other studies that show the presence of Langerin expression in cDC2 CD1C$^+$, a distinct subpopulation of Langerhans cells *(43)*. As these cells were discovered among cDC2s, in this work they were deemed as such, needing more investigations for confirmation.

Monocytes play a role in orchestrating the immune system, not only in homeostatic conditions but also in tumor progression *(44)*, besides being important precursors for macrophages and cDCs *(45)*. They can be derived from hematopoietic stem and progenitor cells in the bone marrow, and their production is enhanced under certain circumstances *(46, 47)*. The adult spleen also contains a reservoir of monocytes that can be rapidly recruited in response to injury or inflammation *(48)*. These monocytes can be pro- or anti-tumoral, regulating a variety of processes from angiogenesis to immune modulation in a context-dependent manner *(49, 50)*. We observed two classical monocyte subpopulations (Mono_CLEC10A and Mono_CXCL10) enriched in the tumor samples. These monocytes have been reported to play pro-tumoral functions once recruited to the TME *(50)*. Our findings show that one subtype, Mono_FCGR3A, is independently associated with a worse prognosis in the basal subtype of breast cancer, which is consistent with the findings of Chen et al. (44) and Jung et al. (45), who have linked non-classical monocytes with immunosuppressive functions. Interestingly, these monocytes were negatively correlated with MonoInter_CXCL10, a macrophage precursor, suggesting that in this context, monocyte differentiation into pro-inflammatory macrophages is hampered *(51)* An antagonistic impact was observed in the Luminal A subtype, reinforcing that MDCs association with cancer prognosis is context-dependent.

Macrophages are present in different tissues, and according to changes in their environment, mainly verified by *in vitro* studies, they can be polarized into at least two different states, namely M1 and M2 *(12, 52, 53)*. However, as evidenced by our results, none of the macrophage states clearly display gene programs associated with the M1 or M2 signatures. In fact, some populations simultaneously expressed genes associated with both signatures. Although the traditional binary division into M1 and M2 states has been adhered to for many decades, it has become clear that this classification does not convey the complex nature imposed by macrophage plasticity, mainly in the TME context. Following a recent effort to standardize nomenclatures, pushed mainly by the popularization of single-cell analyses, we based our annotation mainly on recent signature genes *(22)*, which propose a consensus model of TAM

diversity. Moreover, using our in-depth analysis, we dissected twelve states according to TAM ontogeny and functionality, each with a distinct transcriptional profile. We confirmed all the phenotypes described by Ma et al. *(22)*, except for the cytokine-enriched inflammatory macrophages (Inflam-TAMs), presumably because our data show a multitude of cytokines expressed simultaneously by many macrophages. Beyond the aforementioned confirmations, we reported a new macrophage state (RTM-like MT), which will be discussed in conjunction with the other RTMs. Furthermore, we identified Mac_Hypo in several tumor types (uveal, ovary, lung, and colorectal), which had previously only been linked to breast cancer *(54)*.

Our study brings substantial contributions regarding MDC ontogeny. It allowed distinguishing subpopulations of resident macrophages, which were frequently lumped together in a single group of cells in multiple TME studies *(22, 55)*. We showed for the first time that these resident macrophages can develop highly specialized phenotypes and be derived not only from embryonic progenitors but also from adult monocytes, such as the population called RTM-like_MT that expresses genes associated with both residents and recruited macrophage signatures. These cells also express genes that can regulate metastasis such as VEGF, IL-6, and TNF-α *(56)*. In addition, they express high levels of metallothioneins, which are oxygen-regulated proteins and, therefore, induced by the hypoxic TME *(57, 58)*. There has been substantial debate regarding the development of macrophages. It is believed that the majority of RTM populations derive from both intra- and extra-embryonic precursors that are seeded in target organs before birth, playing an important role in development and homeostasis *(59)*. Under steady-state conditions, RTMs are mostly maintained via self-renewal. However, bone-marrow-derived circulating monocytes can give rise to self-renewing tissue-resident populations if an adequate habitat is provided *(60)*. Although recruited monocytes can be clearly distinguished from RTMs during the initial stages of inflammation, they can progressively acquire a very similar profile to the RTM population *(61)*.

Our data allowed us to draw parallels regarding biological functionality between the recruited and RTM subpopulations. Markers linked with the interferon-primed gene signature connected to an inflammatory profile (*IFIT1*, *IDO1*, and *CXCL9*), for example, were detected in both the RTM_IFN and recruited Mac_IFN states. Similarly, two populations with distinct ontogenies expressing a lipid-associated phenotype were observed. Both express high levels of *TREM2*, which has been linked to systemic metabolism and adipose tissue function *(62)* and has emerged as a marker of pro-tumorigenic macrophages *(63)*.

By analyzing large cohorts of individuals affected by different types of cancer, we were able to assess the clinical impact of all MDC subpopulations. Among the twelve populations we identified to be associated with cancer prognosis, we highlight the RTM_IM subpopulations, which may represent attractive targets for future studies aimed at understanding the roles of specific TAM subpopulations in the TME. Although RTM_IM seems to be more enriched in lung tumors, this cell type is associated with a worse prognosis in OV patients, where it has been identified as an independent prognostic factor. Furthermore, despite the absence of tumor suppressor gene signatures, such as TREM2 and PD-L1, this macrophage subpopulation expresses IL-10, which promotes the release of IL-6, TGF-β, and VEGF-A and contributes to activating processes related to cell proliferation, tumor invasion, tumor growth, angiogenesis, and metastasis *(64)*. Interestingly, RTM_IM seems to use the APOE:SDC3 axis to interact with malignant cells. Apolipoprotein E (apoE) facilitates the cellular transport and metabolism of lipids *(65)* and different studies have linked higher expression of ApoE with tumorigenesis and

cancer progression, such as cell proliferation, angiogenesis, and metastasis *(66, 67)*. Analysis of transcriptome-wide gene expression in ovarian cancer showed that ApoE is upregulated in most high-grade serous carcinomas compared to low-grade serous carcinomas *(68)*. On the other hand, syndecans (Sdc) are transmembrane heparan sulfate proteoglycans *(69)* expressed in TAM, cancer cells, and endothelial cells, and their expression is positively correlated with a macrophage gene signature across several TCGA cancer types *(70)*. The studies regarding the role of SDC3 and apoE in cancer development are sparse. Thus, the uniqueness of the apoE-SCD3 interaction by ovarian tumor cells and RTM-IM needs to be further investigated, especially given the role of these cells in ovarian cancer prognosis as demonstrated here.

Although recent reports have shown that TREM2 is a key phenotypic marker for TAMs and monocytes in a variety of cancer patients and tumor models with potent immunosuppressive activity *(71)*, our findings strongly suggest that distinct TAMs phenotypes expressing *TREM2* can coexist in the TME, exerting different roles, supporting it's unclear dual function on prognosis *(72, 73)*. Of note, the populations expressing high levels of *TREM2* - RTM_LA and Mac_Reg - were respectively associated with poor and good prognoses in LIHC. By analyzing a Brazilian cohort of HGSOC, we found that the presence of TREM2[+] macrophages in the peritumoral space is associated with a better prognosis for OV patients. In contrast to the findings by Binnewies and collaborators (2021), where intratumoral TAM-expressing TREM2 was shown to be associated with T cell exhaustion and anti-PD-1 resistance *(63)*, our data show that the role of TREM2 can be dependent on localization inside the TME and the co-expression of other immunosuppressive markers. Our study highlights the complexity of TREM2[+] macrophages, calling for more investigation to clarify their function in tumors.

Overall, even though there is currently no general method for accurately defining cell identity, we were able to create a high-resolution, dynamic cell atlas of mononuclear phagocytes in the TME through a robust pipeline by combining analysis of cluster-specific gene expression to infer cell type, distribution, function, ontogeny, phenotype (state), possible cellular interactions, and disease associations. This atlas reveals new aspects of MDC identities as well as new avenues for understanding and manipulating their fate in cancer.

MATERIALS AND METHODS

Data download and pre-processing

Pre-processed scRNA-Seq data (Data file S1) from patients with breast cancer (GSE114727), hepatocellular carcinoma (GSE140228, GSE125449), lung cancer (GSE127465), melanoma (GSE115979), ovarian cancer (GSE154600, GSE72056), uveal melanoma (GSE139829), skin (GSE130973) and metastasis from uveal melanoma and lung (GSE158803) samples were obtained from the public repository Gene Expression Omnibus (GEO) using the GEOquery Bioconductor package *(74)*. Additionally, datasets from lung, ovarian, and breast cancer were downloaded from Qian et al. *(42)*, on its platform (blueprint.lambrechtslab.org/). We downloaded datasets from the Human Lung Cell Atlas project *(75)* on the Synapse platform (SYN21560407). Finally, the dataset containing PBMCs (10x Genomics standard) was downloaded from the company platform (support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k).

Raw gene expression (counts) matrices were analyzed with the Seurat package (v4.0) *(76)* for each study. We converted all gene symbols to Ensembl format (hg v38). Different filters of quality control were applied by technology: for 10x Genomics data the cutoff points were <10 to mitochondrial percent, number of counts (nCounts) > 200, and ratio nCounts by number of features (nFeatures) <5; for Smart-Seq2 were <15 to mitochondrial percent, > 200 nCounts, and ratio nCounts/nFeature < 1000; and for inDrop were < 15 to mitochondrial percent, > 200 nCount, and ratio nCounts/nFeature < 1000. The remaining cells were grouped by patient and sample type in each dataset, individually, and submitted to the doublet removal step through the Scrublet package *(77)*. Each group of cells was separately analyzed through the doublet score histogram distribution by samples (since doublets only form from cells from the same sample), and a cutoff value was determined for each case, by manually setting the threshold between the two peaks of the distribution, eliminating cells with a higher probability of being doublets.

Integration and cluster annotation

The datasets were concatenated with Scanpy (1.7.2) *(78)*, and the 3000 highly variably expressed genes were selected using flavor = "Seurat". The integration was performed with scVI (0.6.8) *(79)* from scvi-tools (0.19.0) *(80)*. For training the variational autoencoder neural network, we used the following hyperparameters: n_latent=20, n_layers=4, dropout_rate=0.1. After we trained the scVI and corrected for the batch using study and tissue origin as batch keys in order to remove the technical bias, the clusters were calculated through the Leiden algorithm *(81)*, where after evaluating the range from 0.3 to 2.0, the resolution of 0.6 was defined for the annotation of the first level of clustering. The latent space generated by scVI was then projected to a 2D space using the Uniform Manifold Approximation and Projection (UMAP) dimensional reduction method *(82)*. To determine the broad cell types, we use canonical markers (Data file S5). We found and removed three (3) undefined/contaminated subsets with low quality (clusters 17, 19, and 29). After this first level of annotation, we applied DecontX *(83)*, from the celda package (1.14.1), using default parameters, to estimate and remove ambient contamination in individual cells. For the other levels of annotation, we subset the broad cell types and then rerun highly variable genes and scVI models with the same parameters, except for batch correction, which was performed by patients and technology, and also excluding samples with a contribution of fewer than ten cells. In order to classify clusters in cell types or states, we apply differential gene expression by Wilcoxon rank-sum and MAST *(84)*, using logfc.threshold = 0.25, and min.pct = 0.1. The FDR method was used to adjust p values. In addition, we evaluated the purity of the clusters through the ROGUE (1.0) *(85)* algorithm with default parameters.

Copy Number Variation inference

The copy number variation signal for individual cells was estimated using the inferCNVpy (0.4.0) (https://github.com/icbi-lab/infercnvpy), inspired by the inferCNV *(86)*, by comparing all normal cells in the TME with epithelial cells, in order to classify malignant cells. Epithelial cells were classified according to the copy number variation signal, cells with a score lower than the median score (0.004 for all tumor types) were classified as "normal", while those with a higher score were classified as "malignant".

***In silico* validation**

In order to validate our findings, we correlated the populations we have defined with those from previous works. Another set of pre-processed scRNA-Seq data was downloaded *(25)(24)* (Data file S6). Those two datasets were selected because they are considered robust amounts of data, well-annotated, mainly with regard to mononuclear phagocytes, and with a great diversity of tumor types, including those we have worked with in this paper.

We used the same strategy and quality control filters that were used with our main datasets. Cheng´s and Mulder´s datasets were analyzed separately, using the authors' annotations for macrophages, monocytes, and cDCs subpopulations. The dataset generated by Cheng and colleagues (2021) *(25)* is composed of more than 56,000 mononuclear phagocyte cells retrieved from samples of eight tumor types, tissues adjacent to the tumors, and blood. The subpopulations were identified separately in each tumor type, yielding nine macrophage subtypes, three monocyte subtypes, and nine cDCs subtypes. Meanwhile, the dataset from Mulder and collaborators (2021) *(24)* has more mononuclear phagocytic cells (approximately 133 thousand cells), which were derived from an integrative approach of 13 tumor samples, tissues adjacent to the tumor, and other sample types, totaling eight macrophage clusters, four monocyte clusters, and one cDCs type (for details, see Supplementary Methods). To perform the correlation, expression data was log-normalized using the NormalizedData function of the Seurat package (v4.0) *(76)*, followed by the AverageExpression function from the same package to determine the average expression of each gene for each subpopulation. Cor function *(87)* from the stats R library was used to correlate the expression of all genes between subpopulations, and ggplot2 *(88)* was used to plot the correlation graphs. The purity of clusters was verified by using the ROGUE *(85)* with the same parameters used for our integrated dataset.

Metabolic characterization

We characterized the metabolic pathways of each cluster using Compass *(89)*. Cells were subsetted into clusters prior to running Compass with the following parameters: --calc-metabolites --microcluster-size 10 --lambda 0. Raw reaction penalties were converted to reaction scores as described in the original publication, and reactions with zero variance were removed. Pairwise comparison was performed for each cluster (e.g. Cluster 1 vs. Cluster 2; Cluster 1 vs. Cluster n; etc) applying Wilcoxon's test and calculating the Cohen's d effect size between means. The FDR method was used to adjust p values. Reactions with confidence scores other than 4 (most confident) or zero (unknown) or with an adjusted p-value greater than 0.1 were removed. Cohen's d median scores were calculated for each pairwise comparison and metabolic pathway. Those median scores were again aggregated by median comparing the median Cohen's d value across all clusters to obtain a single value per cluster and metabolic pathway, which were z-scored to compare across macrophage clusters.

Reaction scores were calculated for each myeloid cluster as previously described and used to perform a pairwise comparison between all clusters. Reactions were filtered for adjusted p-value below 0.1 and a measure of the effect size was calculated using the median of Cohen's values for each metabolic pathway. Next, Cohen's medians for each metabolic pathway were summarized for each cluster to reflect the activity of the respective pathway in the cluster.

The metabolic reactions Glycolysis/Gluconeogenesis, Citric Acid Cycle, NAD Metabolism, Oxidative Phosphorylation, Arginine and Proline Metabolism, Transport, Mitochondrial, Fatty Acid Oxidation, Fatty Acid Synthesis, Glycerophospholipid Metabolism,

Glycosphingolipid Metabolism, Sphingolipid Metabolism are used to classified the macrophage based in the metabolic pathway.

Trajectory analysis through pseudo-time

The Seurat object was converted into a CellDataSet to use the UMAP generated in the integration as the partition. We next employed Monocle 3 version 1.3.1 *(90)* and set the parameter use_partition = FALSE of the function "learn graph" in order to create a linear trajectory, with default parameters. To order cells, the root used to infer the pseudotime was the Mono_CD14_FOS- cluster, with default parameters. We also used Monocle 3 to perform gene module analysis based on graph autocorrelation to identify genes that co-vary over pseudotime.

Module and Pathway Enrichment

The top hundred genes identified in the DEG analysis, sorted by the log fold change for each cluster, were selected to perform the functional enrichment analysis from the curated database Reactome (2021), through the clusterProfiler version 3.0 R package *(91–93)*. Significant pathways were those with a p-value $\leq$ 0.05. The authors selected the pathways and genes included in the gene set list based on their scientific experience, and the redundancy was removed from the enriched pathway set by selecting the pathway with the lowest p-value. The gene signature scores were computed by combining the AddModuleScore function and the MSigDB *(94)* following signature.

Cell-cell communication

To perform cell communication analyses, we filter cells with the detection of more than 200 genes and genes with detection in more than 50 cells and use the CellComm algorithm *(95)* between pairs of cell types of macrophages and tumors or macrophages and lymphocytes. We used as a set of possible intercellular communications the 688 ligands and 857 receptors contained in the NicheNet database *(96)* resulting in 12650 possible pairs. In communications between macrophages and tumors, we used cells classified as "MacAlv-like", "Mac_Angio", "Mac_Hypo", "Mac_IFN", "Mac_LA", "Mac_Rec", "Mac_Reg", "RTM-like_MT", RTM_IFN", "RTM_IM" and "RTM_LA" and breast, colorectal, liver, lung, melanoma and ovarian tumor cells. In communications between macrophages and lymphocytes, we used cells classified as macrophages "Mac_IFN", "Mac_LA", "Mac_Rec", "RTM_IFN" and "RTM_IM" and lymphocytes "NKT", "TCD4ex", "TCD4reg", "TCD8em", "TCD8ex" and "TGD". For consideration of intercellular communication, the expression of ligands and receptors should be greater than 0.25 normalized counts and show a p-value less than 0.01 in a permutation test of cluster annotation containing 1000 permutations. To get paths, molecular interactions that connect receptors participating in communications of interest to possibly active transcription factors, CellComm created weighted interaction networks based on the co-expression of genes in single cells and mapping them to protein-protein interaction data from the iRefIndex database. The paths are then obtained using a flow network with a multiple source and multiple sink structure as described in the original work *(95)*. Each receptor-to-regulator path scores consider

the presence and enrichment of regulator regulon genes in cell type-specific signatures based on hypergeometric tests.

Deconvolution analysis

To better understand the heterogeneity of the tumor microenvironment, as well as the cellular composition in each tumor type, we applied deconvolution methods to predict the abundance of cell types by modeling gene expression levels through bulk RNA-seq data. For these analyses, gene expression data (HTSeq counts), and clinical-pathological information from the database portal TCGA (http://cancergenome.nih.gov/) were downloaded using the R environment package TCGAbiolinks *(97)* of nine different tumor types, using the identifiers: OV (n=354), LUAD (n =510), LUSC (n=496), UVM (n=77), SKCM (divided into Metastatic (n=385) and Primary (n= 103) tumor samples), COAD (n=454), READ (n=170), LIHC (n = 369) and BRCA (n = 1195) where it comprises Luminal A (n= 569), Luminal B (n=210), HER2 (n=81) and basal (n= 188) subtypes. Next, we excluded genes from each patient for which the expression sum equals zero and filtered for the 12,854 genes that make up the signatures of the cell types identified by single-cell analysis. In addition, for the cell signature matrix, we filtered for tumor samples and cell types that represent more than 1% of all other cell types in each tumor type.

Therefore, we used the single-cell signatures identified in this work as a reference, to predict the relative frequency of cell types for each patient of each tumor type from TCGA. We considered 47 cell signatures, including cell types identified in the first level of annotation, such as fibroblasts, endothelial cells, mast cells, pDC, malignant cells, and normal epithelial cells, and cell states/subtypes identified in other levels of clustering, including macrophages (except the Mac_Prolif subpopulation), monocytes, cDC, lymphocytes, NK cells, and neutrophil subpopulations.

For this analysis, we use the BayesPrism package *(98)*, version 2.0, in the R environment, which is based on statistical marginalization (a Bayesian method). For each bulk RNA-seq sample, it evaluates the gene expression of each cell state as well as the proportion of each cell state. Because it eliminates batch effects and variations between the scRNA-Seq matrix and the bulk RNA-seq, predicting the expression of each cell type in each bulk RNA-seq sample, it produces a more robust result. As a first step, we removed ribosomal and mitochondrial genes with high expression using the *cleanup.genes()* function. In addition, genes expressed in less than 5 cells were excluded. After that, to get more consistent results, we filtered for protein-coding genes, and used the function *get.exp.stat()* to select differentially expressed genes between cell states of different cell types using the default parameters, except for the parameter *pseudo.count* (0.1 for 10x data and 10 for Smart-seq data). Finally, we set up the prism object using the *new.prism()* function for each tumor type by setting the following parameters:

*reference=**scRNAfiltered_matrix**,mixture=**bulk_matrix**,input.type="**count.matrix**",cell.type.l abels = **celltype**, cell.state.labels = **states**, key="**Malignant Cells**", outlier.cut=**0.01**, outlier.fraction=**0.1**.* And then running BayesPrism, *run.prism(prism = **prism object** , n.cores=**70**).*

Study design, ethical considerations, and patient selection

This retrospective study was approved by the Ethics in Human Research Committee of the Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil, and was conducted following the Good Clinical Practice Guidelines. All women diagnosed with HGSOC at INCA between 2000 and 2019, regardless of adjuvant or neoadjuvant treatment, were identified through the internal database. Clinical data regarding age at diagnosis, staging, surgery, histological subtype, chemotherapy, and survival were retrospectively obtained from the medical records. The TNBC *(99)* cohort was described previously. The study was approved by the institution, reviewed, and approved by the Ethics in Human Research Committee of the INCA, Rio de Janeiro, Brazil, and conducted in accordance with Good Clinical Practice Guidelines. The ethics committee waived the requirement of written informed consent for participation.

Immunohistochemistry

The blocks were cut with a microtome into fine slivers of 4-mm sections to form the tissue microarray which encompasses the most representative areas of greatest tumor cellularity in formalin-fixed paraffin-embedded tissue specimens. The immunohistochemistry was processed using a NOVOLINK™ Polymer Detection Systems (Leica), followed by DAB. Antigen retrieval was performed with the Trilogy (Cell Marque) reagent. For double staining, TREM2 was combined with anti-CD68. Briefly, after completing the first immune reaction with Mach 4 MR-AP (Biocare Medical), the second immune reaction was visualized using DAB.

A total of 63 patients from the INCA, between the years 2001-2017, were investigated for $CD8^+$, $PD-1^+$, $Ki67^+$, $CD68^+$ and $TREM2^+$ expression. The prognostic value of $CD8^+$, PD-1, $CD68^+$, and $TREM2^+$ expression was associated with clinicopathological characteristics by multivariate Cox regression analyses. The samples were immunostained for Ki67 (clone 30-9 at 1:7, Ventana- Roche), CD8 (clone SP57 at 1:7, Ventana- Roche), PD-1 (clone NAT105, Cell Marque, diluted 1:100), CD68 (clone KP-1 at 1:20, Ventana- Roche) and TREM2 (clone D814C, 1:100, Cell Signaling Technology). The entire analysis was carried out at the Division of Pathology-National Institute of Cancer by two experienced pathologists.

Tissue microarrays were assessed by determining the percent of immune cells and their distribution with the markers $TREM2^+$ and $CD68^+$ peritumoral (low = 0 cells, high ≥ 1 cell) and intratumoral (low = 0 - ≤ 1 cell, high > 1 cells), $CD68^+$ peritumoral and intratumoral (low = 0 cells, high ≥ 1 cells) and $PD-1^+$ and $CD8^+$ peritumoral and intratumoral (low = 0 - 1 cell, high > 5-30 cells).

Survival and statistical analysis

Using the results of deconvolution analysis on the bulk RNA-seq samples, we investigated the clinical impact of each cell type identified by the single-cell analyses. To do this, we calculated the hazard ratio (HR) of the relative fractions of each cell population type divided into quartiles, with 95% confidence intervals (CI) based on maximum likelihood estimates for each covariate using a univariate Cox regression model ($p ≤ 0.05$). Subsequently, significant variables were tested in the multivariate Cox regression model ($p ≤ 0.05$), adjusting for covariates known to be prognostic factors, such as age and clinical staging. For these analyses, we used the Survival and Survminer packages for the R environment. In addition, we investigated the correlation between clinical variables and the relative proportions of each cell type predicted by BayesPrism. For this, clinical variables such as response to treatment (responders and non-responders), clinical staging, mutational burden, tumor grade, molecular

subtype, tumor residual disease, tumor event type (recurrence or non-recurrence), and age were investigated. Only results with $p \leq 0.05$ were presented. For the division of high and low groups, the surv_cutpoint function from the survminer package was used. The statistical analyses were done by Kruskal-Wallis for global comparisons, followed by the Wilcoxon test for paired comparisons.

For the Brazilian cohort, overall survival (OS) was calculated from the diagnosis to the date of death of any cause if the patient was known to be alive on the last day of data collection. A further analysis comparing the paired scores of immunohistochemistry factors (IHC markers) for intratumoral and peritumoral components was performed by a multivariate Cox regression model using stage and score of high and low expression. The statistical analyses were conducted using R project version 3.5.3

References and Notes

1. D. Hanahan, Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).

2. A. D. Waldman, J. M. Fritz, M. J. Lenardo, A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat. Rev. Immunol.* **20**, 651–668 (2020).

3. P. Sharma, B. A. Siddiqui, S. Anandhan, S. S. Yadav, S. K. Subudhi, J. Gao, S. Goswami, J. P. Allison, The Next Decade of Immune Checkpoint Therapy. *Cancer Discov.* **11**, 838–857 (2021).

4. E. Peranzoni, V. Ingangi, E. Masetto, L. Pinton, I. Marigo, Myeloid Cells as Clinical Biomarkers for Immune Checkpoint Blockade. *Front. Immunol.* **11**, 1590 (2020).

5. N. Kumari, S. H. Choi, Tumor-associated macrophages in cancer: recent advancements in cancer nanoimmunotherapies. *J. Exp. Clin. Cancer Res.* **41**, 68 (2022).

6. R. Somasundaram, T. Connelly, R. Choi, H. Choi, A. Samarkina, L. Li, E. Gregorio, Y. Chen, R. Thakur, M. Abdel-Mohsen, M. Beqiri, M. Kiernan, M. Perego, F. Wang, M. Xiao, P. Brafford, X. Yang, X. Xu, A. Secreto, G. Danet-Desnoyers, D. Traum, K. H. Kaestner, A. C. Huang, D. Hristova, J. Wang, M. Fukunaga-Kalabis, C. Krepler, F. Ping-Chen, X. Zhou, A. Gutierrez, V. W. Rebecca, P. Vonteddu, F. Dotiwala, S. Bala, S. Majumdar, H. Dweep, J. Wickramasinghe, A. V. Kossenkov, J. Reyes-Arbujas, K. Santiago, T. Nguyen, J. Griss, F. Keeney, J. Hayden, B. J. Gavin, D. Weiner, L. J. Montaner, Q. Liu, L. Peiffer, J. Becker, E. M. Burton, M. A. Davies, M. T. Tetzlaff, K. Muthumani, J. A. Wargo, D. Gabrilovich, M. Herlyn, Tumor-infiltrating mast cells are associated with resistance to anti-PD-1 therapy. *Nat. Commun.* **12**, 346 (2021).

7. J. T. Greene, B. F. Brian 4th, S. E. Senevirathne, T. S. Freedman, Regulation of myeloid-cell activation. *Curr. Opin. Immunol.* **73**, 34–42 (2021).

8. G. Rizzo, J. Gropper, M. Piollet, E. Vafadarnejad, A. Rizakou, S. R. Bandi, P. Arampatzi, T. Krammer, N. DiFabion, O. Dietrich, A. P. Arias-Loza, M. Prinz, M. Mack, K. Schlepckow, C. Haass, J. S. Silvestre, A. Zernecke, A. E. Saliba, C. Cochain, Dynamics of monocyte-derived macrophage diversity in experimental myocardial infarction. *Cardiovasc. Res.* (2022), doi:10.1093/cvr/cvac113.

9. R. Nalio Ramos, Y. Missolo-Koussou, Y. Gerber-Ferder, C. P. Bromley, M. Bugatti, N. G. Núñez, J. Tosello Boari, W. Richer, L. Menger, J. Denizeau, C. Sedlik, P. Caudana, F. Kotsias, L. L. Niborski, S. Viel, M. Bohec, S. Lameiras, S. Baulande, L. Lesage, A. Nicolas, D. Meseure, A. Vincent-Salomon, F. Reyal, C.-A. Dutertre, F. Ginhoux, L. Vimeux, E. Donnadieu, B. Buttard, J. Galon, S. Zelenay, W. Vermi, P. Guermonprez, E. Piaggio, J. Helft, Tissue-resident FOLR2+ macrophages associate with CD8+ T cell infiltration in human breast cancer. *Cell* **185**, 1189–1207.e25 (2022).

10. I. Larionova, N. Cherdyntseva, T. Liu, M. Patysheva, M. Rakina, J. Kzhyshkowska, Interaction of tumor-associated macrophages and cancer chemotherapy. *Oncoimmunology* **8**, 1596004 (2019).

11. H. Aegerter, B. N. Lambrecht, C. V. Jakubzick, Biology of lung macrophages in health and disease. *Immunity* **55**, 1564–1580 (2022).

12. C. D. Mills, K. Kincaid, J. M. Alt, M. J. Heilman, A. M. Hill, M-1/M-2 macrophages and the Th1/Th2 paradigm. *J. Immunol.* **164**, 6166–6173 (2000).

13. K. Kübler, T. H. Ayub, S. K. Weber, O. Zivanovic, A. Abramian, M.-D. Keyver-Paik, M. R. Mallmann, C. Kaiser, N. B. Serçe, W. Kuhn, C. Rudlowski, Prognostic significance of tumor-associated macrophages in endometrial adenocarcinoma. *Gynecol. Oncol.* **135**, 176–183 (2014).

14. S. R. Gordon, R. L. Maute, B. W. Dulken, G. Hutter, B. M. George, M. N. McCracken, R. Gupta, J. M. Tsai, R. Sinha, D. Corey, A. M. Ring, A. J. Connolly, I. L. Weissman, PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. *Nature* **545**, 495–499 (2017).

15. X. Yuan, J. Zhang, D. Li, Y. Mao, F. Mo, W. Du, X. Ma, Prognostic significance of tumor-associated macrophages in ovarian cancer: A meta-analysis. *Gynecol. Oncol.* **147**, 181–187 (2017).

16. A. Mantovani, S. K. Biswas, M. R. Galdiero, A. Sica, M. Locati, Macrophage plasticity and polarization in tissue repair and remodelling. *J. Pathol.* **229**, 176–185 (2013).

17. L. R. Sanchez, L. Borriello, D. Entenberg, J. S. Condeelis, M. H. Oktay, G. S. Karagiannis, The emerging roles of macrophages in cancer metastasis and response to chemotherapy. *J. Leukoc. Biol.* **106**, 259–274 (2019).

18. H. J. Kim, Y. R. Ji, Y. M. Lee, Crosstalk between angiogenesis and immune regulation in the tumor microenvironment. *Arch. Pharm. Res.* **45**, 401–416 (2022).

19. M. Li, L. He, J. Zhu, P. Zhang, S. Liang, Targeting tumor-associated macrophages for cancer treatment. *Cell Biosci.* **12**, 85 (2022).

20. S. Tamoutounour, M. Guilliams, F. Montanana Sanchis, H. Liu, D. Terhorst, C. Malosse, E. Pollet, L. Ardouin, H. Luche, C. Sanchez, M. Dalod, B. Malissen, S. Henri, Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. *Immunity* **39**, 925–938 (2013).

21. T. Doebel, B. Voisin, K. Nagao, Langerhans Cells - The Macrophage in Dendritic Cell Clothing. *Trends Immunol.* **38**, 817–828 (2017).

22. R.-Y. Ma, A. Black, B.-Z. Qian, Macrophage diversity in cancer revisited in the era of single-cell omics. *Trends Immunol.* **43**, 546–563 (2022).

23. J. Schyns, Q. Bai, C. Ruscitti, C. Radermecker, S. De Schepper, S. Chakarov, F. Farnir, D. Pirottin, F. Ginhoux, G. Boeckxstaens, F. Bureau, T. Marichal, Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. *Nat. Commun.* **10**, 3964 (2019).

24. K. Mulder, A. A. Patel, W. T. Kong, C. Piot, E. Halitzki, G. Dunsmore, S. Khalilnezhad, S. E. Irac, A. Dubuisson, M. Chevrier, X. M. Zhang, J. K. C. Tam, T. K. H. Lim, R. M. M. Wong, R. Pai, A. I. S. Khalil, P. K. H. Chow, S. Z. Wu, G. Al-Eryani, D. Roden, A. Swarbrick, J. K. Y. Chan, S. Albani, L. Derosa, L. Zitvogel, A. Sharma, J. Chen, A. Silvin, A. Bertoletti, C. Blériot, C.-A. Dutertre, F. Ginhoux, Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54**, 1883–1900.e5 (2021).

25. S. Cheng, Z. Li, R. Gao, B. Xing, Y. Gao, Y. Yang, S. Qin, L. Zhang, H. Ouyang, P. Du, L. Jiang, B. Zhang, Y. Yang, X. Wang, X. Ren, J.-X. Bei, X. Hu, Z. Bu, J. Ji, Z. Zhang, A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792–809.e23 (2021).

26. V. Sjoelund, M. Smelkinson, A. Nita-Lazar, Phosphoproteome profiling of the macrophage response to different toll-like receptor ligands identifies differences in global phosphorylation dynamics. *J. Proteome Res.* **13**, 5185–5197 (2014).

27. C. M. Daniels, P. R. Kaplan, I. Bishof, C. Bradfield, T. Tucholski, A. G. Nuccio, N. P. Manes, S. Katz, I. D. C. Fraser, A. Nita-Lazar, Dynamic ADP-Ribosylome, Phosphoproteome, and Interactome in LPS-Activated Macrophages. *J. Proteome Res.* **19**, 3716–3731 (2020).

28. L. He, J.-H. Jhong, Q. Chen, K.-Y. Huang, K. Strittmatter, J. Kreuzer, M. DeRan, X. Wu, T.-Y. Lee, N. Slavov, W. Haas, A. G. Marneros, Global characterization of macrophage polarization mechanisms and identification of M2-type polarization inhibitors. *Cell Rep.* **37**, 109955 (2021).

29. J. Etich, M. Koch, R. Wagener, F. Zaucke, M. Fabri, B. Brachvogel, Gene Expression Profiling of the Extracellular Matrix Signature in Macrophages of Different Activation Status: Relevance for Skin Wound Healing. *Int. J. Mol. Sci.* **20** (2019), doi:10.3390/ijms20205086.

30. Y. Katzenelenbogen, F. Sheban, A. Yalin, I. Yofe, D. Svetlichnyy, D. A. Jaitin, C. Bornstein, A. Moshe, H. Keren-Shaul, M. Cohen, S.-Y. Wang, B. Li, E. David, T.-M. Salame, A. Weiner, I.

Amit, Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer. *Cell* **182**, 872–885.e19 (2020).

31. C.-C. Ho, W.-Y. Liao, C.-Y. Wang, Y.-H. Lu, H.-Y. Huang, H.-Y. Chen, W.-K. Chan, H.-W. Chen, P.-C. Yang, TREM-1 expression in tumor-associated macrophages and clinical outcome in lung cancer. *Am. J. Respir. Crit. Care Med.* **177**, 763–770 (2008).

32. F.-C. Yang, P.-Y. Chiu, Y. Chen, T. W. Mak, N.-J. Chen, TREM-1-dependent M1 macrophage polarization restores intestinal epithelium damaged by DSS-induced colitis by activating IL-22-producing innate lymphoid cells. *J. Biomed. Sci.* **26**, 46 (2019).

33. H. Alshetaiwi, N. Pervolarakis, L. L. McIntyre, D. Ma, Q. Nguyen, J. A. Rath, K. Nee, G. Hernandez, K. Evans, L. Torosian, A. Silva, C. Walsh, K. Kessenbrock, Defining the emergence of myeloid-derived suppressor cells in breast cancer using single-cell transcriptomics. *Sci Immunol* **5** (2020).

34. X. Zhang, F. Zink, F. Hezel, J. Vogt, U. Wachter, M. Wepler, M. Loconte, C. Kranz, A. Hellmann, B. Mizaikoff, P. Radermacher, C. Hartmann, Metabolic substrate utilization in stress-induced immune cells. *Intensive Care Med Exp* **8**, 28 (2020).

35. J. Sheng, Q. Chen, I. Soncin, S. L. Ng, K. Karjalainen, C. Ruedl, A Discrete Subset of Monocyte-Derived Cells among Typical Conventional Type 2 Dendritic Cells Can Efficiently Cross-Present. *Cell Rep.* **21**, 1203–1214 (2017).

36. E. M. Wolf, B. Fingleton, A. H. Hasty, The therapeutic potential of TREM2 in cancer. *Front. Oncol.* **12**, 984193 (2022).

37. L. Zhou, M. Wang, H. Guo, J. Hou, Y. Zhang, M. Li, X. Wu, X. Chen, L. Wang, Integrated Analysis Highlights the Immunosuppressive Role of TREM2 Macrophages in Hepatocellular Carcinoma. *Front. Immunol.* **13**, 848367 (2022).

38. G. Qin, X. Wang, S. Ye, Y. Li, M. Chen, S. Wang, T. Qin, C. Zhang, Y. Li, Q. Long, H. Hu, D. Shi, J. Li, K. Zhang, Q. Zhai, Y. Tang, T. Kang, P. Lan, F. Xie, J. Lu, W. Deng, NPM1 upregulates the transcription of PD-L1 and suppresses T cell activity in triple-negative breast cancer. *Nat. Commun.* **11**, 1669 (2020).

39. T. M. Yamawaki, D. R. Lu, D. C. Ellwanger, D. Bhatt, P. Manzanillo, V. Arias, H. Zhou, O. K. Yoon, O. Homann, S. Wang, C.-M. Li, Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling. *BMC Genomics* **22**, 66 (2021).

40. C. V. Jones, S. D. Ricardo, Macrophages and CSF-1: implications for development and beyond. *Organogenesis* **9**, 249–260 (2013).

41. E. Li, X. Yang, Y. Du, G. Wang, D. W. Chan, D. Wu, P. Xu, P. Ni, D. Xu, Y. Hu, CXCL8 Associated Dendritic Cell Activation Marker Expression and Recruitment as Indicators of Favorable Outcomes in Colorectal Cancer. *Front. Immunol.* **12**, 667177 (2021).

42. J. Qian, S. Olbrecht, B. Boeckx, H. Vos, D. Laoui, E. Etlioglu, E. Wauters, V. Pomella, S. Verbandt, P. Busschaert, A. Bassez, A. Franken, M. V. Bempt, J. Xiong, B. Weynand, Y. van

Herck, A. Antoranz, F. M. Bosisio, B. Thienpont, G. Floris, I. Vergote, A. Smeets, S. Tejpar, D. Lambrechts, A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020).

43. V. Bigley, N. McGovern, P. Milne, R. Dickinson, S. Pagan, S. Cookson, M. Haniffa, M. Collin, Langerin-expressing dendritic cells in human tissues are related to CD1c+ dendritic cells and distinct from Langerhans cells and CD141high XCR1+ dendritic cells. *J. Leukoc. Biol.* **97**, 627–634 (2015).

44. A. A. Wolf, A. Yáñez, P. K. Barman, H. S. Goodridge, The Ontogeny of Monocyte Subsets. *Front. Immunol.* **10**, 1642 (2019).

45. Y. Gao, H. Li, Z. Li, L. Xie, X. Liu, Z. Huang, B. Chen, X. Lin, X. Wang, Y. Zheng, W. Su, Single-Cell Analysis Reveals the Heterogeneity of Monocyte-Derived and Peripheral Type-2 Conventional Dendritic Cells. *J. Immunol.* **207**, 837–848 (2021).

46. M. T. Baldridge, K. Y. King, M. A. Goodell, Inflammatory signals regulate hematopoietic stem cells. *Trends Immunol.* **32**, 57–65 (2011).

47. G. Hoeffel, F. Ginhoux, Fetal monocytes and the origins of tissue-resident macrophages. *Cell. Immunol.* **330**, 5–15 (2018).

48. F. K. Swirski, M. Nahrendorf, M. Etzrodt, M. Wildgruber, V. Cortez-Retamozo, P. Panizzi, J.-L. Figueiredo, R. H. Kohler, A. Chudnovskiy, P. Waterman, E. Aikawa, T. R. Mempel, P. Libby, R. Weissleder, M. J. Pittet, Identification of splenic reservoir monocytes and their deployment to inflammatory sites. *Science* **325**, 612–616 (2009).

49. C. Jakubzick, E. L. Gautier, S. L. Gibbings, D. K. Sojka, A. Schlitzer, T. E. Johnson, S. Ivanov, Q. Duan, S. Bala, T. Condon, N. van Rooijen, J. R. Grainger, Y. Belkaid, A. Ma'ayan, D. W. H. Riches, W. M. Yokoyama, F. Ginhoux, P. M. Henson, G. J. Randolph, Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes. *Immunity* **39**, 599–610 (2013).

50. J. Jeong, Y. Suh, K. Jung, Context Drives Diversification of Monocytes and Neutrophils in Orchestrating the Tumor Microenvironment. *Front. Immunol.* **10**, 1817 (2019).

51. R. N. Ramos, C. Rodriguez, M. Hubert, M. Ardin, I. Treilleux, C. H. Ries, E. Lavergne, S. Chabaud, A. Colombe, O. Trédan, H. G. Guedes, F. Laginha, W. Richer, E. Piaggio, J. A. M. Barbuto, C. Caux, C. Ménétrier-Caux, N. Bendriss-Vermare, CD163 tumor-associated macrophage accumulation in breast cancer patients reflects both local differentiation signals and systemic skewing of monocytes. *Clin Transl Immunology* **9**, e1108 (2020).

52. F. O. Martinez, S. Gordon, The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* **6**, 13 (2014).

53. Macrophage M1/M2 polarization. *Eur. J. Pharmacol.* **877**, 173090 (2020).

54. A. Bassez, H. Vos, L. Van Dyck, G. Floris, I. Arijs, C. Desmedt, B. Boeckx, M. Vanden Bempt, I. Nevelsteen, K. Lambein, K. Punie, P. Neven, A. D. Garg, H. Wildiers, J. Qian, A.

Smeets, D. Lambrechts, A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **27**, 820–832 (2021).

55. M. Casanova-Acebes, E. Dalla, A. M. Leader, J. LeBerichel, J. Nikolic, B. M. Morales, M. Brown, C. Chang, L. Troncoso, S. T. Chen, A. Sastre-Perona, M. D. Park, A. Tabachnikova, M. Dhainaut, P. Hamon, B. Maier, C. M. Sawai, E. Agulló-Pascual, M. Schober, B. D. Brown, B. Reizis, T. Marron, E. Kenigsberg, C. Moussion, P. Benaroch, J. A. Aguirre-Ghiso, M. Merad, Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature* **595**, 578–584 (2021).

56. B. Joshi, D. Ordonez-Ercan, P. Dasgupta, S. Chellappan, Induction of human metallothionein 1G promoter by VEGF and heavy metals: differential involvement of E2F and metal transcription factors. *Oncogene* **24**, 2204–2217 (2005).

57. J. A. Raleigh, S. C. Chou, D. P. Calkins-Adams, C. A. Ballenger, D. B. Novotny, M. A. Varia, A clinical study of hypoxia and metallothionein protein expression in squamous cell carcinomas. *Clin. Cancer Res.* **6**, 855–862 (2000).

58. H. Mimata, T. Nomura, F. Sato, M. Yamasaki, Metallothionein is up-regulated under hypoxia and promotes the survival of human prostate cancer cells*Oncology Reports* (2007), doi:10.3892/or.18.5.1145.

59. Y. Lavin, D. Winter, R. Blecher-Gonen, E. David, H. Keren-Shaul, M. Merad, S. Jung, I. Amit, Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).

60. D. Hashimoto, A. Chow, C. Noizat, P. Teo, M. B. Beasley, M. Leboeuf, C. D. Becker, P. See, J. Price, D. Lucas, M. Greter, A. Mortha, S. W. Boyer, E. C. Forsberg, M. Tanaka, N. van Rooijen, A. García-Sastre, E. R. Stanley, F. Ginhoux, P. S. Frenette, M. Merad, Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* **38**, 792–804 (2013).

61. C. Blériot, S. Chakarov, F. Ginhoux, Determinants of Resident Tissue Macrophage Identity and Function. *Immunity* **52**, 957–970 (2020).

62. R.-Y. Li, Q. Qin, H.-C. Yang, Y.-Y. Wang, Y.-X. Mi, Y.-S. Yin, M. Wang, C.-J. Yu, Y. Tang, TREM2 in the pathogenesis of AD: a lipid metabolism regulator and potential metabolic therapeutic target*Molecular Neurodegeneration* **17** (2022), doi:10.1186/s13024-022-00542-y.

63. M. Binnewies, J. L. Pollack, J. Rudolph, S. Dash, M. Abushawish, T. Lee, N. S. Jahchan, P. Canaday, E. Lu, M. Norng, S. Mankikar, V. M. Liu, X. Du, A. Chen, R. Mehta, R. Palmer, V. Juric, L. Liang, K. P. Baker, L. Reyno, M. F. Krummel, M. Streuli, V. Sriram, Targeting TREM2 on tumor-associated macrophages enhances immunotherapy. *Cell Rep.* **37**, 109844 (2021).

64. D. L. Almeida-Nunes, A. Mendes-Frias, R. Silvestre, R. J. Dinis-Oliveira, S. Ricardo, Immune Tumor Microenvironment in Ovarian Cancer Ascites. *Int. J. Mol. Sci.* **23** (2022), doi:10.3390/ijms231810692.

65. Y. Huang, R. W. Mahley, Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiol. Dis.* **72 Pt A**, 3–12 (2014).

66. Z. Zhao, S. Zou, X. Guan, M. Wang, Z. Jiang, Z. Liu, C. Li, H. Lin, X. Liu, R. Yang, Y. Gao, X. Wang, Apolipoprotein E Overexpression Is Associated With Tumor Progression and Poor Survival in Colorectal Cancer*Frontiers in Genetics* **9** (2018), doi:10.3389/fgene.2018.00650.

67. S. B. Kemp, E. S. Carpenter, N. G. Steele, K. L. Donahue, Z. C. Nwosu, A. Pacheco, A. Velez-Delgado, R. E. Menjivar, F. Lima, S. The, C. E. Espinoza, K. Brown, D. Long, C. A. Lyssiotis, A. Rao, Y. Zhang, M. Pasca di Magliano, H. C. Crawford, Apolipoprotein E Promotes Immune Suppression in Pancreatic Cancer through NF-κB-Mediated Production of CXCL1. *Cancer Res.* **81**, 4305–4318 (2021).

68. I.-M. Shih, R. J. Kurman, Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *Am. J. Pathol.* **164**, 1511–1518 (2004).

69. L. Hillemeyer, N. A. Espinoza-Sanchez, B. Greve, N. Hassan, A. Chelariu-Raicu, L. Kiesel, M. Götte, The Cell Surface Heparan Sulfate Proteoglycan Syndecan-3 Promotes Ovarian Cancer Pathogenesis. *Int. J. Mol. Sci.* **23** (2022), doi:10.3390/ijms23105793.

70. E. Prieto-Fernández, L. Egia-Mendikute, A. Bosch, A. García Del Río, B. Jimenez-Lasheras, A. Antoñana-Vildosola, S. Y. Lee, A. Palazon, Hypoxia Promotes Syndecan-3 Expression in the Tumor Microenvironment. *Front. Immunol.* **11**, 586977 (2020).

71. K. Nakamura, M. J. Smyth, TREM2 marks tumor-associated macrophages. *Signal Transduct Target Ther* **5**, 233 (2020).

72. High TREM2 expression correlates with poor prognosis in gastric cancer. *Hum. Pathol.* **72**, 91–99 (2018).

73. X. Cheng, X. Wang, K. Nie, L. Cheng, Z. Zhang, Y. Hu, W. Peng, Systematic Pan-Cancer Analysis Identifies TREM2 as an Immunological and Prognostic Biomarker. *Front. Immunol.* **12**, 646523 (2021).

74. S. Davis, P. S. Meltzer, GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).

75. K. J. Travaglini, A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, G. J. Berry, J. B. Shrager, R. J. Metzger, C. S. Kuo, N. Neff, I. L. Weissman, S. R. Quake, M. A. Krasnow, A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).

76. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

77. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).

78. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

79. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

80. A. Gayoso, R. Lopez, G. Xing, P. Boyeau, K. Wu, scvi-tools: a library for deep probabilistic analysis of single-cell omics data. *bioRxiv* (2021) (available at https://www.biorxiv.org/content/10.1101/2021.04.28.441833v1.abstract).

81. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

82. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*arXiv [stat.ML]* (2018) (available at http://arxiv.org/abs/1802.03426).

83. S. Yang, S. E. Corbett, Y. Koga, Z. Wang, W. E. Johnson, M. Yajima, J. D. Campbell, Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* **21**, 57 (2020).

84. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

85. B. Liu, C. Li, Z. Li, D. Wang, X. Ren, Z. Zhang, An entropy-based metric for assessing the purity of single cell populations. *Nat. Commun.* **11**, 3155 (2020).

86. T. Tickle, I. Tirosh, C. Georgescu, M. Brown, B. Haas, inferCNV of the Trinity CTAT Project. *Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA* (2019).

87. G. Austin, R. A. Becker, J. M. Chambers, A. R. Wilks, The New S Language, A Programming Environment for Data Analysis and Graphics*The Economic Journal* **100**, 650 (1990).

88. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).

89. A. Wagner, C. Wang, J. Fessler, D. DeTomaso, J. Avila-Pacheco, J. Kaminski, S. Zaghouani, E. Christian, P. Thakore, B. Schellhaass, E. Akama-Garren, K. Pierce, V. Singh, N. Ron-Harel, V. P. Douglas, L. Bod, A. Schnell, D. Puleston, R. A. Sobel, M. Haigis, E. L. Pearce, M. Soleimani, C. Clish, A. Regev, V. K. Kuchroo, N. Yosef, Metabolic modeling of single Th17 cells reveals regulators of autoimmunity. *Cell* **184**, 4168–4185.e21 (2021).

90. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

91. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

92. A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C. D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).

93. M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, M. Tanabe, New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595 (2019).

94. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).

95. E. Lummertz da Rocha, C. Kubaczka, W. W. Sugden, M. A. Najia, R. Jing, A. Markel, Z. C. LeBlanc, R. Dos Santos Peixoto, M. Falchetti, J. J. Collins, T. E. North, G. Q. Daley, CellComm infers cellular crosstalk that drives haematopoietic stem and progenitor cell development. *Nat. Cell Biol.* **24**, 579–589 (2022).

96. R. Browaeys, W. Saelens, Y. Saeys, NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).

97. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).

98. T. Chu, Z. Wang, D. Pe'er, C. G. Danko, Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer* **3**, 505–517 (2022).

99. J. L. da Silva, L. Z. de Albuquerque, F. R. Rodrigues, G. G. de Mesquita, P. V. Fernandes, L. C. S. Thuler, A. C. de Melo, Prognostic Influence of Residual Tumor-Infiltrating Lymphocyte Subtype After Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Front. Oncol.* **11**, 636716 (2021).

Author contributions:

Conceptualization: MB, MAM, PMMMV, GRM, GRG
Single-cell integration, and cells annotation: GRM, GRG
Pathways analysis: MAP, NET
Pseudo-time analysis: GRG
In silico validation: LOS
Cell-Cell communication analysis: MF, MMD, ELR
Deconvolution, and clinical association: CET, NGT, GRM
Immunohistochemistry analysis: NGT, NCB, FCM, FRR, CET, AFS
Cohort obtention: CBPC, ACM, JLS, NGT, NET
Funding acquisition: MB, MAM, PMMMV
Supervision: MB, MAM, PMMMV
Writing – original draft: GRM, GRG, NGT, CET, LOS, NET
Writing – review & editing: MB, MAM, PMMMV, ACM, ELR

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data and materials availability: All data used in this work have been listed in Data file S1, including accession numbers. The integrated dataset is accessible through CELL×GENE | Collections.

Figures:
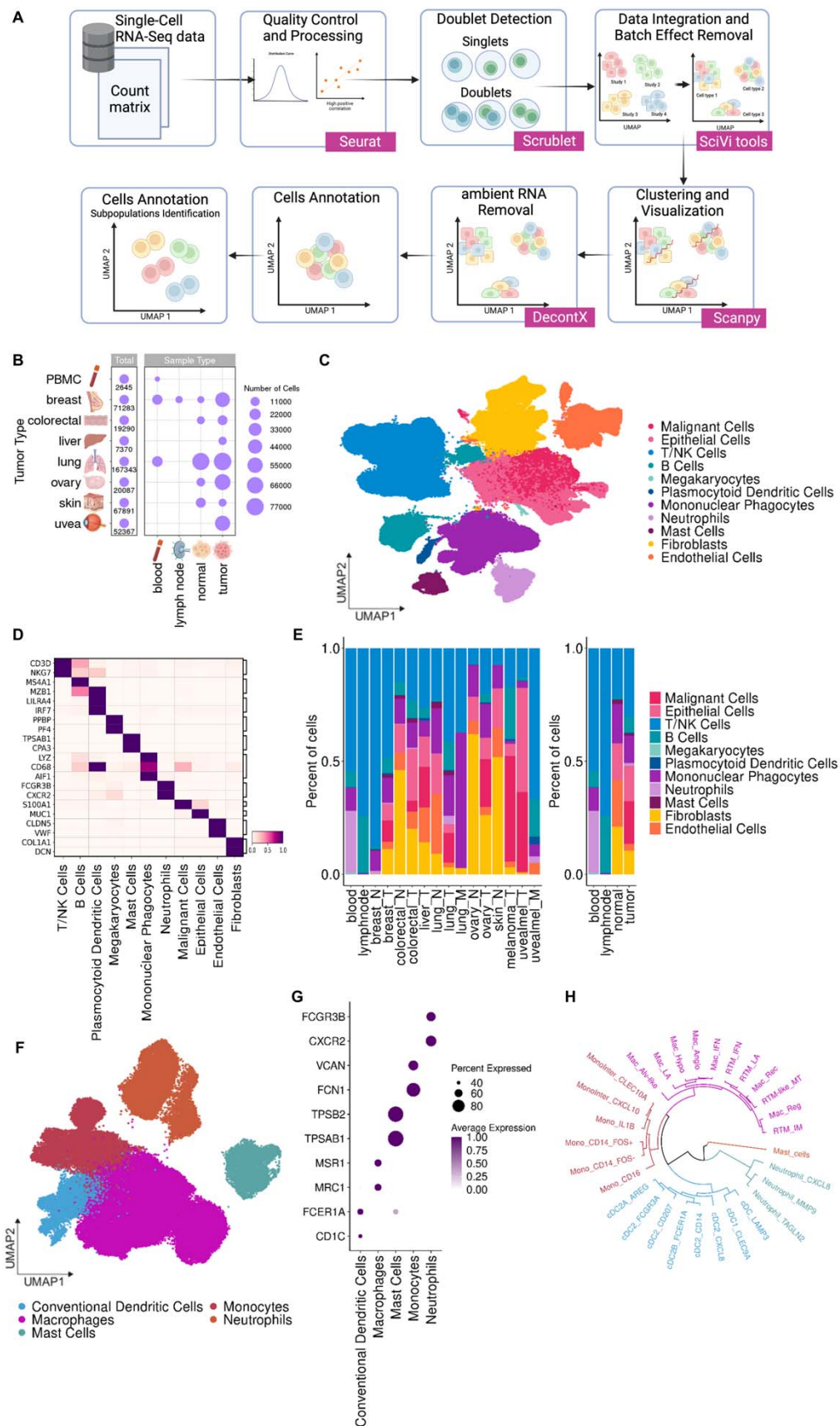
Fig. 1. **Multi-tissue single-cell atlas of tumor and healthy samples.** **(A)** Workflow of experimental design. **(B)** Dot plot showing the distribution of sample types across the eight tissues analyzed in this study. Dot size indicates the number of cells by sample type. **(C)** The latent space generated by scVI was used to construct an unbiased Uniform Manifold Approximation and Projection (UMAP) of 408,476 cells. Broad cell types are color-coded: fibroblasts (yellow), endothelial cells (orange), epithelial (light pink) and malignant cells (pink), pDC cells (dark blue), B cells (green), T and NK cells (blue), neutrophils (lavender), mast cells (eggplant), megakaryocytes (aquamarine), and mononuclear phagocytes (purple). **(D)** Heatmap showing the markers' genes across the 11 broad cell types and the number of cells per type. The color scale reflects the $\log_{10}$-normalized gene expression range. **(E)** Bar plot showing the proportion of clusters across conditions and sample types. Left: Each column represents the sample types for a given condition, if any, such as normal (_N), primary tumor (_T), and tumor metastasis (_M). Right: Proportion of clusters for summarized sample types from the left plot. **(F)** UMAP of 68,259 myeloid-derived cells, color-coded according to cell types. **(G)** Dot plot showing the expression of specific markers by each myeloid-derived subpopulation. Dot size indicates the fraction of expressing cells. **(H)** Dendrogram illustrating the relationship between the mononuclear phagocytes populations.
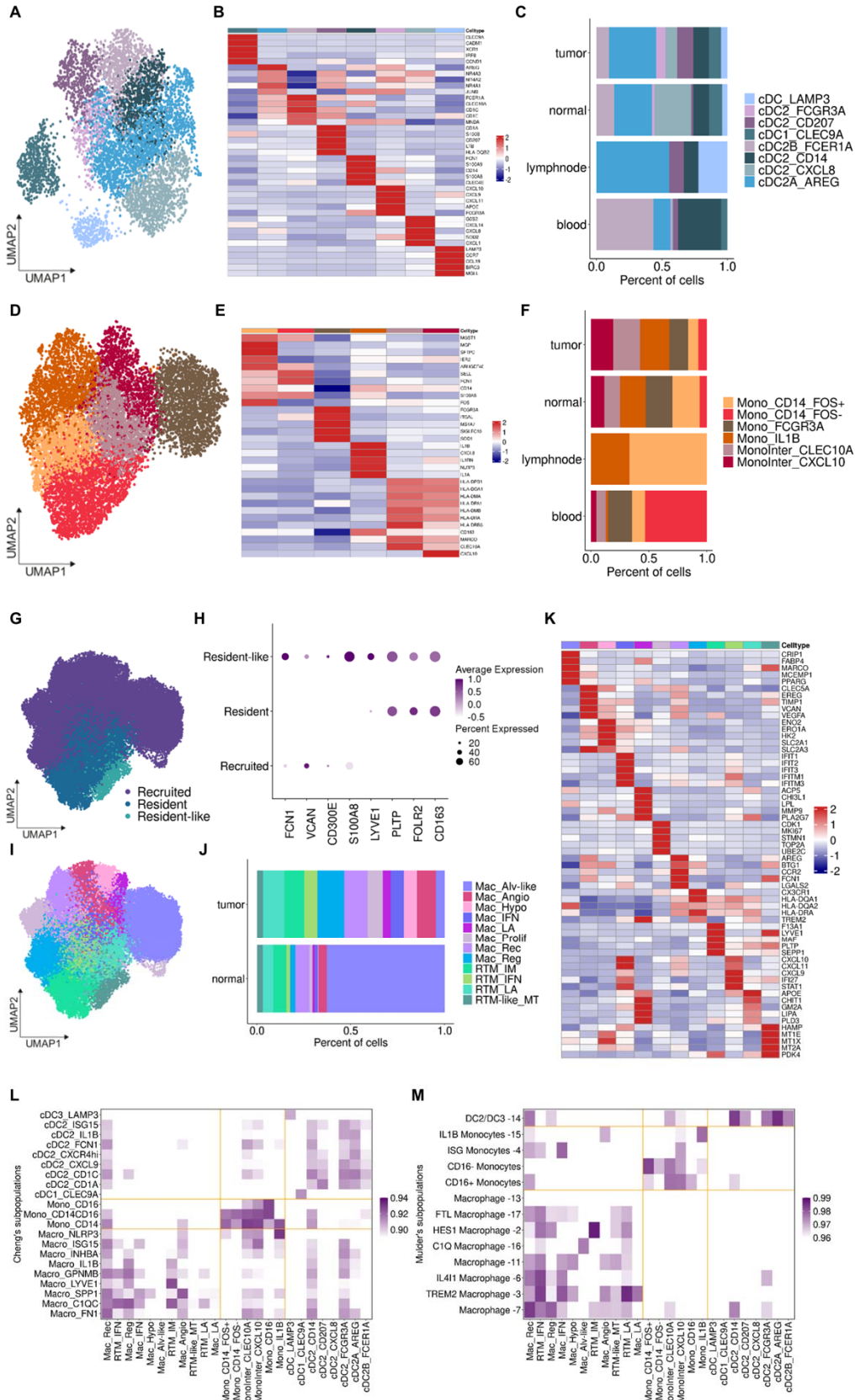
Fig. 2. Mononuclear phagocytes are the major myeloid subpopulations in the TME, displaying several polarization states according to the local cue. **(A)** UMAP of cDC subpopulations colored by the eight states. **(B)** Heatmap showing the average expression of the top five DEGs per cluster. **(C)** Bar plot showing the distribution of cDC cells across sample types. **(D)** UMAP of Mono subpopulations colored by the six states. **(E)** Heatmap showing the average expression of the top five DEGs per cluster. **(F)** Bar plot showing the distribution of Mono across sample types. **(G)** UMAP of Mac color-coded by the ontogeny subdivision. **(H)** Dot plot showing the mean expression of the resident and recruited related markers. Dot size indicates the fraction of expressing cells, colored based on normalized expression levels. **(I)** UMAP represents the twelve Mac states. **(J)** Bar plot showing the distribution of Mac across sample types. **(K)** Heatmap showing the average expression of the top five DEGs per cluster. **(L)** and **(M)** Heatmap showing the correlation between the MDC subpopulations described in the present work and those described by Mulder et al. 2021 (25) and Cheng et al. 2021 (24), respectively.

Figure 3: Macrophages exhibited diverse phenotypes and functional states. Heatmap showing the distribution and score signature of **(A)** M1 and **(B)** M2 markers across the Mac types. **(C)** Ratio of M1 and M2 scores. **(D-I)** Violin plot showing the gene signature of five hallmarks of cancer: **(D)** angiogenesis, **(E)** hypoxia, **(F)** EMT, **(G)** ECM, **(H)** antigen presentation, and **(I)** phagocytosis. **(J)** Heatmap of immunosuppressive gene signature. **(K)** Enrichment analysis of signaling pathways using the Reactome database. **(L)** Heatmap showing main metabolism signature for each Mac phenotype. **(M)** Pseudotime analysis of monocytes, macrophages and dendritic cells derived from monocytes using Monocle 3. **(N)** heatmap showing expression variation of the indicated transcripts (only genes with q-value = 0 and morans_I > 0.25, are depicted. EMT: Epithelial-Mesenchymal Transition; ECM: extracellular matrix.
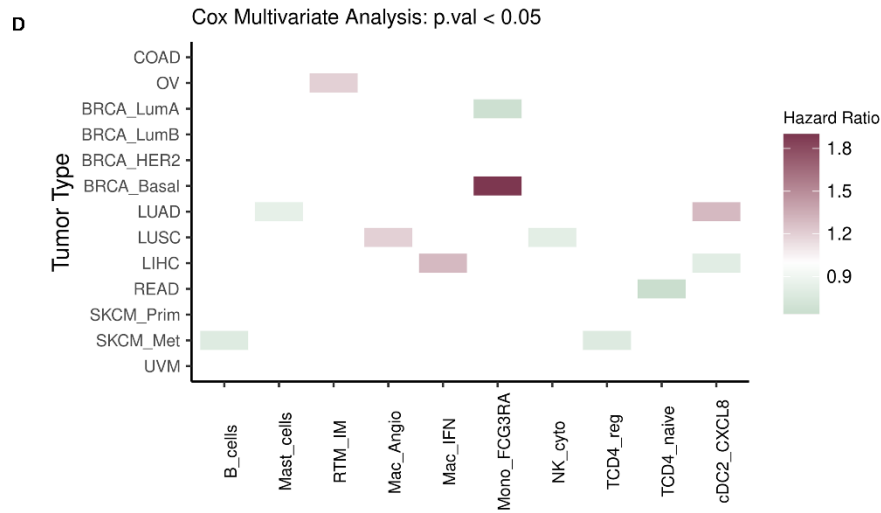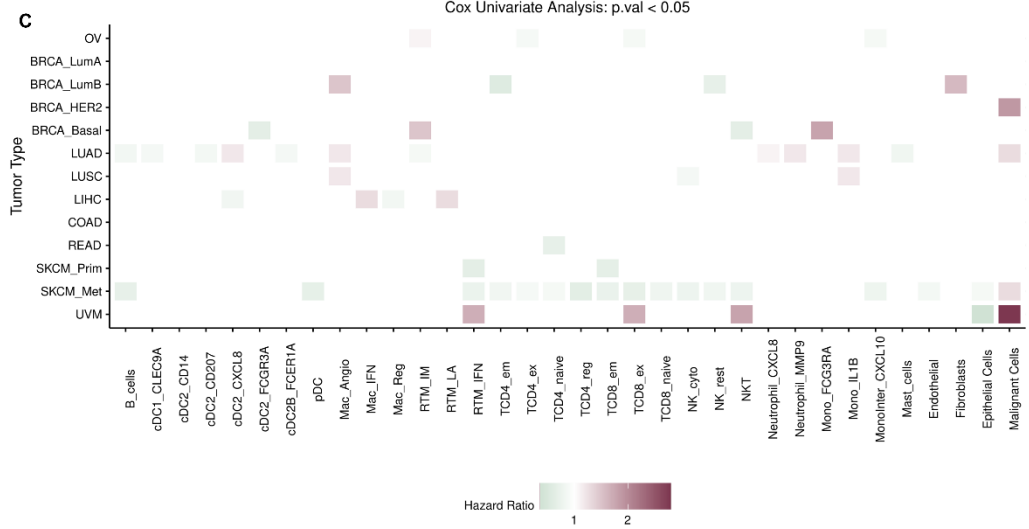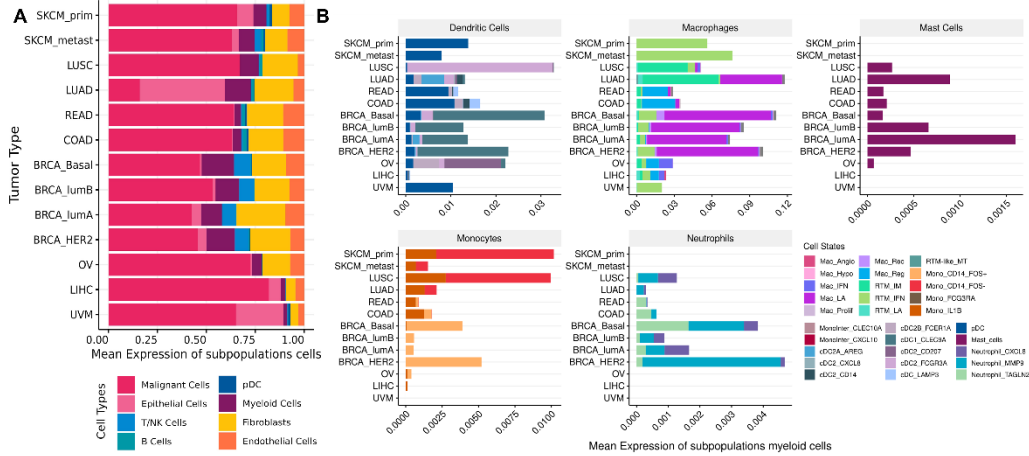
Fig. 4: Macrophages and monocytes are associated with the clinical outcome in different tumor types. (A) Bar plot showing the distribution of cell types across the TCGA database. (B) Bar plot showing the distribution of cell states across the TCGA database. (C) Cox regression univariate and (D) multivariate analysis (p-value log-rank < 0.05).
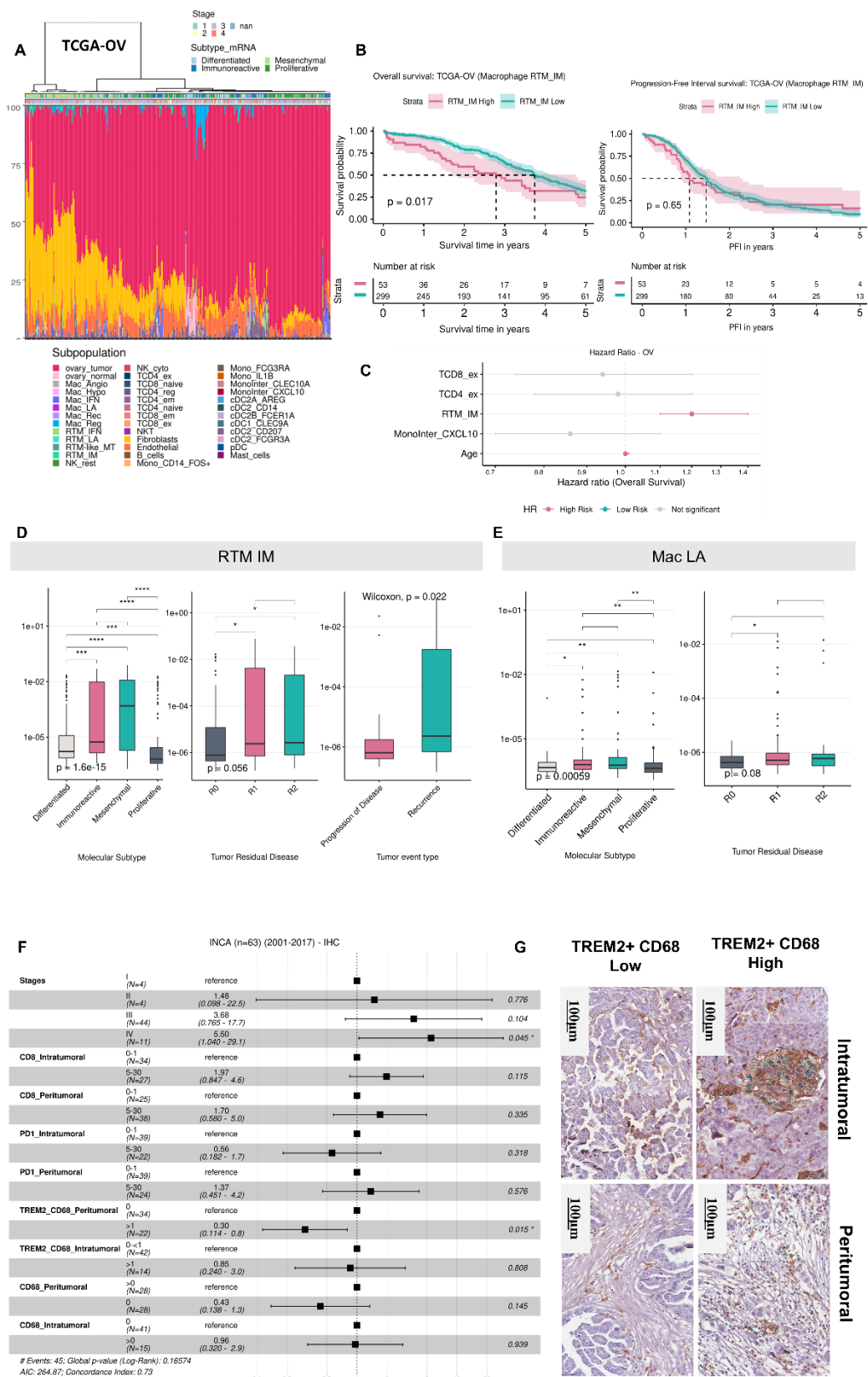
Figure 5: Interstitial Macrophages (RTM_IM) are associated with the clinical outcome in ovary tumors. (A) Bar plot showing the predicted proportion of subpopulation cells for ovarian cancer from TCGA-OV. **(B)** Overall Survival (p-value= 0.01) and Progression-Free Interval (p-value= 0.650 for HIGH and LOW groups of RTM_IM. **(C)** Cox regression multivariate analysis for TCGA-OV cohort. **(D)** Box plots demonstrating the clinical impact of RTM_IM in relation to molecular subtype, and tumor event types in ovary samples. **(E)** Box plots demonstrating the clinical impact of Mac_LA in relation to molecular subtype and tumor residual disease in ovary samples. **(F)** Forest plot representing the survival hazard ratio of relative fractions, divided in quartiles, in the Brazilian cohort of HGSOC. First, the markers CD8[+], PD-1, TREM-2[+] were classified in high and low expression and, then, the percentages of each marker evaluated intratumoral and peritumoral. **(G)** IHC representative of TREM2[+] CD68[+] expression in HGSOC cohort. Image obtained by Aperio ImageScope v12.4.6.5003. Scale bar: 100 µm.
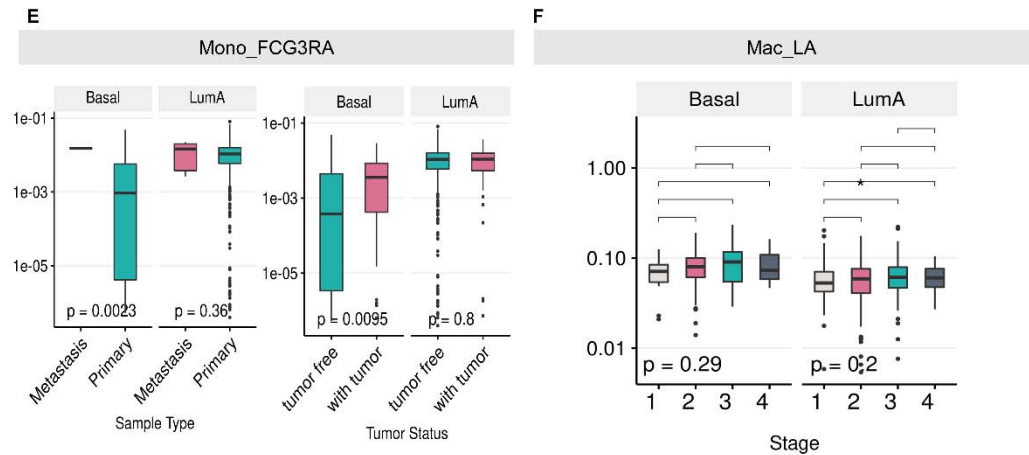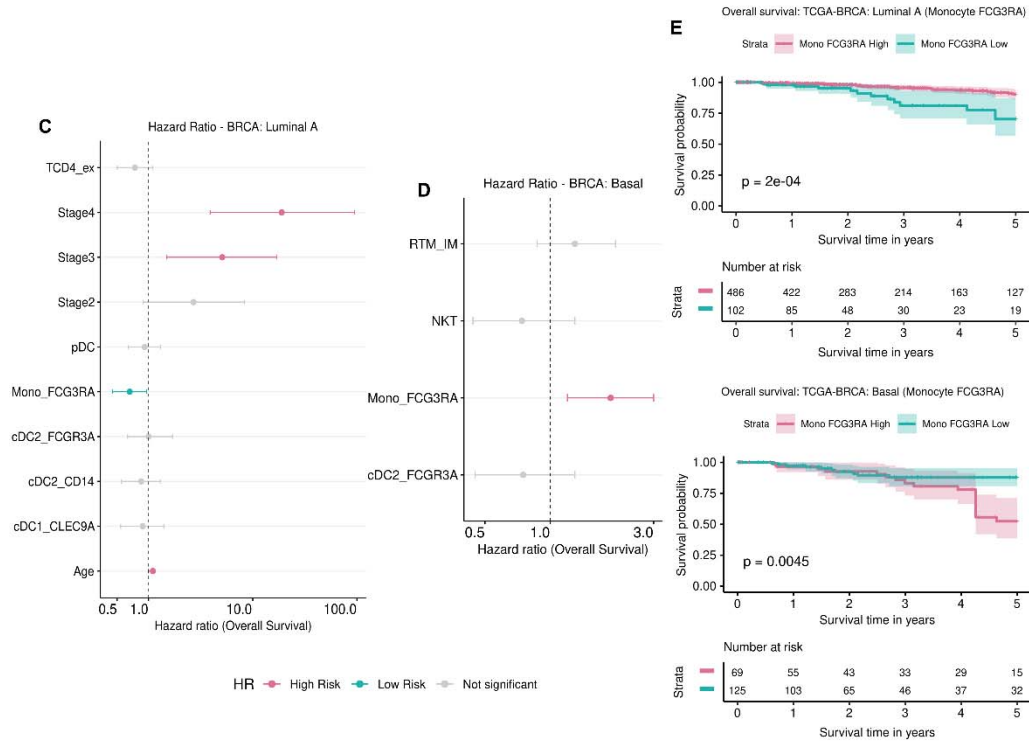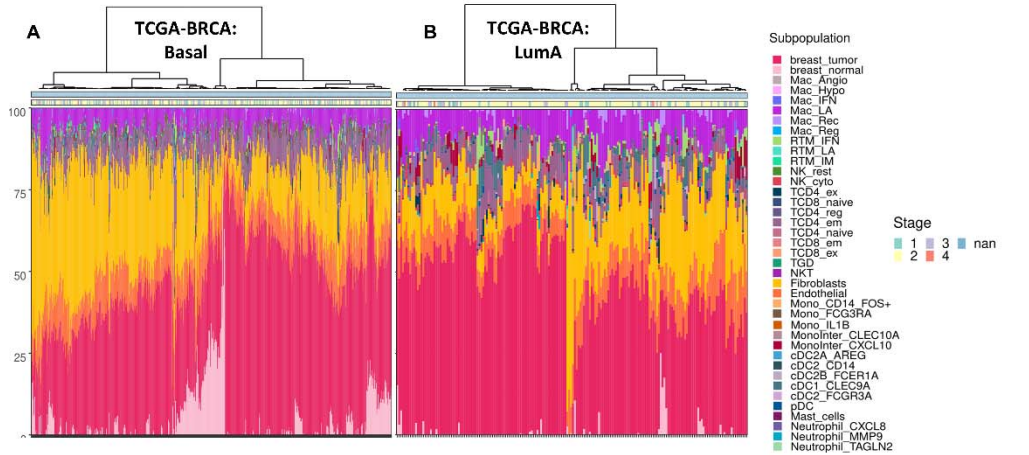
Figure 6: Monocytes FCGR3A (Mono_FCGR3A) are associated with the clinical outcome in breast tumors. (A) Bar plot showing the predicted proportion of subpopulation cells for breast cancer from TCGA-BRCA: Luminal A and **(B)** TCGA-BRCA: Basal. **(C)** Cox regression multivariate analysis for TCGA-BRCA: Luminal A cohort. **(D)** Cox regression multivariate analysis for TCGA-BRCA:Basal cohort. **(E)** Overall Survival for HIGH and LOW groups of Mono_FCGR3A (p-value= $2x10^4$) for Luminal A and Basal subtypes. **(F)** Box plots demonstrating the clinical impact of Mono_FCGR3A in relation to sample type (metastasis and primary situ) and tumor status (tumor free and with tumor) in both breast cancer subtypes Basal and Luminal-A. **(G)** Box plot demonstrating the clinical impact of Mac_LA in relation to stage (1 to 4) in both breast cancer subtypes Basal and Luminal A.