

Regulatory Networks Shaping Human Tumor-Associated Macrophages In Vivo Identify MAFB as a Key Transcriptional Driver

Stanislas Goriely

stanislas.goriely@ulb.be

Université Libre de Bruxelles <https://orcid.org/0000-0002-7005-6195>

Aurélié Detavernier

Université Libre de Bruxelles (ULB)

Emmanuelle Donckier de Donceel

Université Libre de Bruxelles

Francesco Patanè

Université Libre de Bruxelles <https://orcid.org/0009-0001-8619-447X>

Annabelle Pedron

Université Libre de Bruxelles

Valérie Acolty

Université Libre de Bruxelles (ULB)

Muriel Nguyen

Université Libre de Bruxelles

Séverine Thomas

Université Libre de Bruxelles

Vincent Martens

Université Libre de Bruxelles (ULB)

Thila Vanhulst

Université Libre de Bruxelles

Florian Van Horenbeke

Université Libre de Bruxelles

Amandine Collin

Université Libre de Bruxelles

Egor Zindy

Université Libre de Bruxelles

Sanâa El Korchi

Université Libre de Bruxelles

Maxime Melchior

Université Libre de Bruxelles (ULB)

Desiree Venturoli

Université Libre de Bruxelles (ULB)

Florian Szymczak

Université Libre de Bruxelles <https://orcid.org/0000-0002-5431-270X>

Domien Vanneste

Université de Liège

Thomas Marichal

University of Liège <https://orcid.org/0000-0003-1688-6291>

Karen Willard-Gallo

Institut Jules Bordet, Hôpital Universitaire de Bruxelles (H.U.B), Université libre de Bruxelles (ULB)
<https://orcid.org/0000-0002-1150-1295>

Anthony Rongvaux

Fred Hutchinson Cancer Research Center <https://orcid.org/0000-0001-7142-6533>

Martin Bizet

Université libre de Bruxelles (ULB)

Abdulkader Azouz

Université Libre de Bruxelles <https://orcid.org/0000-0001-8626-4370>

Article

Keywords:

Posted Date: January 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8562817/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Regulatory Networks Shaping Human Tumor-associated Macrophages In Vivo Identify MAFB as a Key Transcriptional Driver

Aurélie Detavernier^{1*}, Emmanuelle Donckier de Donceel^{1*}, Francesco Patané^{1*}, Annabelle Pedron¹, Valérie Acolty¹, Muriel Nguyen¹, Séverine Thomas¹, Vincent Martens¹, Thila Vanhulst², Florian van Horenbeke¹, Amandine Collin³, Egor Zindy³, Sanâa El Korchi¹, Maxime Melchior¹, Desire Venturoli¹, Florian Szymczak¹, Domien Vanneste^{4,5}, Thomas Marichal^{4,5,6}, Karen Willard-Gallo², Anthony Rongvaux^{7,8}, Martin Bizet^{1§}, Abdulkader Azouz^{1§}, Stanislas Goriely^{1§}

¹Institute of Medical Immunology (IMI) and Immunobiology lab, ULB Center for Research in Immunology (U-CRI), Gosselies, Belgium

²Molecular Immunology Unit, Institut Jules Bordet, Brussel, Belgium.

³DIAPath, Center for Microscopy and Molecular Imaging, Université Libre de Bruxelles (ULB), Gosselies, Belgium

⁴Laboratory of Immunophysiology, GIGA Institute, University of Liège, Liège, Belgium

⁵Faculty of Veterinary Medicine, University of Liège, Liège, Belgium

⁶Walloon Excellence in Life Sciences and Biotechnology (WELBIO) Department, WEL Research Institute, Wavre, Belgium.

⁷Translational Science & Therapeutics Division, Program in Immunology, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA

⁸Department of Immunology, University of Washington School of Medicine, Seattle, WA 98109, USA

* These authors contributed equally.

§ These authors share senior authorship.

Running title: In vivo modeling of human TAMs

Lead contact:

S. GORIELY, Institute for Medical Immunology, 8 rue Adrienne Bolland, B-6041 Charleroi-Gosselies, Belgium; Phone: 32-2-650-9588; Fax: 32-2-650-9563; E-mail: stanislas.goriely@ulb.be

Abstract

Tumor-associated macrophages (TAMs) are central drivers of resistance to conventional and immune-based therapies. Recent single-cell studies have uncovered an unexpected heterogeneity of macrophage states in human cancers, moving beyond the simplistic “M1–M2” paradigm. However, progress in understanding human TAM biology has been hampered by the lack of physiologically relevant *in vivo* models. Here, we leverage next-generation humanized mouse models implanted with cell line xenografts (CDX) to dissect the transcriptional and epigenetic programs that shape monocyte-to-macrophage differentiation within the tumor microenvironment. We show that these immuno-CDX models faithfully reproduce the phenotypic and molecular hallmarks of TAMs observed in patient tumors. By integrating single-cell transcriptomic and chromatin accessibility profiling with high-dimensional co-expression and enhancer-driven regulatory network analyses, we reconstruct a comprehensive framework of human TAM states. Furthermore, CRISPR-mediated gene editing of hematopoietic stem cells prior to humanization identifies MAFB as a key regulator of TAM plasticity. Our findings establish humanized models as a powerful platform to study TAM biology *in vivo*, uncover fundamental regulators of macrophage plasticity, and highlight new therapeutic targets for next-generation cancer immunotherapies.

Introduction

Tumor-associated macrophages (TAMs) are increasingly recognized as central orchestrators of the tumor microenvironment, contributing to nearly all hallmarks of cancer¹. Across diverse malignancies, high macrophage infiltration consistently associates with poor clinical outcomes and resistance to both conventional therapies and immune checkpoint blockade. Consequently, TAMs have emerged as attractive therapeutic targets, with strategies ranging from depletion to functional reprogramming². Yet, despite encouraging activity in preclinical mouse models, TAM-directed agents have shown limited clinical benefit, with monotherapies rarely exceeding ~5% overall response rates³. This translational gap underscores a critical need for a deeper understanding of TAM biology and heterogeneity in the human setting⁴. However, this has been hindered by the lack of appropriate in vitro and in vivo models.

Macrophage identity is profoundly shaped by the tumor milieu, through cues that reprogram monocytes and tissue-resident cells into TAMs. Historically, these cells were described as “M2-like”, reflecting a simplistic binary polarization framework. However, single-cell transcriptomic studies in both mice and humans have revealed a far greater diversity of macrophage states, governed by complex transcriptional and epigenetic programs that extend well beyond the M1–M2 paradigm^{5,6}. Defining these regulatory networks is essential for developing targeted interventions and patient stratification strategies.

Here, we leverage advanced humanized mouse models bearing cell line–derived xenografts (Immuno-CDX) to interrogate the molecular logic of human TAM differentiation. By integrating single-cell transcriptomic and epigenomic profiling with enhancer-driven regulatory network analyses, we reconstructed a comprehensive framework of human TAM states. Our findings identify MAFB as a central transcriptional regulator of TAM identity, providing new mechanistic insight into human macrophage plasticity in cancer and pointing toward novel therapeutic strategies.

Results

Reconstitution of the tumor immune microenvironment in humanized mice

We developed a humanized mouse platform to study human myeloid cell biology *in vivo*, using the MISTRG strain reconstituted at birth with fetal-derived CD34⁺ hematopoietic stem cells (HSCs) (Fig. S1A). At equivalent irradiation doses (80 cGy) and HSC input, MISTRG and NBSGW mice displayed comparable levels of overall human hematopoietic engraftment in peripheral blood eight weeks post-transplantation (Fig. S1B). However, consistent with prior reports⁷, lineage analysis revealed a striking advantage of the MISTRG model: only these mice supported efficient reconstitution of human myeloid and NK cell compartments (Fig. S1C). Notably, MISTRG mice generated circulating monocyte subsets—including CD14⁺ classical, CD14⁺CD16⁺ intermediate, and CD16⁺ non-classical populations—that closely mirror those observed in human blood (Fig. S1D). These findings validate MISTRG mice as a physiologically relevant system to dissect the biology of human mononuclear phagocytes (MNP).

Having established robust myeloid reconstitution, we next asked whether these cells infiltrate tumors in xenograft models. We generated cell line-derived xenografts (CDX) using melanoma (Mel-LB33, SK-Mel-237), colon carcinoma (LS174T), and breast cancer (MDA-MB-231) cell lines. Flow cytometry of dissociated tumor tissues revealed higher frequencies of CD14⁺ myeloid cells in MDA-MB-231 and SK-Mel-237 tumors compared to LS174T and Mel-LB33. Next, multiplex IHC across these four immuno-CDX models demonstrated the presence of human CD68⁺ macrophages in adjacent tissues and confirmed the consistent infiltration within the tumor microenvironment of SK-Mel-237 and MDA-MB-231 models. A fraction of these infiltrating cells expressed heme oxygenase 1 (HO-1), a canonical marker of tumor-associated macrophages (TAMs)^{8–11} (Fig.1C). Of note, HO-1⁺ macrophages were enriched close to necrotic regions, consistent with stress and hypoxia-driven TAM activation. Together with previous reports¹², these results establish a robust humanized system to explore the differentiation, function, and reprogramming of human TAMs *in vivo*.

We next isolated total human CD45⁺ or myeloid CD33⁺CD14⁺ cells infiltrating tumor tissues, together with matched populations from control organs (bone marrow, spleen, liver, lung) of the same tumor-bearing MISTRG mice, for single-cell RNA sequencing (scRNA-seq). To provide a complementary reference, we also analyzed cells from the bone marrow, spleen, and tumor tissue of humanized NSG mice orthotopically implanted with MDA-MB-231 breast cancer cells. To enhance myeloid reconstitution in this model, NSG mice were transiently

boosted with human cytokines (IL-3, IL-4, IL-15, GM-CSF, FLT3L) via hydrodynamic injection one week prior to tumor engraftment (Fig. S1E). The datasets generated across these experiments, composed of 165K cells, were integrated to establish a comprehensive atlas of human hematopoietic and myeloid populations in tumor-bearing HIS mice (Fig. 1D). We identified and annotated 29 distinct clusters spanning both lymphoid and myeloid lineages, including B cells, CD4⁺ T cells (conventional and regulatory), CD8⁺ T cells, NK cells, plasmacytoid and conventional dendritic cells, monocytes, macrophages, bone marrow-derived progenitors, and mast cells. This demonstrates robust multi-lineage reconstitution of the human immune system in our models and the ability of these populations to infiltrate the tumor microenvironment.

Mono/Macro subsets from immuno-CDX models capture the full spectrum of human TAM states

We next focused on the monocyte/macrophage compartment and identified nine transcriptionally distinct clusters (Fig. 2A). At this stage, we also removed cycling cells and progenitors to simplify the dataset. Among the clusters, we observed classical (VCAN⁺) and non-classical (CX3CR1⁺) monocytes, defined by the expression of *VCAN*, *S100A8/9/12* and *CX3CR1*, *LYN*, respectively. Several intermediate populations were also present, co-expressing canonical monocyte and macrophage markers. These included monocytic cells with high levels of MHC class II molecules (HLA⁺ Mono), cells enriched in cytokines, chemokines, and angiogenesis-related transcripts (IL1B⁺ Mono, expressing *IL1A/B*, *CXCL2/3/8*, *CCL3/4*, *AREG*, *EREG*), and a subset characterized by interferon-responsive genes (CXCL10⁺ Macs, expressing interferon-stimulated genes [ISGs], *CXCL10*, *CD274*, *IDO1*). Additional clusters exhibited broad macrophage gene expression (*CD68*, *C1QA/B/C*, *CD163*, *APOE*, *APOC1*, *MSR1*), and were further distinguished based on their transcriptional states. These included C1Q⁺ Macs (*C1QA*, *C1QB*, *CD163* and genes related to phagocytosis), MT1G⁺ Macs (Metallothionein family genes, *MARCO*, *FTH1* and genes related to response to ROS), TREM2⁺ Macs (*TREM2*, *SPP1*, *TGFB2*, *MMP9* and genes related to lipid metabolism), and GLUT1⁺ TAMs, which expressed genes associated with glucose metabolism and hypoxia (*SLC2A1*, *HILPDA*, heat shock proteins [HSPs], *GPNMB*) (Fig. 2A–C). Projection of these clusters onto the MoMac-VERSE reference atlas¹³ showed strong concordance with macrophage states previously defined in human tissues, supporting the notion that our immuno-CDX models faithfully recapitulate the continuum of human monocyte-to-macrophage differentiation (Fig. 2D).

Notably, myeloid populations in tumors segregated from their counterparts in healthy tissues (Fig. 2E). VCAN^+ and CX3CR1^+ monocytes predominated in non-tumoral organs, whereas TAM-associated states (GLUT1^+ , TREM2^+ , MT1G^+) were enriched in tumors (Fig. 2F), highlighting tumor-driven imprinting of the myeloid compartment. The relative abundance of these subsets varied across the four tumor models analyzed. In MDA-MB-231 tumors, TREM2^+ and C1Q^+ TAMs were the dominant populations, whereas LB-33 tumors contained higher proportions of HLA^+ monocytes and SK-MEL-37 tumors were enriched in CXCL10^+ macrophages. Comparison of MDA-MB-231 tumors grown in MISTRG versus cytokine-boosted NSG (NSG-B) mice revealed a marked reduction in immature subsets (VCAN^+ , CX3CR1^+ , HLA^+ monocytes) in NSG-B, suggesting that the cytokine boost accelerates their differentiation into macrophages (Fig 2G).

Importantly, these transcriptional signatures showed distinct associations with patient outcome in TCGA datasets. As previously reported, the *CXCL9/SPP1* ratio correlated with survival across multiple cancers¹⁴. TREM2^+ , MT1G^+ and GLUT1^+ TAMs and IL1B^+ monocyte signatures were linked to poor prognosis, while CX3CR1^+ and HLA^+ monocytes generally correlated with favorable outcomes (Fig. 2H).

To map TAM spatial organization and evaluate the translational relevance of our immuno-CDX models, we performed multiplex immunohistochemistry (mIHC) in breast cancer (BC) patient samples and in orthotopic MDA-MB-231 tumors from MISTRG mice. Guided by our transcriptomic data (Fig. 3A) and prior literature^{15,16}, we selected four markers to define TAM subtypes: SPP1 (glycolytic/lipid-associated), NLRP3 (inflammatory), IL4I1 (phagocytic/ISG+), and FOLR2 (tissue-resident), in addition to CD68 as a pan-macrophage marker. In BC samples, at least one of these markers was expressed in ~50% of CD68^+ cells (Fig. 3B). Overlap was limited, with most cells expressing a single marker, supporting their utility in distinguishing discrete TAM subsets. This combination thus captures key dimensions of TAM diversity in situ. Across normal breast, DCIS (stage 0), and IDC, we observed a stage-dependent increase in CD68^+ macrophages (Fig. 3C). In normal breast, only FOLR2^+ cells were detected, whereas all four subsets were present in tumors (Fig. 3D). As previously described, FOLR2^+ macrophages were confined to the stroma, furthest from the tumor nests¹⁷. By contrast, IL4I1^+ , NLRP3^+ , and SPP1^+ macrophages localized both in the stroma and within tumor regions, with SPP1^+ cells positioned closest to malignant cells (Fig. 3E). Moreover, FOLR2^+ macrophages decreased with progression, whereas IL4I1^+ subset expanded (Fig. 3F). We also evaluated HO-1 expression in IDC samples (Fig. S2A). HO-1^+ macrophages accumulated in

tumor nests and were broadly distributed across subsets, suggesting integration of stress-associated signals alongside differentiation processes (Fig S2B).

In MDA-MB-231 immuno-CDX tumors, these markers were also largely mutually exclusive (Fig. 3G). IL4I1⁺, NLRP3⁺, and SPP1⁺ macrophages were clearly detected (Fig. 3H), although without clear subregional localization within the TME (Fig. 3I). FOLR2⁺ macrophages were rare, consistent with their origin from tissue-resident populations and their exclusive presence in the stroma rather than the tumor nest (Fig. 3H). Again, when looking at HO-1 which was enriched close to the necrotic core (Fig. 1D), we did not observe clear subset preference. Altogether, these findings provide a robust translational framework to bridge preclinical modeling with patient biology, and to dissect the molecular and spatial determinants of TAM identity and function.

Gene regulatory networks reveal early transcriptional control of TAM differentiation

Cluster-based analysis revealed substantial transcriptional promiscuity and overlap among mononuclear phagocyte states. This continuum and plasticity of differentiation challenged rigid classification schemes and underscored the need for methods capable of capturing shared regulatory programs across transitional cell states. Consistent with this, our in situ data showed that stress-associated markers such as HO-1 were not confined to a single subset. Therefore, to investigate the regulatory architecture shaping TAM identity, we employed *hdWGCNA*¹⁸. This method allowed us to construct a gene-gene correlation network and identify five modules of co-expressed genes (Fig. 4A), highly preserved in the MoMac-VERSE atlas (Fig. S3A) and associated with distinct biological pathways (Fig. S3B). Score for module M-5 is highest in VCAN⁺ mono while M-2 activity culminates in HLA⁺ mono and MT1G⁺ macs clusters (Fig. 4A). Both modules were preferentially active in myeloid populations from healthy tissues, whereas module M-1 (and to a lesser extent M-3 and M-4) was strongly enriched in TAMs as compared to myeloid populations from healthy tissues (Fig. 4A). Consistently, the M-1 score was highest in TAM-associated states such as GLUT1⁺ and TREM2⁺ macrophages (Fig. 4B). Importantly, cluster-by-cluster analysis showed that M-1 genes were upregulated in tumor-derived cells across all nine monocyte/macrophage subsets, including immature VCAN⁺ monocytes, indicating that tumor cues impose a transcriptional program early in the monocyte-to-macrophage trajectory (Fig. 4C).

We identified the transcription factors (TFs) expressed in each module (Fig. S3C) and then inferred which TFs might play dominant regulatory roles based on their regulon activity

scores, focusing on the TAM-associated M-1 module (Fig. 4D). Expression of these TFs was consistently upregulated in TAMs subsets compared to their healthy tissues' counterparts (Fig. S3C). They not only positively regulate M-1 genes, thereby reinforcing TAM-associated transcriptional programs, but also contribute to the activation of M-4 and suppression of M-2 (Fig. 4D). Several of them have well-established roles in myeloid biology, including RBPJ, MAF, NFAT, PRDM1, ETS family members, and NR4A nuclear receptors. Their coordinated activity suggests a regulatory axis through which tumor-derived signals promote TAM identity by stabilizing M-1-driven gene expression and modulating the broader transcriptional landscape to favor TAM-associated states while repressing alternative myeloid programs. Notably, the regulon activity scores of several M-1-associated TFs, including MAFB and BHLHE40, were clearly upregulated in VCAN⁺ monocytes from tumor tissues compared to healthy controls, further supporting their role in driving early TAM identity in response to tumor-derived cues (Fig. 4E).

For modules enriched in healthy tissues, we identified PU.1 and members of the KLF, AP-1, and IRF families (M-2-related), as well as NFE2, POU2F2, and ARID3A (M-5-related), as potential key regulators (Fig. S3C-E). To further explore the regulatory logic linking these modules, we inferred potential inter-module relationships. This analysis revealed a structured network of transcriptional influence: TAM-associated module M-1 reinforced M-4 while inhibiting M-2; M-3 reinforced M-1 and concurrently suppressed M-2, M-4, and M-5; conversely, VCAN⁺ monocyte-associated module M-5 exerted an inhibitory effect on M-1 (Fig. 4F). These interactions highlight a dynamic and antagonistic regulatory landscape, in which TAM-associated programs are both promoted and constrained by competing transcriptional modules.

To identify potential bifurcations during monocyte-to-TAM differentiation, we applied the *Slingshot* algorithm, which uses pseudotime analysis to infer developmental trajectories. Using VCAN⁺ monocytes as the starting population, we identified three distinct trajectories progressing toward GLUT1⁺ (1), MT1G⁺ (2) or CXCL10⁺ (3) Macs (Fig. 4G). These analyses revealed that the transcriptional state corresponding to the HLA⁺ Mono cluster acts as a bifurcation hub, from which cells diverge toward distinct TAM-associated endpoints. This suggests that early transcriptional reprogramming within HLA⁺ monocytes may represent a critical decision point in TAM fate specification. Along these trajectories, we tracked module activity dynamics, observing decrease in M-5 activity early on, followed by preferential usage of M-1 (and M-4) along trajectory 1, whereas M-2 activity predominated along the other two trajectories (Fig. 4H). Taken together, our analysis reveals that tumor-derived signals impose a

TAM-associated transcriptional program early in monocyte differentiation, driven by a coordinated network of TF that stabilize M-1 activity and remodel inter-module interactions, ultimately guiding cells toward distinct TAM states along defined trajectories.

Stepwise modification in chromatin accessibility and TF activity along monocyte to TAM differentiation

To gain deeper insight into the regulatory mechanisms driving TAM differentiation, we integrated scRNA-seq and scATAC-seq datasets from bone marrow (BM)– and tumor-derived MNPs. This multi-omic approach enabled us to study both transcriptional states and the dynamics of chromatin accessibility, providing a more comprehensive view of TF activity across cellular states. We first used ChromVAR to quantify TF motif accessibility across clusters, allowing us to infer TF activity in MNPs from healthy and tumor contexts (Fig. 5A).

We observed an early and consistent increase in the activity of AP-1, NRF, NF- κ B, and MAF families in VCAN⁺ monocytes from tumors compared to BM cells, accompanied by a decrease in CEBP and KLF activities (Fig. 5A and Fig. S4A&B). These changes likely reflect the integration of multiple tumor-derived cues, including oxidative and metabolic stress and cytokine signaling^{19,20}. As differentiation progressed toward C1Q⁺ or TREM2⁺ macrophages, we noted strong enrichment for RBPJ and NFATC2 motifs, followed by increased MITF and HSF1 motif accessibility in GLUT1⁺ macrophages (Fig. 5A and Fig. S4A&C). Overall, TAMs exhibited heightened activity of TF motifs from the M-1 and M-4 modules, highlighting their central role in TAM biology as both regulatory factors and targets. In contrast, motifs associated with the M-2 module, including CEBP family members, showed reduced activity in TAMs. Notably, CXCL10⁺ macrophages displayed specific enrichment of IRF motifs, suggesting a distinct interferon-responsive regulatory program (Fig. 5A and Fig. S4A).

To refine the identification of upstream regulators, we applied SCENIC+, which integrates gene expression, motif enrichment, and chromatin accessibility to infer TF–target relationships with higher confidence²¹. This analysis identified a broad set of M-1–associated transcription factors as key regulators of TAM identity, including ATF3, ZNF331, NFIC, TCF4, MAFB, BNC2, PRDM1, CREM, EPAS1, HES1, BHLHE40, ETV5, and MITF. (Fig. 5B) Many of these TFs showed increased regulon activity in tumor-derived myeloid cells, further supporting their role in driving TAM-associated transcriptional programs. Finally, mRNA expression of these TF was also upregulated along trajectory 1 that culminates in GLUT1⁺ Macs (Fig. S4C).

MAFB acts as a central transcriptional hub in human TAMs

Based on our integrative multi-omic analyses—including scRNA-seq, scATAC-seq, ChromVAR, and SCENIC+, we identified MAFB as a potential central transcription regulator within the TAM-associated M-1 module. MAFB showed increased regulon activity in tumor-derived myeloid cells, elevated motif accessibility in TAMs, and was consistently predicted as an upstream regulator of TAM identity. To validate its relevance in situ, we evaluated MAFB and cMAF expression by multiplex immunohistochemistry in human breast cancer lesions. This confirmed the presence of MAFB⁺ macrophages within the tumor microenvironment, whereas these subsets were absent from normal breast tissue (Fig. 6A). MAFB was over-expressed at comparable levels across TAM subsets in both breast cancer patients and immuno-CDX MISTRG tumors compared to healthy tissues (Fig. 6B). In contrast, cMAF was expressed in both healthy and tumor-associated macrophages, suggesting that MAFB, but not cMAF, is selectively induced in the tumor context (Fig. S5A&B).

Given its consistent association with TAM transcriptional programs and its tumor-specific expression pattern, we therefore focused on MAFB as a candidate regulator of TAM identity. Regulon analysis revealed that MAFB acts upstream of other TFs, including cMAF, SERBF, and RBPJ, suggesting that it may serve as a master regulator coordinating multiple layers of transcriptional control during TAM differentiation (Fig. 6C). To directly assess its genomic binding profile, we performed CUT&RUN in sorted TAMs isolated from MDA-MB-231 immuno-CDX tumors, identifying thousands of MAFB-bound regulatory elements enriched for MAF recognition elements (MARE motifs), as well as PU.1 (a known pioneer factor, that can directly interact with MAFB²²), STATs, and SMADs (Fig. 6 D,E). These targets strongly overlapped with TAM-specific open chromatin regions (bulk ATAC-seq) (Fig. 6F, left panel) and showed increased accessibility in monocyte clusters from the tumor and along TAM differentiation (scATAC-seq) compared to monocytes in bone marrow (Fig. 6F, right panel). Comparison with in vitro monocyte-derived macrophages revealed that ~60% of MAFB sites were shared (Fig. 6G, left panel), while TAM-specific regions were enriched for SMAD and STAT motifs (Fig. 6G, right panel), suggesting that cytokines such as IL-10, IL-6, and TGFβ reroute MAFB genomic occupancy in the tumor microenvironment. Notably, MAFB-bound regions were enriched for M-1 and M-4 module genes, including loci such as *TGFB2*, *SPPI*, *CIQB*, and *MRC1*, which displayed strong MAFB occupancy and chromatin accessibility (Fig. 6H-I). In conclusion, these experiments indicate that MAFB occupies regulatory regions enriched for TAM-specific gene modules. Its binding profile shifts in the tumor

microenvironment, suggesting a role in integrating lineage and cytokine-driven signals to shape TAM identity.

Functional inactivation reveals MAFB as a determinant of TAM identity

To functionally interrogate the role of MAFB *in vivo*, we performed CRISPR-mediated knockout in human HSCs prior to engraftment and implantation of MDA-MB-231 CDX, achieving >95% editing efficiency as confirmed by mRNA analyses in TAMs (Fig. 7A). We also validated decreased expression of MAFB in TAMs at the protein levels by mIHC (Fig. 7B). *MAFB* knockout led to a shift in MNP composition in tumor-infiltrating cells: reduced TREM2⁺, IL1B⁺, and CXCL10⁺ populations, with increased HLA⁺ monocytes and MT1G⁺ macrophages (Fig. 7C,D). To further characterize the impact of MAFB loss, we performed differential gene expression analysis between wild-type and *MAFB* knockout cells. This revealed consistent transcriptional changes both in bulk and cluster-specific comparisons (Fig. 7E and Fig. S6A). Key macrophage-associated genes such as *MERTK*, *SPPI*, *CTSL*, and *HMOX1* were downregulated, along with several chemokine genes (Fig. 7E). Globally, these MAFB-dependent genes mapped to pathways involved in tumor progression and immune modulation (Fig. S6B). In contrast, we observed a striking upregulation of HLA-related genes, as well as *CD1A*, *CD1C*, and *MT1G*-associated transcripts, indicating a shift toward a more antigen-presenting and stress-responsive phenotype. These two signatures showed consistent and opposite associations with patient outcome in TCGA datasets (Fig. 7F). Expression of several TF were also modulated by the absence of MAFB (Fig. S6C). Module-level analysis revealed a global reduction in M-1 and M-4 activity and an increase in M-2 and M-3, underscoring MAFB's role in TAM plasticity and validating predictions from *hdWGCNA* (Fig. 7G). To identify direct transcriptional targets, we applied the *BETA* algorithm, integrating CUT&RUN binding profiles with differential expression data (Fig 7H). This confirmed that MAFB directly regulates key components of TAM identity across multiple modules. In summary, these findings demonstrate that MAFB is essential for establishing TAM identity and functional specialization, acting as a critical determinant of their transcriptional programs and phenotypic diversity.

Discussion

Deciphering the mechanisms that govern human TAM differentiation remains a central challenge in cancer immunology. In this study, we leveraged next-generation humanized immuno-CDX models, which enable the engraftment of human hematopoietic stem cells and the development of a functional human myeloid compartment within a tumor context. These models offer a unique opportunity to interrogate the *in vivo* dynamics of human TAM differentiation, bridging the gap between mechanistic studies and clinical relevance.

A major strength of our approach lies in the integration of multi-omic data, combining single-cell transcriptomics, chromatin accessibility, transcription factor activity, and functional perturbation. This allowed us to move beyond descriptive profiling and toward a systems-level understanding of the regulatory architecture that shapes TAM identity. By integrating scRNA-seq and scATAC-seq, we captured both the transcriptional states and epigenetic landscape of monocyte-to-macrophage differentiation, revealing early transcriptional priming and context-specific regulatory programs. We identified multiple relevant clusters of monocyte-derived TAM states in our immuno-CDX models; however, we likely miss stromal FOLR2⁺ tissue-resident macrophages, which represents a limitation of implantable models and highlights the need for complementary approaches to fully capture the diversity of TAM populations. Our data reveal substantial transcriptional promiscuity and overlap among mononuclear phagocyte states, underscoring the need for analytical frameworks that can capture this plasticity. Importantly, cluster-based approaches, while informative, may be context-specific: markers such as HO-1 were broadly distributed across subsets *in situ*, angiogenic signatures were expressed by distinct populations depending on the microenvironment²³, and TREM2⁺ or FOLR2⁺ macrophages exhibited either immunosuppressive or immunostimulatory roles across cancer types^{17,24–26}. Network-based approaches were therefore essential to identify shared regulatory modules and bifurcation points in TAM differentiation. These analyses revealed that multiple TFs act coordinately to reinforce TAM-specific states while repressing alternative fates and enhancer accessibility patterns, suggesting that the M-1 module might be highly resistant to single-target perturbation.

Among the transcriptional regulators identified, MAFB emerged as a central node in the TAM-associated M-1 module. MAFB, a member of the large MAF transcription factor family, is specifically expressed in the monocyte/macrophage lineage and plays critical roles in hematopoietic stem cell biology, monocyte commitment, and terminal macrophage differentiation²⁷. Recent CRISPR screen in IL-4-stimulated mouse BM-derived macrophages

identified MafB, as one of the few hits that significantly modulated enhancer accessibility²⁸. It regulates genes involved in immune modulation and efferocytosis, including C1QA and MSR1, and contributes to anti-inflammatory programming and suppression of interferon responses^{29,30}. Beyond normal hematopoiesis, MAFB dysregulation has been linked to hematological malignancies, where it influences leukemic self-renewal and disease progression, pointing to its ability to disrupt cellular homeostasis when dysregulated²⁷. In our study, MAFB was selectively upregulated in TAMs, acted upstream of other TFs, and was essential for maintaining key aspects of TAM identity. While its role in macrophage biology and inflammatory conditions is established, its function within the tumor microenvironment remains largely unexplored. Functional knockout of MAFB in human HSCs prior to engraftment led to a redistribution of tumor-infiltrating macrophage subsets and a shift toward antigen-presenting and stress-responsive states. Thus, in vivo targeting of MAFB, as recently shown for ZEB2²⁸, may reprogram TAMs toward antitumoral state. Together, these results position MAFB as a master regulator of TAM plasticity and transcriptional identity. Future studies should explore how MAFB integrates environmental signals and whether its modulation can synergize with immunotherapies to promote anti-tumoral responses.

Through this approach, we also identified a broader set of TFs, including RBPJ, NFATc2, BHLHE40, ZNF331, MITF, and others, strongly associated with TAM-specific modules and chromatin accessibility patterns. These candidates represent promising targets for functional interrogation, with potential to uncover additional layers of regulation and therapeutic leverage points. The ability to directly compare our findings to human breast cancer samples further strengthens the translational relevance of this framework and sets the stage for future strategies aimed at reprogramming TAMs to support anti-tumor immunity.

Material and Methods

Human Hematopoietic Stem and Progenitor Cells (HSPCs)

Anonymized human fetal liver tissues from 14 to 22 weeks gestational age were obtained with informed consent from the donors through Cercle Allocation Services Inc, Irvine, USA, with approval of the Ethics Committee of H.U.B. – Hôpital Erasme (P2020/009). FL samples were cut into small pieces and incubated with collagenase D (Riche, 100 ng/ml) for 45 min. Red blood cells were lysed using ACK (Ammonium-Chloride-Potassium, home-made) lysing buffer. Cell suspensions from were then purified by density gradient centrifugation (Lymphoprep™, Stemcell Technologies) followed by positive magnetic selection using anti-human CD34 microbeads (Miltenyi, 130-100-453). Purity of the CD34⁺ cells suspension was checked by flow cytometry (>90% CD34⁺ cells) and cells frozen in FBS containing 10% DMSO and vials stored in liquid nitrogen for later use.

Mice

NBSGW (NOD.Cg-*Kit*^{W-41J}*Tyr*⁺*Prkdc*^{scid}*Il2rg*^{tm1Wjl/ThomJ}, RRID: IMSR_JAX:026622) mice³¹ were purchased from The Jackson Laboratory. MISTRG mice (M-CSF^{h/h} IL-3/GM-CSF^{h/h} SIRPα^{h/m}TPO^{h/h}*Rag2*^{-/-}*Il2rg*^{-/-})⁷ were kindly provided by Anthony Rongvaux lab and Dirk Amsen lab. These mice are based on a *Rag2 Il2rg* double knockout 129xBALB/c (N2) background, in which several cytokine-encoding genes (*Csf1*, *Il3/Csf2*, and *Thpo*) have been humanized by replacing the corresponding mouse alleles with their human orthologs from the start to the stop codon. The extracellular domain of SIRPα was also knocked in using the human gene sequence. Mice aged between 8 and 15 weeks were used for the experiments. All animals were bred and maintained under specific pathogen-free (SPF) conditions. Experimental procedures were approved by the Institutional Animal Care and Use Committee and the local animal welfare committee of the Biopole ULB Charleroi (BUC).

Humanisation protocol and reconstitution analysis

Newborn mice (within 2 days after birth) were sublethally irradiated (80 cGy) and injected intrahepatically with 10,000 human CD34⁺ cells derived from fetal liver in 20 µL of PBS using a 22G needle mounted on a Hamilton glass syringe, as previously described⁷. All neonates received broad-spectrum antibiotics (enrofloxacin, 0.27 mg/mL) in their drinking water until weaning. Peripheral blood was collected six weeks post-transplantation, and only mice

exhibiting more than 2% human (h) CD45⁺ cells (of total hCD45⁺ and mouse (m) CD45⁺ populations) were selected for subsequent experiments.

Some experiments were performed by TransCure bioServices (TCS, Archamps, France). Female NOD/SCID/IL-2R γ^{null} immunodeficient mice were humanized using CD34⁺ isolated from human cord blood following TCS's proprietary humanization protocol (3 donors, 5 mice from each). Only mice with a humanization rate (hCD45/total CD45) above 25% were used. Mice received a cytokine boost based on the transient expression of human IL-3, IL-4, IL-15, GM-CSF and FLT3L cytokines by hydrodynamic injection 1 week before tumor engraftment.

CDX implantation and monitoring

The human melanoma cell line Mel LB33 was obtained from Pr Pierre Van der Bruggen, Université Catholique de Louvain, Belgium. The SK-Mel-37 cell line was obtained from Dr Taha Merghoub and used under a Material Transfer Agreement with the Memorial Sloan Kettering Cancer Center, New York, USA. The human colorectal cell line LS174T was kindly provided by Professor Marina Kreutz, University Hospital Regensburg, Germany. The human triple negative breast cancer cell line MDA-MB-231 was purchased from Sigma.

SK-Mel-37 cells were grown in RPMI 1640 (Lonza) supplemented with fetal calf serum (FCS) 10%, 2 mM L-glutamine, 25 mM Hepes, 1 mM nonessential amino acids (Lonza), 100 U/ml penicillin, and 100 U/ml streptomycin (Pen-Strep (Lonza)). MDA-MB-231 cells were grown in Leibovitz's L-15 Medium (LB15) supplemented with FCS 10%, 2mM L-glutamine, 100 U/ml penicillin, and 100 U/ml streptomycin (Pen-Strep (Lonza)). Mel-LB-33 cells were grown in IMDM supplemented with FCS 10%, 2mM L-glutamine, 100 U/ml penicillin, and 100 U/ml streptomycin (Pen-Strep (Lonza)). LS174T cells were grown in RPMI supplemented with FCS 10% and 100 U/ml penicillin and 100 U/ml streptomycin (Pen-Strep (Lonza)).

All cell lines were grown to ~80% confluency at 37°C, 5% CO₂, except for the MDA-MB-231 cell line, which was grown at 0% CO₂. 0,5.10E+06 LS174T cells per mouse, or 4.10E+06 MelLB33 cells per mouse, or 5.10E+06 SK-Mel-37 cells per mouse were injected subcutaneously in the flank. 5.10E+06 cells MDA-MB-231 cells per mouse were injected in the mammary fat pads of females under anesthesia. All CDX were implanted in 100 μ L of Matrigel (Corning), using a BD Micro-Fine+ Demi Seringue Insuline 0,3ml 30Gx8mm. The size of the tumors was measured twice a week using a caliper, and the tumor volume was

calculated using the following formula: $\text{volume} = d^2 \times D/2$, where d is the shortest diameter and D is the longest diameter. At the endpoint (when the mean volume of the tumors reached 675 mm³), mice were sacrificed and tissues were collected for single cell suspensions preparation or placed in formol 10% for further immunostaining experiments.

HIS mice cell suspensions preparation

Spleens were harvested and immediately transferred into sterile dishes containing 5 mL of RPMI supplemented with 10% fetal calf serum (FCS) and 2 mM EDTA. A 70 μm cell strainer was positioned on a 50 mL conical tube, and the spleens were mechanically dissociated on the strainer using the plunger of a sterile 3 mL syringe. The tube was then filled with RPMI supplemented with 10% FCS and 2 mM EDTA to a final volume of approximately 20–30 mL. Samples were centrifuged at $300 \times g$ for 7 min at 4 °C. The supernatant was carefully aspirated without disturbing the pellet. Cells were resuspended in 2 mL of homemade ACK (Ammonium-Chloride-Potassium, home-made) RBC lysis buffer and incubated at room temperature for 2 min. The lysis reaction was stopped by adding 15 mL of RPMI supplemented with 10% FCS and 2 mM EDTA, followed by centrifugation at $300 \times g$ for 7 min at 4 °C. RBC lysis was repeated once, for a total of two lysis steps. The final pellet was resuspended in an appropriate volume of RPMI supplemented with 10% FCS and 2 mM EDTA.

Bone marrow cells were isolated from mouse femurs collected under aseptic conditions. Surrounding tissue was removed and the bones were thoroughly cleaned. The extremities of each bone were removed with a scalpel to expose the marrow cavity. Bone marrow was then released by flushing the cavity with sterile medium using an appropriate syringe fitted with a fine-gauge needle, such as a 5 mL syringe fitted with a 25-gauge needle. The recovered marrow was mechanically dissociated to obtain a homogeneous single-cell suspension, which was passed through a 70 μm cell strainer to remove debris and aggregates. Cells were resuspended in 1 mL of homemade RBC lysis buffer and incubated at room temperature for 2 min. The lysis reaction was stopped by adding 14 mL of RPMI supplemented with 10% FCS, followed by centrifugation at $300 \times g$ for 7 min at 4 °C. The final pellet was resuspended in an appropriate volume of RPMI supplemented with 10% FCS.

Livers were perfused with PBS through the portal vein before dissection, then cut in a Petri dish and incubated for 30 min at 37 °C with 40 $\mu\text{g/mL}$ DNase I (Roche) and 20 $\mu\text{g/mL}$ Liberase (Roche), and subsequently neutralized with RPMI supplemented with 5% FCS and 2 mM EDTA. The digested liver was mashed and filtered through a 70 μm cell strainer, and further

purified by centrifugation in a 35% Percoll gradient (2000 rpm, 30 min, room temperature, no brakes). The pellet underwent RBC lysis as described earlier and was resuspended in an appropriate volume of RPMI supplemented with 5% FCS.

After dissection, lungs were incubated for 10 min at 37°C in digestion medium (RPMI supplemented with 5% FCS, 1mg/mL Collagenase A (Roche) and 40 µg/mL DNase I (Roche)), mashed and filtered through a 70 µm cell strainer, then purified by centrifugation in a 20% Percoll gradient (2000 rpm, 10 min, room temperature, no brakes). The pellet underwent RBC lysis as described earlier and was resuspended in an appropriate volume of RPMI supplemented with 5% FCS.

After collection followed by skin and hair removal, tumors were finely minced with scissors in 5 mL RPMI, 40 µg/mL DNase I (Roche) and 20 µg/mL Liberase (Roche) per tumor, in a 50 mL conical tube. Cells were then incubated at 37°C for 40 minutes. The enzymatic digestion was stopped by adding 20 mL of RPMI supplemented with 5% FCS and 2 mM EDTA. Tumors were subsequently passed through a 100 µm cell strainer using the plunger of a sterile 3 mL syringe and centrifuged at $300 \times g$ for 7 min at 4 °C. Cell pellets were resuspended in 10 ml of RPMI and leukocytes were purified by density gradient centrifugation v/v (Lymphoprep™, Stemcell Technologies) in a 50 mL conical tube (3000 rpm, 10 min, room temperature, no brake). The interface layer was collected and transferred to fresh tubes containing RPMI supplemented with 5% FCS and 2 mM EDTA. Cells were washed and resuspended in an appropriate volume of the same medium.

Flow cytometry and cell sorting

Single-cell suspensions were washed in phosphate-buffered saline (PBS) and incubated in Purified Rat Anti-Mouse CD16/CD32 (Mouse BD Fc Block™) and Purified NA/LE Human BD Fc Block™ (both from BD Biosciences) and fluorescently conjugated antibodies for cell surface markers in PBS for 30 min at 4 °C in the dark. Dead cells were excluded with LIVE/DEAD™ Fixable Near IR (876) Viability Kit, for 808 nm excitation (ThermoFisher Scientific). Cell acquisition was performed on a CytoFLEX LX flow cytometer (Beckman Coulter). Data were analyzed using FlowJo v10 software (Tree Star).

Cell sorting

Single-cell suspensions were washed in phosphate-buffered saline (PBS) and incubated in Purified Rat Anti-Mouse CD16/CD32 (Mouse BD Fc Block™) and Purified NA/LE Human BD Fc Block™ (both from BD Biosciences) and fluorescently conjugated antibodies for cell surface markers in PBS for 30 min at 4 °C in the dark. Dead cells were excluded with Live/Dead Fixable Aqua Dead Cell Stain (Life Technologies). Mouse cells were gated out using FITC anti-mouse CD45 (BD). Depending on the experiments, either hCD45⁺ cells or hCD45⁺CD33⁺CD14⁺ (mono-mac) cells were collected in FBS pre-coated 15 ml falcon tube containing 1 ml of RPMI supplemented with 5% FCS and 2mM EDTA. APC-Cy7 anti-hCD45, APC anti-hCD33 and PB anti-hCD14 were purchased from BD Biosciences. Single cells and bulks were sorted on a BD FACSAria™ III.

CRISPR-based gene-editing of CD34⁺ HSPCs in HIS mice

3 guide RNAs (gRNA) were designed to target the only exon of the *MAFB* gene, using the CRISPOR web tool³² to predict off-target and on-target scores, and synthesized by Integrated DNA Technologies (IDT). An approach delivering 3 simultaneous guides in CD34⁺ cells was used to obtain highest efficiency.

CD34⁺ HSPCs were nucleofected based on the protocol described in³³ with some adjustments. Briefly, HSPCs were rapidly thawed 4h before nucleofection in a 37°C water bath. HSPCs in FBS containing 10% DMSO were then washed and sequentially diluted in increasing volumes of DMEM High Glucose (Capricorn Scientific, DMEM-HA) supplemented with 10% FCS to ensure optimum recovery. Cells were centrifuged at 200xg for 10 min at room temperature, and resuspended at 2x10⁶ cells/mL in warm HSPC expansion media (StemSpan SFEM II medium [Stemcell Technologies, 09655], supplemented with 0.1µg/mL of SCF [Peprotech, 300-07], FLT3-L [Peprotech, 300-19], IL-6 [Peprotech, 200-06] and TPO [Peprotech, 300-18], 1% glutamine [Cegrogen Biotech, K0100-670], 0.75µM SR-1 [Stemcell Technologies, 72342] and 35nM UM-729 [Stemcell Technologies, 72332]). Each gRNA was prepared by annealing CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) through incubation at 95°C for 5 min and gradual cooling to room temperature. To assemble the ribonucleoprotein (RNP) complex, 300pmol of each 3 gRNAs were mixed with 100pmol of Cas9 Nuclease V3 (IDT, 1081059) at a gRNA:Cas9 molar ratio of 3:1, and incubated at room temperature for 10-15 min. After blending the 3 RNP complexes, equimolar (300pmol) Electroporation Enhancer (IDT, 1075916) and complete P3 Primary Cell Nucleofector Solution (reconstituted by mixing P3 Nucleofector Solution with the Supplement 1 at a 4.5:1 ratio [Lonza, V4XP-3024]) were

added for a final volume of 100 μ L. HSPCs were centrifuged, and the cell pellet resuspended in 100 μ L of RNP complex. Cells were then transferred into a 100 μ L Nucleocuvette (Lonza, V4XP-3024) and nucleofected in a 4D-Nucleofector X Unit (Lonza, AAF-1003X) using the program EO-100. 500 μ L of warm HSPC expansion media were then added to the cuvette, and the cells transferred to a 12-well plate pre-filled with 1mL of warm HSPC expansion media. The nucleofected HSPCs cells were left at 37°C and harvested 4h later to be used for humanization as described earlier, with the injection adjusted to 20,000 edited CD34⁺ cells per newborn pup in parallel of the usual 10,000 non-edited CD34⁺ cells in control mice.

Single-cell RNA library preparation, sequencing and alignment

FACS-sorted, live human CD45⁺ cells isolated from tumor cell suspensions were filtered through a 40 μ m cell strainer and the concentration adjusted to 1000 cells/ μ L before single cell capture using the Chromium controller (10x Genomics). Single-cell RNA-seq libraries were prepared using the Chromium Single Cell 3' v3.1 Reagent Kit (10x Genomics) according to manufacturer's user guide. Libraries were loaded to Illumina Novaseq 6000 (Brightcore platform) and Cell Ranger (10x Genomics) function count was used to demultiplex the sequencing data, generate gene-barcode matrices, and align the sequences to the GRCh38 genome. Resulted matrices were used to generate a single seurat object using Seurat 5.1.0 R package.

Demultiplexing: Data containing the information about oligo-tagged antibodies and cell barcodes were added as an independent assay in the Seurat object and were normalized with the `NormalizeData()` function from the Seurat package (version 5.1.0), using the "CLR" normalization method which applies a centered log ratio transformation. The `HTODemux()` function was used to reassign individual cells to their original mouse sample. Negative cells, with no oligo-tagged antibody detected, and cells assigned with more than one oligo-tagged antibodies, were removed from downstream analysis.

Filtering: RNA data were filtered by mouse sample with the following thresholds: number of genes expressed > 400 (between 400 and 7500 depending on the mouse sample), fraction of mitochondrial content < 25%. The new data from the filtered Seurat object were normalized with `NormalizeData()` and scaled with `ScaleData()` functions from Seurat package, with the following variables to regress: `ncount_RNA`, mitochondrial and ribosomal content. To remove batch effect and compare clusters between samples, we used `IntegrateLayers()` function with `method = HarmonyIntegration` from Seurat package. Variable features were used in top 30

principal components (PCs) to with *RunUMAP()* function to identify clusters using the original Louvain algorithm with a resolution of 0.8. *FindAllMarkers()* function with default parameters was used to annotate major cell populations. All Myeloid cells were separated from CD45+ cells, re-clustered with the same parameters and projected to a new UMAP with the same number of dimensions. Myeloid clusters were annotated based on differentially expressed genes determined by *FindAllMarkers()* function.

Multomics scRNAxscATAC

Nuclei suspensions from FACS-sorted, life human CD45+ cells were prepared as described in the 10x Genomics demonstrated protocol ‘Nuclei Isolation for Single Cell Multiome ATAC + Gene expression Sequencing CG000365 Rev B’. Libraries were generated using chromium Single Cell Multiome ATAC + Gene expression kits (10X Genomics) following the manufacturer’s protocol (CG000338) and loaded to Illumina Novaseq 6000 (Brightcore platform). Libraries were aligned to hg38 reference genome and transcriptome. Reads were filtered, counted, and peaks called using the cellranger-arc count. Seurat object containing both RNA and ATAC-seq matrices was created using *CreateSeuratObject()* function.

Survival analysis

For each Mono/Macs cluster identified from scRNAseq analysis, a signature was defined as the fifty most specific genes based on *FindAllMarkers()* from Seurat package. Each signature was then submitted to KMplotter webtool (<https://kmplot.com/analysis/>) for survival analysis against TCGA cohorts³⁴. Hazard-ratio and p-value were retrieved for visualization.

Cut&Run experiments

Human CD33⁺CD14⁺ cells were isolated from MDA-MB-231 tumors implanted orthotopically into the mammary glands of MISTRG humanized mice. A total of 100,000 cells were collected for each reaction. Cell fixation, immunoprecipitation, and DNA extraction were performed following the manufacturer’s protocols provided by Cell Signaling Technology (Cat. #86652 and #14209). DNA concentration was quantified using the PicoGreen dsDNA assay (Thermo Fisher Scientific, Cat. #P7589), and DNA libraries were prepared using the Cell Signaling Technology DNA Library Preparation Kit (Cat. #56795). Prepared libraries were subsequently submitted for next-generation sequencing. Immunoprecipitations were carried out using the control antibodies supplied with the kit, as well as an anti-MAFB antibody (Sigma-Aldrich, Cat. #HPA005653).

de novo MAFB motif discovery

Motif discovery within MAFB-bound chromatin regions was conducted using the *rGADEM* package in R [Droit A, Gottardo R, Robertson G, Li L (2024). *rGADEM: de novo motif discovery*. R package version 2.54.0.]. High confidence MAFB peaks ($\log_{10}(\text{qvalue}) > 8$) were selected and converted into fixed 50 bp windows centered on peak summits using the *resize()* function from *GenomicRanges*, ensuring uniform length and consistent anchoring around the most informative region. These curated windows were then supplied to the rGADEM workflow through the *GADEM()* function (*nmotifs*=10), and MAFB motif was selected based on significance and length (<10 bp). The resulting motifs were visually compared to the JASPAR mouse *Ma1b* motif (MA0117.3) and the most similar one was selected. The resulting MAFB Position Weight Matrix (PWM) was then incorporated into the JASPAR2024 object (*collection*="CORE", *species*=9606, last version of each motif) under the identifier "MA0117.3".

CUT&RUN Analysis

Paired-end reads were processed as follows: adapter sequences and low-quality bases were trimmed using Cutadapt v3.5. Trimmed reads were aligned to CRGh38 human genome using Bowtie2 v2.5.1 with the following parameters: -I 0 -X 1000 --fr --dovetail --very-sensitive . Duplicate reads were marked using Picard MarkDuplicates v3.1.0 . MAFB peaks were called by MACS2 with the following parameters: *macs2 callpeak -f BAMPE -g hs --nomodel -SPMR*. For visualization, BAM files were converted to bigwig format using *bamCoverage* function from *deeptools* v3.5.5 package. To evaluate significant association between MAFB-bound genomic loci and differentially expressed genes in TAMs sub-populations, we used Binding and Expression Target Analysis (BETA) v1.0.7 with *beta-basic* function³⁶.

hdWGCNA

Weighted gene co-expression network analysis was carried out using the hdWGCNA framework¹⁸. Only MISTRG-derived Mono-Mac cells were analyzed, excluding proliferating subsets. Genes expressed in >5 % of cells were retained for downstream analysis. Metacells for each cluster were generated using *MetacellsByGroups()* function [parameters: *k*=25, *max_shared*=10, *min_cells*=100, *mode*='average'], grouping by cell types and samples, and integrating all Mono/Mac cell types together for the analysis.

Co-expression networks were constructed using the hdWGCNA adaptation of the blockwise WGCNA pipeline³⁵, optimized for high-dimensional single cell data. A soft thresholding power value (5, see supplemental figure 3G) was automatically estimated using *TestSoftPowers()* function to approximate scale-free topology. Adjacency matrix and TOM (topological overlap matrix) were computed, and modules were identified by hierarchical clustering followed by dynamic tree cutting, obtaining six distinct gene modules (five modules of co-expressed genes, and one module containing other unassigned genes). To determine whether the hdWGCNA modules identified from humanized mouse data were preserved in human myeloid populations, module preservation analysis was performed using the hdWGCNA preservation testing framework via the *ModulePreservation()* function [n_permutations=250], selecting the MoMac-VERSE atlas as the query dataset [GSE178209]¹³.

- Enrichment analysis on module genes

Gene sets derived from hdWGCNA modules were analyzed using *enrichR* package³⁶ to perform an over-representation analysis (ORA) against the Hallmarks database³⁷. Enriched terms were ranked by Enrichment Score and statistical significance.

- Regulatory networks construction

Transcription factors (TFs) within each module were identified based on the catalogue of the 1639 known and likely human TFs and their motifs compiled by Lambert et al³⁸.

Each TF was assigned its most recent human motif from the JASPAR2024 database³⁹. Because JASPAR lacks a validated human MAFB motif, the custom PWM derived from our CUT&RUN motif analysis was used (see “de novo MAFB motif discovery” section). TF binding sites were predicted with motifmatchr’s *MotifScan()* function [Schep A (2025). **motifmatchr: Fast Motif Matching in R**.doi:10.18129/B9.bioc.motifmatchr, R package version 1.32.0, <https://bioconductor.org/packages/motifmatchr>.], scanning promoter regions. Directed TF-target networks were generated with *ConstructTFNetwork()* function from *hdWGCNA*, and using strategy “A”, which selects the top TFs regulating each gene (n_tfs=50).

- Differential regulon analysis

Differences in TF regulon activity between Tumor-associated VCAN and healthy VCAN monocytes were assessed using *FindDifferentialRegulons()* function.

- Module Regulatory Network

Net module-level regulatory interactions were calculated and summarized using *ModuleRegulatoryNetworkPlot()* function.

SCENIC+

SCENIC+ ²¹ was used to integrate matched scRNA-seq and scATAC-seq profiles to infer enhancer-driven regulatory networks and identify context-specific TF activity across bone marrow and tumor conditions.

Data pre-processing included removing proliferating cells and non-Mono/Macro clusters, and processing multiomic data using *pycisTopic* and *Scanpy* according to the documentation. A custom cisTarget database was built on consensus peaks, as suggested by the authors. Finally, SCENIC+ workflow was run using *SnakeMake*. Clusters containing <100 cells were excluded, while only TFs present both as inhibitors and activators, and included in the hdWGCNA modules, were selected among the results.

ChromVAR

We used ChromVAR package to infer TF activity across clusters and tissues within our single cell multiomic data, ⁴⁰. Since ChromVAR allows the creation and utilization of a user-defined motif matrix, we decided to use the same set of motifs defined for the regulatory network inference by hdWGCNA. Motif activity values were obtained using the *RunChromVAR()* function and selecting UCSC hg38 as the reference genome assembly. Finally, the average Z scores of motif SD (standard deviation) was calculated for each cell.

Trajectory analysis

Differentiation trajectories from monocytes to macrophages in tumor and healthy tissues were reconstructed with slingshot⁴¹. Trajectories were computed using harmonized PCA coordinates to preserve the distances between points, and a minimum spanning tree connecting all cells was determined. The fitted curves in PCA coordinates were projected onto UMAP embeddings, all the identified trajectories were visualized in the same UMAP space.

Each trajectory was analyzed independently to better characterize lineage-specific dynamics. for each trajectory, the most dynamic genes were identified using TSCAN's *testPseudotime()* function, focusing on transcription factors within hdWGCNA modules.

MAFB KO efficiency evaluation

To assess the efficiency of CRISPR-mediated MAFB knockout, *pysam* package [<https://github.com/pysam-developers/pysam>]⁴²⁻⁴⁴ was used to parse BAM files, read CIGAR strings, and detect whether deletions theoretically induced by CRISPR-Cas9 were present within the genomic interval defined by the two outermost guide RNAs. Deletions covering this region (or a fraction of it) were classified as CRISPR-edited events. Knockout efficiency was calculated as the proportion of edited reads – those carrying a deletion between the first and the third guide – over the total number of reads mapping across the same interval. To avoid inflation of editing estimates, we applied a basic UMI-based deduplication by removing reads sharing both the same UMI and barcode (BC), thereby excluding PCR duplicates from both the visualization and the efficiency quantification. Based on this computational workflow, we developed a dedicated Python package, CIGAR-bIA, which automates CIGAR parsing, detection of CRISPR-induced indels, UMI deduplication, and visualization of editing events. The source code is openly available at: <https://github.com/francescopatane96/CIGAR-bIA>.

Differential analyses

- Differential expression

Global transcriptional differences between Mono/Macro located within tumoral and healthy tissue ; or between MAFB KO and control cells were assessed using Seurat's *FindMarkers()* function (test.use=wilcoxon). The analysis was then repeated within each cluster to confirm global trends and detect cluster-specific deviations.

- Mono/Macs subtype-specific genes identification

- Differential cell type abundance

Differences in cell type proportions between conditions were evaluated with *scProportionTest*⁴⁵

- Cell type prioritization following perturbation

Perturbation-responsive cell types were identified using *Augur*⁴⁶

- Metascores

For differential module activity, AUCell⁴⁷ was applied to compute hdWGCNA module activity score per cell. Median module activities were compared between conditions using Wilcoxon tests.

For pathways analysis, Seurat *AddModuleScore()* function was used to calculate pathway scores. Those scores were then Z-scaled.

Patient samples

The present human studies were reviewed and approved by the Medical Ethics Committee of the Institut Jules Bordet (CE1981). BC patients gave informed consent prior to their participation in the study. Other residual tissues were obtained anonymously, without access to clinical data. The clinicopathological characteristics of patients are summarized in Table S1.

Table S1: Patient clinicopathological characteristics

	Normal breast	DCIS	IDC
Patients	4	9	8
Molecular subtype a			
Luminal	0	3	4
HER2+ luminal	0	2	1
HER2+	0	3	2
TN	0	1	1
Age			
<50 yr		3	1
> 50 yr		6	7
Menopausal status			
yes		6	7
no		3	1
Histotype			
Ductal		9	8
Grade			
I		0	1
II-III		9	7
Lympho vascular invasion b			
Yes		0	1
No		9	7
Ki67 (Tumor) c			
<15%		/	2
>15%		/	6

Stage			
I-II		9	0
III		0	8
IV		0	0
DCIS presence			
100%		9	0
<10%		0	1
0		0	7

Tissue Collection and Preparation for multiplex immunohistochemistry (mIHC)

Primary tumor tissues were collected during surgery and preserved as formalin-fixed, paraffin-embedded (FFPE) blocks. Tissue sections of 4 μm thickness were cut from FFPE blocks pre-selected by a pathologist based on the ratio of tumor and adjacent tissue areas, determined by examination of hematoxylin and eosin (H&E)-stained sections. Freshly cut slides were air-dried at 37 °C and subsequently incubated for 4 h at 65 °C to minimize tissue detachment during staining.

Multiplex Immunohistochemistry (mIHC)

Multiplex immunohistochemistry was performed following an optimized in-house protocol. Antibodies used are referenced in list xx. Slides were scanned using the PhenoImager HT system (Akoya Biosciences). Initially, whole-slide images were acquired using all five fluorescence filters (DAPI, FITC, Cy3, Texas Red, and Cy5) at 20 $\times$ magnification. Subsequently, regions of interest (ROIs) were selected for multispectral imaging at 40 $\times$ magnification.

To minimize tissue autofluorescence and background signal, multispectral images were spectrally unmixed using *inForm* software (version 2.6.0, Akoya Biosciences). A breast cancer-specific spectral library was generated from the pure emission spectra of each fluorophore and DAPI. Autofluorescence was characterized using an unstained breast cancer tissue section.

Image Analysis and Cell Phenotyping

Quantitative image analysis was performed in *QuPath* (version 5.1)⁴⁸. Tissue and cellular segmentation, as well as phenotyping, were conducted using *QuPath* integrated tools. Tumor regions were annotated based on Pan-Cytokeratin (PanCK) positivity, whereas PanCK-negative regions were designated as adjacent stroma. Necrotic areas were manually annotated and excluded from subsequent analyses. Cell segmentation was carried out using *StarDist*⁴⁹.

Macrophages were identified as CD68⁺ cells. Macrophage subsets were defined as CD68⁺ cells co-expressing one or more subset-specific markers. Cells expressing two subset markers were included in both corresponding categories.

Spatial Analysis

The distance between each cell and the nearest necrotic area was computed in *QuPath*. For each macrophage subset, the median distance to the nearest necrotic region was calculated to assess spatial distribution across the section.

Data availability.

RNA-Seq and ATAC-Seq data that support the findings reported in this study have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus database with the accession code (Pending).

Antibodies used for multiplex IHC

ANTIBODY	Reference	Dilution
PANCYTOKERATIN	Dako-Agilent - M3515	1/400
CD68	Cell signaling – 76437S	1/500
FOLR2	Abcam – ab302532	1/100
IL4I1	Abcam – ab222102	1/200
NRLP3	Sigma – ABF23	1/2000
SPP1	Sigma – HPA027541	1/200
MAFB	Abcam – ab223744	1/500
MITF	Abcam – ab303530	1/500

HO-1

Abcam - ab52947

1/100

References

1. DeNardo, D. G. & Ruffell, B. Macrophages as regulators of tumour immunity and immunotherapy. *Nat Rev Immunol* <https://doi.org/10.1038/s41577-019-0127-6> (2019) doi:10.1038/s41577-019-0127-6.
2. Pittet, M. J., Michielin, O. & Migliorini, D. Clinical relevance of tumour-associated macrophages. *Nat Rev Clin Oncol* 19, 402–421 (2022).
3. Rannikko, J. H. & Hollmén, M. Clinical landscape of macrophage-reprogramming cancer immunotherapies. *Br J Cancer* <https://doi.org/10.1038/s41416-024-02715-6> (2024) doi:10.1038/s41416-024-02715-6.
4. Monnier, M. *et al.* Antitumor strategies targeting macrophages: the importance of considering the differences in differentiation/polarization processes between human and mouse macrophages. *J Immunother Cancer* 10, e005560 (2022).
5. Azizi, E. *et al.* Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* 174, 1293–1308.e36 (2018).
6. Zhang, Y. *et al.* Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell* <https://doi.org/10.1016/j.ccell.2021.09.010> (2021) doi:10.1016/j.ccell.2021.09.010.
7. Rongvaux, A. *et al.* Development and function of human innate immune cells in a humanized mouse model. *Nat Biotechnol* 32, 364–372 (2014).
8. Alaluf, E. *et al.* Heme oxygenase-1 orchestrates the immunosuppressive program of tumor-associated macrophages. *JCI Insight* 5, (2020).
9. Consonni, F. M. *et al.* Heme catabolism by tumor-associated macrophages controls metastasis formation. *Nat Immunol* 22, 595–606 (2021).
10. Luu Hoang, K. N., Anstee, J. E. & Arnold, J. N. The Diverse Roles of Heme Oxygenase-1 in Tumor Progression. *Front Immunol* 12, 658315 (2021).
11. Hegde, S. *et al.* Myeloid progenitor dysregulation fuels immunosuppressive macrophages in tumours. *Nature* 646, 1214–1222 (2025).
12. Voillet, V. *et al.* An In Vivo Model of Human Macrophages in Metastatic Melanoma. *J Immunol* 209, 606–620 (2022).
13. Mulder, K. *et al.* Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* 54, 1883–1900.e5 (2021).
14. Bill, R. *et al.* CXCL9:SPPI macrophage polarity identifies a network of cellular programs that control human cancers. *Science (1979)* 381, 515–524 (2023).
15. Matusiak, M. *et al.* A spatial map of human macrophage niches reveals context-dependent macrophage functions in colon and breast cancer. *Res Sq* <https://doi.org/10.21203/rs.3.rs-2393443/v1> (2023) doi:10.21203/rs.3.rs-2393443/v1.
16. Nasir, I. *et al.* Tumor macrophage functional heterogeneity can inform the development of novel cancer therapies. *Trends Immunol* 44, 971–985 (2023).
17. Nalio Ramos, R. *et al.* Tissue-resident FOLR2+ macrophages associate with CD8+ T cell infiltration in human breast cancer. *Cell* 185, 1189–1207.e25 (2022).
18. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Reports Methods* 3, 100498 (2023).
19. Schaer, D. J. *et al.* Hemorrhage-activated NRF2 in tumor-associated macrophages drives cancer growth, invasion, and immunotherapy resistance. *Journal of Clinical Investigation* 134, (2023).
20. Kloosterman, D. J. & Akkari, L. Macrophages at the interface of the co-evolving cancer ecosystem. *Cell* 186, 1627–1651 (2023).

21. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* 20, 1355–1367 (2023).
22. Saiga, H., Ueno, M., Tanaka, T., Kaisho, T. & Hoshino, K. Transcription factor MafB-mediated inhibition of type I interferons in plasmacytoid dendritic cells. *Int Immunol* 34, 159–172 (2022).
23. Cheng, S. *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* 184, 792–809.e23 (2021).
24. Park, M. D. *et al.* TREM2 macrophages drive NK cell paucity and dysfunction in lung cancer. *Nat Immunol* 24, 792–801 (2023).
25. Hamon, P. *et al.* TREM2 macrophages are associated with enhanced response to PD-1 blockade in human hepatocellular carcinoma. <https://doi.org/10.1101/2025.09.09.675071> (2025) doi:10.1101/2025.09.09.675071.
26. Li, Z. *et al.* Presence of onco-fetal neighborhoods in hepatocellular carcinoma is associated with relapse and response to immunotherapy. *Nat Cancer* 5, 167–186 (2024).
27. Ventura, A. B. *et al.* MAFB: a key regulator of myeloid commitment involved in hematological diseases. *Cell Death Discov* 11, 276 (2025).
28. Sheban, F. *et al.* ZEB2 is a master switch controlling the tumor-associated macrophage program. *Cancer Cell* 43, 1227–1241.e11 (2025).
29. Hamada, M., Tsunakawa, Y., Jeon, H., Yadav, M. K. & Takahashi, S. Role of MafB in macrophages. *Exp Anim* 69, 1–10 (2020).
30. Simón-Fuentes, M. *et al.* MAFB shapes human monocyte-derived macrophage response to SARS-CoV-2 and controls severe COVID-19 biomarker expression. *JCI Insight* 8, (2023).
31. McIntosh, B. E. *et al.* Nonirradiated NOD.B6.SCID Il2ry^{-/-} KitW41/W41 (NBSGW) Mice Support Multilineage Engraftment of Human Hematopoietic Cells. *Stem Cell Reports* 4, 171–180 (2015).
32. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* 46, W242–W245 (2018).
33. Kupperts, D. A. *et al.* Gene knock-outs in human CD34⁺ hematopoietic stem and progenitor cells and in the human immune system of mice. *PLoS One* 18, e0287052 (2023).
34. Posta, M. & Györfy, B. Pathway-level mutational signatures predict breast cancer outcomes and reveal therapeutic targets. *Br J Pharmacol* 182, 5734–5747 (2025).
35. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008).
36. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44, W90–W97 (2016).
37. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* 1, 417–425 (2015).
38. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* 172, 650–665 (2018).
39. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 52, D174–D182 (2024).
40. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* 14, 975–978 (2017).
41. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19, 477 (2018).
42. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 10, (2021).

43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–9 (2009).
44. Bonfield, J. K. *et al.* HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience* 10, (2021).
45. Miller, S. A. *et al.* LSD1 and Aberrant DNA Methylation Mediate Persistence of Enteroendocrine Progenitors That Support BRAF-Mutant Colorectal Cancer. *Cancer Res* 81, 3791–3805 (2021).
46. Skinnider, M. A. *et al.* Cell type prioritization in single-cell data. *Nat Biotechnol* 39, 30–34 (2021).
47. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 14, 1083–1086 (2017).
48. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* 7, 16878 (2017).
49. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-Convex Polygons. in 265–273 (2018). doi:10.1007/978-3-030-00934-2_30.

Author' disclosures: The authors declare no potential conflicts of interest.

Funding information: This work was supported by the National Fund for Scientific Research (FNRS), the European Regional Development Fund (ERDF), the Walloon Region (ERDF 2022-2027 WAL IMAGIN SYST-IMM, WIN4EXCELLENCE GT4HEALTH), the Fondation contre le Cancer, Fonds Julie et Françoise Drion, Fonds Gaston Ithier and by grants from the Fonds Jean Brachet and the Fondation Hoguet. EDD, AP, FvH have been supported by a Télévie grant from the FRS-FNRS, SK is a research fellow and S.G. is a research director of the FRS-FNRS. MM is supported by a FRIA scholarship. .

Acknowledgments

The authors thank Regeneron Pharmaceuticals and the Richard A. Flavell lab at Yale University for generating and sharing the MISTRG mice.

Captions

Figure 1. Reconstitution of the tumor immune microenvironment in humanized mice. (A) Frequency of CD14⁺ cells among hCD45⁺ tumor-infiltrating cells; mean $\pm$ SEM are shown, and each dot represents an individual tumor. (B) Immunohistochemical visualization of selected markers across different cell line–derived xenograft (CDX) models. (C) Area percentage occupied by CD68⁺ and CD68⁺ HO-1⁺ cells in tumor and adjacent tissue. (D) Density of CD68⁺ and CD68⁺ HO-1⁺ cells in proximal (<500 μ m) or distal (>500 μ m from the necrotic core) regions of MDA-MB-231 tumors; mean $\pm$ SEM are shown, and each dot represents an individual slide from independent tumors (**p < 0.01, ****p < 0.0001). (E) scRNA-seq data on human immune cells from HIS mice, showing the proportion of cells from normal or tumor tissues (left), lymphoid, myeloid, and other cells (middle), and from NSG-B and MISTRG models integrated in the HIS atlas. (F) Number and identity of human immune cells infiltrating healthy and tumor tissues. (G) Uniform Manifold Approximation and Projection (UMAP) of all human CD45⁺ cells (n = 164,776) isolated from bone marrow, liver, lung, and spleen of naïve humanized MISTRG mice; bone marrow and spleen of naïve humanized NSG-B mice; and tumor microenvironments of SK-Mel-37, Mel-LB33, MDA-MB-231, and LS174T CDX-engrafted MISTRG mice, as well as MDA-MB-231 CDX-engrafted NSG-B mice.

Figure 2. Mono/Mac subsets from immuno-CDX models capture the full spectrum of human TAM states. (A) UMAP plot showing Mono/Mac clusters. (B) Heatmap displaying Z-score–normalized expression levels of selected marker genes across clusters. (C) Heatmap illustrating Z-score–normalized AUC scores of selected gene signatures across clusters. (D) Heatmap showing the proportion of cells in each cluster overlapping with annotated clusters in the MoMac-Verse reference atlas. (E) UMAP plots showing Mono/Mac clusters split by tissue. (F) Number of cells per cluster, colored by tissue. (G) Bubble heatmap depicting relative hazard ratio values for gene signatures derived from each cluster across different TCGA tumor types; dot size indicates significance and color represents normalized hazard ratio. (H) Stacked bar plot showing the proportion of cells per cluster across tumor CDX, split by mouse strain.

Figure 3. Evaluation of TAMs in breast cancer (BC) patients and MDA-MB-231 tumors in MISTRG humanized mice by multiplex immunohistochemistry (mIHC). (A) Heatmap showing Z-score-normalized gene expression values across HIS MoMac clusters. (B) Representative images illustrating macrophage spatial distribution in invasive BC and the proportion of CD68⁺ TAMs expressing one or more indicated markers assessed by mIHC. (C) Frequency of CD68⁺ macrophages in healthy breast tissue and in BC patients at pre-invasive (DCIS) and invasive (IDC) stages; each dot represents a single patient (*p < 0.05, **p < 0.01, Mann-Whitney test). (D) Frequency of macrophage subsets among CD68⁺ cells in healthy tissue and tumors. (E) Quantification of macrophage subset distribution between tumor nest and adjacent stroma (Wilcoxon test). (F) Comparison of macrophage subset frequencies between pre-invasive (DCIS) and invasive (IDC) stages (two-way ANOVA). (G) Representative images illustrating macrophage spatial distribution in MDA-MB-231 tumors in MISTRG mice and the proportion of CD68⁺ TAMs expressing one or more indicated markers assessed by mIHC. (H) Quantification of macrophage subsets infiltrating MDA-MB-231 tumors in MISTRG mice. (I) Median distance of each macrophage subset to the nearest necrotic region in MDA-MB-231 tumors.

Figure 4. Network analysis of human TAMs. (A) UMAP plots showing expression levels (eigenvalues) of the modules identified by hdWGCNA. (B) Bubble heatmap illustrating the z-score-normalized eigenvalues of the modules across Mono/Macs clusters (*color scale*). Dot size indicates the fraction of expressing cells. (C) heatmap highlighting the differential module expression between tumor and normal samples. Color scale indicates log2-fold-change and stars significance (*** p<0.001, ** p<0.01, *p<0.05). (D) Numbers of primary targets per transcription factor within module M-1. The color indicates the module in which the target gene is present. The direction of the y-axis indicates the regulatory sign (+y for the positive targets, -y for the negative ones). (E) Scatter plot showing the differential activity of the inferred regulons within VCAN⁺ mono between Tumor and Normal tissue. The color of the symbols indicates the module of origin for the TF. Significantly altered regulons are marked by a black contour while diamond shape indicates significantly differential TF expression. Regulons located at the bottom-right part of the plot are positively regulated in the tumoral compartment. (F) Topological overlap matrix (TOM) UMAP Network plot depicting the net inter-module regulatory activity. The color of the edges indicates positive regulation (red) and inhibitory

regulation (blue). (G) UMAP plot showing the inferred differentiation trajectories by slingshot (black lines). (H) Heatmap showing the module eigenvalues along each different trajectory. Blue-to-yellow color scale indicates pseudotime (*blue* early, *yellow* late) and black-to-magenta scale indicates module expression (*black* low, *magenta* high).

Figure 5. TF activity in TAMs based on chromatin accessibility (scATAC-Seq) (A) Heatmap plot depicting ChromVAR motif activity in individual cells (z score value). (B) Dot Heatmap showing SCENIC+ results across clusters & tissues (BM/TAMs). The color of the cell indicates AUCCell-derived regulon expression level, while dot size represents AUC chromatin accessibility score of the transcription factor regulon. Vertical annotation bars indicate the tissue (blue: BM, red: TAMs) while horizontal bars indicate modules.

Figure 6. MAFB is a central transcriptional hub of human TAMs. (A) Frequency of MAFB⁺ CD68⁺ cells among all CD68⁺ cells detected in healthy breast tissue or in BC patients; each dot represents a single patient (empty squares: DCIS, filled squares: IDC; **p < 0.01, Mann–Whitney test). (B) Frequency of MAFB⁺ cells within each macrophage subset in BC patients (top) and in MDA-MB-231 tumors from MISTRG mice (bottom); mean ± SEM are shown, and each dot represents an individual patient or tumor. (C) TOM UMAP network plot representing the inferred MAFB regulatory network using hdWGCNA, with edge colors indicating regulation sign (red: activation, blue: inhibition); MAFB is labeled in red, primary targets in yellow, and secondary targets in black; only transcription factors are labeled. (D) CUT&RUN against MAFB performed on CD14⁺CD33⁺ TAMs isolated from MDA-MB-231 tumors in MISTRG mice, showing the proportion of peaks across genomic region categories and motif enrichment analysis within these regions using HOMER. (E) Bulk ATAC-seq samples from BM monocytes and TAMs from MISTRG mice, identifying differentially accessible regions (DARs) specific to each condition and their overlap with MAFB peaks. (F) Accessibility within MAFB-bound regions in each macrophage cluster assessed by scATAC-seq. (G) CUT&RUN against MAFB performed on human monocyte-derived macrophages (MDMs), showing overlap with MAFB-bound regions in TAMs and Ciiider motif analysis comparing regions specific to TAMs versus those common to TAMs and MDMs. (H) Bar plot showing normalized enrichment scores of MAFB CUT&RUN peaks across hdWGCNA gene modules; significance calculated using a hypergeometric test.

Figure 7. In vivo invalidation of MAFB leads to a shift in TAM subpopulations. (A) Female MISTRG mice were reconstituted at birth with unedited or MAFB-KO fetal CD34⁺ HSCs. After 10–12 weeks, MDA-MB-231 CDX were implanted orthotopically. After 20 days, tumors were collected and processed for scRNA-seq on sorted CD14⁺CD33⁺ myeloid cells and for mIHC. Single-cell read coverage plots highlight CRISPR-deleted regions (red) to illustrate *MAFB* knockout efficiency in TAMs; black vertical dashed lines indicate the start and end positions of the designed RNA guides. (B) Frequency of MAFB⁺CD68⁺ cells detected by mIHC in MDA-MB-231 tumors from control (unedited) or MAFB-KO MISTRG mice (*p < 0.05, Mann–Whitney test). (C) UMAP plots of CTRL and MAFB-KO scRNA-seq samples. (D) Dot plot showing differential cell type abundance between MAFB-KO and CTRL samples; dots on the right indicate overrepresented cell types in the control, with significant results colored in orange. (E) Volcano plot displaying differentially expressed genes between total myeloid cells from control and MAFB-KO samples; red dots on the right are significantly upregulated in the KO, while dots on the left are significantly upregulated in the control. (F) Bubble heatmap depicting relative hazard ratio values for gene signatures derived from top 20 DEG (*MAFBKO* vs unedited) across different TCGA tumor types; dot size indicates significance and color represents normalized hazard ratio. (G) Violin plot comparing hdWGCNA module AUC scores of control (pink) and MAFB-KO (cyan) cells. (H) TOM UMAP network plot representing the MAFB regulatory network inferred by the BETA algorithm based on scRNA-seq and CUT&RUN data; edge color indicates regulatory sign (red: activation, blue: inhibition), while color intensity represents effect size (BETA score); only the top twenty targets per module are labeled.

Figure S1, related to Figure 1. (A) Schematic representation of the humanization procedure and implantation of human tumor cell lines in MISTRG mice. (B) Level of human immune compartment reconstitution in peripheral blood, expressed as the percentage of hCD45/(mCD45 + hCD45), measured 6 weeks post-engraftment for MISTRG mice and 8 weeks for NBSGW mice; each dot represents an individual mouse. (C) Reconstitution levels of the main human immune lineages in the blood of MISTRG and NBSGW strains, expressed as the frequency within hCD45⁺ cells. (D) Representative FACS plot showing blood monocytes in humanized MISTRG mice. (E) Schematic representation of the humanization procedure for NSG-B mice.

Figure S2, related to Figure 3. (A) Expression of HO-1 by CD68⁺ macrophages in invasive ductal carcinoma (IDC) samples, comparing the tumor nest and adjacent stroma. (B) Frequency of HO-1⁺ cells among the indicated macrophage subpopulations in tissue from breast cancer patients and (C) from MISTRG mice. Each dot represents an individual patient or tumor. Data are presented as mean $\pm$ SEM.

Figure S3, related to Figure 4. (A) Dot plot showing module preservation of co-expressed networks within the MoMac-VERSE single-cell atlas. Background shading represents cutoff values for interpreting Z summary statistics, with the darkest shade ($Z < 2$) indicating non-preserved modules in the query dataset; modules with $Z > 10$ are highly preserved. (B) Dot plot representing the top five Hallmark pathways associated with each hdWGCNA gene module. (C) Table listing transcription factors within hdWGCNA modules. (D) Heatmap showing expression of the most central transcription factors (i.e., TFs with the highest number of edges in the regulatory network) differentially expressed between tumor and normal tissue per module; only JASPAR2024-derived TFs are shown. (E) Bar plots illustrating the number and identity (target module) of transcriptionally regulated target genes across transcription factors within hdWGCNA modules. (F) Sequence logo representing MAFB transcription factor binding specificity, derived from CUT&RUN profiling of TAMs and de novo motif discovery using the rGADEM package; column height reflects information content, and nucleotide size indicates frequency at each position. (G) Dot plot depicting how closely the hdWGCNA co-expression network resembles a scale-free graph at different soft power thresholds, with the black dot representing the lowest threshold achieving a Scale-Free Topology Model Fit ≥ 0.8 .

Figure S4, related to Figure 5. (A) Feature plots showing ChromVAR activity for the indicated motif in bone marrow (BM) and tumor samples from MISTRG mice. (B–C) Volcano plots comparing motif activities between the VCAN⁺ Mono cluster from tumor versus BM (B) and between the GLUT1⁺ cluster and the VCAN⁺ cluster from tumor samples (C).

Figure S5, related to Figure 6. cMAF macrophages in healthy breast and breast cancer. (A) Frequency of cMAF⁺ CD68⁺ cells among all CD68⁺ cells detected in healthy breast tissue or breast cancer samples, assessed by mIHC (Mann–Whitney test). (B) Frequency of cMAF⁺ CD68⁺ cells within each macrophage subset in patients with breast cancer.

Figure S6, related to Figure 7. (A) Heatmap illustrating top DEGs (based on average log₂ fold change) between *MAFB-KO* and Mock samples. Genes with positive values are overexpressed in the KO. *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.05. (B) Lollipop plot depicting differentially expressed transcription factors in *MAFB-KO* vs Mock. Genes with positive values are overexpressed in the KO. (C) Radar charts showing overrepresented hallmarks of cancer in downregulated (left) and upregulated (right) genes in the *MAFB-KO* sample.

Figure 1

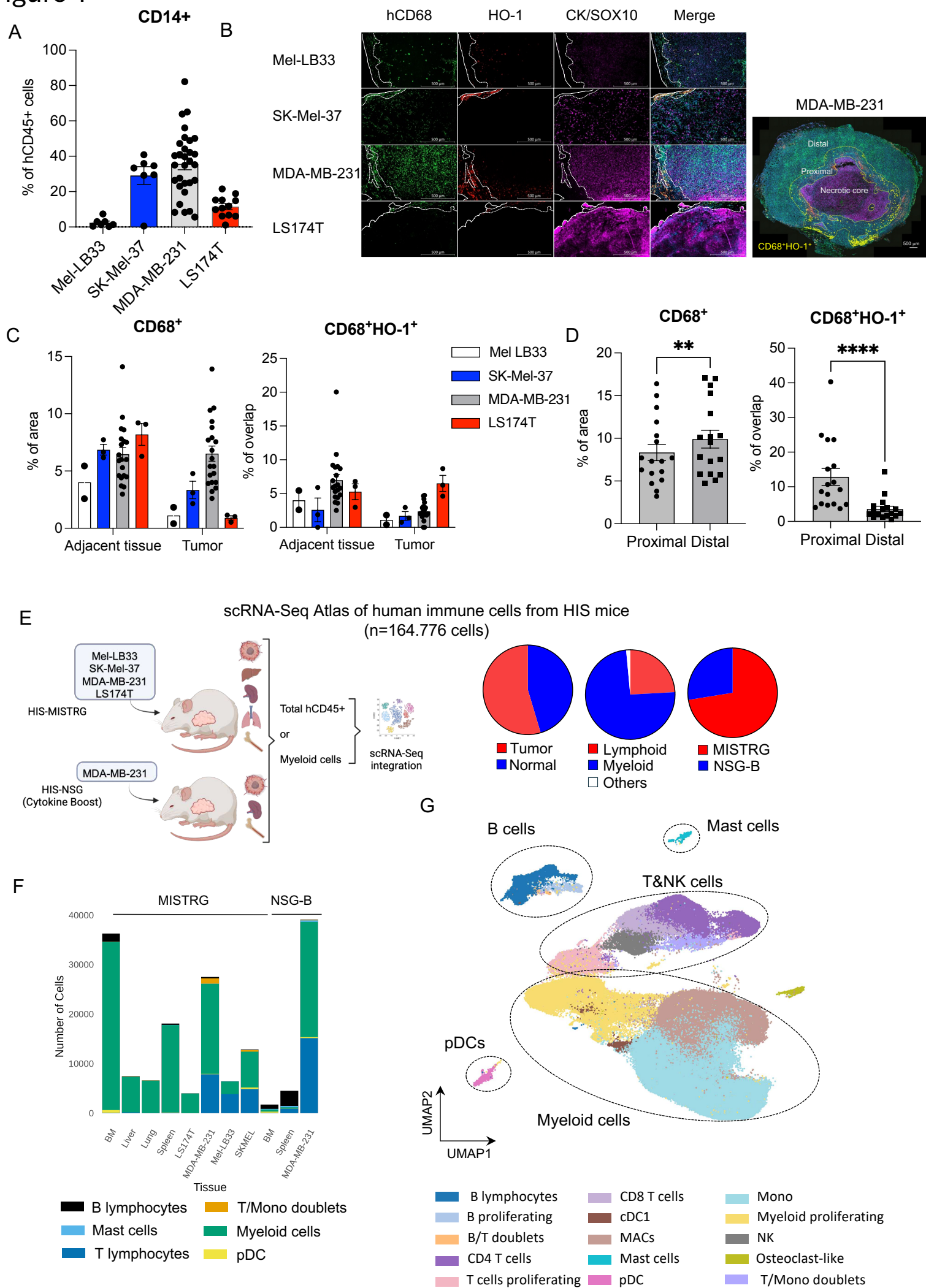
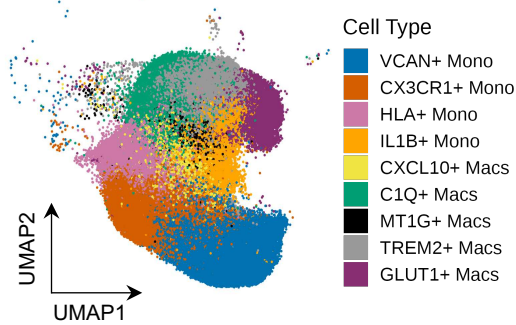
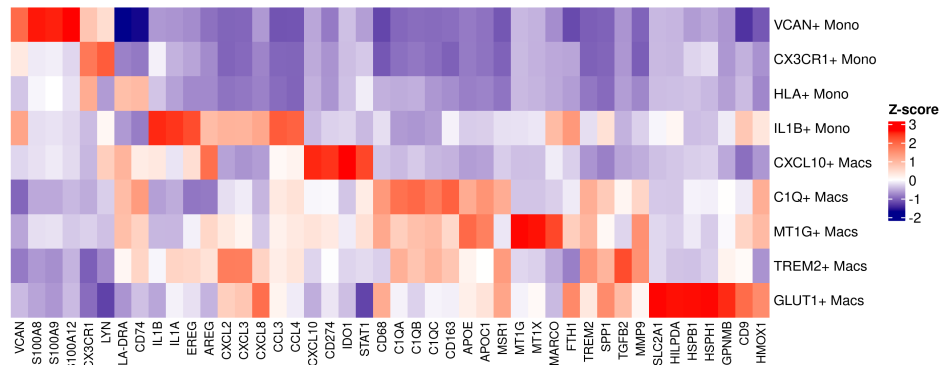


Figure 2

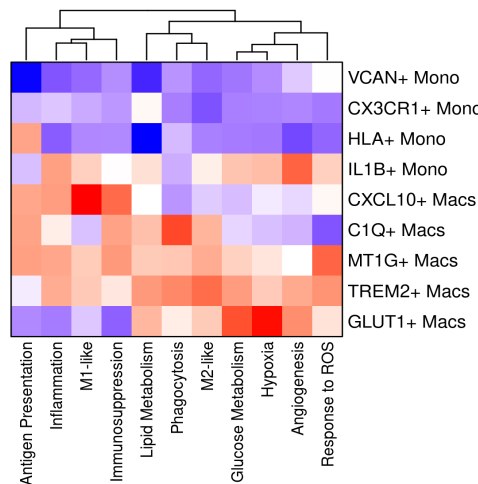
A MoMac Atlas from HIS mice
(n=97,445 cells)



B

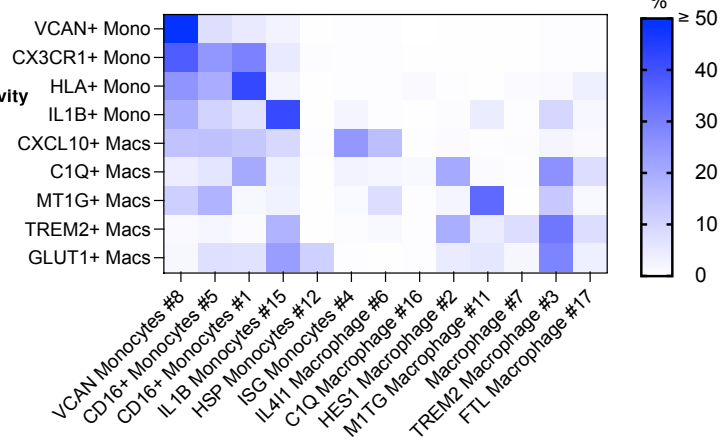


C

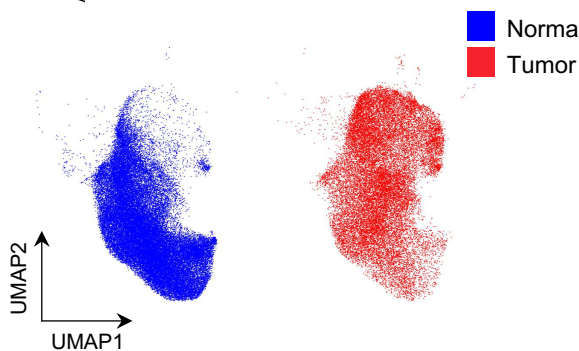


D

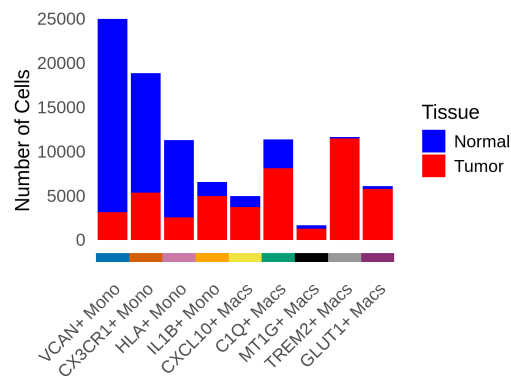
Mapping on the MoMac-VERSE (Mulder *et al* 2021)



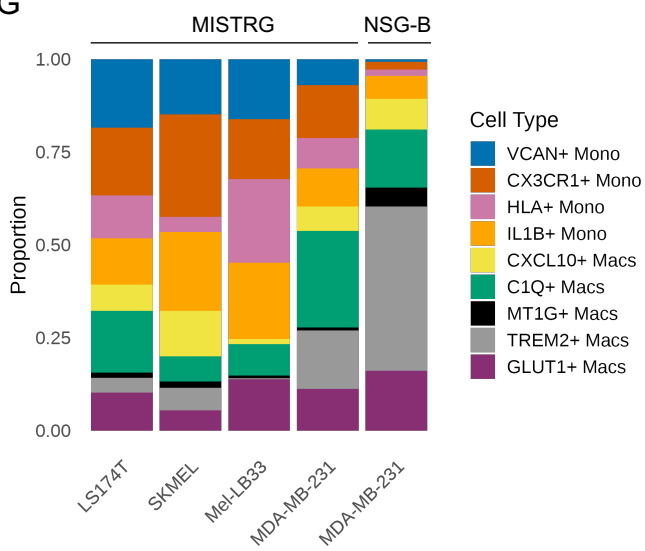
E



F



G



H

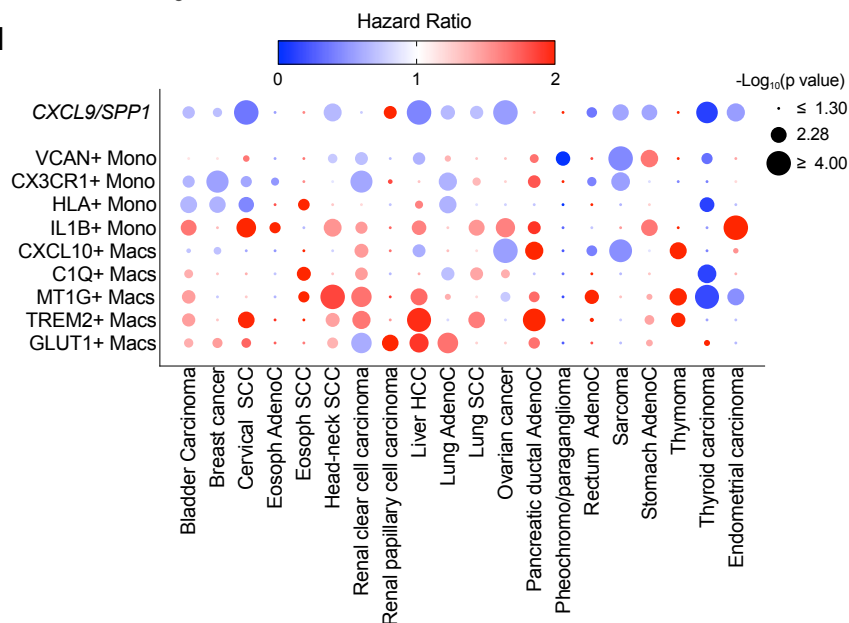


Figure 3

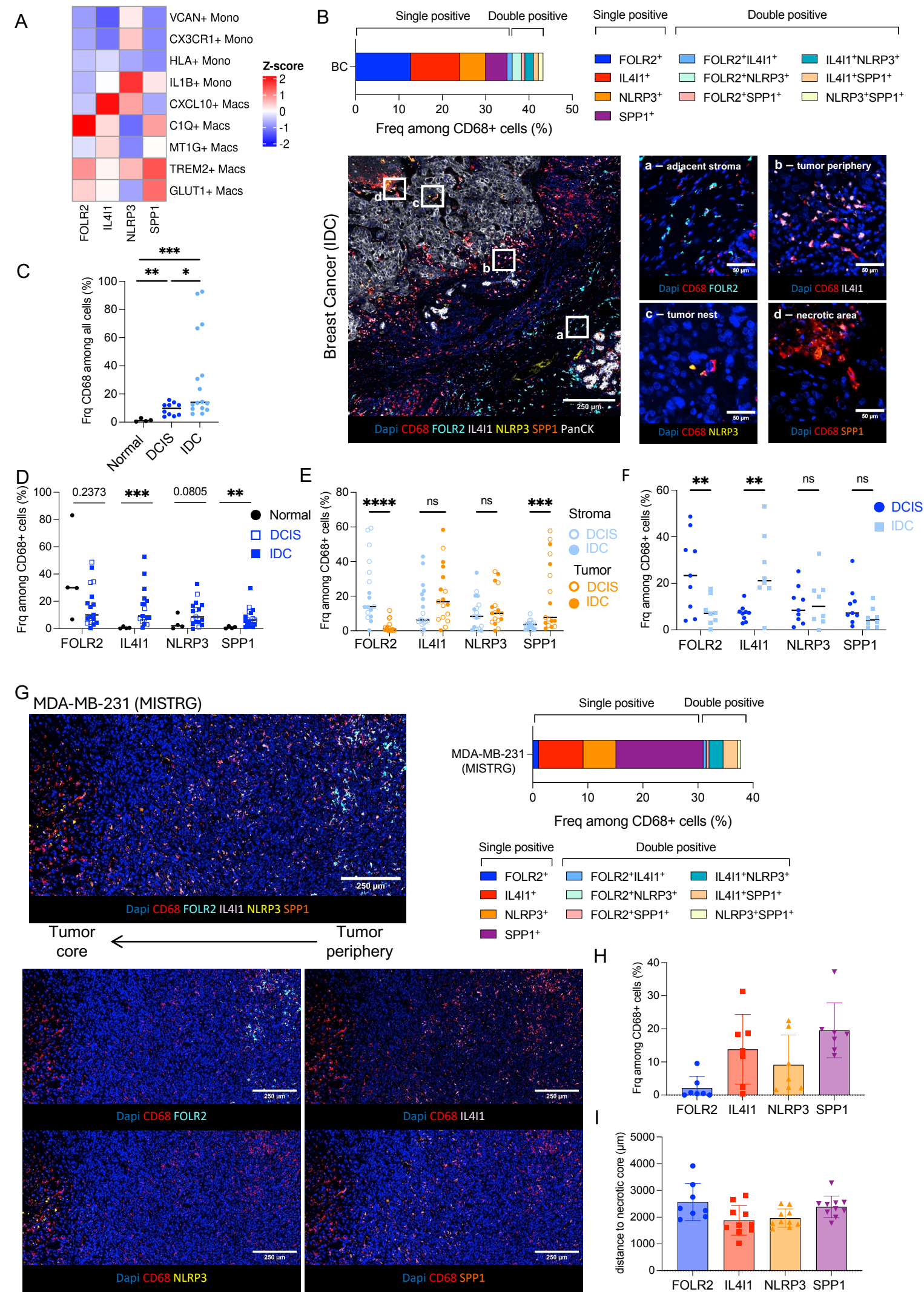


Figure 4

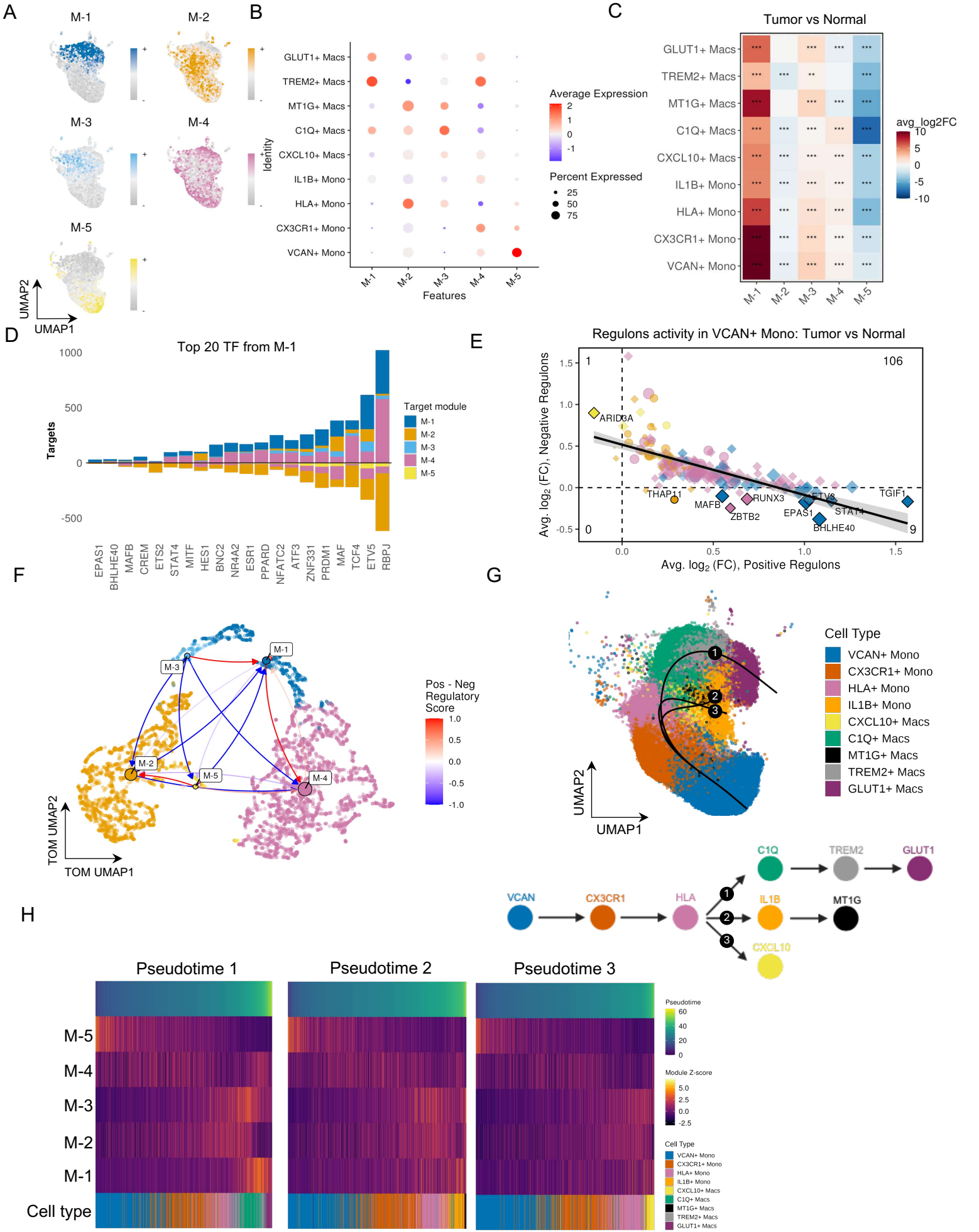


Figure 5

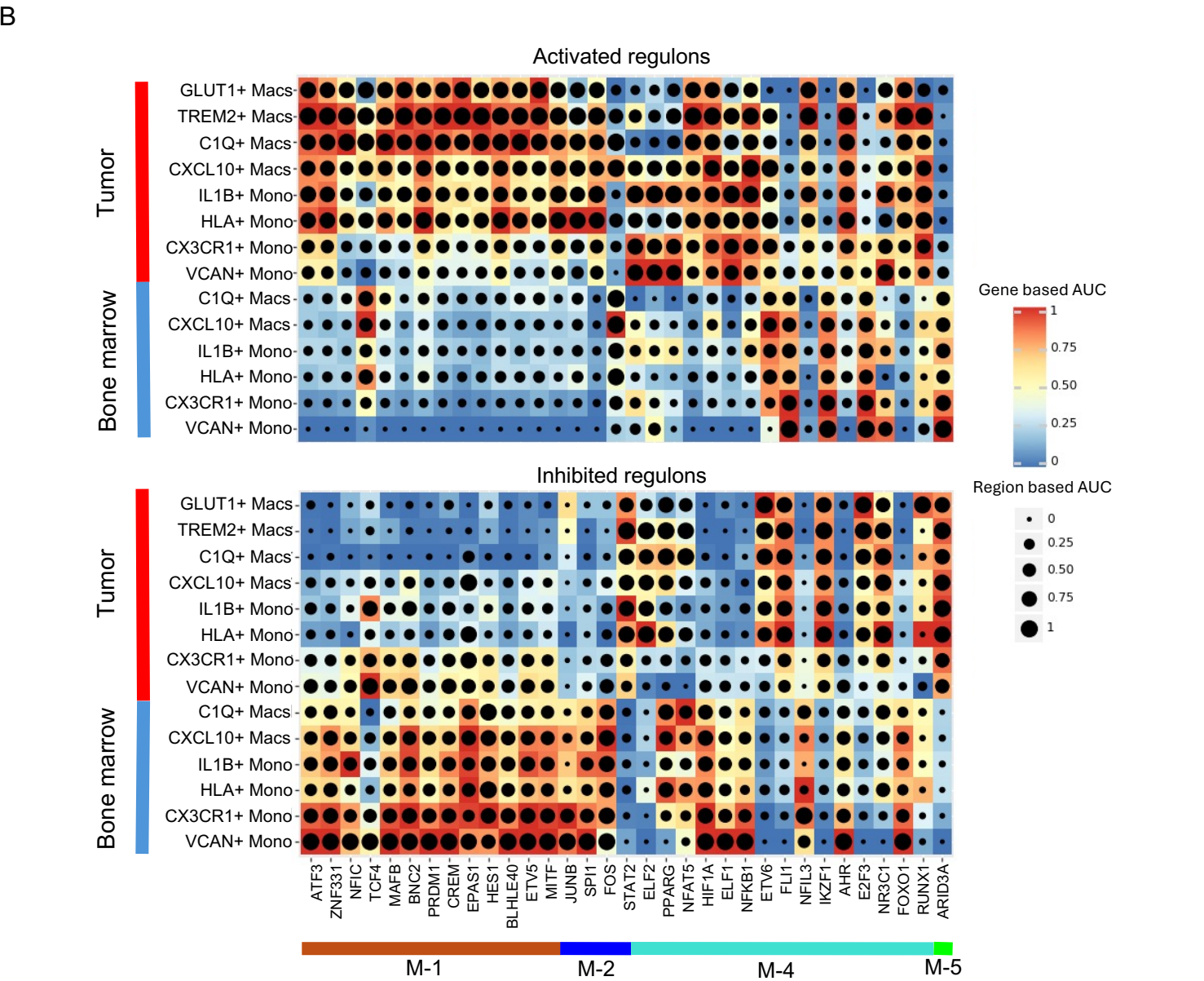
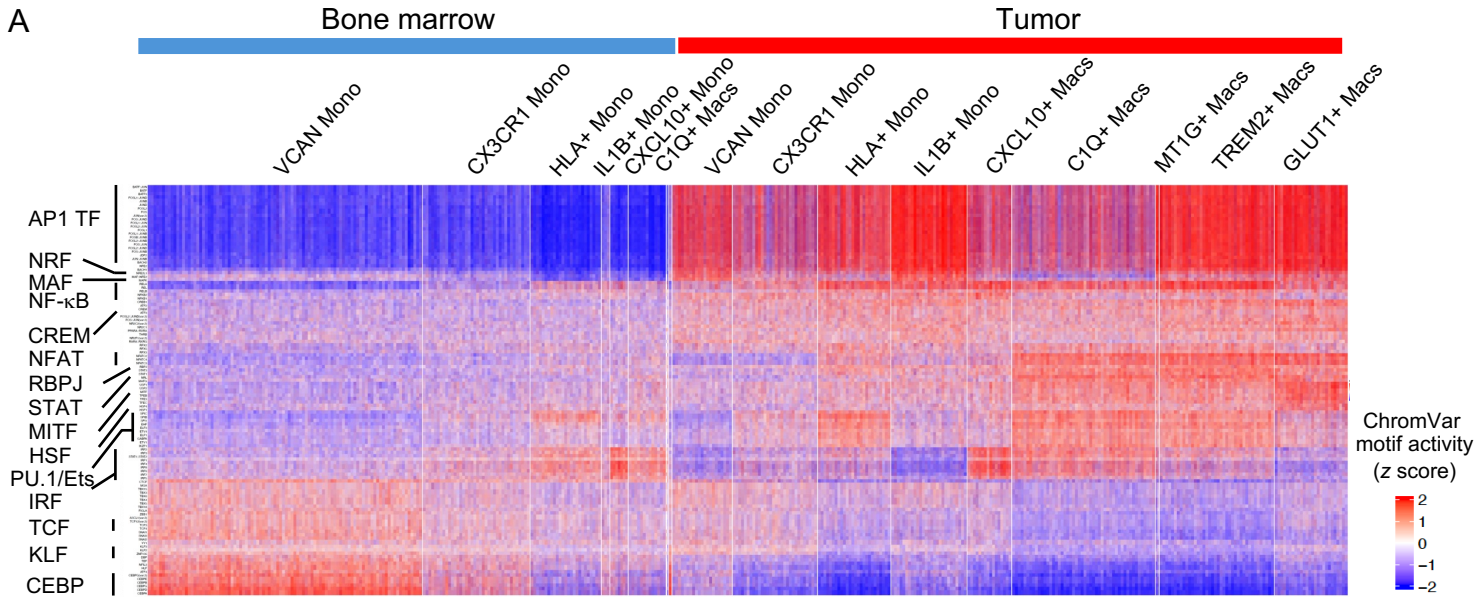


Figure 6

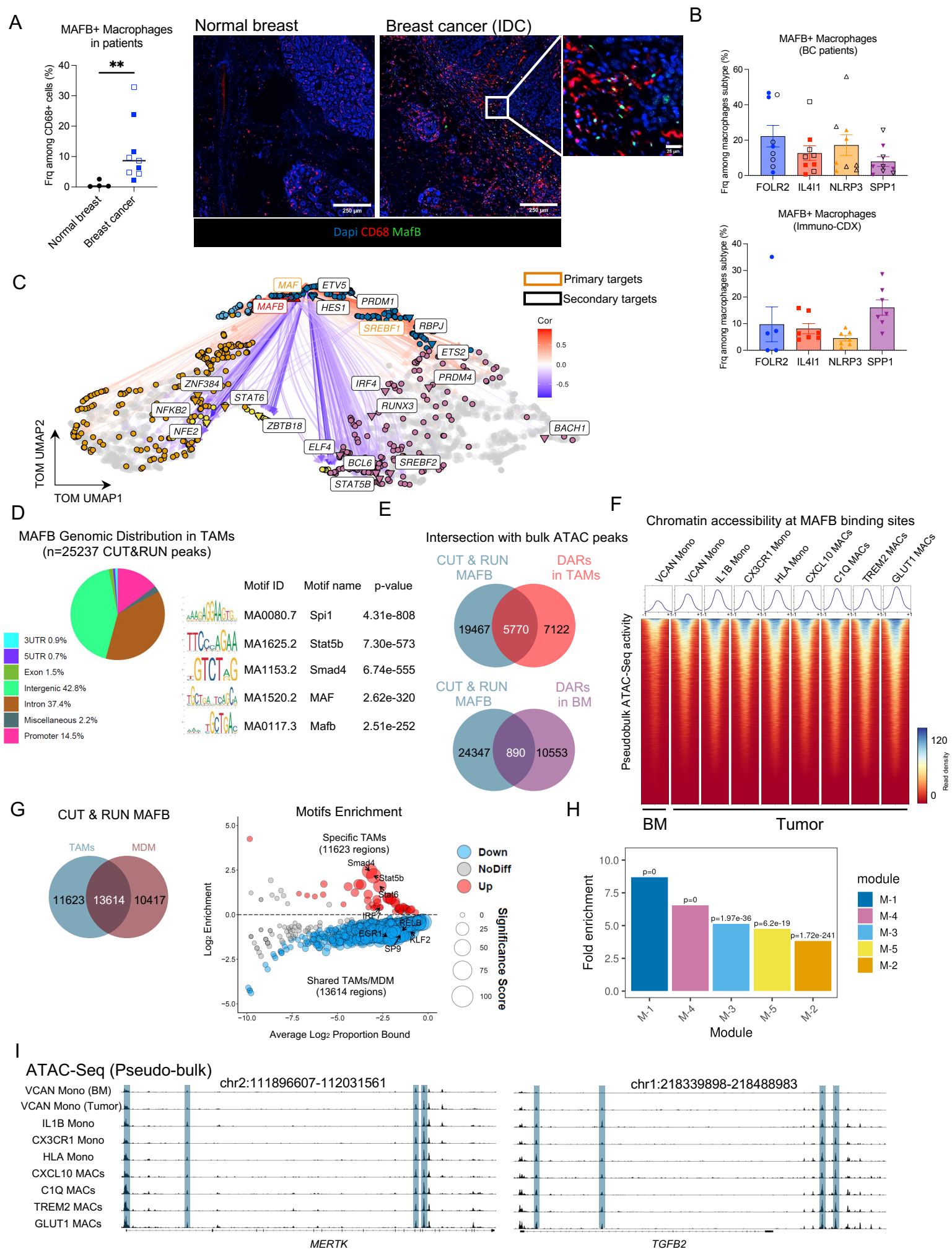
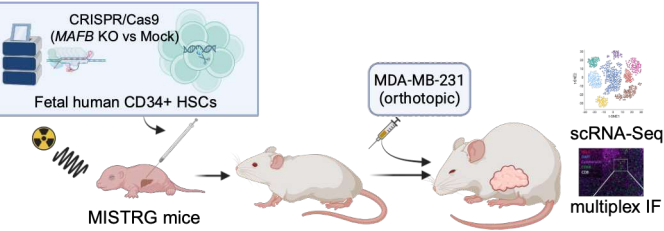
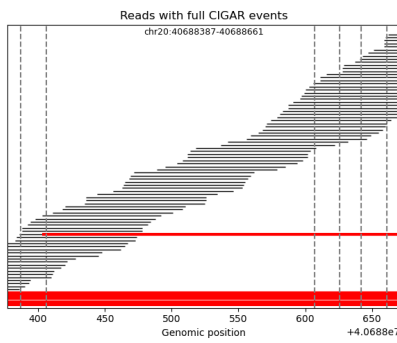


Figure 7

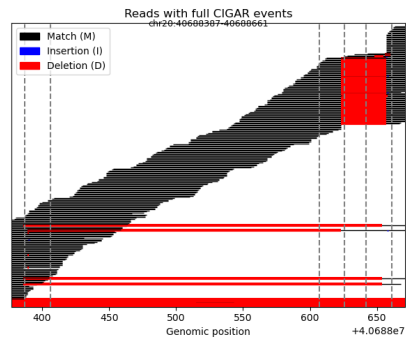
A



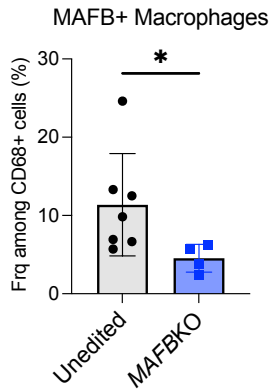
Unedited



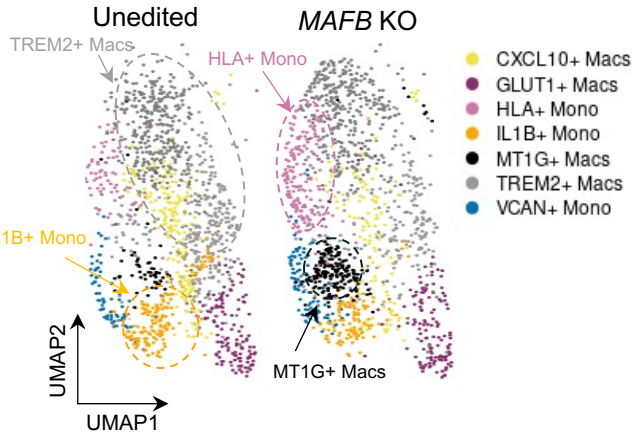
MAFB CRISPR



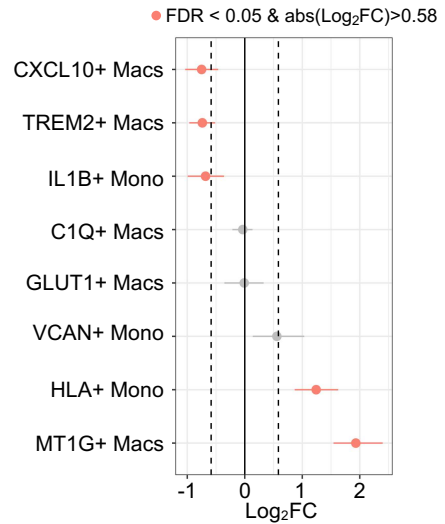
B



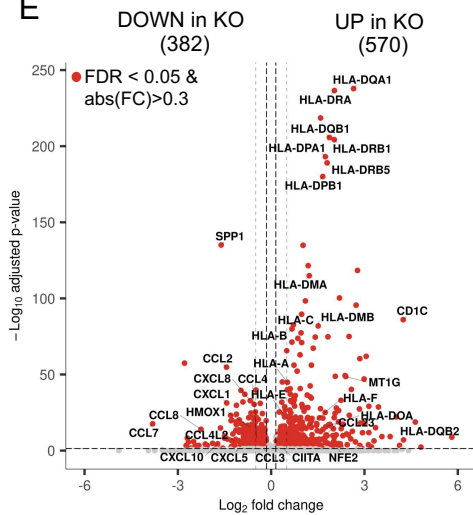
C



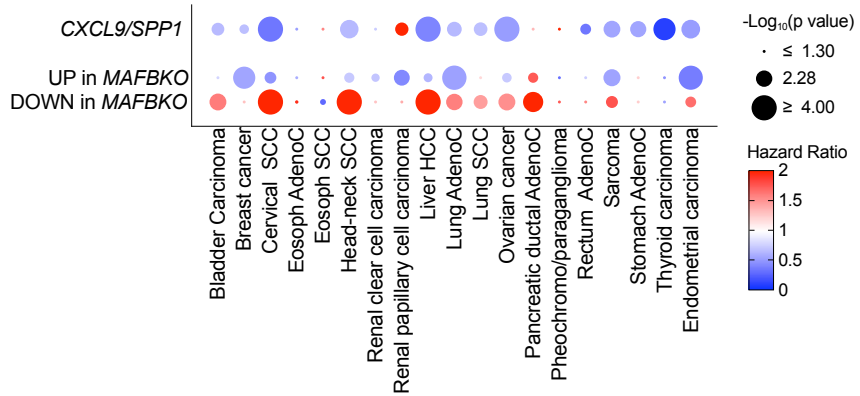
D



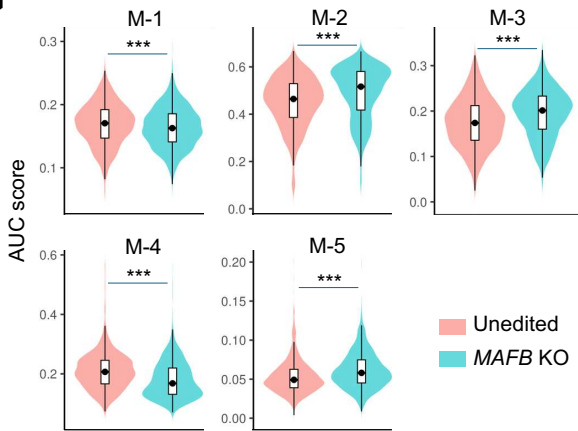
E



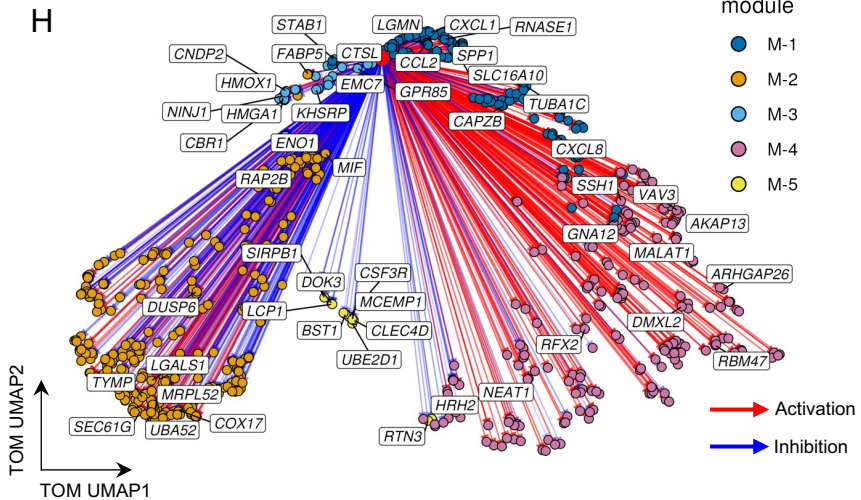
F



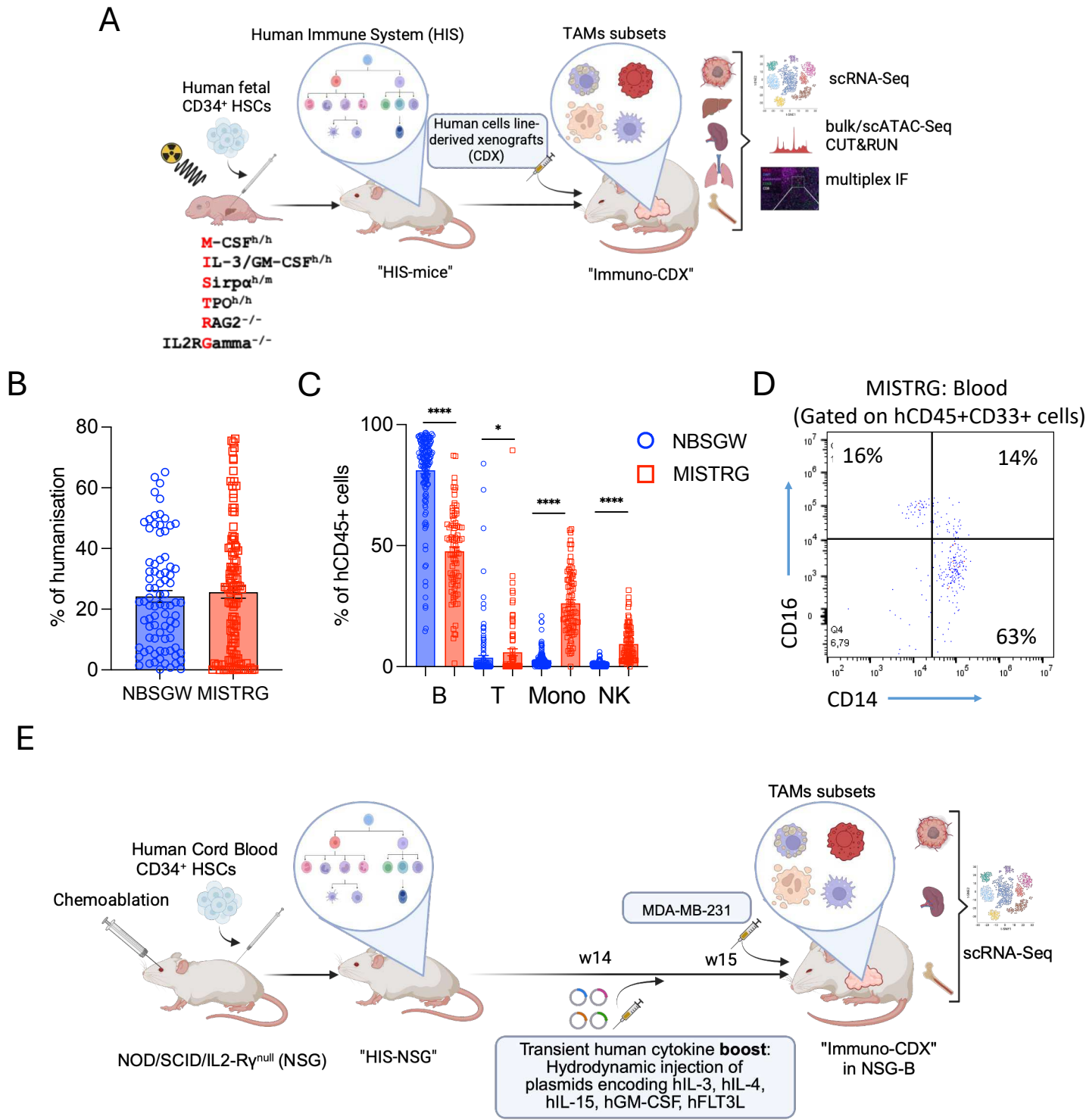
G



H

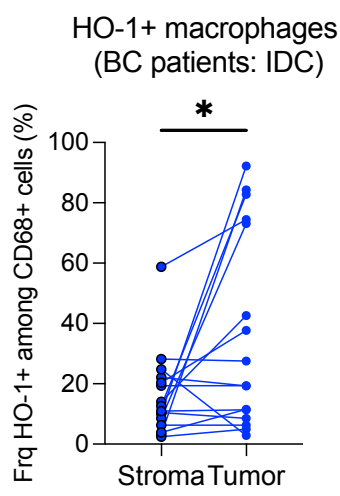


Suppl Figure 1, related to figure 1

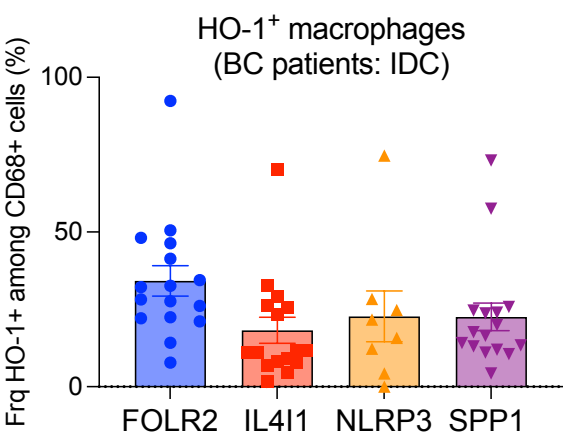


Suppl Figure 2, related to figure 3

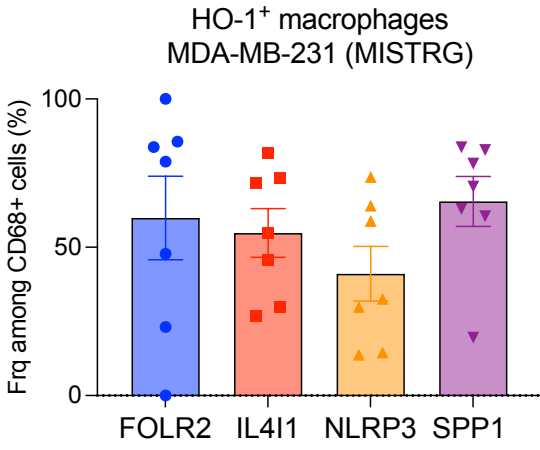
A



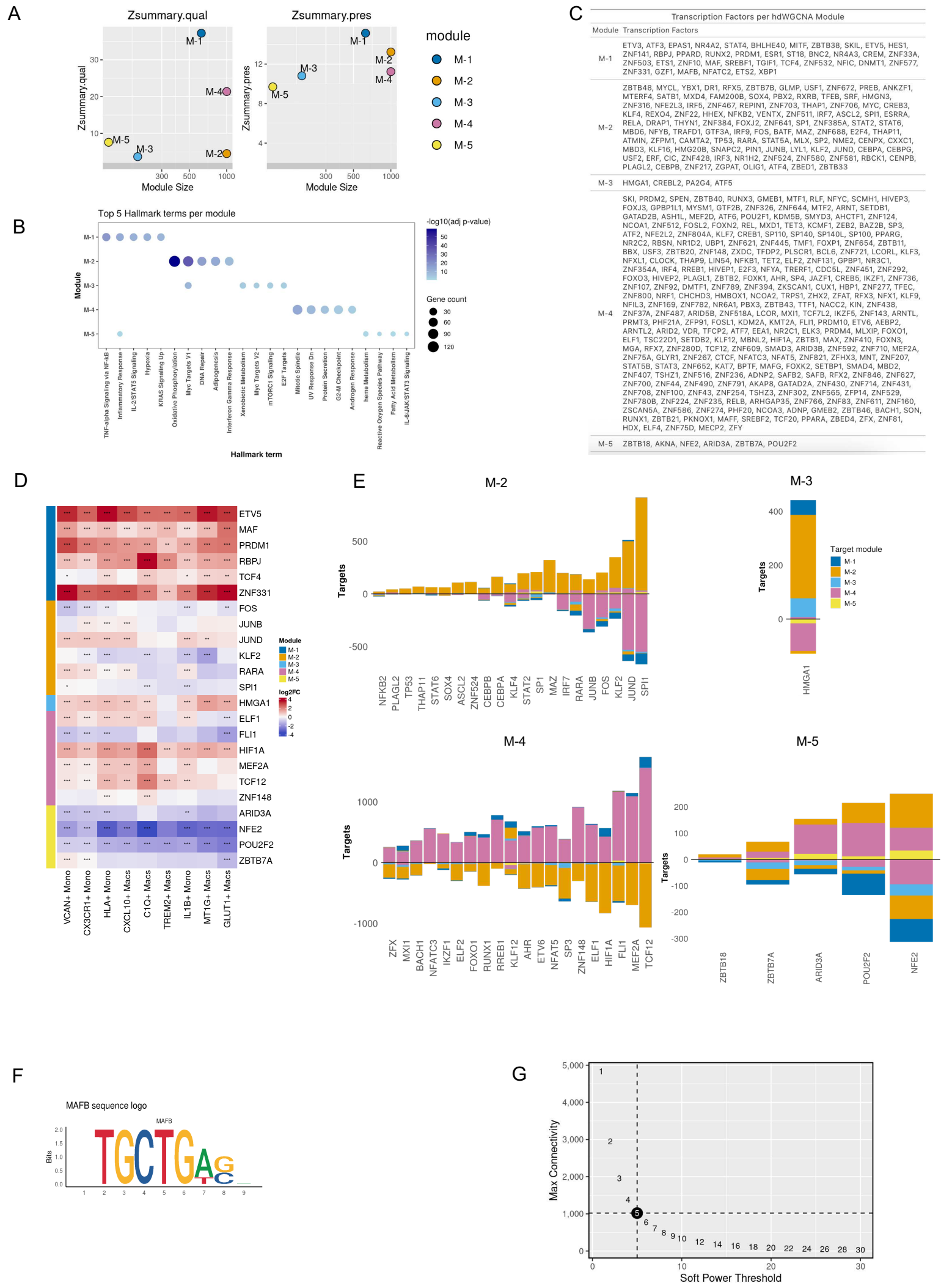
B



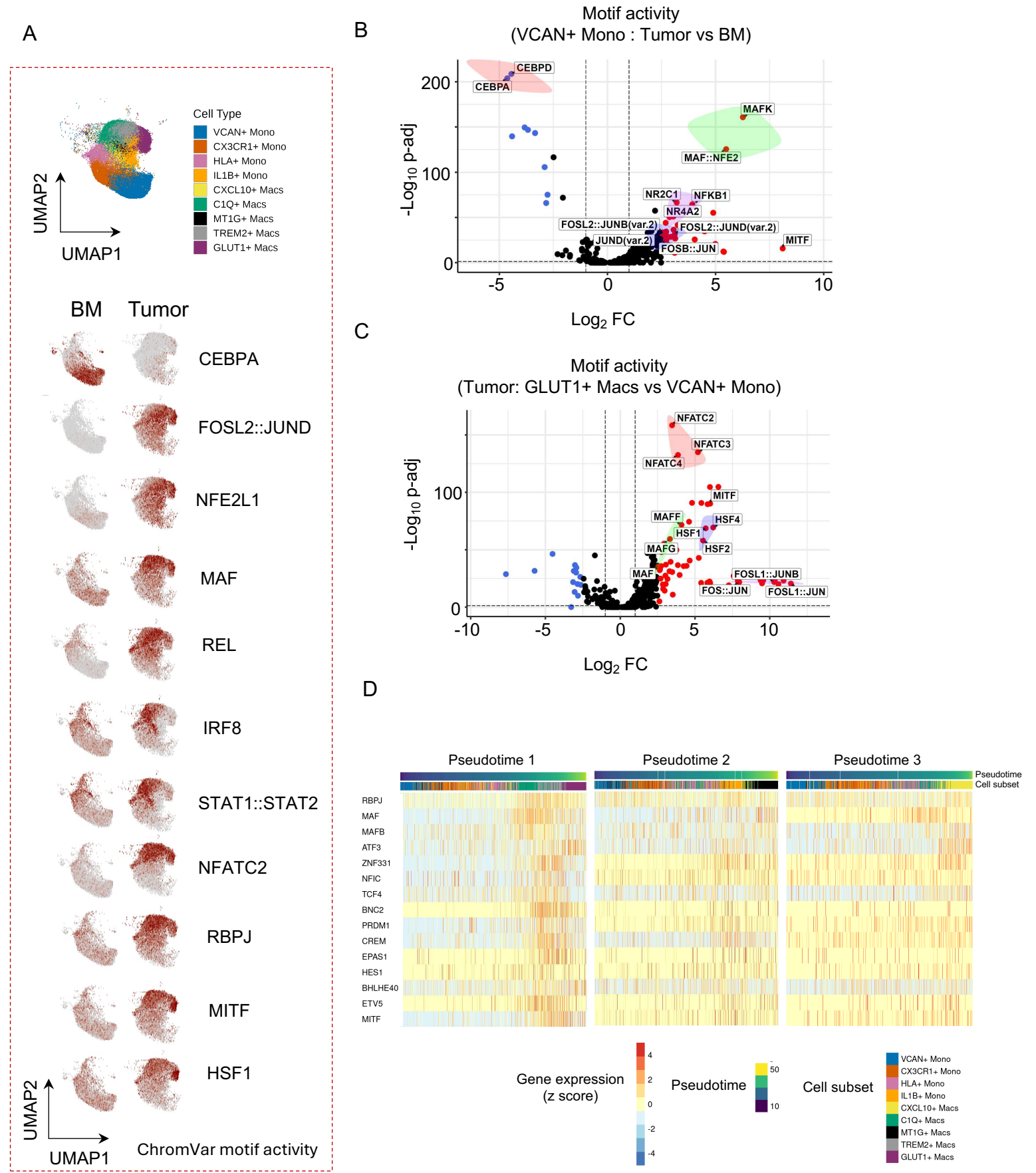
C



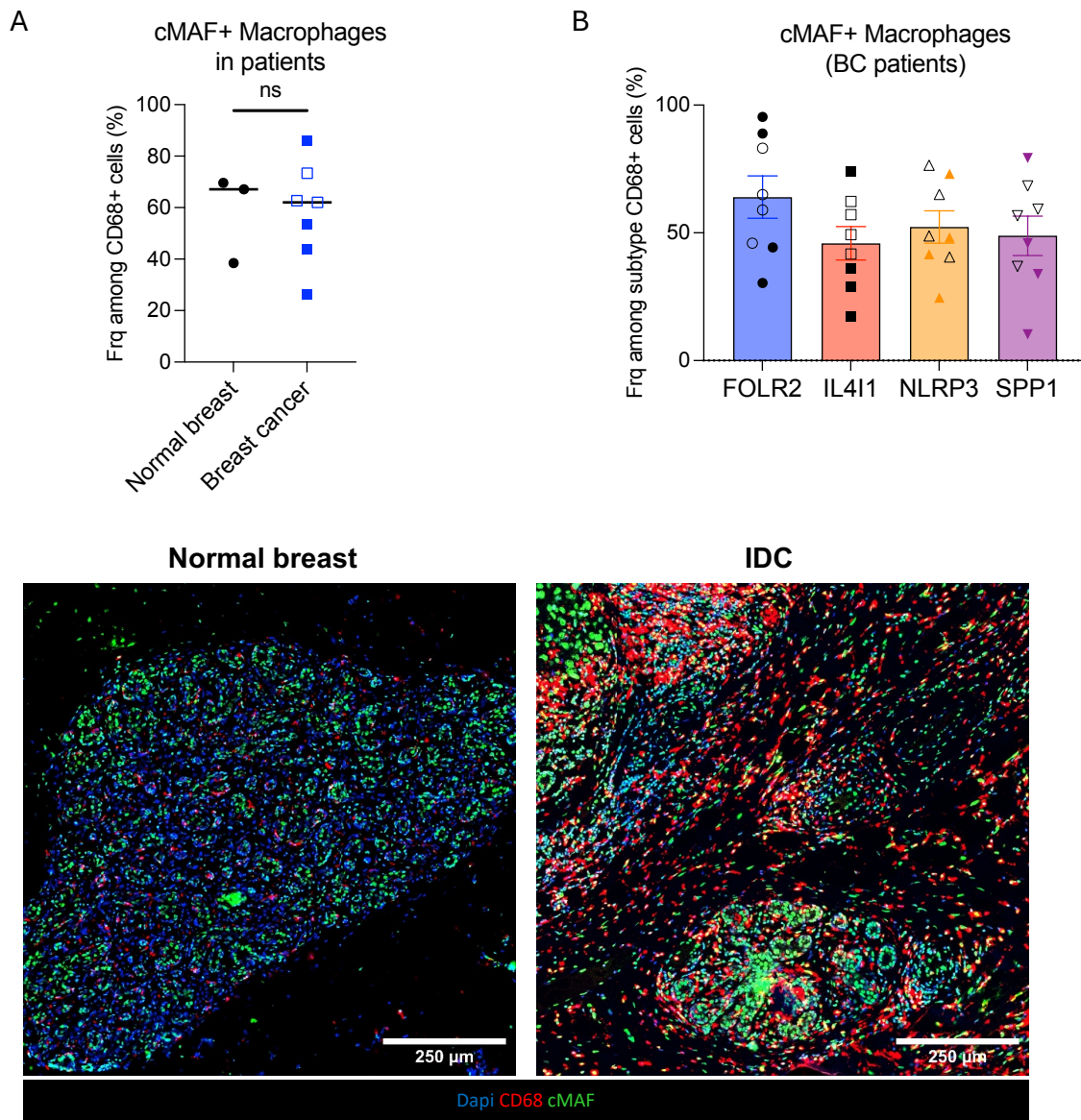
Suppl Figure 3, related to figure 4



Suppl Figure 4, related to figure 5



Suppl Figure 5, related to figure 6



Suppl Figure 6, related to figure 7

