Cell Reports Medicine, Volume *4*

Supplemental information

Single-cell dissection of Merkel cell carcinoma heterogeneity unveils transcriptomic plasticity and therapeutic vulnerabilities

Bhaba K. Das, Aarthi Kannan, Graham J. Velasco, Mikaela D. Kunika, Nils Lambrecht, Quy Nguyen, Haibo Zhao, Jie Wu, and Ling Gao

SUPPLEMENTAL INFORMATION

Supplementary Items and Figures in This File.

Figures S1, S2, S3, S4, S5, S6, S7

Table S1, S2, S5, and S7

Table S1 Summary of patient tumor samples and single cell RNA-seq analysis performed in this study. Related to STAR Methods, Figure 1, 2, and 3.

Table S2 Cancer-associated fibroblast (CAF), MESI-19 and SIG-14 gene list used in this study. Related to STAR Methods and Figure 1 and 5.

Table S5 MCC transcription factors identified by pySCENIC analysis. Related to STAR Methods.

Table S7 TaqMan gene expression primers used for quantitative PCR analyses. Related to STAR Methods.

Supplementary Spreadsheets. (Provided separately in Excel format, due to size.)

Table S3, S4, and S6

Table S3 List of differentially expressed genes used for generating heatmap. Related to Figure 1G.

Table S4 Complete list of Hallmark Pathway enrichment. Related to Figure 1I, 2C, 6B, and 7B.

Table S6 Master gene signature lists for epithelial-to-mesenchymal transition (EMT gene list) and mesenchymal markers (MES gene list). Related to STAR Methods.

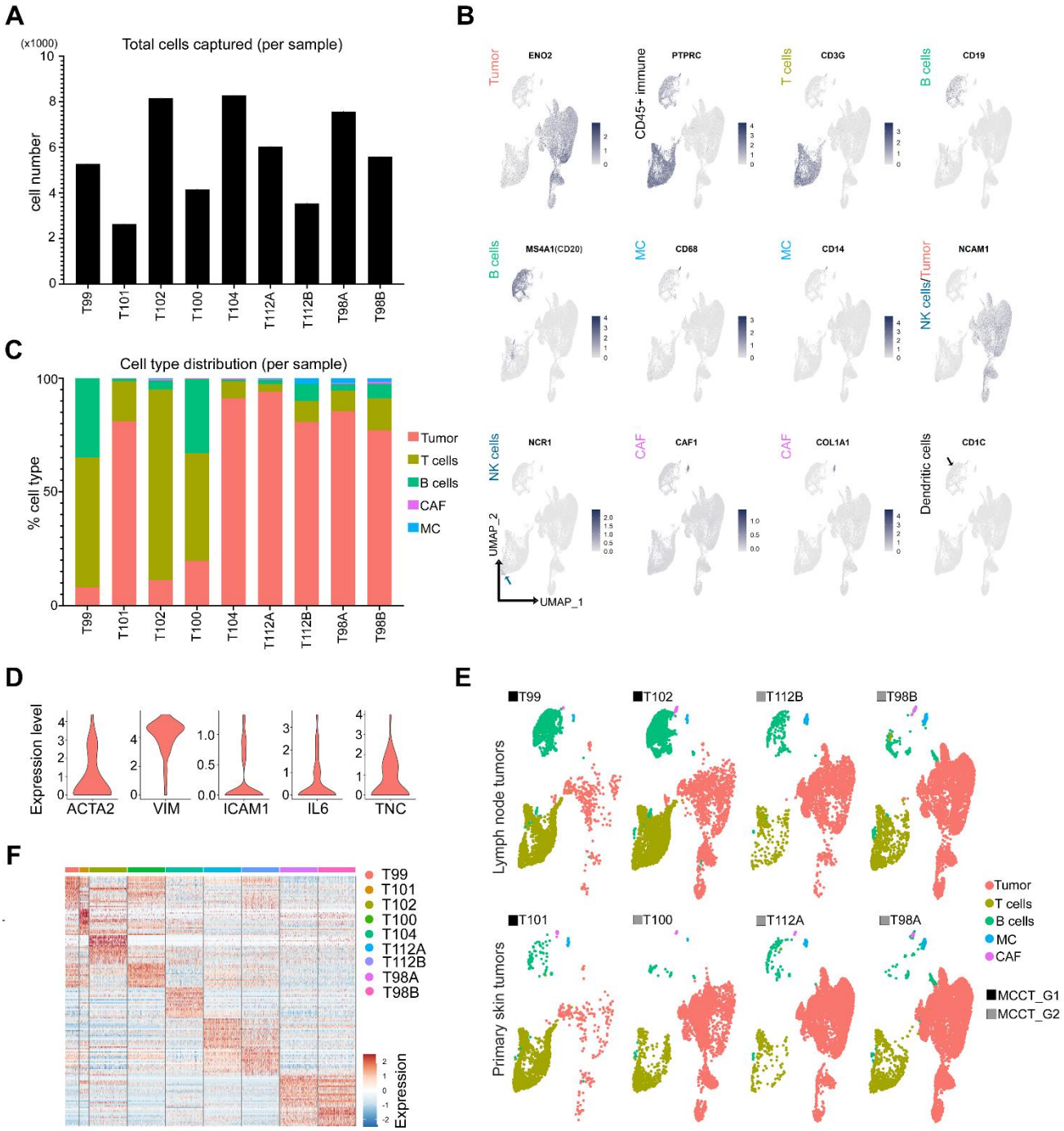

Figure S1. scRNA-seq of treatment-naïve MCC tumors. Related to Figure 1 and Table 1. (**A**) Total cells captured per sample. (**B**) Feature plots depicting distribution of select cell-type specific marker genes in UMAP clusters. MC: macrophages/monocytes; CAF: cancer-associated fibroblasts. (**C**) Percent composition of cell types across samples. (**D**) Violin plot of activation-associated marker gene expression in CAFs. (**E**) Split UMAP clusters of different cell types for lymph node and primary skin tumor samples. (**F**) Heatmap of top 20 variable genes across 500 randomly sampled tumor cells from each sample.

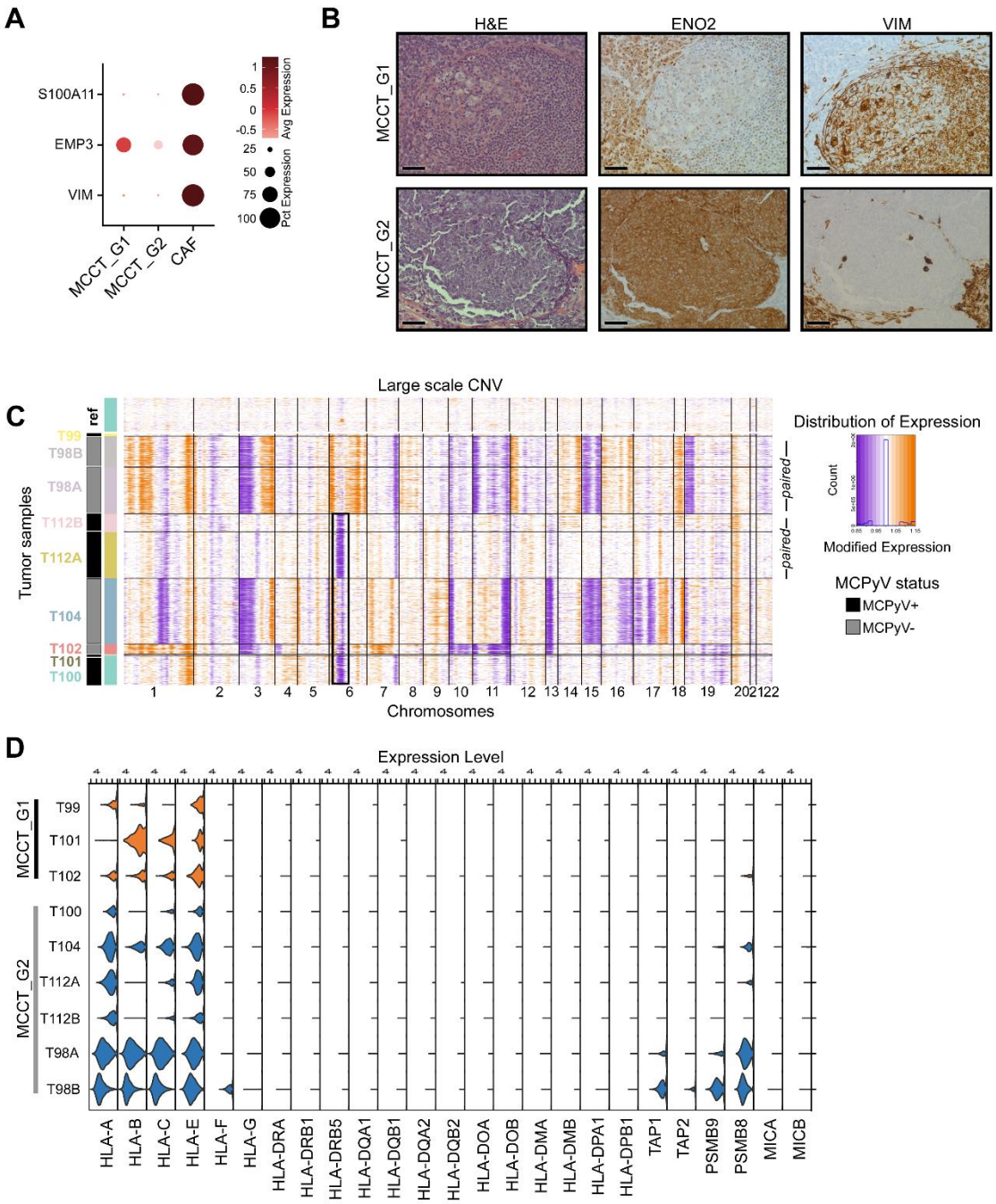

Figure S2. Partial EMT, inferred CNV and HLA expression profile of MCC treatment-naïve tumors. Related to Figure 1. (**A**) Scaled expression of selected mesenchymal markers across MCC tumor cells and CAFs depicting partial EMT in MCCT_G1 tumors. (**B**) Representative brightfield images of H&E (left panels) and immunohistochemistry staining of ENO2 (middle panels) and VIM (right panels) on consecutive sections of MCCT_G1 tumors (top row) and MCCT_G2 tumors (bottom row), at magnification 200x, scale bar = 35 µm. (**C**) Heatmap of inferred copy number variations (CNVs) normalized to the variations in CD3 T cell population demonstrating CNV by chromosome (columns) of individual cells (rows). (**D**) Violin plot depicting expression of MHC molecules in tumor cells across MCCT_G1 and MCCT_G2 tumors.

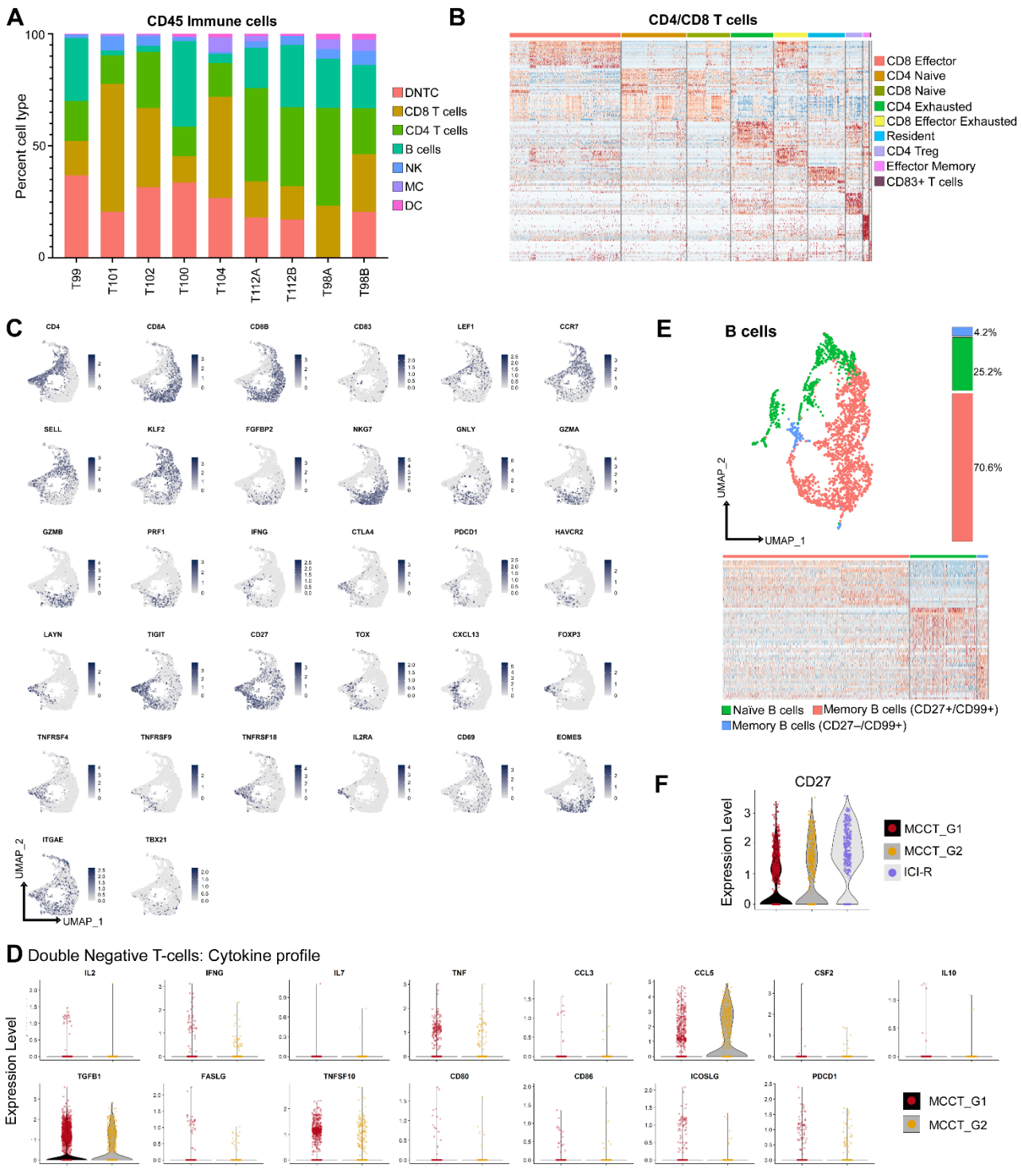

Figure S3. Immune landscape of treatment-naïve MCC tumors. Related to Figure 3. (**A**) Percentage distribution of CD45^{+} immune cells in each tumor sample. (**B**) Heatmap of top 20 variable genes in CD4/CD8 T cell subtypes. (**C**) Feature plots depicting distribution of subtypes and marker genes of functional status in CD4/CD8 T cells. (**D**) Violin plots of select cytokines in DNTCs. (**E**) UMAP clusters for CD19^{+}/CD20^{+} B cells (*n* = 2,331 cells) with color-matched histogram and heatmap of top 20 variable genes across the three B cell subtypes. (**F**) Expression of CD27 in MCC tumor samples.

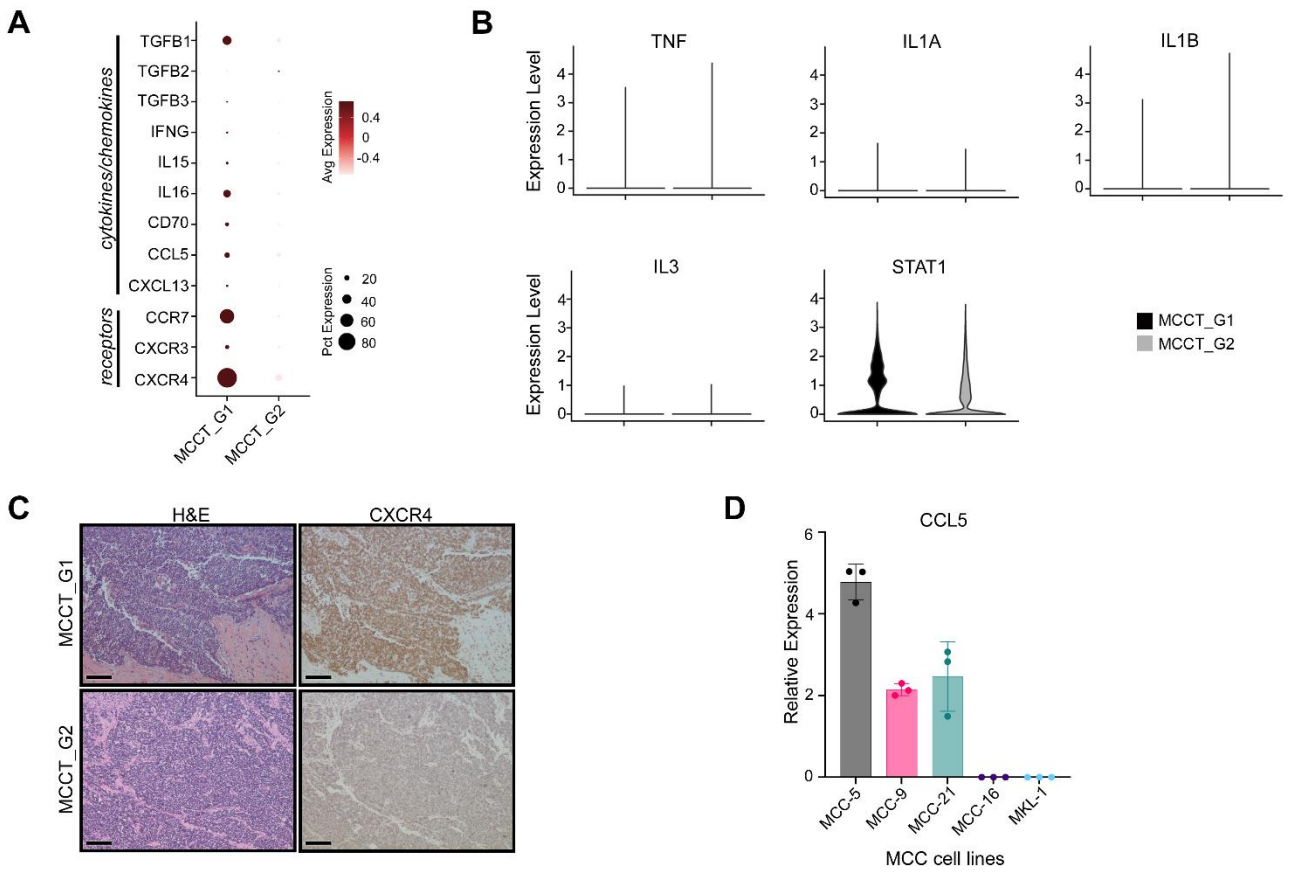

Figure S4. The 'mesenchymal-like' state in MCC is endowed with inflamed phenotype. Related to Figure 3. (**A**) Scaled expression of selected cytokines, chemokines, and chemokine receptors in MCCT_G1 and MCCT_G2 tumors. (**B**) Violin plots of classical inflammatory marker genes. (**C**) Representative brightfield images of H&E and CXCR4 immunohistochemistry staining in MCCT_G1 and MCCT_G2 tumors at 100x magnification, scale bar = 70 µm. (**D**) CCL5 expression in five MCC patient-derived cell lines, as detected by qPCR and normalized to *MRPS2* (triplicate runs, mean ± SD).

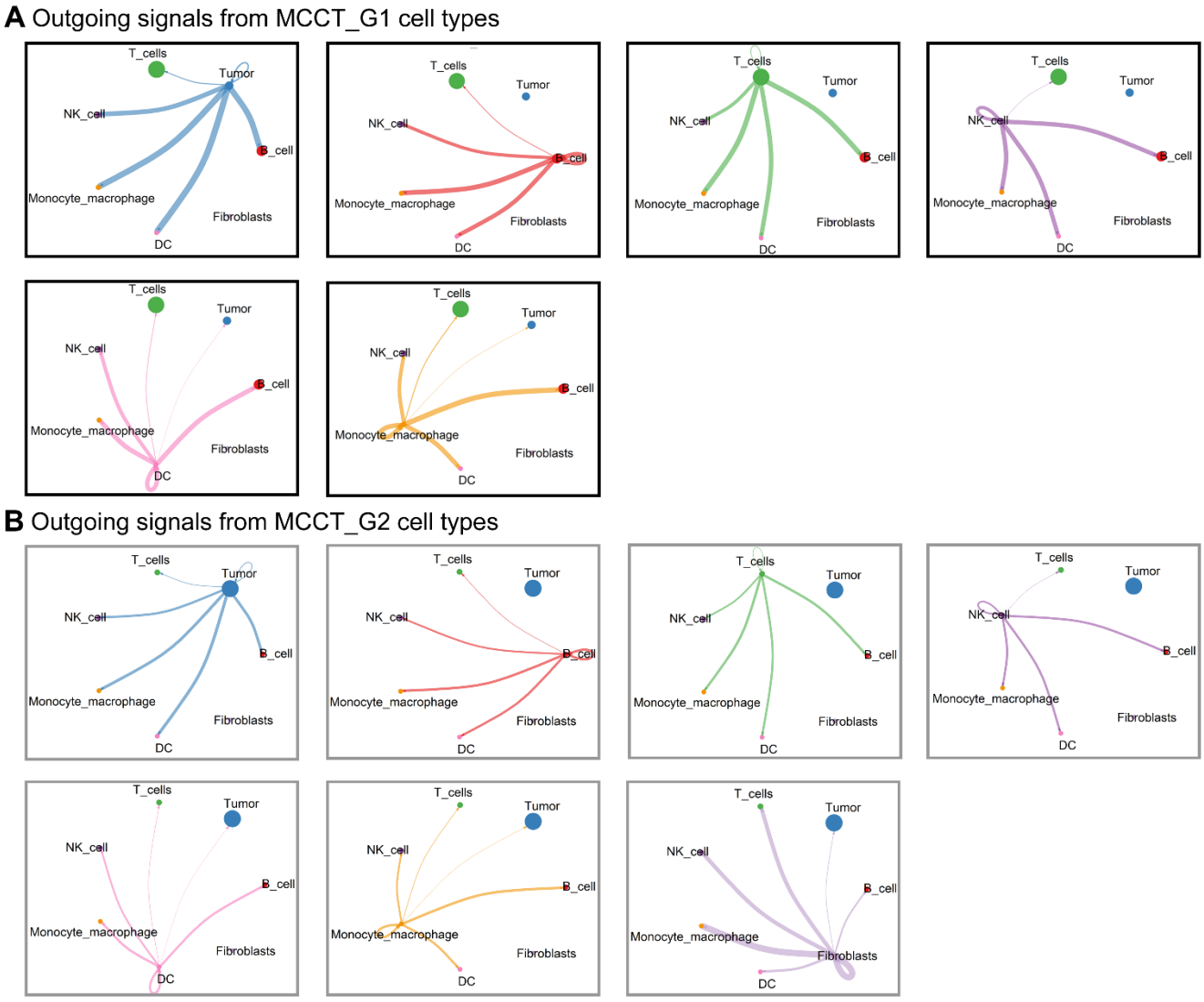

Figure S5. Individual chord diagrams of signals arising from each cell type. Related to Figure 4. Outgoing signals from different cell types in (**A**) MCCT_G1 and (**B**) MCCT_G2 tumors with chord thickness/weight representing signal strength.

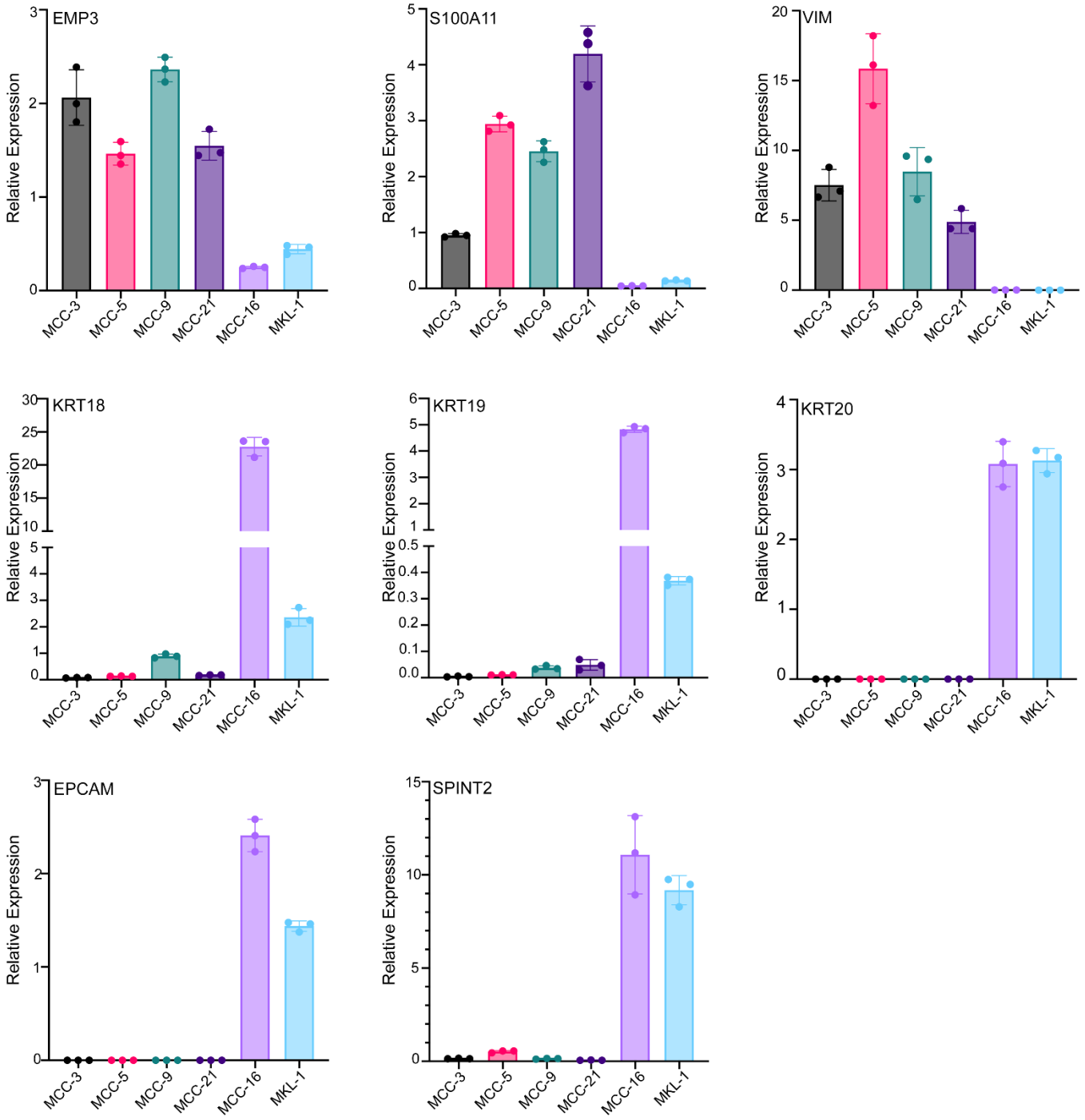

Figure S6. Characterization of patient-derived MCC cell lines established in our laboratory. Related to Figure 6. mRNA expression of selected epithelial and mesenchymal markers in MCC_G1 (MCC-3, MCC-5, MCC-9, MCC-21) and MCC_G2 (MCC-16, MKL-1) cell lines, normalized to *MRPS2*. Data presented as mean ± SD, *n* = 3.

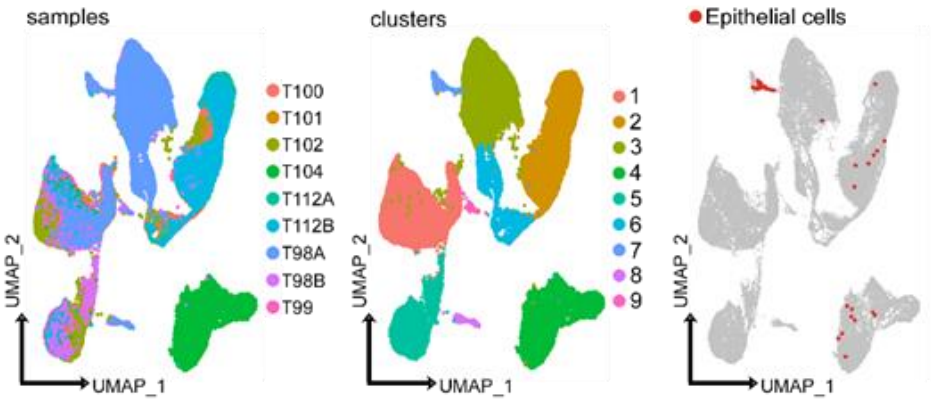

Figure S7. Distribution of KRT14+ epithelial cells in MCC patient tumors. Related to STAR Methods. Utilizing unintegrated tumor dataset, KRT14+ cells (highlighted as red dots) were identified by SingleR.

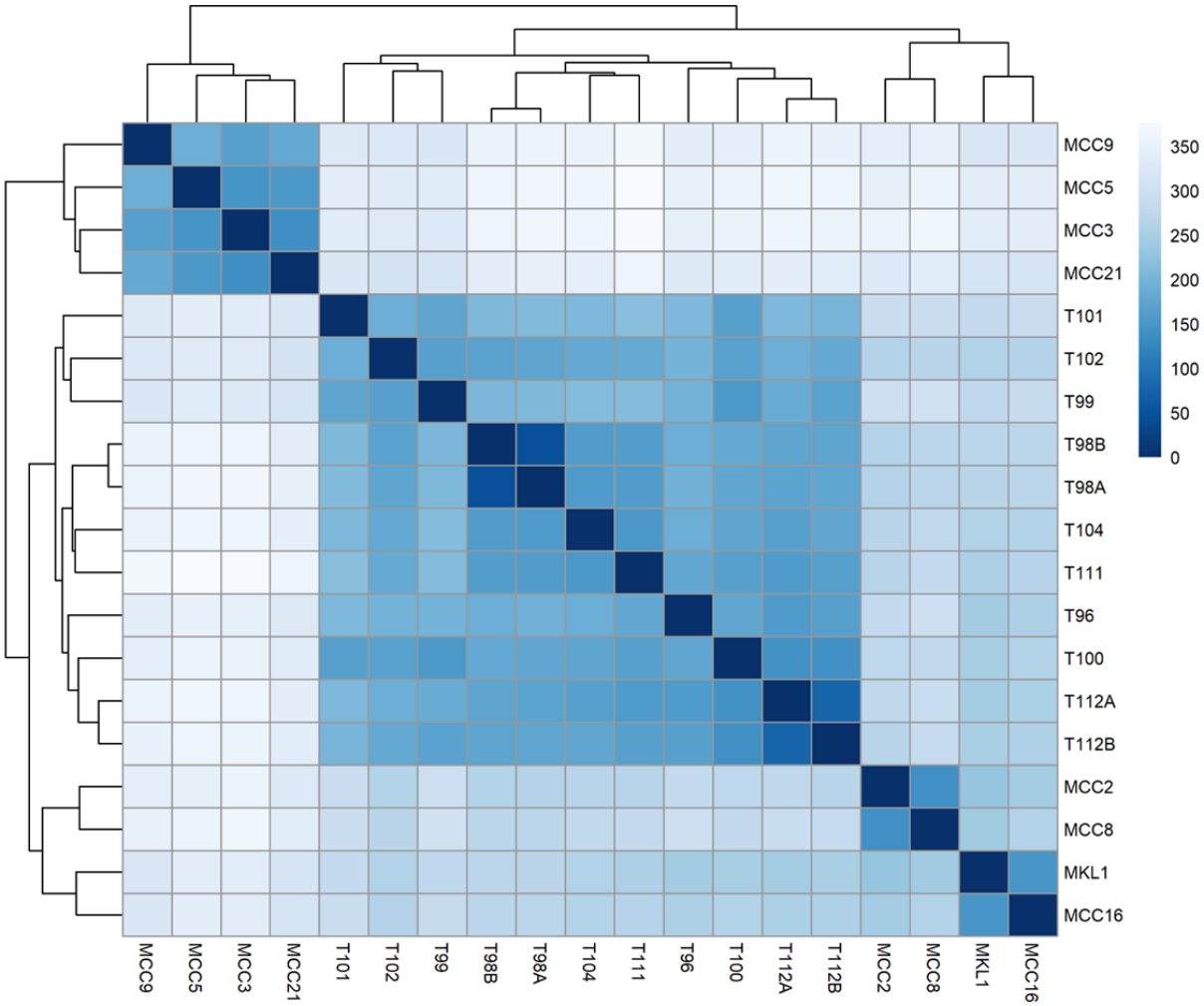

Figure S8. A heatmap of sample-to-sample distance with hierarchical clustering between MCC patient tumors and MCC primary cell lines. Related to STAR Methods. There is a lack of well-established algorithms for integrating bulk RNA-seq and pseudobulk on scRNA-seq datasets. A technically limited analysis depicting sample similarities was performed by coarsely combining the count matrix from of 8 MCC cell lines (bulk RNA-seq) and 11 patient tumors (pseudobulk on scRNA-seq) followed by normalization and distance calculation.

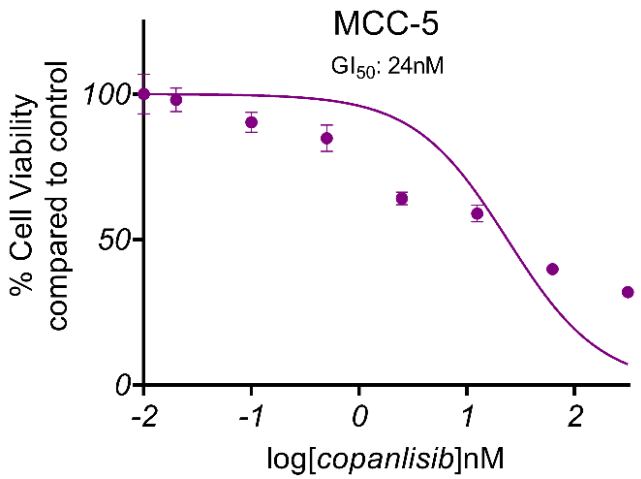

Figure S9. Copanlisib sensitivity in MCC-5 cell line with 'mesenchymal-like' state. Related to STAR Methods and Figure 7. MCC-5 cells were treated with serial concentrations of copanlisib for 72h, then assessed by CCK-8 colorimetric cell proliferation assay. Data presented as mean ± SD for each dose, *n* = 6 per dose, with half maximal growth inhibitory concentration (GI_{50}) as analyzed by nonlinear regression model using GraphPad Prism.

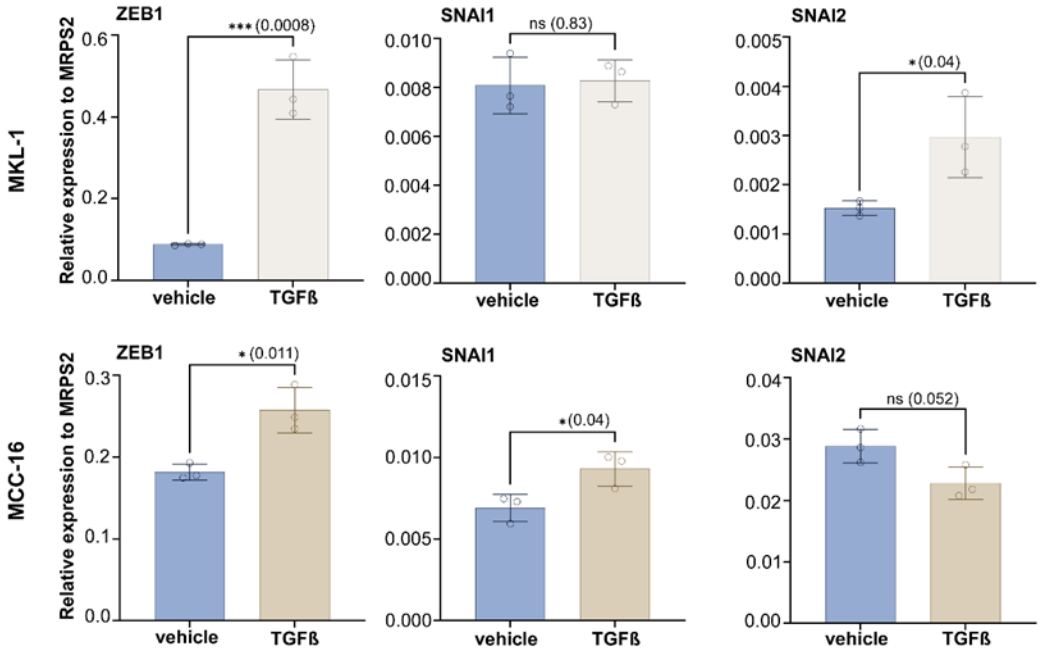

Figure S10. Expression of EMT-TFs upon TGFβ treatment in MKL-1 and MCC-16 cells. Related to Figure 7. MKL-1 and MCC-16 cells were treated with vehicle (DMSO) or 2ng/ml TGFβ for 24 hours, followed by qPCR. Data presented as mean ±SD, n=3. *p < 0.05 ***p < 0.0005 as analyzed by unpaired Student's t-test.

Supplementary Tables 1, 2, 5 and 7

Table S1. Summary of patient tumor samples and single cell RNA-seq analysis performed in this study. Related to STAR Methods, Figure 1, 2, and 3. (A) Patient demographics and tumor characteristics. (B) Summary of Library alignment to human (GRCh38) 2020-A. (C) Summary of each object analyzed in this study.

(A)

Sample	Tumor ID	Resected Tumor Location	Gender	Age	Race	MCPyV status	Prior therapy
1	T96	Lymph node	female	70	white	positive	pembrolizumab
2	T111	Skin	male	72	white	negative	pembrolizumab
3	T98A	Skin	male	80	white	negative	none
	T98B	Lymph node	male	80	white	negative	none
4	T99	Lymph node	male	51	white	positive	none
5	T100	Skin	female	56	unknown	positive	none
6	T101	Skin	male	80	white	positive	none
7	T102	Lymph node	male	74	white	negative	none
8	T104	Parotid gland	female	77	white	negative	none
9	T112A	Skin	female	57	white	positive	none
	T112B	Lymph node	female	57	white	positive	none

(B)

Sample	Estimated number of cells	Mean Reads per cell	Median Genes per cell
T96	6978	91728	3484
T98A	7573	81745	4436
T98B	5589	88191	4335
T99	5272	98381	1545
T100	4157	135101	3352
T101	2636	218000	1930
T102	8161	61443	1805
T104	8287	67708	3340
T111	7318	73905	4102
T112A	6036	73022	3942
T112B	3537	160895	2529

(C)

Object	Description	Number of cells	Mean Reads per cell	Mean Genes per cell
Treatment-naïve object	Contains all cell types from 9 samples	46027	11015	3181
Treatment-naïve Tumor	Contains only tumor cells	22978	13498	3910
Treatment-naïve Immune	Contains CD45+ cells	12796	5808	1850
Treatment-naïve CD4CD8	Contains CD4+/CD8+ cells	3169	5398	1806
Treatment-naïve B cells	Contains CD19+/CD20+ cells	2331	5987	1762
Naïve-ICI-R object	Contains all cell types from 11 samples	58936	11555	3304
Naïve-ICI-R Tumor	Contains only tumor cells	35796	14290	4038
Naïve-ICI-R Immune	Contains CD45+ cells	14895	5751	1825

Table S2. Cancer-associated fibroblast (CAF), MESI-19 and SIG-14 gene list used in this study.
Related to STAR Methods and Figure 1 and 5.

#	CAF gene list
1	COL1A1
2	COL3A1
3	FAP
4	SPARC
5	THY1
6	DCN
7	PDGFRB
8	FBLN1
9	S100A4
10	ITGA5
11	ACTA2
12	COL5A2
13	ADAM12
14	COL6A3
15	LRRC15
16	COL5A1
17	COL1A2
18	VCAN
19	POSTN
20	COL11A1
21	THBS2
22	LUM
23	NTM
24	AEBP1
25	COL6A2
26	PCOLCE
27	GLT8D2
28	ASPN
29	BGN
30	ISLR
31	RARRES2
32	TAGLN
33	CTHRC1
34	P4HA3
35	GREM1
36	MFAP5
37	GAS1
38	COMP
39	EFEMP2

#	MESI-19 gene list
1	CLEC2B
2	CXCR4
3	EMP3
4	FLNA
5	IFITM2
6	IL32
7	JUNB
8	MYH9
9	NR3C1
10	PTPRC
11	S100A11
12	SAMSN1
13	SRGN
14	TIMP1
15	TRIM56
16	VIM
17	WIPF1
18	ZEB2
19	ZYX

#	SIG-14 gene list
1	ISG15
2	CD74
3	ISG20
4	IFI44L
5	B2M
6	BTG1
7	HLA-DRA
8	LTB
9	RPL39
10	IL7R
11	CCR7
12	EMP3
13	CXCR4
14	VIM

40	LOXL1
41	MYL9
42	COL8A2
43	SGCD
44	SCARF2
45	TPM2
46	SPOCK1
47	HTRA1
48	LGALS1
49	ZEB1
50	ZEB2
51	COL6A1
52	COL5A3
53	INHBA
54	COL12A1

Table S5. MCC transcription factors identified by pySCENIC analysis. Related to STAR Methods.

MCCT_G1		MCCT_G2	
regulon	**Z-score**	**regulon**	**Z-score**
ALX1(+)	2.395458542	TAF7(+)	0.312711282
MEF2C(+)	2.150993916	BARHL1(+)	0.265176081
ATOH8(+)	1.684436403	HBP1(+)	0.256969399
FOXO1(+)	1.419356562	ZNF148(+)	0.24670525
ELF1(+)	1.335634403	STAT1(+)	0.245591583
IKZF1(+)	1.289873586	MYC(+)	0.239314246
RUNX3(+)	1.285482879	TLX3(+)	0.236735923
SPIB(+)	1.220197775	IRF9(+)	0.235645606
NEUROD1(+)	1.217931933	PSMD12(+)	0.22530856
STAT5A(+)	1.212667568	YY1(+)	0.221973089
ZEB1(+)	1.205305108	MXI1(+)	0.213696911
STAT6(+)	1.204616792	HOXB4(+)	0.20653584
IRF5(+)	1.1876485	NFE2L2(+)	0.20108852
ELK3(+)	1.147263571	LBX1(+)	0.195877766
FLI1(+)	1.115508028	MAX(+)	0.190698815
MYOG(+)	1.107531862	PRRX2(+)	0.184847361
SPI1(+)	1.103700598	RAX(+)	0.184618116
SP5(+)	1.096817056	FOXK1(+)	0.178893525
IRF4(+)	1.091762814	ARNTL(+)	0.171713951
IRF2(+)	1.068457884	FOXP2(+)	0.171684335
IRF8(+)	1.037399378	ZFX(+)	0.169753284
HOXB2(+)	1.032348847	THAP1(+)	0.169202894
SP1(+)	1.019352929	ATF2(+)	0.168495567
CREM(+)	1.015311736	IRF1(+)	0.16821169

Table S7. TaqMan gene expression primers used for quantitative PCR analyses. Related to STAR Methods.

Gene Symbol /	Source	Assay ID	Cat #
MRPS2	Life Technologies	Hs00211334_m1	4331182
EMP3	Life Technologies	Hs00171319_m1	4331182
S100A11	Life Technologies	Hs01055944_g1	4331182
VIM	Life Technologies	Hs00185584_m1	4331182
KRT18	Life Technologies	Hs02827483_g1	4331182
KRT19	Life Technologies	Hs00761767_s1	4331182
KRT20	Life Technologies	Hs00300643_m1	4331182
CCL5	Life Technologies	Hs00982282_m1	4331182
EPCAM	Life Technologies	Hs00158980_m1	4331182
SPINT2	Life Technologies	Hs01070442_m1	4331182
TWIST1	Life Technologies	Hs04989912_s1	4331182
TWIST2	Life Technologies	Hs02379973_s1	4331182
SNAI1	Life Technologies	Hs00195591_m1	4331182
SNAI2	Life Technologies	Hs00161904_m1	4331182
ZEB1	Life Technologies	Hs01566408_m1	4331182
ZEB2	Life Technologies	Hs00207691_m1	4331182