Title: Single-cell RNA-sequencing of migratory breast cancer cells: discovering genes associated with cancer metastasis

Authors and affiliations:
Yu-Chih Chen[1, 2], Saswat Sahoo[3], Riley Brien[1], Seungwon Jung[1], Brock Humphries[4], Woncheol Lee[1], Yu-Heng Cheng[1], Zhixiong Zhang[1], Kathryn E. Luker[4], Max S. Wicha[2], Gary D. Luker[3,4,5], and Euisik Yoon[*1, 3, 6]

[1] Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, Ann Arbor, MI 48109-2122;
[2] Forbes Institute for Cancer Discovery, University of Michigan, 2800 Plymouth Rd., Ann Arbor, MI 48109, USA;
[3] Department of Biomedical Engineering, University of Michigan, 2200 Bonisteel, Blvd. Ann Arbor, MI 48109-2099, USA
[4] Center for Molecular Imaging, Department of Radiology, University of Michigan, 109 Zina Pitcher Place, Ann Arbor, MI 48109-2200, USA;
[5] Department of Microbiology and Immunology, University of Michigan, 109 Zina Pitcher Place, Ann Arbor, MI 48109-2200, USA;
[6] Center for Nanomedicine, Institute for Basic Science (IBS) and Graduate Program of Nano Biomedical Engineering (Nano BME), Yonsei University, Seoul 03722, Korea.

*Corresponding authors

Yu-Chih Chen
1301 Beal Avenue, Ann Arbor, MI 48109-2122, USA
Tel: 734-272-7113; E-mail: yuchchen@umich.edu.

Euisik Yoon
1301 Beal Avenue, Ann Arbor, MI 48109-2122, USA
Tel: 734-615-4469; E-mail: esyoon@umich.edu.

Keywords:
Cell Migration; RNA-Seq; Single Cell; Microfluidics; Cancer Metastasis.

Supplementary Table 1. Top-ranked 50 pathways (NCI-Nature pathway database) of migratory MDA-MB-231 cells as compared to wild-type ones.

Pathway	**P-value**
Validated targets of C-MYC transcriptional activation	1.03E-07
a6b1 and a6b4 Integrin signaling	0.00001534
ErbB1 downstream signaling	0.0003748
TGF-beta receptor signaling	0.0004387
Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)	0.0006374
Regulation of nuclear beta catenin signaling and target gene transcription	0.0008069
CDC42 signaling events	0.0009868
Class I PI3K signaling events mediated by Akt	0.001656
Integrin-linked kinase signaling	0.002127
HIF-1-alpha transcription factor network	0.002192
Trk receptor signaling mediated by PI3K and PLC-gamma	0.002345
Insulin-mediated glucose transport	0.002349
IGF1 pathway	0.002859
PDGFR-beta signaling pathway	0.002913
Posttranslational regulation of adherens junction stability and dissassembly	0.003242
BARD1 signaling events	0.003447
p38 signaling mediated by MAPKAP kinases	0.003957
Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling	0.005771
Arf6 downstream pathway	0.006451
RAC1 signaling pathway	0.006782
p73 transcription factor network	0.007443
ErbB2/ErbB3 signaling events	0.007487
p75(NTR)-mediated signaling	0.008848
IL2 signaling events mediated by PI3K	0.009019
SHP2 signaling	0.009388
IL6-mediated signaling events	0.009555
IL3-mediated signaling events	0.0103
Arf6 trafficking events	0.0134
Signaling events regulated by Ret tyrosine kinase	0.01343
Signaling mediated by p38-alpha and p38-beta	0.01343
Aurora B signaling	0.01519
Internalization of ErbB1	0.0171
Validated nuclear estrogen receptor alpha network	0.01827
Stabilization and expansion of the E-cadherin adherens junction	0.01917
Role of Calcineurin-dependent NFAT signaling in lymphocytes	0.02009
Signaling events mediated by TCPTP	0.0214
Nephrin/Neph1 signaling in the kidney podocyte	0.02145
E-cadherin signaling in keratinocytes	0.02204
Regulation of Telomerase	0.0235
PDGFR-alpha signaling pathway	0.02585

FOXA1 transcription factor network	0.02638
Direct p53 effectors	0.02654
AP-1 transcription factor network	0.02751
mTOR signaling pathway	0.02751
N-cadherin signaling events	0.0309
PLK1 signaling events	0.03207
IL5-mediated signaling events	0.03394
JNK signaling in the CD4+ TCR pathway	0.03394
Validated transcriptional targets of AP1 family members Fra1 and Fra2	0.03455
FoxO family signaling	0.03851

Supplementary Table 2. Top-ranked 50 pathways (NCI-Nature pathway database) of migratory SUM159 cells as compared to wild-type ones.

Pathway	**P-value**
HIF-1-alpha transcription factor network	2.44E-05
Validated nuclear estrogen receptor alpha network	6.22E-05
IL6-mediated signaling events	7.71E-05
Validated targets of C-MYC transcriptional activation	8.29E-05
Syndecan-4-mediated signaling events	0.000154
Aurora B signaling	0.000277
PDGFR-beta signaling pathway	0.000427
ErbB1 downstream signaling	0.000455
Aurora A signaling	0.000785
Syndecan-2-mediated signaling events	0.001287
Direct p53 effectors	0.002332
E2F transcription factor network	0.002538
PLK1 signaling events	0.004409
AP-1 transcription factor network	0.004725
RAC1 signaling pathway	0.004898
CDC42 signaling events	0.005364
Integrins in angiogenesis	0.006848
IL2 signaling events mediated by PI3K	0.007645
C-MYB transcription factor network	0.008473
Beta1 integrin cell surface interactions	0.00868
Beta3 integrin cell surface interactions	0.00875
Nectin adhesion pathway	0.009066
Signaling events mediated by PRL	0.01041
Signaling events mediated by VEGFR1 and VEGFR2	0.01099
Integrin-linked kinase signaling	0.01181
mTOR signaling pathway	0.01231
Validated transcriptional targets of TAp63 isoforms	0.01231
Alpha4 beta1 integrin signaling events	0.01566
Alpha9 beta1 integrin signaling events	0.0158
FOXM1 transcription factor network	0.01717
Angiopoietin receptor Tie2-mediated signaling	0.01781
Arf1 pathway	0.0183
SHP2 signaling	0.02009
CXCR4-mediated signaling events	0.02067
Class I PI3K signaling events mediated by Akt	0.0216
Signaling events mediated by focal adhesion kinase	0.0225
ATF-2 transcription factor network	0.0225
VEGFR1 specific signals	0.02285
Coregulation of Androgen receptor activity	0.0251
S1P1 pathway	0.02782

BARD1 signaling events	0.03176
Arf6 downstream pathway	0.03264
RhoA signaling pathway	0.03316
PDGFR-alpha signaling pathway	0.03357
Netrin-mediated signaling events	0.03696
E-cadherin signaling in the nascent adherens junction	0.04286
FoxO family signaling	0.04651
S1P2 pathway	0.04714
Plexin-D1 Signaling	0.04714
Internalization of ErbB1	0.0483

Supplementary Table 3. Top-ranked 50 pathways (NCI-Nature pathway database) of migratory SUM149 cells as compared to wild-type ones.

Pathway	P-value
Validated targets of C-MYC transcriptional activation	3.18E-08
Beta1 integrin cell surface interactions	0.000001408
HIF-1-alpha transcription factor network	0.000001408
Integrins in angiogenesis	0.000004806
Beta3 integrin cell surface interactions	0.00003627
Alpha9 beta1 integrin signaling events	0.00007507
Syndecan-4-mediated signaling events	0.0003957
Hypoxic and oxygen homeostasis regulation of HIF-1-alpha	0.0004783
Arf1 pathway	0.0006635
VEGFR1 specific signals	0.000887
p38 signaling mediated by MAPKAP kinases	0.001196
a6b1 and a6b4 Integrin signaling	0.001292
Integrin-linked kinase signaling	0.001292
PLK1 signaling events	0.001522
FOXM1 transcription factor network	0.002377
Plexin-D1 Signaling	0.002537
Alpha4 beta1 integrin signaling events	0.00258
Regulation of nuclear beta catenin signaling and target gene transcription	0.002708
Syndecan-2-mediated signaling events	0.003106
Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling	0.003278
C-MYB transcription factor network	0.003711
HIF-2-alpha transcription factor network	0.003711
E2F transcription factor network	0.003827
Insulin-mediated glucose transport	0.004778
RhoA signaling pathway	0.005101
IL6-mediated signaling events	0.005855
FoxO family signaling	0.007613
Signaling events mediated by focal adhesion kinase	0.007694
Direct p53 effectors	0.008126
AP-1 transcription factor network	0.008285
mTOR signaling pathway	0.008285
a4b7 Integrin signaling	0.009829
PDGFR-beta signaling pathway	0.01053
Validated nuclear estrogen receptor beta network	0.01063
ErbB1 downstream signaling	0.01074
Stabilization and expansion of the E-cadherin adherens junction	0.01074
VEGFR3 signaling in lymphatic endothelium	0.01549
Beta5 beta6 beta7 and beta8 integrin cell surface interactions	0.01683
VEGF and VEGFR signaling network	0.01923
Syndecan-1-mediated signaling events	0.01971

Signaling events mediated by VEGFR1 and VEGFR2	0.02086
Angiopoietin receptor Tie2-mediated signaling	0.02444
IGF1 pathway	0.02468
CDC42 signaling events	0.02475
Alpha6 beta4 integrin-ligand interactions	0.02527
Arf6 trafficking events	0.02708
Lissencephaly gene (LIS1) in neuronal migration and development	0.02838
E-cadherin signaling in the nascent adherens junction	0.02849
Aurora B signaling	0.02849
IFN-gamma pathway	0.03186

Supplementary Table 4. Top-ranked 50 pathways (NCI-Nature pathway database) of migratory GUM36 cells as compared to wild-type ones.

Pathway	**P-value**
Integrins in angiogenesis	0.000001773
Validated targets of C-MYC transcriptional activation	0.000002157
AP-1 transcription factor network	0.000003687
Beta1 integrin cell surface interactions	0.000007659
Arf1 pathway	0.00009448
HIF-1-alpha transcription factor network	0.0001148
C-MYB transcription factor network	0.0001735
Beta3 integrin cell surface interactions	0.0001959
Alpha4 beta1 integrin signaling events	0.0002404
VEGFR1 specific signals	0.0002851
Syndecan-2-mediated signaling events	0.0003185
Hypoxic and oxygen homeostasis regulation of HIF-1-alpha	0.0004707
a6b1 and a6b4 Integrin signaling	0.001217
PDGFR-beta signaling pathway	0.001323
Syndecan-1-mediated signaling events	0.001477
Validated transcriptional targets of AP1 family members Fra1 and Fra2	0.00229
Signaling events mediated by focal adhesion kinase	0.003412
Coregulation of Androgen receptor activity	0.003955
Alpha9 beta1 integrin signaling events	0.004163
Validated transcriptional targets of TAp63 isoforms	0.004838
IL6-mediated signaling events	0.005015
IL3-mediated signaling events	0.005279
TGF-beta receptor signaling	0.005613
HIF-2-alpha transcription factor network	0.007085
Signaling events mediated by TCPTP	0.008382
PDGFR-alpha signaling pathway	0.009027
Signaling events mediated by HDAC Class I	0.009965
Calcium signaling in the CD4+ TCR pathway	0.009982
Trk receptor signaling mediated by PI3K and PLC-gamma	0.01013
Beta5 beta6 beta7 and beta8 integrin cell surface interactions	0.01209
Integrin-linked kinase signaling	0.01321
Validated transcriptional targets of deltaNp63 isoforms	0.01321
Role of Calcineurin-dependent NFAT signaling in lymphocytes	0.01387
Osteopontin-mediated events	0.01446
EPHA2 forward signaling	0.01557
ErbB1 downstream signaling	0.01753
Validated targets of C-MYC transcriptional repression	0.01777
Caspase Cascade in Apoptosis	0.01781
Plasma membrane estrogen receptor signaling	0.01897
E2F transcription factor network	0.01935

Angiopoietin receptor Tie2-mediated signaling	0.01983
Direct p53 effectors	0.02418
Glypican 1 network	0.02476
p38 signaling mediated by MAPKAP kinases	0.02983
mTOR signaling pathway	0.03295
Signaling events regulated by Ret tyrosine kinase	0.04114
BCR signaling pathway	0.04599
Validated nuclear estrogen receptor alpha network	0.04599
Syndecan-4-mediated signaling events	0.04599
EPHB forward signaling	0.04656

Supplementary Table 5. Top-ranked pathways (NCI-Nature pathway database) using overlapping up-regulated and down-regulated genes from four breast cancer cell lines: SUM149, SUM159, MDA-MB-231, and GUM36.

Pathway	**P-value**
HIF-1-alpha transcription factor network	0.0003202
Validated targets of C-MYC transcriptional activation	0.000735
Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling	0.006361
RhoA signaling pathway	0.00771
Aurora B signaling	0.005165
CDC42 signaling events	0.02528
Alpha4 beta1 integrin signaling events	0.03348
IL2 signaling events mediated by PI3K	0.03947
E-cadherin signaling in the nascent adherens junction	0.04803

Supplementary Table 6. The percentage of ALDH+ and CD44+ cells in migratory and wild-type cell populations.

	Wild-Type			**Migratory**		
Cell Type	**Number of ALDH+ Cells**	**Number of CD44+ Cells**	**Total Number of Cells**	**Number of ALDH+ Cells**	**Number of CD44+ Cells**	**Total Number of Cells**
SUM149	1 (0.8%)	18 (13.7%)	131	15 (3.3%)	170 (37.9%)	448
SUM159	0 (0%)	49 (24.3%)	202	1 (0.5%)	72 (33.3%)	216
MDA-MB-231	6 (2.2%)	6 (2.2%)	272	21 (5.3%)	17 (4.3%)	394
GUM36	3 (0.9%)	306 (89.5%)	342	5 (2.0%)	196 (79.4%)	247

Supplementary Table 7. The overlapping up- and down-regulated genes in migratory cells as compared to wild-type cells from all four cell lines.

Up-regulated Genes			Down-regulated Genes		
ACBD3	GOLGB1	MYL12A	ACOT7	IDH3B	PSMC3
ANLN	GOLIM4	NFE2L2	ALG3	MEA1	RP11-778D9.4
ARPC5	HNRNPH3	NUFIP2	APRT	MRPL23	RP11-98L4.1
ASPH	ITGB1	SCAF11	AUP1	MRPL41	RPL18AP3
CALU	MGST3	SNAPC1	CAPNS1	NHP2	RPL19
CENPF	MIA3	SON	CCT7	NME1	RPL23
EEF1A1	MT-ND5	UBXN4	CD151	NPM1	RPS3
			CD320	OTUB1	SNRPB
			CD81	PKM	TIMM10
			CFL1	PPP1CA	TIMM13
			EIF6	PPP1R14B	TUBB4B
			GYG1		

Supplementary Table 8. The overlapping up-regulated genes in migratory cells as compared to wild-type cells from three cell lines.

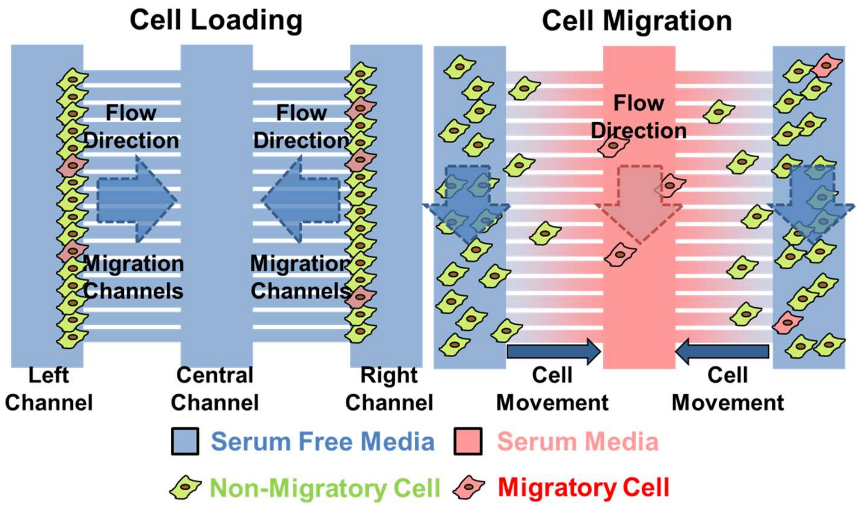
AC004074.3	CLINT1	HDAC2	MATR3	PTPN11	TAF13
ACBD3	CMPK1	HES1	MGST3	PTPN14	TMF1
ACTN4	COL12A1	HIF1A	MIA3	PTTG1IP	TNC
AHNAK	COL5A1	HIPK2	MORF4L2	RANBP2	TOMM20
AKAP9	COX6C	HNRNPA3	MT-CO1	RP11-270C12.3	TPM3
ANLN	CTSL	HNRNPH1	MT-ND5	RP11-302F12.1	TRAM1
ARHGAP11A	CYR61	HNRNPH3	MT-RNR2	RP11-365F18.1	TRIP12
ARID4B	DST	HNRNPU	MYL12A	RP11-75L1.2	TTC3
ARPC5	EEF1A1	HP1BP3	NCL	RPL14P1	TTC37
ASH1L	EFCAB14	HSP90B1	NFE2L2	RPS23P8	TXNP6
ASPH	EIF1AXP1	IGFBP7	NPM1P27	SCAF11	TXNRD1
ASPM	EIF3A	IL6ST	NPM1P39	SCP2	UBE2T
ATP1A1	EIF4G3	ITGB1	NRD1	SDC2	UBXN4
ATP1B1	ENAH	KDELR2	NUFIP2	SEC23A	UHMK1
BLOC1S6	ERLEC1	KHDRBS1	PABPC1	SEC31A	USO1
C12orf23	FAM114A1	KIF14	PABPC1P4	SEC61B	VCAN
CALU	FKBP7	LAMB1	PABPC3	SMARCA5	VIM
CANX	FN1	LIMA1	PALLD	SNAPC1	YBX1
CBX5	FRMD6	LINC00116	PDIA3	SON	YBX1P10
CCDC80	GOLGA2	LINC00657	PGRMC1	SPAG9	YWHAZ
CD46	GOLGB1	LPP	PHF3	SRGN	ZFP36L1
CENPF	GOLIM4	MACF1	PLOD2	SSR1	ZKSCAN1
CEP170	GOLT1B	MALAT1	PRSS23	STT3B	ZNHIT6
CEP350					

Supplementary Table 9. The overlapping down-regulated genes in migratory cells as compared to wild-type cells from three cell lines.

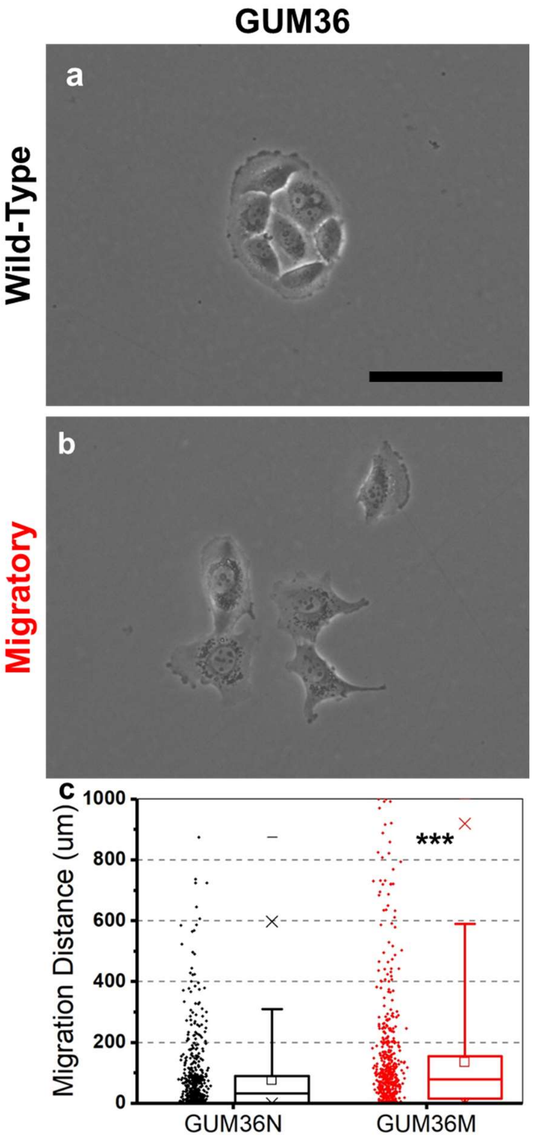
ACOT7	DCTN2	ITGB1BP1	NEDD8	PSMB3	SLC16A3
AIMP2	DRAP1	ITPA	NHP2	PSMC3	SNF8
AKR7A2	EDF1	LDHA	NHP2L1	PSMC5	SNRPB
ALDOA	EIF1	LSM4	NME1	PSMD7	SRM
ALG3	EIF2S2	MCM7	NOP56	PTGES2	STRA13
ANXA2	EIF3F	MEA1	NPM1	PTMS	SYNGR2
AP2M1	EIF3G	MIEN1	NT5C	RAB34	TAF10
AP2S1	EIF3L	MRPL11	ODC1	RAC1	TIMM10
APRT	EIF6	MRPL12	OTUB1	RBCK1	TIMM13
ARL2	EMP1	MRPL20	PCMT1	RNASEH2C	TIMP1
ATF4	ENO1	MRPL23	PFKL	RP11-778D9.4	TMEM147
ATP5G1	FAM192A	MRPL27	PHB	RP11-98L4.1	TMEM258
AUP1	FAM211A	MRPL28	PHLDA2	RPL13	TMEM261
C16orf13	FIBP	MRPL37	PIGT	RPL18A	TNNT1
CAPNS1	FKBP1A	MRPL40	PIN1	RPL18AP3	TRAPPC2L
CCDC124	FOSL1	MRPL41	PKM	RPL19	TUBA1C
CCT7	GADD45GIP1	MRPL52	PLEK2	RPL23	TUBB4B
CD151	GAPDH	MRPS15	POLD2	RPL26	TUFM
CD320	GARS	MT1X	POLE4	RPL29	TXNL4A
CD59	GINS2	MT2A	POLR2G	RPL35	UBA52
CD81	GNB2L1	MT-ATP6	POLR2L	RPL3P4	UBB
CENPM	GPS1	MTCH1	POMP	RPL8	UBE2M
CFL1	GYG1	MTCH2	PPIA	RPS15	UPP1
CLTB	H3F3B	MT-CO2	PPP1CA	RPS2	UQCRQ
CMSS1	HMGA1	MT-ND3	PPP1R14B	RPS3	VKORC1
COA3	HMGB1	MT-TV	PRDX1	RPSA	WDR54
COX4I1	HPCAL1	NARF	PRELID1	RRP7A	XRCC6
COX5A	HSBP1	NDUFS6	PRKCDBP	S100A16	YIF1A
DARS	IDH3B	NDUFS8	PSMB1	SGTA	YWHAQ

Supplementary Table 10. The common pathways (NCI-Nature pathway database) among MDA-MB-231 cells two, four, and six weeks after migration sorting.

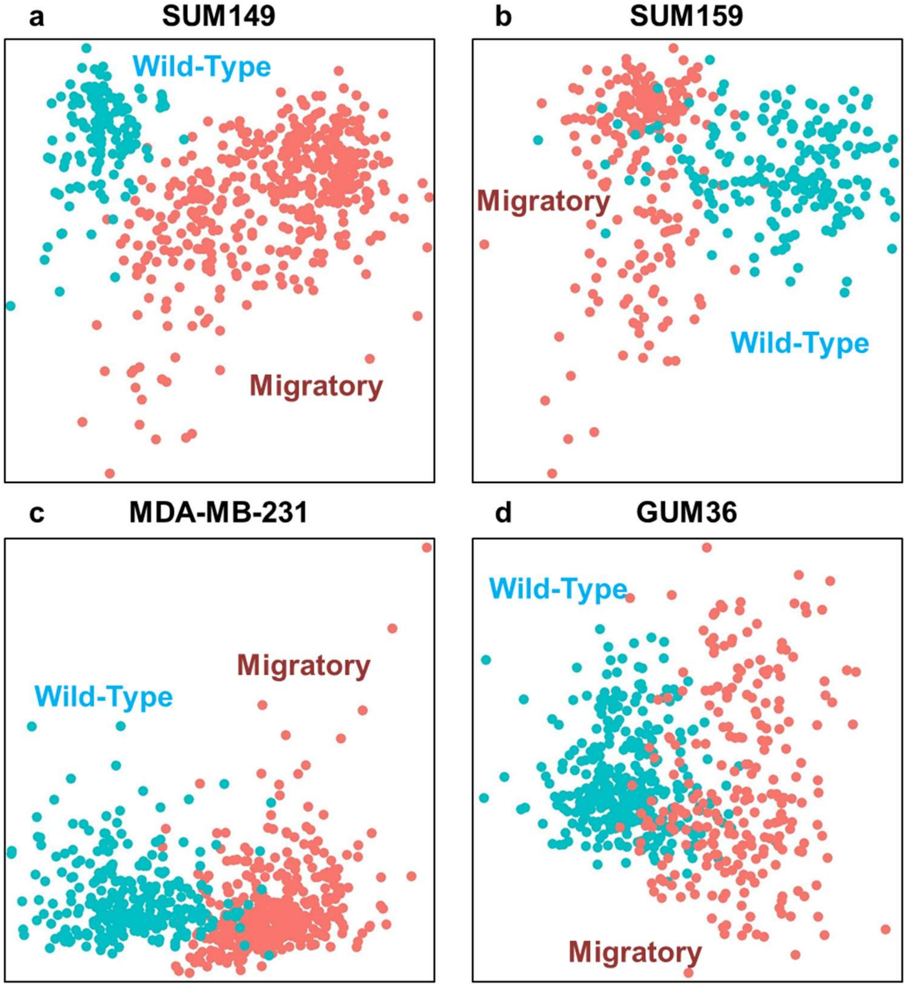
Pathway
a6b1 and a6b4 Integrin signaling
Arf6 downstream pathway
CDC42 signaling events
Class I PI3K signaling events mediated by Akt
ErbB1 downstream signaling
ErbB2/ErbB3 signaling events
FoxO family signaling
HIF-1-alpha transcription factor network
Integrin-linked kinase signaling
mTOR signaling pathway
p38 signaling mediated by MAPKAP kinases
PDGFR-beta signaling pathway
Posttranslational regulation of adherens junction stability and dissassembly
RAC1 signaling pathway
Regulation of nuclear beta catenin signaling and target gene transcription
Regulation of Telomerase
Role of Calcineurin-dependent NFAT signaling in lymphocytes
Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)
Signaling mediated by p38-alpha and p38-beta
TGF-beta receptor signaling
Trk receptor signaling mediated by PI3K and PLC-gamma
Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling
Validated nuclear estrogen receptor alpha network



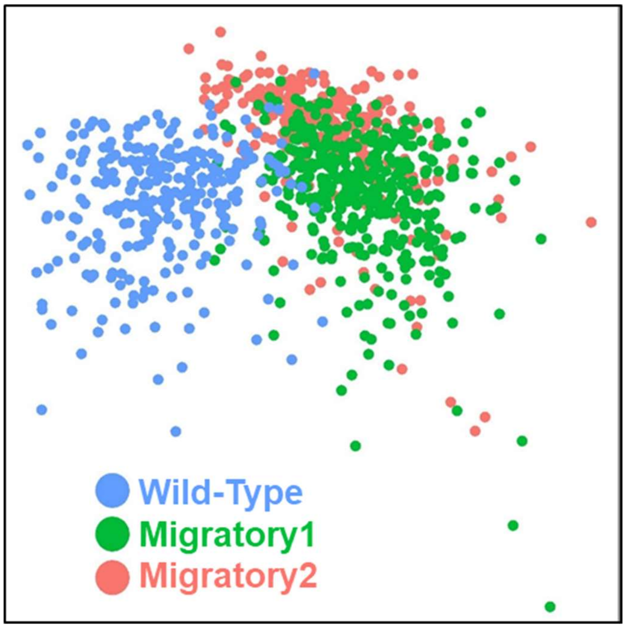
Supplementary Figure 1. Microfluidic high-throughput cell migration chip for migration-based cell selection. We initially load cells into the left/right channels. After cell loading, serum is introduced as a chemoattractant for cell migration into the central channel. Migratory (fast-moving) cells tend to move to the central channel from the left/right channels perfused with serum-free media. Finally, we can selectively retrieve migratory cells from the central channel by trypsinization for further analysis.



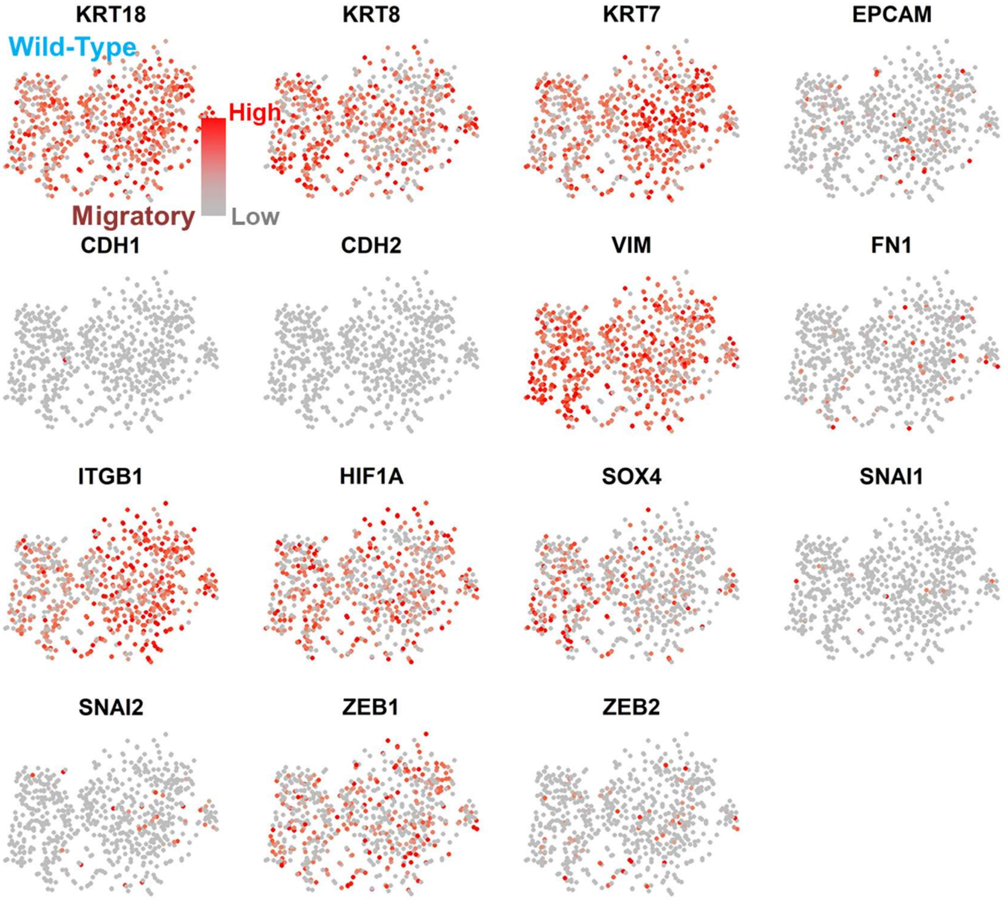
Supplementary Figure 2. (a, b) The different clonal morphologies of migratory and wild-type GUM36 cells. The clones (day 3) derived from wild-type cells are connected, while the ones from migratory cells are disseminated. (scale bar: 100 μm) (c) Microfluidically sorted migratory cell populations have distinct migratory behaviors as compared to wild-type ones. Graphs show positions of individual cells (after migration for 24 hours) and box plot and whiskers summaries for migration of GUM36 cells toward 10% fetal bovine serum (FBS) media. (n = 600 channels). *** refers to $P < 0.001$.



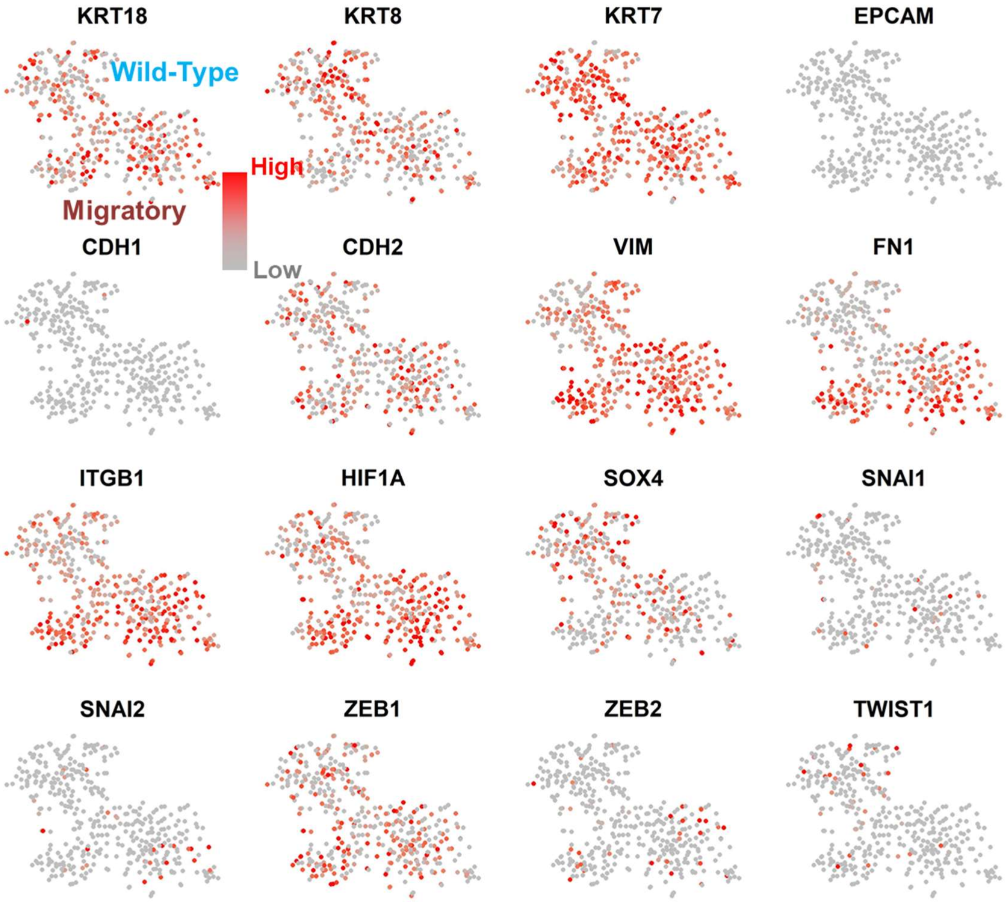
Supplementary Figure 3. PCA clustering demonstrates the separation of migratory and wild-type breast cancer cells. PCA plots of single-cell transcriptome analysis for migratory (red color) and wild-type (blue color) cells of four breast cancer cell lines: (a) SUM149, (b) SUM159, (c) MDA-MB-231, and (d) GUM36. X-axis represents PC1, and Y-axis represents PC2. Each dot represents a cell. Two populations of the same cell line are clearly separated, indicating distinct gene expression profiles.



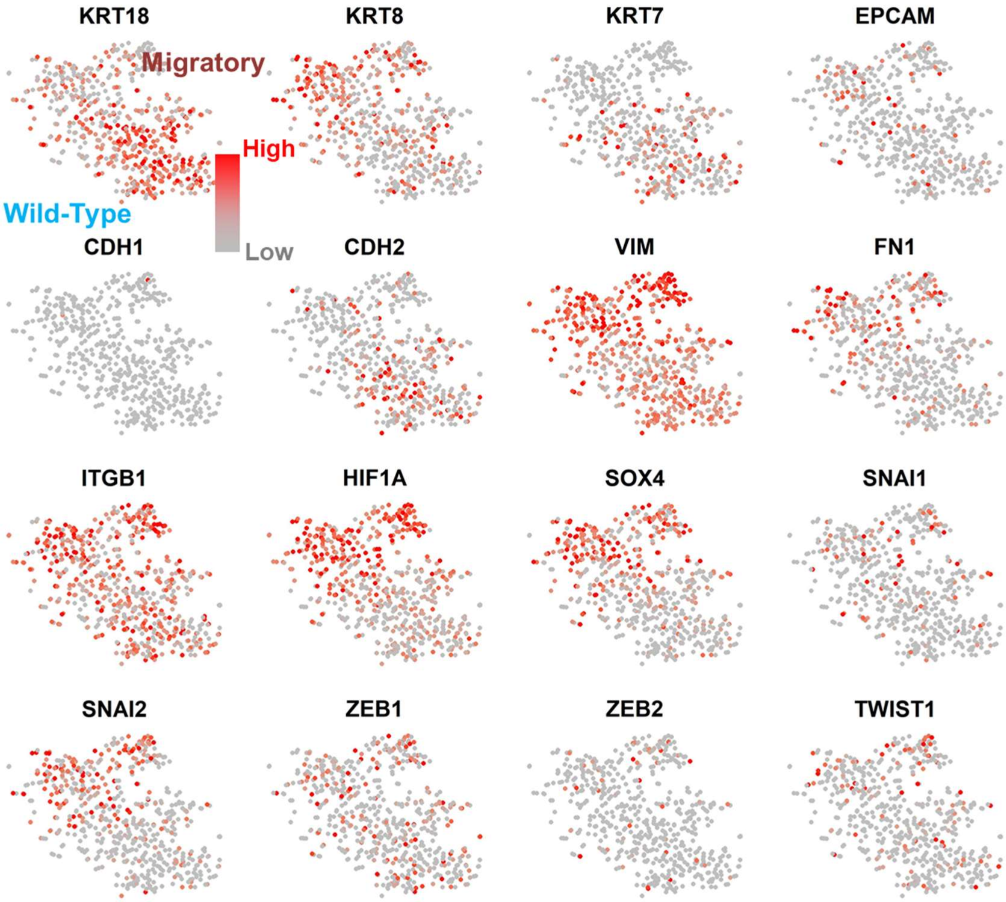
Supplementary Figure 4. Reproducibility of motility-based cell selection. Comparison of migratory MDA-MB-231 cells sorted in two independent experiments. PCA plot suggests that the gene expression profiles of migratory MDA-MB-231 cells sorted in two replicates are similar. X-axis represents PC1, and Y-axis represents PC2. Each dot represents a cell.



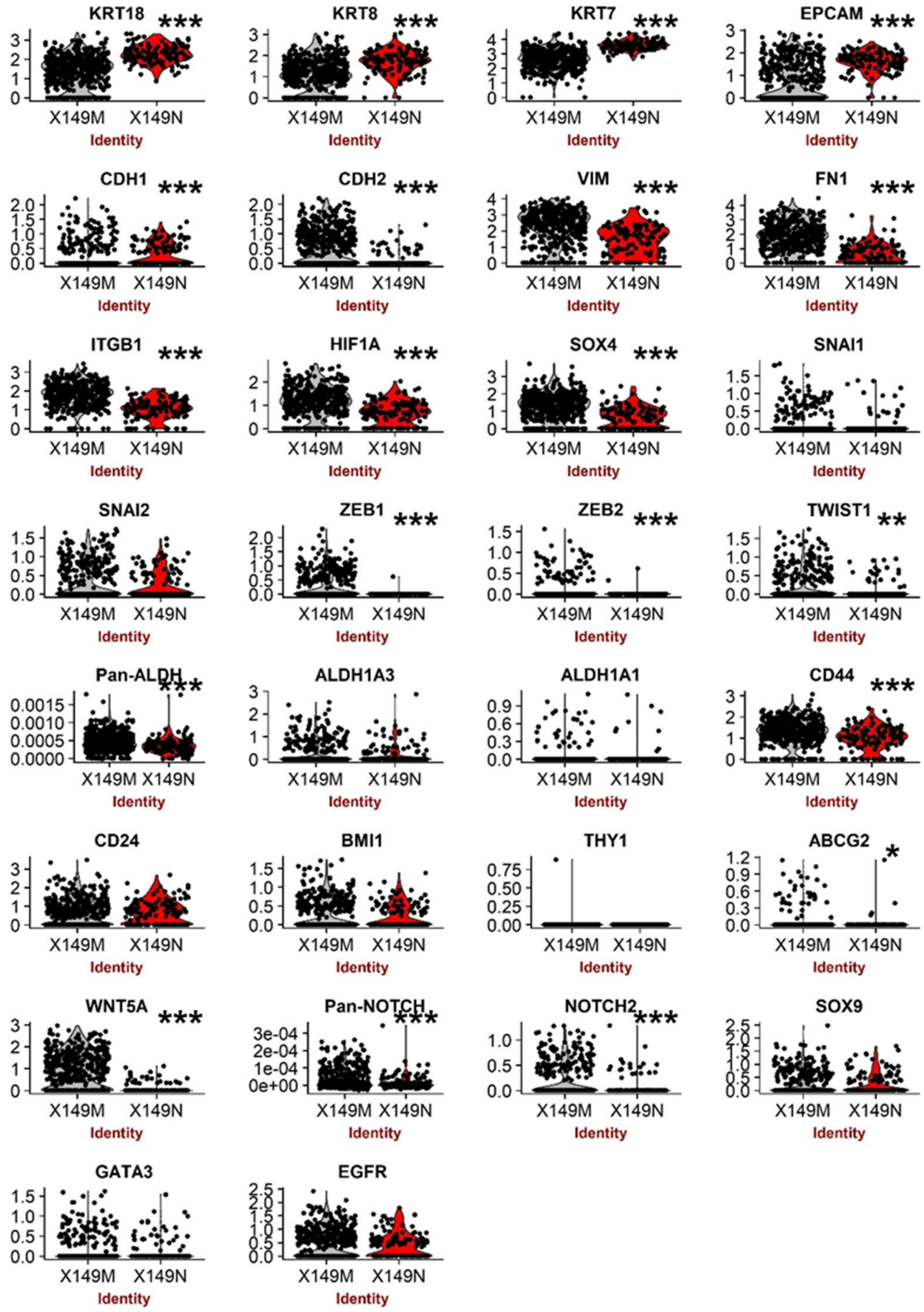
Supplementary Figure 5. EMT gene expression and clustering of migratory and wild-type MDA-MB-231 cells. X-axis represents tSNE1, and Y-axis represents tSNE2. Each dot represents one cell. Red color represents high (90th percentile) expression of a gene, and grey color represents low (10th percentile) expression of a gene. The expression is logarithmically normalized.



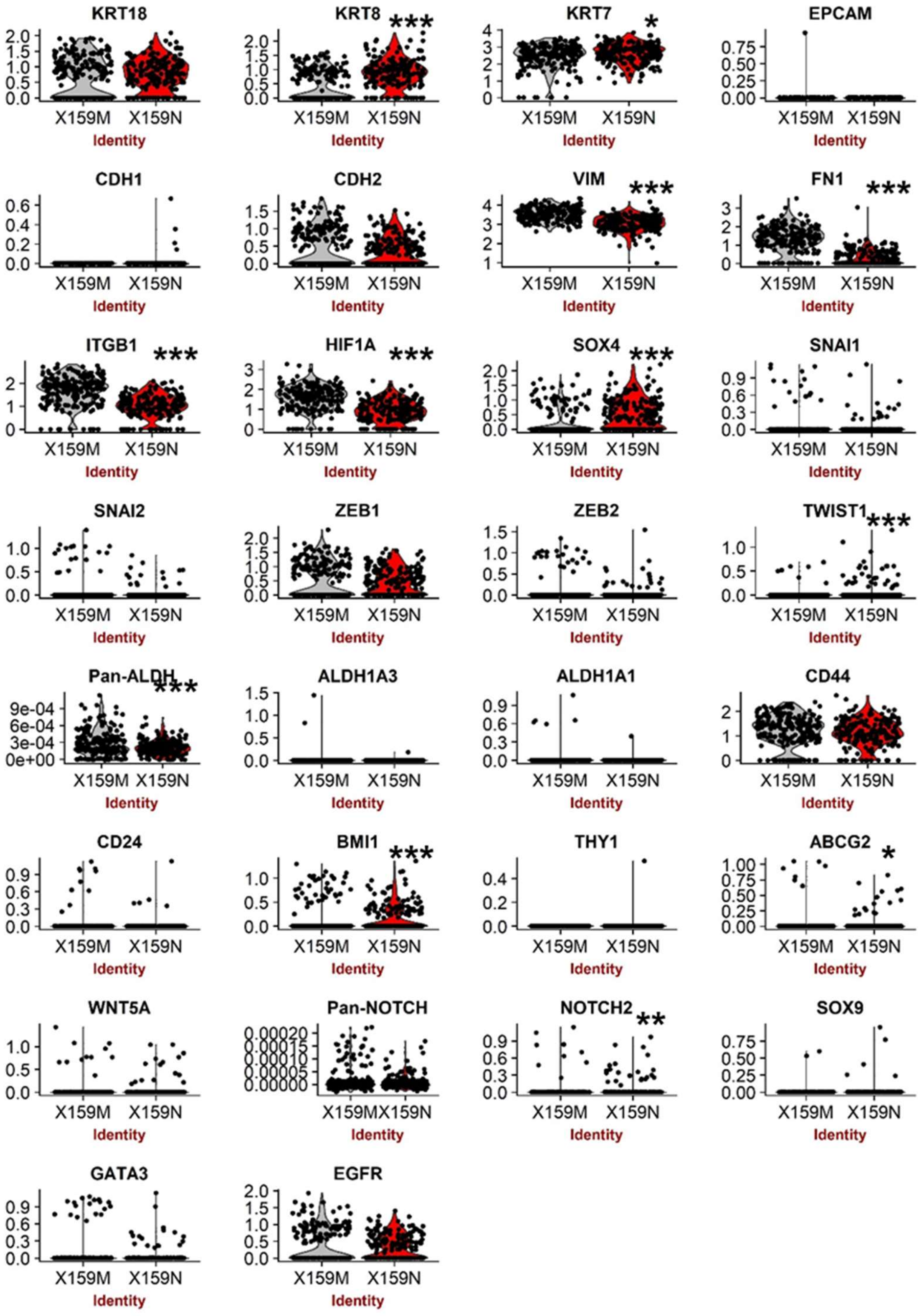
Supplementary Figure 6. EMT gene expression and clustering of migratory and wild-type SUM159 cells. X-axis represents tSNE1, and Y-axis represents tSNE2. Each dot represents one cell. Red color represents high (90th percentile) expression of a gene, and grey color represents low (10th percentile) expression of a gene. The expression is logarithmically normalized.



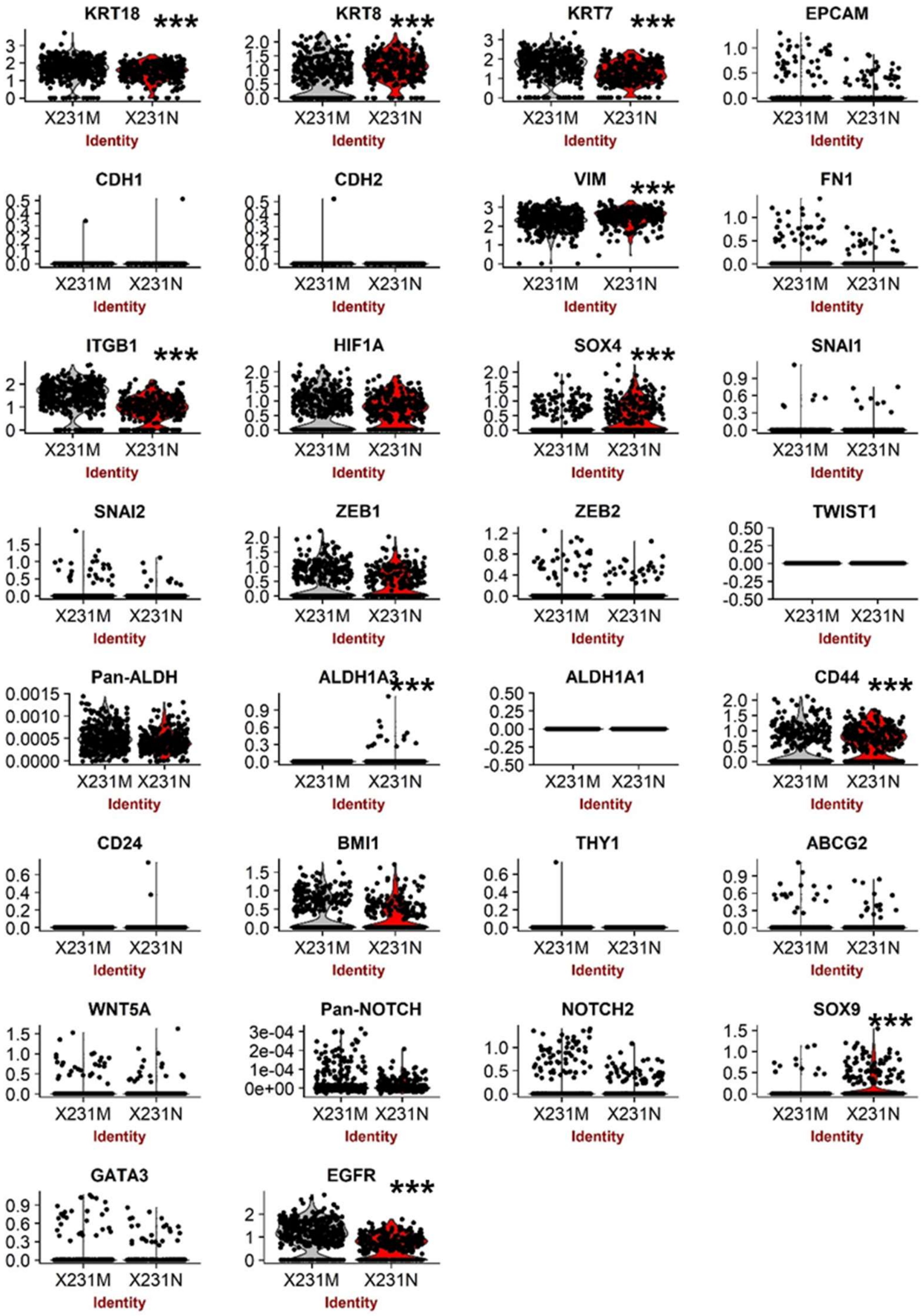
Supplementary Figure 7. EMT gene expression and clustering of migratory and wild-type GUM36 cells. X-axis represents tSNE1, and Y-axis represents tSNE2. Each dot represents one cell. Red color represents high (90th percentile) expression of a gene, and grey color represents low (10th percentile) expression of a gene. The expression is logarithmically normalized.



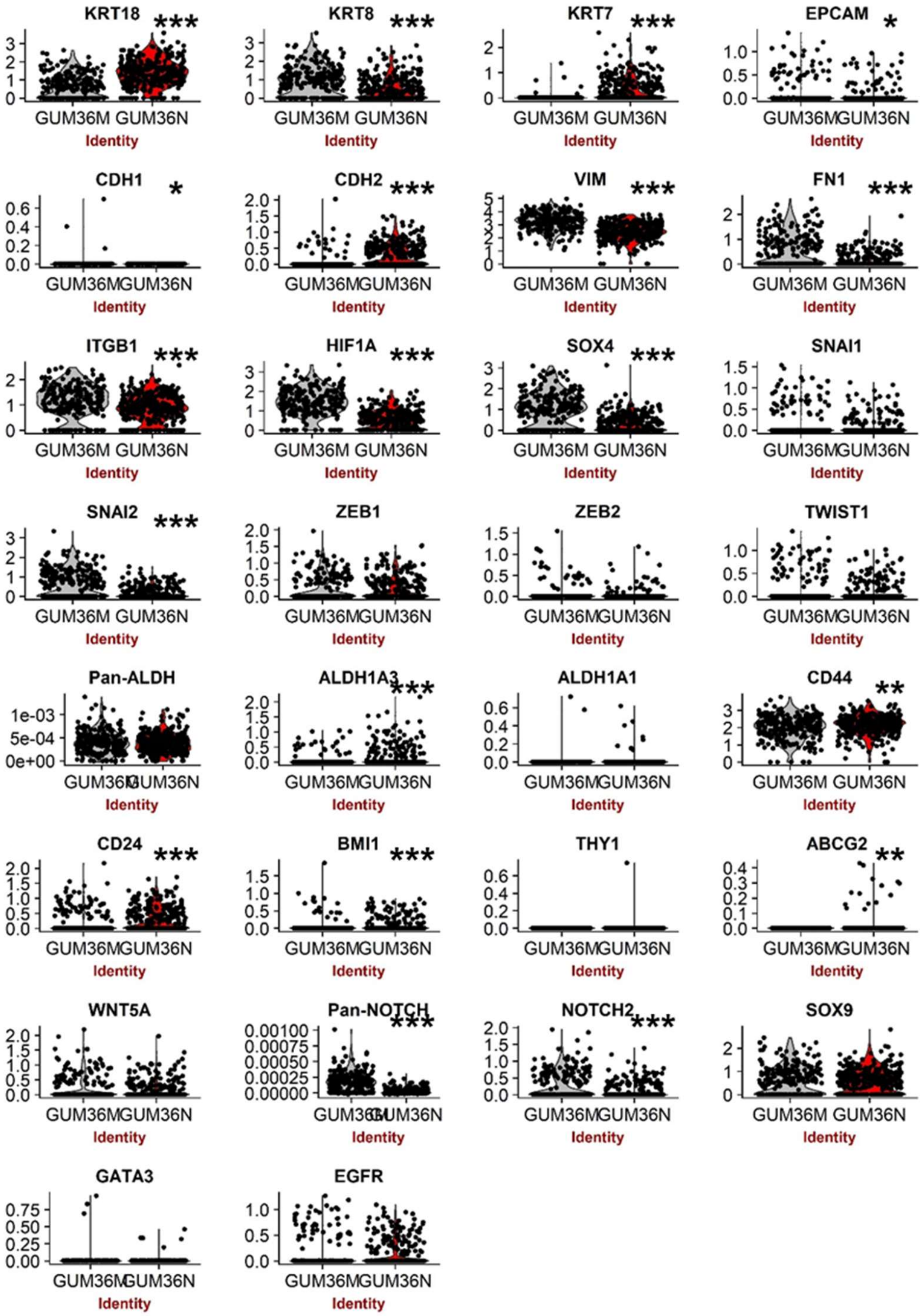
Supplementary Figure 8. Violin plots with statistical test of EMT/CSC related genes for migratory and wild-type SUM149 cells. Y-axis represents gene expression with logarithmic scale. Each dot represents one cell. Wilcoxon rank sum test was used for comparing gene expression. * refers to $P < 0.05$, ** refers to $P < 0.01$, and *** refers to $P < 0.001$.



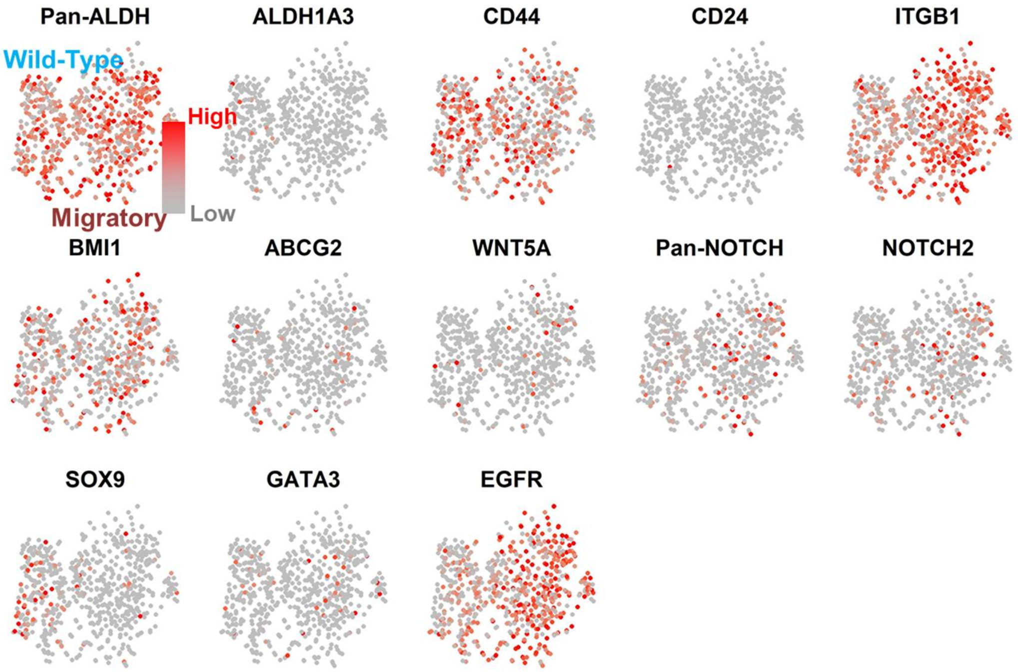
Supplementary Figure 9. Violin plots with statistical test of EMT/CSC related genes for migratory and wild-type SUM159 cells. Y-axis represents gene expression with logarithmic scale. Each dot represents one cell. Wilcoxon rank sum test was used for comparing gene expression. * refers to $P < 0.05$, ** refers to $P < 0.01$, and *** refers to $P < 0.001$.



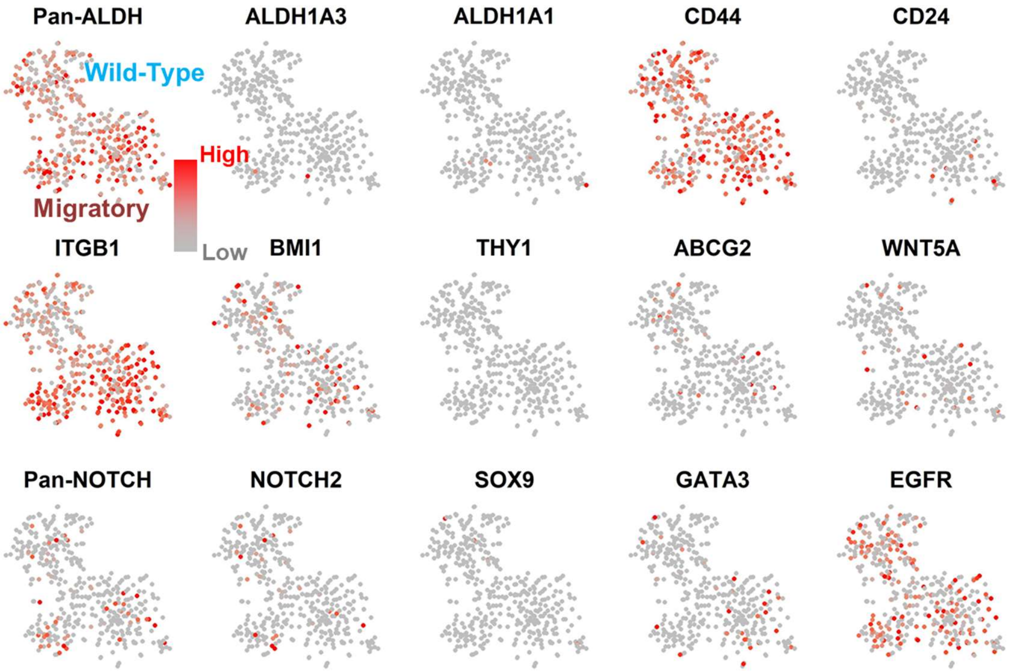
Supplementary Figure 10. Violin plots with statistical test of EMT/CSC related genes for migratory and wild-type MDA-MB-231 cells. Y-axis represents gene expression with logarithmic scale. Each dot represents one cell. Wilcoxon rank sum test was used for comparing gene expression. * refers to $P < 0.05$, ** refers to $P < 0.01$, and *** refers to $P < 0.001$.



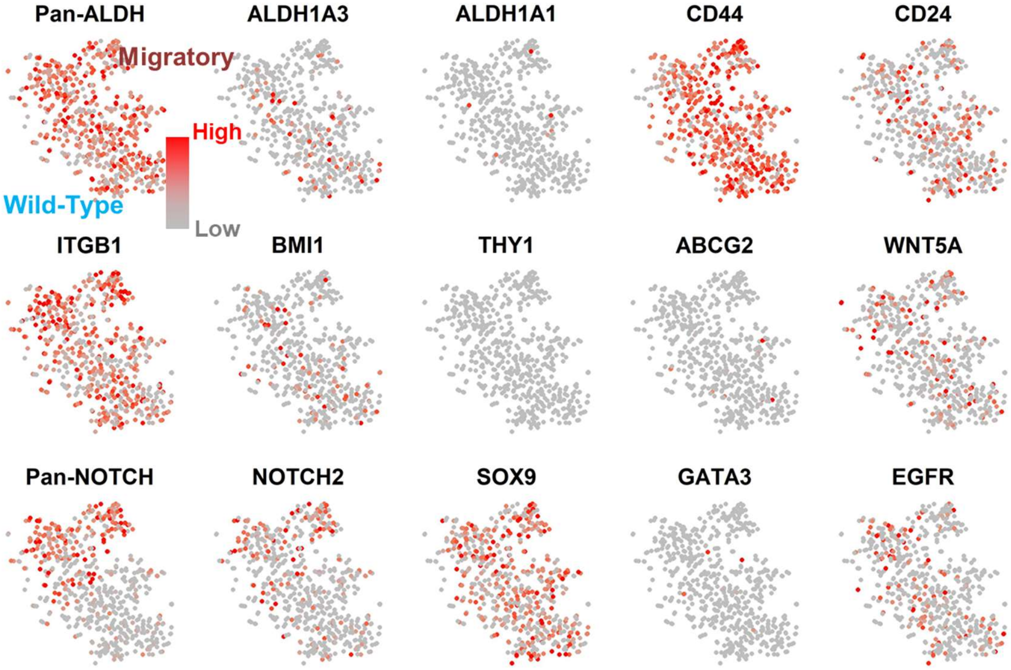
Supplementary Figure 11. Violin plots with statistical test of EMT/CSC related genes for migratory and wild-type GUM36 cells. Y-axis represents gene expression with logarithmic scale. Each dot represents one cell. Wilcoxon rank sum test was used for comparing gene expression. * refers to $P < 0.05$, ** refers to $P < 0.01$, and *** refers to $P < 0.001$.



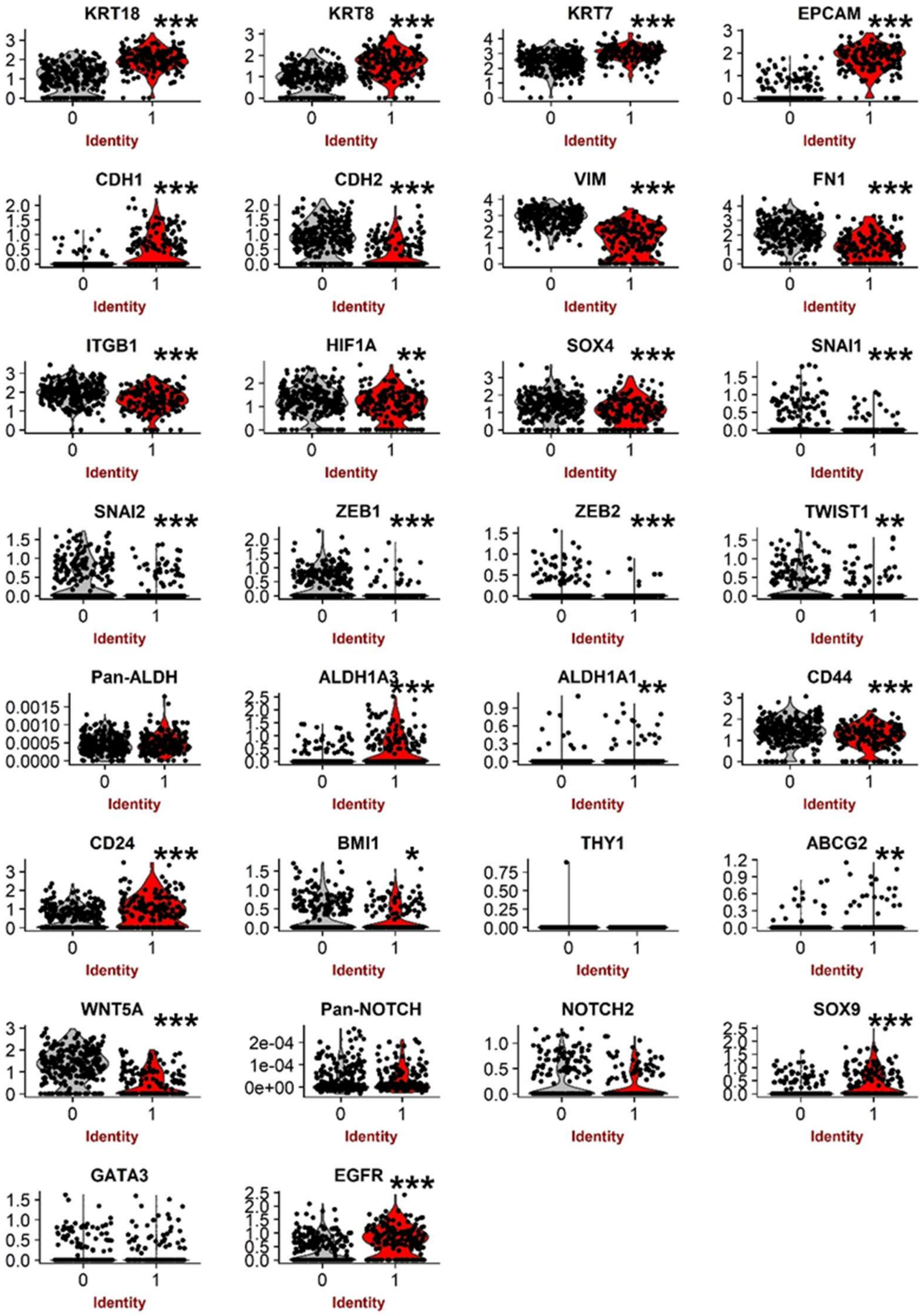
Supplementary Figure 12. CSC gene expression and clustering of migratory and wild-type MDA-MB-231 cells. X-axis represents tSNE1, and Y-axis represents tSNE2. Each dot represents one cell. Red color represents high (90th percentile) expression of a gene, and grey color represents low (10th percentile) expression of a gene. The expression is logarithmically normalized.



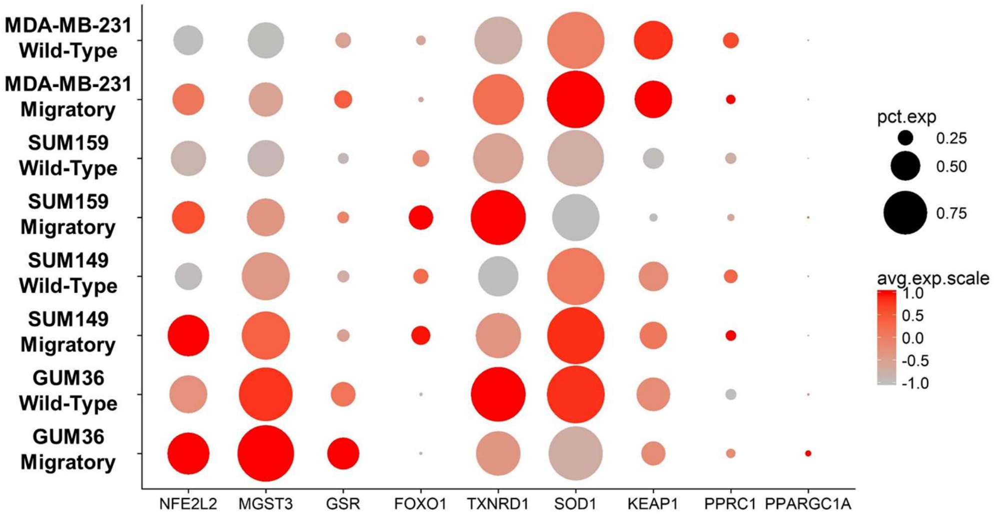
Supplementary Figure 13. CSC gene expression and clustering of migratory and wild-type SUM159 cells. X-axis represents tSNE1, and Y-axis represents tSNE2. Each dot represents one cell. Red color represents high (90th percentile) expression of a gene, and grey color represents low (10th percentile) expression of a gene. The expression is logarithmically normalized.



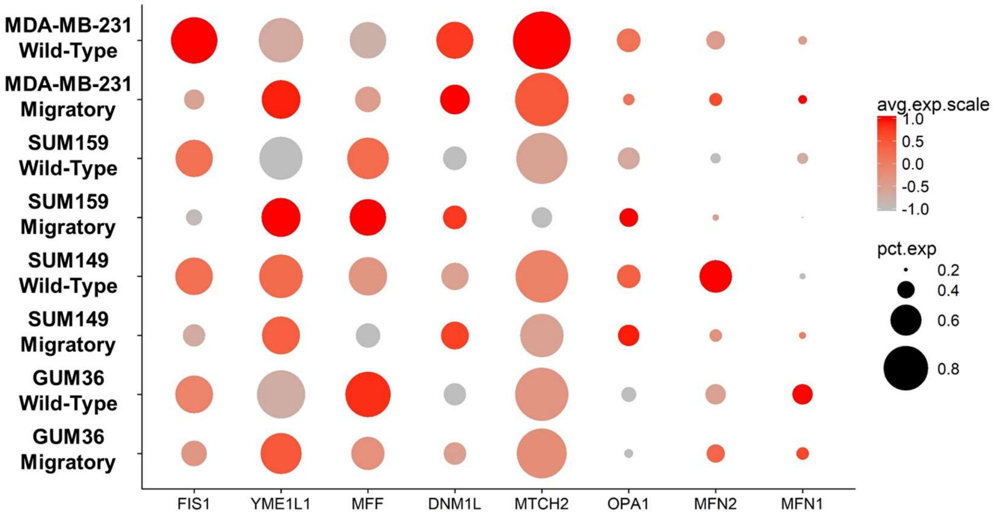
Supplementary Figure 14. CSC gene expression and clustering of migratory and wild-type GUM36 cells. X-axis represents tSNE1, and Y-axis represents tSNE2. Each dot represents one cell. Red color represents high (90th percentile) expression of a gene, and grey color represents low (10th percentile) expression of a gene. The expression is logarithmically normalized.



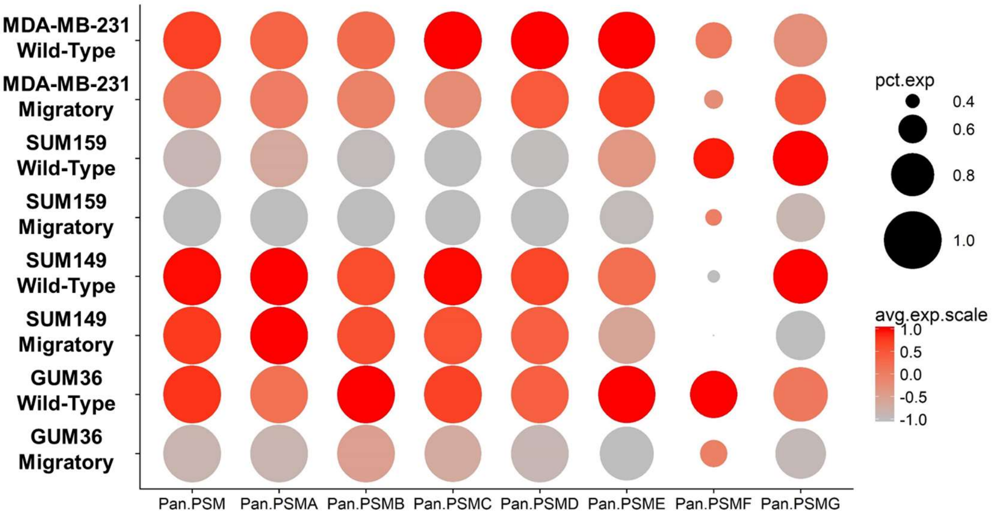
Supplementary Figure 15. Violin plots with statistical test of EMT/CSC related genes for two subpopulations of migratory SUM149 cells. Group 0 represents mesenchymal-like cells, and Group 1 represents epithelial-like cells. Y-axis represents gene expression with logarithmic scale. Each dot represents one cell. * refers to P < 0.05. ** refers to P < 0.01, and *** refers to P < 0.001.



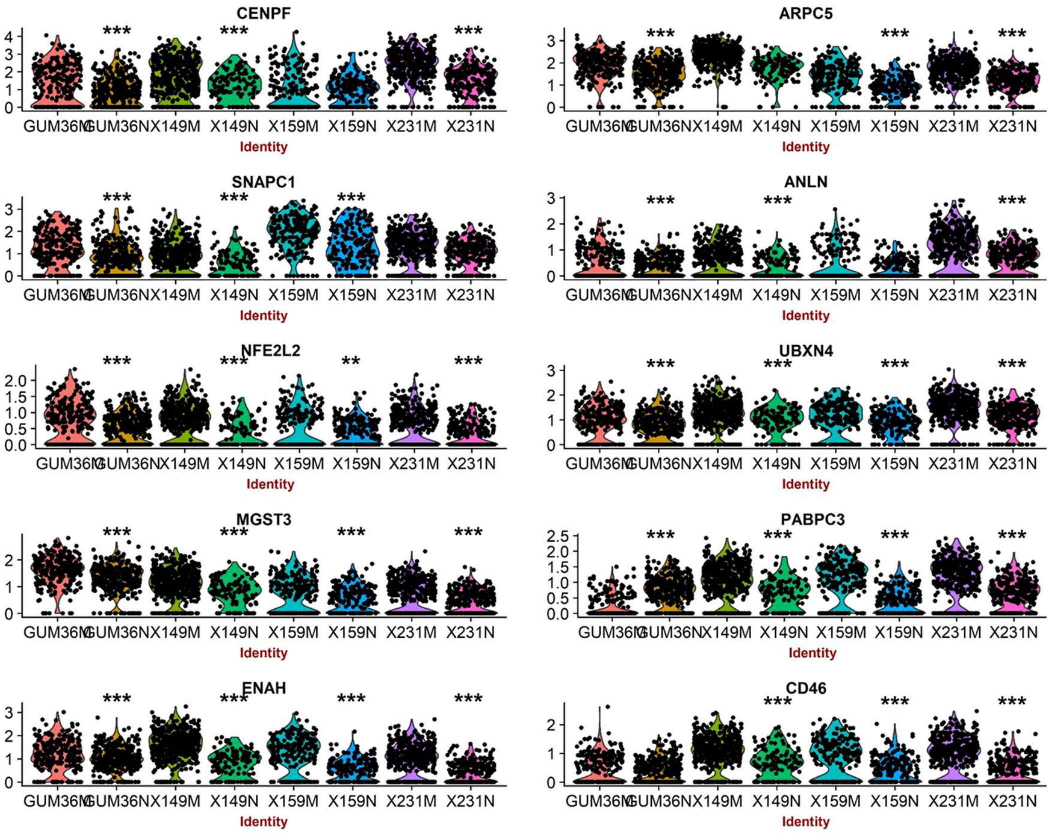
Supplementary Figure 16. Expression of reactive oxygen species (ROS) controlling genes. Dot plot shows the expression of 9 genes among 8 cell populations (migratory and wild-type of 4 cell lines). Larger dot means higher percentage of single cells expressing that gene, and smaller dot means lower percentage of single cells expressing that gene. Grey color represents the lowest expression, and red color represents the highest expression. The expression is logarithmically normalized.



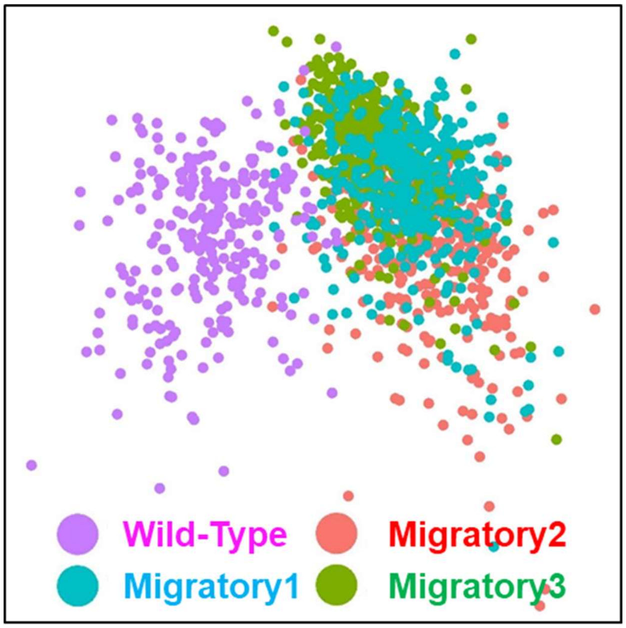
Supplementary Figure 17. Expression of mitochondrial morphology controlling genes. Dot plot shows the expression of 8 genes among 8 cell populations (migratory and wild-type of 4 cell lines). Larger dot means higher percentage of single cells expressing that gene, and smaller dot means lower percentage of single cells expressing that gene. Grey color represents the lowest expression, and red color represents the highest expression. The expression is logarithmically normalized.



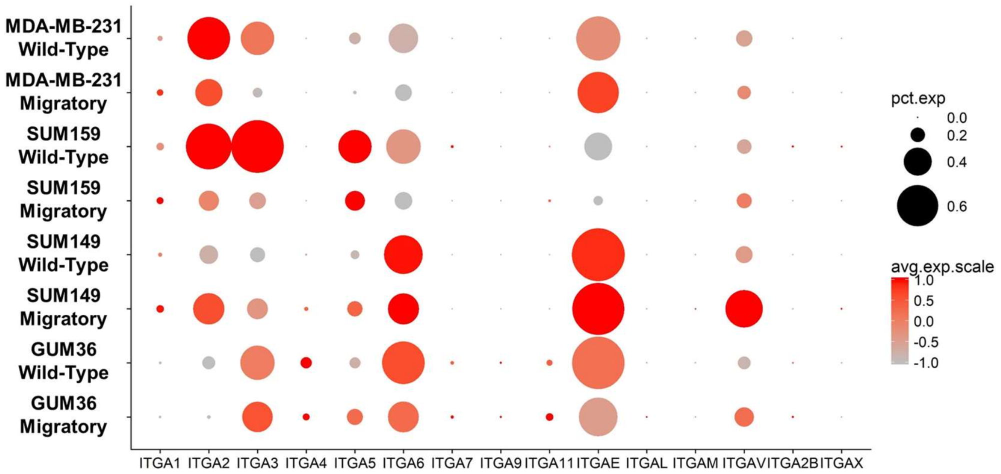
Supplementary Figure 18. Expression of proteasome genes. As there are many proteasome genes, we pooled proteasome genes of the same subunit together (as Pan.PSMA, Pan.PSMB, etc.) and all proteasome genes together (as Pan.PSM). Dot plot shows the expression of 8 Pan-PSM genes among 8 cell populations (migratory and wild-type of 4 cell lines). Larger dot means higher percentage of single cells expressing that gene, and smaller dot means lower percentage of single cells expressing that gene. Grey color represents the lowest expression, and red color represents the highest expression. The expression is logarithmically normalized.



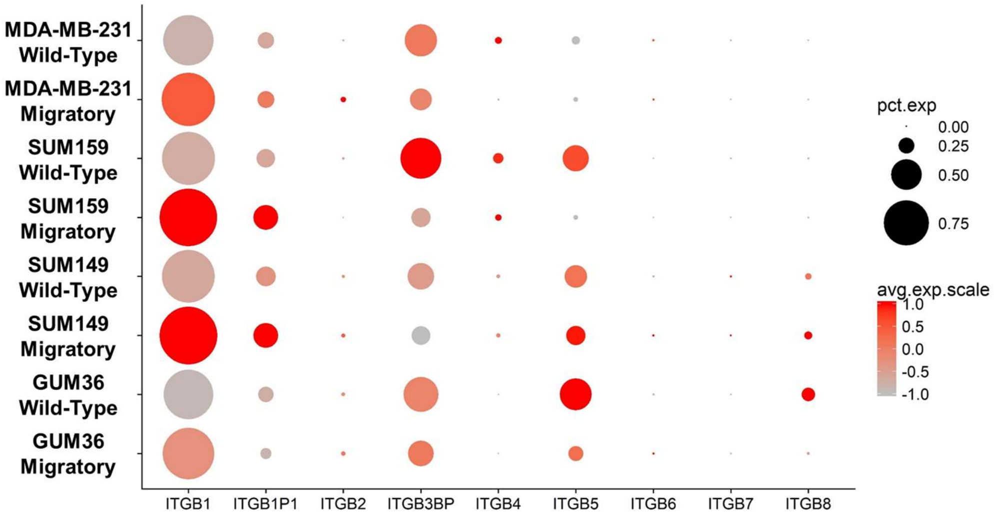
Supplementary Figure 19. Violin plots and statistical test of novel genes associated with cell migration. Y-axis represents gene expression with logarithmic scale. Each dot represents a cell. * refers to $P < 0.05$. ** refers to $P < 0.01$, and *** refers to $P < 0.001$.



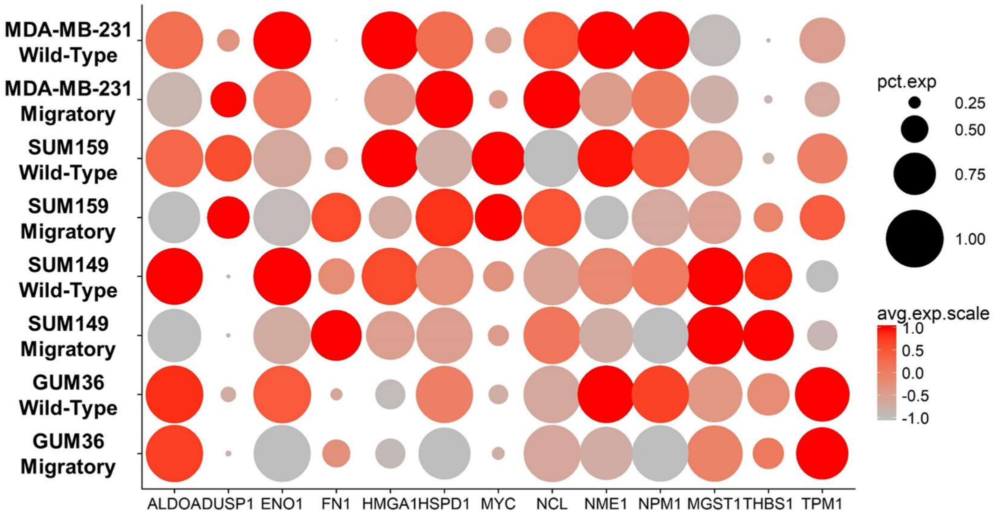
Supplementary Figure 20. The distinct gene expression profile of migratory MDA-MB-231 cells maintains with time. The migratory MDA-MB-231 cells were sequenced three times (two weeks (Migratory1), four weeks (Migratory2), and six weeks (Migratory3)) after microfluidic motility-based sorting. PCA plot suggests that the gene expression profiles of migratory MDA-MB-231 cells at different time points are similar. X-axis represents PC1, and Y-axis represents PC2. Each dot represents a cell.



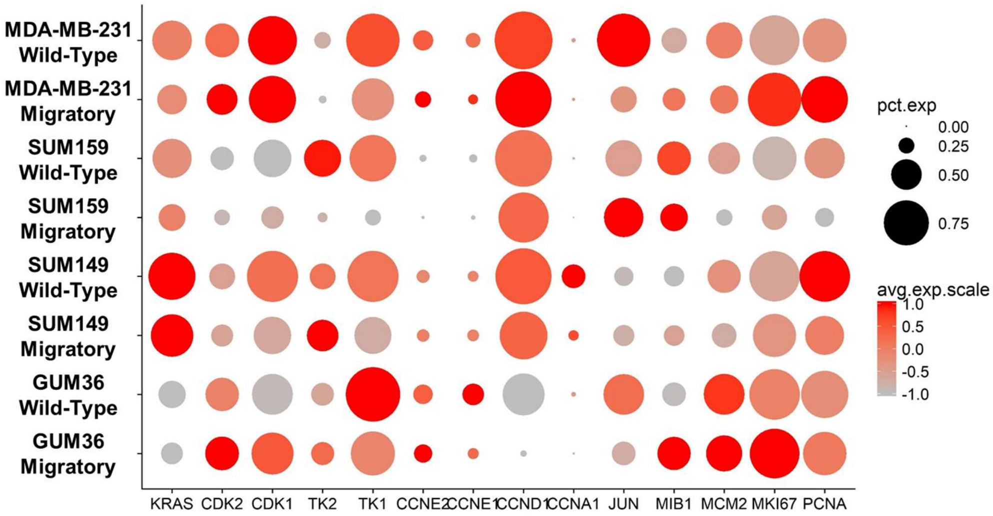
Supplementary Figure 21. Expression of integrin subunit alpha genes. Dot plot shows the expression of 15 genes among 8 cell populations (migratory and wild-type of 4 cell lines). Larger dot means higher percentage of single cells expressing that gene, and smaller dot means lower percentage of single cells expressing that gene. Grey color represents the lowest expression, and red color represents the highest expression. The expression is logarithmically normalized.



Supplementary Figure 22. Expression of integrin subunit beta genes. Dot plot shows the expression of 9 genes among 8 cell populations (migratory and wild-type of 4 cell lines). Larger dot means higher percentage of single cells expressing that gene, and smaller dot means lower percentage of single cells expressing that gene. Grey color represents the lowest expression, and red color represents the highest expression. The expression is logarithmically normalized.



Supplementary Figure 23. Expression of cMyc pathway related genes. Dot plot shows the expression of 13 genes among 8 cell populations (migratory and wild-type of 4 cell lines). Larger dot means higher percentage of single cells expressing that gene, and smaller dot means lower percentage of single cells expressing that gene. Grey color represents the lowest expression, and red color represents the highest expression. The expression is logarithmically normalized.



Supplementary Figure 24. Expression of cell proliferation related genes. Dot plot shows the expression of 14 genes among 8 cell populations (migratory and wild-type of 4 cell lines). Larger dot means higher percentage of single cells expressing that gene, and smaller dot means lower percentage of single cells expressing that gene. Grey color represents the lowest expression, and red color represents the highest expression. The expression is logarithmically normalized.