



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2022

---

## **Mass cytometric and transcriptomic profiling of epithelial-mesenchymal transitions in human mammary cell lines**

Wagner, Johanna ; Masek, Markus ; Jacobs, Andrea ; Soneson, Charlotte ; Sivapatham, Sujana ; Damond, Nicolas ; de Souza, Natalie ; Robinson, Mark D ; Bodenmiller, Bernd

DOI: <https://doi.org/10.1038/s41597-022-01137-4>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-219311>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Wagner, Johanna; Masek, Markus; Jacobs, Andrea; Soneson, Charlotte; Sivapatham, Sujana; Damond, Nicolas; de Souza, Natalie; Robinson, Mark D; Bodenmiller, Bernd (2022). Mass cytometric and transcriptomic profiling of epithelial-mesenchymal transitions in human mammary cell lines. *Scientific Data*, 9:44-60.

DOI: <https://doi.org/10.1038/s41597-022-01137-4>



OPEN

DATA DESCRIPTOR

# Mass cytometric and transcriptomic profiling of epithelial-mesenchymal transitions in human mammary cell lines

Johanna Wagner<sup>1,6</sup>, Markus Masek<sup>2</sup>, Andrea Jacobs<sup>1,3</sup>, Charlotte Soneson<sup>2,4</sup>, Sujana Sivapatham<sup>1,3</sup>, Nicolas Damond<sup>1,3</sup>, Natalie de Souza<sup>1,3,5</sup>, Mark D. Robinson<sup>2,4</sup> & Bernd Bodenmiller<sup>1,3</sup>✉

Epithelial-mesenchymal transition (EMT) equips breast cancer cells for metastasis and treatment resistance. However, detection, inhibition, and elimination of EMT-undergoing cells is challenging due to the intrinsic heterogeneity of cancer cells and the phenotypic diversity of EMT programs. We comprehensively profiled EMT transition phenotypes in four non-cancerous human mammary epithelial cell lines using a flow cytometry surface marker screen, RNA sequencing, and mass cytometry. EMT was induced in the HMLE and MCF10A cell lines and in the HMLE-Twist-ER and HMLE-Snail-ER cell lines by prolonged exposure to TGF $\beta$ 1 or 4-hydroxytamoxifen, respectively. Each cell line exhibited a spectrum of EMT transition phenotypes, which we compared to the steady-state phenotypes of fifteen luminal, HER2-positive, and basal breast cancer cell lines. Our data provide multiparametric insights at the single-cell level into the phenotypic diversity of EMT at different time points and in four human cellular models. These insights are valuable to better understand the complexity of EMT, to compare EMT transitions between the cellular models used here, and for the design of EMT time course experiments.

## Background & Summary

The epithelial-mesenchymal transition (EMT) equips epithelial cells with migratory, survival, and plasticity properties upon loss of epithelial hallmark characteristics. Together with its reverse process, the mesenchymal-epithelial transition, EMT contributes to cancer metastasis, provides resistance to cell death and chemotherapy, confers stemness properties to cancer cells, and interferes with immunotherapy<sup>1–3</sup>. EMT inhibition and elimination of EMT-undergoing cells are therefore investigated as approaches for cancer therapy<sup>4</sup>. However, detecting cancer cells undergoing EMT is challenging due to the intrinsic heterogeneity of cancer cells and the phenotypic diversity of EMT programs<sup>4</sup>.

A hallmark characteristic of epithelial cells is adhesion to neighboring cells and to the basement membrane<sup>1</sup>. To prevent anchorage-independent growth, epithelial cells normally undergo anoikis upon neighbor or matrix detachment<sup>5</sup>. During EMT, normal adhesion complexes, e.g., involving E-Cadherin, epithelial cell adhesion molecule (EpcAM), and laminin receptor integrin  $\alpha$ 6 $\beta$ 1 (CD49f/CD29), are dissolved and resistance to anoikis is established<sup>6,7</sup>. Concomitant cytoskeletal rearrangements break down the epithelial apico-basal orientation and induce a motile front-back polarity, which often includes a replacement of cytokeratins with Vimentin<sup>8</sup>. EMT can further confer stemness properties to epithelial cells<sup>9,10</sup>. Numerous signaling pathways can trigger EMT, including TGF $\beta$ 1, Notch, Hedgehog, WNT, and hypoxia, and activate downstream transcriptional drivers

<sup>1</sup>Department of Quantitative Biomedicine, University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland.

<sup>2</sup>Department of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland.

<sup>3</sup>Institute for Molecular Health Sciences, ETH Zurich, Otto-Stern-Weg 7, 8093, Zurich, Switzerland. <sup>4</sup>SIB Swiss Institute of Bioinformatics, University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland. <sup>5</sup>Institute for Molecular Systems Biology, Department of Biology, ETH-Zurich, Otto-Stern-Weg 3, 8093, Zurich, Switzerland.

<sup>6</sup>Present address: Division of Translational Medical Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, 69120, Heidelberg, Germany. ✉e-mail: [bernd.bodenmiller@dqbm.uzh.ch](mailto:bernd.bodenmiller@dqbm.uzh.ch)

such as Snail family zinc finger transcription factors (TF), Twist family BHLH TFs, zinc finger E-box binding homeobox TFs, and homeobox TF PRRX1<sup>11</sup>. Regulation of EMT occurs by integration of epigenetic, transcriptional, post-transcriptional, and protein stability controls<sup>11,12</sup>. Together, this shows that the phenotypes of EMT-undergoing cells are shaped by complex molecular circuitries.

EMT is increasingly viewed more as a phenotypic continuum with intermediate states and less as a shift between two discrete states, and the concepts of 'partial EMT' and 'hybrid EMT' phenotypes have been introduced<sup>4,13</sup>. A systems biology approach used gene expression profiles of four non-small cell lung cancer cell lines to detect three intermediate states termed 'pre-EMT', 'metastable EMT', and 'epigenetically-fixed'<sup>14</sup>. Transcriptomics of cell lines and clinical samples of cancer was used to rank the resulting spectrum of EMT states, showing that only some were linked to poor survival<sup>15</sup>. However, identification of EMT-undergoing cells in metastatic cancer tissue is still often based on co-expression of a few epithelial and mesenchymal markers<sup>16,17</sup>. This can be misleading as several of the 'mesenchymal' markers, e.g., Vimentin, can also be expressed by non-malignant epithelial cells<sup>18</sup>. It remains an ongoing debate which markers and combination of markers are sufficient to distinguish EMT from other processes *in vitro* and *in vivo*<sup>4,19</sup>. In particular, there remains the need for a comprehensive analysis of EMT phenotypes at the protein level.

To address this need, we applied multiplex single-cell mass cytometry<sup>20</sup> to four non-cancerous human mammary epithelial cell lines that serve as widely-used models of EMT. EMT was induced in the HMLE and MCF10A cell lines by prolonged exposure to TGF $\beta$ 1<sup>9,21</sup> and in the HMLE-Twist-ER (HTER) and HMLE-Snail-ER (HSER) cell lines by treatment with 4-hydroxytamoxifen (4OHT)<sup>9</sup>. In the HTER and HSER cell lines, 4OHT treatment allows the induction of gene expression by murine Twist1 fused to a modified estrogen receptor (ER) or SNAIL1-ER fusion protein, respectively<sup>9</sup>. To design our mass cytometry antibody panel, we conducted a flow cytometry surface protein screen in parallel with a transcriptome analysis at multiple time points of induced EMT. We observed alterations in the surface proteome of EMT-undergoing cells over time and detected distinct gene expression profiles of hybrid epithelial-mesenchymal states compared with epithelial and mesenchymal states. From these analyses, we extracted candidate markers for multiplex mass cytometry, which revealed complex phenotypic transitions in all four EMT models and little phenotypic overlap of EMT states between the cell lines. The data presented here can aid in characterizing the complexity and dynamics of EMT in these widely used *in vitro* models.

## Methods

**Material.** A table listing the material used in this study can be found on Mendeley Data (Mendeley Table 1)<sup>22</sup>.

**Cell lines.** All human breast cancer cell lines were obtained from the American Type Culture Collection (ATCC) and were grown according to ATCC recommendations. The MCF10A human mammary epithelial cell line was obtained from ATCC (CRL-10317) and cultured in DMEM F12 Ham medium (Sigma Aldrich) supplemented with 10  $\mu$ g/ml human insulin (Sigma Aldrich), 20 ng/ml epidermal growth factor (EGF, Peprotech), 500 ng/ml hydrocortisone (Sigma Aldrich), 5% horse serum (Gibco), 100 ng/ml cholera toxin (Sigma Aldrich), and PenStrep (Gibco)<sup>23</sup>. We validated the MCF10A cell line by short tandem repeat (STR) profiling using the ATCC kit (#135-XV). The HMLE, HMLE-Twist-ER (HTER), and HMLE-Snail-ER (HSER) cell lines were a gift from the laboratory of Dr. Robert A. Weinberg at the Massachusetts Institute of Technology and were cultured in a 1:1 mixture of DMEM F12 Ham medium (Sigma Aldrich) supplemented with 10  $\mu$ g/ml human insulin (Sigma Aldrich), 10 ng/ml EGF (Peprotech), 500 ng/ml hydrocortisone (Sigma Aldrich), and PenStrep (Gibco) with the mammary epithelial growth medium (MEGM<sup>TM</sup>) BulletKit<sup>TM</sup> (Lonza)<sup>10</sup>. For the HTER and HSER cell lines, the growth medium was supplemented with 1  $\mu$ g/ml Blastidin S (InvivoGen). All of the cell lines were authenticated upon receipt by comparing them to the originally reported morphological and growth characteristics. They were not tested for mycoplasma. For the HMLE, HTER, and HSER cell lines, growth and morphology as well as protein expression profiles of e.g., cytokeratins, E-Cadherin, Vimentin, CD24, CD44, matched previous reports. None of the cell lines used in this project are among misidentified cell lines listed by the International Cell Line Authentication Committee.

**EMT time courses and cell harvesting.** EMT was induced in the MCF10A cell line by prolonged stimulation with 5 ng/ml TGF $\beta$ 1 (Cell Signaling Technology) for eight days<sup>24</sup>. For this, 0.8 million cells were seeded per 10 cm cell culture dish (Nunc) and incubated at 37°C and 5% CO<sub>2</sub> according to ATCC recommendations. TGF $\beta$ 1 treatment and vehicle treatment using Dulbecco's phosphate buffer saline (PBS, Sigma Aldrich) started 24 hours after seeding and was applied daily together with a growth medium exchange.

EMT was induced in the HMLE cell line by prolonged stimulation with 4 ng/ml TGF $\beta$ 1 (Cell Signaling Technology) for 14 days<sup>9</sup>. For this, 0.5 million cells were seeded per 10 cm cell culture dish (Nunc) and incubated at 37°C and 5% CO<sub>2</sub>. TGF $\beta$ 1 treatment and vehicle treatment using PBS started 24 hours after seeding and was applied daily. The growth medium was exchanged every other day.

EMT was induced in the HTER and HSER cell lines by prolonged stimulation with 4 ng/ml 4-hydroxytamoxifen (4OHT; Sigma Aldrich) for 14 days<sup>9</sup>. For this, 0.5 million cells were seeded per 10 cm cell culture dish (Nunc) and incubated at 37°C and 5% CO<sub>2</sub>. 4OHT treatment and vehicle treatment using methanol (Thommen Furler) started 24 hours after seeding and was applied daily. The growth medium was exchanged every other day.

To avoid over-confluence and senescence during the time course of HMLEs, HTERs, and HSERs, the cells were split and re-seeded on day four and eight. For this, the cells were washed once with pre-warmed PBS, incubated for 5 min at 37°C with 4 ml pre-warmed TrypLE 1X Express (Gibco), quenched with pre-warmed growth medium, pelleted at 350  $\times$  g for 5 min at room temperature, resuspended in pre-warmed growth medium, and re-seeded using 0.5 million cells per 10 cm cell culture dish.

For harvesting, the cells were washed once with pre-warmed PBS, incubated for 5 min at 37 °C with pre-warmed TrypLE 1X Express (Gibco), fixed for 10 min at room temperature with 1.6% paraformaldehyde (PFA, Electron Microscopy Sciences), scraped off the dish using a cell scraper (Sarstedt AG), and quenched using 4 °C growth medium. The cells were pelleted at 600 × g for 4 min at 4 °C, resuspended in 4 °C PBS at a concentration of about 0.5 million cells per ml and frozen at –80 °C. For mass cytometry analysis, 5-Iodo-2'-deoxyuridine (IdU) at 10 μM was added to the medium 20 min before cell harvesting<sup>25</sup>.

**Mass-tag cellular barcoding.** To minimize inter-sample staining variation, we applied mass-tag barcoding to fixed cells<sup>26</sup>. A barcoding scheme composed of unique combinations of four out of nine barcoding metals was used for this study; metals included palladium (<sup>105</sup>Pd, <sup>106</sup>Pd, <sup>108</sup>Pd, <sup>110</sup>Pd, Fluidigm) conjugated to bromoacetamidobenzyl-EDTA (Dojindo) as well as indium (<sup>113</sup>In and <sup>115</sup>In, Fluidigm), yttrium, rhodium, and bismuth (<sup>89</sup>Y, <sup>103</sup>Rh, <sup>209</sup>Bi, Sigma Aldrich) conjugated to maleimido-mono-amide-DOTA (Macrocyclics). The concentrations were adjusted to 20 nM (<sup>209</sup>Bi), 100 nM (<sup>105–110</sup>Pd, <sup>113</sup>In, <sup>89</sup>Y), 200 nM (<sup>113</sup>In), or 2 μM (<sup>103</sup>Rh). Cells were randomly distributed across a 96-well plate and about 0.3 million cells per well were barcoded using a transient partial permeabilization protocol. Cells were washed once with 0.03% saponin in PBS (Sigma Aldrich) prior to incubation in 200 μl barcoding reagent for 30 min at room temperature. Cells were then washed four times with cell staining medium (CSM, PBS with 0.3% saponin, 0.5% bovine serum albumin (BSA, Sigma Aldrich) supplemented with 2 mM EDTA (Stemcell Technologies) and pooled for antibody staining.

**Fluorescence cellular barcoding and flow cytometry surface protein screen.** To apply the flow cytometry surface protein screen to multiple samples simultaneously, we performed fluorescence barcoding of fixed cells. For this, 18 million cells were washed once with CSM prior to incubation in 3 ml barcoding reagent for 20 min at 4 °C in the dark. As barcoding reagents Alexa Fluor-700-NHS-Ester (AF700, Molecular Probes) and Pacific Orange-NHS-Ester (PO, Molecular Probes) dissolved in dimethyl sulfoxide (DMSO) at 200 μg/ml were used. Single stains or a combination of AF700 and PO were performed in CSM at a final concentration of 0.1 μg/ml or 1 μg/ml and 0.4 μg/ml or 2 μg/ml, respectively. Cells were washed twice with CSM before pooling and staining with E-Cadherin-AF647 (clone 67A4, Biolegend) and EpCAM-FITC (clone 9C4, Biolegend) or CD44-FITC (clone IM7, Biolegend) for 20 min at 4 °C in the dark. Cells were washed once with CSM and filtered through a 40 μm cell strainer. About 0.3 million cells in 37.5 μl CSM were loaded in each well of a 96-well plate of the Human Cell Surface Marker Screening (phycoerythrin [PE]) Kit (Biolegend). Each well contained 12.5 μl of diluted PE-conjugated antibody in CSM. The cells were incubated for 30 min at 4 °C in the dark, according to manufacturer's instructions. The cells were then washed twice with CSM, fixed with 1.6% PFA in PBS for 10 min at room temperature in the dark and washed twice with CSM again, prior to flow cytometry analysis using the LSRFortessa Cell Analyzer (BD Biosciences).

**FACS sorting and RNA sequencing.** For live cell FACS sorting, cells were washed once with pre-warmed PBS, incubated for 5 min at 37 °C with 4 ml pre-warmed TrypLE 1X Express (Gibco), pipetted off the cell culture dish, and collected in 4 °C PBS. Cells were pelleted at 350 × g for 5 min at 4 °C, re-suspended in 4 °C PBS with 1% BSA, and stained with E-Cadherin-AF647 (clone 67A4, 5 μg/ 100 μl, Biolegend) and CD44-PE (clone IM7, 1.25 μg/ 100 μl, Biolegend) for 20 min at 4 °C in the dark. Cells were washed once using PBS with 1% BSA and kept on ice until FACS sorting using the FACSAria III (BD Biosciences). For RNA isolation, cells were pelleted at 350 × g for 5 min at 4 °C and lysed in 350 μl RLT buffer of the RNeasy Mini Kit (Qiagen). RNA was isolated according to the manufacturer's instructions. Briefly, RNA was collected on the RNeasy spin column, washed with 70% ethanol (Merck), and DNA was removed by incubation with DNase I (Qiagen). RNA was collected in 30–50 μl diethylpyrocarbonate (DEPC, Sigma Aldrich)-containing water and stored at –80 °C. DEPC water was prepared by dissolving 1 ml DEPC in 1 L ddH<sub>2</sub>O prior to autoclaving. The RNA quality was assessed using a NanoDrop (Thermo Scientific) and Bioanalyzer (Agilent). RNA sequencing was performed using the HiSeq. 2500 System (Illumina) in SR 50 mode (50 base reads) after poly (A) enrichment and stranded library preparation.

**Antibodies and antibody labeling.** All antibodies and corresponding clone, provider, and metal or fluorescence tag are listed in Mendeleev Table 1 and Mendeleev Table 17 on Mendeleev Data<sup>22</sup>. Target specificity of the antibodies was confirmed by the provider and in our laboratory. Antibodies were obtained in carrier/protein-free buffer or were purified using the Magne Protein A or G Beads (Promega) according to manufacturer's instructions. Metal-labeled antibodies were prepared using the Maxpar X8 Multimetal Labeling Kit (Fluidigm) according to manufacturer's instructions. After conjugation, the protein concentration was determined using a NanoDrop (Thermo Scientific), and the metal-labeled antibodies were diluted in Antibody stabilizer PBS (Candor Bioscience) to a concentration of 200 or 300 μg/ml for long-term storage at 4 °C. Optimal concentrations for antibodies were determined by titration, and antibodies were managed using the cloud-based platform AirLab as previously described<sup>27</sup>.

**Antibody staining and cell volume quantification for mass cytometry.** Antibody staining was performed on pooled samples after mass-tag cellular barcoding. The pooled samples were washed once with CSM. Cells were stained with the EMT antibody panel (Mendeleev Table 17 on Mendeleev Data<sup>22</sup>) and incubated for 45 min at 4 °C followed by three washes with CSM. For mass-based cell detection, cells were stained with 500 μM nucleic acid intercalator iridium (<sup>191</sup>Ir and <sup>193</sup>Ir, Fluidigm) in PBS with 1.6% PFA (Electron Microscopy Sciences) for 1 h at room temperature or overnight at 4 °C. Cells were washed once with CSM and once with 0.03% saponin in PBS. For cell volume quantification, cells were stained with 12.5 μg/ml Bis(2,2'-bipyridine)-4'-methyl-4-carboxybipyridine-ruthenium-N-succidimyl ester-bis(hexafluorophosphate) (<sup>96</sup>Ru, <sup>98–102</sup>Ru, <sup>104</sup>Ru, Sigma Aldrich) in 0.1 M sodium hydrogen carbonate (Sigma Aldrich) for 10 min at room temperature as previously described<sup>23</sup>. Cells were then washed twice with CSM, twice with 0.03% saponin in PBS, and twice with ddH<sub>2</sub>O. For mass

cytometry acquisition, cells were diluted to 0.5 million cells/ml in ddH<sub>2</sub>O containing 10% EQ™ Four Element Calibration Beads (Fluidigm) and filtered through a 40 µm filter cap FACS tube. Samples were placed on ice and introduced into the Helios upgraded CyTOF2 (Fluidigm) using the Super Sampler (Victorian Airship) introduction system; data were collected as .fcs files.

For the mass cytometry experiment including fifteen breast cancer cell lines, cells were stained with the following modifications: Purified Galectin-3 (clone Gal397) was applied at 1 µg/ml for 15 min at 4 °C, the cells were washed with CSM, stained with anti-mouse IgG (polyclonal)-<sup>148</sup>Nd for 15 min at 4 °C, washed, and then the EMT antibody panel was applied as above, but using Ki-67 (clone B56) in channel <sup>198</sup>Pt. We observed a strong background signal in the channel <sup>175</sup>Lu (position of Keratin 7) even in Keratin 7-negative cell lines such as MDA-MB-231 and PBMCs and thus drew a gate to exclude this background signal from downstream analyses.

**Mass cytometry data preprocessing.** Mass cytometry data were concatenated using the.fcs File Concatenation Tool (Cytobank, Inc.), normalized using the MATLAB version of the Normalizer tool<sup>28</sup>, and debarcoded using the CATALYST R/Bioconductor package<sup>29</sup>. The .fcs files were uploaded to the Cytobank server (Cytobank, Inc.) for manual gating on populations of interest. The resulting population was exported as .fcs files and loaded into R v4.1.0 (R Development Core Team, 2015) for downstream analysis.

**Flow cytometry surface marker screen data processing.** Flow cytometry data were compensated on the LSRFortessa Cell Analyzer (BD Biosciences) using single-stained samples. The .fcs files were uploaded to the Cytobank server (Cytobank, Inc.) for manual debarcoding and gating on populations of interest. The mean signal intensity per well and population of interest was exported as an excel sheet. The mean signal intensity of the 'Blank' wells of the screen and the signal intensity of the respective 'Isotype control' well were subtracted. From the resulting intensity values, log<sub>2</sub>-transformed fold changes were calculated.

**Dimensionality reduction analyses.** For dimensionality reduction visualizations using the UMAP algorithm<sup>30</sup>, signal intensities (dual counts) per channel were arcsinh-transformed with a cofactor of 5 (counts<sub>transf</sub> = asinh(x/5)) and z scores were calculated. We used the R UMAP implementation package *uwot* (<https://github.com/jlmeville/uwot>) and 1,000 cells per condition and replicate. All markers except IdU, Cyclin B1, Ki-67, and cleaved CASP3/PARP1 were used.

**Clustering analyses and heatmap.** For PhenoGraph<sup>31</sup> clustering of mass cytometry data, the R *RPhenograph* package (<https://github.com/i-cyto/Rphenograph>) was used. PhenoGraph clustering was performed per cell line, using 1,000 cells per condition and replicate and k = 30. All markers except IdU, Cyclin B1, Ki-67, and cleaved CASP3/PARP1 were used. For the heatmap we performed hierarchical clustering on the z scores of the shown markers, using Euclidean distance and ward.D linkage. The z scores were calculated on the arcsinh-transformed data per marker.

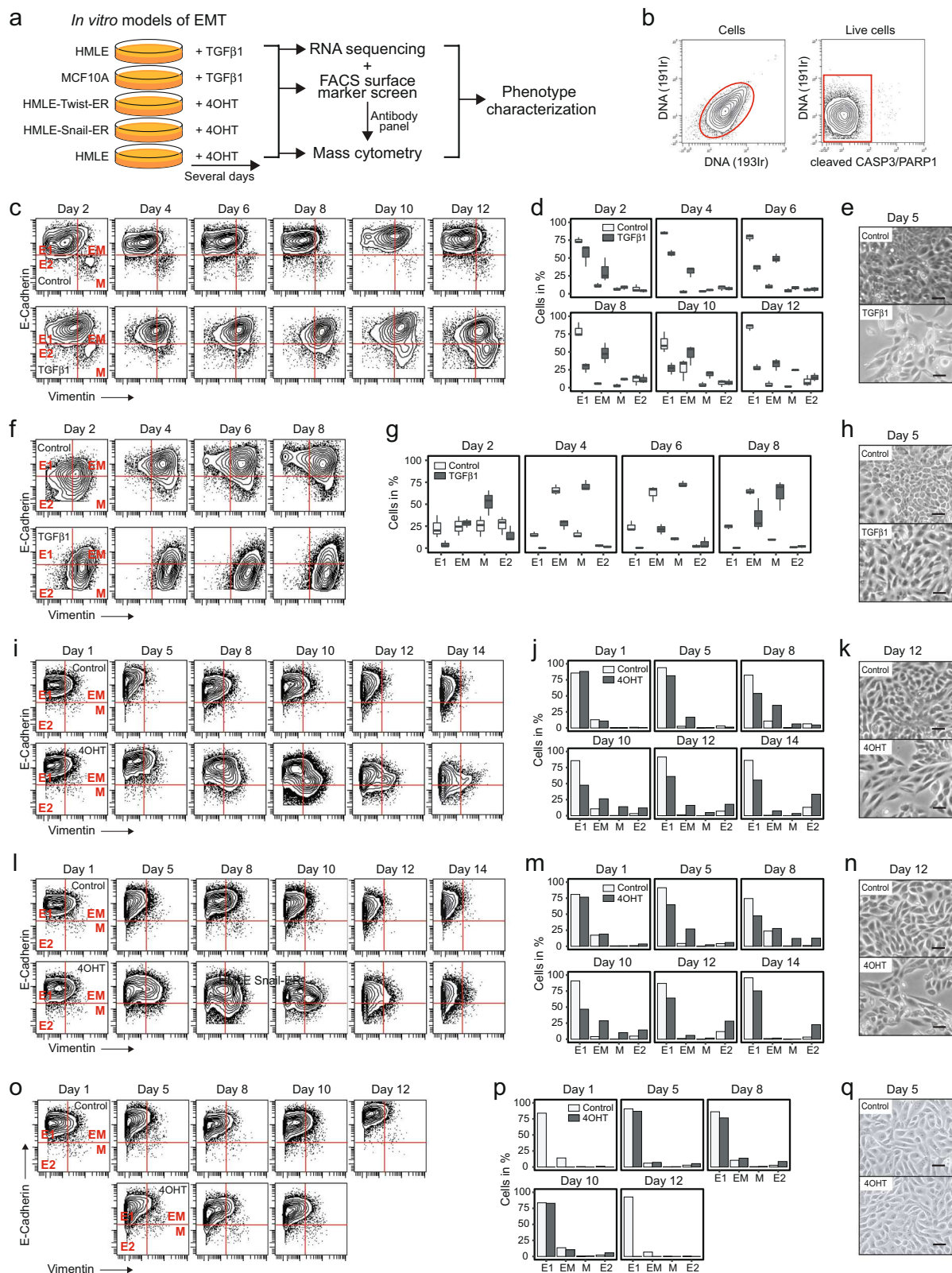
**RNA sequencing data analysis.** The RNA sequencing data was processed using an analysis setup derived from the ARMOR workflow<sup>32</sup>. Quality control of the raw FASTQ files was performed using FastQC v0.11.8 (Andrews S, Babraham Bioinformatics, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Transcript abundances were estimated using Salmon v1.2.0<sup>33</sup>, using a transcriptome index based on Gencode release 34<sup>34</sup>, including the full genome as decoy sequences<sup>35</sup> and setting the k-mer length to 23. For comparison, the reads were also aligned to the genome (GRCh38.p13) using STAR v2.7.3a<sup>36</sup>. Transcript abundances from Salmon were imported into R v4.0.2 and aggregated on the gene level using the tximeta Bioconductor package, v1.6.2<sup>37</sup>. The quasi-likelihood framework of edgeR, v3.30.0<sup>38,39</sup> was used to perform differential gene expression analysis, accounting for differences in the average length of expressed transcripts between samples<sup>40</sup>. In each comparison, edgeR was used to test the null hypothesis that the true absolute log<sub>2</sub>-fold change between the compared groups was less than 1. edgeR was also used to perform exploratory analyses and generate a low-dimensional representation of the samples using multidimensional scaling (MDS). The analysis scripts were run via Snakemake<sup>41</sup> and all the code is available on GitHub (<https://github.com/csoneson/WagnerEMT2020>).

## Data Records

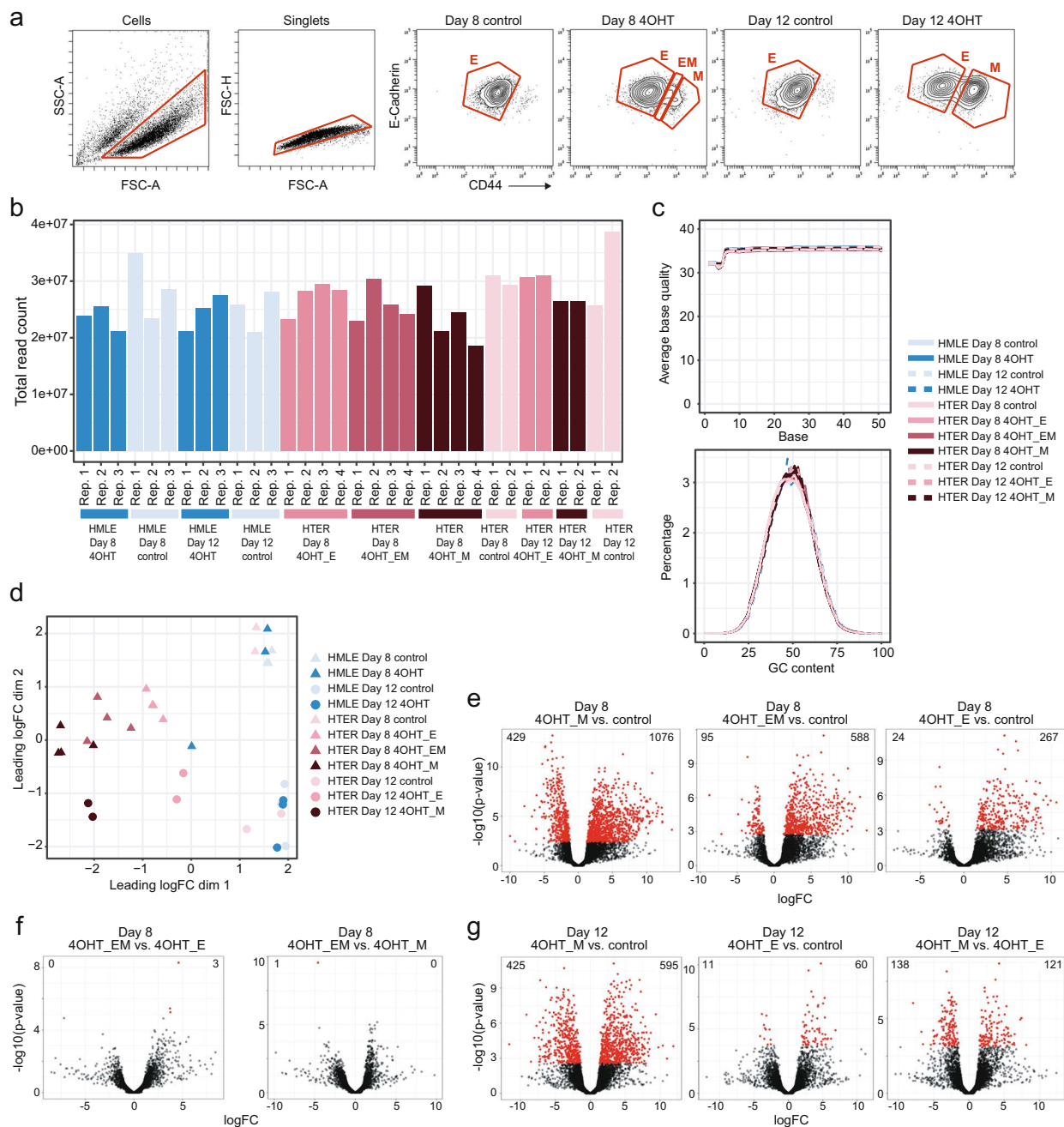
A detailed list of all materials used in this study can be found as Mendeley Table 1 on Mendeley Data<sup>22</sup> (<https://doi.org/10.17632/pt3gmyk5r2.2>). RNA sequencing data have been deposited in the ArrayExpress database at EMBL-EBI with accession number E-MTAB-9365<sup>42</sup>. Tables showing the results of the differential gene expression analyses and a table reporting the RNA quality and RNA sequencing mapping metrics have been deposited as Mendeley Tables 2–13 on Mendeley Data<sup>22</sup>. The code used for RNA sequencing data analysis can be found on GitHub (<https://github.com/csoneson/WagnerEMT2020>). Flow cytometry surface protein screen data as .fcs files and the corresponding data analyses referenced in the text as Mendeley Tables 14–16 have been deposited on Mendeley Data<sup>22</sup>. Furthermore, the Biolegend data sheet corresponding to the flow cytometry screen has been deposited<sup>22</sup>. Mass cytometry .fcs files of cells after debarcoding ('DebarcodedCellsGate') and of live cells ('LiveCellsGate') have been deposited on Mendeley Data<sup>22</sup> together with a table containing.fcs file annotations ('FCS\_File\_Information') and a table corresponding to the antibody panel used (Mendeley Table 17).

## Technical Validation

**Optimizing the time courses for *in vitro* induction of EMT.** We induced EMT in four non-cancerous human mammary epithelial cell lines by prolonged ectopic stimulation with TGFβ1 or 4OHT over several days (Fig. 1a; Methods); all four systems are widely used models of EMT<sup>9,16,21</sup>. We initially carried out a basic characterization of these models and optimized each induction time course to yield the maximum percentage of cells



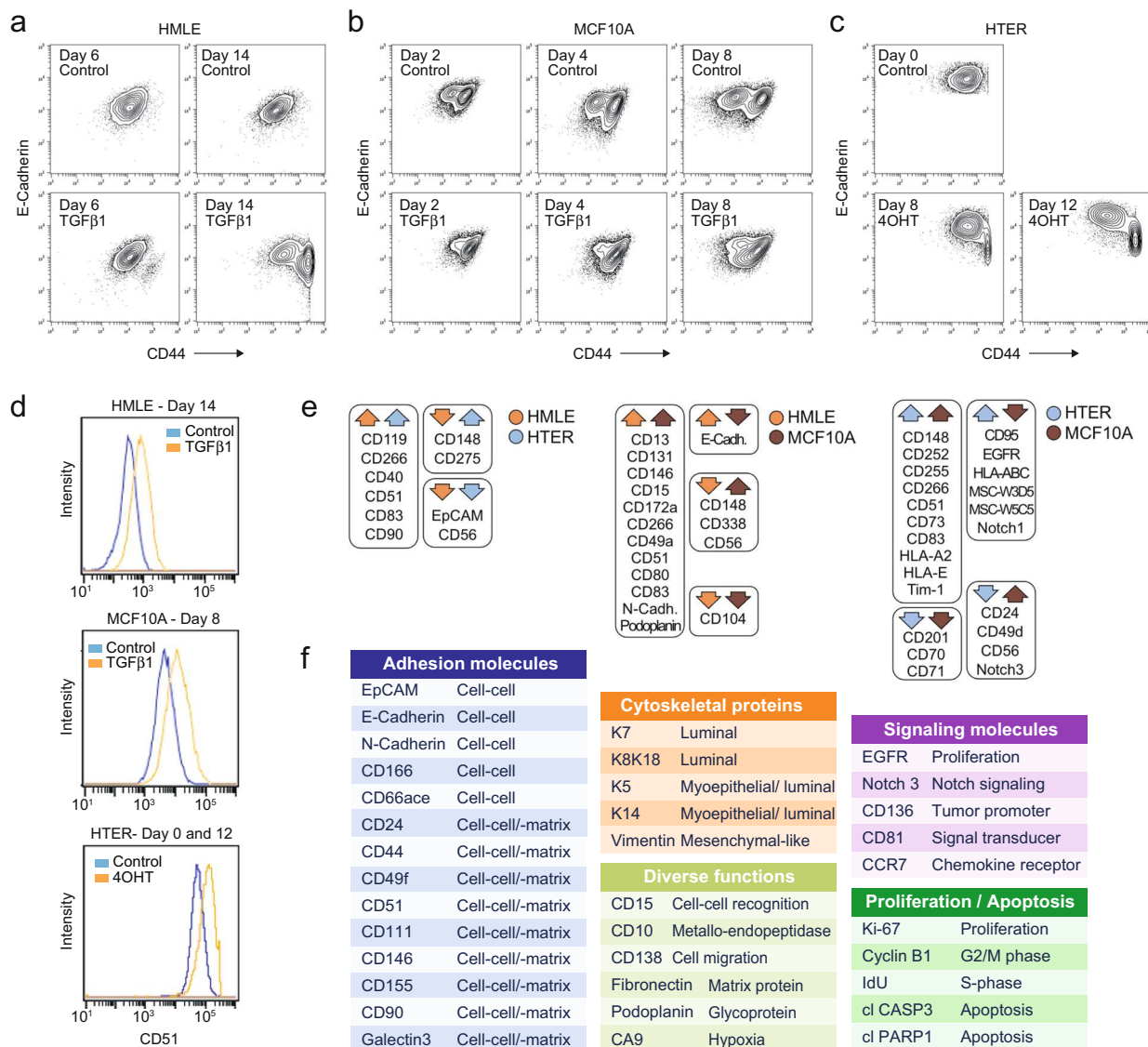
**Fig. 1** Induction of EMT in human mammary epithelial cell lines. **(a)** Experimental workflow. **(b)** Gating to select live cells. **(c)** E-Cadherin and Vimentin expression in HMLEs. Gating to select populations with E1-, E2-, EM-, or M-phenotype. **(d)** Percentages of HMLEs per gate and time point as in **(c)**. **(e)** Phase contrast images of HMLEs. **(f)** E-Cadherin and Vimentin expression in MCF10As. **(g)** Percentage of MCF10As cells per gate and time point as in **(f)**. **(h)** Phase contrast images of MCF10As. **(i)** E-Cadherin and Vimentin expression in HTERs. **(j)** Percentage of HTERs per gate and time point as in **(i)**. **(k)** Phase contrast images of HTERs. **(l)** E-Cadherin and Vimentin expression in HSERs. **(m)** Percentage of HSERs per gate and time point as in **(l)**. **(n)** Phase contrast images of HSERs. **(o)** E-Cadherin and Vimentin expression in HMLEs. **(p)** Percentage of HMLEs per gate and time point as in **(o)**. **(q)** Phase contrast images of HMLEs. Scale bar = 10  $\mu$ m. E1 = epithelial 1, E2 = epithelial 2, EM = hybrid epithelial-mesenchymal, M = mesenchymal.



**Fig. 2** Transcriptomic profiling of EMT-undergoing mammary epithelial cells. **(a)** Gating to select populations of interest of HTERs for RNA sequencing. **(b)** Number of RNA sequencing reads assigned to genes per sample. **(c)** Average base quality (upper panel) and GC content (lower panel) for all samples. **(d)** Multidimensional scaling plot showing the first two dimensions. **(e–g)** Volcano plots showing the indicated differential gene expression analyses. Highlighted in red are genes with an adjusted p-value below 0.05.  $\log_{2}FC = \log_{2}$  fold change, E = epithelial, EM = hybrid epithelial-mesenchymal, M = mesenchymal.

with mesenchymal (M) phenotype, characterized by loss of E-Cadherin and concomitant gain of expression of Vimentin<sup>4</sup>. We excluded apoptotic cells from the analysis (Fig. 1b).

On day 12 of prolonged exposure to TGF $\beta$ 1, the HMLE cell line yielded 25% of cells with an M-phenotype, 33% of cells with a hybrid epithelial-mesenchymal (EM) phenotype with increased Vimentin expression but no downregulation of E-Cadherin, 28% of an E-Cadherin<sup>high</sup>Vimentin<sup>low</sup> phenotype (E1), and 14% of an E-Cadherin<sup>low</sup>Vimentin<sup>low</sup> phenotype (E2) (Fig. 1c,d). In comparison, on day twelve, 2% of control HMLEs exhibited an M-phenotype, 5% an EM-phenotype, 84% an E1-phenotype, and 9% an E2-phenotype (Fig. 1c,d). Control HMLEs with EM- or E2-phenotype were most abundant during sparse growth conditions, such as after splitting (Fig. 1d, Methods), indicating a regulation of E-Cadherin and Vimentin levels by growth density<sup>16,43</sup>.



**Fig. 3** Flow cytometry surface protein profiling of EMT-undergoing mammary epithelial cells. (a–c) Control and treated (TGF $\beta$ 1-treated or 4OHT-treated) cells from the indicated cell lines used for the flow cytometry screen. (d) Histogram overlays comparing CD51 levels between treated and control cells. (e) Proteins that were more than two-fold regulated between treated cells and control in more than one cell line. The arrow direction indicates whether proteins were up- or down-regulated relative to control, and color indicates the corresponding cell line. (f) Antibody panel used for mass cytometry analysis.

As previously reported, treatment with TGF $\beta$ 1 induced spindle-like morphological changes<sup>44</sup> and resulted in lower cell density compared with control<sup>45</sup> (Fig. 1e).

In the MCF10A cell line, induction of EMT by TGF $\beta$ 1 treatment occurred in a different time frame. The percentage of cells with an M-phenotype increased from 54% on day two to 70% on day eight, the percentage of EM cells (28%) and E1 cells (2%) remained stable across the time course, and the percentage of E2 cells dropped from 10% to 2% (Fig. 1f,g). In control, cells with M-phenotype were at 26% on day 2 and 10% on day 8, cells with EM phenotype more than doubled from 25% to 64%, the percentage of E1 cells stayed stable at 22%, and the E2 cells decreased from 29% to 1% over the time course (Fig. 1f,g). As reported, TGF $\beta$ 1-treated MCF10A cells acquired spindle-like morphologies while control cells retained their cobblestone shape (Fig. 1h)<sup>16</sup>. Together, these data show that under sparse growth conditions on day 2, MCF10A cells exhibit mesenchymal-like phenotypes even without TGF $\beta$ 1 treatment, reflecting the basal-like character of the cell line<sup>16</sup>. An increase in cell density over time is accompanied by upregulation of E-Cadherin and therefore loss of the M-phenotype in control, while stimulation with TGF $\beta$ 1 inhibits an E-Cadherin upregulation and induces an upregulation of Vimentin. In TGF $\beta$ 1-treated cells, a decrease in the percentage of cells with M-phenotype on day eight compared with day six, suggests that cell density may inhibit further EMT<sup>46</sup>.

In the HTER and HSER cell lines, EMT was induced by prolonged treatment with 4OHT (Methods). We detected the highest percentage (14%) of 4OHT-treated HTER cells with M-phenotype on day ten, at which



Specificity	Day 6 TGFβ1/Day 6 control	Day 14 TGFβ1/Day 14 control	Day 14 TGFβ1/Day 6 TGFβ1
CD148	-4.10	NA	NA
CD104	-1.83	NA	NA
NPC (57D2)	-1.51	0.58	2.01
CD326 (EpCAM)	-1.31	-2.71	-1.11
CD184 (CXCR4)	-0.95	0.56	1.84
CD300e (IREM-2)	-0.92	-2.35	-0.11
CD275 (ICOSL)	-0.88	-1.82	-0.74
CD56 (NCAM)	-0.24	-1.44	-0.42
CD338 (ABCG2)	-0.23	-1.47	-1.31
CD44	-0.01	3.08	3.17
CD49a	0.15	2.08	1.05
CD166	0.68	2.54	1.81
Podoplanin	0.82	2.90	1.82
CD54	0.86	1.61	1.11
CD90 (Thy1)	0.90	2.88	1.95
CD13	0.92	1.96	1.17
CD263 (TRAIL-R3)	0.96	1.55	0.30
CD80	0.97	1.64	0.37
N-Cadherin	1.25	1.78	0.01
CD146	1.25	2.68	0.51
E-Cadherin	1.44	2.26	-0.53
CD266 (TWEAKR)	1.48	2.02	0.12
CD83	1.49	2.33	0.89
CD119 (IFNgR1)	2.06	1.05	-1.09
CD15 (SSEA-1)	2.07	1.74	0.31
CD182 (CXCR2)	2.66	4.39	0.14
CD51	2.76	NA	0.68
CD172a (SIRPa)	3.51	5.34	0.01
CD162	3.55	NA	0.57
CD134	5.37	NA	-0.14
CD131	NA	4.74	0.46
CD71	NA	NA	-4.36

**Table 1.** Flow cytometry screen results for HMLE cells showing log<sub>2</sub> fold changes selected for at least two-fold differences.

point 26% of cells exhibited an EM-phenotype (Fig. 1i,j,k). The percentage of 4OHT-treated HSER cells with M-phenotype peaked at 12% on day eight and 28% of cells exhibited an EM-phenotype at this time point (Fig. 1l,m). For both cell lines, treatment with 4OHT induced spindle-like morphologies and was accompanied by reduced cell density compared with control (Fig. 1k,n), as previously reported<sup>9</sup>. We then assessed possible effects of the 4OHT treatment on HMLEs in the absence of the Twist1-ER or SNAIL1-ER fusion proteins. As expected, treatment with 4OHT did not induce EMT or morphological changes in HMLEs (Fig. 1o,p,q). In treated and control, the percentage of cells with M-phenotype was below 1% and cells with EM-phenotype at 11% at all time points, indicating a basal-like character of the cell line<sup>9</sup>. The majority of treated and control HMLEs maintained an E1-phenotype throughout the time course (Fig. 1o).

In conclusion, we could induce EMT in four *in vitro* human cell line models of this process. We observed phenotypic variability, including both full and partial EMT phenotypes, in response to 1–2 weeks of prolonged stimulation with TGFβ1 or 4OHT. Each model followed a unique EMT timeline and showed varying extents of transition to the mesenchymal phenotype.

**Transcriptomic profiling of cells undergoing EMT.** We next used RNA sequencing to identify markers that distinguish EMT-undergoing cells from control and markers that distinguish cells with EM-phenotype from cells with E- or M-phenotype. From the resulting markers, candidates were selected to inform a mass cytometry antibody panel design. For RNA sequencing, EMT-undergoing HTER cells on day eight and day twelve were sorted by fluorescence-activated cell sorting (FACS) into three populations: E-Cadherin<sup>high</sup>CD44<sup>low</sup> (E1-phenotype), E-Cadherin<sup>int</sup>CD44<sup>int</sup> (EM-phenotype), and E-Cadherin<sup>low</sup>CD44<sup>high</sup> (M-phenotype) (Fig. 2a, Methods). CD44 served as a surrogate M-phenotype marker for intracellular Vimentin to avoid cell permeabilization and RNA loss<sup>9</sup>. As control, day-matched untreated HTER cells with E1-phenotype were used (Fig. 2a). As a second type of control to monitor possible effects of 4OHT independent of EMT, we included 4OHT-treated and untreated HMLE cells. We included two to four pairs of independent biological replicates per condition and collected high quality RNA for all samples (Mendeley Table 2, Methods).

Specificity	Day 2 TGFβ1/Day 2 control	Day 4 TGFβ1/Day 4 control	Day 8 TGFβ1/Day 8 control
CD201 (EPCR)	-5.58	NA	NA
CD148	-3.46	NA	1.95
CD165	-2.23	-2.51	NA
E-Cadherin	-1.61	-2.17	-1.70
MSC (W3D5)	-1.51	-3.75	-2.72
Notch 1	-1.38	0.97	0.53
CD1a	-1.35	1.52	-2.70
CD9	-1.28	-2.22	-1.61
CD97	-1.17	-2.82	-2.44
CD111	-0.99	-2.00	-2.65
CD70	-0.89	-1.20	-1.44
CD298	-0.87	-1.92	-1.89
Notch 2	-0.79	-1.18	-1.25
CD55	-0.77	-1.14	-1.30
CD96	-0.71	3.61	1.60
CD325	-0.62	1.57	1.44
EGFR	-0.50	-0.33	-1.28
CD56 (NCAM)	-0.45	0.27	2.45
CD46	-0.43	-1.23	-1.45
TCR Vb8	-0.36	-1.48	1.85
CD95	-0.35	-1.02	-1.24
CD11b (activated)	-0.34	NA	1.02
CD338 (ABCG2)	-0.33	2.58	0.45
MSC (W5C5)	-0.30	-1.80	-1.63
Tim-4	-0.25	1.54	0.46
Siglec-10	-0.22	1.05	0.50
DR3 (TRAMP)	-0.22	0.47	1.20
Siglec-9	-0.20	1.03	0.46
CD15 (SSEA-1)	-0.14	1.12	0.66
Notch 3	-0.12	2.35	0.78
CD115	-0.10	0.08	1.78
b2-microglobulin	-0.09	-0.57	-1.11
CD158a/h	-0.08	-1.06	0.07
CD255 (TWEAK)	-0.06	0.84	1.49
CD156c (ADAM10)	0.00	-1.05	-1.04
CD47	0.02	-0.45	-1.05
CD39	0.11	-2.84	0.34
CD49f	0.13	-0.83	-1.07
CD1d	0.14	2.14	0.03
Tim-1	0.15	1.14	0.52
CD88	0.16	0.10	-2.60
CD215 (IL-15Ra)	0.18	1.24	0.58
HLA-E	0.19	1.42	0.56
CD86	0.20	3.31	-0.16
HLA-A2	0.24	1.38	0.08
CD66a/c/e	0.24	-1.93	-3.58
CD24	0.26	0.02	1.11
HER2	0.28	1.60	NA
CD101 (BB27)	0.31	-1.00	-0.63
CD167a	0.31	1.37	1.72
IGF-1R	0.34	0.65	1.11
CD104	0.35	-1.05	-1.27
CD89	0.35	-2.28	-3.49
CD268	0.47	1.39	-0.49
Notch 4	0.47	-0.93	-1.97
CD220	0.48	0.75	-1.07
Continued			

Specificity	Day 2 TGFβ1/Day 2 control	Day 4 TGFβ1/Day 4 control	Day 8 TGFβ1/Day 8 control
CD252 (OX40L)	0.48	1.08	1.37
CD141	0.56	-1.01	0.86
CD318 (CDCP1)	0.60	-0.23	-1.45
CD63	0.64	1.07	1.23
CD114	0.65	0.25	1.48
CD83	0.76	1.54	1.46
CD258	0.77	1.37	-0.75
CD105	0.87	1.14	1.77
CD266	1.02	0.79	0.37
CD80	1.04	-0.18	-0.54
CD49a	1.14	1.31	1.19
TCR Vb23	1.15	1.55	0.62
CD172a (SIRPa)	1.22	0.96	0.57
CD13	1.30	1.44	1.20
CD116	1.36	1.36	0.43
CD146	1.38	2.36	2.07
CD5	1.49	0.00	-0.02
CD1b	1.51	-0.35	0.32
CD138	1.76	0.26	-0.55
CD73	2.48	3.16	1.82
CD131	2.54	0.19	0.73
Podoplanin	2.87	3.33	2.44
CD51	2.95	2.96	2.25
CD49d	6.52	3.79	0.51
CD273	7.98	2.69	4.51
FcRL6	NA	0.04	-1.16
HLA-ABC	NA	NA	-2.26
CD71	NA	NA	-3.19

**Table 2.** Flow cytometry screen results for MCF10A cells showing log<sub>2</sub> fold changes selected for at least two-fold differences.

RNA sequencing yielded above 20 million reads per sample assigned to genes, except one sample with 19 million reads (Fig. 2b, Mendeley Table 2). Mean Phred scores ranged between 35 and 36, indicating high base call accuracy, and GC content distribution across samples did not indicate any noticeable contamination (Fig. 2c, Mendeley Table 2). For all samples, more than 82% of the reads could be uniquely aligned to the human reference genome using STAR<sup>36</sup>. Mapping to the transcriptome index using Salmon<sup>33</sup> showed that more than 86% of fragments were assigned to a transcript, with little variation across samples.

We next assessed the similarity of samples based on global gene expression levels using multidimensional scaling<sup>38,39</sup> (Methods). This showed that the respective pairs of biological replicates were similar (Fig. 2d). Control HTER cells were similar to day-matched 4OHT-treated and control HMLE cells, indicating few effects of 4OHT on transcription independent of EMT. This analysis further revealed that 4OHT-treated HTER cells with E-, EM-, and M-phenotype were all separate from their respective day-matched control (Fig. 2d). Differential gene expression analysis showed that more genes were significantly differentially expressed between HTER cells with M-phenotype or EM-phenotype and control than between E-phenotype and control on day eight (Fig. 2e, Mendeley Tables 3–5). Among differentially expressed genes between M-phenotype and control, we found upregulation of canonical markers of EMT, such as the transcription factors *ZEB1*, *ZEB2*, *FOXC2*, and *PRRX1*, as well as downregulation of typical epithelial markers such as *EPCAM*<sup>1</sup> (Mendeley Table 3). We then asked, which genes were significantly differentially expressed between HTER cells with EM-phenotype and cells with E- or M-phenotype on day eight and found three genes (*HHIP*, *FBN1*, *HHIP-AS1*) and one gene (*KIAA1755*), respectively (Fig. 2f, Mendeley Tables 6 and 7). When comparing HTER cells on day twelve, more genes were significantly differentially expressed between cells with M-phenotype and control than between E-phenotype and control (Fig. 2g, Mendeley Tables 8 and 9).

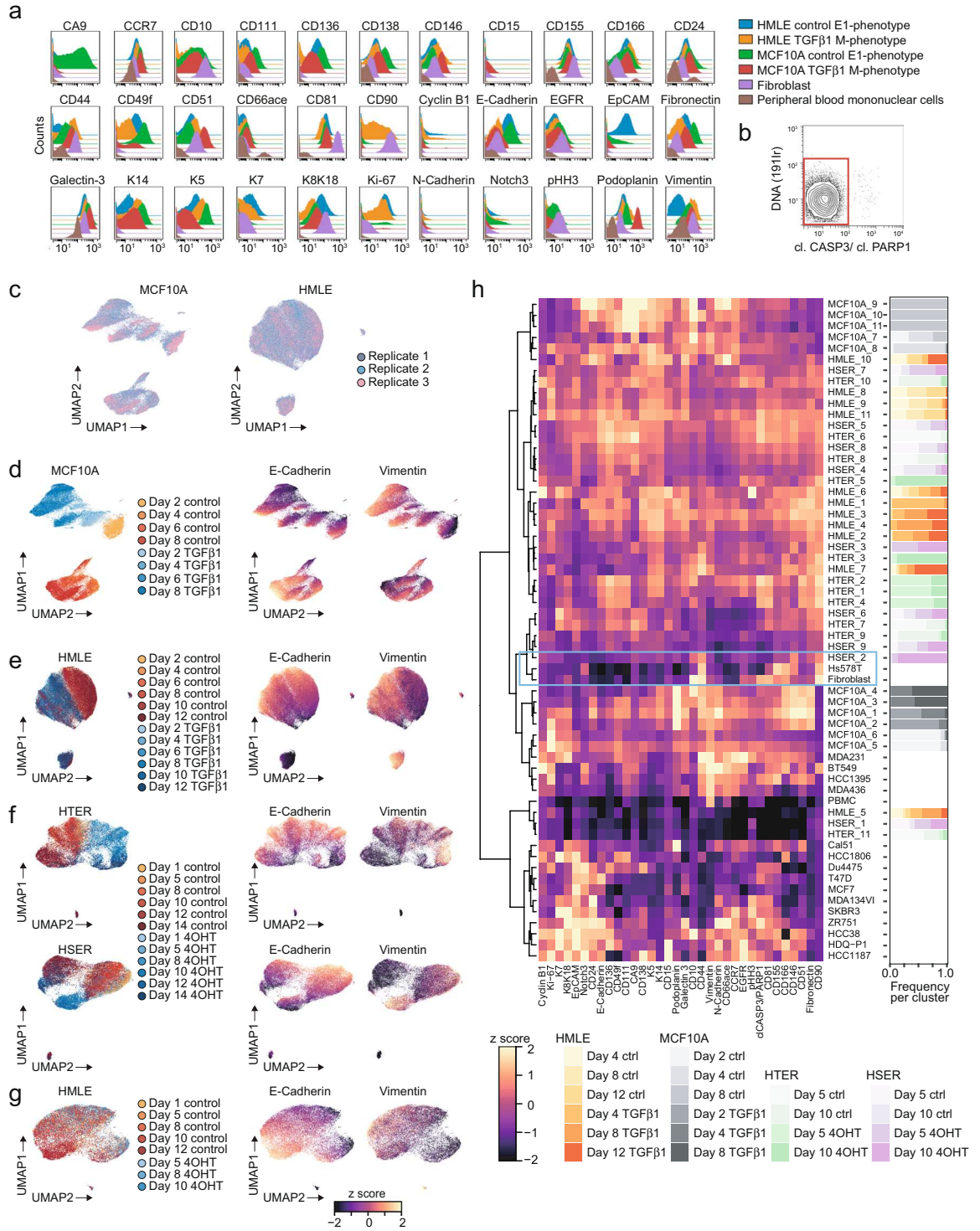
In conclusion, 4OHT-treated HTER cells with M-phenotype or EM-phenotype deviated transcriptionally more from control than cells with E-phenotype. Also, 4OHT-treated cells with E-phenotype are transcriptionally distinct from control cells with E-phenotype.

**Surface protein expression screen during EMT.** We then carried out a flow cytometry surface protein screen to identify further markers that distinguish EMT-undergoing cells from control and to design a mass cytometry antibody panel. Treated and control samples of the HTER, HMLE, and MCF10A cell lines were fixed at multiple time points, fluorescently barcoded, and co-stained with a combination of surface epithelial markers, E-Cadherin and/or EpCAM, and a surface mesenchymal marker, CD44. The resulting flow cytometry data

Specificity	Day 8 4OHT/Day 0 control	Day 12 4OHT/Day 0 control	Day 12 4OHT/Day 8 4OHT
CD20	-4.17	-0.57	3.60
CD49d	-2.67	-3.07	-0.39
CD300F	-2.36	NA	NA
CD28	-2.12	-1.86	0.27
CD201 (EPCR)	-1.99	0.44	2.43
CD56 (NCAM)	-1.84	-0.53	1.31
CD70	-1.43	0.32	1.75
Notch 3	-1.39	-0.11	1.29
CD24	-1.33	-1.23	0.10
EpCAM	-1.32	-3.72	-2.40
CD335 (NKp46)	-1.19	-1.13	0.06
CD1c	-0.93	0.39	1.32
CD340 (HER2)	-0.88	0.70	1.59
CD271	-0.87	0.40	1.28
CD85d (ILT4)	-0.86	0.65	1.51
CD170 (Siglec-5)	-0.82	-2.39	-1.56
CD71	-0.80	-1.40	-0.60
CD275 (ICOSL)	-0.79	1.00	1.78
CD104	-0.61	0.93	1.54
CD109	-0.57	0.86	1.43
HLA-E	-0.40	1.09	1.48
CD95	-0.16	1.09	1.25
CD221 (IGF-1R)	-0.06	1.10	1.16
CD252 (OX40L)	0.06	1.19	1.13
CD119 (IFNgR1)	0.07	1.09	1.02
CD148	0.09	1.19	1.10
CD33	0.20	1.47	1.27
CD73	0.33	1.26	0.93
MAIR-II	0.35	1.17	0.82
EGFR	0.37	1.47	1.11
CD83	0.43	1.53	1.10
Tim-1	0.44	1.45	1.01
CD79b	0.46	1.09	0.63
CD51	0.50	1.14	0.64
HLA-ABC	0.73	1.11	0.37
CD44	0.86	1.17	0.30
MSC (W5C5)	0.90	2.75	1.86
CD90 (Thy1)	0.94	1.45	0.51
CD200 (OX2)	0.95	2.06	1.11
CD255 (TWEAK)	0.97	2.09	1.12
CD93	1.11	1.83	0.72
HLA-A2	1.16	1.16	0.00
CD266 (TWEAK-R)	1.31	0.07	-1.24
MSC (W3D5)	1.56	3.75	2.19
CD10	2.02	3.62	1.60
CD38	5.61	4.69	-0.92
Notch 1	NA	2.44	NA
CD290	NA	NA	-2.71

**Table 3.** Flow cytometry screen results for HTER cells showing log<sub>2</sub> fold changes selected for at least two-fold differences.

were compensated, debarcoded, and gated for cell populations of interest (Fig. 3a–c, Methods). We detected expected surface protein abundance differences between cell populations, such as elevated levels of CD51 in EMT-undergoing cells compared with control<sup>47</sup>, confirming the quality of the screening results (Fig. 3d). We identified multiple surface proteins that were more than two-fold differentially expressed between treated (TGFβ1-treated or 4OHT-treated) and control samples (Tables 1–3, Mendeley Tables 14–16). Several of these were regulated in all three cell lines (CD51, CD83, CD266) or in two cell lines (e.g., CD90, CD146, CD166,



**Fig. 4** Multiplex mass cytometry profiling of EMT phenotypes. **(a)** Histogram overlays showing the antibody panel performance. **(b)** Gating to select live cells. **(c)** UMAPs showing TGFβ1-treated and control MCF10A and HMLE cells colored by biological replicates. **(d–g)** UMAPs showing the indicated cell lines colored either by day and treatment (left) or by marker expression using z score levels (right). For all UMAPs, 1,000 cells per condition and replicate were used. **(h)** Heatmap showing marker expression (columns) using z score levels for each cluster or cell line (rows). Phenograph clustering was applied per EMT cell line. Z scores were calculated on arcsinh-transformed data and per marker. Hierarchical clustering was performed using Euclidean distance and ward.D2 linkage. The blue rectangle highlights that cluster HSER\_2 is phenotypically similar to Hs578T cells and fibroblasts. The bar plot on the right shows the frequency per cluster of cells of the respective day and treatment.

Cell line	Time point	Control	Treatment	Replicates
HMLE	day 2	vehicle	TGF $\beta$ 1	3
HMLE	day 4	vehicle	TGF $\beta$ 1	3
HMLE	day 6	vehicle	TGF $\beta$ 1	3
HMLE	day 8	vehicle	TGF $\beta$ 1	3
HMLE	day 10	vehicle	TGF $\beta$ 1	3
HMLE	day 12	vehicle	TGF $\beta$ 1	3
MCF10A	day 2	vehicle	TGF $\beta$ 1	3
MCF10A	day 4	vehicle	TGF $\beta$ 1	3
MCF10A	day 6	vehicle	TGF $\beta$ 1	3
MCF10A	day 8	vehicle	TGF $\beta$ 1	3
HTER	day 1	vehicle	4OHT	1
HTER	day 5	vehicle	4OHT	1
HTER	day 8	vehicle	4OHT	1
HTER	day 10	vehicle	4OHT	1
HTER	day 12	vehicle	4OHT	1
HTER	day 14	vehicle	4OHT	1
HSER	day 1	vehicle	4OHT	1
HSER	day 5	vehicle	4OHT	1
HSER	day 8	vehicle	4OHT	1
HSER	day 10	vehicle	4OHT	1
HSER	day 12	vehicle	4OHT	1
HSER	day 14	vehicle	4OHT	1
HMLE	day 1	vehicle	—	1
HMLE	day 5	vehicle	4OHT	1
HMLE	day 8	vehicle	4OHT	1
HMLE	day 10	vehicle	4OHT	1
HMLE	day 12	vehicle	-	1
PBMC	—	—	—	1
Fibroblast	—	—	—	1

**Table 4.** Types of samples used for mass cytometry analysis in Fig. 4a–g.

EGFR, N-Cadherin, Notch 3, Podoplanin) and most were regulated in the same direction (up or down) relative to control (Fig. 3e). Based on these flow cytometry screen results and the RNA sequencing analysis, we assembled a panel of candidate targets to assess phenotypic heterogeneity during EMT using a multiplex mass cytometry workflow (Fig. 3f, Mendeley Table 17).

**Mass cytometric profiling of EMT phenotypes.** Mass cytometry is uniquely suited to assess phenotypic heterogeneity during EMT due to its ability to measure about 40 targets at the single-cell level<sup>20,48</sup>. To ensure high data quality, all antibodies against the candidate targets were titrated using samples that represent epithelial phenotypes (HMLE and MCF10A control cells), mesenchymal phenotypes (fibroblasts, TGF $\beta$ 1-treated HMLE and MCF10A cells), and non-epithelial, non-mesenchymal phenotypes (peripheral blood mononuclear cells) (Fig. 4a). We then selected EMT-undergoing and control samples at four to six time points for each of the HMLE, HTER, HSER, and MCF10A cell lines, totaling 92 samples (Table 4). The single-cell suspensions were fixed and mass-tag barcoded<sup>26</sup> to allow pooling and simultaneous antibody staining of the samples (Methods). We used antibodies against cleaved CASPASE-3 (cl. CASP3) and cleaved poly(ADP-ribose)-polymerase 1 (cl. PARP1) to exclude apoptotic cells, yielding more than 1 million live cells for downstream analysis (Fig. 4b). Comparing three biological replicates of the MCF10A or the HMLE cell lines using the dimensionality reduction algorithm Uniform Manifold Approximation and Projection (UMAP)<sup>30</sup> showed a strong similarity of the triplicates for each cell line (Fig. 4c). For MCF10A, the UMAP showed good discrimination of treated and control samples, including differences in E-Cadherin and Vimentin levels (Fig. 4d; Methods). The separation of day 2 control MCF10A cells from other control MCF10A cells is likely caused by the very low expression of epithelial markers and strong expression of the proliferation marker Ki-67 in day 2 cells in contrast to later time points and may reflect growth at lower confluence. Sparse growth conditions have previously been associated with more basal/mesenchymal-like phenotypes in the MCF10A cell line<sup>16</sup>. In the HMLE cell line, TGF $\beta$ 1-treated and control samples were less separable (Fig. 4e). In the HTER and HSER cell lines, we observed a separation of 4OHT-treated cells with E-Cadherin<sup>low</sup>Vimentin<sup>high</sup> phenotype from their respective control on the UMAP (Fig. 4f). In contrast and as expected, 4OHT-treated HMLE cells were indistinguishable from control and displayed only low levels of Vimentin, indicating the absence of an EMT (Fig. 4g). Together, our multiplex mass cytometry data shows that EMT is associated with strong phenotypic changes in all four cell lines.

Cell line	Time point	Control	Treatment	Comment
HMLE	day 4	vehicle	TGF $\beta$ 1	EMT time course
HMLE	day 8	vehicle	TGF $\beta$ 1	EMT time course
HMLE	day 12	vehicle	TGF $\beta$ 1	EMT time course
MCF10A	day 2	vehicle	TGF $\beta$ 1	EMT time course
MCF10A	day 4	vehicle	TGF $\beta$ 1	EMT time course
MCF10A	day 8	vehicle	TGF $\beta$ 1	EMT time course
HTER	day 5	vehicle	4OHT	EMT time course
HTER	day 10	vehicle	4OHT	EMT time course
HSER	day 5	vehicle	4OHT	EMT time course
HSER	day 10	vehicle	4OHT	EMT time course
PBMC	—	—	—	Non-epithelial, non-mesenchymal
Fibroblast	—	—	—	Non-epithelial, mesenchymal
BT-549	—	—	—	Basal breast cancer cell line, Vimentin positive
CAL-51	—	—	—	Basal breast cancer cell line, Vimentin positive
Du4475	—	—	—	Basal breast cancer cell line, Vimentin negative
HCC1395	—	—	—	Basal breast cancer cell line, Vimentin positive
HCC1806	—	—	—	Basal breast cancer cell line, Vimentin negative
HCC38	—	—	—	Basal breast cancer cell line, Vimentin positive
HDQ-P1	—	—	—	Basal breast cancer cell line, Vimentin positive
Hs578T	—	—	—	Basal breast cancer cell line, Vimentin positive
MCF7	—	—	—	Luminal breast cancer cell line, Vimentin negative
MDA-MB-134VI	—	—	—	Luminal breast cancer cell line, Vimentin negative
MDA-MB-231	—	—	—	Basal breast cancer cell line, Vimentin positive
MDA-MB-436	—	—	—	Basal breast cancer cell line, Vimentin positive
SKBR-3	—	—	—	HER2-positive breast cancer cell line, Vimentin negative
T47D	—	—	—	Luminal breast cancer cell line, Vimentin negative
ZR-75-1	—	—	—	Luminal breast cancer cell line, Vimentin negative

**Table 5.** Types of samples used for mass cytometry analysis in Fig. 4h.

We next wanted to assess the phenotypic diversity of EMT-undergoing cells in more detail and in the context of other cell types and cell lines, specifically fibroblasts (i.e., mesenchymal cells), fifteen breast cancer cell lines spanning luminal epithelial, HER2-positive, and basal/mesenchymal-like epithelial phenotypes, and peripheral blood mononuclear cells (PBMC; i.e., neither epithelial nor mesenchymal cells). For this, we repeated the mass cytometry analysis for 35 markers and using a subset of time points for the HMLE, MCF10A, HTER, and HSER cell lines. We included four luminal (MCF-7, T47D, ZR-75-1, MDA-MD-134 VI), one HER2-positive (SKBR-3), eight basal Vimentin-positive (MDA-MB-436, MDA-MB-231, HCC38, HCC1395, BT459, CAL-51, HDQ-P1, and Hs578T), and two basal Vimentin-negative breast cancer cell lines (DU4475, HCC1806), fibroblasts, and PBMCs (Table 5). We then applied the algorithm PhenoGraph<sup>31</sup> to each EMT model individually, which grouped treated and control cells into nine to eleven phenotypically diverse clusters per cell line based on expression of all 35 markers (Fig. 4h, Methods). The majority of clusters contained mostly treated or untreated cells, indicating a treatment-based separation, while other clusters contained cells of both conditions. We observed this separation for all eleven clusters for the MCF10A cell line, six of eleven clusters for HMLE, eight of eleven clusters for HTER, and two of nine clusters for HSER (Fig. 4h). In MCF10A cells, we observed upregulation of CD44, Podoplanin, CD146, and CD51 upon EMT induction compared with control, and concomitant downregulation of E-Cadherin and K5. In the HMLE, HTER, and HSER cell lines, Vimentin, CD44, CD90, CD51, and CD10 were upregulated in EMT-undergoing cells compared with control (Fig. 4h). Several clusters containing EMT-undergoing cells were phenotypically similar to different basal breast cancer cell lines, as determined by hierarchical clustering (Fig. 4h, Methods). For example, cluster HSER\_2 contained 4OHT-treated cells from days 5 and 10 and shared high levels of CD44, CD90, and CD146 and low levels of E-Cadherin, EpCAM, and cytokeratins with the basal Hs578T cell line and with fibroblasts (Fig. 4h, blue rectangle). In another example, the clusters MCF10A\_1–6 contained TGF $\beta$ 1-treated cells from days 4 and 8 or day 2 control cells and shared high levels of Vimentin, CD44, N-Cadherin, and Galectin-3 and low levels of EpCAM and E-Cadherin with the MDA-MB-231, BT549, HCC1395, and MDA-MB-436 basal cell lines. Low levels of epithelial markers in day 2 control cells likely reflects growth at low confluence<sup>16</sup>. In contrast, all luminal and HER2-positive breast cancer cell lines clustered separately from the EMT transition phenotype clusters (Fig. 4h).

In conclusion, we assembled an antibody panel for multiplex mass cytometry characterization of EMT and discovered a vast phenotypic diversity of EMT states among four widely used human *in vitro* models of this process. Several of these EMT states displayed phenotypic similarities with basal breast cancer cell lines and fibroblasts, suggesting that EMT in normal mammary epithelial cells can induce phenotypes observed among aggressive breast cancer cell lines.

## Usage Notes

We provide here a comprehensive characterization of EMT transition phenotypes in four human mammary epithelial cell lines. We characterize transcriptomes and multidimensional protein-level single-cell phenotypes of these cell lines during EMT. We place these transition phenotypes in the context of the multidimensional phenotypes of fifteen luminal, HER2-positive or basal breast cancer cell lines, fibroblasts, and PBMCS. It has previously been shown that EMT in the here used models is associated with increased mammosphere formation<sup>9</sup>, or induction of invasion and migration<sup>49</sup>. A detailed functional assessment of the different molecular phenotypes of EMT-undergoing cells presented here is not part of this study and may be of interest.

## Code availability

The code used for RNA sequencing data analysis can be found on GitHub (<https://github.com/csoneson/WagnerEMT2020>) and can be accessed without restrictions. Please refer to the Methods section above for more details on software versions.

Received: 24 March 2021; Accepted: 7 December 2021;

Published online: 09 February 2022

## References

- Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**(5), 871–90 (2009).
- Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).
- Marjanovic, N. D., Weinberg, R. A. & Chaffer, C. L. Cell plasticity and heterogeneity in cancer. *Clinical Chemistry* **59**(1), 168–179 (2013).
- Nieto, M. A., Huang, R. Y.-J., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **66**(1), 21–45 (2016).
- Paoli, P., Giannoni, E. & Chiarugi, P. Anoikis molecular pathways and its role in cancer progression. *BBA - Molecular Cell Research* **1833**(12), 3481–3498 (2013).
- Desgrosellier, J. S. & Cherech, D. A. Integrins in cancer: Biological implications in therapeutic opportunities. *Cancer, Nat Rev* **10**, 9–22 (2015).
- Cao, Z., Livas, T. & Kyprianou, N. Anoikis and EMT: Lethal “liaisons” during cancer progression. *Crit. Rev. Oncog.* **21**(3–4), 155–168 (2016).
- Micalizzi, D. S., Farabaugh, S. M. & Ford, H. L. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *J. Mammary Gland Biol. Neoplasia* **15**(2), 117–34 (2010).
- Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
- Morel, A. P. *et al.* Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One* **3**(8), e2888 (2008).
- Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178–196 (2014).
- Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nature Medicine* **19**(11), 1438–49 (2013).
- Jordan, N. V., Johnson, G. L. & Abell, A. N. Tracking the intermediate stages of epithelial-mesenchymal transition in epithelial stem cells and cancer. *Cell Cycle* **10**(17), 2865–2873 (2011).
- Thomson, S. *et al.* A systems view of epithelial-mesenchymal transition signaling states. *Clin. Exp. Metastasis* **28**(2), 137–155 (2011).
- Tan, T. Z. *et al.* Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* **6**(10), 1279–1293 (2014).
- Sarrió, D., Rodríguez-Pinilla, S. M., Hardisson, D. & Sarrió, D. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *68*(4), 989–997 (2008).
- Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS* **98**, 10869–10874 (2001).
- Wagner, J. *et al.* A single-cell atlas of the tumor and immune ecosystem of human breast cancer. *Cell* **177**(5), 1330–1345 (2019).
- Zeisberg, M. & Neilson, E. G. Biomarkers for epithelial-mesenchymal transitions. *JCI* **19**(6), 1429–1437 (2009).
- Bandura, D. R. *et al.* Mass cytometry: Technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* **81**, 6813–6822 (2009).
- Elenbaas, B. *et al.* Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* **15**(1), 50–65 (2001).
- Wagner, J. *et al.* Mass cytometric and transcriptomic profiling of epithelial-mesenchymal transitions in human mammary cell lines. *Mendeley Data* <https://doi.org/10.17632/pt3gmyk5r2.2> (2021).
- Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**, 256–268 (2003).
- Brown, K. A. *et al.* Induction by transforming growth factor- $\beta$ 1 of epithelial to mesenchymal transition is a rare event *in vitro*. *Breast Cancer Res.* **6**(3), R215–R231 (2004).
- Rapsomaniki, M. A. *et al.* CellCycleTRACER accounts for cell cycle and volume in mass cytometry data. *Nat. Commun.* **9**(1), 632 (2018).
- Zunder, E. R. *et al.* Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat. Protoc.* **10**, 316–333 (2015).
- Catena, R., Özcan, A., Jacobs, A., Chevrier, S. & Bodenmiller, B. AirLab: A cloud-based platform to manage and share antibody-based single-cell research. *Genome Biol.* **17**, 142 (2016).
- Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytom. Part A* **83A**, 483–494 (2013).
- Chevrier, S. *et al.* Compensation of signal spillover in suspension and imaging mass cytometry. *Cell Syst.* **6**, 612–620.e5 (2018).
- McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform manifold approximation and projection. *J. Open Source Softw.* **3**(29), 861 (2018).
- Levine, J. H. *et al.* Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
- Orjuela, S., Huang, R., Hembach, K. M., Robinson, M. D. & Soneson, C. ARMOR: An automated reproducible modular workflow for preprocessing and differential analysis of RNA-seq data. *G3 Genes, Genomes, Genet.* **9**(7), 2089–2096 (2019).
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).



34. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**(D1), D766–D773 (2019).
35. Srivastava, A. *et al.* Alignment and mapping methodology influence transcript abundance estimation. *Genome Biol.* **21**, 239 (2020).
36. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1), 15–21 (2013).
37. Love, M. I. *et al.* Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLoS Comput. Biol.* **16**(2), e1007664 (2020).
38. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1), 139–140 (2009).
39. Lun, A. T. L., Chen, Y. & Smyth, G. K. It's DE-licious: A recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. in *Methods in Molecular Biology* **1418**, 391–416 (2016).
40. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).
41. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**(19), 2520–2522 (2012).
42. Soneson, C., Wagner, J. & Bodenmiller, B. RNA-seq of two human mammary epithelial cell lines (HMLE and HMLE-Twist-ER) treated with 4-hydroxytamoxifen vs control. *ArrayExpress* <https://identifiers.org/arrayexpress:E-MTAB-9365> (2021).
43. Conacci-Sorrell, M. *et al.* Autoregulation of E-cadherin expression by cadherin-cadherin interactions: The roles of  $\beta$ -catenin signaling, Slug, and MAPK. *J. Cell Biol.* **63**(4), 847–857 (2003).
44. Xu, J., Lamouille, S. & Derynck, R. TGF- $\beta$ -induced epithelial to mesenchymal transition. *Cell Research* **19**, 156–172 (2009).
45. Siegel, P. M. & Massagué, J. Cytostatic and apoptotic actions of TGF- $\beta$  in homeostasis and cancer. *Nature Reviews Cancer* **3**(11), 807–821 (2003).
46. Puliafito, A. *et al.* Collective and single cell behavior in epithelial contact inhibition. *PNAS* **109**(3), 739–744 (2011).
47. Pastushenko, I. *et al.* Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).
48. Di Palma, S. & Bodenmiller, B. Unraveling cell populations in tumors by single-cell mass cytometry. *Current Opinion in Biotechnology* **31**, 122–129 (2015).
49. Kim, E. S., Kim, M. S. & Moon, A. TGF- $\beta$ -induced upregulation of MMP-2 and MMP-9 depends on p38 MAPK, but not ERK signaling in MCF10A human breast epithelial cells. *Int. J. Oncol.* **25**(5), 1375–1382 (2004).

## Acknowledgements

We thank the Bodenmiller lab and the Robinson lab for fruitful discussions. We thank Stéphane Chevrier and the University of Zurich Cytometry Facility for advice and help regarding all flow cytometry experiments. We thank Vito R.T. Zanutelli for advice regarding data visualizations. We thank the ETH Zurich Genomics Facility Basel for excellent RNA sequencing service. We thank the Prof. Dr. Robert A. Weinberg lab (Whitehead Institute for Biomedical Research and Massachusetts Institute of Technology Department of Biology) for their gift of the HMLE, HMLE-Twist-ER, and HMLE-Snail-ER cell lines. BB's research is funded by a SNSF R'Equip grant, a SNSF Assistant Professorship grant, the SystemsX Transfer Project “Friends and Foes”, the SystemsX MetastasiX and PhosphoNetX grants, and by the European Research Council (ERC) under the European Union's Seventh Framework Program (FP/2007–2013)/ERC Grant Agreement n. 336921. MDR acknowledges support from UZH's URPP Evolution in Action.

## Author contributions

JW and BB conceived the study. JW and MM performed the cell culture experiments and flow cytometry surface marker screens with help from AJ and the mass cytometry stainings with the corresponding data processing and interpretation. SS performed the mass cytometry experiment for Fig. 4h. JW performed the FACS sorting and RNA isolation experiments prior to RNA sequencing. CS and MDR performed RNA sequencing data analysis. ND provided extensive help for all data analyses in R. JW, NdS and BB wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to B.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2022