Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Measuring and Modelling the Epithelial Mesenchymal Hybrid State in Cancer: Clinical Implications

Mohit Kumar Jolly[1,*], Ryan J. Murphy[2], Sugandha Bhatia[3,4,5], Holly J. Whitfield[6,7], Melissa J. Davis[6,7,8,*,#], Erik W. Thompson[3,4*,#]

[1]Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore 560012, India

[2]Queensland University of Technology, School of Mathematical Sciences, Brisbane, 4000, Australia

[3]Queensland University of Technology, Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Brisbane, 4102, Australia

[4]Queensland University of Technology, Translational Research Institute, Brisbane, 4102, Australia

[5]The University of Queensland Diamantina Institute, The University of Queensland, Brisbane, 4102, Australia

[6]Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville, 3052, Australia

[7]Department of Medical Biology, University of Melbourne, Parkville, 3010, Australia

[8]Department of Clinical Pathology, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, 3010, Australia

* Denotes co-corresponding authors: mkjolly@iisc.ac.in (MKJ); davis.m@wehi.edu.au (MJD); e2.thompson@qut.edu.au (EWT); Lead contact is e2.thompson@qut.edu.au (EWT)

# Denotes shared senior authors

Key words: epithelial mesenchymal plasticity (EMP), epithelial mesenchymal transition (EMT), mesenchymal epithelial transition (MET), E/M Hybrid, partial-EMP, computational biology, molecular modelling, cancer

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Abstract

The epithelial-mesenchymal (E/M) hybrid state has emerged as an important mediator of the elements of cancer progression facilitated by epithelial mesenchymal plasticity (EMP). We review here the evidence for the presence and prognostic potential of E/M hybrid state in carcinoma, modelling predictions and validations studies to demonstrate stabilised E/M hybrid intermediates along the spectrum of EMP, and computational approaches for characterising and quantifying EMP phenotypes, with particular attention to the emerging realm of single-cell approaches through RNA sequencing and protein-based approaches.

1. Epithelial mesenchymal plasticity, E/M hybrids and cancer progression

E/M hybrid states in cancer

Epithelial–mesenchymal plasticity (EMP) defines a bi-directional axis of phenotypic change: epithelial–mesenchymal transition (EMT) and the reverse process of mesenchymal–epithelial transition (MET). These phenotypic changes have been implicated in embryonic and postnatal development, wound healing, and tissue repair, but are also exploited by cancer cells in avoidance of oncogene-induced senescence, immune-avoidance, invasive progression, survival during dissemination, generation of cancer stem cells, the creation of metastases, and therapeutic resistance [Derynck and Weinberg, 2019, Dongre and Weinberg, 2019, Williams et al., 2019].

Although it is likely that cancer cells can be found to exist at any point across the full EMP spectrum, it is becoming increasingly clear that carcinoma cells often undergo a partial EMT (pEMT) or incomplete EMT, which result in an E/M hybrid state, where cells co-express both epithelial and mesenchymal markers [Huang et al., 2013b, Zhang et al., 2014, Grosse-Wilde et al., 2015, Fustaino et al., 2017, Pastushenko et al., 2018a, Saitoh, 2018, Kroger et al., 2019, McFaline-Figueroa et al., 2019] (Figure 1). The presence of E/M hybrid phenotype is reported in several primary human cancers, including prostate, breast and lung cancer [Livasy et al., 2006, Kolijn et al., 2015, Wang et al., 2017, Zacharias et al., 2018].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

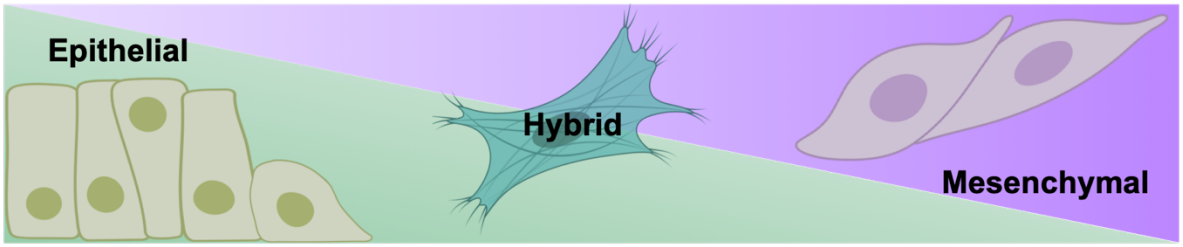


Figure 1: EMP and the E/M hybrid state. *Schematic representation of the Epithelial Mesenchymal Plasticity (EMP) continuum, along which E/M hybrid cells are commonly found in carcinoma tissues. Evidence is growing for the existence of stable intermediates held in check by complex molecular circuitry.*

The E/M hybrid phenotype has been increasingly evident in studies showing coexpression of epithelial (e.g. cytokeratins, E-Cadherin, beta-catenin) and mesenchymal (e.g. vimentin, N-Cadherin, fibronectin) markers in various carcinoma types [Sarrio et al., 2008, Kolijn et al., 2015, Yamashita et al., 2018, Meyer et al., 2019, Navas et al., 2020]. With the advent of molecular profiling, the EMP continuum (and implicit E/M hybrid states) is increasingly well characterised. EMP scores show a spectrum of values derived by composite analysis of expression levels for a range of epithelial and mesenchymal genes in both cell lines and tumour samples, across a large range of cancer origins (e.g. [Tan et al., 2014]). Good concordance exists amongst these different scoring systems, each of which showed that cells can undergo varying extents of EMT across tumour types [Chakraborty et al., 2020a]. These scores, and scRNA-seq across a range of cancer cell lines, indicate a continuum in the gene expression landscape of EMP in both tumours and cell lines [Cursons et al., 2018, Karacosta et al., 2019a, Cook and Vanderhyden, 2020]. Given that EMP often involves changes in cellular morphology too, an EMP continuum in the morphological landscape is also being increasingly characterized recently [Leggett et al., 2016, Devaraj and Bose, 2019]. The trajectories of individual cells in these EMP landscape(s) need not be a cell-autonomous phenomenon; they can be influenced by various biophysical and biochemical modes of cell-cell communication [Mandal et al., 2020, Tripathi et al., 2020b]. Novel analytical tools with single cell RNA-sequencing data are able to align cells along a "pseudospatial" trajectory between these two states [McFaline-Figueroa et al., 2019].

The term metastable state was coined to describe the instability of these E/M hybrid cells [Klymkowsky and Savagner, 2009, Jia et al., 2018], although it may not apply equally to all E/M hybrids. Certain E/M hybrid states appear to be favoured, and these have (i) relative stability / increased proportionality [Biswas et al., 2019], and (ii) enhanced malignant potential, with several studies now having demonstrated a higher malignant potential in E/M hybrid states than more frankly mesenchymal variants [Celia-Terrassa et al., 2012, Huang et al., 2013b, Bierie et al., 2017a, Celia-Terrassa et al., 2018, Pastushenko et al., 2018a, Kroger et al., 2019]. E/M hybrid subpopulations that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

remain more epithelial, with limited additional mesenchymal features, appear to have the greatest malignant and metastatic potential [Jolly et al., 2015, Pastushenko et al., 2018a, Gupta et al., 2019]. Thus, the development of robust and quantifiable methods for estimating the type and extent of E/M hybrid state is becoming an important clinical challenge.

The stability of certain E/M hybrid states is also implied by dynamics of EMP-positive cancer cell systems. Separate analysis of epithelial and mesenchymal transcriptomal components of breast cancer cell lines has shown that high epithelial gene expression often accompanies mesenchymal gene expression in cells designated as mesenchymal, reflecting a predominantly E/M hybrid phenotype, which was also seen in TCGA breast cancer specimens ([Foroutan et al., 2018]; see section 2. Computational approaches, below). The observation of stability states across the continuum also parallels modelling predictions based on plasticity stabilisation factors ([Jolly et al., 2018]; see Section 3. Mechanism-based mathematical models below).

The E/M hybrid phenotype in experimental systems

EMP-related changes occupy a nonlinear landscape in the multi-dimensional space of cancer progression. As mentioned above, physical and functional heterogeneity in cancers resulting from plasticity and changes across the EMP spectrum make it essential to study the E/M hybrid phenotype in cancer studies, for which model systems including cell line systems, preclinical models (e.g. syngeneic mouse cancers), and patient-derived xenografts (PDX) have been employed. For example, tumours arising in a mouse model of metastatic castration-resistant prostate cancer (mCRPC), in which the PI3K/AKT and RAS/MAPK pathways are co-activated in the prostate epithelium, exhibit EMT as detected by a GFP reporter driven off the vimentin promoter [Ruscetti et al., 2016]. Three sub-populations of cells were identified: epithelial ($EpCAM^{+}Vim^{-}$), E/M hybrid ($EpCAM^{+}Vim^{+}$), and mesenchymal ($EpCAM–Vim^{+}$). When cultured, the isolated epithelial and mesenchymal cells remained in their initial cell state, while the E/M hybrid tumour cells exhibited plasticity, transitioning into either epithelial or mesenchymal states within 24 hrs of plating. It has been shown that EMP state subpopulations isolated from systems with inherent EMP tend to revert to a stable equilibrium, as seen with EpCAM-low subpopulations of MMTV-PyMT [Beerling et al., 2016] and PMC42-LA [Bhatia et al., 2019] cells, and either epithelial ($EpCAM^{high}$) or mesenchymal ($EpCAM^{low}$) subpopulations of the HCC-38 human breast cancer cell line [Yamamoto et al., 2017], all of which revert to a tightly controlled epithelial: mesenchymal ratio [Beerling et al., 2016]. In SCC9 cells, $pEMT^{high}$ and $pEMT^{low}$ subpopulations remained distinct at 4 hours and 24 hours of culture after cell sorting, but resembled parental SCC9 cells after 4 days [Puram et al., 2017]. In HNSCC cells, segregated subpopulations returned to unsorted cell proportions over time, with a more rapid plasticity seen in E/M hybrid

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

subpopulations [Pastushenko et al., 2018a]. These data comprehensively suggest that complex dynamics exists amongst the subpopulations of EMP-positive carcinoma cells, supporting the hypothesis of stable E/M hybrid states.

Indeed, evaluation of cancer cell lines in vitro has revealed E/M hybrid states in lung, breast, colorectal, ovarian, prostate and renal cancer cells [Hendrix et al., 1997, Huang et al., 2013a, Bronsert et al., 2014, Andriani et al., 2016, Bierie et al., 2017b, George et al., 2017]. E/M hybrid phenotypes have also been observed in PDX models of human lung, breast and colorectal cancer. In the context of PDX models, serial passages allow the human cancer stroma to be replaced with that of mouse, which provides advantages in clearly distinguishing the changes in the epithelial markers [Chao et al., 2017, Pastushenko et al., 2018b, Mizukoshi et al., 2020]. Aiello *et al*, 2018, using a lineage-labelling strategy in the KPCY mouse model of pancreatic ductal adenocarcinoma (PDAC), reported that pEMT cells were more predominant/proficient in generating epithelial cells compared to cells from tumours that had undergone full EMT [Aiello et al., 2018a]. A more comprehensive picture of EMT spectrum states is provided by the study from Pastushenko *et al* (2018), where they investigated the intermediate states using combinations of the CD106, CD51, and CD61 biomarkers, and found enhanced metastatic competence in relatively stable E/M hybrid states that retained more epithelial features. Further transplantation assays, scRNA seq and methylome analysis from the subpopulations revealed the diversity in EMT features, tumour seeding, dissemination and metastatic capability [Pastushenko et al., 2018b].

E/M hybrid circulating tumour cells

Circulating tumour cells (CTCs) represent a window into the metastatic process and the numbers of CTCs correlate strongly with prognosis and therapy response [Saxena et al., 2019, Burr et al., 2020]. The evidence of E/M hybrid phenotype or partial EMT states from CTC and CTC clusters as part of liquid biopsy assays is gaining hold in terms of their metastatic competence, stemness and therapeutic resistance. The existence of E/M hybrid phenotype has been observed in CTCs of clinical cancer patients at advanced stages [Theodoropoulos et al., 2010, Armstrong et al., 2011]. The E/M hybrid state has been particularly evident in CTCs, in part due to the utilisation of epithelial markers to isolate CTCs from a large background of non-epithelial blood cells, and the subsequent analysis of mesenchymal markers using various methods. Pioneering studies by Yu *et al* (2013), from the Haber lab, employed RNA *in situ* hybridisation of epithelial versus mesenchymal pools of mRNA to show that E/M hybrid EMP states were more common than fully epithelial or mesenchymal states [Yu et al., 2013]. Interestingly, the CTC clusters with E/M hybrid phenotype undergo collective migration, in which cells migrate with retained contact junctions. The leading cells showed increased mesenchymal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

characteristics and actin-mediated mobility, while central cells within the cluster maintained polarity and intercellular junctions, and migrated along with the traction forces generated by leader cells [Revenu and Gilmour, 2009, Aceto et al., 2014, Aceto et al., 2015, Jolly et al., 2015]. These clusters of cancer cells ready to disseminate are often reported in the invasive front of tumours [Yang et al., 2008]. Numerous studies have since confirmed such E/M hybrid enrichment in CTCs (reviewed in [Francart et al., 2018, Saxena et al., 2019]), and CTC phenotype analysis has shown an association between E/M hybrid CTCs and survival of PDAC patients [Sun et al., 2019]. These findings are not universal; a similar study using the CanPatrol RNA-ISH methodology in colorectal cancers found only mesenchymal CTCs to associate with shorter progression free survival, while both total CTCs and mesenchymal CTCs were significantly associated with unfavourable overall survival [Hou et al., 2020]. Nonetheless, CTCs are an abundant source of E/M hybrid cells with strong overall clinical implications.

As detailed more thoroughly in the next section, the last five years has seen a surge in single-cell RNA-sequencing technologies, which more precisely identify the partial EMT or E/M hybrid states from cancer datasets [Sarioglu et al., 2015]. Dong *et al,* 2018 reported the presence of E/M hybrid states in scRNA-seq datasets of primary breast cancer and lung adenocarcinoma patient-derived xenografts [Dong et al., 2018]. Puram *et al,* 2017 compared the scRNA-seq data from the primary and metastatic head and neck squamous cell carcinomas (HNSCCs) and identified a subset of malignant cells with a partial EMT signature [Puram et al., 2017]. *In situ*, these E/M hybrid cells were seen at the leading edge of tumours with cancer-associated fibroblasts. In addition, the E/M hybrid signature correlated with a malignant basal HNSCC subtype and lymph node metastasis, signifying that partial EMT promotes loco-regional invasion [Puram et al., 2017]. The advent of more combinatorial single cell approaches utilising NGS and mass cytometry are beginning to wholly portray integrated genome, transcriptome, DNA methylome [Barros-Silva et al., 2018], and proteome [Abouleila et al., 2019] information from individual cancer cells in relation to their EMT phenotype state [Dey et al., 2015, Macaulay et al., 2017]. It is indeed an exciting time for the E/M hybrid state, attested by these numerous excellent studies and the growing armamentarium of technologies, analytical tools and clinical implications.

Clinical utilisation of the E/M hybrid state

A number of studies have associated evidence of mesenchymal change (gain of mesenchymal markers and/or aberrant epithelial markers) in the tumour parenchyma with poor outcomes, especially prevalent in basal-like breast cancers, many of which are 'triple-negative breast cancers' lacking all 3 predictive receptors (ER, PR, HER2) [Thomas et al., 1999, Sarrio et al., 2008, Jeong et al., 2012, Karihtala et al., 2013]. The unique nature of the E/M hybrid state

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

provides opportunity in assessing EMP from a prognostic / predictive clinical perspective, in terms of both malignant potential and therapy resistance. Advantages of the focus on the E/M hybrid state are two-fold; (i) evidence emerging as described above that the E/M hybrid cells associate better with carcinoma progression than frankly mesenchymal cells, and (ii) the increased certainty that the cells being identified are indeed carcinoma cells exhibiting EMP, since such bi-phenotypic cells are quite restricted outside of malignancy, unlike the abundance of cells exhibiting a mesenchymal phenotype, such as carcinoma-associated fibroblasts, endothelial cells, various immune cells. Several studies have identified basal-like breast cancers co-expressing classical epithelial (e.g. cytokeratin 5/6) and mesenchymal markers (e.g, vimentin) [Thomas et al., 1999, Cakir et al., 2012], and such co-expression has also been seen in colon cancer [Meyer et al., 2019], high grade serous ovarian cancer (HGSOC; [Gonzalez et al., 2018]), and lung carcinoma by novel single cell techniques such as DepArray and CyToff, respectively. High resolution, sub-cellular microscopy has been used to assess E/M hybrid state in a variety of tumours.

Immunohistochemical analysis of co-expression of epithelial and mesenchymal markers provided a more specific assessment of the E/M hybrid phenotype. Such an approach allowed us to identify a trend towards associating co-expression of cytokeratin (CK)-19 and vimentin in cells located in the dense connective tissue stroma with poorer overall survival [Raviraj et al., 2012]. A similar scenario of tumour budding has been reported in the stroma of colon carcinoma and shown to comprise partial EMT [Brabletz et al., 1998, Grigore et al., 2016, Meyer et al., 2019]. Grosse-Wilde *et al*, 2015 [Grosse-Wilde et al., 2015] hypothesised that stemness was associated with E/M hybrid cells, and employed E, M and E/M signatures derived from stable, clonal HMLER sublines, and single cell qPCR analysis to link a mixed EM signature with stemness in i) individual cells, ii) luminal and basal cell lines, iii) in vivo xenograft mouse models, and iv) in all breast cancer subtypes. Co-expression of E and M signatures was associated with poorest outcome in luminal and basal breast cancer patients as well as with enrichment for stem-like cells in both E and M breast cell-lines. Importantly, the prognostic potential of the shared E/M signatures could be related to both intercellular (E:M) cooperativity, as well as enhanced malignancy of the E/M hybrid cells, each of which were proposed as novel therapeutic targets independent of breast cancer-subtype. As detailed in the Modeling section below, an inferential model built on NCI-60 cell line gene

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

expression data screened multiple E/M hybrid predictors against TCGA data across multiple tumour types [George et al., 2017]. The VIM:CDH1 gene expression ratio combined with the expression of CLDN7 provided the best assignment of various tumour cells into three phenotypes — epithelial, E/M hybrid and mesenchymal. Using this approach, metastatic breast tumours could be categorized as either having an epithelial or E/M hybrid phenotype, with the E/M hybrid score, which has tumour type-specific predictive power.

Several recent studies have further demonstrated potential prognostic significance of the E/M hybrid state with various methodologies. Gonzalez et al, 2018, identified clades of E-Cadherin / vimentin co-positive cells in HGOSC, which correlated with metastatic trajectory, and provided a link to a rarer subset of vimentin/cMyc/HE4 co-positive cells that correlated with poor prognosis [Gonzalez et al., 2018]. Karacosta *et al*, 2019 used mass cytometry time-course analysis in experimental systems to resolve lung cancer EMT and MET states [Karacosta et al., 2019b]. They constructed a lung cancer reference map of EMT and MET states referred to as the EMT-MET PHENOtypic STAte MaP (PHENOSTAMP), allowing them to characterise the EMT-MET profile of clinical lung cancers with single-cell resolution. They identified 3 epithelial states and 3 partial EMT states, one mesenchymal state and one MET state, and identified a 'continuous sweep' of these states when assessing clinical NSCLC samples, which resonated with different NSCLC cell lines. A further study [Navas et al., 2020] used high-resolution digital microscopic immunofluorescence analysis (IFA) of β-catenin to quantitate and colocalize E-cadherin and vimentin at subcellular resolution, in an assay they dubbed EMT-IFA. EMT-IFA analysis of core-needle biopsies from various advanced metastatic carcinomas identified E/M hybrid cells, the proportions of which were affected by different carcinoma-specific therapies. The EMT-IFA was proposed as a method for clinical monitoring of tumor adaptation to therapy. Most recently, Godin *et al*, 2020 reported a sequential chromogenic immunohistochemical multiplex (SCIM) technique to quantify E/M hybrid status through pan-CK and VIM analysis in urothelial carcinomas, in relation with cancer aggressiveness [Godin et al., 2020]. Using this methodology, they demonstrated E/M hybrid phenotypes in urothelial carcinomas at the time of diagnosis, and found them to be strongly associated with poor prognosis independent of standard clinicopathological features. Although presence of the E/M hybrid was associated with worse overall and disease-free survival, quantitation of the E/M hybrid did not offer any stratification.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Collectively, the E/M hybrid state appears to offer considerable promise in terms of being (i) a somewhat unique state for which specific assays can be developed using different technologies and (ii) a state with prognostic potential, and also the potential to monitor therapy response, which ultimately may possibly be predictive of therapy response.

2. Computational approaches to reveal regulatory mechanisms underpinning epithelial-mesenchymal plasticity and the E/M hybrid phenotype

Deriving signatures with DE

Plastic phenotypes are often quantified using a workflow that combines a perturbation experiment *in vitro* with a differential expression analysis. To generate an EMT gene expression signature, an EMT program is stimulated using an EMT inducer such as TGFβ [Taube et al., 2010, Du et al., 2016, Cursons et al., 2018] or EGF [Cursons et al., 2015] to enable the identification of a group of genes that are differentially expressed in response to the stimulus, and in correspondence with morphological changes. These differentially expressed genes describe a stimulus-specific gene set for EMT, which can also be separated into "epithelial" and "mesenchymal" components representing the up- and down- regulated genes. Along these lines, Figure 2 displays Cancer Cell Line Encyclopedia (CCLE) breast cancer cell lines and The Cancer Genome Atlas (TCGA) tumour samples across an epithelial-mesenchymal landscape. The landscape shows that Basal and Claudin-low cell lines have lower epithelial scores but higher mesenchymal scores, while the luminal cell lines are less mesenchymal and more epithelial. We would expect that samples expressing an E/M hybrid phenotype would be found in the top right quadrant, with both a high mesenchymal and high epithelial score. Importantly, many Basal and even some of the more fully mesenchymal Claudin-low cell lines exhibit a E/M hybrid transcriptome, as do the vast majority of tumours.

Most experimental studies reviewed here can be used to derive epithelial and mesenchymal gene signatures, however very few have developed gene signatures specific to the E/M hybrid state. Some approximation of this has been made by combining "epithelial" and "mesenchymal" components into one signature. Grosse-Wilde *et al*, 2015 produced a E/M hybrid signature by combining 30 of each of the most extremely expressed genes in the E and M signatures from clonal HMLER human breast cancer cells, and found it to predict

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

poor overall and relapse-free survival in both luminal and basal breast cancer subtypes [Grosse-Wilde et al., 2015]. They found considerable overlap between the HMLER EM signature and a Lung E/M signature also compiled from combining “epithelial” and “mesenchymal” components derived from a microarray dataset of 93 lung cancer cell lines [Loboda et al., 2011]. This Lung E/M signature was found to best differentiate two major unsupervised subpopulations of colon cancers [Loboda et al., 2011], and also associated with poor overall survival in both basal and luminal subtypes of breast cancer. These studies indicated that co-expression of epithelial and mesenchymal markers derived from multiple systems had prognostic value in breast cancer. Signatures from scRNA-seq are discussed below.

Methods for applying signatures

The higher-level biological information captured by gene expression signatures can be mapped onto new datasets using gene set analysis methods, revealing the extent to which they express similar molecular phenotypes. Gene set analysis methods (reviewed and benchmarked recently by Geistlinger and colleagues [Geistlinger et al., 2020]) can assess specific hypotheses, such as “Does this group of samples express an EMT program?”, or competitively test a group of gene signatures i.e. “Which molecular program is upregulated in these samples?”. When applied to differential analysis workflows, EMT gene expression signatures can evaluate the differential enrichment for an EMT program between two groups of samples. Alternatively, single-sample gene set analysis methods such as *ssGSEA* [Barbie et al., 2009] or *Singscore* [Foroutan et al., 2018, Bhuva et al., 2020] can be used to identify individual samples that are concordant with gene expression signatures, and estimate the EMP phenotype of a single sample.

Alternative scoring of EMT phenotypes

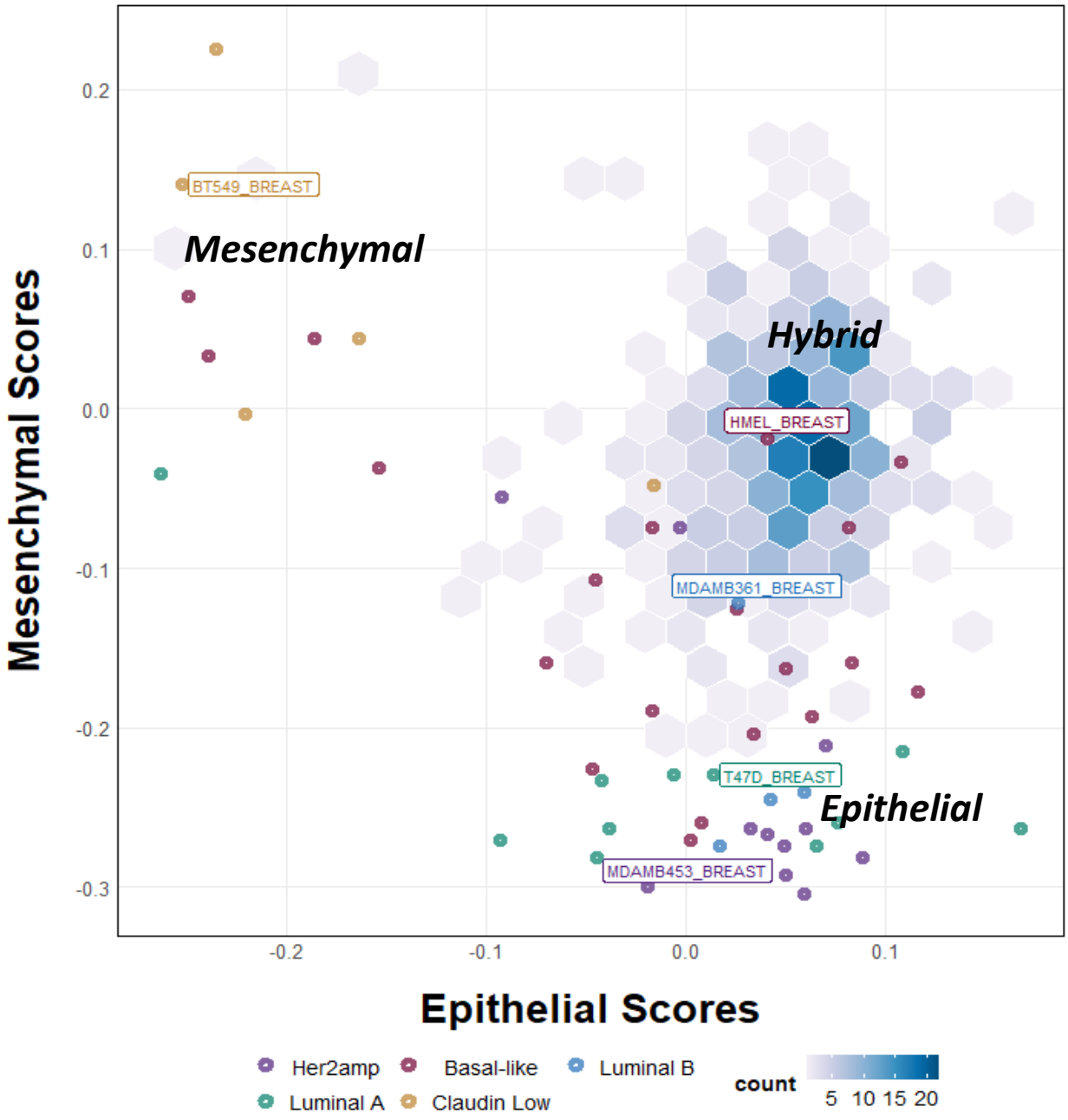
Similarly, various scoring methods have been developed for the quantification of EMT phenotypes in samples independent of a differential expression workflow. These approaches provide scores along a continuous phenotype based on the weighted sum [Byers et al., 2013, Guo et al., 2019] or distribution of gene expression values [Tan et al., 2014, Chakraborty et al., 2020a] for genes that are pro-mesenchymal or pro-epithelial. To

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

place samples on a spectrum inclusive of an E/M hybrid state, a multinomial logistic regression (MLR)-based method was developed using a small number of key regulators to predict a score between 0 and 2 (epithelial to mesenchymal), where a middle value of 1 indicates the E/M hybrid state [Chakraborty et al., 2020a] . Other approaches [Tan et al., 2014, Chakraborty et al., 2020a] assume that the epithelial and mesenchymal states are at two ends of a continuum, and that E/M hybrid states are those that occur between the two extremes. Maximum variability in terms of EMT is associated with increased E/M hybrid signatures [Chakraborty et al., 2020b]. Further, the multinomial logistic regression (MLR)-based metric was suggested to have advantages over the 76 gene signature score and the Kolmogoriv-Smirnov test, as it was capable of distinguishing between the “pure” individual E/M hybrid rather than mixtures of epithelial and mesenchymal cells and uses a small number of predictors to calculate the E/M score.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Figure 2: An epithelial-mesenchymal landscape. *Epithelial and mesenchymal scores for the CCLE breast cancer cell lines (circle markers, coloured by subtype) overlaying a landscape of breast cancer tumour sample scores from TCGA (hexbin density plot) [Barretina et al., 2012, Cancer Genome Atlas, 2012]. Epithelial and mesenchymal signatures were obtained from Tan et al. for both cell lines and tumour samples separately [Tan et al., 2014]. Scores were calculated using singscore and ranked using 5 stable genes [Foroutan et al., 2018, Bhuva et al., 2020].*



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Gene set scoring in single-cell data

Single cell RNA sequencing (scRNA-seq) offers a powerful tool to identify epithelial, mesenchymal and E/M hybrid states in mixed cell populations. Gene set testing methods to characterise molecular phenotypes are well established for bulk analyses [Khatri et al., 2012, Geistlinger et al., 2020], however the knowledgebases and gene signature databases that were developed for bulk sequencing may not be appropriate for the application to single-cell data. Instead, to phenotypically characterise groups of cells in scRNA-seq data, various reference-based cell type annotation methods have been developed. These methods bypass comparison to bulk-derived gene sets by projecting individual cells into a single-cell reference datasets using similarity metrics such as cosine similarity [Kiselev et al., 2018, Srivastava et al., 2018] or Spearman's correlation [Aran et al., 2019], or by applying transfer learning [Lieberman et al., 2018]. However, many of these approaches are dependent upon clustering performance, which introduces a selection bias [Zhang et al., 2019], and may not necessarily infer cell states [Chen et al., 2019]. Alternative methods, such as marker-based phenotyping, rely on existing knowledge of gene expression garnered from protein measurement or bulk RNA sequencing. Additionally, marker-based approaches may not be appropriate for quantifying E/M hybrid phenotypes due to the problem of drop out in scRNA-seq data, where lowly expressed genes may be undetected and falsely quantified as 'unexpressed' [Stegle et al., 2015].

Mapping to references in single-cell data

Lahnemann *et al*, 2020 has recently reviewed single cell "mapping" or cell type annotation [Lähnemann et al., 2020], where the authors argue in favour of reference-based methods. Here, unsupervised clustering followed by cell type annotation is considered "reference-free". The marker genes used in these "reference-free" methods are found on a correlative basis, where the functional relevance of the genes in specific phenotypes is not guaranteed [Yuan et al., 2017]. As discussed by Lahnemann *et al*, we need to enable quantification of the uncertainty in mapping cell type or cell state, however, this requires more sophisticated statistical methods. Although pseudo-bulk analyses [Lun et al., 2016] enable the application of the methods established for bulk sequencing [Tung et al., 2017, Crowell et al., 2019],

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

these generalise large, heterogeneous populations that could otherwise reveal cell-to-cell variation and novel phenotypes.

Single-sample gene set testing in single cell data

Single sample gene set testing methods have been developed for bulk analyses that provide continuous phenotype scores for individual samples. If applied to scRNA-seq data these methods would phenotypically characterise individual cells, independent of clustering. However, previously developed bulk methods do not address transcript drop out and consequential sparsity of single cell data [Hicks et al., 2018]. Imputation methods have been developed [Lähnemann et al., 2020] but can introduce errors in downstream analyses [Andrews and Hemberg, 2018]. Current single sample methods cannot be directly applied to single-cell data, also due to the nuances of scRNA-seq such as the synchronisation of cell cycle stages [Stegle et al., 2015, Chen et al., 2019]. To overcome these challenges many single-cell specific enrichment methods are starting to emerge.

A handful of single-cell single-sample gene set testing methods have been developed that model the count overdispersion and technical noise observed in single-cell data, such as *PAGODA* [Fan et al., 2016] or *VAM* [Frost, 2020]. Alternatively, *Vision* calculates signature scores by summing the expression of "up" genes minus the expression of "down" genes, then standardising based on the cell's expression [DeTomaso et al., 2019] . Although *Vision* does not seem to penalise or account for gene drop out, the aggregate expression score avoids the per-gene technical noise problem. These methods have attempted to adjust gene set testing for scRNA-seq data, however, no benchmarking or comparison between methods has been performed. There is work being done to identify cell-type specific drop out patterns [Qiu, 2020], suggesting that drop out perhaps cannot be modelled during cell type discovery, and that we need an approach for characterising phenotype that is independent of drop out patterns.

Gene expression signature resources and the Hallmarks gene set

Gene expression signatures for a range of molecular phenotypes are available through the C2 collection of gene sets in the Molecular Signature Database (MSigDB) [Liberzon et al., 2015], which contain thousands of experimentally-curated gene sets from the literature.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Knowledgebases such as the Gene Ontology [Ashburner et al., 2000, Consortium., 2019], KEGG [Kanehisa and Goto, 2000] or Reactome [Fabregat et al., 2018] are also available through MSigDB and can provide sets of genes associated with particular pathways or molecular programs. To summarise the large number of gene sets available in MSigDB and reduce redundancy, Liberzon *et al*, 2015 performed consensus clustering on > 8,000 gene sets from the C1-C6 collections to identify 50 clusters of gene sets [Liberzon et al., 2015] . The authors assigned biological themes to each cluster, then curated and refined clusters to produce 50 "Hallmark" gene sets. This includes a 200-gene EMT signature that is representative of 107 individual EMT-related gene sets in MSigDB, including experimentally-derived expression signatures as well as pathways from knowledgebases.

Curated EMT gene signatures

Other EMT gene signatures have been defined using different meta-analysis methods and curation approaches [Groger et al., 2012, Tan et al., 2014, Foroutan et al., 2017]. For example, Foroutan *et al* used two meta-analysis techniques to obtain a stimulus-specific EMT signature using 10 cell line microarray datasets that provided evidence for TGFβ-induced EMT across a range of cell lines [Foroutan et al., 2017]. The signature was then used to assess cell lines from CCLE and patient samples from TCGA for evidence of a TGFβ-driven EMT. Byers *et al* [Byers et al., 2013] curated a 76 gene EMT signature using a lung cancer cell line panel based on gene-gene correlations with one of four key EMT genes (CDH1, VIM, FN1 and CDH2). Byers validated their expression signature on independent datasets where their signature successfully identified cell lines that had undergone EMT. To obtain a directional cancer-specific EMT signature (with epithelial and mesenchymal components), Tan et al stratified pan-cancer cell lines and patients according to known epithelial and mesenchymal markers [Tan et al., 2014]. By applying single-sample gene set enrichment using EMT gene sets from MSigDB, the authors identified EMT signatures that were associated with the most epithelial and most mesenchymal samples.

Need for new gene expression signatures

Many of these EMT gene signatures have been derived from microarray datasets and may not best represent the data we now obtain from modern RNA sequencing platforms [Wang

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

et al., 2014]. High-throughput RNA-sequencing of FACS sorted populations [Fustaino et al., 2017], or even new single-cell sequencing technologies, offer an opportunity for new EMT signatures to be derived. These experiments may better represent the datasets currently being used for EMT scoring, and more accurately reflect the true biology. For example, McFaline-Figueroa and colleagues used single-cell sequencing to demonstrate regulatory checkpoints along the EMT continuum [McFaline-Figueroa et al., 2019]. The authors were able to spatially segregate cells that had undergone EMT from epithelial colonies, and by applying their method *Monocle2,* were able to align cells along a "pseudospatial" trajectory between these two states. Single-cell RNA sequencing experiments enable a finer characterisation of phenotypes and can reveal expression patterns of cells along an EMT continuum, making them suitable for deriving gene expression signatures in rarer, novel states such as the E/M hybrid state, and potentiallly in stable intermediates.

Single-cell sequencing lacks the methodology for signature derivation

Method development for single cell data analysis remains an active area of research, where identifying expression patterns in individual cells has been particularly challenging due to the low number of genes detectable per cell [Hicks et al., 2018]. Consequently, there are no established workflows for deriving gene signatures from single cell data. Using scRNA-seq approaches, E/M hybrid states have been identified based on the presence of both pro-epithelial and pro-mesenchymal markers in skin cancer [Dong et al., 2018], ovarian cancer [Gonzalez et al., 2018], HNSCC [Puram et al., 2017, Karacosta et al., 2019b], and lung, mammary and skin cancer [Pastushenko et al., 2018c]. Despite observations of E/M hybrid states, or partial EMT programs, there is a lack of gene expression signatures for this phenotype.

To instead identify expression signatures of cells in an unsupervised way, Puram *et al* applied non-negative matrix factorisiation (NMF) clustering to reveal a partial EMT gene expression program that expressed both ECM genes and EMT markers in a subset of cells [Puram et al., 2017]. To establish the independence of this partial EMT signature, the authors confirmed the lack of correlation with known EMT signatures and used a highly expressed marker from this signature, TGFβ-induced (TGFBI), to sort SCC9 cells into $pEMT^{high}$ and $pEMT^{low}$ cells and validate the findings. The cells that expressed the partial EMT

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

signature were more invasive, but less proliferative, and were found using immunohistochemistry analysis for the top 10 genes (plus the p63 HNSCC marker) near the leading edge of the tumour adjacent to cancer associated fibroblast (CAF) populations.

Single cell sequencing offers the opportunity to characterise the heterogeneity of EMP and to uncover the particular molecular signatures that regulate the E/M hybrid state. Deriving robust signatures for the E/M hybrid state requires analysis methods that are sensitive to drop out and can address the challenge of mapping between bulk and single-cell data. To ensure the quality and reproducibility of both the signature-derivation methods and gene expression signatures, it is of importance that these are accessible and well communicated, for them to be eventually benchmarked. The application of single cell-derived signatures to bulk-sequencing data will also need to be validated and be able to distinguish E/M hybrid subpopulations from heterogeneous epithelial and mesenchymal populations.

Methods relating to phenotype quantification in single cell

Name	Type	Method	Single Sample	Reference
SingleR	Cell type annotation	Cluster reference + Spearman's correlation	Y	[Aran et al., 2019]
CaSTLe	Cell type annotation	Transfer learning between dataset and reference	N?	[Lieberman et al., 2018]
Cell BLAST	Cell type annotation	Autoencoder	Y	[Cao et al., 2019]
cellHarmony	Cell type annotation	Merges KNN graphs across datasets (Louvain)	Y	[DePasquale et al., 2019]
CellAtlasSearch	Cell type annotation	Cosine similarity to pre-processed reference	Y	[Srivastava et al., 2018]
CHETAH	Cell type annotation	Hierarchical classification tree *	Y	[de Kanter et al., 2019]
CellAssign	Cell type annotation	Probabalistic model	Y	[Zhang et al., 2019]
PAGODA	Gene set testing	PCA projection	Y	[Fan et al., 2016]
Vision	Signature scores	Aggregate expression	Y	[DeTomaso et al., 2019]
iDEA	DE and GSE	Uses summary statistics?	N?	[Ma et al., 2020]
LandSCENT	Cell type annotation	Define 'states' using entropy, then infer cell types of states	N	{Chen, 2019 #9527}

* Decision in branch traversal is based on expression profile correlations, where feature selection finds genes that discriminate between branches (includes "unassigned" or "intermediate" classes)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

3. Mechanism-based mathematical models to characterise signatures and implications of E/M hybrid phenotypes

Role of mathematical modelling in identifying the E/M hybrid state(s)

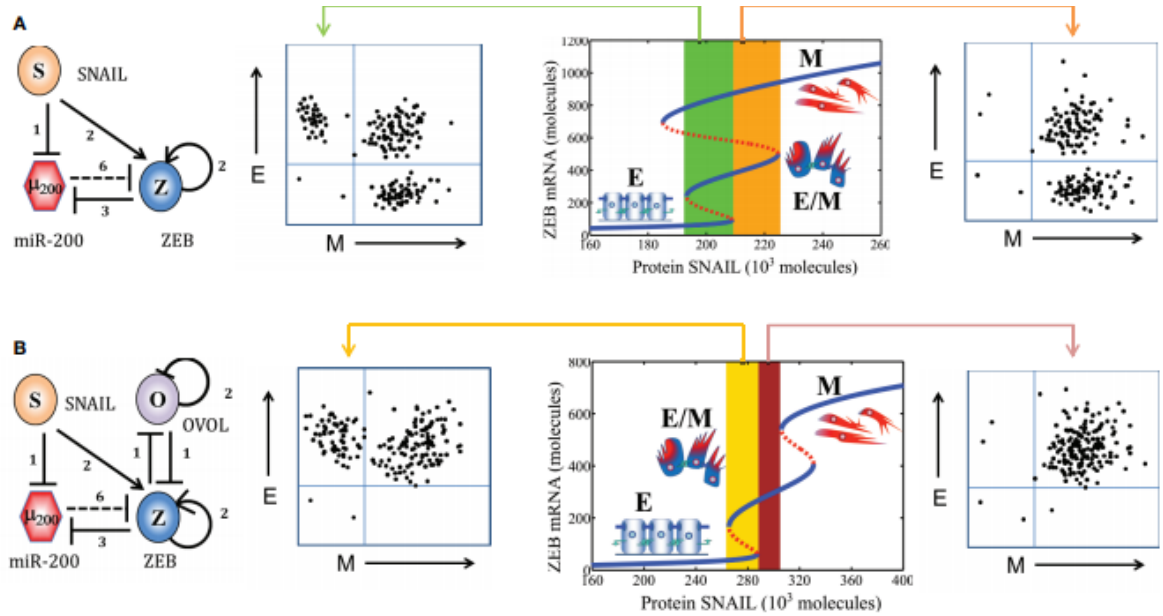
Mathematical models have played a major role in our understanding of the E/M hybrid state. As experiments typically tend to be performed at a single spatial scale (cell-level or population-level) it can be challenging to interpret results from these different spatial scales when processes also occur over multiple overlapping time scales. Mathematical models are a powerful tool with which to integrate the observations from different spatial and time scales, to not only interpret these experimental results, but to also generate and test new hypotheses that can be experimentally tested. This approach has proven extremely insightful in transforming the perception of EMT from a binary (all-or-none) process to a multistable process. Significantly, mathematical modelling studies were among the first to predict the existence of a stable E/M hybrid state characterised by stabilised co-expression of epithelial and mesenchymal markers [Lu et al., 2013, Tian et al., 2013]. This was later confirmed experimentally, both *in vitro* [Bao et al., 2013, Zhang et al., 2014, Pankova et al., 2016] and *in vivo* [Puram et al., 2017, Aiello et al., 2018b, Pastushenko et al., 2018b].

The two mathematical models, by Lu et al. [103] and Tian et al. [141], which initially predicted the existence of the stable E/M hybrid state, focused on signalling pathways within a single cell and built upon decades of observational and functional work by developmental and cancer biologists. Multiple signalling pathways have been implicated in EMT over many studies, but the activities of these pathways can be considered to converge to a core regulatory circuit composed of two families of transcription factors, SNAIL and ZEB, with two families of micro-RNAs, miR-34 AND miR-200, respectively [Nieto et al., 2016]. These two mathematical models used ordinary differential equations (ODEs) to characterise the dynamical behaviour of the core EMT regulatory circuit and observed that a stable E/M hybrid state, distinct from the epithelial and mesenchymal states, was possible. Further extensions to these models have identified factors to stabilise E/M hybrid phenotypes and have since been experimentally validated [Hong et al., 2015, Jolly et al., 2016]. For instance, knockdown of these “phenotypic stability factors” such as OVOL1/2 can drive E/M hybrid cells to a “complete EMT” phenotype during mammary morphogenesis [Watanabe et al.,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

2014] and in lung cancer too [Jolly et al., 2016]. These models tend to display non-linear responses and hysteresis whereby cells *en route* through EMT and MET may take different paths.

Different models of hysteretic patterns in EMT dynamics have been presented. Here, for illustrative purposes, let us consider an example presented in Figure 3 (adapted from Jolly et al. [83]) where ZEB1 mRNA levels vary in response to an inducing signal, in this case the protein SNAIL. In Figure 3A, for a proposed miR-200/ZEB/SNAIL circuit, if a cell starts in an epithelial phenotype and its endogenous SNAIL levels are increased, then it first transitions to the E/M hybrid state, and if SNAIL is further increased the cell transitions to the mesenchymal state. In the reverse direction, by decreasing SNAIL, the cell directly transitions to the epithelial state from the mesenchymal state. In Figure 3B, by including OVOL, a cell initially in the epithelial state transitions to the E/M hybrid state, and then the mesenchymal state with increasing TGFβ. When decreasing TGFβ, the cell first transitions to the E/M hybrid state and then the epithelial state. However, the path followed in both directions is still different – a phenomenon called as 'hysteresis'. Celia-Terrassa et al. [33] further demonstrated mathematically and validated experimentally by applying TGFβ treatment and then TGFβ withdrawal, that most but not all normal and tumour mammary epithelial cells exhibit hysteretic patterns in TGFβ-driven EMT. Statistically significant evidence of hysteresis is also found in TGFβ-treated lung cancer cells [Karacosta et al., 2019b].



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Figure 3: Hysteresis in models of EMT. *(A) Left – miR-200/ZEB/SNAIL circuit. Middle – bifurcation of ZEB mRNA levels in response to protein SNAIL. In the green region, corresponding to a range of SNAIL, cells can be any of three states, E – epithelial, E/M hybrid, M – mesenchymal corresponding to a trimodal distribution as shown in the FACS (fluorescence-activated cell sorting) plot. For the orange region, corresponding to a different range of SNAIL, cells can either be in the E/M hybrid state or mesenchymal state corresponding to a bimodal distribution in the FACS plot. Cells with a value of SNAIL less than in the green region adopt the epithelial state. Cells with a value of SNAIL greater than the orange region adopt the mesenchymal state. (B) Left- miR-200/ZEB/SNAIL/OVOL circuit. Middle – bifurcation of ZEB mRNA levels in response to protein SNAIL. In the yellow region cells can either be in the epithelial or E/M hybrid state. In the red region cells occupy the E/M hybrid state. FACS plots are shown for both yellow and red regions. (Adapted from Figure 5 Jolly et al. [83]).*

Alternative single cell models have focused on larger networks describing the multiple signalling pathways implicated with EMT. While continuous models for large networks can provide information regarding temporal network dynamics, such as phenotypic plasticity and transition rates, it can be experimentally intractable to determine all of the relevant parameters, therefore Boolean-logic approaches, where a gene is considered either active or inactive, have been considered. Using this approach, a study by Steinway et al. [134] identified two stable fixed points corresponding to the epithelial and mesenchymal phenotypes. A subsequent study identified new fixed points that exhibit features of both epithelial and mesenchymal phenotypes thought to be associated to E/M phenotypes and identified putative targets that supress hepatocellular carcinoma EMT. *In vitro* analysis indicated many of the predictions do supress the TGFß-driven EMT, and thus are potential therapeutic targets [Steinway et al., 2015]. Later work by Font-Clos et al. [56] extended the work of Steinway et al. [134] to identify multiple E/M fixed points between the more stable epithelial and mesenchymal fixed points. Hari et al. [74] have since compared the Boolean-logic approach with *RACIPE* (Random Circuit Perturbation), where an ensemble of continuous models is generated with random chosen kinetic parameters for a given network topology and steady state solutions are clustered to identify robust dynamical features of the given network. They show that multistability correlates with the net number of positive feedback loops. In contrast to traditional approaches where nodes in a network are targeted to control EMT plasticity, Hari et al. [74] propose breaking the feedback loops by targeting their links rather than nodes as a possible route to restrict the EMT plasticity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Population dynamics of EMP

Experiments at the population scale have observed epithelial-mesenchymal heterogeneity where cells in the same tissue/population could be either epithelial, mesenchymal, or in an E/M hybrid state, and are able to switch between these states. Recent studies have demonstrated the ability of cells to switch states in the absence of EMT-inducing factors such as TGFß. Ruscetti et al. [127] show that an initial population of E/M hybrid cells can quickly generate a mixture of epithelial, E/M hybrid, and mesenchymal cells. Further, Bhatia et al. [18] demonstrates that an initial population consisting of either epithelial or mesenchymal cells has the ability to regain the phenotypic equilibrium of the parental population. Mathematical models at the population scale for EMT dynamics have tended to be phenomenological. Markov models, where the states of cells in the population at a future time depend only on the states on the current time, have been popular in describing experimental data [Gupta et al., 2011, Mandal et al., 2016, Risom et al., 2018, Chapman et al., 2019, Karacosta et al., 2019b]. However, these models lack integration with the core regulatory network models previously discussed. Therefore, recent work has explored how to link the well-characterised dynamics of EMT regulatory networks to experimentally observed population-level behaviour. In an example of integrating experimental data with known interaction and regulatory networks, Cursons et al. [44] used an experimental model of EMP driven by the well-known ZEB1-miR200 network [Gregory et al., 2011]. Exploring plasticity across an EMP landscape, they integrated experimental data with a known miR-mRNA interaction network and known protein interaction networks to characterise a co-targeted regulatory network regulating EMT and MET.

One approach to incorporate core regulatory networks is to consider a population of cells where each cell contains a copy of the core EMT regulatory circuit [Tripathi et al., 2020a]. The dynamics of the EMT circuit can be calculated for each cell independently to determine the phenotype: epithelial, E/M hybrid, or mesenchymal. Motivated by spontaneous EMT observed in Ruscetti et al. [127], Tripathi et al. [142] developed a model which included cell proliferation, and show that noise in partitioning of signalling pathways during division could be a possible mechanism by which epithelial-mesenchymal heterogeneity can be maintained over extended periods of time involving multiple generations and passages.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

Spatiotemporal dynamics and heterogeneity in EMP: Role of cell-cell communication and cell-ECM crosstalk

An alternative approach to link the cell-level EMT behaviour is to explicitly consider spatial structure. Early evidence of EMT in clinical samples arose from differences in spatial localisation of ß-catenin and E-cadherin across different parts of the tumour [Brabletz et al., 2001]. Related mathematical models tend to be multi-scale models where dynamics of intracellular E-cadherin and ß-catenin signalling drive cellular level properties including cell-cell adhesion. This has been explored using both an off-lattice framework [Ramis-Conde et al., 2008] and an on-lattice cellular Potts model, by integrating CompuCell3D and Bionetsolver [Andasari et al., 2012]. Both models demonstrate properties associated with EMT where cells lose cell-cell adhesion, break off from a primary tumour body, and migrate through and invade surrounding tissues as part of the metastatic cascade. Tumour budding, where numerous finger-like projections, or buds, extend from the primary tumour towards neighbouring stroma in various cancers, has also been identified as marker of invasive behaviour and a possible clinical readout of partial EMT [Bronsert et al., 2014, Grigore et al., 2016]. Other models include features associated with EMT states, such as the different survival times, migration speeds, and interactions with the environment, to develop a novel multi-organ model that explicitly accounts for EMT processes in individual cancer cells in the context of the invasion-metastasis cascade, and are able to produce biologically realistic outcomes regarding tumour shape and metastatic distribution [Franssen and Chaplain, 2019]. Bocci et al. [24], by coupling EMT and cell migration, consider one part of this process – the migration from the primary tumour - to explain individual and collective cancer cell migration. Their model robustly recapitulates circulating tumour cell cluster fractions and size distributions experimentally observed across several cancer types. They identify that mechanisms that increase the population of the E/M hybrid state, rather than a complete EMT, are required to generate large clusters of five to ten cells.

Mesenchymal states and epithelial states of breast cancer stem cells have been observed located at the tumour invasive front and more centrally, respectively [Liu et al., 2014]. By integrating the core EMT regulatory circuit into each cell in a two-dimensional hexagonal grid lattice, spatio-temporal models can be developed to explore the role of cell-cell communication, through Notch-Delta or Notch-Jagged signalling, to show that a gradient of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

EMT-inducing signal, for example TGF-beta, can recapitulate the experimentally observed spatial segregation of different EMT phenotypes [Bocci et al., 2019a]. Mechanosensing through cell-extracellular matrix crosstalk can also influence EMT through an EMT-ECM stiffness feedback loop. Stiffer substrates can results in cells becoming more mesenchymal-like [Matte et al., 2019] and EMT can alter the stiffness of the ECM [Peng et al., 2017]. Kumar et al. [95] use a system of ODEs to describe these relationships and illustrate that the scattering of a heterogeneous population depends on both the ECM density and fraction cells which have undergone EMT. By considering a cell contractility-ECM feedback loop, [Ahmadzadeh et al., 2017], in experimental results and modelling, also find that the driving force underlying EMT is directly proportional to matrix stiffness and identify intermediate ECM stiffness as being optimal for cell invasion. Further experiments by [Margaron et al., 2019] investigate the mechanical properties of single cells in different EMT states and interestingly find that the mechanical properties of E/M hybrid cells are not in between those of epithelial and mesenchymal cells. Instead, they find that E/M hybrid cells produce lower forces and as a consequence of their lower contractility move faster and have a higher invasive potential. Using these observations they identified that triple-negative breast cancer cells had E/M hybrid characteristics rather than mesenchymal characteristics. Their results suggest that the mechanical and structural aspects of the E/M hybrid state are important to understand tumour-cell dissemination.

In summary, mathematical models have served as a powerful tool to unravel new biological insights into EMT/MET, and have driven the next set of experiments, for instance, those characterising the features of stable E/M hybrid state(s). Further development of these models, with close integration and validation with experimental data, will improve understanding of stability of the E/M hybrid states, their spatial patterns, and how they contribute to disease progression.

4. Conclusions, state of play, and potential needs

Clearly there is growing evidence for a preponderance of E/M hybrid states in carcinoma systems, and that these states often are associated with more aggressive malignant / metastatic capacity. As detailed above, a growing number of computational approaches have been developed to illustrate and characterise E/M hybrid states, and theoretical prediction

and support for the stability of particular intermediates has been gleaned from mathematical modelling around some of the well characterised molecular feedback systems. The E/M hybrid state, in a carcinoma context, is somewhat unique, such that opportunities exist in the prognostic and predictive biomarker setting that have started to emerge and show promise, and new technologies are being developed to assess this. To our knowledge, whether these E/M hybrid states show specific sensitivity/resistance profiles to cancer therapies remains to be seen, however it is likely that many of the studies implicating EMT in therapy resistance have been carried out with cells that exhibit an E/M hybrid phenotype. Studies that shed further light on the reasons behind the metastatic advantages of the E/M hybrid state, their biomarker potential based on unique co-expression of epithelial and mesenchymal markers, and opportunities for specificity in therapeutic targeting, will be helpful in determining and achieving the exploitation of EMP in cancer care.

Figure Legends

Figure 1: EMP and the E/M hybrid state. Schematic representation of the Epithelial Mesenchymal Plasticity (EMP) continuum, along which E/M hybrid cells are commonly found in carcinoma tissues. Evidence is growing for the existing of stable intermediates held in check by complex molecular circuitry.

Figure 2: An epithelial-mesenchymal landscape. Epithelial and mesenchymal scores for the CCLE breast cancer cell lines (circle markers, coloured by subtype) overlaying a landscape of breast cancer tumour sample scores from TCGA (hexbin density plot) [Barretina et al., 2012, Cancer Genome Atlas, 2012]. Epithelial and mesenchymal signatures were obtained from Tan et al. for both cell lines and tumour samples separately [Tan et al., 2014]. Scores were calculated using *singscore* and ranked using 5 stable genes [Foroutan et al., 2018, Bhuva et al., 2020].

Figure 3: Hysteresis in models of EMT. (A) Left – miR-200/ZEB/SNAIL circuit. Middle – bifurcation of ZEB mRNA levels in response to protein SNAIL. In the green region, corresponding to a range of SNAIL, cells can be any of three states, E – epithelial, E/M hybrid, M – mesenchymal corresponding to a trimodal distribution as shown in the FACS (fluorescence-activated cell sorting) plot. For the orange region, corresponding to a

different range of SNAIL, cells can either by the E/M hybrid state or M state corresponding to a bimodal distribution in the FACS plot. Cells with a value of SNAIL less than in the green region adopt the epithelial state. Cells with a value of SNAIL greater than the orange region adopt the mesenchymal state. (B) Left- miR-200/ZEB/SNAIL/OVOL circuit. Middle – bifurcation of ZEB mRNA levels in response to protein SNAIL. In the yellow region cells can either be in the epithelial or E/M hybrid state. In the red region cells occupy the E/M hybrid state. FACS plots are shown for both yellow and red regions. (Adapted from Figure 5 Jolly, Boareto et al. (2015)).

Acknowledgements:

MKJ and EWT were supported by a SPARC grant awarded by MHRD, Government of India (SPARC/2018-2019/P303/SL). MEL and EWT received foundational support from the National Breast Cancer Foundation (CG-10-04). MJD is funded by Cancer Council Victoria APP1187825, Cure Brain Cancer and the National Breast Cancer Foundation (CBCNBCF-19-009), and the Betty Smyth Centenary Fellowship in Bioinformatics. SB was supported by a QUT Postgraduate Research Award scholarship.

References

1. Abouleila Y, Onidani K, Ali A, Shoji H, Kawai T, Lim CT, et al. Live single cell mass spectrometry reveals cancer-specific metabolic profiles of circulating tumor cells. Cancer Sci. 2019 Feb;110(2):697-706.
2. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014 Aug 28;158(5):1110-22.
3. Aceto N, Toner M, Maheswaran S, Haber DA. En Route to Metastasis: Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition. Trends in cancer. 2015 Sep;1(1):44-52.
4. Ahmadzadeh H, Webster MR, Behera R, Jimenez Valencia AM, Wirtz D, Weeraratna AT, et al. Modeling the two-way feedback between contractility and matrix realignment reveals a nonlinear mode of cancer cell invasion. Proc Natl Acad Sci U S A. 2017 Feb 28;114(9):E1617-e26.
5. Aiello NM, Maddipati R, Norgard RJ, Balli D, Li J, Yuan S, et al. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. Developmental cell. 2018a 2018/06/18/;45(6):681-95.e4.

6. Aiello NM, Maddipati R, Norgard RJ, Balli D, Li J, Yuan S, et al. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. Dev Cell. 2018b Jun 18;45(6):681-95 e4.
7. Andasari V, Roper RT, Swat MH, Chaplain MA. Integrating intracellular dynamics using CompuCell3D and Bionetsolver: applications to multiscale modelling of cancer cell growth and invasion. PLoS One. 2012;7(3):e33726.
8. Andrews TS, Hemberg M. False signals induced by single-cell imputation. F1000Res. 2018;7:1740.
9. Andriani F, Bertolini G, Facchinetti F, Baldoli E, Moro M, Casalini P, et al. Conversion to stem-cell state in response to microenvironmental cues is regulated by balance between epithelial and mesenchymal features in lung cancer cells. Molecular oncology. 2016 Feb;10(2):253-71.
10. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019 Feb;20(2):163-72.
11. Armstrong AJ, Marengo MS, Oltean S, Kemeny G, Bitting RL, Turnbull JD, et al. Circulating tumor cells from patients with advanced prostate and breast cancer display both epithelial and mesenchymal markers. Molecular cancer research : MCR. 2011 Aug;9(8):997-1007.
12. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9.
13. Bao X, Zhu X, Liao B, Benda C, Zhuang Q, Pei D, et al. MicroRNAs in somatic cell reprogramming. Curr Opin Cell Biol. 2013 Apr;25(2):208-14.
14. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009 Nov 5;462(7269):108-12.
15. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012 Mar 28;483(7391):603-7.
16. Barros-Silva D, Marques CJ, Henrique R, Jerónimo C. Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. Genes. 2018;9(9):429.
17. Beerling E, Seinstra D, de Wit E, Kester L, van der Velden D, Maynard C, et al. Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. Cell Rep. 2016 Mar 15;14(10):2281-8.
18. Bhatia S, Monkman J, Blick T, Pinto C, Waltham M, Nagaraj SH, et al. Interrogation of Phenotypic Plasticity between Epithelial and Mesenchymal States in Breast Cancer. J Clin Med. 2019 Jun 21;8(6).
19. Bhuva DD, Cursons J, Davis MJ. Stable gene expression for normalisation and single-sample scoring. bioRxiv. 2020:2020.05.04.077859.
20. Bierie B, Pierce SE, Kroeger C, Stover DG, Pattabiraman DR, Thiru P, et al. Integrin-beta4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. Proc Natl Acad Sci U S A. 2017a Mar 21;114(12):E2337-E46.
21. Bierie B, Pierce SE, Kroeger C, Stover DG, Pattabiraman DR, Thiru P, et al. Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. Proc Natl Acad Sci U S A. 2017b Mar 21;114(12):E2337-e46.

22. Biswas K, Jolly MK, Ghosh A. Stability and mean residence times for hybrid epithelial/mesenchymal phenotype. Phys Biol. 2019 Jan 29;16(2):025003.
23. Bocci F, Gearhart-Serna L, Boareto M, Ribeiro M, Ben-Jacob E, Devi GR, et al. Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. Proc Natl Acad Sci U S A. 2019a Jan 2;116(1):148-57.
24. Bocci F, Kumar Jolly M, Onuchic JN. A Biophysical Model Uncovers the Size Distribution of Migrating Cell Clusters across Cancer Types. Cancer Res. 2019b Nov 1;79(21):5527-35.
25. Brabletz T, Jung A, Hermann K, Gunther K, Hohenberger W, Kirchner T. Nuclear overexpression of the oncoprotein beta-catenin in colorectal cancer is localized predominantly at the invasion front. Pathol Res Pract. 1998;194(10):701-4.
26. Brabletz T, Jung A, Reu S, Porzner M, Hlubek F, Kunz-Schughart LA, et al. Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. Proc Natl Acad Sci U S A. 2001 Aug 28;98(18):10356-61.
27. Bronsert P, Enderle-Ammour K, Bader M, Timme S, Kuehs M, Csanadi A, et al. Cancer cell invasion and EMT marker expression: a three-dimensional study of the human cancer-host interface. J Pathol. 2014 Nov;234(3):410-22.
28. Burr R, Gilles C, Thompson EW, Maheswaran S. Epithelial-Mesenchymal Plasticity in Circulating Tumor Cells, the Precursors of Metastasis. Adv Exp Med Biol. 2020;1220:11-34.
29. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin Cancer Res. 2013 Jan 1;19(1):279-90.
30. Cakir A, Gonul, II, Uluoglu O. A comprehensive morphological study for basal-like breast carcinomas with comparison to nonbasal-like carcinomas. Diagn Pathol. 2012 Oct 20;7:145.
31. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct 4;490(7418):61-70.
32. Cao Z-J, Wei L, Lu S, Yang D-C, Gao G. Cell BLAST: Searching large-scale scRNA-seq databases via unbiased cell embedding. bioRxiv. 2019:587360.
33. Celia-Terrassa T, Bastian C, Liu DD, Ell B, Aiello NM, Wei Y, et al. Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. Nature communications. 2018 Nov 27;9(1):5005.
34. Celia-Terrassa T, Meca-Cortes O, Mateo F, de Paz AM, Rubio N, Arnal-Estape A, et al. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. J Clin Invest. 2012 May 1;122(5):1849-68.
35. Chakraborty P, George JT, Tripathi S, Levine H, Jolly MK. Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. Frontiers in Bioengineering and Biotechnology. 2020a 2020-March-20;8(220).
36. Chakraborty P, George JT, Tripathi S, Levine H, Jolly MK. Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. Front Bioeng Biotechnol. 2020b;8:220.
37. Chao C, Widen SG, Wood TG, Zatarain JR, Johnson P, Gajjar A, et al. Patient-derived Xenografts from Colorectal Carcinoma: A Temporal and Hierarchical Study of Murine Stromal Cell Replacement. Anticancer Res. 2017 Jul;37(7):3405-12.

38. Chapman MP, Risom T, Aswani AJ, Langer EM, Sears RC, Tomlin CJ. Modeling differentiation-state transitions linked to therapeutic escape in triple-negative breast cancer. PLoS Comput Biol. 2019 Mar;15(3):e1006840.
39. Chen W, Morabito SJ, Kessenbrock K, Enver T, Meyer KB, Teschendorff AE. Single-cell landscape in mammary epithelium reveals bipotent-like cells associated with breast cancer risk and outcome. Commun Biol. 2019;2:306.
40. Consortium. TGO. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019 Jan 8;47(D1):D330-d38.
41. Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. Nature communications. 2020 May 1;11(1):2142.
42. Crowell HL, Soneson C, Germain P-L, Calini D, Collin L, Raposo C, et al. On the discovery of population-specific state transitions from multi-sample multi-condition single-cell RNA sequencing data. bioRxiv. 2019:713412.
43. Cursons J, Leuchowius K-J, Waltham M, Tomaskovic-Crook E, Foroutan M, Bracken CP, et al. Stimulus-dependent differences in signalling regulate epithelial-mesenchymal plasticity and change the effects of drugs in breast cancer cell lines. Cell Communication and Signaling. 2015;13(1):26.
44. Cursons J, Pillman KA, Scheer KG, Gregory PA, Foroutan M, Hediyeh-Zadeh S, et al. Combinatorial Targeting by MicroRNAs Co-ordinates Post-transcriptional Control of EMT. Cell Syst. 2018 Jul 25;7(1):77-91 e7.
45. de Kanter JK, Lijnzaad P, Candelli T, Margaritis T, Holstege FCP. CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing. Nucleic Acids Res. 2019 Sep 19;47(16):e95.
46. DePasquale EAK, Schnell D, Dexheimer P, Ferchen K, Hay S, Chetal K, et al. cellHarmony: cell-level matching and holistic comparison of single-cell transcriptomes. Nucleic Acids Res. 2019 Dec 2;47(21):e138.
47. Derynck R, Weinberg RA. EMT and Cancer: More Than Meets the Eye. Dev Cell. 2019 May 6;49(3):313-16.
48. DeTomaso D, Jones MG, Subramaniam M, Ashuach T, Ye CJ, Yosef N. Functional interpretation of single cell similarity maps. Nature communications. 2019 Sep 26;10(1):4376.
49. Devaraj V, Bose B. Morphological State Transition Dynamics in EGF-Induced Epithelial to Mesenchymal Transition. J Clin Med. 2019 Jun 26;8(7).
50. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol. 2015 Mar;33(3):285-89.
51. Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, et al. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. Genome Biology. 2018 2018/03/14;19(1):31.
52. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol. 2019 Feb;20(2):69-84.
53. Du L, Yamamoto S, Burnette BL, Huang D, Gao K, Jamshidi N, et al. Transcriptome profiling reveals novel gene expression signatures and regulating transcription factors of TGFβ-induced epithelial-to-mesenchymal transition. Cancer medicine. 2016;5(8):1962-72.

54. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018 Jan 4;46(D1):D649-d55.
55. Fan J, Salathia N, Liu R, Kaeser GE, Yung YC, Herman JL, et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat Methods. 2016 Mar;13(3):241-4.
56. Font-Clos F, Zapperi S, La Porta CAM. Topography of epithelial-mesenchymal plasticity. Proc Natl Acad Sci U S A. 2018 Jun 5;115(23):5902-07.
57. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular phenotypes. BMC bioinformatics. 2018 Nov 6;19(1):404.
58. Foroutan M, Cursons J, Hediyeh-Zadeh S, Thompson EW, Davis MJ. A Transcriptional Program for Detecting TGFbeta-Induced EMT in Cancer. Molecular cancer research : MCR. 2017 May;15(5):619-31.
59. Francart ME, Lambert J, Vanwynsberghe AM, Thompson EW, Bourcy M, Polette M, et al. Epithelial-mesenchymal plasticity and circulating tumor cells: Travel companions to metastases. Dev Dyn. 2018 Mar;247(3):432-50.
60. Franssen LC, Chaplain MAJ. A mathematical multi-organ model for bidirectional epithelial-mesenchymal transitions in the metastatic spread of cancer. bioRxiv. 2019.
61. Frost HR. Variance-adjusted Mahalanobis (VAM): a fast and accurate method for cell-specific gene set scoring. bioRxiv. 2020:2020.02.18.954321.
62. Fustaino V, Presutti D, Colombo T, Cardinali B, Papoff G, Brandi R, et al. Characterization of epithelial-mesenchymal transition intermediate/hybrid phenotypes associated to resistance to EGFR inhibitors in non-small cell lung cancer cell lines. Oncotarget. 2017 Nov 28;8(61):103340-63.
63. Geistlinger L, Csaba G, Santarelli M, Ramos M, Schiffer L, Turaga N, et al. Toward a gold standard for benchmarking gene set enrichment analysis. Brief Bioinform. 2020 Feb 6.
64. George JT, Jolly MK, Xu S, Somarelli JA, Levine H. Survival Outcomes in Cancer Patients Predicted by a Partial EMT Gene Expression Scoring Metric. Cancer Res. 2017 Nov 15;77(22):6415-28.
65. Godin L, Balsat C, Van Eycke YR, Allard J, Royer C, Remmelink M, et al. A Novel Approach for Quantifying Cancer Cells Showing Hybrid Epithelial/Mesenchymal States in Large Series of Tissue Samples: Towards a New Prognostic Marker. Cancers. 2020 Apr 8;12(4).
66. Gonzalez VD, Samusik N, Chen TJ, Savig ES, Aghaeepour N, Quigley DA, et al. Commonly Occurring Cell Subsets in High-Grade Serous Ovarian Tumors Identified by Single-Cell Mass Cytometry. Cell Rep. 2018 Feb 13;22(7):1875-88.
67. Gregory PA, Bracken CP, Smith E, Bert AG, Wright JA, Roslan S, et al. An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. Mol Biol Cell. 2011 May 15;22(10):1686-98.
68. Grigore AD, Jolly MK, Jia D, Farach-Carson MC, Levine H. Tumor Budding: The Name is EMT. Partial EMT. J Clin Med. 2016 Apr 29;5(5).
69. Groger CJ, Grubinger M, Waldhor T, Vierlinger K, Mikulits W. Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression. PLoS One. 2012;7(12):e51136.
70. Grosse-Wilde A, Fouquier d'Hérouël A, McIntosh E, Ertaylan G, Skupin A, Kuestner RE, et al. Stemness of the hybrid Epithelial/Mesenchymal State in Breast Cancer and Its Association with Poor Survival. PLoS One. 2015;10(5):e0126522.

71. Guo CC, Majewski T, Zhang L, Yao H, Bondaruk J, Wang Y, et al. Dysregulation of EMT Drives the Progression to Clinically Aggressive Sarcomatoid Bladder Cancer. Cell Rep. 2019 May 7;27(6):1781-93.e4.
72. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011 Aug 19;146(4):633-44.
73. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. Cell Stem Cell. 2019 Jan 3;24(1):65-78.
74. Hari K, Sabuwala B, Subramani BV, La Porta CAM, Zapperi S, Font-Clos F, et al. Identifying inhibitors of epithelial-mesenchymal plasticity using a network topology-based approach. NPJ Syst Biol Appl. 2020 May 18;6(1):15.
75. Hendrix MJ, Seftor EA, Seftor RE, Trevor KT. Experimental co-expression of vimentin and keratin intermediate filaments in human breast cancer cells results in phenotypic interconversion and increased invasive behavior. Am J Pathol. 1997;150(2):483-95.
76. Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics. 2018 Oct 1;19(4):562-78.
77. Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, Dai X. An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. PLoS Comput Biol. 2015 Nov;11(11):e1004569.
78. Hou J, Guo C, Lyu G. Clinical significance of epithelial-mesenchymal transition typing of circulating tumour cells in colorectal cancer. Colorectal Dis. 2020 May;22(5):581-87.
79. Huang RY, Wong MK, Tan TZ, Kuay KT, Ng AH, Chung VY, et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). Cell death & disease. 2013a Nov 7;4(11):e915.
80. Huang RYJ, Wong MK, Tan TZ, Kuay KT, Ng AHC, Chung VY, et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). Cell death & disease. 2013b 11/07
07/24/received
10/03/revised
10/04/accepted;4(11):e915.
81. Jeong H, Ryu YJ, An J, Lee Y, Kim A. Epithelial-mesenchymal transition in breast cancer correlates with high histological grade and triple-negative phenotype. Histopathology. 2012 May;60(6b):E87-95.
82. Jia D, George J, Tripathi S, Kundnani D, Lu M, Hanash S, et al. Testing the gene expression classification of the EMT spectrum. Phys Biol. 2018 Dec 17.
83. Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, et al. Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. Frontiers in oncology. 2015;5:155.
84. Jolly MK, Mani SA, Levine H. Hybrid epithelial/mesenchymal phenotype(s): The 'fittest' for metastasis? Biochim Biophys Acta Rev Cancer. 2018 Dec;1870(2):151-57.

85. Jolly MK, Tripathi SC, Jia D, Mooney SM, Celiktas M, Hanash SM, et al. Stability of the hybrid epithelial/mesenchymal phenotype. Oncotarget. 2016 May 10;7(19):27067-84.
86. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27-30.
87. Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. Nature communications. 2019a 2019/12/06;10(1):5587.
88. Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. Nature communications. 2019b Dec 6;10(1):5587.
89. Karihtala P, Auvinen P, Kauppila S, Haapasaari KM, Jukkola-Vuorinen A, Soini Y. Vimentin, zeb1 and Sip1 are up-regulated in triple-negative and basal-like breast cancers: association with an aggressive tumour phenotype. Breast Cancer Res Treat. 2013 Feb;138(1):81-90.
90. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS computational biology. 2012;8(2):e1002375-e75.
91. Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. Nat Methods. 2018 May;15(5):359-62.
92. Klymkowsky MW, Savagner P. Epithelial-mesenchymal transition: a cancer researcher's conceptual friend and foe. Am J Pathol. 2009 May;174(5):1588-93.
93. Kolijn K, Verhoef EI, van Leenders GJ. Morphological and immunohistochemical identification of epithelial-to-mesenchymal transition in clinical prostate cancer. Oncotarget. 2015 Sep 15;6(27):24488-98.
94. Kroger C, Afeyan A, Mraz J, Eaton EN, Reinhardt F, Khodor YL, et al. Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. Proc Natl Acad Sci U S A. 2019 Mar 25.
95. Kumar S, Das A, Sen S. Extracellular matrix density promotes EMT by weakening cell-cell adhesions. Mol Biosyst. 2014 Apr;10(4):838-50.
96. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome Biol. 2020 Feb 7;21(1):31.
97. Leggett SE, Sim JY, Rubins JE, Neronha ZJ, Williams EK, Wong IY. Morphological single cell profiling of the epithelial-mesenchymal transition. Integrative biology : quantitative biosciences from nano to macro. 2016;8(11):1133-44.
98. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417-25.
99. Lieberman Y, Rokach L, Shay T. CaSTLe - Classification of single cells by transfer learning: Harnessing the power of publicly available single cell RNA sequencing experiments to annotate new experiments. PLoS One. 2018;13(10):e0205499.
100. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterparts. Stem cell reports. 2014 Jan 14;2(1):78-91.
101. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol. 2006 Feb;19(2):264-71.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

102. Loboda A, Nebozhyn MV, Watters JW, Buser CA, Shaw PM, Huang PS, et al. EMT is the dominant program in human colon cancer. BMC medical genomics. 2011 Jan 20;4:9.
103. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proc Natl Acad Sci U S A. 2013 Nov 5;110(45):18144-9.
104. Lun AT, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. 2016 Apr 27;17:75.
105. Ma Y, Sun S, Shang X, Keller ET, Chen M, Zhou X. Integrative differential expression and gene set enrichment analysis using summary statistics for scRNA-seq studies. Nature communications. 2020 Mar 27;11(1):1585.
106. Macaulay IC, Ponting CP, Voet T. Single-Cell Multiomics: Multiple Measurements from Single Cells. Trends Genet. 2017 Feb;33(2):155-68.
107. Mandal M, Ghosh B, Anura A, Mitra P, Pathak T, Chatterjee J. Modeling continuum of epithelial mesenchymal transition plasticity. Integr Biol (Camb). 2016 Feb;8(2):167-76.
108. Mandal M, Ghosh B, Rajput M, Chatterjee J. Impact of intercellular connectivity on epithelial mesenchymal transition plasticity. Biochim Biophys Acta Mol Cell Res. 2020 Jun 23;1867(10):118784.
109. Margaron Y, Nagai T, Guyon L, Kurzawa L, Morel A-P, Pinheiro A, et al. Biophysical properties of intermediate states of EMT outperform both epithelial and mesenchymal states. bioRxiv. 2019:797654.
110. Matte BF, Kumar A, Placone JK, Zanella VG, Martins MD, Engler AJ, et al. Matrix stiffness mechanically conditions EMT and migratory behavior of oral squamous cell carcinoma. J Cell Sci. 2019 Jan 9;132(1).
111. McFaline-Figueroa JL, Hill AJ, Qiu X, Jackson D, Shendure J, Trapnell C. A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. Nature Genetics. 2019 2019/09/01;51(9):1389-98.
112. Meyer SN, Galván JA, Zahnd S, Sokol L, Dawson H, Lugli A, et al. Co-expression of cytokeratin and vimentin in colorectal cancer highlights a subset of tumor buds and an atypical cancer-associated stroma. Hum Pathol. 2019 May;87:18-27.
113. Mizukoshi K, Okazawa Y, Haeno H, Koyama Y, Sulidan K, Komiyama H, et al. Metastatic seeding of human colon cancer cell clusters expressing the hybrid epithelial/mesenchymal state. Int J Cancer. 2020 May 1;146(9):2547-62.
114. Navas T, Kinders RJ, Lawrence SM, Ferry-Galow KV, Borgel S, Hollingshead MG, et al. Clinical Evolution of Epithelial-Mesenchymal Transition in Human Carcinomas. Cancer Res. 2020 Jan 15;80(2):304-18.
115. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. Cell. 2016 Jun 30;166(1):21-45.
116. Pankova D, Chen Y, Terajima M, Schliekelman MJ, Baird BN, Fahrenholtz M, et al. Cancer-Associated Fibroblasts Induce a Collagen Cross-link Switch in Tumor Stroma. Molecular cancer research : MCR. 2016 Mar;14(3):287-95.
117. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018a:1.

118. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018b Apr;556(7702):463-68.
119. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018c 2018/04/01;556(7702):463-68.
120. Peng DH, Ungewiss C, Tong P, Byers LA, Wang J, Canales JR, et al. ZEB1 induces LOXL2-mediated collagen stabilization and deposition in the extracellular matrix to drive lung cancer invasion and metastasis. Oncogene. 2017 Apr 6;36(14):1925-38.
121. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell. 2017 Dec 14;171(7):1611-24.e24.
122. Qiu P. Embracing the dropouts in single-cell RNA-seq analysis. Nature communications. 2020 Mar 3;11(1):1169.
123. Ramis-Conde I, Drasdo D, Anderson AR, Chaplain MA. Modeling the influence of the E-cadherin-beta-catenin pathway in cancer cell invasion: a multiscale approach. Biophys J. 2008 Jul;95(1):155-65.
124. Raviraj V, Zhang H, Chien HY, Cole L, Thompson EW, Soon L. Dormant but migratory tumour cells in desmoplastic stroma of invasive ductal carcinomas. Clin Exp Metastasis. 2012 Mar;29(3):273-92.
125. Revenu C, Gilmour D. EMT 2.0: shaping epithelia through collective migration. Curr Opin Genet Dev. 2009 Aug;19(4):338-42.
126. Risom T, Langer EM, Chapman MP, Rantala J, Fields AJ, Boniface C, et al. Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. Nature communications. 2018 Sep 19;9(1):3815.
127. Ruscetti M, Dadashian EL, Guo W, Quach B, Mulholland DJ, Park JW, et al. HDAC inhibition impedes epithelial-mesenchymal plasticity and suppresses metastatic, castration-resistant prostate cancer. Oncogene. 2016 Jul 21;35(29):3781-95.
128. Saitoh M. Involvement of partial EMT in cancer progression. J Biochem. 2018 Oct 1;164(4):257-64.
129. Sarioglu AF, Aceto N, Kojic N, Donaldson MC, Zeinali M, Hamza B, et al. A microfluidic device for label-free, physical capture of circulating tumor cell clusters. Nat Methods. 2015 Jul;12(7):685-91.
130. Sarrio D, Rodriguez-Pinilla SM, Hardisson D, Cano A, Moreno-Bueno G, Palacios J. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. Cancer Res. 2008 Feb 15;68(4):989-97.
131. Saxena K, Subbalakshmi AR, Jolly MK. Phenotypic heterogeneity in circulating tumor cells and its prognostic value in metastasis and overall survival. EBioMedicine. 2019 2019/08//;46:4-5.
132. Srivastava D, Iyer A, Kumar V, Sengupta D. CellAtlasSearch: a scalable search engine for single cells. Nucleic Acids Res. 2018 Jul 2;46(W1):W141-w47.
133. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet. 2015 Mar;16(3):133-45.
134. Steinway SN, Zanudo JG, Ding W, Rountree CB, Feith DJ, Loughran TP, Jr., et al. Network modeling of TGFbeta signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. Cancer Res. 2014 Nov 1;74(21):5963-77.

135. Steinway SN, Zanudo JGT, Michel PJ, Feith DJ, Loughran TP, Albert R. Combinatorial interventions inhibit TGFbeta-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. NPJ Syst Biol Appl. 2015;1:15014.
136. Sun Y, Wu G, Cheng KS, Chen A, Neoh KH, Chen S, et al. CTC phenotyping for a preoperative assessment of tumor metastasis and overall survival of pancreatic ductal adenocarcinoma patients. EBioMedicine. 2019 Aug;46:133-49.
137. Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY, et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO molecular medicine. 2014 Oct;6(10):1279-93.
138. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci U S A. 2010 Aug 31;107(35):15449-54.
139. Theodoropoulos PA, Polioudaki H, Agelaki S, Kallergi G, Saridaki Z, Mavroudis D, et al. Circulating tumor cells with a putative stem cell phenotype in peripheral blood of patients with breast cancer. Cancer Lett. 2010 Feb 1;288(1):99-106.
140. Thomas PA, Kirschmann DA, Cerhan JR, Folberg R, Seftor EA, Sellers TA, et al. Association between keratin and vimentin expression, malignant phenotype, and survival in postmenopausal breast cancer patients. Clin Cancer Res. 1999;5(10):2698-703.
141. Tian XJ, Zhang H, Xing J. Coupled reversible and irreversible bistable switches underlying TGFbeta-induced epithelial to mesenchymal transition. Biophys J. 2013 Aug 20;105(4):1079-89.
142. Tripathi S, Chakraborty P, Levine H, Jolly MK. A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells. PLoS Comput Biol. 2020a Feb;16(2):e1007619.
143. Tripathi S, Levine H, Jolly MK. The Physics of Cellular Decision Making During Epithelial-Mesenchymal Transition. Annu Rev Biophys. 2020b May 6;49:1-18.
144. Tung PY, Blischak JD, Hsiao CJ, Knowles DA, Burnett JE, Pritchard JK, et al. Batch effects and the effective design of single-cell gene expression studies. Scientific reports. 2017 Jan 3;7:39921.
145. Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, Xu J, et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol. 2014 Sep;32(9):926-32.
146. Wang D, Haley JD, Thompson P. Comparative gene co-expression network analysis of epithelial to mesenchymal transition reveals lung cancer progression stages. BMC Cancer. 2017 Dec 6;17(1):830.
147. Watanabe K, Villarreal-Ponce A, Sun P, Salmans ML, Fallahi M, Andersen B, et al. Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovol2 transcriptional repressor. Dev Cell. 2014 Apr 14;29(1):59-74.
148. Williams ED, Gao D, Redfern A, Thompson EW. Controversies around epithelial-mesenchymal plasticity in cancer metastasis. Nat Rev Cancer. 2019 Dec;19(12):716-32.
149. Yamamoto M, Sakane K, Tominaga K, Gotoh N, Niwa T, Kikuchi Y, et al. Intratumoral bidirectional transitions between epithelial and mesenchymal cells in triple-negative breast cancer. Cancer Sci. 2017 Jun;108(6):1210-22.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2020 doi:10.20944/preprints202008.0023.v1

150. Yamashita N, Tokunaga E, Iimori M, Inoue Y, Tanaka K, Kitao H, et al. Epithelial Paradox: Clinical Significance of Coexpression of E-cadherin and Vimentin With Regard to Invasion and Metastasis of Breast Cancer. Clin Breast Cancer. 2018 Oct;18(5):e1003-e09.
151. Yang MH, Wu MZ, Chiou SH, Chen PM, Chang SY, Liu CJ, et al. Direct regulation of TWIST by HIF-1alpha promotes metastasis. Nat Cell Biol. 2008 Mar;10(3):295-305.
152. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. science. 2013;339(6119):580-84.
153. Yuan GC, Cai L, Elowitz M, Enver T, Fan G, Guo G, et al. Challenges and emerging directions in single-cell analysis. Genome Biol. 2017 May 8;18(1):84.
154. Zacharias M, Brcic L, Eidenhammer S, Popper H. Bulk tumour cell migration in lung carcinomas might be more common than epithelial-mesenchymal transition and be differently regulated. BMC Cancer. 2018 Jul 6;18(1):717.
155. Zhang J, Tian XJ, Zhang H, Teng Y, Li R, Bai F, et al. TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Science signaling. 2014 Sep 30;7(345):ra91.
156. Zhang JM, Kamath GM, Tse DN. Valid Post-clustering Differential Analysis for Single-Cell RNA-Seq. Cell Syst. 2019 Oct 23;9(4):383-92.e6.