UNIVERSITÀ VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA INTERNAZIONALE IN MEDICINA MOLECOLARE

Curriculum in Basic and Applied Immunology and Oncology

Exploring novel mechanisms of plasticity, cell-state transitions and drug resistance in breast and colorectal cancer models

DoS: Prof. Giovanni Tonon

Second Supervisor: Prof. Iain Couzin

Tesi di DOTTORATO di RICERCA di Guido Gatti

matr. 015455

Ciclo di dottorato XXXV

SSD BIO/11 - BIOLOGIA MOLECOLARE

Anno Accademico 2021/2022

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

il sottoscritto / *I* **Guido Gatti**
matricola / *registration number* 015455
nato a / *born at* Romano di Lombardia (BG)
il / *on* 25/02/1992
autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD thesis entitled*: Uncovering new mechanisms of plasticity, cell-state transitions and drug resistance in breast and colorectal cancer models

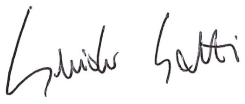
NON AUTORIZZA la consultazione della tesi per 12 mesi/ *NOT AUTHORIZES the Consultation of the thesis for 12 months*
a partire dalla data di conseguimento del titolo e precisamente / *from the PhD thesis date, specifically*
Dal / *from* 18/04/2023 Al / *to* 18/04/2024

Poiché / *because*:

□ l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità / *The whole project or part of it might be subject to patentability*;
X ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione / *Parts of the thesis have already been submitted to a publisher or are in press*;
□ la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione / *The thesis project is financed by external bodies that have rights over it and on its publication*.

È fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *It is not allowed to copy, in whole or in part, the data and the contents of the thesis*.

Data / *Date* 31/01/2023 Firma / *Signature*



DECLARATION

This thesis has been:

- composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.
- has been written according to the editing guidelines approved by the University.

Permission to use images and other material covered by copyright has been sought and obtained. For the following images (Figure 1.1, Figure 1.2, Figure 1.3, Figure 1.4, Figure 1.5, Figure 3.15, Figure 3.17 B), it was not possible to obtain permission and are therefore included in thesis under the "fair use" exception (Italian legislative Decree no. 68/2003).

All the results presented here were obtained by myself, except for:

1) **RNA sequencing**. For instance:
   RNA sequencing experiments to investigate the differential gene expression between basal and luminal subpopulations (Results, figure 3.7), which were performed in collaboration with Dr. Dejan Lazarevic, Center for Omics Sciences, San Raffaele Scientific Institute, Milan, Italy.

All sources of information are acknowledged by means of reference.

AKNOWLEDGMENTS

Guido Gatti was supported by a AIRC three-years fellowship Guglielmina Lucatello e Gino Mazzega *Fellowship for Italy*, ID 26728-2021.

Laugh, and the world laughs with you;
Weep, and you weep alone

Ella Wheeler Wilcox (1850 – 1919)

Abstract

Tumors are composed of heterogeneous mixture of cells whose mutual relationships remain unclear. Previous studies demonstrated that tumor cells exist within a population in different phenotypic states, which are kept in equilibrium through unknown mechanisms, possibly favoring cancer fitness. Exploring the different networks underlying tumor heterogeneity and the derived plasticity could usher novel avenues to target tumor growth and resistance to drugs more effectively. To gain insight into the molecular pathways underlying cell-state transition, we employed SUM149 and SUM159 triple-negative breast cancer cell lines, characterized by the reported co-existence of luminal and basal cell states, as models. We sorted each subpopulation and proved that these different phenotypic states are reinstated within one hour after isolation. This shift is promoted both by transcriptional regulation, and intriguingly by the concomitant protein relocalization of cell-state-specific markers. RNAseq analysis revealed the activation of a gene expression signature driving protein transport. In particular, in basal cells, vesicle transport and protein localization to cell periphery is deregulated, suggesting that protein transport or its regulation may exert a role in the transition velocity. To identify secreted factors that potentially regulate the equilibrium proportion of phenotypically heterogeneous cell lines, conditioned medium assays were performed in isolated cell-states. As we questioned the existence of extrinsic, cooperative mechanisms that govern cell-state proportions in an heterogenous cell environment, we further pondered the possibility of a drug-protective, cooperative mechanism that operates along the resistant-to-sensitive cell axis. To address this issue, we developed a colorectal cancer model and conducted conditioned medium and co-culture experiments, but we discovered no growth benefit when sensitive cells were exposed to enriched conditioned media obtained from resistant cells or resistant cells themselves. In conclusion, our work expanded the understanding of the dynamics of cell-state transition in cell lines, demonstrating that these models are indeed highly plastic and indicating that they are a suitable model for investigating tumor plasticity.

Table of contents

Acronyms and abbreviations **11**

List of figures **13**

1. Introduction **15**

1.1 Tumor heterogeneity 15

1.1.1 Genetic heterogeneity 15

1.1.2 Non-genetic heterogeneity 17

1.1.3 Microenvironmental heterogeneity 21

1.1.4 Heterogeneity and drug resistance 23

1.2 Cell plasticity 26

1.2.1 Plasticity during development and tissue regeneration 26

1.2.2 Epithelial-to-mesenchymal transition (EMT) 27

1.2.3 Epigenetic plasticity and cancer cell-states 31

1.3 Breast cancer and cell-state transitions 35

1.3.1 Breast cancer 35

1.3.2 Breast cancer models of stochastic cell-state transitions 36

2. Aim of the work **39**

3. Results **40**

3.1 Defining plasticity and cell-state dynamics in SUM149 and SUM159 breast cancer cell lines 40

3.1.1 SUM149 and SUM159 cell lines are characterized by the co-existence of different cell-states 40

3.1.2 SUM149 and SUM159 subpopulations return to phenotypic equilibrium over time 43

3.1.3 EPCAM rapidly shuttles from cytoplasm to cell membrane and vice versa in both purified basal and luminal cell-state 47

3.1.4 EPCAM and CD24 are under-expressed in purified basal cells, but highly re-expressed after one hour from sorting 52

3.1.5 Gene expression analysis reveals extensive difference within sorted SUM149 subpopulations 54

3.1.6 Basal cells completely revert to luminal gene expression profiles after one hour from sorting 60
3.1.7 Basal cells actively regulate gene expression of pathways related to protein transport and endocytosis 63
3.1.8 Generation of a reporter SUM149 cell line 66
3.1.9 Factors secreted in basal and luminal conditioned media are not responsible for interconversion between different cell-states 74
3.2 Defining plasticity and cell-state dynamics in other breast cancer cell lines 76
3.2.1 Transcriptional and protein heterogeneity are common features within breast cancer cell lines 76
3.2.2 BT549 cell line is characterized by the co-existence of CD24$^-$ and CD24$^+$ subpopulations displaying a plastic phenotype 84
3.3 Defining non-cell autonomous cooperative mechanisms of drug protection 88
3.3.1 Experimental design and treatment of A375 melanoma cell line 88
3.3.2 Characterization and generation of resistant colorectal cancer cells 92
3.3.3 Generation of GFP-tagged sensitive RKO cells 95
3.3.4 Co-culture of RKO and RKOR cells has no effect on sensitive cells viability upon MAPK inhibitors treatment 98
3.3.5 Experiments with conditioned media experiments did not prove the existence of protective effects from resistant cells 102
4. Discussion **109**
5. Materials and Methods **120**
5.1 Cell culture 120
5.2 Flow cytometry 120
5.3 Immunofluorescence 121
5.4 Imaging analysis 122
5.5 RNAseq data analysis 122
5.6 RNA extraction and real-time quantitative PCR 123
5.7 Conditioned medium experiments 124
5.8 Generation of SUM149 reporter cell line 125
5.8.1 Design of sgRNA 125

5.8.2 Cloning.....125
5.8.3 Design of donor sequence.....126
5.8.4 Transfection.....126
5.9 Analysis of scRNAseq dataset.....127
5.10 Cell sensitivity to drug treatment.....127
5.11 Generation of resistant cells.....128
5.12 Lentiviral infection.....128
5.13 Proliferation studies.....128
5.14 Live imaging and Incucyte®.....129
5.15 Statistical analysis.....129
References.....130

Acronyms and abbreviations

Adj. p-value	Adjusted p-value
CD24	Cluster of Differentiation 24
CD44	Cluster of Differentiation 44
CDH1	Cadherin 1
CDH2	Cadherin 2
CIN	Chromosomal Instability
CM	Conditioned Medium
CNA	Copy Number Alteration
CRC	Colorectal Cancer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSC	Cancer Stem Cell
EMP	Epithelial-to-Mesenchymal Plasticity
EMT	Epithelial-to-Mesenchymal Transition
EPCAM	Epithelial Cell Adhesion Molecule
ER	Estrogen Receptor
EZH2	Enhancer of zeste homolog 2
GFP	Green Fluorescent Protein
GO	Gene Ontology
HER2	Human Epidermal Growth Factor Receptor 2
ITH	Intratumor Heterogeneity
KI	Knock-in
Log_2FC	Log_2 Fold Change
MAPK	Mitogen-activated Protein Kinase
MAPKi	MAPK inhibitor
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
NLS	Nuclear Localization Signal
PAM	Protospacer Adjacent Motif
PCA	Principal Component Analysis
PEST	Proline - Glutamic Acid – Serine - Threonine

PR	Progesterone Receptor
qPCR	Quantitative PCR
RFP	Red Fluorescent Protein
scRNAseq	Single-cell RNA sequencing
SD	Standard Deviation
SEM	Standard Error of the Mean
sgRNA	Single Guide RNA
SNV	Single Nucleotide Variant
TCGA	The Cancer Genome Atlas
TCR	T-Cell Receptor
TK	Thymidine Kinase
TNBC	Triple Negative Breast Cancer
VIM	Vimentin
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

List of figures

Figure 1.1……25

Figure 1.2……30

Figure 1.3……33

Figure 1.4……34

Figure 1.5……38

Figure 3.1……42

Figure 3.2……45

Figure 3.3……46

Figure 3.4……49

Figure 3.5……51

Figure 3.6……53

Figure 3.7……56

Figure 3.8……58

Figure 3.9……61

Figure 3.10……64

Figure 3.11……69

Figure 3.12……70

Figure 3.13……71

Figure 3.14……73

Figure 3.15……75

Figure 3.16……78

Figure 3.17……80

Figure 3.18……81

Figure 3.19……82

Figure 3.20……83

Figure 3.21……..……..86

Figure 3.22……..……..90

Figure 3.23……..……..91

Figure 3.24……..……..93

Figure 3.25……..……..94

Figure 3.26……..……..96

Figure 3.27……..……...100

Figure 3.28……..……...105

Figure 3.29……..……...106

Figure 3.30……..……...107

1. Introduction

1.1 Tumor heterogeneity

Tumors are composed of distinct subpopulations of cells (subclones) that vary in their composition: a phenomenon known as intratumor heterogeneity (ITH). The extent of this variability stems from genetic and non-epigenetic differences and marks all ranges of the tumor phenotype, including proliferation, tumorigenicity, metastatization and resistance to drug treatment (Hausser & Alon, 2020; Marusyk *et al*, 2020). Detailed mapping of the genomic, epigenetic, and transcriptional variations between various tumors has uncovered the magnitude of this heterogeneity, and this is the case of The Cancer Genome Atlas (TCGA) (Koboldt *et al*, 2012), the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (Curtis *et al*, 2012) and a plethora of recent genomic and single-cell studies (Eirew *et al*, 2015; Kim *et al*, 2018; Li *et al*, 2017; Navin *et al*, 2011; Tirosh *et al*, 2016; Patel *et al*, 2014; Kinker *et al*, 2020; Akrap *et al*, 2016; Gambardella *et al*, 2022). These works highlighted the intrinsic complexity of cancer as well as the potential effects of heterogeneity on therapy response. Here, we provide a brief overview of the main sources of ITH: intrinsic, which entails genetic and non-genetic heterogeneity, and extrinsic, or microenvironmental.

1.1.1 Genetic heterogeneity

As a natural consequence of cell division, a few mutational events in the form of nucleotide substitutions or minor deletions are introduced across the genome each time a cell divides (Jones *et al*, 2008). Despite the high efficiency in the proofreading activity of DNA polymerases, the mechanism of DNA replication is not entirely error-free (Vogelstein *et al*, 2013; Martincorena & Campbell, 2015). It is expected that every potential point mutation would exist in at least a few cells of a tumor composed of a large number of cells. The risk of introducing a new mutation could represent a threat for the cells, but also an evolutionary

opportunity to increase their fitness. Notwithstanding, the accumulation of mutations, which poses a significant danger to the life of a cell and, at a broader level, of an organism, typically takes decades. Due to the continuous turnover of highly proliferating cells, such as in the intestinal, skin, or lung epithelia, and the enormous size of cell populations, some random mutational events inevitably affect cancer-relevant genes, causing oncogene activation and tumor suppressor deactivation (Vogelstein & Kinzler, 2004). Alternatively, cancer can be attributed to external sources. The high number of mutations found in certain types of cancer, such as melanoma and lung cancer, typically caused by exposure to UV light and tobacco carcinogens, respectively, is likely due to the long-term exposure of stem cells to these mutagens prior to the first tumorigenic step (Hecht, 1999; Shain & Bastian, 2016). Acquisition of specific genetic mutations is evidently observed in these tumors, which are the direct result of exposure to exogenous mutagenic substances.

Whole-genome (WGS) and whole-exome sequencing (WES) have demonstrated that cancer subclones can be distinguished by the nature of genetic lesions: mutations such as single-nucleotide (SNV) and structural variants, indels (small insertions and/or deletions) or copy number alterations (CNA) (Knösel *et al*, 2004; Jones *et al*, 2008; Negrini *et al*, 2010; Vogelstein *et al*, 2013). Spontaneous tumors develop through a complex process of somatic Darwinian evolution, in which the accumulation of genetic mutations (driver mutations) in the cancer-related genes, which confer a fitness advantage to tumor, is accompanied by many other genetic changes that do not appear to provide a growth advantage, in otherwise renamed passenger mutations. Indeed, according to data from high-throughput sequencing, a tumor has an average of over 10,000 somatic mutations, with about 2-8 alterations in cancer genes and a larger number of passenger mutations in other genes that may impact other cellular activities (Vogelstein *et al*, 2013). Moreover, subclonal heterogeneity is vastly increased by a plethora of genomic instability events including defects in homologous recombination, chromosome instability, and defects associated with DNA mismatch repair and APOBEC-related pathways (Stephens *et al*, 2011; Popova *et al*, 2012; Alexandrov *et al*, 2013; Roberts *et al*, 2013; McGranahan & Swanton,

2015; Kumar *et al*, 2016). For instance, large-scale chromosomal alterations have a major impact on the genome, can disrupt many genes and can be considered macro-evolutionary events that may foster tumor progression (Notta *et al*, 2016). Furthermore, chromosomal instability leading to the loss of genomic material may contribute to genomic heterogeneity (McPherson *et al*, 2016; Murugaesu *et al*, 2015).

Interestingly, different mutational events may appear at distinct stages throughout the tumorigenic pathway (de Bruin *et al*, 2014; Nik-Zainal *et al*, 2012). Mathematical modelling suggests that the clonal mutational burden in certain cancer types, such as lung cancer and melanoma, primarily reflects mutations that occurred prior to cancer initiation, whereas in other cancer types the mutation burden may be more indicative of changes that occur after cancer initiation (Tomasetti *et al*, 2013). As diverse mutational processes can arise at various stages of tumor evolution, the variability of copy-number events among various cancers is a consequence of these aberrant occurrences (Martinez *et al*, 2013). Remarkably, not all genetic changes across tumor cells lead to detectable phenotypic differences. However, it is likely that genetic lesions falling in driver genes sequences (Anderson *et al*, 2011; Gerlinger *et al*, 2012) or in drug resistance-related loci (Ruiz *et al*, 2011) may impact on clinically relevant phenotypes. Additionally, the accumulation of several passenger mutations might have minor effects on cellular phenotypes. For example, by affecting gene expression and epigenetics, mutations that are typically silent may promote the phenotypic plasticity of tumor cells. This concept will be further elaborated in the following sections. In conclusion, genetic heterogeneity is not necessarily the main cause of phenotypic heterogeneity in tumours, but it does provide a substrate for tumorigenesis, tumor evolution and resistance to treatment.

1.1.2 Non-genetic heterogeneity

Except for cells that undergo rare, random somatic mutations, each cell in the body has the same genome. Fascinatingly, non-genetic mechanisms are responsible for the wide variety of cellular and tissue differentiation. In fact,

context-dependent gene patterns influence cell fate decisions, fostering variation among genetically identical cells. Phenotypic heterogeneity in normal tissues can be separated into two categories: deterministic and stochastic (Marusyk *et al*, 2012). Deterministic heterogeneity refers to the variety of cellular phenotypes that can be linked to a specific tissue differentiation hierarchy. In particular, the intestinal crypt (Clevers, 2013) and the hematopoietic stem cells (Doulatov *et al*, 2012) are notable examples. In cancer, this theoretical framework resides on the concept of cancer stem cells (CSCs), despite decades of discussion on the topic (Jordan, 2009; Gilbertson & Graham, 2012; Batlle & Clevers, 2017; Rossi *et al*, 2020). This suggests that the diversity in the cell types within a tumor is similar to the differentiation hierarchies found in normal tissues. In fact, CSCs are a specific subset of tumor cells that have the ability to self-renew and give rise to all the other cell types found in a tumor (Lapidot *et al*, 1994; Al-Hajj *et al*, 2003; Mani *et al*, 2008). They are considered to be responsible for the development, recurrence, and spread of cancer (Batlle & Clevers, 2017). CSCs are thought to be resistant to chemotherapy and radiation, making them difficult to eliminate and leading to the possibility of tumor recurrence (Bao *et al*, 2006; Mathews *et al*, 2011; Desai *et al*, 2014). Although their existence is still a matter of debate (Magee *et al*, 2012), CSCs are considered to be one of the sources responsible for the non-genetic phenotypic heterogeneity in tumors, and their role has become widely accepted in the field of cancer research (Shackleton *et al*, 2009). Phenotypic heterogeneity caused by fluctuations in differentiation states has been found to have a significant impact on clinical outcomes such as treatment resistance and metastatic capability (Clevers, 2011). Therefore, higher stem cell marker expression in cancer cells is indicative of a tumor that is significantly more resistant to therapy and more likely to spread, resulting in fatal clinical consequences. Additionally, the distinction between CSCs and non-CSCs is not always clear. Notwithstanding, experiments conducted in a breast cancer model suggest that oncogenic transformation increases the likelihood that transformed cells will transition to a stem cell state (Chaffer *et al*, 2011). Considerably, brain tumors are one of the finest examples of malignancies fueled by CSCs (Patel *et al*, 2014; Liau *et al*, 2017; Neftel *et al*, 2019). However, phenotypic heterogeneity

cannot be totally recapitulated by the differentiation networks, as cancer cells have a far larger degree of heterogeneity in differentiation states compared to normal tissues (Landau *et al*, 2014; Jenkinson *et al*, 2017).

On the contrary, stochastic heterogeneity refers to the variations in phenotypes that occur in cells that have the same genetic makeup and are in the same deterministic state but are generated by random fluctuations in the biochemical processes within the cells, therefore affecting metabolism (Levine & Hwa, 2007), or by the occasional bursts of gene expression associated with chromatin organization (Wang *et al*, 2019; Tunnacliffe & Chubb, 2020). Transcriptional noise, which refers to the random fluctuations in the gene expression levels in individual cells within a population (Sánchez & Kondev, 2008), plays an important role in the regulation of cell-to-cell variability, and it is a conserved function throughout evolution, from bacteria to mammals (Golding *et al*, 2005; Raj *et al*, 2006). Indeed, every step on the road of the regulation of gene expression can generate noise: translation, post-translational regulation, mRNA and protein degradation (Kærn *et al*, 2005; Balázsi *et al*, 2011). Such variability in gene expression is caused by the stochastic nature of cellular processes, such as the binding of transcription factors to DNA and the consequent initiation of transcription (Coulon *et al*, 2010; Parab *et al*, 2022). This noise-driven variability is thought to play a role in phenotypic diversity in normal tissues and may be altered in cancer cells, leading to increased phenotypic diversity (Han *et al*, 2016; Domingues *et al*, 2020). For instance, it has been demonstrated how noise in the expression levels of drug-resistance-related proteins can affect rare cancer cell subpopulations survival (Shaffer *et al*, 2017). Anyway, in many scenarios noise alone is insufficient to switch between two different cell fates: noise-dependent oscillations may be sharp but transitory, requiring the involvement of mechanisms to maintain the chosen state. Systems that have two stable states that can endure for extended periods of time despite slight perturbations are referred to as bistable systems, or multistable if multiple stable states are involved (Ferrell, 2002). Importantly, bi- or multistable systems frequently display hysteresis, a form of cellular memory that is based on the persistence in cell-state achieved over time that can be induced by external or internal signals (Burrill &

Silver, 2010, 2011). Hysteresis is crucially significant because it permits the maintenance of cellular identity in heterogeneous populations. Being a stable system, hysteresis permits the cell to have long-lasting responses even after stimulus removal (Chang *et al*, 2006; Wang *et al*, 2009). These responses are often maintained by bistable feedback loops on the basis of feedback loops originated from signaling cues (Celià-Terrassa *et al*, 2018) or even epigenetic modifications (Li *et al*, 2020a; Yuan *et al*, 2020). Hysteresis is of great importance in cancer: for instance, long-term exposure to IL-1b generates EMT memory in non-small cell lung cancer via prolonged deposition of H3K27me2 and H3K9ac driven by SNAI2 (Li *et al*, 2020a) (*see also §1.2.3 Epithelial-to-mesenchymal transition (EMT) for further details on EMT).*

Cancer cells are greatly influenced by epigenetics changes and their abnormalities. For instance, Flavahan has demonstrated that disruption of chromatin insulators by loss of CTCF enables a constitutive enhancer to interact with the PDGFRA gene, hence triggering abnormal proliferative signaling (Flavahan *et al*, 2016). Moreover, hypermethylation of the DR4 promoter hinders the initiation of apoptotic pathways in glioma tumors (Elias *et al*, 2009). Another example: evasion of tumor suppressor activity, such as p16/INK4a, has been associated with promoter hypermethylation or EZH2 hyperactivation (Piunti & Shilatifard, 2016, 2021). These works clearly indicate how epigenetic signatures have a major role in defining the tumor phenotypes compared to the impact of genetic changes. Several studies have focused on the measure of DNA methylation to investigate the epigenetic variation within tumors (Mazor *et al*, 2016). Specifically, DNA methylation patterns of regulatory regions that control the expression of cancer related genes, such as promoters, displayed a significant amount of heterogeneity. For instance, high ITH is observed in specific enhancers recognized by the androgen receptor, a pivotal regulator of the survival and proliferation of prostate tumor cells (Brocks *et al*, 2014). Nowadays, single-cell multi-omics profiling of tumors can directly analyze the relationships between genetic and epigenetic alterations coupled with gene expression profiles. Granja has shown common oncogenic signatures across patients and within individual tumors despite the substantial variability and has identified the RUNX1

transcription factor as a major regulator in acute leukemias, using scATAC-seq (Granja *et al*, 2019). Additionally, in Barrett's esophagus, both DNA methylation and genetic mutations have been found to be predictors of progression to esophageal carcinoma, indicating that heterogeneity is a fundamental characteristic of tumors and that heritable epigenetic changes have a similar impact on tumor evolution as genetic mutations (Merlo *et al*, 2006). Finally, epigenetic alterations exert a crucial role in the control of plasticity in cancer cells, a topic that will be examined in further detail in the following sections (*see also §1.2.3 Epigenetic plasticity and cancer cell-states*).

1.1.3 Microenvironmental heterogeneity

Healthy cells elicit phenotypic diversity as a result of developmental processes triggered by microenvironmental signals. In fact, as cells sense and respond to cues coming from their environment, the presence of different gradients of signaling components across space and time may lead to a wide range of cellular responses (Stathopoulos & Iber, 2013). This can result in a high degree of diversity in cellular phenotypes which is reflected in the cellular responses to specific microenvironmental. Although tumor cells may develop aberrant responses such as uncontrolled proliferation and escape from apoptosis, they can also integrate and respond to normal tissue stimuli. Fascinatingly, some cancer cells can mount plastic response to adapt and even change in response to healthy microenvironment. Bussard and Smith have demonstrated how breast cancer cells are able to integrate into the normal mammary tissue and function as part of the epithelium, when co-injected alongside with mammary epithelial cells into mouse fat pads, by taking advantage of the influence of the normal cells surrounding them (Bussard & Smith, 2012). This phenomenon highlights the complexity of cancer and the potential for tumor cells to evade treatment by adapting to their microenvironment. Indeed, tumor formation is a complex process that involves changes in normal cells, as well as the development of abnormal microenvironments: tumorigenesis is not just limited to genetic and epigenetic alterations. In healthy and normal tissues, tissue architecture ensures that all cells

have access to nutrients and oxygen, and that all waste products are properly eliminated. On the contrary, the structure of blood and lymphatic vessels is completely disorganized in tumors and displays significant variations in function and localization (Fernando *et al*, 2008; Shojaei & Ferrara, 2008; Zuazo-Gaztelu & Casanovas, 2018). In addition, changes in the normal tissue architecture due to aging and chronic inflammation can lead to an increased risk of cancer and tumor progression (Fane & Weeraratna, 2020), as in melanoma (Kim *et al*, 2013) and ovarian (Lawrenson *et al*, 2010) cancer. Notably, cancer cells elicit interactions among cells and their surrounding environment that are not observed in healthy tissue (Ma *et al*, 2019; Raghavan *et al*, 2021; Barkley *et al*, 2022). These interactions contribute to the selection and diversification of cancer cell phenotypes. For example, the extracellular matrix which surrounds the cancer cells is altered compared to that of healthy tissue and this may affect tumor proliferation (Naba *et al*, 2012). Another prominent example: the acidification of the microenvironment at the edge of a tumor can promote a change in cell behavior known as epithelial-to-mesenchymal transition, which promotes invasion and metastatization in Lewis lung carcinoma (Suzuki *et al*, 2014) and in colorectal and breast cancer models (Estrella *et al*, 2013).

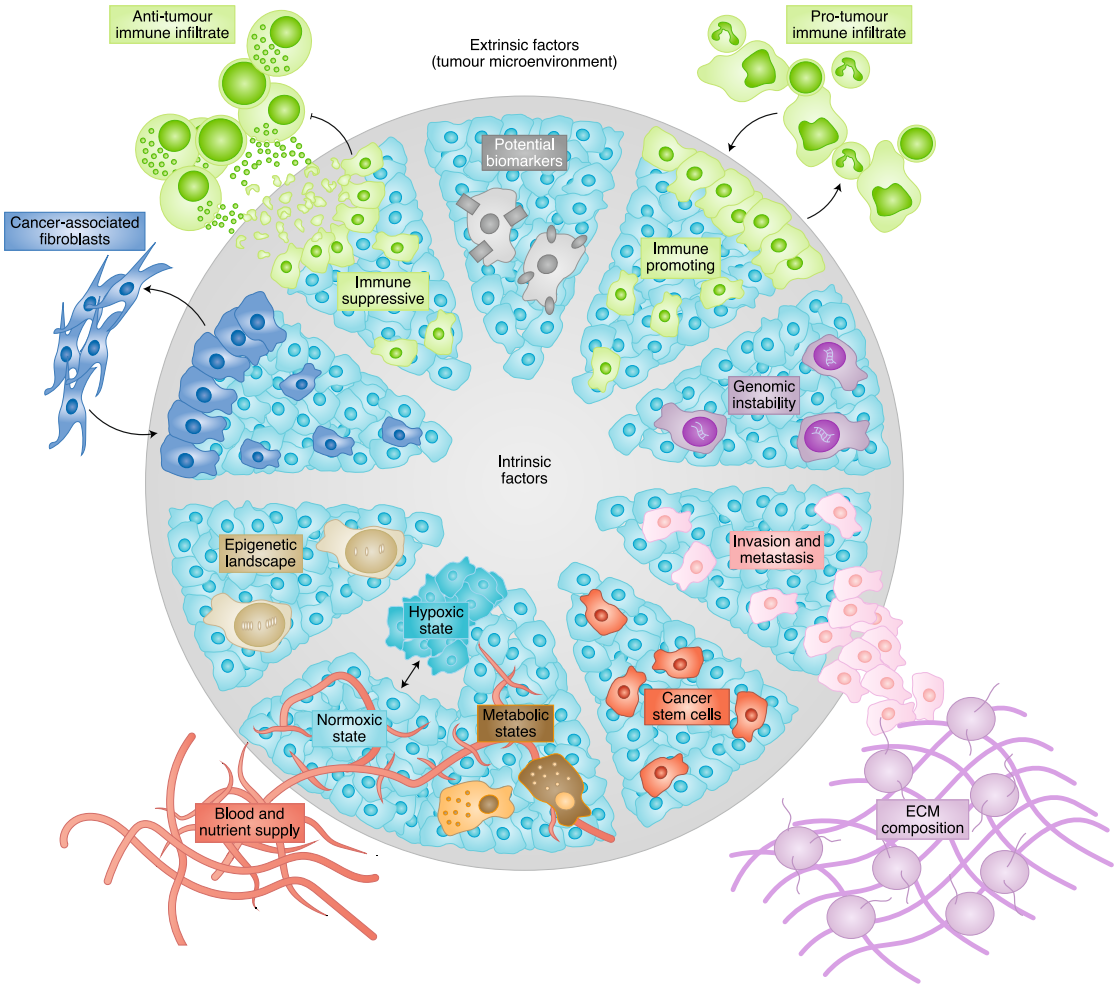
The presence of several types of immune cells, such as leukocytes, which are typically abundant in tumors, is a key source of heterogeneous environment (Del Prete *et al*, 2017). Immunological infiltration is responsible for the establishment of niches where T cells with the ability to target and kill tumor cells can shape the sub-clonal variety of cancer (Zhang *et al*, 2018). Therefore, the spatial localization of T cells harboring a subset of defined T Cell Receptors (TCR) is variable and is associated with the mutational burden within the tumor. In non-small cell lung cancer, for instance, it has been demonstrated that the frequency of amplified intratumor TCRs correlates with the amount of non-synonymous mutations existing in various locations of the tumor (TRACERx consortium *et al*, 2019). In a model of serous ovarian cancer, it was found that a higher number of infiltrating cytotoxic T cells correlated with decreased sub-clonal heterogeneity within the tumor, as a result of the immunological eradication of subpopulation of cancer cells (Zhang *et al*, 2018). In conclusion, tumor heterogeneity is driven

by distinct yet interconnected intrinsic and extrinsic mechanisms. This interconnection reflects the variability and intricacy of the tumor. For instance, it has been showed that inflammation within the tumor microenvironment is associated with increased chromosomal instability and epigenetic plasticity (Korkaya *et al*, 2011) and also triggers inflammatory signaling through the cGAS-STING antiviral pathway (Bakhoum & Cantley, 2018) which can increase the sensitivity of tumor cells to chemotherapy and enhance the effectiveness of the immune system against the tumor.

1.1.4 Heterogeneity and drug resistance

The genetic diversity within a tumor is likely to significantly impact the effectiveness of treatment. Genetic changes intended as mutations are the best-known and most extensively studied cause of resistance to treatment and contribute to both lack of initial response and relapse in various ways. For instance, cytotoxic therapy can lead to the growth of resistant cells, which may have existed prior to treatment or developed through new mutations during treatment. Some studies have shown that relapsed tumors often contain variants that were present in small numbers before therapy began (Roche-Lestienne *et al*, 2003; Mullighan *et al*, 2008; Ding *et al*, 2012). Mechanistically, mutations can hinder the ability of a drug to interact to its target. As a prototypical example, mutations occurring on the BCR-ABL fusion protein have been linked to the relapse of chronic myelogenous leukemia after treatment with imatinib (Deininger, 2005). Although genetic mechanisms of resistance are well-established, in many cases no specific genetic cause has been identified. Numerous evidence indicates that resistance can also be acquired by epigenetic modifications, such as changes in gene expression governed by chromatin structural alterations (Sharma *et al*, 2010; Knoechel *et al*, 2014; Liau *et al*, 2017; Hinohara *et al*, 2018; Marsolier *et al*, 2022). Higher KDM5B levels, for instance, correlate with greater cell-to-cell transcriptome variability, resulting in increased KDM5B activity and, subsequently, endocrine resistance (Hinohara *et al*, 2018). Nevertheless, although epigenetic-mediated drug resistance states are not

permanent, the stochastic nature of the transition and its inheritance within therapeutically relevant time periods permits the establishment of resistant clones as a result of Darwinian selection. Even when cells are genetically identical and reside in the same environment, there is a lot of variation in how individual cells respond to cytotoxic drugs, and this heterogeneous response has been linked to random fluctuations in the levels of proteins that regulate cell death (Cohen *et al*, 2008) (*see also §1.1.2 Non-genetic heterogeneity*). In addition, differences in the TME can lead to variations in the cancer cell phenotype, making some cells less responsive to treatment. For example, low oxygen levels (hypoxia) have been associated with more aggressive cancer cells traits (Ebos *et al*, 2009; Pàez-Ribes *et al*, 2009). As a result, tumor cells in areas with low oxygen levels may respond differently to therapy compared to those in areas with normal oxygen levels. Additionally, this type of resistance can be transmitted for a few cell divisions, establishing a population of resistant cells that can gain resistance through more permanent genetic or epigenetic pathways.



Adapted from Lawson et al. 2018

Figure 1.1 – Tumors are composed of distinct subpopulations of cells that vary in their composition: a phenomenon known as intratumor heterogeneity (ITH), regulated by intrinsic and extrinsic factors. Intrinsic factors consist of genomic and epigenetic instability, variations in differentiation states, including generation and propagation of cancer stem cells (CSCs), and changes in metabolic profiles. Cancer-infiltrated fibroblasts (CAFs) and/or immune cell infiltration, abnormal extracellular matrix composition (ECM), and the infiltration of vascular arteries are extrinsic factors that shape tumor heterogeneity.

1.2 Cell plasticity

Plasticity is the ability of cells to acquire new states, or phenotypes, through differentiation programs. As a fundamental characteristic of biological systems, it is a process tightly controlled by gene regulatory networks. Plasticity is pivotal for physiological developmental processes (Bedzhov & Zernicka-Goetz, 2014; Gandhi *et al*, 2020; Han *et al*, 2020), in the maintenance of adult tissue (Li *et al*, 2020b) and in wound healing (Vibert *et al*, 2018). This ability is commonly exploited by cancer cells, allowing them to acquire new features such as increased proliferation rate and the metastatic capability, which contributes to the development and progression of tumors. Notably, eukaryotic cells have evolved distinct mechanisms that foster plasticity, such as epigenetic regulation (Flavahan *et al*, 2017; Liau *et al*, 2017; Kumar *et al*, 2022), control of mRNA splicing (Li *et al*, 2018; Hall *et al*, 2022), transcriptional activation of alternative promoters (Kupr & Handschin, 2015), and translational control via 3' UTR (Jewer *et al*, 2020). Astonishingly, all cells within multicellular organisms possess the genetic information necessary to undergo plasticity, contained within their genome. In a more artificial context, the well-known Yamanaka factors (SOX2, OCT4, KLF4 and MYC) are able to mount pluripotency – a state of acquired plasticity closely matching a stem cell state (Takahashi & Yamanaka, 2006).

1.2.1 Plasticity during development and tissue regeneration

Seminal papers showed how embryonic cells transplanted to different locations within an embryo have the ability to become different types of cells along their lineage; this differentiation is driven by the local activity of morphogens, signaling molecules responsible for embryo development, a phenomenon known as competence (Solini *et al*, 2017; Chen *et al*, 2018). However, developmental stages exert a pivotal role in regulating such plasticity. An historical work published by the Spemann group demonstrated how, in amphibian embryos, the presumptive neural ectoderm taken in early stages of

gastrulation will be influenced by host environment whilst presumptive neural tissue from late gastrulation will maintain its original fate (Spemann & Mangold, 1924). Although adult tissues preserve a certain degree of plasticity, its potential decreases as differentiation progresses. Indeed, the existence of the so called adult stem cells is responsible for tissue renewal in physiological homeostasis: notable examples are the replenishment of epithelial cells in the intestinal crypts driven by a tiny pool of adult stem cells harboring LGR5 receptor (Barker *et al*, 2007), but also skin (Wang *et al*, 2022a) and hematopoietic cells (Graf, 2002). Not all adult tissues are directly derived from stem cells, therefore plasticity is not confined to stem cells alone (Potente *et al*, 2011).

Metazoans have the ability to repair and regenerate many types of tissue after injury (Addis & Epstein, 2013; Rajagopal & Stanger, 2016; Vibert *et al*, 2018; Li *et al*, 2020b; Rauner *et al*, 2021). This regenerative capability is linked to the increased plasticity of adult stem cells. For instance, when the skin is injured, the stem cells responsible for hair growth can differentiate and gain characteristics of skin stem cells while still maintaining some of their unique features (Ito *et al*, 2005). Another example is liver regeneration: hepatocytes and biliary cells can undergo epithelial lineage plasticity to replenish hepatic epithelium after tissue damage (Michalopoulos *et al*, 2005; Gadd *et al*, 2020). Variations in chromatin accessibility, loss of activity of homeostatic tissue transcription factors, and simultaneous expression of transcription factors specific to other cell lineages might partially explain the mechanisms of trans-differentiation. However, the mechanism behind plasticity will be further explored in the next sections.

1.2.2 Epithelial-to-mesenchymal transition (EMT)

Epithelial-to-mesenchymal transition (EMT) refers to a process by which epithelial cells, which form the lining of organs and glands, lose their cell-to-cell adhesion and acquire mesenchymal properties such as increased mobility and the ability to migrate through tissue. This phenomenon was originally discovered by Greenburg and Hay, when they plated both epithelial adult and embryonic cells embedded in a collagen matrix (Greenburg & Hay, 1982). They demonstrated that

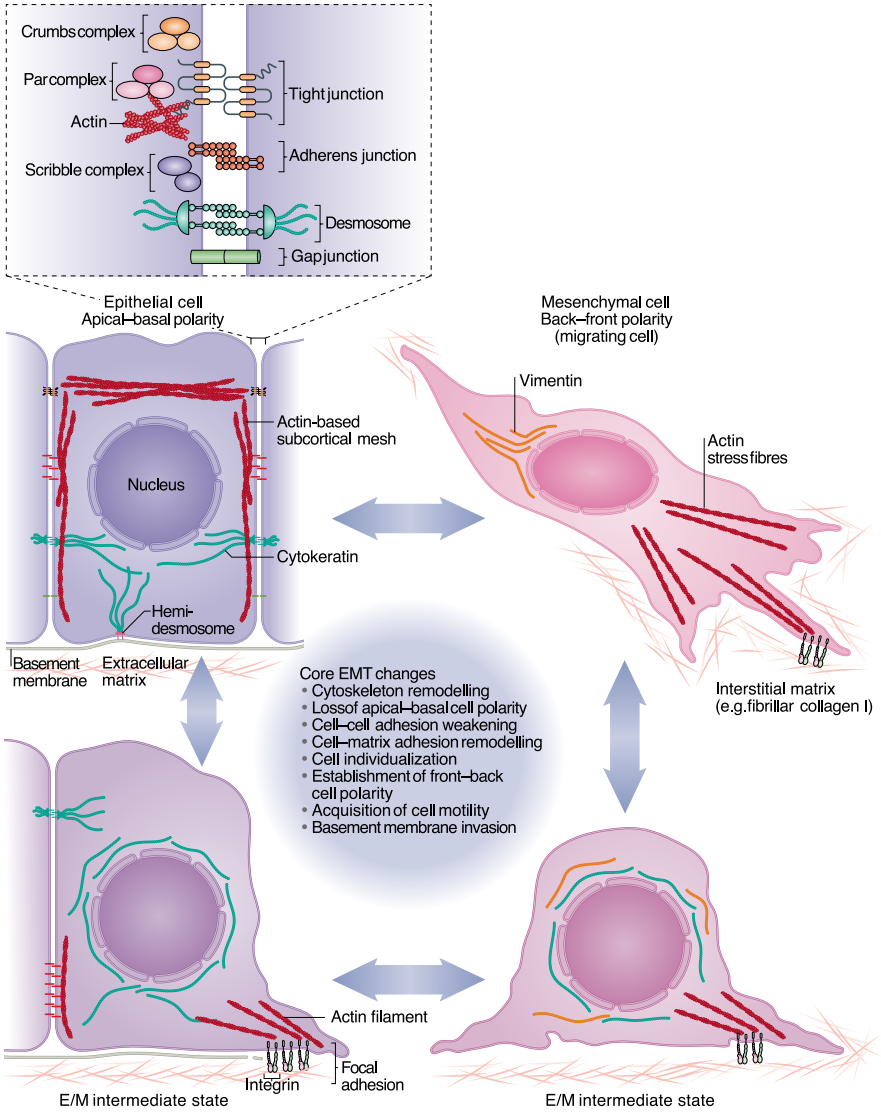
these cells undergo a transformation in which their columnar shape is lost for a more motile and migrating phenotype, like mesenchymal cells. The transition was termed epithelial-to-mesenchymal *transformation*. Subsequently, further works have showed how this phenomenon has a pivotal role during embryo development (Thiery *et al*, 2009). For instance, during metazoan development, cells that originate from epithelium frequently travel to great distances to reach their destination. To achieve this migratory state, cells must modify their gene expression patterns, reorganize their cytoskeleton, and lose their epithelial shape and structure, i.e., they must change their apico-basal polarity. Of note, during gastrulation cells undergo different steps of EMT to form endoderm and mesoderm from ectoderm (Nakaya *et al*, 2008). Another example is the differentiation of cranial neural crest cells in vertebrates, which undergo Twist1-mediated EMT to develop into mesenchymal cells (Soldatov *et al*, 2019).

EMT plays a pivotal role in cancer. In fact, during tumorigenesis, the cancerous epithelial cells may gain certain characteristics of mesenchymal cells, allowing them to invade nearby tissue and eventually spread to other tissue and organs of the body: the activation of the EMT process is responsible for the acquisition of metastatic capability in cancer cells (Ocaña *et al*, 2012; Tsai *et al*, 2012; Chaffer *et al*, 2016). Notwithstanding, cancer cells may be able to move collectively without triggering EMT, potentially taking advantage of similar mechanisms as those used during embryonic development. Notably, the formation of a new tumor in distant site from primary tumor is a consequence of the opposite process of EMT, a phenomenon known as mesenchymal-to-epithelial transition (MET) (Chao *et al*, 2010; Gunasinghe *et al*, 2012). However, cancer nature is much more complicated. In fact, cancer cells might exist in different states other than epithelial or mesenchymal, reflecting hybrid, intermediate situations and expressing a combination of EMT markers, a so called *partial* EMT (Kröger *et al*, 2019; Liao & Yang, 2020; Bornes *et al*, 2021). By generating multiple subsets of diverse clones, partial EMT considerably boosts tumor phenotypic heterogeneity. Indeed, it has been demonstrated how different subpopulations characterized with hybrid EMT signature have distinct chromatin signatures accompanied with specific gene expression patterns (Pastushenko *et*

al, 2018). Moreover, these subpopulations display distinct spatial localization within the tumor. Therefore, EMT reflects a two-way flow between the epithelial and mesenchymal axis, where cells can exist along a continuum of phenotypes (epithelial, mesenchymal, and hybrid states) (Lee *et al*, 2006; Oltean *et al*, 2006). For this reason, nowadays EMT is also known as EMP (epithelial-to-mesenchymal *plasticity*).

In tumor cells, EMT can be induced by a variety of microenvironmental cues, such as hypoxia (Suzuki *et al*, 2014; Thews & Riemann, 2019; Lin & Wu, 2020), immunological response (Jiang & Zhan, 2020) or mechanical stress (Gjorevski *et al*, 2012). In addition, epigenetic alterations might trigger EMT, since aberrant methylation patterns in CpG islands are associated with the transcriptional silencing of miR-200 miRNAs, hence driving EMT (Vrba *et al*, 2010). EMT can be induced by an intricated network of both paracrine and autocrine cues secreted by the stromal component of the tumor and also by other cancer subclones, as a sort of cell cooperative mechanisms (Scheel *et al*, 2011). Examples of these factors: TGFβ (Buonato *et al*, 2015), EGF (Hugo *et al*, 2009), FGF (Kurimoto *et al*, 2016), IL-6 (Miao *et al*, 2014), WNT (Jiang *et al*, 2007), Hedgehog (Yoo *et al*, 2011) and Notch (Yuan *et al*, 2014). Finally, mutational events and alteration in epigenetics marks may prompt EMT (Serrano-Gomez *et al*, 2016; Zhang *et al*, 2022).

Mechanistically, EMT is characterized by reduced cell-to-cell contact through downregulation of E-Cadherin (CDH1). Notably, a subset of E-Cadherin repressors directly binds to *CDH1* promoter, among them: the ZEB family (ZEB1 and ZEB2), SNAI1, TWIST1 and KLF8 (Kang & Massagué, 2004; Peinado *et al*, 2004, 2007). Interestingly, these proteins interact with each other in a complex fashion, with SNAI1 upregulating SNAI2 and TWIST1 (Thuault *et al*, 2008; Smit *et al*, 2009), and SNAI1 and TWIST1 inducing ZEB1 and SNAI2 (Dave *et al*, 2011; Casas *et al*, 2011). In addition, SNAI2 is an inducer of ZEB2 (Thuault *et al*, 2008). Notwithstanding, these proteins not only promote mesenchymal differentiation by repressing cell-to-cell adhesion, but also modulate other cellular programs involved in proliferation, cell survival, and cell adhesion (Barrallo-Gimeno & Nieto, 2005).



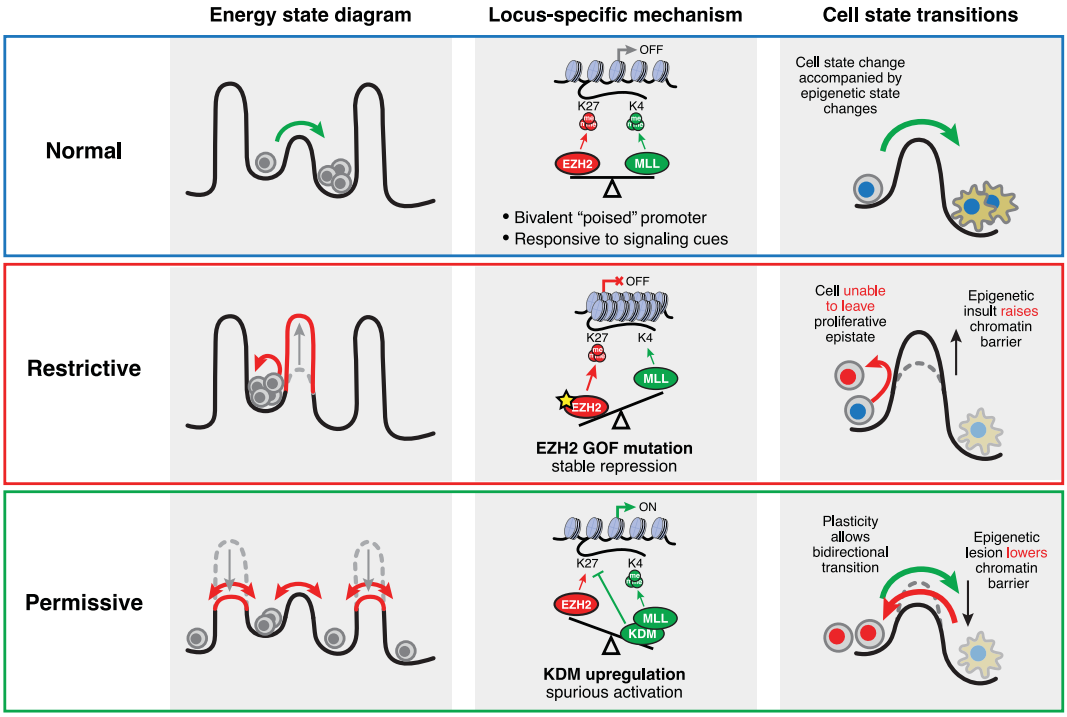
Adapted from Yang et al. 2020

Figure 1.2 – Cells in different EMT states can exhibit traits related to epithelial or mesenchymal state, or a combination of both. Epithelial cells are characterized by the presence of adherens junctions, desmosomes, gap junctions and tight junctions which mediate cell-to-cell adhesion. Polarity complexes, represented here with Par, Crumbs and Scribble complexes define the apico-basal polarity of epithelium. In contrast, mesenchymal cells lack functional epithelial junctions. These cells have vimentin and use integrin-based focal adhesions to connect with the ECM. The loss or gain of E/M traits drives cells towards plastic intermediate states.

1.2.3 Epigenetic plasticity and cancer cell-states

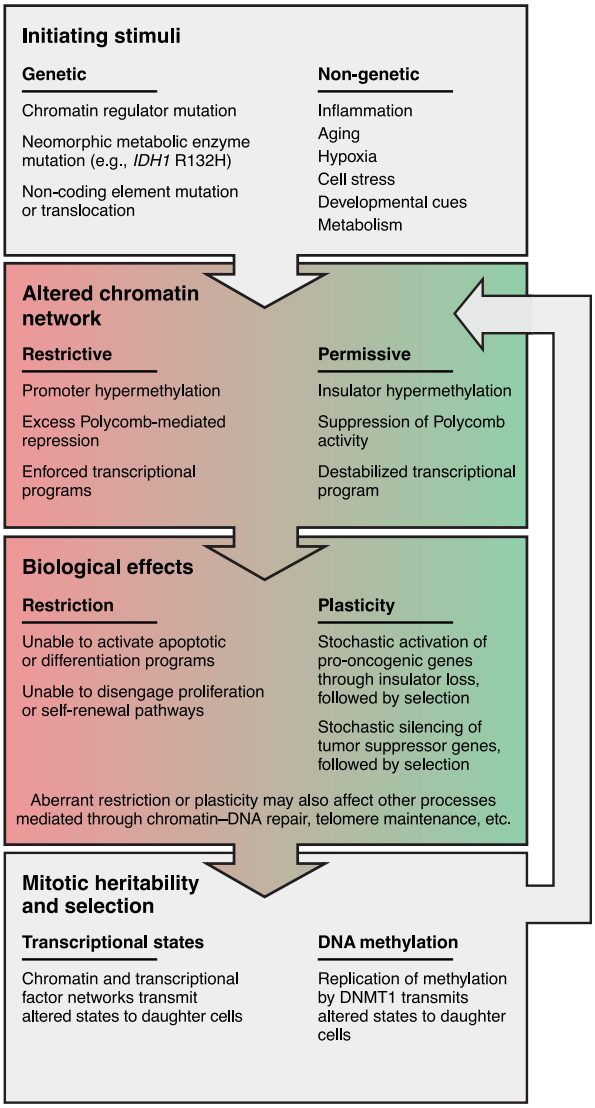
As stated in the previous sections, while cancer is often regarded as a genetic disease, epigenetic changes have a significant and widespread impact. Indeed, cancer cells display genome-wide modifications in DNA methylation patterns, chromatin architecture, and the activation of regulatory elements (Esteller, 2008). In 1957, Waddington envisioned developmental specification as an epigenetic landscape with differentiating cells moving down branching pathways separated by walls and canals that determine lineage and cell identity (Waddington, 1957). Within this model, valleys represent stable cell-states, *attractors*, whereas hills and slopes direct cells towards distinct cell-states. Nowadays, it is established that transcription factors and their networks are the primary determinants of cell fates (Zaret & Mango, 2016; Bradner *et al*, 2017), thereby shaping the topology of the canals. In this context, chromatin modifiers sustain and strengthen cell fates. In Waddington's landscape homology, chromatin structure and its regulators determine the height of the walls separating pathways and prevent cells from changing states. In healthy tissues and cells, chromatin architecture is maintained in a homeostatic equilibrium by epigenetic regulators: for example, the Polycomb repressive complex regulates repressive marks such as H3K27me3 and therefore silences lineage-specific genes with inappropriate activity (Comet *et al*, 2016). Therefore, chromatin modifications reinforce cell identity by limiting changes in gene activity and raising the barriers between different cell states. There are situations in which rapid turnover of nucleosomes impedes the formation of stable repressive structures, leaving many genes involved in development in a *bivalent* state with both active and repressive histone marks (e.g. poised genes) (Dekker & Misteli, 2015). This is the case of embryonic stem cells, which become more committed as soon as their chromatin moves to a more restricted state (Bernstein *et al*, 2006; Perino & Veenstra, 2016). These findings are consistent with a model in which chromatin is inherently resistant to encountering changes, to preserve a homeostatic equilibrium at each stage of development, and disruptions of the homeostatic equilibrium may play a role in the development of cancer. In cancer, different genetic stimuli, such as mutations

in genes encoding for Polycomb repressive complex, Trithorax activating complex or DNA methyltransferase (Piunti & Shilatifard, 2016); or non-genetic stimuli, such as inflammation (Das *et al*, 2021), aging (Wang *et al*, 2022b), cell stress (Johnstone & Baylin, 2010) can disrupt chromatin homeostasis. EZH2 mutations, which have been identified as gain-of-function mutations in numerous lymphoma and melanoma types, are one example (Kim & Roberts, 2016). Fascinatingly, certain stimuli induce more permissive conditions which can be perceived as a lowering activity over the Waddington hills in the epigenetic landscape, thereby promoting cell permanence in specific attractors, dubbed metastable cancer cell states. A permissive condition, for example, may promote stochastic epigenetic modifications that affect cancer cells fitness, such as the inactivation of tumor suppressor genes. For instance, missense mutation in histone H3 on K27 (H3.3 variant) strongly inhibits EZH2 function in pediatric brain cancer (Bender *et al*, 2013; Lewis *et al*, 2013). Therefore, suppression of Polycomb activity by EZH2 inactivation may create an overly permissive chromatin state. In addition, dynamic chromatin behavior permits premalignant or cancerous cells to explore cancer cell-states and developmental pathways, favoring plasticity within subclones. In lung and melanoma models, for example, KDM5A-mediated H3K4 demethylation promotes resistance to antiproliferative treatments by allowing cancer cells to switch to a slow-growing persister state (Roesch *et al*, 2010; Sharma *et al*, 2010). On the same way, KDM6-mediated H3K27 demethylation mediates the reversible transition to a more primitive developmental, persister state in glioblastoma stem cells in response to drug pressure (Liau *et al*, 2017). In conclusion, understanding epigenetic plasticity is of utmost importance: to design more effective therapies, we require a deeper understanding of how cancer cells in various cell-states, driven by plastic chromatin dynamics, interact to cooperate or compete, resulting in enhanced tumor fitness.



Adapted from Flavahan et al. 2017

Figure 1.3 – In healthy cells, chromatin architecture is maintained in a homeostatic equilibrium: cells respond appropriately to signaling inputs and manage epigenetic state changes with precision. In contrast, cells with disrupted chromatin networks are incapable of responding adequately to these stimuli. For example, restrictive chromatin heightens epigenetic barriers that impede cell state transitions. On the contrary, permissive chromatin reduces epigenetic barriers, hence enabling plastic behaviors and bidirectional transition among several cell states. As an example, H3K27 methyltransferase EZH2 and H3K4 methyltransferase MLL are shown.



Adapted from Flavahan et al. 2017

Figure 1.4 – Genetic or non-genetic stressors can disrupt chromatin homeostasis. These triggers may result in an excessively permissive or excessively restrictive chromatin network. For instance, stochastic epigenetic modifications may be enabled by permissive conditions, allowing for the inactivation of tumor suppressor genes. On the contrary, more restrictive conditions may affect the capability to activate apoptotic pathway. In addition, epigenetic modifications may be transmissible to cell progeny.

1.3 Breast cancer and cell-state transitions

1.3.1 Breast cancer

Breast cancer is the most common diagnosed female cancer, with an estimated 2.26 million new cases in 2020, and 684,996 estimated deaths (World Cancer Research Fund International). Breast cancer is an extremely heterogeneous disease with a high degree of intra- and intertumor heterogeneity (Keller *et al*, 2010; Marusyk *et al*, 2012; Chung *et al*, 2017; Barzgar Barough *et al*, 2021; Gambardella *et al*, 2022). Different subtypes of breast cancer have been identified and categorized based on their unique genetic and phenotypic features. In particular, the receptor-based classification consists of the evaluation of hormone receptors expression Estrogen Receptor (ER) and Progesterone Receptor (PR), coupled with Epidermal Growth Factor Receptor 2 (HER2/Neu) expression, as well as additional genomic and transcriptomic profiling. This classification identifies four subtypes: Luminal A (ER+, PR+, HER2–) which comprises 50% of cases; Luminal B (ER+, PR+ some cases, HER–) 15% of cases; HER2 (ER+ some cases, PR+ some cases, HER2 amplified) 20% of cases; triple negative breast cancer (TNBC) (ER–, PR–, HER2–) with 15% of cases, encompassing the subtype with better prognosis (Luminal A) to worse prognosis (TNBC) (Bauer *et al*, 2010; Eliyatkin *et al*, 2015). In addition, molecular profiling performed in the 2000s further expanded breast cancer subclassifications, by identifying: Basal-like, Luminal A, Luminal B, Luminal ER– AR+ (Androgen Receptor), HER2 amplified, Normal breast-like, Claudin-low (Perou *et al*, 2000; Prat & Perou, 2011; Koboldt *et al*, 2012). These categories have a substantial impact on prognosis, since each subtype varies in terms of survival and growth rates, relapse probability, and metastatic potential (Sørlie *et al*, 2001; Cheang *et al*, 2008; Smid *et al*, 2008; Kennecke *et al*, 2010; Sisti *et al*, 2016).

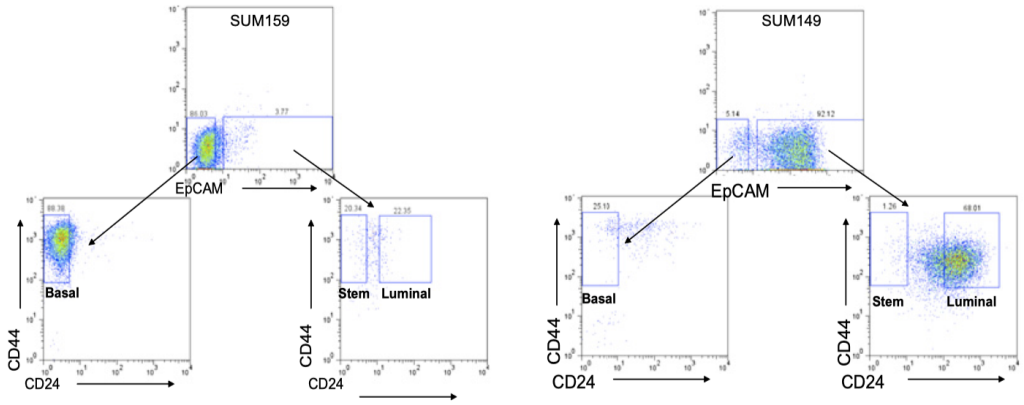
1.3.2 Breast cancer models of stochastic cell-state transitions

(Gupta *et al*, 2011) demonstrated that two triple negative breast cancer cell lines, SUM149 and SUM159, frequently transition between different cell-states, and this has also been observed in other types of cancer (Fillmore & Kuperwasser, 2008; Sacchetti *et al*, 2021). These transitions are stochastic, meaning they occur randomly within the cells. These authors found that after isolating particular differentiation states, namely basal, luminal, and stem-like, pure subpopulations of cells return within six days to the original frequencies of cell-states observed in the parental population, at their steady-state equilibrium. Differentiation states were isolated via cell sorting upon expression of specific cell surface markers (EPCAM, CD24, CD44). Intriguingly, the timeframe needed – merely a few days – shows that this trait is independent of the differential proliferation rate of each cell-state. On the same length, a cell-intrinsic mechanism that regulates such interconversion may be hypothesized. In addition, the use of cell lines reveals that the tumor microenvironment is not a contributing factor to restore equilibrium proportion. (Gupta *et al*, 2011) focused on the generation of a mathematical Markov model to describe and predict biological transitions of purified SUM149 and SUM159 cells. Mathematically, the Markov model assumes that the probability of a cell transitioning to a different cell-state is dependent only on the cell's current state. Therefore, the equilibrium of state frequencies, which can be predicted, is determined only by these transition rates and not by any other factors. When looking at the variety of cells within a tumor it's commonly thought that the co-existence of multiple cellular states is either caused by the cells working together or by the tumor evolving. However, if the cells are in a constant state of transition, the presence of multiple cell-states is a natural outcome and does not require a selective process to occur. Captivatingly, there is growing evidence that the main cellular states that are observed in tumors are also present in some cell lines, such as in liver and non-small cell lung cancer models (Wang *et al*, 2018b; Hayford *et al*, 2021), confirming that stochastic transitions are not a specific trait of 2D-cultures, but also present within tumors, despite the driving role of genetics and microenvironmental factors that boost

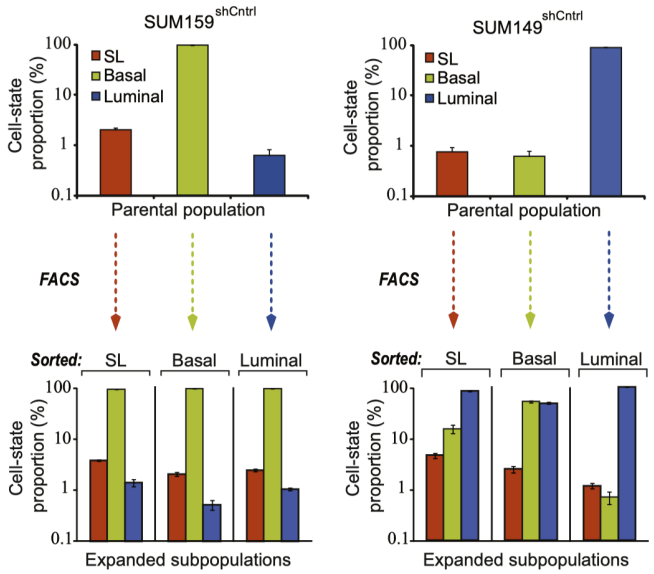
cancer heterogeneity. These results suggest that stochastic transitions are a common characteristic of cancers, and that they may be equally essential in promoting tumor heterogeneity as genetic, non-genetic and microenvironmental forces. Clinical interventions, such as chemo- or targeted therapy, might have a role in influencing the transition rates between cellular states. Hopefully, these treatments may be utilized to target an aggressive cellular state over a less aggressive one, or even better, to halt the processes behind the switch.

Despite (Gupta *et al*, 2011) remarkable work, it is not yet known how the proportions of distinct cell states are maintained within a population or how the transitions occur so rapidly. With these premises, my doctoral thesis work offers an attempt to address the aforementioned issues.

A



B



Adapted from Gupta et al. 2011

Figure 1.5 – **A** Flow cytometry pseudocolor plots reporting the FACS strategy used by Gupta to isolate basal, stem-like and luminal cells in SUM159 and SUM149 cell lines. **B** Barplots from FACS analyses performed by Gupta, on expanded subpopulations confirming the return to phenotypic equilibrium within six days in SUM159 and SUM149 cell lines.

2. Aim of the work

Tumors are heterogeneous entities comprised of distinct cancer subpopulations that co-exist in a variety of phenotypic states, driving proliferation, metastatic capability, and therapeutic resistance. Initially, genetic variability resulting from somatic mutations was assumed to be the driving force behind tumor heterogeneity. However, epigenetic changes have been recognized as one of the key contributors of tumor heterogeneity by promoting phenotypic plasticity, which is responsible for invasion and treatment resistance. In fact, many cancer cells have the potential to undergo transient and reversible alterations in morphology and functions. As these phenotypically diverse populations usually coexist in balance inside a tumor, the mechanisms that govern the proportions of various cell states remain unknown. For this reason, a comprehensive understanding of how tumors govern cancer cell-states is of essential importance and might help the design of more effective treatments. We used two triple-negative breast cancer cell lines as a model for phenotypic plasticity, and BRAF mutant colon cancer cell lines as a model for paracrine, drug-protective mechanisms. Specifically, we sought to address the following issues:

1. Identify the mechanisms underlying the sharp interconversion between cell-states upon perturbations, by time-course experiments and gene expression analysis.
2. Determine the mechanisms behind the maintenance of cell-states equilibrium proportion over time, by assessing the role of paracrine signaling.

3. Results

3.1 Defining plasticity and cell-state dynamics in SUM149 and SUM159 breast cancer cell lines

3.1.1 SUM149 and SUM159 cell lines are characterized by the co-existence of different cell-states

To explore and dissect cell-state dynamics in cancer cells, we took advantage of the system described in the Introduction and in (Gupta *et al*, 2011). Specifically, the SUM149 and SUM159 cell lines present cell subpopulations featuring specific cell-surface markers, which define specific phenotypic states. Once these subpopulations are isolated, they are able to return towards equilibrium proportions within a few days of culture (**Figure 3.1 A**). SUM149 cells have an epithelial appearance, characterized by cobblestone morphology, and tend to attach tightly one to another (**Figure 3.1 B**). On the contrary, SUM159 cells have a spindle-like morphology, resembling more fibroblast-like cells, and are loosely attached. Gupta and others (Shipitsin *et al*, 2007; Fillmore & Kuperwasser, 2008; Fillmore *et al*, 2010) have demonstrated that SUM149 cells express defined membrane markers, such as EPCAM (Epithelial Cell Adhesion Molecule) and CD24 (Cluster of Differentiation 24), as confirmed by our immunofluorescence experiments (**Figure 3.1 B**). EPCAM is a membrane glycoprotein and a known epithelial marker which is broadly expressed in carcinomas. CD24 is another glycoprotein overexpressed in almost 70% of human cancers (Kristiansen *et al*, 2004) and is a marker of poor prognosis in cancer patients (Wang *et al*, 2017; Kwon *et al*, 2015). Conversely, SUM159 cells do not express EPCAM and CD24 on the cell surface. We first sought to confirm the existence of the three distinct subpopulations, namely stem-like, luminal and basal cells, also in our cell lines, identified by the differential expression of EPCAM, CD24 and CD44. We employed antibodies specific to these cell-surface markers and performed flow cytometry on the bulk population. In SUM149, we were able to define a preponderance of $EPCAM^+$ cells, in accordance with the solid epithelial phenotype reflected in 2D cell culture (**Figure 3.1 C**). A small

subset of $EPCAM^{+}$ $CD24^{-}$ $CD44^{high}$ cells is usually described as a stem-like population, while the majority of cells ($EPCAM^{+}$ $CD24^{+}$ $CD44^{high}$ cells) are typically indicated as luminal cells. SUM149 cells also express a small $EPCAM^{-}$ and $CD24^{-}$ population, which is defined basal. In our hands, basal, stem-like and luminal proportions were 3.36%, 1.54% and 95.1%, respectively. On the contrary, SUM159 cells presented for the most part a basal phenotype ($EPCAM^{-}$ $CD24^{-}$ $CD44^{high}$), with only minor subpopulations that were luminal ($EPCAM^{+}$ $CD24^{+}$ $CD44^{high}$) and stem-like ($EPCAM^{+}$ $CD24^{-}$ $CD44^{high}$) (**Figure 3.1 D**). In SUM159, the proportions of basal, stem-like and luminal cells were 99.65%, 0.20% and 0.15%, respectively. All these observations were in accordance with the previous reports (Gupta *et al*, 2011) (see **Figure 1.5 A**).

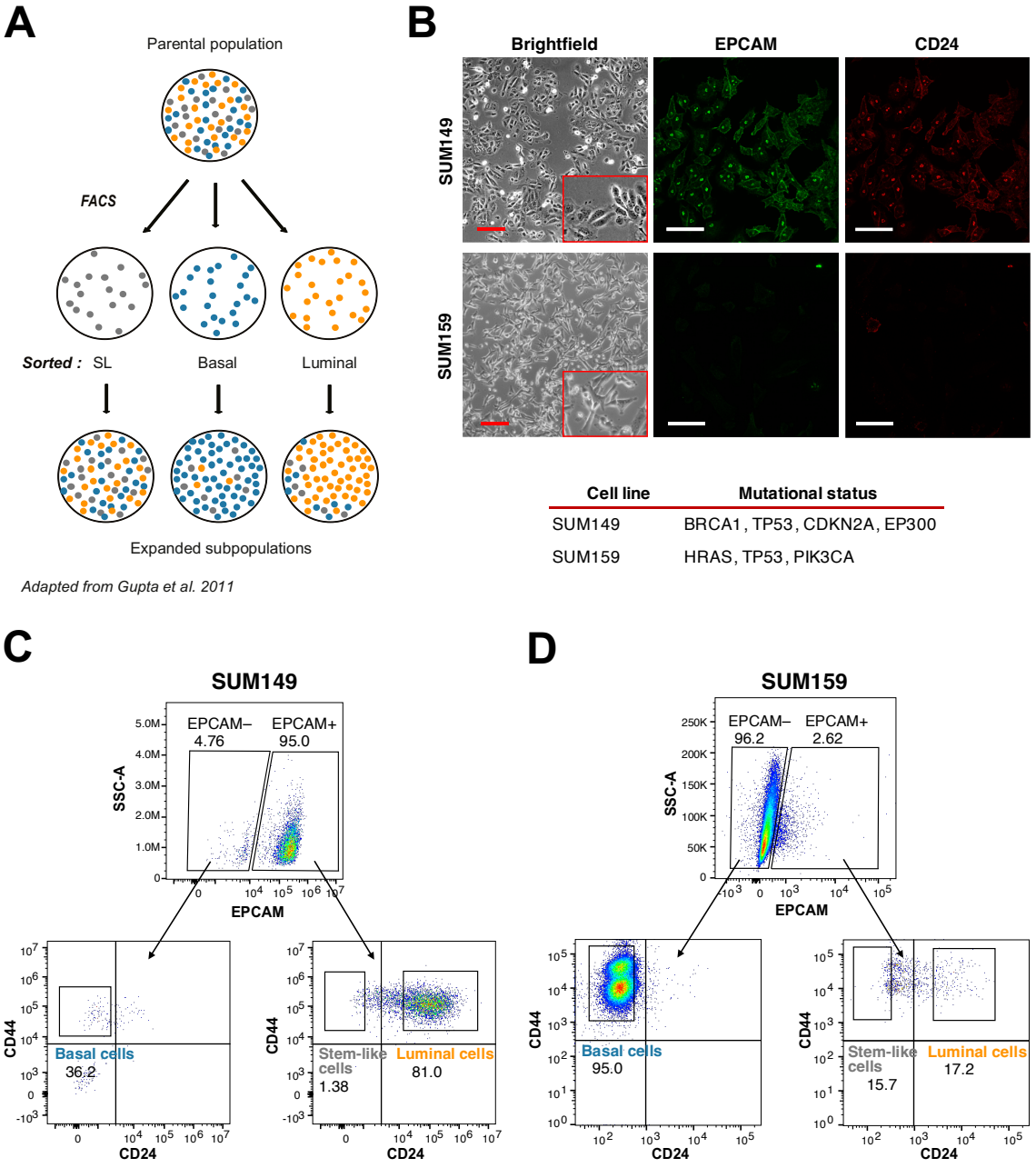


Figure 3.1 – **A** Schematic of experimental procedure from Gupta et al. 2011. See text for further details. **B** Brightfield (with 2X magnification in the corner) and immunofluorescence images showing EPCAM and CD24 staining in SUM149 and SUM159 cells in culture. Scalebar 100 µm. Below a table reporting the mutational status of the two cell lines used. **C-D** Representative flow cytometry pseudocolor plots reporting the FACS strategy used to isolate basal, stem-like and luminal cells in SUM149 (C) and SUM159 (D).

3.1.2 SUM149 and SUM159 subpopulations return to phenotypic equilibrium over time

To assess the plasticity of SUM149 and SUM159 cell lines and the ability of the respective subpopulations to transit between different cell-states, we isolated the three subpopulations in each cell line. To this end, we performed flow cytometry cell sorting and isolated stem-like, basal and luminal cells (**Figure 3.2 A**). Then, we plated each subpopulation and followed the evolution over time by expanding each differentiation state for six days in culture. We measured the cell-state proportions at the end of the timepoint using FACS analysis on expanded subpopulations. As reported in (Gupta *et al*, 2011), we confirmed that each isolated cell-state reinstated the original proportion. This observation is valid for both SUM149 and SUM159 cell lines (**Figure 3.2 B-C**). As before, all these observations were in accordance with previous reports (Gupta *et al*, 2011) (see **Figure 1.5 B**). Intriguingly, Gupta described an *in silico* Markov model that predicted a sharp return of each differentiated states towards equilibrium proportions within 24 hours from sorting in SUM159. Little if no further biological evidence was made after that observation. Therefore, we then sought to further define the transition times: did the shift really needed several days to occur, or even a few hours were enough for the transition to occur? We thus purified subpopulations by FACS and followed the emergence of either luminal or basal cells. In this series of experiments, we focused on the SUM149 cell line, whose cells are mostly luminal, due to a higher number of recovered cells from sorting than SUM159. Basal, luminal and stem-like cell subpopulations were sorted and then grown. To prevent cells from differentiating immediately after isolation, all FACS sorting experiments were performed at 4° C. At different time points after sorting – all within 24 hours (0, 3, 6, 17 and 20 hours) – we assessed whether the various subpopulations were shifting, and in which proportions, by staining cells with EPCAM, CD24 and CD44 antibodies (**Figure 3.3 A**). Strikingly, the time course experiment revealed that purified basal cells restored the original bulk population proportions within 3 hours (**Figure 3.3 B**). Indeed, the percentage of basal cells dropped within 3 hours post sorting, while a sharp increase in the number of both stem-like and, more prominently, luminal cell-

state was evident. Cell state proportions were then maintained over time, as in unsorted SUM149 cells. In contrast, purified SUM149 luminal cells are more stable, as they tend to reach equilibrium proportion through the emergence of basal and stem-like cells within 24 hours but with a slower dynamic compared to basal cells (**Figure 3.3 C**). Indeed, as in SUM149 parental population the percentage of cells in the luminal state is maintained at ~ 95 ± 0.5 %, the transition from 100 % to 95 % can be achieved quickly without a dramatic decrease in cell-state proportions, as observed in basal cells, on the contrary. Taken together, these results indicate that subpopulations within cancer cell lines are highly plastic and tend to transition from one cell-state to another, in order to reinstate the original phenotypic proportions once they are isolated and placed in culture.

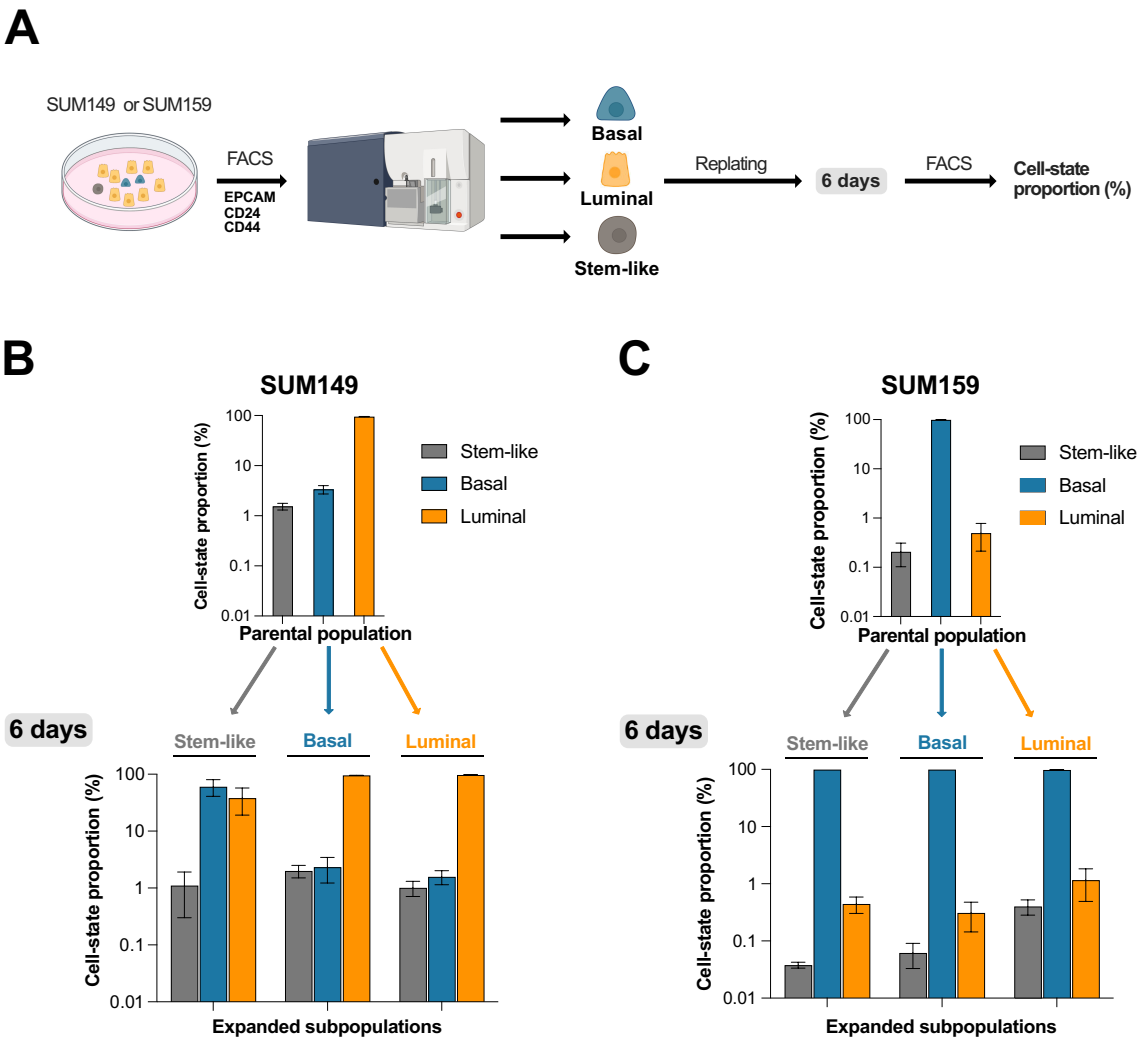


Figure 3.2 – A Schematic of experimental procedure. See text for further details. **B-C** Barplots from FACS analysis on expanded subpopulations confirming the return to phenotypic proportion in SUM149 (B) and SUM159 (C), represented with Standard Deviation (SD), n = 3.

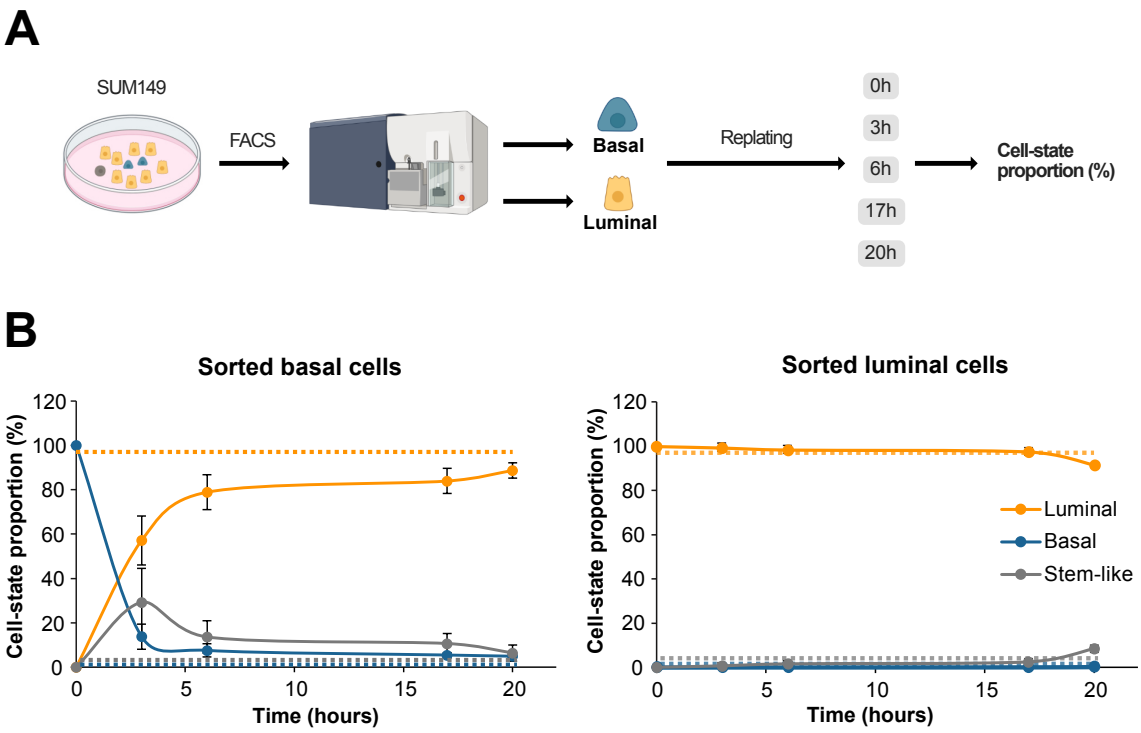
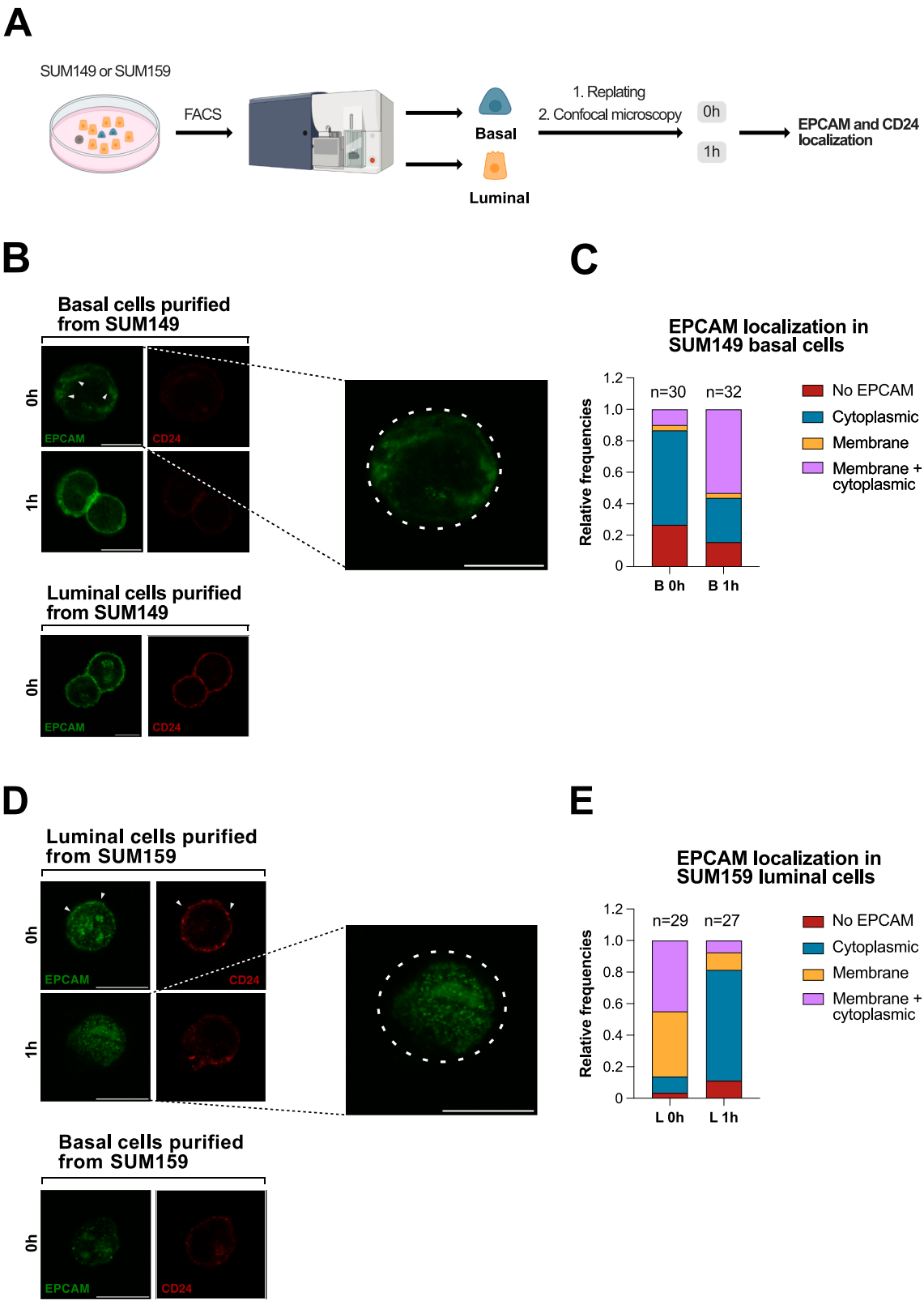


Figure 3.3 – A Schematic of experimental procedure. See text for further details. **B** FACS analyses from time course experiments on sorted subpopulations showing the kinetics of the return to original cell-state proportions in basal (*left*) and luminal (right) cells. Dotted lines represent percentage of steady state levels of luminal, basal and stem-like cells in unsorted, parental cell line. Data are represented with Standard Deviation (SD), n = 3.

3.1.3 EPCAM rapidly shuttles from cytoplasm to cell membrane and vice versa in both purified basal and luminal cell-state

Prompted by these results, we performed immunofluorescence experiments by purifying SUM149 and SUM159 cells with EPCAM and CD24 antibodies (**Figure 3.4 A**). Then, we plated cells and assessed the status of EPCAM and CD24 by confocal microscopy one hour after sorting. In both SUM149 and SUM159, basal cells lack EPCAM and CD24 on the cellular membrane, as we have previously determined with FACS analysis. Immunostaining confirmed that EPCAM is not localized on the cell surface in basal cells taken immediately after sorting (0 hour) (**Figure 3.4 B**). However, when we permeabilized SUM149 cells, we found that EPCAM remained present in these cells, although not on the surface. Strikingly, we found a strong cytoplasmatic signal for EPCAM. This result suggests that in basal cells EPCAM remains expressed at the protein level, re-localizing to the cytoplasm. In basal cells transitioning towards the luminal state, EPCAM moves from the cytoplasm to the cell membrane, within one hour after sorting. Indeed, the percentage of basal cells with surface EPCAM increases from ± 13% to ± 56% after one hour from isolation (**Figure 3.4 C**). We performed the same experiments in SUM159 cells, where cells are almost entirely in the basal state. Also, in SUM159 the isolated luminal cells, which are present in a small proportion, transitioned very quickly, again within one hour, from luminal to basal cell-state (**Figure 3.4 D**). Accordingly, we confirmed that the number of luminal cells with surface-bound EPCAM dropped dramatically from ± 86% to ± 18% (**Figure 3.4 E**). Noteworthy, the kinetics of CD24 localization is clearly slower compared to EPCAM protein, which translocates rapidly from one compartment to another. In fact, CD24 is almost under-expressed in SUM149 basal cells after one hour from sorting. On the contrary, in SUM159 luminal cells, CD24 membrane localization seems to be lost more quickly after one hour (**Figure 3.4 B-D**). Both cell lines grow in adhesion. Upon isolation of the populations by FACS, cells promptly adhere to the plate. Therefore, we asked whether the re-localization of EPCAM protein is simply the result of the adhesion to the plate.

Therefore, we purified SUM149 basal and luminal cells with FACS and plated each cell-state in two different kinds of dishes immediately after sorting. In the first dish, cells were left to adhere for one hour and then processed for immunofluorescence analysis. In the second dish, cells were kept in agitation for one hour to prevent the adhesion, then cells were forced to adhere by centrifugation and immediately processed for immunofluorescence analysis. We observed that SUM149 basal cells kept in suspension equally displayed EPCAM localization to cell surface as in luminal cell-state (**Figure 3.5 A-B**), suggesting that the re-localization is not dependent on cell adhesion. In all, these results suggest that the phenotypic switch from one differentiation state to another is a quick process in both SUM149 and SUM159. Such a high velocity of interconversion can be explained by the rapid re-localization of the EPCAM protein from cytoplasm to cell membrane and *vice versa*. Notably, EPCAM is already present in the cytoplasm of basal cells, although these cells are usually defined as an $EPCAM^{-}$ subpopulation.



See next page for caption.

Figure 3.4 – A Schematic of experimental procedure. See text for further details. **B** Representative immunofluorescence images showing that EPCAM protein is not present in sorted SUM149 basal membrane at time 0 hour post sorting, but it localizes to cell membrane after one hour, as in luminal purified subpopulation. **C** Barplot quantification of EPCAM localization in SUM149 basal cells. **D** Representative immunofluorescence images showing that EPCAM protein localizes at the cell membrane in SUM159 luminal membrane at time 0 hour post sorting, but it translocates to the cytoplasm one hour after sorting, as in basal purified subpopulation. **E** Barplot quantification of EPCAM localization in SUM159 luminal cells. Scalebar 10 μm.

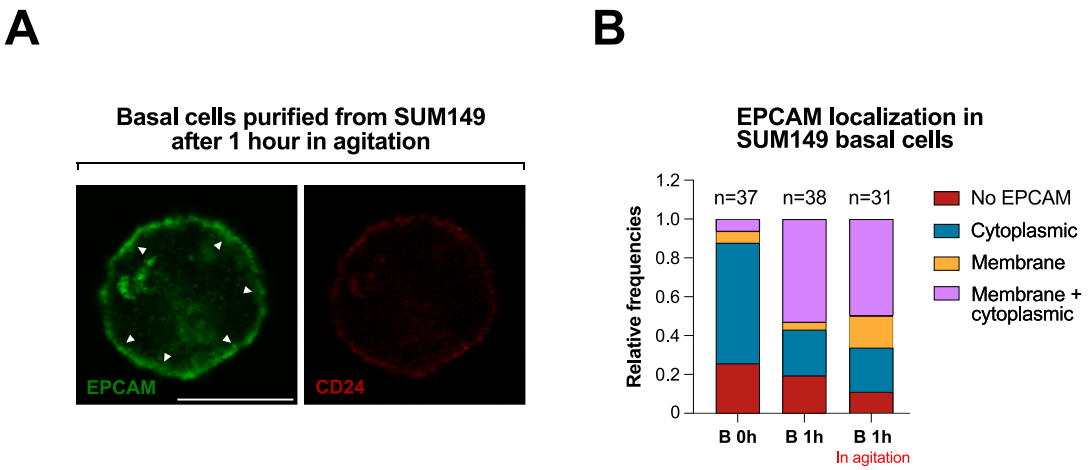


Figure 3.5 – A Representative immunofluorescence images showing that EPCAM and CD24 proteins are localized on cell surface in sorted SUM149 basal membrane after one hour from sorting without adhesion to substrate. **B** Barplot quantification of EPCAM localization in SUM149 basal cells with or without agitation. Scalebar 10 µm.

3.1.4 EPCAM and CD24 are under-expressed in purified basal cells, but highly re-expressed after one hour from sorting

EPCAM seemed to actively shuttle between cytoplasm and cell surface during the interconversion from basal to luminal cell-state, suggesting that the transition between states might represent a process driven by intracellular signaling. On the contrary, CD24 localization showed a slower localization dynamic compared to EPCAM. We asked whether a transcriptional regulation of *EPCAM* and *CD24* was also involved, therefore we assessed the gene expression of these genes at early time points during phenotypic interconversion. We fractionated SUM149 basal and luminal cells and performed qPCR experiments to assess the expression levels of *EPCAM* and *CD24* in these cell subpopulations immediately after sorting (0 hour) and after 1 hour from plating (**Figure 3.6 A**). SUM149 luminal cells, both taken at 0 and 1 hour after sorting, have *EPCAM* and *CD24* mRNA levels comparable with unsorted SUM149 cells (**Figure 3.6 B**). We also observed that *EPCAM* and *CD24* transcript levels are under-expressed in basal cells at 0 hour, compared to unsorted population. Strikingly, after 1 hour, basal cells re-expressed high mRNA levels of both genes, very similarly to unsorted and luminal cells. These data suggest that cells mount a comprehensive reprogramming, engaging transcriptional and signaling programs, to modulate the transitions between states.

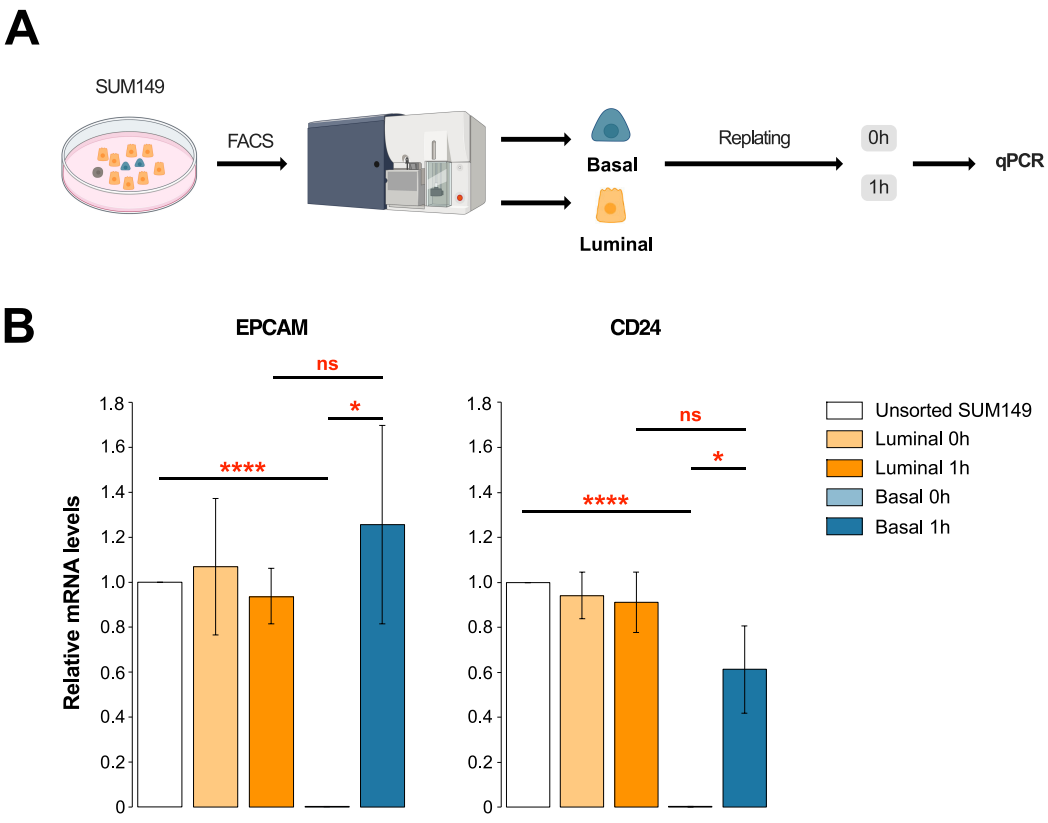
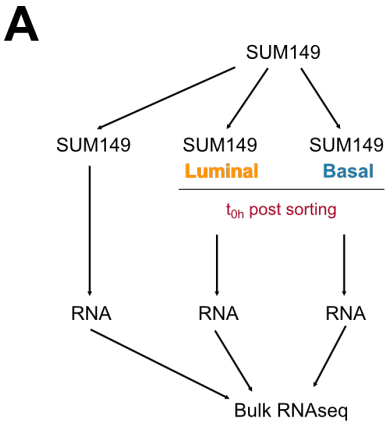


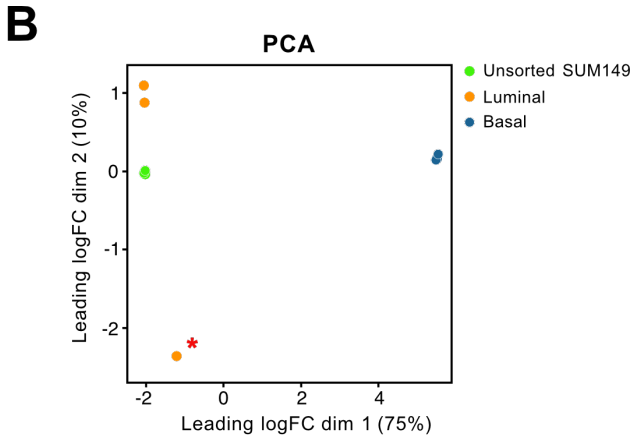
Figure 3.6 – A Schematic of experimental procedure. See text for further details. **B** qPCR analysis of *EPCAM* and *CD24* expression in SUM149 unsorted and purified cell subpopulations showing that *EPCAM* and *CD24* levels are differentially regulated in purified luminal and basal cells. 0h and 1h refer to the timepoints after sorting at which each subpopulation was collected. Data are relative to unsorted SUM149 levels and represented with Standard Error of the Mean (SEM). n = 3.

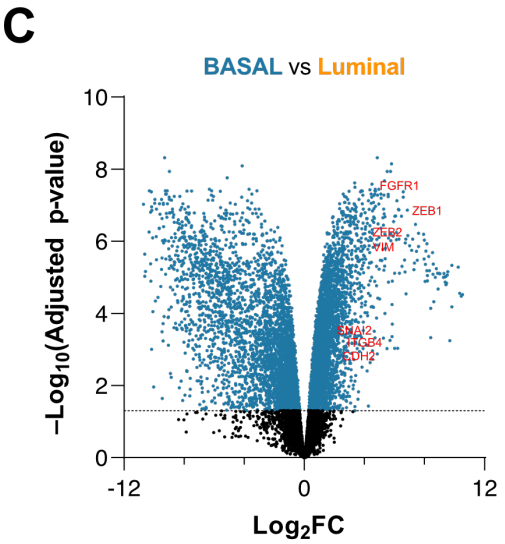
3.1.5 Gene expression analysis reveals extensive difference within sorted SUM149 subpopulations

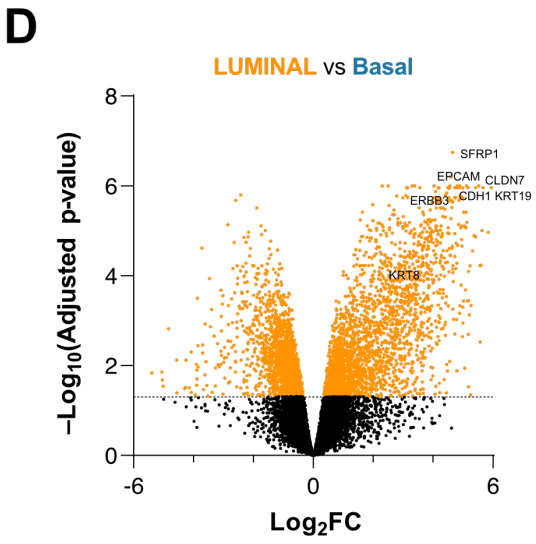
Our data have shown that *EPCAM* and *CD24* underwent a transcriptional rewiring to modulate the transition from basal to luminal state, besides their relocalization from the cytoplasm to the cell surface. Since these differentiation states have distinct gene expression profiles (Sørlie *et al*, 2001; Neve *et al*, 2006; Prat & Perou, 2011; Gutierrez *et al*, 2022) we sought to get a deeper understanding on the transcriptional differences that define SUM149 subpopulations. We obtained bulk RNA-seq profiles from sorted SUM149 luminal and basal subpopulations, with unsorted SUM149 as control (**Figure 3.7 A**). Specifically, we fractionated each cell-state and immediately extracted RNA from cells, in order to avoid any transitions from one state to another. Samples from three different sortings were collected. PCA shows a distinct clusterization of basal and luminal samples (**Figure 3.7 B**). Each point represents a single sorting event. Every cell-state clusterized together, with the exception of one luminal sorting event (in orange) that segregates differently from the other two, as indicated by the asterisk, a likely marginal difference if we consider the low difference in percentage of the second dimensionality (10% on the *y axis*). Basal cells (in blue) are instead tightly clustered. Notably, unsorted SUM149 cells are closer to luminal cells compared to basal cells, as expected. This is a confirmation that luminal and unsorted cells are more similar to each other compared with the basal subpopulations, the latter is undoubtedly transcriptionally different from the other two. The large contribution of the first PCA component (75% on the *x axis*) corroborated this result. Then, we performed a gene expression analysis to investigate the transcriptional heterogeneity of SUM149 cell line. The output of this analysis was a list of differentially expressed genes, that we sorted using the *Adjusted p-value* (threshold value = 0.05). Then, we analyzed changes in expression by ranking the gene list by Log_2FC. Volcano plots revealed 8,164 differentially regulated genes in SUM149 *Basal vs Luminal* and 4,443 genes in *Luminal vs Basal* dataset (**Figure 3.7 C-D**). We found a significant enrichment in the expression of genes associated with the basal state in our sorted $EPCAM^{-}$ $CD24^{-}$ $CD44^{high}$ subpopulation (**Figure 3.7 C**). Specifically, genes such as *ZEB1*,

FGFR1, *ZEB2*, *VIM*, *SNAI2* and *CDH2* were upregulated in basal cells. This result is in accordance with the canonical basal cell-state characterization (**Figure 3.7 F**). On the contrary, we found that genes associated with the luminal phenotype are enriched in our sorted luminal samples, such as *CLDN7*, *SFRP1*, *ERBB3*, *CDH1*, *KRT19* and *KRT8* (**Figure 3.7 D** and **Figure 3.7 F**). We investigated *EPCAM* and *CD24* expression in our sorted subpopulations: we discovered that are both downregulated in basal cells (EPCAM Log_2FC = –10.31287826 *Adj. p-value* = 3.99E-08; CD24 Log_2FC = –6.97218607 *Adj. p-value* = 0.000816845) and upregulated in luminal cells (EPCAM Log_2FC = 5.115309663 *Adj. p-value* = 9.99E-07; CD24 Log_2FC = 4.996516409 *Adj. p-value* = 0.001135302). *EPCAM* and *CD24* expression is consistent with our qPCR experiments. Subsequently, we focused on the basal subpopulation since, in our system, isolated basal cells can rapidly interconvert to luminal state in order to establish equilibrium in a short period of time, whereas luminal cells are more stable in differentiation dynamics (see **Figure 3.3 A**). Therefore, using Enrichr (Chen *et al*, 2013), we found that one of the most downregulated pathways in basal cells belonged to the Regulation of EMT (GO:0010717) Biological Process (**Figure 3.8 A**). Specifically, basal and luminal cells are distinct subpopulations when clustered by EMT-related gene expression, as we have shown in our heatmap (**Figure 3.8 B**). Basal cells exhibited high expression of specific markers of epithelial to mesenchymal transition such as *ZEB1*, *FGF2* and *VIM*. On the contrary, negative regulators of EMT, such as *SFRP1* and *OVOL2*, were upregulated in the luminal subtype. This is consistent with the definition of mammary basal cells as a mesenchymal-like population and luminal cells characterized by an epithelial-like phenotype.









E

	Gene Name	*Log_2FC*	*Adj. p-value*	Cell-state	Reference
BASAL vs Luminal	ZEB1	6.859267771	1.04E-09	Basal	Eger et al. 2005
	FGFR1	4.510879075	2.63E-08	Basal	Shipitsin et al. 2007
	ZEB2	4.030739607	6.14E-07	Basal	Vandewalle et al. 2005
	VIM	4.025217871	1.40E-06	Basal	Shipitsin et al. 2007
	SNAI2	1.61489338	0.000389521	Basal	Gudjonsson et al. 2002
	ITGB4	2.298657903	0.000657842	Basal	Lu et al. 2008
	CDH2	1.984090353	0.001247682	Basal	Gudjonsson et al. 2002

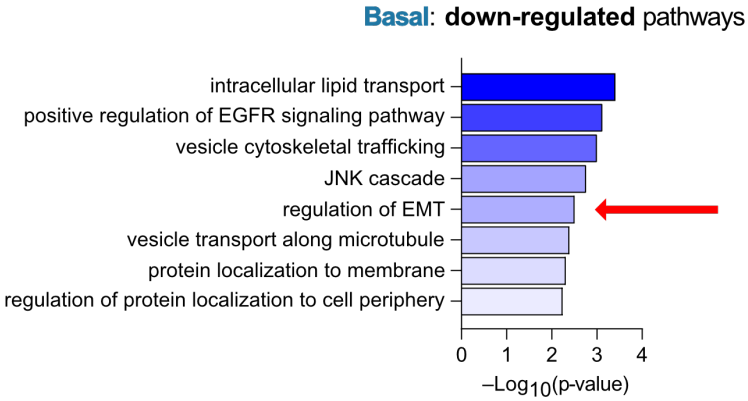
F

	Gene Name	*Log_2FC*	*Adj. p-value*	Cell-state	Reference
LUMINAL vs Basal	EPCAM	5.115309663	9.99E-07	Luminal	Gudjonsson et al. 2002
	CLDN7	5.52090789	9.99E-07	Luminal	Blackman et al. 2005
	SFRP1	4.645055569	1.79E-07	Luminal	Shipitsin et al. 2007
	ERBB3	3.091082321	1.63E-06	Luminal	Jones et al. 2004
	CDH1	4.994021294	1.11E-06	Luminal	Gudjonsson et al. 2002
	KRT19	5.942801357	1.10E-06	Luminal	Gudjonsson et al. 2002
	KRT8	2.262053132	9.01E-05	Luminal	Shipitsin et al. 2007

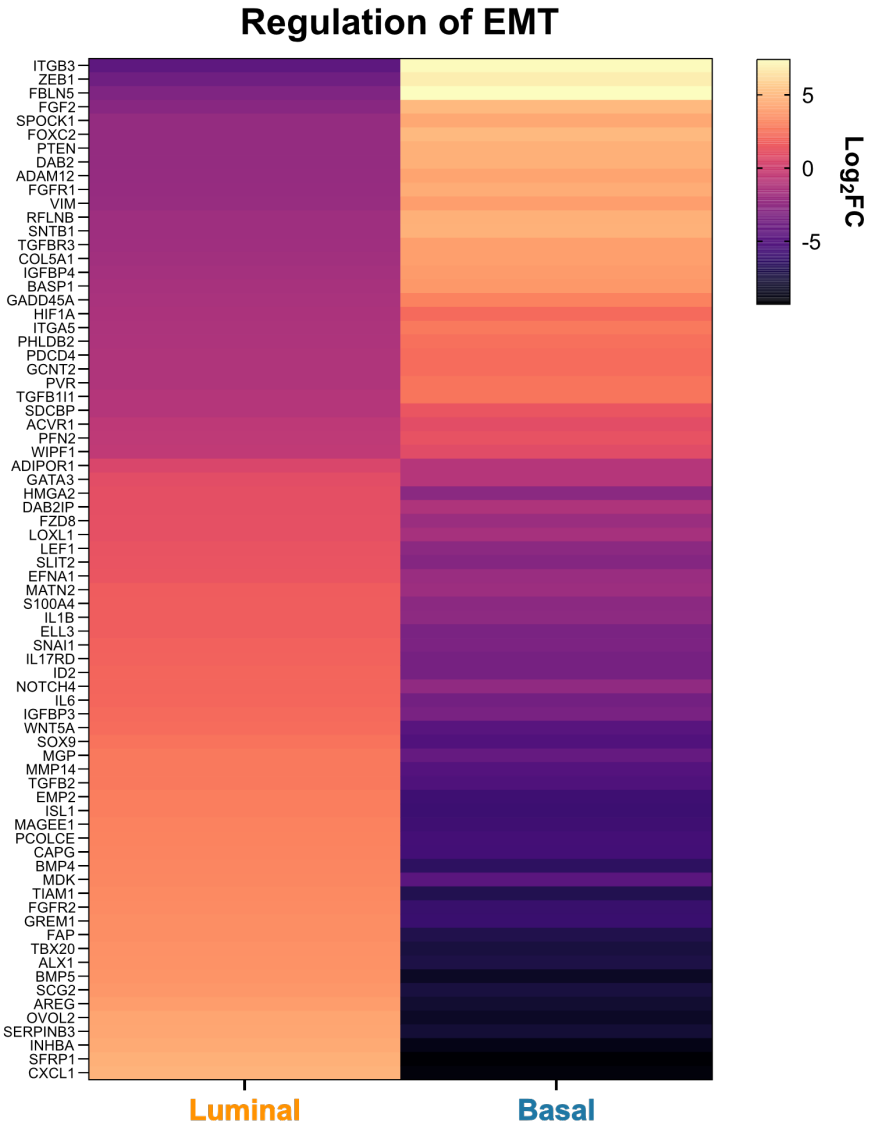
See next page for caption.

Figure 3.7 – **A** Schematic of experimental procedure for bulk RNAseq on SUM149 cell line. See text for further details. **B** PCA plot showing clustered purified basal and luminal samples. Red asterisk showing the luminal sample that did not clustered with the other two samples. **C-D** Volcano plots showing differential gene expression in Basal vs Luminal (C) and Luminal vs Basal (D) samples. **E-F** Representative list of genes defining basal (E) and luminal state (F), with their expression (Log2FC), significance (Adj. p-value) and related references.

A



B



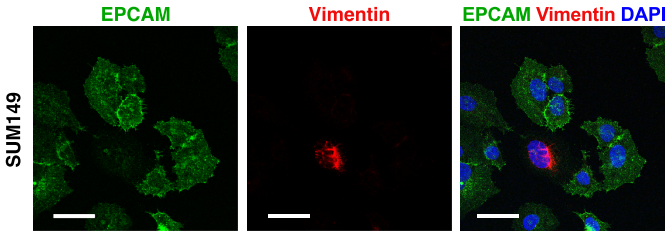
See next page for caption.

Figure 3.8 – A List of GO Biological Process terms enriched in downregulated genes analysed with Enrichr and derived from bulk RNAseq data in SUM149 sorted basal subpopulation. Among these pathways, *regulation of EMT* is significantly downregulated in this sample (indicated by red arrow). **B** Heatmap representing Log_2FC of EMT-related genes in luminal and basal cell-state showing the high degree of differential expression between the two differentiation states.

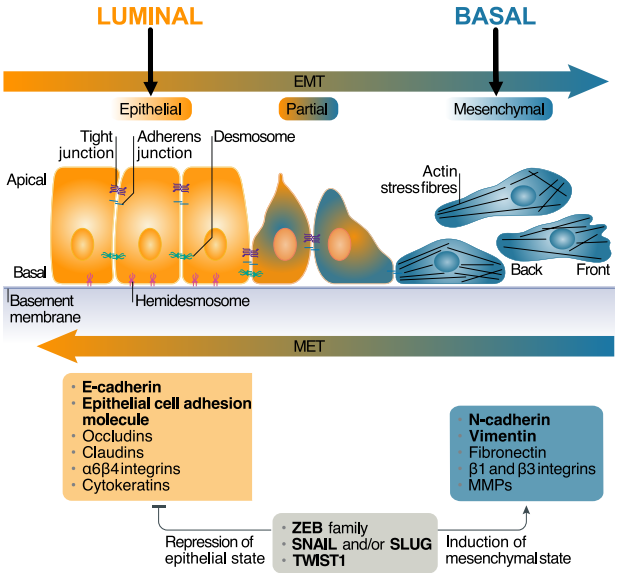
3.1.6 Basal cells completely revert to luminal gene expression profiles after one hour from sorting

We asked whether the transition between states was driven only by changes in few markers or was instead characterized by a broader transcriptional reprograming. We hence took advantage of the gene expression dichotomy in the EMT pathway (as mentioned above, **Figure 3.8 B**) , to address the question. First, immunofluorescence experiments were conducted to confirm the differential expression of mesenchymal and epithelial markers Vimentin and EPCAM in unsorted SUM149 cells. Cells expressing either low or no EPCAM have higher expression of Vimentin compared to cells with high abundance of EPCAM protein, therefore confirming the heterogeneous composition of SUM149 cell lines and the functional characterization of luminal and basal population (**Figure 3.9 A**). We assessed the expression levels of specific genes associated with EMT in basal and luminal cells (genes in bold in **Figure 3.9 B**) via qPCR. Specifically, we compared purified subpopulations at 0 and 1 hour after sorting. *CDH1* (E-Cadherin) and *CDH2* (N-Cadherin), markers of epithelial and mesenchymal states respectively, were differentially regulated in luminal and basal cells taken at 0 hour, as already reported (**Figure 3.9 C**). Accordingly, the mRNA levels of *VIM*, *ZEB1*, *SNAI2* and *TGFβ1* were upregulated in basal cells relative to luminal state, with the exception of *SNAI1*. Noteworthy, each differentially regulated gene in cells within basal state reverted to levels comparable to luminal cells one hour after sorting, as previously proved with *EPCAM* and *CD24*. Notably, there is no difference in gene expression levels between luminal cells taken at 0h and 1h after sorting: SUM149 cell line is nearly entirely composed of luminal cells that, once separated, tend to remain mostly in the luminal state. In conclusion, these data demonstrated that fractionated basal cells revert not only *EPCAM* and *CD24* expression, but also a plethora of different genes after one hour from sorting, engaging a wider transcriptional rewiring to reach luminal cell-state levels.

A

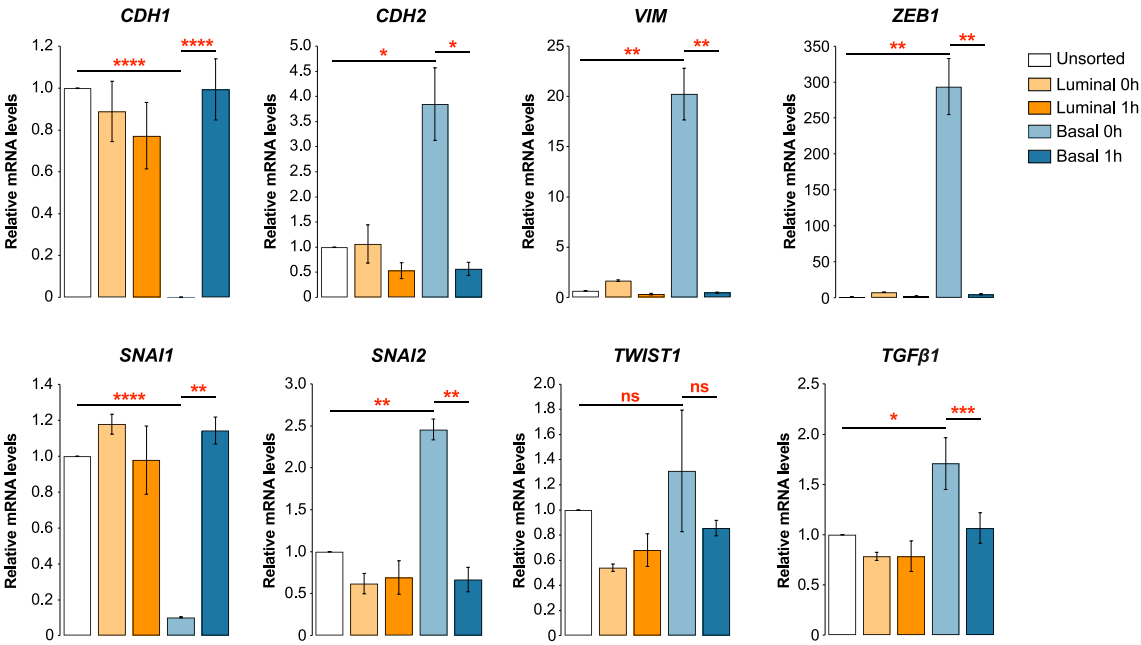


B



Adapted from Dongre et al. 2019

C



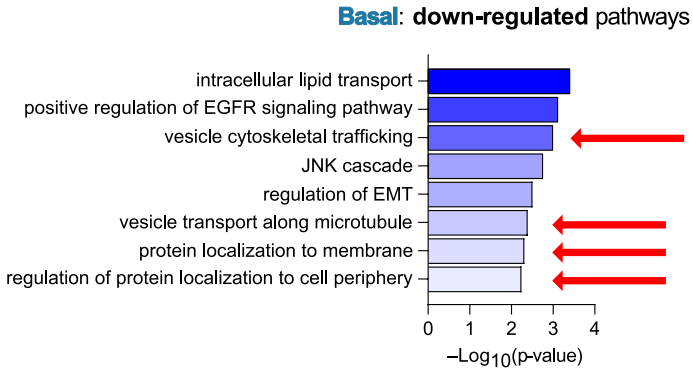
See next page for caption.

Figure 3.9 – A Representative immunofluorescence images showing EPCAM and Vimentin expression in parental SUM149 cell. Scalebar 15 µm. **B** Schematic of EMT with genes validated with qPCR in bold. See text for further details. **C** qPCR analysis of *CDH1* (E-cadherin), *CDH2* (N-cadherin), *VIM*, *ZEB1*, *SNAI1*, *SNAI2*, *TWIST1* and *TGFβ1* expression in SUM149 sorted cell subpopulations relative to expression in parental (unsorted) SUM149 cell line. 0h and 1h refer to the timepoints after sorting at which each subpopulation was collected. Data are represented with Standard Error of the Mean (SEM), n = 3.

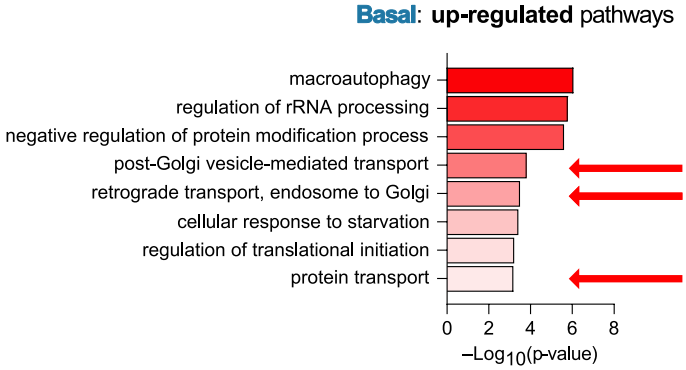
3.1.7 Basal cells actively regulate gene expression of pathways related to protein transport and endocytosis

We have demonstrated that basal cells drastically rewrite their transcriptome in the interconversion to luminal cells just one hour after sorting. Prompted by these results, we performed a thorough analysis with Enrichr on Gene Ontology of Biological Process in basal and luminal cells to find pathways that may be involved in this fast differentiation transition. Strikingly, we found a significant downregulation of *vesicle cytoskeletal trafficking* (GO:0099518), *vesicle transport along microtubule* (GO:0047496), *protein localization to membrane* (GO:0072657), *regulation of protein localization to cell periphery* (GO:1904375) pathways (**Figure 3.10 A**). Conversely, *post-Golgi vesicle-mediated transport* (GO:0006892), *retrograde transport endosome to Golgi* (GO:0042147), *protein transport* (GO:0015031) pathways were significantly upregulated in these cells (**Figure 3.10 B**). On the contrary, luminal cells showed a downregulation of *COPII vesicle coat* (GO:0030127), *clathrin-coated vesicle* (GO:0030136), *transport vesicle membrane* (GO:0030658) pathways (**Figure 3.10 C**). Several genes involved in clathrin-mediated endocytosis, such as *CTLC1*, *AP1S2* and *CTLA* are upregulated in basal cells (**Figure 3.10 D**). These results suggest that basal cells may actively turn on endocytosis and repress protein and cargo movement toward cell periphery, providing a transcriptional link with the translocation of EPCAM from cell membrane to cytoplasm and *vice versa*.

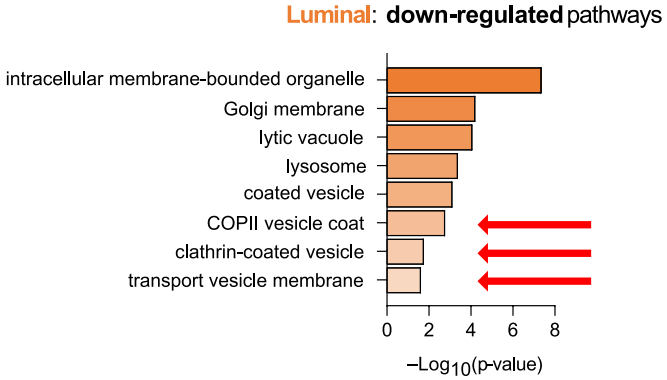
A



B



C



D

Gene Name	*Log₂FC*	*Adj. p-value*	Functional annotation
VTI1A	2.019876524	5.52E-07	SNARE protein involved in intracellular trafficking
AP1S2	2.892220459	6.27E-07	Adaptor Protein complex 1, mediates clhatrin recruitment to cell membrane
CLTCL1	3.045451982	8.68E-07	Clathrin Heavy Chain like 1, member of clathrin family
RAB6B	2.131836822	5.97E-06	Involved in Golgi vesicle transport
RAB9A	1.704375274	1.27E-05	Regulator of exocytosis and protein localization
AP2A1	0.997676964	1.82E-04	Adaptor Protein complex 2, involved in clathrin-mediated trafficking
RAB29	0.954001506	2.62E-04	Involved in receptor recycling
CLTA	1.310534167	0.00041125	Clathrin Light Chain A, member of clathrin family
RAB10	0.739390979	0.00086073	Regulator of intracellular vesicle trafficking
RAB7A	0.583981279	0.00133833	Regulator of intracellular vesicle trafficking
AP4S1	0.730554019	0.00413777	Adaptor Protein complex 4, involved in clathrin-mediated trafficking
SCAMP1	1.338257247	0.00774412	Involved in post-Golgi recycling pathways
CLINT1	0.632374474	0.00993672	Clathrin Interactor 1, involved in the formation of coated vesicles

See next page for caption.

Figure 3.10 – A-C List of GO Biological Process terms enriched in downregulated (A) and upregulated (B) genes analysed with Enrichr and derived from bulk RNAseq data in sorted SUM149 basal and luminal (C) subpopulations. Among them, pathways involved in protein transport, localization and endocytosis process (indicated by red arrows). **D** Table depicting representative selected genes involved in clathrin-mediated endocytosis found in basal cells, with their expression (Log_2FC), significance (*Adj. p-value*) and functional annotation.

3.1.8 Generation of a reporter SUM149 cell line

Our data showed the high interconversion rate among differentiation states in SUM149 cells. Notwithstanding, we had no quantitative evidence of cell-state interconversions in live culturing conditions. Therefore, a closer examination of the behavior of fluorescently tagged endogenous proteins, such as EPCAM and CD24, would be of critical interest to follow how each cell-state emerges and interconverts from subpopulations kept in culture over time. Live fluorescence imaging experiments can monitor the interconversion of cell-states in real time (**Figure 3.11 A**). To this end, to assay the cell status transition in culture over time, in a set of preliminary experiments, we attempted to add antibodies against EPCAM and CD24 directly to plated cells and follow fluorescent signal via live imaging using the Incucyte® S3 (Sartorius) live-imaging microscope. Incucyte® allows the time-course detection of fluorescence (green and red channels) and phase contrast images inside an incubator. We were unable to detect any fluorescent signal, possibly due to EPCAM antibody-mediated internalization and lysosome degradation (Lund *et al*, 2014) (data not shown). Therefore, we decided to adopt a different strategy, adding in-frame fluorescent tags, positioned just before the stop codon of *EPCAM* and *CD24* (TurboRFP and eGFP, respectively), exploiting the CRISPR/Cas9 system. We took advantage of endogenous homologous recombination to generate fusion proteins in a set of knock-in experiments in SUM149 cell line. Briefly, we amplified the fluorescent protein sequence from a lentiviral vector routinely used in our laboratory with primers consisting of two sequences: 1) a sequence homologous to eGFP and TurboRFP; 2) a sequence homologous to *EPCAM* or *CD24* necessary to induce homologous recombination after cutting with Cas9 protein. However, again we were unable to detect any fluorescent cells (data not shown). Possible explanations include the low efficiency of recombination and the reliance of the assay on the generation of chimeric proteins. We hence moved to another CRISPR/Cas9 protocol. We took advantage of the eFLUT tool (Stewart-Ornstein & Lahav, 2016), a protocol to generate a transcriptional reporter bypassing the creation of a fusion protein (**Figure 3.11 B**). Briefly, we inserted a sequence in-frame to the 3' end of *EPCAM* gene (**Figure 3.11 C**). The inserted sequence

contains a T2A cleavage tag between the gene of interest and a fluorescent protein (eGFP) with attached an NLS-PEST cassette, followed by a selectable antibiotic marker (neomycin or blasticidin). The endogenous protein is expressed and translated as a separate polypeptide with a fluorescent protein localized to the nucleus and rapidly degraded thanks to NLS-PEST sequence. The homologous recombination efficiency is still low (0.01%), but selection of modified cells is made possible by antibiotic treatment. The inserted sequence is amplified via PCR from plasmids kindly provided by the authors of the eFLUT paper. To design the sgRNAs that would target the *EPCAM* gene, first we identified the genomic sequence around the stop codon of EPCAM (~ 50 – 100 bp either side) obtained from the UCSC genome browser (**Figure 3.12 A**).

Then, we identified any protospacer adjacent motif (PAM) sequences and their cognate upstream nucleotide sequence (~ 20 bp in length). PAM (Protospacer adjacent motif) is a short DNA sequence, typically consisting of 3 base pairs NGG where N is any nucleotide that follows the Cas9-targeted DNA region. We designed three different sgRNAs: sgRNA_1, sgRNA_2 and sgRNA_3. Next, we cloned each guide into a vector plasmid containing the Cas9 sequence (**Figure 3.12 B**) in the form of double strand oligos (**Figure 3.12 C**). Therefore, we Sanger-sequenced the cloned vectors and we found that the sgRNA_1 and sgRNA_2 were inserted correctly (**Figure 3.13 A**). In order to induce the homology recombination machinery after the cutting with Cas9, we designed homology arms by selecting 40 bp immediately preceding and after the stop codon. To destroy the sgRNA site and the PAM sequence in the donor sequence, we introduced silent mutations into the homology arms to self-inactivate the sgRNA sequence (**Figure 3.13 B**). Consequently, we designed custom DNA primers with modified sequences in order to amplify the 2.2 kb region containing eGFP and neomycin marker (**Figure 3.13 C**). The amplicon length was further measured and validated with agarose gel, then gel-purified (**Figure 3.13 D**). At last, we generated cells expressing nuclear eGFP under control of endogenous *EPCAM* promoter (**Figure 3.14**). Currently, we are implementing more genes, such as *CD24*, as well as other basal and luminal markers, such as *VIM* or *CDH1*. In conclusion, this tool will allow us the tracking

of phenotypic cell-state behaviors directly in culture and to closely follow the timing of the phenotypic switch of a single cell, based on the expression of EPCAM and defining luminal or basal differentiation state.

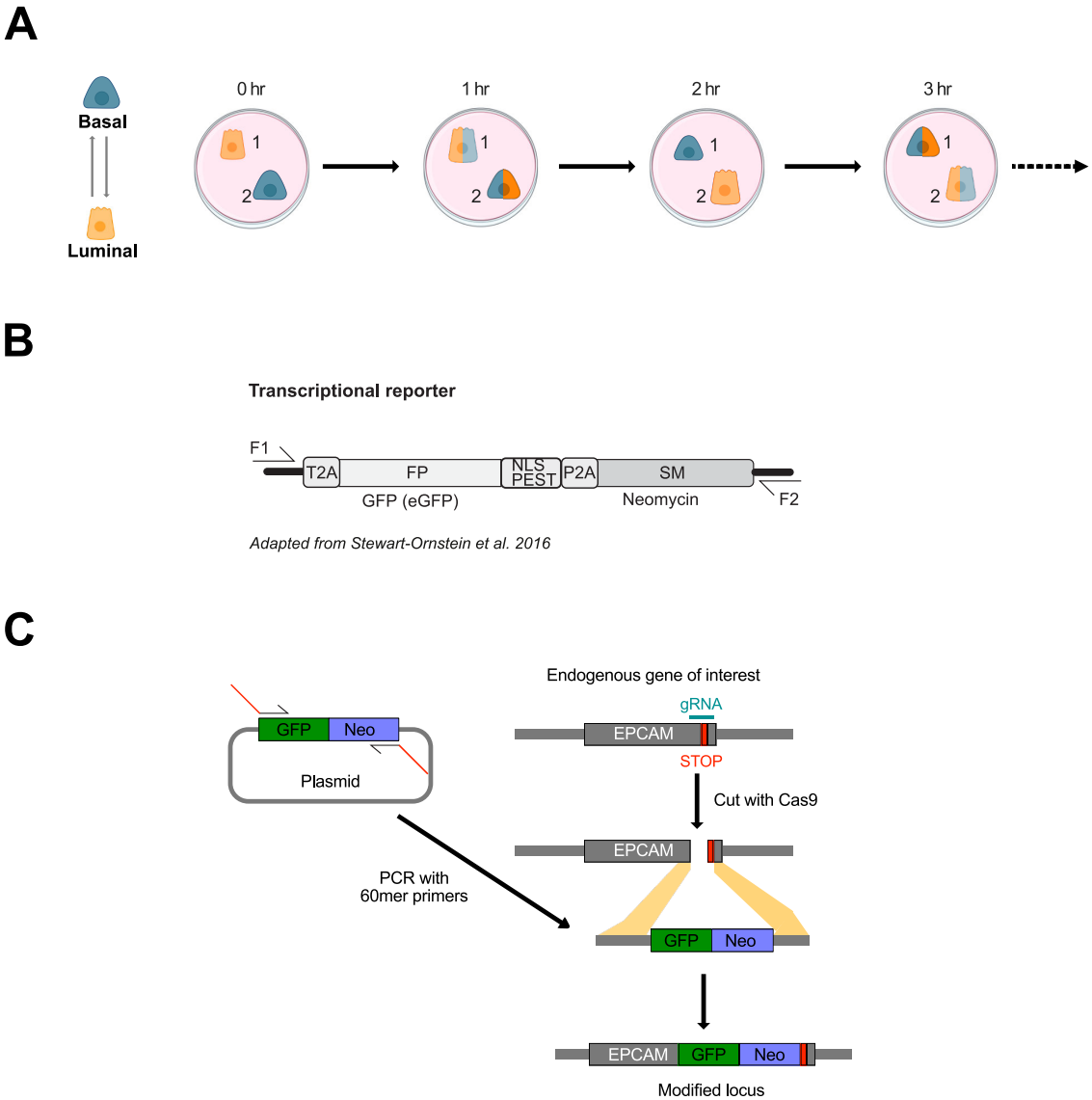


Figure 3.11 – A Schematic of experimental procedure. See text for further details. **B** Schematic of the transcriptional reporters harbouring the T2A and P2A viral cleavage sequences between the ORF and the fluorescent protein (T2A), and between fluorescent protein and the selection marker. FP = Fluorescent Protein. SM = Selection Marker. NLS = Nuclear Localization Sequence. **C** Schematic of the experimental approach, depicting the insertion (knock-in) of the transcriptional reporter via CRISPR/Cas9. See text for further details.

A

RefSeq Gene EPCAM - NM_002354.3

sgRNA_1 EPCAM **GC content: 50%**

agaatgaacaaaagattgaaaaattatttagaatttttttctgtgctttttcctgttt

cagATAAAGGAGATGGGTGAGATGCATAGGGAACTCAATGCATAACTATATAATTTGA

AGATTATAGAAGAAGGGAAATAGCAAATGGACACAAATTACAAATGTGT

sgRNA_2 EPCAM **GC content: 50%**

agaatgaacaaaagattgaaaaattatttagaatttttttctgtgctttttcctgttt

cagATAAAGGAGATGGGTGAGATGCATAGGGAACTCAATGCATAACTATATAATTTGA

AGATTATAGAAGAAGGGAAATAGCAAATGGACACAAATTACAAATGTGT

sgRNA_3 EPCAM **GC content: 40%**

agaatgaacaaaagattgaaaaattatttagaatttttttctgtgctttttcctgttt

cagATAAAGGAGATGGGTGAGATGCATAGGGAACTCAATGCATAACTATATAATTTGA

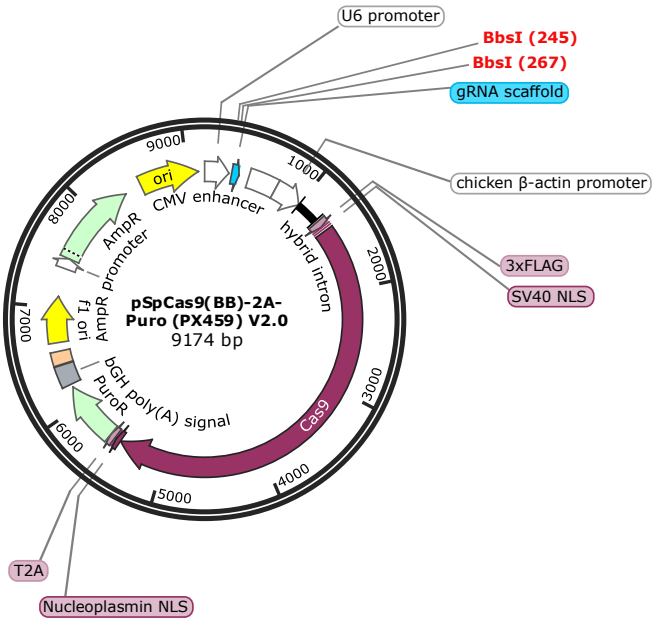
AGATTATAGAAGAAGGGAAATAGCAAATGGACACAAATTACAAATGTGT

Red = PAM sequence

Blue = Stop Codon

Green = Guide sequence

B



C

Oligo form

BbsI - end

5′ CACCGNNNNNNNNNNNNNNNNNNN 3′
3′ CNNNNNNNNNNNNNNNNNNNCAAA 5′

BbsI - end

sgRNA_1 EPCAM

5′ CACCGGGAGATGGGTGAGATGCATA 3′
3′ CCCTCTACCCACTCTACGTATCAAA 5′

sgRNA_2 EPCAM

5′ CACCGAGGAGATGGGTGAGATGCAT 3′
3′ CTCCTCTACCCACTCTACGTACAAA 5′

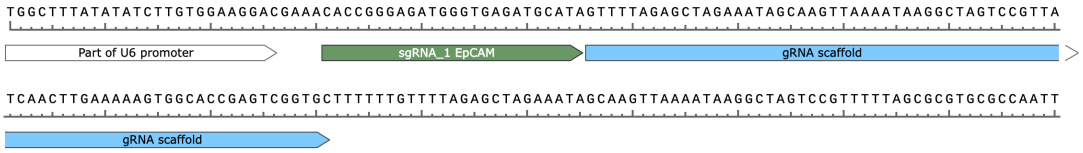
sgRNA_3 EPCAM

5′ CACCGCCTGTTTCAGATAAAGGAGA 3′
3′ CGGACAAAGTCTATTTCCTCTCAAA 5′

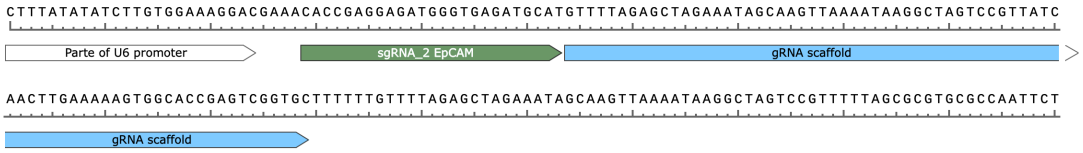
Figure 3.12 – **A** EPCAM genomic sequence surrounding the STOP codon (TAA). Here, the three sgRNAs targeting *EPCAM* are listed. **B** Schematic of pSPCas9(BB)-2A-Puro (PX459) V2.0 plasmid, encoding for Cas9 protein under chicken ß-actin promoter. Each sgRNA is inserted in the vector after cutting with BbsI restriction enzyme. In red are depicted BbsI cutting sites. **C** The three *EPCAM* sgRNAs are designed and synthesized as in the first oligo form, with the sequence complementary for *EPCAM* genomic region (in green) and the sequence compatible with BbsI sticky ends (highlighted in red).

A

sgRNA_1 EPCAM



sgRNA_2 EPCAM



B

WT EPCAM sequence: ATAAAGGAGATGGGTGAGATGCATAGGGAACTCAATGCATAA

AA sequence: I K E M G E M H R E L N A *

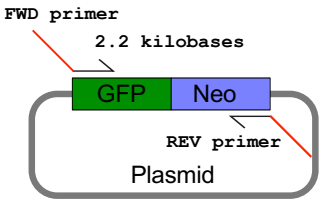
Modified EPCAM sequence: ATAAAGGAAATGGGAGAAATGCACAGAGAACTCAATGCATAA

Disrupted PAM sequence

Blue = Stop Codon

Green = Guide sequence

C



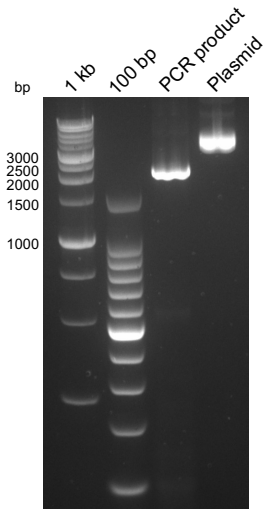
FWD primer

5′ GATAAAGGAAATGGGAGAAATGCACAGAGAACTCAATGCA GGAGCTGGTGCAGGCG 3′

REV primer

5′ GCTATTTCCCTTCTTCTATAATCTTCAAATTATATAGTTAATGGGTGGAGGCGGTTC 3′

D



See next page for caption.

Figure 3.13 – A pSPCas9(BB)-2A-Puro (PX459) V2.0 plasmid Sanger-sequenced showing the correct insertion of *EPCAM* sgRNA_1 and sgRNA_2 under the U6 promoter, after ligation. **B** Schematic illustrating the wild type (WT) *EPCAM* genomic sequence around the STOP codon, the aminoacidic (AA) sequence, and the modified (mutant) *EPCAM* genomic sequence in order to create primers with homology arms and self-disrupting PAM sequence. **C** Forward (FWD) and Reverse (REV) primers consisting of 40 bp immediately preceding the *EPCAM* STOP codon, and 40 bp following TAA but reverse complemented (in red), followed by sequence homologous to the donor plasmid harboring the fluorescent protein and selection marker sequence. **D** Agarose gel showing PCR amplification of the donor (fluorescent protein + selection marker) sequence (~ 2,200 bp), loaded alongside with donor plasmid, not amplified (~ 6,000 bp).

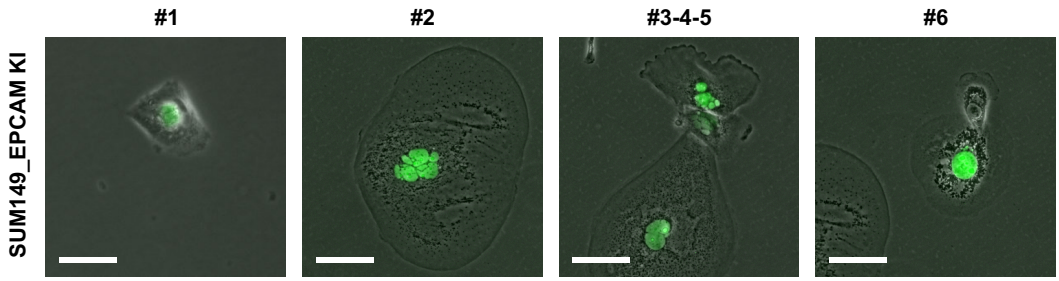


Figure 3.14 – SUM149 cells with *EPCAM*-tagged clones (SUM149 *EPCAM* KI). With the NLS sequence introduced at the C-ter, the GFP under the *EPCAM* promoter is translated and translocated to the nucleus, as expected.

3.1.9 Factors secreted in basal and luminal conditioned media are not responsible for interconversion between different cell-states

Our data suggest that the phenotypic transition in SUM149 cells is an exceedingly fast biological switch, which occurs within few hours. However, how cells are able to maintain cell state proportions over time? We reasoned that the mechanism regulating this kind of maintenance could be determined by extrinsic factors, such as the paracrine influence among cells. To test this hypothesis, that is, that some factors secreted from luminal and/or basal cells may be responsible for the maintenance of all subpopulations, we devised an experiment where we tested whether switching media between the two cell lines could impact on the population structure. More specifically, SUM149 cell line is characterized by a majority of cells in the luminal state, and, conversely, almost all SUM159 cells are in basal populations. We hence reasoned that switching the culturing medium between cell lines could reflect an enrichment of basal factors (from SUM159 medium) or luminal factors (from SUM149 medium) and a modification in cell-state proportions. Thus, we grew SUM149 and SUM159 cells for three days, and then we collected the respective media (**Figure 3.15 A**). Medium derived from SUM149 was called Conditioned Medium 149 (CM149); medium derived from SUM159 was called Conditioned Medium 159 (CM159). Then, we plated both SUM149 and SUM159 cell lines and we added to each cell line: 1) CM149; 2) a combination of half fresh medium and CM149 (CM149/2); 3) CM159; 4) a combination of half fresh medium and CM159 (CM159/2). We used fresh complete medium as control. CM149/2 and CM159/2 media were used to counterbalance the effect of a possible exhaustion of CM149 and CM159 full media. We checked cell-state proportions via FACS after five days of culturing. Interestingly, we did not find any significant differences in subpopulations composition in both SUM149 and SUM159 cell lines grown with different conditioned media, compared to control (**Figure 3.15 B**). These findings suggest that media enriched with factors released by a predominant population of luminal or basal cells may not have the ability to modify the long-term equilibrium of stabilized subpopulations of cancer cells. However, additional precise experiments are necessary to delve deeper into the role of paracrine networks.

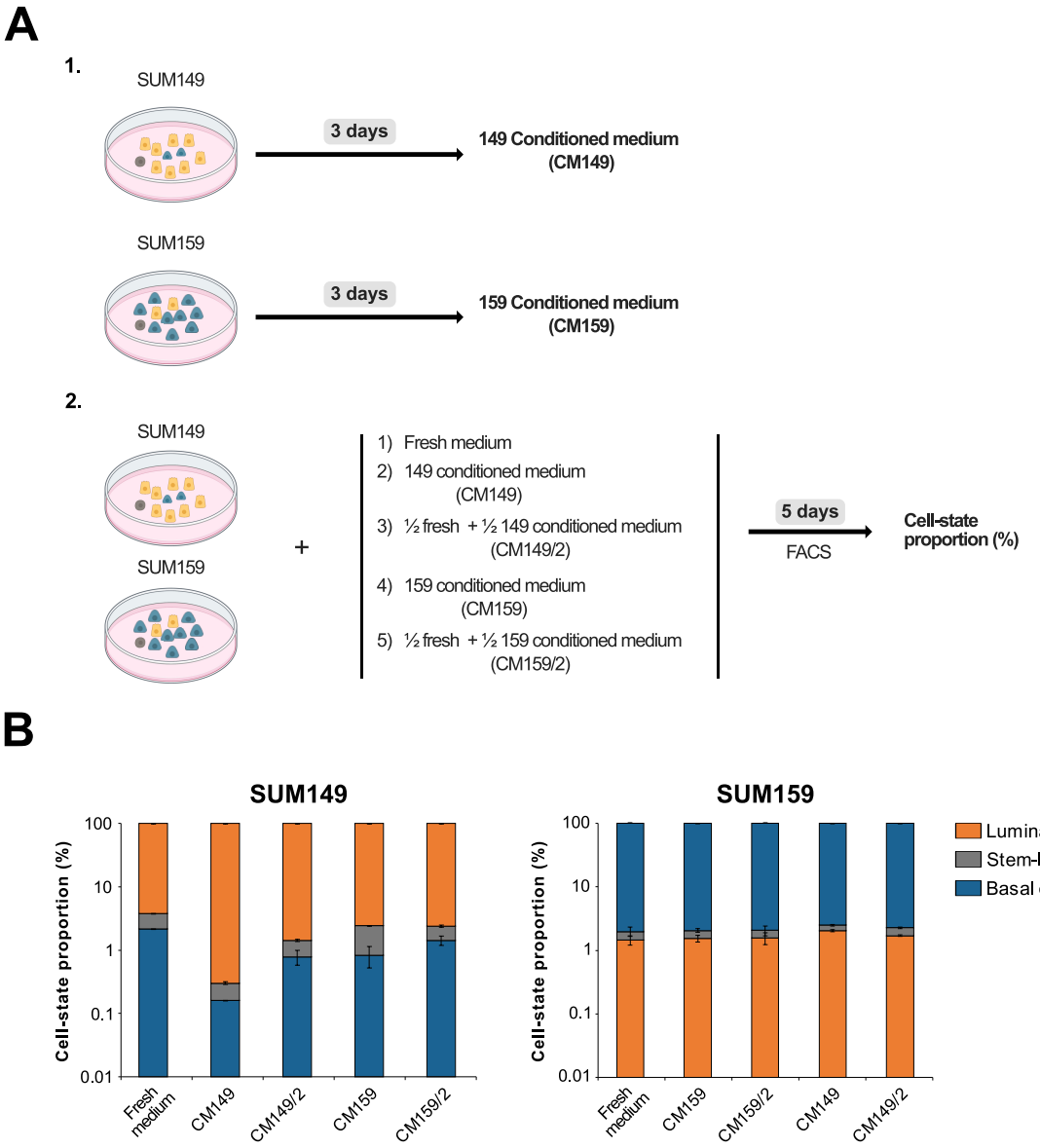


Figure 3.15 – **A** Schematic of experimental procedure for conditioned media experiments. See text for further details. **B** Barplots showing cell-state proportions in SUM149 and SUM159 cell lines cultured with different conditioned media, represented with Standard Deviation (SD). n = 3.

3.2 Defining plasticity and cell-state dynamics in other breast cancer cell lines

3.2.1 Transcriptional and protein heterogeneity are common features within breast cancer cell lines

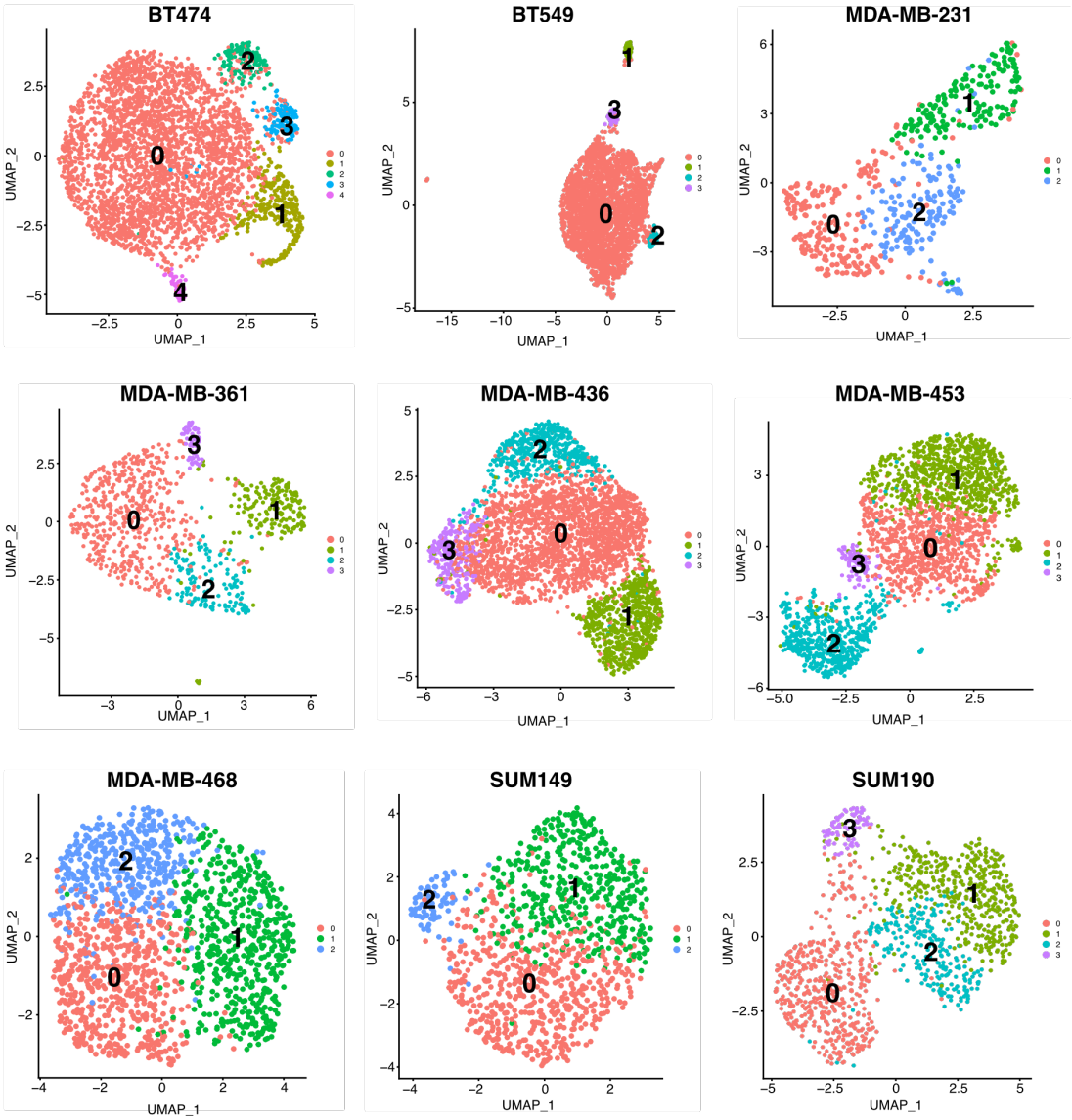
Prompted by the results with SUM149 and SUM159 cell lines, we asked whether the intrinsic heterogeneity and the phenotypic plasticity detected in these cell lines are common features of other breast cancer cell lines. To this aim, we explored a single-cell RNAseq dataset deposited on GEO, which includes a panel of breast cancer cell lines (GSE182694) (**Figure 3.16 A**). We used the statistic software environment *R* to investigate the transcriptomic heterogeneity of these selected cell lines and generate UMAP plots. UMAPs are two-dimensional plots, where each dot corresponds to a cell in a 2D space. We discovered that all breast cancer cell lines analyzed were heterogeneous, independently of breast cancer subtype (**Figure 3.16 B**). Specifically, we could clearly distinguish separate subpopulations based on their transcriptome profiles, in several cell lines, including for example MDA-MB-231, MDA-MB-453 and SUM190 cell lines. Other cell lines such as MDA-MB-436, MDA-MB-468 or SUM149 show a less but still clear degree of transcriptional heterogeneity. Interestingly, all analyzed cell lines have a discrete number of cells clusters, suggesting that the heterogeneity, at least at transcriptional level, is a common feature in stabilized cell cultures. Given these results, we further investigated breast cancer cells heterogeneity by evaluating the expression of the specific cell-surface markers EPCAM, CD24 and CD44 used in our previous experiments in SUM149 and SUM159 cell lines (**Figure 3.17-3.18-3.19**), further revealing the differential expression of these markers within and among each cell line. To corroborate these results, we used a combination of antibodies against EPCAM, CD24 and CD44 in a panel of selected cell lines: BT549, MDA-MB-231, MBA-MB-436 and MDA-MB-468. We introduced also CD49f, or integrin α6, a novel marker that discriminates stem cells (**Figure 3.20**). Using FACS, we observed a certain degree of heterogeneity in the triple negative breast cancer cell line BT549, by means of CD24 expression. This cell line is characterized by the co-existence of cells without CD24 expression ($CD24^{-}$) but other cells expressing CD24 ($CD24^{+}$).

A minor subset of $CD44^-$ cells could be distinguished from the vast majority of $CD44^+$ cells within the MDA-MB-231 cell line. MDA-MB-436 displayed a consistent degree of CD44 and CD49f expression but no difference in terms of EPCAM. MDA-MB-468 showed a minor subpopulation of CD24 positive cells. All together, these results indicated that stabilized breast cancer cell lines retain a certain degree of transcriptomic heterogeneity coupled with differential cell-surface markers expression, suggesting that these subpopulations are kept in an equilibrium status.

A

Cell line	Subtype	References
BT474	Luminal B	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009; Lacroix 2010*
BT549	Triple negative B	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009; Lacroix 2010*
MDA-MB-231	Triple negative B	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009; Lacroix 2010*
MDA-MB-361	Luminal B	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009; Lacroix 2010*
MDA-MB-436	Triple negative A	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009; Lacroix 2010*
MDA-MB-453	Her2+	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009; Lacroix 2010*
MDA-MB-468	Triple negative A	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009; Lacroix 2010*
SUM149	Triple negative B	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009*
SUM190	Her2+	*Riaz 2013; Neve 2006; Hollestelle 2010; Kao 2009*

B



See next page for caption.

Figure 3.16 – A Transcriptomic analyses were performed on the breast cancer cell lines listed in the table. Here, subtypes and relative references are listed. **B** UMAPs depicting the transcriptomic profiles of BT474, BT549, MDA-MB-231, MDA-MB-361, MDA-MB-436, MDA-MB-453, MDA-MB-468, SUM149 and SUM190 cell lines. Each cell line was characterized by a certain degree of transcriptomic heterogeneity: identified subpopulations are marked by a number in bold (e.g. 1, 2, 3 …).

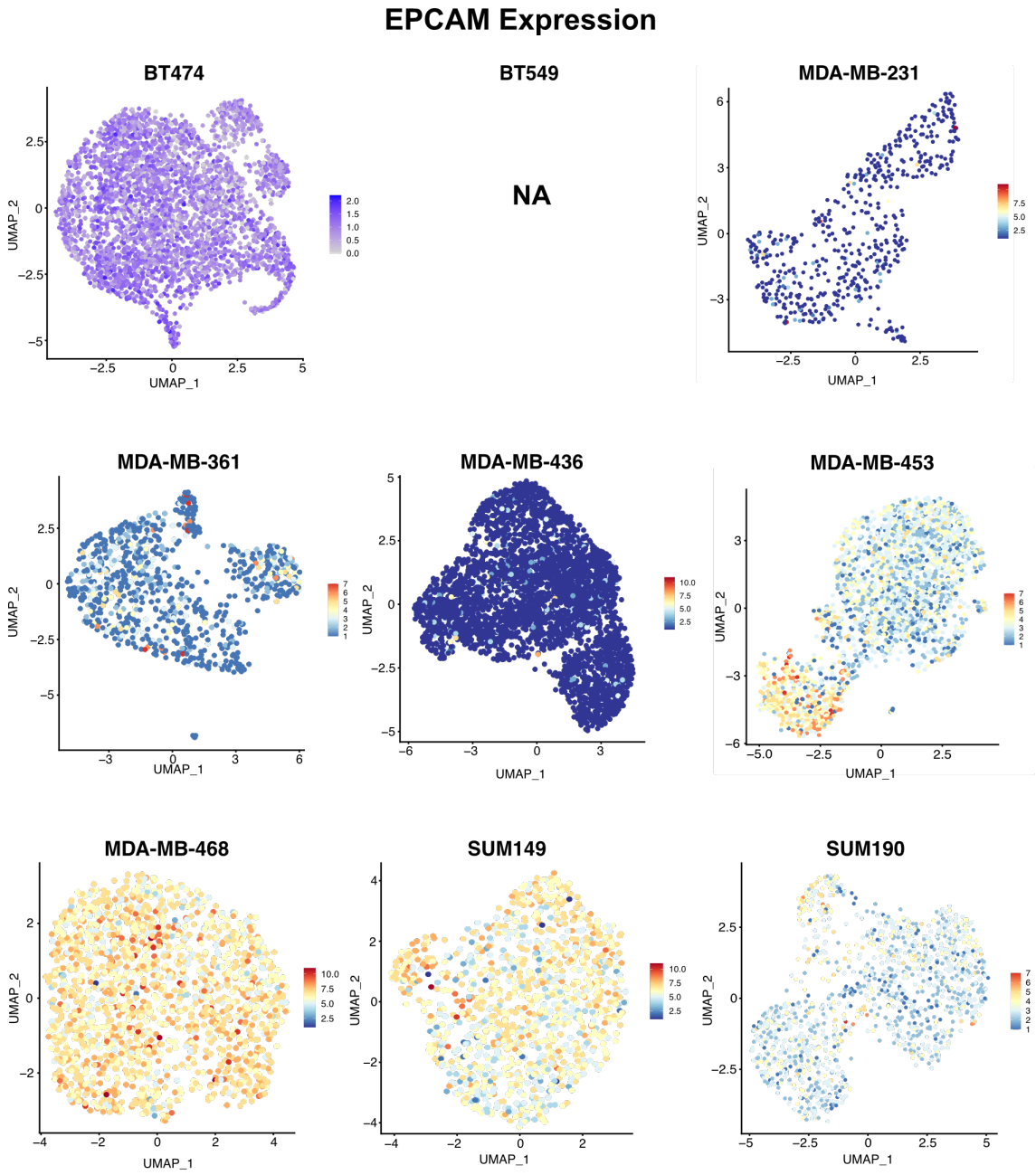


Figure 3.17 – A UMAPs depicting EPCAM expression profiles of BT474, BT549, MDA-MB-231, MDA-MB-361, MDA-MB-436, MDA-MB-453, MDA-MB-468, SUM149 and SUM190 cell lines. NA = Not Available. The color represents the average expression level.

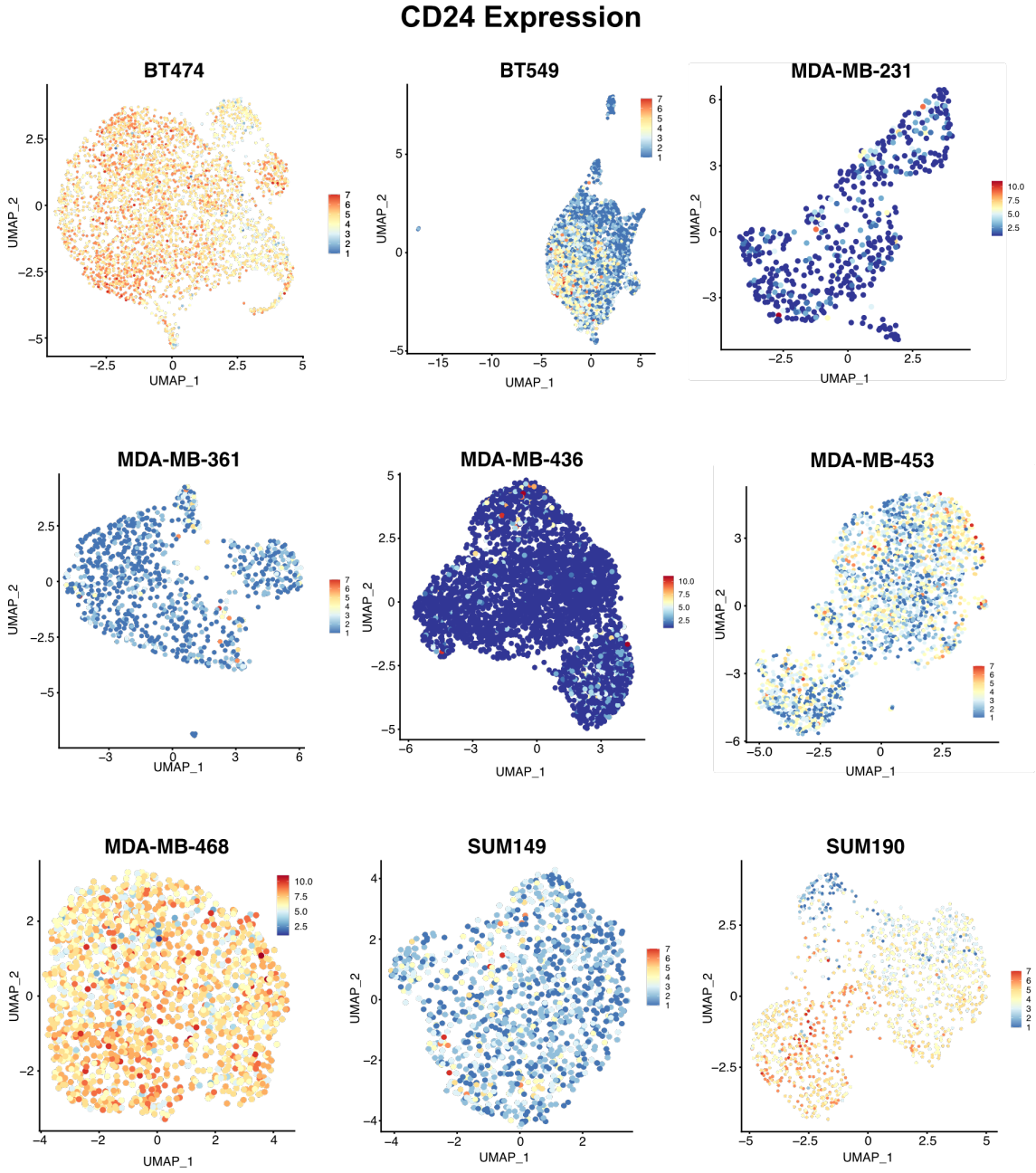


Figure 3.18 – A UMAPs depicting CD24 expression profiles of BT474, BT549, MDA-MB-231, MDA-MB-361, MDA-MB-436, MDA-MB-453, MDA-MB-468, SUM149 and SUM190 cell lines. NA = Not Available. The color represents the average expression level.

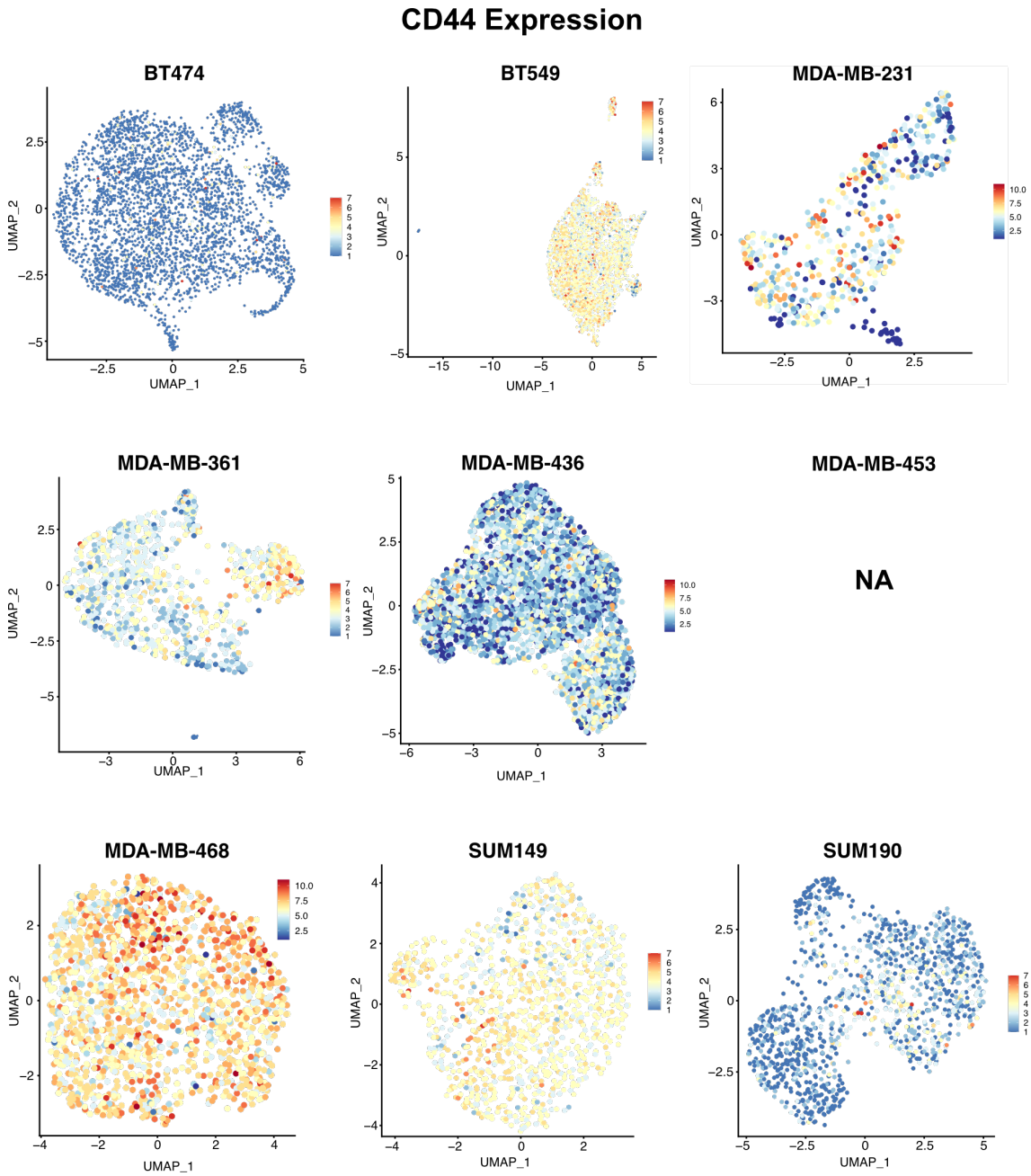


Figure 3.19 – A UMAPs depicting CD44 expression profiles of BT474, BT549, MDA-MB-231, MDA-MB-361, MDA-MB-436, MDA-MB-453, MDA-MB-468, SUM149 and SUM190 cell lines. NA = Not Available. The color represents the average expression level.

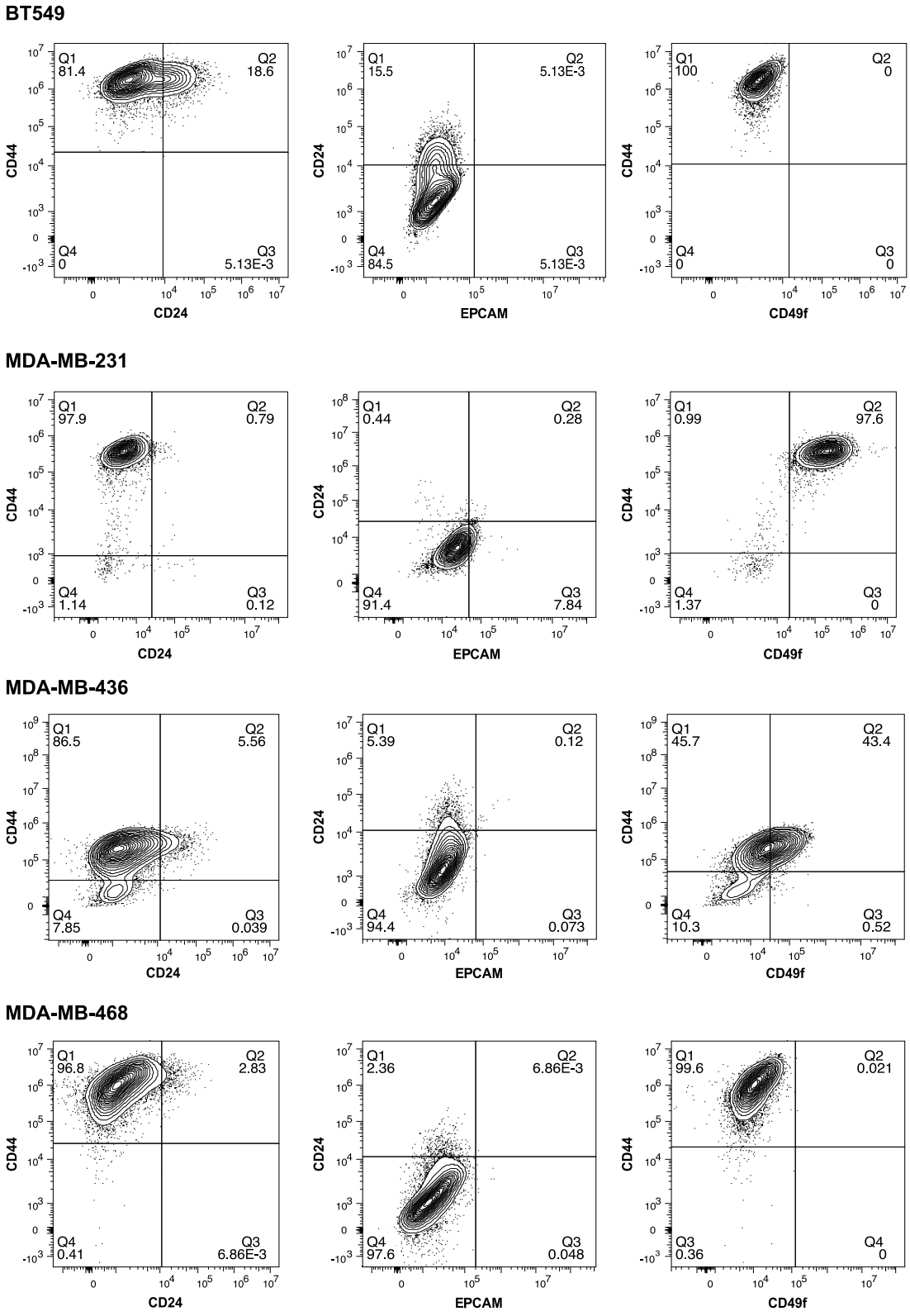
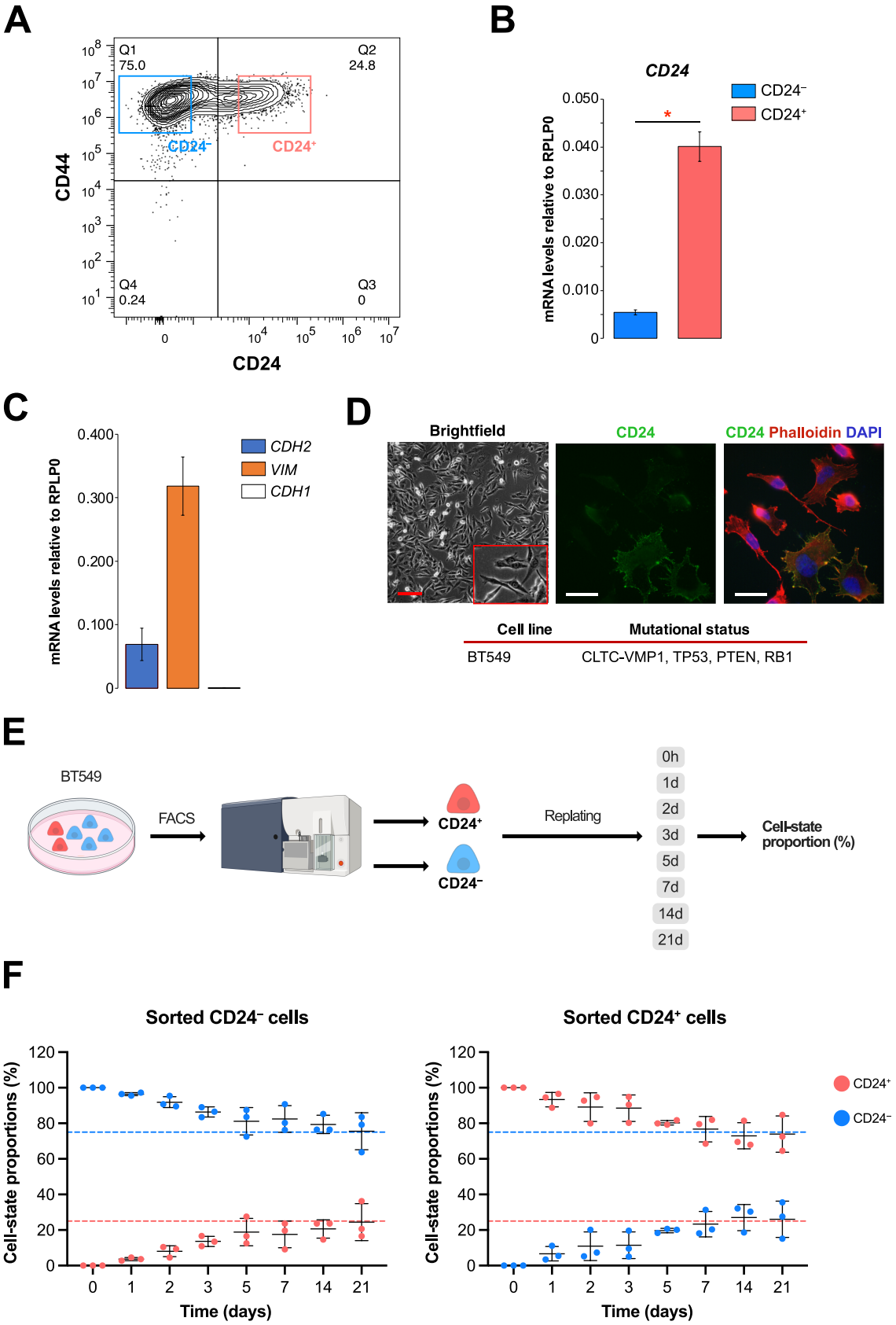


Figure 3.20 – A Representative FACS dot plots reporting EPCAM, CD24, CD44 and CD49f expression on a selected panel of triple negative breast cancer cell lines (BT549, MDA-MB-231, MDA-MB-436, MDA-MB-468).

3.2.2 BT549 cell line is characterized by the co-existence of CD24^{-} and CD24^{+} subpopulations displaying a plastic phenotype

Based on these findings, we focused on BT549, a triple-negative breast cancer cell line, because this cell line exhibited a clear transcriptional CD24 heterogeneity that was paralleled by differential CD24 protein expression. FACS analyses demonstrated that BT549 consisted of two main subpopulations: CD44^{high} CD24^{-} and CD44^{high} CD24^{+} (hereafter called CD24^{-} and CD24^{+}) (**Figure 3.21 A**). Here, CD24^{-} and CD24^{+} proportions are ~ 75% and ~ 25%, respectively. Remarkably, qPCR analyses confirmed the differential expression of *CD24* in CD24^{-} and CD24^{+} subpopulations (**Figure 3.21 B**). Since EPCAM is not expressed (see **Figure 3.20**), we cannot clusterized BT549 cells in luminal, stem-like and basal cell-state. Instead, we could define this cell line as mesenchymal-like, based on the expression of *CDH2* and *VIM*, compared to the expression of the epithelial marker *CDH1* (**Figure 3.21 C**). The spindle-like morphology of these cells supported this classification (**Figure 3.21 D**). We were able to validate the heterogeneous expression of the CD24 protein also with immunofluorescence experiments. As in the representative micrograph in **Figure 3.21 D**, the intensity of CD24 varies among cells in the same field, ranging from no expression, low signal and high intensity, reflecting the broad spectrum of expression detected by FACS. To assess the plasticity of BT549 cell line and the ability of the CD24^{-} and CD24^{+} subpopulation to interconvert between states, we fractionated each specific subpopulation from parental cell line as we have performed in SUM149 and SUM159 cells. Therefore, we isolated CD24^{-} and CD24^{+} via FACS (**Figure 3.21 E**), replated each subpopulation and followed the evolution over time by expanding each lineage in culture for a maximum of 21 days. We next measured cell-state proportions at specific timepoints (0, 1, 2, 3, 5, 7, 14 and 21 days) via FACS analysis on the expanded subpopulations. The time course experiment revealed that purified CD24^{-} cells restore the original bulk subpopulation percentages within five days (**Figure 3.21 F**). Surprisingly, CD24^{+} cells differentiated from CD24^{-} started to emerge only 24 hours after sorting. Following the 5th day, the subpopulation percentages were maintained. On the contrary, purified CD24^{+} cells failed to restore equilibrium proportions

within 21 days: as $CD24^-$ cells differentiated from sorted $CD24^+$ cells within 24 hours, they didn't reach parental proportions in the observed timeframe. This result raised the possibility that purified $CD24^+$ cells might have a slower differentiation dynamic compared to $CD24^-$, so that they need more time to reinstate the original proportion. Our data indicate that only BT549 $CD24^-$ subpopulation can reestablish equilibrium following FACS fractionation, whereas BT549 $CD24^+$ subpopulation, although displaying a plastic behavior, may have a different kinetics.



See next page for caption.

Figure 3.21 – A Representative flow cytometry plot reporting CD24 expression in BT549 cell line. Here, $CD24^-$ (in blue) and $CD24^+$ (in red) subpopulations gating strategy is reported. **B** qPCR analyses of *CD24* expression relative to *RPLP0* in sorted BT549 cell subpopulations. Data are represented with Standard Error of the Mean (SEM), n = 3. **C** qPCR analyses of *CDH2*, *VIM* and *CDH1* expression relative to *RPLP0* in unsorted, parental BT549 cell subpopulations. Data are represented with Standard Error of the Mean (SEM), n = 3. **D** Brightfield (with 2X magnification in the corner) and immunofluorescence images showing CD24 and Phalloidin staining in BT549 cells in culture. Scalebar 100 µm for brightfield image. Scalebar 15 µm for fluorescent microscopy images. Below, a table reporting BT549 mutational status. **E** Schematic of experimental procedure for conditioned media experiments. See text for further details. **F** FACS analysis from time course experiments on sorted subpopulations showing that $CD24^-$ cells return to original cell-state proportions in few days. $CD24^+$ cells, on the other hand, failed to achieve equilibrium proportion over the same period. Dotted lines represent percentage steady state levels of $CD24^+$ and $CD24^-$ cells in unsorted, parental cell line.

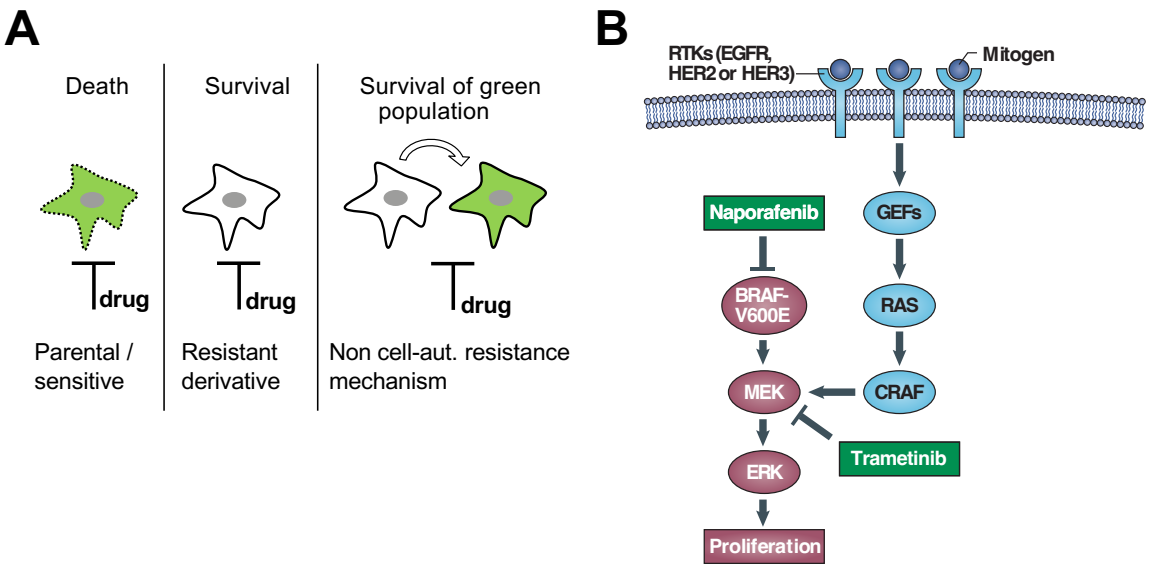
3.3 Defining non-cell autonomous cooperative mechanisms of drug protection

3.3.1 Experimental design and treatment of A375 melanoma cell line

We performed conditioned media experiments in breast cancer cell lines to understand whether factors secreted from a certain subpopulation may influence the other one(s), thus shaping the intricated population ecosystem as a whole, by measuring cell-state proportions. Nonetheless, our studies did not rule out the potential of paracrine control on stable cell-states. However, paracrine signals are responsible for a vast number of cellular processes. In a heterogeneous environment, such as the cancerous one, distinct subclones may release unique signaling molecules that may influence all tumor features, from proliferation to apoptotic evasion, and possibly fostering drug resistance.

A major challenge in cancer therapy is the insurgence of drug resistance (Vasan *et al*, 2019). One explanation for such phenomenon is the positive selection of pre-existing drug-resistant subpopulations carrying advantageous genetic alterations. While such clones can survive and proliferate under drug pressure, the overall resulting drug-resistant tumor subpopulations that result are not necessarily an identical genetic replica of the initial clone, boosting the heterogeneity of tumor. This suggests that forms of non-cell autonomous, cooperative drug protection transmitted by the original resistant clone to the rest of the sensitive cell population might exist. Therefore, to address this issue, we moved to a different cell model in order to identify and characterize potential cell communication mechanisms of drug resistance, when a resistant cell confers drug protection to a sensitive one when co-cultured together (**Figure 3.22 A**), in a system that mimics the cellular heterogeneity within a tumor. First, for this new experimental aim, we needed to generate drug resistant cells, and the tagging of sensitive cells with a fluorescent protein (e.g., GFP). Color tagging would be helpful to visualize and determine the proliferation rate of sensitive cells over the resistant ones, when they are plated in co-culture under a drug treatment regime. Once we have assessed drug protection effects, the next step involved the identification of mediators of drug resistance that act in non-cell autonomous

fashion, or, possibly, in an autocrine way. For our experiments, we chose to employ BRAF mutant colorectal cancer cell lines, because drug resistance derivatives can be easily generated. Specifically, we aimed to generate resistant cells by continuous treatment with a drug combination targeting key nodes of the MAPK signaling pathway, RAF and MEK, targeted with Naporafenib and Trametinib respectively (**Figure 3.22 B**). Naporafenib is a selective B/C-RAF inhibitor, now in Phase II clinical trials (Ramurthy *et al*, 2020). Trametinib is a selective MEK 1/2 inhibitor, used to treat melanoma (Gilmartin *et al*, 2011). Since advanced melanoma patients are routinely treated with these drugs, in a first set of experiments, we tested the toxicity of Trametinib and Naporafenib in A375 melanoma cell line, an established model for toxicology studies, characterized by BRAF V600E mutation and sensitive to MAPK inhibitors treatment (Wang *et al*, 2018a). Therefore, we treated A375 cells for 72 hours with increasing concentration of either Trametinib (0, 4, 8, 16, 32, 62.5 nM), Naporafenib (0, 0.625, 1.25, 2.5, 5, 10 μM) or a combination of the two drugs (**Figure 3.23 A**) and then we checked cell viability. Remarkably, we observed that A375 cells were particularly responsive to both single (**Figure 3.23 B**) and combination drugs (**Figure 3.23 C**), supporting the efficacy of the treatment.



Adapted from Holderfield et al. 2014

Figure 3.22 – A Schematic of experimental design for identification of non-cell autonomous resistance mechanism(s). See text for further details. **B** Schematic depicting the role of Naporafenib and Trametinib in inhibiting the MAPK cascade. Specifically, Naporafenib is a highly specific panRAF inhibitor while Trametinib, acting on the same signaling cascade, is an inhibitor of MEK kinase. The concurrent inhibition of different points within the same signaling cascade has a synergistic effect in the treatment of target cell, thus affecting proliferation and survival.

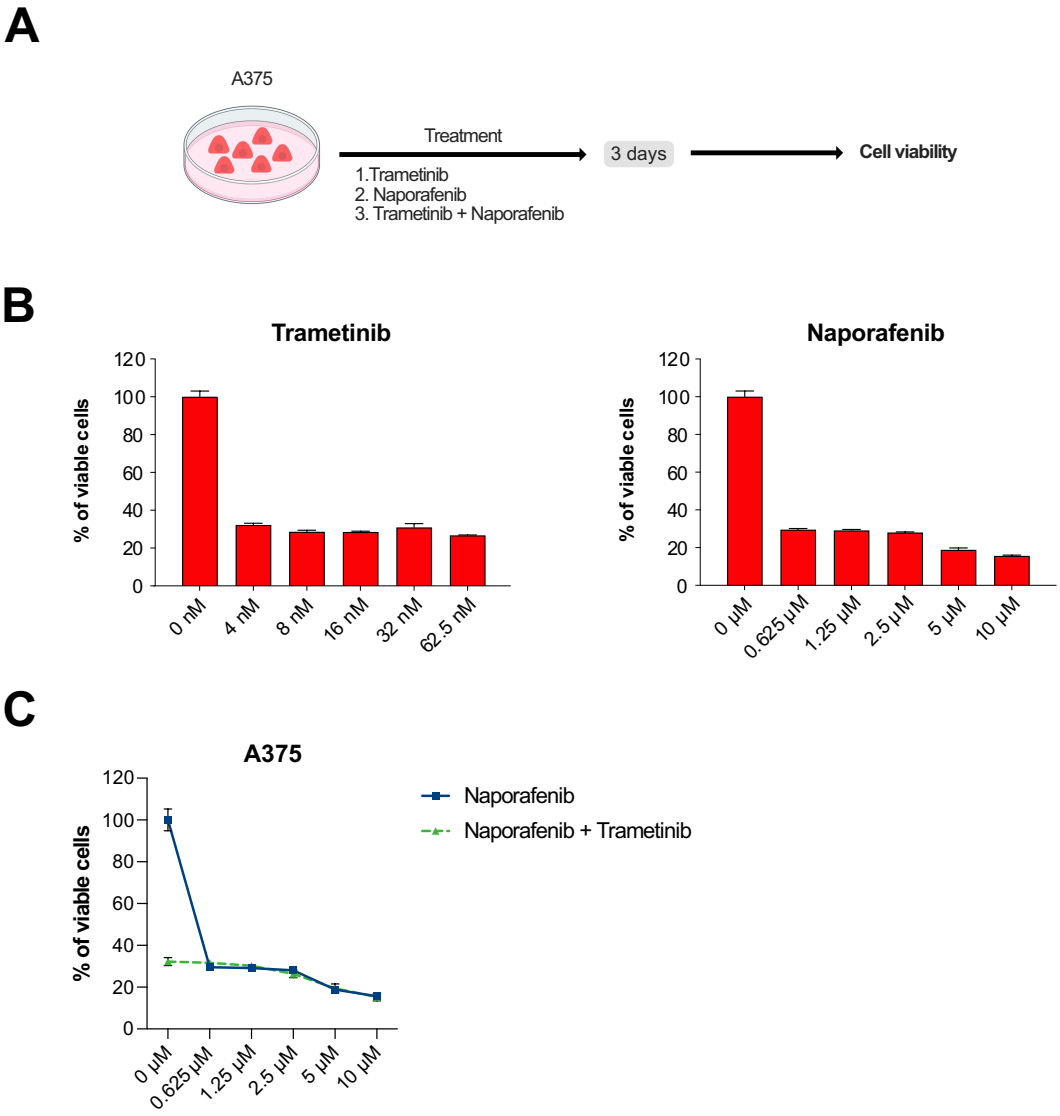


Figure 3.23 – A Schematic of experimental procedure for A375 treatment with MAPK inhibitors. See text for further details. **B** Barplots showing cell viability of A375 cells treated with increasing concentrations of Trametinib or Naporafenib. A375 cells are highly sensitive upon single treatment. **C** Viability curve showing cell viability of A375 cells treated with a combination of Trametinib and Naporafenib, at increasing concentration. A375 cells are highly responsive after combination treatment. Barplots and viability curve are represented with Standard Deviation (SD). n = 3.

3.3.2 Characterization and generation of resistant colorectal cancer cells

After assessing the efficacy of Naporafenib and Trametinib on A375 cells, we moved to RKO, a colorectal cancer cell line characterized by cells with an epithelial appearance and by BRAF V600E mutation (**Figure 3.24 A**). As in A375 cell line, we treated RKO with increasing concentration of both Trametinib (0, 3 nM) and Naporafenib (0, 150, 300, 600 nM) and then we assessed cell viability after 72 hours of treatment (**Figure 3.24 B**). We restricted Trametinib to 3 nM because cells achieved a plateau effect as concentration increased. Remarkably, we noted that RKO cells were sensitive to the combination of the two drugs, with a drop of ~ 50% of viable cells after three days of continuous treatment (**Figure 3.24 C**). Therefore, prompted by the sensitivity of RKO cells to treatment, we decided to generate RKO drug-resistant derivatives by long-term culture in the continuous presence of both inhibitors (**Figure 3.25 A**). To induce resistance, we used 3 nM Trametinib in combination with 300 nM Naporafenib. Colonies of resistant cells, that we renamed RKOR, began to appear over time as indicated by white asterisks in (**Figure 3.25 B**). Cell viability assays were performed in RKOR cells, showing that these cells were highly resistant to combination of MAPK inhibitors compared to sensitive population (**Figure 3.25 C**).

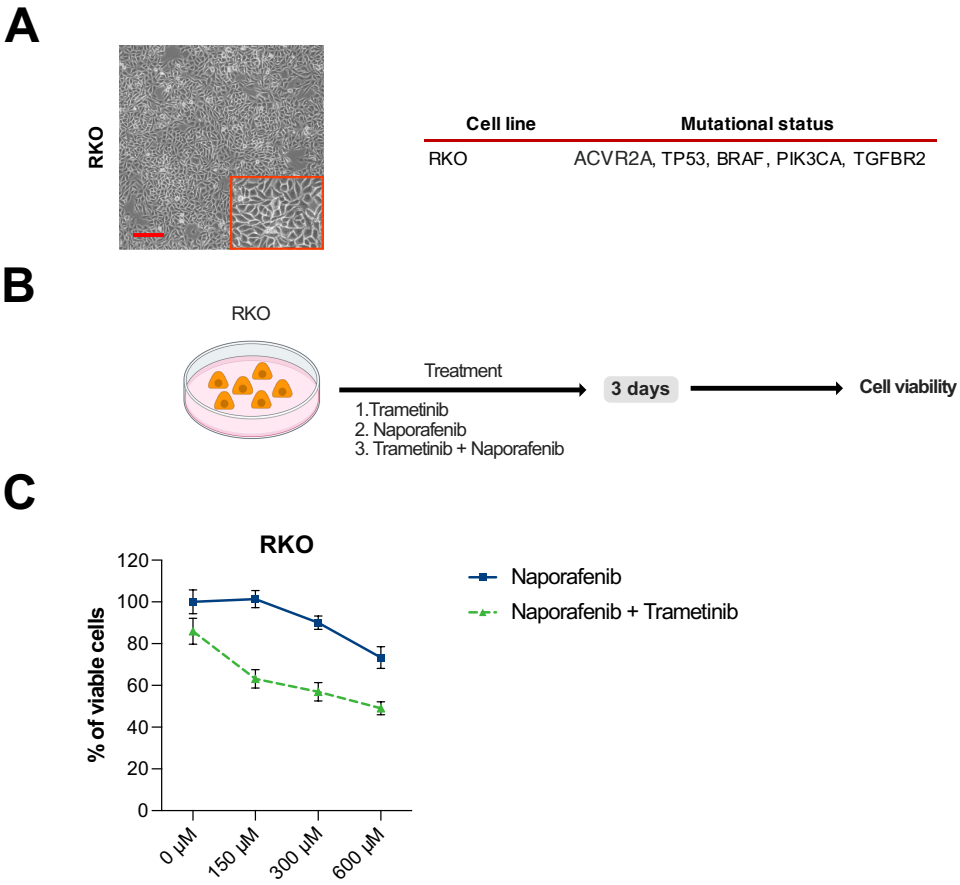


Figure 3.24 – A Brightfield (with 2X magnification in the corner) images showing RKO cells in culture (*right*) and a table reporting the mutational status the cell line (*left*). Scalebar 100 µm. **B** Schematic of experimental procedure for RKO treatment with MAPK inhibitors. See text for further details. **C** Viability curve showing cell viability of RKO cells treated with increasing concentrations of Naporafenib. Trametinib concentration was kept at 3 nM. RKO cells are sensitive to treatment after 72 hours with MAPK inhibitors. Viability curve represented with Standard Deviation (SD). n = 3.

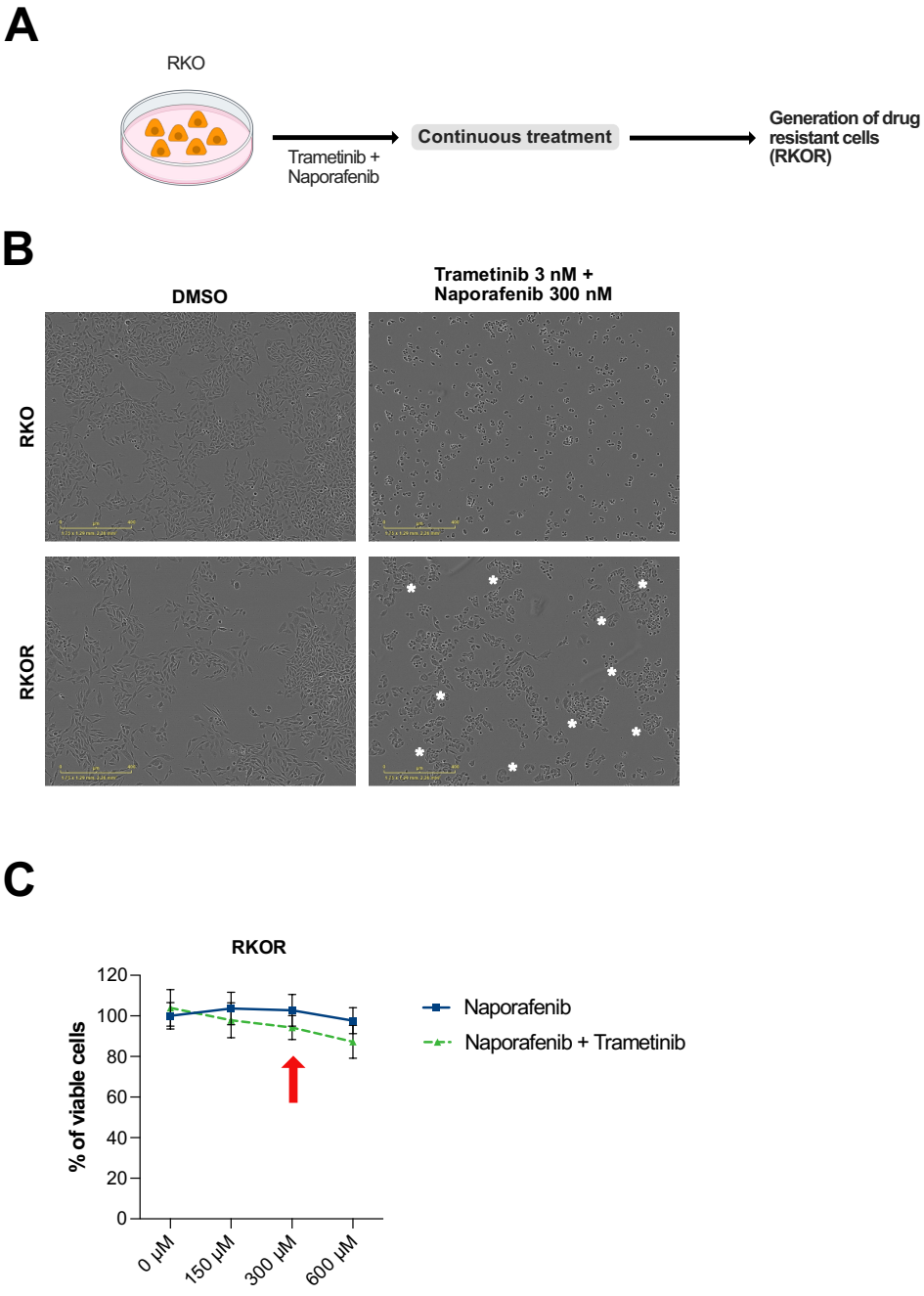
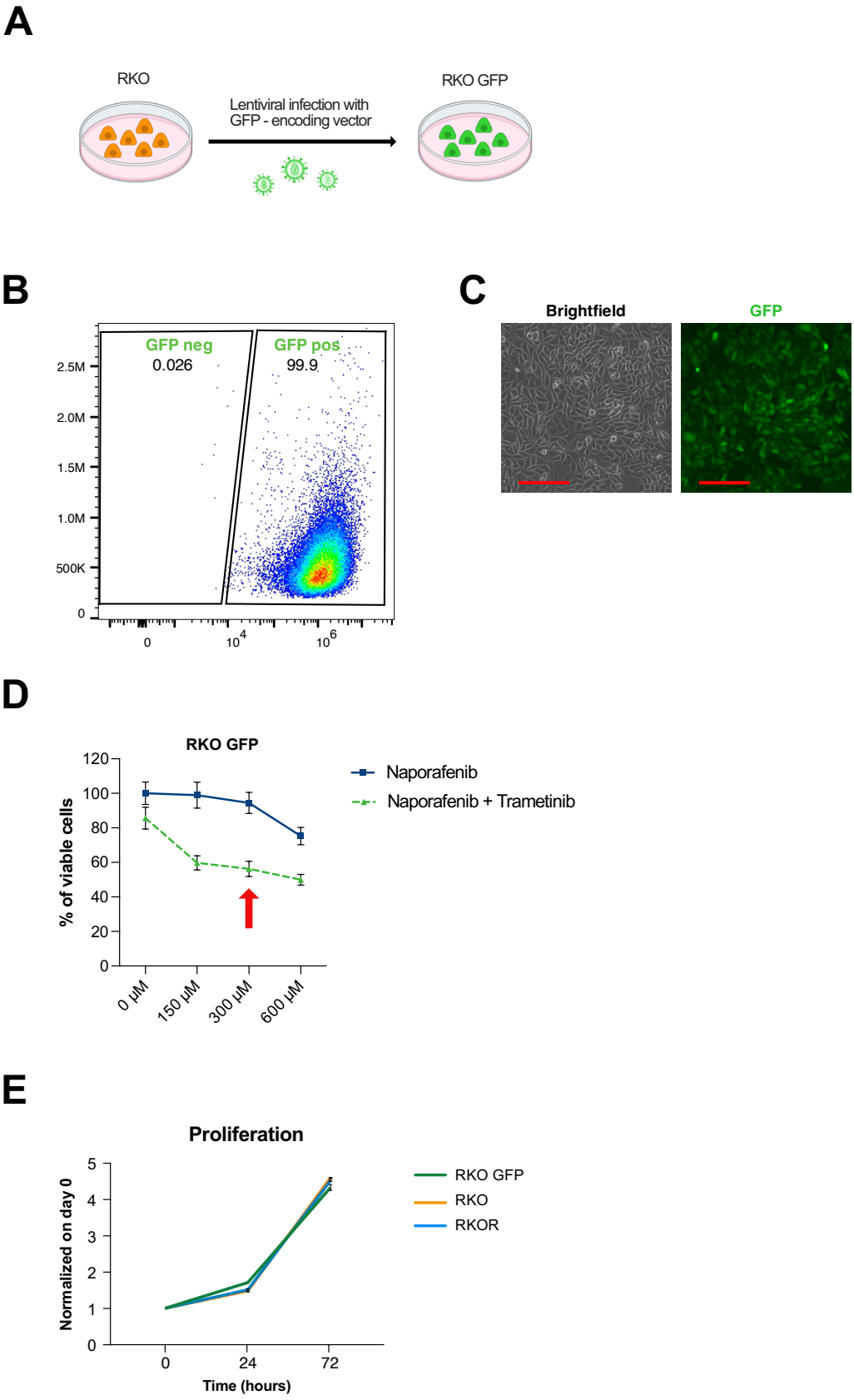


Figure 3.25 – A Schematic of experimental procedure for generation of RKO resistant cells after treatment with MAPK inhibitors. See text for further details. **B** Incucyte® microscope images showing growing colonies of RKO resistant cells (RKOR) after several weeks of continuous treatment with 3 nM Trametinib and 300 nM Naporafenib. Scalebar 400 µm. **C** Viability curve showing cell viability of RKOR cells treated with increasing concentrations of Naporafenib. Trametinib concentration was kept at 3 nM. RKOR cells are insensitive to treatment after 72 hours with MAPK inhibitors. Red arrow pointing at the concentration used to derive resistant cells. Viability curve represented with Standard Deviation (SD). n = 3.

3.3.3 Generation of GFP-tagged sensitive RKO cells

In order to actively monitor the growth of sensitive cells when co-cultured with resistant derivatives, we chose to label RKO cells permanently with a fluorescent protein. To this aim, we infected sensitive RKO cells with a vector encoding for GFP and puromycin resistance cassette and selected with antibiotic treatment for 72 hours; we called the generated reporter cell line RKO GFP (**Figure 3.26 A**). We assessed expression of GFP protein via FACS analysis, that unveiled the presence of a unique population of cells expressing GFP (**Figure 3.26 B**). We further confirmed this result by fluorescent microscopy (**Figure 3.26 C**). We wanted to investigate whether the RKO GFP cells were as sensitive as the parental RKO cells, therefore we measured cell viability three days following MAPK inhibitors administration. As in RKO cells, the parental derivative displayed marked sensitivity to the combination of Trametinib and Naporafenib, also with the concentration used for generation of RKOR cells (3 nM Trametinib + 300 nM Naporafenib), indicated by the red arrow (**Figure 3.26 D**). To exclude random insertion effects related to lentiviral DNA integration, we performed a short-term proliferation experiment and found that RKO and RKO GFP cells had the same growth rate (**Figure 3.26 E**).



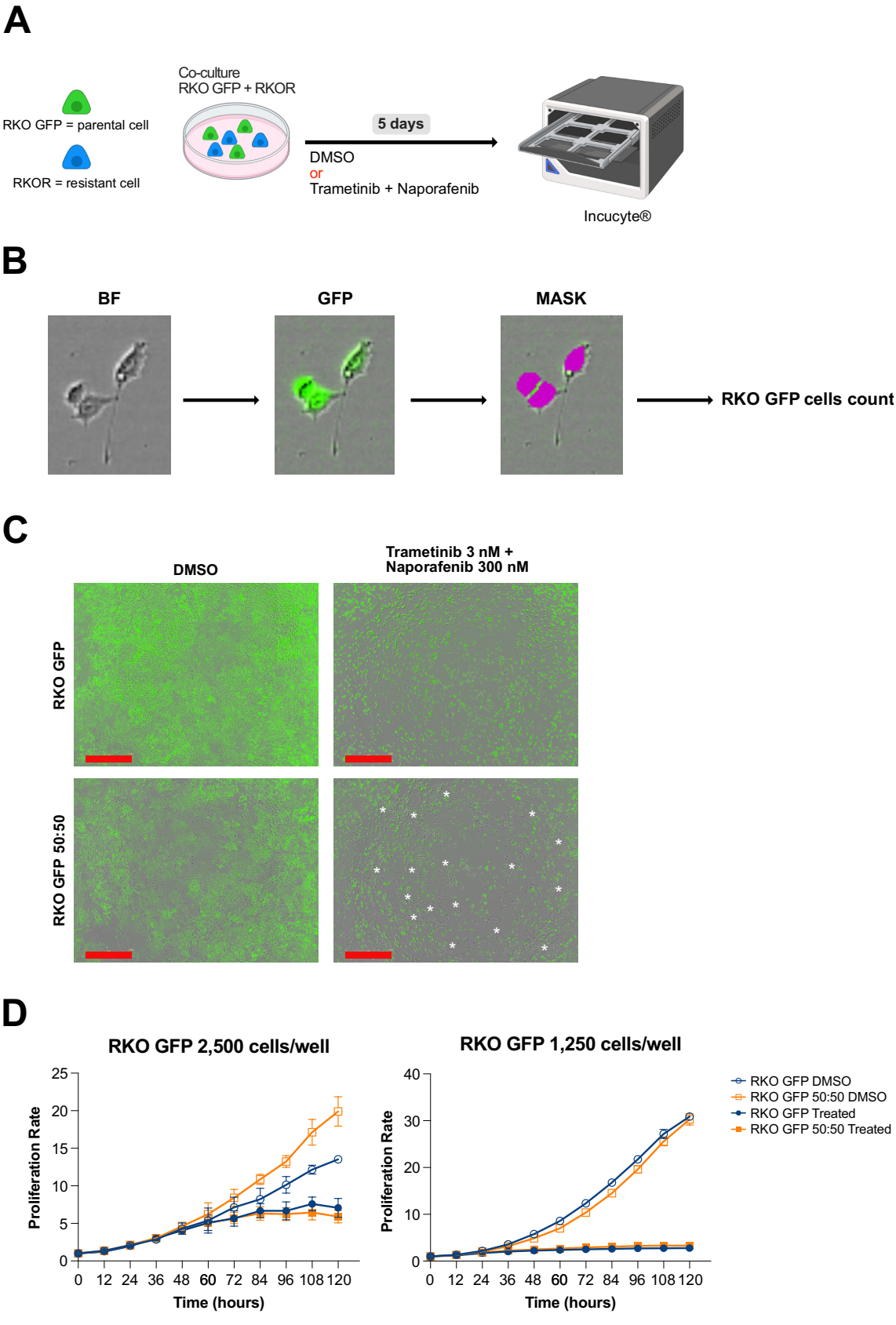
See next page for caption.

Figure 3.26 – A Schematic of experimental procedure for generation of RKO GFP cells with lentiviral infection of a GFP-encoding vector. See text for further details. **B** Representative flow cytometry analysis reporting percentage of GFP^+ cells**. C** Brightfield and fluorescence microscope images showing RKO GFP cells in culture (*right*) and with GFP fluorescence. Scalebar 100 µm. **D** Viability curve showing cell viability of RKO GFP cells treated with increasing concentrations of Trametinib or Naporafenib. RKO GFP cells are sensitive to treatment after 72 hours with MAPK inhibitors. Red arrow pointing at the concentration used to derive RKOR cells. Curve represented with Standard Error of the Mean (SEM). n = 3. **E** Proliferation assay showing that RKO GFP and RKOR cells have the same proliferation rate of the parental RKO cell line. Plot represented with Standard Deviation (SD), n = 3.

3.3.4 Co-culture of RKO and RKOR cells has no effect on sensitive cells viability upon MAPK inhibitors treatment

Stemming by the premise that the resistant derivative may confer a drug protection towards the sensitive population, we hypothesized that the mechanism regulating this protective phenotype could be determined by extrinsic factors, such as paracrine influence, or even juxtracrine factors, among cells. Therefore, to test the *paracrine* hypothesis, that is, that some factors secreted from RKOR cells may have a key role to promote survival of RKO GFP, we performed co-colture experiments. In this experiment, sensitive RKO GFP cells and resistant RKOR cells were plated separately and in a 50:50 combination, followed by five days of continuous treatment with MAPK inhibitors. DMSO-treated cells were used as control. Cell proliferation was monitored with the Incucyte® S3 system. The main objective of this series of experiments is to assess the relative viabilit y and proliferation of sensitive RKO GFP cells (**Figure 3.27 A**). Therefore, we took advantage of the built-in software provided with the Incucyte® and we quantified the number of GFP cells by superimposing a mask over fluorescent cells, and then counting the number of objects identified (**Figure 3.27 B**), i.e. the number of RKO GFP cells. This allowed us to qualitatively evaluate the protective impact of the resistant population over the sensitive one. Thus, we plated and then treated RKO GFP cells and/or RKOR with either 3 nM Trametinib and 300 nM Naporafenib or DMSO, for five days. Representative images taken with the Incucyte® microscope at 72 hours upon treatment administration clearly shows the sensitivity of RKO GFP to MAPK inhibitors, and the presence of colonies of viable resistant RKOR cells (white asterisks) (**Figure 3.27 C**). When co-cultured with RKOR, we noticed that treated RKO GFP did not have a proliferation advantage (**Figure 3.27 D**), when seeded at both 2,500 and 1,250 cells/well. We plated two different cell densities to take in account possible confluency effect on the culture. Intriguingly, we found that RKO GFP cells plated at 2,500 cells/well in co-culture with RKOR have a slight but significant proliferative advantage over RKO GFP cells plated alone, in absence of MAPK inhibitors (*left*). Specifically, the proliferation difference between RKO GFP and RKO GFP + RKOR cells increased at ~ 60 to 72 hours from incubation. On the

contrary, RKO GFP cells seeded alone at 1,250 cells/well have no growth advantage over RKO GFP cells in co-culture with RKOR in absence of treatment (*right*). Overall, these data showed that co-culture of both sensitive and resistant did not confer an advantage to sensitive population when treated with MAPK inhibitors.



See next page for caption.

Figure 3.27 – A Schematic of experimental procedure for co-culture experiments within the Incucyte® system. See text for further details. **B** Imaging analysis strategy to count GFP^+ cells with the built-in software of the Incucyte® microscope. **C** Brightfield combined with fluorescence microscopy images showing RKO GFP and RKO GFP 50:50 (½ RKO GFP + ½ RKOR) treated with DMSO or 3 nM Trametinib and 300 nM Naporafenib. Scalebar 800 µm. **D** Proliferation assays showing that RKO GFP and RKOR cells have the same proliferation rate of the parental RKO cell line. Plot represented with Standard Deviation (SD), n = 3.

3.3.5 Experiments with conditioned media experiments did not prove the existence of protective effects from resistant cells

Co-culture of RKO and RKOR cells did not confer any viability advantage over sensitive population upon treatment. Therefore, we reasoned that co-culturing conditions were suboptimal for seeing protective effects, and we postulated that this system should be stressed. Thus, we asked whether switching media between the two cell lines could impact on the population survival. More specifically, we speculated that using the culturing medium derived from resistant population alone could reflect an enrichment of RKOR *putative* protective factors, thus conferring a positive impact on RKO GFP survival. Consequently, we grew RKO GFP and RKOR cells for three days in presence of DMSO, and then we collected the media (**Figure 3.28 A**). Medium derived from RKO GFP was called Conditioned Medium from Sensitive cells (CS); medium derived from RKOR was called Conditioned Medium from Resistant cells (CR) *(1)*. Then, we plated RKO GFP cell line and we added: 1) a combination of half fresh medium and half CS; 2) a combination of half fresh medium and half CR. Then, we seeded RKO GFP cells at both 2,500 and 1,250 cells/well densities, then treated with either DMSO and a combination of 3 nM Trametinib and 300 nM Naporafenib for five days *(2)*. We assessed RKO GFP survival by checking GFP positive cells proliferation with the Incucyte® microscope. We used fresh complete medium as control. In CS and CR media, we added half fresh media to counterbalance the effect of a possible exhaustion of RKO GFP and RKOR full media. Interestingly, we did not find any significant difference in GFP population survival cultured with the two types of derived media, when treated with MAPK inhibitors: sensitive cells simply stopped proliferating and eventually started dying (**Figure 3.28 B**). We observed the same behavior in cells seeded at both 2,500 and 1,250 cells/well densities, suggesting that cell density is not required, in our hand, for the acquisition of a protective phenotype. On the contrary, DMSO-treated cells displayed an overall increase in proliferation rate. RKO GFP cells exhibited a small but consistent difference in proliferation rate when grown in fresh medium, indicating that the reduction in proliferation rate is negatively affected by proportion of “exhausted” CS and CR medium. In conclusion, our

findings suggested that RKOR-conditioned medium had no protective effect on RKO GFP cells. Notably, in our assays, RKO GFP and RKOR cells were treated with DMSO to mimic a patient's drug-free environment and generate conditioned media. We questioned whether stressing RKOR cells in the presence of MAPK inhibitors would drive the selective production of paracrine molecules that cannot be generated in the absence of treatment. Therefore, RKO GFP and RKOR cells were grown for three days in the presence of Trametinib and Naporafenib, and the medium was then collected. (**Figure 3.29 A**). As before, medium derived from RKO GFP was called Conditioned Medium from Sensitive cells (CS); medium derived from RKOR was called Conditioned Medium from Resistant cells (CR) (**Figure 3.29 A 1.**). We plated RKO GFP cell line and we added: 1) a combination of half fresh medium and half CS; 2) a combination of half fresh medium and half CR. As in the first experiment, we seeded RKO GFP cells at 2,500 and 1,250 cells/well densities, then treated with either DMSO and a combination of 3 nM Trametinib and 300 nM Naporafenib for five days (**Figure 3.29 A 2.**) . Fresh complete medium was used as control, and GFP positive cells proliferation was assessed with Incucyte® microscope. We found that conditioned media derived from pre-treated cells did not confer any proliferative advantage when RKO GFP population is treated with MAPK inhibitors (**Figure 3.29 B**). Again, same outcomes have been achieved with densities of 2,500 and 1,250 cells/well. DMSO-treated cells grown in either fresh or CS media proliferate at the same rate. The CR medium cells had no growth benefit. To corroborate these results, in a set of new experiments, we collected conditioned medium from RKO cells – the sensitive parental population without the GFP – pre-treated with Trametinib and Naporafenib (**Figure 3.30 A**). As before, medium derived from RKO cells was called Conditioned Medium from Sensitive cells (CS). In parallel, we used medium derived from RKOR, recalled Conditioned Medium from Resistant cells (CR) (**Figure 3.30 A 1.**). We plated RKO at 2,500 and 1,250 cells/well densities and we added: 1) a combination of half fresh medium and half CS; 2) a combination of half fresh medium and half CR. Next, we treated cells with either DMSO and a combination of 3 nM Trametinib and 300 nM Naporafenib for five days (**Figure 3.30 A 2.**). Fresh complete medium was used as control. Instead of

assessing GFP positive cells proliferation, we evaluated cell viability by measuring ATP content with luminescence after 72 hours of treatment. We confirmed that pre-treatment of conditioned media with MAPK inhibitors had no effect on RKO cells survival (**Figure 3.30 B**). This occurred at both 2,500 and 1,250 cell densities. Overall, these experiments demonstrated that the addition of a conditioned medium enriched of secreted RKOR factors did not result in a protective phenotype.

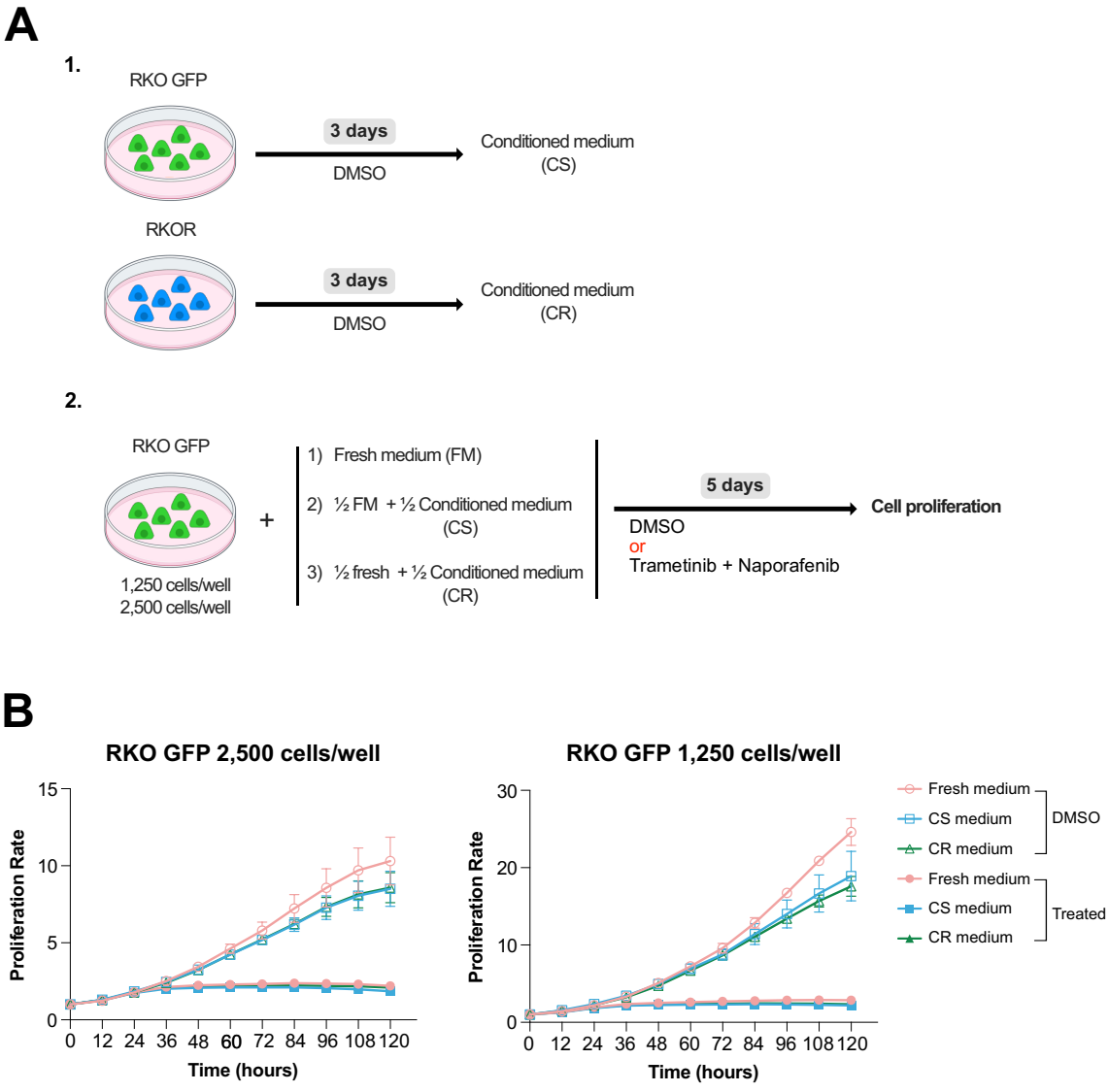


Figure 3.28 – **A** Schematic of experimental procedure for conditioned media experiments. See text for further details. **B** Proliferation assays showing that the addition of conditioned media obtained from RKO (CS medium) and RKOR (CR medium) did not provide a growth advantage to RKO GFP cells seeded at 2,500 cells/well (left) and 1,250 cells/well (right) when treated with DMSO and MAPK inhibitors. Data represented with Standard Deviation (SD), n = 3.

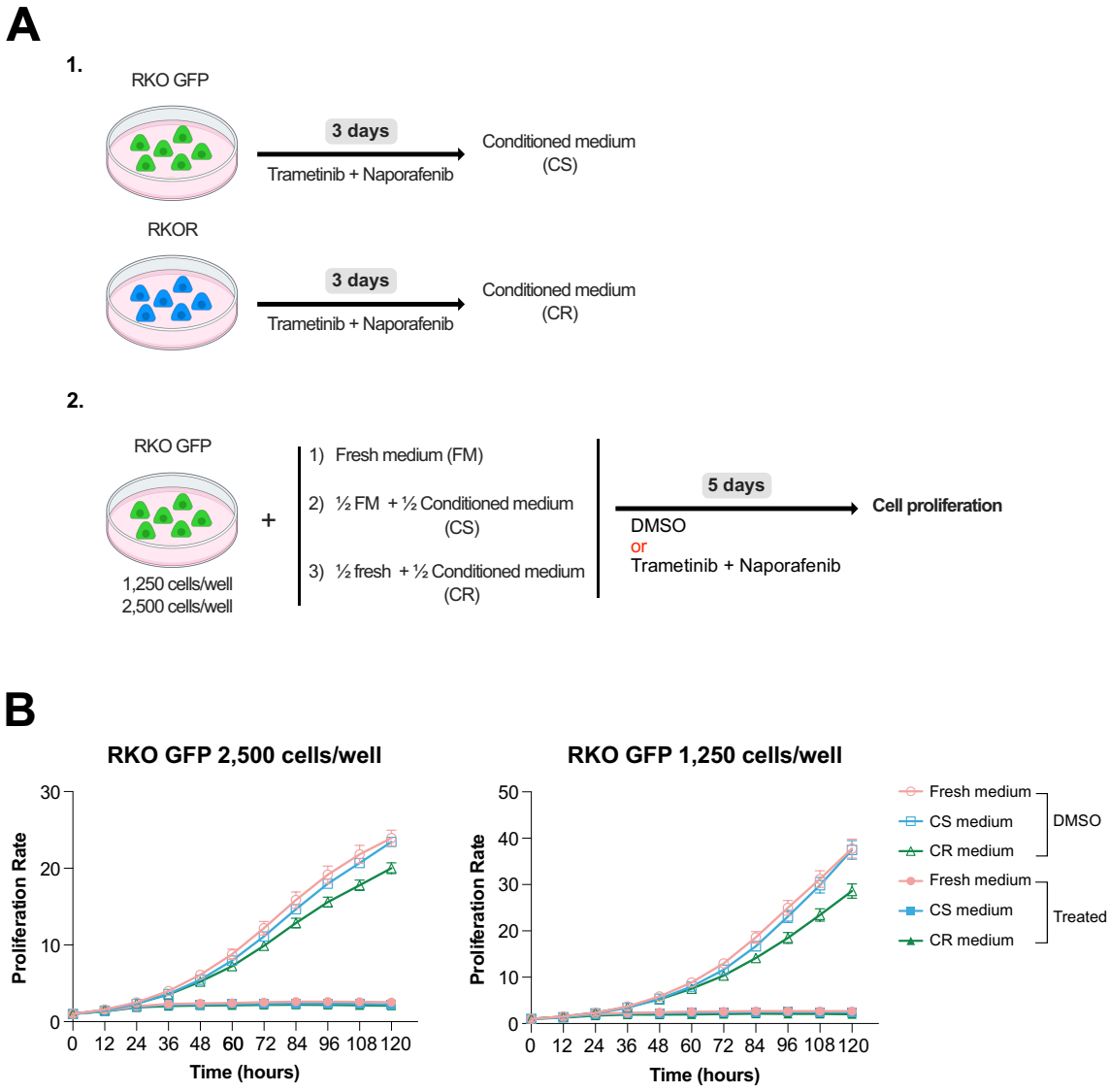
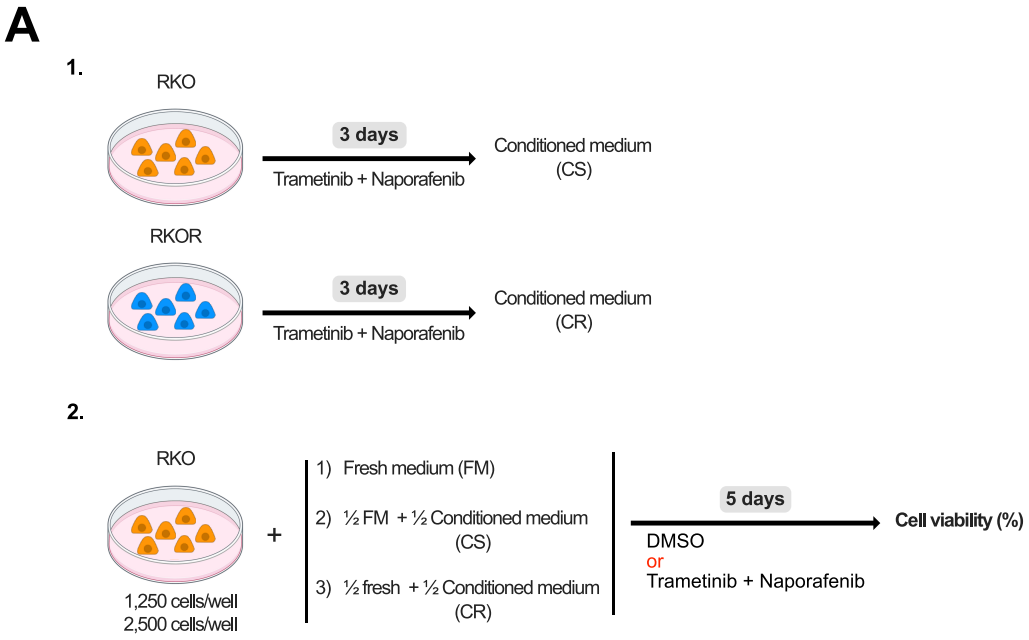
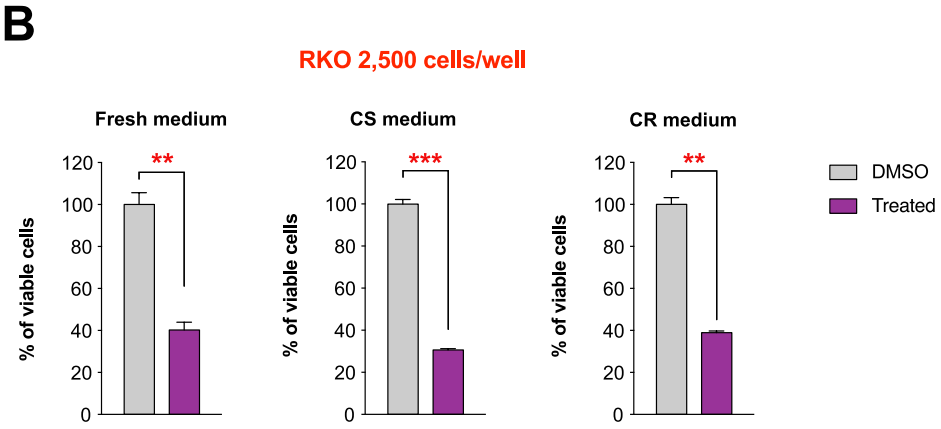
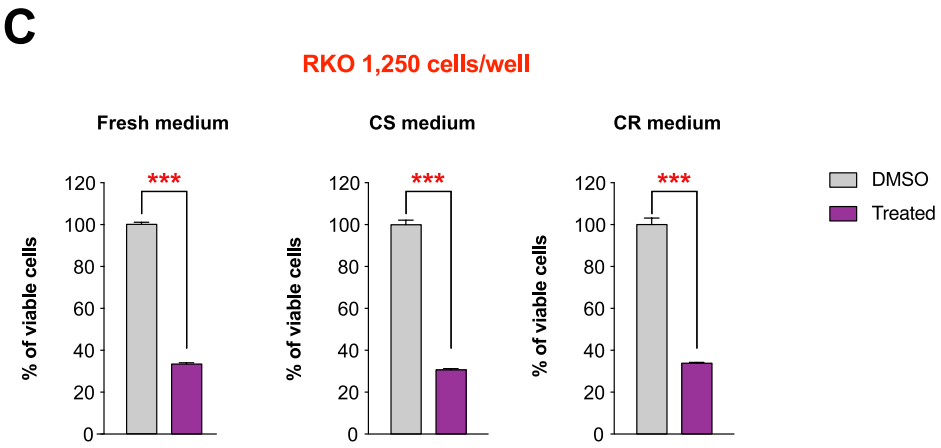


Figure 3.29 – **A** Schematic of experimental procedure for conditioned media experiments with pre-treated cells. See text for further details. **B** Proliferation assays showing that the addition of conditioned media obtained from RKO (CS medium) and RKOR (CR medium) pre-treated with a cocktail of Naporafenib and Trametinib did not provide a growth advantage to RKO GFP cells seeded at 2,500 cells/well (left) and 1,250 cells/well (right) when treated with DMSO and MAPK inhibitors. Data represented with Standard Deviation (SD), n = 3.







See next page for caption.

Figure 3.30 – A Schematic of experimental procedure for conditioned media experiments with pre-treated cells. See text for further details. **B-C** Proliferation assays showing that the addition of conditioned media obtained from RKO (CS medium) and RKOR (CR medium) pre-treated with a cocktail of Naporafenib and Trametinib did not provide a growth advantage to RKO cells seeded at 2,500 cells/well (B) and 1,250 cells/well (C) when treated with DMSO and MAPK inhibitors. Data represented with Standard Error of the Mean (SEM), n = 3.

4. Discussion

In this study, we sought to identify new mechanisms of cell-state interconversions, cancer cell plasticity and drug resistance. Briefly, using breast cancer cell lines, we established that cell-state transition is less than a 60 minutes-long process, likely governed by a rewiring of protein transport and localization pathways, as shown by gene expression profiling. Conditioned media and co-culture experiments were performed in order to define whether secreted factors were involved in the control of phenotypic equilibrium over time in breast cancer cells and, in parallel, of a drug-protective mechanism that could operate along the resistant-to-sensitive cell axis in a colorectal cancer model. Since the role of paracrine factor was not entirely tackled, future experiments along that lines are needed. In conclusion, our work expanded the understanding of the plasticity of cell-state transition within cancer cell lines. Here, we provide a thorough discussion about our findings.

At the beginning of the 2000s, Hanahan and Weinberg described the common traits shared by all cancers, the well-known Hallmarks of Cancer (Hanahan & Weinberg, 2000). Throughout the years, several updates have been made to this compilation (Hanahan & Weinberg, 2011). In the last iteration, *unlocking phenotypic plasticity* has been added (Hanahan, 2022). During development, most normal cells reach and retain a state of terminal differentiation, a consequence of their developmental fate. Conversely, a large body of literature has shown how cancer cells can unlock their differentiation status, thus interconverting among different cell-states (finely reviewed in (Yuan *et al*, 2019; Gupta *et al*, 2019)). In turn, these various kinds of plasticity foster intratumor heterogeneity. Cancer cell plasticity is not a feature restricted to patients. In fact, it has been reported how 2D cell culture still preserves a certain degree of subclonal complexities. Some examples include the generation of cancer stem cells (CSCs) from non-CSCs in basal breast cancer cell (Chaffer *et al*, 2013), the concomitant existence of different cancer cell-states in 2D culture of non-small cell lung cancer (Hayford *et al*, 2021) and breast cancer models (Fillmore & Kuperwasser, 2008; Gupta *et al*, 2011), and the phenotypic plasticity of colorectal cancer cell lines (Sacchetti

et al, 2021). In addition, subpopulations of cancer cells may maintain a condition of phenotypic equilibrium, in which the proportions of different differentiation states are kept at a constant ratio throughout a huge number of cell divisions. However, the mechanism behind such plastic balance remains largely unknown.

Here, we have confirmed how two triple negative breast cancer cell lines, SUM149 and SUM159, are composed by at least three distinct subpopulations, defined by the expression of different cell-surface markers, as reported also in (Fillmore & Kuperwasser, 2008; Gupta *et al*, 2011). How these subpopulations are maintained in equilibrium across cell divisions is an issue that was elegantly modeled by (Gupta *et al*, 2011), proposing that stochastic transitions between differentiation states are responsible for the regeneration of disrupted equilibrium proportion. In this dissertation, we demonstrated that subpopulations of stabilized breast cancer cell lines, once isolated from their respective parental population by flow cytometry cell sorting, are capable of reverting back to their initial frequencies not in timeframe of days, as (Gupta *et al*, 2011) and others have previously reported, but in a time range of minutes. Intriguingly, (Gupta *et al*, 2011) has predicted that these phenotypic transitions may display rapid dynamics, based on an *in silico* analysis of Markov simulations indicating that subpopulations of the SUM159 cell line may recover equilibrium proportions within a day. In time course experiments, we proved that subpopulations of SUM149 cell lines fractionated by FACS can return to equilibrium in significantly less than twenty hours. In particular, FACS re-analysis at specific timepoints indicated that the SUM149 basal subpopulation, which constitutes 3% of the whole parental population, switched to the luminal and stem-like subpopulations nearly immediately after sorting. All FACS sorting procedures were conducted at 4° C, as stated in the Results chapter, to avoid cells from immediately differentiating following isolation. In only one hour we could assist not only to the emergence of luminal and stem-like cells, but also a considerable reduction in basal cells proportion. Intriguingly, we noticed a rapid increase in the number of stem-like cells three hours after sorting, which stabilizes over time to reach the initial proportion, suggesting two hypotheses: 1) the stem-like state is an intermediary stage in the interconversion of basal to luminal cells; 2) the

interconversion between states is driven solely by stochastic processes. To date, we have been working to incorporate data produced from our time course into an expanded Markov model in order to get a deeper comprehension of stochastic transitions. Immunofluorescence experiments have further narrowed the time range of interconversion, indicating that just one hour is required for a significant increase in the number of luminal cells that have differentiated from basal subclones.

Here, we have demonstrated that EPCAM, one of the main cell-surface markers that discriminate each subpopulation, is precisely localized in the cytoplasm or on the cell membrane throughout the transition from the basal to the luminal state and *vice versa*. Notably, we were able to confirm this localization in both SUM149 and SUM159 cell lines, thereby suggesting that the fast modulation of protein localization, with EPCAM or other cell-surface proteins, may be a feature shared in several cancer cell lines. It would be interesting to investigate whether subpopulations of additional cancer cell lines derived from different tumors behave similarly once separated from the parental population. Immunofluorescence images from basal SUM149 cells stained with EPCAM antibody revealed that the EPCAM protein is already present in the cytoplasm, but not on cell-surface, suggesting that EPCAM is actively maintained in the cytosol, removed from cell membrane and *primed* to re-localize to plasma membrane in response to an unidentified signal. The nature of this signal is yet to know. As SUM149 cells are adherent cells that actively attach to a substrate, which is suggestive of their epithelial origin, we wondered whether this epithelial origin had a role in the localization of cell-surface proteins. Consequently, by doing experiment in which cells were prevented from adhering to the dish, we established that the quick interconversion is independent of the ability to attach to substrate.

To characterize cell-states and investigate their phenotypic transition, we have demonstrated that the localization of cell-surface markers is precisely and rapidly controlled during the interconversion between states. Thus, we asked whether, and possibily how, gene expression was affected in purified differentiation states. Therefore, we assessed *EPCAM* and *CD24* expression levels in basal and luminal

cells via qPCR, immediately and after one hour from isolation. Intriguingly, basal cells completely recovered *EPCAM* and *CD24* at expression levels proper of luminal cells. Moreover, these findings demonstrated that *EPCAM*, a well-known epithelial marker, is expressed at extremely low levels in cells with a mesenchymal-like phenotype, such as basal cells. This is consistent with FACS profiling of SUM149 cell lines, which revealed that basal cells lack EPCAM expression on the cell surface. Since FACS analyses were performed on non-permeabilized cells, it was evident that EPCAM was only present on the cell membrane. Here, we demonstrated that, despite the fact that *EPCAM* mRNA expression is constantly repressed in cells in their basal state, compared to the luminal counterpart, the translated protein is present and stored in the cytoplasm. EPCAM is a transmembrane cell adhesion molecule discovered in colorectal cancers (Herlyn *et al*, 1979), expressed practically in all epithelium and employed as a biomarker in the majority of carcinomas since it is overexpressed in many tumors and expressed in tumor-initiating cells (Schön *et al*, 1994; Winter *et al*, 2003; Slanchev *et al*, 2009; van der Gun *et al*, 2010; Imrich *et al*, 2012; Tsaktanis *et al*, 2015). Recently, it has been shown that EPCAM undergoes endocytosis and receptor recycling during mesodermal development and TGFβ-mediated EMT (Pan *et al*, 2021). Using RNAseq, we identified different upregulated pathways involved in *protein transport* and *retrograde transport, endosome to Golgi* in purified SUM149 basal cells. Among them, *RAB5A* and *RAB7A*, with the former involved in the control of membrane-to-early endosome and the latter in the regulation of endo-lysosomal trafficking. Interestingly, Pan et al. demonstrated that EPCAM physically interacts with RAB5 and RAB7 when subjected to endocytosis and recycling. In addition, they indicated that RAB11, a protein involved in recycling endosomes, that is, the pathway that allows the re-localization to cell membrane of receptors or, in general, transmembrane proteins, co-localizes with EPCAM, indicating an active route of trafficking that translocates this cell-surface protein. In addition, our RNAseq analysis pointed out that SUM149 cells in luminal state displayed a downregulation in pathways involved in clathrin-mediated endocytosis and COPII-coated vesicle-mediated transport, suggesting that in luminal cells the transport from the cell surface to

the interior is inhibited, at a steady state level. These observations support a hypothetical model in which SUM149 basal cells actively and promptly reprogram EPCAM localization by intensive trafficking relocalization. We may hypothesize that the same occurs in SUM159 luminal cells, in which EPCAM is actively removed from the cell membrane and relocalized to the cytoplasm; nevertheless, further experiments are needed along this line. Interestingly, it would be fascinating to demonstrate whether RAB family components co-localize with EPCAM in our cells using confocal microscopy. However, since the specific distribution of EPCAM in the cytoplasm of SUM149 and SUM159 basal cells remains uncertain, confocal microscopy coupled with specific markers addressing cell organelles and compartments (endosomes, lysosomes, Golgi, endoplasmic reticulum) could help defining the specific localization of this protein. Based on our RNAseq data, we may hypothesize that EPCAM is localized in the Golgi apparatus, as SUM149 basal cells exhibited an upregulation of genes involved in *retrograde transport, endosome-to-Golgi* pathway. Nonetheless, immunofluorescence experiments against EPCAM and Golgin-97 or GRASP65, markers of the Golgi apparatus, will be crucial to clarify this issue. We are also conducting experiments using inhibitors of clathrin-mediated endocytosis (e.g. PitStop® 2) to interfere with the relocalization process, with the goal of determining whether blocking trafficking may interfere with phenotypic plasticity.

Based on our RNAseq analysis, the regulation of EMT is one of the most differentially regulated pathways in basal and luminal cells. This is consistent with a vast body of literature (Perou *et al*, 2000; Sørlie *et al*, 2001; Dontu *et al*, 2003; Neve *et al*, 2006; Shipitsin *et al*, 2007; Prat & Perou, 2011; Koboldt *et al*, 2012; Barnabas & Cohen, 2013; Prat *et al*, 2013) and with the canonical characterization of basal cells as being in mesenchymal state, and conversely, luminal cells as being in epithelial state. Using the expression of the EMT gene family as a proxy, we found that after one hour from segregation, the purified SUM149 basal fraction restored luminal transcriptional levels, indicating that not only *EPCAM* transcript expression changed, but also a full transcriptional rewiring is involved in the transition to the luminal state in such restricted time

frame. These data indicate that the SUM149 cell line exhibits a high degree of plasticity. It would be crucial to expand RNAseq analyses to include pure subpopulations obtained at other timepoints, such as one hour after sorting, in order to have a more unbiased and exhaustive characterization of the phenotypic interconversion and to confirm the extensive transcriptional rewiring that we observed. Other works have demonstrated how $EPCAM^{high}$ and $EPCAM^{low}$ subpopulations in the heterogenous breast cancer cell line PMC42-LA displayed a plastic behavior when fractionated from parental population, with return to equilibrium proportion (Bhatia *et al*, 2019). Still, the timing to reach the equilibrium is quite long: 6- to 8-weeks after isolation. To date, we have not been able to find studies reporting such a quick interconversion between states, thereby our data provide a novel trait in cell-state transition phenotype.

How does the transition happen so quickly? We surmise various possible explanations. One possibility could relate to paracrine signaling acting among cells, fueling differentiation from one state to another. Notwithstanding, the impact of paracrine factors may be minor, in our opinion, for the following reasons: first, after sorting, the SUM149 and SUM159 subpopulations are relatively pure (almost 99% purity), and when sorted and plated alone, there is no chance to have secreted factors from the other cell-state. Second, given the high velocity to reinstate the equilibrium proportions, a cell signaling cascade acting so rapidly would be rather implausible. Obviously, differential proliferation rate could not be an issue since there is no time for a cell to divide and perform a complete cell cycle, an observation that also (Gupta *et al*, 2011) made.

Activity of bursting genes (Wang *et al*, 2019; Tunnacliffe & Chubb, 2020) leading to transcriptional noise (Sánchez & Kondev, 2008) may have a role in the interconversion; anyway, fluctuations of this kind may be small and transient but not stable enough for maintaining the equilibrium proportion over time (Ferrell, 2002; Losick & Desplan, 2008). In fact, stability requires the establishment of bistable systems, characterized by two stable states that are resistant to mild perturbations and may thus endure for extended time. As explained in the Introduction (*see §1.1.2 Non-genetic heterogeneity for further details*), bistable

systems rely on a type of cell memory called hysteresis: as we have already seen, this is normally sustained by paracrine feedbacks or by epigenetic signature. Therefore, it can be hypothesized the existence of either a paracrine signaling network and/or an epigenetically regulated *cell memory* that could be responsible for the cell-state transitions. Given the high speed of transition from purified basal SUM149 cells toward luminal (and partly, stem-like) cell-state, our data seem to suggest that the transition of basal cells might not be driven by paracrine stimuli but rather a sort of cell-autonomous program that pushes them toward luminal differentiation as soon as they are isolated. We speculated that this cell-autonomous program may be directed by the activity of poised genes, making them accountable for the interconversion of cell-states. Intriguingly, Chaffer et al. have elegantly demonstrated that breast cancer non-cancer stem cells interconvert to cancer stem cells upon environmental stimuli. This plasticity is mediated by ZEB1, whose promoter is maintained in a bivalent chromatin configuration (Chaffer *et al*, 2013). In particular, they have analyzed H3K4me3, H3K27me3, and H3K79me2 histone modifications on the promoter site via ChIP-qPCR. Other works have captivatingly addressed the link between chromatin remodeling and plasticity (Liau *et al*, 2017; Pastore *et al*, 2019; Zhang *et al*, 2022). For all these reasons, we posit that also in our system the transition among cell-states may be preceded and accompanied by epigenetic changes. By taking advantage of a new and efficient epigenomic profiling technique (Kaya-Okur *et al*, 2019), we are planning to perform Cut&Tag experiments on sorted SUM149 subpopulations, taken at different time point, by assessing the chromatin compaction marked by histone modifications, such as H3K4me3 and H3K27me3 for promoter regions (Voigt *et al*, 2013). In parallel, expression profile information will be obtained via bulk RNAseq so that each cell-state information, for both basal and luminal, will be matched with the epigenetic status derived from Cut&Tag data. Notably, since the phenotype is reverted in one hour, we need to take timepoints within 60 minutes.

It is possible that phenotypic proportions could be maintained in *continuous* cell culture by means of the intercellular paracrine networks. To test this hypothesis, we performed co-culture experiments in SUM149 and SUM159 cell

lines by selectively combining their media. We took advantage of the fact that each of these cell lines is dominated by a subpopulation that is in the minority in the other cell line. We did not find any variations in the proportions of cell states upon switching the medium. Due of the rapid rate of interconversion of purified minoritarian subpopulations (e.g. basal population in SUM149 cells), it is impossible to employ authentic conditioned media generated from the same cell line; this is an evident limitation in our assays. An alternative approach could be a proteomic analysis on media from SUM149 and SUM159 to detect differentially paracrine factors; this approach could be paired with an analysis of our RNAseq on isolated subpopulations to determine secreted cues based on their gene expression. Furthermore, experiments involving dialytic chambers may provide useful insights: for instance, utilizing dialysis for refreshing the culture medium could provide a practical solution, as it enables the exchange of low molecular weight substances like waste and/or nutrients, while preserving, or retaining, high molecular weight elements such as growth factors utilized in paracrine interactions. Finally, increasing the number of SUM149 basal cells artificially, perhaps through TGFβ supplementation in the culture medium, may present an opportunity to create genuinely conditioned media suitable for experiments involving conditioned media.

In this discussion, we have tried to answer to the following questions: how do breast cancer cells interconvert among cell-states and what is the mechanism? How is the equilibrium maintained across time? In order to address these issues, we needed to design a tool to follow the transition in cell culture. We generated SUM149 reporter cell line by performing CRISPR/Cas9 knock-in to insert a fluorescent tag downstream of EPCAM gene. This tool would allow us to detect when an $EPCAM^{+}$ cell interconverts into an $EPCAM^{-}$ cell, as a proxy of the basal-to-luminal transition and *vice versa*. By taking advantage of this tool, we will analyze interconversion rates among the lineage of single-cell clones, by plating sorted subpopulations in a single-cell format. In addition, we will determine whether the interconversion rates differ in individual cancer cells of any sorted subpopulation, by tracking cells growing in colonies and increasing or decreasing in the number of fluorescent populations as a readout. These experiments will

allow the tracking of phenotypic cell-state behaviors directly in culture and to closely follow the timing of the phenotypic switch.

What is the biological significance of such heterogenous and plastic population? We posit that the presence of various phenotypic populations within a tumor favors overall cancer fitness. To corroborate this hypothesis we propose the following strategy, for future experiments. The first involves the genetic ablation of populations expressing cell-state defining genes. This approach is used to verify whether specific cell subsets contribute to tumor progression, growth and fitness increase (Chen *et al*, 2012). Interestingly, Sonzogni et al. provided a well-designed method to ablate breast cancer subpopulations by attaching TK (Thymidine Kinase) to Cytokeratin 14 (K14) to selectively target basal cells, or to Cytokeratin 8 (K8) to selectively target luminal cells (Sonzogni *et al*, 2018).

In this work, we have proved how SUM149 and SUM159 cells are a heterogeneous ecosystem in which cells in various cell-state co-exist in equilibrium. Using scRNAseq analysis on deposited datasets, we investigated additional breast cancer cell lines and found that all of them exhibited a degree of transcriptional heterogeneity, independently of their subtype classification. We investigated EPCAM, CD24, and CD44 expression in an effort to identify subpopulations comparable to SUM149 and SUM159. As we expected, not all cell lines expressed these markers, in accordance with their differentiation state (mesenchymal or epithelial cell-state) (Gambardella *et al*, 2022). These results were validated by FACS analysis, demonstrating how transcriptome analysis and protein expression converge to corroborate the heterogeneity of the investigated cell lines. We focus our attention to the triple negative BT549 cell line, characterized by heterogeneity of CD24 expressions. Purified $CD24^-$ and $CD24^+$ subpopulations exhibited a plastic behavior: when plating pure $CD24^-$ and $CD24^+$ cells, both subpopulations arise, although with distinct dynamics. $CD24^-$ cells have been extensively characterized as a compartment enriched of cancer stem cells (Al-Hajj *et al*, 2003; Fillmore & Kuperwasser, 2008; Mani *et al*, 2008; Chaffer *et al*, 2013). This characterization can explain the transition from $CD24^-$ to $CD24^+$; nevertheless, $CD24^+$ giving rise to a pool of $CD24^-$ cells support the

existence of a mechanism of interconversion similar to SUM149 and SUM159, albeit involving different phenotypic subpopulations. Intriguingly, purified $CD24^+$ cells interconvert to $CD24^-$ with a slower interconversion rate compared to the other subpopulation. Here, we propose several lines of hypotheses: 1) pure $CD24^+$ cells are not able to regenerate the equilibrium proportion of the parental population; 2) there are differences in proliferation rate of $CD24^+$ and $CD24^-$ that could explain the delay of the complete phenotypic reversal: $CD24^+$ may need more than 21 days to completely interconvert. To properly describe the two subpopulations, more gene expression experiments (bulk RNAseq, scRNAseq) are needed. Interestingly, scRNAseq experiments could be coupled with barcoded cells to follow lineages, thereby inferring possible differentiation dynamic along cell progenies. Unfortunately, as previous results have shown, SUM149 and SUM159 subpopulations transit within different cell-states in a timeframe incompatible with the biological course of a cell division, so this approach could be unfeasible.

It would be also interesting to expand our phenotypic and plasticity analyses on cancer cell lines of different tumors. In 2021, Sacchetti et al. demonstrated how two colorectal cancer cell lines, HCT116 and SW480, are characterized by a plastic population composed of $EPCAM^{high}$ and $EPCAM^{low}$ subpopulations (Sacchetti *et al*, 2021). The two subpopulations have distinct phenotypic features: $EPCAM^{low}$ cells are motile with highly propension to metastasize, whereas $EPCAM^{high}$ are more reminiscent of an epithelial phenotype. Fascinatingly, the two subpopulations were able to reinstate the equilibrium proportions of the parental cells when FACS purified. These findings clearly show that phenotypic plasticity may be a common feature of different carcinomas.

As we questioned the existence of extrinsic, cooperative mechanisms that control cell-state proportions in an heterogenous cell environment, we further pondered the possibility of a drug-protective, cooperative mechanism that acts along the resistant-to-sensitive cell axis. Taking advantage of a BRAF mutant model of colorectal cancer (RKO), we generated MAPK inhibitors resistant cells (RKOR). Co-culture of RKO and RKOR cells revealed that there is no difference in proliferation rate between the two populations upon drug treatment, in cells

seeded at different densities. Although we detected a minor change in proliferative capacity between co-cultured cells and sensitive cell monoculture in the absence of drugs and with a seeding density of 2,500 cells/well, we did not observe the same result with the other cell density (1,250 cells/well). We hypothesized that this may be a direct result of confluency effects. Experiments using conditioned media failed to demonstrate a pharmacological protective effect via paracrine signaling; this also occurred with pre-treated media, which we expected would have been enriched with paracrine factors directly induced by MAPK inhibitors administration. It has been demonstrated that colorectal cancer cells resistant to cetuximab and panitumumab (EGFR-targeted antibodies) are able to protect the nearby sensitive cells by secreting TGFα and amphiregulin, two EGFR ligands (Hobor *et al*, 2014). Compared to our work, there are several lines of difference: 1) Hobor employed different cell lines: the genetic constitution of a cell may play role in the signaling cascade that leads to the synthesis of paracrine factors in response to drug stimuli; 2) drug targets are different (EGFR-directed antibodies vs MAPK inhibitors); 3) Hobor developed a 3D culture system that allows resistant and sensitive cells to grow together in spheroids. This would take in account short range paracrine and juxtracrine signaling. We could address the last point by performing three-dimensional culturing of RKO and RKOR spheroids and assessed whether forced adhesion might have an impact on drug protection.

5. Materials and Methods

5.1 Cell culture

SUM149 and SUM159 cells were cultured in Ham's F12 medium (Gibco, #21765029) supplemented with 5% fetal calf serum (FCS, HyClone, US origin, #sh30073.03), 1% antibiotics (Pen/Strep, Euroclone, #ECB3001D), insulin (5 µg/mL, Sigma, #I0516), hydrocortisone (1 µg/mL, Sigma, #H0888) and HEPES (10 mM, Gibco, #15630056). BT549 and MDA-MB-436 cells were cultured in RPMI medium (Euroclone, #ECB2000), supplemented with 10% fetal bovine serum (FBS, Gibco, #10270106) and 1% antibiotics (Pen/Strep). A375, MDA-MB-231 and MDA-MB-468 were cultured in DMEM medium (Euroclone, #ECM0728L) supplemented with 10% fetal bovine serum (FBS, Gibco, #10270106) and 1% antibiotics (Pen/Strep). RKO cells were cultured in MEM medium (Gibco, #11095080) supplemented with 10% FBS, sodium pyruvate (1 mM, Sigma, #S8636) and 1% antibiotics (Pen/Strep). All cell lines were grown at 37 °C in a 5% CO_2 incubator.

5.2 Flow cytometry

Subconfluent cell cultures (80-90%) were washed with PBS to remove medium, trypsinized and counted. Cells were stained with antibodies specific for human cell-surface markers: EpCAM–FITC (1:50, mouse monoclonal, clone VU–ID9, AbCAM, #ab8666), CD24–PE (1:25, mouse monoclonal, clone ML5, #555428), CD44–APC (1:25, mouse monoclonal, clone G44-26, BD Biosciences, #559942) and CD49f–BV421 (1:1,500, rat monoclonal, clone GoH3, BD Biosciences, #747725). For each staining reaction, cells were incubated with antibodies diluted in FACS Buffer (PBS 1X, 2.5 mM EDTA, 10 mM HEPES, 2% BSA) for 20 min at room temperature. Unbound antibody was washed off with excess PBS and cells were sorted at 4 °C with BD FACSAria Fusion within one hour post-staining. FACS analyses were performed with CytoFLEX S and CytoFLEX LX. As a control, the following isotypes antibodies were used: FITC-IgG1 κ isotype

control (1:50 mouse monoclonal, clone MOPC-21, BD Biosciences #555909), PE-IgG2α κ isotype control (1:50 mouse monoclonal, clone G155-178, BD Biosciences #555574) and APC-IgG2b κ isotype control (1:50 mouse monoclonal, clone 27-35, BD Biosciences #555745).

5.3 Immunofluorescence

Immunofluorescence techniques were used to study cell-surface protein localization. Briefly, cells were grown on 13 mm coverslip previously coated with poly-L-ornithine (Sigma, #P4957). After 24 hours, cells were washed twice with PBS and fixed in 4% paraformaldehyde (PFA) diluted in H_2O for 15 minutes at room temperature, followed by two washes in PBS, permeabilized and blocked in PBS-T (PBS, Triton X-100 0.1%, 5% Normal Goat Serum and 1% Bovine Serum Albumin) for 1 hour at room temperature. Primary antibodies were diluted in PBS-T and incubated overnight at 4 °C, in agitation. Then, cells were washed 3 times with PBS and incubated with secondary antibodies, rhodamine phalloidin (1:400, excitation wavelength: 557 nm; emission wavelength: 576 nm; Invitrogen, #R415) and DAPI (1:1000, 5 mg/mL, excitation wavelength: 358 nm; emission wavelength: 461 nm; Molecular Probes, #D3571) diluted in PBS for 2 hours at room temperature, in agitation. Finally, Coverslips were washed in PBS three times and in water one time, mounted on microscope slides using ProLong Gold Antifade Reagent (Invitrogen, #P36930). The following primary antibodies were used: rabbit anti-EpCAM (1:1,000, Invitrogen, #PA5-19832), mouse anti-CD24 (1:400, monoclonal, clone SN3, Invitrogen, #MA5-11828), mouse anti-Vim (1:250, monoclonal, clone V9, Santa Cruz Biotechnology, #sc-6260). The following secondary antibodies were used: goat anti-rabbit (1:500 Alexa Fluor secondary antibodies, Molecular Probes), goat anti-mouse (1:500 Alexa Fluor secondary antibodies, Molecular Probes).

5.4 Imaging analysis

SUM149 and SUM159 cells were acquired using confocal microscope with 63X/1.40 NA Oil objective and analyzed with LAS AF software (Leica SP5). Confocal microscope is a particular type of microscope in which the optical resolution and the contrast of an image is enhanced using a point illumination that permits exclusion of out-of-focus light in image formation. In order to enable more sensitive signals discrimination, confocal sections were obtained by acquisition at different levels along the optical axis of z-sections (0.29 µm z-resolution). Comparative imaging analyses were performed in parallel with identical acquisition parameters. Images were acquired with Las X software, processed with Fiji (https://fiji.sc) and Affinity Designer software (https://affinity.serif.com).

5.5 RNAseq data analysis

SUM149 cell line was sorted in triplicate: each cell-state was immediately processed for RNA extraction (see also **5.6 RNA extraction and real-time quantitative PCR**). Library preparation and sequencing were performed by Center for Omics Sciences at San Raffaele Scientific Institute. RNA counts from RNASeq were analyzed in biological triplicates. Log_2FC, *p-value* and *adjusted p-value* were investigated: samples were filtered for *adjusted p-value* < 0.05 and ranked by their Log_2FC value. *Adjusted p-value* is used to determine statistical significance among a pool of samples in a hypothesis test; Log_2FC is defined as the differential change of a quantity that ranges between an initial and a subsequent measure. In this case, we used the base 2 logarithm (Log_2) of the FC (Fold Change), that is, in our case, the ratio between either *basal vs luminal* or *luminal vs basal* counts. Differentially expressed genes were sorted, and their functions were scored: relevant biological pathways were analyzed with the web based EnrichR software (https://maayanlab.cloud/Enrichr) in order to score significant changes between basal and luminal samples. Gene Ontology

Biological process database was used for the analysis. Relevant terms were selected and ranked by $-\log_{10}$ (p-value).

5.6 RNA extraction and real-time quantitative PCR

Total RNA was extracted from unsorted and sorted cells with RNAqueous™-Micro Total RNA Isolation Kit (Life Technologies, #AM1931), according to the manufacturer protocol. RNA was quantified with NanoDrop™ 8000 (ThermoFisher). The same RNA amount for each sample was retrotranscribed with ImProm-II™ Reverse Transcription System (Promega, #A3800) according to the manufacturer protocol. Quantitative Real Time PCR was performed using SYBR® Green Master Mix (Applied Biosystems, #4364346) on the ViiA7 Real Time PCR System (Applied Biosystems). Expression levels of target genes were normalized to an endogenous control (housekeeping gene): the ribosomal protein lateral stalk subunit P0 (rpLP0). The list of primers is provided below:

RPLP0 FWD: TGGTCATCCAGCAGGTGTTCGA
RPLP0 REV: ACAGACACTGGCAACATTGCGG
EPCAM FWD: GCCAGTGTACTTCAGTTGGTGC
EPCAM REV: CCCTTCAGGTTTTGCTCTTCTCC
CD24 FWD: CACGCAGATTTATTCCAGTGAAAC
CD24 REV: GACCACGAAGAGACTGGCTGTT
CDH1 FWD: GCCTCCTGAAAAGAGAGTGGAAG
CDH1 REV: TGGCAGTGTCTCTCCAAATCCG
CDH2 FWD: CCTCCAGAGTTTACTGCCATGAC
CDH2 REV: GTAGGATCTCCGCCACTGATTC
VIM FWD: AGGCAAAGCAGGAGTCCACTGA
VIM REV: ATCTGGCGTTCCAGGGACTCAT
ZEB1 FWD: GGCATACACCTACTCAACTACGG
ZEB1 REV: TGGGCGGTGTAGAATCAGAGTC
SNAI1 FWD: ACTATGCCGCGCTCTTTC
SNAI1 REV: GCTGGAAGGTAAACTCTGGATTA

SNAI2 FWD: CTTCCTGGTCAAGAAGCATTTC
SNAI2 REV: GGGAAATAATCACTGTATGTGTGT
TWIST1 FWD: ACTGGACACACATACAGTGATTAT
TWIST1 REV: GAATGGAGCAGCGGTAGTC
TGFBETA1 FWD: CGTGGAGCTGTACCAGAAATAC
TGFBETA1 REV: CACAACTCCGGTGACATCAA

5.7 Conditioned medium experiments

In SUM149 and SUM159 cell lines, the conditioned medium experiments were performed within two distinct time frames. In the first timeframe, 1 x 10^6 SUM149 and SUM159 cells were seeded in 10 cm dishes in complete medium. Medium was conditioned for 72 hours, generating SUM149 conditioned medium (CM149) and SUM159 conditioned medium (CM159). In a second timeframe, 3 x 10^5 SUM149 and SUM159 cells were seeded in 6-well plates. After 24 hours, cells were washed twice with PBS and covered with 3 mL of media prepared from undiluted CM149, undiluted CM159, half fresh medium and half CM149, half fresh medium and half CM159. Fresh medium was used as control. After 5 days of culturing, cell-state proportions were analyzed with FACS. RKO, RKO GFP and RKOR conditioned medium experiments were performed within two distinct time frames. In the first timeframe, 1 x 10^6 cells of each cell population were seeded in 10 cm dishes in 10 mL of their complete medium. Cells were either pre-treated with DMSO or a combination of 300 nM Naporafenib and 3 nM Trametinib, generating Conditioned Medium from Sensitive (RKO and RKO GFP) cells (CS) and Conditioned Medium from Resistant (RKOR) cells (CR). In a second timeframe, RKO, RKO GFP and RKOR cells were seeded in triplicate in 96-well as 1,250 or 2,500 cells/well density, both alone or a 50:50 mixture of RKO GFP:RKOR or RKO:RKOR. After 24 hours, cells were covered with either fresh medium, half fresh medium and half CS medium or half fresh medium and half CR medium. Then, cells were treated with DMSO or 300 nM Naporafenib and 3 nM Trametinib. After 3 days of culturing, cell viability was assessed by using Incucyte® (Sartorius) (see **5.14 Live imaging and Incucyte®** for further

details) for RKO GFP and RKOR co-culture experiments; by using CellTiter-Glo® Luminescent Cell Viability Assay (Promega, #G7571) for RKO and RKOR co-culture experiments. Luminescence was measured by Mithras plate reader (Berthold Technologies). Results were normalized on day 0.

5.8 Generation of SUM149 reporter cell line

5.8.1 Design of sgRNA

To design the EPCAM sgRNA sequences (three at least), the genomic sequence upstream the stop codon of *EPCAM* was obtained from UCSC genome browser. PAM sequences (NGG) and the 20 nucleotides upstream of this sequence (candidate sgRNA binding sites) were identified. sgRNA sequences were selected by virtue of the nearness to stop codon, the lack of repetitive structures and a high GC content (at least 40%). sgRNAs were synthesized in the form of:

5’ – CACCG**NNNNNNNNNNNNNNNNNN** – 3’

3’ – C**NNNNNNNNNNNNNNNNNN**CAAA – 5’

where the sgRNA sequence is depicted in bold.

5.8.2 Cloning

To clone the sgRNA sequences into the CAS9 vector (pSpCas9(BB)-2A-Puro (PX459) V2.0, Addgene #62988), we digested and dephosphorylated the plasmid with *BbsI* restriction enzyme (ThermoFisher, #D1014) coupled with alkaline phosphatase (ThermoFisher, #F0651) for 30 minutes at 37° C. The digested fragment was gel purified and eluted. Then, each sgRNA oligos were phosphorylated and annealed by using T4 polynucleotide kinase (NEB, #M0201L).

Annealing was performed as depicted below:

37° C, 30 minutes

95° C, 5 minutes, then down to 25° C at 5° C/minute

Ligation of digested plasmid and annealed sgRNA oligos, with a molar ratio of 1:10, respectively, was performed with Quick Ligation Kit™ (NEB, # M2200L). Ligated plasmid was then transformed and amplified into One Shot™ TOP10 *E. coli* (ThermoFisher, #C404010) according to manufacturer protocol. Validation of the ligation was performed with Sanger sequencing using the following primers:

p62988 seq1 FWD: GGACTATCATATGCTTACCG

Maxi-prep to purify the amplified plasmid was performed with the NucleoBond™ Xtra Maxi EF (Macherey-Nagel, #740424.50).

5.8.3 Design of donor sequence

We designed 60 bp oligos composed of 40 bp of homology arm and 20 bp primer sequences to amplify donors containing eGFP and Neomycin resistance cassette. Specifically, we selected the 40 bp region immediately upstream and downstream (this one reverse complemented) of the *EPCAM* stop codon. We amplified the template plasmid by PCR using the FWD and REV primers. Specifically, we used Phusion™ High-Fidelity DNA Polymerase (NEB, #M0530L) with the HF buffer. PCR cycle described below:

Initial denaturation, 98° C, 5 minutes

Denaturation, 98° C, 10 seconds

Annealing, 57° C, 20 seconds

Extension, 72° C, 60 seconds

Final extension, 72° C, 5 minutes

The resulting product was purified via QIAquick PCR Purification Kit (QIAGEN, #28104) and resolved in a 1% agarose gel. See ***3.1.10*** for further details.

5.8.4 Transfection

Approximately 400,000 SUM149 cells were plated in 6-well plates in complete medium. After 24 hours, the day of the transfection, they were 90% confluent. We used Lipofectamine™ 3000 reagent to transfect the donor sequence and the

ligated vector into SUM149 cells, according to manufacturer protocol. Specifically, we transfected 1 µg of plasmid together with 1 µg of donor template (1:1 ratio); after 6 hours we replaced the medium with fresh one. We let cells recover for 72 hours post transfection, then we switched medium to 400 µg/mL G418 (ThermoFisher, #11811023) and selected cells for 7 days. Fluorescence was checked with Zeiss Axio Observer Z1.

5.9 Analysis of scRNAseq dataset

We downloaded data from a publicly available dataset of several breast cancer cell lines from GEO (GSE182694). We took advantage of a pipeline generated in *R* by our bioinformatician to identify several clusters of cells depicted in the UMAP plot generated. Briefly, a PCA was performed on highly variable genes, as a first step of dimensional reduction. UMAP algorithms were then applied to further reduce the dataset dimensionality, and SNN algorithm was applied to cluster the cells into distinct cell subpopulations. *R* (version 4.2.2 – 2022.10.31) and RStudio (version 2022.07.2+576) were employed.

5.10 Cell sensitivity to drug treatment

To study whether A375 and RKO cells were sensitive to BRAFV600E and/or MEK inhibitors, we treated cells with increasing concentration of Naporafenib (Selleckchem, #S8745) and Trametinib (Selleckchem, #S2673), respectively. Naporafenib is a type II ATP-competitive inhibitor that inhibits BRAF kinase; Trametinib is a s a specific MEK1/2 inhibitor. Briefly, we seeded 2,500 cells/well (A375 and RKO) or 1,500 cells/well (RKO only) in white, opaque 96-well plates, in 50 µL of complete medium. After overnight incubation, 50 µL of medium supplemented with DMSO or drugs was added to cells. After 72 hours of incubation, the viability of A375 and RKO cells was assessed using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega, #G7571) that quantifies the amount of ATP within the cell. Luminescence was measured by Mithras plate

reader (Berthold Technologies). Results were normalized to percentage relative to DMSO.

5.11 Generation of resistant cells

$1x10^6$ RKO cells were plated on a 10 cm dish and treated with combination of Naporafenib and Trametinib with a concentration of 300 nM and 3 nM, respectively, until resistance was achieved as per **Figure 3.26 A**.

5.12 Lentiviral infection

To overexpress GFP, pLenti-C-Myc-DDK-IRES-PURO lentiviral vector was used in infection experiments. For lentivirus production, HEK293T cells were transfected with the $CaCl_2$ protocol. A mix containing 10 µg of transfer vector, 6.5 µg of packaging vector ΔR8.74, 3.5 µg of Env VSV-G, 2.5 µg of REV, ddH_2O to 450 µL, 50 µL of 2.5 M $CaCl_2$ and 500 µL of 2X HBS were added dropwise over HEK293T cells seeded on a 15 cm dish. The medium was replaced after 16 hours. 24 hours later the medium containing virus particles was collected and pelleted at 1,200 rpm. Supernatant was filtered on a 0.45 µm filter, and then incubated with PEG-it Virus Precipitation Solution 5X (Systembio, #LV825A-1) overnight at 4° C to precipitate viral particles. After a centrifugation step at 1,500 x*g* at 4° C, lentiviral pellet was resuspended in 1x PBS and stored in aliquots at – 80° C. The day prior to infection, 150,000 RKO cells were plated in a 6-well plate and infected the day after with 10 µL of concentrated virus for 24 hours. Viral medium was replaced the day after and cells were subsequently selected with 3 µg/mL puromycin (Invitrogen, #A1113803) for at least 3 days. Infection outcome was checked via fluorescence microscopy with Zeiss Axio Observer Z1.

5.13 Proliferation studies

To study the proliferation rate of RKO cells and RKO GFP derivative cells, they were plated in triplicate at 2,500 cells/well density in a 96-well plate in

complete medium. After 24 and 72 hours of incubation, the proliferation of each cell population was assessed by quantitation of ATP present, using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega, #G7571). Luminescence was measured by Mithras plate reader (Berthold Technologies). Results were normalized on day 0.

5.14 Live imaging and Incucyte®

RKO GFP, RKOR cells, or a 50:50 mixture of the two populations were plated in triplicate in 96-well as 1,250 or 2,500 cells/well densities. Cells were incubated for 24 hours, then treated with 300 nM Naporafenib and 3 nM trametinib, or with DMSO as control. The plate was then imaged with Incucyte® (Sartorius) for 120 hours, with images taken every 12 hours. Proliferation of RKO GFP was measured through automatic software counting of green cells from images taken with 4X and 10X objectives and normalized on day 0.

5.15 Statistical analysis

Each experiment was repeated three times at least, as biological replicates; means, SD (Standard Deviation) and SEM (Standard Error of the Mean) were calculated. *p-values* obtained by Student t-test, paired, two-tailed: * *p-value* of 0.05, ** *p-value* of 0.01, *** *p-value* of 0.0001.

References

Addis RC & Epstein JA (2013) Induced regeneration--the progress and promise of direct reprogramming for heart repair. *Nat Med* 19: 829–836

Akrap N, Andersson D, Bom E, Gregersson P, Staahlberg A & Landberg G (2016) Identification of Distinct Breast Cancer Stem Cell Populations Based on Single-Cell Analyses of Functionally Enriched Stem and Progenitor Pools. *Stem Cell Reports* 6: 121–136

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, *et al* (2013) Signatures of mutational processes in human cancer. *Nature* 500: 415–421

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ & Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci USA* 100: 3983–3988

Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, Kempski H, Moorman AV, Titley I, Swansbury J, *et al* (2011) Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 469: 356–361

Bakhoum SF & Cantley LC (2018) The Multifaceted Role of Chromosomal Instability in Cancer and Its Microenvironment. *Cell* 174: 1347–1360

Balázsi G, van Oudenaarden A & Collins JJ (2011) Cellular Decision Making and Biological Noise: From Microbes to Mammals. *Cell* 144: 910–925

Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, Dewhirst MW, Bigner DD & Rich JN (2006) Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* 444: 756–760

Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, *et al* (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449: 1003–1007

Barkley D, Moncada R, Pour M, Liberman DA, Dryg I, Werba G, Wang W, Baron M, Rao A, Xia B, *et al* (2022) Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat Genet* 54: 1192–1201

Barnabas N & Cohen D (2013) Phenotypic and Molecular Characterization of MCF10DCIS and SUM Breast Cancer Cell Lines. *International Journal of Breast Cancer* 2013: 1–16

Barrallo-Gimeno A & Nieto MA (2005) The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* 132: 3151–3161

Barzgar Barough N, Sajjadian F, Jalilzadeh N, Shafaei H & Velaei K (2021) Understanding breast cancer heterogeneity through non-genetic heterogeneity. *Breast Cancer* 28: 777–791

Batlle E & Clevers H (2017) Cancer stem cells revisited. *Nat Med* 23: 1124–1134

Bauer K, Parise C & Caggiano V (2010) Use of ER/PR/HER2 subtypes in conjunction with the 2007 St Gallen Consensus Statement for early breast cancer. *BMC Cancer* 10: 228

Bedzhov I & Zernicka-Goetz M (2014) Self-Organizing Properties of Mouse Pluripotent Cells Initiate Morphogenesis upon Implantation. *Cell* 156: 1032–1044

Bender S, Tang Y, Lindroth AM, Hovestadt V, Jones DTW, Kool M, Zapatka M, Northcott PA, Sturm D, Wang W, *et al* (2013) Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer Cell* 24: 660–672

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, *et al* (2006) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125: 315–326

Bhatia, Monkman, Blick, Pinto, Waltham, Nagaraj, & Thompson (2019) Interrogation of Phenotypic Plasticity between Epithelial and Mesenchymal States in Breast Cancer. *JCM* 8: 893

Bornes L, Belthier G & van Rheenen J (2021) Epithelial-to-Mesenchymal Transition in the Light of Plasticity and Hybrid E/M States. *Journal of Clinical Medicine* 10: 2403

Bradner JE, Hnisz D & Young RA (2017) Transcriptional Addiction in Cancer. *Cell* 168: 629–643

Brocks D, Assenov Y, Minner S, Bogatyrova O, Simon R, Koop C, Oakes C, Zucknick M, Lipka DB, Weischenfeldt J, *et al* (2014) Intratumor DNA Methylation Heterogeneity Reflects Clonal Evolution in Aggressive Prostate Cancer. *Cell Reports* 8: 798–806

de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, Jamal-Hanjani M, Shafi S, Murugaesu N, Rowan AJ, *et al* (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 346: 251–256

Buonato JM, Lan IS & Lazzara MJ (2015) EGF augments TGFβ-induced epithelial-mesenchymal transition by promoting SHP2 binding to GAB1. *J Cell Sci* 128: 3898–3909

Burrill DR & Silver PA (2010) Making Cellular Memories. *Cell* 140: 13–18

Burrill DR & Silver PA (2011) Synthetic circuit identifies subpopulations with sustained memory of DNA damage. *Genes Dev* 25: 434–439

Bussard KM & Smith GH (2012) Human Breast Cancer Cells Are Redirected to Mammary Epithelial Cells upon Interaction with the Regenerating Mammary Gland Microenvironment In-Vivo. *PLoS ONE* 7: e49221

Casas E, Kim J, Bendesky A, Ohno-Machado L, Wolfe CJ & Yang J (2011) Snail2 is an essential mediator of Twist1-induced epithelial mesenchymal transition and metastasis. *Cancer Res* 71: 245–254

Celià-Terrassa T, Bastian C, Liu DD, Ell B, Aiello NM, Wei Y, Zamalloa J, Blanco AM, Hang X, Kunisky D, *et al* (2018) Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. *Nat Commun* 9: 5005

Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, Brooks M, Reinhardt F, Su Y, Polyak K, *et al* (2011) Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc Natl Acad Sci U S A* 108: 7950–7955

Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F, D'Alessio AC, Young RA & Weinberg RA (2013) Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity. *Cell* 154: 61–74

Chaffer CL, San Juan BP, Lim E & Weinberg RA (2016) EMT, cell plasticity and metastasis. *Cancer Metastasis Rev* 35: 645–654

Chang HH, Oh PY, Ingber DE & Huang S (2006) Multistable and multistep dynamics in neutrophil differentiation. *BMC Cell Biol* 7: 11

Chao YL, Shepard CR & Wells A (2010) Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition. *Mol Cancer* 9: 179

Cheang MCU, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM & Nielsen TO (2008) Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res* 14: 1368–1376

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR & Ma'ayan A (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14: 128

Chen J, Li Y, Yu T-S, McKay RM, Burns DK, Kernie SG & Parada LF (2012) A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* 488: 522–526

Chen Q, Shi J, Tao Y & Zernicka-Goetz M (2018) Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo. *Nature Communications* 9: 1–11

Chung W, Eum HH, Lee H-O, Lee K-M, Lee H-B, Kim K-T, Ryu HS, Kim S, Lee JE, Park YH, *et al* (2017) Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* 8: 15081

Clevers H (2011) The cancer stem cell: premises, promises and challenges. *Nat Med* 17: 313–319

Clevers H (2013) The intestinal crypt, a prototype stem cell compartment. *Cell* 154: 274–284

Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, Milo R, Cohen-Saidon C, Liron Y, Kam Z, *et al* (2008) Dynamic proteomics of individual cancer cells in response to a drug. *Science* 322: 1511–1516

Comet I, Riising EM, Leblanc B & Helin K (2016) Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer* 16: 803–810

Coulon A, Gandrillon O & Beslon G (2010) On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter. *BMC Syst Biol* 4: 2

Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, *et al* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486: 346–352

Das D, Karthik N & Taneja R (2021) Crosstalk Between Inflammatory Signaling and Methylation in Cancer. *Front Cell Dev Biol* 9: 756458

Dave N, Guaita-Esteruelas S, Gutarra S, Frias À, Beltran M, Peiró S & de Herreros AG (2011) Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition. *J Biol Chem* 286: 12024–12032

Deininger M (2005) Resistance to imatinib: mechanisms and management. *J Natl Compr Canc Netw* 3: 757–768

Dekker J & Misteli T (2015) Long-Range Chromatin Interactions. *Cold Spring Harb Perspect Biol* 7: a019356

Del Prete A, Schioppa T, Tiberio L, Stabile H & Sozzani S (2017) Leukocyte trafficking in tumor microenvironment. *Curr Opin Pharmacol* 35: 40–47

Desai A, Webb B & Gerson SL (2014) CD133+ cells contribute to radioresistance via altered regulation of DNA repair genes in human lung cancer cells. *Radiotherapy and Oncology* 110: 538–545

Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, *et al* (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481: 506–510

Domingues AF, Kulkarni R, Giotopoulos G, Gupta S, Vinnenberg L, Arede L, Foerner E, Khalili M, Adao RR, Johns A, *et al* (2020) Loss of Kat2a enhances transcriptional noise and depletes acute myeloid leukemia stem-like cells. *eLife* 9: e51754

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ & Wicha MS (2003) In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* 17: 1253–1270

Doulatov S, Notta F, Laurenti E & Dick JE (2012) Hematopoiesis: a human perspective. *Cell Stem Cell* 10: 120–136

Ebos JML, Lee CR, Cruz-Munoz W, Bjarnason GA, Christensen JG & Kerbel RS (2009) Accelerated metastasis after short-term treatment with a potent inhibitor of tumor angiogenesis. *Cancer Cell* 15: 232–239

Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, Gelmon K, Chia S, Mar C, Wan A, *et al* (2015) Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 518: 422–426

Elias A, Siegelin MD, Steinmüller A, von Deimling A, Lass U, Korn B & Mueller W (2009) Epigenetic Silencing of Death Receptor 4 Mediates Tumor Necrosis Factor–Related Apoptosis-Inducing Ligand Resistance in Gliomas. *Clinical Cancer Research* 15: 5457–5465

Eliyatkin N, Yalcin E, Zengel B, Aktaş S & Vardar E (2015) Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. *J Breasth Health* 11: 59–66

Esteller M (2008) Epigenetics in Cancer. *N Engl J Med* 358: 1148–1159

Estrella V, Chen T, Lloyd M, Wojtkowiak J, Cornnell HH, Ibrahim-Hashim A, Bailey K, Balagurunathan Y, Rothberg JM, Sloane BF, *et al* (2013) Acidity Generated by the Tumor Microenvironment Drives Local Invasion. *Cancer Research* 73: 1524–1535

Fane M & Weeraratna AT (2020) How the ageing microenvironment influences tumour progression. *Nat Rev Cancer* 20: 89–106

Fernando NT, Koch M, Rothrock C, Gollogly LK, D'Amore PA, Ryeom S & Yoon SS (2008) Tumor Escape from Endogenous, Extracellular Matrix–

Associated Angiogenesis Inhibitors by Up-Regulation of Multiple Proangiogenic Factors. *Clinical Cancer Research* 14: 1529–1539

Ferrell JE (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14: 140–148

Fillmore CM, Gupta PB, Rudnick JA, Caballero S, Keller PJ, Lander ES & Kuperwasser C (2010) Estrogen expands breast cancer stem-like cells through paracrine FGF/Tbx3 signaling. *PNAS*

Fillmore CM & Kuperwasser C (2008) Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Research* 10: 1–13

Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML & Bernstein BE (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529: 110–114

Flavahan WA, Gaskell E & Bernstein BE (2017) Epigenetic plasticity and the hallmarks of cancer. *Science* 357

Gadd VL, Aleksieva N & Forbes SJ (2020) Epithelial Plasticity during Liver Injury and Regeneration. *Cell Stem Cell* 27: 557–573

Gambardella G, Viscido G, Tumaini B, Isacchi A, Bosotti R & di Bernardo D (2022) A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nat Commun* 13: 1714

Gandhi S, Ezin M & Bronner ME (2020) Reprogramming Axial Level Identity to Rescue Neural-Crest-Related Congenital Heart Defects. *Developmental Cell* 53: 300-315.e4

Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, *et al* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 366: 883–892

Gilbertson RJ & Graham TA (2012) Resolving the stem-cell debate. *Nature* 488: 462–463

Gilmartin AG, Bleam MR, Groy A, Moss KG, Minthorn EA, Kulkarni SG, Rominger CM, Erskine S, Fisher KE, Yang J, *et al* (2011) GSK1120212 (JTP-74057) Is an Inhibitor of MEK Activity and Activation with Favorable Pharmacokinetic Properties for Sustained In Vivo Pathway Inhibition. *Clinical Cancer Research* 17: 989–1000

Gjorevski N, Boghaert E & Nelson CM (2012) Regulation of Epithelial-Mesenchymal Transition by Transmission of Mechanical Stress through Epithelial Tissues. *Cancer Microenvironment* 5: 29–38

Golding I, Paulsson J, Zawilski SM & Cox EC (2005) Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* 123: 1025–1036

Graf T (2002) Differentiation plasticity of hematopoietic cells. *Blood* 99: 3089–3101

Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, Parks B, Gars E, Liedtke M, Zheng GXY, *et al* (2019) Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* 37: 1458–1465

Greenburg G & Hay ED (1982) Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *J Cell Biol* 95: 333–339

van der Gun BTF, Melchers LJ, Ruiters MHJ, de Leij LFMH, McLaughlin PMJ & Rots MG (2010) EpCAM in carcinogenesis: the good, the bad or the ugly. *Carcinogenesis* 31: 1913–1921

Gunasinghe NPAD, Wells A, Thompson EW & Hugo HJ (2012) Mesenchymal-epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. *Cancer Metastasis Rev* 31: 469–478

Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C & Lander ES (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146: 633–644

Gupta PB, Pastushenko I, Skibinski A, Blanpain C & Kuperwasser C (2019) Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* 24: 65–78

Gutierrez G, Sun P, Han Y & Dai X (2022) Defining mammary basal cell transcriptional states using single-cell RNA-sequencing. *Sci Rep* 12: 4893

Hall AE, Pohl SÖ-G, Cammareri P, Aitken S, Younger NT, Raponi M, Billard CV, Carrancio AB, Bastem A, Freile P, *et al* (2022) RNA splicing is a key mediator of tumour cell plasticity and a therapeutic vulnerability in colorectal cancer. *Nat Commun* 13: 2791

Han L, Chaturvedi P, Kishimoto K, Koike H, Nasr T, Iwasawa K, Giesbrecht K, Witcher PC, Eicher A, Haines L, *et al* (2020) Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis. *Nat Commun* 11: 4158

Han R, Huang G, Wang Y, Xu Y, Hu Y, Jiang W, Wang T, Xiao T & Zheng D (2016) Increased gene expression noise in human cancers is correlated with low p53 and immune activities as well as late stage cancer. *Oncotarget* 7: 72011–72020

Hanahan D (2022) Hallmarks of Cancer: New Dimensions. *Cancer Discovery* 12: 31–46

Hanahan D & Weinberg RA (2000) The Hallmarks of Cancer. *Cell* 100: 57–70

Hanahan D & Weinberg RA (2011) Hallmarks of Cancer: The Next Generation. *Cell* 144: 646–674

Hausser J & Alon U (2020) Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nature Reviews Cancer* 20: 247–257

Hayford CE, Tyson DR, Iii CJR, Frick PL, Quaranta V & Harris LA (2021) An in vitro model of tumor heterogeneity resolves genetic, epigenetic, and stochastic sources of cell state variability. *PLOS Biology* 19: e3000797

Hecht SS (1999) Tobacco Smoke Carcinogens and Lung Cancer. *JNCI: Journal of the National Cancer Institute* 91: 1194–1210

Herlyn M, Steplewski Z, Herlyn D & Koprowski H (1979) Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proc Natl Acad Sci USA* 76: 1438–1442

Hinohara K, Wu H-J, Vigneau S, McDonald TO, Igarashi KJ, Yamamoto KN, Madsen T, Fassl A, Egri SB, Papanastasiou M, *et al* (2018) KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer Cell* 34: 939-953.e9

Hobor S, Van Emburgh BO, Crowley E, Misale S, Di Nicolantonio F & Bardelli A (2014) TGFα and amphiregulin paracrine network promotes resistance to EGFR blockade in colorectal cancer cells. *Clinical Cancer Research* 20: 6429–6438

Hugo HJ, Wafai R, Blick T, Thompson EW & Newgreen DF (2009) Staurosporine augments EGF-mediated EMT in PMC42-LA cells through actin depolymerisation, focal contact size reduction and Snail1 induction - a model for cross-modulation. *BMC Cancer* 9: 235

Imrich S, Hachmeister M & Gires O (2012) EpCAM and its potential role in tumor-initiating cells. *Cell Adhesion & Migration* 6: 30–38

Ito M, Liu Y, Yang Z, Nguyen J, Liang F, Morris RJ & Cotsarelis G (2005) Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat Med* 11: 1351–1354

Jenkinson G, Pujadas E, Goutsias J & Feinberg AP (2017) Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat Genet* 49: 719–729

Jewer M, Lee L, Leibovitch M, Zhang G, Liu J, Findlay SD, Vincent KM, Tandoc K, Dieters-Castator D, Quail DF, *et al* (2020) Translational control of breast cancer plasticity. *Nat Commun* 11: 2498

Jiang Y & Zhan H (2020) Communication between EMT and PD-L1 signaling: New insights into tumor immune evasion. *Cancer Lett* 468: 72–81

Jiang Y-G, Luo Y, He D, Li X, Zhang L, Peng T, Li M-C & Lin Y-H (2007) Role of Wnt/beta-catenin signaling pathway in epithelial-mesenchymal transition of human prostate cancer induced by hypoxia-inducible factor-1alpha. *Int J Urol* 14: 1034–1039

Johnstone SE & Baylin SB (2010) Stress and the epigenetic landscape: a link to the pathobiology of human diseases? *Nat Rev Genet* 11: 806–812

Jones S, Chen W, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, Traulsen A, Nowak MA, Siegel C, Velculescu VE, *et al* (2008) Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci USA* 105: 4283–4288

Jordan CT (2009) Cancer Stem Cells: Controversial or Just Misunderstood? *Cell Stem Cell* 4: 203–205

Kærn M, Elston TC, Blake WJ & Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6: 451–464

Kang Y & Massagué J (2004) Epithelial-mesenchymal transitions: twist in development and metastasis. *Cell* 118: 277–279

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K & Henikoff S (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10: 1930

Keller PJ, Lin AF, Arendt LM, Klebba I, Jones AD, Rudnick JA, DiMeo TA, Gilmore H, Jefferson DM, Graham RA, *et al* (2010) Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Research* 12: R87

Kennecke H, Yerushalmi R, Woods R, Cheang MCU, Voduc D, Speers CH, Nielsen TO & Gelmon K (2010) Metastatic behavior of breast cancer subtypes. *J Clin Oncol* 28: 3271–3277

Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, Crosetto N, Foukakis T & Navin NE (2018) Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* 173: 879-893.e13

Kim E, Rebecca V, Fedorenko IV, Messina JL, Mathew R, Maria-Engler SS, Basanta D, Smalley KSM & Anderson ARA (2013) Senescent fibroblasts in melanoma initiation and progression: an integrated theoretical, experimental, and clinical approach. *Cancer Res* 73: 6874–6885

Kim KH & Roberts CWM (2016) Targeting EZH2 in cancer. *Nat Med* 22: 128–134

Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, Warren A, Rodman C, Roth JA, Bender SA, *et al* (2020) Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nature Genetics* 52: 1208–1218

Knoechel B, Roderick JE, Williamson KE, Zhu J, Lohr JG, Cotton MJ, Gillespie SM, Fernandez D, Ku M, Wang H, *et al* (2014) An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia. *Nat Genet* 46: 364–370

Knösel T, Schlüns K, Stein U, Schwabe H, Schlag PM, Dietel M & Petersen I (2004) Chromosomal Alterations during Lymphatic and Liver Metastasis Formation of Colorectal Cancer. *Neoplasia* 6: 23–28

Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER, *et al* (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490: 61–70

Korkaya H, Liu S & Wicha MS (2011) Regulation of Cancer Stem Cells by Cytokine Networks: Attacking Cancer's Inflammatory Roots. *Clinical Cancer Research* 17: 6125–6129

Kristiansen G, Sammar M & Altevogt P (2004) Tumour Biological Aspects of CD24, A Mucin-Like Adhesion Molecule. *Histochem J* 35: 255–262

Kröger C, Afeyan A, Mraz J, Eaton EN, Reinhardt F, Khodor YL, Thiru P, Bierie B, Ye X, Burge CB, *et al* (2019) Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proceedings of the National Academy of Sciences* 116: 7353–7362

Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, Etzioni R, Bolouri H, Montgomery B, White T, *et al* (2016) Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat Med* 22: 369–378

Kumar VE, Nambiar R, De Souza C, Nguyen A, Chien J & Lam KS (2022) Targeting Epigenetic Modifiers of Tumor Plasticity and Cancer Stem Cell Behavior. *Cells* 11: 1403

Kupr B & Handschin C (2015) Complex Coordination of Cell Plasticity by a PGC-1α-controlled Transcriptional Network in Skeletal Muscle. *Front Physiol* 6

Kurimoto R, Iwasawa S, Ebata T, Ishiwata T, Sekine I, Tada Y, Tatsumi K, Koide S, Iwama A & Takiguchi Y (2016) Drug resistance originating from a TGF-β/FGF-2-driven epithelial-to-mesenchymal transition and its reversion in human lung adenocarcinoma cell lines harboring an EGFR mutation. *Int J Oncol* 48: 1825–1836

Kwon MJ, Han J, Seo JH, Song K, Jeong HM, Choi J-S, Kim YJ, Lee S-H, Choi Y-L & Shin YK (2015) CD24 Overexpression Is Associated with Poor Prognosis in Luminal A and Triple-Negative Breast Cancer. *PLOS ONE* 10: e0139112

Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S, *et al* (2014) Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 26: 813–825

Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, Minden M, Paterson B, Caligiuri MA & Dick JE (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 367: 645–648

Lawrenson K, Grun B, Benjamin E, Jacobs IJ, Dafou D & Gayther SA (2010) Senescent fibroblasts promote neoplastic transformation of partially transformed ovarian epithelial cells in a three-dimensional model of early stage ovarian cancer. *Neoplasia* 12: 317–325

Lee JM, Dedhar S, Kalluri R & Thompson EW (2006) The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *J Cell Biol* 172: 973–981

Levine E & Hwa T (2007) Stochastic fluctuations in metabolic pathways. *Proc Natl Acad Sci USA* 104: 9224–9229

Lewis PW, Müller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, Garcia BA, Muir TW, Becher OJ & Allis CD (2013) Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* 340: 857–861

Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, Kong SL, Chua C, Hon LK, Tan WS, *et al* (2017) Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 49: 708–718

Li J, Choi PS, Chaffer CL, Labella K, Hwang JH, Giacomelli AO, Kim JW, Ilic N, Doench JG, Ly SH, *et al* (2018) An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. *eLife* 7: e37184

Li R, Ong SL, Tran LM, Jing Z, Liu B, Park SJ, Huang ZL, Walser TC, Heinrich EL, Lee G, *et al* (2020a) Chronic IL-1β-induced inflammation regulates epithelial-to-mesenchymal transition memory phenotypes via epigenetic modifications in non-small cell lung cancer. *Sci Rep* 10: 377

Li W, Li L & Hui L (2020b) Cell Plasticity in Liver Regeneration. *Trends Cell Biol* 30: 329–338

Liao T-T & Yang M-H (2020) Hybrid Epithelial/Mesenchymal State in Cancer Metastasis: Clinical Significance and Regulatory Mechanisms. *Cells* 9: 623

Liau BB, Sievers C, Donohue LK, Gillespie SM, Flavahan WA, Miller TE, Venteicher AS, Hebert CH, Carey CD, Rodig SJ, *et al* (2017) Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. *Cell Stem Cell* 20: 233-246.e7

Lin Y-T & Wu K-J (2020) Epigenetic regulation of epithelial-mesenchymal transition: focusing on hypoxia and TGF-β signaling. *J Biomed Sci* 27: 39

Losick R & Desplan C (2008) Stochasticity and Cell Fate. *Science* 320: 65–68

Lund K, Bostad M, Skarpen E, Braunagel M, Kiprijanov S, Krauss S, Duncan A, Høgset A & Selbo PK (2014) The novel EpCAM-targeting monoclonal antibody 3-17I linked to saporin is highly cytotoxic after photochemical internalization in breast, pancreas and colon cancer cell lines. *mAbs* 6: 1038–1050

Ma L, Hernandez MO, Zhao Y, Mehta M, Tran B, Kelly M, Rae Z, Hernandez JM, Davis JL, Martin SP, *et al* (2019) Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer Cell* 36: 418-430.e6

Magee JA, Piskounova E & Morrison SJ (2012) Cancer Stem Cells: Impact, Heterogeneity, and Uncertainty. *Cancer Cell* 21: 283–296

Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, *et al* (2008) The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* 133: 704–715

Marsolier J, Prompsy P, Durand A, Lyne A-M, Landragin C, Trouchet A, Bento ST, Eisele A, Foulon S, Baudre L, *et al* (2022) H3K27me3 conditions chemotolerance in triple-negative breast cancer. *Nat Genet* 54: 459–468

Martincorena I & Campbell PJ (2015) Somatic mutation in cancer and normal cells. *Science* 349: 1483–1489

Martinez P, Birkbak NJ, Gerlinger M, McGranahan N, Burrell RA, Rowan AJ, Joshi T, Fisher R, Larkin J, Szallasi Z, *et al* (2013) Parallel evolution of tumour subclones mimics diversity between tumours. *J Pathol* 230: 356–364

Marusyk A, Almendro V & Polyak K (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 12: 323–334

Marusyk A, Janiszewska M & Polyak K (2020) Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* 37: 471–484

Mathews LA, Cabarcas SM, Hurt EM, Zhang X, Jaffee EM & Farrar WL (2011) Increased Expression of DNA Repair Genes in Invasive Human Pancreatic Cancer Cells. *Pancreas* 40: 730–739

Mazor T, Pankov A, Song JS & Costello JF (2016) Intratumoral Heterogeneity of the Epigenome. *Cancer Cell* 29: 440–451

McGranahan N & Swanton C (2015) Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 27: 15–26

McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, Ha G, Biele J, Yap D, Wan A, *et al* (2016) Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet* 48: 758–767

Merlo LMF, Pepper JW, Reid BJ & Maley CC (2006) Cancer as an evolutionary and ecological process. *Nat Rev Cancer* 6: 924–935

Miao J-W, Liu L-J & Huang J (2014) Interleukin-6-induced epithelial-mesenchymal transition through signal transducer and activator of transcription 3 in human cervical carcinoma. *Int J Oncol* 45: 165–176

Michalopoulos GK, Barua L & Bowen WC (2005) Transdifferentiation of rat hepatocytes into biliary cells after bile duct ligation and toxic biliary injury. *Hepatology* 41: 535–544

Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA & Downing JR (2008) Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* 322: 1377–1380

Murugaesu N, Wilson GA, Birkbak NJ, Watkins T, McGranahan N, Kumar S, Abbassi-Ghadi N, Salm M, Mitter R, Horswell S, *et al* (2015) Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov* 5: 821–831

Naba A, Clauser KR, Hoersch S, Liu H, Carr SA & Hynes RO (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* 11: M111.014647

Nakaya Y, Sukowati EW, Wu Y & Sheng G (2008) RhoA and microtubule dynamics control cell-basement membrane interaction in EMT during gastrulation. *Nat Cell Biol* 10: 765–775

Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, *et al* (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472: 90–94

Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, Richman AR, Silverbush D, Shaw ML, Hebert CM, *et al* (2019) An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* 178: 835-849.e21

Negrini S, Gorgoulis VG & Halazonetis TD (2010) Genomic instability an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology* 11: 220–228

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe J-P, Tong F, *et al* (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 10: 515–527

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, *et al* (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* 149: 979–993

Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, Denroche RE, Liang S-B, Brown AMK, Kim JC, *et al* (2016) A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 538: 378–382

Ocaña OH, Córcoles R, Fabra Á, Moreno-Bueno G, Acloque H, Vega S, Barrallo-Gimeno A, Cano A & Nieto MA (2012) Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1. *Cancer Cell* 22: 709–724

Oltean S, Sorg BS, Albrecht T, Bonano VI, Brazas RM, Dewhirst MW & Garcia-Blanco MA (2006) Alternative inclusion of fibroblast growth factor receptor 2 exon IIIc in Dunning prostate tumors reveals unexpected epithelial mesenchymal plasticity. *Proc Natl Acad Sci U S A* 103: 14116–14121

Pàez-Ribes M, Allen E, Hudock J, Takeda T, Okuyama H, Viñals F, Inoue M, Bergers G, Hanahan D & Casanovas O (2009) Antiangiogenic therapy elicits malignant progression of tumors to increased local invasion and distant metastasis. *Cancer Cell* 15: 220–231

Pan M, Kohlbauer V, Blancke Soares A, Schinke H, Huang Y, Kranz G, Quadt T, Hachmeister M & Gires O (2021) Interactome analysis reveals endocytosis and membrane recycling of EpCAM during differentiation of embryonic stem cells and carcinoma cells. *iScience* 24: 103179

Parab L, Pal S & Dhar R (2022) Transcription factor binding process is the primary driver of noise in gene expression. *PLoS Genet* 18: e1010535

Pastore A, Gaiti F, Lu SX, Brand RM, Kulm S, Chaligne R, Gu H, Huang KY, Stamenova EK, Béguelin W, *et al* (2019) Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL. *Nat Commun* 10: 1874

Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, Van Keymeulen A, Brown D, Moers V, Lemaire S, *et al* (2018) Identification of the tumour transition states occurring during EMT. *Nature* 556

Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, *et al* (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (New York, NY)* 344: 1396–1401

Peinado H, Marin F, Cubillo E, Stark H-J, Fusenig N, Nieto MA & Cano A (2004) Snail and E47 repressors of E-cadherin induce distinct invasive and angiogenic properties in vivo. *J Cell Sci* 117: 2827–2839

Peinado H, Olmeda D & Cano A (2007) Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* 7: 415–428

Perino M & Veenstra GJC (2016) Chromatin Control of Developmental Dynamics and Plasticity. *Dev Cell* 38: 610–620

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, *et al* (2000) Molecular portraits of human breast tumours. *Nature* 406: 747–752

Piunti A & Shilatifard A (2016) Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* 352: aad9780

Piunti A & Shilatifard A (2021) The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat Rev Mol Cell Biol*

Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, Delattre O, Sigal-Zafrani B, Bollet M, Longy M, *et al* (2012) Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-like Breast Carcinomas with *BRCA1/2* Inactivation. *Cancer Research* 72: 5454–5462

Potente M, Gerhardt H & Carmeliet P (2011) Basic and Therapeutic Aspects of Angiogenesis. *Cell* 146: 873–887

Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M, *et al* (2013) Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat* 142: 237–255

Prat A & Perou CM (2011) Deconstructing the molecular portraits of breast cancer. *Molecular Oncology* 5: 5–23

Raghavan S, Winter PS, Navia AW, Williams HL, DenAdel A, Lowder KE, Galvez-Reyes J, Kalekar RL, Mulugeta N, Kapner KS, *et al* (2021) Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* 184: 6119-6137.e26

Raj A, Peskin CS, Tranchina D, Vargas DY & Tyagi S (2006) Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol* 4: e309

Rajagopal J & Stanger BZ (2016) Plasticity in the Adult: How Should the Waddington Diagram Be Applied to Regenerating Tissues? *Dev Cell* 36: 133–137

Ramurthy S, Taft BR, Aversa RJ, Barsanti PA, Burger MT, Lou Y, Nishiguchi GA, Rico A, Setti L, Smith A, *et al* (2020) Design and Discovery of N-(3-(2-(2-Hydroxyethoxy)-6-morpholinopyridin-4-yl)-4-methylphenyl)-2-(trifluoromethyl)isonicotinamide, a Selective, Efficacious, and Well-Tolerated RAF Inhibitor Targeting RAS Mutant Cancers: The Path to the Clinic. *J Med Chem* 63: 2013–2027

Rauner G, Jin DX, Miller DH, Gierahn TM, Li CM, Sokol ES, Feng Y-X, Mathis RA, Love JC, Gupta PB, *et al* (2021) Breast tissue regeneration is driven by cell-matrix interactions coordinating multi-lineage stem cell differentiation through DDR1. *Nat Commun* 12: 7116

Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, *et al* (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 45: 970–976

Roche-Lestienne C, Laï J-L, Darré S, Facon T & Preudhomme C (2003) A mutation conferring resistance to imatinib at the time of diagnosis of chronic myelogenous leukemia. *N Engl J Med* 348: 2265–2266

Roesch A, Fukunaga-Kalabis M, Schmidt EC, Zabierowski SE, Brafford PA, Vultur A, Basu D, Gimotty P, Vogt T & Herlyn M (2010) A Temporarily Distinct Subpopulation of Slow-Cycling Melanoma Cells Is Required for Continuous Tumor Growth. *Cell* 141: 583–594

Rossi F, Noren H, Jove R, Beljanski V & Grinnemo K-H (2020) Differences and similarities between cancer and somatic stem cells: therapeutic implications. *Stem Cell Research & Therapy* 11: 489

Ruiz C, Lenkiewicz E, Evers L, Holley T, Robeson A, Kiefer J, Demeure MJ, Hollingsworth MA, Shen M, Prunkard D, *et al* (2011) Advancing a clinically relevant perspective of the clonal nature of cancer. *Proc Natl Acad Sci U S A* 108: 12054–12059

Sacchetti A, Teeuwssen M, Verhagen M, Joosten R, Xu T, Stabile R, van der Steen B, Watson MM, Gusinac A, Kim WK, *et al* (2021) Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. *eLife* 10: e61461

Sánchez Á & Kondev J (2008) Transcriptional control of noise in gene expression. *Proc Natl Acad Sci USA* 105: 5081–5086

Scheel C, Eaton EN, Li SH-J, Chaffer CL, Reinhardt F, Kah K-J, Bell G, Guo W, Rubin J, Richardson AL, *et al* (2011) Paracrine and Autocrine Signals Induce and Maintain Mesenchymal and Stem Cell States in the Breast. *Cell* 145: 926–940

Schön MP, Schön M, Klein CE, Blume U, Bisson S & Orfanos CE (1994) Carcinoma-Associated 38-kD Membrane Glycoprotein MH 99/KS 1/4 Is Related

to Proliferation and Age of Transformed Epithelial Cell Lines. *Journal of Investigative Dermatology* 102: 987–991

Serrano-Gomez SJ, Maziveyi M & Alahari SK (2016) Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. *Mol Cancer* 15: 18

Shackleton M, Quintana E, Fearon ER & Morrison SJ (2009) Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell* 138: 822–829

Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, Beqiri M, Sproesser K, Brafford PA, Xiao M, *et al* (2017) Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546: 431–435

Shain AH & Bastian BC (2016) From melanocytes to melanomas. *Nat Rev Cancer* 16: 345–358

Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, *et al* (2010) A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell* 141: 69–80

Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, Nikolskaya T, Serebryiskaya T, Beroukhim R, Hu M, *et al* (2007) Molecular Definition of Breast Tumor Heterogeneity. *Cancer Cell* 11: 259–273

Shojaei F & Ferrara N (2008) Role of the microenvironment in tumor growth and in refractoriness/resistance to anti-angiogenic therapies. *Drug Resistance Updates* 11: 219–230

Sisti JS, Collins LC, Beck AH, Tamimi RM, Rosner BA & Eliassen AH (2016) Reproductive risk factors in relation to molecular subtypes of breast cancer: Results from the nurses' health studies. *Int J Cancer* 138: 2346–2356

Slanchev K, Carney TJ, Stemmler MP, Koschorz B, Amsterdam A, Schwarz H & Hammerschmidt M (2009) The epithelial cell adhesion molecule EpCAM is required for epithelial morphogenesis and integrity during zebrafish epiboly and skin development. *PLoS Genet* 5: e1000563

Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JGM, Foekens JA & Martens JWM (2008) Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 68: 3108–3114

Smit MA, Geiger TR, Song J-Y, Gitelman I & Peeper DS (2009) A Twist-Snail axis critical for TrkB-induced epithelial-mesenchymal transition-like transformation, anoikis resistance, and metastasis. *Mol Cell Biol* 29: 3722–3737

Soldatov R, Kaucka M, Kastriti ME, Petersen J, Chontorotzea T, Englmaier L, Akkuratova N, Yang Y, Häring M, Dyachuk V, *et al* (2019) Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* 364: eaas9536

Solini GE, Dong C & Saha M (2017) Embryonic transplantation experiments: Past, present, and future. *Trends Dev Biol* 10: 13–30

Sonzogni O, Haynes J, Seifried LA, Kamel YM, Huang K, BeGora MD, Yeung FA, Robert-Tissot C, Heng YJ, Yuan X, *et al* (2018) Reporters to mark and eliminate basal or luminal epithelial cells in culture and in vivo. *PLOS Biology* 16: e2004049

Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, *et al* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98: 10869–10874

Spemann H & Mangold H (1924) über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. *Archiv f mikr Anat u Entwicklungsmechanik* 100: 599–638

Stathopoulos A & Iber D (2013) Studies of morphogens: keep calm and carry on. *Development* 140: 4119–4124

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, *et al* (2011) Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* 144: 27–40

Stewart-Ornstein J & Lahav G (2016) Dynamics of CDKN1A in Single Cells Defined by an Endogenous Fluorescent Tagging Toolkit. *Cell Reports* 14: 1800–1811

Suzuki A, Maeda T, Baba Y, Shimamura K & Kato Y (2014) Acidic extracellular pH promotes epithelial mesenchymal transition in Lewis lung carcinoma model. *Cancer Cell Int* 14: 129

Takahashi K & Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126: 663–676

Thews O & Riemann A (2019) Tumor pH and metastasis: a malignant process beyond hypoxia. *Cancer Metastasis Rev* 38: 113–129

Thiery JP, Acloque H, Huang RYJ & Nieto MA (2009) Epithelial-mesenchymal transitions in development and disease. *Cell* 139: 871–890

Thuault S, Tan E-J, Peinado H, Cano A, Heldin C-H & Moustakas A (2008) HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition. *J Biol Chem* 283: 33437–33446

Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, *et al* (2016) Dissecting the multicellular

ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352: 189–196

Tomasetti C, Vogelstein B & Parmigiani G (2013) Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A* 110: 1999–2004

TRACERx consortium, Joshi K, de Massy MR, Ismail M, Reading JL, Uddin I, Woolston A, Hatipoglu E, Oakes T, Rosenthal R, *et al* (2019) Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat Med* 25: 1549–1559

Tsai JH, Donaher JL, Murphy DA, Chau S & Yang J (2012) Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* 22: 725–736

Tsaktanis T, Kremling H, Pavšič M, von Stackelberg R, Mack B, Fukumori A, Steiner H, Vielmuth F, Spindler V, Huang Z, *et al* (2015) Cleavage and cell adhesion properties of human epithelial cell adhesion molecule (HEPCAM). *J Biol Chem* 290: 24574–24591

Tunnacliffe E & Chubb JR (2020) What Is a Transcriptional Burst? *Trends in Genetics* 36: 288–297

Vasan N, Baselga J & Hyman DM (2019) A view on drug resistance in cancer. *Nature* 575: 299–309

Vibert L, Daulny A & Jarriault S (2018) Wound healing, cellular regeneration and plasticity: the elegans way. *Int J Dev Biol* 62: 491–505

Vogelstein B & Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* 10: 789–799

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA & Kinzler KW (2013) Cancer Genome Landscapes. *Science* 339: 1546–1558

Voigt P, Tee W-W & Reinberg D (2013) A double take on bivalent promoters. *Genes Dev* 27: 1318–1338

Vrba L, Jensen TJ, Garbe JC, Heimark RL, Cress AE, Dickinson S, Stampfer MR & Futscher BW (2010) Role for DNA Methylation in the Regulation of miR-200c and miR-141 Expression in Normal and Cancer Cells. *PLoS ONE* 5: e8697

Waddington CH (1957) The Strategy of the Genes 0 ed. Routledge

Wang J, He J, Zhu M, Han Y, Yang R, Liu H, Xu X & Chen X (2022a) Cellular Heterogeneity and Plasticity of Skin Epithelial Cells in Wound Healing and Tumorigenesis. *Stem Cell Rev Rep* 18: 1912–1925

Wang K, Liu H, Hu Q, Wang L, Liu J, Zheng Z, Zhang W, Ren J, Zhu F & Liu G-H (2022b) Epigenetic regulation of aging: implications for interventions of aging and diseases. *Sig Transduct Target Ther* 7: 374

Wang L, Leite de Oliveira R, Huijberts S, Bosdriesz E, Pencheva N, Brunen D, Bosma A, Song J-Y, Zevenhoven J, Los-de Vries GT, *et al* (2018a) An Acquired Vulnerability of Drug-Resistant Melanoma with Therapeutic Potential. *Cell* 173: 1413-1425.e14

Wang L, Walker BL, Iannaccone S, Bhatt D, Kennedy PJ & Tse WT (2009) Bistable switches control memory and plasticity in cellular differentiation. *Proceedings of the National Academy of Sciences* 106: 6638–6643

Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, *et al* (2018b) Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science (New York, NY)* 361

Wang Y, Ni T, Wang W & Liu F (2019) Gene transcription in bursting: a unified mode for realizing accuracy and stochasticity: Gene transcription in bursting. *Biol Rev* 94: 248–258

Wang Z, Wang Q, Wang Q, Wang Y & Chen J (2017) Prognostic Significance of CD24 and CD44 in Breast Cancer: A Meta-Analysis. *Int J Biol Markers* 32: 75–82

Winter MJ, Nagtegaal ID, van Krieken JHJM & Litvinov SV (2003) The Epithelial Cell Adhesion Molecule (Ep-CAM) as a Morphoregulatory Molecule Is a Tool in Surgical Pathology. *The American Journal of Pathology* 163: 2139–2148

Yoo YA, Kang MH, Lee HJ, Kim B, Park JK, Kim HK, Kim JS & Oh SC (2011) Sonic hedgehog pathway promotes metastasis and lymphangiogenesis via activation of Akt, EMT, and MMP-9 pathway in gastric cancer. *Cancer Res* 71: 7061–7070

Yuan S, Natesan R, Sanchez-Rivera FJ, Li J, Bhanu NV, Yamazoe T, Lin JH, Merrell AJ, Sela Y, Thomas SK, *et al* (2020) Global Regulation of the Histone Mark H3K36me2 Underlies Epithelial Plasticity and Metastatic Progression. *Cancer Discovery* 10: 854–871

Yuan S, Norgard RJ & Stanger BZ (2019) Cellular plasticity in cancer. *Cancer Discovery* 9: 837–851

Yuan X, Wu H, Han N, Xu H, Chu Q, Yu S, Chen Y & Wu K (2014) Notch signaling and EMT in non-small cell lung cancer: biological significance and therapeutic application. *J Hematol Oncol* 7: 87

Zaret KS & Mango SE (2016) Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev* 37: 76–81

Zhang AW, McPherson A, Milne K, Kroeger DR, Hamilton PT, Miranda A, Funnell T, Little N, de Souza CPE, Laan S, *et al* (2018) Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell* 173: 1755-1769.e22

Zhang Y, Donaher JL, Das S, Li X, Reinhardt F, Krall JA, Lambert AW, Thiru P, Keys HR, Khan M, *et al* (2022) Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS as regulators of distinct EMT trajectories that contribute differentially to metastasis. *Nat Cell Biol* 24: 554–564

Zuazo-Gaztelu I & Casanovas O (2018) Unraveling the Role of Angiogenesis in Cancer Ecosystems. *Front Oncol* 8: 248

