

RESEARCH

Open Access



EMT-driven plasticity prospectively increases cell–cell variability to promote therapeutic adaptation in breast cancer

Lauriane Muller¹, Frédérique Fauvet², Christelle Chassot², Francesca Angileri^{1,3,4}, Angèle Coutant¹, Cyril Dégletagne⁵, Laurie Tonon⁶, Pierre Saintigny^{1,7,10}, Alain Puisieux^{8,9}, Anne-Pierre Morel², Maria Ouzounova^{2,8} and Pierre Martinez^{1,10*}

Abstract

Cellular plasticity enables cancer cells to adapt non-genetically, thereby preventing therapeutic success. The epithelial-mesenchymal transition (EMT) is a type of plasticity linked to resistance and metastasis. However, its exact impact on population diversity and its dynamics under chemotherapy is unknown. We used single-cell transcriptomics to investigate phenotypic diversity dynamics upon treatment in two in vitro models of triple negative breast cancer (TNBC), where EMT-driven plasticity is either induced or spontaneously occurring. We report that EMT-driven plasticity confers higher phenotypic cell–cell variability ($p < 0.001$) while enriching for stem-like cells. Genetic and phenotypic cell–cell variability were not consistently correlated. High-plasticity populations displayed more pre-adapted cells before treatment ($p = 0.03$). In a population displaying spontaneous EMT and phenotypic variation, pre-adapted cells were a rare minority of high-scoring outliers whose expression patterns correlated with survival in TNBC patients subjected to chemotherapy ($p = 0.03$). Higher plasticity was not associated with a partial EMT status. Our results provide novel insights on how EMT-driven plasticity promotes a prospective diversification process increasing population phenotypic diversity, which can yield rare pre-adapted states before treatment. This highlights the need to tackle phenotypic diversity prior to treatment in high-plasticity tumours.

*Correspondence:

Pierre Martinez
pierre.martinez@lyon.unicancer.fr

¹ Integrated Analyses of Cancer Dynamics Team, Centre de Recherche en Cancérologie de Lyon (CRCL), Institut Convergence PlasCan, INSERM U1052, CNRSUMR 5286, Centre Léon Bérard, Université Claude Bernard Lyon 1, Lyon, France

² EMT and Cancer Cell Plasticity Team, Centre Léon Bérard, Lyon, France

³ Département de Pharmacologie, Physiologie et Toxicologie, Institut des Sciences Pharmaceutiques Et Biologiques (ISPB), Université Claude Bernard Lyon I, Lyon, France

⁴ Unité de recherche 3738 CICLY (Centre Pour l'Innovation en Cancérologie de Lyon), Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux, Université Claude Bernard Lyon I, Lyon, France

⁵ Plateforme de Génomique des Cancers, Centre de Recherche en Cancérologie de Lyon (CRCL), INSERM U1052, CNRS UMR 5286, Centre Léon Bérard, Université Claude Bernard Lyon 1, Lyon, France

⁶ Plateforme de Bioinformatique Gilles Thomas, Synergie Lyon Cancer, Centre Léon Bérard, Lyon, France

⁷ Department of Medical Oncology, Centre Léon Bérard, Lyon, France

⁸ Equipe Labellisée Ligue Contre le Cancer, CNRS UMR 3666, INSERM U1143, Paris, France

⁹ Institut Curie, PSL Research University, Paris, France

¹⁰ CASTING - Cancer dynamics, adaptation and modeling, Inria, Inserm, Ecole Normale Supérieure de Lyon, Centre Léon Bérard, Cnrs, Université Claude Bernard Lyon 1, Lyon, France



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Cellular plasticity is a non-genetic mechanism by which cells are able to adapt their phenotype to external pressures and environmental changes, as well as to cell-intrinsic stresses. This process has been known to foster resistance in cancer [1], and is now recognised as a cancer “hallmark” [2]. The epithelial-mesenchymal transition (EMT) is a particular form of plasticity, during which polarised and adherent epithelial cells acquire a more mobile mesenchymal phenotype. In addition to the facilitation of dispersal and metastasis [3], the EMT has been linked to the acquisition of stem-like cell features [4, 5], increased genomic stability [6] and drug resistance [7, 8] in breast cancer. Partial EMT states, often described as hybrid [9], stable [10] or metastable [11], have been associated with increased plasticity. Patients with triple negative breast cancers (TNBC), which lack targetable alterations in the Her2, ER and PR genes, are presented with few therapeutic options and are thus particularly vulnerable to plasticity-mediated therapeutic resistance.

Phenotypic plasticity is a multifaceted process, whose characteristics and dynamics remain elusive. Key questions have been identified, including whether specific cell states provide high adaptability to the subpopulation of cells in which they occur [12], or whether systemic diversity-generating properties allow resistance to emerge [13]. Furthermore, although plasticity increases the adaptability of cancer cells by definition, it is still unknown whether this occurs through a prospective or reactive process, based on either transcriptional selection or adaptation [14] (Fig. 1). Under the prospective scenario, cells are more capable of generating different phenotypes in absence of external pressure, which can increase the likelihood of already adapted phenotypes

in case of environmental change. Under the reactive scenario, cells instead have an intrinsic capacity to dynamically adapt non-genetically once triggered by environmental change, without relying on an increased phenotypic diversity at the population level. Although different from one another, these two processes are however not necessarily mutually exclusive. Nevertheless, few data-driven approaches have yet provided answers to this complex question.

The availability of single-cell technology, that can nowadays provide multi-omics data on individual cells, allows accurate quantification of phenotypic diversity in a cell population. Here we used single cell transcriptomics on in vitro TNBC models of EMT-driven plasticity to investigate how plasticity influences phenotypic diversity before and during chemotherapy, using cell-state agnostic measures. Our results suggest that plasticity promotes a prospective diversification process increasing population phenotypic diversity, which can yield rare pre-adapted states (cells that are by chance already able to adapt to a given environmental change that hasn't occurred yet) before treatment. This in turn suggests that therapeutic strategies targeting cancer cell plasticity should focus on reducing phenotypic diversity, rather than preventing drug-triggered phenotypic switches.

Results

EMT+ status associated with lower genetic heterogeneity but higher phenotypic cell-cell divergence

We used human mammary epithelial cells (hMEC) in which HRAS^{G12V}-mediated oncogenic transformation and ZEB1-mediated epithelial-mesenchymal transition (EMT) can be exogenously induced, respectively via tamoxifen and doxycycline [6] (Fig. 2A, see Methods).

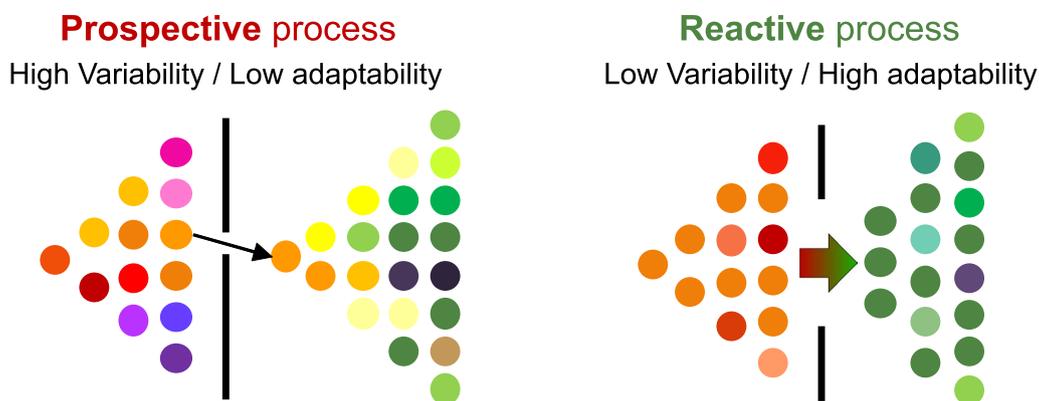


Fig. 1 Types of plasticity-mediated resistance. In a prospective process (left), phenotypic diversity is generated prior to any selective pressure. Upon external pressure, adapted phenotypes can be selected. These adapted phenotypes can correspond to temporary cell states, and can occur in cells that aren't necessarily closely related phylogenetically. In a reactive process (right), populations do not rely on heightened diversity, but on the inherent ability of cells to dynamically transition to a different state upon external pressure, potentially resulting in more adapted cell states

After single-cell cloning, we obtained 2 types of isogenic organoid-like 3D cultures of hMECs, both transformed via Ras activation but either in an EMT context (EMT+, Zeb1-mediated) or in its absence (EMT-). 3D cultures from both conditions showed morphological differences, with EMT+ structures appearing larger and more sparsely spread in the plate (Fig. 2B, Supplementary Fig. 1). 2 replicates of each condition were then subjected to doxorubicin treatment, while 2 were left untreated (8 individual experiments in total). In total, over 48K cells from these models were analysed using UMI-based single cell RNA sequencing (Supplementary Table 1).

We could identify three main clusters across all cells, whose expression of key markers MME (CD10), ITGA6 (CD49f) and EPCAM corresponded to different states of the mammary differentiation cascade [6]: mammary stem cells (MaSC), luminal progenitors (LP) and mature luminal cells (mL) (Fig. 2C, D, Supplementary Fig. 2). We however importantly point out that cell classifications in transformed cells may not be as biologically relevant as in well-regulated homeostatic contexts [15], and that this labelling should be interpreted with caution. Their distribution among replicate experiments highlights a very strong enrichment of the MaSC cluster in the EMT+ population, both before and after doxorubicin treatment (Fig. 2E, $p < 0.001$, Fisher's exact test). In terms of cellular identity, the EMT+ populations thereby demonstrated lower phenotypic diversity than the EMT- ones before and after treatment (Fig. 2F). In addition, EMT+ cells displayed a high enrichment of EMT signatures, which was not impacted by treatment (Supplementary Fig. 3).

We harnessed relative copy number alteration (CNA) inference from single cell transcriptomics data to investigate genetic heterogeneity in our isogenic 3D cultures. We calculated Euclidean distances between each cell and the median CNA profile of each replicate, summarised on a per-cytoband basis (Supplementary Fig. 4, see Methods). This revealed that genetic heterogeneity was lower in the EMT+ populations than in the EMT- ones

($p < 0.001$, two-tailed t-test, Fig. 2G, Supplementary Fig. 5A), in accordance with previous findings on the same models [6]. In addition, genetic cell-cell divergence increased after treatment in all replicates (all $p < 0.001$, two-tailed t-test). Similarly, we quantified phenotypic divergence using Euclidean distances to the mean transcriptomics profile, which revealed that EMT+ cells displayed higher phenotypic divergence than the EMT- ones ($p < 0.001$, two-tailed t-test, Fig. 2H, Supplementary Fig. 5B). This suggests that EMT-driven plasticity, despite specifically enriching for a single phenotype (stem-like cells), provides higher phenotypic diversity via increased cell-cell phenotypic divergence.

Spontaneous EMT followed by phenotypic variation in transformed hMECs

We sought to further characterise the influence of EMT-driven plasticity in a different in vitro TNBC model, where EMT-driven plasticity was observed to spontaneously occur. The model is derived from hMECs which, upon oncogenic transformation by shTP53-HRAS^{G12V}-hTERT, generate distinct EPCAM+/CD24+/CD44- epithelial and EPCAM-/CD24-/CD44+ mesenchymal populations [16] (Fig. 3A). Although cells retain their epithelial or mesenchymal features over time (Supplementary Fig. 6A,B), a fraction of the mesenchymal cells spontaneously started to gradually diverge into an EPCAM-/CD24+/CD44+ "in-between" population (Fig. 3B). When isolated by fluorescence-activated cell sorting (FACS), EPCAM-/CD24+/CD44+ cells could reproduce this gradient variation between CD24- and CD24+ states, which was not observed in standard epithelial nor mesenchymal cells (Supplementary Fig. 6C). Importantly, this rare in-between population appeared stochastically and was not observed in biological replicates of the oncogenic transformation mediated by shTP53-HRAS^{G12V}-hTERT.

As this suggested the presence of a higher-plasticity mesenchymal population, we therefore sorted, clonally expanded and further characterised epithelial (Epi),

(See figure on next page.)

Fig. 2 Characterisation of EMT+ and EMT- inducible 3D models. **A** Experimental scheme. Cells first undergo oncogenic transformation with the induction of the HRAS^{G12V} oncogene, either in a normal epithelial (EMT-) or EMT (EMT+) background, then single-cell cloning to produce EMT- and EMT+ populations. Untreated and doxorubicin-treated replicates are analysed via single cell RNA sequencing. **B** EMT- (top) and EMT+ (bottom) spheroids in culture. **C** UMAP representation and single cell clusters of all cells in EMT- and EMT+ populations (treated & untreated). **D** Expression of mammary lineage markers EPCAM, MME (CD10), ITGA6 (CD49f) and CD24 in each cluster. **E** Cellular composition of each replicate. **F** Cell-identity-based calculation of the Shannon diversity in each replicate. **G** Distributions of genetic distances to median profile for individual cells in EMT- and EMT+ models. Boxplots: Black horizontal bar display the median; boxes display the interquartile range (IQR) between first and 3rd quartiles; whiskers extend to either the minimum between highest value and the first quartile + 1.5 times the IQR (top), and the maximum between the lowest value and the first quartile - 1.5 times the IQR (bottom). **H** Distributions of phenotypic distances to the mean expression profile for individual cells in EMT- and EMT+ populations. NT stands for non-treated, T for treated. Numbers in replicate names indicate the clone identification number in single-cell cloning experiments

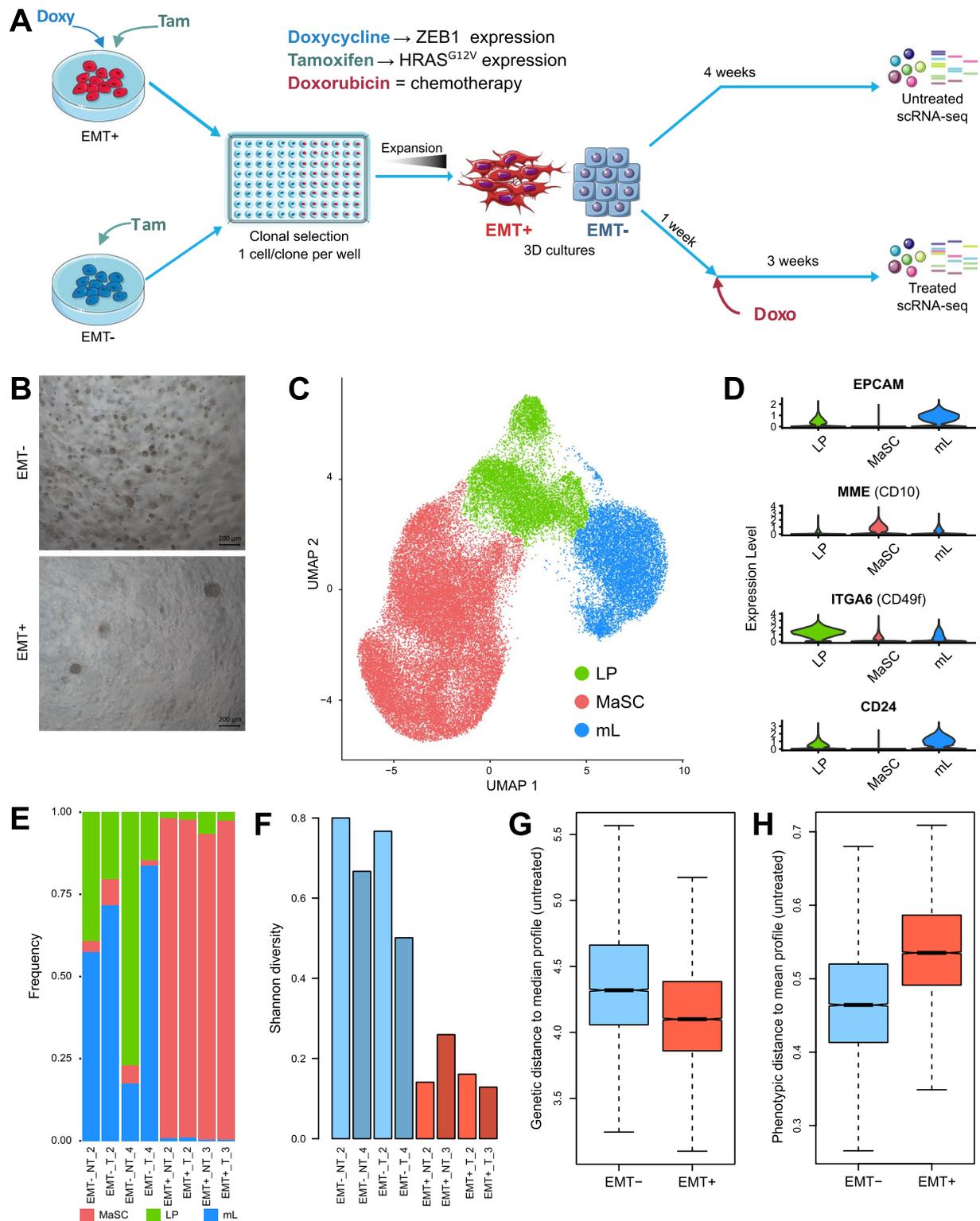


Fig. 2 (See legend on previous page.)

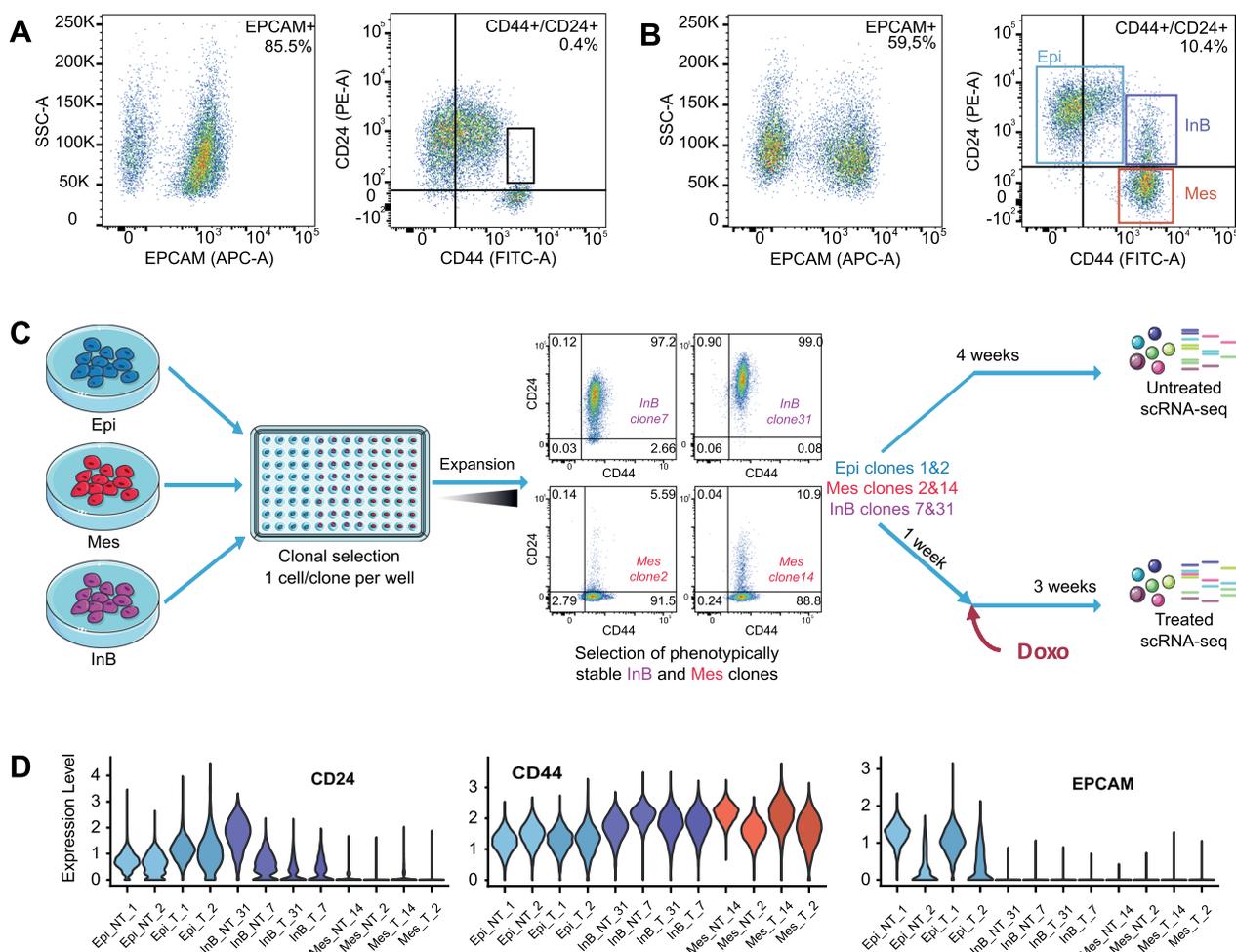


Fig. 3 Spontaneous EMT model and phenotypically variable “in-between” population. **A** Fluorescence-activated cell sorting (FACS) based on the EPCAM (left), CD24 and CD44 markers (right) of the original population after hTERT/HRAS^{G12V}/shTP53 induction. EPCAM+/CD24+/CD44- cells are epithelial, while a second population of spontaneously appearing EPCAM-/CD24-/CD44+ cells is mesenchymal. **B** FACS analysis of a cell culture replicate in which a third population with an EPCAM-/CD24+/CD44+ phenotype spontaneously emerged. Epi (EPCAM+/CD24+/CD44-), Mes (EPCAM-/CD24-/CD44+) and InB “in-between” cells (EPCAM-/CD24+/CD44+) were isolated and analysed separately. **C** Experimental scheme. Epi, Mes and InB cells undergo single cell cloning, before two clones of each were analysed by scRNA-seq with or without doxorubicin treatment. Clones retaining CD24- and CD24+ phenotypes over 4 weeks of culture were specifically selected for Mes and InB cells, respectively. **D** Expression of EPCAM, CD24 and CD44 markers post-scRNA-seq in each replicate. NT stands for non-treated, T for treated. Numbers in replicate names indicate the clone identification number in single-cell cloning experiments

mesenchymal (Mes), and in-between (InB) cells (Supplementary Fig. 6D). Each population was defined solely by FACS-sorting based on EpcAM, CD24 and CD44 expression. As with our inducible 3D models, we analysed isogenic Epi, Mes and InB clones before and during treatment, this time in 2D culture. 2 replicates of each population were subjected to doxorubicin treatment, while 2 were left untreated (Fig. 3C). By following each clone over 4 weeks in culture, we specifically selected clones retaining a CD24+ phenotype as InB, and those retaining a CD24- phenotype as Mes. All three models are isogenic, originating from the same hMEC population. 2 clones of each

population (Epi, InB, Mes) were analysed by scRNA-seq before and after treatment with doxorubicin, resulting in 32K cells analysed (Supplementary Table 2). scRNA-seq analyses confirmed that each population’s phenotype in term of CD24, CD44 and EpcAM expression was conform to cytometry-based expectations before and after treatment (Fig. 3D, Supplementary Fig. 7).

Phenotypically variable mesenchymal cells display highest phenotypic cell-cell divergence

We could identify 5 clusters using the same clustering analyses as for the inducible models, 3 of which were

similar to the MaSC, LP and mL populations previously identified, albeit with unexpectedly strong expression of ITGA6 (CD49f) in all cells (Fig. 4A, B, Supplementary Fig. 8). 2 small isolated clusters could also be identified, and found to be respectively enriched in epithelial (Epithelial outliers) and post-treatment cells (Treated outliers). Once more, we point out that this cellular labelling should be interpreted with caution.

As with our inducible model, an enrichment in MaSC cells was observed in both populations having undergone EMT (Mes and InB), independently of treatment.

Using cell type classification in this model, the Epi population displayed marginally lower diversity compared to the Mes and InB populations, before and after treatment (Fig. 4C, D). Due to their small size (<200 cells each), both the Epithelial and Treated clusters were excluded from further analyses.

Using individual cell divergence from the median genomic and mean transcriptomic profiles (see Methods for details), InB cells were characterised by the highest genetic and phenotypic divergence (Fig. 4E, F, Supplementary Fig. 9–10). Epi cells displayed the lowest

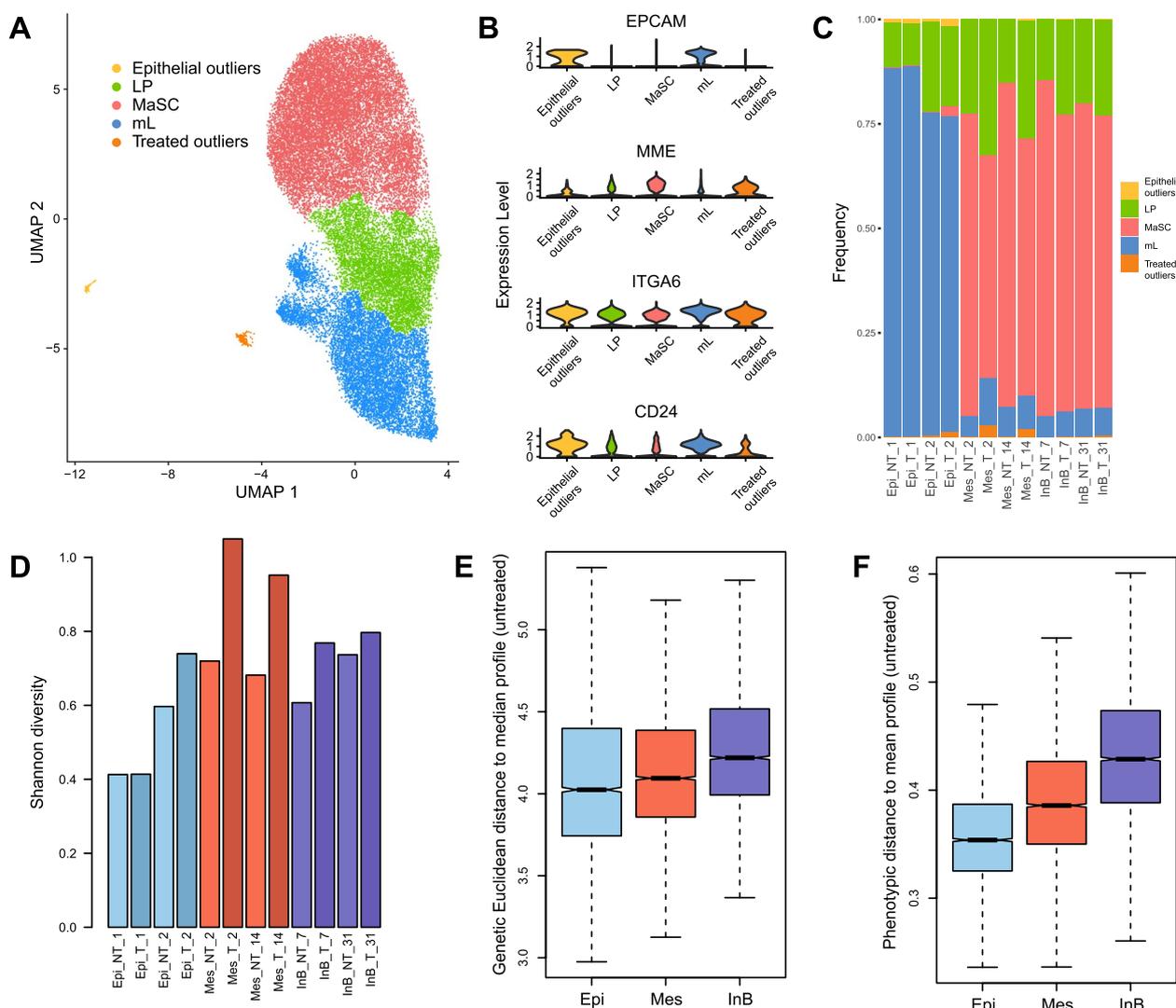


Fig. 4 Characterisation of the spontaneous EMT model. **A** UMAP representation and single cell clusters of all cells in Epi, Mes and InB populations (treated & untreated). **B** Expression of mammary lineage markers EPCAM, MME (CD10), ITGA6 (CD49f) and CD24 in each cluster. **C** Cellular composition of each replicate. **D** Cell-identity-based calculation of the Shannon diversity in each replicate. **E** Distributions of genetic distances to median profile for individual cells in Epi, Mes and InB populations. **F** Distributions of phenotypic distances to the mean expression profile for individual cells in Epi, Mes and InB populations. NT stands for non-treated, T for treated. Numbers in replicate names indicate the clone identification number in single-cell cloning experiments

genetic and phenotypic divergence. As with our inducible models, genetic diversity appeared to increase under treatment for each replicate (Supplementary Figs. 5 and 10, all $p < 0.001$, two-tailed t-test). Our results, based on incomplete relative CNA data, thus cannot support the hypothesis of the selection of subpopulations harbouring specific genomic alterations.

To further account for coverage differences between cell populations, we also performed 5 random downsampling experiments (see Methods). After downsampling, individual and population average profiles were compared using Euclidean distances to quantify intra-population phenotypic divergence. This yielded results similar to the non-downsampled analyses (all $p < 0.001$, Supplementary Fig. 11), thus confirming that EMT-driven plasticity appears to promote cell–cell phenotypic divergence.

InB cells are enriched in rare pre-adapted outliers

We next investigated the impact of treatment on cell–cell phenotypic divergence in all 5 populations. We could not detect specific genomic gains or losses significantly associated with doxorubicin resistance (Supplementary Fig. 12, see Methods). We observed a trend for phenotypic divergence to decrease after treatment in the higher plasticity populations of each model (Fig. 5A). This could indicate stronger selection at the phenotypic level in high-plasticity populations under doxorubicin treatment. To investigate whether this coincided with the presence of pre-existing adapted cells, we first determined post-treatment overexpression signatures that were specific to each of the 5 in vitro models (EMT⁻, EMT⁺, Epi, Mes, InB). Using gene signatures, we then investigated the presence of pre-adapted cells, defined as those with high post-treatment signature enrichment in the untreated population (see Methods, Supplementary Table 3).

Using a threshold corresponding to a Z-score of 1 in the post-treatment distribution of each model (mean of the post-treatment enrichment score distribution + 1 time its standard deviation), we found that, before treatment, high plasticity EMT⁺ and InB populations displayed a >40-fold increase in pre-adapted cells compared to lower plasticity populations (3.03% and 3.20% vs 0.00%, 0.16% and 0.05%, two-tailed t-test: $p = 0.003$, Fig. 5B). More relaxed enrichment score thresholds

corresponding to the top 25% and top 50% of the treated populations also yielded >tenfold increases in EMT⁺ and InB populations (Supplementary Fig. 13). Post-treatment signatures are thus derived from treated cells (compared to gene expression in untreated cells), while pre-adapted cells are untreated cells displaying high expression of these post-treatment signatures, but in absence of treatment.

We then investigated whether the high-scoring pre-adapted cells were outliers in their respective untreated populations. The score distribution for the presence of pre-adapted cells in the untreated InB population had high skewness and kurtosis, indicating a heavy right-tail bias and a high prevalence of outliers (Fig. 5C, D). We report that the EMT⁺ populations displayed fewer high-scoring outliers than other populations ($p < 0.001$, Supplementary Fig. 14), and was more similar to its post-treatment counterpart as a whole ($p < 0.001$, Supplementary Fig. 15, two-tailed Mann–Whitney U test). On the other hand, the InB distribution was heavily skewed towards the right and was the most enriched in high-scoring outliers ($p < 0.001$, two-tailed Mann–Whitney U test, Supplementary Fig. 14). This suggests that the entire EMT⁺ population was more adapted prior to treatment, while InB cells appeared to contain a rare subset of pre-adapted cells that phenotypically stood out from the rest of the population. These results were also reproducible in downsampled analyses (Supplementary Fig. 16). In addition, pre-adapted cells from the untreated InB population showed expression patterns anti-correlated to a dormancy expression signature [17], suggesting that doxorubicin treatment did not select for initially quiescent cells ($p < 0.001$, Pearson correlation; Supplementary Fig. 17).

We analysed the score distributions specific to Epi, Mes, and InB post-treatment signatures in the other two populations, which highlighted that the InB population also displayed strong outliers for the Epi post-treatment signature (Supplementary Fig. 18). Signature enrichment analyses using the same number of genes selected at random (see Methods) furthermore indicated low expectations for skewness and kurtosis in all populations (Supplementary Fig. 19). This suggested that the outlier prevalence observed in the post-treatment scores of untreated InB cells appears to be specific to resistance

(See figure on next page.)

Fig. 5 Pre-adapted cells. **A** Difference in phenotypic divergence before and after treatment in each replicate of each population. Values above 0 indicate an increase in phenotypic divergence after treatment, values below 0 indicate a decrease. **B** Percentage of pre-adapted cells (high-scoring cells based on post-treatment expression signatures) in the untreated populations. **C** Skewness and kurtosis of the respective post-treatment signature enrichment score distributions in each untreated population. Skewness > 0 and < 0 respectively indicate right- and left-tailed distribution biases, while higher kurtosis indicates a higher prevalence of outliers. **D** Density distributions of post-treatment signature enrichment Z-scores in untreated cells. Vertical bars indicate the lower boundaries of the top 5% in each population

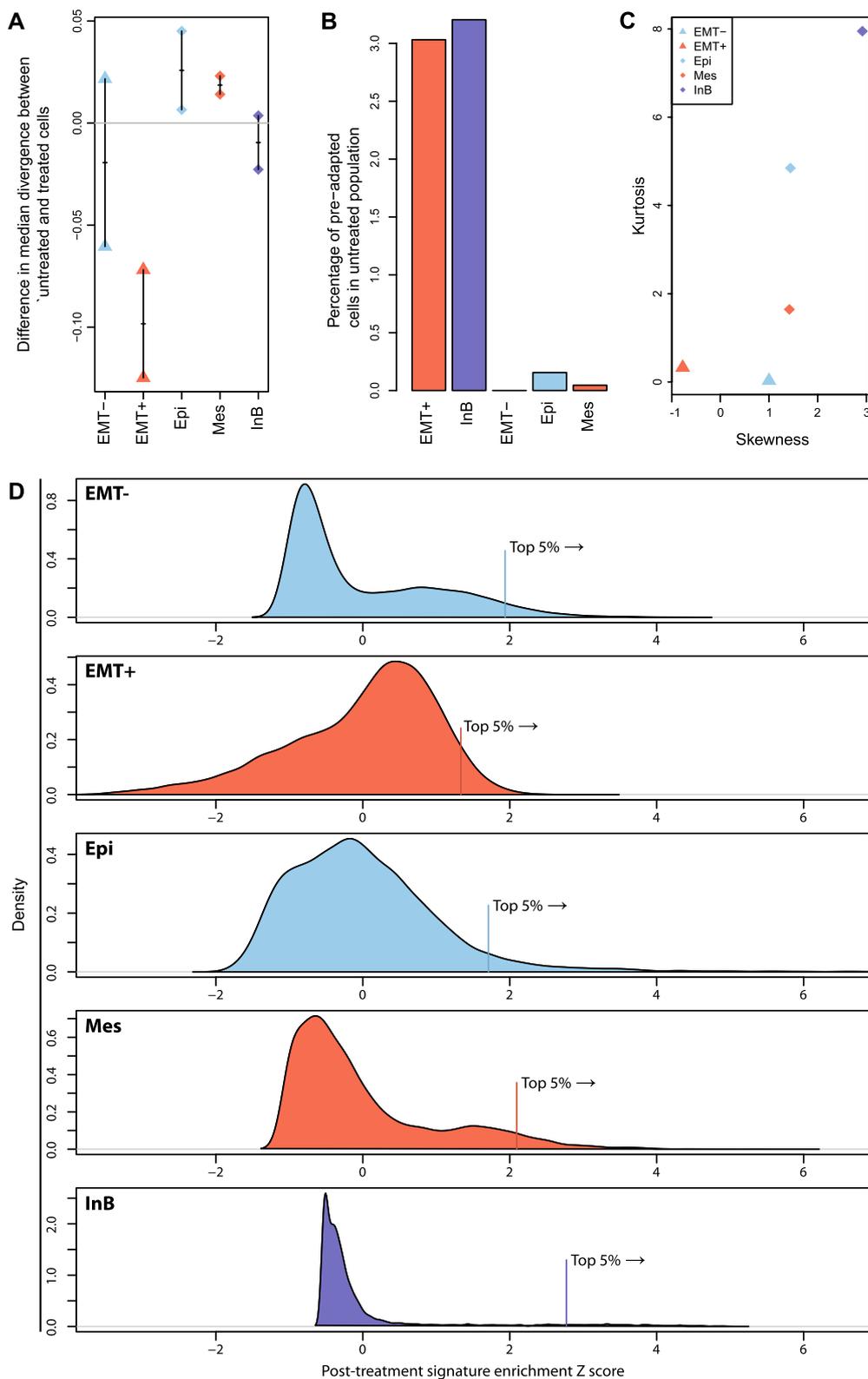


Fig. 5 (See legend on previous page.)

mechanisms in this population, and unlikely to have occurred by chance.

After spontaneously undergoing an EMT and acquiring the ability to increase phenotypic variation, InB cells thus appear to display an extreme form of plasticity. This helps them promote phenotypic variation by increasing the cell–cell divergence in gene expression, which can stochastically facilitate the emergence of better adapted phenotypes. This in turn indicates that plasticity-mediated resistance is a prospective process rather than a reactive one in our model.

Plasticity signature predicts survival in TNBC patients treated by chemotherapy

InB cells displayed more plasticity and more phenotypic divergence than Mes cells in our spontaneous EMT model, while their CD24+ expression status could furthermore correspond to a partial-EMT state. Using signature enrichment to evaluate their mesenchymal status, we observed that InB cells were in fact more mesenchymal than Mes cells (Fig. 6A, Supplementary Fig. 20). Their higher plasticity was therefore unlikely to stem from a partial EMT.

We defined over- and under-expressed genes in MaSC-like InB cells compared to MaSC-like cells from the Epi and Mes populations (see Methods). Further GSEA analyses revealed that multiple pathways linked to multicellular organism regulation and differentiation were impacted, with a prevalence of pathways linked to neurogenesis (Fig. 6B). To understand whether these expression changes could impact resistance in vivo, we investigated their impact on survival in 221 TNBC cases treated by chemotherapy from the METABRIC [18, 19]

cohort. We report that patients with low expression of the genes under-expressed in InB cells displayed significantly worse survival ($p=0.03$, Cox proportional hazard model, $p=0.03$, logrank test, Fig. 6C).

This suggests that our results on the impact of plasticity on phenotypic diversity dynamics and therapeutic adaptation, which are derived from in vitro models, also bear significance in clinical settings.

Discussion

We investigated the impact of EMT-driven plasticity on the dynamics of phenotypic diversity in response to chemotherapy using single-cell transcriptomics and in vitro TNBC models. We report that populations with increased plasticity (EMT+ cells in the inducible EMT model; InB and, to a lesser extent, Mes cells in the spontaneous EMT one) also displayed higher cell–cell phenotypic divergence. In InB cells, characterised by a unique ability to gradually diverge from a spontaneously occurring yet initially stable mesenchymal phenotype, we furthermore report that high cell–cell divergence can facilitate the emergence of rare pre-adapted phenotypes.

It is first worth noting that our results are only representative of the early stages of chemotherapy, as our experiments’ time scale was too short to yield fully resistant clones. This is in line with the absence of detectable genetic selection, as the genetic diversity increased significantly after treatment in all replicates in both inducible and spontaneous EMT models. This was confirmed using pairwise genetic divergence measures (Supplementary Fig. 21), and could in part be explained by the impact of doxorubicin, a DNA-intercalating agent causing DNA damage [20], thus likely to

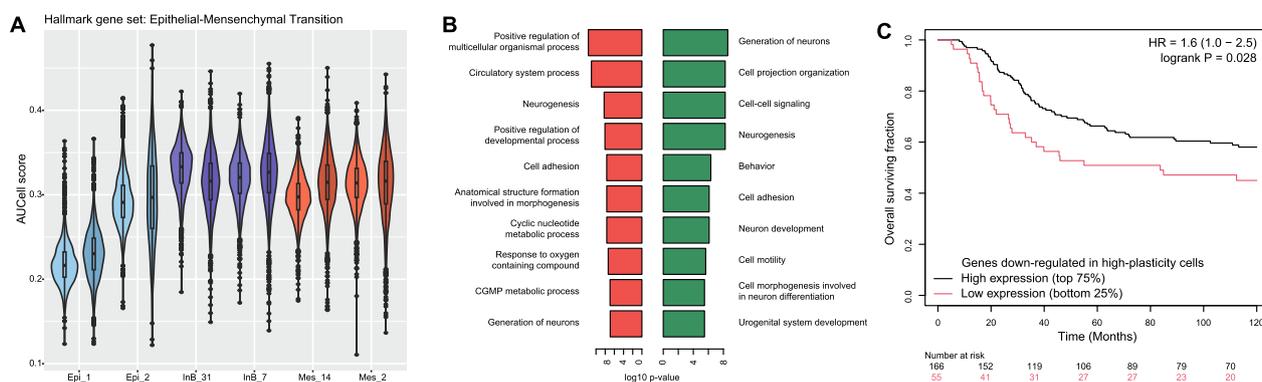


Fig. 6 Plasticity signatures. **A** Gene set enrichment scores for the Epithelial-Mesenchymal Transition hallmark gene set in each replicate of the Epi, Mes and InB populations, using the AUCell software. For each clone: clear-colour samples on the left are untreated, darker-colour samples on the right are treated. **B** Top 10 enriched Gene Ontology (GO) biological processes in genes down- (red, left) up-regulated (green, right) in MaSC cells from the InB population compared to MaSC cells from the Epi and Mes populations. **C** Kaplan–Meier survival curves for TNBC patients treated with chemotherapy in the METABRIC cohort. Patients categorised as “Low expression” display the weakest expression of genes downregulated in MaSC cells from the InB population

increase the presence of cell-specific large copy number alterations and genetic diversity. We only analysed scRNA-seq-derived relative copy number alteration scores, and our analyses therefore lack the precision necessary to accurately reflect the genetics underlying each population's evolution. Our divergence measures are furthermore not a direct measure of selection strength, which may be operating beyond our limited ability to detect it. As such, although our data suggests that prospective phenotypic diversification is a likely scenario in EMT-driven resistance in our breast cancer model, we cannot fully exclude a possible contribution of the reactive adaptation scenario in this context (nor in others). Approaches combining scRNA-seq with lineage tracing [21–23] may thus provide complementary evidence to more fully delineate the impact of EMT on the dynamics of resistance in the future.

We nonetheless demonstrate that phenotypes expressing genes typically overexpressed 3 weeks following treatment were already present in highest plasticity populations of each model, at much higher frequencies than in lower plasticity populations. In InB cells, pre-adapted cells represented a minority of high-scoring cells strongly diverging from the population average phenotype. Our results consequently suggest that cells with the highest level of plasticity are able to more easily foster pre-adapted clones, thanks to a prospective diversification process in which the cell–cell divergence is increased in the population. These pre-adapted phenotypes are not necessarily stable, as stochastically occurring transient states can also foster resistance [24, 25]. Furthermore, phenotypic similarity between stochastically occurring cell states does not imply a recent common ancestor, and selection of a specific phenotype may thus not translate into a signal detectable at the genetic level. Such phenotypic diversification is however consistent with a bet-hedging evolutionary strategy [26, 27], in which cancer cells can increase their adaptation potential, possibly by hijacking the cell–cell variability inherent to normal developmental processes [28] such as the EMT.

The classification of cells that escaped the constraints of normal homeostasis, and may thus not reflect a maintained phenotypic archetype, is a difficult task. Classification-agnostic divergence measures, however, do not depend on the accuracy of either identification or annotation of cellular subtypes. In this work, they provided information that was absent from diversity measures based on traditional cell subtype classifications, whose biological relevance we cannot claim to be universally accepted. This is in agreement with previous suggestions that static cell classifications can provide limited and incomplete information for intra-tumour phenotypic heterogeneity quantification purposes [15], and provides

another argument in favour of classification-agnostic approaches in this aim.

In our spontaneous EMT model, the InB high-plasticity mesenchymal population was characterised by its ability to stably re-express luminal epithelial marker CD24 [29]. Interestingly, CD24+/CD44+ InB cells were however deemed even more mesenchymal than the more classical CD24-/CD44+ Mes population [30, 31], based on scRNA-seq data and using three different expression signatures. They would therefore correspond better to a “hybrid” EMT state rather than to a “transient” or “partial” one, although these alternatives may be appropriate to other cell types and models. This furthermore suggests that increased EMT-driven plasticity is not necessarily associated with an intermediate phenotype between the two end states of the EMT, but can arise from an ability to increase phenotypic variation once a full, unbridled EMT occurred in a cancerous context. Additional studies will be needed to fully characterise the mechanism(s) providing InB cells with increased plasticity, as well as to determine how it can be detected and prevented in cancer. Although cells expressing both classical CD24 and CD44 markers have typically been used to determine hybrid EMT states and their relation to survival [9] and resistance [8] in breast cancer, complementary markers, such as ITGB4 [32] may provide additional and complementary information to identify cells with enhanced plasticity.

In this regard, we could associate a scRNA-seq-derived down-regulated gene signature for high-plasticity, using the specific characteristics of stem-like InB cells. This signature had prognostic value in TNBC patients treated by chemotherapy: low-scoring, high-plasticity patients displayed worse survival. The fact that plasticity appears to facilitate the emergence of already adapted phenotypes further suggests that, rather than aiming to circumvent a reactive mechanism of non-genetic adaptation, therapeutic solutions targeting cancer cell plasticity should focus on reducing phenotypic diversity prior to treatment. Our study therefore highlights that better characterisation of EMT-driven plasticity's impact on phenotypic diversity has strong translational potential to improve the clinical care of breast cancer patients.

Material and methods

Cell culture

Primary human mammary epithelial cells (hMEC) were immortalised by human Telomerase Reverse Transcriptase (hTERT) and named HME.

For the EMT- and EMT+ models, a retroviral pLNCX-2neo-RAS_{ER} expression vector containing an inducible activation of H-RASv12 was included in HME cells. To produce inducible ZEB1, cells were transduced with

pTRIPz-puro-ZEB1 expression vector. Heterogeneous HME-RAS_{ER}-ZEB1 cells were subdivided into one clonal population named TF16. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM)/HAMF-12 medium with 1% glutamax (ThermoFisher Scientific, #31331093) with 10% heat-inactivated foetal calf serum (Eurobio Scientific, #CVFVSF0001), 100 U/mL penicillin/100 µg/mL streptomycin (ThermoFisher Scientific, #15140-122), 10 mg/mL insulin (NovoRapid, NovoNordisk, #7200980), 10 ng/mL human epidermal growth factor (EGF) (Peprotech, #AF-100-15), 0.5 mg/mL hydrocortisone (Sigma-Aldrich, #H0888), 0.5 µg/mL puromycin (Invivogen, #ant-pr-1), and 100 µg/mL geneticin (ThermoFisher Scientific, #10131-027).

For the Epi, Mes and InB models, primary hMEC were infected with pRetroSuper puro-shp53, pBabe neo-RASG12V and pBabe hygro-hTERT (RRID:Addgene_1773), as previously described [16], and named HME-shp53-RAS thereafter. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM)/HAMF-12 medium with 10% heat-inactivated foetal calf serum, 100 U/mL / 100 µg/mL penicillin / streptomycin, 10 mg/mL insulin, 10 ng/mL human EGF, 0.5 mg/mL hydrocortisone, 0.5 µg/mL puromycin, and 100 µg/ml geneticin. All cell lines were kept at 37 °C in a 5% CO₂/95% air incubator, and were routinely tested negative for mycoplasma contamination using the Lonza MycoAlert PLUS Mycoplasma Detection Kit (Lonza, #LT07-318).

Lentiviral and retroviral infections

To produce H-RASV12_{ER} (RAS_{ER}) cells for the EMT- and EMT+ models, 2 × 10⁶ Phoenix cells (ATCC) were transfected by GeneJuice[®] Transfection Reagent (MerckMillipore; #70967-3) with 10 µg of pLNCX2-neo-RAS_{ER}. 48 h post-transfection, the supernatant was collected, filtered, supplemented with 5 µg/ml of polybrene (Sigma-Aldrich; #H9268) and combined with targeted cells for 6 h. Cells were selected 48 h following the infection with 100 µg/ml geneticin. To induce RAS activation in HME-derived cells, (Z)-4-Hydroxytamoxifen (4-OHT) (Sigma-Aldrich, #H7904) was added in the medium at 500 nM every 2–3 days.

To produce inducible ZEB1, 2 × 10⁶ HEK293T cells (RRID:CVCL_0063) were transfected with GeneJuice[®] Transfection Reagent with 13.02 µg of total lentiviral expression vectors (5.1 µg of pCMVdeltaR8.91, 1.32 µg pHCMVG-VSVG and 6.6 µg interest plasmid (pTRIPz-puro-ZEB1)). 48 h post-transfection, the supernatant was collected, filtered, supplemented with 5 µg/ml of polybrene and combined with the targeted cells. Cell selection was made 48 h post-infection with the addition of puromycin at 0.5 µg/mL (Invivogen; #ant-pr-1) in

the medium. To induce ZEB1 expression in HME cells, doxycycline hyclate (Sigma, #D9891) was added in the medium at 1 µg/ml every 2–3 days.

Plasmids

The retroviral plasmid pLNCX2-neo-RAS_{ER} [33] was purchased from Addgene (RRID:Addgene_67844). The inducible pTRIPz-puro-ZEB1 lentiviral plasmid was modified from the pTRIPz-puro-RFP (Openbiosystems, #RHS4750) by replacing the RFP by the full-length human ZEB1 cDNA from pBabe-ZEB1-puro vector.

3D culture

TF16 EMT+ and TF16 EMT- cells were collected and added in Advanced DMEM/F-12 (ThermoFisher Scientific, #12634-010) with 10% R-Spondin conditioned medium (500 ng/mL; homemade), 100 ng/mL Recombinant Human Heregulinβ-1 (Peprotech, #100-03), 5 ng/mL Recombinant Human KGF (FGF-7) (Peprotech, #100-19), 20 ng/mL Recombinant Human FGF-10 (Peprotech, #100-26), 5 ng/mL Human Epidermal Growth (EGF) (Peprotech, #AF-100-15), 100 ng/mL Recombinant Human Noggin (Peprotech, #120-10C), 500 nM A83-01 (Bio-Techne, #2939), 5 mM Y-27632 dihydrochloride (Bio-Techne, #1254/10), 500 nM SB202190 (Sigma, #S7067), 1% B27 supplement (ThermoFisher Scientific, #17504-44), 1,25 mM N-Acetyl-L-cysteine (Sigma, #A9165-5 g), 5 mM nicotinamide (Sigma, #N0636), 1% Supplement GlutaMAX[™] (ThermoFisher Scientific, #35050061), 10 mM hepes (ThermoFisher Scientific, #15680-056), 100 U/mL/100 mg/mL penicillin/streptomycin (ThermoFisher Scientific, #15140-122), and 50 µg/mL primocin (Invivogen, #Ant-pm-05). 40 µL of this suspension containing 20,000 cells was gently mixed with 40 µL of Cultrex Reduced Growth Factor Basement Membrane Extract, Type 2, Select (BME) (Bio-Techne, #3536-005-02) and was placed in a well of a prewarmed ultralow-attachment 96-well plate (Corning, #003474). To allow the mix to harden, the plate was incubated at 37 °C for at least 30 min. Upon complete gelation of the suspension, 100 µL of medium completed with treatments was gently added to each well and the plate was incubated at 37 °C with 5% CO₂. To avoid detachment of the matrix, 10 µL of medium was added twice a week, and medium was completely changed when it became yellow. After 1 week, doxorubicin hydrochloride (Euromedex, #TA-T1020) was added to the medium every 2–3 days in each condition. After 3 weeks of subsequent culture with or without doxorubicin treatment, spheroids were harvested with TrypLE[™] Express (ThermoFisher Scientific, #12605010) and stained with DAPI. A FACSMelody[™] Cell Sorter or FACSaria III (BD Biosciences) was used

to select unique cells and avoid multiplets for single cell experiments. Unique cells collected were brought to the dedicated platform for single cell RNA sequencing.

Flow cytometry sorting

For CD44/CD24/EPCAM sorting, cells were stained with primary antibodies from Miltenyi Biotech: anti-CD44 labelled Fluorescein-5-isothiocyanate (FITC) (clone REA690, #130-113-341, dilution 1:400, RRID:AB_2726117), anti-CD24 labelled phycoerythrin (PE) (clone REA832, #130-112-656, dilution 1:50, RRID:AB_2656557), anti-CD326 (EPCAM) labelled allophycocyanin (APC) (clone REA764, #130-111-000, dilution 1:400, RRID:AB_2657497). hMEC-shp53-RAS cells were washed with PBS, stained with antibodies protected from light at 4°C during one hour, and then incubated with 4',6-Diamidino-2-phenylindole dihydrochloride (DAPI) for 15 min. hMEC-shp53-RAS cells were sorted with a FACSMelody™ Cell Sorter or FACSARIA III (BD Biosciences) under 3 conditions: EpCAM+/CD24+/CD44- (epithelial cells “Epi”), EpCAM-/CD24-/CD44+ (mesenchymal cells “Mes”) and EpCAM-/CD24+/CD44- (in between cells “InB”). TF16 cells were washed with PBS and incubated with DAPI for 15 min. DAPI negative TF16 cells were sorted with a FACSMelody™ Cell Sorter or FACSARIA III (BD Biosciences). One cell per well was incubated in a 96 well plate for each cell line and condition. Medium was changed twice a week. Once clones reached more than 1 million cells, they were deemed viable for future experiments. Sorted hMEC-shp53-RAS clones were cultured and flow cytometry analysis was performed every week for 5 weeks, to confirm the phenotype of each clone. The most stable candidates were selected for subsequent treated/untreated experiments and scRNA-seq analyses.

Flow cytometry analysis

Fresh hMEC-shp53-RAS cells were washed with PBS, stained with EPCAM, CD24 and CD44 antibodies for at least 1 h at 4 °C protected from light, then washed 3 times with PBS and stained 15 min with DAPI. Data acquisition was performed with a FACS BD LSR Fortessa (Becton Dickinson). Daily calibration of the cytometer was assessed using Cytometer, Setup and Tracking (CST) beads (Becton Dickinson) according to the manufacturer's instructions. Sample acquisition was set up to have 10,000 events across all conditions at an adapted flow rate. Post-analysis was done using FlowJo v10 software (FlowJo, RRID:SCR_008520). Antibody validation information is available on the manufacturers' websites.

Cell cytotoxicity assay

MTT assays were assessed on 2D cells to determine IC60 (40% alive cells and 60% dead cells, Supplementary Fig. 22). Cells were plated at 0.5×10^4 per well in 96-well plates and exposed to serial concentrations of doxorubicin for 96 h. Treated and non-treated cells were incubated with 0.5 mg/ml 3-(4,5-Dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) solution (TOCRIS, #5224). After two hours of incubation, the incubation medium was removed and the blue MTT-formazan product was extracted with acidified isopropyl alcohol (0.04 N HCl). After 10 min extraction at room temperature, the absorbance of the formazan solution was read spectrophotometrically at 570 nm.

Doxorubicin treatment

According to results of MTT assays, the determined IC60 concentrations of doxorubicin applied on the inducible 3D models were 19 nM for TF16 EMT+ clones and 17,3 nM for TF16 EMT- clones (inducible model). To find the concentration of doxorubicin adapted for each hMEC-shp53-RAS clone (spontaneous model), replicates were cultured for 5 weeks using different reduced doses of doxorubicin compared to the original TF16 population's IC50: 8, 10, 12, 14 and 16 nM. Doxorubicin doses were added to the medium every 2 days. After 4 weeks of doxorubicin treatment, the replicate that still survived under the highest possible dosage was collected for each clone. The selected dosages were 8 nM doxorubicin for clones 2 (Mes) and 31 (InB), 10 nM for clones 7 (InB) and 14 (Mes), and 14 nM for Epi clones 1 and 2. They were then sorted with negative DAPI staining with FACSARIA III (BD Biosciences) and brought for the scRNA-seq procedure.

Single cell RNA sequencing

Fresh hMEC-shp53-RAS cells were collected and counted to have 2 million of cells in 1 mL of PBS with 0,1% of Bovine Serum Albumin (BSA). Fresh TF16 EMT- and TF16 EMT+ spheroids were harvested, washed with PBS, filtered with 40 µM Flowmi™ cell strainer filter (SP Bel-ART, #136800040) and counted. Cells were incubated for 15 min with DAPI protected from light in PBS with 0,1% BSA. A FACSMelody™ Cell Sorter was used to keep singlet and live cells only. Live cells were counted a second time. The number of live cells was determined with a cell counter to obtain an expected cell recovery population of 5,000 cells per channel, then loaded on a 10×G chip and run on the Chromium Controller system (10×Genomics) according to manufacturer's instruction. Single-cell RNA-seq libraries were generated by the Cancer Genomic Platform of the CRCL with the Chromium Single Cell 3' v.3.1 kit (10XGenomics, no. PN-1000121)

and sequenced on the NovaSeq platform (Illumina) to obtain around 50,000 reads per cell.

Single cell basic analyses and clustering

3D culture experiments with the inducible models were processed individually, while experiments with the spontaneous EMT models were processed in three batches of 4 samples pooled using the CellPlex technology (10X Genomics): the first batch contained the two replicates of the Mes (2 and 14) and InB (7 and 31) populations, untreated; the second batch contained the two replicates of the Mes and InB populations, treated; the third batch contained the two Epi replicates (1 and 2), treated and untreated.

Single cell expression data were analysed using the Seurat package (RRID:SCR_016341) [34]. Inducible cells (EMT+ and EMT- models) with fewer than 1,500 genes or 2,000 UMIs, or with more than 10% of reads mapped to mitochondrial RNA were filtered out. Spontaneous EMT cells (Epi, Mes and InB models) with fewer than 1,500 genes or 5,000 UMIs, or with more than 7% of reads mapped to mitochondrial RNA were filtered out. Inducible (48K) cells and spontaneous cells (32K) were pooled, normalised and analysed separately. Cell cycle regression was not performed, as major differences are expected depending on cells' mammary differentiation subtype. Harmony [35] was used for further batch normalisation. Clusters were obtained by restricting analyses to the genes from the MaSC, LP and mL signatures referenced in Lim et al. [36] (500, 156 and 500 genes, respectively). The 30 closest neighbours were used for UMAP analyses and clusters were defined on UMAP data, using the original Louvain algorithm with a resolution of 0.01 for inducible cells, and 0.02 for spontaneous EMT cells. Gene signature enrichment analyses for all signatures were performed using the AUCell package [37]. EMT expression signatures were obtained from MsigDB [38], Gavish et al. [39] and Tan et al. [11].

Genetic and phenotypic divergence

Copy number alteration (CNA) profiles for all cells were calculated using the inferCNVplus software (<https://github.com/CharleneZ95/infercnvPlus>), based on inferCNV of the Trinity CTAT project: <https://github.com/broadinstitute/inferCNV>), by pooling cells from each individual replicate before and after treatment. scRNA-seq data from normal, untransformed hMEC in vitro cultures were used as normal references, thus using 3 populations per calculation (before treatment, after treatment, reference). Although scRNA-seq-derived CNA profiles are relative and cannot be used to infer exact CNAs, we used them to calculate average profiles for each population, and evaluate how much each

cell diverged from it. Individual CNA profiles were first averaged on a per (minor) cytoband basis, according to Ensembl hg38 annotations. Individual per-cytoband profiles were then summarised using the median value across all cells for each population, separating pre- and post-treatment cells. We finally used Euclidean distances to quantify the intra-population genetic divergence (each individual cell profile compared to the respective median CNA profile, and pairwise distances between all cells).

A similar approach was used to quantify phenotypic divergence, by first averaging the Seurat-normalised counts per gene across all cells in a population, to obtain mean expression profiles. We used the mean rather than the median to avoid obtaining values of 0 for infrequently expressed genes. Pre- and post-treatment cells from the same replicate were analysed separately. To better account for coverage (number of reads per cell) differences between cell populations, we used non-parametric correlation-based distances (Spearman's rho) between individual cell profiles and population average profiles to quantify the intra-population phenotypic divergence.

We performed 5 replicates of a downsampling procedure by randomly sampling (without replacement) at most 5,000 UMIs per cell in the inducible EMT model, and 15,000 UMIs per cell in the spontaneous EMT model. This discrepancy is explained by the higher baseline coverage in the spontaneous EMT model (Supplementary Table 1). Cells with fewer than 5,000 or 15,000 UMIs (respectively in inducible and spontaneous EMT models) were not downsampled. As previously, average transcriptomic profiles were calculated for each population. Phenotypic divergence was then calculated using Euclidean distances between the now similarly covered cells and the average profiles. Given the low variability between the 5 replicates, no further downsampling replicates were performed.

Recurrent gains and losses post-doxorubicin treatment

For each of the 10 clones from the 5 populations (EMT-, EMT+, Epi, Mes, InB), we calculated the post-treatment difference in per-cytoband median CNA score (a positive value meaning an increase for the locus after treatment). We used single-samples t tests on all cytobands to estimate if the post-treatment difference of each locus was significantly different from 0, and applied Bonferroni correction for multiple testing.

Pre-adapted cell identification

Each population (EMT-, EMT+, Epi, Mes, InB) was analysed individually, by pooling clone replicates but separating cells according to treatment status. Differential expression was determined comparing the treated to the untreated cells using the FindMarkers function (Seurat

package). The top 50 genes with $p < 0.001$ and log fold change > 1 in the treated population were defined as post-treatment markers. Single-sample signature enrichment analyses were performed using the AUCell software [37] on all cells (treated and untreated). In each population, respectively, we then defined pre-adapted cells as the untreated cells displaying an enrichment score superior to a given threshold defined on the treated cell score distribution. We used thresholds corresponding to a Z-score of 1 (mean of the treated distribution + 1 time its standard deviation), as well as the top 25% and top 50% of the treated distribution.

Distribution skewness and kurtosis were calculated in untreated cells with of the EnvStats R package (Fisher methods) on the respective Z-normalised post-treatment scores of each population. To assess expectations of skewness and kurtosis of post-treatment score enrichment in each untreated population, 50 genes expressed in at least 5% of cells (treated and untreated) were selected at random 50 times. To account for differences in coverage and number of cells, genes expressed in at least 5% of cells were defined separately for the inducible (EMT-, EMT+) and spontaneous (Epi, Mes, InB) models.

METABRIC survival analysis

Cells classified as MaSC, due to high MME expression, from the InB population were compared to those classified similarly in the Epi and Mes populations. This was achieved with the FindMarkers function (Seurat package), selecting all genes with a corrected $p < 0.001$ and an absolute log fold change > 2 . Genes with a positive fold change were grouped in the UP signature, while genes with a negative fold change were grouped in the DOWN signature. Z-score normalised expression data for the METABRIC cohort [18, 19] was retrieved using the cBioPortal [40–42], along with paired clinical data. To match our in vitro models, we selected only triple negative breast cancer cases (reported Pam50 status differing from “Her2”, “LumA” and “LumB”) treated with chemotherapy (specific information on the exact chemotherapeutic agent used in each case was not available). This resulted in a subcohort of 221 females, with a median age at diagnosis of 49.6 years (range: 26.7–78.3). Gene set enrichment analyses (GSEA) on bulk RNA-seq data were performed using the corto package [43]. Cox proportional hazard models and logrank tests were respectively performed using the surviplot (<https://github.com/aroneklund/surviplot>) and survival packages [44]. Survival analyses were right-censored at 10 years post diagnosis.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12935-025-03637-w>.

Supplementary material 1

Author contributions

Study and experimental design: AP, APM, MO and PM. Experiments: LM, FF, CC, FA, CD. Quality control: LM, FF, CD, LT. Data analysis: LM, AC, LT, PM. Supervision: PS, APM, MO, PM. Manuscript redaction: LM, FF, APM, MO, PM.

Funding

This research was financially supported by the Institut Thématique Multi-Organismes (ITMO) Cancer of AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences & Health) within the framework of the Cancer Plan. AC is funded by a scholarship from the Ligue Nationale contre le Cancer.

Availability of data and materials

All single cell expression data were deposited in the SuperSeries GSE263731 on the GEO database.

Code availability

No ad hoc packages were developed in this study. The following standalone software was used: cellranger version 7.1.0 (10X Genomics). The R language was used for all data analyses, with the following packages: Seurat [34], Harmony [35], AUCell [37], infercnvplus [45], corto [43], surviplot (<https://github.com/aroneklund/surviplot>), survival [44] and EnvStats [46].

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 September 2024 Accepted: 7 January 2025

Published online: 03 February 2025

References

- Black JRM, McGranahan N. Genetic and non-genetic clonal diversity in cancer evolution. *Nat Rev Cancer*. 2021;21:379–92.
- Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov*. 2022;12:31–46.
- Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol*. 2017. <https://doi.org/10.1038/nrclinonc.2017.44>.
- Morel A-PPA-P, Lievre M, Thomas C, Hinkal G, Ansieau S, Puisieux A, et al. Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS ONE*. 2008;3:2888. <https://doi.org/10.1371/journal.pone.0002888>.
- Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*. 2008;133:704–15.
- Morel A-P, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, et al. A stemness-related ZEB1–MSRB3 axis governs cellular pliancy and breast cancer genome stability. *Nat Med*. 2017;23:568–78.
- Nieto MA, Huang RYYJ, Jackson RAA, Thiery JPP. EMT: 2016. *Cell*. 2016;166:21–45.
- Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, et al. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. *NAR Cancer*. 2021. <https://doi.org/10.1093/narcan/zcab027>.

9. Grosse-Wilde A, D'Hérouël AF, McIntosh E, Ertaylan G, Skupin A, Kuestner RE, et al. Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS ONE*. 2015;10:e0126522. <https://doi.org/10.1371/journal.pone.0126522>.
10. Malagoli Tagliazucchi G, Wiecek AJ, Withnell E, Secrier M. Genomic and microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. *Nat Commun*. 2023;14:1–20.
11. Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY-J, et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med*. 2014;6:1279.
12. Torborg SR, Li Z, Chan JE, Tammela T. Cellular and molecular mechanisms of plasticity in cancer. *Trends in Cancer*. 2022;8:735–46.
13. Shah S, Philipp LM, Giaimo S, Sebens S, Traulsen A, Ratz M. Understanding and leveraging phenotypic plasticity during metastasis formation. *Npj Syst Biol Appl*. 2023;9:1–11.
14. Marine JC, Dawson SJ, Dawson MA. Non-genetic mechanisms of therapeutic resistance in cancer. *Nat Rev Cancer*. 2020;20:743–56.
15. Monteiro L, Da Silva L, Lipinski B, Fauvet F, Vigneron A, Puisieux A, et al. Assessing cell activities rather than identities to interpret intra-tumor phenotypic diversity and its dynamics. *iScience*. 2020;23:101061.
16. Morel AP, Hinkal GW, Thomas C, Fauvet F, Courtois-Cox S, Wierinckx A, et al. EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLOS Genet*. 2012;8:e1002723. <https://doi.org/10.1371/journal.pgen.1002723>.
17. Kim RS, Avivar-Valderas A, Estrada Y, Bragado P, Sosa MS, Aguirre-Ghiso JA, et al. Dormancy signatures and metastasis in estrogen receptor positive and negative breast cancer. *PLoS ONE*. 2012;7:e35569. <https://doi.org/10.1371/journal.pone.0035569>.
18. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486:346. <https://doi.org/10.1038/nature10983>.
19. Pereira B, Chin S-FF, Rueda OM, Vollen H-KMKM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun*. 2016;7:11479.
20. Tewey KM, Rowe TC, Yang L, Halligan BD, Liu LF. Adriamycin-induced DNA damage mediated by mammalian DNA topoisomerase II. *Science*. 1984;226:466–8.
21. Umkehrer C, Holstein F, Formenti L, Jude J, Froussios K, Neumann T, et al. Isolating live cell clones from barcoded populations using CRISPRa-inducible reporters. *Nat Biotechnol*. 2021;39:174–8.
22. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell*. 2021;39:1150–62.
23. Fletcher RB, Das D, Ngai J. Creating lineage trajectory maps via integration of single-cell RNA-sequencing and lineage tracing: integrating transgenic lineage tracing and single-cell RNA-sequencing is a robust approach for mapping developmental lineage trajectories and cell fate changes. *Bioessays*. 2018;40:1800056.
24. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. 2017;546:431–5.
25. Shaffer SM, Emert BL, Reyes Hueros RA, Cote C, Harmange G, Schaff DL, et al. Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell*. 2020;182:947–59.
26. Vander Velde R, Yoon N, Marusyk V, Durmaz A, Dhawan A, Miroshnychenko D, et al. Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures. *Nat Commun*. 2020;11:2393.
27. Nichol D, Robertson-Tessi M, Jeavons P, Anderson ARA. Stochasticity in the genotype-phenotype map: implications for the robustness and persistence of bet-hedging. *Genetics*. 2016;204:1523–39.
28. Capp JP, Thomas F. From developmental to atavistic bet-hedging: how cancer cells pervert the exploitation of random single-cell phenotypic fluctuations. *BioEssays*. 2022;44:2200048. <https://doi.org/10.1002/bies.202200048>.
29. Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. Molecular definition of breast tumor heterogeneity. *Cancer Cell*. 2007;11:259–73.
30. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CKK, Surendranath V, et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res*. 2003;13:2363–71. <https://doi.org/10.1101/gr.1680803>.
31. Kristiansen G, Winzer KJ, Mayordomo E, Bellach J, Schlüns K, Denkert C, et al. CD24 expression is a new prognostic marker in breast cancer. *Clin Cancer Res*. 2003;9:4906–13.
32. Brierie B, Pierce SE, Kroeger C, Stover DG, Pattabiraman DR, Thiru P, et al. Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. *Proc Natl Acad Sci U S A*. 2017;114:E2337–46. <https://doi.org/10.1073/pnas.1618298114>.
33. García-Prat L, Martínez-Vicente M, Perdiguero E, Ortet L, Rodríguez-Ubrea JR, et al. Autophagy maintains stemness by preventing senescence. *Nature*. 2016;529:37–42.
34. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2023;42:293–304.
35. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16:1289–96.
36. Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, et al. Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res*. 2010;12:R21.
37. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*. 2017;14:1083–6. <https://doi.org/10.1038/nmeth.4463>.
38. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P, et al. The molecular signatures database hallmark gene set collection. *Cell Syst*. 2015;1:417–25.
39. Gavish A, Tyler M, Greenwald AC, Hoefflin R, Simkin D, Tschernichovsky R, et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature*. 2023;618:598–606.
40. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013. <https://doi.org/10.1126/scisignal.2004088>.
41. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401–4.
42. de Bruijn I, Kundra R, Mastrogiacomo B, Tran TN, Sikina L, Mazor T, et al. Analysis and visualization of longitudinal genomic and clinical data from the AACR Project GENIE biopharma collaborative in cBioPortal. *Cancer Res*. 2023;83:3861–7.
43. Mercatelli D, Lopez-Garcia G, Giorgi FM. corto: a lightweight R package for gene network inference and master regulator analysis. *Bioinformatics*. 2020;36:3916–7. <https://doi.org/10.1093/bioinformatics/btaa223>.
44. Therneau TM, Grambsch PM. Modeling survival data: extending the cox model. *Stat Med*. 2001;20:2053–4. <https://doi.org/10.1002/sim.956>.
45. Zhang C. infercnvPlus: Enhanced “infercnv” package. 2021.
46. Millard SP. EnvStats: an R package for environmental statistics. *EnvStats*. New York: Springer; 2013.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.