Single cell lineage tracing reveals subclonal dynamics of anti-EGFR therapy resistance in triple negative breast cancer.

AUTHORS

Simona Pellecchia[1,4,†], Melania Franchini[1,3,†], Gaetano Viscido[1,2], Riccardo Arnese[1,3], Gennaro Gambardella[1,2,#]

AFFILIATIONS

[1]Telethon Institute of Genetics and Medicine, Naples, Italy
[2]University of Naples Federico II, Department of Chemical, Materials and Industrial Engineering, Naples, Italy
[3]University of Naples Federico II, Department of Electrical Engineering and Information Technology, Naples, Italy
[4]Scuola Superiore Meridionale, Genomics and Experimental Medicine Program, Naples, Italy

[†]These two authors contributed equally to the work.

[#]CORRESPONDENCE: gambardella@tigem.it

ABSTRACT

Although many primary triple negative breast cancers (TNBCs) have enhanced expression of epidermal growth factor receptor (EGFR), EGFR-targeted therapies have shown only variable and unpredictable clinical responses in TNBCs. Here we integrate cellular barcoding and single-cell transcriptomics methods to comprehensively characterize the subclonal dynamics of adaptation of TNBC cells in response to afatinib, tyrosine kinase inhibitor (TKI) that irreversibly inhibits EGFR. Integrated lineage tracing analysis on these cells uncovered a pre-existing subpopulation of cells composed of 192 clones (out of the initial 2,336) with distinct biological features, such as elevated IGFBP2 (Insulin-Like Growth Factor Binding Protein 2) expression levels. We demonstrate that IGFBP2 overexpression is sufficient to make TNBC cells tolerant to afatinib treatment by activating the compensatory IGF1-R signalling pathway. Finally, by using deep learning techniques, we devise an algorithm to predict the afatinib sensitivity of TNBC cells from the transcriptional status of identified marker genes of the afatinib response. Our findings provide a new understanding of EGFR-dependent hierarchy in TNBC.

Introduction

Triple-negative breast cancer (TNBC) is a highly heterogeneous and aggressive breast cancer subtype characterized by metastatic progression, poor prognosis, and the absence of targetable biomarkers (1–3). Neoadjuvant chemotherapy is initially highly effective on TNBCs but about 30%–50% of the patients rapidly develop resistance associated with higher mortality (1, 2, 4). Despite epidermal growth factor receptor (EGFR)-activating mutations and amplifications being rare in TNBC (3, 5–7), the majority of primary TNBCs shows enhanced expression of EGFR because of an increase in gene copy number (7–12) (Supplementary Figure 01A-F), thus representing a valuable vulnerability for TNBC patients (13). However, unlike other tumour types where inhibition of wild-type EGFR by monoclonal antibodies or tyrosine kinase inhibitors (TKIs) is beneficial (14–17), EGFR-targeted therapies in TNBCs have shown variable and unpredictable clinical responses (1.7% to 38.7%) (18–22). Indeed, we and others (12, 20) found no significant correlation between EGFR status (i.e. copy number, mRNA, protein, or phospho-protein levels) and response to anti-EGFR therapies (Supplementary Figure

01G). Moreover, genomic variants that were found to be predictive of anti-EGFR resistance (23) in other tumour types are infrequent in TNBC patients (6, 24, 25) (Supplementary Figure 01H), thus suggesting that non-genomic mechanisms of resistance must be at play in TNBCs. As a result, the lack of predictive biomarkers of response to anti-EGFR therapies in TNBC has hampered the translation of EGFR inhibitors in breast cancer.

Single-cell RNA sequencing (scRNA-seq) technologies have recently emerged as powerful tools to study intra-tumour heterogeneity (26, 27) and to reveal key genes that change in response to an external stimulus such as drug treatment (12, 28). However, when monitoring the emergence of drug-resistant cell populations, it is crucial to identify and compare how surviving lineages (*i.e.*, clones) of cancer cells adapt to the treatment. Recently, bulk RNA sequencing and cellular barcoding, a technique that uses distinct DNA sequences to mark each cell, have provided the possibility of following cancer progression, metastasis dissemination, and cell differentiation at a single clone level (29). Hence, developing novel methods that couple DNA marking techniques with techniques measuring cellular states at single-cell resolution is essential to determine the different adaptive trajectories leading to drug resistance (28, 30, 31). Once these two techniques are coupled, prospective lineage tracing analyses can be used to study single clone dynamics during treatment to identify mechanisms driving drug adaptation, while retrospective analyses can be used to trace backwards in time and reconstruct pre-existing mechanisms of drug resistance by simply comparing the transcriptional states of tolerant and sensitive clones in the untreated condition (Figure 1A).

Here, we investigate the non-genomic mechanisms determining cellular response to EGFR inhibitors in TNBCs. We integrate methods for cellular barcoding and single-cell transcriptomics to enable cell lineage tracing and explore the subclonal evolution of adaptation in an established preclinical model of TNBC (32, 33) in response to incremental concentrations of afatinib, a potent TKI that irreversibly inhibits both EGFR and HER2. Retrospective lineage tracing data analysis uncovered a pre-existing subpopulation of rare afatinib-tolerant cells displaying distinct biological features, such as elevated mRNA levels of IGFBP2 (Insulin-Like Growth Factor Binding Protein 2). We demonstrate by chemical and genetic manipulations that IGFBP2 overexpression is sufficient to make TNBC cells tolerant to afatinib treatment through activation of the compensatory IGF1-R signaling pathway. Finally, by linking reconstructed mechanisms of drug resistance with deep learning techniques we devise an algorithm to computationally predict afatinib response starting from the transcriptional status of TNBC cells. Our findings provide a new understanding of EGFR-dependent hierarchy in TNBC and will be helpful in devising novel strategies for patient stratification and therapeutic intervention.

RESULTS

1. Single cell lineage tracing in EGFR-dependent TNBC cells.

Here, we apply cell lineage tracing to identify novel predictive biomarkers of resistance to EGFR inhibitors in the human TNBC cell line MDAMB468. This cell line is characterized by wild-type EGFR amplification, and it is very sensitive to anti-EGFR

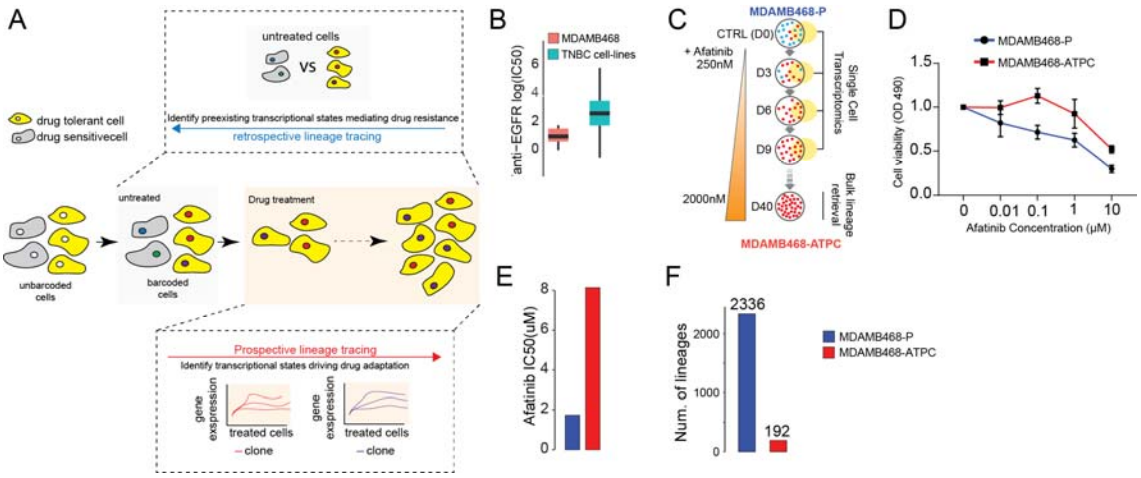


Figure 1 - Platform for the identification of biomarker genes of drug response with single cell lineage tracing. (**A**) An overview of strategies of lineage tracing. Once cells are marked (i.e., barcoded), they can be exposed at a selected drug concentration to drive the selection of resistant clones (yellow cells). Two types of analyses can be then performed. Prospective analyses where resistant clones (i.e., lineages) are followed over time during treatment to reconstruct driving mechanisms of drug adaptation. Retrospective analyses where resistant lineages are retrospectively mapped on the untreated condition to reconstruct pre-existing mechanisms of drug resistance. (**B**) Log IC50 distribution of MDAMB468 cell to anti-EGFR anticancer and of all the other TNBC cell lines. (**B**) Dose-response curve in terms of cell viability following treatment with afatinib at the indicated concentrations on MDAMB468 cells (**C**) Schematics of the time series experiment performed on barcoded MDAMB468 cell line. D0 are untreated cells (i.e., parental MDAMB468) a few hours before afatinib were applied, and D3 trough D40 is the duration of the experiment for which cells where subject to incremental doses of Afatinib ranging from 250 nM to 2000 nM. Cells at D40 were considered as afatinib tolerant persisted cells (ATPC). 10x Chromium sequencing was performed at Day 0 (D0), Day 3 (D3), Day 6 (D6) and Day 9 (D9) of 250 nM of afatinib treatment. Quantseq-Flex to retrieve in bulk afatinib tolerant lineages was performed at Day 40 at 2000 nM of afatinib treatment. (**D**) Dose-response curve in terms of cell viability following treatment with afatinib at the indicated concentrations on either parental MDAMB468 cells (i.e., MDAMB468-P) or MD-MB468 Afatinib tolerant persisted cells (i.e., MDAMB468-ATPC). (**E**) Estimated IC50 for the dose response curve in (G). (**F**) Number of unique lineages present in the parental MDAMB468 cell-line (P) and in the MDA-MB-468 Afatinib tolerant persisted cells (ATPC).

therapies (Figure 1B), including afatinib (Supplementary Figure 02). Interestingly, publicly available whole exome sequencing of MDAMB468 (34, 35) did not report any mutations in genes already linked to resistance to anti-EGFR therapies (23), including KRAS and HRAS. To enable single cell lineage tracing, we engineered MDAMB468 cells with the commercially available Cellecta's CloneTracker XP recorder technology (Methods). Specifically, the engineered MDAMB468 cells express a fluorescent reporter gene (Venus) whose 3'-UTR sequence contains a 48-nucleotide barcode cassette positioned within approximately 70-nt from the polyA sequence (Supplementary Figure 03). As a result, the barcodes are transcribed and can be captured by cDNA first strand synthesis with standard indexed oligo-dT primers used in most library preparation protocols for RNA-sequencing. By means of bulk RNA-sequencing (Methods), we estimated that the pool of barcoded MDAMB468 cells contained 2,336 distinguishable clonal populations (*i.e.*, groups of related cells descended from a single clone).

Next, we exposed barcoded MDAMB468 cells to incremental concentrations of afatinib ranging from 250 nM to 2000 nM for 40 days to select for afatinib-tolerant cells (which we termed MDAMB468-ATPC). Over the course of the experiment, we collected cells at intermediate time points (*i.e.*, every three days) for single cell transcriptomics (Figure 1C). After 40 days, cells were able to grow stably in a medium with 2000 nM afatinib (Figure 1C) and showed a 4-fold increase in afatinib IC_{50} values with respect to the parental population (Figure 1D,E). Barcode lineage retrieval from RNA bulk sequencing of MDAMB468-ATPC at 40 days showed that only 192 out of the initial 2,336 clones

(*i.e*., 8%) originally present in the parental MDAMB468 cell line were selected by afatinib treatment (Figure 1F).

2. Retrospective and prospective lineage tracing to elucidate pre-existing and acquired mechanisms of drug resistance.

Next, we sought to reconstruct pre-existing cellular states mediating the response to afatinib. To this end, we performed single-cell transcriptomics of 4,101 cells collected at day 0, day 3, day 6 and day 9 following afatinib treatment (Figure 2A). By analysing the transcriptional barcodes in single cells, we were able to capture the transcriptional state of 448 distinct clones at day 0, while 186, 206 and 228 distinct clones were captured at day 3, day 6 and day 9, respectively.

We then retrospectively mapped the barcodes of 192 afatinib-tolerant clones present at day 40 onto the single-cell data of untreated MDAMB468 cells at day 0, as shown in Figure 2B. We then asked whether we could identify genes that were differentially expressed in untreated cells at day 0 by comparing cells belonging to afatanib-tolerant lineages with those belonging to the afatinib-sensitive lineages, *i.e.* those that were depleted after 40 days of continuous afatinib exposure. These genes, if present, should highlight pre-existing mechanisms of resistance to afatinib treatment. We found that 374 genes were differentially expressed between these two cell populations (Supplementary Table 01), with 221 genes up-regulated in the afatinib-tolerant lineages and 153 genes up-regulated in the afatinib-sensitive lineages (Figure 2C). Gene Ontology Enrichment Analysis (GOEA) of the 221 over-expressed genes in the afatinib-tolerant cell population highlighted several "canonical" biological processes already associated with resistance to EGFR inhibitors in other cancer types, such as oxidative phosphorylation and fatty acid metabolism (28, 36–38) (Supplementary Table 02). On the other hand, the 153 genes over-expressed in the afatinib-sensitive cell population included EGFR and several genes related to the estrogen signalling pathway (Supplementary Table 02). The most significantly upregulated gene in the untreated afatinib-tolerant lineages was the Insulin-Like Growth Factor Binding Protein 2 (IGFBP2), as shown in Figure 2C. IGFBP2 is a member of a family of six proteins specifically binding insulin-like growth receptors I and II (IGF1-R and IGF2-R) and it has recently been shown to be involved in the progression and metastasis of several tumour types (39, 40). Interestingly, the overexpression of IGFBP2 in afatinib-tolerant lineages already present before drug treatment (*i.e.,* day 0) is maintained during treatment, as shown in (Figure 2D), suggesting that this gene could be required both in drug resistance initiation and maintenance. Indeed, higher expression of the IGFBP2 gene in afatinib-tolerant clones was further confirmed by qRT-PCR in MDAMB468-ATPC (Figure 2E).

Our analysis showed that only 192 out of 2,336 clones (*i.e.*, lineages) initially present in the parental MDAMB468 cell population became fully tolerant to afatinib treatment. Thus, we sought to investigate how these clones evolved over time by identifying key genes driving their expansion. To this end, we computed at each sequenced time point the percentage of cells associated with each afatinib-tolerant lineage and plotted them over time, as shown in Figure 2F (and Supplementary Table 03). Inspection of these profiles revealed that 190 of the 192 lineages were "neutral", maintaining a constant relative

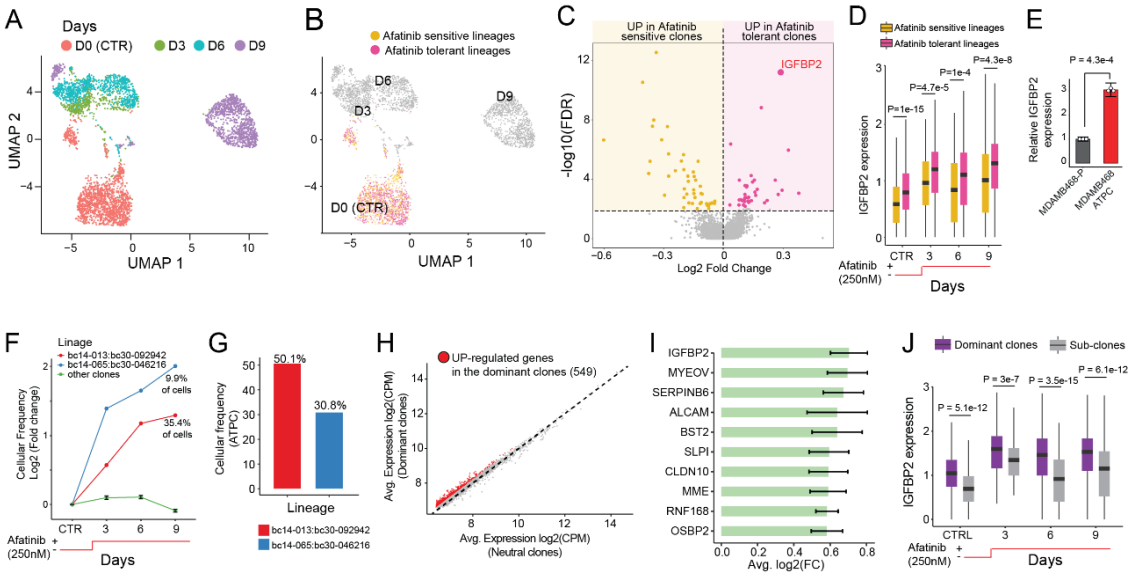


Figure 2 – Retrospective and Prospective linage tracing analyses. (**A**) UMAP representation of MDAMB468 cells colour coded respect to the days of treatment. (**B**) Retrospective projection of the afatinib tolerant persisted lineages on UMAP representation of untreated MDAMB468 cells at Day 0. (**C**) Differential expression between afatinib tolerant persisted lineages versus afatinib sensitive lineages on untreated MDAMB468 cells (i.e., cells from Day 0). (**D**) Expression distribution of IGFBP2 in afatinib tolerant persisted lineages and afatinib sensitive lineages across time. Two tailed Wilcoxon test was used for comparison. (**E**) IGFBP2 expression measured by real-time quantitative PCR in parental MDA-MB-468 cell-line (P) versus MDAMB468 afatinib tolerant persisted cells (ATPC). Two tailed t-test was used for comparisons. (**F**) Cellular frequency of afatinib tolerant clones sequenced time points (i.e., CTRL (Day 0), Day 3, Day 6, Day 9). (**G**) Cellular Frequency of clone bc14-013:bc30-092942 and bc14-013:bc30-092942 in the Drug tolerant MDAMB468-ATPC cell line. (**H**) UP-regulated genes comparing cells belonging to the dominant clones versus cells belonging to the neutral clones. (**I**) Average log2 fold change (FC) of the top ten up-regulated genes from (C). 95% confident interval is reported. (**J**) Expression distribution of IGFBP2 in dominant and neutral clones across time. statistical differences were estimated with two-tailed Wilcoxon test.

population size (*i.e.*, the same percentage) over time, whereas just two clones expanded their relative size frequency covering about half (45%) of the total population after nine days of afatinib treatment (Figure 2F). We then confirmed that these two clones become the "dominant" ones covering 81% of the final (*i.e.*, Day 40) drug-tolerant MDAMB468-ATPC cell line (Figure 2G).

Next, to identify genes driving the cellular expansion of these two dominant clones during the first nine days of afatinib treatment, we divided cells into two groups according to the behaviour of their clone of origin, *i.e.* "dominant" or "neutral". Then, we ordered the two groups of cells along a linear pseudotime (pt) trajectory (41) from day zero to day nine and finally used time-course differential expression analysis (42) to identify genes that were up-regulated in the cells of the dominant clones (see Methods). This clone resolution analysis identified a total of 549 driver genes up-regulated (FDR < 5%) in the dominant clones (Figure 2H and Supplementary Table 04), with IGFBP2 again being the most up-regulated gene (Figure 2I,J), thus suggesting that this gene could be required both in resistance initiation and cellular expansion during afatinib treatment.

To test this hypothesis, we used the CRISPRa and CRISPi systems to generate two novel cell lines starting from the parental MDAMB468 and used them to check for changes in afatinib cytotoxicity: one cell line stably overexpressed IGFBP2 (*i.e.*, MDAMB468-VPH-IGFBP2) while the other cell line stably downregulated IGFBP2 expression (*i.e.*, MDAMB468-KRAB-IGFBP2) (Supplementary Figure 04A,B and Methods). As shown in Figure 3A, IGFBP2 knockdown significantly increased afatinib cytotoxicity while

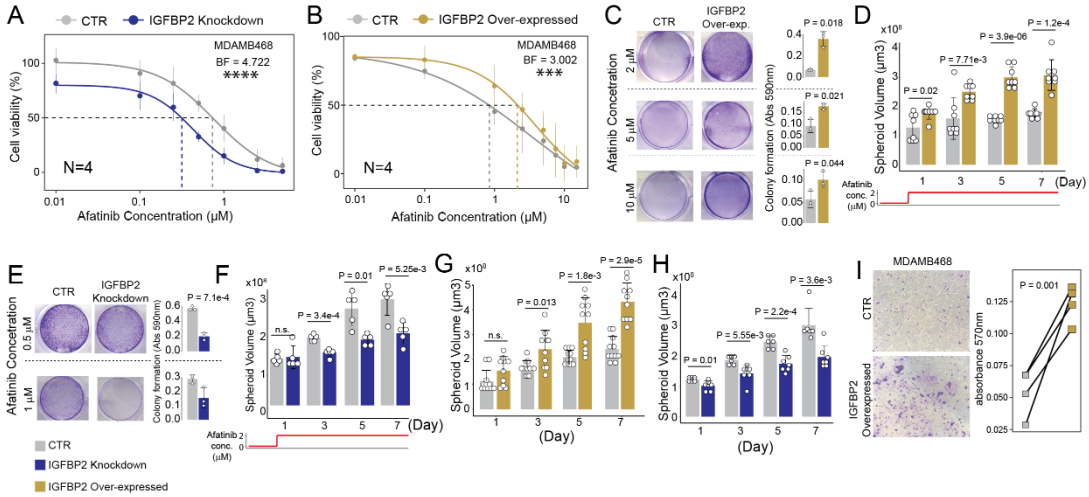


Figure 3 – Identification of the genes driving cellular expansion during afatinib adaptation (**A**) Dose-response curve in terms of cell viability following treatment with afatinib at the indicated concentrations on MDAMB468 IGFBP2 knockdown cells and relative control cells. Asterisks indicate that the shift of the two curves is statistically significant. Significance is assessed by using gaussian processes (Methods) and BF is the estimated Bayesian Factor. (**B**) Dose-response curve in terms of cell viability following treatment with afatinib at the indicated concentrations on MDAMB468 IGFBP2 overexpressing cells and relative control cells. (**C**) Colony assay (representative image) after 10 days of afatinib exposure at 2, 5 and 10µM for MDAMB468 IGFBP2 overexpressing cells and relative control cells (left). Quantification is reported on the right. Experiments were performed in triplicate. (**D**) Spheroid volume growth with 2µM of afatinib (see methods) over time of MDAMB468 IGFBP2 overexpressing cells and relative control cells. (**E**) Colony assay (representative image) after 3 days of afatinib exposure at 0.5 and 1µM for MDAMB468 IGFBP2 knockdown cells and relative control cells (left). Quantification is reported on the right. Experiments were performed in triplicate. (**F**) Spheroid volume growth with 2µM of afatinib over time of MDAMB468 IGFBP2 knockdown cells and relative control cells. (**G**) Spheroid volume growth over time of MDAMB468 IGFBP2 overexpressing cells. (**H**) Spheroid volume growth over time of MDAMB468 IGFBP2 overexpressing cells. (**I**) Transwell migration assay of MDAMB468 IGFBP2 overexpressing cells. Reported p values from panels C to I are estimated using two-tailed t-test.

IGFBP2 overexpression significantly decreased it (Figure 3B). As shown in Figure 3C, IGFBP2 overexpression allowed cell colony growth even with a concentration of afatinib ranging from 2 to 10 µM. These results were further confirmed in 3D cell culture, where IGFBP2 overexpression increased MDAMB468 spheroid growth at 2µM afatinib exposure (Figure 3D). Conversely, IGFBP2 knockdown inhibited colony formation after 3 days of exposure to 1 µM afatinib (Figure 3E) and significantly inhibited MDAMB468 spheroid growth during afatinib treatment (Figure 3F). To test whether IGFBP2 could also confer resistance to afatinib in other TNBC cell lines, we selected additional highly afatinib-responsive TNBC cell lines (HCC1806 and HDQP1) and used again the CRISPRa system to generate cells stably overexpressing IGFBP2 (Supplementary Figure 04C). As shown in Supplementary Figure 05A,B, IGFBP2 overexpression decreased afatinib cytotoxicity in both cell lines.

Finally, to dissect the biological process sustained by IGFBP2 expression that could promote and drive afatinib adaptation, we performed bulk 3' RNA-sequencing of MDAMB468-VPH-IGFBP2 cells. A total of 856 transcripts were significantly modulated by IGFBP2 overexpression (FDR<0.05), of which 373 were up-regulated and 483 down-regulated (Supplementary Table 05). Interestingly, among the up-regulated genes we found several genes related to the insulin signalling and PI3K-AKT pathways, including MET and EGFR genes (40, 43). Indeed, GOEA of the 856 differentially expressed genes showed a strong enrichment for genes related to the PI3K-AKT axis/AMPK signalling but also many genes related to cell adhesion, growth, and migration (Supplementary Table 06). Indeed, IGFBP2 knockdown impairs spheroid growth of untreated

MDAMB468 cells conversely to its overexpression (Figure 3G,H). IGFBP2 overexpression also increases cell migration in MDAMB468 (Figure 3I) and additional TN cell lines such as HCC1806 and HDQP1 (Supplementary Figure 05C,D). Finally, of the 856 genes modulated by IGFBP2 overexpression, 88 are in common with the 549 driver genes upregulated in dominant lineages. Those genes were mainly related to cell growth, adhesion, and metabolic processes (Supplementary Table 07).

Other studies have shown that resistance mechanisms in response to anti-EGFR therapies involve the compensatory activation of redundant signalling pathways sharing downstream effectors of the EGFR signalling cascade, including IGF1-R, MET and the PI3K-AKT-mTOR signalling pathways (23, 44–47). Indeed, co-treatment of parental MDAMB468 cells at different concentrations of afatinib and GSK-1904529A, a selective inhibitor of IGF1-R and Insulin Receptor (INSR), had a strong synergistic effect on these cells (Supplementary Figure 06). To explore a larger combination of the dose response surface, we performed two distinct experiments in triplicate, one in which afatinib varied from 1 to 16 μM (Supplementary Figure 06A,B) and another in which it varied between 31 nM and 1 μM (Supplementary Figure 06C,D). As shown in Supplementary Figure 06, the synergism between afatinib and GSK4529 is evident over a broad range of afatinib concentrations but it is maximal around 1 μM. These results confirm the presence of a pre-existing group of cells characterized by higher activity of the compensatory IGF1-R signalling pathway sustained by IGFBP2.

Overall, we show how our cell lineage tracing strategy enables the identification of genes driving drug resistance and how heterogeneity in gene expression in a genomically stable tumour cell line allows for the coexistence of distinct sub-populations of cells, some of which can be tolerant to targeted anticancer drugs and are thus selected during treatment. Altogether, these results support the role of IGFBP2 as a gene required for resistance initiation and cellular expansion during adaptation to afatinib treatment.

3. Reconstruction of the gene expression dynamics in response to afatinib.

Next, we sought to understand how drug-resistant sub-populations of cancer cells arising over time coordinate gene activity and transcriptional programs. To this end, we clustered the expression patterns of the 549 driver genes upregulated in drug-tolerant lineages over the course of their reconstructed pseudotime trajectory (Figure 4A). This analysis revealed four distinct transcriptional waves of adaptation to afatinib treatment (Figure 4A). Cluster one and two comprise genes whose expression increases over time with a different temporal programme in response to afatinib. Specifically, cluster one includes “early responsive” genes whose expression increases after three days of treatment while cluster two contains “late responsive” genes whose expression is upregulated after six days of treatment. Cluster three is composed of genes whose expression decreases during afatinib adaptation. Finally, genes in cluster four exhibit a “pulse response” characterised by a transient decrease in expression.

Next, we performed a time-resolved Gene Ontology Enrichment Analyses (GOEA) to elucidate the sequential activation of the transcriptional programmes driving afatinib adaptation. To this end, we first performed GOEA considering all 549 driver genes (Supplementary Table 08) and then reconstructed the specific activation time of each

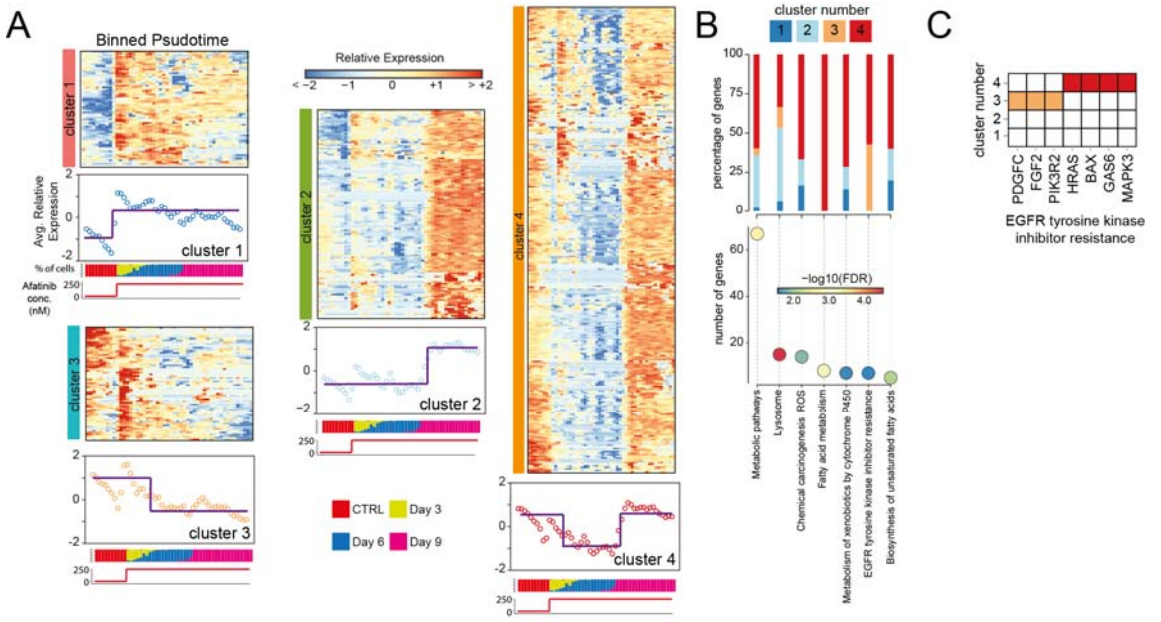


Figure 4 – Gene expression dynamics in response to Afatinib adaptation. (**A**) 549 driver genes clustered according to their expression. Cells are binned according to their reconstructed pseudotime. Below each cluster is reported the average expression profile of the genes in the cluster and the cell composition of each bin as a percentage of cells in each timepoint. (**B**) Gene ontology enrichment analysis of 549 genes (bottom) and for each enriched term the percentage of genes belonging to a specific cluster of (A). (**C**) Genes in the enriched term EGFR tyrosine kinase inhibitor resistance and the cluster of (A) in which they fall.

significant GO term by computing the proportion of genes it included from each cluster. GOEA of the 549 driver genes (Figure 4B) highlighted several biological processes that have been already associated with drug resistance in other cancer types, including lysosome biogenesis, Reactive Oxygen Species (ROS) homeostasis and fatty acid metabolism (48–54), but also specific genes linked to EGFR resistance like PDGF-C (55), FGF2 (56, 57), PIK3R2 (58), HRAS (59), MAPK3 (60, 61), GAS6 (62), and BAX (63) (Figure 4C). Interestingly, PDGF-C and FGF2 are PDGFR-α and FGFR ligands, respectively. They are well-known compensatory pathways activated to overcome EGFR inhibition (56, 59), often accompanied by gain of mesenchymal features (64, 65) that are frequently maintained by the AXL pathway (66, 67) whose activator is GAS6, another EGFR resistance gene we found to be enriched (Figure 4C). On the other hand, lysosomal biogenesis and lipogenesis involvement in drug adaptation has been extensively studied for many anticancer drugs in multitude cancer types, including both standard chemotherapeutics and targeted therapy (48–54). Enhanced lysosomal biogenesis allows the sequestration of hydrophobic weak base compounds thus lowering the cytotoxic potential of a drug by reducing its availability at the site of action (48, 49). Enhanced lipogenesis has been instead reported to contribute to drug adaptation by reducing drug uptake and by participating in antioxidant cell defence by regulating membrane fluidity (50–54).

As shown in Figure 4B, a temporal programme emerges, although most of the pathways have genes that fall within all four clusters. For example, lysosomal genes were particularly enriched in clusters one and two, where gene expression increases in response to afatinib, indicating a possible cellular reaction to drug-induced damage. On the other hand, genes related to fatty acid metabolism, ROS, and EGFR TKI resistance (Figure 4B,C) had expression profiles falling in cluster three and four where gene

expression was high from day 0 (i.e., untreated cells), suggesting that they could be innate properties of drug-tolerant cells which can be either required in later stages of drug adaptation (if they fall within cluster four) or not (if they fall within cluster 3).

Overall, these results sustain the hypothesis that during the response to afatinib, as well as the induction of xenobiotic detoxification mechanisms to reduce drug availability, tolerant clones activate compensatory pathways in addition to IGF1-R to maintain pro-survival signals thus reducing Reactive Oxygen Species (ROS) production (68), while limiting their damaging effects by producing antioxidants and boosting Epithelial to Mesenchymal Transition (EMT) (65, 69).

4. Contrastive learning to predict afatinib response in TNBC cells.

Our retrospective lineage tracing analysis on barcoded MDAMB468 cells identified 374 genes which expression status mediates the afatinib response. Hence, we asked whether these genes are good predictive biomarkers of afatinib therapy for other triple negative breast cancer cells. To this end, we developed scASTRAL (single-cell Afatinib reSponse of TRiple negAtive ceLls), a deep doublet learning approach (Methods) that leverages contrastive learning and Supported Vector Machine (SVM) to predict afatinib sensitivity from single cell data using the expression values of the 374 marker genes of the afatinib response we identified in Figure 2C.

Contrastive learning is a machine learning approach whose goal consists in building an optimized embedding space where similar sample pairs are pushed closer and dissimilar ones farther apart. As shown in Figure 5A, scASTRAL is divided into two main steps. It first uses a contrastive learning approach to build an afatinib-response embedding space where the cosine distance within MDAMB468 control cells belonging to clones of the same (intra-) afatinib-response class is being minimized while at the same time maximized among cells belonging to clones of different (inter-) afatinib-response classes (Figure 5B). In this step, before being embedded into the new space, the transcriptional status of a cell is represented by the expression values of the 374 marker genes of the afatinib response we identified. Finally, during the second step of scASTRAL, we trained an SVM classifier to distinguish afatinib-sensitive and afatinib-tolerant cells in the space learned in the first step.

Next, we first trained and cross-validated scASTRAL on 1,541 MDAMB468 control cells (Methods) and then tested its performance in predicting afatinib sensitivity on 22,724 single-cell transcriptional profiles from 16 TNBC cell lines (Figure 5C). Specifically, single cell data of 11 out of the 16 cell lines (comprising 15,022 cells) were obtained from the cell line breast cancer atlas we released recently (26), while the remaining 5 cell lines (*e.g.*, SUM149PT, SUM185PE, SUM22PE, SUM159PT and SUM1315MO2), comprising 7,702 cells, were de-novo sequenced using the DROP-seq platform (70) (Methods). Then, we used scASTRAL to project the 22,724 triple negative cells into the learned afatinib-response space and classify each cell as sensitive or resistant to afatinib. Finally, to evaluate scASTRAL performance in predicting afatinib sensitivity, we converted predictions from the single-cell level to the cell-line level by computing the fraction of predicted afatinib-sensitive cells in each cell line and correlated these values with the experimentally estimated afatinib IC50 values of each cell line. Experimentally

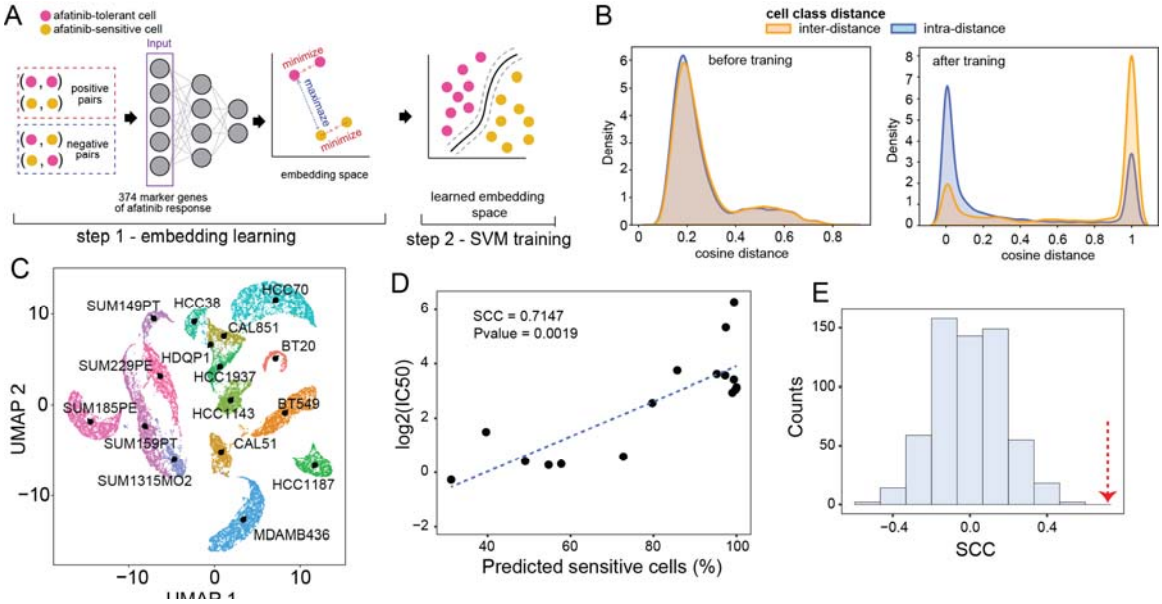


Figure 5 – Single-cell afatinib response prediction in TNBC cells. (**A**) Schematics of scASTRAL. In the first step contrastive learning is used to build and embedding space where positive examples (two cells of the same class) are separated from negative examples (two cells of different classes). In the second step an SVM classifier is trained on the learned embedded space to predict the class (i.e., afatinib-tolerant or afatinib-sensitive) of novel cells. (**B**) Inter- and intra- cosine distance between sensitive and tolerant MDAMB468 cells before and after the training. (**C**) UMAP representation of 22,724 triple negative breast cancer cells from 16 cell-lines. (**D**) Spearman Correlation Coefficient (SCC) between scASTRAL predicted sensitivity to afatinib and experimentally estimated IC50. (**E**) Histogram of SCC values computed between scASTRAL predicted sensitivity to afatinib and experimentally estimated IC50 using 374 random genes. 1,000 simulations were performed. Red arrow indicates SCC value obtained using the 374 marker genes of afatinib response we identified with our retrospective lineage tracing approach.

estimated afatinib IC50 values were retrieved from the Genomics of Drug Sensitivity in Cancer (GDSC) database (34) or measured de-novo when not available (Supplementary Table 09). As shown in Figure 5D, predicted sensitivity significantly correlated with the experimentally estimated one (P=0.0019, SCC=0.71).

Finally, to assess the predictive power of our 374 marker genes we performed Monte Carlo simulations (Methods). Specifically, in each simulation: (*i*) we randomly selected 347 genes; (*ii*) we trained scASTRAL on MDAMB468 control cells; (*iii*) we used the trained model to predict the afatinib response of the 16 TNBC cell lines described above; (iv) we evaluated the performance of scASTRAL correlating predicted sensitivities with the experimentally estimated ones. As shown in Figure 5E, we performed 1,000 simulations, and we never observed a set of 347 genes able to predict the afatinib response better than the 347 marker genes identified with our retrospective lineage tracing strategy.

Altogether, these results show how our cell lineage tracing strategy to retrospectively map drug tolerant clones onto untreated cells enables the identification of biomarker genes predictive of drug response.

Discussion

Due to their molecular heterogeneity and the lack of recurrent oncogenic alterations (5), TNBC still has few targeted therapy options other than conventional chemotherapies (3). Novel approaches are required to identify biomarker genes of drug response that can be effective in successfully stratifying TNBC patients for targeted therapies, such as EGFR targeting agents (18–22). Although clinical studies of EGFR inhibitors across various

tumor types have not shown sufficiently high response rates to warrant their use in unselected patients, durable responses have been observed at low frequencies across several epithelial tumor types (71, 72).

Here, by integrating single-cell transcriptomics, cellular barcoding, and time-resolved computational analyses, we provide the transcriptional characterization at the single-clone level of MDAMB468 cells in response to afatinib treatment, a potent TKI that irreversibly inhibits both EGFR and HER2 proteins. When coupled, we show that single-cell transcriptomics and cellular barcoding can be used to reconstruct both pre-existing and adaptive mechanisms of drug resistance of cancer cells. In our work, retrospective lineage tracing analysis uncovered a pre-existing subpopulation of MDAMB468 afatinib-tolerant cells displaying distinct biological features, such as elevated mRNA levels of the IGFBP2 gene. We provide experimental validation that IGFBP2 overexpression is sufficient to make TNBC cells tolerant to afatinib treatment through activation of the compensatory IGF1-R signalling pathway (47). At the same time, prospective lineage tracing analysis revealed the activation of several mechanisms sustaining drug adaptation, including lysosome biogenesis, reactive oxygen species homeostasis and fatty acid metabolism (48–54). Knowledge of the molecular basis by which clones of cancer cells differentially respond to therapy is essential for gaining new insights into how intra-tumour heterogeneity affects drug response and identifying novel vulnerabilities useful for developing personalized therapies.

Finally, we show our approach is also helpful in automatically identifying gene signatures highly predictive of drug response. Indeed, by using deep learning techniques, we devised an algorithm to computationally predict afatinib sensitivity from the expression levels of the 374 genes mediating afatinib response we identified using our retrospective lineage tracing strategy. When examined in public and de-novo single-cell datasets of triple-negative breast cancer cells, this signature predicted afatinib response with reasonable degree of accuracy. It would be interesting to determine whether transcriptional programs of afatinib resistance are also conserved in other tumor types and shared by other TKI inhibitors.

Methods

Cell culture.

HEK293T, MDAMB468, HCC1806 and HDQP1 were obtained from ATCC biobank. HEK293T and HDQP1 cells were cultured with DMEM supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine, 1% penicillin/streptomycin (Euroclone, Milan, Italy), whereas MDAMB468 and HCC1806 cells were cultured with RPMI supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine, 1% penicillin/streptomycin (Euroclone). Cells were kept at 37 °C under 5% CO2 atmosphere. SUM series cell lines were obtained from BioIVT biobank. SUM149PT, SUM185PE, SUM225CWN, SUM229PE and SUM159PT were cultured with Ham's F12 medium (Gibco) supplemented with 5% heat-incativated FBS (Gibco), 10mM HEPES (Sigma-Aldrich), 1 μg/ml hydrocortisone (Sigma-Aldrich), 5 μg/ml insulin (Sigma-Aldrich) and 1% L-glutamine (Euroclone). SUM1315MO2 were cultured with the Ham's F12 medium supplemented as described above with in addition 10 ng/mL EGF. No penicillin-streptomycin was added to the SUM cell line medium, as recommended by the providers' instruction.

dCAS9 plasmids and sgRNA cloning.

pRDKCE2B-EFS-dCas9-KRAB-2A-Blast, pRDVCE2B-EFS-dCas9-VPH-2A-Blast, pRSG16N-U6-sg-UbiC-TagRFP-2A-Neo and pRSG17H-U6-sg-UbiC-TagGFP2-2A-Hygro were obtained from Cellecta

(#SVKRABC9E2B-PS, #SVVPHC9E2B-PS, #SVCRU616N-L and #SVCRU617H-L respectively). sgRNA used to overexpress or downregulate IGFBP2 gene were designed CRISPick Broad Institute tool (https://portals.broadinstitute.org/gpp/public) (73, 74) (Supplementary Table 10). IGFBP2 protospacer sequences were synthesized, hybridized, phosphorylated, and inserted into pRSU6-gRNA plasmids using BbsI sites.

Lentiviral packaging and transduction of dCAS9 plasmids.

Lentiviral packaging was performed following manufacturer instructions (Cellecta Mountain View, CA). Briefly, 24 hours before transfection, 4 x 10^6 of HEK293T cells were plated in a 100-mm tissue-culture dish. On the day of transfection, 2 µg of the plasmid of interest was combined with 10 µg of the packaging plasmid mix (Cellecta, psPAX2: pMD2.G) in DMEM without serum or antibiotics in the presence of Plus Reagent™ and Lipofectamin™ (Life Technologies). Cells were incubated with the DNA/Plus Reagent™/Lipofectamin™ mix for 24 hours. Viral supernatant was collected 48 hours after transfection, filtered through a 0.45 µm PES filter and concentrated using LentiFuge™ Viral Concentration Reagent (Cellecta) following manufacturer's protocol. The concentrated lentiviral particles were re-suspended in PBS and stored at −80 °C. 5 x 10^5 of MDAMB468 and HCC1806 cells were seeded in 6-well plate and transduced with lentivirus to stably express dCas9-KRAB or dCas9-VHP in the presence of 8 µg/mL of LentiTrans™ Polybrene Transduction Reagent (Cellecta, #LTDR1) at MOI of 0.5. 72 hours after transduction, 2.5 µg/mL of Blasticidin (Thermo Fisher Scientific, Waltham, MA, USA) was added to MDAMB468 cells, while 1 µg/mL of Blasticidin was used for HCC1806 cell line. Once dCas9-KRAB and dCas9-VHP stable cell lines were obtained, those cells were transduced as previously, with pRSU6-gRNA lentivirus to modulate IGFBP2 expression with an MOI of 0.5. After 5 days of selection, sgRNA RFP positive or GFP positive cells were further sorted by using BD FACS ARIA III.

HDQP1 cell transfection.

The pCMV6-AC-IGFBP2-GFP expression vector encoding human IGFBP2 (NM_000597) fused to the GFP in the C-terminal region was purchased from Origene Technologies (#RG202573). This plasmid was used to obtain HDPQ1 cells stably overexpressing IGFBP2 gene. Lipofectamine 2000 (Life Technologies, Grand Island, NY, USA) reagent was used to transfect HDQP1 cells according to the manufacturer's instructions. The stable-transfected HDQ-P1 cells were selected in a medium containing 500 µg/mL of G418 (Life Technologies), respectively. After 7 days of selection, cells were sorted with BD FACS ARIA III in order to select only the transfected population.

CloneTracker XP barcode transduction.

MDAMB468 barcoded cell population was generated by using CellTracker XP 10M Barcode-3' Libraries (Cellecta, #BCXP10M3VP-V). Overall, 50,000 cells were transduced with the library with LentiTrans™ Polybrene Transduction Reagent at low MOI (0.05) to yield only a 5% of infection and obtained a population with 2,500 unique different barcodes. 72 hours post-infection, 1 µg/mL of Puromycin (Thermo Fisher Scientific) was added to the medium to select only transduced MDAMB468 cells. After 5 days, selected cells were checked by measuring the number of cells expressing the Venus using the BD Accuri™ C6 flowcytometer (BD Biosciences, Franklin Lakes, NJ, USA). The cells were 100% positive and thus expanded in cultured medium with 1 µg/mL of Puromycin.

Time series experiment.

5 x 10^5 barcoded cells were plated in triplicate in 6-well plate and allowed to adhere for 24 hours. Control cells were also plated in triplicate and viability checked over the course of the experiment. Cells were then treated with incremental concentration of Afatinib (Selleckchem cat. Num. S1011) starting from 250 nM with each treatment lasting 72 hours. Afatinib concentration was doubled every nine days till 2µM. Every 3 days cell viability was measured for both treated and not treated cells and 1 x 10^4 of treated cells were collected and cryopreserved from each treated replicate as follow. Cells were centrifugate at 1100 rpm for 3 min and then resuspended with the resuspension buffer (RB) (100% of FBS) to obtain a final concentration of 1000 cells/µl. Then, 10 µl of this cell suspension was added to 90 µl of RB to obtain a final concentration of 100 cells/µl (total: 1 x 10^4 cells). Finally, 100 µl of cold freezing buffer (FB) (100% of FBS and 5% DMSO) was added before to place cells at -80°C.

Single cell RNA-sequencing of MDAMB468 cells with 10x chromium platform.

The cryopreserved samples were thawed in a 37 °C water bath and washed multiple times with MACS® Separation Buffer (Miltenyi biotech). Cell viability of the single cell suspension was measured with Triplan Blue using a LUNA-II™ Automated Cell Counter (Logos Biosystems). Cells were then filtered through a 40-µm filter, centrifuged at 300×g for 5 min and resuspended to obtain a final concentration of 1 x 10^3 cells/ µl. The cell suspensions were then processed to generate single-cell libraries using a Chromium Single Cell Gene Expression 3′ Library and Gel Bead kit v3.1 following the manufacturer's instructions (10x Genomics). Briefly, the cells were suspended in reverse transcription reagents and injected into microfluidic chips of the 10x Chromium instrument, along with gel beads, and segregated in a nanoliter-scale Gel Beads-in-Emulsion (GEMs). Reverse transcription was carried out on the GEMs in a MiniAmp Thermal Cycler (Thermo Fisher Scientific) with the following protocol: 53°C for 45 min, 85 °C for 5 min, and hold at 4 °C. After, reverse-transcribed samples were purified with the Recovery agent and isolated with the Dynabeads Cleanup Mix. The cDNA was amplified into the Thermo cycler using the following protocol: 98°C for 3 min, 12 cycles of (98°C for 15 sec, 63°C for 20 sec, 72°C for 1 min), 72°C for 1 min, and hold at 4°C. The quality of the amplified cDNA was quantified with the BioAnalyzer High Sensitivity DNA kit D5000 (Agilent Technologies) and its concentration was measured using the Qubit Fluorometer. Finally, amplified cDNAs were fragmented, end-repaired and A-tailed with SPRIselect Reagent Kit (Beckman Coulter). The post-ligation products were amplified using the following protocol: 98°C for 45 sec, 10 cycles of (98°C for 20 sec, 54°C for 30 sec, 72°C for 20 sec), 72°C for 1 min and holt at 4°C. The sequencing-ready libraries were purified with SPRIselect and the quality were quantified with the BioAnalyzer High Sensitivity DNA kit D1000 and the concentrations were measured using the Qubit Fluorometer. NovaSeq 6000 SP 100 cycles flow cell was used to sequence the libraries.

Specific CloneTracker XP barcode single cell library.

To optimize the capture of the CloneTracker XP barcode at single cell level, a parallel single cell library preparation was performed from the amplified cDNAs by separately amplifying the CloneTracker XP barcode amplicon with a custom P7 primer targeting it (Supplementary table 10). CloneTracker XP barcode was amplified according to the following protocol: 98°C for 45 sec, 14 cycles of (98°C for 20 sec, 54°C for 30 sec, 72°C for 20 sec), 72°C for 1 min. PCR products were then SPRI-purified and the quality were quantified with the BioAnalyzer High Sensitivity DNA kit D1000 and the concentrations were measured using the Qubit Fluorometer. Single cell CloneTracker XP barcode libraries were sequenced as spike-ins alongside the parent RNA-seq libraries. This strategy to use a custom P7 primer with unique i7 index had the advantage to allow to separately analyse the barcoded scRNA-derived cDNA in the NGS data.

CloneTracker XP barcode retrieval from scRNA-Seq reads.

As shown in Supplementary Figure 03, the 48-nucleotide expressed barcode cassette of the cellecta CloneTracker XP barcode library has a composite structure built by juxta positioning 100 different 14-nt sequences (a.k.a. bc14) and 100,000 30-nt sequences (a.k.a. bc30) separated by a 4-nt (TGGT) anchor. Cellecta also provides the whitelist containing the 100 possible bc14, and 100,000 bc30 can appear at both sides of the anchor. Lineage barcodes were identified from the FASTQ file of read two, while cell barcode from read one of both the specific CloneTracker XP barcode single cell library and standard single cell library we produced for each sequenced time point. Specifically, for each fragment present in the read two, the occurrence of the TGGT anchor was first searched allowing one possible mismatch. Then, if some correspondence was found, the 14-nt bases before (*i.e.*, putative bc14) and the 30-nt bases after (*i.e.*, putative bc30) the anchor sequence were extracted. Next, if putative bc14 and bc30 were in the whitelist provided by cellecta, the lineage barcode represented by the string "bc14:bc30" was assigned to the corresponding cell retrieved from read one. On the other hand, if one of the putative lineage barcodes was not found in the cellecta's whitelist, we corrected it with the barcode in the whitelist had a Hamming distance equal to one, but only if there were no other barcodes in the whitelist with a Hamming distance of one. Finally, if both the putative bc14 and bc30 were not found in the cellecta's whitelist, the bc14 barcode sequence is corrected first as above and only if bc14 correction is successful we corrected the corresponding bc30. This strategy allowed us to reduce the high computing demand required by correcting lineage barcodes for possible sequencing errors but without losing sensitivity. At end of this iterative process, each sequenced cell is associated with the most abundant pair of lineage barcodes bc14:bc30 we

have found. Only reads were both bc14 and bc30 after correction matched with the ones provided by cellecta's whitelist were used. This was implemented in an R script (https://github.com/gambalab/CloneTrackerXP_exstraction) that make use of the Biostrings R package to perform string matching and a custom C++ function to compute the Hamming distances between two DNA sequences of the same length.

MDAMB468 scRNA-seq reads alignment and expression quantification.
Data processing was performed using the Cell Ranger tool version 6.1.2 following the standard protocol. Briefly, after demultiplexed raw reads were aligned against Hg38 genome. The GENECODE annotation v32 of Hg38 was used, keeping only the genes with biotypes protein_coding, lincRNA and antisense. Putative cell doublet and cells expressing an high fraction of mitochondrial genes were removed with filterCell function of gficf R package (27, 75, 76). Finally, only highly covered cells with at least 5000 UMI were retained and used for downstream analyses. Cells of day 0 were sequenced with two distinct captures in two distinct flow cells. The possible batch effect was corrected with CombatSeq R package (77). Finally, we excluded genes expressed in less than 50 cell and in less the 5% of sequenced cells only if they had and average expression in the cells where they were expressed less than 1.12 UMI, similar to what done in (78).

Bulk estimation of the number of lineages in MDAMB468 cells.
Bulk estimation of the number of CloneTracker XP barcode presents in parental and ATPC MDAMB468 barcoded cells was performed with QuantSeq-Flex (Lexogen) technology as follow. First, total RNA was isolated from three independent biological replicates by using Qiagen RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. Then, library preparation was performed according to the manufacter's protocol (Lexogene). The target-specific reverse transcription primers to specifically capture the cellecta barcode were designed according to the guidelines outlined in the QuantSeq-Flex manual. All primers were composed of a partial Illumina P7 adapter extension followed by the target-specific sequence. A pool of 4 primers were used at 50 nM of concentration. Primer lengths were in the range of 44–50 nucleotides, as requested by manufacter's protocol. The target-specific first strand synthesis primers can be found in Supplementary Table 10.

Cell viability assay.
Cells were seeded in either 96-well plate (4 x 10^4 cell/well) or 384-well plate (1 x 10^3 cell/well). After overnight incubation at 37°, cells were treated with either drugs, at the selected dilutions, or with DMSO as negative control (in technical triplicate) and incubated at 37° for 72hr. Then, cell viability was evaluated by measuring with the GloMax® Discover instrument (Promega) either luminescence or absorbance (490 nm) respectively with the CellTiter-Glo® luminescent cell viability assay (Promega) or the CellTiter (Promega), as manufacturers' protocol. Background luminescence or absorbance value was measured in wells without cells and with only culture medium. Background values were subtracted from experimental values. All drugs used in this study were purchased from Selleckchem.

Drug combination assay.
4 x 10^4 of MDAMB468 cells were seeded in 96-well plate and incubated overnight at 37°. Afatinib and GSK-1904529A drugs were prepared in different dilutions and then combined in all possible drug pairs to generate a 5 x 5 or 5x6 drug combination matrix. Then, cells were exposed to either single agent drugs or to the drug pairs of the drug combination matrix, while negative control was treated with DMSO (each treatment was performed in triplicate). Following 72hr of incubation at 37°, cell viability was measured with the CellTiter (Promega), and the absorbance was read at 490 nm with the plate reader GloMax® Discover instrument. The Afatinib-GSK-1904529A drug interactions and expected drug responses were calculated with the Combenefit tool (79) using the Loewe additivity model.

Generation of spheroids and image acquisition.
Cells grown as a monolayer were detached to generate a single-cell suspension that was then diluted to 2.5 × 10^4 (for 5000 cells per spheroid) cells per millilitre of ice-cold medium. The Cell Basement Membrane

(ATCC, #ATCC-ACS-3035) was thawed on ice overnight and added at a final concentration of 2.5% with ice-cold pipette tips to the cell suspension. A volume of 200 µl of this cell suspension was added to each well of an Ultra-Low Attachment (ULA) 96-well plate with a round or conical bottom (PerkinElmer, #SPA6055330). The spheroid formation was initiated by centrifugation of the plates at 300g for 10 min. The plates were incubated under standard cell culture conditions at 37° C, 7% CO2 in humidified incubators. When we studied afatinib response the drug was added as follow. After 24h hours, when spheroids were formed, 100 µl of medium was removed and the treatment was performed adding 100 µl of medium with 2 µM of Afatinib into each well. Treatments were renewed after 3 days in a new fresh complete medium. Pictures were taken before adding treatments and then at day 3, 5 and 7 after treatment. Images were acquired with the High Content Analysis System *Operetta CLS* (PerkinElmer) with temperature (37°C) and CO2 (5%) controlled. They were acquired in several planes and the area analyses were performed on maximum projection by Harmony software (PerkinElmer). Volumetric analyses were performed by stack processing with 3d analysis.

RNA extraction, reverse transcription, and quantitative RT-PCR (qPCR).

Total RNA from cell lines were extracted using Qiagen RNeasy Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. A total of 1 µg of total RNA from each sample was used to obtain double strand cDNA with the QuantiTect Reverse Transcription Kit (Qiagen). Quantitative Real-Time PCR (qRT-PCR) was performed and for each PCR reaction, 10 µL of 2× Sybr Green (LightCycler 96, Roche Molecular Systems, Inc.), 200 nM of each primer, and 20 ng of the cDNA, previously generated, were used. Relative gene expression was determined using comparative C(T) method, as described elsewhere. RP18S was used as a housekeeping gene.

Clonogenic survival assay.

5 x 10^4 cells were plated into a 12-well plate and allowed to adhere for 24 hours. Then the cells were treated with different concentrations of Afatinib and incubated for 10 (Figure 3C) or 3 days (Figure 3E). Fresh media was added on the fifth day. On the tenth day, the media was removed from the dishes and washed once with ice-cold PBS. The colonies were fixed and stained with a solution containing crystal violet for 45 minutes on a rocking platform. The dishes were rinsed three times with PBS and air-dried. Then, the stained crystal violet was resolved with PBS-0.1% SDS and absorbance at 590 nm was determined. The data were performed in triplicate and showed as mean ± SD.

QuantSeq 3′ mRNA-Sequencing and bioinformatics analysis.

Total RNA was purified from three biological replicates by Qiagen RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. Total RNA was quantified using the Qubit 2.0 fluorimetric Assay (Thermo Fisher Scientific). Libraries were prepared from 100 ng of total RNA using the QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, Vienna, Austria). Quality of libraries was assessed by using screen tape High-sensitivity DNA D1000 (Agilent Technologies). Libraries were sequenced on a NovaSeq 6000 sequencing system using an S1, 100 cycles flow cell (Illumina Inc.). Bioinformatics analysis of QuantSeq 3′ mRNA-Sequencing data was performed as follow. Briefly, Illumina novaSeq base call (BCL) files were transformed into fastq files through bcl2fastq (version v2.20.0.422, Illumina Inc.), and sequence reads were trimmed using bbduk software (bbmap suite 37.31, Joint Genome Institute, Walnut Creek, CA, USA). Alignment was performed on hg38 reference assembly (Ensembl Assembly 93) with star 2.6.0a (GPL v3, open source) and gene expression levels were determined with htseq-count 0.9.1. Gene expression normalization and differentially expressed genes were identified by using edger package (80) in the R statistical environment. For differential expression analysis only genes with an average CPM more than 5 in at least one of the two conditions were considered.

Differential expression analysis from single cell data.

Differentially expressed genes in Figure 2C between tolerant and sensitive afatinib cells at time 0 (*i.e.*, not treated cells) were identified by using the FindMarkers function implemented in Seurat R package v3 with logfc.threshold and min.pct parameters set to 0.

Pseudotime analysis of tolerant clones and time-course differential expression analysis between cells of the dominant and neutral clones.

The 2,836 single cell expression profiles associated to afatinib-tolerant clones were divided cells into two groups according to the behaviour of their clone of origin, *i.e.*, "dominant" or "neutral". Expression matrix of dominant clones comprised 1,293 cells while the matrix of neutral clones comprised 1,543 cells. Then *psupertime* function of psupertime R package (41) was used to infer the pseudotime order of cells in each independent matrix. Psupertime R function was run with default parameters and using as cell labels their sequencing day (i.e., 0, 3, 6, 9 days). Next, for both matrices the expression profile of each gene was (*i*) ordered according to the reconstructed pseudotime; (*ii*) smoothed using the moving median method and (*iii*) finally divided in 50 expression bins according to the inferred pseudotime. To smooth gene expression with the moving median method, we used *rollmedian* function of the zoo R package with a rolling window equal to 51. After these three steps the expression profiles of each gene in both matrices is represented by 50 pseudotime points. Finally, to identify upregulated genes in the cells associated with dominant clones across the first 9 days of treatment we used the splineTimeR package (42). With this tool we compared the reconstructed time-course profile of a gene across cells of the dominant clones with its reconstructed time-course profile across cells of the neutral clones. Up-regulated genes were defined as the genes with and FDR<5% and a positive average fold change across the 50 binned pseudotime points.

Estimation of EGFR copy number alterations, mutations, and expression in TNBC.

Two independent cohort of TNBC patients were used to estimate gene copy number alterations, mutations and expression reported in Supplementary Figure 01. The first cohort composed by 192 TNBC patients was obtained from the Genomic Data Common (GDC) portal (81). GDC raw bulk expression, mutational data and copy number variation data along with clinical information were retrieved using TCGAbiolinks R package v.2.25.3 (82). The second cohort composed by 465 TNBC patients from Asiatic patients (5) for which gene expression, gene copy number alteration and mutations were available was gathered from the NODE database after author's authorization. The raw expression counts downloaded from GDC were first normalized with edgeR package (80) and transformed in log2(CPM□+□1), while those from the Asiatic cohort were already normalized as FPKM values, and thus we only transformed it into log2(FPKM + 1). A sample in the GDC cohort, was defined to have EGFR copy gain if the number of copies was 3 or 4, while EGFR amplification was defined when this number was greater than 5. For the Asiatic cohort, EGFR gain and amplification were defined based on GISTIC scores, respectively, +1 and +2. For both cohorts, mutations were considered only if they were in the coding region (i.e., missense, non-sense, in-frame and out-frame deletions). Next the deleterious tag was assigned when at least two functional annotations among VEP, PolyPhen and SIFT reported a negative impact of the mutation on protein function. Finally, EGFR activating mutations tag was conferred based on indication in (83).

Cell migration assay

Transwell chambers (8-µm pores) (Euroclone) were used to perform transwell migration assays. Briefly, breast cancer cells were plated in the upper side of the transwell chamber in a serum-deprived medium. Next, as chemoattractant, 300 µl of complete medium was added in the inferior chamber. After 24 h, cells migrated to the lower side of the chambers were stained with crystal violet solution (crystal violet 0.05%, methanol 20%). Then, crystal violet in the chamber was de-stained with PBS-0.1% SDS solution and was read at 570 nm.

scRNA library preparation of Drop-SEQ and bioinformatics analysis.

Single cell transcriptomic of the SUM149PT, SUM185PE, SUM22PE, SUM159PT and SUM1315MO2 cell lines was performed with DROP-seq technology (70) and library preparation as described in Gambardella et al. (26). scRNA libraries were sequenced with NovaSeq 6000 machine using an SP 100 cycles flow cell. Raw reads pre-processing was performed as described in (76). Only high depth cells with at least 5,000 UMI were retained and used to test scATRAL tool. Alignment pipeline can be found at https://github.com/gambalab/dropseq.

Gaussian processes to estimate significance in dose-response curves.

To evaluate the significance of dose-response curves trends difference we used gaussian processes. Gaussian processes computation was performed as in (84). Briefly, we first reconstruct for each condition the dose response trend by interpolating the response values at each tested concentration and then we

identify differences in the two trends under consideration in terms of log-likelihood, by computing the related Bayesian factor.

Estimation of the cellular frequency over time of the 192 afatinib-tolerant clones.
To compute cellular frequencies of an afatinib-tolerant lineage over time, we estimated the percentage of cells associated with it in each sequenced time point. For each time-point, this percentage was simply estimated as the number of cells associated to a specific lineage divided the total number of sequenced cells.

scASTRAL: architecture
Contrastive learning is a machine learning approach that aims to create an embedding space where positive examples (i.e., two cells with the same label) are separated from negative examples (i.e., two cells with different labels). In our case, a contrastive autoencoder was used as a model of scASTRAL. The encoder consisted of an input layer composed of 374 neurons, i.e. the number of biomarker genes of afatinib response identified with their retrospective lineage tracing strategy. The input layer was followed by two hidden layers comprising 64 and 32 neurons, respectively. A ReLU (Rectified Linear Unit) activation function was used. The decoder was symmetrical to the encoder.

scASTRAL: training
We used the 1,541 MDAMB468 control cells labelled as afatinib-tolerant or afatinib-sensitive as a training set. Before to start the training, the matrix of normalized CPM counts was cut to the 374 biomarker genes of afatinib response we identified in Figure 3C and the expression of these genes rescaled using tf-idf transformation (75). This dataset was then randomly split into 5 batches of equal size by ensuring that the number of afatinib-tolerant and afatinib-sensitive cells were similar across the different batches. Next, a contrastive model was trained on each batch follow. First a positive and a negative example is randomly built for each cell of the batch. This means that for each batch the number of examples on which training is performed was always equal to the number of cells in the batch multiplied by two. Second, the contrastive model was trained using a loss function composed by the weighted sum of three terms: the cosine embedding loss, the reconstruction error, and the latent space regularization factor. Specifically, the cosine embedding loss maximize the cosine distance between pairs of cells labelled with -1 (i.e., negative example) while maximizing the cosine similarity between pairs of cells labelled with +1. (i.e., positive example). At the same time, the reconstruction error assures that the decoder was capable to reconstruct original data from the latent representation generated by the encoder. This allows the model to also integrate in the final embedding space the dependencies among the 374 genes we used as input. Third, an SVM classifier is trained using the reconstructed embedding space of the considered batch of cells. Specifically, the SVM classifier was trained with a cosine kernel and with a regularization hyperparameter set to 100. Classification accuracy of the trained SVM of a batch was finally evaluated using the left-out cells. We used the Adam optimizer (85) with a learning rate of 0.0001 while the training was performed using an early stopping criterion, although the maximum number of epochs was set to 250. The metrics considered to compute improvements across epochs were the validation loss and the SVM classification accuracy. If no improvements are obtained considering the last 20 epochs the training stops. The contrastive model hyperparameters were found by a grid search. Once the 5 contrastive models with associated SVM models were trained the model with the highest classification accuracy was used to predict afatinib sensitivity at single cell level as described below. scASTRAL was implemented in Phyton using PyTorch version 1.13.1. Code is available at the following address https://github.com/gambalab/scASTRAL.

scASTRAL performance evaluation on single cell of TNBC cell lines.
To validate scASTRAL and test its performances in predicting afatinib sensitivity of triple negative breast cancer cells, we used 22,724 single-cell transcriptional profiles from 16 TNBC cell lines (Figure 5D). Of these, 11 were obtained from Gambardella et al. (26), while five were de-novo sequenced with the drop-seq platform. Before being fed to scASTRAL method, the raw UMI count matrix of each cell-line was first normalized using sklearn to obtain CPM and then cut on the 374 biomarker genes of afatinib response identified in Figure 3C. Next, the expression values of these genes were rescaled using tf-idf transformation (75) before applying scASTRAL. Finally, a cell was deemed to be sensitive to afatinib if SVM classification probability was greater than 0.95 while it was deemed to be tolerant if the SVM classification probability was greater than 0.90. In all other cases, a cell was labelled as unclassified. Finally, to convert

predictions from the single-cell level to the cell-line level, we computed the fraction of predicted afatinib-sensitive cells in each cell line as the number of cells predicted to be sensitive divided by the total number of cells predicted to be either sensitive or tolerant.

Code availability: scASTRAL code is available at https://github.com/gambalab/scASTRAL. DropSeq Alignment pipeline can be found at https://github.com/gambalab/dropseq. Code for CloneTracker barcode retrieval from sequencing data can be found at https://github.com/gambalab/CloneTrackerXP_exstraction.

Acknowledgments: We thank the TIGEM high content screening facility for helping in 3D-spheroid analysis. We thank the TIGEM bioinformatics core for helping in dose response curves analyses.

Founding: This work was supported by the My First AIRC grant 23162.

Author Contribution: SP performed time series experiment, 10x sequencing and all the experimental validations. MF performed computational analyses. GV setup DropSeq platform and performed sequencing. RA developed the drug prediction method. GG supervised the work, wrote the manuscript, and conceived the original idea.

Competing interests: The authors declare no competing interests.

References

1. Foulkes,W.D., Smith,I.E. and Reis-Filho,J.S. (2010) Triple-Negative Breast Cancer. *N. Engl. J. Med.*, **363**, 1938–1948.
2. Bianchini,G., Balko,J.M., Mayer,I.A., Sanders,M.E. and Gianni,L. (2016) Triple-negative breast cancer: Challenges and opportunities of a heterogeneous disease. *Nat. Rev. Clin. Oncol.*, **13**, 674–690.
3. Yin,L., Duan,J.-J., Bian,X.-W. and Yu,S. (2020) Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.*, **22**, 61.
4. Hoogstraat,M., Lips,E.H., Mayayo-Peralta,I., Mulder,L., Kristel,P., van der Heijden,I., Annunziato,S., van Seijen,M., Nederlof,P.M., Sonke,G.S., *et al.* (2022) Comprehensive characterization of pre- and post-treatment samples of breast cancer reveal potential mechanisms of chemotherapy resistance. *npj Breast Cancer*, **8**, 60.
5. Jiang,Y.Z., Ma,D., Suo,C., Shi,J., Xue,M., Hu,X., Xiao,Y., Yu,K. Da, Liu,Y.R., Yu,Y., *et al.* (2019) Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell*, **35**, 428-440.e5.
6. Koboldt,D.C., Fulton,R.S., McLellan,M.D., Schmidt,H., Kalicki-Veizer,J., McMichael,J.F., Fulton,L.L., Dooling,D.J., Ding,L., Mardis,E.R., *et al.* (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61–70.
7. Perou,C.M., Sørlie,T., Eisen,M.B., van de Rijn,M., Jeffrey,S.S., Rees,C.A., Pollack,J.R., Ross,D.T., Johnsen,H., Akslen,L.A., *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**, 747–752.
8. Ali,R. and Wendt,M.K. (2017) The paradoxical functions of EGFR during breast cancer progression. *Signal Transduct. Target. Ther.*, **2**, 1–7.
9. Bhargava,R., Gerald,W.L., Li,A.R., Pan,Q., Lal,P., Ladanyi,M. and Chen,B. (2005) EGFR gene amplification in breast cancer: Correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. *Mod. Pathol.*, **18**, 1027–1033.
10. Park,H.S., Jang,M.H., Kim,E.J., Kim,H.J., Lee,H.J., Kim,Y.J., Kim,J.H., Kang,E., Kim,S.W., Kim,I.A., *et al.* (2014) High EGFR gene copy number predicts poor outcome in triple-negative breast cancer. *Mod. Pathol.*, **27**, 1212–1222.
11. Tischkowitz,M., Brunet,J.-S., Bégin,L.R., Huntsman,D.G., Cheang,M.C.U., Akslen,L.A., Nielsen,T.O. and Foulkes,W.D. (2007) Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. *BMC Cancer*, **7**, 134.
12. Savage,P., Blanchet-Cohen,A., Revil,T., Badescu,D., Saleh,S.M.I., Wang,Y.C., Zuo,D., Liu,L., Bertos,N.R., Munoz-Ramos,V., *et al.* (2017) A Targetable EGFR-Dependent Tumor-Initiating Program in Breast Cancer. *Cell Rep.*, **21**, 1140–1149.
13. Costa,R., Shah,A.N., Santa-Maria,C.A., Cruz,M.R., Mahalingam,D., Carneiro,B.A., Chae,Y.K., Cristofanilli,M., Gradishar,W.J. and Giles,F.J. (2017) Targeting Epidermal Growth Factor Receptor in triple negative breast cancer: New discoveries and practical insights for drug development. *Cancer Treat. Rev.*, **53**, 111–119.
14. Moore,M.J., Goldstein,D., Hamm,J., Figer,A., Hecht,J.R., Gallinger,S., Au,H.J., Murawa,P., Walde,D., Wolff,R.A., *et al.* (2007) Erlotinib Plus Gemcitabine Compared With Gemcitabine Alone in Patients With Advanced Pancreatic Cancer: A Phase III Trial of the National Cancer Institute of Canada Clinical Trials Group. *J. Clin. Oncol.*, **25**, 1960–1966.
15. Vermorken,J.B., Mesia,R., Rivera,F., Remenar,E., Kawecki,A., Rottey,S., Erfan,J., Zabolotnyy,D., Kienzer,H.-R., Cupissol,D., *et al.* (2008) Platinum-Based Chemotherapy plus Cetuximab in Head and Neck Cancer. *N. Engl. J. Med.*, **359**, 1116–1127.
16. Van Cutsem,E., Peeters,M., Siena,S., Humblet,Y., Hendlisz,A., Neyns,B., Canon,J.-L., Van Laethem,J.-L., Maurel,J., Richardson,G., *et al.* (2007) Open-Label Phase III Trial of Panitumumab Plus Best Supportive Care Compared With Best Supportive Care Alone in Patients With Chemotherapy-Refractory Metastatic Colorectal Cancer. *J. Clin. Oncol.*, **25**, 1658–1664.
17. Cappuzzo,F., Finocchiaro,G., Grossi,F., Bidoli,P., Favaretto,A., Marchetti,A., Valente,M.L., Cseh,A., Clementi,L., Massey,D., *et al.* (2015) Phase II study of afatinib, an irreversible ERBB family blocker, in EGFR FISH-positive non-small-cell lung cancer. *J. Thorac. Oncol.*, **10**, 665–672.
18. Bianchini,G., De Angelis,C., Licata,L. and Gianni,L. (2022) Treatment landscape of triple-negative breast cancer — expanded options, evolving needs. *Nat. Rev. Clin. Oncol.*, **19**, 91–113.

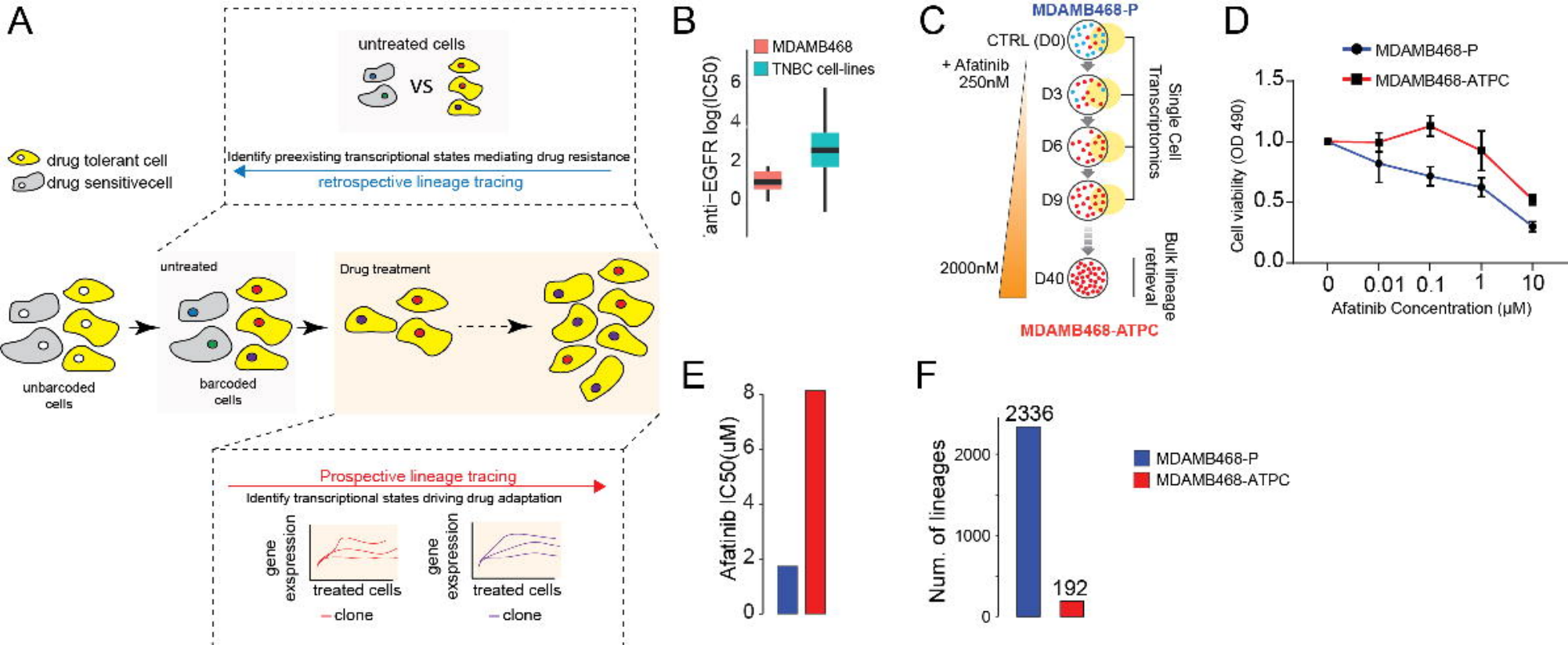
19. Baselga,J. and Arteaga,C.L. (2005) Critical Update and Emerging Trends in Epidermal Growth Factor Receptor Targeting in Cancer. *J. Clin. Oncol.*, **23**, 2445–2459.
20. Carey,L.A., Rugo,H.S., Marcom,P.K., Mayer,E.L., Esteva,F.J., Ma,C.X., Liu,M.C., Storniolo,A.M., Rimawi,M.F., Forero-Torres,A., *et al.* (2012) TBCRC 001: Randomized Phase II Study of Cetuximab in Combination With Carboplatin in Stage IV Triple-Negative Breast Cancer. *J. Clin. Oncol.*, **30**, 2615–2623.
21. von Minckwitz,G., Jonat,W., Fasching,P., du Bois,A., Kleeberg,U., Lück,H.-J., Kettner,E., Hilfrich,J., Eiermann,W., Torode,J., *et al.* (2005) A multicentre phase II study on gefitinib in taxane- and anthracycline-pretreated metastatic breast cancer. *Breast Cancer Res. Treat.*, **89**, 165–172.
22. Nakai,K., Hung,M.-C. and Yamaguchi,H. (2016) A perspective on anti-EGFR therapies targeting triple-negative breast cancer. *Am. J. Cancer Res.*, **6**, 1609–1623.
23. He,J., Huang,Z., Han,L., Gong,Y. and Xie,C. (2021) Mechanisms and management of 3rd generation EGFR TKI resistance in advanced non small cell lung cancer (Review). *Int J Oncol*, **59**, 90.
24. Baselga,J., Albanell,J., Ruiz,A., Lluch,A., Gascón,P., Guillém,V., González,S., Sauleda,S., Marimón,I., Tabernero,J.M., *et al.* (2005) Phase II and Tumor Pharmacodynamic Study of Gefitinib in Patients with Advanced Breast Cancer. *J. Clin. Oncol.*, **23**, 5323–5333.
25. Shah,S.P., Roth,A., Goya,R., Oloumi,A., Ha,G., Zhao,Y., Turashvili,G., Ding,J., Tse,K., Haffari,G., *et al.* (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*, **486**, 395–399.
26. Gambardella,G., Viscido,G., Tumaini,B., Isacchi,A., Bosotti,R. and di Bernardo,D. (2022) A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nat. Commun.*, **13**, 1714.
27. Slovin,S., Carissimo,A., Panariello,F., Grimaldi,A., Bouché,V., Gambardella,G. and Cacchiarelli,D. (2021) Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview BT - RNA Bioinformatics. In Picardi,E. (ed). Springer US, New York, NY, pp. 343–365.
28. Aissa,A.F., Islam,A.B.M.M.K., Ariss,M.M., Go,C.C., Rader,A.E., Conrardy,R.D., Gajda,A.M., Rubio-Perez,C., Valyi-Nagy,K., Pasquinelli,M., *et al.* (2021) Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nat. Commun.*, **12**, 1–25.
29. Sankaran,V.G., Weissman,J.S. and Zon,L.I. (2022) Cellular barcoding to decipher clonal dynamics in disease. *Science (80-.).*, **378**.
30. Oren,Y., Tsabar,M., Cuoco,M.S., Amir-Zilberstein,L., Cabanos,H.F., Hütter,J.-C., Hu,B., Thakore,P.I., Tabaka,M., Fulco,C.P., *et al.* (2021) Cycling cancer persister cells arise from lineages with distinct programs. *Nature*, **596**, 576–582.
31. Simeonov,K.P., Byrns,C.N., Clark,M.L., Norgard,R.J., Martin,B., Stanger,B.Z., Shendure,J., McKenna,A. and Lengner,C.J. (2021) Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell*, 10.1016/j.ccell.2021.05.005.
32. Filmus,J., Pollak,M.N., Cailleau,R. and Buick,R.N. (1985) MDA-468, a human breast cancer cell line with a high number of epidermal growth factor (EGF) receptors, has an amplified EGF receptor gene and is growth inhibited by EGF. *Biochem. Biophys. Res. Commun.*, **128**, 898–905.
33. Thomas,T., Balabhadrapathruni,S., Gardner,C.R., Hong,J., Faaland,C.A. and Thomas,T.J. (1999) Effects of epidermal growth factor on MDA-MB-468 breast cancer cells: Alterations in polyamine biosynthesis and the expression of p21/CIP1/WAF1. *J. Cell. Physiol.*, **179**, 257–266.
34. Iorio,F., Knijnenburg,T.A., Vis,D.J., Bignell,G.R., Menden,M.P., Schubert,M., Aben,N., Gon??alves,E., Barthorpe,S., Lightfoot,H., *et al.* (2016) A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, **166**, 740–754.
35. Barretina,J., Caponigro,G., Stransky,N., Venkatesan,K., Margolin,A.A., Kim,S., Wilson,C.J., Lehar,J., Kryukov,G. V, Sonkin,D., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
36. Yang,Y., Li,S., Wang,Y., Zhao,Y. and Li,Q. (2022) Protein tyrosine kinase inhibitor resistance in malignant tumors: molecular mechanisms and future perspective. *Signal Transduct. Target. Ther.*, **7**, 329.
37. Merino Salvador,M., Gómez de Cedrón,M., Moreno Rubio,J., Falagán Martínez,S., Sánchez Martínez,R., Casado,E., Ramírez de Molina,A. and Sereno,M. (2017) Lipid metabolism and lung cancer. *Crit. Rev. Oncol. Hematol.*, **112**, 31–40.
38. Kim,S., Im,J.H., Kim,W.K., Choi,Y.J., Lee,J.-Y., Kim,S.K., Kim,S.J., Kwon,S.W. and Kang,K.W. (2021) Enhanced Sensitivity of Nonsmall Cell Lung Cancer with Acquired Resistance to Epidermal

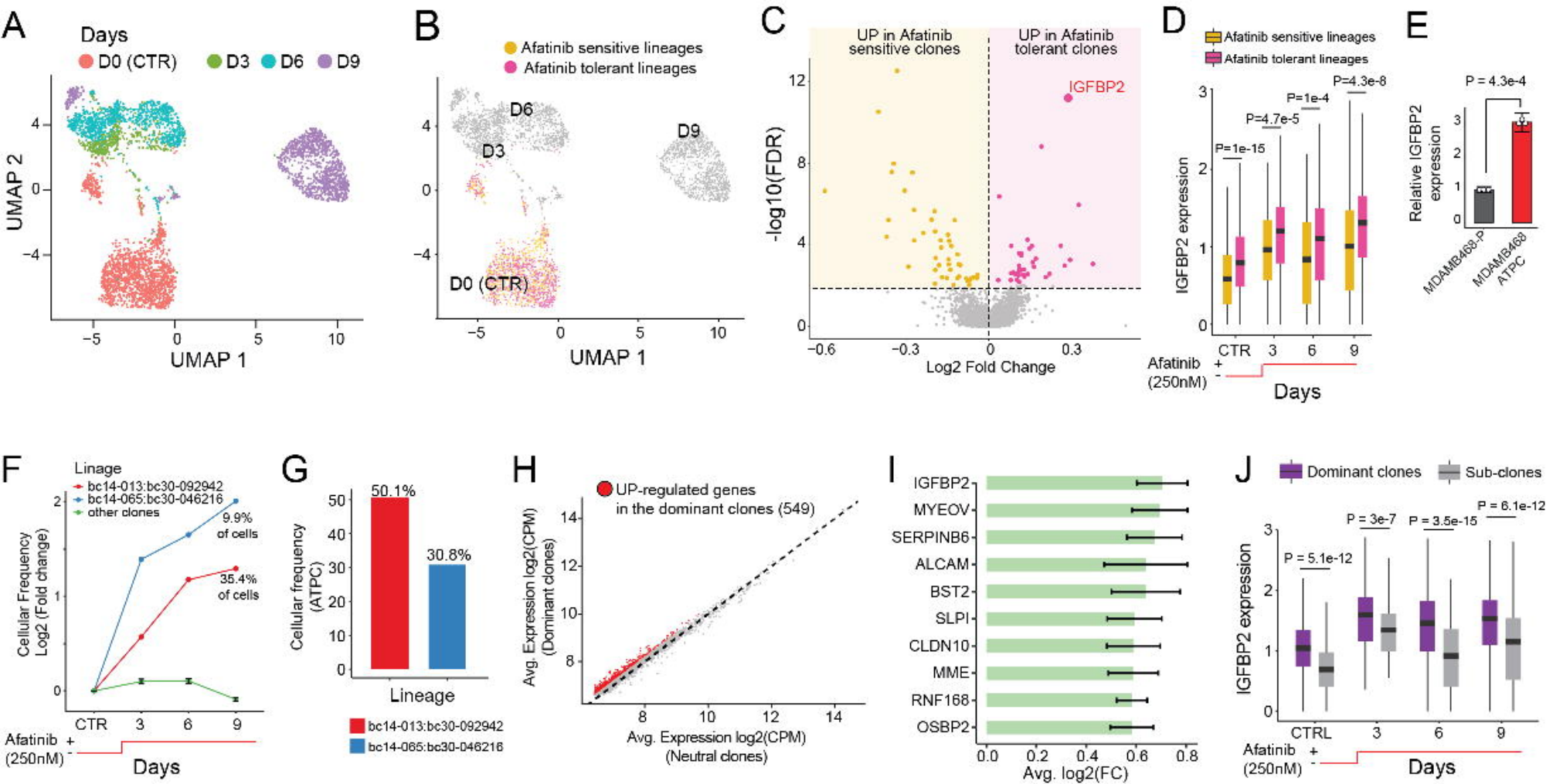
Growth Factor Receptor-Tyrosine Kinase Inhibitors to Phenformin: The Roles of a Metabolic Shift to Oxidative Phosphorylation and Redox Balance. *Oxid. Med. Cell. Longev.*, **2021**, 1–17.

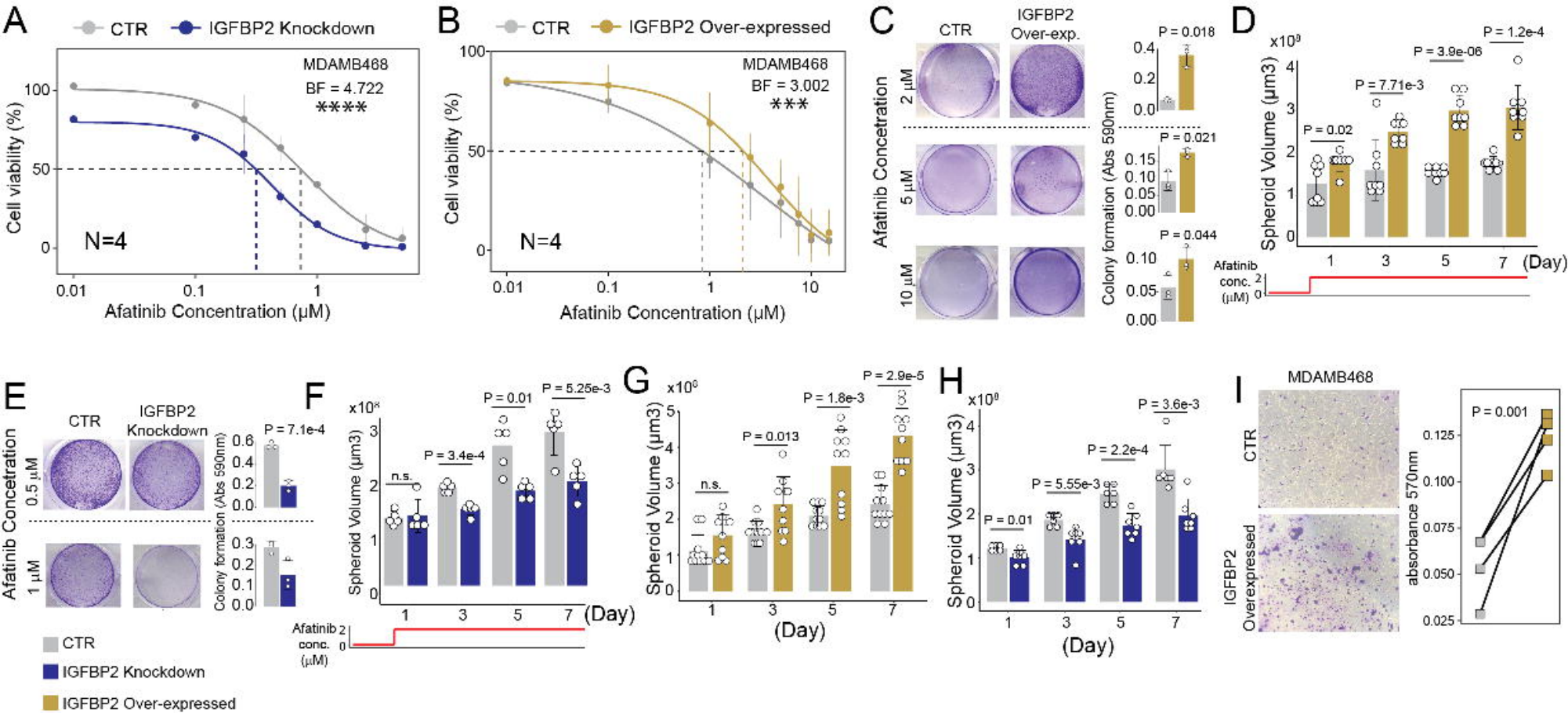
39. Wei,L.F., Weng,X.F., Huang,X.C., Peng,Y.H., Guo,H.P. and Xu,Y.W. (2021) IGFBP2 in cancer: Pathological role and clinical significance (review). *Oncol. Rep.*, **45**, 427–438.
40. Russo,V.C., Azar,W.J., Yau,S.W., Sabin,M.A. and Werther,G.A. (2015) IGFBP-2: The dark horse in metabolism and cancer. *Cytokine Growth Factor Rev.*, **26**, 329–346.
41. Macnair,W., Gupta,R. and Claassen,M. (2022) psupertime: supervised pseudotime analysis for time-series single-cell RNA-seq data. *Bioinformatics*, **38**, i290–i298.
42. Michna,A., Braselmann,H., Selmansberger,M., Dietz,A., Hess,J., Gomolka,M., Hornhardt,S., Blüthgen,N., Zitzelsberger,H. and Unger,K. (2016) Natural Cubic Spline Regression Modeling Followed by Dynamic Network Reconstruction for the Identification of Radiation-Sensitivity Gene Association Networks from Time-Course Transcriptome Data. *PLoS One*, **11**, e0160791.
43. Chua,C.Y., Liu,Y., Granberg,K.J., Hu,L., Haapasalo,H., Annala,M.J., Cogdell,D.E., Verploegen,M., Moore,L.M., Fuller,G.N., *et al.* (2016) IGFBP2 potentiates nuclear EGFR–STAT3 signaling. *Oncogene*, **35**, 738–747.
44. Zhang,J., Jia,J., Zhu,F., Ma,X., Han,B., Wei,X., Tan,C., Jiang,Y. and Chen,Y. (2012) Analysis of bypass signaling in EGFR pathway and profiling of bypass genes for predicting response to anticancer EGFR tyrosine kinase inhibitors. *Mol. Biosyst.*, **8**, 2645–2656.
45. Shi,K., Wang,G., Pei,J., Zhang,J., Wang,J., Ouyang,L., Wang,Y. and Li,W. (2022) Emerging strategies to overcome resistance to third-generation EGFR inhibitors. *J. Hematol. Oncol.*, **15**, 94.
46. Liu,L., Zhu,H., Liao,Y., Wu,W., Liu,L., Liu,L., Wu,Y., Sun,F. and Lin,H. (2020) Inhibition of Wnt/β-catenin pathway reverses multi-drug resistance and EMT in Oct4+/Nanog+ NSCLC cells. *Biomed. Pharmacother.*, **127**, 110225.
47. Lee,Y., Wang,Y., James,M., Jeong,J.H. and You,M. (2016) Inhibition of IGF1R signaling abrogates resistance to afatinib (BIBW2992) in EGFR T790M mutant lung cancer cells. *Mol. Carcinog.*, **55**, 991–1001.
48. Gong,Y., Duvvuri,M., Duncan,M.B., Liu,J. and Krise,J.P. (2006) Niemann-Pick C1 Protein Facilitates the Efflux of the Anticancer Drug Daunorubicin from Cells According to a Novel Vesicle-Mediated Pathway. *J. Pharmacol. Exp. Ther.*, **316**, 242–247.
49. Groth-Pedersen,L., Ostenfeld,M.S., Høyer-Hansen,M., Nylandsted,J. and Ja□a□ttela□,M. (2007) Vincristine Induces Dramatic Lysosomal Changes and Sensitizes Cancer Cells to Lysosome-Destabilizing Siramesine. *Cancer Res.*, **67**, 2217–2225.
50. Robey,R.B., Weisz,J., Kuemmerle,N.B., Salzberg,A.C., Berg,A., Brown,D.G., Kubik,L., Palorini,R., Al-Mulla,F., Al-Temaimi,R., *et al.* (2015) Metabolic reprogramming and dysregulated metabolism: cause, consequence and/or enabler of environmental carcinogenesis? *Carcinogenesis*, **36 Suppl 1**, S203-31.
51. Stine,Z.E., Schug,Z.T., Salvino,J.M. and Dang,C. V (2022) Targeting cancer metabolism in the era of precision oncology. *Nat. Rev. Drug Discov.*, **21**, 141–162.
52. Xu,C., Zhang,L., Wang,D., Jiang,S., Cao,D., Zhao,Z., Huang,M. and Jin,J. (2021) Lipidomics reveals that sustained SREBP-1-dependent lipogenesis is a key mediator of gefitinib-acquired resistance in EGFR-mutant lung cancer. *Cell Death Discov.*, **7**, 353.
53. Munir,R., Lisec,J., Swinnen,J. V. and Zaidi,N. (2019) Lipid metabolism in cancer cells under metabolic stress. *Br. J. Cancer*, **120**, 1090–1098.
54. Rysman,E., Brusselmans,K., Scheys,K., Timmermans,L., Derua,R., Munck,S., Van Veldhoven,P.P., Waltregny,D., Daniëls,V.W., Machiels,J., *et al.* (2010) De novo Lipogenesis Protects Cancer Cells from Free Radicals and Chemotherapeutics by Promoting Membrane Lipid Saturation. *Cancer Res.*, **70**, 8117–8126.
55. Cenciarelli,C., Marei,H.E., Zonfrillo,M., Pierimarchi,P., Paldino,E., Casalbore,P., Felsani,A., Vescovi,A., Maira,G. and Mangiola,A. (2014) PDGF receptor alpha inhibition induces apoptosis in glioblastoma cancer stem cells refractory to anti-Notch and anti-EGFR treatment. *Mol. Cancer*, **13**, 247.
56. Azuma,K., Kawahara,A., Sonoda,K., Nakashima,K., Tashiro,K., Watari,K., Izumi,H., Kage,M., Kuwano,M., Ono,M., *et al.* (2014) FGFR1 activation is an escape mechanism in human lung cancer cells resistant to afatinib, a pan-EGFR family kinase inhibitor. *Oncotarget*, **5**, 5908–5919.
57. Terai,H., Soejima,K., Yasuda,H., Nakayama,S., Hamamoto,J., Arai,D., Ishioka,K., Ohgino,K., Ikemura,S., Sato,T., *et al.* (2013) Activation of the FGF2-FGFR1 Autocrine Pathway: A Novel

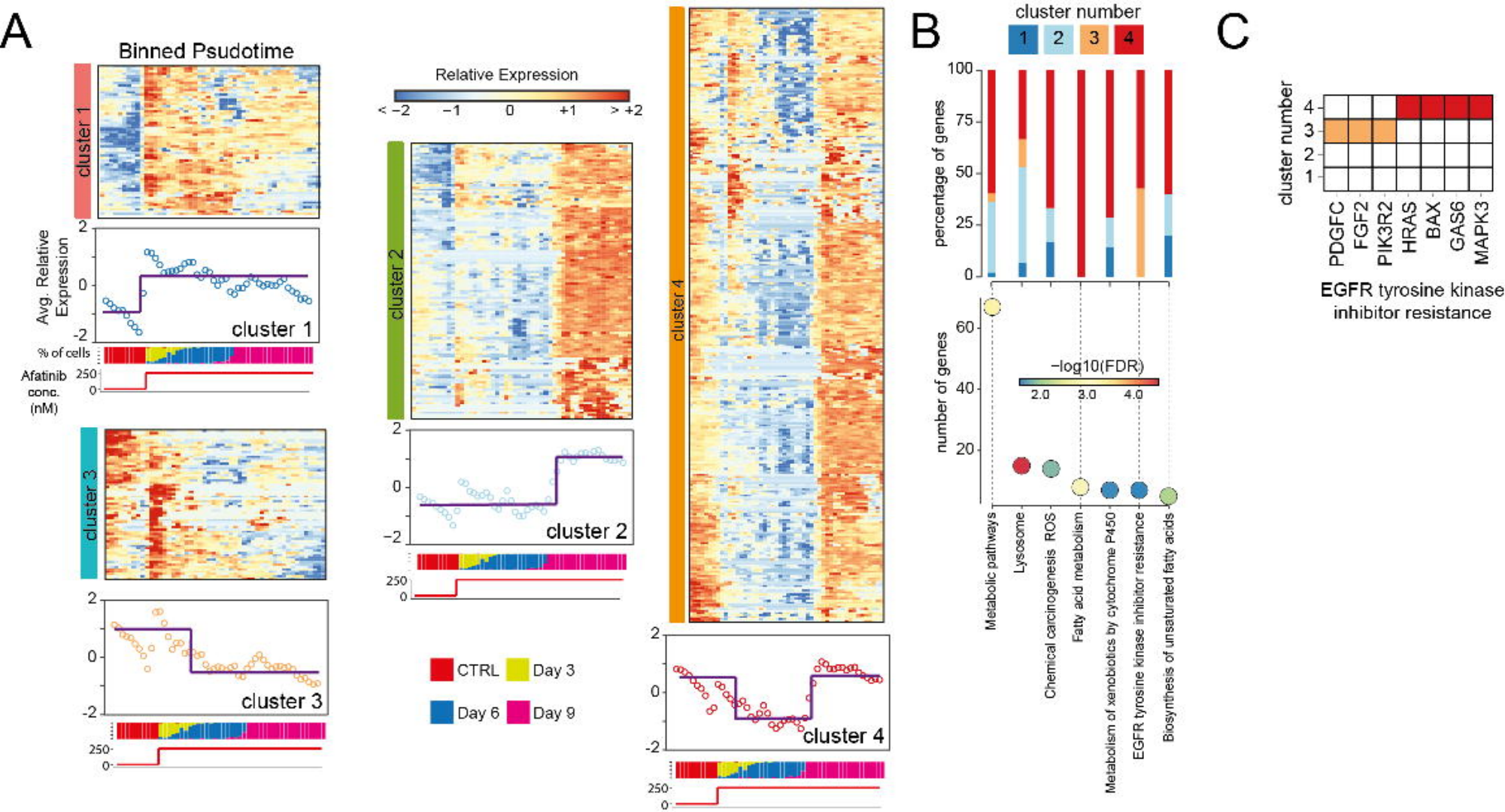
Mechanism of Acquired Resistance to Gefitinib in NSCLC. *Mol. Cancer Res.*, **11**, 759–767.
58. Mak,V.C., Li,X., Rao,L., Zhou,Y., Tsao,S.-W. and Cheung,L.W. (2021) p85β alters response to EGFR inhibitor in ovarian cancer through p38 MAPK-mediated regulation of DNA repair. *Neoplasia*, **23**, 718–730.
59. Leonetti,A., Sharma,S., Minari,R., Perego,P., Giovannetti,E. and Tiseo,M. (2019) Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br. J. Cancer*, **121**, 725–737.
60. Huang,L. and Fu,L. (2015) Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharm. Sin. B*, **5**, 390–401.
61. Roskoski,R. (2019) Targeting ERK1/2 protein-serine/threonine kinases in human cancers. *Pharmacol. Res.*, **142**, 151–168.
62. Zhang,Z., Lee,J.C., Lin,L., Olivas,V., Au,V., Laframboise,T., Abdel-Rahman,M., Wang,X., Levine,A.D., Rho,J.K., *et al.* (2012) Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat. Genet.*, **44**, 852–860.
63. Ma,G., Deng,Y., Qian,L., Vallega,K.A., Zhang,G., Deng,X., Owonikoko,T.K., Ramalingam,S.S., Fang,D.D., Zhai,Y., *et al.* (2022) Overcoming acquired resistance to third-generation EGFR inhibitors by targeting activation of intrinsic apoptotic pathway through Mcl-1 inhibition, Bax activation, or both. *Oncogene*, **41**, 1691–1700.
64. Weng,C.-H., Chen,L.-Y., Lin,Y.-C., Shih,J.-Y., Lin,Y.-C., Tseng,R.-Y., Chiu,A.-C., Yeh,Y.-H., Liu,C., Lin,Y.-T., *et al.* (2019) Epithelial-mesenchymal transition (EMT) beyond EGFR mutations per se is a common mechanism for acquired resistance to EGFR TKI. *Oncogene*, **38**, 455–468.
65. Tulchinsky,E., Demidov,O., Kriajevska,M., Barlev,N.A. and Imyanitov,E. (2019) EMT: A mechanism for escape from EGFR-targeted therapy in lung cancer. *Biochim. Biophys. Acta - Rev. Cancer*, **1871**, 29–39.
66. Taniguchi,H., Yamada,T., Wang,R., Tanimura,K., Adachi,Y., Nishiyama,A., Tanimoto,A., Takeuchi,S., Araujo,L.H., Boroni,M., *et al.* (2019) AXL confers intrinsic resistance to osimertinib and advances the emergence of tolerant cells. *Nat. Commun.*, **10**, 259.
67. Kim,D., Bach,D.-H., Fan,Y.-H., Luu,T.-T.-T., Hong,J.-Y., Park,H.J. and Lee,S.K. (2019) AXL degradation in combination with EGFR-TKI can delay and overcome acquired resistance in human non-small cell lung cancer cells. *Cell Death Dis.*, **10**, 361.
68. Okon,I.S., Coughlan,K.A., Zhang,M., Wang,Q. and Zou,M.-H. (2015) Gefitinib-mediated reactive oxygen specie (ROS) instigates mitochondrial dysfunction and drug resistance in lung cancer cells. *J. Biol. Chem.*, **290**, 9101–10.
69. Bekeschus,S. (2021) Acquired cancer tyrosine kinase inhibitor resistance: ROS as critical determinants. *Signal Transduct. Target. Ther.*, **6**, 437.
70. Macosko,E.Z., Basu,A., Satija,R., Nemesh,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A.R., Kamitaki,N., Martersteck,E.M., *et al.* (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, **161**, 1202–1214.
71. Maubec,E., Petrow,P., Scheer-Senyarich,I., Duvillard,P., Lacroix,L., Gelly,J., Certain,A., Duval,X., Crickx,B., Buffard,V., *et al.* (2011) Phase II Study of Cetuximab As First-Line Single-Drug Therapy in Patients With Unresectable Squamous Cell Carcinoma of the Skin. *J. Clin. Oncol.*, **29**, 3419–3426.
72. Mody,K., Strauss,E., Lincer,R. and Frank,R.C. (2010) Complete response in gallbladder cancer to erlotinib plus gemcitabine does not require mutation of the epidermal growth factor receptor gene: a case report. *BMC Cancer*, **10**, 570.
73. Doench,J.G., Fusi,N., Sullender,M., Hegde,M., Vaimberg,E.W., Donovan,K.F., Smith,I., Tothova,Z., Wilen,C., Orchard,R., *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.
74. Sanson,K.R., Hanna,R.E., Hegde,M., Donovan,K.F., Strand,C., Sullender,M.E., Vaimberg,E.W., Goodale,A., Root,D.E., Piccioni,F., *et al.* (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.*, **9**, 5416.
75. Gambardella,G. and di Bernardo,D. (2019) A Tool for Visualization and Analysis of Single-Cell RNA-Seq Data Based on Text Mining. *Front. Genet.*, **10**.
76. Franchini,M., Pellecchia,S., Viscido,G. and Gambardella,G. (2023) Single-cell gene set enrichment analysis and transfer learning for functional annotation of scRNA-seq data. *NAR Genomics Bioinforma.*, **5**, lqad024.
77. Zhang,Y., Parmigiani,G. and Johnson,W.E. (2020) ComBat-seq: batch effect adjustment for RNA-seq

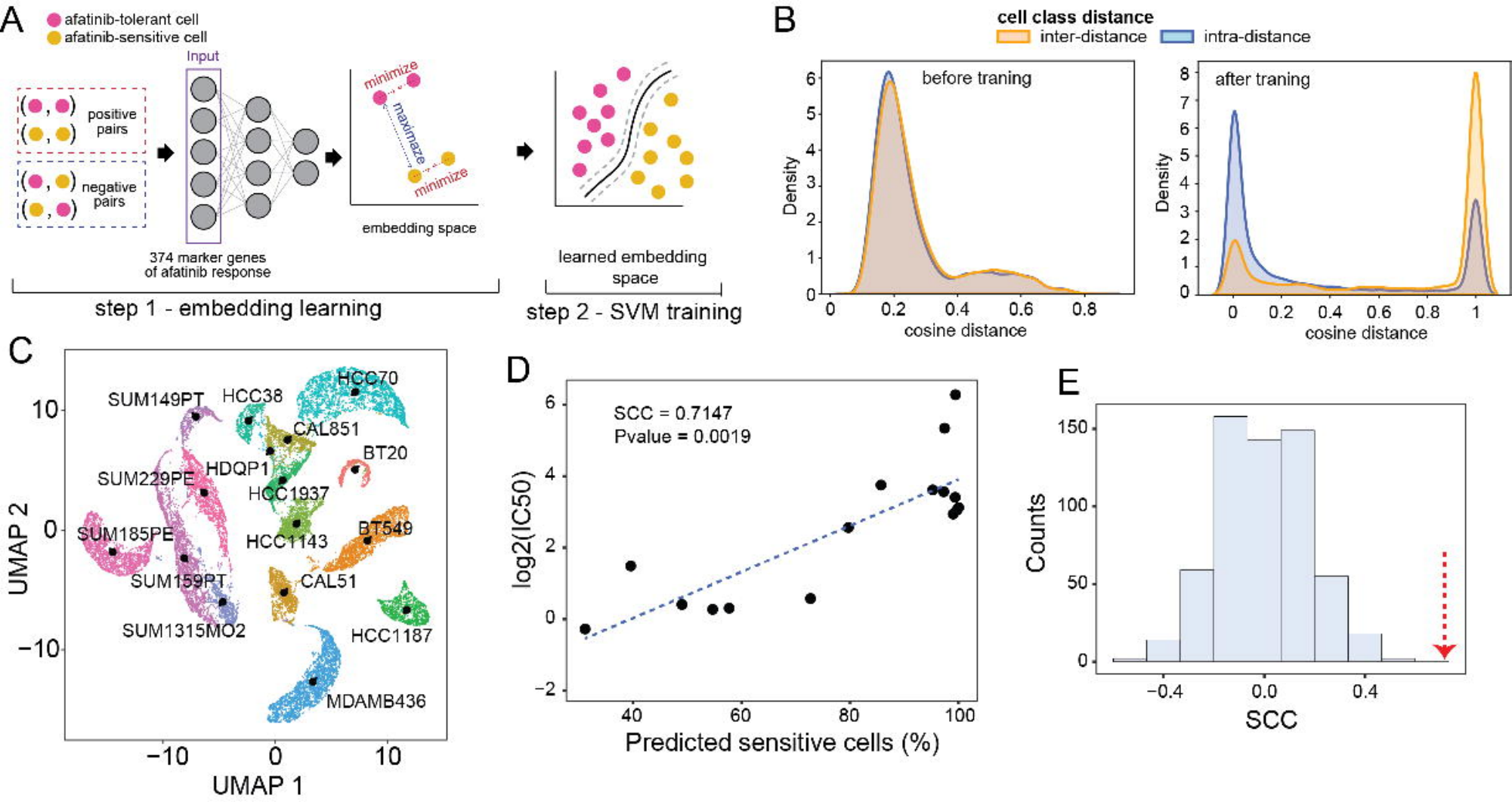
count data. *NAR Genomics Bioinforma.*, **2**, lqaa078.
78. Kleshchevnikov,V., Shmatko,A., Dann,E., Aivazidis,A., King,H.W., Li,T., Elmentaite,R., Lomakin,A., Kedlian,V., Gayoso,A., *et al.* (2022) Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.*, **40**, 661–671.
79. Di Veroli,G.Y., Fornari,C., Wang,D., Mollard,S., Bramhall,J.L., Richards,F.M. and Jodrell,D.I. (2016) Combenefit: an interactive platform for the analysis and visualization of drug combinations. *Bioinformatics*, **32**, 2866–2868.
80. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**.
81. Grossman,R.L., Heath,A.P., Ferretti,V., Varmus,H.E., Lowy,D.R., Kibbe,W.A. and Staudt,L.M. (2016) Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.*, **375**, 1109–1112.
82. Colaprico,A., Silva,T.C., Olsen,C., Garofano,L., Cava,C., Garolini,D., Sabedot,T.S., Malta,T.M., Pagnotta,S.M., Castiglioni,I., *et al.* (2015) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.*, **44**, e71–e71.
83. Dzul Keflee,R., Leong,K.H., Ogawa,S., Bignon,J., Chan,M.C. and Kong,K.W. (2022) Overview of the multifaceted resistances toward EGFR-TKIs and new chemotherapeutic strategies in non-small cell lung cancer. *Biochem. Pharmacol.*, **205**, 115262.
84. Kalaitzis,A.A. and Lawrence,N.D. (2011) A Simple Approach to Ranking Differentially Expressed Gene Expression Time Courses through Gaussian Process Regression. *BMC Bioinformatics*, **12**, 180.
85. Kingma,D.P. and Ba,J. (2017) Adam: A Method for Stochastic Optimization.


A
untreated cells
VS
drug tolerant cell
drug sensitivecell
Identify preexisting transcriptional states mediating drug resistance
retrospective lineage tracing
untreated
Drug treatment
unbarcoded cells
barcoded cells
Prospective lineage tracing
Identify transcriptional states driving drug adaptation
gene exspression
treated cells
clone
B
MDAMB468
TNBC cell-lines
anti-EGFR log(IC50)
C
MDAMB468-P
CTRL (D0)
+ Afatinib 250nM
D3
D6
D9
D40
2000nM
Single Cell Transcriptomics
Bulk lineage retrieval
MDAMB468-ATPC
D
MDAMB468-P
MDAMB468-ATPC
Cell viability (OD 490)
Afatinib Concentration (μM)
E
Afatinib IC50(uM)
F
Num. of lineages
2336
192
MDAMB468-P
MDAMB468-ATPC


A
Days
D0 (CTR)
D3
D6
D9
UMAP 2
UMAP 1
B
Afatinib sensitive lineages
Afatinib tolerant lineages
D6
D3
D9
D0 (CTR)
UMAP 1
C
UP in Afatinib sensitive clones
UP in Afatinib tolerant clones
IGFBP2
-log10(FDR)
Log2 Fold Change
D
Afatinib sensitive lineages
Afatinib tolerant lineages
P=1e-15
P=4.7e-5
P=1e-4
P=4.3e-8
IGFBP2 expression
CTR
3
6
9
Afatinib (250nM)
Days
E
P = 4.3e-4
Relative IGFBP2 expression
MDAMB468-P
MDAMB468 ATPC
F
Linage
bc14-013:bc30-092942
bc14-065:bc30-046216
other clones
9.9% of cells
35.4% of cells
Cellular Frequency Log2 (Fold change)
CTR
3
6
9
Afatinib (250nM)
Days
G
50.1%
30.8%
Cellular frequency (ATPC)
Lineage
bc14-013:bc30-092942
bc14-065:bc30-046216
H
UP-regulated genes in the dominant clones (549)
Avg. Expression log2(CPM) (Dominant clones)
Avg. Expression log2(CPM) (Neutral clones)
I
IGFBP2
MYEOV
SERPINB6
ALCAM
BST2
SLPI
CLDN10
MME
RNF168
OSBP2
Avg. log2(FC)
J
Dominant clones
Sub-clones
P = 5.1e-12
P = 3e-7
P = 3.5e-15
P = 6.1e-12
IGFBP2 expression
CTRL
3
6
9
Afatinib (250nM)
Days


A
CTR
IGFBP2 Knockdown
MDAMB468
BF = 4.722
****
N=4
Cell viability (%)
Afatinib Concentration (μM)
B
CTR
IGFBP2 Over-expressed
MDAMB468
BF = 3.002
***
N=4
Cell viability (%)
Afatinib Concentration (μM)
C
CTR
IGFBP2 Over-exp.
Afatinib Concetration
2 μM
5 μM
10 μM
P = 0.018
P = 0.021
P = 0.044
Colony formation (Abs 590nm)
D
Spheroid Volume (μm3)
P = 0.02
P = 7.71e-3
P = 3.9e-06
P = 1.2e-4
(Day)
Afatinib conc. (μM)
E
CTR
IGFBP2 Knockdown
Afatinib Concetration
0.5 μM
1 μM
P = 7.1e-4
Colony formation (Abs 590nm)
F
Spheroid Volume (μm3)
n.s.
P = 3.4e-4
P = 0.01
P = 5.25e-3
(Day)
Afatinib conc. (μM)
G
Spheroid Volume (μm3)
n.s.
P = 0.013
P = 1.8e-3
P = 2.9e-5
(Day)
H
Spheroid Volume (μm3)
P = 0.01
P = 5.55e-3
P = 2.2e-4
P = 3.6e-3
(Day)
I
MDAMB468
CTR
IGFBP2 Overexpressed
absorbance 570nm
P = 0.001
CTR
IGFBP2 Knockdown
IGFBP2 Over-expressed


A
Binned Psudotime
Relative Expression
< −2
−1
0
+1
> +2
cluster 1
cluster 2
cluster 3
cluster 4
Avg. Relative Expression
% of cells
Afatinib conc. (nM)
CTRL
Day 3
Day 6
Day 9
B
cluster number
1
2
3
4
percentage of genes
number of genes
−log10(FDR)
Metabolic pathways
Lysosome
Chemical carcinogenesis ROS
Fatty acid metabolism
Metabolism of xenobiotics by cytochrome P450
EGFR tyrosine kinase inhibitor resistance
Biosynthesis of unsaturated fatty acids
C
cluster number
PDGFC
FGF2
PIK3R2
HRAS
BAX
GAS6
MAPK3
EGFR tyrosine kinase inhibitor resistance


A
afatinib-tolerant cell
afatinib-sensitive cell
Input
positive pairs
negative pairs
374 marker genes of afatinib response
minimize
maximaze
minimize
embedding space
learned embedding space
step 1 - embedding learning
step 2 - SVM training
B
cell class distance
inter-distance
intra-distance
before traning
after traning
Density
cosine distance
C
SUM149PT
HCC38
HCC70
CAL851
BT20
HDQP1
SUM229PE
HCC1937
SUM185PE
HCC1143
BT549
SUM159PT
CAL51
SUM1315MO2
HCC1187
MDAMB436
UMAP 1
UMAP 2
D
SCC = 0.7147
Pvalue = 0.0019
log2(IC50)
Predicted sensitive cells (%)
E
Counts
SCC