# Epigenetic heritability of cell plasticity drives cancer drug resistance through one-to-many genotype to phenotype mapping

*Javier Fernandez-Mateos[1,*], Salvatore Milite[2,*], Erica Oliveira[1], Georgios Vlachogiannis[1], Bingjie Chen[1], Erika Yara[1], George D Cresswell[1], Chela James[2], Lucrezia Patruno[3], Gianluca Ascolani[3], Ahmet Acar[1,4], Timon Heide[2], Inma Spiteri[1], Alex Graudenzi[3], Giulio Caravagna[5], Trevor Graham[2], Luca Magnani[6,7], Nicola Valeri[1,7,§], Andrea Sottoriva[1,2,§]*

*[1] Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK*
*[2] Computational Biology Research Centre, Human Technopole, Milan, Italy*
*[3] Department of Informatics, Systems and Communication, University of Milan-Bicocca, Milan, Italy*
*[4] Department of Biological Sciences, Middle East Technical University, Ankara, Turkey*
*[5] Department of Mathematics and Geosciences, University of Triest, Italy*
*[6] Division of Breast Cancer Research, The Institute of Cancer Research, London, UK*
*[7] Department of Surgery and Cancer, Imperial College London, UK*

*[*] These authors contributed equally to this work*
*[§] Correspondence to: nicola.valeri@icr.ac.uk and andrea.sottoriva@fht.org*

## Abstract

Drug resistance is a largely unsolved problem in oncology. Despite the explanatory power of the genetic model of cancer initiation, most treatment resistance is unexplained by genetics alone. Even when known resistance mutations are present, they are often found in a small proportion of the cells in the tumour. So where is the cellular memory that leads to treatment failure? New evidence suggests resistance is multi-factorial, resulting from the contribution of heritable genetic and epigenetic changes, but also non-heritable phenotypic plasticity. However, cell plasticity has proven hard to study as it dynamically changes over time and needs to be distinguished from clonal evolution where cell phenotypes change because of Darwinian selective bottlenecks. Here we dissected the contribution of different evolutionary processes to drug resistance by perturbing patient-derived organoids with multiple drugs in sequence. We combined dense longitudinal tracking, single cell multi-omics, evolutionary modelling, and machine learning archetypal analysis. We found that different drugs select for distinct subclones, an essential requirement for the use of evolutionary therapy with sequential drug treatment. The data supports a model in which the cellular memory is encoded as a heritable configuration of the epigenome, which however produces multiple transcriptional programmes. Those emerge in different proportions depending on the environment, giving rise to cellular plasticity. Epigenetically

encoded programmes include reactivation of developmental genes and cell regeneration. A one-to-many (epi)genotype→phenotype map explains how clonal expansions and non-heritable phenotypic plasticity manifest together, including drug tolerant states. This ensures the robustness of drug resistance subclones that can exhibit distinct phenotypes in changing environments while still preserving the cellular memory encoding their selective advantage.

## Introduction

Cancers develop following the Darwinian rules of clonal evolution, where selected subclones with new genetic mutations come to dominate the cellular compartment[1]. The same paradigm is extended to explain the emergence of treatment resistance, arguably the biggest problem in oncology today[2]. However, only a subset of resistance mechanisms has been identified[3–5]. Even when genetic mutations known to confer resistance are detected in the relapsed tumour, they explain only a proportion, sometimes minor, of the resistance phenotype. In many patients, treatment failure remains entirely unexplained by genetic alterations alone. Accumulating evidence suggests that additional Darwinian mechanisms involving non-genetic alterations[3,6,7], as well as non-Darwinian cellular plasticity[8] also contribute to tumourigenesis. Importantly these mechanisms co-exist in the same tumour at the same time and hence therapy resistance is likely multi-factorial.

The identification of epigenetic changes driving cancer evolution has been hindered by the focus on genetic profiling of tumours and the lack of proper controls to compare cancer and normal epigenomes from the same tissue of origin[9]. Also, cancer cell plasticity has been particularly elusive to quantitative characterisation because it dynamically changes over time and it is hard to determine whether the recurrent tumour's cells have changed their phenotype in a plastic manner or by clonal evolution[10]. Elegant experiments base on the Luria-Delbruck approaches have demonstrated in cell lines that indeed one can distinguish plasticity from Darwinian adaptive changes[11]. Such mechanisms have also been identified in patient-derived xenograft models under the pressure of anti-EGFR drugs[12].

However, fundamental open questions remain, such as "to what extent are epigenetic changes heritable upon cell division?", "is plasticity a new cancer programme or some reactivated cellular state?", "is plasticity reversible?", "is the propensity to cell plasticity in itself a heritable trait?". Since therapy resistance is multi-factorial, another important question is: "are different subclones in the same tumour adapting differently to drugs?". Answering these complex questions requires concomitant measurements of genomes, epigenomes and transcriptomes, matched with cell lineage histories to deconvolute Darwinian from non-Darwinian mechanisms. Because drug resistance is multi-factorial in the same tumour, this has to be done at single cell resolution.

In this study we design an evolutionary experiment using patient-derived tumour organoids under the pressure of different sequences of drugs. We used expressed lentiviral barcodes combined with single cell multi-omics to track the evolution of single cell genomes, epigenomes and transcriptomes over time. We study adaptation in persistor cells under the pressure of the drug, and in the same cell lineages when drug pressure is released to understand how the therapeutic microenvironment dynamically perturbs cells. We leverage

on archetypal analysis to chart the mapping between genomes, epigenomes and multiple transcriptomes in individually tracked subclones.

# Results

## Experimental Design

We used an AKT mutant and amplified metastatic colorectal cancer patient-derived organoid previously derived from a clinical trial[13] as model system of oncogene addiction. We infected the organoid cells with a library of 5 million random Cellecta® expressed barcodes, a form of lentiviral barcodes that generate a poly-A transcript that can be captured by single cell RNA sequencing. We ensured to have one barcode per cell through dilution of 0.1 MOI. Single infected cell lineages are subject to high genetic drift due to cell turnover and death, giving rise to a final distribution of barcodes that scales as a power-law distribution (Supplementary Figure 1). This is the expected distribution of lineages we will use as a comparison for our evolutionary experiments.

First, we performed single-cell RNA-seq and barcode analysis at baseline. We refer this as the *parental* sample. We extracted the barcodes both from the single-cell RNA transcripts and from the bulk DNA of the cells, identifying 3,000-5,000 individual cell lineages (Figure 1A). We expanded the barcoded population to 75 million cells and 2.5M cells/well were plated in multiple 6-well plates. We exposed the organoids to a set of first line drugs for 45 days at 2uM concentration. After that, we collected the cells under the influence of the drug to characterise resistant and drug tolerant states, and performed bulk DNA barcode analysis and single-cell RNA-seq. We left the other three wells regrowing in the absence of the drug until confluence and harvest the population again for bulk DNA profiling and single-cell RNA, again extracting the barcodes from both. We re-plated the cells into a new 6-well plate for a second drug and repeated the profiling as before. Notably, we collected floating barcodes by magnetic bead purification from the media change every 2 days during the whole course of the experiment by spinning the supernatant to extract barcodes from the floating DNA. This allowed to have five “solid” timepoints from the system (Figure 1A) but also 58-61 floating barcode timepoints per drug sequence for characterising the system evolution without perturbing it with physical sampling. This is akin to using liquid biopsies to profile circulating tumour DNA in patients. We chose a set of relevant drugs that are clinically relevant for the model system we used, namely an allosteric AKT inhibitor (MK-2206), a competitive AKT inhibitor (capivasertib), and mTOR inhibitor (KU-0063794) and a MEK inhibitor (trametinib). We designed different sequences of drugs as in Figure 1B. We confirmed the cells became resistance to those drugs (Figure 1C).

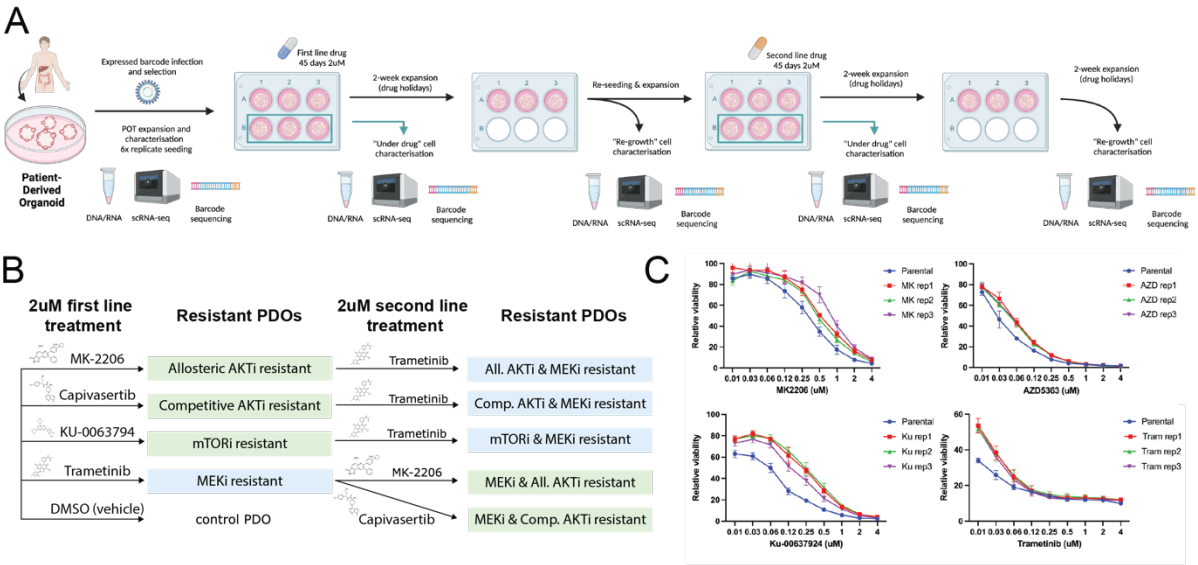



***Figure 1. (A)*** *Experimental design of long-term drug resistance evolution. We performed bulk DNA profiling for genomic characterisation and barcode measurement, as well as single-cell RNA-seq and corresponding single cell barcode extraction of five "solid" timepoints over a five month period: parental, under drug 1, regrowth after drug 1, under drug 2 and regrowth after drug 2. We also collected floating DNA every three days from the supernatant to profile barcodes as a "liquid biopsy".* ***(B)*** *We exposed the cell to 5 different sequences of drugs.* ***(C)*** *Drug response curves show the cells became indeed resistant to pharmacological intervention.*

## Resistance evolution to targeted drugs is heritable, pre-existing and highly repeatable

In all cases we found that drugs selected for one or a few subclones that massively expanded during treatment (Figure 2A). The three biological replicas we seeded per drug showed almost identical composition of selected barcodes, indicating that those subclones were pre-existing in the population and were selected by Darwinian pressure.

When we analysed the floating barcodes that contained dense longitudinal information about the system, also the temporal dynamics were extremely similar between replicas, indicating that evolution is highly repeatable under the strong selective pressures of cancer drugs, depending strongly on the initial conditions of the system (Figure 2B). These data imply that the drug resistant phenotype must be heritable, as the cellular memory encoding the drug resistant features is passed on to the offspring of a given cell lineage together with the lentiviral barcode.

We measured the growth rates over time of different clonal subpopulations (barcodes), showing that most surviving clones were not growing under the drug pressure, but were instead drug tolerant. When the drug pressure was releases, those same subclones massively expanded with a high selective coefficient of up to *s=0.5.*

## Different targeted drugs select for different pre-existing clones

We then investigated what subclones were selected by different drugs. The two AKT inhibitors, despite targeting the same signalling molecule, selected for distinct subclones (Figure 2A). Instead, the mTOR inhibitor selected similar subclones as the allosteric AKT

inhibitor. The MEK inhibitor trametinib selected for an entirely different subclone that became dominant after drug exposure. The selective advantage of this trametinib resistant *blue* subclone was very strong (Figure 2A,B). It also proved to be a persistor in other drug contexts. For example, while during first line treatment of AKT and mTOR inhibition it was not clearly displaying a selective advantage, it remained lingering in the population, until a second line of MEK inhibitor was introduced and the subclone took over with very high selective advantage.

These data indicate that evolutionary adaptation to targeted drugs is highly repeatable, driven by strong selective pressure on pre-existing subclones, that led to evolutionary divergence of subclones under distinct drugs. Therefore, evolution is determined by the initial conditions of the system, making it highly predictable, with the opportunity of exploiting evolutionary trade-offs for subclones in different drug contexts with specific sequences of drugs. These are the necessary conditions to exploit evolutionary *steering* strategies[14].

We performed bulk whole-genome sequencing of all samples at re-growth and compared drug resistant samples with parental to identify which genetic alterations may be responsible for resistance. We found a small set of somatic mutations and copy number alterations that were different in resistant samples (Supplementary Figure 2 and 3). However, none of these variants were convincingly a candidate as a mechanism of drug resistance. We then turned to the analysis of the transcriptomes to understand this cell phenotype.

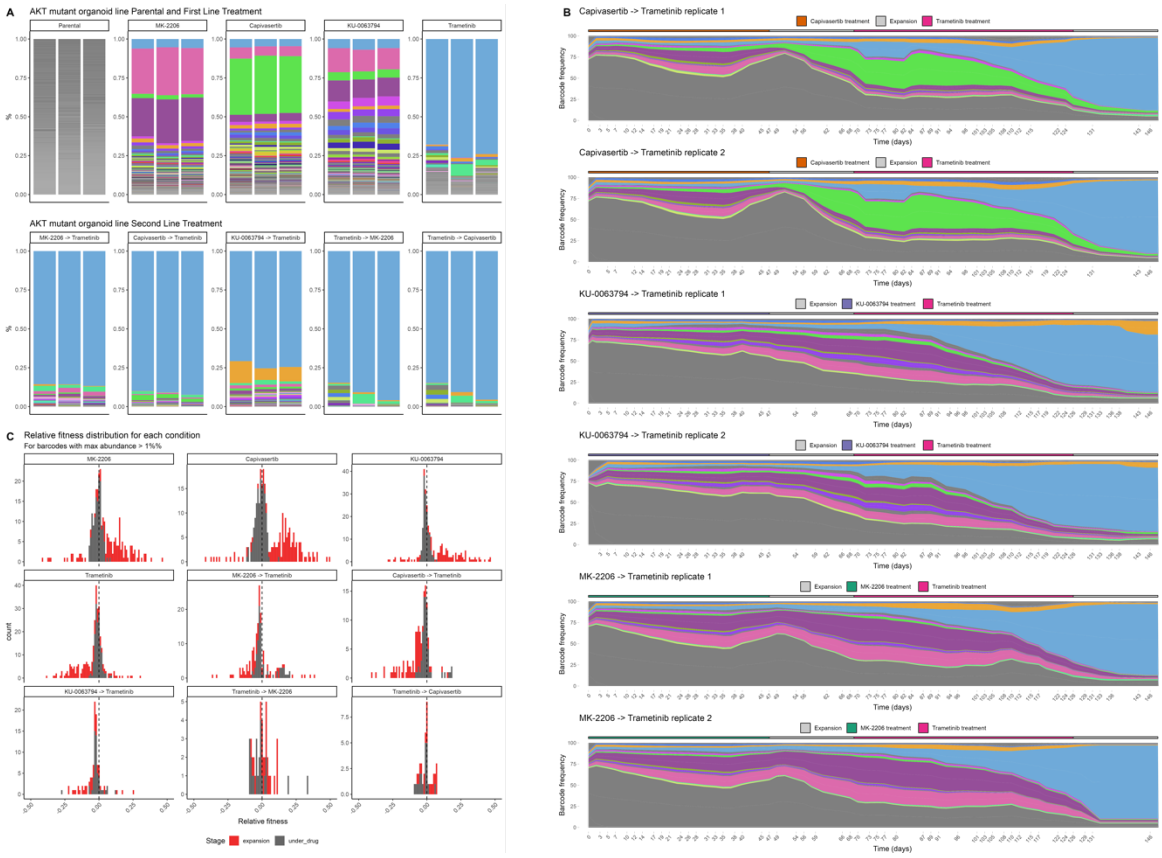



***Figure 2. Evolutionary dynamics of barcoded population*** *(A) Lentiviral barcodes proportion. For each replica and drug condition we quantified barcode proportions from genomic DNA,*

*the only exception being Trametinib for which we used the proportion quantified from 10x scRNA-seq. The top 100 barcodes have a unique colour across the whole experiment, all the others (< 2% abundance) are shown in grey. All the barcodes are quantified after the re-growth period. Selection is evident compared to the parental population. (B) Reconstruction of clonal dynamics using floating barcodes, extracted from culture media every 3 days over the whole length of the experiment. The dynamics shows an evident clonal sweep of the blue barcode after second line treatment with Trametinib. Colour code is the same as in panel (A). Proportions are smoothed by taking a rolling average on a window of 7 points. (C) Relative fitness distribution shows high variability in re-growth phase during first-line treatment. Coefficients were calculated by assuming exponential growth and independence between different clonal populations.*

## Archetype analysis reveals transcriptional programmes

We performed single cell RNAseq with 10x Genomics and single cell barcode extraction from a total of 19 samples. We managed to recovery the barcode of 46% of cells in average per sample [30-64%]. As expected, transcriptomes of cells under treatment clustered separately from parental, reflecting the strong influence of the environment on the transcriptomic profile of a cell (Figure 3A). Cells under different drugs also clustered separately, suggesting each drug influences differently the transcriptome (Figure 3B). Single cell barcodes recapitulated barcode distributions from bulk profiling (Figure 3C). As expected, most cells in S phase belonged to samples from re-growth after drug pressure (Figure 3D). Given the strong selective bottlenecks imposed by drugs, we conducted an initial per-sample analysis neglecting the subclonal composition of each sample. We used markers of known cell types in the intestine to classify the transcriptional programmes active in each sample (Figure 3E). The parental sample was characterised by a small but significant component of LGR5+ stem-like cells and a large component of transient amplifying cells (TA). Under the pressure of first line AKT (capivaserib and MK-2206) and mTOR inhibitors (KU-0063794), which act on the PI3K/AKT pathway, the population maintained a stem cell component but expanded significantly a population with a Paneth signature, similarly to what previously reported for patient-derived xenografts under the pressure of EGFR inhibitors[12]. Inhibiting the MAPK pathway with the MEK inhibitor (trametinib), even after PI3K/AKT suppression, induced instead different transcriptional dynamics where the LGR5+ stem-like component and partly a Goblet like component emerged. Instead, blocking AKT *after* having previously inhibited MEK gave rise to yet another transcriptional programme with low LGR5+ stem component but high ASCL2+ stem-like cells. At regrowth, stem-like programmes were evident and a tendency of microfold cells. Instead for second line drugs a wider range of stem-like programmes emerge, again together with Paneth cell programmes.

Given the heterogeneity of transcriptional programmes we found, we hypothesise that a high level of phenotypic plasticity, possibly a reflection of aberrant differentiation pathways, was present in our cell populations. To deconvolve this signal we used Archetypal Analysis, a dimensionality reduction method that decomposes an input matrix as a convex combination of ideal extreme points called archetypes[15]. In this study we implemented a Deep Archetypal Analysis framework to account for non-linearity in the dimensionality reduction step[16] (Figure 3F and Supplemetary Figure 2). We found that distinct archetypes were enriched for clearly different cell type markers, highlighting the fact that they represent

highly distinct cellular programmes, spanning all eight of the cell types we considered (Figure 3G). Archetype presence clustered distinctively between cells as expected (Figure 3H). These results indicate that archetypes are robust, cluster populations of cells in a meaningful way that is biologically interpretable.

## The drug microenvironment produces plastic shifts in transcription that are reversible

We then deconvoluted the archetypes for each cell dividing cells by barcode. The evolutionary analysis indicated that pre-existing subclones were selected repeatedly by drugs, suggesting a strong cellular memory in those subpopulations that we expected to be different from the rest of the parental population (hence why they were selected) also at the transcriptional level. We found that the selected subclones were indistinguishable in their transcriptome from the rest of the parental population (Figure 3I). Importantly, the subclones did not separate by archetype, and instead cells with the same barcodes represented multiple archetypes at the same time (i.e., archetypes were represented in each barcode), suggesting a “differentiation” hierarchy still present in cell lineages. Surprisingly however, it was the drug environment that would be able to shift the archetype distribution within a subclone, which after drug suspension and re-growth, was reversible. Hence the selected subclones were transcriptionally similar to their background parental population, changed under drug, but returned to their original state after drug pressure was released.

The issue then was reconciling a clear plastic phenotype of the transcriptional programmes with a pre-selected Darwinian dynamics we observed in the barcodes. We thought epigenetics could be the link.

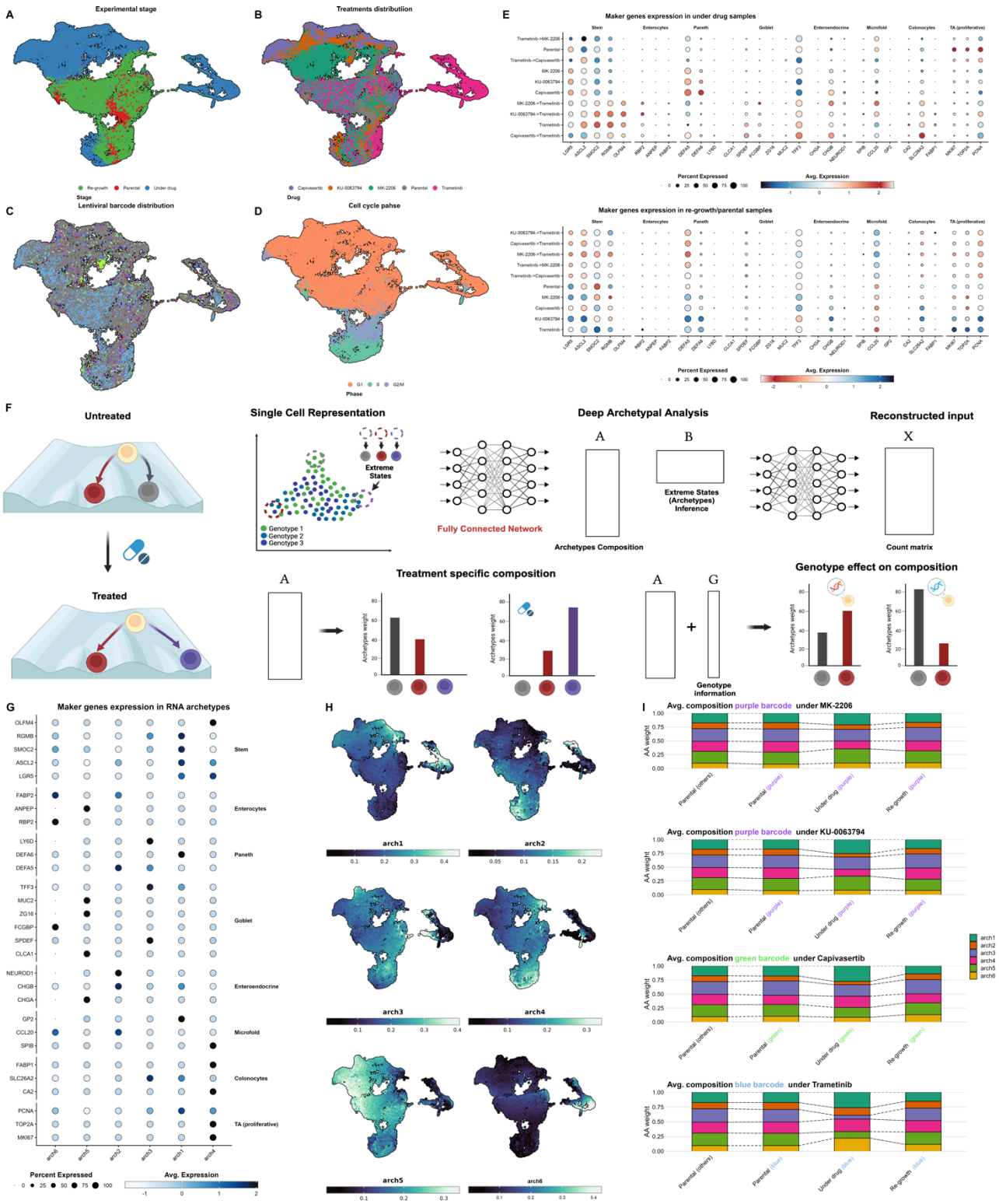



***Figure 3. Transcriptional programmes show plasticity after drug administration*** *(A-D) UMAP of the 37000 cells in the experiment after QC filters coloured respectively by: experimental stage, drug, barcode and cell cycle phase. Cells for which we were not able to extract a valid barcode or with an abundance of less than 1% are shown in grey. It is evident how cells in under drug phase tend to strongly cluster by drug, while they tend to mix back with the parental during re-growth. (E) Z-score distribution of adult colonic cell type markers from ref[17] shows the presence of distinct differentiation programmes inside the organoid. (F) Archetypal Analysis (AA) aims at decomposing the input dataset as a convex combination of*

*extreme points by learning two matrices A and B, which are respectively the archetype weights for each point in the dataset and the matrix that defines the archetypes starting from the input dataset. Here we use a deep implementation of AA. We can then exploit the weights of the A matrix to quantify differences in transcriptional programmes across conditions and genotypes (G) Z-score distribution for the same genes as in panel (E) but computed by archetype. (H) Archetypes weight distribution over UMAP. (I) Average archetype weight for different selected barcodes. The trend is consistent with cells going back to the parental phenotype after re-growth. Colours for barcodes are consistent with Figure 2*

## The epigenome but not the transcriptome of selected subclones reveal their identity

We performed single cell multi-omics of RNA and chromatin accessibility on the same cell with 10x Multiome on four samples: parental, trametinib, capivaserib and capivaserib→trametinib, as those were samples with the strongest bottlenecks. We found the embeddings with UMAP clearly separated the cells by treatment, with a much stronger signal than the previous scRNA data, in particular for the scATAC component (Figure 4A). We inferred archetypes from the ATAC signal and called 3 archetypes that distributed differently in distinct samples. We then used the scATAC signal to infer copy number alterations with CONGAS[18], identified 10 CNA subclones and reconstruct their phylogenetic history (Figure 4B). Again, each subclonal lineage defined in this case by a chromosomal configuration contained multiple ATAC archetype programmes as inferred from the RNA, but were differently distributed in distinct samples due to selection by drugs, as inferred from the barcodes. We then used the lentiviral barcode recovery from the RNA component of the assay to assign barcodes to cells. The efficiency of barcode recovery here was lower (mean per sample = 27% [2-44%]) and we could recovery a strong signal only for the 'blue' barcode selected under trametinib and capivasertib. The proportion of epigenetic archetypes, unlike for scRNA, here was indeed different in the parental between the pre-existing selected barcode and the rest of the population, suggesting that the memory that induced the selection is encoded in the epigenome (Figure 4C). Indeed, this memory is maintained in the clone after it expands under AKT and MEK inhibition (Figure 4C). We wanted to extend the analysis to other clones but as mentioned earlier the recovery of the barcodes was not as high as for scRNA. We then performed subclonal decomposition using the copy number profiles and again compared transcriptional programmes using the RNA archetypes against their inherited CNA 'genotype' (Figure 4D). The transcriptional programmes were very similar when comparing selected versus non-selected subclones, whereas their epigenetic programmes were clearly different (Figure 4E), again indicating that there is a heritable epigenetic memory that encores many transcriptional phenotypes downstream.

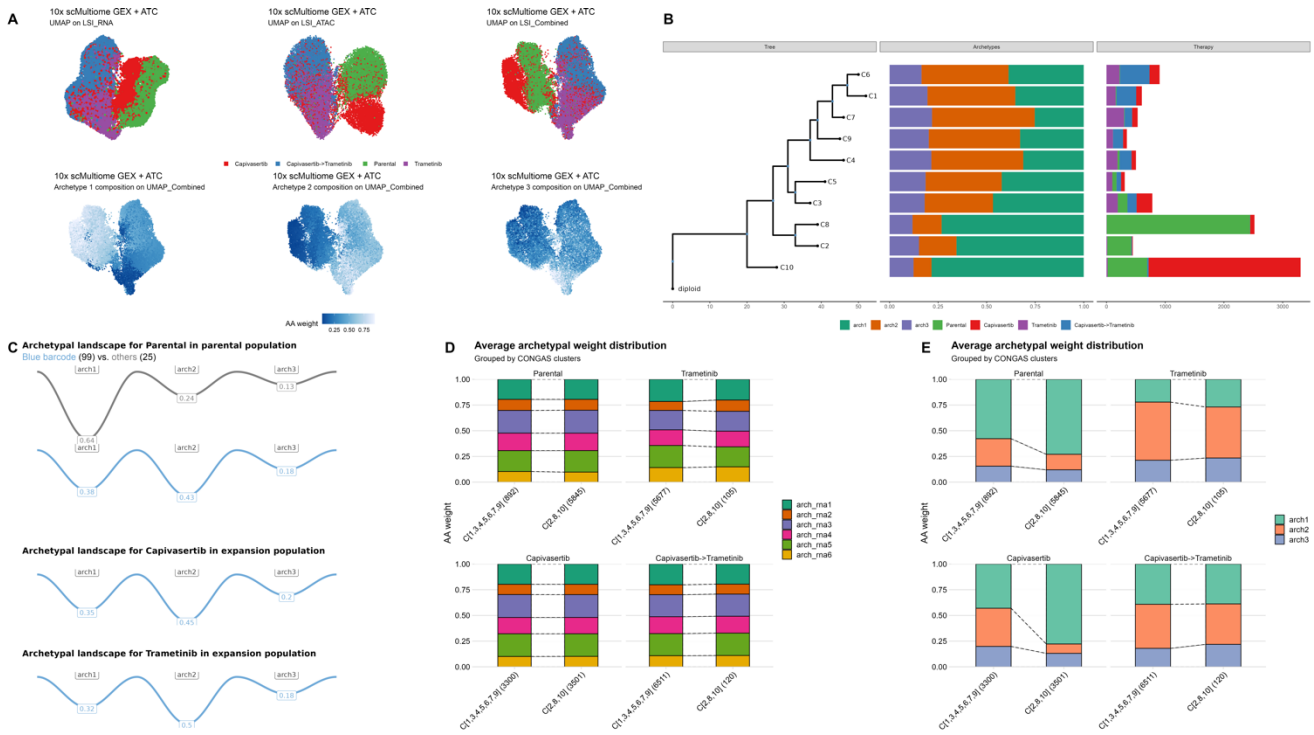



***Figure 4. Epigenetic rewiring in Trametinib resistant population*** *(A) UMAP plots for 10x multiome samples. In the first-row different reduction are shown, respectively UMAP done with Latent Semantic Index (LSI) exploiting just RNA information, LSI with just ATAC and a combined LSI. In the second row we colour the Combined UMAP by archetypal weight (B) Clonal tree constructed starting from Copy Number Alterations inferred from ATAC data (C) Average archetype weights. The blue barcode displays a clear difference in the ATAC profile compared to the others (in grey). (D-E) Average archetype weight for copy number clones, for RNA archetypes (D) and ATAC archetypes (E). We split the tree in two clades, the top one more abundant in the Trametinib samples and the bottom one, more represented in the Parental and Capivasertib samples. The change in archetypal composition is consistent with what we saw with the lentiviral lineage tracing.*

## Multiomic epigenetic analysis reveal heritable rewiring of the epigenome driving plastic phenotypes

Having not found somatic DNA mutations linked to the resistance phenotype, we analyse the epigenome of the four Multiome cases. First, we looked for recurrent focal changes in chromatin accessibility in promoters and enhancers. We found a very small set of differentially accessible peaks in the data (Supplementary Figure X) of which only one was of particular interest: increased accessibility to the promoter of SETBP1, a known epigenetic regulatory hub[19] (Figure 5A). We then also looked at genome-wide rewiring of the epigenome by focussing on accessibility to transcription factor binding sites across the genome. We found that capivaserib→trametinib rewired the epigenome…

We found a significant correlation between gene expression of the TF and associated chanes in the accessibility of the corresponding binding sites for a set of those TFs.

A previous study in pancreatic cancer showed that resistance to trametinib was associated with increase in autophagy and decrease in MYC activity[20]. We find the same here, where

the only increase in autophagy signature was found under trametinib (Figure 5E). However, barcode analysis showed that the signal was not coming from the fully resistant subclone (the 'blue' trametinib resistant barcode), but from the residual persister cells that would die off in the long run, overtaken by the 'blue' barcode.

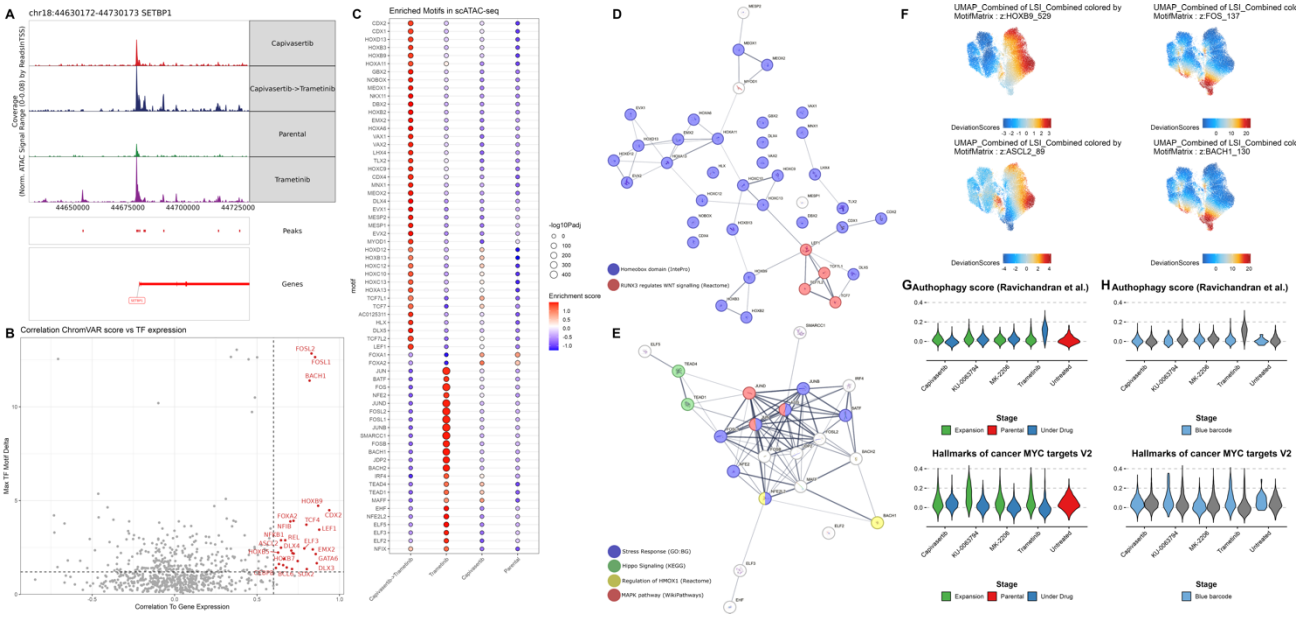



***Figure 5. Biological characterization of Trametinib resistant population*** *(A) SETBP1 chromatin profile in the promoter region. (B) Scatterplot of Transcription Factor (TF) Motifs variance computed using ChromVAR[21] from the ATAC portion of the multiome and expression from the GEX part. Points in red are enriched TFs that also show consistent high change in expression, selected as having correlation > 0.6, adjusted p-value < 0.001 and TFs Motif variance greater than the 10% quantile. (C) Heathmap of enriched Motifs in Marker Peaks. We filtered Enriched Motifs with FDR <= 0.001 & Log2FC >= 1 (D-E) StringDB analysis of Motifs enriched respectively in Capivasetib->Trametinib and Trametinib samples, showing involvement of TFs active in WNT and MAPK signalling pathways as well as TFs containing HomeoBox domains. Analysis was conducted using the browser version 12.0 of StringDB[22] (F) UMAP plots for some of the top enriched Motifs. Sample specific localization is evident. (G-H) Gene Module score for the autophagy signature in Ref.[20] and the MSigDB Hallmark of Cancer MYC targets V2[23], grouped respectively by phase and barcode. The plot shows how in the under-drug phase cells under Trametinib have and high level of autophagy and a correspondingly lower level of Myc activation. The blue resistant barcode however shows an opposite trend and seems to resemble the parental and re-growth behavior even under drug. Panel (H) scores are computed for the under-drug samples.*

## A one-to-many (epi)genotype to phenotype mapping drives resistance

The evolutionary analysis points at clear Darwinian dynamics driving resistance. At the same time the transcriptional programmes do not show such a Darwinian memory. Giving the lack of genetic alterations responsible for resistance, we propose a model in which the combination of genetic alterations as SNVs and copy numbers, but importantly also the epigenetic configuration (Figure 6A) determine the characteristics of a certain subclone, which can be places in a given fitness landscape (Figure 6B). However, for a given subclone

there is a one-to-many map between the cellular memory encoded and heritable in its genome and epigenome, and the many transcriptional *phenotypes* it can produce…

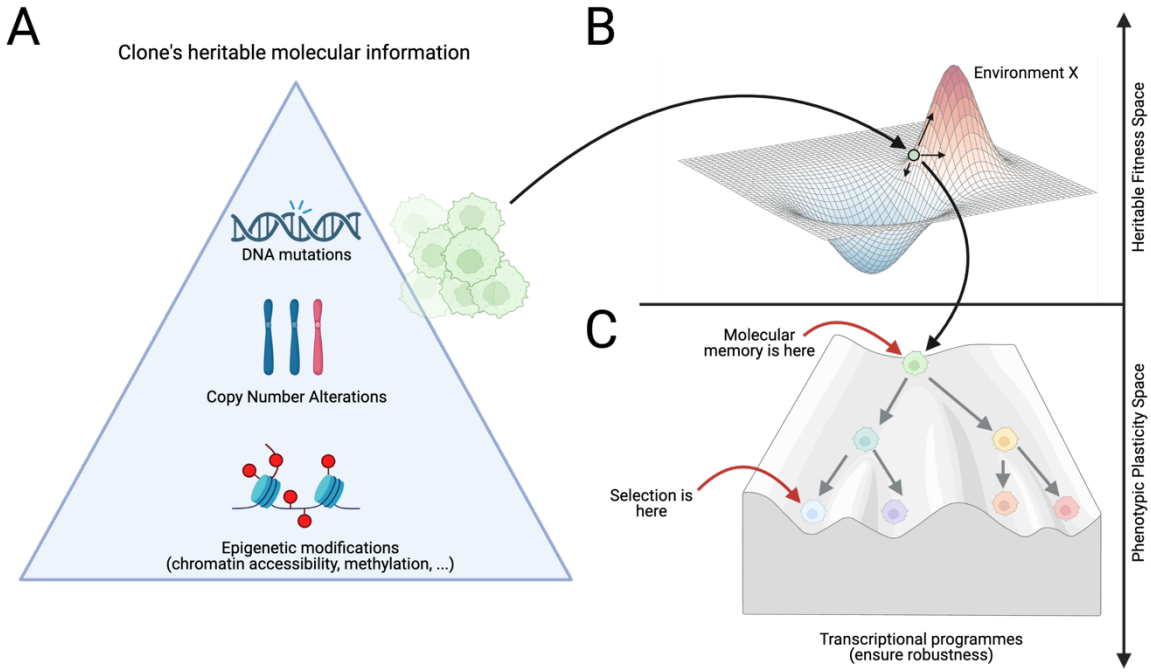



**Figure 6. Heritability and plasticity of cellular phenotypes. (A)** We propose a model in which genetic mutations and copy number alterations, together with heritable chromatin accessibility profiles determine the cellular memory of a certain clone, positioning it within a certain heritable fitness landscape **(B)**. However, the clone does not manifest as a single transcriptional phenotype, but rather as a set of transcriptional programmes that could be represented within a Waddington landscape, similarly to those that regulate development **(C)**. As Darwinian selection occurs at the level of phenotypes, it is likely that the selective pressure is at the bottom of the Waddington landscape, possibly for only a subset of the transcriptional programmes of a clone, despite the molecular memory carrying these programmes also containing as a side effect other plastic phenotypes. This may explain the persistent phenotypic heterogeneity and plasticity of cancer clones, despite the strong selective pressure of treatments that instead should select for a single fittest phenotype.

## Discussion

So far, only a limited set of genetic alterations driving cancer drug resistance has been identified. Typical examples are gatekeeper mutations driving resistance to targeted therapies, such as in EGFR-mutant lung cancer[24], or mutations in signalling pathways downstream of the targeted receptor, such as RAS/RAF mutations[2], or mutations that revert a vulnerability[2]. For endocrine therapy in breast and prostate cancers, genetic alterations in ER[25] and AR[26] respectively have been described. However, in the majority of cases these mechanisms are not dominant (i.e. subclonal) in the tumour, supporting a view of polyclonal drug resistance[27]. Moreover, although chemotherapy is the backbone of cancer treatment in most malignancies, recurrent genetic alterations conferring resistance to chemotherapy have not been documented. This lack of insight is hindering the development of more effective treatments.

# Material and methods

## Organoid culture and passaging

F-016 patient-derived organoid (PDO) was established from a liver metastasis of a colorectal tumour and it has been previously characterised as described in Vlachogiannis et al. 2018[13] Among its pathogenic mutations, *AKT1* c.49G>A_p.Glu17Lys and *AKT1* amplification was of interest for our purpose as this organoid was a good model of oncogenic addition, showing high sensitivity and a selective apoptosis to Akt inhibitors[13].

F-016 PDOs were cultured embedded in Growth Factor Reduced (GFR) Basement Membrane Matrix (Corning), hereinafter referred to as matrigel, and Advanced DMEM/F12 media (Thermo Fisher Scientific), supplemented with 1X B27 and 1X N2 supplements (Thermo Fisher Scientific), 0.01% BSA (Roche), 2mM L-Glutamine (Thermo Fisher Scientific) and 100 units/ml penicillin streptomycin (Thermo Fisher Scientific). Additionally, 12 different growth factors were used to maintain PDOs culture: 50 ng/ml EGF, 100 ng/ml Noggin, 500 ng/ml R-Spondin 1, 10 ng/ml FGF-basic, 10 ng/ml FGF-10 (all from PeproTech), 10 nM Gastrin, 10 µM Y-27632, 4 mM Nicotinamide, 5 µM SB202190 (all from Sigma-Aldrich), 100 ng/ml Wnt-3A (R&D Systems), 1 µM Prostaglandin $E_2$ and 0.5 µM A83-01 (Tocris Bioscience).

Passaging of PDOs was performed using TrypLE 1X diluted in 1mM PBS-EDTA (Thermo Fisher Scientific). In short, after media removal, PDOs in matrigel were harvested by pipetting with 1ml of TryplE1X and they were incubated for 20 min at 37ºC, with mechanical homogenisation every 5 minutes. Then, PDOs were centrifuged at 1,200 rpm for 5 min at 4ºC and washed with HBSS (Thermo Fisher Scientific). Counting and viability measurements were done using 0.4% Trypan Blue staining solution (Thermo Fisher Scientific) and the Countess 3 Automated Cell Counter (Thermo Fisher Scientific). Expected cells were pelleted again and re-seeded in matrigel.

## 3D patient-derived organoids drug screenings

To check baseline sensitivity to Akt inhibitors, initial dose response curves (DRC) were performed with the allosteric inhibitor MK-2206 and the non-allosteric one AZD5363 (capivaserib). Initially, PDOs were dissociated into single cells as the passaging procedure described above, and after an automatic counting by Countess 3 Automated Cell Counter (Thermo Fisher Scientific, 6000 cells were seeded in 30µl of matrigel in 96-well plates. The matrigel was solidified after a 20-minute incubation at 37ºC and 5% $CO_2$, and overlaid with 70 µl of complete human organoid media. 24 hours later, media was removed and replaced with 50 µl of drug-containing media at different concentrations along with DMSO as a vehicle control, and replenished every two days for three times. Finally, drug-containing media was removed and replaced with 10% CellTiter-Blue cell viability assay media (Promega) and after 3-hour incubation at 37ºC and 5% $CO_2$, readings were taken in EnVision plate reader (PerkinElmer). DRC were represented by GraphPad Prism 9 (Dotmatics).

## Patient-derived organoids barcoding

Individual cell barcoding was performed by CloneTracker XP™ 5M Barcode-3' Library in pScribe4M-RFP-Puro (Cellecta). This expressible barcode libraries enable the tracking and profiling of individual clones thanks to the integration in gDNA and barcode transcription in

RNA. To ensure the insertion of a single barcode per cell, lentiviral titration was assessed and a MOI of 0.1 was chosen along with 0.8 µg/ml polybrene, corresponding to 10% of infection. 1 million cells from the F-016 organoids were infected following manufacturer's protocol. After an over-night incubation in media suspension, cells were pelleted and resuspended in 1ml of matrigel in 6-well plates and with 2ml media containing 2.5µg/ml puromycin, as previously calculated by the puromycin killing curve. Following 10 days of puromycin selection, RFP-positive cells were checked under the microscope, and they were expanded with normal growth media. After several passages, parental population (POT) was frozen and characterised. DNA and RNA extractions and libraries sequencing (protocol below), detected around 3,500 different barcodes were in the final POT.

## Generation of drug resistant organoids

Once the POT was expanded, 75 million cells were used for the generation of resistant organoids to previously selected inhibitors. Briefly, 2,5 million cells/well were seeded in 6-well plates in 1ml of matrigel and 2ml of growth media. 24 hours later, media was collected and renewed with media containing high doses of the selected inhibitors corresponding to $IC_{90}$. MK-2206 and capivaserib were chosen as Akt inhibitors, while trametinib, as a MEK inhibitor, and KU-0063794, as an upstream mTOR inhibitor, were also selected. DMSO was always used as the vehicle control.

## Experimental design

Following above-described conditions, individual 6-well plates were used per each drug. Bottom three replicas were used for unique treatments while top three wells were used for a second treatment with drugs involved in different pathways, to perform a crossover screening. F-016 PDOs were treated for 45 days with the respective first treatment and then, the three bottom wells were harvested and mixed to be used for bulk and single-cell characterisation, as well as for barcode amplification from DNA and RNA. Remaining cells were frozen and biobanked in in FBS (Thermo Fisher Scientific), containing 10% DMSO (SigmaAldrich) for future analysis.
Top three wells were expanded without drug for two weeks (drug holidays), and then re-seeded with the same conditions as initial treatment, to be treated with the second inhibitor during additional 45 days. Lately, at the end of the treatment, cells were also expanded without drug for a couple of weeks and characterised and stored like initial time points.

As previously described, a total of 4 inhibitors were selected and a DMSO control plate was maintained until maximum confluency at day 15. For the crossover treatment, cells under Akt or mTOR inhibitors were treated with trametinib as a second treatment, while in cells initially treated with the MEK inhibitor, Akt or mTOR inhibitors were administered. To check the generation of resistant cells, sensitivity assays by dose response curves were performed at the end of each treatment.

## Parental and resistant cells characterisation by bulk and single-cell sequencing

Once treatments concluded, organoids were collected and dissociated into single-cell using the passaging procedure previously described. Half of the cells were used for bulk analysis while the other half was for single-cell experiments.

gDNA and RNA were isolated for bulk characterisation using All Prep DNA/RNA Mini Kit Qiagen (Qiagen) according to manufacturer's recommendations. It mainly consisted of whole genome library preparation from 100ng of gDNA following NEBNext Ultra II FS (New England Biolabs) recommendations, with 20 minutes of enzymatic fragmentation and 5 PCR cycles. NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs, New England Biolabs) were used. After pooling, samples were sequenced for low-pass whole genome sequencing with at least 0.1X coverage in a NovaSeq 6000 (Illumina).

Regarding single-cell approaches, a total of 19 single-cell experiments were performed from the parental (POT) and the 4 inhibitors (MK-2206, capivaserib, trametinib and KU-63794) after drug removal (under drug) and after expansion (re-growth) for the two lines of treatment. After dissociation, single cells were washed with PBS and resuspended in PBS + 0.04% BSA, filtered through a 40µm FlowMi cell strainer (Sigma) and resuspended at a concentration of 1000 cells/µl. Viability was confirmed to be >90% in all samples using 0.4% Trypan Blue dye with Countess 3 Automated Cell Counter (ThermoFisher). Those organoids harvested under drug exposure with less than 70% viability were enriched by Dead Cell Removal Kit (Miltenyi) according to manufacturer's protocol. For an estimation of 5000 cells, single cell suspensions were loaded on a Chromium Next GEM Chip G (10X Genomics) and were run in the Chromium Controller to generate single-cell gel bead-in-emulsions using the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 kit (10X Genomics). After single-cell RNA-seq libraries were prepared and the library quality was confirmed with the TapeStation D1000 Screen Tape (Agilent) and a Qubit 3.0 dsDNA HS Assay Kit (Life Technologies), samples were pooled and sequenced on an Illumina NovaSeq 6000 (Illumina) according to standard 10X Genomics' protocol.

## Barcode amplification and library preparation

Barcoded organoids at each time point before and after treatment were harvested and pelleted. To study the expressible barcodes, gDNA and RNA were isolated using All Prep DNA/RNA Mini Kit Qiagen (Qiagen) according to manufacturer's recommendations. While DNA was directly quantified by Qubit 3.0 dsDNA HS Assay Kit (Life Technologies) and stored for barcode amplification, 300ng of total RNA were taken for cDNA synthesis with SuperScript™ II Reverse Transcriptase (Invitrogen).
Contrary to a 2D culture, organoids are embedded in matrigel and although dead cells remained inside, gDNA is released to the surrounded media. Hence, to track the evolution of each cell lineage during the treatment without replating them, gDNA from apoptotic cells was collected every media change (2 days) to amplify dead cell barcodes, providing us unparalleled temporal resolution. Collected media was incubated for 2 hours at 55ºC with 20mg/ml proteinase K (Roche). Then, gDNA was purified with 3X SPRIselect beads (Beckman Coulter) followed by an 80% ethanol wash step. DNA was eluted in low TE and quantified by Qubit 3.0 dsDNA HS Assay Kit (Life Technologies).

Barcode amplification by PCR was performed using 2x Accuzyme mix (Bioline) and 10 ng of the extracted DNA/cDNA using the primers below (10µM):

Forward XP primer: 5'-ACCGAACGCAACGCACGCA-3'
Reverse XP primer: 5'-ACGACCACGACCGACCCGAACCACGA-3'

Briefly, 25µl reactions PCR steps were: 98ºC 2', and 35 cycles of 95ºC 15'', 71ºC 15'', 72ºC 10'' and a final extension of 72ºC 3'. 4µl of the 128-bp PCR product were later checked on an 1.5% agarose gel and quantified for library generation by TapeStation (Agilent). NGS libraries were prepared from 30ng of the PCR product using the NEBnext Ultra II DNA library preparation kit for Illumina (New England Biolabs) according to manufacturer's recommendations and using NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs, New England Biolabs). Libraries were quantified using Qubit 3.0 dsDNA HS Assay Kit (Life Technologies) and TapeStation D1000 Screen Tape (Agilent Genomics). Up to 384 samples were pooled for sequencing. Due to unspecific library amplification, 3ng/µl samples were pooled considering 250-340bp concentration in TapeStation. Then, pools were dried out to 50ul using a vacuum concentrator and an electrophoresis in 2.5% gel was done. 280pb band was cut and purified by Gel purification kit (Qiagen). 50µl of the eluted product was again quantified by both methods. NGS was performed at the Tumour Profiling Unit of the Institute of Cancer Research using NovaSeq 6000 (Illumina). 150 pair-end reads and 15% of PhiX was considered and approximately 5000 reads per time-point was aimed.

## Phenotypic correlation with barcodes

CloneTracker XP™ barcode enrichment was performed in single-cell gene expression libraries. The 10x Chromium 3' Reagent Kit was developed for amplification of 3'-ends of polyA+ RNAs. Due to the CloneTracker XP BC14-spacer-BC30 barcode design, located at a fixed ~150-200nt distance upstream from the poly-A+ site, the transcribed barcodes are captured in the first step of the 10x Chromium 3' protocol.

To effectively read the barcode sequence an additional PCR amplification step was needed with a barcode-specific primer that is located just upstream of the CloneTracker XP barcode (FBP1). This ensured amplification of the segment of cDNA that contains both the CloneTracker XP barcode and the cell barcode associated with the poly-A+ sequence in the scRNA-Seq protocol. In short, after cDNA generation by 10x Chromium Next GEM Single Cell 3' protocol, 75% of the adaptor ligated product was used for gene expression libraries, while 25% of amplified cDNA was taken for sequential nested PCR barcode amplification. First PCR was performed for 9 cycles using below primers:

Partial 10X Read 1 primer forward: 5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCT- 3'
Specific barcode fragment primer (FBP1) FSeqRNA-BC14-XP Reverse Primer:
5'- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGACCACCGAACGCAACGCACGCA- 3'

Second round of PCR to increase Clone Tracker XP barcoded cDNA sequences used standard 10X p7 and p5 primers for additional 6 cycles to then proceed to indexing and final library generation. After QC by TapeStation (Agilent), barcode amplicon libraries were pooled and sequenced. Then, Clone Tracker XP barcodes and gene expression were correlated thanks to 10X barcodes identifying cellular gene expression with the clone identifiers.

## Barcode bioinformatics analysis

Barcodes from bulk experiments were quantified using *BWA*[1] and *FeatureCounts* [2]. We first generated a fasta reference using the full pool of barcodes provided by Cellecta. We then aligned the amplicons using *bwa-mem* with custom parameters. The bam files obtained from bwa were then used to quantify barcode abundances using the *FeatureCounts*

implementation of the R package *Rsubread* [3]. More specifically we run the command *featureCounts* we filtered reads with mapping quality less than 30 (minMQS = 30).
For barcodes enriched from the 10x Chromium library we used STAR solo [4] with a custom reference index built on the reference fasta used for BWA.

## Barcode evolutionary modelling

We used the floating barcodes abundances to produce an estim1ate of the fitness value of each barcode. To avoid biases due to the differential efficacy in capturing the DNA across different phases of the experiment we worked with the relative abundances of each barcode and to make the calculations tractable for thousands of barcodes we assumed exponential growth. Under this assumption if call as $\omega_i$ the effective growth rate (so the difference between birth rate and death rate) for barcode i we can write its abundance at time T as:

$$N_i(T) = N_i(0)e^{\omega_i T}$$

As we were working with relative abundances it follows, if we assume independence between subclones, that the abundance $f_i(T)$ for barcode I at time T is simply:

$$f_i(T) = \frac{N_i(0)e^{\omega_i T}}{N(0)e^{\overline{\omega} T}} = f_i(0)e^{\Delta\omega T}$$

Where N(0) is the population size at time 0, $\overline{\omega}$ is the average growth rate of the population and $\Delta\omega$ id the difference between the barcoded population growth rate and the average population. From this simple model we derive that the relation between the abundance at sampling time T and time T-1 is (after taking the logarithm of both sides):

$$ln\big(f_i(T)\big) = ln\big(f_i(T-1)\big) + \Delta T \Delta\omega$$

We then calculate the relative growth rate by fitting the beta coefficients of a linear model on the log abundances using the *fastLM* function of the *RcppEigen*[28] R package.

## Single cell RNA sequencing analysis

For the analysis of single-cell RNA sequencing (scRNA-seq) data, the fastq files containing the sequential reads were aligned to the reference genome (assembly hg38) using STARsolo version 2.7.9a.

The matrices produced from the alignment were corrected for ambient RNA using the *adjustCounts* function of *SoupX*[29]. We then used *DoubletFinder*[30] to estimate the doublet status of each cell. For quality control, filters were applied for the minimum number of genes with non-zero expression and for the percentage of expression derived from mitochondrial transcripts (%mt), indicative of dying cells. Specifically, cells with %mt greater than 30%, fewer than 1000 expressed genes and putative doublet from *DoubletFinder* were filtered out.

Data was then analyzed in Seurat[31]. Counts were normalized using the *NormalizeData* function. At this point, the 3000 most variable genes were identified with the FindVariableFeatures routine, which were then scaled (*ScaleData*) and used to compute a Principal Component Analysis (*RunPCA*). The first 25 components, selected based on the amount of variance explained, were used to compute the KNN graph and Uniform Manifold Approximation (UMAP) with the functions *FindNeighbors* and *RunUMAP*.
We compute gene set scores for the Autophagy signature from Ref.[20] and the *MSigDB*[23] Hallmark of Cancer MYC targets V2 using the function *AddModuleScore*.

## Single cell ATAC + GEX analysis

10X multiomics sequencing data has been processed with cellranger-arc version . We then performed all the quality control and analysis using the R package *ArchR*[32] . After generating the arrow files, we used the *filterDoublets* and filtered out cells with *TSSEnrichment* < 6, number of fragments < 10000, putative doublets and cells without a valid UMI for both GEX and ATAC libraries. LSI for RNA and ATAC were computed using the *addIterativeLSI* method using respectively the 2500 most variable genes and the ArchR tile matrix. Motif annotations for *ChromVAR*[21] were added using the *addMotifAnnotations* function with using the *CIS-BP* dataset and scores calculated using the *addDeviationsMatrix* method. To compute correlation, we used the *correlateMatrices* function with input the Motif Matrix and the Gene Expression Matrix. Marker peaks were computed using the *getMarkerFeatures* and setting as bias terms the TSS enrichment score and the log10(number of fragments). To then computer differential Transcription Factor Motif accessibility we used the function *peakAnnoEnrichment*.
Background peaks were computed using the *addBgdPeaks* method.

## Deep Archetypal Analysis

Archetypal Analysis is a dimensionality reduction method that decomposes an input matrix as a convex combination of ideal extreme points called archetypes and it was first introduced in ref[15]. More in detail if we assume $X \in R$ to be our $N\ x\ M$ input matrix, the classical problem of archetypal analysis revolves on optimizing the following loss function:

$$||X - ZA\ ||^2\ with\ Z = XB^T$$

Where B and A are respectively $N\ x\ K$ and $K\ x\ N$ row stochastic matrices with K being the number of archetypes. We define a matrix X to be row stochastic matrix if it satisfies $\mathrm{x}_{ik} \geq 0, \sum_k \mathrm{x}_{ik} = 1$. The matrix B defines the archetype as a convex combination of real data points, while the matrix A represents the weights of each archetype in reconstructing each cell. This is intuitively equivalent to optimally embed the data in a convex polytope with K vertices (archetypes), it then follows that the archetypes define the convex hull of the polytope.

The original algorithm for finding the minima involves alternating optimization steps were we first fix B and optimize for A and then do the do the reverse. This solution is generally considered hard to scale as it needs to use all the whole dataset for each iteration ref[33].

To add the possibility of mini batch learning and to account possible non-linear effects in the archetypal composition we implemented the Deep Archetypal Network presented in ref[16]. In this case we learn both matrices using a feed-forward neural network $A, B = f_\theta(X)$. The input is then reconstructed by a specular decoder network $f_\phi(Z^*A) = \tilde{X}$. To reduce the complexity of the problem $Z^*$ is fixed and set to a standard simplex with K vertices and the condition $Z = XB^T$ is guaranteed by using an additional term in the loss, that becomes:

$$|| X - Z^*A ||^2 + || Z^* - BAZ^* ||^2$$

To fit the scRNA-seq data we first computed the intersection top 2000 most variable genes in the under-drug phase, the 1000 most variable in re-growth and the 500 most variable in the parental sample. We then computed the Z-score across all the cells with a valid barcode and used it as an input for the AA net. For the scATAC-seq from the 10x multiome, given the higher complexity we used the first 30 components of the LSI as input.

### Copy Number Analysis

To infer copy number clones from the ATAC portion of the single-cell multiome we used CONGAS+[18]. To derive a prior segmentation, we used the inferred copy number states from the corresponding low-pass DNA analysis, and we did an intersection of the segments among the 4 samples to obtain a unique profile. We then used the function *segments_selector_congas* to filter for putative multimodal segments. We run the method *fit_congas* and tested a range of clusters going from 1 to 15 for 2000 steps, a learning rate of 0.005 and a temperature parameter for the Gumbel-SoftMax of 10. We choose the best number of cluster that minimized the Integrated Completed Likelihood (ICL). The inferred copy number alteration matrix was used as an input to MEDICC2[34] with parameters *--total-copy-numbers --input-allele-columns* to build a clonal tree.

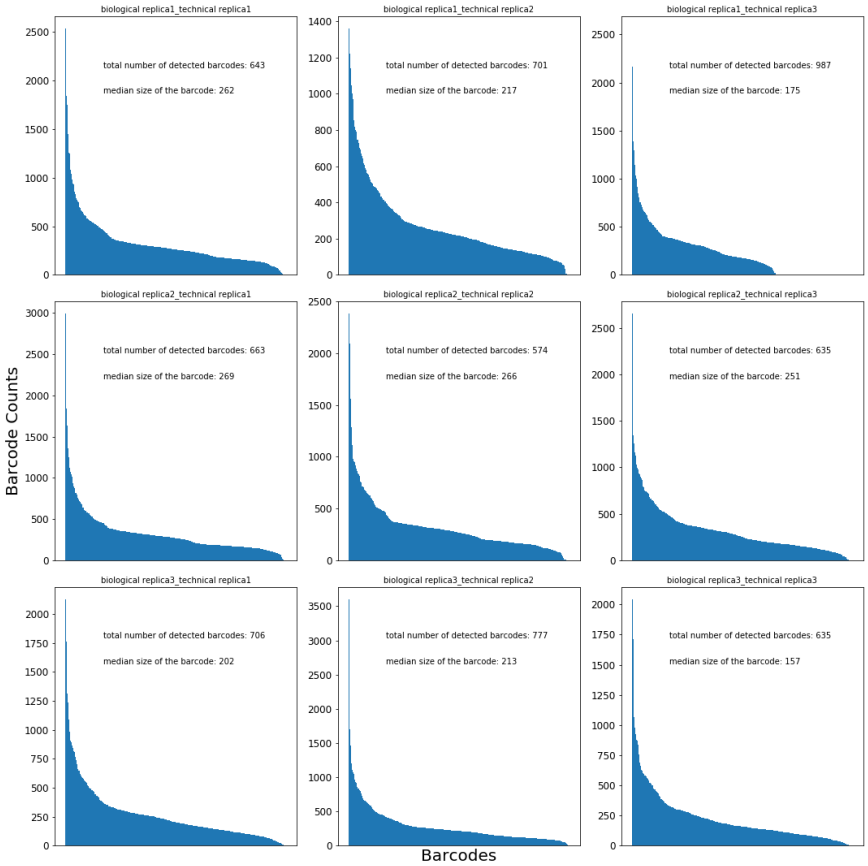
## Acknowledgments

This work was supported by the CRUK Accelerator Award (A26815). A.S. and T.G. are also supported by the Wellcome Trust (202778/B/16/Z and 202778/Z/16/Z respectively) and Cancer Research UK (A22909 to A.S., A19771 & DRCNPG-May21_100001 to T.G.). NV was supported by Cancer Research UK (A18052 and A26815), the National Institute for Health Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, the European Union FP7 (CIG 334261) and the Katherine and Douglas Longden Chair in Oncology at Imperial College London.

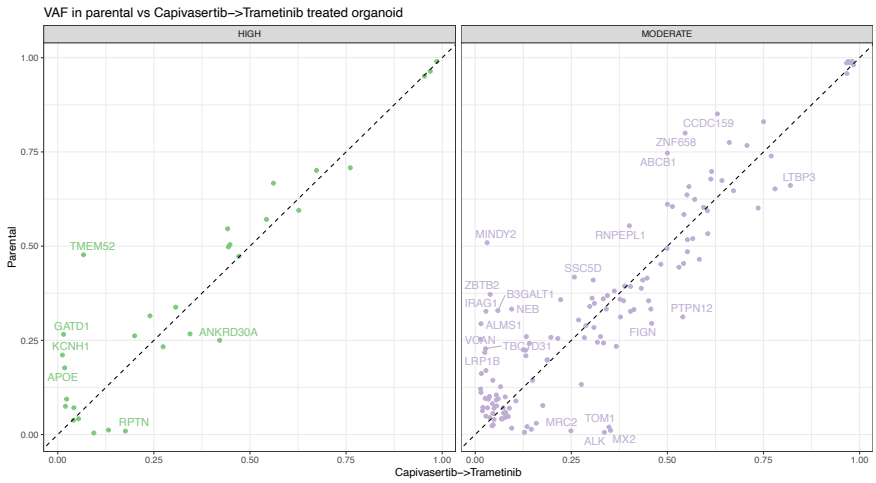
## Conflict of Interest

Between 2018 and 2022 NV received honoraria for lectures from Merck Serono, Pfizer, Bayer, Eli-Lilly and Servier; Research funding (Institutional) from Roche and BenevolentAI and has received paid consultancy fees from BenevolentAI. At the time of submission NV was a full-time employee of AstraZeneca Plc; any work done in relation to this article was undertaken before his current employment.

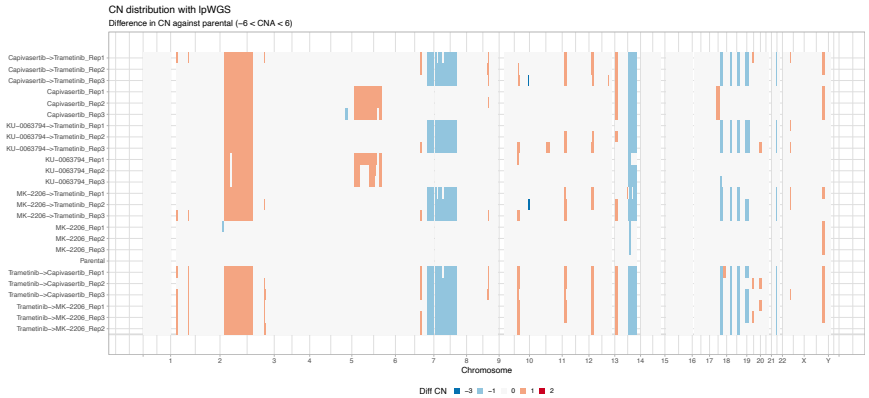
# Supplementary Figures



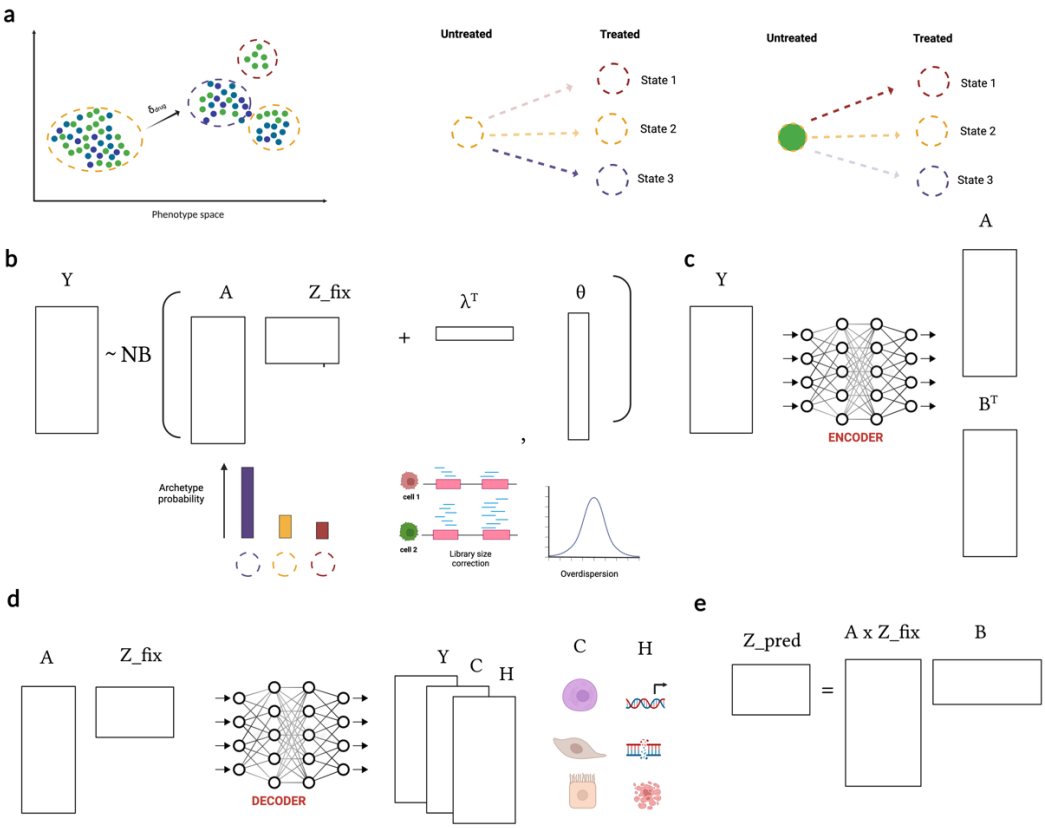

*Supplementary Figure 1*



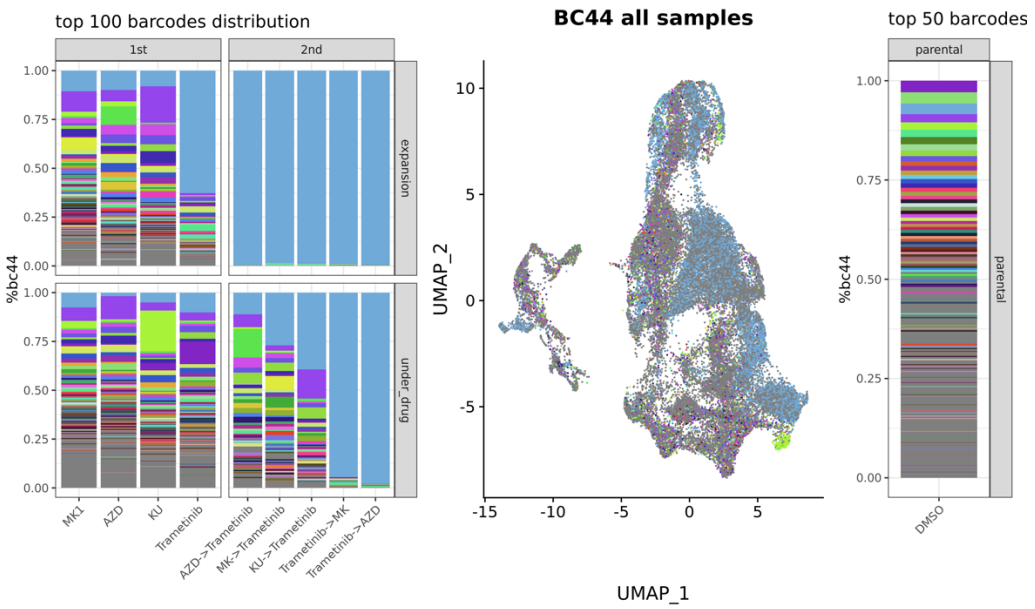

*Supplementary Figure 2*



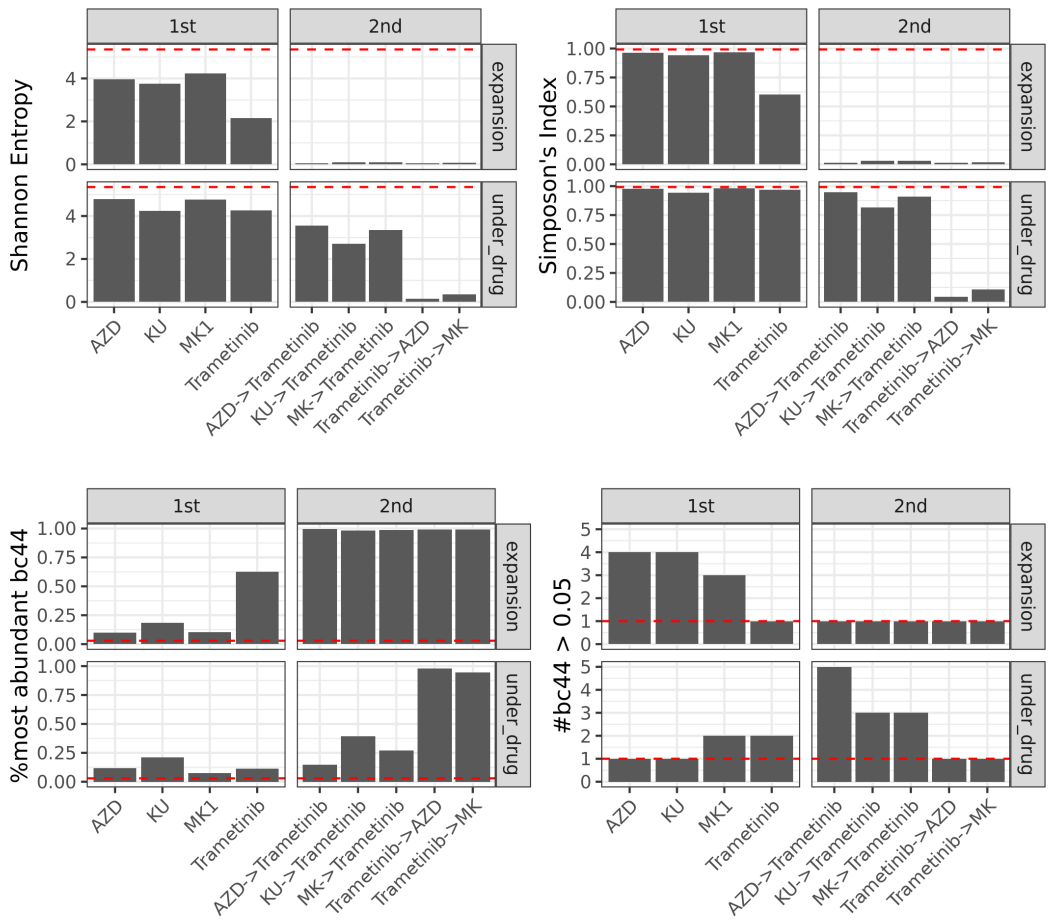

*Supplementary Figure 3*



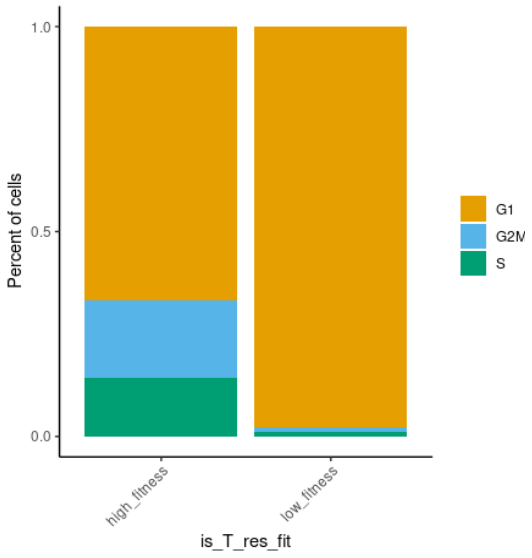

*Supplementary Figure 4*




*Supplementary Figure 3*




*Supplementary Figure 5*




*Supplementary Figure 6*

*Supplementary Figure 7*

## References


1. Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306–313. 10.1038/nature10762.

2. Vasan, N., Baselga, J., and Hyman, D.M. (2019). A view on drug resistance in cancer. Nature *575*, 299–309. 10.1038/s41586-019-1730-1.

3. Black, J.R.M., and McGranahan, N. (2021). Genetic and non-genetic clonal diversity in cancer evolution. Nature Reviews Cancer, 1–14. 10.1038/s41568-021-00336-2.

4. Bell, C.C., and Gilan, O. (2020). Principles and mechanisms of non-genetic resistance in cancer. Br J Cancer *122*, 465–472. 10.1038/s41416-019-0648-6.

5. Marine, J.-C., Dawson, S.-J., and Dawson, M.A. (2020). Non-genetic mechanisms of therapeutic resistance in cancer. Nat Rev Cancer *20*, 743–756. 10.1038/s41568-020-00302-4.

6. Heide, T., Househam, J., Cresswell, G.D., Spiteri, I., Lynn, C., Mossner, M., Kimberley, C., Fernandez-Mateos, J., Chen, B., Zapata, L., et al. (2022). The co-evolution of the genome and epigenome in colorectal cancer. Nature *611*, 733–743. 10.1038/s41586-022-05202-1.

7. Nam, A.S., Chaligne, R., and Landau, D.A. (2021). Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nat Rev Genet *22*, 3–18. 10.1038/s41576-020-0265-5.

8. Househam, J., Heide, T., Cresswell, G.D., Spiteri, I., Kimberley, C., Zapata, L., Lynn, C., James, C., Mossner, M., Fernandez-Mateos, J., et al. (2022). Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature *611*, 744–753. 10.1038/s41586-022-05311-x.

9. Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science (New York, N.Y.) *362*, eaav1898. 10.1126/science.aav1898.

10. Whiting, F.J.H., Househam, J., Baker, A.-M., Sottoriva, A., and Graham, T.A. (2023). Phenotypic noise and plasticity in cancer evolution. Trends in Cell Biology. 10.1016/j.tcb.2023.10.002.

11. Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431–435. 10.1038/nature22794.

12. Lupo, B., Sassi, F., Pinnelli, M., Galimi, F., Zanella, E.R., Vurchio, V., Migliardi, G., Gagliardi, P.A., Puliafito, A., Manganaro, D., et al. (2020). Colorectal cancer residual disease at maximal response to EGFR blockade displays a druggable Paneth cell–like phenotype. Science Translational Medicine *12*, eaax8313. 10.1126/scitranslmed.aax8313.

13. Vlachogiannis, G., Hedayat, S., Vatsiou, A., Jamin, Y., Fernández-Mateos, J., Khan, K., Lampis, A., Eason, K., Huntingford, I., Burke, R., et al. (2018). Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. Science (New York, N.Y.) *359*, 920–926. 10.1126/science.aao2774.

14. Acar, A., Nichol, D., Fernández-Mateos, J., Cresswell, G.D., Barozzi, I., Hong, S.P., Trahearn, N., Spiteri, I., Stubbs, M., Burke, R., et al. (2020). Exploiting evolutionary

steering to induce collateral drug sensitivity in cancer. Nature Communications *11*, 1–14. 10.1038/s41467-020-15596-z.

15. Cutler, A., and Breiman, L. (1994). Archetypal Analysis. Technometrics *36*, 338–347. 10.1080/00401706.1994.10485840.

16. Keller, S.M., Samarin, M., Wieser, M., and Roth, V. (2019). Deep Archetypal Analysis. In Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings (Springer-Verlag), pp. 171–185. 10.1007/978-3-030-33676-9_12.

17. Elmentaite, R., Kumasaka, N., Roberts, K., Fleming, A., Dann, E., King, H.W., Kleshchevnikov, V., Dabrowska, M., Pritchard, S., Bolt, L., et al. (2021). Cells of the human intestinal tract mapped across space and time. Nature *597*, 250–255. 10.1038/s41586-021-03852-1.

18. Milite, S., Bergamin, R., Patruno, L., Calonaci, N., and Caravagna, G. (2022). A Bayesian method to cluster single-cell RNA sequencing data using copy number alterations. Bioinformatics *38*, 2512–2518. 10.1093/bioinformatics/btac143.

19. Piazza, R., Magistroni, V., Redaelli, S., Mauri, M., Massimino, L., Sessa, A., Peronaci, M., Lalowski, M., Soliymani, R., Mezzatesta, C., et al. (2018). SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. Nat Commun *9*, 2192. 10.1038/s41467-018-04462-8.

20. Ravichandran, M., Hu, J., Cai, C., Ward, N.P., Venida, A., Foakes, C., Kuljanin, M., Yang, A., Hennessey, C.J., Yang, Y., et al. (2022). Coordinated Transcriptional and Catabolic Programs Support Iron-Dependent Adaptation to RAS–MAPK Pathway Inhibition in Pancreatic Cancer. Cancer Discovery *12*, 2198–2219. 10.1158/2159-8290.CD-22-0044.

21. Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods *14*, 975–978. 10.1038/nmeth.4401.

22. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research *47*, D607–D613. 10.1093/nar/gky1131.

23. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst *1*, 417–425. 10.1016/j.cels.2015.12.004.

24. Ko, B., Paucar, D., and Halmos, B. (2017). EGFR T790M: revealing the secrets of a gatekeeper. Lung Cancer (Auckl) *8*, 147–159. 10.2147/LCTT.S117944.

25. Lei, J.T., Anurag, M., Haricharan, S., Gou, X., and Ellis, M.J. (2019). Endocrine therapy resistance: new insights. Breast *48*, S26–S30. 10.1016/S0960-9776(19)31118-X.

26. Santer, F.R., Erb, H.H.H., and McNeill, R.V. (2015). Therapy escape mechanisms in the malignant prostate. Seminars in Cancer Biology *35*, 133–144. 10.1016/j.semcancer.2015.08.005.

27. Burrell, R.A., and Swanton, C. (2014). Tumour heterogeneity and the evolution of polyclonal drug resistance. Mol Oncol *8*, 1095–1111. 10.1016/j.molonc.2014.06.005.

28. Bates, D., and Eddelbuettel, D. (2013). Fast and Elegant Numerical Linear Algebra Using the RcppEigen Package. Journal of Statistical Software *52*, 1–24. 10.18637/jss.v052.i05.

29. Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience *9*, giaa151. 10.1093/gigascience/giaa151.

30. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst *8*, 329-337.e4. 10.1016/j.cels.2019.03.003.

31. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29. 10.1016/j.cell.2021.04.048.

32. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet *53*, 403–411. 10.1038/s41588-021-00790-6.

33. Bauckhage, C., and Thurau, C. (2009). Making Archetypal Analysis Practical. In Pattern Recognition Lecture Notes in Computer Science., J. Denzler, G. Notni, and H. Süße, eds. (Springer Berlin Heidelberg), pp. 272–281. 10.1007/978-3-642-03798-6_28.

34. Kaufmann, T.L., Petkovic, M., Watkins, T.B., Colliver, E.C., Laskina, S., Thapa, N., Minussi, D.C., Navin, N., Swanton, C., Loo, P.V., et al. (2021). MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution at bioRxiv, 10.1101/2021.02.28.433227 10.1101/2021.02.28.433227.