nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-021-00837-3

Variability within rare cell states enables multiple paths toward drug resistance

Benjamin L. Emert[1], Christopher J. Cote[2,3], Eduardo A. Torre[4], Ian P. Dardani[3], Connie L. Jiang[5], Naveen Jain[5], Sydney M. Shaffer[3,6,7] and Arjun Raj[2,3] ✉

Molecular differences between individual cells can lead to dramatic differences in cell fate, such as death versus survival of cancer cells upon drug treatment. These originating differences remain largely hidden due to difficulties in determining precisely what variable molecular features lead to which cellular fates. Thus, we developed Rewind, a methodology that combines genetic barcoding with RNA fluorescence in situ hybridization to directly capture rare cells that give rise to cellular behaviors of interest. Applying Rewind to BRAF^{V600E} melanoma, we trace drug-resistant cell fates back to single-cell gene expression differences in their drug-naive precursors (initial frequency of ~1:1,000–1:10,000 cells) and relative persistence of MAP kinase signaling soon after drug treatment. Within this rare subpopulation, we uncover a rich substructure in which molecular differences among several distinct subpopulations predict future differences in phenotypic behavior, such as proliferative capacity of distinct resistant clones after drug treatment. Our results reveal hidden, rare-cell variability that underlies a range of latent phenotypic outcomes upon drug exposure.

Individual cells—even those of ostensibly the same cell type—can differ from each other in several ways. Some of these differences can result in a 'primed' cellular state that can, in a particular context, ultimately lead to biologically distinct behaviors[1,2]. This cellular priming underlies a number of important single-cell phenomena. For instance, when anti-cancer therapeutics are applied to clonally derived cancer cells, most of the cells die; however, a small number of cells survive and proliferate, and these cells drive therapy resistance[3–6]. Yet while this phenomenon suggests the existence of rare, primed cells in the initial population, it remains unclear what distinguishes these cells at the molecular level from the rest of the population.

We and others have shown that rare cells within an isogenic population can exhibit fluctuations in expression of several genes simultaneously, which predict rare cell phenotypes and persist through multiple cell divisions[3,7]. What remains largely unknown, outside of a few cases[6,8,9], is precisely how this variability maps to distinct cellular outcomes following a treatment. As a result, several questions remain unanswered. Is molecular variability in the initial state of cells inconsequential because all cells ultimately funnel into the same cell fate? Can different cell fates arise from otherwise indistinguishable initial molecular states? Or can most differences in ultimate fate be traced back to measurable differences in the initial states of cells? What is the structure of this initial variability? These questions remain largely unanswered because of our limited ability to longitudinally track and profile cells (especially rare ones) from initial state to final fate. Longitudinal profiling by time-lapse microscopy is generally limited in its ability to interrogate large numbers of molecular features simultaneously[8,10]. Barcoding, in which cells are labeled by unique and sometimes mutable nucleic acid sequences[11–16], allows one to track and profile single cells by sequencing or imaging-based readouts[17–20]. However, a key challenge for both of these methodologies is the detection of rare cells (1:1,000 or even more rare), for which neither time-lapse nor single-cell RNA sequencing is particularly effective (new techniques aim to circumvent these limitations[21–24]). Yet many biological phenomena, such as therapy resistance in cancer cells, occur in subpopulations that are at least that rare.

Here we explicitly connect drug-resistant cell fates in melanoma to specific molecular features in rare subsets of cells in the drug-naive population. These connections reveal a rich mapping between previously hidden single-cell variability and several latent cellular behaviors. Our results suggest the existence of a large number of rare subpopulations within seemingly homogenous cells, each with potentially distinct biological behaviors, and set out a path to discover biologically consequential axes of variability.

Results

Rewind enables retrospective identification of rare cell populations. Therapy resistance in cancer provides an excellent system in which to map out the connections between rare cell states and fates. In this context, fates refer to cells that proliferate when treated with targeted therapies, and the states are the molecular profiles of drug-naive cells that will ultimately lead to these resistant fates. These variable profiles can appear even in clonally derived lines and have a non-genetic basis[3–6]. In this study, we focused on BRAF^{V600E}-mutated melanoma, in which we have previously demonstrated that there is a rare, transient subpopulation composed of cells (~1:2,000) that are 'primed' to survive treatment by the targeted therapy vemurafenib[7,25]. These rare, primed cells often express

[1]Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [2]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA, USA. [4]Biochemistry and Molecular Biophysics Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [5]Genetics and Epigenetics, Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [6]Department of Pathology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ✉e-mail: arjunrajlab@gmail.com

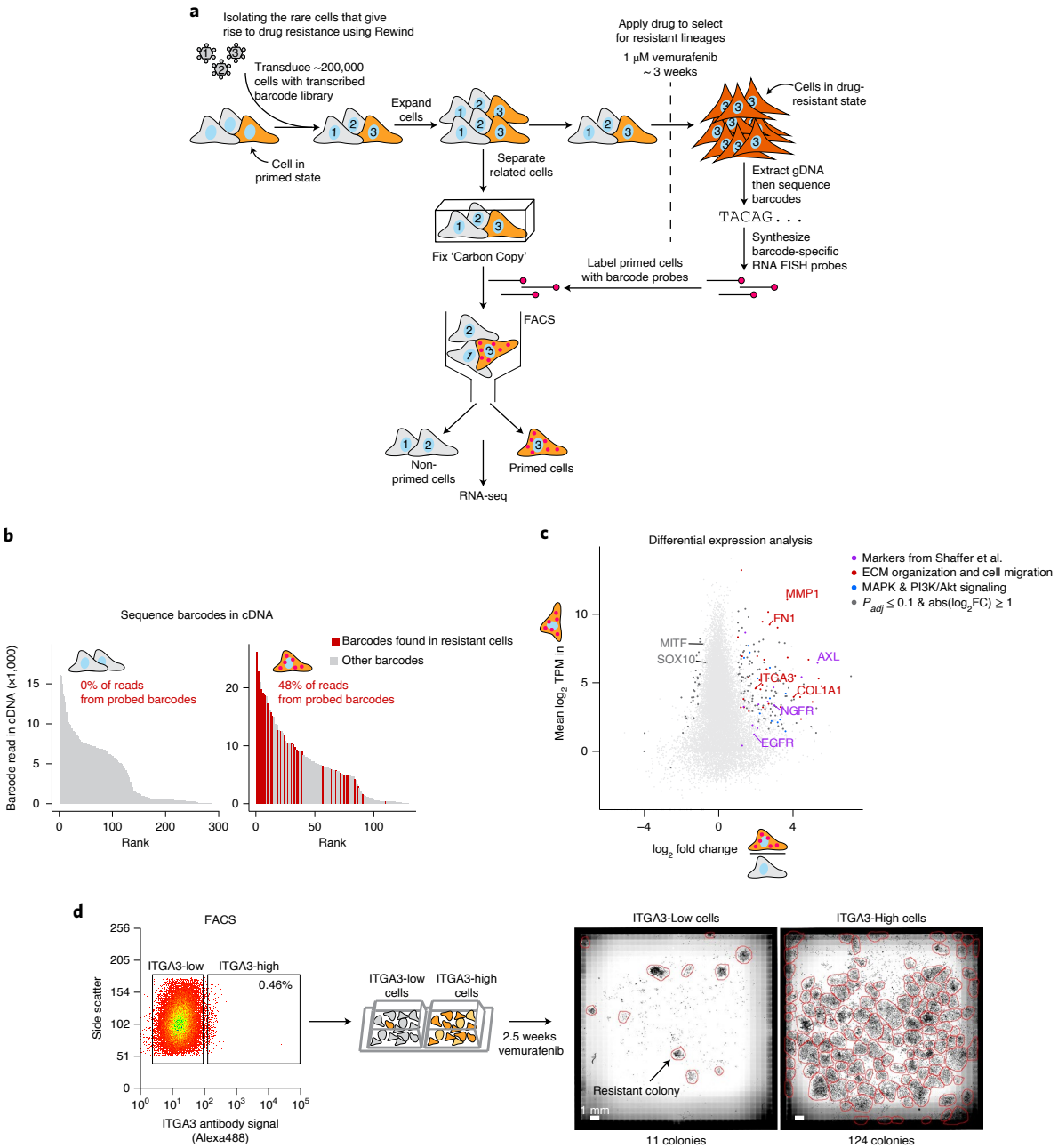


Fig. 1 | Rewind identifies rare cell states giving rise to vemurafenib-resistant colonies. a, Schematic of Rewind approach for isolating the initial primed WM989 A6-G3 melanoma cells that ultimately give rise to vemurafenib-resistant colonies. For the experiment shown, we transduced ~200,000 WM989 A6-G3 cells at an MOI of ~1.0 with our Rewind barcode library. After 11 d (~4 population doublings), we divided the culture in two, fixing half in suspension as a Carbon Copy and treating the other half with 1 μM vemurafenib to select for resistant cells. After 3 weeks in vemurafenib, we extracted gDNA from the resistant cells that remained and identified their Rewind barcodes by targeted sequencing. We then designed RNA FISH probes targeting 60 of these barcodes and used these probes to specifically label cells primed to become resistant from our Carbon Copy. We then sorted these cells out from the population, extracted cellular RNA and performed RNA sequencing. **b**, To assess the sensitivity and specificity of the Rewind experiment in **a**, we performed targeted sequencing to identify barcodes from cDNA generated during RNA sequencing library preparation. Bar graphs show the abundance (*y* axis) and rank (*x* axis) of each sequenced barcode (≥5 normalized reads). Red bars correspond to barcodes targeted by our probe set, and gray bars correspond to 'off-target' barcode sequences. The inset shows the percent of barcode sequencing reads that match a probe-targeted barcode. These data correspond to one of two replicates. **c**, We performed differential expression analysis using DESeq2 of primed versus non-primed sorted cells. Shown is the mean expression level ($\log_2$(TPM)) for protein-coding genes in primed cells (*y* axis) and $\log_2$ fold change in expression estimated using DESeq2 (*x* axis) compared to non-primed cells. Colors indicate differentially expressed genes related to ECM organization and cell migration (red), MAPK and PI3K/Akt signaling pathways (blue) and previously identified resistance markers[6] (purple). Genes were assigned to categories based on a consensus of KEGG pathway and Gene Ontology enrichment analyses (see Methods for details). **d**, We selected the most differentially expressed gene encoding a cell surface protein (*ITGA3*) to validate as a predictive marker of vemurafenib resistance in WM989 A6-G3. After staining cells with a fluorescently labeled antibody targeting ITGA3, we sorted the brightest 0.5% (ITGA3-High) and remaining (ITGA3-Low) populations and then treated both with 1 μM vemurafenib. After approximately 18 d, we fixed the cells, stained nuclei with DAPI and then imaged the entire wells to quantify the number of resistant colonies and cells. The data correspond to one of three biological replicates (see Supplementary Fig. 4 for additional replicates).

higher levels of certain receptor tyrosine kinases (such as *EGFR*, *NGFR* and *AXL*) and lower levels of melanocyte-determining transcription factors (*SOX10* and *MITF*) than the rest of the cells in the population. However, these markers are highly imperfect, with many positive cells being non-resistant and many negative cells being resistant, leaving open the question as to what markers specifically mark the primed state.

The primary technical challenge for studying rare cell processes like drug resistance is the rarity of the cells of interest. Current techniques for retrospective identification require profiling of the entire initial population and then post facto determining which profiles correspond to cells of interest[17,18]. We developed an alternative methodology, dubbed Rewind, to retrospectively isolate or identify rare cell populations of interest for downstream characterization. Rewind works by using a lentiviral library of transcribed barcodes, in which the barcode sequence is incorporated into the 3′ untranslated region of green fluorescent protein (GFP) messenger RNA (mRNA) (Fig. 1a and Supplementary Fig. 1a). After labeling cells with these barcodes, we allowed the cells to divide for a few divisions and then separated the population into two equal groups ('twins') such that most barcoded lineages (>90%) were present in each group (see Methods for discussion and empirical simulations). One group we fix in time as a 'Carbon Copy' of the cells in their initial state, and to the other we apply the treatment to see which cells undergo the rare behavior of interest (for example, becoming resistant to drug). After selecting the cells that undergo the rare behavior, we sequence their DNA to identify their barcodes, and then we use those barcodes to identify their 'twins' in the Carbon Copy by fluorescently labeling the RNA transcribed from those specific barcodes using RNA in situ hybridization techniques (Supplementary Fig. 1b,c,f,h). We verified that the barcode library was sufficiently diverse to label 100,000s of cells with over 99% receiving unique barcodes, thus minimizing spurious identification (see Methods and Supplementary Fig. 2 for experimental details and calculations). Once isolated, we can molecularly profile the Carbon Copy twins to determine what is different about their initial state that led to their distinct fate. Altogether, the Rewind methodology enables the retrospective uncovering of primed cell states that lead to rare cell behaviors.

A critical feature of these rare primed cell states is that they are transient, meaning that cells can fluctuate both into and out of the primed state[3,6]. A key question that is relevant to the ability of Rewind to profile primed cells is whether these cells maintain ('remember') their primed state through several cell divisions. (Memory would be required for the profile of cells isolated from the Carbon Copy to reflect those of their twins that received treatment with vemurafenib.) To empirically test for the existence of such memory, we let a barcoded WM989 A6-G3 culture double 4–5 times, split the culture in two and then separately treated both halves of the population with vemurafenib. We found a large overlap in the barcodes between the two halves, demonstrating that the primed state is maintained for several divisions and that there is sufficient memory in the system for Rewind to effectively profile the primed state (Supplementary Fig. 3).

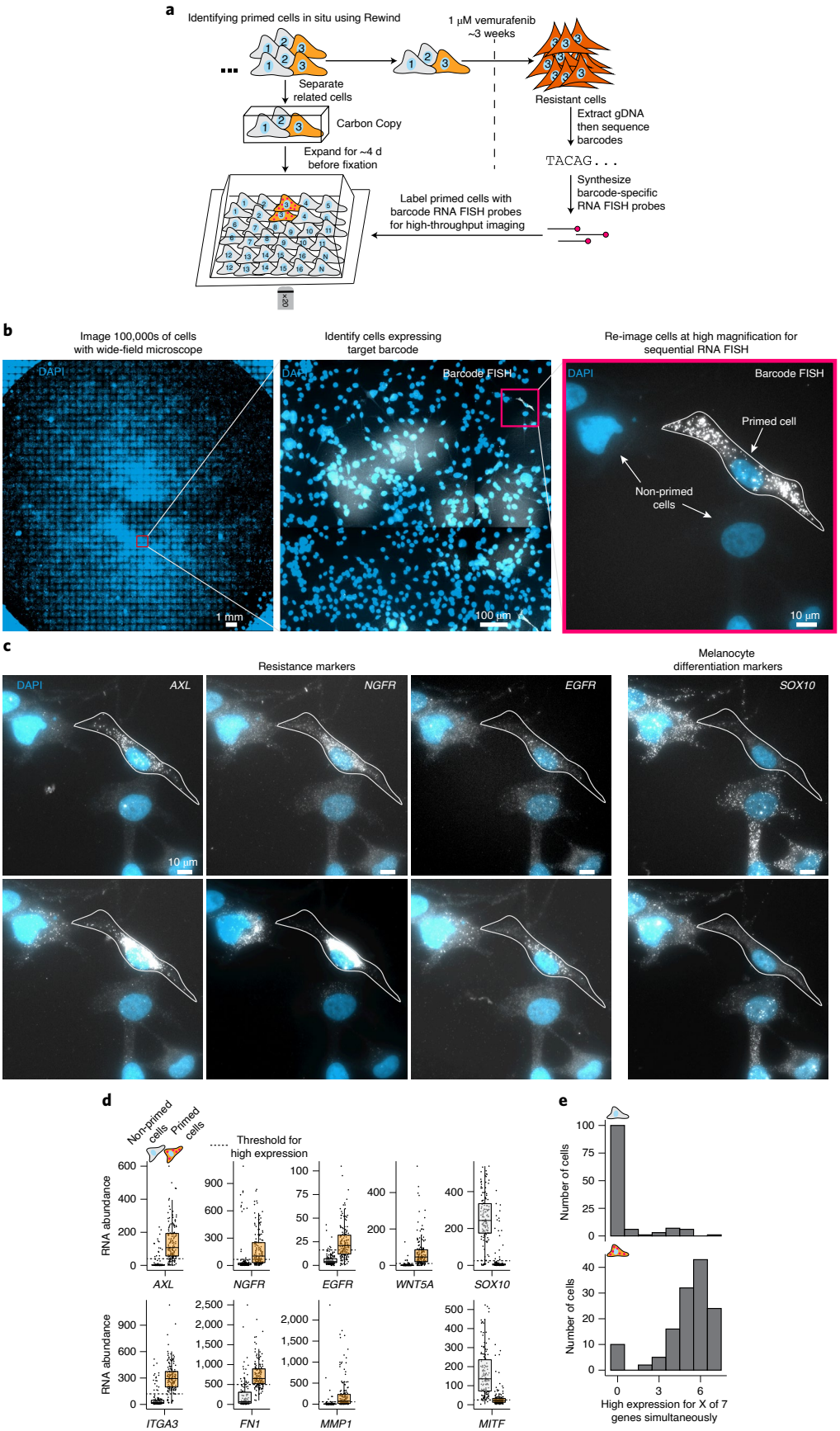
Tracing vemurafenib-resistant melanoma cells back to their rare, drug-naive precursors for gene expression profiling. We applied the Rewind approach to isolate the rare WM989 A6-G3 cells primed for vemurafenib resistance by fluorescence-activated cell sorting (FACS), after which we profiled these primed drug-naive cells by RNA sequencing (Fig. 1 and Supplementary Fig. 4a). Upon sequencing barcodes from complementary DNA (cDNA), we found that ~48% of reads in the sorted primed subpopulation contained probe-targeted barcodes matching those identified in vemurafenib-resistant colonies (versus 0% in the non-primed subpopulation), reflecting an estimated ~1,600-fold enrichment over the baseline frequency of these barcodes in the Carbon Copy (~0.03%; Fig. 1b). (We suspect that the proportion of on-target cells isolated here is lower than in our pilot experiments (Supplementary Fig. 1b,c) due to the lower prevalence of the targeted cells.) Having confirmed that FACS enriched for primed cells, we then looked for differentially expressed genes compared to non-primed cells. Consistent with previous research from our lab and others, we found that primed cells sorted from the Carbon Copy expressed greater than two-fold higher levels of the receptor tyrosine kinases *AXL*, *EGFR* and *NGFR*, as well as lower levels of the melanocyte transcription factors *SOX10* and *MITF* (Supplementary Fig. 4c)[6,26]. Beyond these known markers, the transcriptome profile provided by Rewind enabled us to identify nearly 200 new marker genes whose expression was significantly altered in primed cells. Among these genes, we found a significant enrichment for genes associated with cell adhesion, extracellular matrix (ECM) organization and cell migration (Fig. 1c, Supplementary Fig. 4d and Supplementary Table 6). Longitudinal tracking of primed cells revealed that the expression of most priming marker genes either stayed the same or increased during the acquisition of stable resistance over 3 weeks in vemurafenib treatment, and an additional ~2,800 genes showed a greater than two-fold change in expression during this period (Supplementary Fig. 5 and Supplementary Table 7). Thus, most of the genes that are upregulated in resistant cells are not the genes whose expression marks the primed state, thus motivating the use of Rewind to identify these markers.

Many of these markers have not previously been implicated in cellular priming for vemurafenib resistance and hence represent potentially novel single-cell biomarkers of resistance. An example was *ITGA3*, which was the most differentially expressed cell surface marker identified by Rewind. To verify that it marked primed cells, we prospectively sorted drug-naive WM989 A6-G3 cells expressing high levels of *ITGA3*. These cells gave rise to ten-fold more resistant colonies upon exposure to vemurafenib, confirming that ITGA3 is a marker (Fig. 1d and Supplementary Fig. 4e–h). We also used Rewind to identify markers in another melanoma line, WM983b E9-C6, in which markers of the cells primed for resistance were unknown, revealing and validating that *AXL* was a marker

Fig. 2 | A coordinated primed cell state characterized by high expression of multiple markers gives rise to vemurafenib resistance in WM989 A6-G3 cells. a, We performed Rewind with image-based profiling to identify WM989 A6-G3 cells primed to become vemurafenib resistant in situ and measure gene expression in individual cells using single-molecule RNA FISH. We expanded barcoded cells for ~4 population doublings before dividing the cells into the Carbon Copy or the drug-treated half. **b**, **c**, To identify the rare primed cells, we first imaged Carbon Copies at ×20 magnification and identified primed cells labeled with our barcode RNA FISH probes using a combination of automated image analysis and manual image review. Once identified, we returned to these cells ($n=162$) for re-imaging at high magnification (×60) and quantification of marker gene expression using single-molecule RNA FISH. We additionally imaged multiple randomly selected positions in each well to quantify marker gene expression in 'non-primed' cells ($n=135$). **d**, Quantification of single-cell gene expression in primed and non-primed cell subpopulations. Each point corresponds to an individual cell. We set thresholds for high marker expression based on the observed expression distribution in non-primed cells (see Methods and Supplementary Fig. 7 for details). **e**, Frequency of cells expressing high levels (beyond the thresholds shown in **d**) of 1, 2,…7 markers (out of a total of seven measured) simultaneously in primed and non-primed cell populations. The number of cells from each subpopulation with data for all seven markers are indicated above each histogram. These data correspond to one of two biological replicates (see Supplementary Fig. 7 for the additional replicate).

(Supplementary Fig. 6). Together, these results demonstrate that there are large sets of genes that exhibit rare-cell fluctuations that can lead to drug resistance.

Individual primed cells are marked by coordinated expression of multiple resistance markers prior to vemurafenib treatment. While isolating rare cells that express high levels of these markers



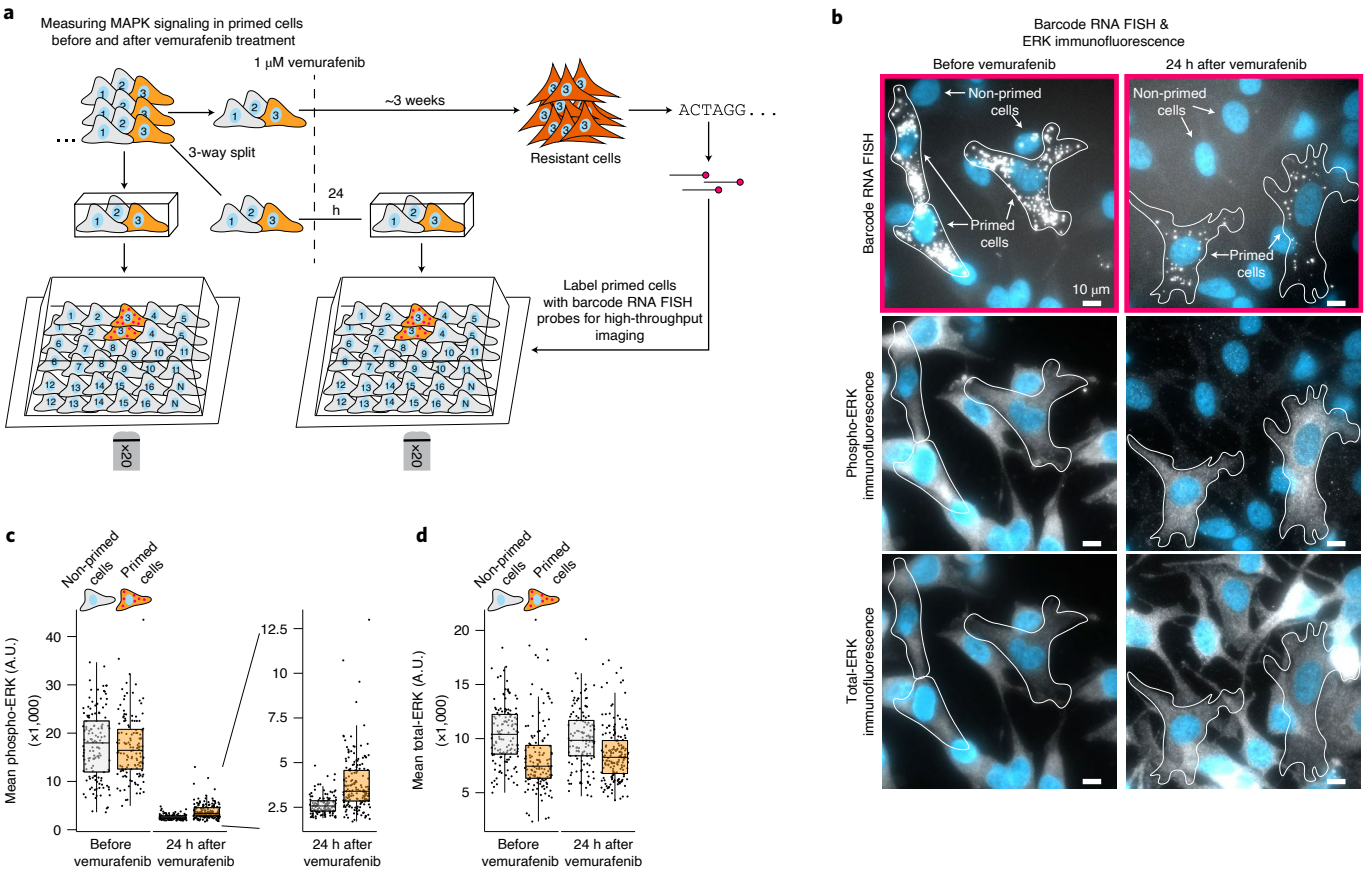


Fig. 3 | Resistance to vemurafenib is associated with single-cell variability in phosphorylated ERK levels 24 h after but not before treatment. a, We used Rewind to quantify doubly phosphorylated pERK (p44/p42) and total ERK levels in primed cells before and 24 h after vemurafenib treatment. To quantify ERK levels over time, we expanded barcoded cells for ~4 population doublings and then plated two Carbon Copies and fixed one 24 h after vemurafenib treatment and the other before treatment. As before, we used barcode RNA FISH probes to identify primed cells in both Carbon Copies and then measured single-cell levels of total ERK and pERK by immunofluorescence ($n = 135$ cells without vemurafenib treatment and $n = 173$ cells with vemurafenib treatment). We additionally imaged multiple randomly selected positions in each well to quantify total ERK and pERK in non-primed cells ($n = 133$ cells without vemurafenib treatment and $n = 125$ cells with vemurafenib treatment). **b**, Barcode RNA FISH and ERK immunofluorescence images of primed cells identified in Carbon Copies fixed before vemurafenib treatment (left) and 24 h after treatment (right). **c**, **d**, Quantification of average pERK and average total ERK immunofluorescence intensity in primed cells and non-primed cells. Each point corresponds to an individual cell. These data correspond to one of two biological replicates (see Supplementary Fig. 8 for the additional replicate). A.U., arbitrary units.

enriched for cells that could become drug resistant, we also observed that most cells that expressed any one marker still died when faced with drug. Thus, there was no one factor whose expression precisely marked the cells that were primed for drug resistance. These facts suggest that the cellular fluctuations that lead to a cell becoming primed for drug resistance may be complex and potentially marked by the fluctuations of several genes in tandem. Indeed, the lack of knowledge of the precise nature of the mapping between fluctuations and outcomes leaves open a rich set of possibilities. In principle, fluctuations of genes associated with a particular rare behavior need not be independent of each other but might take on many correlation structures and substructures, with sets of genes potentially co-fluctuating or anti-fluctuating to demarcate specific subpopulations within the overall rare cell population. A parallel question is whether these different subpopulations all funnel to the same drug-resistant outcome: it is possible that these new axes of variability might represent fluctuations that lead primed cells to adopt phenotypically distinct cellular fates after, say, the addition of drug. Rewind allowed us to look for these new subpopulations.

We first attempted to resolve the question of why most cells that expressed any one particular marker actually did not become resistant to drug. We hypothesized that simultaneous co-expression of multiple markers might more accurately and specifically identify the exact cells that are primed to be resistant. To look for evidence of such structured fluctuations, we used Rewind in combination with RNA imaging to transcriptionally profile primed cells with single-molecule resolution (Fig. 2a,b). In this manner, we located 162 primed cells in situ within a total of ~750,000 cells scanned in our Carbon Copy, which we then probed for expression of nine genes by single-molecule RNA fluorescence in situ hybridization (FISH) (Methods). These cells showed substantially higher expression of *AXL*, *EGFR*, *NGFR*, *WNT5A*, *ITGA3*, *MMP1* and *FN1* and lower expression of *SOX10* and *MITF* than randomly selected cells, consistent with our earlier results from RNA sequencing (Fig. 2c,d). Overall differences in expression capacity were unlikely to explain the increased expression of marker genes in primed cells (Supplementary Fig. 4b, Supplementary Fig. 4h and Supplementary Fig. 7e). Moreover, cells primed for resistance were far more likely to co-express pairs of markers (odds ratios ranging from ~1.5 to ≥58; Supplementary Fig. 7), and ~87% of cells expressed high levels of ≥4 of 7 marker genes simultaneously, in stark contrast to cells not expressing resistant barcodes (Fig. 2e and Supplementary Fig. 7). This apparent coordination suggests that the cell-to-cell differences that lead to distinct cell fates after drug treatment are a consequence

of the coordinated fluctuations of several factors simultaneously, as opposed to sporadic fluctuations of individual genes[7].

Primed melanoma cells are marked by higher levels of phosphorylated ERK shortly after, but not before, vemurafenib treatment. A possible mechanism for how these primed cells survive drug treatment is that the observed increases in expression of multiple receptor tyrosine kinases and their cognate ligands lead to differences in MAPK pathway activation. To address this hypothesis, we measured doubly phosphorylated ERK (pERK) levels in primed and non-primed cells by immunofluorescence (Fig. 3 and Supplementary Fig. 8). We found similar levels of pERK in primed and non-primed cells in Carbon Copies fixed before vemurafenib treatment (Fig. 3b,c and Supplementary Fig. 8a–d). However, in Carbon Copies that underwent vemurafenib treatment for 24 h, we found that primed cells had residual levels of pERK that were on average 40% higher than the rest of the population, with some primed cells having levels nearly five-fold higher than non-primed cells (within the range of untreated cells; Fig. 3b,c and Supplementary Fig. 8a–d). We also observed that, within individual clusters of closely related primed cells, not all cells contained higher levels of pERK, which might reflect pulsatile changes in pERK as documented elsewhere (Supplementary Fig. 8e)[27]. In contrast, single-cell levels of total ERK levels were modestly lower in primed cells compared to non-primed cells, both before and after vemurafenib treatment (Fig. 3d and Supplementary Fig. 8b). These results suggest that primed cells are able to maintain residual MAPK signaling following vemurafenib treatment that may allow them to continue proliferating in the face of drug.

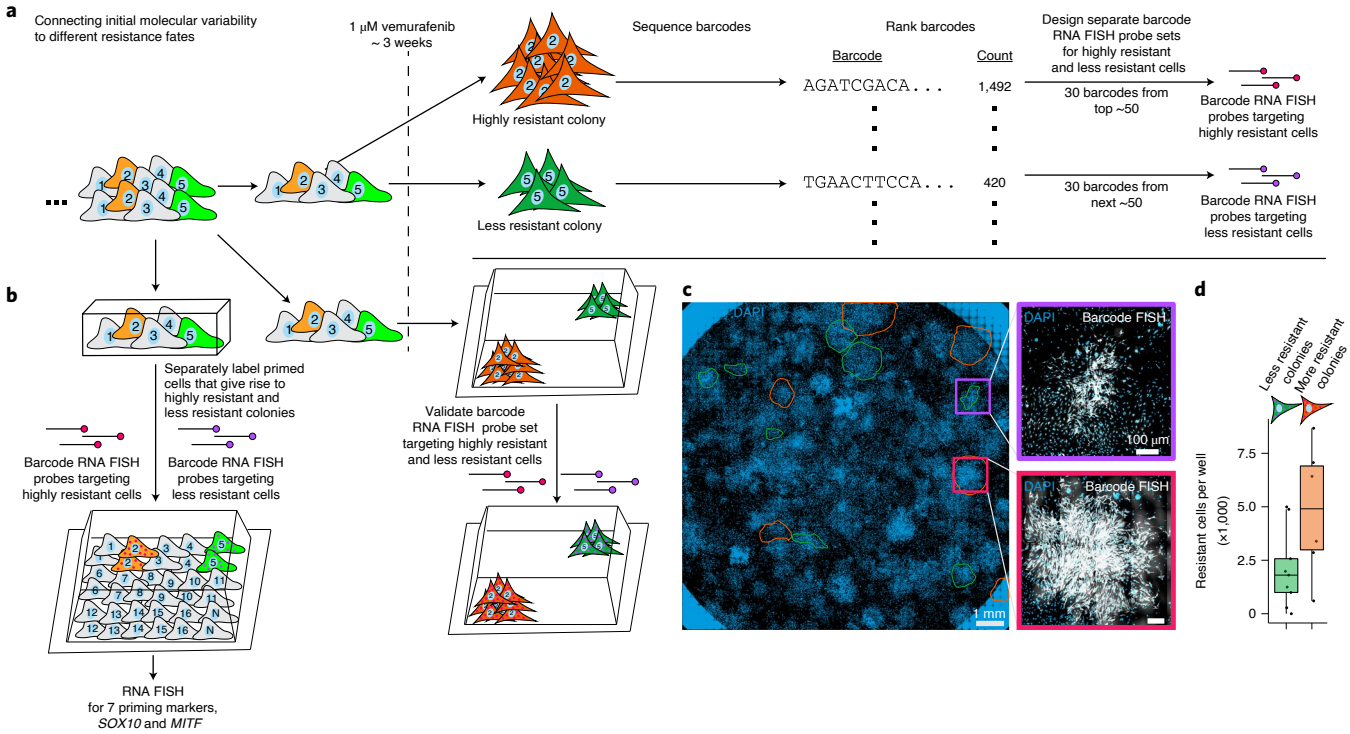
Distinct drug-resistant fates can be traced back to molecular differences within the primed subpopulation. Although these results showed an overall coordination between the different marker genes in primed cells, there were considerable differences in the degree of co-expression among these marker genes in single cells (Supplementary Fig. 7c,d,h,i). These differences suggest the possibility that the expression of specific subsets of genes might delineate specific subpopulations within the overall rare primed population that could in principle have different fates. Evidence for different fates comes from inspection: it was visually clear that different colonies of vemurafenib-resistant cells can show dramatic differences in basic properties, such as the number of cells in the colony. We wondered whether tracing back these differences in fate with Rewind could reveal the molecular profiles that distinguish subsets of the initial primed cell subpopulation with distinct potential. We applied Rewind in the WM989 A6-G3 cell line as before but used the number of barcode reads in the resistant population as a proxy for the number of resistant cells carrying a given barcode (Fig. 4a,b). We then designed RNA FISH probes that distinguished 30 of the most abundant barcodes (that is, 'highly resistant', meaning many resistant cells) from 30 barcodes in the next tier of abundance (that is, 'less resistant'; see Fig. 4a,c,d and Supplementary Fig. 9 for probe set validation). We used these probes to identify their twin cells in a Carbon Copy fixed prior to vemurafenib treatment (Fig. 4b).

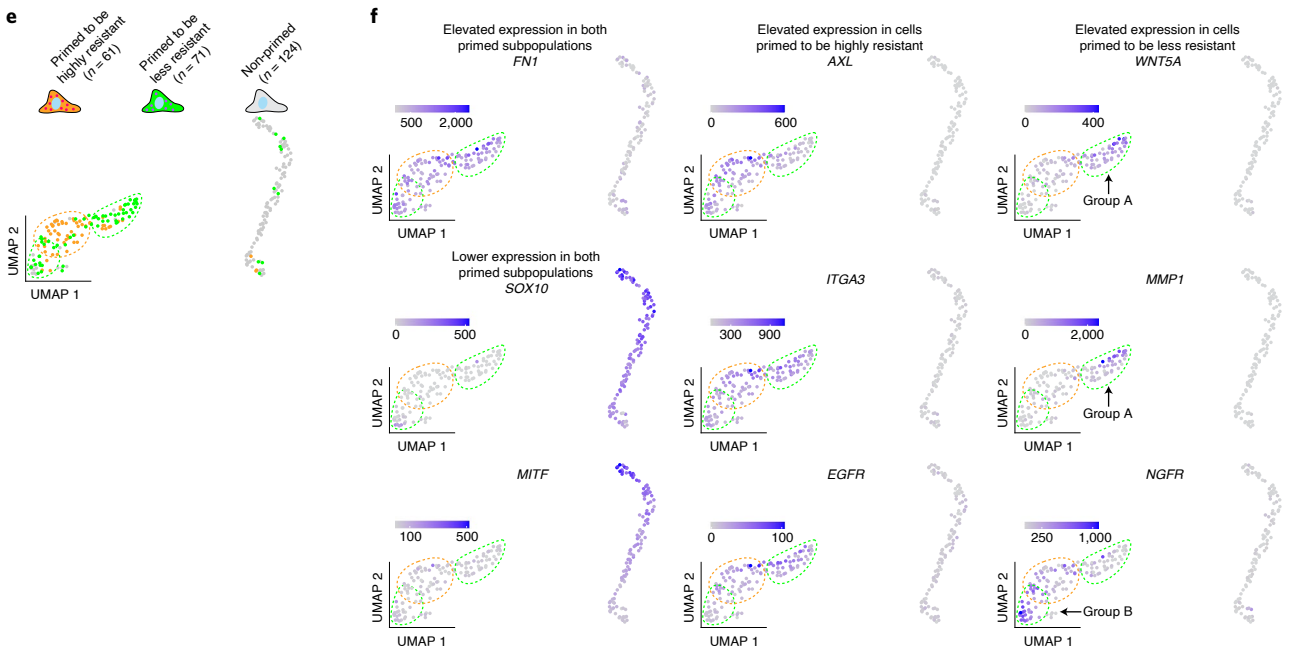
To find transcriptional profiles that predict whether cells are primed to become either highly resistant or less resistant, we measured transcript abundances in individual primed cells by RNA FISH for nine genes, including seven priming markers, *MITF* and *SOX10*. We used the dimensional reduction technique uniform manifold approximation and projection (UMAP) to visualize differences between cells based on expression levels. We then marked individual cells in this visualization based on their ultimate fate as determined by the barcode RNA FISH signal (primed to become highly versus less resistant versus non-primed). We found that non-primed cells clearly separated from all the primed cells, and that, within the primed cells, the highly resistant primed cells grouped together, whereas the less resistant cells formed two distinct groups (Fig. 4e,f). These groupings were also apparent in hierarchical clustering of the single-cell gene expression data, with cluster assignment of each cell roughly corresponding to its resistance fate, suggesting a clear distinction between the groupings (Supplementary Fig. 10c,e).

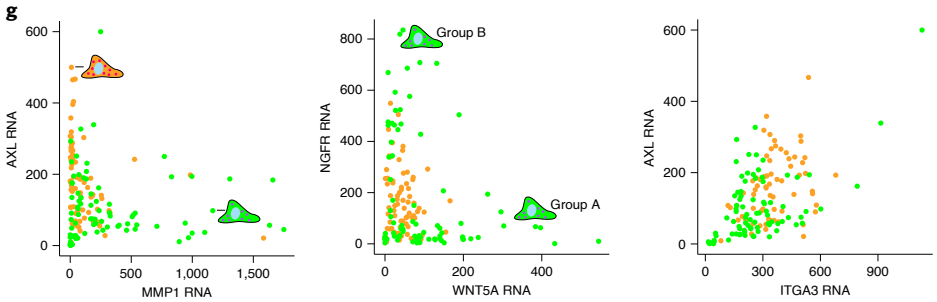
We then asked how expression levels of particular genes corresponded to these groupings. As expected, most (>80%) of the primed cells had markedly decreased levels of both *SOX10* and *MITF* (Fig. 4f, Supplementary Fig. 7 and Supplementary Fig. 10c). We also found that almost all primed cells had increased levels of *FN1* (>98%), thus suggesting that *FN1* is a 'pan' marker of cells primed for vemurafenib resistance (Fig. 4f, Supplementary Fig. 7 and Supplementary Fig. 10c). Co-expression of *AXL*, *ITGA3* and *EGFR* marked cells primed to become highly resistant, but, individually, these genes were also expressed in subsets of cells primed to become less resistant (Fig. 4f and Supplementary Fig. 10c). These subsets could also be distinguished by expression of *WNT5A*, *MMP1* and *NGFR*, with one group (group A) expressing the highest levels of *WNT5A* and *MMP1* and the other (group B) expressing the highest levels of *NGFR* (*NGFR* also had intermediate levels of expression in the cells primed to be highly resistant; Fig. 4f). In addition, quantitative comparison of expression levels between pairs of markers showed that the expression of, for example, *AXL* versus *MMP1* fell along two separate axes of variability (Fig. 4g). Together, these analyses suggest that multiple classes of primed cells with different expression patterns give rise to resistant colonies with different phenotypes.

Fig. 4 | Variation in gene expression among primed cells is associated with differences in resistant cell fate. a, We performed Rewind in WM989 A6-G3 cells and identified barcode sequences enriched in resistant colonies after vemurafenib treatment. We ranked these barcodes by abundance as a proxy for ranking the number of resistant cells carrying each specific barcode. We then designed separate RNA FISH probe sets targeting barcodes from the ~50 most abundant resistant clones ('highly resistant cells') and barcodes targeting the next ~50 resistant clones ('less resistant cells'). Each probe set contained probes targeting 30 distinct barcodes. **b**, We used these separate probe sets to identify corresponding primed cells in the Carbon Copy fixed prior to vemurafenib treatment and then performed sequential rounds of RNA FISH to measure single-cell expression of nine genes. We additionally imaged multiple randomly selected positions to quantify gene expression in non-primed cells. These data are the same as used in Fig. 2, here analyzed using information on which probe set labeled each cell. **c**, **d**, To check whether the separate probe sets label barcode RNA corresponding to distinct resistant fates, we labeled resistant colonies derived from the same population of cells and then quantified the number of resistant cells labeled with each probe set. The number of colonies labeled with each probe set and the average number of cells per colony are shown in Supplementary Fig. 9. These data correspond to one biological replicate. **e**, Using the RNA FISH data from the Carbon Copy in **b**, we applied the UMAP algorithm to the first five principal components to visualize differences in gene expression between primed cells ($n=132$) and non-primed cells ($n=124$). We then colored each cell by its predicted fate based on its barcode. To orient the reader, we circled the largest group of primed cells that give rise to highly resistant colonies in orange and the two separate groups of primed cells that give rise to less resistant colonies in green. **f**, Maintaining the organization provided by UMAP, we colored each cell by its expression of each of the nine genes measured. As noted in the text, ≥98% of primed cells had levels of FN1 RNA that were three-fold higher than the median observed in non-primed cells, and ≥80% of primed cells had levels of SOX10 and MITF RNA that were ≤1/3 of the median levels observed in non-primed cells. **g**, Scatter plots show the single-cell expression for pairs of markers that distinguished the groupings shown in **d**.

Although our labeling scheme did not discriminate between different primed cells that ended up with the same fate, in these imaging data, we were able to use spatial proximity of barcode-positive cells to infer that neighboring barcode-positive cells were likely derived from the same initial cell and therefore belong to a unique subclone (Supplementary Fig. 10b). We could then use the single-cell gene







expression levels to further determine which primed-cell class these cells belonged to and ask whether there were any signs of switching between primed cell classes (including reversion to the non-primed state) (Supplementary Fig. 10). In nearly half of the subclones (11 of 24), all cells fell into a single primed-cell class. Moreover, for most (7 of 13 remaining) subclones containing a mix of cell states, only one cell within the subclone was classified as a separate class (Supplementary Fig. 10d, right). These data suggest that primed cells can transition between states, and these transitions occur on a relatively slow time scale (potentially once per 4 d or ~2–3 cell divisions—slow compared to most expression fluctuations), consistent with recent work quantifying the transcriptional memory of several primed-cell marker genes[7].

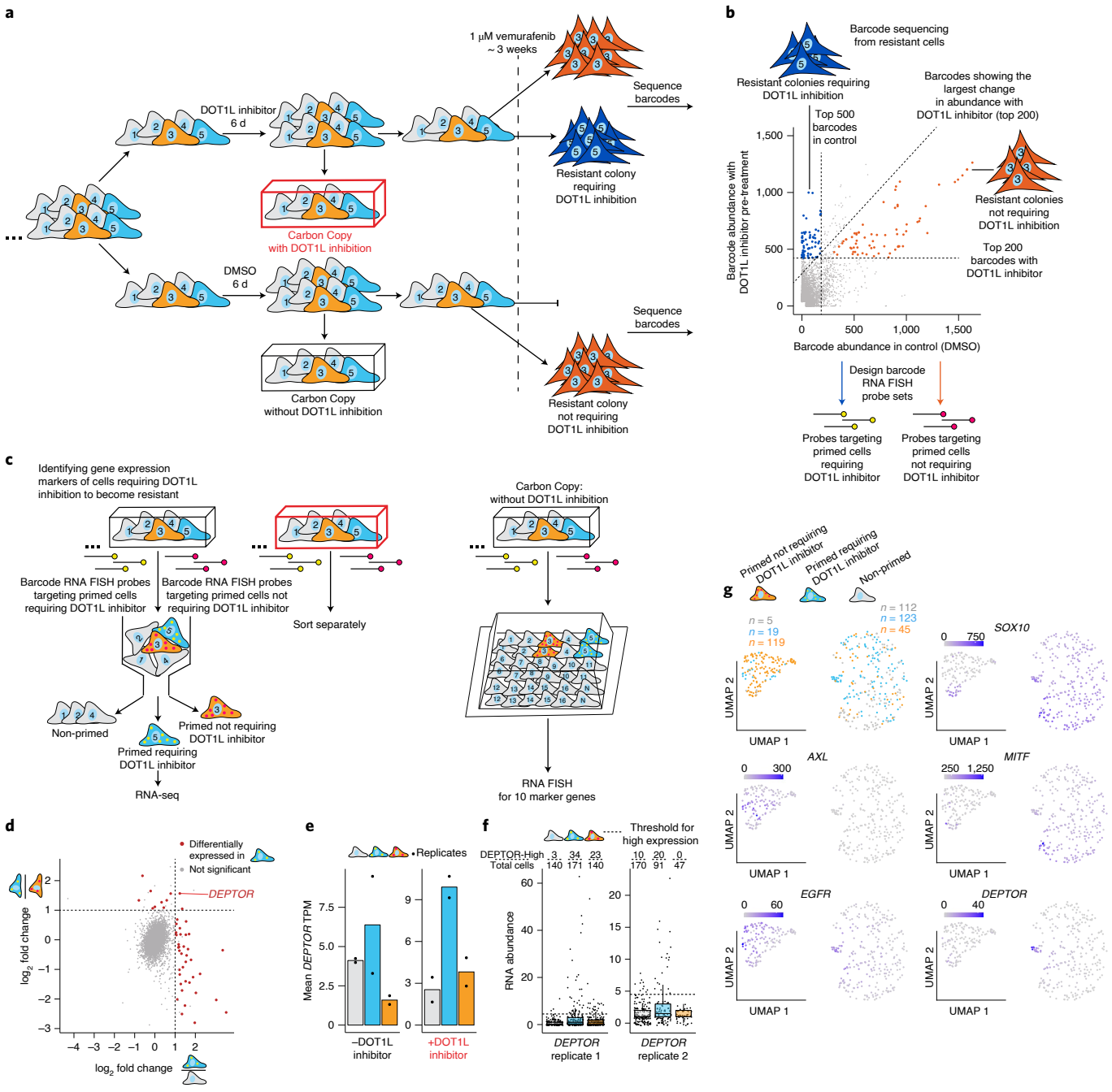
DOT1L inhibition enables a distinct primed subpopulation of melanoma cells to become vemurafenib resistant. These results show that primed cells consist of a complex set of subpopulations that can map to a variety of cell fates. A critical point is that the mapping and hence subpopulations were revealed by the addition of a particular drug. It is possible that there are additional subpopulations present in cells that would normally not survive drug treatment. Furthermore, it may be that the molecular differences that characterize these subpopulations could allow otherwise drug-susceptible cells to become primed for drug resistance in different conditions. Evidence for such a possibility comes from the existence of factors that, when perturbed in drug-naive cells, can reduce or increase the frequency of resistant colony formation, implying an increase or decrease in the number of primed cells within the population[25]. Among these is DOT1L, an H3K79 methylase whose inhibition leads to a three-fold increase in the number of resistant colonies that form upon addition of vemurafenib[25]. Although DOT1L inhibition removes some type of barrier that allows more cells to be primed, this barrier is not removed in all cells because not all cells are able to form resistant colonies. Thus, an important question is what distinguishes the small subset of the cells that become primed for resistance upon DOT1L inhibition from the majority of cells that remain non-resistant to drug. (Barcoding analysis revealed that DOT1L inhibition indeed permits a new subset of cells to enter a primed state rather than affecting proliferation or reversion of primed cells; Supplementary Fig. 11.)

Using Rewind, we sought to reveal the molecular profile specific to the subpopulation of cells that required DOT1L inhibition to survive vemurafenib treatment. To this end, we designed multiple RNA FISH probe sets to separately label the cells that required DOT1L inhibition to become resistant and cells that become resistant irrespective of DOT1L inhibitor treatment (Fig. 5a,b). (We expected these probe sets to label fewer than 1:10,000 cells.) We then used these probes to sort corresponding cells from Carbon Copies fixed before vemurafenib treatment (Fig. 5c, Supplementary Fig. 12 and Supplementary Fig. 13). RNA sequencing of the sorted subpopulations revealed a few dozen genes differentially expressed between cells that required DOT1L inhibition to survive vemurafenib treatment and non-primed cells (Fig. 5d and Supplementary Fig. 14a–e). Interestingly, we observed differences in expression even in the absence of DOT1L inhibition, suggesting that these genes marked a subpopulation that exists independent of the inhibition of DOT1L but nevertheless requires DOT1L inhibition to become resistant (Supplementary Fig. 14). Although most differentially expressed genes were also expressed in 'conventionally primed' cells, there were a few genes whose expression was somewhat specific to cells that were primed for resistance only when DOT1L was inhibited (Fig. 5d,e and Supplementary Fig. 14a–c). Of these, we selected the gene *DEPTOR*, whose expression we sought to characterize in single cells in our Carbon Copy by RNA FISH (Fig. 5f). (We also chose another gene, *MGP*, whose expression was similarly highly elevated but only in one replicate; Supplementary Fig. 15.)

For single-cell analysis, we performed RNA FISH on the Carbon Copies (half treated with DOT1L inhibitor and half treated with vehicle control) for ten total genes: six priming markers, *SOX10*, *MITF*, *DEPTOR* and *MGP*. We scanned through ~2 million cells to find those expressing the targeted barcodes, identifying 850 such cells. Using UMAP, we visualized the expression profiles of cells from the vehicle control-treated Carbon Copy,

Fig. 5 | Rewind identifies a distinct subpopulation of cells that require DOT1L inhibition to become vemurafenib resistant. a, Experimental approach for identifying the subpopulation of cells that require DOT1L inhibition to become vemurafenib resistant. These experiments began with approximately 400,000 WM989 A6-G3 cells transduced at an MOI of ~1.0 and allowed to divide for 10–11 d (~3–4 population doublings) before splitting the culture into two groups. We treated one group with 4 μM DOT1L inhibitor (pinometostat) and the other with vehicle control (DMSO) for another 6 d (~2–3 population doublings). We then split each group again, fixing half as our 'Carbon Copies' and treating the other half with 1 μM vemurafenib for ~2.5 weeks. After vemurafenib treatment, we extracted gDNA from the remaining cells for barcode sequencing. Note that, in principle, DOT1L inhibition might alter cell state (color) even before vemurafenib treatment, which is not depicted here for clarity. **b**, For each barcode identified by sequencing, we plotted its abundance in resistant cells pre-treated with DOT1L inhibitor versus its abundance in resistant cells pre-treated with vehicle control. This comparison revealed a subset of barcodes with a greater relative abundance in resistant cells pre-treated with DOT1L inhibitor (blue points). We used these barcodes to design RNA FISH probes targeting cells that required DOT1L inhibition to become vemurafenib resistant. A separate set of barcodes were highly abundant in resistant cells both with or without DOT1L inhibition (orange points), suggesting that these cells were destined to become resistant whether or not they were pre-treated with DOT1L inhibitor. We used these barcodes to design RNA FISH probes targeting primed cells not requiring DOT1L inhibition to become resistant. The dashed, diagonal line demarcates the 200 barcodes with the largest increase in abundance with DOT1L inhibitor pre-treatment. **c**, Using these probes, we labeled and sorted cells requiring DOT1L inhibition to become vemurafenib resistant (blue), primed cells not requiring DOT1L inhibition (orange) and non-primed cells (gray) from Carbon Copies for RNA sequencing. We separately sorted cells from Carbon Copies treated with DOT1L inhibitor and Carbon Copies treated with vehicle control (two biological replicates each). **d**, To identify markers of cells that require DOT1L inhibition to become resistant, we used DESeq2 to compare their gene expression to non-primed cells (*x* axis) and primed cells not requiring DOT1L inhibition (*y* axis). In this analysis, we included cells sorted from all Carbon Copies (treated with DOT1L inhibitor or vehicle control) from two biological replicates and included DOT1L inhibitor treatment as a covariate in estimating $\log_2$ fold changes. Red points correspond to genes differentially expressed in one or both comparisons ($P_{adj} \leq 0.1$ and $\log_2$ fold change ≥ 1). **e**. Expression of *DEPTOR* in TPM in the subpopulations isolated in **b**. Points indicate TPM values for experimental replicates. **f**, We used the same probe sets as in **b** to identify cells in situ in Carbon Copies fixed before vemurafenib treatment and then measured single-cell expression of *DEPTOR*, *MGP*, *SOX10*, *MITF* and six priming markers by RNA FISH. Shown is the expression of *DEPTOR* in the indicated cell populations identified in the Carbon Copies treated with vehicle control. Each point corresponds to an individual cell. Above each box plot is the proportion of cells with levels of *DEPTOR* RNA above the indicated threshold (~95th percentile in non-primed cells). **g**, We applied the UMAP algorithm to visualize the single-cell expression data from in situ Carbon Copies. These plots include 423 cells from the vehicle control-treated Carbon Copy. In the upper-left plot, points are colored according to the fate of each cell as determined by its barcode, and the number of cells corresponding to each fate are labeled separately above the two largest groupings. For the remaining plots, points are colored by the expression level of the indicated gene in that cell. These data correspond to one of two biological replicates (see Supplementary Fig. 14 for the replicate data).

overlaying the information provided by barcode RNA FISH to label cells by their fates (Fig. 5g). We found that the primed cells that did not require DOT1L inhibition to become resistant separated into a distinct grouping that, as before, expressed the previously identified markers, such as *AXL* and *EGFR* (Fig. 5g and Supplementary Fig. 14f,g). We initially expected the expression of these genes to also be elevated in the cells that required DOT1L inhibition to become resistant, but perhaps to a lesser extent, reflecting a 'subthreshold' state that was unable to survive vemurafenib treatment alone. Contrary to this expectation, the expression profile of this new subpopulation was far more similar to the general population of cells that were not primed for resistance in either condition (Fig. 5g). Although many of these cells were grouped together with the non-primed cells in the UMAP projection, there was a distinct grouping nearby that consisted almost exclusively of cells that were primed for resistance only upon DOT1L inhibition. These cells specifically expressed high levels of *DEPTOR*, along with slightly elevated levels of *EGFR* and lower levels of *MITF*, but showed no differences in the expression levels of the other genes measured compared with non-primed cells (Fig. 5g and Supplementary Fig. 14f–h). (Cells requiring DOT1L inhibition for priming were also enriched for *MGP* in a separate replicate experiment; Supplementary Fig. 15.) Taken together, the identification of a unique molecular state marked by *DEPTOR* expression in the overall absence of established priming markers highlights the existence of a qualitatively distinct rare cell state that can lead to drug resistance when a DOT1L inhibitor is given before vemurafenib. It is noteworthy that many of the primed cells that require DOT1L inhibition to become vemurafenib resistant expressed neither *DEPTOR* nor established markers (for example, *AXL*, *NGFR* and *ITGA3*), and further work is needed to identify features that better distinguish this rare subpopulation.



Although this subpopulation expressed low levels of established priming markers initially, we wondered whether DOT1L inhibition pushed these cells toward a molecular state more similar to the conventional primed cell state (that is, high levels of *AXL*, *EGFR* and *NGFR*; Fig. 6a). To this end, we compared the transcript levels as measured by RNA sequencing from cells sorted from Carbon Copies treated with either DOT1L inhibitor or vehicle control (Fig. 6b). As expected, with vehicle control, cells that require DOT1L inhibition to become vemurafenib resistant clustered separately from primed cells that do not require DOT1L inhibition (Fig. 6c,d). With DOT1L inhibition, these two populations appeared modestly more similar transcriptionally; however, they remained predominantly distinct (Fig. 6c,d). RNA FISH on cells that require DOT1L inhibition to become resistant revealed that DOT1L inhibition did not increase expression of established priming markers and, if anything, modestly decreased their expression (Fig. 6e,f and Supplementary Fig. 16a,b). Overall, these gene expression differences between primed subpopulations both before and after DOT1L inhibition suggest that DOT1L inhibition does not simply convert cells into the previously established primed cell state capable of surviving vemurafenib treatment, but rather, it might reveal a separate route to resistance.

Discussion

Here we have revealed the existence of a rich set of rare subpopulations within seemingly homogenous cells, several of which can lead to phenotypically distinct fates. Despite the population having a clonal origin and being grown in homogeneous cell culture conditions, these subpopulations spontaneously emerge via transient cell state fluctuations that can persist for several cell divisions. It remains unclear how precisely these subpopulations arise, although, intriguingly, they might arise from network interactions between multiple regulatory factors[28]. It is also unclear how these states revert to the population baseline. In this study, we observed states persisting for over 5–6 generations, whereas previous reports based on sorting by individual markers suggested reversion on shorter time scales[6]. It is possible that the more pure primed population identified by Rewind can persist longer than impure populations, which might contain transient intermediates.

For the variability that is associated with priming, it is tempting to imagine single axes of variability for both state and fate, in which cells that have fluctuated further up a putative primed state hierarchy lead to different degrees of resistance. However, our results show that, even for the simple case of heterogeneity in the size of resistant clones, expression of the primed-cell markers *AXL/ITGA3/EGFR* and *WNT5A/MMP1/NGFR* varied along at least two axes prior to the addition of drug, with each axis being associated with either the low-abundance or high-abundance clones. Further use of tools like Rewind, potentially in combination with transcriptome-scale RNA FISH or single-cell RNA sequencing, may help to fully reveal the structure of these fluctuations and consequent subpopulations. Resistant cell fates likely have similarly complex modes of variability, and our results suggest that these modes likely have origins in molecular variability in the initial cell state. The nature of these mappings might help guide therapy, and it may be important to consider the multiple different initial primed cellular states that give rise to resistant cells following distinct treatments, as highlighted by our DOT1L inhibition results.

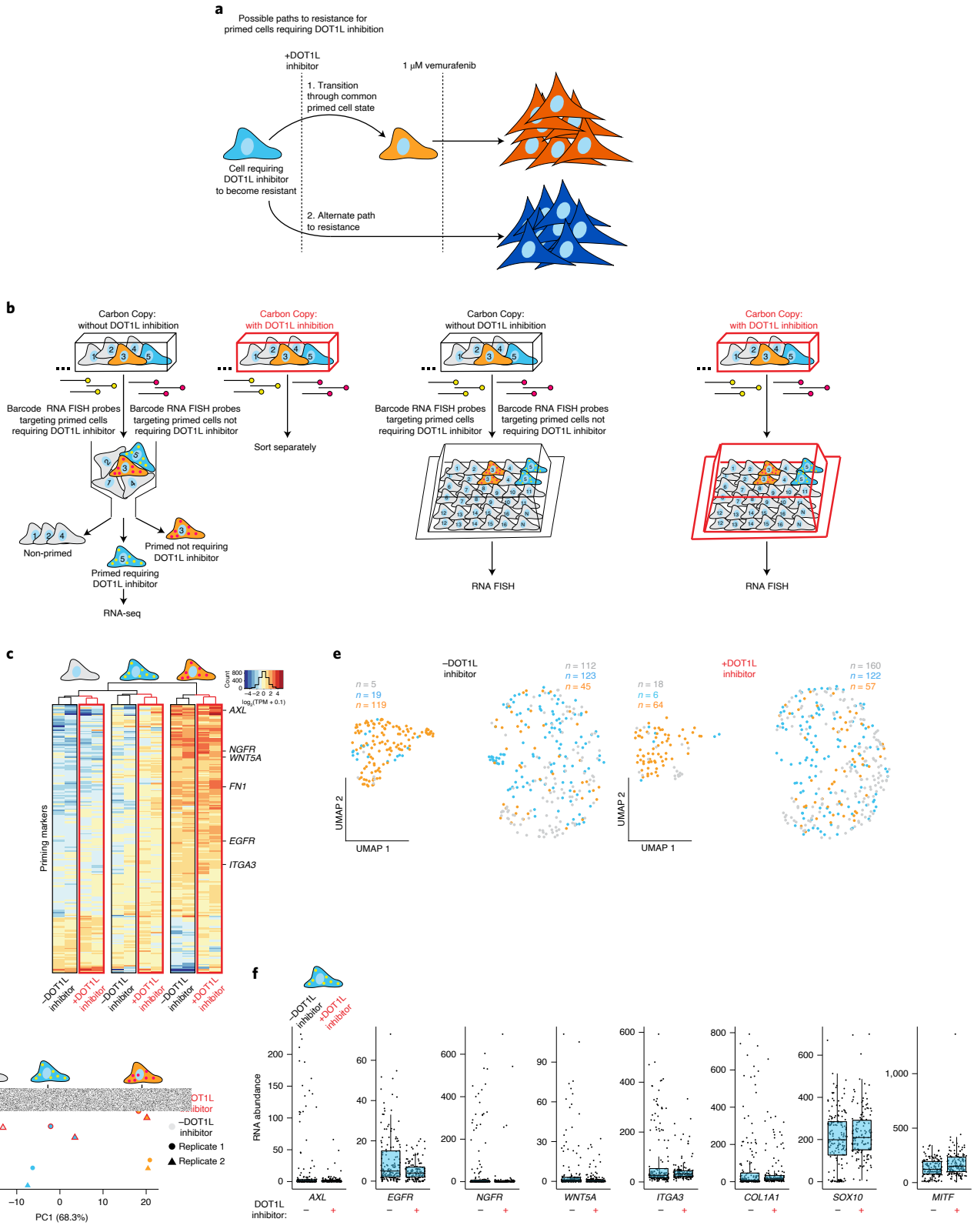
A critical consideration in developing Rewind was minimizing contamination from 'off-target' non-primed cells. These cells could in principle come from probes falsely generating signal in non-primed cells or technical limitations of FACS. These contaminating cells can dramatically dilute measurements of gene expression specific to the targeted, rare subpopulations (Supplementary Fig. 1d,e). We found that barcode detection by FACS was far more prone to contamination than barcode detection by imaging, which had very high precision (estimated to be ~97%; Supplementary Fig. 1f–h); indeed, we believe that it is for this reason that we observe larger magnitude differences by RNA FISH than by RNA sequencing of sorted populations, particularly for markers downregulated in primed cells such as *SOX10* and *MITF*. However, despite these concerns, we discovered and validated the priming markers *ITGA3* and *AXL*, while also identifying previously known markers such as *NGFR* and *EGFR*. We also found that experiment-to-experiment technical variability was relatively minimal: by imaging, we did not see much difference in off-target signal across different probe sets (with rare exceptions of 'dirty' probes), and barcode sequencing of cDNA from sorted subpopulations labeled with different probe sets suggested similar levels of enrichment (Fig. 1, Supplementary Fig. 12 and a notable exception in Supplementary Fig. 6).

The global transcriptional profiles afforded by RNA sequencing of rare primed cells allowed us to ask what pathways might be active in these cells beyond ones such as growth factor receptor signaling, which has already been associated with vemurafenib resistance in melanoma[6,25,29–31]. One of the strongest signatures was the upregulation of cell adhesion proteins and structural components of the ECM. Such signatures suggest the possibility that control of cell state and behavior might have both a component that is autonomous to the cell itself and a component that is instructed by the ECM. Future research may help reveal if and how the ECM is able to influence primed cellular states and, consequently, therapy resistance.

Fig. 6 | DOT1Li inhibition enables a new subpopulation of cells to survive vemurafenib treatment without converting them into the known primed cell state. a, We asked whether DOT1L inhibition enables a new subpopulation of cells to survive vemurafenib treatment by converting them into the previously established primed cell state or whether these cells become resistant via a possible alternative path. **b**, We used Rewind to isolate and perform RNA sequencing on cells requiring DOT1L inhibition to survive vemurafenib treatment (blue), cells not requiring DOT1L inhibition (orange) and non-primed cells (gray) sorted from both Carbon Copies treated with DOT1L inhibitor (red outline) and Carbon Copies treated with vehicle control (gray outline) (two replicates each sorted for RNA sequencing). **c**, Heat map displays expression of established priming markers across sorted subpopulations from control and DOT1L inhibitor pre-treated Carbon Copies. Dendrogram shows hierarchical clustering of samples by expression values. We defined priming markers as protein-coding genes differentially expressed ($P_{adj} \leq 0.1$ and absolute($\log_2$ fold change) ≥ 1) in primed cells not requiring DOT1L inhibition versus non-primed cells isolated from the Carbon Copy treated with vehicle control. **d**, Using expression of priming markers as in **c**, we performed principal component (PC) analysis on primed and non-primed cell populations. The red outline indicates samples sorted from the Carbon Copy treated with DOT1L inhibitor. **e**, We used the same probes as in **b** to identify cell populations in situ in Carbon Copies treated with DOT1L inhibitor or vehicle control. We then used RNA FISH to measure single-cell expression of several established priming markers and visualized the relationship in gene expression between single cells using the UMAP algorithm with the first six principal components. This analysis included expression data from 850 single cells. Points are colored according to the fate of each cell as determined by its barcode, and the number of cells corresponding to each fate are labeled above the largest groupings. **f**, Plotted are single-cell expression data for six priming markers, *MITF* and *SOX10* in cells that require DOT1L inhibition to become vemurafenib resistant. Each point corresponds to an individual cell. Below each box plot, we indicate whether the cells are from the Carbon Copy treated with DOT1L inhibitor (+) or vehicle control (−). The corresponding data for non-primed cells and primed cells not requiring DOT1L inhibition are shown in Supplementary Fig. 16. These data correspond to one biological replicate.

There were also several other expression signatures active in distinct subpopulations of cells. For instance, *DEPTOR* expression marked one set of primed cells. Although *DEPTOR* might not have any functional role in priming, it is known that DEPTOR inhibits mTOR signaling, which might relieve negative feedback on PI3K/Akt signaling and, in turn, bypass the inhibition of BRAF signaling[32]. Further work is needed to establish such potential mechanisms.

In this study, single-cell analysis of ERK signaling revealed that individual cells vary dramatically in ERK activity after vemurafenib treatment, with rare cells reactivating ERK to levels similar to untreated cells. Rewind allowed us to connect these near-term



single-cell signaling dynamics in rare cells to both their initial transcriptional state and their ultimate resistant fate. These connections revealed that the primed melanoma cells that go on to survive vemurafenib treatment both had higher levels of phosphorylated ERK soon after treatment and expressed multiple receptor tyrosine kinases along with their cognate ligands. Of note, other groups have demonstrated that by inhibiting MAPK signaling, vemurafenib treatment can relieve negative feedback on growth factor receptors, thereby allowing ERK reactivation via $BRAF^{V600E}$ -independent routes[27,33]. Thus, it is tempting to speculate that transcriptional state of primed cells enables autonomous ERK reactivation.

We chose to focus on the priming of melanoma cells toward different fates after targeted therapy treatment. However, there are several examples in which non-genetic differences can lead rare cells to undergo important transformations, including the induction of pluripotency in otherwise terminally differentiated cells[34] and transdifferentiation of one cell type into another. Application of techniques like Rewind in these contexts may reveal universal characteristics of priming and reprogramming.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-021-00837-3.

Received: 29 May 2020; Accepted: 18 January 2021;
Published online: 22 February 2021

References

1. Symmons, O. & Raj, A. What's luck got to do with it: single cells, multiple fates, and biological nondeterminism. *Mol. Cell* **62**, 788–802 (2016).
2. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
3. Gupta, P. B. et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
4. Sharma, S. V. et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
5. Roesch, A. et al. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141**, 583–594 (2010).
6. Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
7. Shaffer, S. M. et al. Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell* **182**, 947–959 (2020)
8. Cohen, A. A. et al. Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511–1516 (2008).
9. Spencer, S., Gaudet, S., Albeck, J., Burke, J. & Sorger, P. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432 (2009).
10. Sigal, A. et al. Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
11. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2016).
12. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
13. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).
14. McKenna, A. et al. Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
15. Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
16. Quinn, J. J. et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science* https://doi.org/10.1126/science.abc1944 (2020).
17. Biddy, B. A. et al. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).
18. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020).
19. Al'Khafaji, A. et al. Expressed barcodes enable clonal characterization of chemotherapeutic responses in chronic lymphocytic leukemia. Preprint at *bioRxiv* https://doi.org/10.1101/761981 (2019).
20. Hurley, K. et al. Reconstructed single-cell fate trajectories define lineage plasticity windows during differentiation of human PSC-derived distal lung progenitors. *Cell Stem Cell* **26**, 593–608 (2020).
21. Al'Khafaji, A. M., Deatherage, D. & Brock, A. Control of lineage-specific gene expression by functionalized gRNA barcodes. *ACS Synth. Biol.* **7**, 2468–2474 (2018).
22. Feldman, D. et al. CloneSifter: enrichment of rare clones from heterogeneous cell populations. *BMC Biol.* **18**, 177 (2020).
23. Rebbeck, C. et al. SmartCodes: functionalized barcodes that enable targeted retrieval of clonal lineages from a heterogeneous population. Preprint at *bioRxiv* https://doi.org/10.1101/352617 (2018).
24. Umkehrer, C. et al. Isolating live cell clones from barcoded populations using CRISPRa-inducible reports. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0614-0 (2020).
25. Torre, E. A. et al. Genetic screening for single-cell variability modulators driving therapy resistance. *Nat. Genet.* **53**, 76–85 (2021).
26. Müller, J. et al. Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat. Commun.* **5**, 5712 (2014).
27. Gerosa, L. et al. Receptor-driven ERK pulses reconfigure MAPK signaling and enable persistence of drug-adapted BRAF-mutant melanoma cells. *Cell Syst.* **11**, 478–494.e9 (2020).
28. Schuh, L. et al. Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer. *Cell Syst.* **10**, 363–378 (2020).
29. Sun, C. et al. Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma. *Nature* **508**, 118–122 (2014).
30. Corcoran, R. B. et al. EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer Discov.* **2**, 227–235 (2012).
31. Ji, Z. et al. MITF modulates therapeutic resistance through EGFR signaling. *J. Invest. Dermatol.* **135**, 1863–1872 (2015).
32. Caron, A., Briscoe, D. M., Richard, D. & Laplante, M. DEPTOR at the nexus of cancer, metabolism, and immunity. *Physiol. Rev.* **98**, 1765–1803 (2018).
33. Lito, P. et al. Relief of profound feedback inhibition of mitogenic signaling by RAF inhibitors attenuates their activity in BRAFV600E melanomas. *Cancer Cell* **22**, 668–682 (2012).
34. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).

Methods

Barcode lentivirus library construction. Starting with the LRG2.1T plasmid, kindly provided by Junwei Shi, we derived a lentivirus vector backbone for Rewind by removing the U6 promoter and single guide RNA scaffold and then inserting a spacer sequence flanked by EcoRV restriction sites after the stop codon of GFP. For the barcode insert, we ordered PAGE-purified Ultramer oligonucleotides (IDT) containing 'WSN' repeated for 100 nucleotides (W=A or T, S=G or C, N=Any) flanked by 30 nucleotides homologous to the vector insertion site for Gibson Assembly (see Supplementary Table 1 for barcode insert sequence). We then digested the vector backbone overnight with EcoRV (NEB) and gel purified the linearized vector. We combined 100 ng of linearized vector, 1.08 µl of barcode oligo insert (100 nM in nuclease-free water), 10 µl of Gibson Assembly Master Mix (NEB E2611) and nuclease-free water to a final volume of 20 µl and then incubated the reaction at 50 °C for 1 h. We next column purified the assembled plasmid using Monarch DNA cleanup columns (NEB) according to the manufacturer's protocol and then electroporated 2 µl of the column-purified plasmid into Endura Electrocompetent *Escherichia coli* cells (Lucigen) using a Gene Pulser Xcell (Bio-Rad) with the following settings: 25 ms pulse length, 10 µF capacitance, 600 Ω resistance and 1,800 V voltage. We performed six electroporations using the same plasmid in parallel. Immediately after electroporation, we added 1 ml of pre-warmed (37 °C) recovery media to each electroporation cuvette and then transferred the liquid to 1.5-ml microcentrifuge tubes and placed these tubes on a shaker at 225 r.p.m. and 37 °C for 1 h. After this recovery, we took 10 µl of the culture for plating serial dilutions and transferred the rest to 150–200 ml of 1× LB Broth containing 100 µg ml^{-1} of ampicillin. We incubated these cultures on a shaker at 225 r.p.m. and 32 °C for 12–14 h and then pelleted the cultures by centrifugation and isolated plasmid using the EndoFree Plasmid Maxi Kit (Qiagen) according to the manufacturer's protocol. In some instances, pellets were frozen at −20 °C for several days before plasmid isolation. To estimate transformation efficiency, we counted colonies on the plated serial dilutions and verified barcode insertion by polymerase chain reaction (PCR) from 20–30 colonies per plate. We pooled the plasmids from the six separate cultures in equal amounts by weight before packaging into lentivirus. This protocol is also available online at https://www.protocols.io/view/barcode-plasmid-library-cloning-4hggt3w.

Cell lines and culture. We derived the WM989 A6-G3 melanoma cell line by twice single-cell bottlenecking the WM989 melanoma cell line kindly provided by Dr. Meenhard Herlyn (Wistar Institute)[6,35]. Similarly, we derived WM983b E9-C6 by twice single-cell bottlenecking the WM983b melanoma cell line, also provided by Dr. Meenhard Herlyn. We verified the identity of these cell lines by DNA STR microsatellite fingerprinting at the Wistar Institute.

We cultured both melanoma cell lines in TU2% media consisting of 80% MCDB 153, 10% Leibovitz's L-15, 2% FBS, 2.4 mM $CaCl_2$, 50 U ml^{-1} of penicillin and 50 µg ml^{-1} of streptomycin and passaged cells using 0.05% trypsin-EDTA. For harvesting drug-treated resistant cells, we used 0.1% trypsin-EDTA. For lentivirus packaging, we cultured HEK293FT cells in DMEM containing 10% FBS, 50 U ml^{-1} of penicillin and 50 µg ml^{-1} of streptomycin and passaged cells using 0.05% trypsin-EDTA.

Lentivirus packaging and transduction. Before plasmid transfection, HEK293FT cells were grown to ~90% confluency in six-well plates in DMEM containing 10% FBS without antibiotics. For each six-well plate, we added 80 µl of PEI to 0.5 ml of Opti-MEM (Thermo Fisher Scientific, 31985062) and separately combined 7.5 µg of pPAX2 with 5 µg of VSVG and 7.71 µg of the barcode plasmid library in 0.5 ml of Opti-MEM and then incubated the solutions at room temperature for 5 min. We then mixed the two solutions together with vortexing and incubated the combined solution at room temperature for 15 min. We added 184 µl of the plasmid-PEI solution dropwise to each well of the six-well plate. After 6–8 h, we aspirated the media from the cells, washed the cells once with 1× DPBS and then added fresh culture media (DMEM containing 10% FBS and antibiotics). The next morning, after confirming that most cells were GFP positive, we aspirated the media, washed the cells once with 1× DPBS and then added 1 ml of TU2% to each well. Approximately 12 h later, we transferred the virus-laden media to a Falcon tube and added another 1 ml of of TU2% to each well. We collected virus-laden media twice more over the next ~16 h and, during this time, stored the collected media at 4 °C. After the final collection, we filtered the virus-laden media through a 0.22-µm PES filter and then stored 1–2-ml aliquots at −80 °C.

To transduce WM989 A6-G3 and WM983b E9-C6 cells, we added freshly thawed (on ice) virus-laden media and polybrene (final concentration, 4 µg ml^{-1}) to dissociated cells and then plated the cells onto six-well plates (100,000 cells in 2 ml of media per well) and centrifuged the plate at 1,750 r.p.m. (517*g*) for 25 min. We incubated the cells with virus for 6–8 h and then removed the media, washed the cells once with 1× DPBS and added 3 ml of TU2% to each well. The next day, we passaged the cells to 10-cm dishes (one six-well plate into three 10-cm dishes). For WM989 A6-G3, we split barcoded cells into Carbon Copy and separate vemurafenib treatment groups 11 d after transduction for sort experiments (Fig. 1) or 10 d after transduction for in situ experiments (Figs. 2–4) unless otherwise specified. These time points correspond to 4–5 population doublings since transduction. For WM983b E9-C6, we split barcoded cells into Carbon Copy and separate vemurafenib treatment groups 7 d after transduction (also corresponding to 4–5 population doublings) unless otherwise specified. We cultured in situ Carbon Copies for 4 d before fixation to more easily identify clusters of cells expressing targeted barcodes.

Simulations of experimental conditions used for Rewind. As described above, we expanded barcoded cells for at least four population doublings before splitting off the Carbon Copy and drug-treatment groups for Rewind. As such, there were, on average, ~16 closely related cells for each barcoded clone before the split. For a 50:50 split, the probability that at least one of 16 cells ends up in both groups is ~99.997%, or, in other words, fewer than 0.002% of clones are expected to be 'missing' from either group. However, given the unavoidable variability in cell growth, it is likely that some clones will have divided fewer than four times, and these clones are more likely to be entirely missing from the Carbon Copy. (We note that we do not care about clones that are missing from the drug treatment group because they will not become resistant colonies, and their barcodes will not be selected for probe design). To empirically estimate the proportion of clones present in our Carbon Copy, we sequenced barcode genomic DNA (gDNA) from barcoded WM989 A6-G3 after ~4 population doublings and then computationally 'split' the sequenced barcodes into two halves, after first weighting each barcode by its read count and scaling the average read count to 16. Finally, we calculated the proportion of barcodes present in both halves. Simulating this procedure 10,000 times, we found that ~92.3–92.6% of barcodes were present in both halves, and <4% of barcodes were 'missing' from the simulated Carbon Copy.

We also note that, to eliminate spurious barcodes arising due to PCR or sequencing errors, we merged highly similar barcode sequences as described further below (see 'Computational analyses of barcode sequencing data') and filtered barcodes with fewer than five unique reads. The simulations were robust to a range of read count thresholds ≥2.

We used the same barcode sequencing data to simulate the 'heritability split experiment' for Supplementary Fig. 3d. In this case, we randomly sampled 200 barcodes twice (without replacement and weighting each barcode by its read count) and then calculated the proportion of barcodes shared between the two samples. We performed the same simulation for WM983b E9-C6 (Supplementary Fig. 6b) using sequencing data from barcoded WM983b E9-C6 grown for ~4 population doublings.

The scripts used for these simulations are available on Dropbox at https://www.dropbox.com/s/p5t9onmezasmtty/heritabilitySplitWM989.R?dl=0.

FACS. To isolate ITGA3-High WM989 A6-G3, we first trypsinized and pelleted eight confluent 10-cm plates, washed once with 1× DPBS containing 0.1% BSA (0.1% BSA-PBS) and then split the cells into two equal pellets. We resuspended each pellet in 0.4 ml of 0.1% BSA-PBS containing 1:200 anti-ITGA3 antibody (DSHB clone P1B5 stock concentration, 354 µg ml^{-1}) and then incubated on ice for 1 h. After primary incubation, we pelleted the cells, washed twice with ~5 ml of 0.1% BSA-PBS and then resuspended cells in 0.16 ml of 0.1% BSA-PBS containing 1:500 anti-mouse FAb2 secondary antibody conjugated to Alexa Fluor 488 (Cell Signaling, 4408) and incubated on ice for 30 min. Finally, we pelleted the cells, washed twice with 0.1% BSA-PBS and then resuspended the pellet in 0.1% BSA-PBS containing 100 ng ml^{-1} of DAPI and proceeded with FACS on a MoFlo Astrios (Beckman Coulter). After gating for singlets and live cells, we collected 15,000 events from the brightest 0.3–0.4% ITGA3-High gate and equal numbers from the dimmest ~99% ITGA3-Low gate. We plated two-thirds of the sorted cells onto two-well glass-bottom chamber plates (Nunc Lab-Tek 155380) for treating with vemurafenib (see below) and the rest on a separate two-well glass-bottom chamber plate for verifying *ITGA3* expression by single-molecule RNA FISH.

We followed a similar procedure for isolating AXL-High WM983b E9-C6, starting with ten 10-cm dishes split into two equal cell pellets, performing all incubations and washes with 1% BSA-PBS and staining with 1:50 primary antibody (goat anti-human AXL AF154 from Novus Biologicals) and 1:60 secondary antibody (bovine anti-goat conjugated to Alexa Fluor 647; Jackson ImmunoResearch, 805-605-180). After gating for singlets and live cells, we collected 20,000 events from the brightest ~0.3% AXL-High gate and equal numbers from the dimmest ~20% AXL-Low gate and then plated cells onto two-well glass-bottom plates (10,000 cells per well) for vemurafenib treatment or RNA FISH as above.

Drug treatment experiments. We prepared stock solutions of 4 mM vemurafenib (PLX4032, Selleck Chemicals, S1267), 10 mM pinometostat (Selleck Chemicals, S7062), 100 µM trametinib (Selleck Chemicals, S2673) and 10 mM dabrafenib (Selleck Chemicals, S2807). We prepared all stock solutions in DMSO and divided into small aliquots stored at −20 °C to minimize freeze–thaw cycles. For drug treatment experiments, we diluted the stock solutions in culture medium to a final concentration of 1 µM for vemurafenib, 4 µM for pinometostat, 10 nM for trametinib and 1 µM for dabrafenib unless otherwise specified.

For Rewind experiments in WM989 A6-G3, we treated cells with vemurafenib for 3 weeks, replacing media containing drug every 3–4 d. After vemurafenib treatment, we trypsinized and collected all remaining cells, washed cells once with 1× DPBS and then pelleted and froze 90% of the cells at −20 °C until gDNA

extraction and barcode sequencing as described below. We fixed the remaining 10% of vemurafenib-resistant cells for barcode RNA FISH, FACS and RNA sequencing as described below and in Supplementary Fig. 5a. For DOT1L inhibitor pre-treatment, we treated cells with 4 μM pinometostat for 6 d, replacing media on day 3 and again when splitting off the Carbon Copy on day 5. After the ITGA3 sort, we fixed WM989 A6-G3 cells after 18 d of vemurafenib treatment to more easily quantify numbers of colonies. For Rewind experiments in WM983b E9-C6, we treated cells with vemurafenib for 4 weeks, replacing media containing drug every 3–4 d. Cells surviving drug treatment were harvested and frozen as described above.

Cell quantification. After drug treatment experiments, we fixed cells by incubation for 10 min in 3.7% formaldehyde (Sigma-Aldrich, F1635) diluted in 1× PBS, followed by two washes with 1× PBS and then overnight permeabilization at 4 °C with 70% ethanol. We stained nuclei by incubation in 2× SSC containing 50 ng ml^{-1} of DAPI and then imaged most of each well via a tiling scan at ×20 magnification. To quantify cell and colony numbers, we used custom MATLAB software to stitch the tiled images, identify nuclei and manually circle individual resistant colonies. Software and scripts used for these analyses can be found at https://github.com/arjunrajlaboratory/colonycounting_v2 and https://www.dropbox.com/sh/p279h7mak0rrklx/AACyM_IiVP3prkjdDmd6HqOca?dl=0.

Barcode library preparation and sequencing. We isolated gDNA from barcoded cells using the QIAmp DNA Mini Kit (Qiagen, 51304) according to the manufacturer's protocol. We performed targeted amplification of the integrated barcode vector using custom primers containing Illumina adapter sequences, unique sample indexes, variable-length staggered bases and six random nucleotides ('UMI'; NHNNNN), which, despite not uniquely tagging barcode DNA molecules, appeared to modestly increase reproducibility between replicate libraries and normalize read counts (see Supplementary Table 2 for a complete list of primers). For each sample, we performed multiple PCR reactions (using 20–40% of the total isolated gDNA) each consisting of 1 μg of gDNA, 500 nM primers, 25 μl of NEBNext Q5 HotStart HiFi PCR Master Mix and nuclease-free water to a final volume of 50 μl. We ran the reactions on a thermal cycler with the following settings: 98 °C for 30 s, followed by N cycles of 98 °C for 10 s and then 65 °C for 40 s and, finally, 65 °C for 5 min. After the PCR, we purified libraries using 35 μl (0.7×) of AMPure XP magnetic beads with two 80% ethanol washes followed by final elution in 20 μl of 0.1× TE (1 mM Tris HCl pH 8.0 and 100 μM EDTA). Purified libraries from the same sample were pooled, quantified using the Qubit dsDNA High Sensitivity Assay (Thermo Fisher Scientific) and then sequenced on a NextSeq 500 using 150 cycles for read 1 and eight cycles for each index. For barcoding experiments not requiring RNA FISH probe design, shorter reads (75 cycles) provided sufficient information to identify unique barcodes.

To reduce PCR amplification bias, we determined the number of cycles ('N') for each sample by first performing a separate quantitative PCR (qPCR) reaction and selecting the number of cycles needed to achieve one-third of the maximum fluorescence intensity. We included 0.25 μl of 100× SYBR Green I (10,000× diluted 1:100 in 10 mM Tris pH 8.0, Invitrogen, S7563) per 25 μl of qPCR reaction and, when possible, performed multiple reactions with serial dilutions of gDNA (1:4 and 1:16). For experiments with multiple similar samples (same multiplicity of infection (MOI) and same treatment), we performed qPCR on one of these samples and extrapolated 'N' to the rest.

To test reproducibility of our barcode quantification, for a subset of samples we prepared duplicate libraries with separate indexes and compared barcode read counts among these technical replicates. As shown in Supplementary Fig. 2, we found a high correlation (>95%) in barcode abundance among these technical replicates.

Computational analyses of barcode sequencing data. We recovered barcodes from sequencing data using custom Python scripts available at https://github.com/arjunrajlaboratory/timemachine. These scripts search through each read to identify sequences complementary to our library preparation primers, and, if these sequences pass a minimum length and phred score cutoff, then the intervening barcode sequence is counted. In addition to counting total reads for each barcode, we also count the number of 'UMIs' incorporated into the library preparation primers (see above). Although we do not think that these 'UMIs' tag unique barcode DNA molecules, empirically they appeared to slightly improve the correlation in barcode abundance among replicate libraries and were, therefore, used for most subsequent analyses. Using the STARCODE software[36] (available at https://github.com/gui11aume/starcode), we merged highly similar barcode sequences (Levenshtein distance ≤ 8), summing the counts and keeping only the more abundant barcode sequence.

For selecting barcodes corresponding to resistant colonies, we ranked the barcode sequences by counts and then converted the most abundant 100–200 barcode sequences into FASTA files for probe design, as described below. Barcode sequences with ≥30 bases of homology to the vector backbone were excluded for concerns of generating non-specific FISH probes (we checked for non-specific binding a second time during probe design, as described below).

We selected barcodes corresponding to resistant colonies that require DOT1L inhibition using the following criteria: 1) among the most abundant 200 barcodes in DOT1L inhibitor pre-treated resistant cells; 2) not among the most abundant 500 barcodes in the DMSO pre-treated resistant cells; and 3) greatest difference in abundance between DOT1L inhibitor pre-treated and DMSO pre-treated resistant cells among all barcodes passing criteria 1 and 2. For barcodes corresponding to resistant colonies not requiring DOT1L inhibition, we selected sequences that: 1) were in the top 200 barcodes in both the DOT1L inhibitor and DMSO pre-treated resistant cells and 2) had relatively similar abundances across these two conditions (not among the 500 barcodes with the largest difference in abundance).

Barcode RNA FISH probe design. Using FASTA files of selected barcodes, we design hybridization chain reaction (HCR) probes using RajLab ProbeDesignHD software (code freely available for non-commercial use here: https://github.com/arjunrajlaboratory/ProbeDesign/). For each barcode sequence, we designed two non-overlapping 42mer probes with a target Gibbs free energy for binding of −55 (allowable Gibbs free energy [−65, −45]). We excluded probes with complementarity to repetitive elements, pseudogenes or the vector backbone used to generate the barcode plasmid library. We then split each 42mer probe into two 20mer sequences (removing the middle two nucleotides) and appended split-initiator HCR sequences using custom Python scripts (see Supplementary Table 3 for sequences)[37]. For each 20mer sequence, we measured the maximum complementarity to the vector backbone and other barcodes present in the sample to manually exclude probes with potential for non-specific hybridization. We ordered the final probe sequences synthesized from IDT in picomole-scale 384-well plates. Finally, we resuspended barcode HCR probes to 50 μM in nuclease-free water and then combined these probes into pools each containing 24 different barcode probes at a final concentration of 2 μM each.

For ClampFISH, we designed 30mer probes targeting select barcodes using RajLab ProbeDesignHD software with a target Gibbs free energy of −40 (allowable Gibbs free energy [−50, −30]). As above, we excluded probes with complementarity to repetitive elements, pseudogenes or the vector backbone. We then appended 10mer sequences to the 5′ and 3′ ends of each probe (used for subsequent ligation) and ordered the final probe sequences synthesized from IDT in picomole-scale 384-well plates. We resuspended barcode ClampFISH probes to 100 μM in nuclease-free water and then combined these probes into pools each containing 30 different barcode probes. To these pools, we ligated oligonucleotides (oligos) containing alkyne and azide modifications at their 5′ and 3′ ends, respectively (see Supplementary Table 4 for sequences). For this ligation, we first phosphorylated the 5′ ends of each probe set by combining 4 μl of the pooled oligos with 1 μl of T4 PNK (NEB), 20 μl of T7 DNA ligase reaction buffer (NEB) and 2 μl of nuclease-free water and then incubating at 37 °C overnight. Next, we added the alkyne and azide modified oligos along with complementary bridging 20mer oligos (3 μl each of 400 μM stocks) and heated the reactions to 95 °C for 5 min and then cooled to 12 °C at a rate of −0.1 °C s^{-1}. After cooling, we added 1 μl of T7 ligase (NEB) and incubated overnight at room temperature. We purified the ligated barcode ClampFISH probes using Monarch DNA Cleanup Columns (NEB) according to the manufacturer's protocol. This protocol for generating barcode ClampFISH probes is also available online at https://www.protocols.io/view/invertedclampfish-ligation-qxwdxpe. We prepared amplifier probes MM2B, MM2C, P9B and P9C as described previously[38].

RNA FISH. We designed oligonucleotide probe sets complementary to our genes of interest using custom probe design software written in MATLAB and ordered them with a primary amine group on the 3′ end from Biosearch Technologies (see Supplementary Table 5 for probe sequences). For each gene, we pooled their complementary oligos and coupled the probe set to Cy3 (GE Healthcare), Alexa Fluor 594 (Life Technologies) or Atto647N (ATTO-TEC) *N*-hydroxysuccinimide ester dyes. We performed single-molecule RNA FISH as described in ref. [39] and ref. [6] for multiple cycles of hybridization. We aspirated media from adherent cells, washed the cells once with 1× PBS and then incubated the cells in fixation buffer (3.7% formaldehyde in 1× PBS) for 10 min at room temperature. We next aspirated the fixation buffer, washed samples twice with 1× PBS and then added 70% ethanol and stored samples at 4 °C. For hybridization, we first rinsed samples with washing buffer (10% formamide in 2× SSC) and then applied the RNA FISH probes in hybridization buffer (10% formamide and 10% dextran sulfate in 2× SSC). We covered samples with coverslips and then hybridized samples overnight in humidified containers at 37 °C. The next morning, we washed samples 2 × 30 min with washing buffer at 37 °C, adding 50 ng ml^{-1} of DAPI to the second wash to stain the nuclei. After these washes, we rinsed samples once with 2× SSC and then added new 2× SSC and proceeded with imaging. To strip RNA FISH probes, we incubated samples in stripping buffer (60% formamide in 2× SSC) for 20 min on a hot plate at 37 °C, washed samples 3 × 15 min with 1× PBS on a hot plate at 37 °C and then returned samples to 2× SSC. After stripping RNA FISH probes, we re-imaged all previous positions and excluded dyes with residual signal from subsequent hybridization.

Barcode RNA HCR. We adapted HCR v3.0 (ref. [37]) for barcode RNA FISH as follows. We used 1.2 pmol each of up to 240 barcode RNA FISH probes per 0.3 ml of hybridization buffer. Our primary hybridization buffer consisted of 30% formamide, 10% dextran sulfate, 9 mM citric acid pH 6.0, 50 μg ml^{-1} of heparin, 1×

Denhardt's solution (Life Technologies, 750018) and 0.1% Tween-20 in 5× SSC. For primary hybridization, we used 100 µl of hybridization buffer per well of a six-well plate, covered the well with a glass coverslip and then incubated the samples in humidified containers at 37 °C for 6 h. After the primary probe hybridization, we washed samples 4 × 5 min at 37 °C with washing buffer containing 30% formamide, 9 mM citric acid pH 6.0, 50 µg ml^{-1} of heparin and 0.1% Tween-20 in 5× SSC. We then washed the samples at room temperature 2 × 5 min with 5× SSCT (5× SSC + 0.1% Tween-20) and then incubated the samples at room temperature for 30 min in amplification buffer containing 10% dextran sulfate and 0.1% Tween-20 in 5× SSC. During this incubation, we snap-cooled individual HCR hairpins (Molecular Instruments) conjugated to Alexa Fluor 647 (Alexa647), Alexa Fluor 594 (Alexa594) or Alexa Fluor 546 (Alexa546) by heating to 95 °C for 90 s and then immediately transferring to room temperature to cool for 30 min concealed from light. After these 30 min, we resuspended and pooled the hairpin in amplification buffer to a final concentration of 6 nM each. We added the hairpin solution to samples along with a coverslip and then incubated samples at room temperature overnight (12–16 h) concealed from light. The next morning, we washed samples 5 × 5 min with 5× SSCT containing 50 ng ml^{-1} of DAPI, added SlowFade anti-fade solution (Life Technologies, S36940) and a coverslip and then proceeded with imaging. To remove fluorescent signal for subsequent rounds of RNA FISH or immunofluorescence, we photobleached samples on the microscope or stripped HCR hairpins as described above for RNA FISH probes. We used this modified HCR v3.0 protocol for labeling barcode RNA in all experiments except those indicated in Supplementary Fig. 8, which relied on the ClampFISH protocol described below.

For performing HCR in suspension, we adapted the published protocol[37] as follows. We fixed dissociated cells in suspension by washing the cells with 1× DPBS, resuspending the cell in ice-cold 1× DPBS, adding an equal volume of ice-cold fixation buffer (3.7% formaldehyde 1× PBS) and then incubating with rotation at room temperature for 10 min. We next pelleted fixed cells by centrifugation at 800*g* for 3 min, washed twice with ice-cold 1× PBS and then resuspended in 70% ethanol and stored fixed cells at 4 °C. For primary probe hybridization, we used 0.5 ml of hybridization buffer containing 4 nM of each barcode RNA FISH probe and incubated samples using the same conditions as described above. After primary probe hybridization, we washed samples 4 × 10 min with 0.5 ml of washing buffer and then 2 × 10 min with 0.5 ml of 5× SSCT. We next incubated samples for 30 min in amplification buffer and snap-cooled HCR hairpins as described above. For amplification, we used 15 nM final concentration of each HCR hairpin and incubated samples at room temperature overnight concealed from light. After amplification, we washed samples six times with 5× SSCT and then proceeded with FACS. In between hybridizations and washes, we pelleted cells by centrifugation at 400*g* for 5 min and used low-molecular-weight dextran sulfate (Sigma-Aldrich, D4911) in hybridization and amplification buffers to improve pelleting.

We note that the final hairpin concentrations used in these experiments are four-fold to ten-fold lower than the manufacturer's protocol, which we optimized to reduce non-specific amplification while still enabling sensitive barcode RNA detection at ×20 magnification. At the same time, we have noticed lot-to-lot variation in HCR hairpins purchased from Molecular Instruments, with each lot requiring some testing and optimization for use with Rewind. Finally, we found that hybridization and wash buffers without citric acid, heparin, Denhardt's solution or Tween-20 (that is, using only SSC, formamide and dextran sulfate) appeared to work as well as the manufacturer's recommended buffers for barcode RNA HCR, and we used these minimal buffers for barcode detection before immunofluorescence (Fig. 3).

Barcode RNA ClampFISH. For Supplementary Fig. 8, we adapted the published ClampFISH protocol[38] for labeling barcode RNA as follows. We generated modified primary probes and amplifier probes as described in 'Barcode RNA FISH probe design'. For hybridization, we rinsed fixed samples with washing buffer containing 40% formamide in 2× SSC and then applied the primary ClampFISH probes in primary hybridization buffer containing 40% formamide, 10% dextran sulfate, 1 mg ml^{-1} of yeast transfer RNA (tRNA) (Invitrogen, 15401029), 0.02% BSA and 100 µg ml^{-1} of sonicated salmon sperm DNA (Agilent, 201190-81) in 2× SSC. We included up to 180 ClampFISH probes targeting up to 60 different barcode RNA sequences per hybridization (total probe concentration, 125–250 ng µl^{-1}). We added coverslips to samples and then hybridized for 6–8 h in humidified containers at 37 °C. After hybridization, we added wash buffer containing 40% formamide in 2× SSC to dislodge coverslips and then replaced the wash buffer and incubated the samples for 20 min at 37 °C. We performed a second wash for 20 min at 37 °C using buffer containing 20% formamide and 2× SSC and then performed the second round of hybridization with MM2B and MM2C amplifier probes in amplifier hybridization buffer (20% formamide, 10% dextran sulfate, 1 mg ml^{-1} of yeast tRNA and 0.02% BSA, in 2× SSC; final probe concentration, 10 ng µl^{-1} each). After the second hybridization, we washed samples 2 × 20 min at 37 °C using buffer containing 20% formamide and 2× SSC and then rinsed the sample with 2× SSC. We then performed the copper(I)-catalyzed azide–alkyne cycloaddition ('click' reaction) by adding a solution containing 150 µM BTTAA, 75 µM copper sulfate, 2.5 mM L-ascorbic acid and 0.1% Triton-X 100 in 2× SSC to each sample and incubating at 37 °C for 15–20 min. To prepare this solution, we first combined the BTTAA and copper sulfate, added the 2× SSC containing 0.1% Triton-X and, lastly, added freshly dissolved L-ascorbic acid (19–20 mg of L-ascorbic acid sodium salt dissolved in 1 ml of nuclease-free water). Once the L-ascorbic acid is added, we immediately add the solution to our samples. After the click reaction, we rinsed samples once with 2× SSC and then washed 1 × 20 min at 37 °C with buffer containing 40% formamide in 2× SSC. After this wash, we performed the third round of hybridization with P9B and P9C amplifier probes in the amplifier hybridization buffer, followed by washes and click and post-click wash as described above. We continued with additional amplifier hybridizations (iterating between using MM2B+MM2C amplifier probes on even rounds and P9B+P9C amplifier probes on odd rounds) and washes, performing the click reaction during every odd round (3, 5, 7…).

After the post-click wash for round 7 or round 9, we added RNA FISH hybridization buffer (10% formamide and 10% dextran sulfate in 2× SSC) containing probes targeting P9B and P9C and coupled to Alexa Fluor 594 and Atto647n, respectively (see Supplementary Table 4 for sequences). We hybridized these probes overnight in humidified containers at 37 °C and then washed samples 2 × 30 min with washing buffer (10% formamide and 2× SSC) at 37 °C, adding DAPI to the second wash to stain the nuclei. After these washes, we rinsed samples once with 2× SSC and then replaced the 2× SSC and proceeded with imaging. To remove ClampFISH signal, we stripped dye-coupled probes as described above for RNA FISH.

Immunofluorescence. We performed immunofluorescence using primary antibodies targeting total ERK (L34F12 Cell Signaling, 4696) and pERK (p44/p42 ERK D12.14.4E Cell Signaling, 4370). First, we rinsed cells three times with 5% BSA in PBS (5% BSA-PBS) and then incubated at room temperature for 2 h in 5% BSA-PBS containing 1:100 total ERK and 1:200 pERK antibodies. Next, we washed the cells 5 × 5 min with 5% BSA-PBS and then incubated the cells at room temperature for 1 h in 5% BSA-PBS containing 1:500 donkey anti-mouse secondary antibody conjugated to Cy3 (Jackson ImmunoResearch, 715-165-150) and 1:500 goat anti-rabbit secondary antibody conjugated to Alexa Fluor 594 (Cell Signaling, 8889). After the secondary incubation, we washed the cells 5 × 5 min with 5% BSA-PBS containing 50 ng ml^{-1} of DAPI and then replaced the wash buffer with 2× SSC and proceeded with imaging as described below.

RNA FISH and immunofluorescence imaging. We imaged RNA FISH samples on an inverted Nikon TI-E microscope equipped with a SOLA SE U-nIR light engine (Lumencor), an ORCA-Flash 4.0 V3 sCMOS camera (Hamamatsu), ×20 Plan-Apo λ (Nikon MRD00205), ×40 Plan-Fluor (MRH00401) and ×60 Plan-Apo λ (MRD01605) objectives and filter sets for DAPI, Cy3, Alexa Fluor 594 and Atto647N. For barcode ClampFISH and barcode HCR, we first acquired tiled images in a single *z*-plane (scan) at ×20 or ×40 magnification, and, then, after identifying positions containing cells positive for resistant barcodes, we returned to those positions to acquire a *z*-stack at ×60 magnification. For subsequent rounds of single-molecule RNA FISH and ERK immunofluorescence, we acquired *z*-stacks at ×60 magnification. For scans, we used a Nikon Perfect Focus system to maintain focus across the imaging area.

Image analysis. To identify barcode RNA FISH-positive and GFP-positive cells in Supplementary Fig. 1f–h, we used custom MATLAB scripts to first stitch together scanned images and then identify individual cells using the DAPI nuclear signal. Next, we used a custom graphical user interface (GUI) to zoom in on the stitched image, view the barcode RNA FISH (Alexa647) signal and interactively select barcode RNA FISH-positive cells. After selecting all barcode RNA FISH-positive cells, we repeated the same process with GFP signal to select all GFP-positive cells without knowledge of the cells' barcode RNA FISH status. We then extracted the spatial coordinates, barcode RNA FISH status and GFP status for all cells and plotted the results using custom R scripts available on Dropbox at https://www.dropbox.com/sh/u4sibi0fgorzk0p/AACmLLvqf0iY9GlZBzzuVbtTa?dl=0. MATLAB scripts for stitching scans and the custom GUI are available at https://github.com/arjunrajlaboratory/timemachineimageanalysis.

To identify barcode RNA FISH-positive cells for Rewind, we used custom MATLAB scripts to stitch, contrast and compress scan images (scripts available at https://github.com/arjunrajlaboratory/timemachineimageanalysis) and then manually reviewed these stitched images. This review yielded positions containing candidate barcode RNA FISH-positive cells that we then re-imaged for verification at ×60 magnification in multiple *z*-planes. If we were uncertain about the fluorescence signal in a candidate cell (for example, abnormal localization pattern or non-specific signal in multiple channels), we excluded the cell from imaging during subsequent rounds of RNA FISH or immunofluorescence.

For quantification of RNA FISH images, we used custom MATLAB software available at https://github.com/arjunrajlaboratory/rajlabimagetools. Briefly, the image analysis pipeline includes manual segmentation of cell boundaries, thresholding of each fluorescence channel in each cell to identify individual RNA FISH spots and then extraction of spot counts for all channels and cells. After extracting spot counts, we analyzed RNA levels across single cells using custom R scripts available at https://www.dropbox.com/sh/u4sibi0fgorzk0p/AACmLLvqf0i

Y9GlZBzzuVbtTa?dl=0. In all figures, box plots indicate the 25th, 50th and 75th percentiles, with whiskers extending to 1.5 times the interquartile range. Notably, for some markers, we were not able to quantify expression in a few cells because of grossly abnormal or non-specific fluorescence signal (that is, schmutz) or because we lost a cell during sequential hybridizations. We excluded data from these cells from analyses, and, as a result, some plots might contain slightly different numbers of points for different markers. For analyses involving dimensionality reduction (UMAP) or clustering, we included only cells with data for all assayed markers.

For the UMAP visualizations, we used the Seurat v3.2.0 package (the versions of all dependent packages are documented in the plotting scripts at https://www.dropbox.com/s/v66v41zryogmd78/RsessionInfo.txt?dl=0)[40,41]. For the analysis shown in Fig. 4, we ran the UMAP algorithm on scaled RNA FISH data using the first five principal components and setting n_neighbors = 30 and min_dist = 0.3 (default settings). For the analyses shown in Figs. 5 and 6, we used the first six principal components and set min_dist = 0.6 to better visualize the number of cells expressing high levels of *DEPTOR*.

We adapted the RajLabImageTools pipeline for quantifying immunofluorescence images. After manually segmenting cells, we used custom MATLAB scripts to average fluorescence intensity within cell boundaries for each channel and then took the maximum average fluorescence intensity across *z*-planes. We additionally used DAPI signal to automate nuclei segmentation and separately quantified cytoplasmic and nuclear immunofluorescence intensity. We found qualitatively similar results for both cytoplasmic and nuclear ERK immunofluorescence quantification (Supplementary Fig. 8).

For quantification of cell and colony numbers following vemurafenib treatment, we used custom MATLAB software available at https://github.com/arjunrajlaboratory/colonycounting_v2. The analysis pipeline involves stitching the tiled DAPI images, manually segmenting individual wells and colonies, identifying individual cells based on DAPI signal and then extracting cell counts from the entire well and each colony. We analyzed the extracted cell counts using custom R scripts available at https://www.dropbox.com/sh/u4sibi0fgorzk0p/AACmLLvqf0iY9GlZBzzuVbtTa?dl=0. We used a separate MATLAB script (https://www.dropbox.com/s/xnwtmw8rh8ec3ij/countCellsTimeMachineScans.m?dl=0) to quantify the number of cells imaged in our Carbon Copies.

To assign individual primed cells (marked by barcode RNA FISH signal) to subclones (Supplementary Fig. 10), we first extracted the spatial position of each image in the whole-well scans containing at least one primed cell. We then calculated the Euclidean distance between these images and used these distances to perform hierarchical clustering. Visual inspection of the clustering revealed a clear distance threshold of <2 mm for grouping subclones of closely related (and, therefore, neighboring) primed cells; thus, all primed cells within a group were assigned to the same subclone. To further check our subclone assignments, we manually inspected all barcode RNA FISH images and found that primed cells assigned to the same subclone had similar barcode RNA FISH signal intensity and intracellular patterns, whereas this signal similarity was not observed for primed cells assigned to different subclones. Most primed cells from different subclones were at least 7 mm apart, and, for the few cases of primed cells located between 2 mm and 7 mm apart, we observed that these cells had distinct barcode RNA FISH signal patterns consistent with them belonging to separate subclones. This clear spatial separation gave us confidence in our ability to accurately assign individual cells to particular subclones.

RNA sequencing and analyses. We extracted RNA from fixed cells after barcode RNA FISH and sorting using the NucleoSpin total RNA FFPE XS kit (Takara). We performed cell lysis and reverse cross-linking at 50 °C for 90 min and otherwise followed the manufacturer's protocol. After RNA extraction, we prepared sequencing libraries using the NEBNext single-cell/low-input RNA sequencing library preparation kit for Illumina (NEB) and then performed paired-end sequencing of these libraries (38 cycles read 1 + 37 cycles read 2) on a NextSeq 500 (Illumina). After sequencing, we aligned reads to the human genome (assembly 19; hg19) using STAR[42] v2.5.2a and counted uniquely mapped reads with HTSeq[43] v0.6.1.

We performed differential expression analysis in R v3.6.3 using DESeq2 (ref. [44]) v1.22.2 and with data from at least two biological replicates for each sample and condition. Biological replicates were sorted on separate days using distinct barcode RNA FISH probe sets. We considered a gene to be differentially expressed if the comparison between two conditions yielded a $\log_2$ fold change of ≥ 1 or ≤ -1 and adjusted *P* value (P_{adj}) ≤ 0.1. For determining candidate markers for primed cells requiring DOT1L inhibition (Fig. 5), we compared primed and non-primed subpopulations sorted from both DOT1L inhibitor and vehicle control Carbon Copies and modeled the biological replicate and DOT1L inhibitor treatment as covariates in the design formula for DESeq2. We chose to include data from both DMSO- and DOT1L-inhibitor-treated Carbon Copies (two replicates each) in the analysis and model DOT1L inhibitor treatment as a covariate due to the modest effects of DOT1L inhibitor treatment alone on gene expression (Fig. 6e,f, Supplementary Fig. 14d,e and Supplementary Fig. 16c,d) and our particular interest in identifying gene expression markers that distinguish various subpopulations of primed cells. We performed hierarchical clustering and principal component analysis on $\log_2$-transformed transcripts per million (TPM) values using R v3.6.3.

We tested for enrichment of differentially expressed genes among gene ontologies and pathways (KEGG, REACTOME and WikiPathway) using WebGestaltR. If a differentially expressed gene was included in one or more enriched Gene Ontology term or pathway, we chose a consensus annotation (for example, ECM organization and cell migration) for that gene. Otherwise, we assigned gene annotations by manual review. Our resulting gene annotation can be found in Supplementary Table 8.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All RNA sequencing data generated for this study are available at the Gene Expression Omnibus (accession no. GSE161300). Additional sequencing and imaging data are available on Dropbox at https://www.dropbox.com/sh/mmeg3mckrpridu3/AAALBaMLoJsJiQC2-lrVY0Cva?dl=0 and upon reasonable request to the corresponding author.

Code availability

Software used to segment cells and quantify RNA spots is available at https://github.com/arjunrajlaboratory/rajlabimagetools. Software used to stitch, segment and quantify scan images of resistant colonies is available at https://github.com/arjunrajlaboratory/colonycounting_v2. Additional custom image analysis scripts are available at https://github.com/arjunrajlaboratory/timemachineimageanalysis. The pipeline used for barcode sequencing analysis is available at https://github.com/arjunrajlaboratory/timemachine.

References

35. Hsu, M.-Y., Elder, D. E. & Herlyn, M. Melanoma: the Wistar melanoma (WM) cell lines. In: *Human Cell Culture* 259–274 (Springer, 2002).
36. Zorita, E., Cuscó, P. & Filion, G. J. Starcode: sequence clustering based on all-pairs search. *Bioinformatics* **31**, 1913–1919 (2015).
37. Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).
38. Rouhanifard, S. H. et al. ClampFISH detects individual nucleic acid molecules using click chemistry-based amplification. *Nat. Biotechnol.* **37**, 84–89 (2018).
39. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
40. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
41. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
42. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
43. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements

We thank C. Bartman, A. Anguierra, J. Murray, N. Zhang, L. Cai, B. Stanger, A. Coté, K. Kiani, E. Sanford and Y. Goyal, along with other members of the Raj laboratory, for many useful suggestions. We thank the Flow Cytometry Core Laboratory at the Children's Hospital of Philadelphia Research Institute for assistance in designing and performing FACS, including F. Tuluc for several helpful discussions. B.L.E. acknowledges support from NIH training grants F30 CA236129, T32 GM007170 and T32 HG000046; E.A.T. acknowledges support from R01 CA238237; I.P.D. acknowledges support from NIH 4DN U01 HL129998 and the NIH Center for Photogenomics (RM1 HG007743); C.L.J. acknowledges support from NIH T32 DK007780 and F30 HG010822; N.J. acknowledges support from NIH F30 HD103378; S.M.S. acknowledges support from DP5 OD028144; and A.R. acknowledges support from R01 CA238237, NIH Director's Transformative Research Award R01 GM137425, R01 CA232256, NSF CAREER 1350601, P30 CA016520, SPORE P50 CA174523, NIH U01 CA227550, NIH 4DN U01 HL129998, the NIH Center for Photogenomics (RM1 HG007743) and the Tara Miller Foundation.

Author contributions

B.L.E. and A.R. conceived and designed the project with input from E.A.T. and S.M.S. B.L.E. performed experiments and analyses. I.P.D., C.L.J. and N.J. assisted

with optimizing barcode RNA FISH protocols. C.J.C. assisted with cell and colony segmentation. A.R. supervised the project. B.L.E. and A.R. wrote the manuscript with feedback from S.M.S., C.L.J., N.J. and other members of the Raj laboratory.

Competing interests

A.R. receives consulting income, and A.R. and S.M.S. receive royalties, related to Stellaris RNA FISH probes.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-00837-3.

Correspondence and requests for materials should be addressed to A.R.

Peer review information *Nature Biotechnology* thanks **Leor Weinberger** and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Arjun Raj

Last updated by author(s): Dec 18, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We performed all sequencing for this study using a NextSeq500 and downloaded the demultiplexed fastq files from BaseSpace. We acquired imaging data using a Nikon Ti-E microscope equipped with a SOLA SE U-nIR light engine (Lumencor), an ORCA-Flash 4.0 V3 sCMOS camera (Hamamatsu), 20X Plan-Apo λ (Nikon MRD00205), 40X Plan-Fluor (MRH00401) and 60X Plan-Apo λ (MRD01605) objectives, and filter sets for DAPI, Cy3, Alexa Fluor 594, and Atto647N.
Data analysis	Scripts for all analyses presented in this paper, including all data extraction, processing, and plotting steps are accessible at the following link https://www.dropbox.com/sh/mmeg3mckrpridu3/AAALBaMLoJsJiQC2-lrVY0Cva?dl=0. For RNA sequencing and barcode sequencing analyses, we used the following open source software: STAR (v2.5.2a), HTSeq (v0.6.1), DESeq2 (v1.22.2) and STARCODE (v1.3). The pipeline used for barcode sequencing analysis is available at https://github.com/arjunrajlaboratory/timemachine. Software used to segment cells and quantify RNA spots is available at https://github.com/arjunrajlaboratory/rajlabimagetools. Software used to stitch, segment and quantify scan images of resistant colonies is available on github at https://github.com/arjunrajlaboratory/colonycounting_v2. Additional image analysis scripts are available at https://github.com/arjunrajlaboratory/timemachineimageanalysis. To run UMAP, we used Seurat (v3.2.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The RNA FISH and immunofluorescence data (raw images and processed data), barcode sequencing data, and RNA-sequencing data that support the findings of this study are available on Dropbox at https://www.dropbox.com/sh/mmeg3mckrpridu3/AAALBaMLoJsJiQC2-IrVY0Cva?dl=0. In addition, the RNA sequencing data are available on GEO (accession #GSE161300).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not perform calculations to predetermine sample-sizes for the reported experiments. For all RNA sequencing experiment, we collected samples from at least 2 experiments performed on separate days and used a conservative threshold (abs(fold change) ≥ 2 and adjusted p-value ≤ 0.1) for determining significant differences in gene expression. For imaging-based experiments (RNA FISH and immunofluorescence), we analyzed as many barcode FISH positive (primed) cells as possible (constrained by the rarity of the population) aiming for at least 100 primed cells per experiment. We analyzed a similar number of barcode FISH negative (non primed) cells.
Data exclusions	RNA FISH and immunofluorescence images with grossly abnormal or non-specific fluorescence signal were excluded from all analyses. In addition, cells with missing data for one or more markers measured by RNA FISH (due to the above exclusion) were excluded from dimensionality reduction (UMAP) and clustering analyses. For RNA sequencing analyses, we used DESeq2's collapseReplicates function to combine data from 2-3 technical replicates of non-primed cells sorted from the same Carbon Copy. In 3 experiment, 1 of the 2-3 technical replicates showed highly skewed TPM values for 1,000s of genes (compared to other samples from the same experiment) and were excluded from further analyses. These exclusion criteria were not pre-established.
Replication	The data presented in Fig. 4 and related Supplementary Figs. 9-10 are derived from one independent experiment. Similarly, the data presented in Supplementary Fig. 1b, g, are derived from one independent experiment. We repeated all other experiments reported in this manuscript at least twice and report the results from all replicates.
Randomization	We allocated cells into experimental groups based on barcode RNA FISH signal and drug treatments. To profile barcode RNA FISH negative (non-primed) cells, we imaged random positions as described in the Methods.
Blinding	Blinding was not relevant for this study nor was it possible during data collection.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

anti-AXL (AF154 from Novus Biologicals), anti-ITGA3 (DSHB clone P1B5), anti-total ERK (L34F12 Cell Signaling #4696), anti-phosphorylated-ERK (p44/p42 ERK D12.14.4E Cell Signaling #4370), bovine anti-goat conjugated to Alexa Fluor 647 (Jackson ImmunoResearch 805-605-180), anti-mouse FAb2 conjugated to Alexa Fluor 488 (Cell Signaling #4408), donkey anti-mouse conjugated to Cy3 (Jackson ImmunoResearch 715-165-150), goat anti-rabbit conjugated to Alexa Fluor 594 (Cell Signaling #8889).

Validation

We checked the specificity of AXL and ITGA3 antibodies by sorting high- and low-expressing subpopulations then performing RNA FISH on sorted subpopulations to compare levels of AXL and ITGA3 RNA, respectively. These data are included in the Supplemental Figures. We validated RNA FISH probes by observing spot colocalization of divided probe sets ("odds" and "evens") coupled to distinct fluorophores. As validation of the phosphorylated-ERK antibody, we noted markedly decreased immunofluorescence signal in vemurafenib treated WM989 A6-G3 cells compared to untreated cells (Figure 3 and Supplementary Fig. 8). In contrast, we observed minimal difference in total ERK immunofluorescence signal between vemurafenib treated and untreated WM989 A6-G3 cells. The anti-total ERK antibody was validated for immunofluorescence on human cells by the manufacturer (Cell Signaling) and cited by >300 references on the manufacturer's website.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

We received the WM989 and WM983b cell lines from Meenhard Herlyn (Wistar Institute). These melanoma cell lines are also available from Rockland inc. (WM989-01-000 and WM983B-01-0001). We received the HEK293FT cell line from Roberto Bonasio (University of Pennsylvania). HEK293FT are also available from Invitrogen (R70007).

Authentication

Short tandem repeat DNA profiling through the Wistar Institute.

Mycoplasma contamination

All cell lines tested negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used in this study.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

For each antibody used for FACS, a detailed staining protocol is described in the Methods. Briefly, for anti-AXL staining, we diluted the primary antibody 1:50 in 1%BSA-PBS and the secondary antibody 1:60 in 1%BSA-PBS. For anti-ITGA3 staining, we diluted the primary antibody 1:200 in 0.1%BSA-PBS and the secondary antibody 1:500 in 0.1%BSA-PBS.

Instrument

FACSJazz (BD Biosciences) and MoFloAstrios (Beckman Coulter)

Software

We used FlowJo v10 to analyze FACS data and generate plots for figures. The raw data (fcs files) and original plots from the sorts are available on Dropbox at https://www.dropbox.com/sh/jdp2oz5rd36mmr7/AABQjx0_8zxv1JLQZvY16RNja?dl=0

Cell population abundance

For anti-ITGA3 stained WM989 A6-G3 cells, we assessed the purity of the ITGA3-High subpopulations by measuring expression of ITGA3 in single cells via RNA FISH (Supplementary Fig. 4). On average across replicates, 86% of ITGA3-High sorted cells contained high levels of ITGA3 RNA (>50th percentile of expression in ITGA3-Low cells). For anti-AXL stained WM983b E9-C6 cells, we assessed the purity of the AXL-High subpopulations by measuring expression of AXL in single cells via RNA FISH (Supplementary Fig. 6). On average across replicates, ~67% of AXL-High sorted cells contained high levels of AXL RNA (>50th percentile of expression in AXL-Low cells). For for barcode RNA FISH labeled cells (Rewind experiments), we assessed the purity of the primed subpopulations by performing targeted sequencing of barcodes from cDNA and measuring the proportion of reads matching targeted barcodes. These proportions are reported separately for each experiment in Figure 1 and Supplemental Figures).

Gating strategy

For all FACS experiments, we first excluded doublets by gating on SSC/Trigger Pulse Width and then excluded debris by gating on SSC/FSC. For live-cell sorts (AXL and ITGA3 staining), we excluded dead cells by gating on DAPI, PI or 7-AAD signal. For anti-ITGA3 (Alexa Fluor 488) stained WM989 A6-G3, we gated the brightest 0.3-0.4% (ITGA3-High) and dimmest ~99% (ITGA3-Low) subpopulations. For anti-AXL (Alexa Fluor 647) stained WM983b E9-C6, we gated the brightest 0.3-0.4% (AXL-High) and dimmest ~20% (AXL-Low) subpopulations. For barcode RNA FISH labeled cells (Rewind experiments), we gated the brightest ~.05% (primed) and dimmest ~99% (non-primed) subpopulations.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.