Retrospective identification of rare cell populations underlying drug resistance connects molecular variability with cell fate

Authors:
Benjamin L. Emert[2], Christopher Coté[1], Eduardo A. Torre[2], Ian P. Dardani[1], Connie L. Jiang[2], Naveen Jain[2], Sydney M. Shaffer[1,3] and Arjun Raj[1,*]

[1]Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania
[2]Perelman School of Medicine, University of Pennsylvania
[3]Department of Pathology, Perelman School of Medicine, University of Pennsylvania
*Corresponding author

Abstract:
Molecular differences between individual cells can lead to dramatic differences in cell fate following an applied treatment, such as the difference between death versus survival of cancer cells upon receiving anti-cancer drugs. However, current strategies to retrospectively identify the cells that give rise to distinct rare behaviors and determine their distinguishing molecular characteristics remain limited. Here we describe Rewind, a methodology that combines genetic barcoding with an RNA-based readout to directly capture rare cells that give rise to cellular behaviors of interest, specifically the emergence of resistance to targeted cancer therapy. Using Rewind, we analyzed over 5 million cells to identify differences in gene expression and MAP-kinase signaling in single melanoma cells that mark a rare subpopulation of drug-naive cells (initial frequency of ~1:1000-1:10,000 cells) that ultimately gives rise to drug resistant clones. We further show that even within this rare subpopulation, molecular differences between single cells before the application of drug predict future differences in drug resistant behavior. Similarly, we show that treatments that modify the frequency of resistance can allow otherwise non-resistant cells in the drug-naive population to become resistant, and that these new populations are marked by the variable expression of distinct genes. Together, our results reveal the presence of cryptic variability that can underlie a range of distinct rare-cell phenotypic outcomes upon drug exposure. Applying Rewind to other rare biological phenomena, such as cancer metastasis, tissue regeneration, and stem cell reprogramming, may provide a means to map rare cellular states to the unique cellular fates to which they give rise.

Introduction:

Individual cells—even those of ostensibly the same cell type—can differ from each other in a number of ways. Some of these differences can result in a "primed" cellular state that can, in a particular context, ultimately lead to biologically distinct behaviors (Raj and van Oudenaarden, 2008; Symmons and Raj, 2016). 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 (Gupta et al., 2011; Roesch et al., 2010; Shaffer et al., 2017; Sharma et al., 2010). Yet, while this phenomena implies 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 (Gupta et al., 2011; Shaffer et al., 2018). What remains largely unknown, outside of a few cases (Cohen et al., 2008; Shaffer et al., 2017; Spencer et al., 2009), is 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? 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.

Currently, the two general methodologies for longitudinal tracking and profiling are direct visualization by time lapse microscopy and genetic encoding of lineage information (i.e., "barcoding") (Biddy et al., 2018; Cohen et al., 2008; Raj et al., 2018; Sigal et al., 2006). Time lapse microscopy allows one to directly follow cells over time, providing a definitive cellular lineage, high resolution temporal information, and a direct connection to fate (Cohen et al., 2008; Sigal et al., 2006). However, monitoring the molecular processes occurring in these cells by time-lapse remains challenging: fluorescent reporters of gene expression and signaling activity are difficult to introduce and validate, and the ability to multiplex them is generally limited. Barcoding, in which cells are labeled by unique and sometimes mutable nucleic acid sequences (Alemany et al., 2018; Frieda et al., 2016; Kalhor et al., 2018; McKenna et al., 2016; Raj et al., 2018), allows one to track cells by sequencing. When combined with single cell RNA sequencing methods, one can connect cellular lineages to transcriptomic profiles (Al'Khafaji et al., 2019; Biddy et al., 2018; Hurley et al., 2020; Weinreb et al., 2020). These techniques, however, are difficult to combine with assays that measure molecular features outside of gene expression levels, which may be important to distinguish key subpopulations (Weinreb et al., 2020; Wu et al., 2020). A key challenge for both of these methodologies is the detection of rare cells (1:1000 or even more rare), for which neither time-lapse nor single cell RNA sequencing is particularly effective (new techniques aim to circumvent these limitations; (Al'Khafaji et al., 2018;

Feldman et al., 2019). Yet, many biological phenomena, such as therapy resistance in cancer cells, occur in subpopulations that are at least that rare.

Here, we combine cellular barcoding with RNA Fluorescent In Situ Hybridization (RNA FISH) to selectively isolate rare cells that adopt distinct cellular fates. We apply this methodology, which we call Rewind, to targeted therapy resistance in melanoma, revealing prospective expression markers of cells primed for resistance upon $BRAF^{V600E}$ inhibition. Further, we show that differences in resistance outcome can be traced to distinct cell subpopulations that can be discriminated on the basis of their transcriptome profiles. These findings suggest that cryptic single cell variability within otherwise homogeneous cells can lead to important differences in ultimate cellular behavior in response to a treatment.

Results:

We developed Rewind to profile the rare melanoma cells that ultimately become resistant to targeted therapies in melanoma. We used the $BRAF^{V600E}$ mutated melanoma cell line WM989 A6-G3, in which we have previously demonstrated that there is a rare, transient subpopulation of cells (~1:2000) that are "primed" to survive treatment to the targeted therapy vemurafenib (Shaffer et al., 2017, 2018; Torre et al., 2019). These rare, primed cells often express 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, thus providing a testbed for evaluating Rewind. At the same time, Rewind would enable us to directly determine the connection between variability in the initial cell population and variability in cell fate following vemurafenib treatment.

The primary technical challenge for studying rare cell processes like drug resistance is the rarity of 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 (Biddy et al., 2018; Weinreb et al., 2020). We developed Rewind as a method to directly isolate or identify specifically 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) mRNA (Figure 1A and Supp. Fig 1A). After labeling cells with these barcodes, we allowed the cells to divide into "twins", and then separate the twins into two populations such that each group contained at least one twin. One population 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 (e.g., 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 labelling the RNA transcribed from those specific barcodes using RNA in situ hybridization techniques (see Methods and Supp. Fig 1B-1E). 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 Supp. Fig 2 for experimental details and calculations). We can then molecularly profile the Carbon Copy twins to determine what is different about their initial state that led to their subsequent distinct fate, a methodology we dubbed Rewind for its ability to retrospectively connect the initial state of rare cells with the cells' future fate.

Rewind critically relies on the cells maintaining ("remembering") their primed state through several cell divisions so that the profile of cells isolated from the Carbon Copy 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. We then separately treated both halves of the population with 1μm vemurafenib for three weeks and sequenced genomic DNA from surviving cells to identify their barcodes (Supp. Fig 3A). If the primed cell state giving rise to vemurafenib resistance were maintained through these cell divisions, the twins of cells that survived in one culture would also survive in the parallel culture, in which case the same barcodes should appear upon DNA sequencing of each culture. Consistent with memory over this time scale, 67% percent of the top 200 most abundant barcodes from each culture were identical, despite an expectation of ≤0.002% by chance alone (See Supp. Fig 3 and Methods for dependence on parameters and probability calculations; we expected roughly ~100-500 resistant colonies in each culture). The absence of perfect (100%) overlap may reflect differences in relative growth between twin cells, multiple sources of

technical variability, or slow reversion out of the primed cell state as we have reported previously (Shaffer et al., 2018). As expected, we find no overlap in the most abundant barcode sequences from independently transduced cells treated with vemurafenib (Supp. Fig 3F).

We then applied the Rewind approach to isolate the rare WM989 A6-G3 cells primed for vemurafenib resistance for RNA-sequencing (see methods for details). The first step of Rewind was to identify the barcodes marking the vemurafenib-resistant cells then design RNA FISH probes targeting the most abundant barcodes. Using these probes, we performed RNA FISH (together with signal amplification) on the Carbon Copy population fixed prior to vemurafenib treatment and isolated the brightest ~0.05% of cells expressing resistant barcodes by FACS for transcriptome profiling by RNA sequencing (Fig. 1 and Supp. Fig. 4). Sequencing the cDNA from this subpopulation yielded 33 of the 60 (55%) barcodes we had probed for, demonstrating a greater-than-thousand-fold enrichment of the targeted population over the 60/~200,000 (0.03%) frequency of these barcodes in the overall population (Fig. 1B).

Using the RNA sequencing data, we looked for genes that were differentially expressed in the Carbon Copy cells expressing resistant barcodes to find marker genes specific to cells primed for vemurafenib resistance. Consistent with previous research from our lab and others we found that primed cells sorted from the Carbon Copy expressed greater than 2-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* (Supp. Fig 4B) (Müller et al., 2014; Shaffer et al., 2017). Beyond these known markers, the transcriptome profile provided by Time Machine 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, Supp. Fig 4C 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, while an additional ~2,800 genes showed a greater than 2-fold change in expression during this period (Supp. Fig 5). 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.

To verify that the expression of these newly identified genes specifically marked cells primed for resistance, we sorted drug-naive WM989 A6-G3 cells expressing high levels of *ITGA3* (the most differentially expressed cell surface marker not previously known to be a marker of vemurafenib resistance), and tested their response to vemurafenib. After 18 days of treatment, ITGA3-High cells gave rise to 10-fold more resistant colonies than the rest of the population, confirming *ITGA3* as a marker of the primed cells state (Fig 1D and Supp. Fig 4D-F). Encouraged by these results, we applied Rewind to another melanoma line, WM983b E9-C6, in which markers of the cells primed for resistance were unknown. RNA sequencing identified the gene *AXL* as more highly expressed in the primed subpopulation. Subsequent isolation of AXL-High cells by FACS followed by addition of vemurafenib enriched 4-fold for resistance (Supp. Fig 6). Together these results demonstrate that Rewind can be used to discover new markers of primed cellular states that give rise to drug resistance.

One question is whether each individual primed cell expressed one or a few distinct markers, potentially reflecting “noisy” gene expression, or whether most primed cells expressed most markers simultaneously, reflecting a more coordinated cell state. Having used Rewind to identify

markers of primed cells in aggregate, we next sought to characterize the co-variation in the expression of these markers that underlies the primed cell state in single cells with RNA imaging. Using Rewind 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 9 genes by single molecule RNA FISH (see Methods for details). 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-seq (Fig 2D). Moreover, cells primed for resistance were far more likely to co-express any pair of markers (Odds Ratios ranging from ~1.5 to ≥58; Supp Fig 7), and ~87% percent of cells expressed high levels of ≥4 of 7 marker genes simultaneously, in stark contrast to cells not expressing resistant barcodes (Fig 2E and Supp Fig 7). This apparent coordination suggests that the cell-to-cell differences that lead to distinct cell fates following drug treatment are a consequence of the coordinated fluctuations of several factors simultaneously, as opposed to sporadic fluctuations of individual genes (Shaffer et al., 2018).

Given that we observed increased expression of multiple receptor tyrosine kinases and their cognate ligands in primed cells, we asked whether these cells also exhibited differences in MAPK pathway activation. In a Carbon Copy fixed before vemurafenib treatment, we found similar levels of phosphorylated ERK (pERK) were similar in primed cells vs. non-primed cells (Figure 3C and Supp. Fig 8). However, in a Carbon Copy population that underwent vemurafenib treatment for 24 hours, 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 5-fold higher than non-primed cells (within the range of untreated cells; Fig 3C and Supp. Fig 8). We also observed that within individual clusters of closely related primed cells, not all cells contained higher levels of pERK, which may reflect pulsatile changes in pERK as documented elsewhere (Supp. Fig 8D) (Gerosa et al., 2019). 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.

Following prolonged drug treatment, it was visually clear that different colonies of vemurafenib-resistant cells can show dramatic differences in basic properties like the number of cells in the colony. We wondered whether these differences in fate could also be traced back to differences in the initial molecular state of the primed cells before the addition of drug. (Indeed, our earlier “heritability” split experiments showed a strong correlation in barcode abundance between the parallel cultures, suggesting that heritable differences in initial cell states can predict the extent of proliferation after drug treatment; Supp. Fig 3E). To look for these initial differences, 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. We then designed RNA FISH probes that distinguished 30 of the most abundant barcodes (i.e., “highly resistant”, meaning many resistant cells) from 30 barcodes in the next tier of abundance (i.e., “less resistant”). We used these probes to identify their twin cells in a Carbon Copy fixed prior to vemurafenib treatment (Fig 4A - 4F and Supp. Fig 9 for probe set validation).

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 9 genes, including 7 priming markers, *MITF* and *SOX10*. We used the dimensional reduction technique 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 signal (primed to become highly vs. less resistant vs. 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, while the less resistant cells formed two distinct groups.

We then asked how expression levels of various genes corresponded to these groupings. As expected, most (>80%) of the primed cells had markedly decreased levels of both *SOX10* and *MITF* (Figure 4D and Supp. Fig 7). 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 (Figure 4D and Supp. Fig 7). The cells primed for being highly resistant expressed the highest levels of *AXL*, *ITGA3*, and to some extent *EGFR*. One group (group A) of the cells primed to be less resistant expressed the highest levels of WNT5A and MMP1, whereas the other (group B) expressed the highest levels of NGFR (NGFR also had intermediate levels of expression in the cells primed to be highly resistant; Fig 4E). Quantitative comparison of expression levels between pairs of markers showed that the expression of, for example, AXL vs. MMP1 fell along two separate axes of variability (Fig 4E). Together, these analyses suggest that multiple classes of primed cells with different expression patterns give rise to resistant colonies with different phenotypes.

We then wondered whether there were additional types of initial cell states that could be primed for resistance under different conditions, in principle marked by the expression of unique sets of genes. 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 (Torre et al., 2019). Amongst these is DOT1L, a H3K79 methylase we identified by genetic screening whose inhibition leads to a 3-fold increase in the number of resistant colonies that form upon addition of vemurafenib. While 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.

While we thought it most likely that DOT1L inhibition was enabling a new subset of cells to become vemurafenib resistant, in principle there were three possible explanations for the increased number of primed cells: 1. DOT1L inhibition selectively increases proliferation of primed cells, 2. DOT1L inhibition decreases the rate in which primed cells transition into the non-primed cell state and 3. DOT1L inhibition permits a new subset of cells to enter a primed state (Figure 5A). If either of the first two explanations were correct, we expected that the additional vemurafenib resistant colonies that form following DOT1L-inhibition would contain the same barcodes and arise from the same initial cell state as resistant colonies that grow with or without DOT1L inhibition. In contrast, if DOT1L inhibition permitted a new subset of cells to survive, we expected to find that the additional colonies contained additional barcodes, signifying that they originated from a distinct cell state present in our Carbon Copy (Figure 5A) .

To distinguish these possibilities, we barcoded WM989 A6-G3 and split the culture in two after ~3 cell divisions. We then treated one culture with 4 μm DOT1L-inhibitor (pinometostat) and the other with vehicle control (DMSO) for ~2-3 cell divisions. At that point, we split each culture in two again, fixing half as our Carbon Copies with and without DOT1L-inhibitor pre-treatment, and

treating the other half with 1 μm vemurafenib for 3 weeks to select for resistant cells (Figure 5A and Supp. Fig 10 for diagram of experimental design). Upon sequencing barcodes from vemurafenib resistant cells across all conditions, we found that while some barcodes were highly abundant in both DOT1L-inhibitor and DMSO pre-treated cells (resistant cells not requiring DOT1L inhibition), a separate set of barcodes appeared specifically in the DOT1L-inhibitor pre-treated cells (resistant cells requiring DOT1L inhibition), representing a distinct subpopulation of drug-naive cells that survive vemurafenib treatment only upon DOT1L inhibition (Figure 5A Observed outcome and Supp. Fig 10 and 11). These results suggest that distinct subpopulations are primed for survival in different treatments.

Using Rewind, we sought to identify a gene expression marker 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. (We expect these probe sets to label fewer than 1:10,000 cells.) We used these probes to sort corresponding cells from Carbon Copies fixed prior to vemurafenib treatment (Fig 5B and Supp. Fig 10). RNA sequencing of sorted populations revealed a few genes specific to cells that required DOT1L inhibition to become vemurafenib resistant (Fig 5C-D). Of these, we selected the gene *DEPTOR,* whose expression we sought to characterize in single cells in our Carbon Copy by RNA FISH (Fig 5E-F and Supp. Fig 12). (We also chose another gene, *MGP*, whose expression was similarly highly elevated, but only in one replicate; Supp. Fig 13)

For single cell analysis, we performed RNA FISH on the Carbon Copy (without DOT1L inhibition) for 10 total genes: 6 priming markers, *SOX10*, *MITF*, *DEPTOR*, and *MGP*. We scanned through ~1 million cells to find those expressing the targeted barcodes, identifying 850 such cells. We visualized the expression profiles of these 850 cells using UMAP, overlaying the information provided by the labeled barcode to indicate whether or not they required DOT1L inhibition to become resistant (Fig 5E). 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 like (*AXL*, *EGFR*, etc.; Fig 5E). We initially expected these genes to also have elevated expression 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 profiles of this new subpopulation was far more similar to the general population of cells that was not primed for resistance in either condition (Fig 5E). In our UMAP projection, while many of these cells were grouped together with the non-primed cells, there was another 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 no differences in the expression levels of the other genes measured compared with non-primed cells (Fig. 5E-F). (Cells requiring DOT1L inhibition for priming were also enriched for *MGP* in a separate replicate experiment; Supp. Fig.13.) 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 states that can lead to drug resistance when the DOT1L inhibitor is given prior to vemurafenib. There were also several such cells that did not express *DEPTOR*, potentially due to transient expression of that marker or representing other, uncharacterized subpopulations.

While the cells that require DOT1L inhibition to become vemurafenib resistant expressed distinct markers initially, we wondered whether DOT1L inhibition pushed these cells towards a molecular state more similar to our previously established primed cell state (e.g. high levels of *AXL*, *EGFR*, *NGFR*, etc.; Figure 6A). To this end, we compared the transcript levels as measured by RNA sequencing from cells sorted from Carbon Copies treated either with DOT1L inhibitor or vehicle control (Figure 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 (Figure 6C-6D). With DOT1L inhibition, these two populations appeared modestly more similar transcriptionally, however they remained predominantly distinct (Figure 6C-6D). 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 (Figure 6E-F and Supp. Fig 14). 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 may reveal a separate route to resistance.

Discussion:

We have demonstrated a methodology for directly connecting rare cell behaviors to the initial cell states that give rise to them. Our approach complements recent strategies for connecting single-cell gene expression to cell fate by allowing one to selectively target cells that give rise to rare phenomena and concomitantly measure aspects of cell state beyond gene expression(Al'Khafaji et al., 2019, 2018; Biddy et al., 2018; Feldman et al., 2019; Weinreb et al., 2020). Applying this approach to the phenomena of drug resistance in melanoma, we identified variability in gene expression and MAPK signalling underlying the priming of cells towards the fate of drug resistance.

The global transcriptional profiles afforded by RNA sequencing of these rare primed cells allowed us to ask what pathways might be active in these cells beyond the ones like growth factor receptor signaling that have already been associated with vemurafenib resistance in melanoma (Corcoran et al., 2012; Ji et al., 2015; Shaffer et al., 2017; Sun et al., 2014; Torre et al., 2019). One of the strongest signatures was the upregulation of cell adhesion proteins and structural components of the extracellular matrix. Such signatures suggest the possibility that control of cell state and behavior may have both a component that is autonomous to the cell itself and a component that is instructed by the extracellular matrix. Future research may help reveal if and how the extracellular matrix is able to influence primed cellular states, and consequently, therapy resistance.

The processes involved in the acquisition of stable drug resistance act both on short timescales (such as signaling) and on longer timescales (transcription). For instance, vemurafenib acts by inhibiting MAPK signaling, but the vemurafenib treatment itself relieves negative feedback on growth factor receptor signalling and allows ERK reactivation via $BRAF^{V600E}$-independent routes(Gerosa et al., 2019; Lito et al., 2012). Single cell analysis of ERK signaling has shown that individual cells vary dramatically in ERK activity following vemurafenib treatment with rare cells reactivating ERK to levels comparable 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 had both higher levels of phosphorylated ERK soon after treatment and expressed multiple receptor tyrosine kinases along with their cognate ligands. It is possible that this unique gene expression program enabled autonomous ERK reactivation. Further application of Rewind may help generalize such connections.

Our results reveal some properties of the mapping between initial cellular states and cellular fates upon adding a treatment, but many questions about such mappings remain. The characteristics of the variability that primes cells for different fates upon different stimuli remains mysterious, as does the variability in the fates themselves. 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 rare cell markers *AXL*/*ITGA3*/*NGFR* and *WNT5A*/*MMP1* 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. 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 may help guide

therapy, because 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.

We chose to focus on the priming of melanoma cells towards different fates following 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 (Takahashi and Yamanaka, 2006) and transdifferentiation of one cell type into another. Application of techniques like Rewind in these contexts may reveal universal characteristics of priming and reprogramming.

Materials and Methods:

Barcode Lentivirus Library Construction:

Starting with the LRG2.1T plasmid kindly provided by Dr. Junwei Shi, we derived a lentivirus vector backbone for Rewind by removing the U6 promoter and sgRNA scaffold 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), gel purified the linearized vector. We combined 100ng of linearized vector, 1.08 µL barcode oligo insert (100 nM in nuclease-free water), 10 µL Gibson assembly master mix (NEB E2611) and nuclease free water to a final volume of 20 µL then incubated the reaction at 50ºC for 1 hour. We next column purified the assembled plasmid using Monarch DNA cleanup columns (NEB) according to the manufacturer's protocol then electroporated 2 µL of the column purified plasmid into Endura Electrocompetent *E. coli* cells (Lucigen) using a GenePulserXCell (Biorad) with the following settings: 25msec pulse length, 10 µF capacitance, 600Ω resistance, and 1800V voltage. We performed 6 electroporations using the same plasmid in parallel. Immediately after electroporation, we added 1mL of pre-warmed (37ºC) recovery media to each electroporation cuvette then transferred the liquid to 1.5mL eppendorf tubes and placed these tubes on a shaker at 225rpm and 37ºC for 1 hour. After this recovery, we took 10 µL of the culture for plating serial dilutions and transferred the rest to 150-200mL of 1x LB Broth containing 100 µg/mL ampicillin. We incubated these cultures on a shaker at 225rpm and 32ºC for 12-14 hours then pelleted the cultures by centrifugation and isolated plasmid using the EndoFree plasmid maxiprep kit (Qiagen) according to the manufacturer's protocol. In some instances, pellets were frozen at -20ºC for several days before plasmid isolation. To estimate transformation efficiency, we counted colonies on the plated serial dilutions and verified barcode insertion by PCR from 20-30 colonies per plate. We pooled the plasmids from the 6 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 (Hsu et al., 2002; Shaffer et al., 2017). 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, 2.4mM $CaCl_2$, 50 U/mL penicillin, and 50 µg/mL 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 penicillin and 50 µg/mL streptomycin and passaged cells using 0.05% Trypsin-EDTA.

Lentivirus packaging and transduction:
Prior to plasmid transfection, HEK293FT cells were grown to ~90% confluency in 6-well plates in DMEM containing 10% FBS without antibiotics. For each 6 well plate, we added 80 µL PEI to 0.5 mL Opti-MEM (ThermoFisher 31985062) and separately, combined 7.5 µg pPAX2, with 5 µg VSVG and 7.71 µg of the barcode plasmid library in 0.5 mL OPTIMEM then incubated the solutions at room temperature for 5 minutes. We then mixed the 2 solutions together with vortexing and incubated the combined solution at room temperature for 15 minutes. We added 184 µL of the plasmid-PEI solution dropwise to each well of the 6-well plate. Ater 6-8 hours, we aspirated the media from the cells, washed the cells once with 1xDPBS, then added fresh culture media (DMEM containing 10% FBS and antibiotics). The following morning, after confirming that the majority of cells were GFP positive, we aspirated the media, washed the cells once with 1xDPBS then added 1 mL of TU2% to each well. Approximately 12 hours later, we transferred the virus laden media to a falcon tube and added another 1 mL of TU2% to each well. We collected virus laden media twice more over the next ~16 hours and during this time, stored the collected media at 4ºC. After the final collection, we filtered the virus laden media through a 0.22 µm PES filter then stored 1-2 mL aliquots at -80ºC.

To transduce WM989 A6-G3 and WM983b E9-C6 cells we added freshly thawed (on ice) virus laden media and polybrene (final concentration 4µg/mL) to dissociated cells, then plated the cells onto 6-well plates (100,000 cells in 2mL per well) and centrifuged the plate at 1,750 rpm (517 x g) for 25 minutes. We incubated the cells with virus for 6-8 hours then removed the media, washed the cells once with 1xDPBS and added 3mL of TU2% to each well. The following day, we passaged the cells to 10cm dishes (1 x 6-well plate into 3 x 10 cm dishes). For WM989 A6-G3, we split barcoded cells into Carbon Copy and separate vemurafenib treatment groups 11 days after transduction for sort experiments (Figure 1) or 10 days after transduction for *in situ* experiments (Figures 2-4) unless otherwise specified. For WM983b E9-C6, we split barcoded cells into Carbon Copy and separate vemurafenib treatment groups 7 days after transduction for sort experiments (Supp. Fig 6) unless otherwise specified. We cultured *in situ* Carbon Copies for 4 days before fixation in order to more easily identify clusters of cells expressing targeted barcodes.

Fluorescence Activated Cell Sorting (FACS):
To isolate ITGA3-High WM989 A6-G3, we first trypsinized and pelleted 8 confluent 10cm plates, washed once with 1 x 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 0.1% BSA-PBS containing 1:200 anti-ITGA3 antibody (DSHB clone P1B5 stock concentration 354 µg/mL) then incubated on ice for 1 hour. After primary incubation, we pelleted the cells, washed twice with ~5 mL 0.1% BSA-PBS then resuspended cells in 0.16 mL 0.1% BSA-PBS containing 1:500 anti-mouse FAb2 secondary antibody conjugated to AlexaFluor488 (Cell Signalling #4408) and incubated on ice for 30 minutes. Finally, we pelleted the cells, washed twice with 0.1% BSA-PBS , then

resuspended the pellet in 0.1% BSA-PBS containing 100 ng/mL 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 2-well glass bottom chamber plate (Nunc Lab-Tek 155380) for treating with vemurafenib (see below) and the rest on a seperate 2-well glass bottom chamber plate for verifying *ITGA3* expression by RNA FISH.

We followed a similar procedure for isolating AXL-High WM983b E9-C6 starting with 10 10cm 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 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, then plated cells onto 2-well glass bottom plates (10,000 per well) for vemurafenib treatment or RNA FISH as above.

Drug treatment experiments:
We prepared stock solutions of 4mM vemurafenib (PLX4032, Selleckchem, S1267), 10mM pinometostat (SelleckChem S7062), 100μM trametinib (SelleckChem S2673), and 10mM Dabrafenib (SelleckChem S2807). We prepared all stock solutions in DMSO and divided into small aliquots stored at -20ºC 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 for 3 weeks replacing media containing drug every 3-4 days. For DOT1L inhibitor pre-treatment, we treated cells with 4 μM pinometostat for 6 days, replacing media on day 3 and again when splitting off the Carbon Copy on day 5. Following the ITGA3 sort, we fixed WM989 A6-G3 cells after 18 days of vemurafenib treatment in order to more easily quantify numbers of colonies. For Time Machine experiments in WM983b E9-C6, we treated cells for 4 weeks replacing media containing 1 μM of vemurafenib every 3-4 days.

Cell quantification:
Following drug treatment experiments, we fixed cells by incubation for 10 minutes in 3.7% formaldehyde (Sigma F1635) diluted in 1 x PBS, followed by two washes with 1 x PBS then overnight permeabilization at 4ºC with 70% ethanol. We stained nuclei by incubation in 2xSSC containing 50 ng/mL DAPI then imaged the majority of each well via a tiling scan at 20x 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 analysis can be found:
https://bitbucket.org/arjunrajlaboratory/colonycounting_v2/src/default/ and
https://www.dropbox.com/sh/r5lypm20dd3ei49/AAC7C2hyBWyarC81KcgPtrNta?dl=0.

Barcode library preparation and sequencing:
We isolated genomic DNA (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 6 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 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: 98ºC for 30 seconds, followed by N cycles of 98ºC for 10 seconds then 65ºC for 40 seconds, and finally 65ºC for 5 minutes. After the PCR, we purified libraries using 35 µL (0.7X) Ampure XP magnetic beads with two 80% ethanol washes followed by final elution in 20 µL 0.1X TE (1 mM Tris HCl pH 8.0 100 µM EDTA). Purified libraries from the same sample were pooled, quantified using the Qubit dsDNA High Sensitivity assay (ThermoFisher) then sequenced on a NextSeq 500 using 150 cycles for read 1 and 8 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 qPCR reaction and selecting the number of cycles needed to achieve ⅓ of the maximum fluorescence intensity. We included 0.25 µL 100X SYBR Green I (10,000 X diluted 1:100 in 10 mM Tris pH 8.0; Invitrogen S7563) per 25 µL 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 MOI, 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 between these technical replicates. As shown in Supp. Fig 2, we found a high correlation (>95%) in barcode abundance between these technical replicates.

Computational analyses of barcode sequencing data: We recovered barcodes from sequencing data using custom Python scripts available at: https://bitbucket.org/arjunrajlaboratory/timemachine/src/default/. 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). While we do not believe that these "UMIs" tag unique barcode DNA molecules, empirically they appeared to slightly improve the correlation in barcode abundance between replicate libraries and were therefore used for most subsequent analyses. Using the STARCODE software (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 then converted the most abundant 100-200 barcodes 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 criterias 1 and 2. For barcodes corresponding to resistant colonies not requiring DOT1L inhibition, we selected sequences that fell among the top 200 barcodes in both the DOT1L inhibitor and DMSO pre-treated resistant cells with a relatively small difference in abundance between these two conditions (not among the 500 barcodes with the largest difference in abundance).

Barcode FISH probe design:
Using fasta files of selected barcodes, we design HCR probes using Rajlab ProbeDesignHD software(code freely available for non-commercial use here https://flintbox.com/public/project/50547/) . For each barcode sequence, we designed 2 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 2 20mer sequences (removing the middle two nucleotides) and appended split-initiator HCR sequences using custom python scripts (See Supplementary Table 3 for sequences) (Choi et al., 2018). For each 20mer sequence, we measured the maximum complementarity to the vector backbone and other barcodes present in the sample in order 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 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 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 probeset by combining 4 μL of the pooled oligos with 1 μL T4 PNK (NEB), 20 μL T7 DNA ligase reaction buffer (NEB), and 2 μL nuclease-free water then incubating at 37ºC 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 95ºC for 5 minutes then cooled to 12º C at a rate of -0.1º C/second. After cooling, we added 1 μL 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 (Rouhanifard et al., 2018) .

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 (Supplementary Table 5 for probe sequences). For each gene, we pooled their complementary oligos and coupled the probe set to either Cy3 (GE Healthcare), Alexa 594 (Life Technologies), or Atto647N (Atto-Tec)NHS ester dyes. We performed single molecule RNA FISH as described in (Raj et al., 2008) and (Shaffer et al., 2017) for multiple cycles of hybridization. We aspirated media from adherent cells, washed the cells once with 1x PBS, then incubated the cells in fixation buffer (3.7% formaldehyde 1x PBS) for 10 minutes at room temperature. We next aspirated the fixation buffer, washed samples twice with 1x PBS, then added 70% ethanol and stored samples at 4º C. For hybridization, we first washed samples with washing buffer (10% formamide in 2x SSC) then applied the RNA FISH probes in hybridization buffer (10% formamide and 10% dextran sulfate in 2x SSC). We covered samples with coverslips then hybridized samples overnight in humidified containers at 37ºC. The following morning, we washed samples 2 x 30 minutes with washing buffer at 37ºC, adding 50 ng/mL DAPI to the second wash to stain the nuclei. After these washes, we rinsed samples once with 2xSSC then added new 2xSSC and proceeded with imaging. To strip RNA FISH probes, we incubated samples in stripping buffer (60% formamide in 2x SSC) for 20 minutes on a hot plate at 37ºC, washed samples 3 x 15 minutes with 1xPBS on a hot plate at 37ºC, then returned samples to 2xSSC. After stripping RNA FISH probes, we re-imaged all previous positions and excluded dyes with residual signal from subsequent hybridization.

Barcode RNA ClampFISH:

For *in situ* Barcode FISH using ClampFISH, we adapted the protocol from Rouhanifard et al. 2019 as follows. We generated modified primary probes and amplifier probes as described above. For hybridization, we first washed fixed samples with washing buffer containing 40% formamide in 2x SSC then applied the primary ClampFISH probes in primary hybridization buffer containing 40% formamide, 10% dextran sulfate, 1 mg/mL yeast tRNA (Invitrogen

15401029) , 0.02% BSA, and 100 µg/mL sonicated salmon sperm DNA (Agilent 201190-81) in 2xSSC. We included up to 180 ClampFISH probes targeting up to 60 different barcode RNA sequences per hybridization (total probe concentration 125 ng/uL - 250 ng/µL). We added coverslips to samples then hybridized for 6-8 hours in humidified containers at 37ºC. After hybridization, we added wash buffer containing 40% formamide in 2x SSC to dislodge coverslips then replaced the wash buffer and incubated the samples for 20 minutes at 37ºC. We performed a second wash for 20 minutes at 37ºC using buffer containing 20% formamide and 2x SSC 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 yeast tRNA , 0.02% BSA, and in 2xSSC.; final probe concentration 10 ng/µL each). After the second hybridization we washed samples 2 x 20 minutes at 37ºC using buffer containing 20% formamide and 2x SSC then rinsed the sample with 2xSSC. 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 2x SSC to each sample and incubating at 37ºC for 15-20 minutes. To prepare this solution, we first combined the BTTAA and copper sulfate, add the 2x SSC containing 0.1% Triton-X, and lastly add freshly dissolved L-ascorbic acid (19-20mg of L-ascorbic acid sodium salt dissolved in 1mL nuclease-free water). Once the L-ascorbic acid is added, we immediately added the solution to our samples. Following the click reaction, we rinsed samples once with 2xSSC then washed 1 x 20 minutes at 37ºC with buffer containing 40% formamide in 2x SSC. After this wash, we performed the third round of hybridization with P9B and P9C amplifier probes in amplifier hybridization buffer, followed by washes, 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 x SSC) containing probes targeting P9B and P9C and coupled to AlexaFluor594 and Atto647n, respectively (see Supplementary Table 4 for sequences). We hybridized these probes overnight in humidified containers at 37ºC then washed samples 2 x 30 minutes with washing buffer (10% formamide, 2xSSC) at 37ºC, adding DAPI to the second wash to stain the nuclei. After these washes, we rinsed samples once with 2xSSC then replaced the 2xSSC and proceeded with imaging. To remove ClampFISH signal, we stripped dye-coupled probes as described above for RNA FISH.

Barcode RNA HCR:

For *in situ* Barcode FISH using the Hybridization Chain Reaction (HCR) we adapted the protocol from (Choi et al., 2018) as follows. We used 1.2 pmol each of up to 240 barcode FISH probes per 0.3 mL hybridization buffer. Our primary hybridization buffer consisted of 30% formamide, 10% dextran sulfate, 9 mM citric acid pH 6.0, 50 µg/mL heparin, 1x Denhardt's solution (Life Technologies 750018) and 0.1% tween-20 in 5x SSC. For primary hybridization, we used 100 µL hybridization buffer per well of a 6 well plate, covered the well with a glass coverslip, then incubated the samples in humidified containers at 37ºC for 6 hours. Following the primary probe

hybridization, we washed samples 4 x 5 minutes at 37ºC with washing buffer containing 30% formamide, 9 mM citric acid pH 6.0, 50 µg/mL heparin, and 0.1% tween-20 in 5x SSC. We then washed the samples at room temperature 2 x 5 minutes with 5xSSCT (5xSSC + 0.1% Tween-20), then incubated the samples at room temperature for 30 minutes in amplification buffer containing 10% dextran sulfate and 0.1% Tween-20 in 5xSSC. During this incubation, we snap-cooled individual HCR hairpins (Molecular Instruments) conjugated to either AlexaFluor647, AlexaFluor594 or AlexaFluor546 by heating to 95ºC for 90 second then immediately transferring to room temperature to cool for 30 minutes concealed from light. After these 30 minutes, we resuspended and pooled the hairpin in amplification buffer to a final concentration of 6nM each. We added the hairpin solution to samples along with a coverslip, then incubated samples at room temperature overnight (12-16 hours) concealed from light. The following morning, we washed samples 5 x 5 minutes with 5xSSCT containing 50ng/mL DAPI, added SlowFade antifade solution (Life Technologies S36940) and a coverslip 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.

For performing HCR in suspension, we adapted the published protocol as follows (Choi et al., 2018). We fixed dissociated cells in suspension by washing the cells with 1xDPBS, resuspending the cell in ice cold 1xDPBS, adding equal volume of ice cold fixation buffer (3.7% formaldehyde 1x PBS) then incubating with rotation at room temperature for 10 minutes. We next pelleted fixed cells by centrifugation at 800 x g for 3 minutes, washed twice with ice cold 1xPBS, then resuspended in 70% ethanol and stored fixed cells at 4ºC. For primary probe hybridization we used 0.5 mL 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 x 10 minutes with 0.5 mL washing buffer then 2 x 10 minutes with 0.5 mL 5xSSCT. We next incubated samples for 30 minutes 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 6 times with 5xSSCT the proceeded with FACS. In between hybridizations and washes, we pelleted cells by centrifugation at 400 x g for 5 minutes and used low-molecular weight dextran sulfate (Sigma D4911) in hybridization and amplification buffers to improve pelleting.

We note that the final hairpin concentrations used in these experiments is 4- to 10-fold lower than the manufacturer's protocol, which we optimized to reduce nonspecific amplification while still enabling sensitive barcode RNA detection at 20x 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 prior to immunofluorescence (Figure 3).

Immunofluorescence:
We performed immunofluorescence using primary antibodies targeting total ERK (L34F12 Cell Signalling #4696) and phosphorylated ERK (p44/p42 ERK D12.14.4E Cell Signalling #4370). First, we rinsed cells 3 times with 5% BSA in PBS (5% BSA-PBS) then incubated at room temperature for 2 hours in 5% BSA-PBS containing 1:100 total ERK and 1:200 pERK antibodies. Next, we washed the cells 5 x 5 minutes with 5% BSA-PBS then incubated the cells at room temperature for 1 hour in 5% BSA-PBS containing 1:500 donkey anti-mouse secondary antibody conjugated to Cy3 (Jackson 715-165-150) and 1:500 goat anti-rabbit secondary antibody conjugated to Alexa594 (Cell Signalling #8889). After the secondary incubation, we washed the cells 5 x 5 minutes with 5% BSA-PBS containing 50 ng/mL DAPI, then replaced the wash buffer with 2xSSC and proceeded with imaging as described below.

RNA FISH and Immunofluorescence imaging:
We imaged RNA FISH samples on an inverted Nikon TI-E microscope with a 20x Plan-Apo λ (Nikon MRD00205), 40x Plan-Fluor (MRH00401) or 60x Plan-Apo λ (MRD01605) objective using filter sets for DAPI, Cy3, Alexa594, and Atto647N. For barcode clampFISH and barcode HCR, we first acquired tiled images in a single Z-plane (scan) at 20x or 40x magnification, then after identifying positions containing cells positive for resistant barcodes, we returned to those positions to acquire a Z-stack at 60x magnification. For subsequent rounds of single-molecule RNA FISH and ERK immunofluorescence we acquired Z-stacks at 60x magnification. For scans, we used a Nikon Perfect Focus system to maintain focus across the imaging area.

Image analysis:
To identify Barcode FISH positive cells for Rewind, we used custom MATLAB scripts to stitch, contrast and compress scan images (scripts available at https://bitbucket.org/arjunrajlaboratory/timemachineimageanalysis/src/default/) then manually reviewed these stitched images. This review yielded positions containing candidate Barcode FISH positive cells which we then re-imaged for verification at 60x magnification in multiple Z-planes. If we were uncertain about the fluorescence signal in a candidate cell (e.g. abnormal localization pattern, 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://bitbucket.org/arjunrajlaboratory/rajlabimagetools/wiki/Home. 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/eeu0o9f7bcghm7x/AAAropGHDhSZ7_TYxJsIV8DDa?dl=0/.

For the UMAP visualizations we used the Seurat v3.0 package (Butler et al., 2018; Stuart et al., 2019). For the analysis shown in Figure 4, we ran the UMAP algorithm on scaled RNA FISH

data using the first 5 principal components and setting n_neighbors = 30 and min_dist = 0.3 (default settings). For the analyses shown in Figures 5 and 6, we used the first 6 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 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 (Supp. Fig 8).

For quantification of cell and colony numbers following vemurafenib treatment, we used custom MATLAB software available at: https://bitbucket.org/arjunrajlaboratory/colonycounting_v2/src/default/. The analysis pipeline involves stitching the tiled dapi images, manually segmenting individual wells and colonies, identifying individual cells based on DAPI signal, and then extraction of 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/eeu0o9f7bcghm7x/AAAropGHDhSZ7_TYxJsIV8DDa?dl=0. We used a separate MATLAB script (https://www.dropbox.com/s/5tff7t0x86d80s1/countCellsTimeMachineScans.m?dl=0) to quantify the number of cells imaged in our Carbon Copies.

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 50ºC for 90 minutes 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) 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 v2.5.2a and counted uniquely mapped reads with HTSeq v0.6.1.

We performed differential expression analysis in R v3.6.0 using DESeq2 v1.22.2 . We considered a gene to be differentially expressed if the comparison between 2 conditions yielded a $\log_2$ fold change of ≥1 or ≤ -1 and adjusted p-value of ≤0.1. For determining candidate markers for primed cells requiring DOT1L inhibition (Figure 5) we compared primed and non-primed subpopulations sorted from both DOT1L inhibitor and vehicle control Carbon Copies and modelled the biological replicate and DOT1L inhibitor treatment as covariates in the design formula for DESeq2. We performed hierarchical clustering and principal component analysis on $\log_2$ transformed tpm values using R v3.6.0.

We tested for enrichment of differentially expressed genes among gene ontologies and pathways (KEGG, REACTOME, WikiPathway) using WebGestaltR. If a differentially expressed gene was included in one or more enriched GO term or pathway, we chose a consensus annotation (e.g. ECM organization and cell migration) for that gene. Otherwise, we attempted to assign a gene annotation by manual review. Our resulting gene annotation can be found on dropbox at this link: https://docs.google.com/spreadsheets/d/1XXLC8RUvn8iArpfInRoPzOtIGSEtizWLzKedfjwW2L4/edit?usp=sharing

Software and data availability:

All data and code used for these analyses can be found at https://www.dropbox.com/sh/graf5m2pm7crcdq/AAAyr5liHKpBN7_MEZ2CVEkda?dl=0

Acknowledgements:

We thank Caroline Bartman, Anthony Anguierra, John Murray, Nancy Zhang, Long Cai, Benjamin Stanger, Allison Coté, Karun Kiani, and Eric Sanford, along with other members of the Raj lab 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 Florin Tuluc for several helpful discussions. BLE acknowledges support from NIH training grants F30 CA236129, T32 GM007170 and T32 HG000046, ET acknowledges R01 CA238237, IPD acknowledges NIH 4DN U01 HL129998, NIH Center for Photogenomics (RM1 HG007743), CLJ acknowledges NIH T32 DK007780, NSJ acknowledges NIH Director's Transformative Research Award R01 GM137425, SMS acknowledges support from DP5 OD028144 AR acknowledges 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, NIH Center for Photogenomics (RM1 HG007743), and the Tara Miller Foundation.

References

Alemany, A., Florescu, M., Baron, C.S., Peterson-Maduro, J., and van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. Nature *556*, 108–112.

Al'Khafaji, A., Gutierrez, C., Brenner, E., Durrett, R., Johnson, K.E., Zhang, W., Li, S., Livak, K.J., Neuberg, D., Brock, A., et al. (2019). Expressed barcodes enable clonal characterization of chemotherapeutic responses in chronic lymphocytic leukemia.

Al'Khafaji, A.M., Deatherage, D., and Brock, A. (2018). Control of Lineage-Specific Gene Expression by Functionalized gRNA Barcodes. ACS Synth. Biol. *7*, 2468–2474.

Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., and Morris, S.A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature *564*, 219–224.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*.

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

Corcoran, R.B., Ebi, H., Turke, A.B., Coffee, E.M., Nishino, M., Cogdill, A.P., Brown, R.D., Della Pelle, P., Dias-Santagata, D., Hung, K.E., et al. (2012). 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.

Feldman, D., Tsai, F., Garrity, A.J., O'Rourke, R., Brenan, L., Ho, P., Gonzalez, E., Konermann, S., Johannessen, C.M., Beroukhim, R., et al. (2019). CloneRetriever: retrieval of rare clones from heterogeneous cell populations.

Frieda, K.L., Linton, J.M., Hormoz, S., Choi, J., Chow, K.-H.K., Singer, Z.S., Budde, M.W., Elowitz, M.B., and Cai, L. (2016). Synthetic recording and in situ readout of lineage information in single cells. Nature.

Gerosa, L., Chidley, C., Froehlich, F., Sanchez, G., Lim, S.K., Muhlich, J., Chen, J.-Y., Baker, G.J., Schapiro, D., Shi, T., et al. (2019). Sporadic ERK pulses drive non-genetic resistance in drug-adapted BRAFV600E melanoma cells.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

Hsu, M.-Y., Elder, D.E., and Herlyn, M. (2002). Melanoma: the Wistar melanoma (WM) cell

lines. In Human Cell Culture, (Springer), pp. 259–274.

Hurley, K., Ding, J., Villacorta-Martin, C., Herriges, M.J., Jacob, A., Vedaie, M., Alysandratos, K.D., Sun, Y.L., Lin, C., Werder, R.B., et al. (2020). Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. Cell Stem Cell.

Ji, Z., Erin Chen, Y., Kumar, R., Taylor, M., Jenny Njauw, C.-N., Miao, B., Frederick, D.T., Wargo, J.A., Flaherty, K.T., Jönsson, G., et al. (2015). MITF Modulates Therapeutic Resistance through EGFR Signaling. J. Invest. Dermatol. *135*, 1863–1872.

Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., and Church, G.M. (2018). Developmental barcoding of whole mouse via homing CRISPR. Science *361*.

Lito, P., Pratilas, C.A., Joseph, E.W., Tadi, M., Halilovic, E., Zubrowski, M., Huang, A., Wong, W.L., Callahan, M.K., Merghoub, T., et al. (2012). Relief of profound feedback inhibition of mitogenic signaling by RAF inhibitors attenuates their activity in BRAFV600E melanomas. Cancer Cell *22*, 668–682.

McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole organism lineage tracing by combinatorial and cumulative genome editing. Science aaf7907.

Müller, J., Krijgsman, O., Tsoi, J., Robert, L., Hugo, W., Song, C., Kong, X., Possik, P.A., Cornelissen-Steijger, P.D.M., Geukes Foppen, M.H., et al. (2014). Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. Nat. Commun. *5*, 5712.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Raj, B., Wagner, D.E., McKenna, A., Pandey, S., Klein, A.M., Shendure, J., Gagnon, J.A., and Schier, A.F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat. Biotechnol. *36*, 442–450.

Roesch, A., Fukunaga-Kalabis, M., Schmidt, E.C., Zabierowski, S.E., Brafford, P.A., Vultur, A., Basu, D., Gimotty, P., Vogt, T., and Herlyn, M. (2010). A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. Cell *141*, 583–594.

Rouhanifard, S.H., Mellis, I.A., Dunagin, M., Bayatpour, S., Jiang, C.L., Dardani, I., Symmons, O., Emert, B., Torre, E., Cote, A., et al. (2018). ClampFISH detects individual nucleic acid molecules using click chemistry-based amplification. Nat. Biotechnol.

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

Shaffer, S.M., Emert, B.L., Sizemore, A.E., Gupte, R., Torre, E., Bassett, D.S., and Raj, A. (2018). Memory sequencing reveals heritable single cell gene expression programs associated with distinct cellular behaviors.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell *141*, 69–80.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. (2006). Variability and memory of protein levels in human cells. Nature *444*, 643–646.

Spencer, S., Gaudet, S., Albeck, J., Burke, J., and Sorger, P. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Sun, C., Wang, L., Huang, S., Heynen, G.J.J.E., Prahallad, A., Robert, C., Haanen, J., Blank, C., Wesseling, J., Willems, S.M., et al. (2014). Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma. Nature *508*, 118–122.

Symmons, O., and Raj, A. (2016). What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. Mol. Cell *62*, 788–802.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Torre, E.A., Arai, E., Bayatpour, S., Beck, L.E., Emert, B.L., Shaffer, S.M., Mellis, I.A., Budinich, K.A., Weeraratna, A., Shi, J., et al. (2019). Genetic screening for single-cell variability modulators driving therapy resistance.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*.

Wu, P.-H., Gilkes, D.M., Phillip, J.M., Narkar, A., Cheng, T.W.-T., Marchand, J., Lee, M.-H., Li, R., and Wirtz, D. (2020). Single-cell morphology encodes metastatic potential. Sci Adv *6*, eaaw6938.

Figure legends

Figure 1: Time Machine identifies rare cell states giving rise to vemurafenib resistant colonies. A. Schematic of Time Machine approach for isolating the initial primed WM989 A6-G3 melanoma cells that ultimately give rise to vemurafenib resistant colonies. For the experiment shown, we started with ~ 200,000 cells transduced at an MOI ~ 1.0 with our Rewind barcode library. After 11 days (~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 genomic DNA from the resistant cells that remain 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-seq 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. Inset shows the proportion of barcodes targeted by our probeset detected in each group. These data correspond to 1 of 2 replicates. In the second replicate, 30 out of 50 probed barcodes were detected in the sorted primed population. **C.** We performed differential expression analysis using DESeq2 of primed vs. non-primed sorted cells. Shown is the mean expression level (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 signalling pathways (blue) and previously identified resistance markers (purple; Shaffer et al. 2017). Genes were assigned to categories based on a consensus of KEGG pathway and GO enrichment analyses (See Methods for details). **D**. We selected the most differentially expressed, cell surface ECM-related gene (*ITGA3*) to validate as a predictive marker of vemurafenib resistance in WM989 A6-G3. After staining cells with a fluorescently labelled antibody targeting ITGA3, we sorted the brightest 0.5% (ITGA3-High) and remaining (ITGA3-Low) populations, then treated both with 1 μM vemurafenib. After approximately 18 days, we fixed the cells, stained nuclei with DAPI then imaged the entire wells to quantify the number of resistant colonies and cells. The data correspond to 1 of 3 biological replicates (Supp. Fig 4).

Figure 2: A coordinated primed cell state characterized by high expression of multiple markers gives rise to vemurafenib resistance in WM989 A6-G3. A. We performed Rewind with image-based profiling on WM989 A6-G3 cells, this time fixing the Carbon Copy *in situ* for measuring gene expression in individual cells using single molecule RNA FISH. After sequencing barcodes in resistant colonies, we designed RNA FISH probes targeting these barcodes and used the RNA FISH probes to fluorescently label rare primed cells (frequency ~1:4000) in the Carbon Copy. **B-C.** We then imaged the Carbon Copy at 20x magnification and identified primed cells labeled with our RNA FISH probes using a combination of automated image analysis and manual image review. Once identified, we returned to these cells for re-imaging at high magnification (60x) 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. **D.** Quantification of single-cell gene expression in primed and non-primed cell populations. 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 Supp. Fig 7 for details). Error bars indicate 25th and 75th percentiles of distributions. **E.** Frequency of cells expressing high levels (beyond the thresholds shown in D.) of 1, 2, ...7 markers (out of a total of 7 measured) simultaneously in primed and non-primed cell populations. These data correspond to 1 of 2 biological replicates (Supp. Fig 7).

Figure 3: Resistance to vemurafenib is associated with single-cell variability in phosphorylated ERK levels 24 hours after treatment but not prior to treatment. A. We used Rewind to quantify dual-phospho ERK (p44/p42, pERK) levels in primed cells before and 24 hours after vemurafenib treatment. To quantify pERK levels over time, we plated two Carbon Copies and fixed one 24 hours after vemurafenib treatment and the other prior to treatment. As before, we used barcode RNA FISH probes to identify primed cells in both Carbon Copies then measured single-cell levels of total ERK and pERK by immunofluorescence. We additionally imaged multiple randomly selected positions in each well to quantify total ERK and pERK in non-primed cells. **B.** Barcode RNA FISH and ERK immunofluorescence images of primed cells identified in Carbon Copies fixed before vemurafenib treatment (left) and 24 hours after treatment (right). **C.** Quantification of average pERK immunofluorescence intensity in primed cells and non-primed cells. Each point corresponds to an individual cell. Error bars indicate 25th and 75th percentiles of distributions. These data correspond to 1 of 2 biological replicates (Supp. Fig 8).

Figure 4: Variation in gene expression among primed cells is associated with differences in resistant cell fate. A. We performed Rewind in WM989 A6-G3 and identified barcode sequences enriched in resistant colonies following vemurafenib treatment. We ranked these barcodes by abundance, using abundance as a proxy for 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 then performed sequential rounds of RNA FISH to measure single-cell expression of 9 genes. We additionally imaged multiple randomly selected positions to quantify gene expression in non-primed cells. These data are the same as used in Figure 2, here analyzed using information on which probe set labeled each cell. **C.** We used the UMAP algorithm with the first 5 principal components to visualize the relationship in gene expression between 256 single cells. 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 green, and the two separate groups of primed cells that give rise to less resistant colonies in orange. **D.** Maintaining the organization provided by UMAP, we colored each cell by its expression of each of the 9 genes measured. As noted in the text, ≥98% of primed cells had levels of FN1 RNA that were 3-fold higher than the median observed in non-primed cells, and ≥80% of primed cells had levels of SOX10 and MITF RNA that were ≤ ⅓ the median levels observed in non-primed cells. **E.** We compared expression levels in primed cells of pairs of markers that distinguished the groupings shown in D. **F-G.** We used the same probe sets designed in A to label resistant colonies derived from the same population of cells. We then quantified the number of resistant cells labelled with each probeset.

The number of colonies labeled with each probe set and the average number of cells per colony are in Supp Fig 9. These data correspond to 1 biological replicate.

Figure 5: Rewind identifies a distinct subpopulation of cells that require DOT1L inhibition to become vemurafenib resistant A. WM989 A6-G3 cells transition into and out of a rare primed cell state that gives rise to drug resistant colonies following vemurafenib treatment. We hypothesized that DOT1L inhibition increases the frequency of resistant colonies by either selectively increasing the proliferation of these primed cells, decreasing their transition out of the primed state or enabling a new subpopulation of cells to survive vemurafenib treatment. We distinguished these different possibilities by comparing barcodes from resistant colonies that emerge following vehicle control versus DOT1L inhibitor pre-treatments. We then designed separate RNA FISH probe sets targeting barcodes from resistance that require DOT1L inhibition and barcodes from resistant colonies that emerge regardless of DOT1L inhibition. **B.** Using these probes, we labeled and sorted primed cells requiring DOT1L inhibition to become vemurafenib resistant (purple), primed cells not requiring DOT1L inhibition (green), 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. **C.** 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 2 biological replicates and included DOT1L inhibitor treatment as a covariate in calculating $\log_2$ fold changes using DESeq2. Red points correspond to genes differentially expressed in one or both comparisons (p-adjusted ≤0.1 and $\log_2$ fold change ≥ 1). **D.** Expression of *DEPTOR* in transcripts per million (tpm) in the subpopulations isolated in B. Points indicate tpm values for individual replicates (n=2). **E** We used the same probe sets as in B. to identify cells *in situ* in the Carbon Copies fixed prior to vemurafenib treatment, then measured single cell expression of *DEPTOR*, *MGP*, *SOX10*, *MITF*, and 6 priming markers by RNA FISH. We then applied the UMAP algorithm using the first 6 principal components to the single cell expression data from both vehicle control and DOT1L inhibitor treated Carbon Copies. In this figure, we show the UMAP visualization for the 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. For the remaining plots points are colored by the expression level of the indicated gene in that cell. These data correspond to 1 of 2 biological replicates (Supp. Fig 12). **F.** Single cell expression of *DEPTOR* in the indicated cell populations without DOT1L inhibition or vemurafenib treatment. Each point corresponds to an individual cell. Error bars indicate 25th and 75th percentiles of distributions.. Above each boxplot is the proportion of cells with levels of *DEPTOR* RNA above the indicated threshold (~95th percentile in non-primed cells). 2 of 2 biological replicates are shown.

Figure 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 new 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 cell requiring DOT1L inhibition to survive vemurafenib treatment (purple), cells not requiring DOT1L inhibition (green), and non-primed cells (gray) sorted from both Carbon Copies treated with DOT1L inhibitor (blue outline) and Carbon Copies treated with vehicle control (gray outline) (2

replicates sorted for RNA sequencing). Heatmap 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-adjusted ≤ 0.1 and abs(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. These data include 2 of 2 biological replicates. **D.** Using expression of priming markers as in C., we performed principal component analysis on primed and non-primed cell populations. Blue 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 6 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. **F.** Plotted are single cell expression data for 6 priming markers, *MITF* and *SOX10* in cells that require DOT1L inhibition to become vemurafenib resistant. Each point corresponds to an individual cell. Error bars indicate 25th and 75th percentiles of distributions. Below each boxplot, 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 Figure 14. These data correspond to 1 biological replicate.

Supplementary Figure legends

Supplementary Figure 1. Detections and isolation of cells expressing unique Rewind barcodes using RNA FISH. A. The Rewind construct encodes a 100 nucleotide barcode sequence ("WSN" repeated) in the 3' UTR of GFP downstream of a truncated EF-1 alpha promoter. To optimize barcode RNA detection using RNA FISH, we transduced WM989 A6-G3 cells then derived clonal cell lines each expressing a single barcode. We identified the barcode sequence in each clonal cell line via Sanger sequencing. As these barcodes are encoded on the same transcript as GFP, we could use GFP signal as a ground truth for optimizing our barcode RNA FISH protocol. **B.** We wanted to isolate 100s of primed cells for downstream assays such as RNA sequencing and RNA FISH. In order to capture this many cells using Rewind, we needed to probe for multiple barcodes simultaneously while maintaining a low false positive rate due to non-specific signal. To test this we would combine a clonal GFP+ cell line with a single known barcode with non-transduced GFP- cells then perform barcode RNA FISH using 4 probes targeting the known barcode along with 240 "off-target" probes (designed to target other sequenced barcodes). In parallel, we hybridized a second mix of cells with only the off-target probes. After hybridization, we ran these samples on a FACS instrument and used GFP fluorescence as ground truth for estimating the sensitivity and specificity of our barcode RNA FISH signal. After optimizing probe concentration and hybridization duration (see methods for final protocol), we could detect ~88% of GFP+ cells based on barcode FISH signal (true positives) while excluding 99.9% of GFP- cells (false positives). To isolate rare cell populations with Rewind (i.e. primed cells from our Carbon Copy), we typically used a more conservative Alexa647 gate in an attempt to further minimize false positives. **C-E.** We performed similar mixing experiments for in situ validation of our barcode FISH protocol using either ClampFISH(D) or HCR (E).

Supplementary Figure 2. The Rewind barcode library can uniquely label 100,000s of cells with transcribed barcodes that can be identified via sequencing. A. Critical for Rewind is the ability to uniquely label enough cells with our transcribed barcodes to observe rare phenomena such as drug resistance (frequency < 1:1000). To test this empirically, we separately transduced 2 groups of 150,000 cells at an MOI of ~1.5 , cultured the cells for ~ 1 day then extracted genomic DNA (gDNA) and sequenced their barcodes. Consistent with the starting cell number and MOI, we observed between 210,000 and 253,000 barcodes in the two samples (≥2 UMIs per million; see Methods for description of barcode count normalization) with fewer than 3,500 shared between the two groups. **B-C.** To ensure that we can uniquely distinguish barcodes despite errors during library preparation and sequencing, we estimated the observed barcode diversity in the data from A. We randomly sampled 1,000 barcode sequences and calculated the average and minimum sequence distance (both Hamming and Levenshtein sequence distances) for all 499,500 pairs of barcodes. We repeated this process 500 times and plotted the average and standard deviation of the observed sequence distances. **D.** We tested the reproducibility of our barcode sequencing protocol by preparing separate libraries with unique indexes using the same starting gDNA. As shown in the scatter plots we see a high correspondence in barcode abundances (UMIs per million) between these replicate libraries, even when using different amounts of gDNA (two separate experiments). Plotted are all barcode sequences with at least 50 reads per million in at least one of the two samples. We believe these data also suggest that our barcode sequencing protocol is quantitative, however we acknowledge the possibility that differences between barcode sequences could systematically bias library preparation or PCR. Our validation of barcode RNA FISH probes designed to target

more abundant versus less abundant resistant cells (Figure 4) further suggests that our sequencing data provides a quantitative estimate of clone abundance.

Supplementary Figure 3. Barcode sequencing of "twin" cultures treated with vemurafenib suggests that the primed cell state is maintained through several cell divisions. A. Schematic of the cellular barcoding approach used to test whether the primed cell state is "remembered" through cell division. We transduced ~150,000-200,000 WM989 A6-G3 cells with our Rewind barcode library and allowed the cells to divide for 11-12 days (~4-5 cell divisions). We then split the culture in two and treated both halves with 1 μM vemurafenib for 3 weeks. Finally we sequenced the barcodes in genom- ic DNA extracted from each culture then ranked the barcodes by abundance to identify those likely derived from resistant colonies (expected 100-400 unique barcodes from resistant colonies). In the absence of memory of the primed cell state, we expected to find unique sets of barcodes emerging in the two parallel cultures. In the presence of partial or complete memory, we expected to find some overlap in the barcodes identified in each culture. **B-C.** Heatmap shows the proportion of barcodes shared between the parallel cultures at different rank thresholds. For our Rewind experiments, we selected the top 100-200 barcodes for RNA FISH probe design. **D.** We wanted to rule out the possibility that differences in division rate between cells before adding vemurafenib could skew the distribution of barcodes enough to generate the observed barcode overlap by chance alone. We therefore sequenced barcoded cells after 11 days of growth (before vemurafenib treatment) to estimate the change in the barcode distribution due to differences in cell growth. We then simulated the split and vemurafenib treatment in A by randomly sampling 2 groups of 200 cells each from the observed barcode distribution and calculating the proportion of shared barcodes. The histogram shows the results of repeating this simulation 10,000 times (gray bars) with the red line indicating the experimentally observed proportion of shared barcodes (from B). **E.** We compared the abundance of barcodes from parallel cultures in A-C by plotting all barcodes with at least 100 UMIs per million in at least one sample (see Methods for description of normalization).To better visualize lower abundance barcodes, we binned the barcodes by count and colored each bin by its number of unique barcodes. Based on the observed correlation in barcode abundance between parallel cultures, we reasoned that while the number of cells that make up a vemurafenib resistant clone varies by more than an order of magnitude, these differences are at least partially pre-determined in the initial primed population 3 weeks earlier. **F.** Reassuringly, we do not observe a correlation in barcode counts between vemurafenib resistant cells from independent transductions.

Supplementary Figure 4. RNA sequencing of primed WM989 A6-G3 isolated using Rewind identifies ITGA3 as a prospective marker of vemu- rafenib resistance. A. Shown is flow cytometry data for the Rewind experiment presented in Figure 1. **B.** We compared the expression of genes previously implicated in priming and vemurafenib resistance (Shaffer et al. 2017) in primed and non-primed cells isolated using Rewind. Bargraphs indicate the average log2 fold change in expression in primed versus non-primed cells with individual replicates indicated as points. We previously demonstrated that drug-naive cells expressing the markers in bold (AXL, EGFR, and NGFR) are more likely to form vemurafenib resistant colonies and these same cells express lower levels of SOX10 and MITF (gray bars) (Shaffer et al. 2017). **C.** We found an enrichment for genes associated with ECM organization and cell migration among differentially expressed genes comparing primed cells to non-primed cells (See Methods for pathway enrichment analyses and Supplementary Table 6 for FDR values). Bargraph corresponds to the average log2 fold change in expression in primed versus non-primed cells

with individual replicates indicated as points. We bolded the gene, ITGA3, that we validated as a predictive marker of vemurafenib resistance (Figure 1 and panels D-F). We did not detect expression of ITGA11 in non-primed cells in 1 of 2 replicates (it was detected in primed cells) and the presented data corresponds to the log2 fold change for one replicate. **D.** As described in Figure 1 and Methods, we stained cells with a fluorescently labelled antibody targeting ITGA3, then sorted equal numbers of the brightest ~0.5% (ITGA3-High) and remaining ~99% (ITGA3-Low) cells. We plated ~1/3 of these cells onto one plate for measuring ITGA3 expression by RNA FISH and plated the rest onto a separate plate for treating with 1 μM vemurafenib. After approxi- mately 18 days of treatment, we fixed the cells, stained nuclei with DAPI then imaged the entire wells to quantify the number of resistant colonies and cells. **E.** quantification of ITGA3 RNA by RNA FISH in sorted cells from D. **F.** Whole-well scans of sorted cells from **D.** after vemurafenib treatment.

Supplementary Figure 5. Most priming markers remain transcriptionally altered after 3-weeks of vemurafenib treatment, and an additional ~3000 genes become differentially expressed. A. When performing Rewind in WM989 A6-G3, we collected ~10% of resistant cells for transcriptome profiling alongside cells sorted from our Carbon Copy. This enabled us to ask whether the gene expression changes observed in primed cells persist during their transition into stably resistant cells following vemurafenib treatment. **B.** To address this question, we compared the expression of priming markers in non-primed cells (left), primed cells (middle) and vemurafenib resistant cells (right). We defined priming markers as genes differentially expressed (p-adjusted ≤ 0.1 and abs(log2 fold change) ≥1) in drug-naive primed cells versus non-primed. **C.** The majority of markers with increased expression in primed cells remained highly expressed or became further elevated in resistant cells. In addition, the majority of markers with lower expression in primed cells had lower expression in resistant cells. **D.** We found more than 3,000 genes differentially expressed comparing resistant cells to drug-naive non-primed cells of which the ~200 priming markers represented a small subset. We performed pathway and gene-ontology analyses on the genes differentially expressed only in resistant cells and highlight several recurring annotations (see Supplementary Table 7 for a complete list and FDR values).

Supplementary Figure 6. Rewind on WM983b E9-C6 identifies AXL as a marker of primed cells giving rise to vemurafenib resistance . A. Starting with ~400,000 cells transduced at an MOI ~ 1.0, we repeated our “heritability split experiments” in WM983b E9-C6 to determine if the initial primed cell state giving rise to vemurafenib resistance was maintained through several cell divisions. The venn diagram and heatmap on the right show the proportion of barcodes shared between the parallel cultures at different rank thresholds (See Methods and Supp. Fig 3 for details). **B.** To estimate the probability of seeing the observed fraction of shared barcodes due to chance, we simulated the experiment using data from WM93b E9-C6 cells transduced as in A then cultured for 7 days before sequencing. We simulated the split and vemurafenib treatment by randomly sampling 2 groups of 200 cells each from the barcode distribution and calculating the proportion of shared barcodes. The histogram shows the results of repeating this simulation 10,000 times (gray bars) with the red line indicating the observed proportion of shared barcodes from A. **C.** Using Rewind we isolated primed WM983b E9-C6 cells from a Carbon Copy fixed prior to vemurafenib treatment. We then performed RNA sequencing and barcode sequencing on cDNA prepared from sorted cells. **D.** 16 out of ~200 barcodes recovered from primed cells were among the 60 barcodes found in resistant cells and targeted by our RNA FISH probes. This represents a ≥ 500-fold enrichment of these clones over their frequency in the initial

population (60/400,000). **E.** We used DESeq2 to identify differentially expressed genes (p-adjusted ≤ 0.1 and abs(log2 fold change) ≥1) in primed cells versus non-primed cells (red points). Compared to similar experiments in WM989 A6-G3, few genes passed our significance cutoff which may reflect the shorter memory of the primed cells state (see A) and the lower purity of our Rewind sort (D). We highlighted in blue genes that did not pass our significance cutoff but showed ≥ 2-fold higher expression in primed cells in 2 out of 2 replicates. **F.** As described in Methods, we sorted drug-naive cells expressing high levels of *AXL* and low levels of *AXL* then compared their response to vemurafenib treatment. **G.** With ~ 1/3 of the cells we performed RNA FISH to check AXL expression in the two sorted populations. **H.** We treated the remainder with vemurafenib for 3 weeks then imaged the wells to quantify the number of resistant colonies (see Methods for details).

Supplementary Figure 7. Primed WM989 A6-G3 express high levels of multiple markers simultaneously. A. As described in Figure 2 and Methods, we identified primed cells that give rise to vemurafenib resistance using Rewind then measured single cell expression of 7 priming markers, SOX10 and MITF by RNA FISH. The histograms show the expression distributions for these genes in primed cells (green) and randomly selected non-primed cells (gray). We used these distributions to set thresholds for binning high expressing cells and, in turn, characterizing the co-expression of these markers in single cells. The black vertical lines correspond to the 90th, 95th, 98th and 99th percentiles of expression in non-primed cells and the red vertical line corresponds to the threshold used for Figure 2. **B.** Based on the indicated thresholds, we calculated the number of primed cells (bottom row) and non-primed cells (top row) that express high levels multiple markers simultaneously (number of markers indicated on X-axis). **C.** Using the manual thresh- olds from A (red vertical lines) we calculated the odds ratio for pairs of markers being co-expressed at high levels in the same primed cells. **D-F.** We repeated the analyses in A-C on non-primed and primed cells from a separate experiment. These primed cells correspond to cells that do not require DOT1L inhibition to become vemurafenib resistant from a Carbon Copy treated with vehicle control (DMSO) for 5 days before fixation (see Figure 5 and Methods for further details).

Supplementary Figure 8. Primed cells giving rise to vemurafenib resistance have higher levels of phosphorylated ERK than non-primed cells 24 hours after vemurafenib treatment, but there remains cell to cell variability. A. We used Rewind to quantify dual-phospho ERK (p44/p42, pERK) levels in primed cells (green) and non-primed cells (gray) before and 24 hours after vemurafenib treatment. Shown in A is a biological replicate of the experiment shown in Figure 3. For this second replicate, we amplified the barcode RNA FISH signal using ClampFISH instead of HCR (see Methods for details). Each point corresponds to an individual cell. Average values are calculated for a single Z-plane as described in Methods. **B-C.** In addition to average phosphorylated ERK across the entire cell, we calculated average phosphorylated ERK within just the cell nucleus (left) or just the cytoplasm (right). We found similar results using any of these metrics. **D.** While on average primed cells had higher levels of phosphorylated ERK, we observed several clusters of primed cells (presumably closely related) with variable levels of phosphorylated ERK. Shown is one such example. We speculate that this variability may be a result of pulsatile MAPK signaling, which has been documented in other melanoma cell lines (Gerosa et al. 2019), and our snapshot measurement of ERK phosphorylation via immunofluorescence **E.** After identifying primed cells in situ, we performed both single-molecule RNA FISH and immunofluorescence to measure gene expression and phosphorylated ERK in the same single cells. Shown is the relationship between

phosphorylated ERK levels and AXL (left) or SOX10 (right) expression in individual primed (green points) and non-primed (gray points) cells. Within the primed cell population, we observe a fairly low correlation between phosphorylated ERK and expression of these markers, which we speculate may be a result of MAPK signalling fluctuating on a faster timescale than changes in gene expression.

Supplementary Figure 9. Barcode RNA FISH can distinguish highly resistant clones from less resistant clones. A. As described in the Results for Figure 4, we identified barcodes in vemurafenib resistant cells by sequencing then ranked these barcodes by relative abundance and designed separate RNA FISH probe sets targeting the top ~ 50 and next ~ 50 barcodes. We reasoned that these two groups correspond to clones (cells sharing identical barcodes) with greater and fewer resistant cells (see Supp. Figs 2-3 for further reasoning behind this assumption). We refer to these two groups as “highly resistant” and “less resistant” as these groups roughly correspond to degrees of fitness in drug and our colleagues found the terms “more abundant” and “less abundant confusing. To empirically test that our probe sets distinguish different groups of resistant cells, we labeled resistant colonies derived from the same population of barcoded cells with our two probe sets, each coupled to a differ”ent fluorescent dye (AlexaFluor 546 and AlexaFlu- or647). We then imaged the cells and quantified the number of colonies and resistant cells labeled with each probe set. **B.** Stitched scan images of resistant cells from A. We used custom software as described in Methods to annotate colonies labeled with our “highly resistant” (green) or “less resistant” (orange) probe sets. To check that we were not seeing differences in the number of labeled cells between probe sets due to differences in HCR hairpin and dye, we swapped the initiator sequence (and corresponding hairpin and dye) for the less resistant clone probeset, and labeled an additional 3 wells shown in the bottom row. **C.** Using custom software described in Methods, we quantified the number of colonies and number of cells labeled with each probe set**.**

Supplementary Figure 10. Identification and isolation of cells requiring DOT1L inhibition to become vemurafenib resistant. A. Our approach for identifying the cells that require DOT1L inhibition to become resistant involved first transducing ~400,000 WM989 A6-G3 cells with the Rewind barcode library, letting the cells divide for ~6 days then splitting the culture into two groups. We treated one group with 4 μM DOT1L inhibitor (pinometostat) and the other with vehicle control (DMSO) for 6 days. We then split each group again, fixing half as our “Carbon Copy” and treating the other half with 1 μM vemurafenib for ~2.5 weeks. After this vemurafenib treatment, we extracted genomic DNA from the remaining cells for barcode sequencing. **B.** We compared the relative abundance of each barcode identified in resistant cells pre-treated with DOT1L inhibitor versus resistant cells pre-treated with vehicle control as shown in A. This comparison revealed a subset of barcodes with a greater abundance in resistant cells pre-treated with DOT1L inhibitor than resistant cells pre-treated with vehicle control (purple points). We used these barcodes to design RNA FISH probes targeting primed cells requiring DOT1L inhibition to become vemurafenib resistant. A separate set of barcodes showed similar high abundance with or without DOT1L inhibition (green points), which we used to design RNA FISH probes targeting primed cells not requiring DOT1L inhibition to become resistant. We note that the barcodes with high abundance in both conditions (green points), appear slightly less abundant in resistant cells pre-treated with DOT1L inhibitor. We believe this reflects the greater total number of surviving cells in this condition influencing the normalization of our barcode sequencing data. **C.** We used our barcode RNA FISH probe sets to sort cells from the Carbon Copy pre-treated with DOT1L inhibitor and separately from the Carbon Copy pre-treated with

vehicle control. For our first experimental replicate, we used probes coupled to distinct fluorophores (Alexa647 and Alexa546) to separately isolate primed cells requiring DOT1L inhibitor to become vemurafenib resistant and primed cells not requiring DOT1L inhibitor. We isolated equal numbers of GFP+/Al- exa546-/Alexa647- non-primed cells. For the second experimental replicate (not shown) we divided our Carbon Copies in two and hybridized each half with separate barcode RNA FISH probe sets coupled to Alexa647. After sorting, we prepared libraries for RNA sequencing and with the extra cDNA, performed targeted barcode sequencing. **D.** Bargraphs show the abundance (y-axis) and rank (x-axis) for barcodes (≥ 5 normalized reads) sequenced from sorted cells as described in C. Sequences matching barcodes from our probe set targeting cells requiring DOT1L inhibition to become resistant are colored red. Sequences matching barcodes from our probe set targeting cells not requiring DOT1L inhibition to become resistant are colored blue. All other barcode sequences are colored gray. Inset shows the fraction of barcodes from each probeset identified in each sample.

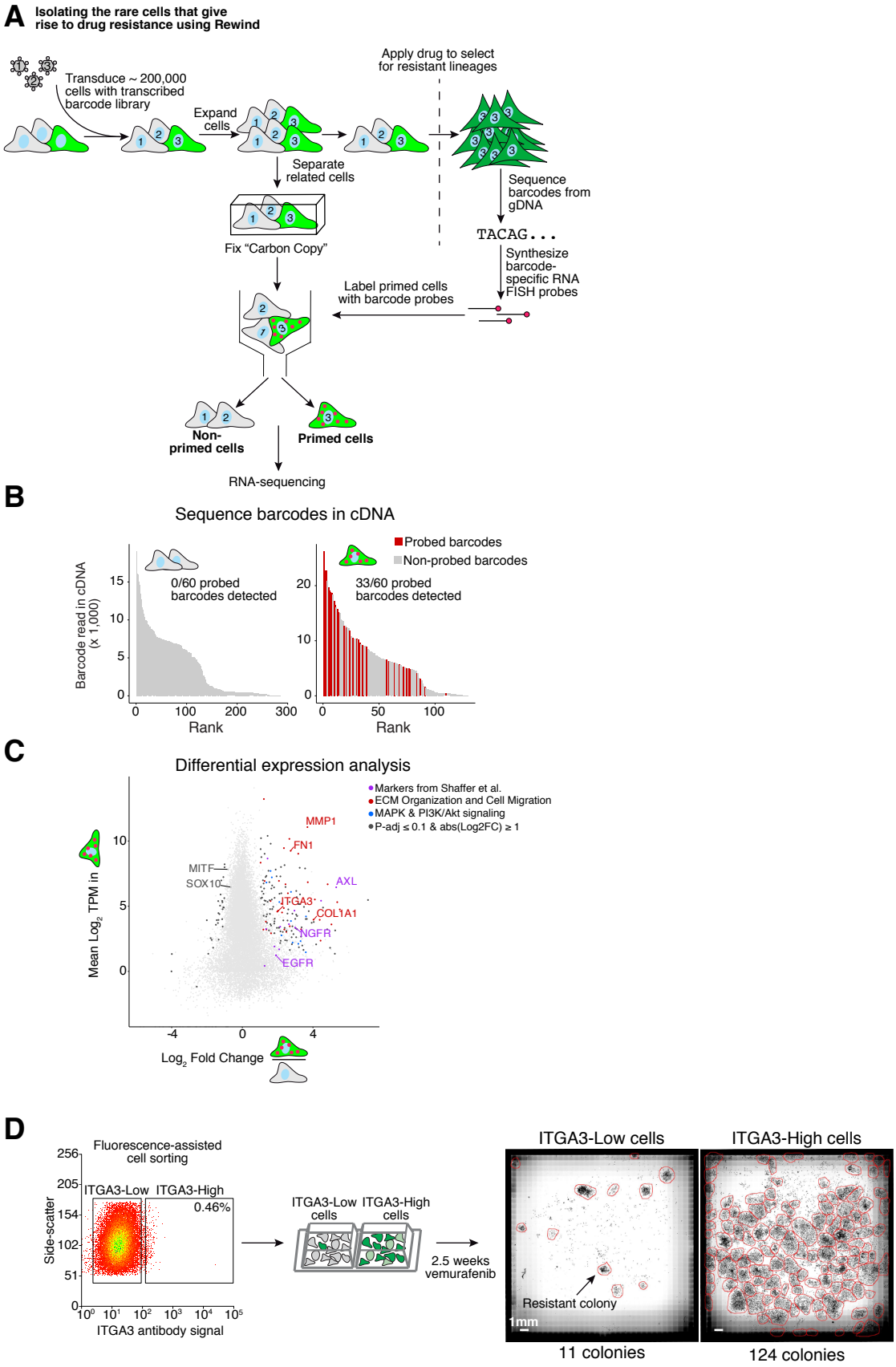
Supplementary Figure 11. Validation of barcode RNA FISH probe sets targeting cells that require DOT1L inhibition to become resistant and cells that do not require DOT1L inhibition. A. When we performed the Rewind experiments on DOT1L inhibitor pre-treated and vehicle control treated cells (see Figure 5 and Supp. Fig 9), we fixed 10% of the vemurafenib resistant colonies for validating our barcode RNA probes. We expected the probes designed to target cells that require DOT1L inhibition to become resistant to label more resistant cells that were pre-treated with DOT1L inhibitor than resistant cells pre-treated with vehicle control. Conversely, we expected the probes designed to target cells that do not require DOT1L inhibition to label a similar fraction of resistant cells in both conditions. As expected the probes targeting cells that require DOT1L inhibition labeled approximately 4x as many resistant cells pre-treated with DOT1L inhibitor compared to resistant cells pre-treated with vehicle control. We observed a minimal difference in labeling using probes designed to target cells that do not require DOT1L inhibition. To verify that these gates corresponded to cells expressing the target-ed barcodes, we sorted the populations and sequenced their barcodes. **B.** Bargraphs show the abundance (y-axis) and rank (x-axis) for barcodes (≥ 5 normalized reads) sequenced from sorted cells as described in A. Sequences matching a barcode from our probe set targeting cells requiring DOT1L inhibition to become resistant are colored red. Sequences matching a barcode from our probe set targeting cells not requiring DOT1L inhibition are colored blue. All other barcode sequences are colored gray. Inset shows the fraction of barcodes from each probeset identified in each sample.

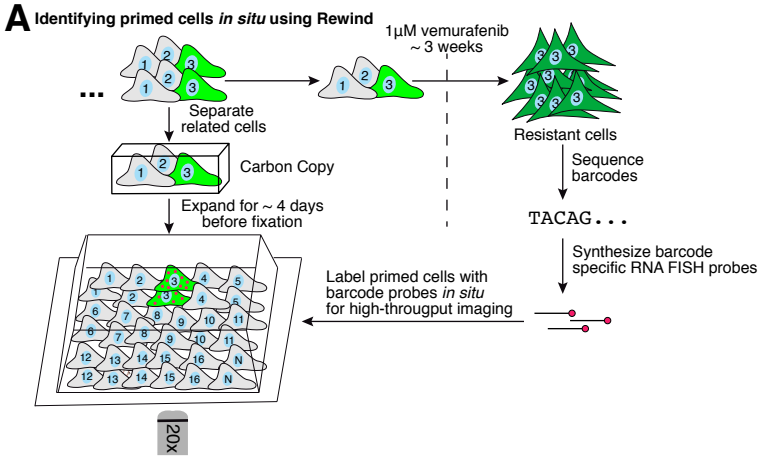
Supplementary Figure 12. The transcriptional profile of cells that require DOT1L inhibition to become vemurafenib resistant is distinct from cells that do not require DOT1L inhibition. A. As described in Results and Methods, we used Rewind to identify cells requiring DOT1L inhibition to become vemurafenib resistant and cells not requiring DOT1L inhibition in Carbon Copies treated with vehicle control (DMSO) and fixed before vemurafenib treatment. We then performed RNA FISH to measure single-cell gene expression of established priming markers. This revealed that cells requiring DOT1L inhibition to become vemurafenib resistant (purple) expressed levels of these markers that were slightly higher than non-primed cells (gray), but far lower than cells not requiring DOT1L inhibition to become resistant (green). Cells requiring DOT1L inhibition expressed intermediate levels of SOX10 and MITF, which may reflect some degree of dedifferentiation. **B.** We repeated this experiment and measured single cell gene expression by RNA FISH on a subset of marker genes. **C.** As shown in Figure 4, RNA sequencing of cells sorted from Carbon Copies found elevated expression of DEPTOR

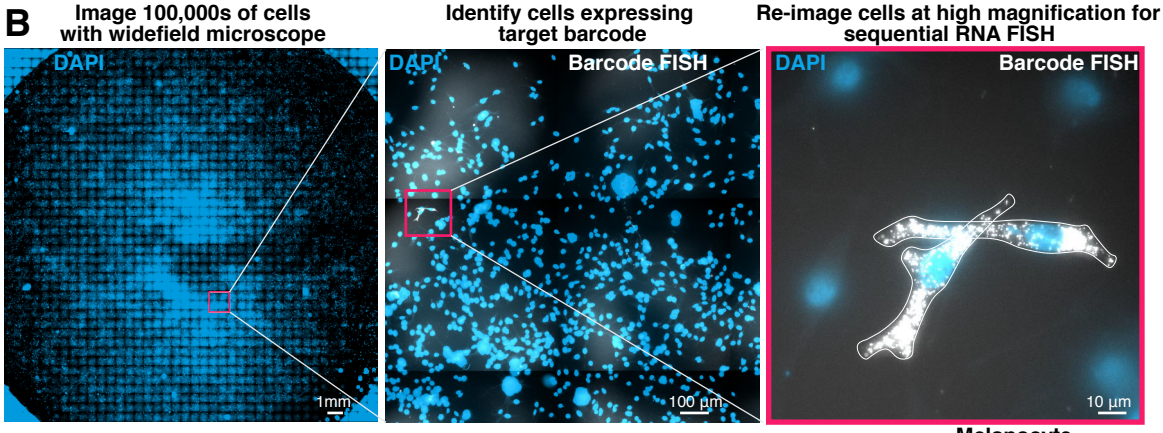
specifically in the cells that require DOT1L inhibition to become resistant. To verify this finding we measured single cell expression of DEPTOR via RNA FISH in both Carbon Copies pre-treated with vehicle control (left) and Carbon Copies pre-treated with DOT1L inhibitor (right). In both these conditions, we found an enrichment of cells expressing high levels of DEPTOR (threshold indicated by dotted line; ~ 95th percentile of non-primed cells) in the subpopulation that requires DOT1L inhibition to become resistant, compared to non-primed cells (in 2 of 2 replicates) and cells that do not require DOT1L inhibition (in 1 of 2 replicates). These expression data implicate DEPTOR as a positive marker of a distinct cell state associated with drug resistance in the context of DOT1L inhibition followed by vemurafenib treatment

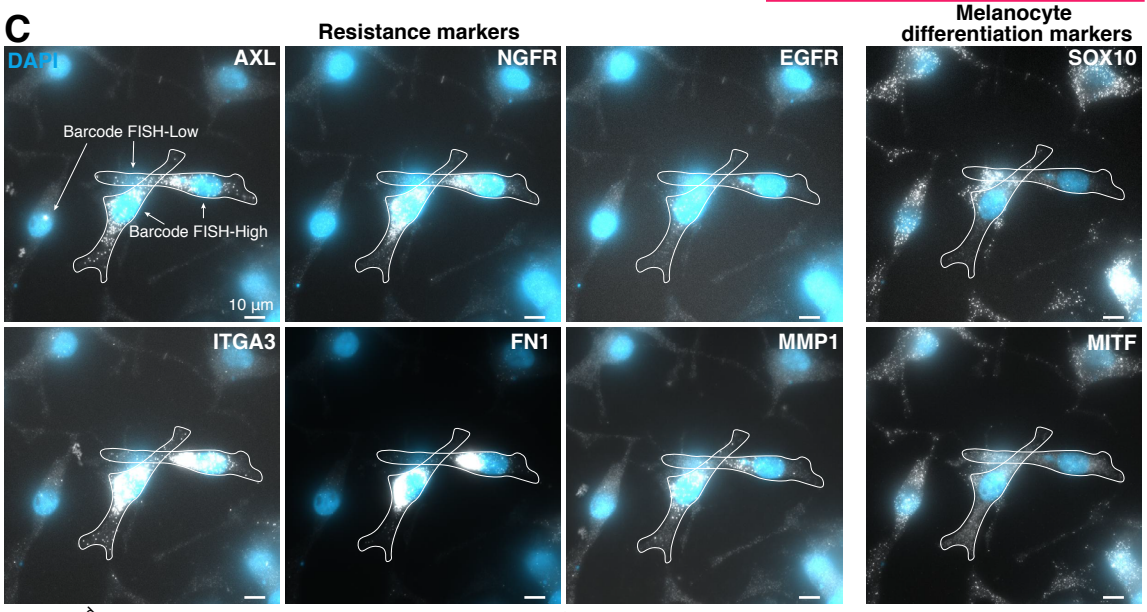
Supplementary Figure 13. MGP is not a consistent marker of cells that require DOT1L inhibition to become vemurafenib resistant. A. As described in the main text and Methods, we sought to use Rewind to identify markers specific to cells that require DOT1L inhibition to become vemurafenib resistant. In our first experimental replicate for RNA-sequencing, we found that primed cells requiring DOT1L inhibition to become resistant (purple) expressed 8-12 fold higher levels of MGP compared to either non-primed cells (gray) or primed cells not requiring DOT1L inhibition (green). This, however, did not replicate in a second experiment. **B.** Based on the magnitude of the initial observation, we nonetheless used RNA FISH to compare single cell expression of *MGP* between each subpopulation in a second set of Carbon Copies. The first RNA FISH experiment revealed an enrichment of cells expressing high levels of MGP in the subpopulation requiring DOT1L inhibition, however this did not replicate in a second experiment.

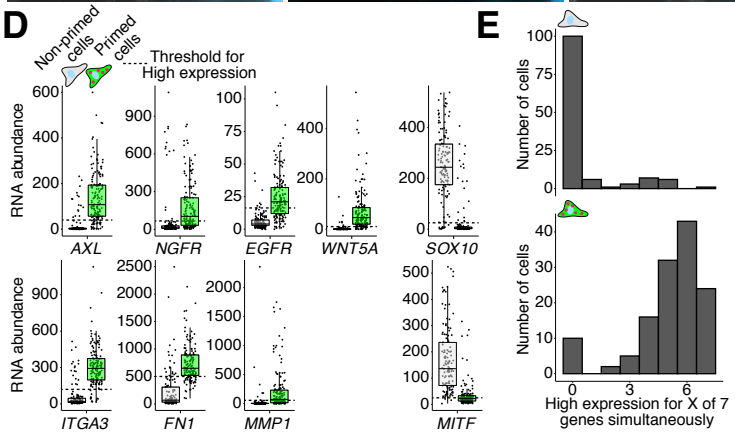
Supplementary Figure 14. DOT1L inhibition partially decouples expression of established priming markers and vemurafenib resistance A. As described in the main text, we asked whether DOT1L inhibition enabled a new subpopulation of cells (purple) to survive vemurafenib treatment by converting them into the previously established primed cell state. To address this question, we measured their expression of 8 genes associated with the primed cell state using RNA FISH performed on both the Carbon Copy treated with DOT1L inhibitor (blue outline) and the Carbon Copy treated with vehicle control (black outline). 6 of these genes (*AXL*, *EGFR*, *NGFR*, *WNT5A*, *ITGA3*, and *COL1A1*) have higher expression in the primed cell state while *SOX10* and *MITF* have lower expression in the primed cell state. At the same time, we measured single cell expression of these same genes in non-primed cells (gray) and primed cells that do not require DOT1L inhibition to become vemurafenib resistant (green). B. We found that DOT1L inhibition modestly increased expression of several genes elevated in the established primed cell state (*AXL*, *NGFR*, *COL1A1*) and decreased expression of *SOX10* and *MITF* specifically in non-primed cells that do not ultimately survive vemurafenib treatment (top row, gray). In contrast, for cells that do ultimately survive vemurafenib treatment (purple and green), DOT1L inhibition appeared to decrease expression of positive markers of the established primed cell state and increase expression of *SOX10* and *MITF*. These transcriptional changes away from the established primed cell state may suggest that compared with vemurafenib treatment alone, cells pre-treated with a DOT1L inhibitor can become vemurafenib resistant via an alternate path. As our RNA sequencing data on these same subpopulations revealed few genes (< 30) induced by DOT1L inhibition, there may not be measurable transcriptional markers for this alternate path.

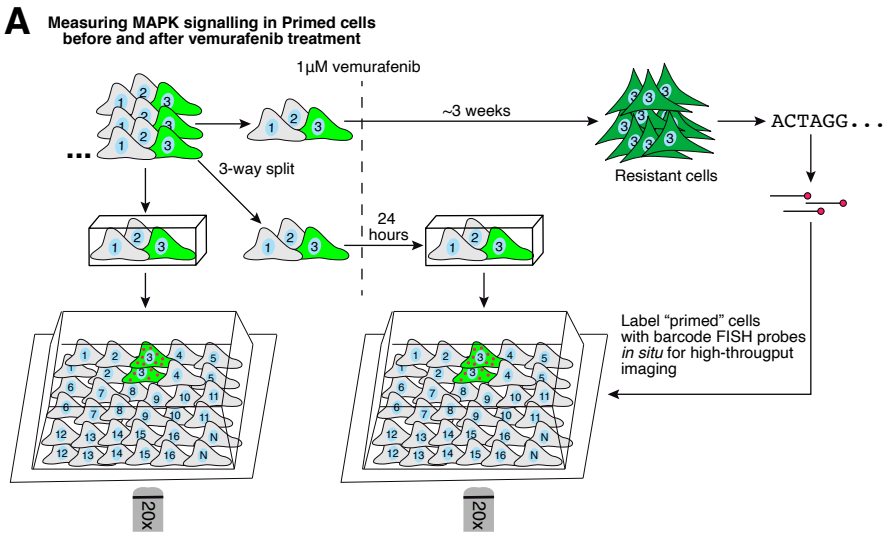


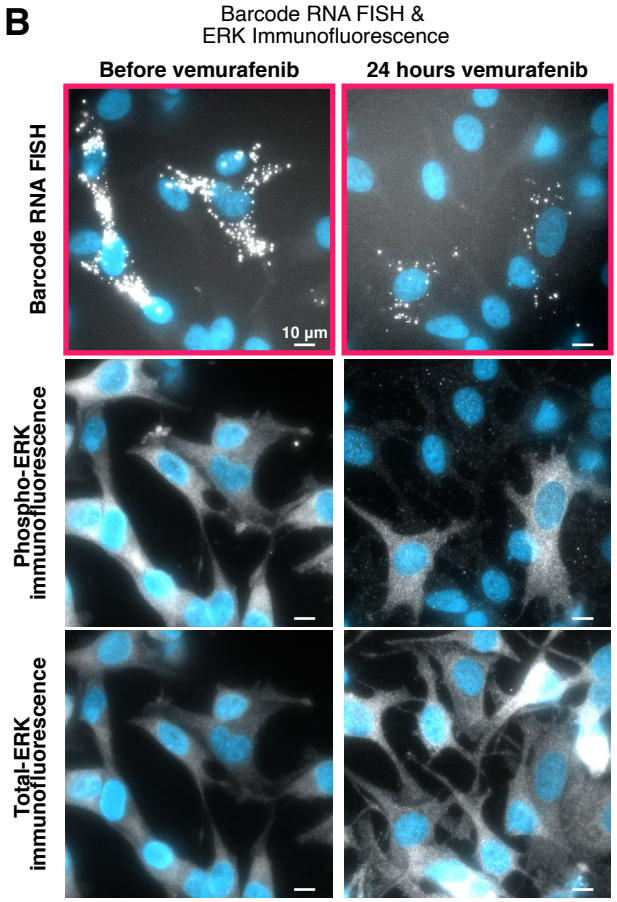


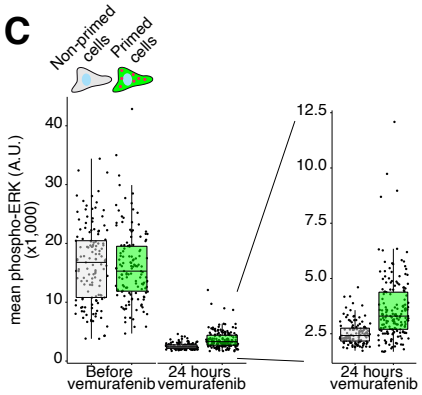


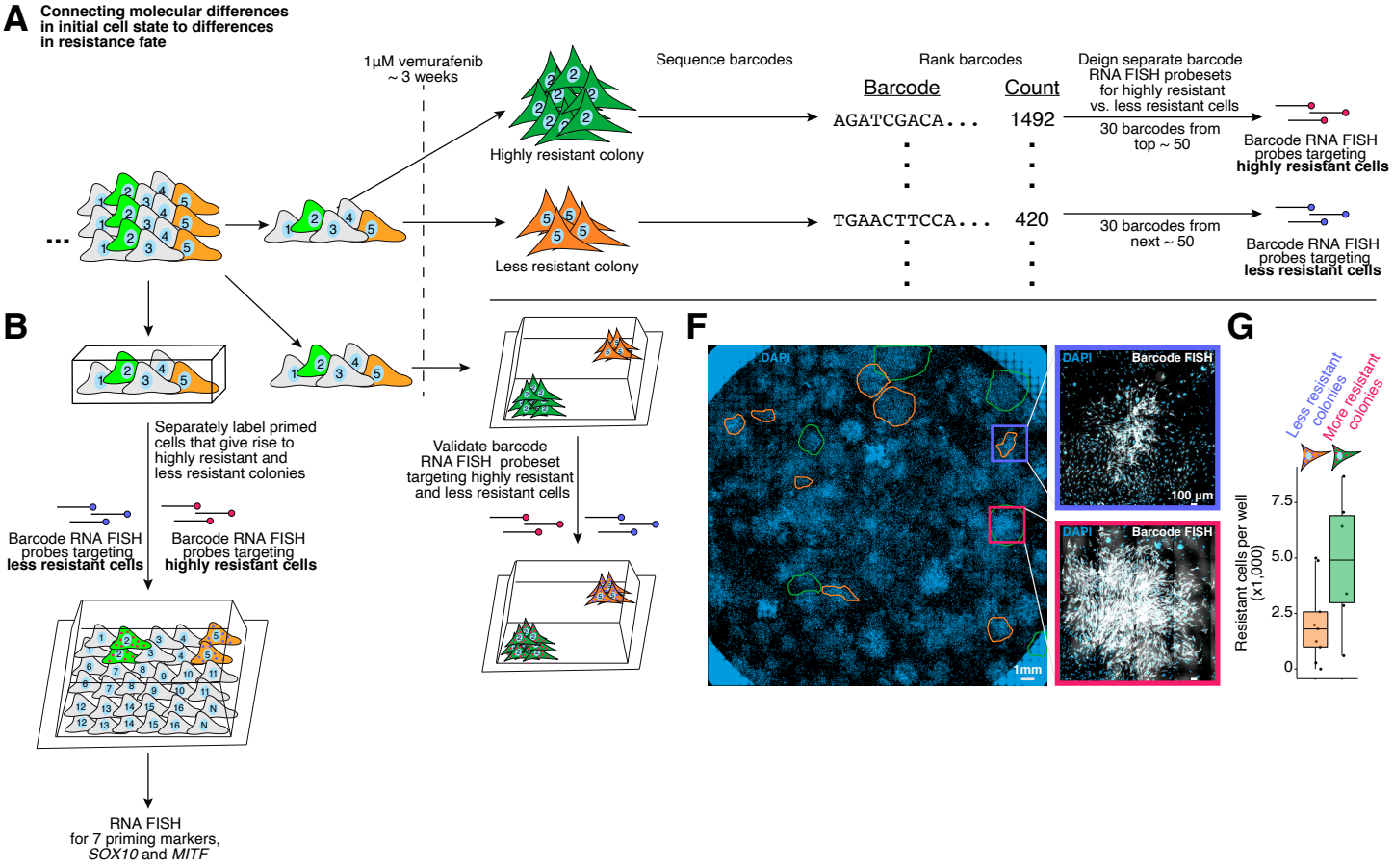


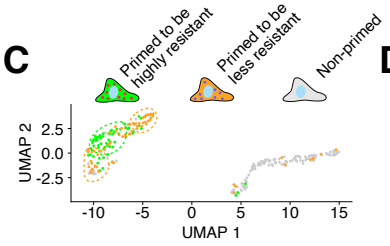


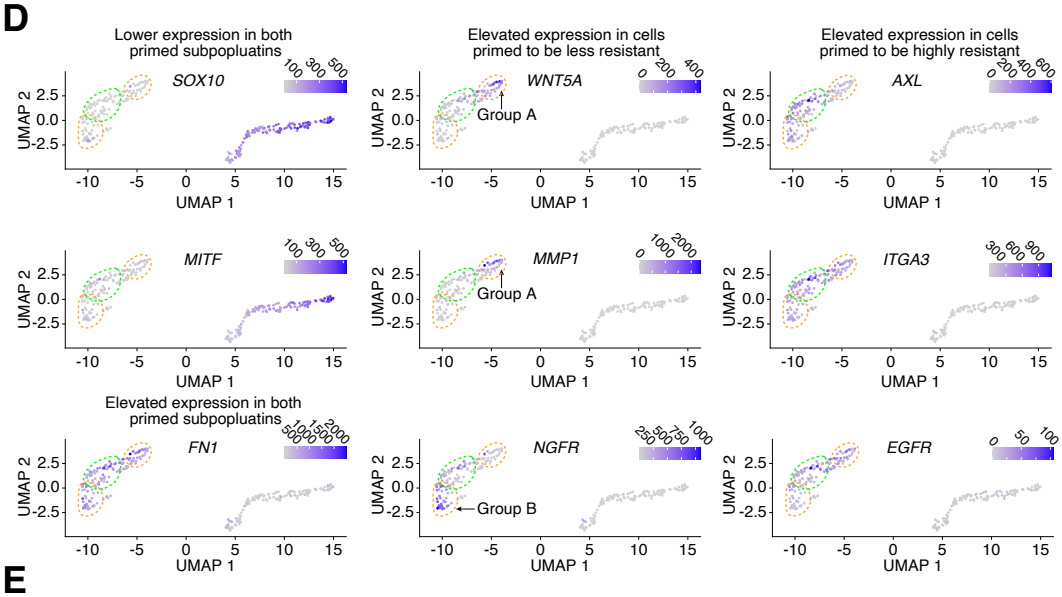


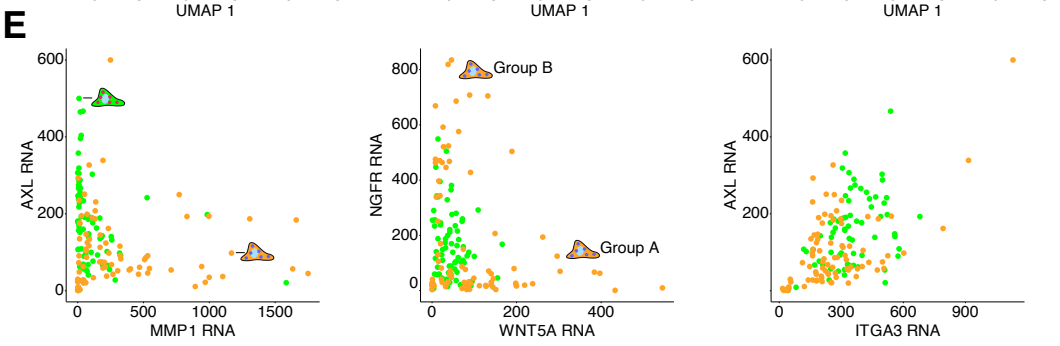


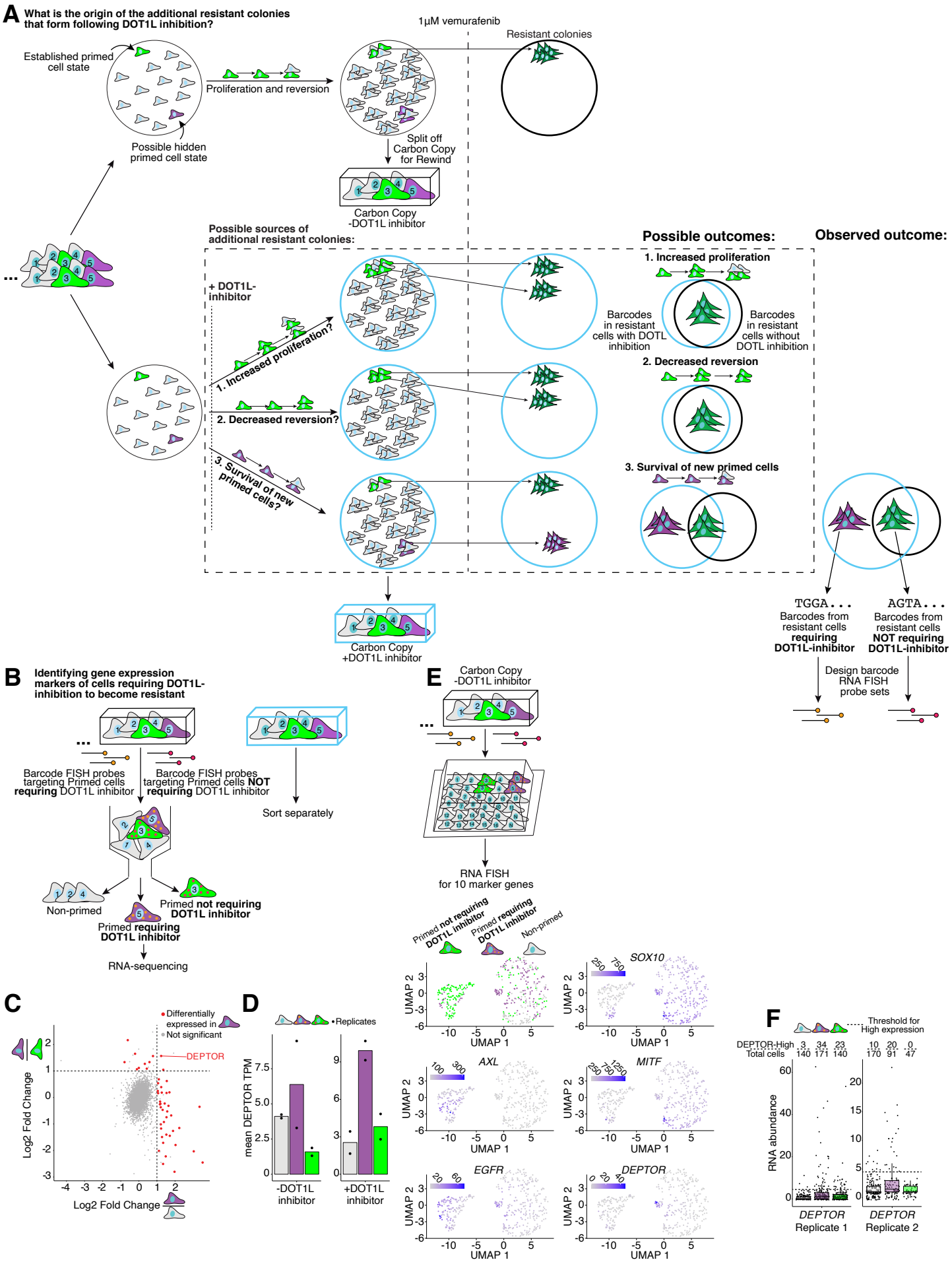


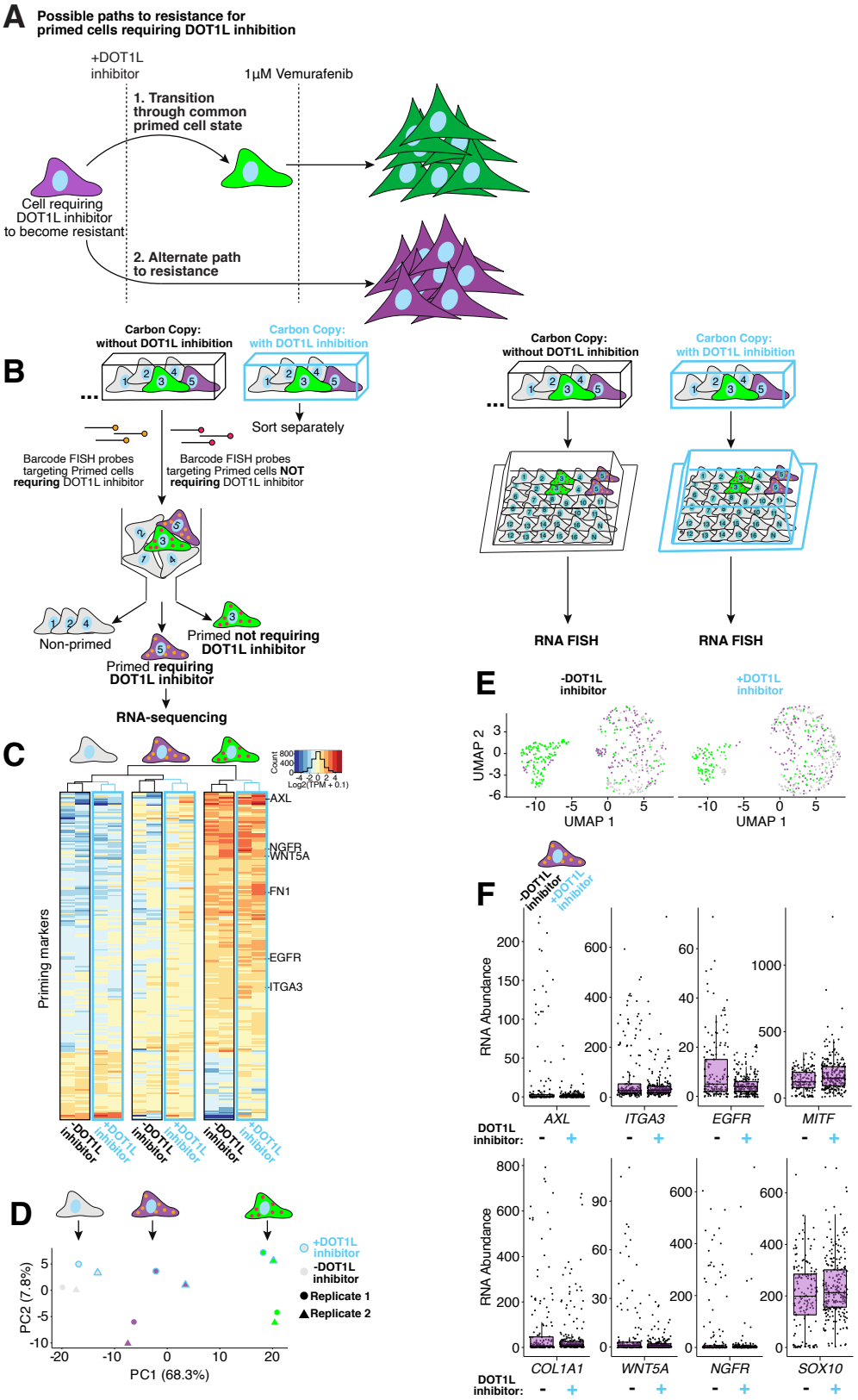




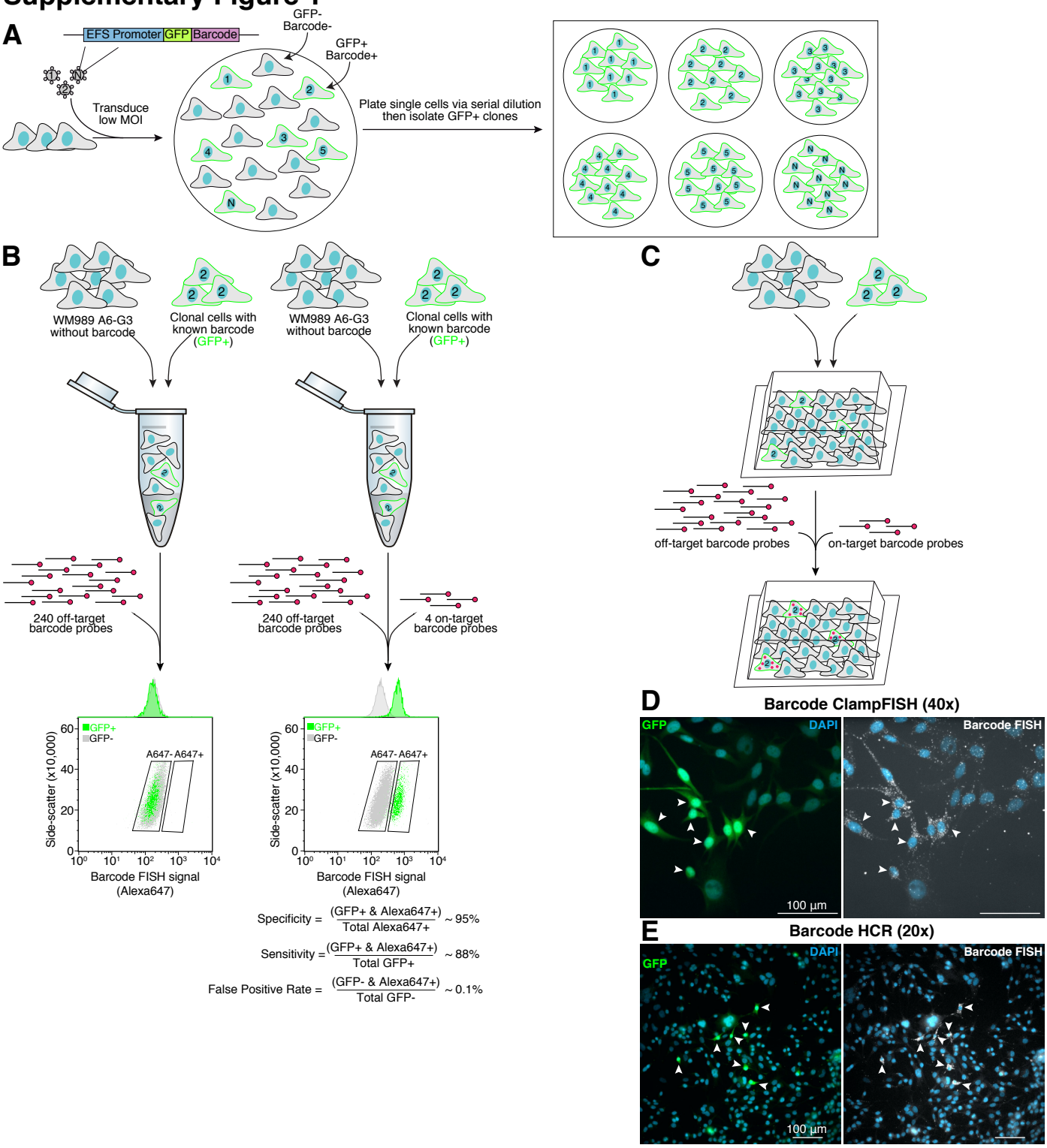






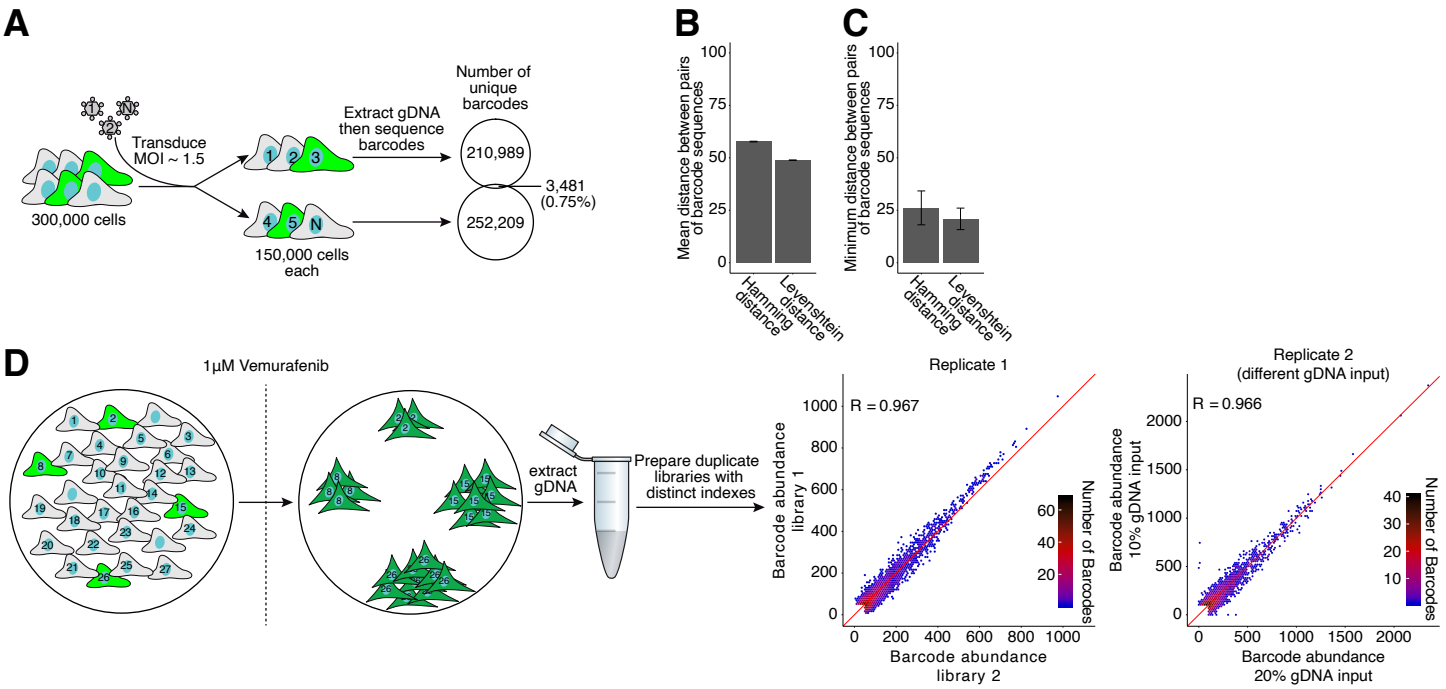


Supplementary Figure 1



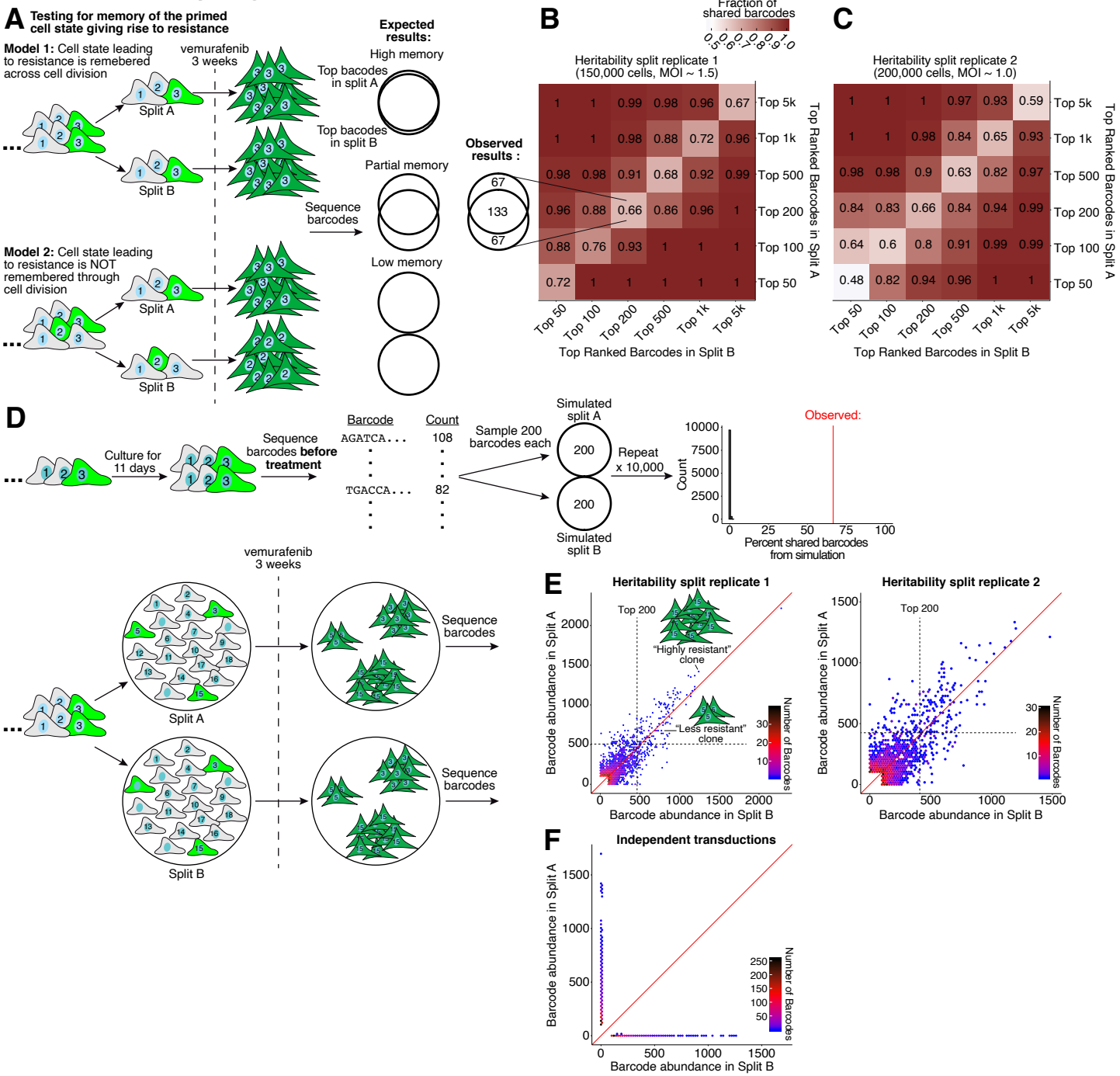
Supplementary Figure 1. Detections and isolation of cells expressing unique Rewind barcodes using RNA FISH. A. The Rewind construct encodes a 100 nucleotide barcode sequence ("WSN" repeated) in the 3' UTR of GFP downstream of a truncated EF-1 alpha promoter. To optimize barcode RNA detection using RNA FISH, we transduced WM989 A6-G3 cells then derived clonal cell lines each expressing a single barcode. We identified the barcode sequence in each clonal cell line via Sanger sequencing. As these barcodes are encoded on the same transcript as GFP, we could use GFP signal as a ground truth for optimizing our barcode RNA FISH protocol. **B.** We wanted to isolate 100s of primed cells for downstream assays such as RNA sequencing and RNA FISH. In order to capture this many cells using Rewind, we needed to probe for multiple barcodes simultaneously while maintaining a low false positive rate due to non-specific signal. To test this we would combine a clonal GFP+ cell line with a single known barcode with non-transduced GFP- cells then perform barcode RNA FISH using 4 probes targeting the known barcode along with 240 "off-target" probes (designed to target other sequenced barcodes). In parallel, we hybridized a second mix of cells with only the off-target probes. After hybridization, we ran these samples on a FACS instrument and used GFP fluorescence as ground truth for estimating the sensitivity and specificity of our barcode RNA FISH signal. After optimizing probe concentration and hybridization duration (see methods for final protocol), we could detect ~88% of GFP+ cells based on barcode FISH signal (true positives) while excluding 99.9% of GFP- cells (false positives). To isolate rare cell populations with Rewind (i.e. primed cells from our Carbon Copy), we typically used a more conservative Alexa647 gate in an attempt to further minimize false positives. **C-E.** We performed similar mixing experiments for in situ validation of our barcode FISH protocol using either ClampFISH(D) or HCR (E).

Supplementary Figure 2



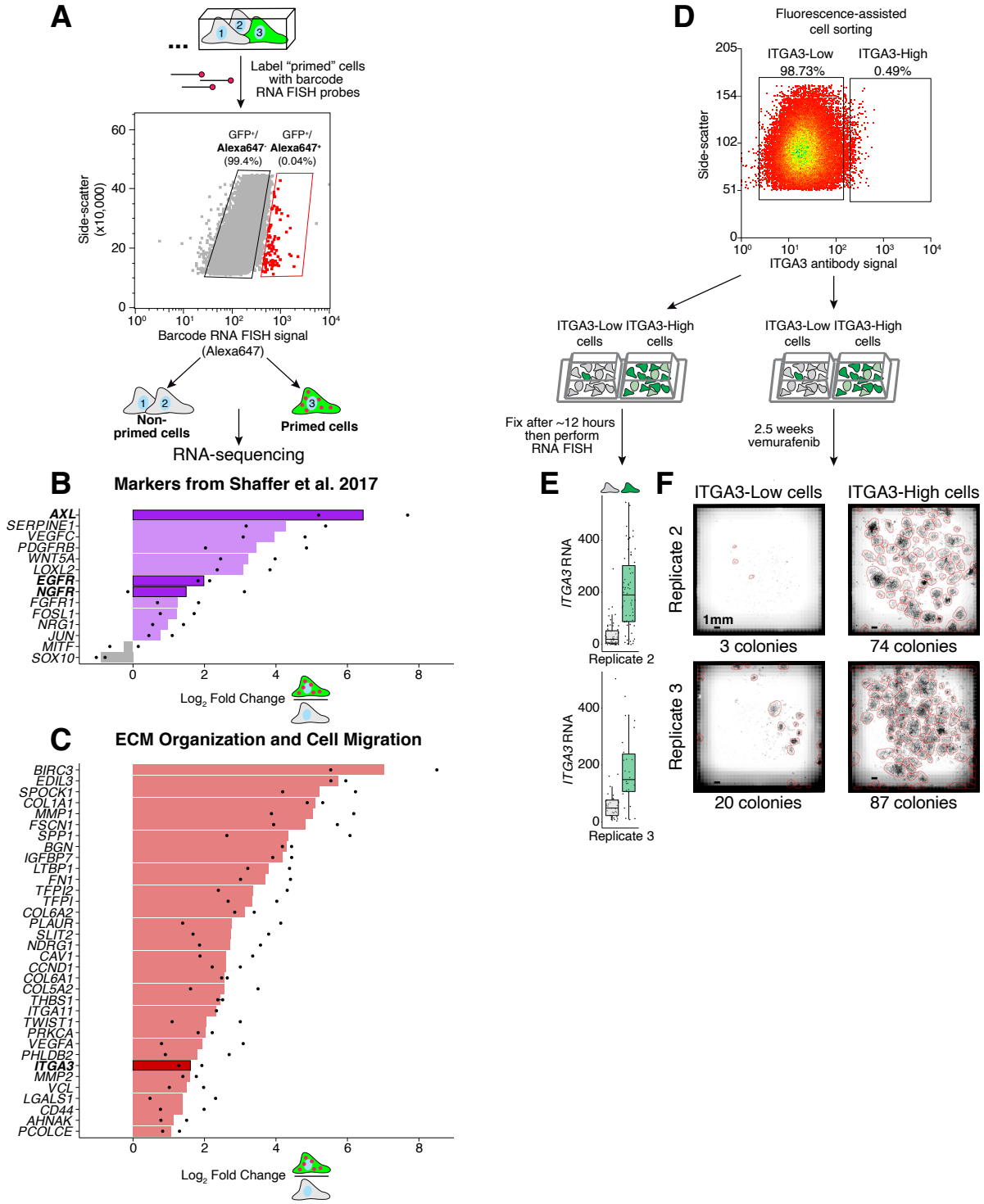
Supplementary Figure 2. The Rewind barcode library can uniquely label 100,000s of cells with transcribed barcodes that can be identified via sequencing. A. Critical for Rewind is the ability to uniquely label enough cells with our transcribed barcodes to observe rare phenomena such as drug resistance (frequency < 1:1000). To test this empirically, we separately transduced 2 groups of 150,000 cells at an MOI of ~1.5 , cultured the cells for ~ 1 day then extracted genomic DNA (gDNA) and sequenced their barcodes. Consistent with the starting cell number and MOI, we observed between 210,000 and 253,000 barcodes in the two samples (≥2 UMIs per million; see Methods for description of barcode count normalization) with fewer than 3,500 shared between the two groups. **B-C.** To ensure that we can uniquely distinguish barcodes despite errors during library preparation and sequencing, we estimated the observed barcode diversity in the data from A. We randomly sampled 1,000 barcode sequences and calculated the average and minimum sequence distance (both Hamming and Levenshtein sequence distances) for all 499,500 pairs of barcodes. We repeated this process 500 times and plotted the average and standard deviation of the observed sequence distances. **D.** We tested the reproducibility of our barcode sequencing protocol by preparing separate libraries with unique indexes using the same starting gDNA. As shown in the scatter plots we see a high correspondence in barcode abundances (UMIs per million) between these replicate libraries, even when using different amounts of gDNA (two separate experiments). Plotted are all barcode sequences with at least 50 reads per million in at least one of the two samples. We believe these data also suggest that our barcode sequencing protocol is quantitative, however we acknowledge the possibility that differences between barcode sequences could systematically bias library preparation or PCR. Our validation of barcode RNA FISH probes designed to target more abundant versus less abundant resistant cells (Figure 4) further suggests that our sequencing data provides a quantitative estimate of clone abundance.

Supplementary Figure 3



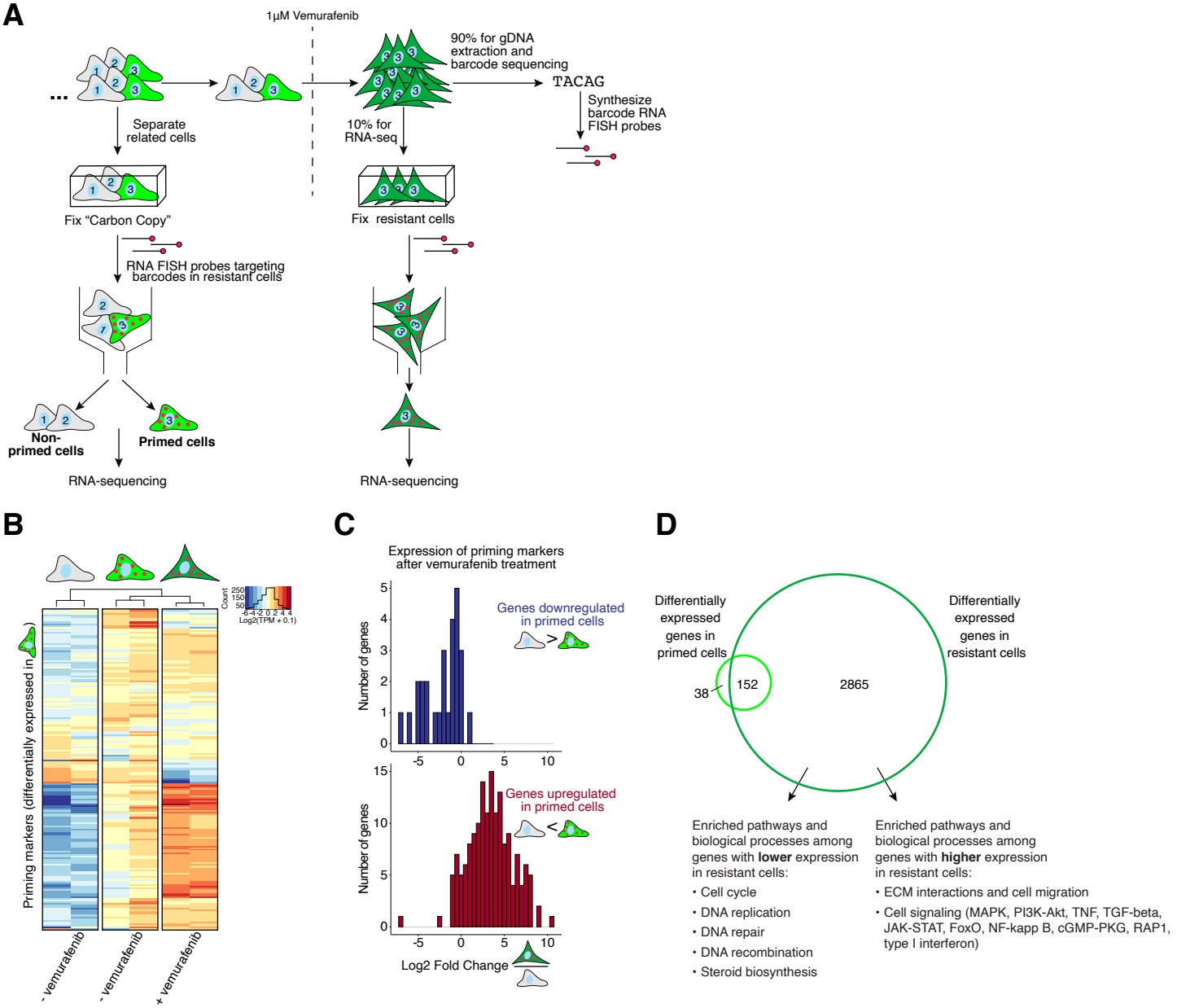
Supplementary Figure 3. Barcode sequencing of "twin" cultures treated with vemurafenib suggests that the primed cell state is maintained through several cell divisions. A. Schematic of the cellular barcoding approach used to test whether the primed cell state is "remembered" through cell division. We transduced ~150,000-200,000 WM989 A6-G3 cells with our Rewind barcode library and allowed the cells to divide for 11-12 days (~4-5 cell divisions). We then split the culture in two and treated both halves with 1 μM vemurafenib for 3 weeks. Finally we sequenced the barcodes in genomic DNA extracted from each culture then ranked the barcodes by abundance to identify those likely derived from resistant colonies (expected 100-400 unique barcodes from resistant colonies). In the absence of memory of the primed cell state, we expected to find unique sets of barcodes emerging in the two parallel cultures. In the presence of partial or complete memory, we expected to find some overlap in the barcodes identified in each culture. **B-C.** Heatmap shows the proportion of barcodes shared between the parallel cultures at different rank thresholds. For our Rewind experiments, we selected the top 100-200 barcodes for RNA FISH probe design. **D.** We wanted to rule out the possibility that differences in division rate between cells before adding vemurafenib could skew the distribution of barcodes enough to generate the observed barcode overlap by chance alone. We therefore sequenced barcoded cells after 11 days of growth (before vemurafenib treatment) to estimate the change in the barcode distribution due to differences in cell growth. We then simulated the split and vemurafenib treatment in A by randomly sampling 2 groups of 200 cells each from the observed barcode distribution and calculating the proportion of shared barcodes. The histogram shows the results of repeating this simulation 10,000 times (gray bars) with the red line indicating the experimentally observed proportion of shared barcodes (from B). **E.** We compared the abundance of barcodes from parallel cultures in A-C by plotting all barcodes with at least 100 UMIs per million in at least one sample (see Methods for description of normalization).To better visualize lower abundance barcodes, we binned the barcodes by count and colored each bin by its number of unique barcodes. Based on the observed correlation in barcode abundance between parallel cultures, we reasoned that while the number of cells that make up a vemurafenib resistant clone varies by more than an order of magnitude, these differences are at least partially pre-determined in the initial primed population 3 weeks earlier. **F.** Reassuringly, we do not observe a correlation in barcode counts between vemurafenib resistant cells from independent transductions

Supplementary Figure 4



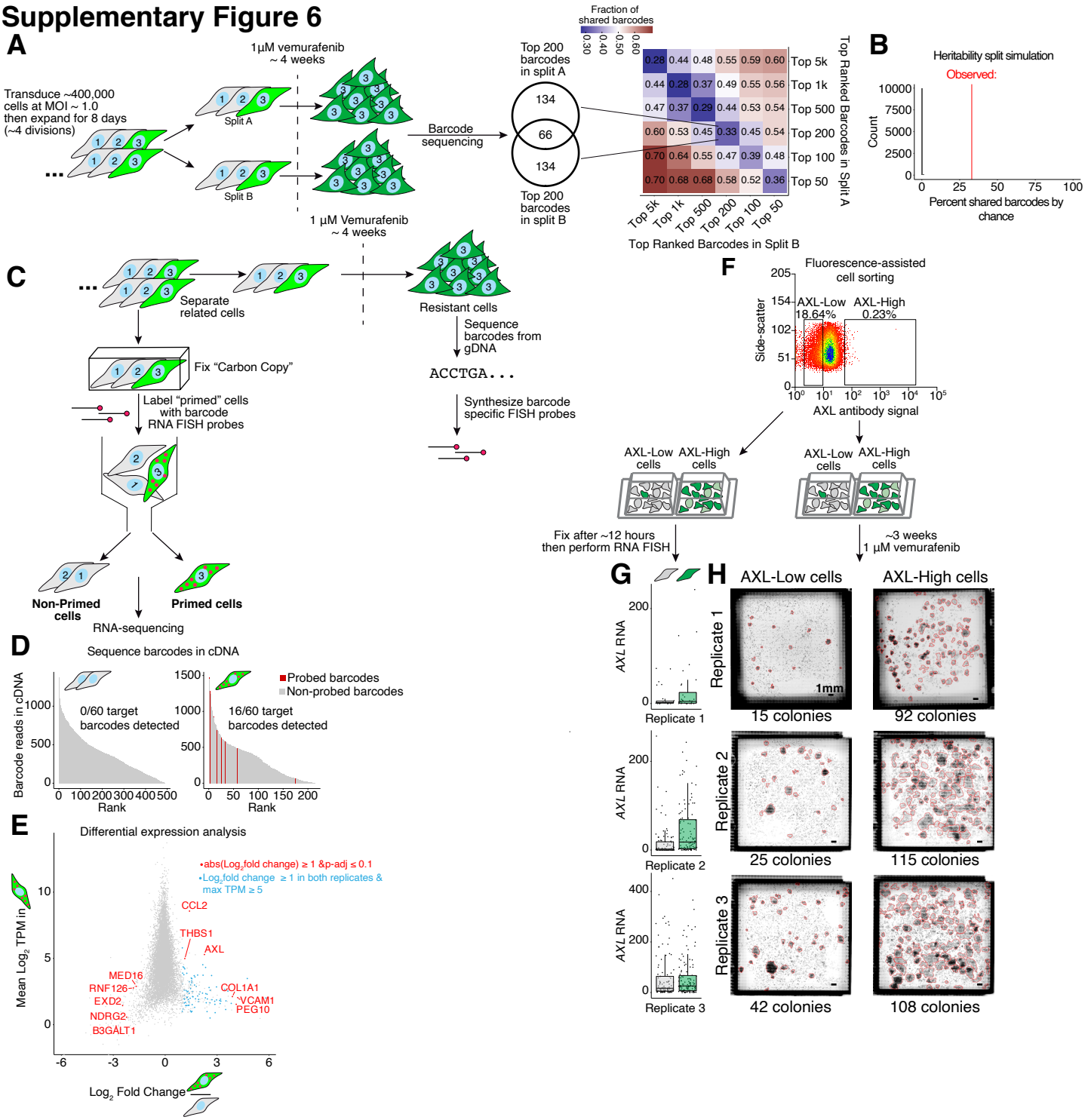
Supplementary Figure 4. RNA sequencing of primed WM989 A6-G3 isolated using Rewind identifies *ITGA3* as a prospective marker of vemurafenib resistance. A. Shown is flow cytometry data for the Rewind experiment presented in Figure 1. **B.** We compared the expression of genes previously implicated in priming and vemurafenib resistance (Shaffer et al. 2017) in primed and non-primed cells isolated using Rewind. Bargraphs indicate the average $\log_2$ fold change in expression in primed versus non-primed cells with individual replicates indicated as points. We previously demonstrated that drug-naive cells expressing the markers in bold (*AXL*, *EGFR*, and *NGFR*) are more likely to form vemurafenib resistant colonies and these same cells express lower levels of *SOX10* and *MITF* (gray bars) (Shaffer et al. 2017). **C.** We found an enrichment for genes associated with ECM organization and cell migration among differentially expressed genes comparing primed cells to non-primed cells (See Methods for pathway enrichment analyses and Supplementary Table 6 for FDR values). Bargraph corresponds to the average $\log_2$ fold change in expression in primed versus non-primed cells with individual replicates indicated as points. We bolded the gene, *ITGA3*, that we validated as a predictive marker of vemurafenib resistance (Figure 1 and panels D-F). We did not detect expression of *ITGA11* in non-primed cells in 1 of 2 replicates (it was detected in primed cells) and the presented data corresponds to the $\log_2$ fold change for one replicate. **D.** As described in Figure 1 and Methods, we stained cells with a fluorescently labelled antibody targeting ITGA3, then sorted equal numbers of the brightest ~0.5% (ITGA3-High) and remaining ~99% (ITGA3-Low) cells. We plated ~1/3 of these cells onto one plate for measuring *ITGA3* expression by RNA FISH and plated the rest onto a separate plate for treating with 1 μM vemurafenib. After approximately 18 days of treatment, we fixed the cells, stained nuclei with DAPI then imaged the entire wells to quantify the number of resistant colonies and cells. **E.** quantification of ITGA3 RNA by RNA FISH in sorted cells from D. **F.** Whole-well scans of sorted cells from D. after vemurafenib treatment.

Supplementary Figure 5



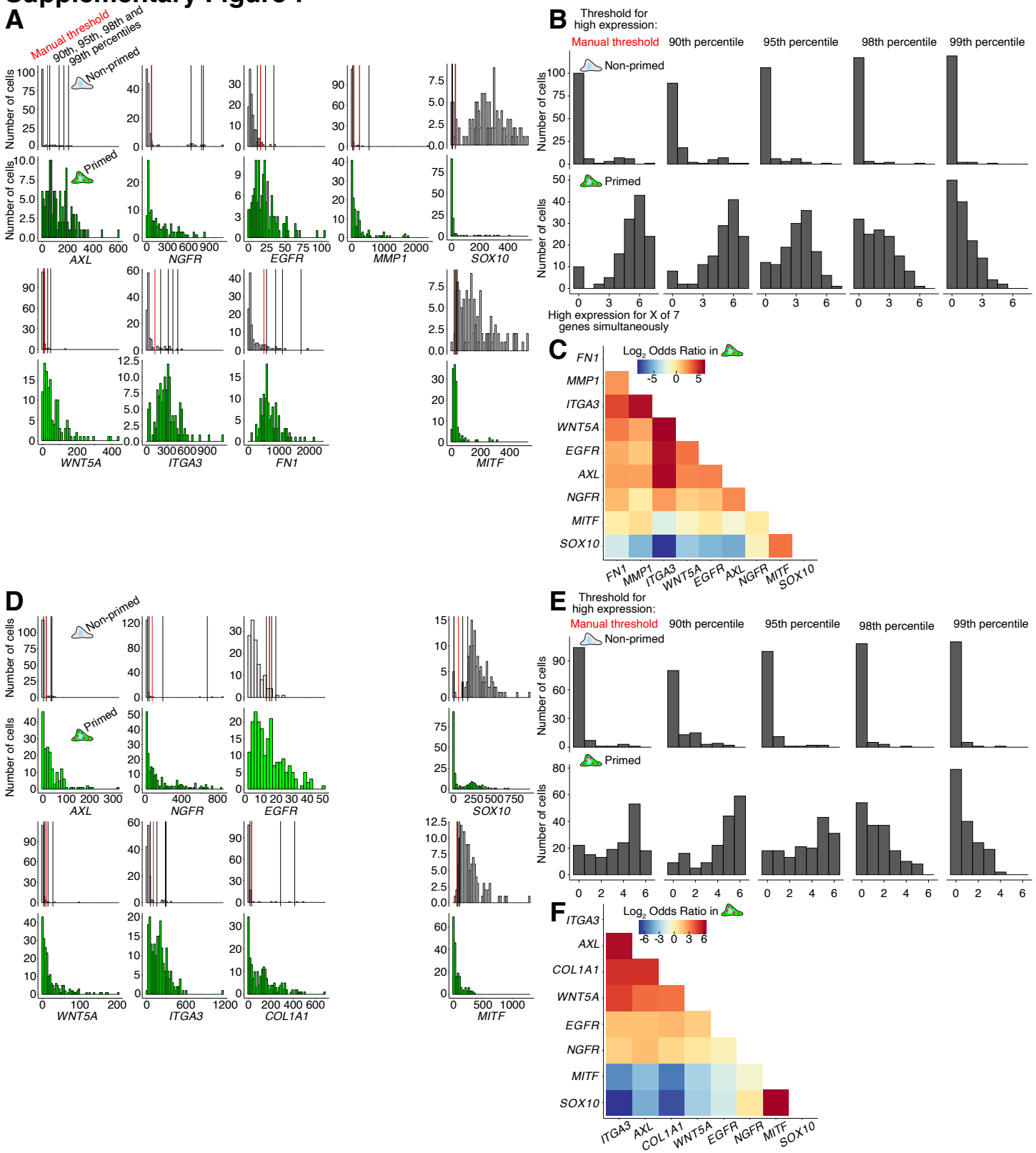
Supplementary Figure 5. Most priming markers remain transcriptionally altered after 3-weeks of vemurafenib treatment, and an additional ~3000 genes become differentially expressed. A. When performing Rewind in WM989 A6-G3, we collected ~10% of resistant cells for transcriptome profiling alongside cells sorted from our Carbon Copy. This enabled us to ask whether the gene expression changes observed in primed cells persist during their transition into stably resistant cells following vemurafenib treatment. **B.** To address this question, we compared the expression of priming markers in non-primed cells (left), primed cells (middle) and vemurafenib resistant cells (right). We defined priming markers as genes differentially expressed (p-adjusted ≤ 0.1 and abs($\log_2$ fold change) ≥1) in drug-naive primed cells versus non-primed. **C.** The majority of markers with increased expression in primed cells remained highly expressed or became further elevated in resistant cells. In addition, the majority of markers with lower expression in primed cells had lower expression in resistant cells. **D.** We found more than 3,000 genes differentially expressed comparing resistant cells to drug-naive non-primed cells of which the ~200 priming markers represented a small subset. We performed pathway and gene-ontology analyses on the genes differentially expressed only in resistant cells and highlight several recurring annotations (see Supplementary Table 7 for a complete list and FDR values).

Supplementary Figure 6



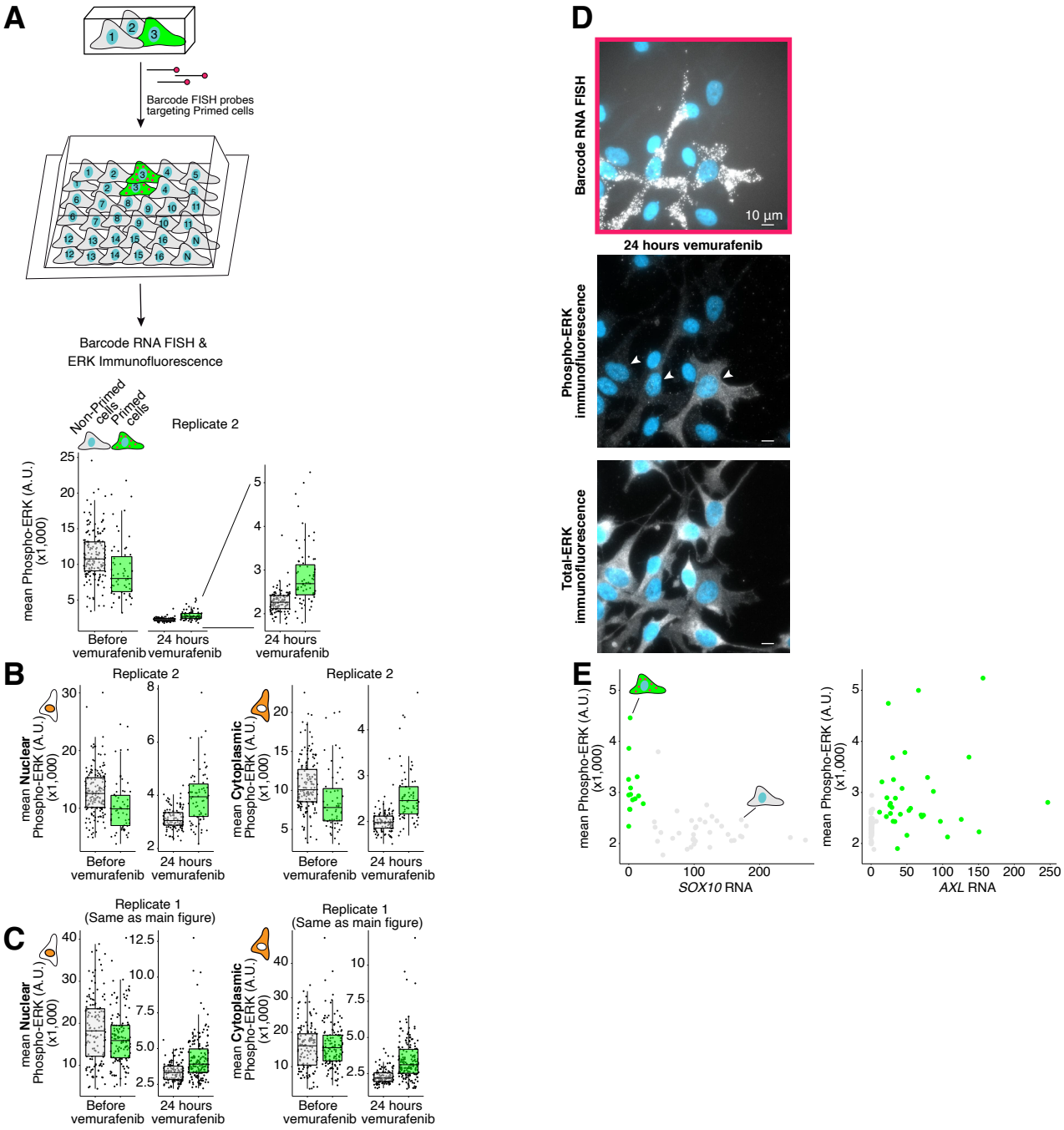
Supplementary Figure 6. Rewind on WM983b E9-C6 identifies AXL as a marker of primed cells giving rise to vemurafenib resistance . A. Starting with ~400,000 cells transduced at an MOI ~ 1.0, we repeated our "heritability split experiments" in WM983b E9-C6 to determine if the initial primed cell state giving rise to vemurafenib resistance was maintained through several cell divisions. The venn diagram and heatmap on the right show the proportion of barcodes shared between the parallel cultures at different rank thresholds (See Methods and Supp. Fig 3 for details). **B.** To estimate the probability of seeing the observed fraction of shared barcodes due to chance, we simulated the experiment using data from WM93b E9-C6 cells transduced as in A then cultured for 7 days before sequencing. We simulated the split and vemurafenib treatment by randomly sampling 2 groups of 200 cells each from the barcode distribution and calculating the proportion of shared barcodes. The histogram shows the results of repeating this simulation 10,000 times (gray bars) with the red line indicating the observed proportion of shared barcodes from A. **C.** Using Rewind we isolated primed WM983b E9-C6 cells from a Carbon Copy fixed prior to vemurafenib treatment. We then performed RNA sequencing and barcode sequencing on cDNA prepared from sorted cells. **D.** 16 out of ~200 barcodes recovered from primed cells were among the 60 barcodes found in resistant cells and targeted by our RNA FISH probes. This represents a ≥ 500-fold enrichment of these clones over their frequency in the initial population (60/400,000). **E.** We used DESeq2 to identify differentially expressed genes (p-adjusted ≤ 0.1 and abs(log_2 fold change) ≥1) in primed cells versus non-primed cells (red points). Compared with similar experiment in WM989 A6-G3, few genes passed our significance cutoff which may reflect the shorter memory of the primed cells state (see A) and the lower purity of our Rewind sort (D). We highlighted in blue genes that did not pass our significance cutoff but showed ≥ 2-fold higher expression in primed cells in 2 out of 2 replicates. **F.** As described in Methods, we sorted drug-naive cells expressing high levels of AXL or low levels of AXL and compared their response to vemurafenb treatment. **G.** With ~ 1/3 of the cells we performed RNA FISH to check AXL expression in the two sorted populations. **H.** We treated the remainder with vemurafenib for 3 weeks then imaged the wells to quantify the number of resistant colonies (see Methods for details).

Supplementary Figure 7



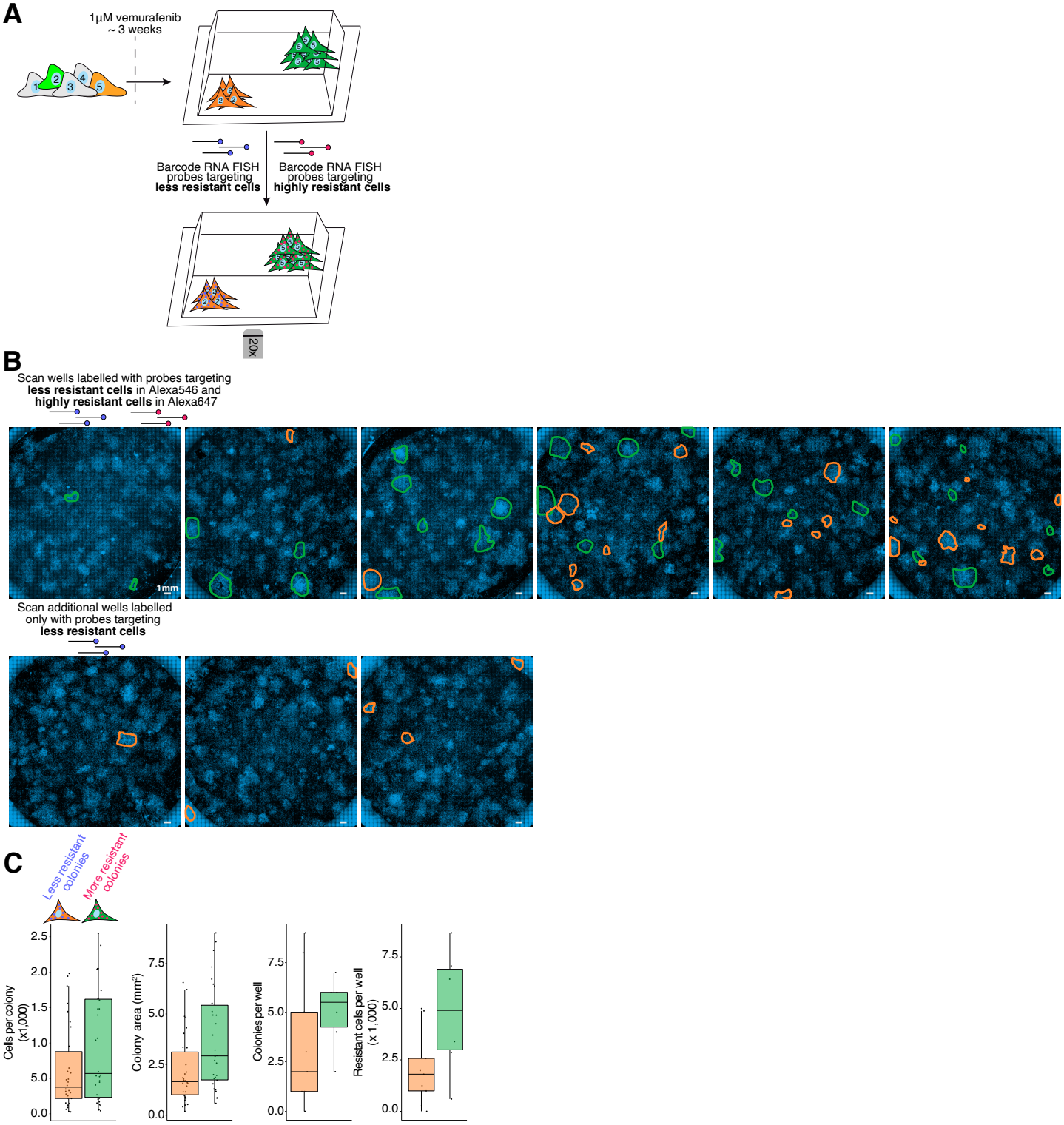
Supplementary Figure 7. Primed WM989 A6-G3 express high levels of multiple markers simultaneously. A. As described in Figure 2 and Methods, we identified primed cells that give rise to vemurafenib resistance using Rewind then measured single cell expression of 7 priming markers, *SOX10* and *MITF* by RNA FISH. The histograms show the expression distributions for these genes in primed cells (green) and randomly selected non-primed cells (gray). We used these distributions to set thresholds for binning high expressing cells and, in turn, characterizing the co-expression of these markers in single cells. The black vertical lines correspond to the 90th, 95th, 98th and 99th percentiles of expression in non-primed cells and the red vertical line corresponds to the threshold used for Figure 2. **B.** Based on the indicated thresholds, we calculated the number of primed cells (bottom row) and non-primed cells (top row) that express high levels multiple markers simultaneously (number of markers indicated on X-axis). **C.** Using the manual thresholds from A (red vertical lines) we calculated the odds ratio for pairs of markers being co-expressed at high levels in the same primed cells. **D-F.** We repeated the analyses in A-C on non-primed and primed cells from a separate experiment. These primed cells correspond to cells that do not require DOT1L inhibition to become vemurafenib resistant from a Carbon Copy treated with vehicle control (DMSO) for 5 days before fixation (see Figure 5 and Methods for further details).

Supplementary Figure 8



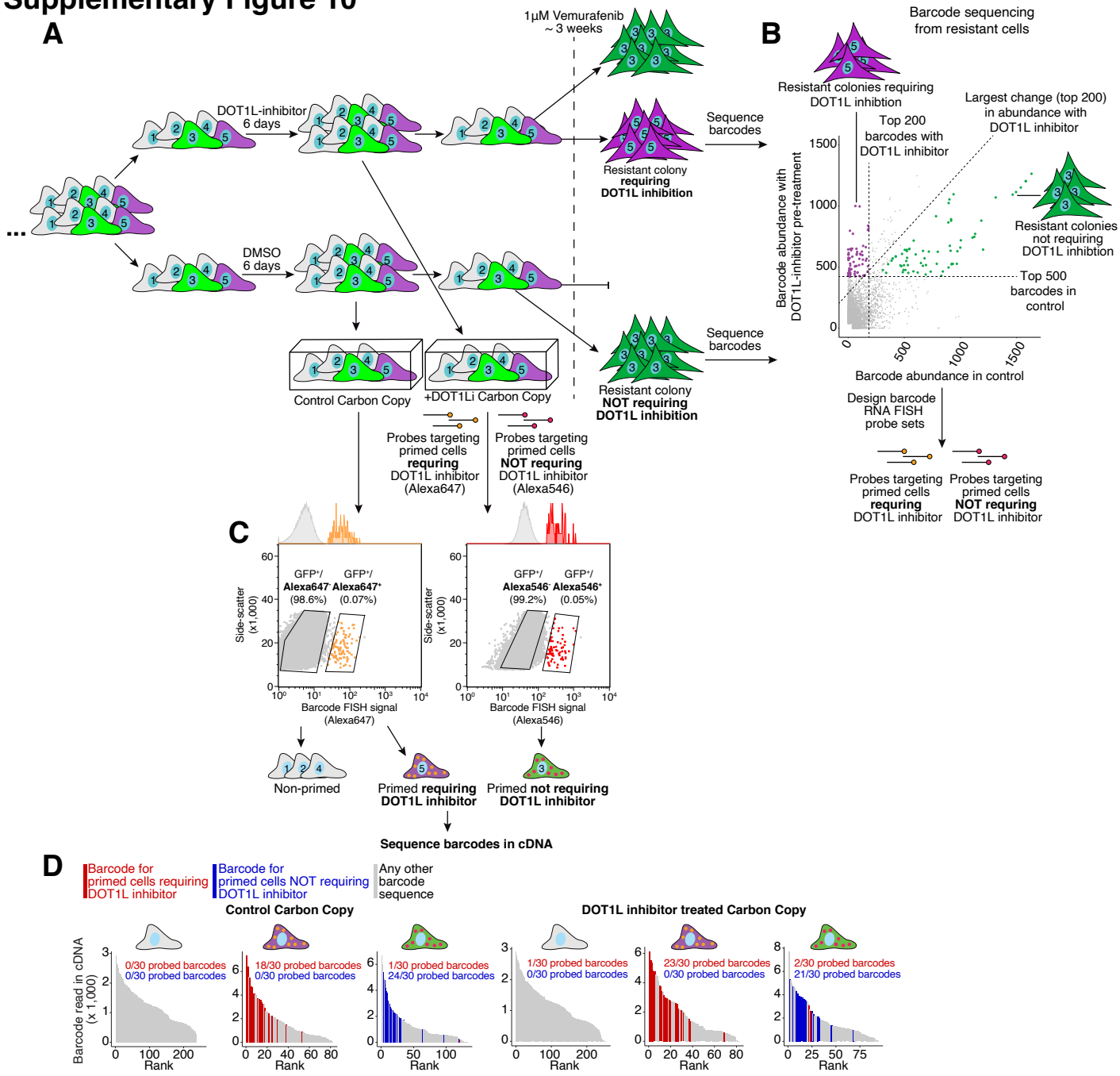
Supplementary Figure 8. Primed cells giving rise to vemurafenib resistance have higher levels of phosphorylated ERK than non-primed cells 24 hours after vemurafenib treatment, but there remains cell to cell variability. A. We used Rewind to quantify dual-phospho ERK (p44/p42, pERK) levels in primed cells (green) and non-primed cells (gray) before and 24 hours after vemurafenib treatment. Shown in A is a biological replicate of the experiment shown in Figure 3. For this second replicate, we amplified the barcode RNA FISH signal using ClampFISH instead of HCR (see Methods for details). Each point corresponds to an individual cell. Average values are calculated for a single Z-plane as described in Methods. **B-C.** In addition to average phosphorylated ERK across the entire cell, we calculated average phosphorylated ERK within just the cell nucleus (left) or just the cytoplasm (right). We found similar results using any of these metrics. **D.** While on average primed cells had higher levels of phosphorylated ERK, we observed several clusters of primed cells (presumably closely related) with variable levels of phosphorylated ERK. Shown is one such example. We speculate that this variability may be a result of pulsatile MAPK signaling, which has been documented in other melanoma cell lines (Gerosa et al. 2019), and our snapshot measurement of ERK phosphorylation via immunofluorescence **E.** After identifying primed cells in situ, we performed both single-molecule RNA FISH and immunofluorescence to measure gene expression and phosphorylated ERK in the same single cells. Shown is the relationship between phosphorylated ERK levels and *AXL* (left) or *SOX10* (right) expression in individual primed (green points) and non-primed (gray points) cells. Within the primed cell population, we observe a fairly low correlation between phosphorylated ERK and expression of these markers, which we speculate may be a result of MAPK signalling fluctuating on a faster timescale than changes in gene expression.

Supplementary Figure 9



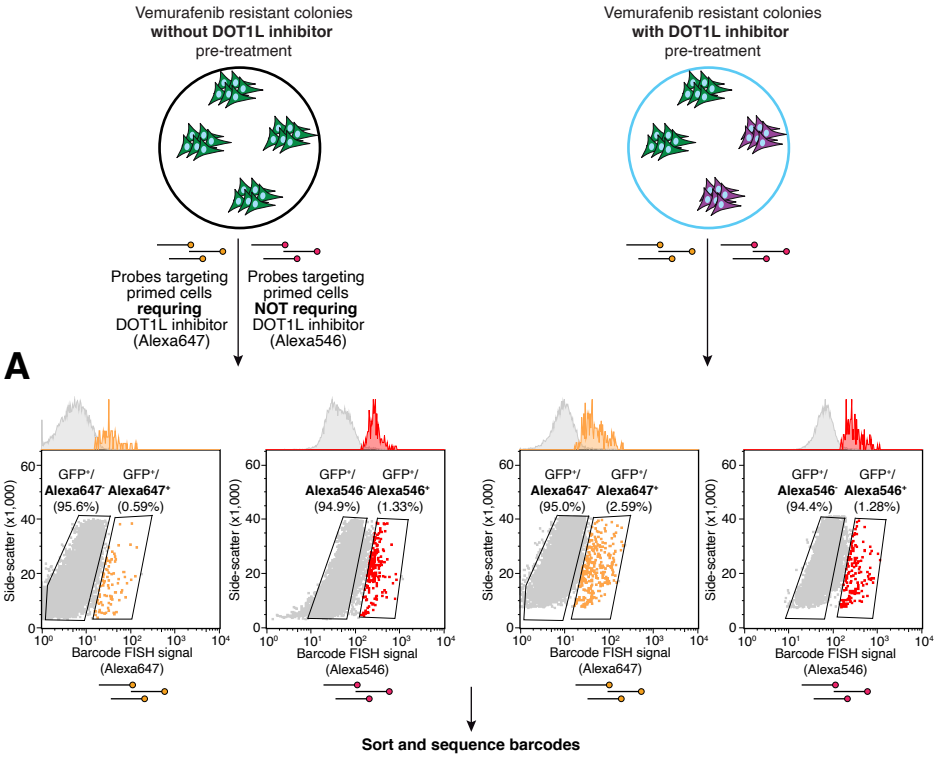
Supplementary Figure 9. Barcode RNA FISH can distinguish highly resistant clones from less resistant clones. A. As described in the Results for Figure 4, we identified barcodes in vemurafenib resistant cells by sequencing then ranked these barcodes by relative abundance and designed separate RNA FISH probe sets targeting the top ~ 50 and next ~ 50 barcodes. We reasoned that these two groups correspond to clones (cells sharing identical barcodes) with greater and fewer resistant cells (see Supp. Figs 2-3 for further reasoning behind this assumption). We refer to these two groups as "highly resistant" and "less resistant" as these groups roughly correspond to degrees of fitness in drug and our colleagues found the terms "more abundant" and "less abundant confusing. To empirically test that our probe sets distinguish different groups of resistant cells, we labeled resistant colonies derived from the same population of barcoded cells with our two probe sets, each coupled to a differ"ent fluorescent dye (AlexaFluor 546 and AlexaFluor647). We then imaged the cells and quantified the number of colonies and resistant cells labeled with each probe set. **B.** Stitched scan images of resistant cells from A. We used custom software as described in Methods to annotate colonies labeled with our "highly resistant" (green) or "less resistant" (orange) probe sets. To check that we were not seeing differences in the number of labeled cells between probe sets due to differences in HCR hairpin and dye, we swapped the initiator sequence (and corresponding hairpin and dye) for the less resistant clone probeset, and labeled an additional 3 wells shown in the bottom row. **C.** Using custom software described in Methods, we quantified the number of colonies and number of cells labeled with each probe set.

Supplementary Figure 10



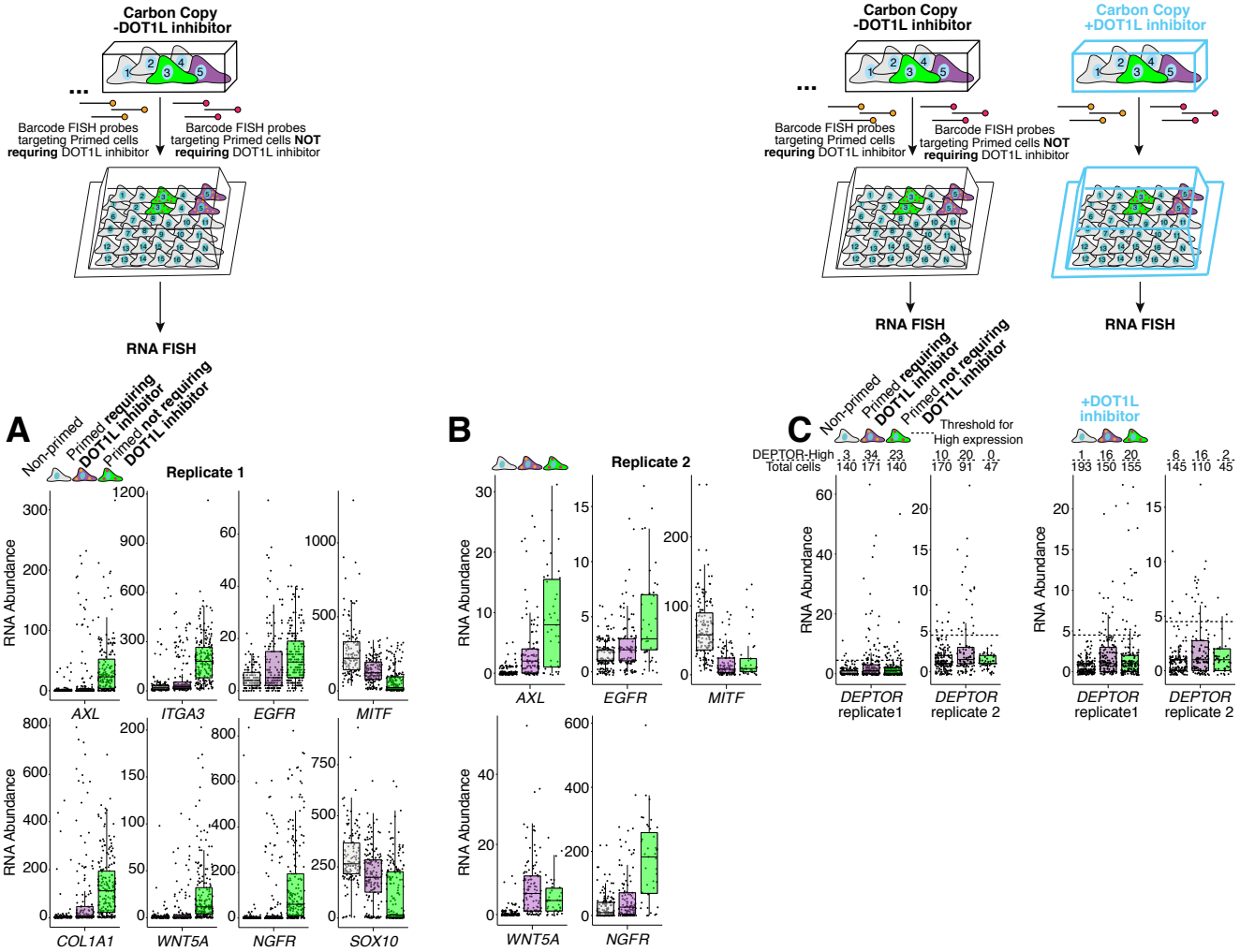
Supplementary Figure 10. Identification and isolation of cells requiring DOT1L inhibition to become vemurafenib resistant. A. Our approach for identifying the cells that require DOT1L inhibition to become resistant involved first transducing ~400,000 WM989 A6-G3 cells with the Rewind barcode library, letting the cells divide for ~6 days then splitting the culture into two groups. We treated one group with 4 μM DOT1L inhibitor (pinometostat) and the other with vehicle control (DMSO) for 6 days. We then split each group again, fixing half as our "Carbon Copy" and treating the other half with 1 μM vemurafenib for ~2.5 weeks. After this vemurafenib treatment, we extracted genomic DNA from the remaining cells for barcode sequencing. **B.** We compared the relative abundance of each barcode identified in resistant cells pre-treated with DOT1L inhibitor versus resistant cells pre-treated with vehicle control as shown in A. This comparison revealed a subset of barcodes with a greater abundance in resistant cells pre-treated with DOT1L inhibitor than resistant cells pre-treated with vehicle control (purple points). We used these barcodes to design RNA FISH probes targeting primed cells requiring DOT1L inhibition to become vemurafenib resistant. A separate set of barcodes showed similar high abundance with or without DOT1L inhibition (green points), which we used to design RNA FISH probes targeting primed cells not requiring DOT1L inhibition to become resistant. We note that the barcodes with high abundance in both conditions (green points), appear slightly less abundant in resistant cells pre-treated with DOT1L inhibitor. We believe this reflects the greater total number of surviving cells in this condition influencing the normalization of our barcode sequencing data. **C.** We used our barcode RNA FISH probe sets to sort cells from the Carbon Copy pre-treated with DOT1L inhibitor and separately from the Carbon Copy pre-treated with vehicle control. For our first experimental replicate, we used probes coupled to distinct fluorophores (Alexa647 and Alexa546) to separately isolate primed cells requiring DOT1L inhibitor to become vemurafenib resistant and primed cells not requiring DOT1L inhibitor. We isolated equal numbers of GFP+/Alexa546-/Alexa647- non-primed cells. For the second experimental replicate (not shown) we divided our Carbon Copies in two and hybridized each half with separate barcode RNA FISH probe sets coupled to Alexa647. After sorting, we prepared libraries for RNA sequencing and with the extra cDNA, performed targeted barcode sequencing. **D.** Bargraphs show the abundance (y-axis) and rank (x-axis) for barcodes (≥ 5 normalized reads) sequenced from sorted cells as described in C. Sequences matching barcodes from our probe set targeting cells requiring DOT1L inhibition to become resistant are colored red. Sequences matching barcodes from our probe set targeting cells not requiring DOT1L inhibition to become resistant are colored blue. All other barcode sequences are colored gray. Inset shows the fraction of barcodes from each probeset identified in each sample.

Supplementary Figure 11



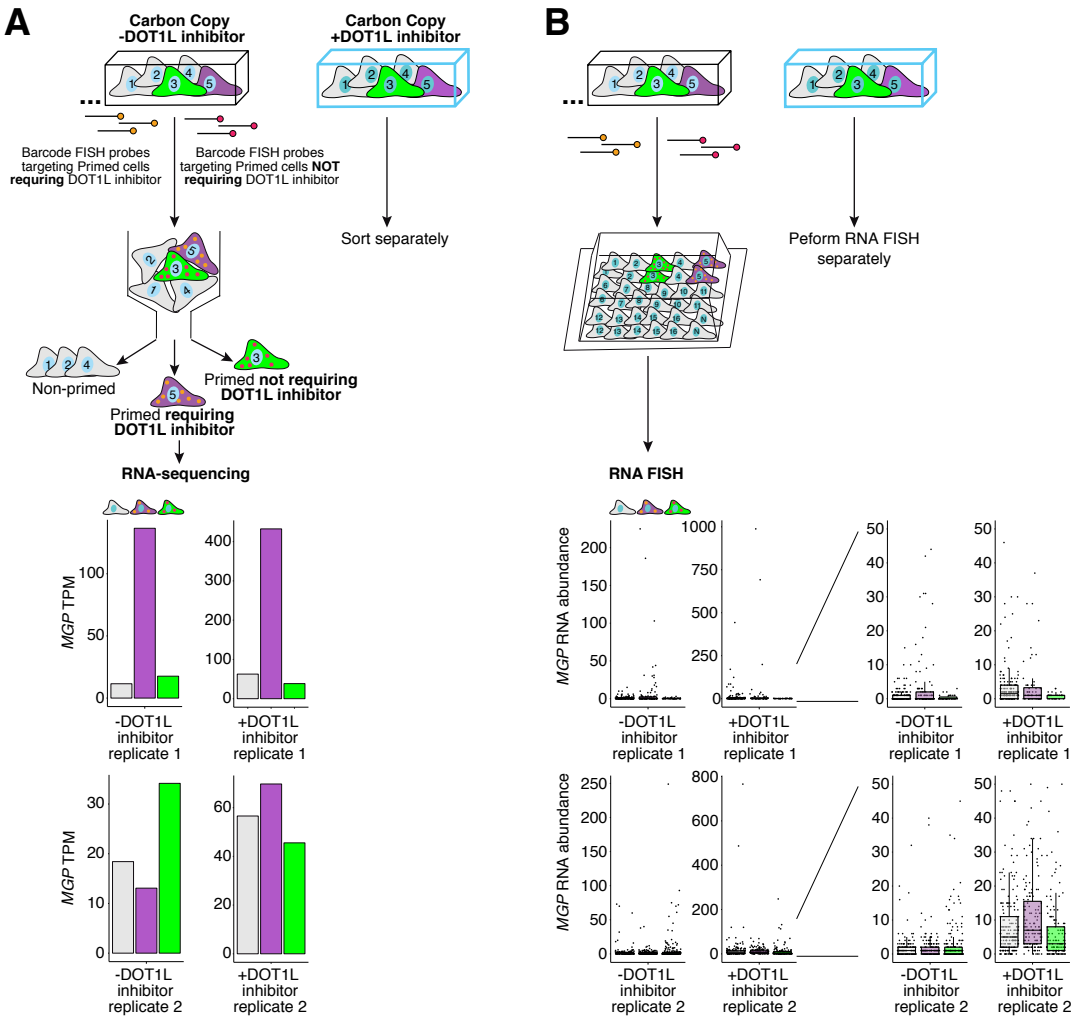
Supplementary Figure 11. Validation of barcode RNA FISH probe sets targeting cells that require DOT1L inhibition to become resistant and cells that do not require DOT1L inhibition. A. When we performed the Rewind experiments on DOT1L inhibitor pre-treated and vehicle control treated cells (see Figure 5 and Supp. Fig 9), we fixed 10% of the vemurafenib resistant colonies for validating our barcode RNA probes. We expected the probes designed to target cells that require DOT1L inhibition to become resistant to label more resistant cells that were pre-treated with DOT1L inhibitor than resistant cells pre-treated with vehicle control. Conversely, we expected the probes designed to target cells that do not require DOT1L inhibition to label a similar fraction of resistant cells in both conditions. As expected the probes targeting cells that require DOT1L inhibition labeled approximately 4x as many resistant cells pre-treated with DOT1L inhibitor compared to resistant cells pre-treated with vehicle control. We observed a minimal difference in labeling using probes designed to target cells that do not require DOT1L inhibition. To verify that these gates corresponded to cells expressing the targeted barcodes, we sorted the populations and sequenced their barcodes. **B.** Bargraphs show the abundance (y-axis) and rank (x-axis) for barcodes (≥ 5 normalized reads) sequenced from sorted cells as described in A. Sequences matching a barcode from our probe set targeting cells requiring DOT1L inhibition to become resistant are colored red. Sequences matching a barcode from our probe set targeting cells not requiring DOT1L inhibition are colored blue. All other barcode sequences are colored gray. Inset shows the fraction of barcodes from each probeset identified in each sample.

Supplementary Figure 12



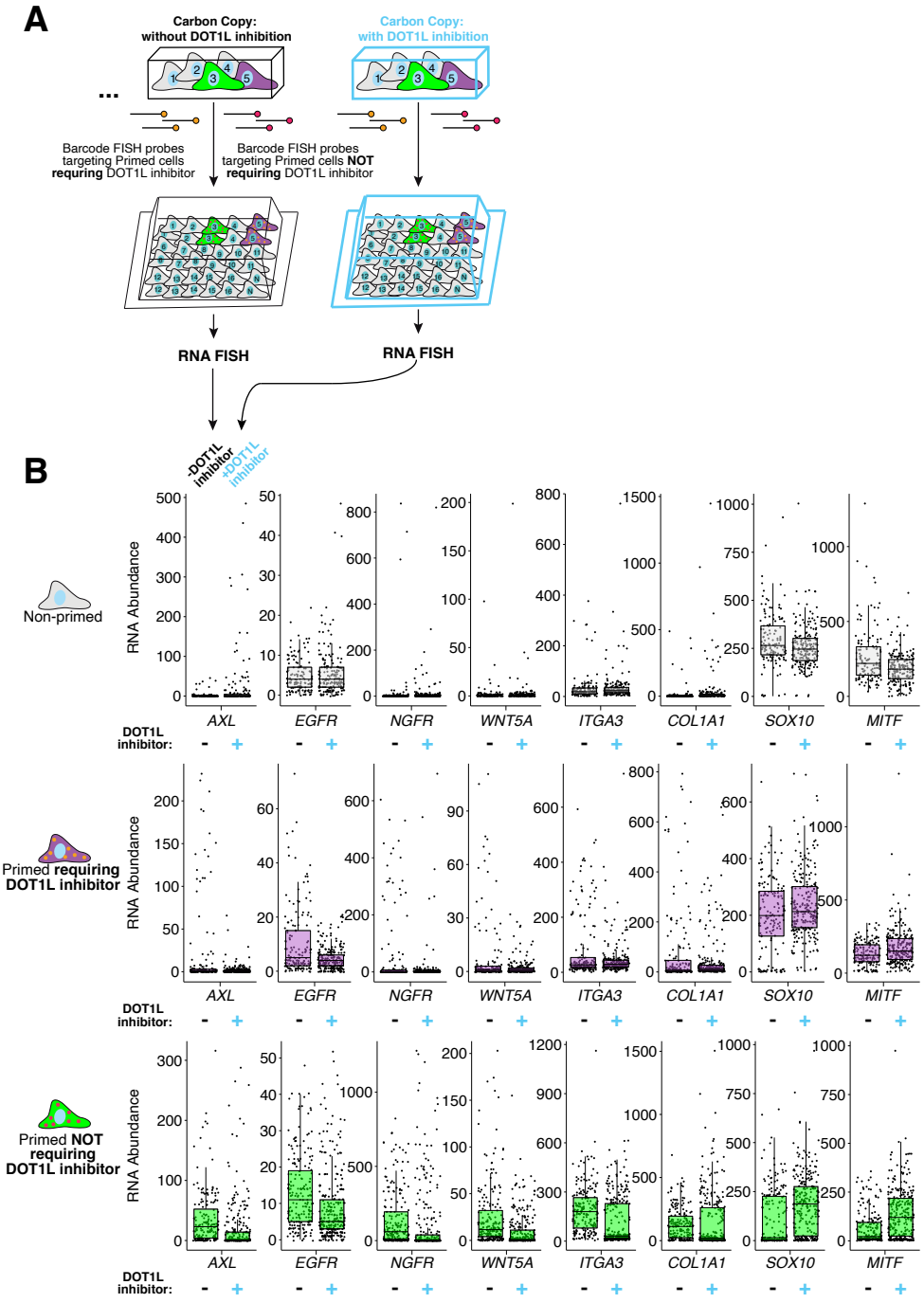
Supplementary Figure 12. The transcriptional profile of cells that require DOT1L inhibition to become vemurafenib resistant is distinct from cells that do not require DOT1L inhibition. **A.** As described in Results and Methods, we used Rewind to identify cells requiring DOT1L inhibition to become vemurafenib resistant and cells not requiring DOT1L inhibition in Carbon Copies treated with vehicle control (DMSO) and fixed before vemurafenib treatment. We then performed RNA FISH to measure single-cell gene expression of established priming markers. This revealed that cells requiring DOT1L inhibition to become vemurafenib resistant (purple) expressed levels of these markers that were slightly higher than non-primed cells (gray), but far lower than cells not requiring DOT1L inhibition to become resistant (green). Cells requiring DOT1L inhibition expressed intermediate levels of *SOX10* and *MITF*, which may reflect some degree of dedifferentiation. **B.** We repeated this experiment and measured single cell gene expression by RNA FISH on a subset of marker genes. **C.** As shown in Figure 4, RNA sequencing of cells sorted from Carbon Copies found elevated expression of *DEPTOR* specifically in the cells that require DOT1L inhibition to become resistant. To verify this finding we measured single cell expression of *DEPTOR* via RNA FISH in both Carbon Copies pre-treated with vehicle control (left) and Carbon Copies pre-treated with DOT1L inhibitor (right). In both of these conditions we found an enrichment of cells expressing high levels of *DEPTOR* (threshold indicated by dotted line; ~ 95th percentile of non-primed cells) in the subpopulation that requires DOT1L inhibition to become resistant, implicating DEPTOR as a positive marker of this subpopulation.

Supplementary Figure 13



Supplementary Figure 13. MGP is not a consistent marker of cells that require DOT1L inhibition to become vemurafenib resistant. A. As described in the main text and Methods, we sought to use Rewind to identify markers specific to cells that require DOT1L inhibition to become vemurafenib resistant. In our first experimental replicate for RNA-sequencing, we found that primed cells requiring DOT1L inhibition to become resistant (purple) expressed 8-12 fold higher levels of MGP compared to either non-primed cells (gray) or primed cells not requiring DOT1L inhibition (green). This, however, did not replicate in a second experiment. **B.** Based on the magnitude of the initial observation, we nonetheless used RNA FISH to compare single cell expression of *MGP* between each subpopulation in a second set of Carbon Copies. The first RNA FISH experiment revealed an enrichment of cells expressing high levels of *MGP* in the subpopulation requiring DOT1L inhibition, however this did not replicate in a second experiment.

Supplementary Figure 14



Supplementary Figure 14. DOT1L inhibition partially decouples expression of established priming markers and vemurafenib resistance A. As described in the main text, we asked whether DOT1L inhibition enabled a new subpopulation of cells (purple) to survive vemurafenib treatment by converting them into the previously established primed cell state. To address this question, we measured their expression of 8 genes associated with the primed cell state using RNA FISH performed on both the Carbon Copy treated with DOT1L inhibitor (blue outline) and the Carbon Copy treated with vehicle control (black outline). 6 of these genes (*AXL*, *EGFR*, *NGFR*, *WNT5A*, *ITGA3*, and *COL1A1*) have higher expression in the primed cell state while *SOX10* and *MITF* have lower expression in the primed cell state. At the same time, we measured single cell expression of these same genes in non-primed cells (gray) and primed cells that do not require DOT1L inhibition to become vemurafenib resistant (green). **B.** We found that DOT1L inhibition modestly increased expression of several genes elevated in the established primed cell state (*AXL*, *NGFR*, *COL1A1*) and decreased expression of *SOX10* and *MITF* specifically in non-primed cells that do not ultimately survive vemurafenib treatment (top row, gray). In contrast, for cells that do ultimately survive vemurafenib treatment (purple and green), DOT1L inhibition appeared to decrease expression of positive markers of the established primed cell state and increase expression of *SOX10* and *MITF*. These transcriptional changes away from the established primed cell state may suggest that compared with vemurafenib treatment alone, cells pre-treated with a DOT1L inhibitor can become vemurafenib resistant via an alternate path. As our RNA sequencing data on these same subpopulations revealed few genes (< 30) induced by DOT1L inhibition, there may not be measurable transcriptional markers for this alternate path.