2016

Adventures With Rna Fish For Diagnosing Viral Infections And Exploring Single-Cell Heterogeneity In Cancer

Sydney Shaffe Shaffer
University of Pennsylvania, sydshaffer@gmail.com

Adventures With Rna Fish For Diagnosing Viral Infections And Exploring Single-Cell Heterogeneity In Cancer

Abstract

Single-molecule RNA FISH is a robust method for visualizing individual molecules of RNA within intact cells that has been used extensively for describing single-cell hetero- geneity in gene expression. In this thesis, we leverage RNA FISH-based technologies for two major problems in biology and medicine: rapid detection of viral infections and understanding mechanisms of resistance to targeted therapy in cancer. Until recently, RNA FISH was not a viable technology for rapid diagnostics, as the hybridization process required a minimum of 6 hours. We start by presenting a modification to the RNA FISH protocol developed by Raj et al. 2008, that enables hybridization in only 5 minutes, and then use these improvements in hybridization time to develop RNA FISH for detection of respiratory viruses. We demonstrate that RNA FISH is capable of de- tecting influenza, rhinovirus, and adenovirus, and propose two probe design strategies with clinical value for discriminating viral strains and detecting many strains at once. Ultimately, we extend these techniques to discriminate single-base differences in the viral sequences, which is clinically useful as single-base mutations can render viruses resistant to our best antiviral medications. In the next section of this thesis, we use RNA FISH for another application: examining single-cell heterogeneity in cancer cells treated with targeted therapy. We first show that melanoma cells can display profound transcriptional variability at the single cell level that predicts which cells will ultimately resist drug. This variability involves infrequent, semi-coordinated transcription of a number of resistance markers at high levels in a very small percentage of cells. The addition of drug then induces an epigenetic reprogramming in these cells, converting the transient transcriptional state to a stably resistant state. This reprogramming is a progressive process consisting of a loss of SOX10-mediated differentiation followed by activation of new signaling pathways, partially mediated by activity of Jun-AP-1 and TEAD. Our work reveals the multistage nature of the acquisition of drug resistance and provides a framework for understanding resistance dynamics. Taken together, these two applications of RNA FISH show its generalizability for exploring many different questions in biology and clinical medicine.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Bioengineering

First Advisor

Arjun Raj

Keywords

Diagnostics, Melanoma, RNA FISH, RNA sequencing

Subject Categories

Biomedical | Cell Biology

ADVENTURES WITH RNA FISH FOR DIAGNOSING VIRAL INFECTIONS AND EXPLORING SINGLE-CELL HETEROGENEITY IN CANCER

Sydney M. Shaffer

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Arjun Raj, Associate Professor of Bioengineering

Graduate Group Chairperson

_____

Jason Burdick, Professor of Bioengineering

Dissertation Committee

Andrew Tsourkas, Professor of Bioengineering

David Issadore, Assistant Professor of Bioengineering

Meenhard Herlyn, Wistar Institute Professor of Dermatology

Ravi Radhakrishnan, Professor of Bioengineering

ADVENTURES WITH RNA FISH FOR DIAGNOSING VIRAL INFECTIONS AND EXPLORING SINGLE-CELL HETEROGENEITY IN CANCER

Acknowledgements

I loved being a PhD student, and it was during this time that I truly fell in love with science. There are many people that shaped this experience and made it wonderful for me.

To my Mom, Dad, Mickey, and Tim, thank you for your unwavering support and love through this long training process. I miss you all everyday that we are not in the same city.

To Jeff and Julien, I feel so lucky to have you both in my life. Thank you both for putting up my late nights working, and for your love and comfort through the challenges of graduate school.

To my hiking partners and best friends, Cyndi and Lisa, our escapes kept me sane. I hope we can be travel companions and friends forever.

To Maggie, Rohan, Stefan, and Eduardo, it has been a blast working with you all in lab. Your contributions to these projects have shaped them into what they are today.

To the members of the Raj lab (past and present), you are a fun bunch of weirdos. Lab was a wonderful environment full of many laughs. Thank you.

And Arjun, I can't imagine a better mentor or better PhD experience. Thank you for everything.

ABSTRACT

ADVENTURES WITH RNA FISH FOR DIAGNOSING VIRAL INFECTIONS AND EXPLORING SINGLE-CELL HETEROGENEITY IN CANCER

Sydeny Shaffer

Arjun Raj

Single-molecule RNA FISH is a robust method for visualizing individual molecules of RNA within intact cells that has been used extensively for describing single-cell heterogeneity in gene expression. In this thesis, we leverage RNA FISH-based technologies for two major problems in biology and medicine: rapid detection of viral infections and understanding mechanisms of resistance to targeted therapy in cancer. Until recently, RNA FISH was not a viable technology for rapid diagnostics, as the hybridization process required a minimum of 6 hours. We start by presenting a modification to the RNA FISH protocol developed by Raj et al. 2008, that enables hybridization in only 5 minutes, and then use these improvements in hybridization time to develop RNA FISH for detection of respiratory viruses. We demonstrate that RNA FISH is capable of detecting influenza, rhinovirus, and adenovirus, and propose two probe design strategies with clinical value for discriminating viral strains and detecting many strains at once. Ultimately, we extend these techniques to discriminate single-base differences in the viral sequences, which is clinically useful as single-base mutations can render viruses resistant to our best antiviral medications. In the next section of this thesis, we use RNA FISH for another application: examining single-cell heterogeneity in cancer cells treated with targeted therapy. We first show that melanoma cells can display profound transcriptional variability at the single cell level that predicts which cells will ultimately resist drug. This variability involves infrequent, semi-coordinated transcription of a number of resistance markers at high levels in a very small percentage of cells. The

addition of drug then induces an epigenetic reprogramming in these cells, converting the transient transcriptional state to a stably resistant state. This reprogramming is a progressive process consisting of a loss of SOX10-mediated differentiation followed by activation of new signaling pathways, partially mediated by activity of Jun-AP-1 and TEAD. Our work reveals the multistage nature of the acquisition of drug resistance and provides a framework for understanding resistance dynamics. Taken together, these two applications of RNA FISH show its generalizability for exploring many different questions in biology and clinical medicine.

Contents

List of Figures

Chapter 1

Introduction

Humans are made up of trillions of individual cells with numerous cell types responsible for performing different functions. Each cell contains essentially the same DNA sequence, but it is through transcription of different RNAs that cells take on their different functions. One method for studying which RNA sequences are present in cells is RNA fluorescence in situ hybridization (FISH), which fluorescently labels each of the RNA transcripts in intact cells. In brief, these methods involve hybridization of fluorescently labeled DNA oligonucleotides to a particular RNA of interest. The sequence of the oligonucleotide is such that it is complementary to its target, and the length and number of such probes can depend on the particular method being used. Our lab uses of clever variation of this technique developed by Raj et al. 2008 in which we use many, usually 15-32, short oligonucleotides, typically 20 base pairs in length, that all bind to different regions of the same RNA species [91]. Any single fluorescent probe will have some off target binding; however, the combined binding of all the different probes at the same site leads to a fluorescent signal that is much brighter than spurious binding to off target sites. In recent years, our lab and others have used RNA FISH for a range of applications including, but not limited to characterizing gene expression in intact tissues, studying the mechanics of transcription, and visualizing chromosome structure [40, 58, 83, 104, 110].

In this thesis, I describe our work applying RNA FISH to two major problems

in biology and medicine: rapid detection of respiratory viruses and understanding mechanisms of resistance to targeted therapy in cancer. First, we developed a variation on the RNA FISH protocol that enables the entire method to be performed in 5 minutes. Second, we applied this new rapid RNA FISH technique for the detection of viruses in intact cells. Then finally, we used RNA FISH in combination with phenomenological cell culture and high-throughput molecular biology techniques to uncover non-genetic contributions to drug resistance in melanoma.

Some of the work presented here has appeared in print in the following publications and has been reprinted with permission:

- S .M. Shaffer and R. P. Joshi et al., *Lab on a Chip*, 2015, 15(15):3170?3182.
- S. M. Shaffer et al., *PLOS One*, 2013, 8(9):e75120.

Chapter 2

Turbo RNA FISH

2.1 Background Turbo RNA FISH

Single molecule RNA FISH offers a number of advantages over other single cell expression quantification tools. In its latest incarnation, it offers the ability to detect individual RNA molecules via fluorescence microscopy, in which each RNA molecule appears in the cell as a bright, diffraction limited spot [35, 91]. Using software to count the spots, one can quantify the absolute number of RNA in individual cells without requiring any amplification, even within the cell's natural developmental context [51, 91]. Moreover, one can analyze spot positions to gain insights into the location of RNA within the cell [49, 71]. Examples include transcriptional dynamics at the site of gene [58, 67], motion at the site of transcription itself [15, 73], and viral RNA localization within the cell [17, 18].

RNA FISH does, however, suffer from some important drawbacks compared to other methods in its current incarnation. One is that it is typically a low-throughput method in the sense that, like RT-qPCR, one can usually only analyze around 5 or so genes at a time, although barcoding schemes can increase this number to many dozens and potentially hundreds [64, 65]. Yet another issue is that most current protocols rely on a long hybridization (often overnight) and series of washes in order to generate adequate and specific signals. The latter limitation hinders the use of RNA FISH in

many scenarios, as it is considerably slower than RT-qPCR in practice, which usually takes on the order of hours to complete. The lack of a rapid version of RNA FISH also places severe restrictions on its use in diagnostic applications, in which timely results are hugely important.

In this thesis and our related publications, we here describe a protocol that enables one to obtain quantifiable single molecule RNA FISH signals in under 5 minutes [106]. We optimized both fixation conditions and hybridization conditions to achieve these results, showing there is a tradeoff between hybridization speed and probe concentration. We showed that these conditions apply across a variety of probes and cell types, and show that the technique is also compatible with our recently developed SNP FISH [57] and iceFISH [58] methods.

2.2 RNA FISH enables single molecule detection

The method we employ for RNA FISH involves the use of several 20-base long single-stranded DNA oligonucleotides, each individually labeled [89, 91] (Fig. 2.1, A). We design these oligonucleotides to bind to different segments of the target RNA via Watson-Crick base pairing, and the combined fluorescence from all the fluorophores at the single RNA leads to a fluorescent spot of intensity much higher than that of the background; we show a representative image for a probe targeting *TBCB* mRNA in Fig. 2.1, B.

2.3 Fixation conditions

Traditionally, we have performed our hybridizations overnight in order to obtain strong signals. In order to perform rapid RNA FISH, we initially reasoned that one could speed the hybridization kinetics by increasing the concentration of probe included

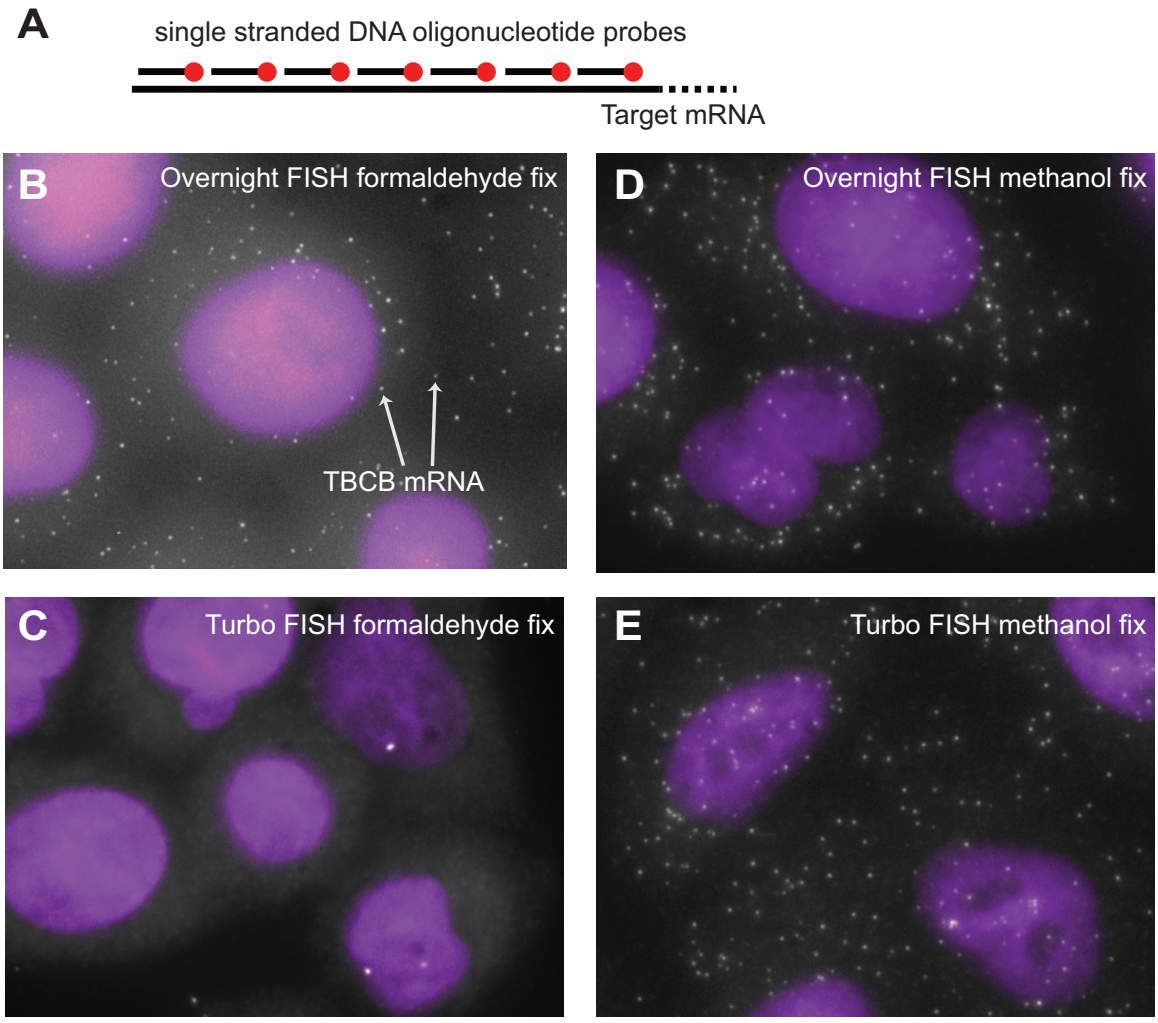

Figure 2.1: Depiction of the RNA FISH scheme and demonstration of rapid hybridization. A. Schematic of the single molecule RNA FISH method, in which we use dozens of short fluorescently labeled oligonucleotides that all target the same RNA molecule. B. Image showing RNA FISH targeting mRNA from the *TBCB* gene under standard overnight hybridization conditions (formaldehyde fixation). Each spot is a single mRNA molecule. C. Image showing RNA FISH signals from an attempt at rapid hybridization (5 minutes) with a high concentration of probe but with formaldehyde fixation. D., E. Traditional overnight hybridization and Turbo RNA FISH hybridization using methanol-fixed cells. All images are maximum projections of a stack of optical sections encompassing the three-dimensional volume of the cell. DAPI (nuclear stain) is in purple.

in the hybridization. Thus, we initially attempted to speed hybridization by simply increasing the amount of probe in our hybridization solution. We found, however, that despite increasing the concentration 20 fold, the signals were greatly diminished at hybridization times of 5 minutes (Fig. 2.2, B, C). Our normal protocol utilizes cells that are fixed with formaldehyde, and we wondered whether the cross-links created by this form of fixation could impede the ability of the oligonucleotide probes to find their targets. To investigate this possibility, we performed hybridization with both ethanol- and methanol- fixed cells (each performed at -20°C), both of which do not generate cross-links. We found that both alcohol-based fixatives performed considerably better (Fig. 2.2, D, E), generating images that were roughly equivalent to those obtained by overnight hybridization with standard conditions. (We note also that we reduced the washing time in these cases to three one-minute washes, for a total of 8 minutes.)

We then quantified the number of mRNA detected in all conditions using software similar in principle to that we have applied previously [91]. We found that after performing overnight hybridization, we obtained roughly the same number of RNA per cell with all fixation methods, but for rapid hybridizations performed at high concentrations, both alcohol-based fixatives gave similar results to those obtained from the overnight hybridizations, whereas the formaldehyde fixed cells performed much more poorly (Fig. 2.2, A, B). We note, however, that the ethanol-fixed cells tended to disintegrate after spending over 48 hours in ethanol solution, so we used methanol-fixed cells for the rapid hybridization experiments in the remainder of the paper.

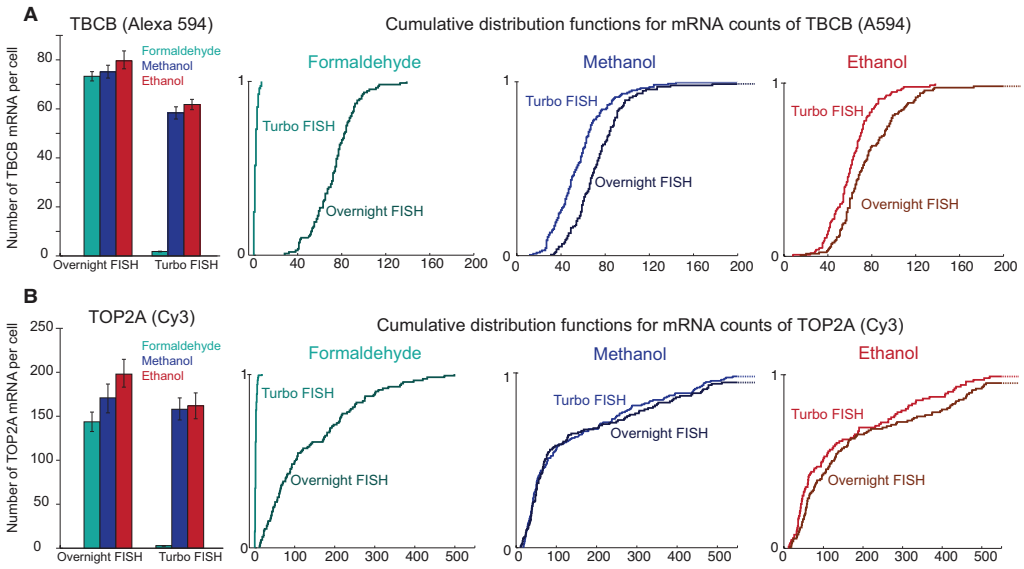

Figure 2.2: Comparison of fixation conditions for both traditional overnight hybridizations and rapid hybridization. A. Comparison of number of spots detected and cumulative distribution functions for the *TBCB* gene with probes labeled with the Alexa 594 fluorophore. Error bars represent the standard error of the mean. No statistically significant differences exist between the overnight RNA FISH samples. Turbo RNA FISH for *TBCB* gene on formaldehyde-fixed cells is statistically different from Turbo RNA FISH on methanol- and ethanol-fixed cells ($p=3.82 \times 10^{-65}$ and $p=4.89 \times 10^{-96}$, respectively; two-tailed t-test). For all conditions, we analyzed spot counts on 100-150 cells. B. Comparison of number of spots detected and cumulative distribution functions for the *TOP2A* gene with probes labeled with the Cy3 fluorophore. Error bars represent the standard error of the mean. Overnight RNA FISH for *TOP2A* gene on formaldehyde-fixed cells is statistically different from overnight RNA FISH on ethanol-fixed cells ($p=0.0067$; two tailed t-test). No other statistically significant differences exist between overnight RNA FISH samples. Turbo RNA FISH for *TOP2A* gene on formaldehyde-fixed cells is statistically different from Turbo RNA FISH on methanol- and ethanol-fixed cells ($p=9.57 \times 10^{-28}$ and $p=4.22 \times 10^{-30}$, respectively; two-tailed t-test). For all conditions, we analyzed spot counts on 100-150 cells. Data shown represents one of two replicate experiments.

2.4 Relationship between concentration and hybridization time

We then explored the degree to which there is a tradeoff between increasing the concentration of the probe and the hybridization time for rapid hybridization in methanol-fixed cells. In order to do so, we needed a means to assess and compare the quality of the signal in these various conditions. Ultimately, we settled on a metric based on the sensitivity of the threshold between signal and background (Fig. 2.2, A). Briefly, we first use a linear filter designed to enhance spot-like signals. We then found all candidate spots by locating all regional maxima. These candidate spots consist of two populations, one corresponding to background spots and one corresponding to the target RNA molecules. When the signals are clear and quantifiable, the intensities of the RNA spots should be nicely separated from those of the background spots (Fig. 2.3, A). However, if the RNA spots are not of high quality, then the spot intensities of the two populations can blend together, making it difficult to accurately quantify the number of true RNA spots within the image (Fig. 2.3, A). To quantify this difference, we measured the degree of separation in the intensities of the two subpopulations by essentially measuring the sensitivity of the threshold separating the two; i.e., once the threshold is set, if we move the threshold slightly higher or lower, we measured the relative change in the number of RNA detected (Fig. 2.3, A). We found that this metric for quantification captured the qualitative visual differences between conditions. We further note that metrics such as spot intensity and average spot count can be somewhat misleading as metrics of the ability to accurately count RNA in single cells (Fig. 2.3, C). For instance, we have found that RNA spot intensity in and of itself need not be particularly high for accurate spot counting; rather, it just needs to be clearly and uniformly higher than the intensity of background spots. Average spot

counts are also problematic because even when thresholds are ill-defined (as in Fig. 2.3, A, right), one could still choose thresholds that yield similar spot counts on average, even though another person might equally well choose a different threshold, giving completely different results. For these reasons, we primarily focused on the sensitivity metric as an objective metric of signal quality.

We here present data from A549 cells, a common cancer cell type that we have found overall to be more difficult to perform rapid hybridizations in (hence providing a stringent test of our method). We performed RNA FISH (targeting *TOP2A* mRNA) over a range of hybridization times from 30 seconds to 10 minutes and probe concentrations ranging from our conventional probe concentration to 100 fold greater (approximately 4.4 μM to 400 μM). Throughout, we compared also to our traditional overnight hybridization protocol (Fig. 2.3, B, C). We found that we were able to obtain readily quantifiable signals after 5 minutes of hybridization in the A549 cells (Fig. 2.3, B, C). We found that there is a clear tradeoff in that higher concentrations of probe in the hybridization solution allow for shorter hybridization times. The exact amount of time and concentration one should use in these cases will of course depend on the constraints of the problem at hand, but we believe that a 5 minute hybridization at a probe concentration of 500 μM would be practical in many real-world scenarios. We also note that it is in some cases possible to perform rapid hybridizations in as little as 30 seconds with high concentrations of probe.

For comparison, we also performed the same analysis by using the concentrations and wash protocol we used for our conventional overnight RNA FISH, except performing the hybridization for various amounts of time. We found that we obtained poorly quantifiable signals (as indicated by the sensitivity metric) once the hybridization time went below 2 hours, which is 24-fold as much time as our rapid hybridization assay (Fig. 2.4, A, B).

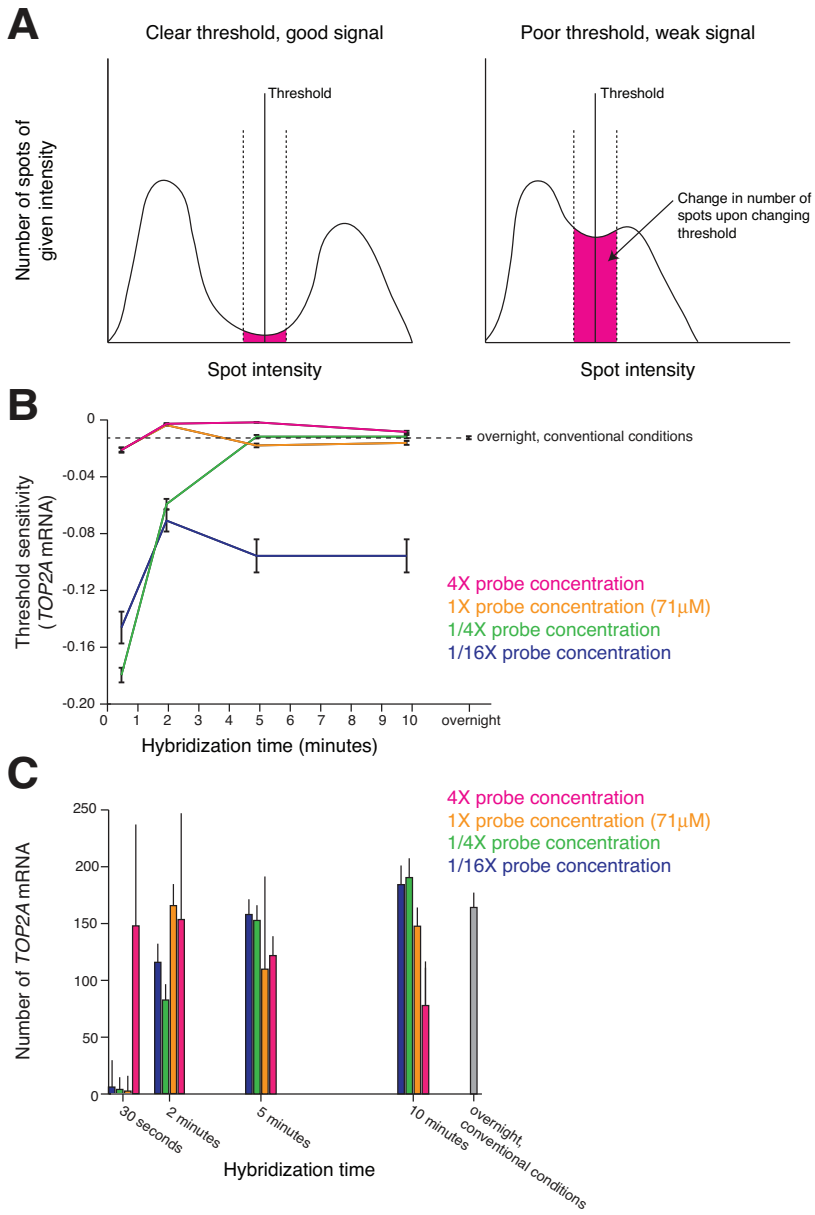

Figure 2.3: Quantification of signal quality and comparison of different hybridization times and probe concentrations. A. Schematic depicting the manner in which we quantify signal quality via threshold sensitivity. B. Sensitivity of threshold measured in varying probe concentrations and hybridization times. The dotted line represents the sensitivity of a traditional overnight RNA FISH. Error bars reflect standard error of the mean. C. Spot counts for the same conditions as in B. Error bars reflect standard deviation. At 10 minutes and for all probe concentrations, the spot counts for Turbo FISH are statistically different from overnight FISH (4X: $p=9.87\text{x}10^{-6}$, 1X: $p=0.0136$, 1/4X: $p=4.86\text{x}10^{-6}$, 1/16X: $p=1.75\text{x}10^{-11}$; two-tailed t-test). For all conditions, we analyzed spot counts and calculated the sensitivity on 80-120 cells. Data shown represents one of two replicate experiments.

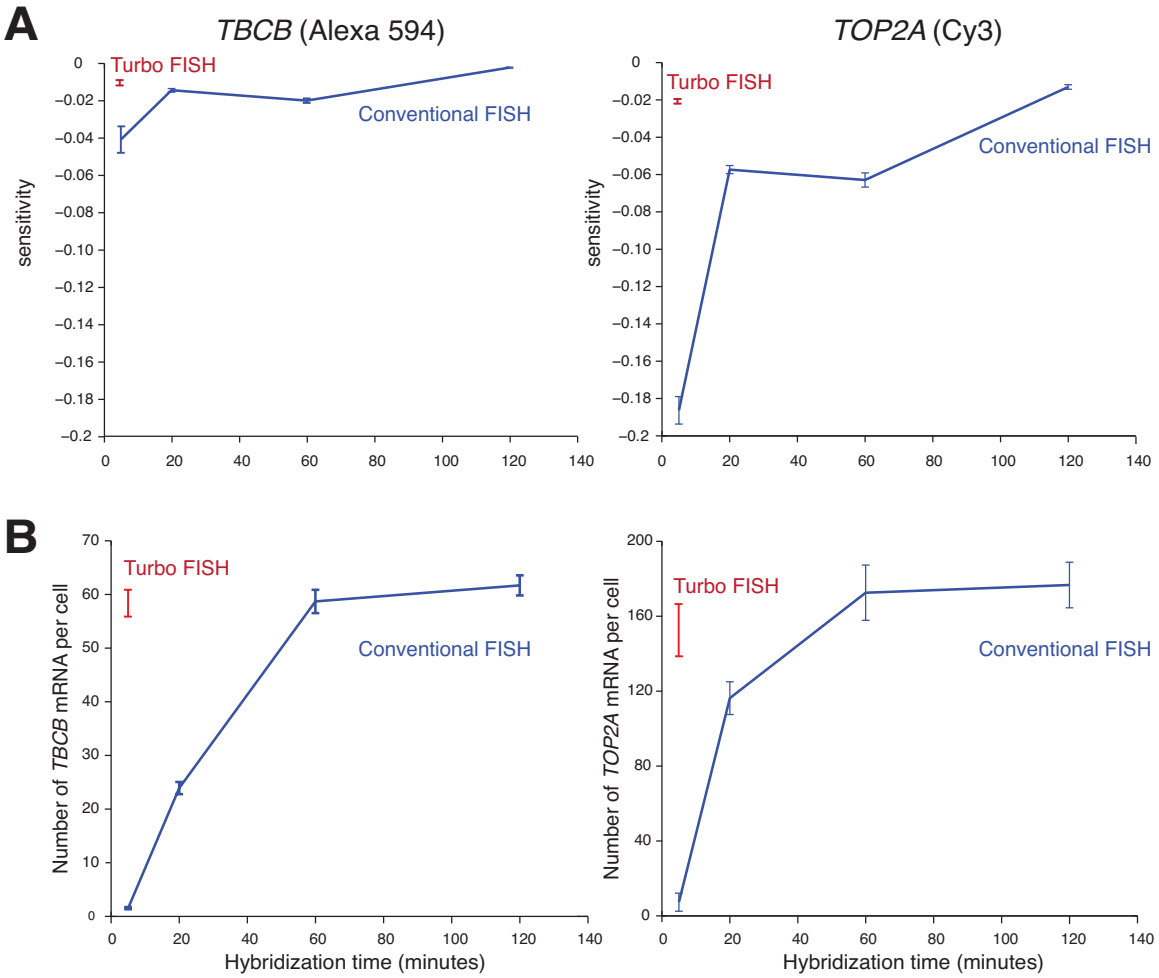

Figure 2.4: Comparison of signal from Turbo RNA FISH (5 minutes; red) to conventional RNA FISH (blue). A. Comparison of RNA FISH signal sensitivity at a range of hybridization times. Error bars reflect standard error of the mean. At 5 minutes, we found a statistically significant difference in signal sensitivity between Turbo FISH and conventional FISH for *TBCB* gene and *TOP2A* gene ($p=4.75 \times 10^{-11}$ and $p=1.19 \times 10^{-74}$, respectively; two-tailed t-test). B. Comparison of RNA FISH spot count at a variety of hybridization times. Error bars reflect standard deviation. At 5 minutes, we found a statistically significant difference in RNA FISH spot count between the Turbo FISH and conventional FISH for *TBCB* gene and *TOP2A* gene ($p=1.69 \times 10^{-68}$ and $p=2.07 \times 10^{-20}$, respectively; two-tailed t-test). For all conditions, we analyzed spot counts and calculated sensitivity on 100-150 cells. Data shown represents one of two replicate experiments.

2.5 iceFISH and SNP FISH

In our lab, we have recently developed two variants of single molecule RNA FISH: 1. a method based on targeting introns that reveals chromosome structure and transcriptional activity (intron chromosomal expression FISH or iceFISH [58]), and 2. a method that utilizes both a new probe design and spot colocalization analysis to enable us to detect single nucleotide differences on individual transcripts (SNP FISH [57]). We wanted to test whether these methods would work in the rapid hybridization format. For iceFISH, we constructed an intron-based chromosomal "paint" that targets chromosome 19. We found that the iceFISH signals were comparable to those obtained via conventional overnight FISH using our rapid hybridization conditions (Fig. 2.5, 5).

For SNP FISH, we used an approach in which we use a single oligonucleotide "SNP detection" probe hybridized to a "mask" oligonucleotide that leaves just a short "toehold" region available to nucleate binding to the target RNA (Fig. 2.6,). The toehold region is short enough (5-10 bases) that it provides discrimination of single-base mismatches, but upon the binding of the correct probe, the mask dissociates via strand displacement [131], leading to the formation of a long (20-30 base) hybrid that provides stability. Meanwhile, we labeled the rest of the target RNA using conventional RNA FISH probes (which we call "guide probes") that tell us where the target RNA are within the cell. Using colocalization between the SNP detection probe and the guide probe, we could assign each RNA based on whether or not it has the SNP. Our previous work demonstrated that this approach works under conventional RNA FISH conditions [57]. To check whether we were able to perform SNP FISH in rapid hybridization conditions, we used higher concentration of probes and shortened hybridization times (5 minutes) in methanol fixed cells. We tested Turbo SNP FISH in WM983b cells (gift of Meenhard Herlyn, Wistar Institute), which are heterozygous for the V600E mutation

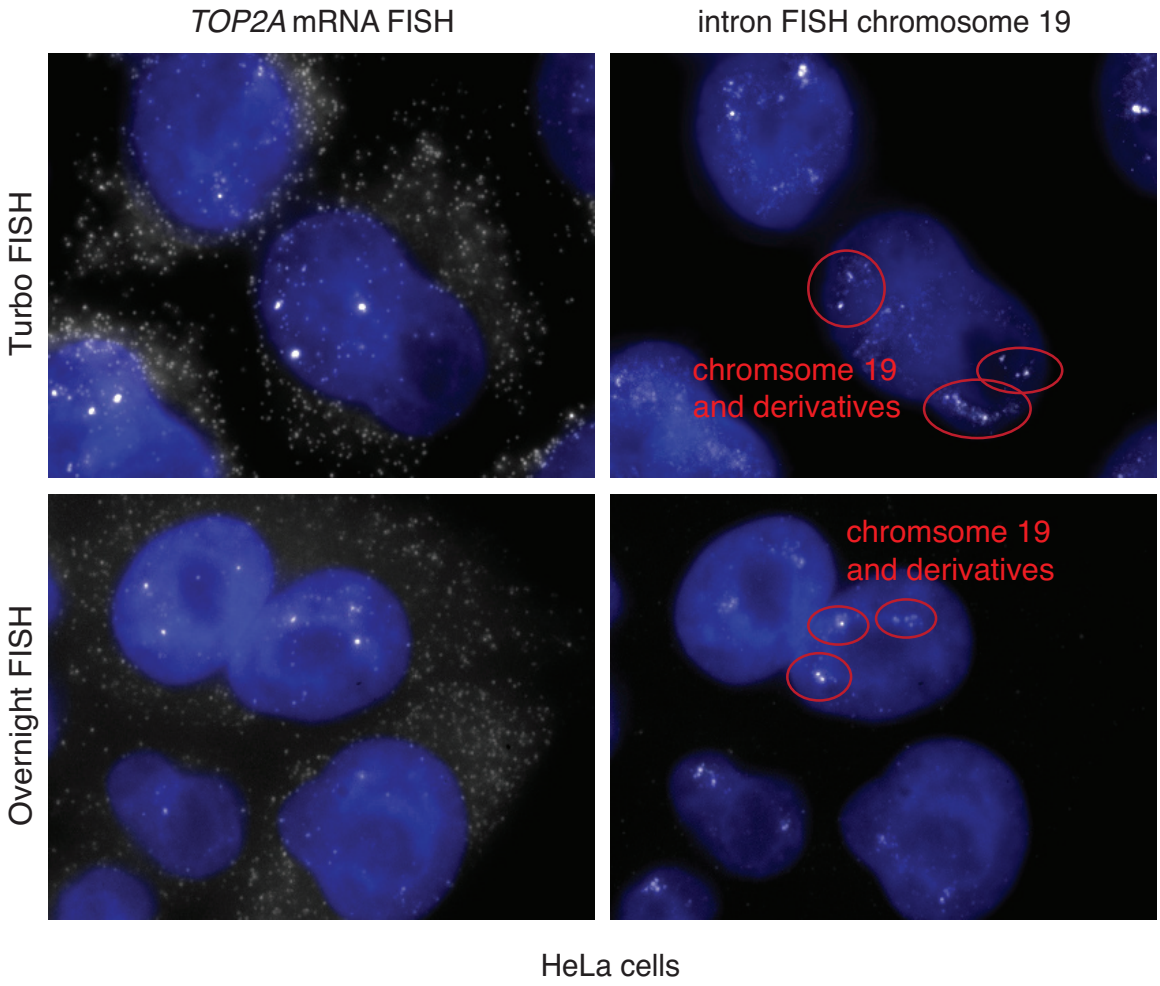

Figure 2.5: Demonstration of Turbo iceFISH. We performed Turbo FISH using iceFISH probes that targeted a total of 20 introns in genes on chromosome 19 (right panels), while simultaneously performing RNA FISH for *TOP2A* mRNA (left panels). We compared both Turbo FISH to conventional RNA FISH performed overnight (top vs. bottom panels). All images are maximum projections of a stack of optical sections encompassing the three-dimensional volume of the cell. DAPI (nuclear stain) is in blue.

in the *BRAF* gene. We used both probes targeting the V600E BRAF mutation or a region common to both alleles on the *BRAF* mRNA as a control for non-specific binding (Fig. 6A). We found that in both Turbo SNP FISH and conventional overnight SNP FISH, the probes targeting the heterozygous base in *BRAF* indeed showed roughly equivalent levels of both mutant and wild-type mRNA (Fig. 2.6, A, top). The probes targeting the region common between the two alleles identified virtually all the mRNA as being wild-type in both turbo and conventional conditions, showing that the rate of cross-hybridization remained low even with rapid hybridization conditions (Fig. 2.6, A, bottom). Quantitatively, the results from both turbo and conventional SNP FISH were similar, both to each other and to our previous results [57] (Fig. 2.6, B).

2.6 Discussion Turbo RNA FISH

In this paper, we have described a protocol that enables rapid and quantitative detection of RNA targets via RNA FISH. We found that alcohol-based fixatives provide the necessary probe accessibility for rapid hybridization via increased probe concentration, potentially enabling hybridizations in as little as 30 seconds.

Our experiments show that there is a straightforward tradeoff between concentration of probe and the speed of hybridization. We have found that increasing probe concentration by 20X compared to our normal overnight protocol yields reliable RNA FISH results after just 5 minutes of hybridization. At first glance, this increased probe concentration may not seem economically viable, considering the increased use of probes (which are the most costly reagent in the RNA FISH protocol). However, we note that because of the decreased time for drying, our protocol uses roughly 10 fold less hybridization solution for the hybridization itself, greatly mitigating such concerns.

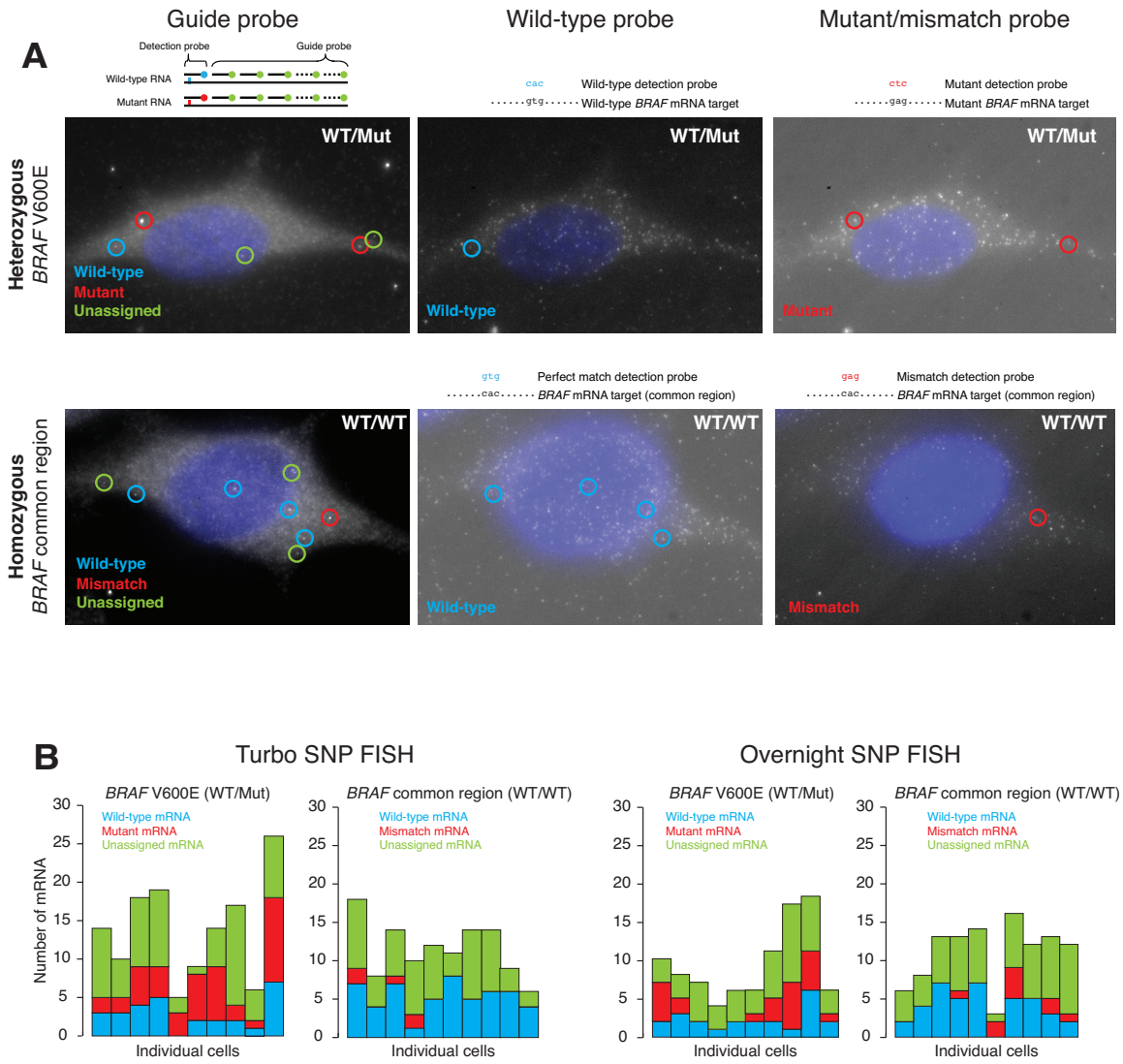

Figure 2.6: Demonstration of Turbo SNP FISH. A. Demonstration of SNP FISH efficacy under Turbo FISH and conventional RNA FISH conditions in WM983b cells. We targeted *BRAF* mRNA with guide probes, and then used detection probes that targeted either the V600E mutation for which *BRAF* is heterozygous in this cell line (top panels) or a common region for which *BRAF* is homozygous in this cell line (bottom panels). Left panels show the signals from the guide probe (that labels the mRNA), the middle panel shows the detection probe that detects the wild-type sequence, and the right panel shows the detection probe that detects the mutant sequence. B. Quantification of RNA as being either mutant or wild type in this cell line. Each bar corresponds to data from a single cell.

We believe that the ultimate choice of how much probe to use and how fast to drive the reaction will depend on the specifics of the application at hand. In some cases, getting a hybridization time of 5-10 minutes may be perfectly fine, in which case one may not need to use large concentrations of probe. However, in some situations, such as during a surgical procedure, the decreased hybridization times may be a benefit that outweighs the cost of increased probe usage.

Of course, even with rapid hybridizations, we have not addressed the issue of the imaging time itself. Typically, image acquisition may require taking image stacks from multiple positions on the slide to obtain enough cells worth of image data to make statistically significant claims about differences in gene expression. Currently, doing so could take on the order of 10-20 minutes per condition. However, we believe that technical advances can reduce the time required for both image acquisition and analysis by at least an order of magnitude. In such a case, one could envision comparing gene expression in two samples in well under 30 minutes from living cells to quantified data.

It may be useful here to make a comparison to other methods such as RT-qPCR. RT-qPCR is the current gold standard for gene expression analysis, widely considered to be the most accurate method for quantifying gene expression to date. It has many benefits, including high dynamic range, low cost per reaction, and the ability to parallelize in 96-well plate format. The qPCR itself usually takes on the order of 1-2 hours to complete, but if one includes both RNA extraction and setup time, the total time required is probably closer to around 3-4 hours. (These extra steps also increase the cost of the experiment as well.) We believe that with rapid hybridization, RNA FISH competes favorably with RT-qPCR on most counts. With respect to quantification, our method provides accurate, absolute counts of gene expression of 3 to 5 genes in individual cells without the explicit need for normalization. Since RNA FISH is a direct detection scheme without any amplification, we are able to detect

even small fold-changes with high precision, differences that would be hard to measure accurately with RT-qPCR, at least not without a large number of replicates. The cost per reaction is probably dominated by the cost of the probe, which is currently around $300-$600 per probe set for 10,000 hybridizations ($0.06 per reaction) and is thus comparable to a molecular beacon or Taq-man RT-qPCR probe. Of course, costs of labor, equipment and other reagents are variables that are hard to predict, but will be of the same order of magnitude, although we note that the labor required for RNA FISH is probably lower, whereas the cost of an automated microscope is admittedly higher than most qPCR machines. Both the accuracy and cost comparisons to RT-qPCR were valid even with overnight RNA FISH.

The time required for previous iterations of RNA FISH, however, was considerably longer than for RT-qPCR, and our new method alleviates that discrepancy. If one is just comparing the expression of a few genes in a few conditions, then we believe our method is unequivocally several times faster than RT-qPCR, especially when one factors in RNA extraction and setup time. For analyzing larger numbers of genes in parallel, though, the imaging time will become a factor. If one assumes 5-10 minutes per condition and triplex RNA detection, then analyzing, say, 20-30 genes could require up to 2 hours. With advances in high throughput imaging, we anticipate that one could reduce this time by an order of magnitude, thus further increasing the speed advantages.

Another major advantage of RNA FISH is that it also provides single cell information, something that is much more difficult to obtain with single cell RT-qPCR approaches. This enables one to measure variability in gene expression from cell to cell. Since the measurements yield absolute numbers of RNA, the measurements do not necessarily require normalization to an internal control (such as *GAPDH*), although one could perform such an analysis if one wished through multiplexing. Normalization

can be difficult to perform with RT-qPCR approaches, since one typically uses all the material for a single qPCR reaction, leaving none for further normalization.

Furthermore, RNA FISH also provides spatial information on the localization of RNA. Such information is important both for examining differences from cell to cell within a tissue and even subcellular spatial localization. In tissues, one can easily identify particular cells by labeled RNA specific to those cells with one color and then looking at the gene of interest in another color. Subcellular information can be of particular importance for RNA that localize to particular regions of the cell, such as many non-coding RNA, in which case RNA FISH can reveal much about its behavior.

We have also shown that one can perform iceFISH and SNP FISH to visualize chromosomes and single base changes, respectively, with rapid hybridization. Such techniques could be useful for rapidly diagnosing chromosomal abnormalities and for rapid genotyping of particular single nucleotide variants.

In summary, our method for rapid hybridization results in orders of magnitude improvements in hybridization time for single molecule RNA FISH, enabling a new set of high throughput and rapid diagnostic applications.

Chapter 3

Developing an RNA FISH-based influenza diagnostic

3.1 Background: Need for rapid and accurate viral diagnostics

Viral infections are the cause of a wide range of clinical diseases. While some viral infections can be diagnosed from signs and symptoms alone, for many viral infections, the clinical signs and symptoms overlap with other diseases and infectious agents [63, 69, 116]. For these reasons, clinicians need laboratory tools to diagnose viral infections. However, currently available viral diagnostics have significant limitations in that they can be very slow, expensive, and are typically not performed at the point-of-care. Overcoming these challenges would enable faster treatment of viral infections, prevent unnecessary doctors office visits, save money, and facilitate large-scale viral surveillance. Here we aim to establish RNA FISH as a methodology for faster, cheaper, and point-of-care viral diagnostic assays.

Most current diagnostic tests target either viral proteins, using immunofluorescence or enzyme-linked immuno assay (ELISA), or viral nucleic acids, using RT-PCR. The vast majority of protein-based diagnostics use antibodies, which require long development

times at high costs. In contrast, nucleic acid detection is highly sensitive and highly specific [116, 130], and allows easier development of new assays as targets evolve. Detecting nucleic acids by RT-PCR, however, requires 1-2 hours and a thermal cycler, which can be a limitation in common clinical practice [41], especially for viruses that are typically treated in a doctor's office or emergency room.

A complementary approach for nucleic acid detection is direct labeling via RNA fluorescence in situ hybridization (RNA FISH). Conventionally, RNA FISH has suffered from three main drawbacks preventing its use as a clinical diagnostic: sensitivity, long assay times (6-12 hours), and many complex steps requiring laboratory training [89]. To overcome sensitivity limitations, we use a variant of RNA FISH develop by Raj et al. that involves hybridization of 20-50 short, fluorescently-labeled oligonucleotide probes to the target RNA [91]. The use of a large number of oligonucleotides amplifies the fluorescence signal to the point where we can readily detect individual molecules of RNA via conventional fluorescence microscopy. This technique has traditionally required 6-12 hours, but recently our lab has overcome this time requirement by developing a rapid hybridization protocol that utilizes alcohol based fixatives and high concentrations of oligonucleotide probe sets [106], see Chapter 2. These improvements have reduced the assay time by orders of magnitude such that it can be performed in under five minutes. This rapid assay time indicates great potential for applications in point of care diagnostics, especially for viruses, which generate large numbers of viral RNA. However, open questions remain as to how well RNA FISH can discriminate clinically relevant viruses and whether the assay itself can be standardized and automated to the point where someone without training could run the assay at the point of care.

In this chapter, we present a complete platform for viral RNA FISH-based rapid diagnostics that includes viral probe design software, microfluidic automation and image processing software (Fig. 3.1). First, we created software to design 20 base

pair DNA oligonucleotides targeting viral RNA. We formulated two different probing strategies: an algorithm to design probe sets that are capable of targeting many input sequences, and an algorithm to design probe sets that differentiate input sequences from each other. Next, the pipeline includes a microfluidic device to standardize the rapid RNA FISH assay and to make it easily parallelizable for interrogating many viral targets. The microfluidic device concentrates cells under a filter, thereby immobilizing the sample for hybridization of the RNA FISH probes and subsequent washes. Finally, we image the RNA FISH-labeled cells on chip and our image processing software classifies the sample as infected or uninfected.

We selected respiratory viruses as a test bed for developing rapid RNA FISH-based diagnostics because clinicians often treat upper respiratory infections (URI) in outpatient or emergency settings where time is limited and most current tests with rapid turnaround time have significant compromises in analytic sensitivity and/or specificity. As URI symptoms are generally non-specific, molecular diagnostics are especially needed for respiratory viruses because the treatments depend upon what virus (or even bacteria) is causing the symptoms [31]. For example, influenza is effectively treated with neuraminidase inhibitors, oseltamivir and zanamivir, while bacterial infections are treated with antibiotics [28, 119]. In addition to providing relief to the patient with treatment, viral diagnostics help reduce inappropriate use of antibiotics, which is important as it can lead to the development of resistant bacteria. We believe RNA FISH would be particularly useful for respiratory viruses because it could provide the molecular specificity to diagnose these infections.

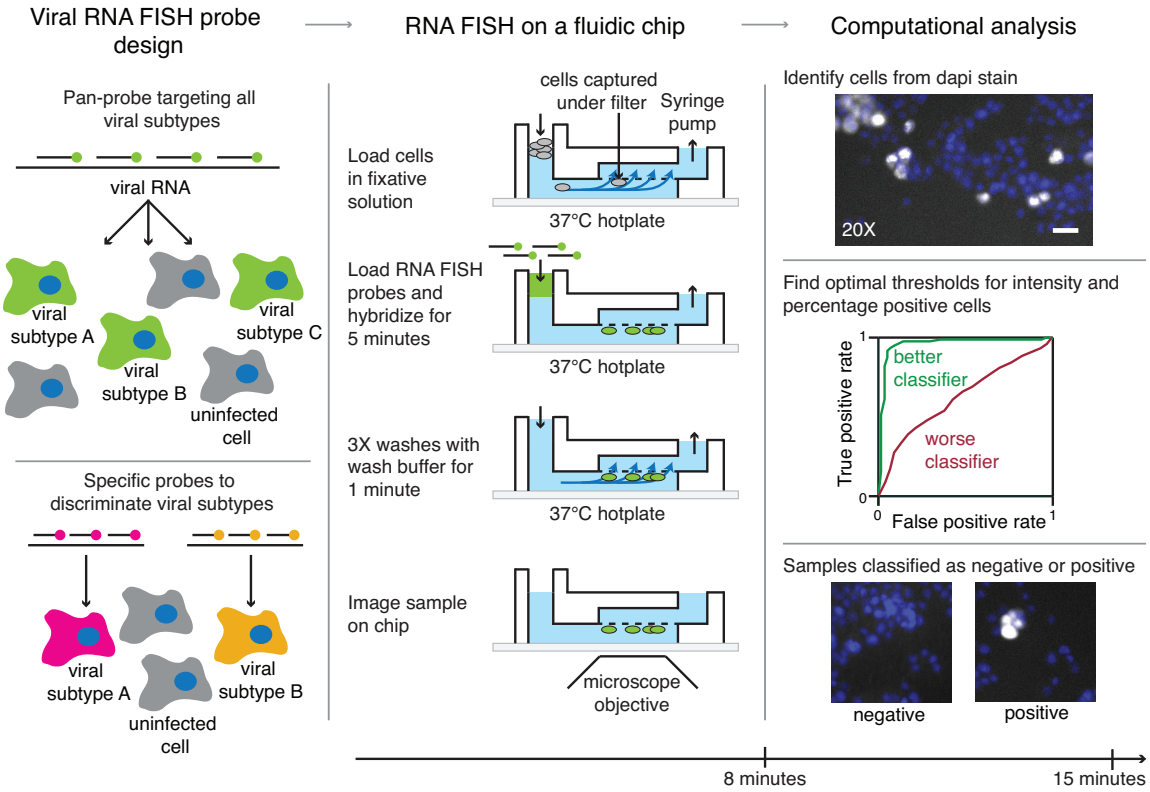

Figure 3.1: RNA FISH platform rapidly determines whether samples are uninfected or infected with a virus. The pipeline includes probe design software to generate subtype specific probe sets or probe sets that label many viral subtypes. Next, we perform RNA FISH on the microfluidic chip which consists of a 5 minute hybridization and 3 one minute wash steps. Finally, we image the sample directly on the chip and then our image processing pipeline determines whether a sample is uninfected or infected.

3.2 Probe design algorithms for discriminating viruses

A robust platform needs to address two dichotomous challenges: the need to differentiate between viral strains that have different clinical characteristics but similar sequences, and the need to simultaneously detect viral strains that have highly diverse sequences but similar clinical characteristics. We created probe design software to address both of these possibilities.

To detect different viral strains, we developed a bioinformatic algorithm for designing RNA FISH probes that are specific for each viral strain while exhibiting minimal cross-targeting to other strains. The user provides the total number of DNA oligonucleotides desired in the probe set and the maximum number of bases for which the software will tolerate cross-targeting to another strain (here, we used 14/20). The software then evaluates every possible 20 base-pair DNA oligonucleotide that can bind to the viral RNA strain of interest. For each oligonucleotide, the software finds the maximum sequence match in the other viral strains, rejecting all oligonucleotides that show a higher degree of sequence complementarity than the user-defined limit. This ensures that none of the oligonucleotides in the strain-specific probe set will cross-hybridize to the other strains (Fig. 3.2, A).

We tested this probe design approach using influenza because multiple influenza strains circulate in the human population at once and discriminating between these strains is clinically useful for treatment and public health purposes. We designed strain-specific influenza RNA FISH probe sets focusing on the strains predicted to be circulating during the 2014-2015 influenza season, specifically A/California/07/2009 H1N1, A/Texas/50/2012 H3N2, and B/Brisbane/60/2008. Influenza is a RNA virus comprised of 8 RNA "segments", each of which gave high levels of signal by RNA

FISH (Fig. 3.3). We targeted the hemagglutinin and nucleoprotein segments with RNA FISH probes for these studies. To test the specificity of our final designs, we infected Madin-Darby canine kidney (MDCK) cells with either H1N1, H3N2, or influenza B. Twenty-four hours after infection, we fixed these cells and performed a five minute hybridization using the RNA FISH probe sets targeting all three viruses. We found that these probe sets specifically bound to cells infected with their corresponding viral strain (Fig. 3.2, B). Furthermore, we found that these strain-specific probes had minimal cross hybridization to the other influenza strains.

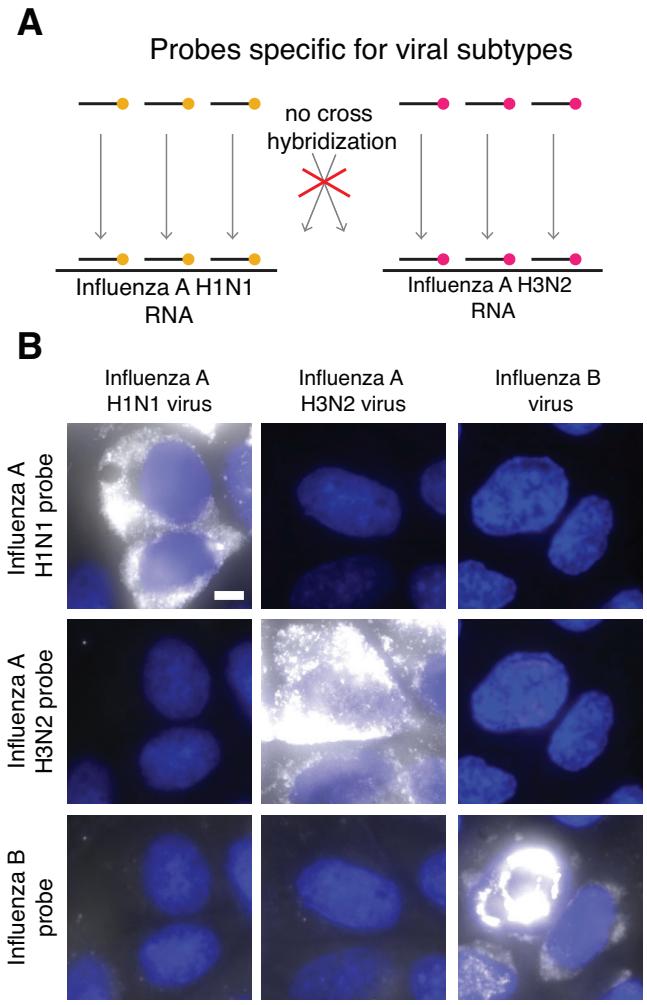

Figure 3.2: Subtype-specific RNA FISH probes discriminate influenza subtypes. A. Overview of our subtype-specific probing strategy, in which we design oligonucleotides that only bind to one influenza subtype and do not cross hybridize to other subtypes. B. We used this software to design probes to target three different influenza subtypes. We then infected MDCK cells with these subtypes and performed RNA FISH with subtype-specific probes. These three strains of influenza that were close enough in sequence similarity to our designs that we would expect our subtype-specific probes to bind. The RNA FISH probes produce bright fluorescent signal in the subtype to which they are designed and do not produce signal in the other subtypes. DAPI stain labels nuclei in blue, and RNA FISH is in white. White scale bar represents 5 microns.

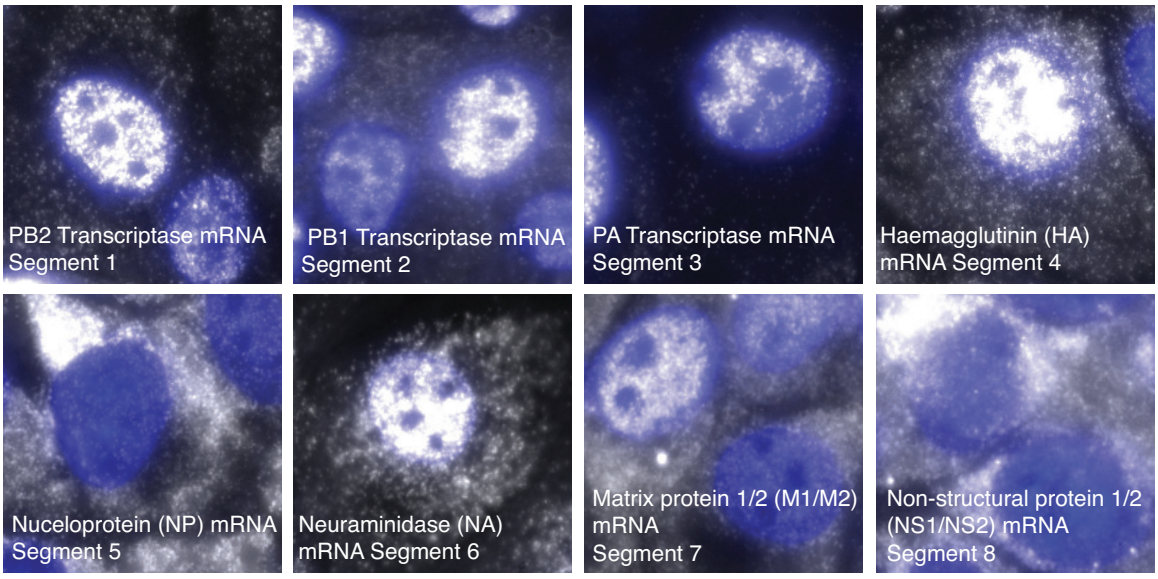

Figure 3.3: RNA FISH for each segment of influenza virus produces bright fluorescent signal. We designed RNA FISH probes to target the viral mRNA from all 8 segments of influenza A. We then infected MDCK cells with influenza A/Puerto Rico/8/1934 H1N1, fixed the cells, and performed rapid RNA FISH for the viral mRNA segments. We found that the RNA from each segment produced bright fluorescent signal. DAPI (nuclear stain) is in blue, and RNA FISH is in white. All images are 100X magnification.

3.3 Probe design algorithms for targeting many viruses at once

Next, we wanted to develop an assay capable of simultaneously labeling all strains of the same virus despite extensive sequence divergence, which requires targeting each sequence with multiple oligonucleotides. Rhinovirus, for example, has such tremendous sequence diversity that designing a separate pool of 10 oligonucleotides for each of the 348 sequences available on NCBI would require a total of 3480 DNA oligonucleotides, which is prohibitive both in cost and assay complexity. Instead, we sought to create a "pan-probe" set of greatly reduced complexity by targeting oligonucleotides to subsequences that exist in multiple strains. Given a sequence alignment generated by Clustal Omega [109], our algorithm seeks to minimize the number of total oligonucleotides in the pool while ensuring that every strain is targeted by at least 10 oligonucleotides. We tested this software on 348 different rhinovirus sequences (including sequences from all three subtypes A, B, and C) available on NCBI, yielding a total of 417 oligonucleotides (Fig. 3.4, B). The maximum number of oligonucleotides targeting any one strain was 31, the minimum was our specified limit of 10, and the mean number of oligonucleotides per strain was 19. The mean number of strains bound by an individual oligonucleotide was 15.7. We also found one oligonucleotide targeting a region in the conserved 5'UTR that binds to 334 of 348 rhinovirus strains.

To test the efficacy of our rhinovirus pan-probe, we infected HeLa cells with one of four different strains of rhinovirus (rhinovirus A1, rhinovirus A16, rhinovirus A81 and rhinovirus B1) and performed RNA FISH on each of the four infected cell lines. With a 5 minute hybridization, we found that the pan-probe was capable of identifying virus in each of the different infections and did not bind to uninfected cells (Fig. 3.4, C). This experiment suggests that we can design a broadly reactive probe set to target all

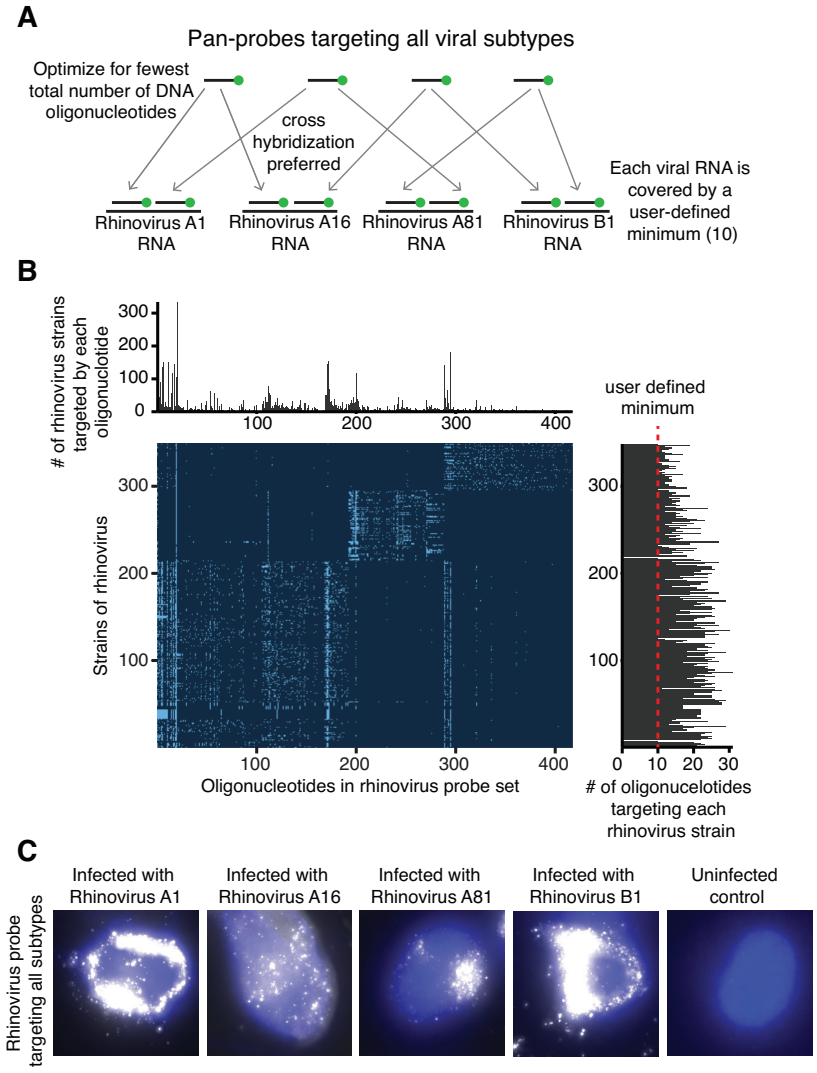

Figure 3.4: Rhinovirus pan-probe targets all rhinovirus strains. A. Overview of pan-probe design, in which we optimize for the minimum number of oligonucleotides that will bind to all sequences. B. We used our pan-probe software to design RNA FISH probes that target 348 rhinovirus sequences with a minimum of 10 oligonucleotides per strain. The results of this design are summarized by the heatmap where each box represents a pair between an RNA FISH oligonucleotide and a viral strain. Light blue boxes indicate that the oligonucleotide is a perfect match somewhere in the virus and thus should bind, while dark blue boxes represent that the oligonucleotide does not bind to that strain. C. We infected HeLa cells with different strains of rhinovirus and performed rapid RNA FISH with the pan-probe for rhinovirus. Cells infected with each strain had bright fluorescent signal by RNA FISH and the uninfected cells remained dark. DAPI (nuclear stain) is in blue, and RNA FISH is in white. White scale bar represents 5 microns.

strains of rhinovirus without off-target binding in uninfected cells. We also used this design software to create probe sets for adenovirus viral mRNA and found that these probe sets gave bright fluorescent signal in infected cells (Fig. 3.5).

3.4 Design and fabrication of a microfluidic device to concentrate samples, automate rapid RNA FISH, and facilitate imaging

To standardize RNA FISH and facilitate its use, we built a microfluidic device capable of automatically concentrating a dilute sample of cells and performing RNA FISH. The microfluidic chip consists of a track-etched polycarbonate micropore filter positioned between two micromachined sheets of adhesive mylar (Fig. 3.6). Cells in suspension enter the device channel (width 2 mm and length 4 mm) and are concentrated by the filter, which also serves as an imaging area for microscopy (area 7.14 mm^2). The filter has a high density of 5 micron pores (10^6 / cm^2), allowing fast flow rates and preventing clogging due to debris. (In particular, if a few of the pores become obstructed, the device continues to operate because fluid can pass through the other pores in the filter.) We anticipate that these features will make the device robust to impurities found in clinical samples such as nasal swabs or sputum [75]. The bottom of the chip consists of a microscopy coverslip (number 1 thickness) allowing us to directly image the cells within the device.

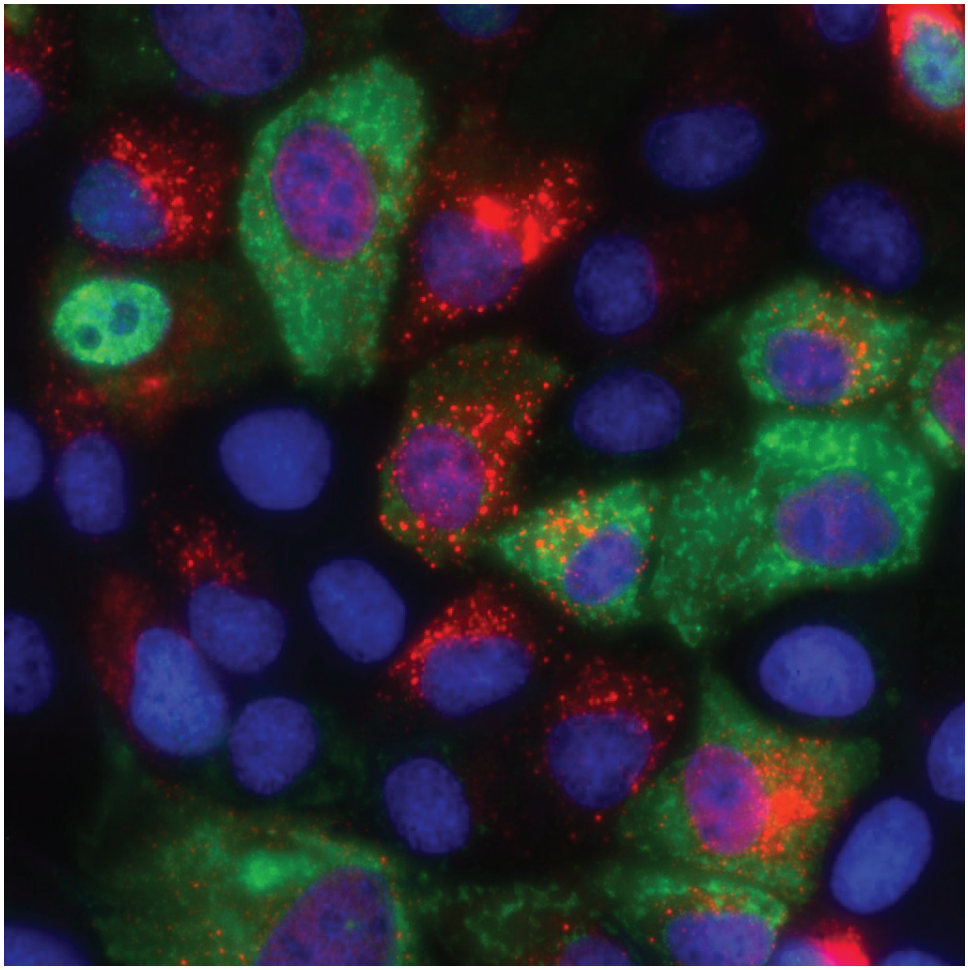

Figure 3.5: RNA FISH for genomic and messenger viral RNA produces bright fluorescent signal. We designed RNA FISH probes to target the viral mRNA and genomic RNA of influenza A. We then infected MDCK cells with influenza A/California/7/2009 H1N1, fixed the cells, and performed rapid RNA FISH for the different RNA. We found that the viral RNA produced bright fluorescent signal in a different spatial distribution from the mRNA. DAPI (nuclear stain) is in blue, genomic viral RNA is in red, and viral mRNA is in green. All images are 100X magnification.

To test our device, we performed RNA FISH on influenza infected MDCK cells and imaged the cells using widefield fluorescence and bright-field microscopy and a 20X objective. We scanned the imaging area of the device to obtain a 5 by 8 grid of images, each of which measured 665 μm by 665 μm. This resulted in images at 40 positions that we then assembled into a large panorama covering 3.3 mm by 5.3 mm of the area on the device where the cells are captured on the filter. At each position in the grid, we acquired one bright-field image and three fluorescence images. The fluorescence images were the RNA FISH probe signal, a dapi nuclear stain, and green fluorescent protein (GFP). During analysis, we used the GFP signal to exclude background autofluorescence in the cells. In some assays we included RNA FISH probe sets for more than one virus and obtained additional images with filters that are designed to specifically detect the probe sets (Fig. 3.10).

3.5 Image processing software and classification of infected and uninfected samples

The next component of our platform is image processing software that aligns and combines the images, finds cells, and classifies them to determine whether samples are positive or negative for virus. Briefly, our image processing module assembles all the individual images into a large panorama and segments the cell nuclei using DAPI, which stains the nucleus (Fig. 3.9, A, steps 1-2). Next, the software calculates the median fluorescence intensity of the RNA FISH signal for each cell (Fig. 3.9, A, step 3). By plotting a histogram of these intensities, we found a peak of intensity values corresponding to negative cells and a peak corresponding to positive cells (Fig. 3.9, A, step 4). We then used an intensity cutoff to designate, at the single cell level, which cells are positive and which cells are negative and thereby calculate the percentage of

infected cells in the sample. Finally, we apply a threshold on the percentage of infected cells in a particular sample to determine whether the overall sample is infected or uninfected. The reason we need this separate threshold is that every sample will have a small number of objects with intensity above the cutoff, but these are essentially always just autofluorescent spots or debris.

We first needed to determine an appropriate intensity cutoff for calling a cell positive and an appropriate threshold of percent cells infected for designating an overall sample as positive or negative. We anticipated that appropriate classification of samples as positive or negative would require simultaneous adjustment of both of these parameters. For instance, if the uninfected cells and infected cells have similar fluorescence intensity, it would be impossible to pick an intensity cutoff that would perfectly designate individual cells as uninfected or infected. In this case, the best intensity cutoff would still designate some uninfected cells as infected, and thus we would need to adjust the percentage infected threshold to avoid designating a negative sample as positive. The percentage of infected cells threshold is also highly dependent upon the total number of infected cells in the sample. For example, if we are assaying a virus that infects every cell in the sample, we would be able to set a high threshold percentage of infected cells to determine whether the overall sample is positive because every positive sample would be above the threshold. By setting a high threshold, we would avoid a situation where a few autofluorescent cells in a negative sample would cause the sample to be incorrectly designated as positive. For these reasons, we performed a series of experiments on our platform using both infected and uninfected samples to establish both assay parameters. We loaded the microfluidic chip with either influenza infected (n=10 at 0.26% and n=9 at 1.87% infected) or uninfected (n=11) MDCK cells and performed RNA FISH for the virus. We imaged the samples on the chip and processed these images with our software to obtain the median fluorescence

intensity for each cell.

We used this dataset to determine an appropriate cutoff intensity value for designating an individual cell as infected or uninfected. To find the optimum, we generated receiver-operator characteristic curves for all possible per cell intensity cutoffs by plotting the true positive rate versus the false positive rate at various thresholds for percentage of infected cells. Each receiver-operator curve shows how well the threshold of percentage of cells infected serves to classify samples and is dependent upon the particular intensity cutoff chosen. Thus, we selected the intensity cutoff that produced the best receiver-operator curve for percentage infected and used the area under the receiver-operator curve as the metric for making this decision (Fig. 3.9, A, step 4, optimal cutoff = 330). From here onward, for any cell with median intensity above 330, we classified this cell as positive and for any cell below 330, we classified the cell as negative. Further, to confirm that this experimentally determined intensity value is consistent between infections and experiments, we separately infected another batch of MDCK cells with influenza and performed RNA FISH using the platform. We found that the intensity histogram of the median RNA FISH signal in infected and uninfected cells was similar to our prior infections (Fig. 3.8).

With the cutoff intensity for calling a cell positive established, we next sought to establish a threshold of percentage of infected cells for classifying samples as positive or negative (Fig. 3.9, B). At an input of 1.87% infected cells, we found that choosing a threshold between 0.22% and 0.66% yielded an essentially perfect classifier, with an area under the curve of 1. As we lowered the percentage of infected cells in the input down to 0.26%, the performance of our classifier deteriorated, with an area under the curve of 0.88 and an optimal threshold of 0.21% (Fig. 3.9, B). However, with approximately 1000 cells loaded onto the device for each run, we found that the decline in performance was likely due to randomly sampling only uninfected cells due

to the low percentage of infected cells (Fig. 3.8).

With our intensity cutoff established, we wanted to verify that our assay operated consistently. Thus, separate from the data set in Fig. 3.9, B, we performed 10 independent influenza RNA FISH experiments in the device with 1.53% infected cells and determined which cells were positive by applying the RNA FISH intensity cutoff previously established. We again considered the percentage of infected cells as the discrimination variable and plotted a receiver operator characteristic curve by varying this parameter. As before, the curve indicated that we are able to perfectly classify uninfected and infected samples. Taken together, these experiments demonstrate that our viral RNA FISH platform detects influenza and discriminates positive and negative samples with high accuracy and reproducibility.

3.6 Multiplex viral detection by integrating specific and pan-probes in the device

One strength of RNA FISH is the ability to multiplex by using multiple probes labeled with differently colored fluorophores on a single sample simultaneously. To demonstrate this ability with our platform, we combined our subtype specific probes and our pan-probes into a single assay that we can run in the device (Fig. 3.10, A). We designed a panel using the subtype specific influenza probe sets (H3N2, H1N1, and influenza B) and rhinovirus pan-probe sets, each labeled with different fluorophores. We then infected MDCK cells with the different influenza strains and infected HeLa cells with rhinovirus. We loaded each of these infected populations into a separate microfluidic device and performed RNA FISH for all four different viruses. For this experiment, we wanted to know if we could identify which virus was causing infection at the single-cell level. By setting the cutoff for calling a cell positive as four standard

deviations from the mean intensity of our uninfected control, we were able to clearly identify cells infected with the correct virus (Fig. 3.10, B). We observed minimal off-target labeling, ranging from 0% to 5.8% of cells incorrectly classified as having another virus; however for each virus, the total percentage of miscalled cells was consistently <10% of the total number of cells. Thus, we conclude that this assay is capable of subtyping one virus, pan-probing another virus, all in the device simultaneously on a single sample.

3.7 Viral SNP FISH detects drug resistant viruses

In addition to having different subtypes, viruses can also acquire important single base mutations such as those that confer resistance to antiviral medications [7, 25, 70]. However, conventional RNA FISH does not have the specificity to discriminate single bases because a one base mismatch does not create a large difference in relative binding affinity for an entire 20 base oligonucleotide. Our lab has recently developed a modification on RNA FISH that enables us to detect single-nucleotide polymorphisms (SNP) in individual RNA transcripts, and we have demonstrated that it is compatible with rapid hybridization [57, 106]. In order to improve specificity, we use a "mask" oligonucleotide that prevents cross hybridization by magnifying the relative difference in binding energy from a one base mismatch (Fig. 3.11, A). While the mask improves specificity of the assay, any single oligonucleotide will still bind to many off-target sites throughout the cell. Thus, we make a "guide" probe consisting of many fluorescently tagged DNA oligonucleotides that brightly and specifically label individual RNA molecules, thereby showing us where to look for SNP probes that are correctly binding to their target. We then only consider SNP probe signal that co-localizes with the spots from our guide probe.

We tested whether our assay could detect a point mutation in the neuraminidase gene (nucleotide position 823 C -> T or amino acid position H274Y in N2 numbering) of influenza A H1N1, which was in circulation prior to 2009 [7, 25]. This mutation alters the oseltamivir binding site in neuraminidase making it ineffective in treating influenza (Fig. 3.11, B) [22].

To test whether our assay could reliably distinguish between drug-resistant and wild-type influenza, we infected MDCK cells with wild-type influenza A/California/07/2009 H1N1 and with A/California/07/2009 H1N1 engineered to possess the H274Y (C823T nucleotide) NA mutation. We then used SNP RNA FISH to classify individual transcripts as wild-type or resistant mutant RNA. Qualitatively, we observed that the cells infected with the wild-type virus had RNA classified as predominantly wild-type and vice versa with the mutant (Fig. 3.11, C and Fig. 3.12). For each cell, we compared the relative amounts of each RNA and calculated the ratio of resistant mutant RNA to wild-type RNA (Fig. 3.12). We used this metric as a binary classifier to determine which virus was infecting each cell and created a receiver operator characteristic curve. Each point on the curve represents a potential threshold ratio of resistant mutant RNA to wild-type RNA (Fig. 3.11, D), with points closest to the upper left-hand corner representing thresholds that generated the most sensitive and specific classification. Applying this analysis to our data, we found that a threshold ratio of around 50:50 mutant RNA to wild-type RNA is optimal, which is also intuitively the ideal choice for this threshold. At this 50:50 threshold, the sensitivity of this assay for individual cells is 0.96, and the specificity is 0.97 (false positive rate 0.03).

3.8 Clinical testing of RNA FISH-based respiratory diagnostics

We designed and executed a small clinical study to evaluate the potential of RNA FISH as a method for rapid influenza diagnostics. We recruited our subjects during the fall 2014 through spring 2015 influenza season from patients admitted to the Children's Hospital of Philadelphia (CHOP). We selected subjects who had a rapid respiratory panel (which is a PCR-based test) sent as part of their clinical care to use this test result as a gold standard for comparison to the RNA FISH. From each subject, we collected a nasal swab and a nasopharyngeal swab for analysis. In total, we collected samples from 7 patients who were infected with influenza as determined by the rapid respiratory panel performed at CHOP. We also collected samples from a handful of subjects uninfected or infected with other viruses, including rhinovirus and adenovirus. Across all of the samples, we had only one where we found RNA FISH signal that we believe is specific to the expected virus. This sample is from a patient infected with adenovirus, and we performed the assay by cytospinning the sample directly onto a coverslip followed by rapid RNA FISH (Fig. 3.13). It is still unclear why the assay worked beautifully on this one sample, but not others. One possibility is that there was something unique to this patient that allowed the virus to persist inside intact cells for longer periods of time, specifically the patient could have had some immunodeficiency or have been treated with an immunosuppressive drug. Without other clinical details about the subject, it is challenging to interpret this result. A huge limitation of this study was the timescales on which we collected samples for RNA FISH relative to the rapid respiratory panel. With most subjects, there was more than 24 hours between the two tests, which could be one of the reasons we did not see signal in these samples. Another problem with the timing arose because routine clinical care at CHOP is to

treat patients with suspected influenza infection with tamiflu, and therefore, all of the influenza infected subjects in our study received tamiflu prior to their RNA FISH test. Thus, it is possible that these subjects cleared the virus from their nasopharyngeal epithelium before we collected the RNA FISH samples. Taken together, we believe that these experiments are inconclusive in evaluating RNA FISH as a viable rapid diagnostic technology as the design of the study had many limitations. To properly evaluate this technology, it would be crucial to perform the gold standard test and the RNA FISH-based test on the same sample or samples collected at the same time.

3.9 Discussion Flu RNA FISH

In this chapter, we outline a platform for rapid viral detection using RNA FISH, including viral probe design software, a microfluidic device to automate RNA FISH, and image-processing and analysis software to discriminate infected and uninfected samples. To tailor RNA FISH to viruses, we developed probe design software that allows the user to design subtype specific probe sets or probe sets that target many viral strains simultaneously and tested this software using influenza and rhinovirus, respectively. The microfluidic device provides simple, semi-automated, and standardized processing of samples for RNA FISH. Finally, our image-processing software and statistical framework enables automatic classification of infected or uninfected samples. We tested our technology on influenza, rhinovirus, and adenovirus, respiratory viruses that require rapid diagnosis in outpatient and emergency settings. Our platform demonstrated high classification performance and the ability to target multiple viruses simultaneously. In addition, the entire assay, including the RNA FISH and the imaging, takes only 15 minutes. These components make rapid RNA FISH a viable technology for viral diagnostics.

The distinct advantages of RNA FISH-based viral diagnostics are its speed, cost per test, and flexibility and specificity of the probe design. The two most relevant diagnostic technologies for comparison are RT-PCR and rapid viral antigen tests. RT-PCR is highly sensitive and specific, but requires 1-2 hours to perform in a pathology laboratory [56]. In contrast, rapid antigen tests require 5-15 minutes and can be performed at the point of care [56], but have variable sensitivity and specificity. We believe that RNA FISH combines some advantages of both of these techniques. The overall time to perform RNA FISH on the microfluidic chip is 8 minutes and the imaging typically requires 5 minutes. However, we aim to reduce the device time and imaging time ever further as hybridization to influenza RNA can take as little as 30 seconds (Fig. 3.14). Newer influenza diagnostics use isothermal amplification and are significantly faster, between 15-30 minutes [21, 78]. For example, the Alere i Influenza test takes 15 minutes and in the initial clinical studies, performed with a sensitivity 91.8-97.8% and specificity 85.6-96.3% [6]. This test was CLIA waived in 2014 and may become widely adopted as it offers the speed of rapid antigen tests with sensitivity and specificity approaching that of conventional RT-PCR. However, unlike RNA FISH, this test does not discriminate between viral strains at the single base level, which can be needed to distinguish drug resistant from drug sensitive strains [6].

A strength of RT-PCR is its sensitivity, which one study showed was able to detect as little as 10-50 copies of virus per sample [8]. While in principle RNA FISH could have sensitivity down to even single molecule resolution, we anticipate that another issue is that if only very few cells are infected, it may lead to false negatives. That said, at low levels of infection, we found that individual cells still produced bright signal, albeit at a reduced frequency (Fig. 3.16). If this holds true in clinical samples, it may be possible to improve the sensitivity by imaging many cells.

A clear strength of RNA FISH is the flexibility of probe design. The software in

our platform makes it easy to generate custom designs to specifically target different viral strains and/or detect many viral strains with one probe set. These features of our software will allow RNA FISH to be performed on emerging viruses and viruses that rapidly mutate, with little development time and cost. While development of probes requires viruses to be sequenced, sequencing is often routine on viruses with a large public health impact and costs are rapidly falling. Combining the probe design and multiplex capabilities as well as the ease of use and low cost of the microfluidic device, RNA FISH could also be used as a surveillance tool to monitor the spread of viral disease. For comparison, antibody based tests, including rapid antigen tests or ELISAs, use high affinity antibodies, the development of which can take months, requires extensive validation, and is not possible for some targets [9]. This is a major concern because antibody-based diagnostics may not be readily available during an outbreak of a new virus or if a virus acquires mutations that prevent binding of the antibody. Similar to RNA FISH, it is possible to quickly develop and deploy new primer sets for conventional RT-PCR assays as was done with the novel H1N1 swine flu in 2009 [87, 125, 129]. However, this level of flexibility with primer design is not yet possible with the newer and faster isothermal amplification techniques. Thus, currently only RNA FISH combines the ability to quickly design new probe sets with a fast assay turnaround time.

Specificity of the probe sets is another strength of RNA FISH-based viral detection assays. Our data demonstrates that RNA FISH can not only differentiate between strains of virus, but also can detect a single nucleotide mutation in influenza conferring resistance to oseltamivir. Detection of single amino acid changes is often not possible with antibody-based diagnostics. While RT-PCR based approaches can detect point mutations, these assays generally have greater complexity, less ease of use, and lower sensitivity than conventional RT-PCR [82, 127].

While still a nascent technology for clinical use, RNA FISH has the potential to be a very cost effective detection assay. Synthesizing a new RNA FISH probe set for a laboratory scale costs between 300−600 for 10,000 tests, which comes to a total of 3-6 cents per test, and at larger scales, the cost of the probe sets would be even lower. Similarly, we expect the microfluidic device to be inexpensive, particularly at large scales where it could be manufactured with injection molding. While the consumables are cost effective, the equipment needed to image these samples is more expensive. However, many research groups are working on developing low cost and portable microscopes and such equipment would be necessary to use this assay at the point of care [10, 132]. For comparison, RT-PCR also requires costly equipment including a thermal cycler and uses enzymes which often have limited shelf-life and licensing fees that raise their costs.

In summary, we developed a pipeline for rapid RNA FISH that demonstrated high classification performance, multiplexing ability, and the ability to distinguish individual strains of a virus. Our work establishes RNA FISH as a viable methodology for rapid detection of viruses with the potential to enable point-of-care diagnostics applications.

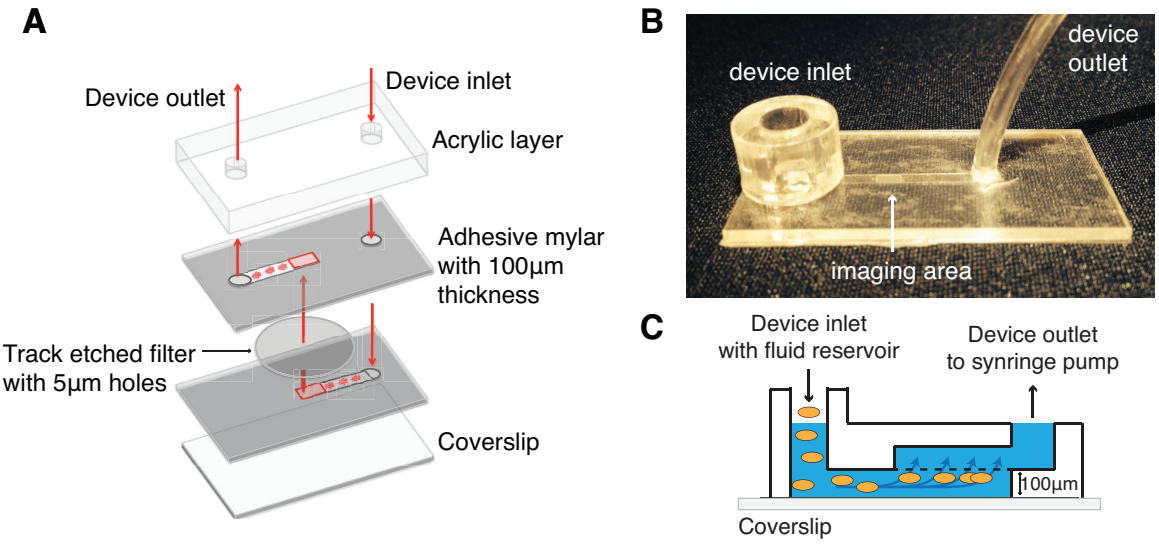

Figure 3.6: Overview of device construction. We constructed the device from laser cut pieces of acrylic, double sided adhesive mylar (thickness 100 μm, 3M company), a polycarbonate track-etched filter (pore diameter = 5 μm, Whatman, Nuclepore), and a number 1 coverslip. A) We assembled the device by layering the pieces as shown. B) Photograph of the assembled device. C) Schematic of the device with cells trapped under the filter.

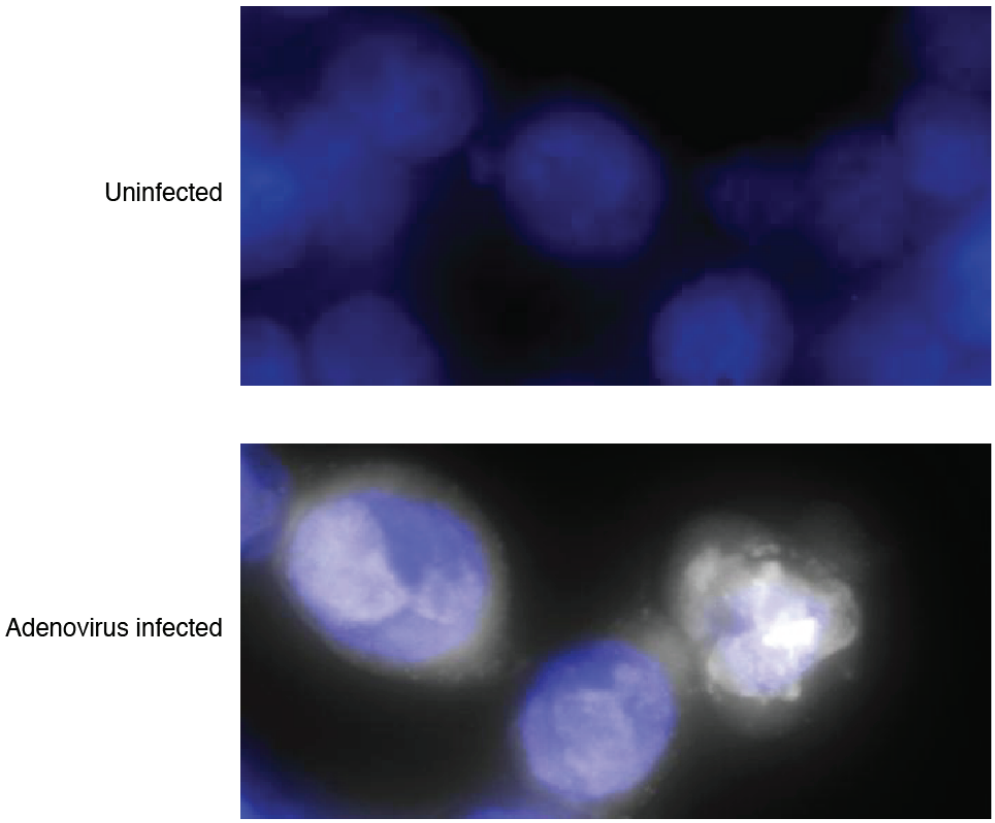

Figure 3.7: *W*e designed adenovirus pan-probes to cover adenovirus serotypes 1-7. We then infected 293T cells with adenovirus, fixed the cells with methanol, and performed rapid RNA FISH for the virus. Infected cells had bright signal from the RNA FISH probes while uninfected cells remained dark. DAPI (nuclear stain) is in blue, and RNA FISH is in white. All images are 100X magnification.

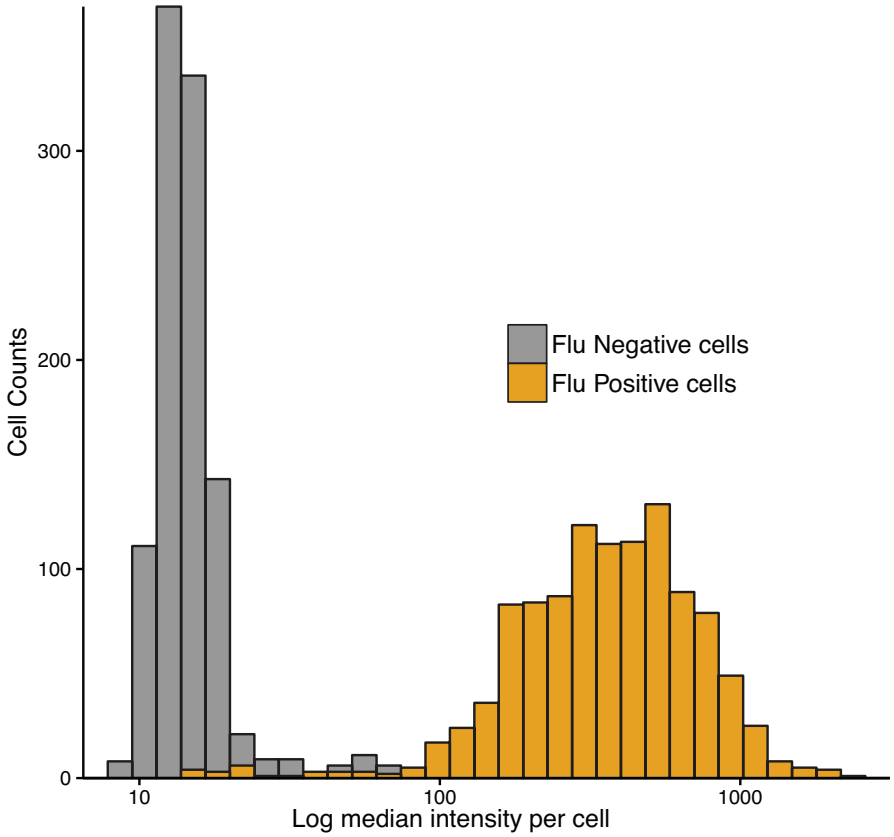

Figure 3.8: The intensity profiles for infected and uninfected cells are reproducible between biological replicates. We loaded the device with either influenza A/Puerto Rico/8/1934 H1N1 infected or uninfected MDCK cells and performed RNA FISH for the virus. We ran our image processing pipeline to extract the fluorescence intensity signal for the viral RNA FISH in each cell and calculated the median intensity. As expected, we found that the median intensity in most infected cells was higher than the median intensity in most of the uninfected cells. In this biological replicate, our intensity cutoff of 330 was again sufficient to discriminate uninfected and infected cells.

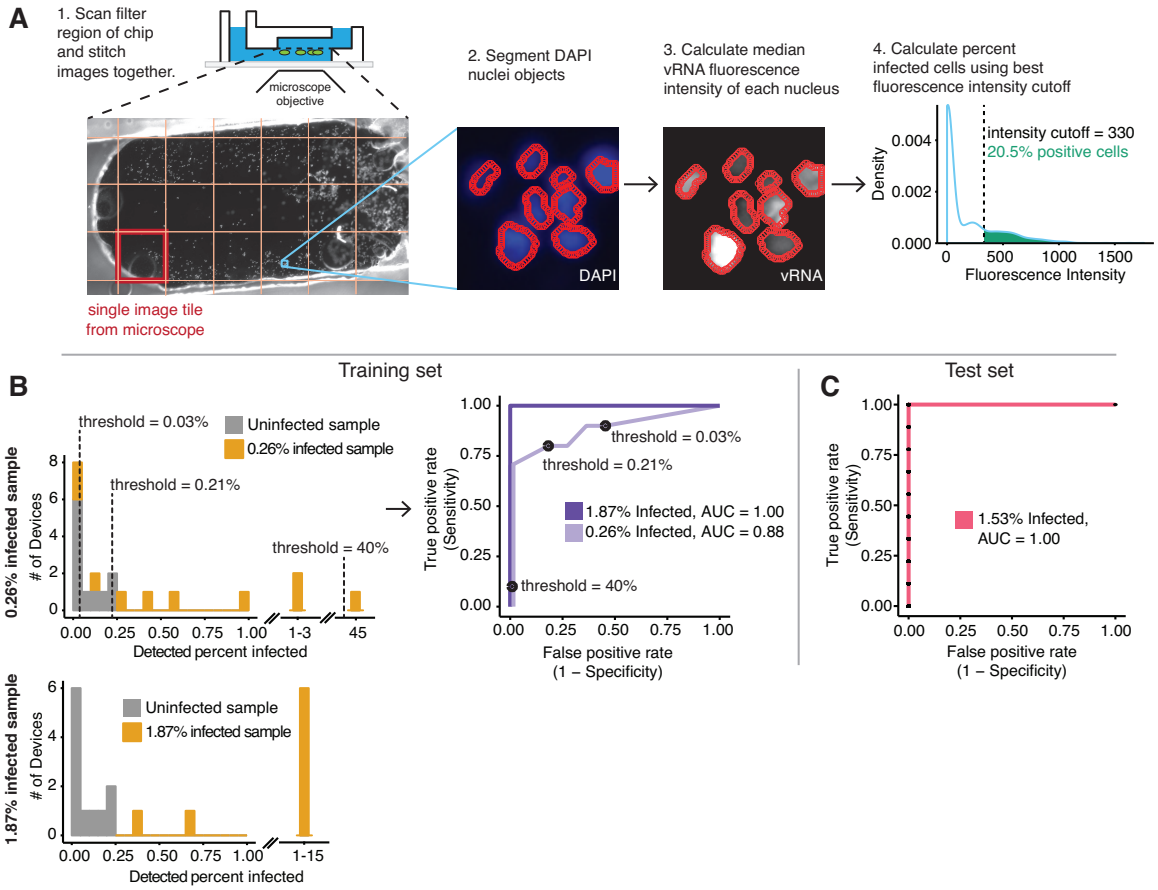

Figure 3.9: Automated computational analysis demonstrates high specificity and sensitivity for viral detection. A. Overview of our image analysis software, in which microscopy images are automatically stitched, viral RNA (vRNA) fluorescence intensity is measured in and surrounding each DAPI-stained nucleus, and the percentage of positive cells is calculated using a cutoff that best distinguishes devices with infected cells from uninfected cells. B. Histogram and receiver operating characteristic curves for devices with 1.87% and 0.26% percent infected samples. C. Optimized intensity cutoff from A and applied to an independent dataset of uninfected devices and devices with 1.53% infected samples. White scale bar represents 25 microns.

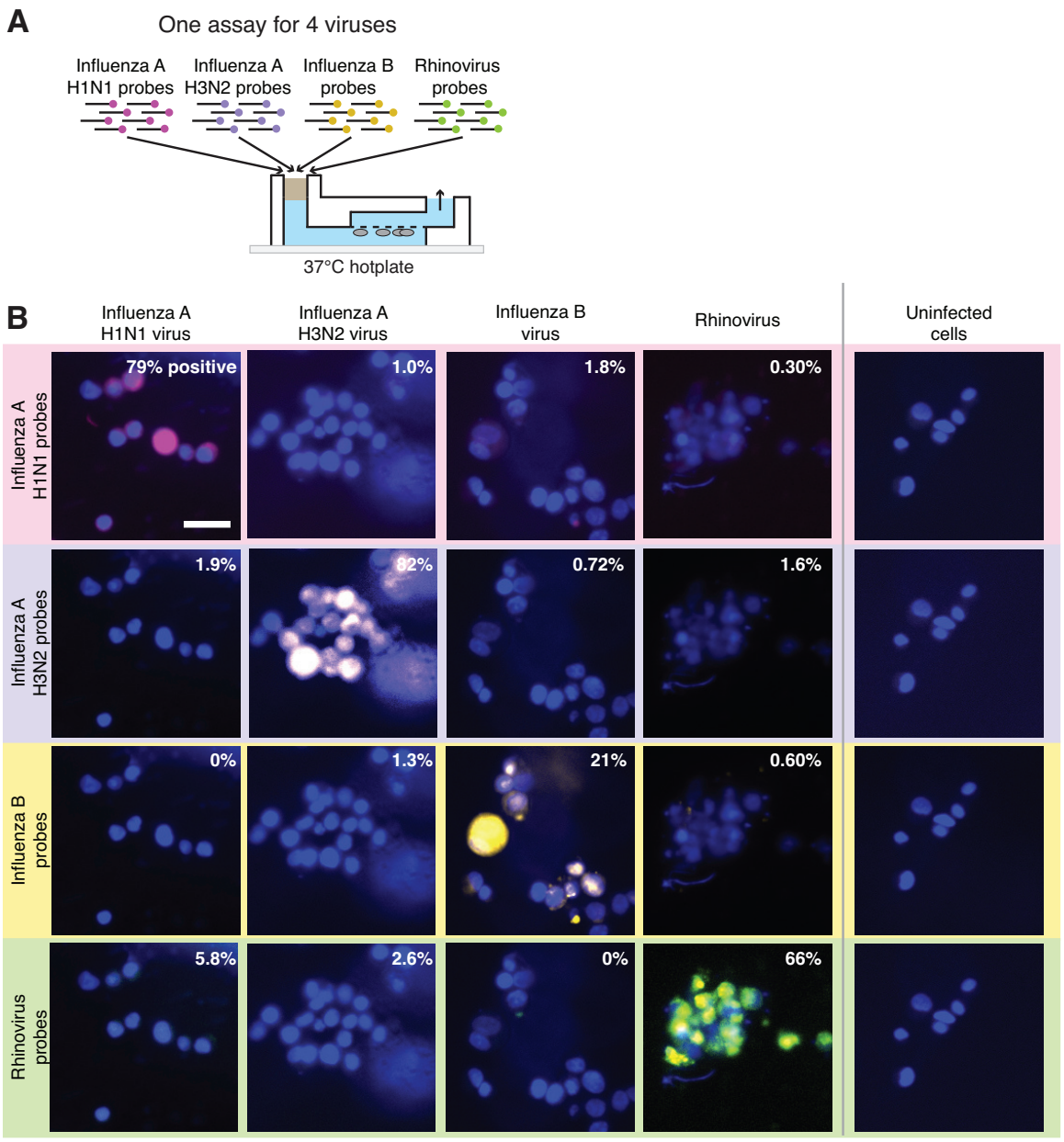

Figure 3.10: Multiplex RNA FISH detects four different viruses in one test. A. We used an RNA FISH probe cocktail including probe sets for influenza A H1N1, influenza A H3N2, influenza B, and rhinovirus labeled with four differ fluorophores. For each virus, we loaded the chip with infected cells and hybridized on the probe cocktail. B. We found that the probe sets brightly labeled the correct virus with minimal fluorescence from the probe sets for other viruses. In the analysis, we set the cutoff for calling a cell positive as four standard deviations from the mean intensity of the uninfected control. The numbers in the upper right-hand corner of each image indicate the percentage of cells designated as positive for that virus. DAPI (nuclear stain) is in blue, and RNA FISH is in white. All images are 20X magnification. White scale bar represents 25 microns.

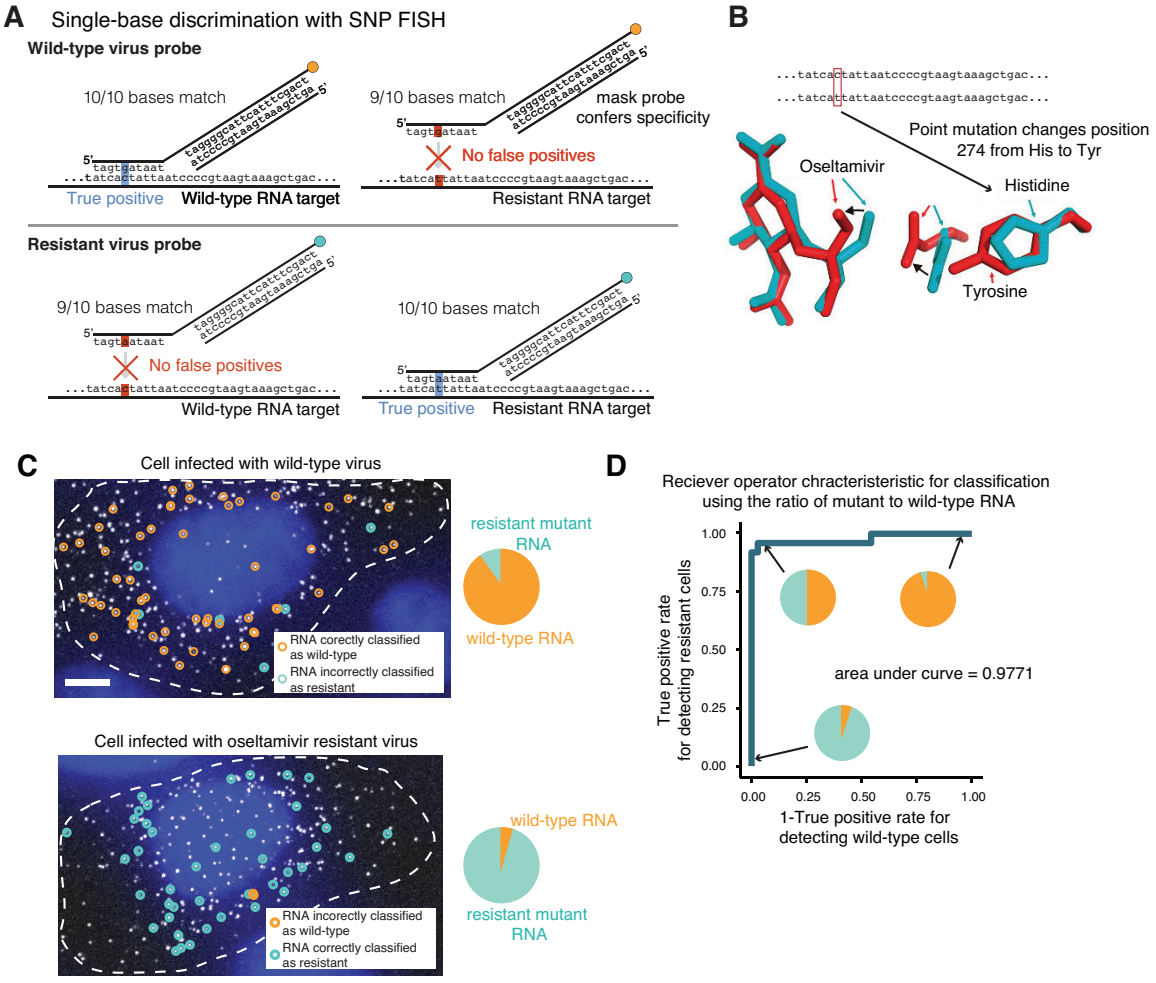

Figure 3.11: Viral SNP FISH detects drug resistant viruses with one base pair sequence difference. A. We use a "mask" oligonucleotide to improve the specificity of a single probe by magnifying the relative energy contribution from a one base mismatch. B. We use the SNP FISH technique to detect a point mutation in the neuraminidase gene that changes a histidine to a tyrosine, which alters the structure of neuraminidase such that the neuraminidase inhibitor oseltamivir is less effective in treating influenza. C. Here, we infected MDCK cells with wild-type influenza A H1N1 and mutant influenza A H1N1 containing the oseltamivir resistance mutation. Representative images of SNP FISH on these cells show that most RNA transcripts are correctly labeled by the probes. DAPI (nuclear stain) is in blue, and RNA FISH is in white. Both images are 100X magnification. White scale bar represents 5 microns. D. A receiver-operator characteristic curve demonstrates that this assay is effective for classifying cells as infected with wild-type or resistant mutant virus.

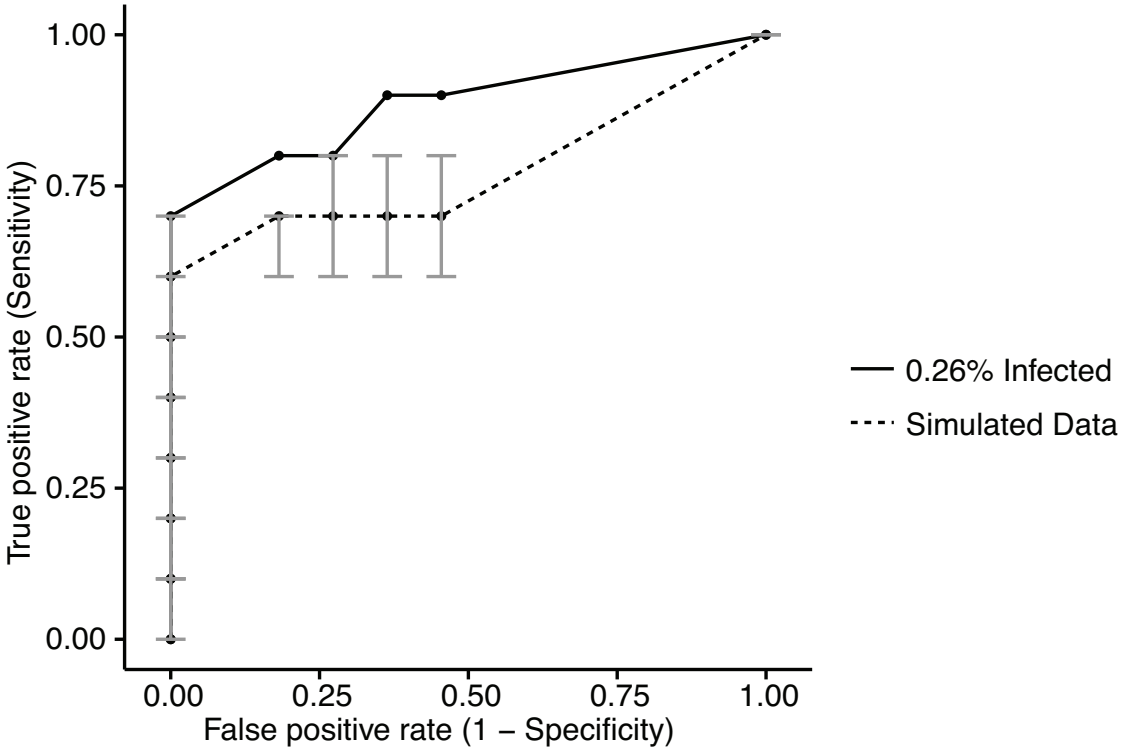

Figure 3.12: Classification using our pipeline approaches the expected classification limit of sampling from small numbers of cells. Simulated data of percent infected cells was generated by sampling 25,000 times from a binomial distribution with probability of trial success equal to 0.26% and number of observations equal to the experimentally determined number of cells in each device with the 0.26% infected sample. For each of the 25,000 simulations, we generated a receiver operating characteristic curve using this simulated data for infected devices and experimental data for uninfected devices. Dotted line represents the median and error bars in gray represent the upper and lower quartile at each percent infected threshold of the receiver operating characteristic curves. Solid black line represents experimental data from uninfected devices and devices with 0.26% infected cells, as shown in Fig. 3.9.

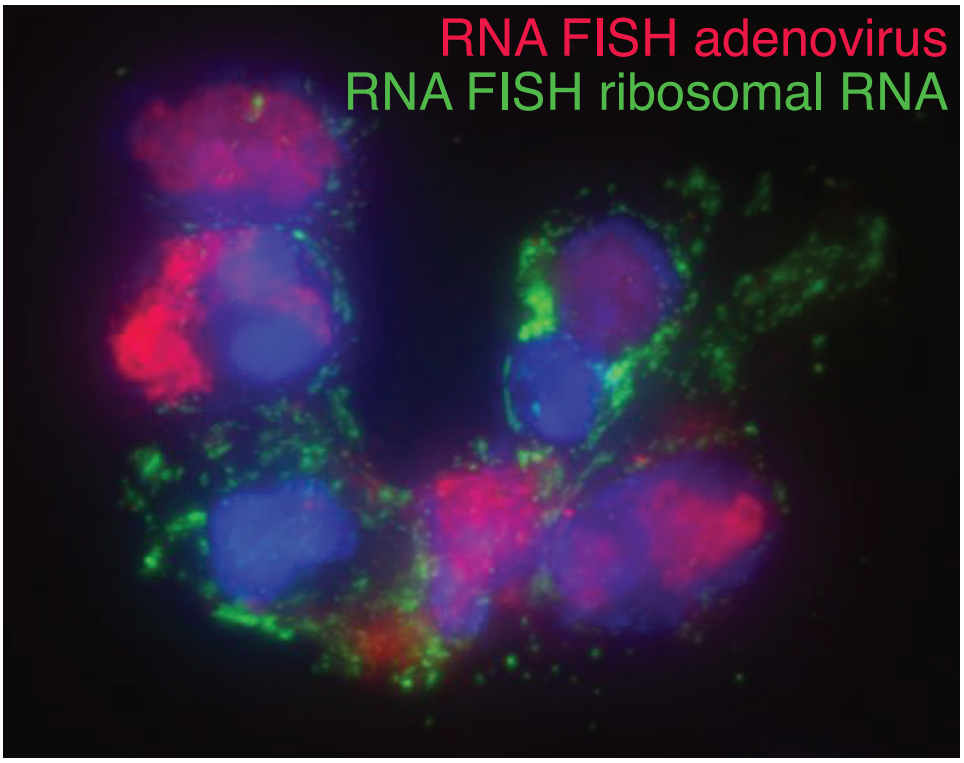

Figure 3.13: Image of rapid RNA FISH for adenovirus in an adenovirus infected patient. We cytospun cells onto a coverslip from a nasopharyngeal sample of an adenovirus infected patient and then performed RNA FISH directly on the coverslip. The red shows adenovirus specific signal, and the green shows human ribosomal RNA. Importantly, the two RNA FISH signals localize to different regions inside the cells, demonstrating that these signals are not from non-specific binding of probe to any target. The cell nuclei are shown in blue.

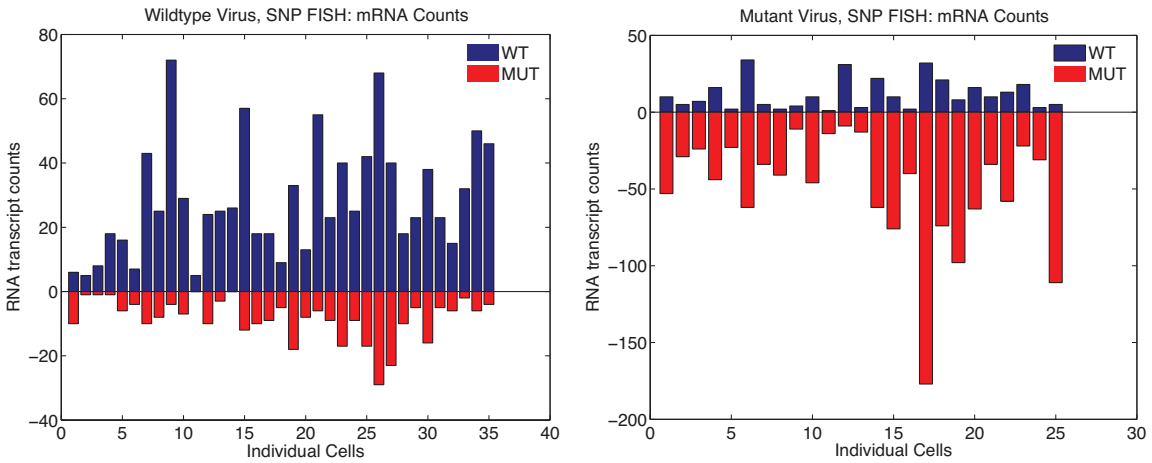

Figure 3.14: Viral RNA molecules in cells infected with wild-type and resistant mutant virus are classified by SNP FISH. We infected MDCK cells with wild-type influenza A/California/07/2009 H1N1 and with A/California/07/2009 H1N1 engineered to possess the H274Y (C823T nucleotide) NA mutation. We used SNP RNA FISH to classify individual transcripts as wild-type or resistant mutant RNA. For each cell, we quantified the amounts of each RNA as shown by the bar graphs. Cells infected with wild-type virus had more wild-type RNA, and cells infected with the resistant mutant virus had more mutant RNA.

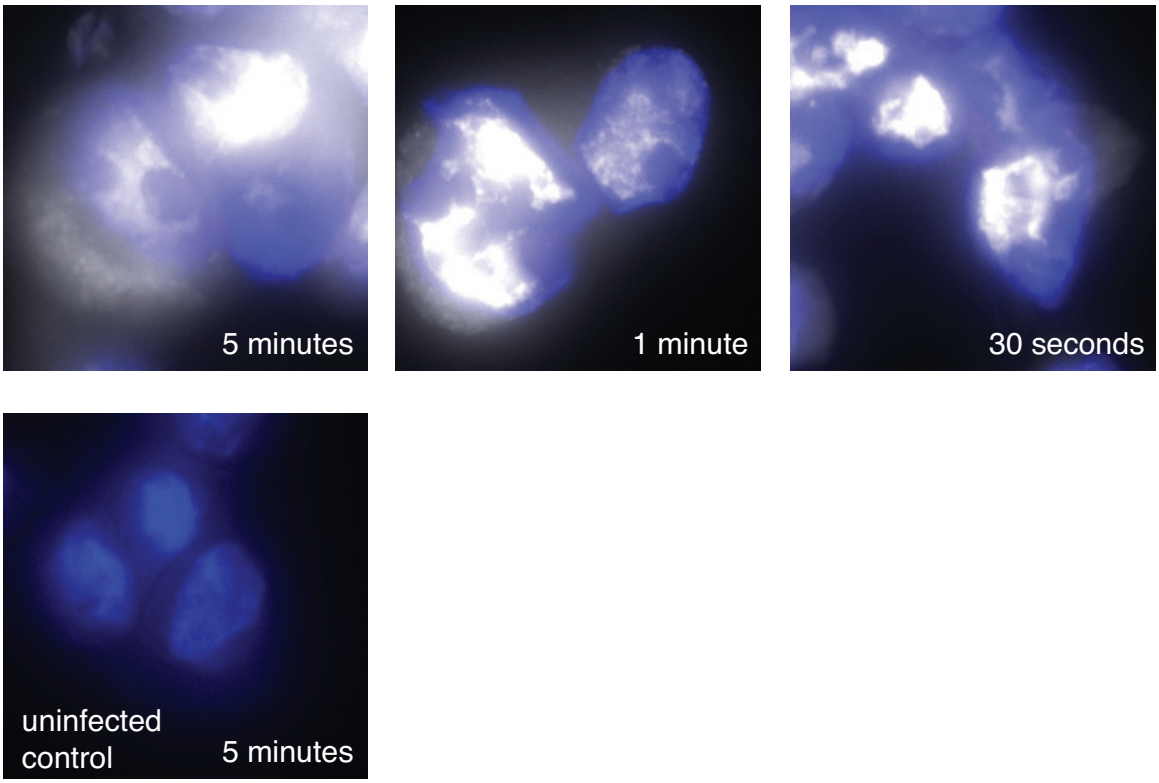

Figure 3.15: Hybridization of Flu RNA FISH probes requires as little as 30 seconds. We infected MDCK cells with influenza A/Puerto Rico/8/1934 H1N1 and fixed the cells 18 hours after infection. We then performed rapid RNA FISH using hybridizations times of 5 minutes, 1 minute, and 30 seconds. For each sample, we used probe sets for the haemagglutinin (HA) and nucleoprotein (NP) segments of influenza A/Puerto Rico/8/1934 H1N1. With each of the hybridization times tested, we observed bright fluorescent signal in infected cells. DAPI (nuclear stain) is in blue, and RNA FISH is in white. All images are 100X magnification.

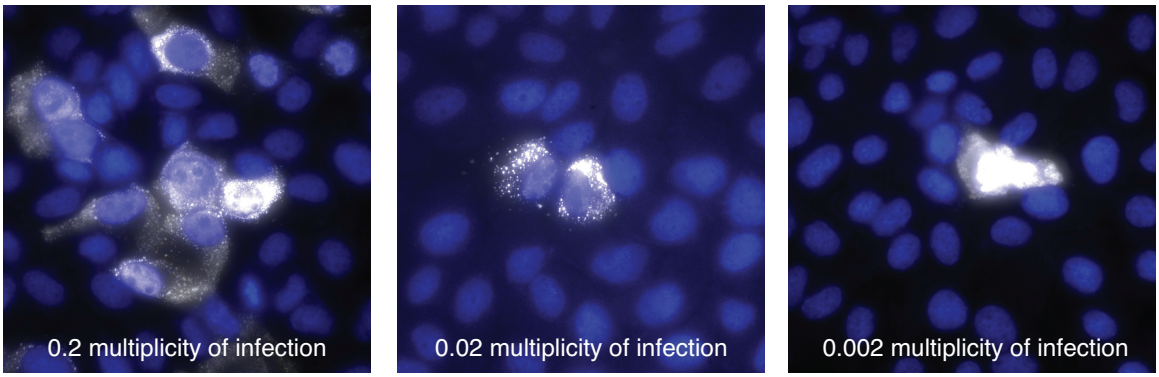

Figure 3.16: At low levels of infection with influenza, individual cells still have bright signal by RNA FISH, but are at low frequency. We infected MDCK cells with influenza A/Puerto Rico/8/1934 H1N1 at multiplicity of infection 0.2, 0.02, and 0.002. Eighteen hours after infection, we fixed the cells and then performed rapid RNA FISH using probes for all 8 segments of influenza A/Puerto Rico/8/1934 H1N1. In the samples with lower multiplicity of infection, we observed fewer infected cells overall, but noted that the individual infected cells in each samples had comparable fluorescence intensity. DAPI (nuclear stain) is in blue, and RNA FISH is in white. All images are 100X magnification.

Chapter 4

Single cell analysis of resistance to targeted therapy

4.1 Background: Single cell heterogeneity in cancer

Biology is replete with examples of individual, seemingly identical, cells behaving in qualitatively different ways. Generally, studies in this area have focused either on random events in gene expression, leading to "noisy" behavior [92, 101, 102, 114] often with biological consequences [33, 68, 88, 112, 114, 128], or rare cell variability, which is typically ascribed to genetic differences-the mutational origins of cancer being a salient example of the latter. However, there are many biological phenomena in which the behavior of rare cells differs from the rest of the population without any genetic underpinning, such as the induction of pluripotent stem cells via expression of particular pluripotency-associated transcription factors [115], cancer stem cells [42], and bacterial persistence [2, 32].

The development of resistance is a prime example of biologically important behavior in rare cells in which a genetic basis is often assumed but not always demonstrated. In most resistance phenomena, a drug or other selective agent is applied to kill or

otherwise inhibit an unwanted cell population, but a small number of cells evade the drug activity and continue to proliferate, ultimately generating a resistant population. The rarity of these cells is often taken as evidence for a genetic basis to resistance, but there are examples of non-genetic mechanisms, such as bacterial persistence and both short-[19, 32, 38, 86, 96, 111, 122] and long-term resistance to targeted therapies in cancer [50, 85, 86, 107]. Yet, despite their importance, the underpinnings of such non-genetic rare-cell resistance behavior remains mysterious, often because of technical limitations in measuring non-genetic differences in large numbers of individual cells; indeed, there have to our knowledge been only very few observations to date of rare fluctuations in gene expression [29, 45, 79], let alone connected to resistant phenotypes.

A paradigmatic example of resistance in cancer is melanoma, a disease which often results from mutations to the BRAF protein leading to uncontrolled cellular proliferation [24]. Treatment with the drug vemurafenib, an inhibitor of the mutated V600E BRAF protein, nearly eradicates the tumors [16, 37], but these therapies rarely result in a permanent cure because a small subset of cancer cells develop resistance to the drug and cause the disease to relapse [124]. Drug resistance in the bulk relapsed tumor is well characterized and can be due to several factors including re-activation of the pathway through increased activity of upstream or downstream effectors, activation of bypass pathways, and alternative splicing altering drug binding [39, 76, 108]. Often, these mechanisms are associated with a genetic mutation, and so the assumption is that at the moment the drug is applied, Darwinian selection of rare cells with these mutations in the original tumor population lead to tumor regrowth [26, 44, 52, 80, 85, 93, 121]. Yet, there is some evidence both in melanoma and other cancers that non-genetic heterogeneity may play a role in why some cells respond to drug while others do not, in the short term due to gene expression "noise" [19, 111] and on longer timescales with subpopulations of cells that go into a slow-cycling drug

tolerant state [98, 99, 107].

Here, we decompose the process of developing resistance to targeted therapies into two distinct phases. The first "priming" phase determines which rare individual cells within the population proliferate even when treated with drug. We used Luria-Delbrck fluctuation analysis [36, 66, 118] to show that the pre-resistant state is transient, and developed a highly-multiplexed single molecule RNA FISH assay to measure expression levels in many tens of thousands of individual cells to show that the primed state is characterized by rare-cell expression of high levels of resistance marker genes. We then show that the addition of drug itself induces a reprogramming phase, in which these primed cells transform into stably resistant cells. This reprogramming occurs in at least two steps, the first of which is dedifferentiation, and the second of which involves the activation of new signaling pathways. Our results provide a framework for understanding transient and epigenetic contributions to the development of resistance and suggest a mechanism by which to understand the efficacy of "drug holiday" treatment regimens [23]. Further, we found that many of the same set of genes show similar rare-cell expression patterns in many different cell types, suggesting the existence of a general rare-cell transcriptional program.

4.2 Heritability of drug resistance

Patient-derived melanoma cells grown in culture provide a model for examining drug resistance. When we cultured cells isolated from two different patients (WM9, WM983B) under normal conditions, they proliferated readily. Upon administering a fractional killing dose of vemurafenib (1 μM, Fig. 4.1), the vast majority of cells stopped growing, but sporadic colonies of resistant cells would form that proliferated even in the presence of drug. (The signature of these surviving cells resembles those of

resistant cells in patients; Fig. 4.2) To examine the population dynamics of resistance at the single cell level, we performed large-scale time-lapse imaging of around 18,000 cells (at the time of drug application) over the course of 28 days. We began our time-lapse 5 days before the addition of drug to build a family tree, then added drug and watched for the formation of resistance clusters of cells (the low frequency of resistance necessitated the large scale of imaging, as we expected roughly 9 resistant colonies total in the entire imaging field). We found that drug resistant colonies derived from single cells (Supplementary video 1) or low numbers of cells (Supplementary video 2), and indeed, even sister cells can have different phenotypes in response to drug (Fig. 4.3). The cells destined to become resistant proliferated normally before the addition of drug, suggesting that they are not in a dormant "persister" cell state [3].

Given the observation that resistant colonies come from single cells, we wanted to know what is different in the cells that develop into the resistant colonies. We considered two possible models for resistance: a genetic mutational model and a transient non-heritable model (Fig. 4.4, A). The primary difference between the mutational and transient models is that in the former "strongly heritable" (i.e., genetic or epigenetic) model, once a cell transitions to the resistant state, the cell cannot revert (at least not in a relevant time frame), whereas in the transient model, cells can easily transition between a pre-resistant and non-resistant state, with the pre-resistant cells being defined as those that will ultimately give rise to resistant colonies upon addition of drug (Fig. 4.4, A). To differentiate these two hypotheses, we used the "fluctuation analysis" framework developed by Luria and Delbrck in 1943 to study bacterial resistance with T1 phage [66]. First, we isolated a single cell from the parental cell line to minimize any existing genetic heterogeneity in the cell line. We expanded this cell for only around 7-8 divisions to minimize heterogeneity and then started several single cell cultures from this small pool. We expanded each of these cultures

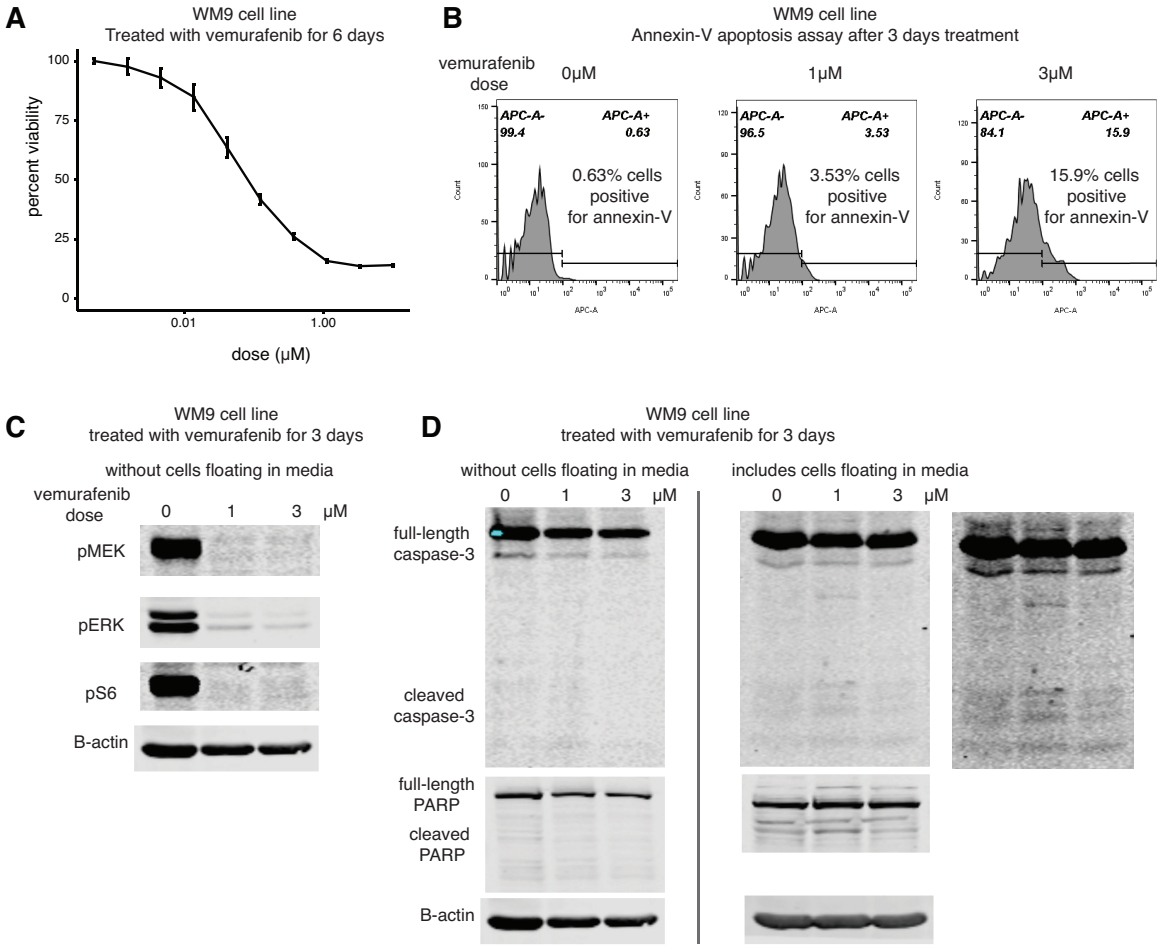

Figure 4.1: Treatment of WM9 melanoma cells with vemurafenib induces cell death. A. Percent viability of WM9 cells treated with vemurafenib for 6 days (MTS assay). B. Annexin-V staining in WM9 cells after 3 days of treatment with vemurafenib measured by flow cytometry. The percentage of cells that are positive for annexin-v is labeled on the plots. C. Western blot for pMEK, pERK, and pS6 after 3 days of treatment with vemurafenib. D. Western blot for caspase-3 and PARP after 3 days treatment with vemurafenib. This demonstrates that WM9 cells are highly responsive to BRAF inhibitor treatment with inhibition of signaling downstream of BRAF and apoptosis of sensitive cells.

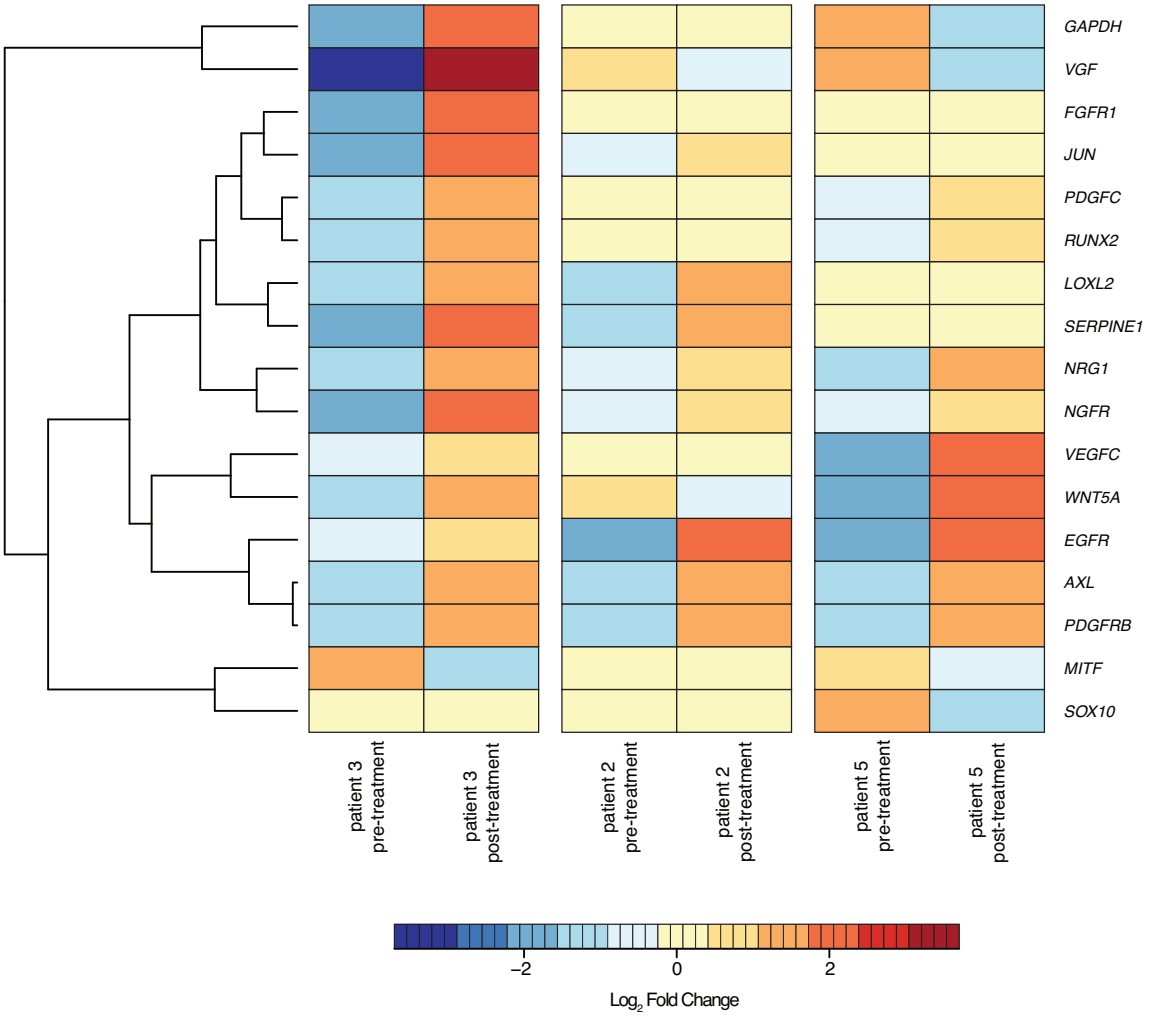

Figure 4.2: RNA sequencing of patient tumors pre-treatment and post-treatment from Sun et al. Nature 2014 shows changes in gene expression for many of the same resistance marker genes found in WM9 cells. Heatmap depicts the $\log_2$ fold change for each gene. Samples are normalized by patient. The genes displayed here are the same panel of genes used for RNA FISH in WM9 cells in Fig. 4.14, A. This analysis demonstrates that there is significant overlap between the transcriptional signature of resistance in WM9 cells and resistant patient samples.

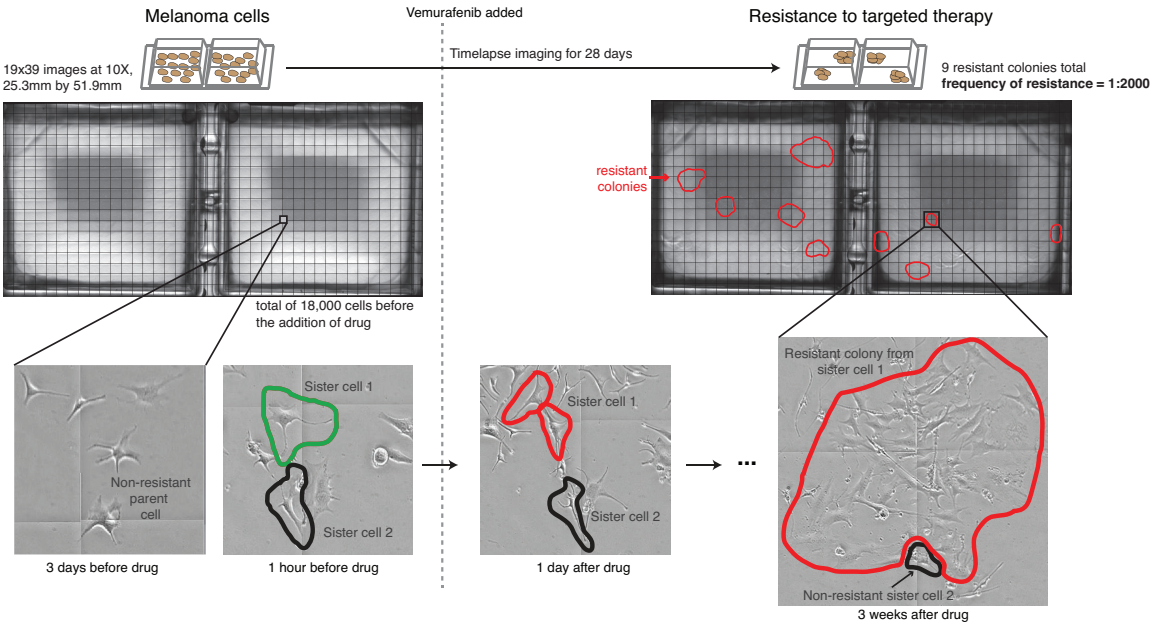

Figure 4.3: Time-lapse imaging reveals that two sister cells can have different phenotypes in response to vemurafenib. Twenty-eight day time-lapse images of WM9 cells before and then after application of cytostatic dose of vemurafenib. Sister cells are labeled in the images. There were approximately 18,000 cells at the time that we applied drug, and a total of 9 resistant colonies formed on the culture dish. We observed instances in which two sister cells exhibited divergent phenotypes, for instance, one would respond to drug while the other would continue growing, eventually forming a resistant colony. These results suggested the possibility of a non-genetic resistance mechanism, although do not constitute proof.

through 20 doubling events, yielding around 1 million cells, after which we added drug to the cells and counted the number of resistant colonies that formed in each culture (Fig. 4.4, A). As per Luria and Delbrck, if the transition to resistance is heritable (such as a DNA mutation or epigenetic change), then mutations would sometimes occur early during the expansion phase before adding drug, leading to some cultures with very large numbers of resistant colonies. If, however, the pre-resistant state is a transient one, then we would expect all cells in any given culture to have a roughly equal chance of forming a resistant colony upon adding drug, greatly reducing the possibility of cultures with large numbers of resistant colonies.

Upon performing this experiment, we found that the lack of outliers in the distribution of number of resistant colonies was incompatible with a strongly heritable resistant phenotype. Simulations of the Luria-Delbrck distribution confirmed the statistical significance of our results with a p-value of 0.0005 and 0.0012 in WM9 cells (biological replicates with 43 and 29 cultures, respectively) and p-value 0.0395 for WM983B cells (WM9 data in Fig. 4.4, B, see Methods for details and Fig. 4.5 and Fig. 4.6 for WM983B data and a more complete statistical analysis). Note that in the purely transient case, one would expect roughly a Poisson distribution of colony counts, but we observed a somewhat super-Poisson distribution. There are multiple experimental parameters that may have introduced additional variance, including variation in plating efficiency between clones, differences in clone proliferation rates prior to drug treatment, and variation in cell motility between clones causing differences in the confluency throughout the dish, as well as the possibility that the pre-resistant state is heritable over small numbers of divisions, but not longer.

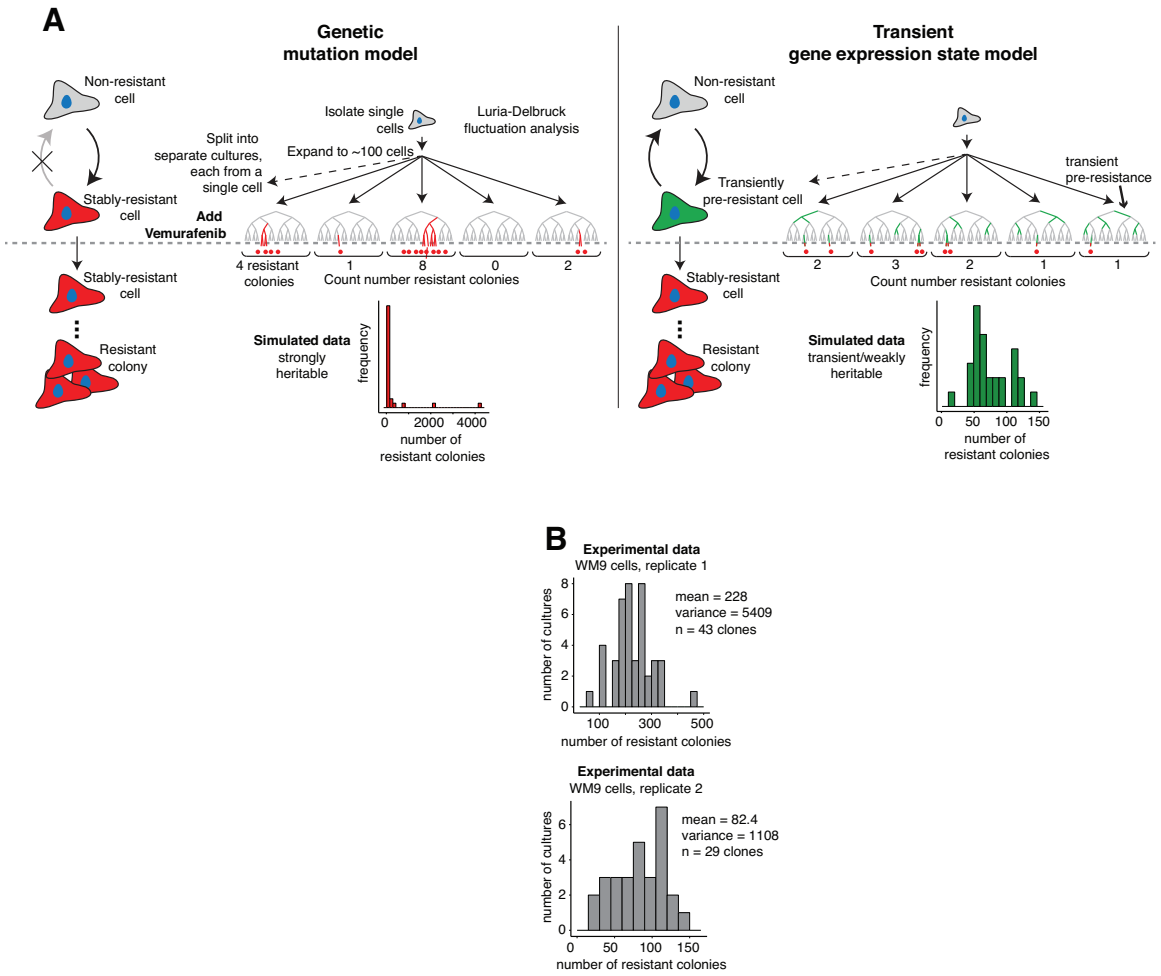

Figure 4.4: Resistance to vemurafenib is not heritable, and pre-existing pre-resistant cells are marked by very high expression of resistance genes. A. Schematic depicting alternative models for heritability of the resistant phenotype and the outcome of the Luria-Delbruck fluctuation analysis given each model. Representative results from simulations shown below. B. Observed distributions of number of resistant colonies in WM9 (biological replicates with sample size = 43 and 29 clones).

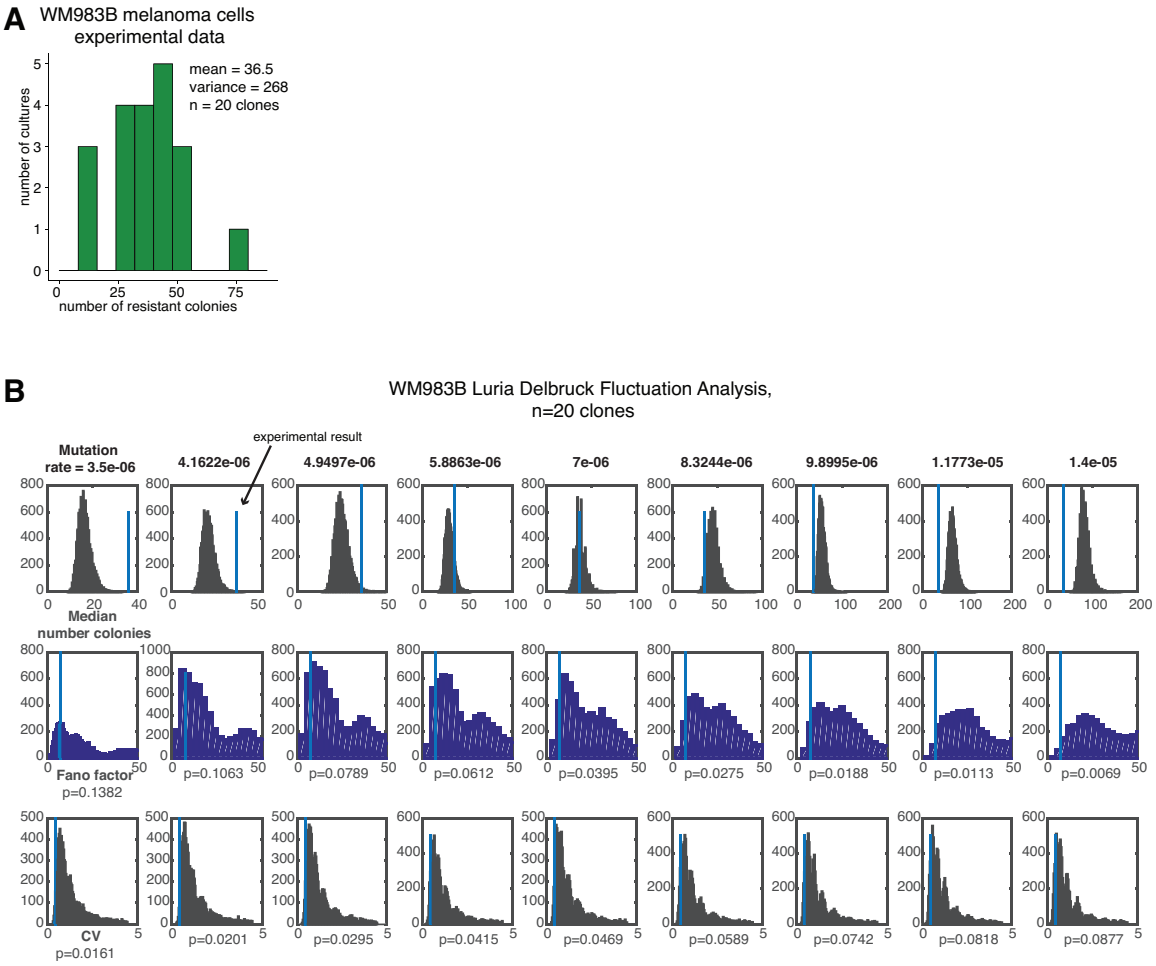

Figure 4.5: Luria Delbruck fluctuation analysis demonstrates that WM983B cells develop drug resistance through a non-heritable mechanism. A. Histogram of the number of resistant colonies from the Luria Delbruck fluctuation analysis in WM983B with a total of 20 clones. B. We simulated the strongly heritable hypothesis for a range of different mutation rates. At each mutation rate, we ran the simulation 10,000 times. We used the parameters specific to this experiment for the WM983B cell line, including the total number of divisions and subsampling of the cultures prior to drug treatment. Each column of plots assumes a different mutation rate which is labeled above. The first row contains histograms of the median number of colonies from each simulation, the second row contains histograms of the Fano factor from each simulation, and the third row contains histograms of the coefficient of variation (CV) from each simulation. In each plot, the value corresponding to our experimental findings are labeled by the vertical line. The p-value to reject the strongly heritable hypothesis based upon the Fano factor or CV at each mutation rate is below the plot.

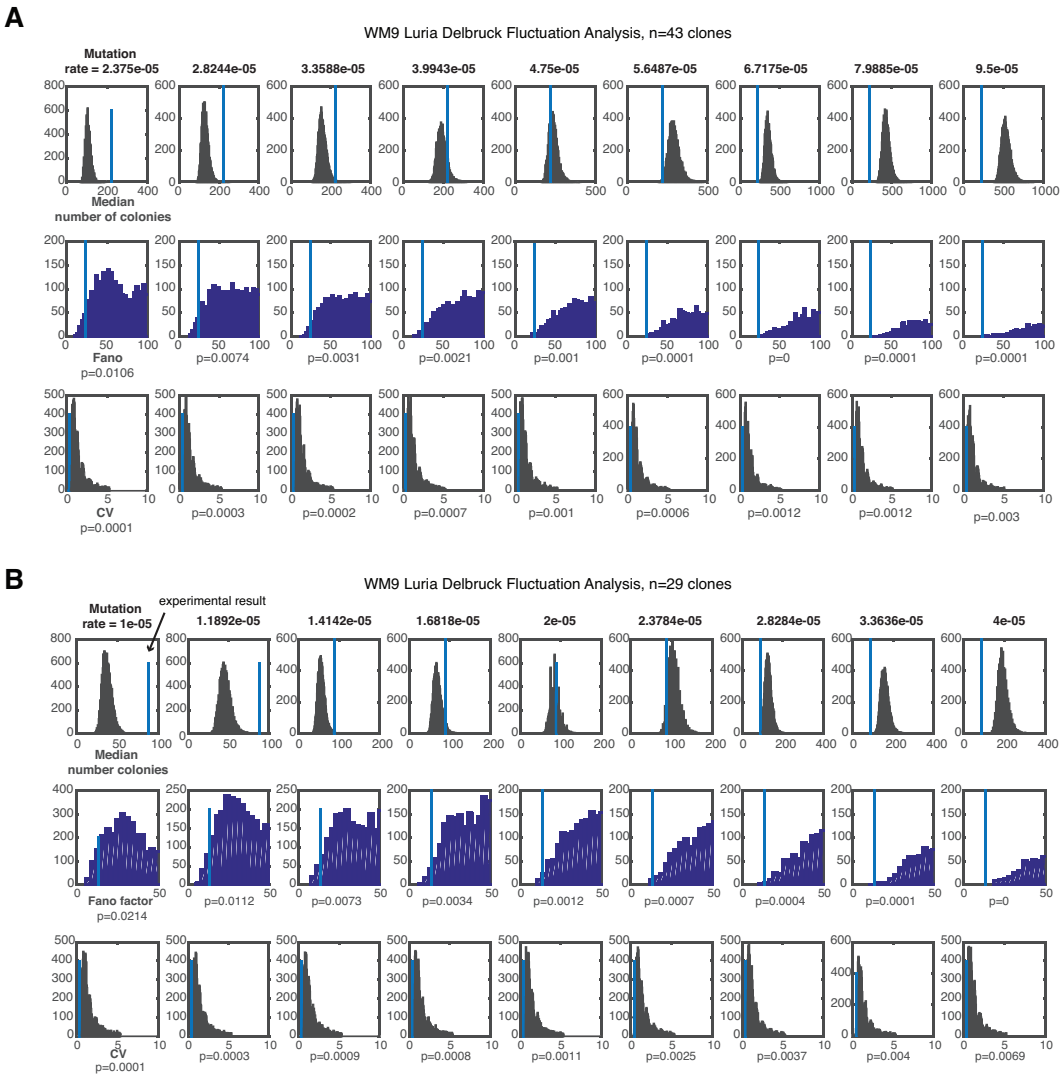

Figure 4.6: Luria Delbruck fluctuation analysis demonstrates that WM9 cells develop drug resistance through a non-heritable mechanism. We performed the Luria-Delbruck fluctuation analysis twice with the WM9 cell line. We simulated the strongly heritable hypothesis for a range of different mutation rates. At each mutation rate, we ran the simulation 10,000 times. We used the parameters specific to this experiment for the WM9 cell line, including the total number of divisions and subsampling of the cultures prior to drug treatment. The plots in panels a and b are from separate biological replicates with a total of 43 and 29 clones. Each column of plots assumes a different mutation rate which is labeled above. The first row contains histograms of the median number of colonies from each simulation, the second row contains histograms of the Fano factor from each simulation, and the third row contains histograms of the coefficient of variation (CV) from each simulation. In each plot, the value corresponding to our experimental findings are labeled by the vertical line. The p-value to reject the strongly heritable hypothesis based upon the Fano factor or CV at each mutation rate is below the plot.

4.3 RNA sequencing of drug resistance

After establishing that pre-resistance to vemurafenib arises in a small, transient subpopulation of melanoma cells, we wondered whether single-cell gene expression differences could underlie the phenotypic differences present in the pre-resistant cells. Since the pre-resistant cells make up only a small portion of the entire population, we expected that population-based gene expression analysis (via RNA sequencing) would not be able to identify the signature of this sub-population. Instead, we hypothesized that some aspects of the signature of stable resistance may already be present in a small, privileged subset of the non-resistant population. To test this hypothesis, we first determined the transcriptional program associated with stable drug resistance in WM9 and WM983B melanoma cells (Fig. 4.7, A and Fig. 4.8, A) via RNA-sequencing. In WM9 and WM983B, we identified a set of marker genes (1456 and 1316 genes, respectively) whose expression significantly increased only in resistant cells and not just upon administration of drug (p-value $<10^{-5}$ and $\log_2$ fold change > 0.5; Fig. 4.8, B). Reassuringly, we found that many well-known markers of drug resistance were specifically associated with resistance, including signaling molecules such as *WNT5A* [1, 81], *AXL* [74] and *EGFR* [113], growth factors such as *PDGFRβ* [76], and transcription factors such as *JUN* [94] (Fig. 4.8, C); furthermore, these signatures are similar to those found in patient samples (Fig. 4.2).

4.4 Single molecule RNA FISH of untreated melanoma cells

Although the expression of these markers was low on average in the untreated population, there could still be rare individual cells with significant expression of these

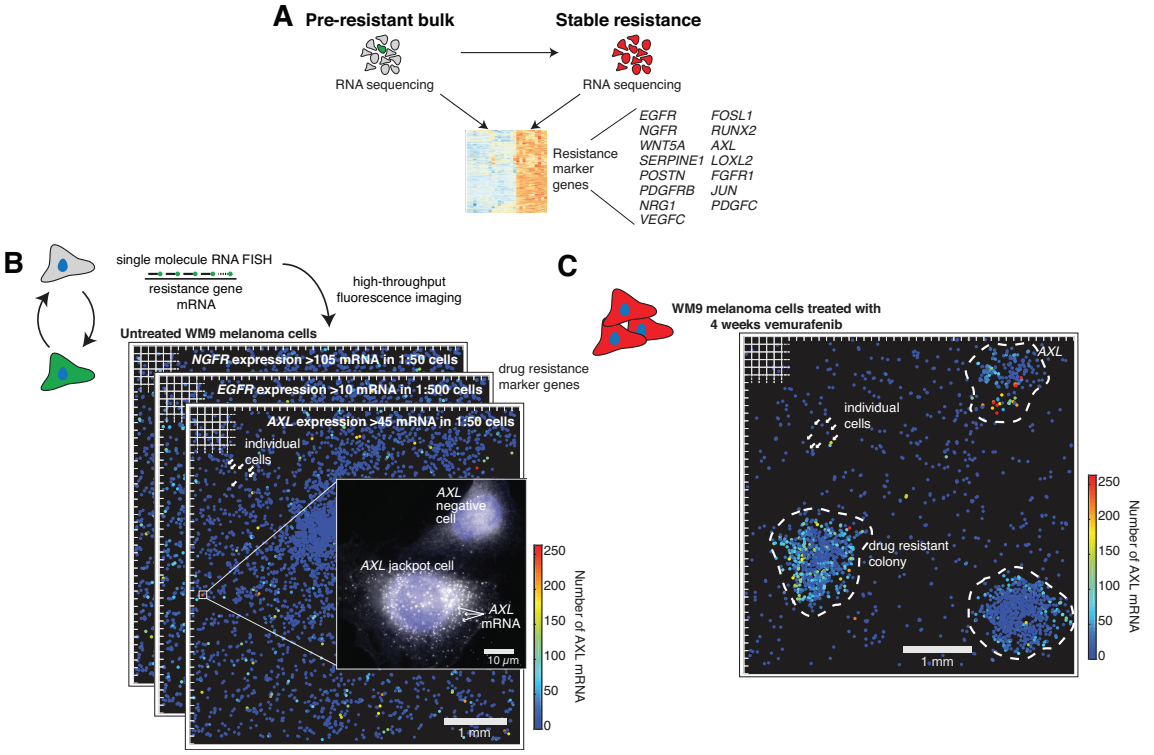

Figure 4.7: Pre-resistant cells are marked by very high expression of resistance genes. A. Schematic of transcriptional profiling experiments. We harvested cells for analysis before adding drug, 48 hours after drug application and then on stably resistant cultures. The heatmap depicts the "marker gene" signature used for follow up analysis in which expression is significantly increased in the resistant cells relative to the untreated. B. Computational representation of single-cell RNA FISH data from 8672 untreated cells shown for *AXL*, *NGFR*, and *EGFR* mRNA levels, with zoomed area showing image data from the indicated subregion. Each cell is represented by a dot on this plot and the color of the dot represents the number of RNA in that particular cell as indicated by the color scale bar. Scale bar on computational representation of image data is 1 mm. Scale bar on image is 10 μm. C. Computational representation of single-cell RNA FISH data shown for *AXL* from 1966 cells after 4 weeks of treatment with 1 μM vemurafenib.

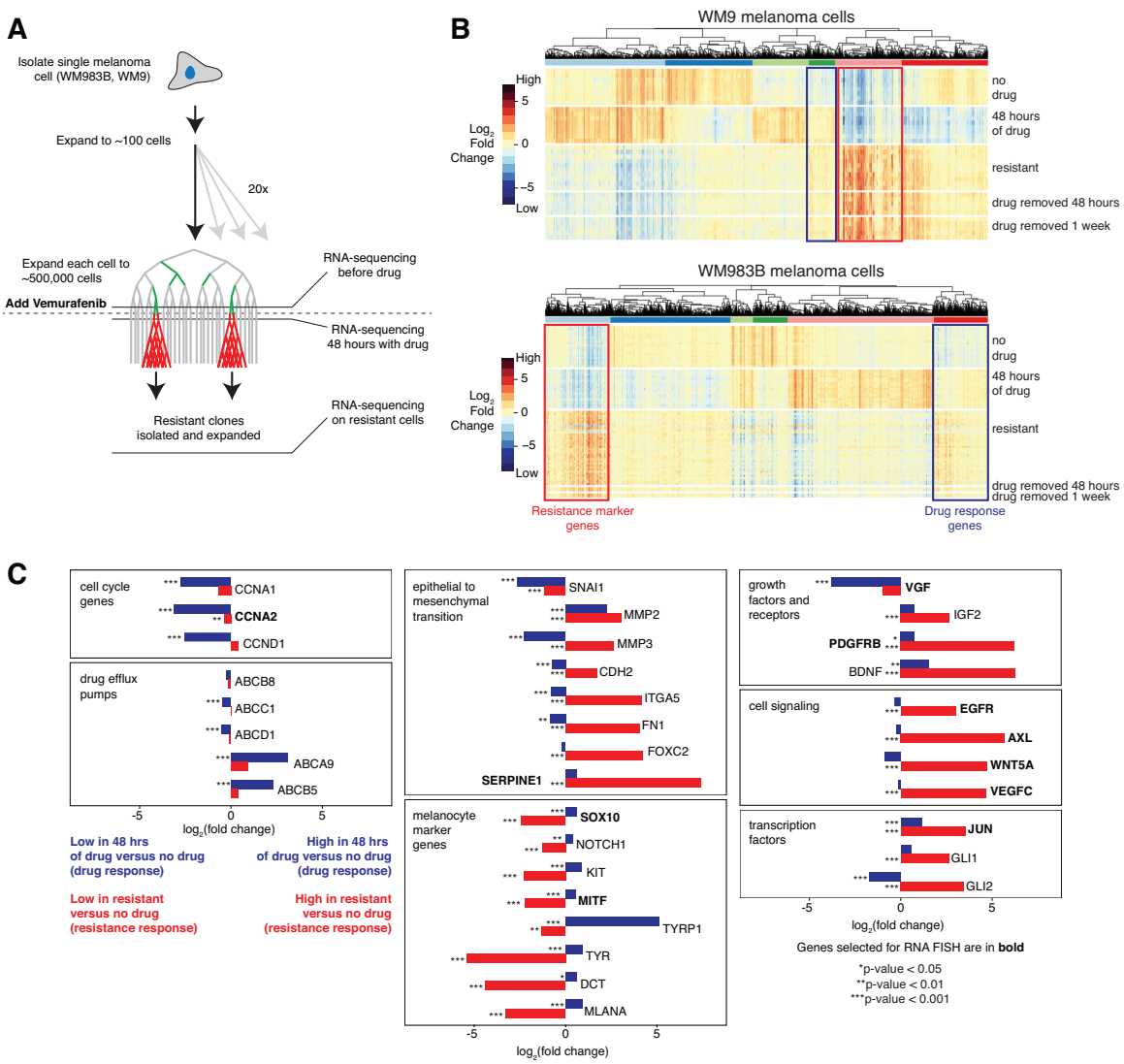

Figure 4.8: RNA-sequencing identifies genes whose expression is specific to resistance. A. Schematic of transcriptional profiling experiments. We harvested cells for analysis before drug, 48 hours after drug application and then on stably resistant cultures. B. Heatmaps depicting expression changes across all differentially expressed genes. Each row represents a separate RNA-sequencing experiment taken from a different Luria-Delbruck subclone. Resistant cultures obtained from subculturing resistant colonies. All genes shown have a greater than 1.4 fold change and adjusted p-value less than 10^{-5} in at least one experimental condition. Color represents $\log_2$ of fold change across the conditions. C. Fold changes in expression in drug response (blue; fold change of 48 hours in drug vs. no drug) and resistance (red; fold change of resistant cells vs. no drug) for WM9 cell line. Bolded gene names are the genes that were selected for analysis by RNA FISH in WM9 cells 4.14. P-values for differential expression analysis are indicated by asterisks as labeled below the plots.

genes. To look for such a population of cells, we needed a strategy to measure with high sensitivity the expression of many resistance genes in many thousands of cells. Thus, we used single molecule RNA FISH and high-throughput fluorescence imaging to capture the expression of a panel of resistance genes in thousands of individual melanoma cells before drug treatment. Surprisingly, we found a subpopulations of cells that expressed very high levels of mRNA resistance genes, while all the other cells in culture expressed very little or none. By counting the molecules of RNA in each cell, we found a population of rare cells (frequencies ranging from 1:50 to 1:500) that express high levels of genes from the resistance program before any treatment with vemurafenib (Fig. 4.7 and Fig. 4.9), and the expression levels in these rare cells was often comparable to that of fully resistant cells (Fig. 4.10, with the exception of *NGFR* and *PDGFRβ*). Note that these cells remained outliers in their expression levels even after normalization by *GAPDH* (Fig. 4.11), suggesting that this rare cell high expression is not a consequence of global extrinsic factors like cellular volume[83]. As expected, after treatment with vemurafenib for 4 weeks, resistant colonies also expressed high levels of these marker genes (Fig. 4.7, C), though, even within a single resistant colony, there was still remarkable heterogeneity in expression.

4.5 Rare "high" expressing cells and drug resistance

The existence of rare cells that express high levels of resistance markers in the population raise the possibility that these cells are the ones that will continue to proliferate despite addition of drug. However, because our single cell RNA FISH analysis required fixing the cells, it was impossible to use the RNA FISH signals to track the cells over time. Thus, we took advantage of the fact that EGFR was

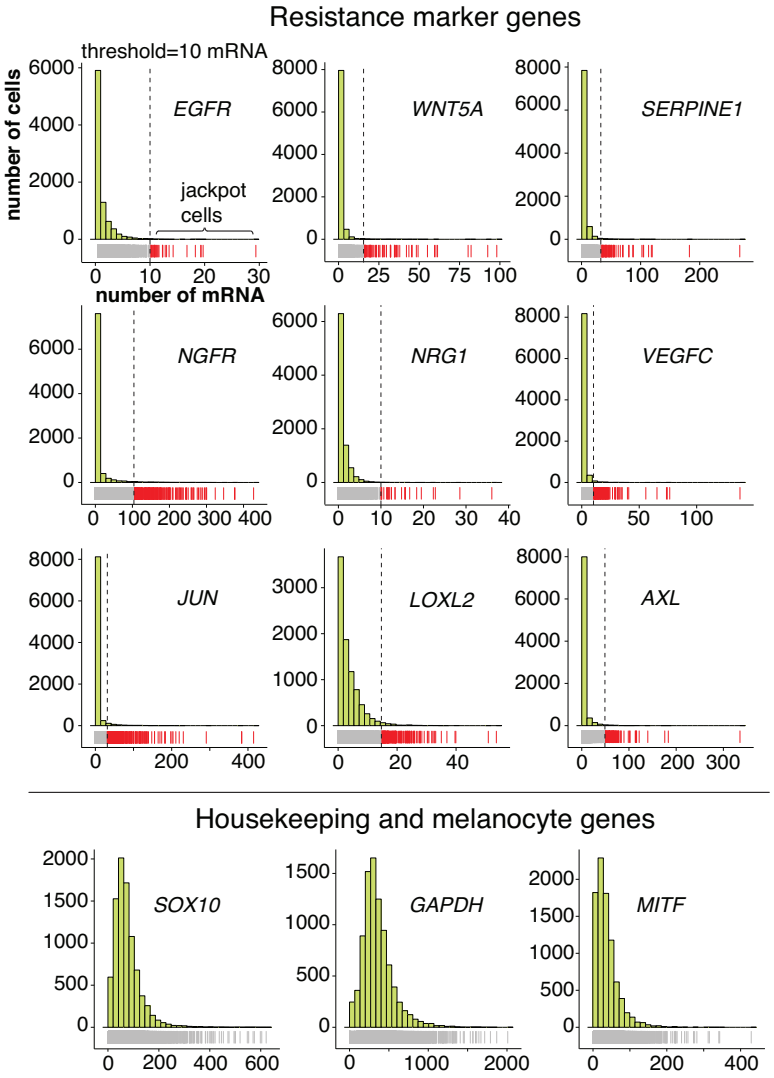

Figure 4.9: RNA FISH on thousands of melanoma cells reveals rare cells that express high levels of resistance marker genes. Histograms of transcript abundance for resistance marker genes (top) and nonresistance markers (bottom). The vertical lines represent the threshold for designating cells as either "high" or "low" expressing for a particular gene. The cells labeled by the red carpet below the histogram are high expressing, and the cells labeled by the gray carpet are low expressing.

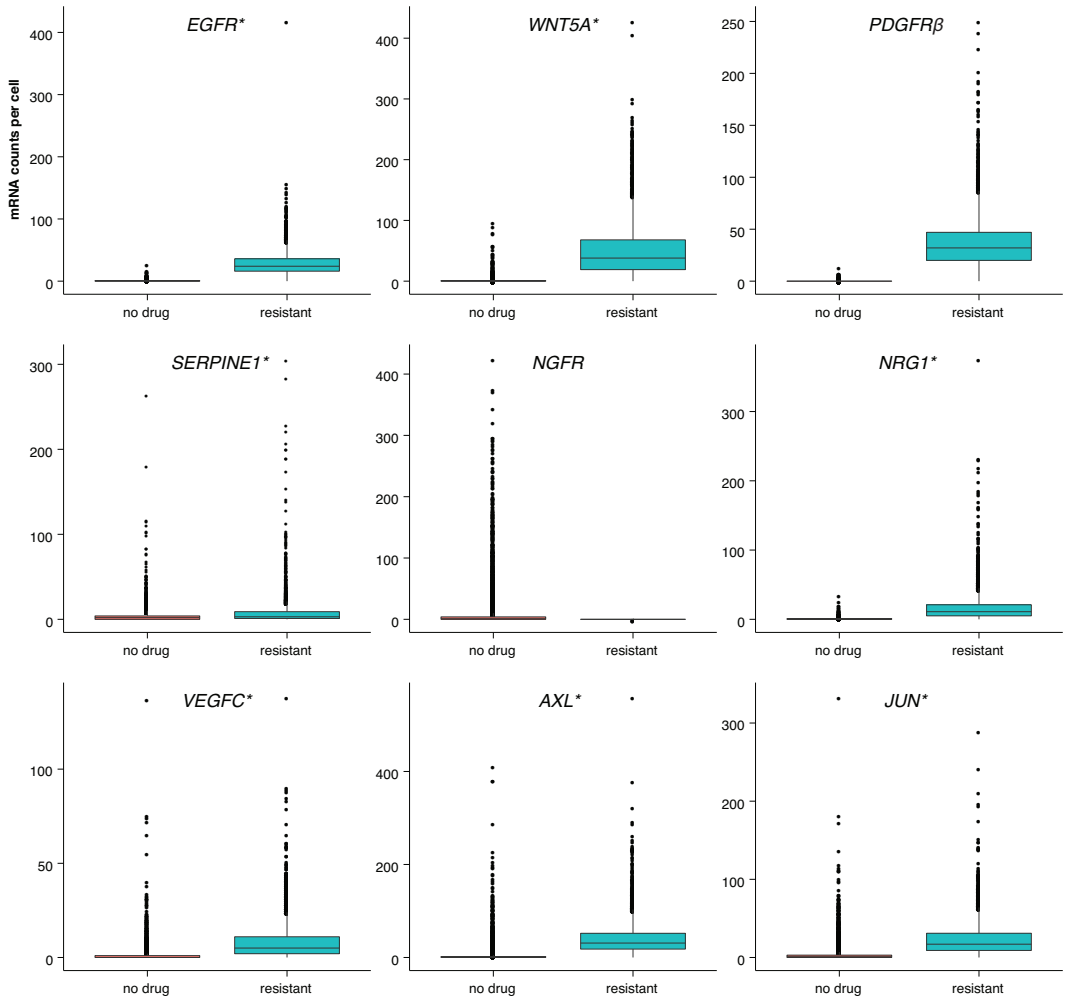

Figure 4.10: In an untreated population of cells, rare cells express resistance marker genes at much higher levels than the population average, sometimes at levels similar to the drug resistant state. Boxplots showing the distribution of mRNA counts per cell for untreated WM9 cells and resistant WM9 cells. The untreated data set is the same data as shown in Fig. **??**, B. For the resistant WM9 cells, we performed iterative RNA FISH with the same panel of genes as in Fig. **??**, B. The untreated data set contains a total of 8672 cells and the resistant data set contains a total of 4082 cells. Asterisks next to the gene names indicates that the max expression of the untreated sample is greater than or equal to the median of the resistant sample, demonstrating that for these 7 of 9 genes, the "high" cells have expression levels potentially equivalent to resistant cells. However, we also point out that given that the sampling of high expressing cells in the untreated samples is low, it is difficult to explicitly compare the distributions to say that the expression in the rare high-expressing cells is equivalent to that in stably resistant cells.

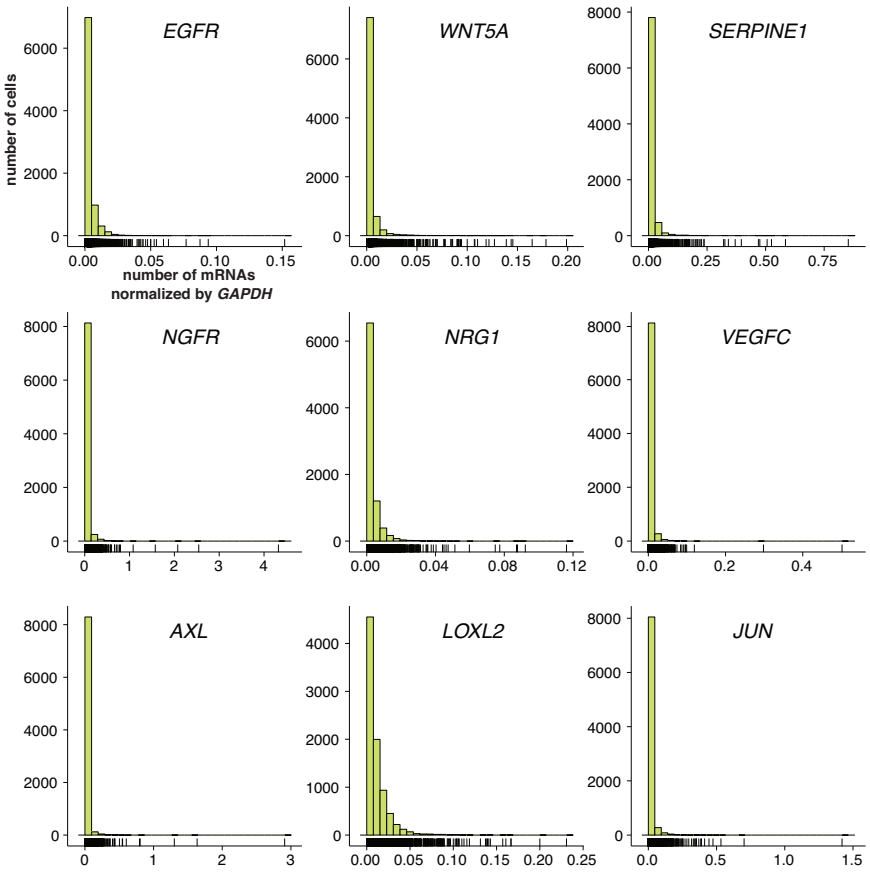

Figure 4.11: Rare cells expressing sporadic but high levels of resistance markers are still present when each gene is normalized by *GAPDH* mRNA counts. Each histogram shows the distribution of *GAPDH* normalized counts for a particular jackpot gene. The counts for each gene in each cells has been divided by the *GAPDH* counts in that same cell. This accounts for any volume-dependent differences between cells. Cells that had *GAPDH* counts less than 50 were dropped from this analysis (these cells were infrequent and gave abnormally high numbers after normalization, thus were dropped). With these cells removed, the data set contains a total of 8477 cells.

one of the genes that displayed the sporadic expression pattern and was also a cell surface receptor for which highly specific antibodies are available, thus allowing us to isolate and track these rare cells over time. Specifically, we stained live WM9 melanoma cells with antibodies targeting EGFR and performed fluorescence-activated cell sorting (FACS) to isolate the top 0.02-0.2% of the EGFR-stained melanoma cells (i.e., EGFR-high cells). After sorting, we waited one day then immediately applied vemurafenib for 3 weeks, and then used a nuclear stain and high-throughput imaging to look for proliferating colonies of resistant cells (Fig. 4.12, A). Strikingly, the EGFR-high population produced 7.9±0.92 (standard deviation) fold more resistant colonies than the EGFR-mixed population (no sorting, thus similar to the initial population, but run through the cell sorting machine as a control) and that these colonies were on average 2.4 fold larger (4.12, A and Fig. 4.13A). This demonstrated that cells exhibiting sporadic expression of EGFR are indeed far more likely to be resistant to vemurafenib treatment, and thus highly enriched for pre-resistant cells. We used RNA FISH to verify that expression of EGFR and other resistance genes (Fig. 4.13, B) was higher in the EGFR-High sorted population. Also, the EGFR-High population also had lower expression of markers including *SOX10* and *MITF* (Fig. 4.13, C), as expected from the previously described relationships between *EGFR*/*SOX10*, *AXL*/*MITF* and *WNT5A*/*MITF*/*SOX10* [1, 27, 74, 113, 126]. Together, these results show that it is the rare cells within the untreated population with high levels of EGFR that are far more likely to become resistant once drug is applied.

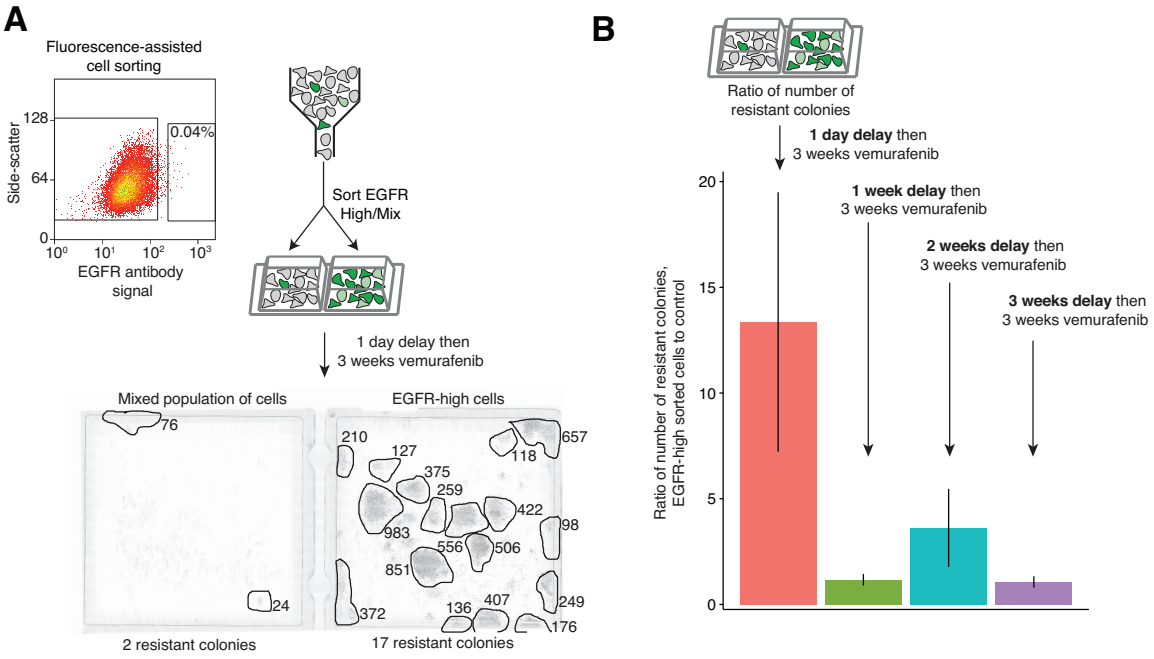

Figure 4.12: EGFR-high cells create more resistant colonies. A. Experimental scheme: we used fluorescence assisted cell sorting of cells labeled with an EGFR-targeting antibody to isolate an EGFR-high cell population and a mixed control, after which we applied vemurafenib. Below is an image of a two-well chamber of EGFR-mix vs. EGFR-high sorted cells after 3 weeks of vemurafenib treatment. Numbers indicate the number of cells in each resistant colony. Total number of resistant colonies is labeled below each well. B. Barplot of the ratio of number of colonies in the EGFR-high well vs. the control mixed well when cells were allowed to grow in culture without drug for varying lengths of time (1 week, 2 weeks, 3 weeks) and applied vemurafenib after the delay. Bars indicate the mean ratio from 3 biological replicates and the error bars represent the standard error of the mean.

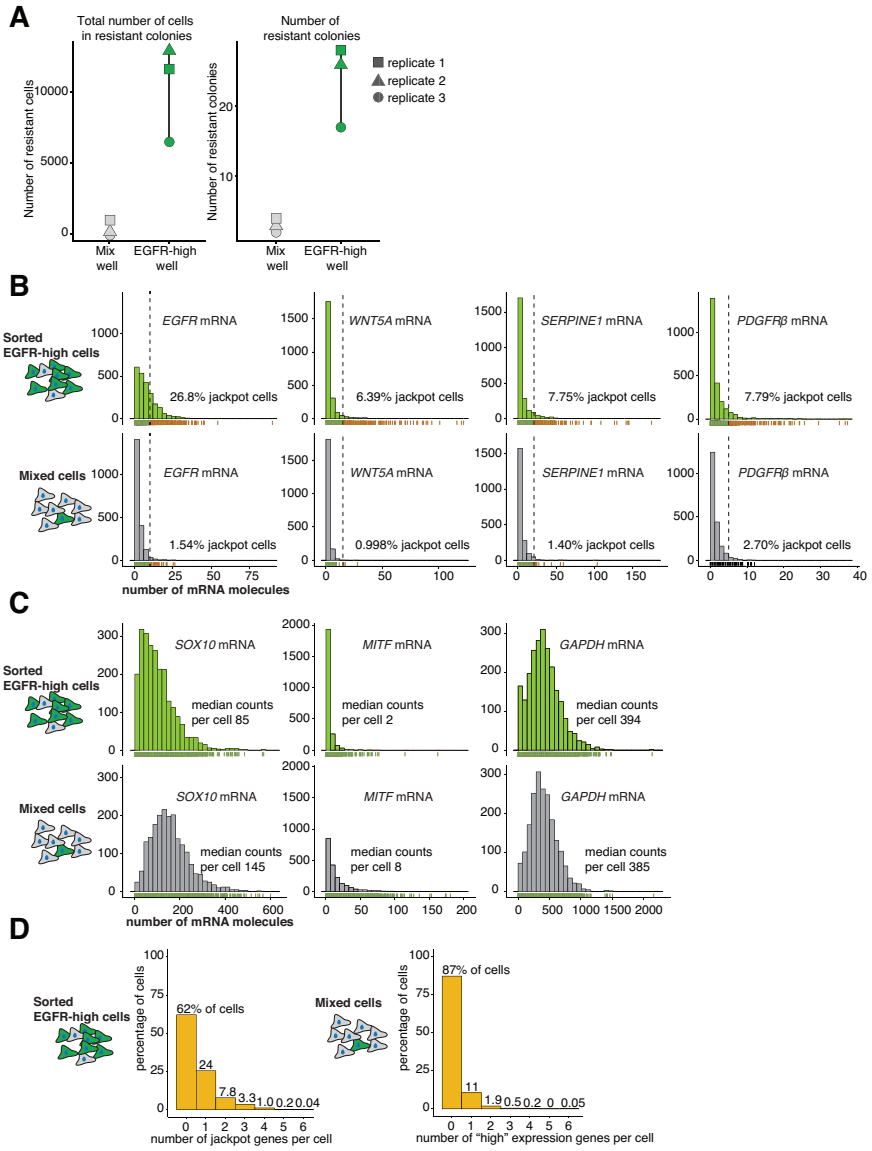

Figure 4.13: Sorting for EGFR-high cells enriches for pre-resistant cells and for cells that express resistance marker genes. A. Quantification of 3 biological replicates of the experiment depicted in Fig. 4.7, D. B and C. Histograms showing the transcript abundance in untreated and FACS sorted EGFR-high and mixed cell populations. The green histograms are from the EGFR-high population and the gray histograms are the mixed population. The percentage of high-expressing cells are labeled on each plot. Panel B shows resistance marker genes *EGFR*, *WNT5A*, *SERPINE1*, and *PDGFRβ*, and panel C shows melanocyte development genes, *SOX10* and *MITF*, and a housekeeping gene, *GAPDH*. D. Histograms of percentage of cells that have high expression of a particular number of genes simultaneously. The left histogram is from the FACS sorted EGFR-high cells, and the right histogram is from the mixed population.

4.6 Heritability of the resistance phenotype in flow sorted EGFR-high cells

Given that the pre-resistant cells are highly enriched in the EGFR-high subpopulation, our Luria-Delbrck results showing that the pre-resistant state is transient would imply that the degree of resistance in this subpopulation would revert to the population average over time in culture without drug. To verify this effect, we stained WM9 melanoma cells with antibodies targeting EGFR and performed FACS to isolate the top 0.02-0.2% of the population. We then allowed the cells to grow in culture without drug for varying lengths of time (1 week, 2 weeks, 3 weeks) and applied vemurafenib after the delay. We found that the ratio of colonies in the EGFR high to the EGFR mixed population significantly decreased from 13.4±10.6 to 1.17±0.47 after 1 week of growth prior to drug treatment, thus demonstrating that the pre-resistant cell state is not heritable on this timescale (Fig. 4.12, B. We did notice, however, that even at the longest time point in culture, 3 weeks, the number of resistant colonies in the EGFR high and the mixed population never became completely equal, indicating that the enrichment for pre-resistant cells did not go away completely, possibly due to differences in paracrine signaling between the cultures magnified once the EGFR-high cells are enriched and brought together in close proximity. Together, these results demonstrate that there is a small subpopulation of pre-resistant cells that continue proliferating when treated with drug, and that this cellular state is transient and associated with the rare expression of resistance marker genes.

4.7 Iterative RNA FISH for expression of 19 genes in the same cells

The finding that rare cells can express high levels of a gene while the remainder of the population had low expression was surprising, and so we wondered how widespread the phenomenon was. We selected a panel of 19 genes for which we performed RNA FISH on up to 40,000 single cells (Fig. 4.14, A). Our panel included a number of housekeeping genes, master regulators of melanocyte differentiation and various resistance markers. In order to measure expression of all of these genes simultaneously in individual cells, we developed a method for iterative RNA FISH using multiple cycles of hybridization: first, we hybridized a set of oligonucleotide probes targeting up to 4 different RNA species, performed large-scale imaging, then chemically stripped these probes off of the sample, then repeated hybridization in the same cells, re-imaged the same cells and so on for 5 rounds [65]. Using computational image alignment, we then assigned expression levels up to 19 genes in our panel to individual cells. (We found that levels of expression were generally comparable whether we probed a gene in the first or last round, showing that iterative hybridization produced reliable results (Fig. 4.15). This provided us with expression levels for these genes in a large number of cells.

4.8 Quantification of "jackpot" heterogeneity with Gini coefficients

To quantify the "jackpot" type of heterogeneity we saw for *AXL*, *EGFR* and *NGFR* in WM9 cells, we used a metric called the Gini coefficient, which was initially developed by sociologist Corrado Gini to describe income inequality, but recently applied to single cell gene expression data [53]. In our context, a Gini coefficient of 0

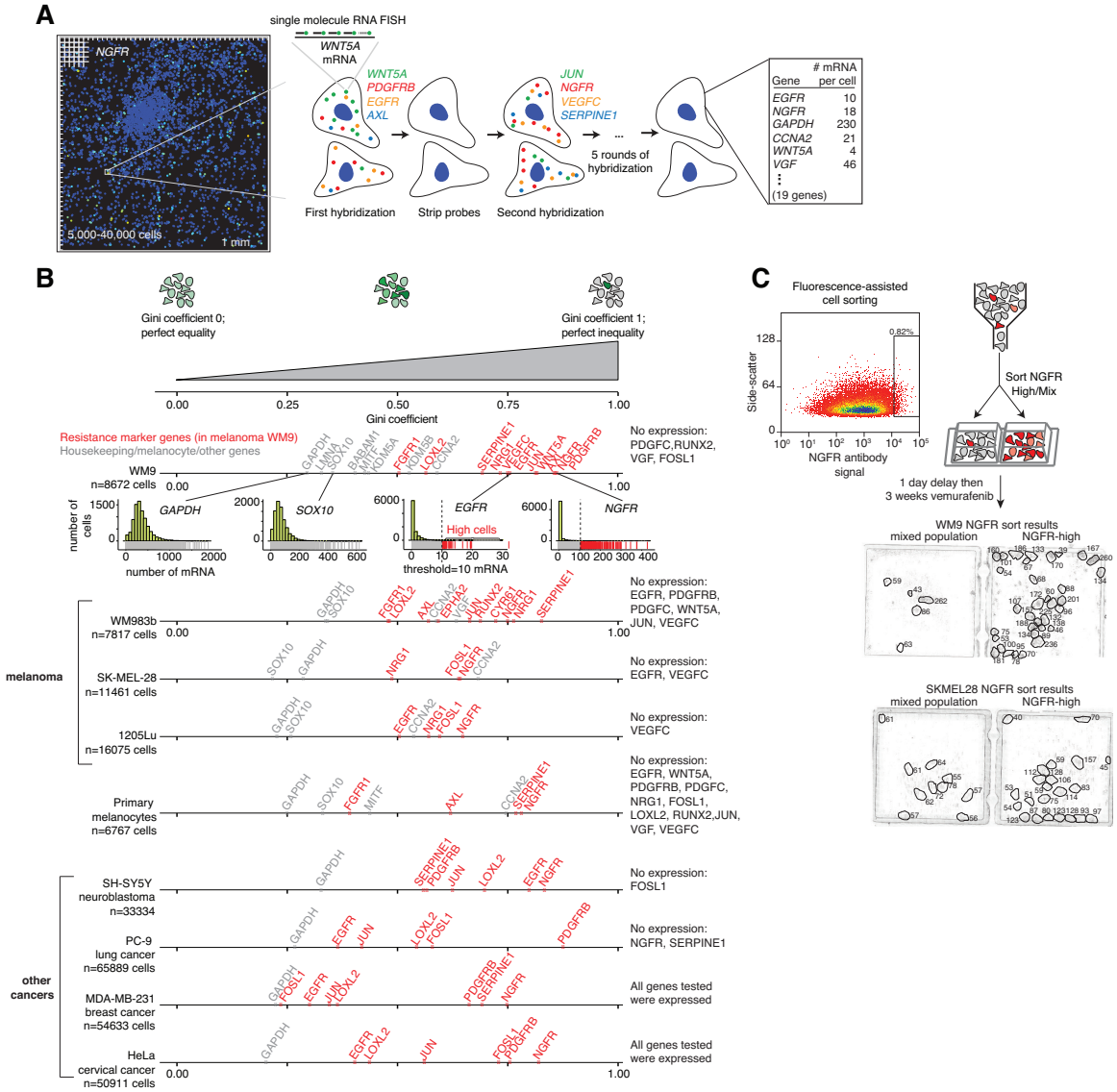

Figure 4.14: Multiplex single cell RNA FISH reveals rare cells with sporadic but intense expression of resistance markers across multiple cell lines and tissue models. A. Schematic of the high-throughput imaging and reiterative hybridization scheme, enabling measurement of 19 genes simultaneously in single cells. B. Gini coefficients capture rare cell population heterogeneity in single-cell RNA FISH data. Gini coefficients plotted for 4 melanoma cell lines (WM9, WM983B, SK-MEL-28, and 1205Lu), primary melanocytes, and 4 other types of cancer cell lines (SH-SY5Y, PC-9, MDA-MB-231, and HeLa). C. We used fluorescence activated cell sorting to isolate NGFR-high cells in WM9 and SK-MEL-28 cell lines, after which we treated with vemurafenib for 3 weeks and then imaged to assay for resistant colonies. Individual resistant colonies are circled and labeled with the number of cells in each colony.

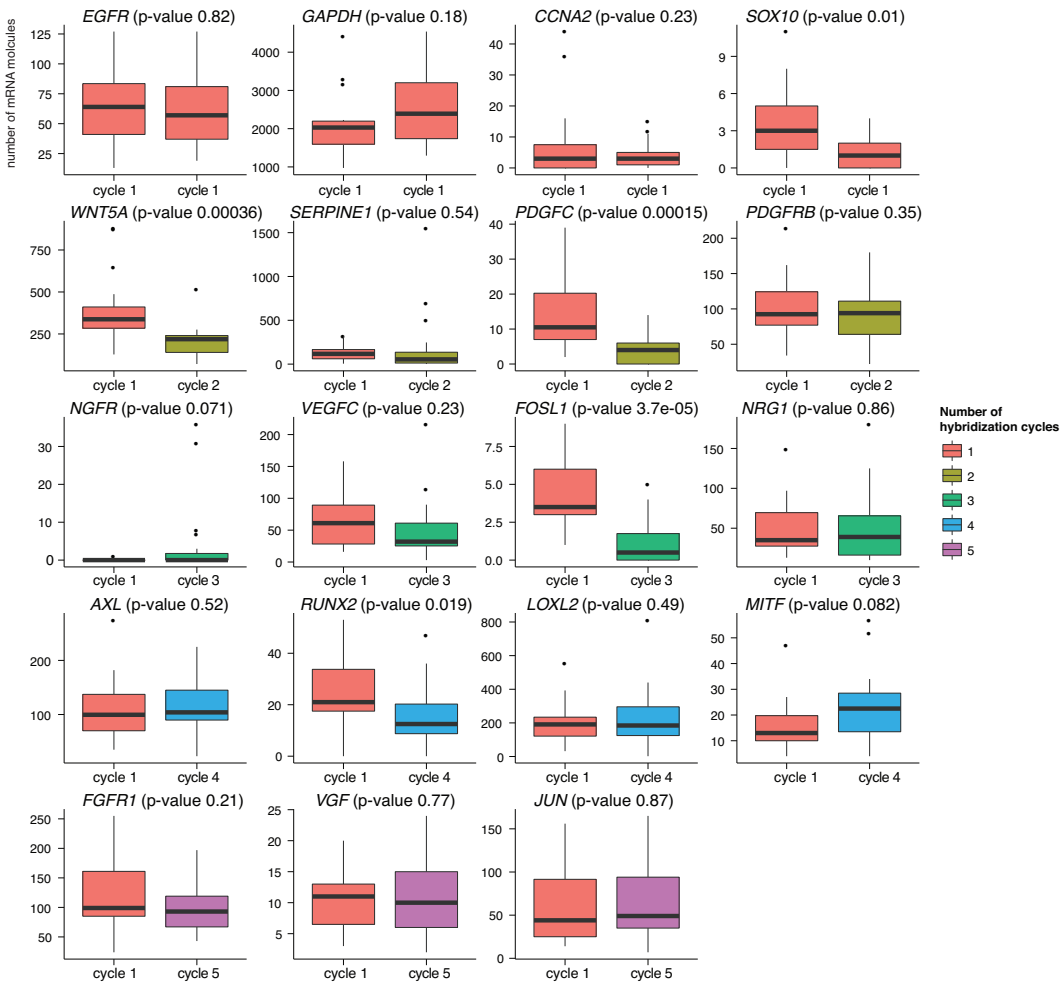

Figure 4.15: RNA counts are consistent whether a gene is probed on the first cycle of iterative RNA FISH or subsequent cycles. Boxplots summarizing RNA FISH mRNA counts for each gene in the 19 gene panel (shown in Fig. 4.14, A). We probed each gene from the panel in resistant WM9 cells without performing iterative hybridizations (note that we used resistant cells because the generally higher expression levels allowed for more robust comparisons). We then performed iterative RNA FISH with all the probes and compared the total mRNA counts. We took image z-stacks of each sample and captured a total 15-25 cells per sample. Expression levels were similar between the first round of hybridization and all subsequent hybridization cycles. The color of the boxplot indicates the hybridization cycle during which we used each probe. The p-value for differences in RNA counts between the cycles are labeled above each plot. Some variability may be due to sampling with genes that have low and/or highly variable expression, and in these instances, we expect some differences in the two count distributions. There is some loss for some genes in later cycles, but we do not believe that affects our qualitative findings of rare, high-expressing cells.

represents perfect equality where all cells have the same number of RNA molecules, while a Gini coefficient of 1 represents perfect inequality where one cells expresses all of the RNA molecules for a particular gene and the others express none (Fig. 4.14, B). On this scale, the resistance genes for which rare cells express high levels of their RNA would have a large Gini coefficient. Performing this analysis on the distributions of 23 genes (this includes the 19 genes in our panel plus 4 additional control genes) from WM9 melanoma cells, we found that 13 genes had a Gini coefficient greater than 0.5, indicating a large degree of inequality in these distributions. As expected, we found that all of our housekeeping genes had Gini coefficients below this 0.5 cutoff and that genes with expected single cell heterogeneity (cell cycle marker CCNA2 and slow-cycling cell marker KDM5B [98, 99]) had Gini coefficients slightly above this cutoff (Fig. 4.14, B).

4.9 Rare-cell high expression behavior in other cell lines

Having established that WM9 melanoma cells exhibit rare-cell expression of these genes, we wanted to know if this phenomenon extended to other melanoma cell lines, to primary cells, or even to other types of cancer. Thus, we first tested a subset of our gene panel in three other melanoma cells lines (WM983B, SK-MEL-28, and 1205Lu) and found that all three of these cell lines harbor rare single cells with much higher levels of expression, creating highly skewed distributions with Gini coefficients as high as 0.82 (*SERPINE2* in WM983B) (Fig. 4.14, B). We next tested primary melanocytes and similarly found rare cells expressing high levels of these genes, demonstrating that this sporadic high expression is not due to some alteration specific to cancer. To test for these rare cells in other types of cancer, we applied a panel of 8 genes (7 resistance

markers and 1 housekeeping gene) to 4 cell lines, each derived from a different type of cancer (SH-SY5Y, PC-9, MDA-MB-231, HeLa), and found that each of these cell lines had some genes with very high Gini coefficients (Fig. 4.14, B). Of note, every cell line we tested exhibited this behavior for at least some of the genes in our panel, suggesting that the rare-cell expression of these genes is not unique to melanoma.

4.10 Rare-cell high expression behavior in tissues

We next wanted to know if rare-cell expression manifested in intact tissues. To evaluate this possibility, we used melanoma tumors from patient-derived xenografts and performed RNA FISH for a subset of the marker genes. By imaging a total of 7791 cells, we found a similar rare cell population expressing *CYR61* and *LOXL2* with Gini coefficients 0.78 and 0.65, respectively (Fig. 4.16). We also found similarly high Gini coefficients in patient-derived xenograft tissue samples created from other melanoma patients (Fig. 4.17).

4.11 NGFR-high cells and drug resistance

Because sorting WM9 cells based upon EGFR staining enriched for pre-resistant cells, we wanted to know if other genes in our resistance panel would similarly enrich for pre-resistant cells in WM9 and other cell lines. To test this, we stained live WM9, SK-MEL-28, and WM983B cell lines with fluorescently tagged NGFR antibodies, performed sorting for the top 1-5% of the population, and then applied 1μM vemurafenib. In WM9 and SK-MEL-28, we found that the NGFR high population of cells was more resistant to the drug (Fig. 4.14, C, replicates in Fig. 4.18, C and D), however, in WM983B, we found that the high population was not more drug resistant (Fig. 4.18, A). Consistent with this result, bulk RNA sequencing of WM983b nonresistant and

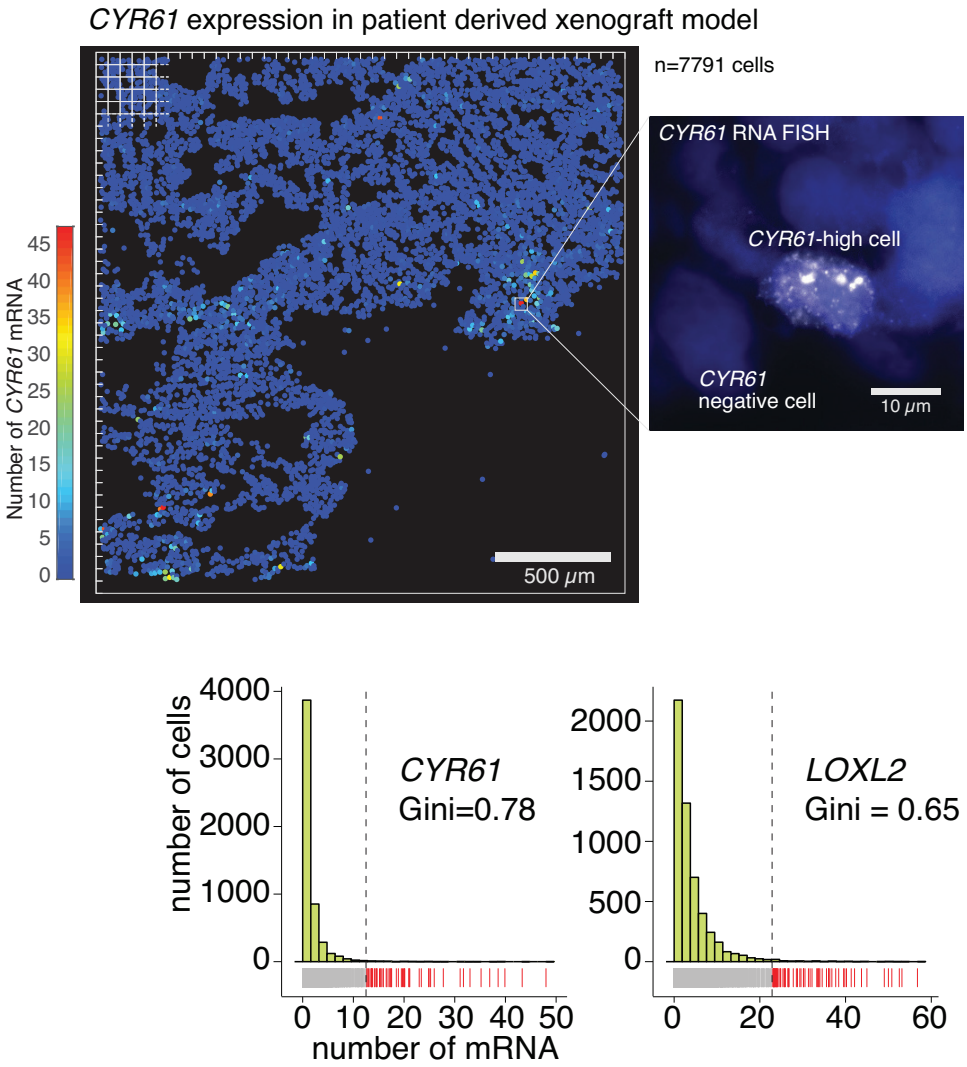

Figure 4.16: "Jackpot" gene expression is present in patient-derived xenografts. Computational representation of *CYR61* mRNA expression in patient-derived xenografts. Each cell is represented by a dot on this plot and the color of the dot represents the number of RNA in that particular cell as indicated by the color scale bar. Image shows an example cell with high expression of *CYR61*. Scale bar on computational representation of the data is 500μm. Scale bar on RNA FISH image is 10μm. Histograms show full distribution of mRNA expression for *CYR61* and *LOXL2* and are labeled with their Gini coefficients.

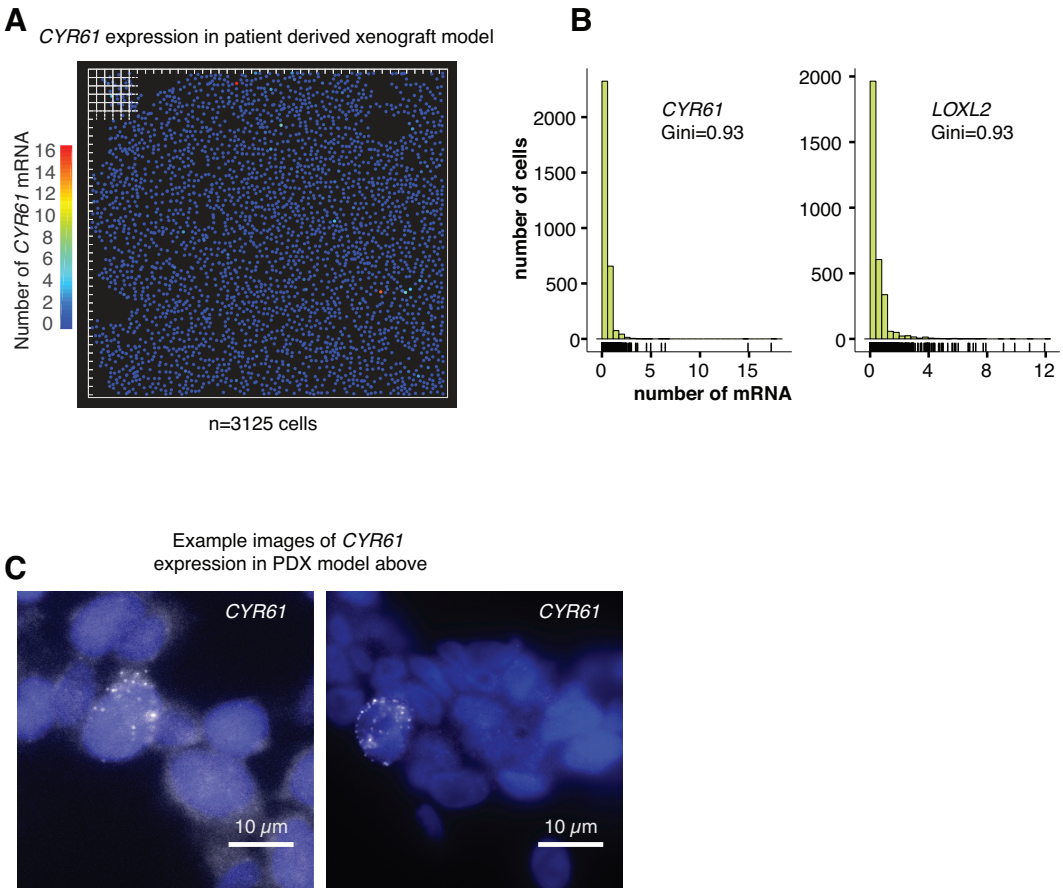

Figure 4.17: Patient-derived xenografts have cells that sporadically express high levels of *CYR61* and *LOXL2*. The samples shown are from a different patient than in Fig. 4.16. A. Computational representation of *CYR61* mRNA expression in patient-derived xenografts. Each cell is represented by a dot on this plot and the color of the dot represents the number of RNA in that particular cell as indicated by the color scale bar. B. Histograms show full distribution of mRNA expression for *CYR61* and *LOXL2* and are labeled with their Gini coefficients. C. Example images of high expression of *CYR61*.

resistant cells revealed that *NGFR* is not a resistance marker (Fig. 4.18, B), confirming that this enrichment for pre-resistant cells is specific to resistance genes.

4.12 Coordination between genes in sporadic high expression events

Given that many of the resistance markers we examined individually showed rare expression patterns and that multiple of them marked pre-resistant cells, we then wondered if these sporadic expression events were coordinated between genes in single cells, and if so, whether this coordination relates to the pre-resistant cellular state. At one extreme, the expression of individual genes could be completely probabilistic and uncoordinated, suggesting that sporadic expression of any one particular gene does not reflect any particular cellular state per se. On the other hand, if the population consisted of two clearly distinct cellular subtypes (e.g., pre-resistant and non-resistant), then one would expect to observe a very strong coordination in sporadic expression between the genes in single cells - i.e., "clustering". In principle, an intermediate between these possibilities could exist as well, in which there is some degree of coordination, but with the coordination showing more of a "continuum" behavior rather than clustering into distinct groups.

To discriminate between these possibilities, we took advantage of the fact that our multiplex RNA FISH assay provided us with mRNA counts for all 19 genes in each individual cell, allowing us to measure coordination in the expression of these genes in single cells. We first examined the pairwise relationships between each of the resistance genes. We set a threshold for labeling each cell as a "high" or "low" expressing for each gene such that the frequencies of "high expressing" cell ranged from 1:50 to 1:500 (example thresholds for *AXL* and *VEGFC* in Fig. 4.19, A). We then determined

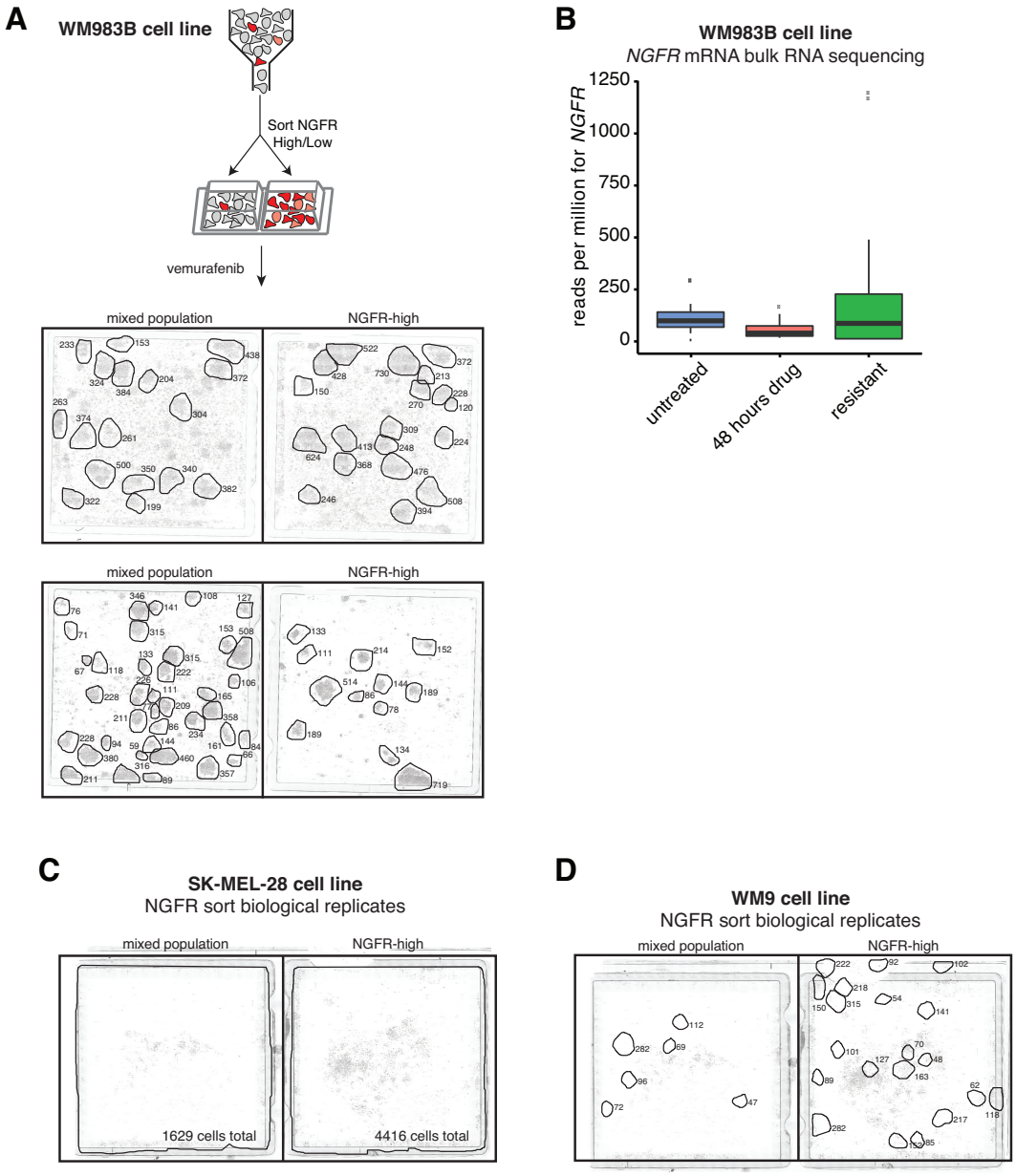

Figure 4.18: Sorting for NGFR-high cells in WM9 and SK-MEL-28 cell lines enriches for pre-resistant cells, but does not enrich for pre-resistance in WM983B. A. We used fluorescence activated cell sorting to isolate NGFR-high cells in WM983B, after which we treated with vemurafenib and then imaged to assay for resistant colonies. Individual resistant colonies are circled and labeled with the number of cells in each colony. B. As described in Fig. 4.7, A and Fig. 4.8, we performed RNA sequencing of untreated WM983B cells, WM983B cells after 48 hours of vemurafenib, and WM983B cells from resistant colonies. Boxplots show the reads per million for *NGFR* in each of the three conditions. C. Replicates of the data in Fig. 4.14, C. These RNA-seq results show that *NGFR* is not a resistance marker for WM983B, which may explain why sorting by *NGFR* does not appear to enrich for resistant cells in that cell line.

the odds ratio between the high and low states for each gene and found that the odds ratios between resistance genes ranged from 4 to 142, which strongly argues against the completely random scenario as these genes have much higher co-expression than expected (Fig. 4.19, B). Moreover, we found two general groups of co-expressing genes: the first consisted of housekeeping genes and melanocyte differentiation factors (*GAPDH*, *SOX10*, *MITF*), and the second consisting of resistance marker genes (*JUN*, *WNT5A*, *NRG1*, etc.) (Fig. 4.19, B). These patterns did not depend on the specific thresholds chosen for determining whether a particular gene was on or off (Fig. 4.20). Principal component analyses of co-expressing genes, both with this panel and another panel of 4 genes including two housekeeping and two resistance markers, revealed two axes of co-variation, one corresponding to housekeeping genes and one corresponding to resistance markers, thus showing that the co-variation shown in the resistance markers is not merely reporting on generic correlations between any pair of genes (Figs. 4.21 and 4.22).

4.13 Double sorting for EGFR-high and NGFR-high cells

Using these thresholds for identifying high expressing cells, it is clear that the frequency of expression for any one of the resistance marker genes is much higher than the frequency of resistance (1:50 to 1:500 for any one marker gene versus 1:2000 for resistance). However, given that cells expressing multiple marker genes exist, we hypothesized that these cells may be the ones that ultimately go on to form the resistant colony. To test this possibility, we stained live WM9 cells with antibodies labeling both NGFR and EGFR and performed FACS to isolate four populations: cells negative for both stains, cells positive for EGFR only, cells positive for NGFR only,

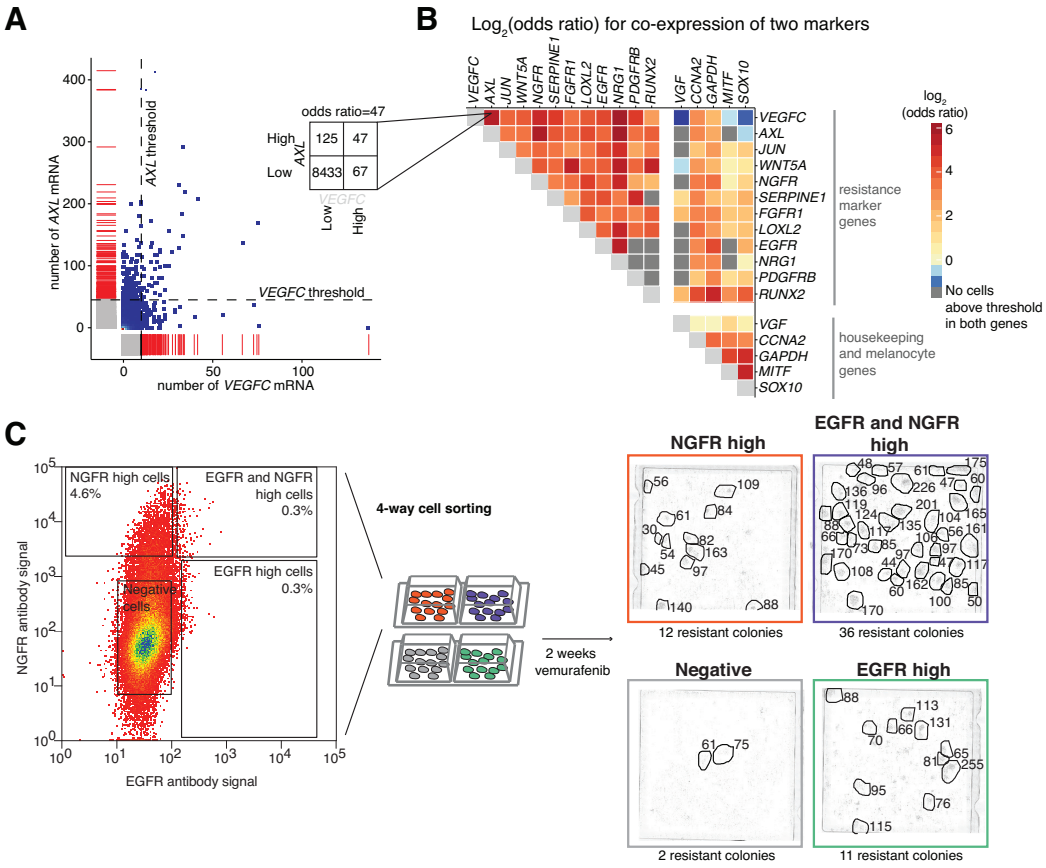

Figure 4.19: Rare cell expression of resistance marker genes is coordinated between genes such that an unexpectedly high number of cells express multiple markers. A. Scatter plot of the number of AXL vs. VEGFC mRNA in individual WM9 melanoma cells. Dotted lines represent the threshold for calling a cell "high" or "low" for each marker. Inset shows the number of cells falling into each quadrant of the plot and is labeled with the odds ratio for co-expression. B. Heatmap shows the odds ratio for co-expression between all pairs of genes. Dark gray boxes label pairs where there were zero cells with counts high expression threshold for both genes. C. Experimental scheme: we co-stained for EGFR and NGFR and then sorted the 4 populations: cells negative for both stains, cells positive for EGFR only, cells positive for NGFR only, and cells positive for both EGFR and NGFR. After sorting, we applied 1μM vemurafenib for 2 weeks and then imaged the samples. Individual resistant colonies are circled in the images and the number of cells in each colony is labeled. The total number of resistant colonies in the entire sample is labeled below each image.

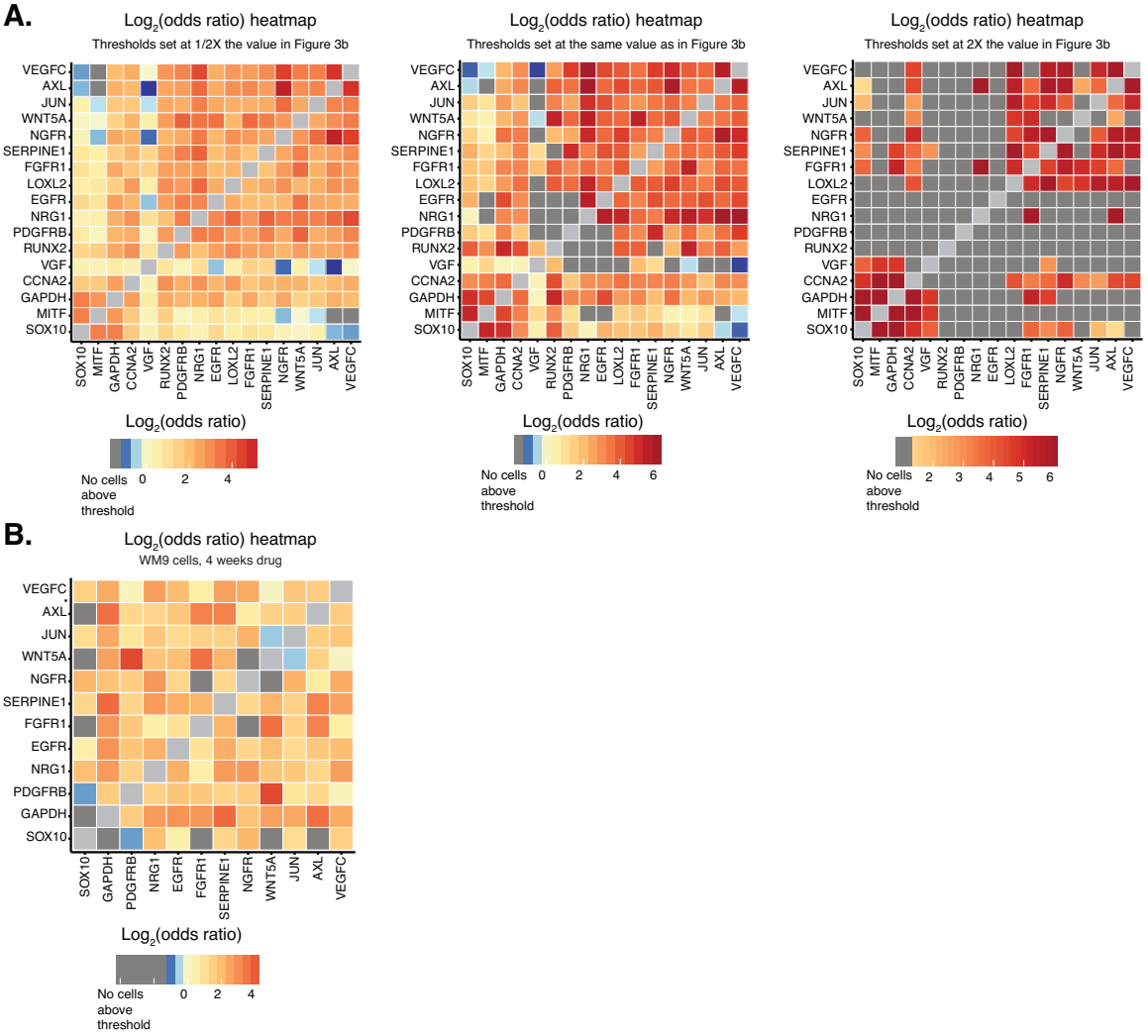

Figure 4.20: Odds ratio between high and low states does not change significantly with different thresholds. A. Heatmap shows the odds ratio for co-expression between all pairs of genes. Dark gray boxes label pairs where there were zero cells with counts high expression threshold for both genes. The heatmap in the middle has the same thresholds for designating cells as "high" or "low" as used in Fig. 4.19, B. Meanwhile, the heatmap on the left shows the same analysis with the thresholds set to 1/2X of the their value in 4.19, B and the heatmap on the right shows this analysis with thresholds set to twice their value in Fig. 4.19, B. When the thresholds are at 1/2X, the result is very similar to that in Fig. 4.19, B. However, increasing the threshold by 2X leads to many gene pairs that do not have any cells that are "high" for both genes (indicated by the dark gray boxes). B. Heatmap showing odds ratios for WM9 data after 4 weeks in drug.

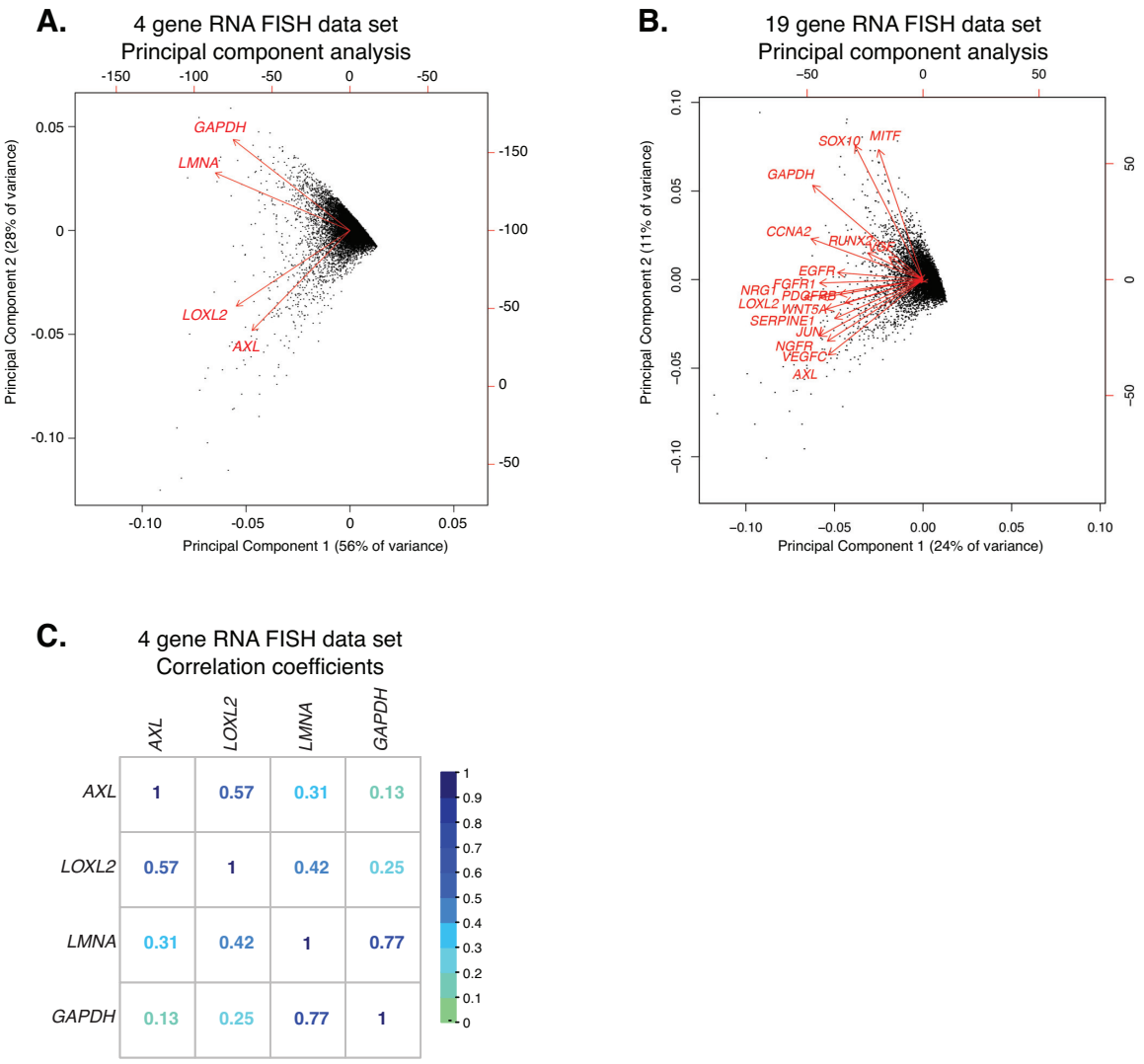

Figure 4.21: Housekeeping genes correlate more with each other than with resistance markers and vice-versa. We performed RNA FISH on 8672 non-drugged cells with probes targeting *LOXL2* and *AXL* (both of which exhibit rare-cell expression) and *LMNA* and *GAPDH*, both of which are control genes not associated with resistance. A. We then performed principal component analysis to determine which genes covary with which other genes. We transformed the vector representing the expression levels of each cell into the space spanned by the first two principal components. Arrows represent transformations of unit vectors of the specified gene into this same space. We observed two rough axes of variability, one corresponding to the *GAPDH* and *LMNA* and the other to *AXL* and *LOXL2*. Thus, these results show that there is substantial covariation in housekeeping genes and in resistance markers, but that these two axes of variation separate. B. Same plot as in panel a, but with the RNA FISH data shown for WM9 in Fig. 4.14, B. C. Pairwise correlations between the genes in A.

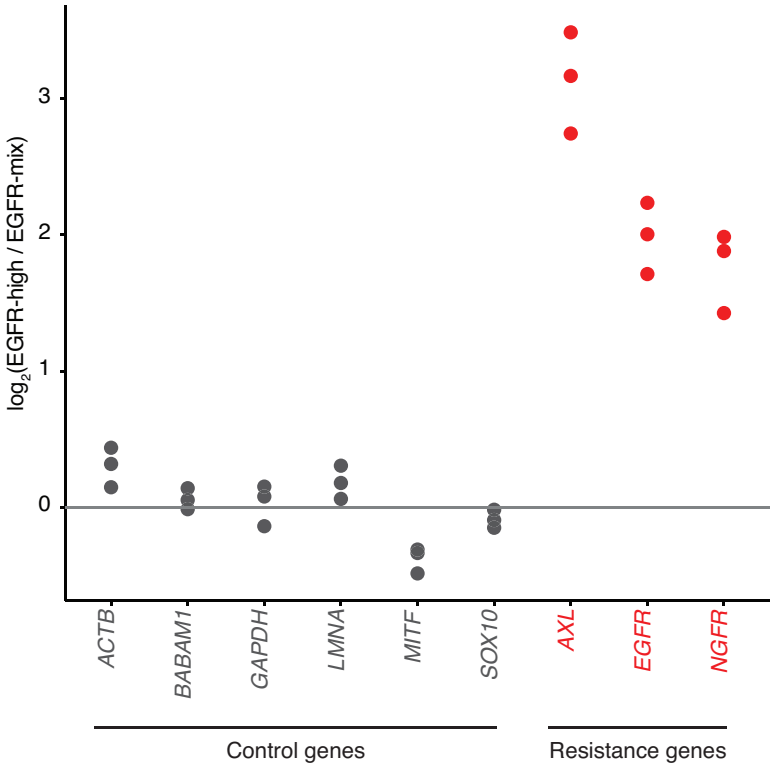

Figure 4.22: Control genes do not show as much enrichment in the EGFR-high subpopulation as the pre-resistance marker genes. We sorted by EGFR antibody to isolate the EGFR-high subpopulation of cells and then performed RNA sequencing on these populations as well as an EGFR-mixed population. Dot plots show the $\log_2$ fold change in gene expression for a set of control genes and a set of resistance marker genes. Each dot represents a separate biological replicate (paired EGFR-high/EGFR-mixed). The horizontal line at y=0 represents no change in the EGFR-high samples relative to the EGFR-mix. For the resistance marker genes (*EGFR*, *AXL*, and *NGFR*), there is significantly more expression in the EGFR-high sample, while the control genes do not show large differences, showing that they do not correlate with the expression of the resistance markers.

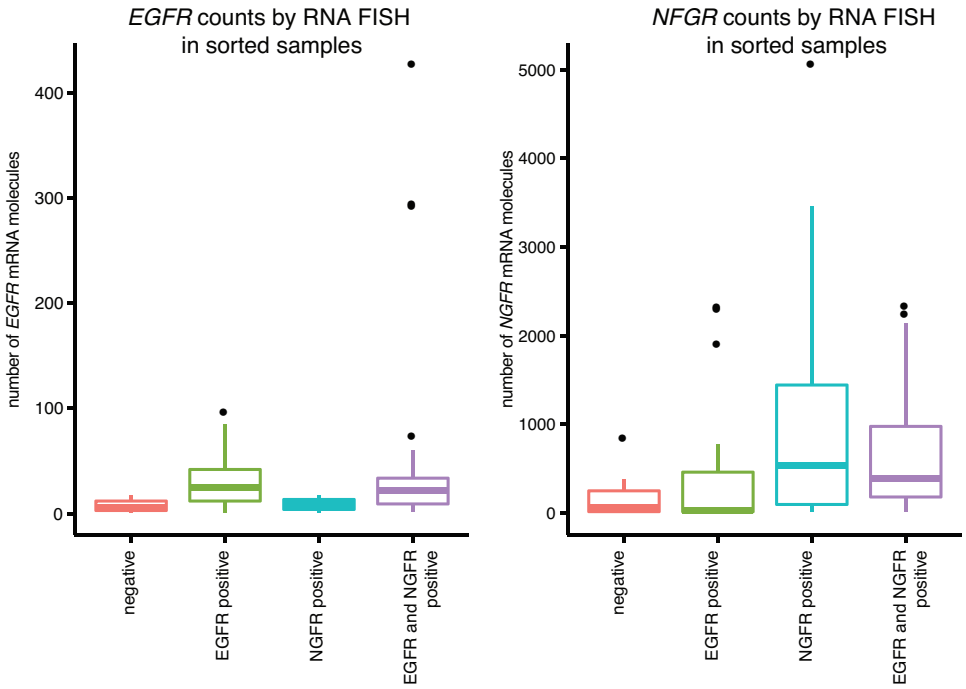

Figure 4.23: Boxplots summarize the single-cell RNA FISH counts for *EGFR* and *NGFR* in flow sorted populations shown in Fig. 4.19, C. These results show that sorting the high populations indeed enriches for *EGFR* and *NGFR* mRNA, thus validating the sort procedure. Furthermore, it shows that the double sorting does not further enrich for either *EGFR* or *NGFR* mRNA alone, showing that the effects of the double sort do not arise from a further enrichment of either *EGFR* or *NGFR*-high cells per se, but rather the combination of both in the same cell.

and cells positive for both EGFR and NGFR. After sorting, we applied vemurafenib for 2 weeks and then imaged the four different cultures. We found that both the EGFR only and NGFR only samples formed more resistant colonies than the negative sample (with 11 and 12 versus 2 in the negative), but that the double positive sample formed the most resistant colonies with 36 total (Fig. 4.19, C). We confirmed by RNA FISH that these sorted populations have the expected expression of *EGFR* and *NGFR*, thus ensuring that the FACS isolated cells are the same populations that we observed by RNA FISH (Fig. 4.23). This finding shows that the cells with high expression for more than one marker gene are the ones that develop into the resistant colonies.

These results further demonstrate that high expression of any one pre-resistant marker does not perfectly correspond to being resistant, suggesting that while there is clearly coordination between markers, the gene expression signature itself may not separate the cells into two clearly distinct states. To explore this possibility further, we next searched for higher order correlations in expression of these resistance markers. A priori, however, finding such cells with high expression of multiple markers would seem implausible if the expression occurred independently of each other. Given a probability of high expression of 1% for a particular marker gene, the probability of a cell expressing 6 of these markers would be essentially zero ($0.01^6 = 10^{-12}$). We found a surprisingly large number of cell expressing high levels of multiple markers, with 13 cells out of 8672 expressing 6 markers simultaneously and 2 expressing as many as 8 markers simultaneously (example images in Fig. 4.24, histograms of higher order coordination Fig. 4.25) - notably, these numbers are close to the expected frequency of resistant colonies. These numbers suggest a strong degree of interdependence between the transcriptional activity of these genes, but a cluster analysis showed that there are no readily discernible clusters, suggesting that the intermediate "continuum" model may fit our data most readily. This continuum is not completely random, and the correlations can reveal potential network structure between the genes interrogated (Fig. 4.26).

4.14 RNA sequencing of resistance reprogramming

Taken together, our results thus far show that jackpot cells with high levels of transcription of resistance markers are the ones that will proliferate when treated with drug, and that this pre-resistant cellular state is transient. At the same time, however,

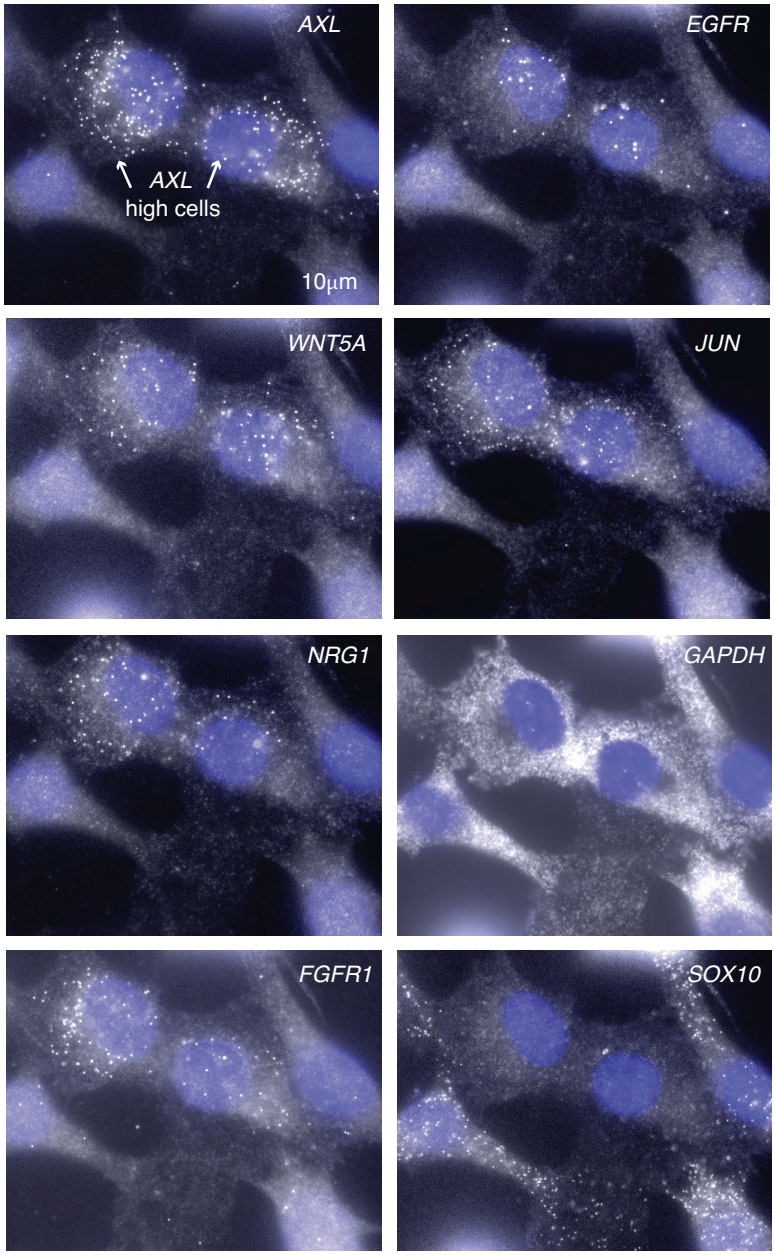

Figure 4.24: Example images of two cells with high expression of 6 resistance marker genes (*AXL*, *EGFR*, *NRG1*, *WNT5A*, *JUN*, *FGFR1*), normal expression of housekeeping gene *GAPDH*, and low expression of melanocyte gene *SOX10*. RNA FISH signal is white, and cell nuclei are shown in blue. Scale bar is 10μm.

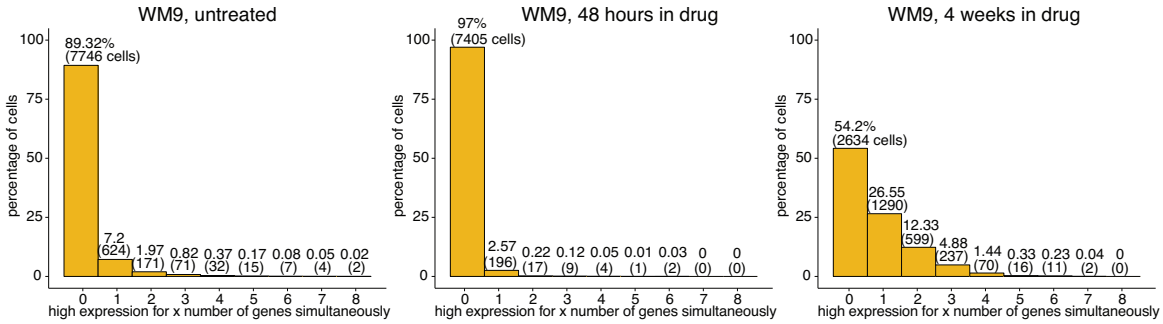

Figure 4.25: There are subpopulations of cells that have high expression of multiple resistance marker genes. Histogram of number/fraction of cells that have high expression for a particular number of genes simultaneously, both before, immediately after and then 4 weeks after application of drug. We found that immediately after adding drug, there was a large general decrease in the amount of high-expressing cells, but a few cells remained that expressed several marker genes at once. This suggests, but does not prove, that these multi-expressing cells may be the pre-resistant cells.

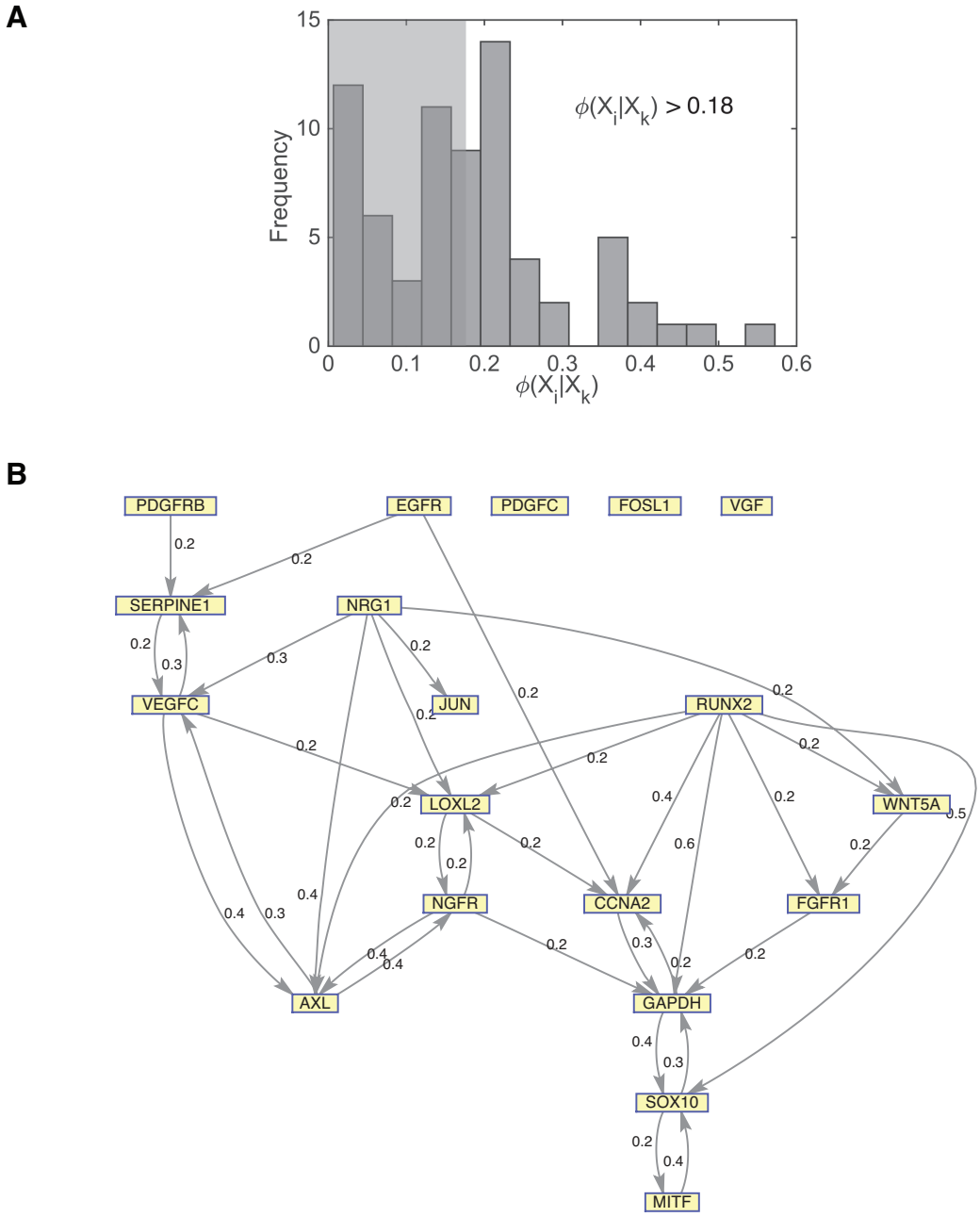

Figure 4.26: Phixer analysis reveals the network structure of rare-cell expression. A. Histogram of the ϕ-mixing coefficient (edge strength) for all edges in the inferred network for melanoma undrugged cancer cells. To illustrate the network we select the 34 most strongest edges (non-shaded portion), and this corresponds to selecting edges with $\phi \geq 0.18$. B. Gene interactions obtained using the phixer algorithm applied to the single-cell RNA FISH data from cancer cells. Each directed edge and its corresponding strength (ϕ-mixing coefficient) quantifies the effect of an upstream gene on the probability of rare-cell expression of a downstream gene.

our time-lapse videos clearly show that once drug treatment begins, the clusters of resistant cells appear to grow stably. This leads to a conceptual paradox: if the resistant state is transient, then how would a colony of resistant cells form? Wouldn't the colony disintegrate as the cells transition back to being non-resistant? Our observations indicate that colonies do not disintegrate, and indeed, the transcriptome profile of resistant cells given a "drug holiday" appear to remain at least partially stable for 3-5 weeks (Fig. 4.8, B), indicating that the cells are at that point maintaining the resistance phenotype. One possible resolution to this paradox is to postulate two separate stage of resistance. The first, demonstrated above, is the generation of transiently pre-resistant cells, characterized by rare but intense transcriptional events. The next stage, we hypothesize, occurs beginning with the addition of the drug itself, which induces a transition from the pre-resistant state to a stably resistant state, thus "burning in" the resistant phenotype through a presumably epigenetic change. Such a two-staged process would be in many ways reminiscent of cellular reprogramming during the induction of pluripotent stem cells [43].

To determine whether the addition of drug led to a reprogramming of the pre-resistant cells, we took advantage of the ability to sort cells by EGFR protein expression levels to isolate pre-resistant cells from the bulk population before adding drug and then their progeny at time points (1 week and 4 weeks) after the addition of drug (Fig. 4.27, A). In a pure selection model, one would expect the profile of pre-resistant cells to be largely the same as the resistant cells; however, in the reprogramming model, one would expect a much more progressive transformation from the pre-resistant cells to the stably resistant state. To discriminate these possibilities, we profiled the transcriptomes of the EGFR-high cells before and after addition of drug, focusing on the set of resistance markers identified by our bulk RNA-sequencing (Fig. 4.7, A), and determined the percentage activation of each resistance gene (using the original EGFR-

mix population as a baseline) throughout the time course. Interestingly, we found that the pre-resistant cells only express a very small fraction (72 of 1456) of the resistance transcriptome near the full level of activation in resistance (Fig. 4.27, B). (As expected, 8 of the 10 genes for which we observed sporadic expression patterns were more highly expressed in the EGFR-High cells than in the EGFR-mixed population; Fig. 4.28) This finding is consistent with our conclusion that the pre-resistant state is a relatively shallow departure from the non-resistant state, and supports the reprogramming model. Furthermore, we confirmed that the EGFR-high cells were proliferating based upon their relative expression of cell cycle genes *CCNA2* and *CCND1* (Fig. 4.28) (which is a requirement for the Luria-Delbruck fluctuation analysis in Fig. 4.4). We also note that the EGFR-high cell subpopulation is molecularly distinct from the slow-cycling subpopulations that express other markers such as *KDM5A* and *KDM5B* [98, 107] (Fig. 4.29).

In stark contrast, after adding drug, the percentage of resistance genes expressed was much higher and increased over time. We found that by week 1 in drug 600 of the 1456 total resistance genes were activated to >80% and that by week 4 this activation increased to 966 of 1456 resistance genes, thus demonstrating a progressive transformation of the transcriptome in resistant cells as they become stably resistant. To validate these expression patterns directly in single cells, we performed RNA FISH on *TXNRD*1, a gene that is not overexpressed in the EGFR-high cells before adding drug, but increases 5.7 and 8.5 fold at weeks 1 and 4, respectively. We found that this gene also expressed in a sporadic pattern before the addition of drug, but that this sporadic expression pattern did not correlate strongly with *EGFR* and instead correlated more strongly with *GAPDH*, suggesting that while variable, its expression is not associated with the expression of pre-resistance markers, as expected based on the RNA-seq (Fig. 4.30).

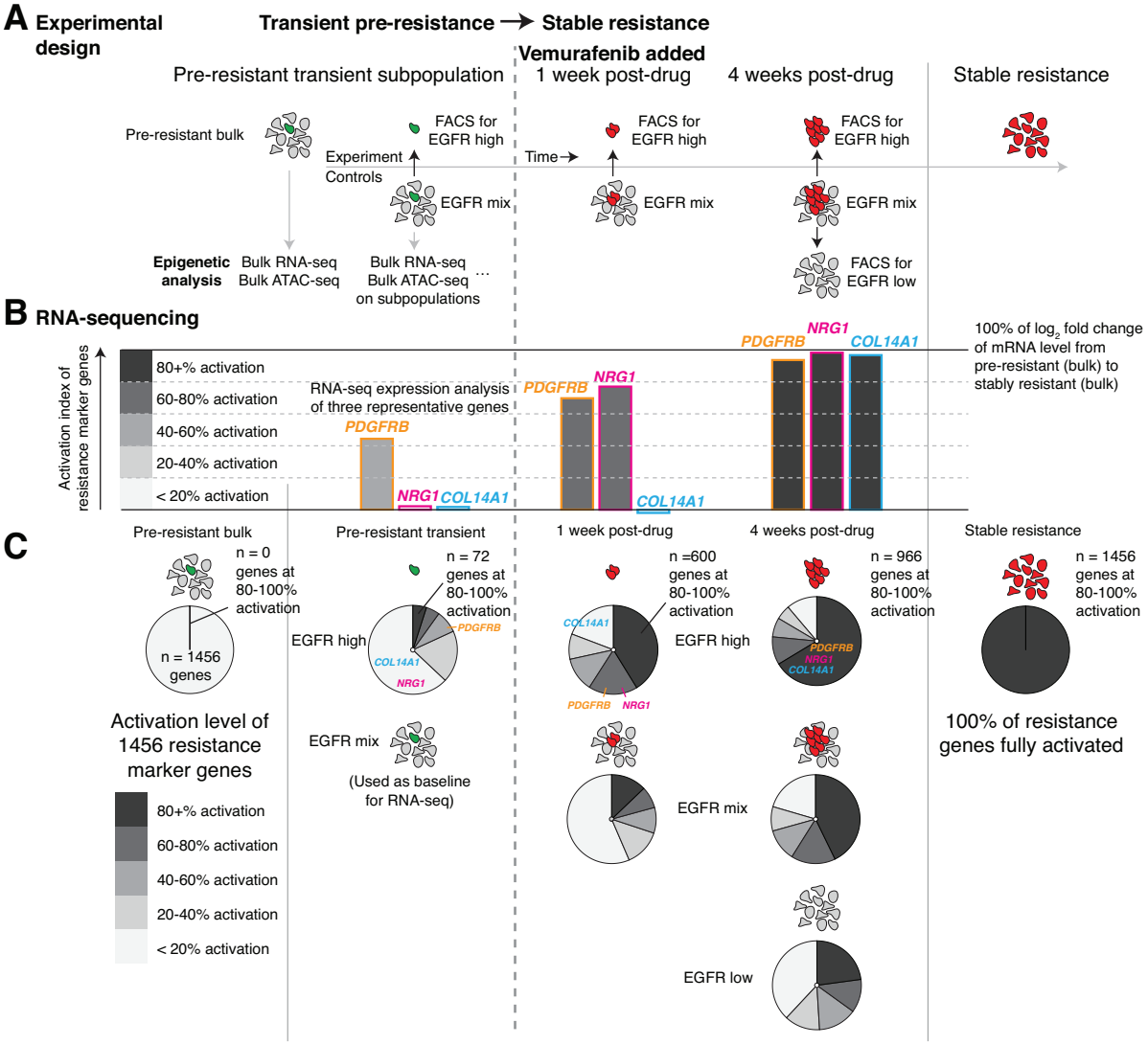

Figure 4.27: Addition of vemurafenib induces a stepwise reprogramming of pre-resistant cells into a stably-resistant state. A. Experimental design: we sorted EGFR-high cells to at different time points in vemurafenib treatment (untreated, 1 week, and 4 weeks) and performed RNA sequencing and ATAC sequencing on the sorted populations. B. Bar plots showing percentage activation for three example resistance genes through the development of stable resistance. Activation index is defined as $\log_2$ of the fold change divided by the total $\log_2$ fold change for the gene between the bulk non-resistant and bulk stably resistant populations. Each gene activates at a different stage. C. Pie charts showing the percentage of the resistance transcriptome that has become activated at different levels.

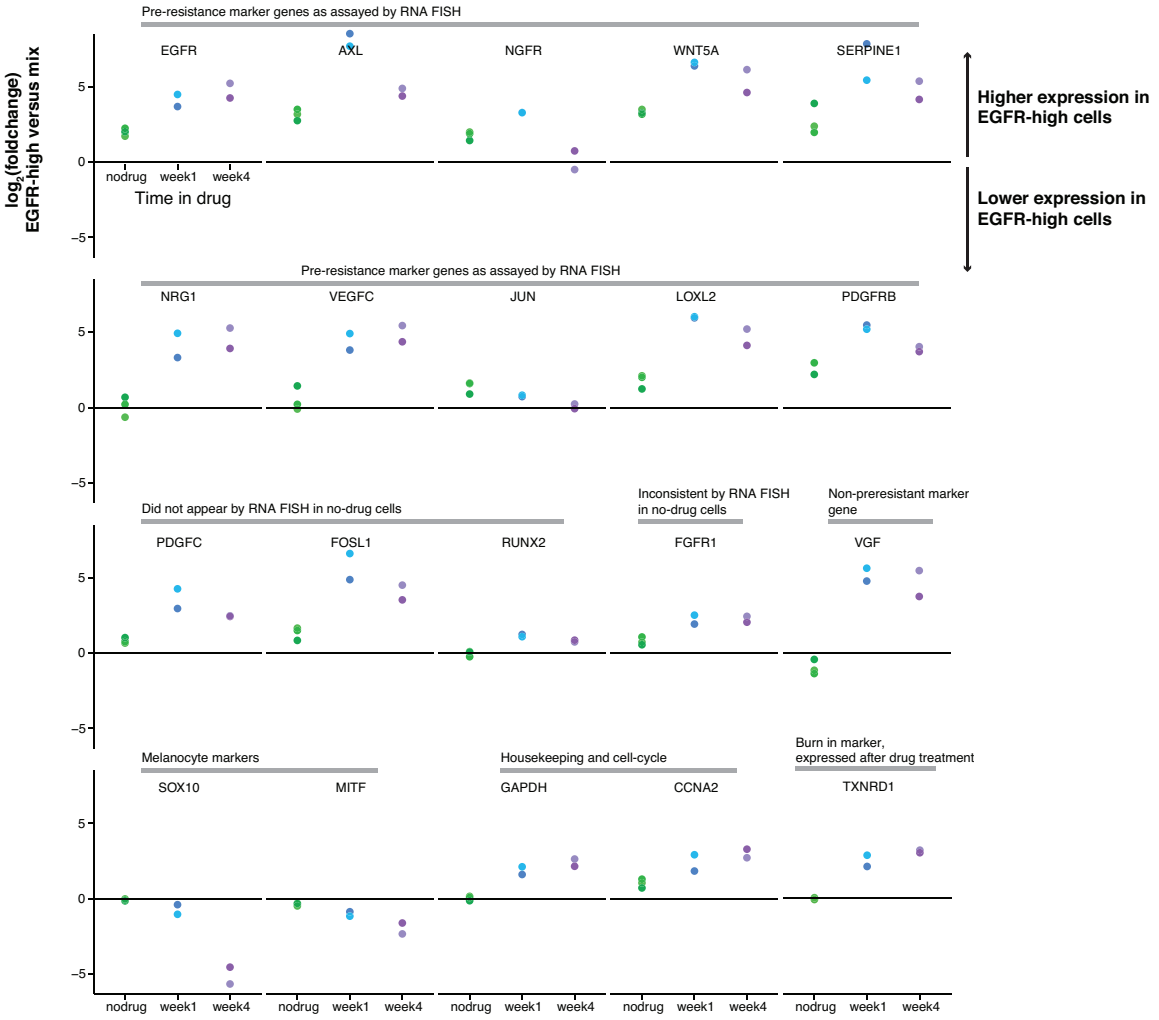

Figure 4.28: RNA-sequencing on FACS sorted EGFR high cells shows that sporadically expressing genes are more highly expressed in EGFR-high cells than the mixed population. Dot plot comparing the gene expression differences between EGFR-high and the mixed cell population. The y-axis shows the $\log_2$ fold change between the EGFR-high and mix cells, and the x-axis shows the different time points in drug (untreated, 1 week, and 4 weeks). Dots that fall above the zero line represent samples that have higher expression in the EGFR-high cells and dots that fall below the zero line represent samples that have lower expression in the EGFR-high cells. The genes summarized here are the same panel of genes used for multiplex RNA FISH in Fig. 4.19. Each dot represents a separate biological replicate. There are 3 biological replicates for the untreated condition, 2 biological replicates for week 1, and 2 biological replicates for week 4. We found that 8 of the 10 genes that exhibited jackpot behavior exhibited increased expression in the EGFR-high cells (*EGFR*, *AXL*, *NGFR*, *WNT5A*, *SERPINE1*, *JUN*, *LOXL2* and *PDGFRβ*).

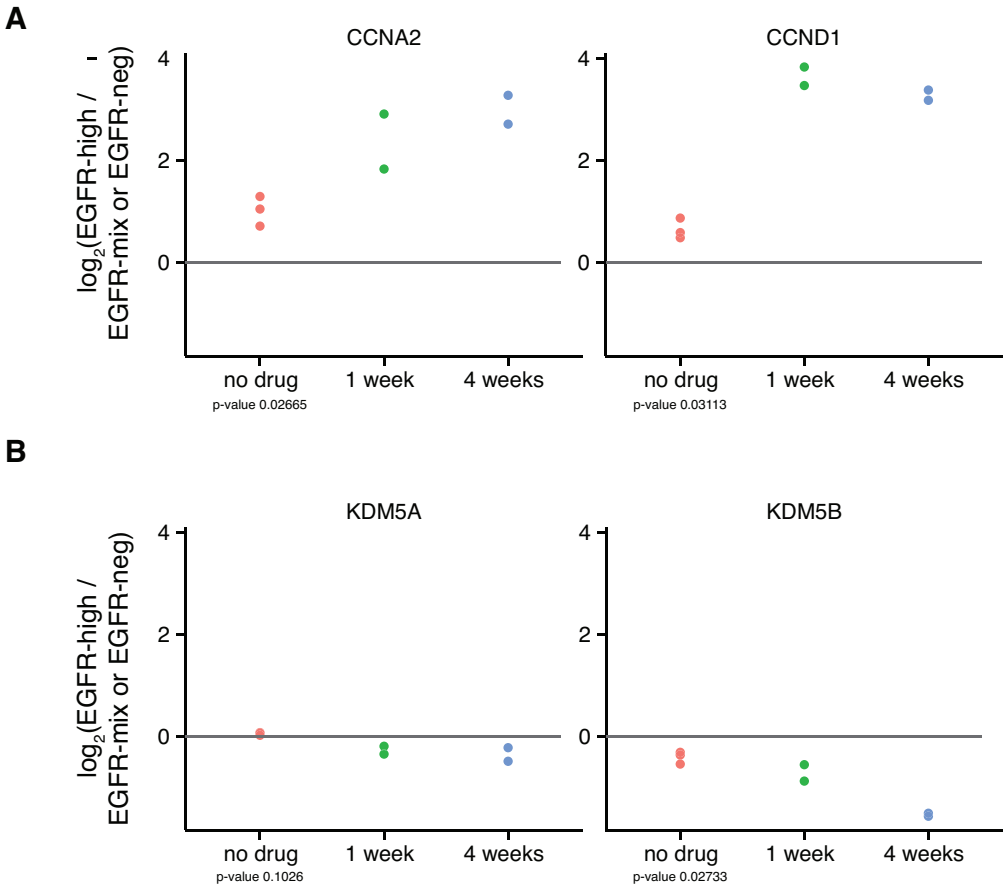

Figure 4.29: EGFR-high cells are proliferating based upon expression of cell cycle markers, and they do not express markers of slow-cycling subpopulations. A. Barplot showing the fraction of max expression for cell cycle genes (*CCNA2* and *CCND1*) across EGFR-high, EGFR-mixed, and EGFR-negative populations at each time point. B. Barplot showing the fraction of max expression for *KDM5A* and *KDM5B,* which are both markers of slow-cycling subpopulations in melanoma (Roesch et al. 2010; Sharma et al. 2010), across EGFR-high, EGFR-mixed, and EGFR-negative populations at each time point. Note that we only collected an EGFR-negative sample at 4 weeks because this was the only time point where the EGFR-high cells represented a significant portion of the total mixed population (>1%).

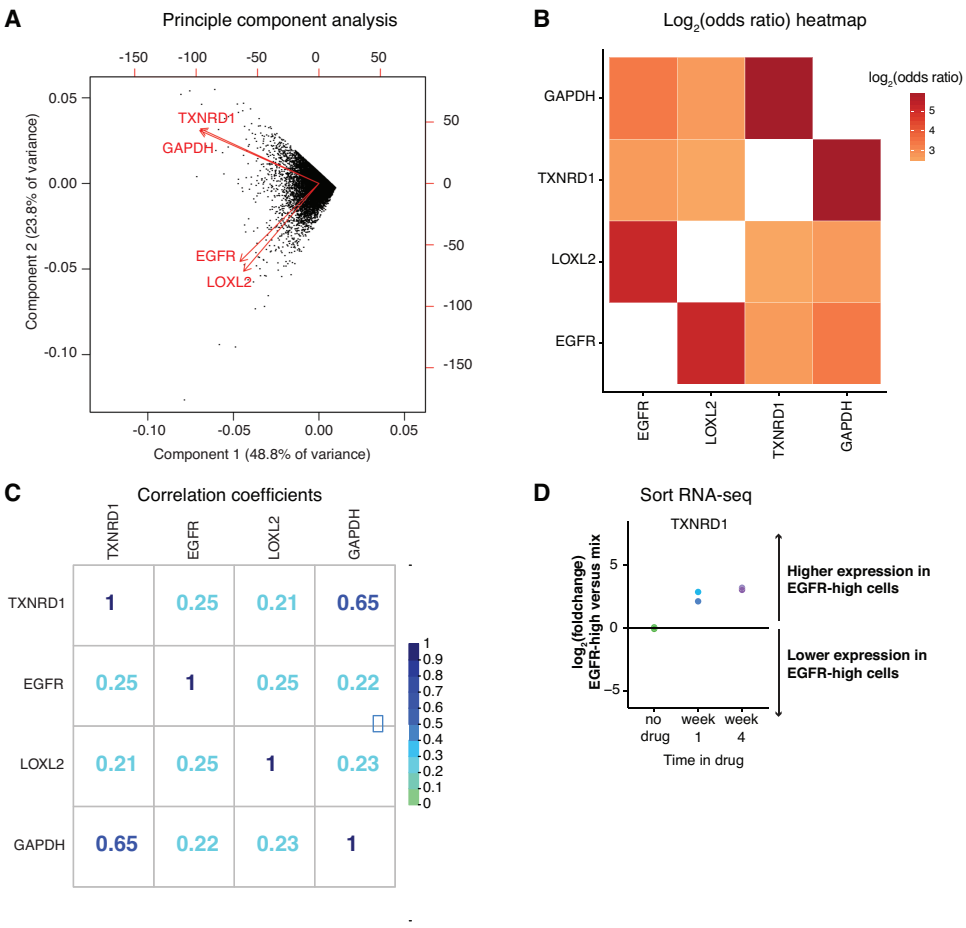

Figure 4.30: Single-cell RNA FISH shows that burn-in marker gene, *TXNRD1*, has variable expression in untreated cells, but that this expression does not correlate with resistance marker genes, *EGFR* and *LOXL2*. We performed RNA FISH on 21,314 untreated WM9 melanoma cells using probes for the burn-in gene *TXNRD1*, resistance marker genes *EGFR* and *LOXL2*, and the housekeeping gene *GAPDH*. A. We performed principal component analysis to determine which genes covary with which other genes. We transformed the vector representing the expression levels of each cell into the space spanned by the first two principal components. Arrows represent transformations of unit vectors of the specified gene into this same space. We observed two rough axes of variability, one corresponding to the *GAPDH* and the burn-in gene *TXNRD1* and the other to the resistance genes *EGFR* and *LOXL2*. Thus, these results show that burn-in gene *TXNRD1* has substantial covariation with housekeeping gene *GAPDH* while the resistance markers have their own separate axis of variation. b. Odds ratio analysis (as in Fig. 3b) on the same data shown in A. C. Pairwise correlations between the genes. D. EGFR-sort RNA-seq data from Fig. 4.27, B showing that *TXNRD1* is not enriched in the untreated EGFR-high cells, but then has higher expression in the EGFR-high subpopulation at 1 week and 4 weeks in drug. Together, these experiments validate the EGFR-sort RNA-seq experiments, and provide a more detailed analysis for *TXNRD1*.

4.15 ATAC sequencing of resistance reprogramming

Given the progressive and broad-ranging changes in the transcriptome in the post-drug stage of resistance, we wondered what gene regulatory changes underpinned these changes. To measure these regulatory differences, we used a genome-wide assay for transposase-accessible chromatin using sequencing (ATAC-seq) [11], which uses the Tn5 transposase to identify accessible regions of DNA (Fig. 4.27, C). This technique is thought to identify regions of DNA to which transcription factors are binding, and so we reasoned that regions that showed differential binding would be regions of DNA involved in the regulatory process governing the epigenetic reprogramming occurring during the post-drug phase of resistance. First, we looked for changes in putative transcription factor occupancy (measured by presence of accessible chromatin by ATAC-seq) between non-resistant and pre-resistant cells, with ATAC-seq revealing only 33 total differentially hypersensitive sites (Fig. 4.31, A). These results are again consistent with our RNA-sequencing findings in showing that the transient pre-drug pre-resistant state is a relatively shallow deviation from the non-resistant state.

Just as our RNA-sequencing data revealed a large, progressive change in the resistant cell transcriptomes upon addition of drug, ATAC-seq revealed similarly large changes in the pattern of transcription factor occupancy. We categorized these changes into gains and losses of hypersensitive sites, which are thought to correspond to increased and decreased transcription factor occupancy, respectively. We found that between the untreated cells and those in drug for 4 weeks, the cells lost 1787 and gained 9143 hypersensitive sites (Fig. 4.31, A), demonstrating a broad reprogramming of the cell. We then decomposed these hypersensitive site changes into hypersensitive site gains and losses over the first week in drug and then subsequent changes from

1 week of drug treatment to 4 weeks. Surprisingly, we found that the predominant change during the first week was the loss of hypersensitive sites (1999 lost vs. 431 gained), after which the change from 1 week to 4 weeks in drug was a gain of new hypersensitive sites (28 lost vs. 8577 gained). In order to identify the dominant factors involved in mediating these gains and losses, we searched for transcription factor binding motifs within these peaks, finding that much of the peak loss at one week of drug treatment resulted from loss of SOX10 binding, whereas the gain of peaks at four weeks resulted from the activation of TEAD and Jun/AP-1 activity, along with other signaling pathways (Fig. 4.31, B). We note that SOX10 is a transcription factor known to regulate neural crest embryonic development in cells such as melanocytes, and that TEAD activity is known to play a role in melanoma phenotype [50, 62, 123]. Thus, our results suggest that the post-drug burning-in stage of resistance consists of an early dedifferentiation stage followed by the activation of new signaling pathways as the cells progress towards becoming stably resistant.

4.16 Functional role of EGFR signaling in resistance

Both our ATAC-seq analysis and other studies [113, 123] have shown that gain of EGFR signaling are important in the development of vemurafenib resistance. Given our findings that the development of resistance consists of the two phases (pre-resistance followed by epigenetic reprogramming), we asked in what stage(s) of resistance EGFR is acting. To perturb EGFR activity in the pre-resistance phase, we applied lapatinib (an inhibitor of EGFR and HER2) for 3 days and then treated with vemurafenib while removing lapatinib, and found that the number of resistant colonies did not change with this pre-treatment. However, when we applied lapatinib in conjunction with

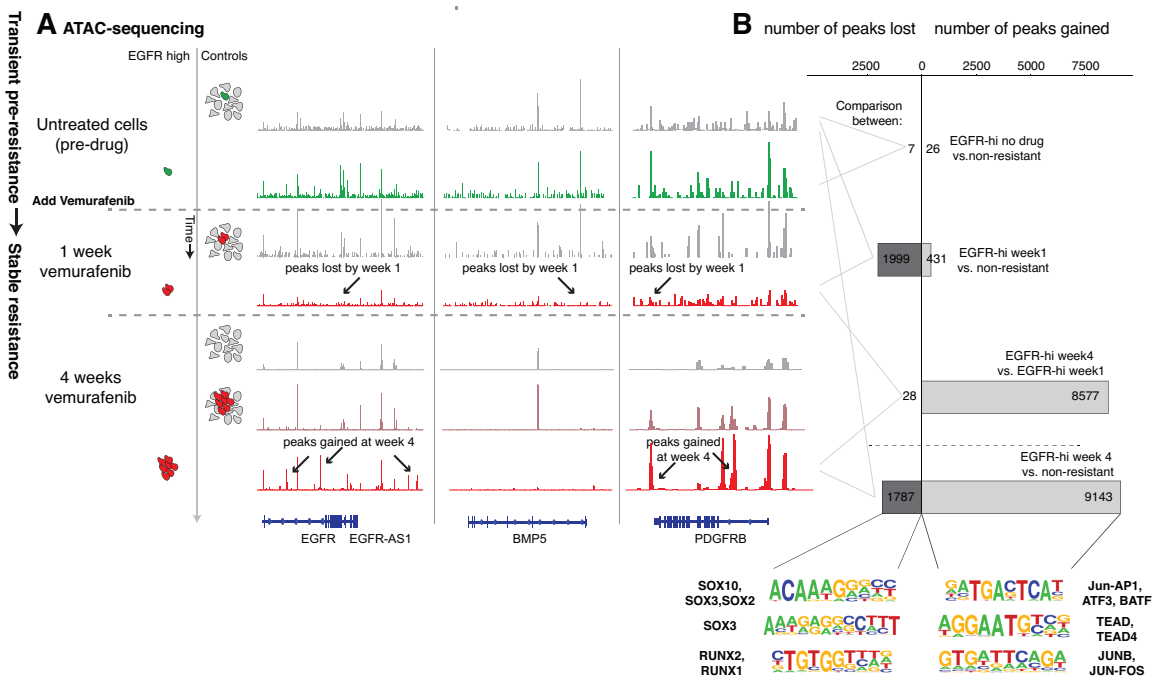

Figure 4.31: ATAC-seq reveals step-wise changes in genome accessibility during the development of vemurafenib resistance. A. ATAC-seq analysis to identify differentially hypersensitive sites between the sorted cell populations; example tracks shown displaying hypersensitive site loss and gain from one of two replicates. B. Quantification of peak loss and gain for the conditions shown in A; comparisons made as annotated on figure. We called a peak lost or gained if we could identify the peak in one of the conditions and saw a change in read count in the peak of 4 fold or higher across both replicates. We found motifs in the sets of peaks gained and lost by using HOMER differential peak calling tools and identified potential transcription factors using HOMER de novo motif tools.

vemurafenib, we found that the number of resistant colonies decreased remarkably from a mean of 10.5 to 0.5 colonies per well (Fig. 4.32, A). Adding lapatinib alone to melanoma cells did not cause significant cell death or affect cellular proliferation (Fig. 4.32, A). Thus, we concluded that EGFR signaling is critical in the post-drug burn-in phases of resistance and plays a role in the epigenetic reprogramming of pre-resistant cells into stably resistant cells.

4.17 Discussion

Here, we have demonstrated that resistance in melanoma, at the single cell level, can occur through non-genetic, non-heritable mechanisms, specifically through rare, coordinated, transient transcription of a small subset of resistance markers. We further showed that these pre-resistant cells transform into stably resistant cells via epigenetic reprogramming. This latter reprogramming phase consists of at least two steps: first, a dedifferentiation stage, and then the activation of new signaling pathways. In many ways, the development of resistance in these cells is analogous to the reprogramming of somatic cells into induced pluripotent stem cells, in which a rare, stochastically determined subpopulation ultimately undergoes deterministic changes to give rise to the transformed cellular state[12, 43]. This multi-phasic acquisition of the resistant phenotype can in many ways resemble the more conventional Darwinian/mutational picture of resistance, but is fundamentally different in nature because the cells do not need to possess a DNA mutation to become resistant.

The pre-resistant cellular state we have identified here is seemingly distinct from other non-genetic drug-resistance mechanisms. In many other cancer studies, the focus has been more on the immediate aftermath of drug treatment, in which some fraction of cells survive; typically, though, the percentage of surviving cells is not particularly

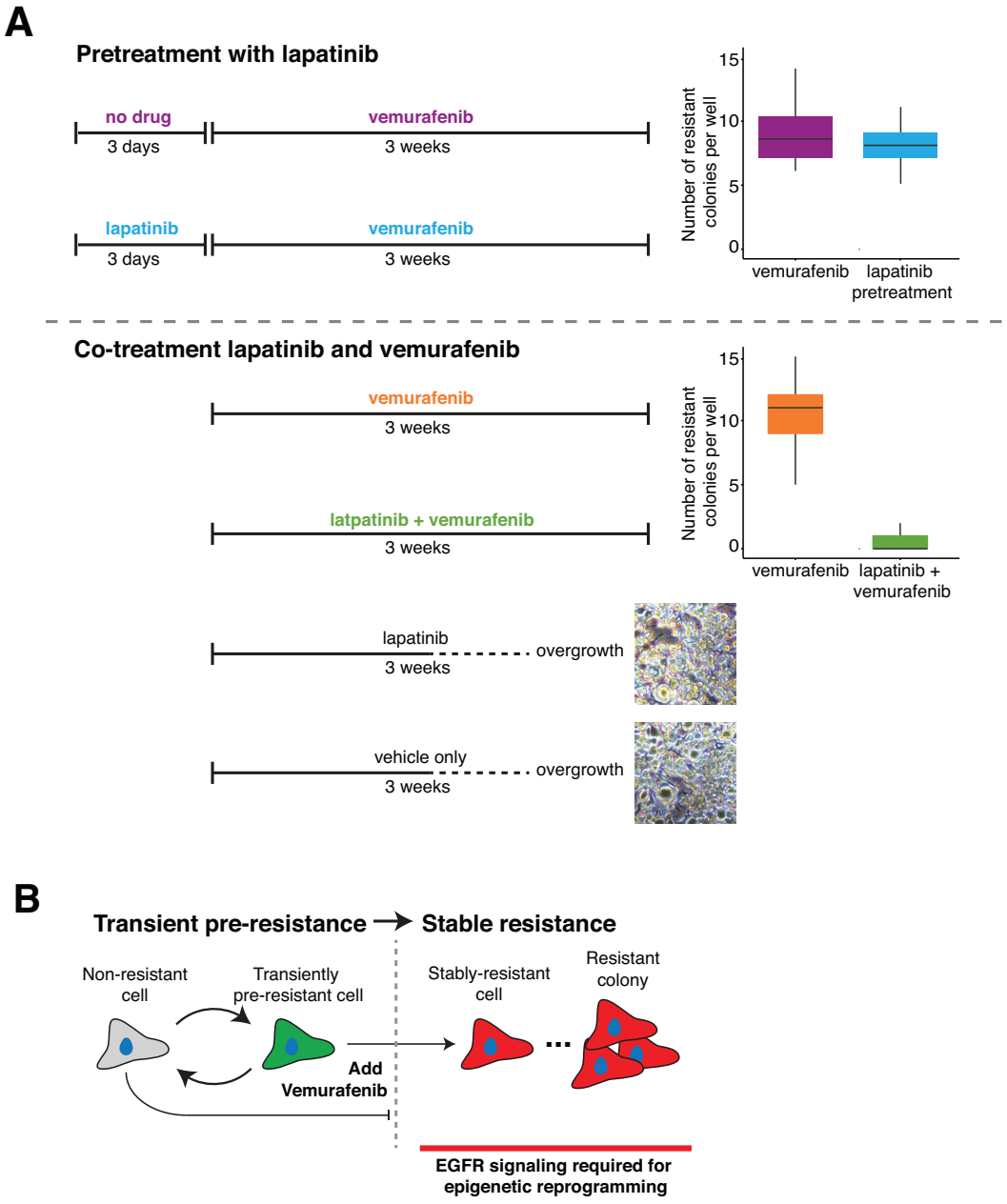

Figure 4.32: EGFR signaling is important for the burn-in phase of vemurafenib resistance. a. Comparison of pre- and co-treatment of cells with lapatinib and vemurafenib and corresponding quantification of number of resistant colonies. In the case of lapatinib and vehicle only, the cells had no growth inhibition, thus making it impossible to quantify colonies, as the images shown depict. Data shown are from 1 of 2 biological replicates. B. Schematic of resistance process with transient pre-resistance phase and epigenetic reprogramming phase involving EGFR signaling.

small [84, 86, 111]. While there are cases in which rare cells survive application of drug or other insult, most involve relatively dormant cells, such as bacterial persisters [3] and slow-cycling cells in melanoma[98, 99]. Here, we have documented a state that occurs rarely, but is proliferative and otherwise not outwardly distinct from the rest of the population, perhaps most conceptually similar to those found in Sharma et al. [107]. Whether such cell states are important for resistance in other cancers remains to be seen.

We were surprised to find that the pre-resistant state was characterized by rare, sporadic expression of a set of resistance marker genes. These extreme levels of transcriptional variability are qualitatively different than most documented examples of cell-to-cell variability in gene expression, much of which may be explained by differences in cell volume [83] or other factors [4]. Indeed, researchers have long hypothesized this "jackpot" rare-cell expression pattern, but examples had proven difficult to find [13]. Intriguingly, a recent study analyzing melanoma tissue via single cell RNA-sequencing found rare-cell expression of many of the same genes that we found in our study [117] (Fig. 4.33), although the genetic and microenvironmental variability in tumor tissue makes it difficult to directly infer that the variable expression of these genes is due to probabilistic transcription that is non-genetic or environmentally controlled.

Beyond melanoma, the observation that many of these same genes expressing in this manner across a variety of unrelated cell types (including primary, non-cancerous melanocytes) suggests the existence of a rare-cell expression program that may exist in normal, healthy cells and is co-opted in the development of resistance. Single cell expression profiling in cells in culture [77] and in tissue [117] may help reveal the full extent of this behavior.

Our results elucidate the epigenetic changes associated with the transformation from a transient pre-resistant to a stably drug-resistant phenotype, which consists of

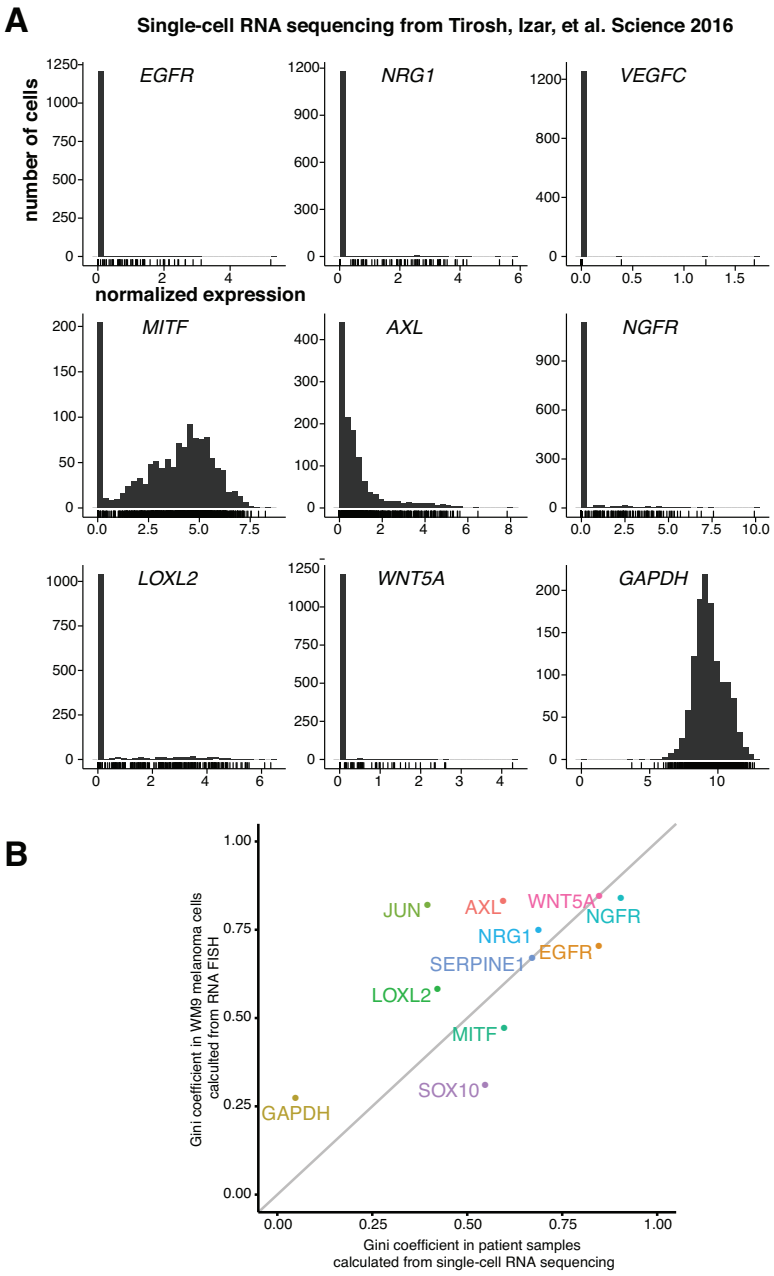

Figure 4.33: RNA sequencing of single melanoma cells from patients in Tirosh, Izar, et al. Science 2016 shows rare cells expressing high levels of resistance marker genes. A. Histograms of normalized single-cell expression data for 9 marker genes, including all malignant cells from this data set. B. Comparison of Gini coefficients calculated from patient single-cell RNA seq data in Tirosh, Izar, et al. Science 2016 and RNA FISH data from untreated WM9 melanoma cells. The gray line represents y=x.

dedifferentiation followed by a transcriptional rewiring. This rewiring may be possible due to the high degree of cellular plasticity that is intrinsic to melanoma and believed to underlie observations of phenotype switching between proliferative and invasive states [47, 48, 81]. Interestingly, it is the application of the drug itself that induces this progression. One question is whether it is the specific targets of the drug itself that are important for this effect, or whether it results from a more general response to cellular stress. Either way, further elucidation of this process may open up new avenues for therapeutic targeting [95]. Our results show several classes of genes associated with this process, including, intriguingly, JUN and targets of JUN/AP-1 signaling [34, 97] as well as several genes involved in cytoskeletal remodeling. Given the obvious morphological differences between resistant and nonresistant cells, we wonder whether cytoskeletal reconfiguration is not just associated with epigenetic transformation, but actually the proximal cause of the transformation to the resistant phenotype [14, 20, 46, 54]. More definitive studies involving manipulation of cell shape may help disentangle cause from consequence in this case.

Melanoma may serve as proof-of-principle for studies in drug resistance. While we use inhibition of mutated BRAF as prime example of highly specific therapy, the phenomenon may be more general. These rare-high expressing cells will likely survive inhibitors of other signaling molecules such as PI3K and AKT or even DNA-damaging chemotherapeutic drugs that are notorious for inducing acquired resistance in the malignant cells. Potentially, these cells have high mitochondrial metabolic rates [100], which will likely require that we need to do systematic drug development studies to effectively target these rare cell populations. Likely, we will need to develop treatment strategies to eliminate two populations, both the bulk of the tumor and the highly drug resistant cells.

Our Luria-Delbrck fluctuation analysis, along with our other results, provide strong

evidence that, at the single cell level, the determination of whether a cell will give rise to a resistant colony does not rely on a heritable difference (in particular, a DNA mutation). We emphasize, however, that our results do not preclude the possibility that genetics can play an important role in resistance, and there are of course many instances in which specific secondary mutations are associated with resistance in melanoma. For instance, in the vemurafenib phase II clinical trial, four of 20 patients acquired mutations in MEK and three of 13 acquired mutations in NRAS that were not present prior to treatment [120]. However, transient and genetic causes of resistance are not mutually exclusive. It is possible that transient sources of resistance provide an initial form of resistance that allows a small subpopulation of tumor cells to survive long enough for them to subsequently acquire secondary mutations, which then drive the reformation of the tumor mass. Delineating the relative contributions of both mechanisms at different stages of resistance will be important, especially because these mechanisms can influence population dynamics in different ways. For instance, one implication of our work is that a brief application of drug (rather than a continuous dose, as is the current standard of care) may be preferable because it would prevent the relatively long burn-in reprogramming process from going to completion, thus leaving cells in a transient state in which they can revert back over time and thus become sensitive to further drug treatment [59, 60]. Such interval dosing regimens have shown some promise, and our findings may further inform such strategies [23, 103, 113].

Appendix A

Experimental and computational methods

A.1 Materials and methods for Turbo RNA FISH

A.1.1 Cell culture

We cultured A549 cells (ATCC CCL-185), HeLa cells [30], and primary human foreskin fibroblasts (ATCC CRL-2097) in Dulbecco's modified Eagle's medium with Glutamax (DMEM, Invitrogen) supplemented with 10% fetal bovine serum and penicillin/streptomyocin. WM983b cells [5] were cultured in melanoma isolation media containing 80% MCDB153, 18% Leibovitz's L-15, 2% fetal bovine serum, 1.68mM CaCl2, and penicillin/streptomyocin.

A.1.2 Formaldehyde fixation

We fixed cells for 10 minutes in 4% formaldehyde/10% formalin in 1X phosphate buffered saline solution at room temperature. Following fixation, we washed cells twice with 1X phosphate buffer solution and then permeabilized the cells with 70% ethanol and stored them at 4°C for at least overnight.

A.1.3 Alcohol fixation

We fixed cells in pre-chilled ethanol or methanol (-20°C) for 10 minutes. Following fixation, we proceeded immediately to RNA FISH or Turbo RNA FISH.

A.1.4 RNA FISH

To perform RNA FISH, we followed the protocol in Raj et al. Nat Meth 2008 [91] with minor modifications. We pre-washed cells with wash buffer containing 10% formamide and 2X saline-sodium citrate (SSC). We then performed hybridization by adding 1 μL of probe to 50 μL of hybridization buffer consisting of 10% formamide, 2X SSC, and 10% dextran sulfate (w/v). For the overnight hybridizations, the final probe concentrations was 3.5 μM for TOP2A probe and 4.9 μM for TBCB probe. We hybridized the samples overnight in a humidified chamber at 37°C. Following hybridization, we washed the samples twice with wash buffer for 30 minutes at 37°C. We then imaged the samples in 2X SSC.

A.1.5 Turbo RNA FISH

For Turbo RNA FISH, we removed the alcohol from previously fixed samples and performed hybridization with 5 μL of hybridization buffer containing 71 μM TOP2A probe and 98 μM TBCB probe (unless otherwise specified), 10% formamide, 2X SSC, and 10% dextran sulfate (w/v). We hybridized the samples for 5 minutes (unless otherwise specified) on a covered hot plate at 37°C. Following hybridization, we washed the samples three times for one minute at 37°C with prewarmed wash buffer. We then imaged the samples in 2X SSC.

A.1.6 Turbo iceFISH

For Turbo iceFISH, we followed the protocol of Levesque and Raj Nat Meth 2013 [58] but with methanol fixed cells, higher probe concentration and shorter hybridization times. We used the probes described in that publication to "paint" chromosome 19, with the chromosome paint labeled with Alexa 594. We performed iceFISH in HeLa cells, which have two normal copies of chromosome 19 and two derivative chromosomes, t(13;19) and t(6;19).

A.1.7 Turbo SNP FISH

For Turbo SNP FISH, we followed the protocol of Levesque et al. Nat Meth 2013 [57] but with with methanol fixed cells, higher probe concentration and shorter hybridization times. Notably, we also performed a 20-minute post-fix in formaldehyde after the hybridization to prevent probes from dissociating. We used the probes described in that publication to detect the BRAF V600E mutation, as well as one that targets a portion of BRAF that is the same on both alleles as a control. Our image analysis was the same as that described in the Levesque et al. Nat Meth 2013 [57] manuscript. We performed all the experiments in WM983b cells (gift of Meenhard Herlyn, Wistar Institute), which are heterozygous for the V600E mutation.

A.1.8 Image acquisition

We imaged all samples on a Nikon Ti-E inverted fluorescence microscope using a 100X Plan-Apo objective (numerical aperture of 1.43) and a cooled CCD camera (Andor iKon 934). We sequentially acquired three-dimensional stacks of fluorescence images in four different fluorescence channels using filter sets for DAPI, Cy3, Alexa 594, and Atto 647N. Our exposure times ranged from 2 to 3 s for most of the dyes

except for DAPI, for which we used 50 ms exposures. The spacing between consecutive planes in our stacks was 0.3 μm. The filter sets we used were 31000v2 (Chroma), 41028 (Chroma), SP102v1 (Chroma), a custom set from Omega as described previously [89], SP104v2 (Chroma) and SP105 (Chroma) for DAPI, Atto 488, Cy3, Alexa 594, Atto 647N and Atto 700, respectively.

A.1.9 Image analysis and quantification

After imaging, we then put our data through an image analysis pipeline for semi-automated spot recognition. We implemented the analysis pipeline in MATLAB. Briefly, our method for analysis involves running the images through a linear filter designed to enhance spots around the size of those we observe, then finding all regional maxima within the filtered image, and then counting the number of regional maxima below a variety of thresholds [91]. We then manually determine a threshold where the number of regional maxima changes the least upon changing the threshold (i.e., the number of spots is least sensitive to moving the threshold). To quantify sensitivity of the threshold, we took the derivative of the logarithm of the graph of the number of regional maxima below varying thresholds. We smoothed the derivative before quantifying to avoid noise due to local variations in the graph.

A.2 Materials and methods for Flu RNA FISH

A.2.1 Influenza viruses

We created viruses expressing the A/California/7/2009 HA and NA or A/Puerto Rico/8/1934 HA and NA via reverse genetics as previously described [72]. All of our reverse genetics-based viruses possessed internal A/Puerto Rico/8/1934 genes. We used QuickChange site-directed mutagenesis kits (Stratagene, La Jolla, CA, USA)

to introduce the D225G HA mutation into the A/California/7/2009 HA to increase virus yields [61]. We also used site-directed mutagenesis to introduce the H274Y NA mutation. All reverse genetics-derived viruses were propagated in 10 day old fertilized chicken eggs.

The A/Victoria/361/2011 strain was obtained from Influenza Reagent Resource, Influenza Division, WHO Collaborating Center for Surveillance, Epidemiology and Control of Influenza, CDC, Atlanta, GA, USA. The B/Florida/4/2006 strain was obtained from BEI Resources, American Type Culture Collection, Manassas, VA, USA. A/Victoria/361/2011 was propagated on Madin-Darby canine kidney cells and B/Florida/4/2006 was propagated in 10 day old fertilized chicken eggs.

For each virus, we used Sanger sequencing to verify that additional mutations did not arise during viral propagation.

A.2.2 Influenza infections

We plated Madin-Darby canine kidney (MDCK) cells on chambered coverglasses (1e5 cells/well) or T75 flasks (1.8e6 cells/well) 18 hours before infections. Immediately before infection, we washed the cells two times with serum-free media. We then added the viruses to the cell layer at multiplicity of infection of 2-10 in serum-free media with L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK)-treated trypsin, HEPES, and gentamicin. Twenty four hours after infection, we fixed cells in 100% methanol.

A.2.3 Rhinovirus infections

We purchased the following rhinovirus strains from ATCC: Human rhinovirus 16 (VR-258), Human rhinovirus 1A (VR-1559), Human rhinovirus 81 (VR-1191), and Human rhinovirus 1B (VR-1645).

We grew HeLa cells in T25 flasks to 70% confluency in Dulbecco?s modified Eagle?s medium with Glutamax (DMEM, Invitrogen) supplemented with 10% fetal bovine serum (FBS), penicillin/streptomycin, 20 mM HEPES, and 40 mM magnesium chloride. We diluted rhinovirus stock at least 1:10 in 2 mL PBS with 0.5% bovine serum albumin. We then washed cells with PBS and added diluted virus. After incubating the cells at 34°C for 1-2 hours, we removed the virus containing medium and added DMEM supplemented with 2% FBS, pen/strep, 20 mM HEPES, and 40 mM MgCl2. We continued to incubate the cells at 34°C for 16-20 hours, after which we washed the cells in PBS and fixed the cells in ice-cold methanol for 10 minutes.

A.2.4 Adenovirus infections

We grew 293T cells in T25 flasks to 70% confluency in DMEM supplemented with 10% FBS and penicillin/streptomycin (D10F). We washed the cells once with PBS and added 1 mL Ad5ΔE1-GFP virus diluted 1:10 in DMEM. We then incubated the cells at 37°C for 1 hour, after which we removed the virus containing medium and added fresh D10F. We continued to incubate the cells at 37°C for 16-20 hours, and then washed the cells in PBS and fixed the cells in ice-cold methanol for 10 minutes.

A.2.5 Device fabrication

We laser cut a piece of clear acrylic for the top layer and two pieces of double sided Mylar tape (thickness = 100 μm, 3M Company) to serve as channel layers (Fig. 3.1 and Fig. 3.6). Between the channel layers, we placed a polycarbonate track-etched film (pore diameter = 5 μm, Whatman, Nuclepore). We assembled the device in a top down fashion by aligning the corners of the acrylic with the laser cut tape and pressing firmly to remove bubbles trapped in the adhesive layer. We placed the track-etched

film on top of the second tape layer and then pressed the first two layers onto the tape, thereby sandwiching the filter. We next attached the bottom layer which is a #1 cover glass (24 mm x 50 mm) to allow for imaging of samples directly on the chip. At the inlet to the device, we secured a laser-cut donut shaped piece of acrylic to create a reservoir for holding liquids. Lastly, we inserted medical tubing into the outlet of the device and secured the tubing with superglue to allow us to pull liquid through the device using a syringe and syringe pump. This allows us to precisely control the flow rate during experimentation.

A.2.6 Rapid RNA FISH on the microfluidic device

We used a pre-programmed syringe pump to automate the fluid handling in all of the following steps and mounted the microfluidic device onto a 37°C hotplate. For each sample, we fixed the cells in methanol prior to loading on the microfluidic device. After fixation, we loaded 200 μL of the cells in methanol into the fluid reservoir of the device and then pulled the cells onto the device at a flow rate of 100 μL/min. Next, we performed a pre-wash step by loading 200 μL of wash buffer containing 10% formamide and 2X SCC and pulling the solution through the device with the syringe pump. To prepare our hybridization solution, we put 1 μL of each viral RNA FISH probe (at stock concentrations ranging from 1288 ng/μL to 5800 ng/μL) into 500 μL of hybridization buffer containing 10% formamide, 2X SCC, and 10% dextran sulfate (w/v). We loaded 80 μL of hybridization solution into the device reservoir and pulled 20 μL onto the sample. We stopped flow through the pump and performed a 5 minute hybridization. At each minute during hybridization, we pulled an additional 4 μL of hybridization solution through the device. After hybridization, we loaded 200 μL of wash buffer into the reservoir and then the pump pulled the buffer through at 200

μL/min. We repeated this wash step 2 additional times for a total of 3 wash steps of 1 minute each. Finally, we applied 200 μL of 2X SSC and imaged the sample. All flow rates were 200 μL/min unless otherwise noted.

A.2.7 Rapid RNA FISH on coverglass

As previously described, we fixed all samples in methanol prior to RNA FISH [106]. We prepared probe solution by combining 50 μL of hybridization buffer containing 10% formamide, 2X SSC, and 10% dextran sulfate with 1 μL of each probe (at stock concentrations ranging from 1288 ng/μL to 5800 ng/μL). For each experiment, we removed all of the methanol from the sample, applied 10 μL of hybridization solution, and covered the sample with a coverslip to disperse the hybridization solution. We hybridized the RNA FISH probes for 5 minutes on a 37°C hotplate. We then washed the samples 3 times for 1 minute at 37°C with wash buffer containing 10% formamide and 2X SSC. Finally, we imaged the cells in 2X SSC.

A.2.8 SNP RNA FISH on coverglass

We adapted SNP RNA FISH protocols previously developed in our lab to target the 823 C -> T mutation in influenza H1N1 [57, 106]. Twenty four hours after infection with the virus, we fixed the infected cells using 100% methanol. We prepared a probe solution by combining 50 μL of hybridization buffer containing 10% formamide, 2X SSC, and 10% dextran sulfate with 1 μL of guide probe, 0.5-2 μL of wild-type and mutant probe (volume adjusted to achieve 1:1 concentration ratio of the probes), and mask oligonucleotide at the volume for 2X the concentration of each SNP probe. We removed the methanol from the fixed cells, applied 10 μL of probe solution, and hybridized the probes for 20 minutes at 37°C on a hotplate. We then washed the

sample 3 times for 1 minute at 37°C with wash buffer containing 10% formamide and 2X SSC. Finally, we imaged the cells in 2X SSC.

A.2.9 Microscopy

After RNA FISH, we imaged each samples on a Nikon Ti-E inverted fluorescence microscope using a 20X Nikon Plan objective (numerical aperture 0.40), 100X Plan-Apo objective (numerical aperture of 1.40) and a cooled CCD camera (Andor iKon 934). For 100X imaging, we sequentially acquired three-dimensional stacks of fluorescence images in five different fluorescence channels using filter sets for DAPI, Cy3, Alexa 594, Atto 647N, and Atto488. Our exposure times ranged from 0.1-2 sec. For SNP FISH specifically, we used exposure times of 3-5 sec. The spacing between consecutive planes in our stacks was 0.3 μm. For 20X imaging, we scanned the a specified area of interest, and at each position, we acquired one fluorescence image in five different fluorescence channels using filter sets for DAPI, Cy3, Alexa 594, Atto 647N, and Atto488. The filter sets we used were 31000v2 (Chroma), 41028 (Chroma), SP102v1 (Chroma), a custom set from Omega as described previously [58], SP104v2 (Chroma) and SP105 (Chroma) for DAPI, Atto 488, Cy3, Alexa 594, Atto 647 N and Atto 700, respectively.

A.2.10 Pan-probe design software

We implemented our pan-probe design software in MATLAB. The software has 4 main steps: 1) sequence cross-matching, 2) probe elimination, 3) oligonucleotide placement memoization, and 4) oligonucleotide solution readout. In step 1, the software reads an aligned FASTA-format file of sequences and compares 20 base pair oligonucleotides between all sequences for starting at each position in the aligned sequences, yielding cross-matches for all oligonucleotides. In step 2, the software eliminates probes

whose hybridization affinities are too far away from a predetermined optimum, determined to be equivalent to around 45% GC [91]. In step 3, the software uses dynamic programming/memoization to globally optimize the placement of oligonucleotides subject to the goals of minimizing the total number of oligonucleotides while still covering all sequences with a user-defined minimum number of oligos. In step 4, the best solution is read from the memoization table and returned to the user.

A.2.11 Subtype-specific design software

We implemented our subtype-specific probe design software in MATLAB. Our software runs similarly to previous design software [91], but in addition to the target sequence, it also accepts a set of other sequences that it tries to avoid. In the context of this manuscript, the target would be the desired viral substrains and the other sequences would be all related substrains that we wish to avoid targeting. Our algorithm screens every candidate oligonucleotide probe (i.e., 20 base oligonucleotides that bind to the target with perfect complementarity) for cross-targeting by excluding all candidates that exhibit more than some user-specified amount of sequence complementarity to all the avoided sequences (in our case, more than 14/20 bases).

A.2.12 Image processing

We implemented our image processing software in MATLAB. First, the software uses phase-correlation to create a montage of individual image tiles. Second, the user defines a region of interest (ROI) to delineate the microscopy viewing area of the device and exclude spurious non-cellular material. The software then stitches together the ROI image for each channel, applying a rolling-ball background subtraction to each included image tile. Third, using the DAPI channel ROI image, nuclei are segmented.

The software applies a wavelet filter to filter out high-frequency noise and examine objects about the size of nuclei, thresholds the image using Otsu's method, and then deletes small, edge, and non-circular objects. Each nuclear object is then increased by a 2 pixel radius to create objects that would include both the nucleus and cytoplasm of a typical cell. Fourth, the GFP channel of the image is used as a mask to delete spurious objects that may be the result of dust or other non-cellular objects being trapped by the device filter. The software removes objects that overlap with GFP pixels that are greater than 3 standard deviations above the mean GFP intensity. Fifth, the software determines median viral RNA (vRNA) fluorescence intensity using pixels in each cellular object, a histogram is created, and the fluorescence intensity of the first maxima of the histogram is subtracted from the vRNA fluorescence intensity data to help eliminate device-to-device variability in washes and image acquisition.

A.2.13 SNP FISH image processing

First, we segmented cells and located individual RNA spots using our lab's custom software described in Levesque et al. For each spot, we fit a two-dimensional gaussian to determine the location of the spot with sub-pixel resolution. We then performed co-localization by finding the nearest SNP probe signal to every guide spot within a 3.0 pixel range. From this, we obtained the median displacement vector field for each SNP FISH and guide pair and subsequently shifted the image to correct for chromatic deviations. We then repeated the search for colocalization between the guide and SNP probe using a 1.5 pixel range to establish the final colocalization.

A.3 Cancer RNA FISH methods

A.3.1 Cell culture, drugs, and fixation

We grew melanoma cell lines (WM9, WM983B, and 1205Lu, SK-MEL-28) from the lab of Meenhard Herlyn, validated in the Herlyn lab by short tandem repeat profiling using AmpFlSTR Identifiler PCR Amplification Kit (Life Technologies) in Tu2% media containing 78% MCDB, 20% Leibovitz?s L-15 media, 2% FBS, and 1.68mM CaCl2 and primary melanocytes isolated from human neonatal foreskin (Fom217-1 from the lab of Meenhard) in Medium 254CF (Life Technologies, M254500) supplemented with Human Melanocyte Growth Supplement (Life Technologies, S0025). We also grew HeLa cells and MDA-MB-231 cells in DMEM supplemented with 10% FBS, PC-9 cells in RPMI supplemented with 10% FBS, and SH-SY5Y cells in DMEM/F12 with 10% FBS. We made stocks of 4 mM vemurafenib (Selleckchem, S1267) and 4 mM lapatinib (Santa Cruz Biotechnology, 202205B) in DMSO, and diluted in media to a final concentration of 1 μM in all drug treatment experiments. For all RNA FISH experiments, we grew cells on two-well Lab-Tek chambered coverglasses. We fixed and permeabilized cells for RNA FISH according to [91].

A.3.2 Time-lapse imaging

We imaged the cells on a Nikon Ti-E enclosed in plexiglass incubation chamber heated to 37°C with 5% CO2. We seeded WM9 cells into a two-well Lab-Tek chambered coverglasses and took brightfield images every 2 hours for 28 days. At each time point, we acquired a total of 702 images over a 39x18 grid of images at 10X magnification to capture the entire culture dish. We stitched each of tiles into one composite image for each time point and then compiled movies from the images using MATLAB.

A.3.3 Luria Delbruck fluctuation analysis

To minimize pre-existing genetic heterogeneity in the cell line, we isolated a single-cell clones from WM9 and WM983B melanoma cell lines. We expanded these clones up to 100-200 cells total and then isolated single cells to derive the subclones for the Luria Delbruck fluctuation analysis. We then allowed the subclones to grow in culture through 20 doublings for WM9 and 22 doublings for WM983B to give approximately 1 million cells and 4 million cells, respectively. After expansion, we trypsinized each subclone, counted the number of cells in the culture using a hemocytometer, and then seeded 600,000 cells into 2 12-well plates (yielding 25,000 cells per well). We had a total of 43 and 29 subclones with the WM9 cell line (biological replicates) and 20 subclones for WM983B. One day after seeding into 12-well plates, we applied 1 μM vemurafenib. Throughout the experiment, we changed the media and drug and counted the number of resistant colonies twice per week. We ended the experiment when the plates stopped developing new resistant colonies or when all the resistant colonies appeared to be daughter colonies from larger ones. Note that upon re-plating the cultures before administration of drug, we observed varying degrees of growth and plating efficiency, all of which served to increase the variance, as we found that cultures with larger numbers of cells following replating had generally higher numbers of resistant cells. Thus, by not taking this into account, our observed variance is likely higher than the actual variance, biasing against the transient pre-resistance hypothesis.

To show that our resulting counts of resistant colonies were likely not the result of a strongly heritable transition to a pre-resistant state, we simulated the strongly heritable Luria-Delbruck process. Briefly, the parameters are the initial culture size (set to one in our case), the ultimate sizes of the cultures (we used the largest multiple of two lower than the actual measured culture sizes, thus biasing against ourselves),

and the mutation rate, which we varied as part of our simulations. For each mutation rate, we ran the simulations 10,000 times, and noted both the Fano factor (variance divided by the mean) and coefficient of variation across the simulated cultures for each iteration. We then computed a p-value for each mutation rate by determining how often the simulated Fano factor or coefficient of variation (separate p-values for each statistic) exceeded our actual measurements (Fig. 4.5 and Fig. 4.6). The p-value we report is based on the most conservative estimate based on both the statistics we computed.

A.3.4 Iterative RNA FISH

We designed oligonucleotide probe sets using the Stellaris probe designer (Biosearch Technologies) and ordered them with an amine group on the 3'end. We pooled the oligonucleotides for each probe set and coupled them to either Cy3 (GE Healthcare), Alexa 594 (Life Technologies), Atto647N or Atto 700 (Atto-Tec). We performed RNA FISH as previously described [90, 91] for each of the cycles of hybridization. We first fixed cells with formaldehyde and permeabilized with 70% ethanol. We next washed once with wash buffer (containing 10% formamide and 2X SSC) and then applied hybridization buffer (containing 10% formamide, 10% dextran sulfate, and 2X SSC) with the specified pool of RNA FISH probes. We hybridized for 6-12 hours and then washed 2 times for 30 minutes with wash buffer.

After imaging, we applied 60% formamide with 2X SSC for 15 minutes on a heat plate kept at 37°C. We then washed the sample 3 times with 1X PBS for 15 minutes also at 37°C to remove residual formamide, which we have found can inhibit further hybridizations. Lastly, we washed once with wash buffer to remove residual 1X PBS and prepare the samples for another RNA FISH hybridization.

A.3.5 RNA FISH on patient derived xenografts

We fixed tissue sections by treating with 4% formaldehyde in PBS for 10 minutes and then permeabilized and stored them in 70% ethanol at 4°C. We performed one cycle of RNA FISH as described above. We mounted the samples for imaging in 2X SSC. We performed these experiments with two biological replicates from different mice and different tissue donors.

A.3.6 RNA FISH imaging

We imaged each sample on a Nikon Ti-E with a 60X Plan-Apo objective and filter sets for DAPI, Cy3, Atto647N, Alexa594, and Atto700. We used Metamorph imaging software (Scan Slide application) to acquire a tiled grid of images (40 by 40 for the data sets shown in Fig. 4.14, A) covering a 8.9 mm by 8.9 mm area of the sample. We used the Nikon Perfect Focus System to ensure that the images remained in focus over the imaging area.

A.3.7 Image analysis

We developed a custom MATLAB pipeline for counting RNA FISH spots in tiled images. First, this software segments the nuclei of individual cells using the DAPI images. Next, the software identifies regional maxima in each tiled image as potential RNA FISH spots and assigns them to the nearest nucleus. Through a MATLAB GUI, the user selects a global threshold for each RNA FISH channel to identify the individual spots. We then visually inspected all cells that were above the jackpot threshold and used GUI editing tools to remove any autofluorescent debris or artifacts from subsequent analysis. Lastly, we extracted the position of every cell in the scan and the number of RNA molecules for each fluorescent channel.

We also developed software to match cells across subsequent hybridizations, which poses a challenge because of slight warping in the tiled image in each acquisition. Our algorithm attempts to match cells locally by shifting cells in the first hybridization to all potential candidates in the subsequent hybridization, choosing the best match as the one that minimizes total distance for nearby cells. We then smooth out this shift and apply it across the entire tiled image field. We then matched cells by proximity, discarding cells that did not match uniquely to a nearby cell in subsequent hybridizations; our yield was typically >90% of cells in the initial hybridization matching in subsequent hybridizations.

We decided whether or not a cell was deemed a "high" expressing cell for a particular gene by determining whether the number of mRNA molecules in the cell exceeded a threshold. To avoid bias, our default was to set a threshold that captured the top 2% of cells. If this percentage did not yield a reasonable threshold, we manually set a more appropriate threshold based on the distribution. Occasionally, autofluorescent debris in the images would be spuriously identified as cells, often with high numbers of false spots in them. Thus, we manually removed such regions, starting with cells with the highest number of identified RNA counts and continuing until we reached a relatively low level of mRNA below which manually evaluating the data was no longer feasible. This procedure ensured that we manually inspected all jackpot cells to verify their expression levels. For genes that did not exhibit sporadic expression patterns (including *GAPDH*, *SOX10*, *CCNA2*), we set thresholds by plotting the distribution and selecting a threshold that captures the tail. We performed all iterative RNA FISH experiments in duplicate with biological replicates. We calculated Gini coefficients on the distributions of RNA FISH counts for each gene using the "ineq" package in R.

A.3.8 RNA sequencing and analysis

We sequenced messenger RNA from WM9 and WM983B melanoma cells. For WM9, we sequenced the RNA from 8 untreated samples, 8 samples treated with 1 μM vemurafenib for 48 hours, 10 resistant samples in 1 μM vemurafenib, 4 resistant samples with drug removed for 48 hours, and 4 resistant samples with drug removed for 1 week. For WM983B, we sequenced RNA from 20 untreated samples, 20 samples treated with 1 μM vemurafenib for 48 hours, 37 resistant samples in 1 μM vemurafenib, 2 resistant samples with drug removed for 48 hours, and 2 resistant samples with drug removed for 1 week. Each sample is a biological replicate. We used the NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra RNA Library Prep Kit for Illumina to extract polyadenylated RNA and prepare barcoded RNA sequencing libraries. We sequenced each sample at a depth of approximately 20 million reads on a HiSeq 2000 (50 base pair length) or NextSeq (75 base pair length). We aligned our reads to hg19 using STAR and quantified reads per gene using HTseq. We then used R to perform differential expression analysis with DESeq2 to identify resistance marker genes.

A.3.9 Generation of patient derived xenografts

We collected tumor biopsies from melanoma patients as previously described by [55]. Fresh biopsies were processed under sterile conditions within 24 hours. For processing, we used a cross blade technique to finely mince the tissue, then briefly digested the tissue in collagenase IV for 20 minutes at 37°C. The tumor tissue was then implanted s.c. with matrigel (Corning Life Sciences) into NSG mice (6-8 weeks, male or female). Tumor grafts were harvested at maximum tumor size and serially transplanted for expansion. Low passage PDX tumors were mounted in OCT immediately after sacrifice.

For RNA FISH analysis, we sectioned the tumors into 7 μm slices and then proceeded with RNA FISH as described above. All sample collection and animal experiments were approved by Wistar IRB and Wistar IACUC, respectively. We analyzed samples from two different patients: the first of which had a BRAF-V600E mutation and which was sensitive to combination BRAF/MEK inhibition, and the second had a NRAS-Q61L mutation and was not sensitive to MEK inhibition (data from the first patient is in Fig. 4.16, and data from the second patient is in Fig. 4.17).

A.3.10 EGFR and NGFR fluorescence assisted cell sorting

We stained WM9 melanoma cells for fluorescence assisted cell sorting using an antibody for EGFR. First, we trypsinized the cells, washed once with 0.1% BSA in 1X PBS, and incubated for 1 hour at 4°C with 1:200 mouse anti-EGFR antibody, clone 225 (Millipore, MABF120) in 0.1% BSA PBS. Next, we washed with 0.1% BSA PBS and then incubated for 30 minutes at 4°C with 1:500 donkey anti-mouse IgG antibody labeled with Alexa Fluor 488 (Jackson Laboratories, 715-545-150). We washed the samples again with 0.1% BSA PBS and resuspended in 1% BSA PBS with 2 mM EDTA and DAPI for fluorescence assisted cell sorting. We used a MoFlo Astrios (Beckman Coulter) to collect the top 0.02-0.2% of cells stained for EGFR. We used the DAPI stain to exclude dead cells and used cells that were not incubated with the primary antibody as a negative control. To stain WM9, WM983B, and SK-MEL-28 cells for NGFR, we used anti-NGFR clone ME20.4 fluorescently labeled with PE/Cy7 (Biolegend, 345110). We incubated the cells with 5 μL of antibody for 10 minutes at 4°C. We then washed the samples and proceeded with sorting (as described above). For our negative control, we used a PE/Cy7 mouse IgG1 (Biolegend, 400126). When sorting either EGFR-high or NGFR-high subpopulations, we also collected a EGFR-mixed or

NGFR-mixed population control by using the same gating for live cells, but without gating on the EGFR or NGFR stain. When staining for both EGFR and NGFR, we performed the EGFR staining first then stained with the NGFR antibody. When sorting for EGFR and NGFR together, we collected all 4 possible populations: cells negative for both stains, cells positive for EGFR only, cells positive for NGFR only, and cells positive for both EGFR and NGFR.

A.3.11 ATAC sequencing and analysis

We performed ATAC sequencing on WM9 melanoma cells according to [11]. Briefly, we lysed the cells and set up the transposition reaction with the Tn5 Transposes (Illumina Catalog, FC121-1030) at 37°C for 30 minutes. We cleaned the reaction with a Qiagen MinElute Kit and then amplified the libraries using the custom Nextera PCR primers described in [11]. We sequenced our libraries on a NextSeq with 75 base pair reads at a depth of approximately 40-70 million reads per sample. We aligned our reads to hg19 with bowtie2 and then used the HOMER package for peak calling, differential peak calling, motif analysis, and gene ontology analysis (code available at https://bitbucket.org/arjunrajlaboratory/rajlabseqtools).

A.3.12 MTS Cell Proliferation Assay

The cell viability was estimated by using CellTiter 96 Aqueous MTS Cell Proliferation Assay (MTS, PR-G1111). Briefly, WM9 cells were seeded in 96-well plates with 2000 cells/well. 24 hours incubation later, cells were cultured in the presence of vemurafenib at serial 3-fold dilution concentration. After 6 days treatment, 20 μl/well MTS reagent were added and incubated for 4 hours. Then read plate at 490 nm wavelength to estimate cell proliferation.

A.3.13 Apoptosis Assay

After treatment with vemuafenib at 1 μM, 3 μM for 3 days, WM9 cells were harvested with trypsin-EDTA, centrifuged into a pellet including all floating cells, and rinsed with phosphate-buffered saline (PBS). Then, the cells were re-suspended in Annexin V binding buffer containing Annexin V APC (Biolegend, 640920) and propidium iodide (Sigma, P4864). The cells were incubated at room temperature for 15 minutes and were analyzed using the FACSCalibur flow cytometry.

A.3.14 Western Blot

WM9 cells were cultured in 1 μM, 3 μM vemurafenib medium for 3 days. The cells were collected with or without floating cells and were lysed with TNE buffer with protease inhibitors. 30 μg protein extracts were electrophoresed on 12% SDS-Page gels and transferred on the Nitrocellulose membranes in Bio-rad Trans-Blot Turbo transfer system. The membranes were blocked with ODYSSEY Blocking Buffer (LI-COR, 927-40000) for 1 hour at room temperature and incubated at 4°C overnight with the following primary antibodies: pMek (Cell signaling, 9121s), pErk (Cell signaling, 4370s), pS6 (Cell signaling, 9121s), caspase3 (Cell signaling, 9662), parp (Cell signaling, 9542s), β-actin (Sigma, A5441). After 2nd antibodies incubation, membranes were visualized by LI-COR Odyssey infrared imaging system.

Bibliography

[1] Jamie N Anastas, Rima M Kulikauskas, Tigist Tamir, Helen Rizos, Georgina V Long, Erika M von Euw, Pei-Tzu Yang, Hsiao-Wang Chen, Lauren Haydu, Rachel A Toroni, Olivia M Lucero, Andy J Chien, and Randall T Moon. WNT5A enhances resistance of melanoma cells to targeted BRAF inhibitors. *J. Clin. Invest.*, 124(7):2877–2890, July 2014.

[2] N Q Balaban. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625, 10 September 2004.

[3] Nathalie Q Balaban, Jack Merrin, Remy Chait, Lukasz Kowalik, and Stanislas Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625, 10 September 2004.

[4] Nico Battich, Thomas Stoeger, and Lucas Pelkmans. Control of transcript variability in single mammalian cells. *Cell*, 163(7):1596–1610, 17 December 2015.

[5] D Becker, C B Meier, and M Herlyn. Proliferation of human malignant melanomas is inhibited by antisense oligodeoxynucleotides targeted against basic fibroblast growth factor. *EMBO J.*, 8(12):3685–3691, 1 December 1989.

[6] Jeremiah Bell, Aleta Bonner, Daniel M Cohen, Robert Birkhahn, Ram Yogev, Wayne Triner, Jason Cohen, Elizabeth Palavecino, and Rangaraj Selvarangan. Multicenter clinical evaluation of the novel alere i influenza A&B isothermal nucleic acid amplification test. *J. Clin. Virol.*, 61(1):81–86, September 2014.

[7] Jesse D Bloom, Lizhi Ian Gong, and David Baltimore. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science*, 328(5983):1272–1275, 4 June 2010.

[8] Guy Boivin, Stéphanie Côté, Pierre Déry, Gaston De Serres, and Michel G Bergeron. Multiplex real-time PCR assay for detection of influenza and human respiratory syncytial viruses. *J. Clin. Microbiol.*, 42(1):45–51, January 2004.

[9] Jennifer Bordeaux, Allison Welsh, Seema Agarwal, Elizabeth Killiam, Maria Baquero, Jason Hanna, Valsamo Anagnostou, and David Rimm. Antibody validation. *Biotechniques*, 48(3):197–209, March 2010.

[10] David N Breslauer, Robi N Maamari, Neil A Switz, Wilbur A Lam, and Daniel A Fletcher. Mobile phone based clinical microscopy for global health applications. *PLoS One*, 4(7):e6320, 22 July 2009.

[11] Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, 10(12):1213–1218, December 2013.

[12] Yosef Buganim, Dina A Faddah, Albert W Cheng, Elena Itskovich, Styliani Markoulaki, Kibibi Ganz, Sandy L Klemm, Alexander van Oudenaarden, and Rudolf Jaenisch. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150(6):1209–1222, 14 September 2012.

[13] Moran N Cabili, Margaret C Dunagin, Patrick D McClanahan, Andrew Biaesch, Olivia Padovan-Merhar, Aviv Regev, John L Rinn, and Arjun Raj. Localization

and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.*, 16(1):20, 29 January 2015.

[14] Gaia Cantelli, Jose L Orgaz, Irene Rodriguez-Hernandez, Panagiotis Karagiannis, Oscar Maiques, Xavier Matias-Guiu, Frank O Nestle, Rosa M Marti, Sophia N Karagiannis, and Victoria Sanz-Moreno. TGF-β-Induced transcription sustains amoeboid melanoma migration and dissemination. *Curr. Biol.*, 0(0).

[15] Jason M Casolari, Christopher R Brown, Suzanne Komili, Jason West, Haley Hieronymus, and Pamela A Silver. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell*, 117(4):427–439, 14 May 2004.

[16] Paul B Chapman, Axel Hauschild, Caroline Robert, John B Haanen, Paolo Ascierto, James Larkin, Reinhard Dummer, Claus Garbe, Alessandro Testori, Michele Maio, David Hogg, Paul Lorigan, Celeste Lebbe, Thomas Jouary, Dirk Schadendorf, Antoni Ribas, Steven J O'Day, Jeffrey A Sosman, John M Kirkwood, Alexander M M Eggermont, Brigitte Dreno, Keith Nolop, Jiang Li, Betty Nelson, Jeannie Hou, Richard J Lee, Keith T Flaherty, Grant A McArthur, and BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.*, 364(26):2507–2516, 30 June 2011.

[17] Yi-Ying Chou, Nicholas S Heaton, Qinshan Gao, Peter Palese, Robert H Singer, Robert Singer, and Timothée Lionnet. Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog.*, 9(5):e1003358, 9 May 2013.

[18] Yi-Ying Chou, Reza Vafabakhsh, Sultan Doğanay, Qinshan Gao, Taekjip Ha, and Peter Palese. One influenza virus particle packages eight unique viral RNAs

as shown by FISH analysis. *Proc. Natl. Acad. Sci. U. S. A.*, 109(23):9101–9106, 5 June 2012.

[19] A A Cohen, N Geva-Zatorsky, E Eden, M Frenkel-Morgenstern, I Issaeva, A Sigal, R Milo, C Cohen-Saidon, Y Liron, Z Kam, L Cohen, T Danon, N Perzov, and U Alon. Dynamic proteomics of individual cancer cells in response to a drug. *Science*, 322(5907):1511–1516, 5 December 2008.

[20] Sam Cooper, Amine Sadok, Vicky Bousgouni, and Chris Bakal. Apolar and polar transitions drive the conversion between amoeboid and mesenchymal shapes in melanoma cells. *Mol. Biol. Cell*, 26(22):4163–4170, 5 November 2015.

[21] Pascal Craw and Wamadeva Balachandran. Isothermal nucleic acid amplification technologies for point-of-care diagnostics: a critical review. *Lab Chip*, 12(14):2469, 2012.

[22] Kalyan Das, James M Aramini, Li-Chung Ma, Robert M Krug, and Eddy Arnold. Structures of influenza a proteins and insights into antiviral drug targets. *Nat. Struct. Mol. Biol.*, 17(5):530–538, May 2010.

[23] Meghna Das Thakur, Fernando Salangsang, Allison S Landman, William R Sellers, Nancy K Pryer, Mitchell P Levesque, Reinhard Dummer, Martin McMahon, and Darrin D Stuart. Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance. *Nature*, 494(7436):251–255, 14 February 2013.

[24] Helen Davies, Graham R Bignell, Charles Cox, Philip Stephens, Sarah Edkins, Sheila Clegg, Jon Teague, Hayley Woffendin, Mathew J Garnett, William Bottomley, Neil Davis, Ed Dicks, Rebecca Ewing, Yvonne Floyd, Kristian Gray,

Sarah Hall, Rachel Hawes, Jaime Hughes, Vivian Kosmidou, Andrew Menzies, Catherine Mould, Adrian Parker, Claire Stevens, Stephen Watt, Steven Hooper, Rebecca Wilson, Hiran Jayatilake, Barry A Gusterson, Colin Cooper, Janet Shipley, Darren Hargrave, Katherine Pritchard-Jones, Norman Maitland, Georgia Chenevix-Trench, Gregory J Riggins, Darell D Bigner, Giuseppe Palmieri, Antonio Cossu, Adrienne Flanagan, Andrew Nicholson, Judy W C Ho, Suet Y Leung, Siu T Yuen, Barbara L Weber, Hilliard F Seigler, Timothy L Darrow, Hugh Paterson, Richard Marais, Christopher J Marshall, Richard Wooster, Michael R Stratton, and P Andrew Futreal. Mutations of the BRAF gene in human cancer. *Nature*, 417(6892):949–954, 27 June 2002.

[25] Nila J Dharan, Larisa V Gubareva, John J Meyer, Margaret Okomo-Adhiambo, Reginald C McClinton, Steven A Marshall, Kirsten St George, Scott Epperson, Lynnette Brammer, Alexander I Klimov, Joseph S Bresee, Alicia M Fry, and Oseltamivir-Resistance Working Group. Infections with oseltamivir-resistant influenza A(H1N1) virus in the united states. *JAMA*, 301(10):1034–1041, 11 March 2009.

[26] Luis A Diaz, Jr, Richard T Williams, Jian Wu, Isaac Kinde, J Randolph Hecht, Jordan Berlin, Benjamin Allen, Ivana Bozic, Johannes G Reiter, Martin A Nowak, Kenneth W Kinzler, Kelly S Oliner, and Bert Vogelstein. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*, 486(7404):537–540, 28 June 2012.

[27] Samudra K Dissanayake, Purevdorj B Olkhanud, Michael P O'Connell, Arnell Carter, Amanda D French, Tura C Camilli, Chineye D Emeche, Kyle J Hewitt, Devin T Rosenthal, Poloko D Leotlela, Michael S Wade, Sherry W Yang, Larry Brant, Brian J Nickoloff, Jane L Messina, Arya Biragyn, Keith S

Hoek, Dennis D Taub, Dan L Longo, Vernon K Sondak, Stephen M Hewitt, and Ashani T Weeraratna. Wnt5A regulates expression of tumor-associated antigens in melanoma via changes in signal transducers and activators of transcription 3 phosphorylation. *Cancer Res.*, 68(24):10205–10214, 15 December 2008.

[28] Joanna Dobson, Richard J Whitley, Stuart Pocock, and Arnold S Monto. Oseltamivir treatment for influenza in adults: a meta-analysis of randomised controlled trials. *Lancet*, 385(9979):1729–1737.

[29] Anne E Dodson and Jasper Rine. Heritable capture of heterochromatin dynamics in saccharomyces cerevisiae. *Elife*, 4:e05007, 12 January 2015.

[30] Margaret S Ebert, Joel R Neilson, and Phillip A Sharp. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods*, 4(9):721–726, September 2007.

[31] Ron Eccles. Understanding the symptoms of the common cold and influenza. *Lancet Infect. Dis.*, 5(11):718–725, November 2005.

[32] Imane El Meouche, Yik Siu, and Mary J Dunlop. Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Sci. Rep.*, 6:19538, 13 January 2016.

[33] Avigdor Eldar and Michael B Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, 9 September 2010.

[34] Mohammad Fallahi-Sichani, Nathan J Moerke, Mario Niepel, Tinghu Zhang, Nathanael S Gray, and Peter K Sorger. Systematic analysis of BRAF(V600E) melanomas reveals a role for JNK/c-Jun pathway in adaptive resistance to drug-induced apoptosis. *Mol. Syst. Biol.*, 11(3):797, March 2015.

[35] Andrea M Femino, Fredric S Fay, Kevin Fogarty, and Robert H Singer. Visualization of single RNA transcripts in situ. *Science*, 280(5363):585, 24 April 1998.

[36] I J Fidler and M L Kripke. Metastasis results from preexisting variant cells within a malignant tumor. *Science*, 197(4306):893–895, 26 August 1977.

[37] Keith T Flaherty, Igor Puzanov, Kevin B Kim, Antoni Ribas, Grant A McArthur, Jeffrey A Sosman, Peter J O'Dwyer, Richard J Lee, Joseph F Grippo, Keith Nolop, and Paul B Chapman. Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.*, 363(9):809–819, 26 August 2010.

[38] Peter L Frick, Bishal B Paudel, Darren R Tyson, and Vito Quaranta. Quantifying heterogeneity and dynamics of clonal fitness in response to perturbation. *J. Cell. Physiol.*, 230(7):1403–1412, July 2015.

[39] Levi A Garraway and Pasi A Jänne. Circumventing cancer drug resistance in the era of personalized medicine. *Cancer Discov.*, 2(3):214–226, March 2012.

[40] Paul Ginart, Jennifer M Kalish, Connie L Jiang, Alice C Yu, Marisa S Bartolomei, and Arjun Raj. Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. *Genes Dev.*, 30(5):567–578, 1 March 2016.

[41] Vladimir Gubala, Leanne F Harris, Antonio J Ricco, Ming X Tan, and David E Williams. Point of care diagnostics: status and future. *Anal. Chem.*, 84(2):487–515, 17 January 2012.

[42] Piyush B Gupta, Christine M Fillmore, Guozhi Jiang, Sagi D Shapira, Kai Tao, Charlotte Kuperwasser, and Eric S Lander. Stochastic state transitions give rise

to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4):633–644, 19 August 2011.

[43] Jacob Hanna, Krishanu Saha, Bernardo Pando, Jeroen van Zon, Christopher J Lengner, Menno P Creyghton, Alexander van Oudenaarden, and Rudolf Jaenisch. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*, 462(7273):595–601, 3 December 2009.

[44] Aaron N Hata, Matthew J Niederst, Hannah L Archibald, Maria Gomez-Caraballo, Faria M Siddiqui, Hillary E Mulvey, Yosef E Maruvka, Fei Ji, Hyo-Eun C Bhang, Viveksagar Krishnamurthy Radhakrishna, Giulia Siravegna, Haichuan Hu, Sana Raoof, Elizabeth Lockerman, Anuj Kalsy, Dana Lee, Celina L Keating, David A Ruddy, Leah J Damon, Adam S Crystal, Carlotta Costa, Zofia Piotrowska, Alberto Bardelli, Anthony J Iafrate, Ruslan I Sadreyev, Frank Stegmeier, Gad Getz, Lecia V Sequist, Anthony C Faber, and Jeffrey A Engelman. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.*, 1 February 2016.

[45] Daniel Hebenstreit, Miaoqing Fang, Muxin Gu, Varodom Charoensawan, Alexander van Oudenaarden, and Sarah A Teichmann. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.*, 7:497, 1 January 2011.

[46] Eishu Hirata, Maria Romina Girotti, Amaya Viros, Steven Hooper, Bradley Spencer-Dene, Michiyuki Matsuda, James Larkin, Richard Marais, and Erik Sahai. Intravital imaging reveals how BRAF inhibition generates drug-tolerant microenvironments with high integrin β1/FAK signaling. *Cancer Cell*, 27(4):574–588, 13 April 2015.

[47] Keith S Hoek, Ossia M Eichhoff, Natalie C Schlegel, Udo Döbbeling, Nikita Kobert, Leo Schaerer, Silvio Hemmi, and Reinhard Dummer. In vivo switching of human melanoma cells between proliferative and invasive states. *Cancer Res.*, 68(3):650–656, 1 February 2008.

[48] Keith S Hoek and Colin R Goding. Cancer stem cells versus phenotype-switching in melanoma. *Pigment Cell Melanoma Res.*, 23(6):746–759, December 2010.

[49] Christine E Holt and Simon L Bullock. Subcellular mRNA localization in animal cells and why it matters. *Science*, 326(5957):1212, 26 November 2009.

[50] Willy Hugo, Hubing Shi, Lu Sun, Marco Piva, Chunying Song, Xiangju Kong, Gatien Moriceau, Aayoung Hong, Kimberly B Dahlman, Douglas B Johnson, Jeffrey A Sosman, Antoni Ribas, and Roger S Lo. Non-genomic and immune evolution of melanoma acquiring MAPKi resistance. *Cell*, 162(6):1271–1285, 10 September 2015.

[51] Shalev Itzkovitz, Anna Lyubimova, Irene C Blat, Mindy Maynard, Johan van Es, Jacqueline Lees, Tyler Jacks, Hans Clevers, and Alexander van Oudenaarden. Single-molecule transcript counting of stem-cell markers in the mouse intestine. *Nat. Cell Biol.*, 14(1):106–114, January 2012.

[52] Leni S Jacob, Sakari Vanharanta, Anna C Obenauf, Mono Pirun, Agnes Viale, Nicholas D Socci, and Joan Massagué. Metastatic competence can emerge with selection of preexisting oncogenic alleles without a need of new mutations. *Cancer Res.*, 75(18):3713–3719, 15 September 2015.

[53] Lan Jiang, Huidong Chen, Luca Pinello, and Guo-Cheng Yuan. GiniClust: detecting rare cell types from single-cell gene expression data with gini index. *Genome Biol.*, 17(1):144, 2016.

[54] Min Hwan Kim, Jongshin Kim, Hyowon Hong, Si-Hyung Lee, June-Koo Lee, Eunji Jung, and Joon Kim. Actin remodeling confers BRAF inhibitor resistance to melanoma cells through YAP/TAZ activation. *EMBO J.*, 14 December 2015.

[55] Clemens Krepler, Min Xiao, Katrin Spoesser, Patricia A Brafford, Batool Shannan, Marilda Beqiri, Qin Liu, Wei Xu, Bradley Garman, Katherine L Nathanson, Xaiowei Xu, Giorgos C Karakousis, Gordon B Mills, Yiling Lu, Tamer A Ahmed, Poulikos Poulikakos, Giordano Caponigro, Markus Boehm, Malte Peters, Lynn M Schuchter, Ashani T Weeraratna, and Meenhard Herlyn. Personalized pre-clinical trials in BRAF inhibitor resistant patient derived xenograft models identify second line combination therapies. *Clin. Cancer Res.*, 16 December 2015.

[56] Swati Kumar and Kelly J Henrickson. Update on influenza diagnostics: lessons from the novel H1N1 influenza a pandemic. *Clin. Microbiol. Rev.*, 25(2):344–361, April 2012.

[57] Marshall J Levesque, Paul Ginart, Yichen Wei, and Arjun Raj. Visualizing SNVs to quantify allele-specific expression in single cells. *Nat. Methods*, 10(9):865–867, September 2013.

[58] Marshall J Levesque and Arjun Raj. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods*, 10(3):246–248, March 2013.

[59] David Liao, Luis Estévez-Salmerón, and Thea D Tlsty. Conceptualizing a tool to optimize therapy based on dynamic heterogeneitythe authors dedicate this paper to dr barton kamen who inspired its initiation and enthusiastically supported its pursuit. *Phys. Biol.*, 9(6):065005, 29 November 2012.

[60] David Liao, Luis Estévez-Salmerón, and Thea D Tlsty. Generalized principles of stochasticity can be used to control dynamic heterogeneity. *Phys. Biol.*, 9(6):065006, December 2012.

[61] Susanne L Linderman, Benjamin S Chambers, Seth J Zost, Kaela Parkhouse, Yang Li, Christin Herrmann, Ali H Ellebedy, Donald M Carter, Sarah F Andrews, Nai-Ying Zheng, Min Huang, Yunping Huang, Donna Strauss, Beth H Shaz, Richard L Hodinka, Gustavo Reyes-Terán, Ted M Ross, Patrick C Wilson, Rafi Ahmed, Jesse D Bloom, and Scott E Hensley. Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013–2014 influenza season. *Proceedings of the National Academy of Sciences*, 111(44):15798–15803, 4 November 2014.

[62] Xiangfan Liu, Huapeng Li, Mihir Rajurkar, Qi Li, Jennifer L Cotton, Jianhong Ou, Lihua J Zhu, Hira L Goel, Arthur M Mercurio, Joo-Seop Park, Roger J Davis, and Junhao Mao. Tead and AP1 coordinate transcription and motility. *Cell Rep.*, 0(0), 28 January 2016.

[63] Michael Loeffelholz and Tasnee Chonmaitree. Advances in diagnosis of respiratory virus infections. *Int. J. Microbiol.*, 2010:126049, 19 October 2010.

[64] Eric Lubeck and Long Cai. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods*, 9(7):743–748, July 2012.

[65] Eric Lubeck, Ahmet F Coskun, Timur Zhiyentayev, Mubhij Ahmad, and Long Cai. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods*, 11(4):360–361, April 2014.

[66] S E Luria and M Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28(6):491–511, November 1943.

[67] Hédia Maamar, Moran N Cabili, John Rinn, and Arjun Raj. linc-HOXA1 is a noncoding RNA that represses hoxa1 transcription in cis. *Genes Dev.*, 27(11):1260–1271, 1 June 2013.

[68] Hédia Maamar, Arjun Raj, and David Dubnau. Noise in gene expression determines cell fate in bacillus subtilis. *Science*, 317(5837):526–529, 27 July 2007.

[69] James B Mahony, Astrid Petrich, and Marek Smieja. Molecular diagnosis of respiratory virus infections. *Crit. Rev. Clin. Lab. Sci.*, 48(5-6):217–249, September 2011.

[70] Le Quynh Mai, Heiman F L Wertheim, Tran Nhu Duong, H Rogier van Doorn, Nguyen Tran Hien, and Peter Horby. A community cluster of Oseltamivir-Resistant cases of 2009 H1N1 influenza. *N. Engl. J. Med.*, 362(1):86–87, 2010.

[71] Kelsey C Martin and Anne Ephrussi. mRNA localization: gene expression in the spatial dimension. *Cell*, 136(4):719–730, 20 February 2009.

[72] Luis Martínez-Sobrido and Adolfo García-Sastre. Generation of recombinant influenza virus from plasmid DNA. *J. Vis. Exp.*, (42), 3 August 2010.

[73] Peter Meister, Benjamin D Towbin, Brietta L Pike, Aaron Ponti, and Susan M Gasser. The spatial dynamics of tissue-specific promoters during c. elegans development. *Genes Dev.*, 24(8):766–782, 15 April 2010.

[74] Judith Müller, Oscar Krijgsman, Jennifer Tsoi, Lidia Robert, Willy Hugo, Chunying Song, Xiangju Kong, Patricia A Possik, Paulien D M Cornelissen-Steijger, Marnix H Geukes Foppen, Kristel Kemper, Colin R Goding, Ultan McDermott, Christian Blank, John Haanen, Thomas G Graeber, Antoni Ribas, Roger S Lo,

and Daniel S Peeper. Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat. Commun.*, 5:5712, 15 December 2014.

[75] Melaku Muluneh, Wu Shang, and David Issadore. Track-etched magnetic micropores for immunomagnetic isolation of pathogens. *Adv. Healthc. Mater.*, 3(7):1078–1085, July 2014.

[76] Ramin Nazarian, Hubing Shi, Qi Wang, Xiangju Kong, Richard C Koya, Hane Lee, Zugen Chen, Mi-Kyung Lee, Narsis Attar, Hooman Sazegar, Thinle Chodon, Stanley F Nelson, Grant McArthur, Jeffrey A Sosman, Antoni Ribas, and Roger S Lo. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature*, 468(7326):973–977, 16 December 2010.

[77] Alexander Nguyen, Mitsukuni Yoshida, Hani Goodarzi, and Sohail F Tavazoie. Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. *Nat. Commun.*, 7:11246, 3 May 2016.

[78] Shuping Nie, Richard B Roth, Jeffrey Stiles, Albina Mikhlina, Xuedong Lu, Yi-Wei Tang, and N Esther Babady. Evaluation of alere i influenza A&B for rapid detection of influenza viruses a and B. *J. Clin. Microbiol.*, 52(9):3339–3344, September 2014.

[79] Daan Noordermeer, Elzo de Wit, Petra Klous, Harmen van de Werken, Marieke Simonis, Melissa Lopez-Jones, Bert Eussen, Annelies de Klein, Robert H Singer, and Wouter de Laat. Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat. Cell Biol.*, 13(8):944–951, 1 August 2011.

[80] Peter C Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28, 1 October 1976.

[81] Michael P O'Connell, Katie Marchbank, Marie R Webster, Alexander A Valiga, Amanpreet Kaur, Adina Vultur, Ling Li, Meenhard Herlyn, Jessie Villanueva, Qin Liu, Xiangfan Yin, Sandy Widura, Janelle Nelson, Nivia Ruiz, Tura C Camilli, Fred E Indig, Keith T Flaherty, Jennifer A Wargo, Dennie T Frederick, Zachary A Cooper, Suresh Nair, Ravi K Amaravadi, Lynn M Schuchter, Giorgos C Karakousis, Wei Xu, Xiaowei Xu, and Ashani T Weeraratna. Hypoxia induces phenotypic plasticity and therapy resistance in melanoma via the tyrosine kinase receptors ROR1 and ROR2. *Cancer Discov.*, 3(12):1378–1393, December 2013.

[82] Darwin J Operario, Michael J Moser, and Kirsten St George. Highly sensitive and quantitative detection of the H274Y oseltamivir resistance mutation in seasonal A/H1N1 influenza virus. *J. Clin. Microbiol.*, 48(10):3517–3524, October 2010.

[83] Olivia Padovan-Merhar, Gautham P Nair, Andrew G Biaesch, Andreas Mayer, Steven Scarfone, Shawn W Foley, Angela R Wu, L Stirling Churchman, Abhyudai Singh, and Arjun Raj. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell*, 58(2):339–352, 16 April 2015.

[84] A O Pisco and S Huang. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: 'what does not kill me strengthens me'. *Br. J. Cancer*, 112(11):1725–1732, 26 May 2015.

[85] A O Pisco and S Huang. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: what does not kill me strengthens me. *Br. J. Cancer*, 112(11):1725–1732, 12 May 2015.

[86] Angela Oliveira Pisco, Amy Brock, Joseph Zhou, Andreas Moor, Mitra Mojtahedi,

Dean Jackson, and Sui Huang. Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat. Commun.*, 4:2467, 2013.

[87] Leo L M Poon, K H Chan, G J Smith, C S W Leung, Y Guan, K Y Yuen, and J S M Peiris. Molecular detection of a novel human influenza (H1N1) of pandemic potential by conventional and real-time quantitative RT-PCR assays. *Clin. Chem.*, 55(8):1555–1558, August 2009.

[88] Arjun Raj, Scott A Rifkin, Erik Andersen, and Alexander van Oudenaarden. Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283):913–918, 18 February 2010.

[89] Arjun Raj and Sanjay Tyagi. Chapter 17 - detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. In Nils G. Walter, editor, *Single Molecule Tools: Fluorescence Based Approaches, Part A*, volume Volume 472, pages 365–386. Academic Press, 2010.

[90] Arjun Raj and Sanjay Tyagi. Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Methods Enzymol.*, 472:365–386, 2010.

[91] Arjun Raj, Patrick van den Bogaard, Scott A Rifkin, Alexander van Oudenaarden, and Sanjay Tyagi. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods*, 5(10):877–879, 21 September 2008.

[92] Arjun Raj and Alexander van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 17 October 2008.

[93] Michael Ramirez, Satwik Rajaram, Robert J Steininger, Daria Osipchuk, Maike A

Roth, Leanna S Morinishi, Louise Evans, Weiyue Ji, Chien-Hsiang Hsu, Kevin Thurley, Shuguang Wei, Anwu Zhou, Prasad R Koduru, Bruce A Posner, Lani F Wu, and Steven J Altschuler. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat. Commun.*, 7:10690, 19 February 2016.

[94] Rachel Ramsdale, Robert N Jorissen, Frederic Z Li, Sheren Al-Obaidi, Teresa Ward, Karen E Sheppard, Patricia E Bukczynska, Richard J Young, Samantha E Boyle, Mark Shackleton, Gideon Bollag, Georgina V Long, Eugene Tulchinsky, Helen Rizos, Richard B Pearson, Grant A McArthur, Amardeep S Dhillon, and Petranel T Ferrao. The transcription cofactor c-JUN mediates phenotype switching and BRAF inhibitor resistance in melanoma. *Sci. Signal.*, 8(390):ra82, 18 August 2015.

[95] Philipp Rathert, Mareike Roth, Tobias Neumann, Felix Muerdter, Jae-Seok Roe, Matthias Muhar, Sumit Deswal, Sabine Cerny-Reiterer, Barbara Peter, Julian Jude, Thomas Hoffmann, Łukasz M Boryń, Elin Axelsson, Norbert Schweifer, Ulrike Tontsch-Grunt, Lukas E Dow, Davide Gianni, Mark Pearson, Peter Valent, Alexander Stark, Norbert Kraut, Christopher R Vakoc, and Johannes Zuber. Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature*, 525(7570):543–547, 24 September 2015.

[96] D Ravindran Menon, S Das, C Krepler, A Vultur, B Rinner, S Schauer, K Kashofer, K Wagner, G Zhang, E Bonyadi Rad, N K Haass, H P Soyer, B Gabrielli, R Somasundaram, G Hoefler, M Herlyn, and H Schaider. A stress-induced early innate response causes multidrug tolerance in melanoma. *Oncogene*, 24 November 2014.

[97] Stefanie Riesenberg, Angela Groetchen, Robert Siddaway, Tobias Bald, Julia Reinhardt, Denise Smorra, Judith Kohlmeyer, Marcel Renn, Bengt Phung, Pia Aymans, Tobias Schmidt, Veit Hornung, Irwin Davidson, Colin R Goding, Göran Jönsson, Jennifer Landsberg, Thomas Tüting, and Michael Hölzel. MITF and c-jun antagonism interconnects melanoma dedifferentiation with pro-inflammatory cytokine responsiveness and myeloid cell recruitment. *Nat. Commun.*, 6:8755, 4 November 2015.

[98] Alexander Roesch, Mizuho Fukunaga-Kalabis, Elizabeth C Schmidt, Susan E Zabierowski, Patricia A Brafford, Adina Vultur, Devraj Basu, Phyllis Gimotty, Thomas Vogt, and Meenhard Herlyn. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell*, 141(4):583–594, 14 May 2010.

[99] Alexander Roesch, Adina Vultur, Ivan Bogeski, Huan Wang, Katharina M Zimmermann, David Speicher, Christina Körbel, Matthias W Laschke, Phyllis A Gimotty, Stephan E Philipp, Elmar Krause, Sylvie Pätzold, Jessie Villanueva, Clemens Krepler, Mizuho Fukunaga-Kalabis, Markus Hoth, Boris C Bastian, Thomas Vogt, and Meenhard Herlyn. Overcoming intrinsic multidrug resistance in melanoma by blocking the mitochondrial respiratory chain of slow-cycling JARID1B(high) cells. *Cancer Cell*, 23(6):811–825, 10 June 2013.

[100] Alexander Roesch, Adina Vultur, Ivan Bogeski, Huan Wang, Katharina M Zimmermann, David Speicher, Christina Körbel, Matthias W Laschke, Phyllis A Gimotty, Stephan E Philipp, Elmar Krause, Sylvie Pätzold, Jessie Villanueva, Clemens Krepler, Mizuho Fukunaga-Kalabis, Markus Hoth, Boris C Bastian, Thomas Vogt, and Meenhard Herlyn. Overcoming intrinsic multidrug resistance

in melanoma by blocking the mitochondrial respiratory chain of slow-cycling JARID1B(high) cells. *Cancer Cell*, 23(6):811–825, 10 June 2013.

[101] Alvaro Sanchez, Sandeep Choubey, and Jane Kondev. Regulation of noise in gene expression. *Annu. Rev. Biophys.*, 42:469–491, 21 March 2013.

[102] Alvaro Sanchez and Ido Golding. Genetic determinants and cellular constraints in noisy gene expression. *Science*, 342(6163):1188–1193, 6 December 2013.

[103] Amélie Clémentine Seghers, Sofie Wilgenhof, Céleste Lebbé, and Bart Neyns. Successful rechallenge in two patients with BRAF-V600-mutant melanoma who experienced previous progression during treatment with a selective BRAF inhibitor. *Melanoma Res.*, 22(6):466–472, December 2012.

[104] Leonardo A Sepúlveda, Heng Xu, Jing Zhang, Mengyu Wang, and Ido Golding. Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science*, 351(6278):1218–1222, 11 March 2016.

[105] Sydney M Shaffer, Rohan P Joshi, Benjamin S Chambers, David Sterken, Andrew G Biaesch, David J Gabrieli, Yang Li, Kristen A Feemster, Scott E Hensley, David Issadore, and Arjun Raj. Multiplexed detection of viral infections using rapid in situ RNA analysis on a chip. *Lab Chip*, 15(15):3170–3182, 17 June 2015.

[106] Sydney M Shaffer, Min-Tzu Wu, Marshall J Levesque, and Arjun Raj. Turbo FISH: a method for rapid single molecule RNA FISH. *PLoS One*, 8(9):e75120, 16 September 2013.

[107] Sreenath V Sharma, Diana Y Lee, Bihua Li, Margaret P Quinlan, Fumiyuki Takahashi, Shyamala Maheswaran, Ultan McDermott, Nancy Azizian, Lee Zou, Michael A Fischbach, Kwok-Kin Wong, Kathleyn Brandstetter, Ben Wittner,

Sridhar Ramaswamy, Marie Classon, and Jeff Settleman. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1):69–80, 2 April 2010.

[108] Hubing Shi, Willy Hugo, Xiangju Kong, Aayoung Hong, Richard C Koya, Gatien Moriceau, Thinle Chodon, Rongqing Guo, Douglas B Johnson, Kimberly B Dahlman, Mark C Kelley, Richard F Kefford, Bartosz Chmielowski, John A Glaspy, Jeffrey A Sosman, Nicolas van Baren, Georgina V Long, Antoni Ribas, and Roger S Lo. Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov.*, 4(1):80–93, January 2014.

[109] Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.*, 7:539, 11 October 2011.

[110] Samuel O Skinner, Heng Xu, Sonal Nagarkar-Jaiswal, Pablo R Freire, Thomas P Zwaka, and Ido Golding. Single-cell analysis of transcription kinetics across the cell cycle. *Elife*, 5:e12175, 29 January 2016.

[111] Sabrina L Spencer, Suzanne Gaudet, John G Albeck, John M Burke, and Peter K Sorger. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, 459(7245):428–432, 21 May 2009.

[112] Gürol M Süel, Rajan P Kulkarni, Jonathan Dworkin, Jordi Garcia-Ojalvo, and Michael B Elowitz. Tunability and noise dependence in differentiation dynamics. *Science*, 315(5819):1716–1719, 23 March 2007.

[113] Chong Sun, Liqin Wang, Sidong Huang, Guus J J E Heynen, Anirudh Prahallad, Caroline Robert, John Haanen, Christian Blank, Jelle Wesseling, Stefan M Willems, Davide Zecchin, Sebastijan Hobor, Prashanth K Bajpe, Cor Lieftink, Christina Mateus, Stephan Vagner, Wipawadee Grernrum, Ingrid Hofland, Andreas Schlicker, Lodewyk F A Wessels, Roderick L Beijersbergen, Alberto Bardelli, Federica Di Nicolantonio, Alexander M M Eggermont, and Rene Bernards. Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma. *Nature*, 508(7494):118–122, 3 April 2014.

[114] Orsolya Symmons and Arjun Raj. Whats luck got to do with it: Single cells, multiple fates, and biological nondeterminism. *Mol. Cell*, 62(5):788–802, 2 June 2016.

[115] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676, 25 August 2006.

[116] Y W Tang, G W Procop, and D H Persing. Molecular diagnostics of infectious diseases. *Clin. Chem.*, 43(11):2021–2038, November 1997.

[117] Itay Tirosh, Benjamin Izar, Sanjay M Prakadan, Marc H Wadsworth, 2nd, Daniel Treacy, John J Trombetta, Asaf Rotem, Christopher Rodman, Christine Lian, George Murphy, Mohammad Fallahi-Sichani, Ken Dutton-Regester, Jia-Ren Lin, Ofir Cohen, Parin Shah, Diana Lu, Alex S Genshaft, Travis K Hughes, Carly G K Ziegler, Samuel W Kazer, Aleth Gaillard, Kellie E Kolb, Alexandra-Chloé Villani, Cory M Johannessen, Aleksandr Y Andreev, Eliezer M Van Allen, Monica Bertagnolli, Peter K Sorger, Ryan J Sullivan, Keith T Flaherty, Dennie T Frederick, Judit Jané-Valbuena, Charles H Yoon, Orit Rozenblatt-Rosen, Alex K

Shalek, Aviv Regev, and Levi A Garraway. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, 352(6282):189–196, 8 April 2016.

[118] T D Tlsty, B H Margolin, and K Lum. Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria-Delbrück fluctuation analysis. *Proc. Natl. Acad. Sci. U. S. A.*, 86(23):9441–9445, December 1989.

[119] John J Treanor, Frederick G Hayden, Peter S Vrooman, Rick Barbarash, Robert Bettis, Dennis Riff, Sudeep Singh, Nelson Kinnersley, Penelope Ward, Roger G Mills, and for the US Oral Neuraminidase Study Group. Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: A randomized controlled trial. *JAMA*, 283(8):1016–1024, 23 February 2000.

[120] Kerstin Trunzer, Anna C Pavlick, Lynn Schuchter, Rene Gonzalez, Grant A McArthur, Thomas E Hutson, Stergios J Moschos, Keith T Flaherty, Kevin B Kim, Jeffrey S Weber, Peter Hersey, Georgina V Long, Donald Lawrence, Patrick A Ott, Ravi K Amaravadi, Karl D Lewis, Igor Puzanov, Roger S Lo, Astrid Koehler, Mark Kockx, Olivia Spleiss, Annette Schell-Steven, Houston N Gilbert, Louise Cockey, Gideon Bollag, Richard J Lee, Andrew K Joe, Jeffrey A Sosman, and Antoni Ribas. Pharmacodynamic effects and mechanisms of resistance to vemurafenib in patients with metastatic melanoma. *J. Clin. Oncol.*, 31(14):1767–1774, 10 May 2013.

[121] Nicholas C Turner and Jorge S Reis-Filho. Genetic heterogeneity and cancer drug resistance. *Lancet Oncol.*, 13(4):e178–85, April 2012.

[122] Darren R Tyson, Shawn P Garbett, Peter L Frick, and Vito Quaranta. Fractional

proliferation: a method to deconvolve cell population dynamics from single-cell data. *Nat. Methods*, 9(9):923–928, September 2012.

[123] Annelien Verfaillie, Hana Imrichova, Zeynep Kalender Atak, Michael Dewaele, Florian Rambow, Gert Hulselmans, Valerie Christiaens, Dmitry Svetlichnyy, Flavie Luciani, Laura Van den Mooter, Sofie Claerhout, Mark Fiers, Fabrice Journe, Ghanem-Elias Ghanem, Carl Herrmann, Georg Halder, Jean-Christophe Marine, and Stein Aerts. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat. Commun.*, 6:6683, 9 April 2015.

[124] Nikhil Wagle, Caroline Emery, Michael F Berger, Matthew J Davis, Allison Sawyer, Panisa Pochanard, Sarah M Kehoe, Cory M Johannessen, Laura E Macconaill, William C Hahn, Matthew Meyerson, and Levi A Garraway. Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J. Clin. Oncol.*, 29(22):3085–3096, 1 August 2011.

[125] Ruixue Wang, Zong-Mei Sheng, and Jeffery K Taubenberger. Detection of novel (swine origin) H1N1 influenza a virus by quantitative real-time reverse transcription-PCR. *J. Clin. Microbiol.*, 47(8):2675–2677, August 2009.

[126] Marie R Webster and Ashani T Weeraratna. A wnt-er migration: the confusing role of β-catenin in melanoma metastasis. *Sci. Signal.*, 6(268):e11, 26 March 2013.

[127] Wei Wang, Zhigang Song, Wencai Guan, Yi Liu, Xiaonan Zhang, Lei Xu, Jianhua Li, Zhenghong Yuan, and Yunwen Hu. PCR for detection of oseltamivir resistance mutation in influenza A(H7N9) virus. *Emerging Infectious Disease journal*, 20(5):847, 2014.

[128] Mathias F Wernet, Esteban O Mazzoni, Arzu Celik, Dianne M Duncan, Ian Duncan, and Claude Desplan. Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, 440(7081):174–180, 9 March 2006.

[129] David M Whiley, Seweryn Bialasiewicz, Cheryl Bletchly, Cassandra E Faux, Bruce Harrower, Allan R Gould, Stephen B Lambert, Graeme R Nimmo, Michael D Nissen, and Theo P Sloots. Detection of novel influenza A(H1N1) virus by real-time RT-PCR. *J. Clin. Virol.*, 45(3):203–204, July 2009.

[130] Samuel Yang and Richard E Rothman. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect. Dis.*, 4(6):337–348, June 2004.

[131] David Yu Zhang and Erik Winfree. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.*, 131(47):17303–17314, 2 December 2009.

[132] Hongying Zhu, Serhan O Isikman, Onur Mudanyali, Alon Greenbaum, and Aydogan Ozcan. Optical imaging techniques for point-of-care diagnostics. *Lab Chip*, 13(1):51–67, 2013.