bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pairwise tests for conditional selection use evolutionary logic to predict intrinsic drug resistance in ALK alterations

Haider Inam[1], Justin Pritchard[1,2,‡]

1. The Pennsylvania State University, Department of Biomedical Engineering, University Park PA, 16802.
2. The Huck Institute for the Life Sciences, Center for the Resistance Evolution, The Pennsylvania State University

‡. To whom correspondence should be addressed: jrp94@psu.edu

Running Title: Pairwise conditional selection predicts intrinsic resistance

Abstract:

Background: Genomic data can facilitate personalized treatment decisions by enabling therapeutic hypotheses in individual patients. Conditional selection (encompassing both mutual exclusivity and co-occurrence) is commonly used to consider rare genomic variants relative to established cancer drivers. However, the direct comparison of p-values across multiple pairs of genes is confounded by the often-dramatic differences in sample size between established driver mutations and novel findings.

Methods: We develop a resampling based method for the direct pairwise comparisons of conditional selection between sets of gene pairs. This effectively creates quantitative positive control guideposts of conditional selection that normalize differences in population size. We applied this method to a transcript variant of ALK in melanoma, termed ALK^{ATI}, which has been the subject of a recent controversy in the literature. We reproduced some of the original cell transformation experiments, performed rescue experiments, and analyzed drug response data to revisit the original ALK^{ATI} findings.

Finding: We found that we are powered to quantitatively compare the degree of relative mutual exclusivity (down to an abundance of 10 patients in a cohort of 500) between novel gene variants and positive controls. We also found that ALK^{ATI} is not as mutually exclusive as BRAF and NRAS are with each other. Our *in vitro* transformation assays and rescue assays suggested that alternative transcript initiation in ALK is not likely to be necessary or sufficient for cellular transformation or growth.

Interpretation: Pairwise comparisons of conditional selection represent a sensitive method of utilizing existing genomic data to directly and quantitatively compare relative levels of conditional selection. The results of using our method in ALK^{ATI} and our experiments strongly disfavor the role of ALK^{ATI} as a targetable oncogenic driver. The progress of other experimental agents in late stage melanoma and our experimental and computational re-analysis led us to conclude that further single agent testing of ALK inhibitors in patients with ALK^{ATI} should be limited to cases where no other treatment hypotheses can be identified.

Correspondence to:
Justin Pritchard
211 Wartik Lab
The Pennsylvania State University
University Park, PA 16802
Jrp94@psu.edu

Data and Code Availability: https://pritchardlabatpsu.github.io/pairwisecomparisons/

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

In cancers, clonal selection influences tumor progression and responses to therapy (1), but in every patient, the process of clonal selection creates independent tumors with independent, parallel evolutionary trajectories. These parallel evolutionary paths can be conditioned upon the occurrence of a previous genetic event. Co-occurrence is when the first variant predicts the presence of a second (2,3). Whereas, mutual exclusivity occurs when the presence of the first predicts the absence of the second—and is the primary focus here. Mutual exclusivity is often viewed qualitatively or used to build large scale networks (2,4–6). Here we aim to create quantitative guideposts of conditional selection that will allow for direct comparisons of mutual exclusivity relative to known druggable oncogene pairs by controlling for cohort size. These guideposts can be used to triage the translational actionability of rare genomic findings.

A strong example of mutual exclusivity in cancer is seen in the evolutionarily ancient mitogen-activated protein kinase (MAPK) pathway that is present from yeast to metazoans (7). In higher eukaryotes, the MAPK pathway is often canonically activated by receptor tyrosine kinases (RTK) that signal through MAPKs to achieve cellular growth and development (8,9). Because of their critical role in growth and division, most known RTK-MAPK mutational events that drive cancer growth lead to mutual exclusivity across patients during parallel evolution. Moreover, many of the most impressive success stories for targeted cancer therapy in the past two decades have centered on this one pathway. Multiple examples of mutual exclusivity have been found between ALK-fusion, EGFR, KRAS, and ERBB2 genes in non-small cell lung cancer (NSCLC) (10–12). This has led to inhibitors of ALK, EGFR, and recently $KRAS_{G12C}$ offering high rates of single agent therapeutic responses (13–17). In thyroid cancer, BRAF mutations have been found to be mutually exclusive with RET fusions, and both have shown high single agent response rates in clinical trials (18). Finally BRAF & NRAS are mutually exclusive in melanoma and single agent use of vemurafenib leads to clinical responses (19–25). Therefore, mutually exclusive activating mutations in the RTK-MAPK pathway have shown impressive clinical sensitivity to single agent targeted therapies (15,26).

There are at least two explanations for mutual exclusivity in gain of function oncogenes in the same pathway: 1) Variants that are sufficient to activate the pathway may be functionally redundant. In this case there is no added benefit to a second activating mutation. 2) If activation of the pathway is sufficient, further activation of that pathway might be harmful to the cell i.e. the two events in the same pathway may be antagonistic because excess pathway activity is selected against (27,28).

While mutual exclusivity is often measured in cancer genomics studies, it can be challenging to interpret. For instance, KRAS and EGFR are well described as mutually exclusive in the literature, and their inhibitors have shown single agent response rates. If a newly discovered variant appears to be mutually exclusive with one or both of these genes, that would give confidence that the variant is important. However, when the variant is not mutually exclusive, the important question becomes, how do we demonstrate a negative result? Answering this question requires estimating the variability in effect size and significance that rarity would bring to positive control genes like KRAS and EGFR. In the absence of such a quantitative test, it is easy to view the data qualitatively (29). Here we seek a direct and quantitative test to rapidly interpret the relative mutual exclusivity of rare variants versus a positive control pair. Since we use positive controls that are genes in the MAPK pathway with previously verified single agent responses to targeted therapeutics (14–16,30), our method can also be used to triage rare genomic findings for potential single agent drug efficacy. If a newly discovered alteration is as mutually exclusive as a therapeutically established mutually exclusive gene pair, it gives greater confidence in the potential for single agent therapy.

As a test case for a method that can quantitatively compare mutual exclusivity between gene pairs, we re-examined the observation that alternative transcription initiation in ALK, termed ALK^{ATI}, is a therapeutically actionable oncogenic target (31). This novel ALK transcript has a transcription initiation site in intron 19 of ALK, just upstream of ALK's kinase domain. Wiesner *et al* posited that this transcript exhibits a novel mechanism of oncogenic activation by overexpressing the kinase domain of ALK. Using *in vitro* transformation assays and inhibitor treatment, they hypothesize that single agent ALK-inhibitor therapy can treat the 2-11% (31,32) of melanoma patients expressing ALK^{ATI}. However, Couts *et al* reported crizotinib (an ALK inhibitor) sensitivity in melanoma patient derived xenografts (PDXs) expressing EML4-ALK but not ALK^{ATI} *in vitro* and *in vivo* (33). Exome sequencing of the melanoma PDXs expressing ALK^{ATI} revealed that 4 out of 6 of these cell lines had transforming mutations in well-established melanoma oncogenes. Furthermore, out of the two ALK^{ATI} melanoma patients treated with ALK-inhibitors so far, one had a modest response that did not rise to the level of an objective response (31) and the other did not

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

respond (33). This literature controversy, and the small sample size of the Couts *et al*. study suggest that reanalysis of the original observation of ALK^{ATI} is important to understanding if further investigation is warranted.

In this paper, we develop a simple and user-friendly resampling method to test the relative conditional selection across sets of gene pairs. This will allow us to quantitatively compare any gene of interest (GOI) such as ALK^{ATI} with positive control gene pairs like BRAF and NRAS in melanoma. Using this method, we show that ALK^{ATI} is not as mutually exclusive with BRAF or NRAS as BRAF or NRAS are with each other. Moreover, by large scale repetition of experiments performed in the Wiesner paper, we uncovered mutations in ALK^{ATI} that suggest that it is not sufficient for growth factor independent transformation. We also find that ALK^{ATI} cannot compensate for oncogenic signaling in melanoma cells, and that ALK kinase domain overexpression does not predict ALK inhibitor sensitivity in the cancer cell line encyclopedia (CCLE) (34). Given the stunning advances in melanoma therapy, and the rich clinical trial landscape, we suggest that relapse/refractory melanoma patients with ALK^{ATI} should not be given ALK inhibitors in an investigational or off-label capacity unless no other options remain.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Pairwise comparisons of conditional selection allow direct comparisons between frequent and rare events

Highly mutually exclusive gene pairs such as KRAS/EGFR, EGFR/EML4-ALK , and NRAS/BRAF predict the success of ALK, EGFR and BRAF inhibition in the clinic (15,26). These mutually exclusive gene pairs are positive controls that could be compared with any gene/alteration of interest (GOI, in Fig 1A). A direct pairwise comparison would allow us to ask whether a "BRAF/NRAS level" of mutual exclusivity exists between any new GOI (such as ALK^{ATI}) and BRAF or NRAS (BRAF and NRAS are an example of a highly mutually exclusive gene pair in melanoma, 20–24) . However, differences in sample size due to mutation frequency complicates our ability to directly compare the p-values and odds ratios between distinct gene pairs. If the GOI is much less abundant than the positive control, how often would the positive control have a similar odds ratio and p-value if it was only as abundant as the test gene of interest? Resampling will allow us to compare mutual exclusivities between pairs of genes with unequal abundances in cancer genomes.

To directly illustrate this idea, we analyzed how sample size affects the p-values and odds ratios observed for a simulated mutually exclusive positive control gene pair. We simulated a mock clinical cohort with two abundant but mutually exclusive genes (odds ratio of 0.025), and performed resampling experiments at different subsample sizes. As expected, decreasing the abundance of one of the genes decreases the measured significance (Fig 1B, left). We also noted a broader range in the observed odds ratios (OR) at smaller sample sizes: OR of 0-0.7 for a subsample size of 5 and OR of 0-0.11 for a subsample size of 20 (Fig 1B, right). Therefore, the range of odds ratios that is observable in highly mutually exclusive genes becomes noisier as sample size decreases. To more directly compare gene pairs with different abundances, we propose resampling the more prevalent gene pairs at an abundance that is matched by the gene of interest.

We term this idea, "pairwise comparisons of conditional selection". We aim to directly compare two gene pairs while controlling for the differences in sample size. Given a highly abundant and mutually exclusive positive control gene pair (Fig 1A, top-right) and a GOI of unknown mutual exclusivity with either of the positive control genes (Fig 1A, top-left), our method uses resampling to create a cohort in which the positive control genes have a similar abundance and frequency as those of the GOI (Fig 1A, bottom). This is done by subsampling the positive control to the same number of hits (abundance) and the same proportion in the cohort (frequency) as the GOI. Alongside this resampling, we also perform bootstrap sampling of the GOI and the other positive control genes. Details of the pairwise comparison strategy are included in the methods, and our GitHub repository.

To explore the sensitivity and specificity of our proposed method, we simulated gene sets (see methods) with varying cohort sizes, GOI abundances, and mutual exclusivity odds ratios (Supplementary Fig 1A). As expected, when we simulated two gene pairs with similar mutual exclusivities, no significant differences were detected (Fig 1C, left panel). Furthermore, large differences in simulated odds ratios required relatively fewer observations of a genetic event. (Fig 1C, Supplementary Fig 1B). Given simulated gene pairs with a range of different mutual exclusivities (middle and right panel of Fig 1C), a statistically significant difference (red asterisks) in their p-value distributions could be observed when the GOI was present in at least 10 patients in a simulated 500 patient cohort. Thus, in a cohort size of 500 and simulated ORs of 0.1-1, an abundance of ~10 events are needed to conclude that a quantitative difference in mutual exclusivity between two pairs of genes exists.

Our method is sensitive enough to detect meaningful differences in pairwise comparisons of mutual exclusivity in ~10/500 (2%) patients in a clinical cohort. Given that this sensitivity should be sufficient for the observed frequency of ALK^{ATI} (2-10%), we decided to re-examine the literature controversy surrounding this putatively oncogenic alteration.

ALK^{ATI} is not as mutually exclusive as other established therapeutic targets in melanoma

The Couts paper contained a sample size of 6 ALK^{ATI} patients, but it identified NRAS and BRAF mutations in patients that harbored the ALK^{ATI} alteration. Thus, the lack of mutual exclusivity of ALK^{ATI} with the transforming melanoma oncogenes BRAF and NRAS has been suggested, but not conclusively and quantitatively demonstrated with an appropriately powered analysis. Pairwise conditional selection clearly showed that ALK^{ATI} is not as mutually exclusive with BRAF or NRAS as they are with each other (p-value from Kolmogorov-Smirnov test: 9.33e-15, Fig 2A-C, see methods for description of data acquisition, sorting, and analysis). Because this analysis could be confounded by the inclusion of all mutations in the BRAF and NRAS genes, we reperformed our analysis using only

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the hot spot mutations of $BRAF_{V600E}$ and $NRAS_{Q61L}$. This hotspot analysis showed a similar result (p-value from KS test: 2.61e-13, Fig 2C, right panel). Thus, ALK^{ATI} is significantly less mutually exclusive with BRAF or NRAS than they are with each other.

We also considered the possibility that the lack of mutual exclusivity of ALK^{ATI} with BRAF and NRAS was due to the definition of the initial filter cutoffs in the original RNA-seq analysis (the exact filters are mentioned in the methods). By systematically changing the cutoffs for all of the filters combinatorically, we varied the sensitivity for ALK^{ATI} detection by orders of magnitude (Fig 3A, top)**.** More stringent filter sets resulted in fewer ALK^{ATI} calls, whereas less stringent filters resulted in more calls (Fig 3A, bottom). Performing pairwise comparisons of conditional selection on all combinations of these filter cutoffs never created mutual exclusivity. This filter analyses indicated that the lack of observed mutual exclusivity in AKL^{ATI} was not due to sub-optimal cutoffs in the heuristic RNA-seq filters that were used.

This analysis shows the utility of tests for pairwise conditional selection to demonstrate quantitative differences in the degree of mutual exclusivity between a positive control gene pair and a new and potentially oncogenic alteration. We also strongly demonstrate a lack of mutual exclusivity between ALK^{ATI} and established MAPK pathway driver mutations.

Kinase domain expression imbalance in ALK is nearly ubiquitous in melanoma and lung cancers

Given the conclusive lack of mutual exclusivity between ALK^{ATI} and transforming melanoma oncogenes, we decided to look deeper at the initial signal, i.e. the bias towards ALK expression in the kinase domain (Exons 20:29). We expected to see a distribution centered at equivalent expression between the kinase domain and the upstream coding region (the diagonal line in Fig 3B). We posited that expression levels in exons 1-19 and 20-29 should be distributed above and below the diagonal, with ALK^{ATI} patients being strong outliers from this expected relationship. However, we observed a significant bias towards overexpression of the kinase domain of ALK across all melanoma patients, p-value from KS-test: 2.2e-16. In fact, almost all the skin cutaneous melanoma (SKCM, n=340) and lung adenocarcinoma (LUAD, n=477) patients in the TCGA (Fig 3B, Supplementary Fig 2A) expressed higher levels of the ALK kinase domain than exons 1-19. The ubiquity of this deviation in all patients led us to suspect that a systematic error could be at play. While multiple interpretations of this signal exist, it is concerning that all patients have some propensity to overexpress the kinase domain. As a control, we also examined whether an imbalance of expression towards the kinase domain is a unique feature of ALK. We did not see this propensity for kinase domain expression in EGFR in SKCM or in EGFR in LUAD, p-value from KS-test:0.66 (Supplementary Fig 2B). Hence, the kinase domain expression imbalance is not a generalizable finding and raises concerns of potential systemic biases in exon specific expression levels in ALK in the TCGA.

ALK^{ATI} is not sufficient for growth/transformation *in vitro*

Our computational re-analysis of ALK expression data in melanoma suggested that ALK^{ATI} is significantly less mutually exclusive with NRAS and BRAF than they are with each other. In the original Wiesner *et al*. study (31), ALK^{ATI} was argued to be sufficient for transformation of Ba/F3s to growth factor independence. When we transduced Ba/F3s with ALK^{ATI} and EML4-ALK (9 independent replicates each), we found that ALK^{ATI} took significantly longer to grow out (two population doublings in 6.7±0.4 days for EML4-ALK and 10.0±1.0 days for ALK^{ATI}, Fig 4A). However, we also reasoned that longer outgrowth times were indicative of a weaker transforming potential. As such, we scaled up the number of transductions to 48 independent replicates by performing many parallel transductions of ALK^{ATI}, EML4-ALK, and vector. The results were striking. Growth factor independence was observed in 100% of EML4-ALK replicate transductions, but only 37.5% ALK^{ATI} replicate transductions, and in 16% of vector controls (Fig 4B). Viral titers were essentially indistinguishable across these three constructs (Supplementary Fig 3); puromycin resistance cassette transfer as indicated by time to outgrowth following puromycin selection was similar as well. Thus, ALK^{ATI} exhibits only a modest increase in transformation efficiency when large scale replication is performed.

Next, we reasoned that growth factor independence in ALK^{ATI} Ba/F3 cells was infrequent because it required a relatively rare second genetic event. To test this hypothesis, we extracted the genomic DNA of the 18 ALK^{ATI} Ba/F3 cell lines that achieved growth factor independence and sequenced their ALK kinase domain. Surprisingly, 3 of the 18 ALK^{ATI} cell lines had well known transforming mutations in the ALK kinase (F1174C, F1174V, and F1174I, Fig 4A, Supplementary Fig 4). These point mutations are the primary cancer causing mutations in ALK-mutated neuroblastoma and have been shown to constitutively activate the ALK kinase (35). These mutations are also sensitive

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to some ALK kinase inhibitors. This gives a concrete rationale for why ALK^{ATI} could transform Ba/F3 cells, and that those cells were sensitive ALK inhibition (31). While the mutations were not found in all cells, their existence in 3 independent selections strongly suggests that the transformation results and therapeutic treatments of ALK^{ATI} can be confounded by secondary genetic events. Alongside our re-analysis of the genomic data and the Couts *et al* data, our transformation data strongly suggested that ALK^{ATI} is not sufficient for growth/transformation *in vitro*, and that ALK inhibitors will not be a good single agent treatment strategy in ALK^{ATI} positive melanoma.

We also stress that plasmids and cell lines that contain these activating point mutations have never been used in our lab (or a neighboring lab) which rules out contamination by another source. While we could only identify secondary mutations in a minority of cell lines, we argue that the selection for 3 distinct activating mutations in the ALK^{ATI} background strongly suggests that ALK^{ATI} is not sufficient for transformation and requires a second genetic event. Moreover, though other transformation experiments were performed in (33), we argue that the large scale replication of the Ba/F3 results in our lab cast significant concerns on the other transformation studies in ALK^{ATI}.

ALK^{ATI} cannot rescue melanoma cell lines from a vemurafenib challenge

While our Ba/F3 analysis suggests that the ALK^{ATI} alteration is not sufficient to transform to growth factor independence, we decided to test the potential oncogenicity of ALK^{ATI} in a more realistic melanoma cell line model. We transduced ALK^{ATI} into two $BRAF_{V600E}$-harboring melanoma cell lines, and challenged them with a $BRAF_{V600E}$ inhibitor, vemurafenib. The experimental rationale was simple, $BRAF_{V600E}$ is necessary for melanoma cell survival, and sufficient for transformation (20,30). The drug vemurafenib inhibits only the mutant BRAF V600E protein as a monomer (16,30,36), while receptor tyrosine kinases signal through dimeric RAF family proteins (37). A candidate oncogenic RTK in melanoma with strong transforming potential should be able to rescue $BRAF_{V600E}$ melanoma from a $BRAF_{V600E}$ inhibitor.

We transduced EML4-ALK, ALK^{ATI}, and vector into two different skin cancer cell lines, SKMEL28 and G361, both of which have a transforming V600E point mutation and are sensitive to vemurafenib (34). When we did this, ALK^{ATI} and vector transduced melanoma cells had statistically indistinguishable vemurafenib dose responses (p-value: 0.49 for SKMEL28 and 0.97 for G361, Fig 4C). The inability of ALK^{ATI} to rescue melanoma cell lines is in line with the notion that ALK^{ATI} is not an oncogenic driver in melanoma.

We proceeded with our vemurafenib challenge in the absence of a positive control because multiple transduction attempts for EML4-ALK and MEK_{Q56P} failed in our melanoma cell lines. Both EML4-ALK and MEK_{Q56P} are established oncogenic drivers in NSCLC and melanoma (10,38). We confirmed our viral titer and infectivity by simultaneously transducing and transforming Ba/F3 to IL-3 independence, as well as infecting Hek293T and selecting for puromycin resistance. Interestingly, virus packaged with the EML4-ALK or MEK_{Q56P} oncogene readily transformed Ba/F3 cells, and could be easily infected into Hek293T cells and selected with puromycin, but we could not successfully select for EML4-ALK or MEK_{Q56P} containing SKMEL28 or G361 cells (Supplementary Table 1). While this is a negative result, we believe that it is a strong one because of our simultaneous infections. We were able to generate both vector control and ALK^{ATI} containing SKMEL28 and G361 cells through puromycin infection each time we attempted infection, and we were able to generate EML4-ALK and MEK_{Q56P} transformed Ba/F3 cells, and puromycin selected Hek293T cells in every transduction attempt. We were never able to generate stable SKMEL28 or G361 cell lines expressing EML4-ALK or MEK_{Q56P} on three separate transduction attempts on three separate days. Thus, we can rule out infection efficiency, and construct function as the source of this result. This is consistent with the idea that conditional selection via mutual exclusivity occurs because two growth pathway activating variants in the same cell can become toxic (27,28).

Because we could not generate a stable positive control, we resorted to transient transfection to show that we could generate resistance to vemurafenib with an upstream oncogenic event. We transfected SOS1Wt, $SOS1^{E846K}$, and MEK^{Q56P} into the $BRAF_{V600E}$ containing Skmel-28 melanoma cells. We used SOS because it has an intermediate level of ERK activation. E846K and Q56P represent activating mutations in SOS1 and MEK and have previously been described to activate MAPK signaling (29). Our rationale was that constructs with upregulated MAPK signaling (Supplementary Fig. 5a) should be able to rescue monomeric $BRAF_{V600E}$-containing SKMEL-28 melanomas from a vemurafenib challenge. Transient transfection of MEK_{Q56P} and $SOS1_{E846K}$ was capable of partially rescuing vemurafenib challenge in Skmel-28 cells compared to Skmel-28 cells transfected with $SOS1_{Wt}$ (Supplementary Fig 5b). Thus, it seems unlikely that ALK^{ATI} is sufficient to act as an oncogene in melanoma cells.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ALK expression imbalance does not predict sensitivity to ALK inhibitors.

While previous experiments suggested that ALK^{ATI} may not be sufficient for oncogenesis, we still wondered if ALK was necessary for *in vitro* survival in some melanoma cell lines. To test this we analyzed ALK expression data and data on dose response to an ALK inhibitor (TAE-684) for 54 melanoma cell lines from the CCLE (34). In this dataset, 2 of the 54 (~4%) cell lines had strong exon imbalance that was ALK^{ATI}-like (exceeded 2/3 filters).Neither of these cell lines were sensitive to TAE-684 (Fig 4D). Furthermore, the degree of exon imbalance in ALK expression (when treated as a continuous variable) did not predict TAE-684 responses in these 54 cell lines (linear regression p-value: 0.51). Although this test would have been helped by a larger number of cell lines or more dramatic ALK^{ATI} like phenotypes, we conclude that the degree of observed exon imbalance is not correlated with single agent sensitivity to ALK-inhibition across 54 melanoma cell lines.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

In summary, we develop a resampling method for the pairwise comparison of two gene pairs to quantitatively interpret a conditional selection signal. Doing this allows us to quantify a negative finding relative to how often a lack of mutual exclusivity would be seen in a positive control gene pair. Our simulations show that pairwise comparisons of conditional selection are a sensitive and specific method to quantitatively compare rare genomic events to positive controls.

We applied our pairwise comparisons of conditional selection to a transcript alteration in ALK^{ATI} because of the controversy in the literature (31,33). Our analysis clearly demonstrates a lack of conditional selection for ALK^{ATI}. ALK^{ATI} is significantly less mutually exclusive than BRAF and NRAS are with each other in melanoma. Moreover, our experiments suggest that ALK^{ATI} is not sufficient for cellular transformation, that kinase domain imbalance does not predict inhibitor response, and that single agent ALK inhibition is unlikely to be therapeutic in melanoma cells.

In their original paper, Wiesner et. al. performed a substantive and detailed description of the ALK^{ATI} event. They found enrichment of H3K4me3 and RNA Pol II near the ATI transcription initiation site of tumors expressing ALK^{ATI} (31). Wiesner et al also confirmed that ALK^{ATI} is expressed at both ALK alleles by comparing DNA, RNA, and H3k4me3 levels. They also performed gene expression profiling of RNA-Seq datasets showing that ALK^{ATI}-like expression is found in 2-11% of melanoma samples and sporadically in various other tumor types and not in normal tissue samples (31,32). The breadth and depth of the analysis leads us to believe that ALK^{ATI} is likely a true transcript variant. Moreover, while we believe our work (alongside the work of Couts et al) provides strong evidence against single agent therapy, it would be unfair to ignore the potential for therapeutic relevance of ALK^{ATI} in contexts other than ALK inhibitor monotherapy. Moreover, from a biological perspective, it would also be unfair to rule out some sort of unknown biological role for ALK^{ATI} that does not fit conventional definitions of driver oncogenes. However, combining Couts' PDX data with the questionable transformation potential of ALK^{ATI}, the lack of objective responses in ALK^{ATI} expressing melanomas to ALK inhibitors, and the clear the lack of mutual exclusivity of ALK^{ATI} in melanoma casts significant doubt upon the single agent therapeutic rationale for ALK^{ATI}.

In the melanoma landscape, dramatic responses to approved and investigational immunotherapy agents are yielding important steps forward for patient care. We have systematically shown that the original ALK^{ATI} data should be re-evaluated in light of our data reproducing the original finding, and in light of the compelling recent reports in PDX models (33,39). Combined with the fact that the patient data in the original manuscript did not achieve the typical clinical criteria for an objective partial response, we strongly recommend that single agent ALK inhibitors receive no further testing in ALK^{ATI} patients when other investigational or off label options exist. Given the weight of evidence, it seems unlikely that refractory patients will benefit from this treatment. We also suggest that pairwise tests for conditional selection will be a useful tool to triage any rare genomic finding in the mountains of cancer sequencing data generated in late stage patients.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

1.1 Pairwise comparisons for conditional selection workflow

Pairwise comparisons for conditional selection is a method that controls for the differences in sample size in a pairwise comparison of the two mutually exclusive or co-occurring gene pairs. Given a highly abundant mutually exclusive positive control gene pair (PC) and a gene pair with the gene of interest (GOI) with variable abundance, our method corrects the frequency and the abundance for the highly abundant gene pair. This is done by bootstrapping the GOI population and resampling the positive control to the same number of hits (abundance) and the same proportion in the population (frequency) as the GOI.

The overall workflow of our pairwise comparisons for conditional selection is split into various steps. Data containing abundances of two genes in a cohort, can be summarized using a contingency table. A Fisher's exact test applied on the summarized counts from the contingency table gives p-values for the mutual exclusivity of a gene pairs. Next, a non-parametric test can compare the p-value distributions of two gene pairs to analyze how different their distributions are. Two gene pairs can be downsampled to correct for bias by controlling for the abundance and frequency of the positive control in the dataset. This requires calculating the subsample size of the bootstrap, and the number of subsampling simulations. A step-by-step description of the workflow process is highlighted in our Github repository.

1.2 Calculating conditional selection in a gene pair

Cohort data containing counts of two genes was summarized in a 2x2 contingency table. The p-value for mutual exclusivity between the two genes was obtained using the Fisher's exact test. P-values were calculated for both the positive control gene pair and the gene of interest gene pair (Fig 1A). Subsequently, the Kolmogorov-Smirnov test (*KS* test), was used to calculate the pairwise difference in mutual exclusivity between two gene pairs by comparing their p-value distributions. KS test is a two sample, non-parametric test of the equality of continuous, one dimensional probability distributions. These pairwise analyses were also repeated using the Wilcoxon signed rank test, another two sample non-parametric test, to yield mostly similar results.

1.3 Abundance and frequency resampling in gene pairs

Pairwise comparisons for conditional selection corrects for the abundance <u>and</u> the frequency of the abundant gene by downsampling it to the abundance and frequency of the rare gene. The abundance of a gene is how many times it appears in a cohort. The frequency of a gene of interest (GOI in Fig 1A), is the proportion of its hits of the gene of interest to the size of the entire population (1).

$$GOI\ frequency = \frac{abundance_{GOI}}{cohort\ size} = \frac{a+b}{a+b+c+d} \tag{1.1}$$

In (eq. 1), it can be seen that the abundance is directly proportional to the frequency if the genes are pulled from the same cohort. Correcting for abundance corrects for frequency by default. Therefore, the positive control genes were downsampled to the abundance of the GOI*.

*It is sometimes impossible to simply downsample the positive control gene to the abundance of the GOI. If the positive control gene is very prevalent in the population (a_1 and c_1 in Fig 1A), there may not be enough events that do not have the positive control (c_1 and d_1 in Fig 1A) to achieve the low frequency of the GOI. In that case, we chose the maximum possible subsample size of the positive control given that its frequency could not exceed the frequency of the GOI:

$$Subsamplesize_{max} = \min\{abundance_{GOI}, \frac{PC1^{Compliment}}{1-frequency_{GOI}} - PC1^{Compliment}, \frac{PC2^{Compliment}}{1-frequency_{GOI}} - PC2^{Compliment}\} \tag{1.2}$$

Once a subsample size was chosen, downsampling was repeated to yield multiple bootstrapped simulated datasets. The number of downsampling simulations did not exceed the maximum possible combinations of sampling with replacement of the abundance of the GOI:

$$Simulations_{max} = \binom{n+k-1}{k} = \binom{2*abundance_{GOI}-1}{abundance_{GOI}} \tag{1.3}$$

No more than 1,000 simulations were allowed for computational efficiency.

2. Generating simulated cohorts:

We simulated cohorts of gene pairs to test our pairwise comparisons method and to characterize its sensitivity. Each simulated gene pair was varied by cohort size, odds ratio, and abundance of the GOI (eq. 2.1-2.3, illustrated in

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Fig. 1A). Given these inputs, the counts of individual elements $\{a_1\}$, $\{b_1\}$, $\{c_1\}$, $\{d_1\}$, $\{a_2\}$, $\{b_2\}$, $\{c_2\}$, $\{d_2\}$, were calculated (elements illustrated in Fig. 1A). The individual elements enabled the calculation of the unknown abundances of the positive control genes (eq. 2.6-2.7). We assumed that both the gene pairs were pulled from cohort of the same size (eq. 6), and that $\{c\}=\{d\}$ for both gene pairs (eq.2.5). We also assume that the positive control takes the same abundance for both gene pairs (eq. 2.6). All count data were rounded to the nearest integer.

$$Cohort\ size = \{a, b, c, d\} = \tag{2.1}$$

$$Odds\ ratio = \frac{a*d}{b*c} \tag{2.2}$$

$$Abundance_{GOI} = \{a_1, b_1\} \tag{2.3}$$

$$Cohort\ size_{GenePair_1} = Cohort\ size_{GenePair_2} \tag{2.4}$$

$$\{c\} = \{d\} \tag{2.5}$$

$$Abundance_{PC_1} = \{a_1, c_1\} = \{a_2, c_2\} \tag{2.6}$$

$$Abundance_{PC_2} = \{a_2, b_2\} \tag{2.7}$$

The following steps outline how the individual elements for gene pair 1 and 2 were calculated. For both the gene pair 1, given the assumption that $c_1=d_1$ (eq. 2.5), eq. 5 can be substituted into (eq. 2.1) to give:

$$c_1 = d_1 = \frac{cohort\ size - abundance_{GOI}}{2} \tag{2.8}$$

Next, substituting (2.2) into (2.3) helps solve for a_1:

$$a_1 = incidence_{GOI} * \frac{OR_{genepair1}}{1+OR_{genepair1}} \tag{2.9}$$

b_1 can be calculated by using (2.3) as $b_1 = abundance_{GOI} - a_1$

Finally, abundance of the positive control 1 (eq. 2.6) can be calculated by adding a_1 and b_1.

The abundance for gene pair 2 can be found in a similar manner. Given the assumption in (eq. 2.5), eq. 2.6 can be substituted into eq. 2.1 to give:

$$c_2 = d_2 = \frac{Cohort\ size - abundance_{PC_1}}{2} \tag{2.10}$$

By substituting (2.2) into (2.6), the odds ratio for the gene pair 2 helps solve for a_2

$$a_2 = incidence_{PC1} * \frac{OR_{Genepair2}}{1+OR_{Genepair2}} \tag{2.11}$$

b_2 can be calculated by using (2.6) as $b_2 = abundance_{PC1} - a_2$

Finally, abundance of the positive control 2 (eq. 2.7) can be calculated by adding a_1 and b_1. These parameters were summarized in two contingency tables (one for each gene pair), and used for downstream analyses.

3. Analysis of public data sets:

We downloaded our level 3 TCGA data from the Broad Institute TCGA GDAC Firehose (http://gdac.broadinstitute.org/) 2016_01_28 run. In our original ALK expression filters (Fig 2), ALK^{ATI} candidates were identified as samples with an ALK expression level of RSEM ≥ 100, ≥ 500 total reads across all ALK exons, and ≥ 10-fold greater average expression in exons 20–29 compared to exons 1–19. When varying filters, ALK^{ATI} patients were identified as samples with an ALK expression level of 100≤RSEM≤500, number of ALK reads≤1000, and 10≤Exon20-29/Exon1-19 ratio≤100 (Fig 3).

4. *In vitro* transformation and drug treatment assays:

We used a standard Ba/F3 transformation protocol for lentiviral transduction. Lentiviral particles were made by transfecting HEK293T cells with the plasmid of interest and with packaging plasmids (3rd generation Lentiviral system from Addgene). Replication incompetent virus was collected using a BL2+ safety protocol. Upon virus collection, Ba/F3 cells at 500k/mL in 4mLs were infected with an equal volume of virus. A minimum of three replicates were used for each infection condition. Three days after infection, the Ba/F3s cells were spun out of virus and selected for IL3 independence and/or puromycin resistance (0.5ug/mL Puromycin was used). An infection was determined to be successful Ba/F3 cells for a given construct grew out for both IL3 independence and puromycin resistance. During selection, cell growth was quantified by doing daily counts of live/dead cells on a daily basis. These live/dead analyses were performed by supplementing 20uLs of cell culture with 0.4% trypan blue and subsequently counting live cells using a hemocytometer. Alternatively, live dead cells were counted using flow cytometry analysis (BD Accuri C6 Sampler Plus).

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

5. Lentiviral Particle Quantification:
For each infection, before transducing Ba/F3s with virus, 500uL of virus was set aside and frozen at -20C. The number of lentiviral particles were quantified using the QuikTiter Lentivirus Quantitation Kit from Cell Biolabs (HIV P24 ELISA). All reported viral titers were within the linear range of the standard curve made using a positive control. Transduction efficiency was verified by counting outgrowth rates using puromycin selections across multiple infections.

6. Cell Lines:
Cell lines used were: SK-MEL-28 (ATCC HTB-72), G361 (ATCC CRL-1424), Ba/F3 (DSMZ ACC-300), HEK-293T (ATCC CRL-1573). Prior to use, all cell lines were tested to be free of mycoplasma using a biochemical-based test (Mycoalert plus, Lonza). Ba/F3 cell lines were cultures in RPMI while HEK-293T, Skmel-28 and G361 cell lines were cultured in DMEM. Media was supplemented with a final concentration of 10% FBS, and 1% Penicillin/Streptomycin/L-Glutamine. Wt Ba/F3 cell lines were grown in media supplemented with 10ng/mL IL3 (R&D systems). Stable transductions were verified for puromycin resistance by using kill curve concentrations ranging from 0.25ug/mL to 2ug/mL.
Cells were split when they were 70-80% confluent. The subculture ratios used for all of our cell lines included splitting ratios of 1:5 every 2-4 days when cells were 70-80% confluent.
$Ba/F3_{Wt}$ cell lines were grown in 10ng/mL IL3. After transduction with an oncogene, cell lines were selected with Puromycin to test transduction efficiency and subsequently with -IL3 to assess growth factor independence.

7. Plasmids:
The sequence of ALK^{ATI} was obtained from the European Nucleotide Archive under the expression number LN964494. The coding sequences, cloned into a pLVX-IRES-Puro backbone, was prepared by Genscript Gene Services.

Note on reproducibility: All of the analyses in this paper were performed in R v3.5.2. GraphPad prism v8 was used for dose response curve analysis. Amazon Web Services were used for the simulations in Fig. 1 and Supplementary Fig. 1. Version controlled html Rmarkdown files that include descriptions of the methods for all figures may be accessed via Github at here. This includes the code for the analysis of mutual exclusivity of gene pairs using pairwise comparisons of conditional selection. The results of all statistical analyses mentioned in this figure are included in the code. The workflow, outlined in our Github repository, requires an input dataset in the form of a table with columns for data on a gene of interest, and two positive control genes. We have also made available a function, named tcga_skcm_data_parser.Rmd, that can generate estimated counts for a gene of interest based on filters applied to RSEM, RPKM, and count data. The entire workflow was tested for reproducibility on simulated datasets and on expression data in LUAD for ALK and EGFR from the TCGA (40).

8. Acknowledgements:
We would like to thank Scott Leighow for his contribution with idea generation, critical analysis, and proofreading of this manuscript. We would also like to thank Lauren Randolph, Kyle McIlroy, Joshua Reynolds, and Yiyun Rao for their help revising previous versions of this manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Bibliography

1. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013 Sep;501(7467):338–45.

2. Mina M, Raynaud F, Tavernari D, Battistello E, Sungalee S, Saghafinia S, et al. Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. Cancer Cell. 2017 Aug 14;32(2):155-168.e6.

3. Whittaker SR, Theurillat J-P, Allen EV, Wagle N, Hsiao J, Cowley GS, et al. A Genome-Scale RNA Interference Screen Implicates NF1 Loss in Resistance to RAF Inhibition. Cancer Discov. 2013 Mar 1;3(3):350–62.

4. Vandin F, Upfal E, Raphael BJ. De novo discovery of mutated driver pathways in cancer. Genome Res. 2012 Feb 1;22(2):375–85.

5. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res. 2012 Feb 1;22(2):398–406.

6. Zhao J, Zhang S, Wu L-Y, Zhang X-S. Efficient methods for identifying mutated driver pathways in cancer. Bioinformatics. 2012 Nov 15;28(22):2940–7.

7. Widmann C, Gibson S, Jarpe MB, Johnson GL. Mitogen-Activated Protein Kinase: Conservation of a Three-Kinase Module From Yeast to Human. Physiological Reviews. 1999 Jan 1;79(1):143–80.

8. Zhang W, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Research. 2002 Mar;12(1):9.

9. Lemmon MA, Schlessinger J. Cell Signaling by Receptor Tyrosine Kinases. Cell. 2010 Jun 25;141(7):1117–34.

10. Takahashi T, Sonobe M, Kobayashi M, Yoshizawa A, Menju T, Nakayama E, et al. Clinicopathologic Features of Non-Small-Cell Lung Cancer with EML4–ALK Fusion Gene. Ann Surg Oncol. 2010 Mar 1;17(3):889–97.

11. Varmus H, Unni AM, Lockwood WW. How Cancer Genomics Drives Cancer Biology: Does Synthetic Lethality Explain Mutually Exclusive Oncogenic Mutations? Cold Spring Harb Symp Quant Biol. 2016 Jan 1;81:247–55.

12. Unni AM, Lockwood WW, Zejnullahu K, Lee-Lin S-Q, Varmus H. Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. Elife. 2015 Jun 5;4:e06907.

13. Bhullar KS, Lagarón NO, McGowan EM, Parmar I, Jha A, Hubbard BP, et al. Kinase-targeted cancer therapies: progress, challenges and future directions. Molecular Cancer. 2018 Feb 19;17(1):48.

14. Muhsin M, Graham J, Kirkpatrick P. Gefitinib [Internet]. Nature Reviews Drug Discovery. [cited 2019 May 26]. Available from: https://www.nature.com/articles/nrd1136

15. Shaw AT, Camidge DR, Engelman JA, Solomon BJ, Kwak EL, Clark JW, et al. Clinical activity of crizotinib in advanced non-small cell lung cancer (NSCLC) harboring ROS1 gene rearrangement. JCO. 2012 May 20;30(15_suppl):7508–7508.

16. Swaika A, Crozier JA, Joseph RW. Vemurafenib: an evidence-based review of its clinical utility in the treatment of metastatic melanoma. Drug Des Devel Ther. 2014 Jun 16;8:775–87.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

17. Fakih M, O'Neil B, Price TJ, Falchook GS, Desai J, Kuo J, et al. Phase 1 study evaluating the safety, tolerability, pharmacokinetics (PK), and efficacy of AMG 510, a novel small molecule KRASG12C inhibitor, in advanced solid tumors. JCO. 2019 May 20;37(15_suppl):3003–3003.

18. Schlumberger M, Tahara M, Wirth LJ, Robinson B, Brose MS, Elisei R, et al. A phase 3, multicenter, double-blind, placebo-controlled trial of lenvatinib (E7080) in patients with 131I-refractory differentiated thyroid cancer (SELECT). JCO. 2014 Jun 20;32(18_suppl):LBA6008–LBA6008.

19. Guan J, Gupta R, Filipp FV. Cancer systems biology of TCGA SKCM: Efficient detection of genomic drivers in melanoma. Scientific Reports. 2015 Jan 20;5:7857.

20. Sensi M, Nicolini G, Petti C, Bersani I, Lozupone F, Molla A, et al. Mutually exclusive NRASQ61R and BRAFV600E mutations at the single-cell level in the same human melanoma. Oncogene. 2006 Jun 8;25(24):3357–64.

21. Poynter JN, Elder JT, Fullen DR, Nair RP, Soengas MS, Johnson TM, et al. BRAF and NRAS mutations in melanoma and melanocytic nevi: Melanoma Research. 2006 Aug;16(4):267–73.

22. Edlundh-Rose E, Egyha??zi S, Omholt K, Mansson-Brahme E, Platz A, Hansson J, et al. NRAS and BRAF mutations in melanoma tumours in relation to clinical characteristics: a study based on mutation screening by pyrosequencing: Melanoma Research. 2006 Dec;16(6):471–8.

23. Marzese DM, Scolyer RA, Roqué M, Vargas-Roig LM, Huynh JL, Wilmott JS, et al. DNA methylation and gene deletion analysis of brain metastases in melanoma patients identifies mutually exclusive molecular alterations. Neuro Oncol. 2014 Nov;16(11):1499–509.

24. Wells SA, Robinson BG, Gagel RF, Dralle H, Fagin JA, Santoro M, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. J Clin Oncol. 2012 Jan 10;30(2):134–41.

25. Roskoski R, Sadeghi-Nejad A. Role of RET protein-tyrosine kinase inhibitors in the treatment RET-driven thyroid and lung cancers. Pharmacological Research. 2018 Feb 1;128:1–17.

26. Gainor JF, Varghese AM, Ou S-HI, Kabraji S, Awad MM, Katayama R, et al. ALK Rearrangements Are Mutually Exclusive with Mutations in EGFR or KRAS: An Analysis of 1,683 Patients with Non–Small Cell Lung Cancer. Clin Cancer Res [Internet]. 2013 May 31 [cited 2018 Dec 10]; Available from: http://clincancerres.aacrjournals.org/content/early/2013/06/20/1078-0432.CCR-13-0318

27. Cisowski J, Sayin VI, Liu M, Karlsson C, Bergo MO. Oncogene-induced senescence underlies the mutual exclusive nature of oncogenic KRAS and BRAF. Oncogene. 2016 Mar;35(10):1328–33.

28. Petti C, Molla A, Vegetti C, Ferrone S, Anichini A, Sensi M. Coexpression of NRASQ61R and BRAFV600E in Human Melanoma Cells Activates Senescence and Increases Susceptibility to Cell-Mediated Cytotoxicity. Cancer Res. 2006 Jul 1;66(13):6503–11.

29. Zhao B, Pritchard JR. Inherited Disease Genetics Improves the Identification of Cancer-Associated Genes. PLOS Genetics. 2016 Jun 15;12(6):e1006081.

30. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. New England Journal of Medicine. 2011 Jun 30;364(26):2507–16.

31. Wiesner T, Lee W, Obenauf AC, Ran L, Murali R, Zhang QF, et al. Alternative transcription initiation leads to expression of a novel *ALK* isoform in cancer. Nature. 2015 Oct;526(7573):453–7.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

32. Busam KJ, Villain RE, Lum T, Busam JA, Hollmann TJ, P.M.Saw R, et al. Primary and metastatic cutaneous melanomas express ALK through alternative transcriptional initiation. Am J Surg Pathol. 2016 Jun;40(6):786–95.

33. Couts KL, Bemis J, Turner JA, Bagby SM, Murphy D, Christiansen J, et al. ALK inhibitor response in melanomas expressing EML4-ALK fusions and alternate ALK isoforms. Mol Cancer Ther. 2017 Jan 1;molcanther.0472.2017.

34. Broad Institute Cancer Cell Line Encyclopedia (CCLE) [Internet]. [cited 2019 Feb 28]. Available from: https://portals.broadinstitute.org/ccle

35. Cosmic. COSMIC - Catalogue of Somatic Mutations in Cancer [Internet]. [cited 2019 Feb 28]. Available from: https://cancer.sanger.ac.uk/cosmic

36. Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H, et al. Clinical efficacy of a RAF inhibitor needs broad target blockade in *BRAF*-mutant melanoma. Nature. 2010 Sep;467(7315):596–9.

37. Freeman AK, Ritt DA, Morrison DK. The importance of Raf dimerization in cell signaling. Small GTPases. 2013 Jul 1;4(3):180–5.

38. Lin AW, Barradas M, Stone JC, Aelst L van, Serrano M, Lowe SW. Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling. Genes Dev. 1998 Oct 1;12(19):3008–19.

39. Uguen A, Uguen M, Guibourg B. ALK Expression in Melanomas: Looking for a Needle in a Haystack. The American Journal of Surgical Pathology. 2016 Oct;40(10):1437.

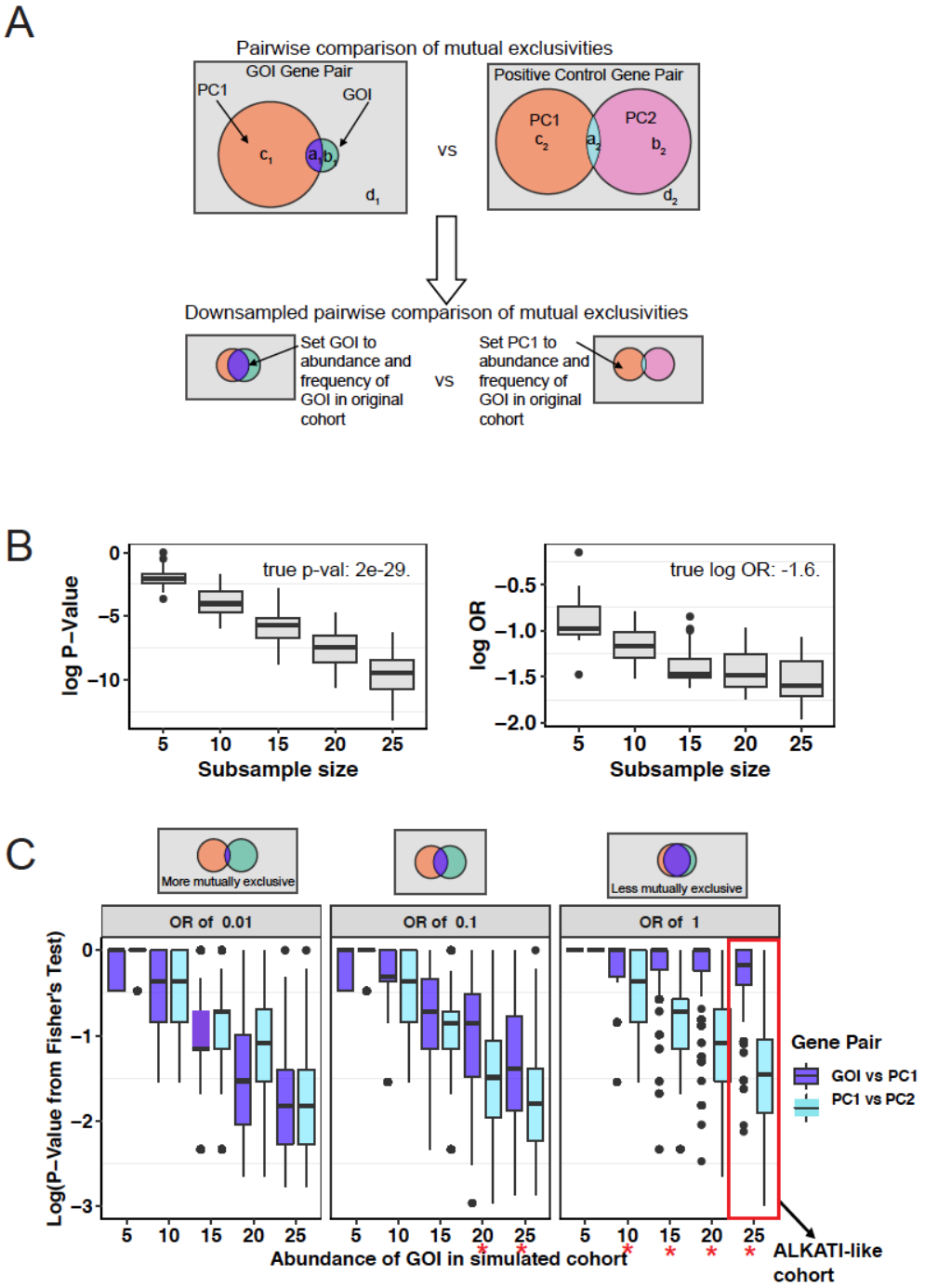
40. The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics. 2013 Sep 26;45:1113–20.

41. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011 Aug 4;12(1):323.

42. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods. 2008 Jul;5(7):621–8.

bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

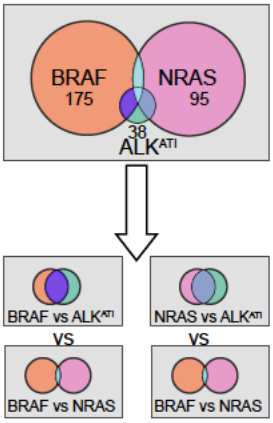
Figure 1: Pairwise comparisons of conditional selection between a GOI gene pair and a positive control gene pair in simulated cohorts. (A) Top: Pairwise comparisons of two gene pairs in which the GOI is rare relative to the positive control genes. Bottom: downsampled pairwise comparisons two gene pairs matches the abundance and frequency of the positive control genes to those of the rare GOI. **(B)** P-value distributions of a gene pair across different subsample sizes of the gene of interest. **(C)** Pairwise comparisons of conditional selection on simulated genes. The positive control gene pair is simulated at a constant, highly mutually exclusive OR of 0.01. Red asterisks mark significantly lower p-value distributions for the positive control gene pair compared to the GOI gene pair (Kolmogorov-Smirnov test).



bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 2: Performing pairwise comparisons of conditional selection on ALK^{ATI} reveals a lack of mutual exclusivity with transforming melanoma oncogenes BRAF and NRAS. (A) An illustration of the parameter space of ALK^{ATI} and the melanoma oncogenes BRAF and NRAS. The overall counts of the original data are: cohort size of 340 patients; abundances of ALK^{ATI}, BRAF, NRAS: 38, 175, and 95. BRAF∩ALK^{ATI}: 12, BRAF∩NRAS: 25, NRAS∩ALK^{ATI}: 8. **(B)** ALK^{ATI} is not mutually exclusive with BRAF or NRAS. BRAF and NRAS are mutually exclusive with each other (Fisher's exact test). **(C)** Pairwise comparisons of the FET p-value distributions of downsampled ALK^{ATI} vs BRAF and NRAS vs BRAF reveals a significant difference in their mutual exclusivities (red asterisks represent a KS test p-value <.05).

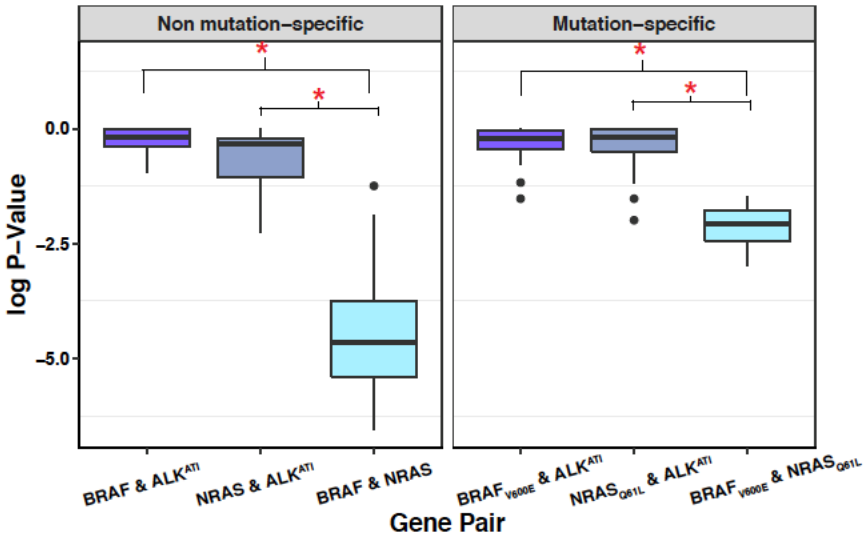
A



B

Comparison	Odds Ratio	P value
ALK^{ATI} vs anyBRAF	0.75	0.41
ALK^{ATI} vs BRAF$_{V600E}$	0.79	0.60
ALK^{ATI} vs any NRAS	1.70	0.13
ALK^{ATI} vs NRAS$_{Q61L}$	1.76	0.36
anyNRAS vs anyBRAF	.025	2.0e-29***
NRAS$_{Q61L}$ vs BRAF$_{V600E}$	0.00	8.3e-3***
Random Array vs BRAF$_{V600E}$	.68	.46

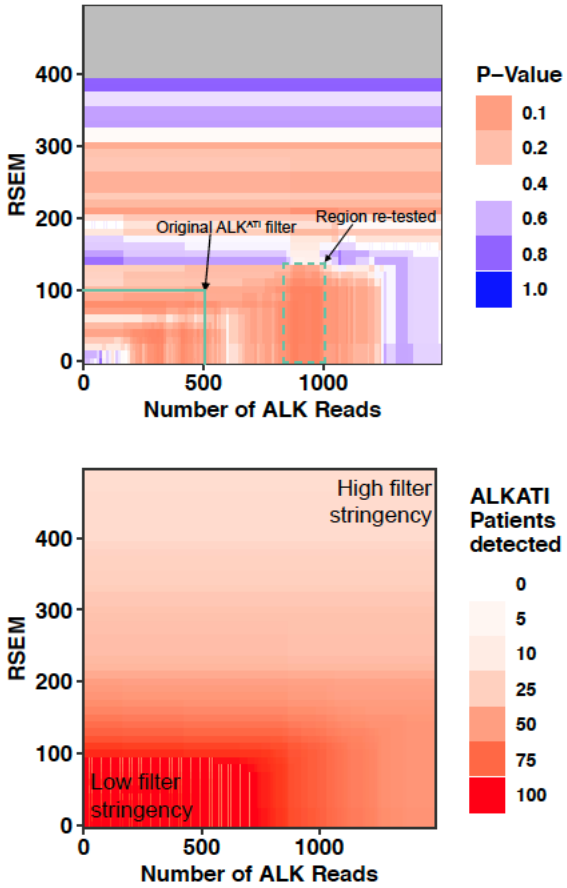
C



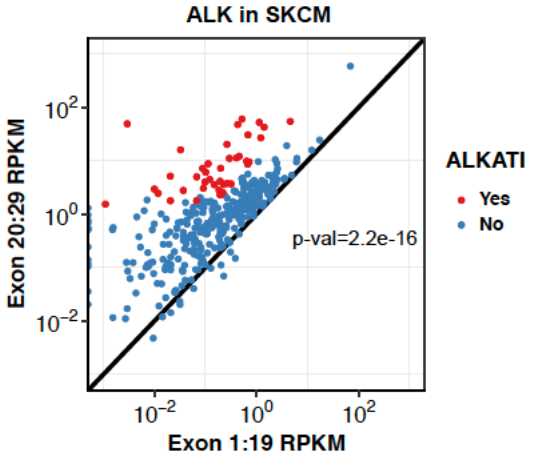
bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3: Changing the RNA-Seq filter cutoffs does not identify mutual exclusivity in ALK^{ATI}. (A) Top: Changing the filter cutoffs for RSEM (41), RPKM (42), and count data did not result in mutually exclusive regions. Filters used were RSEM 10-1000, counts 100-1000, ex20-29/ex1-19 ratio 10-100. Only the minimum p-value observed amongst the exon ratios is displayed. The midpoint of the p-value color gradient is 0.3. Regions that filtered for <10 ALK^{ATI} patients were not included in the pairwise comparison analysis and are colored gray. Bottom: The number of patients that are categorized as ALK^{ATI} positive decreases as the filter stringency increases. **(B)** The kinase domain of ALK (ex. 20-29) is significantly overexpressed in the majority of SKCM patients (χ^2 test).

A

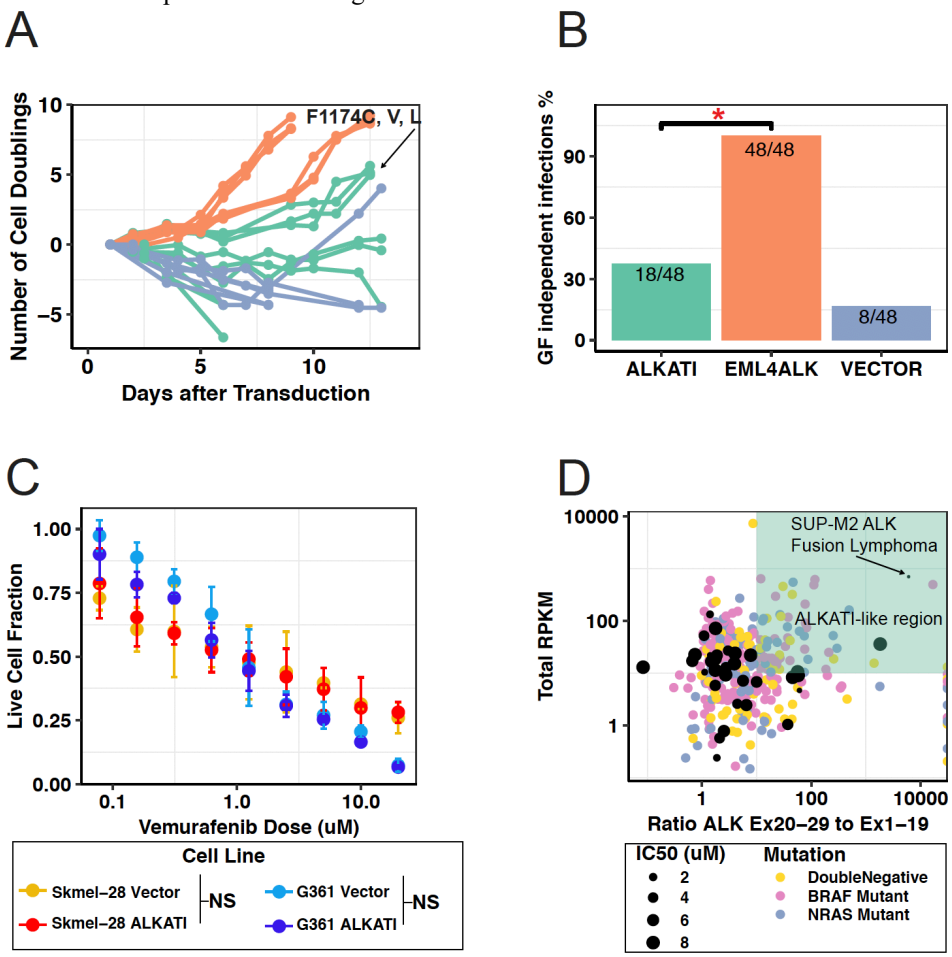


B



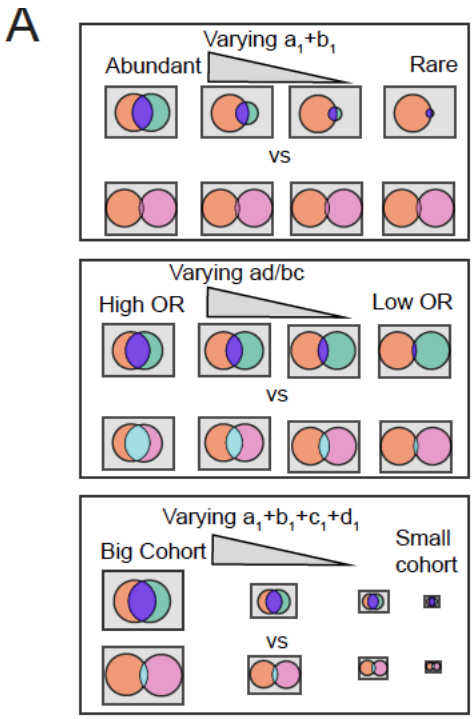
bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

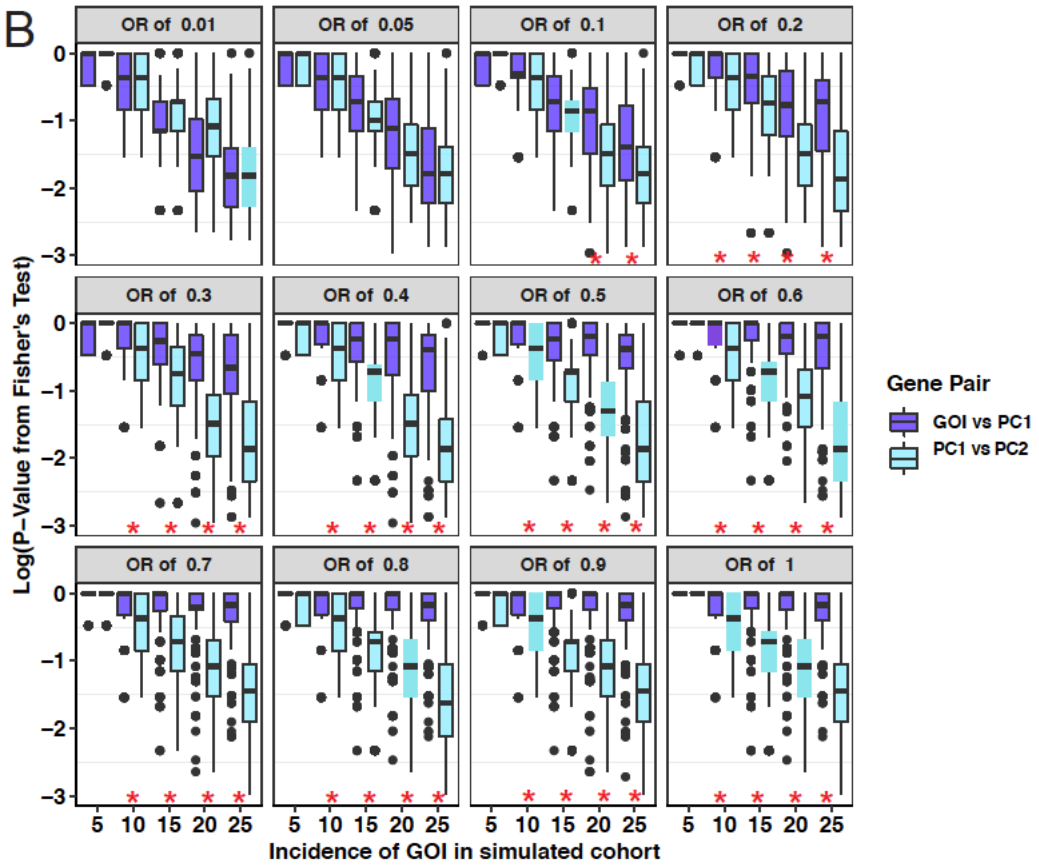
Figure 4: ALKATI is neither necessary nor sufficient in melanoma. (A) Time to growth factor independence, as measured by number of cell doublings, between ALK^{ATI}-transduced Ba/F3 cells (green), EML4ALK-transduced cells (orange), and vector-transduced cells (purple). Activating mutations in the kinase domain of ALK were detected after sequencing the genomic DNA of all growth factor independent outgrowths (see Supp. Fig. 4). **(B)** Proportion of infections that transformed for growth factor independence (χ^2 test p-value for ALK^{ATI} vs EML4ALK: 2.2e-10, ALK^{ATI} vs Vector: .038). **(C)** Dose response of ALK^{ATI} and vector-transduced SKMEL-28 and G-361 cell lines. ALK^{ATI} does not improve the dose response of SKMEL-28 (p-val: 0.49) and G361 (p-val: 0.99) to vemurafenib. P-value calculated using a one-sided paired t-test test between ALK^{ATI} and vector. Error bars represent standard deviation on 3 replicates of 3 different cell lines (9 total replicates). **(D)** No SKCM cell lines that have ALK^{ATI}-like expression levels confer sensitivity to an ALK-inhibitor, TAE684 (n=53). ALKATI-like expression does not predict TAE-684 sensitivity (linear regression p-value: 0.51). An ALK-fusion from a SUP-M2 lymphoma cell line is shown as an example of a sensitizing ALK alteration.



bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

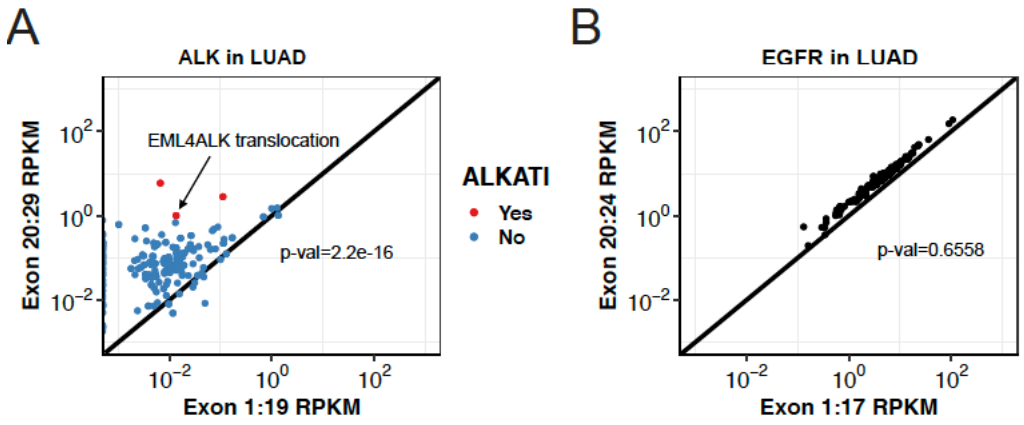
Supplementary 1: Pairwise comparisons of conditional selection in simulated cohorts. **(A)** Schematic of the different parameters that were varied when generating simulated cohorts. **(B).** The minimum required abundance of the GOI decreases as the simulated odds ratios between the two gene pairs increases. The positive control gene pair (PC1 vs PC2) is simulated to be mutually exclusive (OR of 0.01) while the GOI gene pair is simulated at a range of mutual exclusivities (ORs between 0.01 and 1). Cohort is simulated at a size of the cohort 500 patients. Red asterisks mark significantly lower p-value distributions for the PC gene pair compared to the GOI gene pair (Kolmogorov-Smirnov test).





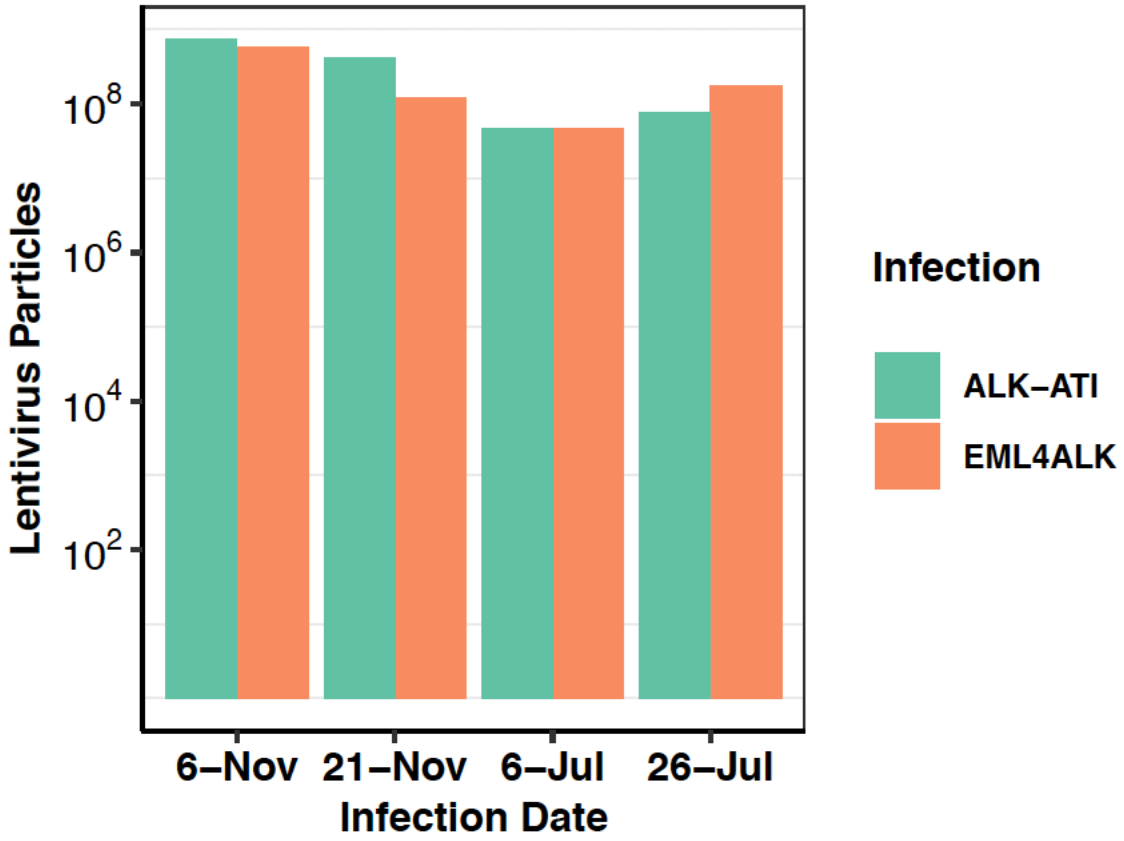
bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary figure 2: Propensity in expression towards kinase domain is ubiquitous in ALK. (A). Expression profile of ALK^{ATI} patients in LUAD showing an exon imbalance towards the kinase domain (ex20-29). One of the patients matching ALKATI expression filters had an EML4ALK translocation. **(B)** Expression profile of EGFR in LUAD showing a lack of exon imbalance towards EGFR's kinase domain (ex20-24).



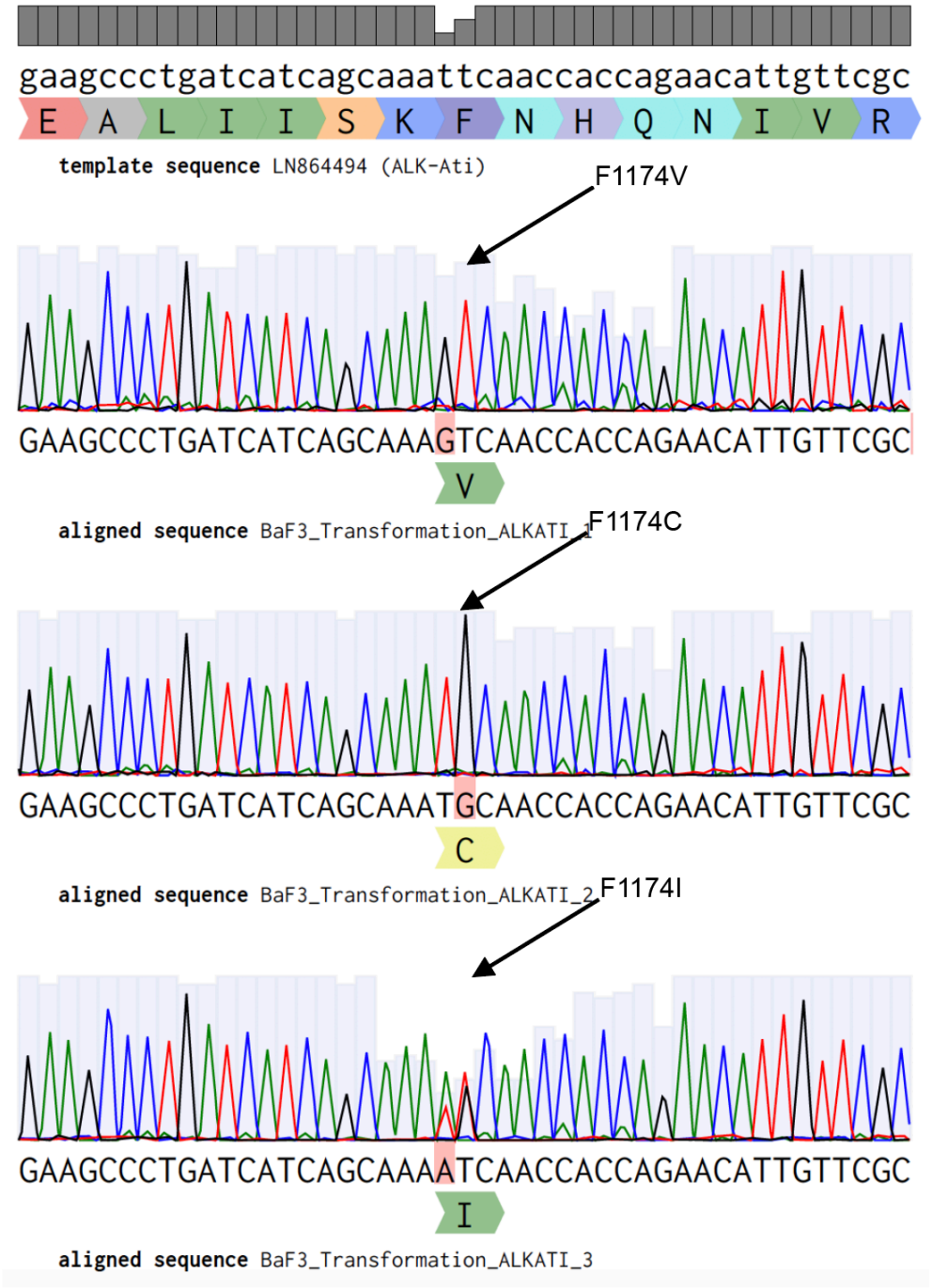
bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary figure 3: Counts of lentiviral particles across multiple infection attempts. Virus samples were collected prior to transducing Ba/F3s and SKML cell lines.



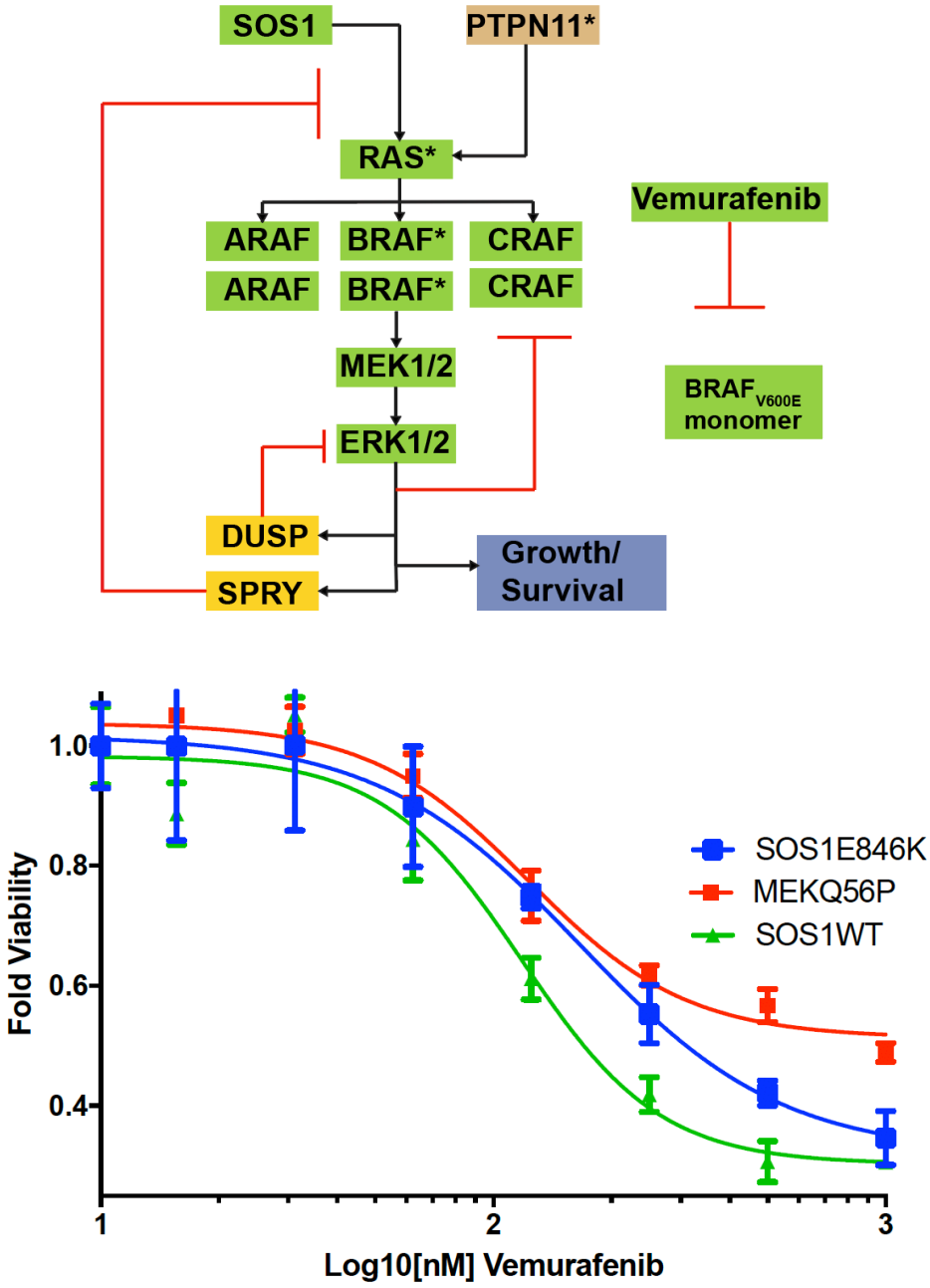
bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary 4: Nucleotide sanger sequencing of a part of the kinase domain of ALK showing the F1174C, V, and I transforming mutations. The three missense mutations represent the only mismatch in a conserved kinase domain.



bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Figure 5: Vemurafenib dose response of MEK$_{Q56P}$-transfected Skmel-28 melanoma cells. (A) Illustration of some of the genes involved in the MAPK signaling pathway. SOS1 is upstream and MEK is downstream of dimeric BRAF. Vemurafenib inhibits monomeric BRAF$_{V600E}$. **(B)** Vemurafenib dose response of Skmel-28 transiently transfected with SOS1Wt, SOS1E486K, and MEKQ56P. Transient transfection of MEK$_{Q56P}$ and SOS1$_{E486K}$ into BRAF$_{V600E}$-contianing Skmel-28 melanoma improves their dose response to vemurafenib (n=3 replicates, error bars are SD, p-value: 0.031 for SOS1Wt and SOS1E846K and 0.004 for SOS1Wt and MEKQ56P, one-sided paired t-test).



bioRxiv preprint doi: https://doi.org/10.1101/696294; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Table 1. Multiple attempts to transduce oncogenes into $BRAF_{V600E}$ melanoma lines fail.

*Note: Replicates represent different infection attempts (biological replicates). All technical replicates for a given infection behaved similarly.
**Replicate 3 not attempted for MEK_{Q56P}

Cell Line	SKMEL28	G361	SKMEL28	G361	SKMEL28	G361	SKMEL28	G361	BAF3	HEK293
Plasmid	Vector	Vector	ALK^{ATI}	ALK^{ATI}	EML4ALK	EML4ALK	MEK_{Q56P}	MEK_{Q56P}	EML4 ALK	EML4AL K
Selection	Puro	Puro	Puro	Puro	Puro	Puro	Puro	Puro	iL3	Puro
Replicate 1*	+	+	+	+	-	-	-	-	+	+
Replicate 2*	+	+	+	+	-	-	-	-	+	+
Replicate 3*	+	+	+	+	-	-	N/A**	N/A**	+	+