Characterization and clustering of kinase isoform expression in metastatic melanoma

David O. Holland[1], Valer Gotea[1], Kevin Fedkenheuer[1], Sushil K. Jaiswal[1], Laura Elnitski[1#]

[1]Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 20892

corresponding author
49 Convent Dr.
Bethesda, MD 20892
elnitski@nih.gov
301-451-0265

Abbreviations:
DE- differential expression
DIR- differential isoform ratios
KI- kinase inhibitor
TCGA- The Cancer Genome Atlas
SKCM- skin cutaneous melanoma project
TK- tyrosine kinase
RTK- receptor tyrosine kinase
nRTK- nonreceptor tyrosine kinase
CAMKs- Ca^{2+}/calmodulin-dependent kinases
CMGCs - Cdk, MAPK, GSK, Cdk-like related kinases
AGCs- protein kinase A, protein kinase G, and protein kinase C related kinases
TKL- tyrosine kinase-like
CK1- casein kinase 1
RGCs- receptor guanylate cyclases
STE- "Sterile" serine/threonine kinases

Abstract

Mutations to the human kinome are known to play causal roles in cancer. The kinome regulates numerous cell processes including growth, proliferation, differentiation, and apoptosis. In addition to aberrant expression, aberrant alternative splicing of cancer-driver genes is receiving increased attention as it could create loss or gain of functional domains, altering a kinase's downstream effects.

The present study quantifies changes in gene expression and isoform ratios in the kinome of metastatic melanoma cells relative to primary tumors. We contrast 538 total kinases and 3042 known kinase isoforms between 103 primary tumor and 367 metastatic samples from The Cancer Genome Atlas (TCGA). We find strong evidence of differential expression (DE) at the gene level in 123 genes (23%). Additionally, of the 468 genes with alternative isoforms, 60 (13%) had differential isoform ratios (DIR). Notably, DE and DIR have little correlation; for instance, although DE highlights enrichment in receptor tyrosine kinases (RTKs), DIR identifies altered splicing in non-receptor tyrosine kinases (nRTKs). Using exon junction mapping, we identify five examples of splicing events favored in metastatic samples.

We cluster isoform expression data and identify subgroups that correlate with genomic subtypes and anatomic tumor locations. Notably, distinct DE and DIR patterns separate samples with *BRAF* hotspot mutations and *(N/K/H)RAS* hotspot mutations, the latter of which lacks effective kinase inhibitor treatments. DE in *RAS* mutants concentrates in CMGC kinases (a group including cell cycle and splicing regulators) rather than RTKs as in *BRAF* mutants. Furthermore, isoforms in the *RAS* kinase subgroup show enrichment for cancer-related processes such as angiogenesis and cell migration. Our results reveal a new approach to therapeutic target identification and demonstrate how different mutational subtypes may respond differently to treatments highlighting possible new driver events in cancer.

Introduction

Melanoma is the deadliest form of skin cancer, with about 232,100 new cases and 55,500 deaths worldwide each year [1]. Although incidence is less than 5% of new cancer cases in the U.S., incidence and deaths worldwide continue to rise, especially in the young adult populations [2]. Stage 1 or 2 disease is easily treated by surgery, where 5-year survival rates are > 90% [1], but if not caught early tumors may metastasize to the nearby lymph nodes and then throughout the body. Once the disease reaches the brain, median survival time decreases to 5 months [3]. Thus, novel systemic treatments for metastatic melanoma are needed.

Kinases have become compelling cancer targets because they contain mutations that produce constitutive kinase activation and dysregulate signaling pathways in cancer. Among the 538 known kinase genes in humans, there are numerous relevant targets. Specifically, mutations have been observed in kinases serving as growth factor receptors [4], cell cycle regulators [5, 6], nuclear signaling [7], and apoptosis regulators [8]. In melanomas, *BRAF* is most commonly mutated, along with other kinases including *NRAS* and *NF1*. Fleuren et al. identified 23 additional kinases harboring driver mutations for melanoma, including the receptor *FGFR3* and cell cycle regulator *CDK4* [9]. Additional targets may remain undiscovered as atypical kinases, which can phosphorylate proteins but lack a typical kinase domain.

Along with chemotherapy and immunotherapy, treatments for advanced melanomas also incorporate small molecule kinase inhibitors (KI). There are currently 37 FDA approved KIs on the market for cancer treatment, with ~150 in clinical trials [10]. Targets of these small molecule KIs include *BRAF*, which occurs in about 50% of melanoma patients [1, 11], and *MEK*, a downstream signaling target of *BRAF* in the *MAPK* pathway. Despite initial successes for these drugs, limitations remain. For example, half of all *BRAF*-mutant tumors treated with *BRAF* inhibitors advance within 6–8 months post-treatment [12] whereas other hotspot mutations, such as in *NRAS*, lack effective KI treatments altogether [13]. Complementary targeted approaches in the form of immune-checkpoint blockers ipilimumab, pembrolizumab, and nivolumab, have recently been shown to significantly improve survival in some patients, even in those with wildtype *BRAF* [14-16]. Although these treatments do not work in the majority of patients [17], combining them with KIs may improve survival prospects. Thus while existing drugs show promise for a

subset of patients, new targets and combination therapies are in dire need to address treatment-resistant tumors, and especially those tumors with wildtype *BRAF*.

There are multiple forms of kinase dysregulation: activating mutations, overexpression, underexpression, copy number alterations, repression, and chimeric translocations; but there has been much less research into gene isoform distributions, in part due to the difficulty of estimating isoform composition from short read RNA sequences [18, 19]. For these data, computational approaches are required to estimate isoform counts prompting development of transcript alignment algorithms such as *RSEM* [20], and faster pseudo-alignment algorithms such as *kallisto* [21]. The gold-standard of isoform analysis might eventually be achieved through "3rd generation" long read sequencing technologies such as PacBio [22] and Oxford Nanopore [23], providing more accurate, contiguous isoform sequences, although these currently have a high error rate and are costly compared to 2nd gen. sequencing [24]. Regardless, long and short read sequencing technologies both discern differential isoform composition to address the question of how alterations in sequential exon continuity can change functional outcomes.

Although isoform distributions are not widely reported in the literature, there is reason to suspect they are altered in cancer tissues. First, alternative splicing is highly abundant under normal conditions where up to 94% of human genes undergo alternative splicing [25], and the dominant isoform depends on cell type [26]. Second, in various cancers, splicing factors can be mutated or mis-regulated [27-31], potentially skewing isoform distributions. Third, somatic DNA mutations – abundant in cancer – may occur on splice sites and splice regulators, favoring or suppressing splicing events. Kinases are known to undergo alternative splicing events in cancer [18] and these are implicated in tumor progression. Examples include *MKNK2* in glioblastoma [32]; *CD44* in breast cancer [33]; and *KLF6* in prostate, lung, and ovarian cancers [34]. Splicing induced losses or gains of functional or regulatory domains, documented in cancers, altering the functions of affected proteins in the cell. Despite these observations, differential isoform usage is an extra level of detail not normally analyzed in cancer studies.

Here we propose to detect and demonstrate the biological relevance of isoform alterations in metastatic melanoma. Notably, a recent study of the human kinome in prostate cancer found that

there was little overlap between genes with differential expression and genes with differential splicing [35], suggesting a study of the latter will yield additional therapeutic targets. Despite our emphasis on differential isoform expression, we include differential expression of genes (i.e., representing a gene locus with a single expression value), to show distinct and relevant findings learned from each type of assessment.

In this study, we analyze RNA-seq data from The Cancer Genome Atlas (TCGA) skin cutaneous melanoma project (SKCM) to study changes to the kinome of metastatic vs. primary tumor melanomas. Important findings include isoforms downregulated in metastatic samples that correspond with known and novel suppressors of metastasis and additional subgroupings of metastatic samples with narrowly focused therapeutic potential. Our results identify characteristics of wildtype *BRAF* tumors, as well as new subdivisions among *BRAF* mutant tumors.

Methods

Human kinome

We obtained Gene IDs for 538 human kinases from the Human Kinome database [36] at http://kinase.com/web/current/kinbase/. Kinases are classified into 10 phylogenetic groups: tyrosine kinases (TKs); "Sterile" serine/threonine kinases (STEs); calmodulin-dependent kinases (CAMKs); Cdk, MAPK, GSK, and Cdk-like related kinases (CMGCs); protein kinase A, protein kinase G, and protein kinase C related kinases (AGCs); tyrosine kinase-like (TKLs); casein kinase 1 (CK1); receptor guanylate cyclases (RGCs); atypical kinases (aPKs); and "other" typical kinases. For our analysis, we further subdivided TKs into receptor tyrosine kinases (RTKs) and non-receptor tyrosine kinases (nRTKs) due to their distinct functional roles.

TCGA data

We obtained RNA-seq data and kinase gene counts – estimated using *HTSeq* [37] – from the National Cancer Institute (NCI)'s Genomic Data Commons (GDC) portal for TCGA's skin cutaneous melanoma (SKCM) project. This included data from 472 samples gathered from 468

5

patients: 367 samples for metastatic tumors, 103 for primary tumors, 1 for an additional metastatic tumor from the same patient, and 1 for solid normal tissue. The latter two samples were not used in our analysis.

The data was processed in 14 batches, with the largest batch (labeled "A18") having 218 of the samples in three plates. The remaining batches had 10-48 samples in a single plate each.

Isoform quantification

For the purpose of quantifying the abundance of isoforms in the human kinome, we used the *kallisto* (v0.45.0) package [21] in conjunction with the transcript sequences of protein coding genes in the Gencode (release 29) annotation of the human genome. We first constructed the *kallisto* index file using the 98,913 FASTA sequences of transcript isoforms of human protein coding genes included in the Gencode annotation (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_29/gencode.v29.pc_transcripts.fa.gz; accessed March 15, 2019). FASTQ-formatted RNA-Seq reads (48-bp, paired-end) for each TCGA SKCM sample were produced from the bam files obtained from the Genomics Data Commons Data Portal. In order to avoid biases in *kallisto* estimates of fragment lengths, for each sample we produced FASTQ files in which the order of the reads was randomized. We then used these randomized reads to perform the *kallisto* "quant" analysis, from which we obtained the transcripts per million (tpm) estimates of each isoform abundance.

Sample quality control

3' bias for each sample was estimated using the QoRTs package [38]. For sample purity, we used the consensus purity estimate from Aran et al. [39]. Samples with purity < 70% were removed to create our "high purity" sample set. Samples with a QoRTs 3' bias score > 0.55 (see ref [38] for Methods) were also removed in our "quality controlled" set. After clustering kinase isoform expression in metastatic samples, we also classified 83 metastatic samples as having amounts of immune infiltrate using k-means clustering with 2 centers (see Clustering of Metastatic Samples below).

6

Differential expression (DE)

We tested differential expression of all genes between primary tumor and metastatic samples using the *DESeq2* toolbox for R [40] with two models: "sample type" and "sample type + batch" to account for batch effects.

Genomic subtypes for 56 primary tumor and 260 metastatic samples were obtained from Akbani et al. [11]. They included *BRAF* hotspot mutants (47%), *RAS [N/H/K]* hotspot mutants (29%), *NF1* mutants (9%), and Triple Wildtype (15%). The remaining 156 samples were added after the study and had no genomic subtype information.

Calculations for differential isoform ratios (DIR)

Transcript isoform counts for the TCGA samples were estimated from RNA-seq data with *kallisto* [21], using isoform information for protein coding loci provided by Gencode v.29 transcriptome annotation. In total, there were 3,042 protein coding isoforms for the human kinome. Isoform counts (in transcripts-per-million or TPM) for each gene were grouped as a vector, and the vector was normalized to sum to 1. One vector per sample was made, ignoring samples with zero counts for all isoforms.

We used two models to test for differential isoform ratios. The first was a permutation method utilizing linear discrimination analysis (LDA). LDA was performed to reduce the space of isoform vectors to the 1D line which best separates sample types, and the LDA statistic

$$\frac{(\mu_{PT} - \mu_{Met})^2}{\sigma_{PT}^2 + \sigma_{Met}^2}$$

was calculated. The sample labels were then randomized n_{iter} times and the statistic recalculated to create a null distribution, from which the p-value was found. This method had the benefit of producing a single p-value without assumptions, but could only find p-values as low as $1/n_{iter}$.

In the second model principal component analysis (PCA) was performed on the space of normalized isoform vectors, providing us with "n-1" components for "n" isoforms. PCs with zero variance were removed. We tested the difference in isoform coordinates between sample types along each PC using one of three different statistical tests (see below) and combined the p-values using Fisher's method. For both models, p-values were adjusted using Benjamini-Hochberg FDR adjustment.

Comparison of statistical tests

Given that the permutation test becomes computationally prohibitive for large datasets and high precision, we attempted to find a statistical test that could reproduced the results obtained through permutations. We used three different tests along the principal components of the space of isoform vectors: the Wilcoxon rank sum test, Welch's t-test, and the general independence test from R's conditional inference (coin) package [41]. We combined the p-values from each principal component with both Fisher's method (FM) and the asymptotically exact harmonic mean (HMP) from DJ Wilson [42]. This resulted in six sets of p-values which we compared to the permutation test results.

We found that the t-test combined with Fisher's method gave the best correlation between p-values ($r = 0.92$) and ranks ($\rho = 0.92$), while the coin test combined with HMP gave the best correlation between the logarithm of p-values ($r = 0.89$). However, total correlation may be of less interest than the sensitivity and specificity of the tests. We calculated Youden's J statistic (sensitivity + specificity – 1) at three significance levels: $p = 0.05, 0.01,$ and 0.001. The t-test combined with Fisher's method performed best at all three levels, with $J = 0.79, 0.80,$ and 0.80 respectively, followed by the coin test with Fisher's method. The geometric mean of these two tests $p_{new} = \sqrt{p_{t-test}p_{coin}}$ performed better, with $J = 0.80, 0.84,$ and 0.86 respectively, and also increased all three correlations. We thus adopted this test for scaling up the number of genes. The Wilcoxon test performed poorly due to difficulties handling ties in the data.

Clustering of metastatic samples

A quasi-Poisson generalized linear model (GLM) was used to test each individual metastatic sample vs. all primary tumor samples for each protein-coding isoform – using TPM counts from *kallisto* – resulting into a 3,042 x 367 matrix of p-values. Before clustering the data was digitized, setting all p < 0.05 to +1 for isoforms with increased expression, all p < 0.20 to -1 for isoforms with decreased expression, and all other entries to 0. The reason we used such a liberal p-value for negative change is because most count data follow a Poisson-like distribution with a low median, which makes decreased expression for individual samples unlikely to test as significant. For example, isoform SLK-202 tests as highly significant for decreased expression (p = 3.4e-9) for all metastatic vs. primary tumor samples but only tests as significant (p = 0.0014 and 0.038) for two individual samples.

After digitizing, we applied k-means clustering to the data matrix, using the elbow method to find an appropriate number of clusters. Enrichment for tumor region, mutation subtype [11], batch ID, and kinase phylogenetic group in each cluster were tested using Fisher's exact test.

Gene biological process and kinase group enrichment

1,572 biological process (BP) annotations were downloaded from the PANTHER database at geneontology.org. Genes were ranked by p-values (for DE or DIR) and significant genes tested for enrichment using the one-sided Fisher's exact test (i.e. hypergeometric test), using the remaining kinase genes as the background. We found that enrichments could differ drastically depending on the p-value threshold chosen for significance, so we searched for BP enrichment at multiple thresholds. Additionally, testing for DE or DIR with small sample sizes produced less extreme p-values than testing with large sample sizes, resulting in comparing >300 significant genes from one set of results (more than half the kinome) to <10 genes in another set of results. So we tested four percentile-based thresholds – the top 5%, 10%, 20% and 40% of all genes with a p-value – to obtain a comparable set of enrichments between sample sets. Results described are for the top 5% of genes unless noted otherwise.

We did not adjust p-values for the biological processes for several reasons. Having discovered a set of significant genes, our question was what functional purpose do these genes serve? Some annotations, such as "protein kinase", will never test as significant because all the genes in our background and foreground are kinases, making the expected false discovery rate lower than assumed in Benjamini-Hochberg (BH) correction. Furthermore, GO terms are highly dependent, making common adjustment methods such as BH inappropriate. Finally, GO terms do not account for individual isoform activities, thus do not address our underlying question.

We did calculate an empirical false discovery rate (see Supplemental Results, Empirical FDR) merely to compare our enrichment results to those of a randomly selected set of "significant" genes.

Kinase phylogenetic group enrichment (see "Human kinome" above) was calculated in the same manner using percentile thresholds, with p-values unadjusted.

Split-read alignment mapping

To evaluate changes in the relative abundance of isoform using an alternative method, we quantified the relative abundance of split reads specifically associated with the isoform of interest. For this purpose, we aligned the RNA-Seq reads using STAR against the hg19 version of the human genome assembly. We used the QoRTs package [38] to quantify split read support for splice junctions. For cases of alternative promoters, we compared the relative abundance of split reads supporting a common exon junction with alternative upstream exons. In cases of isoforms differentiated by a skipped exon, we considered the reads supporting the junction skipping the exon, and the average number of reads supporting the two junctions of the alternatively spliced exon. The relative abundance was expressed as a fraction of reads specifically supporting one isoform out of the total number of reads supporting both isoforms. The difference in the relative abundance was compared between primary tumor and metastatic samples using a one-sided Wilcoxon rank sum test, guided by the expectation set by the output from the *kallisto* tool.

SLK experiment

We transiently overexpressed short-length *SLK* (*SLK-201*), and full-length *SLK* (*SLK-202*) in the metastatic melanoma cell line A375 (ATCC, Manassas, VA). Short-length *SLK* and full-length *SLK* were cloned into the GFP fusion expression vector, p-RECEIVER-M98 (Genecopoeia, Rockville, MD). A375 cells were grown in RPMI Medium 1640 with 10% FPS and 1x Antibiotic-Antimycotic (Gibco, Gaithersburg, MD). A375 cells were transfected with lipofectamine 2000 (Invitrogen, Carlsbad, CA). The cells were stained with Aqua LIVE/DEAD™ (Invitrogen, Carlsbad, CA) and analyzed by FACS at 24h, 48h, and 72h time points. We analyzed percent change in GFP in live cells over time for the empty vector-GFP, short-length *SLK*-201-GFP, and full-length *SLK*-202-GFP. The differential percent change in GFP in the *SLK* constructs compared to the empty vector control was used to measure isoform specific cell death.

Results

We analyzed the 538 kinase genes comprising the human kinome, for changes in total mRNA expression and 3042 isoforms for altered isoform expression, between metastatic and primary tumors. Using computational tools *HTSeq* [37] and *kallisto* [21] with short read sequences, we implemented the data analysis workflow depicted in Figure 1. We included quality control measures to ensure transcript integrity and control for tumor purity. Along with differential expression defined at the gene level and differential isoform ratios calculated within each locus, we performed a clustering analysis to identify pathway, mutational and functional characteristics that define each subgroup.
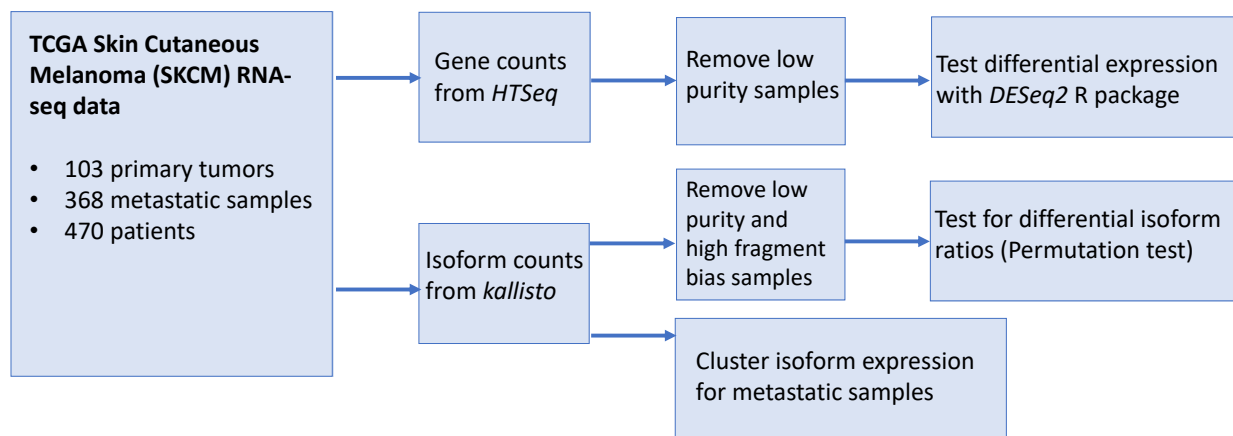

Figure 1: Data analysis workflow

Sample demographics

Primary (n = 103) and metastatic (n = 367) tumors were obtained from the TCGA skin cutaneous melanoma project (SKCM) (Table 1). Primary tumors originated in a number of locations including arms or legs, trunk, head or neck, or other areas, such as armpit, genitalia, etc. Metastatic locations included regional cutaneous or subcutaneous tissue, regional lymph nodes, distant metastases, and unclassified metastases. Samples were skewed towards males, and mostly derived from white individuals. Patient age at time of diagnosis ranged from 15 to 90, with a median of 58-years-old.

Table 1: Sample Demographics

Sample details	Variables	Sample numbers		
type	-	Normal tissue	Primary	Metastatic
number	-	1	103	367
tumor origin	arms or legs	-	41	153
	trunk	1	48	125
	head or neck	-	8	30
	other (armpit, genitalia, etc.)	-	4	9
	unknown	-	2	50
metastatic location	regional cutaneous or subcutaneous tissue	-	-	74
	regional lymph nodes	-	-	221
	distant metastases	-	-	68
	unclassified metastases	-	-	4
genomic subtype	*BRAF* hotspot mutation	-	32	118
	RAS hotspot mutation	-	11	81
	NF1 any mutation	-	5	23
	triple wildtype	-	8	38
	not available	-	47	107
sex	male	1	61	230
	female	-	42	137
age (median)	-	51	65	56
race	white	1	94	353
	Asian	-	7	5
	non-white Hispanic	-	1	2
	black	-	0	1
	unknown	-	1	6
# of batches		1	14	14

Differential expression (DE) dominated by receptor tyrosine kinases

We first tested differential expression at the gene level. Out of 538 kinase genes, 281 (52%) had significant DE ($p_{adj} < 0.05$) between all primary tumor and all metastatic samples (Table S1). The top groups, ranked by p-value, included both non-receptor (nRTKs) and receptor tyrosine kinases (RTKs) (Figure 2). We looked for biological process enrichment in the top 5% and 10% of genes, and found strong enrichment for immune cell activation (both innate and adaptive). Clustering analysis (see Methods) revealed these genes have strongly correlated expression, suggesting their high expression results from immune infiltrate in the metastatic samples, i.e. immune cells attacking tumor cells. Using this approach, we identified 83 metastatic samples with high amounts of putative immune infiltrate (see Methods), which we removed before rerunning the *DESeq2* analysis. This action removed the enrichment for nRTKs, whereas RTK enrichment remained (Figure 2B). We next addressed the impact of impure tumor samples, as measured by the consensus purity estimate developed in Aran et al. [39]. When 168 samples with < 70% estimated purity were removed from the original set, which included 80 of the 83 immune infiltrate samples, again we saw enrichment for nRTKs disappear whereas RTKs remained significant (Figure 2C). We named this filtered group the "high purity" (HP) group. In both assessments, enrichment for RTKs remained significant when assessed as subsets of the top 5% to ~ 20% of genes. There was also a lesser enrichment for the STE kinase group (p = 0.031 at 20% threshold; Figure 2C), which contains kinases upstream of MAPK signaling cascades. When only the high purity (HP) samples were compared we found 197 significant genes including 26 of the 57 RTKs (Table S2). Of these 26, 11 are known to be trafficked to the nuclear membrane [43]: *FGFR1/3*, *FLT1*, *ERBB4*, *INSR*, *TIE1*, *CSF1R*, *EGFR*, *IGF1R*, *MET*, and *KDR*. Absolute fold-changes for significant genes ranged from 0.503 (*KSR*) to 11.7 (*NRK*).

Next we examined differential gene expression when the HP primary tumor and metastatic samples were subdivided into their particular genomic subtypes (*BRAF*, *RAS*, Triple Wildtype and *NF1*) (Table S3). Although this approach reduced sample size for each significance test (Table 3), a similar enrichment pattern emerged. For example, the DE genes for the *BRAF* hotspot mutants, *NF1* mutants, and Triple Wildtype samples were all enriched for RTKs (odds ratios = 4.0, 6.9, and 3.2 respectively at 5% threshold; Figure 2D). The deviant result was the case of the *RAS* hotspot

mutants, where DE was dominated not by RTKs but by CMGC kinases (odds = 4.1, Figure 2E). This group contains both cyclin-dependent kinases – which regulate the cell cycle – and downstream MAP-kinases – which regulate gene expression – as well as kinases directly involved in splicing regulation (i.e., serine arginine protein kinases). Although RTKs (particularly Ephrin receptors; i.e., *EPHA*) remain significantly altered in the RAS mutants, this result suggests a distinct set of alterations are necessary for RAS hotspot mutants to become metastatic.
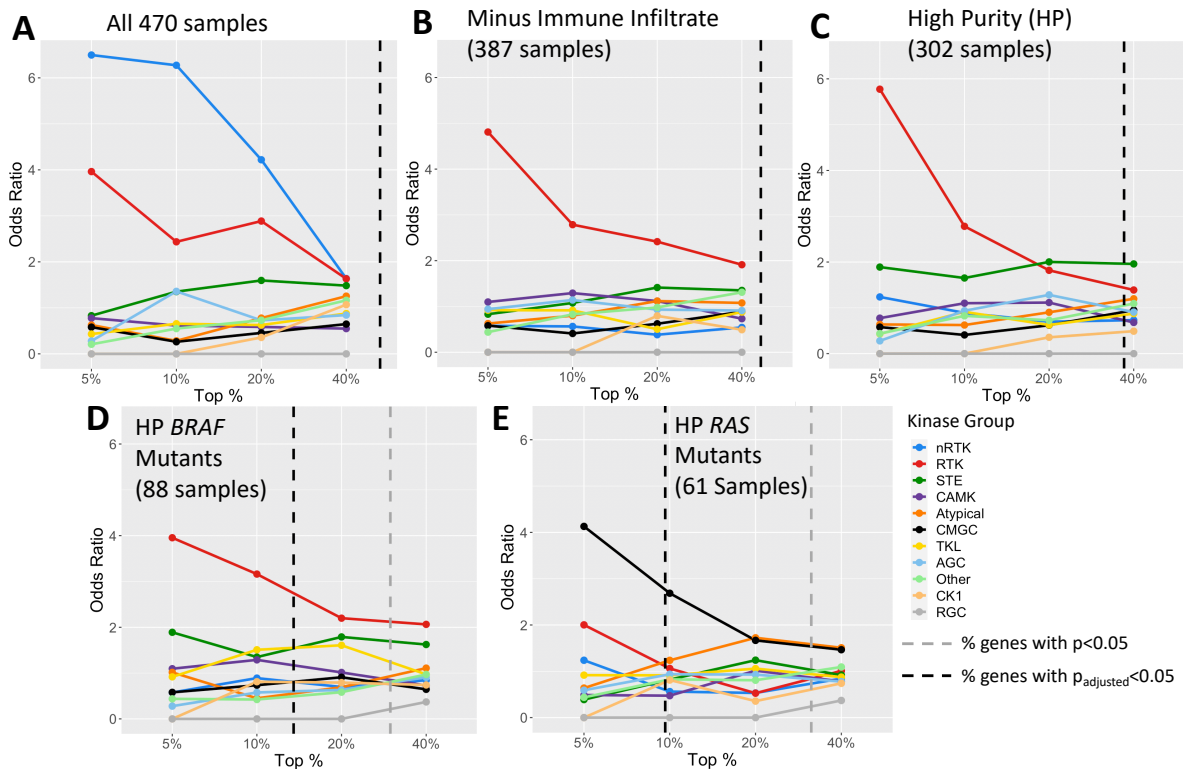

Figure 2: Kinase group enrichment for differential expression of primary and metastatic tumors differs by sample set. Depicted are the odds ratios for each kinase group in the top 5%, 10%, 20%, and 40% of kinase genes, ranked by p-value. This indicates that the strongest DE enrichment is concentrated in nRTKs for all 470 samples, RTKs for high purity and *BRAF* mutant samples, and CMGC kinases for *RAS* mutant samples. Enrichment data collected at the four percentile points are independent of p-value; the percent of genes that have significant DE (p<0.05) before and after p-value adjustment are shown by the gray and black dotted lines respectively. Sample type (primary tumor or metastatic) was the only model variable for the DESeq2 results.

Table 2: Differential expression from primary tumor to metastatic samples in receptor tyrosine kinases (RTKs)

Gene Name	Base Mean	Fold Change	P-value	P-adj	Nuclear Trafficked [43]
EPHA1	149.44	-6.13	8.15e-31	1.82e-28	
FGFR3	496.72	-4.58	4.29e-14	3.72e-12	Yes
LMTK3	84.69	-2.99	4.87e-10	2.97e-08	
EPHA3	671.47	3.20	7.27e-09	3.77e-07	Yes
EPHB6	392.00	-2.79	1.31e-07	5.62e-06	
FLT1	1539.25	1.57	4.10e-06	1.24e-04	
MERTK	993.10	1.86	2.16e-05	5.06e-04	
ROR1	409.93	1.75	4.58e-05	9.44e-04	
FGFR2	232.15	-2.54	6.50e-05	1.25e-03	Yes
EGFR	731.80	-1.95	3.58e-04	4.76e-03	Yes
EPHA2	3375.91	-1.56	1.74e-03	0.0156	
TIE1	895.76	1.42	3.63e-03	0.0267	Yes
EPHA6	16.09	2.24	5.71e-03	0.0371	
DDR2	6622.61	1.36	6.66e-03	0.0416	
RYK	2858.81	1.17	8.31e-03	0.0484	Yes
INSR	2475.26	1.25	8.60e-03	0.0495	Yes

Results are between high purity (>70%) primary and metastatic tumors using sample type and batch ID as model variables. Only RTKs with $p_{adj} < 0.05$ are shown. Negative fold change indicates decreased expression in metastatic samples

Influence of sample batches on differential gene expression

Because not all the samples in the TCGA data set came from the same batch, we also ran *DESeq2* using both sample type and batch ID as model variables. This approach increased p-values, decreasing the number of significant genes. However, 123 kinase genes remained significant at the $p_{adj} < 0.05$ level, including 16 RTKs (Table 2), compared to 197 total genes when only the sample type was the variable. Gene ranking was not substantially altered (Spearman correlation $\rho = 0.81$) and enrichment trends were similar to our prior results for all genomic subtypes (*BRAF, RAS, Triple WT*), with the exception of the *NF1* mutant samples. These could not be assessed due to the small sample size (2 primary and 11 metastatic tumors), where the primary tumors and metastatic samples were not from the same batch. Excepting this subtype, for the remaining analyses we included batch ID as a model variable.

Biological process (BP) enrichment differs by genomic subtype

In addition to the kinase group enrichment, for each set of results (i.e., all samples, high purity, *BRAF* mutants, etc.) we looked for BP enrichment among significant genes, as a hypothesis-free approach to further characterize the metastatic tumors (Table S4). Top genes were highly enriched for immune-related annotations when all 470 samples were used, the highest being "adaptive immune system" (p = 5.1e-9) (Table 3). These enrichments nearly disappeared when samples with < 70% purity were removed. Surprisingly, when only analyzing HP samples BP annotations were depleted, with only four annotations receiving a p-value below 0.05. "Ephrin receptor signaling pathway" (p = 0.033); e.g. *EPHA*, which generates contact-dependent signaling between neighboring cells and transduces signal from the cell exterior to the interior through ligand-induced activation of the kinase domain; was the only non-immune-related enrichment. Ephrin receptors are prototypical RTKs that impact cell shape, adhesion, and movement through activation or repression of the Rho GTPase family [44], suggesting an important role in metastatic processes.

This lack of BP enrichments suggests either that DE is widely distributed among a number of cell processes, or that enrichment patterns differ by genomic subtype and disappear when lumped together. To address this question, we separated the high purity samples into genomic subtypes and found support for the latter hypothesis, where division into individual subtypes revealed enrichment in distinct processes (Table 3). We observed strong BP enrichment among DE genes for samples with *BRAF* mutations, with the most significant annotation being "cell differentiation" (p = 1.3e-4). Neurogenesis and cell projection-related enrichments were also discovered. The DE genes for *RAS* mutants had weaker enrichments, although select examples such as "positive regulation of defense response" and "regulation of angiogenesis" are relevant for cancer. The ephrin receptor signaling pathway was enriched in both the *BRAF* (p = 0.008) and *RAS* (p = 0.035) mutants.

The *NF1* mutant and Triple WT sets had smaller sample sizes (13 and 28 samples respectively). The *NF1* mutants were enriched for "regulation of MAPK cascade" (p = 0.0054), "chemotaxis", and "neuron projection guidance" among others. The Triple WT samples – unlike the other

genomic subtypes – were enriched for responses to cytokine stimulation, especially interleukin-1 (p = 0.0053), as well as the inflammatory response and defense response.

Table 3: Summary of kinase differential expression results.

Sample Set	Sample Size	P<0.05[1]	P$_{adj}$<0.05	Kinase Group Enrichment	Selected BP Enrichments[2]	P-value[2]
All Samples	PT: 103 Met: 367	262 genes	202 genes	nRTK RTK	-Adaptive immune response -Hemopoiesis -Innate immune response	5.1e-9 1.4e-7 6.1e-4
All Samples (Purity>0.7)	PT: 88 Met: 214	203 genes	123 genes	RTK STE	-Ephrin receptor signaling pathway	0.033
BRAF Hotspot Mutation (Purity>0.7)	PT: 27 Met: 61	99 genes	27 genes	RTK	-Cell differentiation -Positive reg. of lipase activity -Biomineralization -Positive reg. of neurogenesis -Positive reg. of cell projection organization -Ephrin receptor signaling pathway	1.3e-4 5.2e-4 5.9e-4 0.0016 0.0058 0.0077
RAS Hotspot Mutation (Purity>0.7)	PT: 9 Met: 52	133 genes	36 genes	CMGC	-Eye morphogenesis -Positive reg. of defense response -Reg. of angiogenesis -Ephrin receptor signaling pathway	0.0034 0.017 0.050 0.035
NF1 Any Mutation (Purity>0.7)	PT: 2 Met: 11	62 genes	12 genes	RTK	-Reg. of MAPK cascade -Eye morphogenesis -Chemotaxis -Neuron projection guidance	0.0054 0.0057 0.013 0.014
Triple Wildtype (Purity>0.7)	PT: 8 Met: 20	41 genes	9 genes	RTK CAMK	-Calcium-mediated signaling -Cellular response to cytokine stimulus -Inflammatory Response -Defense Response	0.0021 0.0031 0.0053 0.013

[1] Both sample type and batch ID were used as model variables for *DESeq2,* except for the *NF1* subtype where only sample type was used
[2] Enrichments are for the top 27 (5%) kinase genes ranked by p-value

Kinase genes exhibit differential isoform usage between primary and metastatic tumors

To complement the usual procedure of DE analysis, we next tested whether multi-isoform kinase genes exhibit differential isoform ratios (DIR) between primary and metastatic tumors. Per the Gencode v.29 annotation, we tested 468 such genes with 3,042 total coding isoforms. We measured significance in 317 (68%) via a permutation test (p$_{adj}$ < 0.05) when all 470 tumor samples were used, more genes than had tested significant for DE. Our complementary PCA test (see Methods) found p-values as low as 5.3e-28 for *LIMK1*.

This high level of observed DIR could be an artefact of sample impurity – since different cell types might express isoforms in different ratios – or experimental artefacts such as fragment sequence bias [45]. Fragment bias results from degraded RNA reads. Because these reads are sequenced from the 3' end following poly(A) enrichment protocols, high levels of degradation results in overestimation of 3' fragment isoforms and underestimation of 5' fragment isoforms (Figure 3A,B), although the total gene count estimate is unaffected.

We inspected of the isoform counts and found that genes with the strongest DIR had 3' fragment isoforms, suggesting samples with high 3' fragment bias could be driving significance. This bias was concentrated in the primary tumor samples (two-sided Wilcoxon, p = 1.3e-8) (Figure 3C). We also found sample impurity was concentrated in metastatic samples (p = 1.6e-4). Thus, both could contribute to the observed levels of significance.
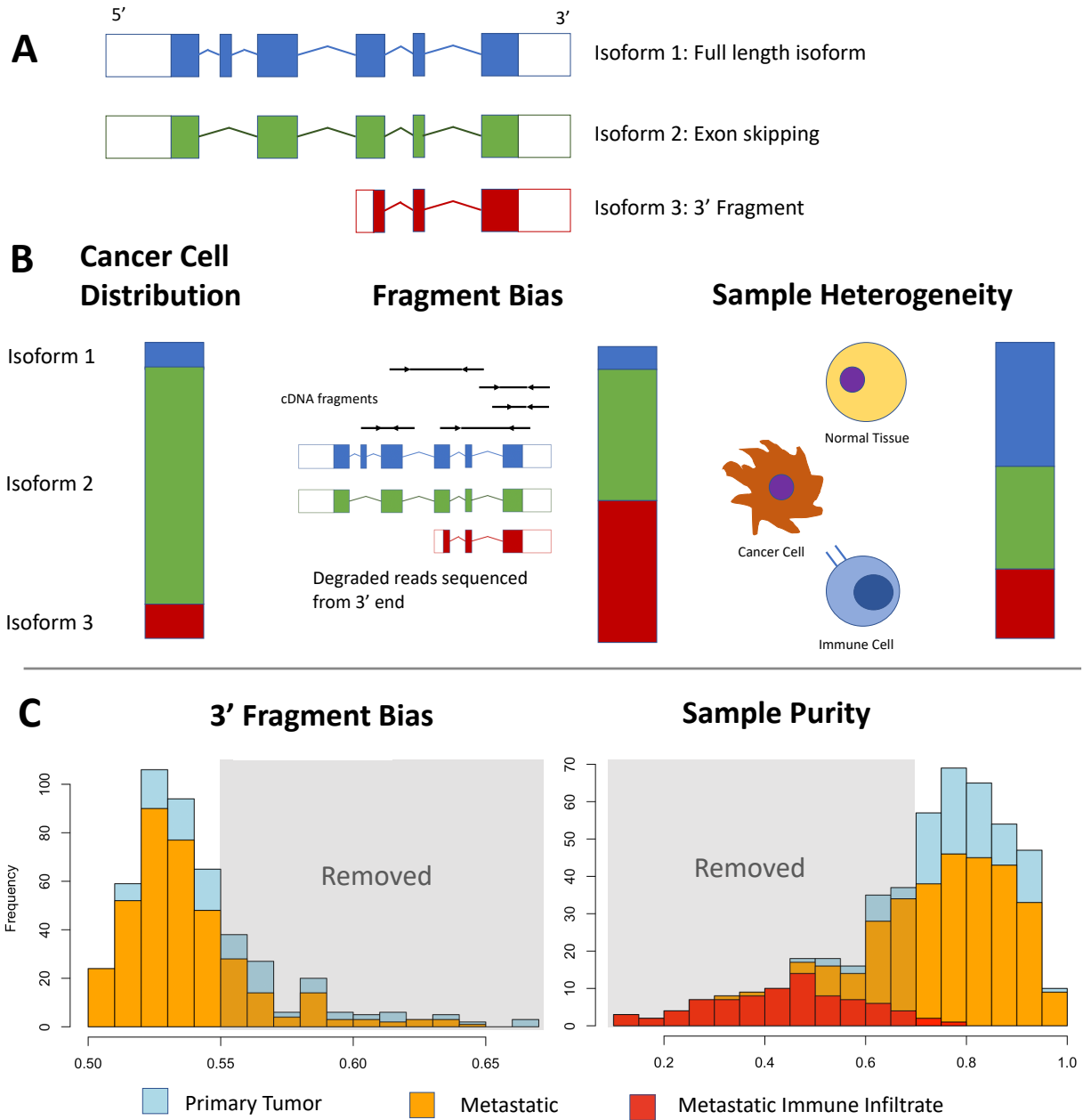
Figure 3: Examining data bias in isoform count estimation (A) Isoform examples. (B) Cancer cell isoform distributions can be altered by effects of fragment bias and cell type heterogeneity (i.e. sample impurity). The bars represent the apparent relative expression level of each isoform, with the left bar indicating the true distribution in a cancer cell and the right two bars indicating how data bias can skew the results. (C) Primary and metastatic tumor samples assessed for 3' fragment bias using QoRTs, where samples scoring above 0.55 were removed, and sample impurity, where samples with < 70% purity were removed. Red bars in the sample purity assessment indicate metastatic samples with observed high immune infiltrate.

Using the histograms as a guide, we removed samples with less than 70% purity or a QoRTs score > 0.55 from further analysis (Figure 3C). This reduced the number of samples to 50 primary tumor and 178 metastatic (Table S3), which we deemed the "quality-controlled" (QC) sample set. In this stringent QC set, only 60 genes had DIR with $p_{adj} < 0.05$ (Permutation test), and the most significant kinase was *SLK* at p = 7e-6 (Table 4, full list in Table S5). As a *post-hoc* analysis we tested the effects on individual genes by removing samples one-by-one to assess the influence of fragment bias or sample impurity. (See Supplemental Results, Figure S1, Figure S2)

Table 4: Example genes with significant changes in isoform ratios after 3' bias and sample impurity filtering.

Gene	DIR, p-value	DIR, p_{adj}	Coding Isoforms	Splicing/isoform changes in metastatic samples	Significant DE?	
SLK	7.00E-06	0.00328	2	Skipping of 13th exon, part of coiled-coil region	**Yes**	↑
TGFBR1[1]	3.40E-05	0.00590	9	Selection for 3rd exon, encodes transmembrane domain		
MAP3K3	5.60E-05	0.00590	5	Skipping of 3rd exon, precedes PB1 domain		
COQ8B[1]	6.80E-05	0.00590	11	Selection for 6th exon, effects on function unknown	**Yes**	↓
ABL1	8.00E-05	0.00590	3	Increase of full-length isoforms, decrease of 5' fragment		
LIMK1	0.000128	0.00750	3	Alternate promoter site. Shortens 1st zinc- binding domain.		
PAN3[1]	0.00027	0.0108	2	Selection for 4th exon		
FGFR3	0.00028	0.0108	7	Decrease of all isoforms, equalizing ratios	**Yes**	↓
FES	0.00037	0.0108	9	Skipping of 11th exon, encodes SH2 domain		
UHMK1[1]	0.00067	0.0154	3	Alternate promoter site favoring longer isoform with ATP-binding region		
PAK6	0.00081	0.0173	14	Decrease of all isoforms except a middle fragment	**Yes**	↓
BLK	0.00215	0.0273	3	Unequal increase of two major isoforms	**Yes**	↑
MKNK2	0.0069	0.0531	8	Alternate splicing at terminal exon, increase of isoform without MAPK binding site	**Yes**	↓

[1]Results supported by *kallisto* counts but not supported by exon junction read mapping

Differential gene expression does not predict differential isoform ratios

Having controlled for fragment bias and impurity, we asked whether genes with differential expression between primary and metastatic tumors were also likely to exhibit DIR. We compared the p-values for DIR from the QC set to the p-values for DE from the high purity set. Genes with $p_{adj} < 0.05$ had non-significant overlap (Fisher's exact test, p = 0.310), with only 15 genes overlapping (Figure 4). Gene rank (by p-value) had no correlation (Spearman, $\rho = 0.038$). The gene *MAP3K3*, for example, had the third highest level of DIR (p = 5.6e-5) but no observed change

in expression (p = 0.77, rank 487). Interestingly, genes with significant DIR were enriched for nRTKs (Figure 5A) but not RTKs, the opposite of what we observed for DE genes. Thus DE and DIR concentrate in different genes.

We separated the QC samples into genomic subtypes, as we did for the DE analysis, and calculated DIR for each subset. Due to the small sample sizes, few genes tested as significant with our permutation test. For example, the *BRAF* mutants revealed only four genes with $p_{adj} < 0.05$ (*SLK*, *MOK*, *ABL2*, and *SYK*) while the other 3 subtypes revealed no significant genes after p-value adjustment (summarized in Table 5). As seen for the full sample set, no ranked gene list for any DIR sample group correlated with its DE counterpart.
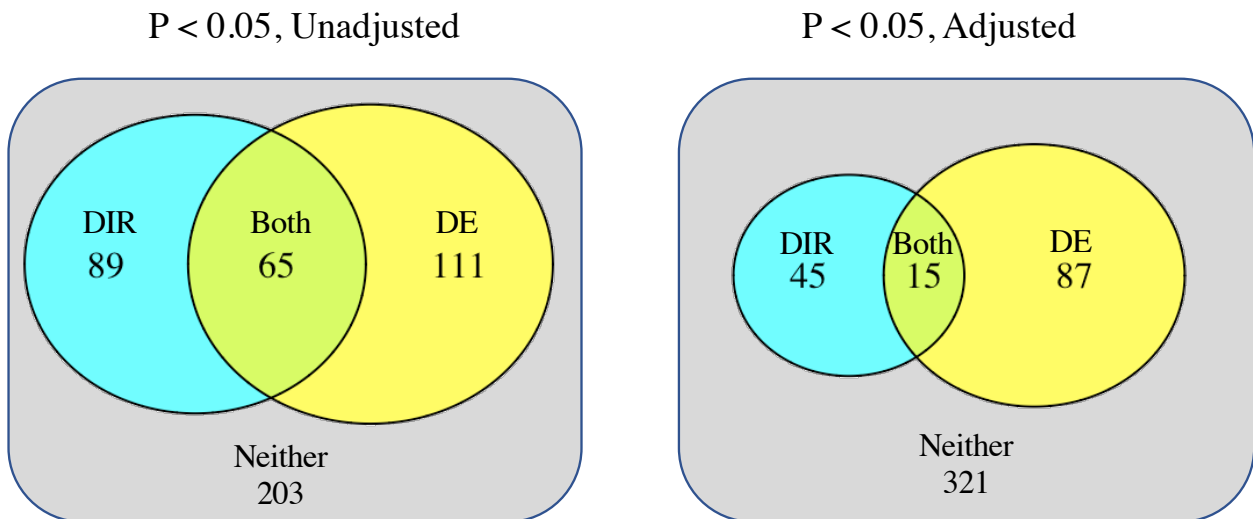

Figure 4: DE does not overlap with DIR. Significance in one does not predict significance in the other (One-sided Fisher's exact test, p = 0.091 (left) and 0.310 (right) Compared are results using high purity samples for DE (model ~ sample type + batch ID) and quality-controlled samples for DIR. Compared are the 468 kinase genes with >1 coding isoform.

Table 5: Summary of kinase differential isoform ratio results.

Sample Set[1]	Sample Size	P<0.05	P_{adj}<0.05	Kinase Group Enrichment	Selected BP Enrichments[2]	P-value
All Samples	PT: 103 Met: 367	330 genes	317 genes	None	-Response to amino acid starvation -Cytoskeletal organization -Positive reg. of lipid kinase activity -Response to fibroblast growth factor -Blood vessel development[3]	3.3e-4 4.5e-4 0.0021 0.0057 8.7e-4
Quality Controlled	PT: 50 Met: 178	154 genes	60 genes	nRTK	-Reg. of endocytosis -Endothelial cell migration -Cell differentiation -Positive reg. of stress fiber assembly -Cytoskeletal organization	3.4e-4 0.0037 0.0050 0.0057 0.0084
BRAF Hotspot Mutation (QC)	PT: 17 Met: 52	57 genes	4 genes	nRTK	-Regulation of protein acetylation -B cell receptor signaling pathway -Reg. of cell motility	0.0084 0.012 0.031
RAS Hotspot Mutation (QC)	PT: 6 Met: 50	29 genes	0 genes	None	-Positive reg. of angiogenesis -Cellular response to VEGF -Chemotaxis	1.6e-4 0.0020 0.0075
Triple Wildtype (QC)	PT: 6 Met: 16	40 genes	0 genes	None	-Reg. of gene expression -Reg. of RNA splicing -Chromatin organization	0.0065 0.015 0.023

[1] Subgroups are comparable to Table 3, except the *NF1* mutant set, which had too few samples (1 primary tumor, 11 metastatic) for reasonable analysis.
[2] Enrichments are for the top 24 (5%) kinase genes with >1 coding isoform ranked by p-value
[3] Blood vessel development is concentrated in the top 47 (10%) genes

DIR affects different biological processes than seen for DE

Because many unadjusted p-values were significant for DIR we elected to search for gene ontology (GO) enrichments. For each sample set, we searched for biological process (BP) enrichment in the top genes (ranked by p-value) using percentile thresholds from 5% - 40% (see Methods). Enrichments are described for the top 5% (24) genes unless noted otherwise.

For comparative purposes, we examined the full sample set first without filtering, which contained low purity and high fragment bias samples, we revealed 221 BP terms with $p < 0.05$ and 10 additional terms with $p<0.001$. The most significant terms included "positive regulation of translation" ($p = 1.5e-4$), "cytoskeletal organization", "response to amino acid starvation", and "blood vessel development" (Table 5). Immune-related enrichments were strongest at the 40%

threshold, indicating putative immune infiltrate may affect DIR results, but the most significant genes were not immune-related.

The QC set had fewer BP enrichments than the full sample set (Table 5). These enrichments included "regulation of endocytosis" (p = 3.5e-4) "cytoskeletal organization", "Endothelial cell migration", "cell differentiation", and "cell cycle arrest" (Figure 5B), all of which have a putative relevance to cancer.

The genomic subtype sets revealed distinct BP enrichments – as they did when testing DE genes. In contrast to the DE genes, the DIR genes between *BRAF* mutant primary and metastatic tumors did not show strong BP enrichments, while the DIR genes between *RAS* mutant samples showed enrichment for 94 BPs. The strongest of these was "positive regulation of angiogenesis" (p = 1.6e-4) and related enrichments such as "vasculature development". Other enrichments included "cell-cell communication", "protein transport", and "membrane organization" (Figure 5C). Such enrichment patterns would not be discovered if DE alone was studied. In contrast, significant genes from the *BRAF* mutants had 27 processes enriched below p = 0.05 – these included 6 cell locomotion-related enrichments – and none below p = 0.008.
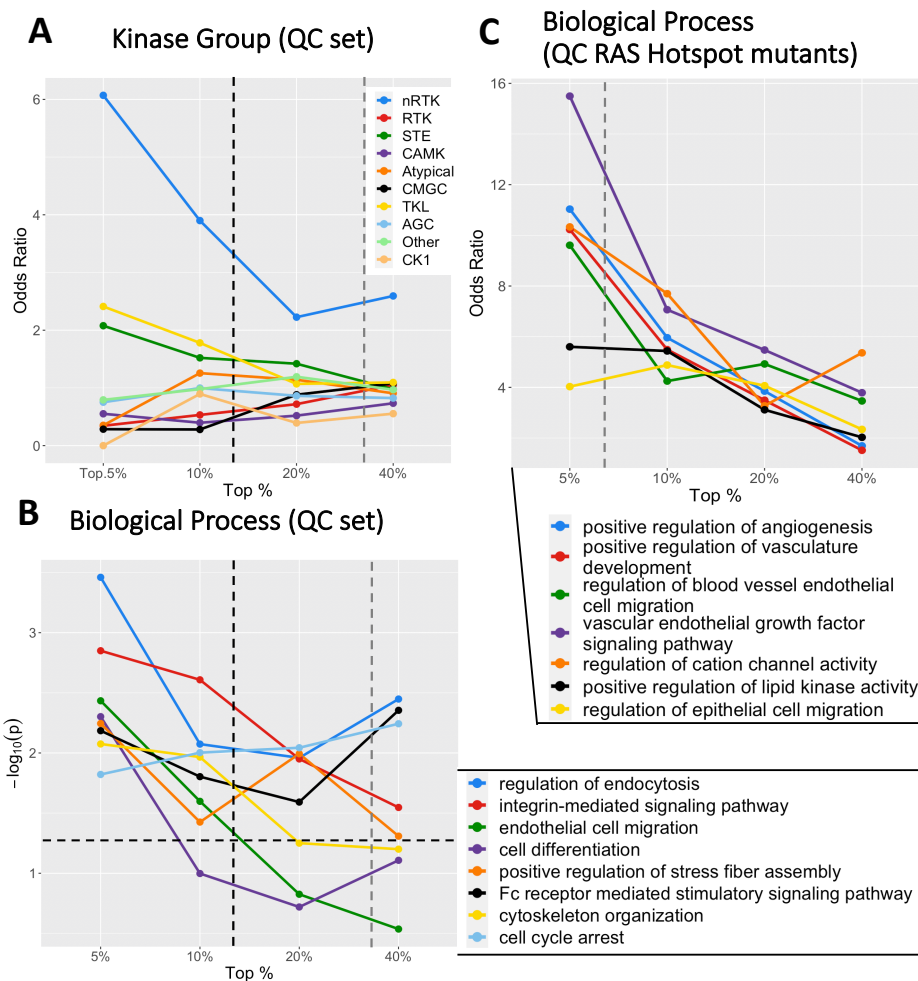
Figure 5: Genes with significant DIR display unique BP enrichment patterns. (A) DIR genes are enriched for non-receptor tyrosine kinases in the QC set, whereas there was no kinase group enrichment in the full sample set. The black dotted line indicates the percent of genes with $p_{adj} <$ 0.05, and the gray dotted line the percent of genes with unadjusted $p < 0.05$ (B) Select biological process enrichments for the QC set. Note that significance is plotted ($-\log_{10}p$), not odds ratios, and the horizontal dotted line indicates $p = 0.05$. (C) When using the QC RAS hotspot mutant samples only, DIR genes were highly enriched for angiogenesis and related annotations.

Resolving alternative splicing events in kinase genes

Focusing on DIR with discrete splicing changes, we identified skipped exons, alternative promoters, and alternative terminal exons (Table 4). For example, *ABL1* has two long isoforms (*ABL1-201* and *-202*), which differ only in their promoter site, that have increased expression in metastatic samples. An additional isoform *(ABL1-203)* encodes a shorter 5' fragment, which decreases in expression. However, *ABL1* does not test as significant in DE between primary and

metastatic samples, indicating that the DIR analysis can reveal aberrations that differential gene expression does not capture.

To test the *kallisto* DIR data for evidence of splicing differences, we directly mapped RNA-seq reads to the nucleotide sequences of exon junctions in several genes from Table 4. This provides a resolved view of exon splicing patterns in the samples which did not rely on *kallisto* (see Methods). Within the melanoma sequence data, we confirmed exon skipping in three genes – *MAP3K3* (exon 3), *FES* (exon 11) (Figure 6A,B) and *SLK* (exon 13) (Figure 7). We also confirmed switching to an alternate terminal exon in *MKNK2* (Figure 6C) and increased use of an alternative promoter in *LIMK1* (Figure 6D).-We illustrate the fraction of split reads, out of all reads, supporting these events.

In *SLK*, the most significant gene on our list, expression of the long isoform *SLK-202* is decreased, whereas the short isoform *SLK-201* increases (Figure 7). The short isoform skips exon 13 predicting a putative role for loss of this exon in cancer. We compared expression of this alternative exon in normal melanocytes using RNA-seq data from Zhang et al. [46]. Exon 13 was absent in the normal cells, and largely specific to primary tumor samples.

Some genes have DIR which coincides with significant DE. For example, 6 of the 7 coding isoforms of *FGFR3* are suppressed in metastatic samples (Figure S3), while the remaining isoform -205 has mildly increased expression, altering isoform ratios. *PAK6*, with 14 isoforms, undergoes a similar alteration. In *BLK*, DIR of 3 isoforms is driven by an unequal increase of 2 major isoforms, rather than all 3.
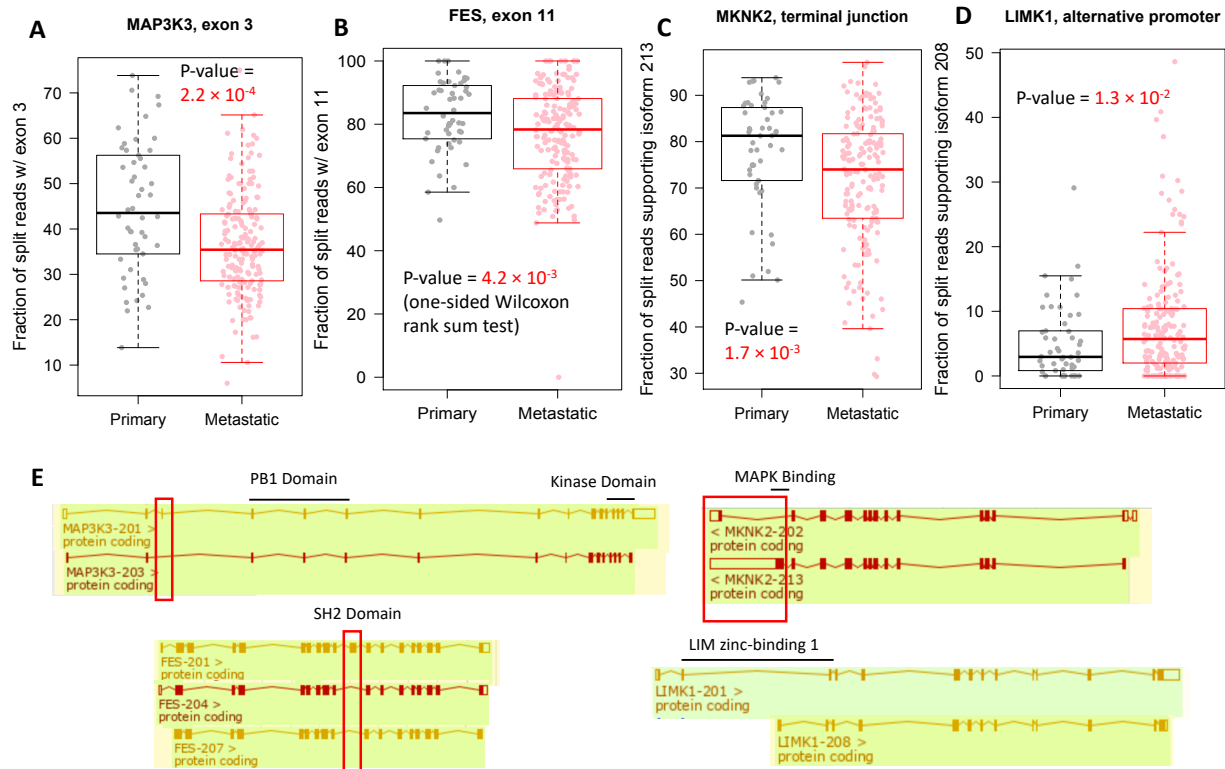
Figure 6: Alternative splicing in *MAP3K3*, *FES*, *MKNK2*, and *LIMK1* in primary and metastatic samples. Shown for each gene are box plots for fraction of split reads aligned to exon junctions in primary tumor and metastatic samples. **(E)** Maps of each isoform identify the differential exon.
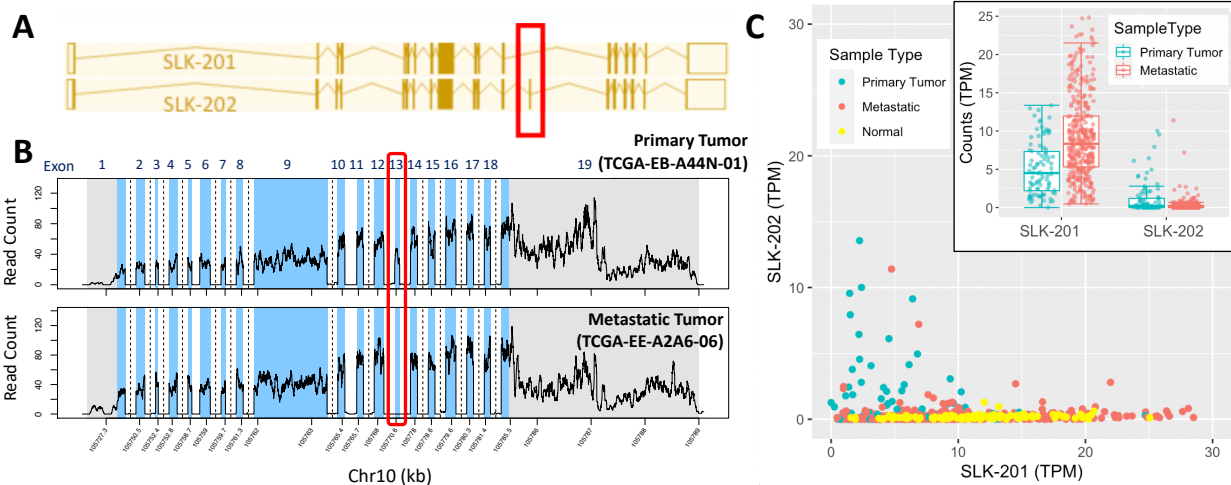

Figure 7: **Alternative splicing in *SLK*** (A) The two coding isoforms of *SLK*, which only differ in the presence or absence of exon 13. (B) Mapped read counts from one primary tumor and one metastatic sample. In the metastatic sample, no reads are mapped to exon 13, indicating that isoform *SLK*-202 is not present. (C) Box and scatter plots for the two isoforms of *SLK*. Also shown in the scatter plot are the normal melanocyte samples from Zhang et al. [46].

To address the functional consequences and biological implications of isoform switching, we matched the alternatively spliced regions in these five genes to domain annotations obtained from UniProt (Table 4). The skipped exon in *SLK* encodes a section of a coiled-coil region in the C-terminal domain. *SLK* uses this domain to dimerize at high concentrations, and these dimers activate apoptosis [47]. *MKNK2* switches to a shortened terminal exon which lacks the MAPK binding site, interfering with downstream signaling. The 11th exon of *FES* encodes the SH2 domain, which is necessary to activate the kinase domain [48]. The 3rd exon of *MAP3K3* is not mapped to any domain, but it precedes the PB1 protein-interaction domain. These data indicate that the isoform changes modulate the usage of important domains in the kinases, which can ultimately affect their function and participation in signaling networks. Finally, the alternative promoter of *LIMK1* utilized in some metastatic samples shortens the first zinc-binding domain, a domain that inhibit the protein's kinase activity [49].

SLK experiment

We wished to see if overexpression of the two *SLK* isoforms could produce cell death in metastatic melanoma. We hypothesized that expression of the full-length isoform (*SLK-202*) would produce more cell death compared to the short-length isoform (*SLK-201*) due to the lack of a dimerization domain (coiled-coil region) in the shorter isoform (Figure S6A). In this experiment, short-length and full-length *SLK* were cloned into the GFP fusion expression vector, p-RECEIVER-M98. We transiently transfected A375 metastatic melanoma cells with short-length *SLK*-GFP, full-length *SLK*-GFP, and empty vector-GFP. Cells were analyzed by FACS at 24h, 48h, and 72h post transfection. We analyzed change in percent GFP in live cells over time (Figure S6B).

We observed a large decrease in change of percent GFP in both *SLK* isoforms compared to the control indicating that both isoforms cause cell death in metastatic melanoma (Figure S6C). This suggests that the cell death pathway is still functional in metastatic melanoma, making *SLK* a candidate cancer therapeutic. No significant difference in cell death was observed between short-length *SLK* and full-length *SLK* (Figure S6C). The lack of differential cell death between isoforms may be due to the fact that both isoforms were expressed at concentrations higher than the biological norm. Short-length *SLK* retains other coiled-coil regions and may still dimerize at high concentrations. To see a difference in cell death between isoforms, these proteins may need to be

expressed at biologically relevant concentrations in *SLK*-/- A375 cells and over a longer time period to facilitate dimerization.

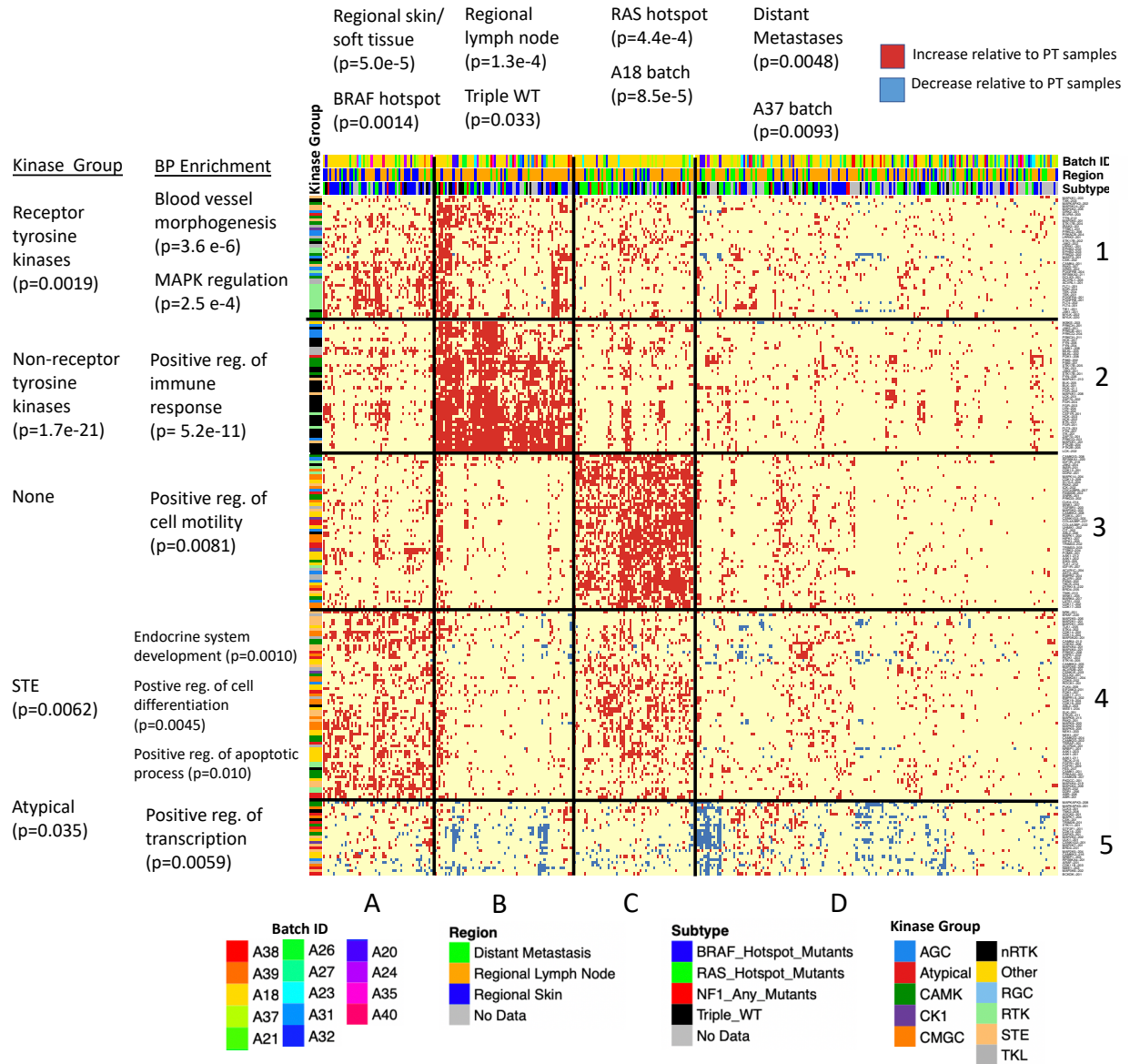## Clustering on DIR identifies correlations with genomic subtype and tumor location

Figure 8: Heatmap of 367 metastatic samples clustered according to kinase isoform counts. Red dots indicate increased expression in metastases (Quasi-Poisson GLM, p < 0.05) while blue dots indicate decreased expression (p < 0.2). Shown are the 367 metastatic samples (columns) and 235 isoforms that were altered in > 13% of samples (rows). P-values for enrichments were calculated using the one-sided Fisher's exact test.

To identify similarities in metastatic samples based on isoform expression patterns, we clustered the samples (columns in Figure 8). Rather than clustering raw expression data, we determined which of the kinase isoforms was significantly upregulated or downregulated in each of the 367 metastatic samples (see Methods) relative to all primary tumor samples. This allowed us to address the simpler question of which isoforms are altered in which samples. To identify correlated patterns of upregulation or downregulation we also clustered the isoforms (rows in Figure 8).

Of the 3,042 protein coding kinase isoforms, 235 had significant altered expression in > 13% of metastatic tumor samples. Clustering this reduced dataset with the k-means elbow method identified 4 sample clusters and 4 isoform groups (Figure S4). However, we found that using k-means with 5 isoform groups strengthened certain BP enrichment patterns. These 5x4 clusters are depicted in Figure 8. For each sample cluster, we tested enrichment for batch ID, region (skin/soft tissue, lymph node, and distant metastasis), and genomic subtype.

Notable enrichments in Cluster A (n = 55 samples) include the tissue location of skin/soft tissue cluster and *BRAF* hotspot mutations. Cluster B (n = 69 samples), was identified as a lymph node cluster with mild enrichment in Triple WT samples. Distant metastases were depleted in both A and B clusters. Cluster C (n = 60 samples) had no region enrichment but was strongly enriched for *RAS* hotspot mutations (Fisher's exact test, p = 4.4e-4, odds = 2.9). Cluster D (n = 183 samples) stood out as a low expression cluster, which had expression largely similar to the primary tumor samples, with little upregulation of isoforms compared to other groups. Moreover, decreased expression of isoforms (shown in blue) occurred in many samples. This cluster was enriched for distant metastases.

The batch ID enrichment analysis identified batch A18 in Cluster C, suggesting batch effects could have influenced our results. To address this issue, we clustered only the 199 metastatic samples (54% of all such samples) in batch A18 (Figure S5), originally found in groups A-D. We found four clusters comparable to the four described above, and Cluster 3 was still significantly enriched for *RAS* hotspot mutants (p = 0.032, odds = 2.1). Clustering all samples *not* in A18, originally present in groups A-D, also revealed 4 clusters and though genomic

subtype was not available for most of these samples, Cluster C still had the highest enrichment for *RAS* mutants (p = 0.16, odds = 2.3). Thus, the *RAS* group enrichment appears to be independent of the batch. Cluster D in our main heatmap was enriched for batch A37, a smaller batch (n = 41 samples), considerably smaller than the cluster it was in.

We also compared the level of 3' bias and sample impurity in each cluster and found that Cluster B had low purity (median 42%) compared to the other three (median of 72%, 80%, and 79% respectively). Median 3' bias did not differ noticeably, although Cluster C had a lowest mean bias (0.517, QoRTs score), indicating higher quality samples. Taken together, these data suggest that metastatic samples have characteristic subgroups related to tumor location and genomic subtype, where isoform expression patterns may help to identify the most similar samples to test as treatment subgroups.

Isoform groups correlate with biological process annotations

We performed a similar analysis on the five isoform groups (i.e., rows), looking for kinase phylogenetic group and BP enrichments compared to the total human kinome. Group 1 was enriched for genes involved in blood vessel morphogenesis (p = 3.6e-6) and related annotations, as well as MAPK regulation. These isoforms are upregulated in Clusters A and B. Since these genes are active in the skin/soft tissue sample cluster and regional lymph nodes, the isoforms may be important in the first transition from primary tumor to metastatic melanoma. This group is also enriched for RTKs.

Group 2 was strongly enriched for nRTKs and contained genes in the category of immune response, for example, used by leukocytes such as T-cells and B-cells (p = 5.2e-11). These isoforms are consistently upregulated in Cluster B. Due to their highly correlated expression and the low estimated purity of the Cluster B samples (median 42%), this group likely arises from immune cells infiltrating the tumor, consistent with previous findings from Akbani et al. [11]. Cluster B is also enriched for samples taken from lymph nodes, a prime location for immune cells to interact with the tumor.

Group 3 was enriched for kinases that regulate cell motility (p = 0.0081). No phylogenetic kinase group enrichments were found, although this group had weak CMGC enrichment compared to the other four groups in Figure 8. These isoforms had the highest expression in Cluster C, containing *RAS* hotspot mutant samples and distant metastases. We note a strong pattern of exclusivity for Group 3 isoforms with the immune infiltrate cluster of Group 2 isoforms, suggesting a novel means of stratifying samples for clinical testing.

Group 4 was enriched for kinases which positively regulate apoptosis (p = 0.010) and cell differentiation (p = 0.0045), and for STE kinases. These isoforms were upregulated in Clusters A and C. This group contains two isoforms of *CDK19*, a gene implicated in cancer proliferation (a third isoform, *CDK19-203*, lacks the seventh exon and decreases in metastatic samples). The function of these isoforms in apoptosis is not explored, on one hand apoptotic processes may occur spontaneously in cancer due to cellular stress and DNA damage [50], on the other hand alternate splicing can modulate pro- and anti-apoptotic functions in the same gene, like *BCLX* [51]. Samples with high levels of immune infiltrate (i.e. Cluster B) appear to have no enrichment of these isoforms, indicating how therapeutics could be specific for one subgroup and be ineffective in another.

Group 5 contained isoforms of genes enriched for regulation of RNA biosynthesis and transcription (p = 0.0059). These isoforms had correlated downregulation in several samples (Clusters B and D), although they are not universally downregulated and in fact increase in some samples. One such gene, *NME1*, is a known suppressor of metastasis [52]. Also in this group are two isoforms of *MAPKAPK3* (-201 and -208), a gene which activates autophagy in response to stress [53] and represses transcription factor E47 [54]. A shorter isoform, -202, is increased in metastatic samples. This isoform lacks the p38 MAPK-binding site, meaning it cannot be activated by p38. This apparent isoform switching was not identified by our DIR analysis because isoform -201 increases in some metastatic samples. *RPS6KA4-201* also significantly decreases, though not the gene's two secondary isoforms -202 and -205. These isoforms lack a nuclear binding site on the 3' end, suggesting it is *RPS6KA4*'s nuclear binding that is selected against. The list of isoforms in Group 5 is given in Table 6, and the full list for each sample cluster and isoform group may be found in Table S6.

Some isoforms had divergent expression patterns depending on cluster. For example, the major isoform of *BRD4*, *BRD4-201*, was found in Group 5, indicating decreased expression in several samples. In contrast, this isoform increased in *RAS*-mutant metastatic samples, as did two shorter isoforms *BRD4-205* (a member of Group 3) and *BRD4-203* (Figure S7). This suggests *BRD4* may be a drug target specific to *RAS*-mutant melanoma; indeed, a recent study found that Vemurafenib-resistant melanoma was susceptible to *BRD4* degradation [55]. Consistent with this observation, DE analysis revealed an 11% increase in *BRD4* expression in metastatic *RAS* mutants, but this increase is not significant ($p_{unadjusted} = 0.452$). Furthermore, we could not confirm *kallisto's* isoform assignments using exon junction alignment, although the reported increase in isoform 205 – a shortened isoform which includes the two bromodomains but not the C-terminal region or NET domain – may suggest some underlying switching effect.

Table 6: Kinase genes with correlated decreased expression in metastatic samples

Identified by isoform clustering (Group 5)[1]	
Gene	**Isoform ID***
MAP2K5	204
MAPK7	204
CAMKK2	216
TNK2	208
ARAF	201
MAP3K6	202
MAP2K7	201
MAPKAPK3	208, 201
ULK1	201
BCKDK	201
CSK	201
BRD4	201
CDK16	220
CDK11B	203
STK11	201
ADCK2	201
RPS6KA4	201
CLK3	201
DAPK3	201
CSNK1G2	201
GTF2F1	201
NRBP1	203
TRIM28	201
MAP2K2	202
NME1	203
MAP2K5	204
MAPK7	204
These isoforms have correlated under-expression in individual samples, but not strong under-expression across all samples, with the exception of *RPS6KA4*, (FC = -1.39, p_{adj} = 0.0011, *DESeq2* with high purity samples) *isoform numbers are from Gencode v.29	

Discussion

Given the rise in melanoma cases across the world, and preliminary success of new therapeutic approaches combing kinase inhibitors and other treatments, we were encouraged to look for differential isoform expression, which has not been intensively studied, and compare it to differential expression identified using conventional approaches (i.e., using the gene locus as a proxy for average expression). We show that both differential expression and altered isoform ratios are prevalent in the human kinome in metastatic melanoma compared to primary tumor melanoma. Furthermore, these changes differ by genomic subtype and tumor location. Affected genes were enriched for several biological processes including immune response, angiogenesis, cell differentiation, chemotaxis, and cell projection organization. Our results provide insight into the regulation of melanoma progression and possible new routes for grouping therapeutic targets.

Different genes were affected by differential expression (DE) and differential isoform ratios (DIR). These genes differed in both phylogenetic groups – e.g. receptor tyrosine kinases in DE vs non-receptor tyrosine kinases in DIR – and biological process enrichments. Thus, isoform analysis may reveal novel information about cancer progression that DE analysis cannot. The drivers behind these splicing events are unknown, but can be multifactorial. For example, mutations in splicing factors can determine outcomes of alternative splicing, but so may somatic mutations or SNPs [56]. Additional determinants derive from epigenetic changes such as aberrant DNA methylation [57] and RNA modifications [58].

Isoform switching may affect protein function

We chose to examine five genes with especially significant isoform switching in greater detail. Metastatic samples showed *SLK* overexpression in our study, something that has been previously observed in other cancer types such as *ErbB2*-driven breast cancer [59]. Knocking down this gene markedly reduces cell migration in *3T3 MEF* cells [60]. Presumably, invasion is the functional benefit acquired by *SLK* overexpression in metastatic melanoma. However, while the short form of *SLK* (*SLK-201*) is overexpressed in metastatic samples, the long form (*SLK-202*) is underexpressed. Overexpression of *SLK* can cause dimerization via the C-terminal coiled-coiled domain; these dimers then activate apoptosis [47]. The short form of *SLK* (*SLK-201)* skips an exon that encodes a coiled-coil region in the C-terminal domain. It is possible the SLK long isoform

dimerizes at lower concentrations and is selected against in metastatic samples. Further experiments are needed to confirm this hypothesis.

MAP3K3 has been identified as an oncogene in various cancers [61-63]. Although we observed no differential expression of the gene (after immune-infiltrate samples were removed), we found that skipping of exon 3 was significant in metastatic samples in the DIR analysis. The functional effect of this skipping is unknown; it precedes, but is not part of, the PB1 protein-protein interaction domain. *MAP3K3* plays important roles in angiogenesis, cell differentiation, and proliferation and may regulate its partners through this structural edit.

In metastatic samples, *MKNK2* was found to switch to a shortened terminal exon which lacks the MAPK binding site. This switching has been previously observed in glioblastoma [32] (compared to normal samples), where the short terminal exon showed pro-oncogenic activity. The authors demonstrated that use of splice switching oligos in glioblastoma reduced the presence of the short terminal isoform and inhibited the oncogenic properties, suggesting this approach might also work in melanoma.

Another notable exon skipping event in *FES*, a non-receptor tyrosine kinase, was found in metastatic tumors at the 11th exon, which encodes the SH2 domain, and is necessary to activate the kinase domain [48]. *FES* has been previously identified as a tumor suppressor in melanoma [64], but we did not observe significant DE in our analysis. We predict that the skipping of the SH2 domain effectively turns off the kinase activity without decreasing the overall gene count. This effect would be consistent with reports of wild type *FES* acting as a tumor suppressor [65]. DE analysis alone would have missed this important effect. Notably, *FES* has several known inhibitors that target the SH2 domain and thus would not be effective against the short isoform [65].

Finally, in some metastatic samples an isoform of *LIMK1* with an abrogated N-terminal LIM domain was expressed. Deleting both LIM domains was previously found to increase kinase activity 3-7 fold [66], suggesting this isoform has greater kinase activity. Targeting *LIMK1* with small molecular inhibitors has been shown to reduce migration and invasion of malignant melanoma [49], suggesting increased activity would promote malignancy. *LIMK1* also did not have significant DE in our dataset.

Expression pattern of *RAS* hotspot mutants

Our various analyses discovered that *RAS* mutants have an expression level pattern distinct from the other three genomic subtypes. *BRAF* and *MEK* inhibitors, while useful for treating *BRAF*-mutant melanoma, have no or limited effectiveness against *RAS*-mutants [67]. *BRAF*-mutants that gain resistance to *BRAF* inhibitors often acquire a secondary *NRAS* mutation [68], meaning any effective *RAS* mutant treatment may also aid in treating drug-resistant *BRAF*-mutants. We found that DE of kinases in *RAS* mutants is concentrated in CMGC kinases (as opposed to receptors as in the other three subtypes) and that DIR is concentrated in kinases involved in angiogenesis. Thus anti-angiogenics [69] are also possible treatments. *Kallisto* counts also identified the bromodomains of *BRD4* as a possible target. A recent study found that Vemurafenib-resistant melanoma was susceptible to *BRD4* degradation [55], and bromodomain inhibitors such as OTX015 and BI-2536 have already had some success treating carcinomas [70]. However, this result was not supported by the *HTSeq* gene counts or exon junction analysis.

Another genomic subtype, Triple Wildtype melanoma, had DE concentrated in Ca^{2+}/calmodulin-dependent protein kinase (in addition to RTKs). These may also serve as a new set of drug targets for this rarer subtype.

Further biological implications

One interesting result from the clustering analysis was the apparent mutual exclusivity of some kinase clusters in metastatic tumors. In particular, the isoform group involved in cell motility (i.e., Group 3) only had high expression in samples lacking in immune response markers (i.e., Cluster C). This sample set was enriched for *RAS* mutants, suggesting not only a unique expression pattern but that they may be better able to evade the immune system. The presence of distant metastases in Clusters C and D, which have little immune marker expression, suggest they may have evaded the initial immune attack in the lymph nodes. Furthermore, cell differentiation and apoptotic markers were highly expressed in regional soft tissue tumors (i.e., Cluster A) and *RAS* mutants (Cluster C), but not lymph node tumors (i.e., Cluster B). Perhaps cells that reach the lymph node have already evaded apoptosis. *BRAFV600E* mutations are present in Clusters A and B, indicating that in addition to the driver mutation, location of the tumor and isoform content is relevant to discern tumor biology and treatment choices. We conclude that the heterogeneity of sample types

displayed in Clusters A-D suggests that the complexity of tumor biology is greater than indicated by driver mutations alone, and that the isoforms in our heatmap may be useful for screening metastatic samples.

Limitations

Our results have several limitations that may impact the interpretations of our data. For example, isoform count estimation is a computational approach to predict isoform expression levels from short read data. Other short read algorithms – using direct alignment approaches such as *RSEM*, *Sailfish*, or *Cufflinks* – may produce different count estimates than *kallisto*. A comparative study of these algorithms found that accuracy decreases as the number of gene isoforms increases [71]. However, we use an additional *kallisto*-independent method to verify *kallisto* findings.

Also, because *kallisto* requires isoform transcript sequences, our method does not account for novel isoforms. 3rd-gen RNA sequencing technologies such as PacBio [22] and Oxford Nanopore [23] are anticipated to provide more accurate knowledge of isoform sequences, both annotated and novel. Currently, novel isoforms may be indirectly inferred by aligning reads to the nucleotide sequences of individual exons [72].

Sample artefacts could also affect our results. As demonstrated in the manuscript, computational estimates of isoform counts are highly impacted by sample impurity or 3' fragment bias. Although our quality-controlled sample set had little difference in purity between primary tumor and metastatic samples (two-sided Wilcoxon, p = 0.88), primary tumor samples still exhibited increased 3' bias compared to metastatic tumors (p = 4.4e-4). Estimates of fragment bias could be incorporated into the existing tools to reduce artefactual results.

With one exception, the TCGA samples are not matched, i.e. the primary tumor and metastatic samples do not come from the same patient. However, our sample size is large enough to make meaningful comparisons between sample categories.

Despite the limitations, the statistical significance of our results along with a variety of quality control measures (e.g. removing problematic samples, resolving DIR events with exon junction reads) provides confidence in our findings.

Summary

We have compared differential gene expression and differential isoform expression to address the hidden effect of differential splicing of kinases in metastatic melanoma. We demonstrate novel, plausible stratification of tumors for clinical testing, for example, immune infiltrate vs. cell migration groups. These groups are consistent with presence of a specific driver mutation, but a mixture of samples could be found in each group (i.e., *BRAF* vs. *RAS*). Additionally, we identified a group of isoforms with significant downregulation in metastatic tumors. These include a known suppressor of metastasis (*NME1*), and may provide a rich source of discovery for additional suppressors. Although we focused here on the kinome in metastatic melanoma, in future work we can expand the analysis to the entire human genome, as well as other cancer types having a rich source of expression data. Further experimental work can confirm links between isoform switching and angiogenesis or other cell processes.

References

1. Schadendorf, D., et al., *Melanoma.* Lancet, 2018. **392**(10151): p. 971-984.
2. Jemal, A., et al., *Cancer statistics, 2010.* CA Cancer J Clin, 2010. **60**(5): p. 277-300.
3. Frinton, E., et al., *Metastatic melanoma: prognostic factors and survival in patients with brain metastases.* J Neurooncol, 2017. **135**(3): p. 507-512.
4. Du, Z. and C.M. Lovly, *Mechanisms of receptor tyrosine kinase activation in cancer.* Mol Cancer, 2018. **17**(1): p. 58.
5. Tsintarakis, A. and A. Zafiropoulos, *Oncogenic Kinases in Cancer*, in *eLS*. 2017, John Wiley & Sons, Ltd: Chichester. p. 1-7.
6. Casimiro, M.C., et al., *Cyclins and cell cycle control in cancer and disease.* Genes Cancer, 2012. **3**(11-12): p. 649-57.
7. Dhillon, A.S., et al., *MAP kinase signalling pathways in cancer.* Oncogene, 2007. **26**(22): p. 3279-90.
8. Lopez-Otin, C. and T. Hunter, *The regulatory crosstalk between kinases and proteases in cancer.* Nat Rev Cancer, 2010. **10**(4): p. 278-92.
9. Fleuren, E.D., et al., *The kinome 'at large' in cancer.* Nat Rev Cancer, 2016. **16**(2): p. 83-98.
10. Bhullar, K.S., et al., *Kinase-targeted cancer therapies: progress, challenges and future directions.* Mol Cancer, 2018. **17**(1): p. 48.
11. The Cancer Genome Atlas Network, *Genomic Classification of Cutaneous Melanoma.* Cell, 2015. **161**(7): p. 1681-96.

12. Johnson, D.B., et al., *Acquired BRAF inhibitor resistance: A multicenter meta-analysis of the spectrum and frequencies, clinical behaviour, and phenotypic associations of resistance mechanisms.* Eur J Cancer, 2015. **51**(18): p. 2792-9.

13. Queirolo, P. and F. Spagnolo, *Binimetinib for the treatment of NRAS-mutant melanoma.* Expert Rev Anticancer Ther, 2017. **17**(11): p. 985-990.

14. Wolchok, J.D., et al., *Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma.* N Engl J Med, 2017. **377**(14): p. 1345-1356.

15. Robert, C., et al., *Pembrolizumab versus Ipilimumab in Advanced Melanoma.* N Engl J Med, 2015. **372**(26): p. 2521-32.

16. Robert, C., et al., *Nivolumab in previously untreated melanoma without BRAF mutation.* N Engl J Med, 2015. **372**(4): p. 320-30.

17. Fisher, R. and J. Larkin, *Vemurafenib: a new treatment for BRAF-V600 mutated advanced melanoma.* Cancer Manag Res, 2012. **4**: p. 243-52.

18. Druillennec, S., C. Dorard, and A. Eychene, *Alternative Splicing in Oncogenic Kinases: From Physiological Functions to Cancer.* Journal of Nucleic Acids, 2012.

19. Steijger, T., et al., *Assessment of transcript reconstruction methods for RNA-seq.* Nat Methods, 2013. **10**(12): p. 1177-84.

20. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.

21. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(5): p. 525-7.

22. Rhoads, A. and K.F. Au, *PacBio Sequencing and Its Applications.* Genomics Proteomics Bioinformatics, 2015. **13**(5): p. 278-89.

23. Lu, H., F. Giordano, and Z. Ning, *Oxford Nanopore MinION Sequencing and Genome Assembly.* Genomics Proteomics Bioinformatics, 2016. **14**(5): p. 265-279.

24. Hardwick, S.A., et al., *Getting the Entire Message: Progress in Isoform Sequencing.* Front Genet, 2019. **10**: p. 709.

25. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.

26. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.

27. Hahn, C.N. and H.S. Scott, *Spliceosome mutations in hematopoietic malignancies.* Nat Genet, 2011. **44**(1): p. 9-10.

28. Harbour, J.W., et al., *Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma.* Nat Genet, 2013. **45**(2): p. 133-5.

29. Hsu, T.Y., et al., *The spliceosome is a therapeutic vulnerability in MYC-driven cancer.* Nature, 2015. **525**(7569): p. 384-8.

30. Seiler, M., et al., *Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types.* Cell Rep, 2018. **23**(1): p. 282-296 e4.

31. Sveen, A., et al., *Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes.* Oncogene, 2016. **35**(19): p. 2413-27.

32.   Mogilevsky, M., et al., *Modulation of MKNK2 alternative splicing by splice-switching oligonucleotides as a novel approach for glioblastoma treatment.* Nucleic Acids Res, 2018. **46**(21): p. 11396-11404.

33.   Watermann, D.O., et al., *Splicing factor Tra2-beta1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene.* Cancer Res, 2006. **66**(9): p. 4774-80.

34.   DiFeo, A., J.A. Martignetti, and G. Narla, *The role of KLF6 and its splice variants in cancer therapy.* Drug Resist Updat, 2009. **12**(1-2): p. 1-7.

35.   Feng, H., T. Li, and X. Zhang, *Characterization of kinase gene expression and splicing profile in prostate cancer with RNA-Seq data.* BMC Genomics, 2018. **19**(Suppl 6): p. 564.

36.   Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

37.   Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

38.   Hartley, S.W. and J.C. Mullikin, *QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments.* BMC Bioinformatics, 2015. **16**: p. 224.

39.   Aran, D., M. Sirota, and A.J. Butte, *Systematic pan-cancer analysis of tumour purity.* Nat Commun, 2015. **6**: p. 8971.

40.   Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

41.   Hothorn, T., et al., *Implementing a Class of Permutation Tests: The coin Package.* Journal of Statistical Software, 2008. **28**(8): p. 1-23.

42.   Wilson, D.J., *The harmonic mean p-value for combining dependent tests.* Proc Natl Acad Sci U S A, 2019. **116**(4): p. 1195-1200.

43.   Carpenter, G. and H.J. Liao, *Receptor tyrosine kinases in the nucleus.* Cold Spring Harb Perspect Biol, 2013. **5**(10): p. a008979.

44.   Lisabeth, E.M., G. Falivelli, and E.B. Pasquale, *Eph receptor signaling and ephrins.* Cold Spring Harb Perspect Biol, 2013. **5**(9).

45.   Love, M.I., J.B. Hogenesch, and R.A. Irizarry, *Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation.* Nat Biotechnol, 2016. **34**(12): p. 1287-1291.

46.   Zhang, T., et al., *Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes.* Genome Res, 2018. **28**(11): p. 1621-1635.

47.   Delarosa, S., et al., *Activity of the Ste20-like kinase, SLK, is enhanced by homodimerization.* Am J Physiol Renal Physiol, 2011. **301**(3): p. F554-64.

48.   Filippakopoulos, P., et al., *Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation.* Cell, 2008. **134**(5): p. 793-803.

49.   Sheng, H., et al., *MiR-429-5p attenuates the migration and invasion of malignant melanoma by targeting LIMK1.* Eur Rev Med Pharmacol Sci, 2020. **24**(5): p. 2625-2631.

50.   Holdenrieder, S. and P. Stieber, *Apoptotic markers in cancer.* Clin Biochem, 2004. **37**(7): p. 605-17.

51.   Paronetto, M.P., I. Passacantilli, and C. Sette, *Alternative splicing and cell survival: from tissue homeostasis to disease.* Cell Death Differ, 2016. **23**(12): p. 1919-1929.

52. Steeg, P.S., et al., *Evidence for a novel gene associated with low tumor metastatic potential.* J Natl Cancer Inst, 1988. **80**(3): p. 200-4.

53. Wei, Y., et al., *The stress-responsive kinases MAPKAPK2/MAPKAPK3 activate starvation-induced autophagy through Beclin 1 phosphorylation.* Elife, 2015. **4**.

54. Neufeld, B., et al., *Serine/Threonine kinases 3pK and MAPK-activated protein kinase 2 interact with the basic helix-loop-helix transcription factor E47 and repress its transcriptional activity.* J Biol Chem, 2000. **275**(27): p. 20239-42.

55. Rathod, D., Y. Fu, and K. Patel, *BRD4 PROTAC as a novel therapeutic approach for the treatment of vemurafenib resistant melanoma: Preformulation studies, formulation development and in vitro evaluation.* Eur J Pharm Sci, 2019. **138**: p. 105039.

56. Antonacopoulou, A.G., et al., *The survivin -31 snp in human colorectal cancer correlates with survivin splice variant expression and improved overall survival.* Cell Oncol (Dordr), 2011. **34**(4): p. 381-91.

57. Chen, Y.C. and L. Elnitski, *Aberrant DNA methylation defines isoform usage in cancer, with functional implications.* PLoS Comput Biol, 2019. **15**(7): p. e1007095.

58. Zhou, K.I., et al., *Regulation of Co-transcriptional Pre-mRNA Splicing by m(6)A through the Low-Complexity Protein hnRNPG.* Mol Cell, 2019. **76**(1): p. 70-81 e9.

59. Roovers, K., et al., *The Ste20-like kinase SLK is required for ErbB2-driven breast cancer cell motility.* Oncogene, 2009. **28**(31): p. 2839-48.

60. Wagner, S., et al., *FAK/src-family dependent activation of the Ste20-like kinase SLK is required for microtubule-dependent focal adhesion turnover and cell migration.* PLoS One, 2008. **3**(4): p. e1868.

61. Fan, Y., et al., *Amplification and over-expression of MAP3K3 gene in human breast cancer promotes formation and survival of breast cancer cells.* J Pathol, 2014. **232**(1): p. 75-86.

62. Hasan, R., et al., *Mitogen activated protein kinase kinase kinase 3 (MAP3K3/MEKK3) overexpression is an early event in esophageal tumorigenesis and is a predictor of poor disease prognosis.* BMC Cancer, 2014. **14**: p. 2.

63. Jia, W., et al., *MAP3K3 overexpression is associated with poor survival in ovarian carcinoma.* Hum Pathol, 2016. **50**: p. 162-9.

64. Olvedy, M., et al., *Comparative oncogenomics identifies tyrosine kinase FES as a tumor suppressor in melanoma.* J Clin Invest, 2017. **127**(6): p. 2310-2325.

65. Hellwig, S., et al., *Small-molecule inhibitors of the c-Fes protein-tyrosine kinase.* Chem Biol, 2012. **19**(4): p. 529-40.

66. Nagata, K., et al., *The N-terminal LIM domain negatively regulates the kinase activity of LIM-kinase 1.* Biochem J, 1999. **343 Pt 1**: p. 99-105.

67. Boespflug, A., et al., *Treatment of NRAS-mutated advanced or metastatic melanoma: rationale, current trials and evidence to date.* Ther Adv Med Oncol, 2017. **9**(7): p. 481-492.

68. Trunzer, K., et al., *Pharmacodynamic effects and mechanisms of resistance to vemurafenib in patients with metastatic melanoma.* J Clin Oncol, 2013. **31**(14): p. 1767-74.

69. Felcht, M. and M. Thomas, *Angiogenesis in malignant melanoma.* J Dtsch Dermatol Ges, 2015. **13**(2): p. 125-36.

70.   Stathis, A., et al., *Clinical Response of Carcinomas Harboring the BRD4-NUT Oncoprotein to the Targeted Bromodomain Inhibitor OTX015/MK-8628.* Cancer Discov, 2016. **6**(5): p. 492-500.

71.   Zhang, C., et al., *Evaluation and comparison of computational tools for RNA-seq isoform quantification.* BMC Genomics, 2017. **18**(1): p. 583.

72.   Roberts, A., et al., *Identification of novel transcripts in annotated genomes using RNA-Seq.* Bioinformatics, 2011. **27**(17): p. 2325-9.