Cell-specific isoform switching in metastatic melanoma

Trevor Manz

May 2020

1 Background

Understanding the functional implications of different gene transcripts is an active area of research. It is believed that gene isoforms play an important role in many biological processes, and that the diversity of transcripts yielded from individual genes is a major contributor of the increased complexity of higher vertebrates [9, 13]. In humans, it is estimated that a gene on average has 6.3 isoforms, of which a subset (3.9) are protein-coding [6]. Gene isoform usage is often tissue-specific and differential usage have been implicated in development and many diseases, especially cancer [2, 8, 9, 10, 14].

Although isoform switching is often hypothesized as an important feature in cancers, historically few studies that characterize differential gene expression do so with isoform resolution [13]. It is suggested that analyses of RNA-seq data at the isoform level are less frequent due to limitations in experimental technologies combined with dedicated downstream computational tools that are designed for counts at the gene level [7, 1]. However, methods have emerged and proved useful in identifying isoform switching, showcasing expression shifts between cell types that were previously masked by standard analyses [7].

A study profiling approx. 5000 cancer patients from TGCA [3], covering 12 solid cancer types, identified common isoform switches with potential functional consequences from bulk RNA-seq data [13] (Figure 1). In particular, several pan-cancer isoform switches were identified, most frequently loss of DNA sequence encoding protein domains, that showed efficacy as biomarkers for predicting patient survival independent of cancer type. Although the methods described offer a mechanism for identifying isoform switches, they fail to provide the resolution at the cellular level to fully characterize differential transcript usage or estimate the potential impact of switches.

The identification of broad, pan-cancer classes of isoform groups (gain or loss of protein domains or signal peptides, loss of protein coding sequence, and changes in propensity for nonsense-mediated decay) may suggest a more general biological mechanisms which are worth characterizing functionally. For example, a previously uncharacterized isoform switch in *AKT1* was identified as a pan-cancer isoform by [13], which is of interest because *AKT1* has been described previously in cancer signaling in melanoma [5].

In contrast, narrower studies have described the specific role of certain isoforms in cancer, but fail to provide insight towards more general mechanisms or place the characterization in the context of complex cellular tissues [4, 15]. One of such cases is an isoform switch for *ALK* which has been previously described to be conversed across melanoma patients, producing a truncated protein and exciting proliferation *in vitro* [15] (Figure 2). These broad spectrum and specific characterizations of differential transcript usage

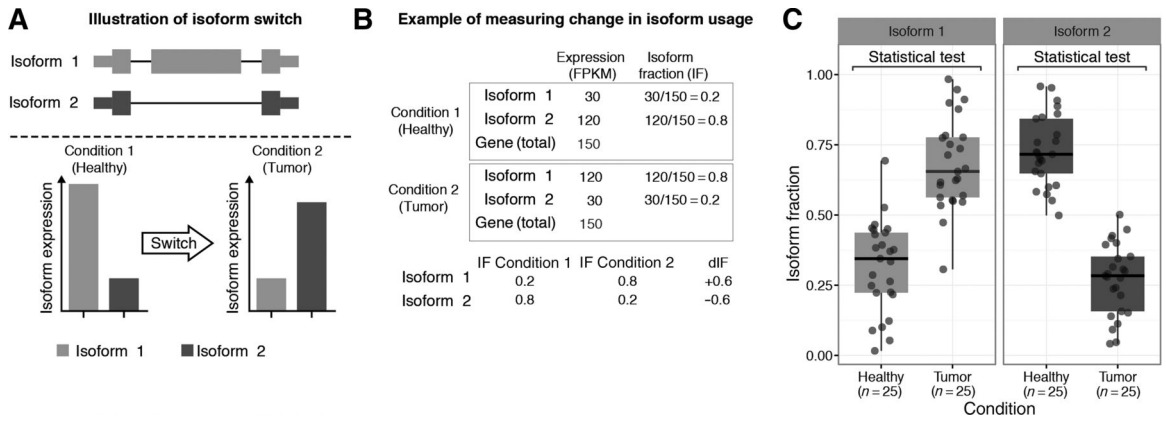

Figure 1: Illustration of isoform switches in cancer [13]. A.) An isoform switch. Under different conditions (healthy vs. tumor), isoforms are differentially expressed. B.) Example computation of expression quantification. When expression is quantified at the gene-level, isoforms that are unique to each condition may go undetected. C.) Example statistical testing of isoform fractions between samples.

in melanoma provide an interesting use case to lay out a spatial mapping of genes in complex tissues.

2 Hypothesis of the present study

I hypothesize that genes exhibit cell type isoform specificity in metastatic melanoma. I propose to use combined spatial and high-throughput single-cell RNA-seq to create a transcriptional map of these complex metastatic tissues at isoform resolution. The goal of this analysis is to 1) identify isoforms which distinguish a spectrum of cancerous cells, and 2) understand the spatial relationship of these cells in their heterogenous context along with functional implications. I hope to identify key differentially expressed isoforms which may be masked at the gene level, providing new insight into the complexity of the transcriptional landscape in metastatic melanoma. The analysis will have two primary stages. The first is to build a spatially resolved transcriptional landscape via combined RNA-seq and identify differentially expressed isoforms. The second is probe the functional significance of cell type isoform specificity derived in phase one by using recently developed experimental methods for isoform screens [11].

3 Specific aims

3.1 Data generation via combined spatial and high-throughput single-cell RNA-seq

Single-cell RNA sequencing (scRNA-seq) technologies have become popular tool for systematically identifying cell populations, providing the ability to quantify cell to cell heterogeneity. The growth of interest in this area has led to the development of many sequencing technologies which each offer unique tradeoffs, primarily in cost, throughput, and resolution. While high-throughput assays (Drop-seq, 10x Genomics, inDrops) can identify rare cell types often missed with lower coverage, they produce 3'-end reads

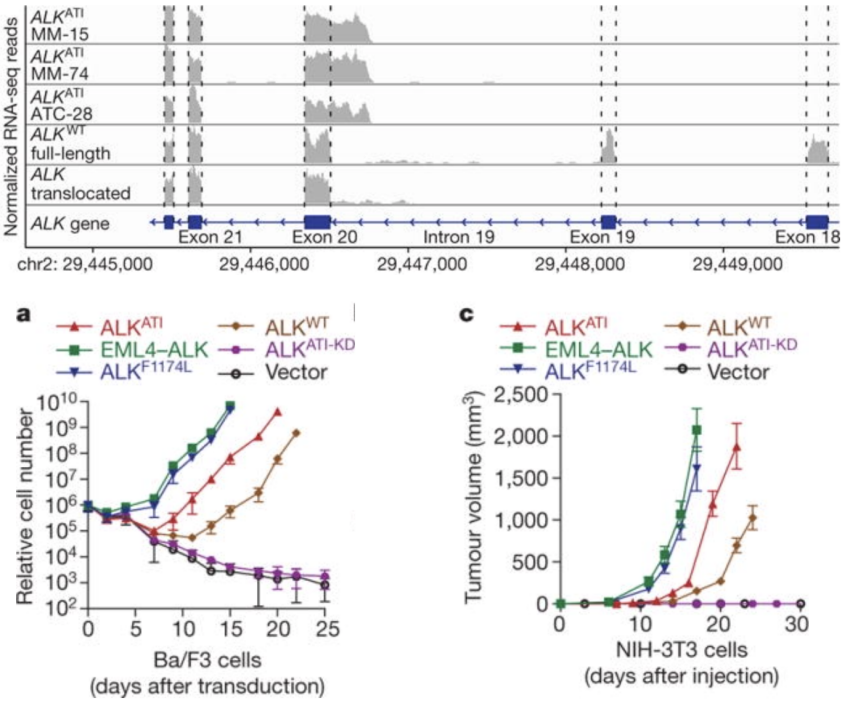

Figure 2: Novel *ALK* isoform promotes proliferation *in vitro* and tumorigenesis in vivo in mice [15]. *De novo* alternative transcription start site in intron 19 leads to ALK^{ATI} isoform that promotes cell proliferation *in vitro* (a) and tumorigenesis *in vivo* (c).

which are gathered in pre-processing to generate counts at the gene-level. In contrast, SMART-Seq scRNA-seq is a full-length scRNA-seq method, which allows for isoform quantification, but is much lower throughput. Spatial method like MERFISH offer an additional dimension to single cell analysis, but are often coupled with droplet-based methods because they are biased towards the probes used and much lower throughput. As a result, there is much debate over what technology is "best", with high-thoughput methods generally falling in favor over SMART-seq.

A recent study from Booeshaghi et al.[10] suggested an integrative approach to using these technologies which identified isoform cell type specificity. The authors surveyed cells from mouse primary motor cortex (MOp) via scRNA-seq, collecting reads from both 6,160 SMART-Seq cells and 90,031 10x Genomics Chromium cells. While the 10x data were stronger in identifying rare cell types, the SMART-Seq data had higher sensitivity and was capable of identifying isoform-specific markers in dozens of cell types. In addition, the authors refined gene-level analysis via MERFISH spatial scRNA-seq by extrapolating SMART-Seq isoform quantification, systematically identifying differentially expressed genes in the MERFISH data and checking whether these genes showed unique isoform markers identified by SMART-Seq analysis.

The above offers a powerful method for creating a detailed isoform atlas in tissues. I will apply a variant of this method, analyzing reads from approximately 500,000 10x Chromium cells and approximately 50,000 SMART-seq cells. The increase in the number of cells compared to the above is due to the diversity of tumors which are surveyed in my study. Tissue selection is important for this atlas, so I plan to procure human melanoma tumors that span a range of clinically relevant backgrounds as previously described in

[12]. Pairing genotypic information with these tumors will allow for the identification of specific oncogene mutations. Following the rapidly translational workflow outlined in [12], I will process tumor tissues immediately after surgical procurement and generate single-cell suspensions. I will use FAC sorting to isolate immune and nonimmune cells via CD45+/CD45- gate. Then I will prepare the libraries for both 10x and SMART-seq protocols. The aim of the droplet-based approach is to help characterize low abundance transcripts, so I will target a sequencing depth of 120,000 reads per cell.

3.2 Creation of isoform atlas and extrapolating spatial specificity

As described in [15], I will use both genetic and trasncriptional states to distinguish between cell types within melanoma tumors. I will quantify gene/isofrom reads using the kallisto suite [1]. To infer large-scale CNVs, I will average gene expression counts over stretches of 100 genes within chromosomes to identify aneuploidy, and thus classisfy cells as malignant. This classification will leverage the many reads provided by the droplet based method to best determine malignant vs non-malignant. Following standard preprocessing and quality filtering of reads, I will select the top 5000 most variable genes (isoforms) and perform dimensionality reduction via UMAP on the first 50 truncated singular value decomposition (SVG) components. As previously described [15], I would expect to see cells classified as malignant to form distinct clusters for each tumor which suggests distinct transcriptional profiles per tumor. However, I expect the expression profiles of non-malignant cells to cluster by cell type and not by tumor.

Similarly with the SMART-seq data, I would assign cell types and malignant vs non-malignant classification. However, next I would perform differential expression between the two classes to identify cell specific isoforms. For each gene/isoform, I will perform a t-test between the malignant and non-malignant classes, as well as between different tumor types, using a significance level of 0.05 and performing Bonferroni correction on all p-value. To identify isoform enrichment that is typically masked, I will also perform this analysis at the gene level by aggregating isoform counts and performing the same tests. I expect that many isoform will not be specific, but I hope to verify both previously described isoform switches, like in *ALK* and *AKT1*, as well as unique switches to malignant melanoma.

Next I will perform MERFISH spatial single-cell RNA-seq on paired tissues from the high-throughput methods. Due to the lack and length of probes required for MERFISH, it is not ideal for profiling isoforms directly; however, [10] showed that meaningful spatial analysis can be made by extrapolating the isoform-level results from SMART-seq to refine the spatial map (Figure 3). Briefly, if specific isoform expression is observed in the SMART-seq data for a particular tissue (for example a single isoform or small set of isoforms), then spatial quantification of the corresponding gene by MERFISH can be refined to spatial quantification of that isoform. In addition, by performing MERFISH following SMART-seq isoform analysis, probes can be designed to detect specific isofroms directly. Although the design of probes in theory just requires a distinct sequence, the technology is not capable of assaying most isoforms [10], and therefore the extrapolation procedure is relevant.

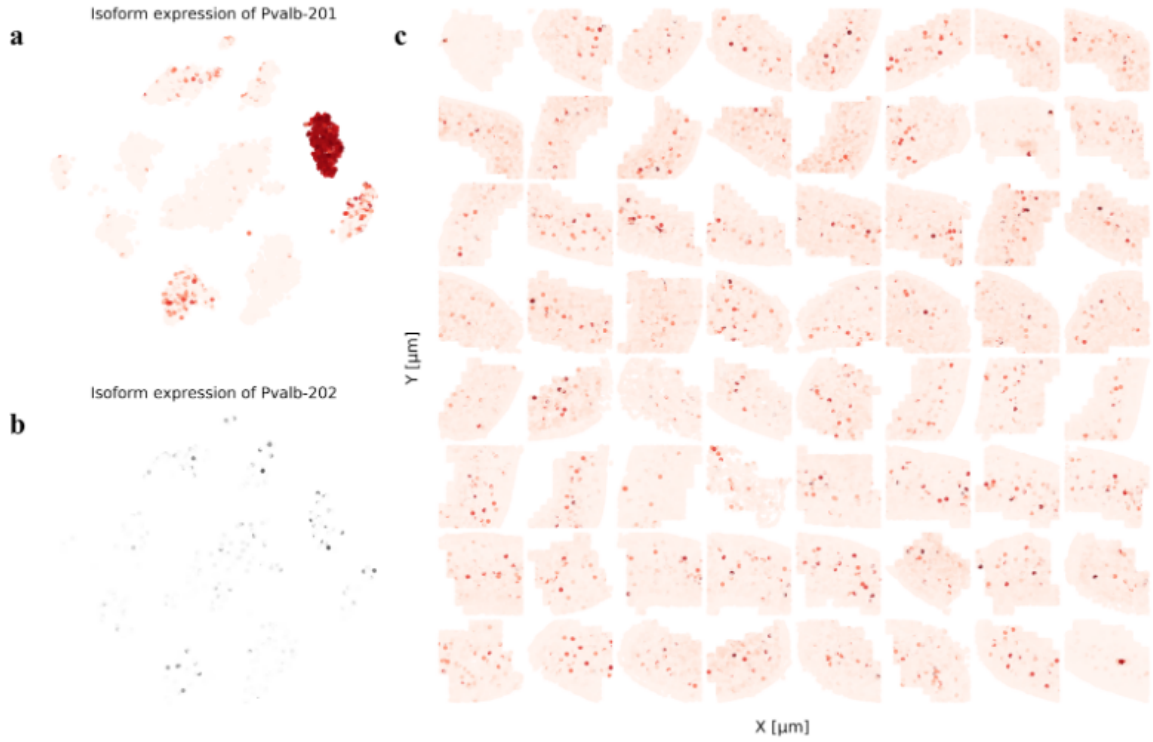

Figure 3: Spatial refinement and extrapolation of isoform expression [10]. If SMART-Seq quantifications reveal that an isoform is uniquely expressed (a, b), then spatial quantifications at the gene-level (via MERFISH) can be refined.

3.3 Functional characterization of cell type isoform specificity

Elucidating the biological impact of different isoform usage is very difficult and yet essential for understanding the significance of their observed diversity in humans. A recent method for screening isoforms, pgFARM (paired guide RNAs for alternative exon removal) [11], is a promising approach for probing this functional significance since it can manipulate isoforms independent of gene inactivation. Briefly, guide RNAs which are designed to unique flank an exon are delivered simultaneously into cells, causing the exon to be excised and the effective skipping of the exon. I propose screening the previously described isoform of *ALK*, ALK^{ATI} [15], as well as novel isoforms identified in Aim 2 in an *in vivo* tumorigenicity assay in mice.

Briefly, using pgFARM I will recapitulate the results from [15] which showed ALK^{ATI} promotes tumorigenesis in mice. I will use the mouse melanocyte (Melan-a) cells line described in [15], but I will deliver deliver pgRNA with CRISPR-Cas9 to introduce the alternative transcription initiation site distinctive of ALK^{ATI}. I will screen the cells via PCR, and then subcutaneously and bilaterally inject the cells into 6-8 week old mice (n=8). Tumor sizes will be measured as previously described with callipers every 2 to 7 days for a period of up to 100 days. An empty vector will be used as a negative control and the WT will be used as a positive control. I expect the ALK^{ATI} to promote tumorgenesis. This will serve as a proof of concept for using pgFARM to screen isoforms specific to malignant melanoma. I will repeat the screen for other malignant cell-specific isoforms identified via the differential analysis in Aim 2 and quantify the functional effect on metastasis. I will use Analysis of variance (ANOVA) to test the significance of the

particular isoform.

References

[1] Nicolas L Bray, Harold Pimentel, Páll Melsted, and Lior Pachter. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, 34(5):525–527, May 2016.

[2] Jean-Philippe Brosseau. Splicing isoform-specific functional genomic in cancer cells. *Applied Cancer Research*, 38(1):18, December 2018.

[3] Cancer Genome Atlas Research Network, John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas Pan-Cancer analysis project. *Nat. Genet.*, 45(10):1113–1120, October 2013.

[4] Giulia Cesi, Demetra Philippidou, Ines Kozar, Yeoun Jin Kim, Francois Bernardin, Guillaume Van Niel, Anke Wienecke-Baldacchino, Paul Felten, Elisabeth Letellier, Sonja Dengler, Dorothee Nashan, Claude Haan, and Stephanie Kreis. A new ALK isoform transported by extracellular vesicles confers drug resistance to melanoma cells. *Mol. Cancer*, 17(1):145, October 2018.

[5] Joseph H Cho, James P Robinson, Rowan A Arave, William J Burnett, David A Kircher, Guo Chen, Michael A Davies, Allie H Grossmann, Matthew W VanBrocklin, Martin McMahon, and Sheri L Holmen. AKT1 activation promotes development of melanoma metastases. *Cell Rep.*, 13(5):898–905, November 2015.

[6] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, September 2012.

[7] Vasilis Ntranos, Lynn Yi, Páll Melsted, and Lior Pachter. A discriminative learning approach to differential expression analysis for single-cell RNA-seq. *Nat. Methods*, 16(2):163–166, February 2019.

[8] Jacqueline Olender and Norman H Lee. Role of alternative splicing in prostate cancer aggressiveness and drug resistance in african americans. *Adv. Exp. Med. Biol.*, 1164:119–139, 2019.

[9] Endre Sebestyén, Michał Zawisza, and Eduardo Eyras. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res.*, 43(3):1345–1356, February 2015.

[10] A Sina Booeshaghi, Zizhen Yao, Cindy van Velthoven, Kimberly Smith, Bosiljka Tasic, Hongkui Zeng, and Lior Pachter. Isoform cell type specificity in the mouse primary motor cortex. March 2020.

[11] James D Thomas, Jacob T Polaski, Qing Feng, Emma J De Neef, Emma R Hoppe, Maria V McSharry, Joseph Pangallo, Austin M Gabel, Andrea E Belleville, Jacqueline Watson, Naomi T Nkinsi, Alice H Berger, and Robert K Bradley. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nat. Genet.*, 52(1):84–94, January 2020.

[12] Itay Tirosh, Benjamin Izar, Sanjay M Prakadan, Marc H Wadsworth, 2nd, Daniel Treacy, John J Trombetta, Asaf Rotem, Christopher Rodman, Christine Lian, George Murphy, Mohammad Fallahi-Sichani, Ken Dutton-Regester, Jia-Ren Lin, Ofir Cohen, Parin Shah, Diana Lu, Alex S Genshaft, Travis K Hughes, Carly G K Ziegler, Samuel W Kazer, Aleth Gaillard, Kellie E Kolb, Alexandra-Chloé Villani, Cory M Johannessen, Aleksandr Y Andreev, Eliezer M Van Allen, Monica Bertagnolli, Peter K Sorger, Ryan J Sullivan, Keith T Flaherty, Dennie T Frederick, Judit Jané-Valbuena, Charles H Yoon, Orit Rozenblatt-Rosen, Alex K Shalek, Aviv Regev, and Levi A Garraway. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, 352(6282):189–196, April 2016.

[13] Kristoffer Vitting-Seerup and Albin Sandelin. The landscape of isoform switches in human cancers. *Mol. Cancer Res.*, 15(9):1206–1220, September 2017.

[14] Sebastien M Weyn-Vanhentenryck, Huijuan Feng, Dmytro Ustianenko, Rachel Duffié, Qinghong Yan, Martin Jacko, Jose C Martinez, Marianne Goodwin, Xuegong Zhang, Ulrich Hengst, Stavros Lomvardas, Maurice S Swanson, and Chaolin Zhang. Precise temporal regulation of alternative splicing during neural development. *Nat. Commun.*, 9(1):2189, June 2018.

[15] Thomas Wiesner, William Lee, Anna C Obenauf, Leili Ran, Rajmohan Murali, Qi Fan Zhang, Elissa W P Wong, Wenhuo Hu, Sasinya N Scott, Ronak H Shah, Iñigo Landa, Julia Button, Nathalie Lailler, Andrea Sboner, Dong Gao, Devan A Murphy, Zhen Cao, Shipra Shukla, Travis J Hollmann, Lu Wang, Laetitia Borsu, Taha Merghoub, Gary K Schwartz, Michael A Postow, Charlotte E Ariyan, James A Fagin, Deyou Zheng, Marc Ladanyi, Klaus J Busam, Michael F Berger, Yu Chen, and Ping Chi. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature*, 526(7573):453–457, October 2015.