Targeted RNAseq Characterization of Kidney Cancer Subtypes

An thesis completed in partial fulfillment of the
requirements for the degree of
Bachelor of Science with Honors
(Cell and Molecular Biology)
The University of Michigan

Jeremy B. Kaplan

Adjunct Associate Professor Scott A. Tomlins, Advisor

Associate Professor Andrzej Wierzbicki, Co-Sponsor

ABSTRACT

Correct diagnostic classification of kidney tumors is critical to providing accurate prognostic information and enabling potential precision medicine opportunities, however this often requires assessing numerous clinical biomarkers by multiple approaches that are both time- and cost-intensive. Here, we sought to determine whether a multiplex-PCR based RNA sequencing (mxRNAseq) panel targeted to kidney cancer relevant genes could facilitate diagnosis of both common and rare kidney tumors as well as inform on prognosis and therapeutic targets. Through assessment of over 146 clinical formalin fixed paraffin embedded (FFPE) samples using a 122 target gene mxRNAseq panel, we demonstrate that this approach enables robust classification of common kidney tumor subtypes using both supervised and unsupervised techniques. To facilitate precision medicine opportunities, we demonstrate that inclusion of cell-cycle related genes enables the calculation of a previously described 31-gene cell-cycle proliferation progression (CCP) score, while immune related genes enables identification of immune "hot" and "cold" tumors. Taken together, these results support the feasibility and broad potential applicability of targeted mxRNAseq for kidney tumor diagnosis, classification, prognosis and therapeutic prediction.

INTRODUCTION

An emerging management strategy for kidney and other tumors is precision medicine – targeting interventions to specific molecular properties. However, such a strategy requires accurate assays for all relevant biomarkers. In some tumors this is relatively straightforward – management of many colon cancers has been revolutionized by the use of a few driving markers to manage care (e.g. KRAS mutations and microsatellite stability status[1]). However, in kidney tumors, such management is made more difficult by the wide variety of subtypes, and the morphologic overlap that often requires extensive immunohistochemistry (IHC), fluorescence in

situ hybridization (FISH), and/or DNA sequencing for complete subtyping. This heterogeneity also complicates the development of broadly applicable prognostic and predictive assays.

IHC remains the gold standard for measuring protein markers, as it measures the actual expressed protein, verses upstream RNA expression or DNA variants that may not impact the actual phenotype. Correct subclassification of kidney tumors often requires a battery of immunostains[2], which makes accurate diagnosis manually intensive and costly. RNA based assays may simplify kidney cancer subtyping – if it can be demonstrated that RNA expression is a true proxy for protein levels and the loss of spatial expression information and expression patterns does not compromise accuracy. Recently, RNAseq has proven to be a method of choice for high-throughput RNA analysis. Most RNAseq analyses involve counting transcripts mapped to the reference genome, however RNAseq data can also be used to identify mutations in exonic regions if probes are designed to capture such regions.[3]

Most RNAseq uses a transcriptomics approach, measuring all transcribed mRNA, typically by selecting for mRNA poly-A tails, or by using capture probes for exonic regions.[4] This requires a relatively large amount of sequencing capacity to be able to gain a large dynamic range but provides data about a sample's entire transcriptional activity. Unfortunately, such approaches are not compatible with RNA isolated from routine FFPE tissues, as it is degraded to formalin induced crosslinking. Importantly, FFPE material is available for essentially all patients with kidney cancer, after fine needle aspiration, kidney biopsy, partial nephrectomy, nephrectomy, or biopsy of metastatic lesions. Likewise, although capture based RNAseq largely alleviates issues with poly-A selection and reverse transcription, if often requires hundreds of nanograms to microgram quantities of RNA[4], which are not available from routine biopsy or fine needle aspiration samples.[4,5]

Typical clinical analysis will only focus on a known subset of targets validated for analysis – all other protein biomarkers or transcripts are irrelevant and wasteful if the cost is too high. One technology enabling sequencing of only targeted regions is multiplexed PCR (after reverse transcription for RNA assessment). Sequencing libraries are prepared by first amplifying only the targets of interest using specific primers, then sequencing the products, and counting reads mapped to genome as with conventional RNAseq. Such a solution enables sequencing more samples/patients at a lower cost as well as lower input requirements by focusing on targets of interest. In addition, multiplex-PCR based amplicon sequencing can be used with smaller and more degraded nucleic acids[6] than conventional poly-A selection[7], as reverse transcription is primed using random primers and mRNA specificity is achieved using highly sensitive PCR with exon-spanning primers. This enables the use of smaller biopsies for clinical or prospective research use and allows for retrospective studies on time and chemically degraded samples.

In this thesis, I describe how a custom target multiplex RNA sequencing panel (mxRNAseq) developed by the Tomlins Lab can be used to characterize features of kidney cancers. I focus primarily on two features in which high-throughput molecular interrogation provides particular value: Subtyping using multiple expression targets, and assessment of proliferation as a potential prognostic feature via a "cell-cycle progression" gene expression module. Additionally, I discuss whether inclusion of many targets may provide strategies for the better management of Kidney cancers that currently go unclassified by typical means.

We focused on the ability to classify four common types of kidney tumors based off availability in our dataset, the presence of reference data in The Cancer Genome Atlas project, and the frequent use of immunohistochemistry to resolve this common differential diagnosis: Clear Cell Renal Cell Carcinomas (RCC), Chromophobe RCC, Papillary RCC, and Urothelial

Carcinomas. Classification of these types is typically performed using IHC of six targets: TP63, PAX8, CA9, KRT7, KIT, and AMACR. How each of these markers contribute to classification is shown in **Figure 1**. TP63, a key marker of urothelial differentiation[8], and PAX8, a PAX family transcription factor regulating organogenesis during fetal development[5)], are used primarily as markers differentiating Urothelial Carcinomas from RCCs. Within RCCs, Carbonic Anhydrase IX (CA9), a carbonic anhydrase expressed in hypoxic environments serves as a strong marker for Clear Cell RCC.[10,11] KIT, a protooncogenic stem cell factor[12,13], is a strong marker for Chromophobe RCC. Cytokerin KRT7/CK7 is also used to differentiate Chromophobe and Papillary RCCs from Clear Cell RCC.[2]

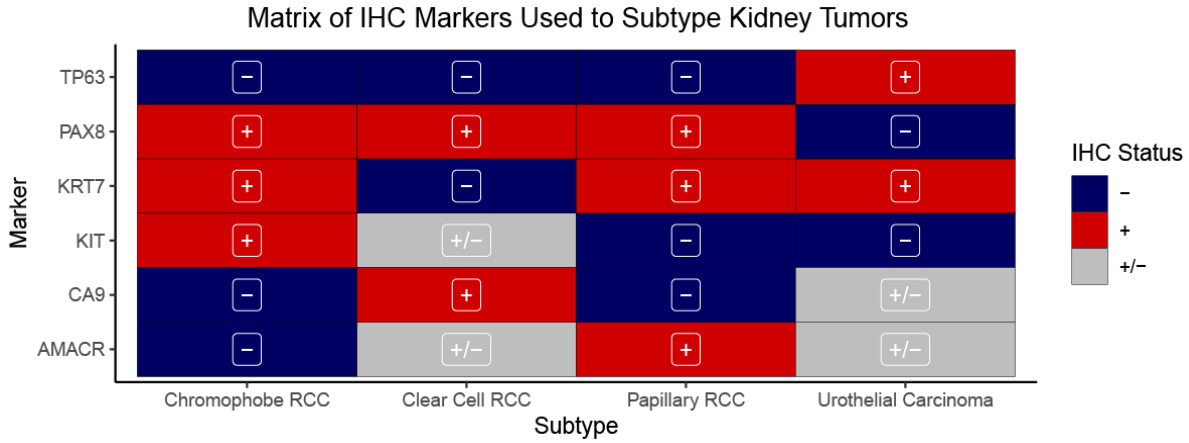

Figure 1: A summary of markers used to classify kidney tumors subtypes discussed. Red represents presence of protein expression, blue represents absence, and grey represents equivocal for a particular subtype.

Beyond classification, high-throughput multi-target assays also present the opportunity for summary signatures that can capture the behavior of many targets belonging to a larger biological process. One example of such a signature is the Cell Cycle Progression (CCP) score, a signature averaging expression of 31 cell cycle related genes, originally validated and commercialized for prostate cancer.[14,15] Our group previously described[3] efforts to port this signature to a mxRNAseq targeted panel for prostate cancer. Given the near universal relevance

of cell cycle/proliferation to molecular oncology (with high proliferation being associated with more aggressive disease[16]), we included these targets on our separate kidney panel to determine the feasibility of characterizing cell cycle/proliferation in kidney tumors.

Here I describe development and validation of a targeted multiplex RNAseq panel capable of assaying both markers relevant for kidney cancer subtyping, and for interrogating larger biological events.

METHODS

Panel Design and Sequencing

Tomlins Lab members identified relevant expression targets from internal and external comprehensive evaluation efforts, and designed a custom Ampliseq mxRNAseq panel using Ion AmpliSeq Designer (Thermo Fisher Scientific, Waltham MA.). The panel consisted of 15 housekeeping genes, and 122 total target genes. Lab members prepared RNA sequencing libraries using the Ion Ampliseq RNA Library kit, and sequenced samples on Ion Torrent Proton or Ion Torrent S5 sequencers as described.[3]

Patient Cohort

Routine FFPE tissues were obtained from the University of Michigan Department of Pathology Tissue Archive with Institutional Review Board (IRB) approval. Diagnostic hematoxylin and eosin–stained slides were reviewed by board-certified anatomic pathologists with areas for microdissection/punching indicated as necessary. For each specimen, 1-3 1mm punches or 3-10 × 10μm FFPE sections were cut from representative blocks, using macrodissection with a scalpel as needed to enrich the tumor content. DNA and RNA were isolated and quantified using the Qiagen AllPrep FFPE DNA/RNA Kit (Qiagen, Valencia, CA) and the Qubit 2.0 fluorometer (Life Technologies), respectively, as described.[3]

Data Processing

Torrent suite version 5.04 was used to process raw sequencing data. End to end read counts for each panel target were obtained from the Torrent Suite Coverage Analysis plugin version 5.0.4.0. All downstream analyses were conducted using R 3.4.1 and Python 3.7.0. All statistical models were implemented using the Scikit-learn package[17] version 0.20.0.

Prior to normalization, sample and target gene amplicon quality was assessed to eliminate technical artifacts. Samples were only included in downstream analyses if they had at least 500,000 mapped reads, with at least 60% being end to end. Target gene amplicons were only included if they had at least 200 reads in at least 2 samples or more than 1000 reads in at least one sample. To ensure robust normalization, performance of all housekeeping genes included on the panel was assessed prior to normalization. Only housekeeping genes with median reads between 50,00-100,000 reads were averaged for normalization.

Raw read counts were normalized as previously described.[3] Briefly, log2 transformed read counts were divided by the geometric mean of the included housekeeping genes. For more intuitive plotting, each sample was then scaled by its median for plotting purposes only.

Data from The Cancer Genome Atlas' Clear Cell RCC, Chromophobe RCC, Papillary RCC, and Urothelial Carcinoma studies[18–21] was also included to increase sample size for model training. RSEM normalized RNAseq data was downloaded from cBioPortal.[22,23] TCGA transcriptome RNAseq data was filtered to only include targets included on the targeted mxRNAseq panel.

To ensure analysis comparability, both the TCGA and mxRNAseq datasets were centered and scaled according to their mean and standard deviation to ensure a unit mean and variance.

Since the TCGA data was used for model training, all data was scaled and centered only according to the properties of the training data. Homogenization was confirmed using principal component analysis and hierarchical clustering (**Figure 2**).

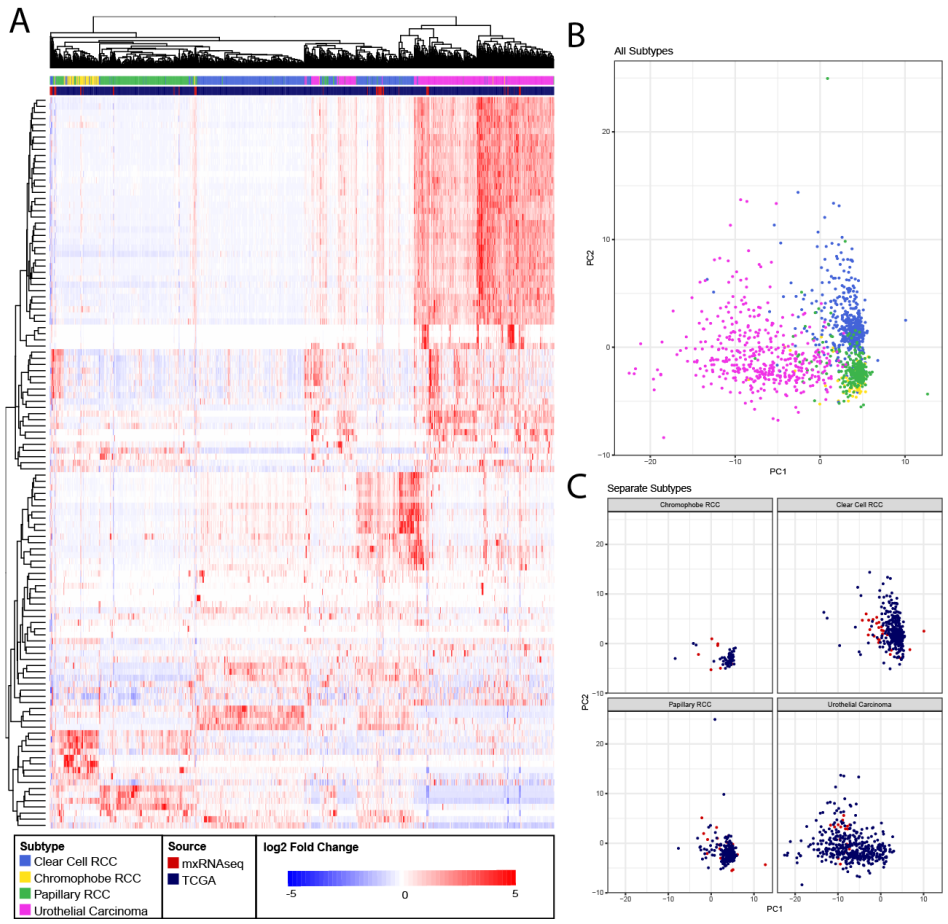

Figure 2: Homogenization of TCGA and mxRNAseq data. Panel A: Merged heatmap of mxRNAseq and TCGA data for available subtypes. B+C: Principal component analysis(PCA) of the merged dataset both combined(B) and separated(C).

Subtype Identifying Models

We used the Scikit-learn[17] package to train statistical models to classify kidney cancer subtypes. To ensure model generalization, we trained the model on 75% of the TCGA cohorts, and validated on the remaining 25%, and the entire mxRNAseq dataset.

We first designed a composite classifier that uses a separate logistic regression classifier for each gene, and classifies each sample based on a composite of each gene's classifications according a pre-provided rule set. Each component model was trained by inferring a training sample's component gene class from the provided rule set. We additionally designed a variant of this composite model including correlated genes in each of the component models.

We also used classification and regression trees[24], as they provide a simple to interpret structure similar to the clinical algorithm. A decision tree was trained using a maximum tree depth of 3 nodes, and a minimum sample split of 5.

RESULTS

Given the foundational importance of classification to kidney tumor management, we first sought to assess the utility of our mxRNAseq panel for diagnosis of common tumor subtypes. In total, we generated mxRNAseq profiles for 146 informative clinical FFPE samples using our panel of 122 target genes. Unsupervised hierarchical clustering of both gene targets and samples is shown in **Figure 3**. Cell-cycle/proliferation targets clustered closely, indicating the ability of the assay to robustly detect a variety of proliferation signals. Targets indicative of immune-oncology response, such as CD8A, GZMA, CTLA4, FOXP3, and others, also showed strong clustering, consistent with the ability of the assay to detect immune infiltration. For example, CD8A – a marker of cytotoxic T lymphocytes clustered directly next to GZMA – a cytotoxic T lymphocyte protease.[25] Samples clustered primarily by cell-cycle/proliferation (unsurprising given the number of related transcripts included in the panel), with highly proliferative Urothelial Carcinoma and Clear Cell RCC separating from most other samples. Within the proliferative sample cluster, Urothelial Carcinoma strongly separated from other subtypes, driven predominantly by urothelial markers PAX8 and TP63.

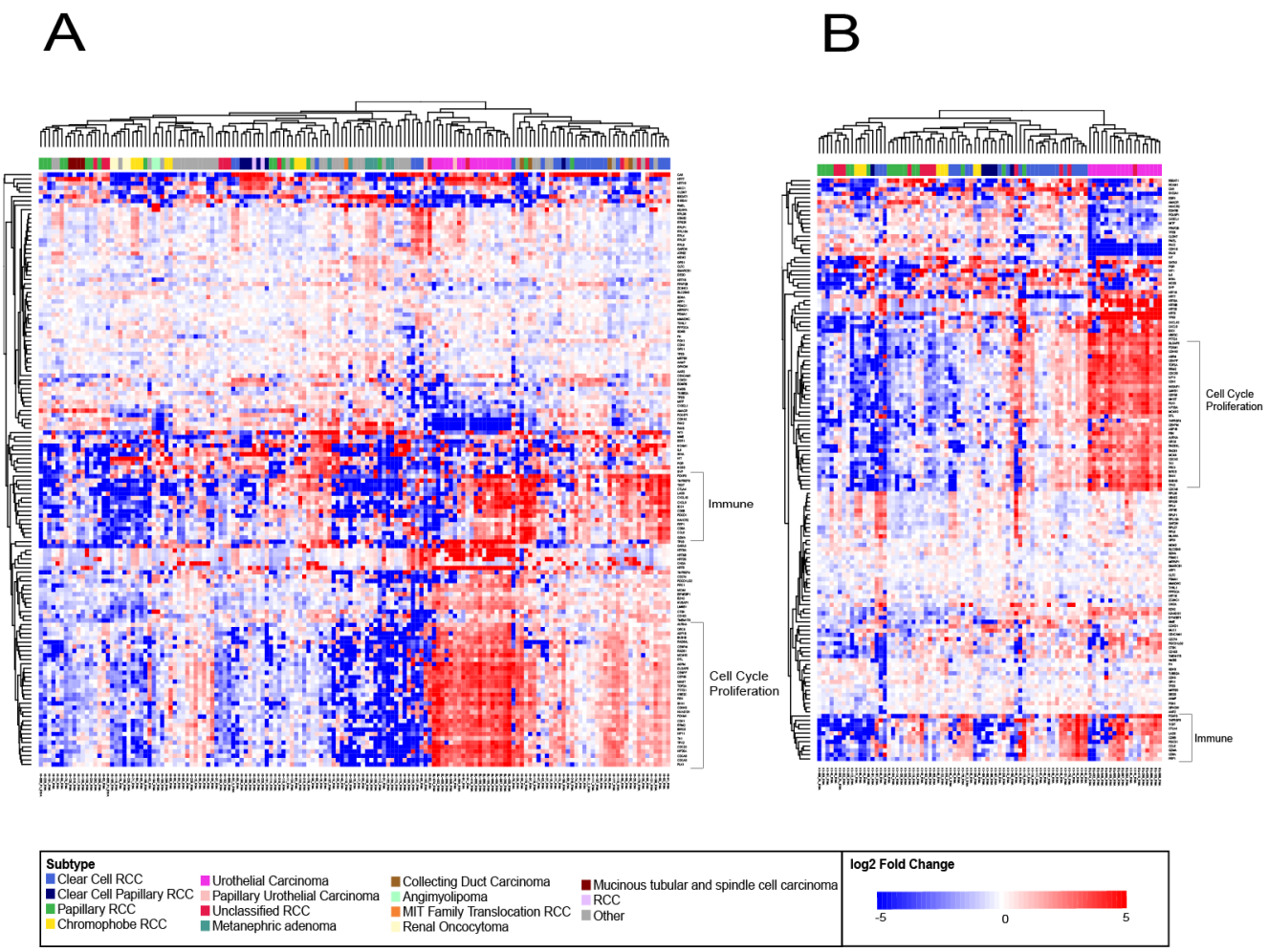

Figure 3: Unsupervised hierarchical clustering of samples and gene targets sequenced on the 122 gene targeted assay. After normalization, genes were centered by their medians, and outlier values were bounded between -5 and 5. Left panel shows all samples, while right shows only the most predominant subtypes.

CCP Scores

Using methodology previously described[3], we calculated a derived Cell Cycle Progression (mxCCP)[15] prognostic score for all samples, shown in **Figure 4** with its composite genes. Urothelial Carcinoma samples showed the highest scores, with Clear Cell RCC samples also trending higher than others, as expected given their clustering behavior. Of note, some samples, for example KI-123, show high mxCCP scores while only showing high expression for a

minority of proliferation targets, supporting the role of multi-gene assays in measuring general proliferation activity.

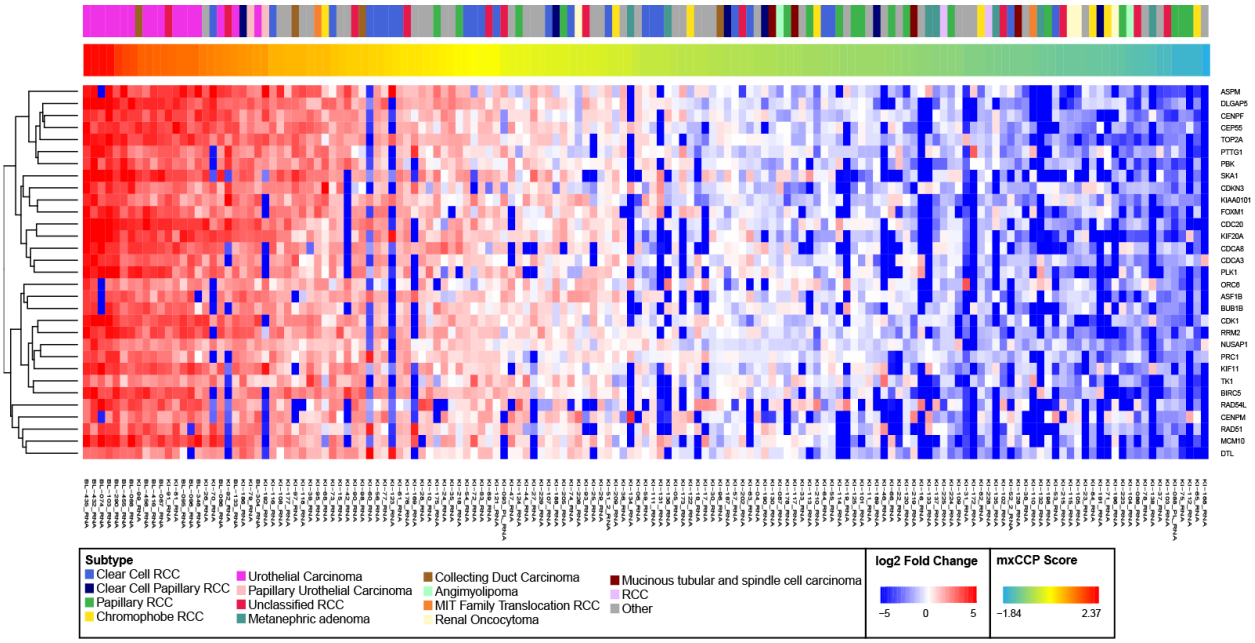

Figure 4: Heatmap of 31 genes that make up the cell cycle progression score. The derived score was calculated as previously described and was centered and scaled by all samples in this cohort to have unit variance and mean. Samples are ordered by score from left (least proliferative) to right (most proliferative). Highly proliferative Urothelial Carcinoma uniformly showed the highest scores, while Chromophobe RCC and Papillary RCC predominantly showed low-proliferation scores.

Although limited in sample number and not the focus of this initial assessment, we did not find significant differences in mxCCP scores across nuclear (Furhman) grades in Clear Cell RCC, as shown in **Figure 5**.

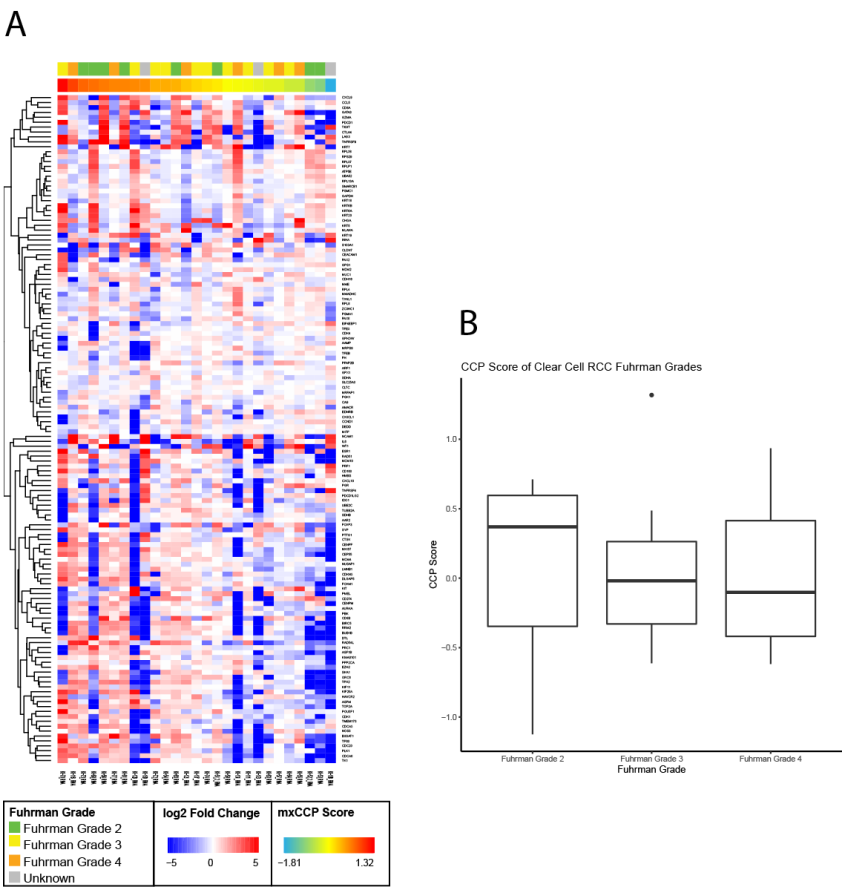

Figure 5: Heatmap and boxplot of Clear Cell RCC samples, annotated with Fuhrman Nuclear Grade. mxCCP score is not predictive of Fuhrman Grade, indicating other processes play a role in poor nuclear differentiation measured by this score.

Subtype Multivariate Modeling

As diagnosis of renal tumors often involves a battery of immunostains, even for pathologists with extensive genitourinary pathology experience, there is the potential to augment or replace an immunohistochemistry (IHC) based approach with mxRNAseq, given this platform's ability to measure many targets simultaneously at a low cost even on very small biopsy samples. Hence, we attempted to characterize the panel's ability to accurately classify primary renal tumor subtype using two approaches. First, we looked at whether measuring the RNA expression of canonical IHC markers[2] could yield an accurate classifier on their own.

Secondly, we sought to see whether the inclusion of other potentially relevant markers included on the targeted panel might allow for better classification.

To assess whether markers traditionally used in IHC diagnostics could provide an accurate diagnosis at the RNA level, we first attempted to replicate the decision rubric used to evaluate canonical IHC antigens[2] (**Figure 1**). This model performed relatively well in our withheld TCGA testing data set with ROC curves with AUC values shown in **Figure 6** but had difficulty generalizing to the mxRNAseq data. Of note, because each gene's positive/negative status is predicted independently of one another, this can produce a pattern that does not match any trained subtype. This occurred in 90 TCGA testing samples, and 32 mxRNAseq validation samples, and were always scored as incorrect for performance validation purposes.

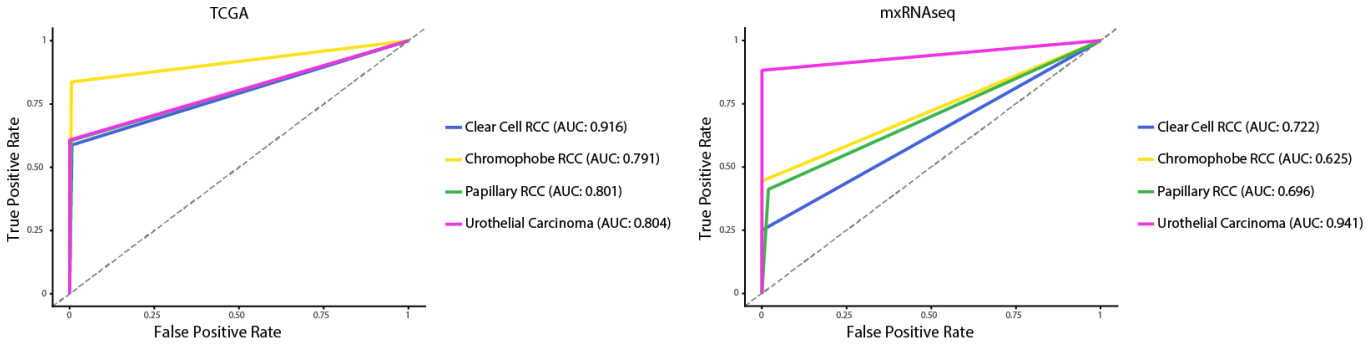

Figure 6: Receiver operating characteristic(ROC) curves measuring the performance of RNAseq replicated IHC decision tree. Separate logistic regression models were trained for each gene involved in IHC-based subtyping, and their predictions for each sample were combined according to the IHC decision tree to yield a predicted subtype. Left panel shows performance on TCGA testing data, right on mxRNAseq validation.

Thus, we next assessed whether statistical learning algorithms could recapitulate the algorithm currently used in clinical practice. We trained a multinomial logistic regression model on just expression of targets measured in clinical IHC-based analysis. The directionality

(positive/negative) of the trained coefficients (**Table 1**) was mostly concordant with the directionality used to subtype tumors using IHC. For example, CA9 and AMACR are both negative in Chromophobe RCC and show negative coefficients for Chromophobe RCC. This approach performed well, with performance gains in both the TCGA and mxRNAseq datasets (**Figure 7**). Of note, Papillary RCC is the only subtype in which this not the case; the directionality of its coefficients do not match the known marker status for TP63 and AMACR.

	Chromophobe RCC	Clear Cell RCC	Papillary RCC	Urothelial Carcinoma
TP63	-1.75828	-3.44541	0.006539	1.405575
PAX8	-0.04699	-0.81781	2.593376	-4.37143
KRT7	-0.36471	-2.34942	-0.70006	1.990789
KIT	0.957877	0.078807	-0.57821	-1.09337
CA9	-2.27608	2.797833	-1.99381	-1.71269
AMACR	-1.12483	-0.13133	0.769411	-0.94736

Table 1 – Multinomial logistic regression coefficients

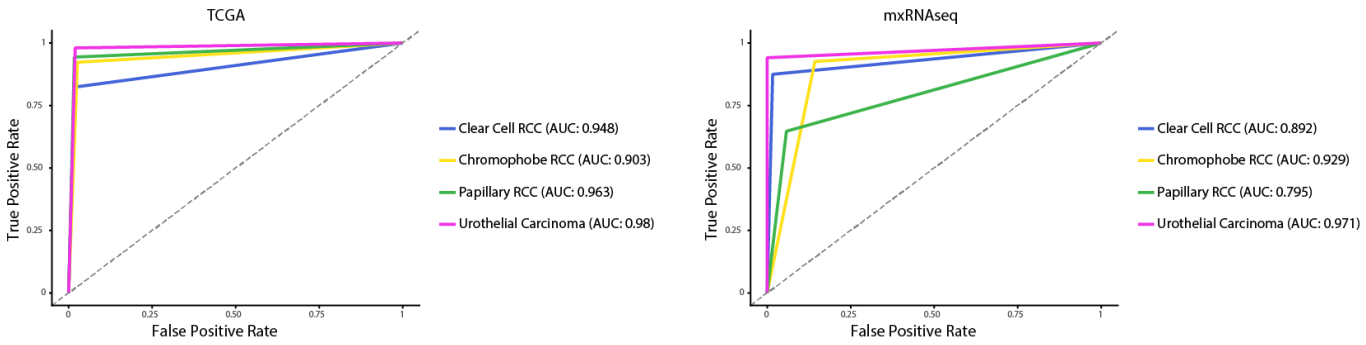

Figure 7: ROC curves measuring the performance of multinomial logistic regression. A multinomial logistic regression models was trained on markers used for IHC analysis. Left panel shows performance on TCGA testing data, right on mxRNAseq validation.

We also attempted to use non-linear models, such as classification and regression trees. These models have the benefit of generating a hierarchical structure similar to ones used in clinical practice, and can capture non-linear relationships that don't appear in the linear models. This model recapitulated the IHC decision tree, while eliminating redundant markers.

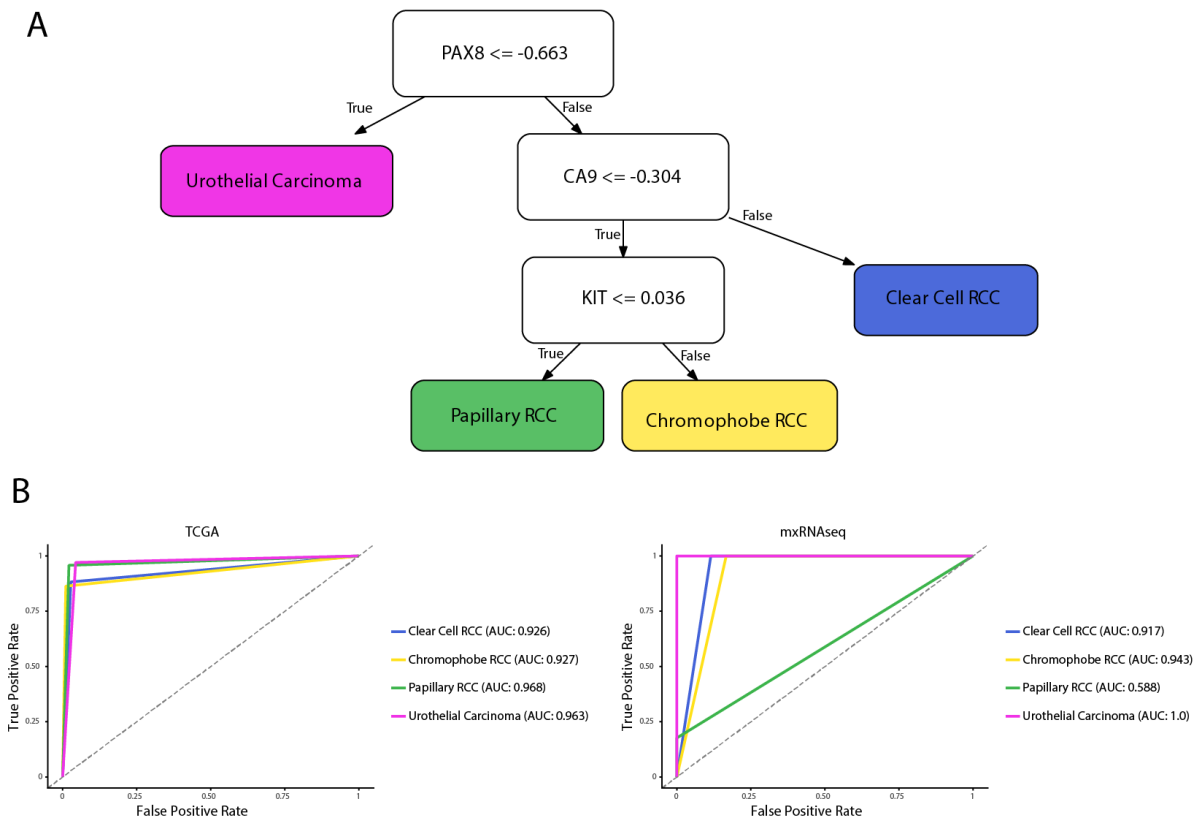

Figure 8: Classification tree and ROC curve performance derived from RNAseq data. Panel A: A learned classification tree from RNAseq analysis of only IHC targets. Urothelial Carcinoma samples are classified solely on PAX8 expression. CA9 expression determined whether a non-UC sample was Clear Cell or not, with KIT expression being the final determinant of the Papillary RCC vs. Chromophore RCC subtype. B: ROC Curves of model performance. Model shows high performance on TCGA data, but generalizes poorly to Papillary RCC in the mxRNAseq cohort.

We next attempted to determine whether analyzing the entire panel (outside of the already known IHC genes) might provide additional insight. We trained a random forest classifier, and adaptive boosted trees classifier on the entire panel, and while it performed well on the TCGA data, it generalized poorly to the mxRNAseq validation data, especially for Chromophobe and Papillary RCC samples (**Figure 9**). One explanation for this discrepancy may

be proliferation - proliferation markers including UBE2C, KIF11, and RAD51 make up 3 of the top genes in the model by gini score, and could be driving classification behaviors on proliferation instead of other biology if there are proliferation differences between the cohorts.

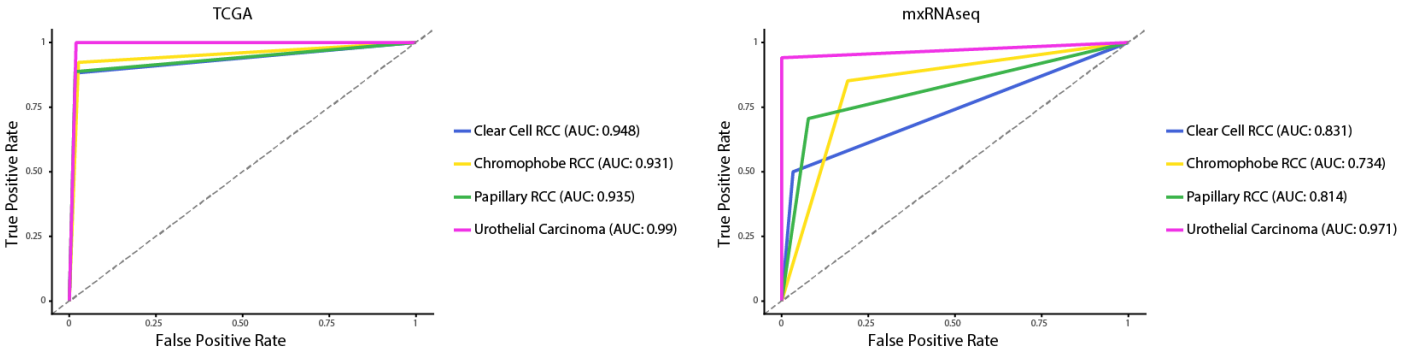

Figure 9: ROC curves of random forest model trained on all targets. Models have strong performance within the TCGA dataset, but generalize poorly to the mxRNA dataset outside of Urothelial Carcinoma.

Unclassified RCCs

A major problem in the management of RCC are tumors that don't meet the classification criteria for any particular subtype, and are therefore deemed "unclassified". Such diagnosis is usually made only after extensive IHC, resulting in a large cost for a diagnosis with little utility to the urologist/medical oncologist. We thus sought to apply our models (trained on classified samples) to these unclassified samples, to see if these models could inform on their potential underlying phenotype. Of the 12 Unclassified RCCs in our cohort, the single gene models only mapped one to the known subtypes – one with Oncocytic features to Chromophobe RCC. The remaining 11 samples had expression patterns that did not match any of the trained subtypes.

Using the well performing multinomial model, we calculated the probability of an unclassified sample belonging to each trained subtype. 6/8 samples with Sarcomatoid features were most likely to be Clear Cell RCC, with the remaining 2 most likely Papillary RCC. One

sample with both Papillary and Sacromatoid features was mostly Clear Cell RCC (77.4%), but had a strong Papillary RCC component (22.4%). While this analysis can't definitively classify these samples, it suggests that RNAseq may enable increased resolution.

DISCUSSION

Our group developed a targeted mxRNAseq panel for the comprehensive profiling of kidney tumors. The panel shows the robust ability to differentiate various kidney cancer subtypes, both unsupervised (as in without any priors as shown in **Figure 3**) and supervised (as in the various models shown in **Figures 6-9**). In addition, we show that a mxRNAseq derived version of the Cell Cycle Progression score is capable of assessing the underlying proliferation of the tumor. We also demonstrate the potential utility of using RNAseq based methods to better characterize unclassified disease through increased molecular resolution.

Our approach has several limitations. The major motivator of this study is to use RNA-based technologies for simpler and higher-throughput classification and characterization of kidney cancers. However, this approach relies on mRNA expression levels of protein targets to be accurate proxies for protein expression. In this case, we show that they are for the purposes of classifying the most common subtypes in our cohort, but the variety of gene regulatory mechanism, including micro-RNA (miRNA) and epigenetic regulation may preclude this approach from working with all subtypes, or different driving molecular alterations.

In addition, the use of TCGA data to train supervised models adds potential bias to our models. We used TCGA data homogenized with our mxRNAseq data to provide substantial training examples for model parameter estimation. However, while our homogenization was

seemingly effective as shown in **Figure 2**, there are artifacts associated with comparing full transcriptome sequenced via a poly-A selection method with targeted amplicons that may not been removed by a standard unit mean and variance homogenization. Additionally, several studies have demonstrated the misclassification of tumors in TCGA data[26], potentially confounding our ability to accurately train model parameters.

The targeted mxRNAseq approach we use poses its own limitations. While targeted mxRNAseq holds promise for its ability to characterize degraded samples and lower cost than traditional transcriptome RNAseq, it also accordingly provides less data than a full transcriptome might. For subtyping or clinical purposes described here, that poses little problem, as that mostly is occurring on a constrained set of genes, However, for characterizations of unclassified RCCs, this technique doesn't fully characterize the expression activity of these cancers. While we have shown that the additional range of detection afforded by RNAseq technologies may allow for deconvolution of known subtypes within unclassified tumors, our panel does not enable the characterization of targets not known to be involved in this cancer. However, future studies involving deep full-transcriptome sequencing of unclassified RCCs may be beneficial to understanding their composition and relationship to known subtypes.

Integration of RNAseq data in this study and DNA somatic single nucleotide and copy number variants may also provide important information for proper characterization of these cancers, as the Tomlins laboratory has shown in other tumor types.[3,27] While mRNA expression provides valuable information about the expression state of the tissue, it doesn't encode the driving mutation(s) responsible for the malignancy.

Importantly, targeted mxRNAseq provides opportunities to simplify and reduce the costs of cancer research. Here we have shown that a kidney cancer specific mxRNAseq assay has the

ability to accurately classify major subtypes using markers clinically used in protein analysis, and the potential for the assay to be used to characterize multi-target phenotypes, such as cell cycle/proliferation.

ACKNOWLEDGEMENTS

Thank you to the entire Tomlins Lab and collaborators for your support over the past few years, especially Komal Kunder and Aaron Udager for sequencing the samples in this study, Dan Hovelson for teaching me basically all I know about bioinformatics analysis, and Dr. Scott Tomlins for giving me the freedom to explore my interests and grow my skillset.

Thank you to my family for your love and support.

REFERENCES

1. Murcia O, Juárez M, Rodríguez-Soler M, et al. Colorectal cancer molecular classification using BRAF, KRAS, microsatellite instability and CIMP status: Prognostic implications and response to chemotherapy. *PLoS ONE*. 2018;13(9):e0203051. doi:10.1371/journal.pone.0203051

2. Reuter VE, Argani P, Zhou M, Delahunt B, Members of the ISUP Immunohistochemistry in Diagnostic Urologic Pathology Group. Best practices recommendations in the application of immunohistochemistry in the kidney tumors: report from the International Society of Urologic Pathology consensus conference. *Am J Surg Pathol*. 2014;38(8):e35-49. doi:10.1097/PAS.0000000000000258

3. Salami SS, Hovelson DH, Kaplan JB, et al. Transcriptomic heterogeneity in multifocal prostate cancer. *JCI Insight*. 2018;3(21). doi:10.1172/jci.insight.123468

4. Cieslik M, Chugh R, Wu Y-M, et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res*. 2015;25(9):1372-1381. doi:10.1101/gr.189621.115

5. Knezevic D, Goddard AD, Natraj N, et al. Analytical validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. *BMC Genomics*. 2013;14:690. doi:10.1186/1471-2164-14-690

6. Thermo Fisher Scientific. Targeted RNA Sequencing by Ion Torrent Next-Generation Sequencing. https://www.thermofisher.com/us/en/home/life-science/sequencing/rna-sequencing/targeted-rna-sequencing-ion-torrent-next-generation-sequencing.html. Accessed March 14, 2019.

7. Illumina. TruSeq Stranded mRNA | Sequence mRNA samples. https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-stranded-mrna.html. Accessed March 14, 2019.

8. Langner C, Ratschek M, Tsybrovskyy O, Schips L, Zigeuner R. P63 Immunoreactivity Distinguishes Upper Urinary Tract Transitional-cell Carcinoma and Renal-cell Carcinoma Even in Poorly Differentiated Tumors. *J Histochem Cytochem*. 2003;51(8):1097-1099. doi:10.1177/002215540305100813

9. Ordóñez NG. Value of PAX 8 Immunostaining in Tumor Diagnosis: A Review and Update. *Advances in Anatomic Pathology*. 2012;19(3):140. doi:10.1097/PAP.0b013e318253465d

10. Genega EM, Ghebremichael M, Najarian R, et al. Carbonic anhydrase IX expression in renal neoplasms: correlation with tumor type and grade. *Am J Clin Pathol*. 2010;134(6):873-879. doi:10.1309/AJCPPPR57HNJMSLZ

11. Liao S-Y, Aurelio ON, Jan K, Zavada J, Stanbridge EJ. Identification of the MN/CA9 Protein As a Reliable Diagnostic Biomarker of Clear Cell Carcinoma of the Kidney. *Cancer Res*. 1997;57(14):2827-2831.

12. Yamazaki K, Sakamoto M, Ohta T, Kanai Y, Ohki M, Hirohashi S. Overexpression of KIT in chromophobe renal cell carcinoma. *Oncogene*. 2003;22(6):847-852. doi:10.1038/sj.onc.1206153

13. Wang H-Y, Mills S. KIT and RCC Are Useful in Distinguishing Chromophobe Renal Cell Carcinoma From the Granular Variant of Clear Cell Renal Cell Carcinoma. *The American Journal of Surgical Pathology*. 2005;29(5):640-646. doi:10.1097/01.pas.0000157943.33903.92

14. Cuzick J, Swanson GP, Fisher G, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes for recurrence and death from prostate cancer: A retrospective study in two cohorts. *Lancet Oncol*. 2011;12(3):245-255. doi:10.1016/S1470-2045(10)70295-3

15. Cuzick J, Berney DM, Fisher G, et al. Prognostic value of a cell cycle progression signature for prostate cancer death in a conservatively managed needle biopsy cohort. *Br J Cancer*. 2012;106(6):1095-1099. doi:10.1038/bjc.2012.39

16. Uhlen M, Zhang C, Lee S, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357(6352). doi:10.1126/science.aan2507

17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825–2830.

18. Davis CF, Ricketts CJ, Wang M, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell*. 2014;26(3):319-330. doi:10.1016/j.ccr.2014.07.014

19. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *New England Journal of Medicine*. 2016;374(2):135-145. doi:10.1056/NEJMoa1505917

20. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013;499(7456):43-49. doi:10.1038/nature12222

21. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507(7492):315-322. doi:10.1038/nature12965

22. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404. doi:10.1158/2159-8290.CD-12-0095

23. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1. doi:10.1126/scisignal.2004088

24. Breiman L, Friedman J., Olshen RA, Stone CJ. Classification and Regression Trees.

25. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell*. 2015;160(1):48-61. doi:10.1016/j.cell.2014.12.033

26. Favazza L, Chitale DA, Barod R, et al. Renal cell tumors with clear cell histology and intact VHL and chromosome 3p: a histological review of tumors from the Cancer Genome Atlas database. *Mod Pathol*. 2017;30(11):1603-1612. doi:10.1038/modpathol.2017.72

27. Hovelson DH, Udager AM, McDaniel AS, et al. Targeted DNA and RNA Sequencing of Paired Urothelial and Squamous Bladder Cancers Reveals Discordant Genomic and Transcriptomic Events and Unique Therapeutic Implications. *Eur Urol*. 2018;74(6):741-753. doi:10.1016/j.eururo.2018.06.047