

Article

Comparative analysis on molecular characteristics of Chromophobe renal cancer and Oncocytoma

Khaled Bin Satter^{1*}, Paul Minh Huy Tran^{1*}, Lynn Kim Hoang Tran¹, Shan Bai¹, Natasha M. Savage⁴, Sravan K. Kavuri⁴, Martha Terris⁵, Jin-Xiong She^{1,2} and Sharad Purohit^{1,2,3}

¹ Center for Biotechnology and Genomic Medicine; Medical College of Georgia, Augusta University, 1120, 15th St, Augusta GA 30912, USA; fbinsatter@augusta.edu (KBS); ptran@augusta.edu (P.M.H.T.); spurohit@augusta.edu (S.P.); lytran@augusta.edu (L.K.H.T.); sbai@augusta.edu (S.B.); jshe@augusta.edu (JXS)

² Department of Obstetrics and Gynecology, Medical College of Georgia, Augusta University, 1120, 15th St, Augusta GA 30912, USA;

³ Department of Undergraduate Health Professionals, College of Allied Health Sciences, Augusta University, 1120, 15th St, Augusta GA 30912, USA

⁴ Department of Pathology, Medical College of Georgia, Augusta University, 1120 15th St, Augusta GA 30912, USA; nsavage@augusta.edu (N.S.); skavuri@augusta.edu (S.K.)

⁵ Department of Urology, Medical College of Georgia, Augusta University, 1120 15th St, Augusta GA 30912, USA; mtorris@augusta.edu (M.T.)

*These authors contributed equally to this work

Correspondence: Sharad Purohit, PhD spurohit@augusta.edu

Abstract (200 words): Chromophobe renal cell carcinoma (chRCC) and oncocytoma (RO) are renal tumor types originating from alpha intercalated cells of the collecting ducts of the kidney. Both tumor types have similar gross histological morphology and increased mitochondria, which leads to difficulties differentiating between these tumors, especially with core biopsy samples. This study aims to apply a machine learning approach to develop a molecular classifier based on transcriptomics data. Here we generated a meta-data set containing 62 chRCC and 45 RO gene expression arrays. Arrays were subjected to quality control steps, and genes were selected based on differential expression and ROC analysis. The final gene list was evaluated with UMAP based dimension reduction followed by density-based clustering with 95.5% accuracy. Molecular profiling by KEGG pathway analysis identified enrichment of fatty acid oxidation pathway in RO. We finally identified and validated the 30-gene signature, with an accuracy of 94.4% to distinguish chRCC from RO on UMAP analysis. Our results show that chRCC and RO have a distinct gene signature that can differentiate these tumors and complement histology for routine diagnosis of these two tumors.

Keywords: Renal cancers, oncocytoma, chromophobe, transcriptomics, machine learning, clustering, gene signature, unsupervised learning

1. Introduction

Chromophobe renal cell carcinoma (chRCC) and oncocytoma are renal tumor types originating from alpha intercalated cells of the collecting ducts of the kidney [1,2]. They account for 5% and 3-7% of all of the renal neoplasms, respectively [3–6]. Oncocytoma is benign and typically requires no treatment [6]. On the other hand, chRCC is a malignant lesion with 5-year overall survival of 90%; it is treated with a partial or radical nephrectomy [3,5]. ChRCC is a well-circumscribed tumor that is frequently larger than any other renal cancer [4,7], whereas oncocytomas can also be large, ranging in size from 1.5 cm to 13 cm, like chRCC [8]. Both have a brown appearance grossly. A central stellate scar is frequently noted in oncocytomas but may not always be present. Consequently, these tumors are difficult to distinguish grossly and by computed tomography (CT) imaging.

Histologically, chRCC is composed of sheets of cells with well-defined cell borders that have darker cytoplasm than conventional clear cell carcinoma [6]. On the other hand, oncocytoma has variable architecture but frequently consists of nests of tumor cells comprised of large, round, eosinophilic cells in loose connective tissue [6]. However, some cases have small cells with hyperchromatic nuclei, termed oncoblasts, which can emulate other eosinophilic renal cancers and pose difficulties in providing an accurate diagnosis from a core needle biopsy, where there is limited histologic context [9].

Electron microscopy may aid in the diagnosis of chRCC, but is not employed routinely in determining diagnosis by pathologists [4]. Immunohistochemical staining for cytokeratin 7, epithelial-mesenchymal antigen, and parvalbumin (PVALB) are occasionally employed, but these three markers have variable sensitivity, from 60-100%, in the diagnosis of chRCC [10,11]. To differentiate oncocytoma, immunohistochemical localization of cytokeratin 7, S100A1 [12], and kidney-specific cadherins are employed, but the overlap between the markers of chRCC and oncocytoma makes it an ineffective method of differentiation. On top of these issues, the tissue collection by core needle biopsy, commonly used in the diagnosis of small renal masses, has a misclassification rate of 20% due to low tissue volume and the presence of oncoblasts [13–15].

Molecular diagnostic methods have been used to identify specific genetic changes in disease and can aid in patient stratification [16]. Previous studies on chRCC and oncocytomas proposed markers like parafibromin, aquaporin 6, and synaptogyrin 3 [17]. Genetic markers like *AP1M2*, *MAL2*, *PROM2*, *PRSS8*, *FLJ20171* [18] and *EGLN2* [19] might be useful as well in addition to the currently available markers. The major drawbacks to these molecular classification studies are the smaller sample size and overlapping expression in chRCC and oncocytoma.

Here, we identified transcriptomic differences distinguishing chRCC from oncocytoma in a meta-dataset combined from multiple studies from the Gene Expression Omnibus (GEO) and ArrayExpress. From the gene expression differences, we elucidated pathway differences between chRCC and oncocytoma. We then implemented machine learning (ML) algorithms and developed an 80-gene transcriptomic signature and ML models to distinguish chRCC from oncocytoma.

2. Materials and Methods

2.1 Dataset search and selection:

ChRCC and oncocytoma transcriptomic microarray studies were identified in GEO and ArrayExpress. We included studies that contained treatment naïve “chRCC” and “oncocytoma” samples and were arrayed using the HG-U133plus 2 mRNA expression array. Based on the selection criteria, 6 studies (GSE11024, GSE11151, GSE12090, GSE2109, GSE8271, GSE19982) were chosen and their expression and phenotype data were downloaded [20,21] with GEOquery [22]. Phenotype data were prepared from the downloaded files and only chRCC and oncocytoma arrays are selected in data preparation steps. The ideal probe set for the study was determined with jetset [23] and all overlapping probes in jetset and all studies were selected (15875 probes). Five studies (GSE11151, GSE11024, GSE19982, GSE2109, GSE8271) were combined and checked for batch variation with boxplots and principal component analysis (PCA). Batch effects were corrected with ComBat from sva [24] using GSE19982 as a reference batch. GSE12090 was kept as the validation dataset. Batch effect correction and normalization were checked with PCA, uniform manifold approximation and projection (UMAP) [25], and boxplots.

2.2 Unsupervised learning, Density-based UMAP (DBU):

Density-based UMAP [26] consists of two algorithms, a dimension reduction and a clustering algorithm. We implemented UMAP as dimension reduction and clustering

was done with density-based clustering (DBSCAN). The optimum parameter for UMAP was identified with a grid-search parameter identifier. This identifier implements different parameters and visual inspection of cluster density and inter-cluster distance was used to identify the optimum parameters. Number of neighbors, minimum neighbor distance, distance metric, and the number of probes were used in the grid search. The final chosen parameters are 20 neighbor, 0.01 minimum distance, 1000 probes, and Manhattan distance metric. DBU was iterated 1000 times with these parameters and the resulting 2-dimensional UMAP coordinates were fed into a density-based clustering algorithm. The optimum parameter for density-based clustering was chosen with an elbow plot (eps = 2, Minimum points 5) from the “fpc” package[27]. Each iteration classified all the arrays into groups and plurality voting was implemented for final group determination. A 70% or higher agreement for one single group was used as a cut-off for group classification. If a sample does not reach 70% for a group, it is called “ambiguous”. Grouping consensus heatmap was generated with gplots [28]. Alluvial plot was generated with the ggalluvial package[29].

2.3 Differential expression and network analysis and Immune cell infiltration:

Differentially expressed genes between chRCC and oncocytoma were assessed with limma [30] and fed into gene set enrichment analysis (GSEA) with fgsea [31], clusterprofiler [32] and ReactomePA [33] to identify networks and pathways involved in chRCC and oncocytoma [34]. Differential expression was presented with a heatmap generated from ComplexHeatmap [35] and the network plot was generated with ReactomePA and enrichplot [28,31]. KEGG plots were created with the “pathview” package for specific pathways [38].

2.4 Gene Signature:

We calculated the sensitivity, specificity, accuracy and AUC for each sample percentile. Genes with the highest sensitivity and specificity were initially selected [39]. We narrowed down the list by comparing their expression levels at the different percentiles. Genes with higher fold change and relevance to chRCC or oncocytoma based on literature review were given preference. We selected a total of 80 genes. The 80 gene signature’s efficiency was evaluated with heatmap, unsupervised learning (UMAP and DBU), and a random forest supervised learning model.

2.5 Statistical analysis:

All statistical analyses in this study were performed using R language and environment for statistical computing [40], version 4.10. All p-values were two-sided and a $p < 0.05$ was considered significant.

3. Results

3.1 Meta-dataset of publicly available oncocytoma and chRCC microarray data

We queried the Gene Expression Omnibus (GEO) and ArrayExpress databases using the search terms “chRCC” and “oncocytoma” and identified 24 microarray studies (Figure 1). After removing duplicates, 17 studies were evaluated for eligibility using the following criteria: 1) the chRCC and oncocytoma tumors used for the array were treatment naïve, 2) the array assessed mRNA expression, and 3) the array used was HGU133 plus2.

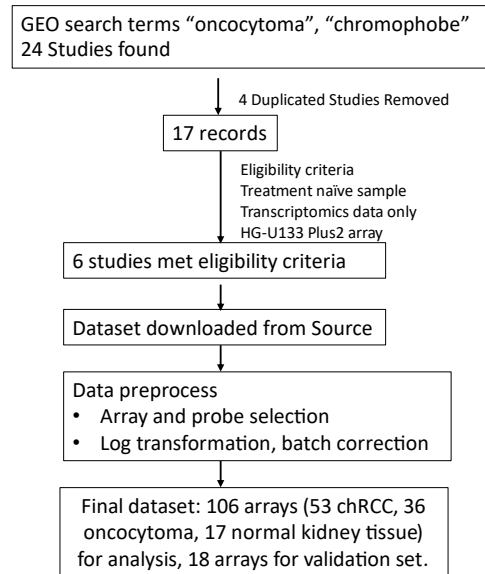


Figure 1: Flow chart depicting selection and preparation of chRCC and Oncocytoma arrays from GEO for meta-analysis.

Based on these eligibility criteria, six studies were selected and downloaded from GEO and ArrayExpress (Table 1). The six studies contained 107 arrays: 45 arrays for oncocytoma tumors and 62 arrays for chRCC. Five studies were used to develop a training-test dataset (53 chRCC and 36 oncocytoma arrays) for unsupervised and supervised machine learning models and sixth dataset (GSE12090, 9 chRCC and 9 oncocytoma arrays) was used for the validation set.

Table 1: GEO datasets selected for chRCC and oncocytoma classification.

GEO Accession ID	Probes	Total Number of samples	Oncocytoma Samples	ChRCC Samples	Normal Kidney sample (N)	Training vs validation	Year
GSE11024	17700	79	7	6	12	Training	2008
GSE11151	54676	67	4	4	5	Training	2008
GSE19982	54676	30	15	15	0	Training	2010
GSE12090	54676	18	9	9	0	Validation	2008
GSE8271	54676	34	10	10	0	Training	2007
GSE2109	17232	2158	0	18	0	Training	2004

All downloaded arrays were merged using the probes which overlapped across all arrays. This resulted in a combined dataset of 89 samples and 15875 probes. Principal component analysis (PCA) demonstrated batch differences among the downloaded studies (Figure 2, A & B). Bayes empirical model was used for batch effect correction with study accession ID as batch and histological types as a covariate. Study GSE19982 was used as a reference batch for the batch correction as it contained the highest number of samples. After batch effect correction, PCA demonstrated global transcriptomic differences between the oncocytoma and chRCC histology types were greater than the transcriptomic differences among the downloaded studies (Figure 2C and D). This resulting dataset was used for further analysis.

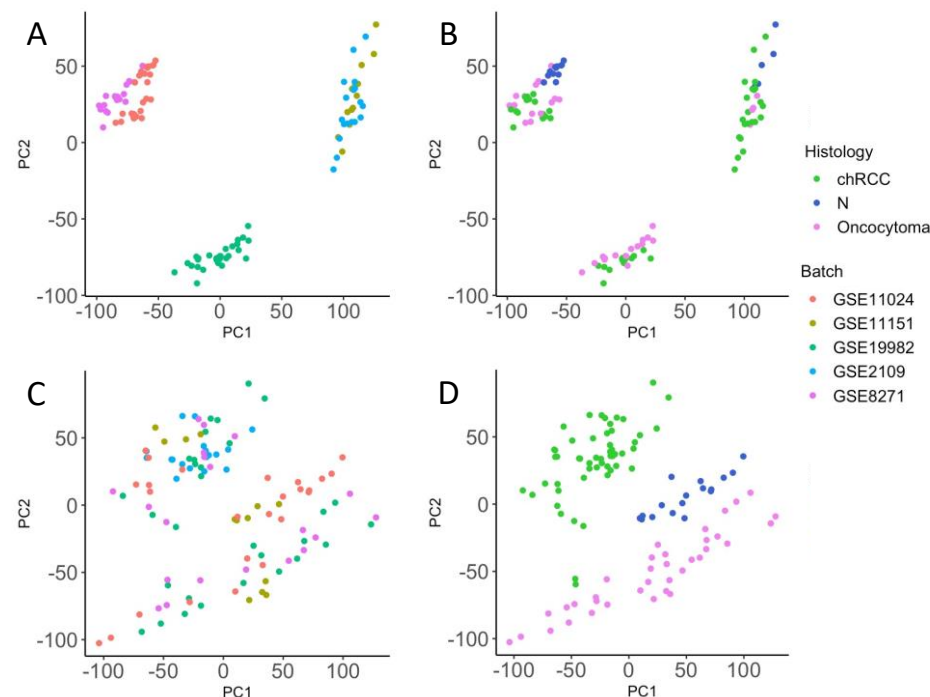


Figure 2: Batch effect correction and normalization of the meta-dataset. A and B: Principal component analysis showing batch and histology for chRCC, oncocytoma and normal kidney tissue arrays before batch effect correction. C and D: Principal component analysis showing batch and histology for chRCC, oncocytoma and normal kidney tissue arrays after removal of batch effect. “N” represents normal kidney tissue.

3.2 Unsupervised clustering using gene expression data largely agrees with histology classification

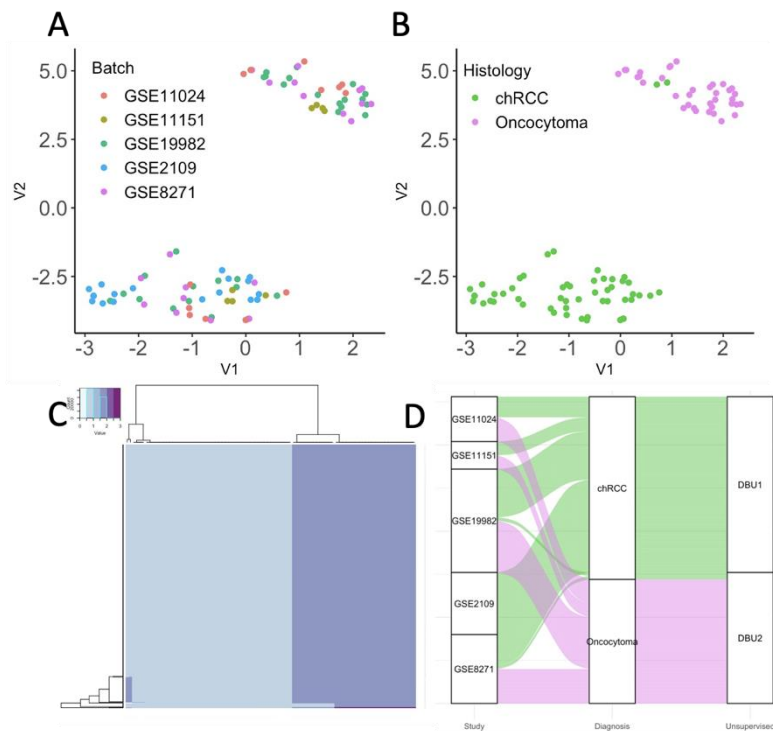


Figure 3: Transcriptomic map of chRCC and oncocytoma arrays in meta-dataset. (A). UMAP plots with Batch annotation. (B). UMAP plots with histology annotation. N represents normal kidney tissue. (C) Group consensus heatmap of 1000 iteration of Density-based UMAP algorithm. Samples are shown in columns and iteration in rows. The color diagram indicates their consensus group. (D) Sankey's diagram for batch and histology and DBU (Density-based UMAP) classification).

Using unsupervised clustering with uniform manifold approximation of projection (UMAP) for all 15875 probes, the 89 arrays formed two distinct clusters with no major batch effects from individual studies (Figure 3A). Transcriptome of ChRCC and oncocy-toma each formed their own clusters (Figure 3B), suggesting

that chRCC and oncocytoma have distinct transcriptomic profiles. Two chRCC samples were clustered with oncocytoma, suggesting sample misclassification. We therefore evaluated the robustness of the unsupervised classification with the density-based UMAP (DBU) algorithm [26]. The DBU algorithm takes a random subset of 1000 genes from the dataset and applies dimension reduction with UMAP to all samples. The UMAP output is then fed into density-based clustering to identify distinct groups.

DBU was iterated 1000 times and the final classification was determined with plu-rality voting (Figure 3C). DBU identified 51 samples in group 1 and 38 samples in group 2 (termed “DBU1” and “DBU2,” respectively). DBU1 contains 51 chRCC samples, while DBU2 contains 36 oncocytoma and 2 chRCC samples (Figure 3D). Two samples were classified as “Ambiguous”. The overall agreement between DBU and histology was 95.5%.

3.3 Differential expression, network and pathway analysis

We performed differential expression analysis between chRCC and oncocytoma after excluding misclassified and ambiguous samples. We found that 8411 out of 15875 probes are differentially expressed

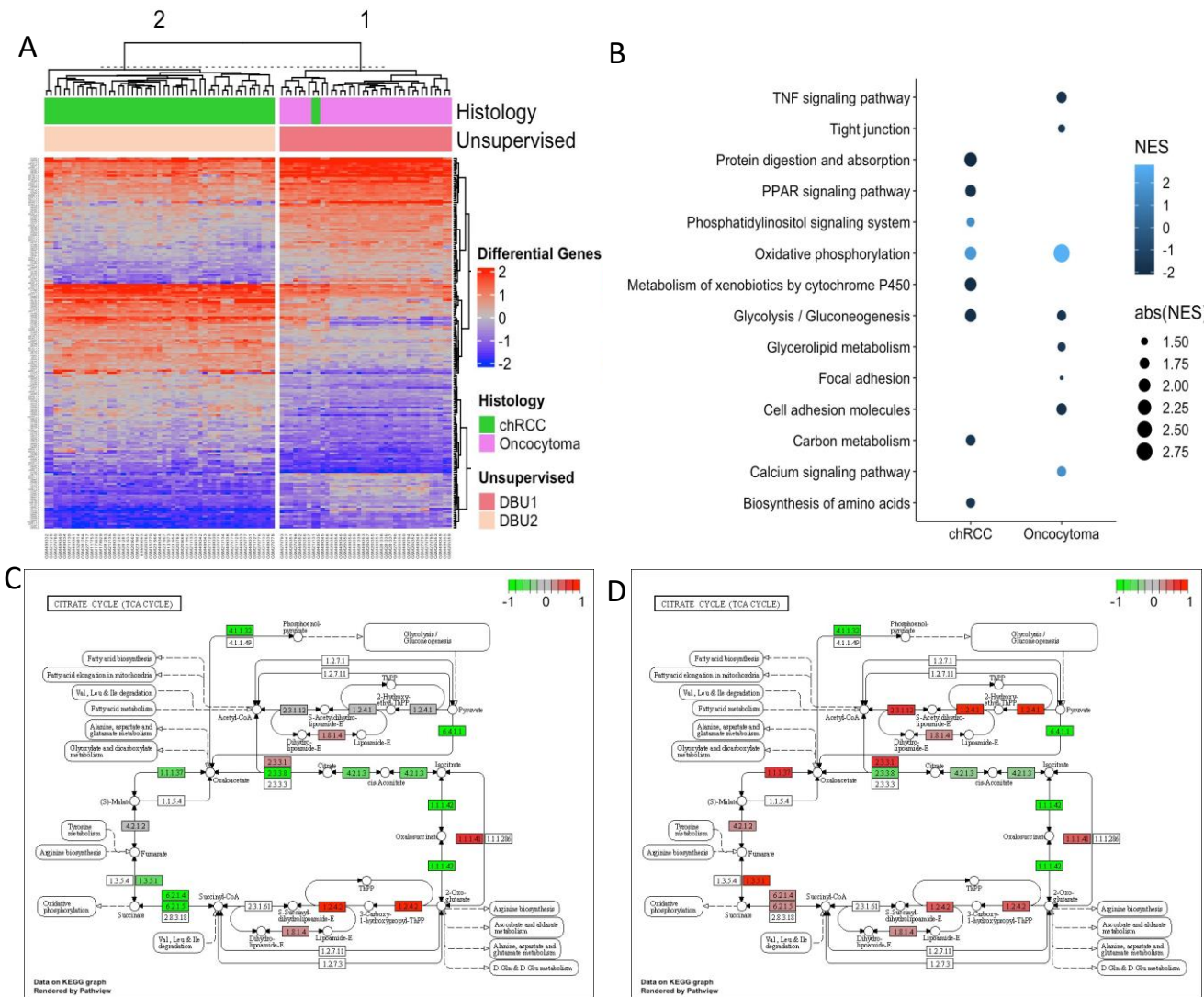


Figure 4: Molecular and pathway analysis. A: Top 194 Differentially expressed genes between chRCC and oncocytoma. Probes are presented in rows and samples are in columns. Row annotation bars are currently us. B: Bubble plot showing top upregulated pathways in chRCC and oncocytoma (x-axis) from Gene set enrichment analysis on canonical pathways (y-axis) for differentially expressed genes, size of the bubble represents normalized enrichment score. C: KEGG pathway analysis for TCA cycle in chRCC D) KEGG pathway analysis of oncocytoma.

between the two tumors (Supplemental Figure 3). Additionally, 299 probes have at least log2 fold change of 1 between the tumors. AUC analysis of these 316 genes identified 194 genes with an AUC of 0.9 or higher between the tumors. Hierarchical clustering with these 194 genes showed a 97.8% agreement with DBU classification (Figure 4A). These results show that there are large transcriptomic differences between chRCC and oncocytoma.

We compared differentially expressed genes (DEG) in both chRCC and oncocytoma with normal kidney cortex using gene set enrichment analysis (GSEA). GSEA using DEG between chRCC and normal kidney tissue identified 67 significantly enriched reactome pathways (adjusted p value < 0.05, Benjamini-Hochberg correction) out of 218 pathways. Notably enriched pathways are transcriptomic regulation by TP53, beta-catenin independent WNT signaling, diseases of signal transduction by growth factor receptors and second messengers, carbohydrate metabolism, interleukins 4 and 13 signaling, glycosaminoglycan metabolism, and extracellular matrix organization.

KEGG pathway analysis on chRCC showed oxidative phosphorylation, phosphatidylinositol signaling system upregulated with normalized enriched score of 1.96 and 1.58 (BH adjusted p value 9.47×10^{-5} and 3.38×10^{-3}). KEGG pathway analysis on oncocytoma identified 36 pathways significantly enriched (BH adjusted), with upregulation of the oxidative phosphorylation and calcium signaling pathways (NES 2.8 and 1.6). Downregulated pathways in oncocytoma are mineral absorption, cell adhesion molecules, glycolysis and gluconeogenesis, TNF signaling pathway (NES -2.2, -1.8, -1.71 and -1.79). Oxidative phosphorylation in chRCC showed upregulated complex I, III, and IV, with downregulation of complex II. Oxidative phosphorylation in oncocytoma showed upregulation of all complexes in the electron transport chain (Supplemental figure). The TCA cycle in chRCC have isocitrate dehydrogenase and oxoglutarate dehydrogenase upregulated and succinate CoA ligase downregulated (Figure 4C). TCA cycle in oncocytoma have upregulated fatty acid metabolism and isocitrate dehydrogenase, oxoglutarate dehydrogenase and succinate CoA ligase. Both tumors showed upregulated arginine biosynthesis, ascorbate and aldarate, alanine, aspartate and glutamate metabolism (Figure 4C).

3.4 Identification of a Gene signature for chRCC/Oncocytoma classification

To develop a robust gene signature to differentiate between chRCC and oncocytoma, we identified genes with maximum diagnostic utility by calculating sensitivity, specificity, AUC, accuracy for all percentiles. We identified 84 genes which have a sensitivity + specificity of > 1.8, an AUC > .92, and an accuracy > .89. Genes with higher fold change, inter-percentile fold change and their relevance to chRCC and oncocytoma were given priority for selection. We selected 30 genes for the final gene signature (GS30). We evaluated the robustness of the signature using 1000 iterations of an unsupervised model. We developed unsupervised models with 10, 15 and 20 genes from GS30 and

identified 20 genes as the optimum number for classification and that less than 20 genes introduce ambiguity in the classification.

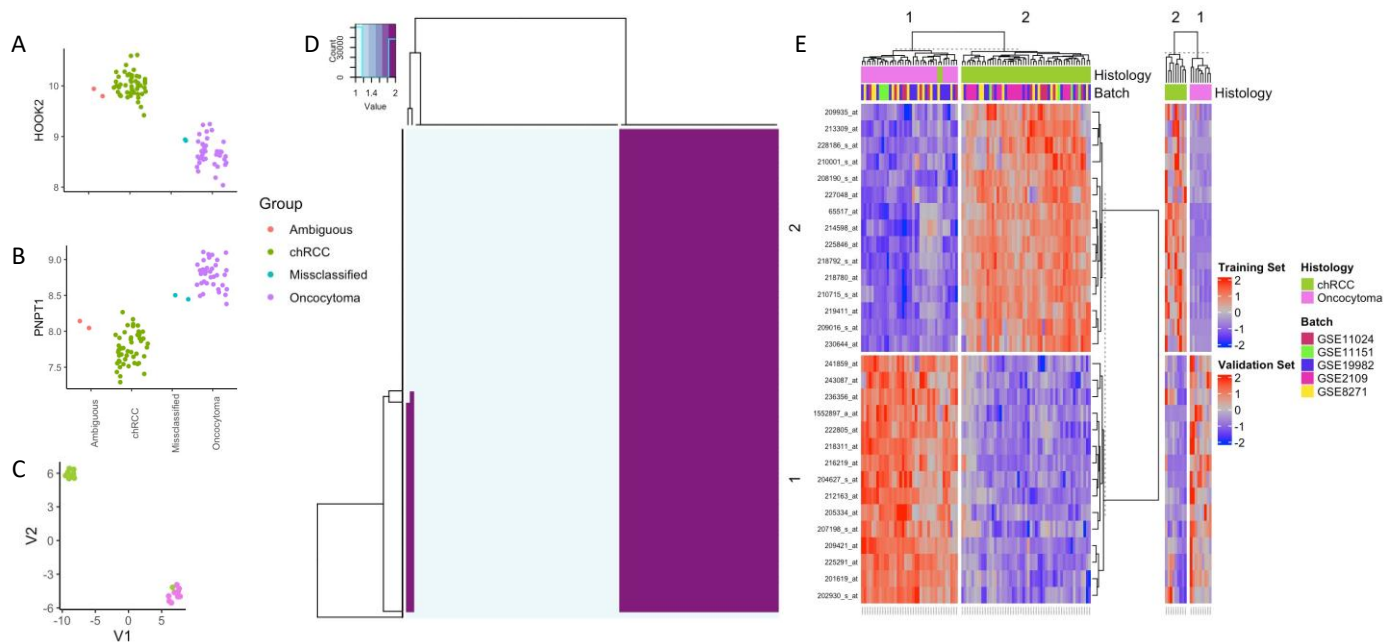


Figure 5: Gene selection and unsupervised model consistency. A-B: Dot plot for 2 representative genes from GS30. Ambiguous samples shown are from DBU with 1000 gene models. Y-axes represent expression. C: UMAP plots for validation cohort with GS30. D: Group consensus heatmap by DBU with 20 random genes from GS30. E: Heatmap of GS30 in training (Left) and validation (Right) set.

DBU with 20 genes from GS30 (Figure 5D) have accuracy of 95.5% in plurality voting for 1000 iterations, similar to 1000 gene model and confirms GS30's ability to capitulate the difference the tumors as good as 1000 gene models. In total, 49 of chRCC and 36 of oncocytoma samples are correctly classified with these models. Two chRCC samples are classified with oncocytoma and two samples were ambiguous, similar to the 1000 gene models.

The accuracy of GS30 assessed on the validation cohort GSE12090 (9 chRCC and 9 Oncocytoma) was 94.4%. Only one chRCC sample was misclassified to oncocytoma. Heatmap and UMAP analysis showed clear distinction between the tumors (Figure 5C and E).

4. Discussion

This study reports a robust way to differentiate chRCC from oncocytoma using transcriptomics. ChRCC and oncocytoma have distinct transcriptomic profiles that can differentiate from each other (Figure 3B). Single molecules from current IHC markers showed an overlap at the transcript level, causing a broad range of sensitivity and specificity. Unsupervised models with 20 genes provided a consistent classification. Based on these models, we developed GS30, with an accuracy of .95 in 1000 unsupervised models. This gene signature shows potential for clinical use for differentiating chRCC from oncocytoma through PCR, microarray, or RNASeq based assays. In addition, these assays do not require large amount of tissue, making this a suitable candidate for a core needle biopsy.

IHC and special stain markers for distinguishing chRCC and oncocytoma include CK7, S100A1, CD117 (KIT), kidney specific cadherins, KAI, cyclin D1, and Hale's colloidal iron are currently used to differentiate the tumors in discrepant cases. CK7 has a wide range of sensitivity and specificity, ranging from 73-100% and 71-100%, respectively [41-45] in different studies. Similarly, S100A1 has sensitivities and specificities at 80-100% and 70-92%, respectively [11,46]. In our study, the sensitivity and specificity for CK7 are 94% and 89% and for S100A1 are 98% and 74% which are similar to previous

studies [42–44,46–48] at transcript level. In comparison, our unsupervised ensemble models have a sensitivity and specificity of 92.4% and 100% for chRCC. This is markedly better than current IHC markers. From our panel, *AP1M2*, *AQP6*, *HOOK2*, *CLDN*, *ESRP1* [18,49] have previously been reported as candidate markers in the literature.

Molecular profiling with KEGG pathway analysis on chRCC identified that calcium signaling, carbon and glycolipid metabolism, PPAR signaling pathway, and TNF signaling pathways (Figure 4B) are distinctive compared with normal kidney tissue. ChRCC and oncocytoma both showed upregulated oxidative phosphorylation with decreased glycolysis and gluconeogenesis suggesting energy metabolic reprogramming in both tumors. Comparison in the TCA cycle pathway shows a dependence of fatty acid metabolism in RO (Figure 4D). Signal transduction by growth factor receptor, TP53 transcriptional regulation, platelet activation, signaling, and aggregation are top differentiating pathways between tumors when comparing functional differences between chRCC and oncocytoma on Reactome pathways.

The strength of our study combining multiple datasets to increase the power of detecting transcriptomic difference. Although transcriptomic studies from The Cancer Genome Atlas contains far largest sample set, our dataset is the largest of chRCC-RO samples ($n = 89$). Our unsupervised algorithm identified 95.5% and 94.4% cases correctly in testing and validation dataset. However, we identified 4 chRCC samples that are not congruent with their histological diagnosis. Two of the samples did not reach consensus of 70%, are ambiguous, with almost evenly split in plurality voting. In case of remaining two samples the transcriptomic profile is indistinguishable from oncocytoma suggesting possible misclassification. Microscopy combined with immunohistochemical analysis are often insufficient in distinguishing chRCC and oncocytoma in biopsies. Hence, transcriptomics is important for further resolving the two pathologies. Since we demonstrated that oncocytoma and chRCC each have unique gene expression and pathways expressed, transcriptomics may aid in distinguishing cases found to be uncertain on biopsy and histologic analysis.

5. Conclusions

We show that chRCC and oncocytoma are transcriptomically distinct tumors. Using our unsupervised learning method, we were able to separate these two tumors into distinct classes. Our results suggest that transcriptomics combined with supervised learning can be a good clinical tool to aid pathologists to distinguish chRCC from oncocytoma tumors for the diagnosis of these two tumors.

Author Contributions: Conceptualization, K.B.S., P.T., J-X.S.; methodology K.B.S., P.T., S.P.; formal analysis, K.B.S., P.T.; investigation, K.B.S., P.T., S.P., J-X.S.; resources, S.P.; data curation, K.B.S.; writing—original draft preparation, K.B.S.; writing—review and editing, K.B.S., P.T., L.T., S.P., J-X.S.; visualization, K.B.S., P.T., S.P.; supervision, S.P.; funding acquisition, J.X.S. All authors have read and agreed to the published version of the manuscript.

Funding: PMHT was supported by NIDDK fellowship F30DK121461.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to the use of publicly available data with identifiers completely removed.

Informed Consent Statement: This study was done with de-identified human tissue expression data from publicly available sources and does not require consent.

Data Availability Statement: Data for this study is publicly available and can be downloaded from the gene expression omnibus or ArrayExpress website using accession number. Histopathology and other clinical information are also available with the datasets.

Acknowledgments: The authors acknowledge the support of the Graduate School, Augusta University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, L.; Henske, E.P. Chromophobe Renal Cell Carcinoma: New Genetic and Metabolic Insights. *Urol Oncol* **2020**, *38*, 678–681, doi:10.1016/j.urolonc.2020.04.035.
2. Lindgren, D.; Eriksson, P.; Krawczyk, K.; Nilsson, H.; Hansson, J.; Veerla, S.; Sjölund, J.; Höglund, M.; Johansson, M.E.; Axelson, H. Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Reports* **2017**, *20*, 1476–1489, doi:10.1016/j.celrep.2017.07.043.
3. Stec, R.; Grala, B.; Mączewski, M.; Bodnar, L.; Szczylić, C. Chromophobe Renal Cell Cancer - Review of the Literature and Potential Methods of Treating Metastatic Disease. *Journal of Experimental & Clinical Cancer Research* **2009**, *28*, 134, doi:10.1186/1756-9966-28-134.
4. Vera-Badillo, F.E.; Conde, E.; Duran, I. Chromophobe Renal Cell Carcinoma: A Review of an Uncommon Entity. *International Journal of Urology* **2012**, *19*, 894–900, doi:https://doi.org/10.1111/j.1442-2042.2012.03079.x.
5. Chao, D.H.; Zisman, A.; Pantuck, A.J.; Freedland, S.J.; Said, J.W.; Belldegrun, A.S. Changing Concepts in the Management of Renal Oncocytoma. *Urology* **2002**, *59*, 635–642, doi:10.1016/S0090-4295(01)01630-2.
6. Atkins, M.; Choueiri, T. Epidemiology, Pathology, and Pathogenesis of Renal Cell Carcinoma - UpToDate Available online: <https://www.uptodate.com/contents/epidemiology-pathology-and-pathogenesis-of-renal-cell-carcinoma> (accessed on 2 March 2021).
7. Amin, M.B.; Paner, G.P.; Alvarado-Cabrero, I.; Young, A.N.; Stricker, H.J.; Lyles, R.H.; Moch, H. Chromophobe Renal Cell Carcinoma: Histomorphologic Characteristics and Evaluation of Conventional Pathologic Prognostic Parameters in 145 Cases. *Am J Surg Pathol* **2008**, *32*, 1822–1834, doi:10.1097/PAS.0b013e3181831e68.
8. Hes, O.; Michal, M.; Sulc, M.; Podhola, M.; Zámečník, M.; Curík, R.; Miculka, P.; Neubauer, L.; Kinkor, Z.; Pavlovský, M. [Oncocytoma of the kidney--morphologic variation in 102 cases]. *Cesk Patol* **2001**, *37*, 51–56.
9. Kryvenko, O.N.; Jorda, M.; Argani, P.; Epstein, J.I. Diagnostic Approach to Eosinophilic Renal Neoplasms. *Arch Pathol Lab Med* **2014**, *138*, 1531–1541, doi:10.5858/arpa.2013-0653-RA.
10. Adley, B.P.; Papavero, V.; Sugimura, J.; Teh, B.T.; Yang, X.J. Diagnostic Value of Cytokeratin 7 and Parvalbumin in Differentiating Chromophobe Renal Cell Carcinoma from Renal Oncocytoma. *Anal Quant Cytol Histol* **2006**, *28*, 228–236.
11. Kim, S.S.; Choi, Y.D.; Jin, X.M.; Cho, Y.M.; Jang, J.J.; Juhng, S.W.; Choi, C. Immunohistochemical Stain for Cytokeratin 7, S100A1 and Claudin 8 Is Valuable in Differential Diagnosis of Chromophobe Renal Cell Carcinoma from Renal Oncocytoma. *Histopathology* **2009**, *54*, 633–635, doi:https://doi.org/10.1111/j.1365-2559.2009.03271.x.
12. Yusenko, M.V.; Zubakov, D.; Kovacs, G. Gene Expression Profiling of Chromophobe Renal Cell Carcinomas and Renal Oncocytomas by Affymetrix GeneChip Using Pooled and Individual Tumours. *Int J Biol Sci* **2009**, *5*, 517–527, doi:10.7150/ijbs.5.517.
13. Leveridge, M.J.; Finelli, A.; Kachura, J.R.; Evans, A.; Chung, H.; Shiff, D.A.; Fernandes, K.; Jewett, M.A.S. Outcomes of Small Renal Mass Needle Core Biopsy, Nondiagnostic Percutaneous Biopsy, and the Role of Repeat Biopsy. *European Urology* **2011**, *60*, 578–584, doi:10.1016/j.eururo.2011.06.021.
14. Frank, I.; Blute, M.L.; Cheville, J.C.; Lohse, C.M.; Weaver, A.L.; Zincke, H. Solid Renal Tumors: An Analysis of Pathological Features Related to Tumor Size. *J Urol* **2003**, *170*, 2217–2220, doi:10.1097/01.ju.0000095475.12515.5e.
15. Lane, B.R.; Babineau, D.; Kattan, M.W.; Novick, A.C.; Gill, I.S.; Zhou, M.; Weight, C.J.; Campbell, S.C. A Preoperative Prognostic Nomogram for Solid Enhancing Renal Tumors 7 Cm or Less Amenable to Partial Nephrectomy. *J Urol* **2007**, *178*, 429–434, doi:10.1016/j.juro.2007.03.106.
16. Dwivedi, S.; Purohit, P.; Misra, R.; Pareek, P.; Goel, A.; Khattri, S.; Pant, K.K.; Misra, S.; Sharma, P. Diseases and Molecular Diagnostics: A Step Closer to Precision Medicine. *Indian J Clin Biochem* **2017**, *32*, 374–398, doi:10.1007/s12291-017-0688-8.

17. Tan, M.-H.; Wong, C.F.; Tan, H.L.; Yang, X.J.; Ditlev, J.; Matsuda, D.; Khoo, S.K.; Sugimura, J.; Fujioka, T.; Furge, K.A.; et al. Genomic Expression and Single-Nucleotide Polymorphism Profiling Discriminates Chromophobe Renal Cell Carcinoma and Oncocytoma. *BMC Cancer* **2010**, *10*, 196, doi:10.1186/1471-2407-10-196.
18. Rohan, S.; Tu, J.J.; Kao, J.; Mukherjee, P.; Campagne, F.; Zhou, X.K.; Hyjek, E.; Alonso, M.A.; Chen, Y.-T. Gene Expression Profiling Separates Chromophobe Renal Cell Carcinoma from Oncocytoma and Identifies Vesicular Transport and Cell Junction Proteins as Differentially Expressed Genes. *Clin Cancer Res* **2006**, *12*, 6937–6945, doi:10.1158/1078-0432.CCR-06-1268.
19. Koeman, J.M.; Russell, R.C.; Tan, M.-H.; Petillo, D.; Westphal, M.; Koelzer, K.; Metcalf, J.L.; Zhang, Z.; Matsuda, D.; Dykema, K.J.; et al. Somatic Pairing of Chromosome 19 in Renal Oncocytoma Is Associated with Deregulated EGLN2-Mediated [Corrected] Oxygen-Sensing Response. *PLoS Genet* **2008**, *4*, e1000176, doi:10.1371/journal.pgen.1000176.
20. Home - GEO - NCBI Available online: <https://www.ncbi.nlm.nih.gov/geo/> (accessed on 3 March 2021).
21. ArrayExpress < EMBL-EBI Available online: <https://www.ebi.ac.uk/arrayexpress/> (accessed on 3 March 2021).
22. mail.nih.gov>, S.D. <sdavis2 at GEOquery: Get Data from NCBI Gene Expression Omnibus (GEO); Bioconductor version: Release (3.13), 2021;
23. Li, Q.; Birkbak, N.J.; Györfy, B.; Szallasi, Z.; Eklund, A.C. Jetset: Selecting the Optimal Microarray Probe Set to Represent a Gene. *BMC Bioinformatics* **2011**, *12*, 474, doi:10.1186/1471-2105-12-474.
24. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Jaffe, A.E.; Storey, J.D. The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* **2012**, *28*, 882–883, doi:10.1093/bioinformatics/bts034.
25. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]* **2020**.
26. Tran, P.M.H.; Tran, L.K.H.; Nechtman, J.; Dos Santos, B.; Purohit, S.; Satter, K.B.; Dun, B.; Kolhe, R.; Sharma, S.; Bollag, R.; et al. Comparative Analysis of Transcriptomic Profile, Histology, and IDH Mutation for Classification of Gliomas. *Sci Rep* **2020**, *10*, 20651, doi:10.1038/s41598-020-77777-6.
27. Hennig, C. *Fpc: Flexible Procedures for Clustering*; 2020;
28. Warnes, G.R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.; Liaw, A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; et al. *Gplots: Various R Programming Tools for Plotting Data*; 2020;
29. Brunson, J.C.; Read, Q.D. *Ggalluvial: Alluvial Plots in "Ggplot2"*; 2020;
30. Smyth, G.; Hu, Y.; Ritchie, M.; Silver, J.; Wettenhall, J.; McCarthy, D.; Wu, D.; Shi, W.; Phipson, B.; Lun, A.; et al. *Limma: Linear Models for Microarray Data*; Bioconductor version: Release (3.12), 2021;
31. Korotkevich, G.; Sukhov, V.; Budin, N.; Sergushichev, A. *Fgsea: Fast Gene Set Enrichment Analysis*; Bioconductor version: Release (3.12), 2021;
32. Yu, G.; Wang, L.-G.; Hu, E.; Chen, M.; Dall'Olio, G. *ClusterProfiler: Statistical Analysis and Visualization of Functional Profiles for Genes and Gene Clusters*; Bioconductor version: Release (3.13), 2021;
33. Yu, G.; Petyuk, V. *ReactomePA: Reactome Pathway Analysis*; Bioconductor version: Release (3.12), 2021;
34. Carlson, M. Org.Hs.Eg.Db Available online: <http://bioconductor.org/packages/org.Hs.eg.db/> (accessed on 3 March 2021).
35. Gu, Z. *ComplexHeatmap: Make Complex Heatmaps*; Bioconductor version: Release (3.12), 2021;
36. Blighe, K.; Rana, S.; Turkes, E.; Ostendorf, B.; Lewis, M. *EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced Colouring and Labeling*; Bioconductor version: Release (3.12), 2021;
37. Guangchuang Yu *Enrichplot*; Bioconductor, 2018;

38. Luo, W. *Pathview: A Tool Set for Pathway Based Data Integration and Visualization*; Bioconductor version: Release (3.13), 2021;
39. Thiele, C. *Cutpointr: Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks*; 2021;
40. R Core Team (2021) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>;
41. Liu, L.; Qian, J.; Singh, H.; Meiers, I.; Zhou, X.; Bostwick, D.G. Immunohistochemical Analysis of Chromophobe Renal Cell Carcinoma, Renal Oncocytoma, and Clear Cell Carcinoma: An Optimal and Practical Panel for Differential Diagnosis. *Arch Pathol Lab Med* **2007**, *131*, 1290–1297, doi:10.1043/1543-2165(2007)131[1290:IAOCRC]2.0.CO;2.
42. Geramizadeh, B.; Ravanshad, M.; Rahsaz, M. Useful Markers for Differential Diagnosis of Oncocytoma, Chromophobe Renal Cell Carcinoma and Conventional Renal Cell Carcinoma. *Indian Journal of Pathology and Microbiology* **2008**, *51*, 167, doi:10.4103/0377-4929.41641.
43. Zhou, J.; Yang, X.; Zhou, L.; Zhang, P.; Wang, C. Combined Immunohistochemistry for the “Three 7” Markers (CK7, CD117, and Claudin-7) Is Useful in the Diagnosis of Chromophobe Renal Cell Carcinoma and for the Exclusion of Mimics: Diagnostic Experience from a Single Institution. *Disease Markers* **2019**, *2019*, e4708154, doi:10.1155/2019/4708154.
44. El-Shorbagy, S.H.; Alshenawy, H.A. Diagnostic Utility of Vimentin, CD117, Cytokeratin-7 and Caveolin-1 in Differentiation between Clear Cell Renal Cell Carcinoma, Chromophobe Renal Cell Carcinoma and Oncocytoma. *Journal of Microscopy and Ultrastructure* **2017**, *5*, 90–96, doi:10.1016/j.jmau.2016.07.005.
45. Wu, S.L.; Kothari, P.; Wheeler, T.M.; Reese, T.; Connelly, J.H. Cytokeratins 7 and 20 Immunoreactivity in Chromophobe Renal Cell Carcinomas and Renal Oncocytomas. *Mod Pathol* **2002**, *15*, 712–717, doi:10.1097/01.MP.0000017566.29755.8A.
46. Conner, J.R.; Hirsch, M.S.; Jo, V.Y. HNF1 β and S100A1 Are Useful Biomarkers for Distinguishing Renal Oncocytoma and Chromophobe Renal Cell Carcinoma in FNA and Core Needle Biopsies. *Cancer Cytopathol* **2015**, *123*, 298–305, doi:10.1002/cncy.21530.
47. Mathers, M.E.; Pollock, A.M.; Marsh, C.; O'Donnell, M. Cytokeratin 7: A Useful Adjunct in the Diagnosis of Chromophobe Renal Cell Carcinoma. *Histopathology* **2002**, *40*, 563–567, doi:10.1046/j.1365-2559.2002.01397.x.
48. ELHAMID, M.A.; Mohebat H. Gouda, M.D., M.A.E., M. Sc.; Nehal S. Zafer, M.D., T.A.E., M.D.; Elbaz, M.D., M. Chromophobe Renal Cell Carcinoma, Oncocytoma and Clear Cell Carcinoma: A Comparative Immunohistochemical Study. *The Medical Journal of Cairo University* **2020**, *88*, 103–110, doi:10.21608/mjcu.2020.93966.
49. Wu, H.; Fan, L.; Liu, H.; Guan, B.; Hu, B.; Liu, F.; Hoher, B.; Yin, L. Identification of Key Genes and Prognostic Analysis between Chromophobe Renal Cell Carcinoma and Renal Oncocytoma by Bioinformatic Analysis. *Biomed Res Int* **2020**, *2020*, 4030915, doi:10.1155/2020/4030915.
50. Davis, C.F.; Ricketts, C.J.; Wang, M.; Yang, L.; Cherniack, A.D.; Shen, H.; Buhay, C.; Kang, H.; Kim, S.C.; Fahey, C.C.; et al. The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. *Cancer Cell* **2014**, *26*, 319–330, doi:10.1016/j.ccr.2014.07.014.