Epigenetic Mutations in clear cell Renal Carcinoma

Alia Shafi

I. Background

Cancer is a disease that is inherited by a selection of cells with either a germ line or somatic genetic or epigenetic alteration of genes that results in growth and proliferation (1). This results in a neoplasm, in which cell multiplication is uncontrolled and progressive. Most cancerous cells found in tumors results from variations of the original mutant cells. Scientists are interested in identifying these mutations that cause cancerous tumors. Studies in the past have attempted to identify common epitopes among different types of carcinomas, however most have been unsuccessful. Mutations in epigenetic regulatory elements such as loss of function mutations or gain of function mutations have a large effect on transcriptome expression. Different epigenetic regulators, such as siRNAs or chromatin remodeling enzymes, have an effect on the levels of expression in cancer genomes. Though there has been research on VHL mutations in RCC, attempts to identify other frequently mutated genes have been unsuccessful.

In recent years, there has been an estimate of 58,000 new cases of kidney cancer among patients in the United States (1). Over 90% of these 58,000 are diagnosed renal cell carcinomas (RCC) (1). Localized clear-cell renal cancer could potentially be curable by removal of all or part of the organ via surgery (2). However, studies about ccRCC show that there seems to be a 30% relapse rate (2). There has yet to be a strategy that is noninvasive and effective for the detection and prognosis of ccRCC (2). Like other cancers, ccRCC does not accompany symptoms until late stages of cancer development (2). However, exploring and analyzing alterations in histone modifiers and other epigenetic mutations may provide a method in prognosis of ccRCC.

The most frequent form of RCC occurs in people that have von-Hipple Lindau (VHL) syndrome (1). This tumor suppressor gene was the first among the RCC genes that was identified as hypermethylated (1). VHL is mutated in approximately 80% of all ccRCC and inactivated in about 10% cases by methylation. This inactivation seems to only been prominent in clear cell RCC. Other attempts to detect mutations in the genome similar to VHL have been unsuccessful (2). However, deregulation of chromatin machinery and large scale sequencing projects has recently emerged as an important tool in studying ccRCC (2). Researchers have identified crucial tumor suppressor genes (TSG) associated with ccRCC (2). These TSG’s were mapped to 3p21 locus and function as epigenetic chromatin and or histone modifications, indicating that epigenetic mutations may play a key role in ccRCC development (2).

The epigenetic alterations in cancer cells include DNA methylation, histone modification and RNA interference (1). The most studied of these alterations is DNA methylation that can occur at the cytosine (1). DNA methyltransferases (DNMT1) generate and maintain methylation patterns (2). There have been some genome wide methylation studies on ccRCC with the use of BeadChip arrays (2). Acetylation or methylation of lysine or arginine and serine phosphorylation is the most common N-terminus tail modification (2). These studies compared DNA methylation profiles in familial and sporadic VHL+/+ ccRCC, which showed more frequent methylated RSSF1, PITX2, CDH13, HS3sT2, TWIST1, TAL1, TUSC3 , and DCC loci. This indicates that there may be differences in tumorigenesis mechanisms that are dependent on VHL (2). In a study conducted by Kluzek et al, patients with BAP1 mutations were significantly more likely to have metastases and a shorter survival rate (2). Similarly, PBRM1 down regulation also had a correlation with advanced tumor stage and a high relapse rate. Nuclease remodeling complexes can also play a role in changing the structure and activity of chromatin by altering the accessibility of regulatory DNA sequences to transcription factors (2).

The interplay between epigenetic mutations and genetic mutations exhibits genetic mutations in cancer (6). Epigenetic mechanisms allow the signal transduction pathways to create a set of progeny from a multi-potent progenitor (4). Epigenetic alterations allow cells to divide, memorize and important signaling events that occur in their early development (4). The use of epigenetic effects is emerging as a common strategy for development of the immune system (4). This paper plans to evaluate the association between the level of immune expression and epigenome deregulation in cancer cells.

II. Data Description

Renal clear cell carcinoma has high levels of inflammation and has ample mRNA expression data. In order to determine all epigenetic mutations in RCC, we gathered data from cBio of all the mutations in key epigenetic regulatory elements identified in prior pan-cancer analyses.

For RCC, mRNA expression data for all tumor samples we mined from synapse using R. The tumors were then separated into two groups based on whether they had epigenetic mutations recorded from cBio.

*About cBio*

The cBio Portal for Cancer Genomics provides visualization, analysis and downloads of large-scale cancer genomic data sets (5). This portal allows researchers to easily and readily analyze genetic alterations across genes, samples and pathways. The website provides visual summaries, gene-level data and software programmatic access. This project uses cBio to gather mutations in important epigenetic regulatory elements identified in prior pan-cancer analyses.

*About Synapse*

Synapse is a database that aims to analyze the predictive value of various molecular phenotypes and computational models in determining the survival of cancers. The TCGA has collected a multitude of data including copy number variation, micro RNA expression, protein expression, clinical covariates and methylation (7). All input data is made available in a directory divided by cancer types. For each cancer type there is a training set, testing set and core set of delimited files for each platform of measurements.

*Retrieval of Data*

With the use of cBio and software R, we used bioinformatics techniques to write an cBio script to obtain all the epigenetic data. The TCGA id’s which identifies the specific tumor in each mutation was observed as well as the type of mutation. The mRNA expression data was analyzed from Synapse using R. The tumors in this sample was filtered into two different groups based on the expression of epigenetic mutations gathered in cBio.

The data we obtained was preprocessed RPKM data. Because this data is already preprocessed, we cannot do quality checks without the raw data. After conducting exploratory analysis and reading on synapse the quality checks that had been done one the raw data, we confirmed that the data was ready for downstream analysis. (preprocessing???)

III. Methods

*Synapse*

We had to create an account with Synapse to mine data from this database. We then used ‘synapseClient’ to retrieve data and download it into our R project. This gave us mRNA expression data for all tumor samples.

*Exploratory Analysis*

In order to retrieve all data from cBio, we had to download the package (‘cgdsr’) from the cBio portal. Once all the data was downloaded, we preformed statistical analysis on the data. We first separated the genes into tumors that had epigenetic mutations and the tumors that did not have epigenetic mutations. We then applied the mean, median, and standard deviation functions to each data fame. We then wrote a function for to find the coefficient of variance, the ration of the standard deviation and mean. The coefficient of variance is often used in studies to analyze the precision and repeatability of an assay. Table 1 displays the exploratory analysis summary for both data frames.

*Differential Expression Analysis*

We used a number of techniques to identify the differential expression in the dataset. We used the Bioconductor package, bioclite(‘qvalue’), ggplot and ggplot2. In order to find a decent cutoff, we plotted a histogram of the RPKM values and decided to place the cutoff at 17% to filter out all the low RPKM data. We then wanted to find the significant genes in the data set. So we calculated the p values, q values and LogFC’s of the filtered RPKM data sets and conducted a T test.

**Figure 1 Histogram of RPKM**

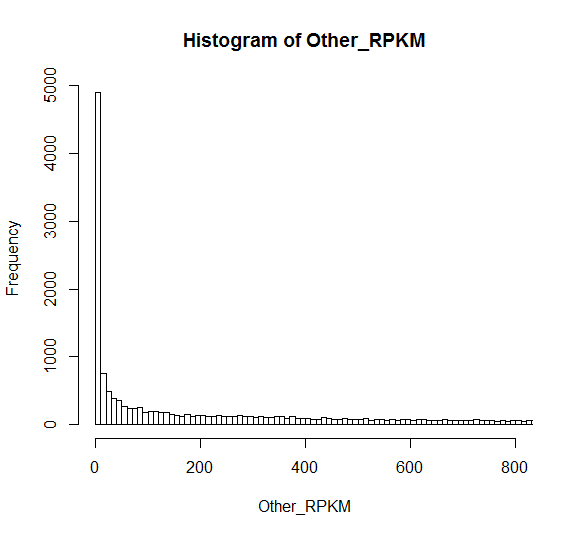


Fig. 1 shows the histogram of RPKM data after we decided the cut off of 17%.

*Volcano Plot*

We used a volcano plot to quickly identify the changes in the dataset. We plotted the graph for epigenetic mutation tumors versus non-epigenetic mutation tumors. Figure 1 displays the plot, which reveals that there is evidence that there are several insignificant values that have a high fold change.

Fig.2 This plot displays a volcano plot for epigenetic mutation tumors vs no-epigenetic mutation tumors which was calculated and constructed in software R. The bottom of the volcano plot reveals that there is several statistically insignificant values that display a large log fold change. Significant genes are clustered around an FC of zero. Because of small intragenic variance and a low RPKM suggests these groups were found significant. Interestingly, there is a lack of genes with a high fold change and small p values. This either means that there is no significant data or there is a fundamental problem in the calculation and assumption of the p value using the T statistic.

*GSEA Analysis*

Gene Set Enrichment Analysis is a method that determines whether a set of genes displays statistically significant, concordant differences between two biological states. The output of GSEA is a list of enriched sets that contain Normalized Enrichment Scores (NES), p value, FDR correction, and FWER. Genes that contribute the most to the NES are outlined with a red box. Table 2 shows positive enrichment of pathways from GSEA analysis. The red box’s mark the pathways that are worth noting. For table 2, we can see that epithelial mesenchymal transition, KAS signaling UP, Xenobiotic Metabolism, Inflammatory response, Apoptosis, Spermatogenesis and DNA repair all seem to contain significant p and q values that are worth noting. Table 3 shows negative enrichment of pathways from GSEA analysis. As we can see from the table, KRAS signaling DN is the most important pathway for negative enrichment. KRAS signaling promotes the Wnt/B-catenin pathway that can drive neoplasia (8). Figure 2 displays the GSEA Enrichment Plots for TNFA, EMT, Inflammatory Response, and KRAS down pathways.

It is a possibility that we have gotten incredibly lucky and the noise we are seeing corresponds to what we expected to see. To solve this, the data was permuted so each sample was randomly placed in a category rather than based on epigenetic mutations. After doing this, our results were confirmed.

*Fisher’s Exact Test*

Fisher’s exact test is a statistical significance test used in the analysis of contingency tables. For this test, we made a top table of all the genes with their p values, q values and Log2FC. Using this table, we isolated the genes into 3 criteria: (1) genes that have a p value less than 0.05, a q value less than 0.25 and a log2FC greater than 0, (2) genes have a log2FC greater than 0.5, (3) and genes that fit both of the criteria. We then gathered data from MSiDB (the same as GSEA) to get the names of genes in the hallmark dataset that are in immune sets, such as inflammatory response, NFKB, IL5, Jak/Stat and others. We then calculated the overlap between the genes from each criterion of the previous step with each MsiDB (or the numbers that they have in common). We then used this numerical data to create a fishers test table which is displayed in table 4.

III. Results

*Exploratory Analysis*

Table 1 shows the data for the exploratory analysis. The tumors with epigenetic mutations have a lower mean, standard deviation, median and coefficient of variance. However, the tumors with epigenetic mutations have a higher maximum that tumors without epigenetic mutations. This is important to note because the tumors with epigenetic mutations should be more prominent than the tumors without epigenetic mutations.

**Table 1. Exploratory Analysis Summary**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | SD | Min | Median | Max | CV |
| Tumors with Epigenetic Mutations | 1157.874 | 4354.882 | 0 | 347.44 | 1246972 | 376.110 |
| Tumors without Epigenetic Mutations | 1224.074 | 5196.56 | 0 | 349.78 | 38554.6 | 424.530 |

*Volcano Plot*

The results of the volcano plot reveal that there are some statistically significant values that have a high fold change and the significant genes are clustered around an FC of zero. Because ther eis a small intragene variance and a low RPKM, these genes were found to be significant.

**Figure 2. Volcano Plot for Epigenetic Mutation Tumors vs No-Epigenetic Mutation Tumors**

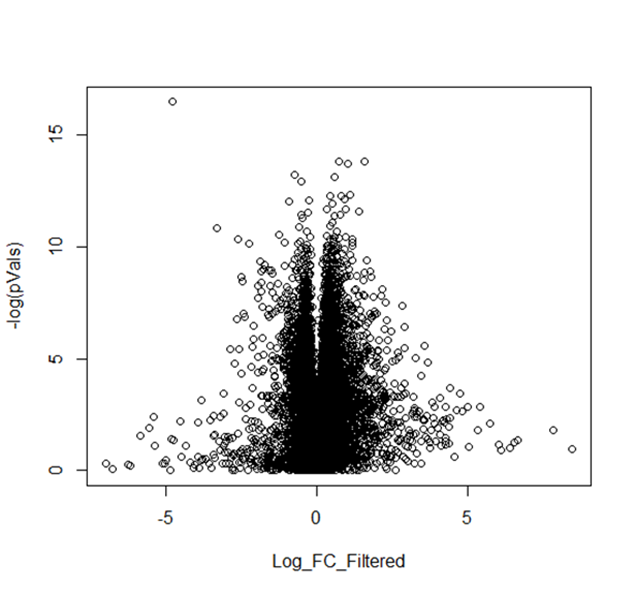


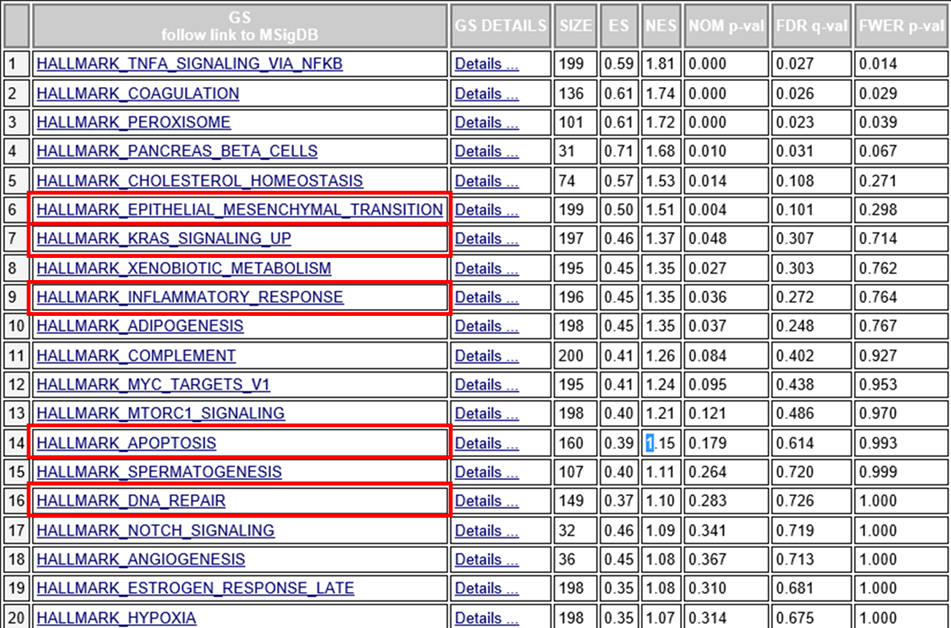
Fig.2 This plot displays a volcano plot for epigenetic mutation tumors vs no-epigenetic mutation tumors which was calculated and constructed in software R. The bottom of the volcano plot reveals that there is several statistically insignificant values that display a large log fold change. Significant genes are clustered around an FC of zero. Because of small intragenic variance and a low RPKM suggests these groups were found significant. Interestingly, there is a lack of genes with a high fold change and small p values. This either means that there is no significant data or there is a fundamental problem in the calculation and assumption of the p value using the T statistic.

*GSEA Analysis*

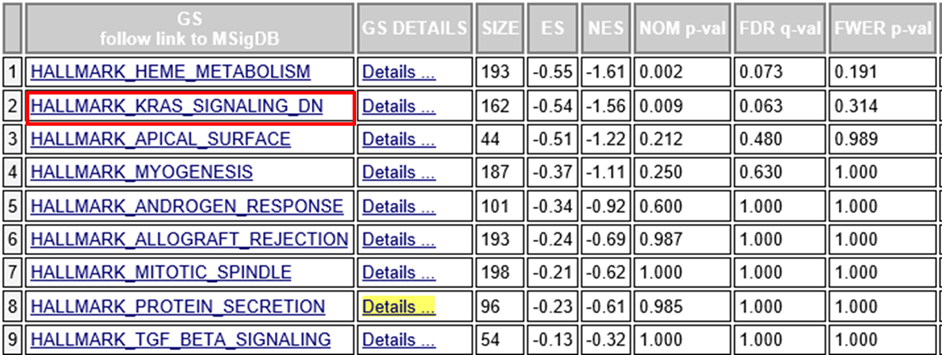
The results of the GSEA analysis show us the NES, p value and FDR correction FWER, and the genes that contribute the most to the NES. The epithelial mesenchymal transition, the KRAS signaling and the inflammatory response are the most important in the positive enrichment (Table 2). Epithelial mesenchymal transition has been prominent in other cancers as well. Epithelial mesenchymal transition is a process when epithelial cells lose their polarity and ability to adhere to other cells and become mesenchymal stem cells that can differentiate into a number of stem cells (9). Nouri *et al* studied the removal of androgens and its effect on epithelial to mesenchymal transition in prostate cancer (9). KRAS signaling has an effect on the Wnt pathway, and inflammatory response also plays a key role in cancer response.

Table 3 shows the negative enrichment of pathways from GSEA analysis. The KRAS signaling down pathway is the most important in this table.

**Table 2. Positive Enrichment of Pathways from GSEA Analysis**



**Table 3. Negative Enrichment of Pathways from GSEA Analysis**



**Figure 3. GSEA Enrichment Plots for TNFA, EMT, Inflammatory Response, and KRAS Down Pathways**

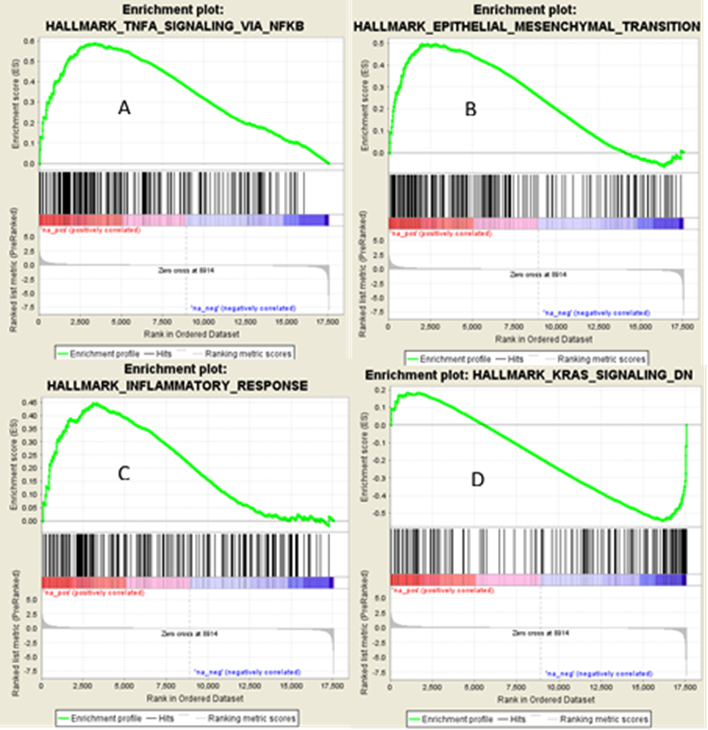


Fig 3. The vertical lines displayed at the bottom of the enrichment plot represent the density of the genes that were used in the enrichment score. A through C are the upregulated pathways, and D is the down regulated pathway. These graphs reveal that most of the genes contributed positively to the enrichment score based on the density of genes that appear before the absolute maximum or minima. Enrichment scores must be normalized first before they can be compared directly to each other.

*Fishers Exact Test*

The fisher exact test is used to examine the significance of the association between two categorical data sets. For this test, we examined the specific hallmark pathways and their overlap with the genes that followed the specific criteria below. The results are displayed below.

**Table 4. Fisher’s Test P Values for Various Differential Expression Heuristics**

|  |  |  |  |
| --- | --- | --- | --- |
| Pathway | Criteria |  |  |
|  | log2FC>0.5 (n=721) | p<0.05 (n=523) | log2FC>0.5 and p<0.05 (n=106) |
| TNFA | 0.0532 | 0.1480 | 0.1213 |
| Inflammatory Response | 0.0126 | 0.4029 | 0.3925 |
| IL2 Stat5 Signaling | 0.0126 | 0.2914 | 0.4315 |
| IL6 Jak Stat 5 Signaling | 0.4472 | 0.1686 | 0.4315 |

*CD177*

Citations

1. Hoffman, Amanda M., and Paul Cairns. “The Epigenetics of Kidney Cancer and Bladder Cancer.” Epigenomics 3.1 (2011): 19–34. PMC. Web. 9 May 2016.

2. Kluzek, Katarzyna, Hans Antonius Bluyssen, & Joanna Wesoly. "The epigenetic landscape of clear-cell renal cell carcinoma."Journal of Kidney Cancer and VHL [Online], 2.3 (2015): 90-104. Web. 9 May. 2016

3. Morris, Mark R, and Eamonn R Maher. “Epigenetics of Renal Cell Carcinoma: The Path towards New Diagnostics and Therapeutics.” Genome Medicine 2.9 (2010): 59. PMC. Web. 9 May 2016.

4. Reiner, S. L. (2005). Epigenetic control in the immune response. Human molecular genetics, 14, R41-6.

5. cBio Portal

6. Yang Z, Jones A, Widschwendter M, Teschendorff AE. An integrative pan-cancer-wide analysis of epigenetic enzymes reveals universal patterns of epigenomic deregulation in cancer. Genome Biol. 2015 Jul 14;16:140. doi: 10.1186/s13059-015-0699-9.

7. Synapse

8. Eser, S., Schnieke, A., Schneider, G., & Saur, D. (2014). Oncogenic KRAS signalling in pancreatic cancer. British Journal of Cancer, 111(5), 817–822. <http://doi.org/10.1038/bjc.2014.215>

9. Nouri M, Ratther E, Stylianou N, Nelson CC. Androgen-targeted therapy-induced epithelial mesenchymal plasticity and neuroendocrine transdifferentiation in prostate cancer: an opportunity for intervention. Front Oncol. 2014 Dec 23;4:370. doi: 10.3389/fonc.2014.00370. eCollection 2014.

10. lalala