**Investigated the list of 50 genes provided (Oncomine – top 10% amplified and top 10% over expressed in renal cancer)**

I ran the genes on my code, and imported results on to the Excel spreadsheet. (If there are genes missing, it is because cbioportal did not have any data relating the gene to any renal cancer studies.)

Many of the renal studies showed amplification of the genes in some of the cases; however, the significance was very low (for example, the percent of CNA (amplified) cases was around 1% of the total number of cases in the study).

I isolated cases that showed a more significant frequency of amplification, and realized that all of these cases were related to studies on Kidney Renal Clear Cell Carcinoma.

There were two studies on this particular form of cancer: Kidney Renal Clear Cell Carcinoma (TCGA, Nature 2013) and Kidney Renal Clear Cell Carcinoma (TCGA, Provisional). I compared the results to investigate similarities and attempt to justify this significance.

“Kidney Renal Clear Cell Carcinoma (TCGA, Nature 2013)” contains 436 cases that were published in the paper, whereas “Kidney Renal Clear Cell Carcinoma (TCGA, Provisional)” contains additional tumors that were considered after the publication. To further confirm that the amplification of a gene may have significance to Kidney cancer, I wanted to check if a linear trend existed – i.e. the number of copy number alterations (specifically amplification) would increase as the number of cases increased.

|  |  |  |
| --- | --- | --- |
| Gene | Nature 2013 (436 cases) | Provisional (528 cases) |
| CANX | 6.65% CNA  6.65% amplified | 16.48% CNA  16.48% amplified |
| FBXL17 | 0.69% CNA  0.69% amplified | 9.09% CNA  9.09% amplified |
| FNIP1 | 4.13% CNA  4.13% amplified | 14.77% CNA  14.77% amplified |
| HSPA4 | 4.59% CNA  4.59% amplified | 14.96% CNA  14.96% amplified |
| KDM3B | 4.82% CNA  4.82% amplified | 14.96% CNA  14.96% amplified |
| LARP1 | 4.82% CNA  4.82% amplified | 14.96% CNA  14.96% amplified |
| NPM1 | 6.19% CNA  6.19% amplified | 16.29% CNA  16.29% amplified |
| RARS | 5.73% CNA  5.73% amplified | 15.91% CNA  15.91% amplified |
| RNF14 | 5.05% CNA  5.05% amplified | 15.15% CNA  15.15% amplified |
| SRP19 | 1.15% CNA  1.15% amplified | 10.61% CNA  10.61% amplified |
| UBE2B | 4.36% CNA  4.36% amplified | 14.77% CNA  14.77% amplified |

From the table, we notice that as the number of cases increases, the number of amplifications also increases.

From these above genes, I checked to see where they are located in the human genome, and found that all of the genes are located on chromosome 5.

* This gives me confidence in my code as it helps me confirm that the data I am retrieving makes sense and is specific (not random).
* I further checked on a few facts about chromosome 5 and noted the following:
  + represents 6% of total DNA in cells
  + contains around 900 genes and about 66 genes are said to be involved in human disease if mutated

I noticed that many of the amplification percentages were very close; thus, I checked to see if they were positioned closely on the chromosome.

* HSPA4, KDM3B, FNIP1, UBE2B, and RNF14 were all located between 5q31.1 and 5q31.3:



* FBXL17 and SRP19 are located closely (21.3 and 22.2, respectively):



When checking P values, none of these genes showed P<0.005.