Intro-

* Clear cell renal cell carcinoma is most common form of kidney cancer
* Renal cell carcinoma is responsible for 13 000 deaths per year in US, clear cell renal cell carcinoma accounts for 80% of renal cell carcinoma
* 30% of surgically treated early stage ccrcc patients due to undetected additional metastases
* 5-year survival of metastatic ccrcc is 10%
* Limited known molecular targets
* Neuronal pentraxin-secreted protein expressed broadly in nervous, testicular, pancreatic, skeletal muscle, heart, involved in synaptic plasticity and neurite outgrowth.
* Binds AMPA glutamate receptors which form ion channels.
* In ccRCC, NPTX2 and AMPA receptor subunit GluR4 are overexpressed, and their interaction facilitates actin cytoskeleton remodelling and subsequent tumour cell migration. NPTX2 expression inhibition reduces tumour proliferation and invasion while increasing apoptosis (main).
* More recently elucidated that NPTX2 mediates tumour progression in prostate cancer (Fibroblast growth factor modulates mast cell recruitment in a murine model of prostate cancer) and one of its subunits, C-reactive protein, with tumours (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7461825/>). More recently elucidated that xenografting of tissue with circ\_0054537 (ceRNA for mir-640, which regulates NPTX2) silenced in mice returned clear cell renal cell carcinoma tumour size by downregulating NPTX2.
* Why do this-
* However, it is unlikely that this protein which the paper focused on is the only potential molecular target which could be found from the generated data. This hallmark microarray dataset should be easily accessible to the cancer genetics community without requiring downloading the dataset and redundant re-analysis. Making an accessible querying web tool also enables access by cancer geneticists who are not experienced in R.
* Usage in the field-Users can search for a gene of interest or gene ontology term and, from the results page, click on links to gene-specific page in order to access statistical significance and fold change information. From a gene page, they can also access related genes by clicking on the associated gene ontology terms, which returns them to the main search page with this term filled in.
* Issues-
* The dataset contains 54675 probes and 143 sample, which corresponds to 7.8 million rows in the expression table. It was initially found that each additional join increased the result return time nonlinearly, such that joining more than 4 tables was fairly impractical for a convenient web interface. Therefore, the database was de-normalized slightly from its original 6-table schema to a 4-table schema. Though this process likely increased database storage size, the gain in runtime efficiency was a worthwhile tradeoff. Even so, some search terms or movements from the search page to a gene page can sometimes approach a minute in runtime, which some impatient users might find undesirable. Though reduction of data size by, for example, removing probes with no significant results, would certainly eliminate this issue for the purposes of the assignment, such data is still informative and this step would almost certainly never be performed in practice. It is possible that use of another database style such as a nonrelational database (perhaps a key-value database with probes used as keys, reducing this access step to O(1)) or data warehouse might improve efficiency (AWS).

Chart

Description automatically generated with medium confidence

Figure 1. Clear cell Renal cell carcinoma microarray volcano plot with p-values<0.05 and fold changes>2 highlighted in purple. Neuronal Pentraxin 2 is colored in red.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | % | PValue | Fold Enrichment | Bonferroni | Benjamini | FDR |
| GO:0070062~extracellular exosome | 27.76204 | 3.87E-20 | 2.668511 | 1.23E-17 | 1.23E-17 | 1.16E-17 |
| CARBOHYD:N-linked (GlcNAc...) asparagine | 40.50992 | 6.66E-19 | 2.005805 | 1.07E-15 | 1.07E-15 | 1.06E-15 |
| KW-0325~Glycoprotein | 43.62606 | 3.41E-16 | 1.725197 | 9.33E-15 | 9.88E-15 | 8.52E-15 |
| GO:0016324~apical plasma membrane | 9.631728 | 3.03E-15 | 5.562387 | 9.53E-13 | 4.82E-13 | 4.53E-13 |
| KW-0732~Signal | 39.66006 | 1.41E-14 | 1.76155 | 3.10E-13 | 3.09E-13 | 2.67E-13 |
| GO:0016323~basolateral plasma membrane | 7.365439 | 1.61E-13 | 6.679042 | 5.13E-11 | 1.71E-11 | 1.61E-11 |
| GO:0005615~extracellular space | 21.52975 | 7.10E-13 | 2.407129 | 2.26E-10 | 5.65E-11 | 5.31E-11 |
| KW-0964~Secreted | 23.51275 | 4.79E-12 | 2.187707 | 1.82E-10 | 1.82E-10 | 1.63E-10 |
| GO:0005576~extracellular region | 22.3796 | 6.02E-12 | 2.253825 | 1.91E-09 | 3.83E-10 | 3.60E-10 |
| KW-1015~Disulfide bond | 34.27762 | 6.22E-11 | 1.694121 | 1.74E-09 | 9.02E-10 | 7.77E-10 |

|  |  |  |  |
| --- | --- | --- | --- |
| PROBEID | GENENAME | fold\_change | p\_value |
| 212143\_s\_at | insulin like growth factor binding protein 3 | 3.009468 | 2.20E-22 |
| 210095\_s\_at | insulin like growth factor binding protein 3 | 2.813106 | 2.36E-22 |
| 205934\_at | phospholipase C like 1 (inactive) | -2.53817 | 6.26E-21 |
| 202856\_s\_at | solute carrier family 16 member 3 | 3.168849 | 3.65E-20 |
| 202856\_s\_at | microRNA 6787 | 3.168849 | 3.65E-20 |
| 202934\_at | hexokinase 2 | 3.056619 | 4.86E-20 |
| 226452\_at | pyruvate dehydrogenase kinase 1 | 2.406077 | 6.45E-20 |
| 223333\_s\_at | angiopoietin like 4 | 3.375204 | 6.45E-20 |
| 201037\_at | phosphofructokinase, platelet | 2.188341 | 9.44E-20 |
| 222774\_s\_at | neuropilin and tolloid like 2 | 2.826774 | 1.93E-19 |