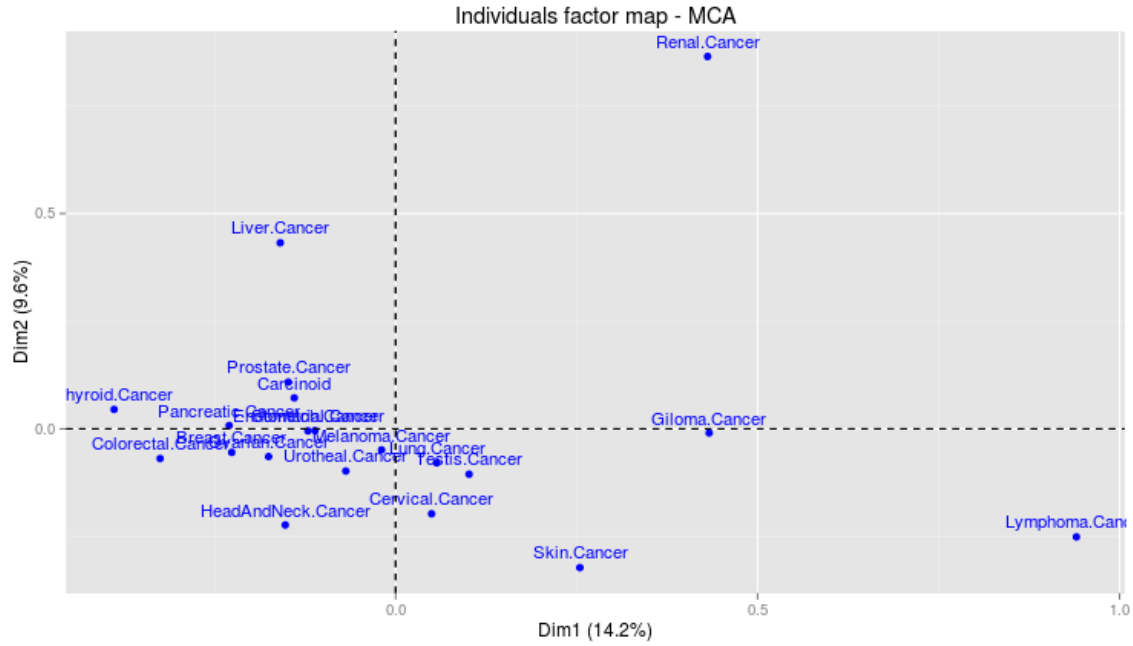
COMPARATIVE ANALYSIS ACROSS CANCER

Problem - What causes that cancer to be unique. How does one cancer vary from the other? What specific biological processes are altered in a cancer?

To answer the above problem we have to look at the basis of differentiation b/w cancers and why to focus on a specific group of cancers.

Before starting this semester I had left it at doing MCA of cancers using all genes.



The main purpose of the above was to see how are the cancers distributed. Whether I can see some separation. Clearly Renal stood separate from the rest. So did Lymphoma. Glioma and Liver also kind of look separate. This excited us and we thought that we should look at different BPS and then pinpoint the BPs responsible for the same.

So we took different BPs. We were expecting for atleast 1 BP the separation would not be there but it was always there. The same more or less repeated every time even questioning my code for the same. Infact I even took various other permutations at time taking intersection of genes of different BPs or unique to that particular BP but picture repeated.

I took genes randomly but to no improvement especially for renal. Then I took reduced set of genes from cell cycle but no improvement. I finally indulged in breaking down that set taking 5-10 genes at a time. I found from that only a few candidates were enough to cause the separation. Concluding that 1-2 genes were enough to separate renal in the MCA plot.

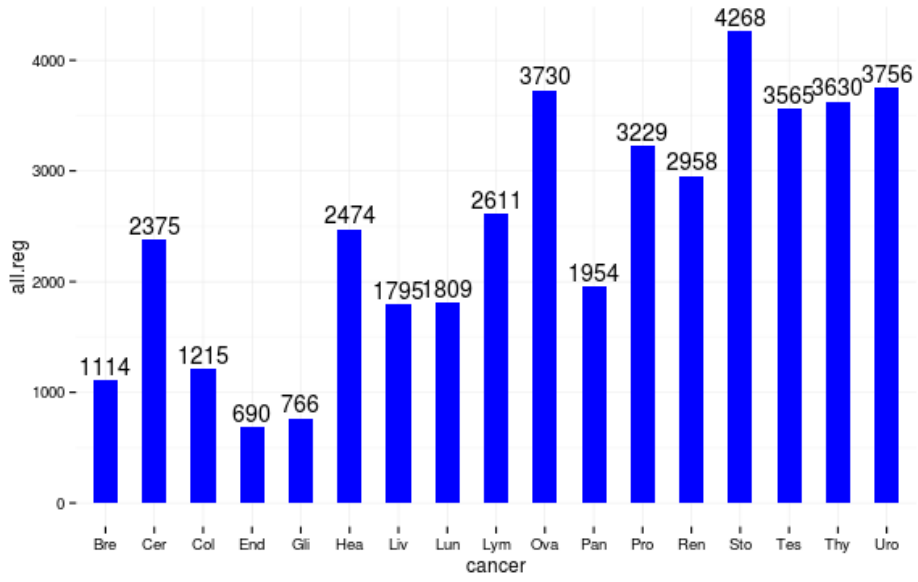
But again is the separation in the MCA worth considering. It separates based on the expression values and that could also happen because of physiology. By that I mean the separation which we see above could also happen because my normal tissues as such are different from each other. I took an intersection of DEGs of 4 cancers which appeared to be different everytime(Liver, Lymphoma, Renal,

Glioma). Though now the separation was not there much but we had only 45 DEGs. I also did MCA using DEGs of each of the above cancers and in them above lymophoma the conclusion remained same about separation but for lymphoma only lymphoma only remained separated.

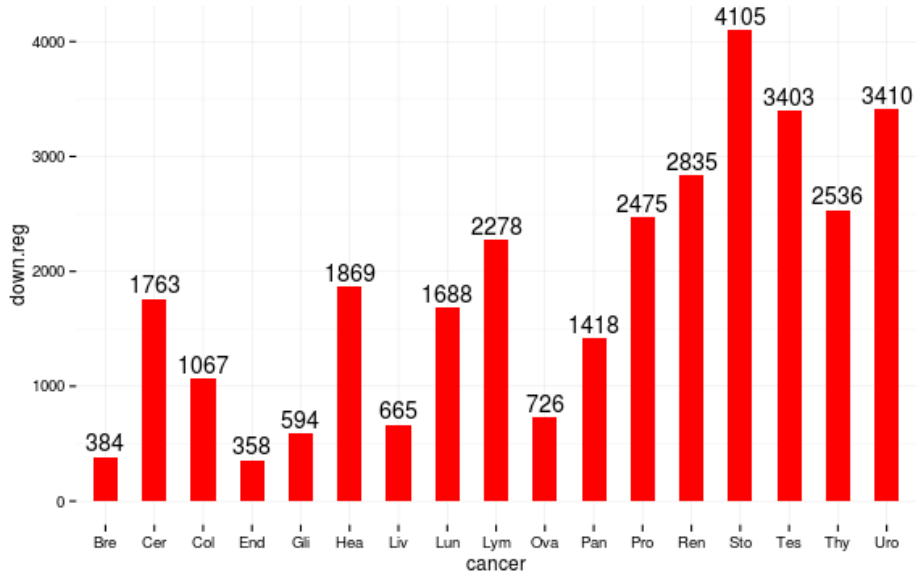
So in general not able to get a clear direction of where to go we went back to a top level looking again at all the cancers and looking numerically. Understanding the MCA plot had always been the region of our interest.

So given below are the images of total DEGs of each cancer.

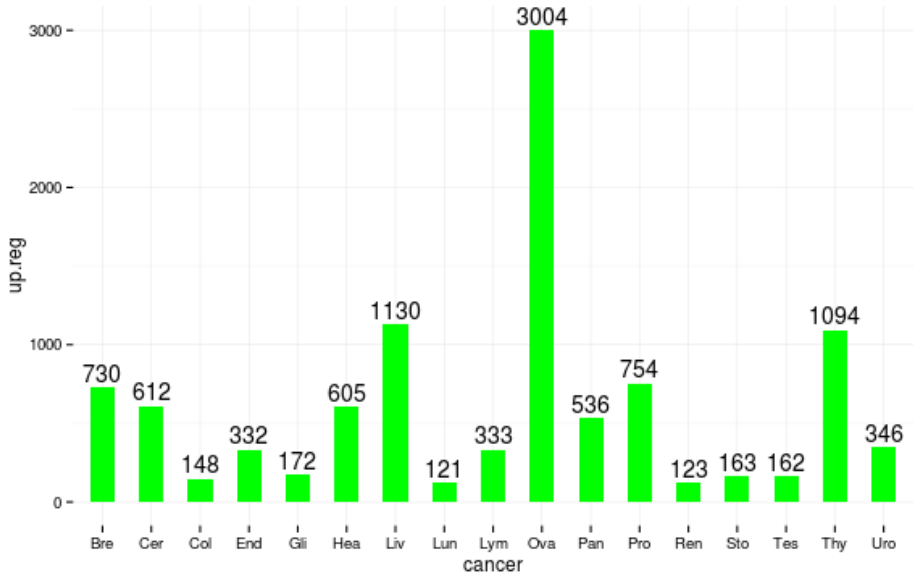
Total DEGs



Down regulated



Up regulated



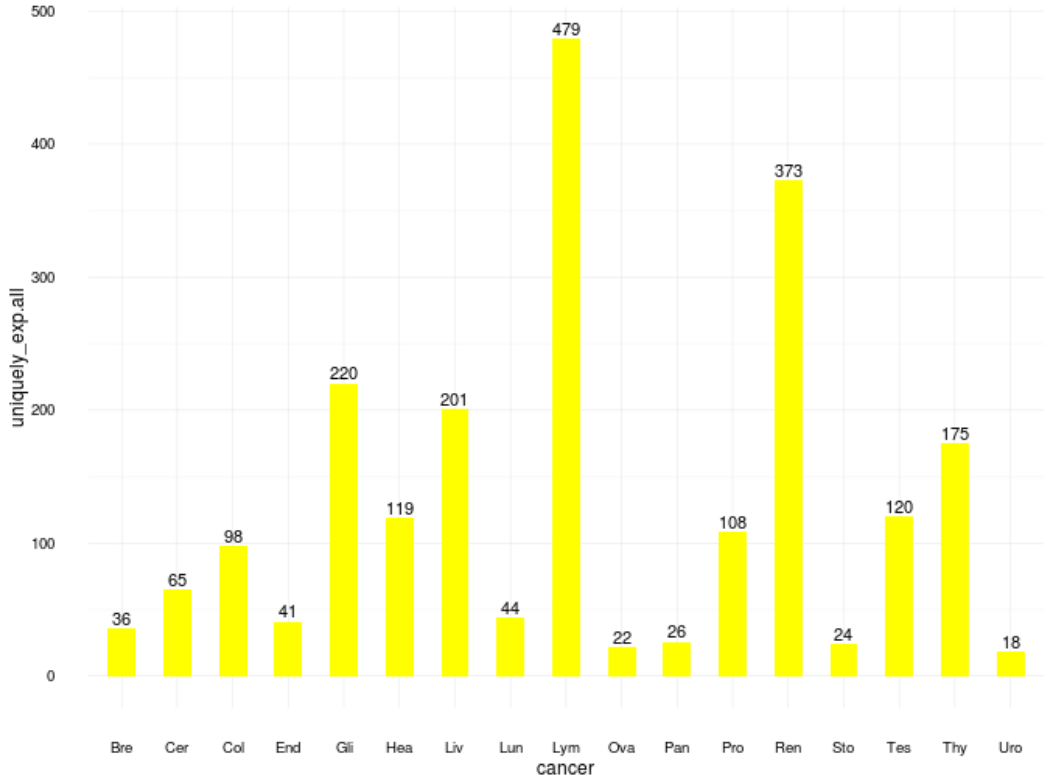
From above we see that Stomach, ovarian, urotheal, Testis have the most DEGs. On the other hand Endometrial, Glioma have the least DEGs.

Further we clearly see Stomach, Testis, Urotheal, Glioma, Endometrial, Colon, Renal, Lymphoma, Lung as predominantly downregulated cancers,

On the other hand Breast, Liver, Ovarian are upregulated. Infact except these all have greater number of DEGs which are downregulated.

One main focus is to understand the MCA plot. Note that it does not differentiate between pathology and physiology of cancers. Natural way of explaining at the number of genes which have the most unique expression level. What do I mean by this unique expression level? Before defining that I would call this set of genes as **uniquely_exp.**

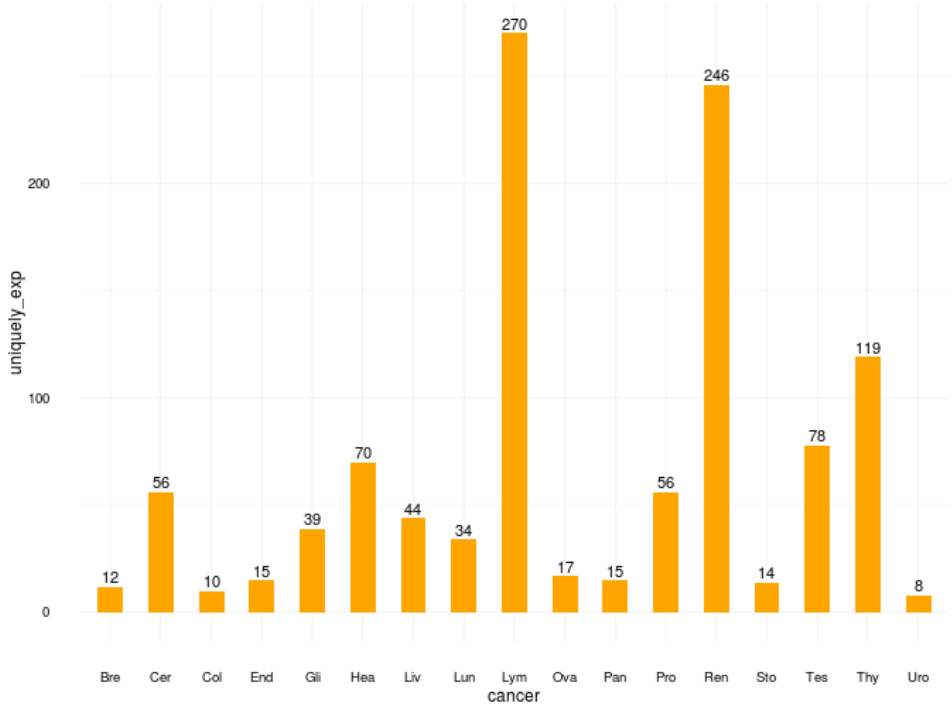
Lets begin with an example. We have a gene X and cancers A,B,C,…. . Now my gene X can have only 2 levels *P,N(Present and Not detected).* If my gene is *P* only across A and *N* across the rest of the cancers or *N* in A and *P* in rest of cancers than my gene X becomes the candidate for **uniquely_exp** for cancer A. Similarly we need to find such candidates for other cancers. In short we find those candidates which have a unique expression level for a cancer. The cancer with maximum of such candidates are likely to be separated the most. Note though we have taken only candidates which had one expression level for 1 cancer and a different expression level for other cancers. If we had a total of N cancers we could represent in the ratio 1:N-1. But the candidates which had a ratio such 2:N-2, 3:N-3 also will be good candidates for separating the cancers. Though we have not looked at them yet. In short though the cancer with the greatest number in 1:N-1 unique expression is definitely going to be separated the most unless and until we have an abnormal count for the other ratios. The picture below validates our statement.



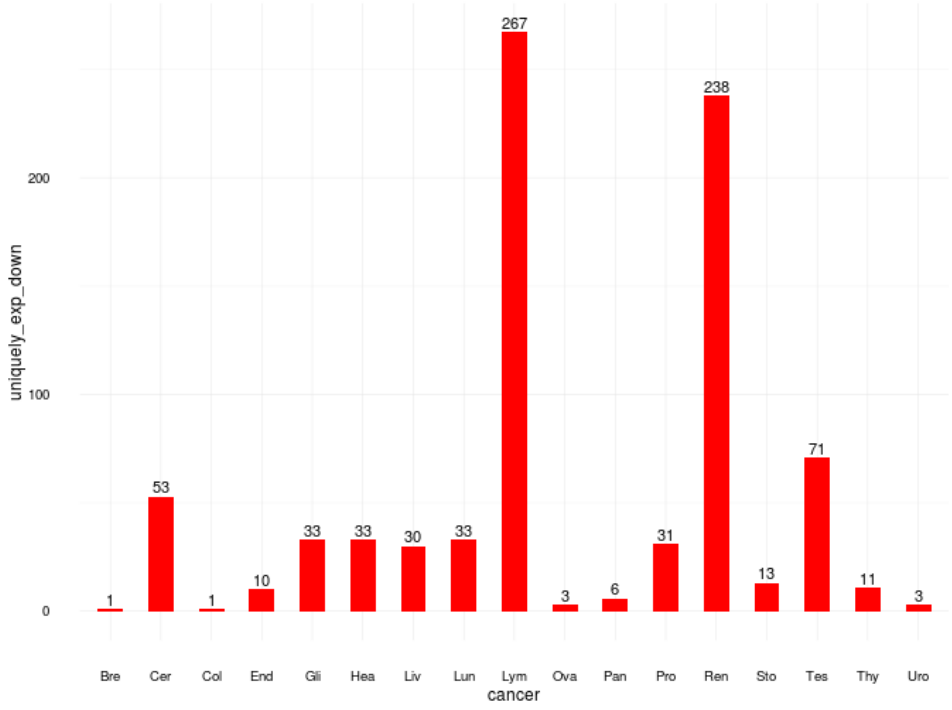
As seen from the diagram we see that renal and lymphoma which are clearly separated have the maximum number followed by glioma and liver which are also separated.

One thing though worth noting as such from above though we got some set of genes which are causing the cancer to separate those might not be a DEG. So out of the above if we take only DEGs than w.r.t to

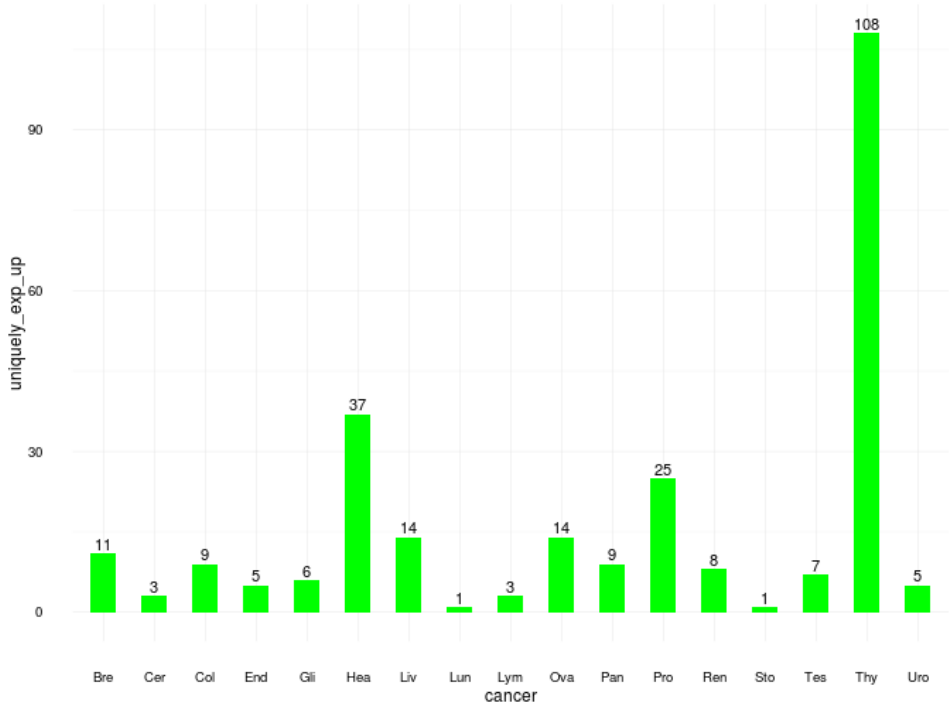
expression level we can make inferences that this particular gene has a unique expression for that cancer not observed in other cancers and also its pathology thus we can say for sure that this gene is a good candidate to look at. So taking only the DEGs from above we are left with this number.



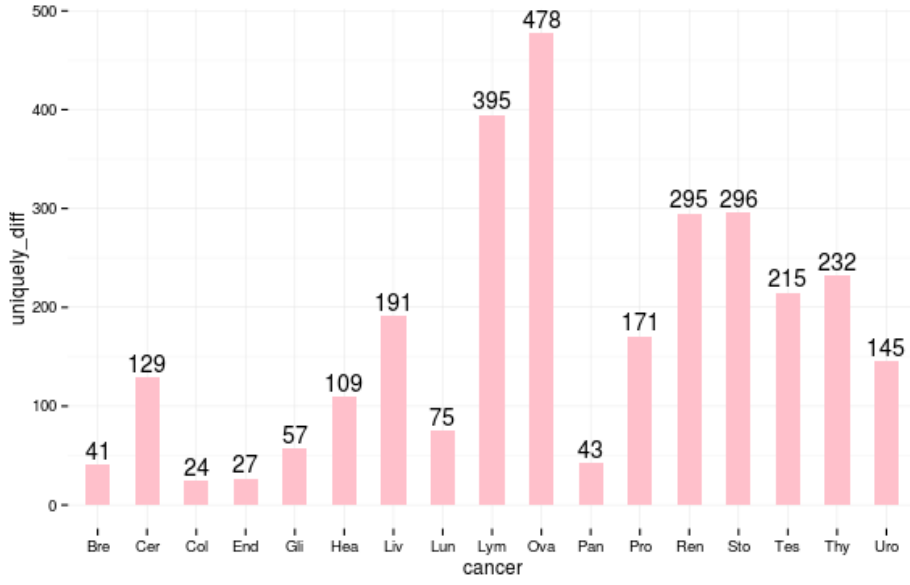
Out of these downregulated:



Upregulated:



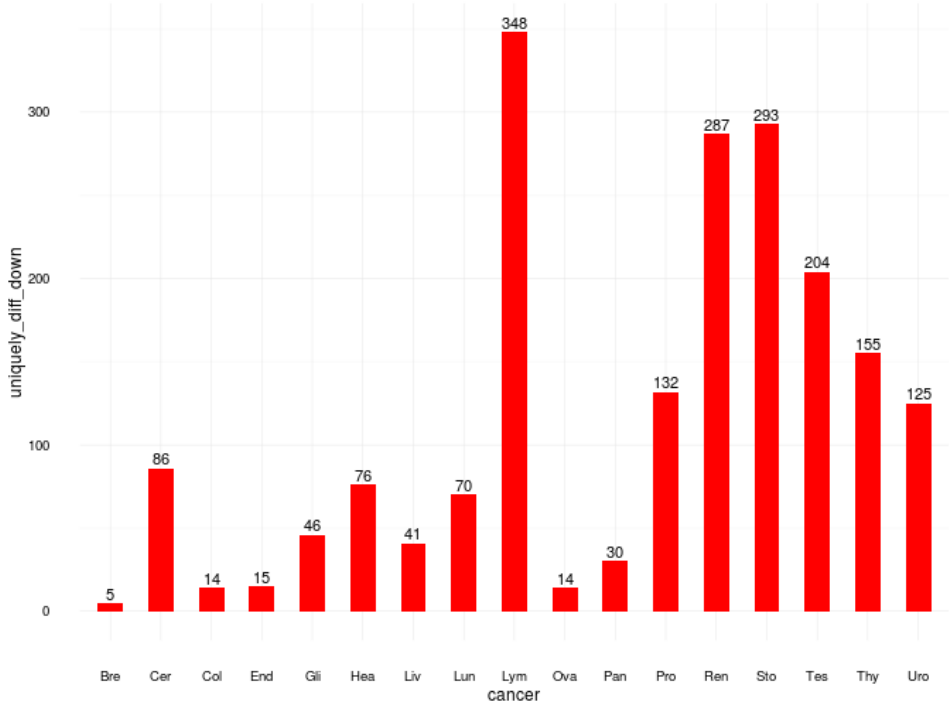
There is another set we take called uniquely differentiated genes. These are those genes which are differentiated only across 1 cancer. Though it might at first seem that we should use these to explain our MCA plot because they are unique but mind you that MCA plot is just based on expression level values and a gene X which though is uniquely differentiated in a cancer A could be lets say have level *P* in A and also *P* in B but not a DEG w.r.t to B because of physiology of that particular cancer B. Thus causing no separation in the MCA plot b/w A and B. Thus if we want to study cancers considering these genes are equally important for our study. Below is the figure of uniquely differentiated genes.



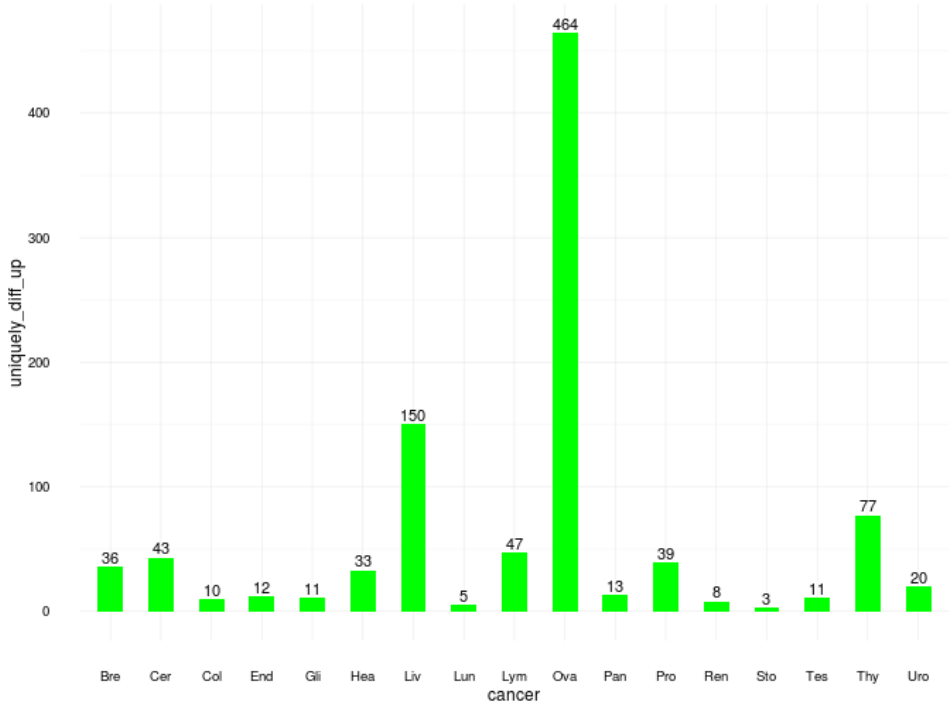
Now looking at above though Renal and Lymphoma still have a considerable amount ovarian cancer somehow dominates which was not very apparent from the MCA plot. The conclusion that could be drawn is that the normal sample of ovarian is very separated from the rest of the majority in terms of gene expression. One thing worth mentioning is that Ovarian also had a very large number of DEGs. Similar inferences could be made for stomach and thyroid.

Now looking at overall nature of the above:

Downregulated



Upregulated



In addition to the inferences made above we can see that these cancers follow the overall pattern. Renal, Lymphoma, Stomach were down regulated cancers and hence thus have predominantly down regulated genes. Liver and Ovarian were upregulated and thus have predominantly upregulated genes.

Without doing any further analysis we can clearly see Lymphoma and Renal cancer as our candidates becaue of the vast number of DEGs they have and the clear separation in the MCA plot. The MCA on normal would make the picture more clear but since each has a further cell type doing MCA on them is tough.

Changes

So each set conveys an important message. The first **unique_exp** tells us that these genes are uniquely expressed meaning that the corresponding processes being contributed to by those genes are kind of uniquely altered w.r.t expression in my cancer. The second **unique_diff** tells us that these genes are uniquely differentiated in my cancer meaning that these are the processes that are sort of new to my cancer such that they being regulated affect only my that particular cancer in some manner.

Though the unique expression level values expresses the MCA plots they do bring to the table something unique but what unique is to be pondered. If a gene is a unique exp but not a unique diff that means that particular gene is present is a DEG in some other cancer also altering a similar pathway though definitely differently.

For the unique_exp we take only DEGs in it. But in general if it is not a DEG then it means that particular tissues physiologically is different from the rest. On the other hand the second(unique_exp) if not unique_diff points to a particular process then that process though even pointed by some other cancers will have a expression for that cancer that is unique for that cancer only. What I mean is if a gene X is a DEG in cancers A, B, C and let's say that it is uniquely_exp in A. Also lets say a gene X affects a process P

then for A it will have an expression that is opposite to that of B, C and for the matter any other cancer. But that would also mean that A and B, C were different physiologically and also that A and rest cancers(except B, C) are same physiologically and B,C are different from rest. The intersection of unique_diff and unique_exp gives us those genes that were same physiologically across all the tissues but are a DEG for my particular tissue.