1) NSAIDs: The last update was on November and there are currently 92 signatures from the list of NSAID shown in the spreadsheet. Attached file.

2)

ChEA
MCF10A
MCF7

I get pubmed alerts for them every Monday. So the spreadsheets are pretty much up to date. I'm just waiting for the next update. For these 3 you had mentioned stopping at what we had for now and use what we already have for the analysis.

Analysis: For MCF10A and MCF7 there are **histograms** for distribution per platform and **PCAs** comparing the GSEs in the 2 platforms with the most entries. Those files are attached.

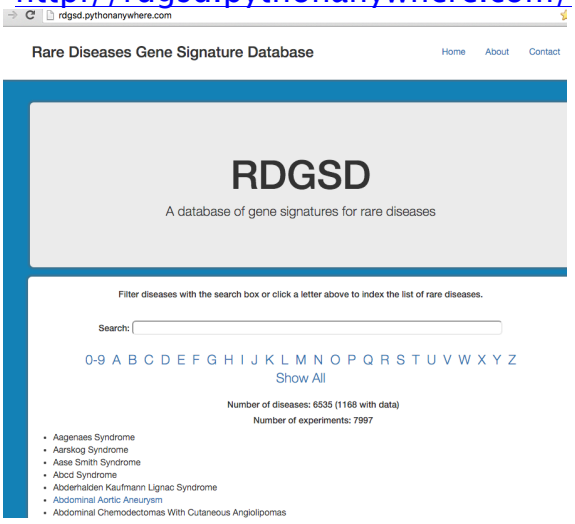
3) Fibroblast_DP: The spreadsheet contains the search results up to December 1.

I drafted a Power Point of it. Please review it and tell me what I should expand/include/take out. Then I'll use your feedback to finish this presentation and make the other ones accordingly.

- The spreadsheet has all the data organized and the 3 most common perturbations found on the mimics are highlighted.

4) RDGS – RARE DISEASES GENE SIGNATURE DATABASE: This was one of the projects I started before coming here. And since the list of tags of interest also includes disease I thought it might be useful to do a little summary about what we already have. Excel file with data and explanations is attached.

http://rdgsd.pythonanywhere.com/



5) Dexamethasone: Was last updated on December 9. Currently have 68 signatures. I used **L1000CDS²** to get the mimics from the G2E gene list. Dexamethasone showed up in 27 and Ketorolac in 15 of the 68 tagged signatures.

Next:
- I have no idea how to analyze this data. My current plan is to extract more signatures and try to get at least 100 before continuing to try and do some sort of analysis.

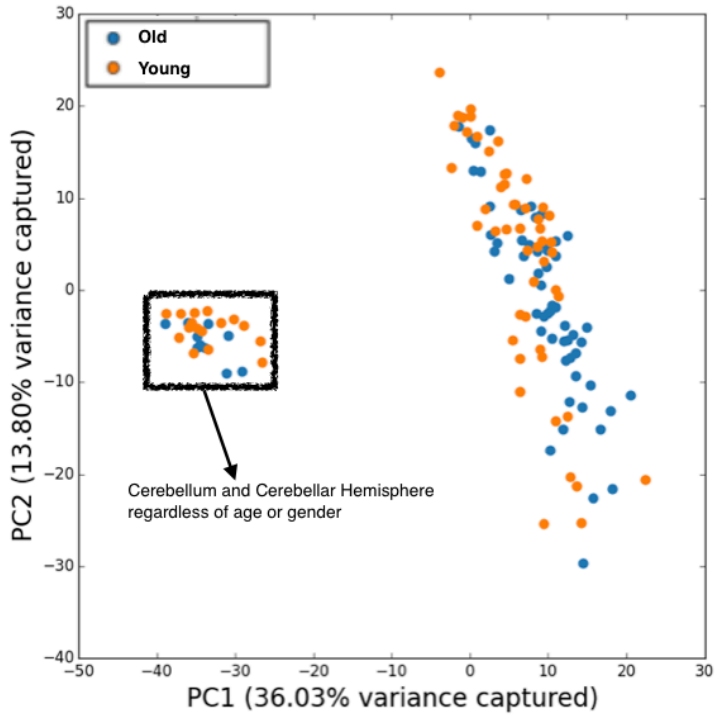
6) Other spreadsheets - There are also gene signatures for;
- Wolfram Syndrome (all available and supported until July 2015)
- Charcot-Marie-Tooth (all available and supported until July 2015)
- Schizophrenia (not all available)
- Huntington's (not all available)

No analysis was done in these. The signatures still have to be uploaded to G2E and Gen3va.

7) GTEx: First I tried using all the Brain samples for 20-39 (young) and 60-79 (old). Ignoring tissue region and gender.

Filtered out the genes that are not detected across all samples (rows that summed to 0) and those for genes that were present in 95% of the samples.taking the the data and normalized it with log 10.

Then I got the following PCA.



To do **NEXT:**
For Brain there were 8 unique young donors (20 to 39 years old) with 60 samples and 105

unique old donors with 641 samples. My plan now is to take the 5 youngest (20-29) and 5 oldest (70-79) and use the RNA-seq data for all the brain regions for those 10 donors.

- Filter by 95%, do a log 10 normalization and plot the 3 and 4th component.

8) GEO Update: I changed the update to monthly since there were not so many entries monthly and usually the first 2 weeks of the month have more than 70% of the entries.

- The script to get the data, clean it and send the email is done.

To do **NEXT:**

1) Make email look nicer (this is proving to be a bit difficult).
2) Add its entries to specific tags (crowdsourcing, drugs, etc)

9) GENES-to-DATES: I noticed that some of the lab data is "excel contaminated", some lists have gene names that are only numbers and if the spreadsheets are not formatted the data ends up with dates (Mar-3, Dec-1, etc) instead of gene numbers. I first noticed this with the ChEA GMT file, so it might be useful to have a list to check the data against and just get a script to substitute the dates for genes.

10) Calendar: I tried to come up with dates for updates for the tags of interest.

Tags

MCF7
MCF10A
ALS
Huntington's
Alcoholic Liver
Aging
Drugs
Genes
Kinases
Diseases
IFN-alpha mimics
iPSC reprogramming
Wnt mimics
Viruses and bacteria
Melanoma
Ketamine
Diabetic nephropathy
Dexamethasone
Hair cells
NSAIDS