

LabBook_24_03_2017

Claire Green

Monday

I spent the day looking over the pathprint methodology. I realised that the files I was using were the ones that John had processed, and I thought I should probably do my own processing with the same files I used for the DEG analysis.

I took the files and ran them through the pathprint script and of course the results were completely different. I thought that they looked a little weird, so I was messing with the threshold when I realised I didn't know what the threshold actually meant. I looked at the source code for diffPathways and it looked like all it was doing was taking the mean of the patients and the mean of controls and then taking the absolute difference. Since I wasn't taking a threshold based on a metric for the DEGs, I figured maybe I should do the same for the pathways.

So, I adapted the diffPathways code to give a table of the absolute differences between average patient and control pathprint. Then I ranked them by this score, and took the top threshold (which at the moment is 200). What I needed to know however is how well I was doing - I wanted to apply the same process to the pathways as I did the genes. What I would need is a way of seeing how well the benchmark list enriched in the pathways available, but extracting the gene lists of a pathway is impossible to automate. I asked Sokratis whether a file existed with all the pathway names and their corresponding gene lists, but apparently that doesn't exist and it does a query between multiple databases every time. Until that happens I can either do it by hand (not many pathways) or wait.

These are the pathways:

Vitamin B12 Metabolism (Wikipathways) Matrix Metalloproteinases (Wikipathways) Irinotecan Pathway (Wikipathways) Complement and Coagulation Cascades (Wikipathways) Complement and coagulation cascades (KEGG) {HSP90AA1,18} (Static Module) Phase I, non P450 (Wikipathways) Aflatoxin B1 metabolism (Wikipathways) cytochrome P450 (Wikipathways) Fructose and mannose metabolism (KEGG) Pentose and glucuronate interconversions (KEGG) Ascorbate and aldarate metabolism (KEGG) Blood Clotting Cascade (Wikipathways) Oxidative phosphorylation (Wikipathways) Type II diabetes mellitus (Wikipathways)

Enrichment is:

Pathway Name	P-Value	Adj. P-Value	Overlap
Vitamin B12 Metabolism (Wikipathways)	3.37E-05	5.10E-04	9/52
Matrix Metalloproteinases (Wikipathways)	1	1	0
Irinotecan Pathway (Wikipathways)	0.093	0.23	1/13
Complement and Coagulation Cascades (Wikipathways)	9.90E-04	0.005	7/52
Complement and coagulation cascades (KEGG)	6.08E-03	0.023	7/69
{HSP90AA1,18} (Static Module)	1	1	0
Phase I, non P450 (Wikipathways)	8.02E-05	0.0006	3/7
Aflatoxin B1 metabolism (Wikipathways)	1	1	0
cytochrome P450 (Wikipathways)	0.44	0.66	2/61
Fructose and mannose metabolism (KEGG)	0.42	0.66	1/36
Pentose and glucuronate interconversions (KEGG)	1	1	0
Ascorbate and aldarate metabolism (KEGG)	1	1	0
Blood Clotting Cascade (Wikipathways)	0.25	0.47	1/24
Oxidative phosphorylation (Wikipathways)	0.22	0.47	3/60
Type II diabetes mellitus (Wikipathways)	0.056	0.17	2/22