MODULE 8 - PART 2: PATHWAY ANALYSIS LAB

Lab Part 1: Enrichment Test

1.1. Let's get a gene list for analysis

ARTICLE

OPEN

doi:10.1038/nature12634

Mutational landscape and significance across 12 major cancer types

Cyriac Kandoth[1]*, Michael D. McLellan[1]*, Fabio Vandin[2], Kai Ye[1,3], Beifang Niu[1], Charles Lu[1], Mingchao Xie[1], Qunyuan Zhang[1,3], Joshua F. McMichael[1], Matthew A. Wyczalkowski[1], Mark D. M. Leiserson[2], Christopher A. Miller[1], John S. Welch[4,5], Matthew J. Walter[4,5], Michael C. Wendl[1,3,6], Timothy J. Ley[1,3,4,5], Richard K. Wilson[1,3,5], Benjamin J. Raphael[2] & Li Ding[1,3,4,5]

The Cancer Genome Atlas (TCGA) has used the latest sequencing and analysis methods to identify somatic variants across thousands of tumours. Here we present data and analytical results for point mutations and small insertions/deletions from 3,281 tumours across 12 tumour types as part of the TCGA Pan-Cancer effort. We illustrate the distributions of mutation frequencies, types and contexts across tumour types, and establish their links to tissues of origin, environmental/carcinogen influences, and DNA repair defects. Using the integrated data sets, we identified 127 significantly mutated genes from well-known (for example, mitogen-activated protein kinase, phosphatidylinositol-3-OH kinase, Wnt/β-catenin and receptor tyrosine kinase signalling pathways, and cell cycle control) and emerging (for example, histone, histone modification, splicing, metabolism and proteolysis) cellular processes in cancer. The average number of mutations in these significantly mutated genes varies across tumour types; most tumours have two to six, indicating that the number of driver mutations required during oncogenesis is relatively small. Mutations in transcriptional factors/regulators show tissue specificity, whereas histone modifiers are often mutated across several cancer types. Clinical association analysis identifies genes having a significant effect on survival, and investigations of mutations with respect to clonal/subclonal architecture delineate their temporal orders during tumorigenesis. Taken together, these results lay the groundwork for developing new diagnostics and individualizing cancer treatment.

Highlights:

- Using the integrated data sets, we identified 127 significantly mutated genes
- Genes under positive selection, either in individual or multiple tumour types,
- tend to display higher mutation frequencies above background.
- Our statistical analysis identified 127 such genes
- The mutational significance in cancer (MuSiC) package was used to identify
- significant genes for both individual tumour types and the Pan-Cancer collective. [Dees et al. MuSiC: Identifying mutational significance in cancer genomes. Genome Res. 2012]
- These significantly mutated genes are involved in a wide range of cellular processes, including transcription factors/regulators, histone modifiers, genome integrity, receptor tyrosine kinase signalling, cell cycle, mitogen-activated protein kinases (MAPK) signalling, phosphatidylinositol-3-OH kinase (PI(3)K) signalling, Wnt/b-catenin signalling, histones, ubiquitin-mediated proteolysis, and splicing (Fig. 2).

Supplementary Data, Table 4

- globally significant, frequency >= 1% for glioblastoma multiforme (GBM): 46

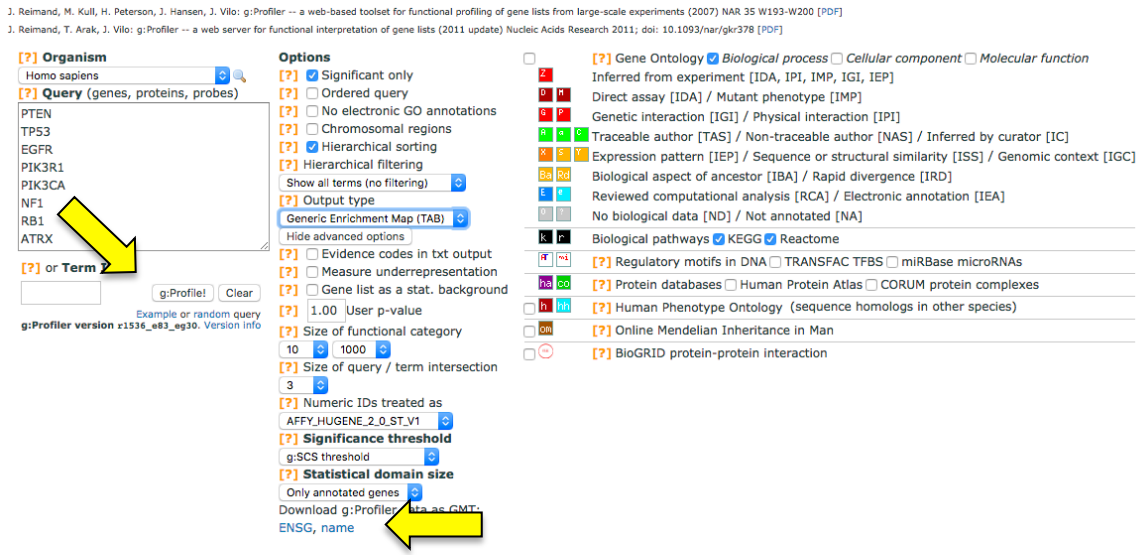
- globally significant, frequency >= 1% for kidney renal clear cell carcinoma (KIRC): 53

1.2. Let's use g:Profiler to obtain enrichment results

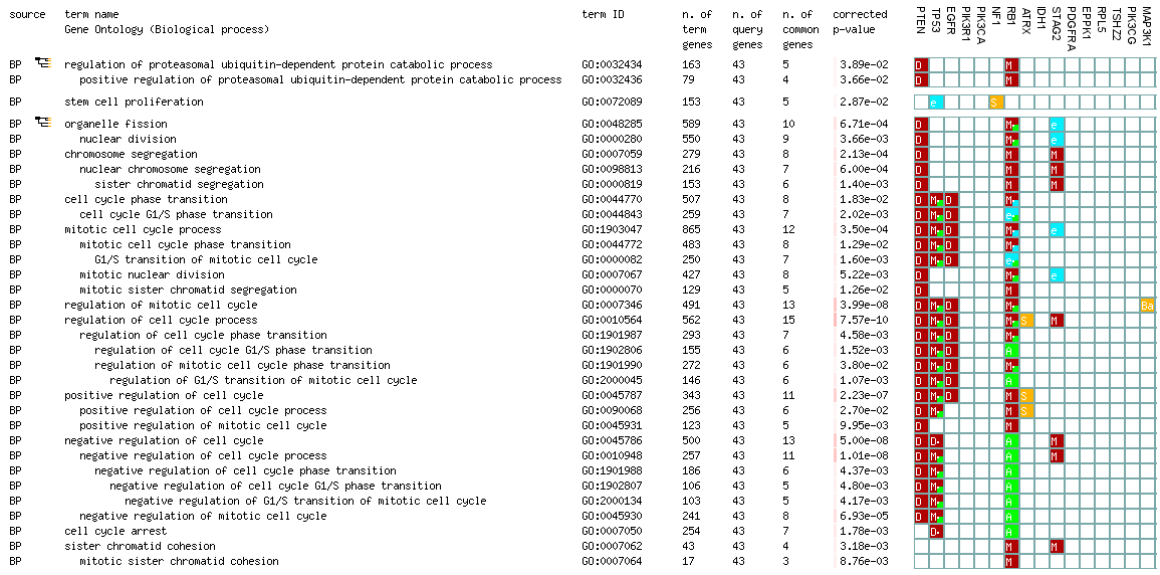
http://biit.cs.ut.ee/gprofiler/

First set the parameters and filter gene sets to be analysed

Then paste in the gene list and press g:Profile to perform the analysis; also download the GMT file with gene symbols, which will be necessary to use Enrichment Map for visualization



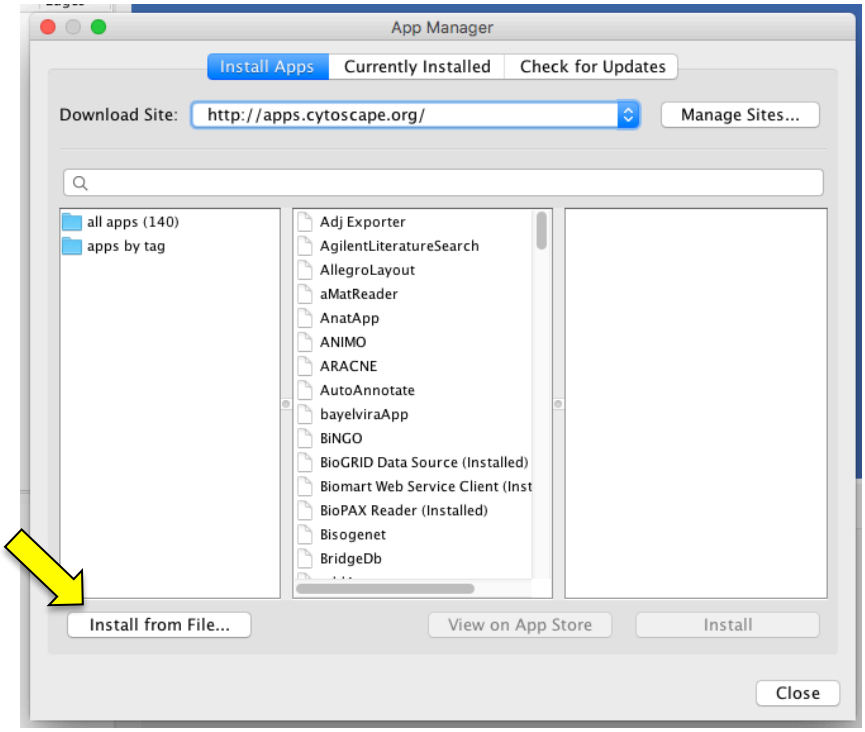
Change Output Type back to Graphical (PNG) and study visualised pathway enrichments. Scroll right and down to see processes and annotated genes.



source	term name Gene Ontology (Biological process)	term ID	n. of term genes	n. of query genes	n. of common genes	corrected p-value
BP	regulation of proteasomal ubiquitin-dependent protein catabolic process	GO:0032434	163	43	5	3.89e-02
BP	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	GO:0032436	79	43	4	3.66e-02
BP	stem cell proliferation	GO:0072089	153	43	5	2.87e-02
BP	organelle fission	GO:0048285	589	43	10	6.71e-04
BP	nuclear division	GO:0000280	550	43	9	3.66e-03
BP	chromosome segregation	GO:0007059	279	43	8	2.13e-04
BP	nuclear chromosome segregation	GO:0098813	216	43	7	6.00e-04
BP	sister chromatid segregation	GO:0000819	153	43	6	1.40e-03
BP	cell cycle phase transition	GO:0044770	507	43	8	1.83e-02
BP	cell cycle G1/S phase transition	GO:0044843	259	43	7	2.02e-03
BP	mitotic cell cycle process	GO:1903047	865	43	12	3.50e-04
BP	mitotic cell cycle phase transition	GO:0044772	483	43	8	1.29e-02
BP	G1/S transition of mitotic cell cycle	GO:0000082	250	43	7	1.60e-03
BP	mitotic nuclear division	GO:0007067	427	43	8	5.22e-03
BP	mitotic sister chromatid segregation	GO:0000070	129	43	5	1.26e-02
BP	regulation of mitotic cell cycle	GO:0007346	491	43	13	3.99e-08
BP	regulation of cell cycle process	GO:0010564	562	43	15	7.57e-10
BP	regulation of cell cycle phase transition	GO:1901987	293	43	7	4.58e-03
BP	regulation of cell cycle G1/S phase transition	GO:1902806	155	43	6	1.52e-03
BP	regulation of mitotic cell cycle phase transition	GO:1901990	272	43	6	3.80e-02
BP	regulation of G1/S transition of mitotic cell cycle	GO:2000045	146	43	6	1.07e-03
BP	positive regulation of cell cycle	GO:0045787	343	43	11	2.23e-07
BP	positive regulation of cell cycle process	GO:0090068	256	43	6	2.70e-02
BP	positive regulation of mitotic cell cycle	GO:0045931	123	43	5	9.95e-03
BP	negative regulation of cell cycle	GO:0045786	500	43	13	5.00e-08
BP	negative regulation of cell cycle process	GO:0010948	257	43	11	1.01e-08
BP	negative regulation of cell cycle phase transition	GO:1901988	186	43	6	4.37e-03
BP	negative regulation of cell cycle G1/S phase transition	GO:1902807	106	43	5	4.80e-03
BP	negative regulation of G1/S transition of mitotic cell cycle	GO:2000134	103	43	5	4.17e-03
BP	negative regulation of mitotic cell cycle	GO:0045930	241	43	8	6.93e-05
BP	cell cycle arrest	GO:0007050	254	43	7	1.78e-03
BP	sister chromatid cohesion	GO:0007062	43	43	4	3.18e-03
BP	mitotic sister chromatid cohesion	GO:0007064	17	43	3	8.76e-03

Lab Part 2: Visualization in Enrichment Map

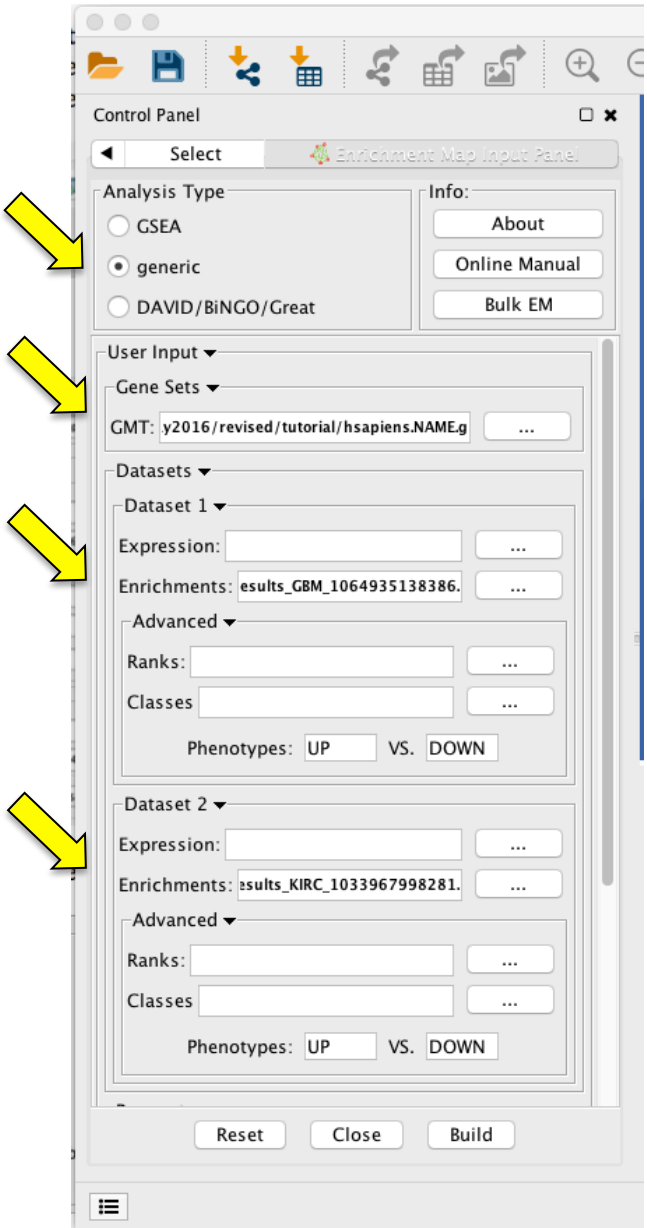
Enrichment Map 2.0.1 needs to be installed prior to analysis (the most recent version 2.1.0 creates flawed maps for datasets generated with g:Profiler). File enrichmentmap-2.0.1.jar is available in the wiki. See Apps > App Manager in menu.



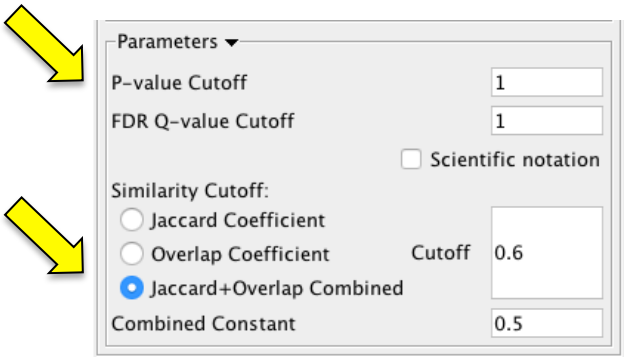
Open Cytoscape 3.2.1, Apps > Enrichment Map > Create enrichment map

First load all the files:

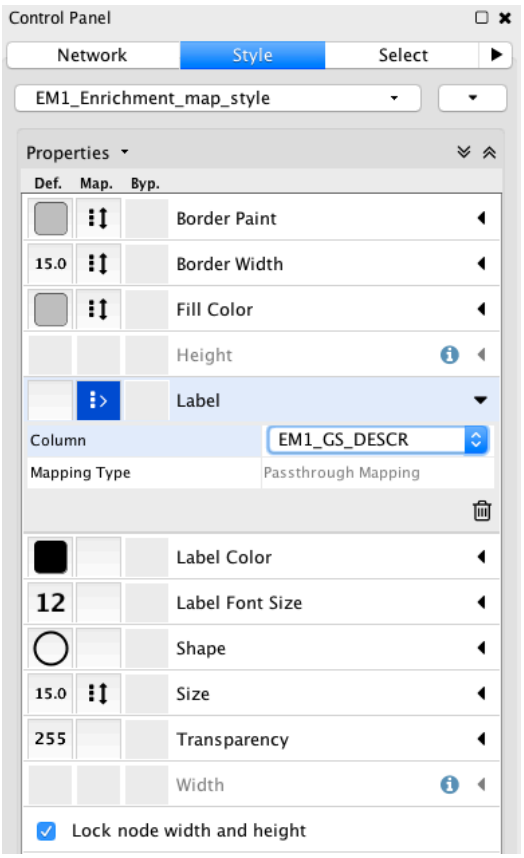
- GMT file: has gene-set definitions
- Enrichment of data-set 1: has enrichment statistics for GBM
- Enrichment of data-set 2: has enrichment statistics for KIRC



Then set the analysis parameters



From the Cytoscape control panel, select the style tab, and map the EM1_GS_DESCR to graphic attribute Label



Then, for both data-sets, reset the color based on FDR.

First, change the attribute mapped to the graphical attribute:

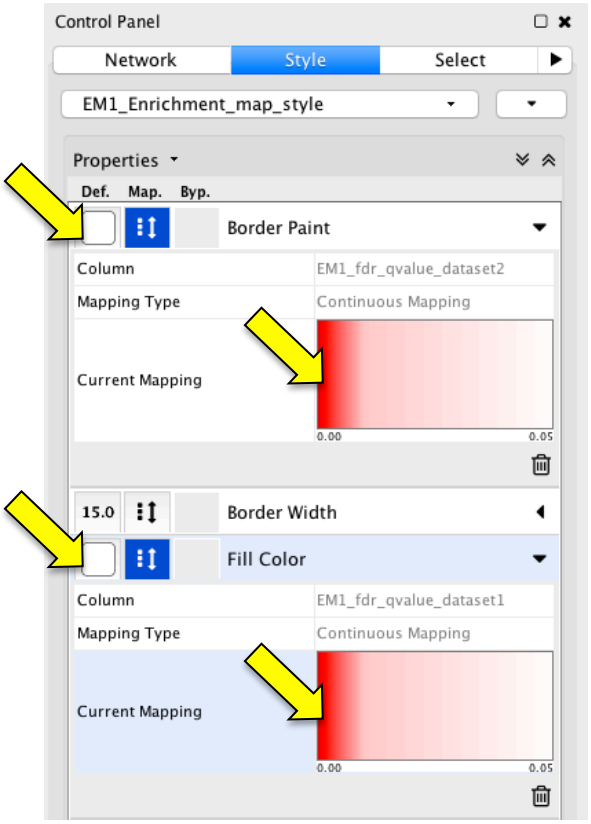
- fill color: EM1_fdr_qvalue_dataset1
- border paint: EM1_fdr_qvalue_dataset2

Second, set the color notches to these values:

- 0.0001: dark red
- 0.001: red

- 0.005: lighter red
- 0.01: lightest red
- 0.05: white

Third, set the default colors of nodes and borders to white.



Finally, we are going to improve the layout.

1. From: Layout > Settings
2. Select: Prefuse Force Directed Layout
3. And reset default spring coefficient to: 1E-6

Tips for interpreting results for these data-sets:

- Entirely red circles represent pathways enriched both in GBM and KIRC
- Red cores with white borders represent pathways only seen in GBM
- White cores with red borders represent pathways only seen in KIRC
- Intensity of red maps to strength of pathway enrichment p-value
- Singletons at the bottom are often redundant with larger clusters; however they sometimes include additional, unique pathways.
- Browse groups of pathways and identify major functional themes characteristic of these groups.