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

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/ $\beta$ –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, mitogenactivated 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

Reimand, M. Kull, H. Peterson, J. Hansen, J. Vilo: g:Prof	iler a web-based toolset for functional profiling of g	ene lists from la	rge-scale experiments (2007) NAR 35 W193-W200 [PDF]		
Reimand, T. Arak, J. Vilo: g:Profiler a web server for	unctional interpretation of gene lists (2011 update) N	ucleic Acids Res	earch 2011; doi: 10.1093/nar/gkr378 [PDF]		
[?] Organism	Options	7	[?] Gene Ontology & Biological process Cellular component Molecular function		
[?] Query (genes, proteins, probes)	[?] Ordered query	DH	Direct assay [IDA] / Mutant phenotype [IMP]		
PTEN TP53 EGFR PIK3R1 PIK3CA NF1 RB1 ATDY	<ul> <li>(?) No electronic GO annotations</li> <li>(?) Chromosomal regions</li> <li>(?) Hierarchical string</li> <li>(?) Hierarchical filtering</li> <li>Show all terms (no filtering)</li> <li>(?) Output type</li> <li>Generic Enrichment Map (TAB)</li> <li>Hide advanced options</li> </ul>	G P A a C X S Y Ba Rd E 0 7	Genetic interaction [IGI] / Physical interaction [IPI] Traceable author [TAS] / Non-traceable author [NAS] / Inferred by curator [IC] Expression pattern [IEP] / Sequence or structural similarity [ISS] / Genomic context [IGI Biological aspect of ancestor [IBA] / Rapid divergence [IRD] Reviewed computational analysis [RCA] / Electronic annotation [IEA] No biological pathways @ KEGG @ Reactome Biological pathways @ KEGG @ Reactome		
[?] or Term g:Profile: Clear Example or random query g:Profiler version x1536_e83_eg30. Version info	[?]       Evidence codes in txt output         [?]       Measure underrepresentation         [?]       Gene list as a stat. background         [?]       Iono         [?]       Size of functional category         [0]       5 1000         [?]       Size of query / term intersection         [?]       Size of query / term intersection         [?]       Size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of size of query / term intersection         [?]       Size of query / term intersection         [?]		Protein databases       Human Protein Atlas       CRUM protein complexes         [?] Human Phenotype Ontology (sequence homologs in other species)       [?] Online Mendelian Inheritance in Man         [?] BioGRID protein-protein interaction		

Change <u>Output Type</u> back to <u>Graphical (PNG)</u> 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	MAP3K1 PH/SCG TSH/Z RPL5 EPPK1 EPPK1 PDGFRA STAG2 IDH1 ATR2 ATR2 ATR2 ATR2 RB1 ATR2 RB1 ATR2 RB1 ATR2 RB1 ATR2 RB1 PH/SCA FR5 RB1 ATR2 ATR2 RB1 ATR2 RB1 ATR2 RB1 ATR2 ATR2 ATR2 ATR2 ATR2 ATR2 ATR2 ATR2
BP ⁼≧ BP	regulation of proteasomal ubiquitin-dependent protein catabolic process positive regulation of proteasomal ubiquitin-dependent protein catabolic process	GD:0032434 GD:0032436	163 79	43 43	5 4	3.89e-02 3.66e-02	
BP	stem cell proliferation	GD:0072089	153	43	5	2.87e-02	
67* 14 67 14 6	stem cell proliferation muclear division muclear division chromosome segregation sister chromosome segregation cell cycle forces mitotic cell cycle phase transition division cell cycle phase transition division cell cycle phase transition division cell cycle phase transition mitotic cell cycle phase transition division of mitotic cell cycle mitotic mitotic nuclear division regulation of cell cycle phase transition regulation of fuctor cell cycle phase transition regulation of fuctor cell cycle phase transition regulation of cell cycle phase transition regative regulation of cell cycle phase transition negative regulation of mitotic cell cycle regulation of mitotic cell cycle regulation of mitotic cell cycle	00:0072609 C0:004285 C0:0000280 C0:000280 C0:0002813 C0:0000819 C0:0004813 C0:0000819 C0:0004813 C0:0004813 C0:0004813 C0:0004813 C0:000706 C0:000706 C0:000706 C0:000706 C0:000706 C0:000707 C0:000706 C0:000707 C0:00007 C0:000707 C0:000707 C0:00007 C0:000707 C0:000707 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00007 C0:00000 C0:00007 C0:00007 C0:00000 C0:00000 C0:00000 C0:00000 C0:00000 C0:00000 C0:000000 C0:000000 C0:000000 C0:000000 C0:000000 C0:000000 C0:000000 C0:000000 C0:000000 C0:0000000 C0:0000000 C0:00000000 C0:0000000000	133           589           550           279           250           507           259           865           483           250           483           250           491           562           293           255           272           146           343           250           257           146           343           500           257           186           103           241	43 43 43 43 43 43 43 43 43 43 43 43 43 4	5 109 8 7 6 8 7 12 8 7 8 5 13 15 7 6 6 6 11 6 5 5 8 7	$\begin{array}{c} 2.07e^{-0.2}\\ 6.74e^{-0.4}\\ 3.66e^{-0.3}\\ 2.13e^{-0.4}\\ 1.40e^{-0.3}\\ 1.38e^{-0.2}\\ 2.02e^{-0.3}\\ 3.55e^{-0.4}\\ 1.28e^{-0.2}\\ 3.55e^{-0.4}\\ 1.28e^{-0.2}\\ 3.55e^{-0.4}\\ 1.28e^{-0.2}\\ 3.55e^{-0.4}\\ 3.55e^{-0.4}\\ 3.55e^{-0.4}\\ 3.55e^{-0.4}\\ 3.56e^{-0.2}\\ 3.99e^{-0.6}\\ 1.28e^{-0.2}\\ 3.99e^{-0.6}\\ 1.28e^{-0.2}\\ 3.99e^{-0.6}\\ 1.01e^{-0.6}\\ 4.37e^{-0.3}\\ 4.37e^{-0.3}\\ 4.37e^{-0.3}\\ 4.37e^{-0.3}\\ 5.9e^{-0.5}\\ 1.28e^{-0.5}\\ 1.28e^{-0.$	
BP BP	sister chromatid cohesion witotic sister chromatid cohesion	GD:0007062 GD:0007064	43 17	43 43	, 4 3	3.18e-03 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 <u>enrichmentmap-2.0.1.jar</u> is available in the wiki. See Apps > App Manager in menu.

• • •	App Manager	
	Install Apps Currently Installed Check for Updates	
Download Site:	http://apps.cytoscape.org/ OManage	Sites
٩		
all apps (140)	<ul> <li>Adj Exporter</li> <li>AgilentLiteratureSearch</li> <li>AllegroLayout</li> <li>aMatReader</li> <li>AnatApp</li> <li>ANIMO</li> <li>ARACNE</li> <li>AutoAnnotate</li> <li>bayelviraApp</li> <li>BiNCO</li> <li>BioGRID Data Source (Installed)</li> <li>Biomart Web Service Client (Inst</li> <li>BioPAX Reader (Installed)</li> <li>Bisogenet</li> <li>BridgeDb</li> <li></li> </ul>	
Install from Fi	View on App Store Insta	all
		Close

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

2	📂 🖹 놓 ╁ 🤄 🖆 🖽	E							
1	Control Panel	×							
	Select 🐇 Enrichment Map Input Panel								
$\wedge$	Analysis Type								
	GSEA About	2							
<u>Σ</u>	generic     Online Manual								
	DAVID/BiNGO/Great Bulk EM								
$\land$	User Input 🗸								
	Gene Sets 🗸								
	GMT: y2016/revised/tutorial/hsapiens.NAME.g								
	Datasets 🗸								
$\land$	Dataset 1 -								
	Expression:								
2	Enrichments: esults_GBM_1064935138386.								
	Advanced -	Advanced -							
	Ranks:								
	Classes								
	Phenotypes: UP VS. DOWN								
$\land$	Dataset 2 🗸								
	Expression:	Expression:							
4	Enrichments: ssults_KIRC_1033967998281								
	Advanced -								
	Ranks:								
	Classes								
	Phenotypes: UP VS. DOWN								
,	Reset Close Build								

Then set the analysis parameters



From the Cytoscape control panel, select the style tab, and map the EM1\_GS\_DESCR to graphic attribute Label

ontrol Panel			);		
Network	Style	Select	►		
EM1_Enrichmen	t_map_style	•	_		
Properties -		≈ :	~		
Def. Map. Byp.					
11	Border Paint	•	1		
15.0 ::	Border Width	•	ł		
<b>i</b> t	Fill Color				
	Height	6	1		
Ð	Label	-	,		
Column	E	M1_GS_DESCR ᅌ			
Mapping Type	Pas	sthrough Mapping			
		Ē	Ì		
	Label Color				
12	Label Font Size				
0	Shape 📢				
15.0 ::	Size 4				
255	Transparency 4				
	Width	0	(		
Lock node	width and heig	ht			

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.

	Control Panel				□ ×			
	Network	Sty	le	Select				
	EM1_Enrichmen	EM1_Enrichment_map_style •						
$\land$	Properties - Def. Map. Byp.			:	* ≈			
		Border Paint						
	Column		EM1_fdr	_qvalue_dataset2				
	Mapping Type	$\wedge$	Continuo	ous Mapping				
	Current Mapping							
					创			
$\diamond$	15.0 <b>il</b>	Border Wi	dth		•			
		Fill Color			•			
	Column		EM1_fdr	_qvalue_dataset1				
	Mapping Type	^	Continuous Mapping					
	Current Mapping		0.00		0.05			
					Ü			

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.