

ROCK: A RESOURCE FOR INTEGRATIVE BREAST CANCER DATA ANALYSIS

SAIF UR-REHMAN
CANCER INFORMATICS
BREAKTHROUGH BREAST CANCER RESEARCH
THE INSTITUTE OF CANCER RESEARCH, LONDON, UK

NETTAB 2012, COMO, ITALY 16/11/2012

Breast Cancer

- Breast cancer rates have increased by more than 50% over the last twenty years.
- Breast cancer is now the most common cancer in the UK, with more than 46,000 women and 200 men diagnosed each year, and more than a million cases worldwide.
- In the last decade numerous experimental approaches have been employed in an attempt to identify sub-types of breast cancer and new molecular targets for pharmaceutical interventions.

Outline

- Introduce issues surrounding breast cancer data integration.
- Define ROCK as a response to some of these issues.
- Illustrate the utility of ROCK via the use of case studies.

Data Types

- Clinical Annotation.
- Gene expression.
- DNA copy number.
- DNA methylation.
- Non-Coding RNA expression.
- Protein expression.
- Mutations/SNPs.

Integration issues

- Lack of consistency in sample classification schemes.
- Mapping between data types e.g. gene to protein.

ROCK

- ROCK Online Cancer Knowledgebase.
- Database containing the results of a large number of high throughput experiments on breast cancer.
- Currently focussed on gene expression and DNA copy number.
- Aimed primarily at bench scientists but is also utilised by bioinformaticians.
- Available at rock.icr.ac.uk

Aims of ROCK

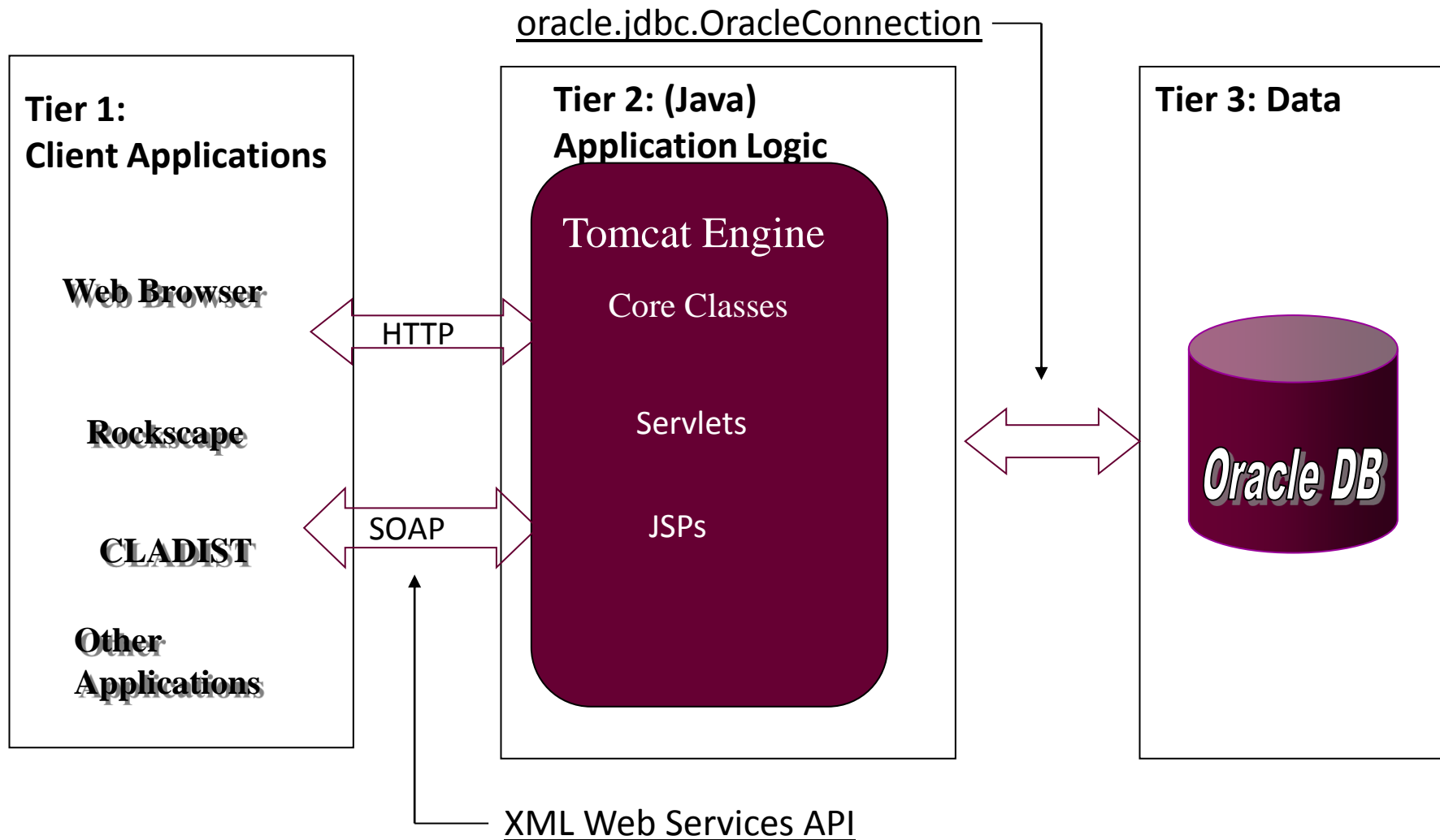
- To provide an integrative framework for breast cancer experimental data.
- To provide functionality allowing bench scientists to use this functionality to test hypotheses in-silico in previously published datasets.

Data in ROCK

Gene Expression Studies:	84	aCGH Projects:	12
Gene Expression Platforms:	54	aCGH Platforms:	9
Analysed Gene Expression Projects:	63	aCGH Samples:	598
Gene Expression Analyses:	216	aCGH CNV Analyses:	40
Differentially Expressed Genes:	21862	aCGH CNVs:	2940
Gene Expression Samples:	7261	Total Genes in CNVs:	19974
Gene Expression Signatures:	38		

Additional data in ROCK

- Data from the Cancer Genome Atlas (TCGA)
- microRNA expression
- Gene expression measured by NGS (RNASeq)
- Human protein protein interaction data from HPRD, BioGrid, Mint (IMEX members) amongst others.

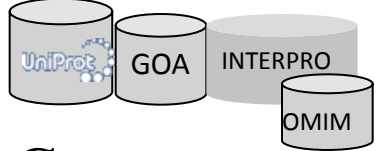


① Aggregate

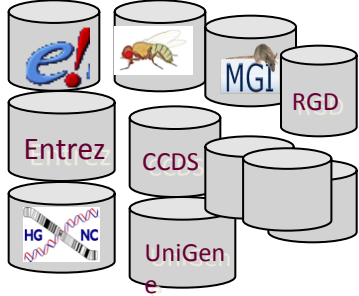
②

③ Analyze

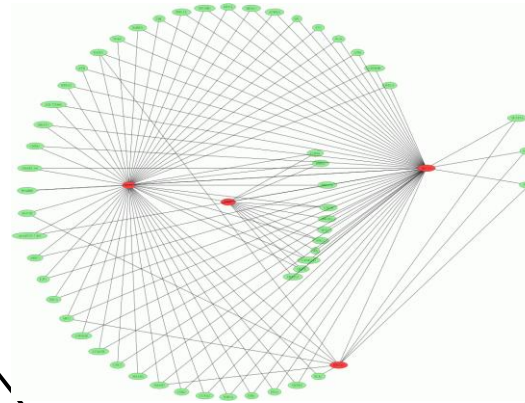
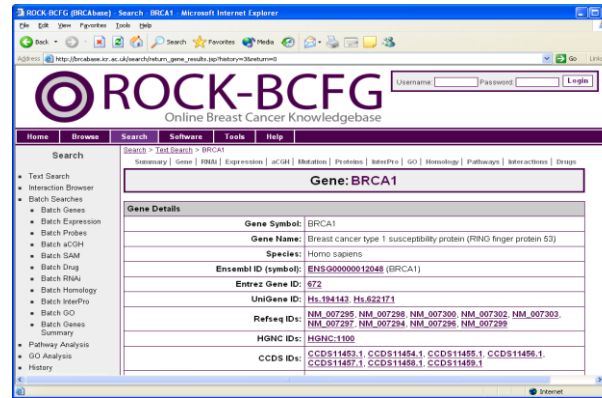
Protein



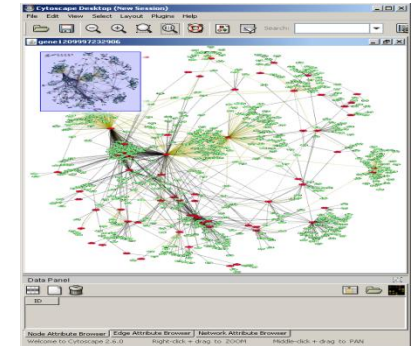
Gene



ROCKORA



Rockscape



Experimental

Expression Arrays
aCGH
RNAi Screens
Sample Annotation
etc.

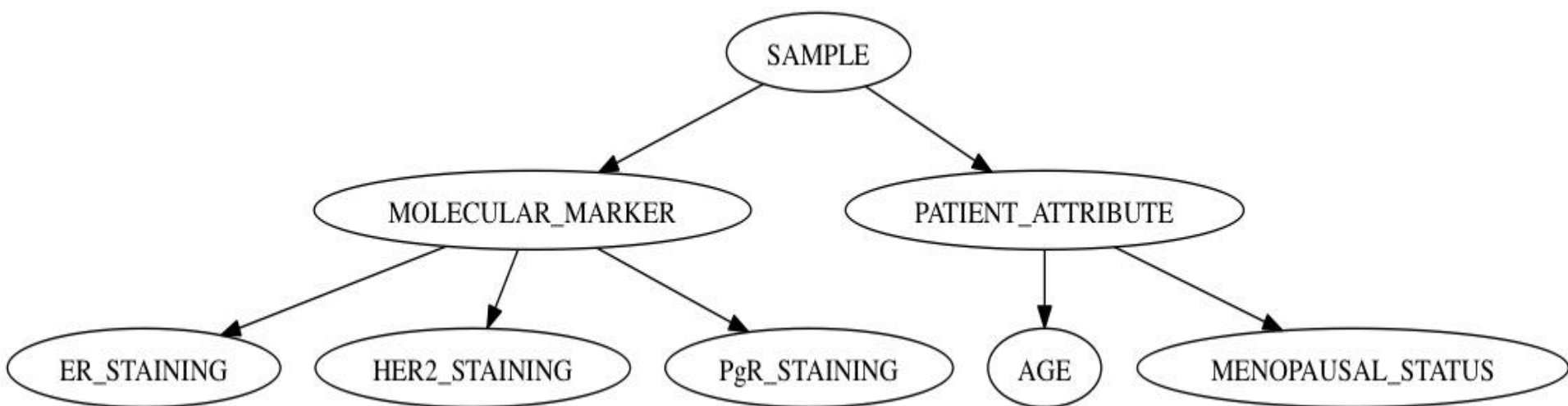
XML Web Services
API (SOAP)

Cladist

PPI/Pathway
Data

Sample Ontology

- The sample is the fundamental data entity in ROCK.
- All samples held in ROCK are classified within an ontological framework.
- Allows comparisons of hypotheses between studies.
- Represented in standard XML.
- Samples can be stratified by up to three annotation terms.





Case Study 1: FZD7

- Frizzled 7 is a cell surface which is an initiating molecule for the Wnt signalling pathway.
- Signalling pathways allow a cell to respond to its immediate environment.
- Faulty Wnt signalling is indicated as possible cause of some breast cancers.

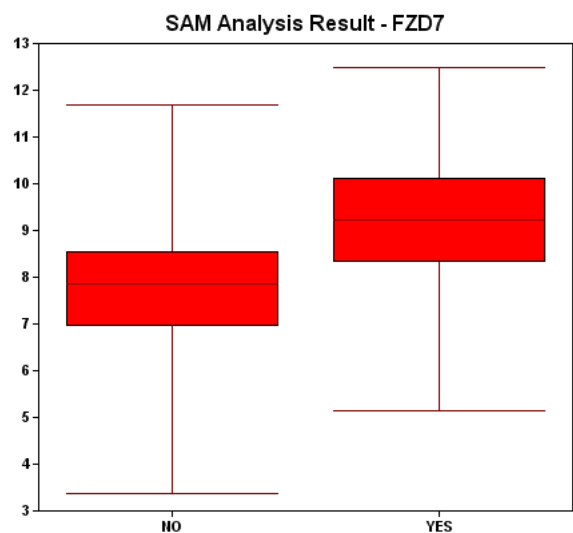
Case Study 1: FZD7 (cont)

- Recent work by Yang *et al.* has shown that FZD7 is upregulated in triple negative breast cancer.
- Triple negative breast cancer is resistant to hormonal treatment and as such carries a comparatively poorer prognosis.
- Yang L, Wu X, Wang Y, Zhang K, Wu J, Yuan YC, Deng X, Chen L, Kim CC, Lau S *et al*: **FZD7 has a critical role in cell proliferation in triple negative breast cancer.** *Oncogene* 2011, **30**(43):4437-4446.

Expression <ul style="list-style-type: none"> Data by Sample Annotations Probes by Gene IDs Probe Annotation arrayCGH <ul style="list-style-type: none"> CNA Types by Studies CNA Types by Sample Annotation Analysis <ul style="list-style-type: none"> Interaction SAM by Annotation SAM by Studies Survival by Annotation and Expression Survival by Annotation only Pathway Analysis GO Analysis RNAi <ul style="list-style-type: none"> Batch RNAi Drugs <ul style="list-style-type: none"> Link to canSAR Batch Drugs Batch Drug Targets <hr/> History	Upload Identifiers	
	Enter list of IDs:	<div>FZD7</div>
	Retrieval Expression Base On:	Median Of Matched Probes ▾
	Affymetrix probeset filtering:	<input checked="" type="radio"/> Yes <input type="radio"/> No
	Expression Study:	Popivici (2010) ▾ ↻
	Platform:	HG-U133A - MAS5 ▾ ↻
	Retrieval Expression Type:	group mean of normalized expression ▾
	Sample annotation term:	Triple Negative ▾ ↻
	Annotation:	<input checked="" type="checkbox"/> NO (207) <input checked="" type="checkbox"/> YES (71) <input type="checkbox"/> Select All
	Operator:	<input checked="" type="radio"/> And <input type="radio"/> By Selected Sample Terms
	Second sample annotation term:	▾ ↻
	Annotation 2:	

	Submitted Item	Matched Gene	Triple -NO	Triple -YES	Fold Change	Response
1	FZD7	<input checked="" type="checkbox"/> <u>FZD7</u>	7.721 	9.0369 	2.4896	Triple -YES

Response	SAM Score	Fold Change	Q Value (%)
YES	4.973127	2.571173	0



Microarray gene expression analyses


- SAM (Significance of microarrays)
- Correlation with known gene signatures for tumour/sample classification (PAM50).
- Co-expression analyses.

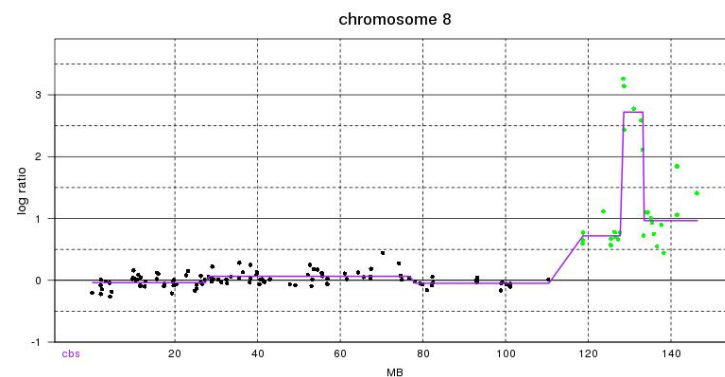
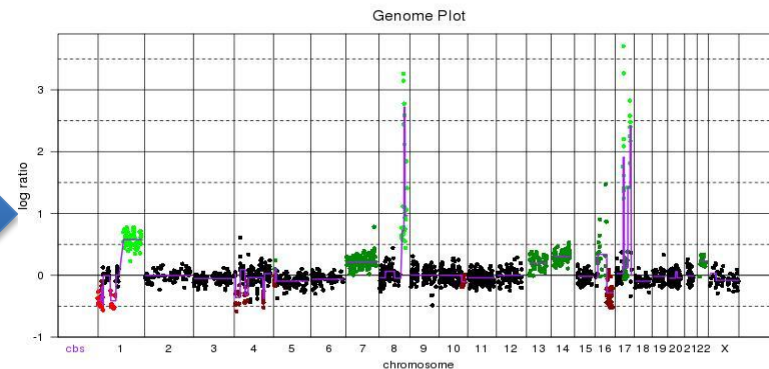
Case Study 2:MYC

- *MYC*/c-MYC is a transcription factor which when over-expressed can drive cell proliferation.
- Chromosomes are frequently altered in various cancer types.
- It is possible to query ROCK as to whether the area of the chromosome containing *MYC* is altered in some studies.
- MYC is located on the q arm of chromosome 8 on the forward strand between the hg19 genomic coordinates 128,747,680 and 128,753,674)

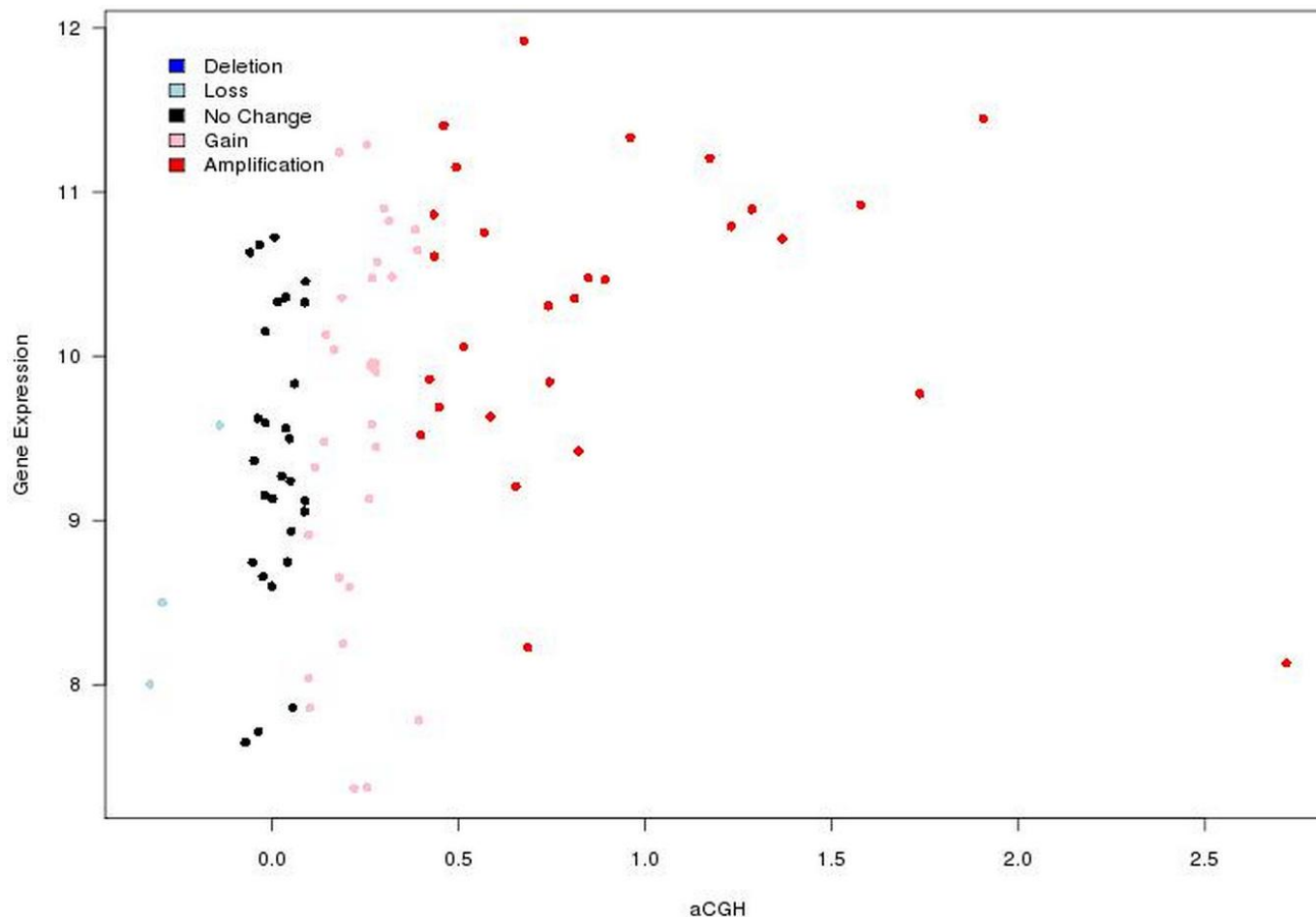
Gene: MYC

Gene Details

Gene Symbol:	MYC
Gene Name:	v-myc myelocytomatosis viral oncogene homolog (avian) [Source:HGNC Symbol Acc:7553]
Species:	Homo sapiens
Ensembl ID (symbol):	ENSG00000136997 (MYC)
Entrez Gene ID:	4609
UniGene ID:	Hs.202453 , Hs.596284
Refseq IDs:	NM_002467
HGNC IDs:	HGNC:7553
CCDS IDs:	CCDS6359
Other Aliases:	cMyc, MRTL, bHLHe39, c-Myc
Chromosomal Location:	8q24.21
PDB:	1MV0 , 1A93 , 1EE4 , 2A93 , 1NKP
OMIM:	Burkitt lymphoma
Links:	The Human Protein Atlas GeneCards Information Hyperlinked Over Proteins (iHOP) WolframAlpha
miRNA IDs:	View miRNA families
 TargetScanHuman prediction of microRNA targets	NM_002467



Gene Expression - DNA Copy Number Correlation Plot for Gene: MYC with rho: 0.4602 and p.value: 0.0000



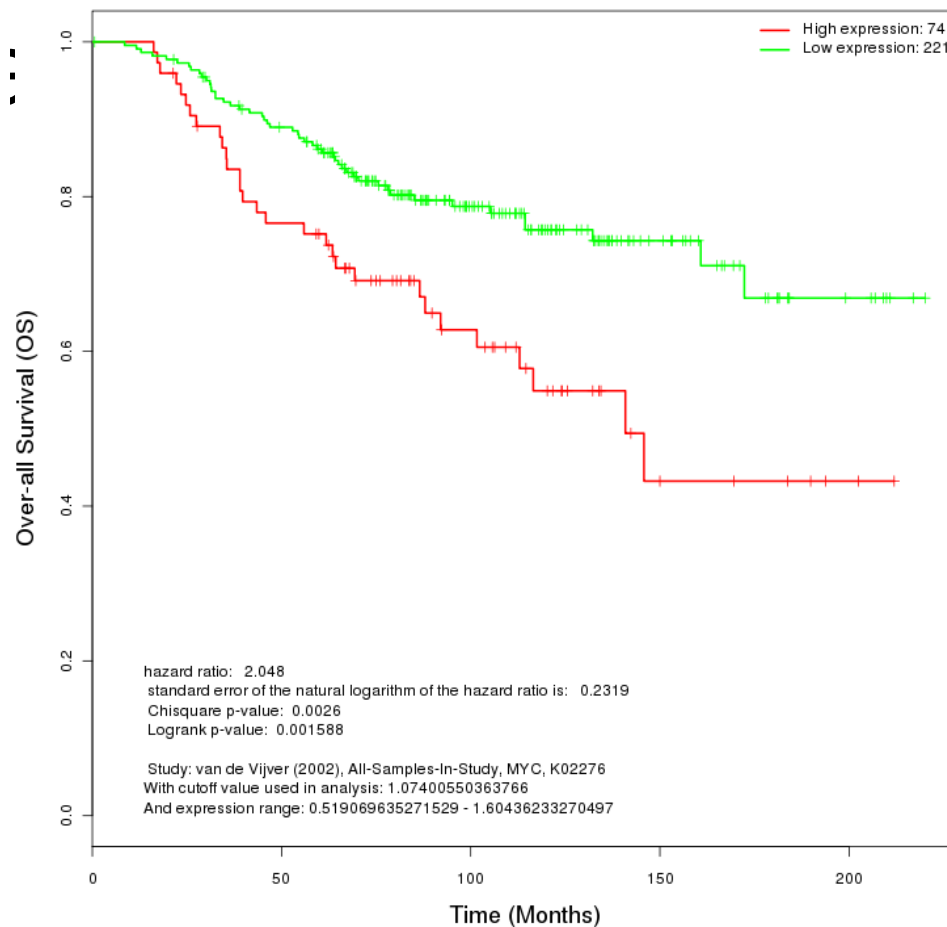
Survival Analysis

- ROCK also provides survival analysis.
- Users can check if the expression level of a given gene is linked to a particular prognosis/outcome.
- Only applicable in studies where survival data is known.

Upload Identifiers	
Enter list of IDs:	MYC
Case Sensitive:	<input type="button" value="No"/>

Kaplan Meyer Survival Analysis	
Gene Expression Study:	van de Vijver (2002) <input type="button" value="↺"/>
Platform & Normalisation:	NKI - mean center <input type="button" value="↺"/>
Method of groups:	top 25% against rest (upper-quartile) <input type="button" value="↺"/>
Length of Time over Survival:	study endpoint <input type="button" value="↺"/>
Survival term:	<input type="button" value="↺"/>
Sample annotation term:	<input type="button" value="↺"/> <input type="button" value="↻"/>
Annotation:	
Operator:	<input checked="" type="radio"/> And <input type="radio"/> By Selected Sample Terms
Second sample annotation term:	<input type="button" value="↺"/> <input type="button" value="↻"/>
Annotation 2:	
Operator:	<input checked="" type="radio"/> And
Third sample annotation term:	<input type="button" value="↺"/> <input type="button" value="↻"/>
Annotation 3:	
<input type="button" value="Analyse"/>	

Kaplan Meier Plot for Gene: MYC



Iterative integration

- ROCK links the results of all analyses.
- This allows a user to undertake an iterative process where the results of one analysis are projected onto another.
- Registered users can save gene lists within ROCK and use them as initiation points for subsequent analyses.

GO/Pathway enrichment

- Sets of genes retrieved via ROCK analyses can be examined for enrichment in GO terms as well as pathway membership.
- KEGG
- REACTOME

Updates/Further Work

- Additional data types being added.
- Methylation/Epigenetic data
- Protein expression data.
- Other cancer types (Prostate/Ovarian)

ROCK Mission

- To provide an integrative framework for breast cancer experimental data.
- To provide functionality allowing bench scientists to use this functionality to test hypotheses in-silico in previously published datasets.

Acknowledgements

Cancer Informatics team

- Alice Gao
- Costas Mitsopoulos
- Jarle Hakas
- Marketa Zvelebil



Funding from
Breakthrough Breast
Cancer

26/11/2012



Thank you for your attention

- Any questions?
- Url: rock.icr.ac.uk