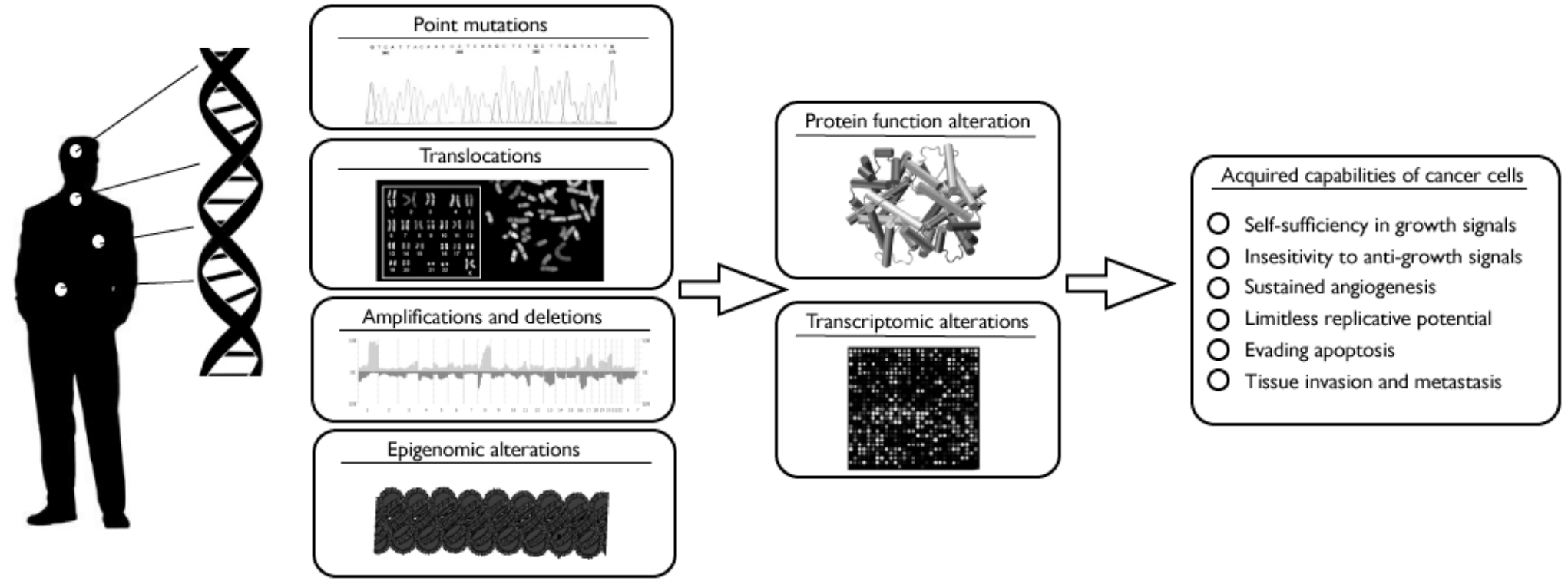
IntOGen: A novel framework for integration and data-mining of multidimensional oncogenomic data

Christian Pérez-Llamas

Biomedical Genomics Group
Universitat Pompeu Fabra

Parc de Recerca Biomedica de Barcelona



intOgen.org

Integrative Onco Genomics

intOgen Overview

Oncogenomics Data

- Transcriptomic alterations
- Copy Number Variation alt.
- Mutations
- Epigenomic alterations

Clinical Annotations

International Classification of Diseases for Oncology

Genomic Annotations

- Ensembl
- GO
- KEGG
- ...

Statistics

- Raw data processing
- Cancer related genes identification
- Cancer related modules identification
- Combinations
- Correlations

DB

Visualization & Exploration

www.intogen.org

GiTools

www.gitools.org

- BioMart service
- DAS server
- SOAP-WS

intOgen Overview

Oncogenomics Data

- Transcriptomic alterations
- Copy Number Variation alt.
- Mutations
- Epigenomic alterations

Clinical Annotations

International Classification of Diseases for Oncology

Genomic Annotations

- Ensembl
- GO
- KEGG
- ...

Statistics

- Raw data processing
- Cancer related genes identification
- Cancer related modules identification
- Combinations
- Correlations

DB

Visualization & Exploration

www.intogen.org

GiTools

www.gitools.org

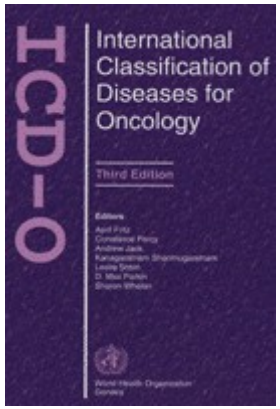
- BioMart service
- DAS server
- SOAP-WS

intOgen Collation of Cancer Data

	Transcriptomic Alterations	Copy Number Variations	Mutations
GEO Gene Expression Omnibus	✓		
ARRAYEXPRESS	✓		
The Cancer Genome Atlas	✓	✓	
progenetix		✓	
Copy Number Analysis from Sanger Institute		✓	
COSMIC			✓
International Cancer Genome Consortium	⚙	⚙	⚙

intOgen Annotation of Cancer Data

Clinical Annotations of Patients and Samples



International Classification of Diseases for Oncology

ICD-O
- Topography
- Morphology

Prepared for any type of annotations:

- Type of treatment
- Age of onset
- Gender
- Cell type
- ...

Sources of Knowledge

Ensembl

KEGG Kyoto Encyclopedia of Genes and Genomes

GO

OBO

intOgen Overview

Oncogenomics Data

- Transcriptomic alterations
- Copy Number Variation alt.
- Mutations
- Epigenomic alterations

Clinical Annotations

International Classification of Diseases for Oncology

Genomic Annotations

- Ensembl
- GO
- KEGG
- ...

Statistics

- Raw data processing
- Cancer related genes identification
- Cancer related modules identification
- Combinations
- Correlations

DB

Visualization & Exploration

www.intogen.org

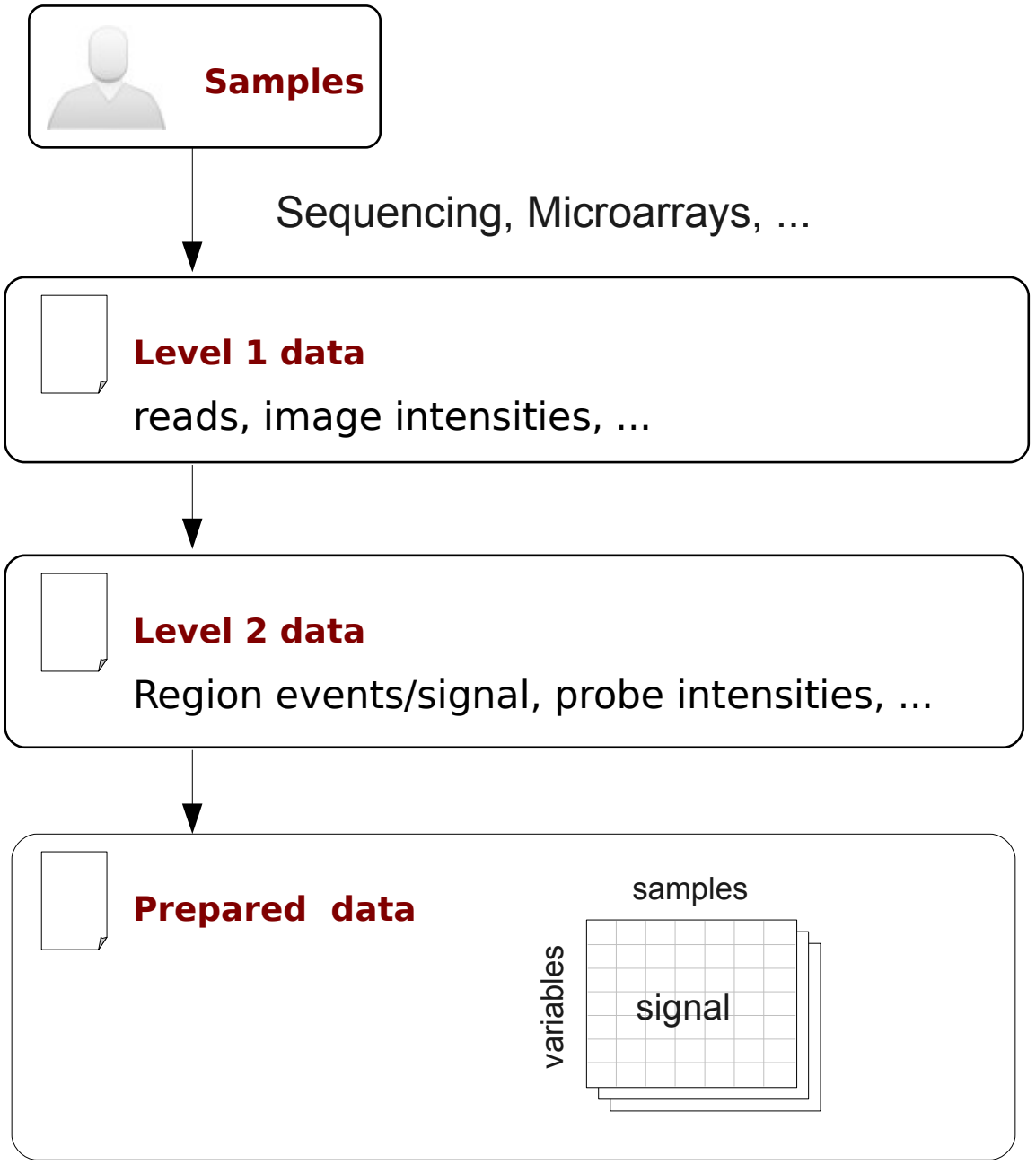
GiTools

www.gitools.org

- BioMart service
- DAS server
- SOAP-WS

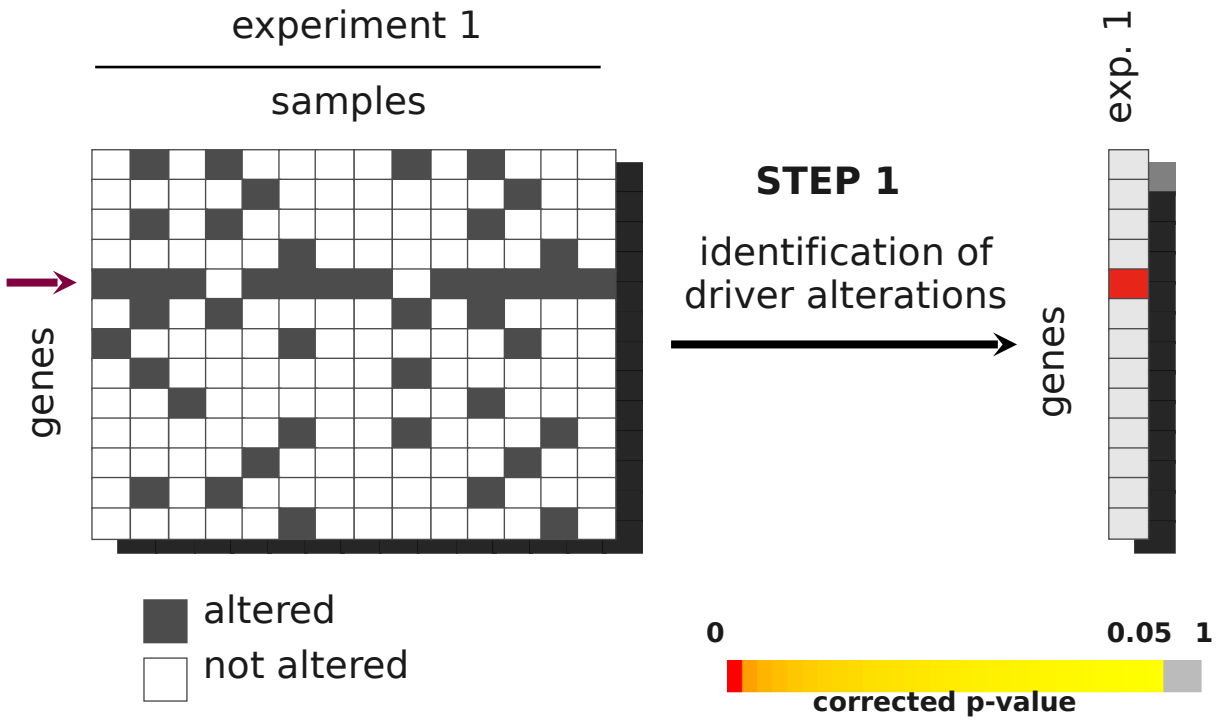
Raw data processing

- **Raw data processing**
- Cancer related genes identif.
- Cancer related modules identif.
- Combinations
- Correlations



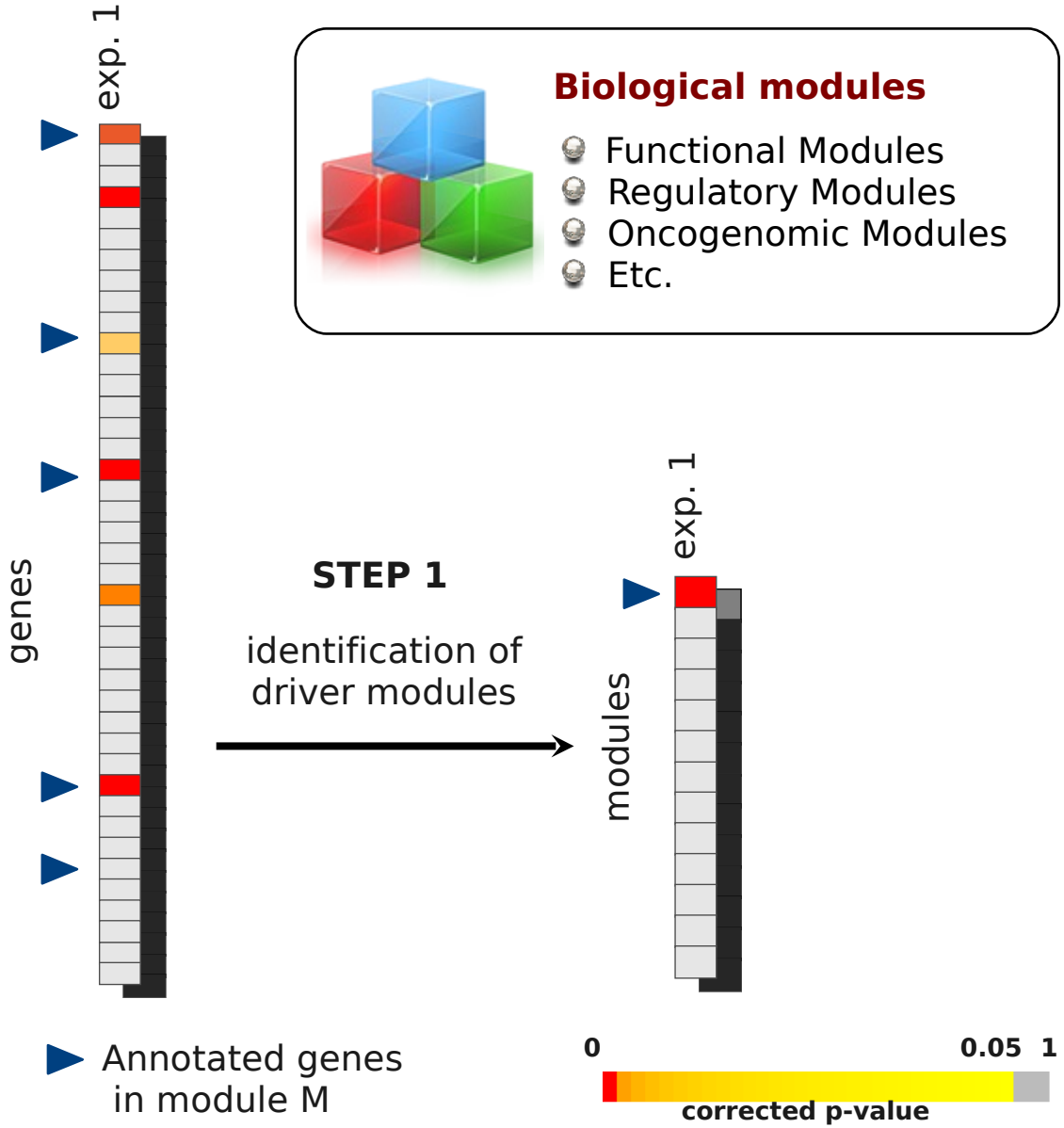
Cancer related genes identification

- Raw data processing
- **Cancer related genes identif.**
- Cancer related modules identif.
- Combinations
- Correlations



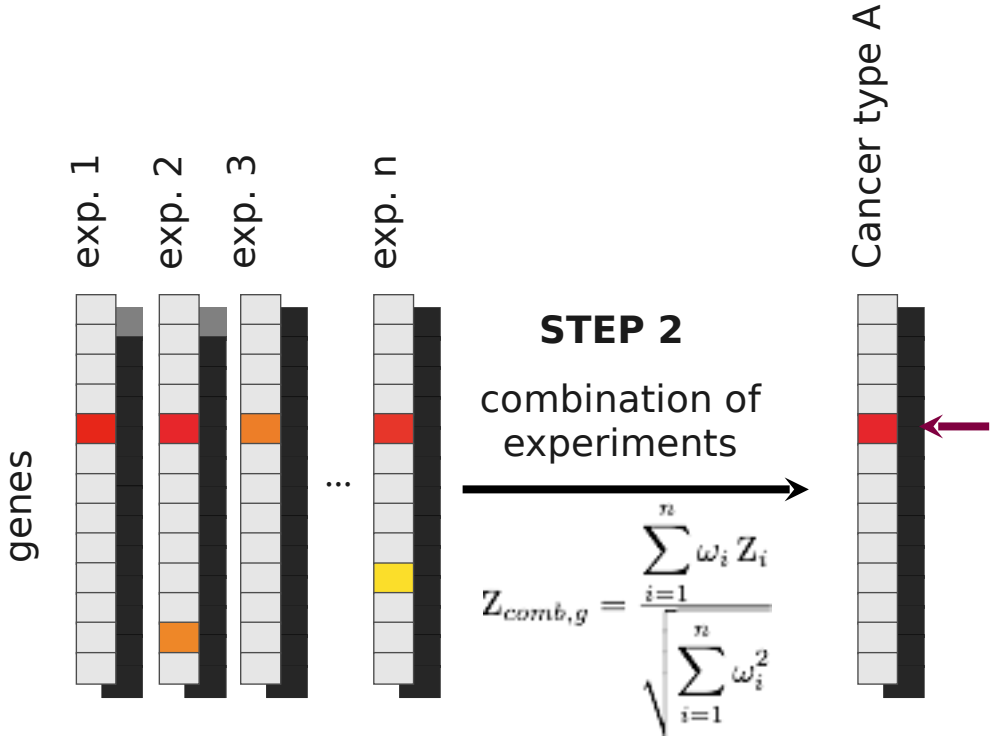
Cancer related modules identification

- Raw data processing
- Cancer related genes identif.
- **Cancer related modules identif.**
- Combinations
- Correlations



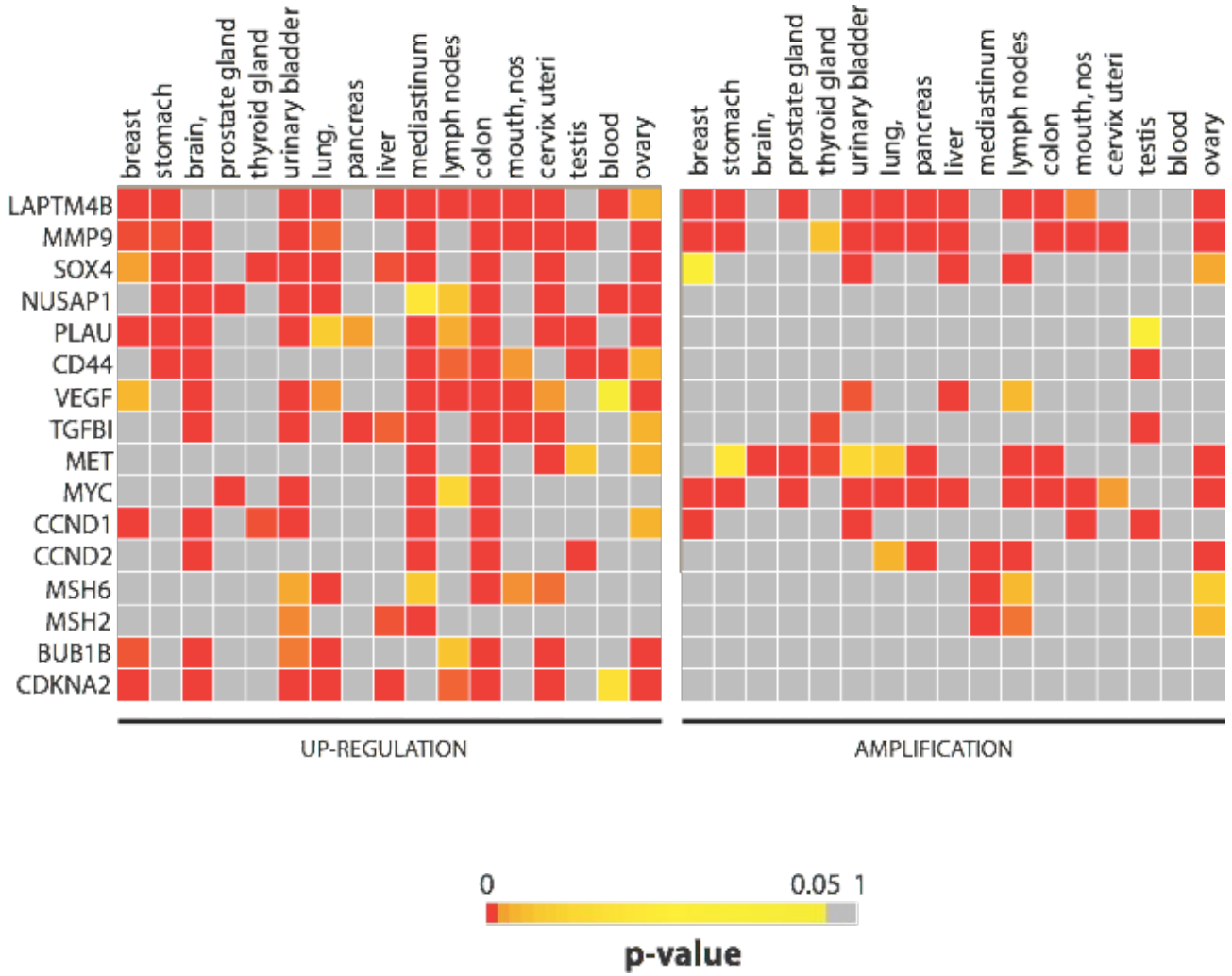
Combinations

- Raw data processing
- Cancer related genes identif.
- Cancer related modules identif.
- **Combinations**
- Correlations

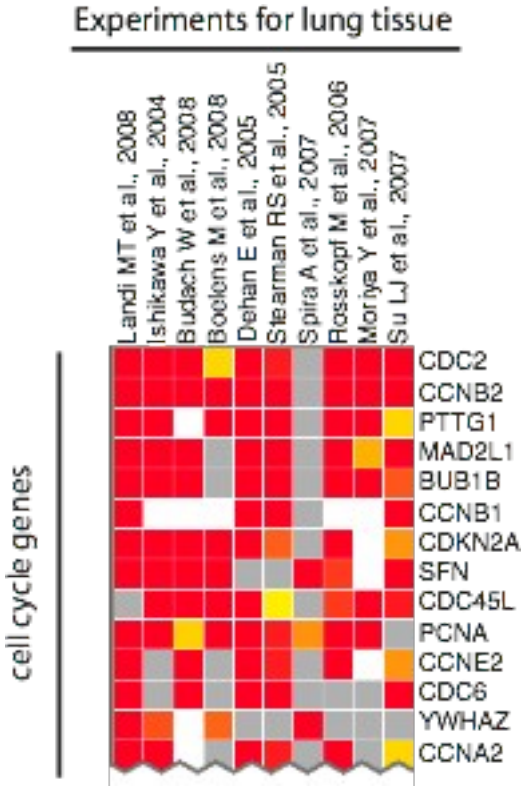


Combinations

- Raw data processing
- Cancer related genes identif.
- Cancer related modules identif.
- **Combinations**
- Correlations



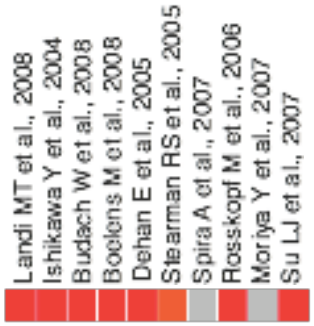
Analysis at different levels



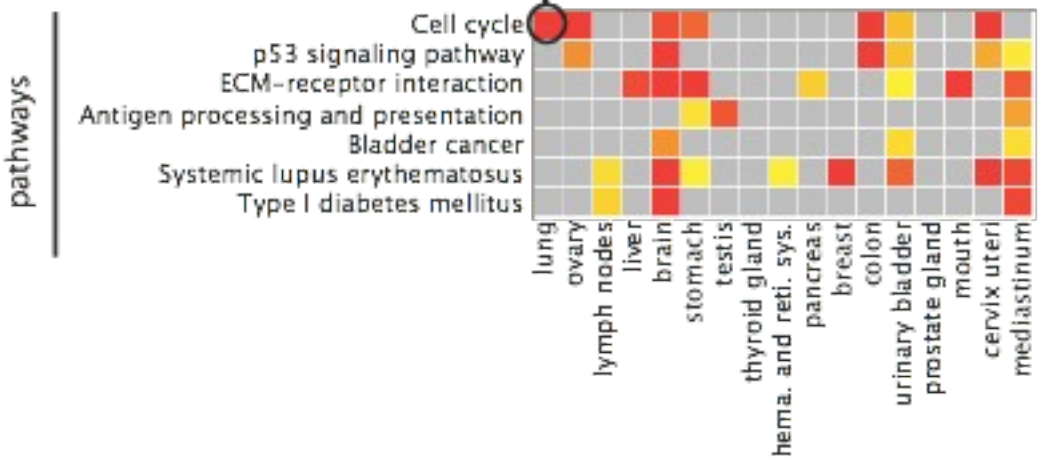
Biological modules

Functional Modules
Regulatory Modules
Oncogenomic Modules
...

MODULE ANALYSIS →

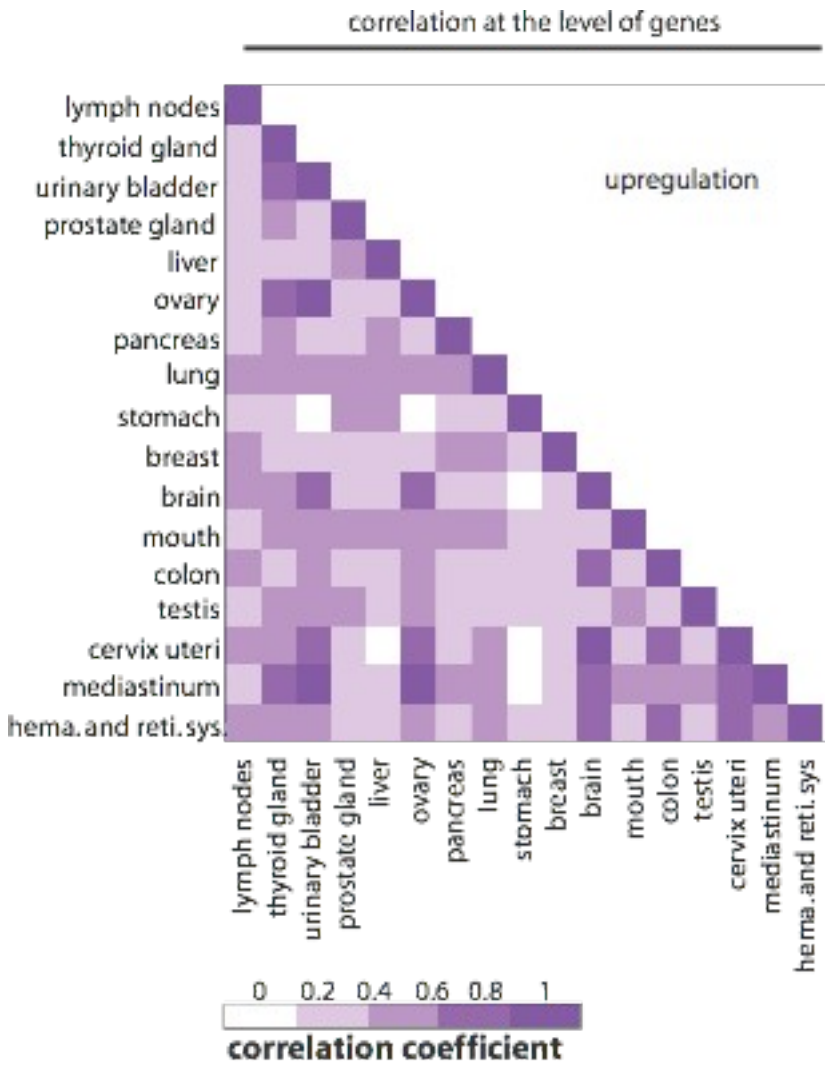


COMBINATION



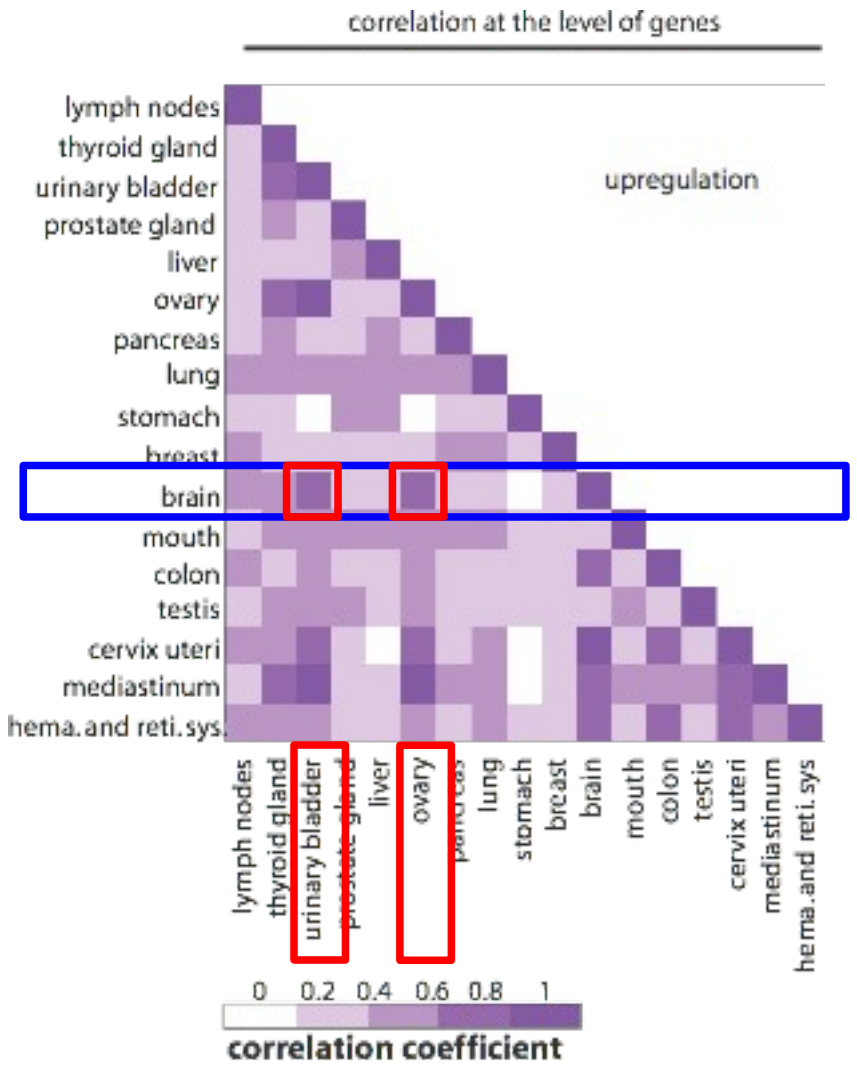
Correlation between cancer types

- Raw data processing
- Cancer related genes identif.
- Cancer related modules identif.
- Combinations
- **Correlations**



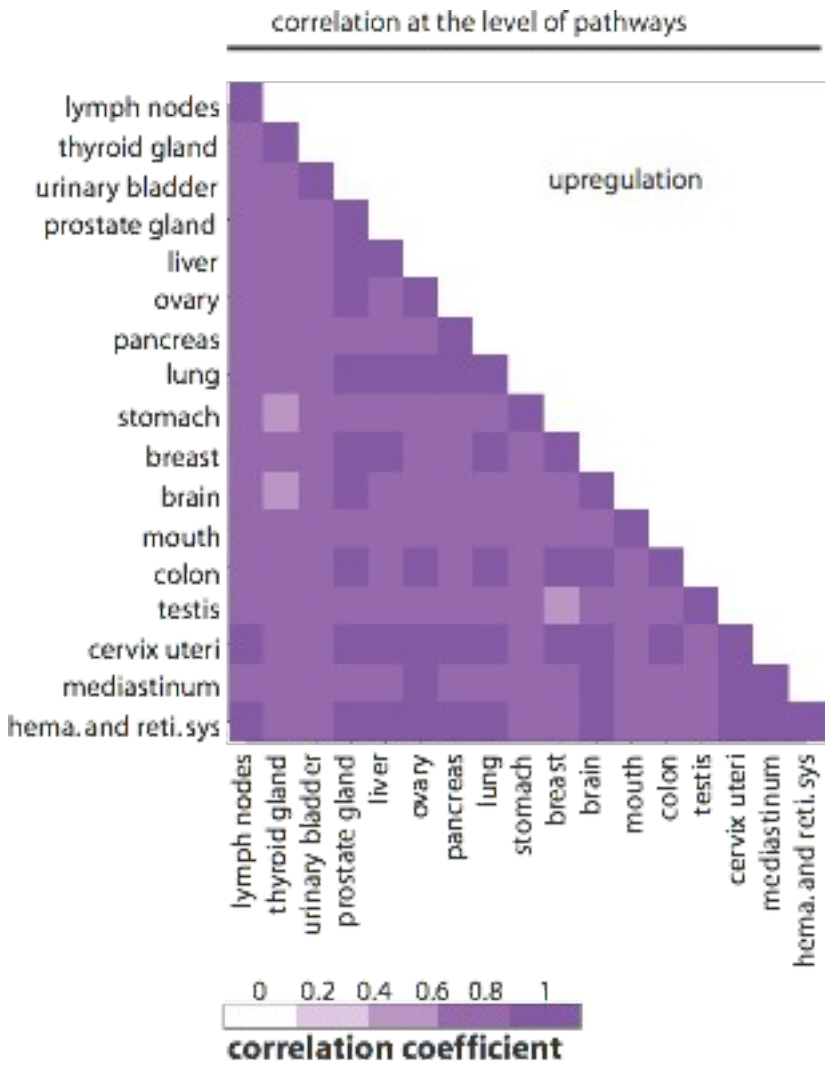
Correlation between cancer types

- Raw data processing
- Cancer related genes identif.
- Cancer related modules identif.
- Combinations
- **Correlations**



Correlation between cancer types

- Raw data processing
- Cancer related genes identif.
- Cancer related modules identif.
- Combinations
- **Correlations**



intOgen Overview

Oncogenomics Data

- Transcriptomic alterations
- Copy Number Variation alt.
- Mutations
- Epigenomic alterations

Clinical Annotations

International Classification of Diseases for Oncology

Genomic Annotations

- Ensembl
- GO
- KEGG
- ...

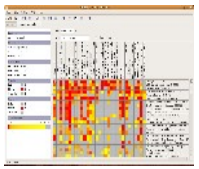
Statistics

- Raw data processing
- Cancer related genes identification
- Cancer related modules identification
- Combinations
- Correlations

DB

Visualization & Exploration

www.intogen.org



GiTools

www.gitools.org

- BioMart service
- DAS server
- SOAP-WS

Integrative OncoGenomics

IntOGen, a discovery tool for cancer researchers, is a resource that integrates multidimensional OncoGenomics Data for the identification of genes and groups of genes (biological modules) involved in cancer development.

Please note: IntOGen is **UNDER DEVELOPMENT**. The results shown are preliminary and the browser is a beta version, errors may be found and we are currently working to fix them.

If you are interested in using IntOGen and would like to be informed when new options are added subscribe to the IntOGen mailing list.

Explore IntOGen

intOgen Discovery tool for cancer research

Gene Search

Search for a gene of interest and have a wide view of the implication of this gene in cancer

Experiment search

Search for a experiment and identify the most significant genes and modules altered

Tumor type search

Search for a cancer type of interest and discover the most significant altered genes and modules

Cancer Gene Prioritization

Prioritize a list of candidate genes based on CGPrio probabilities and oncogenomic alterations

Combine experiments

Choose a group of experiments and combine the statistics to obtain combined significant genes and modules

Analyze your modules

Load your own modules and find out if they are enriched for genes significantly altered in different cancer types

Experiments [ ] all

Tumor type [ ] [ ] all

Genes / Modules [ ] all > Genes > *CDKN2A*

Details | **Tumor type** | Experiments

<< Search

Filters

p-value

Export

CSV file

Tumor type	Alterations				
	Mutations	CNA		Transcrip	
Topography	Fnd / Std	Loss	Gain	Down	Up
cervix uteri	36 / 504				
breast	33 / 604				
brain, nos	329 / 1543				
ovary	70 / 657				
lung	300 / 1896				
colon, nos	9 / 273				
liver	88 / 656				
lymph nodes					
stomach	43 / 526				
stomach	0 / 8				
urinary bladder	226 / 1371				
thyroid gland	22 / 174				
hematopoietic and reticuloendothelial systems	603 / 4051				
mouth, nos					
pancreas	208 / 623				
testis	0 / 33				
mediastinum					
prostate gland	11 / 280				
cerebellum					
nervous system, nos					
tonsil					
pharynx, nos					
unspecified digest. organs					
larynx					
pituitary gland	0 / 71				
adrenal glands	1 / 46				

Showing 1 to 26 of 54 << < 1 2 3 > >>

Experiments [ ] all

Tumor type [breast] [ ] all > breast

Genes / Modules [ ] all > Genes > CDKN2A

Details | Tumor type | Experiments

<< [ ] Search

Filters

p-value

1.00 0.05 1.00

Experiment and platform					Alterations					
					CNA			Transcriptomic		
Authors	Title	Year	Ref	Platform	Loss	Gain	N	Down	Up	N
Yu K et al.	A precisely regulate...	2008	GSE5364	Affymetrix Gene...						183
Finak G et al.	Stromal gene express...	2008	GSE9014	Agilent-012391 ...						52
Richardson A et al.	Expression data fro...	2007	GSE7904	Affymetrix Gene...						43
Richardson AL et al.	X chromosomal abnorm...	2006	GSE3744	Affymetrix Gene...						40
Weigelt B et al.	Molecular portraits ...	2005	GSE2741	Agilent Human 1...						64
Weigelt B et al.	Molecular portraits ...	2005	GSE2741	Agilent-012097 ...						55
Mostafavi, ZB et al.	Gene Expression Anal...	2008	GSE4122	Affymetrix Gene...						53
Casey T et al.	Molecular signatures...	2008	GSE10797	Affymetrix Gene...						28
Davis SR et al.	Gene Expression and ...	2008	GSE7882	NCI Qiagen Homo...						54
Radvanyi L et al.	The gene associated ...	2005	GSE1477	Protein Design ...						51
Persson K et al.	Chromosomal aberrati...	1999	10337995	CGH			27			
Hwang KT et al.	Genomic copy number ...		18649361	aCGH			62			
Aubele M et al.	Intratumoral heterog...	1999	10214356	CGH			23			
Weber-Mangal S et al.	Breast cancer in you...		14520696	CGH			85			
Fridlyand J et al.	Breast tumor copy nu...		16620391	aCGH			67			
Gronwald J et al.	Comparison of genomi...		15540206	CGH			45			
Jones C et al.	Comparative genomic ...		12533827	CGH			28			
Jones C et al.	Molecular cytogeneti...		11141494	CGH			24			
Fung LF et al.	Genetic imbalances i...	2001	11165323	CGH			40			
Vincent-Salomon A et al.	Identification of ty...		17417968	aCGH			59			
Ried T et al.	Comparative genomic ...		7585611	CGH			20			
Teixeira MR et al.	Evaluation of breast...	2001	11494114	CGH			37			
van Beers EH et al.	Comparative genomic ...		15705879	CGH			132			
Moore E et al.	Widespread chromosom...	1999	10398098	CGH			23			
Park SC et al.	Genetic changes in b...		17380298	CGH			32			
Hwang ES et al.	Clonality of lobular...		15197797	CGH			48			

Showing 1 to 26 of 31

<< < 1 2 > >>

Experiments [ ] all

Tumor type [breast] [ ] all > breast

Genes / Modules [ ] all > Genes > CDKN2A

Details | Tumor type | Experiments

<< [ ] Search

Filters

p-value

0.00 0.05 1.00

Experiment and platform					Alterations					
					CNA			Transcriptomic		
Authors	Title	Year	Ref	Platform	Loss	Gain	N	Down	Up	N
Yu K et al.	A precisely regulate...	2008	GSE5364	Affymetrix Gene...						183
Finak G et al.	Stromal gene express...	2008	GSE9014	Agilent-012391 ...						52
Richardson A et al.	Expression data fro...	2007	GSE7904	Affymetrix Gene...						43
Richardson AL et al.	X chromosomal abnorm...	2006	GSE3744	Affymetrix Gene...						40
Weigelt B et al.	Molecular portraits ...	2005	GSE2741	Agilent Human 1...						64
Weigelt B et al.	Molecular portraits ...	2005	GSE2741	Agilent-012097 ...						55
Mostafavi, ZB et al.	Gene Expression Anal...	2008	GSE4122	Affymetrix Gene...						53
Casey T et al.	Molecular signatures...	2008	GSE10797	Affymetrix Gene...						28
Davis SR et al.	Gene Expression and ...	2008	GSE7882	NCI Qiagen Homo...						54
Radvanyi L et al.	The gene associated ...	2005	GSE1477	Protein Design ...						51
Persson K et al.	Chromosomal aberrati...	1999	10337995	CGH			27			
Hwang KT et al.	Genomic copy number ...		18649361	aCGH			62			
Aubele M et al.	Intratumoral heterog...	1999	10214356	CGH			23			
Weber-Mangal S et al.	Breast cancer in you...		14520696	CGH			85			
Fridlyand J et al.	Breast tumor copy nu...		16620391	aCGH			67			
Gronwald J et al.	Comparison of genomi...		15540206	CGH			45			
Jones C et al.	Comparative genomic ...		12533827	CGH			28			
Jones C et al.	Molecular cytogeneti...		11141494	CGH			24			
Fung LF et al.	Genetic imbalances i...	2001	11165323	CGH			40			
Vincent-Salomon A et al.	Identification of ty...		17417968	aCGH			59			
Ried T et al.	Comparative genomic ...		7585611	CGH			20			
Teixeira MR et al.	Evaluation of breast...	2001	11494114	CGH			37			
van Beers EH et al.	Comparative genomic ...		15705879	CGH			132			
Moore E et al.	Widespread chromosom...	1999	10398098	CGH			23			
Park SC et al.	Genetic changes in b...		17380298	CGH			32			
Hwang ES et al.	Clonality of lobular...		15197797	CGH			48			

Showing 1 to 26 of 31

<< < 1 2 > >>

Experiments | all > *Agilent-012391 / Finak G*

Tumor type | breast | infiltrating duct | all > breast > *infiltrating duct carcinoma, nos*

Genes / Modules | all > Genes > *CDKN2A*

Details

Experiment and platform

- **Authors**: Finak G et al. (2008)
- **Title**: Stromal gene expression predicts clinical outcome in breast cancer. (pubmed. 18438415)
- **Journal**: Nature medicine
- **Database**: GEO **Reference**: GSE9014
- **Platform**: Agilent-012391 Whole Human Genome Oligo Microarray G4112A (Feature Number version) **Type**: mRNAexpr

CGPrio predictions

Type	Oncogene	Suppressor
PC-GS	0.361	0.308
PC-GS-PI	0.969	0.942
PC-GS-PD	0.314	0.328
PC-GS-RD	0.55	0.273
PC-GS-PD-PI-RD	0.86	0.941

Transcriptomic Alterations

	Upreg	Downreg
Q-Value	1.466E-17	1.048
P-Value	6.738E-19	1
Observed	46	0
Expected	15.175	14.007
Deviation	3.278	3.199
Samples	52	

Cancer Gene Census

- **Cancer Somatic Mutated**: yes
- **Cancer Germline Mutated**: yes
- **TumourTypes (Somatic Mutations)**: melanoma, multiple other tumour types
- **TumourTypes (Germline Mutations)**: melanoma, pancreatic
- **Cancer Syndrome**: Familial malignant melanoma
- **Tissue Type**: L, E, M, O
- **Cancer Molecular Genetics**: Rec

Ensembl Genes

- **Ensembl ID**: ENSG00000147889
- **Symbol**: CDKN2A
- **Chromosome**: 9
- **Band**: p21.3
- **Description**: Cyclin-dependent kinase inhibitor 2A, isoform 4 (p14ARF) (p19ARF)

Tumor type

- **Morphology**: infiltrating duct carcinoma, nos
- **Topography**: breast

Experiments [ ] all

Tumor type [lung] [ ] all > lung

Genes / Modules [ ] all > Genes

Details | Genes | Modules... | Tumor type | Experiments

<< Search

Filters

Filter Ensembl...

Upload filter...

Predictions

0.0 — 1.0

CGPrio

p-value

0.00 — 0.05 — 1.01

Export

CSV file

Ensembl Genes		CGC	Predictions		Alterations: Mutations	CNA		Transcrip	
Ensembl ID	Symbol	cmg	Onco	Supp	Fnd / Std	Loss	Gain	Down	Up
ENSG00000138413	IDH1	Dom			0 / 11				
ENSG00000147889	CDKN2A	Rec			300 / 1896				
ENSG00000136718	IMP4								
ENSG00000168542	COL3A1								
ENSG00000135446	CDK4	Dom			0 / 70				
ENSG00000145604	SKP2				0 / 11				
ENSG00000146648	EGFR	Dom			4312 / 16115				
ENSG00000135763	KIAA0133								
ENSG00000066248	NGEF				0 / 12				
ENSG00000121879	PIK3CA	Dom			29 / 964				
ENSG00000196712	NF1	Rec			0 / 154				
ENSG00000081479	LRP2								
ENSG00000139687	RB1	Rec							
ENSG00000141510	TP53	Rec			18 / 23				
ENSG00000146350	C6orf170								
ENSG00000165953	SERPINA12				1 / 11				
ENSG00000128641	MYO1B								
ENSG00000146247	PHIP								
ENSG00000162761	LMX1A								
ENSG00000175065	DSG4								
ENSG00000174586	ZNF497								
ENSG00000104499	GML				0 / 11				
ENSG00000198822	GRM3								
ENSG00000159387	IRX6								
ENSG00000169432	SCN9A								
ENSG00000171862	PTEN	Rec			48 / 548				

Showing 1 to 26 of 37

<< < 1 2 > >>

Prototype

Experiments — all

Tumor type — breast — all > lung

Genes / Modules — all > Genes

Details | Genes | Modules... | Tumor type | Experiments

Search

Filters
Upload filter...

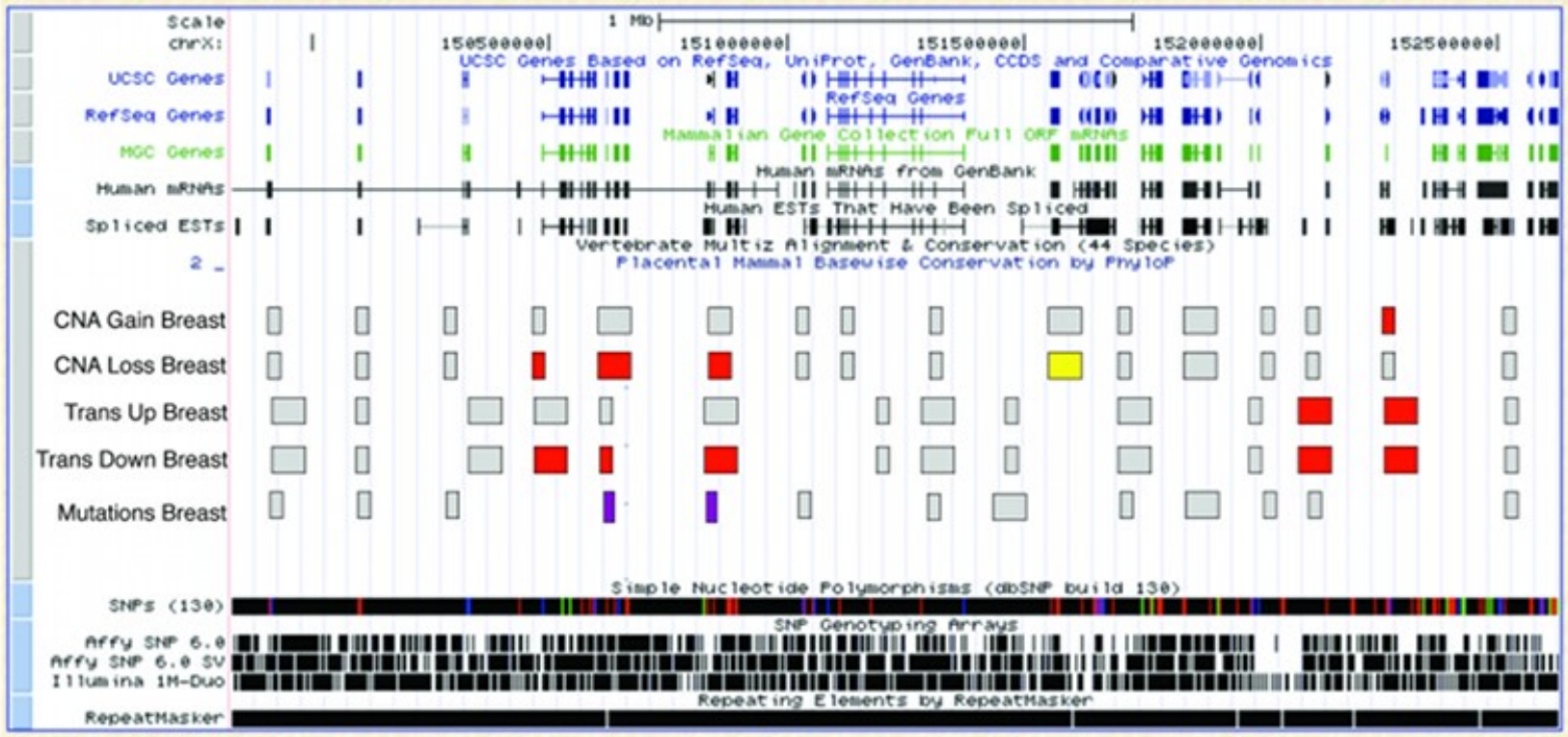
Predictions
0.0 1.0
CGPrio

p-value
0.00 0.05 1.00

Export
CSV file

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chrX:149,829,362-152,627,668 jump clear size 2,798,307 bp. configure

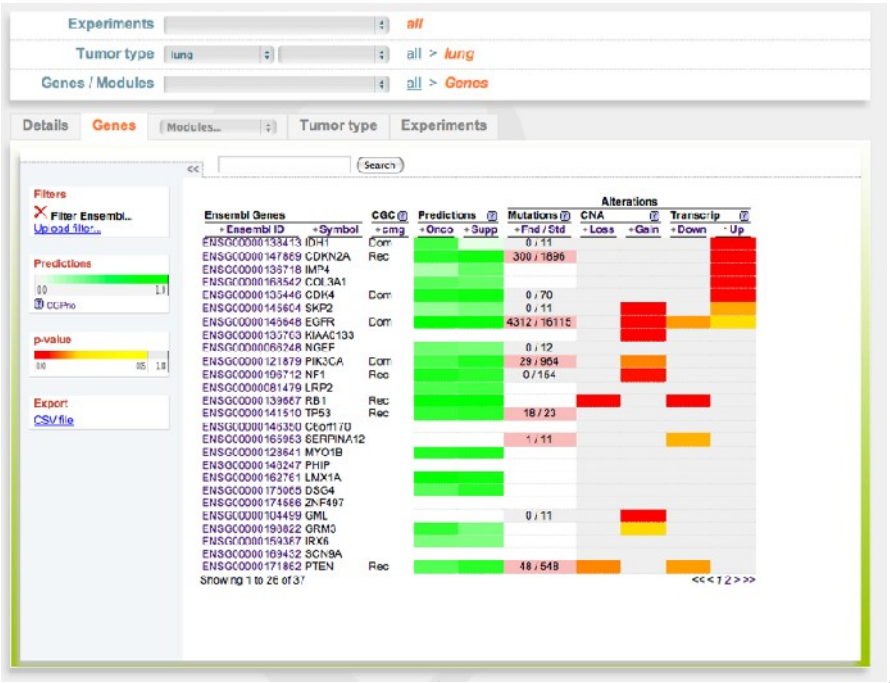


move start < 2.0 >

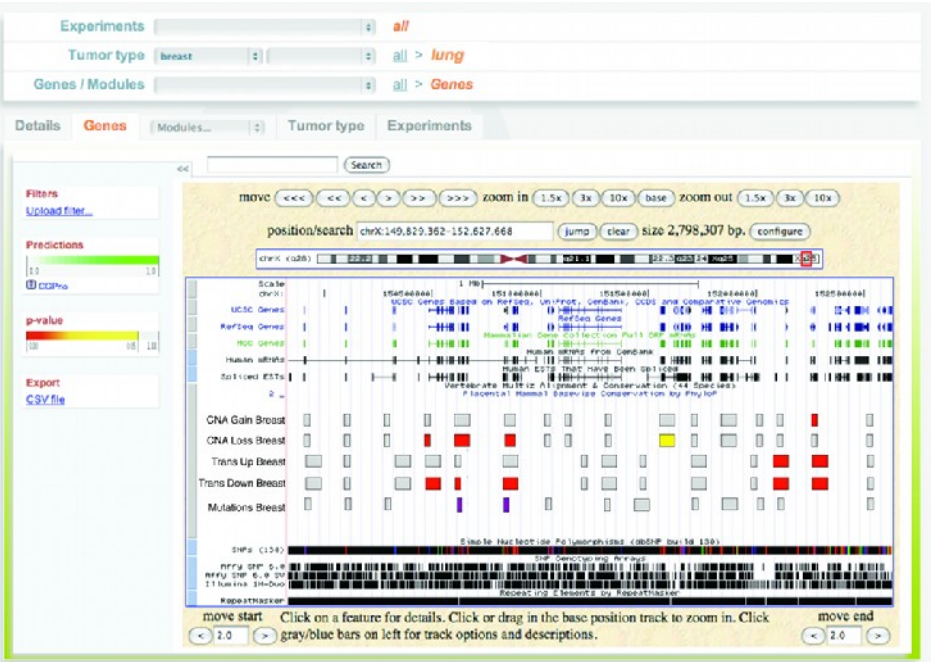
Click on a feature for details. Click or drag in the base position track to zoom in. Click gray/blue bars on left for track options and descriptions.

move end < 2.0 >

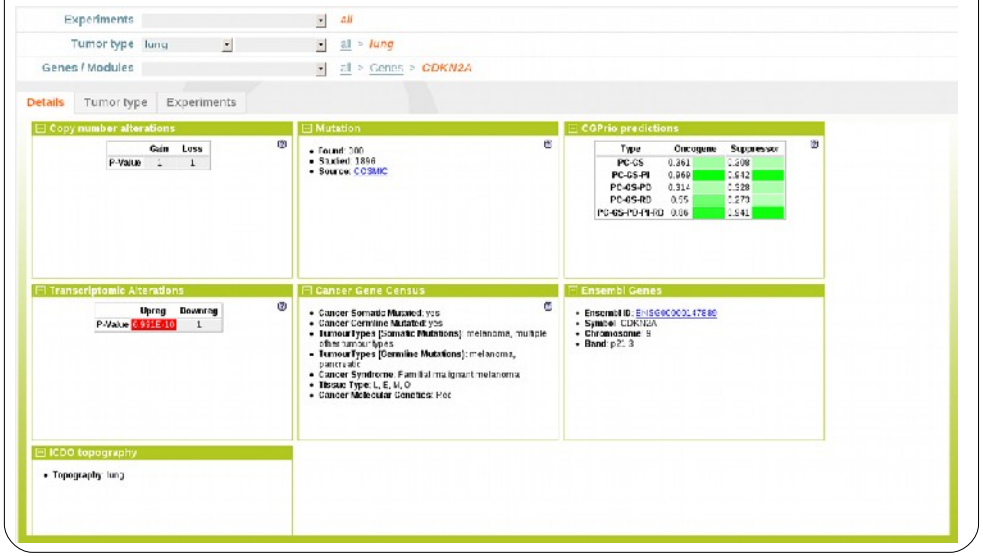
Table view



Genome view



Details view



Concluding remarks

- IntOGen is a novel framework for **Oncogenomics data** integration and analysis
- It **integrates** many tumor types and different types of alterations in a common framework
- It explores the data at **different levels**, from individual experiments to combinations of experiments, and from individual genes to biological modules
- The system is updatable, **flexible** and extensible
- It incorporates an intuitive web system designed to be a **discovery tool** for cancer researchers (www.intogen.org)
- It will allow to store, analyze and visualize data coming from **next generation sequencing** technologies

Acknowledgements

Biomedical Genomics Group

Nuria López-Bigas

Gunes Gundem

Jordi Deu-Pons

Alba Jené-Sanz

Khademul Islam

Michael Schroeder

Sergi Hernandez

Simon J Furney
(University College London)