Integrative Genome-Wide Analysis in Glioblastoma

Marie de Tayrac

Regulation of Transcription and Oncogenesis - CNRS - Jean Mosser
Actual Position INSERM U946 - CEPH, Paris - FRANCE

Gliomas & Glioblastoma

- Highly **complex** & **heterogenous** diseases
- Tumors of the CNS, arising from **glia** - astrocytes & oligodendrocytes
- WHO **classification** - histologic criteria & grading of malignancy

Astrocytomas

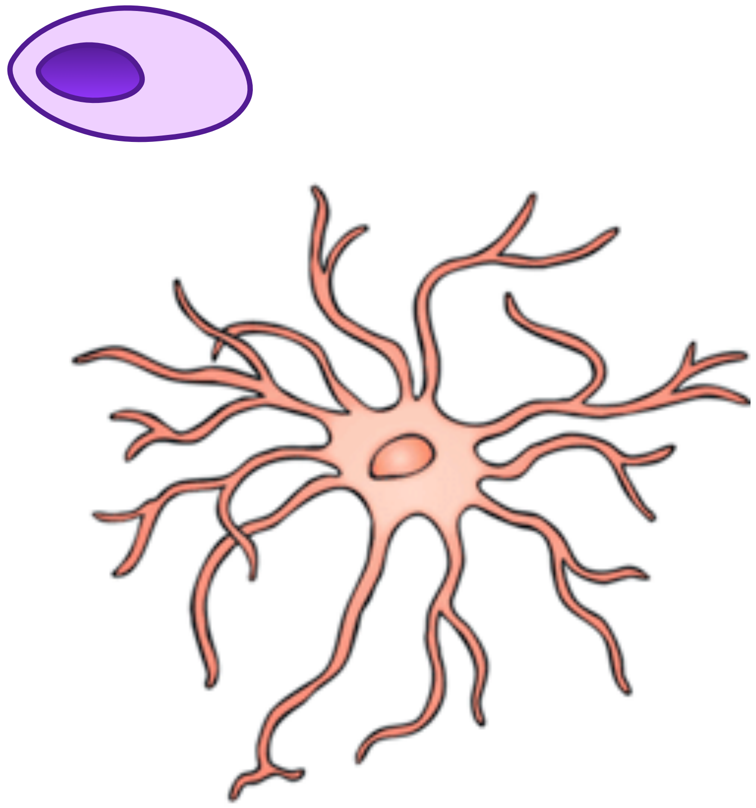
Pilocytic astro	**I**
Fibrillar astro	**II**
Anaplastic astro	**III**
Glioblastoma	**IV**

Oligodendrogliomas

Low-grade oligo	**II**
Anaplastic oligo	**III**

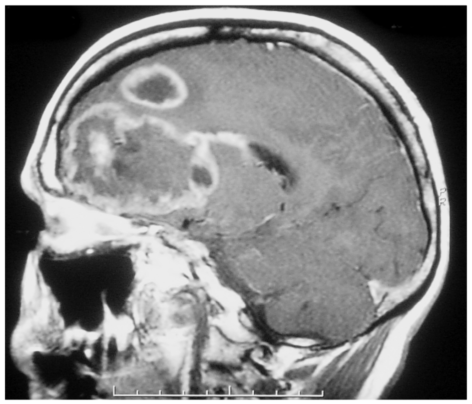
Mixed gliomas

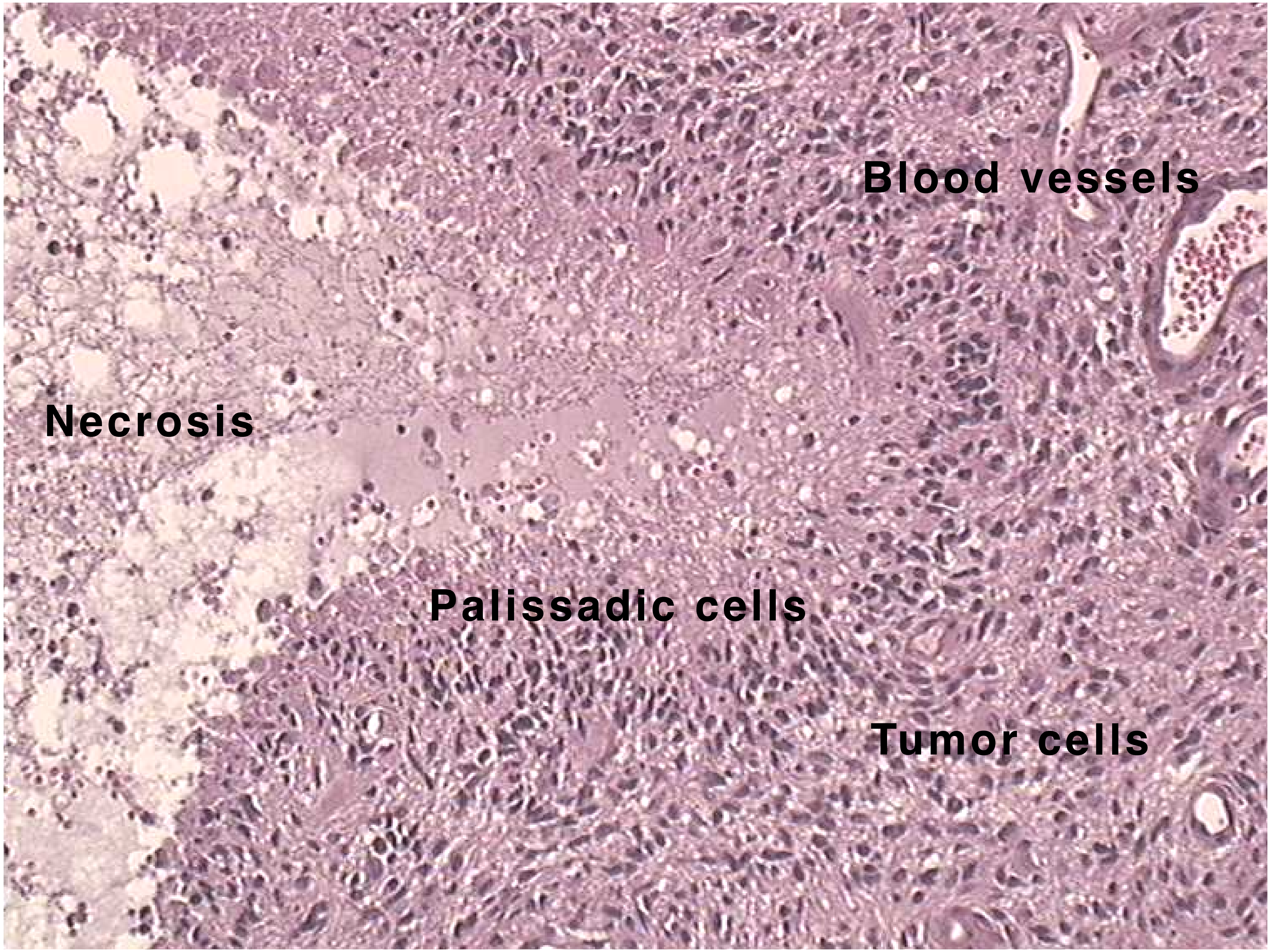
Low-grade oligo-astro	**II**
Anaplastic oligo-astro	**III**



- **Prognostic** ~14 mo with treatment (surgical resection and RT-TMZ)

Histology



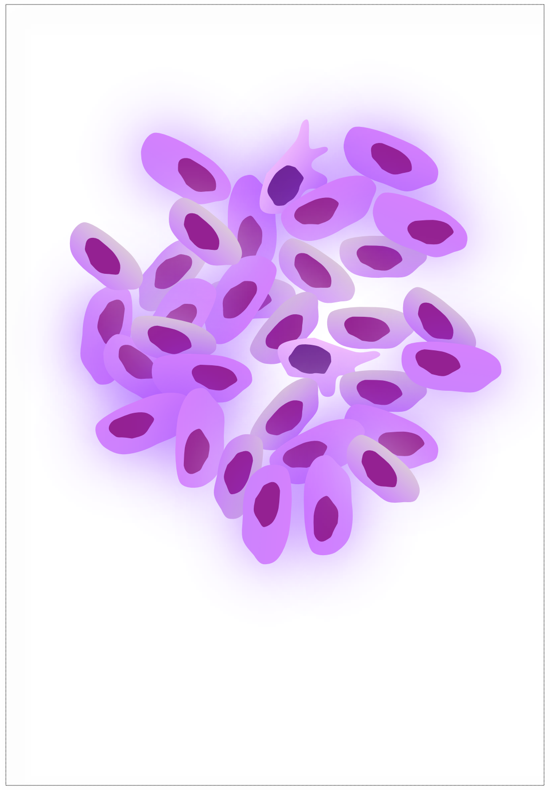


Multistep Model in Cancer

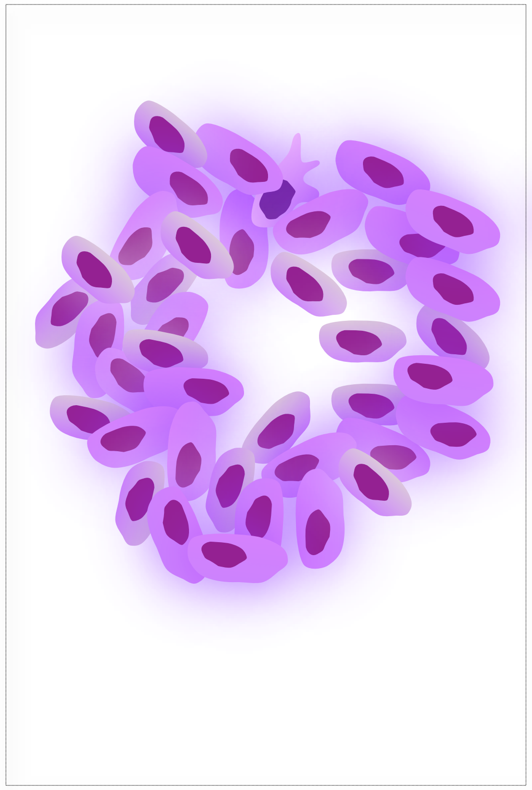
Phenotypic progression

↑↑↑malignant cells

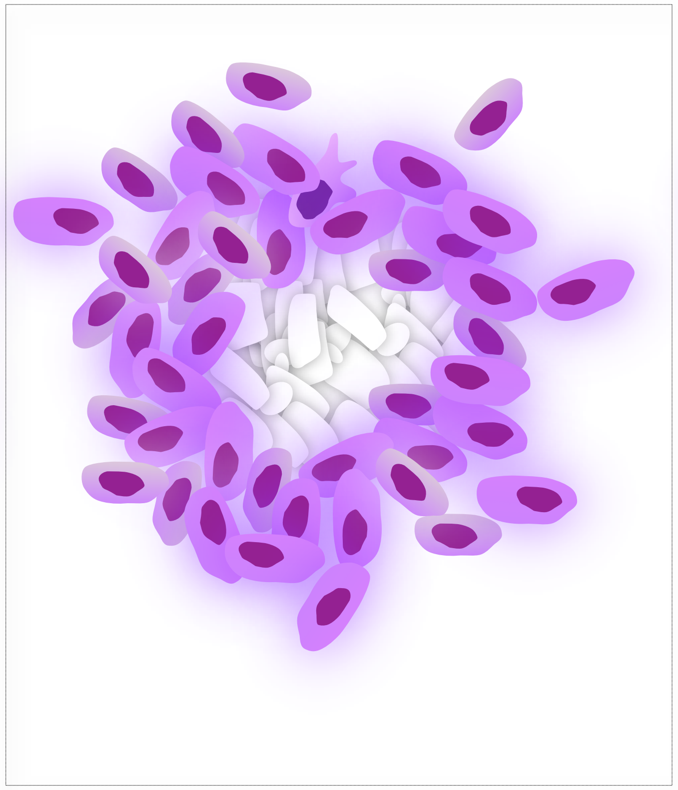
+ Hypoxia



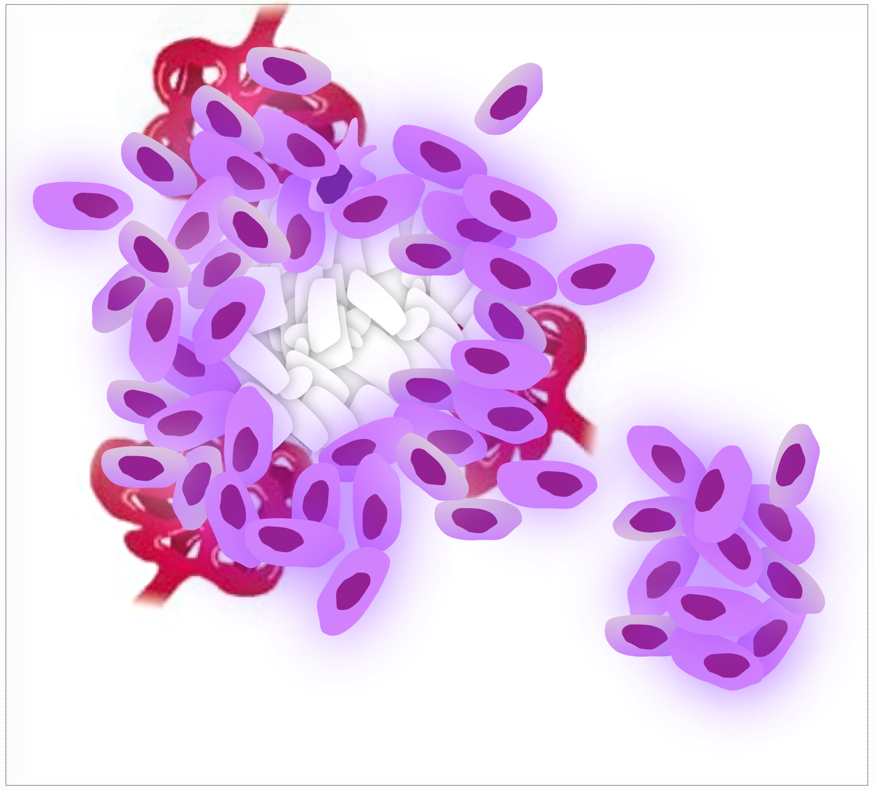
+ Migration



+ Necrosis



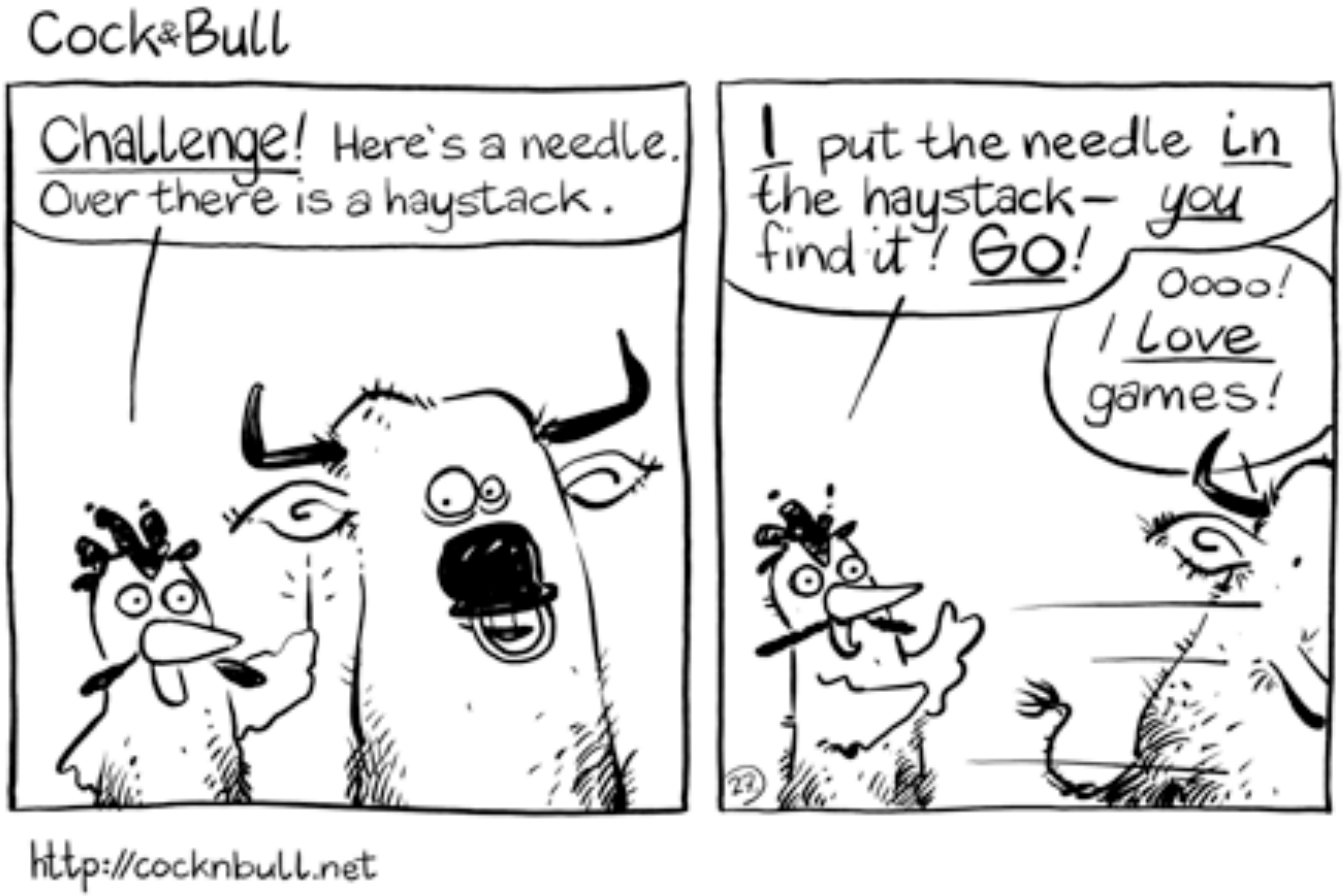
+ Angiogenesis



Genotypic progression

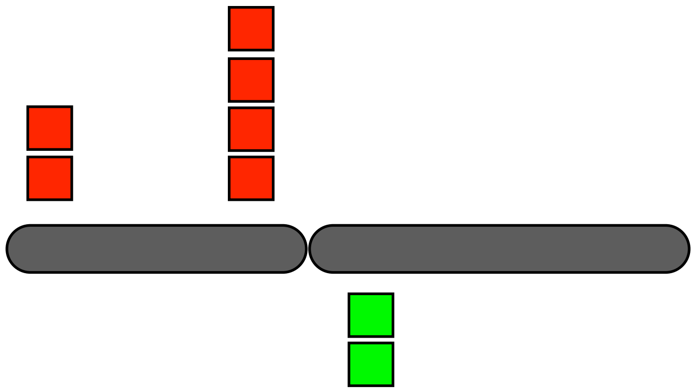
↑↑↑somatic mutations

Looking for a needle in a haystack

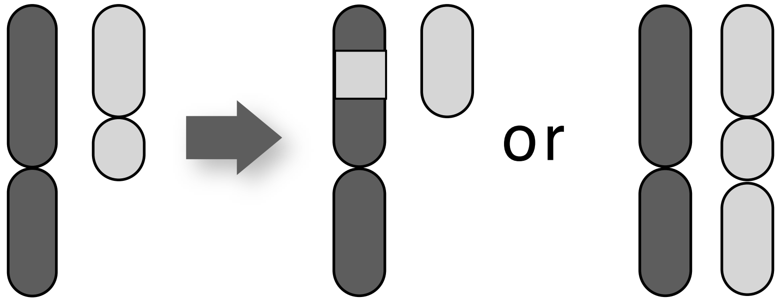


Cancer alterations

Copy number

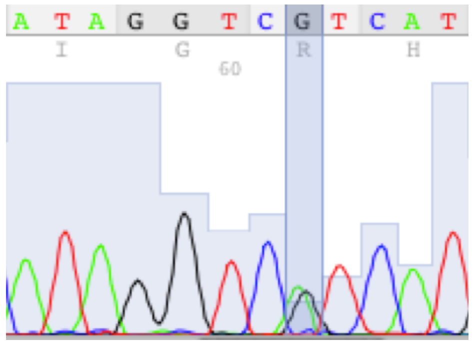


Rearrangements



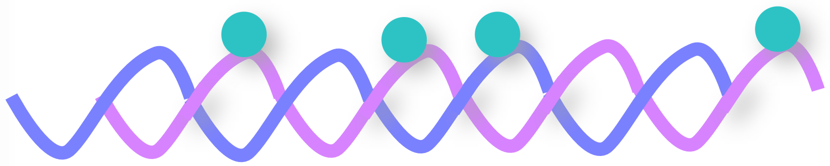
inter & intra

Mutations

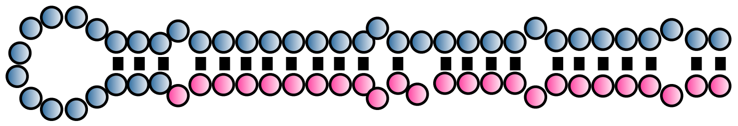


substitution & indels

Epigenetic

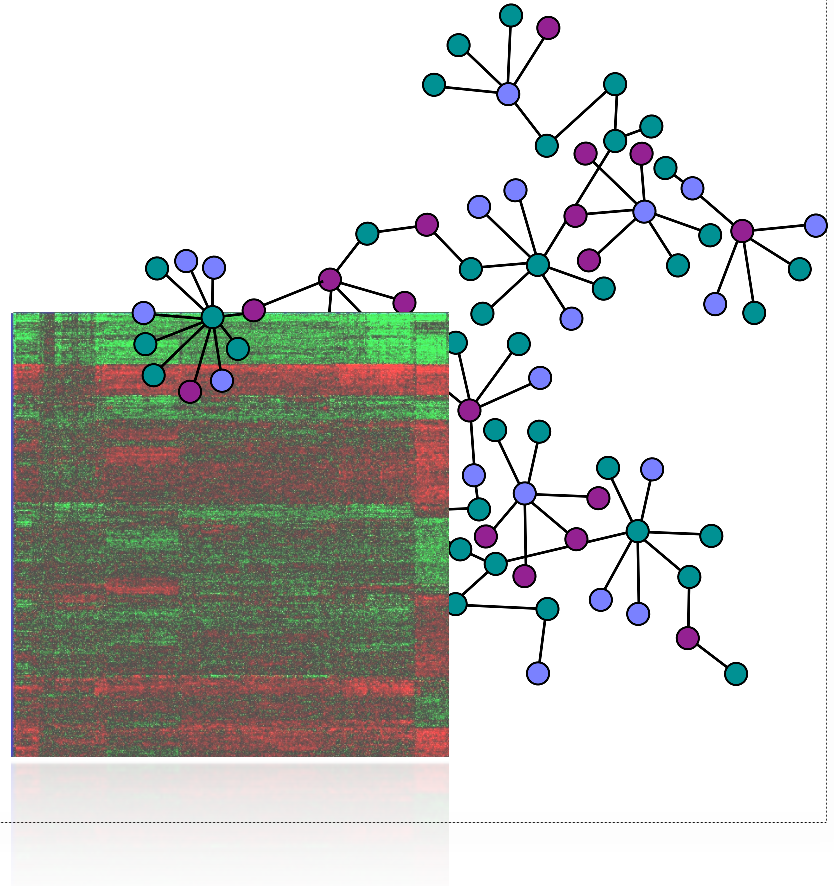


CpG islands

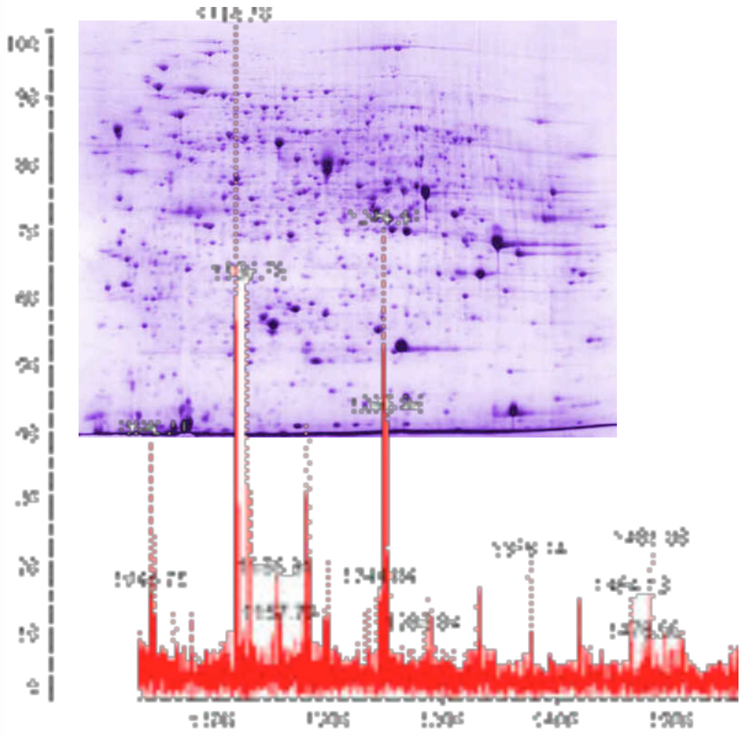


miRNA

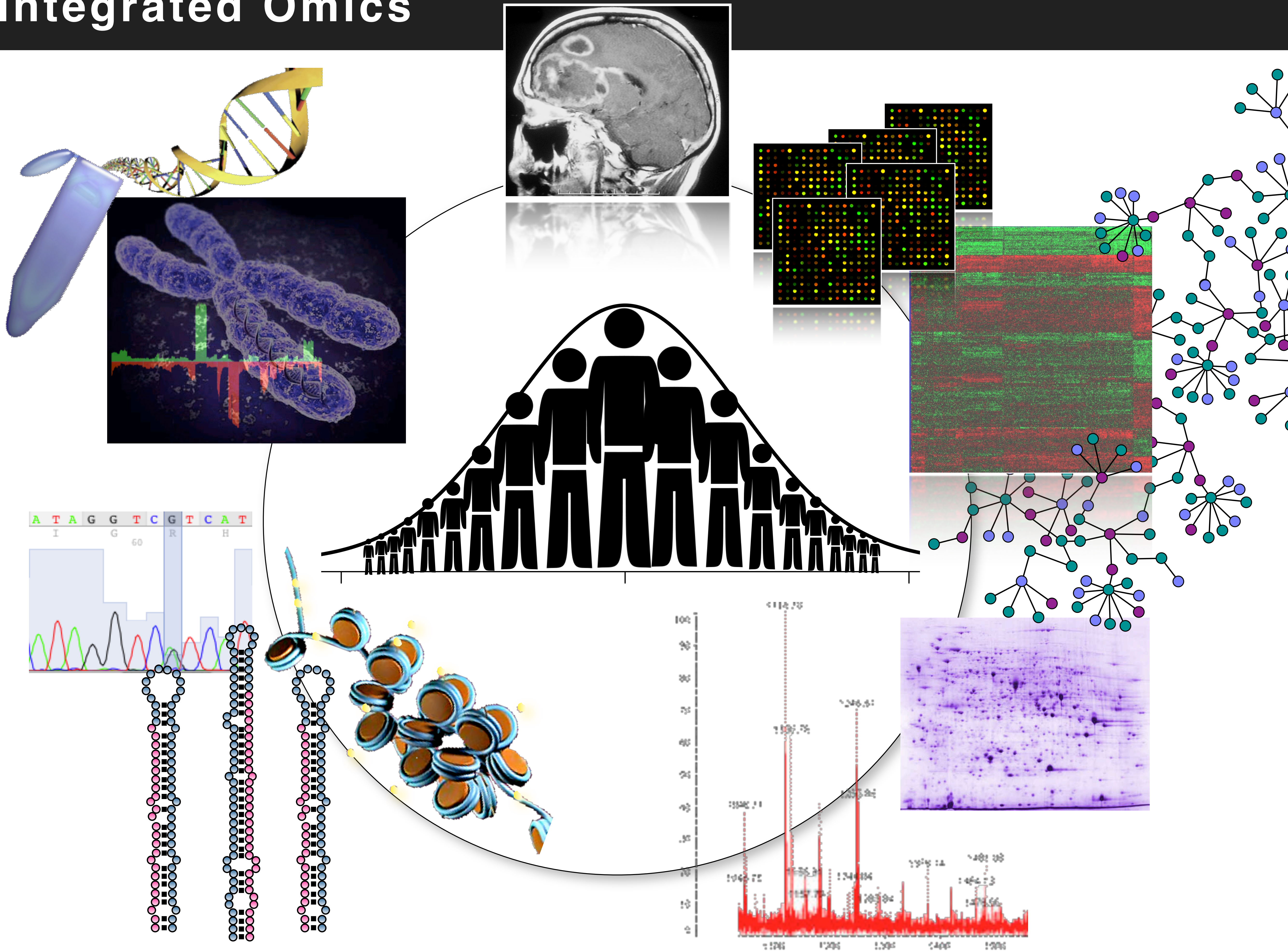
Transcriptome



Proteome



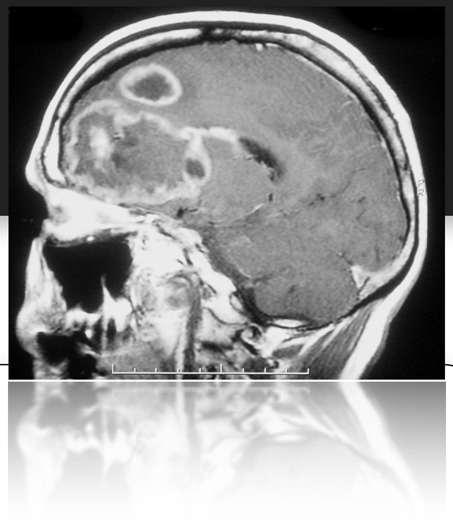
Integrated Omics



This Talk

- Glioblastoma & molecular alterations
- Integrating the genome and the transcriptome
- Adding biological knowledge
- Another way to integrate: meta-analysis

Context

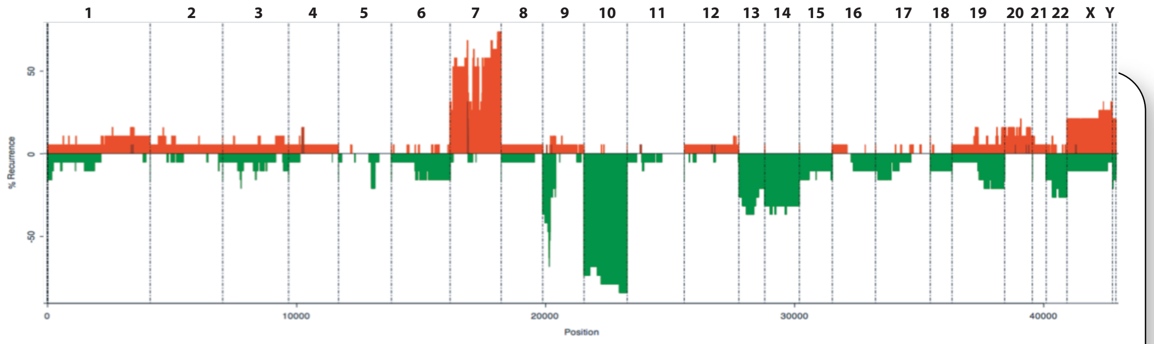


Glioblastoma

- **Complex** and **heterogeneous** disease...

  ... at both **clinical** and **molecular** levels

‘Omics’ analyses



Huge number of regions throughout the genome are altered - LOH10q, EGFR amplification, TP53 & PTEN & IDH1 mutations, NF1 alteration, MDM2/4 amplifications

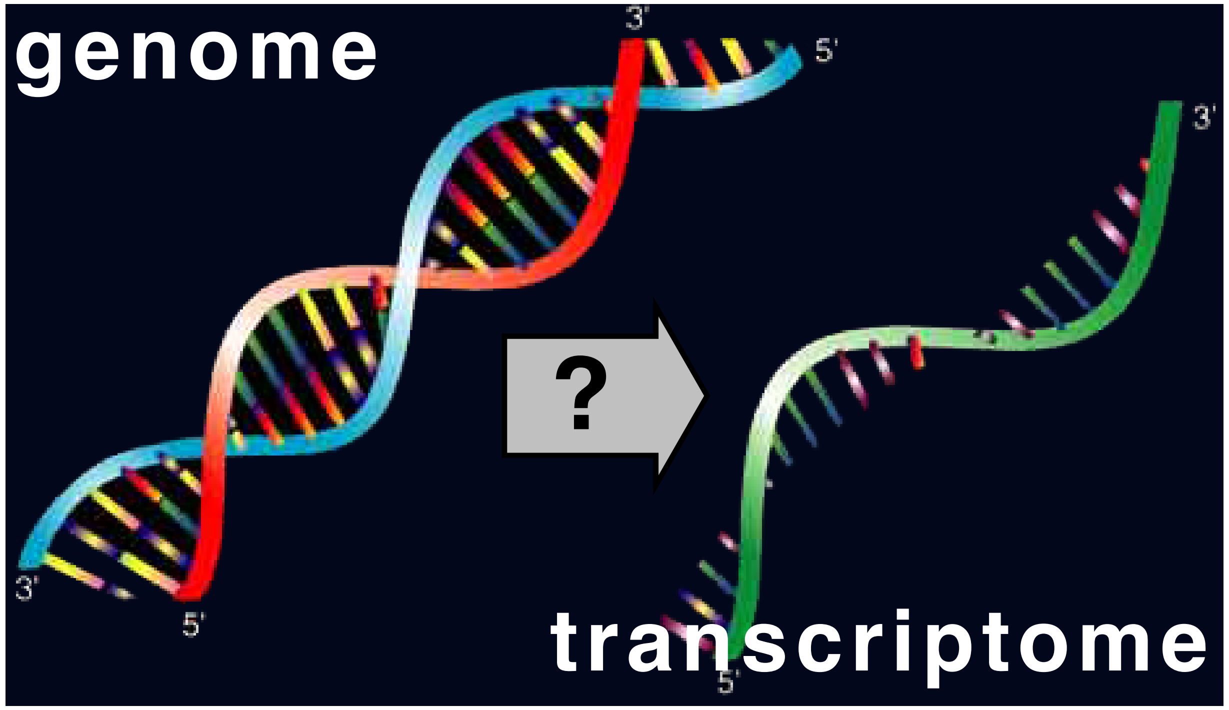
Gene expression changes in key signalling pathways - RB, TP53, PI3K/Akt, Ras/MAPK...

Up to 50% of the genes in the human genome are altered

Which changes are important?

Are their relationships also critical?

Paired genome-transcriptome wide analysis



Finding among the huge amount of copy number alterations of the genome those that will have a direct impact on the gene expression

Which are the losses that drive an under-expression of associated genes?

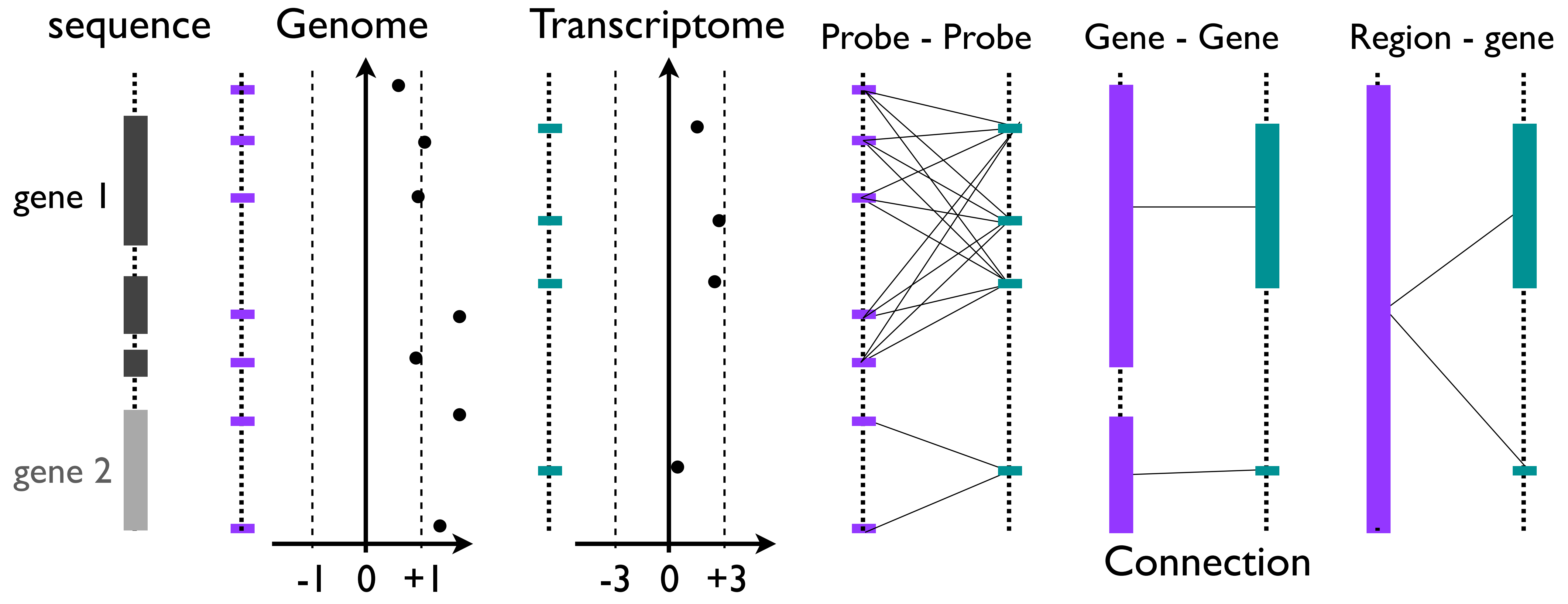
Which are the duplication and amplification that drive an over-expression of associated genes?

How to integrate 'Omics' data?

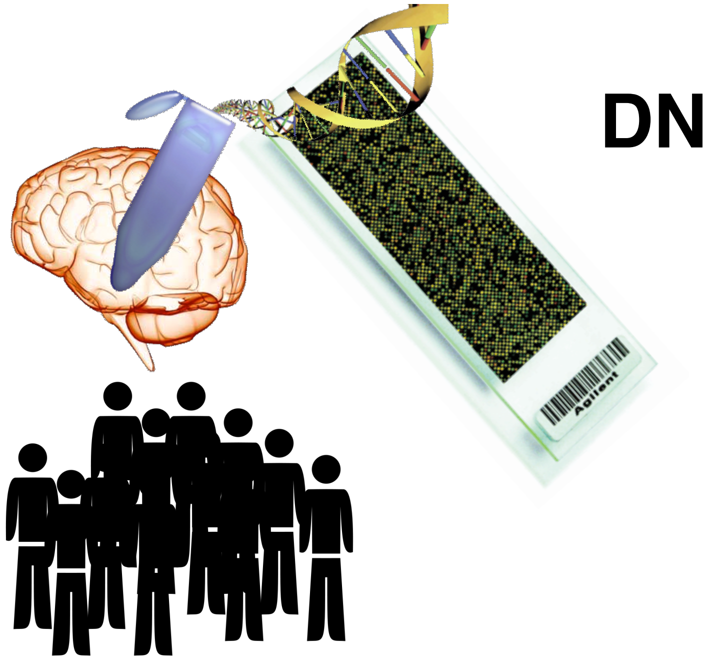
Experimental Design

- **Case selection -** Particularly true for heterogeneous disease
- **Technology -** Including control sample is necessary

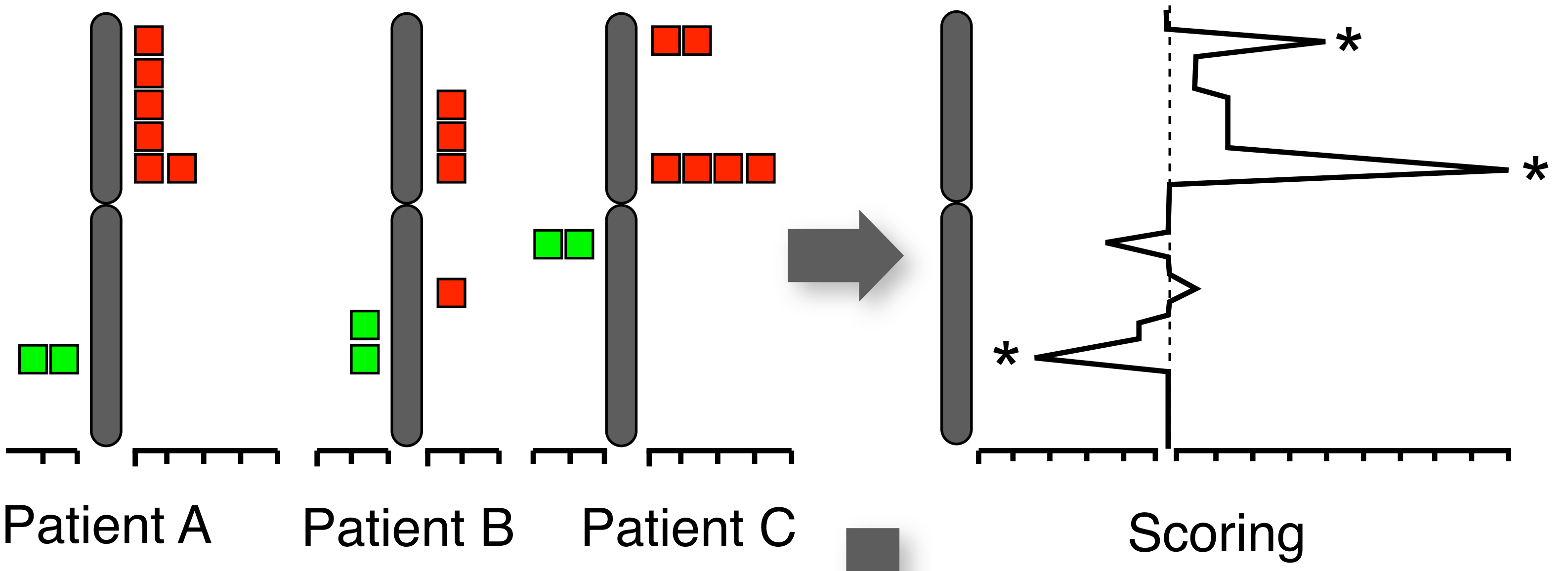
Matching data



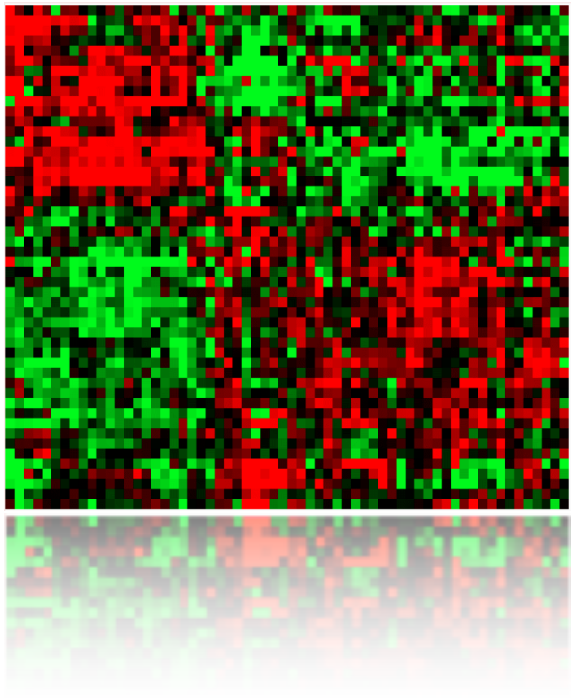
Principle

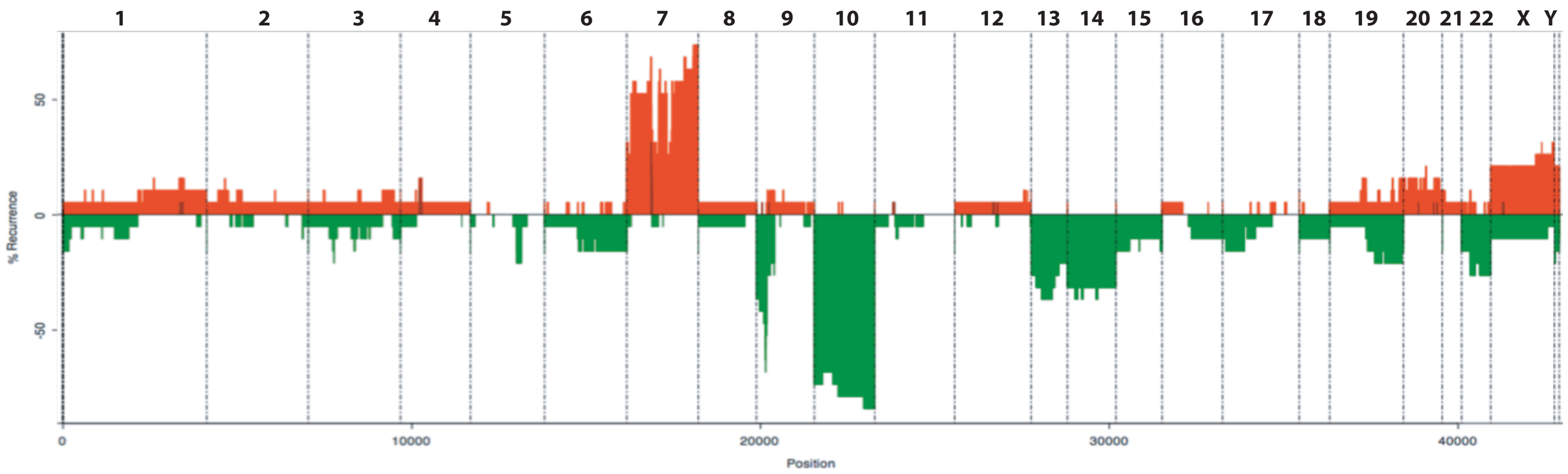


DNA



RNA





On the same tumor samples ➡ paired data

With control samples ➡ comparable data

Non-integrative results

Results

Genome

4816 alterated genes (more than two times)

on 15 chromosomes

Transcriptome

1888 genes ($FC>2$, $p<0.01$)

implicated in 921 Biological Processes

➡ **How to find biomarkers?**

Two approches

GENES, CHROMOSOMES & CANCER 48:55–68 (2009)

Integrative Genome-Wide Analysis Reveals a Robust Genomic Glioblastoma Signature Associated with Copy Number Driving Changes in Gene Expression

Marie de Tayrac,[1,2] Amandine Etcheverry,[3] Marc Aubry,[3] Stephan Saïkali,[4,5] Abderrahmane Hamlat,[6]
Veronique Quillien,[1,7] André Le Treut,[1,2] Marie-Dominique Galibert,[1,2] and Jean Mosser[1,2,3*]

- **Finding correlations**

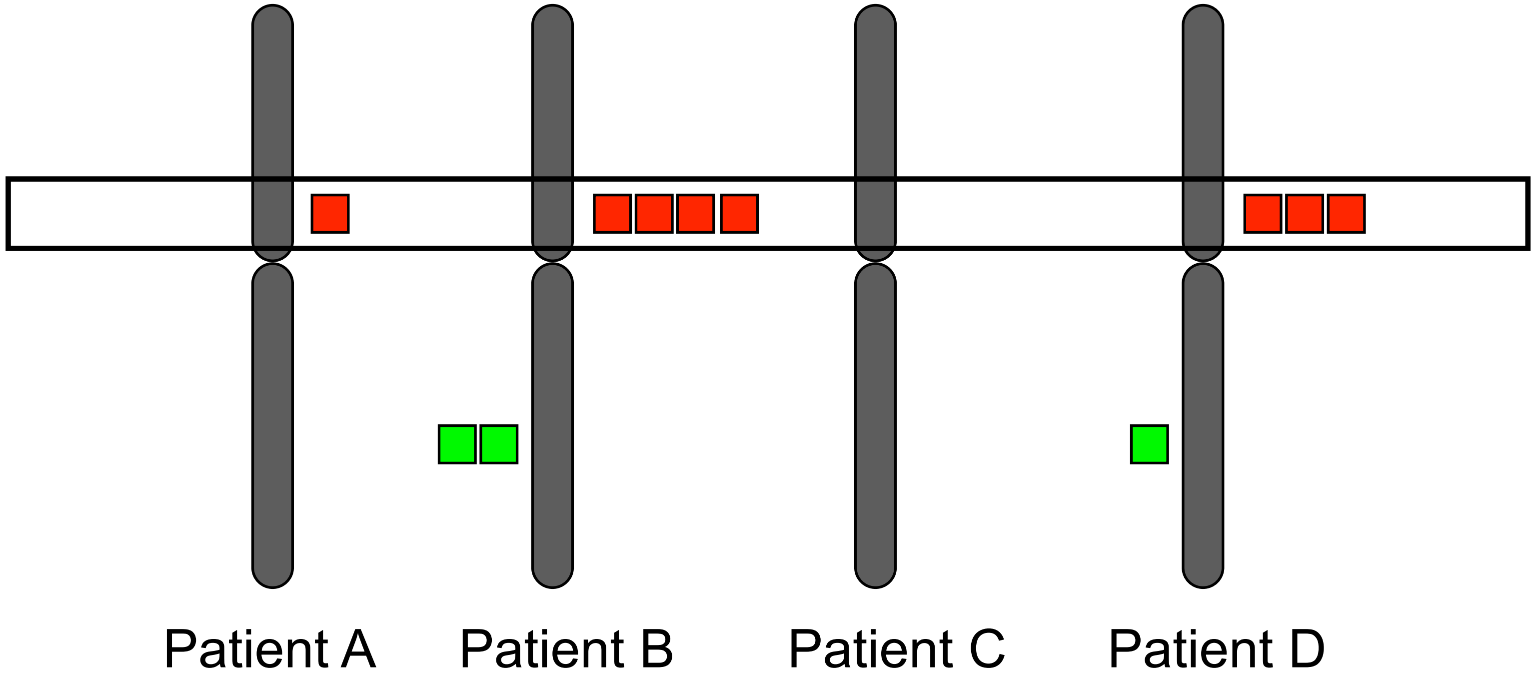
  Genes whose expression is *tightly correlated* with the genome alterations

- **Finding concordances**

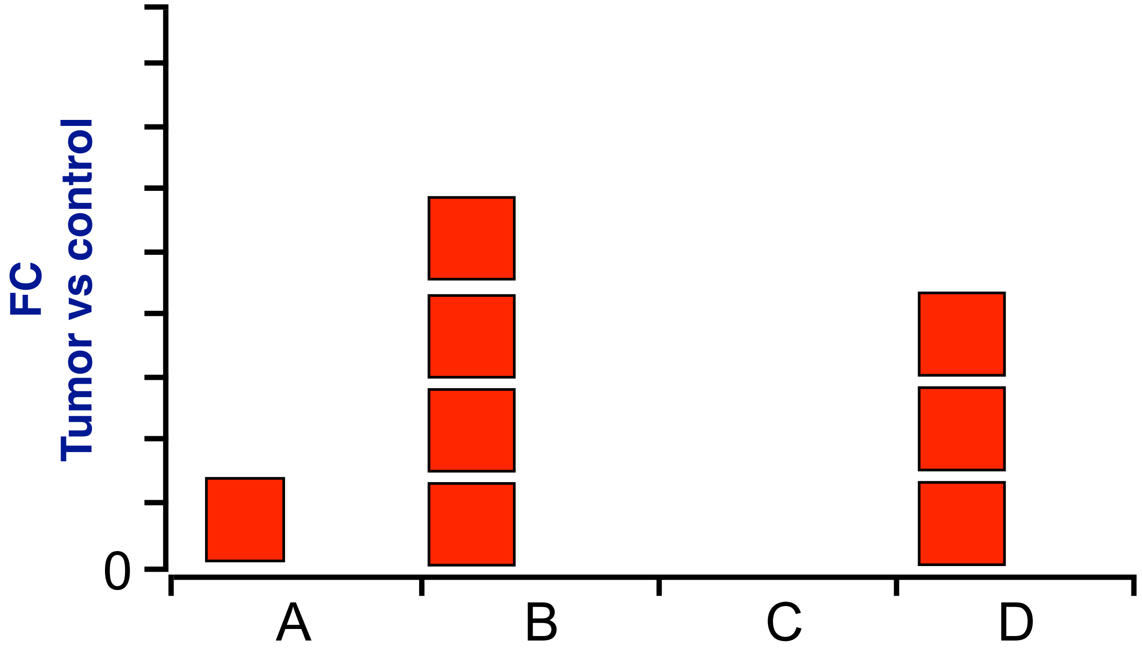
  Genes whose expression is *significantly* modified in *accordance* with the genome alterations

Method

CGH-array analysis

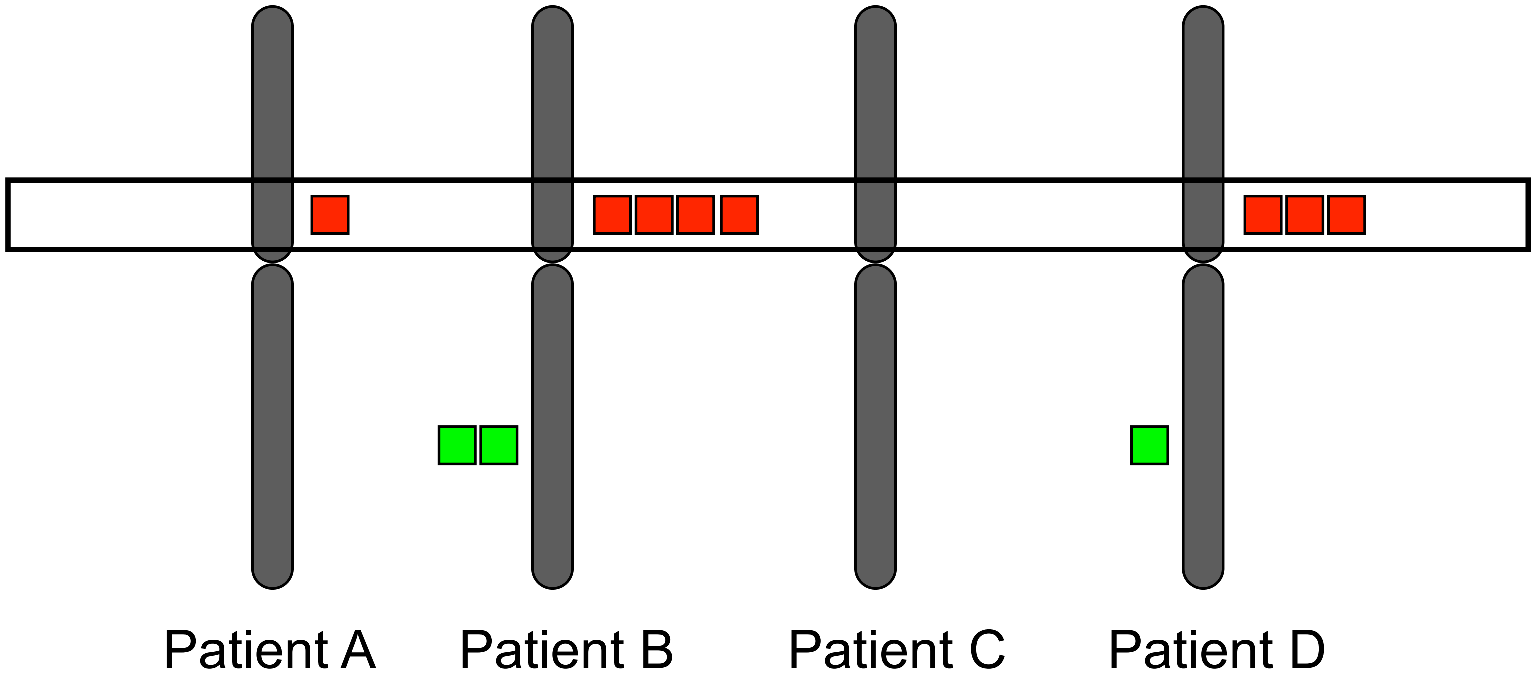


Finding correlations

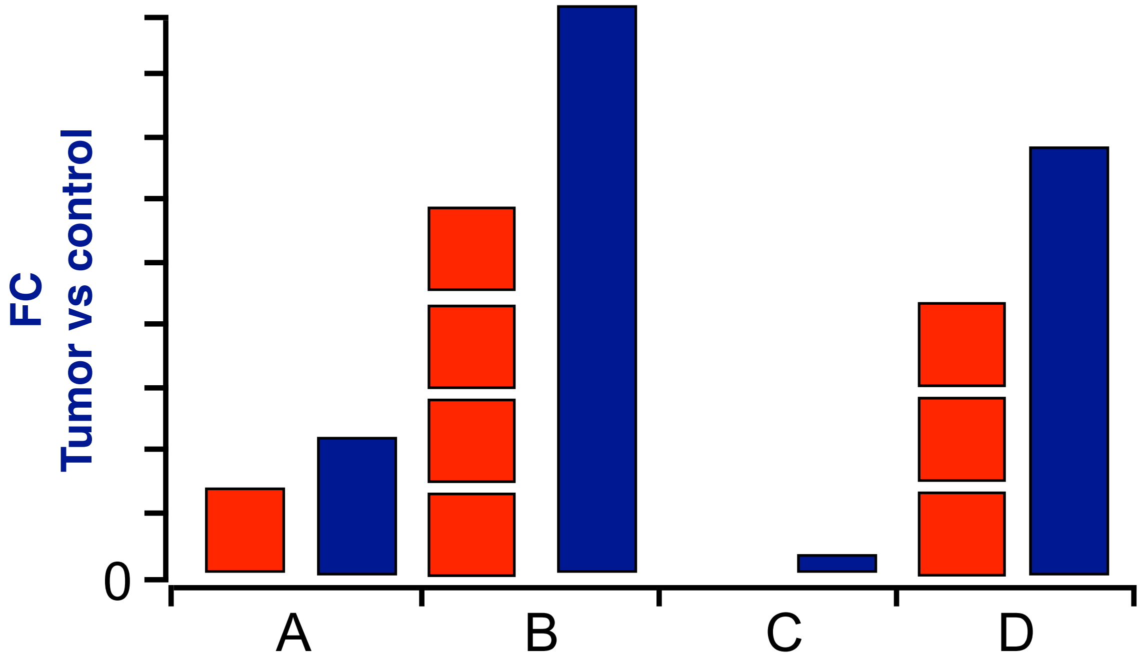


Method

CGH-array analysis



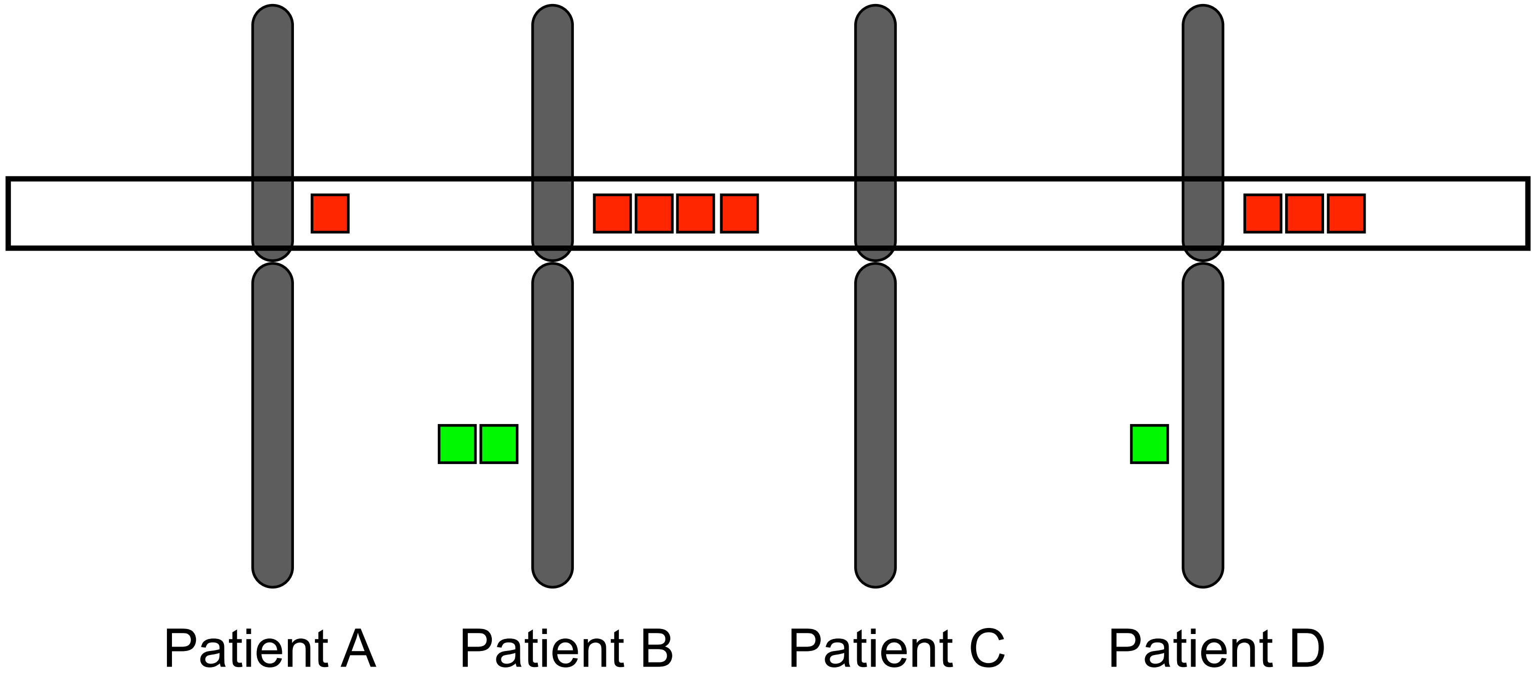
Finding correlations



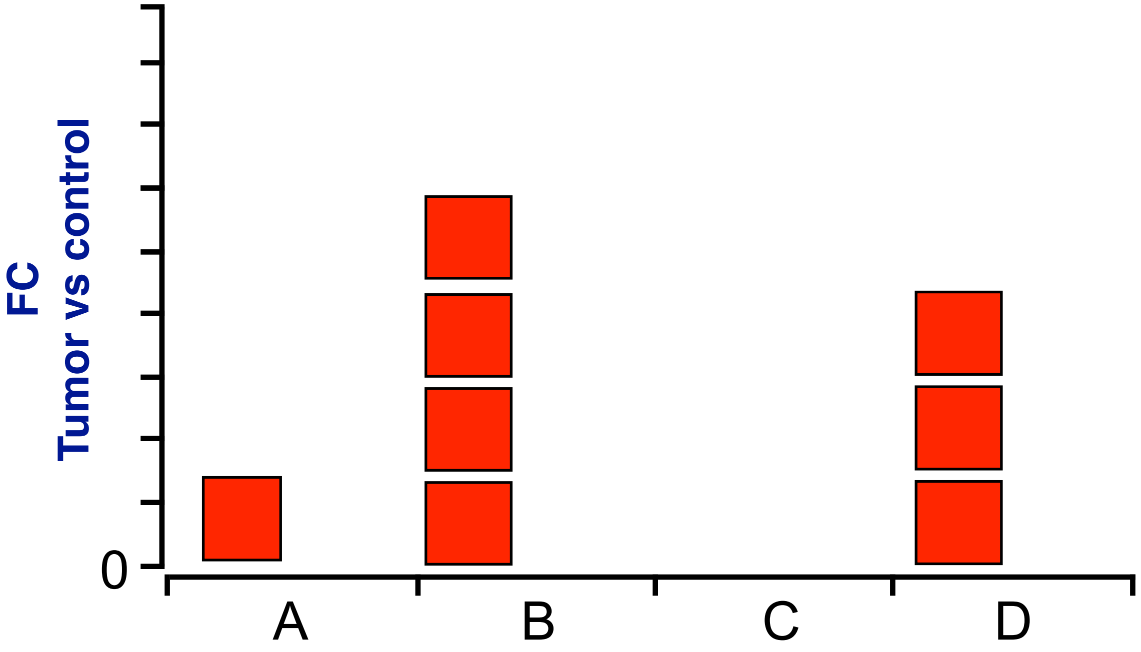
Pearson coefficient > 0.7

Method

CGH-array analysis



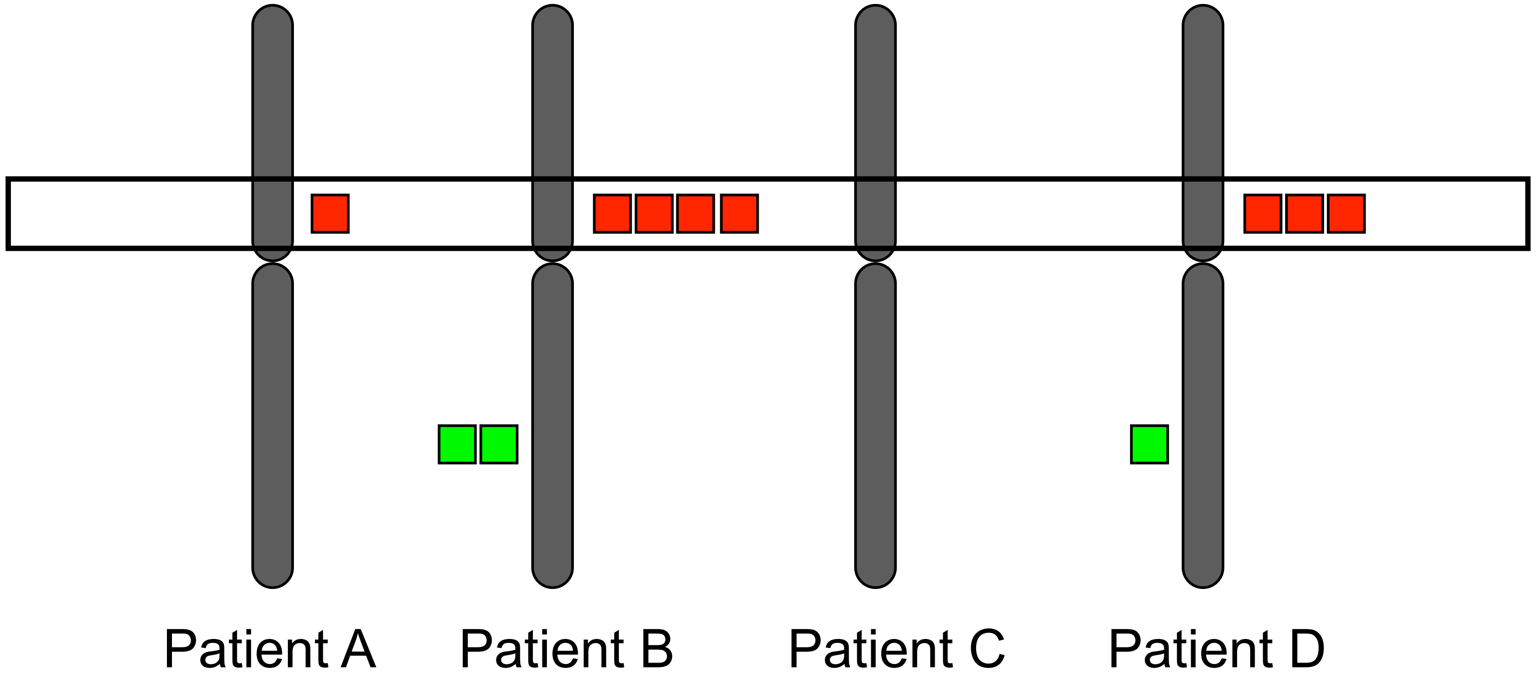
Finding concordances



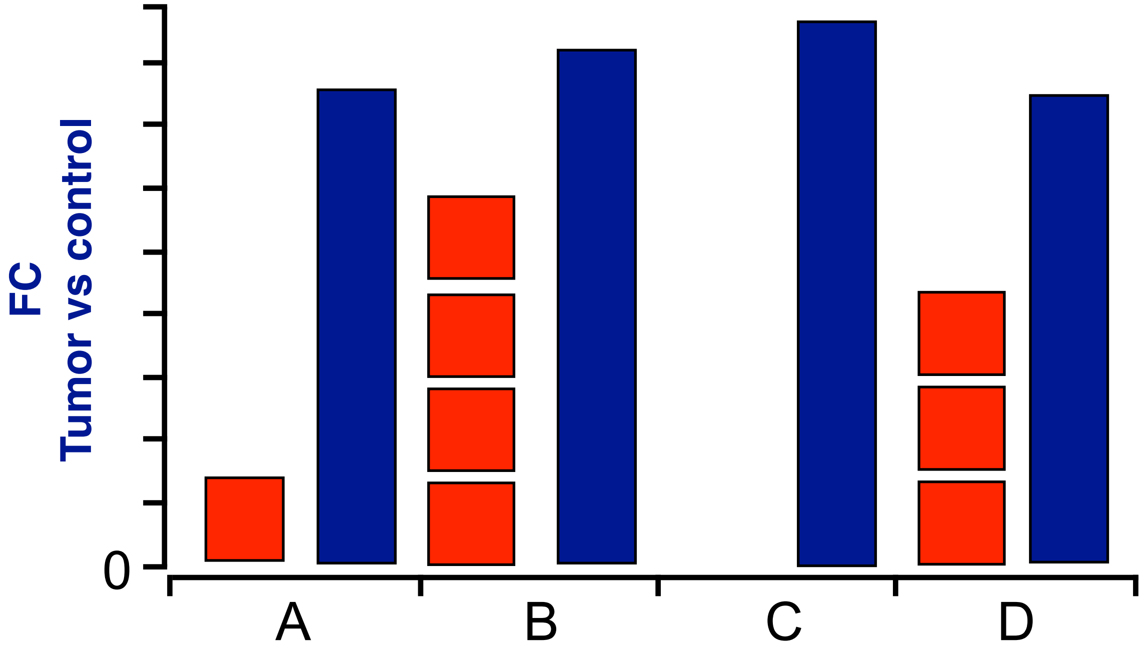
p-value < 0.05 / FC > 2

Method

CGH-array analysis



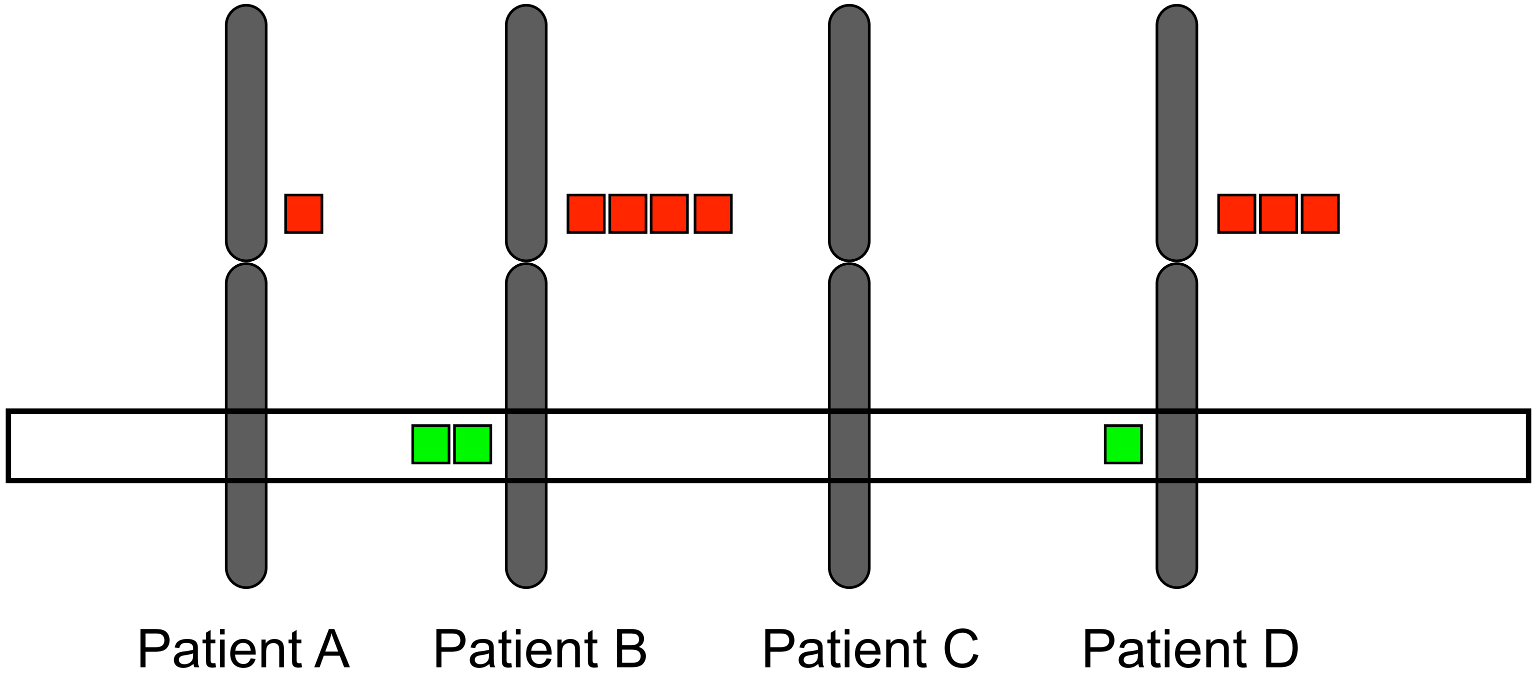
Finding concordances



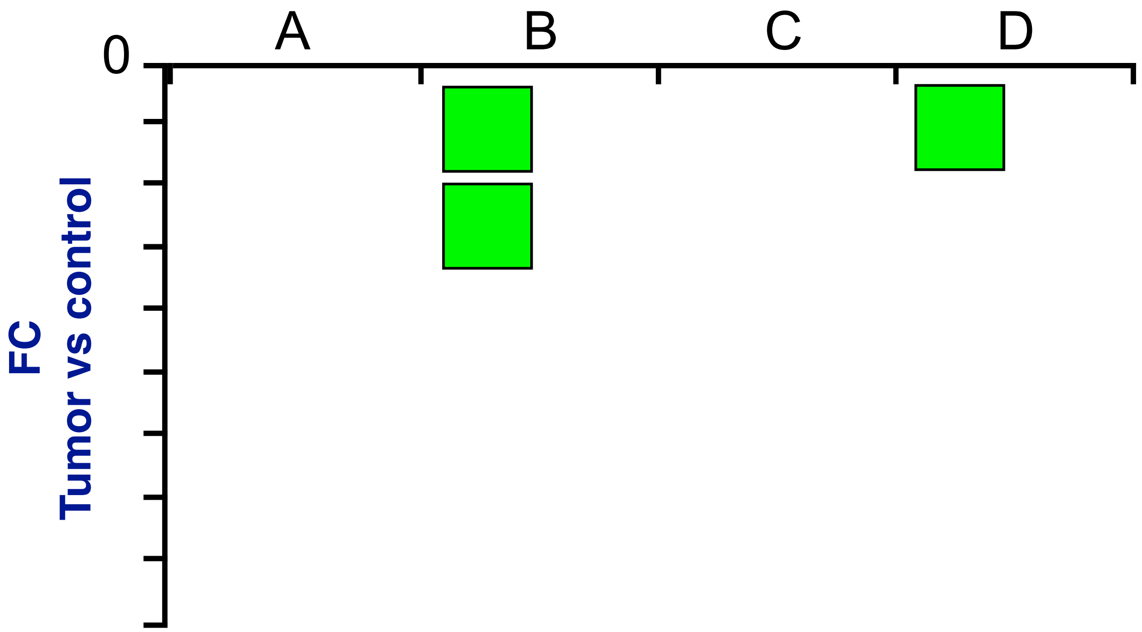
p-value < 0.05 / FC > 2

Method

CGH-array analysis



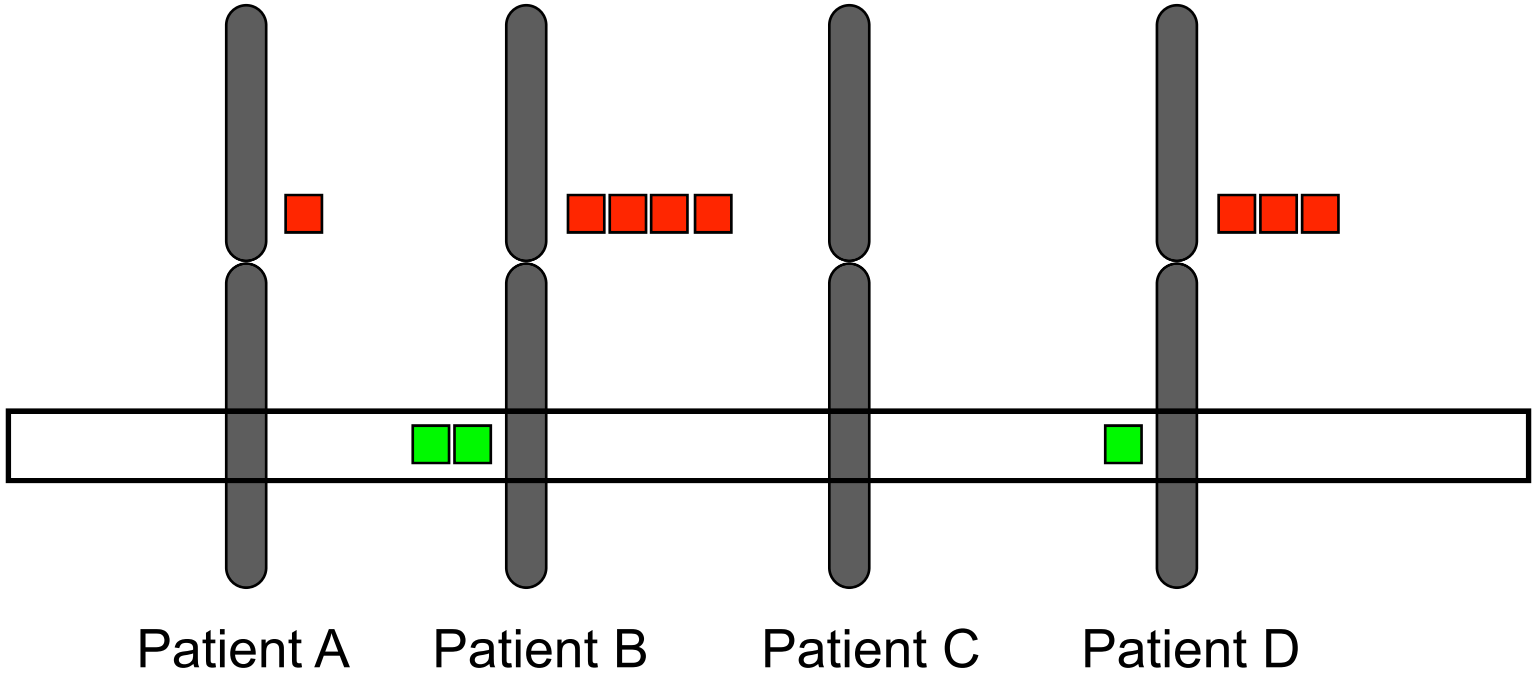
Finding correlations



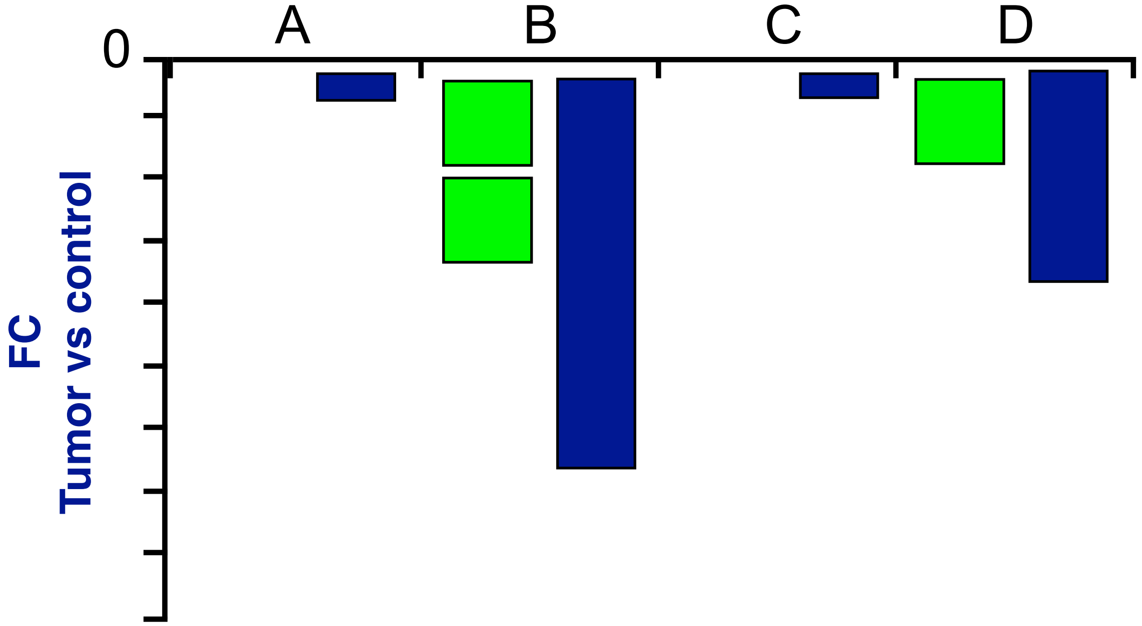
Pearson coefficient > 0.7

Method

CGH-array analysis



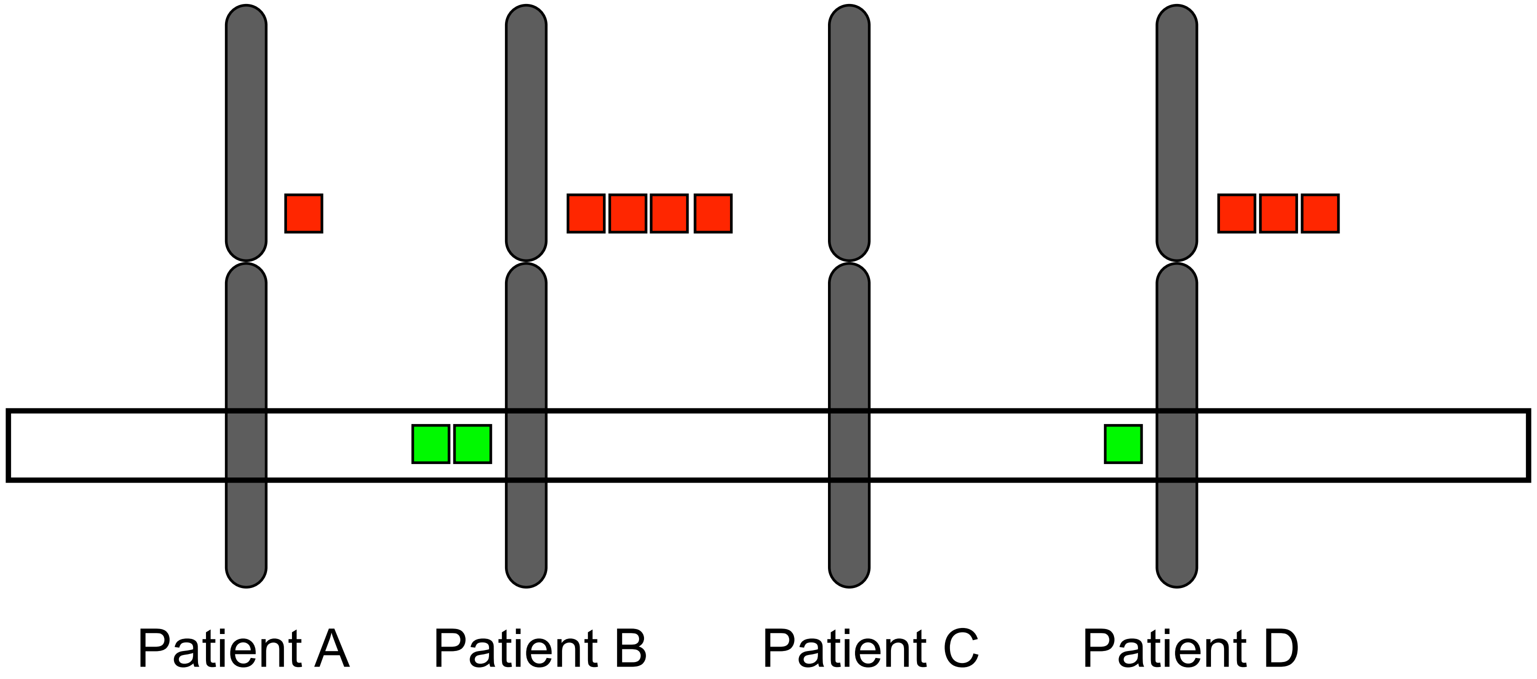
Finding correlations



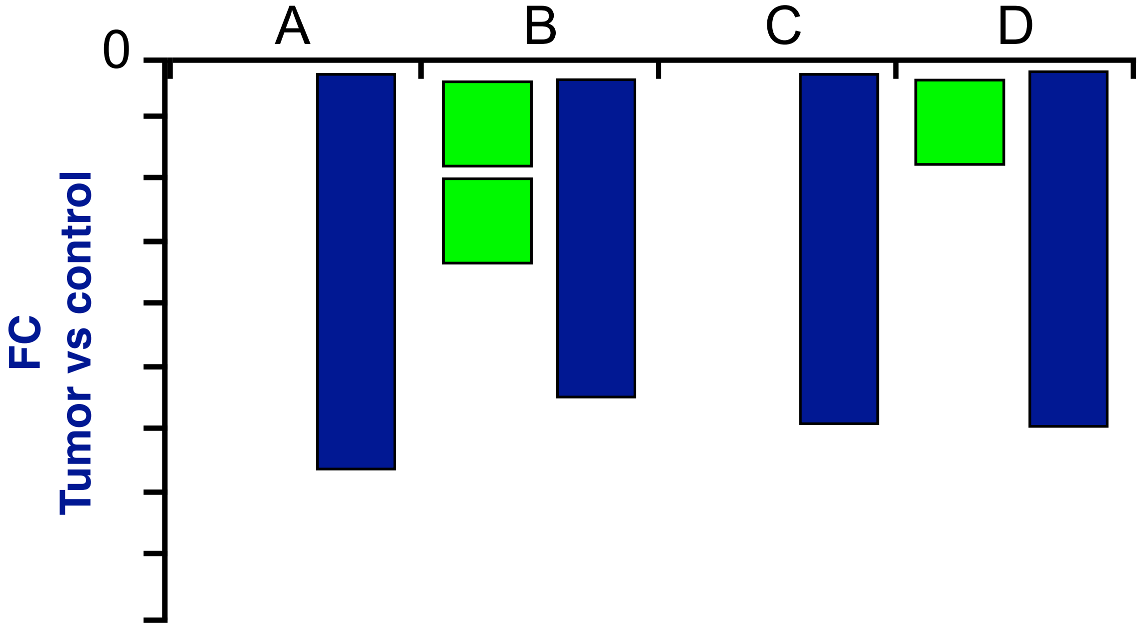
Pearson coefficient > 0.7

Method

CGH-array analysis

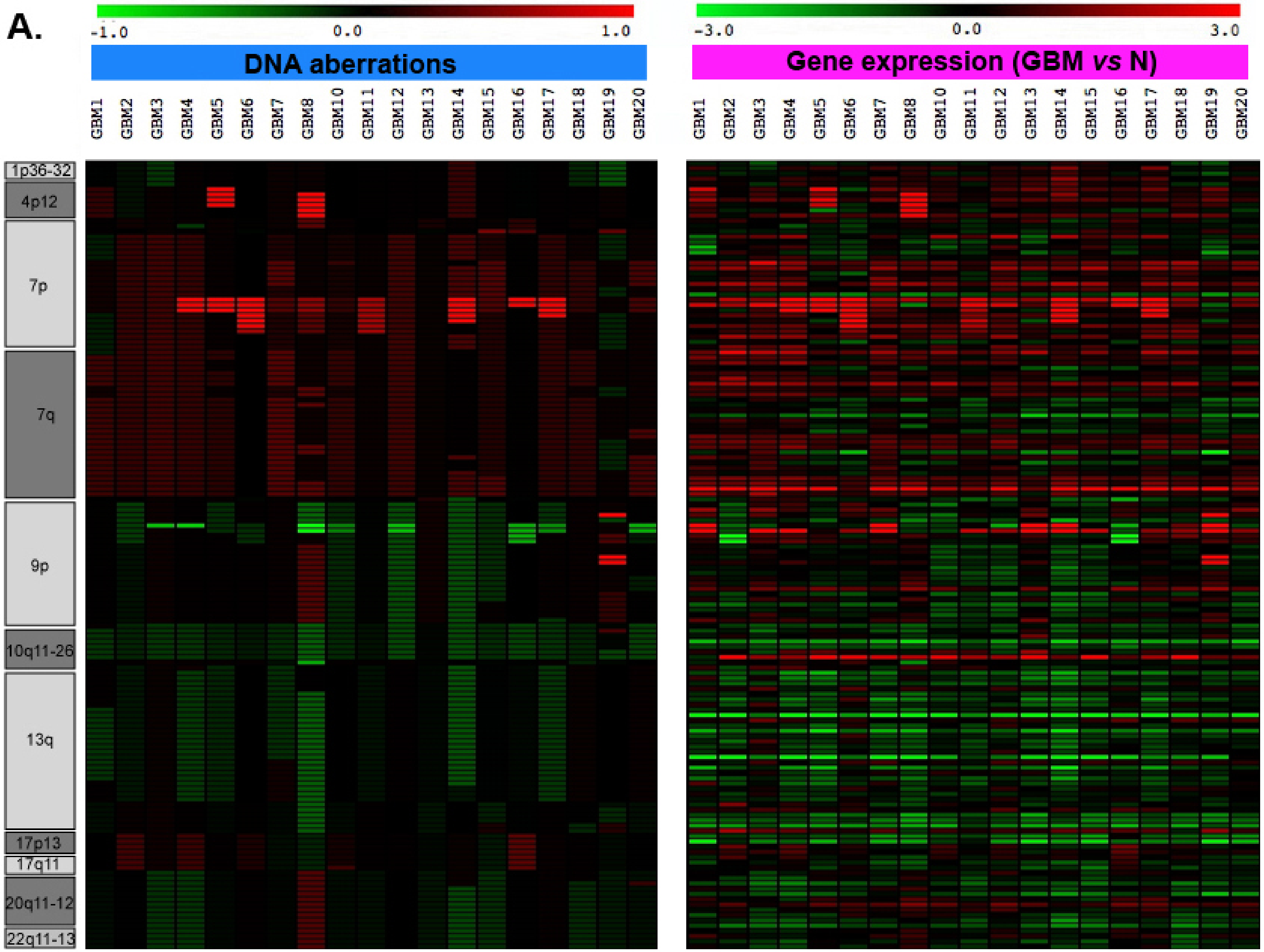


Finding concordances



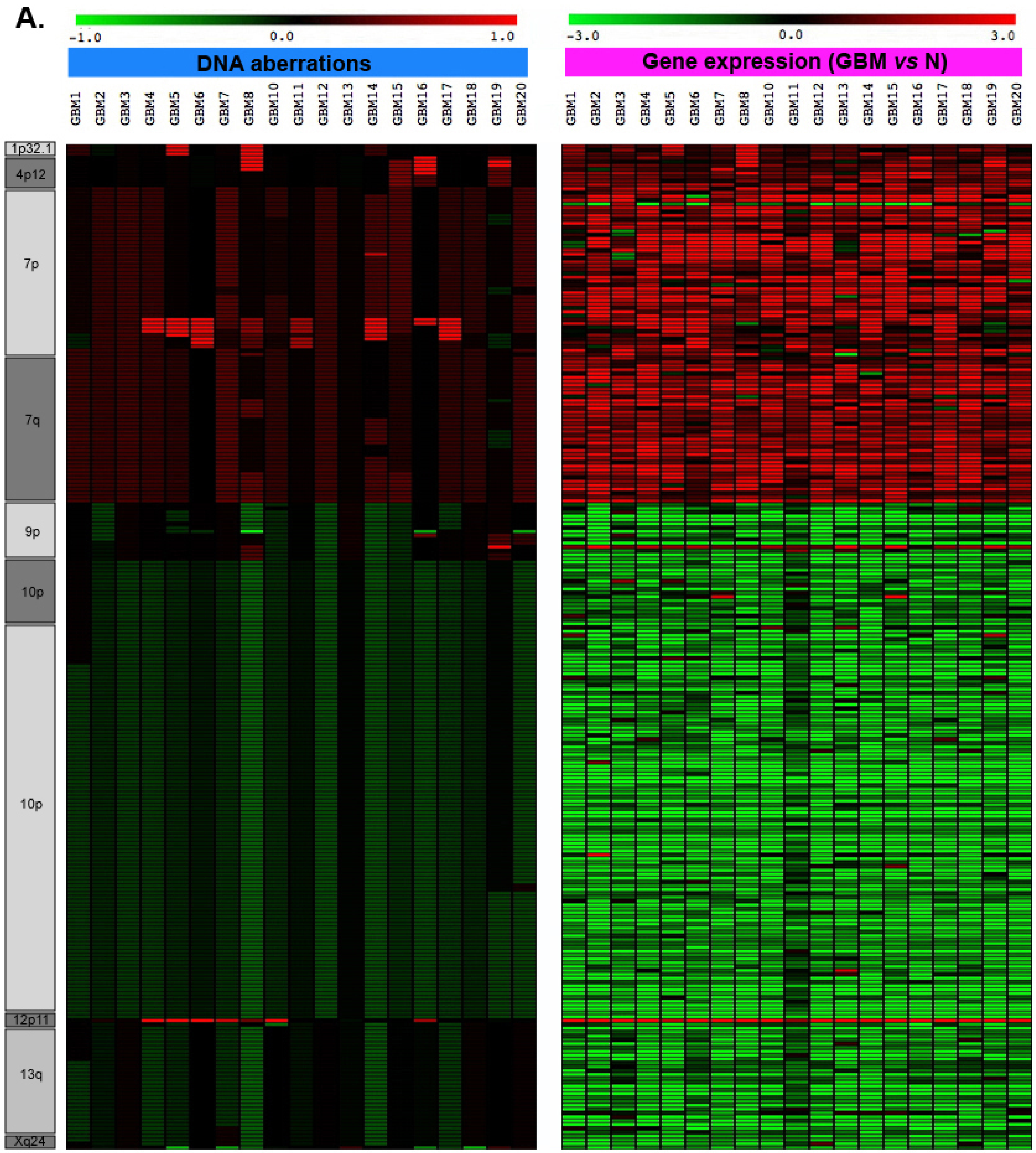
p-value < 0.05 / FC < -2

Finding correlations



4,1% of the genome changes drive correlated gene expression changes (159 genes)

Finding concordances



6,9% of the genome changes drive concordant gene expression changes (261 genes DE)

Over-expression

⬇

Chr1 - Chr4 - Chr7

Under-expression

⬇

Chr9 - Chr10 - Chr13

Validation

Data Set

GEO, Sun et al. 2006

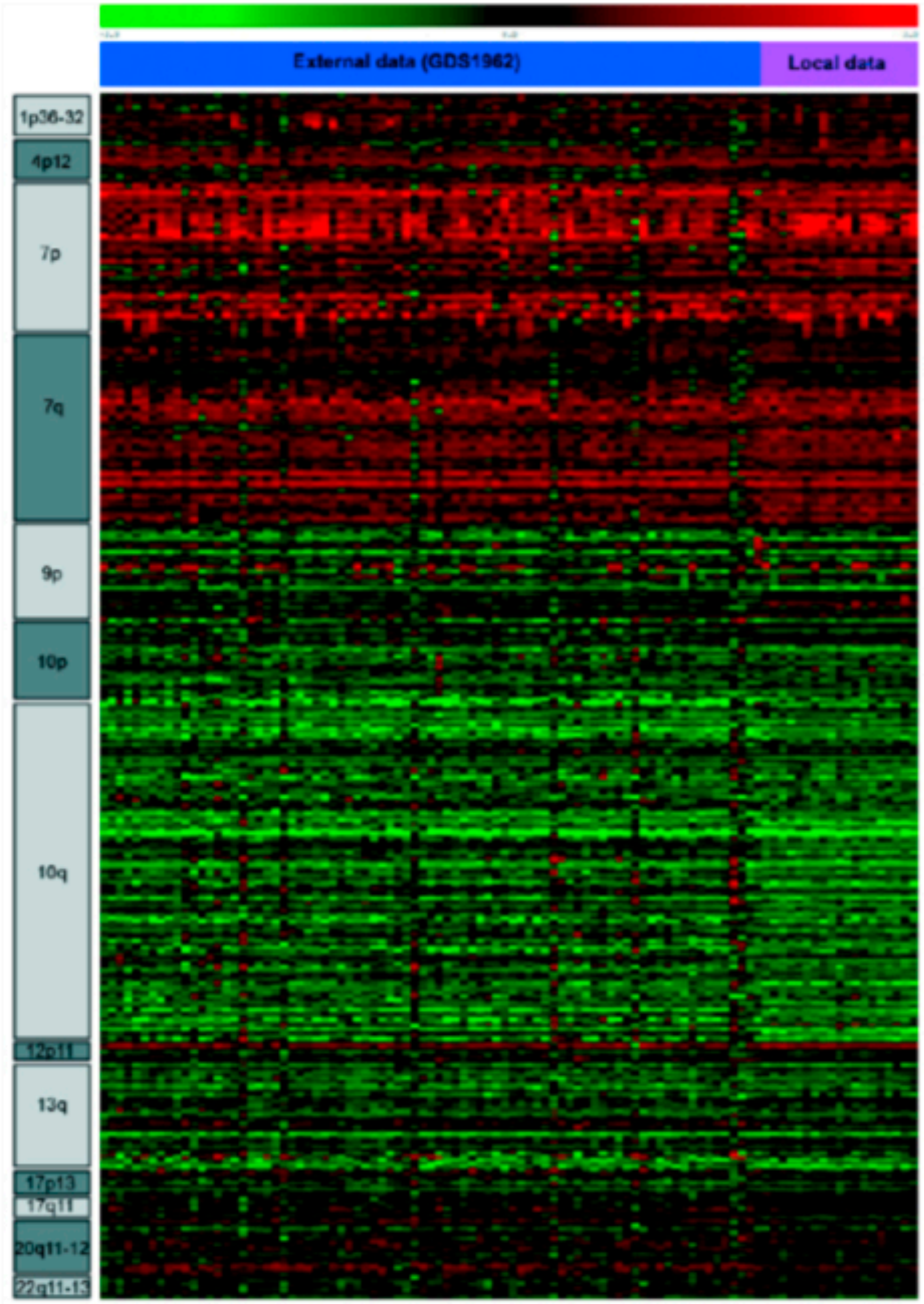
Affymetrix U133 Plus

81 GBMs & 23 controls

Validation

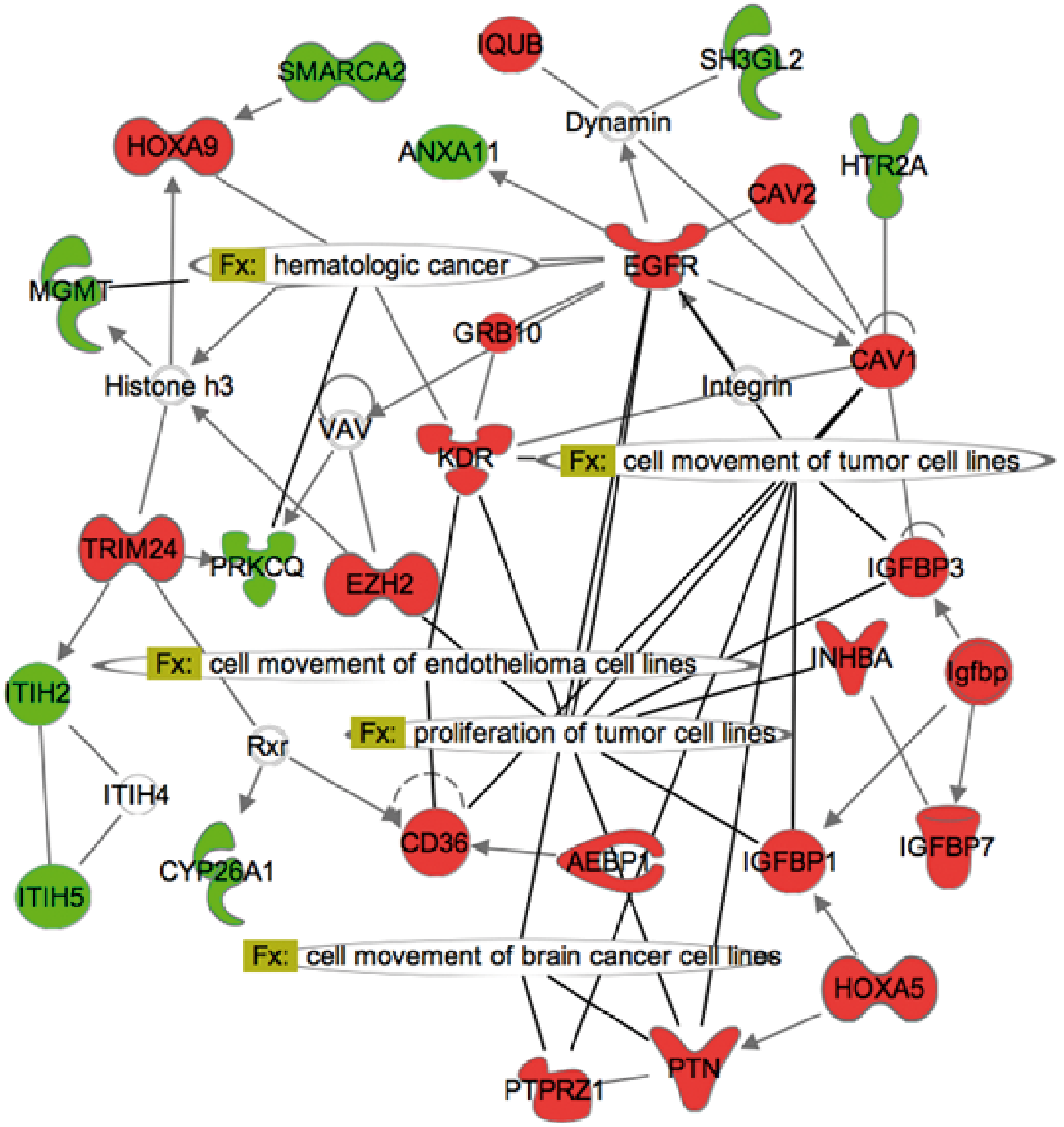
87% genes are valided

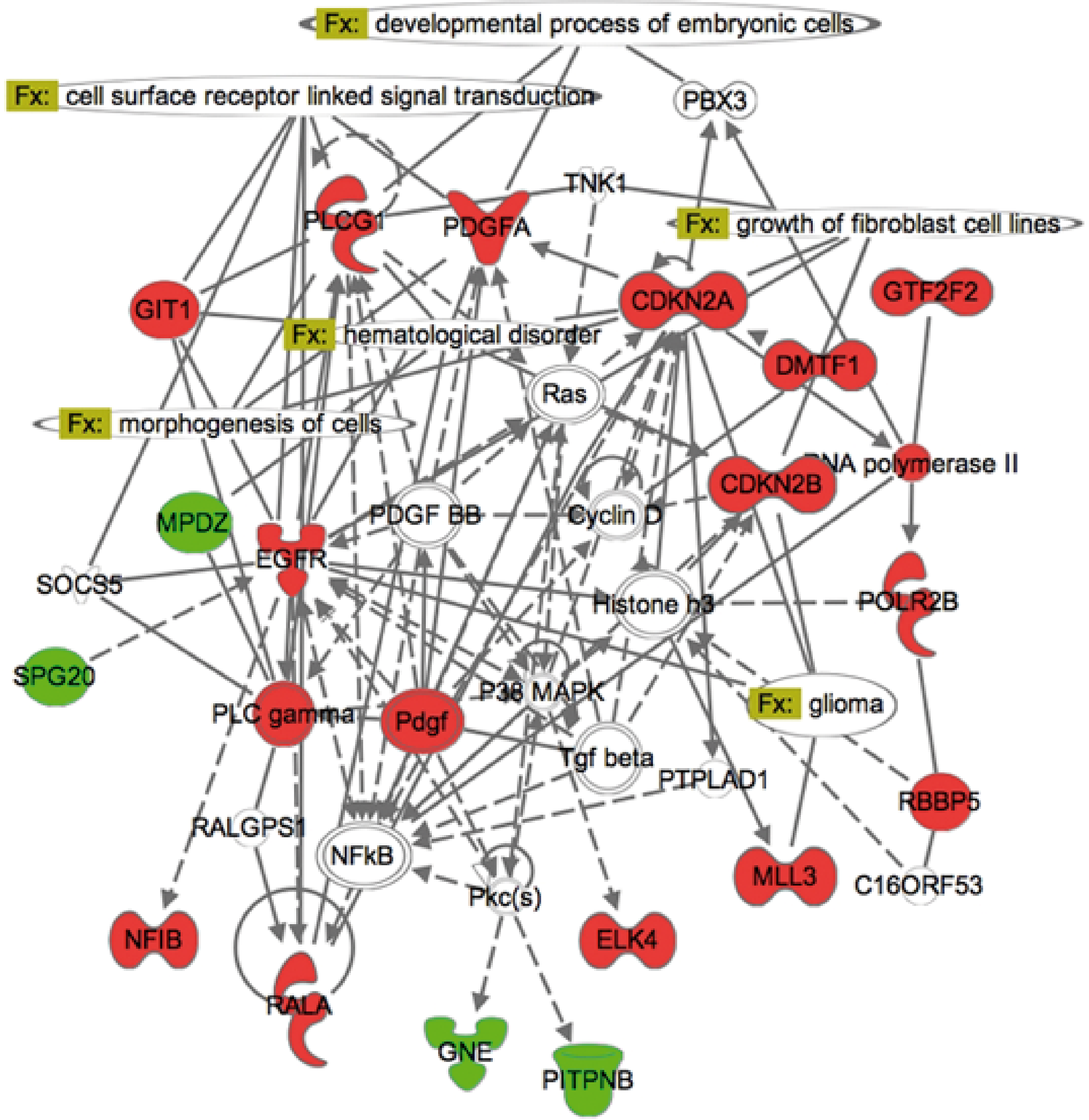
Only 10% of DNA alterations drive direct modification of gene expression



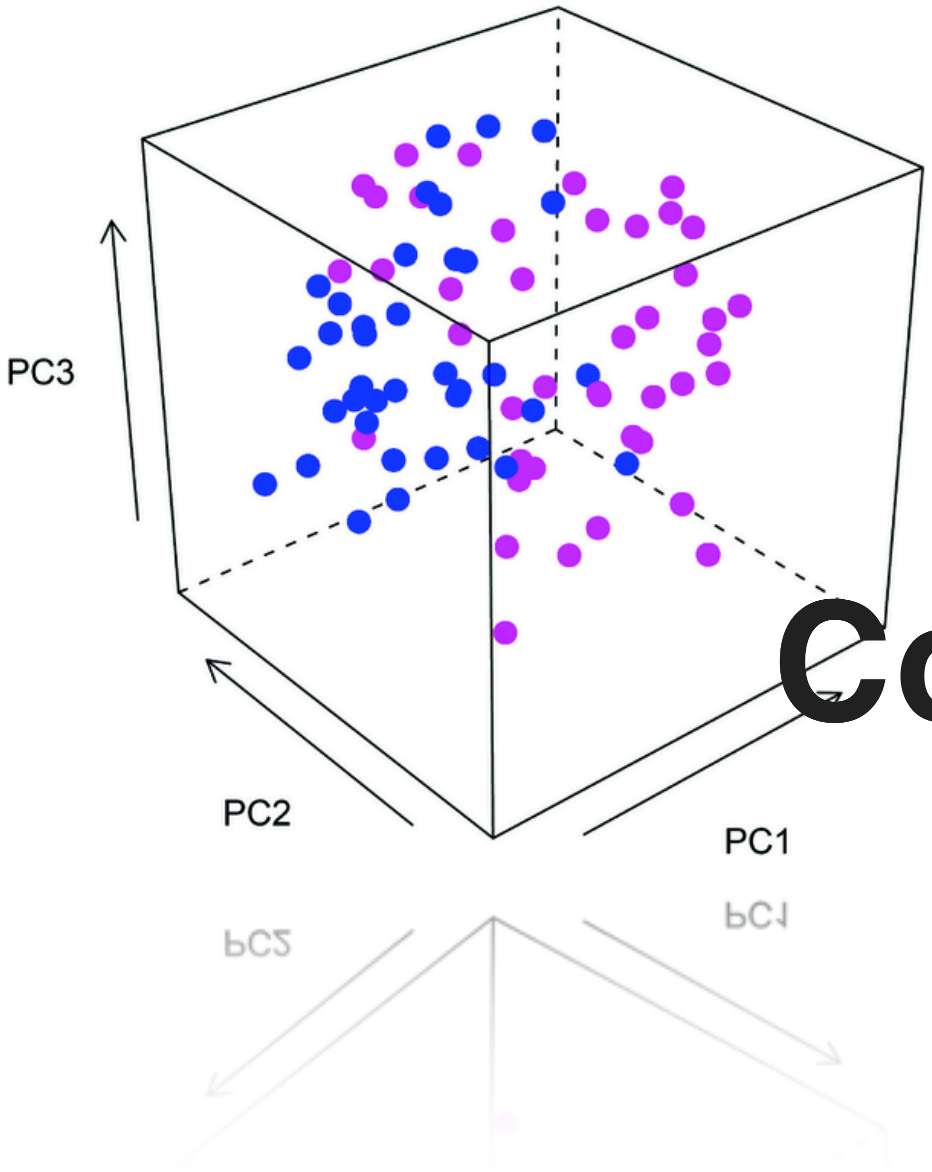
Functionnal Annotation

Cell cycle, apoptosis, cell adhesion, angiogenesis, DNA repair, mRNA processing





Integrating biological knowledge



Biological Contextualisation

... around Principal Component Analysis

de Tayrac et al., *BMC Genomics*, 2009

Principle

BMC Genomics

BioMed Central

Methodology article

Open Access

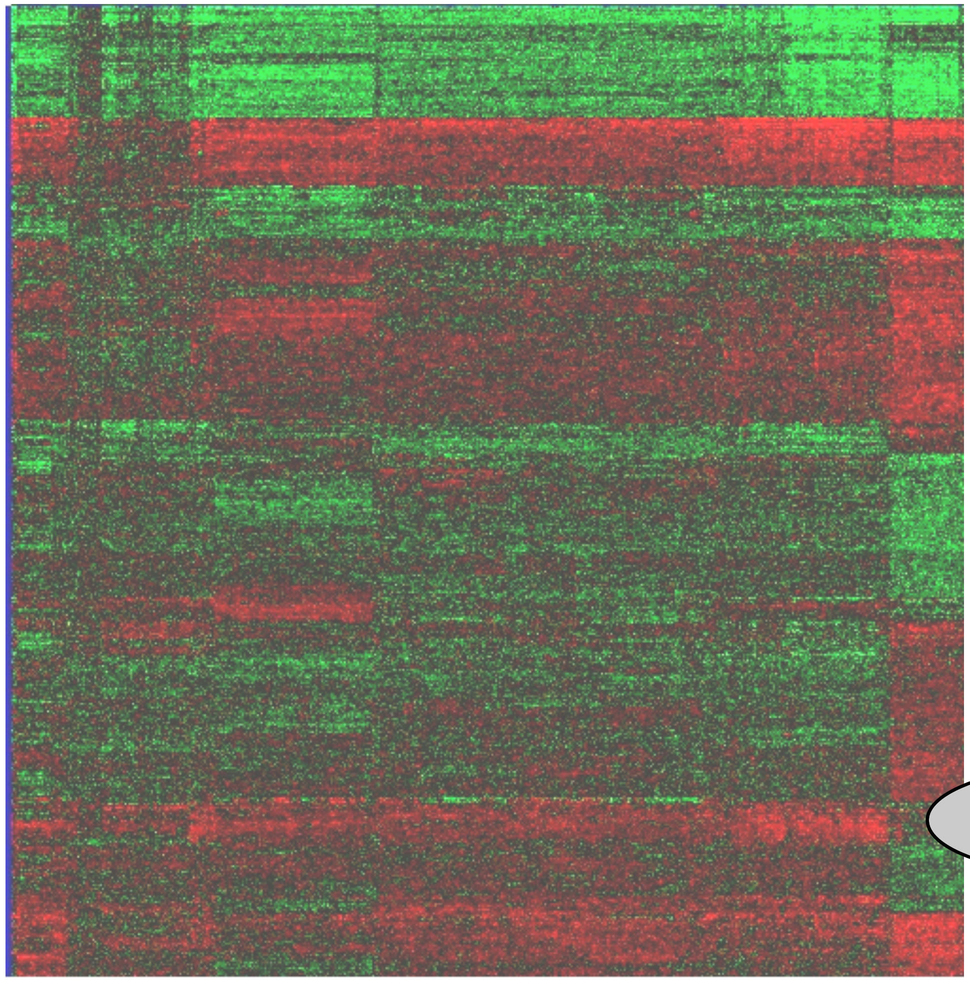
Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach

Marie de Tayrac[1,2], Sébastien Lê*[3], Marc Aubry[4], Jean Mosser*[1,2,4] and François Husson[3]

Prerequisites:

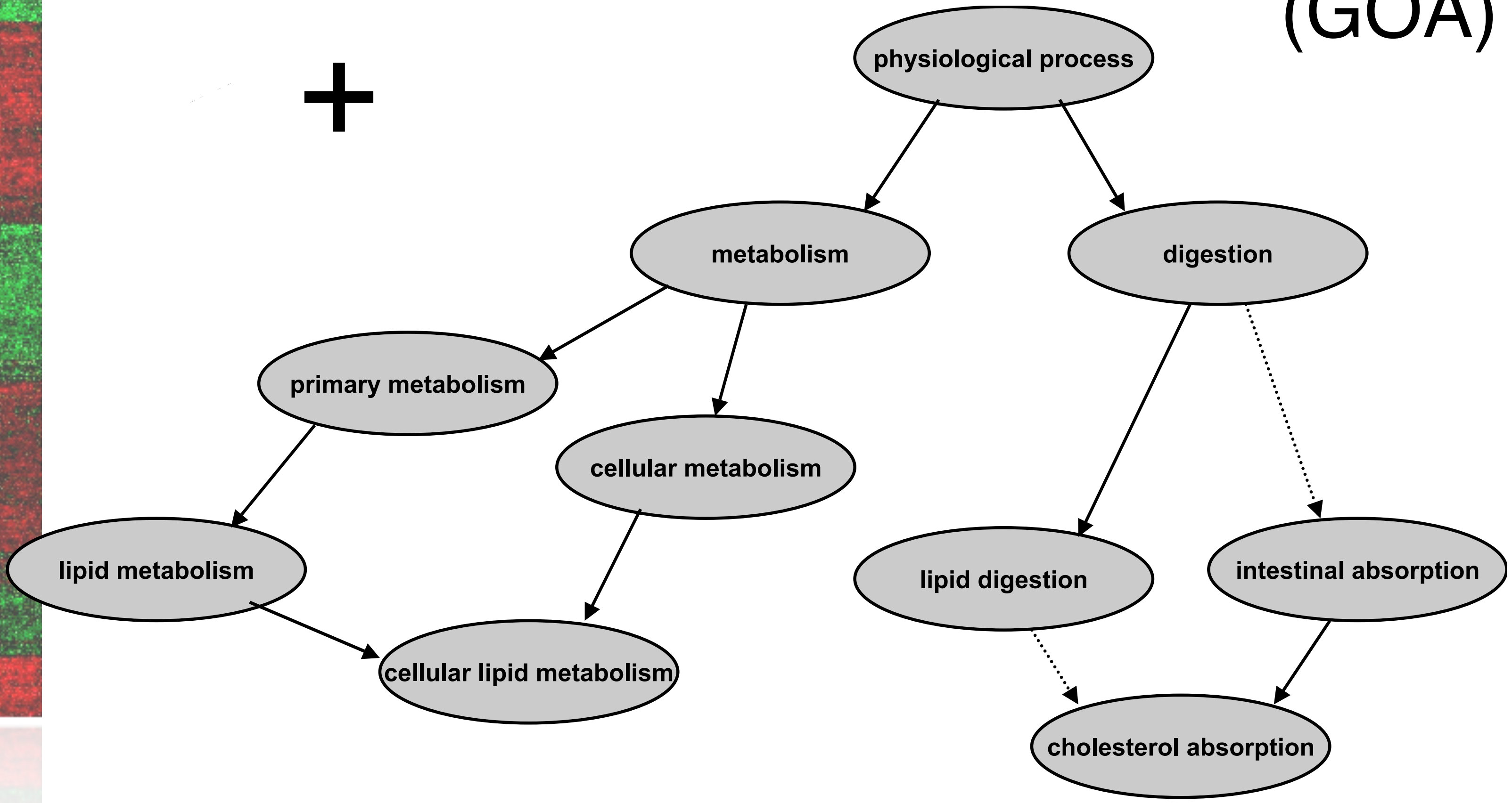
Method without *a priori* to facilitate the functional interpretation of genome-wide data

‘Omic’ data



+

Annotation
Gene Ontology
(GOA)

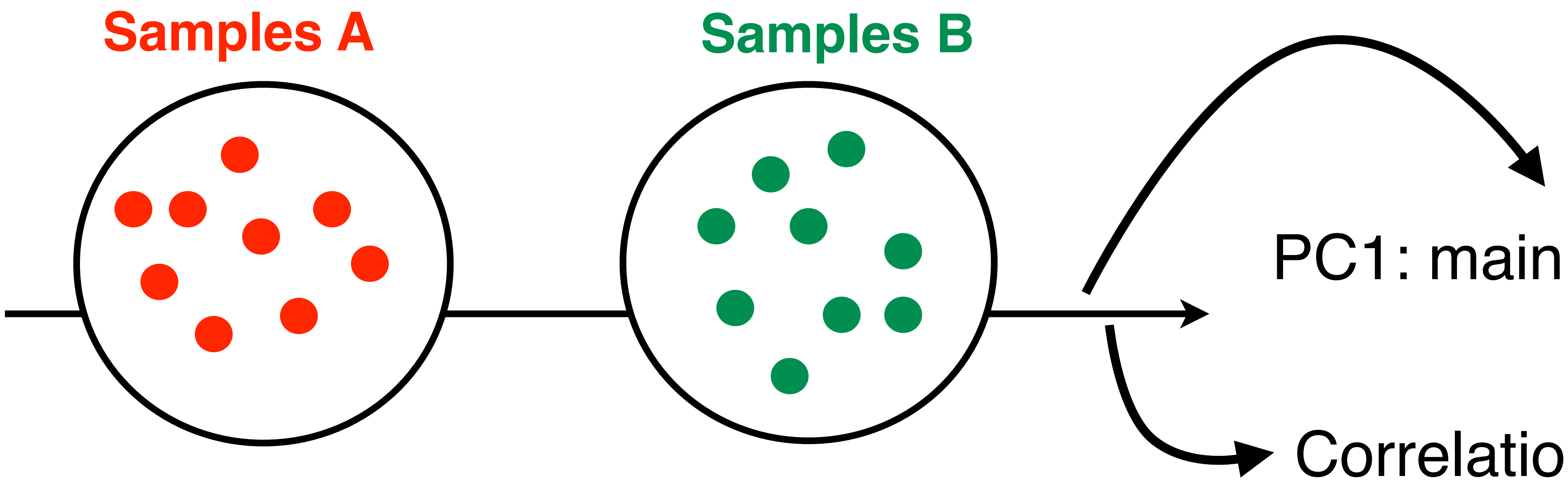


physiological process
metabolism
digestion
primary metabolism
cellular metabolism
lipid metabolism
cellular lipid metabolism
lipid digestion
intestinal absorption
cholesterol absorption

Data-Mining Methodology (MFA)

Investigation of microarray data is supported by a step-by-step sequence of graphical representations

~ Principal Component Analysis (PCA)

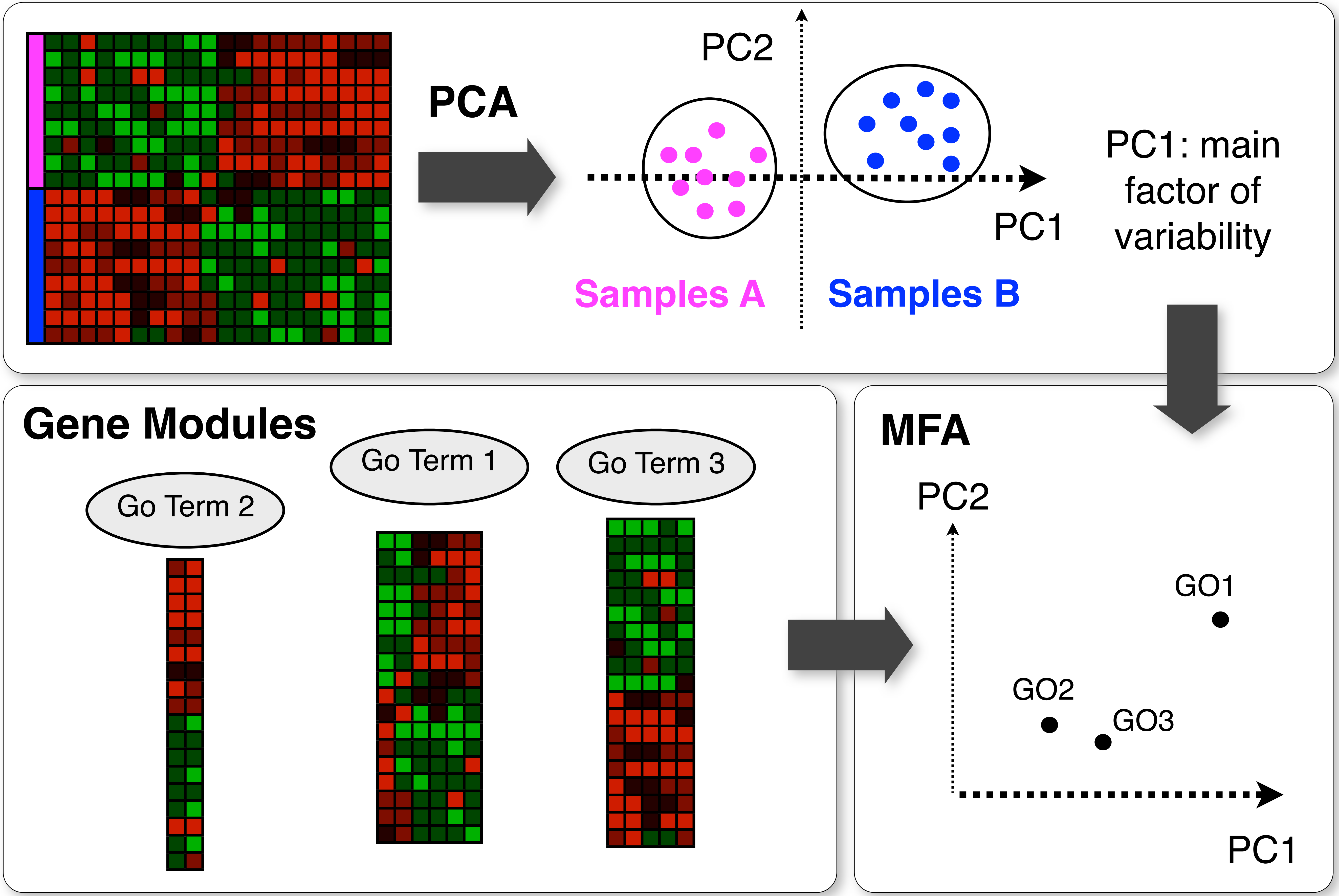


PC1: main factor of variability

Correlation (Annotation / PC)

Gene module used to attach functionally meaningful characteristics to each experimentally interpreted component

Method

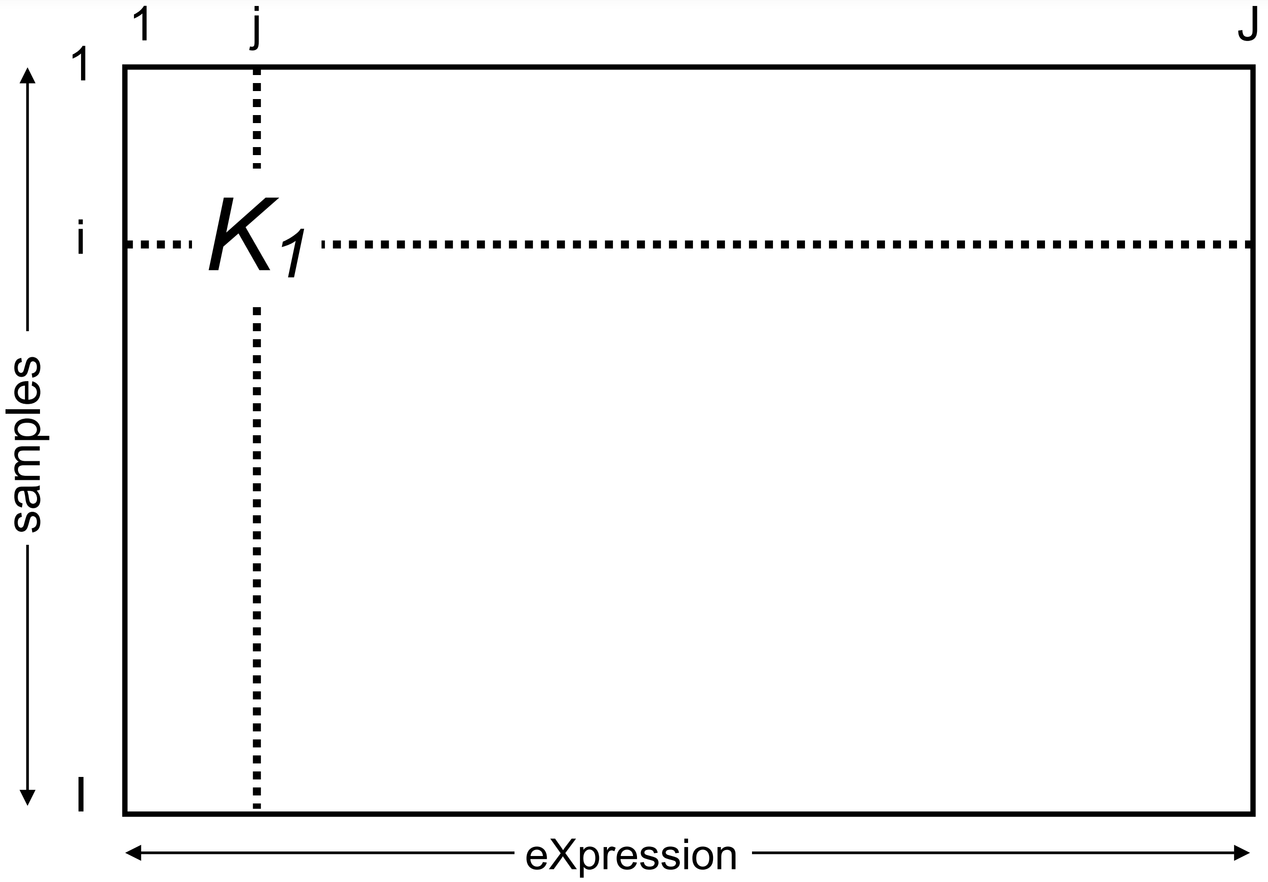

PCA

PC1: main factor of variability

Gene Modules

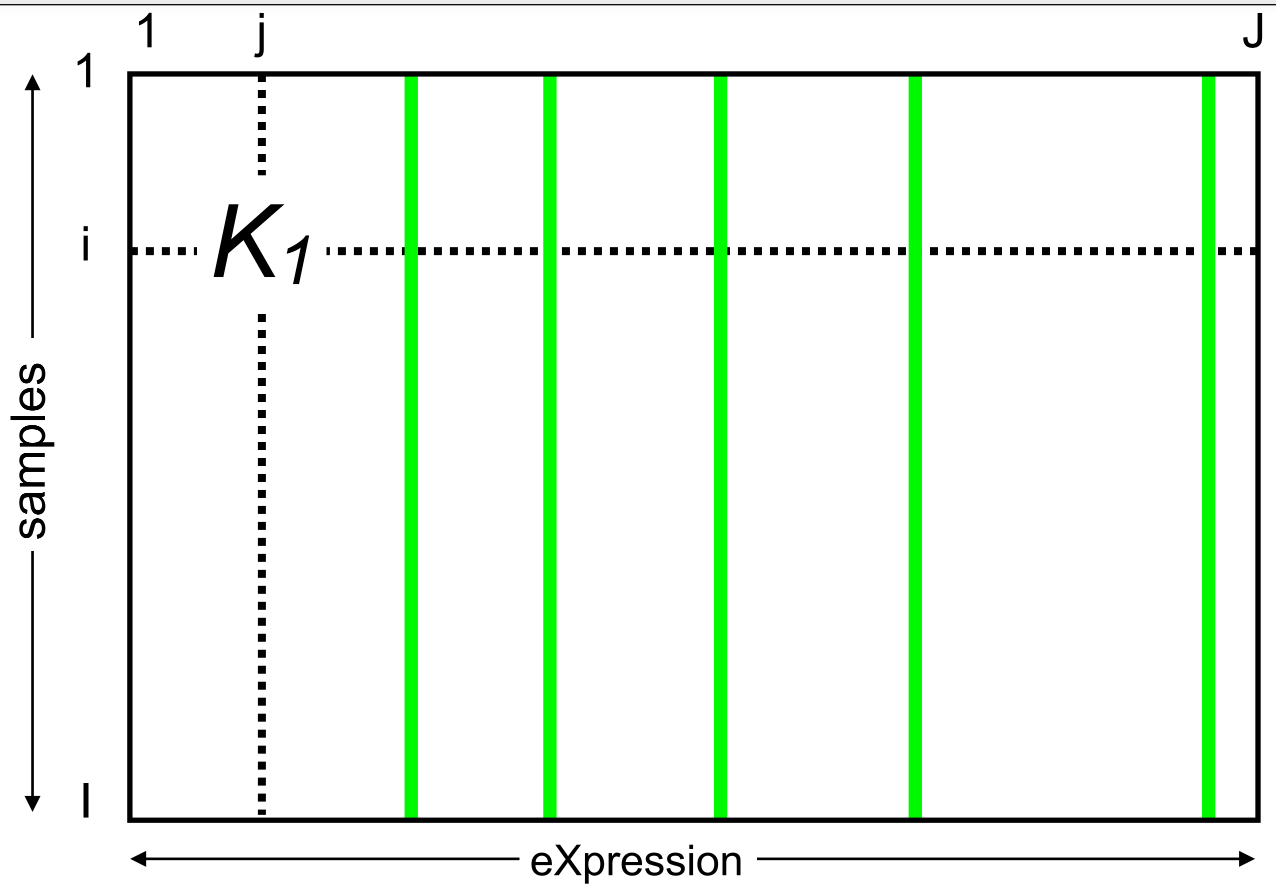
MFA

How modules are built?



- Genes are grouped by processes (**GO term**)
- A process is described by the genes expression (**K_{BP1}**) = module

How modules are built?

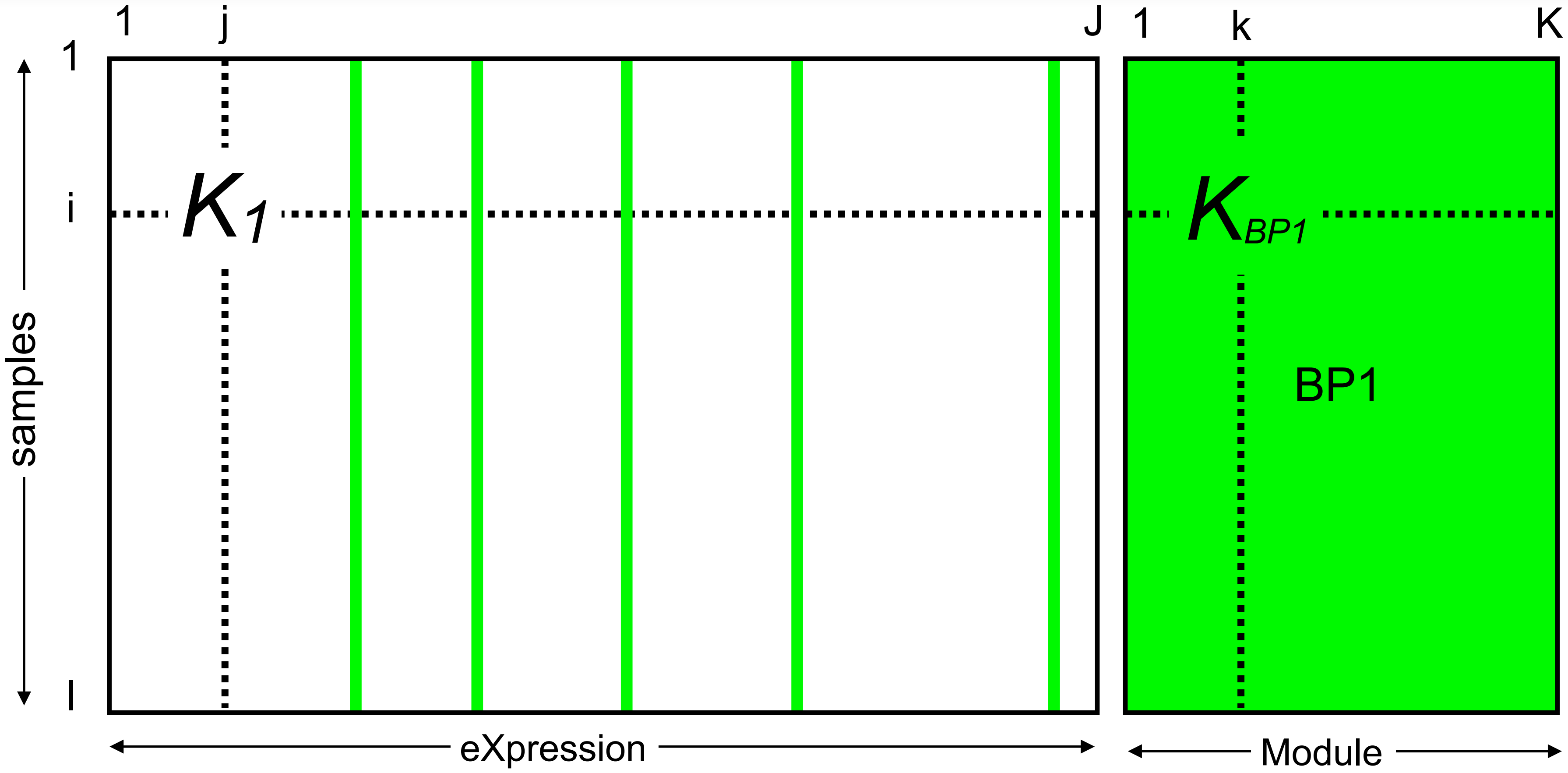


GO term BP1

4,381	4,411	4,472	4,353	4,279
3,154	2,960	3,234	3,049	2,865
3,184	3,207	3,058	3,273	2,923
3,721	3,762	3,676	3,717	3,807
2,480	2,691	2,437	2,681	2,628
3,309	3,202	3,382	3,420	3,161
2,851	2,913	2,917	2,910	3,137
2,764	2,260	2,718	2,088	2,642
1,765	2,512	1,771	3,041	2,428
2,469	2,797	2,410	2,474	2,739
3,337	3,207	3,264	2,829	2,882
3,406	3,420	3,356	3,443	3,233
3,121	3,222	3,020	3,095	3,254
3,896	3,743	3,671	3,878	3,887
...	...	...	...	...
3,286	2,995	3,345	3,360	3,239
3,008	3,166	3,154	3,251	3,054

- Genes are grouped by processes (**GO term**)
- A process is described by the genes expression (**K_{BP1}**) = module

How modules are built?

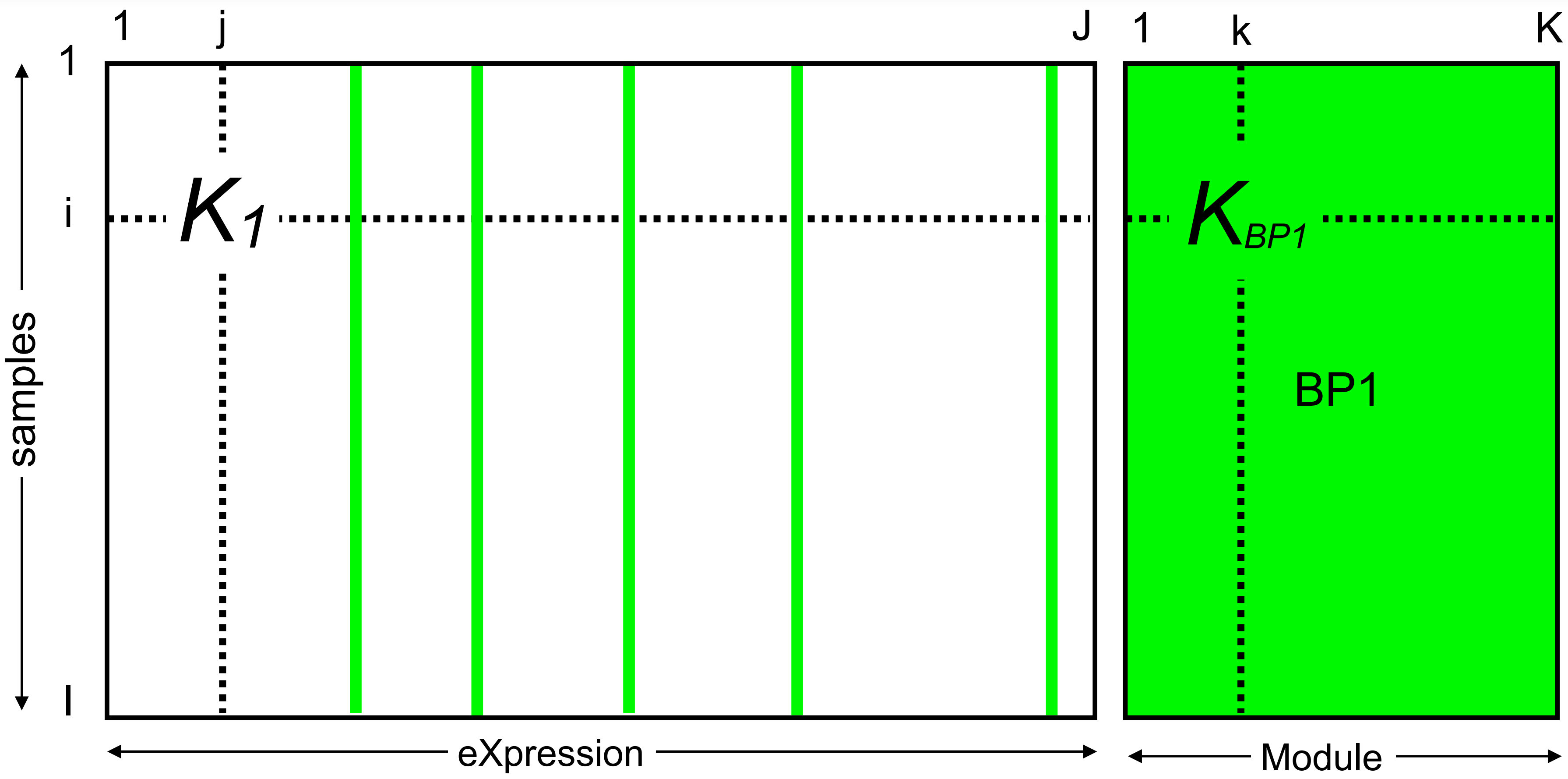


4,381	4,411	4,472	4,353	4,279
3,154	2,960	3,234	3,049	2,865
3,184	3,207	3,058	3,273	2,923
3,721	3,762	3,676	3,717	3,807
2,480	2,691	2,437	2,681	2,628
3,309	3,202	3,382	3,420	3,161
2,851	2,913	2,917	2,910	3,137
2,764	2,260	2,718	2,088	2,642
1,765	2,512	1,771	3,041	2,428
2,469	2,797	2,410	2,474	2,739
3,337	3,207	3,264	2,829	2,882
3,406	3,420	3,356	3,443	3,233
3,121	3,222	3,020	3,095	3,254
3,896	3,743	3,671	3,878	3,887
...	...	...	...	...
3,286	2,995	3,345	3,360	3,239
3,008	3,166	3,154	3,251	3,054

GO term BP1

- Genes are grouped by processes (**GO term**)
- A process is described by the genes expression (**K_{BP1}**) = module

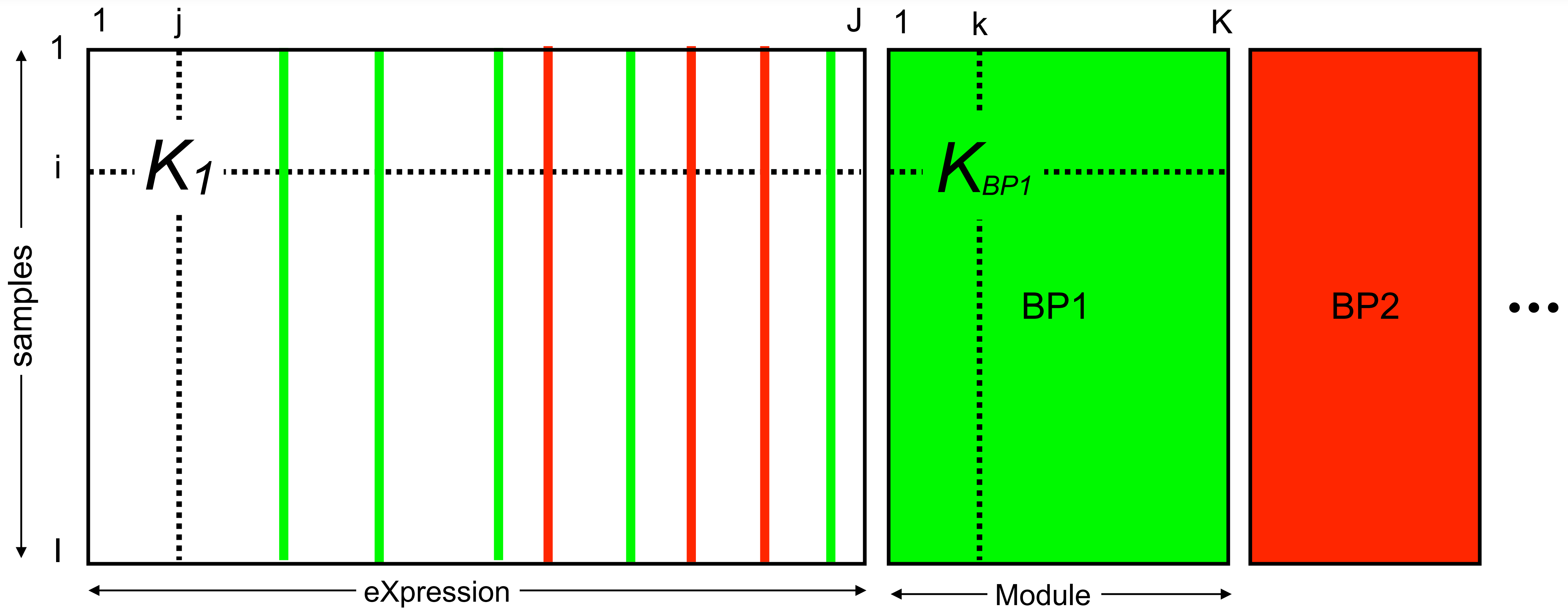
How modules are built?



4,381	4,411	4,472	4,353	4,279
3,154	2,960	3,234	3,049	2,865
3,184	3,207	3,058	3,273	2,923
3,721	3,762	3,676	3,717	3,807
2,480	2,691	2,437	2,681	2,628
3,309	3,202	3,382	3,420	3,161
2,851	2,913	2,917	2,910	3,137
2,764	2,260	2,718	2,088	2,642
1,765	2,512	1,771	3,041	2,428
2,469	2,797	2,410	2,474	2,739
3,337	3,207	3,264	2,829	2,882
3,406	3,420	3,356	3,443	3,233
3,121	3,222	3,020	3,095	3,254
3,896	3,743	3,671	3,878	3,887
...	...	...	...	...
3,286	2,995	3,345	3,360	3,239
3,008	3,166	3,154	3,251	3,054

GO term BP1

How modules are built?



4,413	4,498	4,322
3,420	3,158	2,947
3,093	2,980	3,020
3,570	3,763	3,880
2,497	2,592	3,049
3,438	3,320	3,369
2,805	2,940	3,008
2,819	2,791	2,689
1,715	2,225	2,635
2,447	2,764	2,683
3,127	3,443	2,840
3,364	3,510	3,493
3,021	3,171	3,278
3,823	3,830	3,958
...	...	...
3,272	3,233	3,063
3,172	2,978	3,161

4,381	4,411	4,472	4,353	4,279
3,154	2,960	3,234	3,049	2,865
3,184	3,207	3,058	3,273	2,923
3,721	3,762	3,676	3,717	3,807
2,480	2,691	2,437	2,681	2,628
3,309	3,202	3,382	3,420	3,161
2,851	2,913	2,917	2,910	3,137
2,764	2,260	2,718	2,088	2,642
1,765	2,512	1,771	3,041	2,428
2,469	2,797	2,410	2,474	2,739
3,337	3,207	3,264	2,829	2,882
3,406	3,420	3,356	3,443	3,233
3,121	3,222	3,020	3,095	3,254
3,896	3,743	3,671	3,878	3,887
...	...	...	...	...
3,286	2,995	3,345	3,360	3,239
3,008	3,166	3,154	3,251	3,054

4,413	4,498	4,322
3,420	3,158	2,947
3,093	2,980	3,020
3,570	3,763	3,880
2,497	2,592	3,049
3,438	3,320	3,369
2,805	2,940	3,008
2,819	2,791	2,689
1,715	2,225	2,635
2,447	2,764	2,683
3,127	3,443	2,840
3,364	3,510	3,493
3,021	3,171	3,278
3,823	3,830	3,958
...	...	...
3,272	3,233	3,063
3,172	2,978	3,161

🟩 **GO term BP1**

🟥 **GO term BP2**

- MFA allow to attach functionally characteristics to the factorial map

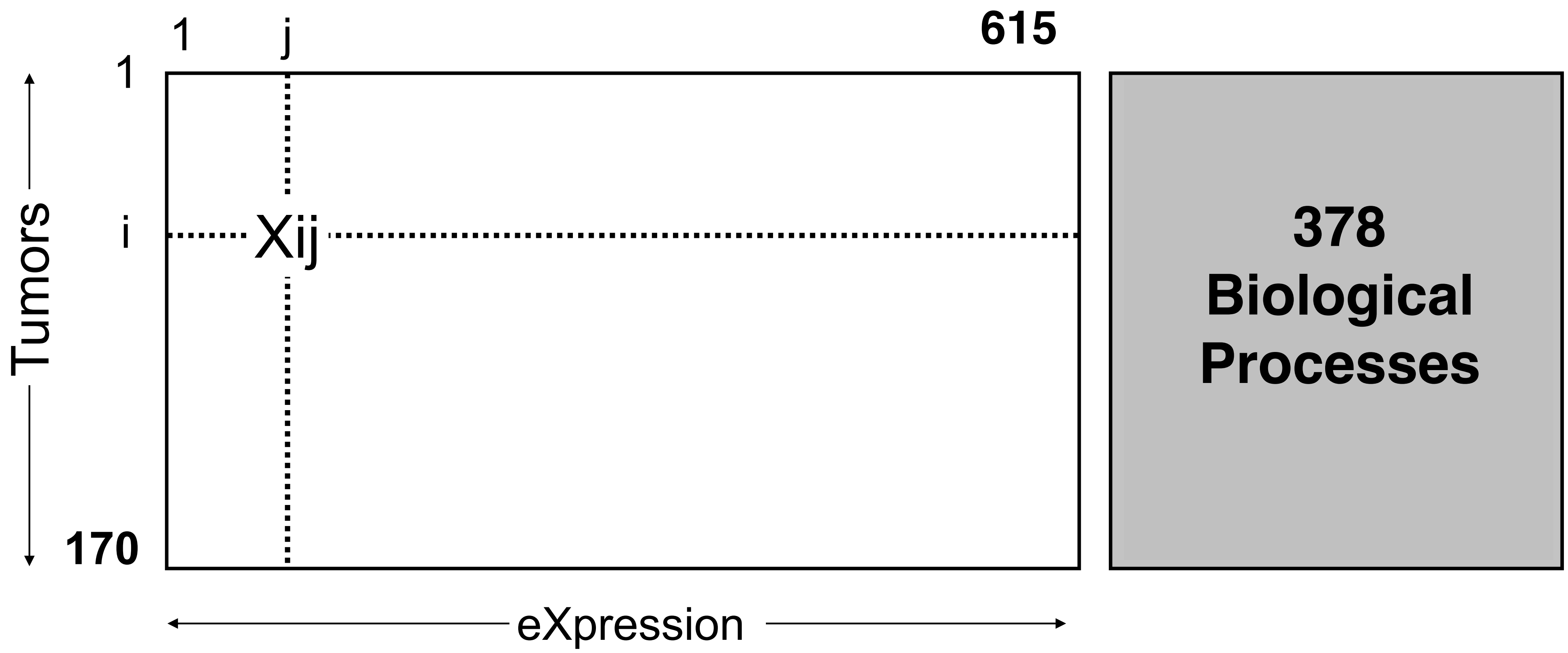
Example - data set

Freije et al., *Cancer Res.* (2004)

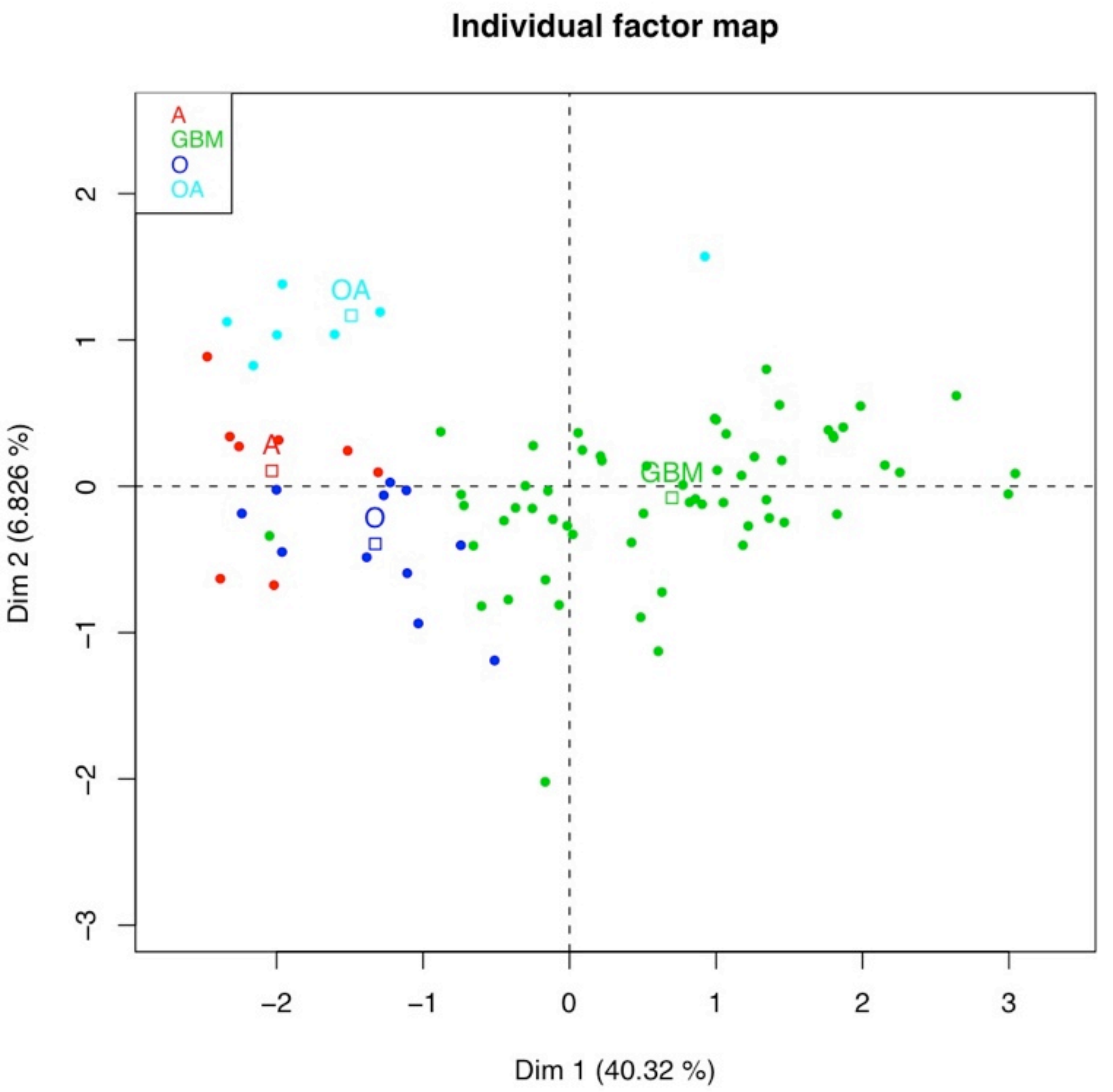
GEO Gene Expression Omnibus

- Affymetrix U133A (GPL96)
- 170 gliomes (GSE4412)

O (oligodendroglioma)
OA (oligoastrocytoma)
A (astrocytoma)
GBM (glioblastoma)

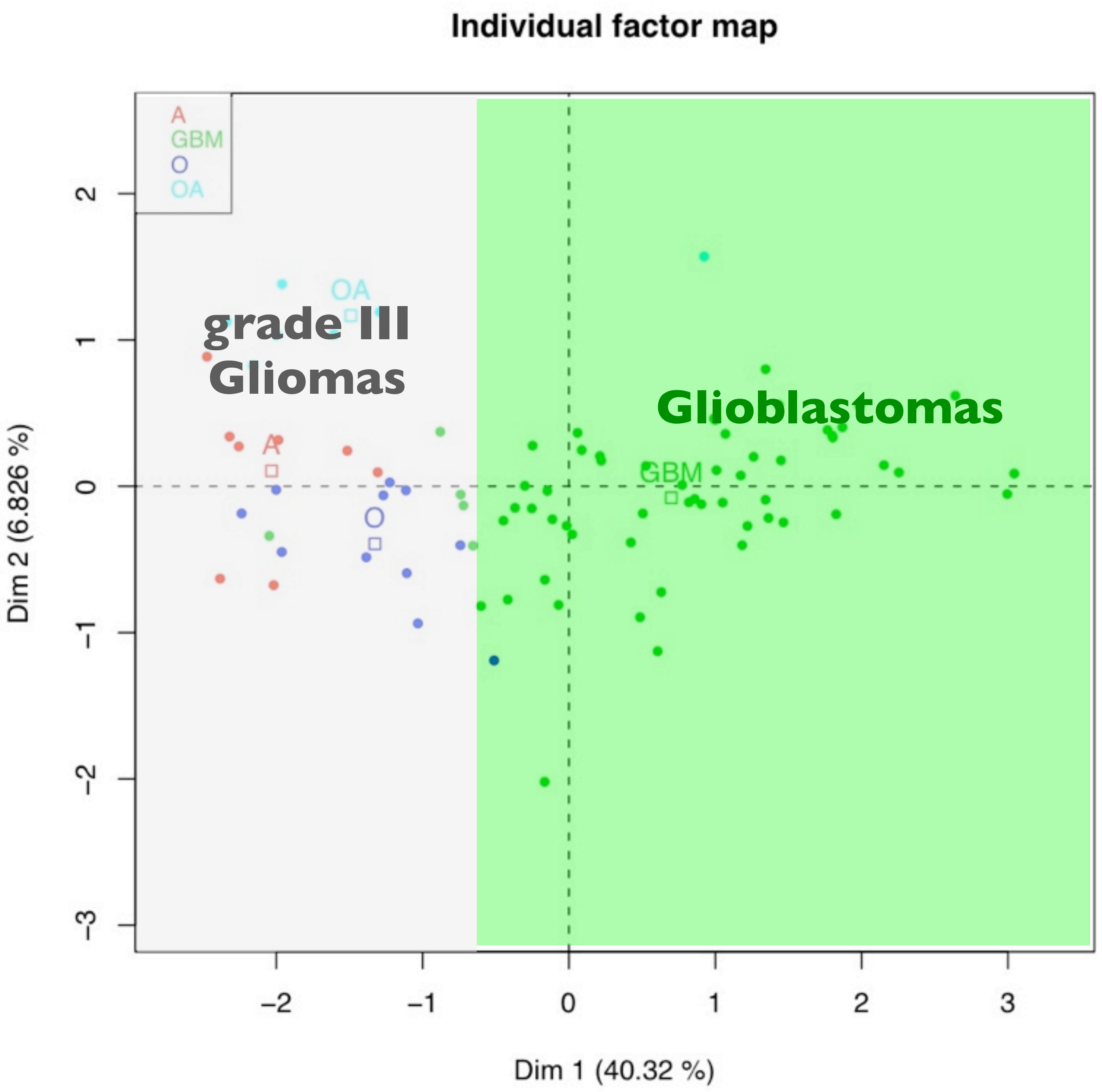


Example - Factorial Map



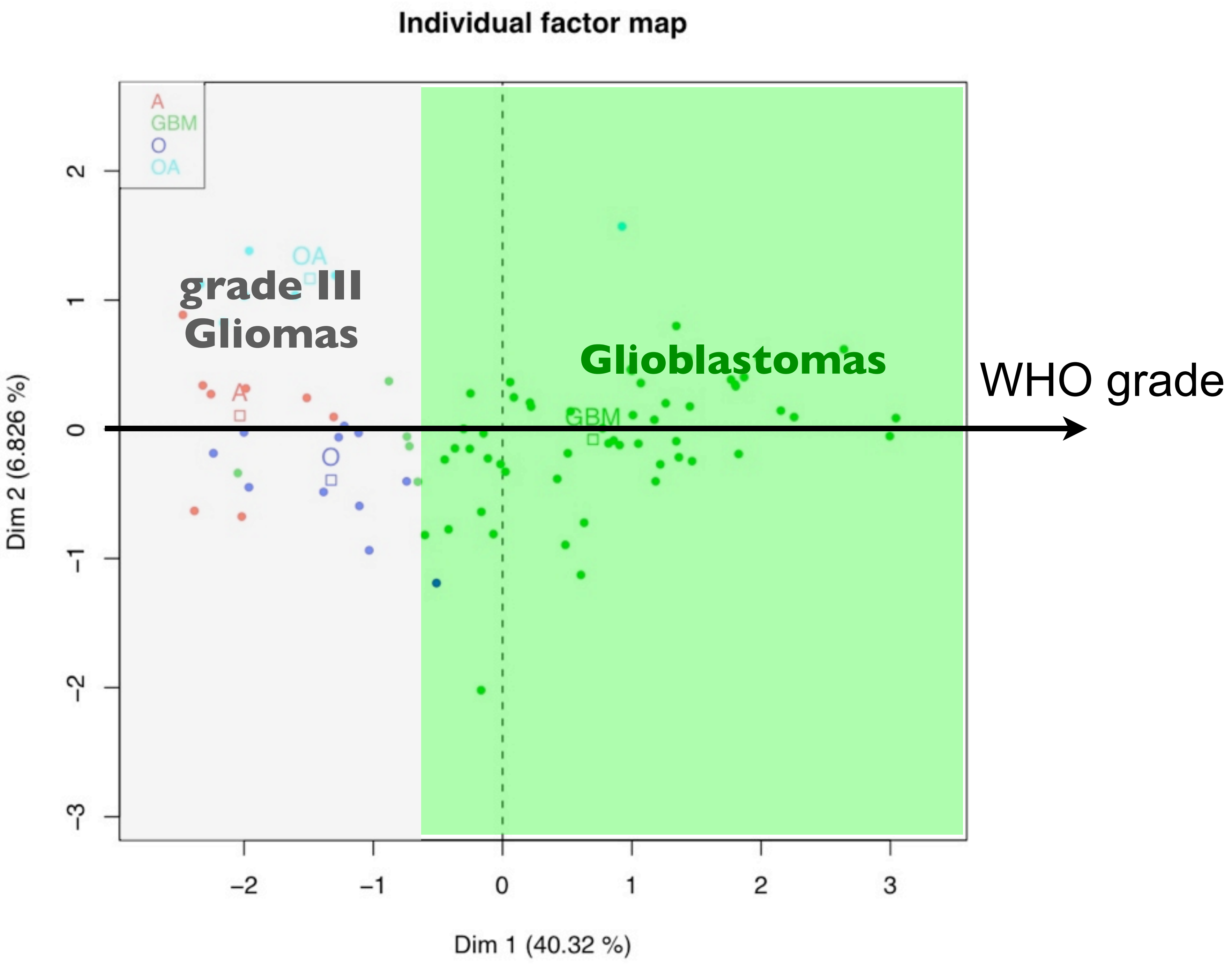
de Tayrac et al., *BMC Genomics*, 2009

Example - Factorial Map



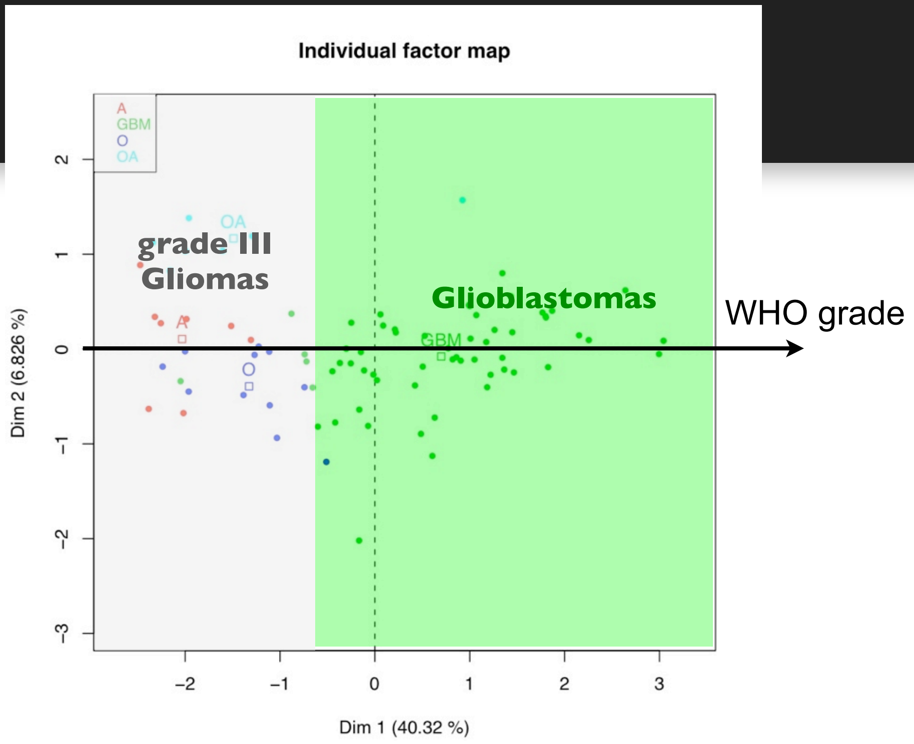
de Tayrac et al., *BMC Genomics*, 2009

Example - Factorial Map



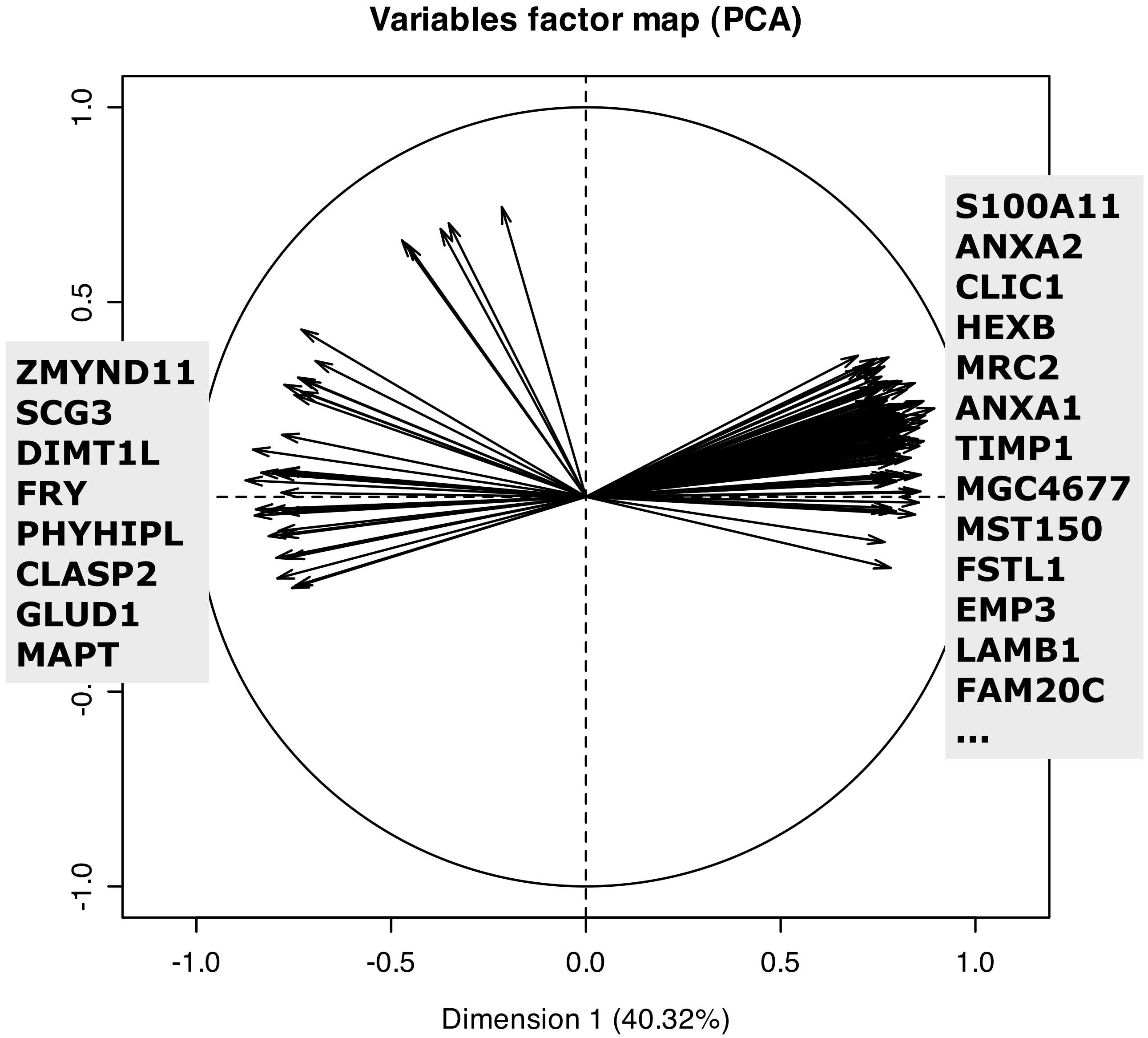
de Tayrac et al., *BMC Genomics*, 2009

Example - Interpretation (1/3)



< *Manual* interpretation >

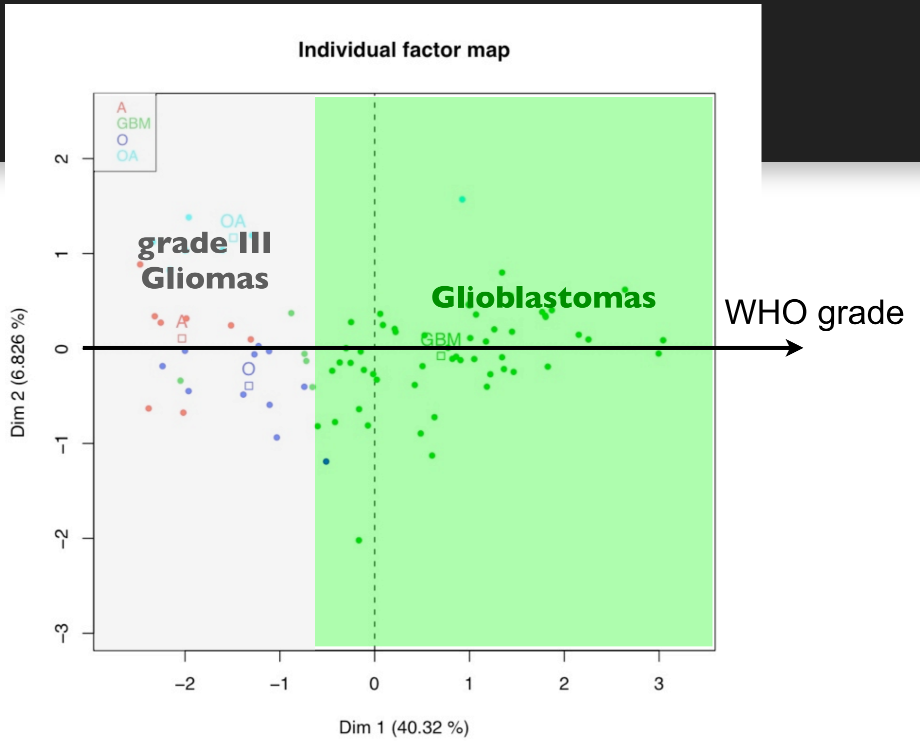
- Gene selection
- Literature search
- Functional annotation

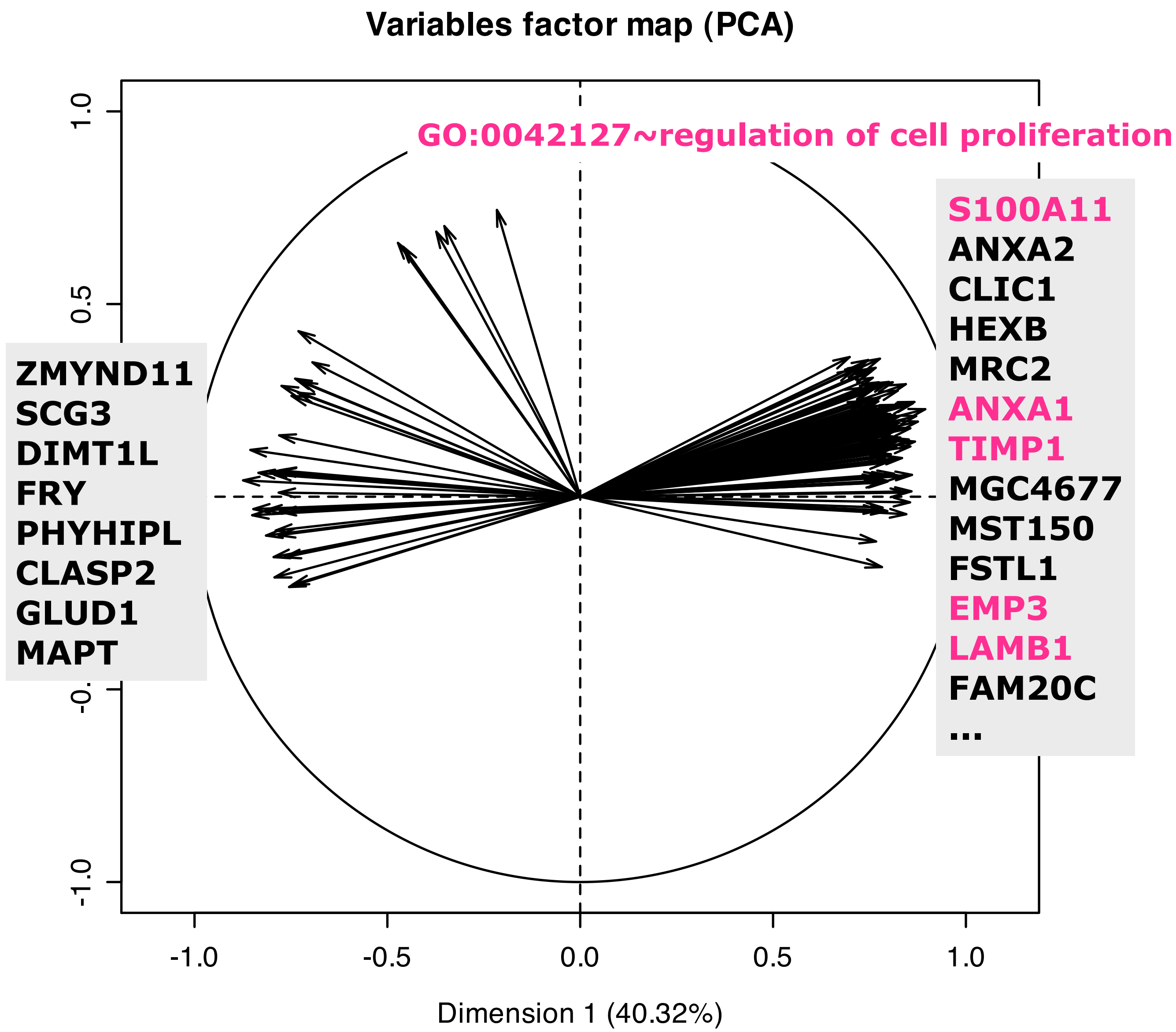


Example - Interpretation (1/3)

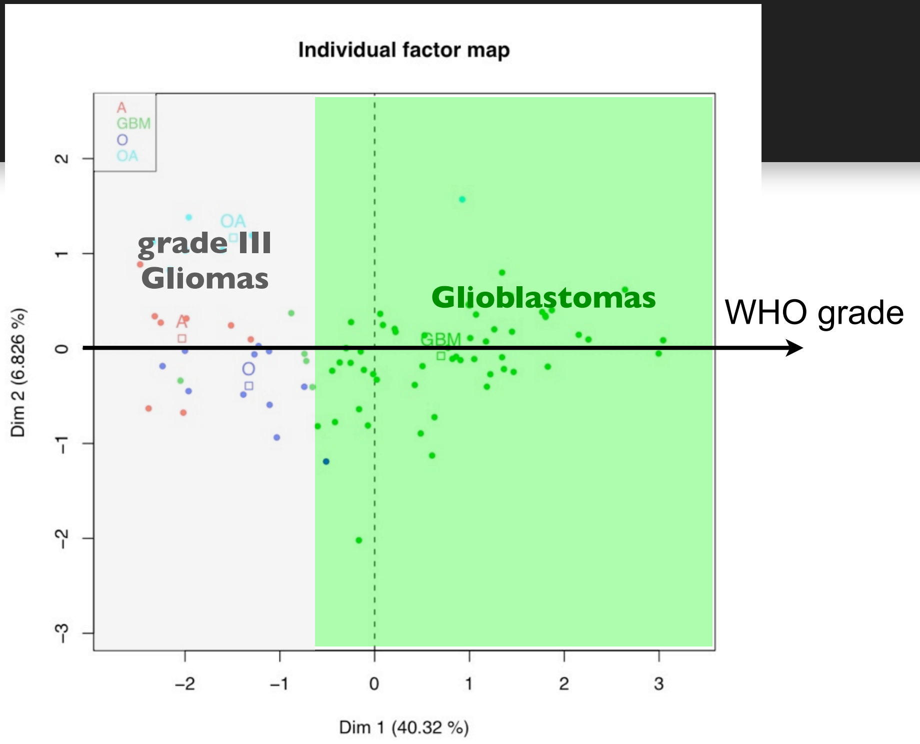
< *Manual* interpretation >

- Gene selection
- Literature search
- Functional annotation



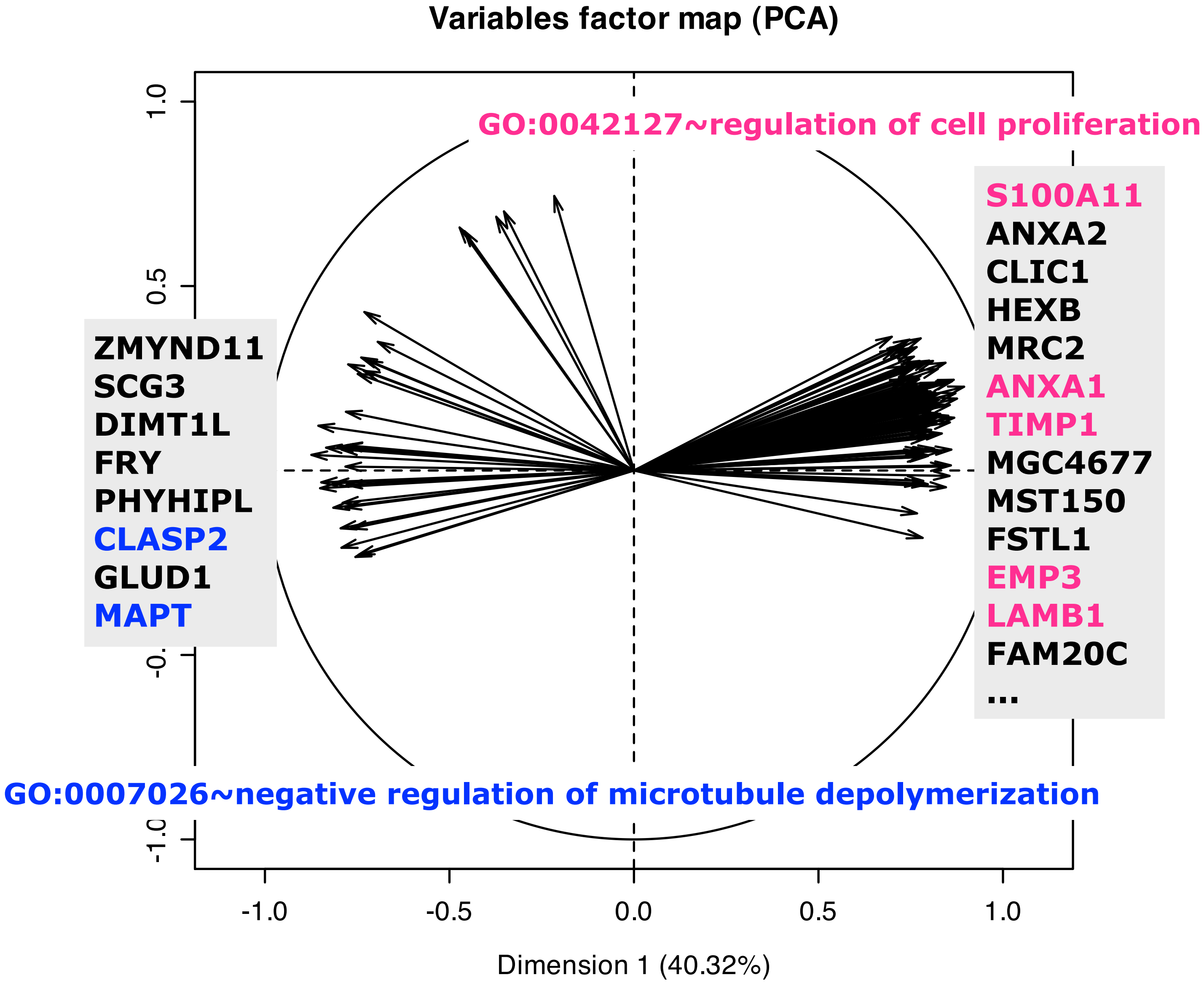


Example - Interpretation (1/3)



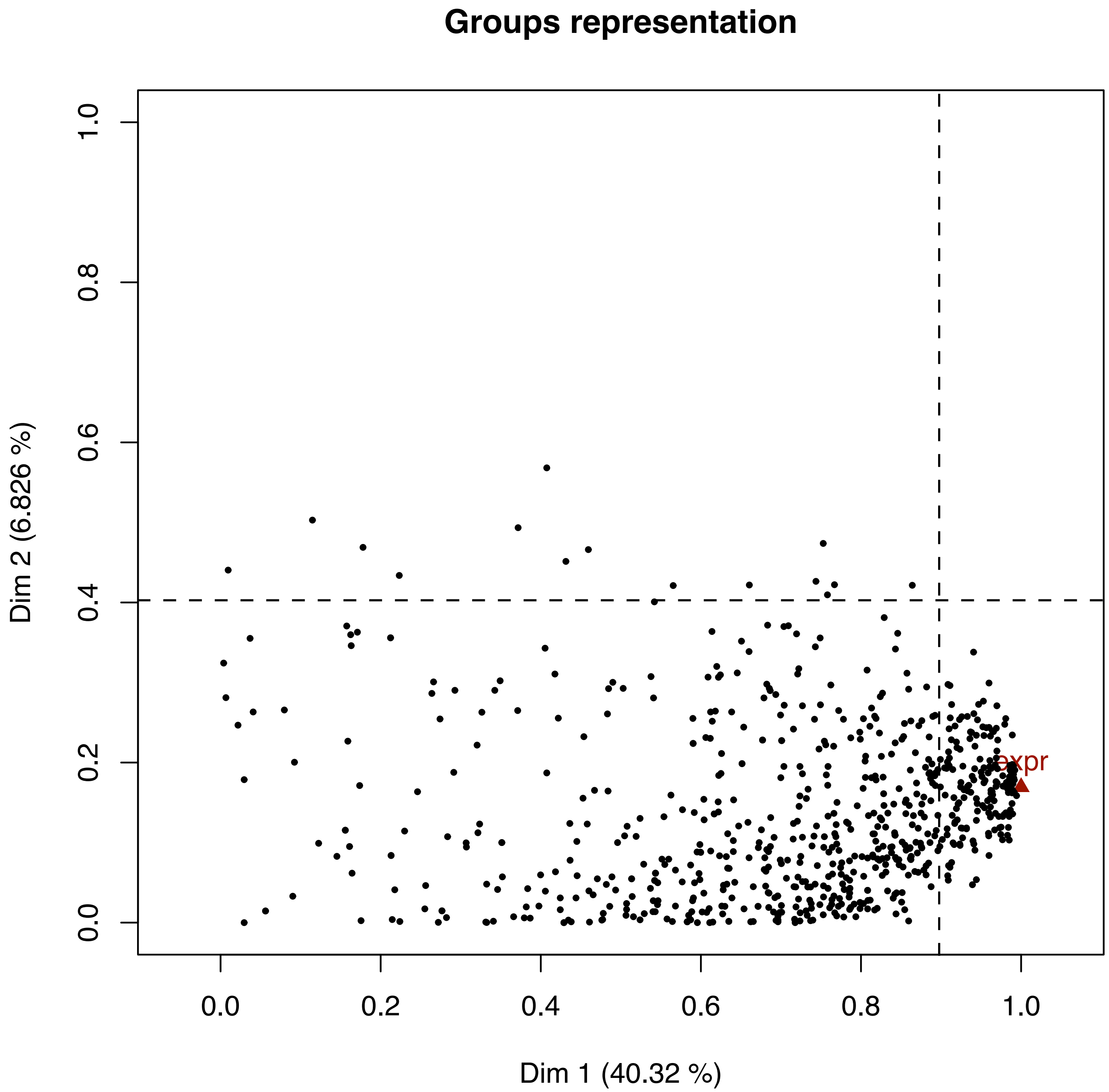
< *Manual* interpretation >

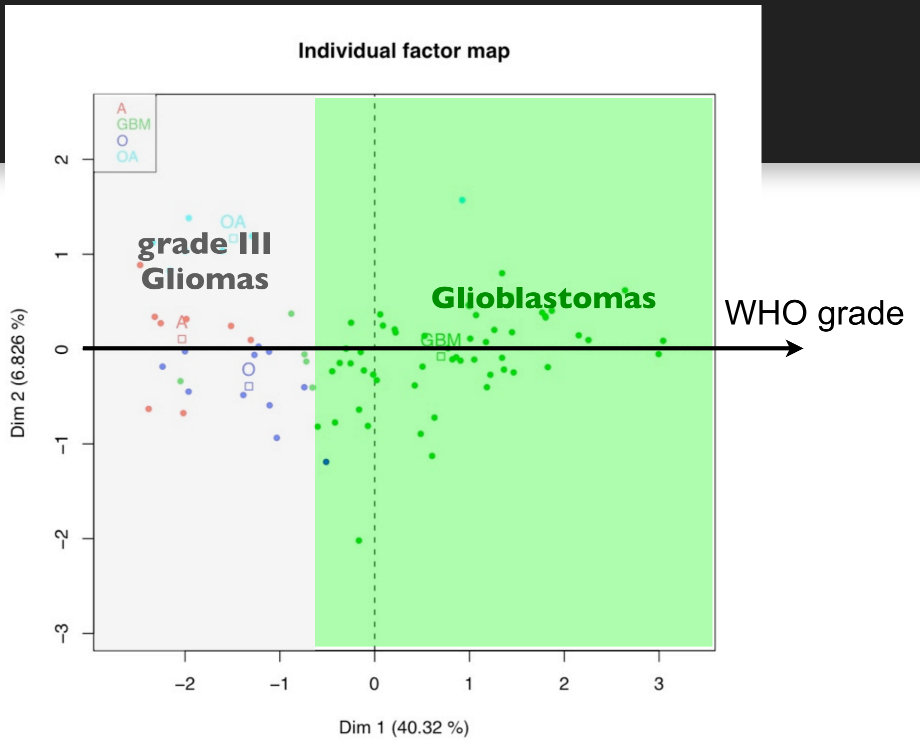
- Gene selection
- Literature search
- Functional annotation



Example - Interpretation (2/3)

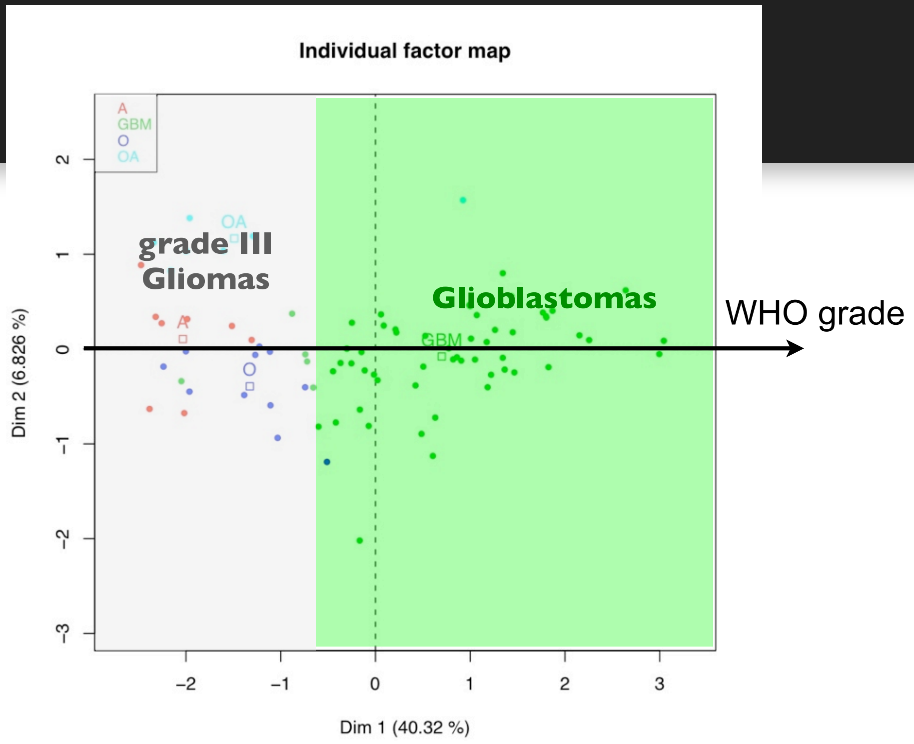
< *Modules* interpretation >

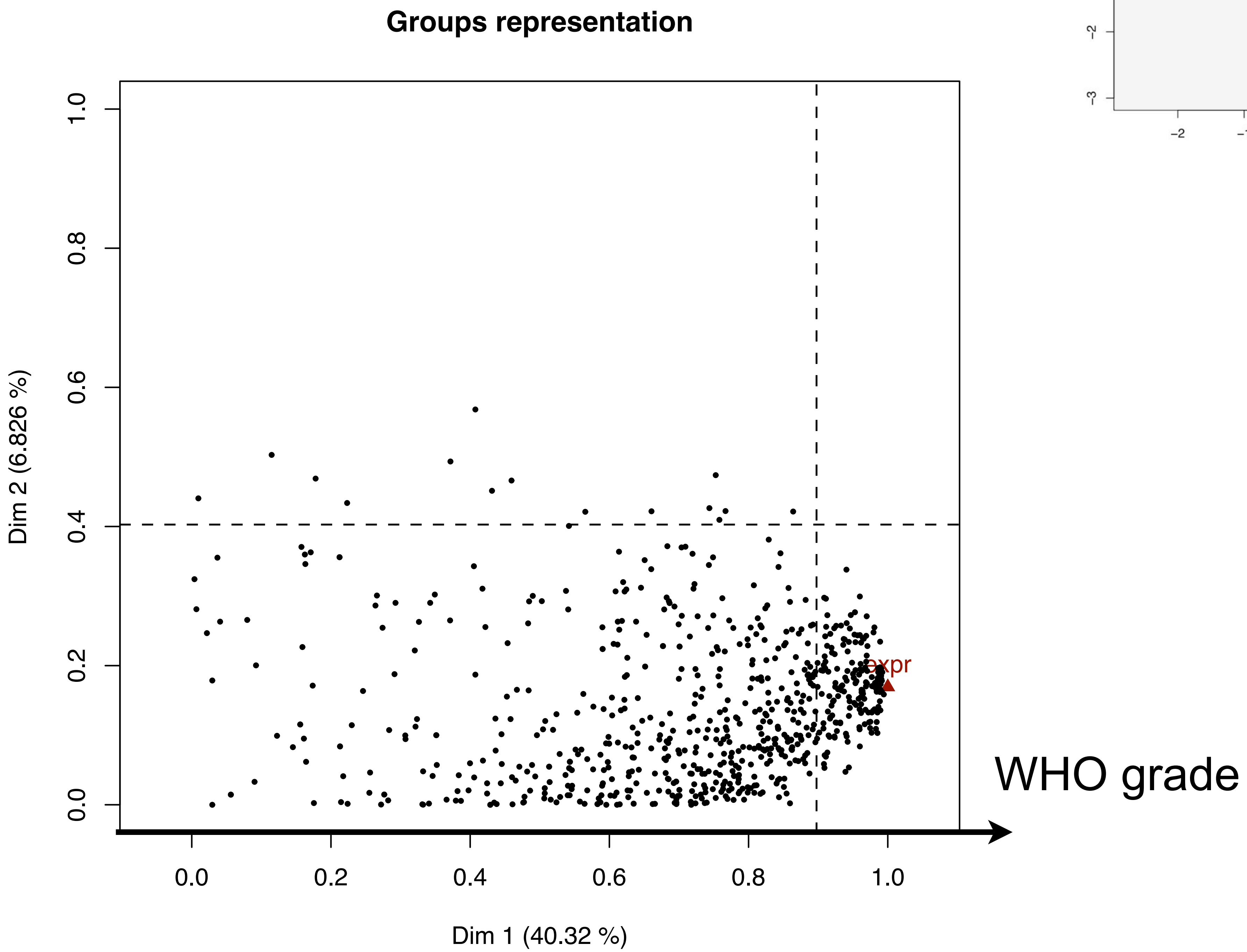




Example - Interpretation (2/3)

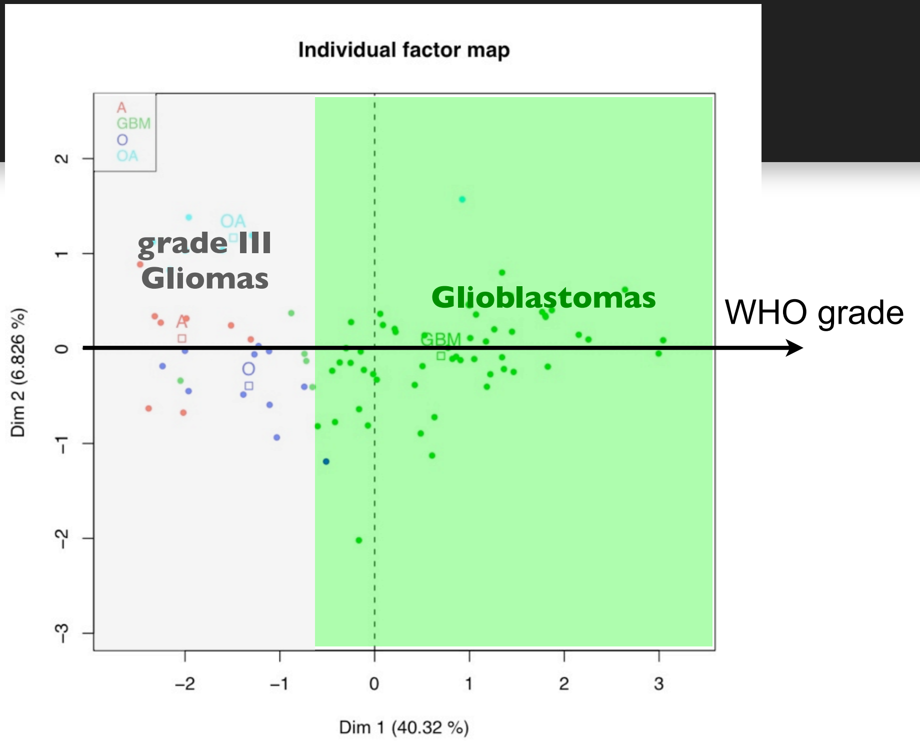
< *Modules* interpretation >

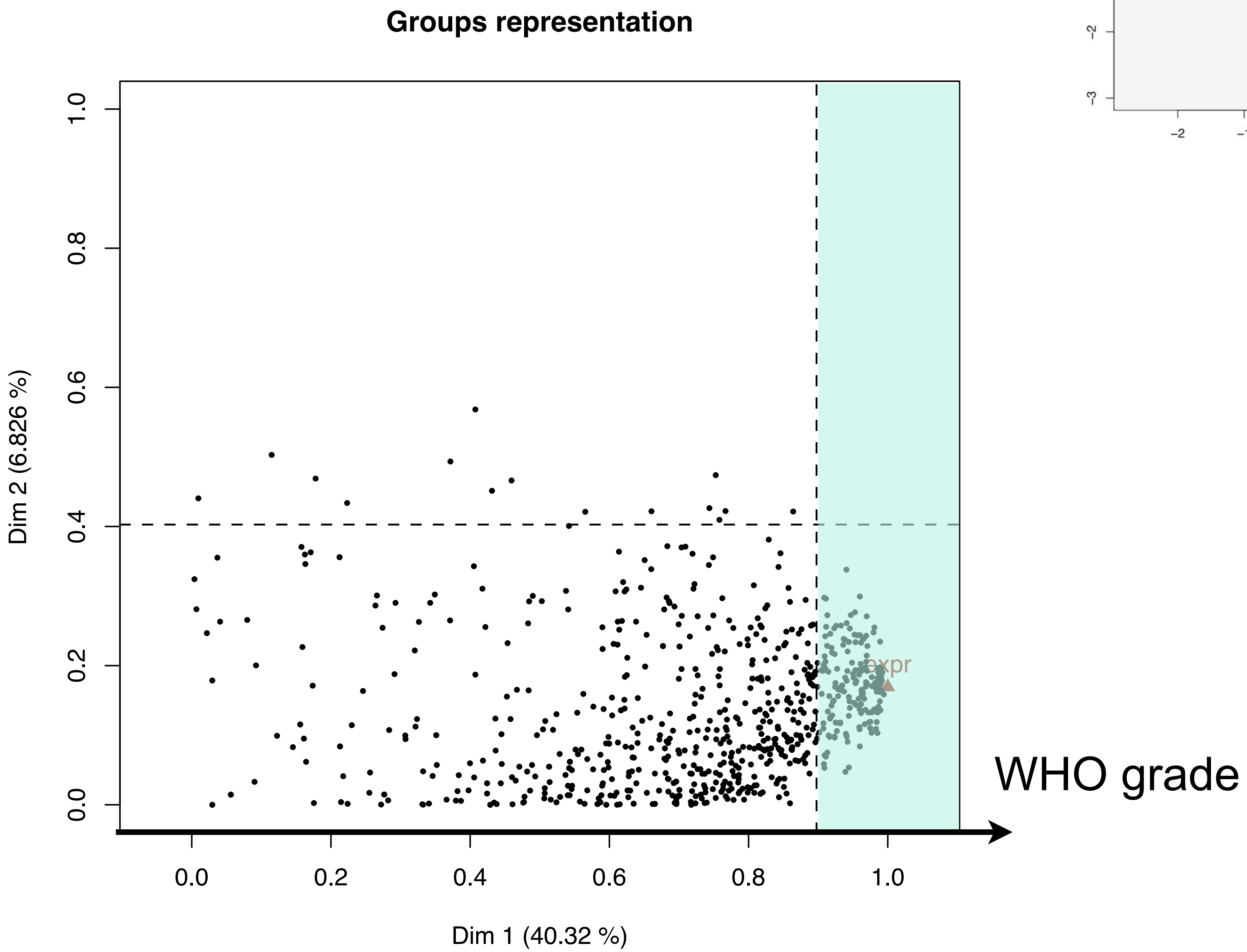




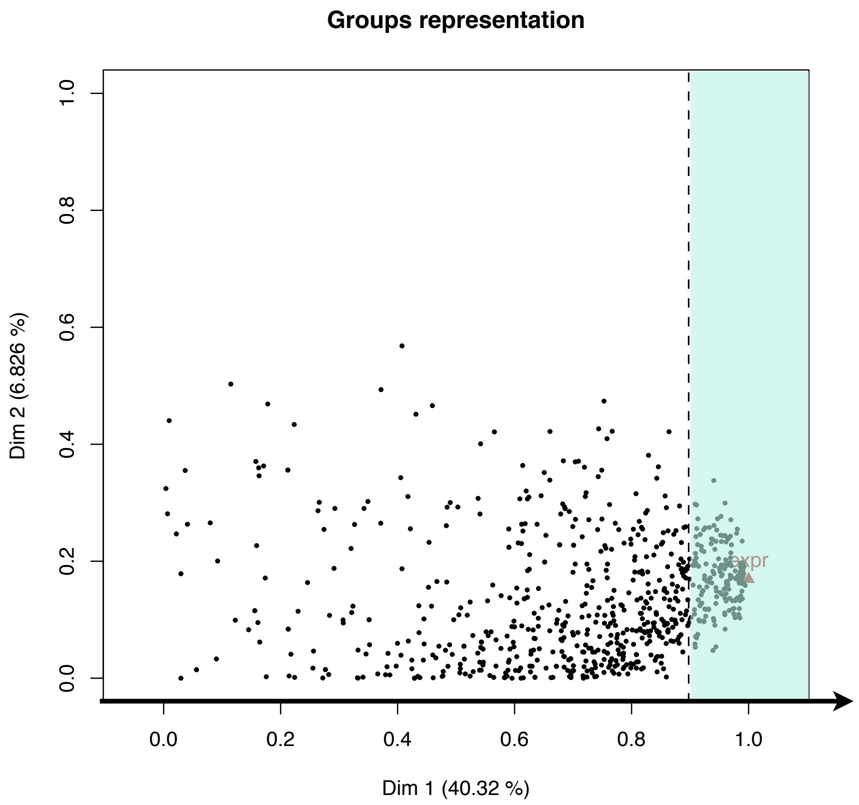
Example - Interpretation (2/3)

< *Modules* interpretation >



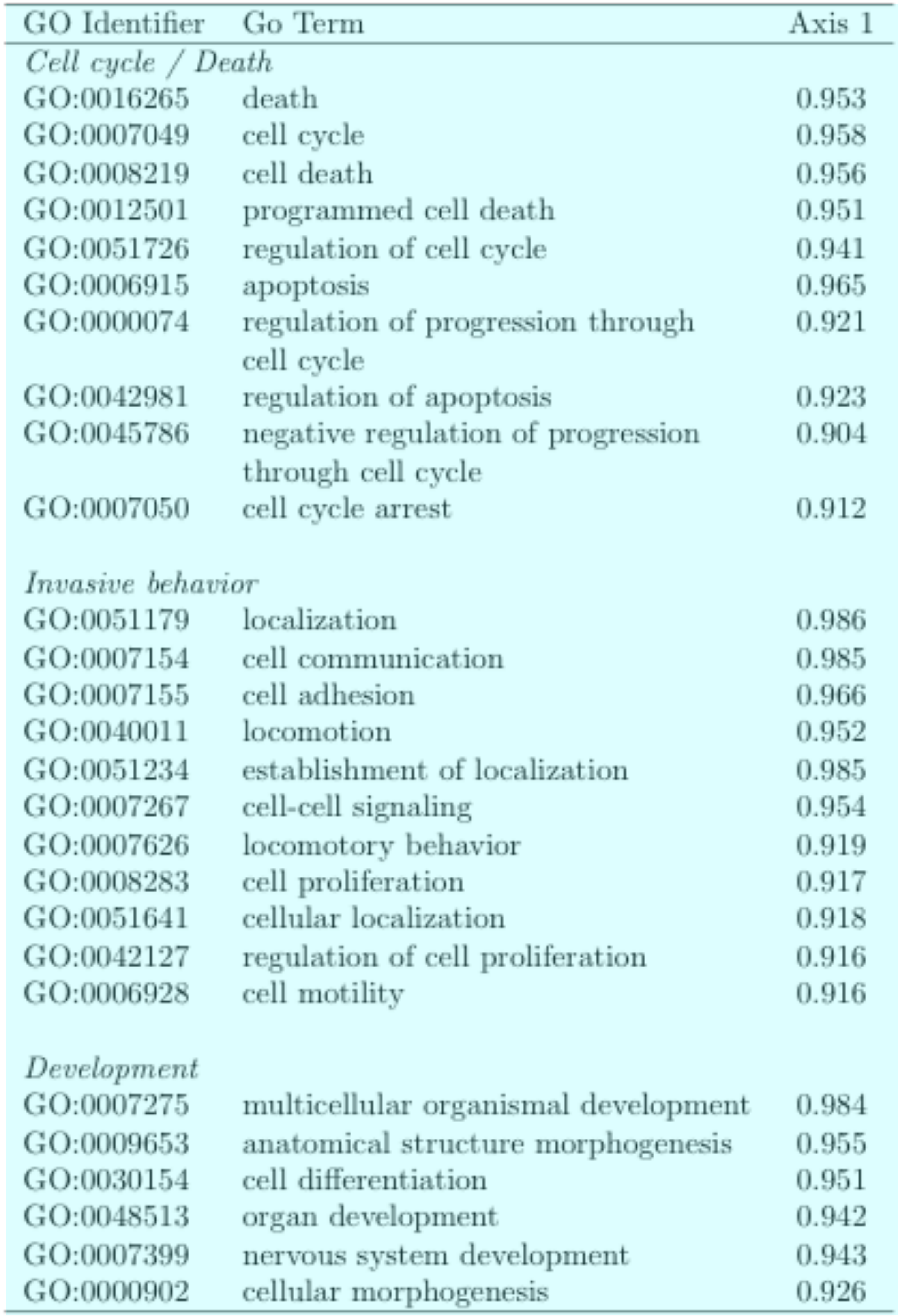


Example - Interpretation (3/3)



Selected biological processes are characteristic of malignant tumors

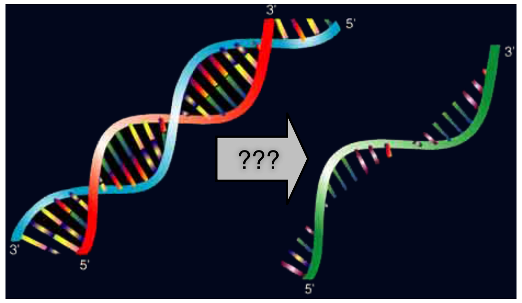
- Proliferation
- Invasion
- Development

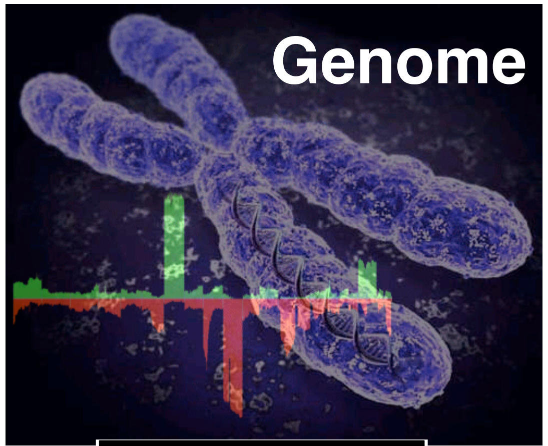


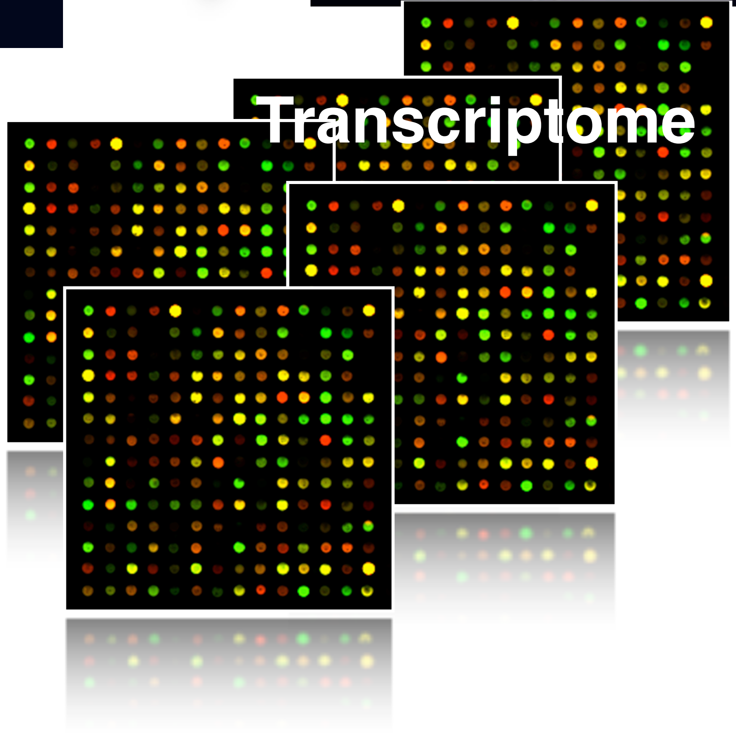
GO Identifier	Go Term	Axis 1
Cell cycle / Death		
GO:0016265	death	0.953
GO:0007049	cell cycle	0.958
GO:0008219	cell death	0.956
GO:0012501	programmed cell death	0.951
GO:0051726	regulation of cell cycle	0.941
GO:0006915	apoptosis	0.965
GO:0000074	regulation of progression through cell cycle	0.921
GO:0042981	regulation of apoptosis	0.923
GO:0045786	negative regulation of progression through cell cycle	0.904
GO:0007050	cell cycle arrest	0.912
Invasive behavior		
GO:0051179	localization	0.986
GO:0007154	cell communication	0.985
GO:0007155	cell adhesion	0.966
GO:0040011	locomotion	0.952
GO:0051234	establishment of localization	0.985
GO:0007267	cell-cell signaling	0.954
GO:0007626	locomotory behavior	0.919
GO:0008283	cell proliferation	0.917
GO:0051641	cellular localization	0.918
GO:0042127	regulation of cell proliferation	0.916
GO:0006928	cell motility	0.916
Development		
GO:0007275	multicellular organismal development	0.984
GO:0009653	anatomical structure morphogenesis	0.955
GO:0030154	cell differentiation	0.951
GO:0048513	organ development	0.942
GO:0007399	nervous system development	0.943
GO:0000902	cellular morphogenesis	0.926

Summary

Defining DNA copy number alteration with direct impact on gene expression

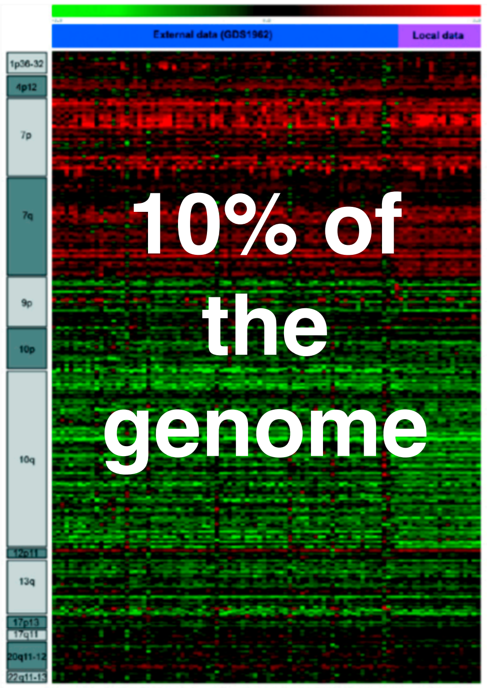




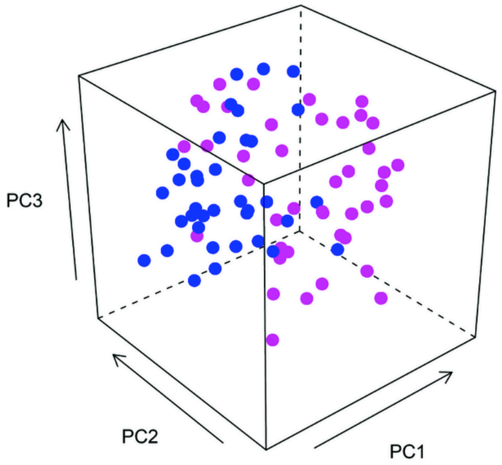


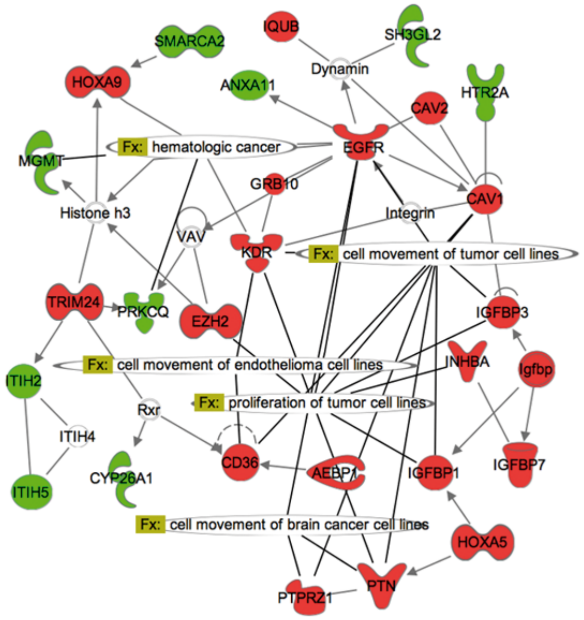
de Tayrac et al., *BMC Genomics* (2009)

Data integration



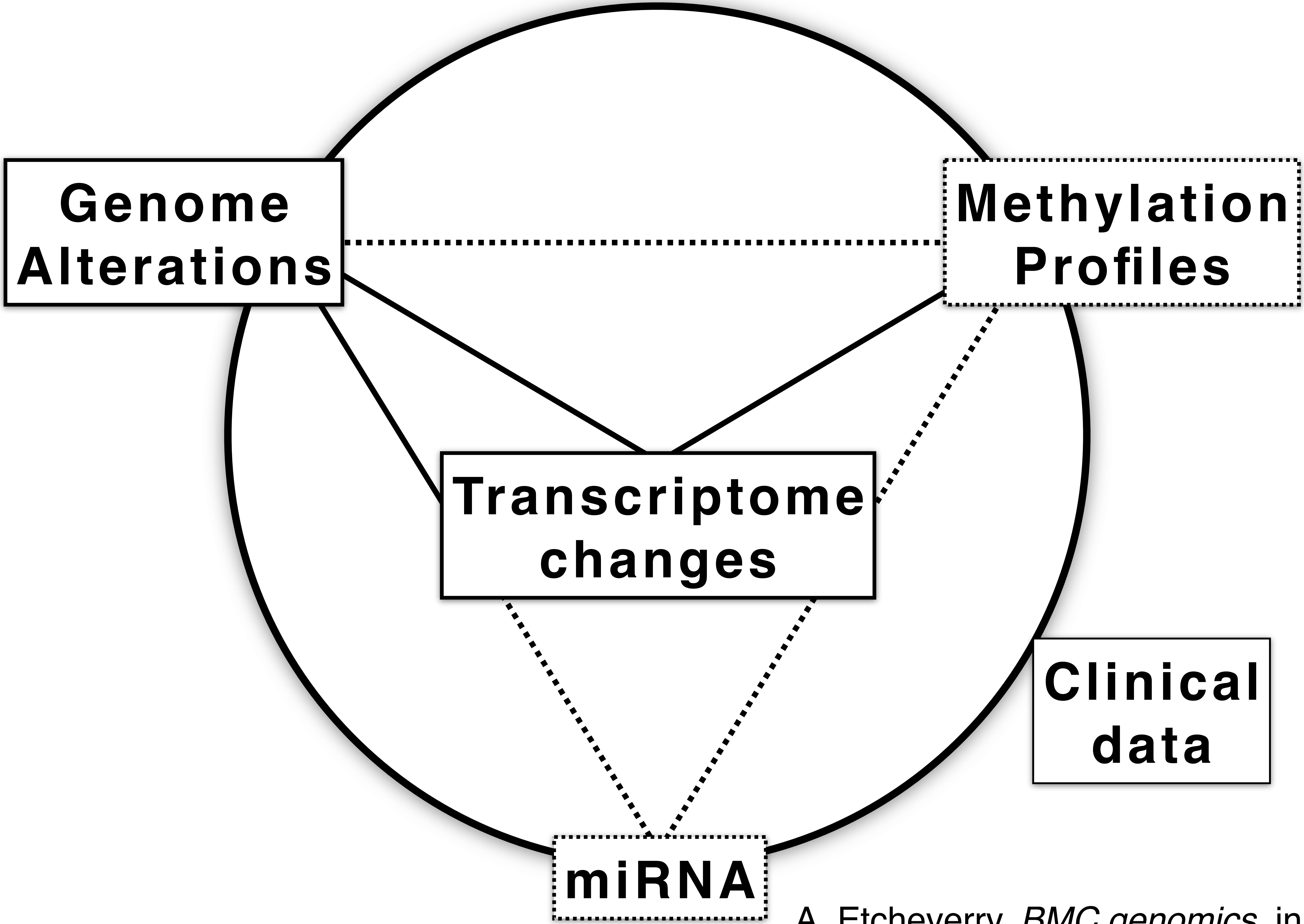
10% of the genome





de Tayrac et al., *Genes Chromosomes Cancer* (2009)

Next Step - Under process



Genome Alterations

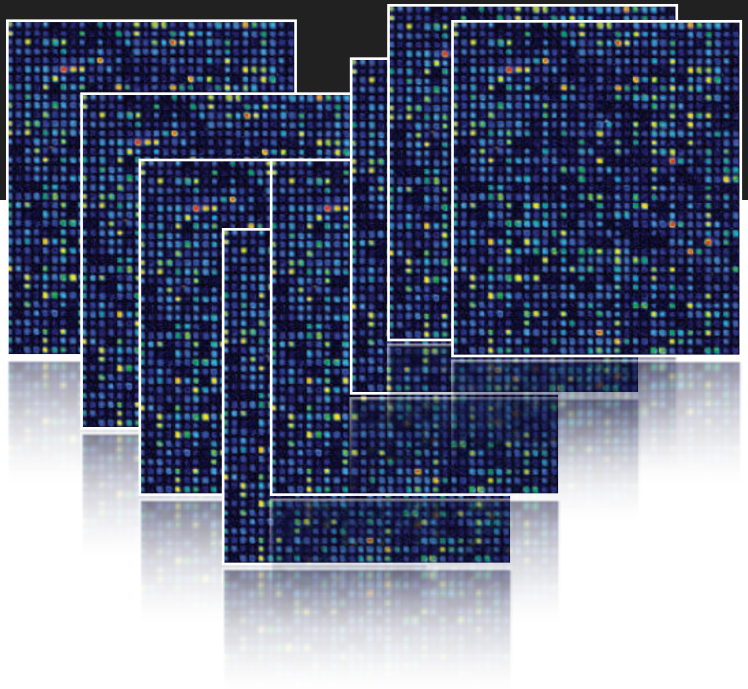
Methylation Profiles

Transcriptome changes

Clinical data

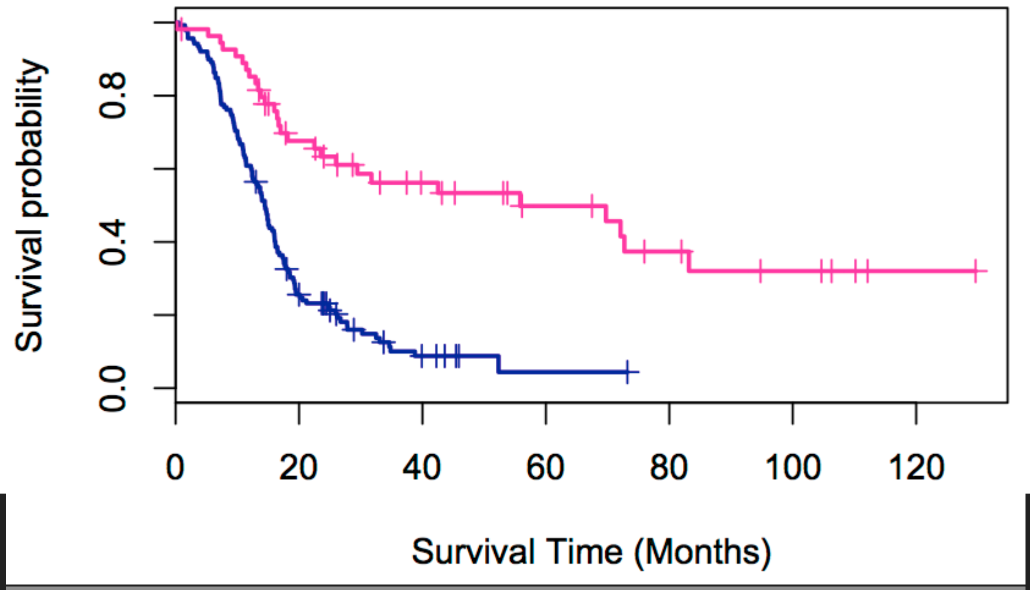
miRNA

A. Etcheverry, *BMC genomics*, in press



Integrative analysis of microarray data

de Tayrac & Aubry,
Clinical Cancer Research, in press



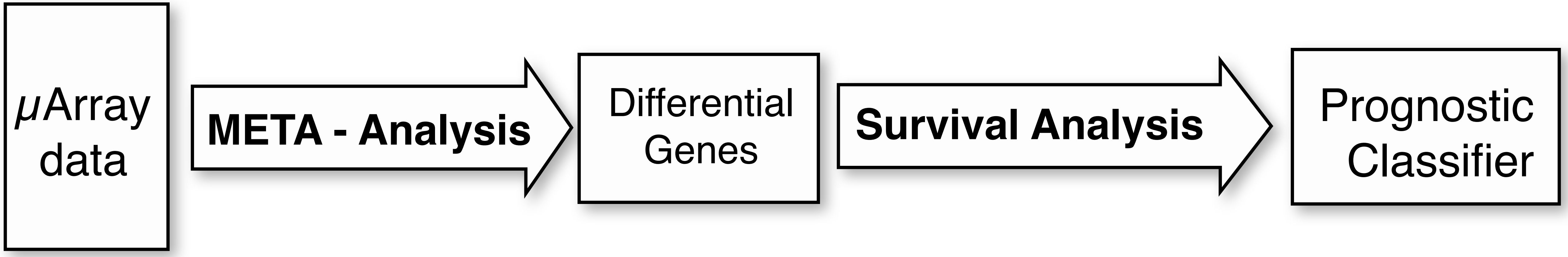
Experimental Design

Gene eXpression data

267 malignant astrocytomas (WHO grade III & grade IV)

➡ 3 studies [GSE4412 - GSE4271 - GSE4019] - Affymetrix

Analysis Workflow



μArray data

META - Analysis

Differential Genes

Survival Analysis

Prognostic Classifier

Validation

56 malignant astrocytomas (III & IV)

➡ [GSE2727] - Affymetrix

194 malignant astrocytomas (III & IV) - Q, RT-PCR analysis

96 malignant astrocytomas (III & IV) - immunohistochemistry

Meta-analysis

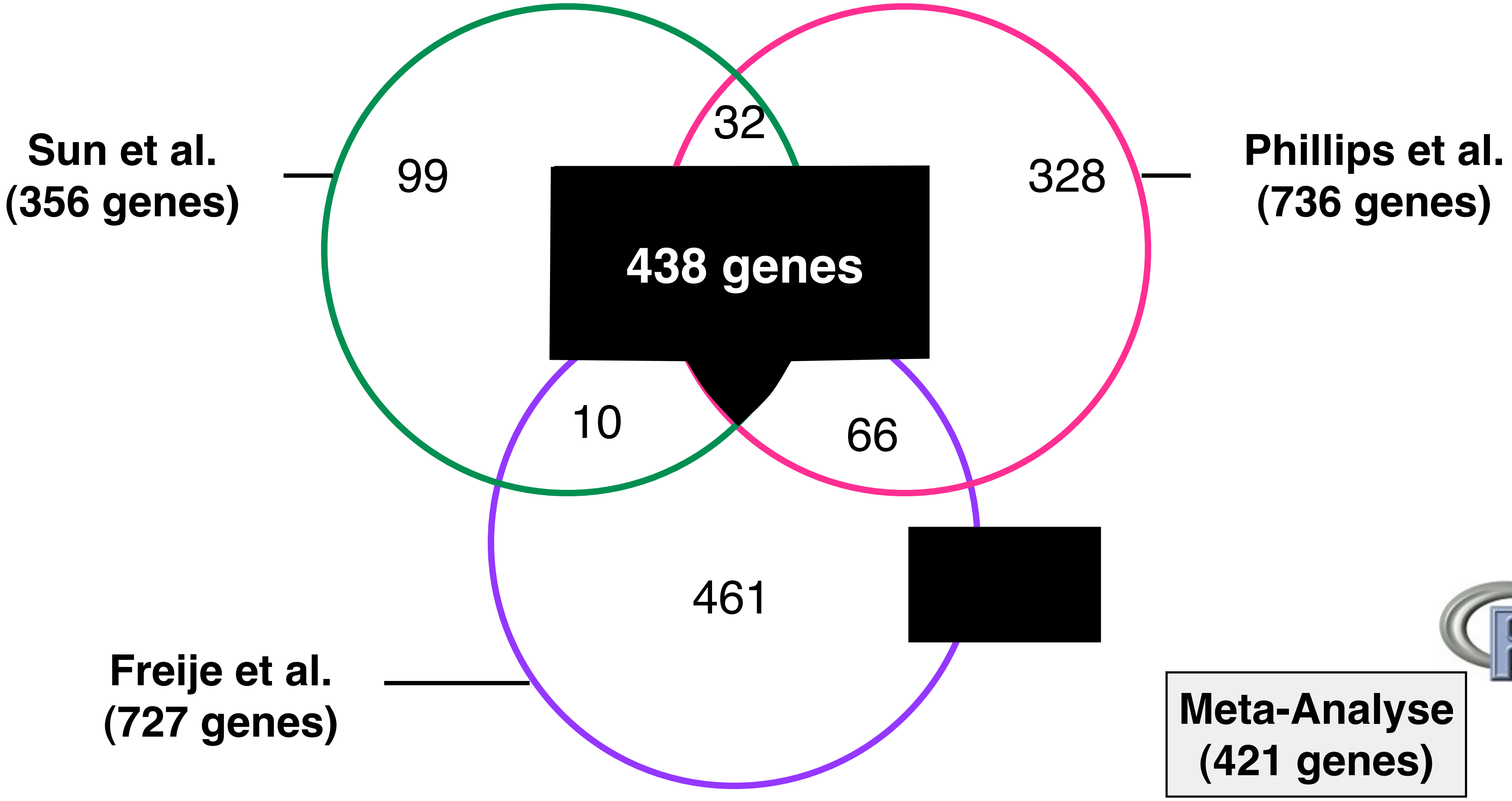
Sun et al. 2006 U133Plus
100 patients
19 grade III
81 grade IV

Phillips et al. 2006 U133A
100 patients
24 grade III
76 grade IV

Freije et al. 2004 U133A
67 patients
8 grade III
59 grade IV

Differential analysis, study by study - grade III vs grade IV

Meta-analysis - grade III vs grade IV



[redacted]

RankProd

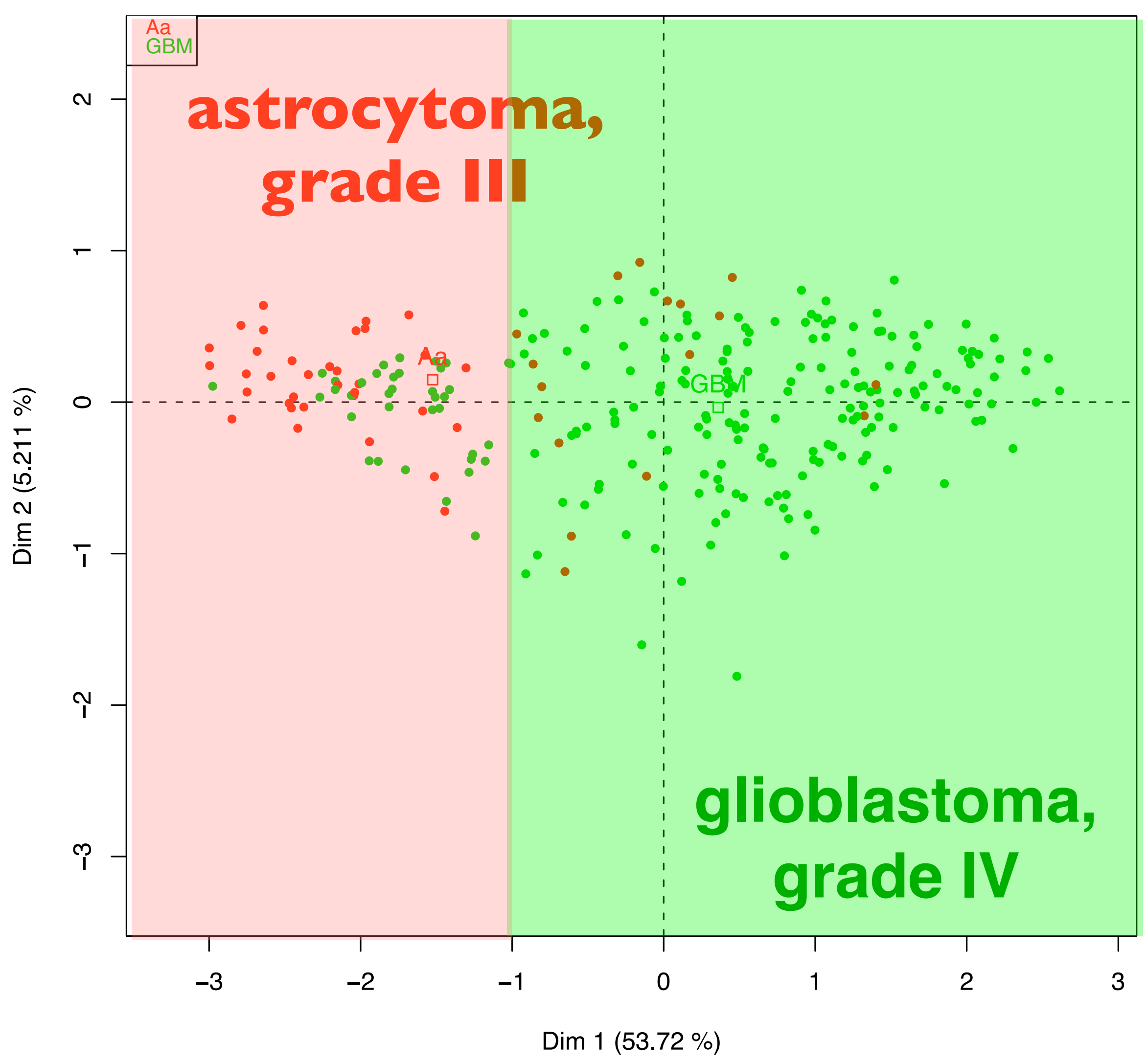
Meta-Analyse
(421 genes)

A 4-gene model associated to GBM outcome

Meta-analysis of microarray data

438 genes - tumor aggressiveness

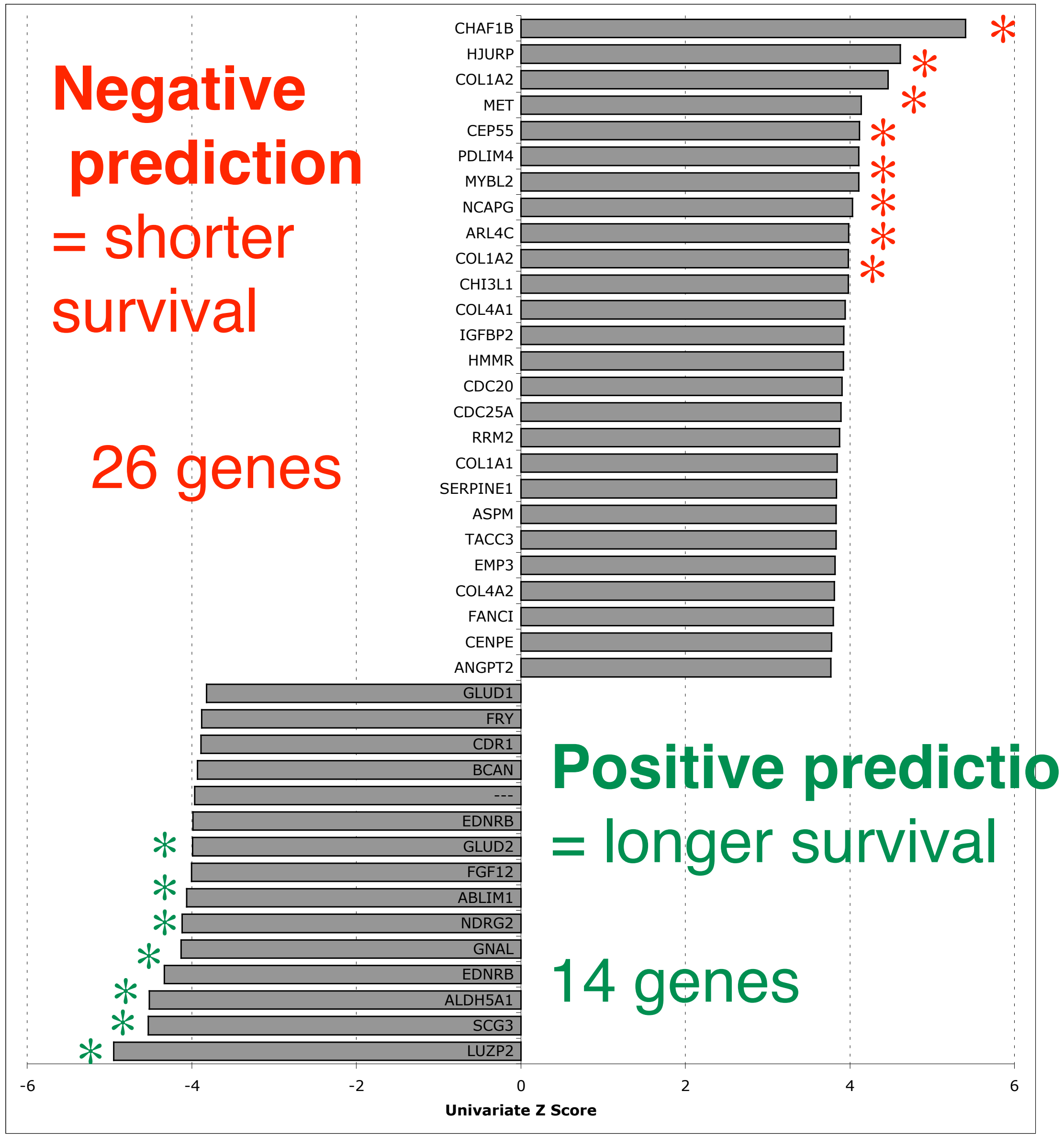
Individual factor map



astrocytoma, grade III

glioblastoma, grade IV

40 genes correlated with outcome



Negative prediction
= shorter survival

26 genes

Positive prediction
= longer survival

14 genes

RISK-SCORE = f(CHAF1B, PDLIM4, EDNRB, HJURP)

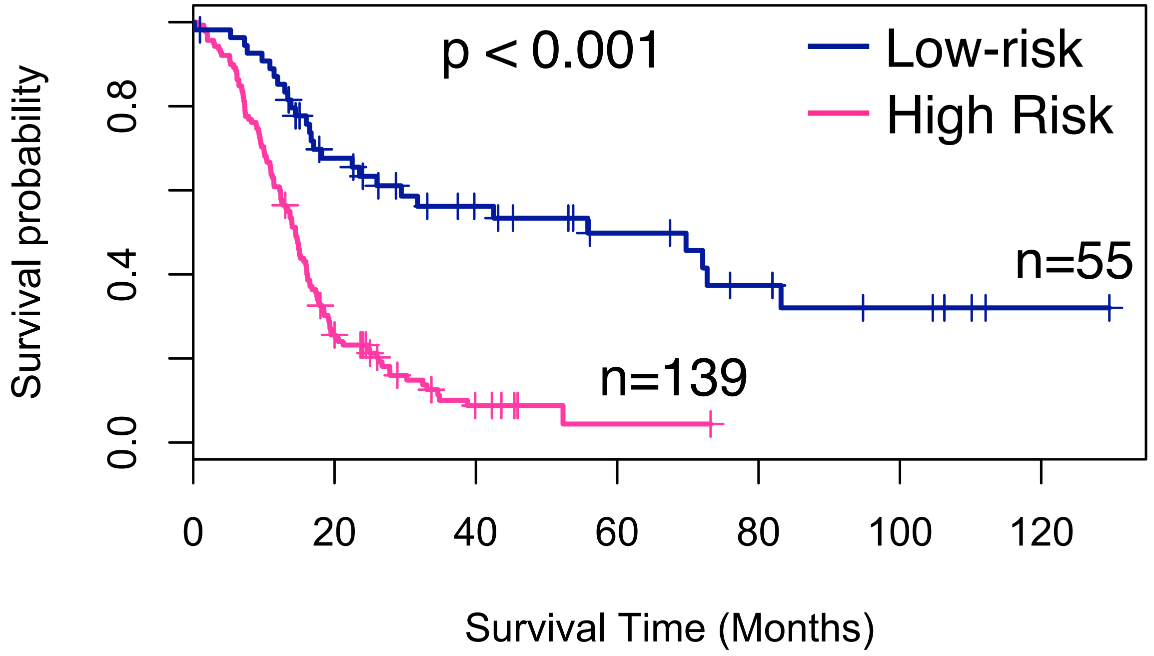
Validation of the 4-gene model

Clinical Data

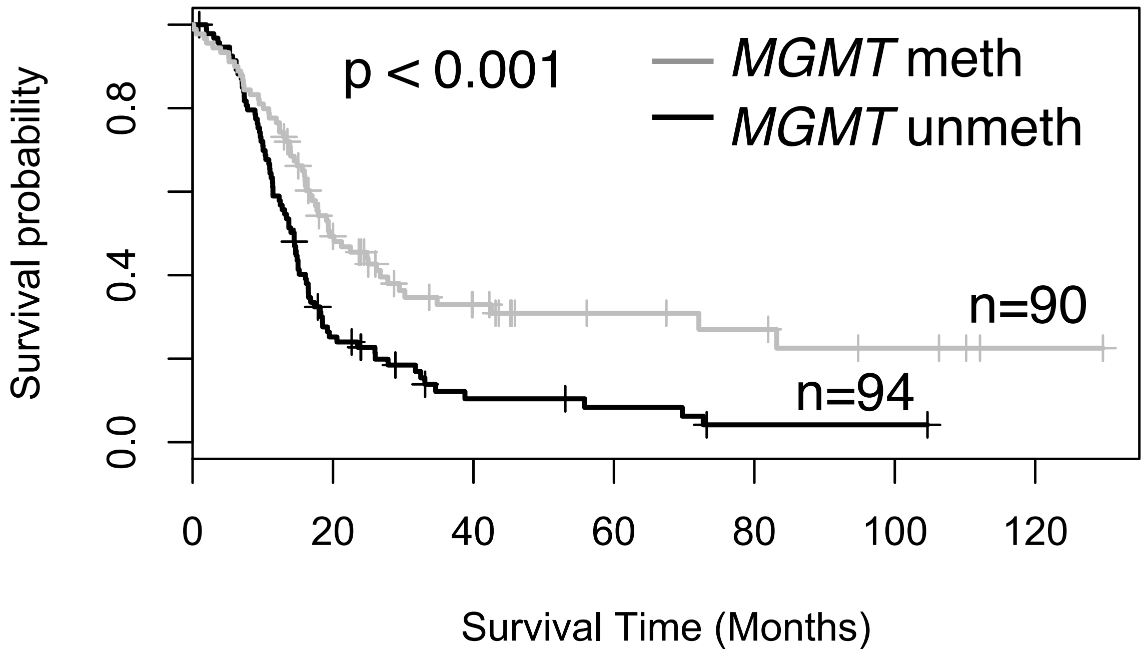
Glioblastoma	145
Anaplastic astrocytoma	38
Anaplastic oligodendroglioma	11

- Age
- Grade
- KPS
- Classes RPA [RTOG]
- Treatment

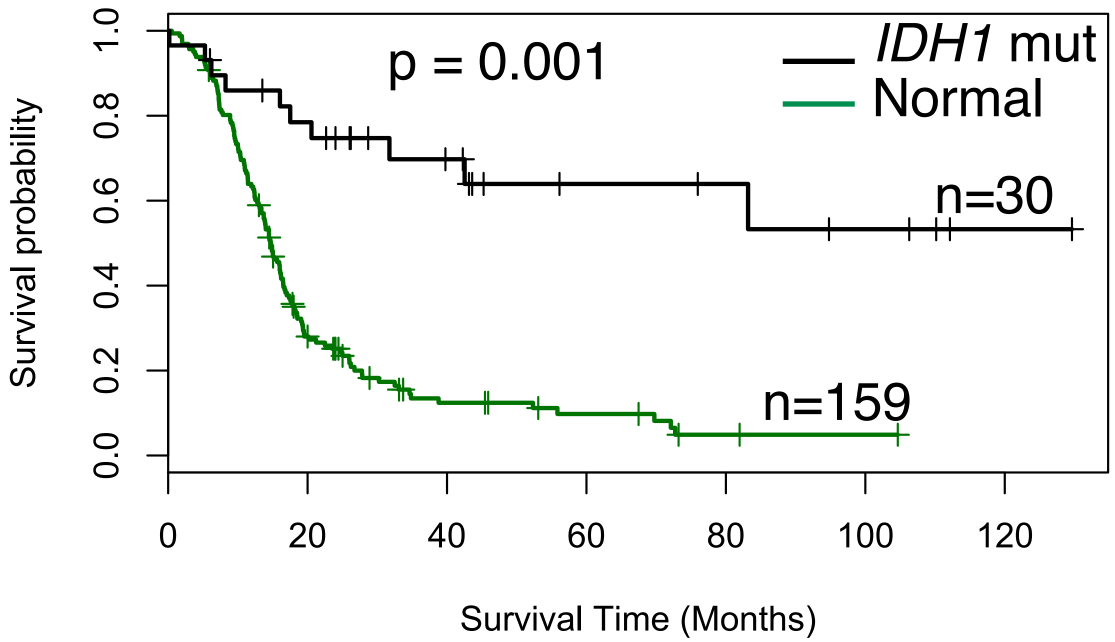
4-genes



MGMT methylation



IDH1 status



Evaluation of the 4-gene model performances

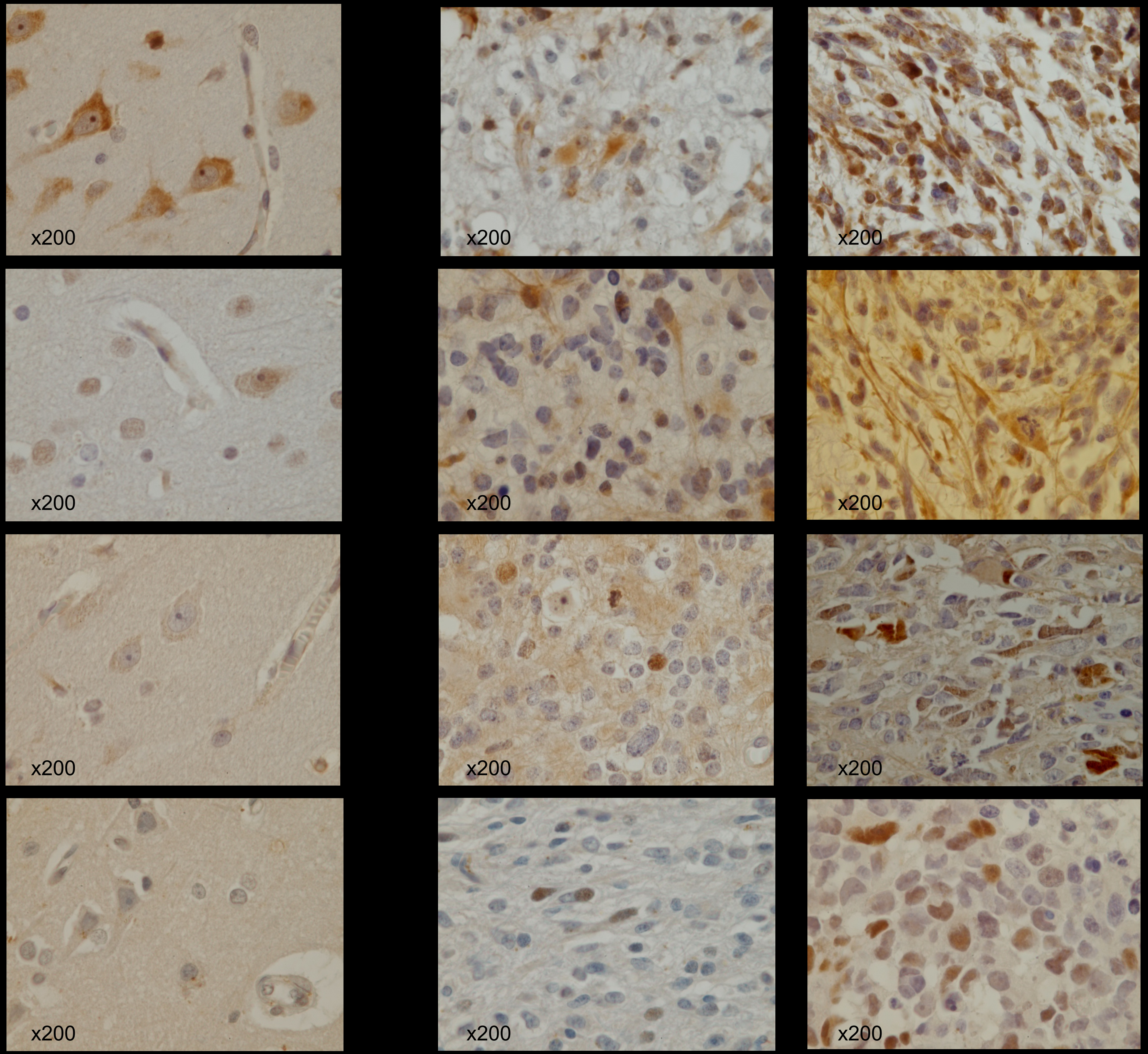
	No *4-genes*	Plus *4-genes*
CLINICAL FACTORS	.	*
CLINICAL F + *IDH1*	**	***
CLINICAL F + *MGMT*	**	***
CLINICAL F + *IDH1 + MGMT*	***	*** .

CLINICAL FACTORS = Age + Treatment + Grade + RTOG RPA

- Highly significant
- Better discrimination
- Good calibration
- HGGs but also significant for glioblastoma treated RT+TMZ

At the protein level...

Cox model $p = 0.0001$



Control Brain

low positivity

Strong positivity

EDN/RB

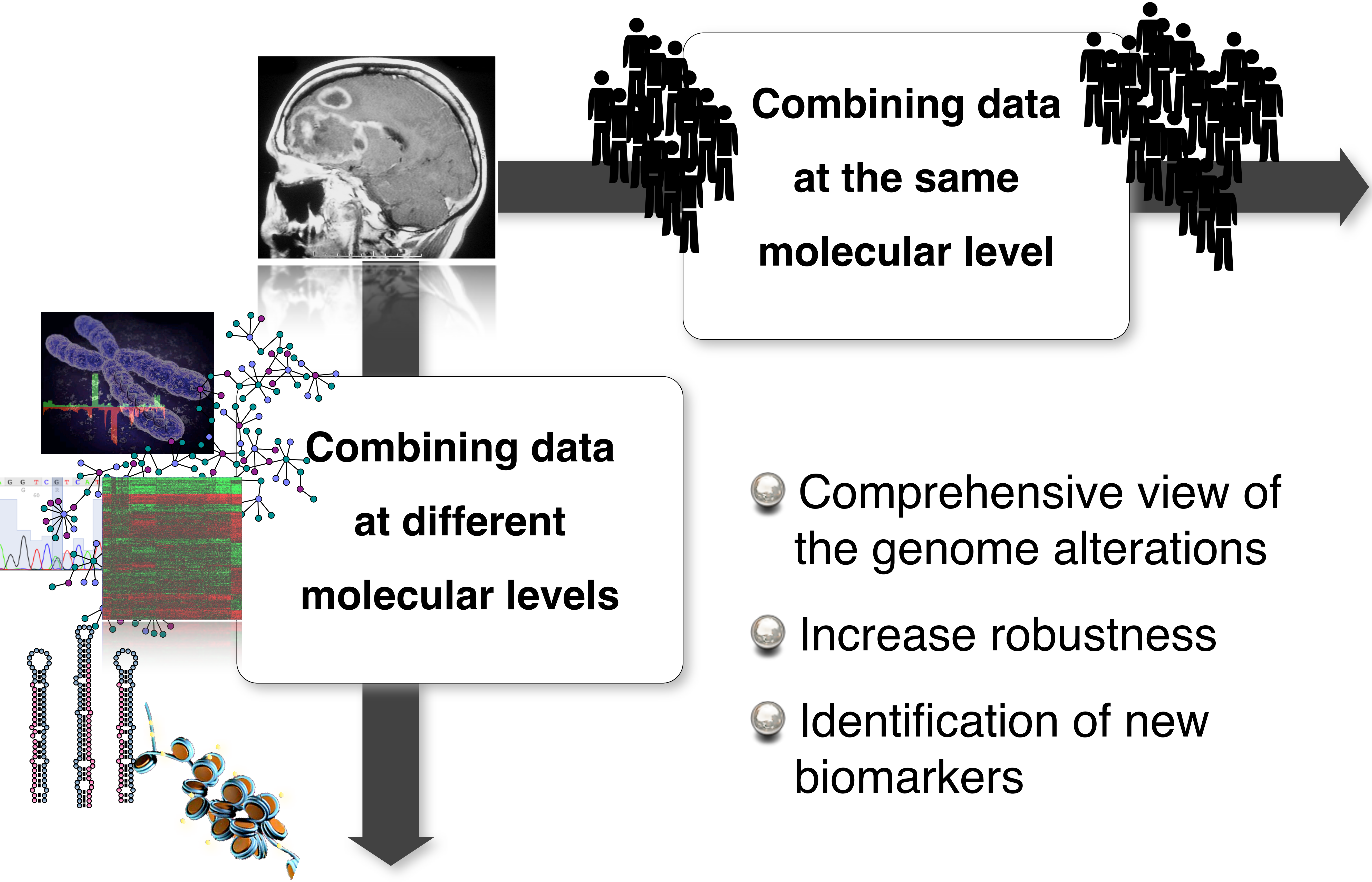
PDLI4

HJURP

p60/CAF-1

de Tayrac & Saikali, in prep

Integrative approaches...



Combining data at the same molecular level

Combining data at different molecular levels

- Comprehensive view of the genome alterations
- Increase robustness
- Identification of new biomarkers

Thanks!

Jean Mosser, Véronique Quillien, Tony Avri

Marc Aubry, Amandine Etcheverry

Sébastien Lê, François Husson

Stephan Saikali, Abderrahmane Hamlat

Thierry Lesimple, Elodie Vauléon

Philippe Menei, Anne Clavreul

Charles Pineau, Emmanuelle Com