

Biocomputación aplicada al análisis de microarrays.

<http://revolutionresearch.uab.es>

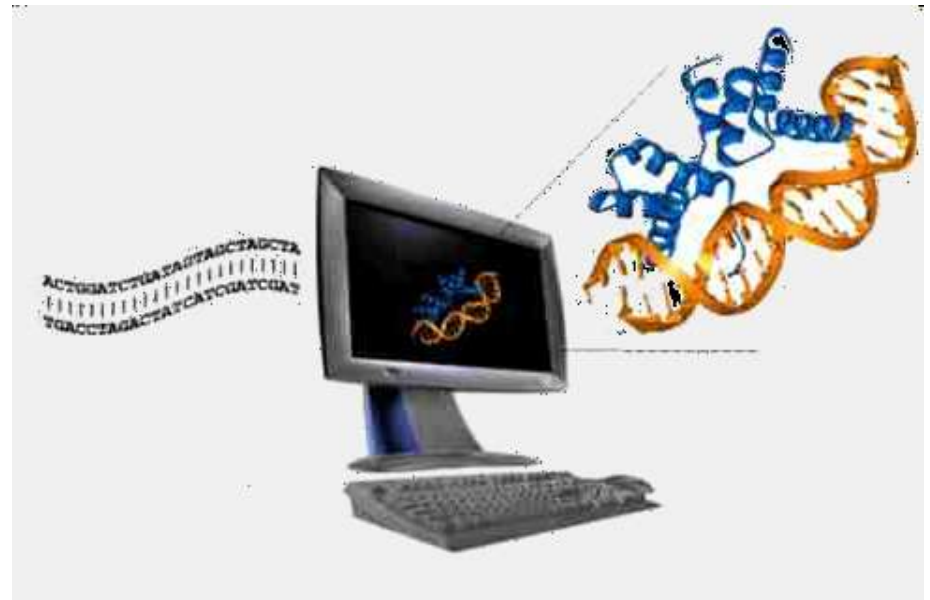
<http://platypus.uab.es>



Biología molecular. Un nuevo tipo de información biológica



ANTES: la naturaleza



AHORA :: Representación de la naturaleza

Dogma de la biologia molecular

Genotip

Informació heretable continguda per tots els éssers vius i codificada en forma de DNA (o RNA) que serveix com a recepta per crear i modelar l'ésser viu.

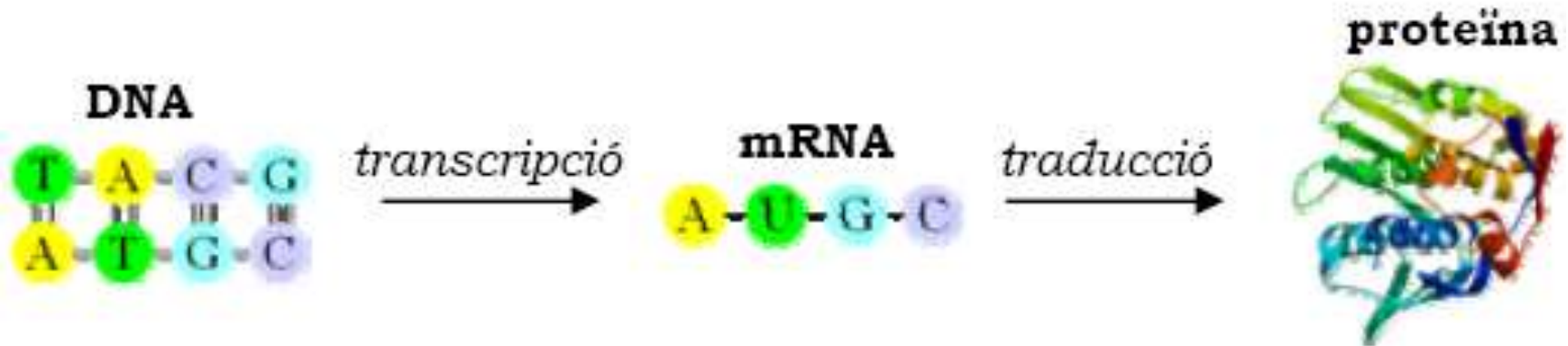


Lamarckisme

Fenotip

Manifestació física de l'organisme (proteïnes, glúcids, lípids...), en funció del seu genotip i de les condicions ambientals en les que es trobi.

DNA->RNA->Proteínas->Fenotipo



DNA - GEN

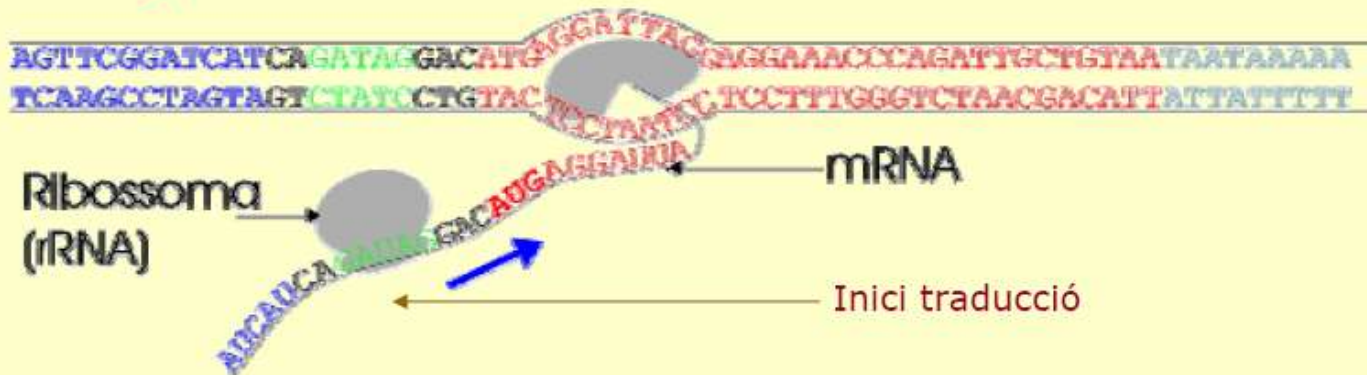


Expresión Gen ->... Proteína->... Función celular.

- Acoblament RNA-polimerasa



- Elongació



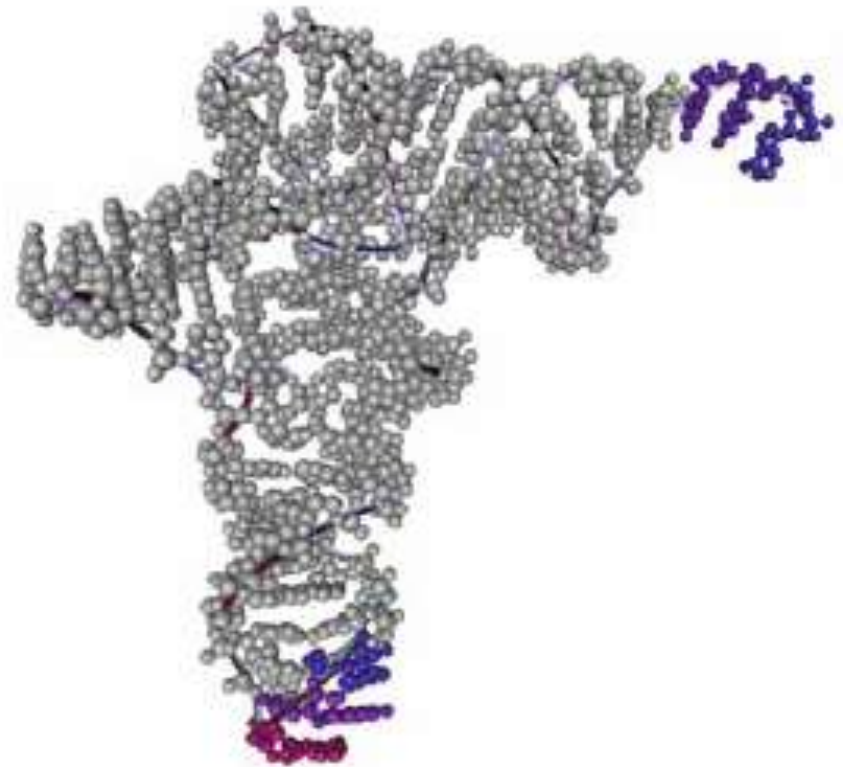
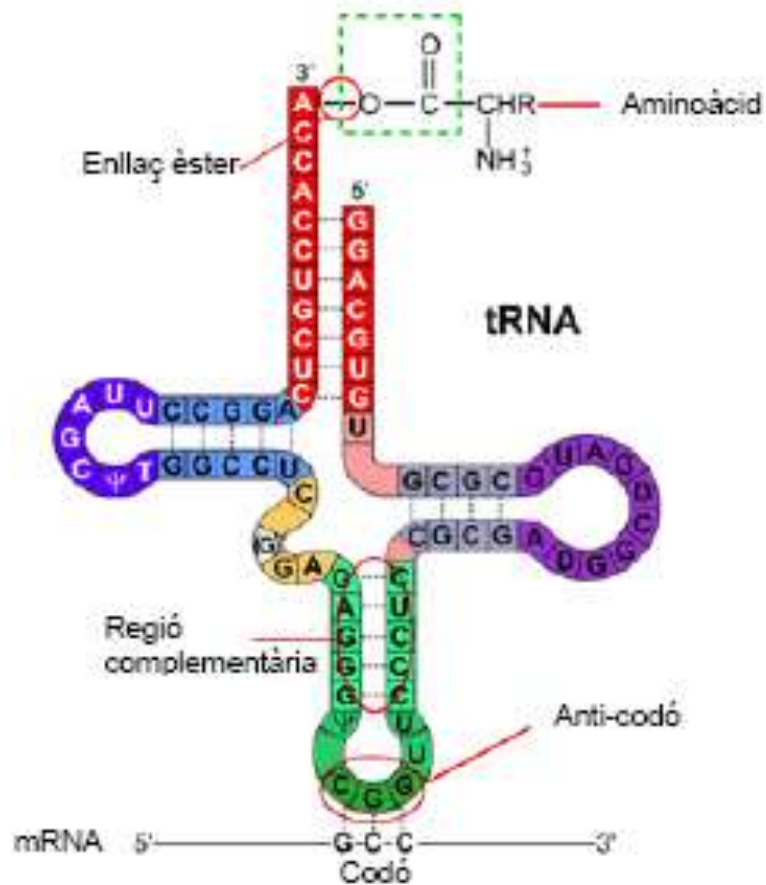
Traducción RNA -> Proteína::

4 bases nitrogenadas -> 20 aminoácidos

		Second Position of Codon					
		T	C	A	G		
RNA	T	TTT Phe [F]	TCT Ser [S]	TAT Tyr [Y]	TGT Cys [C]	T	Third Position
		TTC Phe [F]	TCC Ser [S]	TAC Tyr [Y]	TGC Cys [C]	C	
		TTA Leu [L]	TCA Ser [S]	TAA Ter [end]	TGA Ter [end]	A	
		TTG Leu [L]	TCG Ser [S]	TAG Ter [end]	TGG Trp [W]	G	
	C	CTT Leu [L]	CCT Pro [P]	CAT His [H]	CGT Arg [R]	T	
		CTC Leu [L]	CCC Pro [P]	CAC His [H]	CGC Arg [R]	C	
		CTA Leu [L]	CCA Pro [P]	CAA Gln [Q]	CGA Arg [R]	A	
		CTG Leu [L]	CCG Pro [P]	CAG Gln [Q]	CGG Arg [R]	G	
	A	ATT Ile [I]	ACT Thr [T]	AAT Asn [N]	AGT Ser [S]	T	
		ATC Ile [I]	ACC Thr [T]	AAC Asn [N]	AGC Ser [S]	C	
		ATA Ile [I]	ACA Thr [T]	AAA Lys [K]	AGA Arg [R]	A	
		ATG Met [M]	ACG Thr [T]	AAG Lys [K]	AGG Arg [R]	G	
	G	GTT Val [V]	GCT Ala [A]	GAT Asp [D]	GGT Gly [G]	T	
		GTC Val [V]	GCC Ala [A]	GAC Asp [D]	GGC Gly [G]	C	
		GTA Val [V]	GCA Ala [A]	GAA Gln [E]	GGA Gly [G]	A	
		GTG Val [V]	GCG Ala [A]	GAG Gln [E]	GGG Gly [G]	G	

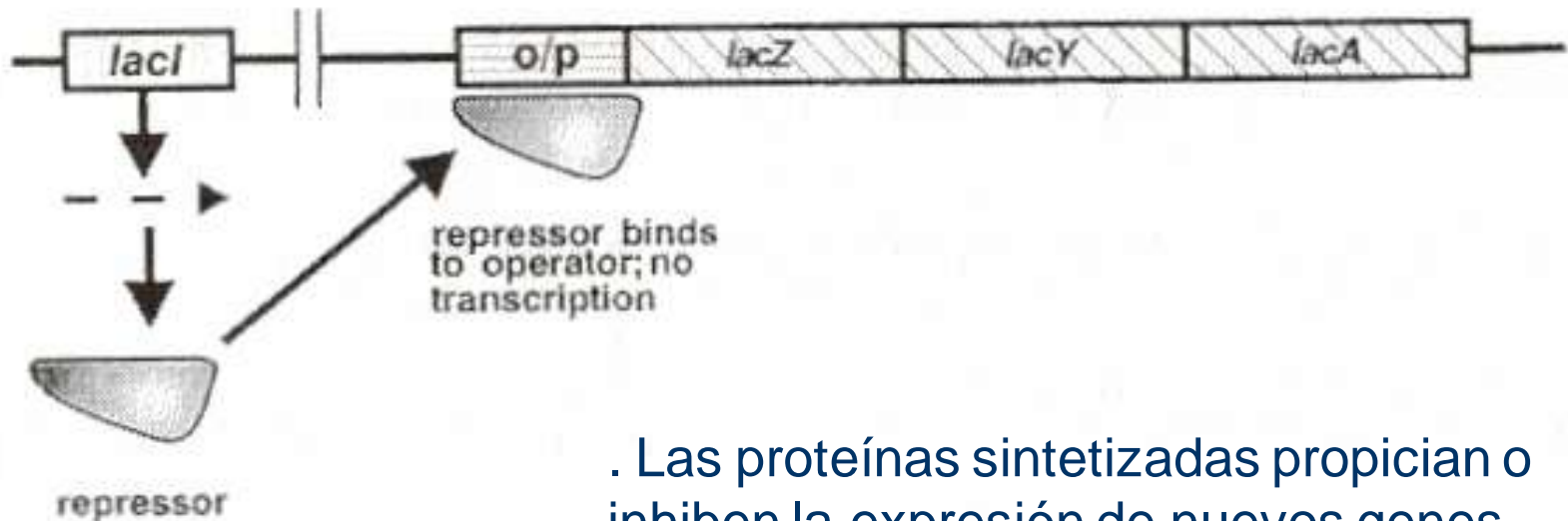
Proteína

Proteínas :: Estructura espacial funcional



Regulación génica :: El sistema se retroalimenta

a) No inducer



- . Las proteínas sintetizadas propician o inhiben la expresión de nuevos genes.
- . Además lo hacen interactuando entre ellas y con señales externas.

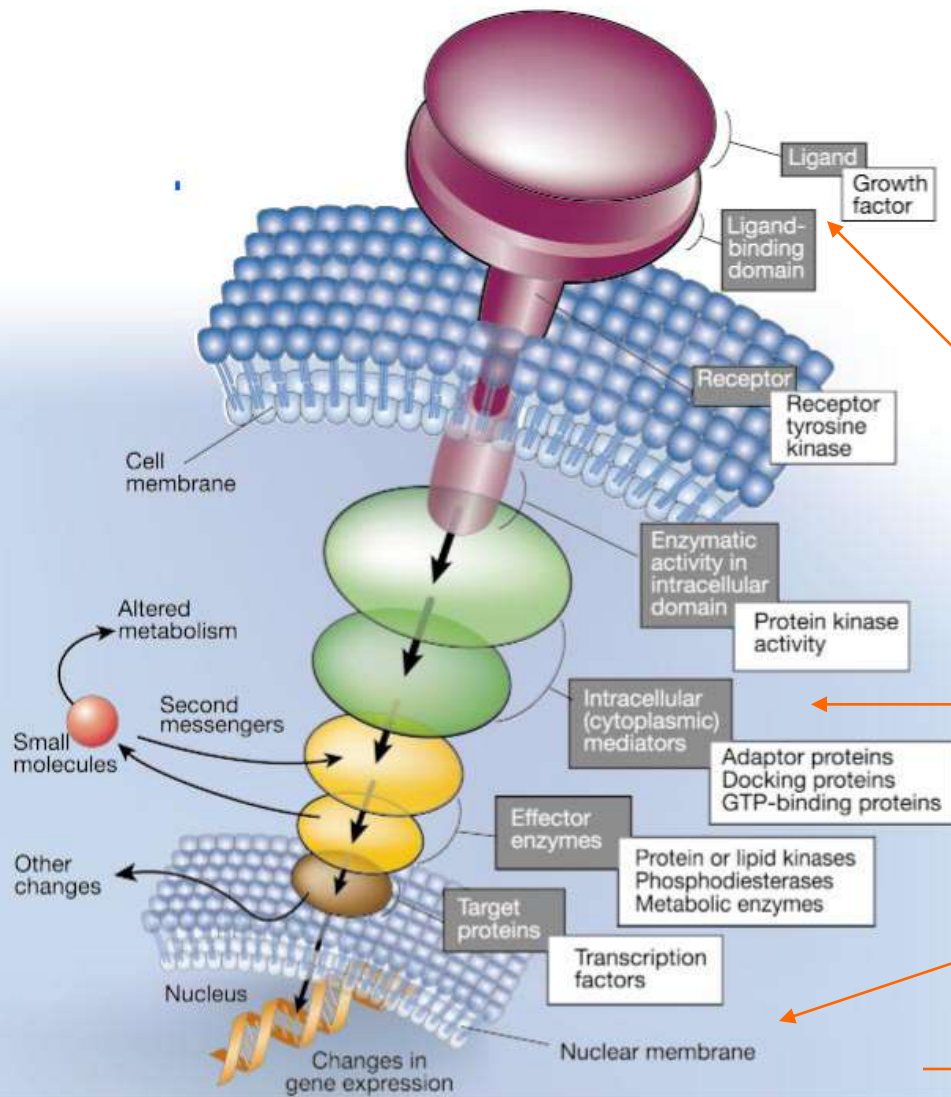
Comunicación con el exterior

Stimulus

Signal Transduction

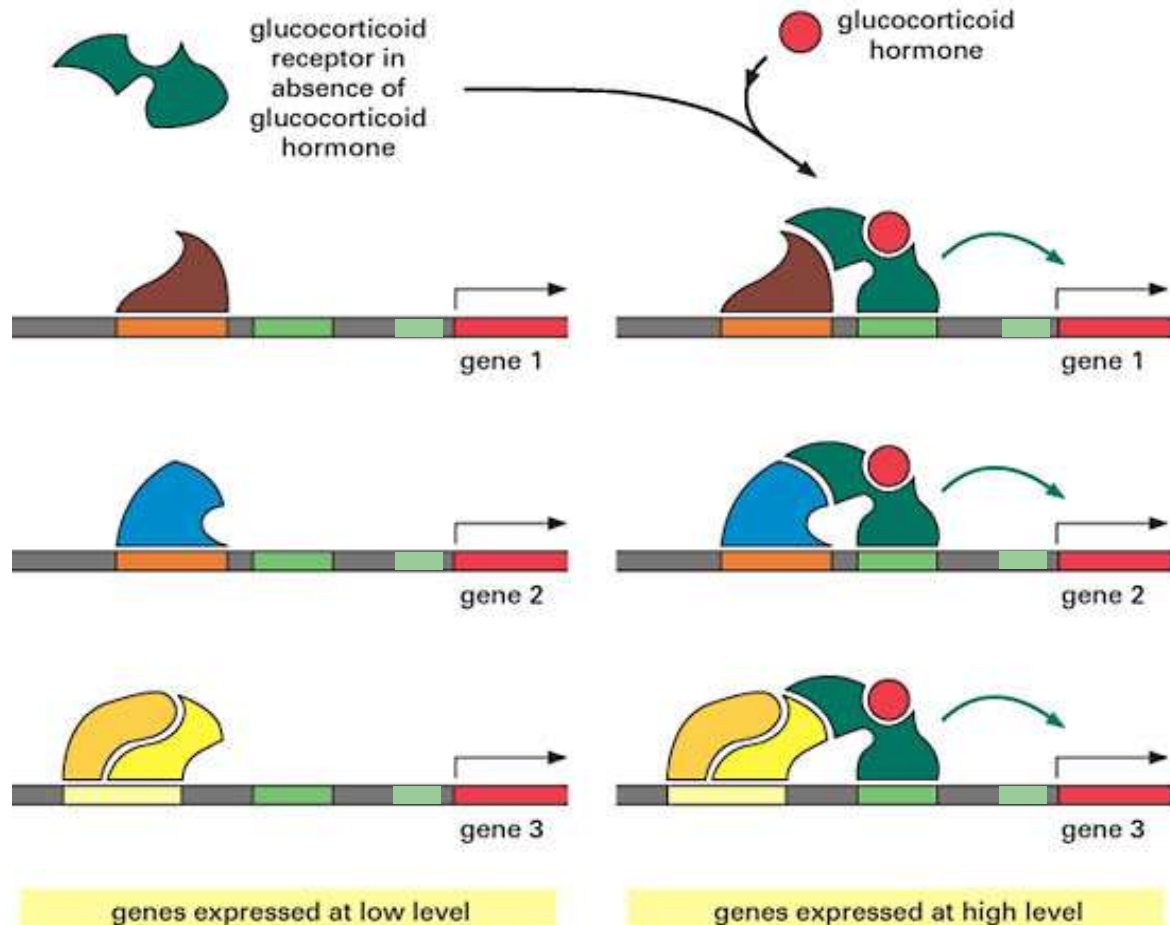
Transcription

mRNA



Por qué varía el grado de expresión de los genes.

Promoter 
Enhancer  



Problemas que afronta la biología computacional.

Problems in Biological Science		Math/Stat/CompSci method
Similarity search		Optimization algorithms <ul style="list-style-type: none"> • Dynamic programming (DP) • Simulated annealing (SA) • Genetic algorithms (GA) • Markov Chain Monte Carlo (MCMC: Metropolis and Gibbs samplers) • Hopfield neural network
	Pairwise sequence alignment Database search for similar sequences Multiple sequence alignment Phylogenetic tree reconstruction Protein 3D structure alignment	
Structure/function prediction	<i>ab initio</i> prediction	RNA secondary structure prediction RNA 3D structure prediction Protein 3D structure prediction
	Knowledge based prediction	Pattern recognition and learning algorithms <ul style="list-style-type: none"> • Discriminant analysis • Neural networks • Support vector machines • Hidden Markov models (HMM) • Formal grammar • CART
Molecular classification		Clustering algorithms <ul style="list-style-type: none"> • Hierarchical, k-means, etc • PCA, MDS, etc • Self-organizing maps, etc
		Superfamily classification Ortholog/paralog grouping of genes 3D fold classification

Genómica :: alineamiento de secuencias.

(A)

YCL030C HIS4

	GCN4	BAS1	PHO2	RAP1	GCN4
SCer	gCAGTCGAACTGACTCTAATAGT	GACTCCGGTAAATTAGT	TAATTAAT	TGCTAAACCCATGCACAGT	GACTCACC
SPar	gCAGTCGAACTGACTCTAATAGT	GACTCCGGTAAATTAGT	TAATTAAT	TGCTAAACCCATGCACAGT	GACTCATG
SMik	gCGGTCAAACCTGACTCTAATAGT	GACTCCGGTAAATTAGT	TAATTAAT	TGCTAAACCCATGCACAGT	GACTCATG
SBay	-TGAACGAACTGACTCTAATAGT	GACTCTGGTAAATTAGT	TAATTAAT	TCTAAACCCATGCACAGT	GACTCATG

* * * * *

	TATA	
SCer	----ATGAACAGTAGTATACTGTGTATATAA	TAGATATGGAACGtATATTCaCCTCCGATGTGTGTTGTACATACATAAA
SPar	----ATGAACAGTAATATACTATGTATATAA	TAGATAAGGAACGtATATTCaCCTTGGATGTGTGTTGTACATACATAAG
SMik	----AAGAACAGTGGTACATGGTGTATATAA	TAGATAAGGAACAT-GTATTCaCTTTTAATGTGAGTTGTCTATA-AGAAG
SBay	caaaATAAATACTAGTGTATGTGTATATAA	CAGAGATGGAACACtGGATTC-CACCTAATGTGTGTTGTCCATACATAGA

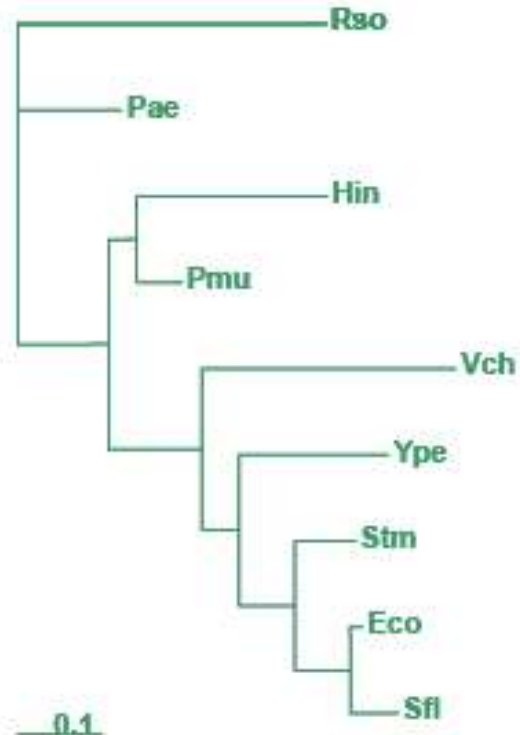
* * * * *

SCer	ATAGTTTACAAaatttTTTTTCGAATA---
SPar	ATAGTTAACAA-----TTTTTTGAATA---
SMik	ACAATTTACAAgatt-TTTTTTGAATA---
SBay	ATAGATTACAAaa---TTTTGGAAAAAaaa

* * * * *

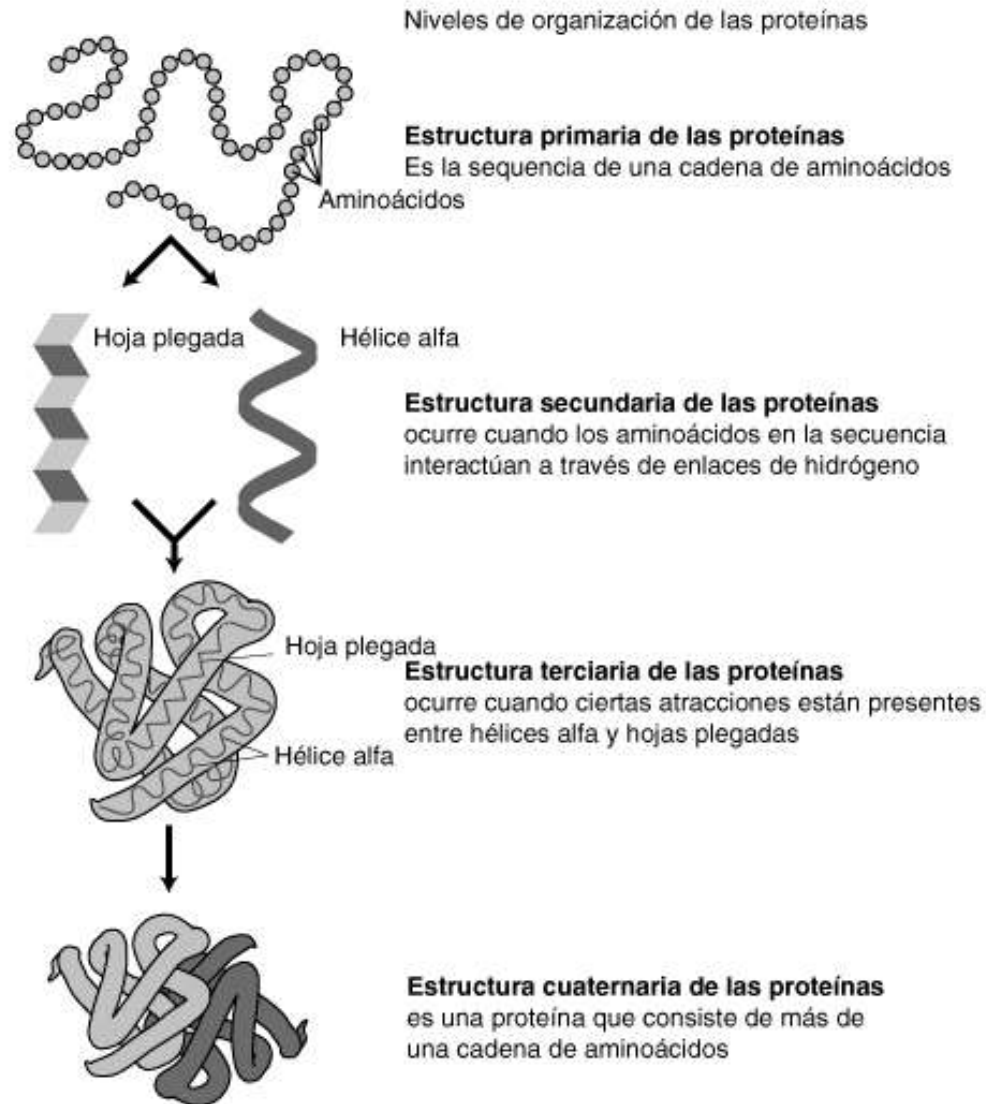
GAP

(B)



Proteómica:: Determinar la funcionalidad, interactividad y posibilidades de la proteína.

La Estructura espacial es clave en la operatividad de la proteína.



Problemas NP complejos -> Métodos heurísticos y aproximación polinomial

Utilizan principios matemáticos, sobre todo el análisis probabilístico para encontrar resultados cercanos al óptimo.

Fixed-parameter tractability : an approach to attacking NP-hard problems with multiple inputs. *Abordaje de un problema NP-complejo usando múltiples parámetros o entradas que acaban reduciendo el problema.*

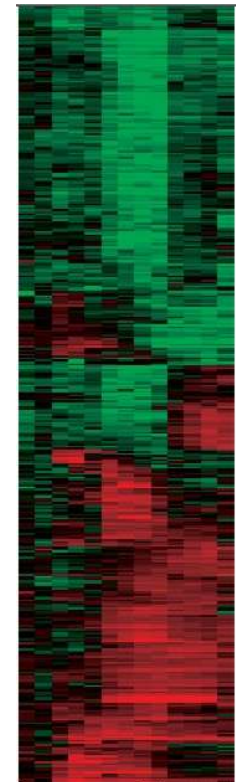
Expresión génica :: el comportamiento celular.

- Tecnología de Microarrays.
 - Permiten estudiar en qué grado y bajo qué circunstancias se van expresando los genes.
 - Ventajas : Tenemos el nivel de expresión de miles de genes bajo las mismas circunstancias.
 - Limitación : No tenemos la interacción de las proteínas producto de esa expresión.

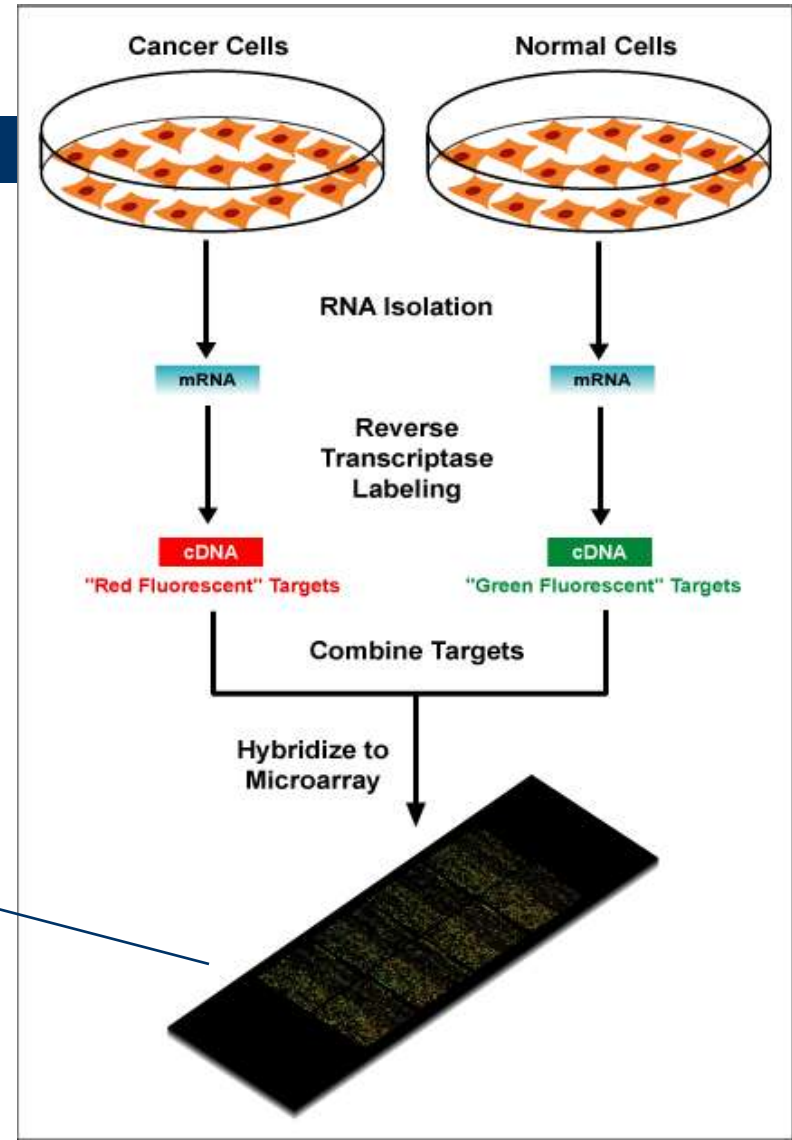
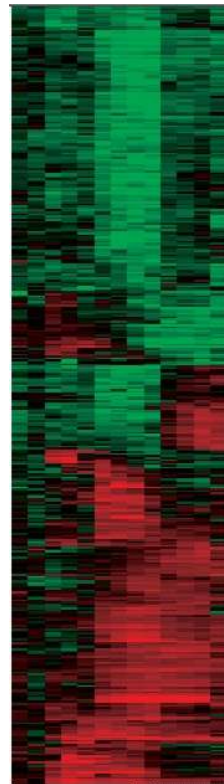
Microarray technology :: miles de genes bajo las mismas condiciones experimentales

Condiciones muestrales

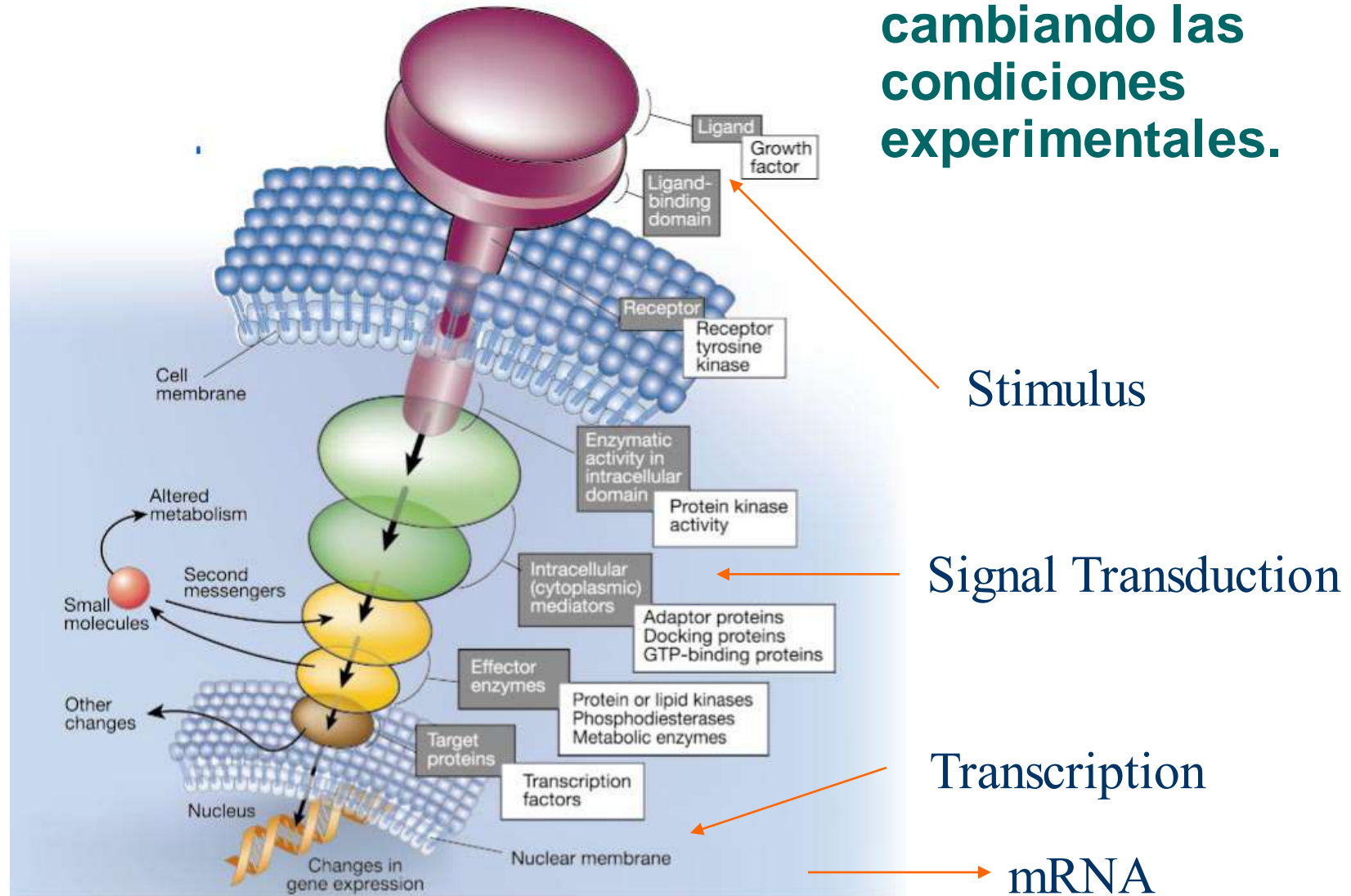
Genes



Condiciones muestrales: comparación de tejidos.



Aplicando drogas,
sustancias, o
cambiando las
condiciones
experimentales.



Downward, *Nature*, 411, 759, 2001

Análisis básicos de datos de Microarray

- Encontrar genes marcadores.
- Agrupar genes y condiciones muestrales.
 - clustering

Gene clustering

- Los genes co-regulados se expresarán igual bajo las mismas circunstancias.
 - Estudio de similitud entre las expresiones de genes.

Estudio de las distancias entre la expresión de dos genes

Minkowski distance

$$d(i, j) = \sqrt[q]{(|x_{i_1} - x_{j_1}|^q + |x_{i_2} - x_{j_2}|^q + \dots + |x_{i_p} - x_{j_p}|^q)}$$

If $q = 1$, d is **Manhattan distance** (semi-metric distance)

$$d(i, j) = |x_{i_1} - x_{j_1}| + |x_{i_2} - x_{j_2}| + \dots + |x_{i_p} - x_{j_p}|$$

If $q = 2$, d is **Euclidean distance** (metric distance)

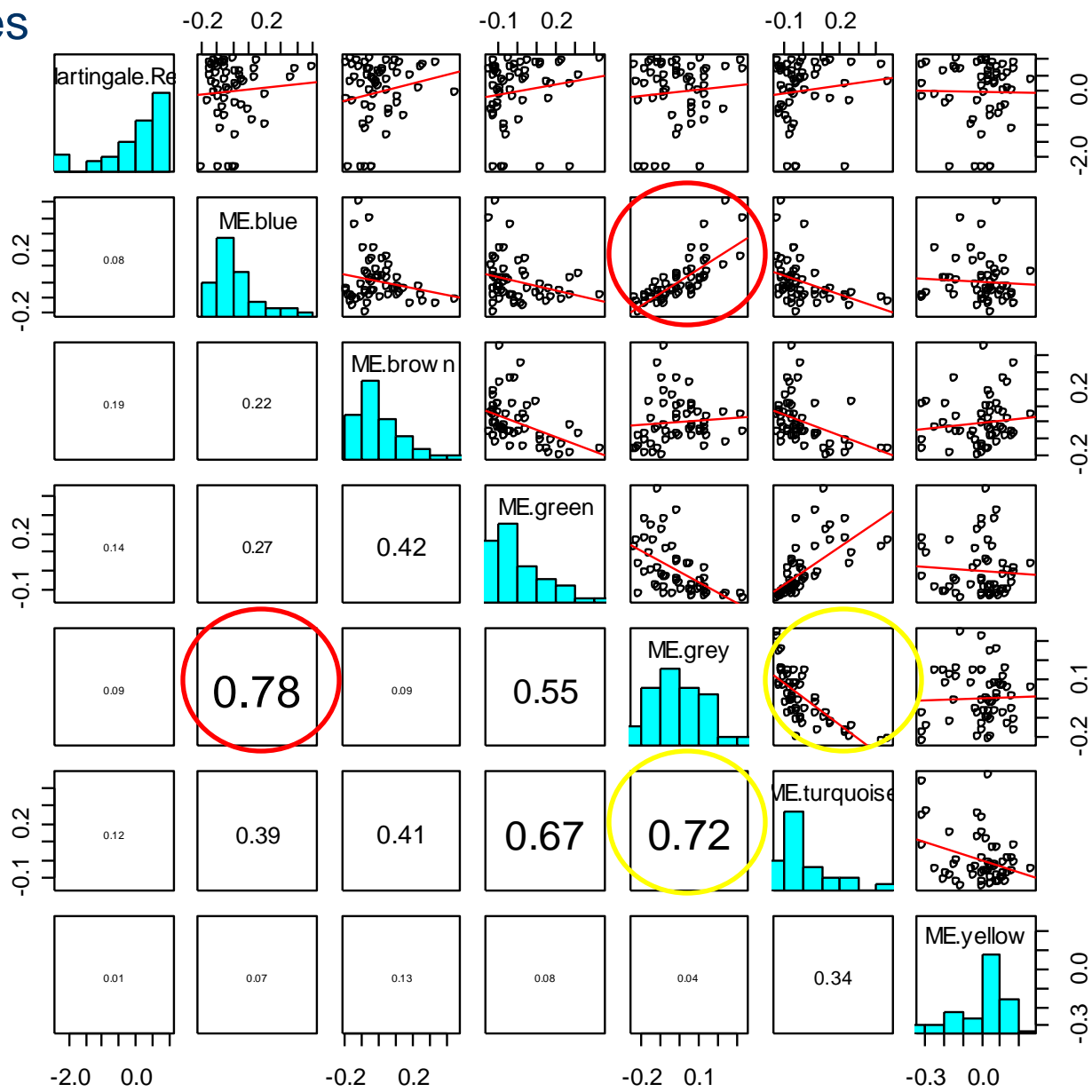
$$d(i, j) = \sqrt{(|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 + \dots + |x_{i_p} - x_{j_p}|^2)}$$

Pearson correlation coefficient (semi-metric distance)

$$d(i, j) = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2} \sqrt{\sum_{i=1}^n (x_{i2} - \bar{x}_2)^2}}$$

Genes
X
Genes

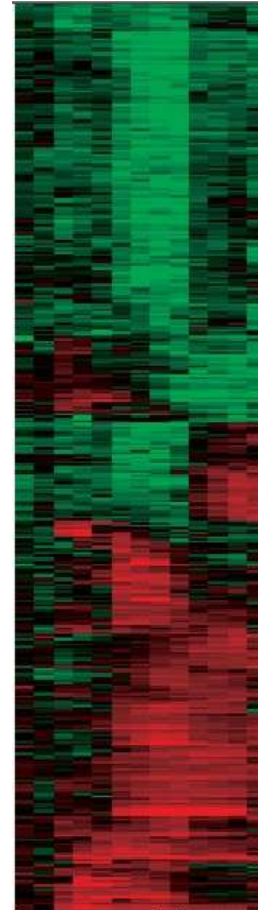
$$a_{ij} = |cor(x_i, x_j)|^\beta$$



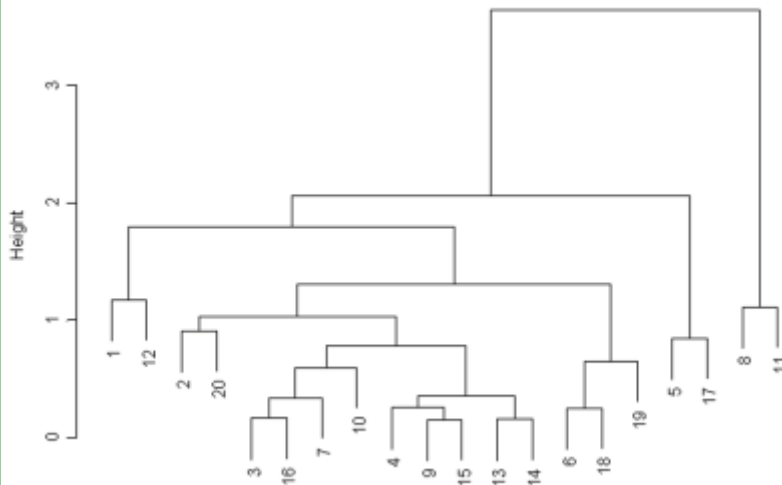
Agrupar los genes por similitud (clustering).

- Para agrupar los genes usaremos técnicas de clustering.

Microarray con los genes agrupados por similitud



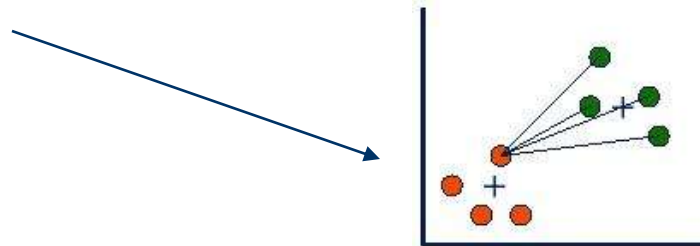
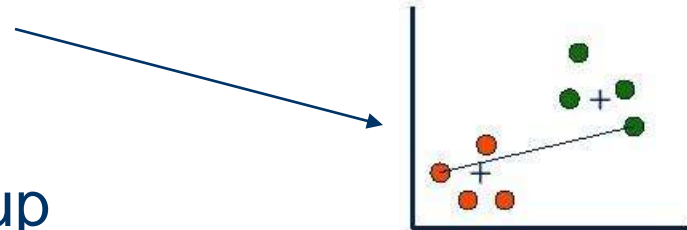
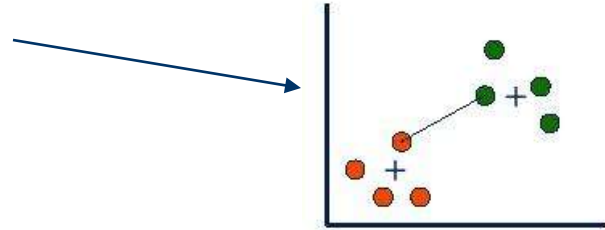
Hierarchical clustering



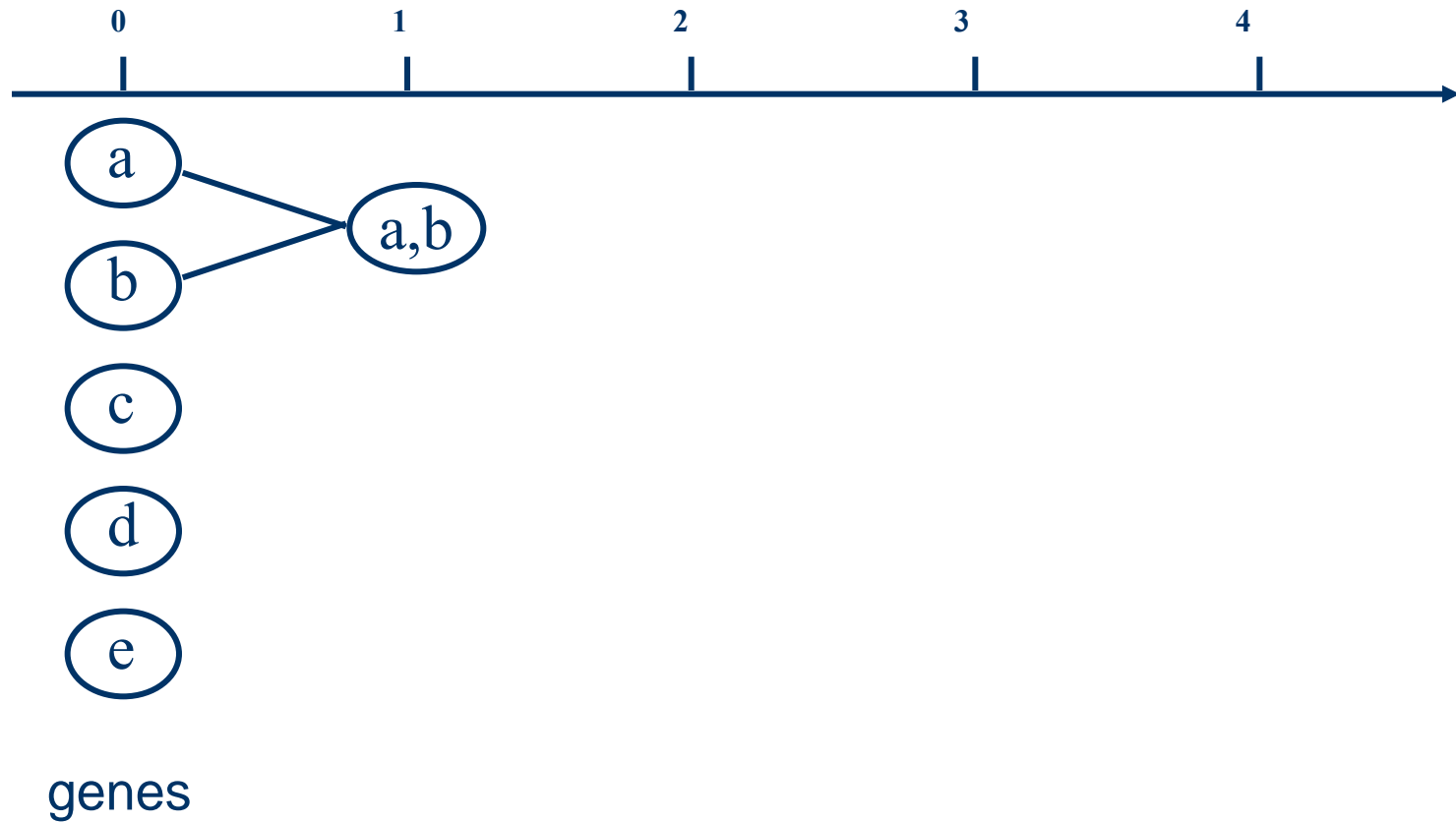
```
*** Agglomerative Hierarchical Clustering ***  
Call:  
agnes(x = menuModelFrame(data = spospread.v12.1000.3col, variables = "<ALL>",  
subset = 1:20, na.rm = T), diss = F, metric = "euclidean", stand = F,  
method = "average", save.x = T, save.diss = T)  
Merge:  
[,1] [,2]  
[1,] -9 -15  
[2,] -13 -14  
[3,] -3 -16  
[4,] -6 -18  
[5,] -4 1  
[6,] 3 -7  
[7,] 5 2  
[8,] 6 -10  
[9,] 4 -19  
[10,] 8 7  
[11,] -5 -17  
[12,] -2 -20  
[13,] 12 10  
[14,] -8 -11  
[15,] -1 -12  
[16,] 13 9  
[17,] 15 16  
[18,] 17 11  
[19,] 18 14  
Order of objects:  
[1] 1 12 2 20 3 16 7 10 4 9 15 13 14 6 18 19 5 17 8 11  
Height:  
[1] 1.1707192 1.7990210 0.9110762 1.0308996 0.1665215 0.3361807 0.5993238  
[8] 0.7855111 0.2544151 0.1539361 0.3562951 0.1572088 1.3038006 0.2524947  
[15] 0.6471538 2.0585785 0.8485635 3.6383437 1.1138825  
Agglomerative coefficient:  
[1] 0.8435272  
  
Available arguments:  
[1] "order" "height" "ac" "merge" "order.lab" "diss"  
[7] "data" "call"
```

Calcular la distancia entre Clusters

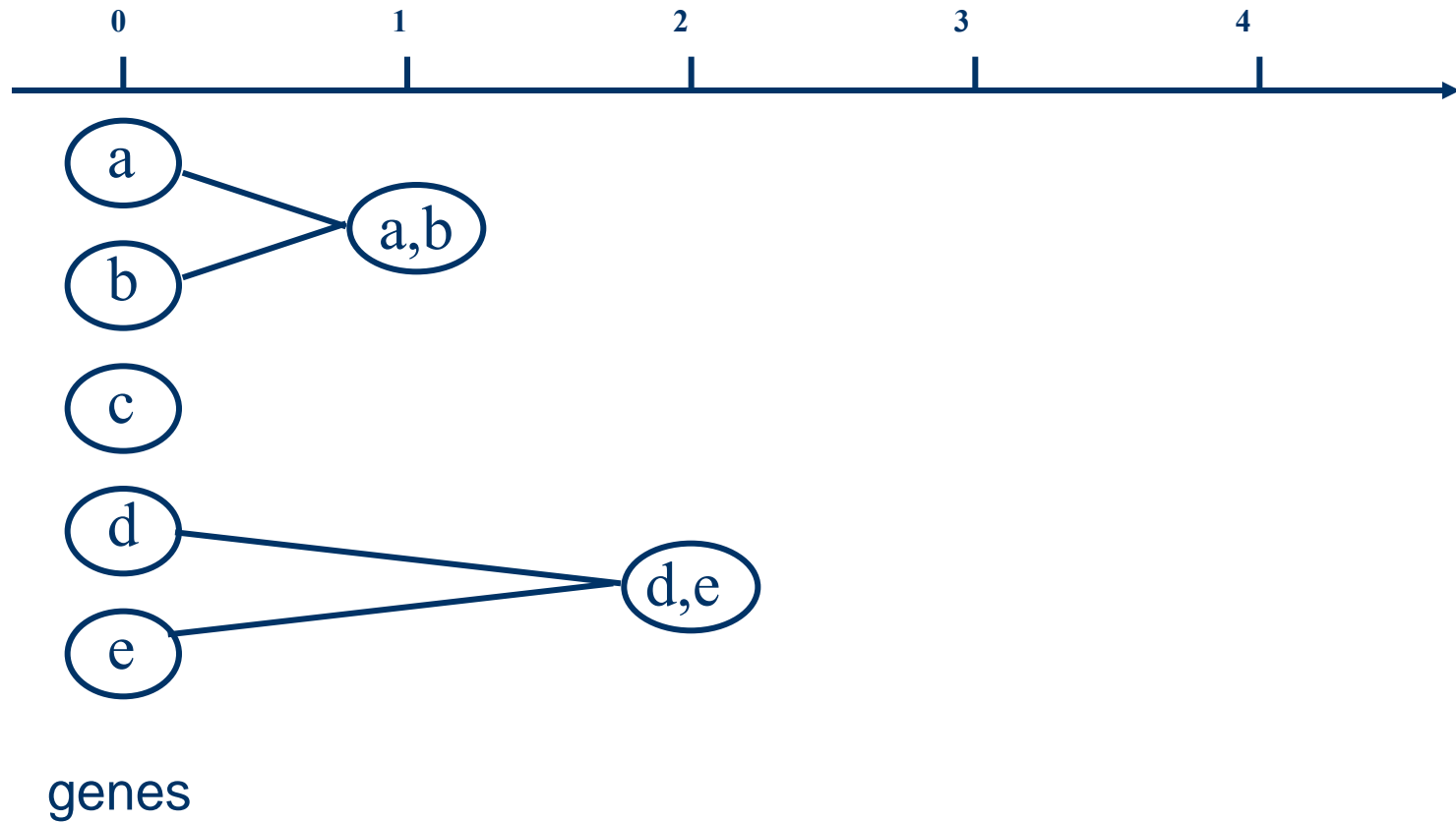
- Single-linkage clustering :: nearest-neighbour
- Complete-linkage :: furthest-neighbour
- Average-linkage :: unweighted pair-group method average (UPGMA)
- ...



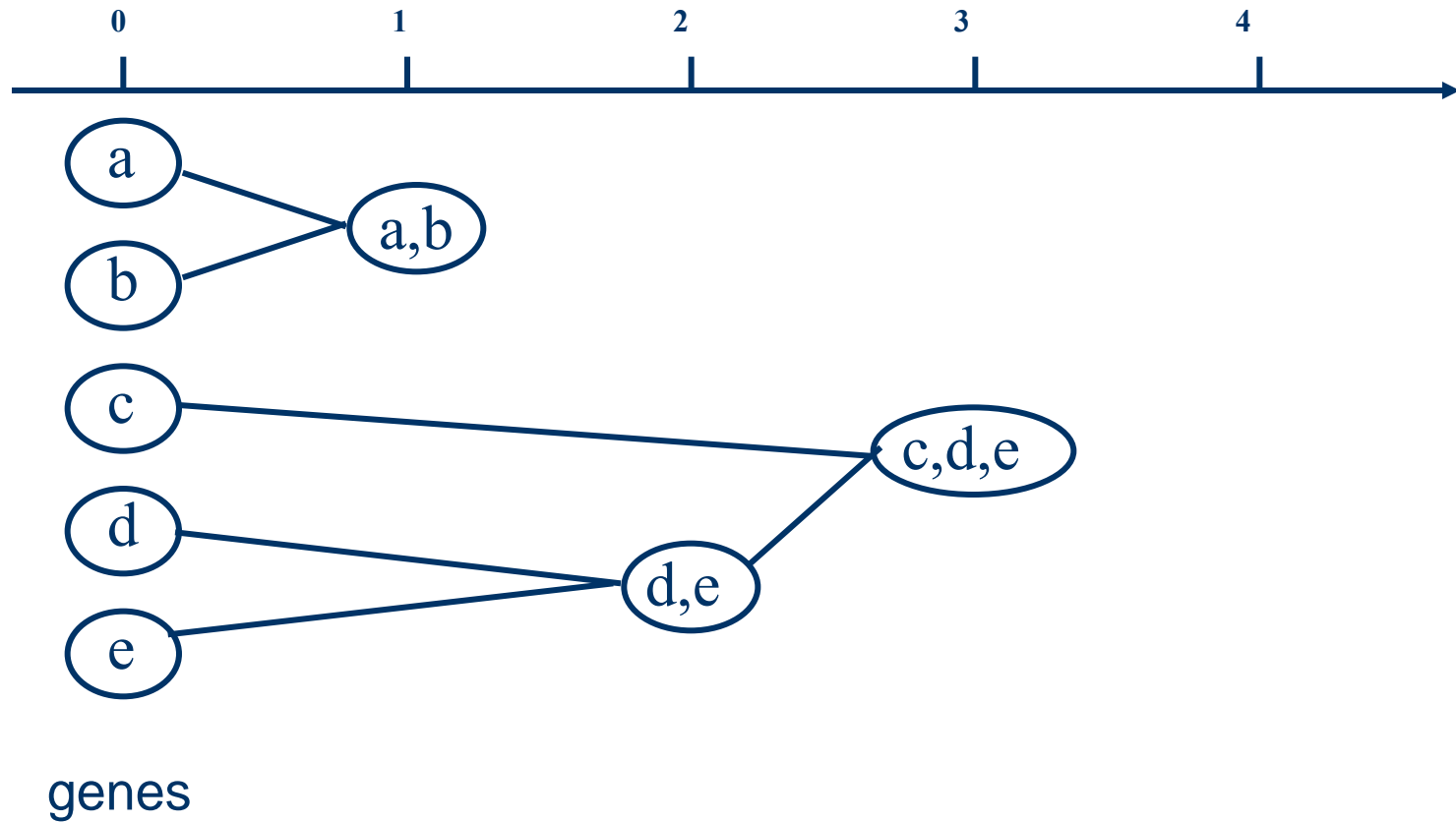
Aglomerative clustering.



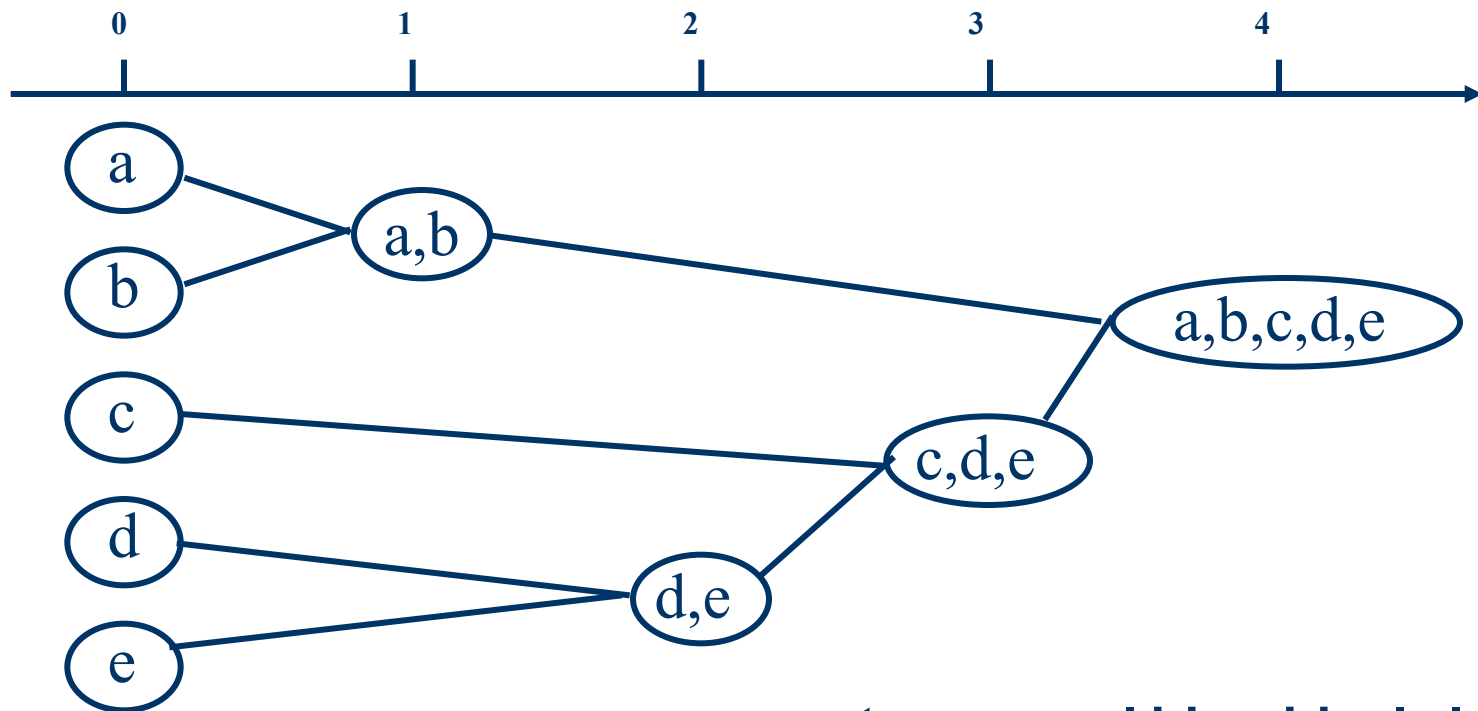
Aglomerative clustering.



Aglomerative clustering.



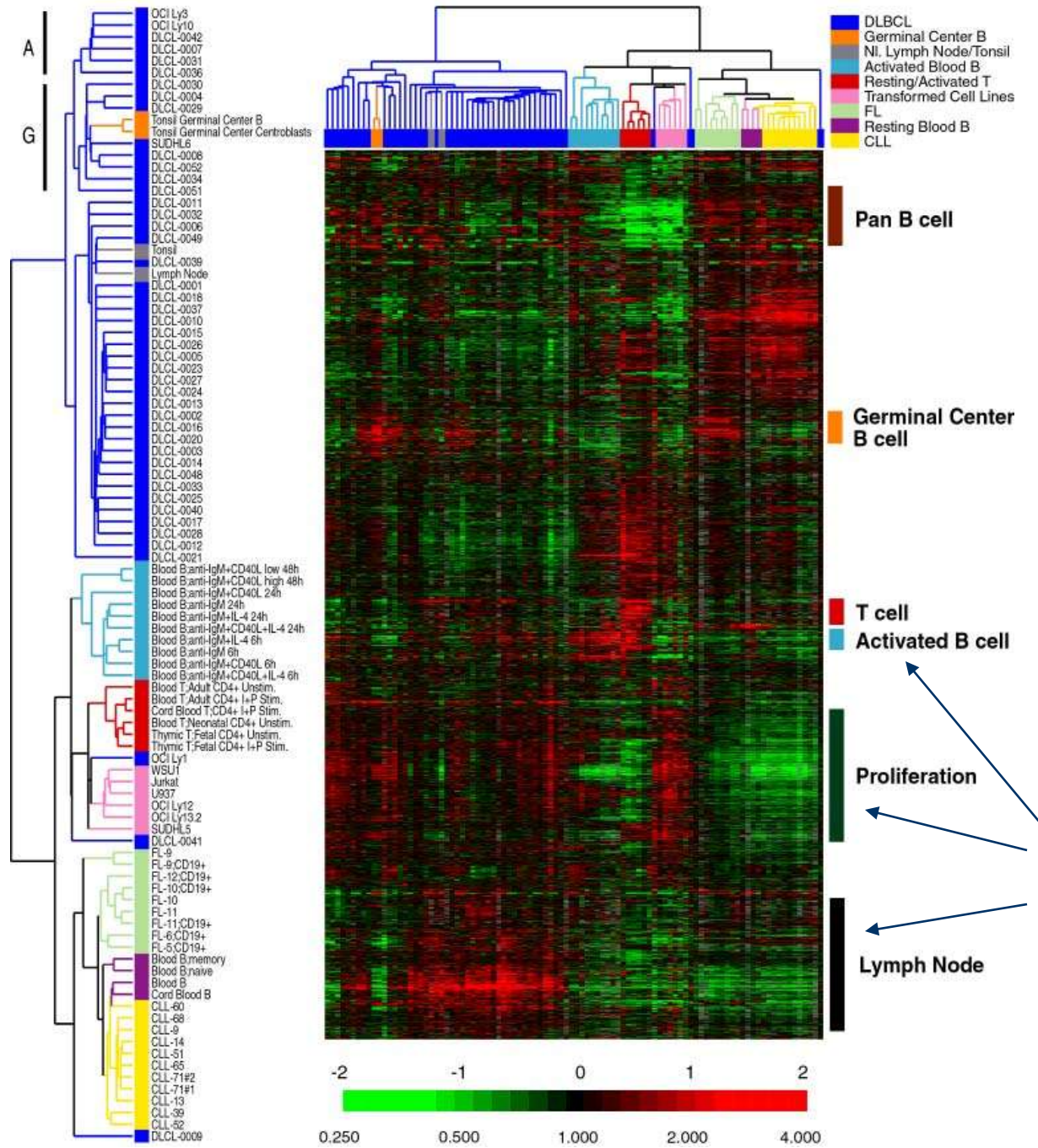
Aglomerative clustering.



...y tenemos el hierchical clustering

genes

Two-way clustering of genes (y-axis) and cell lines (x-axis) (Alizadeh et al., 2000)

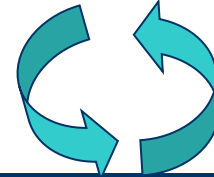


We can extract conclusions!

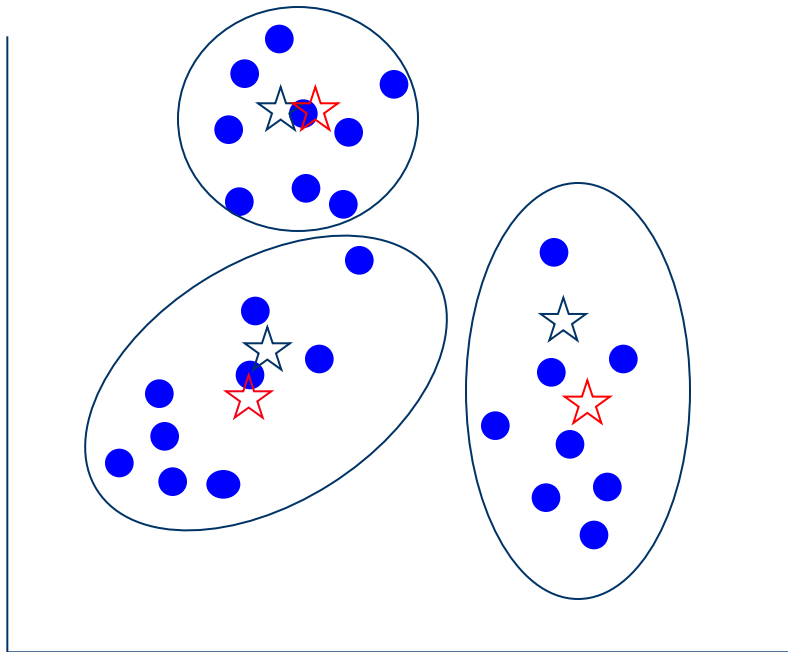
K-means clustering

Step 4 – Points re-assigned to nearest centroid

Step 5 – New centroids calculated



Iterates until
centroids don't
move



$K = 3$

Self-organizing maps (SOM)

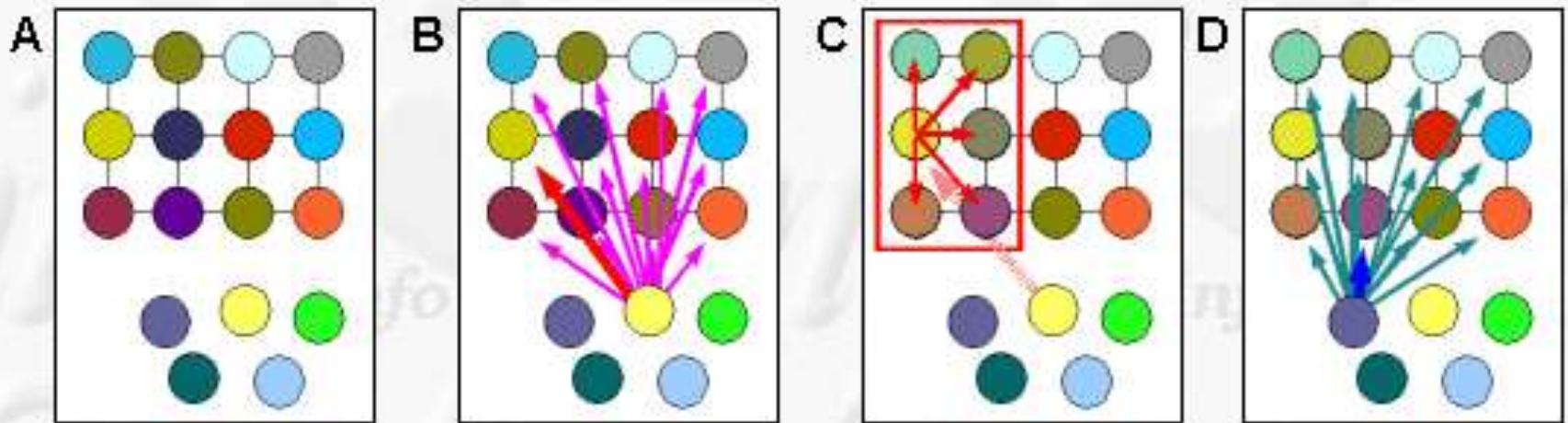
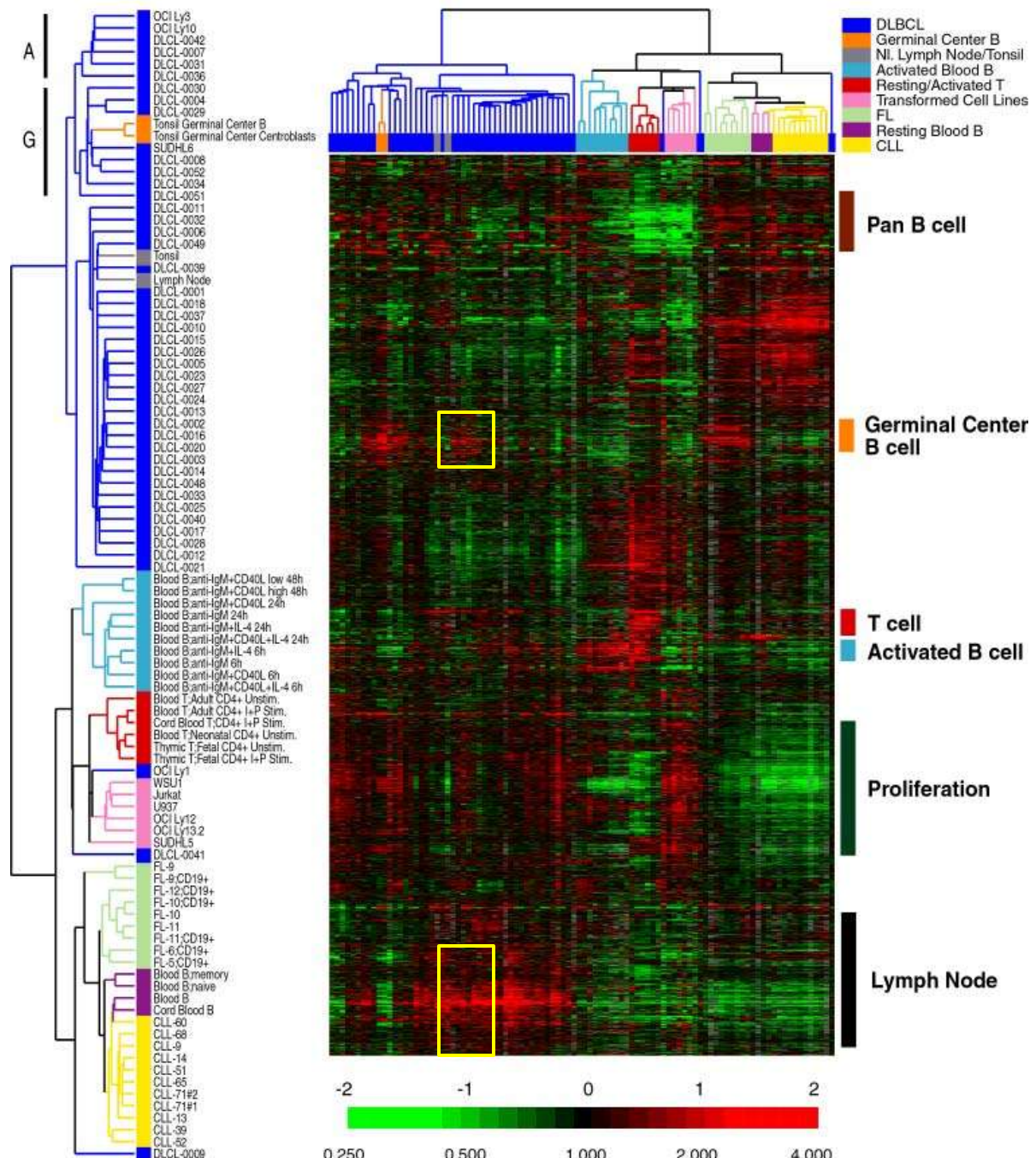


Fig. 1.6.1 Schematic representation of the working of SOM network
(A) The lattice has random values at the beginning. **(B)** The first element is compared with all the nodes of the lattice to get the closest node (red arrow). **(C)** The winning node and its neighbourhood are updated. **(D)** Next item is compared with all the nodes of the lattice as in (B). This process is repeated thousands of times for all the items. .

Bi-clustering



Principal components analysis (PCA)

An exploratory technique used to reduce the dimensionality of the data set to 2D or 3D

For a matrix of m genes \times n samples, create a new matrix of size $n \times n$

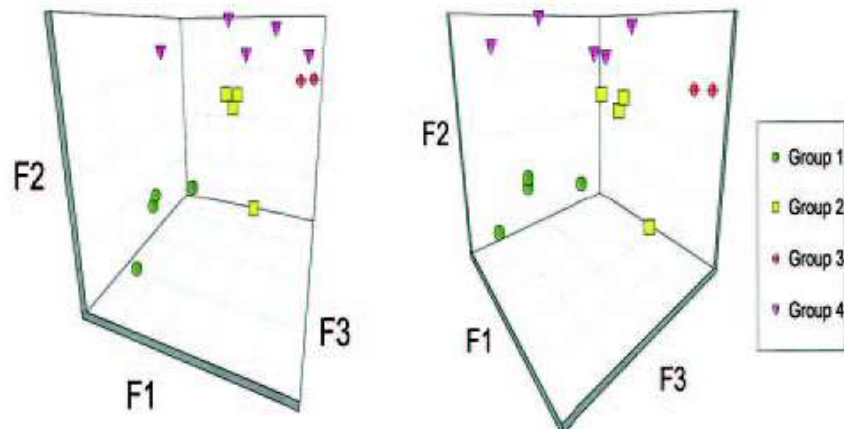
Thus, transform some large number of variables into a smaller number of uncorrelated variables called principal components (PCs).

PCA: objectives

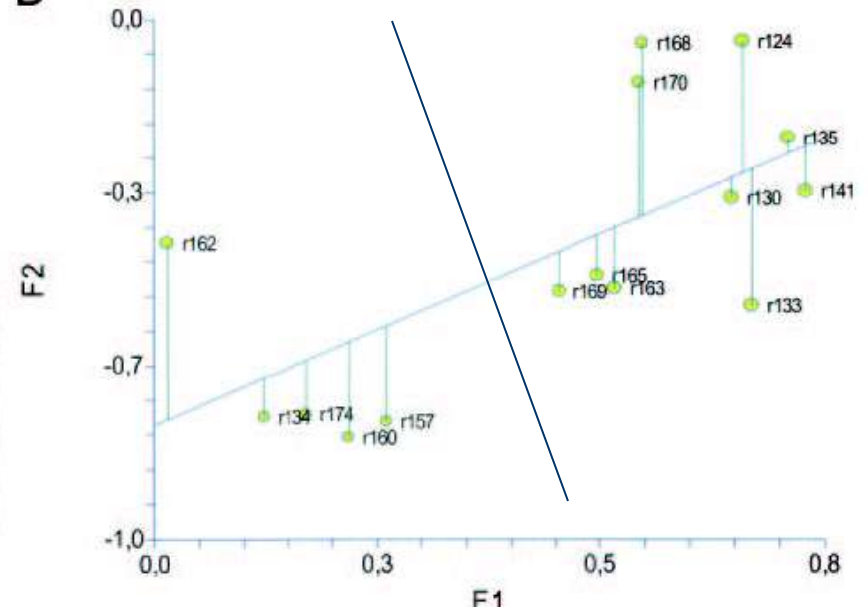
- **to reduce dimensionality**
- **to determine the linear combination of variables**
- **to choose the most useful variables (features)**
- **to visualize multidimensional data**
- **to identify groups of objects (e.g. genes/samples)**
- **to identify outliers**

PCA examples

C

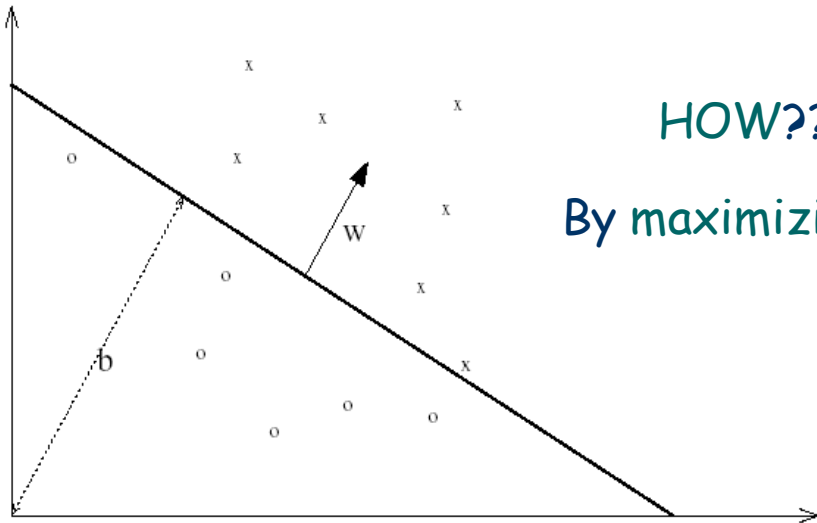


D



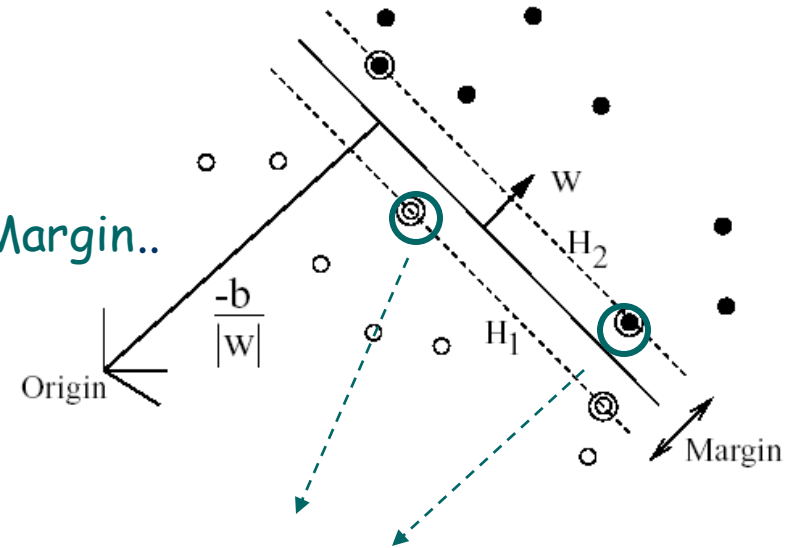
Support Vector Machines (SVM)

- Linear classifiers
- Attempt to avoid overfitting by finding the optimal hyperplane that separates the data



HOW???

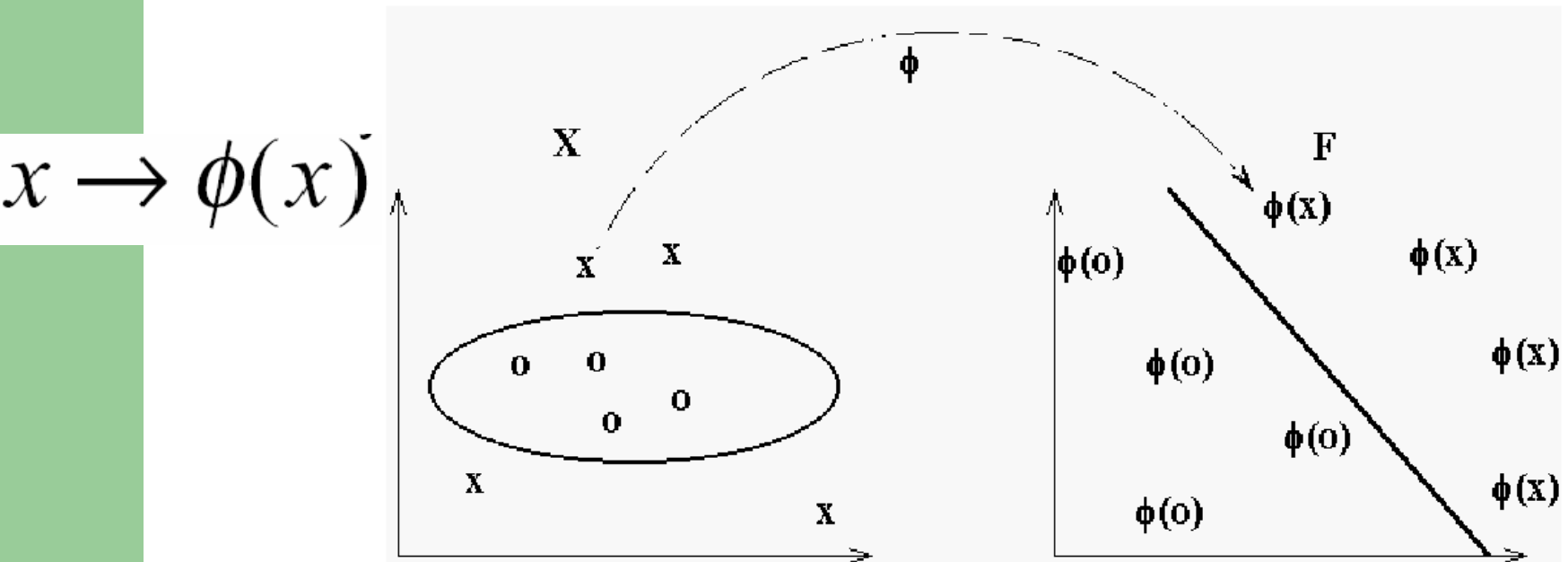
By maximizing the Margin..



Support Vectors

Support Vector Machines (SVM)

- And what about datasets that are not linearly separable??
Map the data into higher dimensional space and make linear classification there (theorem!!)



Support Vector Machines (SVM)

Tipo de SVM	Kernel	Comentarios
Polinómico	$(x^T y + 1)^p$	p especificado por el usuario
RBF	$\exp(-\frac{1}{2\sigma^2} \ x - x_i\ ^2)$	σ^2 especificado por el usuario y común a todos los kernel
Perceptrón de 2 capas	$\tanh(\beta_0 x^T x_i + \beta_1)$	Sólo ciertos valores de β son válidos

SVM: More experiments...

T.furey, N.Cristianini, N. Duffy, D. Bednarski, M. Schummer and D Haussler, "Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expressioin Data", Bioinformatics, 2000.

Kernel	DF	Feature	FP	FN	TP	TN
dot-product 0	25	25	5	4	10	12
dot-product 2	25	25	5	2	12	12
dot-product 5	25	25	4	2	12	13
dot-product 10	25	25	4	2	12	13
dot-product 0	50	50	4	2	12	13
dot-product 2	50	50	3	2	12	14
dot-product 5	50	50	3	2	12	14
dot-product 10	50	50	3	2	12	14
dot-product 0	100	100	4	3	11	13
dot-product 2	100	100	5	3	11	12
dot-product 5	100	100	5	3	11	12
dot-product 10	100	100	5	3	11	12
dot-product 0	97802	97802	17	0	14	0
dot-product 2	97802	97802	9	2	12	8
dot-product 5	97802	97802	7	3	11	10
dot-product 10	97802	97802	5	3	11	12

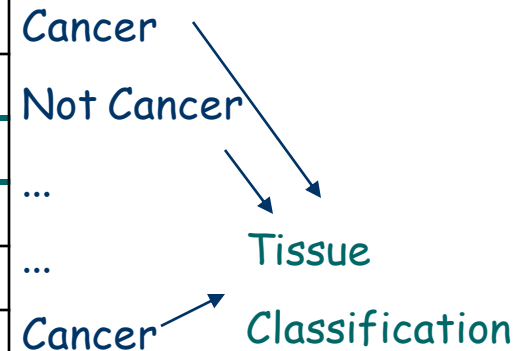
Gene expression data on tissue

97,802 DNA clones

31 tissue samples

Cancer ovarian
Normal ovarian
Normal non-ovarian

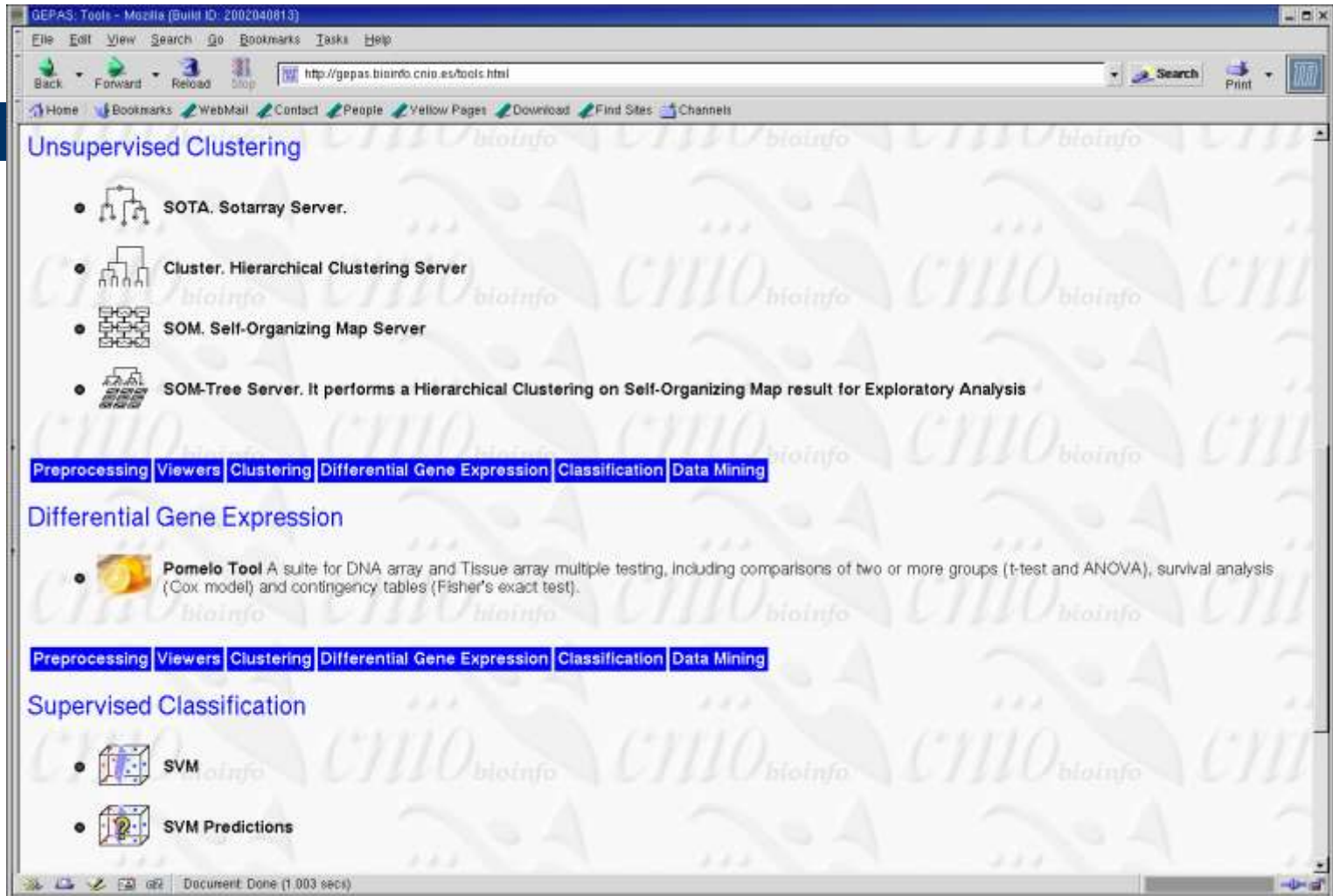
Genes				
Experiment	g-1	g-2	g-n
ex-1				
ex-2				
.....				
.....				
ex-m				










Herramientas para el análisis de microarrays

- Software específico
- Software estadístico
 - Excel
 - MATLAB
 - Octave
 - SAS
 - SPSS
 - S-PLUS
 - Statistica
 - R

Free Software for microarray analysis GEPAS



The screenshot shows a Mozilla browser window displaying the GEPAS Tools website. The address bar shows the URL <http://gepas.bioinfo.cnio.es/tools.html>. The page content is organized into several sections:

- Unsupervised Clustering**
 -  **SOTA. Sotarray Server.**
 -  **Cluster. Hierarchical Clustering Server**
 -  **SOM. Self-Organizing Map Server**
 -  **SOM-Tree Server. It performs a Hierarchical Clustering on Self-Organizing Map result for Exploratory Analysis**
- Navigation Menu:** Preprocessing | Viewers | Clustering | Differential Gene Expression | Classification | Data Mining
- Differential Gene Expression**
 -  **Pomelo Tool** A suite for DNA array and Tissue array multiple testing, including comparisons of two or more groups (t-test and ANOVA), survival analysis (Cox model) and contingency tables (Fisher's exact test).
- Navigation Menu:** Preprocessing | Viewers | Clustering | Differential Gene Expression | Classification | Data Mining
- Supervised Classification**
 -  **SVM**
 -  **SVM Predictions**

The browser's status bar at the bottom indicates "Document Done (1.003 secs)".

Free Software for microarray analysis TM4

The screenshot shows the TM4 Online web interface in a Mozilla browser window. The browser's address bar displays the URL <http://www.tigr.org/software/tm4/mevScreenshots.html>. The interface is organized into a sidebar on the left and a main content area on the right.

Sidebar (Left):

- Documentation
- Screenshots
- Utilities
- Glossary
- FAQ
- Group Members
- Contact Us

Latest Version:

MADAM	V2.7
Spotfinder	V2.2.3
MIDAS	V2.17
MeV	V2.2

TIGR
THE INSTITUTE FOR GENOMIC RESEARCH

Main Content Area (Right):

The main content area is titled "MeV" and displays four analysis tools, each with a hierarchical tree on the left and a visualization on the right:

- Hierarchical Tree:** A vertical bar chart showing a hierarchical structure of data points, with colors ranging from red to green.
- Expression graphs:** A grid of four line graphs showing expression levels over time or across conditions. The graphs are colored blue, yellow, green, and pink.
- Principal Components for Microarrays:** A 3D scatter plot showing data points in a 3D space, with axes and a central point.
- Terrain Map:** A 3D surface plot showing a terrain map with a green base and yellow/orange peaks.

At the bottom of the browser window, a status bar displays "Welcome to TM4!".

Geo – Una base de datos de microarrays.

All Databases PubMed Nucleotide Protein Genome

Search GEO Profiles for breast cancer Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Subgroup effect Send to

All: 607915

Items 1 - 20 of 607915

1: [GDS825 record](#) | [GPL875 2602](#) [Homo sapiens]
Annotation: [FBN1](#): fibrillin 1 FBN, MASS, MFS1, OCTD, SGS, WMS
Reporter: [X63556 1448051](#) (SPOT_ID)
Experiment: **Breast cancer** cell expression profiles (G4100A), gene expression array-based log ratio

2: [GDS825 record](#) | [GPL875 14787](#) [Homo sapiens]
Annotation: [TACSTD1](#): tumor-associated calcium signal transducer 1 CD326, CO17-1A, EGP, EGP40, Ep-CAM,
Reporter: [M32306 2816509](#) (SPOT_ID)
Experiment: **Breast cancer** cell expression profiles (G4100A), gene expression array-based log ratio

3: [GDS825 record](#) | [GPL875 11374](#) [Homo sapiens]
Annotation: [LARS2](#): leucyl-tRNA synthetase 2, mitochondrial KIAA0028, LEURS, MGC26121
Reporter: [D21851 1254922](#) (SPOT_ID)
Experiment: **Breast cancer** cell expression profiles (G4100A), gene expression array-based log ratio

4: [GDS825 record](#) | [GPL875 5316](#) [Homo sapiens]
Annotation: [S100P](#): S100 calcium binding protein P MIG9
Reporter: [X65614 2060823](#) (SPOT_ID)

About Entrez
The GEO site
GEO FAQ
Listing of Contents
Graph caption
Entrez
Help | FAQ

Geo – Buscando genes marcadores

Search GEO Profiles for GDS2250[ACCN] [Save Search](#)

[Limits](#) [Preview/Index](#) [History](#) [Clipboard](#) [Details](#)

Display Summary Show 20 Subgroup effect Send to

All: 54675

Items 1 - 20 of 54675

1: [GDS2250 record](#) | [GPL570 204320_at](#) [Homo sapiens]
Annotation: [COL11A1](#): collagen, type XI, alpha 1 [CO11A1](#), [COLL6](#), [STL2](#)
Reporter: [NM_001854](#)
Experiment: Basal-like breast cancer tumors, gene expression array-based transformed count

2: [GDS2250 record](#) | [GPL570 37892_at](#) [Homo sapiens]
Annotation: [COL11A1](#): collagen, type XI, alpha 1 [CO11A1](#), [COLL6](#), [STL2](#)
Reporter: [NM_001854](#) [NM_080629](#) [NM_080630](#)
Experiment: Basal-like breast cancer tumors, gene expression array-based transformed count

3: [GDS2250 record](#) | [GPL570 204822_at](#) [Homo sapiens]
Annotation: [TTK](#): TTK protein kinase ESK, [FLJ38280](#), [MPS1L1](#), [PYT](#)
Reporter: [NM_003318](#)
Experiment: Basal-like breast cancer tumors, gene expression array-based transformed count

Geo – Buscando genes marcadores

Genes
marcadores

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Subgroup effect Send to

All: 54675

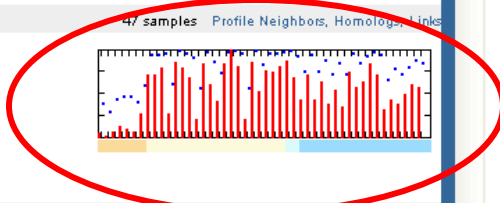
Items 1 - 20 of 54675 Page 1 of 2734 Next

1: GDS2250 record | GPL570 204320_at [Homo sapiens] 47 samples Profile Neighbors, Homologs, Links

Annotation: COL11A1: collagen, type XI, alpha 1 COL11A1, COLL6, STL2

Reporter: NM_001854

Experiment: Basal-like breast cancer tumors, gene expression array-based transformed count

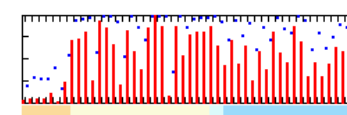


2: GDS2250 record | GPL570 37892_at [Homo sapiens] 47 samples Profile Neighbors, Homologs, Links

Annotation: COL11A1: collagen, type XI, alpha 1 COL11A1, COLL6, STL2

Reporter: NM_001854 NM_080629 NM_080630

Experiment: Basal-like breast cancer tumors, gene expression array-based transformed count

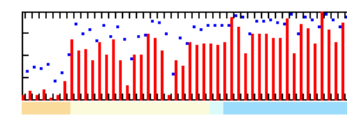


3: GDS2250 record | GPL570 204822_at [Homo sapiens] 47 samples Homologs, Links

Annotation: TTK: TTK protein kinase ESK, FLJ38280, MPS1L1, PYT

Reporter: NM_003318

Experiment: Basal-like breast cancer tumors, gene expression array-based transformed count

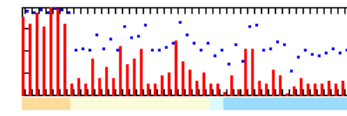


4: GDS2250 record | GPL570 205392_s_at [Homo sapiens] 47 samples Links

Annotation: CCL14: chemokine (C-C motif) ligand 14 CC-1, CC-3, CKb1, HCC-1, HCC-3, MCIF, NCC-2, NCC2, SCYA14, SCYL2, SY14


Reporter: NM_004166

Experiment: Basal-like breast cancer tumors, gene expression array-based transformed count



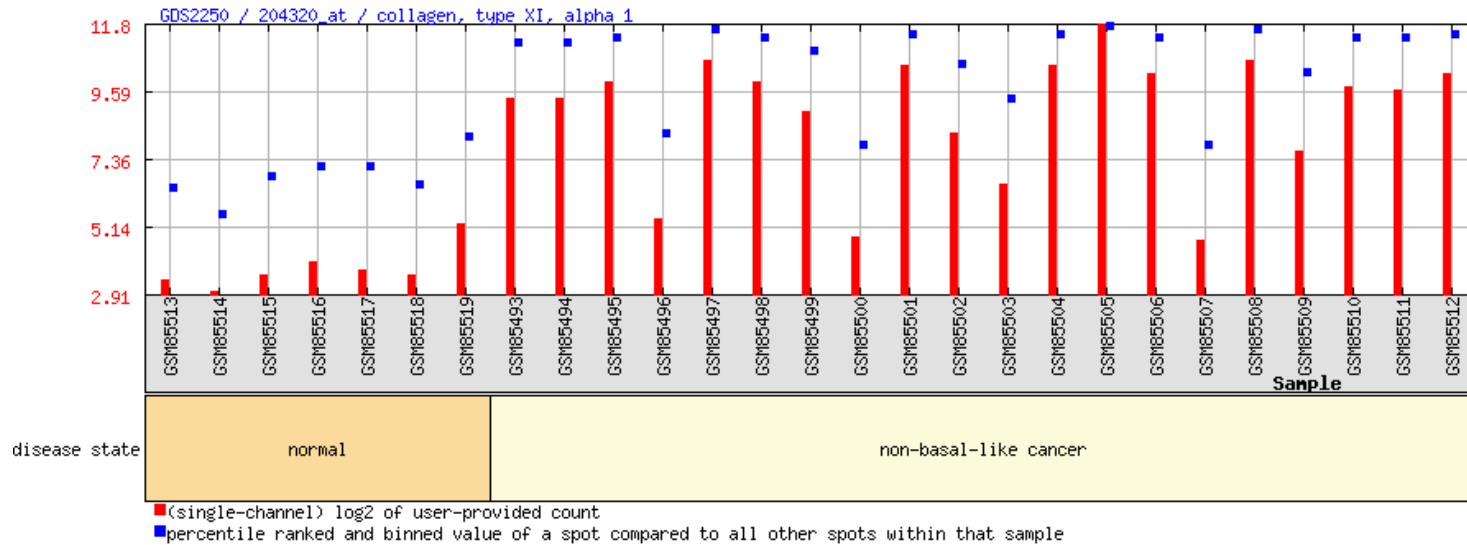
5: GDS2250 record | GPL570 204719_at [Homo sapiens] 47 samples Homologs, Links

Annotation: ABCA8: ATP-binding cassette, sub-family A (ABC1) member 8 KIAA0877

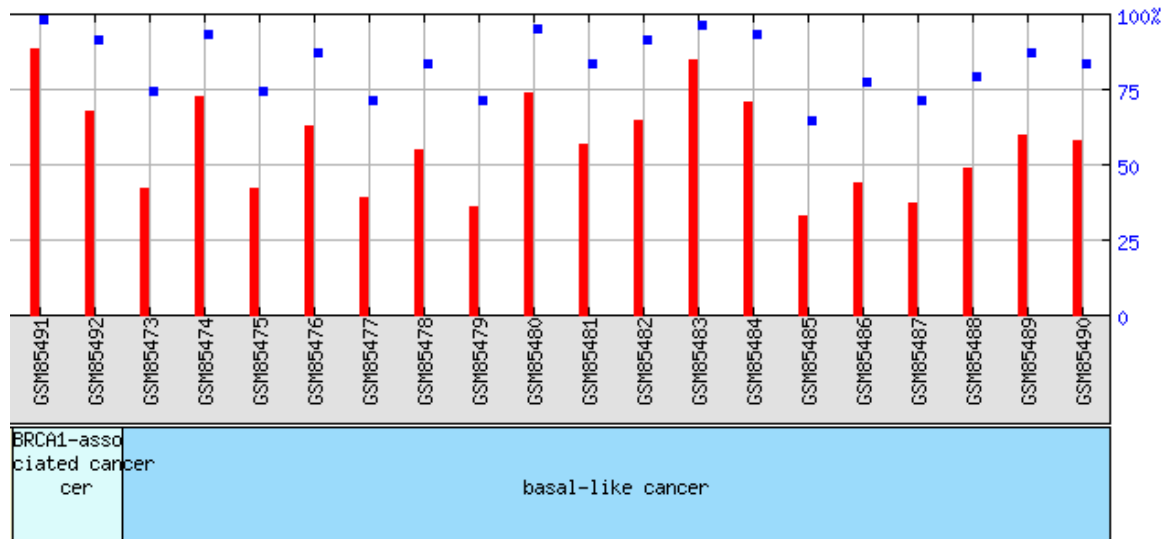


GEO - Expresión de los genes marcadores

- Example: Breast cancer



- Left y-axis is (supposed to be) log two based (must check to verify) expression level.
- Right y-axis is the percentile of this expression level in the entire chip.
- All the chips are normalized.



Geo – Buscando de qué patologías es marcador un gen

The screenshot shows the Entrez GEO website interface. The search bar contains the query "for AFAP AND Breast Cancer". The search results list three items, each with a circled "AFAP" in the annotation field. The first two items are associated with the experiment "Breast cancer cell expression profiles (G4100A), dual channel nucleotide log ratio". The third item is associated with the experiment "DNA copy-number changes (I), dual channel nucleotide log ratio".

Entrez GEO - Gene Expression Omnibus - Microsoft Internet Explorer provided by America Online

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

NCBI Entrez GEO Profiles

My NCBI Welcome rdw2k [Sign Out]

All Databases PubMed Nucleotide Protein Genome Structure PMC Journals Books

Search GEO Profiles for AFAP AND Breast Cancer Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display: Summary Show 20 Subgroup effect Send to

All: 17 Human: 17

Items 1 - 17 of 17 One page.

1: GDS825 record | GPL875 6892 [Homo sapiens] 4 samples - Profile Neighbors, Sequence Neighbors, Links

Annotation: AFAP: actin filament associated protein (AFAP-110)

Reporter: B33248

Experiment: Breast cancer cell expression profiles (G4100A), dual channel nucleotide log ratio

2: GDS825 record | GPL875 11325 [Homo sapiens] 4 samples - Sequence Neighbors, Links

Annotation: AFAP: actin filament associated protein (AFAP-110)

Reporter: AF188700

Experiment: Breast cancer cell expression profiles (G4100A), dual channel nucleotide log ratio

3: GDS82 record | GPL179 3479 [Homo sapiens] 4 samples - Profile Neighbors, Sequence Neighbors, Links

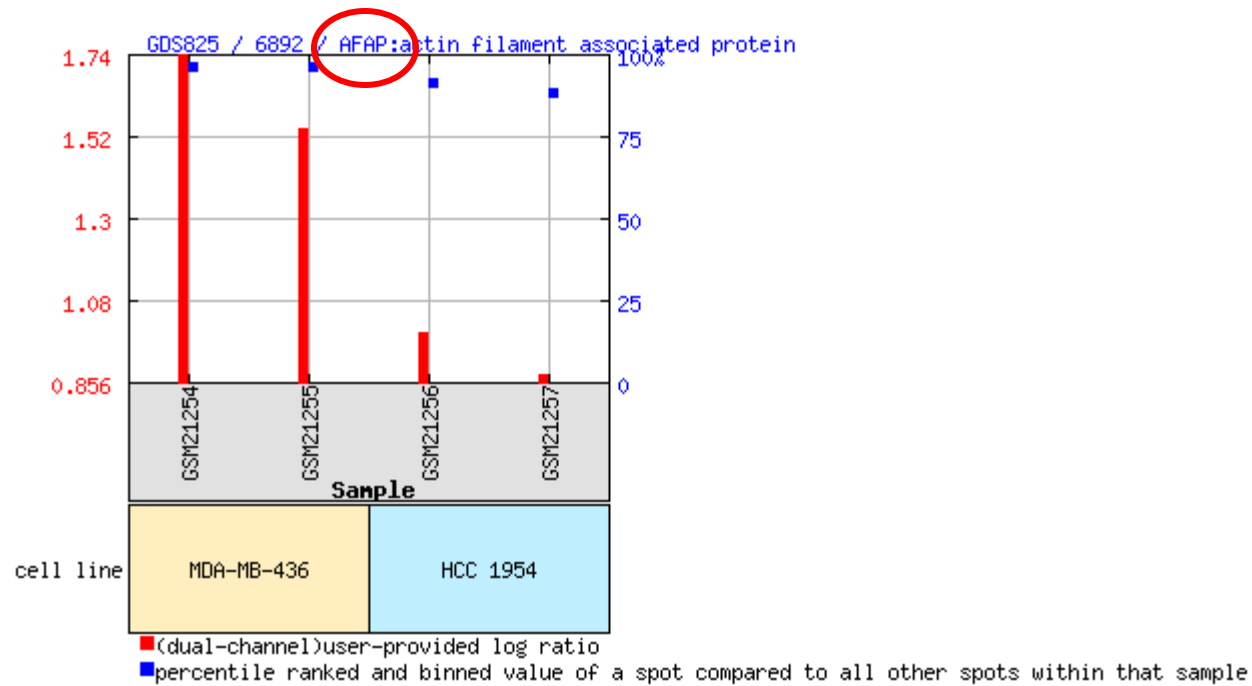
Annotation: AFAP: Hypothetical protein LOC254848

Reporter: R69355 IMAGE:142122 (clone)

Experiment: DNA copy-number changes (I), dual channel nucleotide log ratio

Start Becca Becca Weeld Untitled Micros AFAP Entrez... 10:25 PM

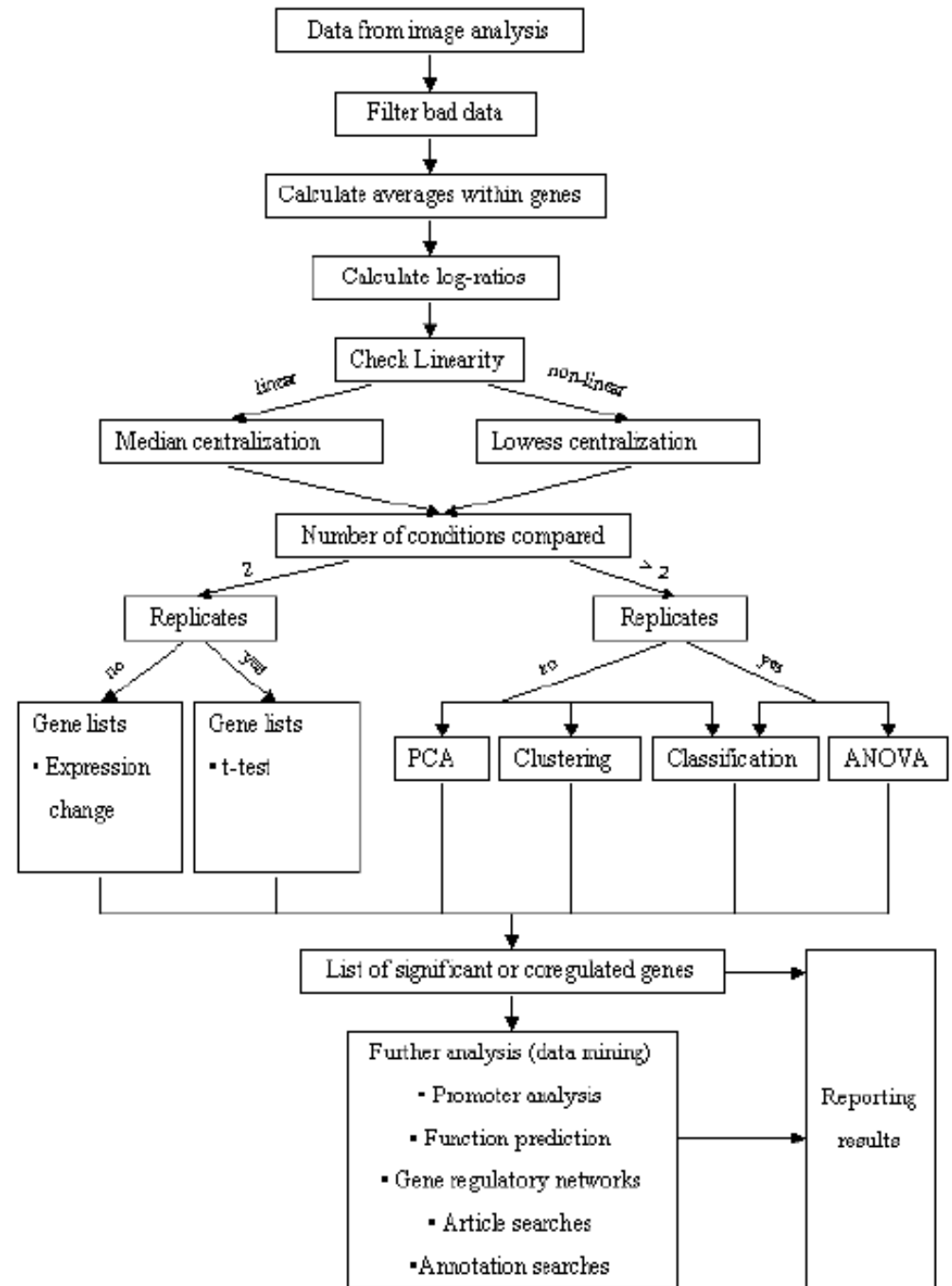
Geo – Buscando de que patologías/procesos es marcador un gen



Microarray Metaanálisis

Gene markers and Global analysis

Metaanálisis: Generación de conocimiento



Herramientas de Microarray meta-analysis y data minning (IBB-UAB).

- Extraer la máxima información útil al gran numero de genes y condiciones muestrales.
- A partir de microarrays con largo número de condiciones muestrales, modelar el comportamiento holístico de la célula/tejido analizado.
- Web server for on-line microarray analysis:
<http://revolutionresearch.uab.es>

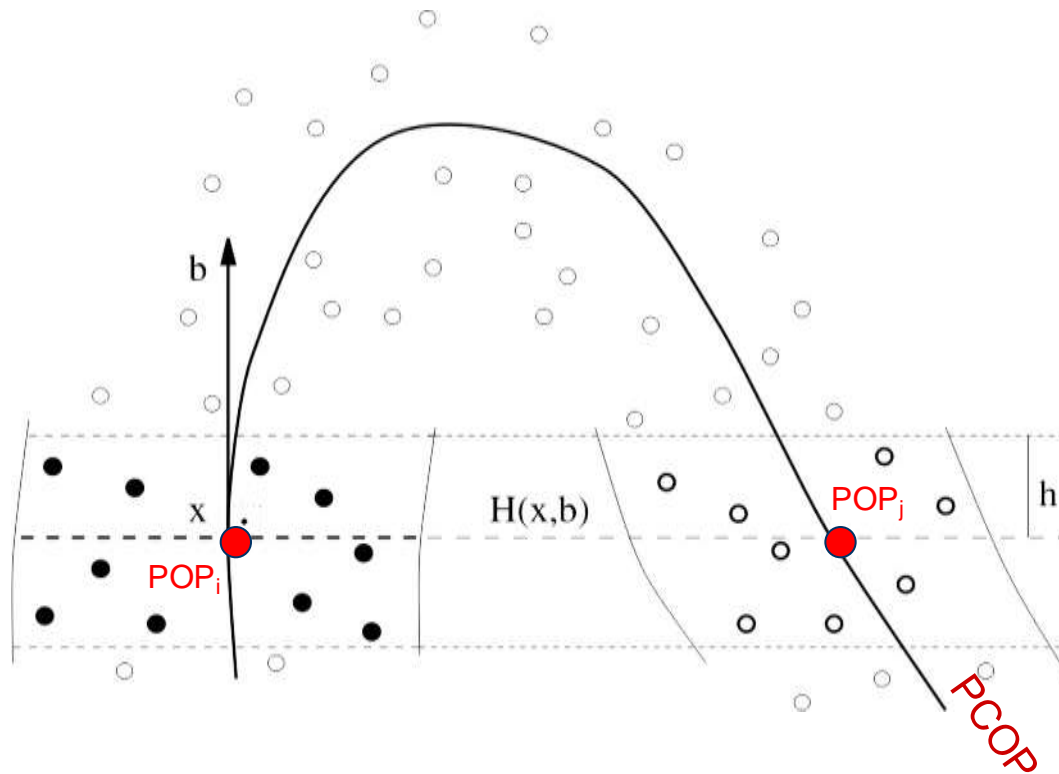
Estrategias para obtener el comportamiento holístico de la célula.

- Estudio de las relaciones LINEALES entre las expresiones de genes.
- Estudio de las relaciones NO LINEALES entre las expresiones de genes.
- Estudio de las relaciones NO CONTÍNUAS entre las expresiones de genes.
- Facilitar la navegación a través de la información generada.
 - Ampliar el análisis progresivamente incluyendo más genes y agrupando las muestras en clases de muestras.

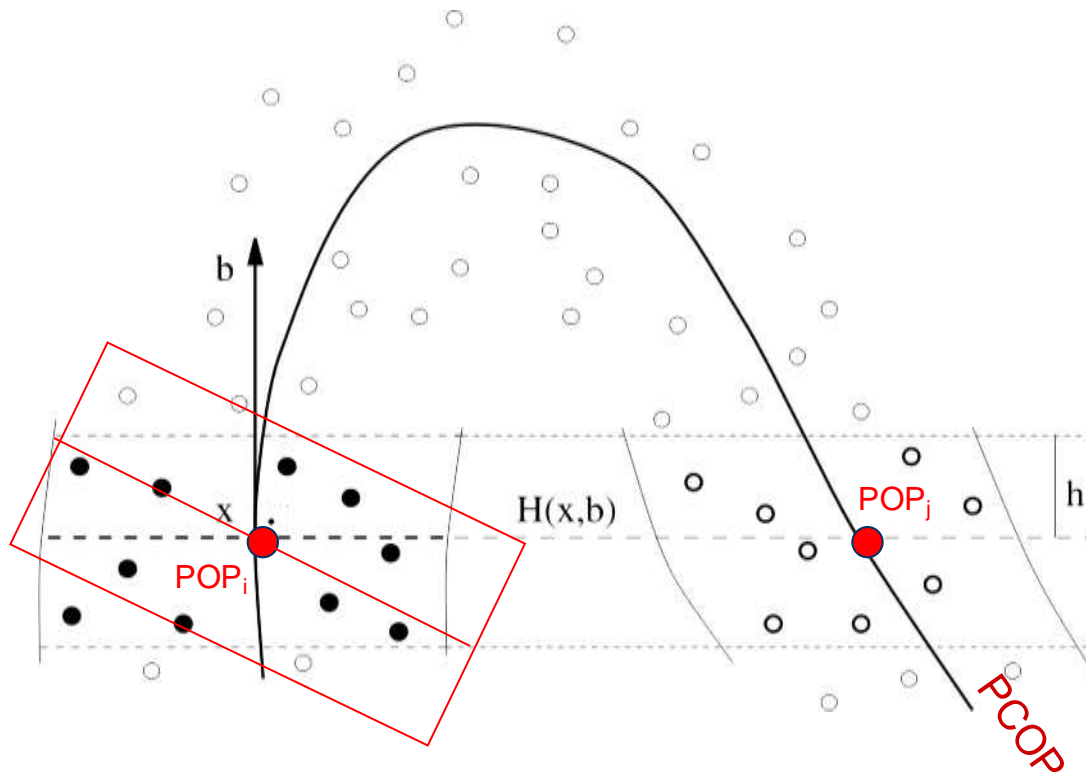
Análisis no lineal de las relaciones entre la expresión de los genes

- Hay genes que están relacionados no linealmente.
- Usaremos las Principal Curve of Oriented Points.

The PCOP is a very suitable analysis for recognising non-linear patterns among independent variables.



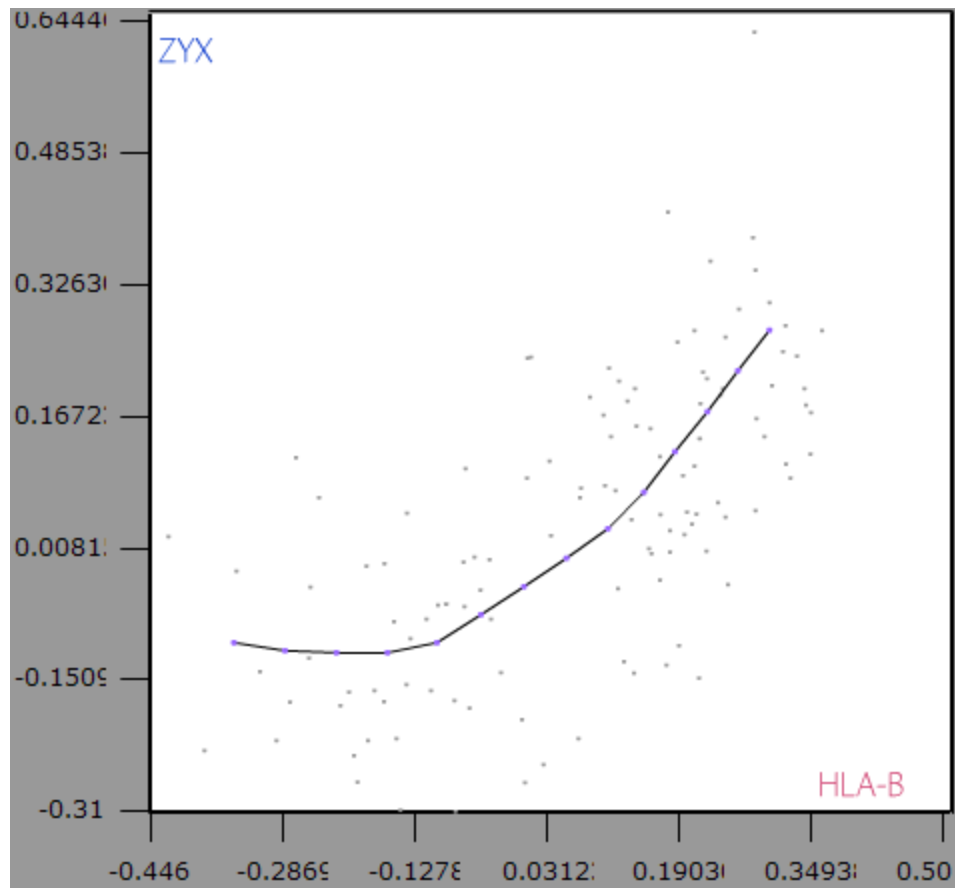
The PCOP is a very suitable analysis for recognising non-linear patterns among independent variables.



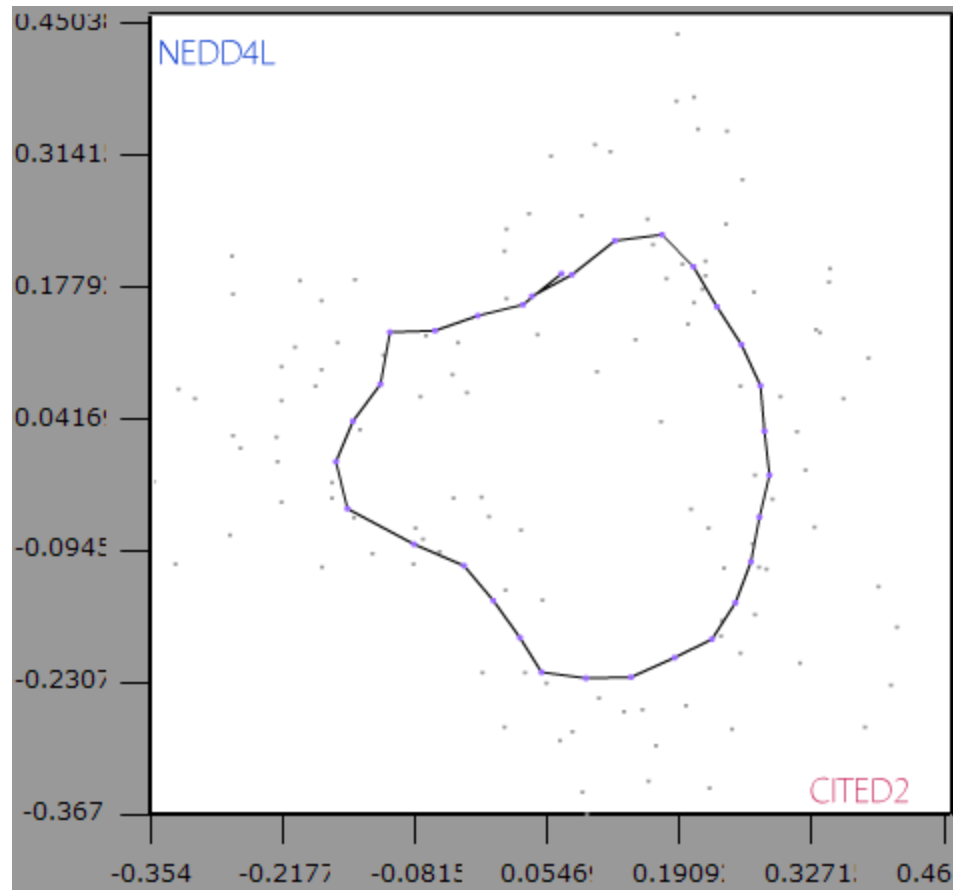
Gene-expression relationships detectable by PCOP analysis.

- Positively co-expressed genes.
- Negatively co-expressed genes.
- Genes mutually excluding in their expression.
 - genes which are dependent in their expression only because one of the genes must be over-expressed or under-expressed to activate the expression fluctuations of the other gene.
- Non-linear correlated genes.

Genes de exclusión mutua detectados



Genes con dependencia no lineal detectados

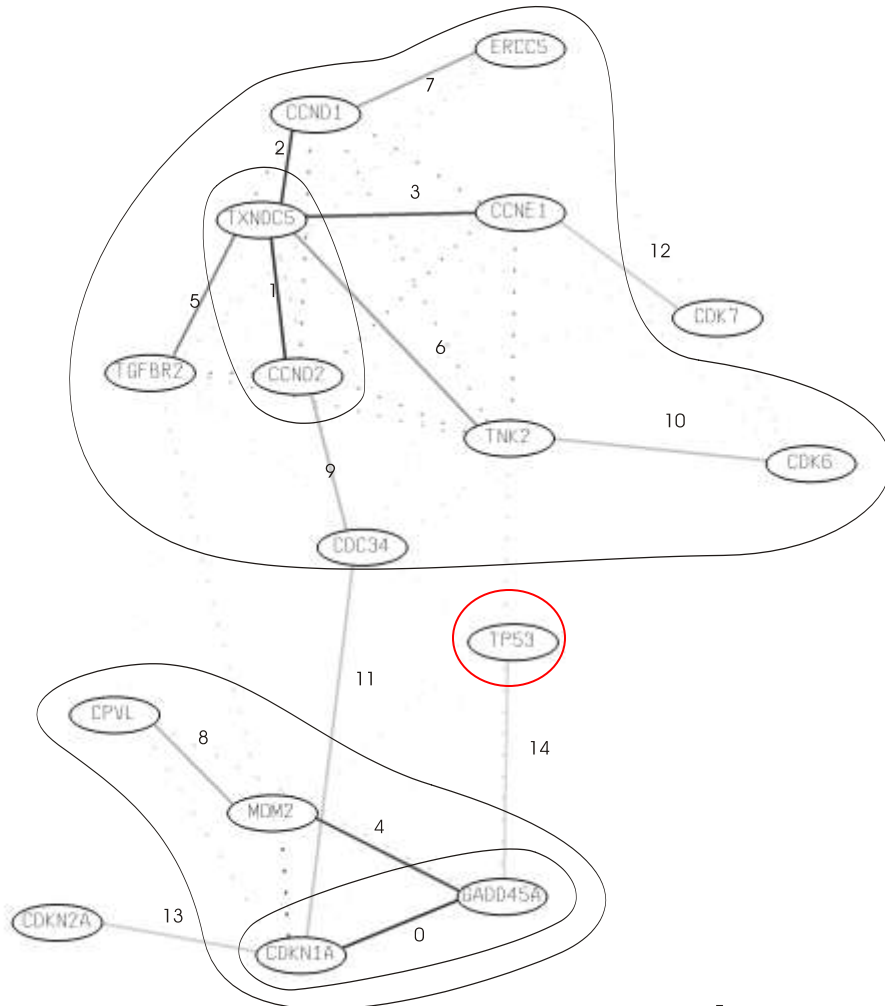


One of the main keys of the PCOP is the calculus of the correlation factor.

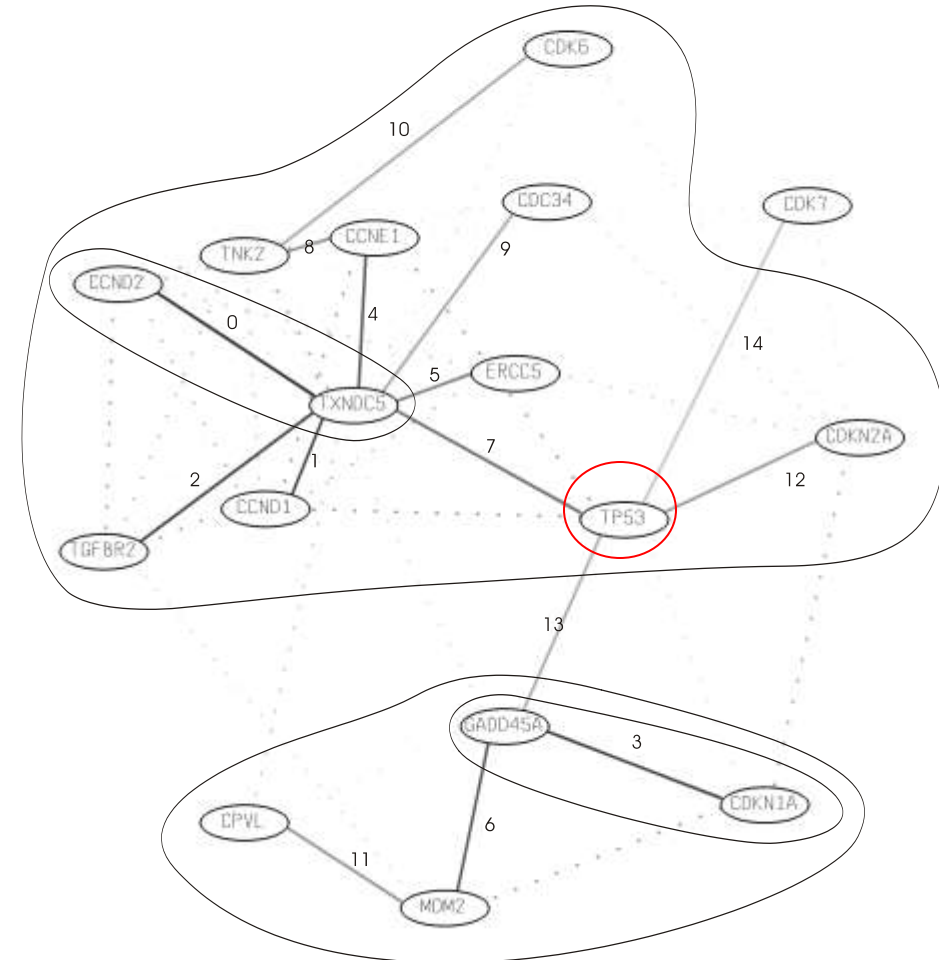
- **The Variance explained by the curve:** Permits one to know if the Principal Curve is able to follow the sample-cloud tendency. Goes up when the sample cloud has a regular behaviour being well identified by the Principal Curve.
- **Residual Variance:** The degree of dispersion of the samples around the Principal Curve.
- **The Generalized Total Variance:** The sum of these two dispersion parameters.
- **f factor:** RV divided by the GTV

Correlations graph and minimum-spanning-tree among the same microarray gene-expressions using R^2 correlation or the f value provided by the PCOP calculus.

Pearson R^2

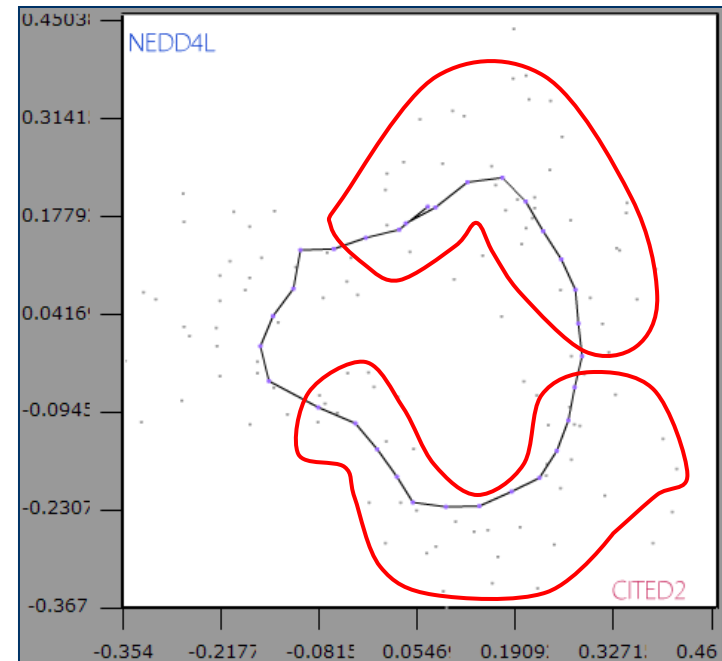
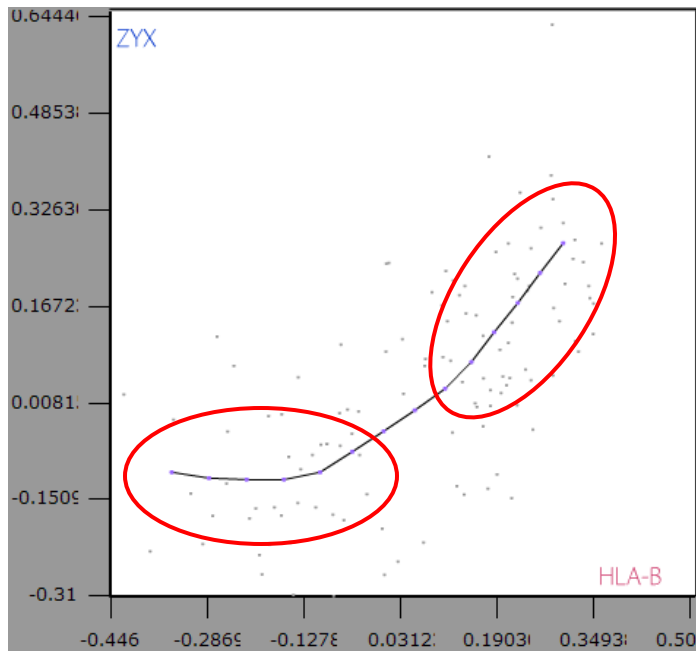


f value (PCOP)



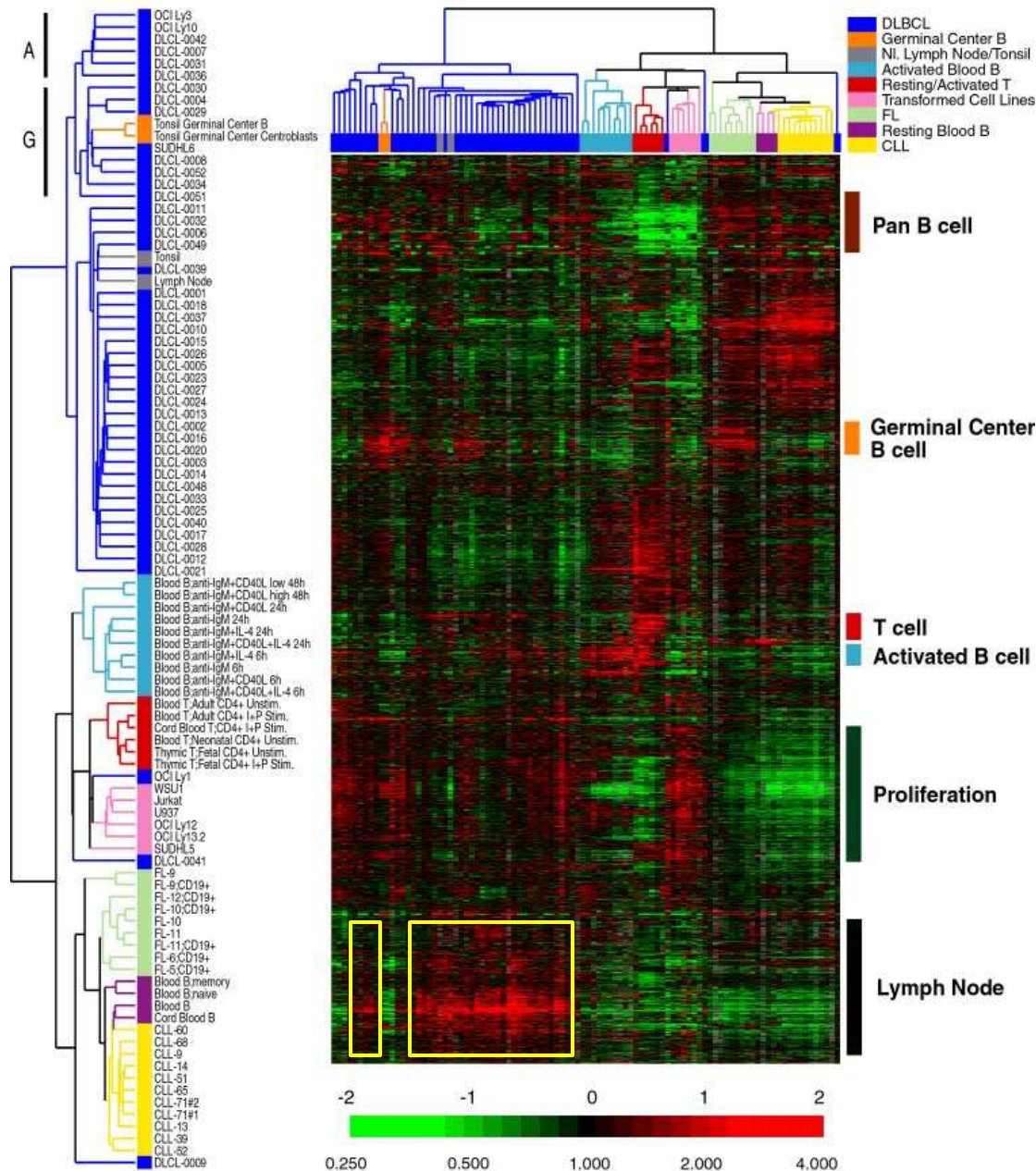
Obtención automática de los cambios fenotípicos

- Y de los genes implicados en el canvio.
 - Clustering local a los genes global para las condiciones muestrales.

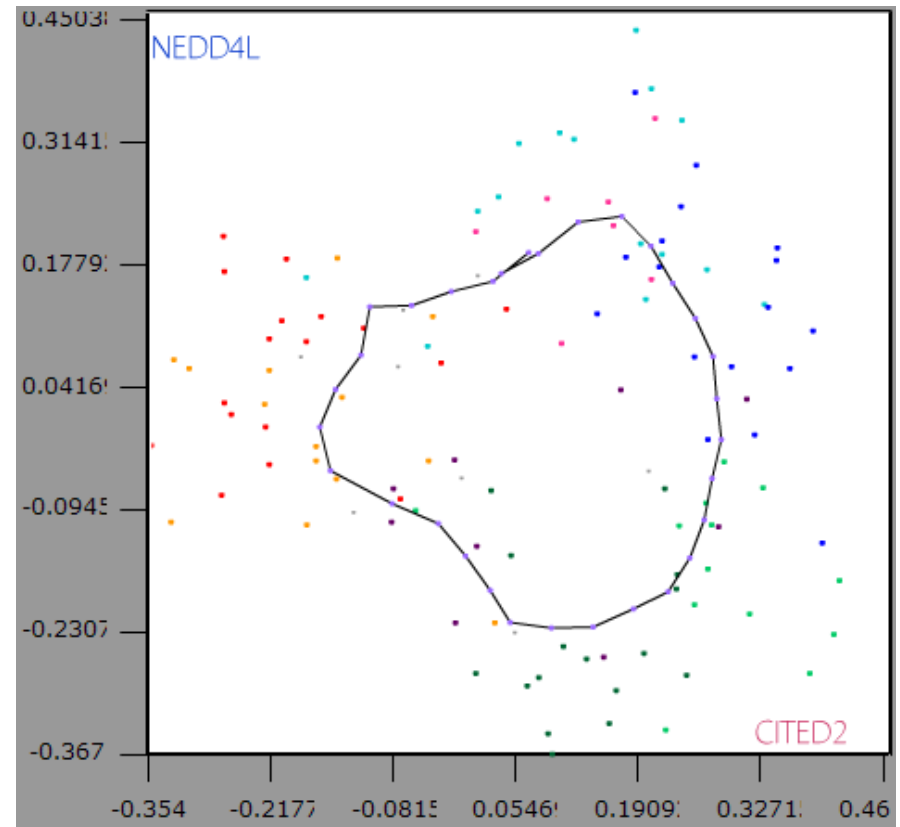
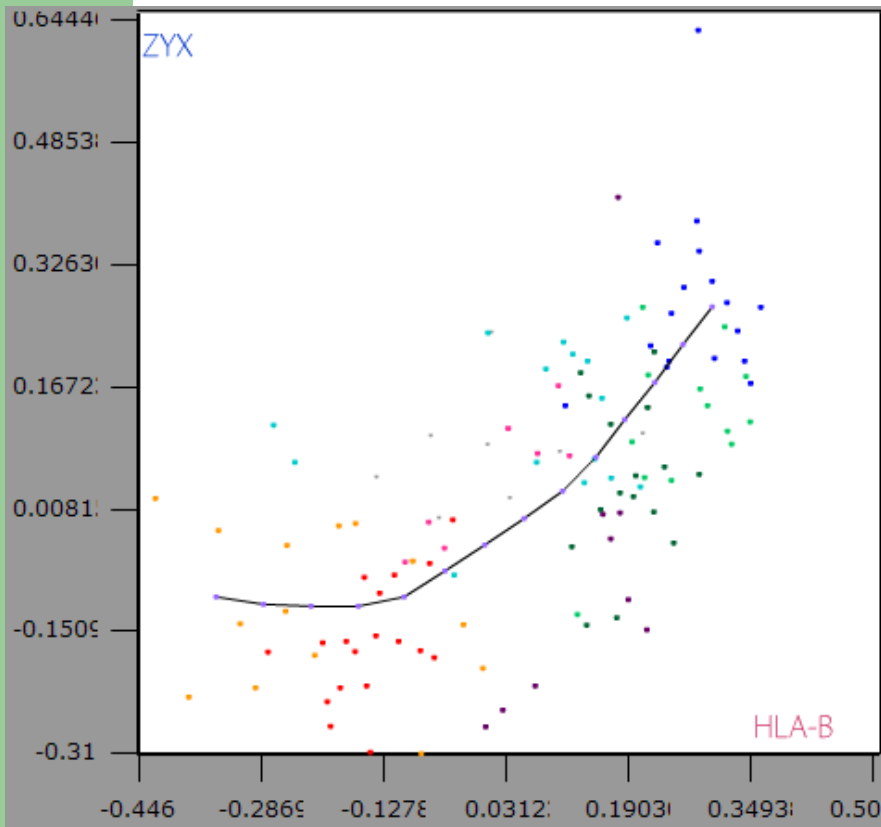


Obtención automática de los cambios fenotípicos basado en las relaciones no lineales

Clustering local a los genes global para las condiciones muestrales



Cruzar fenotipos encontrados con los clusters globales



Y el análisis continúa....

- Obtener genes marcadores para los diferentes fenotipos.
- Cruzar los resultados con otras bases de datos biomédicas.
- Cruzar genes marcadores con genes marcadores de las microarrays de GEO.
- Obtener grafo de cambios fenotípicos con los genes que nos llevan de un fenotipo a otro y los genes que constituyen cada fenotipo.
- ...

Publications

- [Delicado, P.\(2001\) Another look at principal curves and surfaces. Journal of Multivariate Analysis, 77, 84-116.](#)
- [Delicado, P. and Huerta, M. \(2003\): 'Principal Curves of Oriented Points: Theoretical and computational improvements'. Computational Statistics 18, 293-315.](#)
- [Cedano J, Huerta M, Estrada I, Ballllosera F, Conchillo O, Delicado P, Querol E. \(2007\) A web server for automatic analysis and extraction of relevant biological knowledge. Comput Biol Med. 37:1672-1675.](#)
- [Huerta M, Cedano J, Querol E. \(2008\) Analysis of nonlinear relations between expression profiles by the principal curves of oriented-points approach. J Bioinform Comput Biol. 6:367-386.](#)
- [Cedano J, Huerta M, Querol E. \(2008\) NCR-PCOPGene: An Exploratory Tool for Analysis of Sample-Classes Effect on Gene-Expression Relationships Advances in Bioinformatics, vol. 2008.](#)
- [Huerta M, Cedano J, Peña D, Rodriguez A, Querol E. \(2009\) PCOPGene-Net: holistic characterisation of cellular states from microarray data base on continuous and non-continuous analysis og gene-expression relationships. BMC Bioinformatics 2009 May 9;10:138.](#)
- [Huerta M, Fernández-Márquez J, Cabello JL, Medrano A, Querol1 A, Cedano J \(2011\) Studying glucocorticoids' Dual Behaviour and Other Tumour-Progression Paradoxes by means of Exhaustive Analysis of Phenotypic Interdependences, Nature Oncogene \[Accepted\]](#)

Nuestra línea de investigación más ambiciosa: Revertir el dogma de la biología molecular

Genotip

Informació heretable continguda per tots els éssers vius i codificada en forma de DNA (o RNA) que serveix com a recepta per crear i modelar l'ésser viu.



Lamarckisme



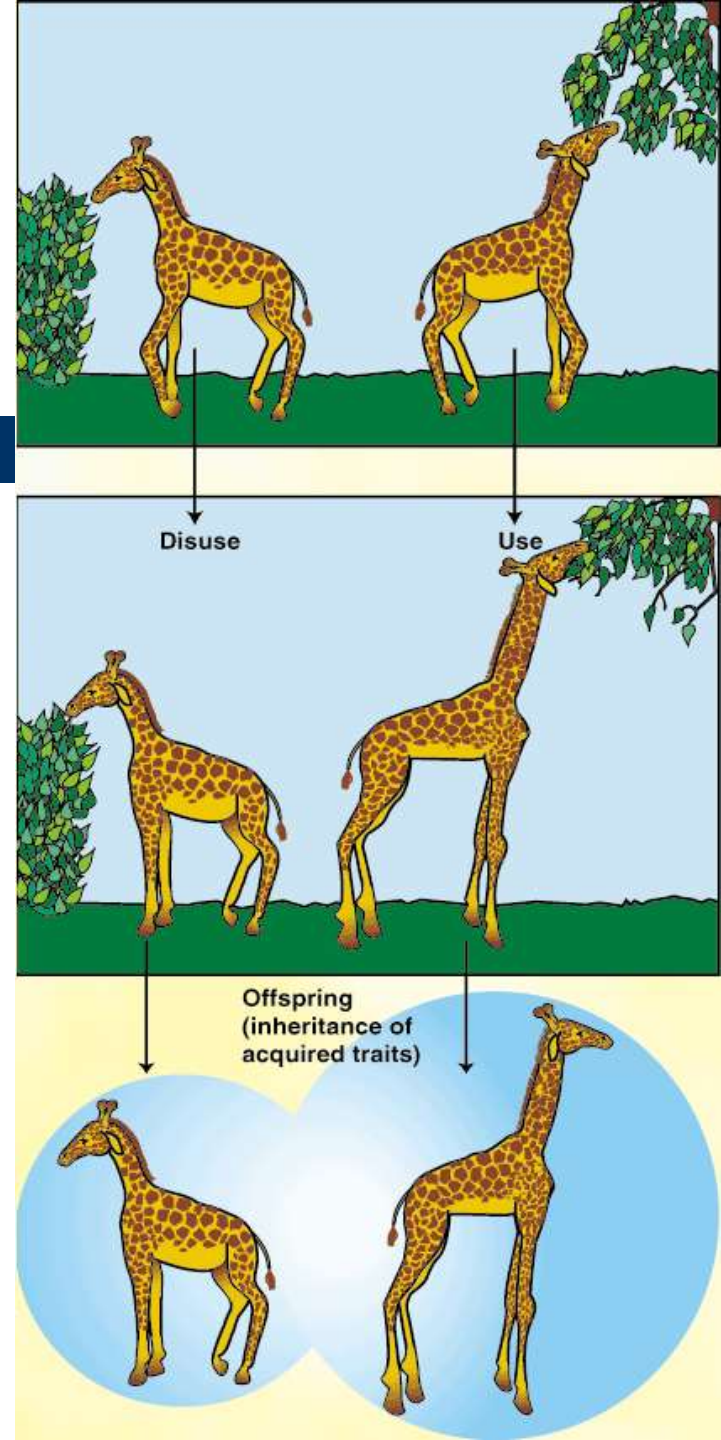
Fenotip

Manifestació física de l'organisme (proteïnes, glúcids, lípids...), en funció del seu genotip i de les condicions ambientals en les que es trobi.

¿La evolución no es producto del azar?

Las mutaciones forman parte de un proceso adaptativo (aprendizaje no supervisado).

Gran parte de las enfermedades que hombres y animales sufren son producto de este proceso adaptativo.



Estrés celular y cáncer.

Mutaciones

