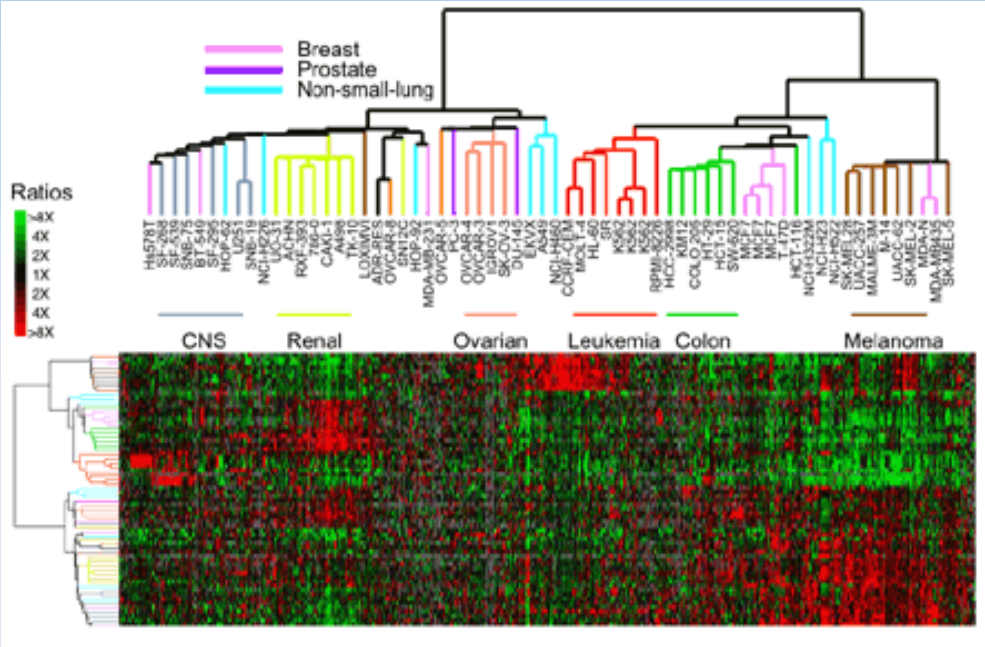
Obtenció de jerarquies d'estat cel·lular via web

Autor: Bernat Gispert

Direcció: Mario Huerta (IBB-UAB) i Jordi Gonzàlez (CVC-UAB)

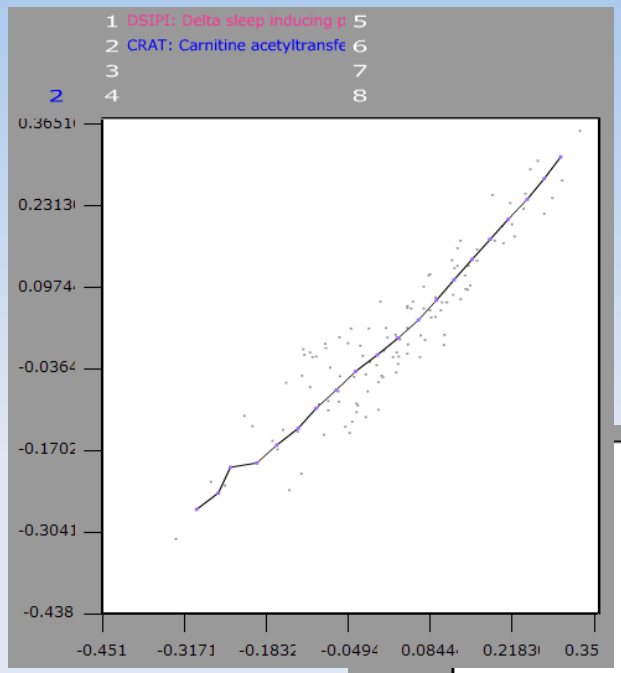
MICROARRAYS

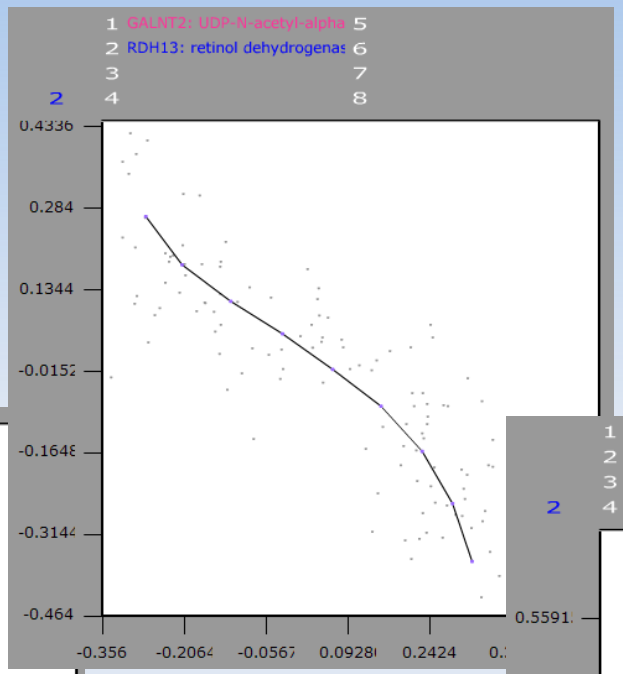


Tecnologia que permet **obtenir** els **nivells d'expressió** d'un gran número de gens per un gran nombre de **condicions experimentals**

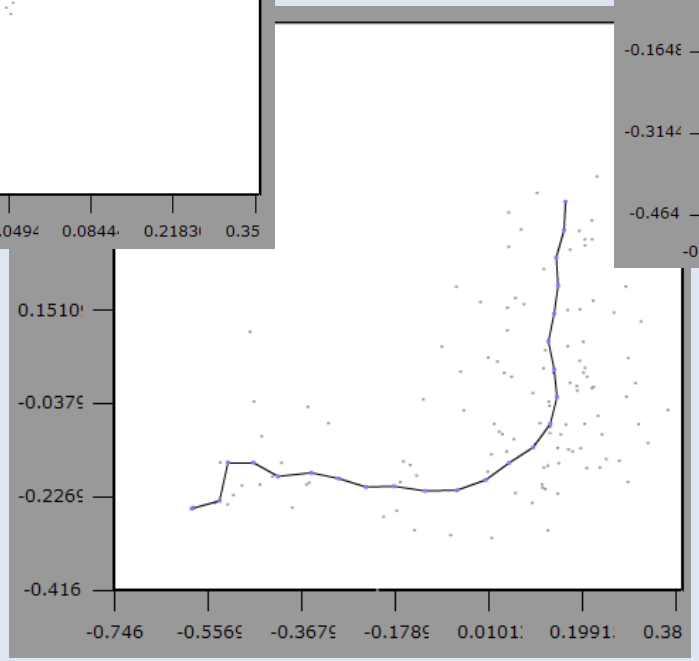
Càlcul de les **RELACIONS D'EXPRESSIÓ** entre parelles de gens

Expressió gènica i canvi d'estat cel·lular

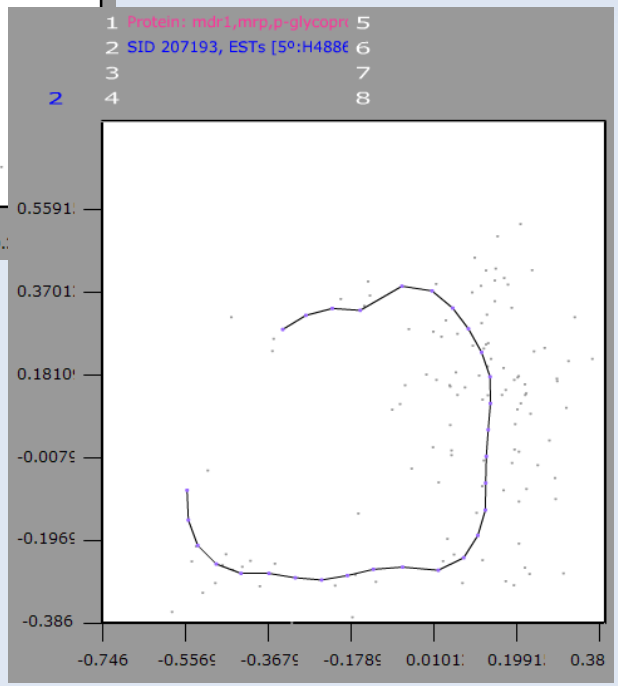




Relacions de coexpressió / inhivició (linials)



Relacions no linials



Estudi només de les relacions d'expressió **NO LINIALS**

Gens involucrats en un mateix canvi fenotípic o cel·lular

OBJECTIUS

Càlcul de les relacions no linials de parelles de gens (Poligonal de la corba PCOP)

Cercar els clusters de condicions mostrals dels diferents canvis fenotípics (clusters de les relacions no linials)

Dissenyar un mètode per buscar distribucions de clusters finals → **Agrupar relacions no linieals SEMBLANTS** → Parelles de gens involucrats TOTS en el mateix canvi fenotípic

Aplicació web per veure les distribucions de clusters finals

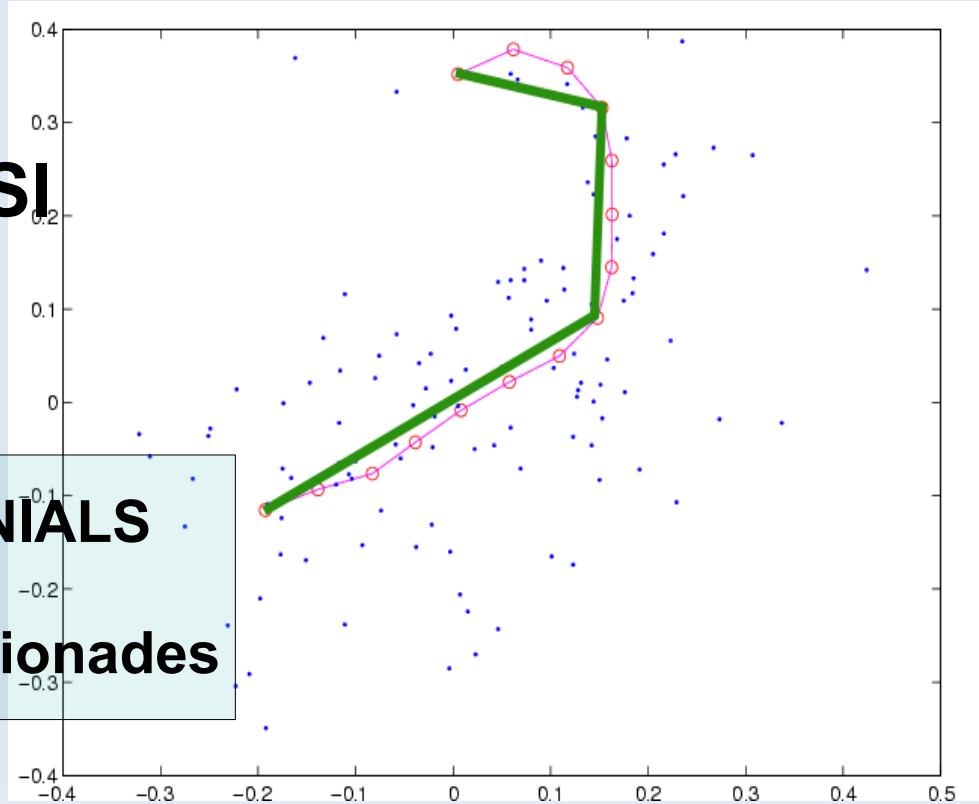
Facilitar poder **comparar resultats** segons els diferents **paràmetres** utilitzats en l'algorisme d'agrupació

Càlcul de PCOP (Principal Curve of Oriented Points)

Components Principals a nivell local n'obtenim els **POPs** (Principal Oriented Points)

La **PCOP** és la corba continua que passa a través d'un núvol multimensional de dades.

Obtenció de la **poligonal de la corba**



CORVES OBJECTE D'ANÀLISI

1- Relació de parelles de gens NO LINIALS

2- Parelles de gens altament correlacionades

Preprocés (I) – Filtres de distribucions de clusters per graus de correlació i grau de corbatura de la PCOP

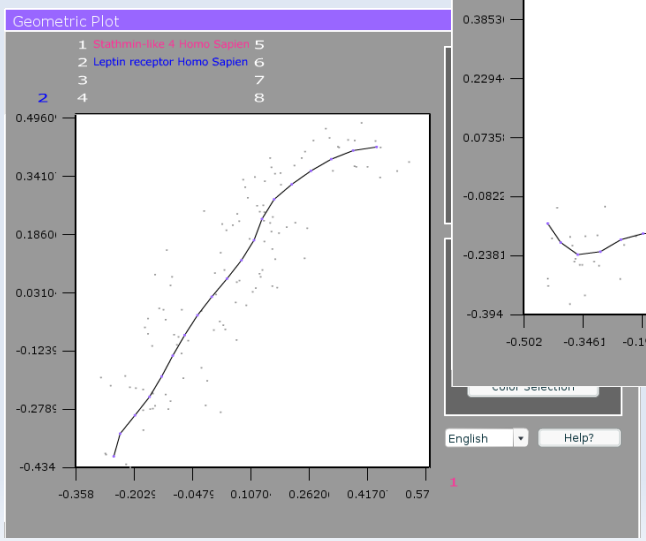
De les dades de la Microarray, càlcul de les PCOPs de la relació de totes les parelles de gens

Càlcul de les poligonals de cada PCOP

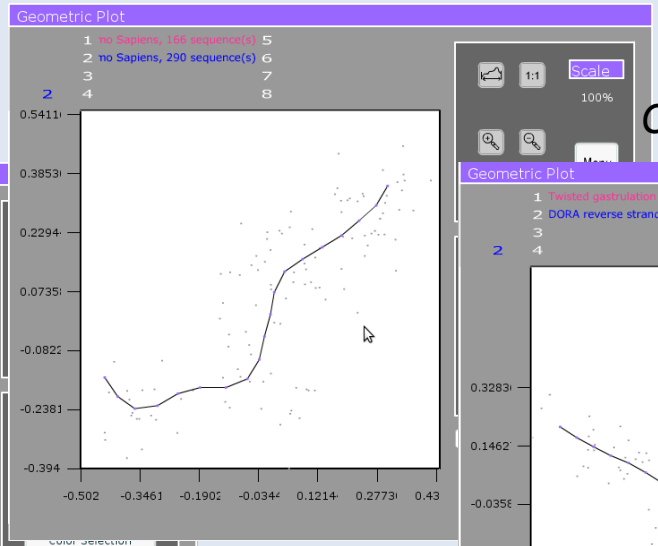
Obtenció dels clusters associats a la poligonal en base a les samples/mostres dels clusters dels POPs que pertanyen a la poligonal

Distribucions de Clusters per correlació de la PCOP i mostres

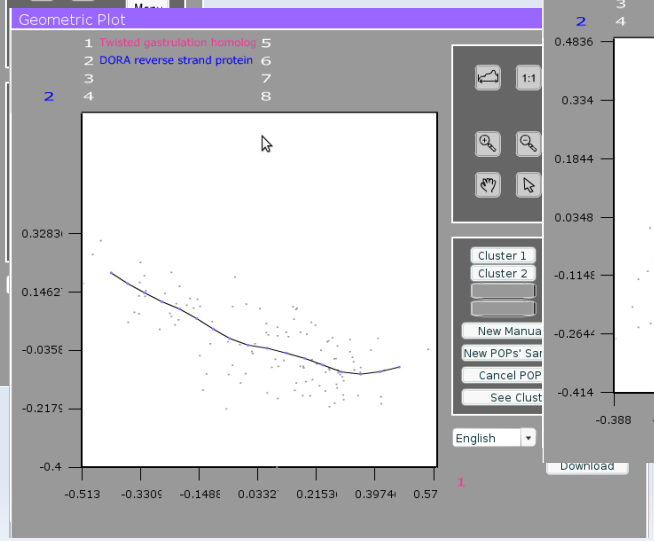
Alta correlació



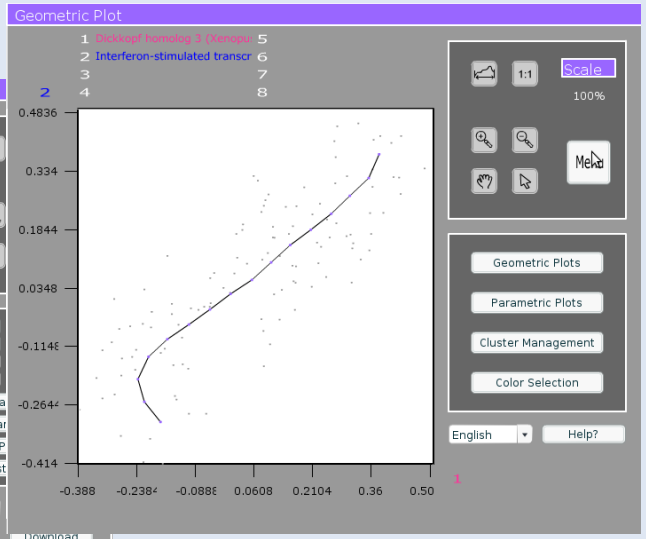
Corrleació mitja i alta corbatura



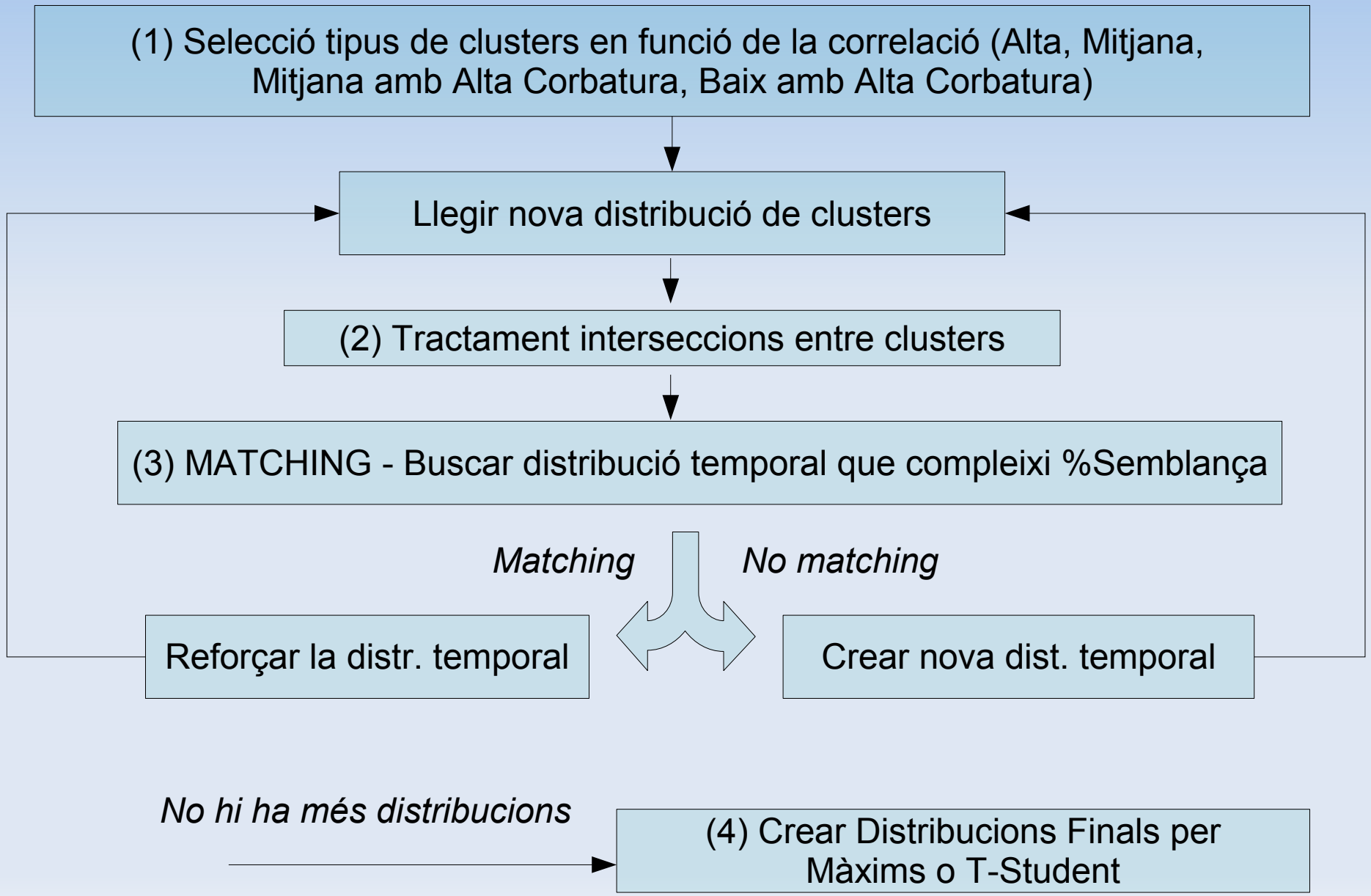
Corrleació mitja



Corrleació baixa i alta corbatura



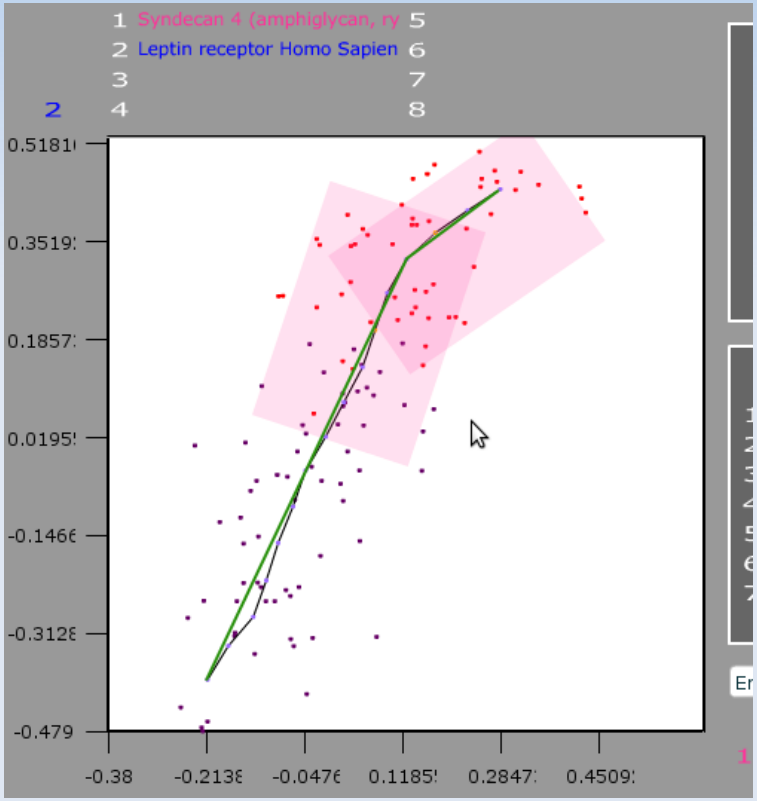
Preprocés – Agrupació de distribucions de clusters semblants



Tractament de les interseccions entre clusters (I)

COM TRACTAR LES ZONES D'INTERSECCIÓ ?

PER FER EL MATCHING ENTRE DISTRIBUCIONS DE CLUSTERS AQUESTES LES VOLEM CLARAMENT DIFERENCIADES



POSSIBLES SOLUCIONS PROPOSADES

DESCARTAT DE LES MOSTRES EN ZONES D'INTERSECCIÓ

ASSIGNACIÓ DE LES MOSTRES AL **CLUSTER MES PETIT**

Tractament de les interseccions entre clusters (II)

Distribució de clusters amb intersecció:

```
[ ************* ********************************************       ***    * * **       *********  ]
[              *      * **    *  ***** ********        *  ***** ************************** **** * ]
```

Interpretació **DESCARTAT** de mostres d'intersecció entre clusters

```
[ ************* ****** * **** **      *        ******** **                              *    *  ]
[              *                                     ***** *   **** * *  *******           ]
```

2) Interpretació **ASSIGNACIÓ** de mostres d'intersecció al **CLUSTER ABSORVIT**

```
[ ************* ****** *  **** **     *        ******** **                              *    *  ]
[              *      * **    *  ***** ********        *  ***** ************************** **** * ]
```

Matching entre distribucions de clusters

UTILITZEM **MATRIU DE SEMBLANÇA**, PER A DIFERENTS PERCENTATGES:

C11:	X				X	X			X	X	X						X	X		X		X	X				=	11 mostres
C12:		X	X	X			X	X				X	X	X	X	X			X		X			X	X	X	=	15 mostres

C21:	X				X	X			X	X	X	X		X			X	X		X		X	X				=	13 mostres
C22:		X	X	X			X	X					X		X	X			X		X			X	X	X	=	13 mostres

	C11	C12					C11	C12
C21	11	2	->	13 samples	->	C21	**85%**	15%
C22	0	13	->	13 samples	->	C22	0%	**100%**

Condició SEMBLANÇA:

Matching si TOTES les files tenen una cel·la >= %matching

Matching si TOTES les columnes tenen una cel·la >= % matching

LA DISTRIBUCIÓ TEMPORAL VARIA INFLUENCIADA SEMPRE PER LA INCORPORACIÓ

Assignació de samples a Distribució de Clusters Finals i Temporals

```
[ *************************  *********** ********   ************   ************************** * ** **         ********** ] - 97
[ ************************************** ******************************************     * ***   ** * **       ********** ] - 99
[  ************************************** ******     ******    **********  ***  * ****   ***** ***  * **       ********** ] - 88
[ ************************************** ****** *   ** **     **  ****** ***   * ** *     ***  *    *        ********** ] - 79
[ ************************************** ********   * ***  ******* ************* ** *   * ***   *  **        ********** ] - 90

[  **                                  * ***  *****  ** ****************** *****  ****************************          ] - 64
[                                      *  *   *****   * ** *************       *  ******************************** **** * ] - 59
[       *  *            *             ************  ************************************* ******************    * *    ] - 72
[ ***                                   **   *************************** ************************************** ** ****   ] - 73
[ **   *                               * **  ******* ********************* ************************************** ****   ] - 75
```

Alta correlació - Matching 75% - Assignació al Cluster Absorvit

(1) ASSIGNACIÓ A LA DISTRIBUCIÓ DE CLUSTERS FINAL PER **MÀXIMS**

```
[ ************************************** ** ***     **  *                ***   * **        *                  ********** ]
[                                      *  *   *****  ** ****************   *** *  ******** ******************           ]
```

(2) ASSIGNACIÓ A LA DISTRIBUCIÓ DE CLUSTERS FINAL PER **T-STUDENT**

*L'**espai mostral** per T-Student és els VALORS DE REFORÇ de cada cluster*
***Ordre assignació**: De cluster petit a més gran, i les mostres que no assignades, assignem per màxim*
***Tall mínim** = 1*

Alfa = 0.003 Tall C1 = C2 = -1,07

```
[    **  ** ************ **************                                                                    *    *  ]
[ ***  **  *            *              ************************************************************************ **** * ]
```

Alfa = 0.009 Tall C1 = C2 = 0

```
[    **  ** ************ **************                                                                    *    *  ]
[ ***  **  *            *              ************************************************************************ **** * ]
```

Alfa = 0.15 Tall C1 = C2 = 1,88

```
[    ************************************** *   *       *                                                    *    ** ]
[ ***                                  * *** ******* ************************************************************ ****   ]
```

Alfa = 0.40 Tall C1 = C2 = 2,40

```
[ * ************************************** *  ***     **                    *                                  *    ** ]
[  *                                   * **   *****  ********************* ************************************** ****   ]
```

Càlcul Error

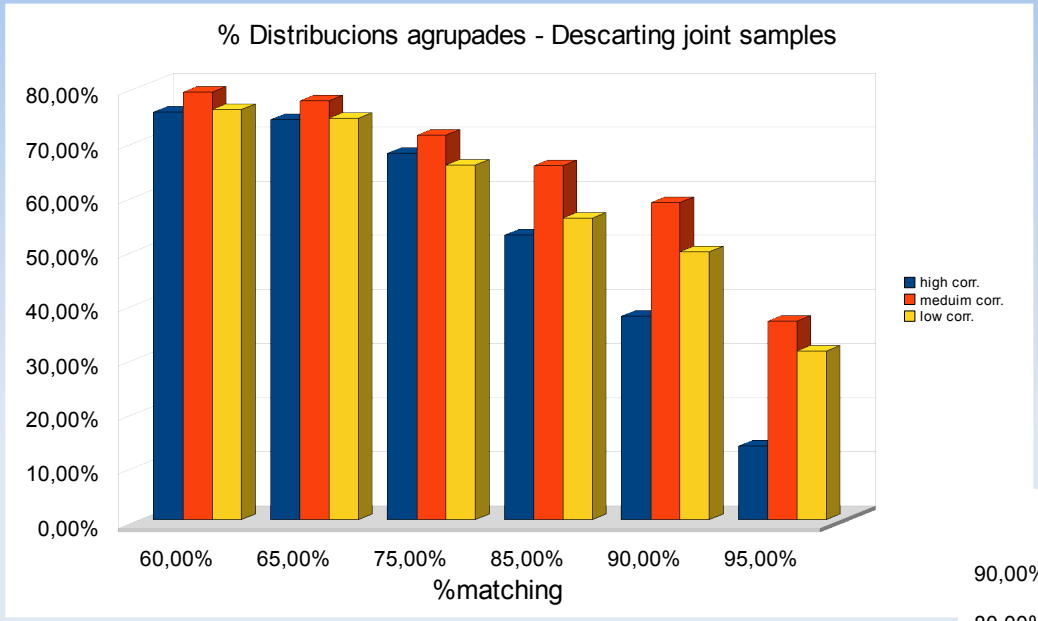
Poder comparar les diferents distribucions de clusters finals

”Mitja” de mostres de les distribucions de clusters d'entrada mal assignades a la distribucio de clusters final

Error segons la Selecció de mostres

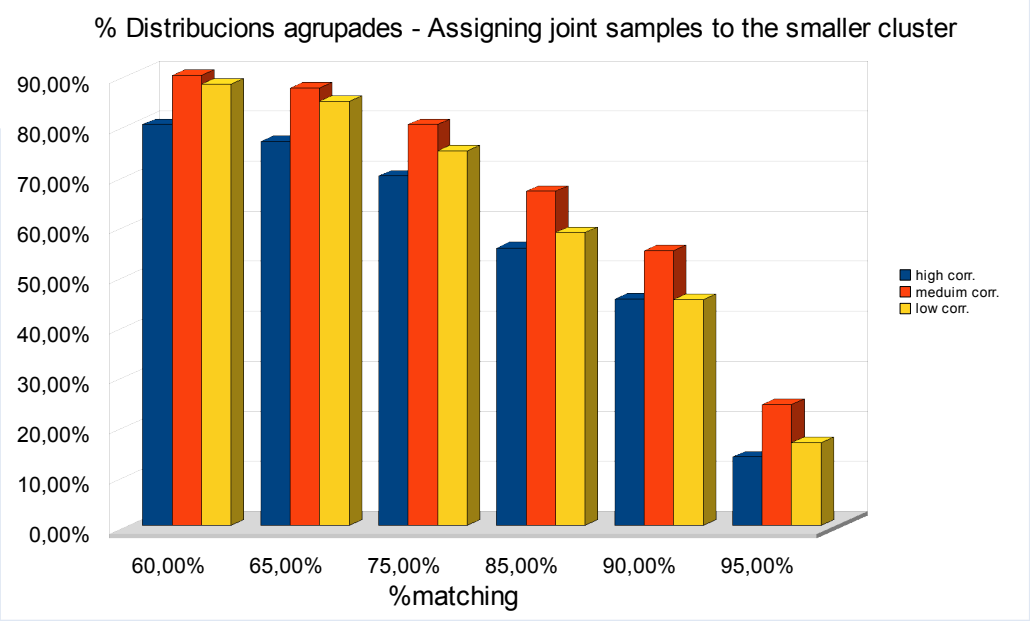
Màxims < Error < T-Student (0,003)

Resultats Agrupació



Descartant intersecció

% distribucions agrupades en funció del tipus de distribucions de clusters d'entrada i el %matching exigit.



Intersecció al cluster més petit

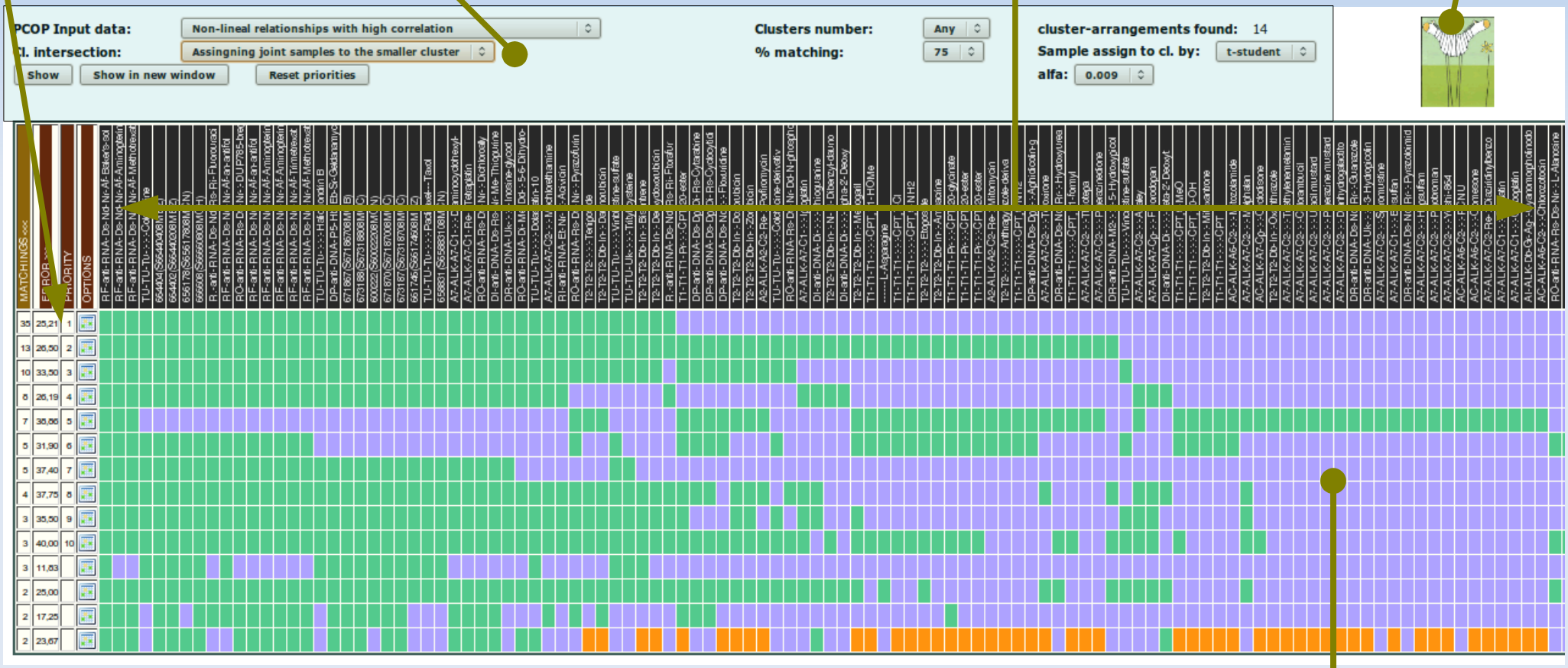
Visualització web

Paràmetres pantalla : PREPROCÉS

Ajuda pantalla

Mostres de la MICROARRAY ordenades per clusters i prioritats de les distribucions finals

Prioritat ordenació clusters



Error acumulat de cada distribució de clusters final

Assignació color a mostra / cluster

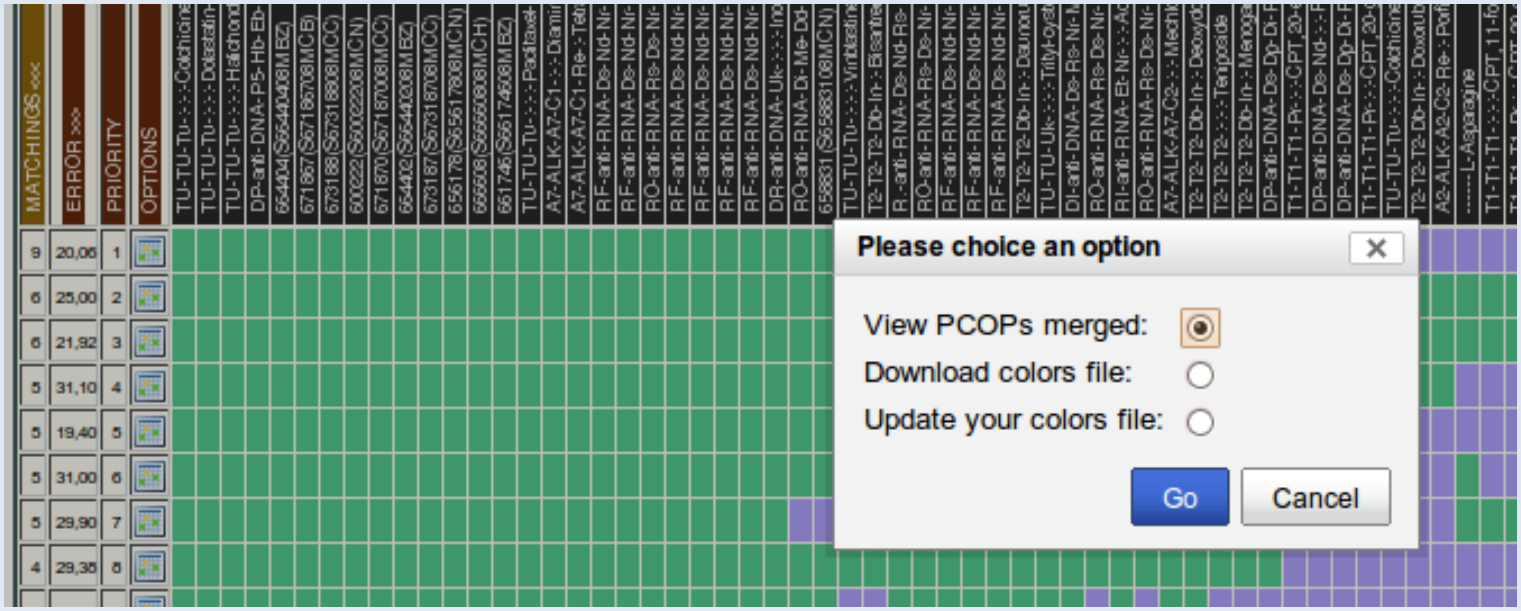
 # Matchings per cada distribució de clusters final

Opcions per a 1 distribució de clusters final

Veure els parells de gens que contribueixen al mateix canvi fenotípic.

Descarregar el fitxer colors d'aquesta distribució per a ús d'altres aplicacions web del servidor d'aplicacions.

Canviar el fitxer colors que per defecte l'investigador associat a la microarray que s'està analitzant



Llistat de distribucions

Dades d'origen del llistat de distribucions

Accés a la BBDD pública GEO (Genne Expression Omnibus) on hi ha informació del gen a altres microarrays

Correlació entre la parella de gens

Accés als gens + correlacionats

Dibuix del tipus de corba i accés a la visualització de la relació d'expressió i clusters del parells de gens, utilitzant aquesta distribució de clusters final.

Número de distribucions detectades que configuren el mateix canvi fenotípic.

85% mergs:6 error:25 t-st 0.009 - Mozilla Firefox

http://revresearch.phpwebhosting.com/applic/gexp/microar

PCOPs merged analysis

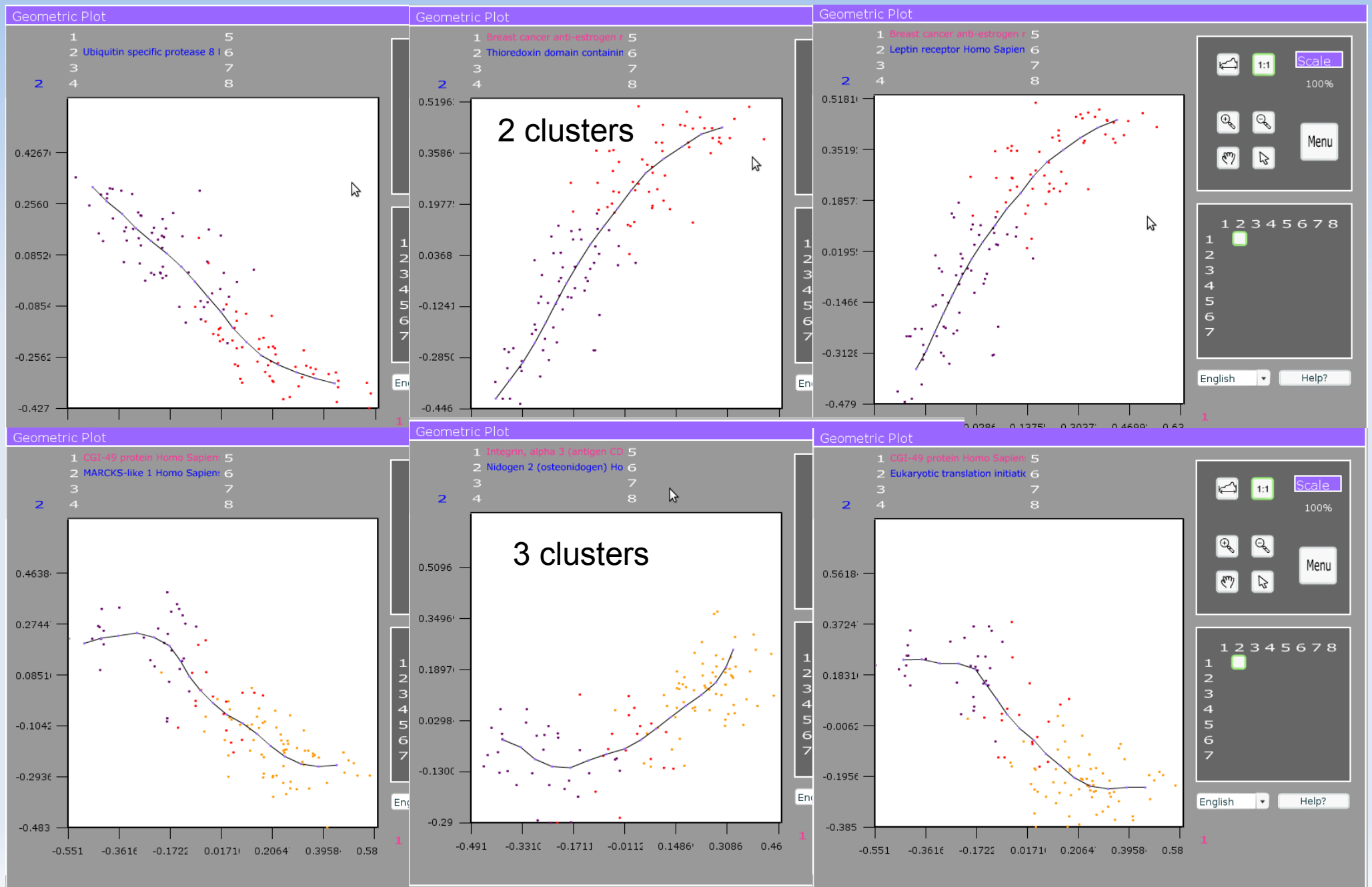
Results

Id	Gene-1	f: uncorrelation	Id	Gene-2	Graphical interface
534	This term is	0.042821	902	USP8	
539	BCAR3	0.043974	776	TXNDC5	
539	BCAR3	0.046067	777	LEPR	
534	This term is	0.049473	776	TXNDC5	
539	BCAR3	0.051065	779	LEPR	
538	STMN4	0.051240	779	LEPR	

Number of merges: 6

Fet

Gràfiques de distribucions finals



CONCLUSIONS

Càlcul de les relacions no linials de parelles de gens (Poligonal de la corba PCOP)

Cercar els clusters de condicions mostrals dels diferents canvis fenotípics (clusters de les relacions no linials)

Buscar distribucions de clusters finals → **Agrupar relacions no linieals SEMBLANTS** → Parelles de gens involucrats TOTS en el mateix canvi fenotípic

Interfície web per veure les distribucions de clusters finals

Possibilitat de poder **comparar resultats** segons els diferents **paràmetres** utilitzats en l'algorisme d'agrupació

Bibliography

http://revolutionresearch.uab.es : Web server for on line microarray analysis supported by the Institute of Biotechnology and Biomedicine of the Autonomous University of Barcelona (IBB-UAB).

Huerta M, Cedano J, Querol E. (2008) Analysis of nonlinear relations between expression profiles by the principal curves of oriented-po 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-Expressic , 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-continuos analysis og gene-expression relationships. BMC Bioinformatics.

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.