

## ENGINYERIA INFORMÀTICA

### MEMÒRIA PRÈVIA DEL PROJECTE

#### **139 Obtenció de Jerarquies d'Estats Cel·lulars vía Web**

Signatura de l'estudiant	Signatura del director/a o directors/es
Nom: Bernat Gispert Pons	Nom/s: Jordi González Sabaté i Mario Huerta Casado
Data: 31 de Març de 2010	Dpt: CVC i IBB
	Data: 31 de Març de 2010

## Introducció

L'anàlisi o estudi dels microarrays no és un problema senzill de resoldre, degut al gran nombre de mostres que inclouen i la relació sovint no lineal que hi ha entre elles. Això fa que hi hagi nombrosos estudis diferents sobre l'anàlisi de microarrays. El que es preten en aquest projecte és aplicar una metodologia d'anàlisi fiable i més precisa que d'altres ja aplicades (la que proposen Delicado, Huerta i Cedano), sobre un conjunt gran de mostres de relació no lineal, i proporcionar una eina per poder estudiar i determinar amb més precisió la relació que hi ha entre l'expressió de cada un dels gens de la cel·lula, davant dels estímuls als que s'ha sotmés aquesta, és a dir, estudiar què és el que provoca aquest procés de canvi d'un estat cel·lular a un altre. Això permetrà que els resultats obtinguts ajudin amb més precisió en aplicacions mèdiques per detectar i corregir certes patologies.

## Breu introducció a l'estat de l'art del tema proposat

Es pot considerar que el conjunt de la societat científica, dona per bó, que l'anàlisi dels microarrays és un objecte d'estudi molt important en la investigació sobre els estats cel·lulars i els processos que els provoquen. La complexitat dels microarrays fa encara, avui en dia, que les metodologies que s'utilitzen no siguin generalitzades, i encara es busqui la manera d'obtenir uns resultats de manera ràpida i precisa. Un dels problemes és la relació no-lineal que hi ha entre les diferents mostres.

Delicado, P. en el seu estudi "Another look at principal curves and surfaces" publicat el 2001 al "*Journal of Multivariate Analysis*"[5], i al 2003 juntament amb Huerta, M. [6] fa que es faci plausible l'estudi sobre si el càlcul per PCOPs (Principal Curve of Oriented Points) en el conjunt de mostres donades pels microarrays es una bona eina per determinar una aproximació d'espais mostrals prou exacte.

La proposta que ens fa M.Huerta, J.Cedano i Enrique Querol [2][3][4][7] és una nova metodologia sobre l'anàlisi dels microarrays utilitzant el càlcul de PCOPs com a eina principal en l'estudi dels estats cel·lulars i els processos que els provoquen.

Actualment disposem d'una serie d'eines web per l'anàlisi de microarrays basades en el càlcul de les PCOP [3][4][7] al servidor de la UAB: <http://revolutionresearch.uab.es> [1], pero no disposem de cap eina per obtindre automàticament els estats celulars utilitzant aquestes PCOPs.

## Objectiu del PFC

L'objectiu del PFC té dues parts clarament diferenciades, que son:

- Classificar els subespais de relació no lineal obtinguts d'una fase anterior al projecte, i obtenir així la jerarquia de subespais mostrals, de tal forma que podrem tenir classificats els estats cel·lulars com a subprocessos o subespais d'altres estats cel·lulars. Aquesta classificació es farà amb metodologia algorísmica de clustering.
- Un cop s'hagin obtingut les jerarquies de subespais mostrals, es desenvoluparà una eina interactiva per poder consultar via web aquests resultats obtinguts.

## Estat actual del projecte i planificació

El projecte d'anàlisi de microarrays que proposen M.Huerta, J.Cedano i E.Querol es troba actualment en la seva etapa final, encara que després, es puguin trobar altres variacions en la metodologia, aplicacions o propostes a fer per a la seva explotació i investigació. En les seves diferents parts, podem dir que:

- La obtenció de la relació lineal o no lineal entre cada parell de gens a partir del càlcul dels PCOP ja està finalitzada.
- El càlcul de la tendència de cada PCOP i la generació dels fitxers que descriuen a partir d'aquestes mostres els subespais trobats per cada una d'aquestes tendències ja està finalitzada.
- La part final del projecte que consta de 2 fases i que és l'objectiu d'aquest PFC son:
  - Classificació de les configuracions de subespais trobats, per obtenir la jerarquia de subprocessos o subespais que formen cada un dels estats cel·lulars.
  - Desenvolupar una interface web que permeti l'estudi d'aquests resultats, podent buscar i analitzar interactivament quins sons els gens que intervenen en cada un dels estats cel·lulars.

	març	abril	maig	juny	juliol
Anàlisi de les configuracions de subespais i la seva classificació (20%)	◆	◆			
Evaluació i correcció del mètode d'anàlisi i classificació		◆◆			
Desenvolupament d'una aplicació web que mostri els resultats obtinguts (0%)			◆	◆	
Testejar aplicacions i validació de resultats (0%)				◆◆	
Recopilació d'informació utilitzada i transcripció de la memòria (0%)					◆◆

## 5. Referencias

[1] <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).

[2] 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.

[3] 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.

[4] 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 of gene-expression relationships. *BMC Bioinformatics*.

[5] Delicado, P.(2001) Another look at principal curves and surfaces. *Journal of Multivariate Analysis*, 77, 84-116 .

[6] Delicado, P. and Huerta, M. (2003): 'Principal Curves of Oriented Points: Theoretical and computational improvements'. *Computational Statistics* 18, 293-315.

[7] Cedano J, Huerta M, Estrada I, Balllllosera 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.