



## 1. Objectius del projecte

Afegir noves funcionalitats a una eina web ja existent per facilitar la recerca als investigadors en l'àmbit de la expressió gènica. Aquesta recerca es basa en l'anàlisi de unes matrius numèriques que ens indiquen el nivell d'expressió de diferents gens sota diferents condicions experimentals. Aquestes matrius son generades mitjançant la tècnica de microarrays.

La primera funcionalitat a afegir consisteix en identificar els gens que mantenen els mateixos tipus de relacions d'expressió respecte la resta de gens i mostrar-los, mitjançant un color diferent, sobre un graf interactiu que mostra tots els gens relacionats entre ells. Aquesta identificació dels gens es realitzarà afegint una nova interfície gràfica a la aplicació web ja existent partint de les dades que ens proporciona un preprocés previ sobre una microarray donada.

Aquest preprocés sobre les microarrays realitza un seguit de càlculs on genera una sèrie de dades que utilitza la aplicació web, un d'aquests càlculs és l'agrupament dels gens en clústers. Tindrè que adaptar el preprocés existent per a generar de manera automàtica una representació gràfica de la relació dels diferents clústers de gens amb els diferents tipus de relacions d'expressió.

La segona funcionalitat consistirà en generar 3 tipus de llistats que es cridaran des de l'aplicació web abans esmentada. En el primer llistat hi constaran tots el gens correlacionats amb el gen seleccionat, indicant el tipus de corba de la relació si aquesta no es lineal. El segon serà un llistat de relacions no lineals entre clústers de gens amb correlació lineal. El tercer serà un llistat de totes les relacions d'expressió detectades d'un tipus de corba concret. Aquests llistats seran totalment integrats a l'aplicatiu.

## 2. Breu introducció a l'estat del art del tema proposat

El meu projecte s'engloba dins del camp de la Bioinformàtica, ciència dedicada a la aplicació de la tecnologia de computadors a la gestió i anàlisi de dades biològiques. L'àrea d'investigació de la bioinformàtica que tractaré és la de anàlisi de la expressió gènica mitjançant microarrays, que són superfícies solides on s'uneixen fragments de ADN i se'n observa el nivell d'expressió sota diferents condicions.

La aplicació web a la que afegiré les noves funcionalitats es diu PCOPGene [1][2]. PCOPGene té com a objectiu estudiar les relacions d'expressió entre gens a partir de les dades obtingudes amb la tècnica de microarrays. Amb aquesta aplicació l'estudi es fa de manera visual i interactiva mitjançant un graf interactiu que es genera per a cada microarray analitzada i un conjunt d'interfícies gràfiques. En [3] s'explica pas per pas com fer servir aquesta eina web.

Aquestes dades dels microarrays son tractades mitjançant un preprocés per a analitzar els tipus de relacions d'expressió entre gens i també per agrupar en clústers els gens pertanyents al microarray. Entre d'altres, es fa una agrupació de gens en clústers en funció del tipus de relacions d'expressió amb els altres gens. Un dels dos objectius del meu projecte és incorporar una interfície gràfica a l'aplicació web PCOPGene-Net. Per a generar les imatges que utilitzarà les interfícies gràfiques faré us de CLUTO [4], una eina de clústering que permet calcular aquests clústers de gens i extreure informació tant estadística com gràfica dels clústers trobats.

PCOPGene-Net esta basat en els llenguatges PHP, CGIs en C++, Java i flash. Per representar el graf de relacions entre gens fa servir una llibreria Java de codi obert anomenada JUNG (Java Universal Network/Graph Framework) [5]. Aquesta llibreria ens proporciona un llenguatge extensible per al modelat, anàlisi i visualització de dades representades per grafs o xarxes. JUNG disposa d'un entorn de treball per la visualització que facilitarà la exploració interactiva del graf. També disposa de mecanismes de filtratge per tractar parts específiques del graf.

### 3. Estudi de viabilitat del projecte

El pla de treball consta de les següents fases:

- **1ª Fase:**

1. Entendre com funciona l'aplicació web PCOPGene-Net i els conceptes relacionats amb aquesta.
2. Anàlisi del funcionament del preprocés i obtenció dels fitxers necessaris a partir del codi existent.

- **2ª Fase:**

1. Generació (mitjançant CLUTO), tractament i gestió de les imatges dels clusters de gens per a integrar-les a la aplicació web PCOPGen-Net.

- **3ª Fase:**

1. Familiaritzar-se amb les llibreries JUNG i analitzar la implementació mitjançant aquestes llibreries que s'ha utilitzat en la aplicació PCOPGen-Net per a la generació del graf interactiu.
2. Creació de la interfície gràfica que permetrà seleccionar els clusters de gens per tipus de relació d'expressió amb la resta de gens i pintar-los sobre el graf interactiu.

- **4ª Fase:**

1. Optimització i correccions del funcionament del pintat d'arestes segons el tipus de relació no lineal que contenen aquestes.
2. Generació dels tres tipus de llistats de relacions que es cridaran des de diferents interfícies gràfiques de l'aplicació web abans esmentada.

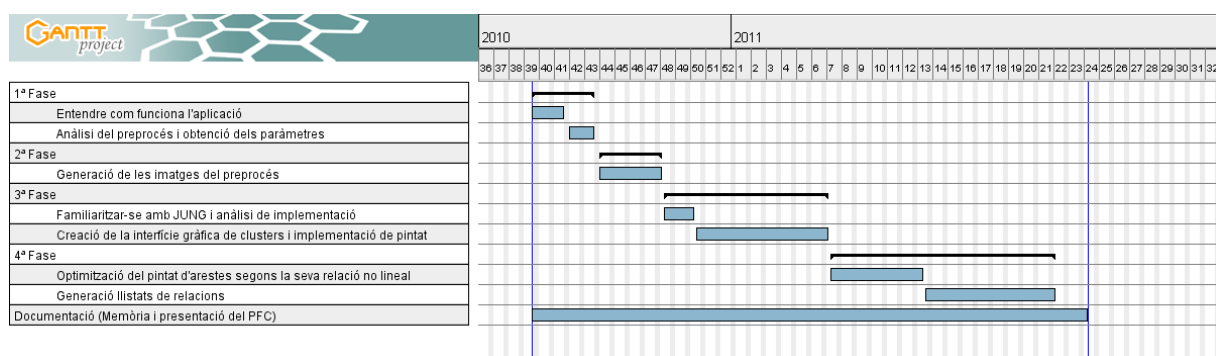
Actualment em trobo a la tercera fase del pla de treball. Considero que podré acomplir totes les fases del pla de treball en el temps estimat.

La viabilitat tècnica no suposa cap problema ja que puc disposar de les eines necessàries per desenvolupar aquest projecte, tant a nivell hardware (ordinador, accés al servidor on està el codi de la aplicació PCOPGen-Net) com software (llibreries JUNG, CLUTO i codi font del preprocés i de PCOPGen-Net).

Tampoc hi ha problemes en la viabilitat operativa, ja que en el pla de treball les fases son totalment seqüencials i els recursos els tindrè sempre a la meua disposició.

No hi ha cap tipus de dependència econòmica en l'accés als recursos, per tant la viabilitat econòmica està garantida.

#### 4. Planificació temporal del treball



#### 5. Bibliografia

- [1] ::: r-evolution research server :::  
<http://revolutionresearch.uab.es/main.html>.
- [2] Huerta M, Cedano J, Peña D, Rodriguez A, Querol E. (2009) PCOPGene-Net: holistic characterisation of cellular states from microarray data based on continuous and non-continuous analysis of gene-expression relationships. *BMC Bioinformatics.*, 9;10:138  
<http://www.biomedcentral.com/1471-2105/10/138>.
- [3] PCOPGene-Net Tour.  
<http://revolutionresearch.uab.es/downloads/PCOPGene/PCOPGene.htm>.
- [4] CLUTO - Software for Clustering High-Dimensional Datasets | Karypis Lab.  
<https://docs.google.com/viewer?url=http://glaros.dtc.umn.edu/gkhome/fetch/sw/cluto/manual.pdf>.
- [5] JUNG - Java Universal Network/Graph Framework  
<http://sourceforge.net/apps/trac/jung/wiki/JUNGManual>