TREBALL DE FI DE MÀSTER

Màster en Estadística i Investigació Operativa

Títol: Cerca de patrons per l'anàlisi de la significació biològica en experiments d'estrès tèrmic

Autor: Anna Díez Villanueva

Director: M. Carme Ruiz de Villa

Departament: Estadística (UB)

Convocatòria:**Juny 2009**

Facultat de Matemàtiques i Estadística

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Cerca de patrons per l'anàlisi de la significació biològica en experiments d'estrès tèrmic

Anna Díez Villanueva

10 de juny de 2009

Índex

1 Introducció **5**
- 1.1 Motivació . . . 5
  - 1.1.1 Dogma central de la Biologia Molecular . . . 6
  - 1.1.2 Chips o arrays d'expressió . . . 8
  - 1.1.3 L'organisme *Saccharomyces cerevisiae* . . . 10
  - 1.1.4 Anàlisi de dades d'expressió genètica . . . 11
  - 1.1.5 Anàlisi de dades de *time course* . . . 11
  - 1.1.6 Mètodes de Clustering . . . 12
  - 1.1.7 Anàlisi d'enriquiment . . . 13
  - 1.1.8 Software . . . 13
- 1.2 Objectius . . . 14
- 1.3 Organització de la Memòria . . . 14

2 Dades **17**

3 Preprocessat **19**
- 3.1 Control de qualitat . . . 19
- 3.2 Filtratge . . . 20
- 3.3 Normalització . . . 20

4 Algorisme DIB-C **23**
- 4.1 Dades . . . 23
- 4.2 Estadístic t moderat . . . 23
- 4.3 Algorisme . . . 25
- 4.4 Output . . . 27

5 Càlcul del nombre òptim de clusters **29**
- 5.1 Matriu d'anotacions de la GO . . . 29
- 5.2 Càlcul de la Informació Mútua (MI) real i aleatòria . . . 31
- 5.3 Càlcul del Z-score . . . 32
- 5.4 Escollir el màxim Z-score . . . 32

6 Anàlisis d'enriquiment **35**

7 Resultats **37**
7.1 Resultats de l'algorisme DIB-C 37
7.2 Resultats de l'anàlisi d'enriquiment 38

8 Conclusions i procediments futurs **41**

Bibliografia **44**

A Gràfics de les dades originals **45**
A.1 BoxPlots de les dades originals sense normalitzar: 45
A.2 Gràfics MA de les dades originals sense normalitzar: 52

B Gràfics de les dades normalitzades **65**
B.1 BoxPlots de les dades normalitzades: 65
B.2 Gràfics MA de les dades normalitzades: 72

C Gràfics dels clusters **85**

D Gràfics dels gens rars **157**

E Taules obtingudes de l'Anàlisi d'Enriquiment **161**

F Implementació **171**
F.1 Fitxer a executar 171
F.2 Funció Preprocessat() 174
F.3 Funció Dibc() 179
F.4 Funció MatGOgen() 184
F.5 Funció CalculZscore() 186
F.6 Funció GrafClus() 188
F.7 Funció GOAnalysis() 194

Capítol 1

Introducció

1.1 Motivació

El procés d'expressió de gens és conegut des de 1963. Però no va ser fins el 1977 que es va desenvolupar un procediment pràctic de mesura, encara que només permetia mesurar un gen. Finalment, el 1989 es van desenvolupar mètodes basats en la tecnologia de construcció de microchips, que permeten fixar milions de cadenes de DNA sobre una placa.

En les últimes dos dècades els chips d'expressió gènica han passat a ser l'eina principal per l'anàlisi d'expressió de gens simultàniament.

L'expressió gènica és el procés pel qual tots els organismes transformen la informació codificada en els àcids nucleics en les proteïnes necessàries per el seu desenvolupament i funcionament.

Un gen és la unitat física i funcional de la herència que es transmet de generació en generació. També es pot definir com un fragment de DNA que codifica una o diverses proteïnes per tal de sintetitzar-les.

El DNA, àcid desoxiribonucleic, (Figura 1.1), és una molècula que s'organitza en una doble cadena d'elements anomenats nucleòtids Els nucleòtids són bases nitrogenades i en el DNA són l'adenina (A), la timina (T), la guanina (G) i la citosina (C). Les bases de les dues cadenes estan disposades de forma que quan en una cadena hi ha una A en l'altre (complementària) hi ha una T i quan una té una C, l'altre té una G. Aquestes bases s'uneixen per ponts d'hidrogen formant una doble hèlix.

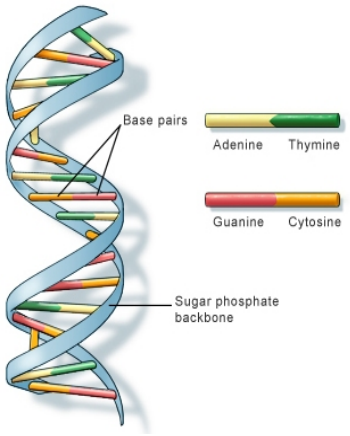


Figura 1.1: Estructura del DNA

El DNA conté tota la informació de l'estructura i funcionament d'un organisme. A cada cèl·lula, només alguns dels gens s'expressen. Això és el que determina la diferència en les funcions específiques de cada cèl·lula a les diferents parts del cos.

1.1.1 Dogma central de la Biologia Molecular

El dogma central de la biologia molecular (Figura 1.2) estableix que la informació codificada en el DNA es transforma en proteïnes a traves del RNA. L'RNA és una molècula formada per una sola cadena de nucleòtids similar al DNA però on la T es substituïda per un Uracil (U). Les fases del dogma central de la biologia molecular són:

1. Fase 1: **Replicació**:

   El DNA es replica en un procés complex en el què intervenen molts enzims catalitzadors.Això permetrà transmetre la informació hereditària.

2. Fase 2: **Transcripció**:

   El fragment de DNA que codifica la proteïna (gen) és copiat en cadenes curtes d'RNA (RNA missatger o mRNA) que contenen la informació codificant. D'aquesta manera, es produeix una cadena d'RNA que conté exons i introns. Els exons són seqüències codificants dels gens i els introns són seqüències no codificants però necessàries pel bon funcionament del gen i per l'augment de la variabilitat genètica.

3. Fase 3: **Splicing**:

   Els introns són eliminats i queden només els exons.

4. Fase 4: **Traducció**:

   Es sintetitzen les proteïnes unint aminoàcids en el mateix ordre en què estan codificats en l'RNA. Cada tres nucleòtids (triplets o codons) codifiquen un aminoàcid. La seqüència d'aminoàcids de la proteïna resultant correspon a la seqüència codificada pel DNA del gen.

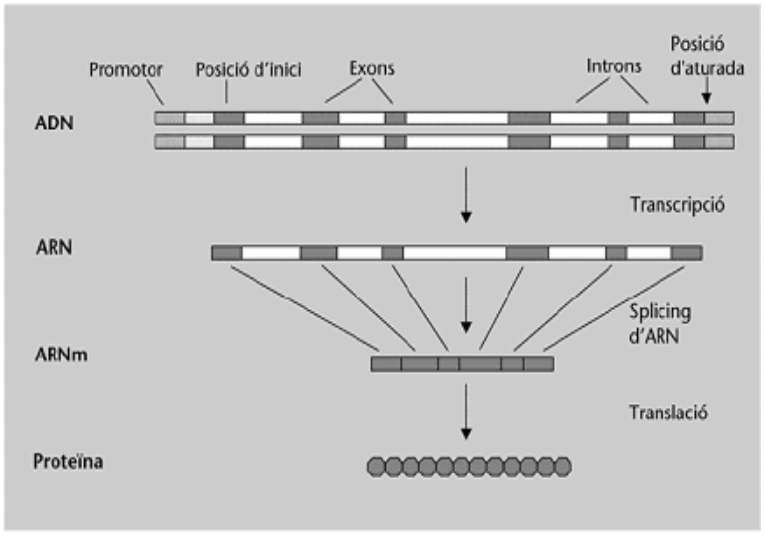


Figura 1.2: Dogma Central de la Biologia Molecular

Quan es volen localitzar nous gens examinant el genoma es fan servir els ORF (*Open reading frame*). Un ORF és una seqüència de nucleòtids que potencialment pot codificar una proteïna ja que està compresa entre una seqüència d'inici (codó d'inici) i una seqüència de terminació (codó d'aturada o codó de stop) i té una llargada considerable. L'existència d'un ORF, especialment quan es tracta d'una seqüència llarga, és un bon indicador de la presència d'un gen als voltants d'aquesta seqüència. La seqüència més llarga sense cap codó d'aturada normalment determina l'ORF corresponent al gen estudiat.

1.1.2 Chips o arrays d'expressió

Per estudiar l'expressió de gens es mesura la quantitat de còpies d'RNA que produeix un gen. El chip de DNA és una eina que ens diu la quantitat d'RNA que cada gen està fabricant. La quantitat de mol·lècules d'RNA que un gen transcriu ens dóna una aproximació del nivell d'expressió del gen.

Els chips d'expressió fan servir el concepte d'hibridació per identificar quines seqüències d'RNA estan presents en una mostra i per mesurar la seva expressió. La hibridació és el procés pel qual es combinen dues cadenes d'àcids nucleics antiparal·leles i amb bases complementàries en una única mol·lècula de doble cadena que pren estructura de doble hèlix.

Bàsicament, el procés per crear i analitzar chips d'expressió és el següent:

1. Es fixa una cadena curta de DNA específica per a un gen, anomenada sonda o *probe*, a la superfície del chip.
2. L'RNA s'aïlla d'una mostra biològica (teixit, sang, ...)
3. Es fan milers de còpies de l'RNA extret i s'etiqueta, normalment, mitjançant fluorescència o radioactivitat.
4. L'RNA etiquetat s'incuba amb les *probes* específiques per fer que s'hibridin de la manera clàssica.
5. Després de la incubació, les mostres no hibridades es netegen i es fa una mesura del senyal emès amb un scanner.

Probablement, els chips d'expressió més coneguts són els microarrays [10]. Aquests chips consisteixen en una superfície sòlida, normalment de vidre, plàstic o silicona, en la què s'uneixen una sèrie de fragments de DNA (*probes*) i que mitjançant un etiquetatge fluorescent, es fan servir per trobar l'expressió de tots els gens del genoma d'un organisme. Hi ha dos tipus bàsics de microarrays:

- **Microarrays de dos colors**: Els microarrays de dos colors, també anomenats microarrays de cDNA, parteixen de dues mostres d'interès que són etiquetades mitjançant dos colors diferents, normalment verd (G) i vermell (R) i són hibridades en un únic chip. Les dues mostres competeixen per unir-se a les *probes* de manera que l'objectiu és comparar, a partir d'un únic chip, l'expressió dels gens de les dues mostres mitjançant, normalment, el seu quocient d'intensitat al ser llegit amb un scanner.

A més, aquestes *probes* es creen a part i s'imprimeixen mecànicament a la superfície. El terme cDNA es fa servir perquè la *probe* es copia complementàriament a partir de la seqüència original i cada una representa un gen.

A la Figura 1.3 podem veure el procés de creació d'aquest tipus de chip:

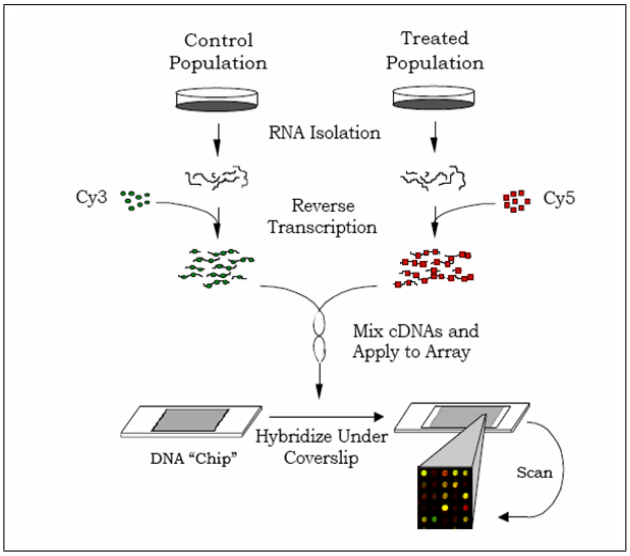


Figura 1.3: Microarray de cDNA

- **Microarrays d'Affymetrix**: Els microarrays d'Affymetrix, també anomenats microarrays d'oligonucleòtids, parteixen d'una sola mostra per cada chip. Això requereix més superficies per a cada experiment i no fa servir hibridació competitiva. Tot i així, la tecnologia és més avançada i el disseny de l'experiment és més simple.

  A més, les *probes* es creen directament sobre la superfície. El terme oligonucleòtid fa referència al fet de què el procés de síntesi només permet crear fragments petits. Així, un gen està representat per una quantitat elevada d'oligonucleòtids (*probe set*).

  A la Figura 1.4 podem veure el procés de creació d'aquest tipus de chip:

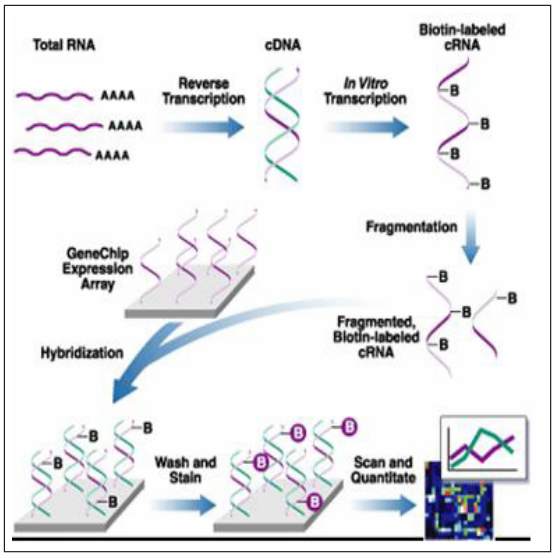


Figura 1.4: Microarray d'oligonucleotids

Els inconvenients alhora d'utilitzar la tècnica dels microarrays són el seu gran cost econòmic, la complexitat de les tècniques de generació del chip i dels mètodes d'anàlisi de les dades.

Un altre tipus de chips d'expressió que es fan servir són els macroarrays, [1] els predecessors dels microarrays, que són més barats i més fàcils de fer servir que els microarrays. Els macroarrays estan fets d'una membrana de nylon reutilitzable i realitzen l'etiquetatge mitjançant radioactivitat. El desavantatge principal d'aquest tipus d'array és que no poden contenir tants gens com els microarrays o altres tipus de chips. La seva densitat de *probes* normalment va de 200 a 10.000 gens comparat amb els 40.000 que pot arribar a tenir un microarray. A més, funcionen de manera semblant als microarrays d'Affymetrix partint d'una sola mostra per a cada chip.

1.1.3 L'organisme *Saccharomyces cerevisiae*

Per poder implementar les tecnologies de chips d'expressió l'organisme que més s'ha fet servir ha estat el llevat (*Saccharomyces cerevisiae*) ja que es coneix la seqüència completa del seu genoma. Aquest organisme té uns 6200 gens i per això no és necessària una densitat de *probes* alta i es poden fer servir els arrays de membrana (macroarrays) enlloc dels microarrays.

1.1.4 Anàlisi de dades d'expressió genètica

Els estudis d'expressió genètica tenen moltes aplicacions d'interès biològic:

- Comparació de grups de gens.
- Descobriment de grups amb igual perfil (clusters).
- Predicció de grups de gens.
- Time Course: perfils d'expressió al llarg del temps.
- Pathway Analysis-(Systems Biology): reconstrucció de xarxes metabòliques a partir de dades d'expressió.
- Whole Genome, CGH, Alternative Splicing: estudis amb dades de diferents tipus (integració).
- Altres aplicacions.

Aquest projecte es centra en el descobriment de grups de gens amb igual perfil d'expressió de l'organisme del llevat en diferents instants de temps (*Time Course Analysis*) i amb rèpliques. L'expressió dels gens s'ha obtingut amb la tècnica dels macroarrays.

1.1.5 Anàlisi de dades de *time course*

Quan les diferents condicions estan representades per diferents punts de temps és important no només entendre perquè un gen s'expressa o no, sinó també quines relacions entre els gens estan basades en els canvis de temps.

Quan el nombre de instants de temps es troba entre 3 i 6 es parla de *short time series* i amb més de 6 instants de temps es parla de *long time series*. També es pot fer la classificació de les dades de *time course* en funció de si les dades són longitudinals o independents (*cross-sectional data*). En les sèries longitudinals els individus són mostrejats en instants de temps diferents, en canvi, en les sèries independents les mostres són d'individus diferents i independents. En aquests últims experiments la mitjana de les rèpliques a cada instant de temps s'analitzarà tenint en compte aquesta independència d'individus diferents.

En el nostre cas, tenim 12 instants de temps i 8 individus diferents, per això parlarem de *long time series* de dades independents.

Normalment, el nombre de instants de temps en les dades de micro - arrays és molt petit per això no es poden aplicar tècniques d'anàlisi de sèries temporals com l'autorregressió (AR), les mitjanes mòbils (MA) o l'anàlisi de Fourier. Per solucionar aquest problema normalment s'usen tècniques de clustering.

1.1.6 Mètodes de Clustering

Les tècniques de clustering permeten fer grups de gens amb un comportament similar i la visualització d'aquests perfils pot revelar patrons biològics significatius.

Tots els mètodes de clustering estan basats en distàncies entre elements del grup i es poden classificar en aquests dos grans grups:

- **Clusters jeràrquics**: els grups estan units en una estructura jeràrquica basada en la distància entre ells.
- **Mètodes de particionament o clusters no jeràrquics**: en els que s'intenta trobar una divisió òptima de les dades a partir d'un nombre prefixat de clusters. En aquest tipus de clustering estan inclosos els mètodes més típics com són el *k-means* o el *Self-Organizing Map* (SOM), ...

Les tècniques de clustering solen ser descriptives i les conclusions no poden avaluar-se en termes de significació estadística. Les limitacions principals dels mètodes habituals de clustering són:

- Necessiten informació prèvia de les dades.
- Els instants de temps diferents es tracten com a successos desordenats en condicions diferents.
- Les rèpliques moltes vegades no s'incorporen.
- La informació visual moltes vegades és poc informativa.

Per aquest motiu, en aquest projecte s'ha decidit aplicar un nou algorisme anomenat DIB-C (*Difference-based clustering*) proposat per Jihoon Kim i Ju Han Kimper [5] per identificar grups de gens que comparteixen un patró temporal. Aquest algorisme pertany als mètodes de particionament i per això s'haurà de calcular prèviament un nombre òptim de clusters.

El DIB-C es basa en trobar els grups de gens mitjançant el càlcul d'uns estadístics basats en la primera i segona diferència entre punts de temps adjacents i, com anirem veient, aquest mètode no genera cap de les limitacions que tenen la resta de mètodes de clustering.

El resultat de l'algorisme és una seqüència de símbols que indiquen el canvi entre instants de temps. Cada gen és assignat a un grup on els seus membres comparteixen el mateix patró.

1.1.7 Anàlisi d'enriquiment

Després d'aplicar els mètodes de clustering a les dades, ens interessarà saber perquè els gens d'un mateix cluster es comporten de manera semblant al llarg del temps. Els gens que mostren un patró similar a l'experiment fan suposar que també comparteixen mecanismes biològics que poden ser la clau per descobrir de quins processos moleculars formen part.

Actualment, un dels enfocaments de la biologia de sistemes més usat per aconseguir aquest propòsit és l'anàlisi d'enriquiment (*Enrichment Analysis* (EA)) que intenta revelar la significació biològica que hi ha dins de cada grup de gens.

1.1.8 Software

Per a la realització del projecte s'ha fet servir el programa R. R conté moltes llibreries o paquets dissenyats per l'anàlisi de microarrays. En concret es farà servir el software Bioconductor aplicat en R. La flexibilitat de programar amb R i amb tots els seus paquets fa que tant l'anàlisi com la generació dels informes es faci de manera automàtica.

El projecte Bioconductor [8] va començar el 2001 com a software obert per a l'anàlisi de dades genòmiques. Des de llavors ha anat creixent fins a obtenir centenars de paquets.[2][9]

La potència del projecte Bioconductor és molt gran ja que els usuaris poden contribuir amb els seus programes i el sistema comprova prèviament que aquests funcionen. D'altra banda, cada tècnica disponible té el seu paquet essent possible que hi hagi tècniques que tinguin més d'un, cosa que dificulta a vegades la seva utilització.

1.2 Objectius

L'objectiu principal del projecte és, a partir de les dades d'expressió gènica obtingudes de l'organisme del llevat mitjançant la tècnica dels macroarrays, trobar grups de gens que es comporten de manera similar. Per fer això, un cop netejades les dades, s'ha implementat l'algorisme DIB-C basat en les primeres i segones diferències entre un instant de temps i el següent. Categoritzant aquestes diferències s'obté un patró per a cada gen que ens permet crear els clusters.

Un cop obtinguts els grups de gens i per ajudar a l'investigador a interpretar els resultats, s'aplicarà un anàlisi d'enriquiment. Aquest anàlisi servirà per definir quines funcions tenen els gens dels clusters en els processos moleculars que es donen dins la cèl·lula, atès que se suposa que els gens d'un mateix cluster tindran un paper semblant.

1.3 Organització de la Memòria

La memòria s'organitza de la següent manera:

En el capitol 2 es fa una breu explicació de les dades obtingudes per l'investigador així com de totes les variables que les acompanyen.

En el capitol 3 es descriu el preprocessat utilitzat pel control de qualitat que consta de dues parts: el filtratge per eliminar dades amb poca qualitat i outliers i la normalització per eliminar la variabilitat causada pel procés d'obtenció d'aquestes.

En el capitol 4 s'implementa l'algorisme anomenat DIB-C per trobar els clusters que contindran els gens amb igual patró.

En el capitol 5 es calcula el nombre òptim de clusters a través d'un estadístic que anomenem Z-score i que es basa en el concepte d'entropia.

En el capitol 6 s'explica l'anàlisi d'enriquiment que ens permetrà trobar la funcionalitat biològica dels grups de gens trobats gràcies a l'algoritme DIB-C i que donarà la informació necessaria al biòleg per poder interpretar els resultats.

En el capitol 7 es mostren els resultats obtinguts tan de l'aplicació de l'algorisme DIB-C com de l'anàlisi d'enriquiment.

En el capitol 8 s'exposen les conclusions i es plantegen futurs procediments per amplificar l'anàlisi de les dades del llevat.

La Figura 1.5 mostra les fases per realitzar un projecte amb la tècnica dels microarrays

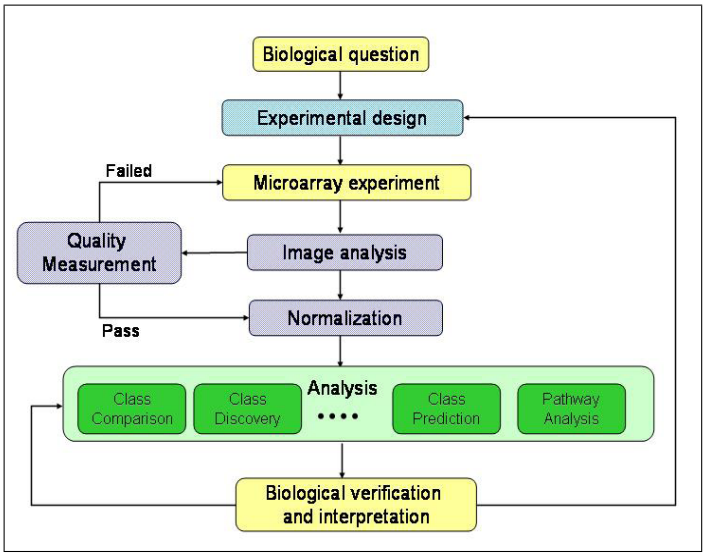


Figura 1.5: Fases en un projecte de microarrays

Capítol 2

Dades

Les dades facilitades per l'investigador corresponen a dades d'expressió de 6144 gens de l'organisme del llevat obtingudes mitjançant la tècnica dels macroarrays.

A cada un d'aquests gens se'ls ha aplicat estrès tèrmic a l'inici de l'experiment i s'ha mesurat la seva expressió en 12 instants de temps diferents (t0, t3, t6, t9, t12, t15, t18, t21, t25, t30, t45 i t60).

L'estrès tèrmic (*heat shock*) consisteix en aplicar a una cèl·lula una temperatura superior a la temperatura normal de l'organisme al qual pertany. D'aquesta manera s'aconsegueix que els gens s'expressin per tal de fer que la cèl·lula no mori.

Per tal de reduir la incertesa, augmentar la precisió i obtenir una potència de test suficient, l'experiment s'ha realitzat 8 cops obtenint 8 rèpliques per cada gen. Aquestes rèpliques s'han obtingut de organismes independents i s'han codificat de la següent manera: A, B, C, D, E i F, G i H.

Per tal de poder comparar el nivell d'intensitat a una escala més petita i per tal de fer que les dades siguin més simètriques i s'aproximin més a una normal, s'ha aplicat la transformació més usual: el logaritme en base dos.

L'esquema de les dades es mostra a la Figura 2.1

A més a més de la lectura d'expressió, per a cada gen tenim la següent informació:

- Position: posició del gen a la membrana del macroarray.
- Length: llargada de la seqüència del gen.

- ORF-name: (Open Reading Frame) nom del codó inicial de la seqüència que codifica la proteïna.
- status: si és un gen o un control
- Gene name: nom del gen
- Molecular function: funció molecular del gen que pot ser desconeguda
- Biological Process: procés biològic en el què actua el gen que pot ser desconegut.

Aquesta informació addicional ens servirà tant per la realització del projecte com per a la interpretació més acurada dels resultats.

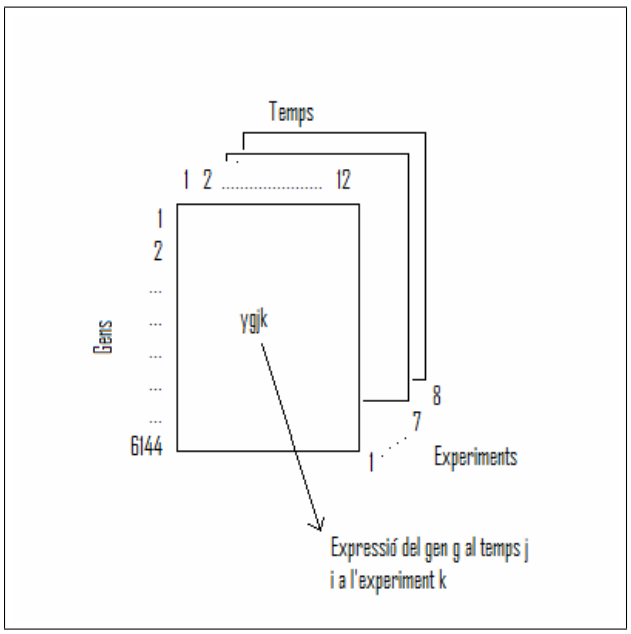


Figura 2.1: Esquema de les dades

Capítol 3

Preprocessat

3.1 Control de qualitat

Els experiments d'arrays produeixen un conjunt d'imatges que són llegides per un scanner que mesura la intensitat del senyal que després es transformarà en un valor numèric.

Al analitzar la qualitat de les lectures el que es pretén és reduir la variabilitat tècnica produïda pel procés de realització de l'array. Hi ha diverses causes de variabilitat tècnica, algunes d'elles poden ser:

- Manca d'eficiència en el procés de marcatge i en la detecció de la radiació.
- Mal funcionament de l'scanner.
- Error d'etiquetatge.
- Diferències en la quantitat de RNA inicial.
- Problemes de manipulació.
- Altres.

El preprocessat de les dades normalment consta de dues fases: el filtratge i la normalització.

3.2 Filtratge

En la lectura dels macroarrays, hi ha moltes proves associades a gens que no estan diferencialment expressats i que per tant obtenen una lectura falsa positiva que pot comportar un augment del biaix en les dades o una lectura que és només soroll i que ens aporta molt poca informació. El filtratge de les dades consisteix en eliminar aquestes lectures que no són bones.

Tot i això, no hi ha un mètode ni un moment concret per realitzar el filtratge de les dades i moltes vegades simplement es mantenen tots els gens en l'anàlisi.

De la informació obtinguda del centre d'obtenció de dades, s'ha considerat que una lectura inferior a 0.1 es podria considerar com a nul·la. Atès que un mateix gen és analitzat al llarg de 12 instants de temps, el fet d'excloure'l per un valor inferior a 0.1 s'ha considerat massa restrictiu i s'ha considerat convenient mantenir el gen amb un minim de 9 instants superiors a 0.1, sempre i quan això s'assolís en el 50% de les rèpliques, és a dir 4. Tot i això, s'han substituït la resta de lectures no filtrades iguals o inferiors a 0.1 per NA's per tal que no es tinguin en compte quan es calculen les mitjanes però que ens aportin informació quan apliquem l'algorisme.

A aquesta conclusió s'ha arribat empíricament després de parlar amb l'investigador i realitzar una seguit de proves.

Així, a partir dels 6144 gens, finalment, s'han mantingut 5833 gens.

3.3 Normalització

El procés de normalització consisteix en eliminar la variabilitat tècnica. Per tal d'aconseguir-ho, s'ha assumit que la majoria de gens no canvien de forma diferencial al llarg del temps i que el nombre de gens que augmenten o disminueixen l'expressió és similar a la mitjana dels 8 experiments per a cada gen i instant de temps.

Atès que en la literatura només existeixen mètodes de normalització basats en informació que no teníem [16], hem decidit adaptar un dels mètodes més utilitzats per normalitzar microarrays.

Com vam explicar a la introducció, hi ha un tipus de microarray anomenat de dos colors amb el que es compara l'expressió dels gens de dues mostres codificades una de vermell (R) i una de verd (G) dins de un únic array. Una primera possibilitat per detectar anomalies seria realitzant un gràfic de R versus G per veure si els punts queden al voltant de la diagonal, si no, voldrà dir que hi ha variabilitat tècnica i s'haurà de solucionar.

Tot i així, el que normalment es fa servir per detectar aquesta variabilitat, és una transformació d'aquest mètode que consisteix en construir un gràfic MA que equival a una rotació de 45º en sentit antihorari del gràfic anterior i que es calcula a partir dels següents valors:

$$M = \log_2 \frac{R}{G} = \log_2 R - \log_2 G \tag{3.1}$$

$$A = \log_2 \sqrt{RG} = \frac{1}{2} \times (\log_2 R + \log_2 G) \tag{3.2}$$

on M és el logaritme de la raó d'un respecte l'altre i A és la mitjana del logaritme i ve a ser una mesura de l'intensitat total.

En el nostre cas, al no tenir dos colors d'intensitat, s'ha suposat que R és el valor de l'expressió per a un gen, en un instant de temps i en un experiment determinat i G és el valor de la mitjana dels 8 experiments per cada gen i per cada instant de temps.

Una vegada realitzada aquesta transformació, i abans de normalitzar, es pot realitzar un gràfic MA per visualitzar la raó d'intensitat. Els gràfics MA col·loquen la M a l'eix de les y's i la A al de les x's. Si les dades són prou bones veurem que els punts es distribueixen a la línia horitzontal del 0, és a dir, s'espera que la raó sigui 1 ($\log(1) = 0$). A l'Annex A es poden veure tant els gràfics MA com els boxplots abans de normalitzar per a cada instant de temps i rèplica. Es pot observar com, en general, els boxplots tenen molta dispersió tant dins de cada rèplica com entre cada una de elles i el núvol de punts que es forma als MA plots es troba per sota de 0 i en molts casos té forma corbada, per tant caldrà normalitzar les dades.

Per normalitzar les dades s'ha aplicat una correcció basada en el càlcul de la regressió suavitzada mitjançant el mètode Lowess (locally weighted scatterplot smooth; Yang et al 2002 [17]). Aquest mètode consisteix en ajustar una regressió local ponderada i calcular un factor de correcció per cada punt, donant més pes als punts d'intensitats més similars. Així, si els punts amb menys intensitat es desvien més del 0 que els de més intensitat seran més modificats.

La regressió Lowess consisteix en calcular per a cada valor x_0, un valor $c(x_0)$ a partir dels següents passos:

1. S'identifiquen els k veïns més propers de x_0 que anomenem veïnatge $N(x_0)$

2. Es calcula la distancia a x_0 del punt més llunyà que està dins del veïnatge $N(x_0)$ i es representa per $\Delta(x_0)$.

3. Per a cada punt x_i del veïnatge $N(x_0)$ es calculen els pesos w_i fent servir la funció de pes tri-cúbica definida per:

$$W(t, x_0) = \left[1 - \left(\frac{|t - x_0|}{\Delta(x_0)}\right)^3\right]^3 \quad sempre \quad que \quad |t - x_0| < (x_0)$$

4. Es defineix el suavitzador $c(x_0)$ com a valor ajustat en x_0 de la regressió ponderada de y versus x en el veïnatge $N(x_0)$, fent servir els pesos definits al pas tres.

Una vegada calculat el valor de la regressió ponderada pels valors d'A (c(A)), es corregeix el gràfic MA a partir de:

$$M_{norm} = M - c(A)$$

on $c(A)$ és la funció depenent de la intensitat.

Finalment, a partir del gràfic MA transformat, és immediat tornar a obtenir els valors de les intensitats ja en escala normalitzada. A l'Annex B es mostren els gràfics MA i els Boxplots per les dades normalitzades. Si observem aquests gràfics, veiem una gran millora amb relació als gràfics de l'Annex A ja que ara els boxplots tenen molta menys dispersió i són molt més semblants entre rèpliques i els MA plots estan centrats a 0 i s'ha eliminat la corbatura.

Capítol 4

Algorisme DIB-C

Una vegada realitzat el preprocessat i la normalització de les dades d'expressió gènica corresponents a l'organisme del llevat, s'ha implementat el següent algorisme per tal de trobar grups de gens amb igual comportament en diferents instants de temps.

4.1 Dades

Les dades obtingudes fins ara corresponen a m=8 rèpliques de n=5833 gens en p=12 instants de temps que estan representades en la següent matriu d'expressió:

$$Y = \{y_{gjk}\} \qquad \forall \quad g = 1, ..., n; \quad j = 1, ..., p; \quad k = 1, ..., m \tag{4.1}$$

4.2 Estadístic t moderat

Els experiments amb microarrays solen tenir poques rèpliques que, al donar pocs graus de llibertat, dificulten l'estimació de la variància per cada gen específic.

L'estadístic t moderat es basa en l'enfocament bayesià i combina la informació de tot l'array i de cada gen individualment per obtenir millors estimacions de l'error.[12][6][4]

Els problemes principals que ens podem trobar si féssim servir l'estadístic t clàssic són:

- L'estimació de la variància basada en pocs graus de llibertat pot ser poc fiable.
- Pot ser particularment problemàtic si el model proposat no és prou bo.
- Les variàncies subestimades ens porten a falsos positius i les sobreestimades ens fan perdre potència per detectar gens diferencialment expressats.

Suposem que tenim, en m arrays, una variable resposta normalitzada per un únic gen que denotem com $Y^T = (y_1, y_2, ..., y_m)$. Suposem que $E(Y) = X\alpha$ on X és la matriu de disseny i α és el vector de coeficients. També suposem que $Var(Y) = W\sigma^2$ on W és una matriu de pesos definida positiva. Els contrastos que ens interessa fer sobre els coeficients estan definits com $\beta = C^T\alpha$. Suposem que la hipòtesis pels contrastos individuals és $\beta_j = 0$.

L'ajust del model lineal a la resposta permet obtenir les estimacions dels coeficients $\hat{\alpha}$, l'estimador s^2 de σ^2 i les matrius de covariances estimades $Var(\hat{\alpha}) = Vs^2$ on V és una matriu definida positiva no depenent de s^2. L'estimació dels contrasts és $\beta = C^T\hat{\alpha}$ amb una matriu de covariances estimada $Var(\hat{\beta}) = C^TVCs^2$.

L'estadístic t clàssic es calcularia de la següent manera: $t_j = \frac{\hat{\beta}}{(s\sqrt{v_j})}$ on v_j és l'element j-èssim de la diagonal de C^TVC.

Per desenvolupar l'estadístic t moderat s'han de fer assumpcions prèvies sobre la distribució de σ^2 i β_j:

$$\frac{1}{\sigma^2} \sim \frac{1}{d_0 s_0^2}\chi^2_{d_0} \quad i \quad \beta_j \mid \sigma^2, \quad \beta_j \neq 0 \sim N(0, v_{0j}\sigma^2)$$

Definim $\tilde{s}^2$ com a mitjana posterior de σ^2 donat s^2. L'estadístic t moderat és:

$$\tilde{t}_j = \frac{\hat{\beta}}{(\tilde{s}\sqrt{v_j})} \tag{4.2}$$

Per al càlcul d'aquest estadístic s'ha fet servir la funció *ebayes* del paquet *limma* d'R que ens dóna un estimador global de la variància (s_0^2) basat en la variància de tots els gens (s_g^2), és a dir, enlloc d'obtenir una estimació per a la variància de cada gen, s'obté una mitjana estimada de la variància de cada gen (s_g^2) i de la global (s_0^2).

4.3 Algorisme

L'algorisme DIB-C representa cada gen amb un patró que identifica les zones de creixement i decreixement de la corba. S'implementa a partir de les dades d'expressió Y i consta dels passos següents:

1. **Anàlisi de la diferència de primer ordre** L'anàlisi de la diferència de primer ordre s'obté calculant un estadístic t moderat per a cada parell de temps j i $j+1$

   La fórmula per calcular aquest estadístic és:

   $$Y^{(1)} = \{y_{gj}^{(1)}\}_{n\times(p-1)} \qquad on \quad y_{gj}^{(1)} = \frac{\beta_{gj}}{\tilde{s}_g\sqrt{v_{gj}}} \tag{4.3}$$

   $$\forall \quad g = 1, ..., n; \quad j = 1, ..., p-1$$

   $y_{gj}^{(1)}$ és l'estadístic t moderat entre els temps j i $j+1$;
   β_{gj} és la mitjana de la diferència entre els temps j i $j+1$ ($\bar{Y}_{gj} - \bar{Y}_{gj+1}$);
   $\tilde{s}_g$ és la mitjana de la desviació estàndard per a cada gen g a posteriori;
   v_{gj} és la variància conjunta del gen g als temps j i $j+1$ (per la diferència).

   Ara ja s'han tingut en compte els gens que tenen valors NA's per obtenir la informació conjunta necessària per calcular l'estadístic t moderat. Aquests gens amb valors NA's no ens permeten categoritzar la diferència de primer ordre en el pas següent, així que els filtrem. A partir d'ara tindrem $n = 5257$ gens.

2. **Matriu F de patró simbòlic** A partir de les diferències de primer ordre es construeix la següent matriu F de patrons categoritzant la matriu 4.3 en tres símbols segons si la diferència de primer ordre d'un instant de temps al següent és creixent (I), decreixent (D) o no significativa (N). La significació s'obté comparant les diferències amb valors de l'estadístic t d'Student.

   $$F = \{f_{gj}\}_{n\times(p-1)} \qquad \forall \quad g = 1, ..., n; \quad j = 1, ..., p-1$$

   $$f_{gj} = \begin{cases} I & si \quad y_{gj}^{(1)} > T(1 - \frac{\alpha_1}{2}; df_{gj}) \\ D & si \quad y_{gj}^{(1)} < -T(1 - \frac{\alpha_1}{2}; df_{gj}) \\ N & altrament \end{cases} \tag{4.4}$$

on T correspon a la distribució t d'Student amb $df_{gj} = d_o + d_g$ graus de llibertat; d_0 són els graus de llibertat obtinguts de la informació a priori de l'estadístic t moderat i d_g són els graus de llibertat obtinguts per cada gen.

3. **Anàlisi de la diferència de segon ordre** L'anàlisi de la diferència de segon ordre es basa en calcular la diferència de l'estadístic $y_{gj}^{(1)}$, per a valors consecutius, obtingut en el pas anterior:

$$Y^{(2)} = \{y_{gj}^{(2)}\}_{n\times(p-2)} \qquad on \quad y_{gj}^{(2)} = y_{g(j+1)}^{(1)} - y_{gj}^{(1)} \tag{4.5}$$

$$\forall \quad g = 1, ..., n; \quad j = 1, ..., p-2$$

4. **Matriu S de patró simbòlic** A partir de les diferències de segon ordre es construeix la següent matriu S de patrons categoritzant la matriu 4.5 en tres símbols segons si la diferència de segon ordre és convexa (V), còncava (A) o no significativa (N) d'una diferència de primer ordre a la següent.

$$S = \{s_{gj}\}_{n\times(p-2)} \qquad \forall \quad g = 1, ..., n; \quad j = 1, ..., p-2$$

$$s_{gj} = \begin{cases} V & si \quad y_{gj}^{(2)} > (1 - \frac{\alpha_2}{2}) \quad quantil \quad de \quad T' \\ A & si \quad y_{gj}^{(2)} < (\frac{\alpha_2}{2}) \quad quantil \quad de \quad T' \\ N & altrament \end{cases} \tag{4.6}$$

on la distribució de l'estadístic $Y^{(2)}$ (T'), atès que és difícil trobar la distribució de la diferència de dos t d'Students, s'ha obtingut mitjançant el càlcul dels seus quantils a través d'una simulació de Montecarlo. Els passos són els següents:

- Es creen dues mostres aleatòries de mida 10.000 d'una distribució t d'Student amb graus de llibertat $m - 1 = 7$.
- Es guarden els αth quantils de la diferència de les dues mostres.
- Es repeteix aquest procediment 1000 vegades.
- S'escull com a valor crític la mediana dels 1000 quantils.

5. **Matriu H de patró simbòlic combinat** Amb la creació de la matriu H s'obté, per cada a gen, una seqüència de $2p - 3 = 21$ lletres que ens indica a quin cluster correspon.

$$H_{n\times(2p-3)} = [F_{n\times(p-1)}|S_{n\times(p-2)}]$$

4.4 Output

El resultat de l'algorisme és un llistat de patrons que ens informa a quin cluster pertany cada gen.

Capítol 5

Càlcul del nombre òptim de clusters

El càlcul del nombre òptim de clusters s'ha fet a partir del càlcul d'un estadístic que anomenem Z-score.

Per calcular aquest estadístic és necessari fer servir la *Gene Ontology* (GO), una base de dades amb les característiques dels gens que volem associar als clusters.

A més, s'ha calculat l'estadístic anomenat d'informació mútua per a les dades experimentals, que ens permetrà mesurar la variabilitat de cada cluster obtingut.

5.1 Matriu d'anotacions de la GO

La *Gene Ontology* (GO) és un projecte de col·laboració que uneix un conjunt de bases de dades de diferents organismes. La GO està estructurada en dos parts: les ontologies i les anotacions.

Una ontologia proporciona un conjunt de termes de vocabulari que cobreixen un domini conceptual. Aquests termes han de tenir una definició rigorosa i han d'estar dins una estructura de relacions. Les tres ontologies de la GO són:

- **Funció molecular** (MF):descriu les activitats que succeeixen a nivell molecular sense especificar ni on ni quan ni en quin context succeeixen.
- **Procés biològic** (BP): descriu una sèrie de successos als que s'arriba a través de un conjunt ordenat de funcions moleculars
- **Component cel·lular** (CC): descriu parts de la cèl·lula que poden ser una estructura anatòmica o un producte de gen.

Cada gen individual pot representar una o més funcions moleculars, ser usat en un o més processos biològics i aparèixer en un o més components cel·lulars

Les anotacions creen un link entre els gens coneguts i els termes de la GO que defineixen les seves funcions la qual cosa crea una xarxa jeràrquica. Cada anotació de la GO té un únic identificador numèric i un únic nom de terme. Cada terme és assignat a una de les tres ontologies.

Per tal de saber quin és el nombre òptim de clusters per dades gèniques s'ha de fer servir la *Gene Ontology* (GO),[7] concretament, la *Saccharomyces Genome Database* (SGD) que és la base de dades de biologia i genètica molecular per l'organisme del llevat (*Saccharomyces cervisiae*).

Per fer servir la GO en R s'ha de baixar el paquet *org.Sc.sgd.db* que conté les anotacions d'ORFs (*Open Reading Frame*) pel llevat. Aquest paquet conté un total de 71.501 anotacions.

Una vegada obtingudes aquestes anotacions, es comparen amb les anotacions de les dades originals (5833 gens). Els gens originals que no tenen la seva corresponent anotació a la GO són eliminats i ens quedem amb un total de 5326 gens. Les anotacions de la GO que no corresponen a cap dels gens també s'eliminen quedant 3802 anotacions. Després d'aquest filtratge es crea la matriu booleana de gens per atributs on 1 ens indica que el gen té aquell atribut de la GO i 0 ens indica que no el té. Aquesta matriu també es filtra traient els gens que tenen algun valor NA. D'aquesta manera, finalment obtenim una matriu de 5257 files per 3802 columnes.

5.2 Càlcul de la Informació Mútua (MI) real i aleatòria

Mitjançant la matriu d'anotacions de la GO es crea una taula de contingència per a cada cluster que ens permetrà calcular la Informació Mútua. Aquesta taula contindrà dues columnes (una pel 0 i l'altre per 1) i tantes files com gens hi hagi en el cluster.

La MI és un estadístic no paramètric robust que utilitzarem per identificar les associacions entre els gens i les seves expressions i que es calcula mitjançant el concepte d'entropia. [3][13][11]

L'entropia és una mesura d'incertesa al predir el valor futur d'una variable aleatòria. La seva fórmula per una variable aleatòria discreta X és:

$$H(X) = -\sum_{x \epsilon X} \rho_X(x) \log_2 \rho_X(x)$$

I la seva estimació si suposem que $\rho_X(x) = P(X = x)$ i que $\hat{P}(X = i) = \frac{n_i}{n}$ és:

$$\hat{H}(x) = -\sum_{i=1}^{|\chi|} \frac{n_i}{n} \log_2 \left(\frac{n_i}{n}\right)$$

on $|\chi|$ és la cardinalitat de X ; n_i és el nombre d'observacions amb $x = i$ i n és el nombre total d'observacions.

La MI s'ha utilitzat per quantificar la informació que té una variable aleatòria (X) sobre una altra variable aleatòria (Y), és a dir, mesura la reducció d'incertesa de X coneixent Y. La fórmula general per calcular la MI entre dues variables aleatòries és:

$$MI(X,Y) = \sum_{x \epsilon X} \sum_{y \epsilon Y} \rho_{XY}(x,y) \log_2 \left(\frac{\rho_{XY}(x,y)}{\rho_X(x)\rho_Y(y)}\right) = H(X) - H(X|Y)$$

on $\rho_{XY}(x,y)$ és la probabilitat conjunta de X i Y; $H(X|Y)$ és l'entropia condicionada ($H(X|Y) = H(X,Y) - H(Y)$) on $H(X,Y)$ és l'entropia conjunta de X i Y

En el nostre cas, per calcular la MI, la variable aleatòria Y correspondria a la variable que ens diu si el gen té la anotació de la GO i per tant prendria valor 0 ó 1 i la variable aleatòria X correspondria als gens del cluster.

L'estimació de la MI és:

$$\widehat{MI}(X,Y) = \sum_{i=1}^{c}\sum_{j=1}^{2}\frac{n_{ij}}{n}\log_2\left(\frac{n_{ij}}{n}\right) - \sum_{i=1}^{c}\frac{n_{i\cdot}}{n}\log_2\left(\frac{n_{i\cdot}}{n}\right) - \sum_{j=1}^{2}\frac{n_{\cdot j}}{n}\log_2\left(\frac{n_{\cdot j}}{n}\right) \tag{5.1}$$

on c és la mida del cluster ; n_{ij} és el nombre d'anotacions que té ($j = 1$) o no té ($j = 2$) cada gen i del cluster ; $n_{i\cdot}$ és el nombre total d'anotacions que tenim per cada gen i del cluster, és a dir, 3802 ; $n_{\cdot j}$ és el nombre d'anotacions que tenen ($j = 1$) o no tenen ($j = 2$) el total de gens del cluster i n és el nombre total d'observacions.

A partir de l'equació 5.1 i dels clusters obtinguts al aplicar l'algorisme, es calcula la MIreal. Una vegada obtingut aquest valor i per saber fins a quin punt és elevat i per tant indica que hi ha molta dispersió, es calcula un valor MIrand. La MIrand es calcula mitjançant l'assignació aleatòria dels gens als clusters tantes vegades com faci falta fins que la distribució de la MI no varii. En aquest cas s'han fet servir 2000 repeticions. Així s'obté una mitjana i una desviació estàndard de la distribució.

5.3 Càlcul del Z-score

El Z-score es defineix com a la distància estandaritzada entre la MIreal obtinguda de l'algorisme de clustering i la MIrand obtinguda d'assignar els gens als clusters a l'atzar. Quant més gran sigui el Z-score millor és el resultat del clustering ja que dista més de la distribució aleatòria d'aquest.

El càlcul del Z-score es fa mitjançant la fórmula:

$$Zscore = \frac{E(MIreal - MIrand)}{\sqrt{Var(MIrand)}} \tag{5.2}$$

5.4 Escollir el màxim Z-score

Per tal d'escollir un bon nombre de clusters s'ha realitzat el càlcul del Z-score per diferents valors dels llindars de l'algorisme (α_1 i α_2). Els valors que s'han provat són:

- $\alpha_1 = \{0.001, 0.005, 0.01, 0.05\}$
- $\alpha_2 = \{0.001, 0.005, 0.01, 0.05\}$

El gràfic 5.1 mostra els resultats d'aplicar l'algorisme i realitzar el càlcul del Z-score per les 16 combinacions de valors dels llindars. Es pot observar que el nombre òptim de clusters a trobar és de 282 obtenint un Z-score màxim de 1.7708486. Aquest Z-score s'obté pels valors de llindar $\alpha_1 = 0.005$ i $\alpha_2 = 0.05$.

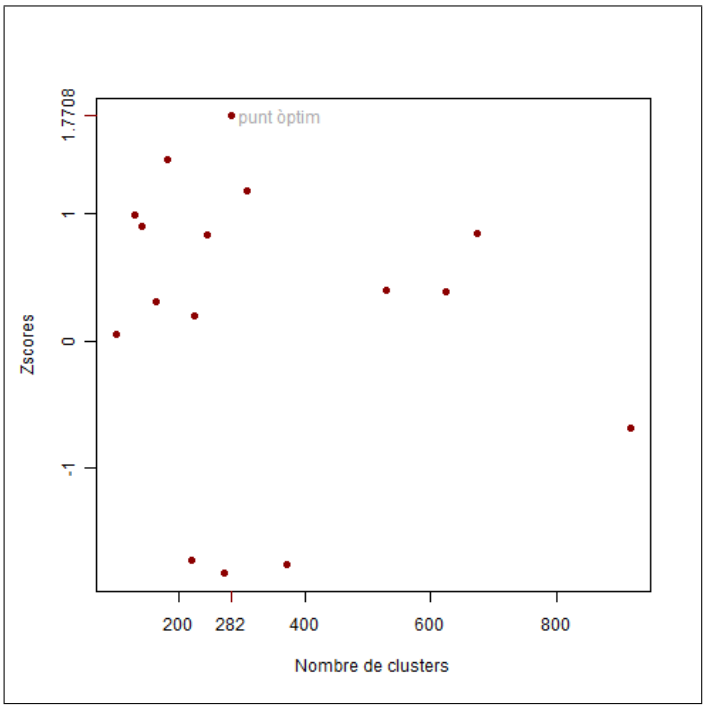


Figura 5.1: Z-scores de les dades

Capítol 6

Anàlisis d'enriquiment

Una vegada obtinguts els grups de gens que es comporten de manera similar al llarg del temps sota estrès tèrmic, volem trobar la funcionalitat biològica dels gens dels clusters mitjançant la GO, que com vam veure al capítol 5 és una base de dades que ens permet organitzar els gens en diferents ontologies.

L'anàlisi d'enriquiment (*Enrichment Analysis* (EA))[15][14] consisteix en trobar l'associació estadística entre els termes de l'ontologia i els gens diferencialment expressats i intenta revelar la significació biològica que hi ha darrera de les dades. La cerca de termes enriquits pot revelar connexions en els processos biològics que ajuden als biòlegs a construir hipòtesi.

Per a realitzar l'EA s'ha d'assumir que els termes de la GO són independents entre si i que els gens han d'estar anotats en més de una classificació biològica.

Normalment l'aplicació de l'EA es fa a través de taules de contingència. En el nostre cas em fet servir la funció de R *hyperGTest* que es troba al paquet *GOstats* i que associa les categories amb les descripcions dels gens mitjançant un test hipergeomètric. Així s'obtenen uns p-valors que ens indiquen si cada terme en una ontologia està sobre o sota representat entre el conjunt de gens especificats. El que fa aquesta funció de R és el següent:

Considerem que tenim N gens al chip. D'aquests N gens, M pertanyen a la categoria A de la GO i N-M no pertanyen a la categoria A. Es seleccionen K gens del total de N i s'assignen a una classe determinada. x gens dels K gens seleccionats estaran a la categoria A.

Les hipòtesi estadístiques a contrastar seran:

H_0:La categoria A de la GO està igual representada en el chip que a la classe de gens diferencialment expressats.

H_1:La categoria A de la GO està més (o menys) representada en el chip que a la classe de gens diferencialment expressats.

Volem saber quina és probabilitat de que hi hagi exactament x gens pertanyents a la categoria A. La distribució hipergeomètrica modela la probabilitat de que una categoria succeeixi x vegades per atzar de una llista de gens diferencialment expressats. Aquesta distribució té els paràmetres (N,M,K)

$$P(X = x) = \frac{\binom{M}{x}}{\binom{N-M}{K-x}} \binom{N}{K} \tag{6.1}$$

Llavors, sota la hipòtesi nul·la, el p-valor de tenir x gens o més a la categoria A es calcularà amb la fórmula:

$$p - value = P(X \geq x | H_0) = \sum_{k=x}^{K} \frac{\binom{M}{x}}{\binom{N-M}{K-x}} \binom{N}{K} \tag{6.2}$$

Un p-valor petit voldrà dir que els termes de la GO estan sobre-representats.

Capítol 7

Resultats

7.1 Resultats de l'algorisme DIB-C

De l'execució de l'algorisme s'han obtingut 282 clusters. Cada un d'aquests clusters conté els gens que han obtingut el mateix patró al llarg dels 12 instants de temps.

Hi ha 200 clusters que contenen només 1 ó 2 gens i que l'investigador no ha volgut que reassignessim a algun cluster semblant atès que volia fer l'anàlisi per separat.

Hi ha també un cluster que conté el 64.4% dels gens (3384 gens) i que és el que correspon al patró nul ((N,N,N,N,N,N,N,N,N,N,N)(N,N,N,N,N,N,N,N,N,N,N)), és a dir, conté els gens que no s'han expressat de forma diferencial al llarg del temps. Així veiem que aquest algorisme no només ens particiona les dades formant grups sinó que també ens està filtrant els gens que no s'han expressat significativament en cap instant de temps.

L'algorisme, doncs, fa servir la discretització conceptual (creixent, decreixent, còncau, convex, sense canvi) per definir un patró. Això fa que cada cluster sigui interpretable.

A l'annex C es mostren dos gràfics per a cada cluster.

El primer gràfic correspon a la mitjana de tots els gens del cluster i de les 8 rèpliques per a cada instant de temps indicant també la desviació típica per veure quanta dispersió hi ha dins del cluster. Aquest gràfic ens serveix per mirar el comportament del cluster i per això està representat en la seva escala.

El segon gràfic correspon a la mitjana de les 8 rèpliques de cada gen del cluster per cada instant de temps. A sota del gràfic està indicat el patró simbòlic i el nombre de gens que conté el cluster. Aquest gràfic ens serveix per fer les comparacions entre clusters i per això està en l'escala on el límit superior el marca el gen de l'array que s'expressa més en mitjana i el límit inferior el que s'expressa menys.

Com es pot observar en aquests gràfics, hi ha clusters que tenen molta dispersió. Això passa perquè hi ha gens que tenen valors d'expressiómolt alts. Per això, al fer els gràfics no s'ha inclòs els gens que passaven d'una mitjana d'expressió de les 8 rèpliques de 10.000 considerant els gens no graficats com a gens rars. Els clusters 65 i 73 contenen només un gen cada un i és un gen rar, per això aquests dos clusters no estan graficats a l'annex C

Aquests gens rars, un total de 22, s'han graficat a part i es mostren a l'annex D. Aquest gràfic correspon a la mitjana de les rèpliques de cada gen rar per cada instant de temps. A sobre del gràfic s'indica a quin cluster pertany el gen i a sota es mostra el seu identificador.

7.2 Resultats de l'anàlisi d'enriquiment

Per a realitzar l'anàlisi d'enriquiment amb els clusters obtinguts s'han eliminat inicialment tots els clusters amb menys de 20 gens ja que l'EA no tindria validesa. També s'ha eliminat el cluster amb patró nul que contenia 3384 gens. Així ens hem quedat amb un total 14 clusters a analitzar.

Per aplicar l'EA a aquests 14 clusters s'ha decidit analitzar només la ontologia corresponent al procés biològic (BP) ja que és la que té més interès.

Per cadascun dels 14 clusters s'ha obtingut una taula amb el següent contingut:

- **Ontology**: Indica a quina de les tres ontologies pertany, en el nostre cas, només BP.
- **GOID**: Identificador de la GO del procés biològic.
- **Term**: Descripció del procés biològic.
- **Gene Names**: Nom comú de tots els gens que estan involucrats en el procés.

- **Size**: Nombre total de gens de l'univers que s'espera que tinguin la anotació.
- **Count**: Nombre total de gens del cluster que tenen l'anotació.
- **ExpCount**: Nombre total de gens del cluster que s'espera que tinguin la anotació.
- **OddsRatio**: Indica quantes vegades més apareix el terme de la GO en el cluster en comparació amb el total de l'array.
- **Pvalue**: P-valor associat a l'odds ratio.
- **OverUnder**: Indica si l'anotació està sobre (*over*) o sota (*under*) representada en el cluster.

Aquestes taules es poden trobar a l'adreça electrònica:
`http://estbioinfo.stat.ub.es/pubs/heatshock/supplementary.html`

A més, a l'annex E també es mostren les taules però sense les columnes corresponents a la Ontology, al Gene Name ni a OverUnder.

A l'annex F es troben tots els programes per a la implementació en R del projecte.

Capítol 8

Conclusions i procediments futurs

A partir de dades corresponents a la mesura de l'expressió de gens del llevat sotmesos a estrès tèrmic en 12 instants de temps i repetint aquest procediment vuit vegades obtenint 8 rèpliques de l'experiment, s'ha aplicat l'algorisme DIB-C i s'ha demostrat totes les avantatges que aporta aquesta nova metodologia de clustering en comparació amb els mètodes ja existents.

Les avantatges principals d'aquest mètode són:

- No necessita informació prèvia de les dades.
- Incorpora la restricció d'ordre entre instants de temps.
- Fa servir un estadístic t moderat que considera una estimació de la variància bayesiana empírica calculada a partir de les rèpliques. També fa servir la mitjana de les rèpliques.
- Genera uns gràfics de dos dimensions fàcilment interpretables a través de la discretització.

Aquest procediment és bo quan es tenen pocs instants de temps però si el nombre de instants és massa alt el temps d'execució de l'algorisme DIB-C augmenta exponencialment i per això serà més útil fer servir tècniques convencionals d'anàlisi de sèries temporals.

Malgrat això, trobar patrons que defineixen grups de gens pot no ser suficient per els biòlegs i per això s'ha d'analitzar cada cluster per separat mitjançant l'anàlisi d'enriquiment.

Així mateix, per poder aplicar l'EA es necessita tenir una mida mínima de cluster, per això aquest anàlisi només s'ha aplicat a 14 clusters.

De cara a procediments futurs, s'hauria de plantejar a l'investigador si l'interessaria reassignar els 200 clusters que tenen 1 ó 2 gens a altres clusters propers que poden tenir un patró semblant.

Potser també seria interessant fer l'anàlisi d'enriquiment per a les tres ontologies de la GO ja que de moment només s'ha analitzat una d'elles, els processos biològics.

Bibliografia

[1] TM. Alberola, J. García-Martínez, O. Antúnez, L. Viladevall, A. Barceló, J. Ariño, and JE. Pérez-Ortín. A new set of dna macrochips for the yeast saccharomyces cerevisiae: features and uses. *International Microbiology*, 3:199–206, 2004.

[2] R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Smyth, G. K., Limma: linear models for microarray data. Chapter 23.

[3] Francis D. Gibbons and Frederick P. Roth. Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Research*, 12:1574–1581, 2002.

[4] Anja von Heydebreck, Wolfgang Huber, and Robert Gentleman. Differential expression with the bioconductor project. *Bioconductor Project Working Papers*, 2004.

[5] Jihoon Kim and Ju Han Kim. Difference-based clustering of short time-course microarray data with replicates. *BMC Bioinformatics*, 8, 2007.

[6] Ingrid Lönnstedt and Terry Speed. Replicated microarray data. *Statistica Sinica*, 12:31–46, 2002.

[7] Gene Ontology. `http://www.geneontology.org/`.

[8] BioConductor Project. `http://www.bioconductor.org/`.

[9] Alex Sanchez and Francesc Carmona. Introducción a R y al análisis de datos con bioconductor. 2007.

[10] Alex Sanchez and M. Carme Ruiz de Villa. A tutorial review of microarray data analysis. 2008.

[11] TD. Schneider. Information theory primer. 1995.

[12] Gordon K. Smyth. Linear models and empirical bayes methods for assessing diferential expression in microarray experiments. *Statistica Sinica*, 3, 2004.

[13] Silke Szymczak, Angelo Nuzzo, Christian Fuchsberger, Daniel F Schwarz, Andreas Ziegler, Riccardo Bellazzi, and Bernd-Wolfgang Igl. Genetic association studies for gene expressions: permutation-based mutual information in a comparison with standard anova and as a novel approach for feature selection. *BMC Proceedings*, 2007.

[14] Ricardo ZN. Vêncio and Ilya Shmulevich. Baygo: Bayesian analysis of ontology term enrichment in microarray data. *BMC Bioinformatics*, 2006.

[15] Ricardo ZN. Vêncio and Ilya Shmulevich. Probcd: enrichment analysis accounting for categorization uncertainty. *BMC Bioinformatics*, 2007.

[16] Yi Xie, Adele Cutler, Bart Weimer, and Andrejus Parfionovas. Statistical methods for spot detection with macroarray data. In *35th Symposium on the Interface*, 2003.

[17] YH. Yang, S. Dudoit, P. Luu, DM. Lin, V. Peng, J. Ngai, and TP. Speed. Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30:4–15, 2002.

Apèndix A

Gràfics de les dades originals

A.1 BoxPlots de les dades originals sense normalitzar:

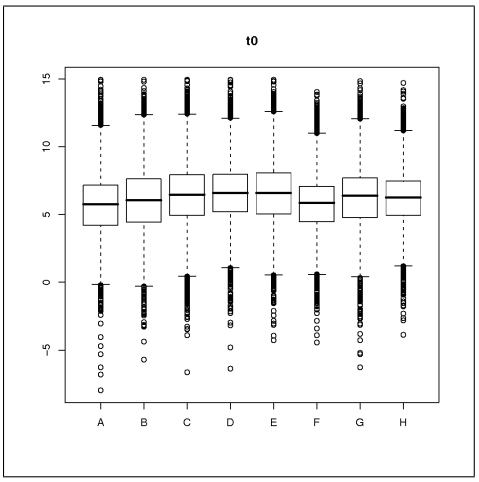


Figura A.1: Boxplot per les 8 rèpliques en l'instant de temps t0

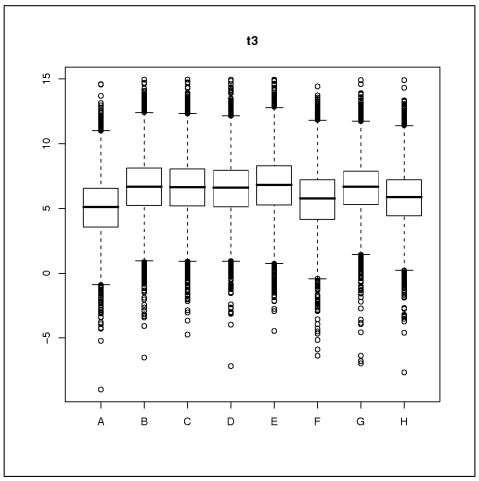


Figura A.2: Boxplot per les 8 rèpliques en l'instant de temps t3

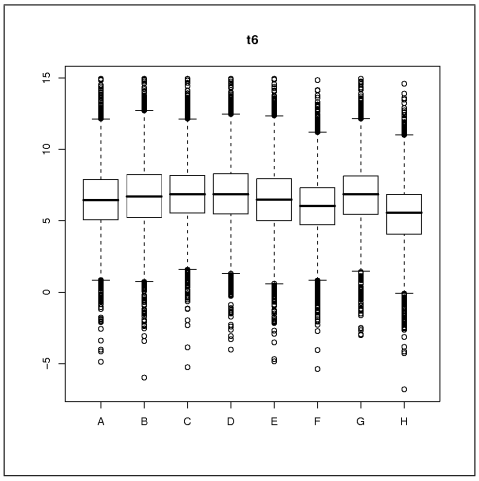


Figura A.3: Boxplot per les 8 rèpliques en l'instant de temps t6

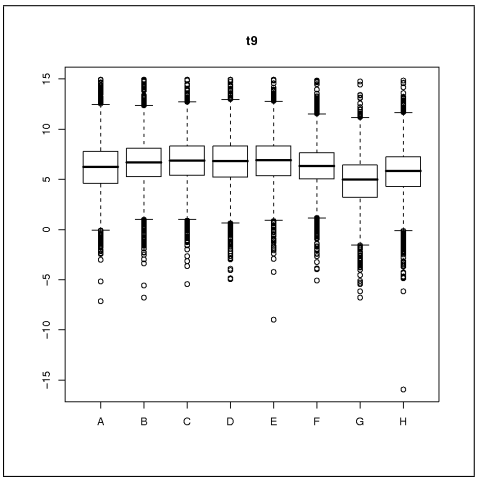


Figura A.4: Boxplot per les 8 rèpliques en l'instant de temps t9

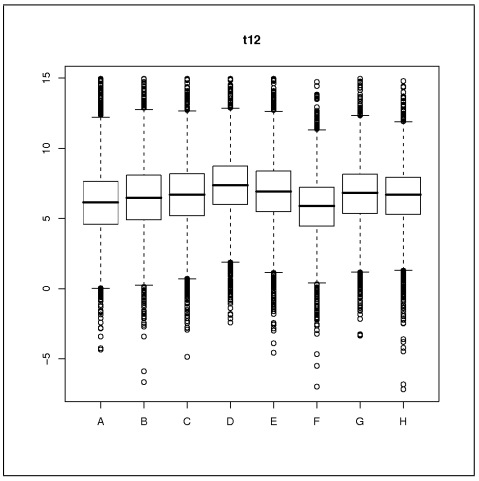


Figura A.5: Boxplot per les 8 rèpliques en l'instant de temps t12

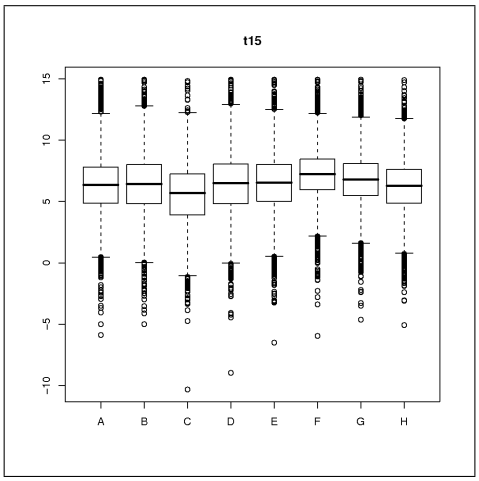


Figura A.6: Boxplot per les 8 rèpliques en l'instant de temps t15

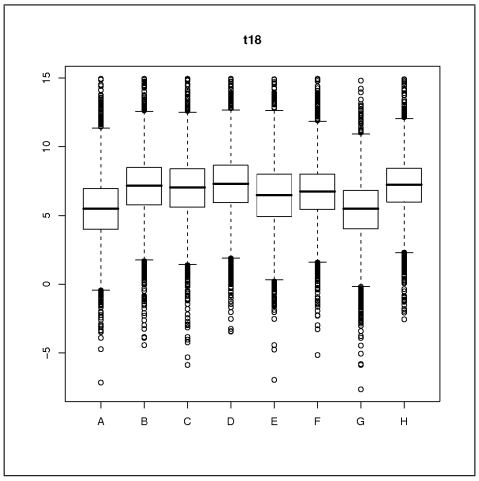


Figura A.7: Boxplot per les 8 rèpliques en l'instant de temps t18

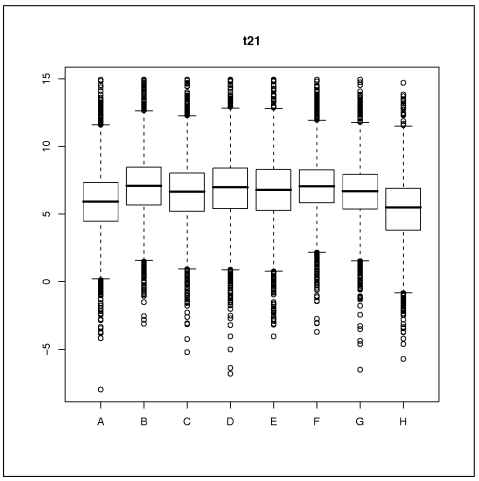


Figura A.8: Boxplot per les 8 rèpliques en l'instant de temps t21

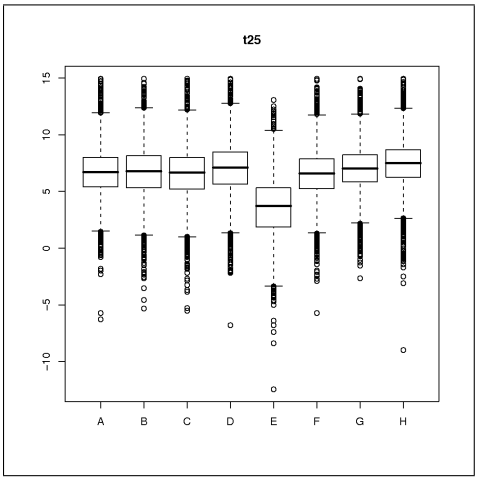


Figura A.9: Boxplot per les 8 rèpliques en l'instant de temps t25

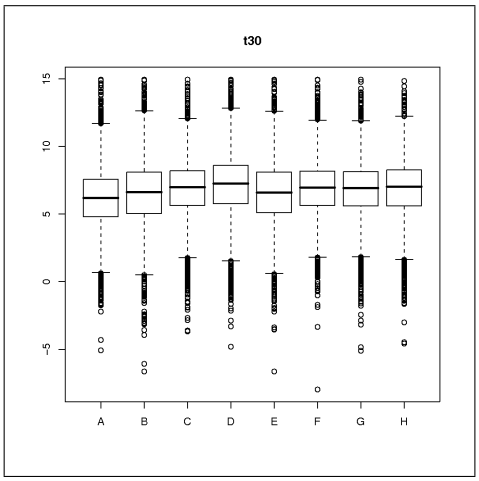


Figura A.10: Boxplot per les 8 rèpliques en l'instant de temps t30

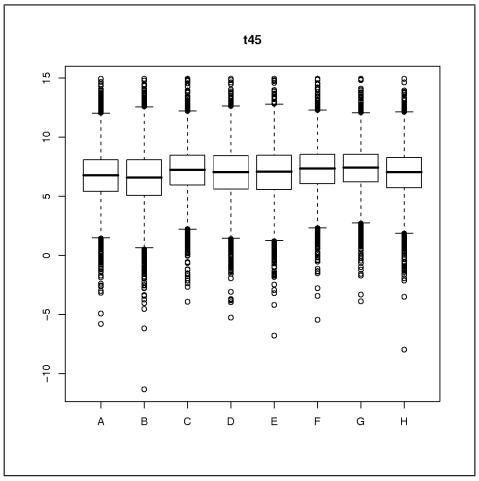


Figura A.11: Boxplot per les 8 rèpliques en l'instant de temps t45

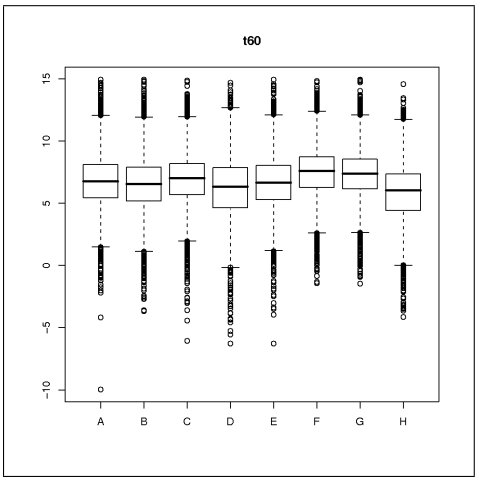


Figura A.12: Boxplot per les 8 rèpliques en l'instant de temps t60

A.2 Gràfics MA de les dades originals sense normalitzar:

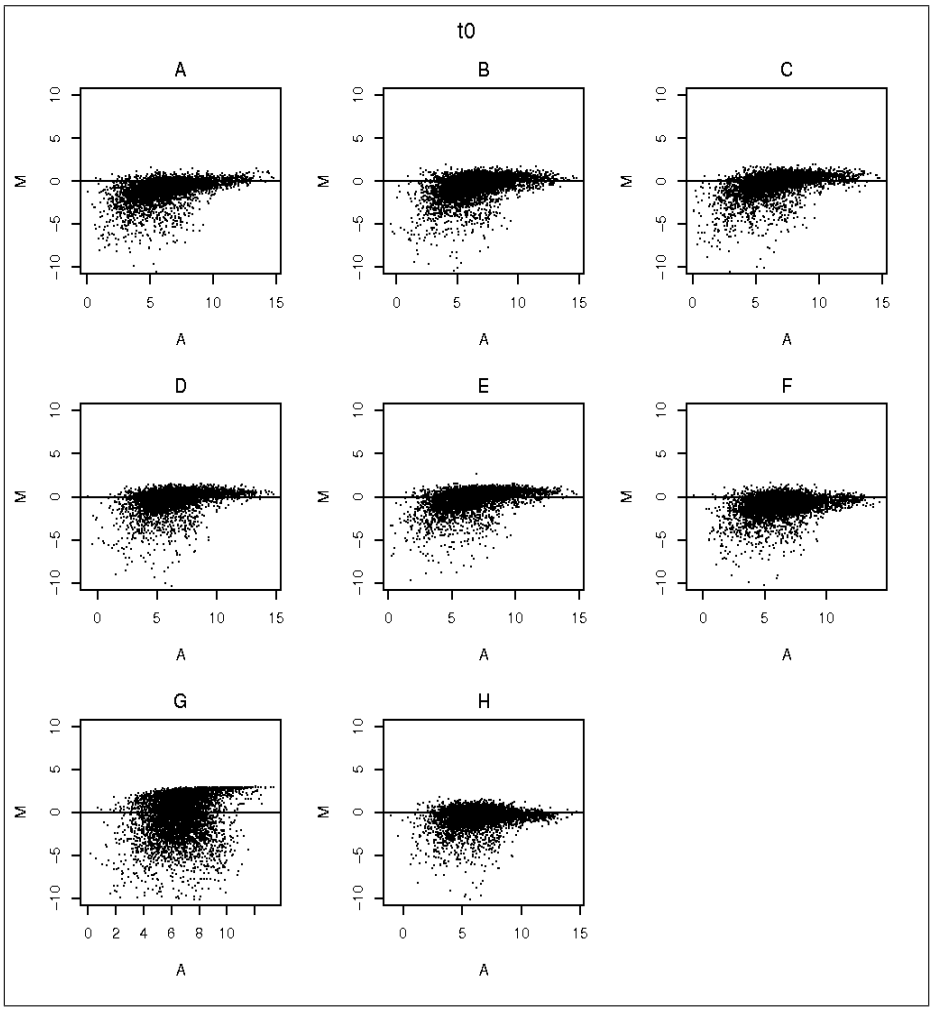


Figura A.13: Gràfics MA per cada rèplica en l'instant de temps t0

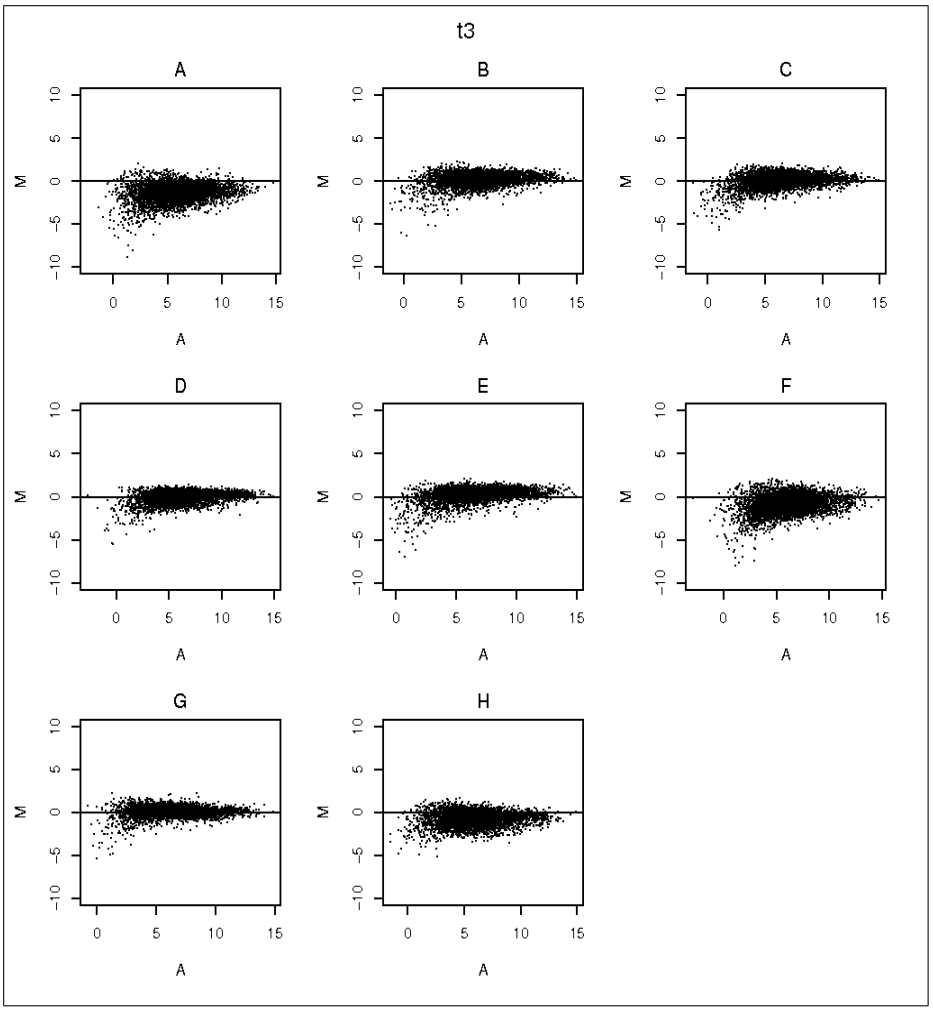


Figura A.14: Gràfics MA per cada rèplica en l'instant de temps t3

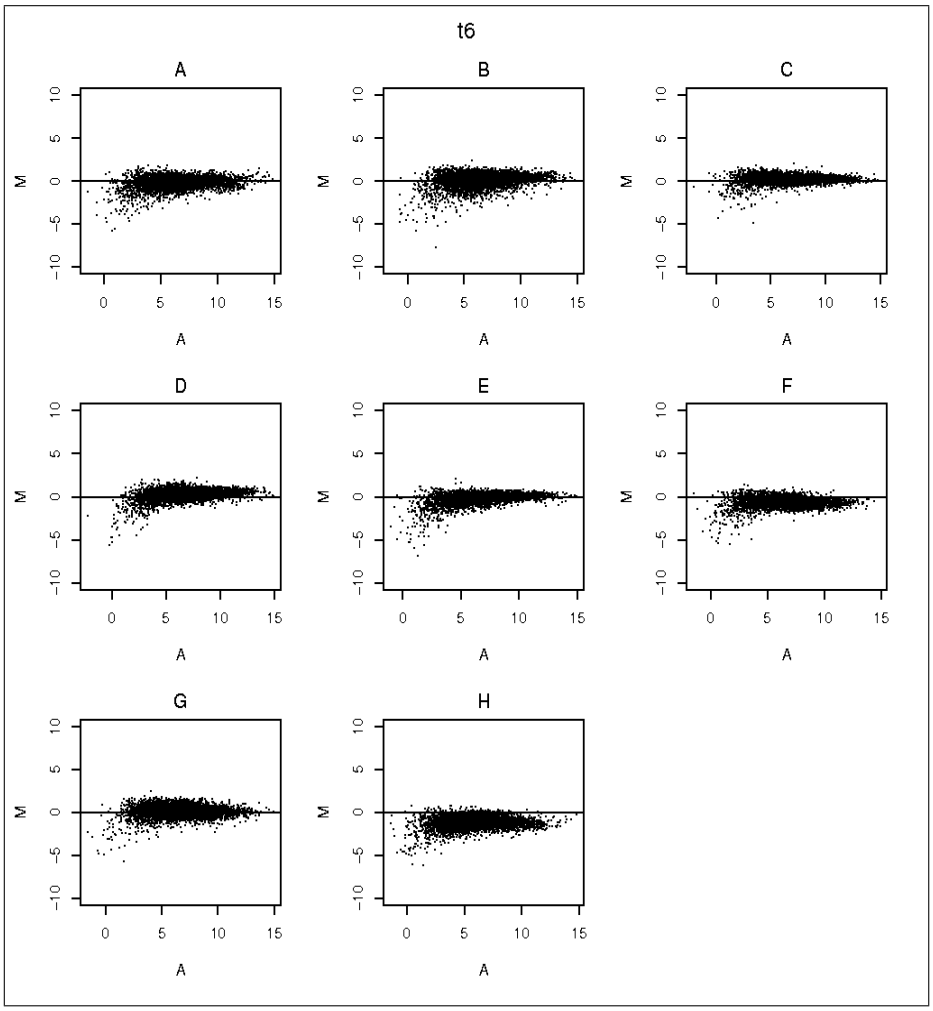


Figura A.15: Gràfics MA per cada rèplica en l'instant de temps t6

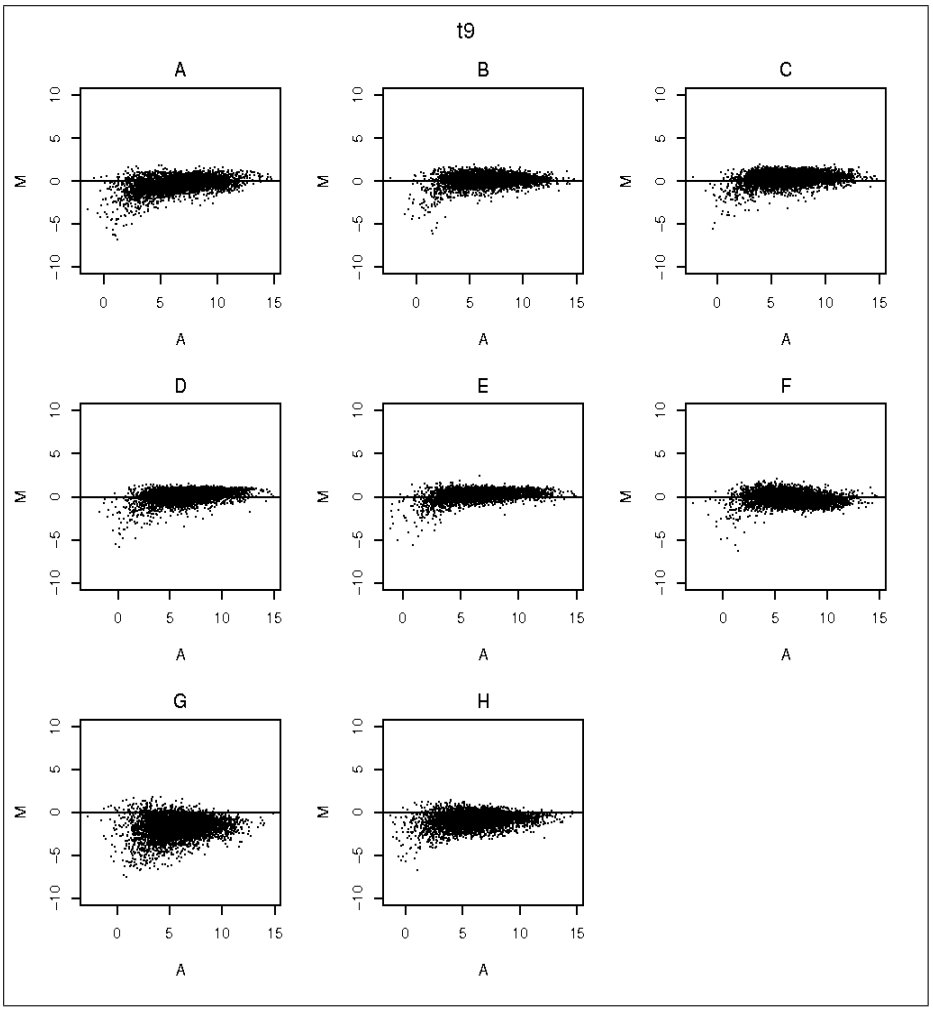


Figura A.16: Gràfics MA per cada rèplica en l'instant de temps t9

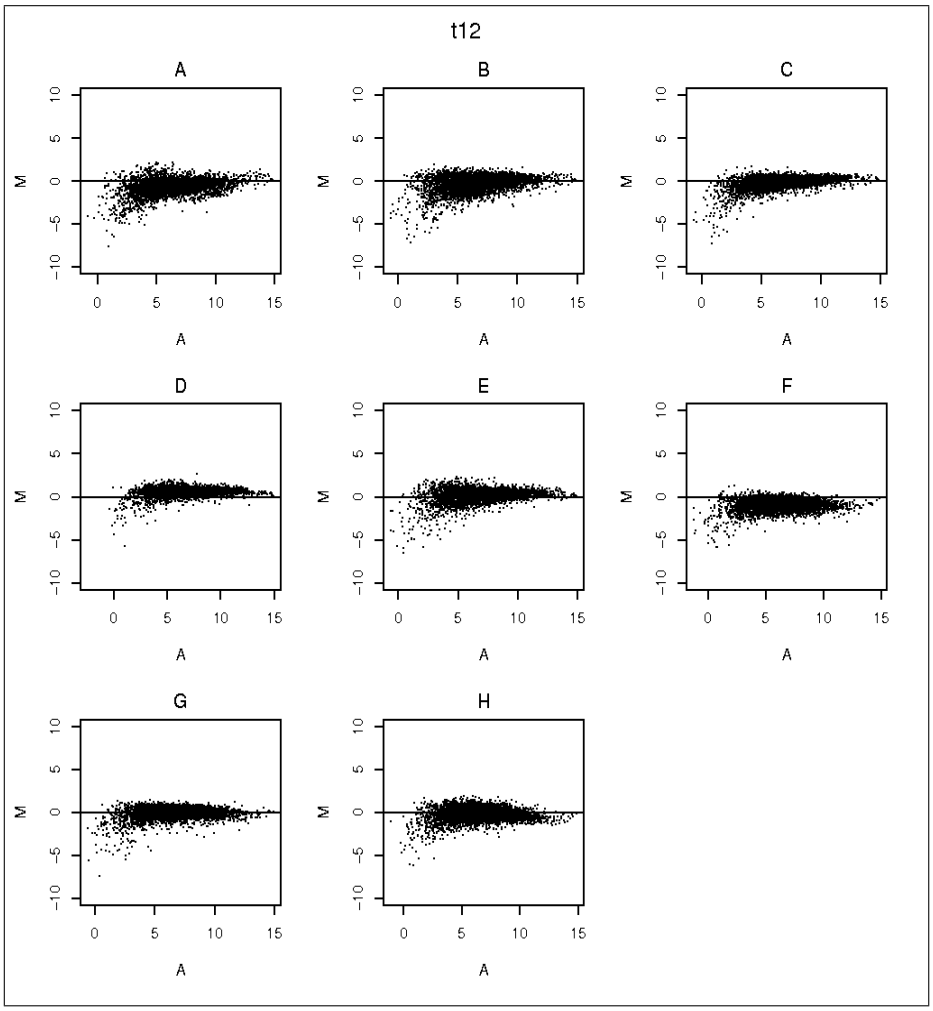


Figura A.17: Gràfics MA per cada rèplica en l'instant de temps t12

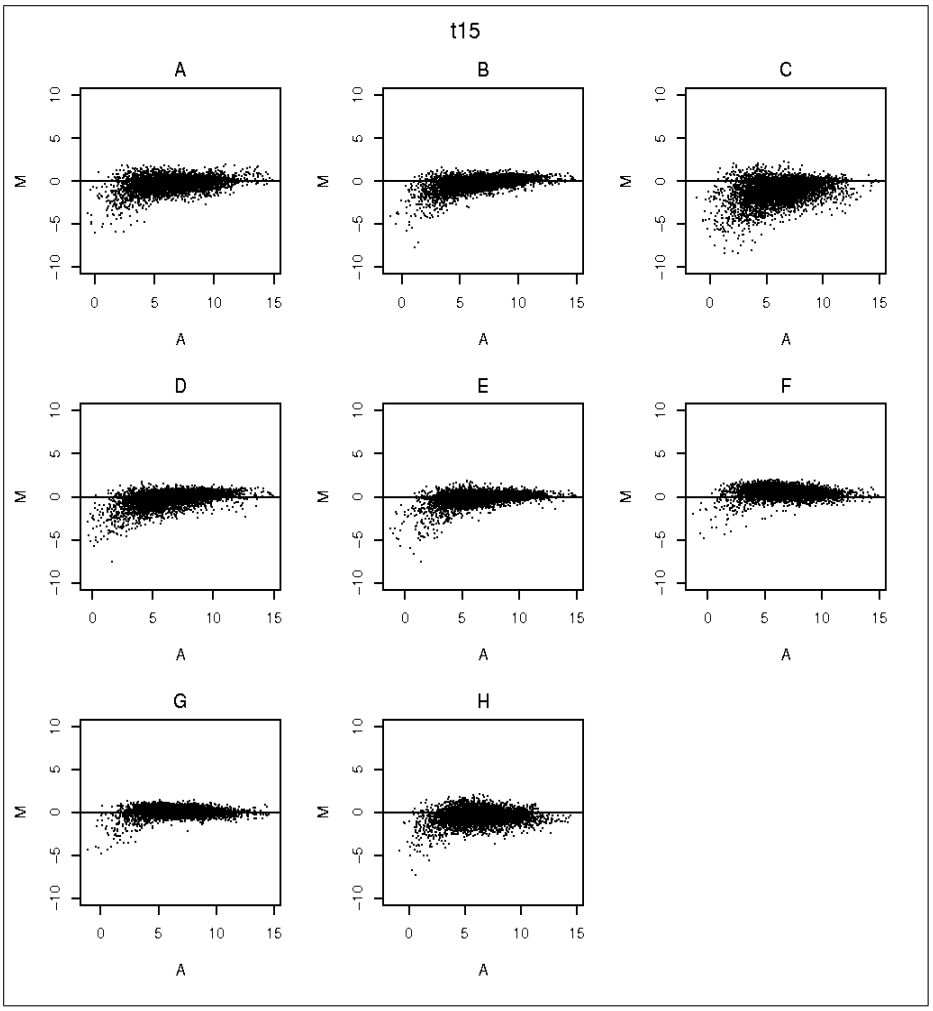


Figura A.18: Gràfics MA per cada rèplica en l'instant de temps t15

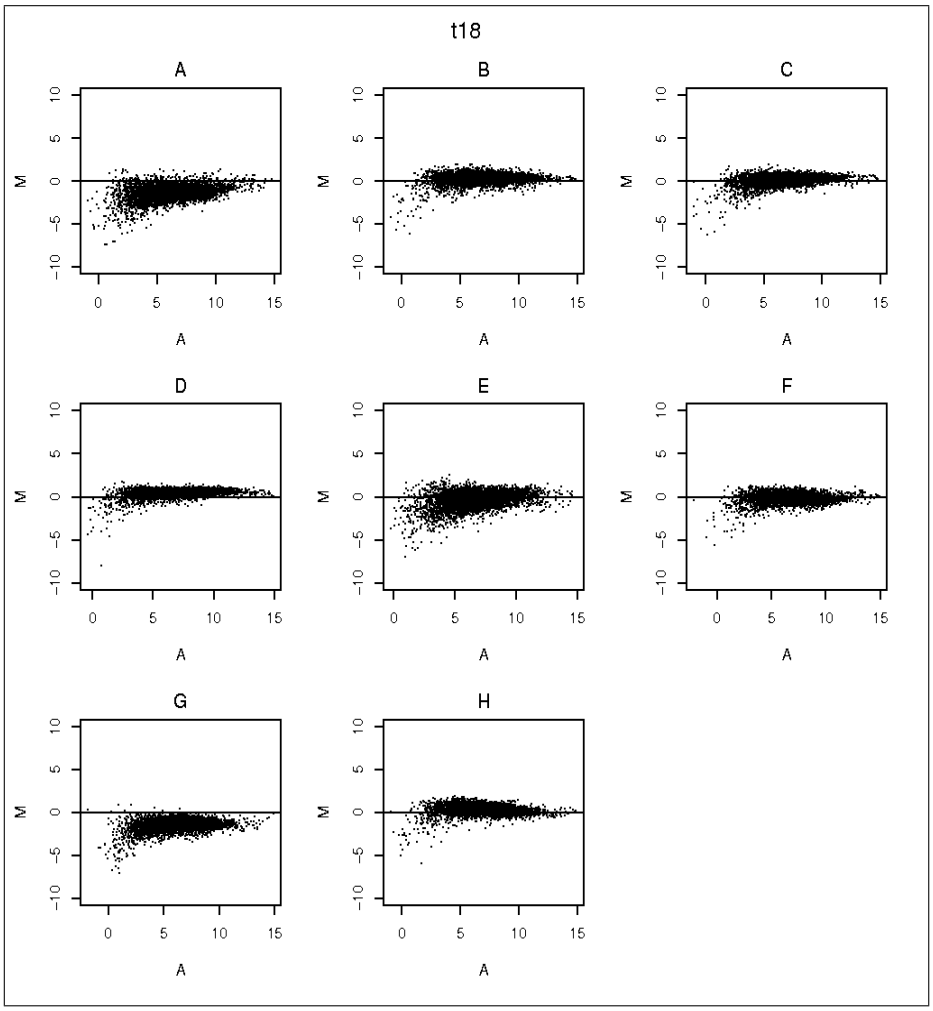


Figura A.19: Gràfics MA per cada rèplica en l'instant de temps t18

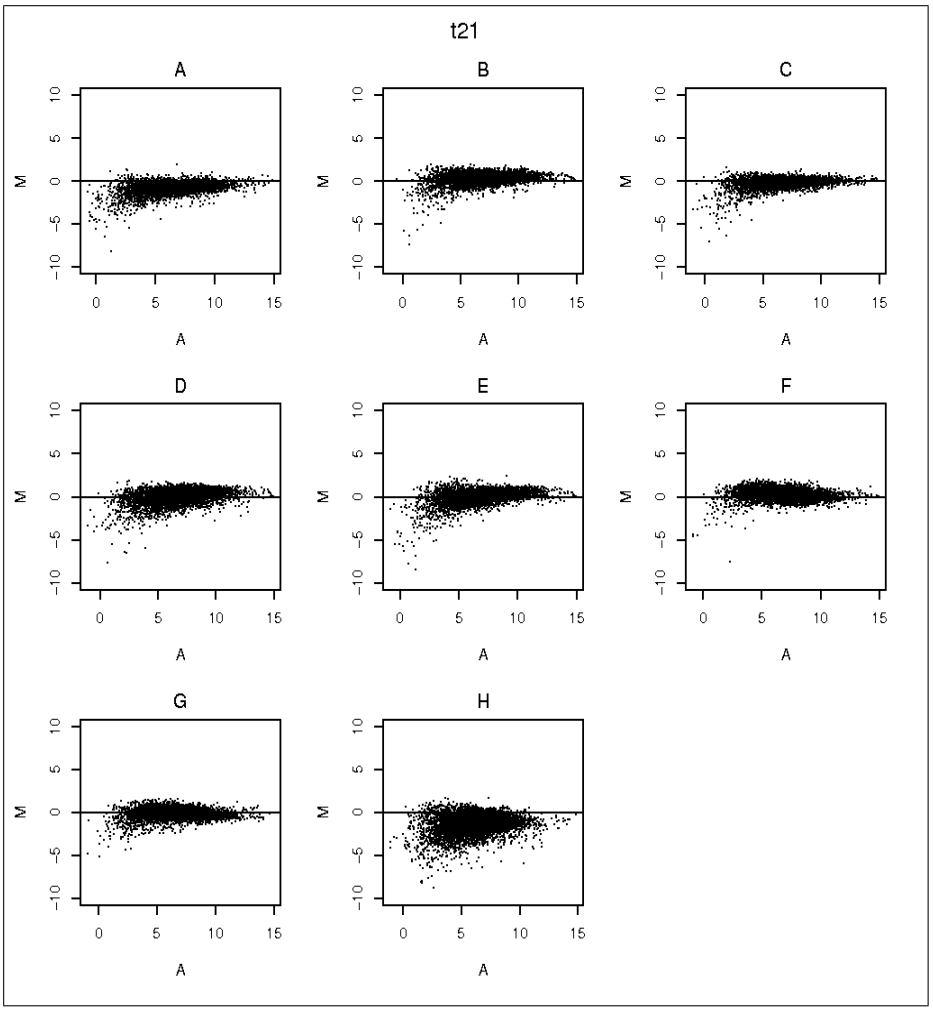


Figura A.20: Gràfics MA per cada rèplica en l'instant de temps t21

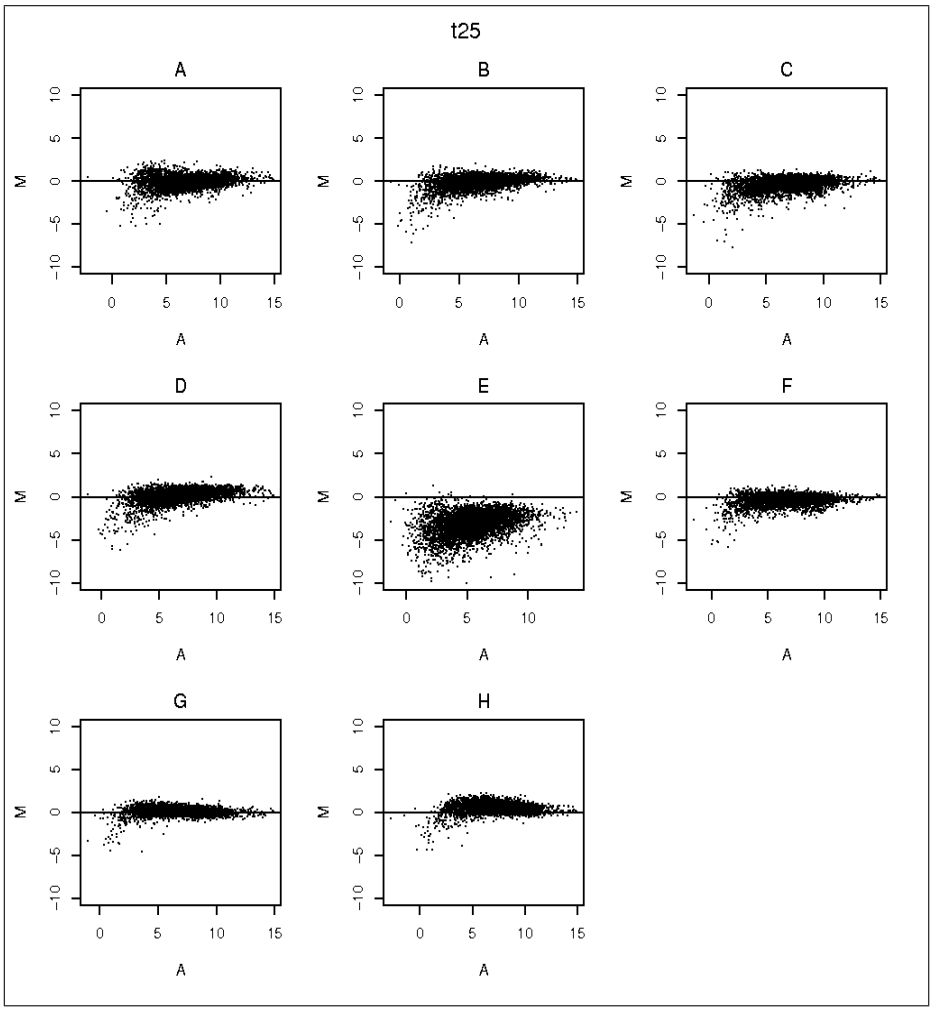


Figura A.21: Gràfics MA per cada rèplica en l'instant de temps t25

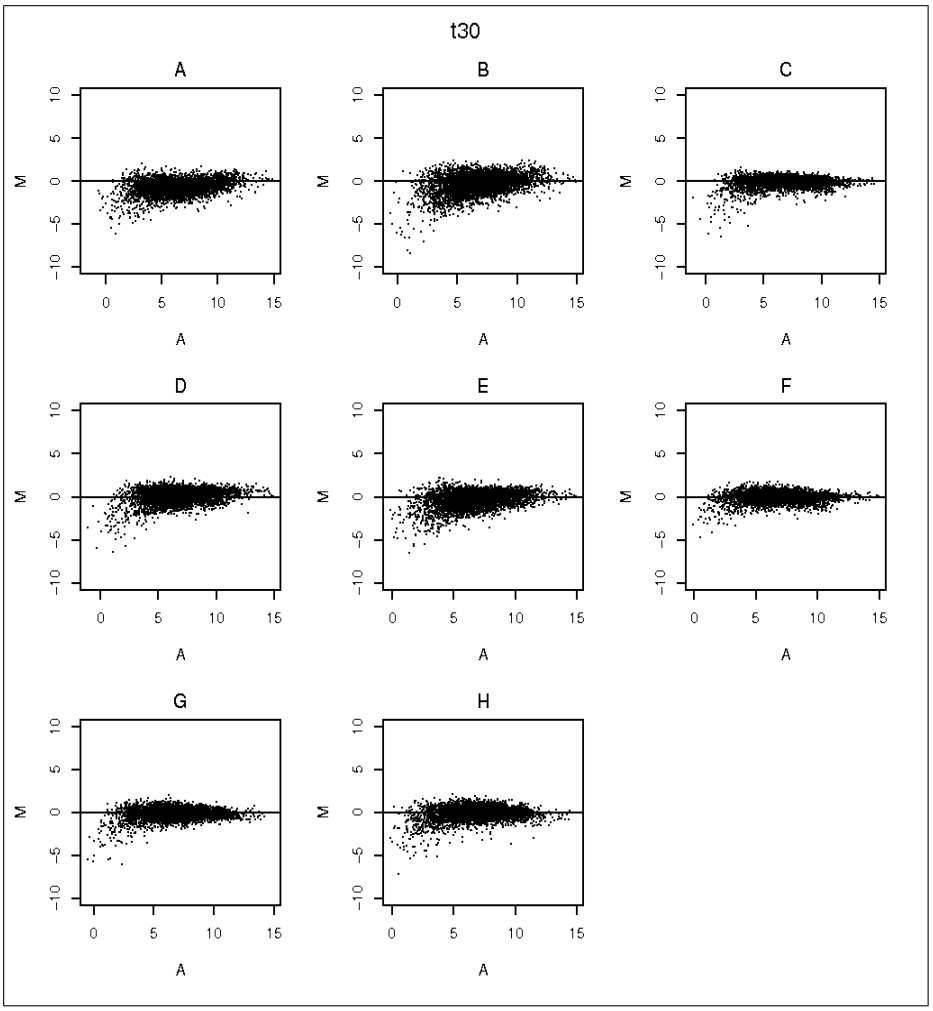


Figura A.22: Gràfics MA per cada rèplica en l'instant de temps t30

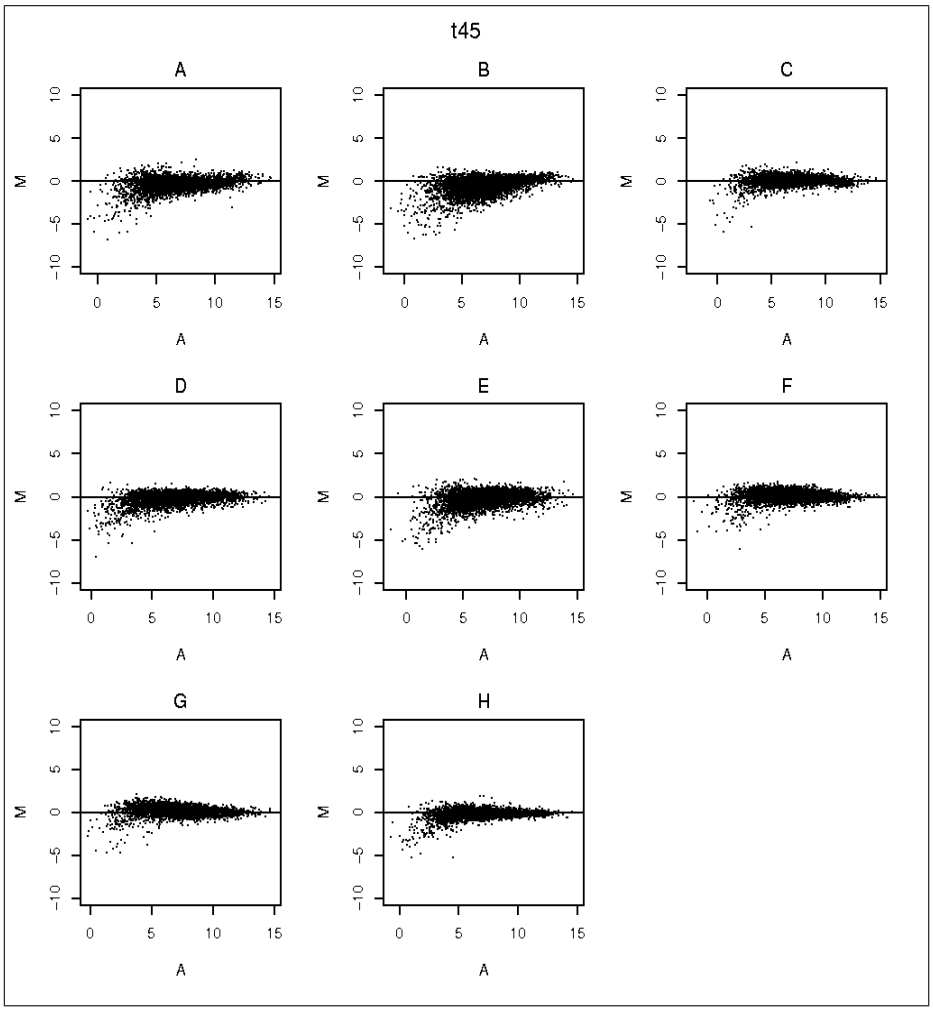


Figura A.23: Gràfics MA per cada rèplica en l'instant de temps t45

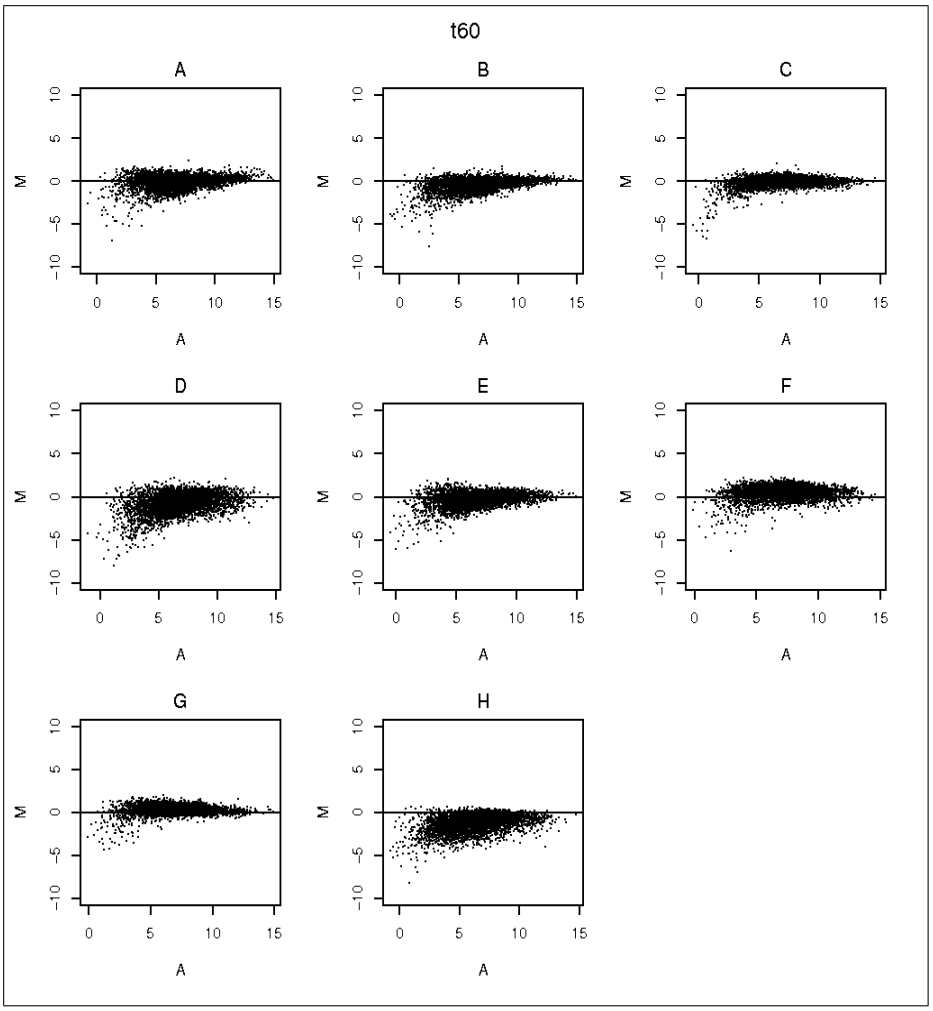


Figura A.24: Gràfics MA per cada rèplica en l'instant de temps t60

Apèndix B

Gràfics de les dades normalitzades

B.1 BoxPlots de les dades normalitzades:

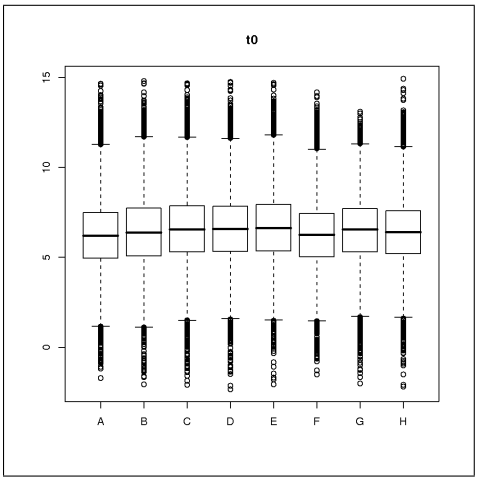


Figura B.1: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t0

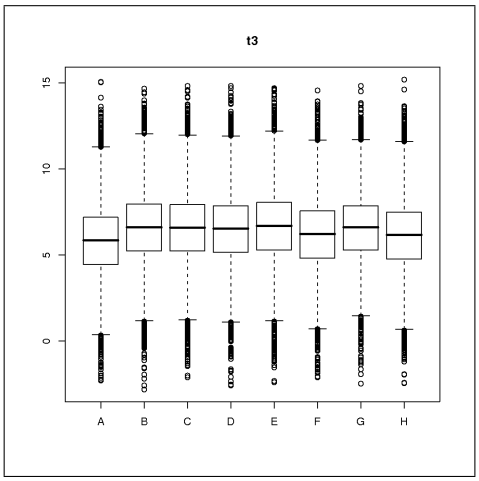


Figura B.2: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t3

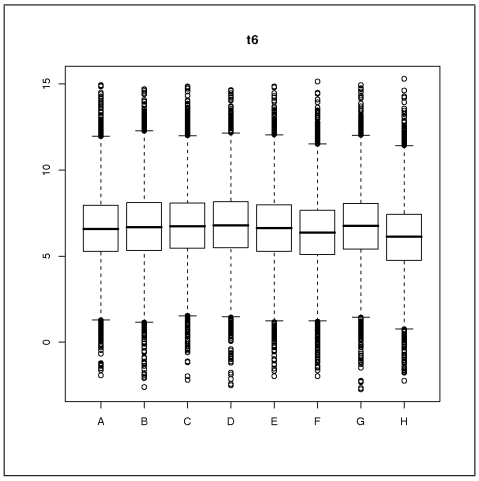


Figura B.3: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t6

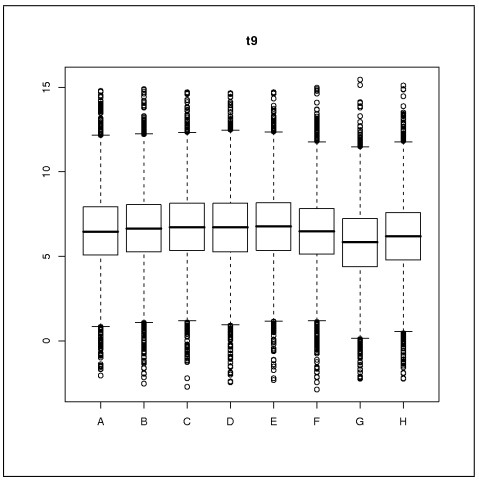


Figura B.4: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t9

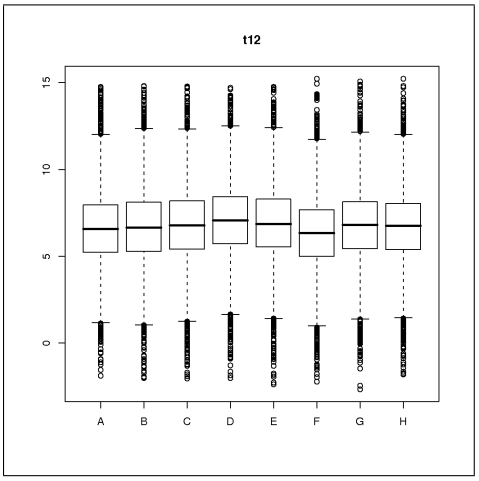


Figura B.5: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t12

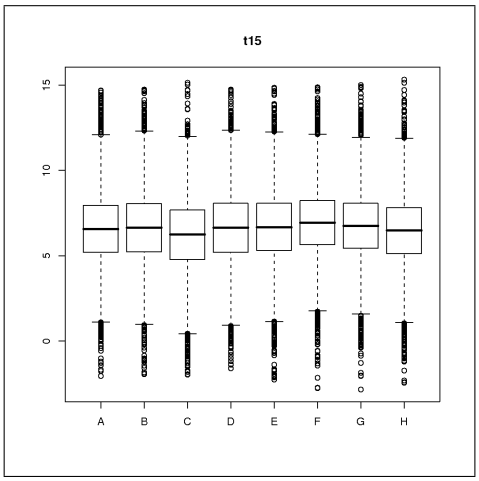


Figura B.6: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t15

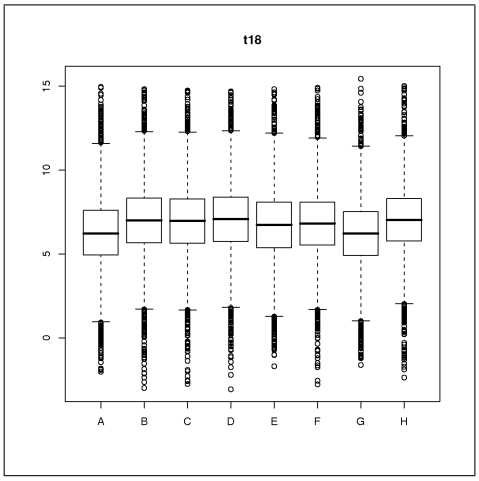


Figura B.7: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t18

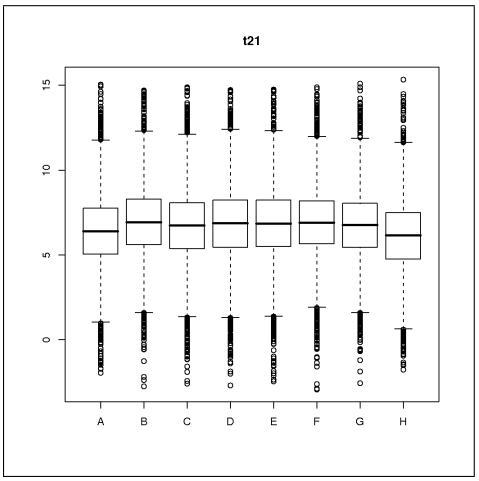


Figura B.8: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t21

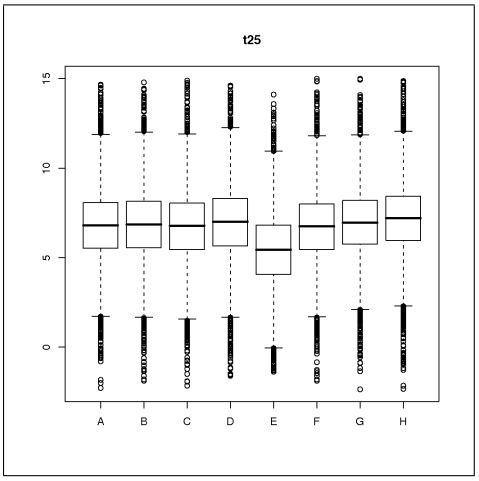


Figura B.9: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t25

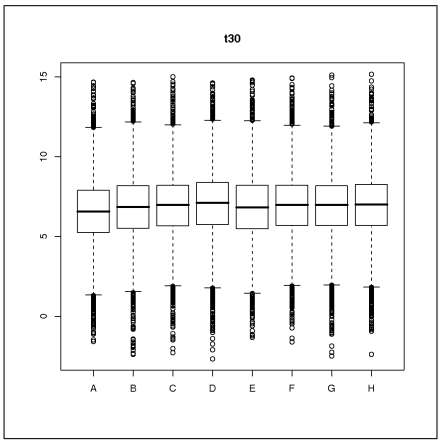


Figura B.10: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t30

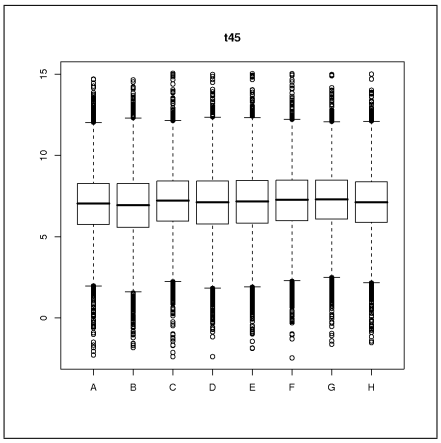


Figura B.11: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t45

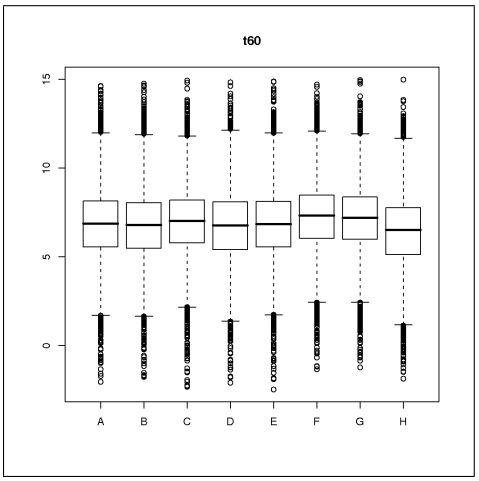


Figura B.12: Boxplot normalitzat per les 8 rèpliques en l'instant de temps t60

B.2 Gràfics MA de les dades normalitzades:

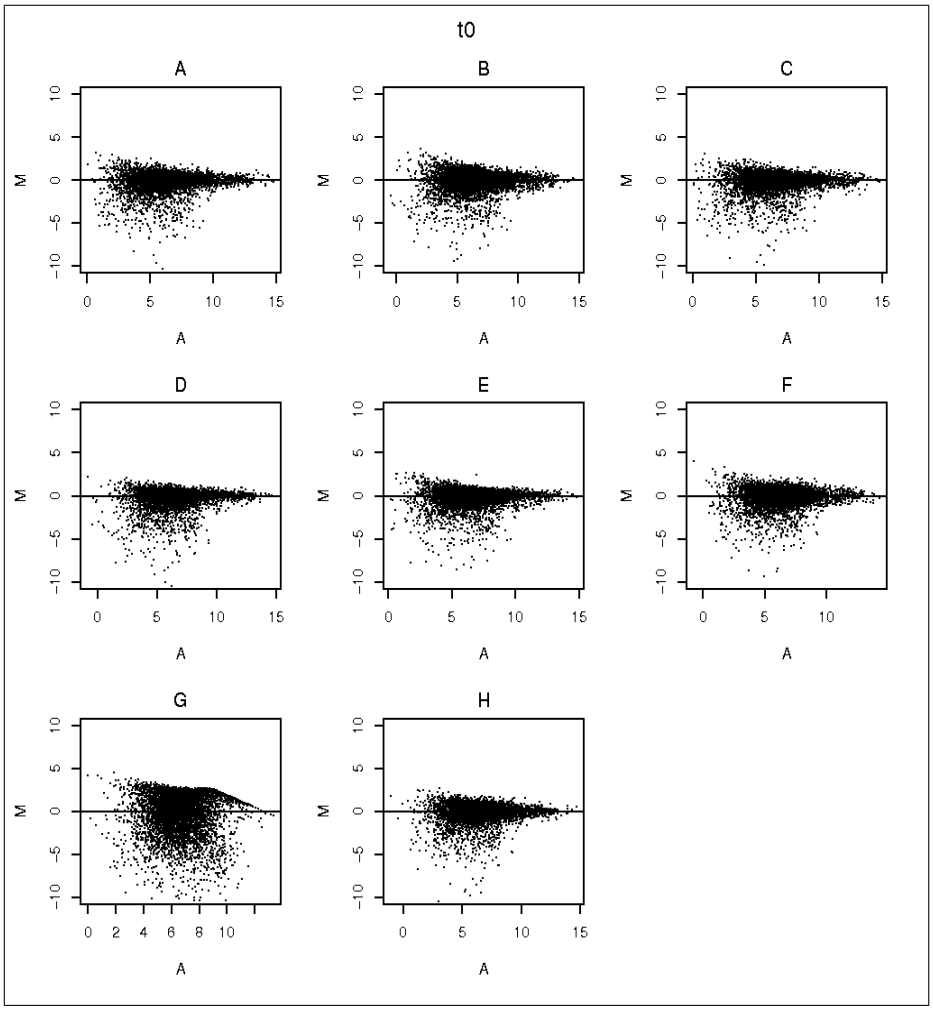


Figura B.13: Gràfics MA normalitzats per cada rèplica en l'instant de temps t0

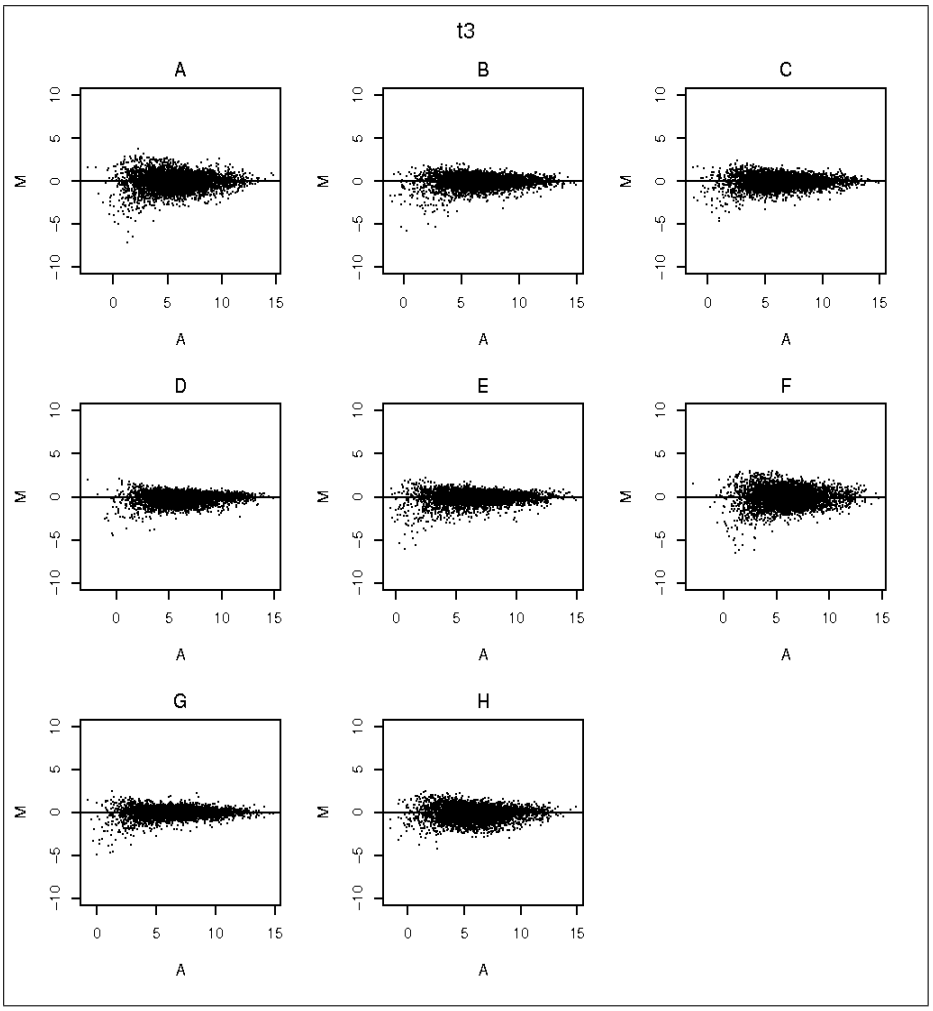


Figura B.14: Gràfics MA normalitzats per cada rèplica en l'instant de temps t3

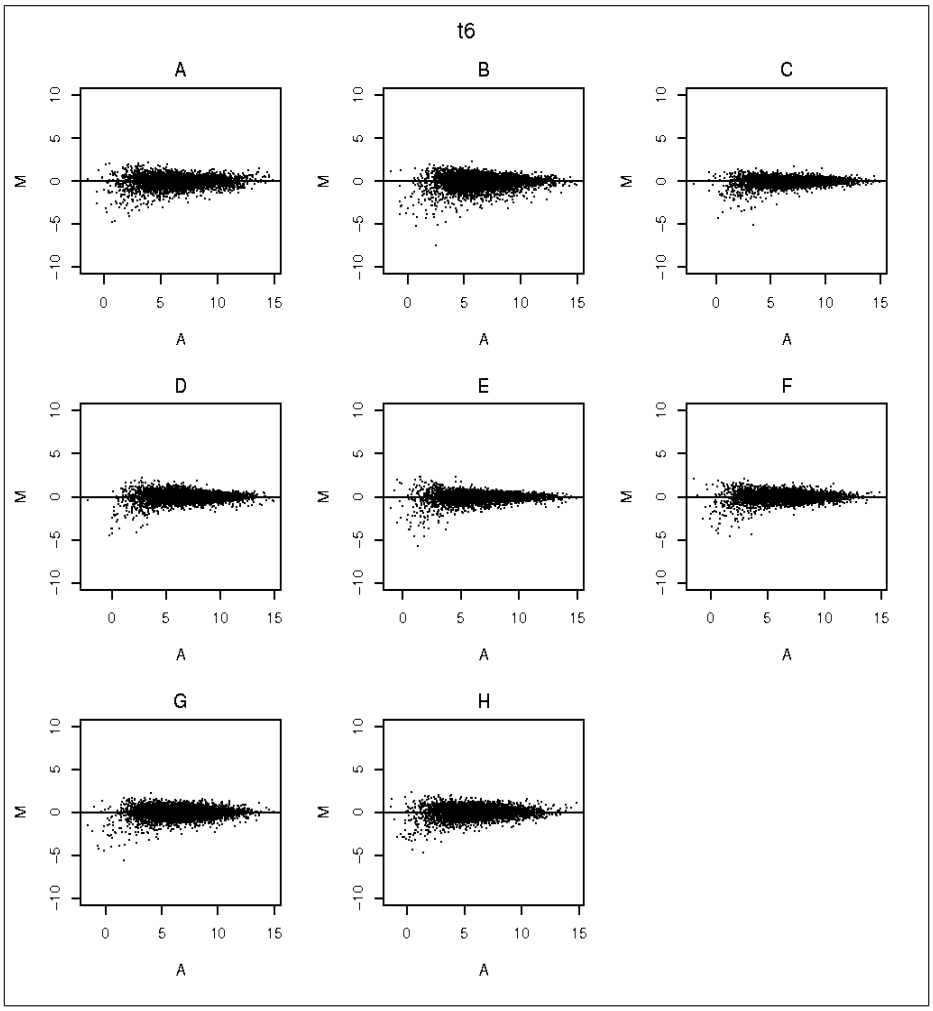


Figura B.15: Gràfics MA normalitzats per cada rèplica en l'instant de temps t6

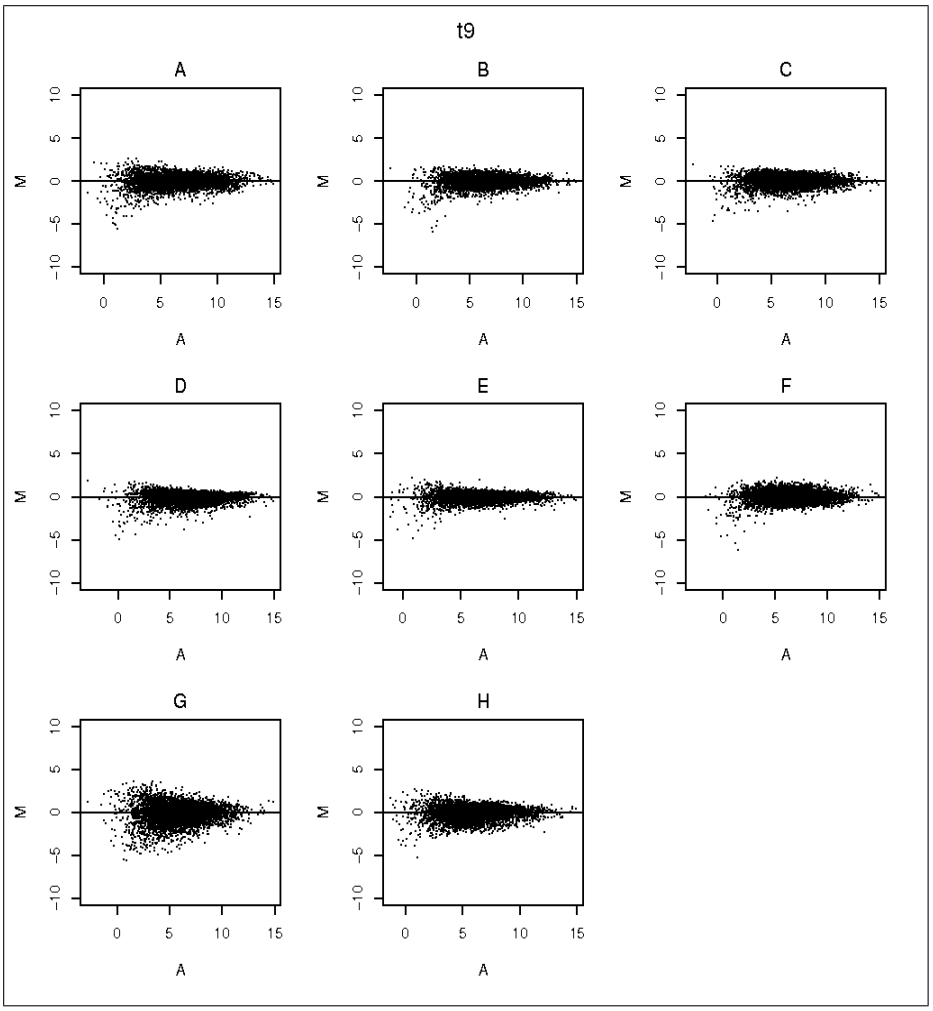


Figura B.16: Gràfics MA normalitzats per cada rèplica en l'instant de temps t9

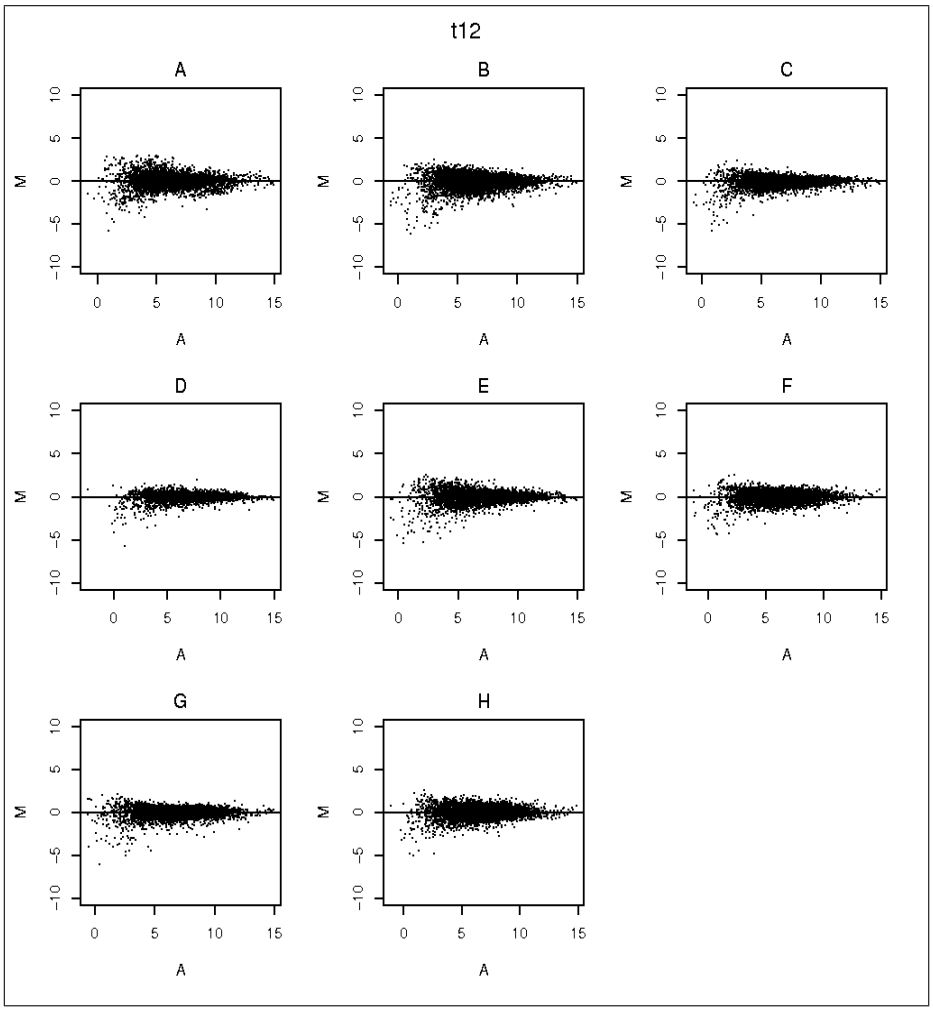


Figura B.17: Gràfics MA normalitzats per cada rèplica en l'instant de temps t12

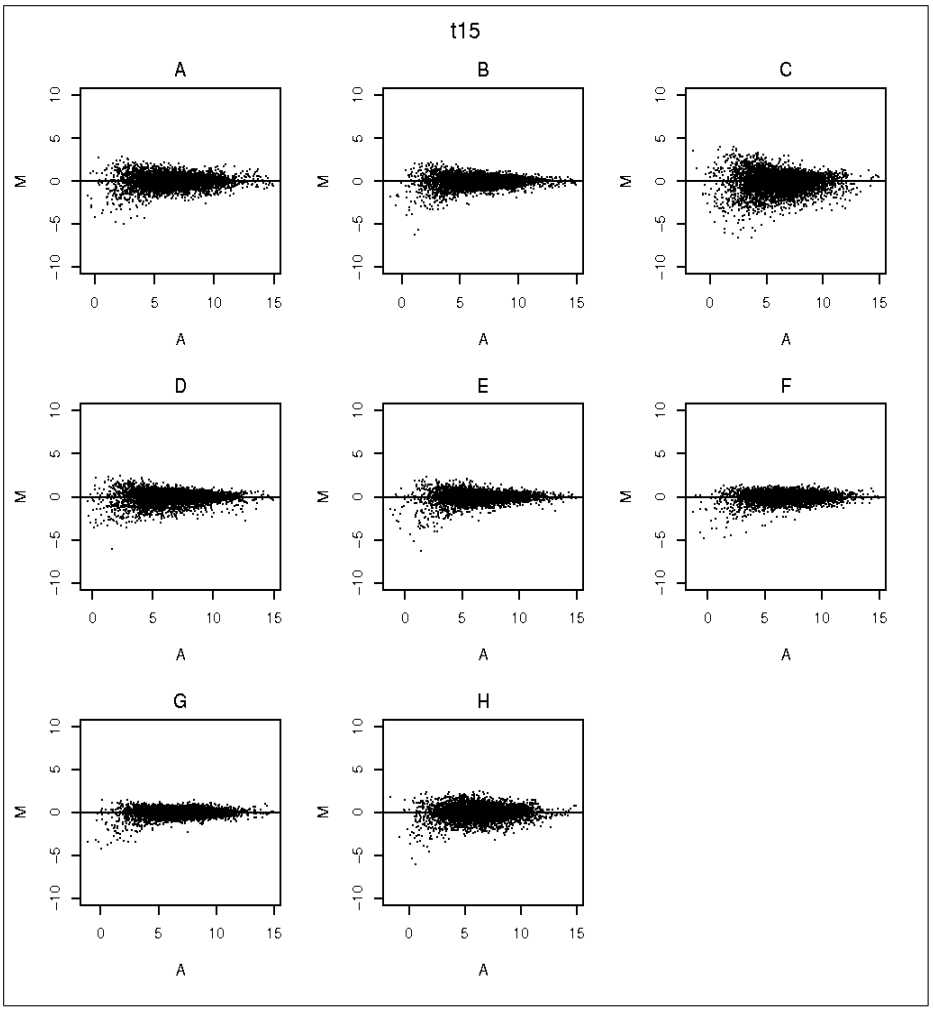


Figura B.18: Gràfics MA normalitzats per cada rèplica en l'instant de temps t15

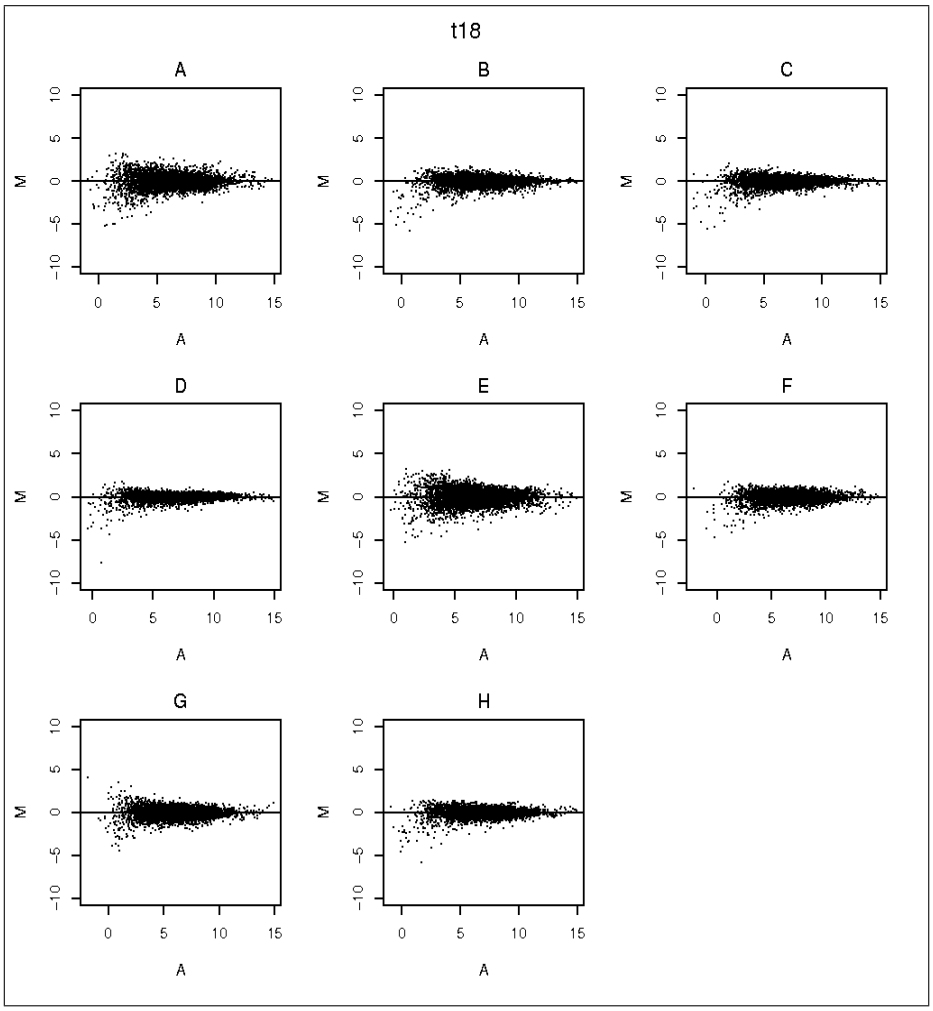


Figura B.19: Gràfics MA normalitzats per cada rèplica en l'instant de temps t18

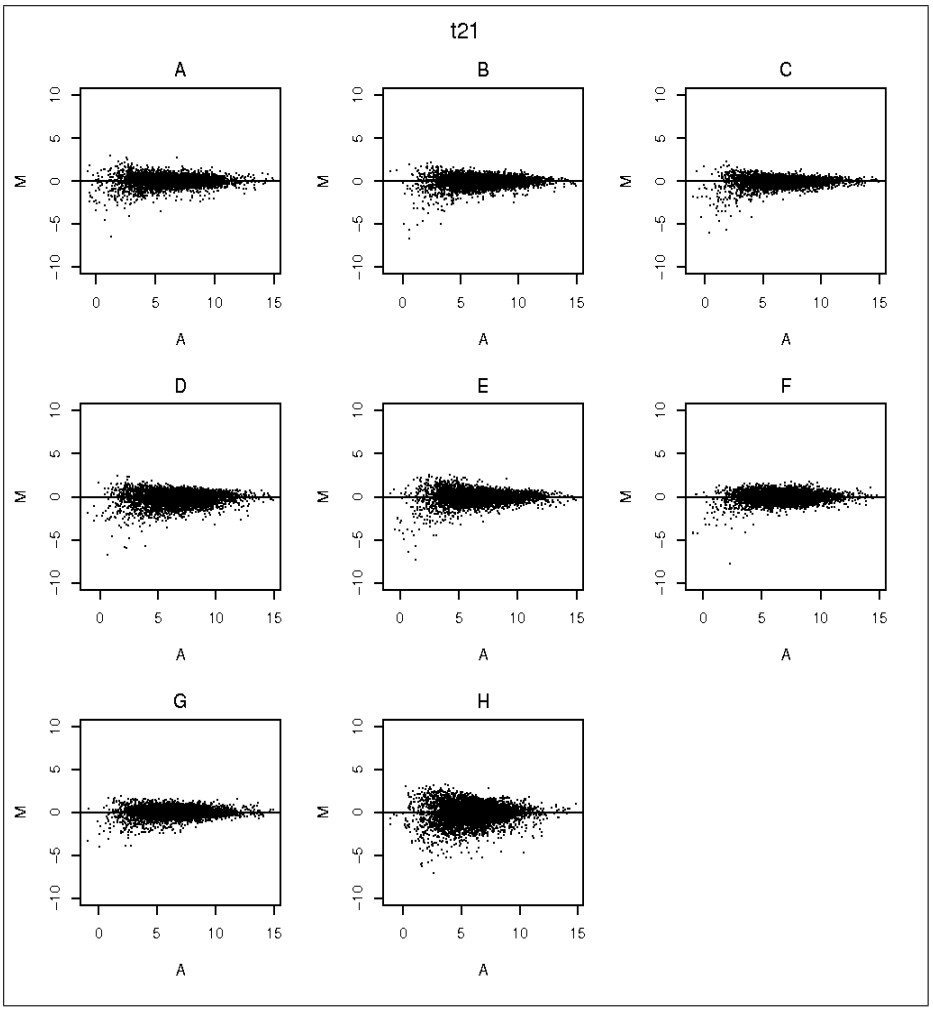


Figura B.20: Gràfics MA normalitzats per cada rèplica en l'instant de temps t21

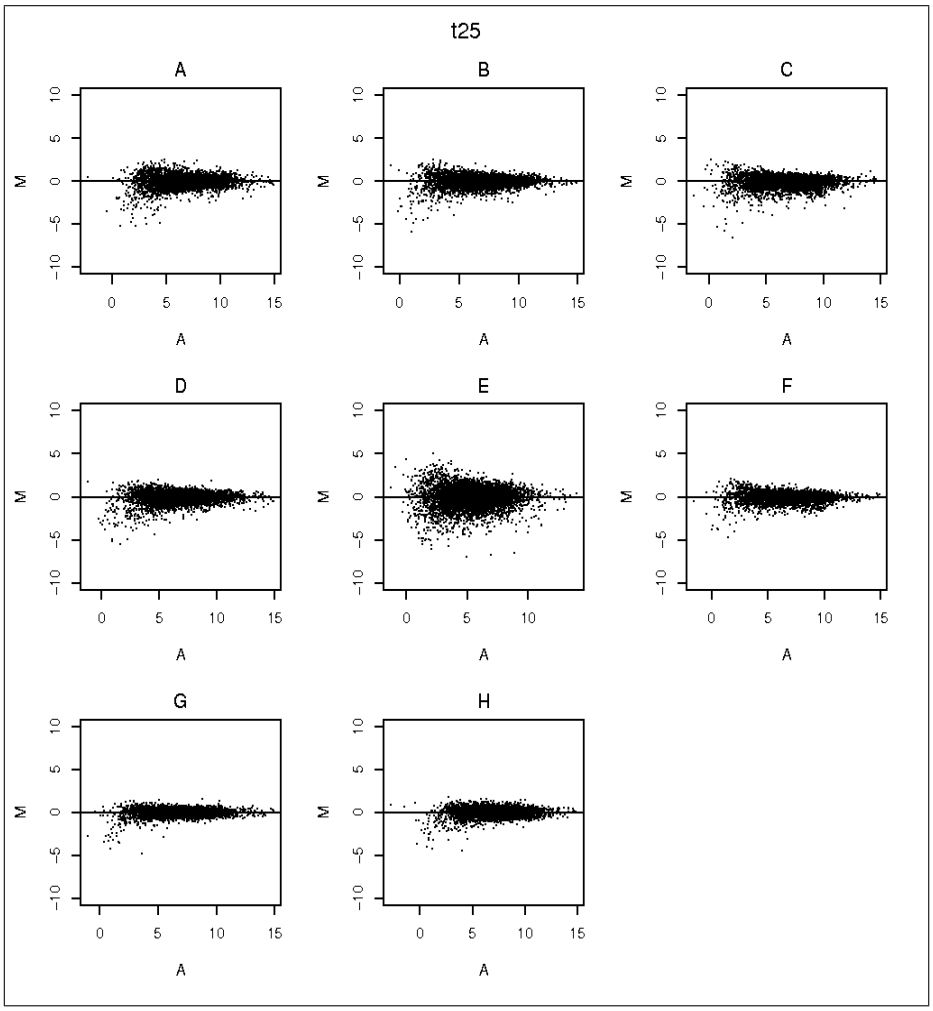


Figura B.21: Gràfics MA normalitzats per cada rèplica en l'instant de temps t25

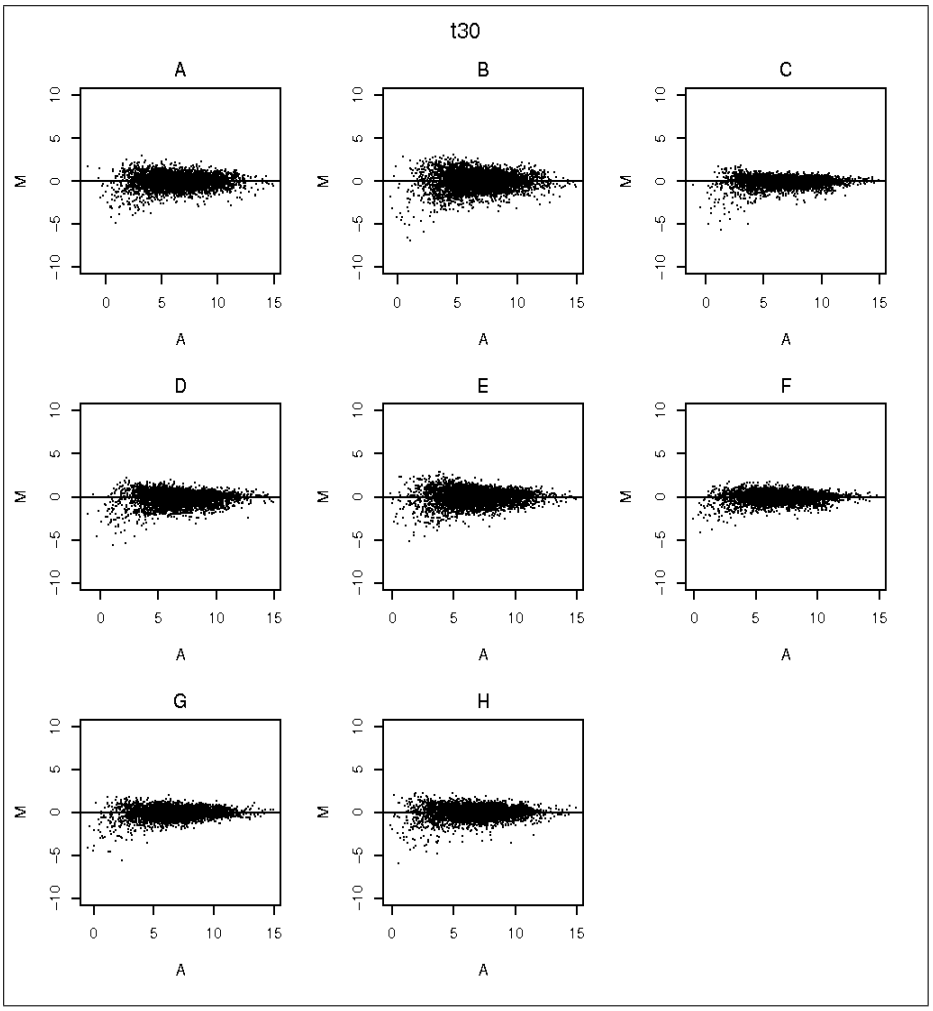


Figura B.22: Gràfics MA normalitzats per cada rèplica en l'instant de temps t30

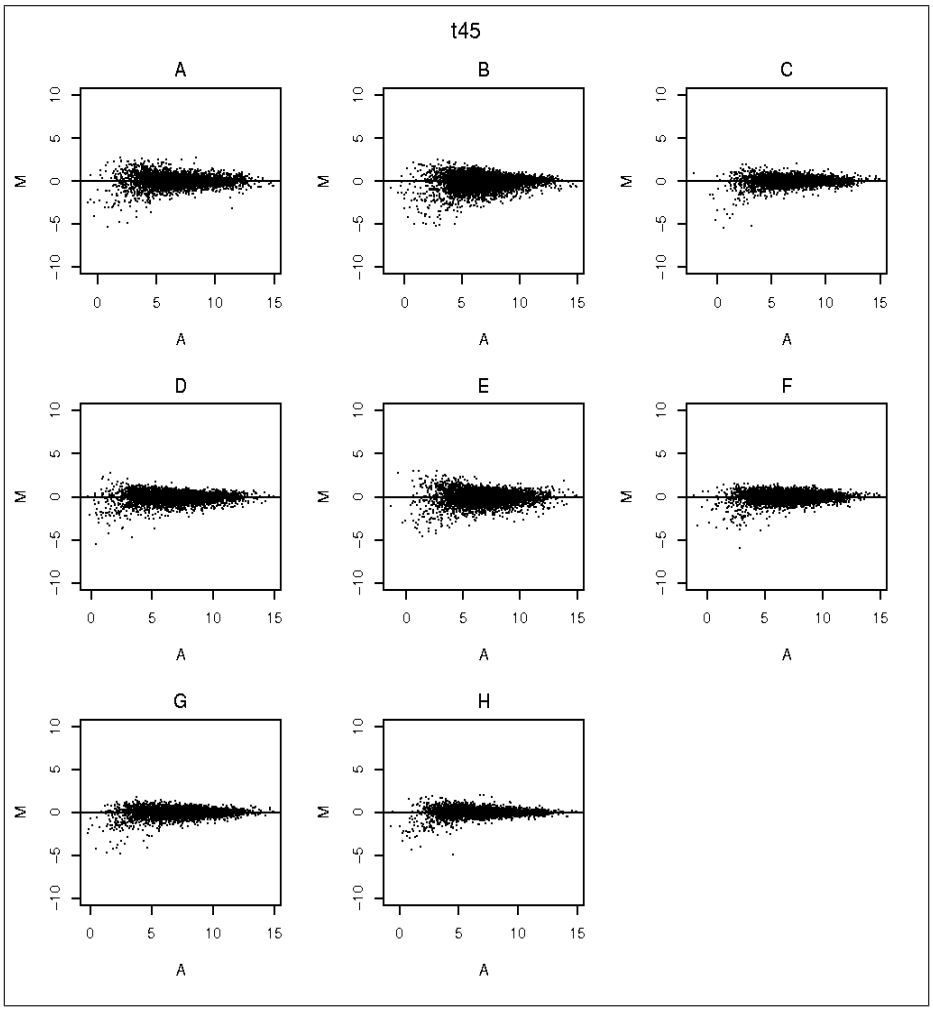


Figura B.23: Gràfics MA normalitzats per cada rèplica en l'instant de temps t45

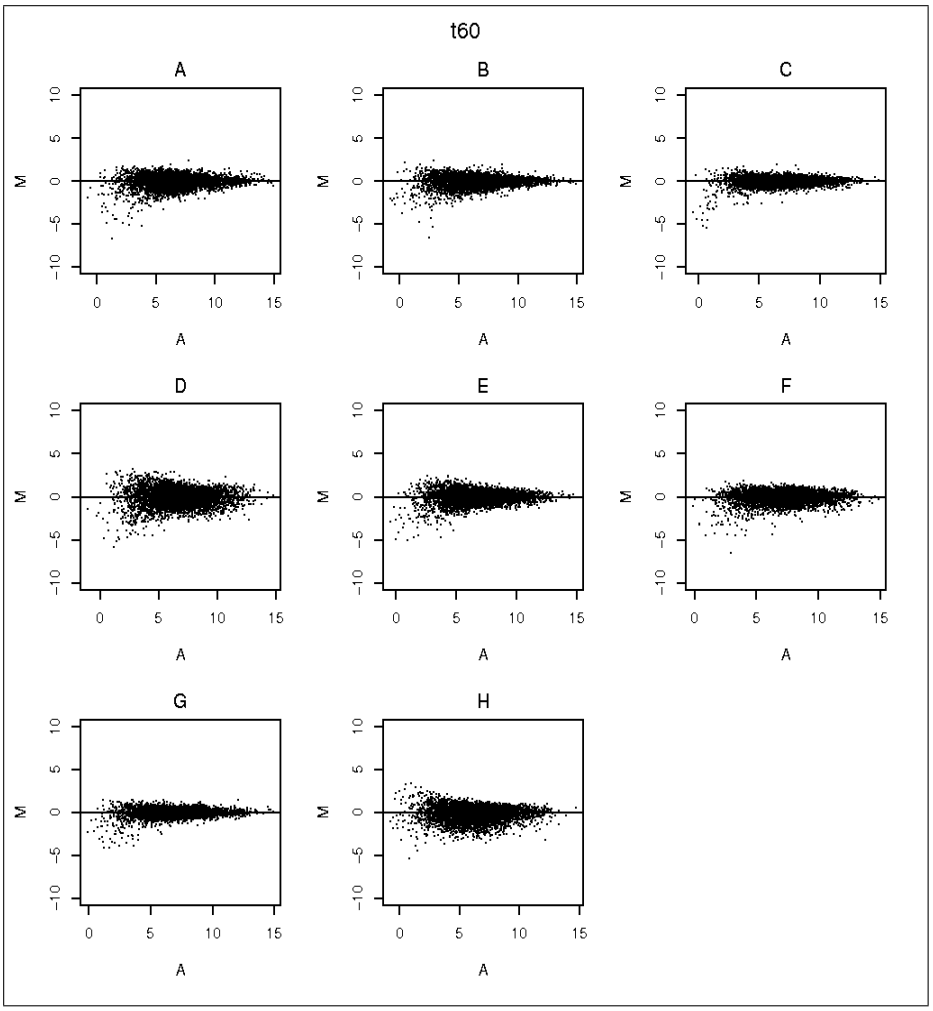
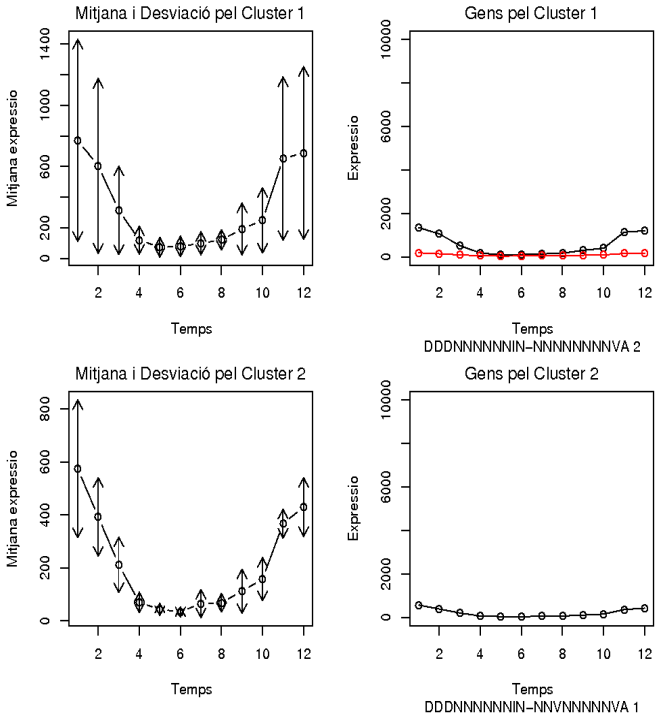
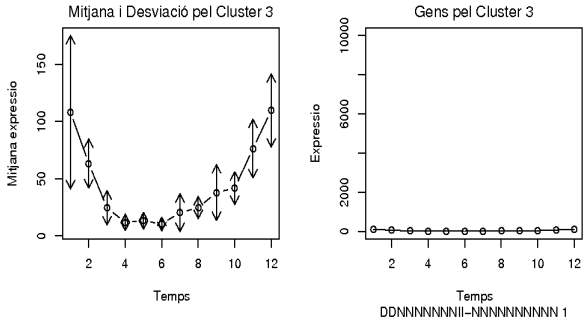


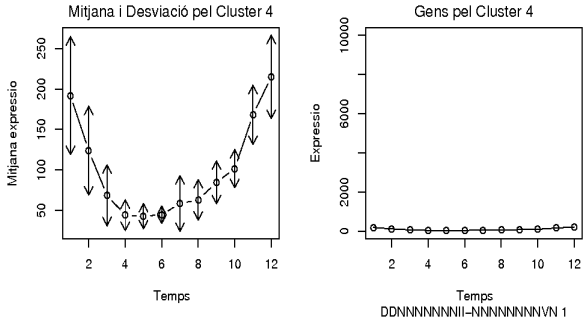
Figura B.24: Gràfics MA normalitzats per cada rèplica en l'instant de temps t60

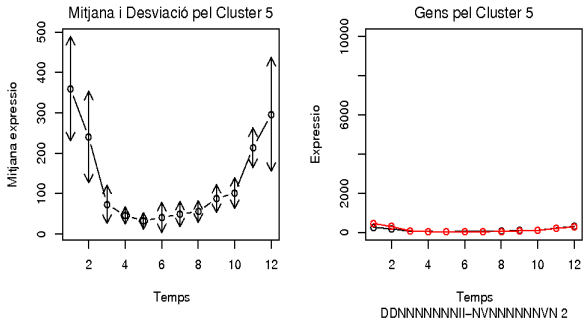
Apèndix C

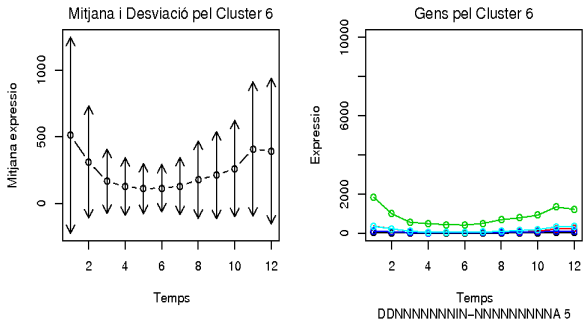
Gràfics dels clusters

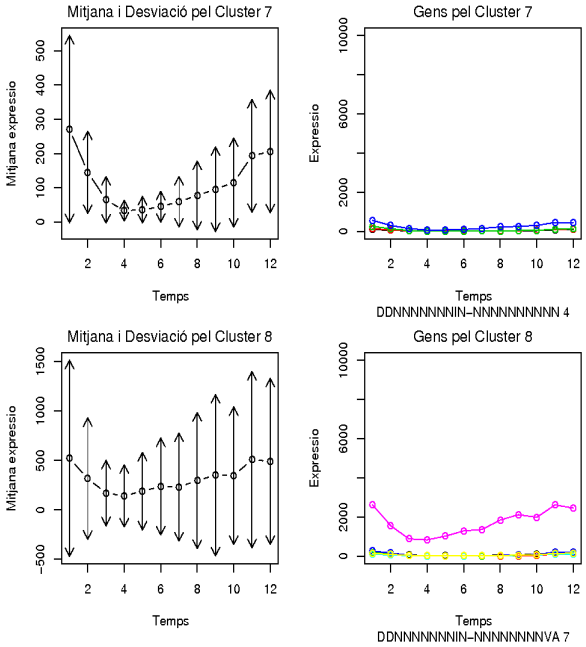


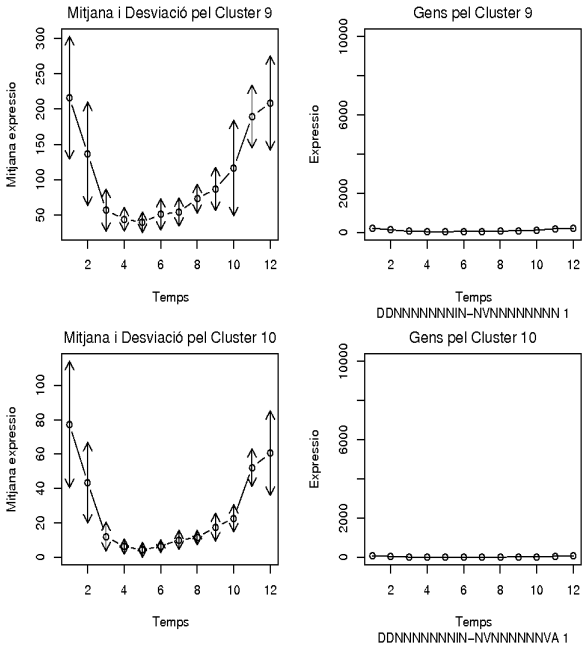

Mitjana i Desviació pel Cluster 3
Mitjana expressio
Temps
Gens pel Cluster 3
Expressio
Temps
DDNNNNNNNII-NNNNNNNNNN 1

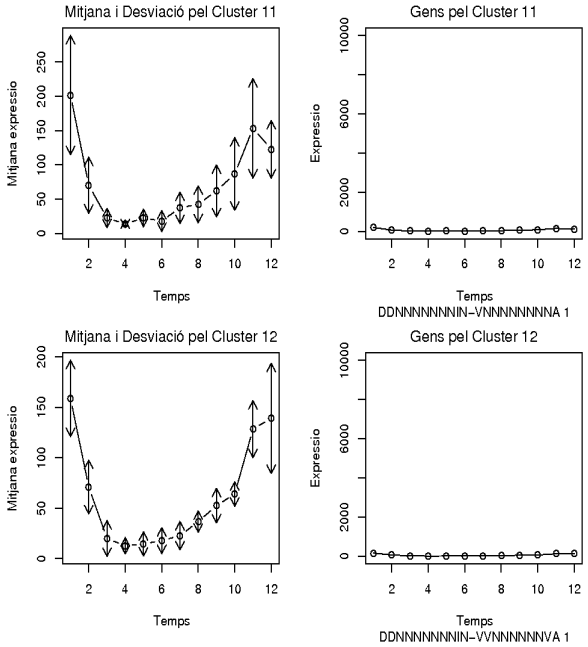

Mitjana i Desviació pel Cluster 4
Mitjana expressio
Temps
Gens pel Cluster 4
Expressio
Temps
DDNNNNNNNII-NNNNNNNNVN 1

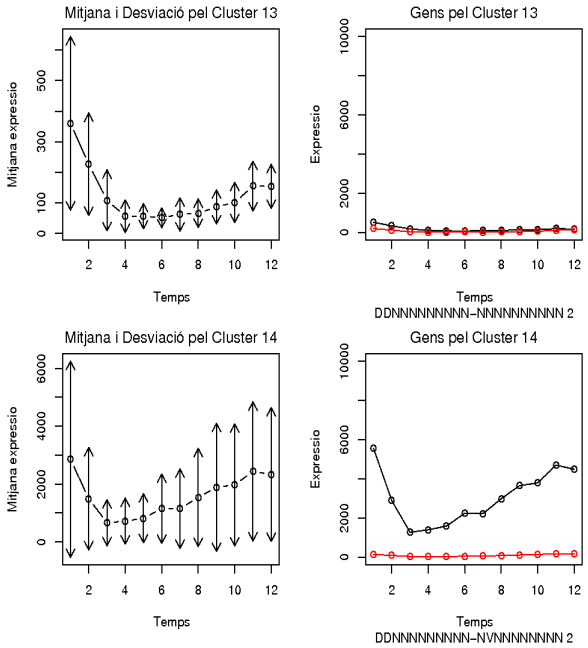

Mitjana i Desviació pel Cluster 5
Mitjana expressio
Temps
Gens pel Cluster 5
Expressio
Temps
DDNNNNNNNII-NVNNNNNNVN 2

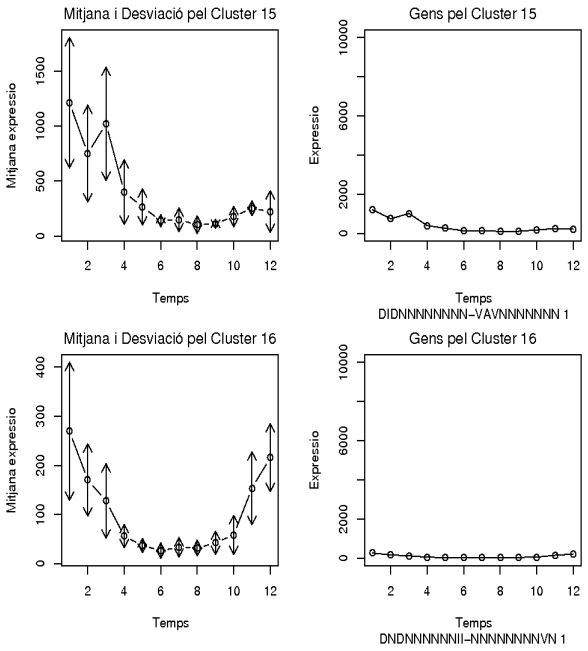

Mitjana i Desviació pel Cluster 6
Mitjana expressio
Temps
Gens pel Cluster 6
Expressio
Temps
DDNNNNNNNIN-NNNNNNNNNA 5

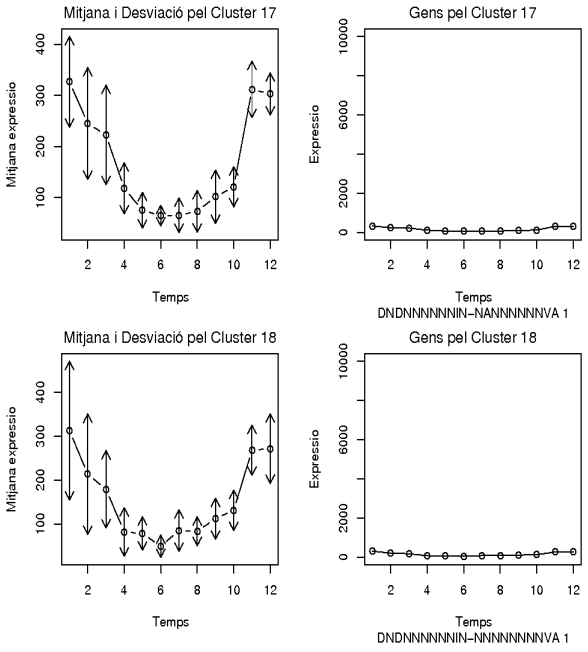

Mitjana i Desviació pel Cluster 7
Mitjana expressio
Temps
Gens pel Cluster 7
Expressio
Temps
DDNNNNNNNIN−NNNNNNNNNN 4
Mitjana i Desviació pel Cluster 8
Mitjana expressio
Temps
Gens pel Cluster 8
Expressio
Temps
DDNNNNNNNIN−NNNNNNNNVA 7

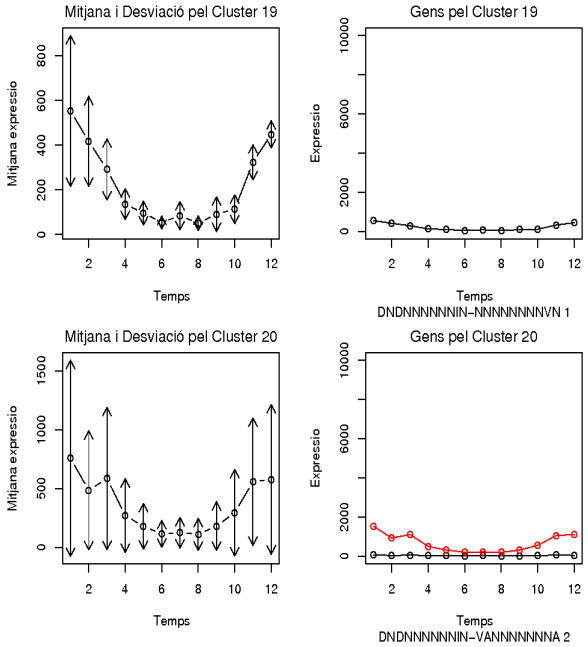

Mitjana i Desviació pel Cluster 9
Mitjana expressio
Temps
Gens pel Cluster 9
Expressio
Temps
DDNNNNNNNIN−NVNNNNNNNN 1
Mitjana i Desviació pel Cluster 10
Mitjana expressio
Temps
Gens pel Cluster 10
Expressio
Temps
DDNNNNNNNIN−NVNNNNNNVA 1

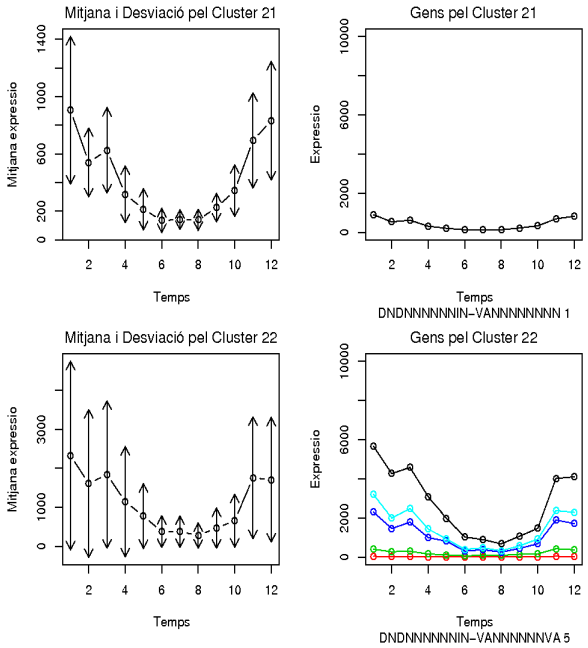

Mitjana i Desviació pel Cluster 11
Mitjana expressio
Temps
Gens pel Cluster 11
Expressio
Temps
DDNNNNNNNIN−VNNNNNNNNA 1
Mitjana i Desviació pel Cluster 12
Mitjana expressio
Temps
Gens pel Cluster 12
Expressio
Temps
DDNNNNNNNIN−VVNNNNNNVA 1

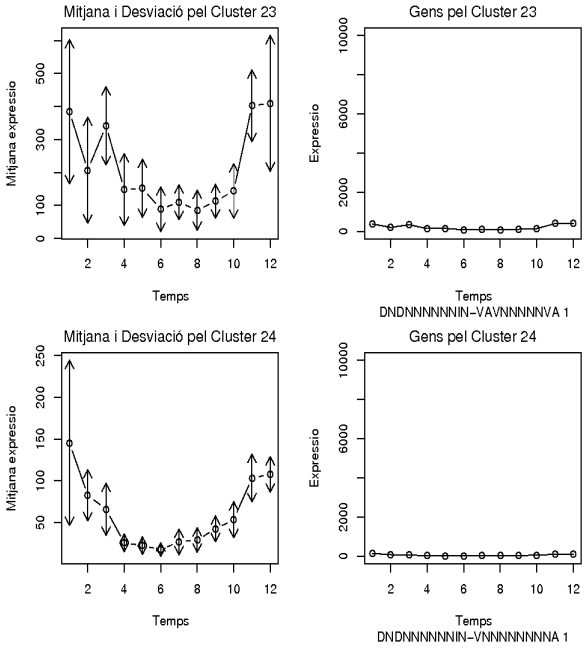

Mitjana i Desviació pel Cluster 13
Mitjana expressio
Temps
Gens pel Cluster 13
Expressio
Temps
DDNNNNNNNNN−NNNNNNNNN 2
Mitjana i Desviació pel Cluster 14
Mitjana expressio
Temps
Gens pel Cluster 14
Expressio
Temps
DDNNNNNNNNN−NVNNNNNNNN 2

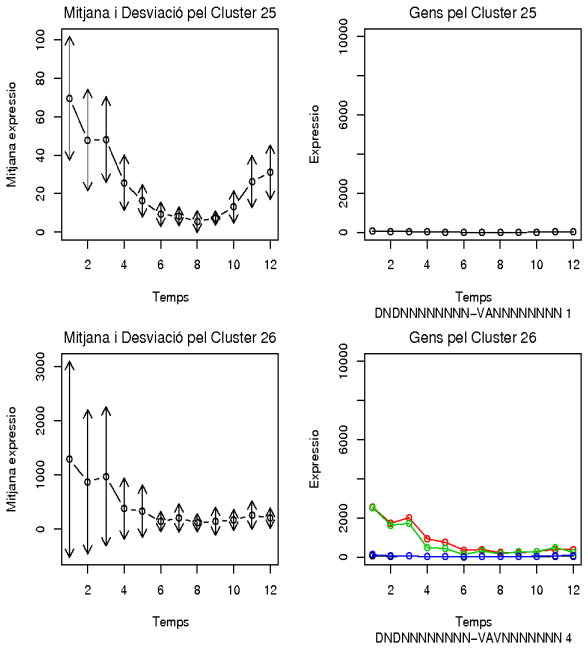

Mitjana i Desviació pel Cluster 15
Mitjana expressio
Temps
Gens pel Cluster 15
Expressio
Temps
DIDNNNNNNNN−VAVNNNNNNN 1
Mitjana i Desviació pel Cluster 16
Mitjana expressio
Temps
Gens pel Cluster 16
Expressio
Temps
DNDNNNNNNII−NNNNNNNNVN 1

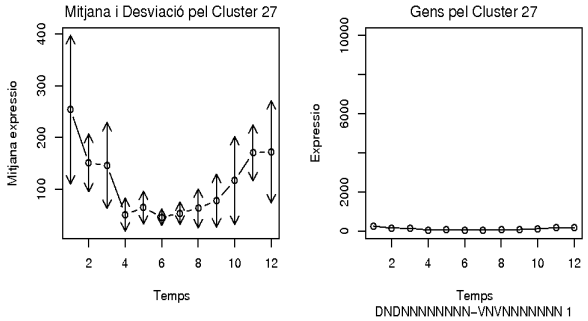

Mitjana i Desviació pel Cluster 17
Mitjana expressio
Temps
Gens pel Cluster 17
Expressio
Temps
DNDNNNNNNIN−NANNNNNNVA 1
Mitjana i Desviació pel Cluster 18
Mitjana expressio
Temps
Gens pel Cluster 18
Expressio
Temps
DNDNNNNNNIN−NNNNNNNNVA 1

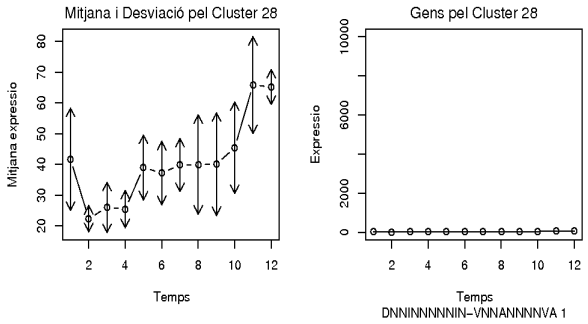

Mitjana i Desviació pel Cluster 19
Mitjana expressio
Temps
Gens pel Cluster 19
Expressio
Temps
DNDNNNNNNIN−NNNNNNNNVN 1
Mitjana i Desviació pel Cluster 20
Mitjana expressio
Temps
Gens pel Cluster 20
Expressio
Temps
DNDNNNNNNIN−VANNNNNNNA 2

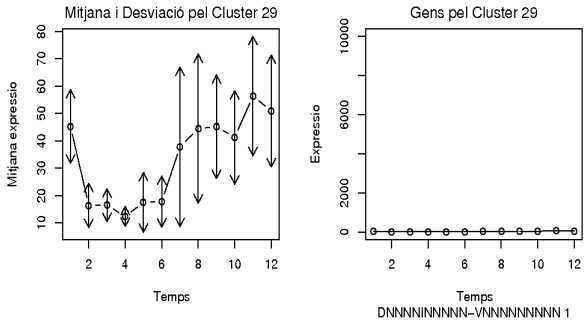

Mitjana i Desviació pel Cluster 21
Mitjana expressio
Temps
Gens pel Cluster 21
Expressio
Temps
DNDNNNNNNIN−VANNNNNNNN 1
Mitjana i Desviació pel Cluster 22
Mitjana expressio
Temps
Gens pel Cluster 22
Expressio
Temps
DNDNNNNNNIN−VANNNNNNVA 5

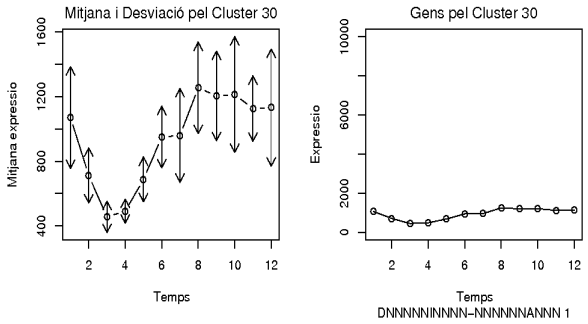

Mitjana i Desviació pel Cluster 23
Mitjana expressio
Temps
Gens pel Cluster 23
Expressio
Temps
DNDNNNNNNIN−VAVNNNNNVA 1
Mitjana i Desviació pel Cluster 24
Mitjana expressio
Temps
Gens pel Cluster 24
Expressio
Temps
DNDNNNNNNIN−VNNNNNNNNA 1

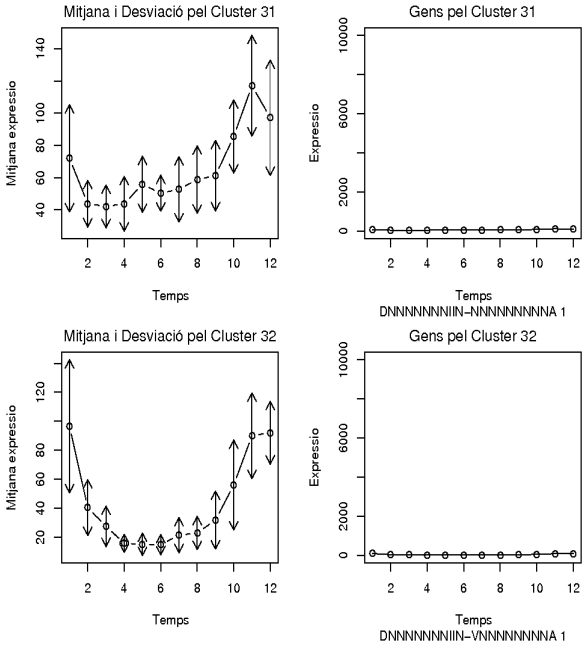

Mitjana i Desviació pel Cluster 25
Mitjana expressio
Temps
Gens pel Cluster 25
Expressio
Temps
DNDNNNNNNNN−VANNNNNNNN 1
Mitjana i Desviació pel Cluster 26
Mitjana expressio
Temps
Gens pel Cluster 26
Expressio
Temps
DNDNNNNNNNN−VAVNNNNNNN 4

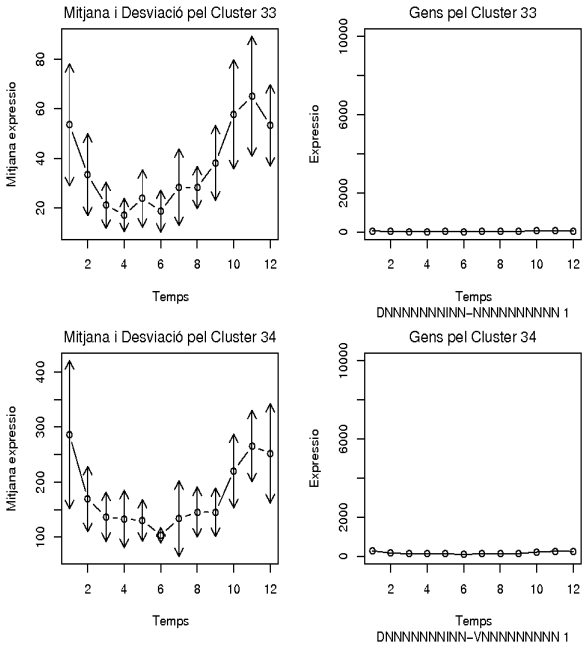

Mitjana i Desviació pel Cluster 27
Mitjana expressio
Temps
Gens pel Cluster 27
Expressio
Temps
DNDNNNNNNNN−VNVNNNNNNN 1

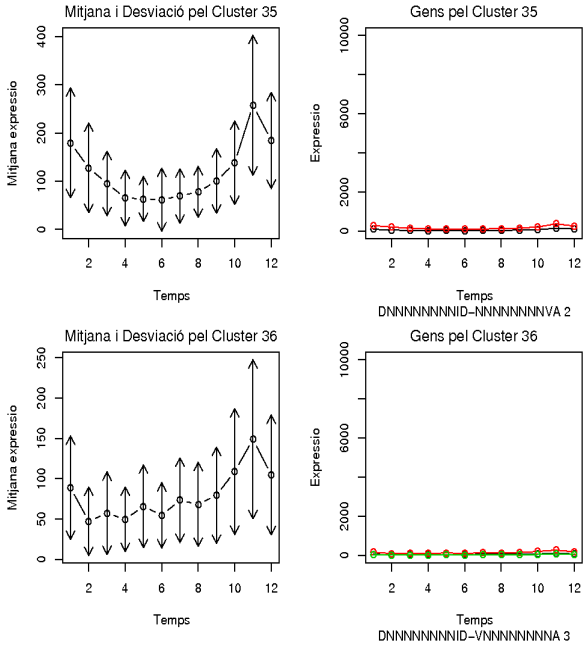

Mitjana i Desviació pel Cluster 28
Mitjana expressio
Temps
Gens pel Cluster 28
Expressio
Temps
DNNINNNNNIN−VNNANNNNVA 1

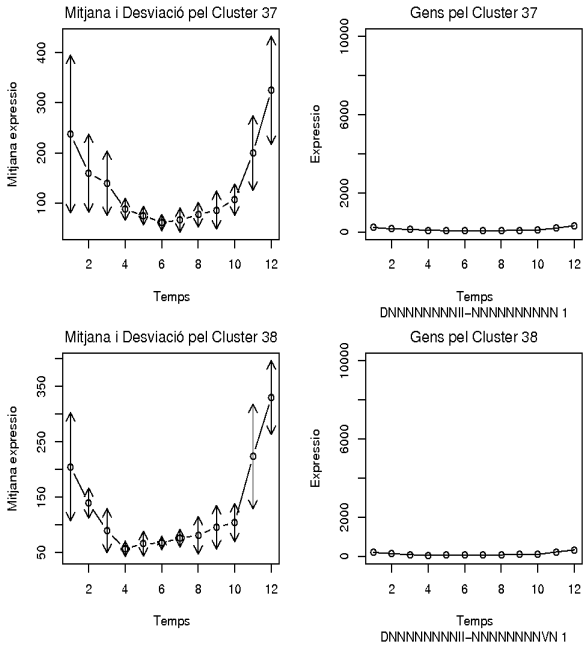

Mitjana i Desviació pel Cluster 29
Mitjana expressio
Temps
Gens pel Cluster 29
Expressio
Temps
DNNNNINNNNN−VNNNNNNNNN 1

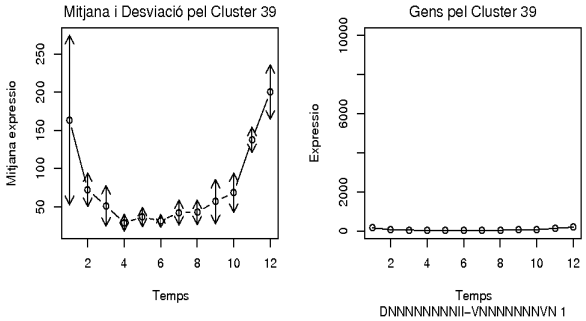

Mitjana i Desviació pel Cluster 30
Mitjana expressio
Temps
Gens pel Cluster 30
Expressio
Temps
DNNNNNINNNN−NNNNNNANNN 1

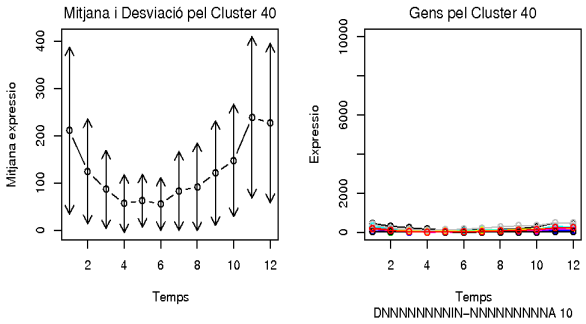

Mitjana i Desviació pel Cluster 31
Mitjana expressio
Temps
Gens pel Cluster 31
Expressio
Temps
DNNNNNNNIIN−NNNNNNNNNA 1
Mitjana i Desviació pel Cluster 32
Mitjana expressio
Temps
Gens pel Cluster 32
Expressio
Temps
DNNNNNNNIIN−VNNNNNNNNA 1

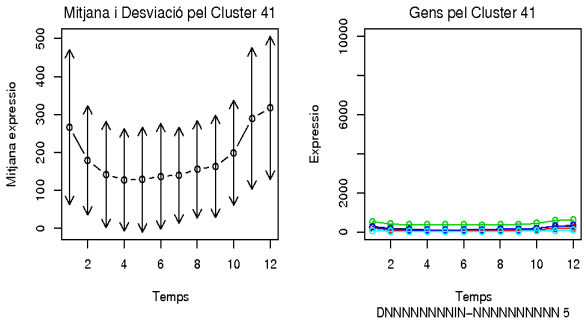

Mitjana i Desviació pel Cluster 33
Mitjana expressio
Temps
Gens pel Cluster 33
Expressio
Temps
DNNNNNNNINN−NNNNNNNNNN 1
Mitjana i Desviació pel Cluster 34
Mitjana expressio
Temps
Gens pel Cluster 34
Expressio
Temps
DNNNNNNNINN−VNNNNNNNNN 1

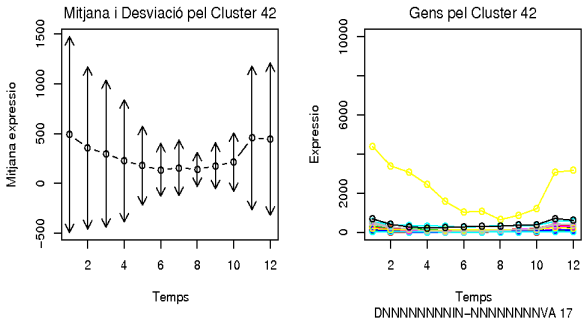

Mitjana i Desviació pel Cluster 35
Mitjana expressio
Temps
Gens pel Cluster 35
Expressio
Temps
DNNNNNNNNID−NNNNNNNNVA 2
Mitjana i Desviació pel Cluster 36
Mitjana expressio
Temps
Gens pel Cluster 36
Expressio
Temps
DNNNNNNNNID−VNNNNNNNNA 3

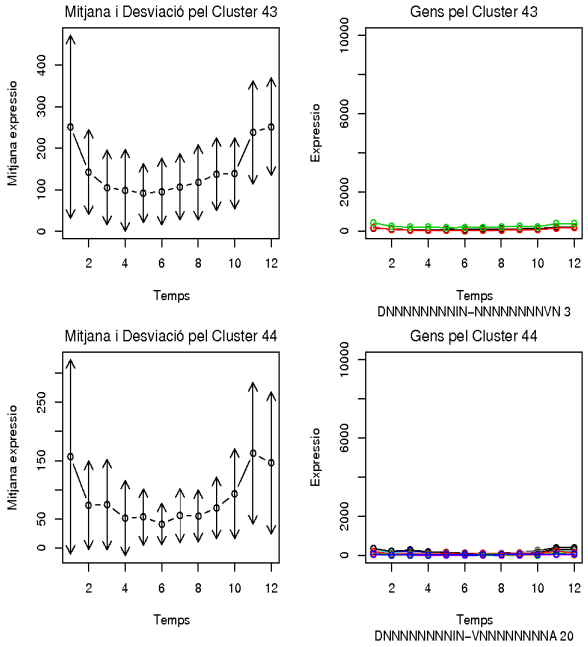

Mitjana i Desviació pel Cluster 37
Mitjana expressio
Temps
Gens pel Cluster 37
Expressio
Temps
DNNNNNNNNII−NNNNNNNNNN 1
Mitjana i Desviació pel Cluster 38
Mitjana expressio
Temps
Gens pel Cluster 38
Expressio
Temps
DNNNNNNNNII−NNNNNNNNVN 1

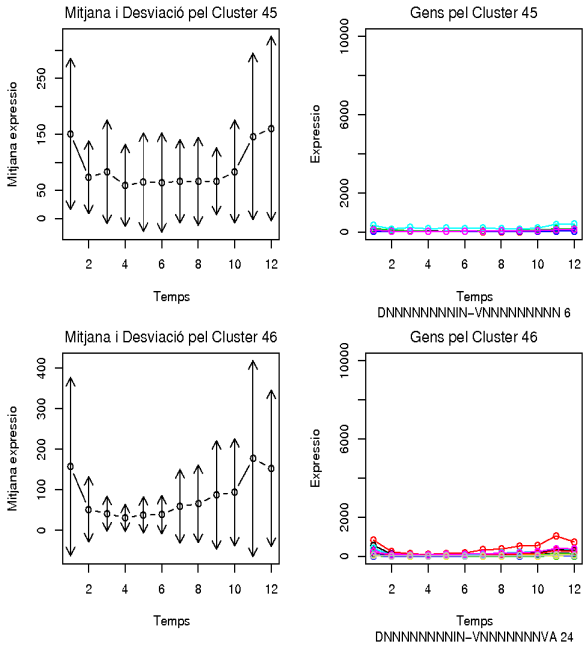

Mitjana i Desviació pel Cluster 39
Mitjana expressio
Temps
Gens pel Cluster 39
Expressio
Temps
DNNNNNNNNII−VNNNNNNNVN 1

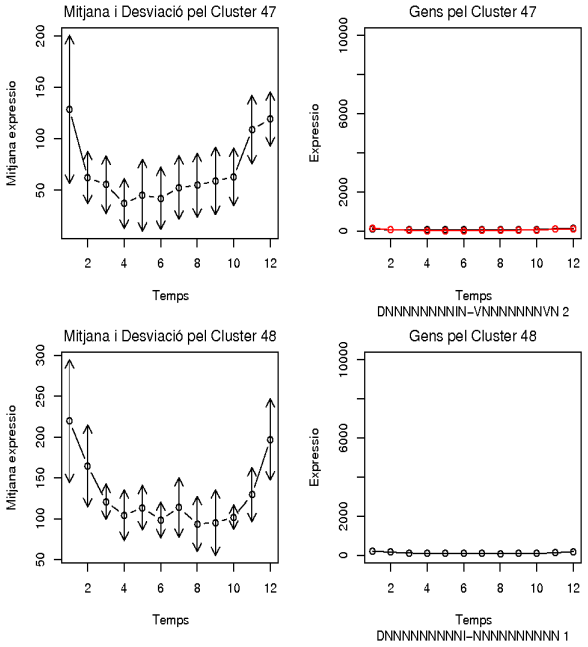

Mitjana i Desviació pel Cluster 40
Mitjana expressio
Temps
Gens pel Cluster 40
Expressio
Temps
DNNNNNNNNIN−NNNNNNNNNA 10

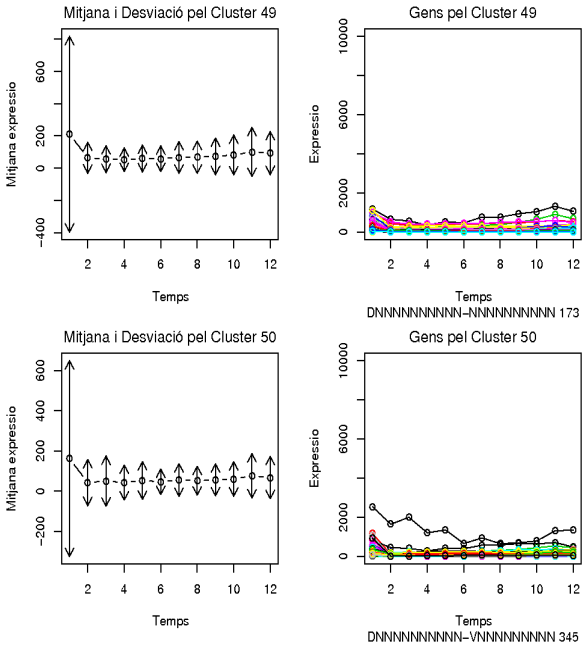

Mitjana i Desviació pel Cluster 41
Mitjana expressio
Temps
Gens pel Cluster 41
Expressio
Temps
DNNNNNNNNIN−NNNNNNNNNN 5

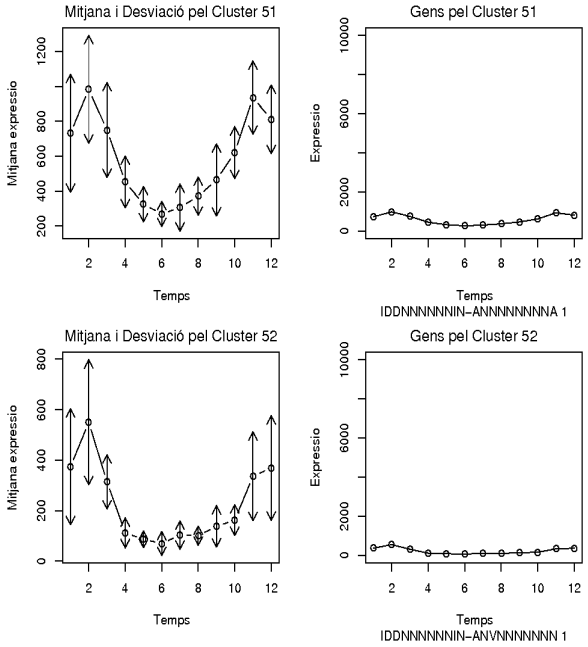

Mitjana i Desviació pel Cluster 42
Mitjana expressio
Temps
Gens pel Cluster 42
Expressio
Temps
DNNNNNNNNIN−NNNNNNNNVA 17

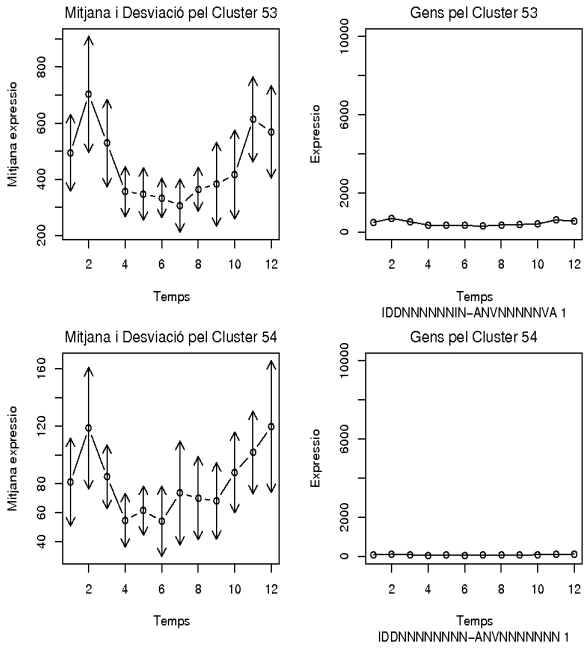

Mitjana i Desviació pel Cluster 43
Mitjana expressio
Temps
Gens pel Cluster 43
Expressio
Temps
DNNNNNNNNIN−NNNNNNNNVN 3
Mitjana i Desviació pel Cluster 44
Mitjana expressio
Temps
Gens pel Cluster 44
Expressio
Temps
DNNNNNNNNIN−VNNNNNNNNA 20

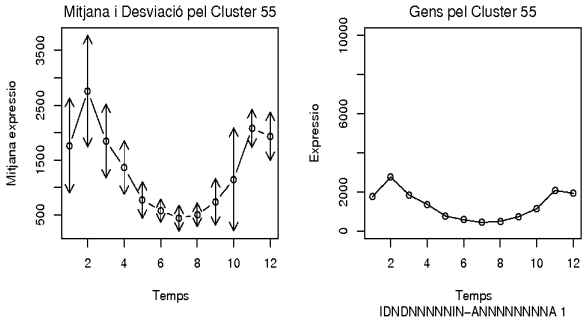

Mitjana i Desviació pel Cluster 45
Mitjana expressio
Temps
Gens pel Cluster 45
Expressio
Temps
DNNNNNNNNIN−VNNNNNNNNN 6
Mitjana i Desviació pel Cluster 46
Mitjana expressio
Temps
Gens pel Cluster 46
Expressio
Temps
DNNNNNNNNIN−VNNNNNNNVA 24

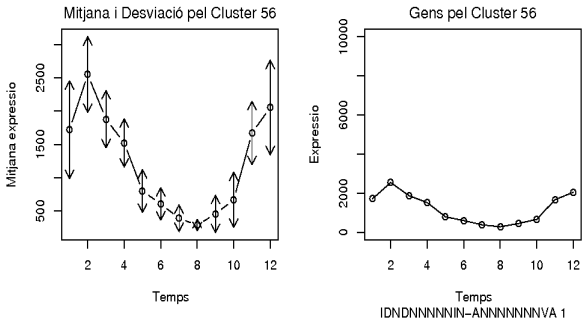

Mitjana i Desviació pel Cluster 47
Mitjana expressio
Temps
Gens pel Cluster 47
Expressio
Temps
DNNNNNNNNIN−VNNNNNNNVN 2
Mitjana i Desviació pel Cluster 48
Mitjana expressio
Temps
Gens pel Cluster 48
Expressio
Temps
DNNNNNNNNNI−NNNNNNNNNN 1

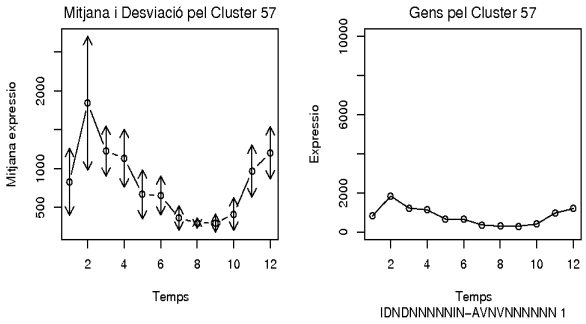

Mitjana i Desviació pel Cluster 49
Mitjana expressio
Temps
Gens pel Cluster 49
Expressio
Temps
DNNNNNNNNNN−NNNNNNNNNN 173
Mitjana i Desviació pel Cluster 50
Mitjana expressio
Temps
Gens pel Cluster 50
Expressio
Temps
DNNNNNNNNNN−VNNNNNNNNN 345

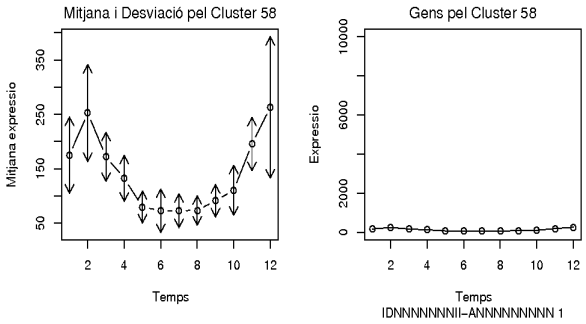

Mitjana i Desviació pel Cluster 51
Mitjana expressio
Temps
Gens pel Cluster 51
Expressio
Temps
IDDNNNNNNIN−ANNNNNNNNA 1
Mitjana i Desviació pel Cluster 52
Mitjana expressio
Temps
Gens pel Cluster 52
Expressio
Temps
IDDNNNNNNIN−ANVNNNNNNN 1

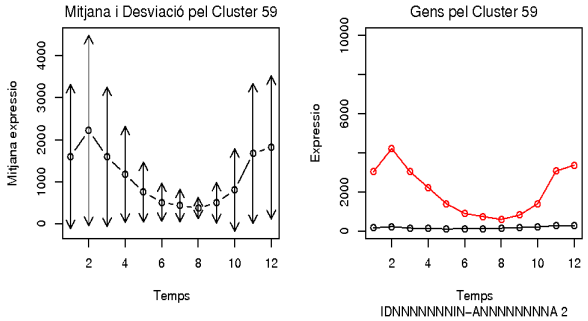

Mitjana i Desviació pel Cluster 53
Mitjana expressio
Temps
Gens pel Cluster 53
Expressio
Temps
IDDNNNNNNIN−ANVNNNNNVA 1
Mitjana i Desviació pel Cluster 54
Mitjana expressio
Temps
Gens pel Cluster 54
Expressio
Temps
IDDNNNNNNNN−ANVNNNNNNN 1


Mitjana i Desviació pel Cluster 55
Mitjana expressio
Temps
Gens pel Cluster 55
Expressio
Temps
IDNDNNNNNIN−ANNNNNNNNA 1

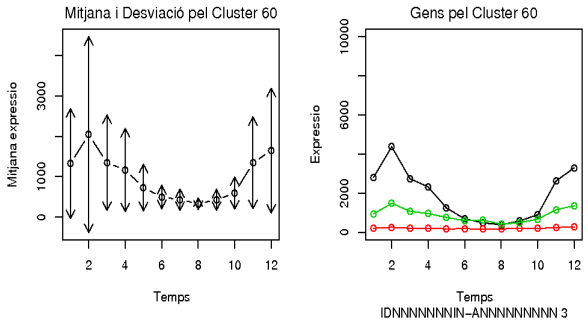

Mitjana i Desviació pel Cluster 56
Mitjana expressio
Temps
Gens pel Cluster 56
Expressio
Temps
IDNDNNNNNIN−ANNNNNNNVA 1


Mitjana i Desviació pel Cluster 57
Mitjana expressio
Temps
Gens pel Cluster 57
Expressio
Temps
IDNDNNNNNIN−AVNVNNNNNN 1

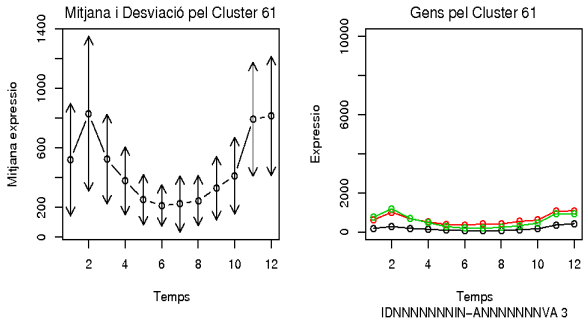

Mitjana i Desviació pel Cluster 58
Mitjana expressio
Temps
Gens pel Cluster 58
Expressio
Temps
IDNNNNNNNII−ANNNNNNNNN 1


Mitjana i Desviació pel Cluster 59
Mitjana expressio
0 1000 2000 3000 4000
2 4 6 8 10 12
Temps

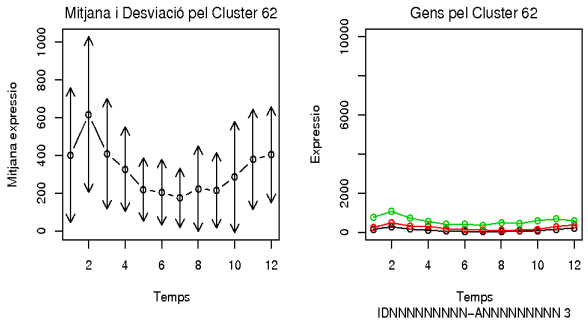
Gens pel Cluster 59
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
IDNNNNNNNIN-ANNNNNNNNA 2


Mitjana i Desviació pel Cluster 60
Mitjana expressio
0 1000 3000
2 4 6 8 10 12
Temps

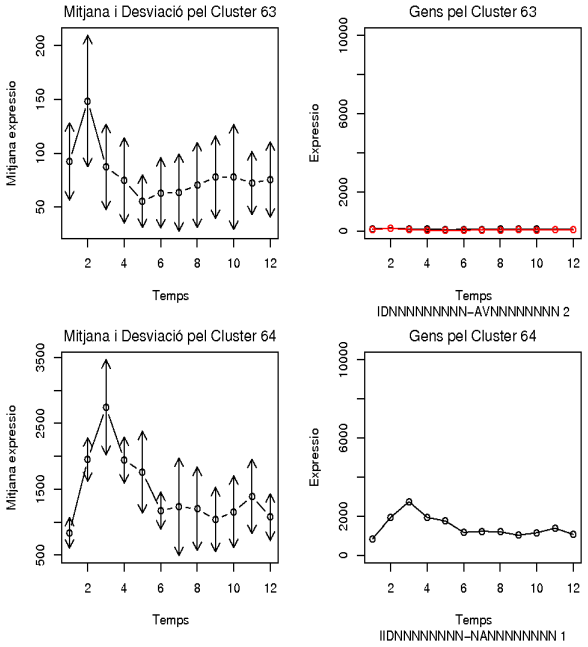
Gens pel Cluster 60
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
IDNNNNNNNIN-ANNNNNNNNN 3

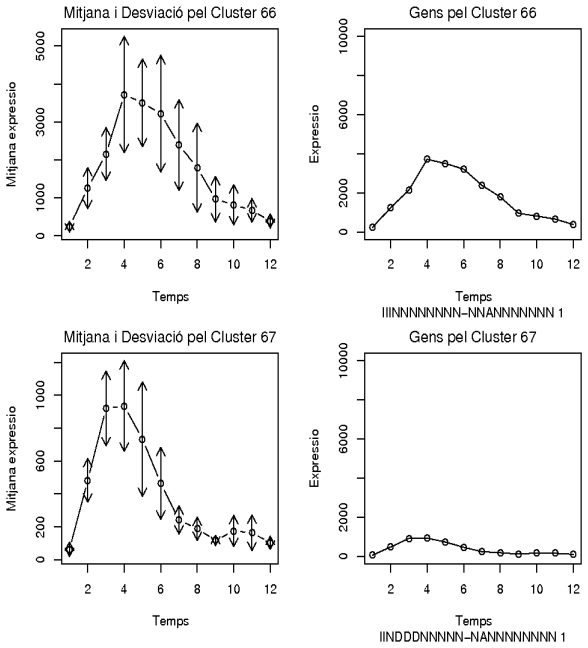

Mitjana i Desviació pel Cluster 61
Mitjana expressio
0 200 600 1000 1400
2 4 6 8 10 12
Temps

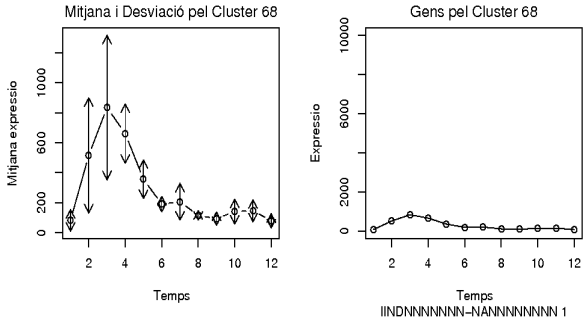
Gens pel Cluster 61
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
IDNNNNNNNIN-ANNNNNNNVA 3

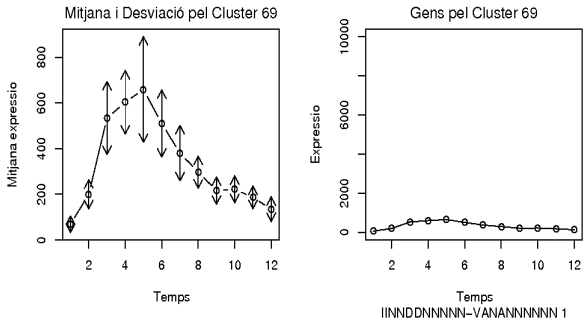

Mitjana i Desviació pel Cluster 62
Mitjana expressio
0 200 400 600 800 1000
2 4 6 8 10 12
Temps

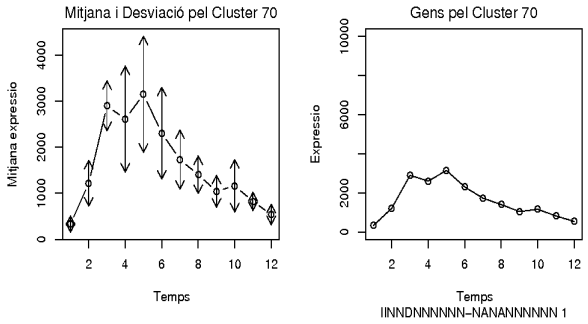
Gens pel Cluster 62
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
IDNNNNNNNNN-ANNNNNNNNN 3

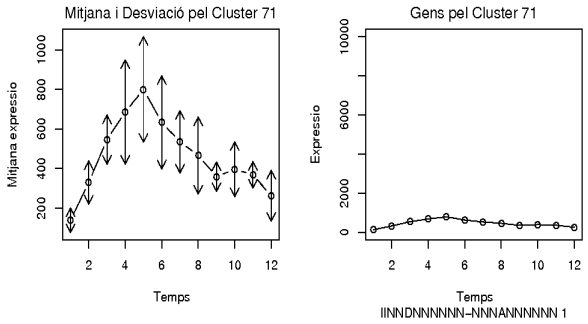

Mitjana i Desviació pel Cluster 63
Mitjana expressio
Temps
Gens pel Cluster 63
Expressio
Temps
IDNNNNNNNNN–AVNNNNNNNNN 2
Mitjana i Desviació pel Cluster 64
Mitjana expressio
Temps
Gens pel Cluster 64
Expressio
Temps
IIDNNNNNNNN–NANNNNNNNN 1

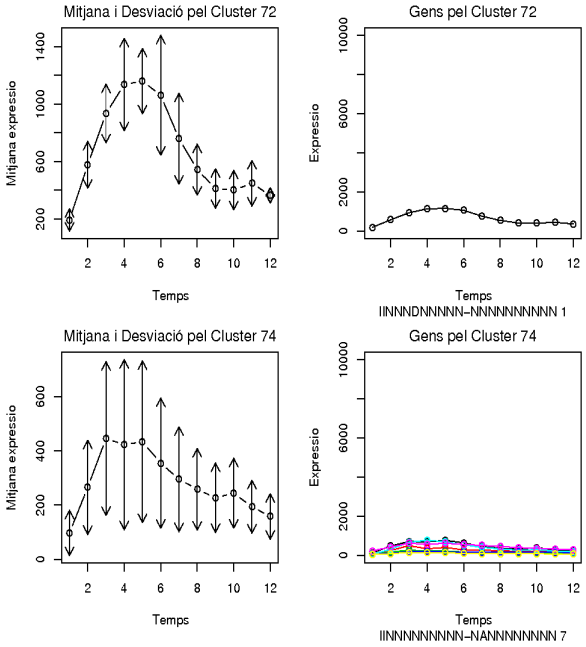

Mitjana i Desviació pel Cluster 66
Mitjana expressio
Temps
Gens pel Cluster 66
Expressio
Temps
IIINNNNNNNN–NNANNNNNNN 1
Mitjana i Desviació pel Cluster 67
Mitjana expressio
Temps
Gens pel Cluster 67
Expressio
Temps
IINDDDNNNNN–NANNNNNNNN 1

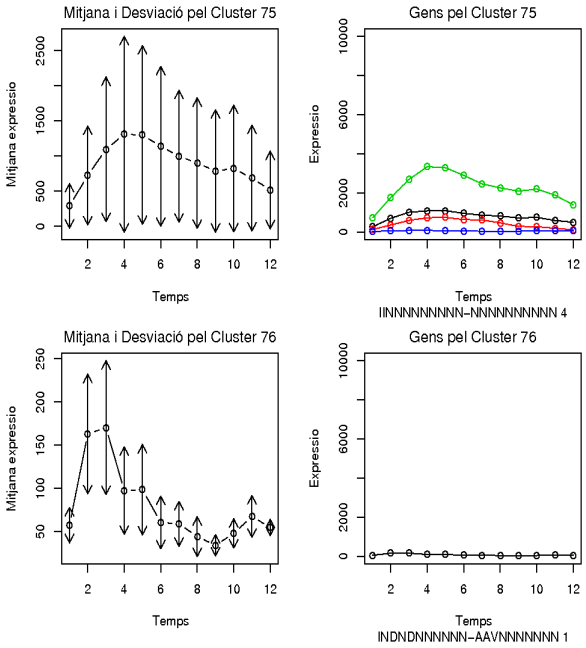

Mitjana i Desviació pel Cluster 68
Mitjana expressio
0
200
600
1000
2
4
6
8
10
12
Temps
Gens pel Cluster 68
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
IINDNNNNNNN–NANNNNNNNN 1

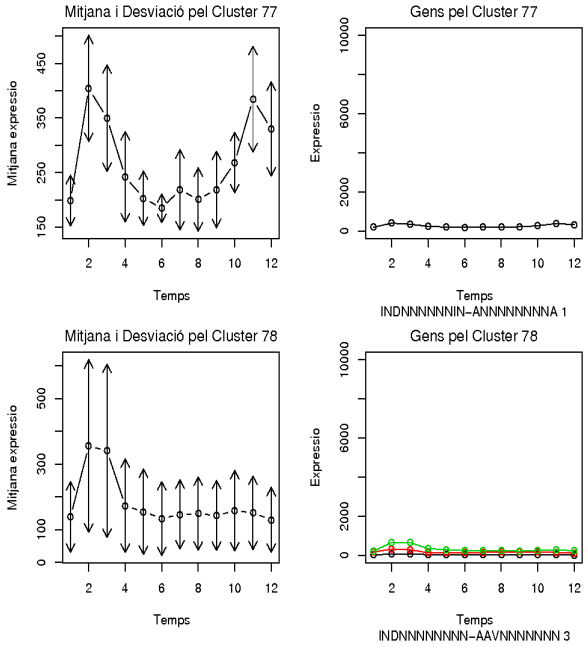

Mitjana i Desviació pel Cluster 69
Mitjana expressio
0
200
400
600
800
2
4
6
8
10
12
Temps
Gens pel Cluster 69
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
IINNDDNNNNN–VANANNNNNN 1

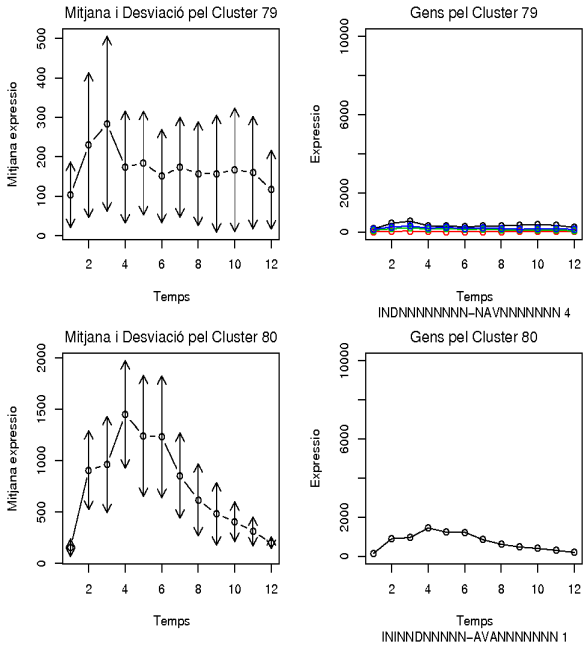

Mitjana i Desviació pel Cluster 70
Mitjana expressio
0
1000
2000
3000
4000
2
4
6
8
10
12
Temps
Gens pel Cluster 70
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
IINNDNNNNNN–NANANNNNNN 1

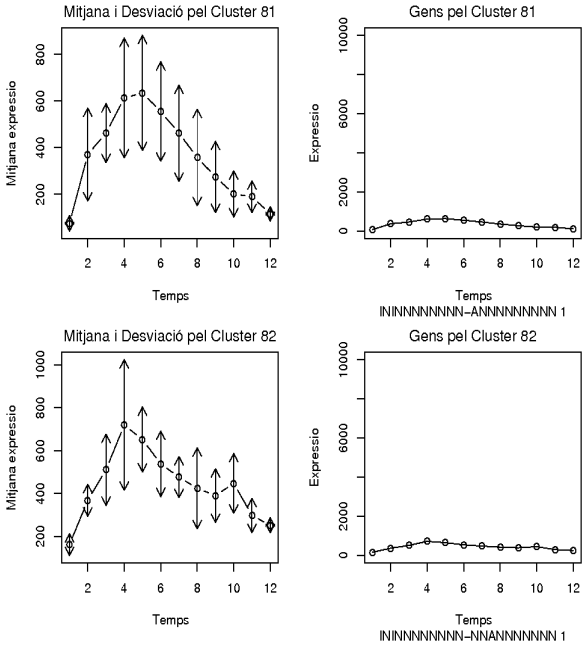

Mitjana i Desviació pel Cluster 71
Mitjana expressio
200
400
600
800
1000
2
4
6
8
10
12
Temps
Gens pel Cluster 71
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
IINNDNNNNNN–NNNANNNNNN 1

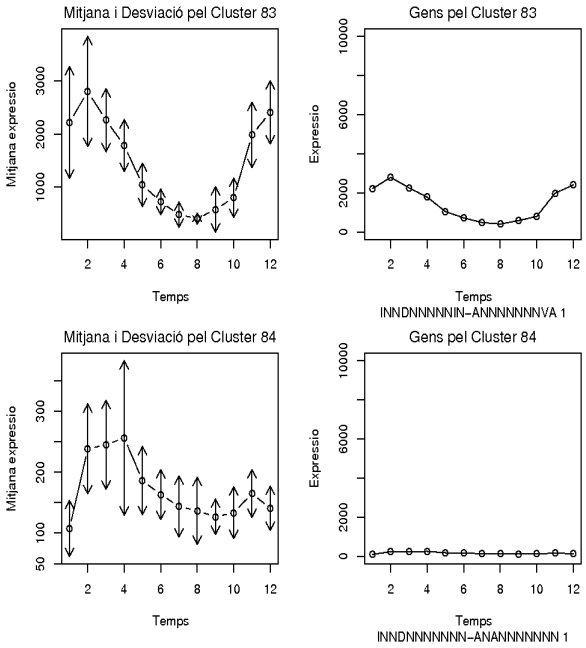

Mitjana i Desviació pel Cluster 72
Mitjana expressio
Temps
Gens pel Cluster 72
Expressio
Temps
IINNNDNNNNN−NNNNNNNNNN 1
Mitjana i Desviació pel Cluster 74
Mitjana expressio
Temps
Gens pel Cluster 74
Expressio
Temps
IINNNNNNNNN−NANNNNNNNN 7

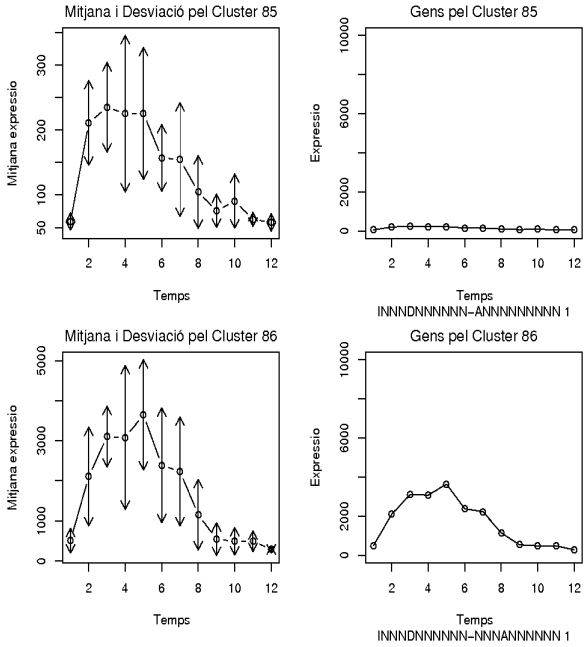

Mitjana i Desviació pel Cluster 75
Mitjana expressio
Temps
Gens pel Cluster 75
Expressio
Temps
IINNNNNNNNN−NNNNNNNNNN 4
Mitjana i Desviació pel Cluster 76
Mitjana expressio
Temps
Gens pel Cluster 76
Expressio
Temps
INDNDNNNNNN−AAVNNNNNNN 1

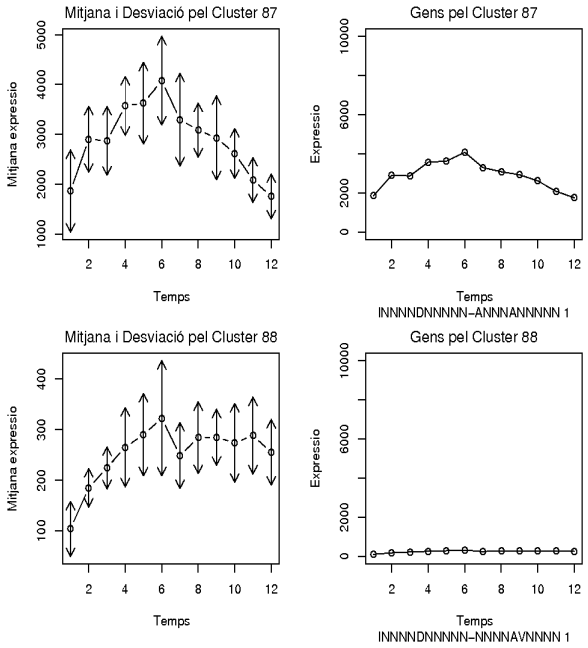

Mitjana i Desviació pel Cluster 77
Mitjana expressio
Temps
Gens pel Cluster 77
Expressio
Temps
INDNNNNNNIN−ANNNNNNNNA 1
Mitjana i Desviació pel Cluster 78
Mitjana expressio
Temps
Gens pel Cluster 78
Expressio
Temps
INDNNNNNNNN−AAVNNNNNNNN 3

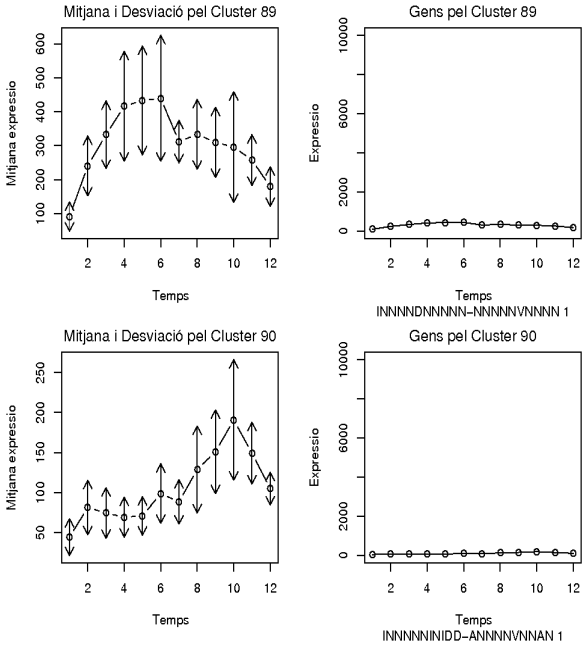

Mitjana i Desviació pel Cluster 79
Mitjana expressio
Temps
Gens pel Cluster 79
Expressio
Temps
INDNNNNNNNN−NAVNNNNNNN 4
Mitjana i Desviació pel Cluster 80
Mitjana expressio
Temps
Gens pel Cluster 80
Expressio
Temps
ININNDNNNNN−AVANNNNNNN 1

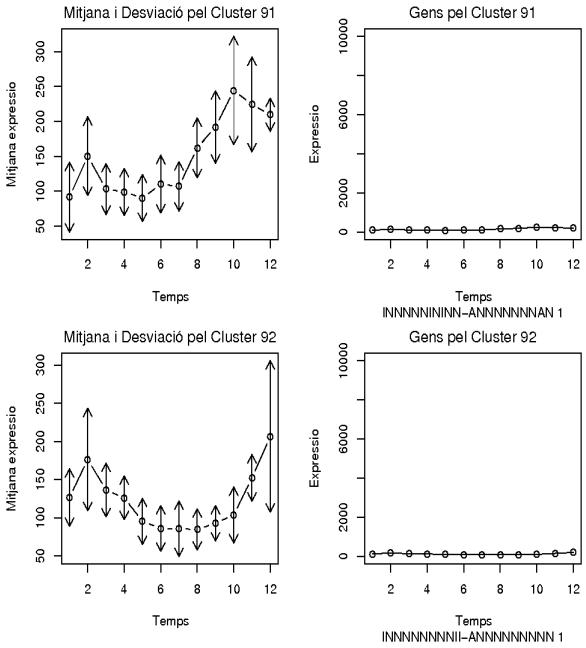

Mitjana i Desviació pel Cluster 81
Mitjana expressio
Temps
Gens pel Cluster 81
Expressio
Temps
ININNNNNNNN−ANNNNNNNNN 1
Mitjana i Desviació pel Cluster 82
Mitjana expressio
Temps
Gens pel Cluster 82
Expressio
Temps
ININNNNNNNN−NNANNNNNNN 1

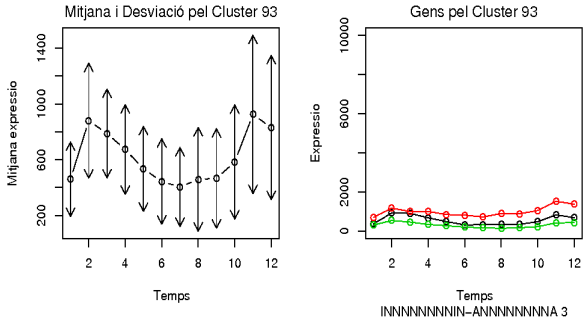

Mitjana i Desviació pel Cluster 83
Mitjana expressio
Temps
Gens pel Cluster 83
Expressio
Temps
INNDNNNNNIN−ANNNNNNNVA 1
Mitjana i Desviació pel Cluster 84
Mitjana expressio
Temps
Gens pel Cluster 84
Expressio
Temps
INNDNNNNNNN−ANANNNNNNN 1


Mitjana i Desviació pel Cluster 85
Mitjana expressio
Temps
Gens pel Cluster 85
Expressio
Temps
INNNDNNNNNN−ANNNNNNNNN 1
Mitjana i Desviació pel Cluster 86
Mitjana expressio
Temps
Gens pel Cluster 86
Expressio
Temps
INNNDNNNNNN−NNNANNNNNN 1

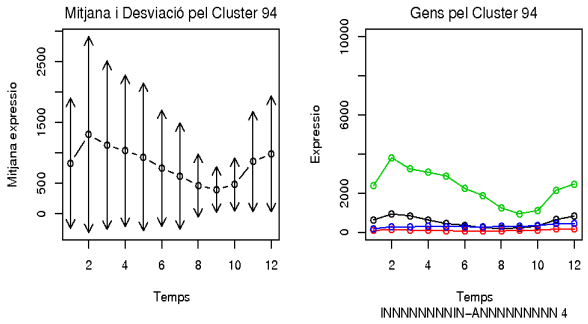

Mitjana i Desviació pel Cluster 87
Mitjana expressio
Temps
Gens pel Cluster 87
Expressio
Temps
INNNNDNNNNN−ANNNANNNNN 1
Mitjana i Desviació pel Cluster 88
Mitjana expressio
Temps
Gens pel Cluster 88
Expressio
Temps
INNNNDNNNNN−NNNNAVNNNN 1


Mitjana i Desviació pel Cluster 89
Mitjana expressio
Temps
Gens pel Cluster 89
Expressio
Temps
INNNNDNNNNN−NNNNNVNNNN 1
Mitjana i Desviació pel Cluster 90
Mitjana expressio
Temps
Gens pel Cluster 90
Expressio
Temps
INNNNNINIDD−ANNNNVNNAN 1

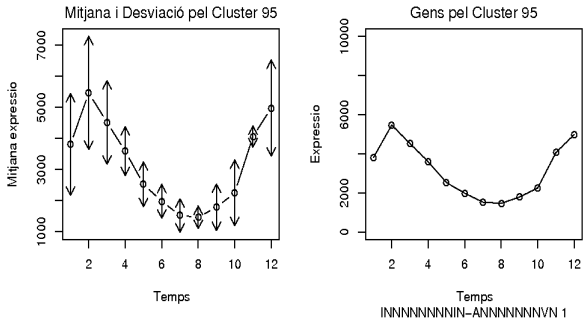

Mitjana i Desviació pel Cluster 91
Mitjana expressio
Temps
Gens pel Cluster 91
Expressio
Temps
INNNNNININN−ANNNNNNNAN 1
Mitjana i Desviació pel Cluster 92
Mitjana expressio
Temps
Gens pel Cluster 92
Expressio
Temps
INNNNNNNNII−ANNNNNNNNN 1


Mitjana i Desviació pel Cluster 93
Mitjana expressio
200
600
1000
1400
2
4
6
8
10
12
Temps

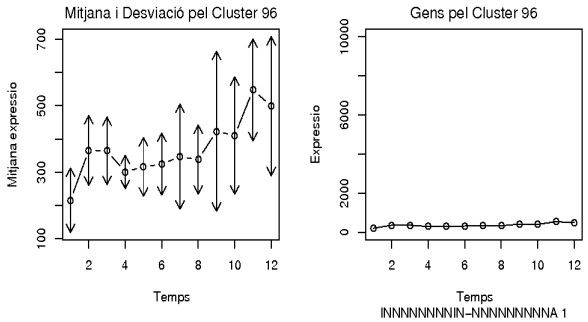
Gens pel Cluster 93
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNIN−ANNNNNNNNA 3


Mitjana i Desviació pel Cluster 94
Mitjana expressio
0
500
1500
2500
2
4
6
8
10
12
Temps

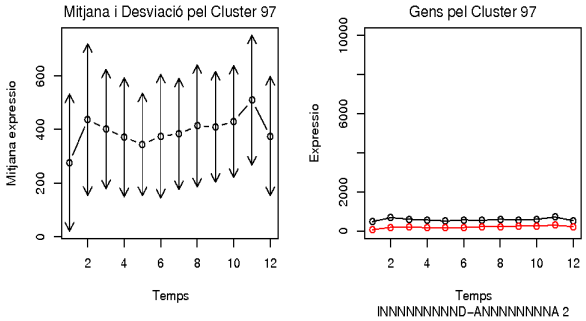
Gens pel Cluster 94
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNIN−ANNNNNNNNN 4

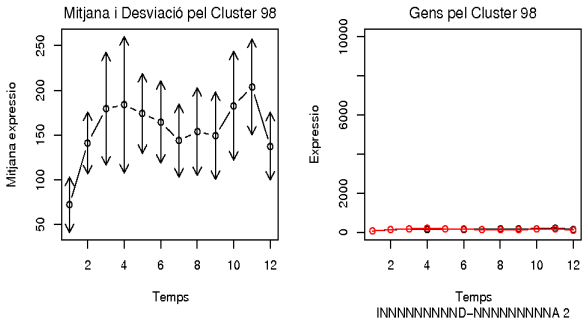

Mitjana i Desviació pel Cluster 95
Mitjana expressio
1000
3000
5000
7000
2
4
6
8
10
12
Temps

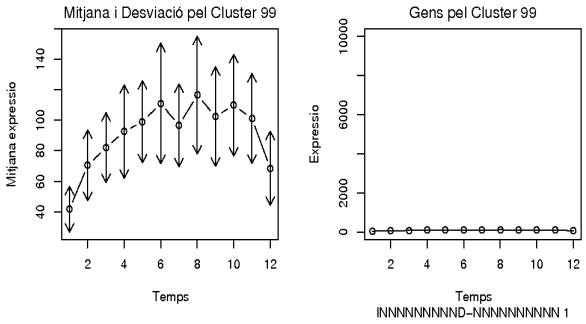
Gens pel Cluster 95
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNIN−ANNNNNNNVN 1

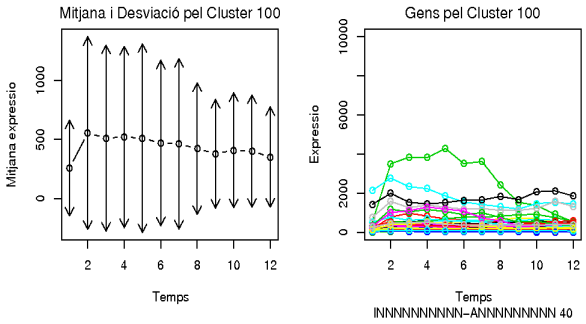

Mitjana i Desviació pel Cluster 96
Mitjana expressio
100
300
500
700
2
4
6
8
10
12
Temps

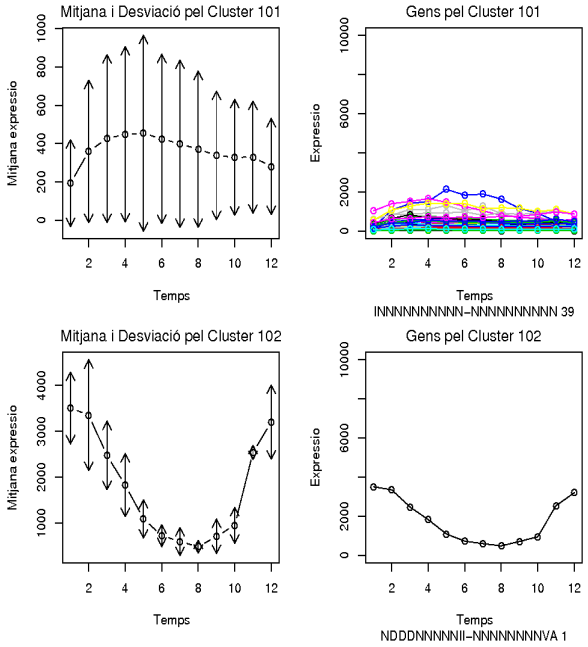
Gens pel Cluster 96
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNIN−NNNNNNNNNA 1

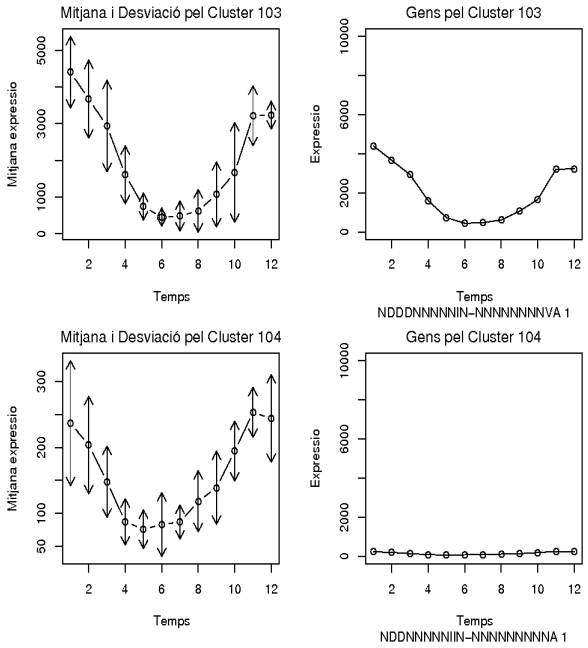

Mitjana i Desviació pel Cluster 97
Mitjana expressio
0
200
400
600
2
4
6
8
10
12
Temps
Gens pel Cluster 97
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNND−ANNNNNNNNA 2

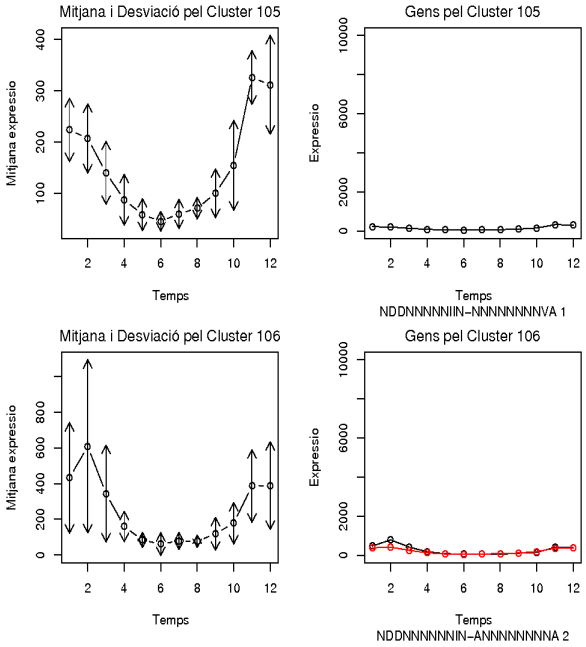

Mitjana i Desviació pel Cluster 98
Mitjana expressio
50
100
150
200
250
2
4
6
8
10
12
Temps
Gens pel Cluster 98
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNND−NNNNNNNNNA 2

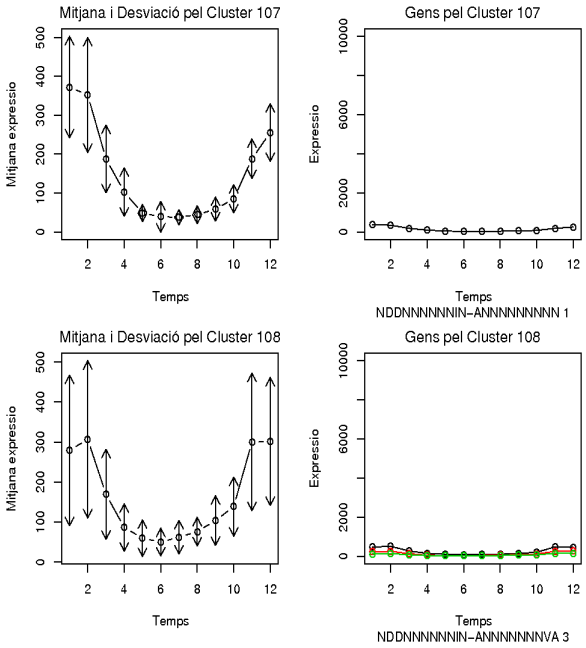

Mitjana i Desviació pel Cluster 99
Mitjana expressio
40
60
80
100
140
2
4
6
8
10
12
Temps
Gens pel Cluster 99
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNND−NNNNNNNNNN 1

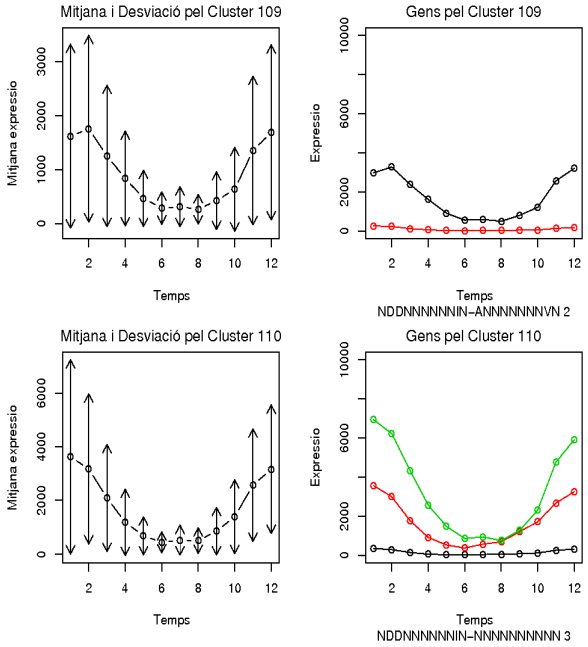

Mitjana i Desviació pel Cluster 100
Mitjana expressio
0
500
1000
2
4
6
8
10
12
Temps
Gens pel Cluster 100
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
INNNNNNNNNN−ANNNNNNNNN 40

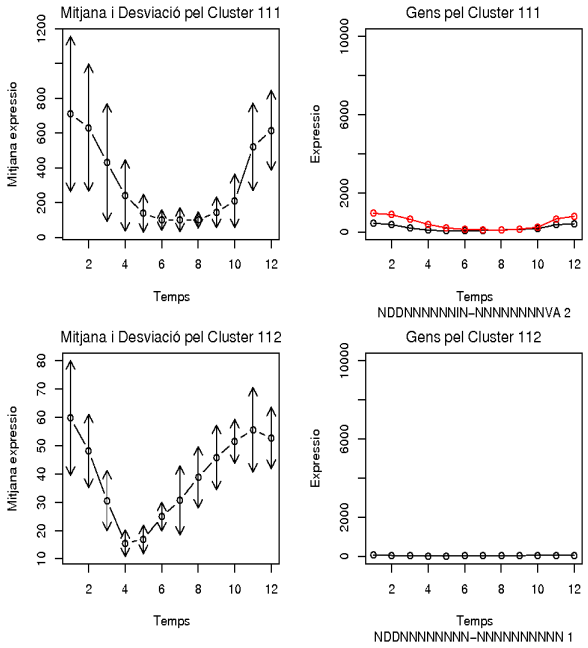

Mitjana i Desviació pel Cluster 101
Mitjana expressio
Temps
Gens pel Cluster 101
Expressio
Temps
INNNNNNNNNN−NNNNNNNNNN 39
Mitjana i Desviació pel Cluster 102
Mitjana expressio
Temps
Gens pel Cluster 102
Expressio
Temps
NDDDNNNNNII−NNNNNNNNVA 1

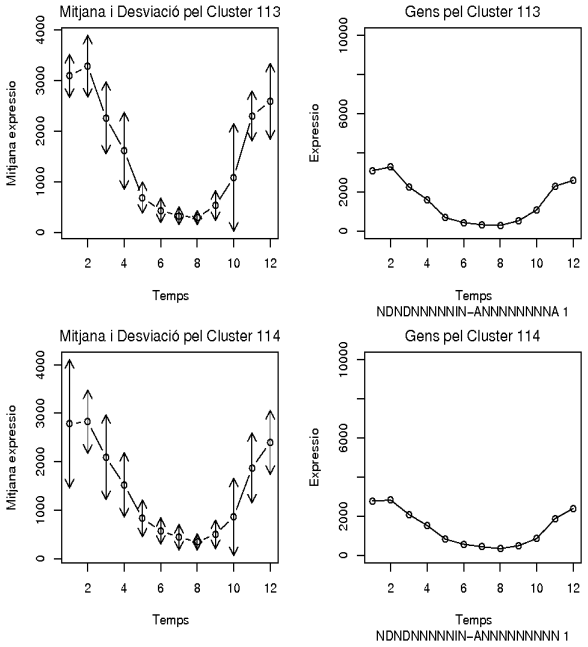

Mitjana i Desviació pel Cluster 103
Mitjana expressio
Temps
Gens pel Cluster 103
Expressio
Temps
NDDDNNNNNIN−NNNNNNNNVA 1
Mitjana i Desviació pel Cluster 104
Mitjana expressio
Temps
Gens pel Cluster 104
Expressio
Temps
NDDNNNNNIIN−NNNNNNNNNA 1

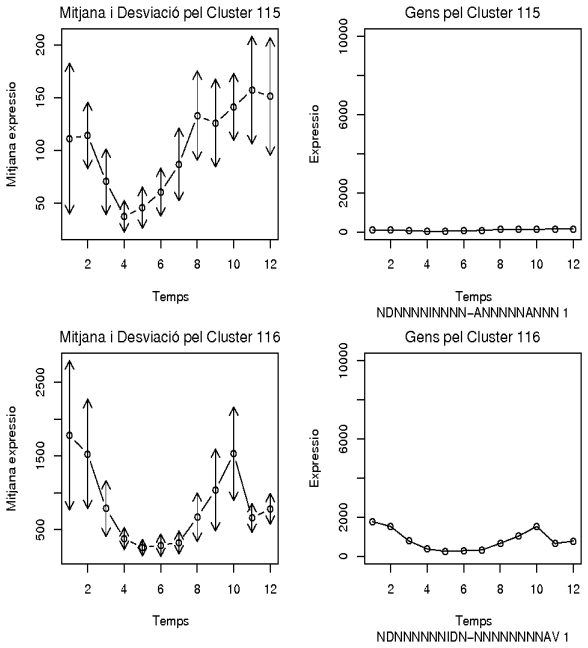

Mitjana i Desviació pel Cluster 105
Mitjana expressio
Temps
Gens pel Cluster 105
Expressio
Temps
NDDNNNNNIIN−NNNNNNNNVA 1
Mitjana i Desviació pel Cluster 106
Mitjana expressio
Temps
Gens pel Cluster 106
Expressio
Temps
NDDNNNNNNIN−ANNNNNNNNA 2

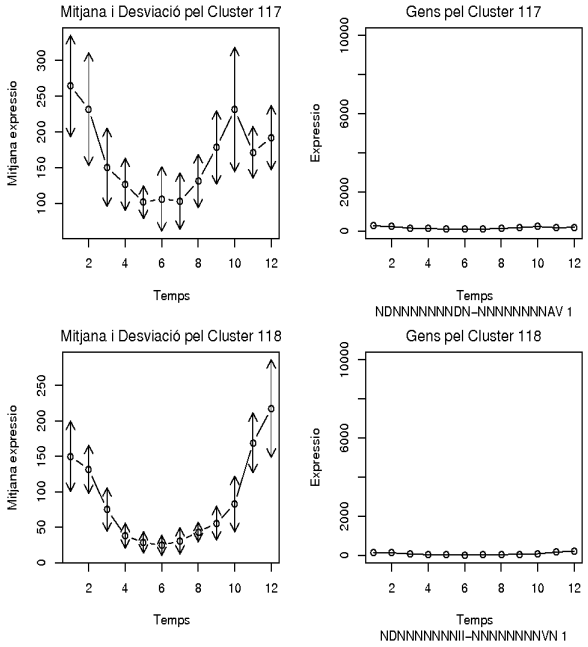

Mitjana i Desviació pel Cluster 107
Mitjana expressio
Temps
Gens pel Cluster 107
Expressio
Temps
NDDNNNNNNIN−ANNNNNNNNN 1
Mitjana i Desviació pel Cluster 108
Mitjana expressio
Temps
Gens pel Cluster 108
Expressio
Temps
NDDNNNNNNIN−ANNNNNNNVA 3

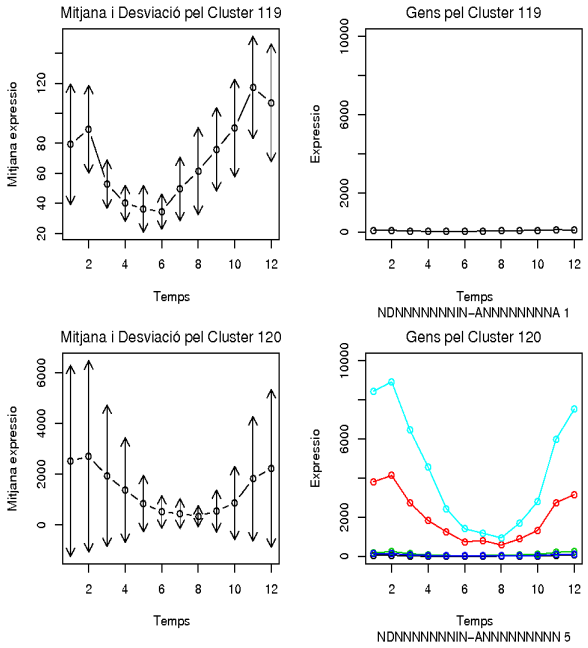

Mitjana i Desviació pel Cluster 109
Mitjana expressio
Temps
Gens pel Cluster 109
Expressio
Temps
NDDNNNNNNIN−ANNNNNNNVN 2
Mitjana i Desviació pel Cluster 110
Mitjana expressio
Temps
Gens pel Cluster 110
Expressio
Temps
NDDNNNNNNIN−NNNNNNNNN 3

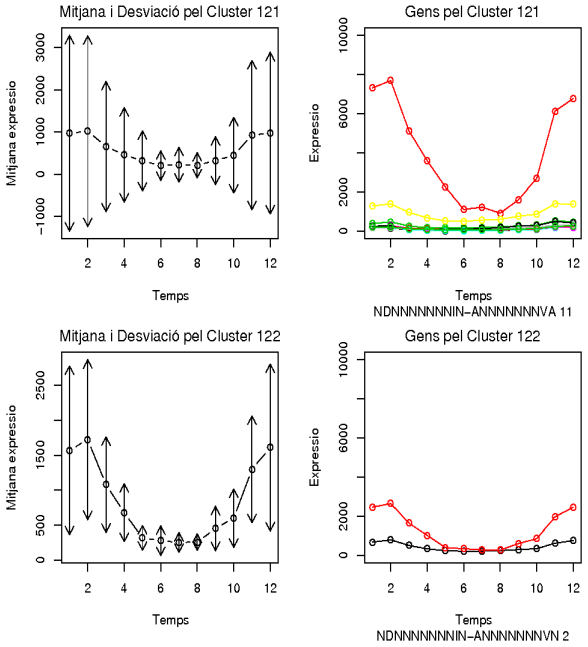

Mitjana i Desviació pel Cluster 111
Mitjana expressio
Temps
Gens pel Cluster 111
Expressio
Temps
NDDNNNNNNIN−NNNNNNNVA 2
Mitjana i Desviació pel Cluster 112
Mitjana expressio
Temps
Gens pel Cluster 112
Expressio
Temps
NDDNNNNNNN−NNNNNNNNN 1

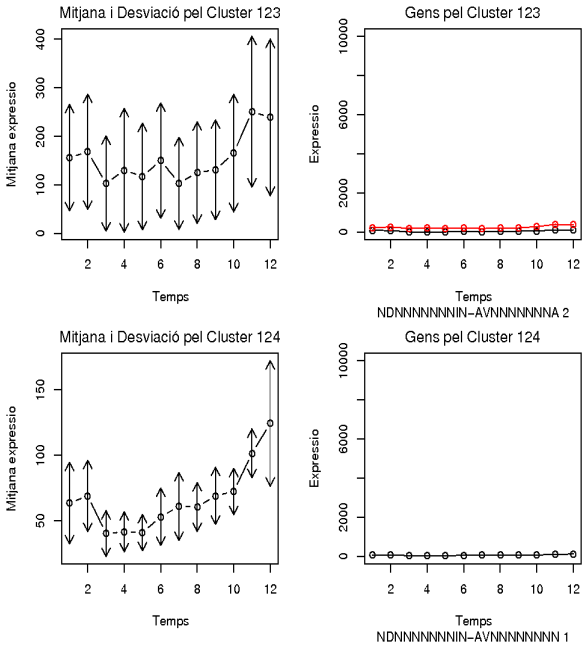

Mitjana i Desviació pel Cluster 113
Mitjana expressio
Temps
Gens pel Cluster 113
Expressio
Temps
NDNDNNNNNIN−ANNNNNNNNA 1
Mitjana i Desviació pel Cluster 114
Mitjana expressio
Temps
Gens pel Cluster 114
Expressio
Temps
NDNDNNNNNIN−ANNNNNNNNN 1

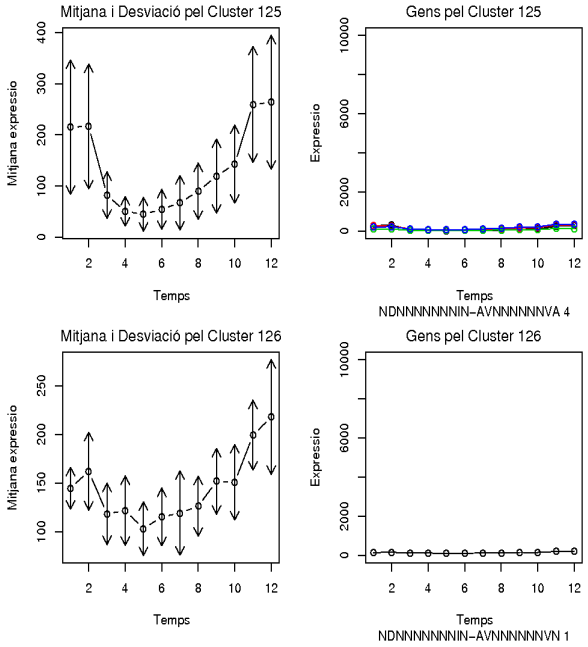

Mitjana i Desviació pel Cluster 115
Mitjana expressio
Temps
Gens pel Cluster 115
Expressio
Temps
NDNNNNINNNN−ANNNNNANNN 1
Mitjana i Desviació pel Cluster 116
Mitjana expressio
Temps
Gens pel Cluster 116
Expressio
Temps
NDNNNNNNIDN−NNNNNNNNAV 1

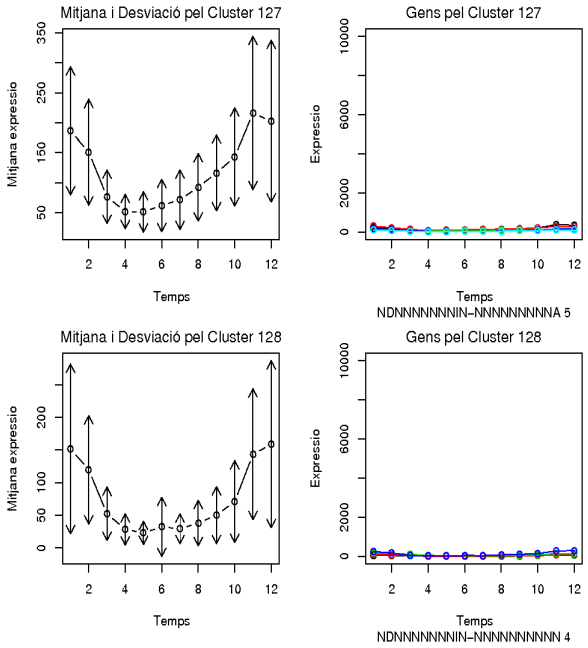

Mitjana i Desviació pel Cluster 117
Mitjana expressio
Temps
Gens pel Cluster 117
Expressio
Temps
NDNNNNNNNDN−NNNNNNNAV 1
Mitjana i Desviació pel Cluster 118
Mitjana expressio
Temps
Gens pel Cluster 118
Expressio
Temps
NDNNNNNNNII−NNNNNNNNVN 1

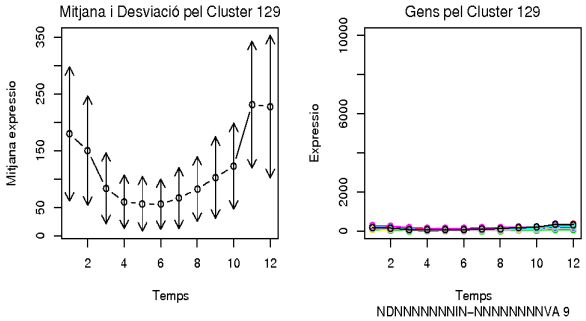

Mitjana i Desviació pel Cluster 119
Mitjana expressio
Temps
Gens pel Cluster 119
Expressio
Temps
NDNNNNNNNIN−ANNNNNNNNA 1
Mitjana i Desviació pel Cluster 120
Mitjana expressio
Temps
Gens pel Cluster 120
Expressio
Temps
NDNNNNNNNIN−ANNNNNNNNN 5

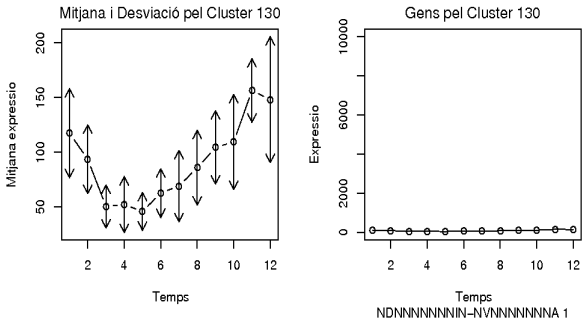

Mitjana i Desviació pel Cluster 121
Mitjana expressio
Temps
Gens pel Cluster 121
Expressio
Temps
NDNNNNNNNIN−ANNNNNNNVA 11
Mitjana i Desviació pel Cluster 122
Mitjana expressio
Temps
Gens pel Cluster 122
Expressio
Temps
NDNNNNNNNIN−ANNNNNNNVN 2

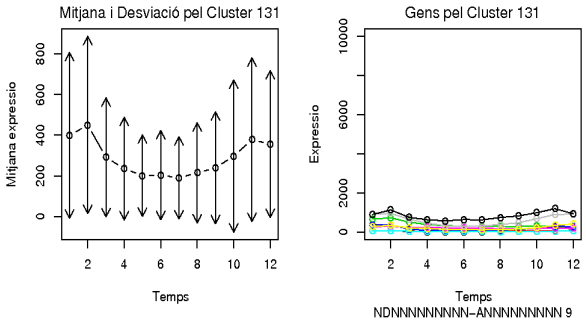

Mitjana i Desviació pel Cluster 123
Mitjana expressio
Temps
Gens pel Cluster 123
Expressio
Temps
NDNNNNNNNIN−AVNNNNNNNA 2
Mitjana i Desviació pel Cluster 124
Mitjana expressio
Temps
Gens pel Cluster 124
Expressio
Temps
NDNNNNNNNIN−AVNNNNNNNN 1

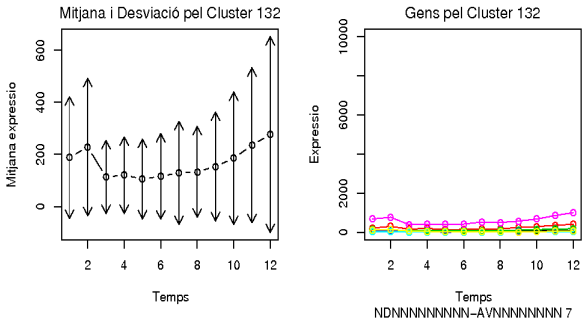

Mitjana i Desviació pel Cluster 125
Mitjana expressio
Temps
Gens pel Cluster 125
Expressio
Temps
NDNNNNNNNIN−AVNNNNNNVA 4
Mitjana i Desviació pel Cluster 126
Mitjana expressio
Temps
Gens pel Cluster 126
Expressio
Temps
NDNNNNNNNIN−AVNNNNNNVN 1

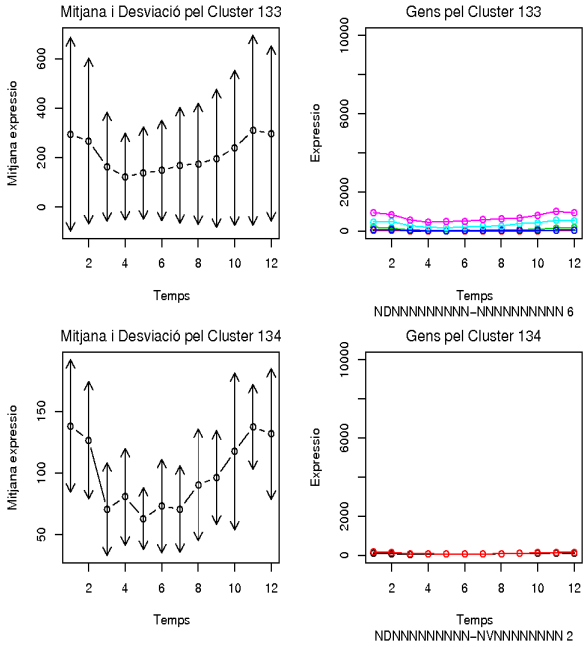

Mitjana i Desviació pel Cluster 127
Mitjana expressio
Temps
Gens pel Cluster 127
Expressio
Temps
NDNNNNNNNIN−NNNNNNNNNA 5
Mitjana i Desviació pel Cluster 128
Mitjana expressio
Temps
Gens pel Cluster 128
Expressio
Temps
NDNNNNNNNIN−NNNNNNNNNN 4

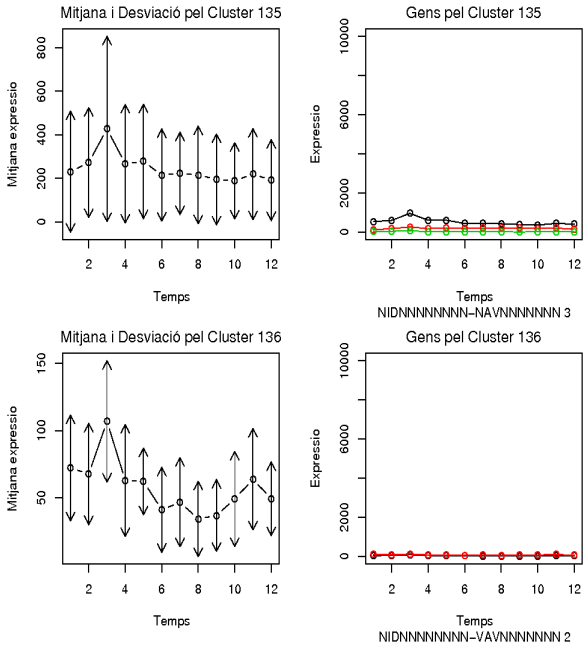

Mitjana i Desviació pel Cluster 129
Mitjana expressio
Temps
Gens pel Cluster 129
Expressio
Temps
NDNNNNNNNIN−NNNNNNNVA 9

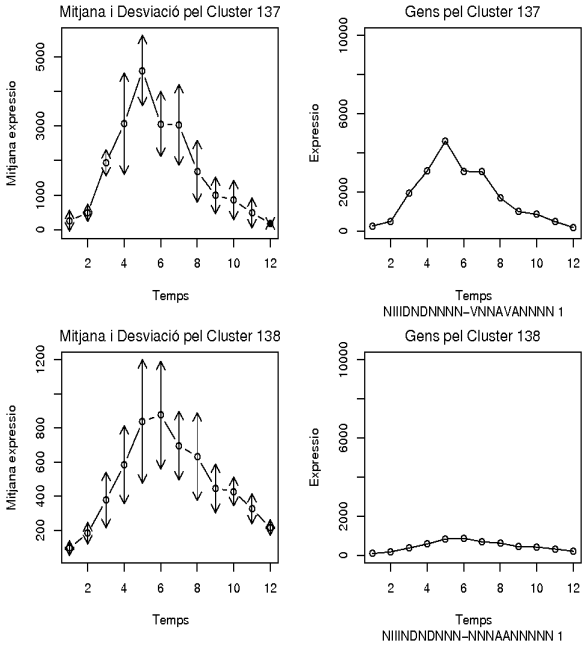

Mitjana i Desviació pel Cluster 130
Mitjana expressio
Temps
Gens pel Cluster 130
Expressio
Temps
NDNNNNNNNIN−NVNNNNNNA 1

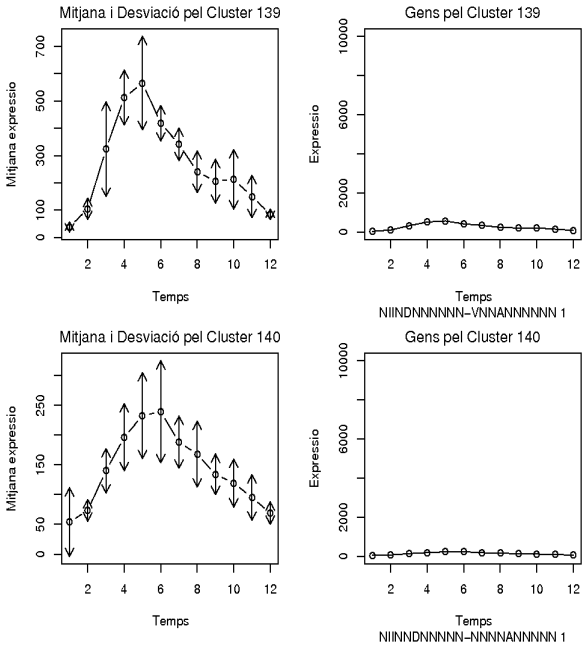

Mitjana i Desviació pel Cluster 131
Mitjana expressio
Temps
Gens pel Cluster 131
Expressio
Temps
NDNNNNNNNN−ANNNNNNNN 9

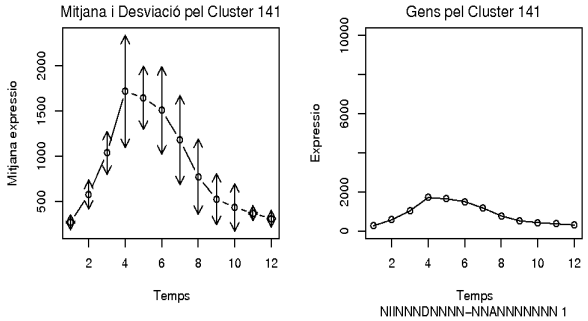

Mitjana i Desviació pel Cluster 132
Mitjana expressio
Temps
Gens pel Cluster 132
Expressio
Temps
NDNNNNNNNN−AVNNNNNNN 7

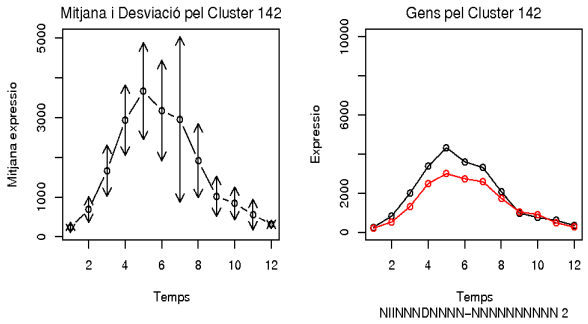

Mitjana i Desviació pel Cluster 133
Mitjana expressio
0
200
400
600
2
4
6
8
10
12
Temps
Gens pel Cluster 133
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
NDNNNNNNNNN−NNNNNNNNNN 6
Mitjana i Desviació pel Cluster 134
Mitjana expressio
50
100
150
2
4
6
8
10
12
Temps
Gens pel Cluster 134
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
NDNNNNNNNNN−NVNNNNNNNN 2

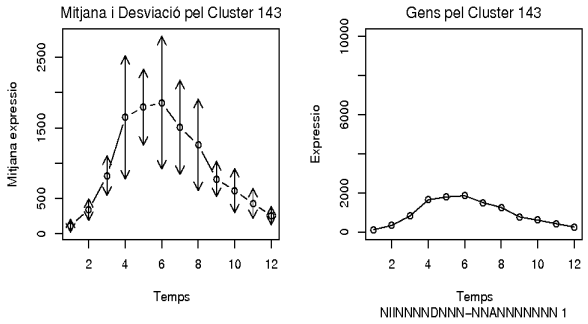

Mitjana i Desviació pel Cluster 135
Mitjana expressio
0
200
400
600
800
2
4
6
8
10
12
Temps
Gens pel Cluster 135
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
NIDNNNNNNNN−NAVNNNNNNN 3
Mitjana i Desviació pel Cluster 136
Mitjana expressio
50
100
150
2
4
6
8
10
12
Temps
Gens pel Cluster 136
Expressio
0
2000
6000
10000
2
4
6
8
10
12
Temps
NIDNNNNNNNN−VAVNNNNNNN 2

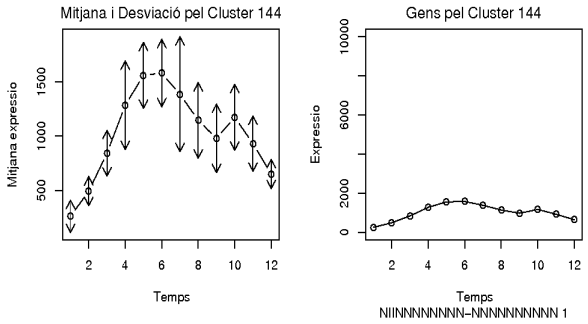

Mitjana i Desviació pel Cluster 137
Mitjana expressio
0 1000 3000 5000
2 4 6 8 10 12
Temps
Gens pel Cluster 137
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NIIIDNDNNNN−VNNAVANNNN 1
Mitjana i Desviació pel Cluster 138
Mitjana expressio
200 400 600 800 1200
2 4 6 8 10 12
Temps
Gens pel Cluster 138
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NIIINDNDNNN−NNNAANNNNN 1

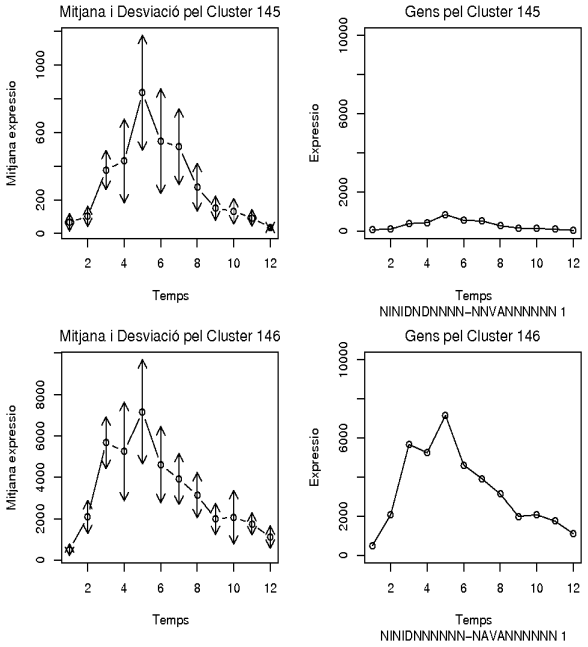

Mitjana i Desviació pel Cluster 139
Mitjana expressio
0 100 300 500 700
2 4 6 8 10 12
Temps
Gens pel Cluster 139
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NIINDNNNNNN−VNNANNNNNN 1
Mitjana i Desviació pel Cluster 140
Mitjana expressio
0 50 150 250
2 4 6 8 10 12
Temps
Gens pel Cluster 140
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NIINNDNNNNN−NNNNANNNNN 1

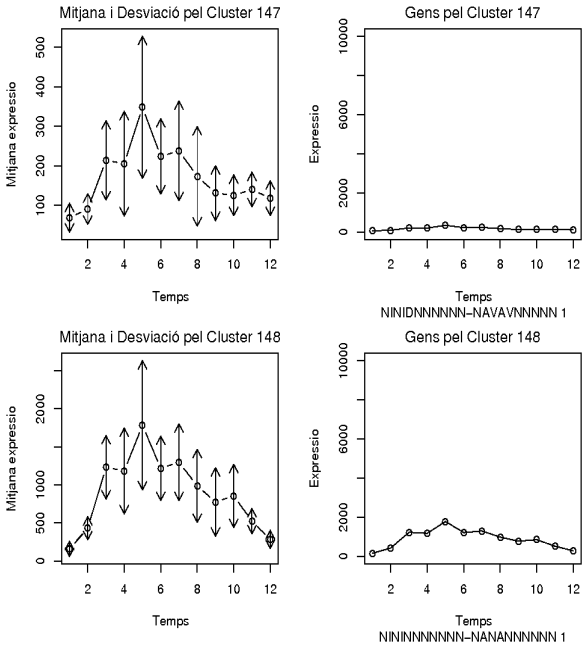

Mitjana i Desviació pel Cluster 141
Mitjana expressio
Temps
Gens pel Cluster 141
Expressio
Temps
NIINNNDNNNN–NNANNNNNNN 1

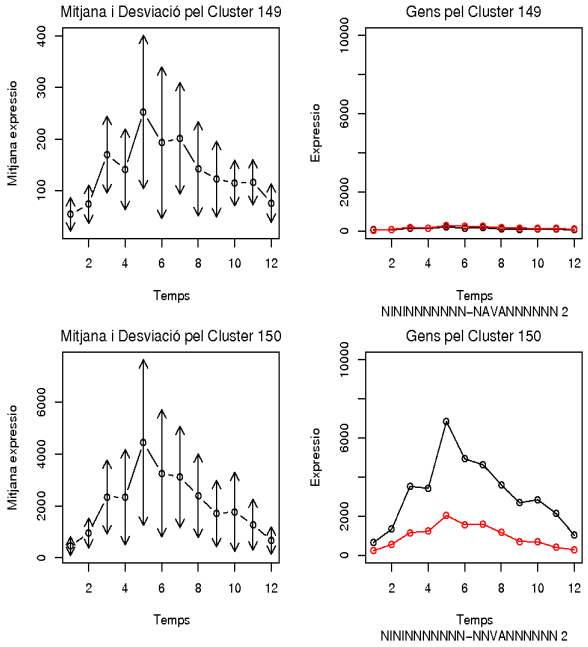

Mitjana i Desviació pel Cluster 142
Mitjana expressio
Temps
Gens pel Cluster 142
Expressio
Temps
NIINNNDNNNN–NNNNNNNNNN 2

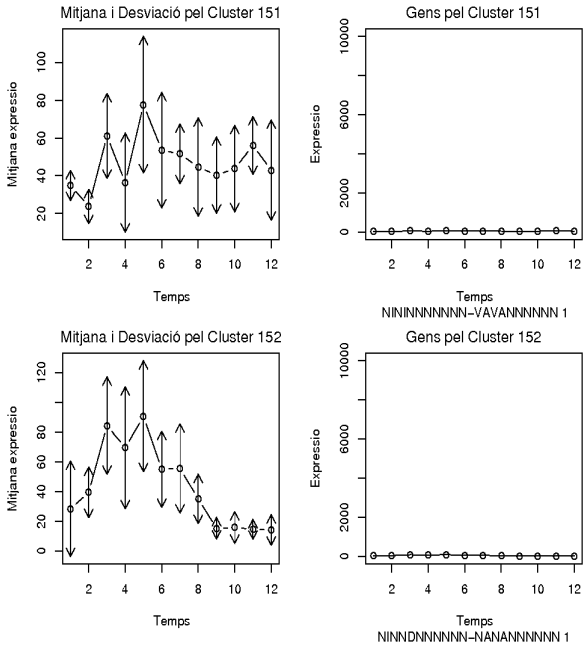

Mitjana i Desviació pel Cluster 143
Mitjana expressio
Temps
Gens pel Cluster 143
Expressio
Temps
NIINNNNDNNN–NNANNNNNNN 1

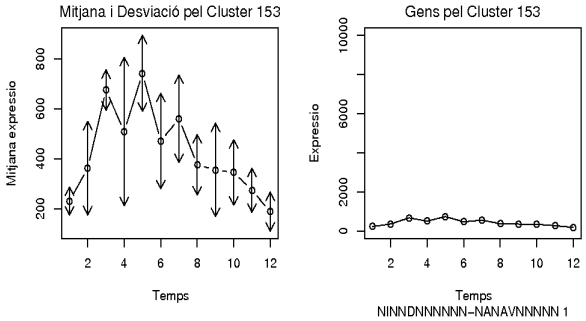

Mitjana i Desviació pel Cluster 144
Mitjana expressio
Temps
Gens pel Cluster 144
Expressio
Temps
NIINNNNNNNN–NNNNNNNNNN 1

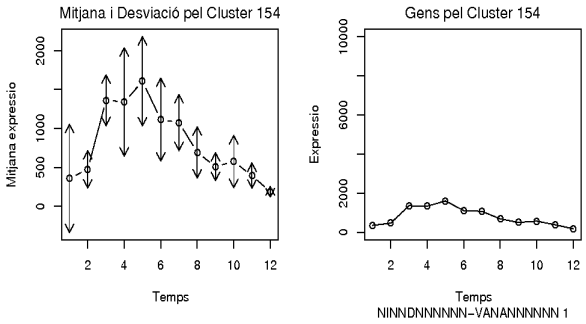

Mitjana i Desviació pel Cluster 145
Mitjana expressio
Temps
Gens pel Cluster 145
Expressio
Temps
NINIDNDNNNN–NNVANNNNNN 1
Mitjana i Desviació pel Cluster 146
Mitjana expressio
Temps
Gens pel Cluster 146
Expressio
Temps
NINIDNNNNNN–NAVANNNNNN 1

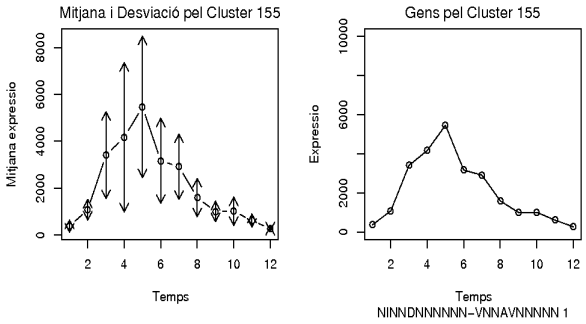

Mitjana i Desviació pel Cluster 147
Mitjana expressio
Temps
Gens pel Cluster 147
Expressio
Temps
NINIDNNNNNN–NAVAVNNNNN 1
Mitjana i Desviació pel Cluster 148
Mitjana expressio
Temps
Gens pel Cluster 148
Expressio
Temps
NININNNNNNN–NANANNNNNN 1

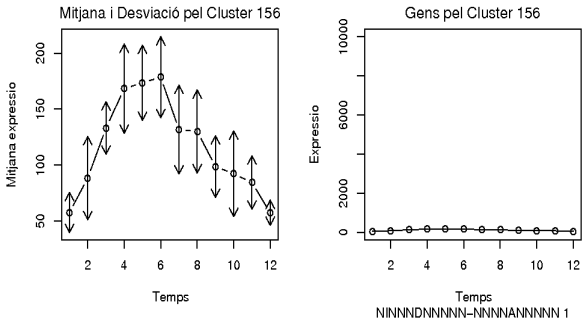

Mitjana i Desviació pel Cluster 149
Mitjana expressio
Temps
Gens pel Cluster 149
Expressio
Temps
NINININNNNNN−NAVANNNNNN 2
Mitjana i Desviació pel Cluster 150
Mitjana expressio
Temps
Gens pel Cluster 150
Expressio
Temps
NINININNNNNN−NNVANNNNNN 2

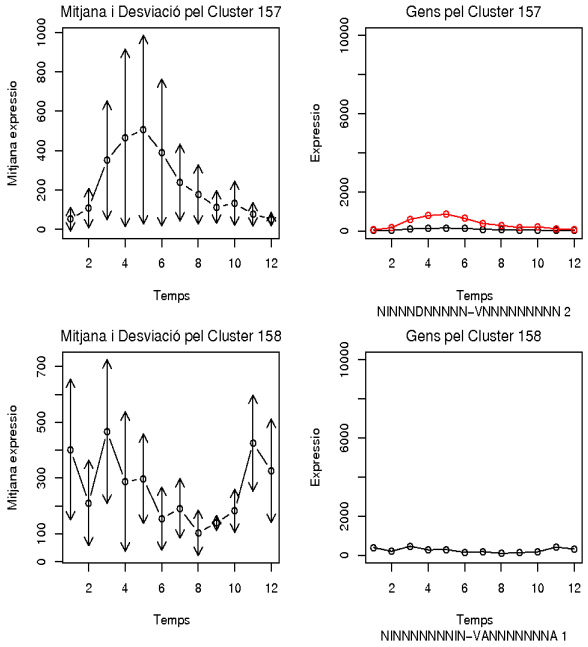

Mitjana i Desviació pel Cluster 151
Mitjana expressio
Temps
Gens pel Cluster 151
Expressio
Temps
NINININNNNNN−VAVANNNNNN 1
Mitjana i Desviació pel Cluster 152
Mitjana expressio
Temps
Gens pel Cluster 152
Expressio
Temps
NINNDNNNNNN−NANANNNNNN 1

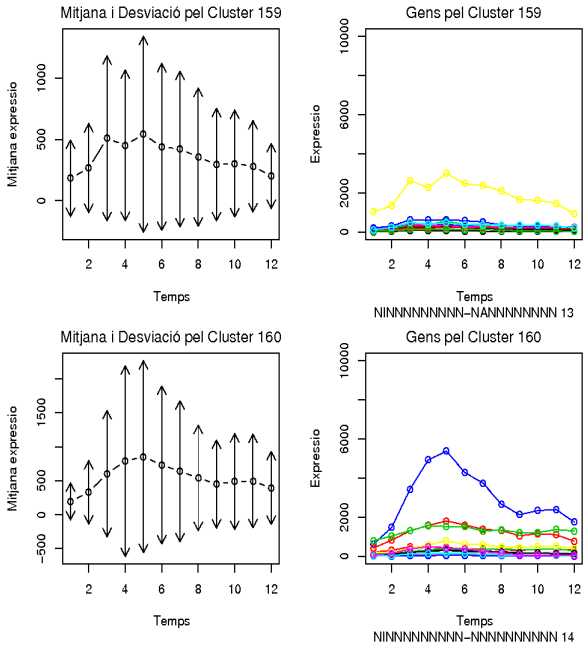

Mitjana i Desviació pel Cluster 153
Mitjana expressio
Temps
Gens pel Cluster 153
Expressio
Temps
NINNDNNNNNN−NANAVNNNNN 1

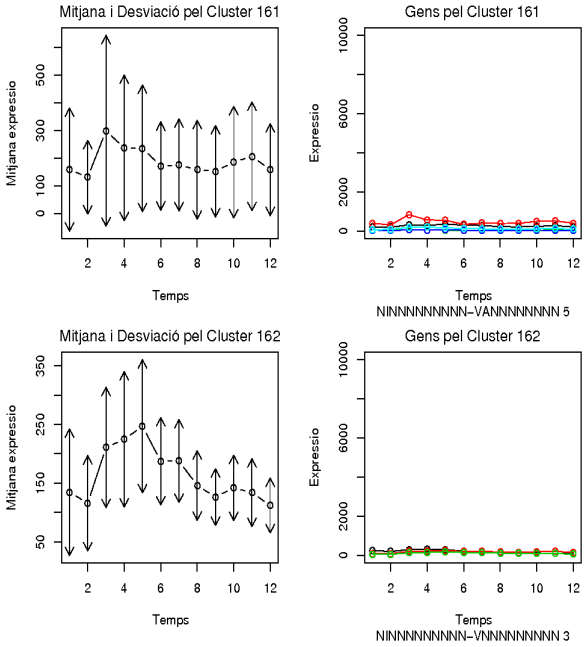

Mitjana i Desviació pel Cluster 154
Mitjana expressio
Temps
Gens pel Cluster 154
Expressio
Temps
NINNDNNNNNN−VANANNNNNN 1

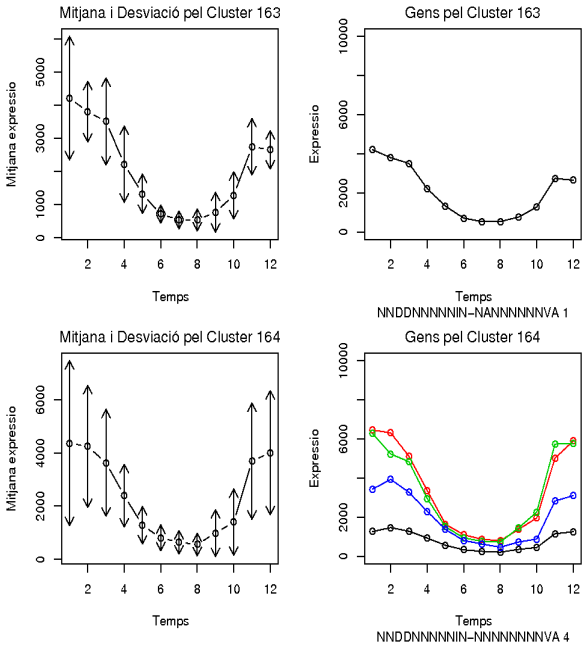

Mitjana i Desviació pel Cluster 155
Mitjana expressio
Temps
Gens pel Cluster 155
Expressio
Temps
NINNDNNNNNN−VNNAVNNNNN 1

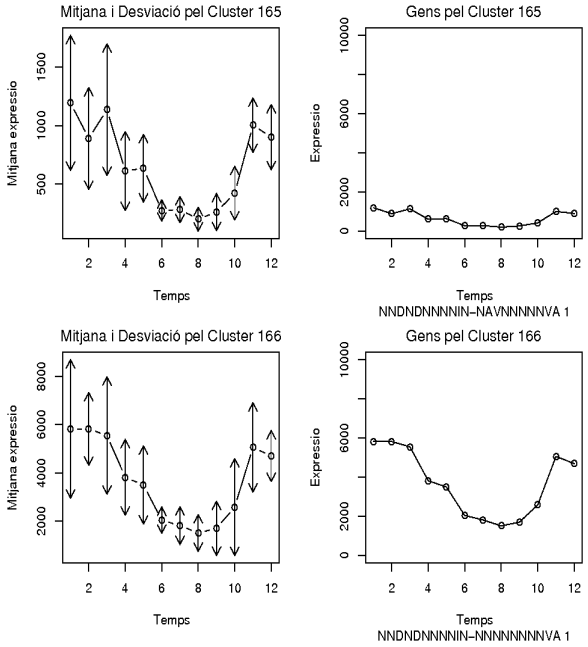

Mitjana i Desviació pel Cluster 156
Mitjana expressio
Temps
Gens pel Cluster 156
Expressio
Temps
NINNNDNNNNN−NNNNANNNNN 1

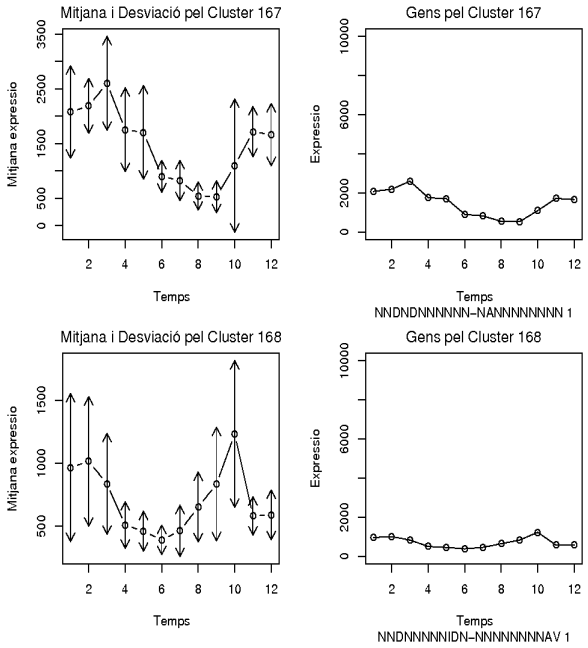

Mitjana i Desviació pel Cluster 157
Mitjana expressio
Temps
Gens pel Cluster 157
Expressio
Temps
NINNNDNNNNN−VNNNNNNNNN 2
Mitjana i Desviació pel Cluster 158
Mitjana expressio
Temps
Gens pel Cluster 158
Expressio
Temps
NINNNNNNNN−VANNNNNNNA 1

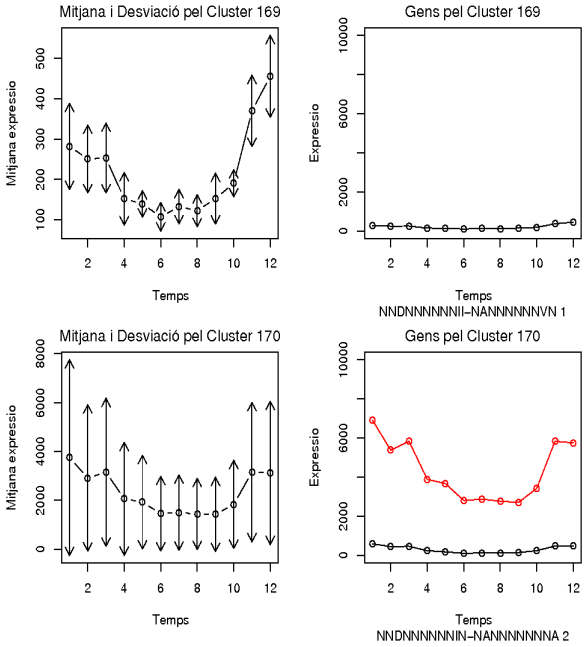

Mitjana i Desviació pel Cluster 159
Mitjana expressio
Temps
Gens pel Cluster 159
Expressio
Temps
NINNNNNNNNN−NANNNNNNNN 13
Mitjana i Desviació pel Cluster 160
Mitjana expressio
Temps
Gens pel Cluster 160
Expressio
Temps
NINNNNNNNNN−NNNNNNNNNN 14

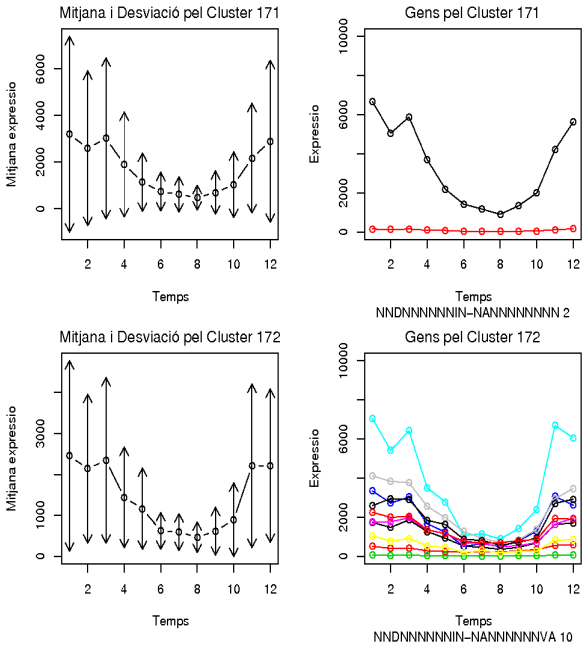

Mitjana i Desviació pel Cluster 161
Mitjana expressio
0 100 300 500
2 4 6 8 10 12
Temps
Gens pel Cluster 161
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NINNNNNNNNN−VANNNNNNNN 5
Mitjana i Desviació pel Cluster 162
Mitjana expressio
50 150 250 350
2 4 6 8 10 12
Temps
Gens pel Cluster 162
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NINNNNNNNNN−VNNNNNNNNN 3

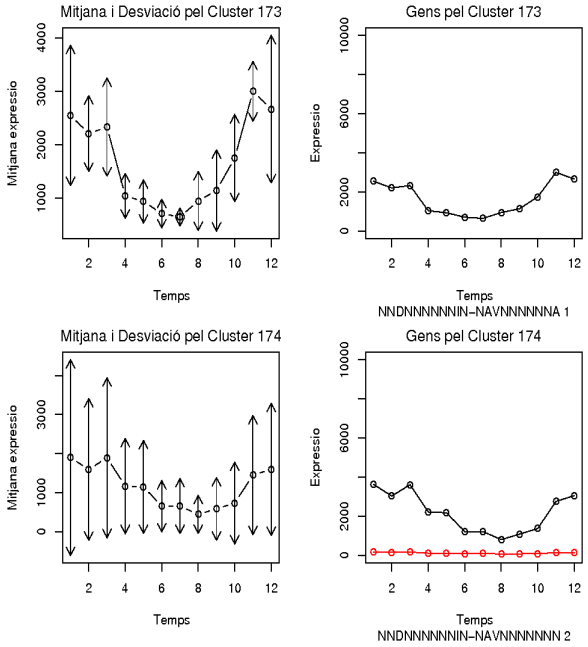

Mitjana i Desviació pel Cluster 163
Mitjana expressio
0 1000 3000 5000
2 4 6 8 10 12
Temps
Gens pel Cluster 163
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NNDDNNNNNIN−NANNNNNNVA 1
Mitjana i Desviació pel Cluster 164
Mitjana expressio
0 2000 4000 6000
2 4 6 8 10 12
Temps
Gens pel Cluster 164
Expressio
0 2000 6000 10000
2 4 6 8 10 12
Temps
NNDDNNNNNIN−NNNNNNNNVA 4

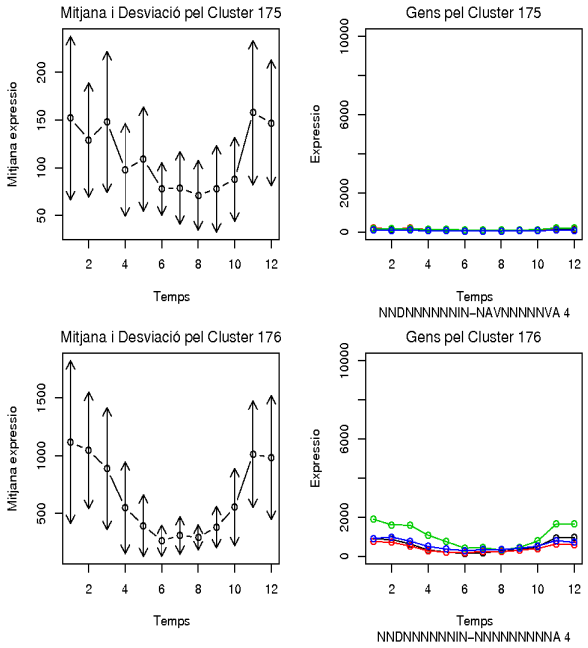

Mitjana i Desviació pel Cluster 165
Mitjana expressio
Temps
Gens pel Cluster 165
Expressio
Temps
NNDNDNNNNIN–NAVNNNNNNVA 1
Mitjana i Desviació pel Cluster 166
Mitjana expressio
Temps
Gens pel Cluster 166
Expressio
Temps
NNDNDNNNNIN–NNNNNNNNVA 1

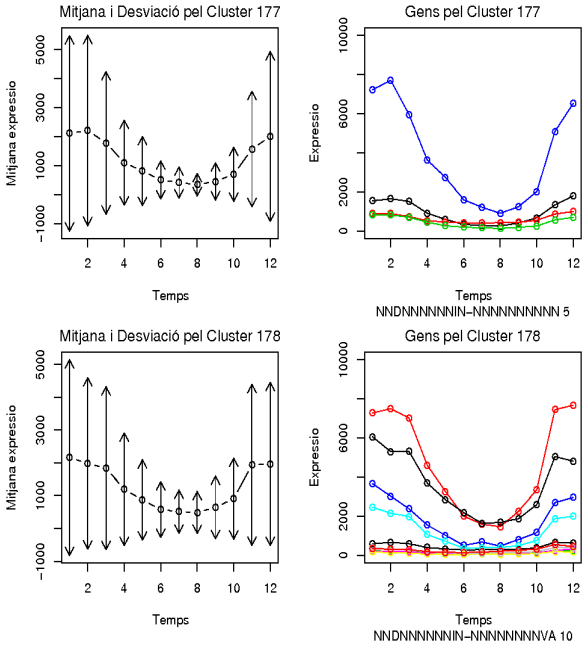

Mitjana i Desviació pel Cluster 167
Mitjana expressio
Temps
Gens pel Cluster 167
Expressio
Temps
NNDNDNNNNNN–NANNNNNNNN 1
Mitjana i Desviació pel Cluster 168
Mitjana expressio
Temps
Gens pel Cluster 168
Expressio
Temps
NNDNNNNNIDN–NNNNNNNNAV 1

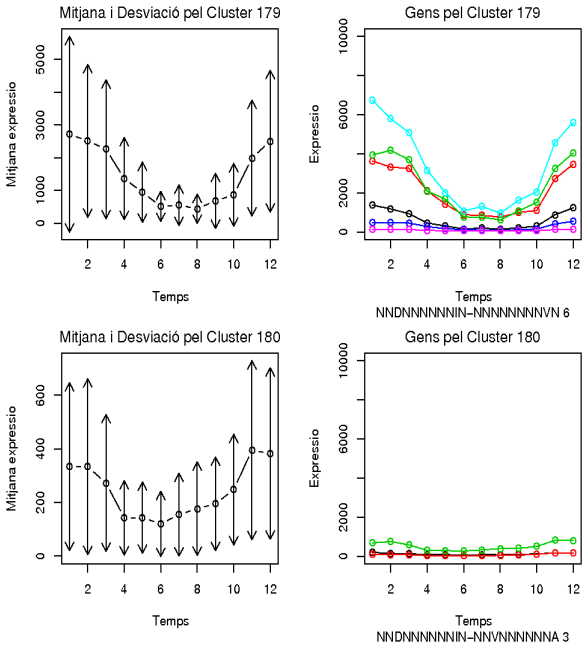

Mitjana i Desviació pel Cluster 169
Mitjana expressio
Temps
Gens pel Cluster 169
Expressio
Temps
NNDNNNNNNII–NANNNNNNVN 1
Mitjana i Desviació pel Cluster 170
Mitjana expressio
Temps
Gens pel Cluster 170
Expressio
Temps
NNDNNNNNNIN–NANNNNNNNA 2

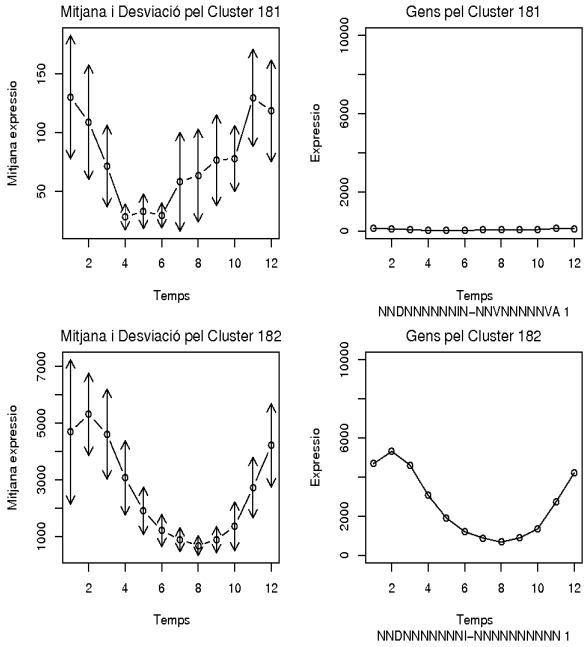

Mitjana i Desviació pel Cluster 171
Mitjana expressio
Temps
Gens pel Cluster 171
Expressio
Temps
NNDNNNNNNIN–NANNNNNNNN 2
Mitjana i Desviació pel Cluster 172
Mitjana expressio
Temps
Gens pel Cluster 172
Expressio
Temps
NNDNNNNNNIN–NANNNNNNVA 10

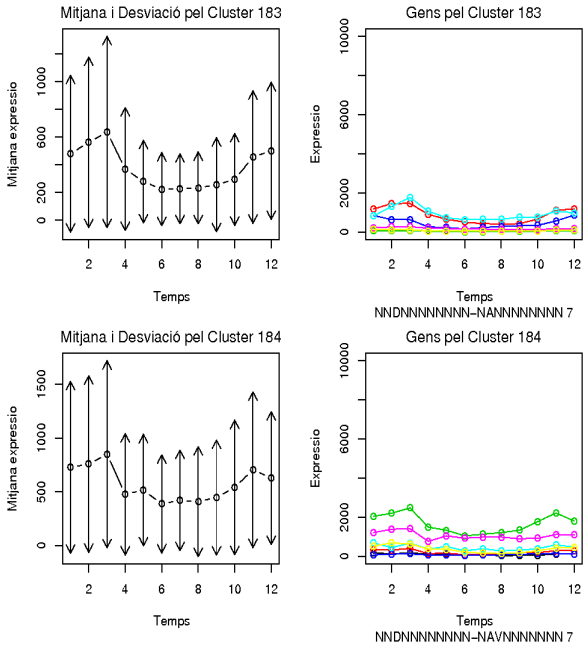

Mitjana i Desviació pel Cluster 173
Mitjana expressio
Temps
Gens pel Cluster 173
Expressio
Temps
NNDNNNNNNIN−NAVNNNNNNA 1
Mitjana i Desviació pel Cluster 174
Mitjana expressio
Temps
Gens pel Cluster 174
Expressio
Temps
NNDNNNNNNIN−NAVNNNNNNN 2

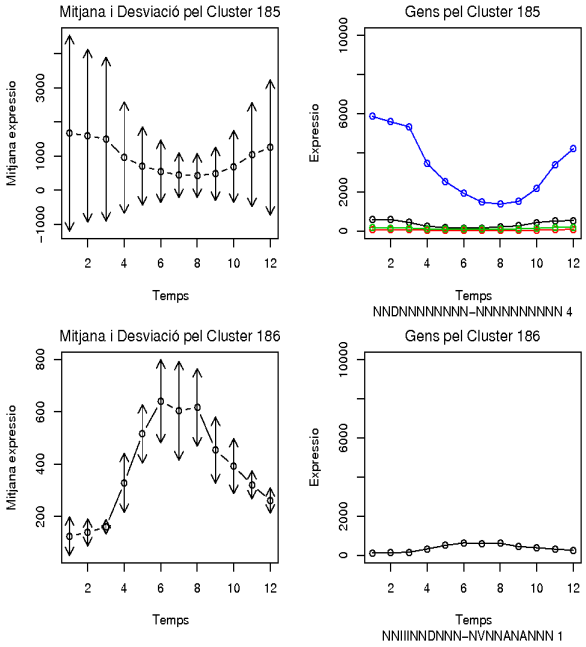

Mitjana i Desviació pel Cluster 175
Mitjana expressio
Temps
Gens pel Cluster 175
Expressio
Temps
NNDNNNNNNIN−NAVNNNNNVA 4
Mitjana i Desviació pel Cluster 176
Mitjana expressio
Temps
Gens pel Cluster 176
Expressio
Temps
NNDNNNNNNIN−NNNNNNNNA 4

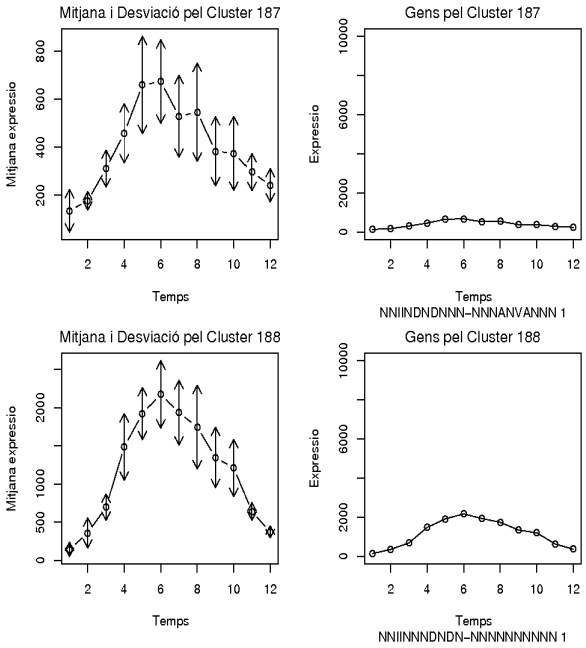

Mitjana i Desviació pel Cluster 177
Mitjana expressio
Temps
Gens pel Cluster 177
Expressio
Temps
NNDNNNNNNIN−NNNNNNNNN 5
Mitjana i Desviació pel Cluster 178
Mitjana expressio
Temps
Gens pel Cluster 178
Expressio
Temps
NNDNNNNNNIN−NNNNNNNNVA 10

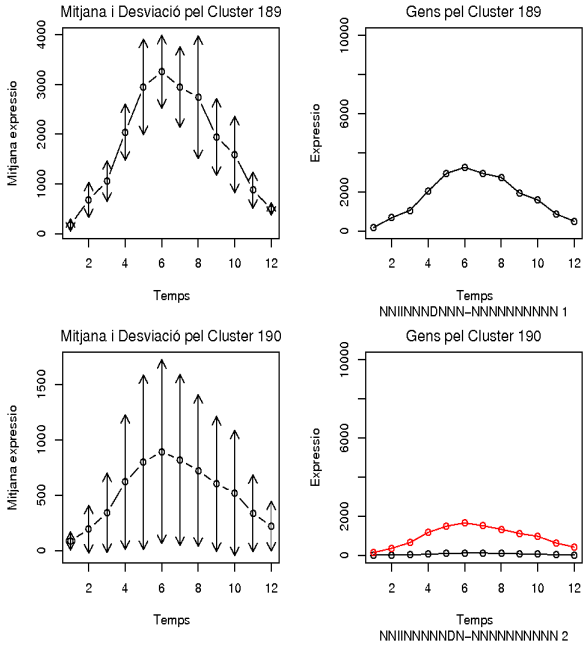

Mitjana i Desviació pel Cluster 179
Mitjana expressio
Temps
Gens pel Cluster 179
Expressio
Temps
NNDNNNNNNIN−NNNNNNNNVN 6
Mitjana i Desviació pel Cluster 180
Mitjana expressio
Temps
Gens pel Cluster 180
Expressio
Temps
NNDNNNNNNIN−NNVNNNNNNA 3

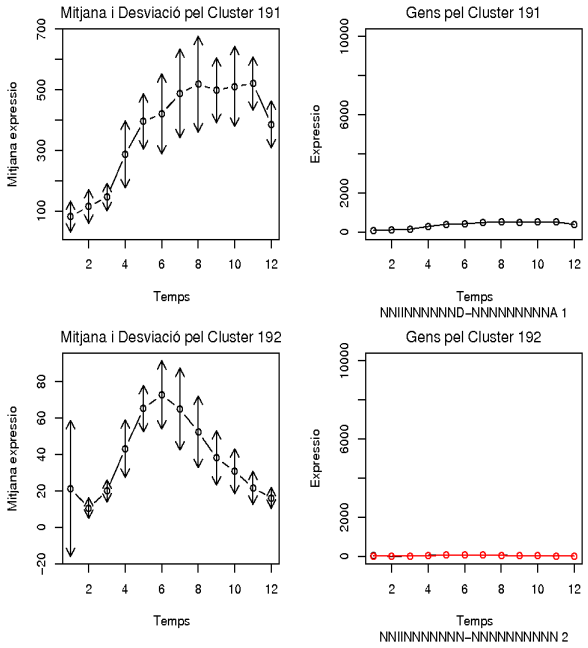

Mitjana i Desviació pel Cluster 181
Mitjana expressio
Temps
Gens pel Cluster 181
Expressio
Temps
NNDNNNNNNIN−NNVNNNNNVA 1
Mitjana i Desviació pel Cluster 182
Mitjana expressio
Temps
Gens pel Cluster 182
Expressio
Temps
NNDNNNNNNNI−NNNNNNNNNN 1

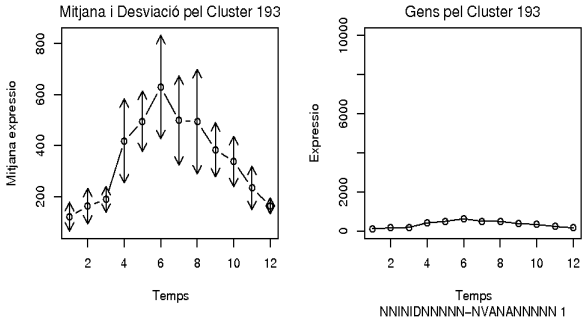

Mitjana i Desviació pel Cluster 183
Mitjana expressio
Temps
Gens pel Cluster 183
Expressio
Temps
NNDNNNNNNNN−NANNNNNNNN 7
Mitjana i Desviació pel Cluster 184
Mitjana expressio
Temps
Gens pel Cluster 184
Expressio
Temps
NNDNNNNNNNN−NAVNNNNNNN 7

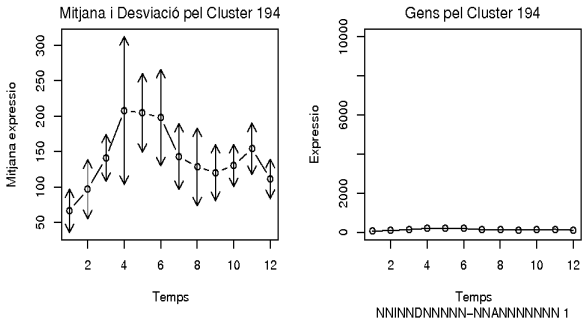

Mitjana i Desviació pel Cluster 185
Mitjana expressio
Temps
Gens pel Cluster 185
Expressio
Temps
NNDNNNNNNNN−NNNNNNNNNN 4
Mitjana i Desviació pel Cluster 186
Mitjana expressio
Temps
Gens pel Cluster 186
Expressio
Temps
NNIIINNDNNN−NVNNANANNN 1

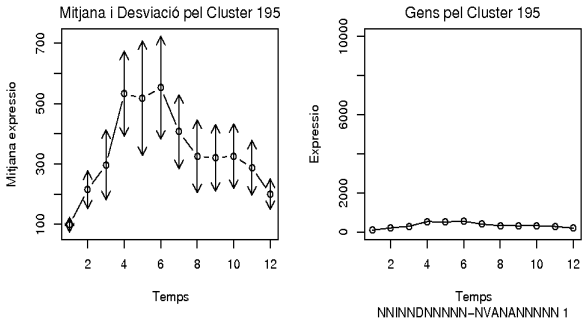

Mitjana i Desviació pel Cluster 187
Mitjana expressio
Temps
Gens pel Cluster 187
Expressio
Temps
NNIIINDNDNNN−NNNANVANNN 1
Mitjana i Desviació pel Cluster 188
Mitjana expressio
Temps
Gens pel Cluster 188
Expressio
Temps
NNIINNNDNDN−NNNNNNNNNN 1

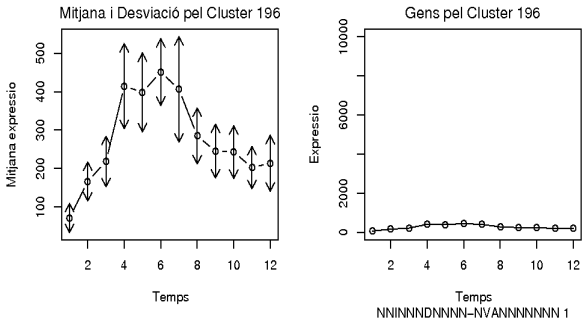

Mitjana i Desviació pel Cluster 189
Mitjana expressio
Temps
Gens pel Cluster 189
Expressio
Temps
NNIINNNDNNN−NNNNNNNNNN 1
Mitjana i Desviació pel Cluster 190
Mitjana expressio
Temps
Gens pel Cluster 190
Expressio
Temps
NNIINNNNNDN−NNNNNNNNNN 2

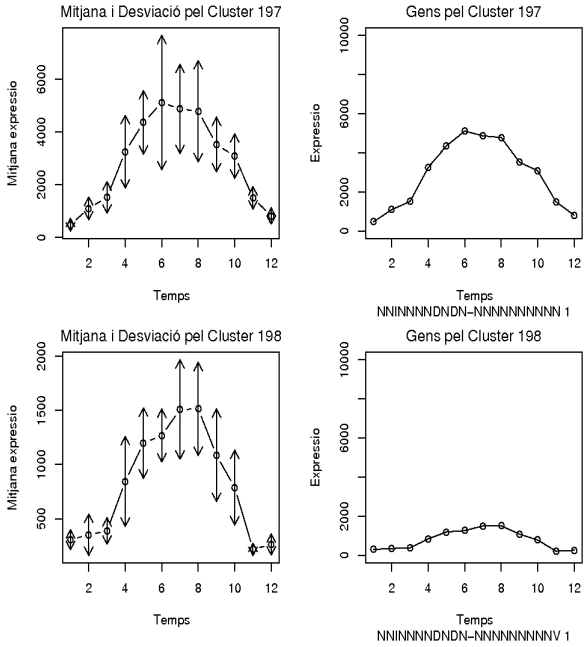

Mitjana i Desviació pel Cluster 191
Mitjana expressio
Temps
Gens pel Cluster 191
Expressio
Temps
NNIINNNNNND−NNNNNNNNNA 1
Mitjana i Desviació pel Cluster 192
Mitjana expressio
Temps
Gens pel Cluster 192
Expressio
Temps
NNIINNNNNNN−NNNNNNNNNN 2

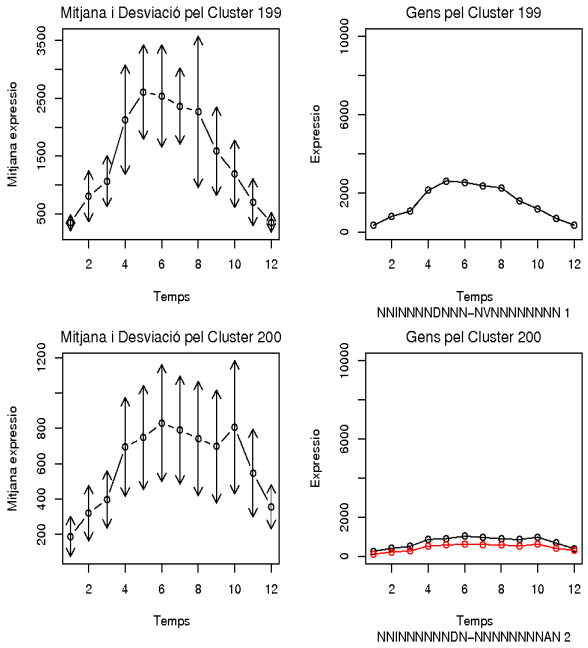

Mitjana i Desviació pel Cluster 193
Mitjana expressio
Temps
Gens pel Cluster 193
Expressio
Temps
NNINIDNNNNN–NVANANNNNN 1

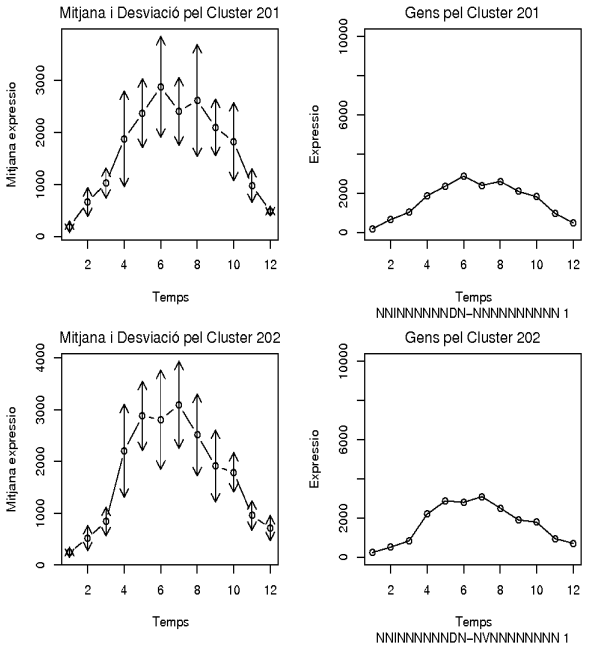

Mitjana i Desviació pel Cluster 194
Mitjana expressio
Temps
Gens pel Cluster 194
Expressio
Temps
NNINNDNNNNN–NNANNNNNNN 1

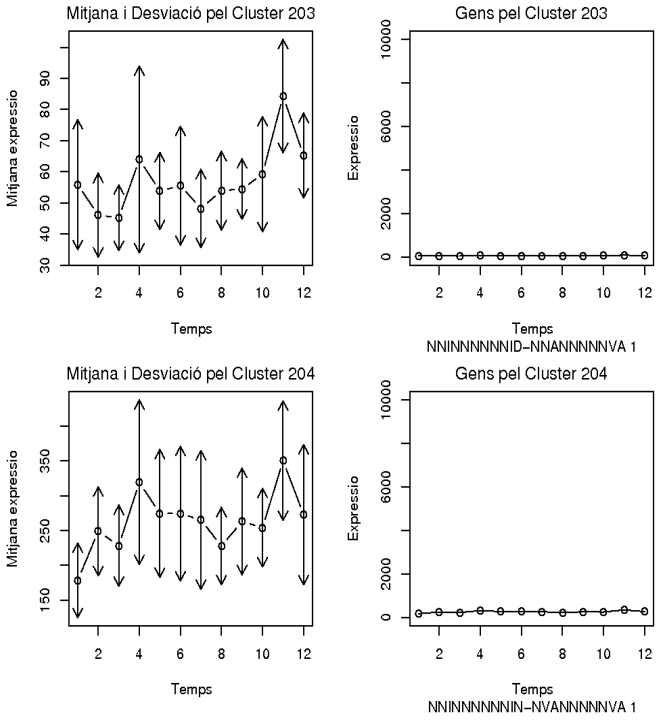

Mitjana i Desviació pel Cluster 195
Mitjana expressio
Temps
Gens pel Cluster 195
Expressio
Temps
NNINNDNNNNN–NVANANNNNN 1

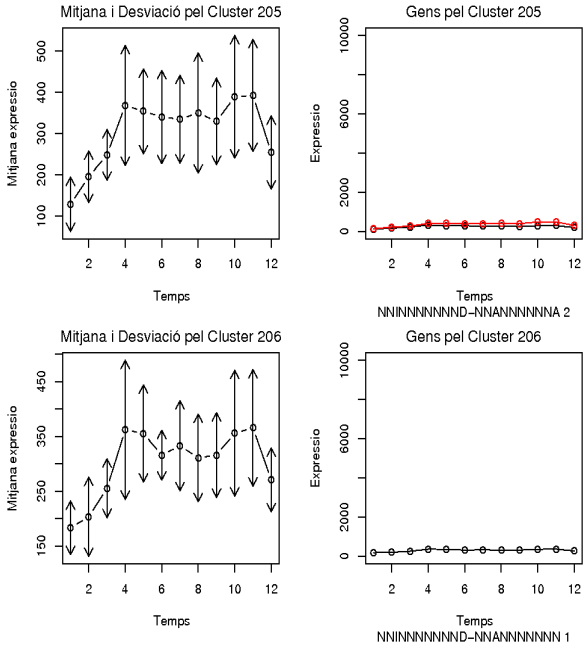

Mitjana i Desviació pel Cluster 196
Mitjana expressio
Temps
Gens pel Cluster 196
Expressio
Temps
NNINNNDNNNN–NVANNNNNNN 1

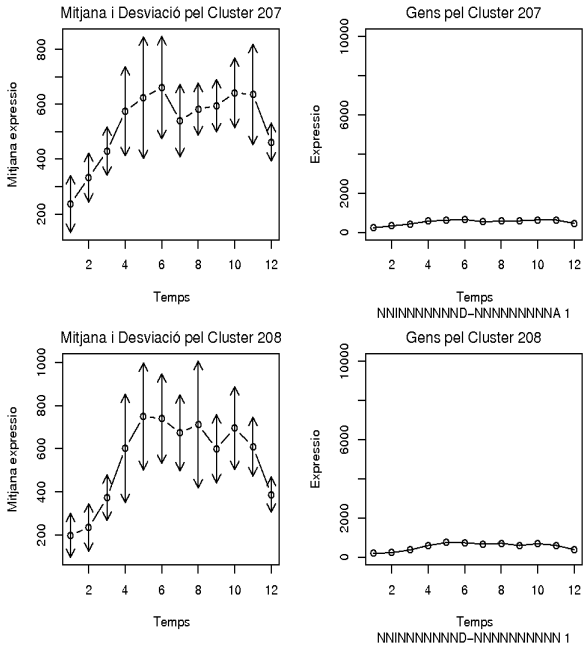

Mitjana i Desviació pel Cluster 197
Mitjana expressio
Temps
Gens pel Cluster 197
Expressio
Temps
NNINNNNDNDN−NNNNNNNNN 1
Mitjana i Desviació pel Cluster 198
Mitjana expressio
Temps
Gens pel Cluster 198
Expressio
Temps
NNINNNNDNDN−NNNNNNNNV 1

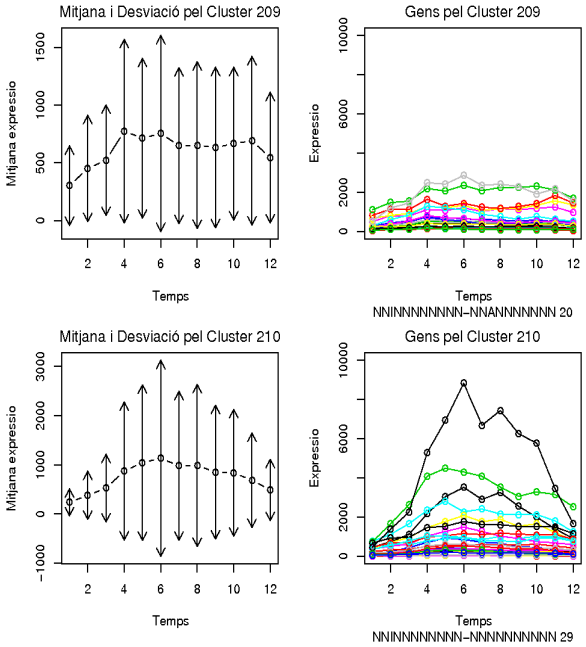

Mitjana i Desviació pel Cluster 199
Mitjana expressio
Temps
Gens pel Cluster 199
Expressio
Temps
NNINNNNDNNN−NVNNNNNNNN 1
Mitjana i Desviació pel Cluster 200
Mitjana expressio
Temps
Gens pel Cluster 200
Expressio
Temps
NNINNNNNNDN−NNNNNNNNAN 2

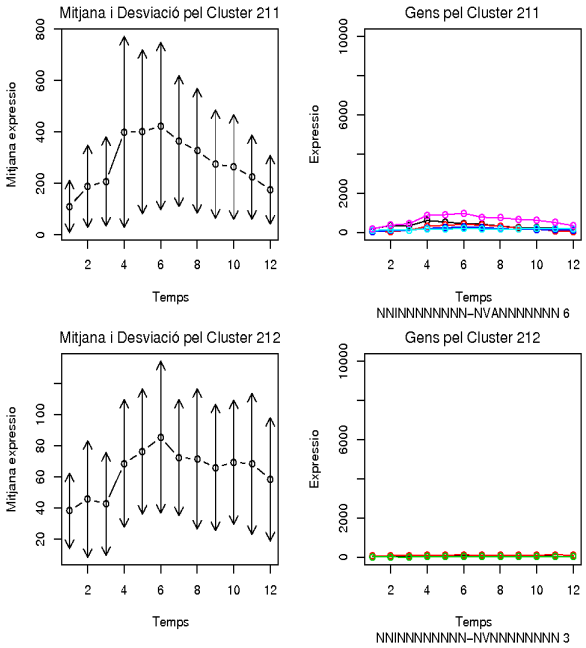

Mitjana i Desviació pel Cluster 201
Mitjana expressio
Temps
Gens pel Cluster 201
Expressio
Temps
NNINNNNNNDN−NNNNNNNNN 1
Mitjana i Desviació pel Cluster 202
Mitjana expressio
Temps
Gens pel Cluster 202
Expressio
Temps
NNINNNNNNDN−NVNNNNNNN 1

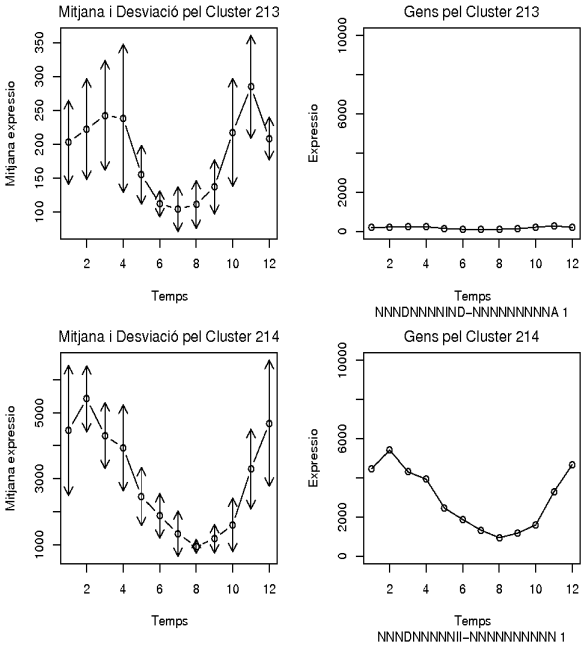

Mitjana i Desviació pel Cluster 203
Mitjana expressio
Temps
Gens pel Cluster 203
Expressio
Temps
NNINNNNNNID−NNANNNNNVA 1
Mitjana i Desviació pel Cluster 204
Mitjana expressio
Temps
Gens pel Cluster 204
Expressio
Temps
NNINNNNNNIN−NVANNNNNVA 1

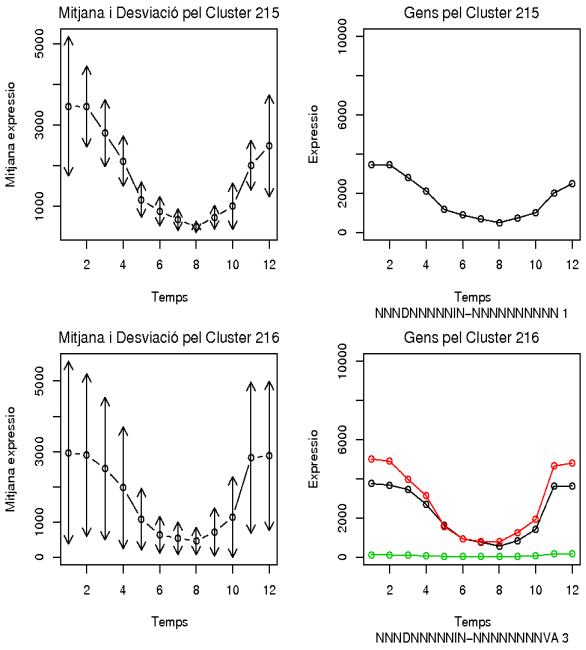

Mitjana i Desviació pel Cluster 205
Mitjana expressio
Temps
Gens pel Cluster 205
Expressio
Temps
NNINNNNNNND−NNANNNNNNA 2
Mitjana i Desviació pel Cluster 206
Mitjana expressio
Temps
Gens pel Cluster 206
Expressio
Temps
NNINNNNNNND−NNANNNNNNN 1

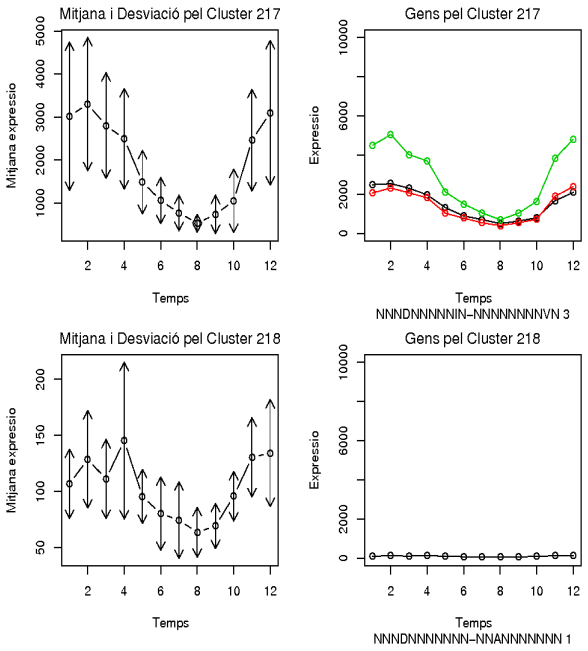

Mitjana i Desviació pel Cluster 207
Mitjana expressio
Temps
Gens pel Cluster 207
Expressio
Temps
NNINNNNNNND−NNNNNNNNNA 1
Mitjana i Desviació pel Cluster 208
Mitjana expressio
Temps
Gens pel Cluster 208
Expressio
Temps
NNINNNNNNND−NNNNNNNNNN 1

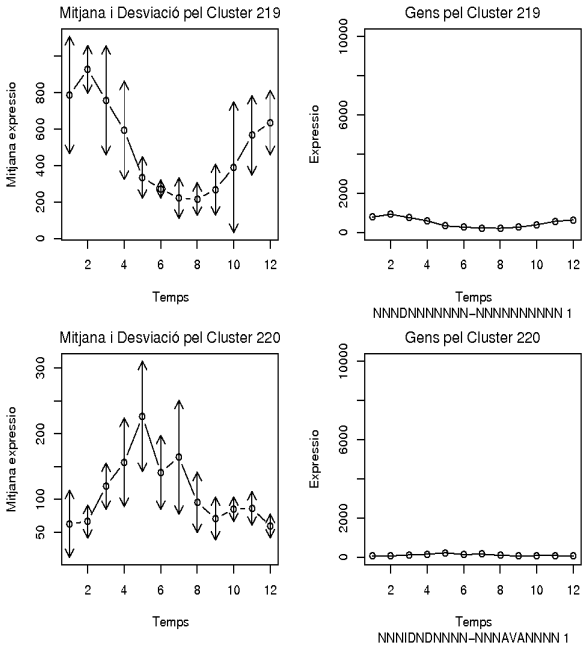

Mitjana i Desviació pel Cluster 209
Mitjana expressio
Temps
Gens pel Cluster 209
Expressio
Temps
NNINNNNNNNN−NNANNNNNNN 20
Mitjana i Desviació pel Cluster 210
Mitjana expressio
Temps
Gens pel Cluster 210
Expressio
Temps
NNINNNNNNNN−NNNNNNNNN 29

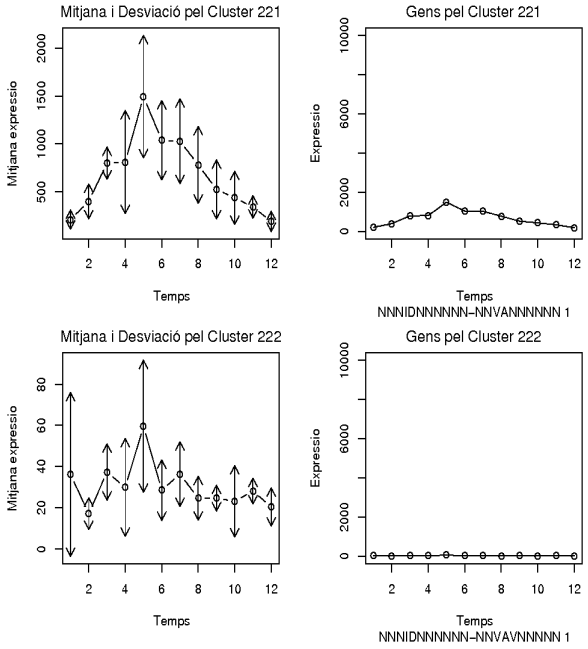

Mitjana i Desviació pel Cluster 211
Mitjana expressio
Temps
Gens pel Cluster 211
Expressio
Temps
NNINNNNNNNN−NVANNNNNNN 6
Mitjana i Desviació pel Cluster 212
Mitjana expressio
Temps
Gens pel Cluster 212
Expressio
Temps
NNINNNNNNNN−NVNNNNNNN 3

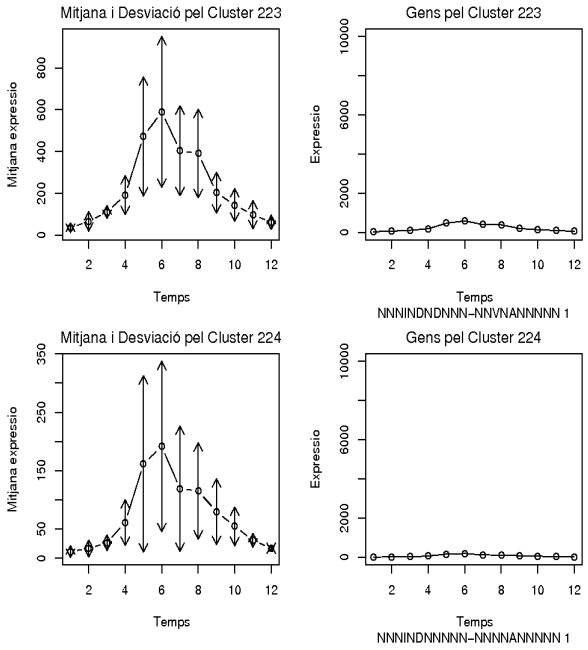

Mitjana i Desviació pel Cluster 213
Mitjana expressio
Temps
Gens pel Cluster 213
Expressio
Temps
NNNDNNNNIND−NNNNNNNNNA 1
Mitjana i Desviació pel Cluster 214
Mitjana expressio
Temps
Gens pel Cluster 214
Expressio
Temps
NNNDNNNNNII−NNNNNNNNNN 1

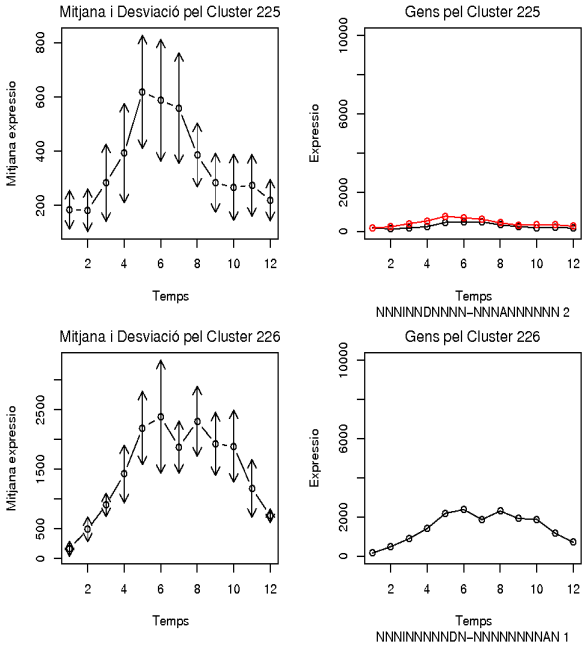

Mitjana i Desviació pel Cluster 215
Mitjana expressio
Temps
Gens pel Cluster 215
Expressio
Temps
NNNDNNNNNIN−NNNNNNNNNN 1
Mitjana i Desviació pel Cluster 216
Mitjana expressio
Temps
Gens pel Cluster 216
Expressio
Temps
NNNDNNNNNIN−NNNNNNNNVA 3

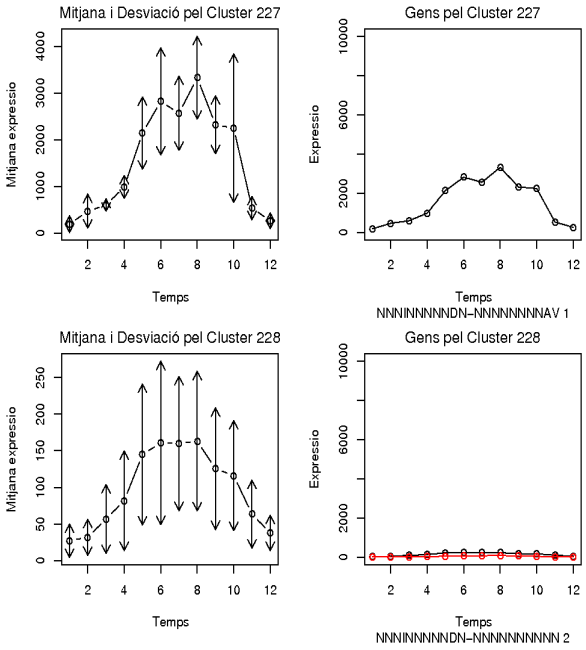

Mitjana i Desviació pel Cluster 217
Mitjana expressio
Temps
Gens pel Cluster 217
Expressio
Temps
NNNDNNNNNIN–NNNNNNNNVN 3
Mitjana i Desviació pel Cluster 218
Mitjana expressio
Temps
Gens pel Cluster 218
Expressio
Temps
NNNDNNNNNNN–NNANNNNNNN 1

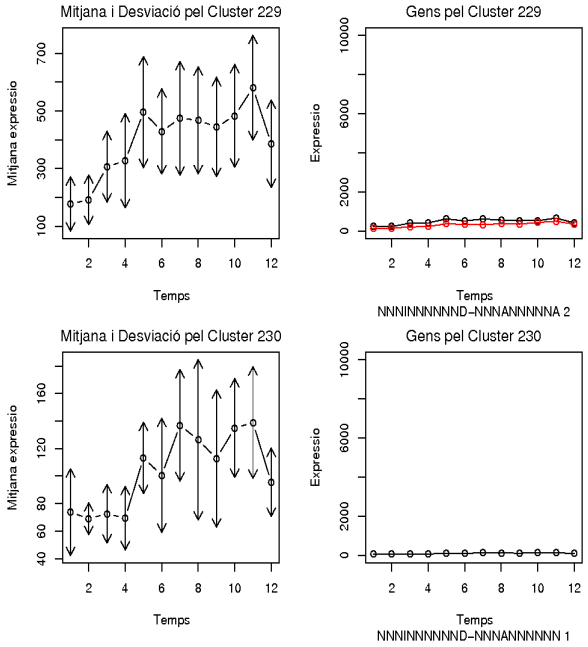

Mitjana i Desviació pel Cluster 219
Mitjana expressio
Temps
Gens pel Cluster 219
Expressio
Temps
NNNDNNNNNNN–NNNNNNNNNN 1
Mitjana i Desviació pel Cluster 220
Mitjana expressio
Temps
Gens pel Cluster 220
Expressio
Temps
NNNIDNDNNNN–NNNAVANNNN 1

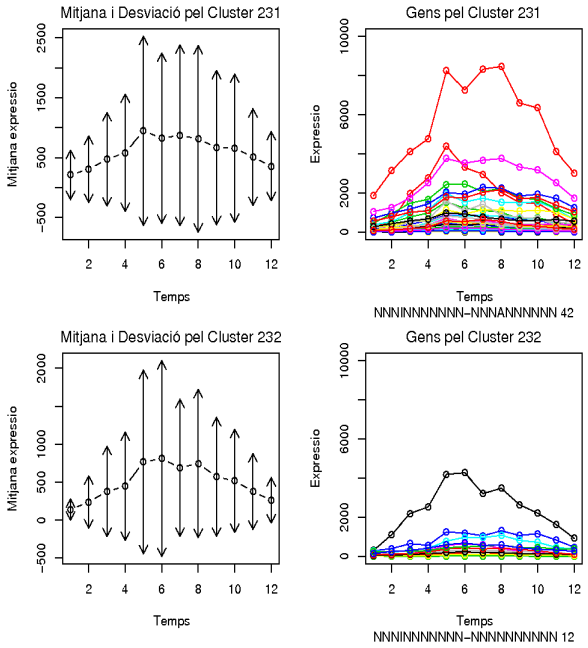

Mitjana i Desviació pel Cluster 221
Mitjana expressio
Temps
Gens pel Cluster 221
Expressio
Temps
NNNIDNNNNNN–NNVANNNNNN 1
Mitjana i Desviació pel Cluster 222
Mitjana expressio
Temps
Gens pel Cluster 222
Expressio
Temps
NNNIDNNNNNN–NNVAVNNNNN 1

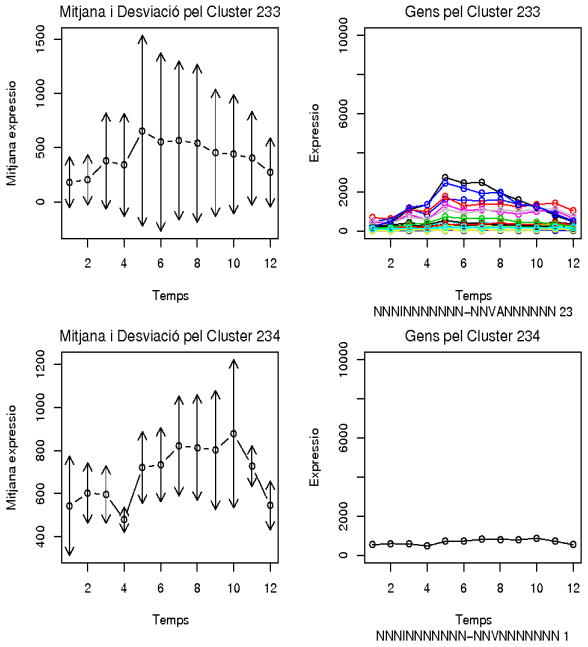

Mitjana i Desviació pel Cluster 223
Mitjana expressio
Temps
Gens pel Cluster 223
Expressio
Temps
NNNINDNDNNN–NNVNANNNNN 1
Mitjana i Desviació pel Cluster 224
Mitjana expressio
Temps
Gens pel Cluster 224
Expressio
Temps
NNNINDNNNNN–NNNNANNNNN 1

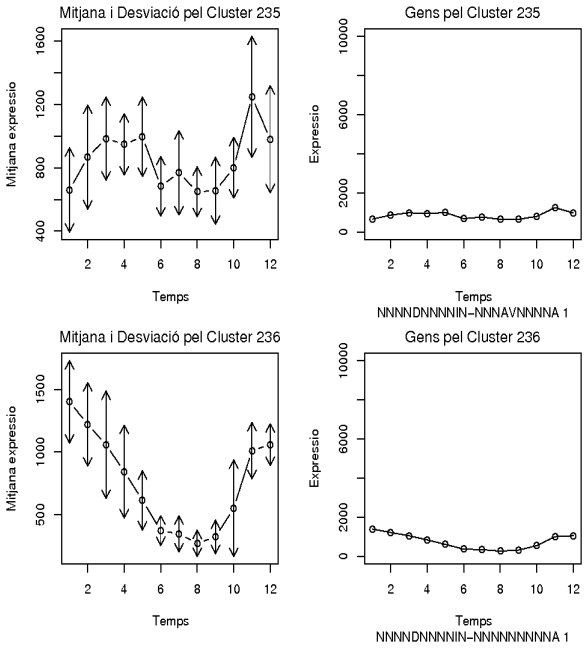

Mitjana i Desviació pel Cluster 225
Mitjana expressio
Temps
Gens pel Cluster 225
Expressio
Temps
NNNINNDNNNN–NNNANNNNNN 2
Mitjana i Desviació pel Cluster 226
Mitjana expressio
Temps
Gens pel Cluster 226
Expressio
Temps
NNNINNNNNDN–NNNNNNNNAN 1

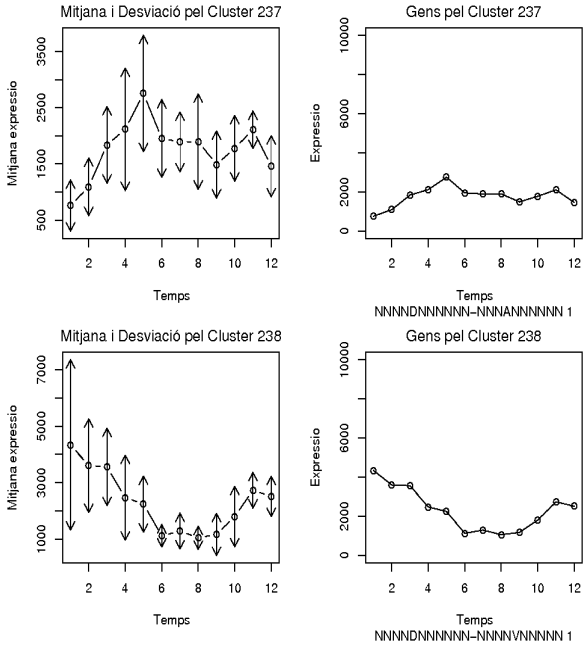

Mitjana i Desviació pel Cluster 227
Mitjana expressio
Temps
Gens pel Cluster 227
Expressio
Temps
NNNINNNNNDN–NNNNNNNNAV 1
Mitjana i Desviació pel Cluster 228
Mitjana expressio
Temps
Gens pel Cluster 228
Expressio
Temps
NNNINNNNNDN–NNNNNNNNNN 2

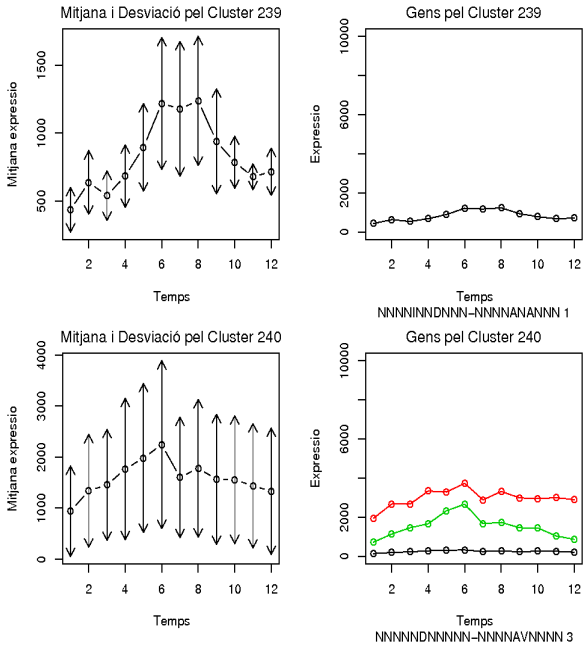

Mitjana i Desviació pel Cluster 229
Mitjana expressio
Temps
Gens pel Cluster 229
Expressio
Temps
NNNINNNNNND−NNNANNNNNA 2
Mitjana i Desviació pel Cluster 230
Mitjana expressio
Temps
Gens pel Cluster 230
Expressio
Temps
NNNINNNNNND−NNNANNNNNN 1

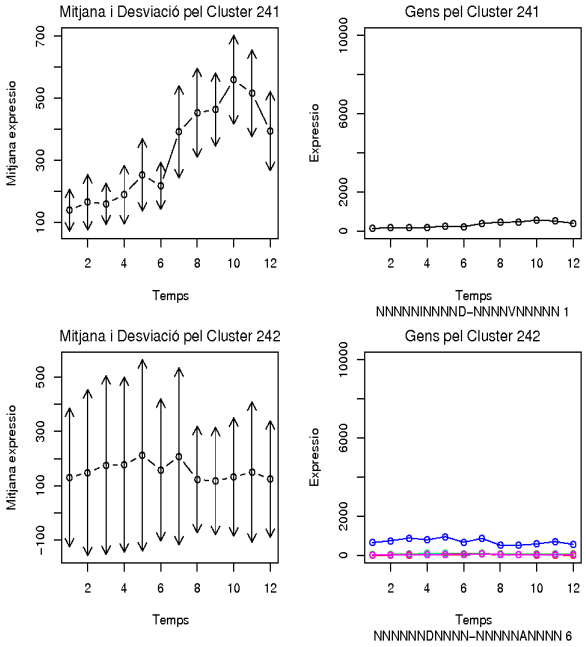

Mitjana i Desviació pel Cluster 231
Mitjana expressio
Temps
Gens pel Cluster 231
Expressio
Temps
NNNINNNNNNN−NNNANNNNNN 42
Mitjana i Desviació pel Cluster 232
Mitjana expressio
Temps
Gens pel Cluster 232
Expressio
Temps
NNNINNNNNNN−NNNNNNNNNN 12

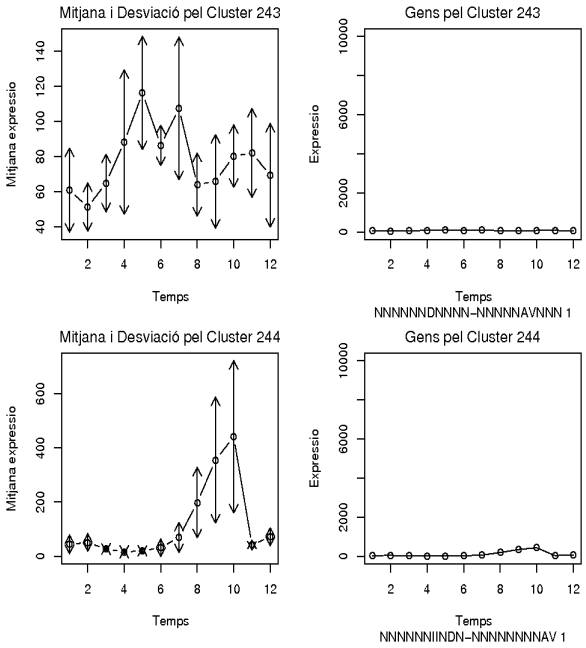

Mitjana i Desviació pel Cluster 233
Mitjana expressio
Temps
Gens pel Cluster 233
Expressio
Temps
NNNINNNNNNN–NNVANNNNNN 23
Mitjana i Desviació pel Cluster 234
Mitjana expressio
Temps
Gens pel Cluster 234
Expressio
Temps
NNNINNNNNNN–NNVNNNNNNN 1

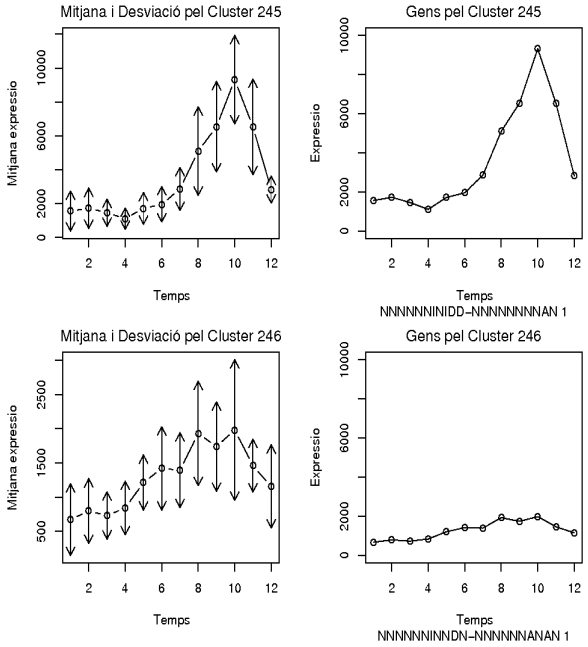

Mitjana i Desviació pel Cluster 235
Mitjana expressio
Temps
Gens pel Cluster 235
Expressio
Temps
NNNNDNNNNIN–NNNAVNNNNA 1
Mitjana i Desviació pel Cluster 236
Mitjana expressio
Temps
Gens pel Cluster 236
Expressio
Temps
NNNNDNNNNIN–NNNNNNNNNA 1

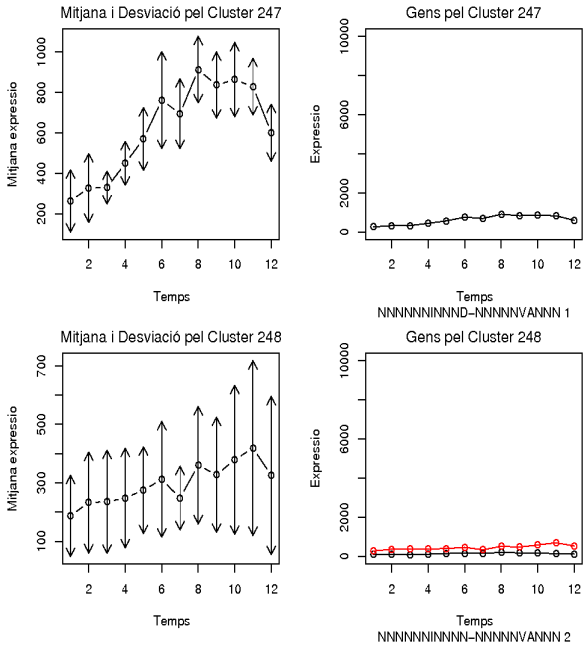

Mitjana i Desviació pel Cluster 237
Mitjana expressio
Temps
Gens pel Cluster 237
Expressio
Temps
NNNNDNNNNN-NNNANNNNNN 1
Mitjana i Desviació pel Cluster 238
Mitjana expressio
Temps
Gens pel Cluster 238
Expressio
Temps
NNNNDNNNNN-NNNNVNNNNN 1

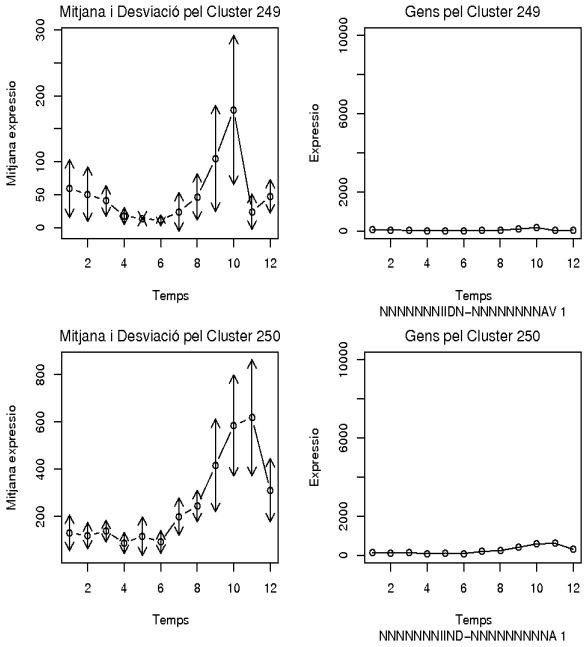

Mitjana i Desviació pel Cluster 239
Mitjana expressio
Temps
Gens pel Cluster 239
Expressio
Temps
NNNNINNDNNN-NNNNANANNN 1
Mitjana i Desviació pel Cluster 240
Mitjana expressio
Temps
Gens pel Cluster 240
Expressio
Temps
NNNNNDNNNNN-NNNNAVNNNN 3

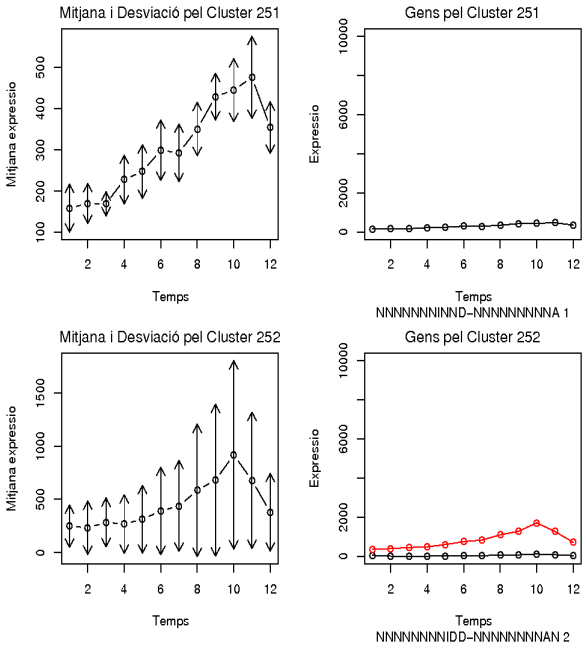

Mitjana i Desviació pel Cluster 241
Mitjana expressio
Temps
Gens pel Cluster 241
Expressio
Temps
NNNNNINNNND−NNNNVNNNNN 1
Mitjana i Desviació pel Cluster 242
Mitjana expressio
Temps
Gens pel Cluster 242
Expressio
Temps
NNNNNNDNNNN−NNNNNANNNN 6

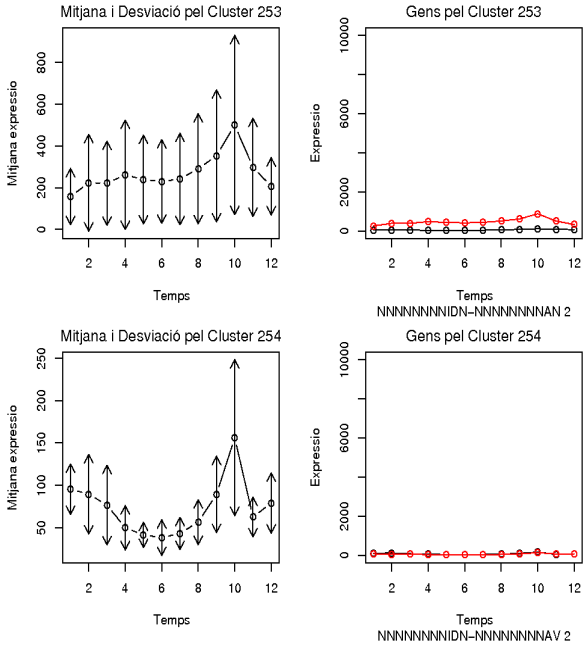

Mitjana i Desviació pel Cluster 243
Mitjana expressio
Temps
Gens pel Cluster 243
Expressio
Temps
NNNNNNDNNNN−NNNNNAVNNN 1
Mitjana i Desviació pel Cluster 244
Mitjana expressio
Temps
Gens pel Cluster 244
Expressio
Temps
NNNNNNIINDN−NNNNNNNNAV 1

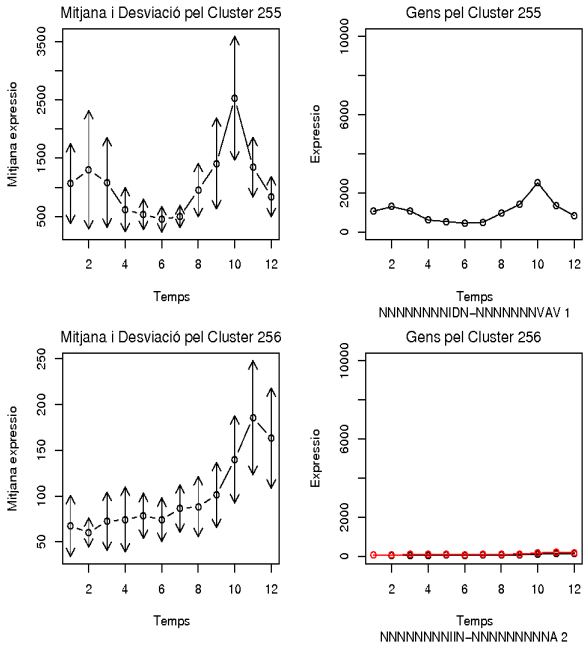

Mitjana i Desviació pel Cluster 245
Mitjana expressio
Temps
Gens pel Cluster 245
Expressio
Temps
NNNNNNINIDD–NNNNNNNNAN 1
Mitjana i Desviació pel Cluster 246
Mitjana expressio
Temps
Gens pel Cluster 246
Expressio
Temps
NNNNNNINNDN–NNNNNNANAN 1

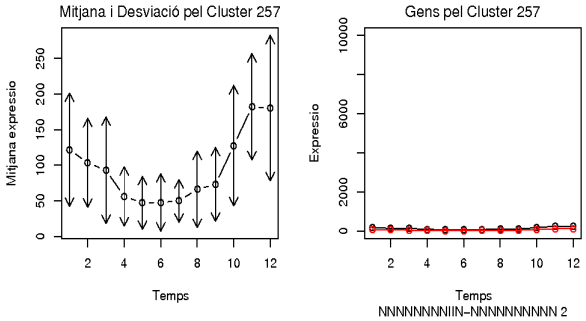

Mitjana i Desviació pel Cluster 247
Mitjana expressio
Temps
Gens pel Cluster 247
Expressio
Temps
NNNNNNINNND–NNNNNVANNN 1
Mitjana i Desviació pel Cluster 248
Mitjana expressio
Temps
Gens pel Cluster 248
Expressio
Temps
NNNNNNINNNN–NNNNNVANNN 2

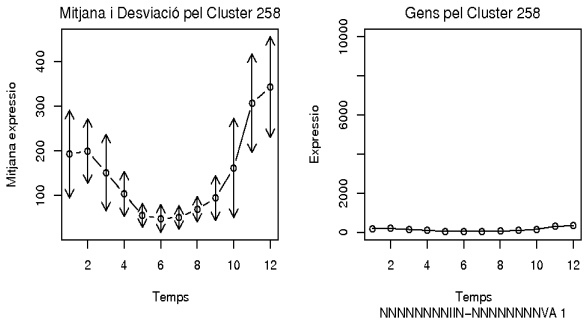

Mitjana i Desviació pel Cluster 249
Mitjana expressio
Temps
Gens pel Cluster 249
Expressio
Temps
NNNNNNNIIDN−NNNNNNNNAV 1
Mitjana i Desviació pel Cluster 250
Mitjana expressio
Temps
Gens pel Cluster 250
Expressio
Temps
NNNNNNNIIND−NNNNNNNNNA 1

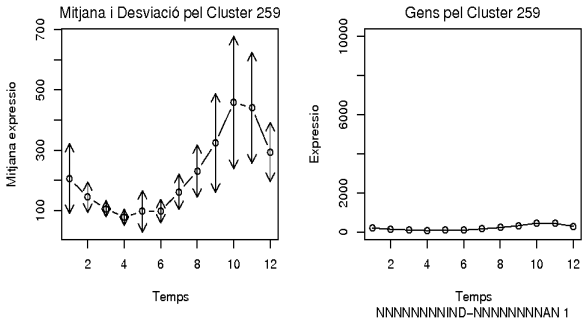

Mitjana i Desviació pel Cluster 251
Mitjana expressio
Temps
Gens pel Cluster 251
Expressio
Temps
NNNNNNNINND−NNNNNNNNNA 1
Mitjana i Desviació pel Cluster 252
Mitjana expressio
Temps
Gens pel Cluster 252
Expressio
Temps
NNNNNNNNIDD−NNNNNNNNAN 2

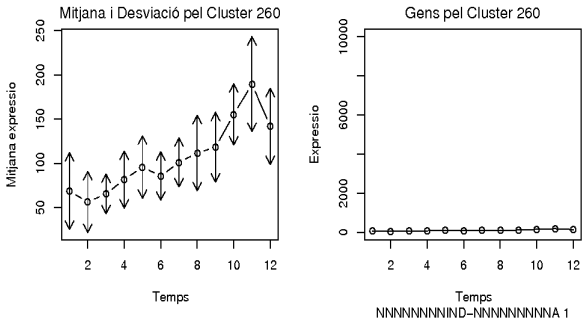

Mitjana i Desviació pel Cluster 253
Mitjana expressio
Temps
Gens pel Cluster 253
Expressio
Temps
NNNNNNNNIDN−NNNNNNNNAN 2
Mitjana i Desviació pel Cluster 254
Mitjana expressio
Temps
Gens pel Cluster 254
Expressio
Temps
NNNNNNNNIDN−NNNNNNNNAV 2

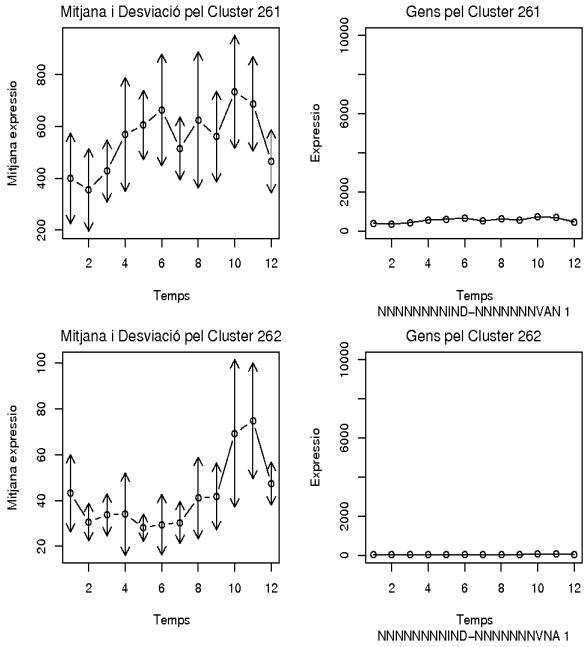

Mitjana i Desviació pel Cluster 255
Mitjana expressio
Temps
Gens pel Cluster 255
Expressio
Temps
NNNNNNNNIDN−NNNNNNNNVAV 1
Mitjana i Desviació pel Cluster 256
Mitjana expressio
Temps
Gens pel Cluster 256
Expressio
Temps
NNNNNNNNIIN−NNNNNNNNNA 2

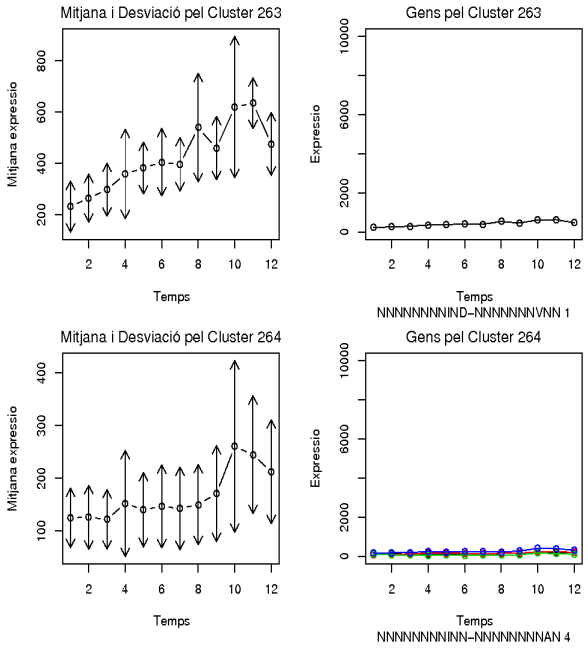

Mitjana i Desviació pel Cluster 257
Mitjana expressio
Temps
Gens pel Cluster 257
Expressio
Temps
NNNNNNNNIIN−NNNNNNNNN 2

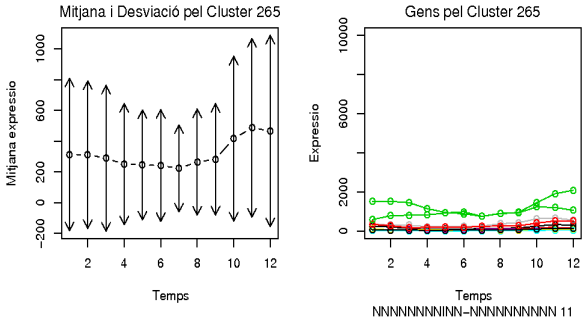

Mitjana i Desviació pel Cluster 258
Mitjana expressio
Temps
Gens pel Cluster 258
Expressio
Temps
NNNNNNNNIIN−NNNNNNNNA 1

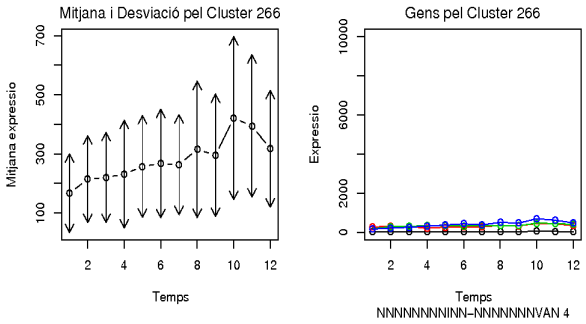

Mitjana i Desviació pel Cluster 259
Mitjana expressio
Temps
Gens pel Cluster 259
Expressio
Temps
NNNNNNNNIND−NNNNNNNNAN 1

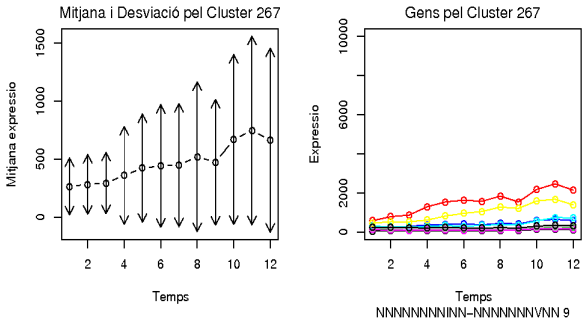

Mitjana i Desviació pel Cluster 260
Mitjana expressio
Temps
Gens pel Cluster 260
Expressio
Temps
NNNNNNNNIND−NNNNNNNNNA 1

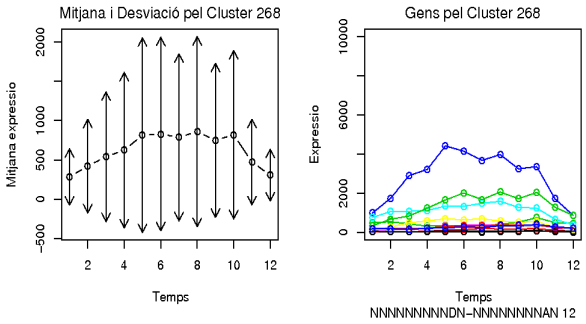

Mitjana i Desviació pel Cluster 261
Mitjana expressio
Temps
Gens pel Cluster 261
Expressio
Temps
NNNNNNNNIND−NNNNNNNVAN 1
Mitjana i Desviació pel Cluster 262
Mitjana expressio
Temps
Gens pel Cluster 262
Expressio
Temps
NNNNNNNNIND−NNNNNNNVNA 1

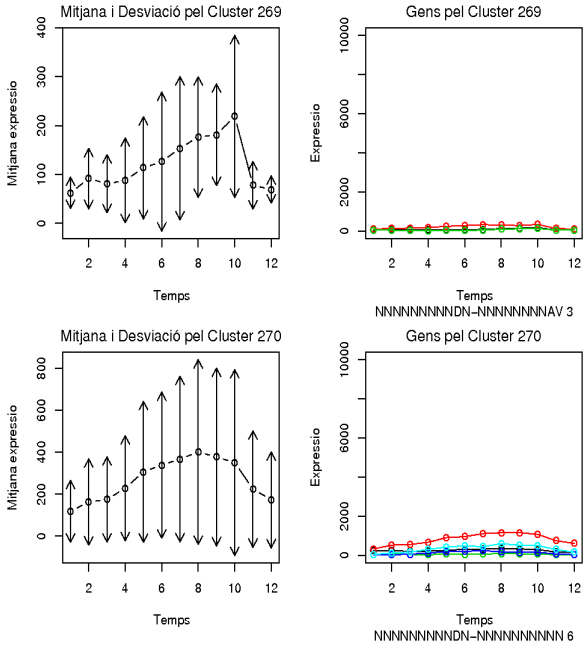

Mitjana i Desviació pel Cluster 263
Mitjana expressio
Temps
Gens pel Cluster 263
Expressio
Temps
NNNNNNNNIND−NNNNNNNVNN 1
Mitjana i Desviació pel Cluster 264
Mitjana expressio
Temps
Gens pel Cluster 264
Expressio
Temps
NNNNNNNNINN−NNNNNNNNAN 4

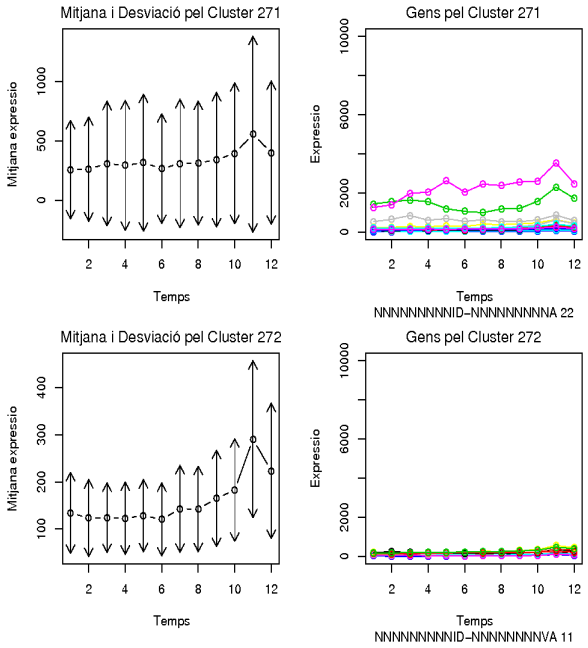

Mitjana i Desviació pel Cluster 265
Mitjana expressio
Temps
Gens pel Cluster 265
Expressio
Temps
NNNNNNNNINN−NNNNNNNNNN 11

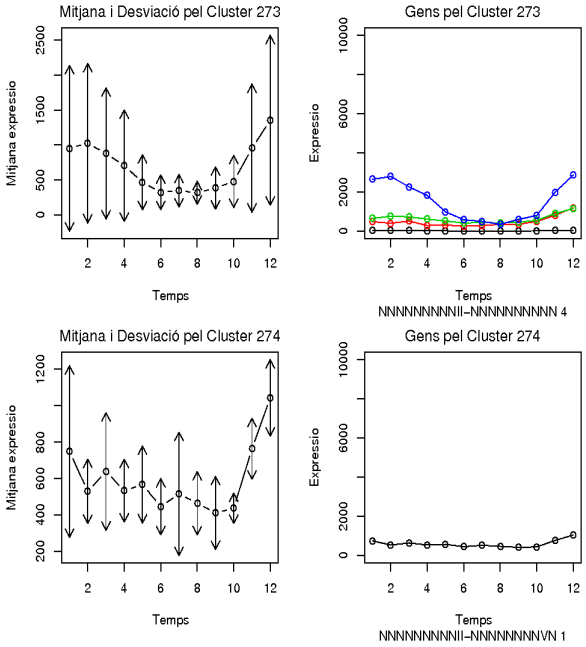

Mitjana i Desviació pel Cluster 266
Mitjana expressio
Temps
Gens pel Cluster 266
Expressio
Temps
NNNNNNNNINN−NNNNNNNVAN 4

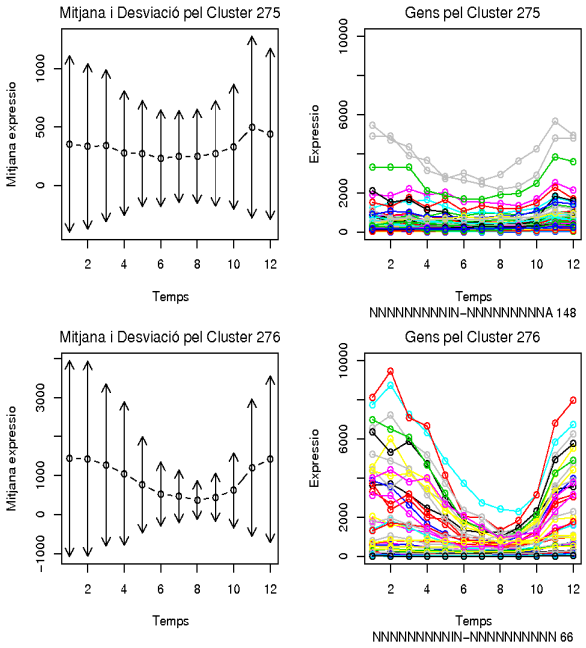

Mitjana i Desviació pel Cluster 267
Mitjana expressio
Temps
Gens pel Cluster 267
Expressio
Temps
NNNNNNNNINN−NNNNNNNVNN 9

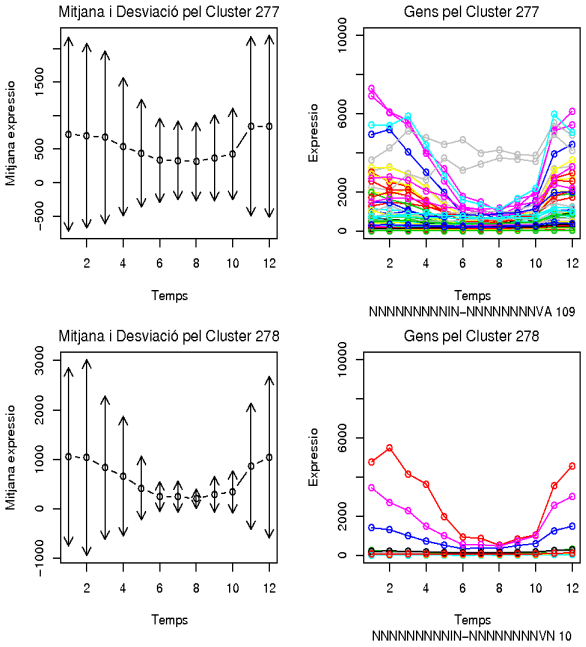

Mitjana i Desviació pel Cluster 268
Mitjana expressio
Temps
Gens pel Cluster 268
Expressio
Temps
NNNNNNNNNDN−NNNNNNNNAN 12

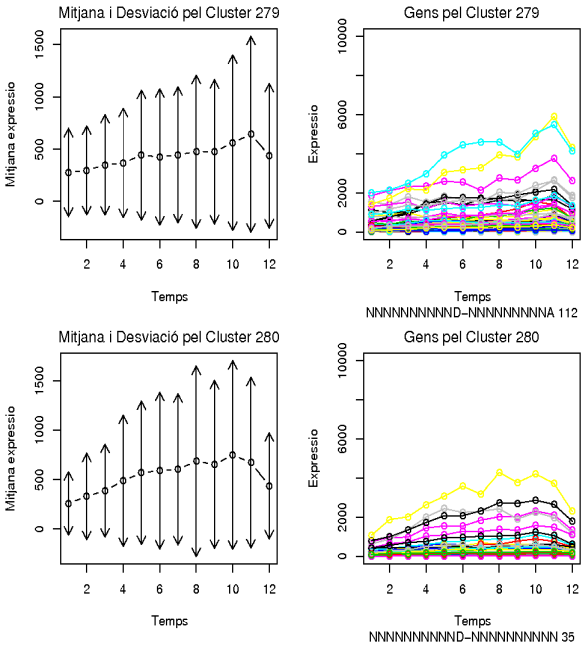

Mitjana i Desviació pel Cluster 269
Mitjana expressio
Temps
Gens pel Cluster 269
Expressio
Temps
NNNNNNNNNDN−NNNNNNNNAV 3
Mitjana i Desviació pel Cluster 270
Mitjana expressio
Temps
Gens pel Cluster 270
Expressio
Temps
NNNNNNNNNDN−NNNNNNNNNN 6

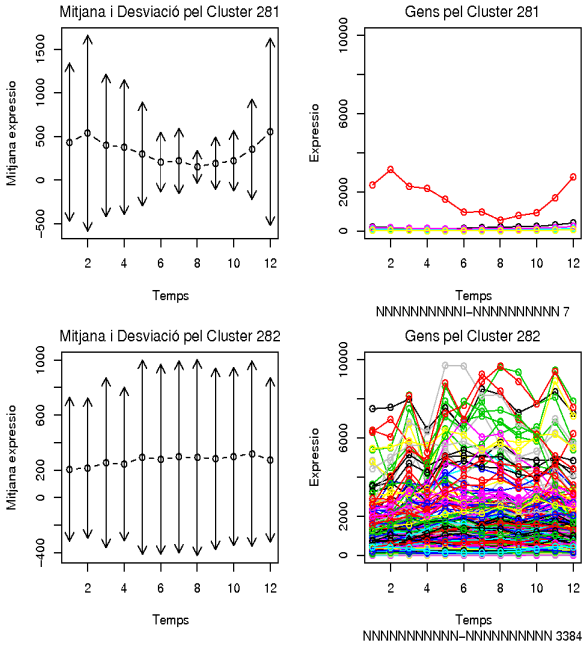

Mitjana i Desviació pel Cluster 271
Mitjana expressio
Temps
Gens pel Cluster 271
Expressio
Temps
NNNNNNNNNID−NNNNNNNNNA 22
Mitjana i Desviació pel Cluster 272
Mitjana expressio
Temps
Gens pel Cluster 272
Expressio
Temps
NNNNNNNNNID−NNNNNNNNVA 11


Mitjana i Desviació pel Cluster 273
Mitjana expressio
Temps
Gens pel Cluster 273
Expressio
Temps
NNNNNNNNNII−NNNNNNNNNN 4
Mitjana i Desviació pel Cluster 274
Mitjana expressio
Temps
Gens pel Cluster 274
Expressio
Temps
NNNNNNNNNII−NNNNNNNNVN 1


Mitjana i Desviació pel Cluster 275
Mitjana expressio
Temps
Gens pel Cluster 275
Expressio
Temps
NNNNNNNNNIN−NNNNNNNNNA 148
Mitjana i Desviació pel Cluster 276
Mitjana expressio
Temps
Gens pel Cluster 276
Expressio
Temps
NNNNNNNNNIN−NNNNNNNNNN 66

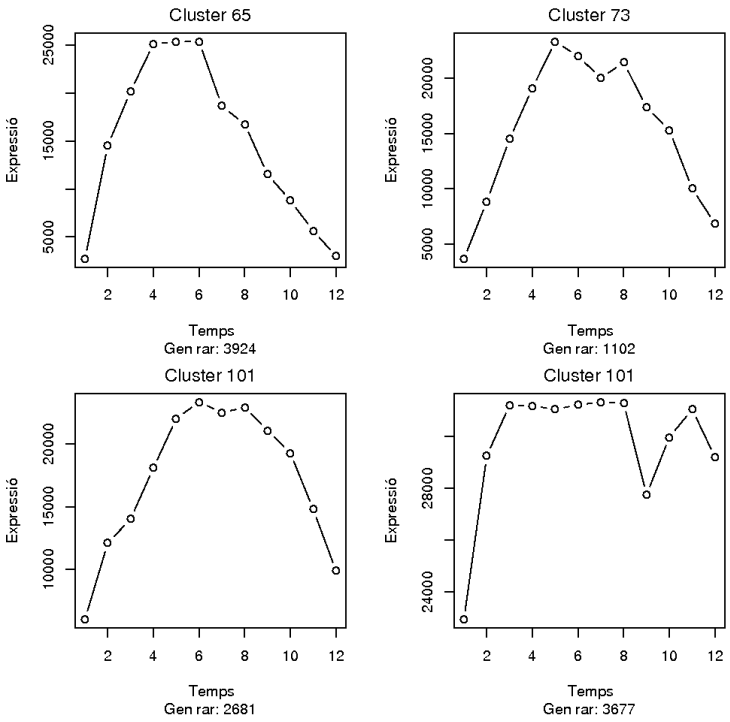

Mitjana i Desviació pel Cluster 277
Mitjana expressio
Temps
Gens pel Cluster 277
Expressio
Temps
NNNNNNNNNIN−NNNNNNNNVA 109
Mitjana i Desviació pel Cluster 278
Mitjana expressio
Temps
Gens pel Cluster 278
Expressio
Temps
NNNNNNNNNIN−NNNNNNNNVN 10

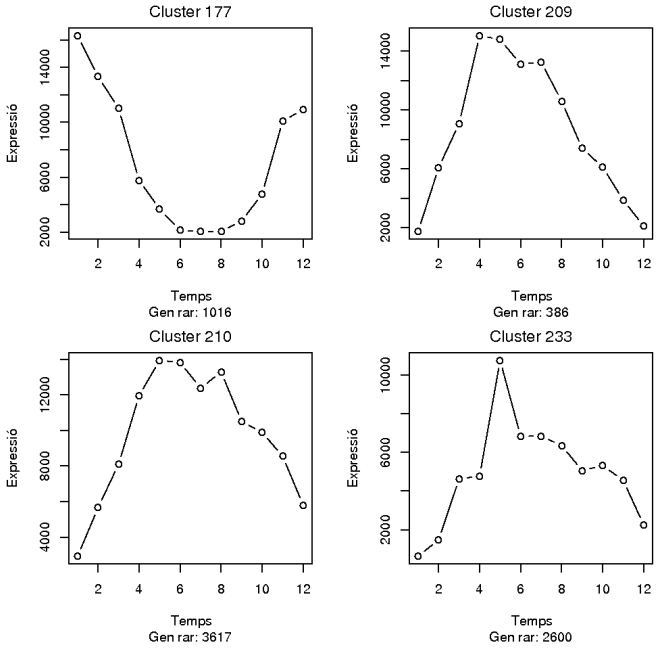

Mitjana i Desviació pel Cluster 279
Mitjana expressio
Temps
Gens pel Cluster 279
Expressio
Temps
NNNNNNNNNND−NNNNNNNNNA 112
Mitjana i Desviació pel Cluster 280
Mitjana expressio
Temps
Gens pel Cluster 280
Expressio
Temps
NNNNNNNNNND−NNNNNNNNNN 35

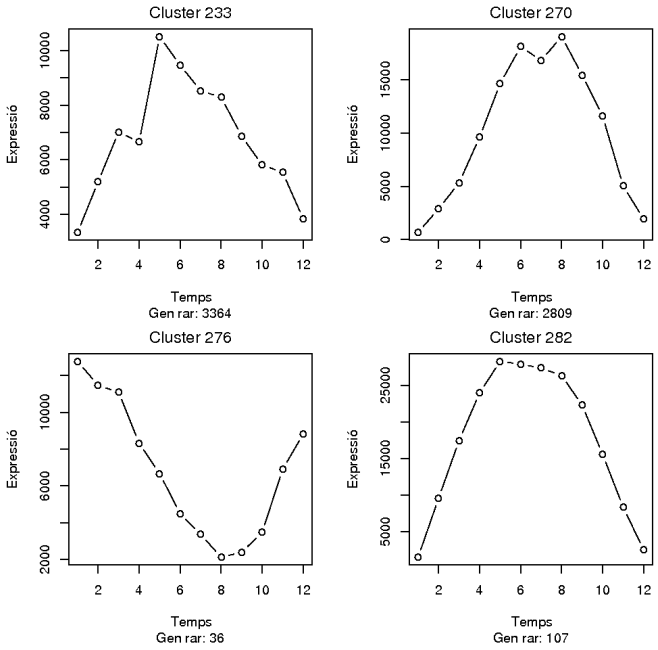

Mitjana i Desviació pel Cluster 281
Mitjana expressio
Temps
Gens pel Cluster 281
Expressio
Temps
NNNNNNNNNNI–NNNNNNNNNN 7
Mitjana i Desviació pel Cluster 282
Mitjana expressio
Temps
Gens pel Cluster 282
Expressio
Temps
NNNNNNNNNNN–NNNNNNNNNN 3384

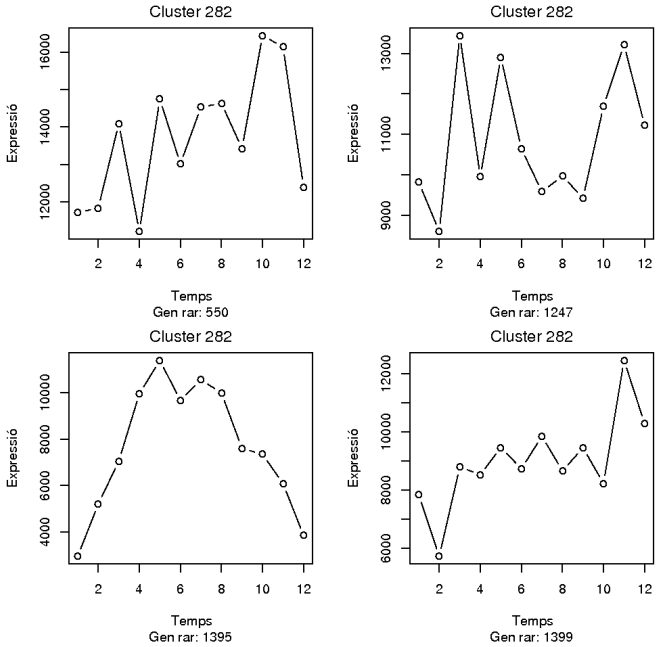
Apèndix D

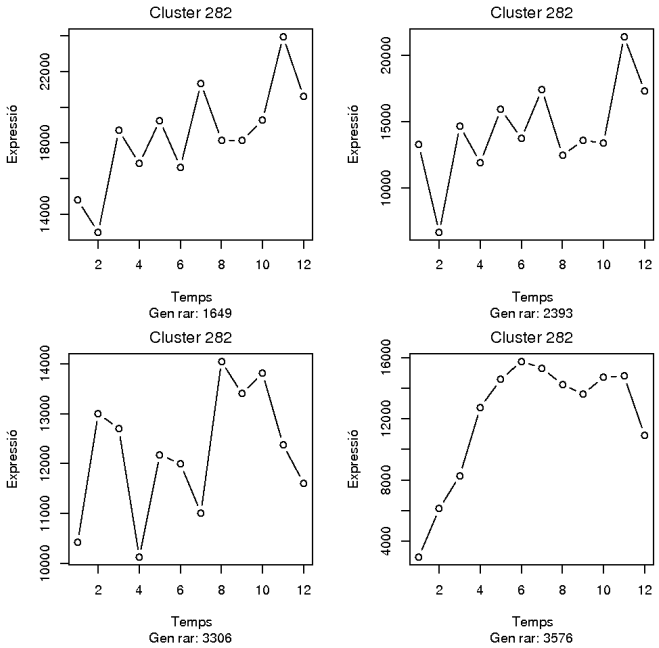
Gràfics dels gens rars

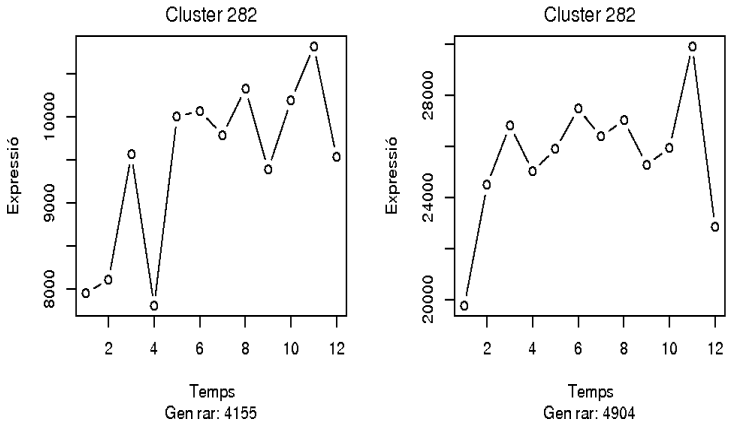



Cluster 177
Expressió
Temps
Gen rar: 1016
Cluster 209
Expressió
Temps
Gen rar: 386
Cluster 210
Expressió
Temps
Gen rar: 3617
Cluster 233
Expressió
Temps
Gen rar: 2600


Cluster 233
Expressió
Temps
Gen rar: 3364
Cluster 270
Expressió
Temps
Gen rar: 2809
Cluster 276
Expressió
Temps
Gen rar: 36
Cluster 282
Expressió
Temps
Gen rar: 107


Cluster 282
Expressió
Temps
Gen rar: 550
Cluster 282
Expressió
Temps
Gen rar: 1247
Cluster 282
Expressió
Temps
Gen rar: 1395
Cluster 282
Expressió
Temps
Gen rar: 1399


Cluster 282
Expressió
Temps
Gen rar: 1649
Cluster 282
Expressió
Temps
Gen rar: 2393
Cluster 282
Expressió
Temps
Gen rar: 3306
Cluster 282
Expressió
Temps
Gen rar: 3576


Cluster 282
Expressió
8000
9000
10000
2
4
6
8
10
12
Temps
Gen rar: 4155
Cluster 282
Expressió
20000
24000
28000
2
4
6
8
10
12
Temps
Gen rar: 4904

Apèndix E

Taules obtingudes de l'Anàlisi d'Enriquiment

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0019219	0.000	1.678	44	66	673	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
GO:0019236	0.001	2.817	6	15	93	response to pheromone
GO:0040029	0.001	2.635	7	16	105	regulation of gene expression, epigenetic
GO:0048869	0.001	1.850	21	36	327	cellular developmental process
GO:0043283	0.002	1.396	158	185	2410	biopolymer metabolic process
GO:0006350	0.002	1.562	47	66	712	transcription
GO:0009451	0.002	2.036	13	24	198	RNA modification
GO:0065004	0.002	2.440	7	16	112	protein-DNA complex assembly
GO:0008643	0.003	3.863	2	8	38	carbohydrate transport
GO:0007165	0.003	1.815	19	32	294	signal transduction
GO:0016072	0.003	1.643	30	45	456	rRNA metabolic process
GO:0006464	0.003	1.520	46	64	704	protein modification process
GO:0016458	0.003	2.435	7	15	105	gene silencing
GO:0030466	0.003	4.218	2	7	31	chromatin silencing at silent mating-type cassette
GO:0006355	0.004	1.547	38	54	580	regulation of transcription, DNA-dependent
GO:0000902	0.005	2.023	11	20	165	cell morphogenesis
GO:0009653	0.005	2.023	11	20	165	anatomical structure morphogenesis
GO:0006270	0.006	3.747	2	7	34	DNA replication initiation
GO:0000750	0.006	4.329	2	6	26	pheromone-dependent signal transduction during conjugation with cellular fusion
GO:0032774	0.006	1.479	45	61	684	RNA biosynthetic process
GO:0000377	0.007	2.395	6	13	92	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
GO:0031137	0.007	4.122	2	6	27	regulation of conjugation with cellular fusion
GO:0043900	0.007	4.122	2	6	27	regulation of multi-organism process
GO:0006260	0.009	1.944	11	19	162	DNA replication
GO:0050789	0.009	1.341	92	111	1414	regulation of biological process

Figura E.1: Taula pel cluster 50 que conté 345 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006139	0.000	1.902	57	83	1743	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
GO:0016072	0.000	2.685	8	20	268	rRNA metabolic process
GO:0042254	0.000	2.243	15	29	470	ribosome biogenesis
GO:0006396	0.001	2.123	14	26	417	RNA processing
GO:0016043	0.001	1.617	63	82	1901	cellular component organization and biogenesis
GO:0006261	0.003	3.390	3	9	90	DNA-dependent DNA replication
GO:0009451	0.005	2.345	7	14	198	RNA modification
GO:0000055	0.006	9.951	0	3	12	ribosomal large subunit export from nucleus
GO:0009062	0.008	8.954	0	3	13	fatty acid catabolic process
GO:0006364	0.009	2.279	6	13	188	rRNA processing
GO:0043283	0.009	1.457	79	95	2410	biopolymer metabolic process
GO:0006298	0.010	5.446	1	4	26	mismatch repair
GO:0000463	0.010	8.139	0	3	14	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)

Figura E.2: Taula pel cluster 49 que conté 173 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006399	0.000	3.951	5	18	191	tRNA metabolic process
GO:0043038	0.000	8.405	1	7	37	amino acid activation
GO:0009058	0.000	1.832	57	78	2010	biosynthetic process
GO:0044260	0.000	1.833	50	71	1781	cellular macromolecule metabolic process
GO:0009101	0.000	4.873	2	9	76	glycoprotein biosynthetic process
GO:0006412	0.000	2.111	19	35	722	translation
GO:0019538	0.000	1.814	50	70	1761	protein metabolic process
GO:0007021	0.000	35.214	0	3	6	tubulin complex assembly
GO:0006432	0.001	Inf	0	2	2	phenylalanyl-tRNA aminoacylation
GO:0043413	0.001	4.504	2	8	72	biopolymer glycosylation
GO:0006409	0.001	7.408	1	5	29	tRNA export from nucleus
GO:0043144	0.002	17.597	0	3	9	snoRNA processing
GO:0006418	0.002	6.212	1	5	34	tRNA aminoacylation for protein translation
GO:0006408	0.004	7.068	1	4	24	snRNA export from nucleus
GO:0006607	0.004	7.068	1	4	24	NLS-bearing substrate import into nucleus
GO:0006608	0.004	7.068	1	4	24	snRNP protein import into nucleus
GO:0006610	0.004	7.068	1	4	24	ribosomal protein import into nucleus
GO:0006407	0.006	6.423	1	4	26	rRNA export from nucleus
GO:0006609	0.006	6.423	1	4	26	mRNA-binding (hnRNP) protein import into nucleus
GO:0006493	0.008	8.788	0	3	15	protein amino acid O-linked glycosylation
GO:0006999	0.008	5.649	1	4	29	nuclear pore organization and biogenesis

Figura E.3: Taula pel cluster 275 que conté 148 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006537	0.002	14.133	0	3	13	glutamate biosynthetic process
GO:0006102	0.004	31.164	0	2	5	isocitrate metabolic process
GO:0006564	0.004	31.164	0	2	5	L-serine biosynthetic process
GO:0051704	0.006	3.194	3	8	129	multi-organism process
GO:0007155	0.006	23.368	0	2	6	cell adhesion
GO:0048610	0.007	2.442	5	12	253	reproductive cellular process
GO:0009064	0.007	4.490	1	5	58	glutamine family amino acid metabolic process
GO:0000747	0.008	3.296	2	7	109	conjugation with cellular fusion
GO:0007039	0.009	3.263	2	7	110	vacuolar protein catabolic process
GO:0030163	0.010	2.248	6	13	297	protein catabolic process

Figura E.4: Taula pel cluster 279 que conté 112 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0042254	0.000	3.142	14	33	658	ribosome biogenesis
GO:0006412	0.000	2.417	16	31	758	translation
GO:0009058	0.001	1.862	42	58	2010	biosynthetic process
GO:0009303	0.001	96.206	0	2	3	rRNA transcription
GO:0016072	0.002	2.276	9	19	456	rRNA metabolic process
GO:0019538	0.004	1.702	37	50	1761	protein metabolic process
GO:0043094	0.005	3.062	3	9	156	metabolic compound salvage
GO:0009100	0.006	3.994	2	6	80	glycoprotein metabolic process
GO:0006490	0.006	24.037	0	2	6	oligosaccharide-lipid intermediate assembly
GO:0044238	0.006	2.019	27	37	1599	primary metabolic process
GO:0009987	0.009	2.319	92	101	4450	cellular process

Figura E.5: Taula pel cluster 277 que conté 109 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006412	0.000	3.959	10	26	758	translation
GO:0009211	0.000	Inf	0	2	2	pyrimidine deoxyribonucleoside triphosphate metabolic process
GO:0009120	0.000	162.188	0	2	3	deoxyribonucleoside metabolic process
GO:0006220	0.001	22.424	0	3	14	pyrimidine nucleotide metabolic process
GO:0019538	0.001	2.267	22	35	1761	protein metabolic process
GO:0044260	0.001	2.228	22	35	1781	cellular macromolecule metabolic process
GO:0009987	0.001	5.873	56	64	4450	cellular process
GO:0009058	0.002	2.081	25	37	2010	biosynthetic process
GO:0006213	0.003	32.413	0	2	7	pyrimidine nucleoside metabolic process
GO:0042254	0.006	2.347	8	15	634	ribosome biogenesis
GO:0007129	0.007	20.246	0	2	10	synapsis
GO:0009262	0.008	17.993	0	2	11	deoxyribonucleotide metabolic process
GO:0042274	0.009	16.681	0	2	12	ribosomal small subunit biogenesis
GO:0000028	0.009	16.191	0	2	12	ribosomal small subunit assembly and maintenance

Figura E.6: Taula pel cluster 276 que conté 66 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0009267	0.000	13.970	0	4	43	cellular response to starvation
GO:0005978	0.000	26.667	0	3	18	glycogen biosynthetic process
GO:0031668	0.001	10.252	0	4	57	cellular response to extracellular stimulus
GO:0009056	0.002	3.088	5	12	610	catabolic process
GO:0019794	0.003	32.544	0	2	10	nonprotein amino acid metabolic process
GO:0031667	0.003	7.313	1	4	78	response to nutrient levels
GO:0009605	0.003	7.214	1	4	79	response to external stimulus
GO:0006073	0.004	10.765	0	3	40	glucan metabolic process
GO:0033692	0.005	9.952	0	3	43	cellular polysaccharide biosynthetic process
GO:0042219	0.005	21.679	0	2	14	amino acid derivative catabolic process
GO:0016051	0.005	6.278	1	4	90	carbohydrate biosynthetic process
GO:0000719	0.008	Inf	0	1	1	photoreactive repair
GO:0018298	0.008	Inf	0	1	1	protein-chromophore linkage
GO:0019547	0.008	Inf	0	1	1	arginine catabolic process to ornithine
GO:0019740	0.009	16.247	0	2	18	nitrogen utilization

Figura E.7: Taula pel cluster 231 que conté 42 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0050789	0.001	2.955	11	21	1441	regulation of biological process
GO:0031323	0.003	2.655	8	16	1063	regulation of cellular metabolic process
GO:0010556	0.004	2.671	7	15	972	regulation of macromolecule biosynthetic process
GO:0010468	0.004	2.644	7	15	980	regulation of gene expression
GO:0000082	0.006	8.919	0	3	50	G1/S transition of mitotic cell cycle
GO:0034401	0.008	Inf	0	1	1	establishment and/or maintenance of chromatin architecture during transcription
GO:0034503	0.008	Inf	0	1	1	protein localization to nucleolar rDNA repeats

Figura E.8: Taula pel cluster 100 que conté 40 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006451	0.007	Inf	0	1	1	translational readthrough

Figura E.9: Taula pel cluster 101 que conté 39 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006511	0.000	12.108	1	10	177	ubiquitin-dependent protein catabolic process
GO:0051603	0.000	11.815	1	10	181	proteolysis involved in cellular protein catabolic process
GO:0043632	0.000	11.605	1	10	184	modification-dependent macromolecule catabolic process
GO:0030163	0.000	9.038	2	12	297	protein catabolic process
GO:0009057	0.000	5.414	3	12	471	macromolecule catabolic process
GO:0044248	0.000	4.258	4	12	582	cellular catabolic process
GO:0000746	0.001	7.967	1	5	112	conjugation
GO:0019953	0.001	7.892	1	5	113	sexual reproduction
GO:0006518	0.002	35.104	0	2	11	peptide metabolic process
GO:0022413	0.004	5.412	1	5	161	reproductive process in single-celled organism
GO:0000754	0.004	24.284	0	2	15	adaptation to pheromone during conjugation with cellular fusion
GO:0006363	0.007	Inf	0	1	1	termination of RNA polymerase I transcription
GO:0006540	0.007	Inf	0	1	1	glutamate decarboxylation to succinate
GO:0022414	0.007	4.013	2	6	262	reproductive process

Figura E.10: Taula pel cluster 280 que conté 35 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0000302	0.002	38.652	0	2	12	response to reactive oxygen species
GO:0006800	0.003	29.715	0	2	15	oxygen and reactive oxygen species metabolic process
GO:0014074	0.006	Inf	0	1	1	response to purine

Figura E.11: Taula pel cluster 210 que conté 29 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0006396	0.000	8.470	2	10	417	RNA processing
GO:0051169	0.000	13.085	1	6	136	nuclear transport
GO:0043628	0.000	62.155	0	3	15	ncRNA 3'-end processing
GO:0006139	0.000	4.935	8	17	1743	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
GO:0016075	0.000	158.485	0	2	5	rRNA catabolic process
GO:0043630	0.000	158.485	0	2	5	ncRNA polyadenylation during polyadenylation-dependent ncRNA catabolic process
GO:0043633	0.000	95.055	0	2	7	modification-dependent RNA catabolic process
GO:0006400	0.001	16.109	0	3	49	tRNA modification
GO:0042273	0.003	12.746	0	3	61	ribosomal large subunit biogenesis
GO:0042254	0.003	5.527	2	6	421	ribosome biogenesis
GO:0051168	0.003	11.753	0	3	72	nuclear export
GO:0034414	0.005	Inf	0	1	1	tRNA 3'-trailer cleavage, endonucleolytic
GO:0043631	0.006	20.593	0	2	25	RNA polyadenylation
GO:0006399	0.006	9.283	0	3	105	tRNA metabolic process
GO:0006409	0.008	17.529	0	2	29	tRNA export from nucleus
GO:0050658	0.008	8.352	0	3	91	RNA transport
GO:0000054	0.008	16.899	0	2	30	ribosome export from nucleus
GO:0033750	0.008	16.899	0	2	30	ribosome localization
GO:0006428	0.009	227.478	0	1	2	isoleucyl-tRNA aminoacylation
GO:0006429	0.009	227.478	0	1	2	leucyl-tRNA aminoacylation
GO:0008033	0.009	15.495	0	2	37	tRNA processing
GO:0006364	0.010	5.488	1	4	188	rRNA processing

Figura E.12: Taula pel cluster 46 que conté 24 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0051261	0.001	62.214	0	2	10	protein depolymerization
GO:0010639	0.001	45.221	0	2	13	negative regulation of organelle organization and biogenesis
GO:0001932	0.004	Inf	0	1	1	regulation of protein amino acid phosphorylation
GO:0051493	0.004	23.642	0	2	23	regulation of cytoskeleton organization and biogenesis
GO:0006538	0.009	237.864	0	1	2	glutamate catabolic process
GO:0030835	0.009	237.864	0	1	2	negative regulation of actin filament depolymerization
GO:0033215	0.009	237.864	0	1	2	iron assimilation by reduction and transport
GO:0051016	0.009	237.864	0	1	2	barbed-end actin filament capping
GO:0032268	0.009	3.983	2	6	432	regulation of cellular protein metabolic process

Figura E.13: Taula pel cluster 233 que conté 23 gens

Gene to GO BP Conditional test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0015866	0.008	249.238	0	1	2	ADP transport

Figura E.14: Taula pel cluster 271 que conté 22 gens

Apèndix F

Implementació

F.1 Fitxer a executar

```
####################
##LECTURA DE DADES##
####################

workingDir <- "C:\\Users\\project"
setwd(workingDir)
dataDir<-workingDir

#paquets necessàries
library(limma)
library(gdata)
library(org.Sc.sgd.db)
library(GOstats)

#dades
annotations <-read.xls("C:\\Users\\project\\annotations1.xls",
                       as.is=T,perl="C:\\Rtools\\perl\\bin\\perl.exe")
save(annotations,file=file.path(dataDir,"annotations.Rda") )
load(file.path(dataDir,"seriesTot.Rda"))
load(file.path(dataDir,"annotations.Rda"))

#Preprocessat (Filtratge i Normalització)
source("C:\\Users\\project\\Preprocessat.r")
Preprocessat(seriesTot)
```

```
#dades normalitzades
load(file.path(dataDir,"seriesNorm.Rda"))

#creació de la matriu booleana de la GO
source("C:\\Users\\project\\MatGOgen.r")
MatGOgen()

#dades noves
load(file.path(dataDir,"seriesNorm2.Rda"))
#matriu booleana
load(file.path(dataDir,"MatGOgen.Rda"))
#ORF dels gens que ens quedem
load(file.path(dataDir,"ORFgens2.Rda"))

#Matriu Mitja per les repliques
mitjaRep<-matrix(0,nrow=n1,ncol=p)
aux<-rep(0,m)
for (j in 1:p){
   for (g in 1:n1){
      for (k in 1:m){
           aux[k]<-c(seriesNorm2[[k]][g,j])
       }
       mitjaRep[g,j]<-mean(aux,na.rm=T)
   }
}
save(mitjaRep, file=file.path(dataDir,"mitjaRep.Rda") )

#index de les dades
p<-length(seriesNorm2[[1]][1,]) #-->temps
n1<-length(seriesNorm2[[1]][,1]) #-->gens
m<-length(seriesNorm2) #-->rèplica

####################################
#ALGORISME DIBC i CALCUL DEL ZSCORE#
####################################
source("C:\\Users\\project\\Dibc.r")
source("C:\\Users\\project\\CalculZscore.r")

#valors per els thresholds:
alf1<-c(0.001,0.005,0.01,0.05)
alf2<-c(0.001,0.005,0.01,0.05)
```

```
numpatro<-matrix(0,nrow=4,ncol=4)
Zscores<-matrix(0,nrow=4,ncol=4)

for (i in 1:4){
    for (j in 1:4){
        numpatro[i,j]<-Dibc(alf1[i],alf2[j])
        load(file.path(dataDir,"clusterTotal.Rda"))
        Zscores[i,j]<-CalculZscore(numpatro[i,j])
    }
}

save(Zscores, file=file.path(dataDir,"Zscores.Rda") )
save(numpatro, file=file.path(dataDir,"numpatro.Rda") )
load(file.path(dataDir,"numpatro.Rda"))
load(file.path(dataDir,"Zscores.Rda"))

#El màxim Zscore(1.77) s'obté per alf1=0.005 i alf2=0.05
#El nombre optims de patrons és 282

png("Zscores.png")
plot(numpatro,Zscores,type="n" ,xlab="Nombre de clusters")
points(numpatro,Zscores,pch=16, col= "darkred")
text(360,1.771,"punt òptim", col="darkgrey")
axis(2,at=c(1.7708), col="darkred")
axis(1,at=c(282), col="darkred")
dev.off()

#Un cop escollit el nombre òptim de clusters fem els gràfics dels clusters:
alf1<-0.005
alf2<-0.05
numpat<-Dibc(alf1,alf2)

#vector que ens indica a quin cluster pertany cada gen
load(file.path(dataDir,"clusterTotal.Rda"))
n<-length(clusterTotal)

#vector de patrons
load(file.path(dataDir,"patrons.Rda"))
#dades noves
load(file.path(dataDir,"seriesNorm2.Rda"))
```

```
#matriu booleana
load(file.path(dataDir,"MatGOgen2.Rda"))
#ORF dels gens que ens quedem
load(file.path(dataDir,"ORFgens2.Rda"))

####################
##GRAFICS CLUSTERS##
####################

load(file.path(dataDir,"mitjaRep.Rda"))
source("C:\\Users\\project\\GrafClus.r")
GrafClus()

######################
#ANALISI ENRIQUIMENT:
######################
ORFgens2
clusterTotal
TaulaGens<-cbind(ORFgens2,clusterTotal)
save(TaulaGens, file=file.path(dataDir,"TaulaGens.Rda") )
load("TaulaGens.Rda")
GOAnalysis()
```

F.2 Funció Preprocessat()

```
Preprocessat<-function(seriesTot){

nom<-c("A", "B", "C", "D", "E", "F", "G", "H")
fileNames<-paste("Serie",nom, sep="")
nfiles<-length(fileNames)
temps<-
c("t0","t3","t6","t9","t12","t15","t18","t21","t25","t30","t45","t60")
n<-nrow(seriesTot[[1]])

########################
#1.-CONTROL DE QUALITAT#
########################
#Definicio de funcions:
```

```
#a) MA i MA gràfic
M <- function(x){
if((!is.na(x[1]) & !is.na(x[2])) &
       (x[1] > 0.1 & x[2] > 0.1))
       log2(x[1]/x[2])
else NA
}

A <- function(x){
if ((!is.na(x[1]) & !is.na(x[2])) &
       (x[1] > 0.1 & x[2] > 0.1))
       log2(sqrt(x[1]*x[2]))
else NA
}

CalculaMiA<-function(x1,x2){
   MiA<-matrix(0,nrow=length(x1),ncol=2)
   colnames(MiA)<-c("M","A")
   Mat<-cbind(x1,x2)
   Mval<-apply(Mat,1,M )
   Aval<-apply(Mat,1,A )
   MiA[,"M"] <- Mval
   MiA[,"A"] <- Aval
   return(MiA)
}
MAGraf<-function(MAObj,Gtitol,titol, limY =c(-10,10)){
  MainTitle<-paste(Gtitol)
  plotMA(MAObj, main=titol, ylim=limY)
  abline(h=0)
  mtext(MainTitle, line=0.5,outer=T)
}

#b)funcio que posa NA si és menor que 0.1
Gthan0.1<-function(x){
   for (i in 1:nrow(x))
       for (j in 1:ncol(x))   if (x[i,j]<=0.1)  x[i,j]<-NA
   return(x)
}

#c)compta per cada experiment el nombre de lectures de 12 < 0.1
TotLT0.1<-function(x) sum(x<0.1)
```

```
#d)compta quants experiments el nombre de lectures de 12 <0.1 és < 9
sel0.1<- function(x) sum(x < 9)

###############
##2.-FILTRATGE#
###############
#Selecció de gens: status="gen" que en cada rèplica tingui màxim 50% "0"
numgen<-seq(1,dim(seriesTot[[1]])[1])
seriesTot2<-list()
for (i in 1:8) seriesTot2[[i]]<-cbind(seriesTot[[i]],numgen)
seriesSelected<-list()
LT0.1<-matrix(0, nrow=n, ncol=8)

#conta,per a cada experiment, el nombre de lectures (de 12) < 0.1
for (i in 1:8)  LT0.1[,i]<-apply(seriesTot2[[i]],1,TotLT0.1)

#conta a quants experiments el valor anterior és < 9
sel0.1<- apply(LT0.1,1,sel0.1)

#Si a més de 4 experiments a on el valor és < 9 seleccionem el gen
Sel<- annotations$status=="gen" & sel0.1 > 4
for (i in 1:8){
   seriesSelected[[i]]<-seriesTot[[i]][Sel,]
   names(seriesSelected)[i] <-fileNames[i]
}
save(seriesSelected, file=file.path(dataDir,"seriesSelected.Rda") )
n<-nrow(seriesSelected[[1]])
grups<-factor(rep(1:8,rep(n,8)))

seiesNA<-list()
RefAll<-matrix(0, nrow=n, ncol=12)
vec<-numeric(8)

seriesNA<-lapply(seriesSelected, Gthan0.1)
for (ig in 1:n){
   for (it in 1:12){
        for (is in 1:8)  vec[is]<-seriesNA[[is]][ig,it]
            RefAll[ig,it]<-mean(vec,na.rm=T) #mitja de les 8 repliques
   }
}
```

```
#########################
##Creació objecte MAList#
#########################

seriesMA<-list()
for (i in 1:12){
   MA <- new("MAList")
   valM<-NULL
   valA<-NULL
   Ref<- RefAll[,i]
   for (j in 1:8){
       S <-seriesSelected[[j]][,i]
       MiA<-CalculaMiA(S,Ref)
       valM <- cbind(valM,MiA[,"M"])
       valA <- cbind(valA,MiA[,"A"])
   }
    MA$M<-valM
    MA$A<-valA
    seriesMA[[i]]<-MA
    names(seriesMA)[i]<-temps[i]
}

##Gràfics inicials:
for (i in 1:12){
  pdf(paste(temps[i],"MAplot.pdf",sep="-") )
  opt<-par(mfrow=c(3,3),pty="m", oma=c(0,0,2,0),mar=c(5,4,2,2), font.main=1)
  for (j in 1:length(nom)){
   MA <- new("MAList")
   MA$M<-  seriesMA[[i]]$M[,j]
   MA$A<-seriesMA[[i]]$A[,j]
   MAGraf(MA,temps[i],nom[j])
  }
  dev.off()
}

for (i in 1:12) {
  pdf(paste(temps[i],"Boxplot.pdf",sep="-") )
  SA <-seriesSelected[[1]][,i]
  SB <-seriesSelected[[2]][,i]
```

```
  SC <-seriesSelected[[3]][,i]
  SD <-seriesSelected[[4]][,i]
  SE <-seriesSelected[[5]][,i]
  SF <-seriesSelected[[6]][,i]
  SG <-seriesSelected[[7]][,i]
  SH <-seriesSelected[[8]][,i]
  dadesBoxplot<-c(SA,SB,SC,SD,SE,SF,SG,SH)
  boxplot(log2(dadesBoxplot)~grups,main=temps[i], names=nom)
  dev.off()
}

############################
##3.-Normalització (Within)#
############################

NormFun<-function(x) normalizeWithinArrays(x, method="loess")
SeriesNormMA<-lapply(seriesMA, NormFun)

##Pas de l'objecte MAlist a la serie original normalitzada##

seriesNorm<-list()
matriu0<-matrix(0,nrow=n,ncol=12)
for (j in 1:8) seriesNorm[[j]]<-matriu0
for (i in 1:12){
    MA2<-SeriesNormMA[[i]]
    valM2<-MA2$M
    valA2<-MA2$A
    for (j in 1:8){
        x1<-(2^valA2[,j])*sqrt(2^valM2[,j])
        seriesNorm[[j]][,i]<-x1
    }
}
save(seriesNorm, file=file.path(dataDir,"seriesNorm.Rda") )

#Gràfics normalitzats
for (i in 1:12){
  pdf(paste(temps[i],"MAplotNorm.pdf",sep="-") )
  opt<-par(mfrow=c(3,3),pty="m", oma=c(0,0,2,0),mar=c(5,4,2,2), font.main=1)
  for (j in 1:length(nom)){
    MA <- new("MAList")
   MA$M<-  SeriesNormMA[[i]]$M[,j]
```

```
   MA$A<-SeriesNormMA[[i]]$A[,j]
   MAGraf(MA,temps[i],nom[j])
  }
  dev.off()
}

for (i in 1:12) {
  pdf(paste(temps[i],"BoxplotNorm.pdf",sep="-") )
  SA <-seriesNorm[[1]][,i]
  SB <-seriesNorm[[2]][,i]
  SC <-seriesNorm[[3]][,i]
  SD <-seriesNorm[[4]][,i]
  SE <-seriesNorm[[5]][,i]
  SF <-seriesNorm[[6]][,i]
  SG <-seriesNorm[[7]][,i]
  SH <-seriesNorm[[8]][,i]
  dadesBoxplot<-c(SA,SB,SC,SD,SE,SF,SG,SH)
  boxplot(log2(dadesBoxplot)~grups,main=temps[i], names=nom)
  dev.off()
}
}
```

F.3 Funció Dibc()

```
Dibc<-function(alf1,alf2){

############################
#Diferència de primer ordre#
############################

#Serie Diferencia de temps:
seriesDifTemps<-list()
matriu0<-matrix(0, nrow=n1  , ncol=p-1)
for (k in 1:m) seriesDifTemps[[k]]<-matriu0
for (k in 1:m){
   for (j in 1:(p-1)){
       seriesDifTemps[[k]][,j]<-seriesNorm2[[k]][,j+1]-seriesNorm2[[k]][,j]
   }
}
```

```
#Juntem la llista obtenint una matriu de nx88
seriesJuntes<-seriesDifTemps[[1]]
for (k in 2:m){
    seriesJuntes<-cbind(seriesJuntes,seriesDifTemps[[k]])
}

#Construccio de la matriu de disseny
X<-matrix(0,nrow=88,ncol=11)
aux<-c(1,rep(0,10))

X[,1]<-rep(aux,8)
for (i in 1:10){
    X[,i+1]<-c(rep(0,i),rep(aux,7),aux[1:(11-i)])
}

#Ajust model lineal:
fit1<-lmFit(seriesJuntes,design=X,na.rm=T)
gllres<-fit1$df.residual


#Calcul estadistic t moderat:
y1<-ebayes(fit1)$t
gllprior<-ebayes(fit1)$df.prior
identgen<-seq(1,n1) #vector identificador del tots els gens
rownames(y1)<-identgen

#treiem els gens amb coefs amb NA
aux<-apply(y1,1,sum)
y1<-subset(y1, !is.na(aux))
for (k in 1:m) seriesNorm2[[k]]<-subset(seriesNorm2[[k]],!is.na(aux))
MatGOgen2<-subset(MatGOgen,!is.na(aux))
ORFgens2<-subset(ORFgens2,!is.na(aux))
n<-dim(y1)[1] #-->gens  no NA
identgen1<- seq(1:n)
gll<-gllres+gllprior
save(seriesNorm2, file=file.path(dataDir,"seriesNorm2.Rda") )
save(MatGOgen2, file=file.path(dataDir,"MatGOgen2.Rda") )
save(ORFgens2, file=file.path(dataDir,"ORFgens2.Rda") )
```

```
############################
#Matriu F de patró simbòlic#
############################

#Calcul del valor critic:

T<-matrix(0,nrow=n ,ncol=p-1)
T<-qt((1-alf1/2),gll)

#Matriu de patrons:
F<-matrix("N",nrow=n,ncol=(p-1))
for (g in 1:n){
   for(j in 1:(p-1)){
      if (y1[g,j]>T[g]) F[g,j]="I"
      if (y1[g,j]<(-T[g])) F[g,j]="D"
   }
}

#Comptatge de quantes lletres D o I tenen els gens
numF<-rep(0,n)
for (g in 1:n){
    for (j in 1:(p-1)){
       if (F[g,j]!="N") numF[g]<-numF[g]+1
    }
}

#Seleccionem els 1ers patrons amb D o I
Fsel<-F[numF>0,]
identgen3<-subset(identgen1,numF>0)#id dels 1ers patrons amb D o I
n3<-length(identgen3)#numero de 1ers patrons amb D o I

#boolea que em diu quins 1ers patrons tenen N, D o I seguides
id2seguits<-matrix(0,nrow=n3,ncol=10)
for (g in 1:n3){
   for (i in 1:10){
     if ((Fsel[g,i]!= "N") & (Fsel[g,i+1]!= "N") )  id2seguits[g,i]<-1
     if ((Fsel[g,i]!= "N") & (Fsel[g,i+1]== "N") )  id2seguits[g,i]<-1
     if ((Fsel[g,i]== "N") & (Fsel[g,i+1]!= "N") )  id2seguits[g,i]<-1
   }
}
```

```
#Vector de subpatró per la primera diferència
F2<-rep(0,n)
for (g in 1:n){
   F2[g]<-paste(F[g,],collapse="")
}


############################
#Diferència de segon ordre#
############################
n2patro<-sum(id2seguits) #numero gens a calcular el segon patró
y1b<-subset(y1,numF>0)
y2<-matrix(0,nrow=n3,ncol=p-2)
for (g in 1:n3) {
   for(j in 1:(p-2)){
      y2[g,j]<-y1b[g,j+1]-y1b[g,j]
   }
}

############################
#Matriu S de patró simbòlic#
############################

#Creacio del valor critic:
q1<-rep(0,1000)
q2<-rep(0,1000)
for (i in 1:1000){
   mostra1<-rt(10000,7)
   mostra2<-rt(10000,7)
   mostra<-mostra1-mostra2
   q1[i]<-quantile(mostra, probs=(alf2/2))
   q2[i]<-quantile(mostra, probs=(1-alf2/2))
}
Q1<-median(q1)
Q2<-median(q2)

#matriu de patró S només pels gens amb patró F amb més de 2 D o I seguides
Saux<-matrix("N",nrow=n3,ncol=p-2)
for (g in 1:n3) {
   for(j in 1:(p-2)){
      if ((id2seguits[g,j]==1) & (y2[g,j]>Q2)) Saux[g,j]="V" #convex
```

```
      if ((id2seguits[g,j]==1) & (y2[g,j]<Q1)) Saux[g,j]="A" #concau
   }
}

S<-matrix("N",nrow=n,ncol=p-2)
for (g in 1:n){
   for (w in 1:n3){
      if (identgen3[w]==identgen1[g]) S[g,]<-Saux[w,]
   }
}


#Vector de subpatró per la segona diferència
S2<-rep(0,n)
for (g in 1:n){
   S2[g]<-paste(S[g,],collapse="")
}


#####################################
#Matriu H de patró simbòlic combinat#
#####################################

H<-cbind(F,"-",S)
identgen<-seq(1,n) #vector identificador dels gens
rownames(H)<-identgen

#Creacio dels patrons com a caracters
H2<-rep(0,n)
for (g in 1:n){
   H2[g]<-paste(H[g,],collapse="")
}

#Calcul de les frequencies per cada patró
taula<-as.data.frame(table(H2))

#####################################
##Assignació de cada gen a un cluster#
#####################################
patrons<-as.vector(taula$H2)
numpat<-length(patrons) #numero de patrons diferents
```

```
clusterTotal<-rep(0,n)
for (g in 1:n){
  for (j in 1:numpat){
    if (H2[g]==patrons[j]) clusterTotal[g]<-j
  }
}

save(clusterTotal, file=file.path(dataDir,"clusterTotal.Rda") )
save(patrons, file=file.path(dataDir,"patrons.Rda") )
return(numpat)
}
```

F.4 Funció MatGOgen()

```
#######################################################
#MATRIU BOOLEANA AMB LES ANOTACIONS DE LA GO I GENS:#
#######################################################
MatGOgen<-function(){

#Obtenció de vector amb els ORF:
load(file.path(dataDir,"seriesSelected.Rda"))
load(file.path(dataDir,"annotations.Rda"))

idGen<-seq(1,dim(seriesTot[[1]])[1])            #id per tots els gens
ORFselec<-seriesSelected[[1]][,13])             #ORF dels gens seleccionats
annotations2<-cbind(annotations$ORF.name,idGen) #ORF i id per tots els gens

#Matriu amb ORF i id pels gens seleccionats
n<-length(seriesNorm[[1]][,1])
ORFgens<-matrix(0,nrow=n,ncol=2)
for (i in 1:length(idGen)){
   for (j in 1:length(ORFselec)){
      if (annotations2[i,2]==ORFselec[j]) ORFgens[j,]<-annotations2[i,]
   }
}

#Guardem els ORF's dels gens
ORFgens2<-ORFgens[,1]
save(ORFgens2, file=file.path(dataDir,"ORFgens.Rda") )
```

```
#Obtenció de les GO annotations:
GOTab<-toTable(org.Sc.sgdGO) #Taula amb tota la info de la GO
GOORF<-GOTab$systematic_name #Vector amb els ORF de la GO
numGOORF<-length(GOORF)
#Vector que em diu si el gen es troba a la GO o no:
NomsGOgen<-rep(0,n)
for (i in 1:n){
   for (j in 1:numGOORF){
      if (ORFgens[i]==GOORF[j]) NomsGOgen[i]<-1
   }
}

#Filtrem els gens i els noms de gens que no tenen nom a la GO:
seriesNorm2<-list()
for (i in 1:8) seriesNorm2[[i]]<-subset(seriesNorm[[i]],NomsGOgen==1)
ORFgens1<-subset(ORFgens,NomsGOgen==0) #gens que treurem
idgens<-seq(1,n)
ORFgens2<-cbind(ORFgens[,1],idgens)
ORFgens2<-subset(ORFgens2,NomsGOgen==1) #gens que ens quedem (5326 gens)
save(ORFgens2, file=file.path(dataDir,"ORFgens2.Rda") )
save(seriesNorm2, file=file.path(dataDir,"seriesNorm2.Rda") )

#Guardem els termes de la GO i el nombre de termes diferents:
GOid<-as.data.frame(table(GOTab$go_id))$Var1
numGO<-dim(as.data.frame(table(GOTab$go_id)))[1]

#Matriu de 0 i 1 dels termes de la GO amb els gens
MatGOgen<-matrix(0,nrow=n,ncol=numGO)
colnames(MatGOgen)<-GOid
for (i in 1:n){
   for (j in 1:numGO){
      aux<-subset(GOTab,GOORF==ORFgens2[i])
      aux2<-dim(aux)[1]
      for (k in 1:aux2){
         if (aux$go_id[k]==colnames(MatGOgen)[j]) MatGOgen[i,j]<-1
      }
   }
}

save(MatGOgen, file=file.path(dataDir,"MatGOgen.Rda") )
}
```

F.5 Funció CalculZscore()

```
#############################################################
#CALCUL DEL ZSCORE PER ESCOLLIR EL NOMBRE OPTIM DE CLUSTERS:#
#############################################################

CalculZscore<-function(numpat){

##################
#Calcul de MIreal:
##################

MI<-rep(0,numpat)
midaclusters<-rep(0,numpat)
for (k in 1:numpat){
   Clus<-subset(mitjaRep,clusterTotal==k)
   c<-dim(Clus)[1]
   midaclusters[k]<-c
   MatGOclus<-subset(MatGOgen2,clusterTotal==k)
   TabGOclus<-matrix(0,nrow=c,ncol=2)
   ni<-rep(0,c)
   nj<-rep(0,2)
   for (i in 1:c){
      TabGOclus[i,]<-table(MatGOclus[i,])
      for (j in 1:2){
         ntot<-sum(TabGOclus)
         ni[i]<-sum(TabGOclus[i,])
         nj[j]<-sum(TabGOclus[,j])
         MIclus<-sum((TabGOclus[i,j]/ntot)*log2(TabGOclus[i,j]/ntot))+
                 sum((ni[i]/ntot)*log2(ni[i]/ntot))+sum((nj[j]/ntot)*
                 log2(nj[j]/ntot))
      }
   }
   MI[k]<-MIclus
}

MIreal<-mean(MIclus)
```

```
##################
#Calcul de MIrand:
##################

MIrep<-rep(0,2000)
for (l in 1:2000){
   clusterRand<-sample(clusterTotal)
   MI<-rep(0,numpat)
   for (k in 1:numpat){
      Clus<-subset(mitjaRep,clusterRand==k)
      c<-dim(Clus)[1]
      MatGOclus<-subset(MatGOgen2,clusterRand==k)
      TabGOclus<-matrix(0,nrow=c,ncol=2)
      ni<-rep(0,c)
      nj<-rep(0,2)
      for (i in 1:c){
         TabGOclus[i,]<-table(MatGOclus[i,])
         for (j in 1:2){
            ntot<-sum(TabGOclus)
            ni[i]<-sum(TabGOclus[i,])
            nj[j]<-sum(TabGOclus[,j])
            MIclus<-sum((TabGOclus[i,j]/ntot)*log2(TabGOclus[i,j]/ntot))+
                    sum((ni[i]/ntot)*log2(ni[i]/ntot))+sum((nj[j]/ntot)*
                    log2(nj[j]/ntot))
         }
      }
      MI[k]<-MIclus
   }
MIrep[l]<-mean(MIclus)
}

MIrand<-mean(MIrep)
MIrand_sd<-sd(MIrep)

Zscore<-(MIrand-MIreal)/MIrand_sd
return(Zscore)
}
```

F.6 Funció GrafClus()

```
GrafClus<-function(){

################################################
#Gràfic de la Mitja i Desviació per cada Cluster#
################################################

temps<-seq(1,p)
ident<-seq(1,n)

pdf(paste("Clusters.pdf",sep="-") )
opt<-par(mfrow=c(2,2),pty="m", oma=c(0,0,2,0),mar=c(5,4,2,2), font.main=1)

for(q in 1:numpat){


#Filtratge per la creació de cada cluster
Clus<-list()
for(k in 1:m){
   Clus[[k]]<-subset(seriesNorm2[[k]],clusterTotal==q)
   c<-dim(Clus[[k]])[1]
}

#Mitja i Desviació per cada instant de temps i cluster
mitjaClus<-rep(0,p)
desvClus<-rep(0,p)
aux<-matrix(0,nrow=c,ncol=m)
for (j in 1:p){
   for (k in 1:m){
      for (g in 1:c){
         aux[g,k]<-c(Clus[[k]][g,j])
      }
   }
   mitjaClus[j]<-mean(as.vector(aux),na.rm=T)
   desvClus[j]<-sd(as.vector(aux),na.rm=T)
}
```

```
######Grafic Mitja i Desviació per cada Cluster

plot(temps,mitjaClus,type='b', xlab='Temps', ylab='Mitjana expressio'
     ,ylim=c(min(-desvClus+mitjaClus),max(desvClus+mitjaClus)))
arrows(x0=temps, y0=desvClus+mitjaClus,x1=temps,
       y1=-desvClus+mitjaClus,code=3, length=0.1)
title(paste('Mitjana i Desviació pel Cluster',q, sep=" "))

#Mitja per les repliques tipificades
ClusMeanRep2<-subset(mitjaRep,clusterTotal==q)

###############################################
#Gràfic de cada Cluster amb tots els gens
#per tal d'identificar els gens rars
###############################################

plot(temps,ClusMeanRep2[1,],type='p',xlab="Temps",ylab="Expressio"
     ,ylim=c(min(mitjaRep,na.rm=T),max(mitjaRep,na.rm=T)))
lines(ClusMeanRep2[1,],col=1)
if (c>1){
  for (i in 2:c){
   points(ClusMeanRep2[i,],col=i)
   lines(ClusMeanRep2[i,],col=i)
  }
}
title(paste('Gens pel Cluster',q, sep=" "),sub=paste(patrons[q],c,sep=" "))

}

dev.off()

################################################
#Anem a veure que passa al treure els gens rars
################################################
temps<-seq(1,p)
ident<-seq(1,n)
lim<-10000
gensrars<-0
```

```
#######################################
#Mateix gràfic traient els gens rars:
#######################################

pdf(paste("Clusters_rars.pdf",sep="-") )
opt<-par(mfrow=c(2,2),pty="m", oma=c(0,0,2,0),mar=c(5,4,2,2), font.main=1)

for(q in 1:numpat){

#Filtratge per la creació de cada cluster
Clus<-list()
for(k in 1:m){
   Clus[[k]]<-subset(seriesNorm2[[k]],clusterTotal==q)
   c<-dim(Clus[[k]])[1]
}


#identificador dels gens del cluster:
idclus<-subset(ident,clusterTotal==q)

#Mitja i Desviació per cada instant de temps i cluster
mitjaClus<-rep(0,p)
desvClus<-rep(0,p)
aux<-matrix(0,nrow=c,ncol=m)
for (j in 1:p){
   for (k in 1:m){
      for (g in 1:c){
         aux[g,k]<-c(Clus[[k]][g,j])
      }
   }
   mitjaClus[j]<-mean(as.vector(aux),na.rm=T)
   desvClus[j]<-sd(as.vector(aux),na.rm=T)
}



#Mitja per les repliques
ClusMeanRep<-matrix(0,nrow=c,ncol=p)
aux2<-rep(0,m)
for (g in 1:c){
   for (j in 1:p){
```

```
      for (k in 1:m){
         aux2[k]<-c(Clus[[k]][g,j])
      }
      ClusMeanRep[g,j]<-mean(aux2,na.rm=T)
   }
}

################################################
#Gràfic de cada Cluster amb només els gens rars#
################################################

ClusMeanRepb<-ifelse(is.na(ClusMeanRep),0,ClusMeanRep)

#variable booleana per trobar els gens rars que superen el valor limit
qwe<-rep(0,c)
for (g in 1:c){
   for (j in 1:p){
      if (ClusMeanRepb[g,j]>lim) qwe[g]=1
   }
}

rars<-subset(idclus,qwe==1) #id dels gens rars
gensrars<-c(gensrars,rars) #anem guardant els id dels gens rars
idclus2<-subset(idclus,qwe==0) #id dels gens no rars

#treiem els gens rars del cluster
ClusMeanRep2<-subset(ClusMeanRep,qwe==0)

rar<-length(rars) #nombre de gens rars

#Grafics dels gens rars:
if (rar!=0) for (i in 1:rar){
   plot(temps,mitjaRep[rars[i],],type="b",xlab="Temps", ylab="Expressió")
   title(paste('Cluster',q,sep=" "),sub=paste('Gen rar:',rars[i],sep=" "))
}

}
dev.off()
```

```
################################################
#GRÀFICS SENSE ELS GENS RARS
################################################

################################################
#Gràfic de la Mitja i Desviació per cada Cluster#
################################################

temps<-seq(1,p)
ident<-seq(1,n)

gensrars<-gensrars[-1]
r<-length(gensrars)

pdf(paste("Clusters_sense_rars.pdf",sep="-") )
opt<-par(mfrow=c(2,2),pty="m", oma=c(0,0,2,0),mar=c(5,4,2,2), font.main=1)

for(q in 1:numpat){

#Filtratge per la creació de cada cluster
Clus<-list()
for(k in 1:m){
   Clus[[k]]<-subset(seriesNorm2[[k]],clusterTotal==q)
   c<-dim(Clus[[k]])[1]
}

idclus<-subset(ident,clusterTotal==q)


#variable booleana per trobar els gens rars del cluster
qwe2<-rep(0,c)
for (g in 1:c){
   for (j in 1:r){
      if (idclus[g]==gensrars[j]) qwe2[g]<-1
   }
}

#traiem els gens rars del cluster
Clus2<-list()
for(k in 1:m){
    Clus2[[k]]<-subset(Clus[[k]],qwe2==0)
```

```
    c1<-dim(Clus2[[k]])[1]
}

if (dim(Clus2[[1]])[1]>0){

#Mitja i Desviació per cada instant de temps i cluster
mitjaClus<-rep(0,p)
desvClus<-rep(0,p)
aux<-matrix(0,nrow=c,ncol=m)
for (j in 1:p){
   for (k in 1:m){
      for (g in 1:c1){
         aux[g,k]<-c(Clus2[[k]][g,j])
      }
   }
   mitjaClus[j]<-mean(as.vector(aux),na.rm=T)
   desvClus[j]<-sd(as.vector(aux),na.rm=T)
}


######Grafic Mitja i Desviació per cada Cluster

plot(temps,mitjaClus,type='b', xlab='Temps', ylab='Mitjana expressio'
 ,ylim=c(min(-desvClus+mitjaClus),max(desvClus+mitjaClus)))
arrows(x0=temps, y0=desvClus+mitjaClus,x1=temps,
      y1=-desvClus+mitjaClus,code=3,length=0.1)
title(paste('Mitjana i Desviació pel Cluster',q, sep=" "))


#Mitja per les repliques del cluster sense gens rars
ClusMeanRep2<-subset(mitjaRep,clusterTotal==q)
ClusMeanRep3<-subset(ClusMeanRep2,qwe2==0)

################################################
#Gràfic tipificat de cada Cluster amb tots els gens
#per tal d'identificar els gens rars
################################################
plot(temps,ClusMeanRep3[1,],type='p',xlab="Temps",ylab="Expressio",ylim=c(0,lim))
lines(ClusMeanRep3[1,],col=1)
if (c1>1){
for (i in 2:c1){
```

```
   points(ClusMeanRep3[i,],col=i)
   lines(ClusMeanRep3[i,],col=i)
    }
}
title(paste('Gens pel Cluster',q, sep=" "),
      sub=paste(patrons[q],c,sep=" "))

}

#variable booleana per trobar els gens rars que superen el valor limit
qwe<-rep(0,c)
for (g in 1:c){
   for (j in 1:p){
      if (ClusMeanRep2[g,j]>lim) qwe[g]=1
   }
}

rars<-subset(idclus,qwe==1) #id dels gens rars
rar<-length(rars)

if (c<=rar){
  plot(temps, type="n")
  text(temps[7],temps[7],"Tots els gens són rars")
  title(paste('Gens pel Cluster',q, sep=" "),
        sub=paste(patrons[q],c,sep=" "))
}
}
dev.off()
}
```

F.7 Funció GOAnalysis()

```
########################
#ANALISI D'ENRIQUIMENT#
########################

GOAnalysis<-function(){

TaulaGens<-as.data.frame (TaulaGens)
tot<-table(clusterTotal)
```

```
tableOrd<-sort(tot, decreasing=T)

clusters<-rownames(tableOrd)

TaulaGens$clusterTotal<- as.character(TaulaGens$clusterTotal)
TaulaGens$V1<- as.character(TaulaGens$V1)

#Creació de la llista amb els ids dels gens dels 14 clusters a escollir
gensInCluster <-list()
j<-1
for (i in clusters[2:15]){
  gensInCluster[[j]]<-TaulaGens$V1[TaulaGens$clusterTotal==i]
  names(gensInCluster)[j]<- i
  j<-j+1
}

#Definició dels paràmetres de la funció
anotPackage<- "org.Sc.sgd.db"
geneUniverse <-TaulaGens$V1

#Funció que fa l'EA
for (i in 1:14){
   numCluster<-names(gensInCluster)[i]
   for (onto in c("BP")){ # o c("BP", "MF", "CC")
      geneIds <- gensInCluster[[i]]
      params <- new("GOHyperGParams", geneIds=geneIds,
              universeGeneIds= geneUniverse, annotation=anotPackage,
              ontology=onto, pvalueCutoff=0.01,
              conditional=TRUE, testDirection="over")
      hgResult <- hyperGTest(params)
      htmlReport(hgResult, file = paste(numCluster,onto,".html", sep="."))
   }
}

}
```