

基因晶片簡介與分析及應用軟體介紹(下)

國家高速網路與計算中心 葉昌偉¹ 謝昌煥²

¹國家高速網路與計算中心 科學計算組 專案助理研究員, c00cwy00@nchc.org.tw

²國家高速網路與計算中心 科學計算組組長 副研究員, c00chh00@nchc.org.tw

摘要

基因晶片(gene chip)可以同時針對生物體內數以千計的基因進行表現量分析，對於生命科學研究者而言，不論是細胞生命週期、生化調控路徑等基礎研究，還是藥物研發中對於藥物作用位置的篩選，到臨床的疾病診斷預測等方向，都是基因晶片可以應用的範圍。基因晶片的分析是相當複雜的程序，由於分析的資料量龐大，需要應用到生物統計及生物資訊相關軟體的協助，本文將會針對基因晶片與其資料分析方法，以及相關軟體與國網中心提供之服務做一個基礎的介紹。

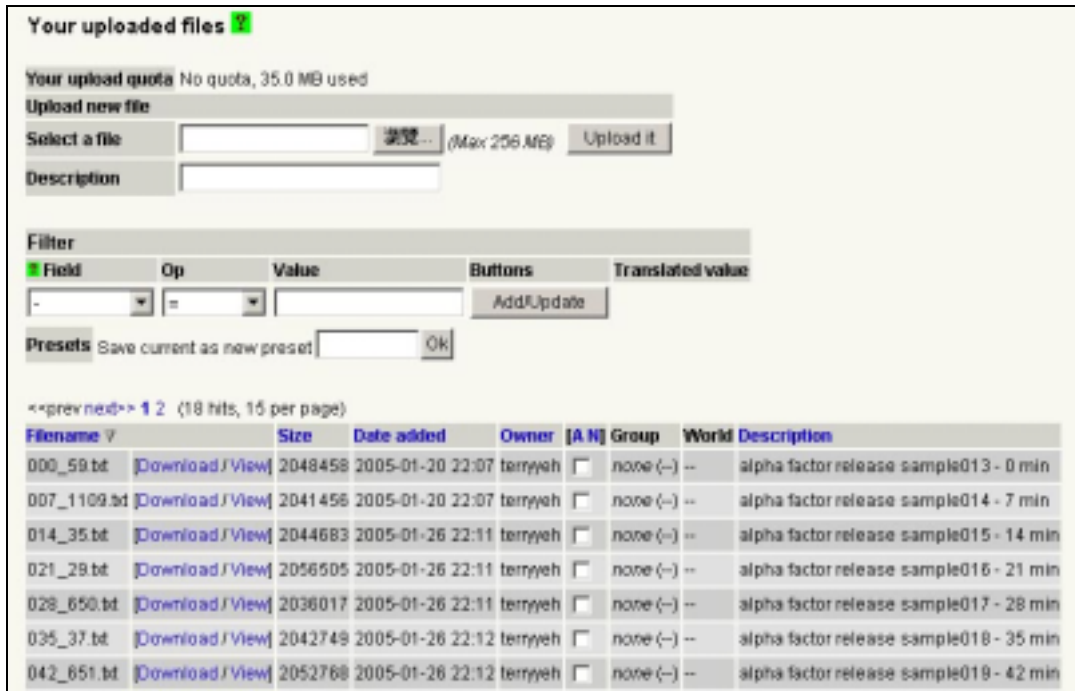
基因晶片分析實例^[1]

由於基因晶片分析方法會依據目的不同而有不一樣的分析策略，本文即以Spellman et al.於1998年所發表關於酵母菌(*Saccharomyces cerevisiae*)細胞週期研究^[2]，所採用之微陣列資料中以alpha-factor為細胞同步因子的資料作為範例，由美國Stanford大學微陣列資料庫取得原始數據資料^[3]，並以國網中心所提供之BASE系統做為分析工具，進行初步的BASE系統使用與基因晶片分析介紹。

在進入BASE系統前，需由國網中心生物知識庫與生物資訊系統網頁(網址為<http://bioinfo.nchc.org.tw/>)申請BASE基因晶片資料庫分析系統之帳號密碼方可使用，輸入帳號密碼便可進入BASE系統。

1. 上傳檔案：

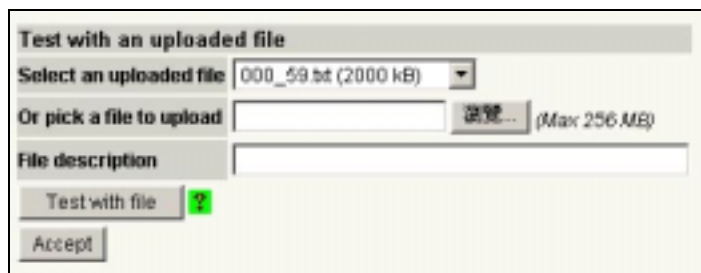
在進行分析前，需利用內建之Uploads將微陣列實驗相關檔案上傳至系統中：點選左方表單中Uploads/My files，則會出現Your uploads files網頁(如圖七)，在此頁可將檔案上傳至系統，也可以設定權限分享給其他使用者，點選”瀏覽...”並且選擇您電腦內的檔案，如果有需要可以加上註解於Description欄位，按下”Upload it”按鈕即可上傳。



圖七 檔案上傳網頁

2. 原始數據結果檔案格式設定

接下來要做的就是讓BASE系統可以辨識你所上傳結果檔案的檔案格式：點選左方表單之Hybridizations/Result file formats，再點選”Add result file format”連結，即可進入New result file format網頁，將網頁下拉至Test with an uploaded file(如圖八)，選擇您先前上傳的檔案，並且按下”Test with file”按鈕。



圖八 按下Test按鈕可測試原始數據結果檔案

點選後，下方會出現上傳檔案的檔案格式(如圖九)，使用者需選擇可辨識檔案的檔案識別列(format identifier row)，點選Use as的下拉式表單，選出Fmt identifier，並選擇上傳檔案的資料開頭列(data header row)，於下拉式表單選出Data header。

Line	Columns	Use as	Spreadsheet				
1	75	-	!ExpId=59				
2	75	-	!Experiment Name=alp				
3	75	-	!Organism=Saccharomy				
4	75	-	!Category=Cell-cycle				
5	75	-	!Subcategory=Alpha f				
6	75	-	!Channel 1 Descripti				
7	75	-	!Channel 2 Descripti				
8	75	-	!Scanning Software=8				
9	75	-	!Software version=2.				
10	75	-	!Scanning parameters				
11	75	-	SPOT	NAME	GENE NAME	CHROMOSOME	STRAI
12	75	-	1	GENOMIC 450NG/UL			
13	75	-	2	GENOMIC 450NG/UL			
14	75	-	3	GENOMIC 50NG/UL			
15	75	-	4	GENOMIC 50NG/UL			
16	75	-	5	POLY A			

圖九 使用者需設定Fmt identifier，以及Data header，做為檔案辨識的依據。

將上述兩項都點選完後，將網頁上拉即可見到檔案資訊自動顯示在 Line number、...must match regexp以及Data header regexp處，而Fields-columns mapping的Key欄位底下可見所有資料開頭列的所有欄位(如圖十)，接下來便要選擇左方Database field與檔案開頭欄位的對應資料，以左方Database field的欄位為主，在Add file column下拉選單中選取檔案開頭欄位的對應資料即可。

Fields-columns mapping			
Refer to columns by name <input type="checkbox"/>			
Database field	Import file columns	Add file column	Key
Array block	65	-	1. SPOT 2. NAME 3. GENE NAME 4. CHROMOSOME 5. STRAND 6. Beginning Coordinate 7. Ending Coordinate 8. PROCESS 9. FUNCTION 10. DS_GENE_1 11. DS_GENE_2 12. SUID 13. CH1I_MEAN 14. CH1D_MEDIAN 15. CH1I_MEDIAN 16. CH1_PER_SAT 17. CH1I_SD 18. CH1B_MEAN 19. CH1B_MEDIAN 20. CH1B_SD 21. CH1D_MEAN 22. CH2I_MEAN
Column in block	67	-	
Row in block	66	-	
Physical x coord.		-	
Physical y coord.		-	
Spot size		-	
Ch1 FG median		-	
Ch1 FG mean	13	-	
Ch1 FG stddev		-	
Ch1 BG median	19	-	
Ch1 BG mean	18	-	
Ch1 BG stddev		-	
Ch1 %pix w/ FG > BG+1SD	43	-	
Ch1 %pix w/ FG > BG+2SD		-	
Ch1 %pix saturated		-	
Ch2 FG median		-	
Ch2 FG mean	22	-	

圖十 使用者需設定資料開頭列與資料庫欄位對應位置

選擇完所有對應的欄位後，並給予這一個檔案格式一個名稱，隨後按下Accept按鈕，即完成結果檔案格式設定。

3. 設定Biomaterials

Biomaterials包含有sample、extracts、labeled extracts三個主要部分，主要為實驗材料的設定，主要的流程如下，先在BASE中設定samples，而sample即可取出作為extracts，並且進行標幟label，標幟完的extracts即可進行雜交實驗，所以在BASE中依照上述順序分碟設定samples、extracts以及labeled extracts三個表單。在設定三個表單前，可以先設定sample origins，增加新的實驗物種生物，以及設定labels，增加新的標幟方法，如Cy3或Cy5。

4. 新增雜交實驗

新增設定完Biomaterials後，即可新增雜交實驗：點選左方Hybridizations/Hybridization，並點選“Add hybridization”連結，即進入New hybridization網頁新增雜交實驗(如圖十一)。

圖十一 由此頁新增雜交實驗資料

先替新增的雜交實驗命名，並且由下拉式選單選擇兩個不同channel的labeled extract，並且填入進行雜交實驗所耗費之labeled extract重量，當labeled extract剩餘重量大於零時，才會顯示在下拉選單中，並且新增雜交實驗後，會由先前的重量扣除此次使用的重量，選擇並填寫完應填的欄位後，即可按下”Accept and add scan”按鈕去新增基因晶片掃描圖片。

此處的圖片資料可以包含任何基因晶片的掃描圖片資料以及所使用掃描器的資訊，可以依照不同channel所掃描的圖片來儲存記錄，也可以存放不同格式如JPEG或TIFF的影像資訊，在新增圖片之前，也要先加入scanner和image processor的資訊。而最重要的一點，在進行分析之前，一定要加入一個scan資料，但不一定要真的加入圖片，只要做過新增scan的動作，表示該雜交實驗已經經過掃描的程序，如此才能新增雜交實驗的原始數據。

5. 新增原始數據

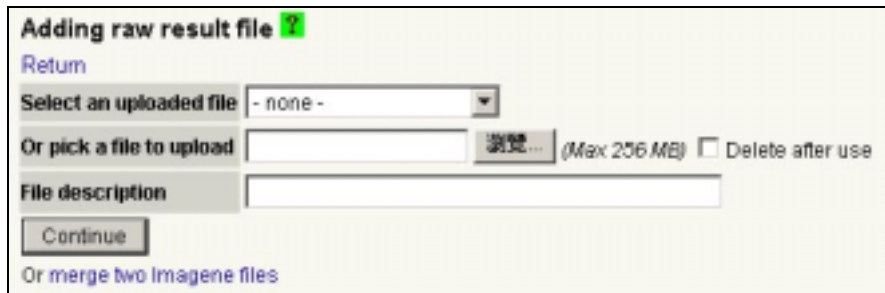
在所有實驗相關資料都建立之後，接下來就要進行原始數據的載入動作，在BASE系統中有兩個地方可以加入原始數據檔，點選左方Hybridizations/Scans連結，再點選先前所新增的Scan名稱，即可進入View scan網頁，在此頁下方有一個”Upload result file [In new window]”連結(圖十二)，或是點選左方Hybridizations/Hybridizations連結，再點選先前所新增的hybridization名稱，進入View hybridization網頁，在Scans表單上，最後有一個Raw data set欄位，其下方即有”new”連結(圖十三)，這兩種方式皆可新增原始數據。



圖十二 點選Upload result file連結新增原始數據檔



圖十三 點選[new]連結新增原始數據檔



圖十四 由此頁選擇原始數據檔以供上傳

當出現Adding raw result file網頁(如圖十四)後，便可以從先前上傳的檔案或是由此處上傳檔案做為原始數據檔的來源，按下”Continue”按鈕，檔案即上傳並且進入下一頁。

Adding raw result file

[Return](#)
[Pick another file](#)

Filename: 000_59.bt

Delete file when done:

Raw data set(RawBioAssay) name: test

Feature extraction software: Scanalyze [\[Add software\]](#)

Feature extraction date (Y-M-D): Today

File format: Yeast alpha factor release sample [\[Define new file format\]](#)

This hybridization is not connected to an array slide, and there are no array designs. Verifying the uploaded file against an array design is therefore impossible. Note that exporting the data in MAQE-ML requires the raw data set to be connected to an array design.

Array design: none [\[And none to choose from\]](#)

Assign positions from: Result file (no features) [\[Choose an array design to change this\]](#)

Reporter ID prefix:

Disallow reporter creation:

圖十五 上傳完原始數據檔，需選擇原始檔案格式及來源。

如圖十五，此處可命名、選取讀取數據的程式，並由下拉式選單選取檔案的格式是屬於何種原始數據檔，選擇完成後再按下“Continue”按鈕，程式便會自動讀取所上傳的原始數據檔。在程式讀取的過程中，網頁會不斷跳動顯示目前讀取的進度，[]內的數字表示目前讀取到原始檔案的列數，如同圖十六箭頭所示，過程中如果有錯誤，也會在箭頭所指處顯示。

Adding raw result file

[Return](#)
[Pick another file](#)

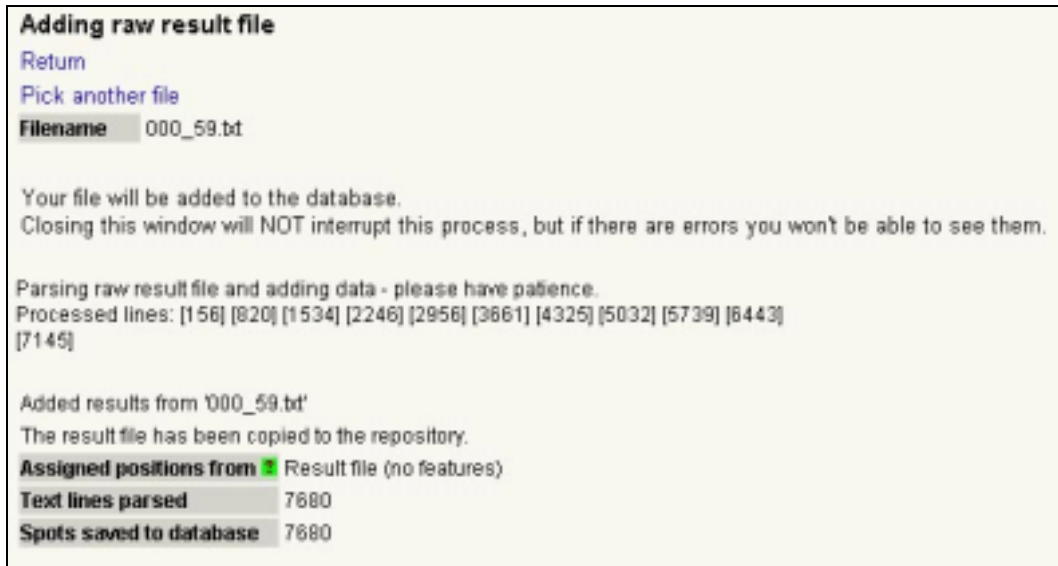
Filename: 000_59.bt

Your file will be added to the database.
 Closing this window will NOT interrupt this process, but if there are errors you won't be able to see them.

Parsing raw result file and adding data - please have patience.
 Processed lines: [156] [820] [1534] [2246] [2956] [3661] [4325] [5032]

圖十六 分析檔案過程中，箭頭所指[]內的數字表示目前讀取到原始檔案的列數。

當檔案完全進入資料庫儲存後，便會顯示如圖十七的資訊，“Added results from 檔案名稱”，並且告知使用者檔案已經複製儲存，以及顯示有多少筆資料存入資料庫當中。



圖十七 經系統分析原始檔案資料後，系統會告知使用者檔案已經複製儲存，以及顯示有多少筆資料存入資料庫

到此已經完成將一筆基因晶片資料載入資料庫的動作，若使用者有多筆資料時，便要重複標題3-5的動作，即可將所有資料載入資料庫中。

6. 設定實驗群組

將所有基因晶片數據載入資料庫後，點選左方Analyze data/Raw data sets連結，即可見所有載入之數據資料，接下來便要設定實驗群組，點選左方Analyze data/Experiments連結，設定實驗相關資料，按下”Accept”按鈕即可建立實驗群組。

實驗群組建立完成後，回到Raw data sets網頁，參閱圖十八，於紅色箭頭處勾選需加入實驗群組的原始資料，並於表格下方藍色箭頭處選取” ‘Add’ marked raw data sets to/from experiment ‘實驗群組名稱’ ”，並按下”Go”按鈕，之後便可以從黑色箭頭處的Experiments欄位看見原始資料被哪一個實驗群組所使用。

Raw data sets

Home Op Make Buttons Translate value

Preview: Save current as new preset [OK]

==prev next== 8/3 (18 kbs, 15 per page)

Raw data set ?	Date added	Owner	[A] Group	World	Samples	Spots	Array design	Pos	Experiments	Functions
alpha factor release sample013	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release, terryye test	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample014	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample015	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample016	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample017	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample018	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample019	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample020	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample021	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample022	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample023	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample024	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample025	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample026	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]
alpha factor release sample027	2005-01-28	terryye	none	-	CHI Alpha-factor block-release, CH2 Alpha-factor block-release	7880	none	No LMS	terryye alpha factor release	[Table] [HTML] [url] [doc] [link] [pdf] [img]

Add marked raw data sets from experiment:

Update marked: Owner: Group: with: access: World: access:

Unkotted / Colored / All

圖十八 將原始數據資料與實驗群組做連結

接下來便要建立BioAssaySet，進行基因晶片分析時使用BioAssays會比使用Raw data sets來的適合，一個BioAssay僅含基因表現量強度值(利用前景強度扣除背景強度即基因表現量)，所以在進行分析時，可以節省硬碟及記憶體空間，加快分析計算的速度。

點選左方Analyze data/Experiments連結，並點選欲分析的實驗群組，便會出現屬於該實驗群組的原始數據表格(圖十九)，於圖十九表格下方黑色箭頭所指之處可選擇計算基因表現強度的方法，當然計算的方式也要依據使用者的原始數據中有的欄位來做選擇，並勾選需加入BioAssaySet的Raw data sets(由於以加入實驗群組的原始數據都要拿來做為分析，所以直接點選圖十九藍色箭頭所指之“A”連結，即可全部選取)，按下“Go”連結即可。接下來網頁便會出現程式處理原始資料的畫面，如圖二十所示。

Experiment terryeh:alpha factor release

Filter

Info Raw data sets Analysis steps Reporter Sets

Raw data set ID	Date added	Owner	Group	Well	Sample	Spots	Assay design	Pos	Function
alpha factor release sample013	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample014	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample015	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample016	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample017	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample018	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample019	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample020	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample021	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample022	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample023	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample024	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample025	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample026	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample027	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample028	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample029	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]
alpha factor release sample030	2005-01-28	terryeh	6006	1-1	Ctrl Alpha factor block-release, C&2 Alpha factor block-release	7080	none	No LMS	[Table] [HTML] [print] [view] [pdf] [log]

Add [view] [raw data sets] [filter] [refresh] [alpha factor release] [Go]

Update filtered: Owner [terryeh] Group [6006] Well [1-1] Spots [7080] Word [none] [Go]

Create new BioAssaySet [alpha] [New] [Raw FO - Median BO] [Go]

圖十九 實驗群組的原始數據表格，在進行分析前一定要設定BioAssaySet。

Experiment terryeh:alpha factor release

Filter

Info Raw data sets Analysis steps Reporter Sets

Processing request - please have patience.
 Processing 'alpha factor release sample013'...
 Processing 'alpha factor release sample014'...
 Processing 'alpha factor release sample015'...
 Processing 'alpha factor release sample016'...
 Processing 'alpha factor release sample017'...
 Processing 'alpha factor release sample018'...
 Processing 'alpha factor release sample019'...
 Processing 'alpha factor release sample020'...
 Processing 'alpha factor release sample021'...
 Processing 'alpha factor release sample022'...
 Processing 'alpha factor release sample023'...
 Processing 'alpha factor release sample024'...
 Processing 'alpha factor release sample025'...
 Processing 'alpha factor release sample026'...
 Processing 'alpha factor release sample027'...
 Processing 'alpha factor release sample028'...
 Processing 'alpha factor release sample029'...
 Processing 'alpha factor release sample030'...
 Counting spots per reporter and position...
 Counting spots, reporter, assays...

BioAssaySet "alpha factor release" [Close]

Edit Delete

Name alpha factor release

Description

Created 2005-03-08 by terryeh

Blockways 10

Created from raw Mean FO - Median BO

Total # of values 136248

Total # of reporters 0072

Extra columns No

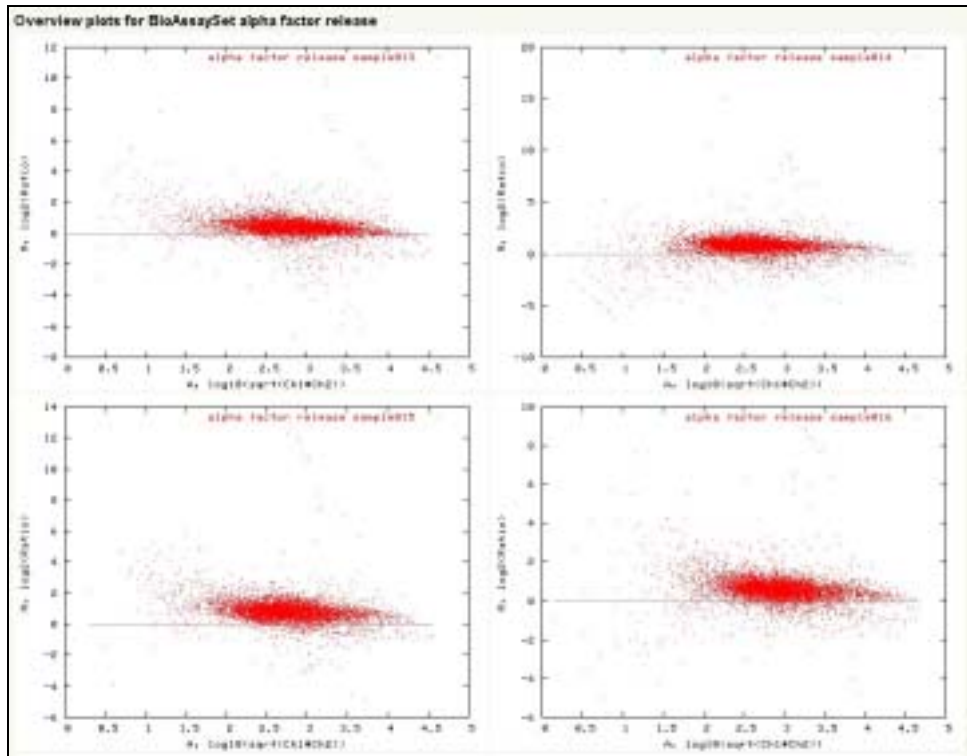
Experiment Explorer

[view] [plot] [HTML] [print]

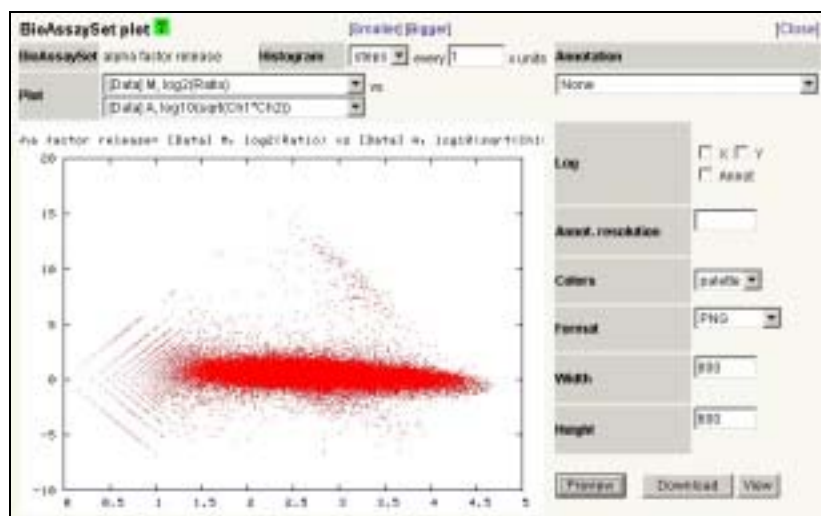
圖二十 BASE系統正在處理同一個實驗群組的資料

7. 基因晶片資料正規化

點選圖二十底下的Overview plot可以瀏覽所有BioAssay的M vs. A分佈圖(其中 $M = \log_2 \frac{R}{G}$, $A = \log_2 \sqrt{R \times G}$)，詳見圖二十一，繪製M vs. A分佈圖的目的主要在判斷基因晶片的資料是否有誤差產生，若有誤差產生時，則必須進行正規化的動作。而點選圖二十底下的HTML plot tool，或是BioAssays表格後面的HTML Plot tool，則會跳出BioAssay Plot的繪製分佈圖工具(圖二十二)，使用者可以依照自我的需求繪製不同的分佈圖。



圖二十一 BioAssay的M vs. A分佈圖



圖二十二 BioAssay Plot的繪製分佈圖工具

整個Analysis steps網頁(圖二十三)共可分為兩大部分，上半部是屬於整個BioAssaySet的資訊部分，下半部是Hierarchical overview of BioAssaySet analyses，當進行分析後，此部份便會依照分析主從架構來呈現分析的結果。

BioAssays					
Name ?	Raw data sets	Parents	Created	Spots [A N]	Functions
alpha factor release sample013	alpha factor release sample013		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample014	alpha factor release sample014		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample015	alpha factor release sample015		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample016	alpha factor release sample016		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample017	alpha factor release sample017		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample018	alpha factor release sample018		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample019	alpha factor release sample019		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample020	alpha factor release sample020		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample021	alpha factor release sample021		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample022	alpha factor release sample022		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample023	alpha factor release sample023		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample024	alpha factor release sample024		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample029	alpha factor release sample029		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]
alpha factor release sample030	alpha factor release sample030		2005-03-09	7680	[Table] [HTML Plot tool] [Java Plot tool]

Filter Run application (on all)

Hierarchical overview of BioAssaySet analyses				
Name	Date	Info	Functions	
alpha factor release	2005-03-09	Mean FG - Median BG, 18 assays, 138240 spots, 8072 reporters	[Table] [HTML Plot tool] [Java Plot tool]	[Delete]

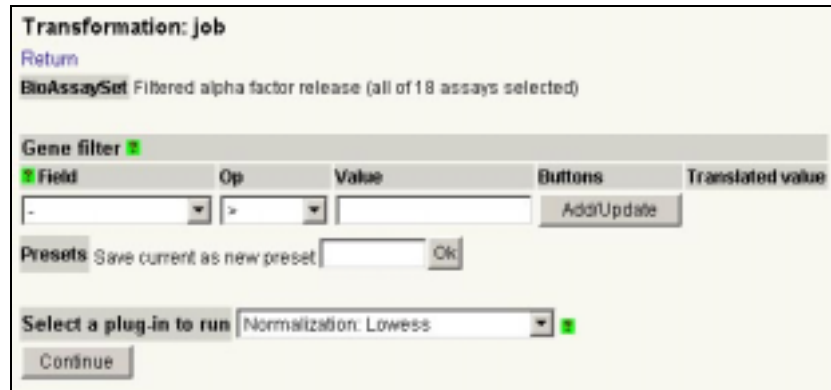
圖二十三 Analysis steps網頁列表分為上下兩部分

在進行正規化之前可以先進行過濾(filter)的動作，剔除掉表現強度為負值的資料或是踢除非必要的參考點，按下”Filter”按鈕，設定欲剔除的標準，按下”Accept”按鈕即可，過濾完後便可在Hierarchical overview看見剛剛所進行的動作(圖二十四)。

Hierarchical overview of BioAssaySet analyses				
Name	Date	Info	Functions	
alpha factor release	2005-03-09	Mean FG - Median BG, 18 assays, 138240 spots, 8072 reporters	[Table] [HTML Plot tool] [Java Plot tool]	[Delete]
Filtering Ratio > 0 and no value gene [F]	2005-03-09	3 spot criteria, 1 reporter criterion	[Delete] [Copy]	[Delete]
Filtered alpha factor release	2005-03-09	18 assays, 113670 spots, 6315 reporters	[Table] [HTML Plot tool] [Java Plot tool]	[Delete]

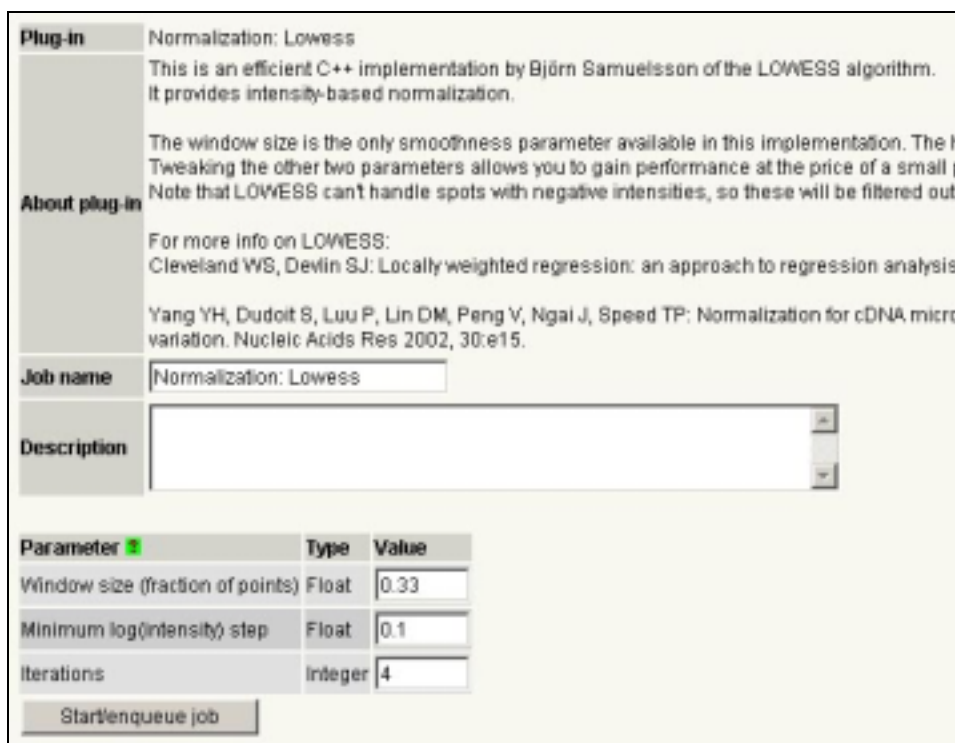
圖二十四 BioAssaySet階層式分析表

點選圖二十四黑色箭頭所指之經過過濾的BioAssaySet連結，進入Analysis steps中BioAssaySet網頁，按下Run application (on all)按鈕或是點選上圖藍色箭頭所指連結，便可以開始選擇分析的程式。



圖二十五 選擇進行分析的應用程式，如Normalization : Lowess

從圖二十五中Select a plug-in to run的下拉式選單選取Normalization: Lowess^[4]，並且按下”Continue”按鈕繼續，則會出現正規化Lowess方法的詳細說明及參數設定頁(見圖二十六)，接下來便可按下”Start/enqueue job”按鈕，送出工作至伺服器計算。



圖二十六 正規化Lowess方法的詳細說明及參數設定頁

送出工作後可點選Check the status of the job連結，檢視目前工作執行狀況，點選圖二十七箭頭處所指[refresh]連結可重新整理工作執行狀況，待執行結束後，即可在Hierarchical overview看見剛剛執行正規化的BioAssaySet。

Transformation 'Normalization: Lowess' [Close]
 Edit Delete

Type Job
Name Normalization: Lowess
Created 2005-03-09 by terryeh
Plug-in Normalization: Lowess
Description
Reporter filter None

Job info

Plug-in Normalization: Lowess
Status Importing data from plugin. [refresh](#) ←
Job ID 26 **Computer** Local **JobController PID** 30355
Submitted at 2005-03-09 22:01:51 **Estimated execution time** 5.00s **CPU time limit** -
Processing started at 2005-03-09 22:01:54 **Plugin execution time** 1.73s **Plugin CPU time** 1.41s
Plugin started at 2005-03-09 22:02:00
Plugin finished at 2005-03-09 22:02:01 **Data matrix size** 239,970 **Export CPU time** 2.49s
Cleanup ended at - **RAM limit** 16.0 MB **Import CPU time** -

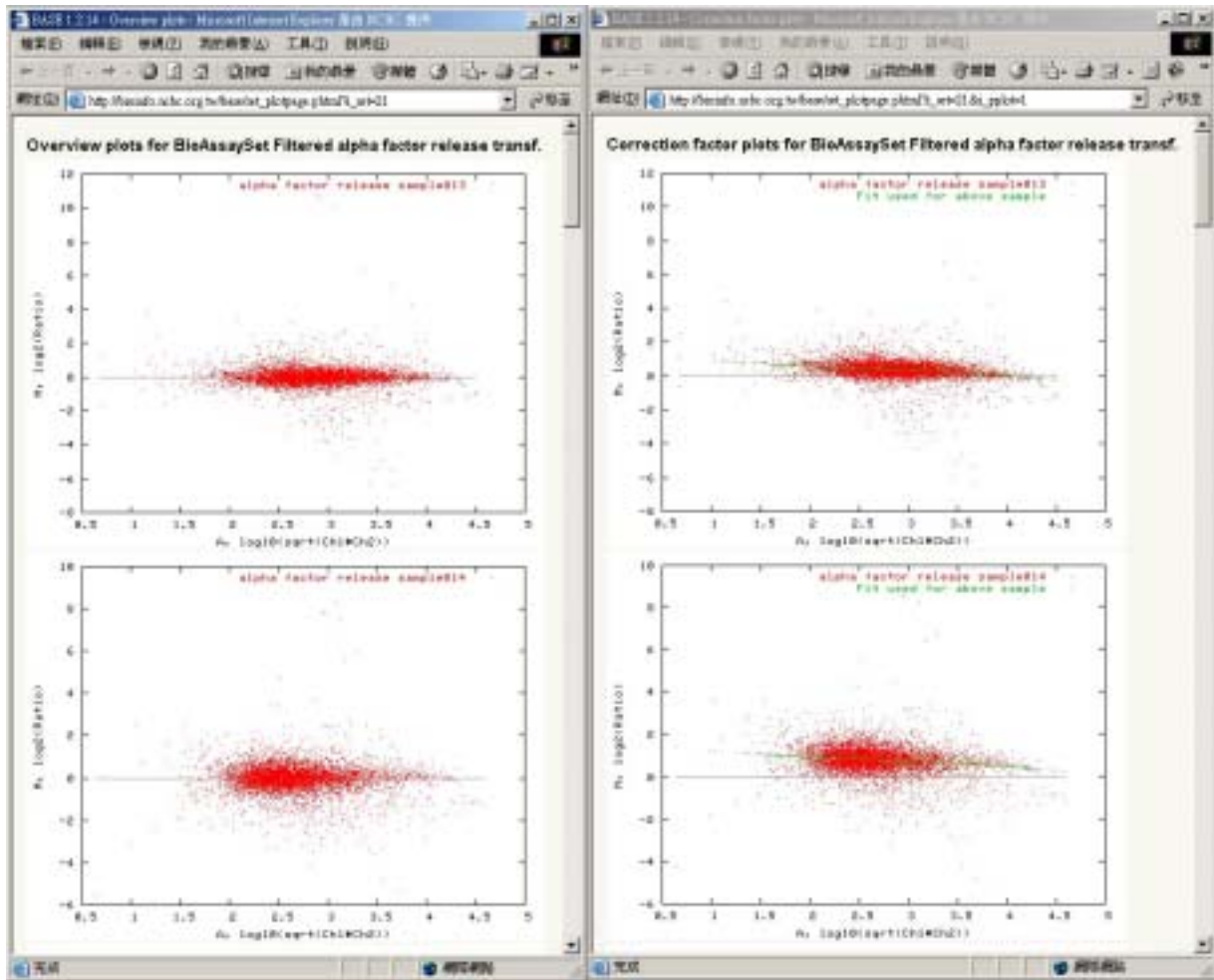
Plug-in output

Parameter	Value
Window size (fraction of points)	0.33
Minimum log(intensity) step	0.1
Iterations	4

Files created by this job
None

圖二十七 目前工作執行狀況

點選進行過正規化之後的BioAssaySet，即可見該BioAssaySet的Analysis steps網頁，再點選中間[Overview plot]，以及[Correction factor plot]連結，會分別跳出網頁顯示正規化前後的M vs. A分佈圖(圖二十八)，在Correction factor plots連結圖中，M vs. A分佈圖中的綠色線條代表的是正規化前數據分佈的趨勢，Lowess法會將這個趨勢拉向M=0的那條線上，於是便產生Overview plots的M vs. A分佈圖。



圖二十八 正規化前後的M vs. A分佈圖，左邊為正規化後，右邊為正規化前。

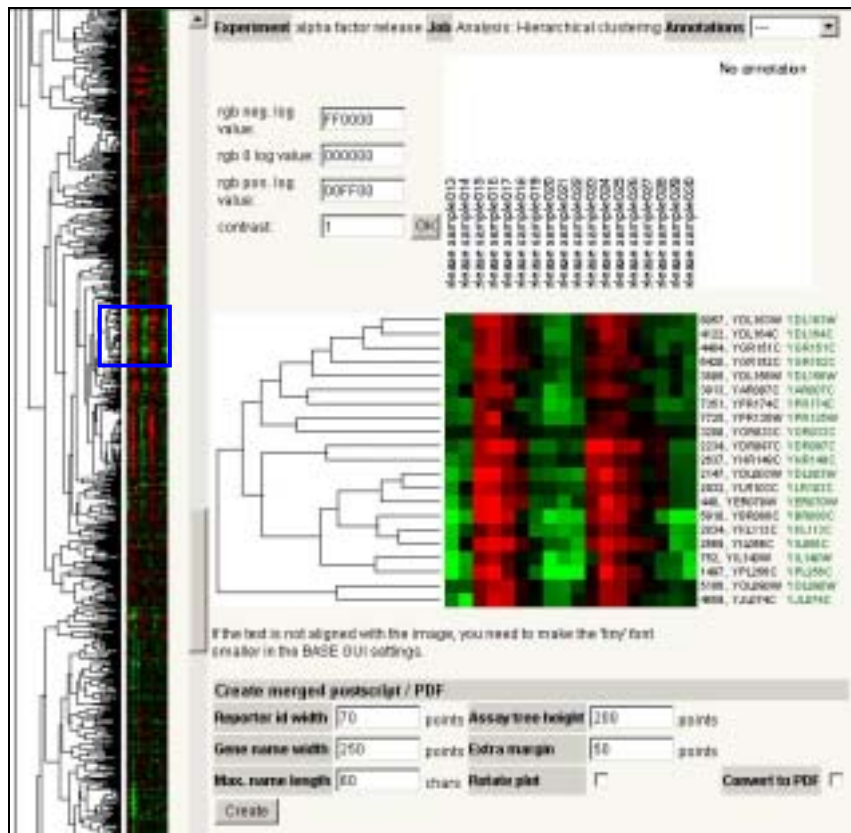
8. 階層式叢集分析法

依照相同的步驟，按下Run application (on all)按鈕，選擇Analysis: Hierarchical clustering，便可以進行階層式叢集分析。以我們選取的酵母菌細胞週期資料做階層式叢集分析時，由於此資料為時間序列資料，故Assay間不需做叢集分析，故參數設定如下，Gene distance metric選定為Pearson correlation (2-r)，以及Assay distance metric選定為No clustering，按下”Start/enqueue job”，即可送出。計算完成後，便會出現如圖二十九之計算結果。

Job info		
Plug-in	Analysis: Hierarchical clustering	
Status	All done (restart)	
Job ID	28	
Submitted at	2005-03-09 23:24:52	
Processing started at	2005-03-09 23:24:52	
Plugin started at	2005-03-09 23:26:07	
Plugin finished at	2005-03-09 23:26:37	
Cleanup ended at	2005-03-09 23:26:37	
Plug-in output	Clustered genes: Distance calculations: 110911376 Clustered assays: Distance calculations: 0	
Parameter	Value	
Gene distance metric	Pearson correlation (2-0)	
Assay distance metric	No clustering	
New cluster point	Center of mass	
Files created by this job		
Name	Size	Functions
assays.bt	72 bytes	New Download Visualize
assays.bt	382 bytes	New Download
assays.ps	6.71 kB	New Download PDF
genenames.bt	85.7 kB	New Download
genes.bt	231 kB	New Download
genes.ps	471 kB	New Download PDF
genes_m.ps	468 kB	New Download PDF
matrix.png	1.02 kB	New Download
stdin.bt	1.008 kB	New Download

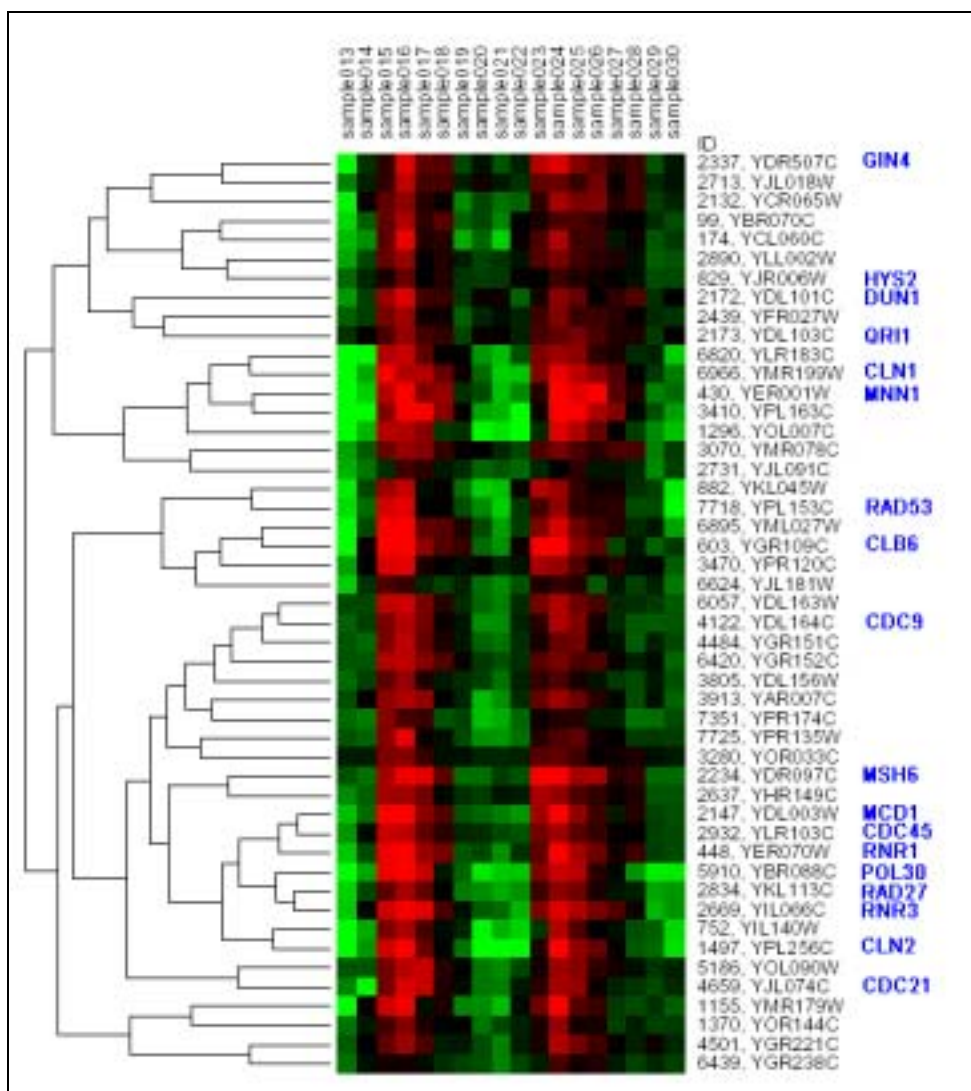
圖二十九 階層式叢集分析法分析結果列表

點選圖二十九箭頭所指之Visualize連結，分析結果則如圖三十以圖形方式呈現。



圖三十 以圖形方式呈現階層式叢集分析法分析結果

利用BASE系統的階層式叢集分析工具對先前提及之酵母菌細胞週期資料^[2]進行分析，僅就文獻中所提及的細胞週期G1階段的相關基因，如CLN1、CLN2、CLB6、RNR1、CDC9、CDC21、CDC45等基因，與此資料來源所發表的文獻比較，經由BASE系統內建的階層式叢集分析工具分析的結果，均可將這些基因分在同一個叢集之中(圖三十一)，所得結果與文獻內分析結果一致，圖三十一中的基因表現叢集，即約等於文獻中之CLN2叢集，由此可以確立此系統分析結果無誤。



圖三十一 經由BASE系統內建的階層式叢集分析工具，可將酵母菌細胞週期G1階段的相關基因分在同一個叢集之中，其位置位於圖三十藍色方框處。

BASE基因晶片資料庫分析系統除了上述的功能之外，尚有其他的功能可供研究人員針對不同目的基因晶片實驗進行分析，亦可以自行開發分析工具放進BASE系統中使用，對於不論是基礎分生實驗人員或是生物資訊工具開發研究人員而言，BASE可以說是一套相當實用的分析工具套件。

結語

由於基因晶片同時可以存在太多的資訊，非一般人力可以輕易分析解決，所以需要利用電腦來輔助人腦計算並分析，而基因晶片可以利用來從事研究的目的與方向非常廣泛，分析方法也日新月異的不斷出現，國網中心一向秉持著服務導向的精神替國內學術界服務，故提供基因晶片分析相關軟體介紹與軟體使用服務，並且目前正計劃發展一套能夠幫助基礎生命科學研究人員的基因晶片分析問題導向建議系統，建議研究人員要進行基因晶片分析時，可以採取的策略與分析工具，以期最終能推動國內生命科學領域進步與發展。

參考資料

1. Saal, L.H., Troein, C., and Vallon-Christersson, J. (2003) BASE v1.2 User Guide.
2. Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 9, 3273-97.
3. <http://genome-www5.stanford.edu/>
4. Workman, C., Jensen, L.J., Jarmer, H., Berka, R., Gautier, L., Nielser, H.B., Saxild, H.H., Nielsen, C., Brunak, S., and Knudsen, S. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* 3, research0048.