

Introduction to R

Outline

R language...

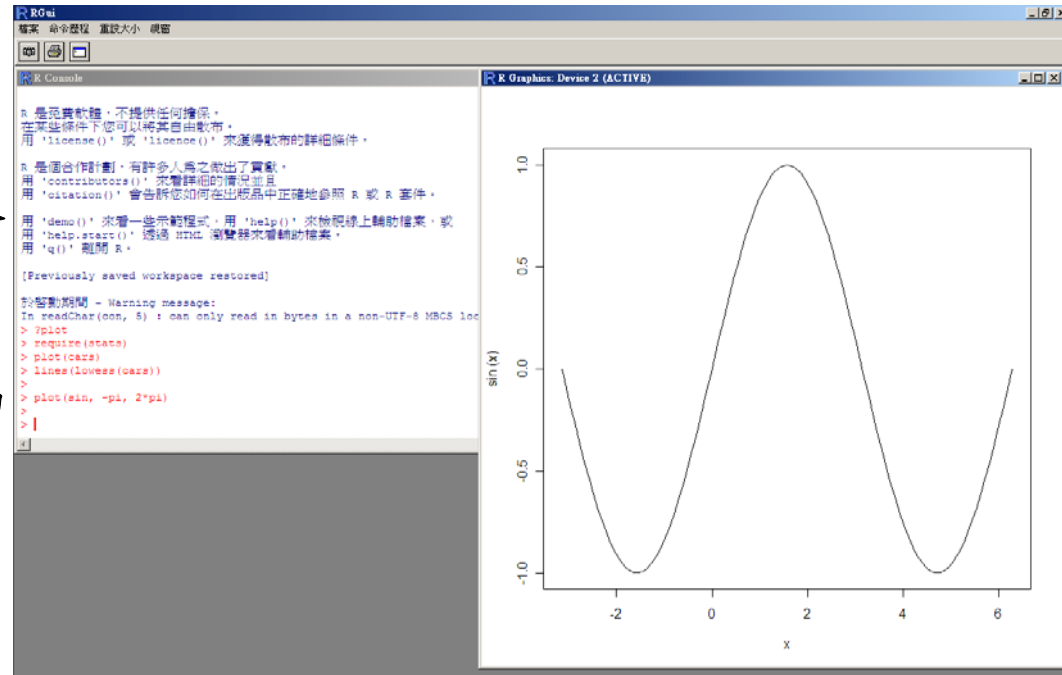
Useful function for analysis

What is R?

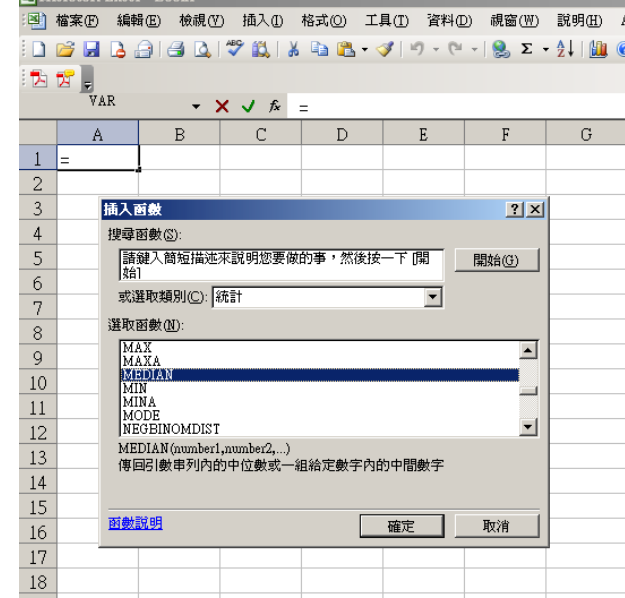
1. 並非專用統計軟體，而是可用來執行統計分析的環境

2. 基本上已經預載了常用的統計函數(sd, mean, t-test)及繪圖功能(boxplot, histogram)，如需使用較特殊的統計方法或是繪圖功能則可以再安裝特殊的package

3. 因開放原始碼及完全免費，所以許多使用者會release自己開發的R code及package，資源取得相當容易

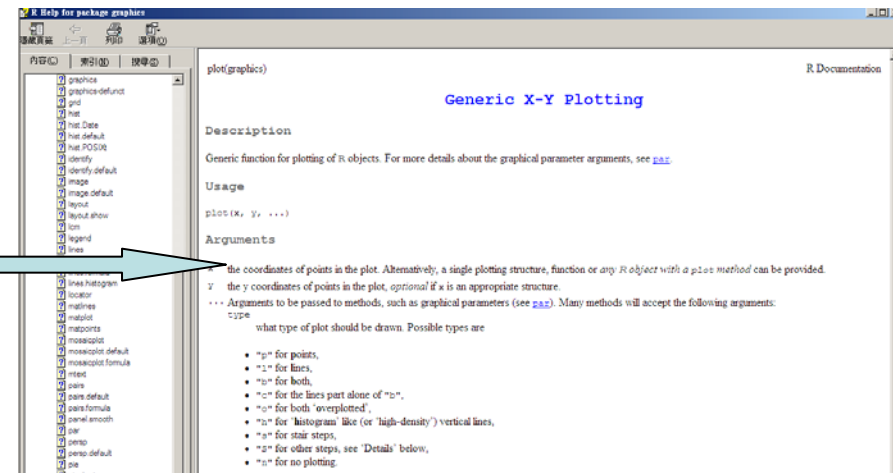


Excel

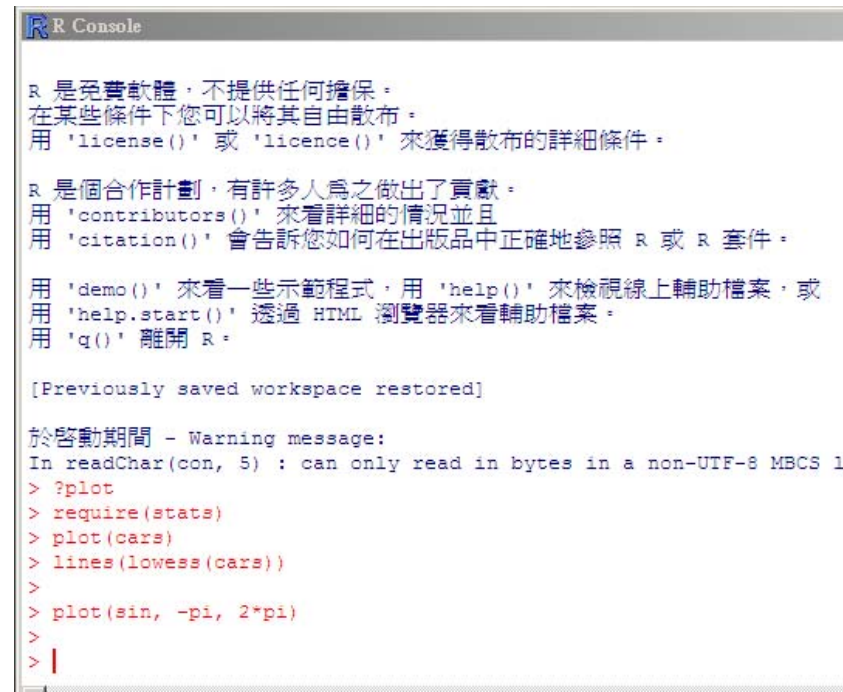


R的優缺點？

- 優點：
 - 免費軟體
 - 完善的說明文件與討論區
 - 程式容易根據使用者需求做修改



- 缺點：
 - 並無容易使用的GUI
 - 需詳知函式名稱與程式編寫邏輯
 - 不同版本的package不見得相容



R的基本運算功能

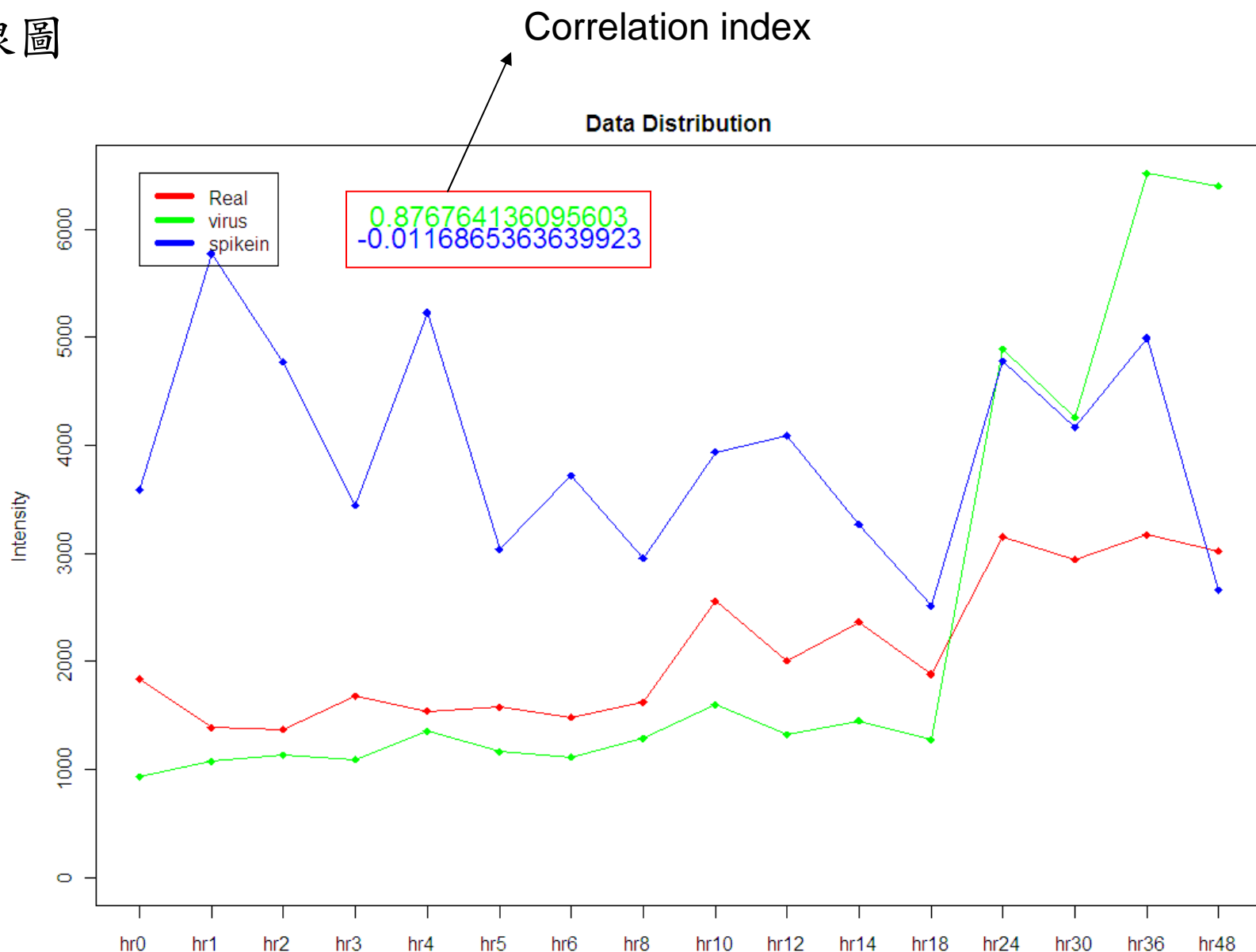
1. 簡單下個指令就可算出每一組數值的sd, mean medium 等等一般的運算
2. 設定觀測值的條件，快速篩選出條件內gene
3. 寫成function後，以後要做同一件事情，只要call function即可
4. 在字串上的處理也相當容易，可以簡單地match不同gene list上同樣具有的gene，也可以直接就gene name上去篩選probe

R 資料編輯器									
	site_index	hr10_F	hr10_B	hr0_F	hr0_B	hr12_F			
1	1	3352.983	58.88606	1933.845	76.30885	2642.689			
2	2	63.2931	58.88105	89.55172	75.45968	58.74627			
3	3	68.66667	58.81323	104.9	75.15175	56.16667			
4	4	57.39683	58.56445	85.66667	74.76758	56.52542			
5	5	63.01587	59.67131	94.2381	76.99203	65.05172			
6	6	63.96429	59.55289	93.60714	78.06986	57.40678			
7	7	59.58929	58.90982	86.625	76.81563	61.56923			
8	8	59.81034	57.9498	84.25862	74.67068	59.21053			
9	9	64.85965	59.38416	96.5614	75.40396	60.57627			
10	10	61.98182	58.2835	84.03636	73.63107	59.58182			
11	11	64.54386	58.17255	89.49123	72.38824	58.56667			
12	12	14	hr10	hr0	hr12	hr1	hr14	hr2	hr18
13	13	12	12001.115	7443.458	9933.370	6478.857	10990.304	6034.872	8591.645
			hr3	hr24	hr4	hr30	hr5	hr36	hr6
			6939.917	14138.864	6961.937	12329.928	6778.253	13460.078	7353.683
			hr48	hr8					
			13133.464	7647.341					

WSSV703	SA_TUS_3736
SA_TUS_4589	SA_TUS_4136
SA_TUC_1720	SA_TUS_1129
SA_TUC_1233	SA_TUS_3518
SA_TUS_4073	SA_TUC_1256
SA_TUC_40	SA_TUS_2141
SA_TUS_1408	WSSV703
SA_TUS_4871	SA_TUS_4589
SA_TUS_4424	SA_TUC_1720
WSSV667	SA_TUC_1233
SA_TUS_3853	SA_TUS_4073
SA_TUS_4044	SA_TUC_40
SA_TUS_860	SA_TUS_1408
SA_TUS_104	SA_TUS_4871
	SA_TUS_4424
	WSSV667

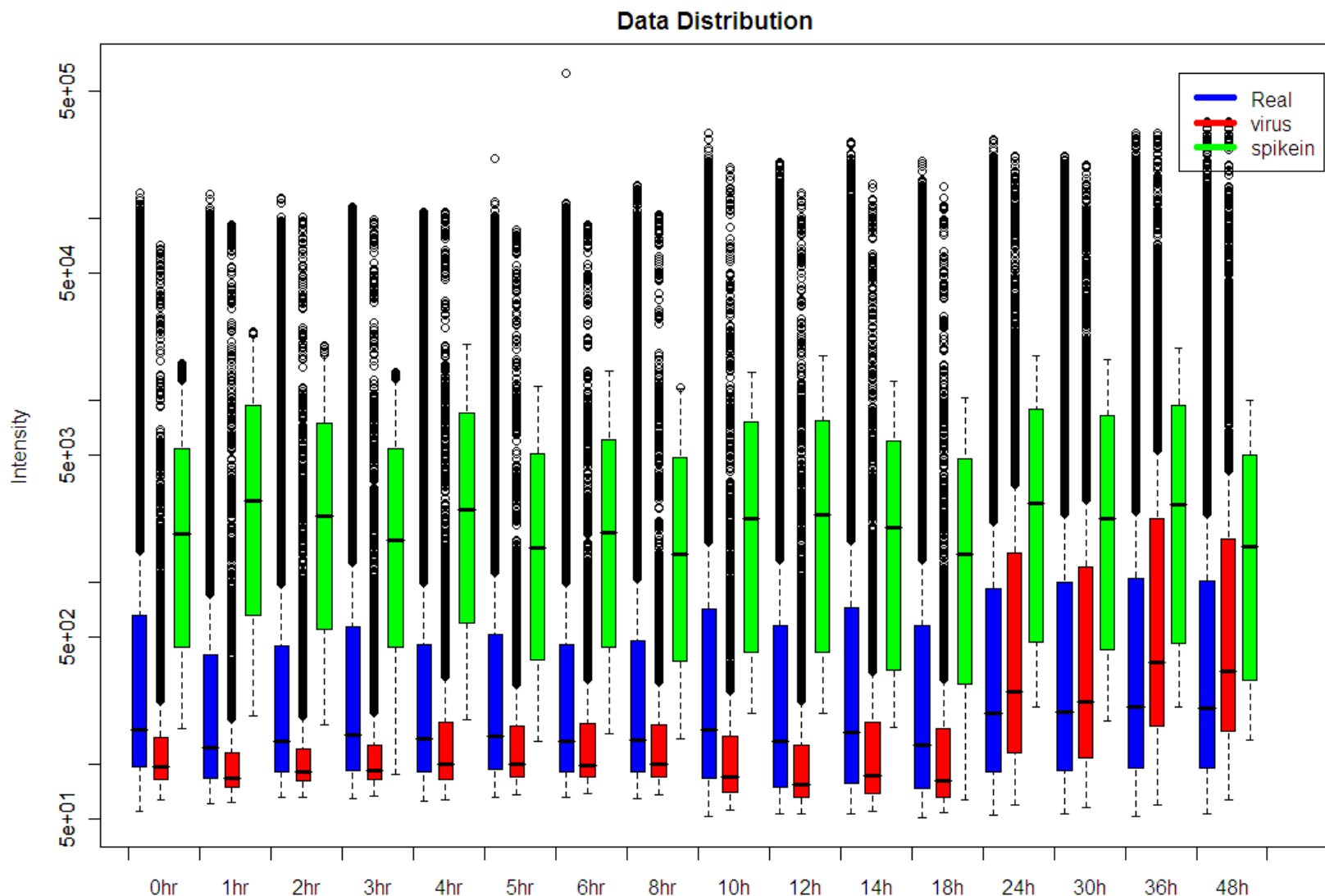
R的基本繪圖功能

折線圖



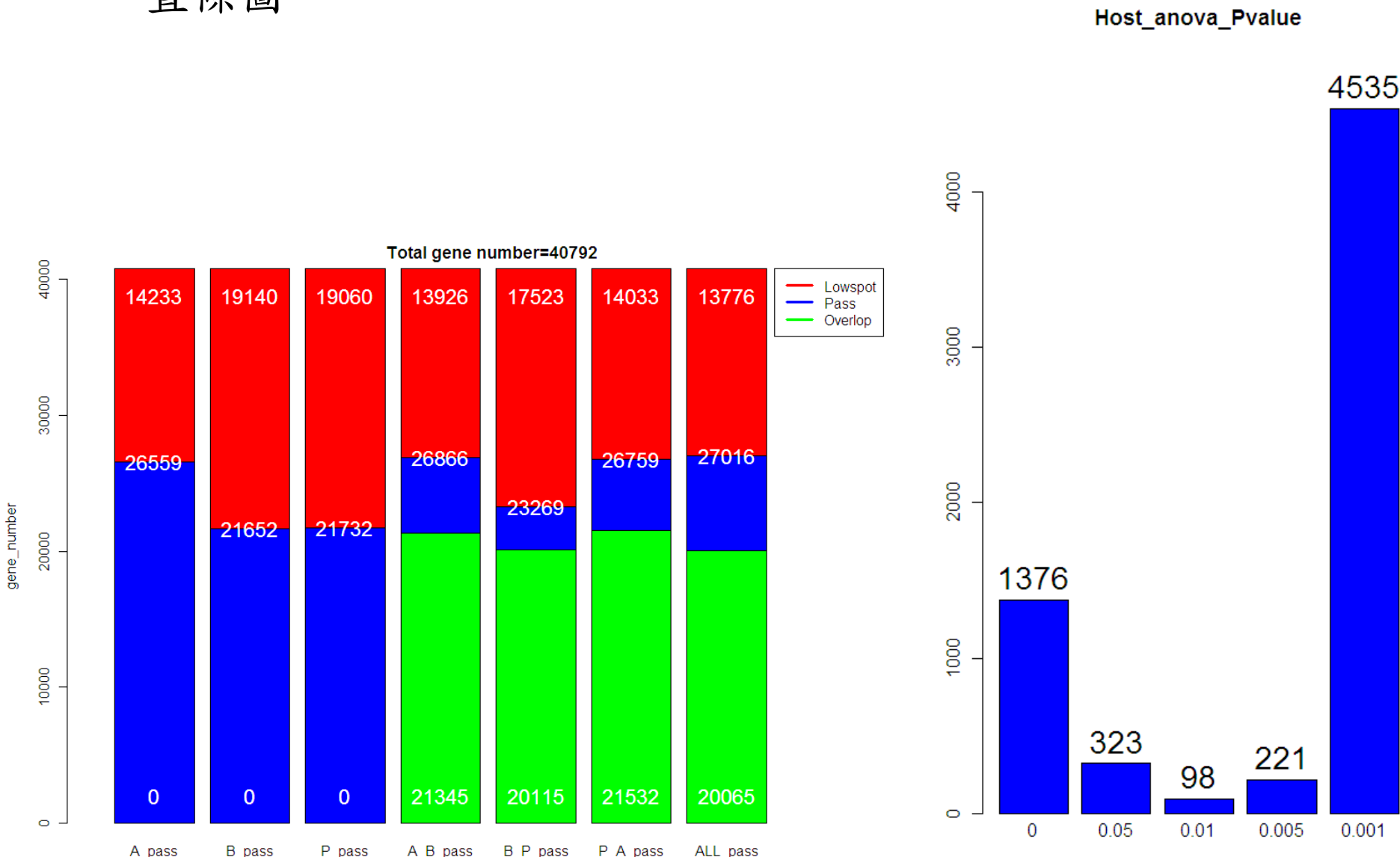
R的基本繪圖功能

boxplot



R的基本繪圖功能

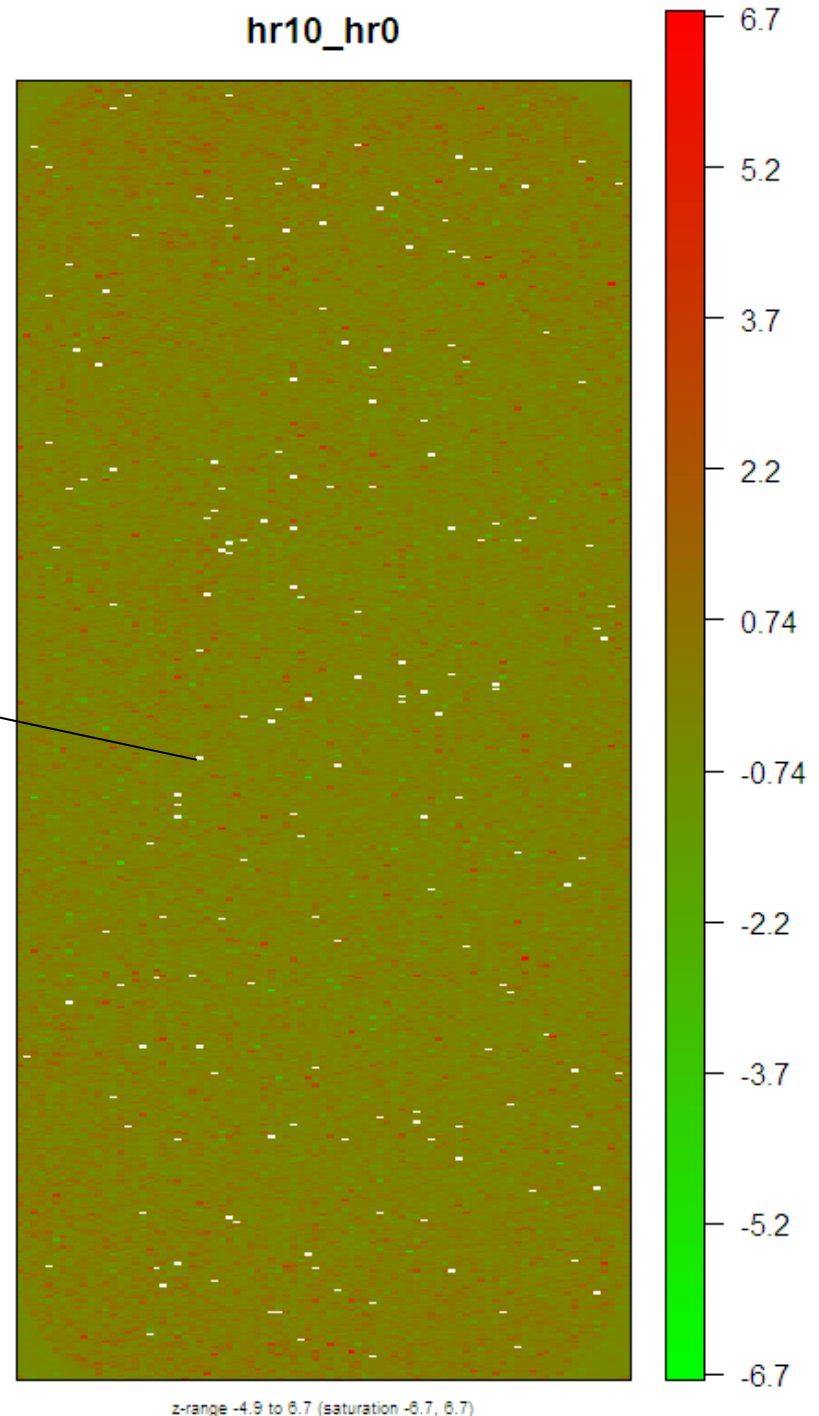
直條圖



*R*的基本繪圖功能

Image plot

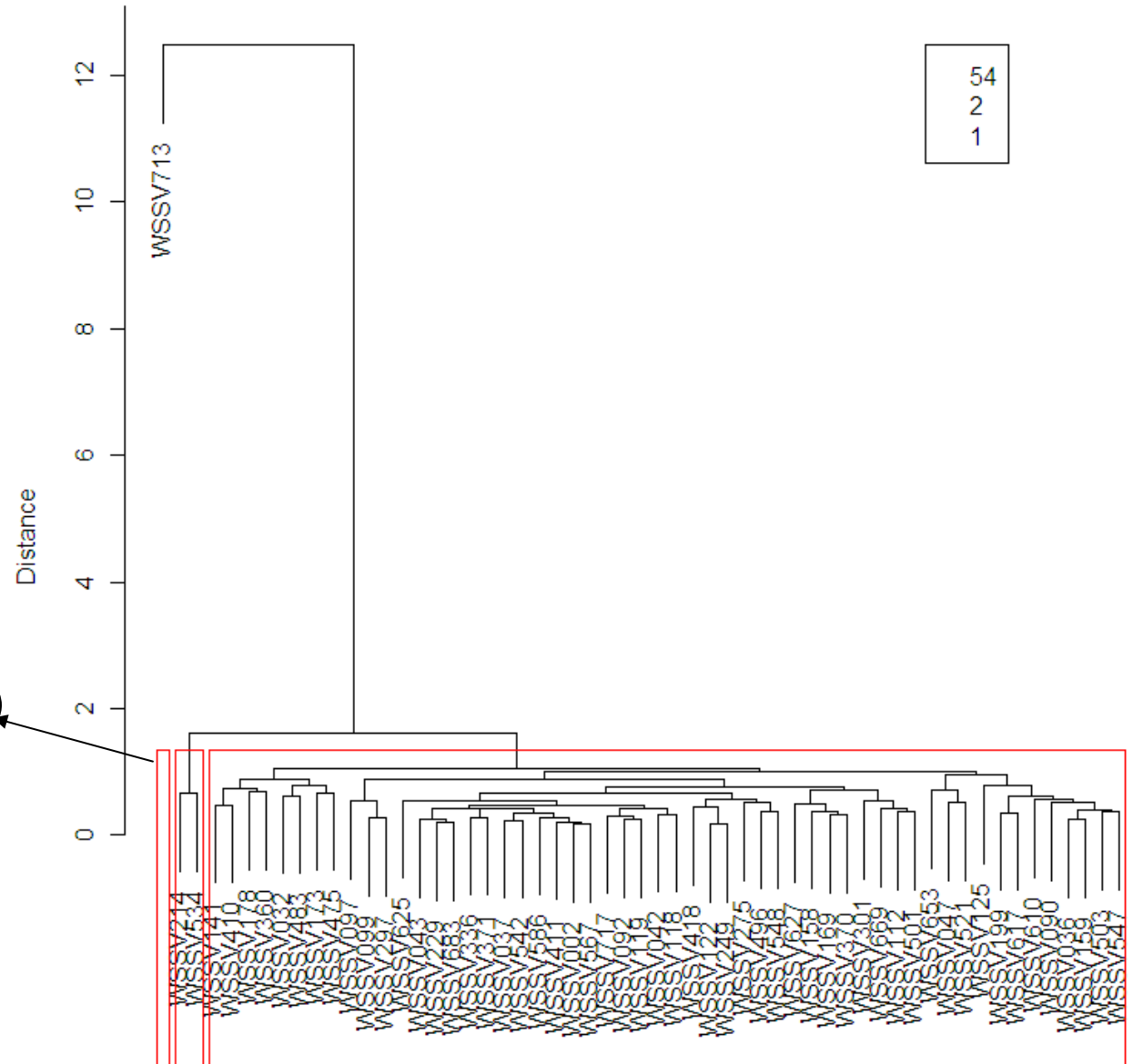
無設計probe在這個點



R的基本繪圖功能

Cluster plot

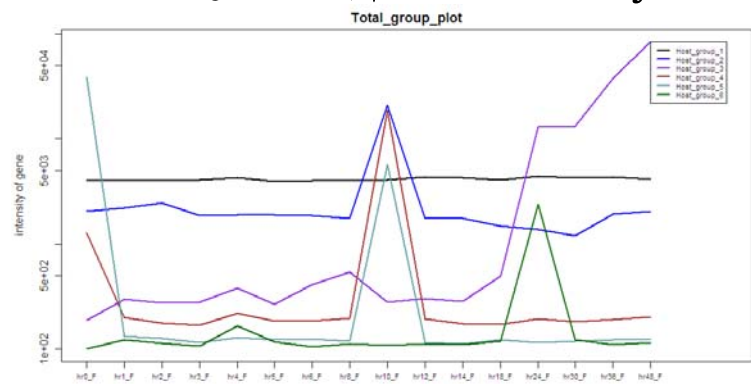
Cluster_Virus_05_trend



Microarray format in R

R is a good tool to process microarray data

- 1. 已有許多R package (limma, marray, affy)可以建立與不同格式的microarray 資料” 相容的資料格式”
- 2. 運用R強大的資料繪圖及統計環境，可以快速的分析microarray data



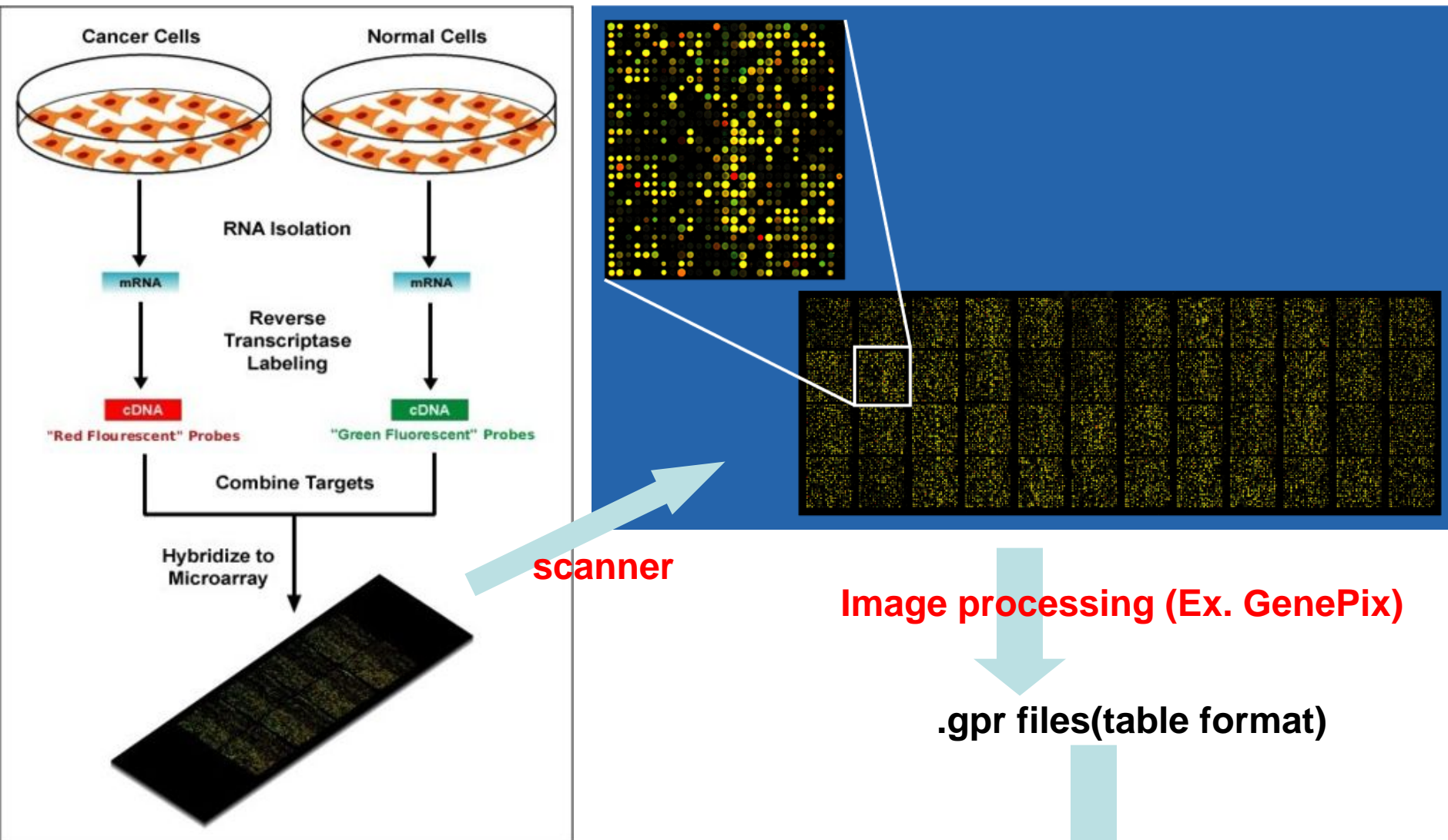
- 3. 與資料庫連結容易，就可以解決大量資料的存放問題，並可有效率的調出想要處理的資料作進一步分析

顯示: 30 筆記錄, 開始列數: 30 > >> 頁碼: 1

顯示為 水平 方式 及 每隔 100 行顯示欄名

←→	site_index	hr10_F	hr10_B	hr0_F	hr0_B	hr12_F	hr12_B	hr1_F	hr1_B	hr14_F	hr14_B	hr2_F
<input type="checkbox"/>	1	3352.983	58.88606	1933.845	76.30885	2642.689	58.57122	1429.033	74.12041	3150.742	61.28612	1750.919
<input type="checkbox"/>	2	63.2931	58.88105	89.55172	75.45968	58.74627	57.83101	79.85075	70.89861	59.1129	58.31144	83.54839
<input type="checkbox"/>	3	68.66667	58.81323	104.9	75.15175	56.16667	57.41406	76.78333	70.50977	59.72131	58.0932	87.16393
<input type="checkbox"/>	4	57.39683	58.56445	85.66667	74.76758	56.52542	58.30237	77.23729	71.66798	57.5625	58.10039	81.10938
<input type="checkbox"/>	5	63.01587	59.67131	94.2381	76.99203	65.05172	58.16567	112.5345	72.07585	60.67797	58.03731	85.61017
<input type="checkbox"/>	6	63.96429	59.55289	93.60714	78.06986	57.40678	57.29528	74.79661	71.84843	57.12698	57.45472	85.15873
<input type="checkbox"/>	7	59.58929	58.90982	86.625	76.81563	61.56923	57.43028	83.87692	69.27291	56.25	56.47348	73.10938
<input type="checkbox"/>	8	59.81034	57.9498	84.25862	74.67068	59.21053	57.26062	78.89474	70.80309	60.41667	57.12891	78.91667

Two-color spotted microarray



Two data in one spot.

Red channel (Cy5) from “Experimental set”

Green channel (Cy3) from “Control set”

Probes from cDNA (complementary) that be amplified by PCR

Microarray data format

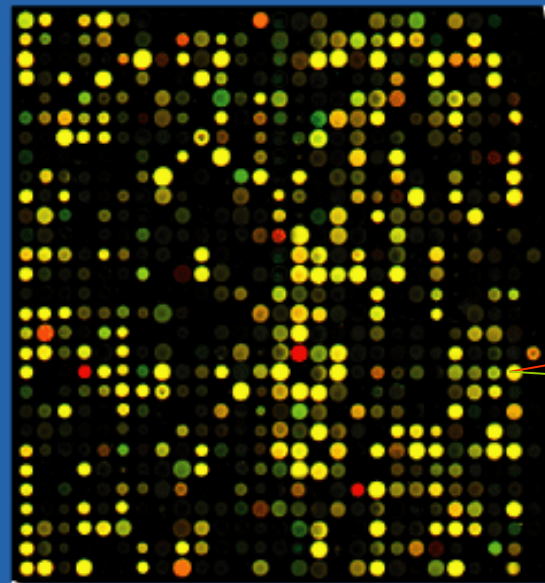
Probe在microarray上的位置

Control點或是實驗點

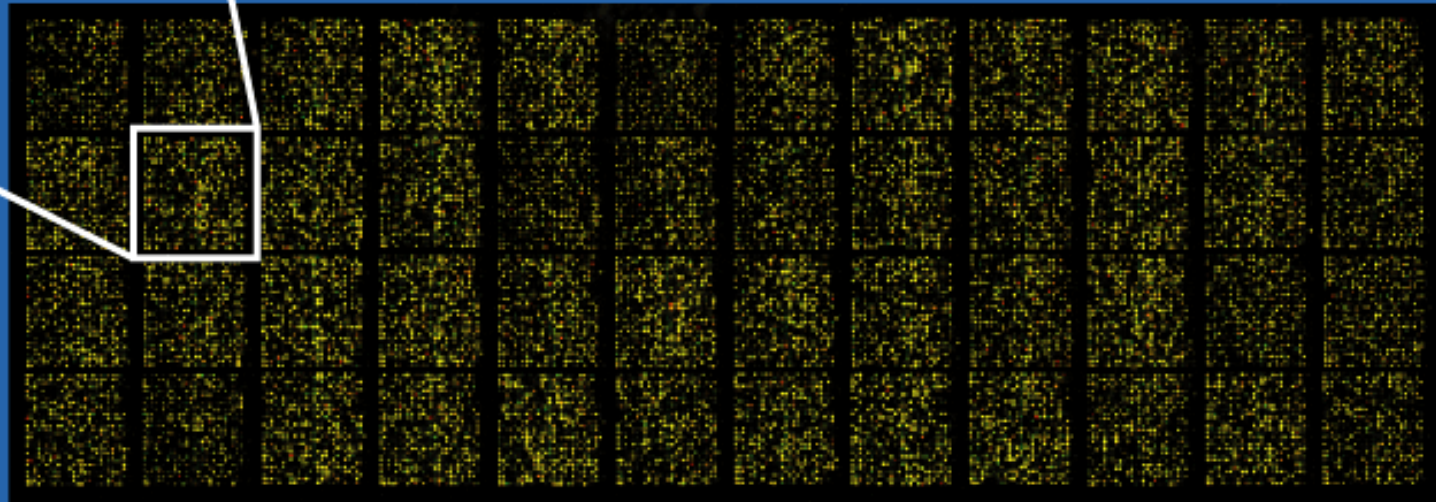
紅光與綠光分別的intensity數值

	A	B	C	D	E	F	G	H	I
9	integer	integer	integer	integer	text	float	float	float	float
10	FeatureNur	Row	Col	ControlType	ProbeName	gMeanSignal	rMeanSignal	gMedianSignal	rMedianSignal
11	1	1	1	1	GE_BrightCorner	1.93E+03	3.35E+03	1936	3341
12	2	1	2	1	DarkCorner	8.96E+01	6.33E+01	87	61.5
13	3	1	3	1	DarkCorner	1.05E+02	6.87E+01	96.5	68
14	4	1	4	1	DarkCorner	8.57E+01	5.74E+01	83	57
15	5	1	5	1	DarkCorner	9.42E+01	6.30E+01	91	62
16	6	1	6	1	DarkCorner	9.36E+01	6.40E+01	92.5	63.5
17	7	1	7	1	DarkCorner	8.66E+01	5.96E+01	82.5	60
18	8	1	8	1	DarkCorner	8.43E+01	5.98E+01	84	60
19	9	1	9	1	DarkCorner	9.66E+01	6.49E+01	95	62
20	10	1	10	1	DarkCorner	8.40E+01	6.20E+01	85	62
21	11	1	11	1	DarkCorner	8.95E+01	6.45E+01	89	64
22	12	1	12	0	SA_TUS_3736	1.69E+02	1.47E+02	167	149
23	13	1	13	0	SA_TUS_4136	1.69E+02	1.21E+02	169	120.5
24	14	1	14	0	SA_TUS_1129	1.12E+02	9.88E+01	109	97
25	15	1	15	0	SA_TUS_3518	1.08E+02	7.70E+01	106	76
26	16	1	16	0	SA_TUC_1256	1.06E+02	9.25E+01	103	89
27	17	1	17	0	SA_TUS_2141	7.68E+02	8.07E+02	721	763.5
28	18	1	18	0	WSSV703	8.23E+01	6.55E+01	83	63
29	19	1	19	0	SA_TUS_4589	1.12E+02	9.30E+01	108	91
30	20	1	20	0	SA_TUC_1720	1.75E+02	1.72E+02	173	166
31	21	1	21	0	SA_TUC_1233	1.80E+02	9.76E+01	181	96

How to get gene intensity?



→ Red light channel: Mean and Median
→ Green light channel: Mean and Median



WSSV Data in microarray

Total gene: 7325

Gene name type	Repeat times	Amount
SA_	6	6508
SO_	6	45
WSSV001~718	6	698
Spike in (E1A_r60)	32	10
Total	45049	7325

Host gene
Host gene
Virus gene
Control gene

Data processing in R

Raw data files: gpr (GenePix); .spot (Spot), .xls (SMD), .txt (Agilent)

Read raw data to R: Table format data

R 資料編輯器						
	site_index	hr10_F	hr10_B	hr0_F	hr0_B	hr12_F
1	1	3352.983	58.88606	1933.845	76.30888	2642.689
2	2	63.2931	58.88105	89.55172	75.45968	58.74627
3	3	68.66667	58.81323	104.9	75.15175	56.16667
4	4	57.39683	58.56445	85.66667	74.76758	56.52542
5	5	63.01587	59.67131	94.2381	76.99203	65.05172
6	6	63.96429	59.55289	93.60714	78.06986	57.40678
7	7	59.58929	58.90982	86.625	76.81563	61.56923
8	8	59.81034	57.9498	84.25862	74.67068	59.21053
9	9	64.85965	59.38416	96.5614	75.40396	60.57627
10	10	61.98182	58.2835	84.03636	73.63107	59.58182
11	11	64.54386	58.17255	89.49123	72.38824	58.56667
12	12	147	58.87174	169.4262	74.59118	138.6032
13	13	121.4333	58.93359	168.8	75.56641	93.44444

marrayRaw RGList:
marray package **limma** package

```
@maLabels
[1] "control" "control" "control" "control" "control"
8443 more elements ...
@maInfo
      ID Name
control control geno1
control.1 control geno2
control.2 control geno3
@maNotes
[1] ""
```

1. 容易對資料做較廣泛的運算
2. 完成分析的程式碼較長
3. 容易匯出在其他軟體下作業

1. 可用有限的簡單指令完成分析
2. 完成分析的程式碼較短
3. 不容易匯出在其他軟體下作業