

## Affymetrix GeneChip 中文快速攻略本

### 簡介

GeneChip 是由美國 Affymetrix 公司所建立的系統，主要係利用 Northern blotting 的概念，將原本轉漬在膜上的 target RNA，轉變成是由 user 提供 RNA 樣品，由我們為您合成出 labeled biotin 的 cRNA。至於 Probe 則是由 Affymetrix 公司事先直接在 Chip 上合成好，再經由 hybridization 的方式，將 labeled 好的 cRNA 雜合到 chip 上頭的 oligonucleotide，即所謂的 probe 上。

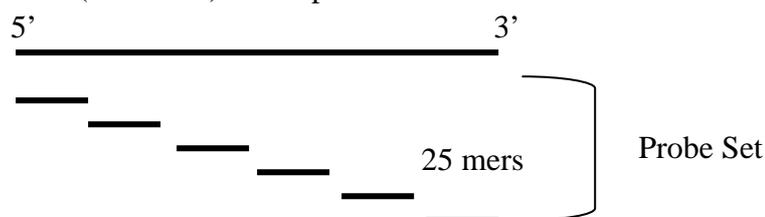
### GeneChip Probe Design

目前可合成的 probe 長度最多為 25 mers，其序列乃是根據 genome project 所定出來的序列，以下利用圖解方式表達比較清楚：

Perfect match (PM)      CAGAATCGATGCTAGTAGTCATCTA  
Mismatch (MM)        CAGAATCGATGC**G**AGTAGTCATCTA



1 Probe Set : 11-20 (PM+MM) Probe pair



當 Labeled biotin 的 target cRNA 在雜合到 probe 上後，經由 streptavidin 染色，經由雷射激發後便可放出訊號，形成黑白兩色的原始圖像。

### 數據的產生

在產生數據之前，affy 的軟體會先調整出適當的 background 值，再經由一些統計運算，將訊號的強度轉換成相對的數值。

Value Category	Absolute Values for Individual Samples
Quantitative	Signal
Qualitative	Detection (e.g. Absent or Present)
Statistical	Detection P-Value

## Data normalization

因為每一片 chip 的訊號強度可能會受到 hybridization 效率等的影響，為了降低這些非生物性因子的影響，我們會先進行 normalization，將頭尾 2% 的訊號數值去掉後，再取剩下數值的 mean 值，令這個 mean 值乘以某個值後的積為 500 (affy default)，而這個值則為 scaling factor。由於每片 chip 的 mean 都不一樣，所以每片的 scaling factor 也不會相同，要注意的是，當 chip 之間要做比較前，要先確定彼此的 scale 設定的值都是一樣的。在我們這邊是採用 affy 所建議的值：500，但其他的機構可能會有不同的設定，請大家要特別注意喔！

## 你所拿到的檔案種類

由於原始檔案大部分只能以 affy 的程式才能讀取，為了大家方便使用，我們會先將檔案轉成 txt 的文字檔，以便各位進行接下來的數據分析。

大致包含兩部分：

**Project folder:** DAT, CEL, CHP, RPT and TXT files。

**3 PDF files:** 包含下面所提到的品管報告外，還有一份使用說明和分析軟體、工具介紹的檔案。

File type	Description
DAT	由 affy 程式所產生的原始圖檔。
CEL	包含訊號強度及 probe 位置等資訊的檔案，可由其他計算方式來產生數據。
CHP	將 CEL 的訊號轉換成數值的檔案，即 raw data。為了讓各位方便使用，我們會直接轉換成.txt 檔。
RPT	紀錄著 hybridization 品質資訊的檔案，我們會將此部分轉成 report.txt 的檔案。
TXT	一份是 raw data 的檔案，另一份 report.txt 則為 RPT 的轉檔。
PDF	為了方便大家快速掌握實驗的品質，我們製作了這一份檔案，其內容為每次實驗流程中各個 check point 的品管報告。

### **TXT (raw data; i.e: File Name = assigned no.txt, e.g. MY9.\*)**

這是從 affy 轉檔的 raw data，其格式為 TXT 檔案。每一個 chip 都會有自己的一個 file，其內容大致如下：

Probe Set ID	Signal	Detection	P-Value	Descriptions
92570_at	64.2	A	0.378184	Cluster Incl AW122482:UI-M-BH2.2-ao...
92571_at	2116.0	P	0.000266	Cluster Incl D85904:Mouse mRNA for ...
92572_at	183.0	P	0.021866	Cluster Incl AI509617:vx14h07.y1 ...
92573_at	4422.7	P	0.000266	Cluster Incl AB021743:Mus musculus ...
92574_at	1928.7	P	0.000219	Cluster Incl AI851046:UI-M-BH0-ajv-..

<b>Probe Set ID</b>	由 affy 所定義出來的 ID，可利用 NetAffy 去找出其代表的基因。
<b>Signal</b>	訊號的強度值，與 sample 中 transcripts 的數量相關。
<b>Detection</b>	提供我們這個訊號的值是否可信，是否具有統計上的意義。A= Absence，也就是在統計上是不具有意義的，（在此可把他當成是沒有表現的基因）；P= Present，也就是在統計上是有意義的，即有表現的基因；M= Marginal，介於 A 和 P 之間的基因，雖然有表現，但在統計上的可信度不夠高。
<b>DetectionP-Value</b>	在統計上計算可信度的-value，也就是用來界定上述 A, P 和 M 的依據。
<b>Descriptions</b>	有關於這一組 probeset 所代表的基因及其相關的資訊，和一些 annotation 的資訊。

### **PDF (Experimental procedure, i.e: MY-1)**

為了讓大家能夠快速掌握此次實驗的品質，在每次實驗結束後，我們將會製作一份有關於此次實驗各個重要步驟的品管報告的彙整。其內容大致可分為下列幾個部分：

<b>基本資料</b>	有關於 sample 的種類、我們所給定的 assigned no、所屬的 project name（在此使用的是各單位的名稱），以及各位登錄的日期等。
<b>RNA QC (total RNA)</b>	在收到各位的 sample 後，為得知 sample 是否在運送的途中發生 degrade，我們會用 Agilent Bioanalyzer 2100 來為您的 RNA sample 做確認。若您所使用的是 mRNA sample，RNA 的 QC 則需等待 IVT 結束後才能得知。

<b>cRNA QC</b>	我們會以測 OD 和電泳分析兩種方式來確認 IVT 完後 cRNA 的品質，需注意的有三項：第一是 RNA 的濃度，第二是 260/280 的 ratio 值要介於 1.9 以上，最後則是合成的 cRNA 在膠上分佈是否呈現 smear 的狀態，至於分佈的 range 範圍，則與各個物種有關。
<b>Fragmentation QC</b>	以 電泳分析的方式來確認 cRNA 是否有被切成適當的大小。
<b>Hybridization</b>	反應的時間記錄。
<b>Wash/ stain</b>	使用的 protocol。
<b>Report</b>	詳細內容請看報告中最右邊一欄的說明。

感謝各位耐心地讀到最後，如果還有疑問的話，歡迎致電：27899590~419 或 e-mail：[Affy@gatesinica.edu.tw](mailto:Affy@gatesinica.edu.tw)給明艷，我將盡力為各位服務。