

專題報導

如何知道基因表現量之高低-基因表現量校正分析

基因表現量僅是一個數值，這些數值大小是否直接代表基因反應的強或弱，一直是個讓科學家困擾的問題。本校資訊工程系白敦文教授、生科系鄒文雄副教授與養殖系吳貫忠助理教授在海大海洋中心支持下進行跨領域合作，透過基因被標註之功能進行基因間的差異分析，找尋最佳之內部控制基因(管家基因)，做為基因表現量高低比較之基準，成功解決此一擾人之難題。



資訊工程學系
白敦文教授



生命科學暨
生物科技學系
鄒文雄副教授



水產養殖學系
吳貫忠助理教授

基因在不同器官或不同時間的表現或受環境條件調控下產生不同反應，不僅啟動生物在各個器官的特殊功能，同時也讓該生物得以適應環境變化的重要反應機制。微陣列晶片(利用已知基因序列與未知序列來進行雜合，希望了解未知序列之種類)與次世代定序技術是目前大量(高通量)分析基因表現量最常使用的實驗方法。但相對於微陣列晶片，次世代定序技術能直接將所有mRNA(全轉錄體)進行定序(RNA-seq)，該方法具有高解析度，可以定序精確到每一個核苷酸，可將mRNA隨環境變化之序列改變內容及相對表現量適當地呈現，具有較高的動態表現。除此之外，次世代定序法不需要目標物種基因體資訊，更適合應用在目前基因體序列未知物種之相關研究。然而，上述兩種實驗方法，在使用前卻存在一個共同需要解決的問題，也就是要如何正確地進行不同資料組的校正，將預備進行比較的兩組基因表現資料數據正確地定量和校正是一項重要議題。

目前針對不同實驗資料集校正(正規化)的方式主要可以分為兩大類，第一類技術是透過統計方法，這種方法會將各別基因資訊忽略，假設整體基因表現量不會有明顯變動，以數值方式計算出實驗條件下基因的總表現量與標準狀態下之正常總表現量比較，得到一校正係數，再透過線性或是非線性的方式對個別基因表現量進行校正。第二類校正技術是透過管家基因進行表現量的校正，這是屬於內部控制的校正方法，也是一種比較具備生物意義的校正方法。所謂管家基因，是指生物體內一群參與細胞維持生存所需基本功能的基因，這些基因的表現與其產物直接關係到細胞的正常運作，因此這群管家基因在所有細胞中皆有表現且表現量亦相對穩定。然而過去實驗常用的方式是取一個或數個管家基因的表現量作為校正的依據，若管家基因在不同實驗條件下表現量是趨於不變，就可以應用管家基因與其他基因的比值作為各別基因表現量指標。過去經常被作為校正標準的管家基因包括GAPDH、 β -actin、 β -tubulin、HPRT等十餘個常用管家基因。有研究指出，使用第一類的統計方式或是使用第二類管家基因作為內部參考基因的方法何者較佳，目前仍然沒有定論。

雖然管家基因的表現量被認為相對穩定，比較不易受環境影響產生變動，但實際上其表現量仍然有可能在某些特殊的時空環境下，出現和原本正常情況下高度差異的表現。例如在血清刺激纖維組織母細胞的相關實驗，經常使用的管家基因包括 β -actin和GAPDH都會變得非常不穩定，同樣的，透過雞隻胚胎時期的實驗發現， β -actin和GAPDH也一樣在不同組織間有著高度的表現差異，因此也不適合作為胚胎發育時期的校正標準。最新的研究甚至指出，由於次世代定序技術產生之RNA-seq在動態範圍與解析度的大幅提升，使得傳統微陣列晶片表現量校正過程所使用的管家基因群，不一定可以適用於RNA-seq的序列資料集。因此本研究建構一套系統可以針對不同實驗特性的轉錄體資料集，自動建議一群相對最穩定的管家基因資料集，透過分析每個管家基因在不同組織及不同胚胎發育時期的穩定性，找出適合作為校正參考的管家基因。目前所建置的系統網址為<http://nhks.cs.ntou.edu.tw/>，主畫面如圖(一)所示。該系統的核心技術是基於每個基因被標註的功能資訊(基因本體論)為基礎進行比較分析。

海洋中心電子報

CENTER OF EXCELLENCE FOR THE OCEANS

基因本體論(Gene Ontology, GO)是一套具備嚴謹架構的詞彙集合，其目的是透過標準化的詞彙描述基因與其產物的功能與特性。我們的系統正是透過標註內容找尋與實驗項目或實驗環境參數最不相關的管家基因群，並作為後續校正的參考。本系統使用兩種方法計算GO標註間的距離：第一種方式是以資訊內容(Information Content, IC)為主的核心理念，將兩個標註與其共同祖先之間的差異性定義為標註之間的距離；第二種方式是透過Czekanowski-Dice距離公式分析兩個GO詞彙所持有“被標註基因”的交集與聯集基因數量，並依據所註解基因的共同性及互斥性計算任兩個GO標註的距離。實驗結果證明本研究所提出新概念的正規化技術，具有重要參考的價值。詳細內容可以參閱本研究發表在Methods期刊論文的完整報告^{註(一)}。

NHKS Gene Expression Normalization by HouseKeeping Selection

Species
Select species of your RNA-Seq expression data
Homo sapiens

Stages of Development
Choose a suitable stage of development for your RNA-seq expression data
 Default (recommended)
 Embryonic stage
 Adult stage

Email
The result(s) will be sent to
twp@mail.ntou.edu.tw

Keyword(s) Optional
Input experiment-related keyword(s) or GO ID(s) or Gene Name(s). Separator: "; "
hypoxia

Upload File Optional
Upload your RNA-seq expression data for normalization
Add file

Parameter Settings
Your species is *Homo sapiens*.
Your period is *allCV*.
Your email is *twp@mail.ntou.edu.tw*.
Warning! Keyword information insufficient, you can add more keywords.
You do not have to upload file

Submit Reset

圖(一) 使用管家基因進行基因表現量正規化之系統主畫面。

註(一)：Chien-Ming Chen, Yu-Lun Lu, Chi-Pong Sio, Guan-Chung Wu, Wen-Shyong Tzou, and Tun-Wen Pai*, “Gene Ontology Based Housekeeping Gene Selection for RNA-seq Normalization,” METHODS (SCI), Vol. 67, Issue 3, pp. 354-363, 2014 (DOI: 10.1016/j.ymeth.2014.01.019)

中心業務報告

1. 海洋中心 103 學年度儀器教育訓練於 8 月底起陸續舉辦，詳情請至中心網站查詢 (http://www.cmbb.ntou.edu.tw/2014_IT)。
2. 海洋中心邀請柳欣博士於 103 年 9 月 1 日至 104 年 8 月 31 日至本校進行研究交流，柳欣博士專長為色素分析，利用色素比例推定浮游植物種類組成。在本校期間將參與相關航次，進行本省沿岸水域特別是淡水河口水域植浮組成研究。