

知識天地

淺談致病基因定位

楊欣洲助研究員（統計科學研究所）

要戰勝敵人必須先找出敵人。

人類許多遺傳性疾病的發生導因於染色體上的某些特定區域或是基因發生變異，例如，跳舞症（Huntington disease）是由於人類第四號染色體短臂末端序列片段發生重複所造成。要想戰勝這些疾病，第一步就是要先找出這些染色體發生變異的位置，然後研究該變異與致病機轉間的關係，最後開發診療技術與藥物，改善國民健康。找出與疾病發生有關的染色體變異位置這項工作稱為「致病基因定位」（disease gene mapping）。

簡單來說，致病基因定位就像是警察要追捕搶匪一樣（故事情節的描述與對比請參見表一），目的都是要找到危害人類的壞蛋。由於人類基因體的高複雜度，要從如此龐雜的基因體中找出與疾病發生相關的位置，猶如大海撈針，難度自然可想而知。

表一、警察捉搶匪與致病基因定位兩個故事情節的比較。

故事	情節一	情節二
主軸內容	警察捉搶匪	致病基因定位
故事背景	搶匪擾亂社會治安	基因變異危害人體健康
主人翁	英勇的警方	強力的統計檢定
主要任務	捉搶匪	找致病基因
藏匿範圍	台灣和澎湖地區	人類基因體
責任分區	二十三個縣市	二十三對染色體
受檢對象	兩千三百萬居民	三十億對的鹼基對
追捕工具	無線電收發器	標誌基因

想像一下，搶匪橫行擾亂治安，警方要從台灣本島和澎湖地區的二十三個縣市裡，約兩千三百萬居民之中，找出那些危害社會治安的壞蛋，遍佈各地的維安警員各個手持無線電收發器，隨時接收或傳送歹徒可能出現的位置，一待訊號出現，就可鎖定搶匪的所在位置，然後集結警網，將搶匪一舉成擒。將這樣的故事情節類比於致病基因定位，就像是統計學家利用統計檢定，試圖要從人類基因體的二十三對染色體裡，約三十億對的鹼基對之中，找出致病基因的所在位置，大量散佈在染色體上的標定物（標誌基因，genetic marker），其在染色體上的位置已經知道，一旦統計檢定發現某些標誌基因與未知的致病基因間很靠近，那就鎖定這些標誌基因的鄰近範圍，檢查更多更密的標誌基因，以進一步定位出致病基因的確切位置，最後找到致病基因。

如何才能有效地逮到這些致病基因？這裡簡介兩種廣泛使用的統計致病基因定位法的概念。第一種方法稱為連鎖分析（linkage analysis），我稱之為指南針定位法，利用這樣方法所得的分析結果，可以提供致病基因大致的方位。連鎖分析的主要概念是透過觀察細胞減數分裂（meiosis）過程中，染色體的某些染色分體發生交換（crossover）後重新組合的現象。一般而言，同一染色分體上，相隔愈近的兩個位點間的重組率（recombination fraction）通常愈低。我們可透過收集罹病家族成員的罹病狀態資料與標誌基因資料，觀察標誌基因的基因型（genotype）由前代傳遞到下一代的模式，然後用統計方法推估出致病基因與每個標誌基因間的重組率，低重組率表示致病基因與標誌基因有連鎖，亦即致病基因就在該標誌基因附近。這樣的方法對於定位孟德爾型疾病（Mendelian disorder）

（這類的疾病的發生通常由一個或少數幾個基因的變異所控制）的致病基因非常有效，過去已經協助找到跳舞症、囊腫纖維症（cystic fibrosis）等遺傳疾病的致病基因。但連鎖分析有兩個主要缺點，第一點是，對於複雜型疾病（complex disorder）的致病基因定位成效不彰，這類疾病的特色是，該疾病的發生是由很多基因變異所共同決定，但由於大多數基因的貢獻都不大，因此不容易被偵測到。第二點是，由於染色分體間發生交換通常約要 0.01 摩根（Morgan）以上的距離，因此連鎖分析的解析度受限，難以進一步縮短致病基因所隱藏的區間，這時就有賴下面介紹的第二種方法的協助。

第二種方法為關聯分析（association analysis），我稱之為衛星定位法，分析結果可以提供致病基因明確的位點。關聯分析的主要概念是透過觀察連鎖不平衡（linkage disequilibrium）現象來定位致病基因。在逢機交配族群中，突變發生之始，存在祖先染色體中的標誌基因與致病基因間的完全連鎖不平衡，在每經過一代繁衍後，就被重組現象逐漸削弱，多代之後，達到連鎖平衡。因此，現今的連鎖不平衡只會在極短的距離內被觀察到，換言之，若觀察到

某標誌基因與致病基因間存在連鎖不平衡，這表示致病基因應該就在該標誌基因附近。關聯分析可以透過收集病例對照型或是家族型個體樣本來進行，利用所收集到的個體樣本的罹病狀態資料與標誌基因的基因型資料，分別計算每個標誌基因與致病基因間的連鎖不平衡或是傳遞不平衡 (transmission disequilibrium)，顯著的不平衡現象表示致病基因就在該標誌基因附近。這樣的方法過去已成功地協助找到一些複雜疾病的致病基因，例如，阿茲海默症的重要致病基因 *APOE* 和克隆氏症的重要致病基因 *NOD2*。近年來，隨著生物科技的快速發展，愈來愈多的標誌基因被發現，特別是單一核苷酸多態性標誌基因 (single nucleotide polymorphism, SNP)，現已發現超過三百萬個以上，這樣稠密且廣佈於人類基因體上的基因地標，正好可與關聯分析相輔相成，為關聯分析致病基因定位提供更高的解析度與統計檢定力。這樣的方法正大量地被應用於定位各種疾病的致病基因。

以上所介紹的統計致病基因定位仰賴標誌基因的資料提供訊息。為了得到標誌基因的基因型資料，必須收集每個個體樣本的去氧核糖核酸 (deoxyribonucleic acids, DNA)，然後進行「個別基因型鑑定實驗」 (individual genotyping experiment)，以取得每個個體樣本在每個標誌基因上的基因型資料。然而，在進行這類基因型鑑定實驗前，成本是很重要的考量，特別是大規模的全基因體關聯分析，這樣的研究往往需動用成千上百的個體樣本，並同時考量數十萬甚至百萬的標誌基因。雖然近年來基因型鑑定的成本愈降愈低，但要進行這類大型計畫仍是索費不貲，常令人望之卻步。舉例說明，以一個病例對照組實驗為例，該研究收集 1,000 個病患樣本和 1,000 個對照組樣本，其中，性別 (男與女) 與年紀 (年長與年輕) 皆等比率。現在針對每個樣本皆進行個別基因型鑑定實驗，所採用的基因晶片 (gene chip) 一次可鑑定超過 100,000 個單一核苷酸多態性標誌基因，以中研院國家基因型鑑定中心一個樣本 21,000 元的學術界鑑定成本價來計算，光是鑑定經費就要超過 40,000,000 元，這對於經費比較拮据的實驗室而言是很高的開銷。

對於無法負擔較昂貴的個別基因型鑑定實驗的研究者，或許可以考慮較節省成本，但仍具效力的基因鑑定實驗，也就是「混合鑑定實驗」 (pooled allelotyping experiment)。顧名思義，混合鑑定實驗就是將許多人的去氧核糖核酸等量混合後才一起進行鑑定，因此大量降低了由於個別鑑定大量個體樣本所造成的成本。仍以前述相同的例子來說明，我們將病例組的 1,000 位病患依性別和年紀分成 4 組，每組中的個體樣本的去氧核糖核酸被等量地混合成 1 管新的混合去氧核糖核酸；同樣地，對照組個體樣本的去氧核糖核酸也被混合成另 4 管新的混合去氧核糖核酸。並且，為了精確性的考量，鑑定實驗將重複兩次，因此總共會有 16 管混合去氧核糖核酸需要被鑑定，所需的鑑定成本為 336,000 元，只有個別基因型鑑定實驗成本的 1/125。

為何混合鑑定實驗可以用來進行關聯性致病基因定位？在病例對照組的例子中，我們主要關心的是，病例組中各個標誌基因的等位基因頻率 (allele frequency) 是否和對照組的有差異，並不在意個人的基因型，而病例組和對照組各自的混合去氧核糖核酸之中，仍保有各自組內等位基因頻率整體的訊息，因此適合用來進行關聯性致病基因定位。如何進行混合型關聯性致病基因定位？透過混合鑑定實驗鑑定每一管混合去氧核糖核酸，結果可得到每一個標誌基因上的等位基因 (allele) 的雜交強度資料 (hybridization intensity)，利用此資料可以估計每個標誌基因上的等位基因頻率，之後透過比較病例組與對照組的等位基因頻率間的差異，即可偵測標誌基因與致病基因間的連鎖不平衡，達到篩選可能的致病基因的目的 (Yang et al., 2005; Yang et al., 2006)。下一階段則是針對這些少數的候選標誌基因進行確認實驗 (confirmatory study)，由於此時所涉及的標誌基因的個數變得很少，因此不會再耗費太多鑑定成本。這樣的混合型關聯致病基因定位已經被證實對於篩選重要的候選致病基因確實有用 (Pearson et al., 2007)。

個別型和混合型關聯致病基因定位各有擅長之處。前者主要優點在於精準度高，但卻十分昂貴；後者主要優點是節省成本，但也有其缺點，由於來自不同個體樣本的去氧核糖核酸在實驗中已經均勻混合，原屬於每個個體特有的訊息幾乎遺失，這造成許多分析難以進行。兩個策略孰優孰劣時有論戰，但個人認為這就好比美伊戰爭一般，個別基因型關聯分析之於混合型關聯分析，就好比船堅砲利的美國之於人人皆兵的伊拉克。如同美國的船堅砲利，戰斧巡弋飛彈總能確實命中攻擊目標，個別基因型關聯分析的優勢在於具有高精度與可靠度，雖然成本相對昂貴；混合型關聯分析的優勢在於節省成本，就如同伊拉克武器雖不精良，但人人皆兵，幼童也可拿槍作戰。從戰爭的目的來看，美伊戰爭的目的在於石油利益 (也許這只是作者的猜測)；致病基因定位的目的則是在找到致病基因。美國用人造衛星探勘石油的所在位置；統計學家則是考慮高解析度的關聯分析。最後，美伊戰爭轟轟烈烈地結束，我想大家都同意，兩個國家都輸了這場戰爭；個別基因型關聯分析與混合型關聯分析間的爭論誰將是最後的贏家？我想合作讓雙方都成為贏家，也唯有跨方法及跨領域的密切合作才可能讓目標更接近。

找到敵人離戰勝敵人還有好長一段路程需要共同努力！

參考文獻：

- Pearson JV et al. (2007) Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide polymorphism association studies. *Am J Hum Genet* **80**, 126-139.
- Yang HC et al. (2005) New adjustment factors and sample size calculation in DNA-pooling experiment with preferential amplification. *Genetics* **169**, 399-410.
- Yang HC et al. (2006) A genome-wide study of preferential amplification/hybridization in microarray-based pooled DNA experiments. *Nucleic Acids Res* **34**, e106.