

生物資訊之蛋白體探索： 認領第四號染色體 台灣躍上國際學術舞台

計畫主持人
宋定懿

蛋白質是基因轉譯的最終產物，在生物體內執行各種不同的功能。在醫藥研究上，人類的蛋白質是最主要的藥物標的。因此在後基因體時代，針對生物體進行大規模蛋白質分析的蛋白體學逐漸受到重視。隨著質譜儀器及實驗技術的精進，Aebersold 和 Mann 兩位學者 2003 年在 *Nature* 期刊上發表一篇「奠基於質譜儀資料分析的蛋白體學」重要論文，象徵蛋白體學跨入重要里程碑；接著 2005 年 Ong 與 Mann 兩位學者在 *Nature Chemical Biology* 期刊上發表了另篇論文，宣告蛋白體學研究邁向定量的研究。近十多年來，液態層析搭配質譜儀的實驗技術成為蛋白體學最重要的實驗方法；此方法首先將樣本中的蛋白質分解成肽，利用液態層析法分離，再送入質譜儀分析。研究者藉由質譜儀所產生的圖譜大數據，來鑑定生物體、組織或細胞內在不同狀態的蛋白質，及分析不同狀態下蛋白質表現量的差異。但由於質譜儀斷裂機制、樣品的品質及高複雜度、實驗中難免的背景雜訊、數個大分子同時流出或斷裂等因素，造成實驗資料難分析。本實驗室於 2003 年底和本院化學所陳玉如教授長期合作，投入質譜大數據分析的生物資訊研究，提出演算法，開發完成三個不同的蛋白定量軟體 Multi-Q、MaXIC-Q 與 IDEAL-Q，提供學界下載使用。

台灣認領人體第四號染色體

隨著蛋白體學研究的蓬勃發展，人類蛋白體學組織在 2011 年提出類似人體基因體計畫的人體蛋白體計畫 (Chromosome-centric Human Proteome Project, C-HPP)，希望描繪人體從基因轉譯過來的所有蛋白質。此計畫由各國負責不同染色體中蛋白體的分析，台灣團隊由陳玉如教授負責召集，認領第四號染色體以在國際學術舞台發聲；我們實驗室負責台灣團隊的生物資訊研究。2013-2015 年 C-HPP 重要的目標是找尋「失蹤蛋白質」，儘管這些蛋白質的序列已由 DNA、RNA 經過嚴謹的推論得知，但在抗體或質譜實驗上未被看到。失蹤蛋白質未在實驗上被確認，可能的原因很多，如：量少、只在某些過渡的狀態下出現、實驗困難。

陳玉如教授實驗室以多個肺癌的細胞株進行質譜實驗，質譜大數據再用定性軟體（如：Mascot）鑑定出其中的蛋白質。儘管軟體可能鑑定出某些失蹤蛋白質，但這不表示就找到了，必須滿足嚴謹的條件以及使用標靶實驗才算確認。根據軟體產生的蛋白質鑑定結果，我們實驗室以資訊專長擔負嚴謹的資料分析。由於大多數蛋白質定性軟體未提供蛋白質的錯誤發現率，因此我們首先以現有的軟體 (PeptideShaker) 進行蛋白質錯誤發現率的分析，刪除錯誤發現率超過 1% 的蛋白質。其次，蛋白質的鑑定是從大量圖譜鑑定出的肽推論而來；為求推論的準確性，這些蛋白質必須包