

多專家線性抗原表位預測系統之設計與實現

Design and realization of a multi-expert linear epitope prediction system



指導教授：葉春超、周信佑、白敦文

00757122張書瑜、00757012黃乃軒、00757026邵安祺

email: sweety89021089@gmail.com、nagin2358@gmail.com、shauangel2000@gmail.com

摘要

有效預測病原菌的抗原表位結合區段是開發具確效性的疫苗及藥品最重要前置分析工作，然而傳統生物實驗方式需耗費相當的實驗經費與時間，至今也尚未出現突破性的生物資訊技術協助抗原表位的有效預測。本研究統合六個使用不同特徵及抗原表位預測技術的專家系統，透過權重式投票機制同步預測抗原蛋白的高抗原胜肽片段，預期此系統可以提升次單位疫苗及專一性檢測套件的開發。本研究使用 Bcipep 與 IEDB 測試資料集進行預測模型的評估，實驗結果顯示本研究所提出的多專家系統的預測表現穩定優於其他各個子系統的單獨表現，具權重式的多專家推薦系統的整合可以提供更穩定的抗原表位預測效能。

研究方法

本研究使用生物資訊技術對多個線性抗原表位預測系統進行實驗數據精確度分析，以此實驗結果計算出各個預測系統對不同抗原序列內容的加權分，以提高系統預測的精準度。

▶ 系統權重參數訓練

本研究採用單一胺基酸及鄰近胺基酸對做為兩類權重係數的分析單位，總共有420個相對應權重參數作為後續六個推薦專家系統的對應推薦效力，以提升各個抗原預測專家系統的推薦有效性。

▶ 權重計算

主要是計算單一胺基酸及鄰近胺基酸對兩類權重係數，對已知抗原表位片段資料集分析六個預測系統對不同單位內容的F1-score預測精確度差異表現，並以該預測結果作為各不同專家預測系統對單一胺基酸及鄰近胺基酸對兩類權重係數的動態權重比例。此分數在經過正規化的計算後，當六個系統皆推薦此抗原胺基酸為抗原表位時，將獲得滿分1分的推薦，若是所有專家系統都沒有推薦時，該抗原胺基酸則獲得0分的推薦分數，也就是極力不推薦。

▶ 多專家投票系統製作

本研究根據上述研究內容開發一個多專家投票系統網頁，系統架構設計如圖1所示。該網頁功能包含六個專家系統之自動預測、權重投票分析及預測結果視覺化呈現。

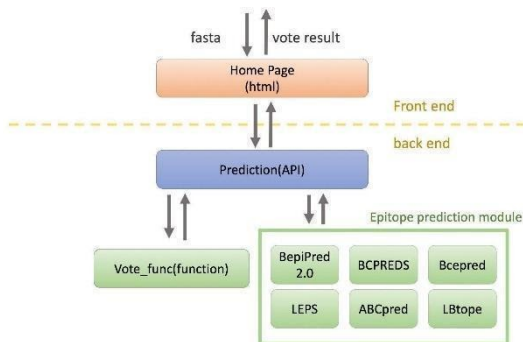


圖1 網頁架構圖

結果與討論

▶ 抗原表位預測結果

我們使用 Bcipep 與 IEDB(fish) 測試資料集進行測試實驗，分別對原先六個專家系統以及本研究開發之多專家投票系統進行F1-Score評測。經過計算後得出本系統最好的表現F1-Score可達0.6858及0.6216，高於任一專家系統。由此可見，基於多專家系統的預測準確率相對於單一專家系統的表現可以提供更好的預測表現。

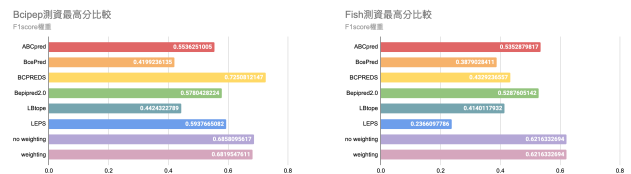


圖2 本系統及原先六系統使用Bcipep測試資料集的F1-Score

圖3 本系統及原先六系統使用 IEDB(fish)測試資料集的F1-Score

▶ 多專家抗原表位預測系統之網頁呈現

多專家抗原表位預測系統



圖4 原始輸入頁面

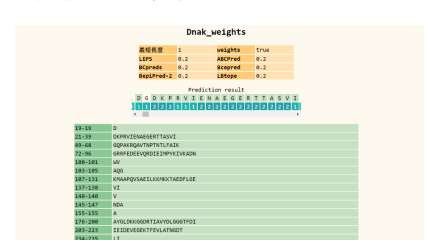


圖5 預測呈現畫面

結論與未來展望

經過比對本研究所提出之權重式多專家預測系統與傳統所使用的六個線性抗原表位預測系統進行預測結果比對，可以證實多專家系統整合後對於次單位疫苗開發將有實質的助益。相信未來持續累積真實實驗證明之線性抗原表位之訓練資料，並嘗試多種機器學習之資料訓練模組，可以有效提升各胺基酸及胺基酸對的有效權重參數，對未來多專家系統的正确預測結果有更多突破的機會。

致謝

本研究感謝科技部贊助周信佑教授的整合型研究計畫(MOST 109-2321-B-019-005)及科技部贊助黃乃軒的大專生專題計畫