

多基因風險分數(polygenic risk score)之介紹與應用

謝宜蓁 教授

台北醫學大學神經醫學博士學位學程

隨著人類基因組計畫(human genome project)成功解密了基因序列，全基因體關聯性分析(genome-wide association study, GWAS)打開了一扇通往研究複雜性疾病的大門，找到許多未曾發現的基因或染色體區域，提供複雜性疾病相關致病機制的線索。近年來，研究者認為單一基因多形性(single nucleotide polymorphism, SNP)影響有限，不足以用來評估個體對某一個複雜性疾病的風險，因此便發展出多基因風險分數(polygenic risk score, PRS)指標，期望此遺傳數據可應用於預測疾病風險，具有潛在臨床效用。

PRS 的概念介紹與定義

PRS 旨在量化多個基因或位點的累積效應，將數個基因組變異訊息濃縮成衡量個體對於表現型(phenotype)或性狀(trait)遺傳傾向的估計。簡單來說，是將每一個個體攜帶變異等位基因之數量(0、1 或 2)的加權總和，而權重為來自 GWAS 數據中變異等位基因與表現型之間的效應量估計值(即二元性狀的 log (OR)或連續性狀的 beta 係數)，並假設遺傳模式為累加模式(additive model)。

其公式如下：

$$PRS_i = \sum_j^M \beta_j \times X_{ij}$$

i : 表示第 i 個個體
 j : 表示第 j 個 SNP
 β : 表示該 SNP 的權重
 X : 表示第 i 個個體第 j 個 SNP 等位基因變異數

建構 PRS

PRS 建構過程包含兩個步驟：首先為選擇變異位點(即 SNP)的過程，以確定哪些易感受性位點會納入模式中；其次為權重估算，以獲得需要附加到所選擇之變異位點的係數或權重的過程。常見 SNP 的加權數值一般是來自於探索階段(discovery stage)的 GWAS 匯總統計資料，這些資料通常也都會公開於美國國家人類基因體研究所和歐洲生物資訊研究所(NHGRI-EBI GWAS Catalog)網站上。此外探索階段的 GWAS 樣本數須足夠大，才能有助於減少樣本誤差，獲得更精準的效應量估計值，同時可納入效應量較小的 SNP。選擇納入建構 PRS 的 SNP 以及效應量估計值也是重要的考量，目前常見的方法如下：

1. Pruning and thresholding (P+T)

此為一標準常用的方法，就是設置一個 p 值作為閾值，並利用連鎖不平衡(linkage disequilibrium, LD)來篩選出要建構之 SNP，亦即在每一個 LD 區塊中選擇具有最大效應量的 SNP。最佳的 P 值閾值取決於樣本大小與表現型的遺傳結構。

2. Bayesian polygenic risk score

此為利用貝式方法透過考慮了基因變異點之間的連鎖平衡程度以及 GWAS 的原始效應大小後，計算 SNP 的聯合效應大小，其又稱為 LDpred，在大樣本量的情況下是較為精準的方法。

3. Stacked clumping and thresholding (SCT)

此堆疊聚集和閾值算法(SCT)為最近開發的機器學習方法。此方法使用 GWAS 摘要統計數據和驗證數據集學習新的 SNP 權重，通過四維參數(LD 相關係數、 p 值閾值、聚集窗口大小和插補品質)執行重複的 P+T 來運算。

4. Meta-scoring

此方法將一個性狀的多個 PRS 組合成一綜合分數(metaGRS)，其假設每個單獨的 PRS 都會在一定程度上受到迴歸稀釋偏差的影響，因此，將它們組合成更強大的綜合分數將減少這種偏差。綜合分數本身被構建為單個 PRS 的線性(或非線性)組合，通常包括目標性狀和相關性狀的分數。

PRS 的驗證

關於 PRS 驗證的黃金標準是將探索階段所發展之 PRS 利用獨立的數據資料庫來進行驗證，並評估 PRS 的預測效果並盡可能的減少 PRS 建構過程中過度擬和(overfitting)的可能性。



圖一 多基因風險分數發展過程簡介

PRS 的應用範圍

1. 辨識高風險群以提供預防性治療

PRS 在複雜性疾病的風險預測上目前被廣應用於找出高風險族群，以提供預防性的治療。例如美國哈佛大學與麻省理工學院合辦的 Broad Institute 利用 PRS 篩選出五種常見的高風險群，分別於人群中識別出 8.0%的冠動脈心臟病、6.1%的心房纖維顫動、3.5%的第二型糖尿病、3.2%的發炎性腸症和 1.5%的乳癌高風險群，發現其罹病的風險均增加了 3 倍以上，故研究人員建議考慮將 PRS 預測納入臨床護理。

2. 優化篩檢策略

PRS 在複雜性疾病篩檢上也應用於優化篩檢方案，提高篩檢成本效益。以乳房篩檢為例，英國國民保健署乳癌篩檢項目建議年齡介於 47~73 歲之間的女性應該接受常規的乳房 X 光攝影檢查。研究估算 47 歲女性發生乳癌的 10 年絕對風險平均值為 2.4%；若利用 PRS 進行風險預測，結果發現年齡介於 32~35 歲之年輕女性，且 PRS 高於 90 百分位者，其發生乳癌的 10 年絕對風險就已經達到 2.4%，因此建議應納入高風險族群進行篩檢；而對於 PRS 低於 90 百分位之女性，她們終其一生罹患乳癌的絕對風險則小於 2.4%。故相較於傳統僅依據年齡範圍納入高風險篩檢的條件，PRS 將更有助於為不同遺傳風險的女性提供更精準的篩檢策略。研究進一步評估了應用 PRS 於乳癌篩檢的成本效益，發現當篩檢限制在 PRS 高於 70 百分位的高風險女性時，每 1 萬人次的篩檢成本將比單純基於年齡的篩檢節省近 54 萬英鎊，過度診斷則會減少 49.9%，而生活品質調整後存活年數(quality-adjusted life year, QALY)會增加 443。因此，利用 PRS 篩選高風險族群將可提供比單純使用年齡範圍更精準的篩查策略，同時避免低風險女性接受不必要的檢查。

3. 協助藥物開發

由日本、歐洲和美國研究團隊透過重新分析組合 GWAS 數據發展出藥物性肝損傷(drug-induced liver injury, DILI)的 PRS，應用於協助藥物開發。對於臨床醫生來說，PRS 將使他們能夠進行快速的基因檢測，以便在開藥之前確定肝損傷風險較高的患者，提示醫生更改劑量或醫囑，以發現肝損害的早期徵兆或完全更換藥物。

PRS 的挑戰

隨著人類基因體研究發展迅速，PRS 在疾病風險辨識、篩檢及精準預防等方面已顯現出其價值，預期臨床應用上將具有相當潛力。然而目前實現 PRS 的廣泛應用仍面臨挑戰。首先為 PRS 辨識出高風險群僅包含全部研究族群中的一小部分，仍有一定程度上歸因於遺傳缺失(missing heritability)，故仍須有更大規模的定序和 GWAS 研究以找出更多常見、低頻率或

罕見的疾病易感受性位點，以提高預測能力。其次為基因與環境對於複雜性疾病發病過程中的交互作用扮演相當重要的角色，且環境因素會隨著時間的遷移而有所變化，故需開發新穎統計學方法將動態的環境暴露訊息納入預測模型，並加以驗證和評估。最後，由於 PRS 是評估個體在未來某一時間段內罹患疾病風險的概率，而非判斷個體是否患有疾病，故臨床人員如何有效地評估疾病風險預測並採取科學性的解讀，以及其所衍伸之可能法律和倫理議題等皆是需要考慮的重要問題。

總而言之，通過推動大規模世代追蹤的流行病學研究，整合基因與環境因子以構建更為完善的預測模型，並利用大規模的人群來進行預測效果的評估，將有助於優化現有疾病篩檢方針、推動個人化精準預防，最終實現降低疾病負擔、提高民眾健康水平的終極目標。

參考文獻

- Babb de Villiers C, Kroese M, Moorthie S. Understanding polygenic models, their development and the potential application of polygenic scores in healthcare. *J Med Genet.* 2020;57(11):725-732.
- Bolli A, Domenico PD, Pastorino R, Busby GB, Bottà G. Risk of Coronary Artery Disease Conferred by Low-Density Lipoprotein Cholesterol Depends on Polygenic Background. *Circulation.* 2021;143(14):1452-1454.
- Choi SW, Mak TSH, O' Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols* 2020;15(9):2759–2772.
- Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics.* 2018;50(9):1219-1224.
- Koido M, Kawakami E, Fukumura J, Noguchi Y, Ohori M, Nio Y, Nicoletti P, Aithal GP, Daly AK, Watkins PB, Anayama H, Dragan Y, Shinozawa T, Takebe T. Polygenic architecture informs potential vulnerability to drug-induced liver injury. *Nature Medicine.* 2020;26(10):1541-1548.
- Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine.* 2020;12(1):44.
- Pashayan N, Morris S, Gilbert FJ, Pharoah PD. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model. *JAMA Oncology.* 2018;4(11):1504-1510.