

行政院國家科學委員會專題研究計畫 成果報告

以 SNP 標記進行全基因體相關掃描之樣本數估算

計畫類別：個別型計畫

計畫編號：NSC91-2320-B-002-109-

執行期間：91年08月01日至92年07月31日

執行單位：國立臺灣大學公共衛生學院流行病學研究所

計畫主持人：李文宗

計畫參與人員：張晉豪、蘇郁翔、陳彥甫

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 92 年 10 月 22 日

行政院國家科學委員會專題研究計畫成果報告

計畫名稱：(中文) 以 SNP 標記進行全基因體相關掃描之樣本數估算
(英文) Sample size calculation for genomewide association scans using SNPs

計畫編號：NSC 91-2320-B-002-109-

執行期限：91/08/01~92/07/31

主持人：李文宗 台灣大學公共衛生學院流行病學研究所

計畫參與人員：張晉豪 研究生 台灣大學公衛學院流行病學研究所
蘇郁翔 研究生 台灣大學公衛學院流行病學研究所
陳彥甫 兼任助理

TEL: (02) 23123456-8357

FAX: (02) 23511955

E-Mail: wenchung@ha.mc.ntu.edu.tw

一、中文摘要：

為探索複雜性疾病之遺傳因素，流行病學中的相關性研究方法會越來越扮演重要的角色。近來，以分析病例雙親三合體資料的傳遞不平衡檢定法（以下簡稱 TDT 法）頗廣為採用。在基因標記和疾病易感受性基因兩者有相關的前提下，TDT 可用來檢定兩者間的連鎖關係，也避開因族群分層所導致檢定偽陽性率偏高的問題。然而，TDT 需同時收集病例以及其雙親，不適用於晚發疾病。對此，本研究發展出兩種新的研究方法，即，惟病例法、以及病例配偶及病例子代法。惟病例法，顧名思義就是只收病例。此法雖然便利，仍無法避免族群分層所產生的效應。對於晚發疾病，

病例配偶 / 子代法不僅收案較 TDT 法容易，且可避免族群分層導致的偏差。本研究的兩種新方法，不僅收案較為容易，而且已發展得到樣本數的估計方法。電腦模擬顯示某些情況下，新法所需的樣本數比 TDT 所需的樣本數少。

關鍵字：病例雙親三合體、惟病例法、病例配偶 / 子代法、遺傳流行病學方法、傳遞不平衡檢定

二、緣由與目的

複雜人類疾病(complex human diseases)的易感受性基因(susceptibility gene)將會越來越倚重流行病學「相關」類型的研究(association approach)作基因定位^{1,2}。其中一個最被重視的方法，係以傳遞不平衡檢定

(transmission/disequilibrium test, 簡稱 TDT)³ 做為基因定位的工具。

TDT 的原理係檢定異合子 (heterozygous) 親代傳遞給有病子代某個標記基因的機率 (transmission probability), 是否有偏離孟德爾遺傳定律的 1/2。TDT 可以視為傳遞基因與未傳遞基因間的配對研究, 是同時檢定疾病基因座與標記基因座間是否有「連鎖」(linkage) 以及「相關」(或「不平衡」, disequilibrium) 的一種檢定。此外, 若在分層族群下, 傳統病例對照的檢定, 偽陽性會偏高, 而 TDT 可以校正這種誤差⁴。然而, TDT 的收案對象必須同時收集病例及其雙親。對於晚發疾病而言, 例如, 非胰島素依賴型糖尿病、心血管疾病、阿滋海默症以及許多的癌症等等, 欲執行 TDT 的收案方式, 並不容易。

有鑑於此, 本研究發展兩種新的研究方法, 即, 惟病例法、以及病例配偶 / 子代法。惟病例法只需要病例的基因資料, 是相當實用的方法。此方法原本是應用在易感受性基因的細部定位 (fine mapping)⁵⁻⁷。本研究則將其推廣, 應用在全基因體掃描 (genome-wide scan) 的場合。

病例配偶 / 子代法, 則是收集病人、其配偶或後代。對於晚發疾病的病人, 其配偶以及子代往往會陪在病人身旁。因而此法的收案過程將會比 TDT 容易許多。

近來由於分子生物學的快速進步, 大量標記基因可以很快速的測定。其中尤以「單點核苷酸多形性」(single nucleotide polymorphism, 簡稱 SNP) 標記最受關注⁸。目前估計基因體上每 1kb 大約即可找到一個 SNP。因此可以預期未來易感受性基因研究,

會越來越倚重 SNP 標記。本研究的兩種新方法, 乃利用 SNP 資料進行分析, 尋找易感受性基因的可能位置。

不論何種研究方法, 樣本數估計是進行研究時一項重要的參考因素。所以本研究除了建立兩種新的研究方法, 也提出其樣本數估計的方式, 使新方法更趨完整。我們也利用電腦模擬, 驗證樣本數公式。並在各種參數下, 計算所需的樣本數大小。

三、結果與討論

電腦模擬結果顯示兩個樣本數公式誤差率極低。在隱性基因頻率 0.5, 或是顯性基因頻率 0.5 時, 惟病例法所需的樣本數皆少於 TDT 法。然而, 此法在分層族群下, 檢定結果的偽陽性可能會偏高。

病例配偶 / 子代法可控制分層族群所導致的偏差。此外, 即使發生率或是其他危險因子的暴露率有性別上的差異, 仍可適用此法。病例配偶法所需的家庭數 (一家兩個案) 與 TDT 法 (一家三個案) 相似。病例子代法, 在單子代家庭的情況下, 則需約 2 倍的家庭數。若欲研究晚發複雜疾病的基因, 此法應是好的替代選擇。然而需符合下列假設。1. 男、女性的基因標記頻率在受精卵初期就已相等。2. 從懷孕過程到出生後各年齡層, 標記基因不會與其他競爭死因有關。

四、計畫成果自評：

尋找疾病未知的易感受性基因, 是遺傳流行病學家專注的一個方向。本研究附有樣本數計算方式的兩種新方法, 相信將會是一個重要的參考指南。此外, 本計畫的研究成果已獲 “American Journal of Epidemiology”⁹⁻¹² 以及 “Journal of Genetics and Molecular Biology” 接受發表¹³。

五、參考文獻

1. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516-7
2. Khoury MJ, Yang Q. The future of genetic studies of complex human diseases: an epidemiologic perspective. *Epidemiology* 1998;9:350-4
3. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52:506-16
4. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision and admixture. *Am J Hum Genet* 1995;57:455-64.
5. Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996;13:399-408
6. Nielsen DM, Ehm MB, Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet* 1999;63:1531-40
7. Jiang R, Dong J, Wang D, et al. Fine-scale mapping using Hardy-Weinberg disequilibrium. *Ann Hum Genet* 2001;65:207-19
8. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998;280:1077-82
9. Lee WC. Searching for disease-susceptibility loci by testing for Hardy-Weinberg disequilibrium in a gene bank of affected individuals. *Am J Epidemiol* 2003;158:397-400
10. Lee WC. Lee responds to "Testing for Hardy-Weinberg Disequilibrium". *Am J Epidemiol* 2003;158:404-405
11. Lee WC. Genetic association studies of adult-onset diseases using the case-spouse and case-offspring designs. *Am J Epidemiol*; in press
12. Lee WC. Lee responds to "Making the Most of Genotype Asymmetries": a weighting scheme for the disequilibrium test. *Am J Epidemiol*; in press
13. Lee WC. Testing for Hardy-Weinberg disequilibrium based on cases only: a general approach for sample size calculation. *J Genet Molecul Biol* 2003;14:159-163

ABSTRACT

The future of genetic studies of complex human diseases will rely more and more on the epidemiologic 'association' paradigm, in particular the application of the transmission/disequilibrium test (TDT) to detect linkage disequilibrium in a case-parents study. TDT is not affected by population stratification. It is difficult, however, when the disease of interest has an age-at-onset in adulthood or older. We propose two new methods, the case-only and case-spouse/offspring studies. The sample size formulae are presented, which are very accurate as revealed by Monte-Carlo simulations. If the disease-susceptibility is recessive with an allele frequency of ≤ 0.5 or dominant with an allele frequency of ≥ 0.5 , the number of subjects needed by the case-only study is smaller than that needed by using the case-parents study. However, the case-only study is prone to

produce false positive and false negative results in a stratified population. With the following two assumptions, the case-spouse/offspring study is valid even if only one sex can get the disease, even if cases/spouses/offspring all have different risk factor profiles: 1. the marker genotype frequencies at conception should be the same for both sexes; and 2. there is no selective attrition of marker allele(s) through gestation and over age. The number of families required in a case-spouse design is almost identical to that in a case-parents design. For the case-offspring study with one offspring per family, the number of families should be doubled.

Keyword: case-parents triads; case-only study; case spouse/offspring study; genetic epidemiologic methods; transmission/disequilibrium test