

行政院國家科學委員會專題研究計畫 成果報告

高三酸甘油酯 / 低高密度脂蛋白膽固醇之家族遺傳研究

(3/3)

計畫類別：個別型計畫

計畫編號：NSC92-2314-B-002-197-

執行期間：92年08月01日至93年07月31日

執行單位：國立臺灣大學醫學院內科

計畫主持人：簡國龍

計畫參與人員：許秀卿，李源德

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 11 月 2 日

# 行政院國家科學委員會補助專題研究計畫成果報告

高三酸甘油酯 / 低高密度脂蛋白膽固醇之家族遺傳研究 (3/3)

計畫類別： 個別型計畫          整合型計畫

計畫編號：NSC 92-2314-B-002-197

執行期間：92 年 8 月 1 日至 93 年 7 月 31 日

計畫主持人：簡國龍

共同主持人：許秀卿, 李源德

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：臺灣大學公衛學院預防醫學研究所

中 華 民 國 93 年 10 月 31 日

# 行政院國家科學委員會專題研究計畫成果報告

## 高三酸甘油酯 / 低高密度脂蛋白膽固醇之家族遺傳研究 (3/3)

### Family genetic studies of hypertriglyceridemia and low HDL-cholesterol (3/3)

計畫編號：NSC 92-2314-B-002-197

執行期限：92年8月1日至93年7月31日

主持人：簡國龍 執行機構及單位名稱 臺灣大學公衛學院預防醫學研究所  
共同主持人：許秀卿, 李源德 執行機構及單位名稱 臺大醫院 內科

#### 一、中文摘要

本年度計畫則是以金山社區的 Triglyceride/HDL-cholesterol 當作一個連續性性狀，在 ApoA1/C3/A4 基因簇集中找到可能與之相關的控制基因，在基本家族資料成員，在所有 456 個家族，共 4175 位家族成員(包括 1436 位 founders 及 2739 位 non-founders)，家族大小平均為 9.16 人(範圍 1~67)，有 26.8% 為 7 人，有 16.9% 為 6 人，有 14.0% 為 8 人，共 2.78 代(範圍 1~5 代)，有 68.9% 為 3 代，有 22.6% 為 2 代，有 5.0% 為 4 代，而真正有基因型資料，約為 20%，而異體基因 (heterozygosis) 約有 40% 左右。參數性連鎖分析可用在以下幾種狀況：藉由估計互組事件的數目來評估基因標識與疾病相關位址的基因距離，如果組合率(Recombination fraction,  $r$ ) 已知，可進一步在基因地圖上作排序，針對常見疾病找出其特定的基因型式。本研究資料結果顯示，其最高 LOD Score 分別發生在組合率( $r$ ) 不同的狀況下，其估計的 full informative meiosis 相等數量分別為 0.1197( $r=0.1$ )，也即是資料提供的訊息量等同於 0.1197 個 full informative meiosis。結果進一步顯示其最大 LOD Score 是在 0.1 之間，雖然結果並無法達到統計上有意義的 LOD Score，但了解其變化及不同的家族供 lod score 有其必要性。作定量性狀遺傳分析，有基因標記之下的流程：先以 Merlin 程式作錯誤偵測程式方式，特別針對是否符合 Mendel 遺傳定律，在每個 SNP 簇集檢查，以最大概似值方法去偵測某些基因型明顯此預期的概似值為大，並且比較以偵測出前 Error 及之後的資

料跑 Merlin，可看到 mismatched 資料之後可得到較高的 LOD Score(在迴歸模型中 LOD Score -0.035 增加為 0.827，在變異數成份模型中 LOD Score 由 0.60 (ApoA75) 變成 0.86，在迴歸模型中 ApoA25 LOD Score 也變大，可以增加其檢定效力，對偵測可能的基因位址有較大成功的機率。

**關鍵詞：**三酸甘油酯/高密度脂蛋白膽固醇、連鎖分析、定量性狀位址分析法

#### Abstract

The basic description of family data was following; Individuals: 4175 (1436 founders, 2739 nonfounders), and total families: 456. The family sizes were average: 9.16 (1 to 67), and the distribution patterns of family were 7 (26.8%), 6 (16.9%) and 8 (14.0%). The generations were average: 2.78 (1 to 5). The estimate of residual nonshared environmental variance is 0.324, and the estimate of polygenic component of variance was 0.192. Thus, the polygenic heritability of the trait,  $LogTG/HDL$  is estimated as  $0.192/(0.192+0.324)=0.372$ . About 37% of the total variance can be explained by a polygenic component, suggesting that the trait is highly heritable. The results of testing each marker locus for linkage to an additive major gene influencing the trait,  $Log TG/HDL$  were not significant as following table. When we tested for association, there is a polygenic effect and that this may map to the locus where our markers are located, so we must include additive components in the null model. The tests of variance are less powerful than the tests of association, it is

possible for an effect to be significant component of the mean but not a significant component of variance. We used orthogonal test of Abecasis et al to evaluate if within-family component of association is significant, and it is robust to stratification. The results showed the most significant result refers to: Marker: ApoA4 (ChiSq: 3.384, p-value: 0.0658). The parametric linkage analysis by LODLINK program did not demonstrated high LOD score among 5 markers.

**Keywords:** Triglyceride/HDL cholesterol, Linkage analysis, Quantitative trait locus,

## 二、緣由,目的,結果

本年度計畫則是以金山社區的 Triglyceride/HDL-cholesterol 當作一個連續性性狀,在 ApoA1/C3/A4 基因簇集中找到可能與之相關的控制基因,在基本家族資料成員,在所有 456 個家族,共 4175 位家族成員(包括 1436 位 founders 及 2739 位 non-founders),家族大小平均為 9.16 人(範圍 1~67),有 26.8% 為 7 人,有 16.9% 為 6 人,有 14.0% 為 8 人,共 2.78 代(範圍 1~5 代),有 68.9% 為 3 代,有 22.6% 為 2 代,有 5.0% 為 4 代,而真正有基因型資料,約為 20%,而異體基因 (heterozygosis)約有 40% 左右。

定量性狀位址分析法(QTL, quantitative trait locus)在最近幾年有突破性的發展,不僅可用在參數模型下的連鎖分析(即 SAGE 程式中 LOKLINK),也可用在大規模家族中的資料。以非參數模型下,一種是以迴歸為基礎的分析方法由基礎的 Haseman-Elston algorithm(計算式)延伸出 Merlin 程式計算,另一種是為變異數成分為基礎的方法,即利用 Covariance structure 的相關結構,此乃是利用 SOLAR 或 Merlin 程式軟體來處理。同時也可運用傳遞連鎖不平衡 TDT(Transmission Disequilibrium

test)的家族相關研究資料,即利用親子三合體(parent-trios)的設計研究基因對連續變數之性狀。

### (一) QTDT 方法:

TDT 程式是同時考慮連鎖及相關的分析兩個基因座連鎖不平衡代表相關的強度,在一隔離的族群,連鎖不平衡距離可高達到 5 cM,但一般族群可能只有幾個 base pair 而已,但族群分層(population stratification)會造成相關的假象,如果資料是來自家族成員,則可避免此分層效果的干擾,而傳統的 TDT 是以有病的子代與雙親的基因型作一配對設計,並且可擴展到多數個位址的狀況。傳統的 TDT 有幾個限制,首先,只能一次處理一個基因位址,其次,此一基因位址必須是對偶基因(di-allelic),最後,只利用家族中有病的小孩的資料,無法充分運用全部的資料。

在作連續變數性狀的 TDT 檢定,有 Allison, Rabinowitz, Sib-TDT 及 Abecasis 等不同模型建立的方法。Allison 提出的 5 種檢定量其中以  $TDT_{Q5}$  最為有用,同時可處理雙親均是異質體(heterozygote),同時可放入其他變數的效果,如年齡、種族等資料,可增加檢定的效力。Rabinowitz 的方法不需參數的假定,也可處理多重對偶基因型的狀況,而 sib-QTDT 則只處理手足的資料,不需要雙親的狀況,可利用混合作模型(mixed effects model)及數列置換排列檢定(permutation test)。

而 Fulker et al.發展的模型在 Abecasis et al.利用 Variance component 方法檢定相關及連鎖強度。此方法可利用所有的家族成員資料,同時符合迴歸程式的直覺,為一概似值為基礎的模型,因此運用上較為廣泛。同時也可處理其他變項的影響,也可處理大家族或多發家族資料。

Allison's 檢定只能用在核心家族資料,因此不適合本研究資料,而 Rabinowitz 方法並不容許有 dominance effects,而 sib-TDT

只利用 full sibs 的資料，而 Abecasis 方法必須在多變數常態分佈之下，因此需將性狀以 logarithm 對數作轉換以符合常態分佈的假設。Abecasis 方法是利用虛無假設模型及相對假設模型的比較，進一步放入各種參數，如變項，imprinting，dominance 等，以 stepwise 方式尋找最佳的模型。QDT 是目前最常用來分析連續性狀 TDT 的軟體。

由於本研究資料為大家族多代資料，因此以 Abecasis 方法估計

- (1) 首先是作遺傳率的評估
- (2) 作變異數成份(variance component)的分析
- (3) 作相關部分的分析

以  $\log(\text{TG}/\text{HDL})$  為研究性狀結果如下：利用非參數連鎖分析主要有二種方法，一則是利用 Haseman-Elston 計算式(1972)，以手足配對的研究下手，以線性模型處理連續變數在手足之間的差異與其 IBD 之間的迴歸模型。另一種則是利用變數成份 (variance component) 分析方法，乃是利用家族成員之間連續變數的變異數區隔的部分，包括由基因控制的部分與其他變數的部分。

利用 VC 方法可避免家族成員內互相相關的困境，但必須符合多變數常態分佈的假設，因此事先檢查資料是很重要。在本研究中以對數轉換來改善其常態分佈的狀況，

(二)參數連鎖分析連續變數的方法 (Parametric linkage analysis), 乃是模型基礎 (model-based) 的連鎖分析, 在連續性狀的參數性連鎖分析可設定各種參數, 如基因頻率, 基因平均值及變異數, 這些參數可由複雜分離分析得到估計數值, 而此模型基礎分析是假設基因位址同時對表現型有主要基因作用及符合孟德爾氏遺傳定律, 此種方法可得到較敏感的 LOD 分數, 但結果較不堅固(Robust)

參數性連鎖分析可用在以下幾種狀況：

1. 藉由估計互組事件的數目來評估基因標識與疾病相關位址的基因距離
2. 如果組合率(Recombination fraction,  $r$ ) 已知, 可進一步在基因地圖上作排序
3. 針對常見疾病找出其特定的基因型式其連鎖統計有意義程度的 P 值常在  $10^{-4}$ , 相當於 LOD Score 為 3, 使其假陽性機率在 5%, 至於其參數的估計, 已由第一年計畫演算出來。

在軟體方法, 雖然傳統的 LINKAGE 程式可以處理連續參數, 但必須利用 2 個常態成份作混合分析估計, 同時設定其易感度 (liability), 目前已被證實統計效力比預期的低。因此目前以 SAGE 軟體中的 LODLINK 程式為最主要使用的軟體。LODLINK 是利用基因型/基因相態消除計算式

(genotype/phase elimination algorithm), 在檢查是否符合孟德氏遺傳定律之後, 進一步估計連續性狀與基因標記之間兩點連鎖分析, 進而得到 LOD Score, 本研究資料結果顯示如下, 其最高 LOD Score 分別發生在組合率( $r$ )不同的狀況下, 其估計的 full informative meiosis 相等數量分別為 0.1197( $r=0.1$ ), 也即是資料提供的訊息量等同於 0.1197 個 full informative meiosis。結果進一步顯示其最大 LOD Score 是在 0

1 之間, 雖然結果並無法達到統計上有意義的 LOD Score, 但了解其變化及不同的家族供 lod score 有其必要性。作定量性狀遺傳分析, 有基因標記之下的流程：

先以 Merlin 程式作錯誤偵測程式方式 (Abecasis et al 2002), 特別針對是否符合 Mendel 遺傳定律, 在每個 SNP 簇集檢查, 以最大概似值方法去偵測某些基因型明顯此預期的概似值為大, 並且比較以偵測出前 Error 及之後的資料跑 Merlin, 可看到 mismatched 資料之後可得到較高的 LOD Score(在迴歸模型中 LOD Score -0.035 增加為 0.827, 在變異數成份模型中 LOD Score

由 0.60 (ApoA75)變成 0.86 , 在迴歸模型中 ApoA25 LOD Score 也變大, 可以增加其檢定效力, 對偵測可能的基因位址有較大成功的機率。

## 五、參考文獻

- (1) Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; 62(5):1198-1211.
- (1) Allison DB. Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 1997; 60(3):676-690.
- (2) Clayton D, Jones H. Transmission/Disequilibrium Tests for Extended Marker Haplotypes. *Am J Hum Genet* 1999; 65(4):1161-1169.
- (3) Statistical analysis for genetic epidemiology, Release 3.1. Computer program package available from the Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland. 1997.
- (4) Sham PC, Curtis D. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Annals of Human Genetics* 1995; 59(Pt 3):323-336.
- (5) Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *Am J Hum Genet* 1993; 52:506-516.