

發明專利說明書 200526789

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※ 申請案號：93129844

※ 申請日期：93.10.1

※IPC 分類：C12Q 1/68,

一、發明名稱：(中文/英文)

G01N 33/574

有關乳癌診斷之材料及方法

MATERIALS AND METHODS RELATING TO BREAST CANCER
DIAGNOSIS

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

新加坡商NCC科技企業公司

NCC TECHNOLOGY VENTURES PTE LIMITED

代表人：(中文/英文)

鄭子維

TAY, NICHOLAS

住居所或營業所地址：(中文/英文)

新加坡新加坡市醫院大道11號

11 HOSPITAL DRIVE, 169610, SINGAPORE, REPUBLIC OF
SINGAPORE

國 籍：(中文/英文)

新加坡 SINGAPORE

三、發明人：(共 2 人)

姓 名：(中文/英文)

1. 余昆
YU, KUN
2. 陳文煒
TAN, PATRICK

國 籍：(中文/英文)

1. 中國大陸 PEOPLE'S REPUBLIC OF CHINA
2. 新加坡 SINGAPORE

四、聲明事項：

☐ 主張專利法第二十二條第二項 ☐ 第一款或 ☐ 第二款規定之事實，其事實發生日期為： 年 月 日。

☒ 申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

☒ 有主張專利法第二十七條第一項國際優先權：

1. 英國；2003年10月03日；0323226.1

2.

☐ 無主張專利法第二十七條第一項國際優先權：

1.

2.

☐ 主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

☐ 主張專利法第三十條生物材料：

☐ 須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

☐ 不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

本發明係關於有關乳癌診斷之材料及方法。詳言之，本發明係關於"低置信度"腫瘤之診斷及/或分類，該等低置信度腫瘤與其"高置信度"對應物相比展現出顯著惡化之整體存活率及更短的遠距離轉移時間。

【先前技術】

現已強烈關注使用基因表現資料來進行生物學分類，尤其是在腫瘤學及醫學領域。此方法之一振奮人心的態樣在於其界定先前藉由較傳統的光學顯微鏡方法(15,16)無法確定之臨床相關亞型癌症之能力。儘管有此潛能，但在基因表現資料用於臨床診斷變成現實之前仍須解決大量問題。舉例而言，需要建構多種演算法，該等演算法除了可傳遞正確的分類外，亦可精確地確定預測之置信度。若分類會影響隨後的治療過程，則此尤為重要；若治療醫師擁有此資訊，則其可衡量有關特定干涉之潛在發病的預測之置信度以作出有根據(informed)的臨床選擇。

將乳腺腫瘤分類為雌激素受體陽性(ER+)及陰性(ER-)亞型係乳癌治療中的一項關鍵性區分。一般而言，ER-腫瘤比其ER+對應物更具臨床侵襲性，且ER+腫瘤通常係使用諸如他莫昔芬(tamoxifen)之抗激素療法進行治療(1)。目前，腫瘤之ER狀況通常係利用ER之抗體藉由免疫組織化學法(IHC)或免疫轉漬法來判定。然而，此技術並不完善，例如，其可能無法偵測出隱藏ER中之基因改變從而使其呈非活性

或構成上活性之腫瘤(2)。因此，至關重要的是發展更精確的方法論以改良乳腺腫瘤之ER亞型分類，從而可隨後施加恰當的治療。目前，許多小組已發表了有關利用表現輪廓資料來將乳癌分類為ER+及ER-類別的報導。在一研究中，已發現ER+及ER-腫瘤之表現輪廓"顯著不同"，此支持先前之理論，即ER+及ER-腫瘤可產生於截然不同的乳上皮細胞類型(3)。

另一小組已報導了使用關於表現資料之監督式研究方法論藉由ER亞型分類乳腺腫瘤(4)。該等研究中的一項共同觀察為：儘管通常可以高可信度將大部分乳腺腫瘤精確地分類為ER+及ER-亞型，但總是存在一組"低置信度"樣品，其被錯誤地歸類或其預測之統計學"置信度"處於邊緣。儘管已提出該等"低置信度"樣品可能反映種群異質性效應(4)，但此等"低置信度"樣品可能在生物學上截然不同於其"高置信度"對應物之假設至今仍未得到完全的探究。

【發明內容】

本發明者考慮到"低置信度"樣品可能具有獨特的生物學特徵之可能性。為了對此進行評定，其使用室內(in-house)所產生的乳癌表現資料集來進行分類分析，且判定：相較於"高置信度"腫瘤，"低置信度"腫瘤在對於ER亞型判別而言重要的多種基因之表現中展現出廣泛的擾動。儘管"高"與"低"置信度腫瘤之間的區分最初係藉由純計算方式得出，但其仍具有臨床意義，因為"低置信度"腫瘤與其"高置信度"對應物相比展現出顯著惡化之整體存活率($p=0.0003$)

及更短的遠距離轉移時間($p=0.0001$)。目前，此區分無法藉由用於偵測ER之習知免疫組織化學策略來辨識。

本發明者已令人驚奇地進一步判定ERBB2受體之高表現量與展現"低置信度"預測之乳腺腫瘤顯著相關，且在自不同患者群/陣列技術所產生的三個獨立獲得之乳癌表現資料集上驗證實了此相關性，且使用不同計算方法進行了分析。ERBB2表現與"低置信度"腫瘤中所觀察到的ER判別基因的廣泛擾動之間的關聯性令人感興趣，因為已知ERBB2活性在乳腺腫瘤及細胞株中均有助於發展對於抗激素治療之抗性(5, 6)，且有助於抑制ER之轉錄活性(5, 7)。

然而，儘管該等"擾動"基因對於ER亞型判別而言很重要，但發明者發現顯著比例的該等"擾動"基因並不具有雌激素反應性，且利用最近所描述的生物資訊學演算法(DEREF)亦證實該等基因在其啟動子中並不含有潛在的雌激素反應要素(estrogen-response element, ERE)。該等結果表明：除了其中ERBB2主要藉由破壞ER之轉錄活性而起作用的現有模型外，ERBB2對乳腺腫瘤之效應中的顯著部分亦可能涉及基因活化之ER-無關機制，其可總體上有助於"低置信度"乳腺腫瘤亞型之臨床侵襲本性。

因此，本發明者已確定用於將乳腺腫瘤樣品分類為"低置信度"腫瘤或"高置信度"腫瘤之基因集("多基因分類器")。發明者首次確定"低置信度"腫瘤組在預後及治療方面具有重大的醫學意義。

對於ER+及ER-中之每一者，發明者已提供大量已改變

"高置信度"與"低置信度"腫瘤之間的表現量之基因。該等基因如表2中所識別。該等擾動基因之表現量可用於判別高置信度與低置信度腫瘤。表S4中識別了另一基因集，相較於高置信度腫瘤，該等基因在低置信度腫瘤中具有不同的表現量。表A1-A4中識別了其它基因集，相較於高置信度腫瘤，該等基因在低置信度腫瘤中具有不同的表現量，而與腫瘤之ER狀況無關。隨後之描述將利用術語"表現輪廓"。此係指來自多基因分類器之基因集之樣品中的表現量。

表現量一般以數字方式表示。因此，表現輪廓一般包括一組數字，每一數字代表一多基因分類器之一基因的表現量。隨後之描述將利用術語"複數種基因"。此術語係指來自一多基因分類器之基因的子集。子集可對應於多基因分類器之子群，例如在ER+低置信度乳腺腫瘤中之上調基因。複數種基因之含量可隨多基因分類器而不同，且對於一特定多基因分類器而言，可隨本發明之不同態樣而不同。此術語可意指一特定多基因分類器之全部基因或其子集。

因此，最一般而言，本發明提供使用多基因分類器來將乳腺腫瘤樣品分類為高或低置信度樣品之新穎診斷方法及檢定。本發明進一步識別了用於分類乳腺腫瘤樣品之多基因分類器及包含一多基因分類器或來自該多基因分類器之複數種基因的裝置。用於本發明之各態樣中的多基因分類器展示於表S4、2、A1、A2、A3及A4中。

表S4列出了當在全面規模上檢查ER+及ER-腫瘤中之每

一者時展現出高置信度與低置信度腫瘤之間顯著差異轉錄調節之基因。

在第一態樣中，本發明提供一種產生乳腺腫瘤樣品之核酸表現輪廓之方法，其包含以下步驟：

- (a)將表現產物與該乳腺腫瘤樣品分離；
- (b)識別選自表S4之複數種基因的表現量；及
- (c)自該等表現量產生該乳腺腫瘤樣品之表現輪廓。

腫瘤樣品可具高置信度及/或低置信度。腫瘤樣品可為ER+高置信度乳腺腫瘤樣品及/或ER+低置信度乳腺腫瘤樣品及/或ER-高置信度乳腺腫瘤樣品及/或ER-低置信度乳腺腫瘤樣品。較佳判定乳腺腫瘤樣品之ER狀況。較佳在該方法之步驟(a)之前判定乳腺腫瘤樣品之ER狀況。乳腺腫瘤樣品之ER狀況可利用如吾人同在申請中之申請案PCT/GB03/000755中所描述的基因表現輪廓分析來判定。

表S4之基因以子集形式展示。相較於ER+低置信度腫瘤，子集(a)中的基因在ER+高置信度樣品中展示出顯著改變的表現。表S4(a)之第一部分係一組相較於ER+高置信度腫瘤在ER+低置信度腫瘤中上調(表S4(a)"上調")之基因。表S4(a)之第二部分展示一組相較於ER+高置信度腫瘤在ER+低置信度腫瘤中下調(表S4(a)下調)之基因。

表S4之(b)部分係相較於ER-高置信度腫瘤在ER-低置信度樣品中展示上調表現之基因。

多基因分類器之各基因的表現輪廓將在獨立樣品之間稍有不同。然而，發明者已認識到：多基因分類器之基因的

表現輪廓提供了在高置信度與低置信度腫瘤之間可識別地不同的特徵表現模式。

藉由自大量已知的高及低置信度樣品創建多基因分類器之大量表現輪廓，可創建高置信度及低置信度樣品之輪廓庫。表現輪廓之數目愈多，則愈容易創建可用作診斷檢定中的對照物之可靠特徵表現輪廓標準(即，包括統計學變化)。因此，標準輪廓可為自複數種個別表現輪廓獲得且在統計學變化內導出以代表高置信度或低置信度樣品輪廓之輪廓。

因此，根據本發明第一態樣之方法可包含以下步驟：

(a)將表現產物與乳腺腫瘤樣品分離；

(b)使該等表現產物與能夠特異且獨立地結合至選自表S4之複數種基因的表現產物之複數個結合組員接觸，以自該等複數種基因之表現量創建腫瘤樣品之第一表現輪廓；

(c)將該表現輪廓與作為高置信度腫瘤及/或低置信度乳腺腫瘤之特徵的表現輪廓相比較。

評定複數種基因之表現量以產生表現輪廓。可絕對地評定表現量，即，對經表現產物的量之量測。可相對地評定表現量，即相較於某些其它因素的表現，該等因素為例如(但不限於)另一基因之表現或該樣品中或一組樣品中的一組基因(較佳為未包括在該方法中所用之多基因分類器中之一組基因)之表現的平均值/中值/模式。舉例而言，基因之表現可根據樣品中複數種基因之平均表現的倍數或分數來度量。表現較佳表示為正或負，以指示表現相對於平均

值的增加或減少。

預測強度較佳使用統計學及/或概率學模型來量測。該模型包含加權表決法(WV)及/或支持向量機。預測強度可使用加權表決法及留一校驗法(參看實例)來確定。當使用2-色cDNA微陣列(例如,彼等用於評定Stanford資料集之微陣列)進行計算時,低置信度可意指量值小於或等於0.4的預測強度。較佳地,低置信度腫瘤之預測強度的範圍為 ≥ -0.4 ,且較佳 ≤ 0.4 。低置信度腫瘤之預測強度可 ≥ -0.35 ,且較佳 ≤ 0.35 。低置信度腫瘤之預測強度可 ≥ -0.3 ,且較佳 ≤ 0.3 。

高置信度樣品較佳具有大於0.4量值之預測強度。較佳地,高置信度腫瘤之預測強度為 ≥ 0.4 ,且較佳 ≤ -0.4 。

然而,高/低置信度腫瘤之預測強度的截止值可隨所使用之資料集及/或陣列技術而不同。舉例而言,在Rosetta資料集中,使用2色寡核苷酸微陣列進行評定時,高置信度腫瘤係彼等具有大於0.7之預測強度量值的腫瘤。高置信度樣品較佳具有大於0.7之預測強度的量值。因此,對於低置信度腫瘤而言,預測強度可 ≥ -0.7 ,且較佳 ≤ 0.7 。對於低置信度腫瘤而言,預測強度可 ≥ -0.6 ,且較佳 ≤ 0.6 。對於低置信度腫瘤而言,預測強度可 ≥ -0.5 ,且較佳地 ≤ 0.5 。更佳地,低置信度腫瘤之預測強度範圍係 ≥ -0.4 ,且較佳 ≤ 0.4 。

當在Stanford與Rosetta資料集中比較乳腺腫瘤群之預測強度時,高與低置信度腫瘤之間的邊界可識別為某些點,於該等點處,資料集中腫瘤的預測強度開始定性地顯示出自該腫瘤群中大部分預測強度減少之預測強度("懸崖點")

(cliff point))。儘管對每一資料集進行獨立分析，但是對於獨立的Rosetta及Stanford資料集而言，低置信度腫瘤所占比例相似。

因此，低置信度腫瘤可落在乳腺腫瘤群中ER預測強度之最低20%內，且更佳落在ER預測強度之最低15-19%內。乳腺腫瘤群較佳包含至少25種腫瘤之最小資料集，更佳為至少25-30種腫瘤，更佳為至少30種腫瘤，更佳為至少50種腫瘤，更佳為至少80種腫瘤，且最佳為約80-100種腫瘤。

表現產物較佳為mRNA、或由該mRNA所製成之cDNA、或cRNA。或者，表現產物可為經表現之多肽。表現輪廓之識別較佳係使用能夠特異地識別表S4中所識別之複數種基因的表現產物之結合組員來進行。例如，若表現產物為cDNA，則結合組員將為能夠特異地與該cDNA雜交之核酸探針。

較佳標記表現產物或結合組員以使得可偵測該等兩種組分之結合。較佳將標記選擇成能夠偵測經表現產物之相對含量/數量及/或絕對含量/數量以基於多基因分類器之個別基因的上調或下調來確定表現輪廓。一般而言，結合組員應不僅能夠偵測表現產物之存在，而且亦能夠偵測其相對豐度(意即，可獲取產物之量)。

然而，目前已出現大量利用"無標記"定量化技術之新技術，例如彼等由Xagros產生之技術。表現產物及/或結合組員可不作標記。與結合組員之結合可藉由量測由兩個引子停靠至目標表現產物上及隨後由聚合酶延伸所引起的電阻

變化來偵測及/或定量化。

核酸表現輪廓之確定可於某些先前所設定之參數內進行，以避免假陽性及假陰性。可使用電腦來確定核酸表現輪廓。

隨後，電腦能夠提供作為上文所討論之低置信度或高置信度乳腺細胞之特徵的表現輪廓標準。隨後，可使用所確定之表現輪廓來分類乳腺組織樣品，以作為一種診斷方法。

因此，在本發明之第二態樣中，提供一種包含高置信度及/或低置信度乳腺腫瘤樣品之複數種基因表現輪廓的表現輪廓資料庫，其中每一基因表現輪廓均係得自選自表S4之複數種基因，且其中該資料庫可擷取地保存於資料載體上。組成該資料庫之表現輪廓較佳係根據第一態樣之方法產生。

基於對多基因分類器的認識，可想出許多用於確定一特定測試樣品中之基因的表現模式或輪廓之方法。例如，可使用標準的分子生物學技術自樣品分離經表現之核酸(RNA、mRNA)。接著，可使用專用於PCR中經表現之序列的核酸引子來擴增對應於來自表S4中所給定的基因識別器之該等複數種基因的經表現之核酸序列。若所分離的表現核酸為mRNA，則其可使用標準方法轉化成cDNA以用於PCR反應。

引子可便利地將標記引入經擴增之核酸中以使得其可被識別。理想地，標記能夠指示擴增事件後所存在的核酸序列之相對數量或比例，以反映原始測試樣品中所存在的相

對數量或比例。例如，若標記具螢光性或放射性，則訊號之強度將指示經表現序列之相對數量/比例或甚至絕對數量。每一基因識別器之表現產物的相對數量或比例將建立測試樣品之特定表現輪廓。藉由將此輪廓與已知輪廓或標準表現輪廓相比較，可判定測試樣品是來自正常的乳腺組織或是來自惡性乳腺組織。如上文所述，引子及/或擴增核酸可不作標記。

或者，表現模式或輪廓可使用能夠結合至基因識別器之表現產物(例如，對應於cDNA或經表現之多肽的mRNA)的結合組員來確定。藉由標記表現產物抑或結合組員，可識別表現產物之相對數量或比例，且可確定基因識別器之表現輪廓。以此方式，可藉由將該表現輪廓與已知輪廓或標準輪廓進行比較而將樣品分類為高置信度或低置信度。結合組員可為補充核酸序列或特異抗體。下文將更具體論述使用此類結合組員之微陣列檢定。

在本發明之第三態樣中，提供一種用於將乳腺腫瘤樣品分類為低置信度或高置信度之方法，該方法包含：提供該乳腺腫瘤樣品之表現輪廓，其中該表現輪廓包含來自表S4之複數種基因的表現量；及基於該表現輪廓將腫瘤分類為高或低置信度腫瘤。

本發明之第三態樣的方法可包含如下步驟：

- (a)獲得來自由患者所獲得之乳腺腫瘤樣品之表現產物；
- (b)藉由使該等表現產物與結合組員相接觸來確定表S4中所識別之複數種基因的表現量，其中每一結合組員均能

夠特異地結合至該等複數種基因之表現產物；及

(c)基於該等表現量來識別該患者中低置信度乳腺腫瘤之存在。

該方法較佳進一步包括較佳在提供腫瘤的表現輪廓之前判定腫瘤之ER狀況的步驟。

確定低置信度乳腺腫瘤存在之步驟可藉由電腦來執行，該電腦能夠將來自測試中之乳腺腫瘤樣品之表現產物的結合輪廓與其它先前所獲得之輪廓及/或作為低置信度腫瘤存在之特徵的先前所確定之"標準"輪廓的資料庫進行比較。可將電腦程式化以報導測試中之輪廓與標準輪廓之間的統計學相似性，從而可進行分類。

分類乳腺腫瘤樣品之步驟可包含使用統計學及/或概率學技術，諸如加權表決法(WV)(13)，其為一種監督式研究技術。在WV中，可執行二分法。將乳腺腫瘤樣品之多基因分類器中的基因之表現量與不同類別中彼基因的平均表現量進行比較。此平均值可(例如)自具有指定類別之表現輪廓計算獲得，例如自高及/或低置信度樣品之表現輪廓的資料庫。較佳地，該等輪廓具有指定的ER狀況。

表現量與若干類別的平均基因表現之間的差值被加權，且其對應於一特定類別之彼基因的"表決(vote)"。對於特定腫瘤而言，將每一類別之所有基因的表決加在一起以創建每一類別之總數。將該腫瘤指定為具有最高表決數之類別。隨後，可將獲勝類別之勝利的邊緣表示為預測強度。

使用包括兩種類別中每一類別的基因之表現量的平均值

及標準偏差之式子來對表現量差值進行加權。一般而言，每一類別之平均值及標準偏差係自具有或代表特定類別腫瘤(例如，高置信度及低置信度)的表現輪廓計算獲得。

此外或替代性地，步驟(c)可包含階層式叢集法之使用，尤其是在腫瘤樣品已使用與用於評定具有指定類別之表現輪廓或與樣品表現輪廓相比較之標準輪廓的技術不同的陣列技術進行評定時。步驟(c)之結果可使用既定的留一校驗(LOOCV)檢定法來驗證(參看實例)。步驟(c)可使用電腦來進行。

在階層式叢集法中，每一表現輪廓均可表示為由 n 種基因組成之向量，其中 $(g_1, g_2 \dots g_n)$ 表示該等基因之表現量。隨後，將每一向量與分析中的所有其它輪廓進行比較，且將彼此具有最高相關性之兩個向量配對，直至將分析中盡可能多的輪廓配對。

此項技術中已知存在許多計算相關性之方法，諸如皮爾遜相關係數法(Pearson's correlation coefficient)(28)。接著，在下一步驟中，自每一對導出合成向量(在平均連接叢集法中，此通常為兩種輪廓之平均值)，且隨後重複配對過程。此持續進行直至無法配對。此方法為"階層式"，因為其自底部(個別輪廓)開始且逐步建立。在本發明中，個別輪廓逐步增加至較佳兩個合成向量，每一向量代表一類別(意即，高置信度及低置信度)。對於未知類別之新樣品而言，該樣品與標準輪廓/樣品叢集。"未知"樣品之類別將基於在反覆巡迴配對的末端其所屬之叢集/向量來判定。

因此，在一實施例中，本發明提供一種識別患者中侵襲性乳腺腫瘤之方法，例如藉由將該腫瘤之表現輪廓與作為腫瘤類別之特徵的輪廓相比較，較佳藉由將該腫瘤之表現輪廓與作為高置信度及/或低置信度腫瘤之特徵的輪廓進行比較。該方法可進一步包含為其中腫瘤具有作為低置信度腫瘤表現輪廓之特徵之表現輪廓的患者指定不良預後之步驟。

此預後可能影響患者的治療過程。在識別該低置信度腫瘤後，可使用治療低置信度腫瘤之侵襲性技術來治療該患者。

不良預後包括相較於患有高置信度腫瘤之患者顯著惡化之整體患者存活率及/或顯著更短的遠距離轉移時間。

如上文所提及，本發明者已識別出相對於高置信度腫瘤在低置信度乳腺腫瘤中具有不同表現模式之若干種關鍵基因，意即其能夠區分高及低置信度類別之乳腺腫瘤。

多基因分類器可包含表S4中所給定之基因。藉由確定測試樣品之表現輪廓且將該表現輪廓與作為低及/或高置信度乳腺腫瘤之特徵之表現輪廓進行比較(及/或使用諸如加權表決之技術來分析該表現輪廓)，可將樣品分類為低置信度或高置信度腫瘤，例如相對於標準模式或高置信度樣品中所見之輪廓，其表現增加或減少。

該等複數種基因可為表S4(a)及/或表S4(b)之基因，或表S4(a)之基因之子集及/或表S4(b)之基因之子集。

該等複數種基因可包括至少10、20、30、40、50、60、

70、80種或全部的表S4(a)之基因。

該等複數種基因可為來自表S4(a)之全部或大體上全部的上調及/或下調基因。該等複數種基因可包含來自表S4a之約30、或約20、或約10、或約5種上調基因，或由上述基因組成。該等複數種基因可包含來自表S4a之約30、或約20、或約10、或約5種下調基因，或由上述基因組成。

較佳地，該等複數種基因包含來自表S4(a)之約80、或約70、或約60、或約50、或約40、或約30、或約20、或約10種基因，或由上述基因組成。該等複數種基因可包含來自表S4(a)之約50、或約40、或約30、或約20或約10或約5種上調基因，或由上述基因組成。

來自表S4(a)之基因較佳係選自上調基因組之上部分及/或下調基因組之上部分。因為每一組中的基因係按顯著性的次序排列，所以上部分較佳為表或組的上半部分。在高置信度與低置信度腫瘤之間展示最大差異表現的基因出現在表S4(a)中每一組的上部分，而彼等以較小差異表現之基因則出現在下部分。

該等複數種基因可包括表S4(a)中不多於80、或70、或60、或50、或40、或30、或20、或10、或5種基因。

該等複數種基因可包含表S4(a)中上調及/或表S4(a)中下調之5至30種基因，或基本上由上述基因組成。該等複數種基因可包含表S4(a)中上調及/或表S4(a)中下調之10至30種基因，或基本上由上述基因組成。該等複數種基因可包含表S4(a)中上調及/或表S4(a)中下調之10至20種基因，或表

S4(a)中上調及/或表S4(a)中下調之20至30種基因，或基本上由上述基因組成。該等複數種基因可包含表S4(a)中上調之5至40種、或5至50種基因，或基本上由上述基因組成。

該等複數種基因可為約10種基因，其可選自表S4(a)中上調及/或表S4(a)中下調之約前40、或約前30、或約前20種基因。該等約10種基因可選自表S4(a)中上調及/或表S4(a)中下調之約前15種基因。該等約10種基因可為表S4(a)中上調及/或表S4(b)中下調的前10種基因。該等複數種基因可為約10種基因，其可選自表S4(a)中上調之約前50、或約前40種基因。

較佳地，該等複數種基因包含表S4(a)中上調及/或表S4(a)中下調之約前30種基因中的約10至20種基因。

該等複數種基因可包含選自由表S4(a)中上調之約前40、或約前30、或約前20或約前10種基因及表S4(a)中下調之約前30、或約前20、或約前10種基因所組成之群的約30、或約20、或約10種基因，或由上述基因組成。該等複數種基因可包含選自由表S4(a)中上調之約前10、或15種基因及表S4(a)中下調之約前10、或15、或約20種基因所組成之群的約10、或約15、或約20種基因，或由上述基因組成。

該等複數種基因可為來自表S4(b)之全部或大體上全部的基因。該等複數種基因可為來自表S4(b)之全部或大體上全部的基因。該等複數種基因可包括表S4(b)之至少10、20、30、40、50或全部的基因。

該等複數種基因可包含來自表S4(b)之約50、或約40、或

約30、或約20、或約10或約5種基因，或由上述基因組成。

來自表S4(b)之基因較佳係選自表的上部分。因為每一組中的基因係按顯著性次序排列，所以該上部分較佳為表的上半部分。在高置信度與低置信度腫瘤之間展示出最大差異表現之基因出現在表S4(b)之上部分，而彼等以較小差異表現之基因出現在下部分。

該等複數種基因可包括表S4(b)之不多於50、或40、或30、或20、或10、或5種基因。

該等複數種基因可包含表S4(b)之5至50種基因，或基本上由上述基因組成。該等複數種基因可包含表S4(b)之10至40種基因，或基本上由上述基因組成。該等複數種基因可包含表S4(b)之10至30種基因，或基本上由上述基因組成。該等複數種基因可包含表S4(b)之10至20種基因或表S4(b)之20至30種基因，或基本上由上述基因組成。

該等複數種基因較佳為約30、或約20、或約10種基因，其可選自表S4(b)之約前40、或約前30、或約前20種基因。約10種基因可選自表S4b之約前15或20種基因。該等約10種基因可為表S4b的前10種基因。

較佳地，該等複數種基因包含表S4(b)的前約30種基因中之約10至20種基因。

如先前所論述，熟習此項技術者將瞭解：與產生特徵性表現輪廓所需之最小顯著性基因的數目相比，產生特徵性表現輪廓需要較少的最大顯著性基因。

將該等複數種基因之數目及選擇選定成提供能夠在高置

信度與低置信度腫瘤之間加以區分之表現特徵。

較佳地，該等複數種基因包括來自表S4(a)及/或表S4(b)之上調及下調基因的混合基因。

將腫瘤分類之步驟可包含評定相較於高置信度腫瘤在低置信度腫瘤中上調之基因。

此外或替代性地，步驟(c)可包含評定相較於高置信度腫瘤在低置信度腫瘤中下調之基因。

組成另一多基因分類器之基因展示於表2中。本發明之第一、第二及第三態樣對表2予以必要的修正，意即該等複數種基因可來自表2。本發明之第一、第二及第三態樣之較佳實施例及可選特性對表2予以必要的修正。

因此，在第四態樣中，本發明提供一種產生乳腺腫瘤樣品之核酸表現輪廓之方法，其包含如下步驟：

- (a)將表現產物與該乳腺腫瘤樣品分離；
- (b)識別來自表2之複數種基因的表現量；及
- (c)自該等表現量產生表現輪廓。

如本發明之第一態樣中所論述，乳腺腫瘤樣品可為任何類別之乳腺腫瘤。較佳地，乳腺腫瘤樣品之ER狀況較佳在步驟(a)之前得以判定。

在本發明之第五態樣中，提供一種包含高置信度及/或低置信度乳腺樣品之複數種基因表現輪廓的表現輪廓資料庫，其中每一表現輪廓均係得自來自表2之複數種基因，且其中該資料庫可擷取地保存於資料載體上。較佳地，組成該資料庫之表現輪廓係藉由根據第四態樣之方法而產生。

表2之基因提供一替代性多基因分類器。

在本發明之第六態樣中，提供一種將乳腺腫瘤樣品分類為低置信度或高置信度之方法，該方法包含：提供該樣品之表現輪廓，其中該表現輪廓包含來自表2之複數種基因的表現量；且基於該表現輪廓將腫瘤分類為高或低置信度腫瘤。

本發明之第六態樣可包含如下步驟：

- (a)獲得來自由患者所獲得之乳腺腫瘤樣品之表現產物；
- (b)藉由使該等表現產物與結合組員接觸來確定表2中所識別之複數種基因的表現量，其中每一結合組員均能夠特異地結合至該等複數種基因之表現產物；及
- (c)基於該等表現量來識別此患者中低置信度乳腺腫瘤之存在。

步驟(c)可包含將結合輪廓與作為低置信度腫瘤之特徵之輪廓相比較。低置信度腫瘤可為ER+或ER-。步驟(c)可包含統計學技術之使用，諸如加權表決法及/或支持向量機(SVM)。

該等複數種基因可包含來自表2之全部或大體上全部的基因，或來自表2a或表2b之全部或大體上全部的基因，或由上述基因組成。

該等複數種基因可包括表2之至少10、20、30、40、50、60、70、80、90種或全部的基因。

較佳地，該等複數種基因包含來自表2a及/或來自表2b之約50、或約40、或約30、或約20、或約10種基因，或由上

述基因組成。因為表2a及2b中每一表之基因均係按顯著性的次序排列，所以來自表2之基因較佳選自表2a及/或表2b之上部分，較佳為上半部分。在高置信度與低置信度腫瘤之間展示出最大擾動之基因出現於表2a及表2b中每一表的上部分，而彼等具有較小擾動之基因則出現在下部分。

熟習此項技術者將瞭解：相較於產生作為低及/或高置信度乳腺腫瘤之特徵之表現輪廓所需的最小顯著性基因之數目，產生該特徵表現輪廓需要較少的最大顯著性基因。舉例而言，與選自表2a之下半部分的基因相比，需要更少的來自該表2a之上半部分的基因。

將該等複數種基因之數目及選擇選定成提供能夠在高置信度與低置信度腫瘤之間加以區分之表現特徵。

該等複數種基因可包括表2a及/或表2b之不多於50種基因。該等複數種基因可包括表2a及/或表2b之不多於40種基因。該等複數種基因可包括表2a及/或表2b之不多於30種基因。該等複數種基因可包括表2a及/或表2b之不多於20種基因。該等複數種基因可包括表2a及/或表2b之不多於10種基因。該等複數種基因可包括表2a及/或表2b之不多於5種基因。

該等複數種基因可包含表2a及/或表2b之5至50種基因，或基本上由上述基因組成。該等複數種基因可包含表2a及/或表2b之10至40種基因，或基本上由上述基因組成。該等複數種基因可包含表2a及/或表2b之10至30種基因，或基本上由上述基因組成。該等複數種基因可包含表2a及/或表2b

之10至20種基因，或表2a及/或表2b之20至30種基因，或基本上由上述基因組成。

該等基因較佳為約10種基因，其可選自表2a之約前40、或約前30、或約前20種基因。該等約10種基因可選自表2a之約前15種基因。該等約10種基因可為表2a的前10種基因。該等基因較佳為約10種基因，其可選自表2b之約前40、或約前30、或約前20種基因。該等約10種基因可選自表2b之約前15種基因。該等約10種基因可為表2b的前10種基因。

該等基因較佳為約10至20種基因，其較佳選自表2a及/或表2b之約前30種基因。

該等複數種基因可包含選自由表2a之約前20種基因與表2b之約前20種基因所組成之群的約30、或約20、或約10種基因，或由上述基因組成。該等複數種基因可包含選自由表2a之約前10種基因與表2b之約前10種基因所組成之群的約10、或約15、或約20種基因，或由上述基因組成。

本發明之方法較佳進一步包含分類之前判定ER+或ER-狀況之步驟。ER狀況可藉由免疫組織化學法(例如，使用ER之抗體)或藉由使用經調適以評定基因表現輪廓之概率學/統計學模型來判定。

發明者已進行了進一步分析，且識別出用於在高與低置信度腫瘤之間加以判別之其它多基因分類器。該等分析之目標在於識別一組最佳基因，其無需考慮腫瘤之ER狀況即可用於分類"高"及"低置信度"腫瘤。使用三種獨立分析方法之系列(微陣列顯著性分析法、基因排列法及威爾科克森測

試法(Wilcoxon test))來識別在二組(LC及HC)之間差異表現之基因。該等分析之結果為展示於表A1、A2、A3及A4中之其它多基因分類器。

在表A1中，存在88種基因，其可用於在高與低置信度腫瘤之間加以判別。表A1之基因係使用SAM(微陣列顯著性分析法)而得以識別。該等基因中的86種基因在低置信度腫瘤中上調，同時該等基因中的2種基因在高置信度腫瘤中上調。

在表A2中，存在251種基因，其可用於在高及低置信度腫瘤之間加以判別。表A2之基因係使用GR(基因排列法)藉由SVM而得以識別。

在表A3中，存在38種基因，其可用於在高及低置信度腫瘤之間加以判別。表A3之基因係在 P -值 < 0.05 且 $a \geq 2$ 倍變化截止值下使用WT(威爾科克森測試法)而得以識別。

在表A4中，存在13種共同基因(意即，在表A1、A2、A3中發現的基因)。該等13種"共同基因"係強健的顯著性標誌，且可達成与其它"完全"標誌集合相當的判別效能。

因此，在第七態樣中，本發明提供一種產生乳腺腫瘤樣品之核酸表現輪廓之方法，其包含以下步驟：

- (a)將表現產物與該乳腺腫瘤樣品分離；
- (b)識別來自表A4及/或表A1及/或表A2及/或表A3之複數種基因的表現量；及
- (c)自該等表現量產生表現輪廓。

如本發明之第一態樣所論述，乳腺腫瘤樣品可為任何類

別之乳腺腫瘤。

在本發明之第八態樣中，提供一種包含高置信度及/或低置信度乳腺樣品之複數種基因表現輪廓的表現輪廓資料庫，其中每一表現輪廓均係得自來自表A4及/或表A1及/或表A2及/或表A3之複數種基因，且其中該資料庫可擷取地保存於資料載體上。較佳地，組成該資料庫之表現輪廓係藉由根據第七態樣之方法產生。

在本發明之第九態樣中，提供一種用於將乳腺腫瘤樣品分類為低置信度或高置信度之方法，該方法包含：提供該樣品之表現輪廓，其中該表現輪廓包含來自表A4及/或表A1及/或表A2及/或表A3之複數種基因的表現量；且基於該表現輪廓將腫瘤分類為高或低置信度腫瘤。

本發明之第九態樣可包含以下步驟：

- (a)獲得來自由患者所獲得之乳腺腫瘤樣品之表現產物；
- (b)藉由使該等表現產物與結合組員接觸來確定表A4及/或表A1及/或表A2及/或表A3中所識別之複數種基因的表現量，其中每一結合組員均能夠特異地結合至該等複數種基因之表現產物；及
- (c)基於該等表現量來識別該患者中低置信度乳腺腫瘤之存在。

步驟(c)可包含將該等表現量與作為低及/或高置信度腫瘤之特徵之輪廓相比較。低置信度腫瘤可為ER+或ER-。步驟(c)可包含統計學技術之使用，諸如加權表決法及/或支持向量機(SVM)。

較佳地，該等複數種基因包含表A4之大體上全部的基因，或基本上由上述基因組成。儘管該等複數種基因可獨立地來自表A1、A2及A3之任一或多個表中，但亦可包括來自表A1、A2及A3中之每一表的其它基因。該等複數種基因未必需要包括表A4之基因。

因此，本發明之第一、第二及第三態樣對上述表A1、A2及A3中之每一表予以了必要的修正，即在本發明之每一態樣中，該等複數種基因可來自表A1及表A2及表A3中之任一或多個表中。本發明之第一、第二及第三態樣之實施例及較佳/可選特徵對表A1、A2、A3及A4予以了必要的修正。

該等複數種基因可包括表A1之至少10、20、30、40、50、60、70、80種或全部的基因。

該等複數種基因可為來自表A1之全部的或大體上全部的"在低置信度中上調"及/或"在高置信度中上調之基因"。該等複數種基因可包含來自表A1的"在低置信度中上調"之約80、或約70、或約60、或約50、或約40、或約30、或約20、或約10、或約5種基因，或由上述基因組成。該等複數種基因可包括來自表A1的"在高置信度中上調"之一或兩種基因。

來自表A1之基因較佳選自"在低置信度中上調"之基因組的上部分。因為該等基因係按顯著性的次序排列，所以該上部分較佳為該表之上半部分。在高置信度與低置信度腫瘤之間展示出最大差異表現的基因出現在表A1之上部分，而彼等以較小差異表現之基因則出現在下部分。

該等複數種基因可包括表A1之不多於80、或70、或60、或50、或40、或30、或20、或10、或5種基因。

該等複數種基因可包含表A1之5至70種基因，或基本上由上述基因組成。該等複數種基因可包含表A1之10至60種基因，或基本上由上述基因組成。該等複數種基因可包含表A1之10至50種、或10至40種、或10至30種基因，或基本上由上述基因組成。

該等複數種基因可為約10至15種基因，其可選自表A1之約前40、或約前30、或約前20種基因。該等複數種基因較佳包含表A1之約前30種基因中的約10至20種基因。

該等複數種基因包括表A2之至少10、20、30、40、50、60、70、80、90、100、110、120、130、140、150種或全部的基因。

該等複數種基因可包括表A2之不多於250、或240、或230、或220、或210、或200、或190、或180、或170、或160、或150、或140、或130、或120、或110、或100、或90、或80、或70、或60、或50、或40、或30、或20、或10、或5種基因。

該等複數種基因可包含表A2之5至200種基因，或基本上由上述基因組成。該等複數種基因可包含表A2之10至150種基因，或基本上由上述基因組成。該等複數種基因可包含表A2之10至100、或10至70、或10至50種基因，或基本上由上述基因組成。

該等複數種基因可為約10至15種基因，其可選自表A2之

約前50、或約前40、或約前30、或約前20種基因。該等複數種基因較佳包含表A2之約前30種基因中的約10至20種基因。

該等複數種基因可包括表A3之至少10、20、30、35種或全部的基因。

該等複數種基因可包括表A3之不多於35、或30、或20、或10、或5種基因。

該等複數種基因可包含表A3之5至35種基因，或基本上由上述基因組成。該等複數種基因可包含表A3之10至30種基因，或基本上由上述基因組成。該等複數種基因可包含表A3之10至20、或20至30種基因，或基本上由上述基因組成。

該等複數種基因可為約10至15種基因，其可選自表A3的前30、或約20種基因。該等複數種基因較佳包含表A3之約前30種基因中的約10至20種基因。

該等複數種基因可包括表A4之至少5、10、15種或全部的基因。

該等複數種基因可包括表A4之不多於10、或8、或6、或5種基因。

該等複數種基因可包含表A4之5至13種基因，或基本上由上述基因組成。該等複數種基因可包含表A4之10至13種基因，或基本上由上述基因組成。

就該等複數種基因而言，術語"約"意指所規定的基因數目加上或減去所規定的基因數目之10%或一基因兩者中之較大者。

如前文所述，表現產物可為經轉錄之核酸序列或經表現之多肽。經轉錄之核酸序列可為RNA或mRNA。表現產物亦可為自該mRNA所產生之cDNA。表現產物可為cRNA。

結合組員可為能夠在適當的雜交條件下特異地結合至經轉錄之核酸的補充核酸序列。通常使用cDNA或寡核苷酸序列。

當表現產物為經表現之蛋白質時，結合組員較佳為對該經表現之多肽具特異性的抗體或包含抗體結合域之分子。

基於偵測之目的，可使用此項技術中已知的標準程序來對結合組員作標記。或者，可在與測試中之樣品分離後對表現產物作標記。一較佳偵測方式係使用可藉由光度計偵測之螢光標記。替代性偵測方式包括電氣訊號法(electrical signalling)。舉例而言，Motorola電子感應器系統具有兩個探針：一個自由浮動的"俘獲探針"；及一個附著至兼任電極表面之固體表面的"訊號探針"。兩個探針均起著表現產物之結合組員的作用。當發生結合時，兩個探針達成彼此緊密接近，從而產生可被偵測之電訊號。

如上所述，結合組員可為用於PCR(例如，多重PCR)中以特異地擴增基因識別器之表現產物的數目之寡核苷酸引子。隨後，可在凝膠上分析該等產物。然而，較佳地，結合組員係固定至固相支持體之單核酸探針或抗體。隨後，可使表現產物在固相支持體上通過，藉此使表現產物與結合組員接觸。固相支持體可為：玻璃表面，例如顯微鏡載玻片；珠粒(Lynx)；或光纖。在珠粒情形中，可將每一結

合組員固定至個別珠粒，且隨後使其在溶液中與表現產物接觸。

此項技術中存在各種用於確定特定基因集之表現輪廓的方法，且該等方法可適用於本發明。舉例而言，基於珠粒之方法(Lynx)或分子條碼法(Surromed)即為已知技術。在該等情形中，每一結合組員均附著至各自可讀取且自由浮動以易於與表現產物接觸之珠粒或"條碼"。結合組員與表現產物(目標)之結合係於溶液中達成，此後，使加注標籤之珠粒或條碼穿過一設備(例如，流式細胞儀)且被讀取。

另一確定表現輪廓之已知方法為由Illumina公司所研發之儀器法，即光纖法。在此情形中，每一結合組員均附著至光纖電纜末端之特異"位址"。表現產物與結合組員之結合可誘發螢光變化，該變化可由光纖電纜另一端上之設備讀取。

本發明者已成功地使用了一包含固定至固相支持體之複數個核酸序列的核酸微陣列。藉由使代表經表現之基因的核酸序列(例如cDNA)穿過微陣列，其能夠自腫瘤樣品及得自乳腺組織之正常細胞創建作為該等表現產物的特徵之結合輪廓。

本發明進一步提供一種用於分類乳腺腫瘤樣品之裝置，較佳為微陣列，其包含附著至固相支持體之複數個結合組員(較佳為核酸序列)，每一結合組員均能夠特異地結合至來自多基因分類器組表S4、表2、表A1、表A2、表A3及表A4中任一或多個表中之基因的表現產物。較佳地，該裝置包

含如先前對每一該等多基因分類器所界定(參見上文)能夠結合至複數種基之的表現產物的結合組員，或基本上由上述結合組員組成。該裝置可包含能夠結合至來自每一該等多基因分類器之複數種基因、或來自該等多基因分類器之一或多者中之複數種基因的表現產物之結合組員，或基本上由上述結合組員組成。

該裝置可包括能夠特異地結合至來自該多基因分類器或來自該多基因分類器之子集的至少5種基因、更佳至少10種基因或至少15種基因之表現產物的結合組員。該多基因分類器之子集可為(例如)來自表2中之ER+/低相對於ER+/高之基因，或來自表S4(a)之ER+/低中之上調組之基因。在一最佳實施例中，固相支持體將收容能夠特異地且獨立地結合至表A4中所識別之所有基因的表現產物之結合組員。

該裝置較佳包括能夠特異地結合至來自多基因分類器之表現產物或結合至其複數種基因的結合組員，且可包括能夠特異地結合至U133A微陣列上不多於14396種基因之表現產物的結合組員。該裝置可包括能夠特異地結合至U133A微陣列上不多於90%基因之表現產物的結合組員。該裝置可包括能夠特異地結合至U133A微陣列上不多於80%、或70%、或50%、或40%、或30%、或20%、或10%、或5%基因之表現產物的結合組員。

此外或替代性地，固相支持體可收容不多於14000、不多於10000、不多於5000、不多於3000、不多於1000、不多於500、或不多於400、或不多於300、或不多於200、或不多

於100、或不多於90、或不多於80、或不多於70、或不多於60、或不多於50、或不多於40、或不多於30、或不多於20、或不多於10、或不多於5種不同基因之結合組員。

高密度核酸序列(通常為cDNA或寡核苷酸)通常係固定於固相支持體之極小的離散區域或位點上。固相支持體通常為經基質(或晶片)塗布之顯微鏡載玻片或膜過濾器。核酸序列通常係藉由機器人系統傳遞(或印刷)至經塗布之固相支持體上，且隨後使其固定不動或固定至支持體。

在一較佳實施例中，通常使用螢光標記對得自樣品之表現產物作標記，然後使其與固定不動之核酸序列接觸。在雜交後，使用諸如高解析度雷射掃描儀之偵測器來偵測螢光標誌。在一替代性方法中，可藉由諸如生物素之非螢光標記來對表現產物加注標籤。在雜交後，接著可藉由結合/鍵結至第一非螢光標記(例如，結合至生物素之經螢光標記的鏈酶抗生物素蛋白)之螢光染料來著色微陣列。

指示基因表現之模式(表現模式或輪廓)之結合輪廓係藉由用數位成像軟體來分析自每一離散位點所發出之訊號而獲得。接著，可將實驗樣品之基因表現的模式與對照物之基因表現的模式(意即，來自高置信度或低置信度樣品之表現輪廓)相比較以用於差異分析。

如上所述，對照物或標準物可為先前判斷為正常或惡性細胞之特徵的一或多種表現輪廓。該或該等多種表現輪廓可擷取地儲存於資料載體上以作為資料庫之一部分。此已在上文中有所論述。然而，亦可將對照物引入檢定程序。

換言之，測試樣品可"摻加(spike)"有一或多種可充當與測試樣品中基因識別器之表現量相比較的對照物之"合成腫瘤"或"合成正常"表現產物。

大多數微陣列都會利用一或兩個螢光團。對於雙色陣列而言，最常使用之螢光團係Cy3(綠色通道激勵)及Cy5(紅色通道激勵)。微陣列影像分析之目的係自每一表現產物提取雜交訊號。對於單色陣列而言，訊號係作為一給定目標之絕對強度來量測(基本上係對於雜交至單一樣品之陣列而言)。對於雙色陣列而言，訊號係作為具有不同螢光標記之兩種表現產物(例如，樣品與對照物(對照物另稱為"參考物"))之比率來量測。

根據本發明之裝置(例如，微陣列)較佳包含複數個離散位點，每一位點均含有一或多個寡核苷酸，且每一位點均代表選自該多基因分類器之基因的表現產物之不同的結合組員。在一實施例中，微陣列將含有若干位點以用於在一或多個多基因分類器中所提供之每一基因。每一位點將包含複數個相同的寡核苷酸，每一寡核苷酸均能夠結合至其所代表之表S4之基因的表現產物，例如 mRNA 或 cDNA。

在本發明之另一態樣中，提供一種用於將乳腺腫瘤樣品分類為高置信度或低置信度之套組，該套組包含結合組員及偵測試劑，其中每一結合組員均能夠特異地結合至該多基因分類器中所識別之複數種基因的表現產物。

多基因分類器之基因係以其Unigene登記號列出(對應於Unigene之建置塊(build)160)。因此，每一基因之序列均可

自 Unigene 資料庫擷取。此外，對於某些基因而言，Affymetrix(www.affymetrix.com)提供了探針集合之實例，包括在固相支持體上使用時能夠偵測基因之表現的探針之序列(意即，寡核苷酸序列形式之結合組員)。探針細節可使用目標基因之 Unigene ID 自 Affymetrix 網站之 U133 部分存取。

若在將來，表中所列的 Unigene ID 中之一 ID 被併入新的 ID 中，或分裂為兩個或兩個以上 ID(例如，在資料庫之新的建置塊中)，或一起被刪除，則本發明者所欲之基因序列可藉由存取 Unigene 之建置塊 160 來擷取。

較佳地，套組中之一或多個結合組員(抗體結合域或核酸序列，例如寡核苷酸)係固定至一或多個固相支持體，例如用於微陣列或光纖檢定之單一載體，或諸如珠粒之多重載體。偵測構件較佳係用於對測試中之樣品的表現產物作標記之標記(具放射性或為染料，例如具螢光性)。套組亦可包含用於偵測且分析測試中之表現產物的結合輪廓之構件。

或者，結合組員可為能夠結合至表現產物以使得其可在 PCR 中得以擴增之核苷酸引子。該等引子可進一步包含偵測構件，意即可用於識別經擴增之序列及其相對於其它經擴增序列之豐度的標記。

套組亦可包含一或多種可擷取地保存於資料載體上用以與測試樣品之表現輪廓相比較之標準表現輪廓。該一或多種標準表現輪廓可根據本發明之第一態樣產生。

乳腺組織樣品可藉由切除性乳腺活組織檢查或細針抽吸

獲得。

同樣地，表現產物較佳為 mRNA、或自該 mRNA 產生之 cDNA、或 cRNA。結合組員較佳為固定至微陣列或珠粒形式之一或多個固相支持體的寡核苷酸(參見上文)。結合輪廓較佳藉由能夠偵測用於標記表現產物之標記的偵測器來分析。對乳癌之存在或其危險性的判定可藉由將樣品之結合輪廓與對照物之結合輪廓(例如，標準表現輪廓)相比較而做出。

在所有上述態樣中，較佳使用能夠特異地結合(且在核酸引子情形中，能夠擴增)該多基因分類器之表現產物的結合組員。其原因在於所有基因之表現量組成了專用於測試中之樣品的表現輪廓。所測試之基因表現量的數目愈多，則表現輪廓之分類愈可靠。因此，較佳評定選自一或多個該等多基因分類器之多於5種基因之表現量，更佳為來自該多基因分類器多於10、多於20、多於30種基因，甚至更佳為多於40種，且較佳為所有基因。舉例而言，結合組員可能夠結合至來自表S4之全部基因、或其中之複數種基因的表現產物，如上文所界定。

已知的微陣列及基因晶片技術允許利用大量的結合組員。因此，更佳之方法將可使用如先前對每一多基因分類器所界定代表該多基因分類器中之全部基因或其中之複數種基因的結合組員。然而，專業人員將瞭解：該等基因之一部分可省略，且該方法仍以可靠且統計學上精確之方式進行。在大部分情形中，較佳使用代表該多基因分類器中

至少70%、80%、或90%基因的結合組員。在本文中，多基因分類器較佳意指表S4之基因或該表之子集或組。多基因分類器可為表A4之基因。

因此，複數種可意指如上文所提及之多基因分類器的至少50%、更佳至少70%且甚至更佳至少90%。

提供基因識別器允許定製診斷工具(例如，核酸微陣列)，且允許使用該等診斷工具來預測、診斷或將腫瘤分型。此外，此等診斷工具可與電腦結合使用，其中該電腦經程式化以確定使用診斷工具(例如，微陣列)所獲得之表現輪廓且將其與作為高置信度相對於低置信度腫瘤之特徵的"標準"表現輪廓進行比較。為此，電腦不僅為使用者提供可用於分類患者之腫瘤類型的資訊，而且同時該電腦亦獲得可用以確定"標準"表現輪廓之另一表現輪廓，從而可更新其自身的資料庫。

因此，本發明首次允許將專用晶片(微陣列)製成含有對應於該等多基因分類器或其中之複數種基因的探針。陣列之確切實體結構可改變，且其範圍係自附著至二維固體基質之寡核苷酸探針至已個別由獨特標記(例如，"條碼")"加注標籤"的自由浮動探針。

可創建一對應於各種生物學分類(例如，高置信度或低置信度 ER^+/ER^-)之資料庫，其將由藉由專用微陣列所確定之各種乳腺組織的表現輪廓組成。然後，該資料庫可經處理及分析以使得其最終將含有：(i)對應於資料庫中每一表現輪廓之數字資料；(ii)充當彼特定分類之典範輪廓之"標準"

輪廓；及(iii)代表所觀察到的個別輪廓相對於"標準"輪廓之統計學變化之資料。

在一實施例中，為了評估患者之樣品，首先分離彼患者之乳腺樣品(經由切除性活組織檢查或細針抽吸獲得)的表現產物，且使用專用微陣列確定彼樣品之表現輪廓。為了分類該患者之樣品，將在上述資料庫中查詢該患者之樣品的表現輪廓。查詢可以直接或間接方式進行。"直接"方式係將患者之表現輪廓直接與資料庫中的其它個別表現輪廓相比較以確定傳遞最佳匹配之輪廓(且因此確定傳遞最佳匹配之分類)。或者，查詢可更"間接"地進行，例如，可簡單地將患者之表現輪廓與資料庫中之"標準"輪廓進行比較。間接方法的優勢在於：因為"標準"輪廓代表許多個別輪廓之總集，所以其將具有小得多的資料密集性，且可儲存於相對廉價的電腦系統中，然後該系統形成根據本發明之套組之一部分(意即，與微陣列相關聯)。在直接方法中，因為必須儲存許多個別輪廓，所以資料載體將很可能具有大得多的規模(例如，電腦伺服器)。

藉由將患者表現輪廓與標準輪廓(間接方法)及該群中預先確定之統計學變化相比較，亦可傳遞有關患者表現輪廓與高或低置信度腫瘤之"標準"典範輪廓緊密匹配程度之"置信級值"。該值將為臨床醫師提供關於分類之可信賴度及(例如)是否應重複分析之有價值的資訊。

如上文所提及，亦可將患者表現輪廓儲存於資料庫上，且其可在任何時候用於更新該資料庫。

現將以實例的方式參考所附圖式說明本發明之各態樣及實施例。熟習此項技術者將明瞭其它態樣及實施例。本文中所提及之所有文獻皆以引用之方式併入本文中。

【實施方式】

結果

使用來自中國患者之表現輪廓藉由ER狀況來對乳腺腫瘤進行分類揭示了一群獨特的"低置信度"樣品

在高加索人群及亞洲人群中乳癌之整體發病模式截然不同(8)，此促使發明者研究是否亦可在其區域患者群中觀察到來自先前報導(3,4)之發現。首先，其使用基因表現輪廓資料藉由ER狀況對一組乳腺腫瘤進行分類。選定55種乳腺腫瘤之訓練組，其中使用IHC預先判定每一腫瘤之ER狀況。測試兩種分類方法：加權表決法(WV)及支持向量機(SVM)，且藉由留一校驗法(LOOCV)來評定分類精確度(補充資訊)。除了將樣品分類外，使用定量化量度來提供對分類之不確定度之評定(材料及方法)。訓練組之整體分類精確度為95%(WV)及96%(SVM)，其中7個樣品以"低置信度"或邊緣預測為特徵(灰色框，圖1a)。為確定此等"低置信度"樣品是否亦可在一組獨立的腫瘤中觀察到，使用具有41種腫瘤之第二組作為獨立測試組。儘管該獨立測試組之整體分類精確度為91%(WV與SVM)，但是9個樣品再次顯示出"低置信度"預測(圖1b)。因此，藉由使用兩種不同的分類方法(WV與SVM)，發現當基於某些乳腺腫瘤之基因表現輪廓藉由其ER狀況來分類時該等乳腺腫瘤展現出獨特的"低置

信度"特徵。

患有"低置信度"腫瘤之患者比患有"高置信度"腫瘤之患者展現示出減少的整體存活率及更短的遠距離轉移時間

因為將腫瘤區分為"高"與"低置信度"亞種群係藉由純計算分析腫瘤基因表現輪廓來達成，所以吾人並不清楚此區分是否具有生物學或臨床意義，及以此方式使用基因表現輪廓與判定乳腺腫瘤之ER狀況之習知免疫組織化學技術相比是否產生任何實質上的優勢。為了解決此問題，發明者研究"低置信度"腫瘤是否可能展現任何區別於其"高置信度"對應物的臨床表現。其使用兩種可公開獲得之乳腺腫瘤表現資料集，其中可獲得與其相關但為不同類型之臨床資訊。第一集合(9)由78種乳癌及7種非惡性樣品的cDNA微陣列資料集組成，其中具有患者之整體存活資訊(稱為Stanford資料集)。第二集合(10)由使用基於寡核苷酸之微陣列進行輪廓分析之71種ER+及46種ER淋巴結陰性腫瘤組成，其中97種樣品具有作為自最初腫瘤診斷至出現新的遠距離轉移之時間間隔的臨床資訊(稱為Rosetta資料集)。發明者使用WV藉由ER亞型來在Stanford與Rosetta資料集中分類該等乳腺腫瘤。與其自身的資料集一致，在Stanford資料集中之56種ER+與18種ER腫瘤中(由於缺少ER狀況資訊，所以除去4種腫瘤)，其觀察到93%的整體LOOCV精確度，其中14種腫瘤被歸類為"低置信度"。類似地，WV分析亦識別出在Rosetta資料集中展現出"低置信度"分類之15種腫瘤，其整體LOOCV精確度為92%。該等數字與發明者之

自身患者種群中所觀察到的數字相當。

接著，發明者使用卡普蘭-梅爾(Kaplan-Meier)分析將"高"與"低置信度"腫瘤群的臨床表現進行比較。如圖2所示，患有"低置信度"腫瘤之患者比其"高置信度"對應物展現出顯著惡化之整體存活($p=0.0003$ ，對數秩測試(log rank test))及更短的遠距離轉移之時間($p=0.0001$ ，對數秩測試)。此結果指示："高"相對於"低置信度"之雙重區分確實具有臨床意義。然後，發明者重複該分析，不同之處在於首先將腫瘤再細分為獨立的ER+及ER-類別。對於ER+腫瘤而言，其再次發現：與"高置信度"ER+腫瘤相比，"低置信度"ER+腫瘤與顯著惡化之整體存活率($p=0.03$ ，對數秩測試)及更短的轉移時間($p=0.004$ ，對數秩測試)相關聯(圖2)。對於ER-腫瘤而言，未觀察到整體存活及轉移時間上的任何統計學顯著差異。該等結果指示：ER+腫瘤可基於"高"與"低置信度"之二分法而再細分為展現不同臨床表現之獨特疾病組。因為目前無法藉由用於ER偵測之習知免疫組織化學方法來在該等二組之間加以區分，所以此結果亦證實了基因表現輪廓資料如何作為習知策略的有用輔助方式以用於乳癌預示與分級。

"低置信度"腫瘤在對於ER亞型判別而言重要之基因的表現中展現出廣泛擾動

用於該等及其它研究之分類演算法(例如，WV、SVM、ANN，參見下文)均依賴於多種判別基因之組合輸入，該等基因之個別貢獻隨後組合以達成特定的分類決定(意即，腫

瘤是ER+或是ER-)。將該等乳腺腫瘤之"低置信度"預測狀況歸因於一些關鍵判別要素之驚人的反調節(意即，特異效應)或歸因於大量判別基因之更細微的擾動(意即，廣泛效應)在形式上均係可能的。為了在該等兩種可能性之間加以區分，發明者將對於在"高"與"低"置信度腫瘤之間的ER亞型判別而言重要的基因之表現量進行比較。首先，為了識別在ER+與ER-腫瘤之間差異調節之ER判別基因，其利用一稱為微陣列顯著性分析(SAM)之統計學技術(11)。

採用其組合資料集(總數=96種腫瘤)，以0%之"檢出率"(FDR)識別總共133種差異調節之基因(SAM-133)(FDR為SAM用以估算假陽性之數目的指數，對於100種基因而言，10%的FDR表示10種基因可能為假陽性)。在該集合中，122種基因在ER+樣品中上調(意即，關於ER狀況呈陽性)，而剩餘11種基因在ER+腫瘤中下調(意即，關於ER狀況呈陰性)。據預測，SAM-133基因集包括大量與ER路徑相關之基因，諸如ESR1、LIV1(雌激素可誘導之基因)及TFF1，且某些基因(例如，GATA-3)經多次識別。在其它人所報導之類似列表中(3，4)亦發現SAM-133列表中的大量基因。

然後，發明者將ER+與ER-腫瘤各自再細分為"高"與"低"置信度類別(意即，ER+/高、ER+/低、ER-/高、ER-/低)，且在各組之間比較SAM-133基因之表現量(圖3)。在SAM-133基因集中關於ER狀況呈陽性之122種基因中，約62%在ER+/低樣品中比在ER+/高腫瘤中展現出顯著更低的平均表現量(稱為"擾動表現")(p < 0.05，圖3a與表2)。具有"擾動"表現

之基因包括ER、GATA3、BCL2、TGF1R及RARA，而諸如TFF1、TFF3及XBP1之其它ER-判別基因則未受影響。類似地，在ER-"高"與"低"置信度樣品中，發明者證明了交互模式，其中122種基因中約42%在ER-/低樣品中比在ER-/高腫瘤中展現出更高的平均表現量($p < 0.05$ ，圖3b及表2)。令人感興趣的是，儘管某些基因(例如，GATA3、BCL2)之表現量在ER+與ER-亞型中的"低"與"高"置信度樣品之間有擾動，但其它基因之擾動似乎具有亞型特異性。舉例而言，ESR1與IGFR1僅在ER+樣品中有擾動，而XBP1僅在ER-樣品中有擾動。最後，在關於ER+狀況呈陰性(意即，在ER-腫瘤中經高表現)之ER-判別基因之表現量存在極小的變化(圖3c及d)。此結果表明：在"低置信度"樣品中所觀察到的表現擾動儘管廣泛，但其主要在表現關於ER呈陽性之基因中觀察到(補充資訊)。

ERBB2腫瘤基因之提高的表現與"低置信度"預測顯著相關

在"低置信度"乳腺腫瘤中所觀察到的表現擾動可歸因於多種原因，其範圍為自實驗變化(例如，不良的樣品品質、腫瘤切除及處理)、分類方法的選擇至種群與樣品異質性。為了獲得對該等表現擾動所潛藏的可能機制之見識，發明者試圖確定是否存在任何可能與"低置信度"狀態相關之特異組織病理學參數。在腫瘤之"低置信度"狀況與患者年齡、淋巴結狀況、腫瘤級別、p53突變狀況或孕激素受體狀況之間未觀察到任何顯著關聯(表1)。然而，發明者發現在腫瘤之ERBB2狀況與"低置信度"預測之間存在顯著的正向

關聯 ($p < 0.001$ ，補充資訊)。然後，使用獨立的測試組樣品來評定使用訓練組資料所觀察到的此關聯性。在獨立測試組中的9個"低置信度"樣品中，8個腫瘤亦為ERBB2+(8/9)，此指示此關聯性並不具有資料集特異性。

本發明者亦研究了是否可藉由將"高"與"低"置信度腫瘤之整體表現輪廓進行比較而獨立地發現"低置信度"預測與高ERBB2表現之間的相關性。首先，其將屬於ER+亞型之"高置信度"與"低置信度"腫瘤進行比較。總共89種基因經識別為經顯著調節 ($FDR = 14\%$)。在ER+"低置信度"樣品中最顯著上調之頂部50種基因中，3種基因，即PMNT(排在第4位)、GRB7V(第8位)及ERBB2(第36位)尤其引人注意(補充資訊)，因為其在實體上均處於17q區，該區為乳癌中DNA擴增之頻繁目標(12)。在一單獨分析中，亦將ER-"高置信度"與ER-"低置信度"樣品相比較。在經識別為差異調節之頂部50種基因中 ($FDR = 4\%$)，發明者再次識別出17q基因PMNT(排在第5位)、GRB7V(第10位)及ERBB2(第28位)在"低置信度"樣品中展現出增加之表現(補充資訊)。總體看來，該等結果表明：對於ER+與ER-亞型兩者而言，與"高置信度"腫瘤相比，"低置信度"乳腺腫瘤與ERBB2之增加的表現顯著相關，此最有可能係由17q座位之DNA擴增引起。然而，請注意：儘管"低置信度"預測與ERBB2+表現之間的關聯性極其顯著，但其並不完善，因為藉由習知IHC標識為ERBB2+之某些腫瘤展現出"高置信度"預測，同時並非所有"低置信度"腫瘤均為ERBB2+。一種可能性為：除了ERBB2

外，其它基因亦可能有助於展現"低置信度"狀態之乳腺腫瘤。

然後，為了驗證其發現，發明者分析了其它獨立獲得之乳癌表現資料集。首先，在Stanford資料集中之9種ERBB2+腫瘤中，該等9種腫瘤均被預測為屬於"低置信度"組($p < 0.001$ ，補充資訊)。第二，在Rosetta資料集中，其再次發現預測之置信級與ERBB2表現之間的顯著關聯性($p < 0.001$ ，補充資訊)。第三，Gruvberger及其同事在具有28種ER+與30種ER-樣品之cDNA微陣列資料集上利用人工神經網路(ANN)來預測乳腺腫瘤之ER狀況(3)。其結果如圖4b所示，其描繪了具有樣品標準偏差(SD)之ANN模型的輸出，如使用ER亞型之頂部100種判別基因所檢定。具有寬SD之樣品類似於WV與SVM方法之"低置信度"狀況。由圖4b可見，ERBB2+樣品(在圖4a中得以判定)趨向於與大SD相關聯，此指示高不確定度，尤其是對於ER腫瘤而言。總體看來，在不同患者群(亞洲人、歐洲人/高加索人)上利用不同微陣列技術(Affymetrix、cDNA及寡核苷酸)於源自不同實驗室的寬範圍之資料集上觀察到ER預測之置信級與ERBB2狀況之間的關聯性，且藉由不同分類演算法(WV、SVM、ANN)對其進行了預測。關於發明者之資料集與可公開獲得之資料集之該等結果的共性表明：大體而言，高ERBB2表現與"低置信度"預測狀況之間的相關性可能係乳癌之固有特性。

已知在低置信度樣品中擾動之基因中之顯著比例並非藉由雌激素調節且在其啟動子中缺乏潛在ERE

在高 ERBB2 量與 "低置信度" 腫瘤中所觀察到的 ER-亞型判別基因之廣泛擾動之間的強烈相關性產生一可能性，即 ERBB2 可能在功能上促成此現象。其可能的一種發生機制係藉由 ERBB2 發訊號，已有人提出利用該機制來抑制 ER 之轉錄活性(參看討論)。在此情形下，吾人可預期在 "高置信度"(ERBB2-) 與 "低置信度"(ERBB2+) 腫瘤之間擾動之基因中的顯著比例將由經 ER 調節之基因組成。發明者以兩種方式測試此假設。首先，其將顯著擾動之基因列表(表 2) 與雌激素(E2)刺激之 MCF-7 細胞(13) 獲得之 SAGE 表現資料進行比較，以確定兩者之間是否存在重疊。在 SAGE 資料與 "擾動" 基因列表之間僅發現兩種共同的基因(STC2、TFF1)，且其中一種基因(TFF1)係以與預期方式相反之方式調節，從而在 ERBB2+ 樣品中展現出更高表現。在細胞株檢定之界限內，此結果表明：在 "低置信度" 腫瘤中許多 "擾動" 基因可能並非藉由雌激素直接調節。第二，因為活體外細胞株研究可能無法充分地概括雌激素在活體內之效應，所以發明者接著採取使用目前所描述之演算法 Dragon Estrogen Response Element Finder(DEREF)之生物資訊學途徑，以在擾動基因之啟動子區域中搜尋假定的雌激素反應要素(ERE)(14)。DEREF 之預測精確度已在大量活體內實例中得以驗證，其在微陣列實驗中以雌激素反應性基因相對於非反應性基因 2.8 倍的頻率在該等基因之啟動子區域中偵測了 ERE 模式，且以屬於雌激素誘導之 SAGE 資料集之基因相對於表現關於乳癌中之 ER 呈陰性之基因 5.4 倍的頻率

在該等基因之啟動子中偵測了ERE模式(補充資訊)。在ER+腫瘤中之頂部50種擾動基因中(表2)，可精確地確定35種基因之轉錄開始部位，且因此隨後藉由DEREF進行分析。在該等35種基因中，僅在12個啟動子中偵測出具有高置信度之ERE(總頻率為34%)(表2)。

相比之下，在ER-腫瘤中之頂部50種擾動基因中，藉由DEREF分析了33種基因，且僅在3種基因中偵測到高置信度ERE(總頻率為9%)(表2)。因此，以ER-腫瘤中之3.7倍高的頻率在ER+腫瘤中之擾動基因的啟動子中偵測ERE。藉由卡方(chi-square)分析可見，此差異係顯著的($p=0.012$)，此表明ERBB2可經由不同的機制影響ER+與ER腫瘤中之轉錄(參見討論)。無論如何，並未偵測到ERE在兩種亞型(ER+與ER-)中的擾動基因中有過度代表，此表明該等基因可能並非ER之直接轉錄目標。該等基因可代表ER的間接目標，或可經由ER無關機制來轉錄地調節。

不管ER亞型而分類低與高置信度腫瘤的最優基因集之界定

此分析之目的係識別可不管腫瘤之ER狀況而用於分類"高"與"低置信度"腫瘤之最優基因集。

細節

總共分析96種腫瘤，其中16種為LC，且80種為HC。使用三種獨立分析方法之系列(SAM、GR與WT，參見下文)來識別在該等兩組(LC與HC)之間經差異調節之基因。使用支持向量機或加權表決法作為分類演算法藉由留一校驗法來評

定該等基因集分類腫瘤之HC或LC狀況的能力。

結果

SAM(微陣列顯著性分析)：以 $< 15\%$ 的FDR(檢出率)，識別了低置信度腫瘤中總共86種上調基因及2種下調基因。藉由使用此基因組，LOOCV檢定產生84%之分類精確度。該等88種基因展示於表A1中。

GR(藉由SVM之基因排列)：總共251種基因經識別具有分類腫瘤之HC或LC狀況之能力，其分類精確度為86%。該等251種基因展示於表A2中。

WT(威爾科克森測試)：在 P 值 < 0.05 且 $a \geq 2$ 倍變化截止值下，識別了總共38種基因。此38種基因之基因組傳遞了80%之LOOCV精確度。該等38種基因展示於表A3中。

然後，識別該等三個基因組(SAM-88，GR-251，WT-38)中之13種"共同"基因。該13員基因藉由LOOCV達成了84%之分類精確度。本質上，該等13種"共同"基因係強健的顯著性標誌，且可達成與其它"完全"標誌集相當之效能。因此，其可視為是"最優"基因。該等13種基因展示於表A4中。

ER陰性"高置信度"相對於"低置信度"腫瘤之臨床成果

此分析之目的係將患有"高置信度"ER陰性腫瘤的患者之臨床預後與患有"低置信度"ER陰性腫瘤的患者相比較。

細節

分析兩個稱作"Rosetta"與"Stanford"資料集之獨立的資料集。Rosetta資料集含有29種ER陰性腫瘤，其中19種為"高置信度"，而10種為"低置信度"。Stanford資料集含有19

種ER陰性腫瘤，其中12種為"高置信度"，且7種為"低置信度"。分析結果展示於圖6(a)與6(b)中。

在兩種情形中，患有"低置信度"腫瘤之患者比患有其"高置信度"對應物之患者展現出更惡化之預後。儘管此差異在統計學上並不顯著，但其可能係由於在該等研究中所分析之患者的數目較少所引起。

討論

此報導中之發現補充且擴展了此領域中有關藉由ER亞型來分類乳腺腫瘤之先前工作。大體而言，該等研究已顯示：儘管可成功地使用基因表現資料來分類大部分腫瘤之ER亞型，但一定存在某些展現低置信度預測且因而無法精確分類之腫瘤群(3,4)。發明者決定藉由深入分析該等"低置信度"腫瘤來研究該等"低置信度"樣品。其得出了大量驚人的發現。其發現：相較於患有"高置信度"腫瘤之患者，患有"低置信度"腫瘤之患者展現出顯著惡化的整體存活及更短的遠距離轉移時間。"高"相對於"低置信度"之分類係藉由計算分析基因表現輪廓而達成，其亦用於將ER+腫瘤分離為展現獨特臨床表現之組(圖2)。因為目前尚無法使用習知免疫組織病理學技術來辨識此等亞組，所以該等結果亦證實藉由表現輪廓與計算分析來分類乳腺腫瘤之ER狀況如何在醫學上具有極大的有用性。

發明者亦得出此驚人發現，即"低置信度"狀況與ERBB2受體之提高的表現顯著相關。然而，其強調：ERBB2與"低置信度"預測之間的聯繫保持著一種關聯性，且在這一點

上，(自其自身的資料)並無跡象表明ERBB2在功能上引起"低置信度"狀態。然而，假設ER與ERBB2係目前乳癌中兩種最具臨床相關性之分子生物標誌，則試圖推測：該等結果表明在乳癌中該等兩種訊號路徑之間存在實質串擾，此亦係其他人已提出的一種可能性(7)。令人感興趣的是，儘管ERBB2+與"低置信度"預測之間的關聯性極其顯著，但其並不完善，因為亦發現某些ERBB2+腫瘤展現出"高置信度"預測，同時並非所有的"低置信度"腫瘤均為ERBB2+。因此，不可能藉由用於偵測ERBB2的習知組織病理學技術(諸如IHC與FISH)來辨識乳腺腫瘤之"低置信度"種群。取而代之，發明者相信：對於藉由常規組織病理學設計ERBB2+之腫瘤而言，進一步檢查該等腫瘤中此等特徵性"表現擾動"之存在係區分以下腫瘤之頗具希望的方法，該等腫瘤相對於將沿相對更遲緩進程發展之彼等腫瘤可能更具臨床侵襲性。

探索此可能性將是未來研究之重要任務。在臨床上，長久以來ER+乳腺腫瘤中之提高的ERBB2表現與對於抗激素療法之減少的敏感度相關聯，且大量實驗論文已報導了可能的解決機制，藉由該等機制ERBB2活性可能引起此效應。一般而言，最風行的模型係其中提高之ERBB2訊號藉由ER基因之轉錄下調(17)、ER的後轉譯修飾(例如，磷酸化)(18)、或經由誘導諸如MTA1之ER結合輔阻遏物(19)來使ER展現減小的轉錄活性。若ERBB2之效應主要係藉由對ER轉錄活性之效應來調節，則吾人可預期其轉錄在ERBB2+

"低置信度"樣品中顯著擾動之基因的實質數目應對應於作為ER之直接目標之基因。然而，發明者發現：在ER+與ER-腫瘤中顯著擾動之基因中之顯著比例先前並未識別為雌激素誘導之基因，且該等基因亦似乎在其啟動子中缺乏潛在ERE。在ER-腫瘤中尤其如此，其中僅發現9%的顯著擾動之基因在其啟動子中含有高置信度假定ERE。儘管發明者無法排除該等擾動基因可為ER的間接目標或可經由非ERE機制而由ER活化之可能性，但是該等發現提出ERBB2活性可以ER無關之方式來調節乳腺腫瘤中顯著部分的基因之可能性。此可經由眾多途徑發生。舉例而言，ERBB2可能藉由活化RAS/MAPK或PI3/Akt路徑來調節除了ER之外的其它轉錄因子(18)。

或者，ERBB2活性可導致可產生更多效的效應之諸如MTA1之染色質因子的誘導(19)。

材料與方法

在已自institution's Repository and Ethics Committees獲得適當核准後，自the Tissue Repository in the institution National Cancer Center of Singapore獲得乳腺組織樣品與患者資料乳腺組織樣品及臨床資料。在外科切除後即刻於手術室中粗解剖樣品，且將其急驟冷凍在流體N₂中。組織學資訊(ER、ERBB2)係由Department of Pathology at Singapore General Hospital提供，且選擇樣品以為每一資料集提供相當數目之ER+與ER-腫瘤(藉由IHC判定)。

如藉由冷凍切片所評定，腫瘤樣品含有>50%之腫瘤含

量。使用 55 種腫瘤 (35 種 ER+ 樣品與 20 種 ER- 樣品) 作為訓練資料，同時單獨一組 41 種腫瘤 (21 種 ER+ 樣品與 20 種 ER- 樣品) 用於盲測試。所有樣品及患者的臨床資料之詳細列表包括在表 S1 中。

樣品製備與微陣列雜交

使用 Trizol 試劑自組織中提取 RNA，且根據製造商的說明使用 U133A 基因晶片對其進行處理以用於 Affymetrix 基因晶片雜交。

資料之預處理

使用基因資料提煉程式 (Genedata Refiner program) 來品質控制生晶片掃描，且將其寄存入中心資料儲存器具中。藉由移除貫穿所有樣品中均缺少表現之基因來預處理表現資料 (意即，"A" 調入)，以使剩餘基因經受 $\log_2$ 轉換，且藉由樣品來調節居中 (mediate-center)。

ER 狀況之預測

使用兩種分類演算法，即加權表決法 (WV) (20) 與支持向量機 (SVM) (21)，以根據 ER 亞型來分類乳腺腫瘤。分類精確度之定義係經正確分類之樣品的數目除以樣品總數。對於 WV 分析而言，分類精確度係使用具有頂部的 50 種 ER 狀況之判別基因之基因集來確定，而基於 SVM 之二分器則利用所有基因。

加權表決法 (WV)：加權表決演算法利用訊雜比 (S2N) 量度來執行二分法。將屬於預測組之各基因指定為一 "表決"，表示為待分類之樣品中之基因表現量與該類別平均表現量

之間的加權差異。加權係使用相關性量度 $P(g,c) = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2}$ 來確

定(其中， μ 與 σ 表示兩種類別中每一類別之基因表現量的平均值與標準偏差)。藉由將類別判別中所用之各基因所得的所有加權表決數相加來計算一特定類別指定之最終表決數。"預測強度"(PS)之定義為： $PS = \frac{V_{WIN} - V_{LOSE}}{V_{WIN} + V_{LOSE}}$ ，其中 V_{WIN} 與

V_{LOSE} 分別為獲勝類別與失敗類別之表決總數。PS反映勝利者之相對邊緣，且因此提供預測確定度之定量化反映。

支持向量機(SVM)：支持向量機係在所利用之試圖最大地分離訓練資料之類別的特徵(基因)空間中界定判別表面之分類演算法(21)。一未知測試樣品相對於判別表面的位置決定其類別。通常在n維基因空間中計算距離，對應於所考慮之基因表現值的總數。發明者使用 SVM-FU(可在 www.ai.mit.edu/projects/cbcl/ 獲得)及線性核來實施SVM分析。如先前所述(22)，每一SVM預測之置信度係基於測試樣品與判別表面之距離。

低置信度腫瘤之識別

由於達成良好預測置信度之臨床重要性，發明者保守地選擇高置信度臨限值以最小化潛在的假陽性分類。基於留一校驗法(LOOCV)之結果，其使用0.4之臨限值，且識別出16種樣品(自總共96種樣品中)係屬於"低置信度"組。若一腫瘤樣品自WV所得之預測強度(PS)小於此臨限值，則將其指定為"低置信度"類別。

經差異表現之基因的選擇與表現擾動之確定的微陣列顯著性分析(SAM)係一種開發用以識別在分離組之間差異表

現之基因的統計學方法(11)。根據基因被調節之統計學可能性來排列基因。SAM演算法亦執行表現資料之置換分析，以估算經識別為因隨機因素而"差異調節"(意即，假陽性)之基因的數目。此數目即為"檢出率"(FDR)。視所要的嚴格度而定，不同的報導已使用在<5%至33%範圍內之FDR(23，24)。

使用 Student t 測試將"高"與"低置信度"組之間的SAM-133基因集中之表現量進行比較。若一基因的p值小於0.05，則將該基因歸類為展現顯著"擾動之表現"。

使用 DEREf(一種計算演算法，即 Dragon ERE Finder (DEREF))之雌激素反應要素(ERE)的計算識別(14)係用於識別作為啟動子中ER之DNA結合部位之假定雌激素反應要素(ERE)(參見 <http://sdmc.lit.org.sg/ERE-V2/index> 中對DEREF之潛在方法論的描述)。基於默認設置，DEREF以83%之敏感度在人類基因組DNA上平均每13,000 nt產生一ERE模式預測。為了降低假陽性之數目，發明者在此報導中施以一額外準則，即17核苷酸之預測ERE模式(14)亦必須(基於BLAST(25)匹配而無容許間隙)與來自至少一其它人類基因啟動子之類似ERE模式匹配，其條件為後一模式可藉由DEREF以97%之敏感度得以預測。在此報導中，亦針對使用 FIE2 程式(26,27)產生且涵蓋相對於基因之5'端的[-3000,+1000]範圍之約11,000個參考人類啟動子序列的資料庫進行ERE搜索。某些待分析之基因並非包含在此啟動子資料庫中，因而未執行對該等基因之ERE搜索。此等基

因在表2中以N/A表示。

在Stanford與Rosetta資料集中具有低預測強度("低置信度")之腫瘤的識別

對兩個獨立的資料集(稱作"Stanford"與"Rosetta"資料集)獨立地進行加權表決與留一校驗法。以類似於圖1之方式繪製其結果，且該等圖展示於圖7中。在兩個資料集中，低置信度腫瘤可被識別為某些點，於該等點處("懸崖點")，腫瘤開始定性地顯現出自大部分腫瘤群減小之預測強度(PS)。儘管對每一資料集進行獨立地分析，但所有資料集之"低置信度"腫瘤的比例均高度相當，其範圍為所有腫瘤之15-19%(圖7(a)中所示之Rosetta資料集為18/117(15.4%)；圖7(b)中所示之Stanford資料集為14/74(18.9%)，吾人之資料集為16/96(16.7%))。

用於產生圖7之資料之不同陣列技術的細節

Stanford資料集：此資料係使用2色cDNA微陣列產生，其中將經PCR擴增之cDNA片段(表示不同基因)自動地寄存於固態基質上以創建微陣列。

Rosetta資料集：此資料係使用2色寡核苷酸微陣列產生，其中將70-80 mer寡核苷酸(表示不同基因)化學地原位合成於固態基質上以創建微陣列。

患者群之細節

Stanford資料集係由78種乳癌(腫瘤)與7種非惡性樣品之cDNA微陣列資料及患者之整體存活資訊組成。

Rosetta資料集係由117種使用基於寡核苷酸之微陣列進

行輪廓分析之早期(淋巴結陰性)乳腺腫瘤組成。

群大小

如上所示，低置信度腫瘤佔據了每一乳腺腫瘤群之約15-19%。為了確信地識別此腫瘤亞種群，較佳需要至少25-30種輪廓、較佳更高(約80-100種腫瘤，如在上述三個資料集中)之最小資料集。

樣品資料

表S7展示了SAM-133基因組之每一基因的加權表決演算法中所使用之平均值(μ)與標準偏差(σ)參數。該等資料可用於在給定SAM-133基因集之基因的表現量集合時將一未知乳腺腫瘤樣品指定為高或低置信度。表2之基因包括在SAM-133基因集中。該等資料專用於施用至來自Affymetrix U133基因晶片之表現資料的加權表決技術。

表S8展示高置信度與低確信樣品上之表A4多基因分類器(13種共同基因)的表現資料。該等資料專用於Affymetrix U133A基因晶片，且已經過資料預處理。表A4多基因分類器之基因表現輪廓可用作訓練資料以建構一預測模型(例如，WV與SVM)，然後該模型可指定未知乳腺腫瘤之置信度。

資料經製表分隔，且具有如下格式：

行：

第一行：預後組基因之探針_ID

第二行：基因名稱

第三行及其它行：基因表現資料

列：

第一列：樣品ID(35種樣品)

第二列：樣品之置信度(高或低)

第三列及其它列：基因表現資料

基因表現資料係如"樣品製備及微陣列雜交"與"資料預處理"中所描述之方式獲得(參見材料及方法部分)。

表S9展示用於表A4基因組之每一基因的加權表決演算法中之平均值(μ)與標準偏差(σ)參數。給定表A4基因集之基因的表現量集合，不管腫瘤之ER狀況，該等資料即可用於將一未知乳腺腫瘤樣品分類為高或低置信度。該等資料專用於施用至來自Affymetrix U133基因晶片之表現資料的加權表決技術。

參考文獻

1. Tavassoli, F. A.及Schnitt S. J. (1992) Pathology of the Breast. 於 (Elsevier)中。
2. Biswas, D.K., Averboukh, L., Sheng, S., Martin, K. Ewaniuk, D.S., Jawde, T.F., Wang, F., Pardee, A.B. (1998) Classification of breast cancer cells on the basis of a functional assay for estrogen receptor. *Mol Med*, 4, 454-467。
3. Gruvberger, S., M. Ringner, Y. Chen, S. Panavally, L. H. Saal, A. Borg, M. Ferno, C. Peterson及P. Meltzer (2001) Estrogen Receptor Status in Breast Cancer is Associated with Remarkably Distinct Gene Expression Patterns. *Cancer Research*, 61, 5979-5984。
4. West, M., Blanchette, C., Dressman, H., Huang, E. , Ishida, S., Spang, R., Zuzan, H. , Olson, J.A. Jr, Marks, J.R., Nevins, J.R. (2001) Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U.S.A.* **98**, 11462-67。
5. Pietras R.J., Arboleda, J., Reese, D.M., Wongvipat, N., Pegram, M.D., Ramos, L., Gorman, C.M., Parker, M.G., Sliwkowski, M.X., Slamon, D.J. (1995) HER-2 tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth in human breast cancer cells. *Oncogene*, 10, 2435-2446。
6. Kurokawa, H.及Arteaga, C.L. (2001) Inhibition of erbB receptor (HER) tyrosine kinases as a strategy to abrogate antiestrogen resistance in human breast cancer. *Clinical Cancer Research*, 12, 4436s-4442s。
7. Bange, J., Zwick, E., 及Ullrich, A. (2001) Molecular targets for breast

cancer therapy and prevention. *Nature Medicine*, 7, 548-552 °

8. Chia, K. S., A. Seow, H. P. Lee, 及K. Shanmugaratnam (2000) Cancer Incidence in Singapore, 1993-1997, 於(Singapore Cancer Registry) 中 °

9. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL. (2001) Gene expression patterns of breast carcinomas distinguish tumour subclasses with clinical implications. *Proc Natl Acad Sci U.S.A.* 98, 10869-74 °

10. Van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530-6 °

11. Tusher, V. G., R. Tibshirani, 及G. Chu (2001) Significance Analysis of Microarrays Applied to the Ionizing Radiation Response. *Proc. Natl. Acad. Sci U.S.A.* **98**, 5116-5121 °

12. Kallioniemi A, Kallioniemi OP, Piper J, Tanner M, Stokke T, Chen L, Smith HS, Pinkel D, Gray JW, Waldman FM. (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U.S.A.* **91**, 2156-60 °

13. Charpentier AH, Bednarek AK, Daniel RL, Hawkins KA, Laflin KJ, Gaddis S, MacLeod MC, Aldaz CM. (2000) Effects of estrogen on global gene expression: identification of novel targets of estrogen action. *Cancer Research*, **60**, 5977-83 °

14. Bajic, V.B., Tan, S.L., Chong, A., Tang, S., Strom, A., Gustafsson, J., Lin, C.Y., Liu, E. (2002) Dragon ERE Finder ver.2: A tool for accurate detection and analysis of estrogen response elements in vertebrate genomes. *Nucleic Acid Res.*, 印刷中。

15. Alizadeh, A.A., M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Truc, Y. Xin, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, L. Lisheng, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, 及 L. M. Staudt (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503-511。

16. Bittner, M., P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendeix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, V. Sondak, N. Hayward, 及 J. Trent (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*, **406**, 536-540。

17. Grunt TW, Saceda M, Martin MB, Lupu R, Dittrich E, Krupitza G, Harant H, Huber H, Dittrich C (1995). Bidirectional interactions between the estrogen receptor and the cerbB-2 signaling pathways: heregulin inhibits estrogenic effects in breast cancer cells. *Int J Cancer*, **63**, 560-567。

18. Stoica GE, Franke TF, Wellstein A, Morgan E, Czubayko F, List HJ, Reiter R, Martin MB, Stoica A (2003) . Heregulin-beta1 regulates the estrogen

receptor-alpha gene expression and activity via the ErbB2/PI 3-K/Akt pathway. *Oncogene*, **22**, 2073-2087 .

19. Mazumdar, A., Wang, R.A., Mishra, S.K., Adam, L., Bagheri-Yarmand, R., Mandal, M., Vadlamudi, R.K., Kumar, R. (2000) Transcriptional repression of oestrogen receptor by metastasis-associated protein 1 corepressor. *Nature Cell Biol*, **3**, 30-37 .

20. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-7 .

21. Vapnik V. (1998) Statistical Learning Theory. Wiley, New York .

22. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. (2001) Multiclass cancer diagnosis using tumour gene expression signatures. *Proc Natl Acad Sci U.S.A.* **98**, 15149-54 .

23. Mueller, A., O'Rourke, J., Grimm, J., Guillemin, K., Dixon, M.F., Lee, A. & Falkow, S. (2003) Distinct gene expression profiles characterize the histopathological stages of disease in Helicobacter-induced mucosa-associated lymphoid tissue lymphoma. *Proc Natl Acad Sci USA*, **100**, 1292 - 1297 .

24. Sanoudou, D., Haslett, J.N., Kho, A.T., Guo, S., Gazda, H.T., Greenberg, S.A., Lidov, H.G.V., Kohane, I.S., Kunkel, L.M., & Beggs, A.H. (2003) Expression profiling reveals altered satellite cell numbers and glycolytic enzyme transcription in nemaline myopathy muscle. *Proc Natl*

Acad Sci U.S.A., **100**, 4666 - 4671 .

25. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W.及Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25**, 3389-3402 .

26. Chong, A., Zhang, G., Bajic, V.B. (2002) Information and sequence extraction around the 5'-end and translation initiation site of human genes, *In Silico Biology*, **2**, 461-465 .

27. Chong, A., Zhang, G., Bajic, V.B. (2003) FIE2: A program for the extraction of genomic DNA sequences around the start and translation initiation site of human genes, *Nucleic Acids Research*, in press .

28. Eisen MB, Spellman PT, Brown PO, Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA.* **95**(25), 14863-14868 .

表 1. 臨床參數與 ER 分類置信度之間的關聯性

訓練資料集(此報導)				Stanford 資料集			
參數	患者編號	平均置信度	P 值	參數	患者編號	平均置信度	P 值
ERBB2			<0.001	ERBB2			<0.001
陽性	18	0.58		陽性	9	0.233	
陰性	37	0.89		陰性	65	0.667	
年齡			0.45	年齡			0.03
<55 歲	25	0.76		<55 歲	33	0.545	
>=55 歲	30	0.81		>=55 歲	41	0.669	
節點			0.98	節點			0.91
0	21	0.787		0	22	0.619	
1-2	30	0.785		1-2	52	0.612	
組織學等級			0.98	組織學等級			0.28
I	7	0.804		I	9	0.727	
II	36	0.784		II	32	0.631	
III-IV	8	0.779		III	32	0.583	
PR			0.03	TP53			0.11
陽性	19	0.88		野生型	38	0.659	
陰性	31	0.71		突變	36	0.567	

表 2. 在 ER+/低與 ER+/高樣品 (a) 之間及在 ER-/低與 ER-/高樣品 (b) 之間顯著擾動之頂部 50 種基因。在 ERE 欄中，"ERE" 指示啟動子含有藉由 DEREf 所預測之高置信度假定 ERE，"非 ERE" 指示未發現假定 ERE，而 "低" 指示發現在中間置信度下之彼啟動子之 ERE。N/A 意指由於無法基於全長轉錄產物確定啟動子之轉錄開始部位故未分析該啟動子。按其高與低置信度樣品之間的 S2N 比率的次序排列基因。

表 2

(a)ER+/低相對於 ER+/高			
基因名稱	UniGene	ERE	排名
雌激素受體 1	Hs.1657	非ERE	1
動力蛋白，纖毛軸絲(axonemal)，輕鏈中間多肽(light intermediate polypeptide)1	Hs.406050	低	2
細胞色素c氧化酶亞單位VIc	Hs.351875	非ERE	3
膜聯蛋白A9	Hs.279928	ERE	4
N-乙醯基轉移酶1(芳基胺N-乙醯基轉移酶)	Hs.155956	ERE	5
細胞色素P450，亞族II B(苯巴比妥可誘導)多肽6	Hs.1360	低	6
視黃酸受體 α	Hs.361071	ERE	7
類胰島素生長因子1受體	Hs.239176	N/A	8
絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員5	Hs.76353	低	9
人類cDNA：FLJ21695 fis，純系COL09653. mRNA序列	Hs.306803	N/A	10
B-細胞CLL/淋巴瘤2	Hs. 79241	ERE	11
GREB 1蛋白	Hs. 193914	非ERE	12
RNB6	Hs.241471	ERE	13
GATA結合蛋白3	Hs.169946	非ERE	14
人類mRNA; cDNA DKFZp564F053(來自純系DKFZp564F053)， mRNA序列	Hs.71968	N/A	15
含有WW域之蛋白1	Hs.355977	非ERE	16
GDNF族受體 α 1	Hs.105445	非ERE	17
染色體1開放讀碼框34	Hs.125783	N/A	18
與AF4相關之淋巴核蛋白	Hs.38070	N/A	19
介白素6訊號轉導子(gp130，制瘤素M受體)	Hs.82065	非ERE	20
G蛋白訊號調節因子11	Hs.65756	ERE	21
類人類胰島素生長因子1受體mRNA，3'序列，mRNA序列	Hs.405998	N/A	22
融合蛋白(hepsin)(跨膜蛋白酶，絲胺酸1)	Hs.823	非ERE	23
Sema域、免疫球蛋白域(Ig)、短鹼性域(short basic domain)分泌之(腦訊號蛋白(semaphorin))3B	Hs.82222	非ERE	24
UDP-葡萄糖神經醯胺葡萄糖基轉移酶；	Hs.432605	ERE	25
細胞色素P450，亞族II B(苯巴比妥可誘導)多肽7	Hs.330780	N/A	26
骨骼，緩慢，肌鈣蛋白T1，	Hs. 73980	N/A	27
微管相關蛋白tau	Hs.101174	非ERE	28

七個失神同系物2(seven in absentia homolog)(果蠅)	Hs.20191	非ERE	29
黃體酮受體	Hs.2905	非ERE	30
KIAA0882蛋白	Hs.90419	N/A	31
假設蛋白FLJ20151	Hs.279916	低	32
ATP-結合盒，亞族A(ABC1)，組員3	Hs.26630	ERE	33
碳酸酐酶XII	Hs.5338	ERE	34
溶質載體蛋白族(solute carrier family)16(一元羧酸轉運子)，組員6	Hs.114924	低	35
假設蛋白FLJ12910	Hs.15929	非ERE	36
假設蛋白FLJ20627	Hs.238270	非ERE	37
毛髮、鼻、指(趾)症候群(trichorhinophalangeal syndrome)1	Hs.26102	非ERE	38
calsyntenin 2	Hs.12079	N/A	39
絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員3	Hs.234726	ERE	40
vav 3致癌基因	Hs.267659	非ERE	41
雌激素調節之LIV-1蛋白	Hs. 79136	N/A	42
人類mRNA；cDNA DKFZp434E082(來自純系DKFZp434E082)，mRNA序列	Hs. 432587	N/A	43
腺苷酸環化酶9	Hs.20196	ERE	44
KIAA0876蛋白	Hs.301011	N/A	45
血紅素結合蛋白1	Hs.294133	ERE	46
司腺鈣蛋白(stanniocalcin)2	Hs.155223	低	47
補體組分4B	Hs.433721	N/A	48
溶質載體蛋白族27(脂肪酸轉運子)，組員2	Hs.11729	N/A.	49
T-盒3(尺骨、乳腺症候群)	Hs.267182	非ERE	50
(b) ER-/低相對於ER-/高			
假設蛋白FLJ20151	Hs.279916	低	1
碳酸酐酶XII	Hs.5338	低	2
GATA結合蛋白3	Hs.169946	非ERE	3
酵母長鏈多元不飽和脂肪酸伸長酶之同系物2	Hs.250175	非ERE	4
含有WW域之蛋白1	Hs.355977	非ERE	5
X盒結合蛋白1	Hs.149923	非ERE	6
特異脂肪2	Hs.74120	低	7
黑色素瘤抗原，族D，2	Hs.4943	N/A	8

前梯度(anterior gradient)2同系物(非洲爪蟾)	Hs. 91011	非ERE	9
細胞色素c氧化酶亞單位VIc	Hs.351875	非ERE	10
醛-酮(aldo-keto)還原酶族7，組員A3(黃麴黴毒素醛還原酶)	Hs.284236	N/A	11
緊密接合蛋白3(tight junction protein 3)(帶狀緊蛋白(zona occludens)3)	Hs.25527	N/A	12
LAG1壽命確保同系物2(釀酒酵母(<i>S. cerevisiae</i>))	Hs.285976	ERE	13
肌醇1,4,5-三磷酸酯受體，類型1	Hs.198443	非ERE	14
果糖-1,6-雙磷酸酯酶1	Hs.574	ERE	15
KIAA0882蛋白	Hs.90419	N/A	16
假設蛋白FLJ12910	Hs. 15929	非ERE	17
雌激素調節之LIV-1蛋白	Hs.79136	N/A	18
甲基巴豆醯基-輔酶A羧基酶2(β)	Hs.167531	非ERE	19
細胞色素P450，亞族IIB(苯巴比妥可誘導)多肽7	Hs.330780	N/A	20
車軸草因子3(腸內)	Hs.82961	低	21
人類純系23948 mRNA序列	Hs.159264	N/A	22
N-乙醯基轉移酶1(芳基胺N-乙醯基轉移酶)	HS.155956	低	23
GREB1蛋白	Hs.193914	非ERE	24
視黃酸誘導3	Hs.194691	非ERE	25
溶質載體蛋白族16(一元羧酸轉運子)，組員6	Hs. 114924	低	26
動力蛋白，纖毛軸絲，輕鏈中間多肽1	Hs.406050	低	27
溶質載體蛋白族7(陽離子胺基酸轉運子，y ⁺ 系統)，組員8	Hs.22891	低	28
WD重複域10	Hs.70202	非ERE	29
calsyntenin 2	Hs.12079	N/A	30
v-myb成髓細胞血症病毒致癌基因同系物(鳥類)	Hs.1334	低	31
車軸草因子1(乳癌，其中表現雌激素可誘導之序列)	Hs. 350470	低	32
假設蛋白MGC2601	Hs.124915	ERE	33
達克斯獵狗(dachshund)同系物(果蠅)	Hs.63931	非ERE	34
跨膜黏蛋白1	Hs.89603	N/A	35
補體組分4B	Hs.433721	N/A	36
富半胱氨酸蛋白1(腸內)	Hs.423190	N/A	37
NPD009蛋白	Hs.283675	低	38
sema域、免疫球蛋白域(Ig)、短鹼性域分泌之(腦訊號蛋白)3B	Hs.82222	非ERE	39
類HRAS抑制因子3	Hs.37189	N/A	40

ATP-結合盒，亞族A(ABC1)，組員3	Hs.26630	低	41
微管相關蛋白tau	Hs.101174	非ERE	42
肌球蛋白VI[人類]，mRNA序列	Hs.385834	N/A	43
CGI-49蛋白	Hs.238126	N/A	44
視黃酸受體 α	Hs.361071	Low	45
vav 3致癌基因	Hs.267659	非ERE	46
染色體1開放讀碼框34	Hs.125783	N/A	
雌激素受體1	Hs.1657	非ERE	48
溶質載體蛋白族27(脂肪酸轉運子)，組員 2	Hs.11729	N/A	49
TBX3-同工蛋白(iso protein)	Hs.332150	N/A.	50

表 S1. 乳腺腫瘤樣品之臨床資訊

表 S1. 吾人資料集之臨床資訊

樣品ID	ER	ERBB2*	PR	年齡	節點	階段	種族
最初收集(55個樣品)							
980177	+	陰性	+	75	2	IIIA	中國人
980178	+	陰性	-	69	1	IIB	中國人
980194	-	陽性	-	58	1	IIB	中國人
980197	+	陽性	+	55	1	IIB	中國人
980203	+	陰性	+	44	0	I	中國人
980208	+	陰性	+	42	1	IIB	中國人
980214	+	陽性	-	49	1	IIIB	中國人
980215	+	陰性	-	54			中國人
980216	-	陰性	-	65	1	IIB	印度人
980217	+	陰性		54	1	IIB	中國人
980220	+	陽性		43	0	IIA	中國人
980221	+	陰性	+	34	1	IV	中國人
980238	-	陽性		62			中國人
980247	-	陰性		35			中國人
980261	+	陰性	-	60			中國人
980338	-	陰性	-	55	0	IIA	中國人
980346	+	陰性	+	54	0	I	中國人
980353	-	陰性	-	59	0	IIA	中國人
980373	-	陽性	-	77	0	IIA	中國人
980380	-	陽性	-	55	0	I	中國人
980383	+	陰性		66	0	IIA	中國人
980391	+	陰性	+	56	0	I	中國人
980395	-	陽性	-	68	1	IIB	中國人
980396	-	陽性	-	66	1	IIB	中國人
980403	+	陰性	+	73	0	IIA	中國人
980404	+	陰性	+	46	1	IIB	中國人
980409	+	陰性	-	48	0	I	中國人
980411	-	陰性	-	72	0	IIA	中國人
980434	+	陰性	+	73	0	IIA	中國人
980441	-	陰性	-	66	1	IIB	中國人

990075	+	陰性	+	66	1	IIB	中國人
990082	+	陰性	+	49	1	IIB	中國人
990107	+	陰性	-	51	1	IIB	印度人
990113	+	陰性	+	70	1	IIIA	中國人
990115	+	陽性	+	38	1	IIB	中國人
990123	+	陰性	+	53	1	IIIA	中國人
990134	-	陽性	-	43	0	IIA	中國人
990148	+	陽性	-	60	1	IIB	中國人
990174	-	陰性	-	56	1	IIB	中國人
990223	+	陽性	-	52	1	IIA	中國人
990262	-	陽性	-	68	1	IIB	中國人
990299	-	陰性	-	58	1	IIIA	中國人
990375	+	陰性	-	38	0	I	中國人
2000209	+	陽性	-	58	0	IIA	中國人
2000422	+	陰性	+	52	1	IIIA	中國人
2000500	-	陰性	-	44	1	IV	中國人
2000683	+	陰性	+	72	0	IIA	中國人
2000759	-	陽性	-	57	0	I	中國人
2000768	+	陰性	+	39	0	IIA	中國人
2000775	+	陰性	-	51	0	IIA	中國人
2000779	+	陰性	-	48	0	IIB	中國人
2000804	+	陰性	+	39	1	IIB	中國人
2000813	-	陽性	-	60	1	IIB	中國人
2000829	-	陽性	-	51	1	IIB	中國人
2000948	+	陰性	-	56	1	IIB	中國人

第二次收集(41個樣品)

980058	+	陰性		72			中國人
980193	-	陰性		49			中國人
980256	-	陰性		46			中國人
980278	+	陰性		64			中國人
980285	-	陰性		49			中國人
980288	+	陽性		45			印度人
980315	-	陰性		59			中國人
980333	+	陰性		51			中國人
980335	-	陽性		33			中國人

2000104	+	陽性		59			中國人
2000171	-	陽性		50			中國人
2000210	-	陽性		50			馬來人
2000215	+	陰性		50			中國人
2000220	+	陰性		52			中國人
2000237	+	陽性		43			中國人
2000272	+	陰性		50			印度人
2000274	+	陰性		40			中國人
2000287	-	陽性		53			中國人
2000320	-	陰性		67			中國人
2000376	-	陽性		65			中國人
2000399	-	陽性		44			中國人
2000401	+	陰性		51			中國人
2000593	-	陰性		60			中國人
2000597	+	陰性		57			中國人
2000609	+	陰性		62			中國人
2000638	-	陰性		60			中國人
2000641	-	陽性		47			馬來人
2000651	+	陰性		45			中國人
2000652	-	陽性		56			中國人
2000675	-	陽性		78			中國人
2000709	-	陽性		45			中國人
2000731	-	陰性		68			印度人
2000787	+	陰性		57,			中國人
2000818	+	陰性		52			中國人
2000880	-	陰性		54			中國人
20020021	+	陰性		64			中國人
20020051	+	陰性		38			馬來人
20020056	+	陰性		71			印度人
20020071	+	陰性		58			中國人
20020090	-	陽性		60			中國人
20020160	+	陰性		82			中國人

*ERBB2狀況之判定：在訓練組(55種樣品)中，ERBB2狀況係藉由習知免疫組織化學法按照表現輪廓分析來判定。

據報導，21種樣品為ERBB2+。對於其它資料集而言，ERBB2狀況係藉由表現輪廓分析及對ERBB2及其它17q連接之基因的分析來判定。

表S2：獨立測試及外部乳癌資料集之分類結果

留一校驗法(LOOCV)：吾人使用標準留一校驗(LOOCV)方法來評定訓練組中的分類精確度。在LOOCV中，最初"省略"訓練組中的一樣品，且在剩餘樣品上執行分類器操作(例如，基因選擇及分類器訓練)。然後，使用經訓練之演算法來分類該"省略"之樣品，且接著對訓練組中的所有樣品重複此過程。

可自 <http://www.omniarray.com/ERClassification.html> 以Excel檔案的形式獲得全部四個資料集(包括PS)之WV分析之輸出及ERBB2表現與預測置信度之關聯性的相應p值。

表S3：對於ER亞型判別而言重要的基因之識別

使用微陣列顯著性分析(SAM)來識別及排列在ER+與ER-腫瘤之間差異調節之133種基因(FDR為0%，≥2倍的表現變化)。其中122種基因在ER+中上調(陽性基因)，且11種基因在ER+中下調(陰性基因)。一特定基因之S2N比率反映低與高置信度樣品之間所觀察到的表現擾動之範圍。

表S3. SAM-133基因列表

122種關於ER+狀況呈陽性之基因					S2N比率	
排名	探針_ID	UG	基因名稱	GB登記號	ER-	ER+
1	205225_at	Hs.1657	雌激素受體1	NM_000125.1	-0.29577	1.273725
2	209603_at	Hs.169946	GATA-結合蛋白3	AI796169_RC	-1.08401	0.863193
3	204508_s_at	Hs.279916	假設蛋白FLJ20151	BC001012.1	-1.78617	0.608118

4	209604_s_at	Hs.169946	GATA結合蛋白3	BC003070.1	-1.45575	0.776251
5	209602_s_at	Hs.169946	GATA結合蛋白3	AI796169_RC	-0.8137	0.654881
6	206754_s_at	Hs.1360	細胞色素P450，亞族IIB(苯巴比妥可誘導)多肽6	NM_000767.2	-0.2593	1.022511
7	203963_at	Hs.5338	碳酸酐酶XII	NM_001218.2	-1.46907	0.598453
8	214164_x_at	Hs.5344	銜接物相關之蛋白錯合物1， γ 1亞單位	BF752277	-1.38937	0.650127
9	212956_at	Hs.90419	KIAA0882蛋白	AI348094_RC	-0.64903	0.68526
10	215867_x_at	Hs.5344	銜接物相關之蛋白錯合物1， γ 1亞單位	AL050025.1	-1.63678	0.613887
11	210735_s_at	Hs.5338	碳酸酐酶XII	BC000278.1	-1.44687	0.484214
12	214440_at	Hs.155956	N-乙醯基轉移酶1(芳基胺N-乙醯基轉移酶)	NM_000662.1	-0.52605	1.043165
13	202089_s_at	Hs.79136	雌激素調節之LIV-1蛋白	MM_012319.2	-0.61899	0.528173
14	210085_s_at	Hs.279928	膜聯蛋白A9	AF230929.1	-0.24463	1.123041
15	205862_at	Hs.193914	KIAA0575基因產物	NM_014668.1	-0.51927	0.883508
16	202088_at	Hs.79136	雌激素調節之LIV-1蛋白	AI635449_RC	-0.5332	0.584697
17	211712_s_at		人類純系MGC：1925，mRNA，完全cds	BC005830.1		
18	206401_s_at	Hs.101174	微管相關蛋白tau	J03778.1	-0.33797	0.700836
19	215304_at	Hs.159264	人類純系23948 mRNA序列	U79293.1	-0.52908	0.19541
20	218195_at	Hs.15929	假設蛋白FLJ12910	NM_024573.1	-0.62769	0.590894
21	212195_at	Hs.71968	人類mRNA；cDNA DKFZp564F053(來自純系DKFZp564F053)	AL049265.1	-0.22898	0.854505
22	203928_x_at	Hs.101174	微管相關蛋白tau	AI870749_RC	-0.35356	0.682993
23	209460_at	Hs.283675	NPD009蛋白	AF237813.1	-0.18444	0.451265
24	212960_at	Hs.90419	KIAA0882蛋白	BE646554_RC	-0.58169	1.072165
25	209443_at	Hs.76353	絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員5	J02639.1	0.065273	0.94045
26	209173_at	Hs.91011	前梯度2(非洲爪蟾)	AF088867.1	-0.80392	-0.25677
27	203071_at	Hs.82222	sema域、免疫球蛋白	NM_004636.1	-0.39014	0.726153

			域(Ig)、短鹼性域分泌之(腦訊號蛋白)3B			
28	203571_s_at	Hs.74120	特異脂肪2	NM_006829.1	-0.81429	0.240008
29	205354_at	Hs.81131	乙酸酞酯N-甲基轉移酶	NM_000156.3	-0.01557	0.074452
30	213712_at	Hs.30504	人類mRNA；cDNA DKFZp434E082(來自純系DKFZp434E082)	BF508639_RC	0.008265	0.522867
31	41660_at		叢集(Cluster)Incl. AL031588: dJ1163J1.1(小鼠跨膜受體 Celsr1之直向同源物，類似於大鼠MEG之含有KIAA0279 LIKE EGF類域之蛋白)			
32	220744_s_at	Hs.70202	WD重複域10	NM_018262.1	-0.48046	0.159954
33	204798_at	Hs.1334	v-myb鳥類成髓細胞血症病毒致癌基因同系物	NM_005375.1	-0.46303	0.284211
34	215552_s_at	Hs.272288	來自染色體6q25.1-26上之純系RP1-6315之人類DNA序列。含有新穎基因之三部分及雌激素受體1之ESR1基因之外顯子(NR3A1，雌二醇受體)，ESTs、STSs及GSSs	AI073549_RC	-0.19227	0.946801
35	209339_at	Hs.20191	七個失神(果蠅)同系物2	U76248.1	-0.0458	0.698282
36	210272_at	Hs.330780	人類細胞色素P450-IIB (h11B3) mRNA，完全cds	M29873.1	-0.58159	0.717949
37	205186_at	Hs.33846	動力蛋白，纖毛軸絲，輕鏈中間多肽	NM_003462.2	-0.49548	1.221071
38	207414_s_at	Hs.170414	成對鹼性胺基酸裂解系統4	NM_002570.1	-0.00943	0.222009
39	205009_at	Hs.1406	車軸草因子1(乳癌，其中表現雌激素可誘導之序列)	NM_003225.1	-0.44277	0.213135
40	203628_at	Hs.239176	類胰島素生長因子1受體	H05812_RC	0.241512	0.748503

41	211323_s_at	Hs.198443	肌醇1,4,5-三磷酸酯受體，類型1	L38019.1	-0.72886	0.116021
42	201825_s_at	Hs.238126	CGI-49蛋白	AL572542_RC	-0.32444	0.398111
43	211234_x_at	Hs.1657	雌激素受體1	AF258449.1	0.268077	0.482442
44	209459_s_at	Hs.283675	NPD009蛋白	AF237813.1	-0.40497	0.048419
45	212196_at	Hs.71968	人類mRNA；cDNA DKFZp564F053(來自純系DKFZp564F053)	AW242916_RC	-0.0843	0.516679
46	203438_at	Hs.155223	司腺鈣蛋白2	AI435828_RC	-0.15925	0.456003
47	217838_s_at	Hs.241471	RNB6	NM_016337.1	0.38602	0.872588
48	204041_at	Hs.82163	單胺氧化酶B	NM_000898.1	0.050799	0.120203
49	203929_s_at	Hs.101174	微管相關蛋白tau	AI056359_RC	-0.27747	0.427658
50	200670_at	Hs.149923	X-盒結合蛋白1	NM_005080.1	-0.83621	0.279976
51	219414_at	Hs.12079	calsyntenin-2	NM_022131.1	-0.47893	0.553864
52	203627_at	Hs.239176	類胰島素生長因子1受體	AI830698_RC	0.088492	0.976305
53	208451_s_at	Hs.278625	補體組分4B	NM_000592.2	-0.42162	0.448767
54	213419_at	Hs.324125	澱粉樣蛋白 β (A4)前驅體蛋白結合，族B，組員2(類Fe65)	U62325.1	-0.01491	-0.06708
55	205768_s_at	Hs.11729	脂肪酸輔酶A連接酶，極長鏈1	NM_003645.1	-0.26778	0.41298
56	204862_s_at	Hs.81687	非轉移性細胞3，其中表現蛋白	NM_002513.1	-0.24568	0.320418
57	210480_s_at	Hs.22564	肌球蛋白VI	U90236.2	-0.3344	-0.15111
58	205696_s_at	Hs.105445	GDNF族受體 α 1	NM_005264.1	0.013863	0.846687
59	203685_at	Hs.79241	B-細胞CLL淋巴瘤2	NM_000633.1	0.385651	0.915025
60	218976_at	Hs.260720	含有J域之蛋白1	NM_021800.1	-0.17876	0.280663
61	219197_s_at	Hs.222399	CEGP1蛋白	AI424243_RC	-0.09661	0.157384
62	202996_at	Hs.82520	聚合酶(DNA-引導)， δ 4	NM_021173.1	0.158087	0.060137
63	205734_s_at	Hs.38070	與AF4相關之淋巴核蛋白	AI990465_RC	0.187651	0.796703
64	211235_s_at	Hs.1657	雌激素受體1	AF258450.1	0.269909	0.7271
65	211000_s_at	Hs.82065	介白素6訊號轉導子(gp130，制瘤素M受體)	AB015706.1	0.204138	0.785104
66	217190_x_at	Hs.247976	雌激素受體 {外顯子	S67777	0.17102	0.653981

			6}人類他莫昔芬抵抗 乳腺腫瘤17，基因組 突變體，187 nt			
67	202752_x_at	Hs.22891	溶質載體蛋白族 7(陽 離子胺基酸轉運子， γ +系統)，組員 8	NM_012244.1	-0.48423	0.153806
68	201754_at	Hs.74649	細胞色素c氧化酶亞 單位VIc	NM_004374.1	-0.79843	1.207003
69	204623_at	Hs.82961	車軸草因子3(腸內)	NM_003226.1	-0.53903	0.149093
70	207038_at	Hs.114924	溶質載體蛋白族16(一 元羧酸轉運子)，組員6	NM_004694.1	-0.50672	0.593732
71	212637_s_at	Hs.324275	人類mRNA；cDNA DKFZp434D2111(來 自純系 DKFZp434D2111)	AU155187_RC	-0.851	0.852788
72	208682_s_at	Hs.4943	肝細胞癌相關蛋白； 乳癌相關基因1	AF126181.1	-0.80969	-0.06845
73	218502_s_at	Hs.26102	毛髮、鼻、指(趾)症候 群I	NM_014112.1	-0.26191	0.571226
74	202376_at	Hs.234726	絲胺酸(或半胱胺酸) 蛋白酶抑制劑，分化 枝 A(α -1抗蛋白酶， 抗胰蛋白酶)，組員3	NM_001085.2	0.02888	0.549323
75	215616_s_at	Hs.301011	KIAA0876蛋白	AB020683.1	-0.00184	0.507129
76	211233_x_at	Hs.1657	雌激素受體1	M12674.1	0.360947	0.949046
77	205081_at	Hs.17409	富半胱胺酸蛋白1(腸 內)	NM_001311.1	-0.41153	-0.05483
78	214428_x_at	Hs.170250	補體組分4A	K02403.1	-0.22882	0.346824
79	209696_at	Hs.574	果糖-1,6-雙磷酸酯酶 1	D26054.1	-0.68072	0.137814
80	219682_s_at	Hs.332150	TBX3-同工蛋白	NM_016569.1	-0.26452	0.412502
81	212496_s_at	Hs.301011	KIAA0876蛋白	BE256900	-0.272	0.841331
82	203108_at	Hs.194691	視黃酸誘導3	NM_003979.2	-0.51766	0.212322
83	206107_at	Hs.65756	G-蛋白訊號之調節因 子11	NM_003834.1	-0.0233	0.778074
84	218806_s_at	Hs.267659	vav 3致癌基因	AF118887.1	-0.3126	0.544105
85	209581_at	Hs. 37189	類似於大鼠HREV107	BC001387.1	0.37261	0.359298
86	213412_at	Hs. 25527	緊密接合蛋白3(帶狀 緊蛋白3)	NM_014428.1	-0.76231	0.227893

87	212638_s_at	Hs.324275	人類mRNA；cDNA DKFZp434D2111(來自純系 DKFZp434D2111)	BF131791	-0.76733	0.888627
88	206469_x_at	Hs.284236	醛-酮還原酶家族7，組員 A3(黃麴黴毒素醛還原酶)	NM_012067.1	0.77705	0.278936
89	210652_s_at	Hs.125783	DEME-6蛋白	BC004399.1	-0.29655	0.806265
90	216381_x_at	Hs.284236	醛-酮還原酶族7，組員 A3(黃麴黴毒素醛還原酶)	AL035413	-0.61275	0.253454
91	216092_s_at	Hs.22891	溶質載體蛋白族7(陽離子胺基酸轉運子，y ⁺ 系統)，組員8	AL365347.1	-0.67193	0.152525
92	208788_at	Hs.250175	酵母長鏈多元不飽和脂肪酸伸長酶之同系物2	AL136939.1	-0.87121	0.346787
93	204792_s_at	Hs.111862	KIAA0590基因產物	NM_014714.1	0.085973	0.134751
94	207847_s_at	Hs. 89603	跨膜黏蛋白1	NM_002456.1	-0.42941	-0.24975
95	213201_s_at	Hs.73980	骨骼，緩慢，肌鈣蛋白 T1	AJ011712	-0.11892	0.71764
96	204497_at	Hs.20196	腺苷酸環化酶 9	AB011092.1	0.007184	0.509774
97	222314_x_at	Hs.205660	ESTs	AW970881_RC	-0.1322	0.201872
98	222212_s_at	Hs.285976	腫瘤轉移抑制因子	AK001105.1	-0.74148	0.357607
99	219919_s_at	Hs.279808	假設蛋白FLJ10928	NM_018276.1	0.085456	0.152147
100	214053_at	H5.7888	人類純系23736 mRNA序列	AW772192_RC	-0.21533	0.32841
101	204934_s_at	Hs.823	融合蛋白酶(跨膜蛋白酶，絲氨酸1)	NM_002151.1	-0.03851	0.743961
102	216109_at	Hs. 306803	人類cDNA：FLJ21695 fis，純系COL09653	AK025348.1	-0.03594	0.921802
103	203749_s_at	Hs. 250505	視黃酸受體 α	AI806984_RC	-0.3159	1.006049
104	220329_s_at	Hs.238270	假設蛋白FLJ20627	NM_017909.1	0.068053	0.588123
105	204881_s_at	Hs.152601	UDP-葡萄糖神經醯胺葡萄糖基轉移酶	NM_003358.1	-0.248	0.724338
106	208305_at	Hs.2905	黃體酮受體	NM_000926.1	0.145722	0.687258
107	209623_at	Hs 167531	甲基巴豆醯基-輔酶A羧基酶2(β)	AW439494_RC	-0.61293	0.369239
108	218450_at	Hs.108675	血紅素結合蛋白	NM_015987.1	-0.07982	0.486745

109	204343_at	Hs.26630	ATP-結合盒，亞族A(ABC1)，組員 3	NM_001089.1	-0.36256	0.648789
110	219051_x_at	Hs.124915	假設蛋白MGC2601	NM_024042.1	-0.43578	0.112222
111	205471_s_at	Hs.63931	達克斯獵狗(果蠅)同系物	AW772082_RC	-0.43168	-0.26408
112	203439_s_at	Hs.155223	司腺鈣蛋白2	BC000658.1	-0.28836	0.67174
113	204863_s_at	Hs.82065	介白素6訊號轉導子(gp130，制瘤素M受體)	BE856546_RC	0.259289	0.691633
114	203289_s_at	Hs.19699	與 α 珠蛋白簇相調聚之保守基因	BE791629	-0.18036	0.122646
115	221765_at	Hs.23703	ESTs	AI378044_RC	-0.0539	0.714017
116	219001_s_at	Hs.317589	假設蛋白MGC10765	NM_024345.1	-0.28755	0.64098
117	220581_at	Hs. 287738	假設蛋白FLJ23305	NM_025059.1	-0.13763	0.781039
118	211596_s_at		膜糖蛋白LIG-1之人類mRNA，完全cds	AB050468.1		
119	205645_at	Hs.80667	含有RALBP1相關之Eps域之2	NM_004726.1	-0.29164	0.308819
120	219663_s_at	Hs.157527	假設蛋白MGC4659	NM_025268.1	0.059072	-0.06016
121	205380_at	Hs.15456	含有域PDZ之1	NM_002614.1	0.094959	0.486972
122	201508_at	Hs.1516	類胰島素生長因子-結合蛋白質4	NM_001552.1	0.102433	0.237825

11種關於ER+狀況呈陰性之基因

1	215729_s_at	Hs. 9030	TONDU	BE542323	0.729732	-0.40161
2	201983_s_at	Hs.77432	表皮生長因子受體(鳥類成紅血細胞白血病病毒(v-erb-b)致癌基因同系物)	AW157070_RC	0.183968	-0.10873
3	204914_s_at	Hs.32964	SRY(性別決定區Y)-盒11	AW157202_RC	-0.3552	-0.61822
4	204913_s_at	Hs.32964	SRY(性別決定區Y)-盒11	AI360875_RC	-0.54222	-0.6594
5	205646_s_at	Hs.89506	成對盒基因6(無虹膜，角膜炎)	NM_000280.1	0.667994	-0.15217
6	207030_s_at	HS.10526	富半胱氨酸及甘氨酸蛋白2	NM_001321.1	0.526203	-0.44193
7	204915_s_at	Hs 32964	SRY(性別決定區Y)-盒11	AB028641.1	-0.4419	-0.47414
8	203021_at	Hs.251754	分泌性白細胞蛋白酶抑制劑(抗白細胞蛋白酶)	NM_003064.1	-0.08293	-1.00559
9	209800_at	Hs.115947	角蛋白16(病灶性(focal)非表皮鬆懈性掌蹠角皮症)	AF061812.1	0.573263	-0.29962

10	203234_at	Hs.77573	尿苷磷酸化酶	NM_003364.1	0.30456	0.307505
11	201984_s_at	Hs.77432	表皮生長因子受體(鳥類成紅血細胞白血病毒(v-erb-b)致癌基因同系物)	NM_005228.1	0.416409	0.086073

關於ER+狀況呈陰性之頂部54種ER判別基因

由於ER陰性基因之數目有限，所以吾人降低了SAM之臨限值以獲得具有0%FDR之54種基因。該等陰性基因用於圖2c)與d)。

表S4：比較"高"與"低置信度"腫瘤之整體表現輪廓

使用SAM來識別在a)ER+"高"與"低"置信度腫瘤之間及b)ER-"高"與"低"置信度腫瘤之間差異調節之基因。對於ER+比較而言，相較於ER+/高腫瘤，識別出50種基因在ER+/低腫瘤中上調且39種基因下調。對於ER-比較而言，相較於ER-/高腫瘤，識別出50種基因在ER-/低腫瘤中上調，且未識別出任何基因下調。

表S4. 在低/高置信度樣品中差異表現之排列於頂部的基因

a)ER+/低相對於ER+/高

在ER+/低中上調之基因	UniGene	排名	染色體
鈣激活之氯化物通道，族組員2	Hs.241551	1	
ESTs，較弱地類似於人類假設蛋白	Hs.106642	2	
v-myc鳥類骨髓細胞瘤病毒相關之致癌基因，衍生自神經母細胞瘤	Hs.25960	3	
苯乙醇胺N-甲基轉移酶	Hs.1892	4	17q21-q22
具有鋅指域之Alu結合蛋白	Hs.289104	5	
成纖維細胞生長因子受體4	Hs. 165950	6	
KIAA0300蛋白	Hs.173035	7	
生長因子受體結合蛋白7	Hs.86859	8	17q21.1
肌球蛋白，重鏈多肽4，骨骼肌	Hs.272207	9	

阿撲黏蛋白(apomucin)	Hs.103707	10	
脯胺酸氧化酶同系物	Hs. 274550	11	
S100鈣結合蛋白A8(鈣粒蛋白A)	Hs. 100000	12	
甘胺酸C-乙醯基轉移酶(2-胺基-3-酮基丁酸酯輔酶A連接酶)	Hs.54609	13	
受磷蛋白	Hs.85050	14	
CGI-96蛋白	Hs.239934	15	
瘦蛋白(鼠科肥胖症同系物)	Hs. 194236	16	
假設蛋白FLJ14146	Hs.103395	17	
犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)	Hs.107318	18	
抑制素 β B(活化素AB β 多肽)	Hs.1735	19	
羥基類固醇(17 β)脫氫酶2	Hs.155109	20	
脂肪酸結合蛋白7，腦	Hs.26770	21	
血清類黏蛋白2	Hs.278388	22	
分泌性白細胞蛋白酶抑制劑(抗白細胞蛋白酶)	Hs.251754	23	
肌動蛋白 γ 2，平滑肌，腸	Hs.78045	24	
人類mRNA；cDNA DKFZp564G112(來自純系DKEp564G112)	Hs.51515	25	
肽基精胺酸脫亞胺酶(peptidylarginine deiminase)類型III	Hs.149195	26	
肌球蛋白，重鏈多肽 11，平滑肌，	Hs.78344	27	
S100鈣結合蛋白A9(鈣粒蛋白B)	Hs.112405	28	
人類純系23809 mRNA序列	Hs.6932	29	
整合素 β 6	Hs.123125	30	
脂多糖-結合蛋白	Hs.154078	31	
離子型麩胺酸受體，AMPA 3	Hs.100014	32	
人類PAC純系RP5-1093O17，來自7q11.23-q21	Hs. 193606	33	
KIAA1102蛋白	Hs. 202949	34	
跨膜4超族組員3	Hs.84072	35	
v-erb-b2鳥類成紅血細胞白血病毒致瘤基因同系物2(衍生自神經膠母細胞瘤之致瘤基因同系物)	Hs.323910	36	17q11.2-q12
蛋白磷酸酯酶1，調節性(抑制劑)亞單位1A	Hs.76780	37	
HGC6.1.1蛋白	Hs.225962	38	
黏蛋白及類鈣黏著蛋白	Hs.165619	39	
同源盒A9	Hs.127428	40	
4-羥基苯基丙酮酸雙加氧酶	Hs.2899	41	

乳運鐵蛋白	Hs.105938	42	
KIAA1069蛋白	Hs.193143	43	
葉酸水解酶(前列腺特異性膜抗原)1	Hs.1915	44	
精胺基琥珀酸合成酶	Hs.160786	45	
角蛋白7	Hs.23881	46	
血管緊張素受體2	Hs.3110	47	
類鈣調蛋白皮膚蛋白	Hs.180142	48	
電子轉移黃素蛋白， α 多肽(戊二酸尿II)	Hs.169919	49	
S100鈣結合蛋白A7(牛皮癬素1)	Hs.112408	50	

在ER+/低中下調之基因

佛波醇12-肉豆蔻酸酯13-乙酸酯誘導之蛋白1	Hs.96	1
動力蛋白，纖毛軸絲，輕鏈中間多肽	Hs.33846	2
細胞色素P450，亞族IIB(苯巴比妥-可誘導)，多肽6	Hs.1360	3
雌激素受體1	Hs.1657	4
艾特民(artemin)	Hs.194689	5
癌胚抗原相關之細胞黏附分子1(膽汁糖蛋白)	Hs.50964	6
ESTs	Hs.23703	7
KIAA0575基因產物	Hs.193914	8
視黃酸受體 α	Hs.250505	9
膜聯蛋白A9	Hs.279928	10
Cas-Br M(鼠科)異位逆轉錄病毒轉換序列c(ectropic retroviral transforming sequence c)	Hs.156637	11
GATA結合蛋白質3	Hs.169946	12
假設蛋白FLJ12650	Hs.4243	13
砷酸鹽抗性蛋白ARS2	Hs.111801	14
亨廷頓相互作用蛋白(huntingtin interacting protein)2	Hs.155485	15
假設蛋白FLJ13134	Hs.99603	16
鋅指蛋白165	Hs.55481	17
人類cDNA: FLJ21695 fis, 純系COL09653	Hs.306803	18
類胰島素生長因子1受體	Hs.239176	19
融合蛋白酶(跨膜蛋白酶，絲氨酸1)	Hs.823	20
二孔鉀通道KT3.3	Hs.203845	21
UDP-葡萄糖神經鹽胺葡萄糖基轉移酶	Hs.152601	22
人類染色體P450-IIB (hIIB3) mRNA, 完全cds	Hs.330780	23
sema域、免疫球蛋白域(Ig)、短鹼性域分泌之(腦訊號蛋白)3B	Hs.32981	24

微管相關蛋白tau	Hs.101174	25
磷脂醯基絲胺酸特異性磷脂酶A1 α	Hs.17752	26
類似於假設蛋白PRO2831 [人類]，mRNA序列	Hs.406646	27
細胞色素c氧化酶亞單位VIc	Hs.74649	28
腺苷酸環化酶9	Hs.20196	29
類人類細胞因子核因子n-pac mRNA，完全cds	Hs.331584	30
來自染色體6q25.1-26上的純系RP1-63I5之人類DNA序列。含有新穎基因之三部分及雌激素受體1的ESR1基因之外顯子(NR3A1，雌二醇受體)，ESTs、STSs及GSSs	Hs.272288	31
calsyntenin-2	Hs.12079	32
介白素6訊號轉導子(gp130，制瘤素M受體)	Hs.82065	33
A激酶(PRKA)錨定蛋白10	Hs.75456	34
N-乙醯基轉移酶1(芳基胺N-乙醯基轉移酶)	Hs.155956	35
假設蛋白FLJ13687	Hs.278850	36
胱抑蛋白(cystatin)SA	Hs.247955	37
熱衝擊27kD蛋白1	Hs.76067	38
突觸小泡磷酸酶(synaptojanin)2	Hs.61289	39

b) ER-/低相對於ER-/高

在ER-/低中上調之基因	UniGene	排名	染色體
UDP-N-乙醯基- α -D-半乳糖胺多肽N-乙醯基胺基半乳糖基轉移酶6(GalNAc-T6)	Hs.151678	1	
醛脫氫酶4族，組員A1	Hs.77448	2	
染色體6開放讀碼框29	Hs.334514	3	
黑色素瘤抗原，族D，2	Hs.4943	4	
苯乙醇胺N-甲基轉移酶	Hs.1892	5	17q21-q22
含有三重基序之蛋白3	Hs.321576	6	
假設基因MGC9753	Hs.91668		
ATP-結合盒，亞族C(CFTR/MRP)，組員6	Hs.274260	8	
富SH3域結合麩胺酸之類蛋白	Hs.14368	9	
生長因子受體結合蛋白7	Hs.86859	10	17q21.1
3-羥基-3-甲基戊二醯-輔酶A合酶2(線粒體)	Hs.59889	11	
成纖維細胞生長因子受體4	Hs.165950	12	
脂肪酸合酶	Hs.83190	13	
跨膜黏蛋白1	Hs.89603	14	
phafin 2	Hs. 29724	15	

肉鹼乙醯基轉移酶	Hs.12068	16	
假設蛋白FLJ20151	Hs.279916	17	
GATA結合蛋白3	Hs.169946	18	
包含WW域之蛋白1	Hs. 355977	19	
轉錄因子AP-2 β (活化增強子結合蛋白2 β)	Hs.33102	20	
KIAA0882蛋白	Hs.90419	21	
四穿膜區(tetraspan)1	Hs. 38972	22	
過氧化物酶體生源因子(peroxisomal biogenesis factor)11A	Hs.31034	23	
溶質載體蛋白族4，碳酸氫鈉共轉運子，組員8	Hs.132136	24	
假設基因MGC9753	Hs.91668	25	
叉頭盒(forkhead box)A1	Hs.70604	26	
水通道蛋白3	Hs.234642	27	
v-erb-b2成紅血細胞白血病毒致癌基因同系物2，衍生自神經/膠母細胞瘤之致癌基因同系物(鳥類)	Hs.323910	28	17q11.2-q12
肌醇1,4,5-三磷酸酯受體，類型1	Hs.198443	29	
假設蛋白PRO1489	Hs.197922	30	
醛脫氫酶3族，組員B2	Hs.87539	31	
假設蛋白[人類]，mRNA序列	Hs.381412	32	
雙特異型磷酸酯酶6	Hs.180383	33	
碳酸酐酶XII	Hs.5338	34	
NAD(P)H脫氫酶，醌1	Hs.406515	35	
甘露糖苷酶 α ，類別1C，組員1	Hs.8910	36	
KIAA0703基因產物	Hs 6168	37	
硬脂醯基輔酶A去飽和酶(δ -9-去飽和酶)	Hs.119597	38	
果糖-1,6-雙磷酸酯酶1	Hs.574	39	
硫酸芳酯酶D	Hs.326525	40	
X-盒結合蛋白1	Hs.149923	41	
甲基巴豆醯基-輔酶A羧基酶2(β)	Hs.167531	42	
突觸體小相關蛋白23kDa	Hs.184376	43	
類海怪(kraken-like)	Hs.301947	44	
前梯度2同系物(非洲爪蟾)	Hs.91011	45	
假設蛋白FLJ20174	Hs.114556	46	
含有伴侶蛋白之TCP1，亞單位2(β)	Hs.432970	47	
免疫球蛋白重鏈常數 γ 3(G3m標誌)	Hs.300697	48	
跨膜4超族組員3	Hs.84072	49	
山梨醇脫氫酶	Hs.878	50	

使用 DRAGIN-ERE Finder(DEREF)來識別基因啟動子中的假定ERE

使用 DEREf 演算法來界定屬於各種類別之基因之啟動子中的潛在 ERE(參見 <http://sdmc.lit.org-sg/ERE-V2/index> 中對 DEREf 的潛在方法論之描述)。參考文獻 14 之原稿可經由 <http://www.omniarray.com/ERClassification.html> 存取。使用 3 小時倍數增加 ≥ 2 及 3 小時 p 值 < 0.005 之臨限值，自 (<http://143.111.133.249/ggeg/>，參見參考文獻 13) 獲得雌激素誘導之 SAGE 資料集。選出 65 個 SAGE 標籤。該等 65 個 SAGE 標籤匹配 68 種基因，使該等 68 種基因進一步經受 ERE 分析。關於 ER 狀況呈陰性之頂部 100 種基因之基因集係使用 SAM 獲得。表 S6a 描繪了其結果。

表 S6a：各資料集上之 ERE 預測：E2-誘導之 SAGE 資料集；關於 ER+ 呈陰性之基因及 SAM-133 基因集

資料集	非ERE	低	高	具有高置信度之ERE命中率	'N/A'
E2-誘導之SAGE	21	15	21	41.18%	11
ER-陰性基因	50	22	6	7.69%	22
SAM-133	15	15	17	36.17%	23

表 S6b. 文章主體部分之表 2 中所列之基因經由 DEREf 所預測之 ERE 模式

表 2 之 ERE 模式

基因名稱	排名	ERE 模式
在 ER+ 中擾動之 50 種基因中具有高置信度之 12 種 ERE		
膜聯蛋白 A9	4	PP 2783 CA-GGGCA-CCC-CAGCC-TG 新 CCTGTTGGGGCACATACCAGCAGGGCACCCCAGCCT GCACCCAGAGGGGGTCCCAG 21

N-乙醯基轉移酶1(芳基胺N-乙醯基轉移酶)	5	PP 150 AA-GGTGA-CAA-TAACC-AA 新 CCACCTTCAAATCATACTACAAGGTTACAATAACCAA AACAGCGTGGTACTGATACA 21
視黃酸受體 α	7	PP 2149 GA-GGTCC-CTC-TGCCC-CT 新 TGAAGTTGATCTGTTGTATTGAGGTCCCTCTGCCCCCT ATATTTATCCTAAATGGTAT 21
B-細胞CLL/淋巴瘤2	11	PP 647 CA-GGGCA-CAG-TGGCT-CA 新 GACAAAATAAAGATGTCAGGCAGGGCACAGTGGCTC ATGTCTGTAATCCCAGCACTT 21
RNB6	13	PP 1920 TT-GGTCA-GGC-TGGTC-TC 已知 AAAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCT CGAACTTCTGACCTCAGGTGA 21
G-蛋白訊號之調節因子11	21	PP 847 CG-GGTCA-CTG-CAACC-TC 新 GGAGTGCAATGGTGCAATCTCGGGTCACTGCAACCT CCGCCTCCTGGGTTCAAGCGA 21
UDP-葡萄糖神經醯胺葡萄糖基轉移酶	25	PP 466 TG-AGTCA-CCG-TGCCC-AG 新 AAGTGCTGGGATTACAGGCGTGAGTCACCGTGCCCA GCCAATGGCTTGTGGTTTTCT 21
ATP-結合盒，亞族A (ABC1)，組員3	33	PP 1363 CA-GGGCA-CAG-TGGCT-CA 新 GCACAGAGATAAAACCTCGGCAGGGCACAGTGGCTC ACGCCTGTAATCCCCACACTT 21
碳酸酐酶XII	34	PP 1376 TA-GGCCA-AAC-TAACC-TT 新 TCCTTATTCATTCCTGGGCATAGGCCAAACTAACCTT AGAAAGGAATTCAGTTTATG 21
絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員3	40	PP 2408 TT-GGTCG-GAC-TGGTC-TT 新 AGAGACAGGGTTTCACCTTGTTGGTCGGACTGGTCT TGAACCTCCTGACCTCGTGATC 21
腺苷酸環化酶9	44	PP 710 TT-GGTCA-GGC-TGGTC-TC 已知 AGAGATGGGGTTTCTCCGTGTTGGTCAGGCTGGTCT CGAACTCCCGACCTCAGGTGA 21
血紅素結合蛋白1	46	PP 1738 GA-GGTCC-GGG-TGGCC-GC 新 AAAGAGCAGAGGCGCCCGTAGAGGTCCGGGTGGCC GCTGCTGTAAACATCCATCACT 21
在ER-中擾動之50種基因中具有高置信度之3種ERE		
LAG1壽命確保同系物 2(釀酒酵母)	13	PP 3662 CA-GGCCA-GGG-CAACC-CC 新 CCCAAGCCACAGGACGCGTCCAGGCCAGGGCAACC CCGCGGGCCGCTGCCAGGGTGG 21
果糖-1,6-雙磷酸酯酶1	15	PP 776 TT-GGTCA-GGC-TGGTC-TC 已知 AGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCT CGAGCTCCCAACCTCAGGTGA 21
假設蛋白MGC2601	33	PP 966 CT-GGTCA-GGC-TGGTC-TT 新 AGAGACGAGGTTTCTCCATGCTGGTCAGGCTGGTCT TGAACCTCCGACCTCAGGTGA 21

表 S7：SAM-133 基因集之表現資料的平均值(μ)與標準偏差(σ)之加權表決參數

_ID	基因名稱	ER-		ER+	
		平均值	SD	平均值	SD
0_at	X-盒結合蛋白1	0.786506	0.716285	4.265411	1.422852
8_at	類胰島素生長因子結合蛋白4	-0.34357	1.388805	2.57045	0.925761
4_at	細胞色素c氧化酶亞單位VIc	-1.58027	1.870693	1.927493	1.237708
5_s_at	CGI-49蛋白	3.371655	1.153737	5.720964	0.582412
3_s_at	表皮生長因子受體(鳥類成紅血細胞白血病毒(v-erb-b)致癌基因同系物)	-0.23687	1.75591	2.753161	0.803569
4_s_at	表皮生長因子受體(鳥類成紅血細胞白血病毒(v-erb-b)致癌基因同系物)	-1.44281	0.960058	2.42027	2.337701
8_at	雌激素調節之LIV-1蛋白	1.312524	1.221556	3.870357	0.929939
9_s_at	雌激素調節之LIV-1蛋白	1.734565	1.093064	4.085214	0.81537
6_at	絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員3	2.023548	1.032196	4.420661	0.934515
2_x_at	溶質載體蛋白族7(陽離子胺基酸轉運子，y ⁺ 系統)，組員8	1.981605	1.049118	4.149982	0.712426
6_at	聚合酶(DNA-引導)， δ 4	0.786499	1.029001	3.014232	0.865812
1_at	分泌性白細胞蛋白酶抑制劑(抗白細胞蛋白酶)	0.355523	0.675879	3.16287	1.761351
1_at	sema域、免疫球蛋白域(Ig)、短鹼性域分泌之(腦訊號蛋白)3B	1.825558	0.726706	4.052804	1.145816
8_at	視黃酸誘導3	-2.75146	0.887259	-0.09227	1.606679
4_at	尿苷磷酸化酶	-2.68964	1.552946	0.243702	1.641435
9_s_at	與 α 珠蛋白簇相調聚之保守基因	3.20195	0.718557	5.197518	0.987453
8_at	司腺鈣蛋白2	-1.29648	1.055361	0.795528	0.993152
9_s_at	司腺鈣蛋白2	-1.57332	1.345545	0.998514	1.454402
1_s_at	特異脂肪2	0.233895	0.988328	2.283714	1.060332
7_at	類胰島素生長因子1受體	0.141016	0.610073	2.127288	1.174363
8_at	類胰島素生長因子1受體	2.29995	0.509475	3.833107	0.788714

5_at	B-細胞CLL淋巴瘤2	-1.10751	1.324287	1.15701	1.355875
9_s_at	視黃酸受體 α	-1.58118	1.167735	0.537334	1.268906
8_x_at	微管相關蛋白tau	0.359852	0.516477	1.888305	0.821962
9_s_at	微管相關蛋白tau	-2.59884	0.565755	-0.00962	2.145673
3_at	碳酸酐酶XII	1.190756	3.229512	4.402	1.181501
1_at	單胺氧化酶B	-3.13061	1.085626	-0.75919	1.755041
3_at	ATP-結合盒，亞族A(ABC1)，組員 3	-0.29571	1.843682	2.228971	1.512369
7_at	腺苷酸環化酶9	-2.34613	1.534418	-0.05573	1.429526
8_s_at	假設蛋白FLJ20151	-3.52135	1.303031	-0.87495	2.10528
3_at	車軸草因子3(腸內)	-0.37083	1.33889	1.50405	0.899477
2_s_at	KIAA0590基因產物	-0.9475	1.745737	1.257564	1.170708
8_at	v-myb鳥類成髓細胞血症病毒致癌基因同系物	1.288571	1.107004	3.060625	0.97928
2_s_at	非轉移性細胞3，其中表現蛋白	-1.44821	0.786716	0.388854	1.271171
3_s_at	介白素6訊號轉導子(gp130，制瘤素M受體)	-0.10956	1.179102	1.970259	1.431009
1_s_at	UDP-葡萄糖神經醯胺葡萄糖基轉移酶	-1.39262	1.195462	1.156751	2.153286
3_s_at	SRY(性別決定區Y)-盒11	-2.53383	1.536914	-0.16571	1.727001
4_s_at	SRY(性別決定區Y)-盒11	-1.8799	1.273909	0.144791	1.375233
5_s_at	SRY(性別決定區Y)-盒11	0.484505	1.125341	2.823356	1.941558
4_s_at	融合蛋白酶(跨膜蛋白酶，絲胺酸1)	0.462278	0.985428	2.501289	1.570414
9_at	車軸草因子1(乳癌，其中表現雌激素可誘導之序列)	-1.98675	1.39922	-0.14861	0.959657
1_at	富半胱氨酸蛋白1(腸內)	0.366598	1.124549	1.87895	0.590829
6_at	動力蛋白，纖毛軸絲，輕鏈中間多肽	-2.39302	0.959482	-0.48343	1.433455
5_at	雌激素受體1	-1.62943	1.558096	0.486988	1.459551
4_at	乙酸酞酯N-甲基轉移酶	0.719039	0.547264	2.096279	0.868384
0_at	含有PDZ域之1	-0.92507	1.254295	1.252606	1.789471
1_s_at	達克斯獵狗(果蠅)同系物	1.676963	0.591793	3.169036	1.05951
5_at	含有RALBP1相關之Eps域之2	-0.63258	1.838056	2.053427	2.368533

6_s_at	成對盒基因6(無虹膜，角膜炎)	-0.06075	0.836545	1.524428	1.119938
6_s_at	GDNF族受體 $\alpha 1$	3.8834	1.041947	5.212661	0.43379
4_s_at	與AF4相關之淋巴核蛋白	-1.3702	1.00987	0.420671	1.393757
8_s_at	脂肪酸輔酶A連接酶，極長鏈1	0.5008	0.790296	2.069968	1.166292
2_at	K1AA0575基因產物	2.848348	1.291904	4.670661	1.303459
7_at	G-蛋白訊號之調節因子11	-1.36697	1.337414	0.179662	0.681822
1_s_at	微管相關蛋白tau	-3.3514	1.637863	-1.01214	2.020108
9_x_at	醛-酮還原酶族7，組員A3(黃麴黴毒素醛還原酶)	0.948475	0.99349	2.289914	0.621401
4_s_at	細胞色素P450，亞族IIB(苯巴比妥-可誘導)，多肽 6	-0.71324	1.775643	1.082716	0.869708
207030_s_at	富半胱胺酸及甘胺酸蛋白2	-2.03214	1.126525	-0.19338	1.540646
207038_at	溶質載體蛋白族16(一元羧酸轉運子)，組員6	0.374876	0.580637	1.790818	1.094049
207414_s_at	成對鹼性胺基酸裂解系統4	0.341324	1.065353	2.062852	1.376036
207847_s_at	跨膜黏蛋白1	0.247008	1.354516	2.257601	1.737215
208305_at	黃體酮受體	-1.24605	0.974745	0.384022	1.29497
208451_s_at	補體組分4B	-4.78762	1.049086	-2.66361	2.080728
208682_s_at	肝細胞癌相關蛋白；乳癌相關基因1	-1.959	0.821013	-0.3239	1.382716
208788_at	酵母長鏈多元不飽和脂肪酸伸長酶之同系物2	0.152008	0.660975	1.523099	1.038038
209173_at	前梯度2(非洲爪蟾)同系物	-4.28803	0.661578	-2.56017	1.677193
209339_at	七個失神(果蠅)同系物 2	1.270858	1.066389	2.646046	0.849767
209443_at	絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員5	4.667825	0.671724	5.873446	0.804606
209459_s_at	NPD009蛋白	1.072112	1.457092	2.973341	1.645057
209460_at	NPD009蛋白	-0.96002	1.349904	0.607753	1.04472
209581_at	類似於大鼠HREV107	-0.56188	0.872894	0.668399	0.727131
209602_s_at	GATA-結合蛋白3	2.019065	1.056594	3.416464	0.940078
209603_at	GATA-結合蛋白3	1.985985	0.863569	3.186089	0.674166
209604_s_at	GATA-結合蛋白3	2.395052	1.790175	4.34208	1.519527
209623_at	甲基巴豆醯基-輔酶A羧基酶2(β)	-1.00419	1.154041	0.445889	1.017354

209696_at	果糖-1,6-雙磷酸酯酶1	-1.68104	0.963742	-0.1215	1.377052
209800_at	角蛋白16(病灶性非表皮鬆懈性掌蹠角皮症)	2.324715	1.562155	4.012295	1.229197
210085_s_at	膜聯蛋白A9	2.4829	1.125042	4.043161	1.290489
210272_at	人類細胞色素P450-IIB (h11B3)mRNA，完全cds	1.01495	0.91653	2.191543	0.64021
210480_s_at	肌球蛋白VI	-0.14392	1.616287	1.455335	1.006298
210652_s_at	DEME-6蛋白	1.251577	0.889677	2.556116	0.970199
210735_s_at	碳酸酐酶XII	1.213425	2.03426	3.084783	1.272118
211000_s_at	介白素6訊號轉導子(gp130，制瘤素M受體)	-3.02427	1.43442	-1.18813	1.697067
211233_x_at	雌激素受體1	-0.0459	1.740133	1.544577	0.867934
211234_x_at	雌激素受體1	0.044649	1.763802	1.765441	1.206805
211235_s_at	雌激素受體1	-2.24335	1.765844	-0.48324	1.306074
211323_s_at	肌醇1,4,5-三磷酸酯受體，類型1	2.749775	0.789763	3.855643	0.652063
211596_s_at	膜糖蛋白 LIG-1之人類mRNA，完全cds	0.451307	1.03825	1.691284	0.751559
211712_s_at	人類，純系MGC：1925，mRNA，完全cds	0.615955	1.516076	2.069047	0.790366
212195_at	人類mRNA；cDNA DKFZp564F053(來自純系DKFZp564F053)	0.66476	0.873729	1.797193	0.663081
212196_at	人類mRNA；cDNA DKF2p564F053(來自純系DKFZp564F053)	1.370605	0.637597	2.49272	0.820267
212496_s_at	KIAA0876蛋白	2.9339	0.874367	4.097768	0.756001
212637_s_at	人類mRNA；cDNA DKFZp434D2111(來自純系DKFZp434D2111)	-1.88266	1.081913	-0.63578	0.780821
212638_s_at	人類mRNA；cDNA DKFZp434D2111(來自純系DKFZp434D2111)	2.261515	1.394089	3.785398	1.192581
212956_at	KIAA0882蛋白	-2.7829	1.397052	-0.86347	2.046812
212960_at	KIAA0882蛋白	-0.50333	1.45485	0.947772	1.02444
213201_s_at	肌鈣蛋白T1，骨骼，緩慢	-1.9544	1.210569	-0.40381	1.441706
213412_at	緊密接合蛋白3(帶狀緊蛋白3)	2.951875	0.714379	4.007446	0.711117

213419_at	澱粉樣蛋白 β (A4)前驅體蛋白結合, 族B, 組員2 (Fe65-類)	-2.21361	1.478023	-0.51415	1.591816
213712_at	人類mRNA ; cDNA DKFZp434E082(來自純系 DKFZp434E082)	0.270749	0.847277	1.499404	1.020576
214053_at	人類純系23736 mRNA序列	-0.39205	1.186238	0.845048	0.820314
214164_x_at	銜接物相關之蛋白錯合物1, γ 1 亞單位	-1.08541	1.111223	0.178117	0.95879
214428_x_at	補體組分4A	0.533406	0.838849	1.642348	0.807099
214440_at	N-乙醯基轉移酶1(芳基胺N-乙醯 基轉移酶)	-0.99962	0.684062	0.154358	0.999297
215304_at	人類純系23948 mRNA序列	2.4353	0.529481	3.488893	0.879103
215552_s_at	來自染色體6q25.1-26上之純系 RP1-63I5之人類DNA序列。含有 新穎基因之三部分及雌激素受體 1之ESR1基因之外顯子(NR3A1, 雌二醇受體), ESTs、STSs及GSSs	-4.0518	1.024367	-2.20072	2.254477
215616_s_at	KIAA0876 蛋白	2.582125	0.659442	3.570411	0.700552
215729_s_at	TONDU	1.641575	0.849076	2.756482	0.863148
215867_x_at	銜接物相關之蛋白錯合物1, γ 1 亞單位	-0.42352	0.884606	0.727052	0.926142
216092_s_at	溶質載體蛋白族7(陽離子脂肪酸 轉運子, y+系統), 組員8	0.063651	1.352604	1.366287	0.918248
216109_at	人類cDNA: FLJ21695 fis, 純系 COL09653	-1.17386	1.143511	0.232514	1.345207
216381_x_at	醛-酮還原酶族7, 組員A3(黃麴黴 毒素醛還原酶)	0.46636	0.383625	1.657506	1.251032
217190_x_at	雌激素受體{外顯子6}人類, 他 莫昔芬抵抗乳癌17, 基因組突 變體, 187 nt	0.899139	0.533766	2.030393	1.097631
217838_s_at	RNB6	-1.31066	0.930532	-0.16453	0.933916
218195_at	假設蛋白 FLJ12910	0.847629	0.786234	2.077682	1.202885
218450_at	血紅素結合蛋白	0.080843	0.82158	1.234993	1.027254
218502_s_at	毛髮-鼻-指(趾)症候群I	-1.57325	1.012703	-0.27651	1.276184
218806_s_at	vav 3致癌基因	1.662298	0.790643	2.689179	0.799202
218976_at	含有J域之蛋白1	-1.84709	1.306292	-0.43267	1.374615

219001_s_at	假設蛋白MGC10765	-2.18314	1.146729	-0.93169	1.100879
219051_x_at	假設蛋白MGC2601	-1.64776	1.079359	-0.04531	1.917545
219197_s_at	CEGP1蛋白	3.017955	0.866409	4.110571	0.929583
219414_at	calsyntenin-2				
219663_s_at	假設蛋白MGC4659				
219682_s_at	TBX3-同工蛋白	-2.31967	2.774285	-5.24093	1.743328
219919_s_at	假設蛋白FLJ10928	1.5957	1.348698	-0.22476	1.003375
220329_s_at	假設蛋白FLJ20627	1.476165	1.643622	-0.81183	1.617203
220581_at	假設蛋白FLJ23305	0.707923	1.691725	-1.11592	1.188481
220744_s_at	WD重複域10	-1.15664	1.569856	-2.79242	0.859538
221765_at	ESTs	1.266316	0.936218	-0.08462	0.892242
222212_s_at	腫瘤轉移抑制因子	0.105187	1.541242	-1.65582	1.335109
222314_x_at	ESTs	2.914925	1.476344	1.290308	1.093452
41660_at	叢集Incl. AL031588: dJ1163J1.1(小鼠跨膜受體Celsr1之 直向同源物，類似於大鼠MEG之 含有KIAA0279 LIKE EGF類域之 蛋白)	-1.50101	2.986928	-3.88453	1.411412
		-0.50993	0.923661	-1.93244	1.140847
		0.987597	0.893199	-0.11725	0.498882

表 S8: 表 A4之基因(13種共同基因)之基因表現資料

UID 名稱	2000683T+ 陰性	2000775T+陰性	2000804T+ 陰性	980346T+陽性	980383T +陰性						
990082T+陰性	980177T+ 陰性	980178T+陰性	980403T+ 陰性	980434T+陰性	990075T +陰性						
990113T+陰性	990107T+ 陰性	980203T+陰性	980208T+ 陽性	980220T+陽性	980221T +陰性						
990115T+陽性	990375T+ 陰性	980404T+陰性	980409T+ 陰性	990123T+陰性	2000422T +陰性						
2000787T-LA	2000818T- LA	20020021T-LA	20020051T- LA	20020056T-LA	980197T +陽性						
980215T+陰性	980217T+ 陰性	980261T+陰性	980391T+ 陰性	2000768T+陽性	2000779T +陰性						
2000948T+陰性	20020160T- LA	2000401T-LA	20020071T- LA	2000215T-正常-類							
2000220T-LA	980333T- LA	980058T-LA	980278T- LA	980288T-ERBB2	2000597T -LA						
2000609T-LA	2000272T- LA	2000274T-正常-類 980285T-基礎 2000593T-基礎									
2000638T-基礎	2000641T- ERBB2	2000675T-ERBB2	2000287T- ERBB2	2000320T-基礎							
2000880T-基礎	2000731T- 基礎	980353T-陰性	2000829T- 陽性	980373T-陽性	2000500T -陰性						
2000759T-陽性	980238T- 陽性	980395T-陽性	980396T- 陽性	980411T-陰性	980441T -陰性						
990262T-陰性	980216T- 陰性	980194T-陽性	980247T- 陽性	980338T-陰性	990174T -陰性						
990299T-陰性	2000210T- ERBB2	980315T-LA	980335T- ERBB2	980193T-基礎							
980256T-基礎	980214T+ 陽性	990148T+陽性	2000209T+ 陽性	990223T+陽性							
2000104T-ERBB2	2000651T-正常-類 2000237T-ERBB2 2000652T-ERBB2 2000376T-ERBB2										
2000399T- ERBB2	20020090T- ERBB2	2000709T- ERBB2	2000813T- 陽性	980380T- 陽性	990134T -陽性						
2000171T-ERBB2											
置信度			高	高	高	高	高	高	高	高	高
高	高	高	高	高	高	高	高	高	高	高	高
高	高	高	高	高	高	高	高	高	高	高	高
高	高	高	高	高	高	高	高	高	高	高	高
高	高	高	高	高	高	高	高	高	高	高	高
高	高	高	高	高	高	高	高	高	高	高	高
高	高	高	高	高	高	高	高	高	高	高	低
低	低	低	低	低	低	低	低	低	低	低	低
低	低	低									

201525_at		載脂蛋白D			2.749	7.332	2.111	2.803	1.752	1.958	1.75
2.712	4.541	3.009	3.613	4.291	1.486	4.204	2.849	3.388	3.262	3.603	
3.097	7.419	5.491	4.873	1.444	2.954	1.296	3.352	2.856	2.266	5.145	
4.695	4.072	6.963	4.804	2.886	0.7888	3.226	0.3389	1.921	2.803	4.261	
4.993	4.251	0.785	6.066	4.539	2.019	5.235	1.808	4.592	0.09904	2.77	2.85
3.059	3.353	1.229	1.679	1.879	2.77	0.9126	4.246	6.957	3.753	7.109	4.31
1.624	2.986	2.603	0.984	4.797	0.5836	5.433	2.722	1.66	3.161	2.94	
0.3395	1.008	4.023	2.417	4.21	4.833	5.118	0.7322	7.893	5.443	5.369	
1.104	6.198	2.819	3.773	1.536	1.673	6.562	4.973	6.796	6.121		
202991_at		含有 START 域之 3				0.1623	0.7959	-0.3925	3.014	0.4513	
0.2522	0.3208	-0.2599	0.5714	-0.5644	0.5246	0.8061	0.6035	-0.3416	2.886	0.8943	
-0.6905	2.991	0.6204	0.4511	-0.4408	-0.2534	0.07863	1.517	0.6792	0.6636	0.2455	
-0.1443	2.871	-0.3209	-0.05486		1.605	0.1314	2.252	0.002929		0.9972	
0.08306	2.623	0.4914	0.4794	-0.02506		0.1142	0.3137	0.5399	3.005	0.2001	
2.758	0.1815	0.1945	-0.05305		0.6643	0.5267	2.002	0.462	3.014	0.2885	
0.1389	-0.05295		-1.923	1.882	0.5175	0.09324	1.667	3.328	2.384	3.651	
1.299	0.1444	0.158	1.234	2.21	0.1798	-0.1465	0.411	0.5087	3.457	1.745	
3.551	-0.2846	0.158	2.62	3.53	3.728	3.149	0.2238	-0.9861	-0.3033	3.286	
-0.07757		2.736	3.579	2.466	1.495	2.523	3.703	3.77			

203628_at		類人類胰島素生長因子1受體mRNA，3'序列，mRNA序列								
2.795	2.381	5.773	1.45	3.568	3.288	2.631	2.062	2.515	4.693	2
2.984	3.098	4.667	2.513	2.232	2.442	0.5148	2.452	3.675	4.111	2.55
3.705	1.115	1.538	1.731	2.76	3.559	2.259	1.855	0.6405	3.657	4.928
2.664	6.732	6.752	0.5081	2.53	1.503	1.872	4.124	1.466	3.48	2.903
0.2213	3.556	1.22	1.193	3.206	-0.1502	0.07299	0.3962	0.5347	0.7098	0.06693
0.09198	0.3905	-0.02844		-0.009415		1.025	0.7389	2.194	-0.4784	1.723
0.222	0.05793	0.573	3.054	1.338	0.6058	1.426	1.54	0.9868	0.84	0.1264
0.2324	-0.258	1.21	-0.8171	1.998	1.449	-0.1467	0.3772	1.21	-0.4615	1.451
0.1205	-0.1947	-0.9146	1.441	-0.8475	0.04923	0.4557	-2.688	0.2235	0.5537	
205307_s_at		犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)							-0.117	-1.011
-2.489	-0.9037	-1.085	-1.12	-1.219	-1.735	-1.829	-1.721	-1.433	-0.02038	
1.167	-1.694	-1.571	1.055	-2.743	0.03987	0.01731	0.1225	0.1203	-1.484	-0.591
-1.35	-0.2275	0.7435	-1.218	-0.4883	-0.8609	-0.7848	-0.2848	-1.499	-0.3403	-1.388
-0.9036	-0.3888	-0.4186	-1.082	-1.261	-1.201	-0.1329	-1.222	-1.679	-0.2855	0.5551
-1.587	-0.1132	-1.485	-1.13	-0.7033	-0.7773	0.7705	0.008025		-0.2992	0.06924
-0.3291	-2.038	-1.017	-3.967	-0.4769	0.8039	-1.589	-0.7423	-0.4919	-1.328	0.2971
-1.549	-0.7277	1.643	-1.604	0.5154	-0.09918		-0.6515	-0.8327	-0.986	-0.04337
-0.95	-0.273	-0.3601	-2.266	1.182	0.7985	-0.8065	1.063	2.302	-0.6945	-1.219
0.9502	-0.894	0.7855	-1.668	0.1515	-0.3956	-1.677	0.22	1.595		

210761_s_at		生長因子受體結合蛋白7						0.4452	1.205	1.412	2.858	
1.493	1.508	0.3961	0.7703	1.033	0.922	0.4947	1.016	1.668	1.669	2.906		
1.568	0.889	3.42	1.335	0.6151	0.7453	0.6185	1.248	1.748	2.238	0.6557		
0.7697	1.296	4.588	0.7527	0.5559	0.7794	0.9863	1.981	1.503	0.3864	0.5489		
3.704	0.7039	1.561	0.9271	0.6039	0.9461	1.471	3.699	1.334	1.981	0.6054		
0.5662	1.051	1.677	1.507	3.042	1.307	4.472	1.189	0.7615	0.228	0.6253		
3.214	1.966	0.6688	2.263	3.093	2.839	1.988	1.721	1.684	0.6625	1.159	2.94	
1.063	0.1599	1.04	0.2849	3.697	2.31	3.887	0.6321	0.7463	3.728	5.268		
3.912	3.666	1.984	0.7088	0.5511	3.982	5.042	4.321	4.339	4.248	2.174		
3.317	4.032	4.736										
210930_s_at		v-erb-b2成紅血細胞白血病毒致癌基因同系物2，										
衍生自神經/膠母細胞瘤之致癌基因同系物(鳥類)								-0.8461	-2.708	-0.9694	0.3187	
-1.475	-1.568	0.3559	-1.343	-2.559	-0.9886	-1.727	-1.466	-0.1998	-0.8977	0.3377		
-0.3748	-1.943	1.36	-1.455	-1.361	-1.218	-1.374	-0.4494	1.16	0.7238	-0.4209		
-2.201	-0.4352	1.833	-1.829	-0.6478	-4.138	-0.5983	0.6215	-1.066	-1.07	-0.332		
1.556	-0.5345	-0.8175	-0.2384	-1.649	-0.837	0.487	1.322	-0.7451	0.7285	-0.9136		
-1.812	-3.225	-0.1626	-1.19	1.542	-0.4326	1.705	0.2116	-0.2503	-1.408	-1.292		
1.544	-0.8231	-1.735	0.4762	0.09548	-0.7243	-0.7869	-1.927	-1.524	-2.637	-4.457		
-0.278	-2.773	-2.013	-1.611	-2.056	1.532	0.08922	2.774	-0.2269	-1.08	1.078	2.7	
1.397	1.554	-1.5	-0.9627	-0.8952	2.069	1.728	3.212	3.121	3.149	1.108		
-0.7891	0.9288	2.864										

211657_at		癌胚抗原相關之細胞黏附分子6(非特異性交叉反應性抗原)										
			3.887	1.127	5.069	1.162	4.256	2.372	0.06854	2.496		
0.534	1.805	0.6949	4.237	3.755	-0.05911		1.471	1.388	1.548	1.032		
4.176	0.407	3.742	3.638	4.006	3.88	5.988	1.433	0.1368	2.179	3.537		
0.7946	0.4718	3.327	-0.02141		1.842	0.3149	5.084	0.3826	1.889	-0.9834		
2.416	0.3955	0.08346	1.603	2.92	3.158	0.7611	5.397	-0.485	0.3396	0.1982		
0.2382	1.376	4.494	0.6605	4.674	4.38	-0.2242	0.2056	-0.3151	3.863	0.983		
0.8939	1.474	0.5326	3.265	-0.034	-0.8774	-0.5614	2.687	5.257	4.683	0.7389		
0.7168	0.8051	4.189	4.894	4.905	1.134	0.431	0.5341	3.92	5.643	4.536		
4.869	3.96	0.6223	5.275	4.33	3.687	4.673	0.2819	1.224	2.126	5.62		
3.871	0.6072											
213557_at		ESTs，微弱地類似於普遍轉錄之三十四肽重複基因，Y染色體；Y染色體上普遍轉錄之TPR基因[人類]										
[人類]		1.252	1.184	0.5043	3.153	1.387	1.868	0.5293	-0.2155	0.3275		
0.5276	1.395	1.851	1.543	0.5434	2.397	1.591	0.1861	1.623	1.723	0.7596		
0.5377	0.3335	1.596	2.154	1.513	1.603	0.1632	1.181	3.969	0.5737	1.136		
2.645	0.6143	2.339	0.2645	0.7221	0.6219	3.499	0.5513	1.099	0.9166	1.378		

0.6302	0.9299	3.592	0.9732	3.427	0.7249	0.7654	0.586	1.397	-1.58	3.088	
0.7145	4.663	0.5107	1.368	1.251	0.8759	1.862	2.072	1.048	0.8533	3.836	
2.693	4.055	1.126	0.493	0.3712	1.462	1.211	0.621	1.516	0.4326	1.09	2.63
2.419	0.667	0.5337	0.3296	3.749	3.494	3.834	3.956	1.295	-0.3071	0.5377	
0.8307	1.086	2.534	3.733	3.321	2.127	0.05067	3.98	4.461			

214451_at	轉錄因子AP-2 β (活化增強子結合蛋白2 β)									
-3.097	2.467	-3.372	3.439	0.1365	-1.298	2.39	1.441	2.839	2.516	-1.258
-2.597	-0.5943	1.978	-0.9813	-1.202	1.496	3.43	3.001	-1.562	2.541	-4.519
2.889	0.6659	1.661	-2.472	1.623	3.059	-2.935	3.575	1.469	-4.59	3.603
3.517	-3.813	-0.1878	4.003	-0.4031	0.88	2.51	-4.28	2.753	1.234	-4.588
3.173	-4.705	1.066	-1.809	1.967	-2.498	1.153	0.279	2.117	3.623	-0.005383
1.745	-4.141	-1.479	-1.257	1.798	4.45	-1.547	2.506	3.646	-3.226	-0.913
-3.058	-3.123	3.658	-1.289	3.548	-0.2634	-1.531	-4.923	2.247	1.723	-2.025
3.197	-2.015	-0.7008	4.068	3.333	-1.154	4.028	3.88	0.3311	3.34	2.444
2.631	3.682	3.38	3.92	3.618	4.305	3.96	4.973			
215465_at	ATP-結合盒，亞族A(ABC1)，組員12								-5.53	-0.2993
-2.982	-1.196	-1.515	-1.129	1.018	-2.386	-0.3181	-1.932	-1.838	0.7215	-1.211
-1.273	-1.483	-0.995	-1.928	-1.288	-1.39	-0.7415	-0.23	-2.464	-1.478	-0.2715
-1.114	-2.064	1.22	-2.498	-0.9399	-2.507	-0.4786	-2.321	-0.5358	-2.004	-2.388
-2.234	0.078	-1.043	1.185	-1.93	-1.992	-2.169	-2.156	-2.18	0.381	-4.889
1.702	-1.345	-1.946	-1.149	-0.7878	-0.6671	-1.429	-0.559	-1.242	-2.897	-2.329
-1.631	-2.476	-0.6065	0.4199	-2.905	-0.8082	-1.942	-1.804	-1.404	-1.384	-3.471
0.2961	-0.6596	-0.5091	-2.246	-2.386	-2.697	-1.245	0.4357	-0.7417	-0.01172	
-1.168	-2.224	-0.5227	1.617	-0.04832		0.4729	-0.4882	-2.002	-0.5482	1.449
-1.664	0.7275	0.8683	-2.091	0.14	0.4634	1.916	0.7919			

219429_at	脂肪酸羧化酶				-1.539	-0.2486	-0.06329		-0.606	-1.426
-1.273	0.05695	0.4841	0.3636	-0.7702	-1.403	-0.7	-1.611	-0.5367	0.6557	-0.5048
-0.9159	0.8194	-1.687	-1.037	-0.6167	-0.1531	-1.306	0.1918	-0.531	0.2454	0.7654
-1.344	0.7986	0.2327	-0.9519	-0.8758	-1.052	-0.6758	0.8207	-0.1432	-0.4994	-0.0002446
-0.2944	-1.152	-0.2746	-1.314	0.3005	-0.5842	0.218	-0.5254	-0.7197	-0.6967	-0.2
-0.8899	-0.2978	0.2625	1.562	-1.044	1.383	-0.5091	-0.3997	-0.8286	-3.217	-0.2482
0.5994	0.06282	0.06886	0.1471	0.9134	0.1739	0.6888	-1.575	0.3812	-0.6085	0.7442
-0.7528	-0.5949	-0.4236	-0.7073	1.218	-0.4363	1.209	0.3444	-0.969	0.2863	0.9532
0.7178	1.296	0.6456	-0.4466	1.152	0.4512	1.933	1.497	-0.3116	0.1834	0.142
1.228	1.876	1.35								

220149_at		假設蛋白FLJ22671				-0.585	-1.416	-0.7662	2.221	-0.3646
-0.8895	-0.6838	-0.5557	-0.4347	-0.4597	-0.07175		-0.09613		-0.4148	-0.781
-1.112	-0.482	-1.328	-0.6111	-2.445	-1.028	-0.6113	-0.08989		-1.397	-0.5025
-0.3443	-1.424	-0.3695	-0.8427	0.4616	-1.052	-1.163	-0.9368	-0.3882	0.7431	-0.04467
-0.4188	-0.7193	2.204	-1.393	-0.7435	-1.423	-0.5707	-0.4196	-0.6552	2.686	-0.6905
4.914	-0.3156	-0.9062	-0.1168	0.2261	0.1723	0.386	1.191	2.885	-0.7671	-2.42
-0.2398	-1.799	2.044	0.8819	-0.3224	3.604	1.023	3.736	2.807	-0.5473	-1.357
0.3665	-0.2828	-0.246	-0.01971		0.4476	-0.5921	-0.2366	1.906	-0.3266	2.079
0.2249	-0.5295	0.08667	2.691	1.636	1.349	-0.3243	-1.536	1.435	4.099	-0.8161
1.734	2.641	1.301	1.355	-1.242	1.708	3.096				

39248_at		水通道蛋白 3		0.4769	-0.2623	-0.7927	1.948	0.03186	2.194	0.6044
2.335	-0.1663	0.4244	1.476	3.025	0.6734	2.102	3.241	-0.5173	0.8267	3.789
2.556	-0.07496		2.804	1.786	-1.024	0.4586	2.795	0.6762	0.07351	0.3396
0.4198	0.7147	1.677	2.114	-0.1301	0.06363	3.336	3.314	0.1946	1.919	-0.1613
0.8785	-0.1946	-0.1926	-1.876	3.881	0.3148	-1.082	-0.852	0.0508	0.3455	-0.9268
0.2052	0.2611	0.8294	2.1	1.987	3.696	0.8302	1.104	-1.175	3.041	0.07521
3.434	3.543	0.13	1.305	0.1424	2.271	1.841	0.7022	4.044	4.959	0.2898
0.4821	1.642	0.9258	1.169	-0.382	-0.8969	0.8155	1.156	3.712	2.333	1.722
1.466	3.247	1.128	1.167	3.68	4.088	4.324	-0.5153	2.505	5.002	0.05894
5.292	0.9251									

表 S9：表 A4(13種共同基因)基因集之表現資料的平均值

 (μ) 與標準偏差(σ)之加權表決參數

探針_ID	基因名稱	參考序列全長	Unigene	高置信度		低置信度	
				平均值	SD	平均值	SD
在低置信度腫瘤中上調							
201525_at	載脂蛋白D	NM_001647	Hs.75736	3.213993	1.711066	4.43395	2.23157
202991_at	含有 START 域之3	NM_006804	Hs.77628	0.838735	1.186229	2.215114	1.621765
205307_s_at	犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)	NM_003679	Hs.107318	-0.75339	0.924201	0.105819	1.199695
210761_s_at	生長因子受體結合蛋白7	NM_005310	Hs.86859	1.512564	1.051211	3.500556	1.421506
210930_s_at	v-erb-b2 成紅血細胞白血病毒致瘤基因同系物2，衍生自神經/膠母細胞瘤之致瘤基因同系物(鳥類)	NM_004448	Hs.323910	-0.71309	1.339254	1.297613	1.591897
211657_at	癌胚抗原相關之細胞黏附分子6 (非特異性交叉反應抗原)	NM_002483	Hs.73848	1.948209	1.842322	3.452838	1.859184
213557_at	ESTs，微弱地類似於普遍轉錄之三十四肽重複基因，Y染色體；Y染色體[人類]上普遍轉錄之TPR基因[人類]	---	Hs.14691	1.359728	1.098941	2.417623	1.605763
214451_at	轉錄因子AP-2 β (活化增強子結合蛋白2 β)	NM_003221	Hs.33102	0.234429	2.657284	3.171194	1.547226
215465_at	ATP-結合	NM_015657	Hs.134585	-1.35669	1.237705	0.067599	1.228661

	盒，亞族 A(ABC1)，組 員12						
219429_at	脂肪酸羥化 酶	---	Hs.249163	-0.32527	0.827988	0.809581	0.722212
220149_at	假設蛋白 FLJ22671	NM_024861	Hs.193745	-0.05674	1.363225	1.200829	1.596251
39248_at	水通道蛋白 3	NM_004925	Hs.234642	1.076674	1.458035	2.508421	1.755277
在高置信度腫瘤中上調							
203628_at	類人類胰島 素生長因子1 受體 mRNA，3' 序列，mRNA 序列	---	Hs.405998	1.956068	1.625758	0.129864	1.072433

表 A1：SAM(微陣列顯著性分析)：以 < 15% 之 FDR(檢出率)，總共識別出在低置信度腫瘤中之 86 種上調基因及 2 種下調基因。使用此基因集，LOOCV 檢定產生 84% 之分類精確度。

在低置信度腫瘤中上調之基因				
基因名稱	計分(d)	q-值(%)	Unigene	參考序列全長
206793_at	4.1852709	1.3837984	Hs.1892	NM_002686//苯乙醇胺N-甲基轉移酶
211237_s_at	4.071839	1.3837984	Hs.165950	NM_002011//成纖維細胞生長因子受體4異型1前驅體///NM_022963//成纖維細胞生長因子受體4異型2前驅體
210761_s_at	3.9001438	1.3837984	Hs.86859	NM_005310//生長因子受體結合蛋白7
206164_at	3.8109161	1.3837984	Hs.241551	NM_006536//鈣激活氯化物通道2
204913_s_at	3.4806716	1.3837984	Hs.32964	NM_003108//SRY(性別決定區Y)-盒11
210930_s_at	3.4544924	1.3837984	Hs.323910	NM_004448//v-erb-b2成紅血細胞白血病毒致癌基因同系物2，衍生自神經/膠母細胞瘤之致癌基因同系物
204910_s_at	3.3311974	1.3837984	Hs.321576	NM_006458 //含有三重基序之3異型 α ///NM_033278 //含有三重基序之3異型 β ///NM_033279 //含有三重基序之3異型 γ
214451_at	3.2935388	1.3837984	Hs.33102	NM_003221 //轉錄因子AP-2 β (活化增強子結合蛋白2 β)
217562_at	3.2344498	1.3837984	Hs.106642	---
217276_x_at	3.0703975	1.3837984	Hs.301947	NM_014509 //類海怪
215686_x_at	3.0323791	1.3837984	---	---
215559_at	3.0225718	1.3837984	Hs.274260	NM_001171 // ATP-結合盒，亞族C，組員6
206827_s_at	2.9342047	1.3837984	Hs.302740	NM_014274 //瞬時感受器電位陽離子通道，亞族V，組員6///NM_018646 //瞬時感受器電位陽離子通道，亞族V，組員6
208893_s_at	2.9089684	1.3837984	Hs.180383	NM_001946 //雙特異型磷酸酯酶6異型a///NM_022652 //雙特異型磷酸酯酶6異型b
203619_s_at	2.8107802	1.3837984	Hs.182859	---
203824_at	2.7813798	1.3837984	Hs.84072	NM_004616//跨膜4超族組員3
221811_at	2.747613	1.3837984	Hs.91668	---
216202_s_at	2.7319622	1.3837984	Hs.59403	NM_004863 //絲胺酸棕櫚醯基轉移酶，長鏈鹼亞單位2

209757_s_at	2.7152502	1.3837984	Hs.25960	NM_005378 // v-myc骨髓細胞瘤病毒相關之致癌基因，衍生自神經母細胞瘤
219429_at	2.665359	1.3837984	Hs.249163	---
215465_at	2.628031	1.3837984	Hs.134585	NM_015657 // ATP-結合盒，亞族A，組員12異型b /// NM_173076 // ATP-結合盒，亞族A，組員12異型a
214203_s_at	2.6018018	1.3837984	Hs.343874	NM_005974///NM_016335//脯胺酸脫氫酶(氧化酶)1
202942_at	2.5652724	1.3837984	Hs.74047	NM_001985 //電子轉移黃素蛋白， β 多肽
205478_at	2.545305	1.3837984	Hs.76780	NM_006741 //蛋白磷酸酯酶1，調節性(抑制劑)亞單位1A
203722_at	2.5390254	1.3837984	Hs.77448	NM_003748 //醛脫氫酶4A1前驅體///NM_170726 //醛脫氫酶4A1前驅體
202991_at	2.5022628	1.3837984	Hs.77628	NM_006804 //類固醇產生急性調節蛋白相關
205104_at	2.4827654	1.3837984	Hs.323833	NM_014723 //syntaphilin
215659_at	2.4619073	1.3837984	Hs.306777	---
220622_at	2.407245	1.3837984	Hs.114005	NM_024727 //假設蛋白FLJ23259
208083_s_at	2.3715062	1.3837984	Hs.57664	NM_000888 //整合素， $\beta 6$
206043_s_at	2.3543638	1.3837984	Hs.6168	NM_014861 // KIAA0703基因產物
221345_at	2.3351396	1.3837984	Hs.248056	NM_005306 // G蛋白偶合受體43
39248_at	2.3213986	1.3837984	Hs.234642	NM_004925 //水通道蛋白3
205766_at	2.3057935	1.3837984	Hs.343603	NM_003673 //視松蛋白(telethonin)
211682_x_at	2.2991204	1.3837984	Hs.137585	NM_053039 //UDP糖基轉移酶2族，多肽B28
210571_s_at	2.2806771	1.3837984	Hs.24697	XR_000114//
219233_s_at	2.2752973	1.3837984	Hs.19054	NM_018530 //假設蛋白PRO2521
204818_at	2.2720676	1.3837984	Hs.155109	NM_002153 //羥基類固醇(17- β)脫氫酶2
211828_s_at	2.2270979	1.3837984	Hs.170204	---
205916_at	2.2142817	1.3837984	Hs.112408	NM_002963 //S100鈣結合蛋白A7
209522_s_at	2.2117774	1.3837984	Hs.12068	NM_000755 //肉鹼乙醯基轉移酶前驅體，異型1 /// NM_004003 //肉鹼乙醯基轉移酶，異型2 /// NM_144782 //肉鹼乙醯基轉移酶前驅體，異型3
209016_s_at	2.2112214	1.3837984	Hs.23881	---
209505_at	2.2006627	1.3837984	Hs.374991	---

200831_s_at	2.1927228	1.3837984	Hs.119597	NM_005063 //硬脂醯基輔酶A去飽和酶(δ -9-去飽和酶)
207802_at	2.1832898	1.3837984	Hs.54431	NM_006061 //特異型顆粒蛋白(28 kDa)
216633_s_at	2.1766477	1.3837984	Hs.193143	---
214614_at	2.1670563	1.3837984	Hs.37035	NM_005515 //同源盒HB9
204607_at	2.1402505	1.3837984	Hs.59889	NM_005518 //3-羥基-3-甲基戊二醯基-輔酶A合酶2(線粒體)
220149_at	2.1400852	1.3837984	Hs.193745	NM_024861//假設蛋白FLJ22671
219756_s_at	2.1391208	1.3837984	Hs.267038	NM_024921//卵巢早衰1B

213674_x_at	2.1351759	1.3837984	Hs.300697	---
211657_at	2.1231572	1.3837984	Hs.73848	NM_002483 //癌胚抗原相關細胞黏附分子6(非特異性交又反應抗原)
204941_s_at	2.1178907	1.3837984	Hs.87539	NM_000695 //醛脫氫酶3B2
214133_at	2.0836401	3.5733527	Hs.99918	---
210663_s_at	2.0766057	3.5733527	Hs.169139	NM_003937 //犬尿胺酸酶(L-犬尿胺酸水解酶)
220414_at	2.0543228	3.5733527	Hs.180142	NM_017422 //類鈣調蛋白皮膚蛋白
205808_at	2.0365629	3.5733527	Hs.283664	NM_004318 //天門冬胺酸酯 β -羥化酶異型a /// NM_020164 //天門冬胺酸酯 β -羥化酶異型e /// NM_032466 //天門冬胺酸酯 β -羥化酶異型c /// NM_032467 //天門冬胺酸酯 β -羥化酶異型d /// NM_032468 //天門冬胺酸酯 β -羥化酶異型b
203365_s_at	2.0185514	3.5733527	Hs.80343	NM_002428 //基質金屬蛋白酶15前蛋白原
206509_at	2.0114514	3.5733527	Hs.99949	NM_002652 //催乳素誘導之蛋白
213557_at	1.9942427	3.5733527	Hs.14691	---
214971_s_at	1.9917977	3.5733527	Hs.2554	NM_003032 //唾液酸轉移酶1異型a /// NM_173216 //唾液酸轉移酶1異型a /// NM_173217//唾液酸轉移酶1異型b
211899_s_at	1.9768615	4.5901604	Hs.8375	NM_004295//TNF受體相關之因子4異型1//NM_145751//TNF受體相關之因子4異型2
220615_s_at	1.9216703	4.5901604	Hs.100895	NM_018099//假設蛋白FLJ10462
206915_at	1.8471141	7.400989	Hs.355454	NM_002509 // NK2轉錄因子相關，座位2
201388_at	1.8446012	7.400989	Hs.9736	NM_002809 //蛋白酶體26S非ATP酶亞單位3
205307_s_at	1.8282052	7.400989	Hs.107318	NM_003679 //犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)

209616_s_at	1.8059335	7.400989	Hs. 76688	NM_001266 // 羧基酯酶1(單核細胞/巨噬細胞絲胺酸酯酶1)
205910_s_at	1.7828285	7.400989	Hs.406160	NM_001807 // 羧基酯脂肪酶前驅體
201525_at	1.7490382	7.400989	Hs.75736	NM_001647//載脂蛋白D前驅體
201729_s_at	1.7197176	9.106286	Hs.151761	---
204304_s_at	1.6603865	9.106286	Hs. 112360	NM_006017 //類普若敏(prominin-like)1
220225_at	1.6559087	9.106286	Hs. 196927	NM_016358 //艾庫絲(iroquois)同源盒蛋白4
209560_s_at	1.6357376	10.248328	Hs. 169228	NM_003836 // δ 類同系物
207131_x_at	1.6311017	10.248328	Hs.401847	NM_005265 // γ -麩胺醯轉移酶1//NM_013421// γ -麩胺醯轉移酶1前驅體//NM_013430// γ -麩胺醯轉移酶1
220972_s_at	1.6233436	10.248328	Hs.307010	NM_030975 //角蛋白相關蛋白9.9
209641_s_at	1.6169812	10.248328	Hs.90786	NM_003786 //ATP-結合盒，亞族C，組員3異型MRP3 /// NM_020037 // ATP-結合盒，亞族C，組員3異型MRP3A /// NM_020038 //ATP-結合盒，亞族C，組員3異型MRP3B

211588_s_at	1.6135313	10.246328	Hs.381618	---
201946_s_at	1.5784917	10.248328	Hs.432970	NM_006431 //含有伴侶蛋白之TCP1，亞單位2(β)
205029_s_at	1.5779091	10.248328	Hs.26770	NM_001446 //脂肪酸結合蛋白7，腦
201942_s_at	1.5530281	11.432502	Hs.5057	NM_001304 //羧基肽酶D前驅體
213913_s_at	1.5514129	11.432502	Hs.11912	---
207102_at	1.5436816	11.432502	Hs.201667	NM_005989 // 醛-酮還原酶家族1，組員D1
214624_at	1.5133976	11.432502	Hs.159309	NM_007000 //泌尿血小板溶素(uroplakin)1A//NM_032896//
206714_at	1.5040028	11.432502	Hs.111256	NM_001141 //花生四烯酸酯15-脂氧合酶，第二類型
205765_at	1.4589879	12.831585	Hs. 104117	NM_000777 //細胞色素P450，族3，亞族A，多肽5
213043_s_at	1.4469888	12.831585	Hs.23106	NM_014815 //甲狀腺激素受體相關之蛋白
在高置信度腫瘤中上調之基因				
基因名稱	計分(d)	q-值(%)		
204286_s_at	-3.429773	1.3837984	Hs.96	NM_021127 //佛波醇12-肉豆蔻酸酯13-乙酸酯誘導之蛋白1
203628_at	-2.907564	1.3837984	Hs.405998	---

表 A2：GR(藉由 SVM 之基因排列)：以 86% 之分類精確度，總共識別出 251 種基因具有分類腫瘤之 HC 或 LC 狀況的能力。該等基因係根據其判別強度來排列，而其判別強度則係藉由基因特異性之錯誤分類率來計算。此基因排列 -SVM 套件係由 GeneData™ (Basel, Switzerland) 提供。

探針ID	基因描述	Unigene ID
205225_at	雌激素受體1	Hs.1657
206165_s_at	鈣活化之氯化物通道，族組員2	Hs.241551
202917_s_at	S100 鈣結合蛋白A8(鈣粒蛋白A)	Hs.100000
210761_s_at	生長因子受體結合蛋白7	Hs.86859
202376_at	絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員3	Hs.234726
211657_at	癌胚抗原相關細胞黏附分子6(非特異性交叉反應抗原)	Hs.73848
206509_at	催乳素誘導之蛋白	Hs.99949
201650_at	角蛋白19	Hs.182265
204734_at	角蛋白15	Hs.80342
203627_at	類人類胰島素生長因子1受體mRNA，3'序列，mRNA序列	Hs.405998
39248_at	水通道蛋白3	Hs.234642
209603_at	GATA結合蛋白3	Hs.169946
204508_s_at	假設蛋白FLJ20151	Hs.279916
215470_at	人類cDNA FLJ36630 fis，純系TRACH2018278，mRNA序列	Hs.14658
203749_s_at	視黃酸受體 α	Hs.361071
210930_s_at	v-erb-b2成紅血細胞白血病病毒致癌基因同系物2，衍生自神經/膠母細胞瘤之致癌基因同系物(鳥類)	Hs.323910
219233_s_at	假設蛋白PRO2521	Hs.19054
204475_at	基質金屬蛋白酶1(間質膠原酶)	Hs.83169
203875_at	染色質之SWI/SNF相關、基質相關聯、肌動蛋白相依性調節因子，亞族a，組員1	Hs.152292
211699_x_at	血紅蛋白， α 1	Hs.272572
205239_at	雙調蛋白(衍生自神經鞘瘤之生長因子)	Hs.270833
205009_at	車軸草因子1(乳癌，其中表現雌激素可誘導之序列)	Hs.350470

221811_at	假設基因MGC9753	Hs.91668
218541_s_at	染色體8開放讀碼框4	Hs.283683
203628_at	類人類胰島素生長因子1受體mRNA，3'序列，mRNA序列	Hs.405998
209301_at	碳酸酐酶II	Hs. 155097
219263_at	假設蛋白FLJ23516	Hs.9238
203917_at	柯薩奇(coxsackie)病毒及腺病毒受體	Hs.79187

203980_at	脂肪酸結合蛋白4，脂肪細胞	Hs.391561
207076_s_at	精胺基琥珀酸合成酶	Hs. 160786
203408_s_at	富特定AT序列結合蛋白1(結合至核基質/支架相關DNA)	Hs.74592
203060_s_at	3'磷酸腺苷-5'-磷酸硫酸酯合成酶2	Hs.274230
63825_at	類似於假設蛋白PRO2831[人類]，mRNA序列	Hs.406646
222303_at	ESTs	Hs.292477
211959_at	未知[IMAGE: 4183312之蛋白][人類]，mRNA序列	Hs.380833
217776_at	視黃醇脫氫酶11(全反式及9-順式)	Hs.179817
204863_s_at	介白素6訊號轉導子(gp130，制瘤素M受體)	Hs. 82065
202887_s_at	HIF-1反應性RTP801	Hs.111244
201841_s_at	熱衝擊27kDa蛋白1	Hs.76067
207847_s_at	跨膜黏蛋白1	Hs.89603
215294_s_at	染色質之SWI/SNF相關、基質相關聯、肌動蛋白相依性調節因子，亞族a，組員1	Hs. 152292
218677_at	S100鈣結合蛋白A14	Hs.288998
201931_at	電子轉移黃素蛋白， α 多肽(戊二酸尿II)	Hs. 169919
202991_at	含有START域之3	Hs.77628
210633_x_at	角蛋白10(表皮鬆懈性角化過度症；掌蹠角化症)	Hs.99936
203571_s_at	特異脂肪2	Hs.74120
220625_s_at	類E74因子5(ets域轉錄因子)	Hs.11713
205567_at	碳水化合物(硫酸角質素Gal-6)磺基轉移酶1	Hs. 104576
212202_s_at	DKFZP564G2022蛋白	Hs.16492
202888_s_at	丙胺鹽基(膜)胺基肽酶(胺基肽酶 N，胺基肽酶M，微粒體胺基肽酶，CD13，p150)	Hs.1239
207023_x_at	角蛋白 10(表皮鬆懈性角化過度症；掌蹠角化症)	Hs.99936
204913_s_at	SRY(性別決定區Y)-盒11	Hs.32964

204404_at	溶質載體蛋白族12(鈉/鉀/氯化物轉運子)，組員2	Hs. 110736
211719_x_at	纖連蛋白1	Hs.287820
216510_x_at	免疫球蛋白重鏈常數mu	Hs.153261
218772_x_at	假設蛋白FLJ10493	Hs.279610
201951_at	活化白血球細胞黏附分子	Hs.10247
209250_at	退化性精母細胞同系物，脂質去飽和酶(果蠅)	Hs.185973
214745_at	KIAA1069蛋白	Hs.193143
201946_s_at	含有伴侶蛋白之TCP1，亞單位2(β)	Hs.432970
205916_at	S100鈣結合蛋白A7(牛皮癬素1)	Hs. 112408
212736_at	假設基因BC008967	Hs.6349

213438_at	人類cDNA FLJ34019 fis，純系FCBBF2002898，mRNA序列	Hs.7309
205518_s_at	胞核甘單磷酸-N-乙醯基神經胺酸羥化酶(CMP-N-乙醯基神經胺酸單加氧酶)	Hs.24697
221728_x_at	人類cDNA FLJ30298 fis，純系BRACE2003172，mRNA序列	Hs.351546
205943_at	色胺酸2,3-雙加氧酶	Hs. 183671
207431_s_at	退化性精母細胞同系物，脂質去飽和酶(果蠅)	Hs. 185973
209267_s_at	單核細胞中之BCG誘導之基因，純系103	Hs.284205
204018_x_at	血紅蛋白， α 1	Hs.272572
212204_at	DKFZP564G2022蛋白	Hs.16492
202310_s_at	膠原蛋白，類型I， α 1	Hs. 172928
201998_at	唾液酸轉移酶1(β -半乳糖甙 α -2，6-唾液酸轉移酶)	Hs.2554
208792_s_at	簇蛋白(補體溶離抑制劑，SP-40,40,硫酸化糖蛋白2，睪丸激素壓製之前列腺訊息2，載脂蛋白J)	Hs.75106
204731_at	轉換生長因子， β 受體III(β 聚糖，300 kDa)	Hs.342874
204881_s_at	UDP-葡萄糖神經醯胺葡萄糖基轉移酶	Hs.432605
205242_at	趨化因子(C-X-C基序)配位子13(B-細胞趨化物)	Hs.100431
200601_at	肌動蛋白， α 4	Hs. 182485
202037_s_at	分泌之捲曲相關蛋白1	Hs.7306
219795_at	溶質載體蛋白族6(神經遞質轉運子)，組員14	Hs. 162211
217028_at	趨化因子(C-X-C基序)受體4	Hs.89414

205066_s_at	胞外核苷酸焦磷酸酯酶/磷酸二酯酶1	Hs.11951
202357_s_at	B-因子，備解素	Hs.69771
202743_at	磷酸肌醇-3-激酶，調節性亞單位，多肽3(p55， γ)	Hs.372548
203874_s_at	染色質之SWI/SNF相關、基質相關聯、肌動蛋白性依性調節因子，亞族a，組員1	Hs. 152292
210072_at	趨化因子(C-C基序)配位子 19	Hs. 50002
202990_at	磷酸化酶，糖原；肝(赫爾斯病(Hers disease)，糖原儲存病類型VI)	Hs.771
206115_at	早期生長反應3	Hs.74088
205498_at	生長激素受體	Hs.125180
212789_at	KIAA0056蛋白	Hs.13421
222155_s_at	假定G蛋白偶合受體GPCR41	Hs.6459
218776_s_at	假設蛋白FLJ23375	Hs.285996
200820_at	蛋白酶體(前體，macropain)26S亞單位，非ATP酶，8	Hs.78466
203337_x_at	整合素胞漿域相關蛋白1	Hs. 173274
214218_s_at	人類XIST，編碼序列'a' mRNA(座位 DXS399E)，mRNA序列	Hs.352403
201729_s_at	KIAA0100基因產物	Hs.151761

204285_s_at	佛波醇12-肉豆蔻酸酯13-乙酸酯誘導之蛋白1	Hs.96
214451_at	轉錄因子AP-2 β (活化增強子結合蛋白2 β)	Hs.33102
218313_s_at	UDP-N-乙醯基- α -D-半乳糖胺：多肽N-乙醯基胺基半乳糖基轉移酶7(GalNAc-T7)	Hs.246315
217838_s_at	RNB6	Hs.241471
209189_at	v-fos FBJ鼠科骨肉瘤病毒致瘤基因同系物	Hs.25647
201131_s_at	鈣黏著蛋白1，類型1，E-鈣黏著蛋白(上皮)	Hs. 194657
203058_s_at	3'-磷酸腺苷5'-磷酸硫酸合成酶2	Hs.274230
213557_at	ESTs，微弱地類似於普遍轉錄之三十四肽重複基因，Y染色體；Y染色體[人類]上普遍轉錄之TPR基因[人類]	Hs. 14691
215465_at	ATP-結合盒，亞族A(ABC1)，組員12	Hs. 134585
213693_s_at	跨膜黏蛋白1，	Hs.89603
202218_s_at	脂肪酸去飽和酶 2	Hs. 184641
207175_at	脂肪最豐富之基因轉錄產物1	Hs.80485
205798_at	介白素7受體	Hs.362807
200916_at	轉凝蛋白2	Hs.406504

216623_x_at	含有三核苷酸重複之9	Hs.110826
211776_s_at	紅細胞膜蛋白帶4,1-類3	Hs. 103839
204472_at	在骨骼肌中過度表現之GTP結合蛋白	Hs. 79022
220149_at	假設蛋白FLJ22671	Hs. 193745
219517_at	假設蛋白FLJ22637	Hs.296178
208653_s_at	CD164抗原，唾液黏蛋白	Hs.43910
202457_s_at	蛋白磷酸酯酶 3(先前的2B)，催化亞單位， α 異型(鈣調磷酸酶A α)	Hs.272458
222108_at	---	---
200648_s_at	麩胺酸-氨連接酶(麩胺醯胺合成酶)	Hs.170171
203287_at	拉啉尼(ladinin)1	Hs.18141
219429_at	脂肪酸羧化酶	Hs.249163
212934_at	人類 cDNA FLJ30096 fis，純系BNGH41000045，mRNA 序列	Hs. 155572
205307_s_at	犬尿胺酸 3-單加氧酶(犬尿胺酸 3-羧化酶)	Hs. 107318
212686_at	KIAA1157蛋白	Hs.21894
204623_at	車軸草因子3(腸內)	Hs.82961
209459_s_at	NPD009蛋白	Hs.283675
203827_at	假設蛋白FLJ10055	Hs.9398
201952_at	活化白血球細胞黏附分子	Hs.10247
202047_s_at	色盒同系物(Chromobox homolog)6	Hs. 107374

206036_s_at	v-rel網狀內皮增生病毒致癌基因同系物(鳥類)	Hs.44313
205048_s_at	磷酸絲胺酸磷酸化酶類	Hs.369508
211527_x_at	血管內皮生長因子	Hs.73793
202660_at	次要組織相容性抗原HA-1	Hs.196914
210495_x_at	纖連蛋白1	Hs.287820
216442_x_at	纖連蛋白1	Hs.287820
212865_s_at	膠原蛋白，類型XIV， $\alpha 1$ (粗纖維調節素(undulin))	Hs.403836
221765_at	UDP-葡萄糖神經醯胺葡萄糖基轉移酶	Hs.432605
210538_s_at	含有桿狀病毒IAP重複之3	Hs.127799
204151_x_at	醛-酮還原酶家族1，組員C1(二氫二醇脫氫酶 1；20- α (3- α)-羥基類固醇脫氫酶)	Hs.306098
213836_s_at	假設蛋白FLJ10055	Hs.9398

202724_s_at	叉頭盒O1A(橫紋肌肉瘤)	Hs.170133
202404_s_at	膠原蛋白，類型1， $\alpha 2$	Hs.179573
202871_at	TNF受體相關因子4	Hs.8375
204455_at	大皰性類天疱瘡抗原1，230/240 kDa	Hs.198689
203640_at	類肌盲(muscleblind)蛋白MBLL39	Hs.283609
823_at	趨化因子(C-X3-C基序)配位子1	Hs.80420
214203_s_at	脯胺酸脫氫酶(氧化酶)1	Hs.343874
201963_at	脂肪酸輔酶A連接酶，長鏈2	Hs.154890
221730_at	膠原蛋白，類型V， $\alpha 2$	Hs.82985
217047_s_at	具有序列相似性之族13，組員A1	Hs.177664
203814_s_at	NAD(P)H脫氫酶，醌2	Hs.73956
202581_at	熱衝擊70 kDa蛋白1B	Hs.274402
218640_s_at	phafin2	Hs.29724
201752_s_at	內收蛋白3(γ)	Hs.324470
221558_s_at	淋巴增強子結合因子1	Hs.44865
211798_x_at	免疫球蛋白 λ 接合之3	Hs.102950
218400_at	2'-5'-寡腺苷酸合成酶3，100 kDa	Hs.56009
203549_s_at	脂蛋白脂肪酶	Hs.180878
201525_at	載脂蛋白D	Hs.75736
203207_s_at	具有聚脯胺酸區之雞軟骨細胞蛋白之可能的直向同源物	Hs.170198
201397_at	磷酸甘油酸脫氫酶	Hs.3343
217996_at	血小板-白細胞C激酶底物(pleckstrin)同源類域，族A，組員 1	Hs.82101
211479_s_at	5-羥基色胺(血清素)受體 2C	Hs.46362

213287_s_at	角蛋白 10(表皮鬆懈性角化過度症；掌蹠角化症)	Hs.99936
221517_s_at	Sp1轉錄活化所需之輔助因子，亞單位 6，77 kDa	Hs.22630
212775_at	KIAA0657蛋白	Hs.6654
217791_s_at	吡咯啉-5-羧酸酯合成酶(麩胺酸 γ -半醛合成酶)	Hs.114366
215250_at	人類cDNA FLJ12140 fis，純系MAMMA1000340，mRNA序列	Hs.287491
208733_at	RAB2，組員RAS致癌基因族	Hs.78305
219629_at	假設蛋白FLJ20635	Hs.265018
205542_at	前列腺之6個跨膜上皮抗原	Hs.61635
208682_s_at	黑色素瘤抗原，族D，2	Hs.4943
218729_at	latexin蛋白	Hs.109276
205376_at	肌醇多磷酸酯4-磷酸酯酶，類型II，105 kDa	Hs.153687

203953_s_at	密蛋白(claudin)3	Hs.25640
206916_x_at	酪胺酸胺基轉移酶	Hs.161640
212196_at	人類mRNA；cDNA DKFZp564F053(來自純系DKFZp564F053)，mRNA序列	Hs.71968
211000_s_at	介白素6訊號轉導子(gp130，制瘤素M受體)	Hs.82065
212254_s_at	大皰性類天庖瘡抗原1，230/240 kDa	Hs.198689
204914_s_at	SRY(性別決定區Y)-盒 11	Hs.32964
221505_at	富亮胺酸酸性核蛋白類	Hs.71331
208498_s_at	澱粉酶 α 1A，唾液	Hs.274376
201694_s_at	早期生長反應1	Hs.326035
201936_s_at	真核轉譯起始因子4 γ ，3	Hs.25732
203090_at	基質細胞衍生因子2	Hs.118684
37117_at	Rho GTP酶活化蛋白8	Hs.102336
202770_s_at	週期蛋白G2	Hs.429880
209522_s_at	肉鹼乙醯基轉移酶	Hs.12068
212451_at	KIAA0256基因產物	Hs.118978
201839_s_at	腫瘤相關之鈣訊號轉導子1	Hs.692
218309_at	假設蛋白PRO1489	Hs.197922
212450_at	KIAA0256基因產物	Hs.118978
221589_s_at	醛脫氫酶6族，組員A1	Hs.293970
217281_x_at	免疫球蛋白重鏈常數 γ 3(G3m標誌)	Hs.300697
217388_s_at	犬尿胺酸酶(L-犬尿胺酸水解酶)	Hs.169139
203336_s_at	整合素胞漿域相關蛋白1	Hs.173274
217704_x_at	---	---

201563_at	山梨醇脫氫酶	Hs.878
208151_x_at	DEAD/H(Asp-Glu-Ala-Asp/His)盒多肽 17，72 kDa	Hs.349121
217880_at	細胞分裂週期27	Hs.406631
213229_at	切酶1，Dcr-1同系物(果蠅)	Hs.87889
219768_at	假設蛋白FLJ22418	Hs.36563
200602_at	澱粉樣蛋白 β (A4)前驅體蛋白(蛋白酶微管連接蛋白(protease nexin) II，阿茲海默氏症)	Hs.177486
201082_s_at	動力蛋白活化蛋白1(p150，膠合同系物(glued homolog)，果蠅)	Hs.74617
214774_x_at	含有三核苷酸重複之9	Hs.110826
208654_s_at	CD164抗原，唾液黏蛋白	Hs.43910

202018_s_at	乳運鐵蛋白	Hs.105938
212915_at	小鼠semaF胞漿域相關蛋白之可能的直向同源物3	Hs.177635
202196_s_at	dickkopf同系物3(非洲爪蟾)	Hs.4909
221024_s_at	溶質載體蛋白族2(輔助型葡萄糖轉運子)，組員10	Hs.305971
211702_s_at	泛素特異性蛋白酶	Hs.155787
205110_s_at	成纖維細胞生長因子13	Hs.6540
219956_at	UDP-N-乙醯基- α -D-半乳糖胺：多肽N-乙醯基胺基半乳糖基轉移酶6(GalNAc-T6)	Hs.151678
202687_s_at	腫瘤壞死因子(配位子)超族，組員10	Hs.83429
205882_x_at	內收蛋白3(γ)	Hs.324470
203476_at	滋養層糖蛋白	Hs.82128
208991_at	人類cDNA FLJ35646 fis，純系SPLEN2012743，mRNA 序列	Hs.381933
204866_at	KIAA0215基因產物	Hs.82292
208180_s_at	H4組蛋白族，組員H	Hs.421737
219410_at	假設蛋白FLJ10134	Hs.104800
209290_s_at	核因子I/B	Hs.33287
202718_at	類胰島素生長因子結合蛋白2，36 kDa	Hs.433326
205862_at	GREB1蛋白	Hs.193914
203895_at	人類mRNA；cDNA DKFZp434E235 (來自純系DKFZp434E235)，mRNA序列	Hs.348724
212171_x_at	血管內皮生長因子	Hs.73793
217762_s_at	RAB31，組員RAS致癌基因族	Hs.223025
208891_at	雙特異型磷酸酯酶6	Hs.180383
221543_s_at	染色體8開放讀碼框2	Hs.125849
218834_s_at	假設蛋白FLJ20539	Hs.118552
201852_x_at	膠原蛋白，類型III， α 1(艾丹氏(Ehlers-Danlos)症候群類型IV，常染色體顯性)	Hs.119571
211965_at	鋅指蛋白36，類C3H類型-類1	Hs.85155

202015_x_at	甲硫胺醯基胺基肽酶 2	Hs.78935
203348_s_at	ets變異基因5(ets-相關分子)	Hs.43697
202783_at	菸鹼醯胺核苷酸轉氨酶	Hs.18136
202403_s_at	膠原蛋白，類型I， α 2	Hs.179573
214440_at	N-乙醯基轉移酶1(芳基胺N-乙醯基轉移酶)	Hs.155956
211748_x_at	前列腺素D2合成酶21 kDa(腦)	Hs.8272

215073_s_at	人類，純系 IMAGE: 5287010，mRNA，mRNA 序列	Hs.288869
215806_x_at	T細胞受體 γ 常數2	Hs.274509
205158_at	核糖核酸酶，RNase A 族，4	Hs.283749
221841_s_at	人類cDNA FLJ38575 fis，純系 HCHON2007046，mRNA 序列	Hs.376206
214858_at	人類純系 24566 mRNA 序列	Hs.133342
212464_s_at	纖維蛋白1	Hs.287820
206510_at	sine oculis同源盒同系物2(果蠅)	Hs.101937
216246_at	核糖體蛋白S20	Hs.173717
200923_at	凝集素，半乳糖苷結合，可溶性，3結合蛋白	Hs.79339
221989_at	核糖體蛋白L10	Hs.29797
211284_s_at	顆粒體蛋白	Hs.180577
209173_at	前梯度2同系物(非洲爪蟾)	Hs.91011
200924_s_at	溶質載體蛋白族3(二元及中性胺基酸轉運子之活化劑)，組員2	Hs.79748
212859_x_at	---	---
213109_at	KIAA0551蛋白	Hs.170204

表 A3：WT(威爾科克森測試)：在P值<0.05且a 2倍變化截止值下，總共識別了38種基因。此38種基因集傳遞了80%的LOOCV精確度。該等基因係藉由其顯著性(P值)來排列。

探針	基因描述	Unigene
210761_s_at	生長因子受體結合蛋白7	Hs.86859
201931_at	電子轉移黃素蛋白， α 多肽(戊二酸尿II)	Hs.169919
219429_at	脂肪酸羥化酶	Hs.249163
204285_s_at	佛波醇12-肉豆蔻酸酯13-乙酸酯誘導之蛋白1	Hs.96
209603_at	GATA結合蛋白3	Hs.169946
206165_s_at	鈣活化之氯化物通道，族組員2	Hs.241551
216836_s_at	v-erb-b2成紅血細胞白血病病毒致癌基因同系物2，衍生自神經/膠母細胞瘤之致癌基因同系物(鳥類)	Hs.323910
203627_at	類人類胰島素生長因子1受體mRNA，3'序列，mRNA序列	Hs.405998
205225_at	雌激素受體1	Hs.1657
215465_at	ATP-結合盒，亞族A(ABC1)，組員12	Hs.134585
203628_at	類人類胰島素生長因子1受體mRNA，3'序列，mRNA序列	Hs.405998
202991_at	含有START域之3	Hs.77628
208891_at	雙特異型磷酸酯酶6	Hs.180383
214451_at	轉錄因子AP-2 β (活化增強子結合蛋白2 β)	Hs.33102
204508_s_at	假設蛋白FLJ20151	Hs.279916
202376_at	絲胺酸(或半胱胺酸)蛋白酶抑制劑，分化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員3	Hs.234726
200832_s_at	硬脂醯基輔酶A去飽和酶(δ -9-去飽和酶)	Hs.119597
205307_s_at	犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)	Hs.107318
203060_s_at	3'-磷酸腺苷5'-磷酸硫酸合成酶2	Hs.274230
201963_at	脂肪酸輔酶A連接酶，長鏈2	Hs.154890
209602_s_at	GATA結合蛋白3	Hs.169946
211138_s_at	犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)	Hs.107318
39248_at	水通道蛋白3	Hs.234642
220149_at	假設蛋白FLJ22671	Hs.193745
55616_at	假設基因MGC9753	Hs.91668
205306_x_at	犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)	Hs.107318
205862_at	GREB1蛋白	Hs.193914
217388_s_at	犬尿胺酸酶(L-犬尿胺酸水解酶)	Hs.169139

204942_s_at	醛脫氫酶3族，組員B2	Hs.87539
202218_s_at	脂肪酸去飽和酶2	Hs.184641
213557_at	ESTs，微弱地類似於普遍轉錄之三十四肽重複基因，Y染色體； Y染色體[人類]上普遍轉錄之TPR基因[人類]	Hs.14691
211657_at	癌胚抗原相關細胞黏附分子6(非特異性交叉反應抗原)	Hs.73848
214598_at	密蛋白8	Hs.162209
218532_s_at	假設蛋白FLJ20152	Hs.82273
202917_s_at	S100鈣結合蛋白A8(鈣粒蛋白A)	Hs.100000
208792_s_at	簇蛋白(補體溶離抑制劑，SP-40,40硫酸化糖蛋白2，載脂蛋白J)	Hs.75106
215659_at	人類cDNA：FLJ21521 fis，純系COL05880，mRNA序列	Hs.306777
201525_at	載脂蛋白D	Hs.75736

表 A4：接著識別該等三個基因集 (SAM-88、GR-251、WT-38) 中之 13 種 "共同" 基因。該 13 種組成基因達成了 84% 的 LOOCV 分類精確度。本質上，該等 13 種 "共同基因" 為強健的顯著性標誌，且可達成與其它 "完全" 標誌集相當之效能。

探針_ID	Unigene	參考序列全長	位置
39248_at	Hs.234642	NM_004925 // 水通道蛋白 3	Chr: 9p13
201525_at	Hs.75736	NM_001647 // 載脂蛋白D前驅體	Chr:3q26.2-qter
202991_at	Hs.77628	NM_006804 // 類固醇產生急性調節性蛋白相關	Chr: 17q11-q12
203628_at	Hs.405998	---	---
205307_s_at	Hs.107318	NM_003679 // 犬尿胺酸3-單加氧酶(犬尿胺酸3-羥化酶)	Chr:1q42-q44
210761_s_at	Hs.86859	NM_005310 // 生長因子受體結合蛋白7	Chr: 17q21.1
211657_at	Hs.73848	NM_002483 // 癌胚抗原相關細胞黏附分子6(非特異性交叉反應抗原)	Chr: 19q13.2
213557_at	Hs. 14691	---	---
214451_at	Hs.33102	NM_003221 // 轉錄因子AP-2 β (活化增強子結合蛋白 2 β)	Chr: 6p12
215465_at	Hs. 134585	NM_015657 // ATP-結合盒，亞族A，組員12異型 b /// NM_173076 // ATP-結合盒，亞族A，組員12異型 a	Chr: 2q35
219429_at	Hs.249163	---	Chr: 16q23
220149_at	Hs. 193745	NM_024861 // 假設蛋白FLJ22671	Chr: 2q37.3
210930_s_at	Hs.323910	NM_004448 // v-erb-b2成紅血細胞白血病病毒致癌基因同系物2，衍生自神經/膠母細胞瘤致癌基因同系物	Chr: 17q11.2-q12

表 L1： SAM-133基因之ID查找表

SAM-133

排名	探針_ID	Unigene	GenBank
1	205225_at	Hs.1657	NM_000125.1
2	209603_at	Hs. 169946	AI796169
3	204508_s_at	Hs.279916	BC001012.1
4	209604_s_at	Hs. 169946	BC003070.1
5	209602_s_at	Hs.169946	AI796169
6	206754_s_at	Hs.1360	NM_000767.2
7	203963_at	Hs.5338	NM_001218.2
8	214164_x_at	Hs.5344	BF752277
9	212956_at	Hs.90419	AI348094
10	215867_x_at	Hs.5344	AL050025.1
11	210735_s_at	Hs.5338	BC000278.1
12	214440_at	Hs.155956	NM_000662.1
13	202089_s_at	Hs.79136	NM_012319.2
14	210085_s_at	Hs.279928	AF230929.1
15	205862_at	Hs.193914	NM_014668.1
16	202088_at	Hs.79136	AI635449
17	211712_s_at		BC005830.1
18	206401_s_at	Hs.101174	J03778.1
19	215304_at	Hs.159264	U79293.1
20	218195_at	Hs 15929	NM_024573.1
21	212195_at	Hs. 71968	AL049265.1
22	203928_x_at	Hs.101174	AI870749
23	209460_at	Hs.283675	AF237813.1
24	212960_at	Hs. 90419	BE646554
25	209443_at	Hs.76353	J02639.1
26	209173_at	Hs.91011	AF088867.1
27	203071_at	Hs.82222	NM_004636.1
28	203571_s_at	Hs.74120	NM_006829.1
29	205354_at	Hs.81131	NM_000156.3
30	213712_at	Hs.30504	BF508639
31	41660_at		
32	220744_s_at	Hs.70202	NM_018262.1
33	204798_at	Hs.1334	NM_005375.1
34	215552_s_at	Hs.272288	AI073549

35	209339_at	Hs.20191	U76248.1
36	210272_at	Hs.330780	M29873.1
37	205186_at	Hs.33846	NM_003462.2
38	207414_s_at	Hs.170414	NM_002570.1
39	205009_at	HS. 1406	NM_003225.1
40	203628_at	Hs.239176	H05812
41	211323_s_at	Hs.198443	L38019.1
42	201825_s_at	Hs.238126	AL572542
43	211234_x_at	Hs.1657	AF258449.1
44	209459_s_al	Hs.283675	AF237813.1
45	212196_at	Hs.71968	AW242916
46	203438_at	Hs.155223	AI435828
47	217838_s_at	Hs.241471	NM_016337.1
48	204041_at	Hs.82163	NM_000898.1
49	203929_s_at	Hs.101174	AI056359
50	200670_at	Hs.149923	NM_005080.1
51	219414_at	Hs.12079	NM_022131.1
52	203627_at	Hs.239176	AI830698
53	208451_s_at	Hs.278625	NM_000592.2
54	213419_at	Hs.324125	U62325.1
55	205768_s_at	Hs. 11729	NM_003645.1
56	204862_s_at	Hs.81687	NM_002513.1
57	210480_s_at	Hs.22564	U90236.2
58	205696_s_at	Hs.105445	NM_005264.1
59	203685_at	Hs.79241	NM_000633..1
60	218976_at	Hs.260720	NM_021800.1
61	219197_s_at	Hs.222399	AI424243
62	202996_at	Hs.82520	NM_021173.1
63	205734_s_at	Hs.38070	AI990465
64	211235_s_at	Hs.1657	AF258450.1
65	211000_s_at	Hs.82065	AB015706.1
66	217190_x_at	Hs.247976	S67777
67	202752_x_at	Hs.22891	NM_012244.1
68	201754_at	Hs.74649	NM_004374.1
69	204623_at	Hs.82961	NM_003226.1
70	207038_at	Hs. 114924	NM_004694.1
71	212637_s_at	Hs.324275	AU155187
72	208682_s_at	Hs.4943	AF126181.1

73	218502_s_at	Hs. 26102	NM 014112.1
74	202376_at	Hs. 234726	NM_0010852
75	215616_s_at	Hs. 301011	AB020683.1
76	211233_x_at	Hs.1657	M12674.1
77	205081_at	Hs. 17409	NM_001311.1
78	214428_x_at	Hs. 170250	K02403.1
79	209696_at	Hs.574	D26054.1
80	219682_s_at	Hs.332150	NM_016569.1
81	212496_s_at	Hs. 301011	BE256900
82	203108_at	Hs.194691	NM_003979.2
83	206107_at	Hs.65756	NM_003834.1
84	218806_s_at	Hs.267659	AF118887.1
85	209581_at	Hs.37189	BC001387.1
86	213412_at	Hs.25527	NM_014428.1
87	212638_s_at	Hs.324275	BF131791
88	206469_x_at	Hs. 284236	NM_012067.1
89	210652_s_at	HS.125783	BC004399.1
90	216381_x_at	Hs.284236	AL035413
91	216092_s_at	Hs22891	AL365347.1
92	208788_at	Hs.250175	AL136939.1
93	204792_s_at	Hs.111862	NM_014714.1
94	207847_s_at	Hs.89603	NM_002456.1
95	213201_s_at	Hs.73980	AJ011712
96	204497_at	Hs.20196	AB011092.1
97	222314_x_at	Hs.205660	AW970881
98	222212_s_at	Hs.285976	AK001105.1
99	219919_s_at	Hs.279808	NM_018276.1
100	214053_at	Hs.7888	AW772192
101	204934_s_at	Hs.823	NM_002151.1
102	216109_at	Hs.306803	AK025348.1
103	203749_s_at	Hs. 250505	AI806984
104	220329_s_at	HS238270	NM_017909.1
105	204881_s_at	Hs.152601	NM_003358.1
106	208305_at	Hs.2905	NM_000926.1
107	209623_at	Hs.167531	AW439494
108	218450_at	Hs. 108675	NM_015987.1
109	204343_at	Hs.26630	NM_001089.1
110	219051_x_at	Hs. 124915	NM_024042.1

111	205471_s_at	Hs.63931	AW772082
112	203439_s_at	Hs.155223	BC000658.1
113	204863_s_at	Hs.82065	BE856546
114	203289_s_at	Hs.19699	BE791629
115	221765_at	Hs.23703	AI378044
116	219001_s_at	Hs.317589	NM_024345.1
117	220581_at	Hs.287738	NM_025059.1
118	211596_s_at		AB050468.1
119	205645_at	Hs.80667	NM_004726.1
120	219663_s_at	Hs.157527	NM_025268.1
121	205380_at	Hs.15456	NM_002614.1
122	201508_at	Hs.1516	NM_001552.1
1	215729_s_at	Hs.9030	BE542323
2	201983_s_at	Hs.77432	AW157070
3	204914_s_at	Hs.32964	AW157202
4	204913_s_at	Hs.32964	AI360875
5	205646_s_at	Hs.89506	NM_000280.1
6	207030_s_at	Hs.10526	NM_001321.1
7	204915_s_at	Hs.32964	AB028641.1
8	203021_at	Hs.251754	NM_003064.1
9	209800_at	Hs.115947	AF061812.1
10	203234_al	Hs.77573	NM_003364.1
11	201984_s_at	Hs.77432	NM_005228.1

表 L2：表 2 基因之查找表

表 2

探針_ID	Unigene	GenBank
205225_at	Hs.1657	NM_000125.1
205186_at	Hs.406050	NM_003462.2
201754_at	Hs.351875	NM_004374.1
210085_s_at	Hs.279928	AF230929.1
214440_at	Hs.155956	NM_000662.1
206754_s_at	Hs.1360	NM_000767.2
203749_s_at	Hs.361071	AI806984
215552_s_at	Hs.239176	AI073549
209443_at	Hs.76353	J02639.1
216109_at	Hs.306803	AK025348.1
203685_at	Hs.79241	NM_000633.1
205862_at	Hs.193914	NM_014668.1
217838_s_at	Hs.241471	NM_016337.1
209603_at	Hs.169946	AI796169
212195_at	Hs.71968	AL049265.1
212637_s_at	Hs.355977	AU155187
205696_s_at	Hs.105445	NM_005264.1
210652_s_at	Hs.125783	BC004399.1
205734_s_at	Hs.38070	AI990465
211000_s_at	Hs.82065	AB015706.1
206107_at	Hs.65756	NM_003834.1
203628_at	Hs.405998	H05812
204934_s_at	Hs.823	NM_002151.1
203071_at	Hs.82222	NM_004636.1
204881_s_at	Hs.432605	NM_003358.1
210272_at	Hs.330780	M29873.1
213201_s_at	Hs.73980	AJ011712
206401_s_at	Hs.101174	J03778.1
209339_at	Hs.20191	U76248.1
208305_at	Hs.2905	NM_000926.1
212956_at	Hs.90419	AI348094
214164_x_at	Hs.279916	BF752277
204343_at	Hs.26630	NM_001089.1
203963_at	Hs.5338	NM_001218.2

207038_at	Hs.114924	NM_004694.1
218195_at	Hs.15929	NM_024573.1
220329_s_at	Hs.238270	NM_017909.1
218502_s_at	Hs.26102	NM_014112.1
219414_at	Hs.12079	NM_022131.1
202376_at	Hs.234726	NM_001085.2
218806_s_at	Hs.267659	AF118887.1
202089_s_at	Hs.79136	NM_012319.2
213712_at	Hs.432587	BF508639
204497_at	Hs.20196	AB011092.1
215616_s_at	Hs.301011	AB020683.1
218450_at	Hs.294133	NM_015987.1
203438_at	Hs.155223	AI435828
208451_s_at	Hs.433721	NM_000592.2
205768_s_at	Hs.11729	NM_003645.1
219682_s_at	Hs.267182	NM_016569.1
204508_s_at	Hs.279916	BC001012.1
203963_at	Hs.5338	NM_001218.2
209603_at	Hs.169946	AI796169
208788_at	Hs.250175	AL136939.1
212637_s_at	Hs.355977	AU155187
200670_at	Hs.149923	NM_005080.1
203571_s_at	Hs.74120	NM_006829.1
208682_s_at	Hs.4943	AF126181.1
209173_at	Hs.91011	AF088867.1
201754_at	Hs.351875	NM_004374.1
206469_x_at	Hs.284236	NM_012067.1
213412_at	Hs.25527	NM_014428.1
222212_s_at	Hs.285976	AK001105.1
211323_s_at	Hs.198443	L38019.1
209696_at	Hs.574	D26054.1
212956_at	Hs.90419	AI348094
218195_at	Hs.15929	NM_024573.1
202089_s_at	Hs.79136	NM_012319.2
209623_at	Hs.167531	AW439494
210272_at	Hs.330780	M29873.1
204623_at	Hs.82961	NM_003226.1
215304_at	Hs.159264	U79293.1

214440_at	Hs.155956	NM_000662.1
205862_at	Hs.193914	NM_014668.1
203108_at	Hs.194691	NM_003979.2
207038_at	Hs.114924	NM_004694.1
205186_at	Hs.406050	NM_003462.2
202752_x_at	Hs.22891	NM_012244.1
220744_s_at	Hs.70202	NM_018262.1
219414_at	Hs.12079	NM_022131.1
204798_at	Hs.1334	NM_005375.1
205009_at	Hs.350470	NM_003225.1
219051_x_al	Hs.124915	NM_024042.1
205471_s_al	Hs.63931	AW772082
207847_s_al	Hs.89603	NM_002456.1
208451_s_al	Hs.433721	NM_000592.2
205081_at	Hs.423190	NM_001311.1
209459_s_al	Hs.283675	AF237813.1
203071__at	Hs.82222	NM_004636.1
209581_at	Hs.37189	BC001387.1
204343_at	Hs.26630	NM_001089.1
206401_s_at	Hs.101174	J03778.1
210480_s_at	Hs.385834	U90236.2
201825_s_at	Hs.238126	AL572542
203749_s_at	Hs.361071	AI806984
218806_s_at	Hs.267659	AF118887.1
210652_s_at	Hs.125783	BC004399.1
205225_at	Hs.1657	NM_000125.1
205768_s_at	Hs.11729	NM_003645.1
219682_s_at	Hs.332150	NM_016569.1

表 L3: 表 S4 基因之查找表

Unigene	GenBank
Hs.106642	BF589529
Hs.25960	AF320053.1
Hs.1892	NM_002686.1
Hs.289104	NM_014274.1
Hs.165950	NM_002011.2
Hs.173035	AF338650.1
Hs.86859	AB008790.1
Hs.272207	NM_017533.1
Hs.103707	AW192795
Hs.274550	AA074145
Hs.100000	AW238654
Hs.54609	NM_014291.1
Hs.85050	NM_002667.1
Hs.239934	AL022316
Hs.194236	NM_000230.1
Hs.103395	NM_024709.1
Hs.107318	NM_003679.1
Hs.1735	NM_002193.1
Hs.155109	NM_002153.1
Hs.26770	NM_001446.1
Hs.278388	NM_000608.1
Hs.251754	NM_003064.1
Hs.378774	NM_001615.2
Hs.51515	AA053967
Hs.149195	NM_016233.1
Hs.78344	AI889739
Hs.112405	NM_002965.2
Hs.417091	AF052117.1
Hs.57664	NM_000888.3
Hs.154078	NM_004139.1
Hs.100014	NM_007325.1
Hs.193606	AA343027
Hs.202949	AK027231.1
Hs.84072	NM_004616.1
Hs.323910	AF177761.2

Hs.76780	NM_006741.1
Hs.225962	NM_014354.1
Hs.165619	NM_017717.2
Hs.127428	AI246769
Hs.2899	NM_002150.1
Hs.105938	NM_002343.1
Hs.193143	AK022610.1
Hs.1915	NM_004476.1
Hs.160786	NM_000050.1
Hs.23881	AI920979
Hs.3110	MM_000686.2
Hs.180142	NM_017422.2
Hs.169919	NM_000126.1
Hs.112408	NM_002963.2
Hs.96	NM_021127.1
Hs.33846	NM_003462.2
Hs.1360	NM_000767.2
Hs.1657	NM_000125.1
Hs.194689	AF120274.1
Hs.50964	NM_001712.1
Hs.23703	BF970427
Hs.193914	NM_014668.1
Hs.250505	AI806984
Hs.279928	AF230929.1
Hs.156637	NM_012116.1
Hs.169946	AI796169
Hs.4243	NM_024522.1
Hs.111801	NM_015908.1
Hs.155485	NM_005339.2
Hs.99603	NM_024701.1
Hs.55481	NM_003447.1
Hs.306803	AK025348.1
Hs.239176	NM_000875.2
Hs.823	NM_002151.1
Hs.203845	NM_022358.1
Hs.432605	NM_003358.1
Hs.330780	M29873.1
Hs.32981	U38276

Hs.101174	NM_016835.1
Hs.17752	NM_015900.1
Hs.406646	未找到資料
Hs.351875	NM_004374.1
Hs.20196	AB011092.1
Hs.331584	AF326966.1
Hs.272288	AI073549
Hs.12079	NM_022131.1
Hs.82065	NM_002184.1
Hs.372446	NM_007202.1
Hs.155956	NM_000662.1
Hs.278850	NM_024935.1
Hs.247955	NM_001322.1
Hs.76067	NM_001540.2
Hs.61289	AL157424.1
Unigene	
Hs.334514	NM_032794
Hs.4943	NM_177433
Hs.1892	NM_002686
Hs.321576	NM_006458
Hs.91668	BF033007
Hs.274260	NM_001171
Hs.14368	NM_003022
Hs.86859	NM_005310
Hs.59889	NM_005518
Hs.165950	NM_002011
Hs.83190	NM_004104
Hs.89603	NM_002456
Hs.29724	NM_024613.1
Hs.12068	NM_000755
Hs.279916	NM_017689
Hs.169946	NM_002051
Hs.355977	NM_007013
Hs.33102	NM_003221
Hs.90419	XM_093895
Hs.38972	NM_005727
Hs.31034	NM_003847
Hs.132136	NM_004858

Hs.91668	BF033007
Hs.70604	NM_004496
Hs.234642	NM_004925
Hs.323910	NM_004448
Hs.198443	NM_002222
Hs.197922	NM_018584.1
Hs.87539	NM_000695
Hs.381412	未找到資料
Hs.180383	NM_001946
Hs.5338	NM_001218
Hs.406515	NM_000903
Hs.8910	NM_020379
Hs.6168	NM_014861
Hs.119597	NM_005063
Hs.574	NM_000507
Hs.326525	NM_009589
Hs.149923	NM_005080
Hs.167531	NM_022132
Hs.184376	NM_003825
Hs.301947	NM_014509
Hs.91011	NM_006408
Hs.114556	NM_017699
Hs.432970	NM_006431
Hs.300697	AK090461
Hs.84072	NM_004616
Hs.878	NM_003104

【圖式簡單說明】

圖 1：具有低預測強度("低置信度")之腫瘤的識別。

相對於樣品之預測強度(PS, y軸)繪製訓練組(a)及測試組(b)中之每一樣品(x軸)。訓練資料集由55種腫瘤組成，且測試資料集由41種腫瘤組成。將展現高正PS值的樣品分類為ER+，而具有高負PS值的樣品為ER-。藍色樣品經正確地分類，而紅色樣品則經錯誤分類。大體而言，在訓練及測試

腫瘤中均觀察到一組"低置信度"樣品(灰色框)。

圖2：比較"高"與"低置信度"腫瘤之臨床表現的卡普蘭-梅爾分析。

(a)與(b)中之整體存活資料係自Stanford資料集獲得(9)，而(c)與(d)中之遠距離轉移時間係自Rosetta資料集獲得(10)。將患有"高置信度"腫瘤之患者描繪為綠色，而將患有"低置信度"腫瘤之患者描繪為粉紅色。a)患有"高"(60位患者)與"低置信度"(14位患者)腫瘤(不管ER狀況)之患者的整體存活；b)患有ER+"高"(48)及"低置信度"(7)腫瘤之患者的整體存活；c)患有"高"(82)及"低置信度"(15)腫瘤(不管ER狀況)之患者自最初腫瘤診斷至出現遠距離轉移之時間；d)患有ER+"高"(63)及"低置信度"(5)腫瘤之患者自最初腫瘤診斷至出現遠距離轉移之時間。

圖3：在低相對於高置信度樣品中的ER-相關基因之廣泛擾動。

所描繪之(a)與(b)係在(a)ER+/高(黃色)及ER+/低(青綠色)與(b)ER-/高(深藍色)及ER-/低(粉紅色)樣品中關於ER+狀況呈陽性之頂部122種ER判別基因(自SAM-133基因集獲得，參見正文)之相對表現量。沿x軸的122種基因之次序係藉由其S2N比率來確定(參看材料及方法)。一特定基因之S2N量度考慮到兩種類別之間平均表現量的差值及正比較之每一類別中的彼基因之表現的標準偏差。應注意，在(a)與(b)中之122種基因的特定次序視其S2N比率而不同(表2)。所描繪之(c)與(d)係在(c)ER+/高(黃色)及ER+/低(青綠

色)及(d)ER-/高(深藍色)及ER-/低(粉紅色)樣品中關於ER+狀況呈陰性之頂部54種ER判別基因(11種屬於SAM-133基因集，參見對於細節之補充資訊)之相對表現量。觀察到比(a)與(b)中顯著較少的擾動。

圖4：ERBB2+與多個乳癌表現資料集上之"低置信度"預測相關聯。資料係取自參考文獻3。a)表現高含量ERBB2之腫瘤樣品(行)及實體連接至17q ERBB2染色體座位之其它基因(MLN64，GRB7)(列)的識別。高表現係由紅色方框表示。腫瘤樣品5141、8443、7636、4527、5955、10444、5985、6936展現出ERBB2及ERBB2連接之基因的高表現，而6080與10188展現出提高但較弱之表現。b)對ER分類之ANN模型的總結(自參考文獻3之圖1b改編)。歸類為ER+之腫瘤樣品為藍色，而ER-腫瘤為橙色。預測置信度係由每一樣品之標準偏差(SD)來表示，其中"低置信度"樣品具有高的SD。描繪了8個"高表現"ERBB2+ve樣品(在樣品SD之左側或右側之ERBB2)。應注意，具有高SD之腫瘤樣品趨向於成為ERBB2+ve。

圖5：主組分分析(PCA)，一種提供複數資料集在減小的、易於視覺化之空間中的投影且提供對基於SAM-133基因集判別樣品之清晰度的有用視覺評定之數學技術。ER+與ER-腫瘤可彼此清晰地區分，而ERBB2+樣品處於中部空間中。顏色編碼機制：ER+ERBB2-，黃色；ER+ERBB2+，青綠色；ER-ERBB2-，藍色；且ER-ERBB2+，粉紅色。顏色編碼機制：ER+ERBB2-，黃色；ER+ERBB2+，青綠色；ER-ERBB2-，

藍色；且ER-ERBB2+，粉紅色。X軸為主組分1，且Y軸為組分2。除去兩個ER-樣品，位於紅線左側之樣品均為ER+樣品；而除去一個錯誤分類，右側的樣品均為ER-樣品。靠近邊界之樣品(在方框中)均為ERBB2+。

圖6展示患有"高置信度"ER陰性腫瘤之患者相對於隱藏(harboring)"低置信度"ER陰性腫瘤之患者的臨床預後。分析兩個稱作"Rosetta"與"Stanford"資料集之獨立的資料集。圖6(a)展示Rosetta腫瘤：量測無復發存活。11/19(58%)之高置信度患者在5年內產生遠距離轉移；而在低置信度ER-中，數目為8/10(80%)。圖6(b)展示Stanford腫瘤：量測整體存活。7/12(58%)之高置信度患者死亡；而在低置信度ER-中，數目為5/7(71%)。

圖7展示在Stanford與Rosetta資料集中具有低預測強度("低置信度")之腫瘤的識別。

五、中文發明摘要：

將乳腺腫瘤分類為雌激素受體陽性及陰性(ER+及ER-)亞型係乳癌治療中的一項重要區分。ER定型經常係利用基因之表現輪廓來進行，其中已知該等基因之表現將受ER活性的影響。基於此等表現資料無法確信地將某些腫瘤指定為特定的ER類型。本發明者已發現此等"低置信度"腫瘤構成了一獨特生物亞型之乳腺腫瘤，與高置信度腫瘤相比，其與顯著惡化之整體存活率相關聯。本發明提供能夠區分低置信度與高置信度腫瘤之基因集以及用於執行乳腺腫瘤的恰當分類之方法及裝置。

六、英文發明摘要：

十、申請專利範圍：

1. 一種用於將一乳腺腫瘤樣品分類為"低置信度"或"高置信度"之方法，該方法包含：提供該乳腺腫瘤樣品之表現輪廓，其中該表現輪廓包含一包含來自表S4之至少5種基因的多基因分類器之表現量；且基於該表現輪廓將該腫瘤分類為高或低置信度腫瘤。
2. 如請求項1之方法，其包含判定該樣品之雌激素受體(ER)狀況。
3. 如請求項1或2之方法，其包含如下步驟：
 - (a)獲得來自由一患者所獲得之乳腺腫瘤樣品之表現產物；
 - (b)藉由使該等表現產物與結合組員接觸來確定一包含表S4中所識別的至少5種基因之多基因分類器之表現量，其中每一結合組員均能夠特異地結合至該多基因分類器之一表現產物；及
 - (c)基於該等表現量來識別該患者中低置信度乳腺腫瘤之存在。
4. 如請求項3之方法，其中該等表現產物為cDNA，且該等結合組員為能夠特異地與該cDNA雜交之核酸探針。
5. 如請求項3之方法，其中該等表現產物為RNA或mRNA，且該等結合組員為能夠特異地與該RNA或mRNA雜交且在一PCR中將其擴增之核酸引子。
6. 如請求項3之方法，其中該等表現產物為多肽，且該等結合組員為能夠特異地結合至該等多肽之抗體結合域。

7. 如請求項3之方法，其包含：將來自測試中之該乳腺腫瘤樣品的該等表現產物之結合輪廓與一具有其它先前所獲得之輪廓及/或一作為低置信度腫瘤之存在的特徵之先前所確定之"標準"輪廓的資料庫進行比較。
8. 如請求項7之方法，其中該比較係藉由一電腦來執行，該電腦經程式化以報導測試中之該輪廓與該等標準輪廓之間的統計學相似性，從而可進行歸類。
9. 如請求項1或2之方法，其中分類該乳腺腫瘤樣品之該步驟包含使用加權表決法、支持向量機及/或階層式叢集法。
10. 如請求項1或2之方法，其中該多基因分類器包含來自表S4(a)之基因、來自表S4(b)之基因、或上述兩者中任一者之子集。
11. 如請求項10之方法，其中該子集之基因係源自表S4(a)或表S4(b)之上半部分。
12. 如請求項10之方法，其中該多基因分類器包含來自表S4(a)及/或表S4(b)之上調及下調基因的混合物。
13. 一種將一乳腺腫瘤樣品分類為"低置信度"或"高置信度"之方法，該方法包含：提供該乳腺腫瘤樣品之表現輪廓，其中該表現輪廓包含一包含來自表2之至少5種基因的多基因分類器之表現量；且基於該表現輪廓將該腫瘤分類為高或低置信度腫瘤。
14. 如請求項13之方法，其包含判定該樣品之雌激素受體(ER)狀況。
15. 如請求項13或14之方法，其包含如下步驟：

(a)獲得來自一患者所獲得之乳腺腫瘤樣品之表現產物；

(b)藉由使該等表現產物與結合組員接觸來確定一包含表2中所識別的至少5種基因之多基因分類器的表現量，其中每一結合組員均能夠特異地結合至該多基因分類器之一表現產物；及

(c)基於該等表現量來識別該患者中低置信度乳腺腫瘤之存在。

16. 如請求項15之方法，其中該等表現產物為cDNA，且該等結合組員為能夠特異地與該cDNA雜交之核酸探針。
17. 如請求項15之方法，其中該等表現產物為RNA或mRNA，且該等結合組員為能夠特異地與該RNA或mRNA雜交且在一PCR中將其擴增之核酸引子。
18. 如請求項15之方法，其中該等表現產物為多肽，且該等結合組員為能夠特異地結合至該等多肽之抗體結合域。
19. 如請求項13或14之方法，其包含：將來自測試中之該乳腺腫瘤樣品的該等表現產物之結合輪廓與一具有其它先前所獲得之輪廓及/或一作為低置信度腫瘤之存在的特徵之先前所確定之"標準輪廓"的資料庫進行比較。
20. 如請求項19之方法，其中該比較係藉由一電腦來執行，該電腦經程式化以報導測試中之該輪廓與該等標準輪廓之間的統計學相似性，從而可進行分類。
21. 如請求項13或14之方法，其中分類該乳腺腫瘤樣品之該步驟包含使用加權表決法、支持向量機及/或階層式叢集

法。

22. 如請求項13或14之方法，其中該多基因分類器包含來自表2(a)之基因、來自表2(b)之基因或上述兩者中任一者之子集。
23. 如請求項22之方法，其中該子集之基因係源自表2(a)或表2(b)之上半部分。
24. 如請求項22之方法，其中該多基因分類器包含表2(a)及/或表2(b)之上調及下調基因的混合物。
25. 一種用於將一乳腺腫瘤樣品分類為"低置信度"或"高置信度"之方法，該方法包含：提供該乳腺腫瘤樣品之表現輪廓，其中該表現輪廓包含一包含來自表A1及/或表A2及/或表A3及/或表A4中之至少5種基因的多基因分類器之表現量；及基於該表現輪廓將該腫瘤分類為高或低置信度腫瘤。
26. 如請求項25之方法，其包含如下步驟：
 - (a)獲得來自由一患者所獲得之乳腺腫瘤樣品之表現產物；
 - (b)藉由使該等表現產物與結合組員接觸來確定一包含表A1及/或表A2及/或表A3及/或表A4中所識別之至少5種基因的多基因分類器之表現量，其中每一結合組員均能夠特異地結合至該多基因分類器之一表現產物；及
 - (c)基於該等表現量來識別該患者中低置信度乳腺腫瘤之存在。
27. 如請求項26之方法，其中該等表現產物為cDNA，且該等

結合組員為能夠特異地與該cDNA雜交之核酸探針。

28. 如請求項26之方法，其中該等表現產物為RNA或mRNA，且該等結合組員為能夠特異地與該RNA或mRNA雜交且在一PCR中將其擴增之核酸引子。
29. 如請求項26之方法，其中該等表現產物為多肽，且該等結合組員為能夠特異地結合至該等多肽之抗體結合域。
30. 如請求項26至29中任一項之方法，其包含將來自測試中之該乳腺腫瘤樣品的該等表現產物之結合輪廓與一具有其它先前所獲得之輪廓及/或一作為低置信度腫瘤之存在的特徵之先前所確定之"標準"輪廓的資料庫進行比較。
31. 如請求項30之方法，其中該比較係藉由一電腦來執行，該電腦經程式化以報導測試中之該輪廓與該等標準輪廓之間的統計學相似性，從而可進行分類。
32. 如請求項25至29中任一項之方法，其中分類該乳腺腫瘤樣品之該步驟包含使用加權表決法、支持向量機及/或階層式叢集法。
33. 如請求項25至29中任一項之方法，其中該多基因分類器包含來自表A4之基因或其子集。
34. 一種產生一乳腺腫瘤樣品之核酸表現輪廓之方法，其包含如下步驟：
 - (a)將表現產物與該乳腺腫瘤樣品分離；
 - (b)識別一包含選自表S4、表2、表A1、表A2、表A3及表A4中任一表中之至少5種基因的多基因分類器之表現量；

(c)自該等表現量產生該乳腺腫瘤樣品之一表現輪廓。

35. 如請求項34之方法，其包含如下步驟：

(a)將表現產物與一乳腺腫瘤樣品分離；

(b)使該等表現產物與一包含至少5個結合組員之多基因分類器接觸以自該多基因分類器之該等表現量創建一腫瘤樣品之第一表現輪廓，其中該等至少5個結合組員均能夠特異且獨立地結合至選自表S4或表2、或獨立地選自表A1及/或表A2及/或表A3及/或表A4中之複數種基因的表现產物；

(c)將該表現輪廓與一作為一高置信度腫瘤及/或一低置信度乳腺腫瘤之特徵的表現輪廓進行比較。

36. 一種基因表現輪廓資料庫，其包含高置信度及/或低置信度乳腺腫瘤樣品之複數種基因表現輪廓，其中每一基因表現輪廓均係得自一包含選自表S4或表2、或獨立地選自表A1及/或表A2及/或表A3及/或表A4中的至少5種基因之多基因分類器，且其中該資料庫可擷取地保存於一資料載體上。

37. 如請求項36之表現輪廓資料庫，其中組成該資料庫之該等表現輪廓係藉由請求項34或請求項35之方法產生。

38. 一種用於將一乳腺腫瘤樣品分類為"高置信度"或"低置信度"之裝置，其包含附著至一固相支持體之複數個結合組員，每一結合組員均能夠特異地結合至一包含來自表S4、表2、表A1、表A2、表A3及表A4中之一或多個表中的至少5種基因的多基因分類器之一表現產物。

39. 如請求項38之裝置，其包含能夠結合至來自該等表中每一者中之複數種基因的表現產物之結合組員。
40. 如請求項38或39之裝置，其包含能夠特異且獨立地結合至表A4中所識別之所有基因的表現產物之結合組員。
41. 如請求項38或39之裝置，其包含一微陣列，其中該等結合組員為能夠特異地與RNA或mRNA表現產物或自其衍生之cDNA雜交之核酸序列。
42. 一種用於將一乳腺腫瘤樣品分類為"高置信度"或"低置信度"之套組，該套組包含複數個結合組員及一偵測試劑，其中每一結合組員均能夠特異地結合至一包含表S4、表2、表A1、表A2、表A3及表A4中任一或多個表中所識別的至少5種基因的多基因分類器之一表現產物。
43. 如請求項42之套組，其中該等結合組員為固定至一或多個固相支持體之抗體結合域或核酸序列。
44. 如請求項43之套組，其包含一微陣列。
45. 如請求項42之套組，其中該等結合組員為能夠結合至該等表現產物以使其可在一PCR中擴增之核酸引子。
46. 如請求項42至45中任一項之套組，其進一步包含可擷取地保存於一資料載體上用以與一測試樣品之表現輪廓進行比較的一或多個標準表現輪廓。
47. 如請求項46之套組，其中該一或多個標準表現輪廓係根據請求項34或請求項35之方法產生。

十一、圖式：

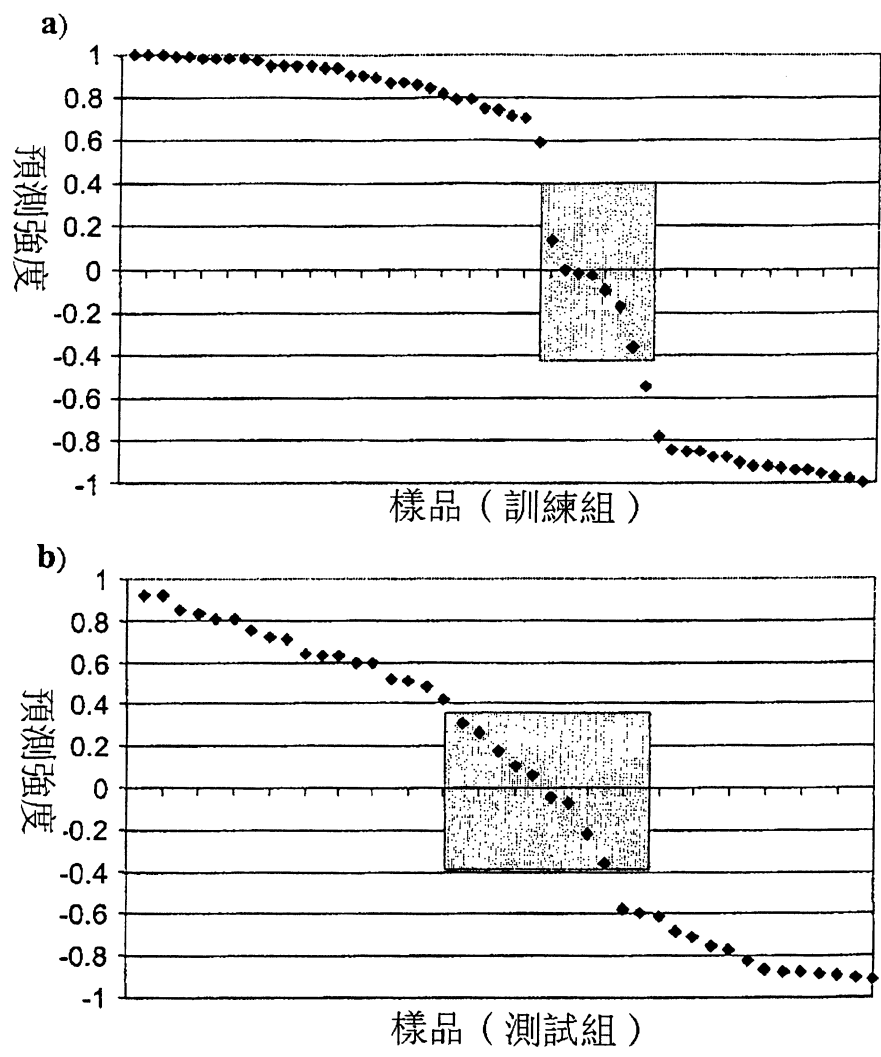


圖 1

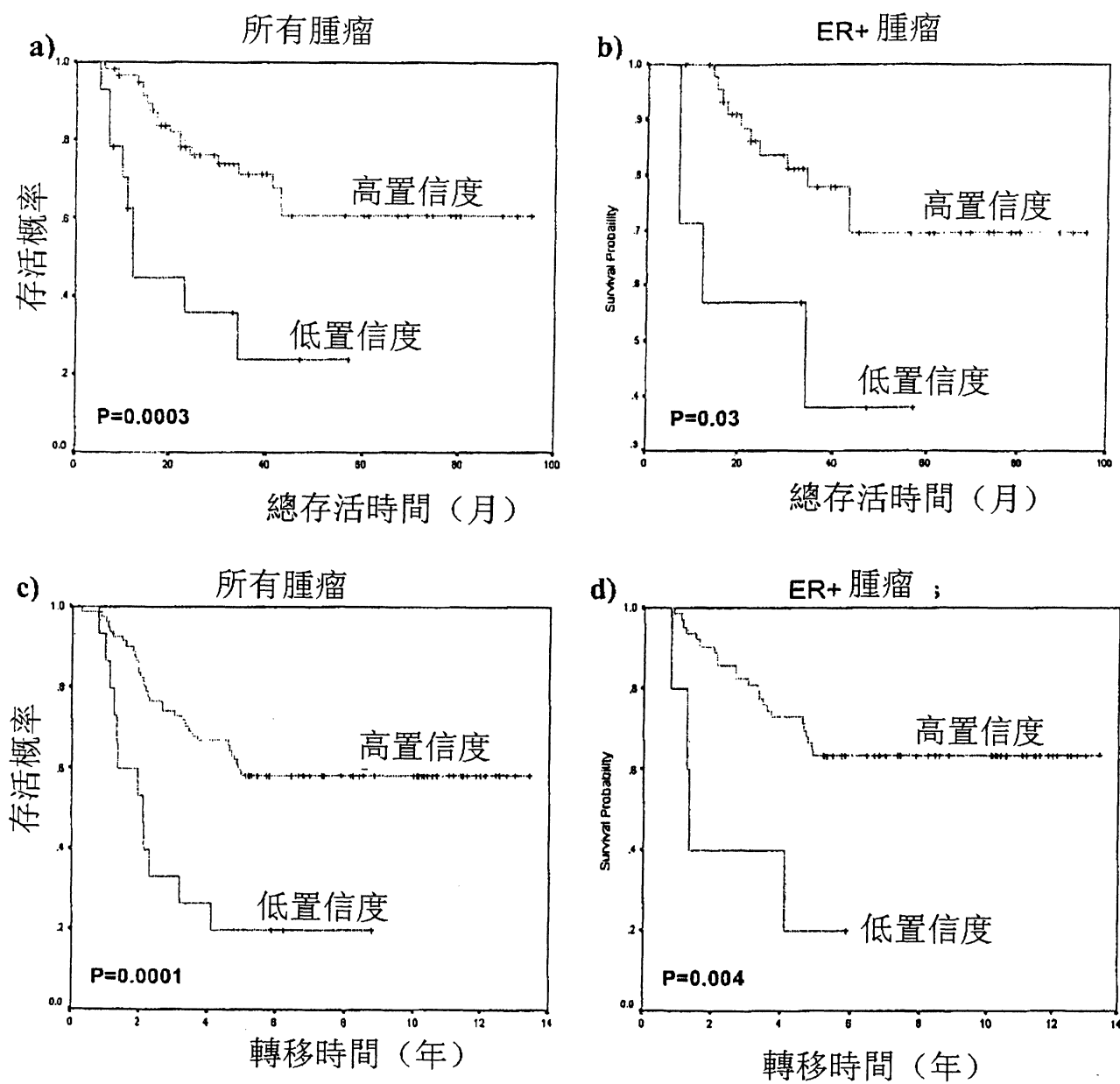


圖 2

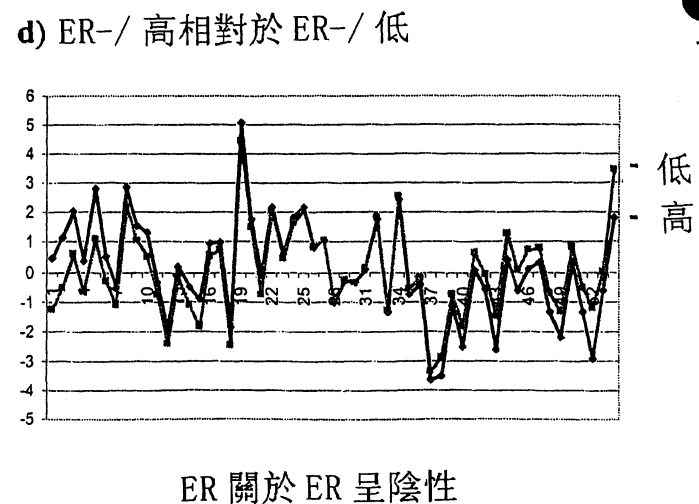
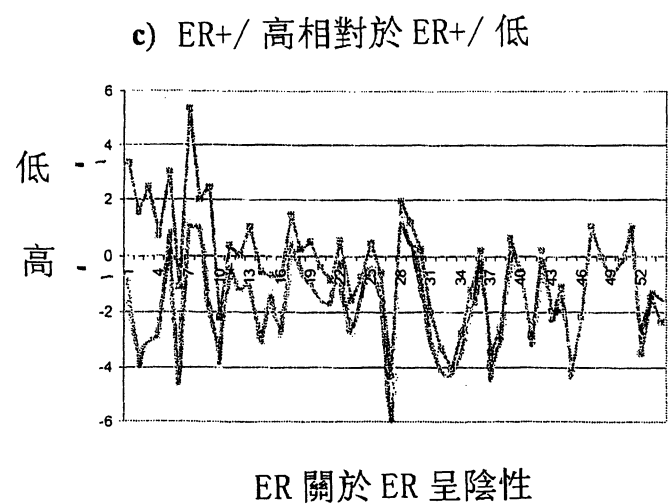
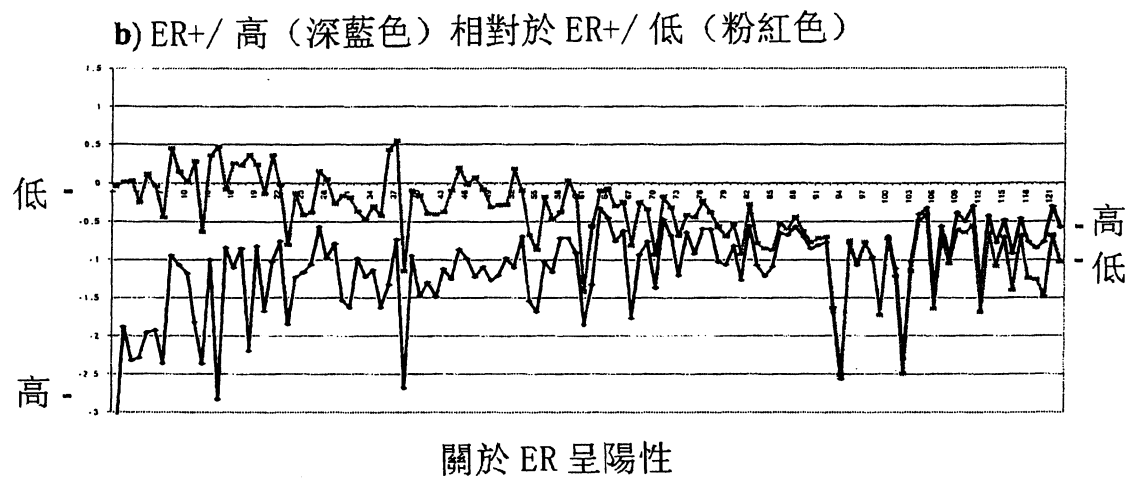
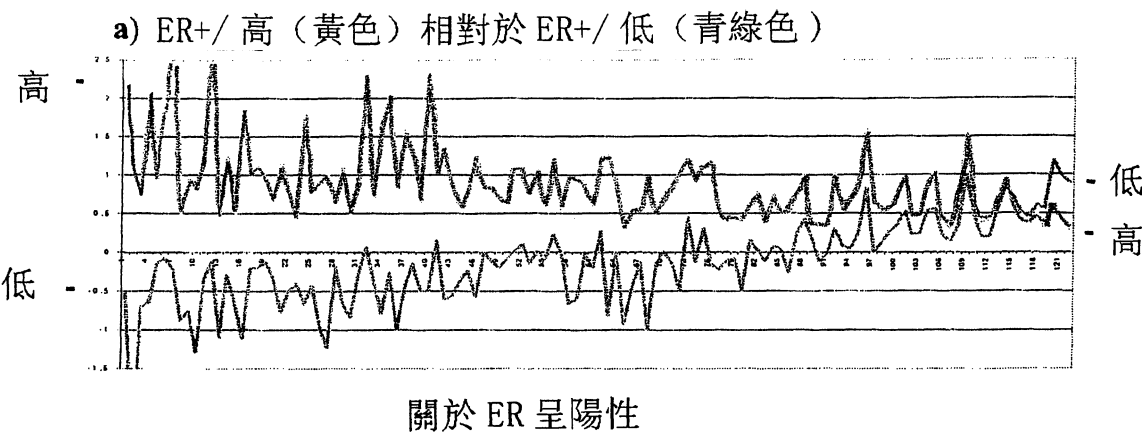


圖 3

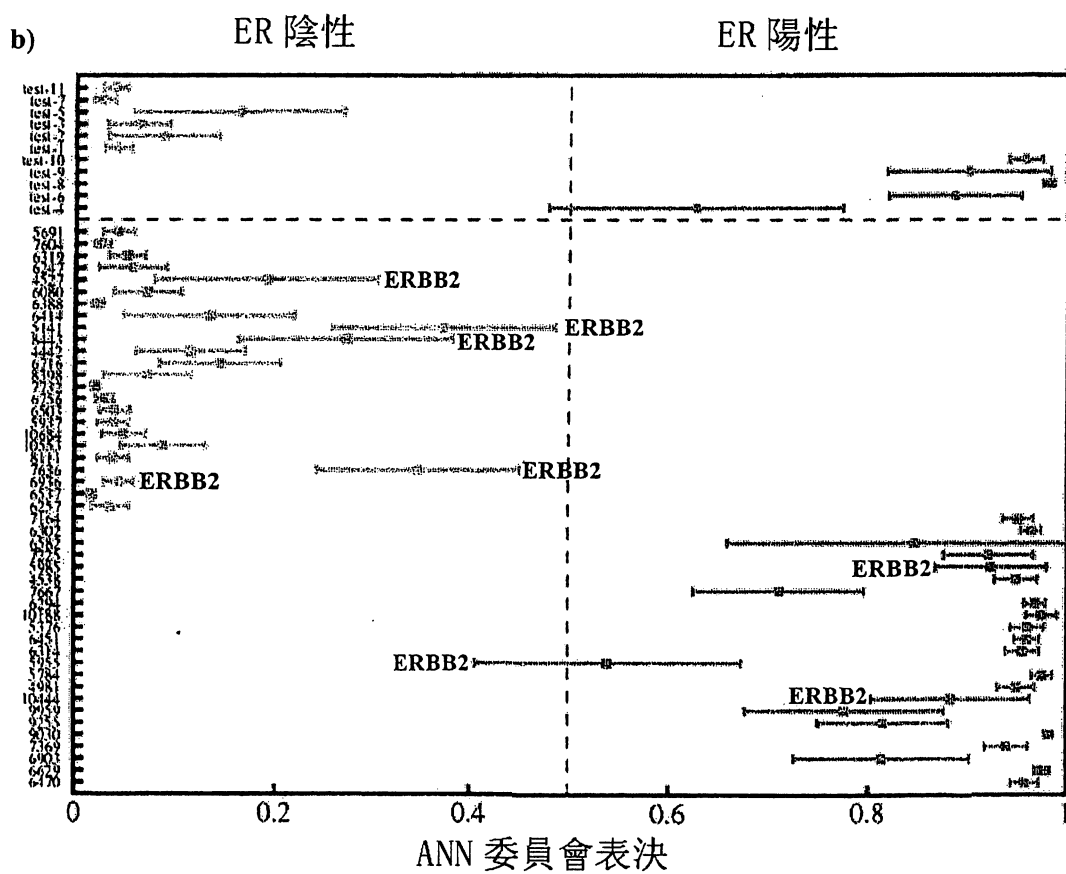
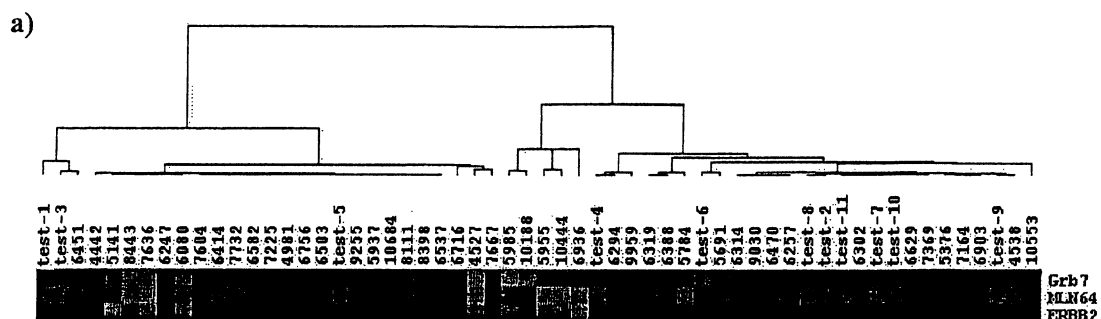


圖 4

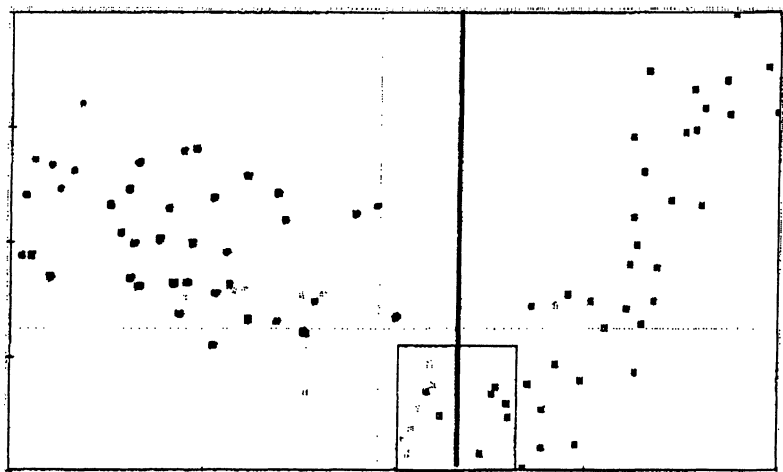


圖 5

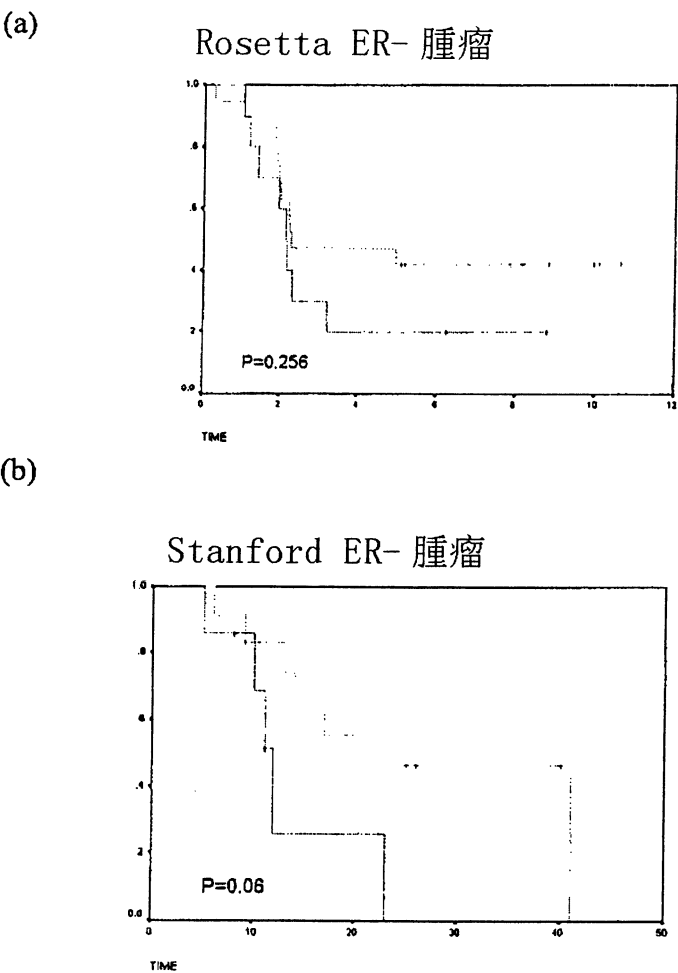
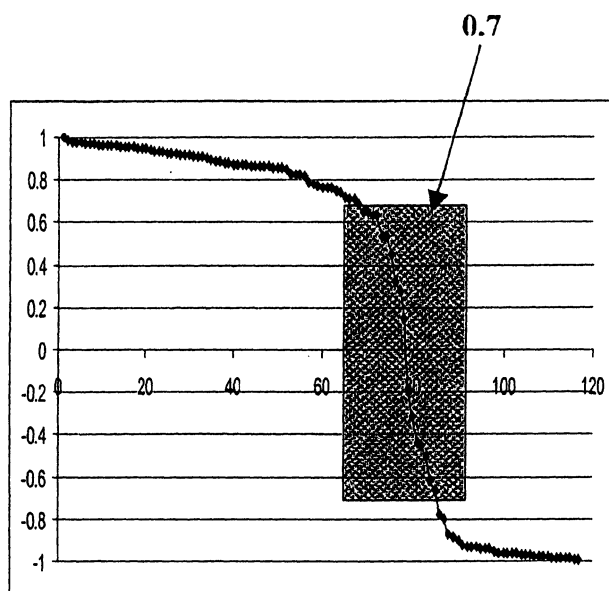


圖 6

(a) Rosetta 資料集



(b) Stanford 資料集

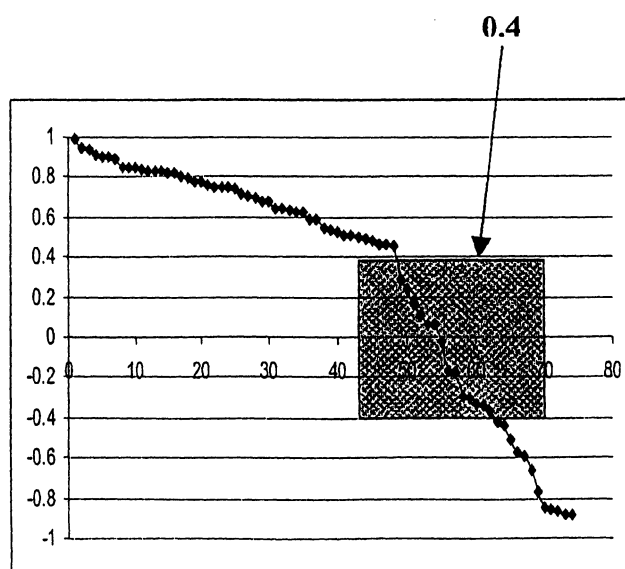


圖 7

七、指定代表圖：

(一)本案指定代表圖為：第 (6) 圖。

(二)本代表圖之元件符號簡單說明：

(無元件符號說明)

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

(無)