

發明專利說明書

(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：93130044

※ 申請日期：93.10.1

※IPC 分類：G01N³³/574
C12Q 1/68

一、發明名稱：(中文/英文)

有關乳癌分類之材料及方法

MATERIALS AND METHODS RELATING TO BREAST CANCER
CLASSIFICATION

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；0323225.3

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分類之材料及方法。特定言之，本發明係關於乳癌之預後的測定。

【先前技術】

已對使用基因表現資料以供生物分類(特別是在腫瘤學及醫學領域)產生了強烈興趣。此研究之一令人激動的態樣已在定義癌症之臨床相關亞型方面(其先前為更傳統之光顯微鏡研究所不能理解)表現出了能力。儘管具有此潛力，在可實現使用基因表現資料以供臨床診斷之前仍不得不解決許多問題。例如，需要實施演算法，其除提供正確分類以外，亦可精確測定預測之置信度。若該分類會影響隨後治療過程則此特別重要—若供給此資訊，則然後治療醫師可權衡預測置信度與特定干預之潛在發病率以作出精明的臨床選擇。

諾丁漢預後指數(NPI)為一基於腫瘤尺寸、組織學等級及淋巴結狀態之分類系統，其在歐洲及英國廣泛用於將預後指派給乳房腫瘤(1-5)。儘管其具有效用，吾人承認習知組織病理學參數(諸如腫瘤等級)及細胞形態學之使用亦與特定侷限性相關聯。即使在標準化嘗試後(6)，許多此等變數(舉例而言，等級)仍經受顯著之觀察者間的變化性。NPI標度可達到值2與8之間。當正在量測之參數係於值(7)之連續範圍內記分時(諸如NPI)，經常難以定義合適截止點。

因此，該指數視一系列主觀標準而定，此在所指派之預

後方面可導致觀察者之間產生差異。

NPI為一值標度；一具有比另一病人更低之NPI值的病人通常具有比該另一病人之預後更佳的預後。通常使用諸如經一特別時間標度之存活機會及/或在一特別時間標度(雖然不必與存活之時間標度相同)內之遠距離轉移之機會的因子來定義預後。因此，通常而言，隨NPI值增加病人前景減少。

在測定病人治療之類型及程度時，測定病人預後為一重要因子。當一未來治療程式可與預後相關聯時，所指派之預後的準確度因此而變得關鍵。例如，van't Veer等人(10)已識別70個基因之"預後識別簽名"(PES)，其能預測乳房腫瘤之無病存活(DFS)狀態。

【發明內容】

本發明者研究一組乳房腫瘤之表現資料，但最初不能識別其表現與NPI相關聯之一組基因。發明者假定在亞型之間基因表現方面可存在顯著差異("亞型間差異")，其潛在地遮掩了亞型內變化之更微細樣式("亞型內差異")。已提議，乳癌內在基因表現變化之顯著比例可歸因於屬於區別性"分子亞型"(諸如ER+及ER-(其中ER為"雌激素受體")(8-9, 14))之不同腫瘤。

使用非監督聚類技術將資料集分離為個別分子子範疇(ER+、ER-、ERBB2+)。各個分子亞型作為獨立資料集來處理。各個亞型內之腫瘤經獨立分析以定義其表現水平與NPI相關聯的一組基因。

臨床醫師通常將NPI標度劃分為三個種類："良好"預後、"中等"預後及"不良"預後。定義種類邊界之值視臨床醫師而變化。一典型邊界組之實例為：良好預後 $NPI < 3.4$ ；中等預後 $3.4 \leq NPI \leq 5.4$ ；且不良預後 $NPI > 5.4$ 。熟習此項技術者將明白此等邊界可變化。

本發明者已識別出一組62個基因，其差別表現於具有不同預後之腫瘤中，舉例而言，差別表現於具有高NPI(且因此具有不良預後)之腫瘤與相比較之具有低NPI(且因此具有良好預後)之腫瘤。

雖然該基因組係於根據其NPI對樣品進行分類後而得以識別，但是亦已發現使用此等基因之表現水平來對腫瘤樣品進行分類與預後之其它量測相關聯(舉例而言，無病存活)。

因此，對可獲得該樣品之病人的預後及治療而言，腫瘤樣品內此等基因之表現水平具有顯著醫學含義。尤其可使用其作為該病人之預後的指標來分類腫瘤樣品。

NPI標度上自3.8變化至4.6之值可用作"良好"與"不良"預後之間的截止點，且使用各個截止值來識別該相同組之62個差別表現基因。

此指示出，雖然NPI覆蓋了自2至8之值的連續譜，但來自具有62個基因之組的基因表現水平能夠將腫瘤樣品分類為離散種類。因此，顯示出基於組織病理學參數之連續NPI值的樣品可在分子水平上分離為離散種類。

此外，使用(i)本發明之方法及(ii)臨床技術(通常為組織

病理學技術)對指派給乳房腫瘤病人之預後所作的比較揭示出：基於諸如DFS及Kaplan-Meier存活曲線之病人資料，本發明之方法可提供比組織病理學技術更精確之預後。

該等62個基因在表S6中得以識別。下列描述將使用術語"表現輪廓"。此意謂樣品內一組基因之表現水平。除非本文另有規定，否則該基因組將包括表S6中所識別之某些或全部62個基因。

本文所識別之62個基因與van't Veer等人(10)之PES中所識別的基因僅重疊一個基因(DC13或Hs. 6879)。該PES為231個Rosetta基因(10)之擴展基因組的前70個基因(該等基因在可顯示不同疾病自由存活率的群之間展現出了最顯著的表現差異)。對表S6之62個基因及該等231個Rosetta基因而言具有8個共同基因，該等8個基因在表S13中列出。

表S6中兩個基因高度表現在低NPI腫瘤方面("陰性基因")，而60個基因則高度表現在高NPI腫瘤方面("陽性基因")。

因此，最通常地，本發明提供了一種獲得一組差別表現基因之方法。本發明亦提供了用以分類及/或將預後指派給一乳房腫瘤樣品之方法及檢定。本發明可識別一組基因，且提供使用一乳房腫瘤樣品內某些或全部彼等基因的表現水平以將一預後指派給可獲得該樣品之病人。

在第一態樣中，本發明提供了一種測定乳癌病人之預後的方法，該方法包括基於一組基因(下文稱作"預後組")在該病人之乳房腫瘤內的表現水平將一預後指派給該病人，其中該預後組包括來自表S6之複數個基因。

本發明進一步提供使用該預後組以測定乳癌病人之預後。較佳地，本發明提供了使用一表現輪廓以測定乳房腫瘤病人之預後，該表現輪廓代表該預後組之基因在腫瘤內的表現水平。

"預後"意欲具有其最通常意義，且可為定量或定性。其可表現為通常術語，諸如"良好"或"不良"預後，及/或表現為可能臨床結果之術語，諸如無病存活(DFS)之持續時間、定義時間段之存活可能性及/或定義時間段內遠距離轉移之可能性。預後之定量量測通常將具有概率性。另外或其它，且特別對於向開業醫生或開業醫生之間傳達該預後而言，該預後可表現為預後之另一指標之術語，諸如NPI標度。

通常，將可能以習知治療攝生法來治療具有"良好預後"腫瘤之病人。可能以替代或更積極攝生法來治療具有"不良預後"腫瘤之病人。該"不良預後"病人通常將不必等待習知治療攝生法失敗而進入更積極的攝生法。此外，對病症之可能臨床過程瞭解可使病人得以為將來準備現實計劃，此是癌症治療之重要社會態樣。

為避免疑問，術語"測定"不必暗含預後中之絕對確定性。相反，腫瘤內預後組之表現水平通常將指示該病人之可能預後。

該等表現水平通常將以數字代表。因此，表現輪廓通常將包括一組數字，各個數字代表該預後組之一個基因之表現水平。

一種根據本發明之第一態樣之方法可包括步驟：

提供一表現輪廓，其代表該預後組之基因在腫瘤內之表現水平，且

基於該表現輪廓將一預後指派給該病人。

該提供步驟可包括自一預先存在之資料集來提取關於該預後組之基因之表現水平的資訊，其亦可包括其它表現水平(舉例而言，代表該腫瘤內其它基因之表現水平的資料)。或者，其可包括以實驗方式測定該等表現水平。

該測定步驟可包括步驟：

(a)自病人獲得一乳房腫瘤樣品；

(b)量測該預後組之基因在該樣品內的表現水平。

基因表現水平之量測(且尤其是以表現輪廓作為其代表)可為絕對術語或相對於某些其它因子，諸如(但不限於)該樣品內或一樣品群間的另一基因之表現或一基因群(較佳為該預後組之外的基因，但可能包括該預後組之基因)之表現水平之平均值、中值或模式。例如，基因表現可作為樣品內複數個基因之平均表現的倍數或分數來量測或代表。較佳地，該表現以表現輪廓代表為正或負以指示相對於平均值的表現之增加或減少。

在一非較佳實施例中，以一組數值之形式的表現輪廓資訊被轉換為該預後組之基因之排列表，其中該等基因以表現水平之順序排列，隨後將該等個別基因之排列順序用作分析參數(而非該基因之表現水平)。

較佳地，步驟(b)包括以能夠結合至可指示該預後組基因表現之表現產物的複數個結合組員來接觸獲自該樣品之該

等表現產物，其中可量測此結合。

通常，該等結合組員能夠不僅偵測一表現產物之存在而且能夠偵測其相對豐度(意即，可得到產物之量)。該表現輪廓可使用能夠結合至該預後組之表現產物的結合組員而得以測定，舉例而言 mRNA，其對應於 cDNA 或 cRNA 或表現之多肽。藉由標記該表現產物或該結合組員，可識別該等表現產物之相對量或比例且測定該預後組之表現輪廓。該等結合組員可為互補核酸序列或特定抗體。

指派一預後之步驟可藉由以下步驟進行：比較測試之表現輪廓與其它先前獲得之與已知預後相關聯之輪廓及/或一個(或多個)先前測定之表示一個(或多個)特別預後之特徵的"標準"輪廓。一特別預後之標準輪廓可由來自彼預後之複數個腫瘤的表現輪廓產生。

該比較通常將藉由或借助於電腦執行。

較佳地，將該表現輪廓與具有不同已知預後之已知或標準輪廓(較佳為標準輪廓)進行比較。待指派給該病人之預後為已知或標準輪廓之預後，其中測試之表現輪廓最接近地類似。

較佳地，該比較係藉由被分類為兩個不同預後(舉例而言，"良好"及"不良"、或高及低 NPI(較佳地截止點介於 3.8 與 4.6 之間))之已知或標準輪廓(較佳為標準輪廓)來進行。該等已知或標準輪廓將已由已知預後之樣品產生，其可以任何便利方式測定一藉由除去該樣品後該病人之實際臨床結果或藉由其它預後技術，舉例而言組織病理學技術，舉

例而言使用NPI標度。

該比較可涉及可歸因於該預後之置信度水平的評定，其基於統計技術。該等標準輪廓對可獲得其之特別材料及方法(舉例而言，微陣列)而言通常係特定的。若採用一新穎材料及/或方法(舉例而言，一新型微陣列)，則已知預後之標準輪廓較佳再次使用該預後組而得以獲得。

根據本發明之第一態樣之方法可包括將乳房腫瘤之樣品分類為(例如)具有高NPI或低NPI，或分類為具有良好或不良預後。

如先前所提及，指派一預後之步驟可藉由以下步驟進行：比較來自測試之乳房腫瘤樣品的表現輪廓與先前獲得之輪廓及/或一先前測定之"標準"輪廓，其表示一特別預後之特徵，例如一"良好"及/或"不良"預後及/或至少一個NPI值及/或NPI值之至少一個範圍。該等先前獲得之輪廓可作為輪廓之資料庫而儲存。

較佳地，該資料庫包括表示一特別預後之特徵的基因表現輪廓。該等基因表現輪廓較佳由與本發明之第一態樣之預後組相同的預後組(表S6之基因之亞組)或一充分重疊該第一態樣之預後組以便為該等表現水平之比較提供統計方面顯著之基礎的預後組(潛在地，為一與以上不同之亞組)之表現水平產生。電腦可經程式化以報告測試之輪廓與標準輪廓之間的統計相似性，使得可指派一預後。

有利地，使用一基因表現輪廓以指派一預後可減少或甚至可消除用以將一預後指派給一腫瘤樣品的臨床程序之主

觀本性。由於該方法需要較佳定量地在分子水平來評定表現產物，所以該方法提供一更客觀且因此潛在地更可信之用以指派一預後之方式。如早些所提及，該預後組能夠將乳癌腫瘤樣品分離為離散種類且因此減少甚或消除臨床預後指派之主觀分析。此外，可將置信度指派給該預後，使得可做出關於病人治療的精明的選擇，此視該預後之"力"而定。

在相似預後之獨立樣品之間，該預後組之表現輪廓可稍稍不同。然而，本發明者已意識到，當組合使用時能組成該預後組之特別基因的表現輪廓提供了一腫瘤樣品內之表現樣式(表現輪廓)，該樣式表示腫瘤之預後的特徵。

本發明者已發現，該預後組能夠將腫瘤樣品分解為高NPI及低NPI類。高NPI意謂較佳至少3.4、較佳至少3.5、更佳至少3.6、更佳至少3.7、更佳至少3.8、更佳至少3.9及最佳至少4.0之NPI。高NPI可為至少4.1、至少4.2、至少4.3、至少4.4、至少4.5或至少4.6。高NPI與低NPI之間的較佳截止值為3.8-4.6之間。

在歷史上，NPI之"良好"、"中等"及"不良/差"種類可使用大的臨床研究而得以測定，其中屬於此等不同群之病人在整個存活過程中顯示出統計方面顯著之差異。例如，良好預後之病人可具有約83%之十年存活率，"中等"預後之病人可具有約52%之十年存活率，且"差"或"不良"預後之病人可具有約13%之十年存活率(4)。

詳言之，該預後組似乎最強烈地與雌激素受體陽性腫瘤

(ER+)內之腫瘤預後(如NPI所反映)相關聯。

在乳癌治療中，將乳房腫瘤分類為雌激素受體陽性(ER+)及陰性(ER-)亞型為一重要區別。ER-腫瘤通常比其ER+對應物臨床上更積極，且ER+腫瘤係使用諸如他莫西芬(tamoxifen)(21)之抗激素療法而按常規治療。可使用組織學技術(舉例而言，使用特定用於受體之抗體)或使用基因表現技術將乳房腫瘤分類為ER+或ER-。目前，腫瘤之ER狀態係藉由免疫組織化學(IHC)或使用ER之一抗體進行免疫印跡而按常規測定。

本發明之第一態樣較佳包括測定腫瘤樣品之ER狀態之步驟。可使用基因表現分析或藉由使用組織病理學技術來測定該ER狀態。較佳地，本發明之第一態樣進一步包括作為一初始步驟之測定腫瘤樣品之ER狀態，且僅當該狀態為ER+時才繼續進行。

較佳地，該腫瘤樣品之ER狀態係使用基因表現輪廓而得以測定，如吾人之同在申請中之申請案PCT/GB03/000755中所描述。基因表現輪廓能夠以高置信度將乳房腫瘤分類為ER+或ER-。然而，亦存在第三乳癌種類，其不能以顯著統計確定性("低置信度"腫瘤)被分類為ER+或ER-。ERBB2+之向上調節常常與低置信度腫瘤相關聯。較佳地，僅有以高置信度識別之ER+腫瘤(較佳被分類為ER+，其具有如使用PCT/GB03/000755之方法而測定之大於0.4的量值預測力)係使用根據本發明之第一態樣之方法而進行分析。

將一預後指派給乳房腫瘤樣品之步驟可包括使用統計及

/或概率性技術，諸如加權表決(WV)(13)、一有監督的學習技術。在WV中，可執行二進位分類。意即，可使用該技術以將一樣品指派給兩類中之一類。將在該乳房腫瘤樣品之預後組中各個基因之表現水平與不同類間彼基因之表現水平之平均值進行比較。該平均值可(例如)由具有一指派預後之表現輪廓(舉例而言，"已知"預後之表現輪廓的資料庫)來計算。

該表現水平與該等類間平均基因表現之差異被加權，且其對應於一特別類之彼基因的一"投票"及與另一類相抵之彼基因的一相等但負的投票。對一特別腫瘤而言，為各個類而將全部基因之投票(正及負)合計在一起，從而為各個類得出總計。將該腫瘤指派給具有最高(正)總計之類。然後可將該取勝類之取勝餘量表達為預測力。

表現水平之差異係使用一公式而得以加權，該公式包括該等兩類中之各類的基因表現水平之平均及標準偏差。通常，各個類之平均及標準偏差由具有或代表一特別預後(舉例而言，高NPI及低NPI)之表現輪廓來計算。

另外或其它，指派一預後之步驟可包括使用分級群集，特別是若已使用不同材料及/或由用以測定具有"已知"預後之表現輪廓或已與該樣品表現輪廓進行比較之標準輪廓來測定表現輪廓的方法來測定該腫瘤樣品內表現水平。

可使用既定的留一交叉確認(LOOCV)檢定來確認經指派之預後(參見實例)。可使用電腦執行步驟(c)。

在分級群集中，各個表現輪廓可代表為一個由n個基因組

成之向量，其中(g_1 、 g_2 、 $\dots$ 、 g_n)代表該等基因之表現水平。然後將各個向量與該分析中各個其它輪廓之向量比較，且將彼此具有最高相關性之兩個向量配對，直至已經將該分析中盡可能多的輪廓配對。

存在此項技術中已知之許多方式來計算該相關性，諸如Pearson之相關係數(22)。在下一步驟中，然後自各個配對獲得一合成向量(在平均連接群集中此通常為兩個輪廓之平均)，然後重複該配對過程。此繼續進行直至已將全部向量配對在一起，從而組裝一可代表全部輪廓之"樹"。此過程為"分級"過程，因為吾人可自底部(個別輪廓)起始並構造。在本發明中，個別輪廓構造至較佳兩個合成向量，其中各個向量代表一類(意即，良好或不良預後)。對一未知類之新穎樣品而言，將該樣品與標準輪廓/樣品群集。"未知"樣品之類別將基於在配對迭代循環之末端處其所屬之群集/向量而得以測定。

具有一個或多個"已知"或所指派之預後的表現輪廓意謂已指派有或已獲得一預後之表現輪廓。該預後可已經：自基因表現資料計算出來；自執行於源樣品之臨床技術(舉例而言，組織病理學技術)獲得；或基於可獲得該表現輪廓之病人實際病症進展/結果而得以追溯性地指派。該第三選擇最佳，因為可基於該病人之隨後結果、由該病人之醫療記錄來指派一精確預後(對於獲得該樣品之時間點而言)。在此追溯性指派中，事後認識之使用提供了準確度。

本發明之方法可用以評定治療乳癌病人之功效。可在治

療前及/或治療之早期階段指派該病人之預後並將該預後與在治療後(或在治療之後階段)指派給該病人之預後進行比較。治療前及/或後之預後較佳係使用根據本發明之方法來指派。若該治療包括若干階段，則該表現輪廓可在各個階段後得以測定，以繪製該治療之進展。治療後之一經改良預後指示了一成功或至少部分成功之治療。該治療可為化學療法。

本發明之方法可包括在治療前及治療後比較乳房腫瘤樣品內預後組之表現水平，以偵測能指示一改良預後或惡化預後的表現輪廓之改變。

該方法可包括偵測在表S6中指示為"向上調節"之預後組內基因的向下調節及/或在表S6中指示為"向下調節"之預後組內基因的向上調節。比較標準值(舉例而言，具有不同預後之樣品範圍的平均表現水平)及/或比較先前值(例如，能指示"不良"預後或表示此之特徵的標準輪廓)，該等基因可呈向下調節/向上調節。"向上調節"基因的向下調節及/或"向下調節"基因的向上調節指示出一良好或中等預後。調節(in regulation)改變之程度可指示出治療之功效。

本發明者已發現表現輪廓中朝向良好預後腫瘤的改變指示出成功治療。顯示出表現輪廓之此改變的腫瘤具有最佳預後(舉例而言，最佳存活率、最佳無病存活率)。可將前治療階段及後治療階段之腫瘤表現輪廓與具有已知預後之標準輪廓進行比較。

因此，該方法可包括將乳房腫瘤之表現輪廓指派給良好

或不良預後類(或者高NPI類或低NPI類)，及將第二表現輪廓(在較晚治療階段自該腫瘤而測定)指派給良好或不良預後類(或者高NPI類或低NPI類)，及偵測類改變，其中自不良預後至良好預後(或高NPI至低NPI)之改變指示出一有效治療。另外或其它，指派良好或不良預後類(或者高NPI類或低NPI類)之統計置信度水平的改變可指示出治療之功效。指派一可指示不良預後之類的置信度的減少可暗示成功或至少部分成功之治療。

評定治療功效之方法可包括測定腫瘤之ER狀態之步驟。然而，該等評定功效之方法對評定ER+、ER-及ERBB2+腫瘤之治療功效頗為有效，意即不管該腫瘤之ER狀態如何。

表現輪廓代表腫瘤內一群基因之表現水平。各個表現輪廓之基因不需要相同但在各個表現輪廓之基因之間應具有足夠之重疊以容許對該等表現輪廓進行比較及群聚。

可使用此項技術中已知之標準程序為偵測目的而對結合組員進行標記。或者，可在隔離測試樣品後對表現產物進行標記。一較佳偵測方法為使用一螢光標記，其可藉由一測光表來偵測。替代偵測方法包括電訊號傳輸。例如，Motorola(Pasadena, California)e-sensor(電子感測器)系統具有兩個探針，一個能自由浮動之"俘獲探針"及一個連接至一可加倍以作為電極表面之固體表面的"訊號傳輸探針"。兩個探針均充當表現產物之結合組員。當結合發生時，將兩個探針彼此緊密接近從而導致產生可偵測到之電訊號。

然而，存在大量近來出現之較新技術，其利用"無標記"技術以供定量，例如彼等藉由 Xagros(Mountain View, California)而產生之技術。引子及/或經擴增之核酸可缺乏任何標記。可藉由量測由於兩個引子對接至一目標表現產物及藉由聚合酶之隨後擴展而產生之電阻改變來評定定量。

如以上所討論，該等結合組員可為寡核苷酸引子以用於PCR(舉例而言，多重PCR)從而特定地擴增遺傳識別符(genetic identifier)之表現產物之數目。然後將在凝膠上分析該等產物。然而，較佳地，該結合組員為一固定至載體之單一核酸探針或抗體。然後將該等表現產物通過該載體，藉此使其接觸該結合組員。該載體可為玻璃表面，舉例而言顯微鏡載物片；串珠(Lynx)；或光導纖維。在串珠之情況中，可將各個結合組員固定至一個別串珠且然後將其接觸溶液中的表現產物。

在此項技術中存在用以測定特別基因組之表現輪廓的各種方法，且此等方法可應用於本發明。例如，基於串珠之研究(Lynx)或分子條形碼(Surromed)為已知技術。在此等情況中，各個結合組員被附著至可個別讀取且可自由浮動之串珠或"條形碼"以放鬆與表現產物之接觸。該等結合組員至該等表現產物(目標)之結合係於溶液中達成，隨後使該等已標記串珠或條形碼穿過一設備(舉例而言，一流動血細胞計數器)且進行讀取。

一種測定表現輪廓之進一步已知方法為由Illumina(San

Diego, California)開發之測試裝備，即光導纖維。在此情況中，各個結合組員被附著至光導纖維電纜之末端處之一特定"地址"。該表現產物至該結合組員之結合可引發由該光導纖維電纜之另一端處的設備來讀取之螢光改變。

本發明者已經成功使用一核酸微陣列，其包括固定至一載體之複數個核酸序列。藉由將代表表現基因之核酸序列(舉例而言，cDNA)通過該微陣列，其能夠自一具有特別預後之腫瘤樣品(尤其是一具有良好預後之腫瘤樣品或一具有不良預後之腫瘤樣品或一具有高NPI之腫瘤樣品或一具有低NPI之腫瘤樣品)產生表現產物之結合輪廓特徵。

在第二態樣中，本發明提供裝置(較佳為一微陣列)以用於將一預後指派給乳房腫瘤樣品，該裝置包括一附著有複數個結合組員之載體，各個結合組員能夠特定結合至該預後組之基因之表現產物。較佳地，附著至該載體之該等結合組員能夠特定及獨立地結合至具有至少5個基因、更佳至少10個基因或至少15個基因及最佳至少20或30個基因(識別於表S6中)之表現產物。附著至該載體之該等結合組員能夠特定地結合至具有20至30個基因(識別於表S6中)之表現產物。

在一實施例中，能夠特定及獨立地結合至表S6中所識別之全部基因的表現產物之結合組員被附著至該載體。該載體可僅已經附著至能夠特定及獨立地結合至表S6所識別之基因的表現產物或其中一預後組之表現產物的結合組員。

該裝置較佳包括能夠特定結合至來自該預後組之表現產

物或其複數個基因的結合組員，且其可包括能夠特定結合至U133A微陣列上所代表之基因的不完全亞組之表現產物的結合組員(雖然其亦可包括U133A微陣列上所未代表之其它基因的結合組員)。據信，U133A微陣列代表約14397個區別性基因。因此，該裝置較佳包括U133A微陣列上之不多於14396個基因的結合組員。該裝置可包括能夠特定結合至U133A微陣列上之不多於90%的基因之表現產物的結合組員。該裝置可包括能夠特定結合至U133A微陣列上之不多於80%或70%或50%或40%或30%或20%或10%或5%的基因之表現產物的結合組員。

另外或其它，該載體可容納不多於14000、或不多於10000、或不多於5000、或不多於3000、或不多於1000、或不多於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個不同基因之結合組員。

較佳地，該等結合組員為核酸序列，且該裝置為一核酸微陣列。

列出了表S6之基因，其中其單基因登記號(accession number)對應於單基因資料庫之建置160。因此可自國家衛生研究院(NIH)之單基因資料庫(<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>)來擷取各個基因之序列。

此外，對於全部該等基因而言，Affymetrix(Santa Clara，

California)(www.affymetrix.com)提供探針組之實例，包括當用於載體上時能夠偵測該基因之表現的該等探針(意即，呈寡核苷酸序列之形式的結合組員)之序列。使用目標基因之單基因ID，該等探針詳情可自Affymetrix網站之U133A部分存取。

若將來該表所列出之單基因ID中的一個將被合併為一新ID、或分成兩個或兩個以上ID(舉例而言，在資料庫之新建置中)、或完全刪除，則如本發明者所希望，該基因之序列係可藉由存取單基因之建置160而得以擷取。

通常，高密度核酸序列(通常為cDNA或寡核苷酸)被固定至載體之很小的離散區域或點上。該載體經常為一蓋玻片面或一薄膜過濾器，其塗佈有一基材(意即一"晶片")。該等核酸序列通常藉由一機器人系統而被傳遞(或印刷)至該經塗佈之載體上，且然後使其不動或固定至該載體。

在一較佳實施例中，獲自該樣品之表現產物通常使用一螢光標記物而得以標記出來，且然後使其與該固定不動之核酸序列接觸。雜交之後，使用偵測器(諸如高解析度雷射掃描儀)偵測該等螢光標示物。在一替代方法中，可以非螢光標記物(舉例而言，生物素)來標記該等表現產物。雜交之後，然後以能結合/鍵結至第一非螢光標記物(舉例而言，結合至生物素的經螢光標記之鏈酶親和素)之螢光染料對該微陣列進行"染色"。然而，如以上所討論，該等表現產物可為無標記。

藉由使用數位成像軟體分析自各個離散點發射之訊號來

獲得一能指示基因表現之一樣式(表現樣式或輪廓)的結合輪廓。然後可將實驗樣品之基因表現樣式與標準輪廓(意即，來自一具有(例如)已知良好或不良預後或已知NPI值或NPI值之已知範圍的組織樣品的表現輪廓)之樣式進行比較以供差別分析。

該標準可獲自一個或多個表現輪廓，其先前經判斷是以一特別預後(舉例而言"差"或"良好"預後)為特徵及/或表示一特別NPI範圍(諸如高及/或低NPI)之特徵及/或表示一個或多個NPI值或值之一或多個範圍之特徵。該標準可獲自一個或多個表現輪廓，其先前經判斷表示一特別NPI值或值之範圍(或預後標度上之其它定義值)之特徵。該標準可包括表示一正常樣品之特徵的表現輪廓。此等/此標準表現輪廓可作為資料庫之一部分而被可擷取地儲存於資料載體上。

大多數微陣列利用一個或兩個螢光團。對於兩色陣列而言，最通常使用之螢光團為Cy3(綠色通道激發)及Cy5(紅色通道激發)。微陣列影像分析之目的係自各個表現產物提取雜交訊號。對於單色陣列而言，訊號係作為給定目標之絕對強度來量測(基本上對於雜交為單一樣品之陣列而言)。對於兩色陣列而言，訊號係作為具有不同螢光標記物的兩種表現產物(舉例而言，樣品及對照(對照可另外通稱為參考))之比率來量測。

根據本發明之裝置較佳包括複數個離散點，各個點包含一個或多個寡核苷酸且各個點代表一選自表S6之基因之表現產物的不同結合組員。在一實施例中，該微陣列將包含

表S6所提供之各個基因的點。各個點將包括複數個相同寡核苷酸，其均能夠結合至其所代表之表S6之基因的表現產物(舉例而言，mRNA或cDNA)。各個基因較佳由複數個不同寡核苷酸代表，對於該基因而言較佳為Affymetrix U133A探針組。

在本發明之第三態樣中，提供了一套組以將一預後指派給乳癌病人，該套組包括能夠特定結合至該預後組之基因之表現產物的複數個結合組員及一偵測試劑。該套組可包括一資料分析工具，較佳呈電腦程式之形式。該資料分析工具較佳包括一經調適以辨別腫瘤之表現輪廓與差異預後的演算法。較佳地，該演算法經調適以辨別"良好"預後與"差"預後，最佳辨別高NPI與低NPI腫瘤。該演算法較佳為如以上所描述之加權表決演算法。

在一實施例中，該套組包括本發明之第二態樣之裝置。

該套組可包括具有已知預後之乳房腫瘤樣品之表現輪廓(如以上所討論)，及/或表示一特別預後之特徵的基因表現輪廓(如以上所討論)，其較佳儲存於一資料載體或其它記憶體設備上。該等輪廓可已經得以分析或統計成群，例如經計算之平均表現水平及/或基因加權。

較佳地，該套組中該一個或該等多個結合組員(抗體結合域或核酸序列，舉例而言寡核苷酸)被固定至一個或多個載體，舉例而言，用於微陣列或光導纖維檢定之單一載體或諸如串珠之多個載體。偵測構件較佳為一用於標記測試樣品之表現產物的標記物(放射性或染料，舉例而言螢光)。該

套組亦可包括用於測試及分析測試表現產物之結合輪廓的試劑。

或者，該等結合組員可為能夠結合至表S6中所識別之基因之表現產物的核苷酸引子，使得可在PCR中將其擴增。該等引子可進一步包括偵測構件，意即可用以識別經擴增之序列及其相對於其它經擴增之序列之豐度的標記物。

該乳房腫瘤樣品可作為切除乳房活組織檢查或細針抽吸物而獲得。

藉由自大量腫瘤樣品產生該預後組之大量表現輪廓(其均具有一已指派之預後(較佳基於預後標度))，可生成良好及不良預後之輪廓庫。表現輪廓之數目越大，越容易形成可在預後檢定中用作標準之可靠特徵性表現輪廓標準(意即，包括統計變化)。因此，標準輪廓可為自複數個個別表現輪廓設計出及在統計變化內設計出以代表(例如)"良好"或"差"預後或高NPI或低NPI的輪廓。

在第四態樣中，提供了一種產生乳房腫瘤樣品之核酸表現輪廓的方法，其包括步驟

- (a)自該乳房腫瘤樣品隔離表現產物；
- (b)識別該基因預後組之表現水平；且
- (c)自該等表現水平產生該乳房腫瘤樣品之表現輪廓。

可將該表現輪廓添加至基因表現輪廓資料庫。該方法進一步包括將該表現輪廓與第二表現輪廓(或複數個第二表現輪廓)進行比較之步驟。可使用大體上相同之預後組自一第二乳房腫瘤樣品(或多個樣品)產生該(等)第二表現輪

廓，其中一預後已經被指派給該(等)第二樣品或已為其測定。該(等)第二表現輪廓可為一個(或多個)表現一特別預後(例如"良好"預後或"差"預後、或高NPI或低NPI、或至少一個特別NPI值或NPI值之至少一個範圍)之特徵的標準輪廓。

較佳地，該預後呈預後量測之形式，較佳為臨床可接受之預後分類系統，諸如NPI。再次，可自基因表現資料預測該預後，該基因表現資料獲自臨床技術(諸如組織病理學技術)，或是基於捐獻可獲得該第二輪廓之樣品的病人之疾病結果而追溯性地指派給該第二表現輪廓。

已知該預後組後，可設計出許多方法來測定一特別測試樣品內基因之表現樣式或輪廓。例如，可使用標準分子生物學技術自該樣品隔離表現核酸(RNA、mRNA)。然後可使用特定用於PCR中之表現序列的核酸引子來擴增對應於表S6中所給出的遺傳識別符之基因組員的表現核酸序列。若該經隔離之表現核酸為mRNA，則可使用標準方法將此轉換為cDNA以用於PCR反應。

該等引子可便利地將一標記物引入該經擴增之核酸，使得可識別其。理論上，該標記物能夠指示擴增事件後存在的核酸序列之相對量或比例，此反映原始測試樣品中存在的相對量或比例。例如，若該標記物為螢光或放射性標記物，則訊號強度將指示表現序列的相對量/比例甚或絕對量。各個遺傳識別符之表現產物之相對量或比例將為該測試樣品產生一特別表現輪廓。

根據本發明之第四態樣的方法可包括步驟：

(a)自第一乳房腫瘤樣品隔離表現產物；以能夠特定及獨立地結合至該預後組之表現產物的複數個結合組員來接觸該等表現產物；且自該腫瘤樣品內該預後組之表現水平產生第一表現輪廓；

(b)自已知預後(如先前所定義)之第二乳房腫瘤樣品隔離表現產物；以能夠特定及獨立地結合至步驟(a)之預後組之表現產物的複數個結合組員來接觸該等表現產物，以便產生一乳房腫瘤樣品之可比較第二表現輪廓；

(c)比較第一及第二表現輪廓以測定該第一乳房腫瘤樣品之預後。

在本發明之第五態樣中，提供了一表現輪廓資料庫，其包括乳房腫瘤樣品之複數個基因表現輪廓，其中該等基因表現輪廓獲自該基因預後組之表現水平，該資料庫可擷取地保存於一資料載體上。該資料庫較佳藉由根據本發明之第四態樣之方法產生。

該等表現輪廓較佳為核酸表現輪廓。該核酸表現輪廓之測定可計算機化，且可在特定先前設定之參數內進行以避免假陽性或假陰性。

該資料庫可包括表示一特別預後(諸如良好或不良預後)之特徵或表示一特別預後值(較佳NPI值(舉例而言高NPI、低NPI或特定定性值或值範圍))之特徵的表現輪廓。可根據源腫瘤之ER狀態(意即ER+或ER-)對該等表現輪廓進行分類。然後可處理及分析該資料庫，使得其最終將包含(i)對應於該資料庫中各個表現輪廓之數字資料；(ii)充當特別預

後指派(舉例而言，良好或不良預後、或值或值範圍，較佳來自NPI)之典範輪廓的"標準"輪廓；及(iii)代表個別輪廓至該"標準"輪廓之所觀察的統計變化的資料。

然後電腦可能夠提供一具有特別預後(舉例而言，良好預後及/或不良預後及/或高NPI及/或低NPI)之乳房腫瘤樣品之表現輪廓標準特徵。如較早所陳述，然後可使用已測定之表現輪廓以將一預後指派給該乳房組織樣品，此較佳使用區別演算法，最佳為加權表決演算法，如以上所描述。

所測試之基因表現水平之數目越大，表現輪廓之分類越可靠。已知微陣列及基因晶片技術容許利用大量結合組員。因此，更佳方法將為使用可代表表S6中之全部基因的結合組員。然而，熟習者將理解，可省略此等基因之比例，且仍以可靠及統計精確之方式進行該方法。

本發明之任何態樣中的預後組可包括或由來自表S6之全部或大體上全部基因，或全部或大體上全部之該等陽性基因及/或全部該等陰性基因組成。在本發明之態樣間，該基因預後組可獨立地在含量及數目上變化。

該預後組可包括表S6之至少5個、10個、20個、30個、40個、50個、60個或全部基因。

較佳地，該預後組包括或由來自表S6之約60個或約50個或約40個或約30個或約20個或約10個或約5個基因組成。來自表S6之陽性基因較佳選自表S6中陽性基因之清單的上面部分、較佳上面一半，因為該等基因係按重要性之順序排列。

該預後組可包括來自表S6之一個或兩個陰性基因或可由來自表S6之兩個陰性基因組成。

選擇基因之數目及精選品以便提供一預後組，其至少能夠區別具有良好預後之腫瘤與具有不良預後之腫瘤(或具有高NPI之腫瘤及具有低NPI之腫瘤)。

該預後組可包括表S6之不多於60個基因。該預後組可包括表S6之不多於50個基因。該預後組可包括表S6之不多於40個基因。該預後組可包括表S6之不多於30個基因。該預後組可包括表S6之不多於20個基因。該預後組可包括表S6之不多於10個基因。該預後組可包括表S6之不多於5個基因。

該預後組可包括或基本上由表S6之5至60個基因組成。該預後組可包括或基本上由表S6之10至40個基因組成。該預後組可包括或基本上由表S6之10至30個基因組成。該預後組可包括或基本上由表S6之10至20個基因、或表S6之20至30個基因、或表S6之較佳30至40個基因組成。

該預後組(較佳約10個或約20個或約30個基因)可選自表S6之前約40個、或約30個、或約20個基因。約10個基因可選自表S6之前約15個基因。該等約10個基因可為表S6之前10個基因。

該預後組可包括或基本上由選自由以下數目之基因組成之群的約40個或約30個或約20個或約10個基因組成：表S6之前約40個或約30個或約20個或約10個陽性基因及視情況表S6之一個或兩個陰性基因。該預後組可包括或由以下數

目之基因組成之群的約30個基因組成：表S6之前約30個或約40個陽性基因、及視情況表S6之一個或兩個陰性基因。

該預後組中與U133A微陣列相同之基因之數目較佳如以上所描述而受到局限。

術語"約"較佳意謂所陳述之基因的數目加上或減去較大為10%的所陳述之基因的數目。

該預後組之供應容許診斷工具(舉例而言，核酸微陣列)定做且用以預測、診斷或將腫瘤分亞型。此等診斷工具可進一步與一經程式化之電腦聯合使用，以測定使用該診斷工具(舉例而言微陣列)而獲得之表現輪廓，且如以上所討論將其與一"標準"表現輪廓或"已知"預後之表現輪廓之資料庫進行比較。在此情況下，該電腦不僅為使用者提供用以診斷病人腫瘤之存在或類型的資訊，且同時該電腦獲得可藉由其測定該"標準"表現輪廓的進一步表現輪廓且因此可更新其擁有的資料庫。

因此本發明首次容許製造包含對應於該預後組之探針的專門化晶片(微陣列)。該陣列之精確物理結構可變化，且自附著至一個2維固體基材之寡核苷酸探針變化至已經以獨特標記物(舉例而言，"條形碼")而被個別地"標記"之自由浮動探針。

可以直接或間接方式來查詢具有已知預後之表現輪廓的資料庫。"直接"方式為，將病人之表現輪廓與該資料庫中其它個別表現輪廓進行直接比較以測定哪一個輪廓(及因此哪一個預後)傳遞最佳匹配。或者，該查詢可更"間接"完

成，例如，可將該病人表現輪廓與該資料庫中用於一特別預後指派(舉例而言"，不良")或一預後值或值範圍(較佳來自NPI，舉例而言，高NPI)之"標準"輪廓進行簡單比較。間接研究之優勢為，"標準"輪廓因其代表許多個別輪廓之集合而將呈小得多的資料密集型且可儲存於相對便宜的資料載體或其它記憶體設備上(舉例而言，電腦系統)，然後其可根據本發明形成該套組之部分(意即，與該等微陣列相關聯)。

在直接研究中，資料載體可能將為具有大得多的標度(舉例而言電腦伺服器)之資料載體，因為將不得不儲存許多個別輪廓。

藉由將病人表現輪廓與標準輪廓(間接研究)及總體中預測定之統計變化進行比較，亦將可能傳遞關於病人表現輪廓如何接近地匹配該"標準"典範輪廓之"置信度值"，如以上所討論。此值將提供給臨床醫師關於該預後之可信賴性及(例如)是否應重複該分析的有價值資訊。

如以上所提及，亦可能將病人表現輪廓儲存於該資料庫上，且任何時候均可使用此等表現輪廓以更新該資料庫。

在第六態樣中，本發明提供了一種方法以用於識別差別表現於一腫瘤群內之一組基因，該方法包括自該群之複數個腫瘤中的各個提供一表現輪廓；根據腫瘤之分子亞型對該等輪廓進行分類，且分析一亞型內之表現輪廓以識別該組基因(其中該等基因差別表現於彼亞型中)。

此方法與van't Veer等人(10)之方法區別在於，van't Veer

等人中偶發淋巴結陰性乳房腫瘤之初始選擇涉及藉由臨床分析分亞型，而非以分子水平分亞型。

當然，本發明之此態樣及隨後態樣與前述態樣緊密相關。除非本文清楚地另有規定，否則所揭示之前述態樣的較佳特性因此亦可應用於此態樣及隨後態樣。

在本發明之第六、第七及第八態樣之情形中，術語"表現輪廓"並不侷限於該預後組之基因。相反，其通常係指該群之腫瘤內基因的表現水平，包括(但不必僅僅)差別表現於一分子亞型內之基因的表現水平。

藉由本發明之第六態樣而獲得之該組基因(下文中為"區別組")的差別表現可指示該群之腫瘤的一特別表型或遺傳型或表示其之特徵。該方法較佳包括使該區別組之差別表現與一特別表型及/或遺傳型相關聯。具有不同但已知表型及/或遺傳型之大量樣品中的該區別組之表現輪廓可經測定以在該區別組之一特別基因表現輪廓與一特別表型及/或遺傳型之間確立相關性。

該差別表現可表示被指派給腫瘤之臨床參數或醫學類的特徵，此作為治療或診斷腫瘤病人之部分，舉例而言預後(諸如NPI值或NPI類)之量測。該區別組之差別表現可容許將一腫瘤樣品指派給至少兩個不同遺傳型或表型類之一。

本發明之第六態樣之方法可進一步包括用以將一類指派給病人之腫瘤樣品的步驟，其中該區別組之基因之差別表現可表示該類之特徵，該等步驟包括提供該區別組之樣品內的表現水平，且基於該等表現水平將一類指派給該腫瘤。

指派該類之步驟可包括使用統計技術，諸如(但不限於)加權表決、支撐向量機或分級群集，如先前所討論。較佳地，該方法包括識別該腫瘤樣品之分子亞型及使用對該亞型特定之區別組的步驟。

另外或其它，本發明之第六態樣之方法可包括測定一腫瘤樣品內該區別組之表現水平、自該等表現水平測定一表現輪廓及將該輪廓添加至資料庫的步驟。該腫瘤樣品之分子亞型亦較佳被識別，且較佳被添加至該資料庫。

可表示一特別類之特徵的標準輪廓可獲自已知類之至少兩個表現輪廓，其中該等表現輪廓獲自該區別組之基因。該標準輪廓較佳對類及分子亞型特定。另外或其它，將已知類(及視情況，亞型)之表現輪廓添加至該資料庫。

另外或其它，第六態樣之方法可進一步包括用以核對治療過程中該腫瘤之類的改變。在一實施例中，自治療不同階段(舉例而言，治療起始及治療結束)之腫瘤提供表現輪廓，且比較其以測定類的改變，其中該等表現輪廓獲自該區別組基因之表現水平。較佳將該等表現輪廓與標準及/或已知輪廓進行比較以測定該類。

較佳使用能直接量測腫瘤樣品內基因表現產物之水平的技術(諸如，組織病理學(舉例而言，免疫學)技術或基因表現技術)來執行根據分子亞型之分類。最佳為基因表現技術。然而，亦可使用能夠精確區別分子亞型之臨床技術。

該等腫瘤較佳為乳房腫瘤且該分子亞型較佳對應於該腫瘤之ER(雌激素受體)狀態(舉例而言ER+)。然而，可將該方

法應用於其它群之腫瘤(舉例而言，肺腫瘤、卵巢腫瘤及淋巴瘤)及/或其它分子亞型(舉例而言，擴散大B細胞淋巴瘤中的胚中心類似物及活化B細胞類似物)。對表現輪廓之類進行執行以測定差別表現之基因的分析較佳包括微陣列之顯著分析(SAM，文獻12)，其可識別在進行比較的樣品之間表現水平顯著變化之基因。較佳地，該分析涉及統計分析，例如使用加權表決、支撐向量機及/或分級群集(稍後可見此等技術之解釋)。

在本發明之第七態樣中，提供了藉由本發明之第六態樣而獲得之該組基因。

在本發明之第八態樣中，提供了使用該區別組以將一腫瘤樣品指派給一特別類。

現在將藉由實例並參考隨附圖式來說明本發明之態樣及實施例。對於熟習此項技術者而言，進一步態樣及實施例將顯而易見。本文中所提及之全部文獻均以引用之方式併入本文中。

【實施方式】

材料及方法

乳房組織及臨床資訊

人類乳房組織在NCC博物館及倫理委員會適當批准後自NCC組織博物館獲得。腫瘤狀態之組織學證實及雌激素受體(ER)及ERBB2免疫組織化學狀態由新加坡綜合醫院之病理學部門提供(參見臨床資訊之補充資訊)。樣品包含至少50%腫瘤含量。NPI狀態按如下計算：腫瘤尺寸(cm)*0.2+

等級+淋巴結點(陰性結=1個點；陽性結，1至3個陽性=2個點；陽性結，4個或4個以上=3個點)。因為Stanford資料集中腫瘤尺寸係使用CAT系統而得以定義，所以吾人為各個CAT等級指派一大約值(意即T1=2 cm，T2=3.5，T3=5，T4=3.5)。

樣品製備及微陣列雜交

使用Trizol試劑自組織提取RNA，且對該RNA進行處理以根據製造者指令使用U133A基因晶片進行Affymetrix基因晶片雜交。

資料處理及分析

使用Genedata Refiner對原始基因晶片掃描進行質量控制且藉由除去在全部基因中不存在表現之基因來對該等原始基因晶片掃描進行過濾(意即"A"呼叫)。表現值經受log2轉換，且藉由由各個樣品來對全部剩餘基因進行中值定心而得以規格化。使用Genedata Expressionist或習知電子表格應用來執行資料分析。未監督之資料集(圖1，a-b)包含能在全部經良好量測之樣品間顯示出>1.5之標準偏差(SD)的基因。用於基因選擇之變化過濾器的較小變化亦會產生非常類似之結果(P. Tan，未出版之資料)。自分析中除去相同基因之雙重複探針，從而留下用於每個基因之一個探針。平均連接分級群集係使用CLUSTER而得以執行，且藉由使用TREEVIEW而得以顯示。實施微陣列之重要性分析(SAM)(12)以識別經差別調節之基因。"錯誤發現率"對於圖1c而言為0.1%，且對於圖2而言為15%。可如在Golub等人

(13)(補充資訊)中計算加權表決(WV)、留一交叉確認(LOOCV)檢定及預測力(PS)。使用SPSS產生Kaplan-Meier存活曲線，且對數秩測試用於計算存活曲線之間的差異之統計顯著性。藉由卡方分析來測定基因表現與臨床變數之間的統計關聯性。

加權表決(WV)及留一交叉確認(LOOCV)檢定之描述

加權表決(WV)：加權表決演算法利用一訊雜比(S2N)度量以執行二進位分類。屬於一預估組之各個基因被指派一"投票"，其表現為待分類之樣品內基因表現水平之間的加權差異及平均類之平均表現水平。加權係使用相關性度量而得

以測定：
$$P(g,c) = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2}$$
 (μ 及 σ 指示兩類中之各類的基因表現水平之平均及標準偏差)。一特別類指派之最終投票藉由對由在類區分中所使用之各個基因所得到的加權投票進行加

和來計算。"預測力"(PS)被定義為：
$$PS = \frac{V_{WIN} - V_{LOSE}}{V_{WIN} + V_{LOSE}}$$
 其中 V_{WIN} 及 V_{LOSE} 分別為取勝及失敗類之投票總數。PS反映取勝之相對餘量，且因此提供了預測確定性之定量反映。

留一交叉確認(LOOCV)：吾人使用標準留一交叉確認(LOOCV)研究來評定訓練組之分類精確度。在LOOCV中，最初"省去"該訓練組中一個樣品，且對剩餘樣品執行分類器操作(舉例而言，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"省去"樣品，且然後對該訓練組中全部樣品重複此過程。

結果及討論

使用未監督群集定義乳癌之分子亞型

已經提議，乳癌中內在基因表現變化之顯著比例可歸因於屬於區別性"分子亞型"(舉例而言，ER+及ER-腫瘤)(8-9，14)之不同腫瘤。在其中腫瘤經處理而不管亞型的初始分析中，吾人不可使人信服地識別與NPI關聯之表現簽名。吾人假定此可能歸因於亞型間基因表現方面的驚人差異(亞型間差異)，其潛在地遮蔽亞型內更微細之變化樣式(亞型內差異)。為避開此問題，吾人實施一方法學，其中各個分子亞型係作為獨立資料集而得以處理。簡言之，首先使用多種未監督之群集技術以根據其各自之"分子亞型"種類來廣泛隔離一組乳房腫瘤表現輪廓。其次，然後獨立分析各個亞型內之腫瘤以定義可能與NPI或其組份元素關聯之表現簽名。

使用 Affymetrix U133A 基因晶片，吾人產生了可獲自吾人本地最主要之中國病人人口的98個偶發乳房腫瘤之表現輪廓。資料規格化及預處理後，吾人應用一標準偏差過濾器以識別367個基因之組，其在腫瘤系列間展現出高的基因表現變化程度，且使用此基因組以使用未監督分級群集基於其總體相似性使該等腫瘤表現輪廓成群。該等乳房腫瘤自身隔離為3個主要亞群，其分別稱為ER+、ER-及ERBB2+(圖1a)。此隔離樣式係使用主要組份分析(PCA)而得以證實，該PCA為一獨立分析技術(圖1b)，其傳遞高度相似之結果。為穩固地識別此等成群，吾人使用SAM(12)以識別差別表現於該等亞型之間的基因。在0.1%之FDR("錯誤

發現率")時，吾人識別了以亞型特定方式而被顯著調節之409個基因(圖1c)。

表S5之清單代表藉由SAM而得以識別為在各個分子亞型(ER⁺、ER⁻、ERBB2⁺)中受顯著調節之頂部50個基因。該等基因藉由其S2N相關性比率(其反映在不同群中所觀察到的表現擾動之程度)而得以排列等級。

在此等基因與其它研究(文獻8-11)所報告之相似清單之間存在良好重疊。

該409個基因之組的大約69%展現出ER⁺亞群內的增加表現，包括雌激素受體基因ESR1及經雌激素調節之基因(諸如LIV1、TFF1及MYB)(補充資訊)。與其它研究相一致，在此亞型中亦觀察到GATA3、HNF3a、膜聯蛋白A9及XBP1之高表現水平(8-9，11)。ER⁻亞群與基部乳房上皮標示物(角蛋白5及17)、基底膜蛋白質laminin 1、絲胺酸蛋白酶KLK5(其已經與差疾病預後相關聯(15))及絲胺酸蛋白酶抑制因子maspin(一種他莫西芬誘導之基因，先前已經報告其表現於與ER(16)相反之樣式)之高表現相關聯。最後，ERBB2⁺亞型與ERBB2受體及物理連接至17q基因座之其它基因(諸如GRB7及PMNT(14))的高表現水平相關聯，此暗示存在DNA擴增。然而，特定在ERBB2⁺亞型中展現增加之表現的大多數基因並不受限於17q基因座而是發現於整個基因體中，諸如S100鈣結合族(S100A8、A9)之成員。總體而言，吾人之結果可確認及證實先前報告，即大多數乳癌腫瘤的確可基於其整體基因表現輪廓而被細分為區別性分子亞型。

與ER+腫瘤內NPI相關聯之預後組之識別

吾人集中在屬於ER+分子亞型之34個腫瘤，且嘗試識別此亞型內表現可能與NPI狀態關聯之基因。傳統上，通常藉由NPI將乳癌病人分層為3個主要群—"良好"預後(NPI < 3.4)、"中等"預後(NPI 3.4-5.4)及"差"預後(NPI > 5.4)(2)。其它研究已經為可定義此等群(17)之截止值而提議了稍稍不同之值，此可反映出不同計分病理學家之間變化性的效果。為避免在測定合適NPI截止值時產生任何潛在偏差，吾人進行一移動臨限值分析，其中藉由一可自2.3-7.8穩定增加之NPI臨限值將ER+腫瘤劃分為一系列二進位群。在各個臨限值處，識別能在兩個群之間顯示表現之顯著變化的基因。吾人發現，使用3.8至4.6之NPI截止值產生了一具有62個差別表現基因之基因組(圖2a)，其大部分在ER+樣品中以高NPI而展現出增加之表現(圖2b)。吾人將此62個組員之基因組稱作"NPI表現簽名"或NPI-ES，其顯示於表S6中。屬於該NPI表現簽名之基因與牽涉於腫瘤發生之廣泛種類的細胞功能相關聯，該等細胞功能包括DNA複寫及細胞分裂(APRT、MCM4、KNSL 1、CDC2)、細胞訊號傳輸(向化性激動素配位子1、Met、ShC)、細胞凋零(生存蛋白質(survivin)、CD27結合蛋白質)及細胞黏附(大的圓盤同系物(discs-large homolog)7、tetraspan 1)。在個別NPI組份(腫瘤尺寸、腫瘤等級、淋巴結狀態)中，腫瘤等級似乎代表NPI-ES之分子組成的主要貢獻者(補充資訊)。

藉由NPI-ES而對腫瘤之分類定義兩個離散分子群

使用分子輪廓以進行腫瘤分類之一個經提議之優勢為能夠以數學方式量化該分類之置信度水平(11)，若該分類會影響隨後治療過程則其尤為重要。在此情景中，治療醫師接著可權衡預測之置信度水平與特定干涉之潛在發病。值得注意的是，雖然吾人之資料集內ER+樣品與標準NPI值(2-8)之連續譜相關聯，但是使用NPI-ES進行群集分析似乎將該等ER+腫瘤分離為兩個明顯離散之群(圖2b)，此提高了能展現基於組織病理學參數之連續值的樣品仍然可在分子水平上分離為離散種類的可能性。

為更佳定義NPI-ES能確信地辨別此等兩類之能力，吾人使用加權表決(13)(一種受監督之學習演算法)以辨別能展現NPI-ES之高及低表現之腫瘤，且使用確定的留一交叉確認(LOOCV)檢定來測試受訓練演算法之分類準確度。除分類準確度以外，亦如Golub等人(13)所描述計算定量度量(預測力，PS)以提供預測置信度之分析(圖2c)。該WV分析顯示NPI-ES能傳遞91%之LOOCV分類準確度，其具有3個錯誤分類。在被錯誤分類之3個樣品中，2個與低預測力($PS < 0.3$)相關聯，且因此代表"低置信度"或"非確定"分類。實際上，在與"高置信度"分類($PS > 0.3$)相關聯之(34個中之)29個ER+腫瘤中，僅有一個樣品被錯誤分類。此等結果暗示，可使用NPI-ES以高置信度將吾人之資料集中大多數ER+腫瘤分類為離散群。

使用全部腫瘤衍生NPI表現簽名，而不管是否為亞型

吾人使用兩步方法學定義NPI-ES。最初，使用未監督之

群集以根據其各自"分子亞型"(意即ER+、ER-、ERBB2+)來群集腫瘤。為可能與NPI相關聯的表現簽名分析各個亞型內之腫瘤)。此處，吾人顯示執行第一步(定義區別性分子亞型)在NPI-ES之識別方面頗為重要。

吾人組裝了一由全部79個腫瘤(而不管是否為分子亞型)組成之資料集，且執行了一移動NPI臨限值分析以定義一"合適"NPI臨限值，如上所述(參見圖2a)。吾人發現，使用4之NPI臨限值產生了總共44個差別表現之基因。在此44個基因之組中，16個(35%)亦屬於NPI-ES(其獲自ER+樣品)。

吾人使用加權表決(WV)及交叉確認(LOOCV)檢定來分析此44個基因之組確信地將該等腫瘤樣品分類為離散群之能力。如在圖4中所見，比較圖2c，低置信度($PS < 0.3$ ，紅色區域)樣品之數目以及錯誤分類比率(對該44個基因之組而言為9%)均顯著減少。此結果指示該44個基因之組(基於全部79個腫瘤)在預測腫瘤之NPI狀態方面不如NPI-ES對ER+腫瘤有效。

在圖4中，樣品按照其NPI值(X軸)而得以分類，加權表決用以分類樣品，且各個樣品之預測力(Y軸)基於Golub等人(13)而得以計算。認為預測力 < 0.3 之樣品分類是"非確定"或"低置信度"(灰色區域)。比較圖2c，觀察到較高數目的"非確定"(低PS)樣品及錯誤分類樣品。

在預測一獨立資料集內的NPI狀態方面，獲自全部腫瘤(而不管亞型)之44個基因之組亦不如NPI-ES有效。將Rosetta資料集用作盲目測試組，吾人將該44個基因之組應

用至Rosetta資料集中所發現之49個ER+腫瘤，且使用學生之t測試來測定能表現該44個基因之組的高水平且擁有一高NPI的ER+腫瘤之間的關聯的重要性。吾人獲得了該44個基因之組的為0.29之p值，此與NPI-ES的為0.0004之p值相比較顯著性差很多。

有趣的是，NPI-ES儘管獲自ER+腫瘤之分析，但是即使當應用於Rosetta資料集中全部78個腫瘤時仍勝過該44個基因之組。為說明此點，將該等78個Rosetta腫瘤分別劃分為NPI<3.4(良好預後)及>3.4(中等預後)之兩個群。然後使用加權表決以由NPI-ES或該44個基因之組來分類該等Rosetta腫瘤。如表S3中可見，NPI-ES傳遞80%之分類準確度，相比較該44個基因之組傳遞70%分類準確度。

與組織學等級(1&2對比3)關聯之基因

因為經典NPI為一獲自腫瘤等級、腫瘤尺寸及淋巴結狀態之複合度量，所以吾人定義了此等個別元素中之各個元素對NPI-ES之分子組成所作的貢獻。使用SAM以識別與三個組織病理學變數中之各個變數相關聯之基因，吾人不能夠令人信服地識別其表現與腫瘤尺寸或淋巴結狀態顯著關聯的基因。相反，在組織學等級之情況中，發現顯著大量之基因差別表現於等級1或2與等級3腫瘤之間，且此等級關聯基因組內之基因顯示與NPI-ES之實質重疊(66%)(表S6)。此等結果暗示，能展現不同組織學等級之腫瘤可在生物學方面區別開，且腫瘤等級為NPI表現簽名之關鍵貢獻者，而剩餘的兩個參數(腫瘤尺寸及淋巴結狀態)則傳遞相對較少之

貢獻。

NPI-ES在多個獨立乳癌表現資料集的應用

為在一系列盲目"測試組"中測試NPI-ES預測NPI狀態及疾病預後兩者之能力，吾人使用可公開得到之兩個獨立乳癌資料集。第一資料集(稱為Rosetta資料集)由使用基於寡核苷酸之微陣列來測繪的78個淋巴結陰性乳房腫瘤組成，且亦包含各個病人的"無病存活"(DFS)之持續時間(自起始腫瘤診斷至出現一新的遠距離轉移之時間)(10)。重要的是，若干研究先前已經顯示即使在結點陰性乳癌中NPI仍具有預後價值(18、19)。第二資料集由使用cDNA微陣列來測繪之78個乳房癌與全部病人存活資訊組成(稱為Stanford資料集)(14)。此等資料集之可獲得性容許吾人獨立測試NPI-ES之預測能力，因為Rosetta及Stanford資料集在多個方面與吾人之資料不同，包括I)病人人口，II)樣品處理協定，III)計分病理學家及IV)陣列技術及探針組之選擇(在Rosetta及Stanford資料集中為兩色且吾人之資料集中為單色)。

Rosetta乳癌資料集：在藉由SAM分析而識別之能定義ER+、ER-及ERBB2+亞型的409個基因中，在Rosetta微陣列上發現了276個基因(67%)。吾人將此基因組應用至該等78個Rosetta腫瘤輪廓，且識別了屬於ER+分子亞型之49個腫瘤(參見圖5)。為將NPI-ES應用至此等腫瘤，吾人測定來自62個基因之屬於NPIES的46個亦存在於該Rosetta微陣列上。因為該Rosetta資料集係基於一不同於吾人之陣列技

術，所以不可能直接應用在吾人資料集上發展起來的受訓練加權表決模型來分類該等Rosetta腫瘤。

然而，遵循Ramaswamy等人(20)中所描述之用以比較不同陣列技術間的基因組之策略，吾人使用分級群集以使用46個基因之重疊NPI-ES組來將該等49個ER+Rosetta腫瘤分群。該群集分析將該等49個ER+ Rosetta腫瘤劃分為2個群，其分別由24個及25個能展現NPI-ES之"高"及"低"表現水平的腫瘤組成(見圖6)。

吾人比較此等兩個亞群內之腫瘤以測定其是否與其NPI值差異相關聯。使用兩種區別性統計研究，其中該等腫瘤NPI值作為連續梯度(學生之T測試)或作為兩個離散群(卡方分析，使用為3.4之經典NPI截止點)來處理，可展現NPI-ES之高表現的腫瘤與可表現NPI-ES之低水平的腫瘤相比較一貫顯示出具有顯著更高的NPI值(對於連續分析 $p=0.0004$ ，對於二進位分析 $p=0.0087$)(表1a)。此分析指示即使在藉由一不同陣列技術而產生之獨立資料集內NPI-ES之表現仍與ER+腫瘤內經典NPI狀態顯著關聯。

為比較NPI-ES之預後力與經典NPI系統分級，可執行優勢率計算(表1b)。表現NPI-ES之高水平的ER+腫瘤的病人與表現NPI-ES之低水平的腫瘤相比較具有對於遠距離轉移而言五年內為10.3之優勢率(95% CI 2.4至44.0， $p<0.001$)。比較起來，具有 >3.4 ("中等"預後)之經典NPI指數之ER+腫瘤的病人與 <3.4 ("良好"預後)之NPI指數的ER+腫瘤相比較具有對於遠距離轉移而言為6.1之較低優勢率(95% CI

1.6-23.4, $p=0.06$)。吾人亦使用Kaplan-Meier存活分析比較NPI-ES及NPI之預後效能(圖3)。與其它研究相一致, 具有低NPI(<3.4)之腫瘤的病人與高NPI(>3.4)的病人相比較展現出更佳DFS($p=0.007$, 圖3a)。當藉由NPI-ES重新分層此相同人口時, 具有能展現NPI-ES之高表現之腫瘤的病人與具有能表現NPI-ES之低水平的腫瘤的病人相比較展現出更佳無復發存活($p=0.0007$)。總體而言, 此資料暗示, 對於ER+腫瘤而言, NPI表現簽名之預後力可勝過經典NPI分級系統。

Stanford 資料集: 使用一相似研究以測試Stanford資料集上之NPI-ES(參見圖7)。在用以定義ER+、ER-及ERBB2+亞型之SAM-409基因組中, 在Stanford微陣列上發現136個基因(<http://genome-www5.stanford.edu/MicroArray/SMD/>), 且使用此等基因以群集該等Stanford腫瘤從而識別屬於該ER+分子亞型之46個腫瘤(來自丟棄6個腫瘤之類似正常腫瘤亞群之後的72個腫瘤, 該亞群很可能由於污染非惡性組織之存在而引起)。

然後使用NPI-ES(在Stanford微陣列上31個配對)將此等46個腫瘤群集(參見圖8)為"高NPI-ES"(13個腫瘤)及"低NPI-ES"群(33個腫瘤)。再次, 學生之t測試揭示高及低表現NPI-ES亞群與經典NPI狀態(表1a)之間的顯著關聯性($p=0.001$)。另外, 在低NPI-ES表現腫瘤之病人與高NPI-ES表現腫瘤之病人相比較中, KM存活分析亦證明顯著($p=0.0493$)總體存活優勢(圖3d)。

有趣的是, 在能表現NPI-ES之高水平之ER+腫瘤與Sorlie

等人(14)中所識別之"魯米那(Luminal)C"分子亞型之間似乎存在強烈相關性，儘管屬於NPI-ES之62個基因中沒有一個已被報告為表現於後者中。有趣的是，Sorlie等人(文獻14)先前報告了基於500個基因之一"內在"組的"魯米那C"亞型之識別。在"魯米那C"腫瘤與能表現NPI-ES之高水平之腫瘤之間似乎具有一強烈重疊(96%)，儘管如以上所提及屬於NPI-ES之62個基因中沒有一個發現於此"內在"組中。此說明於表S11中。

NPI-ES之預後力可與先前描述之乳癌的"預後簽名"相比較

在Van Veer等人(10)之相同研究中，作者亦識別了一能預測乳房腫瘤之DFS狀態的70個基因之"預後"表現簽名(PES)。有趣的是，在屬於NPI-ES與PES之基因之間存在最小重疊，因為在該兩者之間僅發現一個共同基因。為比較NPI-ES及PES對Rosetta ER+腫瘤之預後效能，吾人使用KM存活分析以比較藉由NPI-ES(圖3b)或PES(圖3c)而得以分層的病人之DFS。由PES($p=0.0001$)與NPI-ES($p=0.0007$)進行比較觀察到稍稍較佳的效能。然而，與該PES關聯之邊際改良出人意外，因為該PES之識別係直接基於此等相同腫瘤之表現輪廓及臨床資訊。因此，該等Rosetta腫瘤對該PES而言並非"盲目"，而在NPI-ES之情況中，該等Rosetta腫瘤代表一真實獨立測試組。實際上，當將PES及NPI-ES應用於Stanford ER+腫瘤時，對5年內復發而言兩個分子簽名均傳遞高度相似之優勢率(PES為3.9對比NPI-ES為4.17)(表

1c)。因此，此等結果暗示，NPI-ES與PES之預後力為相對可比較。

NPI-ES分子簽名之表現預測化學療法反應

在此分析中，吾人在化學療法之前及之後檢查配對之乳房腫瘤樣品內的NPI-ES分子簽名之表現，且將此簽名之表現與最終臨床反應相關聯。

利用一公開可獲得之乳癌資料集("Stanford")，其由20對樣品組成，其在14週阿黴素治療(8)"之前"及"之後"獲得。在NPI-ES中發現之62個基因中，31個基因亦在Stanford微陣列上發現，且該31個基因之組的表現在該等配對樣品中檢查。

在該等20個"之前"樣品中，10個樣品展現NPI-ES表現之高水平(H)，且10個展現低表現水平(L)。如圖10中所示，在該等前10個樣品中，6個在化學療法後保持高表現水平(H->H，以紅色描繪)，而4個在治療後展現低表現水平(H->L，以黃色描繪)。然後將各個群之死亡數字(5年之後)製表，如表S12中所示。

然後執行Kaplan-Meier無復發存活分析，且將其顯示於圖11中。吾人發現，與其它群相比較"H->L"腫瘤具有最佳存活結果($p=0.022$)，而"H->H"腫瘤則具有最差預後。此結果暗示，高表現NPI-ES腫瘤內NPI-ES之向下調節可作為化學療法反應之標示物。

總之，吾人已經識別一可潛在充當NPI之分子替代品的62個基因之表現簽名。藉由顯示其可預測由不同中心產生之

兩個獨立腫瘤組的NPI狀態及疾病預後兩者，獲得了NPI-ES之可靠性的置信度。一個自此研究出現的引人興趣之概念為，在組織病理學水平上顯示明顯連續變數之樣品仍然可在分子水平上分離為離散種類。此可在癌症組織病理學方面提出一主要挑戰，即當正在計分之參數具有連續本性時定義臨床上合適之截止值的困難。藉由承認在可完全評定NPI-ES之臨床效用之前需要執行更多工作，吾人作出結論。首先，NPI-ES之預測力顯然需要進行測試以抗衡遠遠更大之腫瘤群。

其次，儘管吾人已經證明NPI-ES在ER+分子亞型中的適用性，但是NPI-ES之表現似乎又未與其它分子亞型(ER-、ERBB2+)所關聯之NPI值(補充資訊)相關聯。

樣品資料

表 S14 顯示具有差異 NPI 值之樣品間基因預後組 (或 NPI-ES) 的表現資料。該資料特定用於 Affymetrix U133A 基因晶片，且已完成了資料預處理。該預後組之該等基因表現輪廓可用作訓練資料以建構一預測模型 (舉例而言，WV 及 SVM)，然後其可指派一未知腫瘤之 NPI 類。

該資料係製表而定界的，且具有下列格式：

行：

第 1 行：預後組基因之探針_ID

第 2 行：基因名稱

第 3 及其它行：基因表現資料

列：

第1列：樣品ID(35個樣品)

第2列：NPI指數

第3及其它列：基因表現資料

該基因表現資料係如"樣品製備及微陣列雜交"及"資料預處理"(參見材料及方法部分)中所描述而得到。詳言之，原始基因表現資料值藉由用以量測該微陣列之儀器(通常為一微陣列掃描儀，舉例而言Affymetrix)來計算。

表S15顯示平均(μ)及標準偏差(σ)參數以用於各類中預後組之各個基因的加權表決演算法。此等資料可用以指派給定該預後組基因之一組表現水平的未知乳房腫瘤樣品之預後。該資料對應用於來自Affymetrix U133A基因晶片之表現資料的加權表決技術而言具有特定性。

文獻

1. Elston、C. W.與I. O. Ellis. Pathological prognostic factors in breast cancer: I. The value of histological grade in breast cancer - Experience from a large study with long- term follow-up. Histopathology 19, 403-410, 1991.
2. Galea、M. H.、R. W. Blamey、C. W. Elston與I. O. Ellis. The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat. 22, 207-219, 1992.
3. Ellis、I. O., M. Galea、N. Broughton、A. Locker、R. W. Blamey與C. W. Elston. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. Histopathology 479-489, 1992.

4. Balslev 、 I., C. K. Axelsson 、 K. Zedeler 、 B. B. Ramussen 、 B. Carstensen與 H. T. Mouridsen. The Nottingham Prognostic Index applied to 9,419 patients from the studies of the Danish Breast Cancer Cooperative Group (DBCG). *Breast Cancer Res. Treat.* 32, 281-290, 1994.
5. Sauerbrei 、 W., K. Hubner 、 C. Schmoor 、 與 M. Schumacher. Validation of existing and development of new prognostic classification schemes in node negative breast cancer. *Breast Cancer Res. Treat.* 42, 149-163, 1997.
6. Gilchrist 、 K. W. 、 L. Kalish 、 V. E. Gould 、 S. Hirschl 、 J. E. Imbriglia 、 W. M. Levy 、 A. S. Patchefsky 、 D. W. Penner 、 J. Pickren, J. A. Roth及 等人 。 Interobserver reproducibility of histopathological features in stage II breast cancer. An ECOG study. *Breast Cancer Res. Treat.* 5, 3-10, 1985.
7. Buettner 、 P., C. Garbe與 Guggenmoos-Holzmann. Problems in defining cutoff points of continuous prognostic factors: Example of tumour thickness in primary cutaneous melanoma. *J Clin. Epidemiology* 50, 1201-1210, 1997.
8. Perou 、 C. M., T. Sorlie 、 M. B. Eisen 、 v. d. R. M. 、 S. S. Jeffrey 、 C. A. Rees 、 J. R. Pollack 、 D. T. Ross 、 H. Johnsen 、 L. A. Akslen 、 O. Fluge 、 A. Pergamenschikov 、 C. Williams 、 S. X. Zhu 、 P. E. Lonning 、 A. L. Borresen-Dale 、 P. O. Brown與 D. Botstein. Molecular Portraits of Human Breast Tumours. *Nature* 406, 747-752, 2000.

9. Gruvberger 、 S., M. Ringner 、 Y. Chen 、 S. Panavally 、 L. H. Saal 、 A. Borg 、 M. Ferno 、 C. Peterson與 P. Meltzer. Estrogen Receptor Status in Breast Cancer is Associated with Remarkably Distinct Gene Expression Patterns. *Cancer Research* 61, 5979-5984, 2001.
10. van't Veer 、 L. J. 、 H. Dai 、 M. J. van de Vijver 、 Y. D. He 、 A. A. M. Hart 、 M. Mao 、 H. L. Peterse 、 K. van der Kooy 、 M. J. Marton 、 A. T. Witteveen 、 G. J. Schreiber 、 R. M. Kerkhoven 、 C. Roberts 、 P. S. Linsley 、 R. Bernards與 S. H. Friend. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536, 2002.
11. West 、 M., C. Blanchette 、 H. Dressman 、 E. Huang 、 S. Ishida 、 R. Spang 、 H. Zuzan 、 J. A. J. Olson 、 J. R. Marks與 J. R. Nevins. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci* 98, 11462-11467, 2001.
12. Tusher 、 V. G., R. Tibshirani與 G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci* 98, 5116-5121, 2001.
13. Golub 、 T. R. 、 D. K. Slonim 、 P. Tamayo 、 C. Huard 、 J. P. Gaasenbeek 、 H. Coller 、 M. L. Loh 、 J. R. Downing 、 M. A. Caligiuri 、 C. D. Bloomfield與 E. S. Molecular Classification of Cancer : Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* 286, 531-537, 1999.
14. Sorlie 、 T. 、 C. M. Perou 、 R. Tibshirani 、 T. Aas 、 S. Geisler 、 H. Johnsen 、 T. Hastie 、 M. B. Eisen 、 M. van de Rijn 、 S. S. Jeffrey 、

- T. Thorsen 、 H. Quist 、 J. C. Matese 、 P. O. Brown 、 D. Botstein 、 P. E. Lonning 與 A. L. Borresen-Dale. Gene Expression Patterns of Breast Carcinomas Distinguish Tumour Subclasses with Clinical Implications. *Proc. Natl. Acad. Sci.* 98, 10879-10874, 2001.
15. Yousef 、 G. M. 、 A. Scorilas 、 L. G. Kyriakopoulou 、 L. Rendl 、 M. Diamandis 、 R. Ponzzone 、 N. Biglia 、 M. Gai 、 R. Roagna 、 P. Sismondi 與 E. P. Diamandis. Human kallikrein gene 5 (KLK5) expression by quantitative PCR : an independent indicator of poor prognosis in breast cancer. *Clin Chem* 48, 1241-1250, 2002.
16. Martin 、 K. J. 、 B. M. Kritzman 、 L. M. Price 、 B. Koh 、 C. P. Kwan 、 X. Zhang 、 A. Mackay 、 M. J. O'Hare 、 C. M. Kaelin 、 G. L. Mutter 、 A. B. Pardee 與 R. Sager. Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res.*, 60, 2232-2238, 2000.
17. Sundquist 、 M. 、 S. Thorstenson 、 L. Brudin 與 B. Nordenskjold. Applying the Nottingham Prognostic Index to a Swedish breast cancer population. *Breast Cancer Res Treat* 53, 1-8, 1999.
18. Barbareschi 、 M. 、 O. Caffo 、 S. Veronese 、 R. D. Leek 、 P. Fina 、 S. Fox 、 M. Bonzanini 、 S. Girlando 、 L. Morelli 、 C. Eccher 、 F. Pezzella 、 C. Doglioni 、 P. Dalla Palma 與 A. Harris. Bcl-2 and p53 expression in node-negative breast carcinoma : a study with long-term follow-up. *Hum. Pathol.* 27, 1149-1155, 1996.
19. Frkovic-Grazio 、 S. 與 M. Bracko. Long term prognostic value of Nottingham histological grade and its components in early

- (pTINOMO) breast carcinoma. J Clin Pathol 55, 88-92, 2002 .
20. Ramaswamy 、 S. 、 K. N. Ross 、 E. S. Lander與T. R. Golub. A molecular signature of metastasis in primary solid tumours. Nat Genet 33, 49-54, 2003.
21. Travassoli 、 F. A.與Schnitt S. J. (1992) Pathology of the Breast In (Elsevier).
22. 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 1a)Rosetta及 Stanford ER+腫瘤內 NPI-ES表現及 NPI狀態之關聯。第一行代表能表現NPI-ES之高或低水平之腫瘤的數目。

	學生之t測試(連續)		卡方(二進位)		
Rosetta	平均(偏差)	P=0.0004	低(<3.4)	高	P=0.0087
高(24*)	3.1±0.4		13	11	
低(25)	2.3±0.6		22	3	
Stanford		P=0.001			
高(13)	5.3±0.5				
低(33)	4.5±0.6				

*圓括號內數字代表樣品之數目。

表 1b)作為基於經典NPI分級及NPI-ES表現之Rosetta ER+腫瘤內首次事件的五年內遠距離轉移的優勢率

	ER+腫瘤		優勢率*
	無遠距離轉移>5年	<5年	(95% CI)
NPI (p=0.06)			6.08(1.58-23.39)
低(<3.4)	27	8	
高(>=3.4)	5	9	
NPI-ES(p<0.001)			10.27(2.40-43.94)
低	22	3	
高	10	14	

*優勢率係使用標準2乘2表格計算得出。CI代表"置信度區間"。

表1c)作為基於PES表現及NPI-ES表現之Stanford ER+腫瘤內首次事件的五年內復發之優勢率。一個樣品不擁有復發資訊，且已自分析移除(剩餘45個ER+腫瘤)。

	ER+腫瘤		優勢率
	無	復發	(95% CI)
PES(p=0.053)			3.90(0.94-16.25)
低	26	8	
高	5	6	
NPI-ES(p=0.040)			4.17(1.05-16.48)
低	25	7	
高	6	7	

表 S1. 乳房腫瘤之組織病理學*

	年齡	尺寸 (mm)	等級	結	NPI	ER	PR	亞型	LVI	DCIS
ER+										
2000220	52	60	3	34之30	7.2	陽性	陰性	管狀	是	最小
980278	64	40	3	20之14	6.8	陽性	陰性	管狀/ Micropap	是	最小
2000597	57	40	2	12之0	3.8	陽性	陰性	管狀	可能	廣泛
2000609	62	70	2	17之17	6.4	陽性	陽性	管狀	是	無
20020071	58	28	3	16之0	4.56	陽性	陽性	管狀	否	無
20020160	86	120	3	10之0	6.4	陽性	陽性	小葉片狀	否	無
2000787	57	60	3	9之0	5.2	陽性	陽性	管狀	是	無
2000818	52	10	2	11之0	3.2	陽性	陰性	管狀	否	最小
20020051	38	50	3	25之1	6	陽性	陽性	管狀	否	無
20020056	71	20	1	17之2	3.4	陽性	陰性	管狀	否	最小
980197	55	30	3	4之2	5.6	陽性	陽性	管狀	是	最小
980261	60	15	2	9之0	3.3	陽性	陰性	管狀	否	最小
980391	56	20	2	7之0	3.4	陽性	陽性	管狀	否	無
2000768	39	40	3	17之0	4.8	陽性	陽性	管狀	否	最小
2000779	48	55	3	14之0	5.1	陽性	陰性	管狀	否	無
990123	54	55	3	11之7	7.1	陽性	陽性	管狀	否	無
2000422	51	63	3	7之3	6.26	陽性	陽性	管狀	否	最小
2000683	72	35	2	17之0	3.7	陽性	陽性	管狀	否	最小
2000775	51	25	2	12之0	3.5	陽性	陰性	管狀	否	最小
2000804	39	40	3	21之5	6.8	陽性	陽性	管狀	是	最小
980346	52	20	3	4之0	4.4	陽性	陽性	管狀	可能	最小
980383	64	30	2	16之0	3.6	陽性	陽性	管狀	否	最小
990082	49	34	2	16之3	4.68	陽性	陽性	管狀	否	最小
980177	75	26	2	13之6	5.52	陽性	陽性	管狀	是	無
980178	69	32	3	15之2	5.74	陽性	陰性	管狀	否	最小
980403	73	30	3	9之0	4.6	陽性	陽性	管狀	可能	最小
980434	73	30	3	16之0	4.6	陽性	陽性	管狀	否	最小
990075	66	25	3	21之5	6.5	陽性	陽性	管狀	是	無
990113	70	90	3	15之11	7.8	陽性	陽性	管狀	否	最小
990107	50	40	1	18之1	3.8	陽性	陰性	混和盆狀 (tub-mixed)	是	最小

980208	42	25	3	20之5	6.5	陽性	陽性	管狀	否	無
980220	40	37	2	5之0	3.74	陽性	陽性	管狀	是	最小
980221	33	65	3	13之1	6.3	陽性	陽性	管狀	否	無
990375	38	15	1	10之0	2.3	陽性	陰性	管狀	否	廣泛
ER-										
980193	49	25	3	23之3	5.5	陰性	陰性	管狀	否	最小
980216	65	45	2	20之5	5.9	陰性	陰性	管狀	否	無
980256	46	36	3	12之1	5.72	陰性	陰性	管狀	否	無
980285	49	40	3	7之1	5.8	陰性	陰性	管狀	是	最小
980338	55	30	3	7之0	4.6	陰性	陰性	管狀	否	無
980353	58	45	3	25之0	4.9	陰性	陰性	化生	否	無
980411	69	30	2	9之0	3.6	陰性	陰性	管狀	否	無
980441	66	30	3	14之4	6.6	陰性	陰性	管狀	是	無
990174	55	45	2	24之3	5.9	陰性	陰性	管狀	是	最小
2000320	67	20	3	21之20	6.4	陰性	陰性	管狀	是	無
2000500	44	75	3	6之6	7.5	陰性	陰性	管狀	是	無
980247	35	45	3	19之1	5.9	陰性	陰性	管狀	是	最小
990299	58	55	3	17之7	7.1	陰性	陰性	管狀	可能	最小
2000593	60	41	3	15之0	4.82	陰性	陰性	管狀	否	無
2000638	60	40	1	15之0	2.8	陽性	陰性	小葉片狀	否	無
2000731	68	51	3	29之1	6.02	陽性	陰性	管狀	否	最小
2000880	55	15	2	26之0	3.3	陰性	陰性	管狀	否	無
ERBB2										
980194	58	50	3	32之25	7	陰性	陰性	管狀	是	無
980214	49	60	2	13之5	6.2	陽性	陰性	管狀	否	廣泛
980238	62	-20	3	21之7	6.4	陰性	陰性	管狀	否	廣泛
980288	45	60	3	15之13	7.2	陽性	陰性	管狀	是	廣泛
980335	33	3	3	7之3	5.06	陰性	陰性	管狀	是	廣泛
980373	77	30	3	14之0	4.6	陰性	陰性	管狀	否	最小
980380	56			6之0		陰性	陰性			
980395	68	30	3	10之1	5.6	陰性	陰性	管狀	是	無
980396	66	35	3	12之10	6.7	陰性	陰性	管狀	是	廣泛
990115	38	28	3	10之9	6.56	陽性	陽性	管狀	是	廣泛
990134	43	40	3	19之0	4.8	陰性	陰性	管狀	否	無
990148	60	40	2	19之6	5.8	陽性	陰性	管狀	是	最小
990223	52	5	3	21之1	5.1	陽性	陰性	管狀	否	廣泛

2000104	59					陽性	陰性	管狀		
2000171	50	25	2	9之0	3.5	陰性	陰性	管狀	否	無
2000209	58	50	3	7之0	5	陽性	陰性	管狀	否	無
2000210	50	40	3	6之3	5.8	陰性	陰性	管狀	是	無
2000237	43	47	3	40之23	6.94	陽性	陽性	管狀	是	最小
2000287	53	40	3	8之0	4.8	陰性	陰性	管狀	可能	無
2000399	44	40	2	8之0	3.8	陰性	陰性	管狀	否	最小
2000641	47	60	3	24之16	5.2	陰性	陰性	管狀	是	最小
2000652	56	25	3	21之6	6.5	陰性	陰性	管狀	否	最小
2000675	78	55	3	16之16	7.1	陰性	陰性	管狀	是	最小
2000709	45	30	3	16之0	4.6	陰性	陰性	管狀	否	無
2000759	57	7	3	12之0	4.14	陰性	陰性	管狀	否	廣泛
2000813	60	23	3	17之16	6.46	陰性	陰性	管狀	是	廣泛
2000829	51	45	2	10之10	5.9	陰性	陰性	管狀	是	廣泛
20020090	60	45	3	27之19	6.9	陰性	陰性	管狀	是	最小

*此清單包含此研究使用之98個腫瘤中的79個腫瘤之臨床資訊。剩餘19個腫瘤之臨床資訊不完整且不包括於此清單中。僅使用具有完整臨床資訊之79個樣品以用於隨後NPI-ES分析。

表S3，NPI-ES傳遞80%之分類準確度，與可傳遞70%之分類準確度的44個基因之組相比較。

表S3：78個Rosetta腫瘤上之NPI-ES或44個基因之組的分類準確度

	NPI分類(<3.4或>3.4)
	錯誤分類數目(準確度)
44個基因	23(70%)
NPI-ES	15(80%)

表 S5：ER+、ER-及 ERBB2+ 分子亞型中前 50 個經顯著調節之基因的清單

此清單代表由 SAM 識別為在各個分子亞型 (ER+、ER-、ERBB2+) 內經顯著調節之前 50 個基因。該等基因藉由其 S2N 相關性比率而按順序排列，該比率反映不同群之間所觀察到的表現擾動之程度。在此等基因與由其它研究 (文獻 8-11)(正文) 所報告之相似清單之間存在良好重疊。

基因描述	單基因	染色體
ER+ 分子亞型		
雌激素受體1	Hs.1657	Chr:6q25.1
GATA結合蛋白質3	Hs.169946	Chr:10p15
膜聯蛋白 A9	Hs.279928	Chr:1q21
KIAA0882蛋白質	Hs.90419	Chr:4q31.1
碳酸酐酶XII	Hs.5338	Chr:15q22
細胞色素P450，亞族IIB(魯米那誘導)，多肽6	Hs.1360	Chr:19q13.2
動力蛋白，基因絲、輕中間多肽1	Hs.406050	Chr:1p35.1
Sema域，免疫球蛋白域(Ig)，短基本域，經分泌，(semaphorin)3B	Hs.82222	Chr:3p21.3
N-乙醯基轉移酶1(芳基胺N-乙醯基轉移酶)	Hs.155956	Chr:8p23.1-p21.3
絲胺酸(或半胱胺酸)蛋白酶抑制因子，進化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員5	Hs.76353	Chr:14q32.1
細胞色素c氧化酶亞單位Vlc	Hs.351875	Chr:8q22-q23
人類mRNA；cDNA DKFZp564F05(來自純系DKFZp564F053)，mRNA序列	Hs.71968	---
LIV-1蛋白質，經雌激素調節	Hs.79136	Chr:18q12.1
肌鈣蛋白T1，骨骼，慢	Hs.73980	Chr:19q13.4
假想蛋白質FLJ20151	Hs.279916	Chr:15q21.3
Calsyntenin 2	Hs.12079	Chr:3q23-q24
B細胞CLL/淋巴瘤2	Hs.79241	Chr:18q21.3
胍基乙酸酯N-甲基轉移酶	Hs.81131	Chr:19p13.3
與微管關聯之蛋白質 τ	Hs.101174	Chr:17q21.1
假想蛋白質FLJ12910	Hs.15929	Chr:6q25.1
包含WW域之蛋白質1	Hs.355977	Chr:8q21
UDP-葡萄糖神經醯胺葡糖基轉移酶	Hs.432605	Chr:9q31
GREB1蛋白質	Hs.193914	Chr:2p25.1

RNB6	Hs.241471	Chr:14q32.32
人類類胰島素生長因子1受體mRNA, 3'序列, mRNA序列	Hs.405998	---
白細胞間介素6訊號轉換器(gp130, 製瘤素M受體)	Hs.82065	Chr:5q11
LAG1長壽保證同系物2(S.壓縮啤酒釀母)	Hs.285976	Chr:1q21.2
鈣黏附素, EGF LAG七渡G型受體2(火烈鳥同系物, 果蠅)	Hs.57652	Chr:1p21
成對基本胺基酸裂開系統4	Hs.170414	Chr:15q26
G-蛋白質訊號11之因子	Hs.65756	Chr:16p13.3
UDP-葡萄糖神經醯胺葡萄糖基轉移酶	Hs.432605	Chr:9q31
NPD009蛋白質	Hs.283675	Chr:16p13.2
V-myb成髓細胞血症病毒致癌基因同系物(鳥類)	Hs.1334	Chr:6q22-q23
白細胞間介素6訊號轉換器(gp130, 製瘤素M受體)	Hs.82065	Chr:5q11
圓盤, 大(果蠅)同系物5	Hs.170290	Chr:10q23
人類mRNA; cDNA DKFZp434E082 (來自純系DKFZp434E082), mRNA序列	Hs.432587	---
細胞色素P450, 亞族IIB(魯米那誘導), 多肽7	Hs.330780	Chr:19q13.2
HSPC009蛋白質	Hs.16059	Chr:17q21
KIAA1025蛋白質	Hs.4084	Chr:12q24.22
蛋白質酪胺酸磷酸酶類型IVA, 組員2	Hs.82911	Chr:1p35
CGI-49蛋白質	Hs.238126	Chr:1q44
染色體20開啟讀取框架35	Hs.256086	Chr:20q13.11
佛波醇-12-十四烷酸酯-13-乙酸誘導之蛋白質1	Hs.96	Chr:18q21.31
KIAA0876蛋白質	Hs.301011	Chr:19p13.3
假想蛋白質FLJ20152	Hs.82273	Chr:5p15.1
假想蛋白質FU22318	Hs.22753	Chr:5q35.3
三葉草因子1(乳癌, 雌激素誘導之序列表現於其中)	Hs.350470	Chr:21q22.3
聚合酶(受DNA指導), $\delta 4$	Hs.82520	Chr:11q13
推定脯胺酸4-羥化酶	Hs.348198	Chr:3p21.31
GDNF族受體 $\alpha 1$	Hs.105445	Chr:10q26
ERBB2+分子亞型		
氯化物通道, 經鈣活化, 族組員2	Hs.241551	Chr:1p31-p22
v-erb-b2成紅細胞白血病病毒致癌基因同系物2, 獲自致癌基因同系物(鳥類)之神經/成膠質細胞瘤	Hs.323910	Chr:17q11.2-q12
經生長因子受體結合之蛋白質7	Hs.86859	Chr:17q21.1

雙專一性磷酸酶6	Hs.180383	Chr:12q22-q23
包含3之START域	Hs.77628	Chr:17q11-q12
瞬時受體電位陽離子通道，亞族V，組員6	Hs.302740	Chr:7q33-q34
S100鈣結合蛋白質A8(鈣粒蛋白A)	Hs.100000	Chr:1q21
蛋白質磷酸酶1，調節(抑制因子)亞單位1A	Hs.76780	Chr:12q13.13
成纖維細胞生長因子受體4	Hs.165950	Chr:5q35.1-qter
SRY(性決定區域Y)-盒11	Hs.32964	Chr:2p25
未知蛋白質[人類]，mRNA序列	Hs.106642	——
分裂1之類轉導蛋白強化因子(E(sp1)同系物，果蠅)	Hs.28935	Chr:9q21.32
假想基因MGC9753	Hs.91668	Chr:17q21.1
促分裂素活化之蛋白質激酶激酶激酶5	Hs.151988	Chr:6q22.33
KIAA1102蛋白質	Hs.202949	Chr:4p13
脂肪酸羧化酶	Hs.249163	Chr:16q23
轉錄因子AP-2 β (活化強化因子結合蛋白質2 β)	Hs.33102	Chr:6p12
S100鈣結合蛋白質A9(鈣粒蛋白B)	Hs.112405	Chr:1q21
脂肪酸輔酶A連接酶，長鏈2	Hs.154890	Chr:4q34-q35
假想蛋白質FLJ22671	Hs.193745	Chr:2q37.3
犬尿胺酸3-單加氧酶(犬尿胺酸3-羧化酶)	Hs.107318	Chr:1q42-q44
KIAA0644基因產物	Hs.21572	Chr:7p15.1
天冬胺酸 β -羧化酶	Hs.283664	Chr:8q12.1
電子傳遞黃素蛋白， α 多肽(戊二酸酸尿II)	Hs.169919	Chr:15q23-q25
分泌白細胞蛋白酶抑制因子(抗白細胞蛋白酶)	Hs.251754	Chr:20q12
異檸檬酸脫氫酶1(NADP+)，可溶	Hs.11223	Chr:2q33.3
苯乙醇胺N-甲基轉移酶	Hs.1892	Chr:17q21-q22
假想蛋白質FLJ14146	Hs.103395	Chr:1q42.11
墨角藻糖基乳糖3(半乳糖苷3(4)-L-墨角藻糖基乳糖，包括Lewis血型)	Hs.169238	Chr:19p13.3
角蛋白，頭髮，鹼性，1	Hs.32952	Chr:12q13
包含2之PDZ域	Hs.173035	Chr:5p13.3
精胺基琥珀酸合成酶	Hs.160786	Chr:9q34.1
特定顆粒蛋白質(28 kDa)	Hs.54431	Chr:6p12.3
人類 cDNA：FLJ21521 fis，純系COL05880，mRNA序列	Hs.306777	——
犬尿胺酸酶(L-犬尿胺酸水解酶)	Hs.169139	Chr:2q22.1
假想蛋白質FLJ20539	Hs.118552	Chr:11q12.1
脯胺酸脫氫酶(氧化酶)1	Hs.343874	Chr:22q11.21

v-myc髓細胞組織增生病毒相關之致癌基因，成神經細胞瘤所得(鳥類)	Hs.25960	Chr:2p24.1
整合素， $\beta 6$	Hs.57664	Chr:2q24.2
假想蛋白質MGC3077	Hs.433404	Chr:7p15-p14
解偶聯蛋白質2(線粒體，質子載體)	Hs.80658	Chr:11q13
肌球蛋白X	Hs.61638	Chr:5p15.1-p14.3
角蛋白7	Hs.23881	Chr:12q12-q21
類固醇硫酸酯(微粒體)，芳基硫酸酯酶C，同工酶S	Hs.79876	Chr:Xp22.32
包含1之甲酸精同源2域	Hs.95231	Chr:16q22
ATP結合盒，亞族C(CFTR/MRP)組員3	Hs.90786	Chr:17q22
軟骨素 $\beta 1$ ，4 N-乙醯基胺基半乳糖基轉移酶	Hs.11260	Chr:8p21.3
KIAA0485蛋白質	Hs.89121	——
類Kraken	Hs.301947	Chr:22q13
膠原質，XIII型， $\alpha 1$	Hs.211933	Chr:10q22
ER-分子亞型		
角蛋白16(病灶非表皮松解掌蹠角皮症)	Hs.432448	Chr:17q12-q21
γ -胺基丁酸(GABA)A受體，pi	Hs.70725	Chr:5q33-q34
TONDU	Hs.9030	Chr:Xq26.3
角蛋白6B	Hs.432677	Chr:12q12-q13
絲胺酸(或半胱胺酸)蛋白酶抑制因子，進化枝B(卵清蛋白)，組員5	Hs.55279	Chr:18q21.3
角蛋白5(單純性大皰性表皮鬆懈症，Dowling-Meara/Kobner/Weber-Cockayne類型)	Hs.433845	Chr:12q12-q13
SRY(性決定區域Y)-盒10	Hs.44317	Chr:22q13.1
黑色素瘤抑制活性	Hs.279651	Chr:19q13.32-q13.33
基質金屬蛋白酶7(基質分解素，子宮)	Hs.2256	Chr:11q21-q22
分泌型捲曲相關(frizzled-related)之蛋白質1	Hs.7306	Chr:8p12-p11.1
B細胞CLL/淋巴瘤11A(鋅指蛋白質)	Hs.130881	Chr:2p15
人類 cDNA FLJ11796 fis，純系HEMBA1006158，高度類似於人類轉錄因子類叉頭7(FKHL7)基因，mRNA序列	Hs.284186	——
溶質載體族6(神經遞質轉運蛋白)，組員14	Hs.162211	Chr:Xq23-q24
Desmuslin	Hs.10587	Chr:15q26.3
鋸齒狀同系物1	Hs.271977	Chr:2q13-q21
核糖體蛋白質，大P2	Hs.153179	Chr:11p15.5-p15.4
包含三部分模體之29	Hs.82237	Chr:11q22-q23
類調鈣蛋白之皮膚蛋白質	Hs.180142	Chr:10p15.1
橋粒芯膠蛋白 2	Hs.239727	Chr:18q12.1

ropporin、rhopilin關聯之蛋白質	Hs.194093	Chr:3q21.1
晶狀體蛋白， α B	Hs.391270	Chr:11q22.3-q23.1
包含三部分模體之2	Hs.12372	Chr:4q31.23
表皮生長因子受體(成紅細胞白血病毒(v-erb-b)致癌基因同系物，鳥類)	Hs.77432	Chr:7p12
富白胺酸酸性核內蛋白類似物	Hs.71331	Chr:1q21.2
鉀通道，亞族K，組員5	Hs.127007	Chr:6p21
激肽釋放酶5	Hs.50915	Chr:19q13.3-q13.4
前膠原C-肽鏈內切酶強化因子2	Hs.8944	Chr:3q21-q24
假想蛋白質[人類]，mRNA序列	Hs.66762	——
LIM域，僅4	Hs.3844	Chr:1p22.3
角蛋白17	Hs.2785	Chr:17q12-q21
橋粒芯糖蛋白3(天疱瘡尋常抗原)	Hs.1925	Chr:18q12.1-q12.2
角蛋白6A	Hs.367762	Chr:12q12-q13
唾液酸轉移酶8A(α -N-乙醯基神經胺酸酯： α -2,8-唾液酸轉移酶，GD3合酶)	Hs.82527	Chr:12p12.1-p11.2
類Kruppel因子5(腸內)	Hs.84728	Chr:13q21.32
Rho鳥嘌呤核苷酸交換因子(GEF)4	Hs.6066	Chr:2q22
激肽釋放酶6(甘油磷酸鈣，酶)	Hs.79361	Chr:19q13.3
前列腺素-內過氧化物合酶2(前列腺素G/H合酶及環加氧酶)	Hs.196384	Chr:1q25.2-q25.3
染色體20開啟讀取框架42	Hs.180479	Chr:20p12.3
糖蛋白M6B	Hs.5422	Chr:Xp22.2
尿苷磷酸化酶	Hs.77573	Chr:7
Ladinin 1	Hs.18141	Chr:1q25.1-q32.3
多形態腺瘤基因類似物1	Hs.75825	Chr:6q24-q25
橋粒芯膠蛋白3	Hs.41690	Chr:18q12.1
人類 cDNA FLJ30869 fis，純系 FEBRA2004224，mRNA序列	Hs.349611	——
類似於HRAS之抑制因子	Hs.36761	Chr:3q29
半胱胺酸及富甘胺酸蛋白質2	Hs.10526	Chr:12q21.1
瘙癢病反應蛋白質1	Hs.7122	Chr:4q31-q32
澱粉樣蛋白 β (A4)前驅體蛋白質結合，族A，組員2(X11類似物)	Hs.26468	Chr:15q11-q12
急沖同系物類似物(小鼠)	Hs.105940	Chr:11q21
轉化生長因子， α	Hs.170009	Chr:2p13

表 S6：屬於 NPI-ES 之基因 (62 個基因)

DC 13 蛋白質為可在 Rosetta 70 個基因之 "預後" 簽名 (PES，見正文) 中配對的 NPI-ES 之唯一基因，其中 42 個存在於 Affymetrix U133A 晶片內。

基因描述	單基因	生物過程(GO)
陽性基因(60)(高度表現於高NPI腫瘤中)		
腺苷酸轉磷酸核糖基酶	Hs.28914	9116//核苷新陳代謝//擴展：自電子注釋推斷；Pribosyltran；5 e-44
MCM4微染色體保持缺乏4(S.壓縮啤酒釀母)	Hs. 154443	6260//DNA複寫//預測/計算
核酸外切酶1	Hs.47504	6310//DNA重組//實驗證據//6281//DNA修復//實驗證據//6298//錯誤配對修復//預測/計算
類似於金屬硫蛋白1H之蛋白質[人類]，mRNA序列	Hg.367850	——
人類，純系 IMAGE：5270727，mRNA，mRNA序列	Hs.319215	——
DC 13蛋白質	Hs.6879	——
HSPC037蛋白質	Hs.433180	——
H2A組蛋白族，組員Z	Hs.119192	——
圓盤，大同系物7(果蠅)	Hs.77695	7267//細胞-細胞訊號//擴展：未知；GKAP；2.1 e-05
RNA解螺旋酶相關之蛋白質[人類]，mRNA序列	Hs.381097	——
驅動蛋白類似物1	Hs.8878	7067//有絲分裂//實驗證據///7052//有絲分裂紡錘體總成//實驗證據
染色體20開啟讀取框架1	Hs.9329	7067//有絲分裂//預測/計算///8283//細胞增殖//預測/計算
KIAA0095基因產物	Hs.155314	——
解螺旋酶，淋巴特定	Hs.203963	——
相似盒HB9	Hs.37035	6959//體液免疫反應//實驗證據///6357//自Pol II促進子之轉錄調節//預測/計算///7345//胚胎發生及形態發生//實驗證據
染色體X(獨特)9879上的DNA片段表現序列	Hs.18212	——
MAD2有絲分裂阻止缺乏類似物1(酵母菌)	Hs.79078	7067//有絲分裂//預測/計算///7093//有絲分裂檢測點//實驗證據

真核翻譯起始因子4E結合蛋白質1	Hs.433317	6445//翻譯之調節//預測/計算
組織蛋白酶C	Hs.10029	6508//蛋白質水解及肽分解//未記錄///6955//免疫反應//實驗證據
H2B組蛋白族，組員J	Hs.249216	——
蛋白質降解體(前體，類巨蛋白質(macropain))亞單位，β類型，8(大的多功能蛋白酶7))	Hs.180062	6508//蛋白質水解及肽分解//未記錄
假想蛋白質FLJ20105	Hs.89306	——
染色體10開啟讀取框架3	Hs.14559	——
不典型骨髓蛋白質BM039	Hs.283532	——
富小鼠基因群集之可能直向同源基因，C8基因	Hs.30114	——
細胞分裂循環2，G1至S及G2至M	Hs.334562	74//細胞循環之調節//未記錄///7089//有絲分裂細胞循環之起始對照點//未記錄
金屬硫蛋白2A	Hs.118786	6878//銅內穩態//預測/計算
類比蛋白質(geminin)，DNA複寫抑制因子	Hs.234896	7050//細胞循環停止//預測/計算///8156//DNA複寫之陰性調節//預測/計算
低密度脂蛋白受體相關之蛋白質8，載脂蛋白，受體	Hs.54481	7165//訊號轉換//預測/計算///6629//脂類代謝//預測/計算
血液學及神經學表現之1	Hs.109706	——
H1組蛋白族，組員2	Hs.7644	——
nudix(核苷二磷酸連接之部分X)類型模體1-	Hs.388	6979//對氧化性壓力之反應//預測/計算///6281//DNA修復//未記錄
金屬硫蛋白1X	Hs.374950	——
H2B組蛋白族，組員T	Hs.247817	——
Tetraspan 1	Hs.38972	8283//細胞增殖//未記錄///8583//神秘細胞命運差異(sensu果蠅)//預測/計算//7155//細胞黏附//未記錄///6928//細胞運動//未記錄
金屬硫蛋白1H	Hs.2667	——
H3組蛋白族，組員K	Hs.70937	——
核糖核苷酸還原酶M2多肽	Hs.75319	——

包含桿狀病毒 IAP 重複之 5(生存蛋白質)	Hs.1578	86//G2/有絲分裂細胞循環之M 轉換//實驗證據///7048//腫瘤發生//預測/計算///6916//抗細胞凋零//實驗證據
F盒，僅蛋白質5	Hs.272027	6508//蛋白質水解及肽分解//預測/計算
絲胺酸(或半胱胺酸)蛋白酶抑制因子，進化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員1	Hs.297681	---
與溶菌酶關聯之蛋白質橫跨膜4 β	Hs.296398	---
向化性激動素(C-X3-C模體)	Hs.80420	7165//訊號轉換//實驗證據///6954//炎症反應//未記錄///6935//趨化性//實驗證據///6955//免疫反應//未記錄///7155//細胞黏附//實驗證據///7267//細胞-細胞訊號//實驗證據
CD27-結合(Siva)蛋白質	Hs.112058	8624//由細胞外訊號之細胞凋零之誘導//預測/計算///6952//防禦反應//預測/計算
LGN蛋白質	Hs.278338	7186//G-蛋白質偶合之受體蛋白質訊號路徑//預測/計算
小鼠乳房腫瘤病毒受體同系物1	Hs.18686	---
叉頭盒M1	Hs.239	6366//自Pol II促進子轉錄//實驗證據///6979//對氧化性壓力之反應//實驗證據
Met原癌基因(肝細胞生長因子受體)	Hs.316752	7048//腫瘤發生//實驗證據//8283//細胞增殖//預測/計算///7165//訊號轉換//預測/計算
丁醯蛋白，亞族3，組員A2	Hs.87497	---
SBBI26蛋白質	Hs.26481	---
小鼠Shc SH2-域結合蛋白質1之可能直向同源基因	Hs.123253	---
H3組蛋白族，組員B	Hs.143042	---
三葉草因子3(腸內)	Hs.82961	6952//防禦反應//預測/計算///7586//消化//預測/計算
免疫球蛋白 λ 基因座	Hs.405944	---
DNA複寫因子	Hs.122908	---

人類cDNA FLJ30781 fis，純系 FEBRA2000874，mRNA 序列	Hs.301663	——
向化性激動素(C-C模體)配位子18(肺部及活化調節)	Hs.16530	7165//訊號轉換//實驗證據 ///7154//細胞通訊//預測/計算 ///6935//向化//實驗證據 ///6955//免疫反應//預測/計算 ///6960//抗微生物體液反應(sensu無脊椎動物)//預測/計算 ///9607//對生物刺激之反應//預測/計算 ///7267//細胞-細胞訊號//實驗證據
免疫球蛋白K常量	Hs.406565	——
Ty 4同系物1之抑制因子(S.壓縮啤酒釀母)	Hs.79058	6355//轉錄調節，DNA-依賴//預測/計算 ///6357//自Pol II促進子之轉錄//預測/計算 ///6338//染色質模型//預測/計算
父方表現之10	Hs.137476	——
陰性基因(2)(高度表現於低NPI腫瘤中)，BTG族，組員2	Hs.75462	8285//細胞增殖之陰性調節//預測/計算 ///6281//DNA修復//預測/計算 ///6976//DNA損壞反應，p53之活化//預測/計算
細胞色素P450，亞族IVF，多肽8	Hs.268554	6118//電子傳輸//擴展：未知；p450；1.9e-142 ///6693//前列腺素代謝//預測/計算

表S7：行SAM以識別與等級顯著關聯之68個基因(14%之FDR， ≥ 2 倍改變)。此等基因之45個(66%)亦屬於NPI分類器，其在NPI-ES行中標記為"YES"。

基因名稱	NPI-ES
在等級3腫瘤中向上調節之基因	
RAD51-相互作用蛋白質	
DC 13蛋白質	YES
HSPC037蛋白質	YES
相似盒HB9	YES
細胞週期蛋白82	
胞質分裂之蛋白質調節因子1	
富小鼠基因群集之可能直向同源基因，C8基因	YES
驅動蛋白類似物1	YES

H2A組蛋白族，組員Z	YES
DNA複寫因子	YES
MCM4微染色體保持缺乏4(S.壓縮啤酒釀母)	YES
圓盤，大同系物7(果蠅)	YES
ZW10相互作用因子	
MAD2有絲分裂阻止缺乏類似物1(酵母菌)	YES
金屬硫蛋白1H類似蛋白[人類]，mRNA序列	YES
染色體10開啟讀取框架3	YES
核糖核苷酸還原酶M2多肽	YES
細胞分裂週期2，G1至S及G2至M	YES
叉頭盒M1	YES
不典型骨髓蛋白質BM039	YES
解螺旋酶，淋巴特定	YES
RNA解螺旋酶相關之蛋白質[人類]，mRNA序列	YES
金屬硫蛋白1X	YES
人類，純系IMAGE:5270727，mRNA，mRNA序列	YES
金屬硫蛋白2A	YES
金屬硫蛋白1H	YES
KIAA0095基因產物	YES
包含桿狀病毒IAP重複之5(生存蛋白質)	YES
類比蛋白質(geminin)，DNA複寫抑制因子	YES
Zeste同系物2之強化因子(果蠅)	
組織蛋白酶C	YES
nudix(核苷二磷酸連接之部分X)類型模體1	YES
假想蛋白質FLJ10719	
向化性激動素(C-X3-C模體)配位子1	YES
Tetraspan 1	YES
早期細胞凋亡蛋白酵素(proapoptotic caspase)調適因子或蛋白質	
免疫球蛋白λ基因座	YES
H2B組蛋白族，組員J	YES
三葉草因子3(腸內)	YES
CD27-結合(Siva)蛋白質	YES
局部異構酶(DNA)II α 170kDa	
免疫球蛋白λ接合3	
真核翻譯起始因子4E結合蛋白質1	YES
H3組蛋白族，組員K	YES

向化性激動素(C-C模體)配位子18(肺部及活化調節)	YES
溶菌酶關聯之蛋白質橫跨膜4β	YES
小鼠乳房腫瘤病毒受體同系物1	YES
LGN蛋白質	YES
免疫球蛋白□常量	YES
羧肽酶B1(組織)	
met原癌基因(肝細胞生長因子受體)	YES
H2B組蛋白族，組員T	YES
RAB38，組員RAS致癌基因族	
H1組蛋白族，組員2	YES
來自EUROIMAGE 2021883之假想蛋白質	
載脂蛋白B mRNA編輯酵素，催化多肽類似物3B	
H3組蛋白族，組員B	YES
免疫球蛋白重常量γ3(G3m標示物)	
類似於bK246H3.1(類似於免疫球蛋白λ的多肽1，前B細胞特定)	
免疫球蛋白λ輕鏈[人類]，mRNA序列	
免疫球蛋白□輕鏈變數區域[人類]，mRNA序列	
絲胺酸(或半胱胺酸)蛋白酶抑制因子，進化枝A(α-1抗蛋白酶，抗胰蛋白酶)，組員1	YES
蛋白脂質蛋白質1(髓鞘蛋白編碼基因缺陷病，痙攣截癱2，不複雜)	
鈉通道，非電壓門控之1，β(徵候群)	
H4組蛋白族，組員H	
多配體蛋白聚糖2(硫酸類肝素蛋白多糖1，細胞表面關聯，纖維多糖)	
神經菌毛素(NRP)及tolloid(TTL)類似物2	
等級3腫瘤中向下調節之基因	
假想蛋白質FLJ22418	

	魯米那 A	魯米那 C
低 NPI-ES	30	0
高 NPI-ES	2	10

表 S11：具有高及低 NPI-ES 表現之魯米那 A 及魯米那 C 腫瘤的相關性(魯米那腫瘤係基於 Sorlie 等人之結果(2001)而得以識別)

表 S12：然後將各個群死亡數字(5年後)製表，如下：

	H->H	H->L	L->L	L->H
總數	6	4	10	N/A
死亡	4	0	3	N/A
AWD*	1	0	2	N/A

AWD*：帶病活著

表 S13：預後組與 Rosetta 231 基因之間重疊的基因

登記#	相關性	基因名稱	描述
NM_020188	-0.40007	DC13	DC13 蛋白質
NM_001168	-0.33813	BIRC5	包含桿狀病毒 IAP 重複之 5(生存蛋白質)
NM_006763	0.345013	BTG2	BTG 族，組員 2
NM_012177	-0.32571	FBXO5	F 盒僅蛋白質 5
NM_013296	-0.30129	HSU54999	LGN 蛋白質
Contig41413_RC	-0.30837	RRM2	核糖核苷酸還原酶 M2 多肽
NM_018455	-0.33103	BM039	不典型骨髓蛋白質 BM039
NM_002358	-0.30251	MAD2L1	MAD2(有絲分裂阻止缺乏，酵母菌，同系物)類似物 1

圖 S14: 具有不同 NPI 值之樣品間基因預後組 (或 NPI-ES) 之表現資料

UID	NAME	2000220	980278	2000597	2000609	20020071	20020160	2000787	200081
8	20020051	20020056							
980197	980261	980391	2000768	2000779	990123	2000422	2000683	2000775	2000804
	990082	980177	980178						980346
980403	980434	990075	990113	990107	980208	980220	980221	990375	980383
NPI		7.2	6.8	3.8	6.4	4.56	6.4	5.2	3.2
5.6	3.3	3.4	4.8						6
5.1	7.1	6.26	3.7	3.5	6.8	4.4	3.6	4.68	5.52
4.6	6.5	7.8	3.8						5.74
6.5	3.74	6.3	2.3						4.6
200853_at		" H2A 組蛋白族, 組員 Z		"	-0.1454	1.29	-0.2888	-0.1469	0.3389
0.3976	-1.025	0.7639							1.274
-0.7213	1.395	-0.5183	-0.1454	1.481	1.149	1.102	1.105	-0.9016	-0.2015
	0.3702	-0.78	0.7502					-0.2015	0.6147
-0.1024	1.684	0.4969	0.5195	-0.319	0.1196	0.7886	-0.002354	-0.2928	-0.0726
201236_s_at		" BTG 族, 組員 2	"	-0.006272	0.5032	0.9142	-0.1329	0.7774	0.2717
	-0.3218	0.9	-0.4893	2.126					
-0.2778	1.747	1.955	0.1703	-0.09297	-0.8116	0.2803	1.573	-0.4571	0.2552
	0.8867	0.5263	0.278						0.5244
0.6472	1.158	0.1387	-0.09749	0.4156	1.328	0.4434	1.355	0.3473	0.866
201483_s_at		"Ty 4 同系物 1 之抑制因子 (S. 壓縮啤酒釀母)"					0.6097	1.482	-0.0874
1.818	0.9257	0.2263							1.187
0.4099	0.7686	-0.174	0.603	-0.8021	0.4711	0.8151	1.052	0.8867	0.6619
2	1.6	-1.372	-0.8661						-0.7083
-1.684	1.396	0.4893	1.347	-0.3128	-0.5101	-0.09044	0.4318	2.904	0.4475
7	0.01991								-0.391
201487_at		組織蛋白酶 C	0.07473	-0.138	-0.7108	-0.2718	0.6703	1.105	0.9386
3	0.9838	-0.7759	0.2844						-0.274
-1.244	-0.8704	2.864	-1.201	1.285	0.546	-0.9224	1.085	0.1034	-0.2643
	1.502	0.4309	0.9151						-1.447
-0.6552	1.012	-0.6763	-1.624	1.46	-0.292	-0.01074	-0.688		-1.158
201890_at		核糖核苷酸還原酶 M2 多肽							
0.8083	0.7456	-5.399					-0.7399	0.9706	-0.3813
								0.1577	0.621

0.7672	-1.632	-0.244	-0.8654	-0.1484	0.5024	0.8568	1.374	-0.2848	-1.812	0.2609	1.347
0.729	-0.3775	-2.774									
-0.1699	-0.3712	0.3715	0.09703	-0.7396	-0.327	-0.01902		0.8041	-0.9266	-2.221	-1.183
202095_s_at		包含桿狀病毒 IAP 重複之 5 (生存蛋白質)						0.4877	1.084	-0.3761	-0.090
17	-0.1052	1.08									
0.7607	-2.821	2.266	-4.419	-0.4761	0.824	0.1905	0.7446	0.1633	1.969	1.562	-1.104
	0.04571	0.743	0.8446								
-0.009848		-3.205	1.153	-0.6422	2.755	-0.519	0.6679	0.3284	-0.1171	0.3173	-0.689
9	-0.7857	-0.5155									
202188_at		KIAA0095 基因產物									
4	-1.711	0.679	-1.065	1.582		-1.796	2.314	-1.991	-1.603	1.976	-0.6295
-1.575	-0.9511	0.02638	1.178	0.9636	1.625	0.2826	-0.007729		-1.634	-0.9197	-1.993
	-1.106	2.364	2.902	1.597							
2.523	2.11	-1.844	-0.00351	1.418		-0.7783	2.405	-1.969			
202580_x_at		叉頭盒 M1				-0.6508	0.6023	-0.5555	-1.427	0.5444	1.37
-3.152	0.4569	-2.362									-0.08859
0.1443	0.08023	-0.4678	0.585	0.957	0.422	0.7627	-0.8553	-0.333	2.135	0.8166	0.2332
	-3.414	0.3564	0.2976								
0.2955	-0.1124	0.1875	0.1209	-0.427	0.719	-0.4973	-1.537	-1.016			
202833_s_at	絲胺酸 (或半胱胺酸)	蛋白質抑制因子, 進化枝 A (α -1 抗蛋白酶, 抗胰蛋白酶), 組員 1									
0.2746	-0.5269	-1.078	2.707	-0.2308	-1.418	-0.1188	-0.8583	-0.9274	-0.6578	1.595	-2.001
	-0.7663	0.5393	-1.519	4.702							
3.99	-1.54	-0.02002		-0.155	-0.2741	-0.7656	-1.361	1.912	1.678	-0.154	2.179
2.225	4.768	-0.2613									
0.8987	2.563	2.903	-1.034								
203362_s_at		MAD2 有絲分裂阻止缺乏類似物 1 (酵母菌)									
	0.4871	1.438	1.036					0.5555	1.23	0.3254	0.1155
-1.972	1.544	-0.8522	0.8364	0.1556	-0.1722	1.08		0.6537	0.6816	0.5234	-1.458
1	1.186	0.6807	0.2946	-1.65							-0.768
-0.08355		-0.8345	1.333	0.09134	-0.4979	-0.4036	-0.4871	0.6176	-0.915	-1.052	-1.223
203510_at		met 原癌基因 (肝細胞生長因子受體)							-0.6627	0.2747	-3.044
	-2.481	0.6526	-4.04								
0.4437	-3.992	1.427	-2.922	0.8314	-1.178	-2.988	-3.458	-3.411	-5.32	-0.736	-2.917
	-2.182	-3.039	-0.05372								
-3.493	-2.429	-1.026	1.1	-2.892	-2.466	0.4109	-0.7238	-4.735	-3.73	-0.0802	1.504

- 73 -

- 74 -

-3.669	0.342	1.33	1.492	-3.361	0.1566	0.0934	-1.052	0.5683	-4.269	-0.4344	1.387
-0.772	1.844	0.7642									
-0.7578	1.344	0.897	0.5398	-2.669	-0.7838	1.814	1.461	0.9016	0.5718	1.511	0.2541
	0.4053	2.276	0.9434								
0.5052	-1.327										
209114_at		tetraspan 1	0.8555	-0.07584			-2.888	0.727	1.287	0.618	1.377
-0.4489	-0.4048	-0.05417									
1.687	0.3291	-0.4002	-0.4801	2.494	1.825	2.19	1.306	-1.028	0.923	-0.6499	1.26
-1.03	1.344	1.158									
0.9383	3.609	-0.4017	0.3422	-0.6132	0.7248	0.5805	2.243	-1.668			
209398_at		" H1 組蛋白質, 組員2		"	0.6575	1.078	-1.211	2.027	2.668	3.317	
0.775	-0.02843	2.147									
0.8804	0.3597	-1.194	-0.7188	3.047	3.613	-0.5534	-0.2921	-0.7163	1.459	0.1181	2.258
-1.025	1.194	1.332	1.659								
-0.6719	1.729	0.1807	0.586	1.717	3.717	0.9074	1.742	0.5849			
209806_at		" H2B 組蛋白質, 組員1		"		0.04495	1.146	-1.895	0.5906	1.685	1.715
-0.007187		0.05995	1.35								
-0.4676	1.172	-0.5088	-1.877	1.592	2.085	-0.9202	-1.274	-0.3533	1.334	0.1805	1.377
-1.049	0.6906	0.3721	1.173								
-0.003976		0.04341	-0.4338	0.003975		0.4399	1.955	0.2045	0.4444	-0.08685	
209832_s_at		DNA 複寫因子			-0.2093	0.4368	-0.6791	-3.906	1.28	0.9618	0.448
-2.259	1.283	-0.784	0.785								
-0.3879	-0.2683	0.792	1.165	0.8799	0.953	-0.4351	0.2743	1.248	1.03	-0.5027	-1.445
	0.5995	-0.5054	1.063								
-0.009595		-0.244	-0.3794	-0.1792	0.2804	-1.287	-0.2545	-0.2157			
209924_at		向化性激動素 (C-C 模體) 配位子 18 (肺部及活化調節)									
48	-2.039	-0.5577									0.0032
-1.466	0.6619	-3.446	-0.2178	-0.8463	1.369	-1.233	0.7177	-0.5422	-1.256	0.8261	-0.469
3	0.8621	0.746	-0.07244								
0.1775	0.7399	-0.003248	-3.226	-0.8553	0.7689	-0.2968	0.6344	-1.833	0.8685	-2.811	
	-3.967	0.8904	-1.524								
-0.7571	-0.8797										
210052_s_at		染色體 20 開啟讀取框架 1									
0.7662	0.7948	-1.894	1.671			0.03423	0.7286	0.4886	-0.3575	1.121	
-1.204	-0.1785	-0.3447	-0.0865	0.7136	0.9764	1.282	1.064	-1.092	-0.1238	1.548	0.5728
	0.244	-1.466	1.214								

0.08015	1.559	0.1766	-0.2576	0.3938	0.1252	1.078	-0.47	-0.636	-0.7279	-0.2508	-0.771
210559_s_at	"	細胞分裂循環2, G1至S及G2至M"						-0.5529	0.8735		
7	1.068	0.9656	1.144								
-1.567	1.246	-1.044	0.6194	-0.3937	-0.2029	0.3902	1.055	1.288	1.092	-2.252	-0.066
58	2.025	0.7638									
0.6873	-1.305	0.6587	-0.9514	1.169	0.4542	0.02187	0.7038	0.1117	0.3919	0.1868	0.1606
	-0.3175										
210576_at	"	細胞色素P450, 亞族 IVF, 多肽 8						-1.268	-0.8577	-1.438	-1.47
-0.9421	-0.8446	-0.5397									
-1.359	-0.6011	3.761	-0.9546	3.704	4.427	-0.8408	-0.8623	1.306	-1.195	0.05855	9.803
-0.3689	-0.5783	5.219	2.246								
-0.02554		0.2243	-0.5602	0.0802	-0.842	0.004669		2.291	0.9273	-1.093	-1.182
210792_x_at	4.035	CD27 結合 (Siva) 蛋白質						-0.6963	-0.7867	-0.898	-0.06822
6	2.469	-0.1134	-0.5174	1.706							0.0900
-0.2906	0.323	-1.999	0.2819	2.967	0.3527	2.022	1.753	-0.4607	0.6391	0.2633	-0.683
1	-1.031	-2.234	2.18	1.734							
0.7072	0.2691	1.524	0.3786	0.8489	1.1	-0.4151	0.8751	-0.08786			
211456_x_at	"	類似於金屬硫蛋白 1H之蛋白質 [人類], mRNA 序列"						-1.117	-0.041		
23	-2.591	-0.6507									
0.01816	-0.01816	-1.794	-3.482	-1.126	-3.411	-0.8509	-1.1	-2.47	0.43	0.1692	
-1.199	0.5589	1.738	-2.633								
5	0.5023	0.1271	-1.565	-1.591	-0.5746	1.378	0.1608	-0.1149	-3.006	-0.9624	-0.876
-2.313		0.03099	0.7018								
212094_at	父方表現之 10							-0.672	-3.155	-3.211	0.9864
-1.438	1.204	-2.245	-1.56								0.05423
-3.69	-2.33	-0.05982		-5.556	-4.073	1.828	-0.1291	0.8363	-2.39	-0.2446	-3.414
	2.494	2.876	-0.3911	2.367							
3.409	0.03678	2.613	-1.918	-0.8758	-2.307	2.694	2.508				
212141_at	MCM4 微染色體保持缺乏 4 (S. 壓縮啤酒釀母)										-0.01048
2.045	0.6207	0.2774	1.282								
1.395	1.012	0.2859	2.278	-2.371	0.9675	-3.468	0.08903	1.907	0.5054	1.39	1.754
-0.05179		0.1601	1.165								
1.149	0.05549	-0.7758	1.347	0.5721	2.048	0.1982	0.2689	0.3733	0.6741	0.8496	-2.634
	0.1973	-2.22									

212185_x_at	金屬硫蛋白 2A	-0.3496	0.1074	-1.298	-0.8511	0.7438	0.6982	-1.544
-1.558	-0.8231 -1.389							
-0.6833	-0.213	-0.4286	0.8156	0.5282	-0.1693	1.705	-2.072	-0.1536
-1.32	-0.8457	0.08054	1.222					0.3755
0.4245	0.2713	-1.085	-0.3512	-0.3473	0.7846	0.1873	0.5453	-1.284
212484_at	小鼠乳房腫瘤病毒受體同系物 1	-2.627	-1.996	-2.477	-1.915			
1.076	1.051 -0.5181							
-3.237	-1.786	-0.2226	0.05274	-2.979	0.01185	0.3771	0.2132	0.9695
94	0.7631	0.4243						0.4707
-2.427	-0.2257	0.1713	1.835	0.2063	1.064	-0.05172	0.6024	0.3308
1.666	-1.209						-0.9625	0.3044
212613_at	"丁鹽蛋白，亞族 3，組員 A2"	0.3185	-3.738	-1.563	0.1766	1.357		
0.6323	-3.472	-2.94						
-0.948	-2.16	-0.9575	-2.186	-3.444	0.9747	-0.4156	-0.1155	-2.642
-1.869	-0.5386	-0.05145						-0.3187
0.4147	1.095	0.8142	0.03735	1.204	-0.5544	-2.84	1.699	-0.8163
213245_at	"人類 cDNA FLJ30781 fis，純系 FEBRA2000874，mRNA 序列"	-0.9554	-0.9554					
-0.2654	0.3636 -0.09124							-1.177
0.2931	-0.01804	-0.01614	-0.4445	0.6931	0.1661	0.3127	0.06183	0.3379
4	0.3865	1.577	0.5969					-0.390
-0.228	-0.449	2.261	-0.01159	-0.1097	-0.09017	2.64	1.719	0.0225
2.159	3.16	-0.7841						2.535
-0.1771	0.2709	1.316	-0.02205					
213892_s_at	腺嘌呤轉磷酸核糖基酶	-0.5899	0.858	-1.057	0.1746	1.372		
0.4337	0.6628	-0.6729						
0.1435	-0.8664	0.9885	0.2662	-0.3091	0.342	0.8098	0.5261	0.3701
9	0.5974	-0.7273	0.04012	1.121				-0.9586
1.028	1.872	0.01646	0.3821	-0.3695	-0.267	1.251	0.0724	0.7799
214472_at	"H3 組蛋白族，組員 B"	-0.4576	0.9796	-3.75	0.1232	2.512	2.872	
0.3387	0.1235	1.964	1.03					
0.2118	-3.128	0.6922	1.271	3.664	-0.3823	-2.636	-1.03	0.4435
3	1.02	-0.1162	1.411					1.095
0.0753	0.02653	-0.948	1.585	0.251	2.84	0.02776	0.06318	0.6362
214614_at	相似盒 HB9	0.3303	0.6661	-4.35	0.4488	-0.9459	0.2757	0.05505
2.021	-3.871	0.9336						-2.666
-0.2894	-1.361	1.553	1.158	1.157	0.008503	-2.796	-0.2623	0.4189
								0.2761
								-3.09

- 78 -

218542_at	1.48	0.8353	-3.057	1.639	0.7738	0.8823	-0.3105	-0.2223	0.2337
-2.889	0.9011	0.01222	0.5653	0.3149	0.9069	0.7584	0.9998	-1.406	0.6297
0.4355	-1.843	-0.01727							1.085
-0.6697	0.7572	-0.1458	0.0329	-0.2066	-0.384	0.9562	-1.544	-0.2706	-0.3654
218875_s_at		F盒僅蛋白質5			0.1373	0.4371	-0.2461	-0.004244	0.6533
1.102		-0.5123	2.342						0.6225
-0.7599	0.6561	-0.7147	-0.04562	1.172	0.3008	1.113	0.6165	-0.5506	0.006706
0.7875	0.6787	0.2788							
-0.7371	0.5419	-0.9452	0.9993	-0.3711	0.8086	-0.6675	-0.4852	0.5996	-0.8644
219061_s_at		染色體 X (獨特) 9879 表現序列的 DNA 片段							-2.121
-0.3947	0.6783	1.636	1.591						0.1001
-0.3655	-0.254	0.3603	-0.1915	0.8592	-1.293	0.2111	2.012	0.7631	1.2
1	0.02172	0.1617	1.074						0.6261
-0.05658		0.2486	1.981	1.841	1.55	0.5562	-0.8898	-0.3676	0.07326
-0.0437		-0.3224							0.927
219493_at		小鼠 Shc SH2 域結合蛋白質 1 之可能直向同源基因							0.4476
-0.6428	1.684	0.2368							0.3727
0.003738		1.107	-0.5274	1.644	-0.08562	-1.648	-0.1838	-0.6264	0.3935
8	2.026	-0.4948	-0.5828						0.0276
0.05786	0.282	1.483	0.2727	-1.421	0.4901	-0.9642	1.269	0.1536	-0.1922
4	1.013	-0.3574	-0.8402						-0.1069
-2.008									-0.626
219555_s_at		不典型骨髓蛋白質 BM039							
6	4.215	-0.09475		2.603				-0.2336	2.786
-1.184	3.628	-0.9118	3.769	0.1243	-0.8211	0.06107	0.2847	2.663	3.131
3.039	0.03131	2.418							-0.5802
-0.6508	0.9633	1.56	4.613	-0.07838	3.06	-0.7089	-0.2662	0.3197	-0.9291
-1.049									2.801
219650_at		假想蛋白質 FLJ20105			0.4417	0.7085	-0.9848	-1.671	-0.6439
1.205		-1.405	1.856						0.9707
-1.188	0.8473	-1.826	-0.0391	0.3461	0.911	1.13	1.107	-0.7248	0.7535
8	0.7382	-1.113	1.668						1.691
0.1232	0.8186	0.5107	0.2778	0.2186	-0.4108	-0.2544	-1.474	-0.4748	-0.7022
220085_at		解螺旋酶, 淋巴特定			"	1.098	2.117	-0.3944	-0.1505
2.545	-1.554	2.94							2.926
									2.196

-0.6154	-0.6666	-1.443	0.9973	1.975	3.076	3.464	2.717	-0.6921	-0.3206	-0.6188	2.428
2.27	-1.005	0.9854									
-0.4835	2.768	2.083	-0.9158	-0.3327	-0.8196	3.488	-1.969	-0.8297	-0.0946		
220238_s_at		SBBI 蛋白質		-0.5356	-1.869	0.2827	0.4126	1.225	1.143	0.855	-0.508
7	0.01481	-2.392	-0.7889								
0.06984	-1.5	-0.04597		-0.1081	-0.1832	-0.02539		-4.754	-0.8441	-0.2432	1.218
-0.8203	-4.598	1.751									
-0.1621	0.3076	0.3174	-0.1157	-0.4725	-3.034	-4.345	-3.132	1.094	-0.5442		
221436_s_at		" 富小鼠基因群集之可能直向同源基因, C8 基因 "							-0.0811	1.385	-0.414
5	-1.594	0.7122	1.042								
-0.1976	-2.491	0.9193	-2.686	1.146	0.3385	-0.1127	1.256	0.8854	0.3978	1.538	-0.148
1	0.1661	2.272	0.9145								
0.5171	-2.271	1.161	1.425	1.627	1.09	1.094	0.0009523	-0.1412	0.9679	-0.781	
3	-1.46	-1.893									
221521_s_at		HSPC037 蛋白質		0.5415	0.9347	-0.2726	-1.719	2.054	0.7876	1.801	-0.146
4	2.363	-3.6	1.422	1.45							
-0.04945		0.04416	2.338	2.103	2.45	-1.793	0.377	2.219	0.3689	0.1082	-0.274
1	1.391	1.063	3.74	1.458							
1.522	1.135	0.5133	2.025	0.005946		0.3726	-0.2652				
221539_at		真核翻譯起始因子 4E 結合蛋白質 1								-0.07309	
2.738	-1.713	0.6294									
-0.158	1.168	0.05806	-1.008	-0.1794	-1.094	-0.09421		-0.7066	-0.399	1.699	0.229
3.751	0.7476	-0.6211									
-0.8323	0.4994	-0.2218	-0.5206	0.08783	1.111	4.2	0.9997	0.1046	-1.671	-0.9075	-0.376
2	2.44	-0.2567	0.1609								
-0.2192											
222037_at		" 人類, 純系 IMAGE: 5270727, mRNA, mRNA 序列 "								0.3059	1.569
0.3614	0.8596	1.005	1.397								
0.742	-2.44	1.361	-3.674	1.354	-1.237	0.4627	1.613	0.441	0.8522	1.526	-0.224
8	0.9008	1.122	1.145								
0.03042	0.09205	0.326	0.3462	1.786	-0.4235	1.052	-0.2547	-0.4347	0.1949	0.04239	-0.156
1	-0.5869										

表 S15：該預後組基因之表現資料之平均值(μ)及標準偏差(σ)的加權表決參數

探針_ID	基因名稱	低-NPI		高-NPI	
		平均值	SD	平均值	SD
213892_s_at	腺嘌呤轉磷酸核糖基酶	-0.4139	0.419865	0.5261	0.5756
212141_at	MCM4微染色體保持缺乏4(S.壓縮啤酒釀母)	0.05549	1.527753	1.012	0.771858
204603_at	核酸外切酶1	-0.7394	0.414899	0.3089	0.546392
211456_x_at	類似於金屬硫蛋白1H的蛋白質[人類], mRNA序列	-2.313	1.10771	-0.01816	1.061529
222037_at	人類, 純系 IMAGE:5270727, mRNA, mRNA序列	-0.2248	1.360941	0.8596	0.648812
218447_at	DC 13蛋白質	-0.7617	0.497934	0.3587	0.655529
221521_s_at	HSPC037蛋白質	-0.04945	1.328055	1.422	1.13546
200853_at	H2A組蛋白族, 組員2	-0.2015	0.437181	0.7502	0.667011
203764_at	圓盤, 大同系物7(果蠅)	-0.518	0.626375	0.3234	0.711794
217165_x_at	RNA解螺旋酶相關蛋白質[人類], mRNA序列	-1.315	1.126665	0.4527	1.042786
204444_at	驅動蛋白類似物1	-0.7489	0.817308	0.6377	0.760632
210052_s_at	染色體20開啟讀取框架1	-0.3447	0.713083	0.7286	0.785951
202188_at	KIAA0095基因產物	-1.065	0.858421	1.178	1.616733
220085_at	解螺旋酶, 淋巴特定	-0.6154	1.198542	2.083	1.619802
214614_at	相似盒HB9	-2.666	1.462508	0.2757	1.583945
219061_s_at	染色體X(獨特)9879表現之序列上的DNA片斷	-0.1915	0.461491	0.6783	0.795975
203362_s_at	MAD2有絲分裂阻止缺乏類似物1(酵母菌)	-0.7681	0.74839	0.6176	0.842842
221539_at	真核翻譯起始因子4E結合蛋白質1	-0.6211	0.442172	0.229	1.408505
201487_at	組織蛋白酶C	-0.7759	0.729779	0.4309	0.950128
208546_x_at	H2B組蛋白族, 組員J	0.4872	1.894009	0.9474	1.009994
209040_s_at	蛋白質降解體(前體, 類巨蛋白酶(macropain))亞單位, β 型, 8(大的多功能蛋白酶7)	-0.7578	1.8346	0.588	1.159099
219650_at	假想蛋白質FLJ20105	-0.7248	0.85837	0.5107	0.893847
218542_at	染色體10開啟讀取框架3	-0.3654	1.305871	0.7584	0.82541
219555_s_at	不典型骨髓蛋白質BM039	-0.5802	1.164774	1.56	1.763962
221436_s_at	富小鼠基因群集之可能直向同源基因, C8基因	-0.1481	1.137308	0.9679	1.10724

210559_s_at	細胞分類循環2，G1至S及G2至M	-0.2508	0.844298	0.7038	0.805354
212185_x_at	金屬硫蛋白2A	-1.284	0.725732	0.1074	0.798804
218350_s_at	類比蛋白酶(geminin)，DNA複寫抑制因子	-0.9141	0.51298	-0.06399	0.926376
208433_s_at	低密度脂蛋白受體相關蛋白質8，載脂蛋白e受體	-1.55	1.219961	-0.2532	1.04719
217755_at	血液學及神經學表現之1	-0.1708	0.614723	0.4835	0.951001
209398_at	H1組蛋白族，組員2	-0.02843	1.093238	1.332	1.299819
204766_s_at	nudix(核苷二磷酸連接之部分X)類型模體1	-1.462	1.152307	0.6079	1.516876
208581_x_at	金屬硫蛋白1X	-1.11	0.696985	0.1739	0.997649
209806_at	H2B組蛋白族，組員T	-0.3533	0.961244	0.5906	0.913624
209114_at	Tetraspan 1	-0.4002	1.24355	0.923	1.133855
206461_x_at	金屬硫蛋白1H	-0.7782	1.051675	0.1177	0.916536
206110_at	H3組蛋白族，組員K	-0.3704	1.40578	0.3631	1.411458
201890_at	核糖核苷酸還原酶M2多肽	-0.8654	1.559316	0.3715	1.024143
202095_s_at	包含桿狀病毒IAP重複之5(生存蛋白質)	-0.3761	1.515513	0.6679	1.21519
218875_s_at	F盒僅蛋白質5	-0.5123	0.409105	0.6165	0.900364
202833_s_at	絲胺酸(或半胱胺酸)蛋白酶抑制因子，進化枝A(α -1抗蛋白酶，抗胰蛋白酶)，組員1	-0.7663	1.176481	0.5393	1.901084
208767_s_at	溶菌酶關聯之蛋白質橫跨膜4 β	-0.5525	0.938047	0.5525	1.011665
203687_at	向化性激動素(C-X3-C模體)配位子1	-2.375	1.081471	-1.073	1.154088
210792_x_at	CD27-結合(Siva)蛋白質	-0.4151	0.800901	0.3786	1.230555
205240_at	LGN蛋白質	-1.249	1.72051	1.297	1.446916
212484_at	小鼠乳房腫瘤病毒受體同系物1	-0.3862	1.394896	0.2132	1.187908
202580_x_at	叉頭盒M1	-0.4973	1.022497	0.3564	1.104339
203510_at	met原癌基因(肝細胞生長因子受體)	-2.988	1.352621	-0.736	2.009295
212613_at	丁醯蛋白，亞族3，組員A2	-1.563	1.383434	0.1766	1.475442
220238_s_at	SBBI26蛋白質	-0.8441	1.574483	-0.04597	1.556341
219493_at	小鼠Shc SH2域結合蛋白質1之可能直向同源基因	-0.5274	0.594225	0.282	1.007135
214472_at	H3組蛋白族，組員B	0.1235	1.581567	0.8844	1.40927
204623_at	三葉草因子3(腸內)	0.2033	1.408904	1.662	1.554202
215214_at	免疫球蛋白 λ 基因座	-0.6629	2.409822	-0.107	2.500735

209832_s_at	DNA複寫因子	-0.4351	0.674077	0.5995	1.153719
213245_at	人類cDNA FLJ30781 fis， 純系FEBRA2000874， mRNA序列	-0.02205	0.369593	0.3127	1.16657
209924_at	向化性激動素(C-C模體) 配位子18(肺部及活化調節)	-0.8797	1.267438	0.003248	1.311969
214768_x_at	免疫球蛋白□常量	-1.158	1.997589	0.1494	2.246666
201483_s_at	Ty 4同系物1之抑制因子 (S.壓縮啤酒釀母)	-0.0874	0.541135	0.7686	1.030094
212094_at	父方表現之10	-2.245	1.918298	0.03678	2.405576
201236_s_at	BTG族，組員2	1.328	0.70948	0.2717	0.438693
210576_at	細胞色素P450，亞族IVF， 多肽8	3.704	3.447008	-0.6011	0.891116

表L1：預後組基因之ID的查找表格

NPI-ES

探針_ID	GenBank	單基因
200853_at	NM_002106.1	Hs. 119192
201483_s_at	BC002802.1	Hs.79058
201487_at	NM_001814.1	Hs.10029
201890_at	NMJ01034.1	Hs.75319
202095_s_at	NM_001168.1	Hs.1578
202188_at	NM_014669.1	Hs.155314
202580_x_at	NM_021953.1	Hs.239
202833_s_at	NM_000295.1	Hs.297681
203362_s_at	NM_002358.2	Hs.79078
203510_at	BG170541	Hs.316752
203687_at	NM_002996.1	Hs.80420
203764_at	NM_014750.1	Hs.77695

204444_at	NM_004523.2	Hs.8878
204603_at	NM_003686.1	Hs.47504
204623_at	NM_003226.1	Hs.82961
204766_s_at	NM_002452.1	Hs.388
205240_at	NM_013296.1	Hs.278338
206110_at	NM_003536.1	Hs.70937
206461_x_at	NM_005951.1	Hs.2667
208433_s_at	NM_017522.1	Hs.54481
208546_x_at	NM_003524.1	Hs.249216
208581_x_at	NM_005952.1	Hs.374950
208767_s_at	AW149681	Hs.296398
209040_s_at	U17496.1	Hs.180062
209114_at	AF133425.1	Hs.38972
209398_at	BC002649.1	Hs.7644
209806_at	BC000893.1	Hs.247817
209832_s_at	AF321125.1	Hs.122908
209924_at	AB000221.1	Hs. 16530
210052_s_at	AF098158.1	Hs.9329
210559_s_at	D88357.1	Hs.334562
210792_x_at	AF033111.1	Hs.112058
211456_X_at	AF333388.1	Hs.367850
212094_at	BE858180	Hs.137476
212141_at	X74794.1	Hs.154443
212185_x_at	NM_005953.1	Hs.118786

212484_at	BF974389	Hs.18686
212613_at	AI991252	Hs.87497
213245_at	AL120173	Hs.301663
213892_S_at	AA927724	Hs.28914
214472_at	NM_003530.1	Hs.143042
214614_at	AI738662	Hs.37035
214768_x_at	BG540628	Hs.406565
215214_at	H53689	Hs.405944
217165_x_at	M10943	Hs.381097
217755_at	NM_016185.1	Hs.109706
218350_s_at	NM_015895.1	Hs.234896
218447_at	NM_020188.1	Hs.6879
218542_at	NM_018131.1	Hs.14559
218875_s_at	NM_012177.1	Hs.272027
219061_s_at	NM_006014.1	Hs.18212
219493_at	NM_024745.1	Hs.123253
219555_s_at	NM_018455.1	Hs.283532
219650_at	NM_017669.1	Hs.89306
220085_at	NMJJ18063.1	Hs.203963
220238_s_at	NM__018846.1	Hs.26481
221436_s_at	NM_031299.1	Hs.30114
221521_s_at	BC003186.1	Hs.433180
221539_at	AB044548.1	Hs.433317
222037_at	AI859865	Hs.319215

201236_s_at NM_006763.1 Hs.75462

210576_at AF133298.1 Hs.268554

【圖式簡單說明】

圖1展示了藉由整體表現輪廓之偶發(sporadic)乳房腫瘤之群集a)使用能展示基因表現中最高變化之最前面376個基因的98個乳房腫瘤之未監督分級群集，b)使用該376個基因之組的主組份分析(PCA)。類似分子成群可在如a)中觀察到，c)使用SAM-409基因組的樣品之分級群集，其由在乳癌亞型間被顯著調節之基因組成。SAM-409基因組中大約三分之二的基因在ER+腫瘤中展示增加之表現。

圖2展示了與NPI相關聯之表現簽名(NPI-ES)的識別：

a)使用移動NPI臨限值來測定差別表現基因。可展示顯著差別表現之基因(y軸)於各個臨限值(x軸)處得以識別。使用4次傳遞差別調節基因之最高數目的臨限值，

b)使用NPI-ES的ER+樣品之分級群集。紅條指示了低NPI(<4)之樣品；而藍條指示了高NPI之樣品，

c)使用NPI-ES的ER+腫瘤樣品之分類及預測置信度。樣品根據其NPI值(X軸)而得以分類。使用加權表決以分類該等樣品及基於Golub等人(13)計算得出之各個樣品(Y軸)的預測力。預測力<0.3之樣品分類被看作"不確定"或"低置信度"(灰色區域)。

圖3展示了比較ER+腫瘤上不同分類方案之預後力的KM存活分析。綠線代表(a)低NPI、(b)低NPIES表現水平、或(c)低"預後"簽名(PES)表現水平，而粉線代表高水平。(a)藉由

經典NPI而被分層為"良好"預後($NPI < 3.4$)(35個腫瘤)及"中等"預後($NPI > 3.4$)(14個腫瘤)群之49個Rosetta ER+腫瘤。(b)藉由NPI-ES而被分層為能表現NPI-ES之高(24個腫瘤)對比低水平(25個腫瘤)之群的相同49個Rosetta ER+腫瘤。(c)藉由70個基因"預後"簽名而被分別分層為"良好預後"群(27個腫瘤)對比"差預後"群(22個基因)之相同49個Rosetta ER+腫瘤。(d)藉由NPI-ES而被分層為能表現NPI-ES之高(13個腫瘤)對比低(33個腫瘤)水平之群的46個Stanford ER+腫瘤。

圖4展示了基於全部腫瘤(而不管亞型)使用44個基因之組的腫瘤樣品之分類及預測置信度。

圖5展示了來自Rosetta資料集之基因表現資料的分級群集。頂部)顯示腫瘤間類似性之樹狀圖。彩色編碼條將亞型指示給對應基因簽名。左側)具有三種區別性基因群集之276個基因的完全群集。注意某些ERBB2腫瘤似乎與ER+腫瘤(紅條)隔離，但經仔細檢查與ERBB2+相關之基因的表現(群集團clustergram之擴增)其被識別為ERBB2+。此歸因於Rosetta微陣列擁有比ERBB2亞型高得多的與ER+亞型相關之基因的數目。

圖6展示了基於NPI-ES之表現水平的Rosetta ER+樣品(49)之分級群集(62個基因中在Rosetta資料發現了46個配對)。該彩色條如圖2b中所定義。

圖7展示了Stanford乳房腫瘤之群集。頂部)顯示腫瘤間類似性之樹狀圖。彩色編碼條將亞型指示給對應之基因簽名。左側)具有三種區別性基因群集之136個基因的完全群

集。

圖8展示了使用NPI-ES的Stanford46個ER+樣品之分級群集(62個基因中31個配對)。彩色條如圖2b)所定義。

圖9展示了ER-及ERBB2+分子亞型中NPI-ES表現與NPI狀態之間的關係。ER-及ERBB2腫瘤之NPI狀態通常高於ER+腫瘤。不同於ER+腫瘤之情況，吾人不能夠藉由在ER-及ERBB2+亞型之高對比低NPI腫瘤內經差別調節之SAM基因來識別。並且，NPI-ES似乎亦不與其它分子亞型所關聯之NPI值相關聯。

圖10展示了20對樣品，其於14週阿黴素治療(Perou等人，2000)"之前"及"之後"獲得。在20個"之前"樣品中，10個樣品展示出高NPI-ES表現水平(H)，且10個展示出低表現水平(L)。在該等前10個樣品中，6個在化學治療後保持高表現水平(H->H，以紅色描述)，而4個在治療後展示出低表現水平(H->L，以黃色描述)。

圖11展示出Kaplan-Meier無復發存活分析曲線，其使用捐獻圖10之20個樣品的病人。

五、中文發明摘要：

本發明者闡明識別一組用作與諾丁漢預後指數(NPI)相關聯之乳房腫瘤的預後標示物的基因。最初其不能識別其表現與NPI相關之單一基因組。然而，在將該資料集分離成分子子範疇(雌激素受體陽性、雌激素受體陰性及ErbB2陽性)後，其可識別一組在具有不同預後之腫瘤中差別表現的62個基因。提供了一種用以測定預後之方法及裝置。亦提供了用以測定腫瘤對化學療法之反應的方法，其涉及比較治療前及治療後之預測基因組的表現水平。

六、英文發明摘要：

十、申請專利範圍：

1. 一種測定乳癌病人之該預後的方法，該方法包括基於該病人之乳房腫瘤內一預後組基因之表現水平將一預後指派給該病人，其中該預後組包括來自表S6之至少5個基因。
2. 如請求項1之方法，其中該預後組包括表S6之至少10個、20個、30個、40個、50個、60個或全部基因。
3. 如請求項1或2之方法，其進一步包括測定該腫瘤樣品之雌激素受體(ER)狀態的步驟。
4. 如請求項3之方法，其進一步包括測定該腫瘤樣品之ErbB2狀態。
5. 如請求項1或2之方法，其包括下列步驟：
 - (a) 自該病人獲得一乳房腫瘤樣品；且
 - (b) 量測該樣品內該預後組之該等基因的該等表現水平。
6. 如請求項5之方法，其中步驟(b)包括以複數個結合組員接觸自該樣品獲得之該等表現產物，該等結合組員能夠結合至能指示該預後組之基因之該表現的表現產物，其中可量測此結合。
7. 如請求項6之方法，其中該等結合組員為互補核酸序列或特定抗體。
8. 如請求項1或2之方法，其包括將該乳房腫瘤樣品分類為具有高NPI或低NPI之樣品，或分類為具有良好或不良預後之樣品。
9. 如請求項1或2之方法，其中該指派一預後之步驟藉由將

來自該測試乳房腫瘤樣品之該表現輪廓與先前獲得之輪廓及/或一先前測定之表示一特別預後的標準輪廓進行比較。

10. 如請求項9之方法，其中該等先前獲得之輪廓係作為輪廓之一資料庫而得以儲存。
11. 如請求項1或2之方法，其進一步包括比較治療前及治療後該乳房腫瘤樣品內該預後組之該等表現水平，以偵測能指示一改良預後或惡化預後之該表現輪廓的一改變。
12. 一種將一預後指派給一乳房腫瘤樣品之裝置，該裝置包括一附著有複數個結合組員之載體，各個組員能夠特定及獨立地結合至一預後組基因中的一個基因之一表現產物，其中該預後組包括來自表S6之至少5個基因。
13. 如請求項12之裝置，其中該預後組包括表S6之至少5個、10個、20個、30個、40個、50個、60個或全部基因。
14. 如請求項12或13之裝置，其中該載體已僅附著至能夠特定及獨立地結合至表S6中所識別之該等基因之表現產物的結合組員。
15. 如請求項12或13之裝置，其包括一核酸微陣列，其中該等結合組員為核酸序列。
16. 一種將一預後指派給一乳癌病人之套組，該套組包括能夠特定結合至一預後組基因之基因的表現產物的複數個結合組員及一偵測試劑，其中該預後組包括來自表S6之至少5個基因。
17. 如請求項16之套組，其中該預後組包括表S6之至少10

- 個、20個、30個、40個、50個、60個或全部基因。
18. 如請求項16或17之套組，其進一步包括一資料分析工具，其中該資料分析工具為一電腦程式。
 19. 如請求項18之套組，其中該資料分析工具包括一演算法，該演算法經調適以辨別具有不同預後之腫瘤的該等表現輪廓。
 20. 如請求項16或17之套組，其包括來自具有已知預後之乳房腫瘤樣品的表現輪廓及/或表示一特別預後之特徵的表現輪廓。
 21. 如請求項16或17之套組，其包括如請求項12至15中任一項之裝置。
 22. 如請求項16或17之套組，其包括能夠結合至該預後組之該等基因之該等表現產物的核苷酸引子，使得其可於一PCR中被擴增。
 23. 一種產生一乳房腫瘤樣品之一核酸表現輪廓的方法，其包括步驟
 - (a) 自該乳房腫瘤樣品隔離表現產物；
 - (b) 識別一預後組基因之該等表現水平，其中該預後組基因包括來自表S6之至少5個基因；且
 - (c) 自該等表現水平產生該乳房腫瘤樣品之一表現輪廓。
 24. 如請求項23之方法，其中該預後組包括表S6之至少10個、20個、30個、40個、50個、60個或全部基因。
 25. 如請求項23或24之方法，其包括將該表現輪廓添加至一基因表現輪廓資料庫。

26. 如請求項23或24之方法，其進一步包括將該表現輪廓與表示一特別預後之特徵的一第二表現輪廓或複數個第二表現輪廓進行比較。
27. 如請求項26之方法，其包括下列步驟：
 - (a) 自一第一乳房腫瘤樣品隔離表現產物；以能夠特定及獨立地結合至該預後組之表現產物的複數個結合組員接觸該等表現產物；且自該腫瘤樣品內該預後組之該等表現水平產生一第一表現輪廓；
 - (b) 自一具有已知預後之第二乳房腫瘤樣品隔離表現產物；以能夠特定及獨立地結合至步驟(a)之該預後組之表現產物的複數個結合組員接觸該等表現產物，以便產生一乳房腫瘤樣品之一可比較的第二表現輪廓；
 - (c) 比較該第一表現輪廓及該第二表現輪廓以測定該第一乳房腫瘤樣品之該預後。
28. 一種表現輪廓資料庫，其包括乳房腫瘤樣品之複數個基因表現輪廓，其中該等基因表現輪廓獲自一預後組基因之該等表現水平，其中該預後組基因包括來自表S6之至少5個基因，該資料庫可擷取地保持於一資料載體上。
29. 如請求項28之表現輪廓資料庫，其中該預後組包括表S6之至少10個、20個、30個、40個、50個、60個或全部基因。
30. 如請求項28或29之表現輪廓資料庫，其中該等表現輪廓為核酸表現輪廓。
31. 如請求項28或29之表現輪廓資料庫，其中該等表現輪廓

係根據該腫瘤源之ER狀態而得以分類。

32. 一種識別於一群腫瘤內差別表現之一組基因的方法，該方法包括提供一來自該群之複數個腫瘤之各個腫瘤的表現輪廓，根據腫瘤之分子亞型將該等輪廓分類，及分析一亞型內之表現輪廓以識別一區別基因組，其中該區別組之該等基因差別表現於彼亞型內。
33. 如請求項32之方法，其進一步包括將一類指派給來自一病人之一腫瘤樣品的步驟，其中該區別組之基因的差別表現表示該類之特徵，該等步驟包括提供該區別組之該樣品內之表現水平，及基於該等表現水平將一類指派給該腫瘤。
34. 如請求項32或33之方法，其包括測定一腫瘤樣品內該區別組之該等基因之該等表現水平、自該等表現水平測定一表現輪廓及將該輪廓添加至一資料庫的步驟。
35. 如請求項32或33之方法，其中該腫瘤樣品之該分子亞型亦被識別且添加至該資料庫。
36. 如請求項32或33之方法，其包括提供來自該腫瘤之不同治療階段的表現輪廓且該等表現輪廓用以測定預後類之一改變，其中該等表現輪廓係獲自該區別組之基因的該等表現水平。
37. 如請求項32或33之方法，其中該等腫瘤為乳房腫瘤且該分子亞型對應於該腫瘤之ER狀態。

七、指定代表圖：

(一)本案指定代表圖為：第(3)圖。

(二)本代表圖之元件符號簡單說明：

(無元件代表符號)

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)