I532843

(19)中華民國智慧財產局

(12)發明說明書公告本 (11)證書號數：TW I532843 B

(45)公告日： 中華民國 105 (2016) 年 05 月 11 日

(21)申請案號：100144098 (22)申請日： 中華民國 100 (2011) 年 11 月 30 日

(51)Int. Cl.： *C12Q1/68 (2006.01)* *G06F19/18 (2011.01)* *G06F19/20 (2011.01)*

(30)優先權：2010/11/30 美國 61/418,391

2011/08/31 美國 61/529,877

(71)申請人：香港中文大學 (香港地區) THE CHINESE UNIVERSITY OF HONG KONG (HK) 香港

(72)發明人：盧煜明 LO, YUK MING DENNIS (GB)；陳君賜 CHAN, KWAN CHEE (HK)；趙慧君 CHIU, WAI KWUN ROSSA (AU)；江培勇 JIANG, PEIYONG (CN)

(74)代理人：陳長文

(56)參考文獻：

2001 年 08 月 01 日 Taback, B., et al.,"Prognostic Significance of Circulating Microsatellite Markers in the Plasma of Melanoma Patients", Cancer Res., 2001, Vol.61, Issue 15, P.5723-5726.

2005 年 11 月 08 日 Diehl, F., et al.,"Detection and Quantification of Mutations in the Plasma of Patients with Colorectal Tumors", Proc. Natl. Acad. Sci. USA., 2005, Vol.102, No.45, P.16368-16373.

審查人員：林桂滿

申請專利範圍項數：34 項 圖式數：29 共 119 頁

(54)名稱

與癌症有關之基因或分子變異之檢測

DETECTION OF GENETIC OR MOLECULAR ABERRATIONS ASSOCIATED WITH CANCER

(57)摘要

本發明提供測定有機體之生物學試樣中之基因或分子變異的系統、裝置及方法。分析包括無細胞之 DNA 片段之生物學試樣以鑑別腫瘤中之染色體區域中之不均衡，例如因缺失及/或擴增所致。對於各染色體區域，使用多個基因座。然後可使用此等不均衡來診斷(篩選)癌症患者，以及預測癌症患者，或檢測患者癌變前病況之存在，或監測患者癌變前病況之進展。可使用不均衡之嚴重程度以及展現不均衡之區域之數目。基因組之非重疊區段之系統分析可提供試樣之一般篩選工具。另外，可隨時間測試患者以追蹤一或多個染色體區域之每一者之嚴重程度及染色體區域之數目以能夠篩選及預測以及監測進展(例如在治療後)。

Systems, apparatus, and methods are provided for determining genetic or molecular aberrations in a biological sample from an organism. Biological samples including cell-free DNA fragments are analyzed to identify imbalances in chromosomal regions, e.g., due to deletions and/or amplifications in a tumor. Multiple loci are used for each chromosomal region. Such imbalances can then be used to diagnose (screen) a patient for cancer, as well as prognosticate a patient with cancer, or to detect the presence or to monitor the progress of a prematignant condition in a patient. The severity of an imbalance as well as the number of regions exhibiting an imbalance can be used. A systematic analysis of non-overlapping segments of a genome

can provide a general screening tool for a sample. Additionally, a patient can be tested over time to track severity of each of one or more chromosomal regions and a number of chromosomal regions to enable screening and prognosticating, as well as monitoring of progress (e.g. after treatment)

指定代表圖：

符號簡單說明：

(無元件符號說明)

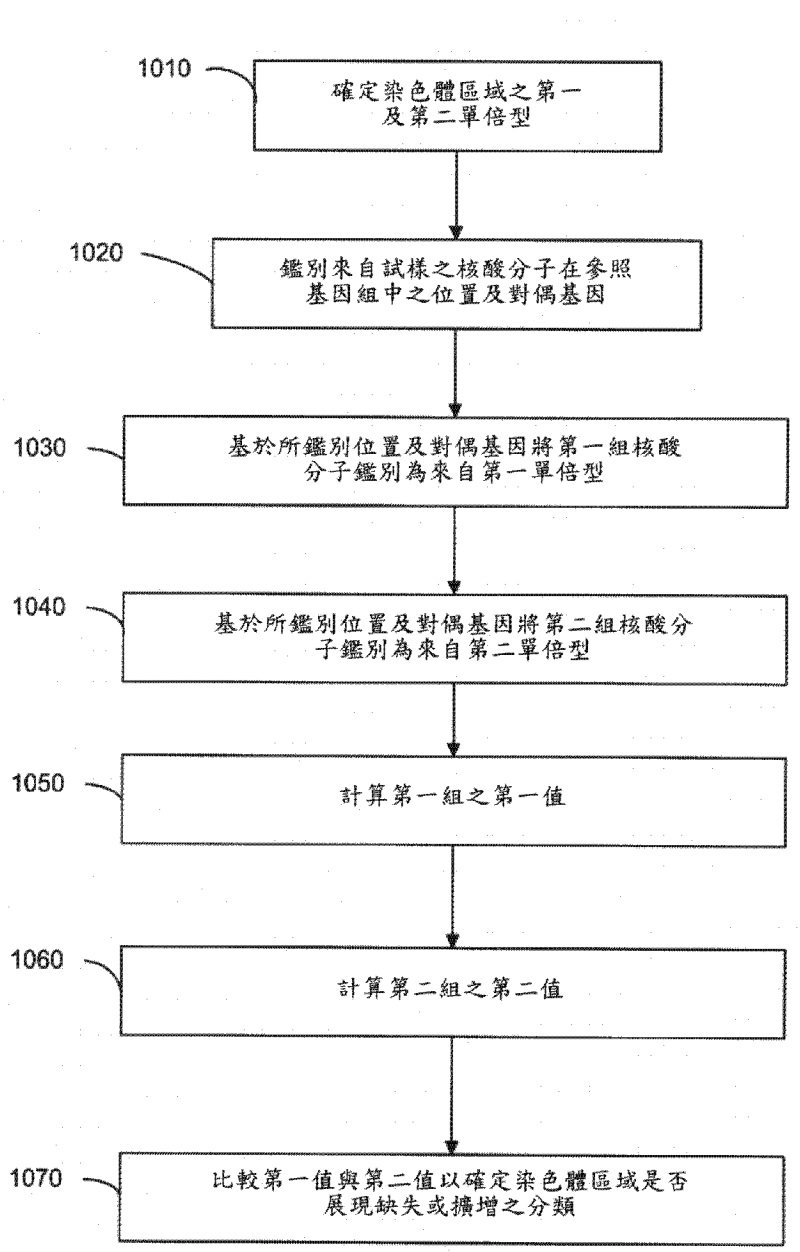

圖10

六、發明說明：

本申請案主張優先於2010年11月30日申請之標題為「DETECTION OF GENETIC ABERRATIONS ASSOCIATED WITH CANCER」之美國臨時申請案第61/418,391號及2011年8月31日申請之標題為「DETECTION OF GENETIC OR MOLECULAR ABERRATIONS ASSOCIATED WITH CANCER」之美國臨時申請案第61/529,877號，且本申請案係該等臨時申請案之非臨時申請案，該等案件之全部內容出於所有目的以引用方式併入本文中。

本申請案係關於共同擁有之頒予Lo等人之標題為「Size-Based Genomic Analysis」之美國專利申請案第12/940,992號(美國公開案2011/0276277)(代理人案號80015-794101/006610US，2010年11月5日申請)及頒予Lo等人之標題為「Fetal Genomic Analysis From A Maternal Biological Sample」之美國專利公開案第12/940,993號(美國公開案2011/0105353)(代理人案號80015-794103/006710US，2010年11月5日申請)，該等案件之揭示內容之全文以引用方式併入本文中。

【先前技術】

癌症係影響許多人之常見疾病。經常直到表現嚴重症狀時才鑑別出癌症。對於常見癌症類型而言，存在鑑別可能患有癌症之患者之篩選技術。但此等技術經常不可靠或使患者經受輻射。對於許多其他癌症類型而言，尚無有效篩選技術。

已在罹患肺癌及頭頸癌患者之循環DNA中之特定基因座檢測出異型接合性喪失(LOH)(Chen XQ等人，Nat Med 1996; 2: 1033-5；Nawroz H等人Nat Med 1996; 2: 1035-7)。然而，可藉助檢查特定基因座檢測之LOH相對少量已阻礙此等技術。甚至在使用數位PCR時，該等方法仍會遭遇到不能檢測少量LOH。此外，此等技術仍受限於對已知出現於具體癌症類型中之特定基因座之研究。因此，通常尚不可能或有效地對癌症進行篩選。

除對癌症之存在進行篩選外，現有技術亦缺乏提供對癌症患者之預後及監測治療效應(例如手術或化學療法或免疫療法或靶向療法後之恢復)。此等技術經常花費多(例如成像技術)、不精確、效率低、不靈敏或可能使患者經受用於成像技術之輻射。

因此，人們期望提供針對癌症篩選、預測及監測患者之新穎技術。

【發明內容】

實施例提供確定與癌症有關之基因變異之系統、裝置及方法。分析包括無細胞DNA片段之生物學試樣以鑑別腫瘤中之染色體區域中因(例如)缺失及/或擴增所致之不均衡。使用具有多個基因座之染色體區域可達成較大效率及/或精度。然後可使用此等不均衡來針對癌症診斷或篩選患者以及預測癌症患者。可使用不均衡之嚴重程度以及展現不均衡之區域之數目。另外，可隨時間測試患者以追蹤一或多個染色體區域中之每一者之嚴重程度及染色體區域之數

目以使得能夠篩選及預測以及監測進展(例如在治療後)。

根據一個實施例中，提供分析有機體之生物學試樣之與癌症有關之染色體缺失或擴增的方法。生物學試樣包括源自正常細胞及可能源自與癌症有關之細胞之核酸分子。至少一些核酸分子在試樣中係無細胞的。確定有機體之正常細胞於第一染色體區域之第一及第二單倍型。第一染色體區域包括第一複數個異型接合基因座。試樣中複數個核酸分子中每一者在所鑑別有機體之參照基因組中具有位置且具有所確定各別對偶基因。使用位置及所確定對偶基因自第一單倍型確定第一組核酸分子且自第二單倍型確定第二組核酸分子。電腦系統計算第一組之第一值及第二組之第二值。每一值界定各別組核酸分子之性質(例如該組中分子之平均大小或數目)。比較第一值與第二值以確定第一染色體區域在與癌症有關之任何細胞中是否展現缺失或擴增之分類。

根據另一實施例，提供分析有機體之生物學試樣之方法。生物學試樣包括源自正常細胞及可能源自與癌症有關之細胞之核酸分子。至少一些核酸分子在試樣中係無細胞的。鑑別有機體之複數個非重疊染色體區域。每一染色體區域包括複數個基因座。試樣中複數個核酸分子中每一者在所鑑別有機體之參照基因組中具有位置。對於每一染色體區域而言，基於所鑑別位置將各別組核酸分子鑑別為來自染色體區域。每一各別組包括至少一個位於染色體區域之複數個基因座中每一者處之核酸分子。電腦系統計算各

別組之各別值，其中各別值界定各別組核酸分子之性質。比較各別值與參照值以確定染色體區域是否展現缺失或擴增之分類。然後確定分類為展現缺失或擴增之染色體區域之量。

根據另一實施例，提供使用包括源自正常細胞及可能源自與癌症有關之細胞之核酸分子的生物學試樣來確定有機體中染色體變異之進展的方法。核酸分子中之至少一些在生物學試樣中係無細胞的。鑑別有機體之參照基因組之一或多個非重疊染色體區域。每一染色體區域包括複數個基因座。分析於不同時間取自有機體之試樣以確定進展。對於試樣而言，試樣中複數個核酸分子中之每一者在所鑑別有機體之參照基因組中具有位置。對於每一染色體區域而言，基於所鑑別位置將各別組核酸分子鑑別為來自染色體區域。各別組包括至少一個位於染色體區域之複數個基因座中每一者處之核酸分子。電腦系統計算各別組核酸分子之各別值。各別值界定各別組核酸分子之性質。比較各別值與參照值以確定第一染色體區域是否展現缺失或擴增之分類。然後，使用於複數個時刻對每一染色體區域之分類來確定有機體中染色體變異之進展。

本發明之其他實施例係關於與本文所述方法有關之系統、便攜式消費型器件及電腦可讀取媒體。

可參照以下實施方式及附圖獲得對本發明之性質及優點之較佳理解。

【實施方式】

5

定義

本文所用術語「*生物學試樣*」係指取自個體(例如，人類，患有癌症者、疑似患有癌症者，或其他有機體)且含有一或多個所關注核酸分子之任何試樣。

術語「*核酸*」或「*多核苷酸*」係指去氧核糖核酸(DNA)或核糖核酸(RNA)及其呈單鏈或雙鏈形式之聚合物。除非明確限制，否則該術語涵蓋含有天然核苷酸之已知類似物的核酸，該等類似物與參照核酸具有類似結合特性且以類似於天然核苷酸之方式代謝。除非另有說明，否則特定核酸序列亦隱含地涵蓋其保守修飾變體(例如，簡並密碼子取代)、對偶基因、直向同源物、單核苷酸多態性(SNP)、拷貝數變體及互補序列以及明確指出之序列。具體而言，可藉由產生一或多個所選(或全部)密碼子之第三位經混合鹼基及/或去氧次黃嘌呤核苷殘基取代的序列達成簡並密碼子取代(Batzer等人，*Nucleic Acid Res.* 19:5081 (1991)；Ohtsuka等人，*J. Biol. Chem.* 260:2605-2608 (1985)；及Rossolini等人，*Mol. Cell. Probes* 8:91-98 (1994))。術語核酸涵蓋(但不限於)：基因、cDNA、mRNA、非編碼小RNA、微型RNA (miRNA)、Piwi-相互作用RNA及由基因或基因座編碼之短髮夾RNA (shRNA)。

術語「*基因*」意指參與產生多肽鏈或轉錄RNA產物之DNA區段。其可包括在編碼區域之前及之後之區域(前導區域及尾隨區域)以及在個別編碼區段(外顯子)之間之間插序列(內含子)。

本文所用術語「*臨床相關核酸序列*」或「*臨床相關染色體區域*」(或所測試之區域/區段)可指多核苷酸序列，其對應於較大基因組序列中正測試可能的不均衡之區段或該較大基因組序列本身。實例包括經缺失或擴增(包括簡單倍增)或可能經缺失或擴增之基因組區段、或包括該區段之亞區之較大區域。在一些實施例中，可使用多個臨床相關核酸序列或臨床相關核酸序列之多個等效標記物來提供檢測區域中不均衡之數據。舉例而言，來自染色體上5個非連續序列之數據可以相加方式用來確定可能的不均衡，從而將所需試樣容積有效地減至1/5。

本文所用術語「*參照核酸序列*」或「*參照染色體區域*」係指定量特徵或大小分佈用於與測試區域進行比較之核酸序列。參照核酸序列之實例包括不包括缺失或擴增之染色體區域、整個基因組(例如經由總經測序標籤計數正規化)、來自一或多個已知正常之試樣之區域(其可為所測試試樣之相同區域)或特定單倍型之染色體區域。此等參照核酸序列可以內源性方式存在於試樣中，或在試樣處理或分析期間以外源性方式添加。在一些實施例中，參照染色體區域顯示代表無病健康狀態之大小特徵。在再一實施例中，參照染色體區域顯示代表無病健康狀態之定量特徵。

本文所用術語「*基於*」意指「*至少部分地基於*」且係指一個用於確定另一值(或結果)之值，例如在方法之輸入與該方法之輸出之關係中出現。本文所用術語「導出」亦係指方法之輸入與該方法之輸出之關係，例如在導出係計算

公式時出現。

本文所用術語「*參數*」意指表徵定量數據集及/或定量數據集間之數值關係之數值。舉例而言，第一核酸序列之第一量與第二核酸序列之第二量間之比率(或比率之函數)係一參數。

本文所用術語「*基因座*(*locus*)」或其複數形式「*基因座*(*loci*)」係可在基因組間具有差異之任一長度之核苷酸(或鹼基對)之位置或地址。

本文所用術語「*序列不均衡*」或「*變異*」意指臨床相關染色體區域之量與參照量間如藉由至少一個截止值界定之任一顯著偏差。序列不均衡可包括染色體劑量不均衡、對偶基因不均衡、突變劑量不均衡、拷貝數不均衡、單倍型劑量不均衡及其他類似不均衡。作為一實例，對偶基因不均衡可出現於腫瘤在其基因組內缺失基因之一個對偶基因或擴增基因之一個對偶基因或差式擴增兩個對偶基因，由此在試樣中特定基因座處產生不均衡時。作為另一實例，患者可在腫瘤抑制基因內具有遺傳突變。然後患者可繼續發生腫瘤，其中缺失腫瘤抑制基因之非突變對偶基因。因此，腫瘤內存在突變劑量不均衡。當腫瘤將其DNA釋放至患者血漿中時，腫瘤DNA將在血漿中與患者之組成型DNA(來自正常細胞)混合。藉助使用本文所述方法，可在血漿中檢測出此DNA混合物之突變劑量不均衡。

本文所用術語「*單倍型*」係指在相同染色體或染色體區域上共同傳遞之多個基因座上之對偶基因的組合。單倍型

可指少至一對基因座或指染色體區域或指整個染色體。術語「對偶基因」係指相同物理基因組基因座上可產生或可不產生不同表型性狀之替代性DNA序列。在各染色體具有兩個拷貝(雄性人類個體中性染色體除外)之任一特定二倍體有機體中，各基因之基因型皆包含存在於該基因座上在同型合子中相同且在異型合子中不同之對偶基因對。在不同個體間，一組或一種有機體通常在各基因座上包括複對偶基因。將在群中發現一個以上對偶基因之基因組基因座稱為多態位點。基因座上之對偶基因差異可以群中存在之對偶基因數目(即，多態性程度)或異型合子之比例(即，異型接合性比率)來量測。本文所用術語「多態性」係指人類基因組內與其頻率無關之任一個體間差異。此等差異之實例包括(但不限於)單核苷酸多態性、簡單串聯重複多態性、插入-缺失多態性、突變(其可為致病性)及拷貝數差異。

術語「*經測序標籤*」係指自全部或一部分核酸分子(例如，DNA片段)確定之序列。經常，僅對片段一端實施測序，例如，約30 bp。然後可將經測序標籤與參照基因組比對。另一選擇為，可對片段兩端實施測序以產生兩個經測序標籤，該等標籤可提供較大比對精度且亦提供一定長度之片段。

術語「*通用測序*」係指將銜接體添加至片段末端，且將測序用引物附接至銜接體之測序。因此，可利用相同引物對任何片段測序，且因此該測序可係隨機的。

術語「*大小分佈*」係指代表對應於特定組(例如特定單倍型或特定染色體區域之片段)之分子之長度、質量、重量或其他大小量度的任一值或值集。各實施例可使用多種大小分佈。在一些實施例中，大小分佈係指一個染色體之片段相對於其他染色體之片段之大小的分級(例如，平均值(average)、中值(median)或均值(mean))。在其他實施例中，大小分佈可指染色體片段之實際大小的統計值。在一實作中，統計值可包括染色體片段之任一平均值、均值或中值大小。在另一實作中，統計值可包括低於截止值之片段之總長度(可將其除以全部片段之總長度)或至少低於較大截止值之片段之總長度。

本文所用術語「*分類*」係指與試樣之特定性質有關之任何數值或其他符號。舉例而言，「+」符號(或詞語「正」)可表示將試樣分類為具有缺失或擴增。分類可為二元(例如，正或負)或具有較多分類級別(例如，1至10或0至1之量表)。術語「*截止值*」及「*閾值*」係指操作中所用之預定數值。舉例而言，截止大小可指在高於時排除片段之大小。閾值可為在高於或低於時特定分類適用之值。任一該等術語中皆可用於任一該等情形下。

術語「*癌症等級*」可指是否存在癌症、癌症之階段、腫瘤大小、所涉及染色體區域缺失或擴增之數目(例如倍增或增至三倍)及/或癌症嚴重程度之其他量度。癌症等級可為數值或其他符號。該等級可為0。癌症等級亦包括與缺失或擴增有關之癌變前或癌前病況。

癌性組織(腫瘤)可具有染色體區域之變異，例如缺失或擴增。腫瘤可將DNA片段釋放至機體中之流體中。實施例可藉由分析DNA片段以鑑別染色體區域中DNA相對於正常(預期)值之變異來鑑別腫瘤。

缺失或擴增之確切大小以及位置可有所有不同。有時可能已知特定區域一般在癌症中或在特定類型之癌症中顯示變異(由此可診斷特定癌症)。當特定區域未知時，可採用分析整個基因組或該基因組之大部分之系統方法來檢測變異區域，該等變異區域可分散遍佈基因組且其大小(例如缺失或擴增之鹼基數目)有所不同。可隨時間追蹤染色體區域以鑑別變異之嚴重程度或顯示變異之區域數目之變化。此追蹤可為篩選、預測及監測腫瘤(例如在治療後或用於檢測復發或腫瘤進展)提供重要資訊。

此說明首先始於癌症中染色體變異之實例。然後，討論藉由檢測及分析生物學試樣中之無細胞DNA來檢測染色體變異之方式之實例。在建立檢測一個染色體區域中之變異之方法後，闡述以系統方式使用檢測多個染色體區域中之變異之方法來篩選(診斷)並預測患者。此說明亦闡述隨時間追蹤自對一或多個區域中染色體變異之測試獲得之數值指標以提供篩選、預測及監測患者的方法。然後討論實例。

I.癌症中染色體變異之實例

通常在癌細胞中檢測到染色體變異。此外，可在所選類型之癌症中發現染色體變異之特徵性模式。舉例而言，通

常在肝細胞癌(HCC)中檢測到染色體臂1p、1q、7q、15q、16p、17q及20q中DNA之增加以及3p、4q、9p及11q上DNA之喪失。先前研究已證實亦可在癌症患者之循環DNA中檢測到此等基因變異。舉例而言，已在罹患肺癌及頭頸癌患者之循環DNA分子中之特定基因座檢測到異型接合性喪失(LOH)(Chen XQ等人，Nat Med 1996; 2: 1033-5；Nawroz H等人，Nat Med 1996; 2: 1035-7)。血漿或血清中檢測到之基因改變與彼等發現於腫瘤組織中者相同。然而，由於腫瘤源DNA僅佔總循環無細胞DNA之一小部分，因此藉由腫瘤細胞之LOH引起之對偶基因不均衡通常較小。許多研究人員已研發數位聚合酶鏈反應(PCR)技術(Vogelstein B, Kinzler KW. Proc Natl Acad Sci U S A. 1999; 96: 9236-41；Zhou W等人，Nat Biotechnol 2001; 19: 78-81；Zhou W等人，Lancet. 2002; 359: 219-25)來精確量化循環DNA分子中基因座之不同對偶基因(Chang HW等人，J Natl Cancer Inst. 2002; 94: 1697-703)。數位PCR在檢測藉由腫瘤DNA中特定基因座之LOH引起之小對偶基因不均衡方面比即時PCR或其他DNA量化方法更為靈敏。然而，數位PCR可能仍難以鑑別特定基因座之極小對偶基因不均衡，且因此本文所述實施例以共同方式分析染色體區域。

本文所述技術亦可用於檢測癌變前或癌前病況。此等病況之實例包括肝硬化及宮頸上皮內瘤。前一病況係肝細胞癌之癌變前病況，而後一病況係宮頸癌之癌變前病況。已報導，此等癌變前病況業已具有若干可發展成為惡性腫瘤

之分子改變。舉例而言，染色體臂1p、4q、13q、18q上存在LOH及同時喪失3個以上基因座與肝硬化患者發生HCC之風險增加有關(Roncalli M等人Hepatology 2000; 31:846-50)。此等癌變前病變亦會將DNA釋放至循環中，但可能濃度較低。該技術可允許藉由分析血漿中之DNA片段並量測血漿中循環癌變前DNA之濃度(包括分數濃度)來檢測缺失或擴增。檢測此等變異之容易性(例如測序深度或檢測到之此等變化之數目)及濃度將預測進展成完全成熟癌性病況之可能性或速度。

A.染色體區域之缺失

圖1繪示癌細胞中展現缺失變異之染色體區域。正常細胞顯示具有兩種單倍型，即Hap I及Hap II。如所顯示，Hap I與Hap II二者皆在複數個異型接合基因座110中之每一者處具有序列(亦稱作單核苷酸多態性SNP)。在與癌症有關之細胞中，Hap II缺失染色體區域120。作為實例，與癌症有關之細胞可來自腫瘤(例如，惡性腫瘤)、腫瘤之轉移灶(例如局部淋巴結中或遠端器官中)或癌前或癌變前病變，例如，如上文所提及。

在癌細胞中缺失兩種同源單倍型中之一者之染色體區域120中，全部異型接合SNP 110皆因在對應缺失同源染色體上喪失另一對偶基因而表現為同型接合。因此，此類染色體變異稱為異型接合性喪失(LOH)。在區域120中，該等SNP之非缺失對偶基因將代表可在正常組織中發現之兩種單倍型中之一者。在圖1所示之實例中，LOH區域120上之

單倍型I(Hap I)可藉由對腫瘤組織實施基因分型來確定。另一單倍型(Hap II)可藉由比較正常組織與癌症組織之表觀基因型來確定。Hap II可藉由接合全部缺失對偶基因來構築。亦即，確定正常細胞之區域120中在癌細胞之區域120中不會出現的全部對偶基因皆位於相同單倍型(即Hap I)上。藉助此分析，可對腫瘤組織中展現LOH之全部染色體區域確定患者(例如，肝細胞癌HCC患者)之單倍型。此一方法僅用於患有具有癌細胞時，且僅用於確定區域120中之單倍型，但提供對缺失染色體區域之良好闡釋。

B.染色體區域之擴增

圖2繪示癌細胞中展現擴增變異之染色體區域。正常細胞顯示具有兩種單倍型，即Hap I及Hap II。如所顯示，Hap I與Hap II二者皆在複數個異型接合基因座210之每一者處具有序列。在腫瘤細胞中，Hap II之染色體區域220擴增兩次(倍增)。

類似地，對於在腫瘤組織中具有單對偶基因擴增之區域而言，SNP 210處之擴增對偶基因可藉由諸如微陣列分析等方法檢測。兩種單倍型中之一者(圖2所示實例中之Hap II)可藉由將染色體區域220中之全部擴增對偶基因接合在一起來確定。特定基因座處之擴增對偶基因可藉由比較基因座處之每一對偶基因之數目確定。然後，另一單倍型(Hap I)可藉由將非擴增對偶基因接合在一起來確定。此一方法僅用於患有具有癌細胞時，且僅用於確定區域220中之單倍型，但提供對擴增染色體區域之良好闡釋。

擴增可源自具有2個以上染色體，或一個染色體中一個基因重複出現。一個區域可串聯倍增，或區域可為含有該區域之一或多個拷貝之微小染色體。擴增亦可源自一個染色體之一基因拷貝並插入不同染色體或相同染色體中不同區域中。此等插入係一種擴增類型。

II.染色體區域之選擇

由於癌症組織將提供該等無細胞DNA(及可能細胞DNA)中之至少一些，因此可在諸如血漿及血清等試樣中檢測到癌症組織之基因組變異。檢測變異之問題在於，腫瘤或癌症可能相當小，因此提供之癌細胞DNA相對較少。因此，具有變異之循環DNA之量相當少，因此使檢測極為困難。在基因組中單一基因座處之DNA可能不足以檢測變異。本文所述方法可藉由以下方式來克服此困難：分析包括複數個基因座(單倍型)之染色體區域處之DNA，由此在對單倍型進行合計時將一個基因座處之微小變化改變為可察覺之差異。因此，分析區域之複數個基因座可提供較大精度，且可減少假正性及假負性。

同樣，變異區域可相當小，由此使得難以鑑別變異。若僅使用一個基因座或特定基因座，則將漏掉不在該等基因座處之變異。本文所述之一些方法可研究整個區域以在區域子集內發現變異。當分析區域跨越基因組時，可因此分析全基因組以發現不同長度及位置之變異，如下文所更詳細闡述。

為闡釋該等要點，如上文所示，區域可具有變異。但區

域必須經選擇用於分析。區域之長度及位置可改變結果，且因此影響分析。舉例而言，若分析圖1中之第一區域，則不會檢測到變異。若分析第二區域，則可使用(例如)本文所述方法檢測到變異。若分析包括第一區域與第二區域二者之較大區域，則會遇到僅較大區域之部分具有變異從而可使得更難以鑑別任何變異之困難，以及鑑別變異之確切位置及長度之問題。各實施例可解決該等困難中之一些及/或全部。對選擇區域之說明同樣適用於使用相同染色體區域之單倍型或使用兩個不同染色體區域之方法。

A.選擇特定染色體區域

在一個實施例中，可基於對癌症或患者之瞭解來選擇特定區域。舉例而言，已知該區域通常可在許多癌症或特定癌症中展現變異。可藉由參照關於對癌症類型或具有特定風險因子之患者熟知者之文獻來確定該區域之確切長度及位置。另外，可獲得患者之腫瘤組織並分析以鑑別變異區域，如上文所述。當前，儘管此一技術需要獲得癌細胞(此對於僅診斷之患者可能不現實)，但可使用此一技術來鑑別用於在相同患者中隨時間監測之區域(例如，在手術以移除癌性組織後，或在化學療法或免疫療法或靶向療法後，或用於檢測腫瘤復發或進展)。

可鑑別一個以上特定區域。可獨立地使用對此等區域中每一者之分析，或可共同分析不同區域。另外，可將區域細分以在定位變異時提供較大精度。

圖3顯示表300，其繪示不同類型之癌症及有關區域及其

對應變異。310欄列示不同癌症類型。本文所述實施例可用於任一類型之涉及變異之癌症，且因此此列表僅係實例。320欄顯示增加(擴增)與相同列之特定癌症有關之區域(例如，大區域，例如7p或17q25之更特定區域)。330欄顯示可發現喪失(缺失)之區域。340欄列示討論該等區域與特定癌症之關聯之參考文獻。

該等具有可能染色體變異之區域可用作根據本文所述方法分析之染色體區域。癌症中改變之其他基因組區域之實例可參見the Cancer Genome Anatomy Project (cgap.nci.nih.gov/Chromosomes/RecurrentAberrataions)及the Atlas of Genetics and Cytogenetics in Oncology and Haematology (atlasgeneticsoncology.org Tumors/Tumorliste.html)之數據庫。

如可發現，鑑別區域可相當大，而其他區域可能較具體。變異可能不包括表中所鑑別之整個區域。因此，此等關於變異類型之線索無法確切定位特定患者中變異實際存在於何處，但可另外用於大致指出大區域以供分析。此等大區域可包括在較大區域(其詳細內容闡述於本文中)中個別地以及共同地分析之許多亞區(其可具有相等大小)。因此，實施例可基於欲測試癌症之特定情形組合選擇大區域之態樣，但亦可採用更多一般技術(例如，測試亞區)，如下文所述。

B.選擇任意染色體區域

在另一實施例中，任意地選擇所分析染色體區域。舉例而言，可將基因組分成長度為一百萬鹼基(Mb)或其他預定

區段長度(例如500 Kb或2 Mb)之區域。若區域為1 Mb，則在人類基因組中存在約3,000個區域，此乃因在單倍體人類基因組中存在約30億個鹼基。然後可對該等區域各自分析，如下文所更詳細討論。

此等區域可不基於對癌症或患者之任一瞭解，而是基於將基因組系統分割為欲分析區域來確定。在一實作中，當染色體之長度不為預定區段之倍數(例如，不可被1百萬鹼基整除)時，染色體之最後區域可小於預定長度(例如小於1 MB)。在另一實作中，可基於染色體之總長度及欲產生區段之數目(其通常會在染色體間有所不同)將各染色體分成相等長度(或近似相等-在捨入誤差內)之區域。在此一實作中，各染色體區段之長度可有所不同。

如上文所提及，可基於所測試特定癌症鑑別特定區域，但然後可將特定區域細分成較小區域(例如跨越特定較大區域之相等大小之亞區)。以此方式，可指出變異。在下文討論中，任何一般提及之染色體區域可為明確鑑別之區域、任意選擇之區域或二者之組合。

III.特定單倍型中變異之檢測

此部分闡述藉由分析包括無細胞DNA之生物學試樣來檢測單一染色體區域中之變異之方法。在此部分之實施例中，單一染色體區域在該區域中複數個基因座處係異型接合(不同對偶基因)，由此提供可藉由瞭解給定基因座處之特定對偶基因來區別之兩種單倍型。因此，可將給定核酸分子(例如，無細胞DNA片段)鑑別為來自兩種單倍型中之

特定一者。舉例而言，可對片段測序以獲得與染色體區域比對之序列標籤，且隨後可鑑別對偶基因所屬異型接合基因座處之單倍型。兩種一般技術類型闡述於下文中，用於確定特定單倍型(Hap)之變異，具體而言用於標籤計數及大小分析。

A.確定單倍型

為區分兩種單倍型，首先確定染色體區域之兩種單倍型。舉例而言，可確定圖1之正常細胞中所示之兩種單倍型Hap I及Hap II。在圖1中，單倍型包括異型接合之第一複數個基因座110，且允許兩種單倍型之間有差異。此第一複數個基因座跨越所分析染色體區域。可首先確定不同異型接合基因座(het)上之對偶基因且隨後定相以確定患者之單倍型。

可藉由單分子分析法來確定SNP對偶基因之單倍型。此等方法之實例已由Fan等人(Nat Biotechnol. 2011; 29:51-7)、Yang等人(Proc Natl Acad Sci U S A. 2011; 108:12-7)及Kitzman等人(Nat Biotechnol. 2011 Jan; 29:59-63)闡述。另一選擇為，可藉由分析家族成員(例如父母、同胞及子女)之基因型來確定個體之單倍型。實例包括由Roach等人(Am J Hum Genet. 2011; 89(3):382-97)及Lo等人(Sci Transl Med. 2010; 2:61ra91)闡述之方法。在又一實施例中，可藉由比較腫瘤組織及基因組DNA之基因分型結果來確定個體之單倍型。可藉由微陣列分析(例如使用t)來實施該等個體之基因型分析。

5

亦可藉由彼等熟習此項技術者所熟知之其他方法來構築單倍型。此等方法之實例包括彼等基於以下者：單分子分析(例如數位PCR(Ding C及Cantor CR. Proc Natl Acad Sci USA 2003; 100: 7449-7453；Ruano G等人Proc Natl Acad Sci USA 1990; 87: 6296-6300))、染色體分選或分離(Yang H等人Proc Natl Acad Sci U S A 2011; 108: 12-17；Fan HC等人Nat Biotechnol 2011; 29: 51-57)、精子單倍型分析(Lien S等人Curr Protoc Hum Genet 2002；第1章：第1.6單元)及成像技術(Xiao M等人Hum Mutat 2007; 28: 913-921)。其他方法包括彼等基於以下者：對偶基因特異性PCR(Michalatos-Beloin S等人Nucleic Acids Res 1996; 24: 4841-4843；Lo YMD等人Nucleic Acids Res 19: 3561-3567)、選殖及限制酶消化(Smirnova AS等人Immunogenetics 2007; 59:93-8)等。又一些方法係基於群中單倍型區塊之分佈及連鎖不平衡結構，此允許自統計評估推斷個體之單倍型(Clark AG. Mol Biol Evol 1990; 7:111-22; 10:13-9；Salem RM等人Hum Genomics 2005; 2:39-66)。

若可獲得腫瘤組織，則確定LOH之區域之單倍型的另一方法係藉由對個體之正常組織及腫瘤組織實施基因分型。在LOH存在下，對於顯示LOH之區域內之全部SNP基因座而言，具有極高分數濃度之腫瘤細胞之腫瘤組織會顯示表觀同型接合性。該等SNP基因座之基因型會包含一種單倍型(圖1中LOH區域之Hap I)。另一方面，正常組織會指示，對於LOH之區域內之SNP基因座而言，個體係異型接

合。存在於正常組織而非腫瘤組織中之對偶基因會包含另一單倍型(圖1中LOH區域之Hap II)。

B.相對單倍型劑量(RHDO)分析

如上文所提及，具有染色體區域之單倍型之一者之擴增或缺失的染色體變異會導致兩種單倍型在腫瘤組織之染色體區域中之劑量不均衡。在具有腫瘤生長者之血漿中，一部分循環DNA係源自腫瘤細胞。由於在癌症患者之血漿中存在腫瘤源DNA，因此其血漿中亦會存在此等不均衡。可藉助對來自各單倍型之分子之數目計數來檢測兩種單倍型之劑量不均衡。

對於在腫瘤組織(例如圖1之區域120)中觀察到LOH之染色體區域而言，由於缺乏來自腫瘤組織之Hap II之貢獻，因此在與Hap II比較時，Hap I會在循環DNA分子(片段)內過度呈現。對於在腫瘤組織中觀察到拷貝數擴增之染色體區域而言，由於釋放額外劑量之來自腫瘤組織之Hap II，因此在與Hap I比較時，Hap II會在受Hap II之單對偶基因擴增影響之區域過度呈現。為確定過度呈現或呈現不足，確定試樣中之某些DNA片段來自Hap I或Hap II，此可藉由多種方法進行，例如藉由實施通用測序及比對或使用數位PCR及序列特異性探針。

在對來自癌症患者之血漿(或其他生物學試樣)之複數個DNA片段測序以生成經測序標籤後，可鑑別對應於兩種單倍型上之對偶基因的經測序標籤並計數。然後可比較對應於兩種單倍型中每一者之經測序標籤之數目以確定兩種單

5

倍型是否相等地呈現於血漿中。在一個實施例中，可使用逐次機率比測試(SPRT)來確定兩種單倍型在血漿中之呈現是否顯著不同。統計學顯著差異表明在所分析染色體區域存在染色體變異。另外，可使用兩種單倍型在血漿中之定量差異來估計腫瘤源DNA在血漿中之分數濃度，如下文所述。

測定本案中所述DNA片段之身份(例如其在人類基因組中之位置)的診斷方法不限於使用大規模平行測序作為檢測平臺。該等診斷方法亦可應用於(例如(但不限於)微流體數位PCR系統(例如Fluidigm數位陣列系統、微滴數位PCR系統(例如來自RainDance及QuantaLife者)、BEAM-ing系統((即珠粒、乳液PCR、擴增及磁力)(Diehl等人ProcNatlAcadSci USA 2005; 102: 16368-16373)、即時PCR、基於質譜之系統(例如SequenomMassArray系統)及多重連接依賴性探針擴增(MLPA)分析。

正常區域

圖4繪示根據本發明之實施例，癌細胞內不展現變異之染色體區域以及在血漿中進行之測量。可藉由任一方法來選擇染色體區域410，例如，基於欲測試之特定癌症，或基於使用跨越大部分基因組之預定區段的一般篩選。為區分兩種單倍型，首先測定兩種單倍型。圖4顯示正常細胞之染色體區域410之兩種單倍型(Hap I及Hap II)。該等單倍型包括第一複數基因座420。此第一複數基因座420跨越所分析之染色體區域410。如所示，該等基因座在正常細胞

中係異型接合。亦顯示癌細胞之兩種單倍型。在該癌細胞中，無區域缺失或擴增。

圖4亦顯示各單倍型上之各該基因座420之對偶基因計數之數目。亦提供染色體區域410之某些亞區之累積總數。對偶基因計數數目對應於DNA片段數目，該等DNA片段對應於各特定基因座之特定單倍型。舉例而言，針對Hap I對包括第一基因座421且具有對偶基因A之DNA片段計數。而且針對Hap II對具有對偶基因T之DNA片段計數。對片段比對處(即其是否包括特定基因座)及其所含對偶基因之測定可以如本文所提及之各種方式測定。可使用兩種單倍型之計數比率來測定是否存在統計學顯著差異。此比率在本文中稱為優勢比。亦可使用兩個值間之差；該差可藉由片段總數正規化。該比率及差(及其函數)係與閾值比較以決定是否存在變異之分類之參數實例。

RHDO分析可利用相同單倍型上之全部對偶基因(例如累積計數)來確定在血漿中是否存在兩種單倍型之任何不均衡，例如，如可在母體血漿中進行，如上文所提及Lo專利公開案12/940,992及12/940,993中所述。此方法可顯著增加用於確定是否存在任何不均衡之DNA分子之數目，且因此獲得更佳的統計檢力，以用於區分因癌症存在所致之不均衡與在癌症或癌變前病況不存在下對偶基因計數之隨機分佈。與單獨地分析多重SNP基因座不同，RHDO方法可利用對偶基因在兩條染色體上之相對位置(單倍型資訊)，使得可一起分析位於相同染色體上之對偶基因。在單倍型資

5

訊不存在下，不能將不同SNP基因座之對偶基因計數相加在一起以在統計學上確定單倍型在血漿中過度呈現抑或低度呈遞。對偶基因計數之量化可藉由(但不限於)以下來實施：大規模平行測序(例如使用藉由合成系統之Illumina測序、藉由連接技術(SOLiD)之測序(Life Technologies)、Ion Torrent測序系統(Ion Torrent and Life Technologies)、奈米孔測序(nanoporetech.com)及454測序技術(Roche))、數位PCR(例如藉由微流體數位PCR(例如Fluidigm (fluidigm.com))或BEAMing(珠粒、乳液PCR、擴增、磁力學(inostics.com))或液滴PCR(例如藉由QuantaLife (quantalife.com)及RainDance (raindancetechnologies.com))及即時PCR。在該技術之另一實作中，可利用使用溶液內捕獲(in-solution capture)(例如使用Agilent SureSelect系統、Illumina TruSeq Custom Enrichment套組(illumina.com/applications/sequencing/targeted_resequencing.ilmn)或藉由MyGenostics GenCap Custom Enrichment系統(mygenostics.com/))或基於陣列之捕獲(例如使用Roche NimbleGen系統)之富集目標測序。

在圖4中所示之實例中，前兩個SNP基因座(對於第一SNP，24對26；及對於第二SNP，18對20)觀察到輕微對偶基因不均衡。然而，對偶基因計數之數目在統計學上不足以確定是否存在真正對偶基因不均衡。因此，將對相同單倍型上之對偶基因之計數相加在一起直至兩種單倍型之累積對偶基因計數足以在統計學上得出以下結論，染色體區

域410(對於此實例，第五SNP)之兩種單倍型之間不存在對偶基因不均衡。在達成統計學顯著分類後，將累積計數復位(對於此實例，於第六SNP)。然後確定累積計數直至兩種單倍型之累積對偶基因計數再次足以在統計學上得出以下結論，區域410之該特定亞區之兩種單倍型之間無對偶基因不均衡。總累積計數亦可用於整個區域，但與整個區域410相反，先前方法可測試不同亞區，此可在確定變異之位置上提供較大精度(即亞區)。用於確定是否存在真正對偶基因不均衡之統計測試之實例包括(但不限於)逐次機率比測試(Zhou W等人Nat Biotechnol 2001; 19: 78-81；Zhou W等人Lancet. 2002; 359: 219-25)、t測試及卡方測試(chi-square test)。

檢測缺失

圖5繪示根據本發明之實施例，癌細胞內染色體區域510之缺失以及在血漿中進行之確定缺失區域之量測。圖5顯示正常細胞之染色體區域510之兩種單倍型(Hap I及Hap II)。該等單倍型包括跨越所分析染色體區域510之第一複數個異型接合基因座520。亦顯示癌細胞之兩種單倍型。在癌細胞中，Hap II缺失區域510。與圖4一樣，圖5亦顯示每一基因座520之對偶基因計數之數目。染色體區域510內某些亞區亦保持累積總數。

由於腫瘤組織通常含有腫瘤細胞與非腫瘤細胞之混合物，因此可藉由偏移區域510內基因座處兩個對偶基因之量之比率來表現LOH。在此一情形下，可藉由組合基因座

5

520來確定區域510中之缺失單倍型Hap II，在與正常組織上之對應基因座比較時，基因座520顯示DNA片段量之相對降低。具有更經常出現之片段之單倍型係保留於腫瘤細胞中之Hap I。在某些實施例中，可能期望實施會富集腫瘤細胞在腫瘤試樣中之比例，以使得更容易地確定缺失及保留單倍型之程序。此一程序之一個實例係顯微解剖(手動或藉由雷射捕獲技術)。

理論上，在腫瘤組織中，對於展現LOH之染色體區域而言，Hap I上之每一對偶基因會在循環DNA中過度呈現且對偶基因不均衡程度會取決於腫瘤DNA在血漿中之分數濃度。然而，同時，兩個對偶基因在任一循環DNA試樣中之相對豐度亦會由蔔瓦松分佈(Poisson distribution)決定。可實施統計分析以確定所觀察到之對偶基因不均衡係歸因於癌症組織中存在LOH抑或歸因於偶然性。檢測癌症中與LOH有關之真正對偶基因不均衡之能力取決於所分析循環DNA分子數目及腫瘤DNA之分數濃度。腫瘤DNA之較高分數濃度及較大分析分子數目會產生檢測真正對偶基因不均衡之較高靈敏度及特異性。

在圖5中所示之實例中，前兩個SNP基因座(對於第一SNP，24對22；及對於第二SNP，18對15)觀察到輕微對偶基因不均衡。然而，對偶基因計數之數目在統計學上不足以確定是否存在真正對偶基因不均衡。因此，將對相同單倍型上之對偶基因之計數相加在一起直至兩種單倍型之累積對偶基因計數足以在統計學上得出以下結論，在區域

510(對於此實例，第五SNP)中存在兩種單倍型之間之對偶基因不均衡。在一些實施例中，僅已知不均衡，不確定具體類型(缺失或擴增)。然後確定累積計數直至兩種單倍型之累積對偶基因計數再次足以在統計學上得出以下結論，區域510之該特定亞區之兩種單倍型之間存在對偶基因不均衡。總累積計數亦可用於整個區域，且可以本文所述任一方法實施。

檢測染色體區域之擴增

圖6繪示根據本發明之實施例，癌細胞內染色體區域610之擴增以及在血漿中進行之確定擴增區域之量測。除LOH外，亦經常在癌症組織中觀察到染色體區域之擴增。在圖6所示之實例中，將癌細胞中染色體區域610中之Hap II擴增至3個拷貝。如所顯示，區域610僅包括6個異型接合基因座，此與先前圖中所示較長區域相反。第六基因座中之擴增經檢測為統計學顯著，其中確定過度呈現為統計學顯著。在一些實施例中，僅已知不均衡，且不確定具體類型(缺失或擴增)。在其他實施例中，可獲得癌細胞並分析。此分析可提供關於不均衡歸因於缺失(對於缺失區域而言，癌細胞為同型接合)抑或擴增(對於擴增區域而言，癌細胞係異型接合)之資訊。在其他實作中，可使用第IV部分之方法分析整個區域(即並非個別地單倍型)來確定存在缺失抑或擴增。若區域過度呈現，則變異係擴增；且若區域係呈現不足，則變異係缺失。亦分析區域620且累積計數證實不存在不均衡。

用於血漿RHDO分析之SPRT分析

對於具有異型接合基因座之任何染色體區域而言，可使用RHDO分析來確定血漿中是否存在兩種單倍型之任何劑量不均衡。在該等區域中，血漿中存在單倍型劑量不均衡表明血漿試樣中存在腫瘤源DNA。在一個實施例中，可使用SPRT分析來確定Hap I及Hap II之經測序讀取物(read)數目之差是否在統計學上顯著。在SPRT分析之此實例中，首先確定來自兩種單倍型中每一者之經測序讀取物數目。然後可確定代表可能過度呈現之單倍型所貢獻經測序讀取物之比例量之參數(例如分數)(例如一種單倍型之讀取物數目除以另一單倍型之讀取物數目之分數)。可能過度呈現之單倍型在LOH之情形下為非缺失單倍型，且在染色體區域之單對偶基因擴增之情形下為擴增單倍型。然後，比較此分數與基於虛無假設(即不存在單倍型劑量不均衡)及備擇假設(即存在單倍型劑量不均衡)構築之兩個閾值(上限及下限)。若分數大於上限，則其指示血漿中存在兩種單倍型之統計學顯著不均衡。若分數低於下限，則其指示不存在兩種單倍型之統計學顯著不均衡。若分數介於上限與下限之間，則其指示無足夠統計檢力來得出結論。可逐次增加所分析區域之異型接合基因座之數目直至可進行成功的SPRT分類。

用於計算SPRT之上限及下限之方程式係：

上限=[(ln 8)/N-ln δ]/ln γ；下限=[(ln 1/8)/N-ln δ]/ln γ，其中 $\delta=(1-\theta_1)/(1-\theta_2)$ 且 $\gamma=\frac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)}$，$\theta_1$係在血漿中存在對偶基因不

均衡時，來自可能過度呈現之單倍型之經測序標籤的預期分數，

θ_2係在不存在對偶基因不均衡時，兩種單倍型中任一者之預期分數，即0.5，N係Hap I及Hap II之經測序標籤之總數，*ln*係代表自然對數(即$\log_e$)之數學符號。θ_1將取決於預期(或認為)存在於血漿試樣中之腫瘤源DNA之分數濃度(F)。

在LOH之情形下，$\theta_1=1/(2-F)$。在單對偶基因擴增之情形下，$\theta_1=(1+zF)/(2+zF)$，其中z代表腫瘤中染色體區域經擴增之額外拷貝數。舉例而言，若一個染色體倍增，則會有特定染色體之一個額外拷貝。則z等於1。

圖7顯示根據本發明之實施例，在HCC患者之血漿DNA中對腫瘤組織中位於染色體1p處且顯示單對偶基因擴增之區段的RHDO分析。綠色三角形代表患者之數據。經測序讀取物總數隨所分析SNP之數目增加而增加。來自腫瘤中擴增單倍型之經測序讀取物之總分數隨所分析經測序讀取物之總數的增加而變化且最終達到高於上限之值。此指示顯著單倍型劑量不均衡且因此支持血漿中存在此與癌症有關之染色體變異。

對HCC患者在腫瘤組織中顯示擴增及缺失之全部染色體區域實施使用SPRT之RHDO分析。結果如下，已知922個區段具有LOH且已知105個區段具有擴增。對於LOH而言，利用SPRT對922個區段分類，且將該等區段中之921個正確地鑑別為在血漿中具有單倍型劑量不均衡，以提供

99.99%之精度。對於單對偶基因擴增而言，利用SPRT對105個區段分類，且將該等區段中之105個正確地鑑別為在血漿中具有單倍型劑量不均衡，以提供100%之精度。

C.相對單倍型大小分析

作為對與兩種單倍型比對之片段之劑量計數的替代，可使用各別單倍型之片段大小。舉例而言，對於特定染色體區域而言，可比較一種單倍型之DNA片段大小與另一單倍型之DNA片段大小。可分析對應於該區域之第一單倍型之異型接合基因座處任一對偶基因之DNA片段的大小分佈，且比較其與對應於第二單倍型之異型接合基因座處任一對偶基因之DNA片段的大小分佈。可使用大小分佈之統計學顯著差異以與計數數目可達成類似之方式來鑑別變異。

已報導，癌症患者之總(即腫瘤與非腫瘤)血漿DNA之大小分佈有所增加(Wang BG等人Cancer Res. 2003; 63: 3966-8)。然而，若明確地研究腫瘤源DNA(而非DNA之總(即腫瘤與非腫瘤)量)，則觀察到腫瘤源DNA分子之大小分佈比源自非腫瘤細胞之分子之大小分佈更短(Diehl等人Proc Natl Acad Sci U S A. 2005; 102:16368-73)。因此，可使用循環DNA之大小分佈來確定是否存在與癌症有關之染色體變異。大小分析原理顯示於圖8中。

圖8顯示根據本發明之實施例，在存在含有缺失之腫瘤時染色體區域之兩種單倍型之片段的大小分佈變化。如圖8中所繪示，腫瘤組織中缺失T對偶基因。因此，腫瘤組織僅會將A對偶基因之短分子釋放至血漿中。腫瘤源短DNA

分子會導致血漿中A對偶基因之大小分佈總體縮短，從而導致A對偶基因在血漿中之大小分佈與T對偶基因相比更短。如先前部分中所討論，可一起分析位於相同單倍型上之全部對偶基因。換言之，可比較載有位於一種單倍型上之對偶基因之DNA分子的大小分佈與載有另一單倍型上之對偶基因之DNA分子的大小分佈。腫瘤組織中之缺失單倍型會在血漿中顯示較長大小分佈。

大小分析亦可應用於檢測與癌症有關之染色體區域之擴增。圖9顯示根據本發明之實施例，在存在含有擴增之腫瘤時，染色體區域之兩種單倍型之片段的大小分佈變化。在圖9中所示實例中，腫瘤中載有對偶基因T之染色體區域倍增。因此，血漿中會釋放增加量之載有T對偶基因之短DNA分子，從而導致T對偶基因之大小分佈與A對偶基因之大小分佈相比總體縮短。大小分析可共同地應用於位於相同單倍型上之全部對偶基因。換言之，單倍型擴增在腫瘤組織中之大小分佈會比未擴增單倍型在腫瘤中之大小分佈更短。

檢測循環DNA之大小分佈之縮短

可藉由(但不限於)末端配對大規模平行測序來確定源自兩種單倍型(即Hap I及Hap II)之DNA片段的大小。在對DNA片段之末端測序後，可將經測序讀取物(標籤)與參照人類基因組比對。可自各末端最外核苷酸之坐標推斷測序DNA分子之大小。可使用分子之經測序標籤來確定測序DNA片段源自Hap I抑或Hap II。舉例而言，一種經測序標

籤可包括所分析染色體區域中之異型接合基因座。

因此，對於各測序分子而言，可確定大小與其源自Hap I抑或Hap II二者。基於與每一單倍型比對之片段之大小，電腦系統可計算Hap I與Hap II二者之大小分佈特徵(例如平均片段大小)。可使用適當統計分析比較來自Hap I及Hap II之DNA片段之大小分佈以確定大小分佈之差異何時足以鑑別變異。除末端配對大規模平行測序外，可使用其他方法來確定DNA片段之大小，包括(但不限於)對全DNA片段測序、質譜及光學方法，其用於觀察及比較所觀察DNA分子與標準物之長度。

接下來，引入兩種實例方法用於檢測與腫瘤之基因變異有關之循環DNA之縮短。該兩種方法旨在提供對兩群DNA片段之大小分佈之差異的定量量測。兩群DNA片段係指對應於Hap I及Hap II之DNA分子。

短DNA片段之分數差

在一實作中，使用短DNA片段之分數。設定截止大小(w)以界定短DNA分子。截止大小可有所不同且經選擇以適合不同診斷目的。電腦系統可確定等於或適於大小截止之分子數目。然後可藉由用短DNA數除以總數DNA片段來計算DNA片段分數(Q)。DNA分子群之大小分佈會影響Q值。較短總體大小分佈表示較高比例之DNA分子為短片段，因此，獲得較高Q值。

然後可使用Hap I與Hap II間之短DNA片段之分數差。Hap I及Hap II之短片段之分數差(ΔQ)可反映來自Hap I及

Hap II之DNA片段之大小分佈之差異。$\Delta Q=Q_{HapI}-Q_{HapII}$，其中$Q_{HapI}$係Hap I DNA片段之短片段之分數；且$Q_{HapII}$係Hap II DNA片段之短片段之分數。$Q_{HapI}$及$Q_{HapII}$係兩組來自各單倍型之片段之大小分佈之統計值的實例。

如先前部分中所闡釋，當腫瘤組織中缺失Hap II時，Hap I DNA片段之大小分佈會比Hap II DNA片段之大小分佈更短。因此，會觀察到正ΔQ值。可比較正ΔQ值與閾值以確定ΔQ是否大至足以認為存在缺失。Hap I之擴增亦會顯示正ΔQ值。當腫瘤組織中之Hap II倍增時，Hap II DNA片段之大小分佈會比Hap I DNA片段之大小分佈更短。因此，ΔQ值會變為負。在不存在染色體變異時，Hap I及Hap II DNA片段在血漿/血清中之大小分佈係類似的。因此，ΔQ值約為0。

可比較患者之ΔQ與正常個體以確定該值是否正常。另外或另一選擇為，可比較患者之ΔQ值與自患有類似癌症之患者獲得之值以確定該值是否異常。此比較可涉及與本文所述閾值之比較。在疾病監測之情形下，可隨時間連續監測ΔQ值。ΔQ值之變化可指示腫瘤DNA在血漿/血清中之增加分數濃度。在此技術之所選實作中，可將腫瘤DNA之分數濃度與疾病之腫瘤階段、預測及進展相關聯。使用不同時刻之量測進行之此等實作更詳細地討論於下文中。

由短DNA片段所貢獻總長度之分數差

在此實作中，使用由短DNA片段所貢獻總長度之分數。電腦系統可確定一組DNA片段(例如來自給定區域之特定

5

單倍型之片段或僅來自給定區域之片段)在試樣中之總長度。可選擇低於時將DNA片段定義為「短片段」之截止大小(w)。截止大小可有所不同且經選擇以適合不同診斷目的。然後，電腦系統可藉由將所隨機選擇等於或短於截止大小之DNA片段之長度相加確定短DNA片段之總長度。然後可將由短DNA片段所貢獻總長度之分數計算如下：$F=\sum^{w}$長度$/\sum^{N}$長度，其中$\sum^{w}$長度代表長度等於或小於w (bp)之DNA片段之長度的總和；且$\sum^{N}$長度代表長度等於或小於預定長度N之DNA片段的總和。在一個實施例中，N係600個鹼基。然而，可使用其他大小限制(例如150個鹼基、180個鹼基、200個鹼基、250個鹼基、300個鹼基、400個鹼基、500個鹼基及700個鹼基)來計算「總長度」。

可選擇600個鹼基或以下之值，此乃因Illumina基因組分析儀系統不能對長於600個鹼基之DNA片段有效地擴增及測序。另外，將分析限制於短於600個鹼基之DNA片段亦可避免源自基因組之結構差異之偏誤。在結構差異(例如重排)(Kidd JM等人，Nature 2008; 453:56-64)存在下，當藉由將DNA片段末端定位至參照基因組以生物資訊方式估計大小時，可能高估DNA片段之大小。另外，>99.9%全部成功地測序且定位至參照基因組之DNA片段小於600個鹼基且因此(包括等於及短於600個鹼基之全部片段)會提供DNA片段在試樣中之大小分佈的代表性估計。

因此，可使用Hap I與Hap II間由短DNA片段所貢獻總長度之分數差。Hap I與Hap II DNA片段之F值之差可反映其

間之大小分佈之干擾程度。此處，將$F_{Hap\ I}$及$F_{Hap\ II}$定義為分別由Hap I及Hap II之短DNA片段所貢獻總長度之分數。可將Hap I與Hap II間由短DNA片段所貢獻總長度之分數差(ΔF)計算為：$\Delta F=F_{Hap\ I}-F_{Hap\ II}$。$F_{HapI}$及$F_{HapII}$係兩組來自各單倍型之片段之大小分佈之統計值的實例。

與先前部分中所闡釋實施例類似，腫瘤組織中缺失Hap II會導致Hap I DNA片段之大小分佈在與Hap II DNA片段比較時明顯縮短。此會產生正ΔF值。當Hap II倍增時，會觀察到負ΔF值。在不存在染色體變異時，ΔF值約為0。

可比較患者之ΔF與正常個體以確定該值是否正常。可比較患者之ΔF與自患有類似癌症之患者獲得之值以確定該值是否異常。此比較可涉及與本文所述閾值之比較。在疾病監測之情形下，可連續監測ΔQ值。ΔF值之變化可指示腫瘤DNA在血漿/血清中之增加分數濃度。

D. 一般方法

圖10係繪示根據本發明之實施例，分析有機體之生物學試樣之單倍型，以確定染色體區域是否展現缺失或擴增之方法的流程圖。生物學試樣包括源自正常細胞及可能源自與癌症有關之細胞之核酸分子(亦稱為片段)。該等分子在試樣中可為無細胞的。有機體可屬於具有一個以上染色體拷貝之任何類型，即，至少二倍體有機體，但可包括較高多倍體有機體。

在本文所述此方法及任何其他方法之一個實施例中，生物學試樣包括無細胞DNA片段。儘管已使用對血漿DNA之

分析來闡釋本申請案中所述之不同方法，但該等方法亦可應用於在含有正常DNA與腫瘤源DNA之混合物之試樣中檢測與腫瘤有關之染色體變異。其他試樣類型包括唾液、眼淚、胸膜液、腹水液、膽汁、尿液、血清、胰液、糞便及宮頸塗片試樣。

在步驟1010中，確定有機體之正常細胞於第一染色體區域之第一及第二單倍型。可藉由任一適宜方法(例如彼等本文所提及者)確定單倍型。可經由任一方法(例如，本文所述方法)選擇染色體區域。第一染色體區域包括異型接合之第一複數個基因座(例如，區域410之基因座420)。異型接合基因座(het)可彼此相距較遠，例如，基因座可與第一複數個基因座之另一基因座彼此相距500個或1000個鹼基(或更多)。其他het可存在於第一染色體區域中，但不使用。

在步驟1020中，在各分子之位置及對偶基因方面對生物學試樣中之複數個核酸分子實施表徵。舉例而言，可鑑別核酸分子在有機體之參照基因組中之位置。此定位可以各種方式實施，包括對分子實施測序(例如經由通用測序)，以獲得該分子之一或多個(末端配對)經測序標籤且隨後將經測序標籤與參照基因組比對。此比對可使用諸如基本局部比對搜尋工具(BLAST)等工具實施。可將位置鑑別為染色體臂中之編號。可使用一個異型接合基因座(het)處之對偶基因來確定片段所屬單倍型。

在步驟1030中，基於所鑑別位置及確定對偶基因將第一

組核酸分子鑑別為來自第一單倍型。舉例而言，可將包括圖4之具有對偶基因A之基因座421之片段鑑別為來自Hap I。第一組可因包括至少一個位於第一複數個基因座中每一者處之核酸分子來跨越第一染色體區域。

在步驟1040中，基於所鑑別位置及確定對偶基因將第二組核酸分子鑑別為來自第二單倍型。舉例而言，可將包括圖4之具有對偶基因T之基因座421之片段鑑別為來自Hap II。第二組包括至少一個位於第一複數個基因座中每一者處之核酸分子。

在步驟1050中，電腦系統計算第一組核酸分子之第一值。第一值界定第一組核酸分子之性質。第一值之實例包括第一組中分子之數目之標籤計數及第一組中分子之大小分佈。

在步驟1060中，電腦系統計算第二組核酸分子之第二值。第二值界定第二組核酸分子之性質。

在步驟1070中，比較第一值與第二值以確定第一染色體區域是否展現缺失或擴增之分類。存在缺失或擴增之分類可提供關於具有與癌症有關之細胞之有機體之資訊。比較之實例包括獲得兩個值之差或比率及比較結果與一或多個閾值，如本文所述。舉例而言，可在SPRT分析中比較比率與閾值。實例分類可包括正(即檢測到擴增或缺失)、負及未分類、以及不同程度之正及負(例如，使用介於1與10之間整數或介於0與1之間之實數)。擴增可包括簡單倍增。此一方法可檢測與癌症有關之核酸之存在，該等核酸

包括腫瘤DNA及來自癌前病變(即癌症前體)之DNA。

E.深度

分析深度係指在指定精度內提供分類或另一確定所需分析之分子之量。在一個實施例中，可基於已知變異計算深度，且隨後可實施具有該深度之量測及分析。在另一實施例中，分析可持續至作出分類為止，且可使用作出分類時之深度來確定癌症等級(例如，癌症之階段或腫瘤大小)。以下提供一些涉及深度之計算之實例。

偏差可指本文所述任一差或比率。作為實例，偏差可為第一值與第二值之間之偏差或為參數與閾值或腫瘤濃度之偏差，如本文所述。若偏差增加一倍，則需要量測之片段數減少到¼。更通常而言，若偏差增加到N倍，則需要量測之片段數係$1/N^2$。作為推論，若偏差減少到1/N，則欲測試之片段數增加到N^2。N可為實數或整數。

假定試樣(例如血漿)的10%為腫瘤DNA之情形，且假定自對1000萬個片段測序發現統計學顯著差異。則假如舉例而言，實施富集程序而使得試樣中現在存在20%腫瘤DNA，則所需片段數為2,500,000個片段。以此方式，可將深度與腫瘤DNA在試樣中之百分比相關聯。

擴增量亦會影響深度。對於拷貝量為該區域中拷貝量之兩倍(例如為4，與之相比，正常為2)之區域而言，假定需要分析數目X之片段。若區域之拷貝量為正常拷貝量之4倍，則此區域將需要X/4量之片段。

F.閾值

可使用參數(例如每一單倍型之值之差或比率)與正常值之偏差量來提供診斷，如上文所述。舉例而言，偏差可為來自區域之一種單倍型之片段之平均大小與來自另一單倍型之片段之平均大小的差。若偏差高於某一量(例如，如自正常試樣及/或區域確定之閾值)，則鑑別為缺失或擴增。但高於閾值之程度可提供資訊，其可導致使用多個閾值，每一者對應於不同癌症等級。舉例而言，與正常之較高偏差可提供癌症所處階段(例如階段4之不均衡程度會高於階段3)。較高偏差亦可歸因於腫瘤較大且因此釋放許多片段及/或區域擴增多倍。

除提供不同癌症等級外，不同閾值亦可允許有效檢測具有變異之區域或特定區域。舉例而言，可設定高閾值以主要尋找3倍及更高之擴增，此會產生較一種單倍型之缺失更大之不均衡。亦可檢測區域之兩拷貝之缺失。同樣，可使用較低閾值來鑑別可能具有變異之區域，且然後可進一步分析該等區域以證實變異是否確實存在及其位置。舉例而言，可利用階層中之較低等級並使用較高閾值來實施二元搜尋(或更高階搜尋，例如八叉樹)。

圖11顯示根據本發明之實施例，癌細胞中缺失亞區1130之區域1110以及在血漿中進行之確定缺失區域之量測。可藉由本文所提及任一方法(例如藉由將基因組分裂成相等大小之區段)來選擇染色體區域1110。圖11亦顯示每一基因座1120之對偶基因計數之數目。區域1140(正常區域)及區域1130(缺失區域)亦分別保持累積總數。

若選擇區域1110進行分析，則在11個基因座內累積計數之數目係258(對於Hap I)及240(對於Hap II)，差為18。此一差作為計數總數之百分比小於僅分析缺失亞區1130時。此係合理之原因在於區域1110之約一半為正常，而癌細胞中缺失全部亞區1130。因此，端視所用閾值而可能漏掉區域1110中之變異。

為允許檢測亞區之缺失，實施例可使用相對較大區域之下限(對於此實例而言，假定區域1110與欲鑑別之缺失區域之大小相比相對較大)。儘管下限會鑑別更多區域，此可包括一些假正性，但其會減少假負性。現在，可藉助進一步分析來去除假正性，此亦可指出變異。

在已將區域標明用於進一步分析後，可將該區域劃分成亞區以供進一步分析。在圖11中，可將11個基因座分裂成兩半(例如使用二元樹)以提供6個基因座之亞區1140及具有5個基因座之亞區1130。可利用相同閾值或更嚴格之閾值來分析該等區域。在此實例中，然後可將亞區1140鑑別為正常且將亞區1130鑑別為包括缺失或擴增。以此方式，可排除不具有變異之較大區域，且可將時間用於進一步分析疑似區域(高於下限之區域)以鑑別顯示具有高置信度(例如使用上限)之變異之亞區。儘管此處使用RHDO，但大小技術同樣適用。

可基於欲檢測變異之大小選擇第一搜尋級之區域之大小(及樹中之較低級之亞區之大小)。已發現癌症顯示10個具有10 MB長度之變異之區域。患者亦已具有100 MB展現變

異之區域。稍後階段之癌症可具有較長變異部分。

G.變異在區域內之位置的精確定位(Refinement)

在最後一部分中，討論基於樹搜尋將區域分成亞區。此處，討論其他用於分析亞區及指出區域內變異之方法。

圖12顯示根據本發明之實施例，可如何使用RHDO分析定位變異之位置。水平顯示染色體區域，其中將非癌細胞之單倍型標記為Hap I及Hap II。將Hap II在癌細胞中之缺失區域標記為LOH。

如所顯示，RHDO分析自假想染色體區域1202之左側開始至右側。各箭頭代表RHDO分類區段。每一區段可視為其自身區域，具體而言具有較大區域之het之子集之亞區。在可確定分類之前，RHDO分類區段之大小取決於基因座數(及基因座之位置)。各RHDO區段中所包括基因座數取決於分析用於每一區段之分子數目、期望精度(例如SPRT分析中之優勢比)及腫瘤源DNA在試樣中之分數濃度。當分子數目足以確定在兩種單倍型之間存在統計學顯著差異時，將作出分類，如圖4及圖5中闡釋之實例中所述。

各實心水平箭頭代表顯示DNA試樣中不存在單倍型劑量不均衡之RHDO分類區段。在腫瘤中不具有LOH之區域內，作出6個RHDO分類且各指示單倍型劑量不均衡之不存在。下一RHDO分類區段1210橫跨具有LOH之區域與不具有LOH之區域之間之接合部1205。在圖12之下部中，顯示RHDO區段1210之SPRT曲線。黑色豎直箭頭指示具有LOH

之區域與不具有LOH之區域之間之接合部。隨著來自具有LOH之區域之漸增數據的累積，此區段之RHDO分類指示存在單倍型劑量不均衡。

各白色水平箭頭代表指示單倍型劑量不均衡之存在的RHDO分類區段。右側上隨後4個RHDO亦指示DNA試樣中存在單倍型劑量不均衡。可推斷具有LOH之區域與不具有LOH之區域間之接合部的位置在顯示RHDO分類變化(即自存在單倍型劑量不均衡至不存在單倍型劑量不均衡或反之亦然)之第一RHDO區段內。

圖13顯示根據本發明之實施例，始於另一方向之RHDO分類。在圖13中，顯示自二個方向之RHDO分類。自始於左側之RHDO分析，可推斷具有LOH之區域與不具有LOH之區域間之接合部在顯示存在單倍型劑量不均衡之第一RHDO區段1310內。自始於右側之RHDO分析，可推斷接合部在指示不存在單倍型劑量不均衡之第一RHDO區段1320內。組合以兩個方向實施之RHDO分析之資訊，可推斷具有LOH之區域與不具有LOH之區域間之接合部的位置1330。

IV.變異之非特異性單倍型檢測

RHDO方法依賴於使用異型接合基因座。現在，二倍體有機體之染色體將具有一些差異，從而導致兩種單倍型，但異型接合基因座之數目可有所不同。一些個體可具有相對較少之異型接合基因座。本部分所述實施例亦可用於同型接合基因座，其中比較兩個區域且並非比較相同區域之

兩種單倍型。因此，可獲得更多數據點，但根據與兩個不同染色體區域之比較可能存在一些缺陷。

在相對染色體區域劑量方法中，比較來自一個染色體區域之片段數(例如，如藉由對與該區域比對之經測序標籤計數所確定)與預期值(其可來自參照染色體區域或來自另一已知健康試樣中之相同區域)。以此方式，可對染色體區域之片段計數，無論經測序標籤來自哪種單倍型。因此，仍可使用不含有het之經測序標籤。為實施比較，實施例可將標籤計數正規化，然後進行比較。藉由至少兩個基因座(其彼此分開)來界定各區域，且該等基因座處之片段可用於獲得關於該區域之集合值。

用於特定區域之經測序讀取物(標籤)的正規化值可藉由用與該區域比對之經測序讀取物數目除以可與全基因組比對之經測序讀取物總數來計算。此正規化標籤計數允許比較一個試樣之結果與另一試樣之結果。舉例而言，正規化值可為預期來自特定區域之經測序讀取物之比例(例如，百分比或分數)，如上文所述。但可進行許多其他正規化，如熟習此項技術者會明瞭。舉例而言，可藉由用一個區域之計數數目除以參照區域(在上述情形下，參照區域即為全基因組)之計數數目來正規化。然後可比較此正規化標籤計數與閾值，該閾值可自一或多個不展現癌症之參照試樣確定。

隨後比較該測試例之正規化標籤計數與一或多個參照個體(例如無癌症者)之正規化標籤計數。在一個實施例中，

藉由計算該例之特定染色體區域之z-計分來進行比較。使用以下方程式計算z-計分：z-計分=(該例之正規化標籤計數-均值)/S.D.，其中「均值」係與參照試樣之特定染色體區域比對之均值正規化標籤計數；及S.D.係與參照試樣之特定區域比對之正規化標籤計數之數目的標準偏差。因此，z-計分係測試例之染色體區域之正規化標籤計數偏離一或多個參照個體之相同染色體區域之均值正規化標籤計數的標準偏差數。

在所測試有機體患有癌症之情況下，腫瘤組織中擴增之染色體區域會在血漿DNA中過度呈現。此會導致正z計分值。另一方面，腫瘤組織中缺失之染色體區域會在血漿DNA中呈現不足。此會導致負z計分值。z計分數值係由幾個因素決定。

一個因素係腫瘤源DNA在生物學試樣(例如血漿)中之分數濃度。腫瘤源DNA在試樣(例如血漿)中之分數濃度愈高，則測試例與參照例之正規化標籤計數間之差愈大。因此，會獲得較大z計分數值。

另一因素係在一或多個參照例中正規化標籤計數之差異。在測試例之生物學試樣(例如血漿)中染色體區域之過度呈現程度相同時，在參照組中正規化標籤計數之較小差異(即較小標準偏差)會導致較高z計分。類似地，在測試例之生物學試樣(例如血漿)中染色體區域之呈現不足程度相同時，在參照組中正規化標籤計數之較小標準偏差會導致較高之負z計分。

另一因素係腫瘤組織中之染色體變異數值。染色體變異數值係指特定染色體區域之拷貝數變化(增加或損失)。腫瘤組織中之拷貝數變化愈高，則特定染色體區域在血漿DNA中之過度呈現或呈現不足程度將愈高。舉例而言，喪失染色體之兩個拷貝會導致染色體區域在血漿DNA中之呈現不足大於喪失染色體之兩個拷貝中之一者之情形，且因此導致負性較高之z計分。通常，癌症中存在多種染色體變異。各癌症中之染色體變異可進一步在以下改變：其性質(即擴增或缺失)、其程度(單或多個拷貝增加或喪失)及其程度(在染色體長度上之變異大小)。

所分析分子之數目影響量測正規化標籤計數之精度。預期在分數濃度分別為約12.5%、6.3%及3.2%時，將需要分析15,000個、60,000個及240,000個分子來檢測具有一個拷貝變化(增加或喪失)之染色體變異。用於檢測癌症之不同染色體區域之標籤計數的其他細節闡述於頒予Lo等人之標題為「Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing」之美國專利公開案第2009/0029377號中，該案件之全部內容出於所有目的以引用方式併入本文中。

實施例亦可使用大小分析代替標籤計數方法。亦可使用大小分析代替正規化標籤計數。大小分析可使用如本文及美國專利申請案第12/940,992號中所提及之各種參數。舉例而言，可使用來自上文之Q或F值。此等大小值無需藉由來自其他區域之計數之正規化，此乃因該等值不會隨讀取

物數目而縮放。單倍型特異性方法之技術亦可用於非特異性方法。舉例而言，可使用涉及區域之深度及精確定位之技術。在一些實施例中，可在比較兩個區域時考慮特定區域之GC偏誤。由於RHDO方法使用相同區域，因此無需此一校正。

V.多個區域

儘管某些癌症通常可以在特定染色體區域中具有變異之形式存在，但此等癌症不會總是僅存在於相同區域中。舉例而言，其他染色體區域可顯示變異，且此等其他區域之位置可能未知。此外，在篩選患者以鑑別癌症之早期階段時，可能期望鑑別寬範圍之癌症，其可顯示存在於整個基因組內之變異。為應對該等情況，實施例可以系統方式分析複數個區域以確定顯示變異之區域。可使用變異數及變異位置(例如其是否鄰接)基於展現變異之各區域之數目及位置來(例如)證實變異，確定癌症之階段，提供對癌症之診斷(例如若數目大於閾值)及提供預後。

因此，實施例可基於顯示變異之區域之數目來鑑別有機體是否患有癌症。因此，可測試複數個區域(例如，3000個)以鑑別展現變異之區域之數目。該等區域可涵蓋整個基因組或僅部分基因組，例如，非重複區域。

圖14係根據本發明之實施例，使用複數個染色體區域分析有機體之生物學試樣之方法1400的流程圖。生物學試樣包括核酸分子(亦稱為片段)。

在步驟1410中，鑑別有機體之複數個非重疊染色體區

域。每一染色體區域包括複數個基因座。如上文所提及，區域之大小可為1 Mb，或某另一相等大小。然後整個基因組可包括約3,000個各具有預定大小及位置之區域。同樣，如上文所提及，可改變此等預定區域以適應特定染色體之長度或指定數目之欲用區域、及本文所提及任一其他標準。若區域具有不同長度，則此等長度可用於將結果正規化，例如，如本文所述。

在步驟1420中，對於複數個核酸分子中之每一者而言，鑑別核酸分子在有機體之參照基因組中之位置。可以本文所提及任一方式確定位置，例如，藉由對片段測序以獲得經測序標籤及將經測序標籤與參照基因組比對。對於單倍型特異性方法而言，亦可確定分子之特定單倍型。

對每一染色體區域實施步驟1430-1450。在步驟1430中，基於所鑑別位置將各別組核酸分子鑑別為來自染色體區域。各別組包括至少一個位於染色體區域之複數個基因座中每一者處之核酸分子。在一個實施例中，該組可為與染色體區域之特定單倍型比對之片段，例如，如上文RHDO方法中所述。在另一實施例中，該組可為與染色體區域比對之任一片段，如第IV部分中所述方法中所述。

在步驟1440中，電腦系統計算各別組核酸分子之各別值。各別值界定各別組核酸分子之性質。各別值可為本文所提及任一值。舉例而言，該值可為組中之片段數或組中片段之大小分佈之統計值。各別值亦可為正規化值，例如，用區域之標籤計數除以試樣之標籤計數總數或參照區

域之標籤計數之數目。各別值亦可為與另一值之差或比率(例如，在RHDO中)，由此提供區域之差異之性質。

在步驟1450中，比較各別值與參照值以確定第一染色體區域是否展現缺失或擴增之分類。此參照值可為本文所述任一閾值或參照值。舉例而言，參照值可為針對正常試樣確定之閾值。對於RHDO而言，各別值可為兩種單倍型之標籤計數之差或比率，且參照值可為用於確定存在統計學顯著偏差之閾值。作為另一實例，參照值可為另一單倍型或區域之標籤計數或大小值，且比較可包括獲得差或比率(或其函數)及隨後確定差或比率是否大於閾值。

參照值可基於其他區域之結果而變化。舉例而言，若相鄰區域亦顯示偏差(儘管與一個閾值相比較小，例如，z計分為3)，則可使用下限。舉例而言，若3個連續區域皆高於第一閾值，則較有可能為癌症。因此，此第一閾值可低於自非連續區域鑑別癌症所需之另一閾值。有3個(或3個以上)區域甚至具有小偏差可具有足夠低之偶然性效應之機率，該偶然性效應可維持靈敏度及特異性。

在步驟1460中，確定分類為展現缺失或擴增之染色體區域之量。所計數染色體區域可具有限制。舉例而言，僅可對與至少一個另一區域鄰接之區域計數(或可能需要鄰接區域具有某一大小，例如，4個或更多個區域)。對於區域不相等之實施例而言，數目亦可為各別長度(例如，數目可為變異區域之總長度)。

在步驟1470中，比較該量與閾值量以確定對試樣之分

類。作為實例，分類可為有機體是否患有癌症、癌症之階段及癌症之預後。在一個實施例中，對全部變異區域計數且使用單一閾值，無論該等區域出現在何處。在另一實施例中，閾值可基於所計數區域之位置及大小而變化。舉例而言，可比較特定染色體或染色體臂上區域之量與該特定染色體(或臂)之閾值。可使用多個閾值。舉例而言，特定染色體(或臂)上變異區域之量必須大於第一閾值，且基因組中變異區域之總量必須大於第二閾值。

區域量之此閾值亦可取決於所計數區域之不均衡強度。舉例而言，用作確定癌症分類之閾值之區域量可取決於用於檢測各區域中之變異之特異性及靈敏度(變異閾值)。舉例而言，若變異閾值為低(例如z計分為2)，則可將閾值之量選擇為高(例如，150)。但若變異閾值為高(例如，z計分為3)，則閾值之量可為較低(例如，50)。顯示變異之區域之量亦可為加權值，例如，一個顯示高不均衡之區域可經加權高於僅顯示微小不均衡之區域(即對於變異存在多於僅正及負之分類)。

因此，可使用顯示正規化標籤計數(或組之性質之另一各別值)之顯著過度呈現或呈現不足之染色體區域的量(其可包括數目及/或大小)來反映疾病之嚴重程度。具有變異正規化標籤計數之染色體區域之量可藉由兩個因素確定，即染色體變異在腫瘤組織中之數目(或大小)及腫瘤源DNA在生物學試樣(例如血漿)中之分數濃度。較晚期之癌症傾向於展現較多(及較大)染色體變異。因此，會在試樣(例如

血漿)中可能檢測到較多與癌症有關之染色體變異。在患有較晚期之癌症之患者中，較高腫瘤負載會導致腫瘤源DNA在血漿中之較高分數濃度。因此，會較容易在血漿試樣中檢測到與腫瘤有關之染色體變異。

在癌症篩選或檢測之情況下，可使用展現正規化標籤計數(或另一值)過度呈現或呈現不足之染色體區域之量來確定測試個體患上癌症之可能性。使用±2(即z計分>2或<-2)之截止值時，預期約5%測試區域會僅因偶然性而產生顯著偏離對照個體之均值之z計分。當將全基因組劃分成1 Mb區段時，全基因組會有約3,000個區段。因此，預期約150個區段會具有>2或<-2之z計分。

因此，可使用用於z計分為>2或<-2之區段之數目之150的截止(閾值)值來確定是否存在癌症。用於具有變異z計分之區段之數目的其他截止值(例如，100、125、175、200、250及300)可經選擇以適合診斷目的。較低截止值(例如100)會達成較靈敏之測試，但特異性較低，且較高截止值會具有較高特異性，但靈敏度較低。可藉由增加z計分之截止值來減少假正性分類數。舉例而言，若將截止值增加至3，則僅0.3%區段會為假正性。在此情況下，可使用3個以上具有變異z計分之區段來指示癌症之存在。亦可選擇其他截止值，例如1、2、4、5、10、20及30，以適合不同診斷目的。然而，檢測與癌症有關之染色體變異之靈敏度會隨進行診斷所需變異區段數之增加而降低。

一種可能改良靈敏度而不犧牲特異性之方法係考慮毗鄰

染色體區段之結果。在一個實施例中，用於z計分之截止值仍為>2且<-2。然而，僅在兩個連續區段顯示相同類型之變異，例如兩個區段皆具有>2之z計分時，會將染色體區域分類為可能變異。若正規化標籤計數之偏差係隨機誤差，則具有兩個在相同方向上為假正性之連續區段之機率係0.125%(5%×5%/2)。另一方面，若染色體變異涵蓋兩個連續區段，較低截止值會使對該等區段在血漿試樣中之過度呈現或呈現不足之檢測較靈敏。由於正規化標籤計數(或另一值)與對照個體之均值之偏差並非歸因於隨機誤差，因此連續分類需要不會對靈敏度具有顯著不良效應。在其他實施例中，可使用較高截止值將相鄰區段之z計分相加在一起。舉例而言，可對3個連續區段之z計分求和且可使用5之截止值。此概念可延伸至3個以上連續區段。

量與變異閾值之組合亦可取決於分析目的及對有機體之任何預先瞭解(或其缺乏)。舉例而言，若針對癌症篩選正常健康群體，則通常可能會在區域量(即區域數目之高閾值)及變異閾值(對於在將區域鑑別為具有變異時而言)二者上使用高特異性。但在具有較高風險之患者(例如受腫塊或家族病史、吸煙者、HPV病毒、肝炎病毒或其他病毒困擾之患者)中，閾值可能較低以具有較高靈敏度(較低假負性)。

在一個實施例中，若使用1-Mb解析度及6.3%腫瘤源DNA之較低檢測限來檢測染色體變異，則每一1-Mb區段中之所需分子數目將為60,000。對於全基因組而言，此可理

解為約1億8千萬個(60,000個讀取物/Mb×3,000 Mb)可比對讀取物。

圖15顯示表1500，其繪示根據本發明之實施例，不同區段數所需之深度及腫瘤源片段之分數濃度。1510欄提供來自試樣之腫瘤細胞之片段的濃度。濃度愈高，檢測變異愈容易，因此需要分析之分子數目愈少。1520欄提供每區段所需之估計分子數目，其可經由上文關於深度之部分中所述之方法來計算。

較小區段大小會使檢測較小染色體變異具有較高解析度。然而，此會增加對欲分析之分子總數之需要。較大區段大小會以降低解析度為代價減少分析所需分子數目。因此，僅可檢測到較大變異。在一實作中，可使用較大區域，可將顯示變異之區段細分且分析該等亞區以獲得較佳解析度(例如，如上文所述)。1530欄提供每一區段之大小。該值愈小，所用區域愈多。1540欄顯示全基因組中欲分析之分子數目。因此，若已作出估計(或欲檢測之最低濃度)，則可確定欲分析之分子數目。

VI.隨時間之進展

隨著腫瘤進展，腫瘤片段量將有所增加，此乃因腫瘤將釋放較多DNA片段(例如，因腫瘤生長、較多壞死或較高血管供應)。來自腫瘤組織之較多DNA片段進入血漿將增加血漿中之不均衡程度(例如，RHDO中兩種單倍型之間之標籤計數差將有所以增加)。另外，由於腫瘤片段數增加，因此可較容易地檢測存在變異之區域之數目。舉例而

言，區域之腫瘤DNA之量可能過小而使得不能檢測變異，此乃因當腫瘤較小且釋放少量癌症DNA片段時，由於未分析足量片段而不能建立統計學顯著差異。較多片段甚至可在腫瘤較小時進行分析，但此可能需要大試樣(例如大量血漿)。

對癌症進展之追蹤可使用在一或多個區域中之變異量(例如，藉由不均衡或所需深度)或展現變異之染色體區域之量(數目及/或大小)。在一個實例中，若一個區域(或若干區域)之變異量之增加速度比其他區域之變異更快，則該(等)區域可用作較佳標記物來監測癌症。此增加可歸因於腫瘤較大且因此釋放許多片段及/或區域擴增多倍。亦可監測手術後之變異值(例如變異量或顯示變異之區域之數目或其組合)以證實已適當地移除腫瘤。

在該技術之各實作中，使用腫瘤DNA之分數濃度之確定值對癌症之進展進行分期、預測或監測。量測進展可提供關於癌症之當前階段及癌症生長或擴散之速度之資訊。癌症之「階段」係指以下中之全部或一些：腫瘤大小、組織學外觀、存在/不存在淋巴結侵犯及存在/不存在遠端轉移。對癌症之「預測」涉及估計疾病進展之幾率及/或自癌症存活之幾率。其亦可涉及對患者無臨床進展之時間或存活持續時間之估計。對癌症之「監測」會涉及檢查以觀察癌症是否已進展(例如大小已增加，淋巴結之侵犯已增加，或已擴散至遠端器官，即轉移)。監測亦可涉及檢查腫瘤是否已受治療控制。舉例而言，若治療有效，則可觀

5

察到腫瘤大小之減小、轉移或淋巴結侵犯之消退、患者之一般健康狀況之改良(例如體重增加)。

A.確定癌症DNA之分數濃度

一種追蹤一或多個區域之變異增加量之方式係確定(該)等區域之癌症DNA之分數濃度。然後可使用癌症DNA之分數濃度變化來隨時間追蹤腫瘤。可使用此追蹤來診斷，例如，第一量測可提供背景量(其可對應於一般變異程度)且稍後量測可觀察到將表明腫瘤生長(因此癌症)之變化。亦可使用癌症DNA之分數濃度變化來預測治療實施情況。在該技術之其他實作中，腫瘤DNA在血漿中之分數濃度之增加會指示患者之差預後或腫瘤負載之增加。

可以各種方式確定癌症DNA之分數濃度。舉例而言，一種單倍型與另一單倍型相比(或一個區域與另一區域相比)之標籤計數差。另一方法係觀察到統計學顯著差異前之深度(即分析片段數)。對於較早期實例，可使用單倍型劑量差藉由分析喪失異型接合性之染色體區域來確定腫瘤源DNA在生物學試樣(例如血漿)中之分數濃度。

已顯示，腫瘤源DNA之量與癌症患者之腫瘤負載正相關(Lo等人 Cancer Res. 1999; 59:5452-5及Chan等人 Clin Chem. 2005; 51:2192-5)。因此，可使用藉由RHDO分析連續監測腫瘤源DNA在生物學試樣(例如血漿試樣)中之分數濃度來監測患者之疾病進展。舉例而言，可使用監測治療後腫瘤源DNA在連續採集之試樣(例如血漿)中之分數濃度來確定治療之成功。

圖16顯示根據本發明之實施例，藉由RHDO分析量測腫瘤源DNA在血漿中之分數濃度之原理。確定兩種單倍型間之不均衡且可使用不均衡程度來確定腫瘤DNA在試樣中之分數濃度。

Hap I及Hap II代表非腫瘤組織中之兩種單倍型。腫瘤組織中之Hap II部分地缺失亞區1610。因此，血漿中檢測到之對應於缺失區域1610的與Hap II有關之片段係由非腫瘤組織貢獻。另一方面，腫瘤組織與非腫瘤組織二者中皆存在Hap I中之區域1610。因此，Hap I及Hap II之讀取物計數間之差將代表腫瘤源DNA在血漿中之量。

腫瘤源DNA之分數濃度(F)可自受LOH影響之染色體區域之缺失及非缺失染色體的經測序讀取物(標籤)數使用以下方程式計算：$F=(N_{HapI}-N_{HapII})/N_{HapI}\times 100\%$，其中$N_{HapI}$係對應於位於LOH所影響染色體區域中之異型接合SNP之Hap I上之對偶基因的經測序讀取物數目；且N_{HapII}係對應於位於LOH所影響染色體區域1610中之異型接合SNP之Hap II上之對偶基因的經測序讀取物數目。

上式等效於將p定義為位於不包括缺失之染色體區域(Hap I)上之異型接合基因座的累積標籤計數且將q定義為包括缺失之染色體區域(Hap II)1610之累積標籤計數，其中將腫瘤DNA在試樣中之分數濃度(F)計算為$F=1-q/p$。對於圖11中繪示之實例而言，腫瘤DNA之分數濃度係14%(1-104/121)。

在腫瘤切除之前及之後採集腫瘤源DNA在HCC患者之血

5

漿試樣中之分數濃度。在腫瘤切除之前，給定染色體區域之第一單倍型之N_{HapI}係30,443，且染色體區域之第二單倍型之N_{HapII}係16,221，其使得F為46.7%。在腫瘤切除之後，N_{HapI}係31,534，N_{HapII}係31,089，其使得F為1.4%。此監測顯示腫瘤切除係成功的。

亦可使用循環DNA大小特徵之變化程度來確定分數濃度。在一實作中，可確定源自腫瘤組織與非腫瘤組織二者之血漿DNA之確切大小分佈，且隨後介於兩個已知分佈之間之量測大小分佈可提供分數濃度(例如使用腫瘤組織與非腫瘤組織之大小分佈之兩個統計值之間的線性模型)。另一選擇為，可使用對大小變化之連續監測。在一個態樣中，將大小分佈變化確定為與腫瘤DNA在血漿中之分數濃度成比例。

亦可以類似方式使用不同區域間之差，即，上述非特異性單倍型檢測方法。在標籤計數方法中，可使用若干參數來監測疾病進展。舉例而言，可使用展現染色體變異之區域之z計分量值來反映腫瘤源DNA在生物學試樣(例如血漿)中之分數濃度。特定區域之過度呈現或呈現不足程度與腫瘤源DNA在試樣中之分數濃度及腫瘤組織中拷貝數變化之程度或數目成比例。Z計分量值係與對照個體相比特定染色體區域在試樣中之過度呈現或呈現不足程度之量度。因此，z計分量值可反映腫瘤DNA在試樣中之分數濃度及因此患者之腫瘤負載。

B.追蹤區域之數目

如上文所提及，可使用展現染色體變異之區域之數目來篩選癌症，且亦可使用其來進行監測及預測。作為實例，可使用監測來確定癌症之當前階段、是否已出現癌症及治療是否已起作用。隨著腫瘤進展，腫瘤之基因組組成會降格較多。為鑑別此持續降格，可使用追蹤區域(例如，1 Mb之預先界定區域)之數目之方法來鑑別腫瘤之進展。處於癌症較晚期階段之腫瘤則會具有較多展現變異之區域。

C.方法

圖17係繪示根據本發明之實施例，使用包括核酸分子之生物學試樣確定有機體中染色體變異之進展之方法的流程圖。在一個實施例中，至少一些核酸分子係無細胞的。作為實例，染色體變異可來自惡性腫瘤或癌變前病變。同樣，變異之增加可歸因於有機體隨時間而具有愈來愈多含有染色體變異之細胞，或歸因於有機體具有一定比例之每細胞含有增加量之變異之細胞。作為減少之實例，治療(例如手術或化學療法)可移除或減少與癌症有關之細胞。

在步驟1710中，鑑別有機體之一或多個非重疊染色體區域。每一染色體區域包括複數個基因座。可藉由任一適宜方法(例如，彼等本文所述者)來鑑別區域。

針對複數個時刻中之每一者實施步驟1720-1750。每一時刻對應於自有機體獲得試樣之不同時刻。當前試樣係針對給定時段分析之試樣。舉例而言，可在6個月內每月獲得試樣，且可在獲得試樣後立即進行分析。另一選擇為，可在若干時段內進行若干次量測後進行分析。

在步驟1720中，分析有機體之當前生物學試樣以鑑別核酸分子在有機體之參照基因組中之位置。可以本文所提及任一方式確定位置，例如，藉由對片段測序以獲得經測序標籤及將經測序標籤與參照基因組比對。對於單倍型特異性方法而言，亦可確定分子之特定單倍型。

對一或多個染色體區域中之每一者實施步驟1730-1750。當使用複數個區域時，可使用第V部分之實施例。在步驟1730中，基於所鑑別位置將各別組核酸分子鑑別為來自染色體區域。各別組包括至少一個位於染色體區域之複數個基因座中每一者處之核酸分子。在一個實施例中，該組可為與染色體區域之特定單倍型比對之片段，例如，如上文RHDO方法中所述。在另一實施例中，該組可為與染色體區域比對之任一片段，如第IV部分中所述方法中所述。

在步驟1740中，電腦系統計算各別組核酸分子之各別值。各別值界定各別組核酸分子之性質。各別值可為本文所提及任一值。舉例而言，該值可為組中之片段數或組中片段之大小分佈之統計值。各別值亦可為正規化值，例如，將區域之標籤計數除以試樣之標籤計數總數或參照區域之標籤計數之數目。各別值亦可為與另一值(例如，RHDO中)之差或比率，由此提供區域之差之性質。

在步驟1750中，比較各別值與參照值以確定第一染色體區域是否展現缺失或擴增之分類。此參照值可為本文所述任一閾值或參照值。舉例而言，參照值可為針對正常試樣

確定之閾值。對於RHDO而言，各別值可為兩種單倍型之標籤計數之差或比率，且參照值可為用於確定存在統計學顯著偏差之閾值。作為另一實例，參照值可為另一單倍型或區域之標籤計數或大小值，且比較可包括獲得差或比率(或其函數)及隨後確定差或比率是否大於閾值。可根據任一適宜方法及標準確定參照值，例如，如本文所述。

在步驟1760中，使用於複數個時刻對每一染色體區域之分類來確定有機體中染色體變異之進展。可使用進展來確定有機體是否患有癌症、癌症之階段及癌症之預後。該等確定中之每一者可涉及癌症分類，如本文所述。

可以各種方式實施此癌症分類。舉例而言，可對變異區域之量計數且與閾值比較。對區域之分類可為數值(例如，腫瘤濃度，其中各別值及參照值為不同單倍型或不同區域之值)且可確定濃度變化。可比較濃度變化與閾值以確定已發生顯著增加，由此發出存在腫瘤之信號。

VII.實例

A.使用SPRT之RHDO

在此部分中，顯示對肝細胞癌(HCC)患者使用相對單倍型劑量(RHDO)分析且使用SPRT之實例。在此患者之腫瘤組織中，觀察到兩個染色體4中一者之缺失。此導致染色體4上SNP之異型接合性之喪失。在對此患者進行單倍型分析時，分析該患者、其妻子及其兒子之基因組DNA且確定3名個體之基因型。然後自患者之基因型導出其組成型單倍型。實施大規模平行測序且鑑別對應於染色體4之兩

種單倍型SNP之對偶基因的經測序讀取物並計數。

RHDO及SPRT之方程式及原理已闡述於上文中。在一個實施例中，當兩種單倍型中之一者擴增或缺失時，RHDO分析經程式化以檢測(例如)DNA試樣中單倍型劑量之10%之差，其對應於存在10%腫瘤源DNA。在其他實施例中，RHDO分析之靈敏度可經設定以檢測DNA試樣中之2%、5%、15%、20%、25%、30%、40%及50%等之腫瘤源DNA。可在用於計算SPRT分類曲線之上限及下限之參數中調節RHDO分析之靈敏度。可調節參數可為期望檢測限值(例如應可檢測之腫瘤濃度百分比，其影響所分析分子數目)及分類用閾值，例如，使用優勢比(一種單倍型之標籤計數相對於另一單倍型之標籤計數之比率)。

在此RHDO分析中，虛無假設係，染色體4之兩種單倍型以相同劑量存在。備擇假設係，生物學試樣(例如血漿)中兩種單倍型之劑量相差10%以上。根據兩種假設以統計學方式比較具有對應於兩種單倍型之SNP對偶基因之經測序讀取物數目，作為來自不同SNP累積之數據。當累積數據足以確定兩種單倍型劑量以相等量存在或在統計學上相差至少10%時，作出SPRT分類。染色體4之q臂上之典型SPRT分類區塊顯示於圖18A中。此處僅出於闡釋目的使用10%之閾值。亦可檢測其他差異程度(例如0.1%、1%、2%、5%、15%或20%)。一般而言，期望檢測之差異程度愈低，需要分析之DNA分子愈多。相反地，期望檢測之差異程度愈大，需要分析且仍達成統計學顯著結果之DNA分

子數目愈少。在進行此分析時，將優勢比用於SPRT，但可使用其他參數，例如z計分或p-值。

在HCC患者於診斷時獲得之此血漿試樣中，染色體4之p臂及q臂分別存在76個及148個成功RHDO分類。全部RHDO分類指示於診斷時獲得之血漿試樣中存在單倍型劑量不均衡。作為比較，亦分析患者於手術切除腫瘤後獲得之血漿試樣，如圖18B中所顯示。對於治療後試樣而言，染色體4之p臂及q臂分別存在4個及9個成功RHDO分類。全部4個RHDO分類皆指示血漿試樣中沒有>10%之可觀察單倍型劑量不均衡。在染色體4q之9個RHDO分類中，7指示不存在單倍型劑量不均衡且2指示存在不均衡。在腫瘤切除後，顯示>10%之劑量不均衡之RHDO區塊數已顯著降低，從而指示治療後試樣與治療前試樣相比，顯示>10%之劑量不均衡之染色體區域的大小顯著更小。該等結果表明在手術切除腫瘤後腫瘤DNA在血漿中之分數濃度已有所降低。

當與非單倍型特異性方法比較時，RHDO分析可提供腫瘤DNA之分數濃度之較精確估計且尤其可用於監測疾病進展。因此，預期具有疾病進展之情形會展現腫瘤DNA在血漿中之分數濃度之增加；而具有穩定疾病之情形或彼等腫瘤大小已消退或減小者之腫瘤DNA在血漿中之分數濃度會有所降低。

B.靶向分析

在所選實施例中，DNA片段之通用測序可依照靶富集方

法實施。此一方法在本文中亦稱作富集靶測序。此一方法之一個實施例係使用溶液內捕獲系統(例如Agilent SureSelect系統、Illumina TruSeq Custom富集套組(illumina.com/applications/sequencing/targeted_resequencing.ilmn)或藉由MyGenostics GenCap Custom富集系統(mygenostics.com/))或基於微陣列之捕獲系統(例如Roche NimbleGene系統)預先選擇片段。儘管可捕獲一些其他區域，但優先捕獲某些區域。此等方法可允許以較多深度(例如，可利用數位PCR對較多片段進行測序或分析)及/或以較低成本分析此等區域。較大深度可增加區域中之靈敏度。可基於片段大小及甲基化模式實施其他富集方法。

因此，以全基因組方式分析DNA試樣之替代方式係靶向所關注檢測常見染色體變異之區域。靶向方法可潛在地改良此方法之成本有效性，此乃因分析過程將主要集中於可能存在染色體變異之區域或具有將尤其為特定腫瘤類型之特性之變化的區域、或彼等具有將在臨床上尤其重要之變化之區域。後一者之實例包括會於特定癌症類型之腫瘤形成中早期發生之變化(例如存在1q及8q之擴增及8q之缺失係HCC之早期染色體變化- van Malenstein等人Eur J Cancer 2011; 47:1789-97)；或與好或差預後有關之變化(例如，在腫瘤進展期間觀察到6q及17q處之增加、及6p及9p處之喪失，且18q、8p及17p處存在LOH與結腸直腸癌患者之較差存活有關-Westra等人Clin Colorectal Cancer 2004; 4:252-9)；或可預測治療反應之變化(例如7p處存在增加可預測具有

表皮生長因子受體突變之患者之對酷胺酸激酶抑制劑之反應- Yuan等人J ClinOncol 2011; 29: 3435-42)。癌症中改變之基因組區域之其他實例可參見許多在線數據庫(例如the Cancer Genome Anatomy Project database (cgap.nci.nih.gov/Chromosomes/RecurrentAberrataions)及the Atlas of Genetics and Cytogenetics in Oncology and Haematology (atlasgeneticsoncology.org//Tumors/Tumorliste.html))。相反，在非靶向全基因組方法中，將以與具有可能變異之區域相同之程度分析不可能發生染色體變異之區域。

已應用靶富集策略來分析3名HCC患者及4名健康對照個體之血漿試樣。使用來自Agilent之SureSelect捕獲系統實施靶富集(Gnirke等人Nat. Biotechnol 2009. 27: 182-9)。選擇SureSelect系統作為可能之靶富集技術之實例。亦可使用其他溶液相(IlluminaTruSeq Custom富集系統)或固相(例如Roche-Nimblegen系統)靶捕獲系統及基於擴增子之靶富集系統(例如QuantaLife系統及RainDance系統)。捕獲探針經設計位於在HCC中通常及很少顯示變異之染色體區域上。在靶捕獲後，然後在IlluminaGAIIx分析儀上藉由流動池之一個泳道對每一DNA試樣測序。使用很少出現擴增及缺失之區域作為參照來與較常出現擴增及缺失之區域比較。

在圖19中，顯示HCC中發現之常見染色體變異(圖改編自Wong等人(Am J Pathol 1999; 154:37-43))。染色體表意圖(chromosome ideogram)之右側線代表染色體增加且左側

5

線代表個別患者試樣之染色體喪失。粗線代表高度增加。長方形代表靶捕獲探針之位置。

靶向標籤計數分析

在檢測染色體變異時，首先計算具有可能變異之區域及參照區域之正規化標籤計數。然後針對該等區域之GC含量校正正規化標籤計數，如先前由Chen等人(PLoS One 2011; 6:e21791)所述。在當前實例中，選擇染色體8之p臂作為可能變異區域且選擇染色體9之q臂作為參照區域。使用Affymetrix SNP6.0陣列來分析3名HCC患者之腫瘤組織之染色體變異。對於3名患者而言，8p及9q在腫瘤組織中之染色體劑量變化顯示於下文中。HCC 013患者之8p有所喪失且9q無變化。HCC 027患者之8p有所增加且9q無變化。HCC 023患者之8p有所喪失且9q無變化。

然後使用靶向分析計算3名HCC患者及4名健康對照個體之chr 8p與9q間之正規化標籤計數的比率。圖20A顯示HCC患者及健康患者之正規化標籤計數比率之結果。對於HCC 013情形及HCC 023情形而言，觀察到減小之8p與9q間之正規化標籤計數比率。此與腫瘤組織中染色體8p之喪失的發現一致。對於HCC 027情形而言，觀察到增加之比率且與此情形之腫瘤組織中染色體8p之增加一致。虛線代表具有4個正常情形之均值之兩個標準偏差之區域。

靶向大小分析

在先前部分中，闡述藉由確定血漿DNA片段在癌症患者中之大小特徵來檢測與癌症有關之改變的原理。亦可利用

靶富集方法來檢測大小改變。對於3個HCC情形(HCC 013、HCC 027及HCC 023)而言，在將經測序讀取物與參照人類基因組比對後確定各測序DNA片段之大小。自兩個末端之最外核苷酸之坐標推斷測序DNA片段之大小。在其他實施例中，將對全DNA片段測序且然後可自測序長度直接確定片段大小。比較與染色體8p比對之DNA片段之大小分佈與與染色體9q比對之DNA片段之大小分佈。在檢測兩群DNA之大小分佈之差異時，首先確定當前實例中各群之短於150 bp之DNA片段之比例。在其他實施例中，可使用其他大小截止值，例如80 bp、110 bp、100 bp、110 bp、120 bp、130 bp、140 bp、160 bp及170 bp。然後將ΔQ值確定為兩個比例之差。$\Delta Q=Q_{8p}-Q_{9q}$，其中Q_{8p}係與染色體8p比對之短於150 bp之DNA片段的比例；且Q_{9q}係與染色體9q比對之短於150 bp之DNA片段的比例。

由於DNA片段之較短大小分佈會使短於截止值(即在當前實例中為150 bp)之DNA之比例值較高，因此較高(正性較高)ΔF值將代表相對於彼等與染色體9q比對者與染色體8p比對之DNA片段之較短分佈。相反，較小(或負性較高)結果將指示相對於彼等與染色體9q比對者與染色體8p比對之DNA片段之較長大小分佈。

圖20B顯示在靶富集及大規模平行測序後，3名HCC患者及4名健康對照個體之大小分析之結果。4名健康對照個體之ΔQ之正值指示，與染色體8p比對之DNA片段之大小分佈略短於彼等與染色體9q比對者。虛線代表在兩個相對於

4名對照個體之均值之標準偏差內之ΔQ區間。HCC 013情形及HCC 023情形之ΔQ值比對照個體之均值低超過兩個標準偏差。該兩種情形在腫瘤組織中缺失染色體8p。對於此染色體區域而言，腫瘤中缺失8p會導致腫瘤源DNA對血漿之貢獻減少。由於循環中之腫瘤源DNA短於源自非腫瘤組織之DNA，因此此會導致與染色體8p比對之血漿DNA片段明顯較長之大小分佈。此與該兩種情形中較低(負性較高)之ΔQ值一致。相反，8p在HCC 027情形中之擴增會導致與此區域比對之DNA片段明顯較少短之分佈。因此，將認為較高比例之與8p比對之血漿DNA片段較短。此與以下觀察一致：HCC 027之ΔQ值比健康對照個體之正性更高。

C.多個用於檢測腫瘤源染色體變異之區域

通常在腫瘤組織中檢測到染色體變異，包括某些染色體區域之缺失及擴增。在不同類型之癌症中觀察到染色體變異之特徵性模式。此處，使用若干實例來闡釋用於檢測癌症患者之血漿中該等與癌症有關之染色體變異的不同方法。吾人之方法亦可用於篩選癌症及監測疾病進展及對治療之反應。分析來自一名HCC患者及兩名鼻咽癌(NPC)患者之試樣。對於HCC患者而言，在手術切除腫瘤之前及之後採集靜脈血樣。對於兩名NPC患者而言，在診斷時採集靜脈血樣。另外，分析一名慢性肝炎B攜帶者及一名在血漿中具有可檢測艾伯斯坦-巴爾(Epstein-Barr)病毒DNA之個體之血漿試樣。該兩名個體未患有任何癌症。

利用微陣列分析來實施對腫瘤源染色體變異之檢測。具

體而言，使用Affymetrix SNP6.0微陣列系統來分析自HCC患者之血細胞及腫瘤試樣提取之DNA。使用Affymetrix Genotyping Console v4.0來確定血細胞及腫瘤組織之基因型。使用Birdseed v2算法來確定包括增加及缺失在內之染色體變異，該算法基於SNP之不同對偶基因之強度及微陣列上之拷貝數差異(CNV)探針。

基於計數之分析

為在血漿中實施經測序標籤計數分析，自各個體採集10毫升靜脈血。對於各血樣而言，在離心試樣後分離血漿。使用QIAmp血液微型套組(blood mini Kit) (Qiagen)自4-6 mL血漿提取DNA。如先前所述(Lo YMD. Sci Transl Med 2010, 2:61ra91)構築血漿DNA文庫，且隨後使用Illumina基因組分析儀平臺實施大規模平行測序。實施對血漿DNA分子之末端配對測序。對各分子之兩個末端中每一者之50 bp測序，因此每分子總共為100 bp。使用SOAP2程式(soap.genomics.org.cn/)將各序列之兩個末端與非重複遮蔽人類基因組(自UCSCgenome.ucsc.edu下載之Hg18 NCBI.36)比對(Li R等人Bioinformatics 2009, 25:1966-7)。

然後將基因組劃分成多個1百萬鹼基(1-Mb)區段且確定與各1-Mb區段比對之經測序讀取物數目。然後根據各庫(bin)之GC含量利用基於局部加權散點平滑(LOESS)回歸之算法來校正各庫之標籤計數(Chen E等人PLoS One 2011, 6: e21791)。此校正旨在將因不同基因組區段間之GC含量差所產生與測序有關之定量偏誤最小化。上文所提及成1-Mb

區段之劃分用於闡釋目的。亦可使用其他區段大小，例如2 Mb、10 Mb、25 Mb或50 Mb等。一般而言，亦可基於特定患者中特定腫瘤之基因組特性及特定類型之腫瘤來選擇區段大小。此外，若對於(例如)單分子測序技術(例如Helicos系統(www.helicosbio.com)或Pacific Biosciences單分子即時系統(www.pacificbiosciences.com))而言，測序方法可顯示具有低GC偏誤，則可略去GC校正步驟。

在先前研究中，已對來自未患有癌症之個體之57個血漿試樣測序。該等血漿測序結果用於確定各1-Mb區段之標籤計數之參照範圍。對於各1-Mb區段而言，確定57名個體之標籤計數之均值及標準偏差。然後，將研究個體之結果表示為z計分，如使用以下方程式所計算：z計分=(該情形之經測序標籤數-均值)/S.D，其中「均值」係與參照試樣之特定1-Mb區段比對之經測序標籤之均值數；且S.D.係與參照試樣之特定1-Mb區段比對之經測序標籤數之標準偏差。

圖21-24顯示4名研究個體之經測序標籤計數分析結果。1-MB區段顯示於曲線邊緣。人類染色體數及表意圖(最外環)係呈順時針方向之定向pter-qter(著絲粒係以黃色顯示)。在圖21中，內環2101顯示變異(缺失或擴增)區域，如自分析腫瘤所確定。內環2101顯示具有5個量表。量表係-2(最內線)至+2(最外線)。-2之值代表對應區域喪失兩個染色體拷貝。-1之值代表喪失兩個染色體拷貝中之一者。0之值代表無染色體增加或喪失。+1之值代表增加一個染色體拷貝且+2代表增加兩個染色體拷貝。

中環2102顯示血漿分析結果。如可發現，結果反映內環。中環2102係較多量表線，但進展相同。外環2103顯示來自分析治療後血漿之數據點，且該等數據點係灰色(證實無過度/呈現不足-無變異)。

經測序標籤在血漿中過度呈現之染色體區域(z計分>3)係由綠色點2110代表。經測序標籤在血漿中呈現不足之區域(z計分<-3)由紅色點2120代表。血漿中未檢測到顯著染色體變異之區域(z計分介於-3與3之間)係由灰色點代表。藉由計數總數將過度/呈現不足正規化。對於在測序前進行擴增而言，正規化可考慮GC偏誤。

圖21顯示HCC患者之Circos圖，其繪示根據本發明之實施例，來自血漿DNA之經測序標籤計數之數據。自內至外之跡線：藉由微陣列分析檢測到之腫瘤組織之染色體變異(紅色及綠色分別代表缺失及擴增)；在手術切除腫瘤前及切除後1個月獲得之血漿試樣之z計分分析。在腫瘤切除前，血漿中檢測到之染色體變異與彼等藉由微陣列分析在腫瘤組織中所鑑別者密切相關。在腫瘤切除後，血漿中大多數與癌症有關之染色體變異消失。該等數據反映此一方法對於監測疾病進展及治療效果之價值。

圖22顯示根據本發明之實施例，無HCC之慢性HBV攜帶者之血漿試樣之經測序標籤計數分析。與HCC患者(圖21)相反，在此患者之血漿中未檢測到與癌症有關之染色體變異。該等數據反映該方法用於癌症篩選、診斷及監測之價值。

圖23顯示根據本發明之實施例，對患有階段3 NPC之患者之血漿試樣的經測序標籤計數分析。在治療前獲得之血漿試樣中檢測到染色體變異。具體而言，在染色體1、3、7、9及14中鑑別出顯著變異。

圖24顯示根據本發明之實施例，對患有階段4 NPC之患者之血漿試樣的經測序標籤計數分析。在治療前獲得之血漿試樣中檢測到染色體變異。當與患有階段3疾病之患者(圖23)比較時，檢測到較多染色體變異。經測序標籤計數亦偏離對照之均值較多，即z計分偏離0較多(正或負)。與對照相比增加之染色體變異數及更高之經測序標籤計數偏差度反映在較晚期階段之疾病中更顯著程度之基因組改變，且因此反映此一方法用於對癌症進行分期、預測及監測之價值。

基於大小之分析

在先前研究中，已顯示，源自腫瘤組織之DNA之大小分佈短於彼等源自非腫瘤組織者之大小分佈(Diehl F等人 Proc Natl Acad Sci USA 2005, 102(45):16368-73)。在先前部分中，已概述藉由對血漿DNA之大小分析檢測血漿單倍型不均衡之方法。此處，使用HCC患者之測序數據來進一步闡釋此方法。

出於闡釋目的，鑑別兩個區域進行大小分析。在一個區域(染色體1 (chr1)；座標：159、935、347至167、219、158)中，在腫瘤組織中檢測到兩條同源染色體中之一者倍增。在另一區域(染色體10 (chr10)；座標：100、137、050

至101、907、356)中，在腫瘤組織中檢測到兩條同源染色體中一者之缺失(即LOH)。除確定測序片段所來自之單倍型外，亦使用參照基因組中測序片段之最外核苷酸之坐標以生物資訊方式確定測序片段之大小。然後，確定來自兩種單倍型中每一者之片段之大小分佈。

對於Chr10之LOH區域而言，腫瘤組織中缺失一種單倍型(缺失單倍型)。因此，與此缺失單倍型比對之全部血漿DNA片段皆源自非癌症組織。另一方面，與腫瘤組織中不缺失之單倍型(非缺失單倍型)比對之片段可源自腫瘤組織或非腫瘤組織。由於腫瘤源DNA之大小分佈較短，因此會預期來自非缺失單倍型之片段與彼等來自缺失單倍型者相比大小分佈更短。可藉由繪製片段之累積頻率對DNA片段之大小之曲線來確定兩種大小分佈之差異。具有更短大小分佈之DNA群會具有更豐富之短DNA且因此在大小譜之短末端處累積頻率增加更快。

圖25顯示根據本發明之實施例，血漿DNA之累積頻率對腫瘤組織中展現LOH之區域之大小之曲線。X軸係片段以鹼基對表示之大小。Y軸係大小低於X軸上之值之片段之百分比。與來自缺失單倍型之序列相比，來自非缺失單倍型之序列在低於170 bp之大小下具有更快增加及更高累積頻率。此指示來自非缺失單倍型之短DNA片段更豐富。此因來自非缺失單倍型之短腫瘤源DNA之貢獻而與上文預測一致。

在一個實施例中，大小分佈之差異可量化為兩群DNA分

子之累積頻率之差。將ΔQ定義為該兩群之累積頻率之差。$\Delta Q=Q_{非缺失}-Q_{缺失}$，其中$Q_{非缺失}$代表來自非缺失單倍型之測序DNA片段之累積頻率；且$Q_{缺失}$代表來自缺失單倍型之測序DNA片段之累積頻率。

圖26顯示對於LOH區域，ΔQ對經測序血漿DNA之大小。根據本發明之實施例，在130 bp之大小下，ΔQ達到0.2。此指示使用130 bp作為截止值用於上文方程式中來界定短DNA係最佳的。使用此截止值時，與來自缺失單倍型之群相比，來自非缺失單倍型之群之短DNA分子之豐度高20%。然後可比較此百分比差(或以類似方式導出之值)與得自未患有癌症之個體之閾值。

對於具有染色體擴增之區域而言，在腫瘤組織中一種單倍型倍增(擴增單倍型)。由於額外量之來自此擴增單倍型之短腫瘤源DNA分子會釋放至血漿中，因此來自擴增單倍型之片段之大小分佈會短於來自非擴增單倍型之片段之大小分佈。與LOH情形類似，可藉由繪製片段之累積頻率對DNA片段之大小之曲線來確定大小分佈之差異。具有更短大小分佈之DNA群會具有更豐富之短DNA且因此在大小譜之短末端處累積頻率增加更快。

圖27顯示根據本發明之實施例，血漿DNA之累積頻率對腫瘤組織中具有染色體倍增之區域之大小之曲線。與來自非擴增單倍型之序列相比，來自擴增單倍型之序列在低於170 bp之大小下具有更快增加及更高累積頻率。此指示來自擴增單倍型之短DNA片段更豐富。此與下文所示之預測

一致，此乃因較大量之短腫瘤源DNA源自擴增單倍型。

與LOH情形類似，大小分佈之差異可量化為兩群DNA分子之累積頻率之差。將ΔQ定義為該兩群之累積頻率之差。

$\Delta Q = Q_{擴增} - Q_{非擴增}$，其中$Q_{擴增}$代表來自擴增單倍型之測序DNA片段之累積頻率；且$Q_{非擴增}$代表來自非擴增單倍型之測序DNA片段之累積頻率。

圖28顯示根據本發明之實施例，對於擴增區域，ΔQ對經測序血漿DNA之大小。根據本發明之實施例，在126 bp之大小下，ΔQ達到0.08。此指示使用126 bp作為截止值來界定短DNA時，與來自非擴增單倍型之群相比，來自擴增單倍型之群之短DNA分子之豐度高8%。

D.其他技術

在其他實施例中，可使用序列特異性技術。舉例而言，寡核苷酸可經設計以與特定區域之片段雜交。然後可以與經測序標籤計數類似之方式對寡核苷酸計數。此方法可用於展現特定變異之癌症。

VIII.電腦系統

本文所提及任一電腦系統皆可利用任一適宜數目之子系統。此等子系統之實例顯示於圖29中之電腦裝置900中。在一些實施例中，電腦系統包括單一電腦裝置，其中子系統可為電腦裝置之組件。在其他實施例中，電腦系統可包括多個具有內部組件之電腦裝置，各裝置為子系統。

顯示於圖29中之子系統經由系統匯流排2975互連。顯示

5

其他子系統，例如列印機2974、鍵盤2978、固定磁碟2979、監視器2976(其耦聯至顯示適配器2982)及其他。可藉由任一數目之業內已知構件(例如串聯埠2977)將耦聯至I/O控制器2971之周邊設備及輸入/輸出(I/O)器件連接至電腦系統。舉例而言，可使用串聯埠2977或外部介面2981(例如乙太網路、Wi-Fi等)將電腦系統2900連接至廣域網路(例如網際網路)、鼠標輸入器件或掃描儀。經由系統匯流排2975之互連使中央處理器2973與各子系統通信並控制來自系統記憶體2972或固定磁碟2979之指令之執行、以及子系統間之資訊交換。系統記憶體2972及/或固定磁碟2979可體現電腦可讀取媒體。本文所提及任一值皆可自一個組件輸出至另一組件且可輸出至使用者。

電腦系統可包括例如藉由外部介面2981或藉由內部介面連接在一起之複數個相同組件或子系統。在一些實施例中，電腦系統、子系統或裝置可經由網路通信。在此等情形下，可將一台電腦視為用戶端且將另一電腦視為伺服器，其中每一者可為相同電腦系統之一部分。用戶端及伺服器各自可包括多個系統、子系統或組件。

應理解，可使用硬體及/或使用電腦軟體以控制邏輯形式以模組或整合方式來實作本發明之任一實施例。基於本文所提供之揭示及教示內容，熟習此項技術者應認識並瞭解使用硬體及硬體與軟體之組合實作本發明實施例之其他方式及/或方法。

本申請案中所述之任一軟體組件或功能皆可以欲藉由處

理器執行之軟體程式碼形式實作，其中使用任一適宜電腦語言(例如，Java、C++或Perl)且使用(例如)習用或物件導向技術。軟體程式碼可以一系列指令或命令之形式儲存於用於儲存及/或傳輸之電腦可讀取媒體上，適宜媒體包括隨機存取記憶體(RAM)、唯讀記憶體(ROM)、磁性媒體(例如硬碟或軟碟)或光學媒體(例如光碟(CD)或DVD(數位多功能碟))、快閃記憶體及諸如此類。電腦可讀取媒體可為此等儲存或傳輸器件之任一組合。

此等程式亦可經編碼並使用適於經由遵循多種協議之有線網路、光纖網路及/或無線網路(包括網際網路)傳輸之載波信號傳輸。因此，本發明之實施例之電腦可讀取媒體可使用經此等程式編碼之數據信號產生。經程式碼編碼之電腦可讀取媒體可與相容器件封裝在一起或與其他器件分開提供(例如，經由網際網路下載)。任一此類電腦可讀取媒體皆可駐留於單一電腦程式產品(例如硬蝶、CD或整個電腦系統)之上或之內，且可存在於系統或網路內不同電腦程式產品之上或之內。電腦系統可包括監視器、列印機或用於將本文所提及任一結果提供給使用者之另一適宜顯示器。

本文所述任一方法皆可利用可經組態以實施該等步驟之電腦系統(包括處理器)來完全地或部分地實施。因此，實施例可係關於經組態以實施本文所述任一方法之步驟之電腦系統，可能其中不同組件實施各別步驟或各別組之步驟。儘管以編號步驟提供，但本文方法之步驟可同時或以

不同順序實施。另外，該等步驟之部分可與來自其他方法之其他步驟之部分一起使用。同樣，步驟之全部或部分可係可選的。另外，任一方法之任一步驟皆可利用模組、電路或用於實施該等步驟之其他構件來實施。

特定實施例之具體細節可以任一適宜方式組合，此並不背離本發明實施例之精神及範圍。然而，本發明之其他實施例可係關於與各個別態樣或該等個別態樣之具體組合有關之具體實施例。

已出於闡釋及說明目的提供上文對本發明實例性實施例之說明。並不意欲進行窮盡列舉或將本發明限於所述確切形式，且可根據上文教示內容進行許多修改及改變。該等實施例經選擇及闡述以最佳地解釋本發明之原理及其實際應用，藉以使其他熟習此項技術者能夠以適合於所構想特定應用之各種實施例形式及各種修改來最佳地利用本發明。

除非明確指明相反情況，否則所列舉「一(a, an)」或「該」意欲指「一或多者」。上文所提及之所有專利、專利申請案、出版物及說明皆出於所有目的以全文引用方式併入本文中。不承認任一者為先前技術。

【圖式簡單說明】

圖1繪示癌細胞中展現缺失變異之染色體區域。

圖2繪示癌細胞中展現擴增變異之染色體區域。

圖3顯示表300，其繪示不同類型之癌症及有關區域及其對應變異。

圖4繪示根據本發明之實施例，癌細胞內不展現變異之染色體區域以及在血漿中進行之量測。

圖5繪示根據本發明之實施例，癌細胞內染色體區域510之缺失以及在血漿中進行之確定缺失區域之量測。

圖6繪示根據本發明之實施例，癌細胞內染色體區域610之擴增以及在血漿中進行之確定擴增區域之量測。

圖7顯示根據本發明之實施例，在肝細胞癌(HCC)患者之血漿DNA中對腫瘤組織中位於染色體1p處且顯示單對偶基因擴增之區段的RHDO分析。

圖8顯示根據本發明之實施例，在存在含有缺失之腫瘤時染色體區域之兩種單倍型之片段的大小分佈變化。

圖9顯示根據本發明之實施例，在存在含有擴增之腫瘤時染色體區域之兩種單倍型之片段的大小分佈變化。

圖10係繪示根據本發明之實施例，分析有機體之生物學試樣之單倍型，以確定染色體區域是否展現缺失或擴增之方法的流程圖。

圖11顯示根據本發明之實施例，癌細胞中缺失亞區1130之區域1110以及在血漿中進行之確定缺失區域之量測。

圖12顯示根據本發明之實施例，可如何使用RHDO分析定位變異之位置。

圖13顯示根據本發明之實施例，始於另一方向之RHDO分類。

圖14係根據本發明之實施例，使用複數個染色體區域分析有機體之生物學試樣之方法1400的流程圖。

圖15顯示表1500，其繪示根據本發明之實施例，不同區段數所需之深度及腫瘤源片段之分數濃度。圖15提供對欲分析癌症源DNA在試樣中之分數濃度之不同百分比之分子數目的估計。

圖16顯示根據本發明之實施例，藉由相對單倍型劑量(RHDO)分析量測腫瘤源DNA在血漿中之分數濃度之原理。Hap I及Hap II代表根據本發明之實施例非腫瘤組織中之兩種單倍型。

圖17係繪示根據本發明之實施例，使用包括核酸分子之生物學試樣確定有機體中染色體變異之進展之方法的流程圖。

圖18A顯示對癌症患者之染色體4之q臂上之區段實施RHDO分析的SPRT曲線。各點代表在各別異型接合基因座後之累積計數比率。圖18B顯示對治療後患者之染色體4之q臂上之區段實施RHDO分析的SPRT曲線。

圖19顯示HCC中發現之常見染色體變異。

圖20A顯示使用靶向分析獲得之HCC患者與健康患者之正規化標籤計數比率之結果。圖20B顯示在靶富集及大規模平行測序後，3名HCC患者及4名健康對照個體之大小分析之結果。

圖21顯示HCC患者之Circos圖，其繪示根據本發明之實施例，來自血漿DNA之經測序標籤計數之數據。

圖22顯示根據本發明之實施例，對無HCC之慢性肝炎B病毒(HBV)攜帶者之血漿試樣之經測序標籤計數分析。

圖23顯示根據本發明之實施例，對患有階段3鼻咽癌(NPC)之患者之血漿試樣的經測序標籤計數分析。

圖24顯示根據本發明之實施例，對患有階段4 NPC之患者之血漿試樣的經測序標籤計數分析。

圖25顯示根據本發明之實施例，血漿DNA之累積頻率對腫瘤組織中展現異型接合性喪失(LOH)之區域之大小之曲線。

圖26顯示對於LOH區域，ΔQ對經測序血漿DNA之大小。根據本發明之實施例，在130 bp之大小下，ΔQ達到0.2。

圖27顯示根據本發明之實施例，血漿DNA之累積頻率對腫瘤組織中具有染色體倍增之區域之大小之曲線。

圖28顯示根據本發明之實施例，對於擴增區域，ΔQ對經測序血漿DNA之大小。

圖29顯示可與本發明之實施例之系統及方法一起使用之實例性電腦系統900之方塊圖。

【主要元件符號說明】

2971	輸入/輸出控制器
2972	系統記憶體
2973	中央處理器
2974	列印機
2975	系統匯流排
2976	監視器
2977	串聯埠

5

2978	鍵盤
2979	固定磁碟
2981	外部介面
2982	顯示適配器

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：100144098

※ 申請日： 100.11.30　　※IPC 分類：C12Q 1/68 (2006.01)
G06F 19/18 (2011.01)
G06F 19/20 (2011.01)

一、發明名稱：（中文/英文）

與癌症有關之基因或分子變異之檢測

DETECTION OF GENETIC OR MOLECULAR ABERRATIONS ASSOCIATED WITH CANCER

二、中文發明摘要：

本發明提供測定有機體之生物學試樣中之基因或分子變異的系統、裝置及方法。分析包括無細胞之DNA片段之生物學試樣以鑑別腫瘤中之染色體區域中之不均衡，例如因缺失及/或擴增所致。對於各染色體區域，使用多個基因座。然後可使用此等不均衡來診斷(篩選)癌症患者，以及預測癌症患者，或檢測患者癌變前病況之存在，或監測患者癌變前病況之進展。可使用不均衡之嚴重程度以及展現不均衡之區域之數目。基因組之非重疊區段之系統分析可提供試樣之一般篩選工具。另外，可隨時間測試患者以追蹤一或多個染色體區域之每一者之嚴重程度及染色體區域之數目以能夠篩選及預測以及監測進展(例如在治療後)。

三、英文發明摘要：

Systems, apparatus, and methods are provided for determining genetic or molecular aberrations in a biological sample from an organism. Biological samples including cell-free DNA fragments are analyzed to identify imbalances in chromosomal regions, e.g., due to deletions and/or amplifications in a tumor. Multiple loci are used for each chromosomal region. Such imbalances can then be used to diagnose (screen) a patient for cancer, as well as prognosticate a patient with cancer, or to detect the presence or to monitor the progress of a premalignant condition in a patient. The severity of an imbalance as well as the number of regions exhibiting an imbalance can be used. A systematic analysis of non-overlapping segments of a genome can provide a general screening tool for a sample. Additionally, a patient can be tested over time to track severity of each of one or more chromosomal regions and a number of chromosomal regions to enable screening and prognosticating, as well as monitoring of progress (e.g. after treatment)

四、指定代表圖：

(一)本案指定代表圖為：第（ 10 ）圖。

(二)本代表圖之元件符號簡單說明：

(無元件符號說明)

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

(無)

七、申請專利範圍：

1. 一種分析有機體之生物學試樣中與癌症有關之染色體缺失或擴增的方法，該生物學試樣包括源自正常細胞及可能源自與癌症有關細胞之核酸分子，其中在該生物學試樣中至少一些該等核酸分子係無細胞的，該方法包含：

測定該有機體之正常細胞於第一染色體區域之第一及第二單倍型，該第一染色體區域包括第一複數基因座，其中該第一及第二單倍型於該第一複數基因座中之每一者係異型接合(heterozygous)；

對於該生物學試樣中多數該等核酸分子之每一者：

鑑別該核酸分子在該有機體之參照基因組中之位置；及

測定該核酸分子之各別對偶基因；

基於所鑑別位置及所測定對偶基因將第一組核酸分子鑑別為來自該第一單倍型，該第一組包括至少一個核酸分子位於該第一複數基因座中每一者；

基於所鑑別位置及所測定對偶基因將第二組核酸分子鑑別為來自該第二單倍型，該第二組包括至少一個核酸分子位於該第一複數基因座中每一者；

利用電腦系統計算該第一組核酸分子之第一值，該第一值界定該第一組核酸分子之性質；

利用該電腦系統計算該第二組核酸分子之第二值，該第二值界定該第二組核酸分子之性質；及

比較該第一值與該第二值以確定該第一染色體區域在

與癌症有關之任何細胞中是否展現缺失或擴增之分類。

2. 如請求項1之方法，其中將該第一染色體區域分類為在與癌症有關之細胞中展現缺失或擴增，且其中該等與癌症有關之細胞包括惡性腫瘤及/或來自癌變前病變。

3. 如請求項1之方法，其中該第一複數基因座中之每一基因座與該第一複數基因座之另一基因座相距至少500個鹼基。

4. 如請求項1之方法，其中該比較包括測定該第一值與該第二值間之差或比率，且其中該分類包括比較該差或該比率與至少一個閾值。

5. 如請求項4之方法，其中該至少一個閾值係得自健康有機體或不具有缺失或擴增之區域，其中使用逐次機率比測試、t測試或卡方測試(chi-square test)來測定該至少一個閾值。

6. 如請求項1之方法，其中該第一值對應於該第一組核酸分子之大小分佈的統計值，且該第二值對應於該第二組核酸分子之大小分佈的統計值。

7. 如請求項6之方法，其中該第一值係該第一組核酸分子之平均大小，且該第二值係該第二組核酸分子之平均大小。

8. 如請求項6之方法，其中該第一值Q_{HapI}係該第一組中短於截止大小(cutoff size)之核酸分子之分數，且該第二值Q_{HapII}係該第二組中短於該截止大小之核酸分子之分數。

9. 如請求項8之方法，其中該比較包括測定該第一值與該

S

第二值間之差，其中該差係$\Delta Q=Q_{HapI}-Q_{HapII}$，且其中正$\Delta Q$值大於閾值指示該第二單倍型在該有機體中之腫瘤組織中包括缺失或該第一單倍型在該有機體中之腫瘤組織中包括擴增。

10. 如請求項8之方法，其中該比較包括測定該第一值與該第二值間之差，其中該差係$\Delta Q=Q_{HapI}-Q_{HapII}$，且其中$\Delta Q$近似值為0指示在該第一染色體區域中不存在缺失或擴增。

11. 如請求項6之方法，其中對於各別單倍型，將該第一值$F_{Hap\ I}$及該第二值$F_{Hap\ II}$定義為$F=\sum^{w}$長度$/\sum^{N}$長度，其中$\sum^{w}$長度代表對應組中長度等於或小於截止大小w之核酸分子之長度的總和；且$\sum^{N}$長度代表該對應組中長度等於或小於N個鹼基之核酸分子之長度的總和，其中N大於w。

12. 如請求項11之方法，其中該比較包括測定該第一值與該第二值間之差，且其中該差係$\Delta F=F_{Hap\ I}-F_{Hap\ II}$，且其中正$\Delta F$值大於閾值指示該第二單倍型在該有機體之腫瘤組織中包括缺失。

13. 如請求項11之方法，其中該比較包括測定該第一值與該第二值間之差，且其中該差係$\Delta F=F_{Hap\ I}-F_{Hap\ II}$，且其中負$\Delta F$值大於閾值指示該第二單倍型在該有機體之腫瘤組織中包括擴增。

14. 如請求項1之方法，其中該第一組之第一值對應於位於該第一複數基因座之核酸分子之數目，且該第二組之第

二值對應於位於該第一複數基因座之核酸分子之數目。

15. 如請求項14之方法，其進一步包含：

計算該第一值與該第二值之比率以測定癌症DNA在該生物學試樣中之分數濃度。

16. 如請求項15之方法，其進一步包含：

於多個時間測定癌症DNA在該生物學試樣中之分數濃度；及

使用該多個時間之分數濃度來對該有機體之癌症等級之進展進行診斷、分期、預測或監測。

17. 如請求項1之方法，其進一步包含：

對於多個其他染色體區域中之每一者：

重複如請求項1之方法；

測定展現缺失或擴增之染色體區域之第一數目；及

比較該第一數目與一或多個閾值以確定該有機體之癌症等級。

18. 如請求項17之方法，其進一步包含：

測定經鑑別為展現缺失或擴增之各染色體區域之缺失或擴增之量；及

比較該量與一或多個閾值以確定該有機體之該癌症等級。

19. 如請求項17之方法，其進一步包含：

於多個時間重複如請求項17之方法；及

使用該多個時間之第一數目來對該有機體之癌症等級之進展進行診斷、分期、預測或監測。

S

20. 如請求項19之方法，其中使用該多個時間之第一數目對該有機體之癌症等級之進展進行診斷、分期、預測或監測包括測定癌變前病況之存在或進展。

21. 如請求項17之方法，其中該等染色體區域中之每一者具有預定長度。

22. 一種分析有機體之生物學試樣之方法，該生物學試樣包括源自正常細胞及可能源自與癌症有關細胞之核酸分子，其中在該生物學試樣中至少一些該等核酸分子係無細胞的，該方法包含：

鑑別該有機體之多個非重疊染色體區域，各染色體區域包括複數基因座；

對於該有機體之生物學試樣中多數核酸分子之每一者：

鑑別該核酸分子在該有機體之參照基因組中之位置；

對於各別之該多個非重疊染色體區域：

基於所鑑別位置將各別組核酸分子鑑別為來自該染色體區域，該各別組包括至少一個核酸分子位於該染色體區域之該複數基因座中每一者；

利用電腦系統計算該各別組核酸分子之各別值，該各別值界定該各別組之核酸分子之性質；及

比較該各別值與參照值以確定該染色體區域是否展現缺失或擴增之分類；

測定分類為展現缺失或擴增之染色體區域之量；及

比較該量與閾值以確定該有機體是否患有癌症之分類。

23. 如請求項22之方法，其中該等染色體區域中之每一者具有預定長度。

24. 如請求項22之方法，其中該多個非重疊染色體區域跨越該有機體之基因組。

25. 如請求項22之方法，其中僅在染色體區域包含於至少一個指定數目之具有超過參照值之各別值的連續染色體區域中，而指定數目為2或更多時，該染色體區域分類為展現缺失或擴增。

26. 如請求項22之方法，其中該閾值具有至少1之值，且其中比較該量與閾值包括決定該量是否大於閾值。

27. 一種使用包括源自正常細胞及可能源自與癌症有關細胞之核酸分子的生物學試樣來測定有機體中染色體變異之進展之方法，其中在該等生物學試樣中至少一些該等核酸分子係無細胞的，該方法包含

鑑別該有機體之參照基因組之一或多個非重疊染色體區域，各染色體區域包括複數基因座；

於多個時間之每一時間：

對於該有機體之生物學試樣中多數核酸分子之每一者：

鑑別該核酸分子在該參照基因組中之位置；

對於各別之該一或多個非重疊染色體區域：

基於所鑑別位置將各別組核酸分子鑑別為來自該染色體區域，該各別組包括至少一個核酸分子位於

S

該染色體區域之該複數基因座中每一者；

利用電腦系統計算該各別組核酸分子之各別值，該各別值界定該各別組之核酸分子之性質；

比較該各別值與參照值以確定該第一染色體區域是否展現缺失或擴增之分類；及

藉由以下使用該多個時間中之該等一或多個染色體區域之一或多個分類來測定該等染色體變異在該有機體中之進展：

計算從一時間至另一時間之分類之數值變化；及

比較該數值變化與閾值。

28. 如請求項27之方法，其中該分類係反映各時間各染色體區域之該各別值與該參照值間之差的數值。

29. 如請求項27之方法，其中一或多個非重疊染色體區域係多個染色體區域。

30. 如請求項29之方法，其進一步包含：

於該多個時間之每一時間：

測定分類為展現缺失或擴增之染色體區域之量；且其中使用該等量來監測癌症在該有機體中之進展。

31. 如請求項27之方法，其中該參照值係得自鑑別為來自參照染色體區域之一組核酸分子，該參照染色體區域在獲得該等核酸分子時不展現缺失或擴增。

32. 如請求項27之方法，其進一步包含：

測定該有機體之正常細胞於該一或多個非重疊染色體區域之第一染色體區域的第一及第二單倍型，該第一染

色體區域包括第一複數基因座，其中該兩種單倍型於該第一複數基因座之每一者係異型接合，

其中該第一染色體區域之該各別值係得自鑑別為來自該第一單倍型之第一組核酸分子，且

其中該第一染色體區域之該參照值係得自鑑別為來自該第二單倍型之第二組核酸分子。

33. 一種電腦程式產品，其包含電腦可讀取媒體，該電腦可讀取媒體儲存多數用於控制處理器以實施操作本文所述方法之指令，該等指令包含如請求項1至32中任一項之步驟。

34. 一種電腦系統，其包括經程式化以實施如請求項1至32中任一項之方法之任一者之處理器。

S