



(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 I636255 B

(45)公告日：中華民國 107 (2018) 年 09 月 21 日

(21)申請案號：102122036

(22)申請日：中華民國 102 (2013) 年 06 月 20 日

(51)Int. Cl. : **G01N33/50 (2006.01)****C40B20/04 (2006.01)****G06F17/18 (2006.01)****G06F19/18 (2011.01)**

(30)優先權：2012/06/21 美國

61/662,878

2012/08/13 美國

61/682,725

2012/08/31 美國

61/695,795

2012/10/08 美國

61/711,172

2013/03/13 美國

13/801,748

(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)；趙慧君 CHIU, ROSSA WAI KWUN
(AU)；陳君賜 CHAN, KWAN CHEE (HK)；江培勇 JIANG, PEIYONG (CN)

(74)代理人：陳長文

(56)參考文獻：

TW 200700558A1

WO 2012071621A1

審查人員：林桂滿

申請專利範圍項數：70 項 圖式數：25 共 112 頁

(54)名稱

癌症檢測之血漿DNA突變分析

MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION

(57)摘要

可比較經歷癌症篩選或監測之個體之生物試樣(例如血漿或血清)中與同一個體之組成DNA中的體細胞突變頻率。可自該等頻率衍生一參數且用於測定癌症等級之分類。可由要求任何變體基因座具有至少一個指定數目之變體序列讀段(reads)(標籤)來濾出偽陽性，藉此提供更正確參數。可分析不同變體基因座之相對頻率以測定患者中腫瘤之異質性程度。

A frequency of somatic mutations in a biological sample (e.g., plasma or serum) of a subject undergoing screening or monitoring for cancer, can be compared with that in the constitutional DNA of the same subject. A parameter can derived from these frequencies and used to determine a classification of a level of cancer. False positives can be filtered out by requiring any variant locus to have at least a specified number of variant sequence reads (tags), thereby providing a more accurate parameter. The relative frequencies for different variant loci can be analyzed to determine a level of heterogeneity of tumors in a patient.

指定代表圖：

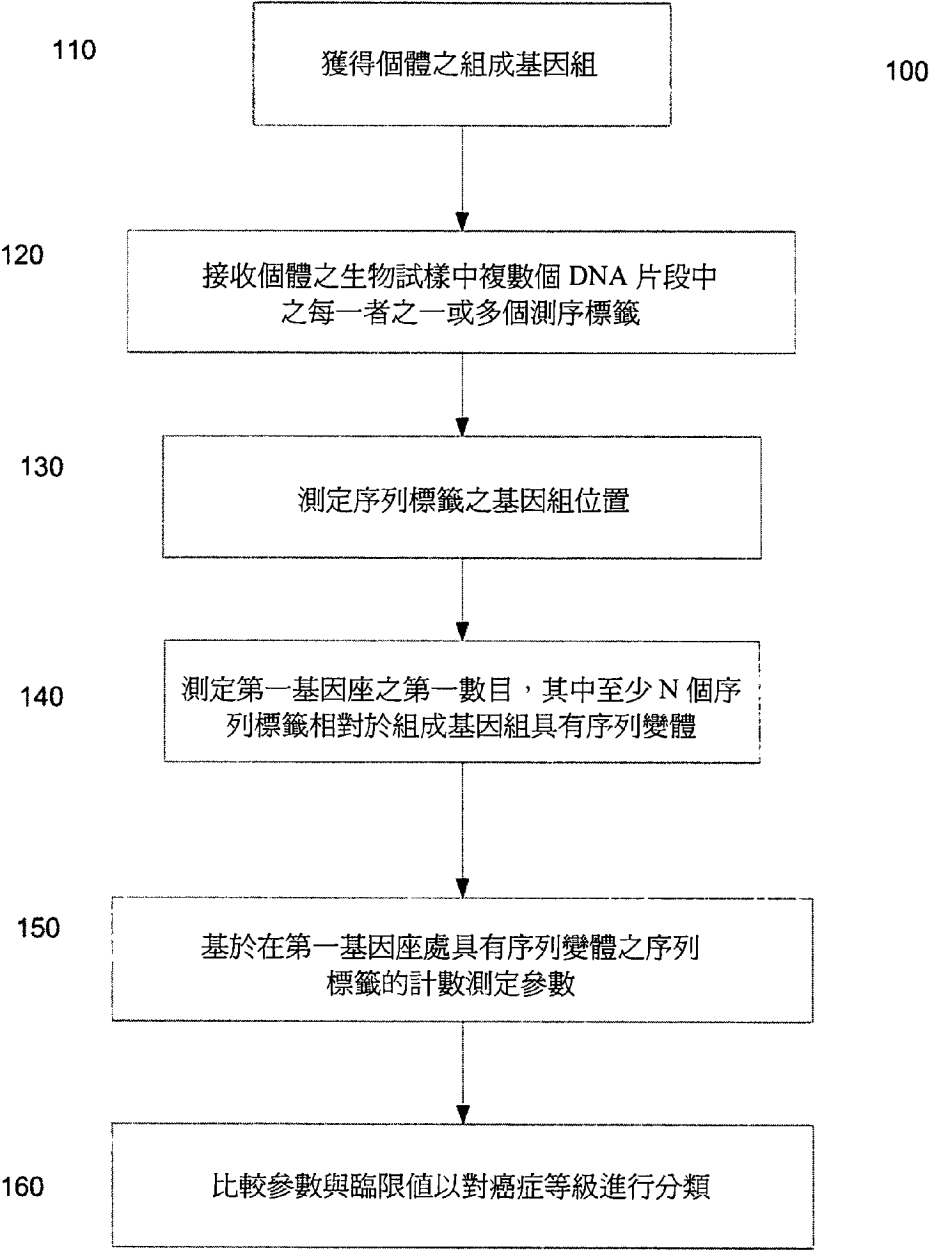


圖 1

發明專利說明書

(本說明書格式、順序，請勿任意更動)

【發明名稱】

癌症檢測之血漿DNA突變分析

MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER
DETECTION

相關申請案之交叉參照

本申請案係非臨時申請案且主張以下申請案之權益：於2012年6月21日提出申請之標題為「MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION」之美國臨時專利申請案第61/662,878號；於2012年8月13日提出申請之標題為「MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION」之美國臨時專利申請案第61/682,725號；於2012年8月31日提出申請之標題為「MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION」之美國臨時專利申請案第61/695,795號；及於2012年10月8日提出申請之標題為「MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION」之美國臨時專利申請案第61/711,172號，該等案件出於所有目的全文以引用方式併入本文中。

【先前技術】

已顯示腫瘤源DNA存於癌症患者之無細胞之血漿/血清中(Chen XQ等人 Nat Med 1996; 2: 1033-1035)。大多數當前方法係基於已知與癌症相關突變之直接分析(Diehl F等人 Proc Natl Acad Sci 2005; 102: 16368-16373; Forsheo T等人 Sci Transl Med 2012; 4: 136ra68)。另一方法已研究藉由對血漿DNA進行隨機測序檢測之癌症相關之拷貝數差異(Lo等人之美國專利公開案2013/0040824)。

已知隨時間流逝，一個以上癌細胞可獲得生長且產生多個子細胞純系。最後，腫瘤生長及/或其轉移病灶可含有一簇純系癌細胞群。此現象通常稱作腫瘤異質性(Gerlinger M等人 N Engl J Med 2012 ; 366: 883-892 ; Yap TA 等人 Sci Transl Med 2012 ; 4: 127ps10)。

已知癌症高度異質，即相同組織類型之癌症之突變概況可廣泛變化。因此，特異性突變之直接分析通常可僅檢測已知與彼等特異性突變相關之特定癌症類型內之病例之亞組。另外，腫瘤源DNA通常係人類血漿中次要物種之DNA；血漿中之DNA之絕對濃度較低。因此，即使在患有已知具有靶向突變之癌症的患者中，血漿或血清中一個或一小群癌症相關突變的直接檢測亦可達成低分析靈敏度。此外，已顯示甚至在單一腫瘤內，在突變方面存在顯著瘤內異質性。可僅在腫瘤細胞子群中發現突變。原發性腫瘤與轉移性病灶之間之突變概況差異甚至更大。瘤內及原發性轉移異質性之一個實例涉及患有結腸直腸癌之患者中之*KRAS*、*BRAF*及*PIK3CA*基因(Baldus等人 Clin Cancer Research 2010. 16:790-9.)。

在患者患有原發性腫瘤(帶有*KRAS*突變而非*PIK3CA*突變)及隱藏之轉移性病灶(帶有*PIK3CA*突變而非*KRAS*突變)之情況下，若主要關注原發性腫瘤中之*KRAS*突變之檢測，則無法檢測隱藏之轉移性病灶。然而，若在分析中包括兩個突變，則可檢測原發性腫瘤及隱藏之轉移性病灶二者。因此，涉及兩個突變之測試可在殘餘腫瘤組織之檢測中具有較高靈敏度。在篩選癌症時，及在對可能發生之突變類型瞭解極少或毫無所知時，該簡單實例變得更複雜。

因此，期望提供對癌症實施廣泛篩選、檢測或評定之新技術。

【發明內容】

實施例可觀察經歷癌症篩選或監測之個體之生物試樣(例如，血

漿或血清)中的體細胞突變頻率，此時與相同個體之組成DNA中之體細胞突變頻率進行比較。可使用隨機測序來測定該等頻率。參數可源自該等頻率且用於測定癌症等級之分類。可藉由需要任何變體基因座具有至少指定數目之變體序列讀段(標籤)濾出偽陽性，藉此提供更精確參數。可分析不同變體基因座之相對頻率以測定患者中之腫瘤之異質性程度。

在一個實施例中，可將該參數與源自一組無癌症或處於低癌症風險之個體的相同參數進行比較。自測試個體獲得之參數與自無癌症或處於低癌症風險之個體之群組獲得之參數的顯著差異可指示測試個體患有癌症或癌變前病況或將來可發生癌症之風險增加。因此，在一個實施例中，血漿DNA分析可在無腫瘤之先前基因組資訊情況下進行。因此，該實施例尤其可用於篩選癌症。

在另一實施例中，實施例亦可用於監測治療後之癌症患者並觀察是否存在殘餘腫瘤或腫瘤是否復發。舉例而言，具有殘餘腫瘤或腫瘤復發之患者較無殘餘腫瘤或未觀察到腫瘤復發之患者可具有較高頻率之體細胞突變。監測可涉及在治療後之多個時間點時自癌症患者獲得試樣以確定體液或具有無細胞之核酸之其他試樣(例如血漿或血清)中之腫瘤相關之遺傳畸變的時間變化。

根據一個實施例，方法檢測個體中之癌症或癌變前變化。獲得個體之組成基因組。接收個體之生物試樣中複數個DNA片段中之每一者之一或多個序列標籤，其中生物試樣包括無細胞DNA。測定序列標籤之基因組位置。比較序列標籤與組成基因組以測定第一基因座之第一數目。在每一第一基因座處，相對於組成基因組具有序列變體之序列標籤的數目高於截止值，其中截止值大於1。基於在第一基因座處具有序列變體之序列標籤的計數測定參數。將比較參數與臨限值以測定個體之癌症等級之分類。

根據另一實施例，方法分析個體之一或多種腫瘤之異質性。獲得個體之組成基因組。接收個體之生物試樣中複數個DNA片段中之每一者之一或多個序列標籤，其中生物試樣包括無細胞DNA。測定序列標籤之基因組位置。比較序列標籤與組成基因組以測定第一基因座之第一數目。在每一第一基因座處，相對於組成基因組具有序列變體之序列標籤的數目高於截止值，其中截止值大於1。基於該組第一基因組位置之各別第一數目計算一或多種腫瘤之異質性量度。

根據另一實施例，方法測定包括無細胞DNA之生物試樣中之腫瘤DNA的分數濃度。接收生物試樣中複數個DNA片段中之每一者之一或多個序列標籤。測定序列標籤之基因組位置。對於複數個基因組區域中之每一者，自基因組位置在基因組區域內之序列標籤測定基因組區域內之DNA片段之各別量。各別量經正規化以獲得各別密度。比較各別密度與參照密度以鑑別基因組區域是否呈現1個拷貝喪失或1個拷貝增加。根據鑑別為呈現1個拷貝喪失之各別密度或根據鑑別為呈現1個拷貝增加之各別密度計算第一密度。藉由比較第一密度與另一密度以獲得差來計算分數濃度，其中利用參照密度正規化該差異。

其他實施例係關於與本文所述方法相關之系統及電腦可讀取媒體。

可參照以下詳細說明及附圖獲得對本發明之性質及優點之較佳理解。

【圖式簡單說明】

圖1係根據本發明實施例檢測個體之癌症或癌變前變化之方法100的流程圖。

圖2顯示根據本發明實施例直接比較試樣基因組(SG)與組成基因組(CG)的方法之流程圖。

圖3顯示根據本發明實施例使用參照基因組(RG)比較試樣基因組

(SG)與組成基因組(CG)的方法300之流程圖。

圖4係表400，其顯示根據本發明實施例，在試樣中之腫瘤源DNA之分數濃度假定為10%時，使用不同發生次數作為將突變分類為存於試樣中之準則正確鑑別的癌症相關單核苷酸突變的數目。

圖5係表500，其顯示在試樣中之腫瘤源DNA之分數濃度假定為5%時鑑別之偽陽性基因座之預計數目及突變之預計數目。

圖6A係圖表600，其顯示在具有10%及20%血漿分數濃度之腫瘤源DNA之血漿中，使用4次及6次發生次數(r)作為稱為潛在癌症相關突變之準則檢測到癌症相關突變之比率。圖6B係圖表650，其相對於測序深度顯示使用4、5、6及7次發生次數(r)之準則誤分類為具有核苷酸變化之核苷酸位置的預計數目。

圖7A係圖表700，其顯示在試樣中之腫瘤源DNA之分數濃度假定為5%時，具有不同測序深度之癌症相關突變位點及偽陽性位點的數目。圖7B係圖表750，其顯示偽陽性位點之預測數目且涉及對全基因組(WG)及所有外顯子之分析。

圖8係表800，其顯示根據本發明實施例，4名HCC患者在治療之前及之後之結果，包括血漿中之腫瘤源DNA之分數濃度。

圖9係表900，其顯示根據本發明實施例，在16名健康對照個體中對HCC相關之SNV之檢測。

圖10A顯示根據本發明實施例，HCC患者之腫瘤試樣之序列讀段密度的分佈曲線。圖10B顯示根據本發明實施例，HCC患者之血漿中所有庫(bin)之 z 分值的分佈曲線1050。

圖11顯示根據本發明實施例，HCC患者之血漿之 z 分值的分佈曲線1100。

圖12係根據本發明實施例測定包括無細胞DNA之生物試樣中腫瘤DNA的分數濃度之方法1200的流程圖。

圖13A顯示根據本發明實施例，在診斷時分析患有卵巢癌及乳癌之患者血漿中之突變之表1300。

圖13B顯示根據本發明實施例，在腫瘤切除後分析患有雙側卵巢癌及乳癌之患者血漿中之突變之表1350。

圖14A係表1400，其顯示HCC1之血漿DNA中之單核苷酸變異的檢測。圖14B係表1450，其顯示HCC2之血漿DNA中之單核苷酸變異的檢測。

圖15A係表1500，其顯示HCC3之血漿DNA中之單核苷酸變異的檢測。圖15B係表1550，其顯示HCC4之血漿DNA中之單核苷酸變異的檢測。

圖16係表1600，其顯示患有卵巢癌(及乳癌)之患者之血漿DNA中之單核苷酸變異的檢測。

圖17係表1700，其顯示不同發生次數要求及測序深度之預測靈敏度。

圖18係表1800，其顯示不同截止值及不同測序深度之偽陽性基因座的預測數目。

圖19顯示圖解說明在不同腫瘤位點中檢測到之突變數目的樹狀圖。

圖20係表2000，其顯示治療前及治療後血漿試樣中帶有腫瘤源突變之片段的數目。

圖21係圖表2100，其顯示在單一腫瘤位點中檢測到之突變及在所有四個腫瘤位點中檢測到之突變在血漿中之發生次數的分佈。

圖22係圖表2200，其顯示來自異質腫瘤之突變在血漿中之發生次數的預測分佈。

圖23展示募集16名健康對照個體之實施例之特異性。

圖24係根據本發明實施例分析個體之一或多種腫瘤之異質性的

方法2400之流程圖。

圖25顯示可與本發明實施例之系統及方法一起使用之實例性電腦系統2500之方塊圖。

【實施方式】

定義

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。「庫(*bin*)」係基因組中之預定長度的區域。複數個庫可具有相同第一長度(解析度)，而不同複數個可具有相同第二長度。在一個實施例中，庫彼此並不重疊。

本文所用術語「*隨機測序*」係指測序，藉此在測序程序之前尚未特異性鑑別或預定測序之核酸片段。不需用以靶向特異性基因座之序列特異性引物。術語「*通用測序*」係指測序可始於任何片段之測序。在一個實施例中，將銜接體添加至片段末端，且將測序用引物附接至銜接體。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。

本文所用術語「*序列標籤*」(亦稱作序列讀段)係指自任何部分或全部核酸分子測序之核苷酸串。舉例而言，測序標籤可為自核酸片段測序之短的核苷酸串(例如，約30)、核酸片段兩端處之短的核苷酸串或生物試樣中存在之整個核酸片段的測序。核酸片段係較大核酸分子之任一部分。片段(例如基因)可與較大核酸分子之其他部分分開地(即未連接)存在。

術語「*組成基因組*」(亦稱作CG)包括基因組內之基因座處之共有核苷酸，且因此可視為共有序列。CG可覆蓋個體之全基因組(例如，人類基因組)，或僅覆蓋基因組之部分。組成基因組(CG)可自細胞之DNA以及無細胞DNA獲得(例如，如可在血漿中發現)。理想地，

共有核苷酸應指示基因座對於一個對偶基因同型或對於兩個對偶基因異型。異型基因座通常含有兩個作為遺傳多態性之成員之對偶基因。作為實例，用於測定基因座是否異型之準則可為兩個對偶基因之臨限值，該兩個對偶基因各自以至少預定百分比(例如，30%或40%)之與基因座比對之讀段出現。若一個核苷酸以足夠百分比(例如，70%或更大)出現，則基因座可測定為在CG中同型。儘管由於在細胞分裂期間自發出現隨機突變，一個健康細胞之基因組可不同於另一健康細胞之基因組，但在使用該共有基因組時，CG應不變化。一些細胞可具有含基因組重排之基因組，例如，B及T淋巴球，例如涉及抗體及T細胞受體基因。該等大規模差異仍可為血液中之總有核細胞群體之相對較小群體，且因此該等重排將不會影響血細胞足夠取樣(例如，測序深度)之組成基因組之測定。其他細胞類型(包括頰細胞、皮膚細胞、毛囊或各種正常身體組織之活檢)亦可用作CG之來源。

術語「組成DNA」係指反映個體出生帶有之遺傳組成之DNA的任何來源。對於個體，「組成試樣」(可自其獲得組成DNA)之實例包括健康血細胞DNA、頰細胞DNA及毛根DNA。該等健康細胞之DNA定義個體之CG。例如，在已知人員不患有癌症或可自不可能含有癌症或癌變前細胞(例如，在懷疑肝癌時，毛根DNA)之組織獲得試樣時，細胞可以多種方式鑑別為健康。作為另一實例，可在患者無癌症時獲得血漿試樣，且比較所測定組成DNA與隨後血漿試樣(例如，1年或更長時間後)之結果。在另一實施例中，含有<50%腫瘤DNA之單一生物試樣可用於推論組成基因組及腫瘤相關之遺傳變化。在該試樣中，腫瘤相關之單核苷酸突變之濃度可低於CG中之異型SNP之每一對偶基因的濃度。該試樣可與用於測定下文所述試樣基因組之生物試樣相同。

本文所用術語「生物試樣」係指取自個體(例如，人類、患有癌

症者、懷疑患有癌症者或其他有機體)且含有一或多個所關注無細胞之核酸分子之任何試樣。生物試樣可包括無細胞DNA，其中之一些可源自健康細胞且一些可源自腫瘤細胞。舉例而言，可在血液或其他流體(例如，尿、胸水、腹水、腹膜液、唾液、眼淚或腦脊髓液)中發現腫瘤DNA。非流體實例係糞便試樣，其可與腹瀉液混合。對於一些該等試樣，生物試樣可以非侵入方式獲得。在一些實施例中，生物試樣可用作組成試樣。

術語「試樣基因組」(亦稱作SG)係與基因組(例如，人類基因組)之定位比對之序列讀段之集合。試樣基因組(SG)並非共有序列，但包括僅可在足夠數目之讀段(例如，至少2或3或更高截止值)中出現的核苷酸。若對偶基因出現足夠次數且其並非CG之部分(即，並非共有序列之部分)，則該對偶基因可指示「單核苷酸突變」(亦稱作SNM)。亦可使用本發明檢測其他類型之突變，例如涉及兩個或更多個核苷酸之突變(例如影響微衛星或簡單串聯重複多態性中之串聯重複單元之數目)、染色體易位(其可為染色體內或染色體間)及序列倒置。

術語「參照基因組」(亦稱作RG)係指可與生物試樣及組成試樣之序列讀段比對及比較的單倍體或二倍體基因組。對於單倍體基因組，在每一基因座處僅存在一個核苷酸。對於二倍體基因組，可鑑別異型基因座，該基因座具有兩個對偶基因，其中任一對偶基因可容許與基因座比對匹配。

術語「癌症等級」可指是否存在癌症、癌症階段、腫瘤大小及/或癌症之嚴重性之其他量度。癌症等級可為數值或其他符號。該等級可為0。癌症等級亦包括與突變或突變數目相關之癌變前或癌前病況(狀態)。可以多種方式使用癌症等級。舉例而言，篩選可檢查先前已知無癌症之某人中是否存在癌症。評定可研究已診斷患有癌症之某

人。檢測可意指「篩選」或可意指檢查具有癌症之指示特徵(例如症狀或其他正性測試)的某人是否患有癌症。

提供藉由並非直接取自腫瘤且包括無細胞之核酸的生物試樣(例如，血漿/血清試樣)之分析檢測癌症之實施例。無細胞之核酸可源自全身之各種類型之組織。以此方式，可實施各種癌症之檢測的寬分析。

在癌症發生期間，遺傳畸變(包括單核苷酸突變、缺失、擴增及重排)在腫瘤細胞中累積。在實施例中，可使用大規模平行測序來檢測並定量體液(例如血漿、血清、唾液、腹水、胸水及腦脊髓液)中之單核苷酸突變(SNM) (亦稱作單核苷酸變異(SNV))，以便檢測並監測癌症。SNM (或其他類型突變)之數目之定量可提供鑑別早期癌症作為篩選測試之部分的機制。在各個實作中，小心區分測序誤差並區分健康細胞中發生之自發突變(例如，藉由需要欲在特定基因座處鑑別多個SNM，例如，至少3、4或5個)。

一些實施例亦提供腫瘤異質性之分析之非侵入性方法，其可涉及體內之相同腫瘤內之細胞(即瘤內異質性)或不同腫瘤之細胞(來自相同位點或來自不同位點)。舉例而言，可非侵入性分析該腫瘤異質性之純系結構，包括含有每一突變之相對腫瘤細胞質量之估計。以較高相對濃度存在之突變存於體內較大數目之惡性細胞中，例如，在致腫瘤過程期間相對於仍存於體內之其他惡性細胞更早出現之細胞(Welch JS等人 Cell 2012; 150: 264-278)。該等突變由於其較高相對豐度而預計較彼等具有較低相對豐度者呈現用於檢測癌症DNA之較高診斷靈敏度。突變之相對豐度變化的連續監測可容許非侵入性監測隨著疾病進展或因應治療自發地出現之腫瘤之純系架構變化。該資訊可用於評定預後或用於對治療之腫瘤抗性之早期檢測。

I. 引言

由於DNA複製及/或DNA修復誤差，可在細胞分裂期間發生突變。一類該等突變涉及單核苷酸變化，其可涉及基因組之不同部分之多個序列。通常據信癌症係由於獲得生長優勢之單一癌細胞之純系擴增。此純系擴增可導致在源自祖癌細胞之所有癌細胞中突變(例如單核苷酸突變)累積。該等子代腫瘤細胞可共享一組突變(例如單核苷酸突變)。如本文所述，癌症相關之單核苷酸突變在癌症患者之血漿/血清中可檢測。

一些實施例可有效地篩選生物試樣(例如，血漿或血清)中之所有突變。由於突變之數目並不固定(可檢測來自不同腫瘤細胞子群之數百、數千或數百萬癌症相關突變)，故實施例可較檢測特異性突變之技術提供更好靈敏度。突變之數目可用於檢測癌症。

為提供許多或所有突變之該篩選，實施例可實施生物試樣(例如，體液，包括血漿及血清)之遺傳變化的搜索(例如，隨機搜索)，該生物試樣可含有腫瘤源DNA。使用試樣(例如血漿)會消除實施腫瘤或癌症之侵入性活檢的需要。同樣，由於篩選可覆蓋基因組之所有或大區域，故篩選並不限於任何可數及已知突變，但可使用任何突變之存在。此外，由於跨越基因組之所有或大部分對突變之數目求和，故可獲得較高靈敏度。

然而，在人類基因組中存在多態位點，包括單核苷酸多態性(SNP)，其不應計數於突變中。實施例可確定已檢測到之遺傳變化是否可能為癌症相關突變或在基因組中具有多態性。舉例而言，作為在癌症相關突變與基因組中之多態性之間進行測定之部分，實施例可測定組成基因組，其可包括多態性。組成基因組(CG)之多態性可限於在測序數據中以足夠高之百分比(例如，30-40%)呈現之多態性。

隨後可將自生物試樣獲得之序列與組成基因組進行比對並鑑別係單核苷酸突變(SNM)或其他類型突變之變化。該等SNM可為並不包

括於已知多態性中之變化，且因此可標記為癌症相關且並非組成基因組之部分。健康人員由於健康細胞間(例如)在細胞分裂期間產生之隨機突變可具有特定數目之SNM，但患有癌症之人員可具有更大數目。

舉例而言，對於患有癌症之人員而言，在體液中可檢測之SNM之數目可高於同一人員之組成基因組中存在之多態性。可在含有腫瘤源DNA之體液試樣與含有大部分組成DNA之DNA試樣中檢測到之變化之量之間進行比較。在一個實施例中，術語「大部分」可意指超過90%。在另一較佳實施例中，術語「大部分」可意指超過95%、97%、98%或99%。在體液中之變化量超過具有大部分組成DNA之試樣的變化量時，體液可含有腫瘤源DNA之機率增加。

一種可用於隨機搜尋DNA試樣之變化的方法係隨機或鳥槍法測序(例如，使用大規模平行測序)。可使用任何大規模平行測序平臺，包括藉由連接測序之平臺(例如，Life Technologies SOLiD平臺)、Ion Torrent/Ion Proton、半導體測序、Roche 454、單分子測序平臺(例如，Helicos、Pacific Biosciences及奈米孔(nanopore))。但是，已知測序誤差可發生且可誤解為組成DNA變化或源自腫瘤DNA之突變。因此，為改良所提出方法之特異性，可藉由(例如)使用適當測序深度以及在基因座處需要至少指定數目(例如，2或3個)之檢測對偶基因以將其計數為SNM來慮及測序誤差或分析誤差之其他組份的機率。

如本文所述，在試樣中存在之隨機檢測之遺傳變化之量超過組成DNA所期望之量及由於分析誤差(例如，測序誤差)可無意地檢測之變化時，實施例可證實在生物試樣(例如，體液)中存在腫瘤源DNA。該資訊可用於篩選、診斷、預測及監測癌症。在以下章節中，闡述可用於檢測血漿/血清或其他試樣(例如，體液)中之單核苷酸突變之分析步驟。體液可包括血漿、血清、腦脊髓液、胸水、腹水、乳頭溢液、唾液、枝氣管肺泡灌洗液、痰、眼淚、汗液及尿。除體液外，該技術

亦可適於糞便試樣，此乃因後者已顯示含有結腸直腸癌之腫瘤DNA (Berger BM, Ahlquist DA. Pathology 2012 ; 44: 80-88)。

II. 一般篩選方法

圖1係根據本發明實施例檢測個體之癌症或癌變前變化之方法100的流程圖。實施例可分析個體生物試樣中之無細胞DNA以檢測可能自腫瘤產生之無細胞DNA的變化。分析可使用個體之組成基因組以慮及作為健康細胞之部分的多態性且可慮及測序誤差。方法100及本文所述任何方法可完全或部分地利用包括一或多個處理器之電腦系統實施。

在步驟110中，獲得個體之組成基因組。可自所測試個體之組成DNA測定組成基因組(CG)。在各個實施例中，CG可自記憶體讀取或藉由(例如)分析組成DNA之序列讀段主動測定，該組成DNA可存於包括無細胞DNA之試樣的細胞中。舉例而言，在懷疑非血液惡性腫瘤時，可分析血細胞以測定個體之組成DNA。

在各個實作中，可使用大規模平行測序、基於陣列之雜交、基於探針之溶液中雜交、基於連接分析、引物延伸反應分析及質譜實施組成DNA之分析。在一個實施例中，可在個體壽命中之一個時間點時(例如，出生時或甚至在產前期中測定CG(其可使用胎兒細胞或經由無細胞DNA片段進行，參見美國公開案2011/0105353))，且隨後在個體壽命之其他時間時獲得體液或其他試樣時提及。因此，可自電腦記憶體簡單讀取CG。可讀出組成基因組作為基因座之列表，其中組成基因組不同於參照基因組。

在步驟120中，接收個體之生物試樣中複數個DNA片段中之每一者之一或多個序列標籤，其中生物試樣包括無細胞DNA。在一個實施例中，一或多個序列標籤係自生物試樣中之DNA片段之隨機測序產生。在實施末端配對測序時可獲得一個以上序列標籤。一個標籤可對

應於DNA片段之每一末端。

可分析試樣(例如，血漿、血清或其他體液)中之無細胞DNA以搜尋遺傳變化。可使用與使用分析組成DNA相同之分析平臺分析無細胞DNA。或者，可使用不同分析平臺。舉例而言，可使用大規模平行測序對無細胞DNA試樣進行測序或可在大規模平行測序之前捕獲或富集基因組之部分。若使用富集，則可(例如)使用基因組之選擇部分之溶液相或固相捕獲。隨後，可在所捕獲DNA上實施大規模平行測序。

在步驟130中，測定序列標籤之基因組位置。在一個實施例中，將序列標籤與自一或多個其他個體獲得之參照基因組進行比對。在另一實施例中，將基因組序列標籤與所測試個體之組成基因組進行比對。可使用彼等熟習此項技術者已知之技術(例如使用基本局部比對搜尋工具(BLAST))實施比對。

在步驟140中，測定基因座之第一數目，其中至少N個序列標籤相對於組成基因組(CG)具有序列變體。N等於或大於2。如下文更詳細論述，可藉由使N等於2、3、4、5或更高移除測序誤差以及細胞中隨機發生之體細胞突變(例如，由於細胞分裂)。滿足一或多個指定準則之基因座可鑑別為突變(變體)或突變基因座(變體基因座)，而具有變體但不滿足一或多個準則之基因座(例如，作為僅一個變體序列標籤)稱作潛在或假定突變。序列變體可針對僅一個核苷酸或多個核苷酸。

N可測定為與絕對值相對之基因座之總標籤的百分比。舉例而言，在自變體讀段推斷之腫瘤DNA之分數濃度測定為等於或大於10%(或一些其他百分比)時，可鑑別變體基因座。換言之，在基因座由200個序列讀段覆蓋時，將變體定義為突變可需要至少10個顯示變體對偶基因之序列讀段之準則。變體對偶基因之10個序列讀段及野生型對偶基因之190個讀段可產生10%之腫瘤DNA之分數濃度($2 \times$

10/(10+190))。

在一個實施例中，可直接比較序列標籤(統稱為試樣基因組)與CG以測定變體。在另一實施例中，經由參照基因組(RG)比較試樣基因組(SG)與CG以測定變體。舉例而言，可比較CG及SG二者與RG以測定呈現變體之基因座之各別數目(例如，組)，且隨後可採用差異以獲得基因座之第一數目。第一數目可簡單地以數字形式獲得或可對應於基因座之特定組，隨後可對其進行進一步分析以自第一基因座處之序列標籤測定參數。

在一個實作中，比較組成DNA及血漿DNA之測序結果以測定血漿DNA中是否存在單核苷酸突變。可分析組成DNA同型之區域。出於闡釋目的，假定特定基因座之基因型在組成DNA中係同型的且係AA。則在血漿中，除A外之對偶基因之存在可指示在特定基因座處潛在存在單核苷酸突變(SNM)。指示潛在存在SNM之基因座可在步驟140中形成第一數目之基因座。

在一個實施例中，靶向已知在特定癌症類型中或在特定亞組之群體中尤其易於突變的基因組之部分可為有用的。與後一態樣相關，實施例可尋找在具體群體組中尤其普遍之類型之突變，例如在係B型肝炎(對於肝癌)或人類乳頭狀瘤病毒(對於子宮頸癌)攜帶者或具有體細胞突變之遺傳傾向的個體或在DNA失配修復基因中具有種系突變之個體中尤其常見的突變。該技術亦可用於篩選具有*BRCA1*或*BRCA2*突變之個體之卵巢癌及乳癌中的突變。該技術可類似地用於篩選具有*APC*突變之個體之結腸直腸癌中的突變。

在步驟150中，基於在第一基因座處具有序列變體之序列標籤的計數測定參數。在一個實例中，參數係基因座之第一數目，其中至少N個DNA片段在基因座處相對於組成基因組具有序列變體。因此，該計數可簡單地用於確保基因座具有N個以上在包括於第一數目中之前

鑑別之特定變體拷貝。在另一實施例中，參數可為或包括在第一基因座處相對於組成基因組具有序列變體之序列標籤的總數目。

在步驟160中，比較個體之參數與臨限值(例如，源自一或多個個體)以測定個體中之癌症等級之分類。癌症等級之實例包括個體是否患有癌症或癌變前病況，或發生癌症之機率增加。在一個實施例中，可自從個體先前獲得之試樣測定臨限值。

在另一實施例中，可測定一或多個其他個體不患有癌症或低癌症風險。因此，臨限值可為正常值、正常範圍，或指示偏離正常值或範圍之統計上顯著之偏差。舉例而言，相對於具體個體之CG之無癌症或具有低癌症風險之個體血漿中可檢測之突變之數目可用作正常範圍以測定所測試個體中檢測到之突變的數目是否正常。在另一實施例中，可已知其他個體患有癌症，且因此類似數目之突變可指示癌症。

在一個實作中，其他個體可經選擇以具有與測試個體之臨床特徵匹配之臨床特徵，例如性別、年齡、飲食、吸菸習慣、用藥史、先前疾病、家族病史、所需基因組基因座之基因型、病毒感染(例如B或C型肝炎病毒或人類乳頭狀瘤病毒或人類免疫缺陷病毒或愛潑斯坦-巴爾病毒(Epstein-Barr virus)感染)或經其他感染物(例如細菌(例如幽門螺旋桿菌(*Helicobacter pylori*))及寄生蟲(例如華支睪吸蟲(*Clonorchis sinensis*))感染之狀態等。舉例而言，作為B或C型肝炎病毒攜帶者之個體具有發生肝細胞癌之增加風險。因此，作為B或C型肝炎之攜帶者具有類似突變數目或型式之測試個體可視為具有發生肝細胞癌之增加風險。另一方面，由於使用適當基線(即相對於另一肝炎患者)，較另一肝炎患者呈現更多突變之B或C型肝炎患者可適當地鑑別為具有較高癌症等級之分類。類似地，作為人類乳頭狀瘤病毒感染之攜帶者之個體具有子宮頸癌及頭頸癌之增加風險。感染愛潑斯坦-巴爾病毒與鼻咽癌、胃癌、霍奇金氏淋巴瘤(Hodgkin's lymphoma)及非霍奇金

氏淋巴瘤相關。感染幽門螺旋桿菌與胃癌相關。感染華支睪吸蟲與膽管上皮癌相關。

不同時間點時突變數目之變化的監測可用於監測癌症之進展及治療反應。該監測亦可用於記錄處於個體可發生癌症之風險中之癌變前病況或變化的進展。

顯示變化之序列標籤之量亦可用於監測。舉例而言，可使用基因座處之變體讀段的分數濃度。在一個實施例中，在連續監測期間試樣中之腫瘤相關之遺傳畸變的分數濃度增加可表示疾病進展或危急復發。類似地，在連續期間監測試樣中之腫瘤相關之遺傳畸變的分數濃度減少可表示對治療有反應及/或緩解及/或良好預後。

III. 測定基因組

下文更詳細解釋上文論述之各種基因組。舉例而言，論述參照基因組、組成基因組及試樣基因組。

A. 參照基因組

參照基因組(RG)係指個體之單倍體或二倍體基因組或群體之共有基因組。參照基因組係已知的且因此其可用於比較新患者之測序讀段。可比對並比較患者試樣之序列讀段以鑑別RG之讀段的變化。對於單倍體基因組，在每一基因座處僅存在一個核苷酸，且因此每一基因座可視為半合子。對於二倍體基因組，可鑑別異型基因座，該基因座具有兩個對偶基因，其中任一對偶基因可容許用於與基因座比對之匹配。

參照基因組可在個體群體中相同。此相同參照基因組可用於健康個體以測定欲用於對患者進行分類(例如，患有或無癌症)之適當臨限值。然而，不同參照基因組可用於不同群體，例如，用於不同種族或甚至用於不同家族。

B. 組成基因組

個體(例如，人類或其他二倍體有機體)之組成基因組(CG)係指個體之二倍體基因組。CG可詳細說明異型基因座，其中第一對偶基因係來自第一單倍型且不同第二對偶基因係來自第二單倍型。應注意，不需已知覆蓋兩個異型基因座之兩個單倍型之結構，即，不需已知一個異型基因座上之對偶基因與另一異型基因座之對偶基因在相同單倍型上。僅在每一異型基因座處存在兩個對偶基因即可足矣。

由於多態性，CG可不同於RG。舉例而言，RG上之基因座對於T同型，但CG對於T/A異型。因此，CG可在此基因座處呈現變化。由於遺傳突變(例如，在家族中運行)或新生突變(發生於胎兒中，但其不存於其母體中)，CG亦可不同於RG。遺傳突變通常稱作「種系突變」。該等突變中的一些與癌症傾向相關，例如家族中運行之*BRCA1*突變。該等突變不同於在人員壽命期間可由於細胞分裂發生且可將細胞及其子代推至將要變為癌症的「體細胞突變」。

測定CG之目的係自試樣基因組(SG)之突變移除該等種系突變及新生突變以鑑別體細胞突變。隨後可使用SG中之體細胞突變之量來評定個體之癌症之機率。可進一步過濾該等體細胞突變以移除測序誤差，且潛在地移除極少發生(例如，僅一種顯示變體之讀段)之體細胞突變，因此體細胞突變不可能與癌症相關。

在一個實施例中，可使用細胞(棕黃色塗層DNA)測定CG。然而，亦可自無細胞DNA (例如血漿或血清)測定CG。對於大部分細胞係非惡性之試樣類型(例如健康個體之棕黃色塗層)，則大多數或共有基因組係CG。對於CG，每一基因組基因座由所取樣組織中之大多數細胞所具有之DNA序列組成。測序深度應足以闡明組成基因組內之異型位點。

作為另一實例，血漿可用作組成試樣以測定CG。舉例而言，對於血漿中之腫瘤DNA少於50%且SNM係呈異型狀態(例如，突變係添

加新對偶基因)之情形，則新對偶基因可具有小於25%之濃度。而CG中之SNP之異型對偶基因的濃度應總計約50%。因此，可在體細胞突變與CG之多態性之間加以區分。在一個實作中，在使用具有顯著腫瘤濃度之血漿或其他混合物時，對於測定多態性之體細胞突變適宜截止值可介於30-40%之間。腫瘤DNA濃度之量測可用於確保血漿中之腫瘤DNA少於50%。本文闡述測定腫瘤DNA濃度之實例。

C. 試樣基因組

試樣基因組(SG)並非僅為單倍體或二倍體基因組，如RG及CG之情形一般。SG係試樣之讀段之集合，且可包括：對應於CG之組成DNA之讀段、腫瘤DNA之讀段、相對於CG顯示隨機突變之健康細胞的讀段(例如，由於自細胞分裂產生之突變)及測序誤差。可使用各種參數來精確控制SG中包括何種讀段。舉例而言，需要對偶基因以在至少5個讀段中出現可減少SG中存在之測序誤差，以及減少由於隨機突變之讀段。

作為實例，假定個體健康，即不患有癌症。出於闡釋目的，1000個細胞之DNA存於1 ml自此個體獲得之血漿(即1000個基因組當量之DNA)中。血漿DNA通常由約150 bp之DNA片段組成。由於人類基因組係 3×10^9 bp，故每個單倍體基因組可存在約 2×10^7 個DNA片段。由於人類基因組係二倍體，故每ml血漿可存在約 4×10^7 個DNA片段。

由於每單位時間數百萬至數十億細胞在血漿中釋放其DNA且該等細胞之片段可在循環期間混合在一起，故 4×10^7 個DNA片段可來自 4×10^7 個不同細胞。若該等細胞彼此不具有最近(與距離相對，例如，初始合子)純系關係(即，其不共享最近祖細胞)，則在統計學上可能在該等片段間不可觀察到一次以上突變。

另一方面，若在1000基因組當量/ml血漿DNA中，存在特定百分比之共享最近祖細胞(即，其彼此純系相關)的細胞，則可自此純系觀

察到突變在血漿DNA中優先表示(例如在血漿中呈現純系突變概況)。該等純系相關之細胞可為癌細胞或將要變為癌症但尚未成為癌症(即腫瘤前)之細胞。因此，需要突變以出現一次以上可移除試樣中鑑別之「突變」之此自然差異，其可留下更多突變相關之癌細胞或腫瘤前細胞，藉此容許檢測、尤其早期檢測癌症或癌前病況。

在一個近似情況中，已闡明平均在每次細胞分裂後基因組中將累積一個突變。先前工作已顯示，大部分血漿DNA係來自造血細胞(Lui YY等人 Clin Chem 2002: 48: 421-427)。已估計造血幹細胞每25-50週複製一次(Catlin SN等人 Blood 2011; 117: 4460-4466)。因此，作為過分簡單化之近似情況，健康40歲個體可在每個造血幹細胞累積約40至80個突變。

若存在1000基因組當量/ml此人血漿，且若每一該等細胞皆係源自不同造血幹細胞，則預計在 4×10^{10} 個DNA片段(即 4×10^7 個DNA片段/基因組及1000基因組當量/ml血漿)中可具有40,000至80,000個突變。然而，由於每一突變可觀察一次，故每一突變仍可低於檢測限值(例如，若截止值N大於1)，且因此可濾出該等突變，藉此容許分析主要關注更可能自癌症病況產生之突變。截止值可為大於1之任何值(整數或非整數)，且可對於不同基因座及區域具有動態。測序深度及腫瘤DNA之分數濃度亦可影響檢測癌細胞或腫瘤前細胞之突變的靈敏度(例如，可檢測突變之百分比)。

IV. 直接比較SG與CG

一些實施例可鑑別CG係同型但其中SG中之少數物質(即腫瘤DNA)係異型之核苷酸位置。在以高深度(例如，50倍以上覆蓋)對位置進行測序時，可檢測在健康細胞及癌細胞之DNA混合物中於該位置處是否存在兩個對偶基因。在存在兩個檢測對偶基因時，(1) CG係異型或(2) CG係同型但SG係異型。該兩種情況可藉由尋找主要及次要對

偶基因之相對計數來區分。在前一情況下，兩個對偶基因可具有類似數目之計數；但對於後一情況，其計數數目可存在大的差異。測試試樣之讀段之相對對偶基因計數的此比較係比較序列標籤與組成基因組之一個實施例。方法100之第一基因座可測定為基因座，其中對偶基因之數目低於上臨限值(對應於CG中之多態性之臨限值)且高於下臨限值(對應於以與癌症病況無關之足夠低比率發生之誤差及體細胞突變的臨限值)。因此，可同時測定組成基因組及第一基因座。

在另一實施例中，鑑別突變之方法可首先測定CG，且隨後測定相對於CG具有足夠突變數目之基因座。可自不同於測試試樣之組成試樣測定CG。

圖2顯示根據本發明實施例直接比較試樣基因組(SG)與組成基因組(CG)的方法200之流程圖。於框210處，獲得個體之組成基因組。可自(例如)先前及時採取之試樣或在即將執行方法200之前獲得並分析之組成試樣獲得組成基因組。

於框220處，接收個體生物試樣中複數個DNA片段中之每一者之一或多個序列標籤。可使用如本文中提及之各種技術實施測序。序列標籤係據信為何種片段之序列的量測。但是，序列標籤之一或多個鹼基可有誤差。

在框230處，比對至少一部分序列標籤與組成基因組。比對可慮及在不同基因座處CG係異型的。比對可不需精確匹配以使可檢測變體。

在框240處，鑑別在基因座處相對於組成基因組具有序列變體之序列標籤。序列標籤可具有一個以上變體。可追蹤每一基因座及每一序列標籤之變體。變體可為並不存於CG中之任何對偶基因。舉例而言，CG對於A/T可為異型且變體可為G或C。

於框250處，對於具有變體之每一基因座，電腦系統可對與基因

座比對且在基因座處具有序列變體之序列標籤的各別第一數目進行計數。因此，每一基因座可具有於基因座處觀察之變體數目的相關計數。通常，例如，由於腫瘤DNA濃度小於50%，與對應於CG之序列標籤相比，在基因座處將觀察到較少變體。然而，一些試樣可具有大於50%之腫瘤DNA濃度。

於框260處，基於各別第一數目測定參數。在一個實施例中，若各別數目大於截止值(例如，大於2)，則可將各別數目添加至和，該和係參數或用於測定參數。在另一實施例中，具有大於截止值之各別數目之基因座的數目用作參數。

於框270處，比較參數與臨限值以對癌症等級進行分類。如上文所述，可自其他個體之試樣的分析測定臨限值。端視該等其他個體之健康或癌症狀態而定，可測定分類。舉例而言，若其他個體具有4期癌症，則若當前參數接近(例如，在特定範圍內)自其他個體獲得之參數之值，則當前個體可分類為具有4期癌症。然而，若參數超過臨限值(即，端視如何定義參數而定，大於或小於)，則分類可鑑別為小於4期。在其他個體不具有癌症時，可進行類似分析。

可使用多個臨限值測定分類，其中每一臨限值係自一組不同個體測定。每一組個體可具有常見癌症等級。因此，可比較當前參數與每一組個體之值，其可提供與組中之一者之匹配或提供範圍。舉例而言，參數可約等於針對癌前或2期個體獲得之參數。作為另一實例，當前參數可屬於可能與若干不同癌症等級匹配之範圍。因此，分類可包括一個以上癌症等級。

V. 使用參照基因組

可比較組成DNA及來自生物試樣之DNA二者的基因組序列與人類參照基因組。在與參照基因組相比血漿試樣中較組成DNA存在更多變化時，則存在較高癌症機率。在一個實施例中，研究參照基因組中

之同型基因座。比較組成DNA及來自生物試樣之DNA二者中之異型基因座的量。在自生物試樣之DNA檢測到之異型位點的量超過組成DNA之異型位點的量時，癌症機率較高。

分析亦可限於在CG中同型之基因座。亦可針對異型基因座定義SNM，但此通常可需要產生第三變體。換言之，若異型基因座係A/T，則新變體可為C或G。鑑別同型基因座之SNM通常較容易。

在與健康個體中觀察到之變化之比率相比時，相對於組成DNA之生物試樣DNA中之異型基因座之量增加可表明癌症或癌變前狀態。舉例而言，若該等位點中之增加程度超過健康個體中觀察到之增加程度達特定臨限值，則可將該數據視為表明癌症或癌變前狀態。在一個實施例中，確定無癌症之個體中之突變之分佈且可取臨限值為標準偏差之特定數目(例如，2或3個標準偏差)。

在對基因座進行計數之前，一個實施例可在基因座處需要至少指定數目之變體。另一實施例甚至提供對基於觀察到一次變化之數據之測試。舉例而言，當在血漿中觀察到之變化(誤差 + 真突變或多態性)之總數目在統計學上顯著高於組成DNA中之變化總數目時，則證實存在癌症。

圖3顯示根據本發明實施例使用參照基因組(RG)比較試樣基因組(SG)與組成基因組(CG)的方法300之流程圖。方法300假定已經獲得RG，且已經接收生物試樣之序列標籤。

於框310處，比對至少一部分序列標籤與參照基因組。在檢測變化時，比對可容許失配。參照基因組可來自與個體類似之群體。比對序列標籤有效地包含試樣基因組(SG)

於框320處，鑑別潛在變體(例如，單核苷酸突變(SNM))之第一數目(A)。潛在SNM係基因座，其中SG之序列標籤顯示不同於RG之核苷酸。可使用其他準則，例如，顯示變化之序列標籤之數目應大於截

止值且基因座在RG中是否同型。在藉由將基因座儲存於記憶體中鑑別並追蹤特定基因座時，潛在SNM組可表示為A組。可測定具體基因座可僅測定該等SNM之數目。

於框330處，藉由將藉由對組成試樣之DNA片段進行測序獲得的序列標籤與參照基因組進行比對來測定組成基因組。此步驟可在先前之任何時間且使用在先前之任何時間獲得之組成試樣來實施。CG可僅為記憶體之讀段，其中先前進行比對。在一個實施例中，組成試樣可為血細胞。

於框340處，鑑別基因座之第二數目(B)，其中CG之比對序列標籤於基因座處相對於參照基因組具有變體(例如，SNM)。若特定追蹤一組基因座，則僅與數目相對，B可表示該組。

與框350處，自A組減去B組以鑑別存於試樣基因組而非CG中之變體(SNM)。在一個實施例中，SNM組可限於CG同型之核苷酸位置。為達成此過濾，其中CG同型之特定基因座可在C組中鑑別。在另一實施例中，若CG在基因座處不同型，則未以第一數目A或第二數目B對基因座計數。在另一實施例中，可濾出任何已知多態性(例如藉由其在SNP數據庫中之存在)。

在一個實施例中，框350中之扣減可僅為數目之扣減，且因此未移除特定潛在SNMs，而是僅減去一值。在另一實施例中，扣減為A組與B組間之差(例如其中B組係A組之亞組)以鑑別不在B組中之特定SNMs。在邏輯值中，此可表示為[A且非(B)]。所得鑑別變體組可標記為C。該參數可測定為數目C或自C組測定。

在一些實施例中，可考慮突變之性質且不同加權歸因於不同突變種類。舉例而言，通常與癌症相關之突變可導致較高加權(在提及基因座之相對加權時，亦稱作重要值)。該等突變可發現於腫瘤相關突變之數據庫，例如 Catalogue of Somatic Mutations in Cancer

(COSMIC)(www.sanger.ac.uk/genetics/CGP/cosmic/)。作為另一實例，非同義變化相關之突變可導致較高加權。

因此，第一數目A可測定為一加權和，其中在一個基因座顯示變體之標籤計數較另一基因座之標籤計數可具有不同加權。第一數目A可反映此加權和。可對B實施類似計算，且因此數目C及該參數可反映此加權。在另一實施例中，在測定C組之特定基因座時進行加權。舉例而言，可針對C組之基因座之計數測定一加權和。該等加權可用於本文所述之其他方法。

因此，與一臨限值比較以測定癌症等級分類的參數可為呈現SG及CG相對於RG之變化之基因座數目。在其他實施例中，可計數顯示變化之DNA片段之總數目(如經由序列標籤計數)。在其他實施例中，該等數目可用於另一式中以獲得該參數。

在一個實施例中，該變體在各基因座之濃度可為一參數並與一臨限值進行比較。此臨限值可用於測定一基因座是否係潛在變體基因座(除顯示該變體之讀段之特定數目的截止值以外)，且隨後計數該基因座。該濃度亦可用作SNMs和中之一個加權因子。

VI. 使用截止值減少偽陽性

如上文所提及，可在大基因組區域(例如全基因組)之大量無細胞DNA片段(例如血漿中循環DNA)或大量基因組區域中勘察單核苷酸突變以改良方法之靈敏度。然而，分析誤差(例如測序誤差)可影響此方法之可行性、精度及特異性。此處，使用大規模平行測序平臺作為實例以闡釋測序誤差之重要性。Illumina合成測序平臺(sequencing-by-synthesis)之測序誤差率係約0.1%至0.3% / 測序核苷酸(Minoche等人 *Genome Biol* 2011, 12:R112)。可使用任何大規模平行測序平臺，包括藉由連接測序之平臺(例如，Life Technologies SOLiD平臺)、Ion Torrent/Ion Proton、半導體測序、Roche 454、單分子測序平臺(例

如，Helicos、Pacific Biosciences及奈米孔)。

在關於肝細胞癌之先前研究中顯示，整個癌症基因組中存在約3,000個單核苷酸突變(Tao Y等人 2011 Proc Natl Acad Sci USA ; 108: 12042-12047)。假定循環總DNA中僅10%源自腫瘤細胞且以一倍單倍體基因組覆蓋之平均測序深度對血漿DNA進行測序，由於測序誤差可遇到9百萬($3 \times 10^9 \times 0.3\%$)個單核苷酸變異(SNV)。然而，預計大部分單核苷酸突變僅發生在兩條同源染色體中之一者上。在具有100%腫瘤DNA之試樣的一倍單倍體基因組覆蓋之測序深度下，預計僅檢測到3,000個突變之一半，即1,500個突變。在將含有10%腫瘤源DNA之血漿試樣測序至一倍單倍體基因組覆蓋時，預計僅檢測150 ($1,500 \times 10\%$)個癌症相關之單核苷酸突變。因此，檢測癌症相關突變之信噪比係1/60,000。此極低信噪比表明，若僅使用生物試樣(例如，血漿)中之所有單核苷酸變化作為參數，則使用此方法區分正常及癌症情形之精度會極低。

預計隨著測序技術之進展，測序誤差率可連續減少。亦可使用一個以上測序平臺並經由比較交叉-平臺測序結果分析相同試樣，指出可能受測序誤差影響之讀段。另一方法係分析在不同時間時自相同個體獲取之兩個試樣。然而，該等方法耗時。

在一個實施例中，一種在癌症患者血漿中之單核苷酸突變之檢測中增強信噪比之方式係僅在試樣中多次發生相同突變時對突變進行計數。在所選測序平臺中，涉及特定核苷酸取代之測序誤差可更常見且可影響測試個體及對照個體之測試試樣及組成DNA試樣的測序結果。然而，一般而言，測序誤差隨機發生。

在多個DNA片段中之相同核苷酸位置處觀察相同變化時，具有測序誤差之機會指數級較低。另一方面，檢測試樣中真癌症相關之突變變化之機會受測序深度及試樣中之腫瘤DNA之分數濃度影響。觀察

多個DNA片段中之突變之機會可隨測序深度及腫瘤DNA之分數濃度增加。在使用具有無細胞之腫瘤DNA之試樣(例如血漿中)的各個實施例中，分數濃度可為5%、10%、20%及30%。在一個實施例中，分數濃度小於50%。

圖4係表400，其顯示本發明實施例，使用不同發生次數作為將突變分類為存於試樣中之準則正確鑑別的癌症相關單核苷酸突變的數目。亦顯示基於相同分類準則由於測序誤差誤鑑別為具有突變之核苷酸位置的數目。測序誤差率假定為0.1% (Minoche等人 *Genome Bio* 2011, 12:R112)。試樣中之腫瘤源DNA之分數濃度假定為10%。

圖4顯示血漿中檢測到之癌症相關突變之數目與偽陽性稱為之數目之間的比率可隨次數增加而增加，在試樣中之腫瘤源DNA之分數濃度假定為10%時，在定義突變之試樣中觀察到相同變化。換言之，癌症突變檢測之靈敏度及特異性二者均可改良。另外，檢測癌症相關突變之靈敏度受測序深度影響。在測序之100倍單倍體基因組覆蓋情況下，甚至可使用試樣中之至少4個DNA片段中之特定突變的發生次數準則檢測3,000個突變中之2,205個(73.5%)。可使用最小數目之片段之其他值，例如3、5、8、10及大於10。

圖5係表500，其顯示在試樣中之腫瘤源DNA之分數濃度假定為5%時鑑別之偽陽性基因座之預計數目及突變之預計數目。試樣中之腫瘤源DNA之分數濃度愈低，將需要愈高之測序深度以達成檢測癌症相關突變之相同靈敏度。亦可需要更嚴格準則來維持特異性。舉例而言，試樣中可需要使用至少5個DNA片段中之特定突變之發生次數準則，而非在10%腫瘤DNA部分之情況下至少4次發生之準則。表400及500提供假設可如本文所述假定或量測之覆蓋倍數及腫瘤DNA濃度下使用截止值之指南。

使用檢測單核苷酸變化一次以上以定義突變之準則的另一優勢

在於，由於非惡性組織中之單核苷酸變化，預計此可使偽陽性檢測最小化。由於核苷酸變化可在正常細胞之有絲分裂期間發生，故體內之每一健康細胞可具有多個單核苷酸變化。該等變化可潛在地導致偽陽性結果。然而，在細胞死亡時，細胞之變化可存於血漿/血清中。儘管預計不同正常細胞帶有不同突變組，但一個細胞中發生之突變可能存於血漿/血清中之數個拷貝中。此與腫瘤細胞內之突變相對，其中由於腫瘤生長在性質上係純系的，故預計在血漿/血清中可觀察到多個拷貝。因此，來自純系之多個細胞可死亡並釋放代表純系之簽名突變。

在一個實施例中，可在測序之前實施特定基因組區域之靶標富集。此靶標富集步驟可增加所關注區域之測序深度，實施相同總量之測序。在又一實施例中，可首先實施具有相對較低測序深度之一輪測序。隨後可針對具有較高覆蓋倍數之第二輪測序富集顯示至少一個單核苷酸變化之區域。隨後，多次發生之準則可適於定義具有靶標富集之測序結果之突變。

VII. 動態截止值

如上文所述，支持變體(潛在突變)之讀段之數目之截止值 N 可用於測定基因座是否對於欲計數突變(例如，SNM)合格。使用該截止值可減少偽陽性。下文論述提供選擇不同基因座之截止值的方法。在以下實施例中，假定存在單一主要癌症純系。對於涉及向血漿中釋放不同量之腫瘤DNA之癌細胞的多個純系的情況，可實施類似分析。

A. 血漿中檢測到之癌症相關突變之數目

血漿中可檢測之癌症相關突變的數目可受多個參數影響，例如：(1) 腫瘤組織中之突變之數目(N_T) - 腫瘤組織中存在之突變之總數目係患者血漿中可檢測之腫瘤相關之突變的最大數目；(2) 血漿中之腫瘤源DNA之分數濃度 (f) - 血漿中之腫瘤源DNA之分數濃度愈

高，則檢測血漿中之腫瘤相關之突變之機會愈高；(3) 測序深度(D) - 測序深度係指測序區域由序列讀段覆蓋之次數。舉例而言，10倍之平均測序深度意指測序區域內之每一核苷酸由10個序列讀段覆蓋。在測序深度增加時，檢測癌症相關突變之機會可增加；及(4) 血漿中檢測到之核苷酸變化的最小次數以便將其定義為潛在癌症相關突變(r)，其係用於區分測序誤差與真實癌症相關突變之截止值。

在一個實作中，可使用葛瓦松分佈(Poisson distribution)預測血漿中檢測到之癌症相關突變之數目。假定在兩條同源染色體中之一條上之核苷酸位置中存在突變，測序深度為D，血漿中存在之預計突變次數(M_p)計算為： $M_p = D \times f/2$ 。

於特定突變位點處檢測血漿中之突變之機率(P_b)計算為：

$$P_b = 1 - \sum_{i=0}^{r-1} \text{葛瓦松}(i, M_p)$$

其中r (截止值)係血漿中觀察到之核苷酸變化之次數以便將其定義為潛在腫瘤相關之突變；葛瓦松(i, M_p)係具有*i*發生次數與 M_p 之平均數之葛瓦松分佈機率。

預計在血漿(N_p)中檢測到之癌症相關突變的總數目可計算為： $N_p = N_t \times P_b$ ，其中 N_t 係腫瘤組織中存在之突變數目。以下圖表顯示使用稱為潛在突變及不同測序深度之發生次數(r)的不同準則之預計在血漿中檢測到之腫瘤相關之突變之百分比。

圖6A係圖表600，其顯示在具有10%及20%血漿分數濃度之腫瘤源DNA之血漿中，使用4次及6次發生次數(r)作為稱為潛在癌症相關突變之準則檢測到之癌症相關突變之比率。在相同r下，血漿中較高分數濃度之腫瘤源DNA可產生更高數目之可在血漿中檢測到之癌症相關突變。在血漿中之相同分數濃度之腫瘤源DNA下，較高r可產生較小數目之檢測突變。

B. 由於誤差之單次檢測之偽陽性的數目

血漿DNA測序數據中之單核苷酸變化可由於測序及比對誤差發生。可基於二項分佈以數學方式預測具有偽陽性單核苷酸變化之核苷酸位置的數目。影響偽陽性位點(N_{FP})之數目之參數可包括：(1) 測序誤差率(E) - 測序誤差率定義為測序核苷酸不正確之比例；(2) 測序深度(D) - 在較高測序深度下，顯示測序誤差之核苷酸位置之數目可增加；(3) 定義潛在癌症相關突變之相同核苷酸變化的最小發生次數(r)；及(4) 所關注區域內之核苷酸位置之總數目(N_I)。

突變之發生通常可視為隨機過程。因此，隨著定義潛在突變之發生次數之準則增加，偽陽性核苷酸位置之數目可隨 r 以指數減少。在一些現存測序平臺中，某些序列上下文更易於具有測序誤差。該等測序上下文之實例包括GGC基序、均聚物(例如AAAAAAA)及簡單重複(例如ATATATATAT)。該等序列上下文可實質上增加單核苷酸變化或插入/缺失假像(Nakamura K等人 *Nucleic Acids Res* 2011; 39,e90及 Minoche AE等人 *Genome Biol* 2011;12,R112)。另外，重複序列(例如均聚物及簡單重複)可以計算方式引入比對模糊，且因此，導致單核苷酸變異之偽陽性結果。

所關注區域愈大，則可觀察到之偽陽性核苷酸位置之數目愈大。若尋找全基因組中之突變，則所關注區域可為全基因組且所涉及核苷酸之數目可為30億。另一方面，若主要關注外顯子，則編碼外顯子之核苷酸之數目(即約4500萬)可構成所關注區域。

可基於以下計算測定與測序誤差相關之偽陽性核苷酸位置之數目。由於測序誤差在相同位置處具有相同核苷酸變化之機率(P_{Er})可計算為：

$$P_{Er} = C(D, r)E \left(\frac{E}{3}\right)^{r-1}$$

其中 $C(D, r)$ 係用於自總共 D 要素選擇 r 要素之可能組合之數目； r 係定義潛在突變之發生次數； D 係測序深度；且 E 係測序誤差率。 $C(D, r)$

可計算為：

$$C(D, r) = \frac{D!}{r!(D-r)!}$$

對於突變偽陽性之核苷酸位置(N_{FP})之數目可計算為：

$$N_{FP} = N_I P_{Er}$$

其中 N_I 係所關注區域中之核苷酸位置之總數目。

圖6B係圖表650，其相對對測序深度顯示使用4、5、6及7次發生次數(r)之準則誤分類為具有核苷酸變化之核苷酸位置的預計數目。在此計算中，所關注區域假定為全基因組(30億核苷酸位置)。測序誤差率假定為測序核苷酸之0.3%。如可見， r 之值對偽陽性具有明顯影響。但是，如可自圖6A所見，較高 r 值亦減少所檢測突變之數目，至少直至使用顯著較高測序深度為止。

C. 選擇最小發生次數(r)

如上文所論述，真實癌症相關突變位點及由於測序誤差之偽陽性位點之數目可隨測序深度增加。然而，其增加速率可不同。因此，可利用測序深度及 r 值之選擇以使真實癌症相關突變之檢測最大化同時保持偽陽性位點之數目於低值下。

圖7A係圖表700，其顯示具有不同測序深度之真實癌症相關突變位點及偽陽性位點的數目。腫瘤組織中之癌症相關突變之總數目假定為3,000且血漿中之腫瘤源DNA之分數濃度假定為10%。測序誤差率假定為0.3%。在圖例中，TP表示腫瘤組織中存在相應突變之真實性位點，且FP表示腫瘤組織中不存在相應突變且測序數據中存在之核苷酸變化係由於測序誤差之偽陽性位點。

自圖表700，於110倍之測序深度下，若使用6之最小發生次數作為用以定義血漿中之潛在突變位點之準則($r=6$)，則可檢測約1,410個真實癌症相關突變。使用此準則，僅可檢測約20個偽陽性位點。若使

用7次發生次數($r=7$)作為用以定義潛在突變之準則，則可檢測之癌症相關突變之數目可減少470至約940。因此， $r=6$ 之準則可使得血漿中之癌症相關突變之檢測更靈敏。

另一方面，於200倍之測序深度下，若分別使用6及7之最小發生次數(r)之準則來定義潛在突變，則所檢測真實癌症相關突變之數目可為約2,800及2,600。使用該兩個 r 值，偽陽性位點之數目可分別為約740及20。因此，於200倍之測序深度下，使用定義潛在突變之 $r=7$ 之更嚴格準則可大大減少偽陽性位點之數目，而不顯著有害地影響檢測真實癌症相關突變之靈敏度。

D. 定義血漿中之潛在突變之測序數據的動態截止值

所關注區域內之每一核苷酸的測序深度可不同。若採用核苷酸變化之發生次數之固定截止值以定義血漿中之潛在突變，則與具有較低測序深度之核苷酸相比，由更多序列讀段(即較高測序深度)覆蓋之核苷酸可在不存在腫瘤組織中之該變化下具有由於測序誤差誤標記為具有核苷酸變異的更高機率。克服此問題之一個實施例係根據特定核苷酸位置之實際測序深度且根據稱為偽陽性變化之機率之期望上限施用 r 之動態截止值至不同核苷酸位置。

在一個實施例中，最大可容許之偽陽性比率可固定為 1.5×10^{-8} 個核苷酸位置中之1個。利用此最大可容許偽陽性比率，全基因組中可鑑別之偽陽性位點之總數目可小於20。可根據圖6B中所示曲線測定不同測序深度之 r 值且該等截止值示於表1中。在其他實施例中，可使用其他不同之最大容許偽陽性比率，例如 3×10^{-8} 之1、 10^{-8} 之1或 6×10^{-7} 之1。偽陽性位點之相應總數目可分別為小於10、30及50。

特定核苷酸位置之測序深度	用以定義潛在突變之血漿DNA測序數據中存在之核苷酸變化的最小發生次數(r)
<50	5
50 - 110	6
111 - 200	7
201 - 310	8
311 - 450	9
451 - 620	10
621 - 800	11

表1. 定義特定核苷酸位置之不同測序深度之潛在突變的血漿中存在之核苷酸變化的最小發生次數(r)。最大偽陽性比率固定為 1.5×10^8 個核苷酸中之1個。

E. 靶標富集測序

如圖7A中所示，較高測序深度對於檢測癌症相關突變可產生更好靈敏度，同時藉由容許使用較高 r 值保持偽陽性位點之數目較低。舉例而言，於110倍之測序深度下，在血漿中使用6之 r 值可檢測到1,410個真實癌症相關突變，而在測序深度增加至200倍且施用7之 r 值時，所檢測真實癌症相關突變之數目可為2,600。兩組數據可產生約20之偽陽性位點之預計數目。

儘管全基因組之測序至200倍之深度在目前相對昂貴，但達成該測序深度之一種可能方式係主要關注更小所關注區域。可藉由雜交藉由(例如，但不限於)使用DNA或RNA引誘物來捕獲所關注基因組區域達成靶區域之分析。隨後藉由(例如)磁力方式下拉所捕獲區域且隨後使其經受測序。可使用(例如) Agilent SureSelect靶標富集系統、Roche Nimblegen靶標富集系統及Illumina靶向再測序系統實施該靶標捕獲。另一方法係實施靶區域之PCR擴增且隨後實施測序。在一個實

施例中，所關注區域係全外顯子組(exome)。在該實施例中，所有外顯子之靶標捕獲皆可在血漿DNA上實施，且隨後可對針對外顯子區域富集之血漿DNA進行測序。

除具有較高測序深度外，主要關注特定區域而非分析全基因組可顯著減少搜尋空間中之核苷酸位置之數目且假定相同測序誤差率下可導致偽陽性位點之數目減少。

圖7B係圖表750，其顯示涉及全基因組(WG)及所有外顯子之分析的偽陽性位點之預測數目。對於每一類型之分析，使用兩個不同 r 值，即5及6。於200倍之測序深度下，若使用 $r = 5$ 定義血漿中之突變，則對於全基因組及所有外顯子之偽陽性位點之預測數目分別係約23,000及230。若使用 $r = 6$ 定義中血漿之突變，則偽陽性位點之預測數目分別係750及7。因此，所關注區域中之核苷酸之數目之限值可顯著減少血漿突變分析中之偽陽性之數目。

在外顯子-捕獲或甚至全外顯子組-捕獲測序中，搜尋空間中之核苷酸之數目減少。因此，即使容許較高偽陽性比率用於檢測癌症相關突變，但偽陽性位點之絕對數目可保持為相對較低等級。較高偽陽性比率之容許可容許用於定義欲使用血漿中之單核苷酸變異之最小發生次數(r)之較不嚴格準則。此可產生檢測真實癌症相關突變之較高靈敏度。

在一個實施例中，可使用 1.5×10^6 之最大可容許偽陽性比率。在此偽陽性比率下，靶向外顯子內之偽陽性位點之總數目可僅為20。使用 1.5×10^6 之最大可容許偽陽性比率的不同測序深度之 r 值示於表2中。在其他實施例中，可使用其他不同之最大可容許偽陽性比率，例如 3×10^6 之1、 10^6 之1或 6×10^5 之1。偽陽性位點之相應總數目可分別為小於10、30及50。在一個實施例中，不同種類之突變可導致不同加權，如上文所述。

特定核苷酸位置之測序深度	用以定義潛在突變之血漿DNA測序數據中存在之核苷酸變化之最小發生次數(r)
<50	4
50 - 125	5
126 - 235	6
236 - 380	7
381 - 560	8
561 - 760	9

表2. 定義特定核苷酸位置之不同測序深度之潛在突變的血漿中存在之核苷酸變化的最小發生次數(r)。最大偽陽性比率固定為 1.5×10^6 個核苷酸中之1個。

VIII. 癌症檢測

如上文所提及，與對癌症等級進行分類之臨限值相比，變體基因座處序列標籤之計數可以各種方式使用以測定參數。一基因座或許多基因座處變體讀段相對於所有讀段之分數濃度係可使用之另一參數。以下計算參數及臨限值之一些實例。

A. 參數之測定

若於第一對偶基因之特定基因座處CG同型且在生物試樣(例如，血漿)中觀察到變體對偶基因，則分數濃度可計算為 $2p / (p+q)$ ，其中 p 係具有變體對偶基因之序列標籤之數目且 q 係具有CG之第一對偶基因之序列標籤之數目。此式假定僅腫瘤之單倍型中之一者具有變體，其通常可如此。因此，對於每一同型基因座，可計算分數濃度。可將分數濃度平均化。在另一實施例中，計數 p 可包括所有基因座之序列標籤之數目，且類似地對於計數 q ，用以測定分數濃度。現闡述實例。

探索4名HCC患者之血漿中之腫瘤源單核苷酸變體(SNV)之全基因組檢測。對腫瘤DNA及棕黃色塗層DNA測序至分別29.5倍(範圍，

27倍至33倍)及43倍(範圍，39倍至46倍)單倍體基因組覆蓋之平均深度。比較來自4名HCC患者之每一者之腫瘤DNA及棕黃色塗層DNA之MPS數據，且利用嚴格生物資訊演算法使腫瘤DNA中而非棕黃色塗層DNA中存在之SNV最小化。此演算法需要至少臨限值數目之測序腫瘤DNA片段中(即相應測序標籤中)存在假定SNV，之後可將其分類為真實SNV。藉由慮及(例如)如本文所述之特定核苷酸之測序深度及測序誤差率測定臨限值數目。

圖8係表800，其顯示根據本發明實施例，4名HCC患者在治療之前及之後之結果，包括血漿中之腫瘤源DNA之分數濃度。在4個HCC病例中，腫瘤相關之SNV之數目在1,334至3,171範圍內。在治療之前及之後列舉可在血漿中檢測到之該等SNV之比例。在治療之前，在血漿中檢測到15%-94%腫瘤相關之SNV。在治療之後，百分比介於1.5%-5.5%之間。因此，所檢測SNV之數目與癌症等級有關。此顯示SNV之數目可用作對癌症等級進行分類之參數。

藉由突變體關於總(即，突變體加上野生型)序列之分數計數測定血漿中之腫瘤源DNA之分數濃度。式係 $2p/(p+q)$ ，其中2解釋在腫瘤上突變之僅一個單倍型。該等分數濃度與彼等利用全基因組合計對偶基因喪失(GAAL)分析所測定者充分有關(Chan KC等人 Clin Chem 2013;59:211-24)且其在手術後降低。因此，分數濃度亦顯示為測定癌症等級之有用參數。

SNV分析之分數濃度可傳送腫瘤負載。具有較高腫瘤負載(例如，較高推斷分數濃度)之癌症患者將具有較具有較低腫瘤負載者高之體細胞突變頻率。因此，實施例亦可用於預測。一般而言，具有較高腫瘤負載之癌症患者較彼等具有較低腫瘤負載者具有較差預後。因此，前一組可具有因該疾病死亡之較高機會。在一些實施例中，若可測定(例如使用實時PCR或螢光計)生物試樣(例如血漿)中之DNA之絕

對濃度，則可測定腫瘤相關之遺傳畸變之絕對濃度且使用其進行臨床檢測及/或監測及/或預測。

B. 臨限值之測定

表800可用於測定臨限值。如上文所提及，SNV之數目及藉由SNV分析測定之分數濃度與癌症等級有關。可基於個體測定臨限值。舉例而言，可使用治療前之值測定臨限值。在各個實作中，臨限值可為偏離治療前之絕對值之相對變化。適宜臨限值可為SNV之數目或分數濃度之數目減少50%。該臨限值可提供表800中之每一病例之較低癌症等級的分類。注意，該臨限值可取決於測序深度。

在一個實施例中，臨限值可在試樣中使用，且可慮及或可不慮及參數之治療前之值。舉例而言，可使用100個SNV之臨限值以將個體分類為無癌症或具有低癌症等級。藉由表800中之四個病例之每一者滿足100個SNV之此臨限值。若分數濃度用作參數，則1.0%之臨限值可將HCC1-HCC3分類為實際零癌症等級，且1.5 %之第二臨限值可將HCC4分類為低癌症等級。因此，可使用一個以上臨限值獲得兩個以上分類。

為闡釋其他可能之臨限值，對腫瘤相關之SNV之健康對照之血漿進行分析。可對健康個體進行多個量測以測定自生物試樣預計相對於組成基因組有多少變化之範圍。

圖9係表900，其顯示本發明實施例，在16名健康對照個體中對HCC相關之SNV之檢測。表900可用於估計SNV分析方法之特異性。16個健康對照列舉為不同列。該等列研究針對特定HCC患者檢測之SNV，且顯示具有變體對偶基因之變體基因座處之序列讀段的數目及具有野生型對偶基因(即，CG之對偶基因)之序列讀段的數目。舉例而言，對於HCC1而言，對照C01在該等變體基因座處具有40個變體讀段，但野生型對偶基因為31,261個讀段。最後一列顯示針對HCC1患

者跨越所有SNV之總分數濃度。由於HCC相關之SNV對於HCC患者具有特異性，故HCC相關之SNV之存在代表偽陽性。若如本文所述截止值適於該等表面序列變體，則可濾出所有該等偽陽性。

16個健康對照之血漿中之小數目之該等假定腫瘤相關之突變的存在代表此方法之「隨機噪音」且可能係由於測序誤差。自該噪音估計之平均分數濃度係0.38%。該等值顯示健康個體之範圍。因此，HCC之零癌症等級之分類的臨限值可為約0.5%，此乃因最高分數濃度係0.43%。因此，若自HCC患者移除所有癌細胞，則可預計該等低分數濃度。

再次參見表800，若0.5 %用作零癌症等級之臨限值，則HCC1及HCC3之治療後血漿數據可基於SNV分析測定為具有零等級。HCC2可能分類為自零高一個等級。HCC4亦可分類為自零高一個等級或稍微更高等級，但與治療前試樣相比仍相對較低等級。

在參數對應於變體基因座之數目之一個實施例中，臨限值可為零(即，一個變體基因座可指示非零癌症等級)。然而，在許多設定(例如，深度之設定)下，臨限值可較高，例如，絕對值為5或10。在治療後對人員進行監測之一個實作中，臨限值可為試樣中顯現之SNV之特定百分比(藉由直接分析腫瘤鑑別)。若基因座處所需變體讀段之數目的截止值足夠大，則僅具有一個變體基因座可指示非零癌症等級。

因此，生物試樣(例如，血漿)之DNA之變化(例如，單核苷酸變異)之定量分析可用於診斷、監測及預測癌症。對於癌症之檢測，可將所測試個體之血漿中檢測之單核苷酸變異之數目與一組健康個體中者進行比較。在健康個體中，血漿中之表面單核苷酸變異可由於測序誤差、血細胞及其他器官之非純系突變。已顯示正常健康個體中之細胞可具有小數目之突變(Conrad DF等人 Nat Genet 2011；43:712-4)，如表900中所顯示。因此，一組表面健康個體之血漿中之表面單核苷

酸變異的總體數目可用作參照範圍以測定所測試患者血漿中是否具有對應於非零癌症等級之異常高之單核苷酸變異數目。

用於建立參照範圍之健康個體在年齡及性別方面可與所測試個體匹配。在先前研究中，已顯示體細胞中之突變之數目可隨年齡增加 (Cheung NK等人, JAMA 2012;307:1062-71)。因此，隨著年齡增長，則對於累積細胞之純系者可為「正常的」，即使其在大多數時間相對良性，或變得臨床上明顯可需花費極長時間。在一個實施例中，可針對不同個體組(例如不同年齡、性別、種族及其他參數(例如吸菸狀態、肝炎狀態、酒精、用藥史))產生參照等級。

參照範圍可基於所用截止值(即，基因座處所需變體序列標籤之數目)以及假定偽陽性比率及其他變量(例如，年齡)變化。因此，可針對特定組之一或多個準則測定參照範圍，且相同準則可用於測定試樣之參數。隨後，可比較參數與參照範圍，此乃因二者均係使用相同準則測定。

如上文所提及，實施例可使用多個臨限值來測定癌症等級。舉例而言，因參數低於臨限值，第一等級可測定無癌症體徵，且至少第一癌症等級可為腫瘤前等級。其他等級可對應於不同癌症階段。

C. 實驗變量之依賴性

測序深度對於建立少數(例如腫瘤)基因組之最小檢測臨限值可為重要的。舉例而言，若使用10個單倍體基因組之測序深度，則可甚至在無任何誤差情況下利用測序技術檢測之最小腫瘤DNA濃度係1/5，即20%。另一方面，若使用100個單倍體基因組之測序深度，則可下降至2%。此分析係參照僅分析一個突變基因座之情況。然而，在分析更多突變基因座時，最小腫瘤DNA濃度可較低且係由二項機率函數決定。舉例而言，若測序深度係10倍且腫瘤DNA之分數濃度係20%，則檢測突變之機會係10%。然而，若具有10個突變，則檢測至少一個

突變之機會可為 $1 - (1 - 10\%)^{10} = 65\%$ 。

存在增加測序深度之若干效應。測序深度愈高，則可觀察到更多測序誤差，參見圖4及5。然而，在較高測序深度下，可更容易地區分測序誤差與由於細胞(例如癌細胞)子群之純系擴增發生的突變，此乃因測序誤差可在基因組中隨機發生，但突變可在給定細胞群體之相同位置處發生。

測序深度愈高，則可鑑別「健康細胞」之更多突變。然而，在該等健康細胞無純系擴增且其突變概況不同時，則該等健康細胞之突變可因其於血漿中之發生頻率(例如，藉由使用所需數目之呈現突變之讀段的截止值N，例如使N等於2、3、4、5或更大)而與突變區分開來。

如上文所提及，臨限值可取決於可純系擴增之健康細胞之突變量，且因此可能不經由其他機制濾出。可藉由分析健康個體獲得可預計之此差異。隨著隨時間流逝發生純系擴增，患者之年齡可影響健康個體中可看到之差異，且因此，臨限值可取決於年齡。

D. 與靶向方法之組合

在一些實施例中，隨機測序可與靶向方法組合使用。舉例而言，在呈現癌症患者時，可實施血漿試樣之隨機測序。可針對拷貝數畸變及SNV分析血漿DNA之測序數據。可出於連續監測目的靶向顯示畸變(例如，擴增/缺失或高密度SNV)之區域。監測可有效地作為單一程序在一段時間內進行或在隨機測序後立刻進行。對於靶向分析，基於溶液相雜交之捕獲方法已成功用於富集血漿DNA用於非侵入性產前診斷(Liao GJ等人 Clin Chem 2011；57:92-101)。該等技術係如上文所提及。因此，靶向及隨機方法可組合用於癌症檢測及監測。

因此，可使用上文提及之非靶向、全基因組方法實施發現潛在突變之基因座之靶向測序。可使用溶液相或固相雜交技術(例如使用

Agilent SureSelect、NimbleGen序列捕獲或Illumina靶向再測序系統)實施該靶向測序，之後實施大規模平行測序。另一方法係實施靶向測序之擴增(例如基於PCR)系統(Forsheew T等人 Sci Transl Med 2012；4:135ra68)。

IX. 分數濃度

腫瘤DNA之分數濃度可用於測定在基因座鑑別為突變之前，基因座處變化之所需數目之截止值。舉例而言，若已知分數濃度相對較高，則高截止值可用於濾出更多偽陽性，此乃因已知相對較高數目之變體讀段應對於真實SNV存在。另一方面，若分數濃度較低，則可能需要較低截止值以便不會漏掉一些SNV。在此情形下，可藉由與SNV分析不同之方法測定分數濃度，其中該分數濃度用作參數。

可使用各種技術測定分數濃度，本文闡述其中之一些技術。該等技術可用於測定混合物(例如含有腫瘤細胞與惡性細胞之混合物之活檢試樣或含有自腫瘤細胞釋放之DNA及自非惡性細胞釋放之DNA之癌症患者的血漿試樣)中之腫瘤源DNA的分數濃度。

A. GAAL

全基因組合計對偶基因喪失(GAAL)分析喪失異型接合性之基因座(Chan KC等人 Clin Chem 2013；59:211-24)。對於異型組成基因組CG之位點而言，腫瘤經常具有缺失一個對偶基因之基因座。因此，該基因座之序列讀段將較一個對偶基因更多地顯示另一對偶基因，其中差異與試樣中之腫瘤DNA之分數濃度成正比。該計算之實例如下。

利用Affymetrix Genome-Wide Human SNP Array 6.0系統對來自HCC患者之棕黃色塗層及腫瘤組織之DNA萃取物進行基因分型。利用Affymetrix Genotyping Console 4.1版處理微陣列數據。利用Birdseed 2版演算法實施基因分型分析及單核苷酸多態性(SNP)稱為。棕黃色塗層及腫瘤組織之基因分型數據用於鑑別異型接合性喪失(LOH)區域且

用於實施拷貝數分析。利用具有Affymetrix之默認參數及100 bp之最小基因組區段大小及片段內之最少5個遺傳標記的Genotyping Console實施拷貝數分析。

具有LOH之區域鑑別為腫瘤組織中具有1個拷貝且棕黃色塗層中具有2個拷貝之區域，該等區域內之SNP在棕黃色塗層中係異型的但在腫瘤組織中係同型的。對於在腫瘤組織中呈現LOH之基因組區域而言，將在棕黃色塗層中存在但在腫瘤組織中不存在或具有減少強度之SNP對偶基因視為染色體區域之缺失區段上之對偶基因。將在棕黃色塗層及腫瘤組織二者中存在之對偶基因視為源自染色體區域之非缺失區段。對於腫瘤中具有單一拷貝喪失之所有染色體區域，對帶有缺失對偶基因及非缺失對偶基因之序列讀段的總數目進行計數。該兩個值之差用於使用以下等式推斷試樣中之腫瘤源DNA之分數濃度(F_{GAAL})：

$$F_{GAAL} = \frac{N_{\text{非缺失}} - N_{\text{缺失}}}{N_{\text{非缺失}}}$$

其中 $N_{\text{非缺失}}$ 代表帶有非缺失對偶基因之序列讀段的總數目且 $N_{\text{缺失}}$ 代表帶有缺失對偶基因之序列讀段的總數目。

B. 使用基因組代表之估計

關於GAAL技術之問題在於鑑別特定基因座(即呈現LOH者)且僅使用與該等基因座比對之序列讀段。該要求可增加額外步驟，且因此有花費。現闡述僅使用拷貝數(例如，序列讀段密度)之實施例。

通常在癌症基因組中觀察到染色體畸變(例如擴增及缺失)。癌症組織中觀察之染色體畸變通常涉及亞染色體區域且該等變異可比1 Mb短。而且，癌症相關之染色體畸變在不同患者中係異質的，且因此在不同患者中可影響不可區域。對於癌症基因組中發現之數十、數百或甚至數千拷貝數畸變而言，其亦不常見。所有該等因素使得測定

腫瘤DNA濃度困難。

實施例涉及自腫瘤相關之染色體畸變產生之定量變化的分析。在一個實施例中，使用大規模平行測序藉由(例如) Illumina HiSeq2000測序平臺對含有源自癌細胞及正常細胞之DNA的DNA試樣進行測序。源DNA可為血漿或其他適宜生物試樣中之無細胞DNA。

在腫瘤組織中擴增之染色體區域可具有進行測序之增加機率且在腫瘤組織中缺失之區域可具有進行測序之減少機率。因此，與擴增區域比對之序列讀段之密度可增加且與缺失區域比對之序列讀段之密度可減少。變化程度與DNA混合物中之腫瘤源DNA之分數濃度成正比。腫瘤組織之DNA之比例愈高，則可由染色體畸變引起之變化愈大。

1. 具有高腫瘤濃度之試樣中之估計

自四名肝細胞癌患者之腫瘤組織萃取DNA。使用Covaria DNA音波處理系統使DNA片段化且如所述用Illumina HiSeq2000平臺進行測序(Chan KC等人 Clin Chem 2013 ; 59:211-24)。比對序列讀段與人類參照基因組(hg18)。隨後將基因組分成1 Mb庫(區域)且在在如所述調節GC偏誤後計算每一庫之序列讀段密度(Chen EZ等人 PLoS One. 2011 ; 6:e21791)。

在比對序列讀段與參照基因組後，可計算各個區域之序列讀段密度。在一個實施例中，序列讀段密度係測定為定位至特定庫(例如，1 Mb區域)之讀段之數目除以可與參照基因組(例如，參照基因組中之獨特位置)比對之總序列讀段的比例。預計與在腫瘤組織中擴增之染色體區域重疊之庫具有較來自無該等重疊之庫之序列讀段密度高之序列讀段密度。另一方面，預計與缺失之染色體區域重疊之庫較無該等重疊之序列讀段密度低之序列讀段密度。具有染色體變異與無染色體畸變之區域之間之序列讀段密度之差的量值主要受試樣中之腫瘤

源DNA之比例及腫瘤細胞中之擴增/缺失程度影響。

可使用各種統計模型來鑑別具有對應於不同類型之染色體畸變之序列讀段密度的庫。在一個實施例中，可使用正常混合物模型(McLachlan G及Peel D. Multivariate normal mixtures. In Finite mixture models 2004: p81-116. John Wiley & Sons Press)。亦可使用其他統計模型，例如二項混合物模型及蔔瓦松回歸模型(McLachlan G及Peel D. Mixtures with non-normal components, Finite mixture models 2004: p135-174. John Wiley & Sons Press)。

可使用如自棕黃色塗層DNA之測序測定之相同庫的序列讀段密度正規化庫之序列讀段密度。不同庫之序列讀段密度可受特定染色體區域之序列上下文影響，且因此，正規化可有助於更精確地鑑別顯示畸變之區域。舉例而言，不同染色體區域之可定位性(其係指將序列比對回其初始位置之機率)可不同。另外，拷貝數之多態性(即拷貝數差異)亦可影響庫之序列讀段密度。因此，利用棕黃色塗層DNA之正規化可潛在地使與不同染色體區域之間之序列上下文差異相關之變化最小化。

圖10A顯示根據本發明實施例，HCC患者之腫瘤試樣的序列讀段密度之分佈曲線1000。在自HCC患者手術切除後獲得腫瘤組織。x軸代表患者之腫瘤組織與棕黃色塗層之間之序列讀段密度的比率(R)的 $\log_2$ 。y軸代表庫之數目。

可使用正常混合物模型將峰擬合至分佈曲線以代表具有缺失、擴增及無染色體畸變之區域。在一個實施例中，可藉由Akaike資訊準則(AIC)在不同可信值間測定峰之數目。 $\log_2 R = 0$ (即 $R = 1$)之中心峰代表無任何染色體畸變之區域。左峰(相對於中心峰)代表具有一個拷貝喪失之區域。右峰(相對於中心峰)代表具有一個拷貝擴增之區域。

代表擴增及缺失區域之峰之間之距離可反映腫瘤源DNA之分數

濃度。距離愈大，試樣中之腫瘤源DNA之分數濃度可愈高。可藉由此基因組代表方法(表示為 F_{GR})使用以下等式測定試樣中之腫瘤源DNA之分數濃度： $F_{GR} = R_{右} - R_{左}$ ，其中 $R_{右}$ 係右峰之 R 值且 $R_{左}$ 係左峰之 R 值。最大差可為1，對應於100%。自HCC患者獲得之腫瘤試樣中之腫瘤源DNA的分數濃度估計為66%，其中 $R_{右}$ 及 $R_{左}$ 之值分別係1.376及0.712。

為驗證此結果，亦使用採用全基因組合計對偶基因喪失(GAAL)分析之另一方法獨立地測定腫瘤DNA之比例之分數濃度(Chan KC等人 Clin Chem 2013；59:211-24)。表3使用基因組代表(F_{GR})及GAAL(F_{GAAL})方法顯示四名HCC患者之腫瘤組織中之腫瘤源DNA之分數濃度。藉由該兩種不同方法測定之值彼此非常一致。

HCC腫瘤	F_{GAAL}	F_{GR}
1	60.0%	66.5%
2	60.0%	61.4%
3	58.0%	58.9%
4	45.7%	42.2%

表3顯示藉由GAAL及基因組代表(GR)測定之分數濃度。

2. 具有低腫瘤濃度之試樣中之估計

上述分析已顯示，在50%以上之試樣DNA係腫瘤源DNA時，即在腫瘤DNA係大多數比例時，可使用基因組代表方法量測腫瘤DNA之分數濃度。在前述分析中，已顯示此方法亦可適於腫瘤源DNA代表微小比例(即，低於50%)之試樣。可含有微小比例之腫瘤DNA之試樣包括(但不限於)癌症患者之血液、血漿、血清、尿、胸水、腦脊髓液、眼淚、唾液、腹水及糞便。在一些試樣中，腫瘤源DNA之分數濃度可為49%、40%、30%、20%、10%、5%、2%、1%、0.5%、0.1%或更低。

對於一些試樣，代表具有擴增及缺失之區域之序列讀段密度的峰可不如如上文所闡釋具有相對較高濃度之腫瘤源DNA之試樣中明顯。在一個實施例中，可藉由與已知不含癌症DNA之參照試樣進行比較鑑別具有癌細胞中之染色體畸變之區域。舉例而言，無癌症之個體之血漿可用作參照以測定染色體區域之序列讀段密度的規範性範圍。可比較所測試個體之序列讀段密度與參照組之值。在一個實施例中，可測定序列讀段密度之平均及標準偏差(SD)。對於每一庫而言，比較所測試個體之序列讀段密度與參照組之平均值以使用下式測定z分

值：

$$z \text{ 分 值 } = \frac{(GR_{\text{測試}} - \overline{GR_{\text{參照}}})}{SD_{\text{參照}}}$$

，其中 $GR_{\text{測試}}$ 代表癌症患者之序列讀段密度； $\overline{GR_{\text{參照}}}$ 代表參照個體之序列讀段密度且 $SD_{\text{參照}}$ 代表參照個體之序列讀段密度之SD。

z分値 < -3之區域表示癌症患者中之特定庫之序列讀段密度顯著過低，表明腫瘤組織中存在缺失。z分値 > 3之區域表示癌症患者中之特定庫之序列讀段密度之顯著過高，表明腫瘤組織中存在擴增。

隨後，可構築所有庫之z分値之分佈以鑑別具有不同數目之拷貝增加及喪失(例如缺失1或2個拷貝之染色體)及擴增(從而產生1、2、3及4個額外拷貝之染色體)的區域。在一些情形下，可涉及一個以上染色體或染色體之一個以上區域。

圖10B顯示根據本發明實施例，HCC患者之血漿中之所有庫之z分值的分佈曲線1050。峰(自左至右)代表1個拷貝喪失，無拷貝變化，1個拷貝增加及2-拷貝增加，將其擬合至z分値分佈。隨後可使用(例如)如上文所述正常混合物模型鑑別具有不同類型之染色體畸變之區域。

隨後可自呈現1個拷貝增加或1個拷貝喪失之庫之序列讀段密度

推斷試樣中之癌症DNA之分數濃度(F)。針對特定庫測定之分數濃度

可計算為 $F = \frac{|GR_{\text{測試}} - \overline{GR_{\text{參照}}}| \times 2}{GR_{\text{參照}}} \times 100\%$ 。此亦可表示為： $F = \frac{|z - score \times SD_{ref}|}{GR_{ref}} \times 2$ ，
其可重新寫為： $F = |z - score| \times CV \times 2$ ，其中CV係參照個體之序列讀段密度之量測之變化係數；且 $CV = \frac{SD_{ref}}{GR_{ref}}$ 。

在一個實施例中，將庫之結果組合。舉例而言，可將顯示1個拷貝增加之庫之z分值平均化或將所得F值平均化。在另一實作中，藉由統計模型測定用於推斷F之z分值之值且由圖10B及圖11中顯示之峰表示。舉例而言，右峰之z分值可用於測定呈現1個拷貝增加之區域之分數濃度。

在另一實施例中，z分值< -3及z分值> 3之所有庫皆可歸因於分別具有單一拷貝喪失及單一拷貝增加之區域，此乃因該兩種染色體畸變最常見。在具有染色體畸變之庫之數目相對較小且正常分佈之擬合可不精確時，此近似情況最有用。

圖11顯示根據本發明實施例，HCC患者之血漿之z分值的分佈曲線1100。儘管與染色體畸變重疊之庫之數目相對較小，但將z分值 < -3及z分值 > 3之所有庫分別擬合至單一拷貝喪失及單一拷貝增加之正常分佈。

使用GAAL分析及此基於GR之方法測定四名HCC患者之血漿中之腫瘤源DNA之分數濃度。結果示於表4中。如可見，所推斷分數代表在GAAL分析與GR分析之間充分相關。

	血漿中之腫瘤源DNA之	
分數濃度 試樣	GAAL分析	GR分析
病例11	4.3%	4.5%
病例13	5%	5.5%

病例23	52%	62%
病例27	7.6%	6.1%

表4. 藉由染色體畸變之分析推斷之血漿中之腫瘤源DNA的分數濃度.

C. 測定分數濃度之方法

圖12係根據本發明實施例測定包括無細胞DNA之生物試樣中之腫瘤DNA的分數濃度之方法1200的流程圖。方法1200可經由各種實施例(包括上述實施例)實施。

於框1210處，接收生物試樣中複數個DNA片段中之每一者之一或多個序列標籤。框1210可如本文針對其他方法所述實施。舉例而言，可自血漿試樣對DNA片段之一端進行測序。在另一實施例中，可對DNA片段之兩端進行測序，藉此容許估計一段片段。

於框1220處，測定序列標籤之基因組位置。基因組位置可如本文所述藉由(例如)比對序列標籤與參照基因組來測定。若對片段之兩端進行測序，則可成對比對配對標籤，且該兩個標籤之間之距離限制在指定距離(例如，500或1,000個鹼基)以內。

於框1230，對於多個基因組區域中之每一者，自具有基因組位置在該基因組區域內之序列標籤測定該基因組區域內之DNA片段之各別量。該等基因組區域可為參照基因組中相等長度之非重疊庫。在一個實施例中，可計數與一庫比對之標籤之數目。因此，每一庫可具有相應數目之比對標籤。可計算一直方圖說明該等庫具有特定數目之比對標籤的頻率。可針對各具有相同長度之基因組區域(例如1 Mb庫)實施方法1200，其中該等區域不重疊。在其他實施例中，可使用可慮及之不同長度，且該等區域可重疊。

於框1240，正規化各別量以獲得各別密度。在一個實施例中，正規化各別量以獲得各別密度包括使用相同總數目之比對參照標籤以



測定各別密度及參照密度。在另一實施例中，可將各別量除以比對參照標籤之總數目。

於框1250，比較各別密度與參照密度以鑑別該基因組區域是否呈現1個拷貝喪失或1個拷貝增加。在一個實施例中，計算各別密度與參照密度間之差(例如作為測定z分値之部分)並與一截止值比較。在各個實施例中，參照密度可獲自健康細胞之試樣(例如獲自棕黃色血球層(buffy coat))或獲自各別量本身(例如取中值或平均值，在假設大部分區域不呈現喪失或增加之下)。

於框1260，自一或多個鑑別呈現1個拷貝喪失之各別密度或自一或多個鑑別呈現1個拷貝增加之各別密度計算第一密度。第一密度可對應於僅一個基因組區域，或可自多個基因組區域之密度測定。舉例而言，第一密度可自具有1個拷貝喪失之各別密度計算。該等各別密度提供腫瘤中該區域之缺失所產生之密度差量的量度，給予該腫瘤濃度。類似地，若第一密度係來自具有1個拷貝增加之各別密度，則可獲得腫瘤中該區域之倍增所產生之密度差量的量度。上述部分闡述可如何使用多個區域之密度來測定欲用於第一密度之平均密度之各種實例。

於框1270處，藉由比較第一密度與另一密度以獲得差來計算分數濃度。利用參照密度正規化該差，此可在框1270中完成。舉例而言，可利用參照密度藉由用該差除以參照密度正規化該差。在另一實施例中，可在前述框正規化該差。

在一個實作中，另一密度係參照密度，例如，如上述部分2中所述。因此，計算分數濃度可包括將該差乘以2。在另一實作中，另一密度係自鑑別為呈現1個拷貝喪失之各別密度計算的第二密度(其中第一密度係使用鑑別為呈現1個拷貝增加之各別密度計算)，例如，如上述部分1所述。在此情形下，可藉由計算第一密度與參照密度之第一

比率(例如, $R_{右}$)並計算第二密度與參照密度之間之第二比率($R_{左}$)測定正規化差, 其中該差係第一比率與第二比率之間之差異。如上文所述, 可藉由將峰擬合至各別密度之直方圖之分佈曲線實施呈現1個拷貝喪失或1個拷貝增加之基因組區域之鑑別。

總之, 實施例可分析不同染色體區域中之血漿DNA之基因組代表, 以同時測定染色體區域在腫瘤組織中是否擴增或缺失且若該區域擴增或缺失則使用其基因組代表來推斷腫瘤源DNA之分數濃度。一些實作使用正常混合物模型來分析不同區域之基因組代表之總體分佈, 以便測定與不同類型之畸變相關之基因組代表, 即1、2、3或4個拷貝增加及1或2個拷貝喪失。

實施例相對於其他方法(例如全基因組合計對偶基因喪失(GAAL)方法(美國專利申請案13/308,473; Chan KC等人 Clin Chem 2013; 59:211-24)及腫瘤相關之單核苷酸突變之分析(Forsheew T等人 Sci Transl Med. 2012; 4:136ra68)具有若干優勢。定位至具有染色體畸變之區域之所有序列讀段皆可用於測定區域之序列讀段密度, 且因此, 可提供關於腫瘤DNA之分數濃度的資訊。另一方面, 在GAAL分析中, 僅覆蓋在個體中異型之單核苷酸且位於具有染色體增加或喪失之染色體區域內的序列讀段可提供資訊。類似地, 對於癌症相關突變之分析而言, 僅覆蓋突變之序列讀段可用於推斷腫瘤DNA濃度。因此, 實施例可容許更成本有效地使用測序數據, 此乃因在與其他方法相比時, 可能需要相對較少測序讀段以在腫瘤源DNA之分數濃度之估計中達成相同程度之精度。

X. 替代方法

除使用在序列標籤上觀察到之特定突變之次數作為鑑別基因座為真實突變的準則(藉此調節陽性預測值)外, 可利用其他技術替代或另外使用在鑑別癌性突變中提供更大預測值之截止值。舉例而言, 在

處理測序數據時可藉由(例如)慮及所測序核苷酸之品質評分使用不同嚴格度之生物資訊過濾器。在一個實施例中，可使用具有不同測序誤差概況之DNA測序儀及測序化學方法。具有較低測序誤差率之測序儀及化學方法可產生較高陽性預測值。亦可使用相同DNA片段之重複測序以增加測序精度。一種可能之策略係Pacific Biosciences之圓形共有測序策略。

在另一實施例中，可將測序片段上之大小資訊納入數據之解釋中。由於在血漿中腫瘤源DNA比非腫瘤源DNA短(參見美國專利申請案第13/308,473號)，故含有潛在腫瘤源突變之較短血漿DNA片段的陽性預測值可高於較長血漿DNA片段的陽性預測值。若實施血漿DNA之末端配對測序，則可容易地獲得大小數據。作為替代，可使用具有長讀段長度之DNA測序儀，因此產生全長之血漿DNA片段。亦可在DNA測序之前實施血漿DNA試樣之大小分級分離。可用於大小分級分離之方法之實例包括凝膠電泳，使用微滴方法(例如Caliper LabChip XT系統)及尺寸排除離心管。

在又一實施例中，若主要關注血漿中之較短DNA片段，則預計具有非血液癌症之患者之血漿中之腫瘤相關之突變的分數濃度可增加。在一個實作中，可比較兩個或更多個不同大小分佈之DNA片段中之血漿中之腫瘤相關之突變的分數濃度。在與較大片段比較時，具有非血液癌症之患者將在較短片段中具有較高分數濃度之腫瘤相關之突變。

在一些實施例中，可組合相同血樣之兩個或更多個等分樣品或在相同時機或不同時機時獲取之兩個或更多個試樣的測序結果。在一個以上等分樣品或試樣中觀察到之潛在突變可具有較高陽性預測值之腫瘤相關之突變。陽性預測值可隨顯示該突變之試樣之數目增加。在不同時間點時獲取之血漿試樣中存在之潛在突變可視為潛在突變。

XI. 實例

以下係實例性技術及數據，不應將其認為限於本發明之實施例。

A. 材料及方法

關於試樣收集，募集肝細胞癌(HCC)患者、B型慢性肝炎之攜帶者及患有同步乳癌及卵巢癌之患者。所有HCC患者皆具有Barcelona臨床肝癌A1期疾病。將所有參與者之周邊血樣收集至含有EDTA之管中。HCC患者之腫瘤組織係在其癌症切除手術期間獲得。

於4°C下將周邊血樣以1,600 g離心10 min。於4°C下將血漿部分以16,000 g重新離心10 min且隨後於80°C下儲存。根據QIAamp DSP DNABlood Mini Kit (Qiagen)之血液及體液方案自4.8 mL血漿萃取無細胞DNA分子。將血漿DNA利用SpeedVac濃縮器(Savant DNA120; Thermo Scientific)濃縮成每情形40- μ l最終體積用於DNA-測序文庫之隨後製備。

根據QIAamp DSP DNA Blood Mini Kit之血液及體液方案自患者之棕黃色塗層試樣萃取基因組DNA。利用QIAamp DNA Mini Kit (Qiagen)自腫瘤組織萃取DNA。

根據製造商之說明利用末端配對試樣製備套組(Paired-End Sample Preparation Kit) (Illumina)構築基因組DNA試樣之測序文庫。簡言之，首先利用Covaris S220焦點超音波發生器將1至5微克基因組DNA剪切成200-bp片段。其後，利用T4 DNA聚合酶及Klenow聚合酶重新末端配對DNA分子；隨後使用T4多核苷酸激酶磷酸化5'端。利用3'-至-5'外切酶缺陷性Klenow片段產生3'突出端。將Illumina銜接體寡核苷酸連接至黏稠端。利用12個循環PCR富集銜接體連接之DNA。由於血漿DNA分子係短的片段且血漿試樣中之總DNA之量相對較小，故在自血漿試樣構築DNA文庫時，忽略片段化步驟且使用15個循環

PCR。

使用Agilent 2100 Bioanalyzer (Agilent Technologies)檢查銜接體連接之DNA文庫之品質及大小。隨後根據製造商之說明藉由KAPA文庫定量套組(Kapa Biosystems)量測DNA文庫。將DNA文庫稀釋並雜交至末端配對測序流動小室。在具有TruSeq PE Cluster Generation Kit v2 (Illumina)之cBot簇生成系統(Illumina)上生成DNA簇，之後在具有TruSeq SBS Kit v2 (Illumina)之HiSeq 2000系統(Illumina)上進行51_2個循環或76_2個循環之測序。

藉助短寡核苷酸比對程式2 (Short Oligonucleotide Alignment Program 2, SOAP2)以末端配對模式分析末端配對測序數據。對於每一末端配對讀段，比對每一端之50 bp或75 bp與非重複遮蔽之參照人類基因組(hg18)。對於每一端之比對容許至多2個核苷酸失配。隨後分析2個末端之該等潛在比對之基因組坐標以測定是否任何組合皆可容許2個末端與相同染色體進行比對，該染色體具有正確定向，跨越小於或等於600 bp之插入物大小，且定位至參照人類基因組中之單一位置。倍增讀段定義為末端配對讀段，其中插入物DNA分子顯示人類基因組中之相同起始及末端位置；如前文所述移除倍增讀段(Lo等人 *Sci Transl Med* 2010；2: 61ra91)。

在一些實施例中，對配對腫瘤及組成DNA試樣進行測序以鑑別腫瘤相關之單核苷酸變體(SNV)。在一些實作中，主要關注組成DNA (在此實例中為棕黃色塗層DNA)中之同型位點處發生之SNV。原則上，腫瘤組織之測序數據中檢測到但組成DNA中不存在之任何核苷酸變異可為潛在突變(即，SNV)。然而，由於測序誤差(0.1%-0.3%測序核苷酸)，若將腫瘤組織之測序數據之任何核苷酸變化的單次發生視為腫瘤相關之SNV，則可在基因組中鑑別到數百萬偽陽性。一種減少偽陽性之數目之方式可為在將稱為腫瘤相關之SNV之前，制定觀察腫

瘤組織中之測序數據之相同核苷酸變化的多次發生的準則。

由於測序誤差之發生係隨機過程，故由於測序誤差產生之偽陽性之數目可隨將所觀察SNV定性為腫瘤相關之SNV所需之發生次數的增加而以指數方式減少。另一方面，偽陽性之數目可隨測序深度增加而增加。可利用葛瓦松及二項分佈函數預測該等關係。實施例可測定將所觀察SNV定性為腫瘤相關之發生次數的動態截止值。實施例可慮及腫瘤測序數據中特定核苷酸之實際覆蓋率、測序誤差率、所容許之最大偽陽性比率及突變檢測之期望靈敏度。

在一些實例中，設定極嚴格準則以減少偽陽性。舉例而言，可能需要組成DNA測序中完全不存在突變，且特定核苷酸位置之測序深度必須為20倍。在一些實作中，發生次數之截止值達成小於 10^{-7} 之偽陽性檢測比率。在一些實例中，亦濾出在著絲粒、末端著絲粒及低複雜性區域內之SNV以使由於比對假像之偽陽性最小化。另外，亦移除定位至dbSNP構建135數據庫中之已知SNP之假定SNV。

B. 切除之前及之後

圖13A顯示根據本發明實施例，在診斷時分析患有卵巢癌及乳癌之患者血漿中之突變之表1300。因此，證實患有雙側卵巢癌及乳癌之患者的實例。比較血漿之測序數據與患者(棕黃色塗層)之組成DNA之測序結果。將血漿中存在但組成DNA中不存在之單核苷酸變化視為潛在突變。在兩個位點處對患者右側及左側之卵巢癌各自取樣，從而得到總共四個腫瘤試樣。腫瘤突變係在所有四個卵巢腫瘤組織中於四個不同位點處檢測之突變。

藉由測序在血漿中檢測到超過360萬個單核苷酸變化至少一次。在該等變化中，在腫瘤組織中僅檢測到2,064個變化，從而產生0.06%之陽性預測值。使用在血漿中檢測到至少兩次之準則，潛在突變之數目顯著減少99.5%至18,885。腫瘤突變之數目僅減少3%至2,003，且陽

性預測值增加至11%。

使用在血漿中檢測至少5次之準則，僅檢測到2,572個潛在突變，且其中，在所有腫瘤組織中檢測到1,814個突變，因此產生71%之陽性預測值。端視所需靈敏度及陽性預測值而定，發生次數(例如2、3、4、6、7、8、9、10等)之其他準則可用於定義潛在突變。用作準則之發生次數愈高，則靈敏度可減少之陽性預測值愈高。

圖13B顯示根據本發明實施例，在腫瘤切除後分析患有雙側卵巢癌及乳癌之患者血漿中之突變之表1350。實施患者之手術切除。在切除卵巢腫瘤及乳癌後一天取血樣。隨後對血漿DNA進行測序。對於此實例，僅分析卵巢癌之突變。血漿試樣中至少檢測到一次超過300萬個潛在突變。然而，使用具有至少5次發生次數之準則，潛在突變之數目減少至238。在與於診斷時所取試樣之潛在突變的數目比較時且使用5個突變之相同準則，觀察到顯著減少。

在一個實施例中，血漿中檢測到之單核苷酸變化之數目可用作檢測、監測並預測癌症患者之參數。不同發生次數可用作準則以達成期望靈敏度及特異性。預計具有較高腫瘤負載且因此預後較差之患者具有血漿中觀察到之較高突變負載。

對於該分析而言，可針對不同類型之癌症建立突變負載曲線。出於監測目的，可看到，對治療有反應之患者血漿中之突變負載可減小。若腫瘤在(例如)復發期間再次出現，則預計突變負載將增加。該監測可容許監測對於患者之選擇治療模式之功效並檢測對特定治療之抗性之出現。

經由可在血漿DNA測序結果中觀察到之特定突變之分析，亦可鑑別可預測靈敏度(例如表皮生長因子受體基因中之突變及對酪胺酸激酶抑制劑治療之反應)及對特定靶向治療之抗性(例如結腸直腸癌中之*KRAS*突變及對帕尼單抗(panitumumab)及西妥昔單抗(cetuximab)治

療之抗性)之靶標，且可指導治療方案之計劃。

以上實例係針對雙側卵巢癌。亦可對乳癌之突變實施相同分析且隨後將能夠追蹤血漿中之該兩種癌症類型之突變。亦可使用類似策略追蹤原發性癌症及其轉移(metastasis或metastase)之突變。

實施例可用於篩選表面健康個體或具有特定風險因子(例如吸菸狀態、病毒狀態(例如肝炎病毒攜帶者、人類乳頭狀瘤病毒感染之個體))之個體之癌症。該等個體血漿中可觀察到之突變負載可產生個體可在特定時間框內發生症狀性癌症之風險。因此，可預計血漿中具有較高突變負載之個體較彼等具有較低突變負載者具有較高風險。此外，血漿中之該突變負載之時間概況亦可為風險之有力指標。舉例而言，若個體每年實施一次血漿突變負載且若突變負載逐漸增加，則應參照此個體以使用(例如)胸部X射線、超音波、電腦斷層掃描術、磁共振成像或正電子發射斷層掃描術進行癌症之額外篩選模式。

C. 用以自測序血漿推斷突變之動態截止值

募集四名患有肝細胞癌(HCC)之患者及一名患有卵巢癌及乳癌之患者用於此研究。對於後一患者，主要關注卵巢癌之分析。在手術切除腫瘤之前及之後自每一患者收集血樣。亦收集所切除腫瘤組織。使用HiSeq2000測序系統(Illumina)對自腫瘤組織萃取之DNA、術前血樣之白血球及術前及術後血漿試樣進行測序。使用短寡核苷酸分析包2(SOAP2)比對測序數據與參照人類基因組序列(hg18) (Li R等人 Bioinformatics 2009; 25: 1966-1967)。將白血球之DNA序列視為每一研究個體之組成DNA序列。

在此實例中，首先自與腫瘤組織無關之血漿DNA測序數據及CG推斷腫瘤相關之SNM。隨後，比較自血漿推斷之結果與自腫瘤組織產生之測序數據(作為黃金標準)以確定推斷結果之精度。就此而言，黃金標準係藉由比較腫瘤組織之測序數據與組成序列以檢查腫瘤組織中

之突變來製得。在此分析中，主要關注所研究個體之組成DNA係同型的核苷酸位置。

1. 非靶向全基因組分析

每一患者之白血球、腫瘤組織及血漿DNA的測序深度示於表5中。

病例	中值測序深度(次)			
	白血球	腫瘤組織	術前血漿	術後血漿
HCC1	39	29	23	24
HCC2	39	29	25	28
HCC3	46	33	18	21
HCC4	46	27	20	23
卵巢癌患者	44	53	37	28

表5. 4個HCC病例之不同試樣的中值測序深度.

如表1中所示定義血漿突變之最小發生次數(r)之動態截止值用於鑑別每一患者之血漿中之突變。由於每一基因座之測序深度可變，故截止值可變，其有效地提供截止值對基因座之讀段之總數目的依賴性。舉例而言，儘管中值深度小於50（表5），但個別基因座之測序深度可變化很大且包含>100次。

除測序誤差外，另一誤差來源可為比對誤差。為使此類誤差最小化，使用Bowtie比對程式重新比對帶有突變之序列讀段與參照基因組(Langmead B等人 Genome Biol 2009, 10:R25)。僅可藉由SOAP2及Bowtie與參照基因組之獨特位置比對的讀段用於血漿突變之下游分析。亦可使用基於不同演算法之比對軟體包之其他組合。

為進一步使實際測序數據中之測序及比對誤差最小化，施用兩種額外過濾演算法以稱為顯示序列讀段中之單核苷酸變異的核苷酸位置：(1) 可使用Bowtie重新比對≥ 70%帶有突變之序列讀段與相同基

因組坐標，其中定位品質 $\geq Q20$ (即未比對機率 $<1\%$)；(2) $\geq 70\%$ 帶有突變之序列讀段並不在序列讀段之兩端(即5'及3'端)之5 bp內。制定此過濾規則，此乃因測序誤差於序列讀段之兩端處更普遍。

在未預先瞭解腫瘤基因組情況下，亦研究影響腫瘤之推斷之因子。一個該參數係血漿中之腫瘤源DNA之分數濃度。此參數可視為另一黃金標準參數且出於參照目的在預先瞭解腫瘤基因組情況下使用GAAL進行推斷。

表6顯示治療之前及期間血漿中檢測到之核苷酸變異。對於HCC1而言，在未預先瞭解腫瘤基因組之情況下，檢測總共961個單核苷酸變異。在血漿中檢測到之該等核苷酸變異中，828個係癌症相關突變。在手術切除HCC後，核苷酸變異之總數目減少至43且其皆非癌症相關突變。

出於參照目的，術前血漿試樣中之腫瘤源DNA之分數濃度係53%且在預先瞭解腫瘤基因組情況下進行推斷。對於HCC2、HCC3及HCC4而言，在未預先瞭解腫瘤基因組情況下，對於術前血漿試樣，血漿中之單核苷酸變異之數目推斷為在27至32範圍內。該等結果與數學預測相容，在約20倍之測序深度下，可在血漿中檢測到極低百分比之癌症相關突變且血漿中檢測到之大部分序列變異係由於測序誤差。在腫瘤切除後，所檢測序列變異之數目顯著變化。出於參照目的，血漿中之腫瘤源DNA之分數濃度推斷為在2.1%至5%範圍內且在預先瞭解腫瘤基因組情況下進行推斷。

	術前血漿			術後血漿		
	腫瘤源DNA之分數濃度	單核苷酸變異之總數目	所鑑別癌症相關突變之數目	腫瘤源DNA之分數濃度	單核苷酸變異之總數目	所鑑別癌症相關突變之數目
HCC1	53%	961	828	0.4%	43	0

HCC2	5%	32	0	0.6%	49	0
HCC3	2.1%	29	0	0.2%	32	0
HCC4	2.6%	27	0	1.3%	35	1
卵巢癌(及乳癌)患者	46%	1718	1502	0.2%	2	0

表6. 血漿中檢測到之核苷酸變異.

2. 外顯子之靶富集

如上文所論述，增加所關注區域之測序深度可增加鑑別血漿中之癌症相關突變之靈敏度及特異性，且因此，增加癌症患者與非癌症個體之間之區分能力。儘管全基因組之測序深度之增加仍花費極大，但一種替代方式係富集特定區域用於測序。在一個實施例中，所選外顯子或實際上整個全外顯子組可靶富集用於測序。此方法可顯著增加靶區域之測序深度，而不增加序列讀段之總量。

使用用於全外顯子組之靶富集之Agilent SureSelect All Exon套組捕獲HCC患者及患有卵巢癌(及乳癌)之患者之血漿DNA的測序文庫。隨後使用HiSeq 2000測序系統對外顯子富集之測序文庫進行測序。比對序列讀段與人類參照基因組(hg18)。在比對後，針對單核苷酸變異分析獨特地定位至外顯子之序列讀段。對於全外顯子組捕獲分析之血漿中之單核苷酸變異的鑑別，使用表2中所示之動態截止值。

圖14A係表1400，其顯示HCC1之血漿DNA中之單核苷酸變異的檢測。在未預先瞭解腫瘤基因組情況下，自靶向測序數據推斷總共57個血漿中之單核苷酸變異。在自腫瘤組織獲得之測序數據之隨後驗證中，發現55個突變係真實腫瘤相關之突變。如上文所論述，術前血漿中之腫瘤源DNA之分數濃度係53%。在腫瘤切除後，在自血漿獲得之靶向測序數據中未檢測到單核苷酸變異。該等結果指示血漿中之單核苷酸變異之數目的定量分析可用於監測癌症患者之疾病進展。

圖14B係表1450，其顯示HCC2之血漿DNA中之單核苷酸變異的檢測。在未預先瞭解腫瘤基因組情況下，自血漿之靶向測序數據推斷總共18個單核苷酸變異。所有該等突變皆係在腫瘤組織中發現。如上文所論述，術前血漿中之腫瘤源DNA之分數濃度係5%。在腫瘤切除後，血漿中未檢測到單核苷酸變異。與血漿中具有較高分數濃度之腫瘤源DNA的HCC1相比，在涉及HCC2之病例之血漿中檢測到較少單核苷酸變異。該等結果表明血漿中之單核苷酸變異之數目可用作參數以反映血漿中之腫瘤源DNA之分數濃度，且因此反映患者中之腫瘤負載，此乃因已顯示血漿中之腫瘤源DNA之濃度與腫瘤負載正相關(Chan KC等人 Clin Chem 2005；51:2192-5)。

圖15A係表1500，其顯示HCC3之血漿DNA中之單核苷酸變異的檢測。在未預先瞭解腫瘤基因組情況下，未自靶向測序數據觀察到切除前及切除後血漿試樣中之任何單核苷酸變異。此可能係由於此患者血漿中腫瘤源DNA之相對較低分數濃度(2.1%)。在具有較低分數濃度之腫瘤源DNA之病例中，預測測序深度進一步增加會改良檢測癌症相關突變之靈敏度。

圖15B係表1550，其顯示HCC4之血漿DNA中之單核苷酸變異的檢測。在未預先瞭解腫瘤基因組情況下，自血漿之靶向測序數據推斷總共3個單核苷酸變異。所有該等突變皆係在腫瘤組織中發現。與在血漿中具有較高分數濃度之腫瘤源DNA的HCC1及HCC2相比，在血漿中之分數腫瘤DNA為2.6%的病例HCC4之血漿中檢測到較少單核苷酸變異。該等結果表明血漿中之單核苷酸變異之數目可用作參數以反映血漿中之腫瘤源DNA之分數濃度及患者中之腫瘤負載。

圖16係表1600，其顯示患有卵巢癌(及乳癌)之患者之血漿DNA中之單核苷酸變異的檢測。在未預先瞭解腫瘤基因組情況下，自血漿之靶向測序數據推斷總共64個單核苷酸變異。其中，在卵巢腫瘤組織中

發現59個單核苷酸變異。血漿中之卵巢腫瘤源DNA的估計分數濃度為46%。在切除卵巢癌後，在血漿中檢測到單核苷酸變異之總數目顯著減少。

除使用 SureSelect 靶富集系統 (Agilent) 外，亦使用 Nimblegen SeqCap EZ Exome+UTR靶富集系統(Roche)以自外顯子富集序列用於測序。Nimblegen SeqCap系統覆蓋基因組之外顯子區域以及5’及3’未轉譯區域。分析四名HCC患者、兩名健康對照個體及兩名無癌症之B型慢性肝炎攜帶者的治療前血漿試樣(表7)。在其他實施例中，可使用其他靶富集系統(包括但不限於彼等使用溶液相或固相雜交者)。

	治療前血漿			治療後血漿	
	藉由GAAL 分析之血 漿中之腫 瘤源DNA 的分數濃 度	血漿中 檢測到 之滿足 動態截 止值之 序列變 化的數 目	與相應腫瘤 組織中檢測 到之突變重 疊之序列變 化的數目	血漿中 檢測到 之滿足 動態截 止值之 序列變 化的數 目	與相應腫瘤 組織中檢測 到之突變重 疊之序列變 化的數目
HCC1	53%	69	64	1	1
HCC2	5%	51	47	3	0
HCC3	2.1%	0	0	1	0
HCC4	2.6%	8	7	0	0

表7. 使用用於序列捕獲之Nimblegen SeqCap EZ Exome+UTR靶富集系統之四名HCC患者(HCC1-4)之全外顯子組測序結果。由於PCR-倍增讀段之較高百分比，HCC3之治療前血漿之測序分析係次優的。

在兩名B型慢性肝炎攜帶者及兩名健康對照個體中，檢測到一個或更少滿足動態截止值準則之單核苷酸變異(表8)。在四名HCC患者中之三名中，血漿中檢測到之滿足動態截止值要求之序列變異的數目

係至少8。在HCC3中，未檢測到滿足動態截止值之SNV。在此試樣中，在測序讀段中存在高比例之PCR-倍增讀段，從而導致非倍增測序讀段之數目較低。在手術切除腫瘤後觀察到血漿中可檢測之SNV之顯著減少。

	血漿中檢測到之滿足動態截止值之序列變化的數目
HBV1	0
HBV2	1
Ctrl1	1
Ctrl2	1

表8. 使用用於序列捕獲之Nimblegen SeqCap EZ Exome+UTR靶富集系統之2名B型慢性肝炎B攜帶者(HBV1及HBV2)及2名健康對照個體(Ctrl1及Ctrl2)的全外顯子組測序結果.

XII. 腫瘤異質性

生物試樣(例如，血漿/血清)中之單核苷酸突變的定量亦可用於分析腫瘤異質性，即腫瘤內及腫瘤間異質性。腫瘤內異質性與相同腫瘤內之腫瘤細胞之多個純系之存在相關。腫瘤間異質性與相同組織類型、但存於不同位點中(在相同器官中或在不同器官中)之兩種或更多種腫瘤之多個純系之存在相同。在某些類型之腫瘤中，腫瘤異質性之存在係差預測指標(Yoon HH等人 J Clin Oncol 2012；30: 3932-3938；Merlo LMF等人 Cancer Prev Res 2010；3: 1388-1397)。在某些類型之腫瘤中，腫瘤異質性程度愈高，則在靶向治療後腫瘤進展或抗性純系發育之機會愈高。

儘管據信癌症係自一個腫瘤細胞之純系擴增產生，但癌症之生長及演化可導致癌症之不同部分中之新的不同突變累積。舉例而言，在癌症患者發生轉移時，位於初始器官中之腫瘤及轉移性腫瘤可共享

許多突變。然而，兩個位點之癌細胞亦可帶有一組在其他腫瘤位點中不存在之獨特突變。預計由兩個位點共享之突變較彼等僅在一個腫瘤位點中觀察到之突變以較高濃度存在。

A. 實例

分析患有雙側卵巢癌及乳癌之患者的血漿。兩個卵巢腫瘤皆係血清腺癌。左側腫瘤之最長尺寸量測為6 cm且右側腫瘤量測為12 cm。在結腸及網膜處亦存在多個轉移性病灶。使用來自Illumina之合成測序平臺對自白血球萃取之DNA進行測序至平均44倍單倍體基因組覆蓋。對於血漿中之單核苷酸突變，進一步分析僅顯示一個對偶基因(即同型)之核苷酸位置。

自左側及右側腫瘤之四個不同位點萃取DNA且使用Illumina測序平臺對其進行測序。兩個位點(位點A及B)係來自右側腫瘤且其他兩個位點(位點C及D)係來自左側腫瘤。位點A與B間隔約4 cm。位點C與D之間之距離亦係約4 cm。在手術切除卵巢腫瘤之前及之後自患者收集血漿試樣。隨後自患者之血漿萃取DNA。來自位點A、B、C及D之腫瘤以及血漿試樣之測序深度示於表9中。

試樣	原始測序讀段之數目	比對讀段之數目	單倍體基因組覆蓋次數
來自棕黃色塗層之組成DNA	1,091,250,072	876,269,922	43.81
右側卵巢腫瘤(位點A)	1,374,495,256	1,067,277,229	53.36
右側卵巢腫瘤(位點B)	934,518,588	803,007,464	40.15
左側卵巢腫瘤(位點C)	1,313,051,122	1,036,643,946	51.83

左側卵巢腫瘤 (位點D)	1,159,091,833	974,823,207	48.74
手術之前收集之 血漿試樣	988,697,457	741,982,535	37.10
手術之後收集之 血漿試樣	957,295,879	564,623,127	28.23

表9. 來自位點A、B、C及D之腫瘤之測序深度。

在當前實例中，對於定義單一腫瘤相關之單核苷酸突變，在腫瘤組織中對核苷酸位置測序至少20次且在組成DNA中測序30次。在其他實施例中，可使用其他測序深度，例如35、40、45、50、60、70、80、90、100及>100倍。測序成本降低可容許顯著更容易地增加所實施之深度。核苷酸位置在組成DNA中係同型的，而在腫瘤組織中觀察到核苷酸變化。腫瘤組織中之核苷酸變化之發生次數的準則取決於腫瘤組織中之特定核苷酸位置之總測序深度。對於20至30倍核苷酸覆蓋而言，核苷酸變化之發生次數(截止值)係至少5次。對於31至50倍覆蓋而言，核苷酸變化之發生次數係至少6次。對於51至70倍覆蓋而言，所需發生次數係至少7次。該等準則源自對檢測真實突變的靈敏度之預測及使用荀瓦松分佈預計之偽陽性基因座之數目。

圖17係表1700，其顯示不同發生次數要求及測序深度之預測靈敏度。靈敏度可對應於在特定倍深度下使用特定截止值檢測到之真實突變之數目。測序深度愈高，則對於給定截止值更可能檢測到突變，此乃因可獲得更多突變序列讀段。對於更高截止值，較不可能檢測到突變體，此乃因準則更嚴格。

圖18係表1800，其顯示不同截止值及不同測序深度之偽陽性基因座的預測數目。偽陽性之數目隨測序深度增加而增加，此乃因獲得更多序列讀段。然而，對於5或更大之截止值、甚至高達70之測序深

度預測無偽陽性。在其他實施例中，可使用不同發生次數準則以便達成期望靈敏度及特異性。

圖19顯示圖解說明在不同腫瘤位點中檢測到之突變數目的樹狀圖。藉由直接對腫瘤進行測序來測定突變。位點A具有71個對該腫瘤具有特異性之突變，且位點B具有122個位點特異性突變，即使其僅間隔4 cm。在位點A及B二者中觀察到10個突變。位點C具有168個對該腫瘤具有特異性之突變，且位點D具有248個位點特異性突變，即使其僅間隔4 cm。在位點C及D二者中觀察到12個突變。在不同腫瘤位點之突變概況中存在顯著異質性。舉例而言，僅在位點D腫瘤中檢測到248個突變，但在其他三個腫瘤位點中未檢測到該等突變。在所有位點中觀察到總共2,129個突變。因此，在不同腫瘤之間共享許多突變。因此，存在7個SNV組。該四個區域在拷貝數畸變方面無可觀察到之差異。

圖20係表2000，其顯示治療前及治療後血漿試樣中帶有腫瘤源突變之片段的數目。亦顯示帶有各別突變之腫瘤源DNA的推斷分數濃度。突變之類別係指檢測到突變之腫瘤位點。舉例而言，A類突變係指僅存於位點A中之突變，而ABCD類突變係指存於所有四個腫瘤位點中之突變。

對於存於所有四個腫瘤位點中之2,129個突變，2,105個(98.9%)可在至少一個血漿DNA片段中檢測到。另一方面，對於僅存於四個腫瘤位點中之一者中之609個突變，僅77個(12.6%)可在至少一個血漿DNA片段中檢測到。因此，血漿中單核苷酸突變之定量可用於反映腫瘤組織中該等突變之相對豐度。此資訊可用於癌症異質性之研究。在此實施例中，在測序數據中觀察到一次潛在突變時，稱為潛在突變。

利用每一SNV組測定循環腫瘤DNA之分數濃度。如藉由所有4個區域(即，ABCD組)共享之SNV所測定，手術之前及之後血漿中之腫

瘤DNA之分數濃度分別係46%及0.18%。上述該等百分比與彼等於GAAL分析中獲得者(即46%及0.66%)密切相關。由所有4個區域(即ABCD組)共享之突變在血漿中之腫瘤源DNA之分數濃度中貢獻最高。

自AB及CD組利用SNV測定之術前血漿中之腫瘤源DNA的分數濃度分別係9.5%及1.1%。該等濃度與右側及左側卵巢腫瘤之相對大小一致。利用區域獨特SNV (即，彼等於A、B、C及D組中者)測定之腫瘤源DNA之分數濃度通常較低。該等數據表明，對於癌症患者中之總腫瘤負載之精確量測，與靶向特異性腫瘤相關之突變之更傳統方法相比，使用全基因組鳥槍方法可提供更具代表性圖片。對於後一方法，若僅腫瘤細胞之亞組具有靶向突變，則可丟失關於由不具有靶向突變之腫瘤細胞引起之危急復發或疾病進展的重要資訊，或可丟失治療抗性純系之出現。

圖21係圖表2100，其顯示在單一腫瘤位點中檢測到之突變及在所有四個腫瘤位點中檢測到之突變在血漿中之發生次數的分佈。條形圖2100顯示兩類突變之數據：(1) 僅在一個位點中檢測到之突變及(2) 在所有四個腫瘤位點中檢測到之突變。水平軸係在血漿中檢測到突變之次數。垂直軸顯示對應於水平軸上之特定值之突變的百分比。舉例而言，約88%類型(1)突變在血漿中僅出現一次。如可見，在一個位點中出現之突變大部分檢測到一次，且不超過四次。與存於所有四個腫瘤位點中之突變相比，在血漿中檢測到單一腫瘤位點中存在之突變之頻率遠較低。

此技術之一個應用可為容許臨床醫師估計帶有不同種類突變之腫瘤細胞的負載。一定比例之該等突變可潛在地利用靶向試劑治療。可預計靶向由較高比例之腫瘤細胞帶有之突變之試劑具有更顯著治療效應。

圖22係圖表2200，其顯示來自異質腫瘤之突變在血漿中之發生

次數的預測分佈。腫瘤含有兩組突變。基於兩個位點代表每一卵巢腫瘤之近似情況，一組突變存於所有腫瘤細胞中且另一組突變僅存於 $\frac{1}{4}$ 腫瘤細胞中。血漿中之腫瘤源DNA之總分數濃度假定為40%。假定對血漿試樣測序至每核苷酸位置50次之平均深度。根據血漿中之發生次數之此預測分佈，存於所有腫瘤組織中之突變可因其於血漿中發生區分於僅存於 $\frac{1}{4}$ 腫瘤細胞中之突變。舉例而言，6次發生次數可用作截止值。對於存於所有腫瘤細胞中之突變，92.3%突變可存於血漿中至少6次。與之相比，對於存於 $\frac{1}{4}$ 腫瘤細胞中之突變，僅12.4%突變可存於血漿中至少6次。

圖23係證實16名健康對照個體之實施例之特異性的表2300。對其血漿DNA試樣測序至30倍之中值覆蓋。在該等健康個體之血漿試樣中實施存於上述卵巢癌患者之血漿中之突變的檢測。在健康對照個體之血漿之測序數據中極不經常檢測到存於卵巢症患者之腫瘤中之突變且各類突變皆不具有>1%之表觀分數濃度。該等結果顯示此檢測方法具有高度特異性。

B. 方法

圖24係根據本發明實施例分析個體之一或多種腫瘤之異質性的方法2400之流程圖。可如本文所述實施方法2400之某些步驟。

於框2410處，獲得個體之組成基因組。於框2420處，接收個體生物試樣中複數個DNA片段中之每一者之一或多個序列標籤，其中生物試樣包括無細胞DNA。於框2430處，測定序列標籤之基因組位置。於框2440處，比較序列標籤與組成基因組以測定第一基因座之第一數目。於每一第一基因座處，相對於組成基因組具有序列變體之序列標籤之數目高於截止值，其中截止值大於1。

於框2450處，基於第一基因組位置之組之各別第一數目計算一或多種腫瘤之異質性量度。在一個態樣中，量度可提供代表由腫瘤共

享之突變之數目相對於不由腫瘤共享之突變之數目的值。此處，各種腫瘤可以單一物件形式存在，物件內之腫瘤不同，其可代表通常稱作腫瘤內異質性者。量度亦可與和許多或大部分腫瘤中之突變相比一個或幾個腫瘤中是否存在一些突變有關。可計算異質性之一個以上量度。

於框2460處，可比較異質性量度與臨限值以測定異質性程度之分類。可以各種方式使用一或多個量度。舉例而言，一或多個異質性量度可用於預測腫瘤進展之機會。在一些腫瘤中，異質性愈大，則進展之機會愈高且在治療(例如靶向治療)後抗性純系之出現機會愈高。

C. 腫瘤異質性量度

異質性量度之一個實例係血漿中之不同組突變之「濃度帶」的數目。舉例而言，若患者內存在兩個顯著腫瘤純系，且若該等純系以不同濃度存在，則可預計在血漿中觀察到具有不同濃度之兩個不同突變。可藉由測定不同突變組之分數濃度計算該等不同值，其中每一組對應於腫瘤中之一者。

該等濃度中之每一者皆可稱作「濃度帶」或「濃度種類」。若患者具有更多純系，則將觀察到更多濃度帶/種類。因此，帶愈多，則異質性愈大。可藉由針對各種突變之繪製分數濃度之曲線觀察濃度帶之數目。可針對各種濃度製得直方圖，其中不同峰對應於不同腫瘤(或一種腫瘤之不同純系)。大峰將可能針對由所有或一些腫瘤(或腫瘤之純系)共享之突變。可分析該等峰以測定組合哪些較小峰以測定較大峰。可使用(例如)類似於圖10B及11之擬合程序之擬合程序。

在一個實作中，直方圖係Y軸係基因座之量(例如，數目或比例)且x軸係分數濃度之曲線。由所有或一些腫瘤共享之突變可產生較高分數濃度。峰大小可代表產生特定分數濃度之基因座的量。低及高濃度下之峰之相對大小可反映腫瘤(或腫瘤之純系)之異質性程度。高濃

度下之較大峰反映大部分突變係由大部分或所有腫瘤(或腫瘤之純系)共享且指示較低腫瘤異質性程度。若低濃度下之峰較大，則大部分突變係由幾個腫瘤(或幾個腫瘤之純系)共享。此可指示較高腫瘤異質性程度。

所存在峰愈多，則位點特異性突變愈多。每一峰可對應於不同組之突變，其中該組突變係來自腫瘤(例如，僅一個或兩個腫瘤-如上文所闡釋)之亞組。對於圖19之實例，可能存在總共7個峰，即可能具有最小濃度(端視腫瘤之相對大小而定)之4個唯一位點峰、針對AB位點及CD位點之兩個峰及由所有位點共享之突變之峰。

峰之位置亦可提供腫瘤之相對大小。較大濃度可與較大腫瘤有關，此乃因較大腫瘤可釋放更多腫瘤DNA至試樣中(例如至血漿中)。因此，可估計帶有不同種類之突變的腫瘤細胞之負載。

異質性量度之另一實例係與具有相對較多變體讀段(例如，9-13個)之突變讀段之比例相比，具有相對較少變體讀段(例如，4、5或6個)之突變位點的比例。再次參見圖22，可看到位點特異性突變具有較少變體讀段(其亦產生較小分數濃度)。共享突變具有更多變體讀段(其亦產生較大分數濃度)。6 (較小計數)下之第一比例除以10 (較大計數)下之第二比例的比率傳送異質性量度。若該比率較小，則存在較少具有位點特異性之突變，且因此異質性程度較低。若該比率較大(或至少大於自己知樣品校正之值)，則異質性程度較大。

D. 測定臨限值

可自腫瘤經活檢(例如，如上文所述)之個體測定臨限值以直接測定異質性程度。該等級可以各種方式定義，例如位點特異性突變對共享突變之比率。隨後可分析生物試樣(例如，血漿試樣)以測定異質性量度，其中生物試樣之異質性量度可與藉由直接分析腫瘤之細胞測定之異質性程度相關。

該程序可相對於異質性程度校正臨限值。若測試異質性量度在兩個臨限值之間，則異質性程度可估計為在對應於臨限值之等級之間。

在一個實施例中，可在自活檢測定之異質性程度與自血漿試樣(或另一試樣)測定之相應異質性量度之間計算校正曲線。在該實施例中，異質性程度係數字，其中該等數字值可對應於不同分類。數字值之不同範圍可對應於不同診斷，例如，不同癌症階段。

E. 使用基因組代表之分數濃度之方法

亦可使用(例如)如使用方法1200之實施例測定之分數濃度分析腫瘤異質性。呈現一個拷貝喪失之基因組區域可能來自不同腫瘤。因此，端視僅一個腫瘤或多個腫瘤中是否存在擴增(或缺失，針對1個拷貝喪失)而定，針對不同基因組區域測定之分數濃度可不同。因此，相同異質性量度可用於經由方法1200之實施例測定之分數濃度。

舉例而言，一個基因組區域可鑑別為對應於1個拷貝喪失，且分數濃度可僅自該基因組區域處之各別密度測定(各別密度可用作分數濃度)。可藉由對具有各種密度之區域之數目進行計數自各種各別密度測定直方圖。若僅一個腫瘤或一個腫瘤純系或一個腫瘤沈積之特定區域增加，則該區域之密度可小於多個腫瘤或多個腫瘤純系或多個腫瘤沈積增加之區域的密度(即，共享區域中之腫瘤DNA之分數濃度可大於位點特異性區域中之腫瘤DNA之分數濃度)。因此，上述異質性量度可適於使用各個區域中之拷貝數增加或喪失鑑別之峰，與不同位點之分數濃度顯示分數濃度之分佈一樣。

在一個實作中，若各別密度用於直方圖，則可分開增加與喪失。可藉由僅針對增加產生直方圖來單獨分析顯示增加之區域，且可僅針對喪失產生單獨直方圖。若使用分數濃度，則可一起分析喪失及增加之峰。舉例而言，分數濃度使用與參照密度之差(例如，以絕對

值形式)，且因此，增加及喪失之分數濃度可促成相同峰。

XIII. 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。此等子系統之實例示於圖25中之電腦裝置2500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。

示於圖25中之子系統經由系統匯流排2575互連。顯示其他子系統，例如列印機2574、鍵盤2578、固定磁碟2579、監視器2576（其耦聯至顯示介面卡2582）及其他。可藉由任一數目之業內已知構件（例如序列埠2577）將耦聯至I/O控制器2571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，可使用序列埠2577或外部介面2581（例如乙太網路(Ethernet)、Wi-Fi等）將電腦系統2500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2575之互連使中央處理器2573與各子系統通信並控制來自系統記憶體2572或固定磁碟2579之指令之執行、以及子系統間之資訊交換。系統記憶體2572及/或固定磁碟2579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2581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（例如應用專用積體電路或現場可程式化閘陣列）及/或以模組或整合方式使用電腦軟體及通用可程式化處理器以控制邏輯形式來實作本發明之任一實施例。如本文所用，處理器包括

同一積體晶片上之多核心處理器或單一電路板上或網路化之多個處理單元。基於本文所提供之揭示內容及教示內容，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)。任一此電腦可讀取媒體皆可駐留於單一電腦程式產品(例如硬碟、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(包括一或多個處理器)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」。

本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。

【符號說明】

2571	I/O控制器
2572	系統記憶體
2573	中央處理器
2574	列印機
2575	系統匯流排
2576	監視器
2577	序列埠
2578	鍵盤
2579	固定磁碟

2581	外部介面
2582	顯示介面卡

I636255

發明摘要

※ 申請案號：102122036

※ 申請日：102/06/20

※IPC 分類：G01N 33/50 (2006.01)
C40B 20/04 (2006.01)
G06F 17/18 (2006.01)
G06F 19/18 (2011.01)

【發明名稱】

癌症檢測之血漿DNA突變分析

MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER
DETECTION

【中文】

可比較經歷癌症篩選或監測之個體之生物試樣(例如血漿或血清)中與同一個體之組成DNA中的體細胞突變頻率。可自該等頻率衍生一參數且用於測定癌症等級之分類。可由要求任何變體基因座具有至少一個指定數目之變體序列讀段(reads)(標籤)來濾出偽陽性，藉此提供更正確參數。可分析不同變體基因座之相對頻率以測定患者中腫瘤之異質性程度。

【英文】

A frequency of somatic mutations in a biological sample (e.g., plasma or serum) of a subject undergoing screening or monitoring for cancer, can be compared with that in the constitutional DNA of the same subject. A parameter can be derived from these frequencies and used to determine a classification of a level of cancer. False positives can be filtered out by requiring any variant locus to have at least a specified number of variant sequence reads (tags), thereby providing a more accurate parameter. The relative frequencies for different variant loci can be analyzed to determine a level of heterogeneity of tumors in a patient.

【代表圖】

【本案指定代表圖】：第（1）圖。

【本代表圖之符號簡單說明】：

無

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

無

申請專利範圍

1. 一種檢測個體之癌症或癌前變化的方法，該方法包含：
 - 獲得該個體之組成基因組；
 - 接收該個體之生物試樣中多個DNA片段中每一者之一或多個序列標籤，該生物試樣包括無細胞DNA；
 - 測定該等序列標籤之基因組位置；
 - 比較該等序列標籤與該組成基因組以測定第一基因座之第一數目，其中：
 - 於各該等第一基因座，相對於該組成基因組具有序列變體之該等序列標籤的數目高於一截止值(cutoff value)，該截止值大於1；
 - 基於在該等第一基因座具有序列變體之序列標籤的計數測定一參數；及
 - 比較該參數與一臨限值(threshold value)以測定該個體之癌症等級之分類。
2. 如請求項1之方法，其中該臨限值係自一或多個其他個體之一或多個試樣測定。
3. 如請求項1之方法，其中一個基因座之截止值取決於具有一個基因組位置在該基因座之序列標籤的總數目。
4. 如請求項1之方法，其中對於該等第一基因座中之至少二者使用不同截止值。
5. 如請求項4之方法，其進一步包含：
 - 動態測定該等第一基因座中之一者之第一截止值，該等第一基因座之一個位於第一區域內。
6. 如請求項5之方法，其中該第一截止值係基於該等第一基因座中

之一者之測序深度測定。

7. 如請求項5之方法，其中該第一截止值係基於偽陽性率測定，該偽陽性率取決於測序誤差率、該第一區域之測序深度及該第一區域中核苷酸位置之數目。
8. 如請求項7之方法，其中該第一截止值係基於該第一區域中真陽性之數目測定。
9. 如請求項8之方法，其進一步包含：

基於該生物試樣中腫瘤源DNA之該第一區域之該測序深度D及分數濃度f計算該第一截止值之真陽性之數目。
10. 如請求項9之方法，其中使用葛瓦松分佈(Poisson distribution)機率根據下式計算該真陽性數目： $P_b = 1 - \sum_{i=0}^{r-1} \text{Poisson}(i, M_p)$ ，其中Pb係檢測到真陽性之機率，r係該第一截止值， $M_p = D \times f / 2$ 。
11. 如請求項5之方法，其中該第一截止值係使用以下準則中任一者測定：

若測序深度小於50，則該第一截止值係5，
若該測序深度係50至110，則該第一截止值係6，
若該測序深度係111至200，則該第一截止值係7，
若該測序深度係201至310，則該第一截止值係8，
若該測序深度係311至450，則該第一截止值係9，
若該測序深度係451至620，則該第一截止值係10，及
若該測序深度係621至800，則該第一截止值係11。
12. 如請求項1之方法，其中該參數係第一基因座之該第一數目之加權和(weighted sum)，其中各該等第一基因座之貢獻係基於分配給各別第一基因座之重要值進行加權。
13. 如請求項1之方法，其中該參數包括指示在該第一數目之第一基因座之序列變體的該等序列標籤之和。

14. 如請求項13之方法，其中該和係加權和，且其中該等第一基因座中之一者具有第一加權不同於該等第一基因座中之第二者之第二加權。
15. 如請求項14之方法，其中該第一加權大於該第二加權，且其中該等第一基因座中之該一者與癌症相關，該等第一基因座中之該第二者與癌症無關。
16. 如請求項1之方法，其中該參數係第一基因座之該第一數目。
17. 如請求項1之方法，其中測定序列標籤之基因組位置包括：

將至少一部分該等序列標籤與參照基因組進行比對，其中該序列標籤之比對容許該序列標籤與該參照基因組之間有一或多個失配。
18. 如請求項17之方法，其中比較該等序列標籤與該組成基因組包括：

比較該組成基因組與該參照基因組以測定相對於該參照基因組具有變體之第二基因座之第二數目；

基於該比對，測定第三基因座之第三數目，其中：

在各該等第三基因座，相對於該參照基因組具有序列變體之該等序列標籤之數目高於一截止值；及

取該第三數目與該第二數目之差以獲得第一基因座之該第一數目。
19. 如請求項18之方法，其中取該第三數目與該第二數目之差來鑑別該等第一基因座。
20. 如請求項19之方法，其中測定該參數包括：

對於該第一數目之第一基因座之各基因座：

計數與該基因座比對且在該基因座具有序列變體之序列標籤；及

基於該等各別計數測定該參數。

21. 如請求項1之方法，其中該組成基因組係源自該個體的含有50%以上組成DNA之組成試樣。

22. 如請求項1之方法，其中測定序列標籤之基因組位置包括：

將至少一部分該等序列標籤與該組成基因組進行比對，其中該序列標籤之比對容許該序列標籤與該組成基因組之間有一或多個失配。

23. 如請求項22之方法，其中比較該等序列標籤與該組成基因組包括：

基於該比對，鑑別相對於該個體之組成基因組於基因組位置具有序列變體的序列標籤；

對於呈現序列變體之各基因組位置：

計數與該基因組位置比對且於該基因組位置具有序列變體之序列標籤的各別數目；及

基於該等各別數目測定一參數。

24. 如請求項23之方法，其中基於該等各別數目測定該參數包括：

總計該等各別數目以獲得第一和；及

使用該第一和以測定該參數。

25. 如請求項24之方法，其中使用該第一和測定該參數包括：

自該第一和減去該等呈現序列變體之基因組位置之數目。

26. 如請求項24之方法，其中使用該第一和測定該參數包括：

基於比對序列標籤之量正規化該第一和。

27. 如請求項1之方法，其進一步包含：

獲得該個體之含有90%以上組成DNA之組成試樣；

對該組成試樣中之DNA片段實施隨機測序，以獲得該組成試樣中多個DNA片段中之每一者之一或多個第二序列標籤；

將至少一部分該等第二序列標籤與參照基因組進行比對，其中該第二序列標籤之比對容許該第二序列標籤與該組成基因組之間在M個或更少基因組位置有失配，其中M係等於或大於1之整數；及

基於該等第二序列標籤及該比對構築該組成基因組。

28. 如請求項27之方法，其中該組成試樣係該生物試樣，且其中構築該組成基因組包括：

測定共有序列，其包括測定具有兩個對偶基因之同型基因座或異型基因座；及

使用該共有序列於該組成基因組。

29. 如請求項1之方法，其進一步包含：

接收該個體之生物試樣；及

對該生物試樣中之DNA片段實施隨機測序，以產生該生物試樣中多個DNA片段中之每一者之一或多個序列標籤。

30. 如請求項1之方法，其中該一或多個序列標籤係自該生物試樣中之DNA片段之隨機測序產生。

31. 如請求項1之方法，其中該參數係腫瘤來源DNA之分數濃度。

32. 一種分析個體之一或多種腫瘤之異質性的方法，該方法包含：

獲得該個體之組成基因組；

接收該個體之生物試樣中多個DNA片段中之每一者之一或多個序列標籤，該生物試樣包括無細胞DNA；

測定該等序列標籤之基因組位置；

比較該等序列標籤與該組成基因組以測定第一基因座之數目，其中：

於各該等第一基因座，相對於該組成基因組具有序列變體之該等序列標籤的第一數目高於一截止值，該截止值大於1；

及

基於該等第一基因座之序列標籤之各別第一數目計算該一或多種腫瘤之異質性量度。

33. 如請求項32之方法，其進一步包含：

比較該異質性量度與一或多個臨限值以測定異質性程度之分類。

34. 如請求項33之方法，其中該一或多個臨限值係自一或多個其他個體測定，該等個體之腫瘤已經活檢且經分析以測定該等活檢腫瘤中之突變以測定異質性程度，且其中使用一或多個其他個體之包括無細胞DNA之生物試樣的異質性量度來測定一臨限值。

35. 如請求項34之方法，其中比較該異質性量度與一或多個臨限值包括：

將該異質性量度輸入校正函數中，該校正函數基於該異質性量度輸出異質性程度。

36. 如請求項32之方法，其中該異質性量度包括第一基因座之總數目，於該等第一基因座一個以上DNA片段與該組成DNA進行比對，同時在該基因座具有序列變體。

37. 如請求項32之方法，其中計算多個異質性量度，其中計算該等異質性量度包括：

對於各第一基因座，計算具有序列變異之序列標籤的比例；

針對該等比例值產生第一基因座數目之直方圖；及

鑑別該直方圖中之多個峰。

38. 如請求項37之方法，其中多個異質性量度之一個對應於所鑑別峰之數目。

39. 如請求項37之方法，其中多個異質性量度之一個包括兩個峰之高度比。

40. 如請求項37之方法，其中該等比例各代表於特定第一基因座所測量之腫瘤DNA之分數濃度。
41. 如請求項32之方法，其中該異質性量度對應於具有第一指定量之具有序列變異之序列標籤的第一基因座之第一比例與具有第二指定量之具有序列變異之序列標籤的第一基因座之第二比例之比率。
42. 如請求項41之方法，其中該第一指定量小於該第二指定量。
43. 如請求項42之方法，其中該第一指定量係第一範圍且該第二指定量係第二範圍，該第一範圍低於該第二範圍。
44. 如請求項41之方法，其中該第一指定量及該第二指定量對應於具有序列變異之序列標籤之分數濃度或絕對數目。
45. 如請求項32之方法，其中該異質性量度係自對應於該等各別第一數目中之每一者之第一基因座之數目的直方圖測定。
46. 如請求項45之方法，其中計算多個異質性量度，其中該組第一基因座包括第一亞組及第二亞組，其中該等多個量度包括對應於該第一亞組之該等各別第一數目的第一直方圖及對應於該第二亞組之該等各別第二數目的第二直方圖。
47. 如請求項32之方法，其中該異質性量度包括具有高於指定值之相應第一數目之第一基因座的比例。
48. 一種測定包括無細胞DNA之生物試樣中腫瘤DNA之分數濃度的方法，該方法包含：

接收該生物試樣中多個DNA片段中之每一者之一或多個序列標籤；

測定該等序列標籤之基因組位置；

對於多個基因組區域中之每一者：

自具有基因組位置在該基因組區域內之序列標籤測定該基

因組區域內之DNA片段的各別量；

正規化該各別量以獲得各別密度；及

比較該各別密度與參照密度以鑑別該基因組區域是否呈現1個拷貝喪失或1個拷貝增加；

自一或多個基因組區域之一或多個鑑別為呈現1個拷貝喪失之各別密度或自一或多個基因組區域之一或多個鑑別為呈現1個拷貝增加之各別密度計算第一密度；及

藉由以下計算該生物試樣中腫瘤DNA之該分數濃度：

比較該第一密度與另一密度以獲得差，其中用該參照密度正規化該差。

49. 如請求項48之方法，其進一步包含：

產生具有各種各別密度之基因組區域之數目的直方圖；

鑑別該直方圖中之多個峰；及

自一或多個第一峰之基因組區域之數目對一或多個第二峰之基因組區域之數目的比率計算異質性量度。

50. 如請求項49之方法，其中在產生該直方圖時該直方圖使用各別針對各基因組區域所測定之分數濃度值。

51. 如請求項49之方法，其中該等第一峰具有第一指定量之各別密度，且該等第二峰具有第二指定量之各別密度。

52. 如請求項51之方法，其中該第一指定量係第一範圍且該第二指定量係第二範圍，且其中該第一範圍小於該第二範圍。

53. 如請求項48之方法，其中比較該各別密度與該參照密度以鑑別該基因組區域是否呈現1個拷貝喪失或1個拷貝增加包括：

計算該各別密度與該參照密度之間之差；及

比較該差與一截止值。

54. 如請求項48之方法，其中用該參照密度正規化該差係由：

該差除以該參照密度。

55. 如請求項48之方法，其中該另一密度係該參照密度。
56. 如請求項55之方法，其中計算該分數濃度進一步包括該差乘以2。
57. 如請求項48之方法，其中該第一密度係使用鑑別呈現1個拷貝增加之各別密度計算，且其中該另一密度係自鑑別呈現1個拷貝喪失之各別密度計算的第二密度。
58. 如請求項57之方法，其中用該參照密度正規化該差係由：
 - 計算該第一密度與該參照密度之第一比率；
 - 計算該第二密度與該參照密度之第二比率，
 - 其中該差係介於該第一比率與該第二比率之間。
59. 如請求項57之方法，其中比較該各別密度與該參照密度以鑑別該基因組區域是否呈現1個拷貝喪失或1個拷貝增加包括：
 - 將峰擬合至該等各別密度之直方圖之分佈曲線，其中該第一密度對應於第一峰且該第二密度對應於第二峰。
60. 如請求項48之方法，其中經測定呈現該各別密度相對於該參照密度在統計學上顯著增加的所有基因組區域鑑別為呈現1個拷貝增加。
61. 如請求項48之方法，其中區域的參照密度係自不同試樣測定，或係使用自其他區域之各別密度測定。
62. 如請求項48之方法，其中測定序列標籤之基因組位置包括：將至少一部分該等序列標籤與參照基因組進行比對。
63. 如請求項62之方法，其中序列標籤之比對容許該序列標籤與該參照基因組之間有一或多個失配(mismatch)。
64. 如請求項62之方法，其中正規化該各別量以獲得該等多個基因組區域之第一基因組區域之該各別密度包括使用相同總數目之

比對標籤以測定該各別密度及該參照密度。

65. 如請求項62之方法，其中正規化該各別量以獲得該等多個基因組區域之第一基因組區域之各別密度包括該各別量除以比對標籤之總數目。
66. 如請求項48之方法，其中該多個基因組區域各具有相同長度。
67. 如請求項48之方法，其中該等基因組區域不重疊。
68. 如請求項1之方法，其中該生物試樣係尿、胸水、腹水、腹膜液、唾液、腦脊髓液或糞便試樣。
69. 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用於控制電腦系統以實施如請求項1-28及30-67中任一項之多個指令。
70. 一種電腦系統，其包含一或多個處理器，該處理器經程式化以實施如請求項1-28及30-67中任一項之多個指令。