



(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 I829031 B

(45) 公告日：中華民國 113 (2024) 年 01 月 11 日

(21) 申請案號：110140554

(22) 申請日：中華民國 105 (2016) 年 01 月 22 日

(51) Int. Cl. : C12Q1/6883 (2018.01)

G16B30/00 (2019.01)

G16B40/00 (2019.01)

(30) 優先權：2015/01/23 美國

62/107,227

(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)；余 燦妍 YU, CHEUK YIN JANDY (HK)

(74) 代理人：陳長文

(56) 參考文獻：

US 2013237431A1

WO 2014190286A2

審查人員：張維纓

申請專利範圍項數：19 項 圖式數：11 共 59 頁

(54) 名稱

基於大小及數目之組合分析母體血漿以偵測胎兒亞染色體畸變

(57) 摘要

可藉由分析胎兒及母體 DNA 之樣本識別胎兒基因組中之畸變。使用基於數目及基於大小之方法判定亞染色體區域中是否存在畸變(擴增或缺失)之分類。可組合使用數目分類及大小分類以判定是僅胎兒或僅母親抑或兩者在該亞染色體區域中具有該畸變，藉此避免當母親具有該畸變而胎兒並不具有該畸變時之假陽性。

An aberration in a fetal genome can be identified by analyzing a sample of fetal and maternal DNA. Classifications of whether an aberration (amplification or deletion) exists in a subchromosomal region are determined using count-based and size-based methods. The count classification and the size classification can be used in combination to determine whether only the fetus or only the mother, or both, have the aberration in the subchromosomal region, thereby avoiding false positives when the mother has the aberration and the fetus does not.

指定代表圖：

符號簡單說明：

100:表

110:行

120:行

130:行

140:行

| 母親 | 胎兒 | 基於數目  | 基於大小 |
|----|----|-------|------|
| 重複 | 重複 | 過度表示  | 正常   |
| 缺失 | 缺失 | 不充分表示 | 正常   |
| 正常 | 重複 | 過度表示  | 更短   |
| 正常 | 缺失 | 不充分表示 | 更長   |
| 重複 | 正常 | 過度表示  | 更長   |
| 缺失 | 正常 | 不充分表示 | 更短   |

【圖1】



I829031

【發明摘要】

【中文發明名稱】

基於大小及數目之組合分析母體血漿以偵測胎兒亞染色體畸變

【英文發明名稱】

COMBINED SIZE- AND COUNT-BASED ANALYSIS OF  
MATERNAL PLASMA FOR DETECTION OF FETAL  
SUBCHROMOSOMAL ABERRATIONS

【中文】

可藉由分析胎兒及母體DNA之樣本識別胎兒基因組中之畸變。使用基於數目及基於大小之方法判定亞染色體區域中是否存在畸變(擴增或缺失)之分類。可組合使用數目分類及大小分類以判定是僅胎兒或僅母親抑或兩者在該亞染色體區域中具有該畸變，藉此避免當母親具有該畸變而胎兒並不具有該畸變時之假陽性。

【英文】

An aberration in a fetal genome can be identified by analyzing a sample of fetal and maternal DNA. Classifications of whether an aberration (amplification or deletion) exists in a subchromosomal region are determined using count-based and size-based methods. The count classification and the size classification can be used in combination to determine whether only the fetus or only the mother, or both, have the aberration in the subchromosomal region, thereby avoiding false positives when the mother has the aberration and the fetus does not.

【指定代表圖】

圖1

【代表圖之符號簡單說明】

100:表

110:行

120:行

130:行

140:行

【特徵化學式】

無

## 【發明說明書】

### 【中文發明名稱】

基於大小及數目之組合分析母體血漿以偵測胎兒亞染色體畸變

### 【英文發明名稱】

COMBINED SIZE- AND COUNT-BASED ANALYSIS OF  
MATERNAL PLASMA FOR DETECTION OF FETAL  
SUBCHROMOSOMAL ABERRATIONS

### 【技術領域】

### 【先前技術】

母體血漿中之無細胞DNA包含胎兒及母體DNA之混合物。以非侵入性方法產前量測母體血漿結果可用於藉由計數來自亞染色體區域之DNA片段來偵測亞染色體複本數畸變(copy number aberration; CNA)。然而，該計數並未區分胎兒DNA與母體DNA。因此，藉由計數DNA片段偵測之畸變可能源自胎兒或母親。因此，當母親自身為CNA之攜帶者時，吾人無法辨別其胎兒是否已遺傳CNA。另外，當亞染色體區域分析的越多時，假陽性結果將變得更普遍。

實施例可解決此等及其他問題。

### 【發明內容】

實施例使用一種將包括母體及胎兒DNA之母體樣本的基於數目及基於大小之分析相組合以偵測胎兒亞染色體複本數畸變(CNA)之策略。可使用基於數目之分析偵測區域中之CNA。亦可使用DNA分子之基於大小之分析來分析經判定具有CNA之區域，其中基於大小之分析可用於區分來源於胎兒或母親，或來自兩者之畸變。

其他實施例係針對與本文中所描述之方法相關聯的系統及電腦可讀媒介。

可參考以下詳細描述及隨附圖式來獲得對本發明實施例之性質及優勢的較佳理解。

### 【圖式簡單說明】

圖1展示根據本發明之實施例的針對基於數目及基於大小之結果的組合之六種情境的表100。

圖2展示根據本發明之實施例的針對胎兒中之2號染色體上之區域中具有複本數增益之情況的基於數目及大小之組合分析的實例過程流程。

圖3為根據本發明之實施例的藉由分析來自懷有胎兒之雌性個體的生物樣本識別胎兒之胎兒基因組中之亞染色體畸變的方法300之流程圖。

圖4展示根據本發明之實施例的針對基於數目及基於大小之結果的組合以及 $F_{CNA}$ 值之六種情境的表400。

圖5為展示用於說明本發明之實施例之準確性的六種母體血漿DNA樣本之基於數目之分值及基於大小之分值的表500。

圖6為展示關於CNA源自胎兒或母親，或兩者之六個案例之資訊的表600。

圖7展示根據本發明之實施例針對表600中之六個案例之基於數目及基於大小之組合分析母體血漿DNA。

圖8為展示根據本發明之實施例之在每一案例中未偵測到CNA之測試區域的基於數目及基於大小之z分值的表800。

圖9為展示根據本發明之實施例之由藉由分段處理判定之100 kb區間組成之擴增區域的曲線圖900。

圖10為展示根據本發明之實施例之由藉由分段處理判定之100 kb區間組成之缺失區域的曲線圖1000。

圖11展示可與根據本發明之實施例的系統及方法一起使用之實例電腦系統10的方塊圖。

### 術語

如本文所使用，術語「生物樣本」係指獲自個體(例如，人類，諸如懷孕婦女)且含有所關注之一或多個核酸分子的任何樣本。實例包括血漿、唾液、胸膜液、汗液、腹水、膽液、尿液、血清、胰液、糞便及子宮頸抹片樣本。

術語「核酸」或「多核苷酸」係指單鏈或雙鏈形式之脫氧核糖核酸(DNA)及其聚合物。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，該等已知類似物具有與參考核酸類似的結合性質且係以與天然存在之核苷酸類似的方式代謝。除非另外指示，否則特定核酸序列亦隱含地涵蓋其經保守修改的變體(例如，簡併密碼子取代)、對偶基因、直系同源物、單核苷酸多形現象(SNP)，及互補序列以及經明確指示之序列。具體言之，可藉由產生其中一或多個選擇的(或全部)密碼子之第三位置被混合鹼基及/或脫氧肌核苷取代之序列獲得簡併密碼子取代(Batzer MA等人，*Nucleic Acid Res* 1991；**19**:5081；Ohtsuka E等人，*J Biol Chem* 1985；**260**:2605-2608；及Rossolini GM等人，*Mol Cell Probes* 1994；**8**:91-98)。

術語「序列讀段」係指獲自核酸分子之全部或部分之序列，例如，DNA片段。在一項實施例中，片段僅一個末端加以定序，例如，約30個鹼基，接著將經定序讀段與參考基因組對準。替代地，片段之兩個末端皆

可經定序以產生兩個定序讀段，該兩個定序讀段可在對準中提供更大準確性且亦提供片段之長度。在又另一實施例中，可(例如)藉由連接反應環化線狀DNA片段，且可定序跨越連接反應位點之部分。

術語「分數胎兒DNA濃度」可與「術語胎兒DNA比例」及「胎兒DNA分數」互換使用，且係指存在於源自胎兒之母體血漿或血清樣本中的DNA分子之比例(Lo YMD等人，*Am J Hum Genet* 1998;62:768-775；Lun FMF等人，*Clin Chem* 2008;54:1664-1672)。

術語「大小分佈」通常係關於生物樣本中之DNA片段的大小。大小分佈可為以多種大小提供一定量之DNA片段之分佈的直方圖。各種統計參數(亦被稱為大小參數或僅稱為參數)可用於區分一個大小分佈與另一大小分佈。一個參數為相對於所有DNA片段或相對於另一大小或範圍之DNA片段的特定大小或大小的範圍之DNA片段的百分比。

如本文所使用，術語「參數」意謂表徵定量資料集之數值及/或定量資料集之間的數值關係。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之間的比率(或比率之函數)為參數。

如本文所使用，術語「分類」係指與樣本之特定性質相關聯的任何數字或其他字符(包括字)。舉例而言，「+」符號可表示將樣本分類為具有缺失或擴增(例如，複製)。術語「截止值」及「臨限值」係指使用於運算之預定數字。舉例而言，截止大小可指其上排除片段之大小。臨限值可為高於或低於特定分類適用之值。可在此等上下文中之任一者中使用此等術語中之任一者。

「亞染色體區域」為小於染色體之區域。亞染色體區域之實例為大小為100 kb、200 kb、500 kb、1 Mb、2 Mb、5 Mb或10 Mb之彼等。亞

染色體區域之另一實例為對應於一或多個分帶、或子分帶或染色體之臂中之一者的一者。分帶或子分帶為在細胞遺傳學分析中觀測到之特徵。亞染色體區域可關於參考人類基因組序列被稱為其基因組座標。

### 【實施方式】

相關申請案之交叉參考

本申請案主張2015年1月23日申請之題為「Combined Size- and Count-Based Analysis Of Maternal Plasma For Detection Of Fetal Subchromosomal Aberrations」的美國臨時申請案第62/107,227號之優先權，該申請案之全部內容出於所有目的以引用之方式併入本文中。本申請案亦相關於2008年7月23日由Lo等人申請之題為「Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing」的美國專利公開案2009/0029377；及2010年11月5日由Lo等人申請之題為「Size-Based Genomic Analysis」的美國專利第8,620,593號，該等申請案之揭示內容出於所有目的以全文引用之方式併入。

基於母體血漿DNA之非侵入性產前測試已擴展至包括偵測某些亞染色體複本數畸變(CNA)，亦稱為複本數畸變(CNA)。然而，假陽性結果較為普遍，尤其當分析更多亞染色體區域時。即使具有高偵測率及低假陽性率，由於不足夠高的陽性預測性價值，使用母體血漿中之無細胞DNA針對胎兒亞染色體非整倍體的非侵入性產前測試(NIPT)當前亦並未廣泛地用作篩選測試。

下文之描述表明基於大小之分析可用作驗證由基於數目之分析偵測之CNA的獨立方法。另外，展示基於大小及基於數目分析之組合可判定胎兒是否已自自身為CNA之攜帶者的其母親遺傳CNA。使用基於大小及

基於數目分析之組合的實施例可區別藉由分析母體生物樣本(例如，使用定序及分析定序結果)偵測之畸變的來源(亦即，胎兒、母體或兩者)。此策略改善當前測試之專一性。結果展示，實施例藉由能夠識別CNA之來源提供改善，僅單獨地使用基於數目之技術或基於大小之技術不可能實現該改善。

## I. 引入

母體血漿中之無細胞DNA包含胎兒及母體DNA之混合物。針對胎兒染色體非整倍體之非侵入性產前測試(NIPT)使用之母體血漿中之無細胞DNA的大規模平行定序技術(MPS)已在產前保健中廣泛地採用(1、2)。此等方法係基於計數母體血漿中定位至基因組之不同區域的DNA片段，且因此被稱為「計數方法」(3)。近期研究已表明，此方法可藉由使用更高定序深度及適合之生物資訊分析來偵測胎兒亞染色體異常(4至8)。實際上，多個開始對多種臨床重要且相對常見的亞染色體異常(諸如，迪喬治症候群、貓叫(Cri-du-chat)症候群、普拉德-威利/安格爾曼症候群及1p36缺失症候群)提供NIPT (9)。

計數方法列舉母體樣本中之胎兒及母體DNA分子兩者。其相對於懷有正常胎兒之健康懷孕女性群組之對應值比較懷孕婦女之血漿中之特定基因組區域的相對表示。因此，來自基於數目之方法的異常結果可能起因於一個以上臨床場景，即(i)胎兒、(ii)母親或(iii)兩者中存在複本數畸變(CNA) (8、10)。如本文中所使用，母親可指親生母親或代孕母親。術語懷孕雌性個體亦指代兩者。

因此，當母親攜帶CNA時，吾人可不能辨別胎兒是否已遺傳畸變。實際上，母體複本數變體之存在為干擾NIPT結果的所報導之原因中之一

者(11)。Snyder等人在兩個案例中表明，不一致NIPT結果可歸因於母體複本數變體之存在(11)。在近期研究中，Yin等人報導母體複本數變體存在於在其1,456個樣本定組中具有假陽性NIPT結果的55個樣本中之35個中(63.7%) (12)。基於此發現，Yin等人建議跟蹤測試母體DNA以在胎兒亞染色體畸變具有陽性NIPT結果之情況下排除母體複本數變體。因此，母體複本數變體之存在造成使用胎兒及母體DNA之混合物的胎兒亞染色體CNA偵測中之不準確性。

近來，包括本發明之發明者的群組開發一種方法，該方法利用母體血漿中之胎兒與母體DNA分子之間的大小差異以用於胎兒非整倍體偵測(13)。源自胎兒之DNA分子(亦稱為片段)相較於源自母親之彼等具有更短大小分佈(14、15)。因此，胎兒三染色體中之額外胎兒染色體的存在將減少源自彼染色體之母體血漿中之DNA的大小分佈。此基於大小之方法偵測來自血漿中之非整倍體染色體之短片段的增加比例。此方法已允許對多種類型之胎兒全染色體非整倍體(包括21、18、13號染色體及單染色體X)之偵測具有高準確性(13)。當存在母體複本數變體時，單獨使用此基於大小之方法在使用胎兒及母體DNA之混合物偵測胎兒亞染色體CNA中亦將不準確。

本發明展示基於大小及基於數目分析之組合經展示能夠區別藉由母體血漿DNA定序偵測之畸變的來源(亦即，胎兒、母體或兩者)。若胎兒及母親兩者在特定亞染色體區域中皆具有CNA，則相較於不具有任何CNA之另一亞染色體區域時，彼區域之大小分佈中將不存在淨差。另一方面，若當與特定亞染色體區域中之母體DNA相比較時(諸如，當(i)胎兒具有微重複而母親正常；或(ii)母親具有微缺失而胎兒正常時)存在胎兒DNA之相

對過度表達，則整個大小分佈將縮短。相反地，若當與特定亞染色體區域中之母體DNA相比較時(諸如，當(i)胎兒具有微缺失而母親正常；或(ii)母親具有微重複而胎兒正常時)存在胎兒DNA之不充分表示，則整個大小分佈將延長。以此方式，當與基於數目之方法組合時，基於大小之方法可用於判定亞染色體CNA之來源。

## II. 獨立分析

為了識別亞染色體CNA，可將全部基因組劃分為亞染色體區域(亦稱為區間)。在一些實施例中，區間可小於區域，其中具有CNA之亞染色體區域可包括多個區間。具有畸變之連續區間可定義具有CNA之區域。在其他實施例中，區域可對應於一個區間。如稍後更詳細解釋，可合併區間以識別具有CNA之區域(區段)。

舉例而言，此等區間可具有約100 kb、200 kb、500 kb、1 Mb、2 Mb、5 Mb或10 Mb之大小。亞染色體區域亦可包括分帶、子分帶及臂。在一項實施中，可使用2,687 1 Mb的區間。可排除基因組之某些部分，例如，重複區域。可使用基於數目之方法及基於大小之方法分析此等亞染色體區域以判定區域是包括擴增抑或缺失。畸變可不對應於全部區域，但區域可經測試以識別畸變是否出現在區域中之某處。

作為分析之部分，母體樣本中之無細胞DNA片段經分析以(例如)關於參考基因組判定基因組中之DNA片段的位置。舉例而言，無細胞DNA片段可經定序以獲得序列讀段，且該等序列讀段可定位(對準)至參考基因組。若生物體為人類，則參考基因組將為可能來自特定亞群之參考人類基因組。作為另一實例，可(例如，在PCR或其他擴增之後)利用不同探針分析無細胞DNA片段，其中每一探針對應於不同基因組位置。在一些實施

例中，可藉由獲得對應於無細胞DNA片段之序列讀段或其他實驗資料執行對無細胞DNA片段之分析，且接著使用電腦系統分析該實驗資料。

#### A. 基於數目之分析

為了執行基於數目之分析，實施例可計數來自位於每一區域中之母體樣本的DNA片段的數目。區域中之DNA片段之數目可與一或多個數目臨限值比較以判定該區域是否呈現CNA。可基於健康對照區域中之對應數目判定數目臨限值，以便鑑別CNA及不具有CNA之區域。若數目高於高臨限值，則識別擴增。若數目低於低臨限值，則可識別缺失。熟習此項技術者將瞭解如何判定此等臨限值。

DNA片段之數目經歸一化使得可在不同個體中進行對比，其中可分析不同量之DNA片段。可以許多方式執行歸一化，例如，藉由將區域之各別數目除以一或多個其他區域(潛在地全部基因組)之各別數目的總和。一或多個其他區域之各別數目的此對比亦可藉由始終分析來自不同樣本之相同數目的DNA片段來完成，該對比使得總和始終相同。因此，可直接使用區域之計數數目(例如，由於總和可有效地包括於臨限值中)。因此，仍在此實施例中執行數目之比率。因此，實施例可利用第一區域之第一數目及第二區域之第二數目計算數目參數。

區域中之歸一化數目的DNA片段亦可被稱為基因組表示(GR)。藉由不同區域之第二數目個DNA片段的比率執行歸一化。舉例而言，區域之GR可為位於由分析中所使用之所有DNA片段劃分之區域中的多個DNA片段。區域之GR可對應於任何量，例如，多個DNA片段、與DNA片段重疊之多個鹼基或區域中之其他量值的DNA片段。

在一些實施例中，數目參數可(例如)以以下方式對應於分值。可判定

對照組之測試區域之基因組表示(GR)的均值及標準差(SD)可使用下列等式針對每一樣本之測試區域計算基於數目之z分值(7)：

$$\text{基於數目之z分值} = \frac{GR_{\text{樣本}} - \text{均值 } GR_{\text{對照組}}}{SD GR_{\text{對照組}}}$$

在本文中之實例中，將>3且<-3的基於數目之z分值分別用作分別指示複本數增益及複本數損失之數目臨限值。在此實例中，z分值可對應於相較於3或-3之數目參數。可使用其他數目臨限值(例如，除了3以外之值)。在其他實例中， $GR_{\text{樣本}}$ 為數目參數且可將另一項目移動至等式之左側以用作數目臨限值之部分。可在美國專利公開案2009/0029377中發現基於數目之方法的其他細節。

#### B. 基於大小之分析

如上所述，胎兒DNA片段小於母體DNA片段。此大小差異可用於偵測胎兒中之CNA。若胎兒在第一區域中具有擴增，則區域之DNA片的平均大小將低於不具有擴增之第二區域；第一區域中之額外、更小胎兒DNA將減小平均大小。類似地，對於缺失，區域之較少胎兒片段將使得平均大小大於正常區域之平均大小。大小之實例包括長度或質量。

可使用其他統計值，例如，給定大小之累積頻率或不同大小之DNA片段之量的各種比率。累積頻率可對應於具有給定大小或更小之DNA片段的一部分。統計值提供有關DNA片段之大小之分佈的資訊以與健康對照個體之一或多個大小臨限值相比較。如同數目臨限值，熟習此項技術者將知曉如何判定此臨限值。

因此，為了執行基於大小之分析，實施例可計算位於第一亞染色體區域中之核酸分子之大小的第一統計值，且計算位於參考區域中之核酸分子之大小的參考統計值。可在第一統計值與參考統計值之間判定分離值

(例如，差值或比率)。亦可利用其他值判定分離值。舉例而言，可利用多個區域之統計值判定參考值。分離值可與大小臨限值相比較以獲得大小分類(例如，DNA片段是更短、更長抑或與正常區域相同)。

一些實施例可針對測試區域中之每一者計算參數(分離值)，使用下列等式將該參數定義為測試區域與參考區域之間的短DNA片段之比例的差值(13)：

$$\Delta F = P(\leq 150bp)_{\text{測試}} - P(\leq 150bp)_{\text{參考}}$$

其中  $P(\leq 150bp)_{\text{測試}}$  表示源自大小 $\leq 150$  bp之測試區域之定序片段的比例，且  $P(\leq 150bp)_{\text{參考}}$  表示源自大小 $\leq 150$  bp之參考區域之定序片段的比例。在其他實施例中，可使用其他大小臨限值，例如(但不限於)，100 bp、110 bp、120 bp、130 bp、140 bp、160 bp及166 bp。在其他實施例中，可在鹼基、或核苷酸，或其他單元中表達大小臨限值。在一些實施中，可將參考區域定義為除測試區域外之所有亞染色體區域。在其他實施中，參考區域可為除測試區域外之亞染色體區域的僅一部分。

基於數目之分析中所使用之對照組的相同群組可用於基於大小之分析中。可使用對照組之  $F$  之均值及SD值計算測試區域的基於大小之z分值(13)。

$$\text{基於大小之z分值} = \frac{\Delta F_{\text{樣本}} - \text{均值 } \Delta F_{\text{對照組}}}{SD \Delta F_{\text{對照組}}}$$

在一些實施例中， $>3$ 之基於大小之z分值指示測試區域之短片段的增加比例，而 $<-3$ 之基於大小之z分值指示測試區域之短片段的減少比例。可使用其他大小臨限值。可在美國專利第8,620,593號中發現基於大小之方法的其他細節。

為了判定DNA片段之大小，至少一些實施例可與其中可分析染色體

來源及分子長度之任何單分子分析平台共同作用，該單分子分析平台為（例如）電泳、光學方法（例如，光學映射及其變體，[en.wikipedia.org/wiki/Optical\\_mapping#cite\\_note-Nanocoding-3](http://en.wikipedia.org/wiki/Optical_mapping#cite_note-Nanocoding-3)，及 Jo 等人，Proc Natl Acad Sci USA 2007;104:2673-2678）、基於螢光之方法、基於探針之方法、數位 PCR（基於微流體或基於乳化液，例如 BEAMing (Dressman 等人，Dressman et al. Proc Natl Acad Sci USA 2003;100:8817-8822) 、 RainDance ([www.raindancetech.com/technology/pcr-genomics-research.asp](http://www.raindancetech.com/technology/pcr-genomics-research.asp)))、滾環擴增、質譜法、熔融分析(或熔融曲線分析)、分子篩分等。作為質譜法之實例，更長分子將具有更大質量(大小值之實例)。

在一項實例中，可使用成對末端定序協定隨機定序 DNA 分子。兩個末端處之兩個讀段可定位(對準)至重複序列遮蔽的參考基因組。可利用兩個讀段定位至的基因組位置之間的距離判定 DNA 分子之大小。

### III. 基於數目及大小之方法的組合

基於數目之方法相對於懷有正常胎兒之健康懷孕女性群組比較特定基因組區域之相對表示。因此，來自基於數目之方法的異常結果將告知胎兒或母親，抑或兩者具有複本數畸變。另一方面，取決於 DNA 分子之來源，基於大小之方法係基於母體樣本中之 DNA 分子之大小分佈中的差異。因此，源自胎兒之 DNA 分子相較於源自母親之彼等將具有更短大小分佈。因此，若胎兒及母親在特定亞染色體區域中皆具有複本數畸變，則與不具有複本數畸變之另一亞染色體區域比較時，彼區域中之大小分佈中將不存在淨差。

另一方面，若當與特定亞染色體區域中之母體 DNA 相比較時(諸如，

當(i)胎兒具有微重複(或更大範圍之微擴增)而母親正常；或(ii)母親具有微缺失而胎兒正常時)存在胎兒DNA之過度表示，則大小分佈將縮短。相反地，若當與特定亞染色體區域中之母體DNA比較時(諸如，當(i)胎兒具有微缺失而母親正常；或(ii)母親具有微重複而胎兒正常時)存在胎兒DNA之不充分表示，則大小分佈將延長。

圖1展示根據本發明之實施例的針對基於數目及基於大小之結果的組合之六種情境的表100。在一些實施例中，當統計值大致等於期望值時，正常的基於大小之分類對應於(例如)參考對照組。行110展示如重複(或更大範圍之擴增)、缺失或正常之母親的各種狀態。行120展示如重複(或更大範圍之擴增)、缺失或正常之胎兒的各種狀態。行130展示過度表示(對應於上述實例中之基於數目之正z分值)及不充分表示(對應於上述實例中之基於數目之負z分值)的基於數目之分類。行140展示正常(兩個區域中之相同大小)、更短(對應於上述實例中之基於大小之正z分值)及更長(對應於上述實例中之基於大小之負z分值)的基於大小之分類。

查看表100，具有僅源自胎兒之CNA的情況在與基於數目之z分值相同的方向上將具有基於大小之z分值。舉例而言，基於數目之z分值的正值指示彼區域之DNA片段之過度表示，且基於大小之z分值的正值指示彼區域之DNA片段更短，藉此表明僅胎兒之擴增(例如，重複)。相反地，基於數目之z分值及基於大小之z分值兩者的負值分別指示彼區域之不充分表示及更長DNA片段，藉此表明僅胎兒之缺失。

對於母親攜帶CNA之情況，基於大小之分析將適用於判定胎兒是否已自母親遺傳畸變。胎兒已自母親遺傳畸變之情況將具有正常範圍內之基於大小之z分值，此係因為相較於其他基因組區域，受影響區域之胎兒

DNA及母體DNA之相對比例沒有改變。舉例而言，基於數目之分析的過度表示分類及基於大小之分析的正常分類展示胎兒遺傳擴增。及，基於數目之分析的不充分表示分類及基於大小之分析的正常分類展示胎兒遺傳缺失。

另一方面，CNA僅存在於母親中之情況在與基於數目之 $z$ 分值相對的方向上會具有基於大小之 $z$ 分值。因此，基於數目之正 $z$ 分值(過度表示分類)及基於大小之負 $z$ 分值(更長分類)指示母體具有重複而胎兒是正常狀態。相反地，基於數目之負 $z$ 分值(不充分表示分類)及基於大小之正 $z$ 分值(更短分類)指示母體具有缺失及胎兒是正常狀態。

#### A. 執行組合分析

圖2展示根據本發明之實施例的針對胎兒中之2號染色體上之區域中具有複本數增益之情況的基於數目及大小之組合分析的實例過程流程。對於2號染色體母體細胞展示為正常，且對於區域202胎兒細胞展示為具有重複。

無細胞DNA片段展示為存在於母體血漿205中。源自胎兒之DNA分子(粗紅色片段220)具有比源自母親之彼等(黑色片段)更短之大小分佈。獲取母體血漿之樣本210。如圖所示，執行成對末端定序以獲得序列讀段。(例如)當序列讀段定位至參考基因組時，成對末端定序(包括定序全部DNA片段)可用於判定DNA片段之大小以及其位置。

在基於數目之分析中，在區塊230處，計數對準至2號染色體之各別區間232的DNA片段。區域202之區間經識別為具有比正常區域204之區間更高的量。在此實例中，用於計數之區間小於用於分析 $z$ 分值之區域。在其他實例中，可針對每一區間進行獨立判定(因此，區間將與區域大小相

同)。在一些實施例中，為展示同一畸變需要多個連續區間，例如，如美國專利公開案第2014/0195164號，其以全文引用的方式併入本文中。因此，即使正常區域204中之一個區間具有與區域202中之兩個區間一樣高的數目，該區間仍包括於正常區域204中。

在區塊240處，區域202之基因組表示(GR)經判定為區域202中之序列讀段之計數之數目除以序列讀段總計數。在其他實施例中，分母可為僅某些區間之序列讀段的計數。

在區塊245處，判定基於數目之z分值。可使用來自對照組250之z分值的值(例如)來判定對照組之均值及標準差(SD)。對照值可用於對照個體中之相同測試區域或用於類似大小之其他區域。使用指示大於數目臨限值之正分值的向上箭頭展示基於數目之z分值。

在區塊260處，基於大小之分析可獲得展示過度表示之區域202的識別。基於大小之分析展示區域202及參考區域之大小分佈。如圖所示，區域202之大小分佈小於參考區域之大小分佈。該等大小分佈之間的此關係之判定可使用大小分佈之統計值在以下區塊中得以判定。

在區塊270處，利用  $P(\leq 150bp)_{\text{測試}}$  及  $P(\leq 150bp)_{\text{參考}}$  判定分離值 $\Delta F$ 。其他統計值可用於其他實例。由於區域202具有150個鹼基或更少之DNA片段的較高比例，因此分離值展示為正值。

在區塊280處，判定基於大小之z分值。可使用來自對照組250之z分值的值(例如)來判定對照組之均值及標準差(SD)。對照值可用於對照個體中之相同測試區域或用於類似大小之其他區域。使用指示大於大小臨限值之正分值的向上箭頭展示基於大小之z分值，該基於大小之z分值對應於小於參考區域之區域202的DNA片段。

根據圖1之表100，過度表示之基於數目之分類及較短的基於大小之分類指示僅胎兒具有區域202之擴增。以此方式，實施例可用於判定胎兒、母親抑或兩者具有經識別畸變。

## B. 方法

圖3為根據本發明之實施例的藉由分析來自懷有胎兒之雌性個體的生物樣本識別胎兒之胎兒基因組中之亞染色體畸變的方法300之流程圖。生物樣本包括來自雌性個體及胎兒之無細胞DNA分子。可完全或部分地藉由電腦系統執行方法300。

在區塊310處，量測生物樣本中之DNA分子中之至少一些的大小。由於DNA分子為全部基因組之片段以及染色體之片段，因此DNA分子亦被稱為片段。可經由任何合適之方法(例如，上文描述之方法)量測大小。

在區塊320處，核酸分子中之每一者源自其的參考基因組中識別位置。位置可為基因組的任何部分，該基因組對於所提供之實例為人類，但可用於其他基因組。舉例而言，由於可藉由基因組座標(例如，具體座標或座標之範圍)定義，因此位置可為染色體之一部分。

在一項實施例中，可藉由定序及比較序列資訊與參考人類基因組序列來執行識別。在另一實施例中，可藉由將一組探針與已知染色體來源雜交執行此識別。在微陣列格式中抑或在溶液中，可使用一或多個螢光標記來標記探針。在又另一實施例中，可藉由一組探針在溶液中抑或在固體表面上捕獲核酸分子，且接著定序所捕獲之(或其餘未捕獲之)核酸分子。

在區塊330處，使用基於數目之分析在第一亞染色體區域之生物樣本中偵測畸變，例如，如部分II.A中所描述。舉例而言，可將參考基因組劃分為區間，且可計數定位至每一區間之DNA片段。基於數目，可判定區

域是過度表示抑或不充分表示為偵測畸變之部分。若既未判定過度表示亦未判定不充分表示，則可將區域識別為正常。

在一些實施例中，可使用區塊320中所判定之位置判定位於第一亞染色體區域中之DNA分子的第一量。作為實例，第一量可對應於全部位於第一亞染色體區域內、部分地位於第一亞染色體區域內之DNA分子之數目，及DNA分子與第一亞染色體區域重疊之基因組位置之數目。

可判定位於第二區域中之DNA分子的第二量。在各種實例中，第二區域可為整個基因組、僅一個亞染色體區域、染色體(其可包括第一亞染色體區域)及不相交亞染色體區域(例如，未經測試之所有其他區域)。可利用第一量及第二量計算數目參數。數目參數可與一或多個數目臨限值相比較以判定第一亞染色體區域之生物樣本中現有的畸變類型之數目分類。

畸變之類型之實例為缺失、重複及較高階擴增。畸變中之每一者可對應於不同數目臨限值。舉例而言，當數目參數小於低臨限值時可判定缺失，該低臨限值將低於不存在畸變之區域的彼臨限值。當數目參數大於高臨限值時可判定擴增，該高臨限值將高於不存在畸變之區域之彼臨限值。如上所提及，第二量可包括於數目臨限值中，該數目臨限值僅改變臨限值之值，且因此與利用第一量及第二量判定數目參數相同。

在區塊340處，使用基於大小之分析判定大小分類以用於第一亞染色體區域。大小分類可指示位於第一亞染色體區域中之DNA分子的大小分佈是短於、長於抑或相同於參考區域之彼大小分佈。由於胎兒DNA分子較小，大小分佈之分析可指示相較於參考區域是否存在更多(比參考區域更短大小)或更少(比參考區域更大大小)之胎兒DNA比例，藉此指示相對於參考區域是存在過多、相同抑或不足之胎兒DNA。在各種實例中，參

考區域可為整個基因組、僅一個亞染色體區域、染色體(其可包括第一亞染色體區域)，及不相交亞染色體區域(例如，未經測試之所有其他區域)。如上文關於圖1所解釋，大小分類可用於鑑別胎兒及母親中之畸變的不同可能性。

在一些實施例中，可計算位於第一亞染色體區域中之DNA分子之大小的第一統計值。統計大小之實例為平均大小、大小分佈之峰值處的大小、大小分佈之眾數、給定大小處之累積頻率及類似者。可判定位於參考區域中之DNA分子之大小的參考統計值以用於與第一統計值相比較。可判定第一統計值與參考統計值之間的分離值。分離值可提供相對於參考區域第一亞染色體區域中之胎兒DNA分子之相對比例的量測值。分離值可與一或多個大小臨限值相比較以獲得大小分類。

在區塊350處，基於大小分類及數目分類判定胎兒在第一亞染色體區域中是否具有畸變。可使用圖1進行判定，其中大小分類可用於鑑別各自用於過度表示及不充分表示之數目分類的三個可能性。舉例而言，當胎兒、母親或兩者具有擴增時，可能出現過度表示分類。

若僅胎兒具有擴增，則相較於參考區域，第一亞染色體區域中將存在胎兒DNA分子之更大比例，且第一亞染色體區域之大小分佈將更短。若僅母親具有擴增，則相較於參考區域，第一亞染色體區域中將存在胎兒DNA分子之更小比例，且第一亞染色體區域之大小將更長。若母親及胎兒兩者皆具有擴增，則胎兒DNA與母體DNA兩者具有升高量，因此其比例將相同，且因此大小分佈將相同，從而得到正常分類。

存在於樣本中之許多明顯畸變將為非病原性複本數變異(CNV)，該等變異一般存在於人類種群中。因此，在樣本中所偵測到之畸變可藉由與

各種資料庫相比較進一步評分或評級。此類資料庫具有關於CNV是否存在於特定人類種群之關注區域中、CNV之類型(缺失或重複；增益或損失)、CNV之頻率，及關注區域中是否報告病原性畸變的資訊。舉例而言，自血漿DNA識別之畸變的短清單可與千人基因組(<http://www.1000genomes.org/>)中識別之CNV、變異資料庫(DGV, <http://dgv.tcag.ca/dgv/app/home>)中管理之CNV及/或包含於DECIPHER資料庫([decipher.sanger.ac.uk/](http://decipher.sanger.ac.uk/))中記錄的發育病症之專家評議之微缺失及微重複症候群清單相比較。

在一項實施例中，樣本中所識別之與已知病原性畸變重疊的畸變將被指派較高分值，而樣本中識別之與已知非病原性CNV重疊之畸變將被指派較低分值。可組合每一畸變區域之分值以提供總體病原性分值。

#### IV. 含畸變分數

異常區域之基於數目之z分值的量值與含有畸變之血漿DNA的比例相關(17)。舉例而言，若僅胎兒具有畸變，則含有畸變之血漿DNA的比例將與血漿樣本中之胎兒DNA的比例相關。含有畸變之血漿DNA的比例可用作識別母親具有畸變之情況的額外篩檢。

血漿中之含畸變之分數( $F_{CNA}$ ，亦稱為AcF)係指源自具有CNA之細胞之血漿DNA的比例。理論上，若僅胎兒攜帶畸變，則僅彼等源自胎兒之血漿DNA分子將含有畸變；且 $F_{CNA}$ 將等於血漿中之胎兒DNA分數。類似地，若僅母親攜帶畸變，則僅彼等源自母體之血漿DNA分子將含有畸變；且 $F_{CNA}$ 將等於血漿中之母體DNA分數。另一方面，若母親及胎兒兩者皆攜帶本質上並非嵌合體之畸變(亦即，細胞中之僅一些攜帶畸變)，則所有血漿DNA分子將源自含有畸變之細胞；且 $F_{CNA}$ 將為100%。

在一些實施例中，為了計算 $F_{CNA}$ ，可將整個基因組劃分為2,687 1 Mb的區間，亦稱為區間。如上所述，可計算基於數目之z分值以用於每一區間。具有展示CNA之區域中之最高z分值之1 Mb的區間可用於計算 $F_{CNA}$ 。若區域僅具有一個區間，則將使用彼區間。可如下計算 $F_{CNA}$  (7)：

$$F_{CNA} = \left| \frac{GR_{\text{樣本}} - \text{均值 } GR_{\text{對照組}}}{\text{均值 } GR_{\text{對照組}}} \right| \times 100\% \times 2$$

$GR_{\text{樣本}}$ 為在用於測試案例之具有受影響區域中的最高z分值之1 Mb的區間的基因組表示，且均值 $GR_{\text{對照組}}$ 為對照組中之彼區間之基因組表示的均值。 $F_{CNA}$ 為數目參數之另一實例，該 $F_{CNA}$ 可與數目臨限值相比較以判定是否存在畸變。然而，亦可以其他方式使用 $F_{CNA}$ 。

可針對展示CAN之每一區域計算 $F_{CNA}$ 。 $F_{CNA}$ 可用於判定母親是否存在畸變。倘若超過99%之母體血漿樣品將具有小於50%之胎兒DNA分數(17、18)， $F_{CNA} > 50\%$ 的情況將表明母親攜帶複本數畸變，且因此其不大可能為僅胎兒畸變。對於 $F_{CNA} < 50\%$ 之情況，CNA潛在存在於胎兒中。在一項實施例中，若 $F_{CNA}$ 小於50% (或其他截止值)，則判定畸變為母體中之胎兒或嵌合體。

圖4展示根據本發明之實施例的針對基於數目及基於大小之結果的組合以及 $F_{CNA}$ 值之六種情境的表400。表400類似於圖1之表100。行410展示如正常、複本數增益(擴增)及複本數損失(缺失)之母親的各種狀態。行420展示胎兒之各種狀態。

行430展示基於數目之分類。綠色上指箭頭指示大於高數目臨限值(例如，>3)的基於數目之正z分值。紅色下指箭頭指示小於低數目臨限值(例如<-3)的基於數目之負z分值。雙箭頭指示大量值基於數目之z分值。大量值基於數目之z分值可對應於 $F_{CNA}$ 之某一臨限值，例如(但不限於)，

>40%、>45%、>50%、>55%及>60%。不同截止值可用於定義「大量值z分數」，例如(但不限於)，對於正z分數>10、>15、>20、>25、>30、>35、>40、>45、>50、>55、>60且對於負z分數<-10、<-15、<-20、<-25、<-30、<-35、<-40、<-45、<-50、<-55。

行440展示基於大小之分類。綠色上指箭頭指示大於高大小臨限值(例如，>3)的基於大小之正z分數。紅色下指箭頭指示小於低數目臨限值(例如<-3)的基於數目之負z分數。

行435針對組合中之每一者展示 $F_{CNA}$ 是大於抑或小於50%。如圖所示，當 $F_{CNA}$ 大於50%時，母親將具有畸變。因此，可計算 $F_{CNA}$ 以判定母親是否存在畸變。 $F_{CNA}>50\%$ 表明母親攜帶複本數畸變。在其他實施例中，可使用 $F_{CNA}$ 之其他臨限值，例如(但不限於)，>40%、>45%、>55%及>60%。

因此，可判定第一亞染色體區域之第一區間的第一區間數目參數(例如， $GR_{\text{樣本}}$ )。第一亞染色體區域可包括一或多個區間。可自位於由位於另一區域中之DNA分子的量歸一化之第一區間中之DNA分子的量判定第一區間數目參數，該另一區域可為與用於第一亞染色體區域之歸一化的相同區域或不同區域。可使用對照樣本計算第一區間之對照區間數目參數(例如，均值 $GR_{\text{對照組}}$ )的均值。可藉由減去對照區間數目參數之均值獲得結果及將該結果除以對照區間數目參數之均值來計算第一區間之第一分數(例如， $F_{CNA}$ )，例如，如上文所示。第一分數(例如，絕對值)是否大於截止值(對於上文之實例而言，50%)可用於識別雌性個體是否具有第一亞染色體區域之畸變。取決於第一分數之定義可使用其他截止值，例如，是否使用因素2及100%。

## V. 結果

以下結果確認實施例正確地識別所偵測到之亞染色體畸變是來自胎兒、母親抑或兩者的能力。此類結果展示優於現有技術之改善，現有技術將把所有亞染色體畸變錯誤分類為來自胎兒，此導致假陽性。

### A. 第一組結果

藉由大小分析方法分析自先前研究具有已知胎兒微缺失及微重複之六種母體血漿DNA樣本的成對末端定序資料。在六種測試情況中，存在涉及染色體3q、4q及22q的源自胎兒之亞染色體缺失或重複的五種情況，及22q上母系遺傳微重複之一種情況。利用成對末端讀段之開始及末端座標判定每一定序DNA片段的大小。

對於每一測試情況，將目標區域定義為藉由基於數目之分析識別的含有複本數畸變之區域。參考區域(方法300之第二區域)涵蓋不含畸變之體染色體上之所有不受影響的基因組區域。將具有基於數目之分析的先前研究中所使用之正常胎兒核型的八個單胞胎懷孕案例之同一群組作為參考對照組應用於大小分析中(7)。每一測試樣本之目標區域的基於大小之z分值則計算缺失。

圖5為展示用於說明本發明之實施例之準確性的六種母體血漿DNA樣本之基於數目之分值及基於大小之分值的表500。基於數目之z分值將畸變展示為存在。基於數目之分值的範圍對應於區域之1 Mb的區間。

使用圖4之表400，具有僅源自胎兒之複本數畸變之情況將具有在與基於數目之z分值相同之方向上的基於大小之z分值，即正數將表明擴增而負數將表明缺失。使用來自均值之3個SD之z分值截止值，在案例01至04及06中之由基於數目之方法偵測之所有複本數畸變經獨立確認為源自胎兒

(如表500中所展示)。

案例05為涉及已自其母親遺傳染色體22q上之2.4 Mb的微重複之胎兒的妊娠。由於母親自身攜帶微重複，因此存在所涉及三個1 Mb之區間的基於數目之極高z分值(範圍，39.7至71.7)。然而，此分析本身並不展現胎兒是否已自母親遺傳微重複。

結合基於數目之分析使用基於大小之分析，合併3 Mb之區間在正常範圍內展示基於大小之z分值。當畸變係源自胎兒及母親兩者時，此觀測結果與保持不變(胎兒及母親之相對比重並未更改)之受影響區域中之母體血漿DNA的大小分佈一致。另一方面，若胎兒並未自母親遺傳微重複，則受影響區域中之短片段的比例將減少，產生與基於數目之正z分值相反的基於大小之負z分值。

#### *B. 第二組結果*

圖6為展示關於CNA源自胎兒或母親，或兩者之六個案例之資訊的表600。該等案例中之三者已包括於僅評估基於數目之方法的先前研究中(7)。其餘三者為之前尚未分析的新案例。將具有正常胎兒及母體核型之單胞胎妊娠案例用作對照組。由於三個新案例及已包括於先前研究中之三個案例係使用不同庫製備套組而製備，因此對照組之兩個不同組使用於測試案例之此等兩個群組之分析中。每一組係使用相同庫製備套組來製備且使用與對應之測試案例群組相同之通道數目來定序。

行610展示母親是存在抑或不存在的畸變的已知狀態。行620展示胎兒是存在抑或不存在的畸變之已知狀態。為了測試實施例是否可預測已知狀態，針對每一案例對四個目標區域執行基於數目及基於大小之組合分析，該等四個目標區域包括：4號染色體上之兩個2 Mb的區域、12號染色體上

之一個4 Mb的區域，及22號染色體上之一個3 Mb的區域。

圖7展示根據本發明之實施例針對表600中之六個案例之基於數目及基於大小之組合分析母體血漿DNA。可使用圖4之表400分析圖7中之分值以預測畸變是在胎兒、母親抑或兩者中。行710展示母親是存在抑或不存在畸變的預測狀態。行720展示胎兒是存在抑或不存在畸變之預測狀態。行710及720中之預測對應於行610及620中之已知狀態。

對於CNA僅存在於胎兒中之兩個案例(M10219及HK310)，基於大小之方法確認由基於數目之方法偵測之畸變。對於母親自身攜帶畸變之四個案例，實施例成功地推論該等胎兒中之兩個已遺傳畸變且其他兩個胎兒並未遺傳畸變。在此定組中並未觀測到假陽性。

對於M10219，22號染色體上之3 Mb的微重複經偵測具有為13.4的基於數目之z分值。對於HK310，3 Mb的微缺失在相同區域中經偵測具有為-8.2的基於數目之z分值。此等兩個案例之 $F_{CNA}$ 分別為21.3%及15.1%。對於M10219及HK310，此區域之基於大小之z分值分別為6.9及-6.3，該等z分值指示之受影響區域在M10219中具有更短大小分佈且在HK310中具有更長大小分佈。在兩個案例中，基於大小之z分值在與基於數目之z分值相同之方向上，從而指示胎兒為母體血漿中偵測之畸變的唯一來源。此等結果與圖6中之兩個案例的臨床資訊一致。

對於M14-13489-F1，在4號染色體上偵測到之2 Mb的微重複具有為93.8的基於數目之z分值。對於DNA 11-04530，在4號染色體上之另一區域中偵測到之2 Mb的微缺失具有為-61.9的基於數目之z分值。 $F_{CNA}$ 分別為69.1及82.5。對於具有異常的基於數目之z分值的區域，對應的基於大小之z分值對於M14-13489-F1為-3.6而對於DNA 11-04530為5.1因此，在兩

個案例中，基於大小之z分值在與基於數目之z分值相對的方向上，表明畸變將僅存在於母親中。此等結果與圖6中之臨床資訊一致。此等結果展示在將所有畸變指派給胎兒之當前技術上進行改善之個例。在此，避免此類假陽性。

對於PW503及M11879，在22號染色體上偵測3 Mb的微重複(基於數目之z分值：71.6)，且在12號染色體上偵測4 Mb的微重複(基於數目之z分值：154.5)。此等兩個案例之 $F_{CNA}$ 分別為100%及99.6%。基於大小分析展示在兩個案例之正常範圍內之基於大小之z分值(基於大小之z分值：對於PW503為0.9且對於M11879為0.0)的目標區域，指示此等兩個案例中之母親及胎兒兩者包含微重複。此等結果亦與圖6中之臨床資訊一致。

PW503及M11879為涉及已分別自其母親遺傳染色體22q上之2.4 Mb之微缺失及3.5 Mb之微缺失的胎兒的妊娠。由於母親自身攜帶微重複，判定基於數目之極高z分值。然而，基於數目之分析本身並未展現胎兒是否已自母親遺傳微重複。使用基於大小之方法，兩個案例之3 Mb的及4 Mb的區域分別展示在正常範圍內之基於大小之z分值。此觀測結果與保持不變之受影響區域中之母體血漿DNA的大小分佈一致，此係因為當畸變係源自胎兒及母親兩者時，相較於其他不受影響的區域，受影響區域中之胎兒及母親的相對比重並未更改。另一方面，如在M14-13489-F1中，當胎兒並未自母親遺傳微重複時，受影響區域中之短片段的比例將減少，產生與基於數目之正z分值相反的基於大小之負z分值。

圖8為展示根據本發明之實施例之在每一案例中未偵測到CNA之測試區域的基於數目及基於大小之z分值的表800。使用組合分析在每一案例之其他測試區域中並未偵測到畸變。僅藉由基於數目之方法，使用基於數目

之z分值6.61在M14-13489-F1中之22號染色體上之3 Mb的區域中偵測過度表示。 $F_{CNA}$ 為14.8%且基於大小之z分值为-0.82。因此，由基於數目之分析偵測到之畸變並未由基於大小之分析確認，且與陣列CGH分析一致地將此區域分類為正常。

因此， $F_{CNA}$ 可用於區別母親及胎兒皆具有畸變之例項與基於數目之分析中之假陽性。在關於M14-13489-F1之以上實例中， $F_{CNA}$ 為14.8%，其遠低於50%。因此，母親及胎兒兩者皆呈現畸變係不大可能的，其將對應於基於數目之分析為陽性而基於大小之分析展示正常。以此方式， $F_{CNA}$ 可用作進一步檢查。因此，在一些實施例中，當第一分值小於截止值時且當：數目分類指示擴增且大小分類指示第一亞染色體區域中並不存在畸變，或數目分類指示缺失且大小分類指示第一亞染色體區域中並不存在畸變時，可判定第一亞染色體區域中並不存在畸變。

以以下方式取樣及處理關於第二組結果之上述資料。在書寫知情同意書及機構倫理委員會批准之情況下，自香港威爾斯親王醫院( the Prince of Wales Hospital)及廣華醫院(Kwong Wah Hospital)之婦產科部門及荷蘭內梅亨(Radboud)大學醫療中心招募單胞胎妊娠的女性。如先前所述採集並處理母體末梢血液樣本(16)。使用QIAamp DSP DNA血液微型套件自血漿提取DNA (16)。

以以下方式執行血漿DNA定序。吾人遵循製造商的說明使用KAPA庫製備套件(Kapa生物系統)製備新案例之DNA庫。由以12為週期的PCR富集接附子接合血漿DNA。藉由HiSeq 1500或HiSeq 2500定序器(Illumina)上之流量槽之兩個通道定序每一庫。吾人執行50個週期的成對末端定序。如先前所述對準且過濾成對末端讀段(13)。在對準之後，自成

對末端讀段之開始及末端座標判定每一定序DNA片段的大小。

對於已包括於先前研究中之三個案例，如下文所述重分析此等母體血漿DNA樣本之成對末端定序資料。此等案例之血漿DNA庫係預先使用成對末端定序樣本製備套件(Illumina)製備且使用HiSeq 2000定序器(Illumina)上之流量槽之一個通道來定序。

## VI. 合併區段

如上所述，區間可小於異常區域，且展示畸變之連續區間可經組合以識別異常區域。除了使用具有數目參數之連續區間以外，實施例可使用諸如二元循環分段法及隱式馬爾可夫模型之其他技術來識別對應於異常區域之區間之群組。可合併區間以形成對應於異常區域之合併區段。

### A. 合併區間

作為實例，使用特定大小之窗口將人類基因組劃分為非重疊區間。窗口大小之實例為10 kb、50 kb、100 kb、500 kb及1000 kb等。在一些實施例中，濾出具有低可定位性(例如，小於10%)之區間。可判定每一區間之DNA分子的量，其中GC校正可用於利用原始計數判定量(Chen EZ等人，PLoS One，2011;6(7):e21791)。可定位性對應於藉由對準將來源於區域之讀段指派或識別返回至正確原始基因組位置之能力。某些區域具有低可定位性，例如，歸因於獨特核苷酸內容不足。此類區域在定序深度中無法充分表示。

在對準至區間*i*的GC校正(被稱為基因組表示，GR)之後的讀段比例可被判定且被稱為 $GR_i$ 。 $GR_i$ 可進一步變換為測試樣本之z分值統計， $Z_i$ ：

$$Z_i = \frac{GR_i - GR_{i0}}{SD_{i0}},$$

其中 $GR_{i0}$ 及 $SD_{i0}$ 分別為對應於懷有整倍體胎兒之健康孕婦(正常個體)

之群組中的區間 $i$ 之GR之均值及標準差(SD)。

可接著沿每一染色體對 $Z_i$ 應用分段步驟。此分段步驟可將在相同方向上呈現基因組表示改變(例如，相對過度表示、相對不充分表示或無改變)之連續區間合併為更大區段，稱為合併區段。可以基因組座標之升序或降序次序執行分段。各種技術可用於分段步驟。

在一項實施例中，二元循環分段法及隱式馬爾可夫模型(<https://www.bioconductor.org/packages/3.3/bioc/manuals/snapCGH/man/snapCGH.pdf>)演算法可用於實施此分段步驟。可將展示相較於自不受影響之對照組或個體建立之參考範圍統計學上顯著增大之正 $z$ 分值的值之合併區段識別為候選微重複(或更一般而言為微擴增)。可將展示相較於自不受影響之對照組或個體建立之參考範圍統計學上顯著減小之負 $z$ 分值的值之合併區段識別為候選微缺失。術語「候選」可指區域為胎兒畸變的候選者，其可使用大小分析確認。

可藉由僅大於臨限值定義出正常範圍之顯著偏差，例如，如方法300中所描述。舉例而言，可分析合併區段的大小以判定合併區段是否大於長度臨限值，例如至少3兆鹼基(Mb)。其他長度臨限值之實例包括1 Mb、2 Mb、4 Mb、5 Mb，10 Mb等。

此外，可分析合併區段之偏差的量值以判定該量值是否超過偏差臨限值。舉例而言，合併區段(亦即，包括合併區段中之所有區間)之絕對平均 $z$ 分值被要求要大於偏差臨限值(例如，1.5)。其他偏差臨限值之實例包括1、2、3、5等。量值為數目參數之實例或可經判定為數目參數與數目臨限值之比較的部分，該數目臨限值可為參考範圍。平均 $z$ 分值可為個別 $z$ 分分之平均值或使用整個合併區段之DNA分子總量判定之 $z$ 分值。

在一些實施例中，可執行區間之初始分析以識別可能形成滿足長度臨限值及/或偏差臨限值之合併區段的畸變區間。此初始分析亦可使用z分值分析。初始分析之臨限值可不同(例如，大)於用於整個合併區段之偏差臨限值。一旦識別彼此接近(例如，在指定長度內，諸如小於500 kb之無間距)之足夠數目的畸變區間，可對區域中之區間使用分段處理以識別適合區域。接著，可(例如)使用長度臨限值及/或偏差臨限值分析該區域。長度或偏差可單獨加以測試以識別候選者，或兩者需要被滿足。

### B. 含畸變分數(AcF)

如上所述，含畸變分數可用作數目參數以判定第一亞染色體區域之生物樣本中現有的畸變之類型之數目分類。因此，可使用含畸變分數而非區域之z分值或(例如)藉由合併區段定義之區域之區間的平均z分值。含畸變分數因此可用作來自合併區段之參考範圍之偏差。含畸變分數可對應於樣本中含有畸變之等效細胞的比例。

可以各種方式(例如，使用部分IV中之定義)定義含畸變分數(表示為 $F_{CNA}$ )。然而，可以其他方式定義在此吾人表示為AcF之含畸變分數。

在一項實施例中，可使用下列等式定義AcF：

$$AcF = \frac{|GR' - GR_0| \times 2 \times 100\%}{GR_0}$$

其中 $GR'$ 為測試樣本中展示微缺失或微重複之合併區段(區域)的GR，且 $GR_0$ 為對照(參考)樣本中之合併區段的均值GR。

在另一實施例中，可使用下列等式定義AcF：

$$Or\ AcF = |Z' \times CV_0| \times 2$$

其中 $Z'$ 為測試樣本中展示微缺失或微重複之合併區段(區域)之z分值。  $CV_0$ 為正常個體(亦稱為對照或參考個體或樣本)中之對應區域的變異

係數。在一項實施例中，(例如)根據部分II.A中定義之 $z$ 分值藉由比較測試樣本中之區域之GR與正常個體中之對應區域之GR的均值及標準差來重新計算合併區段之 $Z'$ 。在另一實施例中，亦可藉由將100 kb之 $z$ 分值的總和除以所涉及的多個區間之平方根而利用屬於合併區段內之區間的一系列個別 $z$ 分值來估計 $Z'$ 。

AcF可反映畸變來源之潛在組織。舉例而言，若畸變僅來源於胎兒，則AcF將等於胎兒DNA分數。若畸變僅來源於母親，則AcF將遠大於胎兒DNA分數，大體而言，此係因為胎兒DNA相當於血漿中之少數比例。若畸變來源於兩者，則所分析之AcF將接近於100%。

因此，AcF與胎兒DNA分數之間的分離值(例如，差值或比率)可用於分類樣本中可見之基因組畸變之來源的組織。可使用諸如基於SNP、基於大小及基於chrY之方法的各種技術來計算樣本之胎兒DNA分數，例如，如美國專利公開案2013/0237431中所描述。

下文為AcF與胎兒DNA分數之間的差值可如何用於推斷畸變是否來源於母親或胎兒的一些實例。舉例而言，若AcF與胎兒DNA分數之間的差值小於低臨限值(例如，2%)，則畸變將被分類為「源自胎兒之畸變」。臨限值之其他實例包括1%、3%、4%及5%。

若AcF與胎兒DNA分數之間的差值大於高臨限值(例如，20%)，則畸變將被分類為「涉及母親之畸變」。高臨限值之其他實例包括10%、30%、40%及50%。當超過高臨限值時，畸變可僅在母親中或在母親及胎兒兩者中。

由於來自背景母體細胞之畸變可為嵌合體(亦即，提供血漿DNA之母體細胞的僅一比例(<100%)含有畸變)，因此在畸變僅源自母親或母親及

胎兒兩者時AcF與胎兒DNA分數之間的差值可具有類似值。在一項實施例中，展示AcF超過某一臨限值(例如，4%、5%、6%等)之僅一區域可被視為候選微缺失或微重複。

第一亞染色體區域之第一區間的第一區間數目參數可對應於整個區域之GR，例如，當區域僅具有一個區間時。可藉由自第一區間數目參數減去對照區間數目參數之均值來判定區域之第一分值(AcF或FCNA)。在各種實施例中，減法運算之結果可除以正常個體之對照區間數目參數的均值或標準差。

如上所述，可在生物樣本中量測胎兒DNA濃度。可計算第一分值與胎兒DNA濃度之間的差值。判定第一分值是否大於截止值可包括判定差值是否大於高臨限值，例如，如上文所述。此外，可比較差值與低臨限值以判定當差值低於低臨限值時僅胎兒具有第一亞染色體區域之畸變，例如，如上文所述。

在其中第一亞染色體區域包括複數個區間之實施例中，可計算該複數個區間中之每一者的各別分值。可使用各別分值之總和判定第一分值。舉例而言，可採用平均值。作為另一實例，第一分值为總和除以用於判定總和之多個區間之平方根。

### C. 結果

圖9為展示根據本發明之實施例之由藉由分段處理判定之100 kb的區間組成之擴增區域的曲線圖900。曲線圖900提供微重複識別及整合解釋之實例。將每一染色體劃分為100 kb的區間。在曲線圖900中，每一圓點表示100 kb的區間中之z分值。

執行二元循環分段法及基於隱式馬爾可夫模型(HMM)之分段兩者。

在22號染色體中，兩種分段演算法展示一致結果。使用合併區段的大小 $>3$  Mb之截止值及平均z分値之量值 $>1.5$ 之截止值將如陰影區的區域中所展示之合併區段(chr22：17,000,000至20,000,000)識別為候選微重複。根據合併區段之基於數目之合併z分値72將含畸變分數(AcF)判定為100%。

使用 FetalQuant 演算法 (Jiang P 等人，Bioinformatics，2012;28(22):2883-90)胎兒DNA分數為22%。AcF遠大於胎兒DNA分數，其指示此區域中將存在母親之畸變。合併區域之基於大小之z分値經判定為0.9，當與截止值3比較時表明正常大小分佈。因此，根據圖1及圖4，大小分析表明在此區域中母親及胎兒兩者皆具有微重複。藉由與資料庫相比較，發現此候選微重複與22q11重複症候群重疊。

圖10為展示根據本發明之實施例之由藉由分段處理判定之100 kb的區間組成之缺失區域的曲線圖1000。曲線圖1000提供微缺失識別及整合解釋之實例。將每一染色體劃分為100 kb的區間。在曲線圖1000中，每一圓點表示100 kb的區間中之z分値。

執行二元循環分段法及基於隱式馬爾可夫模型(HMM)之分段兩者。在4號染色體中，兩種分段演算法展示一致結果。使用合併區段的大小 $>3$  Mb之截止值及平均z分値 $>1.5$ 之量值將如陰影區的區域中所展示之合併區段(chr4：158,000,000至198,000,000)識別為候選微缺失。根據合併區段之基於數目之合併z分値-74.5含畸變分數(AcF)經判定為14.7%。

使用 FetalQuant 演算法 (Jiang P 等人，Bioinformatics，2012;28(22):2883-90)胎兒DNA分數為13%。AcF非常接近胎兒DNA分數(例如，小於2%之低臨限值)，其指示在此區域中存在僅胎兒具有微缺失。合併區段之基於大小之z分値經判定為-14，表明與截止值-3相比較之

顯著更長大小分佈。因此，根據圖1及圖4，大小分析表明在此區域中胎兒將具有微缺失。

## VII. 概述

即使具有高偵測率及低假陽性率，歸因於不充分高正預測性值，針對胎兒亞染色體非整倍體使用母體血漿中之無細胞DNA的NIPT當前亦未廣泛地用作篩選測試。若包括亞染色體CNA，由於此等條件之個別成份甚至比全染色體非整倍體更罕見，則測試之正預測性值將被預期為甚至更低。另外，歸因於多個比較，假陽性之數目將隨著更多目標經測試而增加。如由Yin等人所報導，其55個假陽性樣本中之20個可歸因於定序及統計誤差(12)。如本發明之實施例中所展示，基於大小之分析可充當確認由基於數目之分析偵測之畸變的獨立方法。結果展示歸因於使用基於數目及基於大小之組合方法的統計誤差，吾人可將假陽性數目減至最低。

在一些實施例中，在胎兒DNA分數為5%時，為了達成具有95%之敏感度及99%之特異度之胎兒亞染色體CNA偵測的2 Mb之分辨率，基於數目及基於大小之方法兩者將皆需要分析約2億個分子(7)。另一方面，由於早期妊娠中之中位胎兒DNA分數為約15% (17、19)，故約2千萬分子可用於達成相同效能。此估計係基於預先所報導之數學關係，其中胎兒DNA分數之每兩倍增加將引起相同測試效能要求之分子降低4倍(20)。由於同一組定序資料可用於兩種分析類型，相較於要求用於單末端定序之試劑的僅計數協定，實施例僅要求用於成對末端定序之額外試劑。另外，兩個協定需要之用於生物資訊處理之時間需求為相當的。

總體而言，吾人已表明對懷孕女性中之血漿DNA的大小分析可準確地偵測胎兒亞染色體CNA。基於大小及基於數目之方法之組合使用可進

一步判定是胎兒、母親抑或其兩者攜帶畸變。此組合方法在幫助臨床醫師解釋NIPT結果中係極有價值的。

## VIII. 電腦系統

本文中提及之任何電腦系統均可利用任何適合數目之子系統。此等子系統之實例展示於圖11中之電腦裝置10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，電腦系統可包括具有內部組件之多個電腦裝置，其各自為子系統。電腦系統可包括桌上型及膝上型電腦、平板電腦、行動電話及其他行動器件。

圖11中所展示之子系統經由系統匯流排75互連。展示額外子系統，諸如印表機74、鍵盤78、儲存器件79、耦接至顯示配接器82之監視器76及其他。耦接至I/O控制器71之周邊裝置及輸入/輸出(I/O)器件可藉由所屬領域中已知之多種構件(諸如輸入/輸出(I/O)埠77(例如，USB、FireWire®))連接至電腦系統。舉例而言，I/O埠77或外部介面81(例如，乙太網、Wi-Fi等)可用於將電腦系統10連接至廣域網路(諸如網際網路)、滑鼠輸入器件或掃描儀)。經由系統匯流排75之互連允許中央處理器73與每一子系統通信且控制來自系統記憶體72或儲存器件79(例如，固定磁碟，諸如硬碟機或光碟)之指令的執行以及子系統之間資訊的交換。系統記憶體72及/或儲存器件79可實施電腦可讀媒介。另一子系統為資料採集器件85，諸如，攝影機、麥克風、加速計及類似者。本文所提及之任何資料可自一個組件輸出至另一個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81或藉由內部介面連接在一起的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。在此等情況下，可將一個電腦視為用戶端且將另一電腦視

為伺服器，其中每一者可為同一電腦系統之一部分。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，本發明之任何實施例可使用硬體(例如，特殊應用積體電路或場可程式閘極陣列)以控制邏輯之形式實施，及/或使用具有通用可程式化處理器之電腦軟體以模組或整合式方式實施。如本文中所使用，處理器包括單核心處理器、位於同一整合式晶片上之多核心處理器或位於單一電路板上或網路化之多個處理單元。基於本發明及本文所提供之教示，一般熟習此項技術者將知道及瞭解使用硬體及硬體與軟體之組合來實施本發明之實施例的其他方式及/或方法。

本申請案中所描述之任何軟體組件或功能可實施為待由使用任何適合之電腦語言(諸如，Java、C、C++、C#、Objective-C、Swift)或腳本語言(諸如Perl或Python)的處理器使用(例如)習知或目標定向技術來執行的軟體程式碼。軟體程式碼可儲存為用於儲存及/或傳輸之電腦可讀媒介上的一系列指令或命令，適合之媒介包括隨機存取記憶體(RAM)、唯讀記憶體(ROM)、磁性媒介(諸如硬碟機或軟性磁碟)或光學媒介(諸如光碟(CD)或DVD(數位化通用光碟))、快閃記憶體及其類似者。電腦可讀媒介可為此類儲存或傳輸器件之任何組合。

此等程式亦可使用適用於經由有線、光學及/或符合多種協定之無線網路(包括網際網路)傳輸的載波信號來編碼及傳輸。因此，可使用以此等程式編碼之資料信號產生根據本發明之實施例的電腦可讀媒介。以程式碼編碼之電腦可讀媒介可與相容器件一起封裝或與其他器件分開提供(例如，經由網際網路下載)。任何此等電腦可讀媒介可存在於單一電腦產品(例如，硬碟機、CD或整個電腦系統)上或其內部，且可存在於系統或網

路內之不同電腦產品上或其內部。電腦系統可包括用於向使用者提供本文所提及之任何結果的監視器、印表機、或其他適合之顯示器。

本文所描述之任何方法可完全或部分地使用電腦系統來進行，該電腦系統包括一或多個處理器，該等處理器可經組態以執行該等步驟。因此，實施例可針對經組態以執行本文所描述之任何方法之步驟的電腦系統，潛在地使用不同組件執行各別步驟或各別步驟組。儘管本文中方法之步驟以經編號之步驟呈現，但其可同時或以不同次序執行。另外，此等步驟之部分可與其他方法之其他步驟之部分一起使用。另外，步驟之全部或部分可視情況選用。另外，任何方法之任何步驟皆可使用執行此等步驟之模組、電路或其他構件來執行。

可在不脫離本發明之實施例的精神及範疇的情況下以任何適合之方式組合特定實施例之特定細節。然而，本發明之其他實施例可針對與各個別態樣或此等個別態樣之特定組合相關的特定實施例。

已出於說明及描述之目的呈現本發明之實例實施例的上述描述。該描述並不意欲為窮盡性的或將本發明限制於所描述之精確形式，且鑒於上文教示，許多修改及變化係可能的。

除非明確指示相反，否則「一(a/an)」或「該(the)」之敘述欲意謂「一或多個(種)」。除非明確指示相反，否則「或」之使用欲意謂「包括或」而並非「互斥或」。

本文所提及之所有專利、專利申請案、公開案及描述均出於所有目的以全文引用之方式併入。不容許任一者為先前技術。

## IX. 參考文獻

1. Bianchi DW, Wilkins-Haug L. Integration of noninvasive

DNA testing for aneuploidy into prenatal care: What has happened since the rubber met the road? *Clin Chem* 2014;60:78-87.

2. Wong FCK, Lo YMD. Prenatal diagnosis innovation: Genome sequencing of maternal plasma. [Epub ahead of print] *Annu Rev Med* October 15, 2015 as doi:10.1146/annurev-med-091014-115715.

3. Chiu RWK, Cantor CR, Lo YMD. Non-invasive prenatal diagnosis by single molecule counting technologies. *Trends Genet* 2009;25:324-31.

4. Peters D, Chu T, Yatsenko SA, Hendrix N, Hogge WA, Surti U, et al. Noninvasive prenatal diagnosis of a fetal microdeletion syndrome. *N Engl J Med* 2011;365:1847-8.

5. Jensen TJ, Dzakula Z, Deciu C, van den Boom D, Ehrich M. Detection of microdeletion 22q11.2 in a fetus by next-generation sequencing of maternal plasma. *Clin Chem* 2012;58:1148-51.

6. Srinivasan A, Bianchi DW, Huang H, Sehnert AJ, Rava RP. Noninvasive detection of fetal subchromosome abnormalities via deep sequencing of maternal plasma. *Am J Hum Genet* 2013;92:167-76.

7. Yu SCY, Jiang P, Choy KW, Chan KCA, Won H-S, Leung WC, et al. Noninvasive prenatal molecular karyotyping from maternal plasma. *PLoS ONE* 2013;8:e60968.

8. Zhao C, Tynan J, Ehrich M, Hannum G, McCullough R, Saldivar J-S, et al. Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma. *Clin Chem*

2015;61:608-16.

9. Hayden EC. Prenatal-screening companies expand scope of DNA tests. *Nature* 2014;507:19-19.

10. Lau TK, Jiang FM, Stevenson RJ, Lo TK, Chan LW, Chan MK, et al. Secondary findings from non-invasive prenatal testing for common fetal aneuploidies by whole genome sequencing as a clinical service. *Prenat Diagn* 2013;33:602-8.

11. Snyder MW, Simmons LE, Kitzman JO, Coe BP, Henson JM, Daza RM, et al. Copy-number variation and false positive prenatal aneuploidy screening results. *N Engl J Med* 2015;372:1639-45.

12. Yin A-H, Peng C-F, Zhao X, Caughey BA, Yang J-X, Liu J, et al. Noninvasive detection of fetal subchromosomal abnormalities by semiconductor sequencing of maternal plasma DNA. *Proc Natl Acad Sci U S A* 2015; 112: 14670-5.

13. Yu SCY, Chan KCA, Zheng YWL, Jiang P, Liao GJW, Sun H, et al. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc Natl Acad Sci U S A* 2014;111:8583-8.

14. Chan KCA, Zhang J, Hui ABY, Wong N, Lau TK, Leung TN, et al. Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem* 2004;50:88-92.

15. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* 2010;2:61ra91.

16. Chiu RWK, Chan KCA, Gao Y, Lau VYM, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2008;105:20458-63.

17. Chiu RWK, Akolekar R, Zheng YWL, Leung TY, Sun H, Chan KCA, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ* 2011;342:c7401.

18. Palomaki GE, Kloza EMM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, et al. DNA sequencing of maternal plasma to detect Down syndrome: An international clinical validation study. *Genet Med* 2011;13:913-20.

19. Hudcova I, Sahota D, Heung MMS, Jin Y, Lee WS, Leung TY, et al. Maternal plasma fetal DNA fractions in pregnancies with low and high risks for fetal chromosomal aneuploidies. *PLoS ONE* 2014;9:e88484.

20. Lo YMD, Lun FMF, Chan KCA, Tsui NBY, Chong KC, Lau TK, et al. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc Natl Acad Sci U S A* 2007;104:13116-21.

#### 【符號說明】

10:電腦系統

71: I/O控制器

72:系統記憶體

73:中央處理器

74:印表機

75:系統匯流排

76:監視器

77: I/O埠

78:鍵盤

79:儲存器件

81:外部介面

82:顯示配接器

85:資料採集器件

100:表

110:行

120:行

130:行

140:行

202:區域

204:正常區域

205:母體血漿

210:樣本

220:粗紅色片段

230:區塊

232:區間

240:區塊

245:區塊

250:對照組

260:區塊

270:區塊

280:區塊

300:方法

310:區塊

320:區塊

330:區塊

340:區塊

350:區塊

400:表

410:行

420:行

430:行

435:行

440:行

500:表

600:表

610:行

620:行

700:表

710:行

720:行

800:表

900:曲線圖

1000:曲線圖

## 【發明申請專利範圍】

### 【請求項1】

一種藉由分析來自懷有胎兒雌性個體之生物樣本來識別該胎兒的胎兒基因組中亞染色體複本數畸變之方法，該生物樣本包括來自該雌性個體及該胎兒之無細胞DNA分子，該方法包含：

對於該生物樣本中之複數個DNA分子中之每一者：

識別參考基因組中該DNA分子之位置，從而獲得該複數個DNA分子之位置；

對於該參考基因組中之複數個非重疊區間(non-overlapping bin)中之每一者：

利用該複數個DNA分子之位置來判定位於各別非重疊區間中之該複數個DNA分子之各別量；

利用該各別量計算各別區間數目參數(bin count parameter)；及

比較該各別區間數目參數與一或多個區間數目臨限值，以判定該各別非重疊區間之畸變類型之區間數目分類；

執行分段處理，該分段處理基於具有相同區間數目分類之連續非重疊區間將該等連續非重疊區間合併為合併區段；

比較該合併區段之長度與長度臨限值；

利用該合併區段之非重疊區間之區間數目參數計算合併區段之數目參數，或利用位於該合併區段之非重疊區間中之該複數個DNA分子之量計算合併區段之數目參數；

比較該數目參數與一或多個偏差臨限值以判定該畸變類型之數目分類；以及

基於該合併區段之長度與該長度臨限值之比較及該數目分類判定該胎兒在對應於合併區段之亞染色體區域中是否具有該亞染色體複本數畸變，其中當該合併區段之長度超過該長度臨限值且該數目分類指示存在畸變時，判定該胎兒具有亞染色體複本數畸變。

**【請求項2】**

如請求項1之方法，其中該分段處理包括至少下列其中一者：二元循環分段法(binary circular segmentation)及隱式馬爾可夫模型(Hidden Markov model)。

**【請求項3】**

如請求項1之方法，其中該區間數目分類為過度表示或未充分表示。

**【請求項4】**

如請求項1之方法，其中該亞染色體複本數畸變為缺失或重複。

**【請求項5】**

如請求項1之方法，其中當該數目參數大於該一或多個偏差臨限值之偏差臨限值時，判定該胎兒在該亞染色體區域中具有擴增。

**【請求項6】**

如請求項1之方法，其中當該數目參數小於該一或多個偏差臨限值之偏差臨限值時，判定該胎兒在該亞染色體區域中具有缺失。

**【請求項7】**

如請求項1之方法，其中該等非重疊區間之大小相等。

**【請求項8】**

如請求項1之方法，其中該等非重疊區間排除具有小於預定百分比之可定位性之區間(bin)。

**【請求項9】**

如請求項1之方法，其中判定各別量包含利用GC校正。

**【請求項10】**

如請求項1之方法，其中計算各別區間數目參數包含計算z分值，且其中該z分值包含利用自懷有整倍體胎兒之健康孕婦之群組判定之量。

**【請求項11】**

如請求項1之方法，其中該一或多個區間數目臨限值大於該一或多個偏差臨限值之偏差臨限值。

**【請求項12】**

如請求項1之方法，其中該合併區段之數目參數係來自於位於該合併區段之非重疊區間中之該複數個DNA分子之量。

**【請求項13】**

如請求項1之方法，其進一步包含：

當該數目參數高於該一或多個偏差臨限值之偏差臨限值時，判定該胎兒及該雌性個體具有亞染色體複本數畸變。

**【請求項14】**

如請求項12之方法，其進一步包含：

判定生物樣本中之胎兒DNA分數，以及

利用該胎兒DNA分數判定該一或多個偏差臨限值之偏差臨限值。

**【請求項15】**

如請求項14之方法，其進一步包含：

比較該數目參數與該一或多個偏差臨限值之第一偏差臨限值及第二偏差臨限值，以及

當該數目參數大於該第一偏差臨限值，且小於該第二偏差臨限值時：

判定該胎兒具有亞染色體複本數畸變，及  
判定該雌性個體不具有亞染色體複本數畸變。

**【請求項16】**

如請求項14之方法，其進一步包含：

比較該數目參數與該一或多個偏差臨限值之第一偏差臨限值及第二偏差臨限值，以及

當該數目參數大於該第一偏差臨限值，且大於該第二偏差臨限值時：

判定該雌性個體具有亞染色體複本數畸變。

**【請求項17】**

一種包含電腦可讀媒介的電腦產品，該電腦可讀媒介儲存用於控制電腦系統以執行如請求項1至16中任一項之方法之操作的複數個指令。

**【請求項18】**

一種用於識別胎兒的胎兒基因組中亞染色體複本數畸變之系統，該系統包含：

如請求項17之電腦產品；及

一或多個處理器，其用於執行儲存於該電腦可讀媒介上的指令。

**【請求項19】**

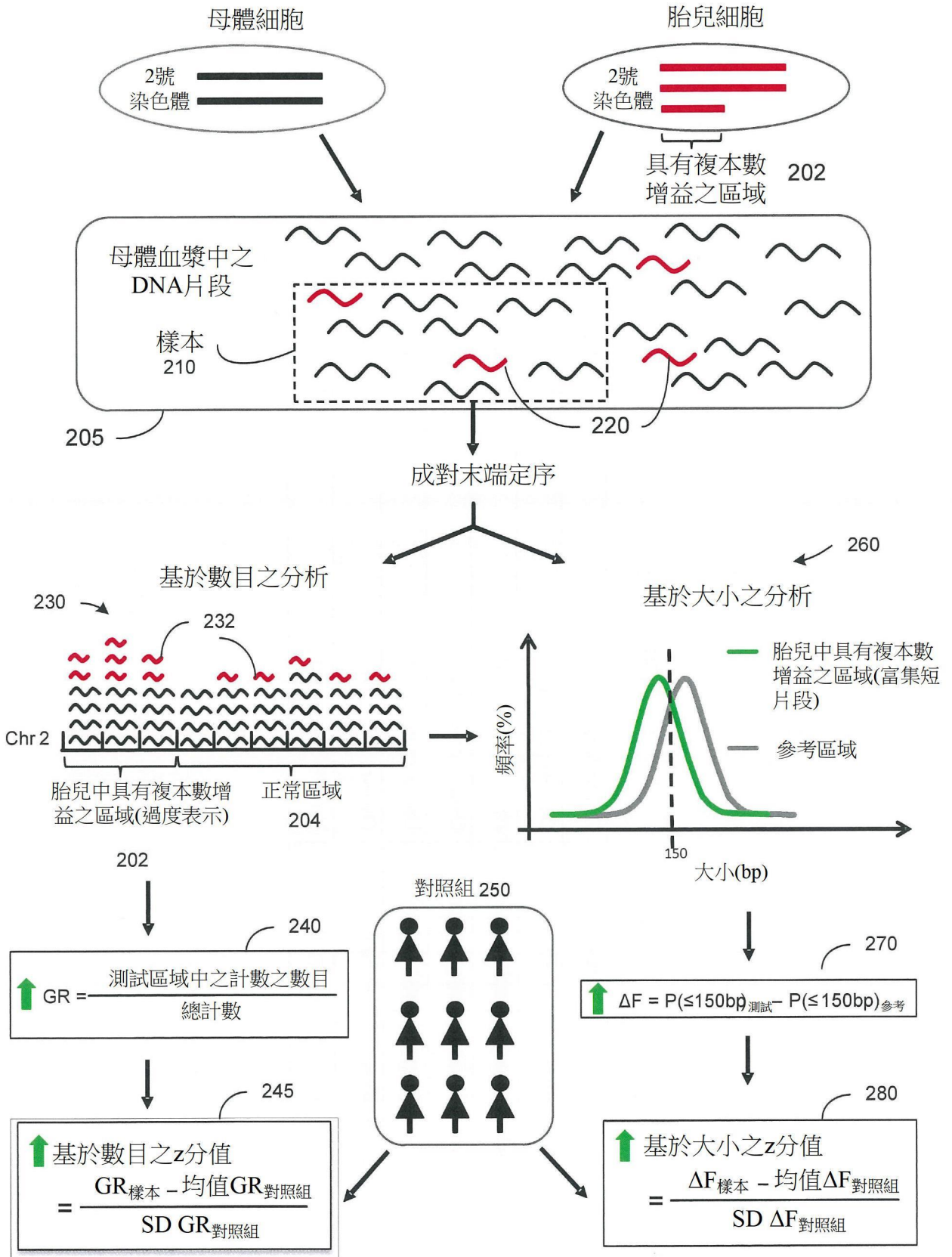
一種用於識別胎兒的胎兒基因組中亞染色體複本數畸變之系統，該系統包含用於執行如請求項1至16中任一項之方法之構件。

【發明圖式】

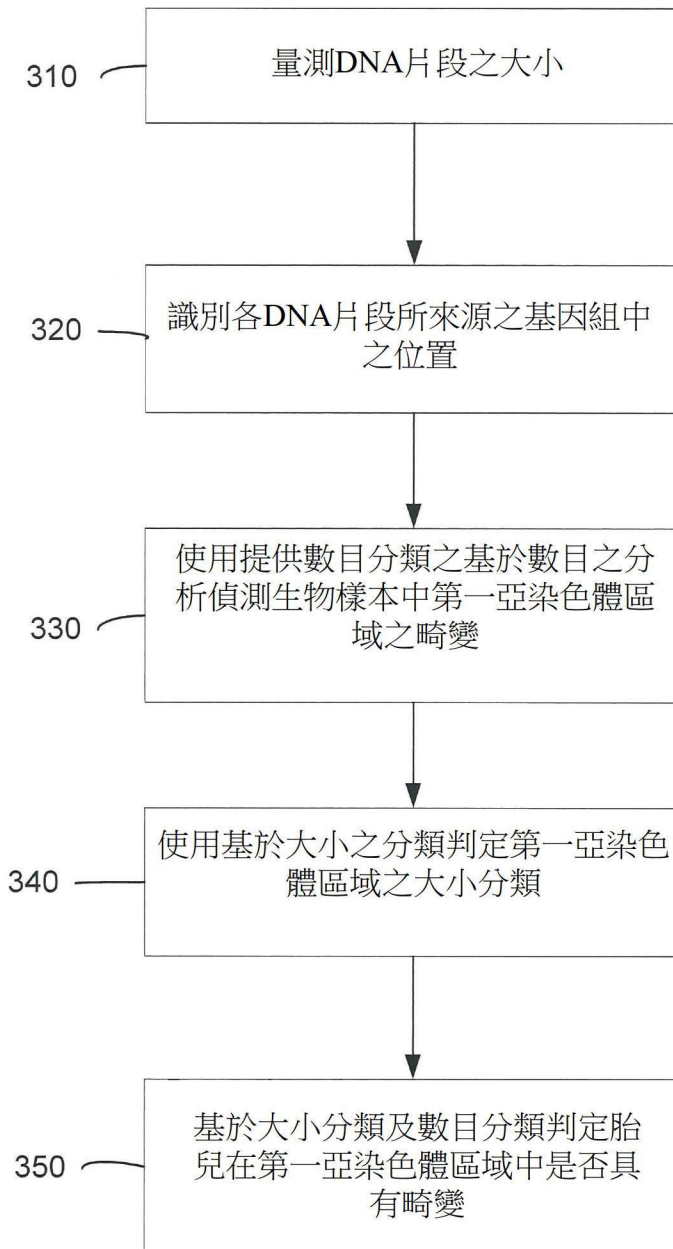
100

|     |     |       |      |
|-----|-----|-------|------|
| 110 | 120 | 130   | 140  |
| 母親  | 胎兒  | 基於數目  | 基於大小 |
| 重複  | 重複  | 過度表示  | 正常   |
| 缺失  | 缺失  | 不充分表示 | 正常   |
| 正常  | 重複  | 過度表示  | 更短   |
| 正常  | 缺失  | 不充分表示 | 更長   |
| 重複  | 正常  | 過度表示  | 更長   |
| 缺失  | 正常  | 不充分表示 | 更短   |

【圖1】



【圖2】



【圖3】

400

| 410<br>母親 | 420<br>胎兒 | 430<br>基於數目之<br>z分値 | 435<br>FcNA | 440<br>基於大小之<br>z分値 |
|-----------|-----------|---------------------|-------------|---------------------|
| 正常        | 複本數增益     | ←                   | < 50 %      | ←                   |
| 正常        | 複本數損失     | →                   | < 50 %      | →                   |
| 複本數增益     | 複本數增益     | ←                   | > 50 %      | 正常                  |
| 複本數損失     | 複本數損失     | →                   | > 50 %      | 正常                  |
| 複本數增益     | 正常        | ←                   | > 50 %      | →                   |
| 複本數損失     | 正常        | →                   | > 50 %      | ←                   |

【圖4】

500

| 案例 | 藉由標籤計數法偵測之複本數畸變之<br>基因組座標(大小)           | 基於數目之z分數   | 基於大小之z分數 |
|----|---|------------|----------|
| 01 | 22號染色體：17,000,000 – 20,000,000 (3 Mb)   | -3.7至-8.3  | -4.22    |
| 02 | 22號染色體：17,000,000 – 20,000,000 (3 Mb)   | -7.3至-13.2 | -5.76    |
| 03 | 22號染色體：17,000,000 – 20,000,000 (3Mb)    | -5.2至-9.6  | -6.56    |
| 04 | 22號染色體：17,000,000 – 20,000,000 (3 Mb)   | 6.8至13.0   | 6.95     |
| 05 | 22號染色體：17,000,000 – 20,000,000 (3 Mb)   | 39.7至71.7  | 0.34     |
| 06 | 3號染色體：194,000,000 – 199,000,000 (5 Mb)  | 3.2至10.8   | 8.10     |
|    | 4號染色體：158,000,000 – 189,000,000 (31 Mb) | -3.8至-12.3 | -14.40   |

【圖5】



| 案例                  | 孕齡<br>(週數) | 畸變  | 母親  | 胎兒  |
|---------------------|------------|---|-----|-----|
| M10219 <sup>a</sup> | 12 3/7     | 22號染色體上之2.4 Mb增益(藉由陣列CGH <sup>b</sup> 確認) | 不存在 | 存在  |
| HK310 <sup>a</sup>  | 22 5/7     | 22q11.2上之微缺失(藉由QF-PCR及FISH確認)             | 不存在 | 存在  |
| M14-13489-F1        | 12 2/7     | 4號染色體上之1.6 Mb增益(藉由陣列CGH <sup>b</sup> 確認)  | 存在  | 不存在 |
| DNA11-04530         | 14         | 4號染色體上之1.33 Mb損失(藉由陣列CGH <sup>b</sup> 確認) | 存在  | 不存在 |
| PW503 <sup>a</sup>  | 20 2/7     | 22號染色體上之2.4 Mb增益(藉由陣列CGH <sup>b</sup> 確認) | 存在  | 存在  |
| M11879              | 22 5/7     | 12號染色體上之3.5 Mb增益(藉由陣列CGH <sup>b</sup> 確認) | 存在  | 存在  |

<sup>a</sup>已包括於先前研究中之案例(7)。

<sup>b</sup>陣列CGH，陣列比較性基因組雜交。

【圖6】

700

710

720

| 案例           | 展示CNA之區域的基因組座標(區域之大小)                  | 基於數目之z分數 | 含畸變分數(%) | 畸變存在於母親中? | 基於大小之z分數 | 胎兒受影響? |
|--------------|--|----------|----------|-----------|----------|--------|
| M10219       | 22號染色體: 17,000,000 – 20,000,000 (3 Mb) | 13.4     | 21.3     | 否         | 6.9      | 是      |
| HK310        | 22號染色體: 17,000,000 – 20,000,000 (3 Mb) | -8.2     | 15.1     | 否         | -6.3     | 是      |
| M14-13489-F1 | 4號染色體: 60,000,000 – 62,000,000 (2 Mb)  | 93.8     | 69.1     | 是         | -3.6     | 否      |
| DNA11-04530  | 4號染色體: 97,000,000 – 99,000,000 (2 Mb)  | -61.9    | 82.5     | 是         | 5.1      | 否      |
| PW503        | 22號染色體: 17,000,000 – 20,000,000 (3 Mb) | 71.6     | 100      | 是         | 0.9      | 是      |
| M11879       | 12號染色體: 12,000,000 – 16,000,000 (4 Mb) | 154.5    | 99.6     | 是         | 0.0      | 是      |

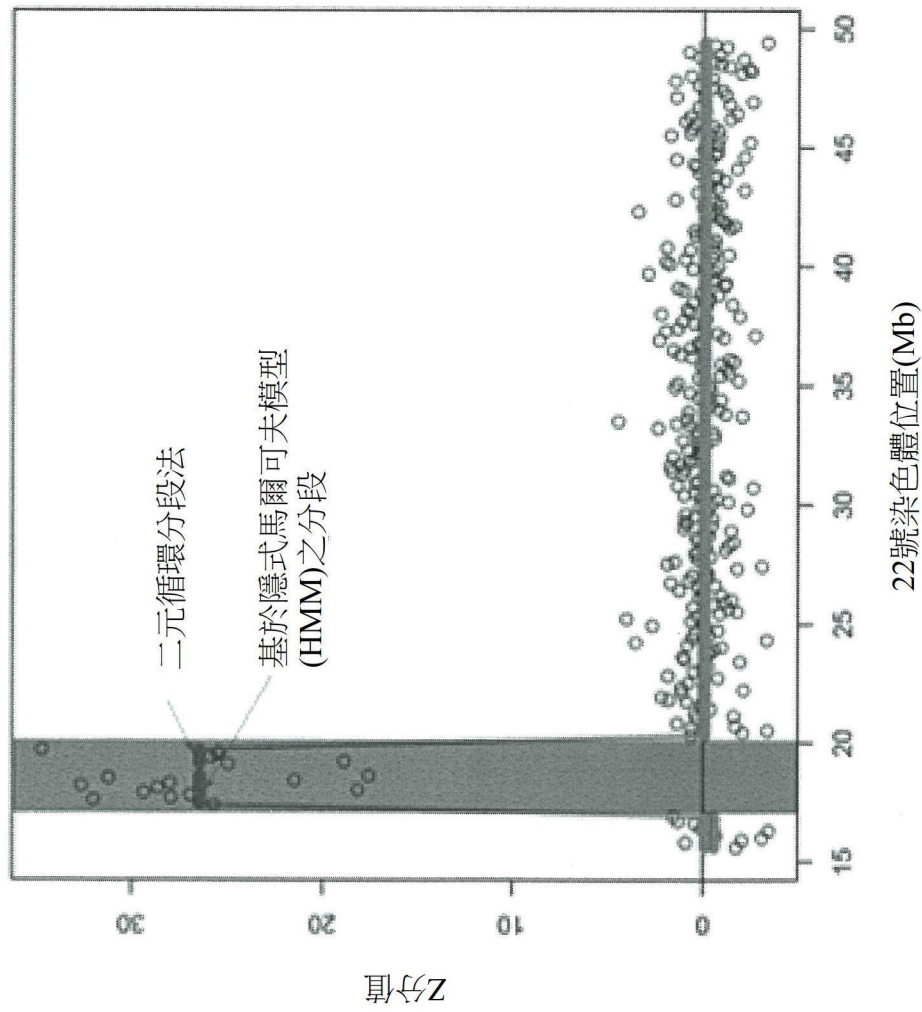
【圖7】

800

| 案例           | 測試區域的基因組座標<br>(區域之大小)                    | 基於數目之<br>z分値 | 基於大小之<br>z分値 |
|--------------|--|--------------|--------------|
| M10219       | 4號染色體 : 60,000,000 - 62,000,000 (2 Mb)   | 0.84         | -1.31        |
|              | 4號染色體 : 97,000,000 - 99,000,000 (2 Mb)   | 0.56         | -2.05        |
|              | 12號染色體 : 12,000,000 - 16, 000,000 (4 Mb) | 0.41         | 0.07         |
| HK310        | 4號染色體 : 60,000,000 - 62,000,000 (2 Mb)   | 0.28         | 1.95         |
|              | 4號染色體 : 97,000,000 - 99,000,000 (2 Mb)   | -0.00        | 1.05         |
|              | 12號染色體 : 12,000,000 - 16, 000,000 (4 Mb) | 0.66         | 0.28         |
| M14-13489-F1 | 4號染色體 : 97,000,000 - 99,000,000 (2 Mb)   | 0.04         | -0.45        |
|              | 12號染色體 : 12,000,000 - 16, 000,000 (4 Mb) | -0.04        | -1.26        |
|              | 22號染色體 : 17,000,000 - 20,000,000 (3 Mb)  | 6.61         | -0.82        |
| DNA11-04530  | 4號染色體 : 60,000,000 - 62,000,000 (2 Mb)   | 1.92         | -2.69        |
|              | 12號染色體 : 12,000,000 - 16, 000,000 (4 Mb) | -2.28        | 0.74         |
|              | 22號染色體 : 17,000,000 - 20,000,000 (3 Mb)  | -0.68        | 2.63         |
| PW503        | 4號染色體 : 60,000,000 - 62,000,000 (2 Mb)   | 2.69         | -1.20        |
|              | 4號染色體 : 97,000,000 - 99,000,000 (2 Mb)   | 0.12         | -1.34        |
|              | 12號染色體 : 12,000,000 - 16, 000,000 (4 Mb) | 0.19         | -0.80        |
| M11879       | 4號染色體 : 60,000,000 - 62,000,000 (2 Mb)   | -0.42        | -2.02        |
|              | 4號染色體 : 97,000,000 - 99,000,000 (2 Mb)   | 0.55         | 0.24         |

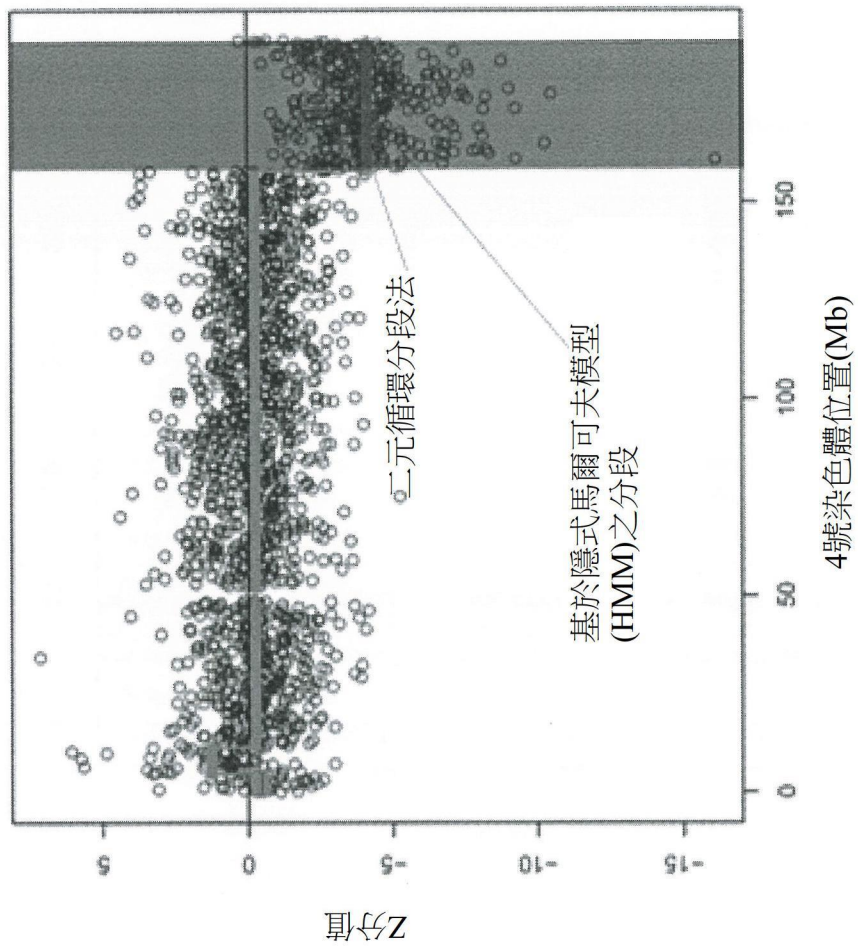
【圖8】

900

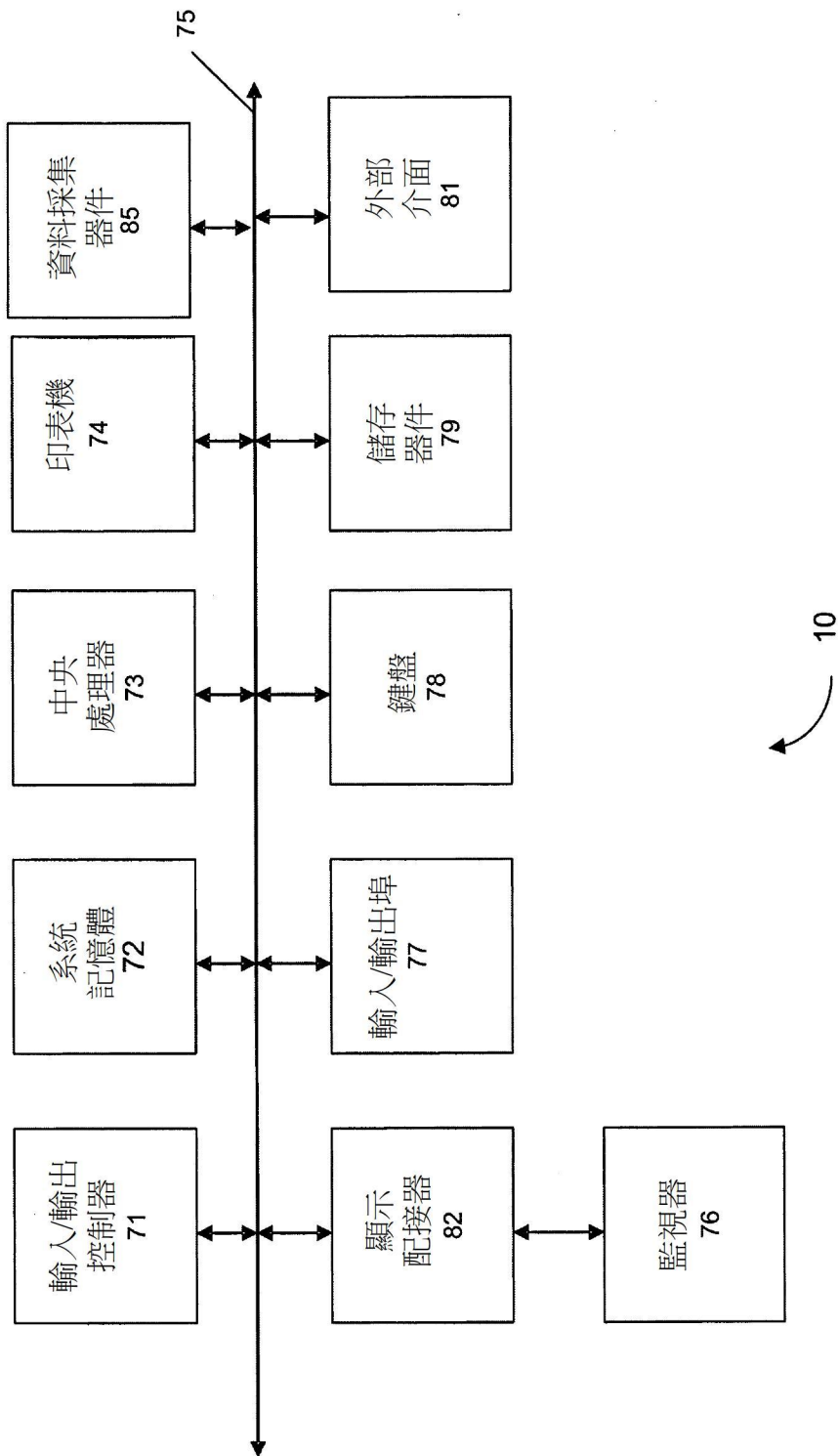


【圖9】

1000



【圖10】



【圖11】