201720932

(19)中華民國智慧財產局

(12)發明說明書公開本 (11)公開編號：TW 201720932 A

(43)公開日： 中華民國 106 (2017) 年 06 月 16 日

(21)申請案號：105130685 (22)申請日： 中華民國 105 (2016) 年 09 月 22 日

(51)Int. Cl.： *C12Q1/68 (2006.01)*

(30)優先權：2015/09/22 美國 62/222,157

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

(74)代理人：陳長文

申請實體審查：無 申請專利範圍項數：59 項 圖式數：19 共 92 頁

(54)名稱

藉由母體血漿ＤＮＡ之淺深度測序以準確定量胎兒ＤＮＡ含量

ACCURATE QUANTIFICATION OF FETAL DNA FRACTION BY SHALLOW-DEPTH SEQUENCING OF MATERNAL PLASMA DNA

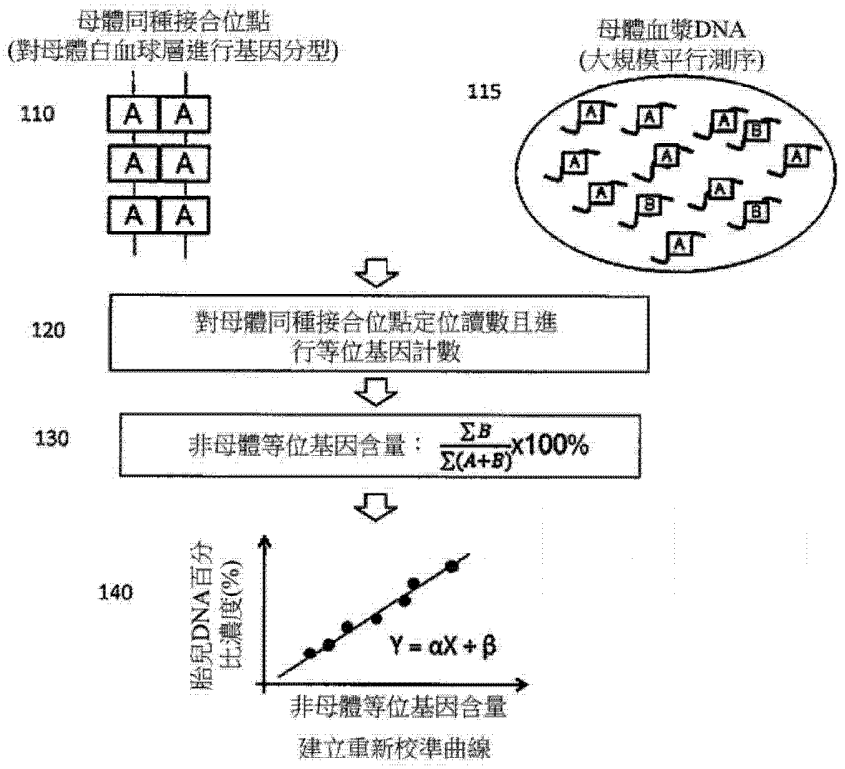
(57)摘要

本發明之實施例提供在不使用父本或胎兒基因型下用於推導母體血漿中胎兒 DNA 含量之方法、系統及設備。母體基因型資訊可自僅僅母體 DNA 樣品獲得或可自具有母體與胎兒 DNA 分子之生物樣品之淺深度測序來假設。因為測序可在淺深度下，所以基因座可具有僅僅少量讀數且可能無法展現非母體等位基因，即使存在非母體等位基因。然而，表徵所測序之非母體等位基因的歸一化參數可用於提供胎兒 DNA 含量之準確估值，即使非母體等位基因之量有誤差。本文所述之方法可無需特定區域之高深度測序或富集。因此，此等方法可整合至廣泛使用之無創產前測試及其他診斷學中。

Embodiments of the present invention provide methods, systems, and apparatus for deducing the fetal DNA fraction in maternal plasma without using paternal or fetal genotypes. Maternal genotype information may be obtained from a maternal-only DNA sample or may be assumed from shallow-depth sequencing of a biological sample having both maternal and fetal DNA molecules. Because sequencing may be at shallow depths, a locus may have only few reads and may fail to exhibit a non-maternal allele even if a non-maternal allele is present. However, normalized parameters that characterize non-maternal alleles sequenced can be used to provide an accurate estimate of the fetal DNA fraction, even if the amount of non-maternal alleles is in error. Methods described herein may not need high-depth sequencing or enrichment of specific regions. As a result, these methods can be integrated into widely used non-invasive prenatal testing and other diagnostics.

指定代表圖：

TW 201720932 A



【圖1】

【發明說明書】

【中文發明名稱】

藉由母體血漿ＤＮＡ之淺深度測序以準確定量胎兒DNA含量

【英文發明名稱】

ACCURATE QUANTIFICATION OF FETAL DNA FRACTION BY SHALLOW-DEPTH SEQUENCING OF MATERNAL PLASMA DNA

【技術領域】

【先前技術】

母體血漿中循環之不含細胞之胎兒DNA的發現[Lo YM等人(1997), *Lancet,* 350: 485-487]已催生一系列用於無創產前診斷之新方法，包括胎兒RhD血型基因分型[Lo YM等人(1998), *N Engl J Med,* 339: 1734-1738；Finning K等人(2008), *BMJ,* 336: 816-818]、性連鎖遺傳病之胎兒性別確定[Costa JM, Benachi A, Gautier E (2002), *N Engl J Med*, 346: 1502]、染色體非整倍性偵測[Lo YM等人(2007), *Proc Natl Acad Sci U S A*. 104: 13116-13121；Chiu RW等人(2008), *Proc Natl Acad Sci U S A* 105: 20458-20463；Chiu RW, Cantor CR, Lo YM (2009), *Trends Genet*, 25: 324-331；Fan HC等人(2008), *Proc Natl Acad Sci U S A*, 105: 16266-16271；Chiu RW等人(2011), *BMJ* 342: c7401；Yu SC等人(2014), *Proc Natl Acad Sci U S A*, 111: 8583-8588]及單基因性疾病之偵測[Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91；Lam KW等人(2012), *Clin Chem.*；New MI等人(2014), *The Journal of Clinical Endocrinology & Metabolism,* 99: E1022-E1030；Yoo S-K等人(2015), *Clinical Chemistry*；

Ma D等人(2014), *Gene,* 544: 252-258；Tsui NB等人(2011), *Blood,* 117: 3684-3691]。

在以上提及之應用中，準確推導胎兒DNA含量(亦稱為胎兒DNA百分比濃度或胎兒DNA百分比)對於在統計學上準確解釋經由使用血漿DNA之無創產前診斷之結果而言具有重要意義，尤其在視此參數而定之統計模型用於偵測染色體非整倍體[Sparks AB等人, (2012), *Am J Obstet Gynecol*, 206: 319 e311-319]及確定單基因性疾病遺傳性[Lo YM等人(2007), *Proc Natl Acad Sci U S A*. 104: 13116-13121；Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91；Lam KW等人(2012), *Clin Chem.*；New MI等人(2014), *The Journal of Clinical Endocrinology & Metabolism,* 99: E1022-E1030；Yoo S-K等人(2015), *Clinical Chemistry*；Tsui NB等人(2011), *Blood,* 117: 3684-3691]時。舉例而言，胎兒DNA含量為用於準確確定哪種母體單倍型傳遞至胎兒之相對單倍型劑量分析(Relative Haplotype Dosage Analysis，RHDO)中的重要參數[Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91；Lam KW等人(2012), *Clin Chem.*；New MI等人(2014), *The Journal of Clinical Endocrinology & Metabolism,* 99: E1022-E1030]。在此診斷方法中，基礎原理為傳遞至胎兒之母體單倍型之相對劑量將比未傳遞之單倍型略微過度表示，且胎兒DNA含量用於確定過度表示之統計顯著性。

迄今，已研發許多方法來估計懷孕女性之母體血漿中胎兒DNA百分比濃度。舉例而言，來源於Y染色體之特定信號用於推導攜帶男性胎兒之孕婦中的胎兒DNA含量[Chiu RW等人(2011), *BMJ* 342: c7401；Lo YM等人(1998), *Am J Hum Genet*, 62: 768-775；Lun FM等人(2008), Clin Chem 54: 1664-1672；Hudecova I等人(2014), *Plos One*, 9: e88484]。然而，基於Y

染色體特定信號之方法不適用於攜帶女性胎兒之孕婦。一種替代方法為使用單核苷酸多型現象(single nucleotide polymorphism，SNP)，從而計算胎兒特異性等位基因與共享等位基因之比率以推斷胎兒DNA含量。在此方法中，必須已知基因型資訊，且應符合以下情形之一：(a)母親為純合接合，而胎兒為雜合接合；(b)父本與母體基因型均為純合接合，但具有不同等位基因[Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91；Liao GJ等人(2011), *Clin Chem*, 57: 92-101]。然而，一方面，在實際臨床情境中，在無創產前診斷期間，胎兒基因型無法預先獲得。另一方面，關於全世界父本偏差之流行病學研究提出，父本偏差之發生率可高達30% [Bellis MA, Hughes K, Hughes S, Ashton JR (2005) *J Epidemiol Community Health,* 59: 749-754]，其限制父本DNA用於數據分析之可利用性。即使研發不依賴親本基因型之演算法，藉由跨越不同SNP位點，利用母體血漿DNA之高深度測序(例如母體血漿DNA之靶向測序)來排除額外基因型資訊之先決條件[Jiang P等人(2012), *Bioinformatics,* 28: 2883-2890]，亦需要其他工作，例如經由使用基於雜交或擴增子之富集系統，捕捉一組所關注區域[Sparks AB等人, (2012), *Am J Obstet Gynecol*, 206: 319 e311-319；Liao GJ等人(2011), *Clin Chem*, 57: 92-101]。

除SNP依賴性方法之外，亦研究不依賴於SNP之方法。舉例而言，母體血漿DNA之片段尺寸可用於估計胎兒DNA含量[Yu SC等人(2014), *Proc Natl Acad Sci U S A*, 111: 8583-8588；Kim SK等人(2015), Prenatal diagnosis: n/a-n/a]，因為胎兒來源之DNA一般比母體來源之DNA短[Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91]。然而，一些其他病狀將影響基於尺寸之胎兒DNA含量估計的準確性，例如全身性紅斑狼瘡[Chan

RW等人(2014), *Proc Natl Acad Sci U S A*, 111: E5302-5311]。作為替代方案，證實胎兒特定之表觀遺傳改變，諸如甲基化RASSF1A及未甲基化SERPINB5序列，為與基因型資訊無關的胎兒DNA含量預測之胎兒標記物[Chan KC等人(2006), *Clin Chem*, 52: 2211-2218；Chim SS等人(2005), *Proc Natl Acad Sci U S A,* 102: 14753-14758]。然而，用於定量此等表觀遺傳標記物之分析程序涉及亞硫酸氫鹽轉化或用甲基化敏感性限制酶消化，且因此可能影響此等方法之精確度。

因此，需要自母體血漿提供胎兒DNA含量資訊之新技術。

【發明內容】

本發明之實施例提供用於定量母體血漿中胎兒DNA含量之方法、系統及設備。胎兒DNA含量可在不特別測定父本或胎兒基因型下測定。可測定另一參數，且校準曲線可用於測定實際胎兒DNA含量。舉例而言，可確定具有標稱上鑑別為非母體等位基因之等位基因之讀數的量與具有標稱上鑑別為母體等位基因之等位基因之讀數的量之比率。作為另一實例，如自另一資料集確定，可確定展現標稱非母體等位基因之基因座之量與純合接合母體基因座之量的比率。亦可使用讀數尺寸之差異。基因座(位點)可限於群體中已知之雜合接合基因座。

母體基因型資訊可自僅僅母體之DNA樣品獲得，或可自具有母體與胎兒DNA分子之生物樣品之測序(例如在淺深度下)假設。實際或假設之母體基因型資訊可與來自生物樣品之DNA分子之測序組合。即使可能不確定母親在特定基因座是否為純合接合或胎兒是否為雜合接合，實施例亦可使用此類位點之讀數確定另一參數，此為與先前技術之差異。任何誤差展示一致，且因此，藉由校準曲線抵消，校準曲線可使用另一技術產生一次以確

定胎兒DNA含量。

因為測序可在淺深度下，所以基因座可具有僅僅少量讀數且可能無法展現非母體等位基因，即使存在非母體等位基因。然而，表徵所測序之非母體等位基因之歸一化參數可用於提供胎兒DNA含量之準確估計，即使在基因座或所有基因座之非母體等位基因之量不代表胎兒DNA含量。此等歸一化參數可包括非母體等位基因下序列讀數之量或具有非母體等位基因之基因座之量。本文所述之方法可無需特定區域之高深度測序或富集。因此，此等方法可整合於廣泛使用之無創產前測試及其他診斷學中。

一些實施例係針對與本文所描述之方法相關的系統及電腦可讀媒體。

更好地瞭解本發明實施例之性質及優點。

【圖式簡單說明】

圖1為根據本發明之實施例，使用母體基因型量測胎兒DNA百分比濃度之示意圖。

圖2A展示根據本發明之實施例，自來自第一資料集之訓練資料集構築之實際胎兒DNA含量及非母體等位基因含量的線性回歸模型。

圖2B展示根據本發明之實施例，使用獨立資料集之圖2A中回歸模型之驗證。

圖3A展示根據本發明之實施例，自來自第二資料集之訓練資料集構築之實際胎兒DNA含量及非母體等位基因含量的線性回歸模型。

圖3B展示根據本發明之實施例，使用獨立資料集之圖3A中回歸模型之驗證。

圖4A展示根據本發明之實施例，第一資料集之實際胎兒DNA含量與估計之胎兒DNA含量之間的偏差。

圖4B展示根據本發明之實施例，第二資料集之實際胎兒DNA含量與估計之胎兒DNA含量之間的偏差。

圖5展示根據本發明之實施例，相對預測誤差與實際胎兒DNA含量之間的關係之圖。

圖6A、**6B**、**6C**及**6D**展示根據本發明之實施例，在各種測序深度下胎兒DNA含量預測之準確性。

圖7展示一種根據本發明之實施例，使用讀數之量量測懷有胎兒之女性之生物樣品中胎兒DNA含量的方法。

圖8展示根據本發明之實施例，在不獲得母體基因型、父本基因型或僅僅含有母體DNA分子之生物樣品下量測胎兒DNA含量之圖示。

圖9展示一種根據本發明之實施例，使用基因座之量量測懷有胎兒之女性之生物樣品中胎兒DNA含量的方法。

圖10A展示根據本發明之實施例，來自胎兒DNA含量及表觀等位基因偏差(apparent allelic discrepancy，AAD)值之線性回歸模型的校準曲線。

圖10B展示根據本發明之實施例，基於胎兒DNA含量及短DNA分子之比例的線性回歸之圖。

圖10C展示根據本發明之實施例，藉由AAD值確定之胎兒DNA含量針對基於得自Y染色體之讀數之比例的胎兒DNA含量的圖。

圖11展示一種根據本發明之實施例，量測懷有胎兒之女性之生物樣品中胎兒DNA含量的方法。

圖12A、**12B**、**12C**及**12D**說明根據本發明之實施例，母體及非母體等位基因之DNA分子尺寸之間的關係。

圖13展示一種根據本發明之實施例，使用尺寸值量測懷有胎兒之女性

之生物樣品中胎兒DNA含量的方法。

圖14為根據本發明之實施例，針對六組不同雙胞胎計算之胎兒DNA含量表。

圖15為根據本發明之實施例，胎兒DNA含量與展示尺寸差異之基因座之間的關係之圖。

圖16展示一種根據本發明之實施例，使用具有一定尺寸DNA分子之基因座之量，量測懷有胎兒之女性之生物樣品中胎兒DNA含量之方法。

圖17展示可與根據本發明之實施例的系統及方法一起使用的一個實例電腦系統的方塊圖。

圖18展示根據本發明之實施例之測序系統。

圖19展示根據本發明之實施例之電腦系統。

【實施方式】

相關申請之交叉引用

本申請案主張2015年9月22日申請之美國申請案62/222,157之優先權益，該案之內容以引用的方式併入本文中以達成所有目的。

術語

如本文所用，術語「*基因座*(locus)」或其複數形式「*基因座*(loci)」為跨基因組具有變化的任何長度之核苷酸(或鹼基對)的位置或位址。「序列讀數(sequence read)」係指在核酸分子之任何部分或全部中所測序的核苷酸串。舉例而言，序列讀數可為自核酸片段測序之短核苷酸串(例如約20-150)、位於核酸片段之一或兩個末端的短核苷酸串，或存在於生物樣品中之整個核酸片段之測序。序列讀數可以多種方式獲得，例如使用測序技術或使用探針，例如雜交陣列或捕捉探針，或擴增技術，諸如聚合酶鏈反

應(polymerase chain reaction，PCR)或使用單一引子之線性擴增或等溫擴增。

「*生物樣品*」係指自個體(例如人類，諸如孕婦、癌症患者或疑似患有癌症者、器官移植接受者，或疑似具有涉及器官(例如心肌梗塞中之心臟或中風中之腦或貧血中之造血系統)之疾病過程的個體)獲得且含有所關注之一或多種核酸分子的任何樣品。生物樣品可為體液，諸如血液、血漿、血清、尿液、陰道流體、來自水囊腫(例如睪丸)之流體、陰道沖洗液、胸膜液、腹水、腦脊髓液、唾液、汗液、淚液、痰、支氣管肺泡灌洗液、乳頭排放液、來自身體不同部分(例如甲狀腺、乳房)之抽吸流體等。亦可使用糞便樣品。在各種實施例中，富集不含細胞之DNA之生物樣品(例如經由離心方案獲得之血漿樣品)中的大部分DNA可不含細胞，例如大於50%、60%、70%、80%、90%、95%或99%的DNA可不含細胞。離心方案可包括例如3,000 g×10分鐘，獲得流體部分，且在例如30,000 g下另外再離心10分鐘以移除殘餘細胞。樣品中不含細胞之DNA可來源於各種組織之細胞，且因此樣品可包括不含細胞之DNA之混合物。

「*核酸*」可指去氧核糖核苷酸或核糖核苷酸及其呈單股或雙股形式之聚合物。該術語可涵蓋含有已知核苷酸類似物或經修飾主鏈殘基或鍵聯之核酸，其為合成的、天然產生的及非天然產生的，具有與參考核酸類似之結合性質，且以類似於參考核苷酸之方式代謝。此類類似物之實例可包括(但不限於)硫代磷酸酯、胺基磷酸酯、膦酸甲酯、對掌性膦酸甲酯、2-O-甲基核糖核苷酸、肽核酸(PNA)。

除非另外指明，否則特定核酸序列亦含蓄地涵蓋其經保守修飾之變異體(例如簡併密碼子取代)及互補序列，以及經明確指示之序列。特定言之，簡併密碼子取代可藉由產生其中一或多個所選(或所有)密碼子之第三位置

經混合鹼基及/或去氧肌苷殘基取代的序列來達成(Batzer等人, *Nucleic Acid Res.* 19:5081 (1991)；Ohtsuka等人, *J. Biol. Chem.* 260:2605-2608 (1985)；Rossolini等人, *Mol. Cell. Probes* 8:91-98 (1994))。術語核酸可與基因、cDNA、mRNA、寡核苷酸及聚核苷酸互換使用。

除非上下文另外清楚地指出，否則術語「*核苷酸*」除指代天然存在之核糖核苷酸或去氧核糖核苷酸單體之外，應瞭解亦指其關於核苷酸正使用之特定情況(例如，與互補鹼基雜交)在功能上同等的相關結構變異體，包括衍生物及類似物。

「*序列讀數*」係指在核酸分子之任何部分或全部中所測序的核苷酸串。舉例而言，序列讀數可為存在於生物樣品中之整個核酸片段。序列讀數可自單分子測序獲得。

「*分類*」係指與樣品之特定性質相關的任何數字或其他特徵。舉例而言，符號「+」(或詞語「陽性」)可表示樣品歸類為具有缺失或擴增。分類可為二元(例如陽性或陰性)或具有更多分類層級(例如量表1至10或0至1)。術語「*截止值*」及「*閾值*」係指用於操作之預定數字。舉例而言，截止尺寸可指排除超過其之片段之尺寸。閾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。

術語「*尺寸型態*」通常係指生物樣本中DNA片段之尺寸。尺寸型態可為提供多種尺寸之DNA片段之量的分佈的直方圖。各種統計參數(亦稱作尺寸參數或僅參數)可用於區分一個尺寸型態與另一尺寸型態。一個參數為相對於所有DNA片段或相對於另一尺寸或範圍之DNA片段的特定尺寸或尺寸範圍之DNA片段的百分比。

詳細說明

使用母體血漿DNA大規模平行測序之無創產前測試(Noninvasive

prenatal testing，NIPT)日益視為現代產前診斷中之必需組分且已迅速地用於全世界臨床使用中。為確保此種無創產前診斷之精確解釋，胎兒DNA含量變成待量測之關鍵參數。雖然已研發各種方法估計此參數，但很少有通用及廣泛適用之方法。

一些實施例允許使用實際或假設之母體基因型及母體血漿之隨機大規模平行測序準確估計胎兒DNA含量。胎兒DNA含量可與表徵生物樣品中非母體物質之量的參數相關。非母體物質之量可計算為非母體之序列讀數之片段，或具有非母體等位基因之基因座之比例。在此等計算中之任一者中，參數可能不準確表示實際胎兒DNA含量。測序可在淺深度下進行，從而並非所有存在之非母體等位基因可進行測序。另外，序列讀數之非母體片段之計算可包括在不具有非母體等位基因之位點之讀數。接著在非母體片段之計算中包括此等序列讀數將包括可能純合接合位點，該等位點通常不用於習知方法中來確定胎兒DNA含量。類似地，計算具有非母體等位基因之基因座之比例可包括使用許多可能純合接合基因座，通常認為該等基因座在計算胎兒DNA含量之習知方法中不重要。

然而，發現胎兒DNA含量與懷孕女性之血漿中來源於母體純合接合基因座之非母體等位基因之片段相關，即使當非母體等位基因之片段未準確說明所有非母體等位基因時。另外，發現胎兒DNA含量與具有非母體等位基因之基因座的比例相關，即使在序列讀數未展現具有非母體等位基因之所有基因座時。此等方法用實驗數據證實。使用淺深度測序，方法可比習知方法更有效及低成本。另外，此等方法不依賴於胎兒中父本基因型或特定基因性狀，且因此該等方法可廣泛應用於任何懷孕女性。此等方法亦可進一步增強無創產前測試之臨床解釋。

I. 使用序列讀數之量分析DNA

僅僅母體DNA可進行測序且與含有母體DNA與胎兒DNA之樣品中的DNA相比，以估計胎兒DNA含量。僅僅母體DNA可進行測序以鑑別純合接合位點。接著含有母體DNA與胎兒DNA之混合物的樣品(例如母體血漿或血清)可進行測序。在混合物中，一些鑑別之純合接合位點可具有非母體等位基因之序列讀數，而其他鑑別之純合接合位點可具有僅僅與母體等位基因一致之等位基因之序列讀數。非母體等位基因及與母體等位基因一致之等位基因的此等讀數可用於計算非母體片段。在分母中可包括具有僅僅與母體等位基因一致之等位基因之位點的此非母體片段可能不等同於實際胎兒DNA含量。然而，此非母體片段可與胎兒DNA含量相關。胎兒DNA含量愈高，非母體片段可愈高。胎兒DNA含量及非母體片段之校準曲線可用於使胎兒DNA含量之估計值與針對樣品計算之非母體片段相關。

然而，含有母體DNA與胎兒DNA之樣品可在淺深度下測序，其中基因座可能僅僅具有一個或兩個讀數。即使母體DNA與胎兒DNA之樣品中讀數展示與母體等位基因相同之等位基因，位點亦不可在高統計學信賴度下確定為在胎兒中純合接合，因為胎兒非母體等位基因可已存在，但剛好未展現在少數讀數中。接著淺深度測序可低估胎兒DNA中非母體等位基因之實際數目。

即使非母體等位基因之片段可能不為非母體等位基因之實際片段，此片段亦可用於校準曲線以獲得準確胎兒DNA含量。發現胎兒DNA含量與非母體片段相關，即使該片段低估或者不包括非母體等位基因之準確計數。較高胎兒DNA含量增加測序之非母體等位基因之可能性，從而增加非母體片段。因此，甚至在淺深度下，非母體片段與胎兒DNA含量之間的關係亦

可在校準曲線中表示且用於估計胎兒DNA含量。

A.　非母體片段及校準曲線

含有母體DNA及胎兒DNA之樣品的非母體片段為具有非母體等位基因之讀數之第一量與讀數之第二量的比率。讀數之兩種量可在母體基因組中之某些位點，包括已知具有雜合接合高度可能性之位點(亦即具有SNP之位點)。讀數之第二量可包括來自具有母體等位基因之DNA混合物之樣品的讀數。在一些實施例中，第二量可為在該等位點之讀數之總量，其中總量為第一量與具有母體等位基因之讀數的總和。

此非母體片段可能不等於生物樣品中存在之非母體等位基因之實際片段。實際上，非母體片段反映作為母體血漿中非母體等位基因之測序讀數。因此，非母體片段可視測序誤差、基因分型誤差、母親為純合接合且胎兒為雜合接合之位點(資訊性SNP位點)的基礎數目及胎兒DNA含量而定。結果展示測序誤差、基因分型誤差及資訊性位點之基礎數目相對恆定。因此，胎兒DNA含量可自非母體等位基因含量確定。

圖1展示使用非母體片段確定片段胎兒DNA。鑑別母體DNA中純合接合位點。母體血漿進行測序且在鑑別之純合接合位點計數等位基因之讀數。自此等位點之等位基因之讀數的總和計算非母體片段，即使在一些位點，無非母體等位基因進行測序。接著所得非母體片段可與先前量測之胎兒DNA百分比濃度及先前計算之非母體片段的校準曲線相比。可獲得估計之胎兒DNA百分比濃度。

在部分110，例如藉由使用基於微陣列之基因分型技術分析膚色血球層或頰黏膜拭子樣品，自母體組織獲得母體基因型。在其他實施例中，亦可使用包含胎兒與母體DNA之混合物的樣品進行母體基因型分析。

部分110展示母體基因型中純合接合位點。各位點具有兩個A等位基因，展示為盒。A等位基因可為或可不為具有A核苷酸之SNP之結果。雖然部分110展示母體基因組中無雜合接合位點，但雜合接合位點可在純合接合位點之間。純合接合位點可僅僅限於已知具有單核苷酸多型現象(SNP)之位點，該等位點可在諸如dbSNP或HapMap之資料庫中鑑別。呈現純合接合之母體位點可自基因分型資訊鑑別且與參考基因組比對。可使用任何適合之基因分型技術，例如使用測序(其可包括與參考基因組比對)、靶向測序、基於擴增子之測序、質譜分析、微滴式數位PCR、雜交陣列或微陣列進行基因分型。

使用之純合接合位點之數目可視使用之微陣列平台而定。舉例而言，對於Affymetrix，存在約700,000個純合接合位點且對於BeadChip，存在約2百萬個。因此，存在足夠位點用於實施例來集中於SNP位點而非整個基因組中之任何位點，不過後者亦為可能。

在部分115，母體血漿DNA進行測序。母體血漿包括許多可包括來自鑑別之純合接合位點之等位基因的DNA片段。部分115展示具有A等位基因(存在於純合接合位點之母體等位基因)之片段，及具有B等位基因(不存在於純合接合位點之非母體等位基因)之片段。母體血漿DNA可用大規模平行測序進行測序。母體血漿DNA可在淺深度下測序。舉例而言，測序讀數之數目可小於單倍體人類基因組之0.1x、0.2x、0.3x、0.4x、0.5x、0.8x、1x、1.5x、2x、3x、4x、5x及10x覆蓋範圍。讀數之數目可小於或等於5000萬個讀數，包括小於或等於3000萬個讀數、2000萬個讀數、1500萬個讀數、小於或等於1000萬個讀數、小於或等於800萬個讀數、小於或等於500萬個讀數、小於或等於400萬個讀數、小於或等於200萬個讀數或小於或等於100

萬個讀數。在部分110中獲得之序列讀數亦可在淺深度下確定。因此，純合接合之基因型可能不準確(亦即女性可在鑑別為純合接合之位點之一具有B)，但結果展示此類不準確在樣品中為一致的，從而允許校準曲線以所需準確性提供胎兒DNA含量。

在部分120，來自母體血漿(或具有母體與胎兒DNA之混合物的其他樣品)之序列讀數定位至參考基因組。可僅僅對母體基因組中鑑別之純合接合位點進行定位。與純合接合位點之比對辨別通常來自父本貢獻之B非母體等位基因，但可涉及測序誤差、重新突變及本文中提及之其他實例。如上所述，在部分110之基因分型在淺深度下進行的情況下，B等位基因亦可來自母親。

接著，針對A母體等位基因與B非母體等位基因，求得來自母體血漿之序列讀數總數。求得在鑑別之純合接合位點具有B非母體等位基因之序列讀數的總數。求得在鑑別之純合接合位點具有A母體等位基因之序列讀數的總數，即使在特定位點無B非母體等位基因進行測序。

在部分130，確定非母體等位基因含量。為計算非母體等位基因含量，自部分120中之讀數，獲得在純合接合位點具有B非母體等位基因之序列讀數之總數ΣB。自部分120中之讀數，獲得具有B非母體等位基因或A母體等位基因之序列讀數之總數Σ(A+B)。非母體等位基因含量計算為具有B非母體等位基因之序列讀數之總數與具有A母體等位基因或B非母體等位基因之序列讀數之總數的比率，且比率轉化成百分比：

$$\frac{\sum B}{\sum(A+B)} \times 100 \quad (1)$$ 。

可使用其他相關片段或百分比。舉例而言，具有B非母體等位基因之序列讀數之總數可除以僅僅具有A母體等位基因之序列讀數之總數的總

和。亦可使用本文所述之任何片段之倒數。

實際上，非母體等位基因之片段由胎兒DNA含量以及測序及基因分型誤差控制。限制條件為來源於基因分型及測序平台之誤差為在某一情形下相對恆定之系統性誤差，胎兒DNA百分比濃度與母體血漿中量測之非母體等位基因之片段成比例。胎兒DNA含量可藉由分析非母體等位基因之片段來預測。

在部分140，展示用於自非母體等位基因含量獲得胎兒DNA百分比濃度之校準曲線。校準曲線可具有各種函數形式，例如線性、二次或任何多項式。部分140展示線性校準曲線，其中X為非母體片段，自等式(1)計算，Y為胎兒DNA百分比濃度，α為直線斜率，且β為直線之y截距。

為建立校準曲線，實施例可使用一系列具有已知之胎兒DNA含量(例如自Y染色體估計、基於資訊性SNP位點等)之樣品。對於具有已知之胎兒DNA含量之各樣品，量測非母體等位基因含量。可確定已知之胎兒DNA含量值與量測之非母體片段的函數擬合且用作校準曲線。此等樣品可稱為校準樣品。

在各種實施例中，校準值可對應於自校準樣品確定之校準資料點之校準值或自其確定之任何校準值，例如模擬校準資料點之校準函數之校準值。一或多種校準樣品可包括或可不包括用於確定較佳終止位點之任何其他樣品。

對於一或多種校準樣品中之每一者而言，可使用例如組織特異性等位基因量測第一組織類型之對應比例份額。對應相對豐度可使用在對應於第一組基因組位置之複數個窗內終止之不含細胞之DNA分子的對應數目來確定。所量測比例份額及相對豐度可提供校準資料點。一或多個校準資料

點可為形成模擬複數個校準資料點之校準函數的複數個校準資料點。校準值使用之其他細節可見於美國專利公開案2013/0237431中。

在確定非母體等位基因含量時，在位點具有非母體等位基因之每個讀數均可計數，即使不知胎兒是否真地具有非母體等位基因或是否其為誤差。在一些實施方案中，在使用位點前可不需要非母體等位基因之最少數目，此將另外用作確定等位基因不為誤差之測試。另外，不具有非母體等位基因之序列讀數的位點仍可用於確定非母體等位基因含量。舉例而言，即使母體血漿DNA中之一些位點僅僅具有母體等位基因之序列讀數，母體等位基因之此等讀數仍可呈現在等式(1)中之分母中以計算非母體等位基因含量。接著所得非母體等位基因含量可能不反映實際非母體等位基因含量，因為計算包括可能不具有非母體等位基因之位點。

為確保更大準確性，實施例可濾出針對dbSNP資料庫中未標註之位點攜帶等位基因的讀數，例如假設所有使用之SNP為雙等位基因。舉例而言，SNP位點在dbSNP資料庫中標註為A/C。將濾出血漿中看到攜帶「G」之讀數，但位點仍可用作所分析之其他讀數的參考。此可降低測序誤差影響。此外，可濾出在未標註為SNP位點之位點的所有讀數。

B. 校準曲線之訓練及驗證

母體血漿樣品用於驗證非母體等位基因含量評估胎兒DNA含量之用途。一些樣品用作訓練資料集，以產生實際胎兒DNA含量相對於非母體等位基因含量之校準曲線。對於剩餘樣品，確定各樣品之非母體等位基因含量，且接著基於自初始樣品產生之校準曲線，估計胎兒DNA含量。接著針對剩餘樣品估計之胎兒DNA含量與實際胎兒DNA含量相比，以檢驗使用非母體等位基因含量之準確性。

1. 資料集

使用兩個資料集測試關於胎兒DNA濃度是否可自母體血漿中量測之非母體等位基因之片段確定的假設。對於第一資料集，存在總共35個樣品，其在Affymetrix基因分型微陣列(Affymetrix全基因組人類SNP陣列6.0系統)上進行基因分型且用Genome Analyzer IIx (Ilumina)上雙端模式之36次循環來測序，分別如[Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91]及[Yu SC等人(2014), *Proc Natl Acad Sci U S A*, 111: 8583-8588]中所述。平均在Affymetrix基因分型平台上詢問之906,600個SNP中獲得671,206 (範圍635,378-682,501)個純合接合位點。同時，在使用SOPA2將雙端測序讀數定位至參考人類基因組後[Yu SC等人(2014), *Proc Natl Acad Sci U S A*, 111: 8583-8588；Li RQ等人(2009), *Bioinformatics*, 25: 1966-1967]，對於樣品，獲得中值為12,961,498 (範圍7,728,645-23,454,296)之可比對及非複製讀數用於隨後分析。幾乎1300萬個讀數之中值對應於大約0.3x覆蓋範圍。

第二資料集具有比第一資料集更高數目之讀數及樣品。對於第二資料集，存在總共70個樣品，其在BeadChip陣列(Illumina)上進行基因分型且在HiSeq 2000測序儀(Illumina)上測序(50 bp×2)[Stephanie C等人(2013), *Clinical Chemistry*]。平均在BeadChip陣列(Illumina)上詢問之2,351,072個SNP中獲得1,940,577 (範圍1,925,282-1,949,532)個純合接合基因座。比對後，對於樣品，獲得中值為69,959,574 (範圍26,036,386-94,089,417)之可比對及非複製讀數。幾乎7000萬個讀數對應於約2.3x之覆蓋範圍。為評估胎兒DNA含量預測之效能，將估計之胎兒片段與經由使用胎兒基因型作為標準確定之胎兒DNA含量(稱為實際胎兒DNA含量)比較。

2. 非母體等位基因含量計算

使用等式(1)，計算各樣品之非母體等位基因含量。對於自對僅僅母體樣品進行基因分型所鑑別之純合接合位點，計數來自對應母體血漿樣品之讀數之數目。在鑑別之純合接合位點具有非母體等位基因之讀數數目的總和除以在純合接合位點之讀數總數(亦即具有非母體等位基因或母體等位基因之讀數)，且接著轉化成百分比。

3. 胎兒DNA百分比濃度估計

為證實胎兒DNA百分比濃度與母體血漿中非母體等位基因之片段成比例，將各資料集隨機劃分，其中一些樣品在訓練集中且剩餘樣品在獨立驗證集中。線性回歸用於藉由分別分析第一資料集及第二資料集之訓練集中12個及23個樣品，模擬母體血漿中實際胎兒DNA含量(因變數Y)與非母體等位基因之片段(自變數X，藉由等式(1)計算)之間的關係。根據下式，藉由分析與其中母親為純合接合且胎兒為雜合接合之SNP重疊之讀數，推斷實際胎兒DNA含量(F)[Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91]。

$$F = \frac{p}{p+q} \quad (2)$$

其中p為胎兒特異性等位基因之測序讀數之數目，且q為共享等位基因之讀數計數。等式(2)不同於等式(1)，因為等式(2)包括僅僅來自其中母親為純合接合且胎兒為雜合接合之位點的讀數，而等式(1)亦可包括來自其中母親與胎兒均呈現純合接合之位點的讀數。在其他實施例中，F可以2來縮放，以對應於所有胎兒DNA之總胎兒片段。亦可使用其他比率，例如p/q。

因此，假設F為實際胎兒DNA含量且自其中母親為純合接合且胎兒為雜合接合之位點估計。雜合接合性可藉由在對應位點對胎盤組織進行基因分型來確定。用於確定實際胎兒DNA含量之樣品用於展示使用本發明之實

施例推斷之F準確。

4. 結果

圖2A展示使用第一資料集之訓練資料集構築的線性模型(Y = 11.9X - 1.4)。使用先前獲得之胎兒基因型確定的實際胎兒DNA含量展示在y軸上，且非母體等位基因含量展示在x軸上。調整之R^2為0.97 (p值＜0.0001)。

圖2B展示估計之胎兒DNA含量高度類似於第一資料集中之實際胎兒DNA含量。使用來自圖2A之線性模型，估計之胎兒DNA含量展示在y軸上。使用先前獲得之胎兒基因型確定的實際胎兒DNA含量展示在y軸上。線性回歸與資料擬合，其中所得擬合具有在0.99下的調整之R^2 (p值<0.0001)。

圖3A展示線性模型(Y = 18.9X - 6.6，調整之R^2為0.99且p值 < 0.0001)自第二資料集之訓練集中的24個樣品構築。使用先前獲得之胎兒基因型確定的實際胎兒DNA含量展示在y軸上，且非母體等位基因含量展示在x軸上。

圖3B展示估計之胎兒DNA含量高度類似於第二資料集中之實際胎兒DNA含量。使用來自圖2B之線性模型，估計之胎兒DNA含量展示在y軸上。使用先前獲得之胎兒基因型確定的實際胎兒DNA含量展示在y軸上。線性回歸與資料擬合，其中所得線性擬合具有0.99的調整之R^2 (p值<0.0001)。

圖2B及圖3B中之驗證集展示基於非母體等位基因含量及實際胎兒DNA含量高度之校準曲線的估計之胎兒DNA含量與實際胎兒DNA含量高度相關。圖2B與圖3B中驗證集之線性擬合具有0.99之R^2 (p值<0.0001)。高R^2值指示技術精確。圖2B及圖3B中之點亦接近y=x線，此將表明實際胎兒

DNA含量之完美估計。

圖**4A**及**4B**展示與實際胎兒DNA含量之中值偏差。圖4A及圖4B中之x軸為實際胎兒DNA含量。y軸為驗證資料集中各樣品之估計之胎兒DNA含量與實際DNA片段之間的偏差，呈百分比。y軸上之正值對應於大於實際胎兒DNA含量之估計之胎兒DNA含量。y軸上之負值對應於小於實際胎兒DNA含量之估計之胎兒DNA含量。圖4A展示對於第一資料集之驗證集，中值偏差為-0.14%且在-0.7%至1.7%範圍內。圖4B展示對於第二資料集，中值偏差為-0.22%且在-1.5%至0.98%範圍內。兩個資料集之兩個校準曲線之結果的差異可歸因於所使用之不同平台。圖4A及圖4B證實2%之最大偏差及-0.14%與-0.22%之間的中值偏差可能來自使用非母體等位基因含量之估計之胎兒DNA含量。

相對預測誤差(E%)進一步用於量測自驗證資料集構築之模型之準確性，其由下式定義：

$$E\% = \frac{|\hat{F}-F|}{F}\ x\ 100\% \qquad (2)$$

其中$\hat{F}$表示估計之胎兒DNA百分比濃度，且F表示實際胎兒DNA濃度。舉例而言，E%=5%表明若實際胎兒DNA含量為10%，則讀數將在9.95%與10.05%之間(10%±0.05)。發現第一資料集及第二資料集之E%之平均值分別為1.7% (範圍：0.7%-2.9%)及3.8% (範圍：1.3%-14.9%)。

實驗結果證實非母體等位基因含量可用於準確及精確地量測胎兒DNA含量。估計之胎兒DNA含量之準確性及精確度在NIPT測試常常要求之範圍內。

C. 胎兒DNA含量估計之準確性視實際胎兒DNA百分比濃度而定

預測準確性視所分析之實際胎兒DNA含量而定。亦即，胎兒DNA含

量愈高，估計愈準確。第二資料集用於研究實際胎兒DNA含量與相對預測誤差之間的關係，因為存在更多在產後收集的涉及小於5%之胎兒DNA含量的資料點(圖3B)。

圖5展示相對預測誤差相對於實際胎兒DNA濃度之散點圖。呈百分比之相對預測誤差展示在y軸上，且母體血漿中之非母體等位基因含量呈百分比展示在x軸上。散點圖顯示極其明顯之「L」形狀，其中具有高胎兒DNA含量之情況展現低預測誤差，且具有低胎兒DNA含量之情況展示相對較高之預測誤差。E%將接近5%，甚至對於實際胎兒DNA含量，為5% (圖5)。

D. 測序深度與胎兒DNA含量估計準確性之間的關係

為進一步證實測序深度如何影響胎兒DNA含量，對第二資料集執行減少取樣分析，因為第二資料集中之樣品的測序深度比第一資料集高，所以允許取樣分析進行多次。對於20個樣品中之各樣品，隨機選擇不同數目之序列讀數，且自第二資料集中之20個樣品隨機選擇雙端讀數，其中每個樣品100萬個、200萬個、400萬個、600萬個及800萬個。重複胎兒DNA含量預測之以上提及之分析。隨機選擇之序列讀數之數目為100萬個、200萬個、400萬個及800萬個。

圖6A展示100萬個讀數下估計之胎兒DNA含量相對於實際胎兒DNA含量。與資料之線性回歸擬合具有0.9946之R^2及小於0.001之p值。

圖6B展示200萬個讀數下估計之胎兒DNA含量相對於實際胎兒DNA含量。與資料之線性回歸擬合具有0.9918之R^2及小於0.001之p值。

圖6C展示400萬個讀數下之關係。與資料之線性回歸擬合具有0.9927之R^2及小於0.001之p值。

圖6D展示800萬個讀數下之關係。與資料之線性回歸擬合具有0.9924

之R^2及小於0.001之p值。

無論讀數數目如何，R^2值均超過0.99。無論讀數數目如何，p值保持小於0.001。結果展示即使使用100萬個讀數亦可允許吾人進行預測，預測結果與使用200萬個、400萬個、600萬個或800萬個讀數之結果同等優良。

E. 方法適用性

懷孕女性之母體血漿中存在之非母體等位基因之片段可用於估計胎兒DNA含量。高R^2下胎兒DNA百分比濃度與非母體等位基因之片段之間的線性關係存在於母體血漿中，此證實基因分型及測序誤差在一致平台應用於相同資料集之條件下相對恆定。此方法之預測能力已在獨立資料集中證實。使用不同測序或基因型平台之最新校準曲線可提高準確性。第二資料集中提高之R^2可歸因於基因分型及測序系統之準確性提高[Yu SC等人(2014), *Proc Natl Acad Sci U S A*, 111: 8583-8588；Lo YMD等人(2010), *Sci Transl Med*, 2: 61ra91]。然而，在兩個資料集之間觀測到的不同相對誤差(E%)可能歸因於在第二資料集中展示相對較低胎兒DNA含量之更多樣品。

特別地，如減少取樣分析中所證實，測序深度並非影響胎兒DNA含量之準確性的關鍵因素。此方法可準確推廣至具有不同測序深度之樣品。根本原因可能在於母體血漿中展示非母體等位基因之基因座部分將隨著測序深度變化按比例增加或降低。因而，母體血漿中非母體等位基因之片段在不同測序深度中可為恆定值。因此，此方法可消除對測序深度之高度需求且可容易應用於實際臨床操作，因為在無創產前診斷中常規可實現約1500萬之測序讀數之量[Kim SK等人(2015), Prenatal diagnosis: n/a-n/a]。

胎兒DNA預測之準確性應高於兩個先前基於非多型現象之方法[Yu

SC等人(2014), *Proc Natl Acad Sci U S A*, 111: 8583-8588；Kim SK等人(2015), Prenatal diagnosis: n/a -n/a]，因為R^2統計資料之先前值分別為0.83及0.93 [Yu SC等人(2014), *Proc Natl Acad Sci U S A*, 111: 8583-8588；Kim SK等人(2015), Prenatal diagnosis: n/a -n/a]，低於此研究之對應值(第二資料集中R^2值為0.99)。此外，此演算法能夠準確確定5%之低胎兒DNA含量(圖5)。此量測低胎兒DNA含量之能力特別重要，因為相當大部分(約5%)之母體血漿樣品具有小於5%之胎兒DNA百分比濃度[Chiu RW等人(2011), *BMJ* 342: c7401；Palomaki GE等人(2011), *Genet Med,* 13: 913-920]。接著胎兒DNA含量之準確估計可允許在品質控制步驟中精確濾去具有低胎兒DNA含量之樣品[Palomaki GE等人(2011), *Genet Med,* 13: 913-920]。另外，來自與胎兒非整倍性有關之染色體的母體血漿DNA之量的改變程度展示與胎兒DNA含量相關。認為數據在相關性曲線外之樣品更可能為假陽性。評估胎兒DNA含量之實施例可幫助鑑別假結果。另一方面，某些妊娠相關之病狀，諸如先兆子癇及第18對染色體三體症，與母體血漿中擾亂之胎兒DNA含量相關。因此，胎兒DNA含量之更佳估計將允許更靈敏地偵測與擾亂量之胎兒DNA百分比濃度相關之彼等病狀。

因為基於大規模平行測序之臨床診斷日益得到認可且逐漸應用於臨床實踐，所以個人化基因型可用於各個體。因此，母體基因型輔助之胎兒DNA含量估計可容易整合於無創產前診斷中使用之當前現有方法中。使用等位基因之序列讀數的實施例提供一種準確估計胎兒DNA百分比濃度之通用方法。因為缺乏研發在基於隨機測序之無創產前診斷中準確估計胎兒DNA百分比濃度之方法，所以此方法藉由允許更準確地在統計學上解釋母體血漿DNA之測序結果，將提供一種可用於增強胎兒染色體非整倍體之無

創產前偵測的最迅速採用之臨床效用之一[Agarwal A等人(2013), *Prenat Diagn*, 33: 521-531]。

F. 使用讀數之量量測胎兒DNA含量之例示性方法

圖7展示一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之實例方法700。生物樣品包括母體DNA分子及胎兒DNA分子。方法700可使用電腦系統來進行。

在方塊702處，方法700包括鑑別複數個具有序列資訊之位點，該序列資訊指示女性在該複數個位點之各位點為第一等位基因純合接合。作為實例，序列資訊可自同一樣品(例如樣品之不同讀數集或不同等分試樣)確定，樣品可為血漿或胎兒與母體DNA之其他混合物；或自女性之不同樣品(例如膚色血球層之樣品、頰黏膜拭子或血漿之不同樣品)。無論樣品來源如何，序列資訊可包含DNA分子讀數之另一資料集，例如來自相同樣品或不同樣品之其他讀數。在一些實施例中，可僅僅基於在特定位點偵測到第一等位基因指示女性為純合接合。在其他實施例中，該指示可允許具有不同等位基因之一些讀數，但在該位點具有其他等位基因之讀數之數目低於閾值(例如在特定統計學準確性內稱該位點為純合接合之閾值)。當僅僅母體樣品(例如膚色血球層)用於獲得序列資訊時可進行此類實施例。序列資訊可經由任何適合之技術，例如測序獲得或如本文所述證明。

該女性實際上在該複數個位點可為純合接合。然而，在一些實施例中，該女性可在淺深度下測序，從而在該等位點僅僅讀取幾個等位基因(例如一個或兩個)，且即使該女性在該位點為雜合接合，該女性亦可呈現純合接合。鑑別該複數個位點可包括自該生物樣品之DNA分子獲得複數個讀數。在其他實施例中，鑑別女性呈現純合接合之該複數個位點可包括自不

包括胎兒DNA樣品之另一生物樣品(亦即第二生物樣品)之複數個讀數鑑別複數個位點。舉例而言，此第二生物樣品可為母體膚色血球層或頰黏膜拭子。鑑別女性呈現純合接合之該複數個位點可包括對來自女性之第二生物樣品中的細胞進行基因分型。在一些實施例中，母體基因型之分析不需要高度準確，且可自母體膚色血球層之淺深度測序獲得，例如(但不限於)小於單倍體人類基因組之0.1x、0.2x、0.3x、0.4x、0.5x、0.8x、1x、1.5x、2x、3x、4x、5x及10x覆蓋範圍。在一些實施例中，該複數個讀數可僅僅限於出現在已知具有SNP之位點之參考資料庫中的第二複數個位點之讀數。

在方塊704處，方法700包括自生物樣品之DNA分子獲得複數個讀數。該複數個讀數可自測序裝置或自資料儲存裝置獲得。方法700亦可包括在獲得讀數前接收生物樣品。讀數可限於在資料庫中鑑別為對應於雙等位基因位點之位點，亦即具有SNP之位點。可對生物樣品中之複數個DNA分子測序以獲得讀數。在其他實施例中，可使用探針微陣列分析該生物樣品中複數個DNA分子以獲得讀數。

在方塊705處，方法700包括在參考基因組中鑑別複數個讀數之位置。舉例而言，不含細胞之DNA分子可經測序以獲得序列讀數，且該等序列讀數可定位(對準)至參考基因組。若生物體為人類，則參考基因組將為潛在地來自特定子群體之參考人類基因組。作為另一實例，可用不同探針(例如以下PCR或其他擴增)分析不含細胞之DNA分子，其中各探針對應於基因組位置。

在方塊706處，方法700包括確定讀數之第一量。讀數之第一量之各讀數位於該複數個位點之一位點，且各讀數在該位點展現不同於女性之第一

等位基因的第二等位基因。在一些情況下，讀數之第一量為非母體等位基因之讀數之量。第二等位基因可限於在資料庫中鑑別為對應於雙等位基因位點之等位基因。並非複數個位點之所有位點均可包括展現第二等位基因之讀數。實際上，複數個位點之位點一部分可不包括展現第一等位基因之讀數。

在方塊708處，可確定在複數個位點之讀數之第二量。讀數之第二量之各讀數位於該複數個位點之一位點，且各讀數在該位點展現第一等位基因。在一些實施例中，第二量可包括與展現不同於該女性之等位基因之等位基因的讀數之第一量求和的一致等位基因之讀數。換言之，第二量可為如圖1中所示之A+B之總和，或第二量可為A之總和。確定讀數之第二量可暗中自讀數總數確定。讀數總數可為讀數之第二量。

在方塊710處，非母體等位基因含量可自第一量及第二量確定。非母體等位基因含量可包括第一量除以第二量。非母體等位基因含量可包括轉化成百分比之數值。在一些實施例中，非母體等位基因含量可包括第二量除以第一量。

在方塊712處，可獲得使用具有已知之胎兒DNA含量之另一樣品及量測之非母體等位基因含量確定的校準點。校準點可為複數個校準點之一個校準點，且複數個校準點可構成校準曲線。校準曲線可藉由確定來自複數個懷孕女性之複數個其他樣品之胎兒DNA含量計算。該複數個其他樣品之各其他樣品的胎兒DNA含量可包括鑑別第二複數個位點，其中在各位點，胎兒為雜合接合且懷孕女性為純合接合。在一些實施例中，胎兒DNA含量可使用男性胎兒之Y染色體確定。可獲得來自其他樣品之DNA分子之複數個讀數。複數個讀數可等於或類似於來自第一生物樣品之DNA分子的複數

個讀數之數目。可確定在該第二複數個位點具有胎兒特異性等位基因之讀數的第三量；可確定在該第二複數個位點具有共享等位基因之讀數的第四量。可使用第三量及第四量確定胎兒DNA含量。可計算複數個樣品之非母體片段。胎兒DNA含量及非母體片段可與線性或其他函數擬合。線性或其他函數可描述校準曲線。

在方塊714處，可使用校準點及非母體等位基因含量計算胎兒DNA含量。非母體等位基因含量可與校準曲線之校準點相比。計算之胎兒DNA含量可等於對應於校準曲線上相同或類似非母體等位基因含量之胎兒DNA含量。若校準曲線由等式表示，則胎兒DNA含量可為非母體等位基因含量取代至等式中之計算結果。

G. 使用DNA分子之特性量測胎兒DNA含量之例示性方法

圖11展示一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之實例方法1100。方法1100可使用界定一組DNA分子之特性的值。特性可包括該組分子之尺寸參數或該組分子之序列讀數之量。

在方塊1102處，方法1100包括鑑別複數個具有序列資訊之位點，該序列資訊指示女性在該複數個位點之各位點為第一等位基因純合接合。鑑別該複數個位點可藉由本文所述之任何操作，包括方法700中描述之操作。

在方塊1104處，方法1100包括自DNA分子獲得複數個讀數。獲得該複數個讀數可藉由本文所述之任何操作，包括方法700中描述之操作。

在方塊1105處，方法1100包括在參考基因組中鑑別複數個讀數之位置。參考基因組可為人類基因組。鑑別讀數位置可包括將讀數與參考基因組比對或使用探針。鑑別位置可藉由本文所述之任何操作，包括方法700中對方塊705所描述之操作。

在方塊1106處，方法1100包括確定第一組DNA分子之第一值。第一組DNA分子之各DNA分子可包括位於該複數個位點之一位點的讀數。各讀數可在該位點展現不同於第一等位基因之第二等位基因。第一值可界定該第一組DNA分子之特性。舉例而言，如方法700中，第一值可為位於該複數個位點且具有第二等位基因之讀數之數目。確定第一值可進一步包括量測第一組DNA分子之尺寸，其中第一尺寸值具有第一組DNA分子之第一尺寸分佈。在實施例中，第一值可為尺寸參數。尺寸參數可為尺寸在一定範圍內之分子數目，或在一定尺寸下分子之累積頻率，例如具有第一組DNA分子中最大尺寸之DNA分子的第一累積頻率。

在方塊1108處，方法1100包括確定第二組DNA分子之第二值。第二組DNA分子之各DNA分子可包括位於該複數個位點之一位點的讀數。各讀數在該位點可展現第一等位基因。第二組DNA分子亦可來自與第一組DNA分子相同之生物樣品，或可來自另一生物樣品(例如僅僅母體DNA之樣品，諸如膚色血球層或頰黏膜拭子)。確定第二值可進一步包括量測第二組DNA分子之尺寸，其中第二尺寸值具有第二組DNA分子之第一尺寸分佈。第二值可界定該第二組DNA分子之特性。舉例而言，若第一值為尺寸參數，則第二值亦可為尺寸參數。該第二值可為位於複數個位點且具有第一等位基因之讀數的第二數目。

在方塊1110處，參數之參數值可自第一值及第二值確定。參數可包括第一值除以第二值之比率。

在方塊1112處，方法1100可包括將參數值與使用至少一個具有已知之胎兒DNA含量之其他樣品(例如校準樣品)及對應於該至少一個其他樣品中參數之另一量測值之校準值確定的校準點比較。校準點可為複數個校準點

之一個校準點，且複數個校準點可構成校準曲線。校準曲線可類似於在方法700之方塊712中的操作，藉由確定來自複數個懷孕女性之複數個其他樣品之胎兒DNA含量計算。可計算複數個其他樣品之參數值。胎兒DNA含量及參數可與線性或其他函數擬合。線性或其他函數可描述校準曲線。

在方塊1114處，可基於該比較計算胎兒DNA含量。計算之胎兒DNA含量可等於對應於校準曲線上相同或類似參數值之胎兒DNA含量。若校準曲線由等式表示，則胎兒DNA含量可為參數值取代至等式中之計算結果。

H. 使用尺寸參數量測胎兒DNA含量

具有非母體等位基因之DNA分子之尺寸及/或具有母體等位基因之DNA分子之尺寸可用於估計胎兒DNA含量。已發現胎兒DNA比母體血漿中母體DNA短(Lo YMD等人Sci Transl Med. 2010;2:61ra91)。因此，具有非母體等位基因之DNA分子應平均比母體血漿中具有母體等位基因之DNA分子短。

作為一個實例，具有20%胎兒DNA含量之母體血漿藉由微陣列法(Illumina)基因分型。鑑別母體DNA為純合接合且非母體等位基因存在於血漿中之位點。比較母體等位基因及非母體等位基因下具有此等位點之DNA分子的尺寸。

圖12A展示具有母體等位基因及非母體等位基因之DNA分子之尺寸分佈。x軸為鹼基對中DNA分子之尺寸。y軸為呈百分比之既定尺寸之頻率。線1202為具有母體等位基因之DNA分子的尺寸分佈，且線1204為具有非母體等位基因之DNA分子的尺寸分佈。線1204一般在線1202左邊，表明具有非母體等位基因之DNA分子一般比具有母體等位基因之DNA分子短。

圖12B展示來自圖12A之DNA分子之尺寸的累積頻率。x軸為鹼基對中

DNA分子之尺寸。y軸為呈百分比之累積頻率。線1206為具有母體等位基因之DNA分子之尺寸的累積頻率曲線。線1208為具有非母體等位基因之DNA分子之尺寸的累積頻率曲線。線1208在線1206上方，表明具有非母體等位基因之DNA分子比具有母體等位基因之DNA分子短。

圖12C展示ΔS，兩個累積頻率曲線(線1206及線1208)之間的差異。x軸為鹼基對中DNA分子之尺寸。y軸為ΔS，兩個累積頻率曲線之間的差異。ΔS之最大值在約150 bp下。因此，對於尺寸小於或等於150 bp之DNA分子，DNA分子將相對富含非母體等位基因。在150 bp下之ΔS表示為ΔS_{150}，針對具有800萬個雙端序列讀數之32個樣品定量其，以測試其評估胎兒DNA含量之適合性。

圖12D展示32個樣品之ΔS_{150}與胎兒DNA含量之間的關係。x軸為呈百分比之胎兒DNA含量。y軸為ΔS_{150}，在150 bp長度下具有非母體等位基因之DNA分子之累積頻率曲線與具有母體等位基因之DNA分子之累積頻率曲線之間的差異。ΔS_{150}與胎兒DNA含量正相關。換言之，較高量之攜帶非母體等位基因之短DNA分子表明較高胎兒DNA含量。線性回歸與資料擬合。線性擬合具有0.81之R^2 ($p<0.01$)。具有母體等位基因之DNA分子包括作為胎兒DNA，但仍攜帶母體等位基因之DNA分子。因此，不預期ΔS_{150}反映母體DNA與胎兒DNA之間的實際尺寸差異。

在一些實施例中，ΔS可在除150 bp外之尺寸下。舉例而言，ΔS可在50、60、70、80、90、100、110、120、130、140、160、170、190、200或210 bp下。亦可使用其他尺寸參數。尺寸差異可在兩組之尺寸分佈之任何統計值之間。舉例而言，可使用第一組DNA片段與第二組DNA片段之中值尺寸之差異。作為另一實例，第一組與第二組之間的尺寸之累積頻率之

最大值。可使用美國專利公開案2011/0276277及2013/0237431中描述之任何尺寸值。

可使用尺寸參數與胎兒DNA含量之間的校準曲線。校準曲線可將胎兒DNA含量與其他樣品之尺寸參數相關。其他樣品之胎兒DNA含量可藉由本文所述之任何方法確定。接著尺寸參數可針對其他樣品量測且針對胎兒DNA含量繪圖。線性回歸或其他回歸將與該資料擬合以確定校準曲線。接著未知胎兒DNA含量之生物樣品之尺寸參數可與校準曲線相比以估計胎兒DNA含量。

在此等實施例中，基於非母體等位基因DNA及母體等位基因DNA之尺寸的尺寸參數可用於評估胎兒DNA含量，即使參數不反映胎兒DNA及母體DNA之尺寸。

I. 使用尺寸參數量測胎兒DNA含量之例示性方法

圖13展示一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之實例方法1300。方法1300可使用界定一組DNA分子之尺寸的值。值可為尺寸參數之值。

在方塊1302處，方法1300包括基於序列資訊鑑別複數個位點，該序列資訊指示女性在該複數個位點之各位點為第一等位基因純合接合。鑑別複數個位點可藉由本文所述之任何操作，包括如方法700中。

在方塊1304處，方法1300包括自生物樣品之DNA分子獲得複數個讀數。獲得該複數個讀數可藉由本文所述之任何操作，包括如方法700中。

在方塊1305處，方法1300包括鑑別參考基因組中複數個讀數之位置且確定生物樣品之DNA分子之尺寸。鑑別位置可藉由本文所述之任何操作，包括方法700中針對方塊705所描述之操作。量測尺寸可藉由電泳或電腦模

擬進行。

在方塊1306處，方法1300包括確定第一組DNA分子之第一尺寸值。第一組DNA分子之各DNA分子可包括位於該複數個位點之一位點的讀數。各讀數可在該位點展現不同於第一等位基因之第二等位基因。第一尺寸值可對應於第一組DNA分子之第一尺寸分佈之統計值。第一尺寸值可為尺寸參數。尺寸參數可為尺寸在一定範圍內之分子之數目，或在一定尺寸下之分子之累積頻率。作為其他實例，尺寸值可為第一組DNA分子之中值尺寸、尺寸分佈之眾數或平均尺寸。

在方塊1308處，方法1300包括確定第二組DNA分子之第二值。第二組DNA分子之各DNA分子可包括位於該複數個位點之一位點的讀數。各讀數在該位點可展現第一等位基因。第二組DNA分子亦可來自生物樣品或可來自另一生物樣品(例如僅僅母體DNA之樣品，諸如膚色血球層或頰黏膜拭子)。作為其他實例，尺寸值可為第二組DNA分子之中值尺寸、尺寸分佈之眾數或平均尺寸。第二尺寸值可對應於第二組DNA分子之第二尺寸分佈之統計值。舉例而言，若第一值為尺寸參數，則第二值亦可為尺寸參數。

在方塊1310處，可自第一值及第二值確定參數值。參數可包括第一值除以第二值之比率。

在方塊1312處，方法1300可包括將參數值與使用至少一個具有已知之胎兒DNA含量之其他樣品(例如校準樣品)及對應於該至少一個其他樣品中參數之另一量測值之校準值確定的校準點比較。校準點可為複數個校準點之一個校準點，且複數個校準點可構成校準曲線。校準曲線可類似於方法700中所述，藉由確定來自複數個懷孕女性之複數個其他樣品之胎兒DNA含量計算。可計算複數個其他樣品之參數值。胎兒DNA含量及參數可與線

性或其他函數擬合。線性或其他函數可描述校準曲線。

在方塊1314處，可基於該比較計算胎兒DNA含量。參數值可與校準曲線之校準點相比。計算之胎兒DNA含量可等於對應於校準曲線上相同或類似參數值之胎兒DNA含量。若校準曲線由等式表示，則胎兒DNA含量可為參數值取代至等式中之計算結果。

II. 使用基因座之量分析DNA

為確定母體基因型或胎兒基因型，一些實施例不需要分析來自僅僅母體DNA、僅僅胎兒DNA或來自僅一名個體之任何DNA之樣品的讀數。實際上，一些實施例無需包括關於母體基因型之高度準確資訊。舉例而言，無需在高統計學信賴度或甚至任何統計學信賴度下已知確定母體基因型上何種基因座為純合接合。實際上，方法可藉由含有母體DNA與胎兒DNA之樣品中僅僅一個或幾個等位基因之存在，假設某些基因座為純合接合。此等方法常常具有淺度測序深度，該等淺度測序深度視為不足以信賴地評估基因座存在之等位基因。舉例而言，確定基因座為純合接合可僅僅基於基因座之一個或兩個讀數。因此，鑑別為純合接合之位點僅僅可似乎為純合接合，因為位點尚未以足夠深度測序。

另外，分析DNA之實施例可包括分析含有母體DNA與胎兒DNA之樣品中替代等位基因(例如非母體等位基因)之明顯純合接合基因座。分析替代等位基因之樣品亦可在淺度測序深度下進行。淺度測序深度可在基因座產生少數讀數，有時僅僅一個或兩個讀數。基因座之讀數之低數目可導致無法測序實際上存在於基因座之任何替代等位基因，或存在於基因座之替代等位基因之比例計數不足。因為此等可能誤差，所以將不預期使用淺度測序深度之技術準確量測胎兒DNA含量或生物樣品之其他特徵。

此外，將不預期作為確定胎兒DNA含量之手段的鑑別基因座上替代等位基因對任何單個基因座起作用。對於任何單個基因座，替代等位基因將存在或不存在。此二元結果將不提供足夠用於量測胎兒DNA含量或生物樣品之其他特徵的資訊。

然而，當進行淺度測序時本文所述之方法意外地可準確量測胎兒DNA含量或生物樣品之其他特徵。此等方法可藉由使用若干基因座，將結果求平均值以將測序及其他誤差降至最低且使用校準數據提供關於生物樣品之有用資訊。此等方法優於習知方法，習知方法僅僅對男性胎兒奏效，可能需要父母任一者或兩者之基因型資訊或可能需要高測序深度。

A. 一般方法

圖8展示不獲得母體基因型、父本基因型或僅僅含有母體DNA分子之生物樣品的量測胎兒DNA含量之方法800的圖示。

方塊802開始於一個生物樣品或多個生物樣品。生物樣品可為來自懷有胎兒之女性的血漿、血清、血液、唾液、汗水、尿、淚液或其他體液。生物樣品可具有1%、2%、3%、4%或5%之最小胎兒DNA分子片段。生物樣品含有母體DNA分子與胎兒DNA分子。生物樣品可藉由醫學專業人員投與之針獲得。生物樣品亦可作為常規醫學約診之部分無創獲得。

方塊804展示自來自生物樣品之DNA分子獲得測序讀數。任何資料集之測序讀數可為淺度或低深度。舉例而言，測序讀數之數目可小於單倍體人類基因組之0.1x、0.2x、0.3x、0.4x、0.5x、0.8x、1x、1.5x、2x、3x、4x、5x及10x覆蓋範圍。DNA分子之測序可藉由任何適合之測序技術或系統進行。測序或讀數可限於具有已知及常見SNP，包括參考資料庫(例如dbSNP或HapMap)中之SNP之位點。

方塊806及808展示自一或多個生物樣品獲得之測序讀數之兩個資料集。兩個資料集可為來自以下之數據：兩個生物血漿DNA複本(亦即在約相同時間來自相同患者之兩次不同血液抽取)；一個血漿樣品拆分成兩個等分試樣；一個血漿樣品及一個構成基因組DNA樣品(例如母體膚色血球層DNA、頰黏膜拭子DNA)；或一個電腦模擬隨機拆分成兩個測序資料集的血漿/血清測序資料集。因此，兩個樣品可在方塊802中獲得，其中分別針對各樣品獲得序列讀數。

方塊810描繪鑑別第一組基因座之各基因座中互斥等位基因。為達成說明之目的，讀數812展示第一組基因座814，其特徵為在基因座a-h之各基因座互斥之單個等位基因。圖8中等位基因由白色或黑色正方形表示。在讀數812中，第一組基因座814包括基因座a-h。基因座a-h可不為連續基因座。此等基因座為明顯純合接合，因為無基因座展示存在兩個不同等位基因。在讀數812中任何基因座僅僅一個或兩個讀數下，表徵任何基因座為純合接合無法在高統計學信賴度下進行。實際上，對於具有單個讀數之基因座，將通常不考慮該基因座在任何信賴度下表徵為純合接合。此等基因座可限於具有已知及常見SNP之基因座。

方塊816展示鑑別展現替代等位基因之第二組基因座。自第一組基因座814內鑑別第二組基因座。讀數818展示自第二資料集測序之等位基因。基因座a、c、f及g展示針對相同基因座，等位基因不同於讀數814中等位基因之讀數。此等基因座展示替代等位基因，因為等位基因為第一資料集中等位基因之替代。基因座b、d、e及h展示等位基因為與讀數814中相同之等位基因的讀數。因此第二組基因座鑑別為基因座a、c、f及g。

方塊820確定基因座之第一量。基因座之第一量可自第一組基因座確

定。在其他實施例中，因為在第二資料集中分析第一組基因座，所以基因座之第一量可自第二資料集確定。第一量可為具有等位基因之基因座的數目或具有等位基因之讀數的數目。若第一量為基因座之數目，則對於讀數812，說明之基因座之數目為8。第一量可限於來自具有某一尺寸或某一特性之DNA分子的讀數。舉例而言，第一量可為具有某一絕對尺寸或相對於其他DNA分子某一尺寸之DNA分子的基因座之數目。具有等位基因之讀數之數目可為等位基因讀數之計數。在讀數812中，具有等位基因之讀數之數目為11。在某些實施例中，若各基因座平均有約一個等位基因讀數，則具有等位基因之讀數之平均數目可等於基因座數目。

方塊822確定第二組基因座中基因座之第二量。第二量可為具有等位基因之基因座的數目或具有等位基因之讀數的數目。第二量應與第一量相稱且具有相同單位，但在一些實施例中，第一量及第二量可具有不同單位。若第二量為基因座之數目，則自第二組讀數818確定之第二量為4。若第二量為具有等位基因之讀數之數目，則讀數818中之第二量為6。因為第二資料集中第一組基因座係針對第二組基因座分析，所以除第一資料集之外，在一些情況下可考慮自第二資料集確定基因座之第一量。

方塊824自第一量及第二量確定表觀等位基因偏差(AAD)。AAD為定量在第二資料集中展示替代等位基因且不存在於第一資料集之位點之比例的參數。AAD可藉由第二量除以第一量計算，如方塊824中所示。在其他實施例中，AAD可自第二量除以僅僅母體等位基因之量(亦即第二量與第一量之間的差值)計算。計算AAD可包括所述計算之乘法因子及/或倒數。AAD可視為第二量之歸一化參數。

方塊826展示使用AAD分析生物樣品。分析生物樣品可包括使用校準

曲線自AAD計算胎兒DNA含量。校準曲線描述胎兒DNA含量與AAD之間的關係，如圖828中所示。校準曲線可基於來自其他生物樣品之實際胎兒DNA含量及AAD值確定。校準曲線中資料點之測序讀數之數目可類似於具有未知胎兒DNA含量之生物樣品中測序讀數之數目。換言之，來自具有已知之胎兒DNA含量的AAD資料應在與具有未知之胎兒DNA含量的生物類似或相同測序深度下。舉例而言，校準曲線可在來自生物樣品之DNA分子之測序深度的1x、5x、10x、15x或20x內的測序深度下。在一些實施例中，校準曲線可限於具有與母親、父親或胎兒類似基因背景之樣品。舉例而言，校準曲線可縮至來自具有相同或類似種族基團之人之樣品的AAD數據。校準曲線亦可限於特定單倍型或單倍型方塊。因此，若干校準曲線可用於相同樣品、若干基因組區域，包括單倍型。在一些實施例中，AAD可用於雙接合子之無創測試。

因為測序深度增加，所以鑑別為具有非母體等位基因之基因座之比例將隨著對愈來愈多基因座之母體的等位基因測序而增加。因此，在高度測序深度下，將不預期具有非母體等位基因之基因座之比例及AAD值隨胎兒DNA含量變化。測序深度可限於5x、10x、15x、20x或25x覆蓋範圍之最大值，以避免其中AAD值不視胎兒DNA含量而定之區域。生物樣品仍可在超過此最大值之高測序深度下測序，但所得數據接著可隨機減少取樣以產生測序深度低於最大值之測序讀數資料集。

B. 量測胎兒DNA含量之例示性方法

圖9展示一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之方法900。生物樣品包括母體DNA分子及胎兒DNA分子。生物樣品可為本文所述之任何生物樣品。

在方塊902中，方法900包括接收來自懷有胎兒之女性之第一樣品之DNA分子的複數個讀數之資料集。資料集可由來自測序裝置或資料儲存裝置之電腦系統接收。第一樣品可為或可不為生物樣品。第一樣品可具有僅僅母體DNA且無胎兒DNA，諸如膚色血球層或頰黏膜拭子。

在方塊904處，方法900包括在參考基因組中鑑別複數個讀數之位置。鑑別該等位置可藉由本文所述之任何操作，包括方法700中之方塊704中針對鑑別位點所描述之操作。

在方塊906中，方法900包括基於第一資料集及鑑別之位置鑑別第一組基因座。第一組基因座中無基因座展現超過一個等位基因。換言之，第一組基因座之各基因座為單偶合基因，且呈現純合接合。第一組基因座可選自來自參考資料庫之一組基因座。換言之，第一組基因座可為來自參考資料庫之該組基因座之子集，其中該組基因座中第一組基因座之各基因座來自參考資料庫。可已知該組基因座包括單核苷酸多型現象(SNP)或高雜合接合性情況。參考資料庫可包括短基因變化資料庫(dbSNP)或HapMap資料庫。該組基因座可縮至已知在類似於母親或胎兒之種族或基因組之某些種族或其他基因組中具有高概率雜合接合性之某些基因座。

複數個讀數可在淺深度下。舉例而言，在實施例中讀數深度可小於或等於10x、小於或等於5x、小於或等於4x、小於或等於3x、小於或等於2x、小於或等於1x或小於或等於0.5x。對於單倍體人類基因組，1x覆蓋範圍為50 bp之尺寸，大約5000萬個讀數。讀數之數目可小於或等於5000萬個讀數，包括小於或等於3000萬個讀數、2000萬個讀數、1500萬個讀數、小於或等於1000萬個讀數、小於或等於800萬個讀數、小於或等於500萬個讀數、小於或等於400萬個讀數、小於或等於200萬個讀數或小於或等於100

萬個讀數。基因座可具有總共一個或兩個讀數。第一組基因座中複數個基因座，包括第一組基因座中所有基因座超過10%、超過20%、超過30%、超過40%、超過50%、超過60%、超過70%、超過80%或超過90%可具有一個或兩個讀數之最大值。第一組基因座之任何基因座中讀數之最大數目可為1、2、3、4、5、6、7、8、9或10。

在方塊908中，方法900包括確定基因座之第一量。第一量可為來自第一資料集之第一組基因座中基因座之數目或可為第一組基因座中等位基因讀數之總數。在其他實施例中，第一量可自來自第二資料集之第一組基因座中基因座之數目確定。

在方塊910中，可接收來自生物樣品之DNA分子之複數個讀數的第二資料集。第二資料集可由來自測序裝置或資料儲存裝置之電腦系統接收。

在方塊912處，方法900包括在參考基因組中鑑別第二複數個讀數之位置。鑑別該等位置可藉由本文所述之任何操作，包括方法700中之方塊704中針對鑑別位點所描述之操作。

在方塊914中，基於第二資料集及鑑別之位置鑑別第二組基因座。第二組基因座之各基因座展現不同於在第一組基因座中展現之等位基因的等位基因。換言之，第二組基因座之各基因座可展示非母體等位基因，而第一組基因座中各對應基因座可展示僅僅母體等位基因。第二資料集中各讀數可不同於第一資料集中各讀數。在一些實施例中，第一資料集可為複數個讀數之一半，且第二資料集可為複數個讀數之另一半。

在方塊916中，方法900包括確定第二資料集中該組基因座中基因座之第二量。第二量可為第二組基因座中基因座之數目或可為第二組基因座中等位基因讀數之總數。第二量可限於來自具有某一尺寸之DNA分子的讀

數。舉例而言，用於第二量之讀數可限於來自與第二複數個DNA分子具有最小平均尺寸差異之第一複數個DNA分子的讀數。第二複數個DNA分子可包括用於第一組基因座之序列讀數的DNA分子。第二複數個DNA分子可包括在生物樣品中測序之所有DNA分子。在實施例中最小尺寸差異可為5 bp、10 bp、20 bp、30 bp或40 bp。可量測第一複數個或第二複數個DNA分子之尺寸，或可接收第一複數個或第二複數個DNA分子之尺寸。

在方塊918中，可確定第一量及第二量之歸一化參數。在一些實施例中，歸一化參數可包括第二量除以第一量。接著歸一化參數可為明顯非母體等位基因之數目與母體等位基因之數目的比率。在其他實施例中，歸一化參數可包括第二量除以第一量與第二量之總和。在此等實施例中，歸一化參數可為明顯非母體等位基因之數目與等位基因之總數的比率。歸一化參數亦可為任一此等計算之倒數。AAD為歸一化參數之一實例。

在方塊920處，方法900包括將參數值與使用至少一個具有已知之胎兒DNA含量之其他樣品(例如校準樣品)及對應於至少一個其他樣品中參數之另一量測值之校準值確定的校準點比較。校準點可為複數個校準點之一個校準點。複數個校準點可構成校準曲線。校準曲線可為與已知胎兒DNA含量及針對不同生物樣品確定之歸一化參數之資料點擬合的曲線。圖828為校準曲線之一個實例。校準曲線可為資料點之線性回歸。校準曲線可具有不等於1且可小於1之斜率。

校準曲線可使用已知之胎兒DNA含量及歸一化參數(亦即第二歸一化參數)確定，該歸一化參數藉由與來自當前分析之生物樣品的歸一化參數(亦即第一歸一化參數)類似的方法自另一生物樣品確定。第二歸一化參數亦可自類似於方塊902至918之操作確定。與來自其他生物樣品之資料集中

的基因座相關之讀數之數目可約等於當前生物樣品中讀數之數目。讀數之數目可在彼此1x、5x或10x內。

在方塊922中，基於該比較計算胎兒DNA含量。胎兒DNA含量可為校準曲線中對應於歸一化參數之相同值之胎兒DNA含量。在一些實施例中，胎兒DNA含量可內插在歸一化參數之兩個值之兩個胎兒DNA含量之間。在其他實施例中，校準曲線可為形式為y=mx+b之線性等式，其中y為胎兒DNA含量，x為歸一化參數，且m及b為與校準曲線擬合之參數。

C. 胎兒DNA含量之實驗結果

使用來自24個攜有男性胎兒之懷孕女性的24個血漿樣品，測試使用AAD量測胎兒DNA含量，其中各樣品具有平均810萬個序列讀數(範圍：710萬~1030萬)。在24個樣品中，14個樣品用於建立模擬實際胎兒DNA含量含量與AAD值之間的關係之校準曲線。實際胎兒DNA含量藉由來源於Y染色體之讀數之比例確定(Hudecova I等人, PLoS One. 2014；9:e88484)。為計算AAD值，14個樣品之各樣品隨機劃分成兩個資料集。在第一資料集中，鑑別展示一個等位基因及僅僅一種類型等位基因之第一組基因座。在第二資料集中，分析第一組基因座之各基因座以確定是否存在替代等位基因。具有替代等位基因之基因座構成第二組基因座。AAD計算為第二組基因座中基因座之數目除以第一組基因座中基因座之數目乘以100%。

圖10A展示來自胎兒DNA含量及AAD值之線性回歸模型的校準曲線。y軸展示來源於Y染色體之胎兒DNA含量，且x軸展示AAD值。線性回歸具有11.61之斜率及-109.93之y截距。R^2值為0.8795。

圖10B展示基於胎兒DNA含量及短DNA分子之比例之線性回歸。y軸展示來源於Y染色體之胎兒DNA含量，且x軸展示樣品中尺寸小於150 bp

之DNA分子之百分比。已基於DNA分子之尺寸估計胎兒DNA含量(Yu SC等人, Proc Natl Acad Sci USA. 2014;111: 8583-8)。線性回歸具有1.9247之斜率及-3.7911之y截距。R^2值為0.3593。

對此資料集，自AAD值確定胎兒DNA含量得出比如R^2值所指示，自更短DNA分子之比例確定胎兒DNA含量更高的相關性。在更高R^2值下，基於AAD之胎兒DNA含量估計將比基於尺寸型態之方法更準確。

為測試圖10A之基於AAD之校準曲線的普遍性，對來自10個懷孕女性之剩餘10個樣品測序。10個樣品之各樣品隨機劃分成兩個資料集。在第一資料集中，鑑別展示一個等位基因及僅僅一種類型等位基因之第一組基因座。在第二資料集中，分析第一組基因座之各基因座以確定是否存在替代等位基因。具有替代等位基因之基因座構成第二組基因座。AAD計算為第二組基因座中基因座之數目除以第一組基因座中基因座之數目乘以100%。

自圖10A中之校準曲線確定來自10個樣品之AAD值的胎兒DNA含量。另外，藉由來源於Y染色體之讀數之比例，確定10個樣品之胎兒DNA含量。

圖10C在y軸上展示藉由AAD值確定之胎兒DNA含量，其針對基於來源於Y染色體之讀數之比例的胎兒DNA含量。藉由AAD值估計之胎兒DNA含量與實際胎兒DNA含量非常相關，其中R^2為0.896。與實際胎兒DNA含量之中值偏差僅僅為0.8%，表明實現高準確性之胎兒DNA含量預測。因此，觀測到基於AAD之校準曲線很好地推廣至一組新樣品。

基於AAD之胎兒DNA含量估計的準確性可隨樣品中胎兒DNA含量更高、測序誤差率降低及使用基於來自具有類似基因型態之個體之樣品的校準曲線而增加。

D. 用AAD對雙胞胎接合性進行分類

AAD可用於對雙胞胎為單接合還是雙接合進行分類。雙接合雙胞胎具有具不同基因型之胎兒。具有不同基因型之基因座意謂至少一個胎兒具有非母體等位基因。具有雙接合胎兒之血漿樣品中具有非母體等位基因之基因座之比例將高於具有單個胎兒之血漿樣品中基因座之比例。對於單接合雙胞胎，因為胎兒之基因型一致，所以將不預期具有單接合胎兒之血漿樣品中基因座之比例高於具有單個胎兒之血漿樣品中基因座之比例。接著將預期與單接合胎兒相比，對於雙接合胎兒，具有非母體等位基因之基因座之比例較高。因此，將預期自具有非母體等位基因之基因座之比例計算的AAD較高，且將預期雙接合胎兒之計算之胎兒DNA含量較高。

圖14展示針對六組不同雙胞胎計算之胎兒DNA含量。三組雙胞胎為單接合且三組雙胞胎為雙接合。藉由兩種方法估計胎兒DNA含量。在第一方法中，基於DNA分子之尺寸估計胎兒DNA含量(Yu SC等人, Proc Natl Acad Sci USA. 2014;111: 8583-8)。不預期DNA分子之尺寸基於胎兒之接合性變化。在第二方法中，自基因座之量估計胎兒DNA含量(例如如針對使用AAD值之實施例所述)。在第二方法中，自AAD值估計胎兒DNA含量。預期AAD值基於胎兒之接合性變化。圖14展示與單接合雙胞胎相比，雙接合雙胞胎的基於AAD之胎兒DNA含量與基於尺寸之胎兒DNA含量之間的差異更大。胎兒DNA含量估值之此差異可用於將胎兒分類為單接合或雙接合。

為對多胞胎之接合性進行分類，如本文所述，可使用AAD值，估計生物樣品之胎兒DNA含量。接著此第一胎兒DNA含量可與截止值相比。該截止值可確定為比該生物樣品之第二胎兒DNA含量大一定值。第二胎兒DNA

含量可藉由其中估計之胎兒DNA含量不基於樣品中胎兒DNA之接合性變化的方法估計。舉例而言，估計之胎兒DNA含量可基於生物樣品中DNA分子之尺寸型態。截止值可為比第二胎兒DNA含量大某一絕對百分比。舉例而言，圖14中，截止值可在比基於尺寸之胎兒DNA含量大2絕對百分比與4絕對百分比之間。截止值可為超過第二胎兒DNA含量之絕對百分比、超過第二胎兒DNA含量之相對百分比或超過第二胎兒DNA含量之標準偏差之倍數。

若計算之胎兒DNA含量超過截止值，則胎兒可歸類為雙接合。若計算之胎兒DNA含量小於截止值，則胎兒可歸類為單接合。在一些實施例中，可使用兩種截止值，其中第一截止值大於第二截止值。若計算之胎兒DNA含量大於或等於第一截止值，則胎兒可歸類為雙接合。若計算之胎兒DNA含量小於或等於第二截止值，則胎兒可歸類為單接合。若計算之胎兒DNA含量在兩個截止值之間，則胎兒可歸類為接合性不確定。接著胎兒可進行進一步接合性測試。

E. 使用具有某一尺寸之基因座之AAD

AAD之計算可基於經由除序列讀數外之特徵鑑別非母體等位基因。舉例而言，如上文所解釋，胎兒DNA比母體DNA短。因此，長DNA分子可能包括母體等位基因，而短DNA分子可能包括非母體等位基因。表明非母體等位基因之特徵可與基因座中DNA分子之尺寸參數相關。尺寸參數可為某一絕對尺寸或相對於其他DNA分子之某一尺寸。

鑑別具有非母體等位基因之基因座可基於與母體等位基因之尺寸差異。胎兒DNA含量更大可與其中一個資料集中分子展示與另一資料集中分子至少一定尺寸差異之位點的比例更大相關。

作為一個實例，分析來自具有來自懷孕女性之母體DNA的一個等分試樣之淺深度序列資料，且鑑別具有長度超過166 bp之DNA分子的第一組基因座。具有來自同一懷孕女性之母體及胎兒DNA的第二等分試樣在淺深度下測序。在來自第二等分試樣之資料中，鑑別具有尺寸參數(尺寸值)比143 bp短之DNA分子的第二組基因座。換言之，在既定基因座下兩個等分試樣之DNA分子之尺寸參數之間的差異為至少23 bp。第二組基因座中基因座之數目除以第一組基因座中基因座之數目得出具有至少23 bp之尺寸差異的基因座之比例。亦確定懷孕女性之胎兒DNA含量。針對23個其他懷孕女性重複此過程且繪製結果。計算亦可藉由首先確定尺寸值低於尺寸閾值之基因座，且接著確定不同等分試樣中尺寸值高於第二閾值之彼等基因座之比例來進行。

圖15展示胎兒DNA含量與展示尺寸差異之基因座之間的關係。x軸為比例，呈展示兩個等分試樣之間至少23 bp之尺寸差異的基因座之百分比。y軸為胎兒DNA含量。在胎兒DNA含量與展示尺寸差異之基因座之比例之間看到正關係。R^2為0.62 (p=0.0011).

展示尺寸差異之基因座之比例與胎兒DNA含量之間的相關性說明展示尺寸差異之基因座之比例可用作估計胎兒DNA含量之參數(類似於AAD)。尺寸差異不必為23 bp。在其他實施例中，尺寸差異可為至少10、20、30、40或50 bp。每組基因座之數據可不來自兩個不同等分試樣。數據可自相同生物樣品獲得。

最小尺寸差異可用作鑑別非母體等位基因之另一因子。在淺深度測序下，若在第二資料集中發現不同於第一資料集中母體等位基因之等位基因，則第二資料集中等位基因可為非母體等位基因。然而，第二資料集中

等位基因或者可為由於淺深度而在第一資料集中未測序的母體等位基因。若第二資料集中等位基因為與母體等位基因類似之尺寸，則等位基因可能為第二母體等位基因。因此，考慮第二資料集中等位基因之尺寸差異可提高對具有非母體等位基因之基因座之鑑別。

F. 用具有展示尺寸差異之DNA分子之基因座量測具有胎兒DNA含量的例示性方法

圖16展示一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之方法1600。生物樣品包括母體DNA分子及胎兒DNA分子。生物樣品可為本文所述之任何生物樣品。

在方塊1602中，方法1600包括接收來自第一複數個DNA分子之第一複數個讀數的資料集。資料集可由來自測序裝置或資料儲存裝置之電腦系統接收。第一複數個DNA分子可在或可不在生物樣品中。第一複數個DNA分子可來自不具有胎兒DNA之生物樣品。

在方塊1603中，方法1600包括鑑別參考基因組中第一複數個讀數之位置，及確定對應於該第一複數個讀數之DNA分子之尺寸。

在方塊1604中，方法1600包括鑑別第一資料集中第一組基因座。包含第一組基因座每一者中讀數之第一複數個DNA分子具有第一尺寸分佈且具有超出第一尺寸閾值之第一分佈之第一尺寸值。在一些實施例中，包含第一組基因座中讀數之所有DNA分子均超過第一尺寸閾值。第一組基因座可選自來自參考資料庫之一組基因座或考慮其他因素，如方法900中所述。複數個讀數可在淺深度下。

在方塊1606中，方法1600包括確定基因座之第一量。第一量可為來自第一資料集之第一組基因座中基因座之數目。

在方塊1608中，可接收來自生物樣品之第二複數個DNA分子之第二複數個讀數的第二資料集。第二資料集可由來自測序裝置或資料儲存裝置之電腦系統接收。方法1600可包括量測第二複數個DNA分子之尺寸，或接收第二複數個DNA分子之尺寸資訊。

在方塊1609中，方法1600包括鑑別參考基因組中第二複數個讀數之位置，及確定對應於該第二複數個讀數之DNA分子之尺寸。

在方塊1610中，鑑別來自該複數個讀數之第二資料集中第二組基因座。第二組基因座之各基因座為第一組基因座之基因座。包含第二組基因座每一者中讀數之DNA分子具有第二尺寸分佈且在與第一尺寸值超出第一尺寸閾值相反的方向上具有超出第二尺寸閾值之第二分佈之第二尺寸值。

第一尺寸值可大於第一尺寸閾值且第二尺寸值可小於第二尺寸閾值，且第二尺寸閾值可小於第一尺寸閾值。在其他實施例中，第一尺寸值可小於第一尺寸閾值，第二尺寸值可大於第二尺寸閾值，且第二尺寸閾值大於第一尺寸閾值。

在方塊1612中，方法1600包括確定第二資料集中該組基因座中基因座之第二量。第二量可為第二組基因座中基因座之數目。

在方塊1614中，可確定第一量及第二量之歸一化參數。在一些實施例中，歸一化參數可包括第二量除以第一量。接著歸一化參數可為DNA分子小於某一尺寸之基因座之數目與DNA分子大於某一尺寸之基因座之數目的比率。在其他實施例中，歸一化參數可包括第二量除以第一量與第二量之總和。在此等實施例中，歸一化參數可為DNA分子具有更小尺寸之基因座之數目與基因座總數的比率。歸一化參數亦可為任一此等計算之倒數。

歸一化參數可為一種類型AAD。

在方塊1616處，方法1600可包括將參數值與使用至少一個具有已知之胎兒DNA含量之其他樣品(例如校準樣品)及對應於該至少一個其他樣品中參數之另一量測值之校準值確定的校準點比較。校準點可為複數個校準點之一個校準點。複數個校準點可構成校準曲線。校準曲線可為與已知胎兒DNA含量及針對不同生物樣品確定之歸一化參數之資料點擬合的曲線。校準曲線可為資料點之線性回歸。校準曲線可具有不等於1之斜率。

校準曲線可使用已知之胎兒DNA含量及歸一化參數(亦即第二歸一化參數)確定，該歸一化參數藉由與來自當前分析之生物樣品的歸一化參數(亦即第一歸一化參數)類似的方法自另一生物樣品確定。第二歸一化參數亦可自類似於方塊1602至1614之操作確定。與來自其他生物樣品之資料集中的基因座相關之讀數之數目可約等於當前生物樣品中讀數之數目。讀數之數目可在彼此1x、5x或10x內。

在方塊1618中，基於該比較計算胎兒DNA含量。胎兒DNA含量可為校準曲線中對應於歸一化參數之相同值之胎兒DNA含量。在一些實施例中，胎兒DNA含量可內插在歸一化參數之兩個值之兩個胎兒DNA含量之間。在其他實施例中，校準曲線可為形式為y=mx+b之線性等式，其中y為胎兒DNA含量，x為歸一化參數，且m及b為與校準曲線擬合之參數。

III.其他具體實例

實施例1包括一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之方法，該生物樣品包括母體DNA分子及胎兒DNA分子，該方法包含：自該生物樣品之DNA分子獲得複數個讀數；鑑別該女性為純合接合之複數個位點；確定在該複數個位點展現非母體等位基因之讀數之第一量；確定該複

數個位點之讀數之總量；自該第一量及該總量確定非母體等位基因含量；獲得校準曲線，該校準曲線係使用已知之胎兒DNA含量及量測之非母體等位基因含量確定；以及使用該校準曲線及該非母體等位基因含量計算該胎兒DNA含量。

實施例2包括實施例1之方法，其進一步包含：藉由以下來計算該校準曲線：確定來自複數個懷孕女性之複數個其他樣品的胎兒DNA含量；計算該複數個樣品之非母體片段；以及將該等胎兒DNA含量及該等非母體片段與線性函數擬合。

實施例3包括實施例2之方法，其中確定另一樣品之胎兒DNA含量包括：鑑別其中該胎兒為雜合接合且該懷孕女性為純合接合之第二複數個位點；自另一樣品之DNA分子獲得複數個讀數；確定在該第二複數個位點具有胎兒特異性等位基因之讀數之第二量；確定在該第二複數個位點具有共享等位基因之讀數之第三量；以及使用該第二量及該第三量確定該胎兒DNA含量。

實施例4包括實施例1之方法，其中該非母體等位基因限於在資料庫中鑑別為對應於雙等位基因位點的等位基因。

實施例5包括實施例1之方法，其中鑑別該女性為純合接合之該複數個位點包括對來自該女性之細胞之樣品進行基因分型。

實施例6包括實施例1之方法，其進一步包含：接收該生物樣品；以及對該生物樣品中複數個DNA分子測序以獲得該等讀數。

實施例7包括實施例1之方法，其進一步包含：接收該生物樣品；以及使用探針微陣列分析該生物樣品中複數個DNA分子以獲得該等讀數。

實施例8包括一種電腦產品，其包含儲存複數個指令以便控制電腦系

統來執行實施例1至7之任一方法之操作的電腦可讀媒體。

實施例9包括一種系統，其包含：實施例8之電腦產品；以及一或多個處理器，其用於執行儲存於該電腦可讀媒體上之指令。

實施例10包括一種系統，其包含用於執行實施例1至7之任一方法的構件。

實施例11包括一種系統，其經組態以執行實施例1至7之任一方法。

實施例12包括一種系統，其包含分別執行實施例1至7之任一方法之步驟的模組。

IV. 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。此類子系統之實例展示於**圖17**之電腦系統10中。在一些實施例中，電腦系統包括單一電腦設備，其中子系統可為電腦設備之組件。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設備，其各自為一個子系統。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裝置。

展示於圖17中之子系統藉助於系統匯流排75互連。展示額外子系統，諸如印表機74、鍵盤78、儲存裝置79、耦接至顯示配接器82之監測器76及其他。耦合至輸入/輸出(I/O)控制器71之周邊裝置及I/O裝置可藉由此項技術中已知之任何數目之構件(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 (例如乙太網路、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)、滑鼠輸入裝置或掃描儀。經由系統匯流排75之互連允許中央處理器73與各子系統通信且控制來自系統記憶體72或儲存裝置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/

或儲存裝置79可體現電腦可讀媒體。另一子系統為資料採集裝置85，諸如攝影機、麥克風、加速計及其類似物。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接口81或藉由內部接口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、子系統或設備可經網路通信。在該等情況下，可將一個電腦視為用戶端而另一個電腦視為伺服器，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。

應瞭解本發明之任一實施例可以控制邏輯形式以模組或積體方式使用硬體(例如特殊應用積體電路或場可程式閘極陣列)及/或使用電腦軟體用可程式化處理器來執行。如本文中所使用，處理器包括位於同一積體晶片上之單核處理器、多核處理器，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。

描述於本申請案中之任何軟體組件或功能可作為待由處理器執行的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)的軟體程式碼使用例如習知或面向對象之技術來執行。軟體程式碼可以一系列指令或命令形式儲存於電腦可讀媒體上用於儲存及/或傳輸。適合的非暫時性電腦可讀媒體可包括隨機存取記憶體(RAM)、唯讀記憶體(ROM)、磁性媒體(諸如硬碟機或軟碟機)，或光學媒體，諸如壓縮光盤(CD)或數位化通用光碟(DVD)、快閃記憶體及其類似物。電腦可讀媒體可為該等儲存或傳輸裝置之任何組合。

該等程式亦可使用適合於經由有線、光學及/或符合多種方案之無線網

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體。經程式碼編碼之電腦可讀媒體可與相容裝置一起封裝或與其他裝置分開提供(例如經由網際網路下載)。任何此類電腦可讀媒體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測器、印表機、或其他適合之顯示器。

圖18展示一種示例性測序系統。圖18中描繪之系統包含測序裝置1802及作為電腦系統1806之一部分的智慧模組1804。測序裝置1802可包括本文所述之任何測序裝置。電腦系統1806可包括電腦系統10之部分或所有。資料集(測序讀數資料集)經由網路連接或直接連接自測序裝置1802傳送至智慧模組1804或反過來。資料集可例如進行處理以鑑別某些基因座。鑑別及確定步驟可藉由儲存於電腦系統1806之硬體上之軟體執行。資料集可藉由在處理器上運作且儲存於智慧模組之儲存裝置上之電腦編碼處理，且在處理後傳送回分析模組之儲存裝置，其中經修改之數據可顯示於顯示裝置上。在一些實施例中，智慧模組亦可在測序裝置中執行。

圖19展示電腦系統1900可包含接收構件1910，其可包括例如接收自測序裝置獲得之測序資料。電腦系統1900亦可包括鑑別構件1920，其用於自來自DNA分子之複數個讀數鑑別第一資料集中之第一組基因座。電腦系統1900亦可包括確定構件1930，其用於確定第一資料集中第一組基因座中基因座之第一量。電腦系統1900可進一步包括鑑別構件1940，其用於鑑別來自複數個讀數之第二資料集中第二組基因座。電腦系統1900亦可包括確定構件1950，其用於確定第二資料集中第二組基因座中基因座之第二量。電

腦系統1900可進一步包括確定構件1960，其用於確定第一量及第二量之歸一化參數。電腦系統1900可另外包括獲得構件1970，其用於獲得使用已知之胎兒DNA含量確定的校準點。電腦系統1900亦可包括計算構件1980，其用於使用校準點及歸一化參數計算胎兒DNA含量。

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，任何方法中之任何步驟可使用用於執行此等步驟的模組、單元、電路或其他構件來執行。

可在不脫離本發明之實施例之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於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。其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，且鑒於以上教示，許多修改及變化為可能的。

在先前描述中，出於解釋之目的，已闡述多個細節以便提供對本發明技術之各種實施例之理解。然而，熟習此項技術者將清楚，某些實施例可在無此等細節中之一些下或在其他細節下實踐。

在描述若干實施例之後，熟習此項技術者將認識到，在不背離本發明之精神情況下，可使用各種修改、替代構造及同等物。另外，為避免不必

要地混淆本發明，未描述多個熟知方法及元件。另外，任何具體實施例之細節可能不會一直存在於實施例之變體中或可添加至其他實施例中。

在提供值之範圍下，應瞭解除非上下文另外明確規定，否則亦特別揭示在該範圍上限與下限之間的各插入值，精確至下限單位之十分位。涵蓋在所述範圍內任何陳述值或插入值之間的各更小範圍及所述範圍內之任何其他陳述或插入值。此等更小範圍之上限及下限可獨立地包括或排除在該範圍內，且任一界限、無界限或兩個界限包括於更小範圍中之各範圍亦涵蓋於本發明內，服從所述範圍中任何特別排除之界限。在所述範圍包括界限中之一或兩者下，亦包括排除彼等所包括之界限之任一者或兩者的範圍。

除非專門相反地指出，否則「一(a/an)」或「該(the)」之敍述欲意謂「一或多個」。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。不承認任一者為先前技術。

【符號說明】

10	電腦系統
71	輸入/輸出控制器
72	系統記憶體
73	中央處理器
74	印表機
75	系統匯流排
76	監測器
77	輸入/輸出埠

78	鍵盤
79	儲存裝置
81	外部介面
82	顯示配接器
85	資料採集裝置
1802	測序裝置
1806	電腦系統
1804	智慧模組
1900	電腦系統
1910	接收構件
1920	鑑別構件
1930	確定構件
1940	鑑別構件
1950	確定構件
1960	確定構件
1970	獲得構件
1980	計算構件

201720932
專利案號：105130685
201720932
申請日：105/09/22
IPC分類：C12Q 1/68 (2006.01)

【發明摘要】

【中文發明名稱】

藉由母體血漿ＤＮＡ之淺深度測序以準確定量胎兒DNA含量

【英文發明名稱】

ACCURATE QUANTIFICATION OF FETAL DNA FRACTION BY SHALLOW-DEPTH SEQUENCING OF MATERNAL PLASMA DNA

【中文】

本發明之實施例提供在不使用父本或胎兒基因型下用於推導母體血漿中胎兒DNA含量之方法、系統及設備。母體基因型資訊可自僅僅母體DNA樣品獲得或可自具有母體與胎兒DNA分子之生物樣品之淺深度測序來假設。因為測序可在淺深度下，所以基因座可具有僅僅少量讀數且可能無法展現非母體等位基因，即使存在非母體等位基因。然而，表徵所測序之非母體等位基因的歸一化參數可用於提供胎兒DNA含量之準確估值，即使非母體等位基因之量有誤差。本文所述之方法可無需特定區域之高深度測序或富集。因此，此等方法可整合至廣泛使用之無創產前測試及其他診斷學中。

【英文】

Embodiments of the present invention provide methods, systems, and apparatus for deducing the fetal DNA fraction in maternal plasma without using paternal or fetal genotypes. Maternal genotype information may be obtained from a maternal-only DNA sample or may be assumed from shallow-depth sequencing of a biological sample having both maternal and fetal DNA molecules. Because sequencing may be at shallow depths, a locus may have only few reads and may fail to exhibit a non-maternal

allele even if a non-maternal allele is present. However, normalized parameters that characterize non-maternal alleles sequenced can be used to provide an accurate estimate of the fetal DNA fraction, even if the amount of non-maternal alleles is in error. Methods described herein may not need high-depth sequencing or enrichment of specific regions. As a result, these methods can be integrated into widely used non-invasive prenatal testing and other diagnostics.

【指定代表圖】

圖1

【代表圖之符號簡單說明】

無

【發明申請專利範圍】

【第1項】

一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之方法，該生物樣品包括母體DNA分子及胎兒DNA分子，該方法包含藉由電腦系統執行以下：

基於序列資訊，鑑別複數個位點，該序列資訊指示該女性在該複數個位點之各位點為第一等位基因純合接合；

自該生物樣品之DNA分子獲得複數個讀數；

鑑別參考基因組中該複數個讀數之位置；

基於該複數個讀數確定第一組DNA分子之第一值，其中：

該第一組DNA分子之各DNA分子包括位於該複數個位點之一個位點的讀數且在該位點展現不同於該第一等位基因之第二等位基因，且

該第一值界定該第一組DNA分子之特性；

確定第二組DNA分子之第二值，其中：

該第二組DNA分子之各DNA分子包括位於該複數個位點之一個位點的讀數且在該位點展現該第一等位基因，且

該第二值界定該第二組DNA分子之該特性；

自該第一值及該第二值確定參數之參數值；

將該參數值與使用具有已知之胎兒DNA含量之至少一個其他樣品及對應於該至少一個其他樣品中該參數之另一量測值之校準值確定的校準點比較；以及

基於該比較，計算該胎兒DNA含量。

【第2項】

如請求項1之方法，其中：

該第一值為位於該複數個位點且具有該第二等位基因之讀數的第一數目，

該第二值為位於該複數個位點且具有該第一等位基因之讀數的第二數目，且

該參數為非母體等位基因含量。

【第3項】

如請求項1之方法，其中確定該第一值及該第二值進一步包含：

量測該第一組DNA分子及該第二組DNA分子之尺寸，其中該第一值為該第一組DNA分子之第一尺寸分佈之第一尺寸值，且其中該第二值為該第二組DNA分子之第二尺寸分佈之第二尺寸值。

【第4項】

一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之方法，該生物樣品包括母體DNA分子及胎兒DNA分子，該方法包含藉由電腦系統執行以下：

基於序列資訊，鑑別複數個位點，該序列資訊指示該女性在該複數個位點之各位點為第一等位基因純合接合；

自該生物樣品之DNA分子獲得複數個讀數；

鑑別參考基因組中該複數個讀數之位置；

確定讀數之第一量，其中：

讀數之該第一量之各讀數位於該複數個位點之一個位點，且

各讀數在該位點展現與該第一等位基因不同之第二等位基因；

確定在該複數個位點之讀數之第二量，其中：

讀數之該第二量之各讀數位於該複數個位點之一個位點，且

各讀數在該位點展現該第一等位基因；

自該第一量及該第二量確定非母體等位基因含量；

獲得校準點，該校準點係使用具有已知之胎兒DNA含量之另一樣品及量測之非母體等位基因含量確定；以及

使用該校準點及該非母體等位基因含量計算該胎兒DNA含量。

【第5項】

如請求項4之方法，其中：

該校準點為複數個校準點之一個校準點，且

該複數個校準點構成校準曲線，

該方法進一步包含：

藉由以下計算該校準曲線：

確定來自複數個懷孕女性之複數個其他樣品之胎兒DNA含量；

計算該複數個樣品之非母體等位基因含量；以及

將該等胎兒DNA含量及該等非母體等位基因含量與線性函數擬合，其中該線性函數描述該校準曲線。

【第6項】

如請求項4之方法，其中確定在該複數個位點之讀數之該第二量包含確定該複數個讀數之讀數總數，其中該複數個讀數之該讀數總數為讀數之該第二量。

【第7項】

一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量之方法，該生物樣品包括母體DNA分子及胎兒DNA分子，該方法包含藉由電腦系統執行以

下：

基於序列資訊，鑑別複數個位點，該序列資訊指示該女性在該複數個位點之各位點為第一等位基因純合接合；

自該生物樣品之DNA分子獲得複數個讀數；

鑑別參考基因組中該複數個讀數之位置；

確定該生物樣品之DNA分子之尺寸；

確定第一組DNA分子之第一尺寸值，其中：

該第一組DNA分子之各DNA分子包括位於該複數個位點之一個位點的讀數且在該位點展現不同於該第一等位基因之第二等位基因，且

該第一尺寸值對應於該第一組DNA分子之第一尺寸分佈之統計值；

確定該第二組DNA分子之第二尺寸值，其中：

該第二組DNA分子之各DNA分子包括位於該複數個位點之一個位點的讀數且在該位點展現該第一等位基因，且

該第二尺寸值對應於該第二組DNA分子之第二尺寸分佈之統計值；

自該第一尺寸值及該第二尺寸值確定參數值；

將該參數值與使用具有已知之胎兒DNA含量之至少一個其他樣品及對應於該至少一個其他樣品中該參數之另一量測值之校準值確定的校準點比較；以及

基於該比較，計算該胎兒DNA含量。

【第8項】

如請求項7之方法，其中：

該第一尺寸值為在該第一組DNA分子中具有最大尺寸之DNA分子的第一累積頻率，且

該第二尺寸值為在該第二組DNA分子中具有該最大尺寸之DNA分子的第二累積頻率。

【第9項】

如請求項7之方法，其中：

該第一尺寸值為該第一組DNA分子之中值尺寸、平均尺寸或眾數，且

該第二尺寸值為該第二組DNA分子之中值尺寸、平均尺寸或眾數。

【第10項】

如請求項7之方法，其中：

該校準點為複數個校準點之一個校準點，且

該複數個校準點構成校準曲線，

該方法進一步包含：

藉由以下計算該校準曲線：

確定來自複數個懷孕女性之複數個其他樣品之胎兒DNA含量；

計算該複數個樣品之參數值；以及

將該等胎兒DNA含量及該等非母體等位基因含量與線性函數擬合，其中該線性函數描述該校準曲線。

【第11項】

如請求項5或10之方法，其中確定該複數個其他樣品之各其他樣品之該胎兒DNA含量包括：

鑑別第二複數個位點，其中在該第二複數個位點之各位點，該複數個懷孕女性之一懷孕女性為純合接合且該懷孕女性之胎兒為雜合接合；

自該其他樣品之DNA分子獲得複數個讀數；

確定在該第二複數個位點展現胎兒特異性等位基因之讀數之第三量；

確定在該第二複數個位點展現共享等位基因之讀數之第四量；以及

使用該第三量及該第四量確定該胎兒DNA含量。

【第12項】

如請求項1、4或7之方法，其中：

該複數個讀數為第一複數個讀數，

該生物樣品為第一生物樣品，

鑑別該複數個位點包含自來自第二生物樣品之DNA分子的第二複數個讀數鑑別該複數個位點，且

該第二生物樣品不包含胎兒DNA分子。

【第13項】

如請求項1、4或7之方法，其中：

該複數個位點為第一複數個位點，

該複數個讀數限於來自第二複數個位點之讀數，其中該第二複數個位點為在資料庫中鑑別為對應於雙等位基因位點之位點。

【第14項】

如請求項1、4或7之方法，其中該第二等位基因限於在資料庫中鑑別為對應於雙等位基因位點之等位基因。

【第15項】

如請求項1、4或7之方法，其中鑑別該複數個位點包括對來自該女性之細胞之樣品進行基因分型。

【第16項】

如請求項1、4或7之方法，其中：

該複數個讀數為第一複數個讀數，

該生物樣品為第一生物樣品，且

鑑別該複數個位點包含自來自不同於該第一生物樣品之第二生物樣品之DNA分子獲得第二複數個讀數。

【第17項】

如請求項1、4或7之方法，其中：

該複數個讀數為第一複數個讀數，且

鑑別該複數個位點包含自該生物樣品之DNA分子獲得第二複數個讀數。

【第18項】

如請求項1、4或7之方法，其中：

該複數個位點為第一複數個位點，且

該方法進一步包含：

自參考資料庫提供第二複數個位點，其中：

該第二複數個位點包含具有已知之單核苷酸多型現象之位點，且

該第一複數個位點中之每一者為該第二複數個位點之一。

【第19項】

如請求項1、4或7之方法，其進一步包含：

接收該生物樣品；以及

對該生物樣品中複數個DNA分子進行測序以獲得該等讀數。

【第20項】

如請求項1、4或7之方法，其進一步包含：

接收該生物樣品；以及

使用探針微陣列分析該生物樣品中之複數個DNA分子以獲得該等讀

數。

【第21項】

如請求項1、4或7之方法，其中該複數個讀數包含小於或等於5000萬個讀數。

【第22項】

如請求項1、4或7之方法，其中該複數個讀數處於或小於單倍體人類基因組之1x覆蓋範圍下。

【第23項】

如請求項1、4或7之方法，其中該複數個位點之該等位點的一部分不包含展現不同於該第一等位基因之第二等位基因的讀數。

【第24項】

一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量的方法，該生物樣品包括母體DNA分子及胎兒DNA分子，該方法包含：

接收包括來自該懷有胎兒之女性之第一樣品之DNA分子的第一複數個讀數的第一資料集；

鑑別參考基因組中該第一複數個讀數之位置；

基於該第一資料集及該等鑑別之位置鑑別第一組基因座，其中該第一組基因座中無基因座展現超過一個等位基因；

確定該第一組基因座中基因座之第一量：

接收包括來自該生物樣品之DNA分子之第二複數個讀數的第二資料集；

鑑別該參考基因組中該第二複數個讀數之位置；

基於該第二資料集及該等鑑別之位置鑑別第二組基因座，其中：

該第二組基因座中之每一者為該第一組基因座之一，且

該第二組基因座之各基因座展現不同於該第一組基因座中展現之該等位基因的等位基因；

確定該第二資料集中該第二組基因座中基因座之第二量；

確定該第一量及該第二量之歸一化參數；

將該參數值與使用具有已知之胎兒DNA含量之至少一個其他樣品及對應於該至少一個其他樣品中該參數之另一量測值之校準值確定的校準點比較；以及

基於該比較，計算該胎兒DNA含量。

【第25項】

如請求項24之方法，其中該第一樣品為該生物樣品。

【第26項】

如請求項24之方法，其中該第一樣品不為該生物樣品，且其中該第一樣品不包括胎兒DNA分子。

【第27項】

如請求項24之方法，其中不同於該第一組基因座中展現之該等位基因的該等位基因限於在資料庫中鑑別為對應於雙等位基因位點的等位基因。

【第28項】

如請求項24之方法，其中該第一量為基因座之數目，且該第二量為該第二資料集中基因座之數目。

【第29項】

如請求項24之方法，其中確定該第一量包含確定該第二資料集中該第一量。

【第30項】

如請求項24之方法，其進一步包含：

量測該第一複數個DNA分子之尺寸，以及

量測該第二複數個DNA分子之尺寸，

其中：

該第一複數個DNA分子具有第一尺寸值，

該第一尺寸值對應於該第一複數個DNA分子之第一尺寸分佈之統計值，

該第二複數個DNA分子具有第二尺寸值，

該第二尺寸值對應於該第二複數個DNA分子之第二尺寸分佈之統計值，以及

該第一尺寸值以最小差值大於該第二尺寸值。

【第31項】

如請求項24之方法，其中：

包含位於該第一組基因座之讀數之第一複數個DNA分子大於第一尺寸，

包含位於該第二組基因座之讀數之第二複數個DNA分子小於第二尺寸，以及

該第一尺寸與該第二尺寸之間的差值超過最小差值。

【第32項】

如請求項30或31之方法，其中該最小差值為5 bp。

【第33項】

如請求項24之方法，其中該女性懷有複數個胎兒，該方法進一步包含：

將該胎兒DNA含量與截止值比較，以及

若該計算之胎兒DNA含量低於該截止值，則將該複數個胎兒分類為單接合子性，或

若該計算之胎兒DNA含量高於該截止值，則將該複數個胎兒分類為雙接合子性。

【第34項】

如請求項33之方法，其中：

該胎兒DNA含量為第一胎兒DNA含量，

該截止值確定為超過該生物樣品之第二胎兒DNA含量之值，且

在不使用該歸一化參數下確定該第二胎兒DNA含量。

【第35項】

如請求項34之方法，其中使用該生物樣品中DNA分子之尺寸型態，估計該第二胎兒DNA含量。

【第36項】

一種量測懷有胎兒之女性之生物樣品中胎兒DNA含量的方法，該生物樣品包括母體DNA分子及胎兒DNA分子，該方法包含：

接收來自第一複數個DNA分子之第一複數個讀數的第一資料集；

鑑別參考基因組中該第一複數個讀數之位置；

確定對應於該第一複數個讀數之該等DNA分子之尺寸；

鑑別該第一資料集中第一組基因座，其中包含該第一組基因座每一者中讀數之DNA分子具有第一尺寸分佈且該第一尺寸分佈之第一尺寸值超出第一尺寸閾值；

確定該第一組基因座中基因座之第一量：

自來自該生物樣品之第二複數個DNA分子的第二複數個讀數接收第二資料集；

鑑別該參考基因組中該第二複數個讀數之位置；

確定對應於該第二複數個讀數之該等DNA分子之尺寸；

鑑別該第二資料集中第二組基因座，其中：

該第二組基因座中之每一者為該第一組基因座之一，且

包含該第二組基因座每一者中讀數之DNA分子具有第二尺寸分佈且在與該第一尺寸值超出該第一尺寸閾值相反之方向，具有超出第二尺寸閾值之該第二分佈之第二尺寸值；

確定該第二資料集中該第二組基因座中基因座之第二量；

確定該第一量及該第二量之歸一化參數；

將該參數值與使用具有已知之胎兒DNA含量之至少一個其他樣品及對應於該至少一個其他樣品中該參數之另一量測值之校準值確定的校準點比較；以及

基於該比較，計算該胎兒DNA含量。

【第37項】

如請求項36之方法，其中該第一尺寸值大於該第一尺寸閾值且該第二尺寸值小於該第二尺寸閾值，且其中該第二尺寸閾值小於該第一尺寸閾值。

【第38項】

如請求項36之方法，其中該第一尺寸值小於該第一尺寸閾值且該第二尺寸值大於該第二尺寸閾值，且其中該第二尺寸閾值大於該第一尺寸閾值。

【第39項】

如請求項36之方法，其中該第一尺寸值與該第二尺寸值之間的差值為

10 bp。

【第40項】

如請求項24或36之方法，其中該第一資料集包含來自該生物樣品之DNA分子的讀數。

【第41項】

如請求項24或36之方法，其中：

該生物樣品為第一生物樣品，且

該第一資料集包含來自不包含胎兒DNA之第二生物樣品之DNA分子的讀數。

【第42項】

如請求項24或36之方法，其進一步包含：

接收該生物樣品；以及

對該生物樣品中複數個DNA分子進行測序以獲得該第二複數個讀數。

【第43項】

如請求項24或36之方法，其進一步包含：

接收該生物樣品；以及

使用探針微陣列分析該生物樣品中複數個DNA分子以獲得該第二複數個讀數。

【第44項】

如請求項24或36之方法，其中該第二複數個讀數包含小於或等於5000萬個讀數。

【第45項】

如請求項24或36之方法，其中該第二複數個讀數處於或小於單倍體人

類基因組之1x覆蓋範圍下。

【第46項】

如請求項24或36之方法，其中該歸一化參數包含該第二量除以該第一量，或其中該歸一化參數包含該第二量除以該第一量與該第二量之和。

【第47項】

如請求項24或36之方法，其中：

該校準點為複數個校準點之一個校準點，且

該複數個校準點構成校準曲線。

【第48項】

如請求項47之方法，其中該校準曲線為線性的。

【第49項】

如請求項47之方法，其中：

該歸一化參數為第一歸一化參數，

該生物樣品為第一生物樣品，

該校準點係使用該已知之胎兒DNA含量及第二歸一化參數確定，該第二歸一化參數以與該第一歸一化參數對應的方法，自來自第二生物樣品之DNA分子之第三複數個讀數的資料集中基因座組確定，且

該第二複數個讀數包含單倍體人類基因組之第一覆蓋範圍，在該第二複數個讀數之讀數的第二覆蓋範圍之10x覆蓋範圍內。

【第50項】

如請求項24或36之方法，其中：

該歸一化參數為第一歸一化參數，

該生物樣品為第一生物樣品，

獲得該校準點包含使用該已知之胎兒DNA含量及第二歸一化參數，該第二歸一化參數以與該第一歸一化參數對應的方法自第二生物樣品確定。

【第51項】

如請求項24或36之方法，其進一步包含：

自參考資料庫提供第三組基因座，其中：

該第三組基因座包含具有已知之單核苷酸多型現象的基因座，且

該第一組基因座中之每一者為該第三組基因座之一。

【第52項】

如請求項1至10、24至31及33至39中任一項之方法，其中資料集之序列讀數之覆蓋深度小於5x覆蓋範圍。

【第53項】

如請求項1至10、24至31及33至39中任一項之方法，其中資料集之序列讀數之覆蓋深度小於1x覆蓋範圍。

【第54項】

如請求項1至10、24至31及33至39中任一項之方法，其中資料集中序列讀數之數目小於5000萬。

【第55項】

一種包含電腦可讀媒體之電腦產品，該電腦可讀媒體儲存用於控制電腦系統執行如請求項1至54中任一項之方法之操作的複數個指令。

【第56項】

一種系統，其包含：

如請求項55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體上之指令。

【第57項】

一種系統，其包含用於執行如請求項1至54中任一項之方法的構件。

【第58項】

一種系統，其經組態以執行如請求項1至54中任一項之方法。

【第59項】

一種系統，其包含分別執行如請求項1至54中任一項之方法之步驟的模組。