







## 【發明說明書】

### 【中文發明名稱】

來自血漿之胚胎或腫瘤甲基化模式組(METHYLOME)之非侵入性測定

### 【英文發明名稱】

NON-INVASIVE DETERMINATION OF METHYLOME OF FETUS OR TUMOR FROM PLASMA

### 【技術領域】

本發明大體上係關於DNA之甲基化模式(甲基化模式組(methylome))之測定，且更特定言之係關於分析包括來自不同基因組(例如，來自胎兒及母親，或來自腫瘤及正常細胞)之DNA之混合物的生物樣品(例如血漿)以測定少數基因組之甲基化模式(甲基化模式組)。亦描述所測定之甲基化模式組之用途。

### 【先前技術】

胚胎及胎兒發育係複雜過程，且涉及一系列高度調諧的遺傳及表觀遺傳事件。癌症發展亦為通常涉及多個遺傳及表觀遺傳步驟之複雜過程。發育過程之表觀遺傳控制方面的異常牽涉不育症、自發性流產、子宮內生長異常及產後結果。DNA甲基化為最常研究的表觀遺傳機制之一。DNA甲基化主要在甲基加入CpG二核苷酸中之胞嘧啶殘基之5'碳的情形下發生。胞嘧啶甲基化對基因轉錄及DNA功能增加一層控制。舉例而言，富含CpG二核苷酸(稱為CpG島)之基因啟動子之高甲基化通常與壓制基因功能相關。

儘管表觀遺傳機制在介導發育過程方面起重要作用，但不容易獲得

人類胚胎及胎兒組織用於分析(可能同樣無法獲得腫瘤)。實際上不可能在人類產前期間在健康及疾病情況下研究該等表觀遺傳過程之動態變化。可作為產前診斷性程序之一部分或在生產後獲得之胚胎外組織，尤其是胎盤，已為該等研究提供主要途徑之一。然而，該等組織需要侵入性程序。

人類胎盤之DNA甲基化概況已引起研究者的興趣有數十年。人類胎盤展現大量涉及DNA甲基化之特殊生理學特徵。在整體層面上，當與大部分體細胞組織相比時，胎盤組織經低甲基化。在基因層面下，所選基因組基因座(genomic loci)之甲基化狀態為胎盤組織之特異性標誌。整體及基因座特異性甲基化概況兩者均顯示妊娠年齡依賴性變化。印記基因，亦即表現視等位基因之親本來源而定之基因，在胎盤中發揮關鍵功能。該胎盤已描述為假惡性且已觀測到若干腫瘤抑制基因之高甲基化。

對胎盤組織之DNA甲基化概況之研究已洞悉妊娠相關或發育相關疾病(諸如子癩先兆及子宮內生長限制)之病理生理學。基因組印記方面之病症與發育病症相關，諸如普威二氏症候群(Prader-Willi syndrome)及天使症候群(Angelman syndrome)。已在由輔助生殖技術產生之妊娠中觀測到胎盤及胎兒組織中之基因組印記及整體DNA甲基化之經改變概況(H Hiura等人2012 Hum Reprod; 27: 2541-2548)。許多環境因素與後代之表觀遺傳畸變相關，諸如母體抽菸(KE Haworth等人2013 Epigenomics; 5: 37-49)、母體飲食因素(X Jiang等人2012 FASEB J; 26: 3563-3574)及母體代謝狀態，諸如糖尿病(N Hajj等人, Diabetes. doi: 10.2337/db12-0289)。

儘管已付出數十年努力，但尚未有任何實際手段可用於研究胎兒或腫瘤甲基化模式組且監測整個妊娠或疾病過程(諸如惡性疾病)中之動態變化。因此，需要提供非侵入性分析所有或部分胎兒甲基化模式組及腫瘤甲

基化模式組的方法。

### 【發明內容】

實施例提供用於測定並使用各種組織及樣品之甲基化概況的系統、方法及設備。提供實例。可基於血漿甲基化(或具有細胞游離DNA之其他樣品，例如尿液、唾液、生殖器洗滌物)與母親/患者之甲基化概況之比較來推斷胎兒/腫瘤組織之甲基化概況。當樣品具有DNA混合物時，可使用組織特異性等位基因測定胎兒/腫瘤組織之甲基化概況以鑑別來自胎兒/腫瘤之DNA。甲基化概況可用於測定胎兒/腫瘤之基因組中的複本數目變化。已經由各種技術鑑別胎兒之甲基化標記物。可藉由測定DNA片段之大小分佈之大小參數來測定甲基化概況，其中大小參數之參考值可用於測定甲基化程度。

另外，甲基化程度可用於測定癌症程度。在癌症之情形下，血漿中甲基化模式組變化之量測值可允許偵測癌症(例如用於篩選目的)、用於監測(例如，用於偵測抗癌治療後之反應；及用於偵測癌症復發)及用於預測(例如用於量測體內癌細胞負載或用於分期目的或用於評定由疾病或疾病進展或轉移性過程所致之死亡概率)。

可參考以下[實施方式]及所附圖式更透徹地理解本發明實施例之性質及優勢。

### 【圖式簡單說明】

圖1A展示根據本發明實施例之母體血液、胎盤及母體血漿之定序結果之表100。

圖1B展示根據本發明實施例之經定序樣品之1 Mb窗口中之甲基化密度。

圖2A至2C展示 $\beta$ 值相對於甲基化指數之圖：(A)母體血細胞；(B)絨毛膜絨毛樣品；(C)足月胎盤組織。

圖3A及3B展示自成年男性及未妊娠成年女性收集之血漿及血細胞中之甲基化CpG位點之百分比的直方圖：(A)常染色體；(B)染色體X。

圖4A及4B展示血細胞DNA及血漿DNA中之相應基因座之甲基化密度的圖：(A)未妊娠成年女性；(B)成年男性。

圖5A及5B展示自妊娠收集之樣品中之甲基化CpG位點之百分比的直方圖：(A)常染色體；(B)染色體X。

圖6展示母體血液、胎盤及母體血漿之人類基因組之不同重複序列類別之甲基化程度的直方圖。

圖7A展示前三個月樣品之Circos圖700。圖7B展示後三個月樣品之Circos圖750。

圖8A至8D展示比較資訊性單核苷酸多態性周圍之CpG位點之基因組組織DNA相對於母體血漿DNA之甲基化密度的圖。

圖9為說明根據本發明實施例之用於測定生物體之生物樣品之第一甲基化概況之方法900的流程圖。

圖10為說明根據本發明實施例之用於測定生物體之生物樣品之第一甲基化概況之方法1000的流程圖。

圖11A及11B展示根據本發明實施例之使用母體血漿資料及部分胎兒DNA濃度之預測演算法之效能的圖式。

圖12A為展示根據本發明實施例之用於甲基化預測之15個所選基因組基因座之詳情的表1200。圖12B為展示15個所選基因組基因座之推斷類別及其在胎盤中之相應甲基化程度的圖式1250。

圖13為用於偵測懷有至少一個胎兒之女性個體之生物樣品之胎兒染色體異常之方法1300的流程圖。

圖14為根據本發明實施例之用於藉由比較胎盤甲基化概況與母體甲基化概況來鑑別甲基化標記物之方法1400的流程圖。

圖15A為展示使用前三個月資料參考33個先前報導之前三個月標記物之DMR鑑別演算法之效能的表1500。圖15B為展示使用後三個月資料且與分娩時所獲得之胎盤樣品相比較之DMR鑑別演算法之效能的表1550。

圖16為展示基於直接分析母體血漿亞硫酸氫鹽定序資料而預測為高甲基化或低甲基化之基因座之數目的表1600。

圖17A為展示母體血漿、未妊娠女性對照血漿、胎盤及周邊血液DNA之大小分佈的圖1700。圖17B為母體血漿、成年女性對照血漿、胎盤組織及成年女性對照血液之大小分佈及甲基化概況的圖1750。

圖18A及18B為根據本發明實施例之血漿DNA分子之甲基化密度及大小的圖。

圖19A展示成年未妊娠女性之定序讀段之甲基化密度及大小的圖1900。圖19B為展示母體血漿中之胎兒特異性及母體特異性DNA分子之大小分佈及甲基化概況的圖1950。

圖20為根據本發明實施例之用於評估生物體之生物樣品中之DNA甲基化程度之方法2000的流程圖。

圖21A為展示肝細胞癌(HCC)患者之手術前血漿及組織樣品之甲基化密度的表2100。圖21B為展示每份樣品之序列讀段之數目及所達到之定序深度的表2150。



圖22為展示健康對照之血漿樣品中之常染色體甲基化密度(在71.2%至72.5%範圍內)的表220。

圖23A及23B展示HCC患者之白血球層、腫瘤組織、非腫瘤肝組織、手術前血漿及手術後血漿之甲基化密度。

圖24A為展示得自HCC患者之手術前血漿之甲基化密度的圖2400。圖24B為展示得自HCC患者之手術後血漿之甲基化密度的圖2450。

圖25A及25B展示使用四個健康對照個體之血漿甲基化模式組資料作為參考針對染色體1對HCC患者之手術前(圖2500)及手術後(圖2550)血漿樣品之血漿DNA甲基化密度的z-評分。

圖26A為展示手術前及手術後血漿之z評分之資料的表2600。圖26B為展示使用四個健康對照個體作為參考針對所有常染色體之所分析之1 Mb區間對HCC患者之手術前及手術後血漿樣品之血漿DNA甲基化密度之z評分的Circos圖2620。圖26C為展示HCC患者之手術前及手術後血漿樣品兩者中之整個基因組之1 Mb區間之z評分之分佈的表2640。圖26D為展示當使用CHH及CHG情形時與一些對照血漿樣品重疊之腫瘤組織及手術前血漿樣品之甲基化程度的表2660。

圖27A至27H展示根據本發明實施例之8個癌症患者之甲基化密度的Circos圖。圖27I為展示每份樣品之序列讀段之數目及所達到之定序深度的表2780。圖27J為展示患有不同惡性疾病之患者之血漿中整個基因組之1 Mb區間z評分之分佈的表2790。CL=肺腺癌；NPC=鼻咽癌；CRC=結腸直腸癌；NE=神經內分泌癌；SMS=平滑肌肉瘤。

圖28為根據本發明實施例之分析生物體之生物樣品以確定癌症程度分級之方法2800的流程圖。

圖29A為展示參考個體中之甲基化密度分佈的圖2900，假定此分佈遵循正態分佈。圖29B為展示癌症個體中之甲基化密度分佈的圖2950，假定此分佈遵循正態分佈且平均甲基化程度比截止值低2倍標準差。

圖30為展示健康個體及癌症患者之血漿DNA之甲基化密度分佈的圖3000。

圖31為展示健康個體之血漿DNA平均值與HCC患者之腫瘤組織之間的甲基化密度差之分佈的圖式3100。

圖32A為展示當血漿樣品含有5%或2%腫瘤DNA時降低定序深度之效應的表3200。

圖32B為展示四個健康對照個體之血漿、HCC患者之白血球層、正常肝組織、腫瘤組織、手術前血漿及手術後血漿樣品中之重複序列元件及非重複序列區域之甲基化密度的圖式3250。

圖33展示可與根據本發明實施例之系統及方法一起使用之實例電腦系統3300的方塊圖。

圖34A展示全身性紅斑狼瘡(SLE)患者SLE04中之血漿DNA之大小分佈。圖34B及34C展示對來自SLE患者SLE04 (圖34B)及HCC患者TBR36 (圖34C)之血漿DNA之甲基化分析。

圖35為根據本發明實施例之基於CpG島之高甲基化確定癌症程度分級之方法3500的流程圖。

圖36為根據本發明實施例之使用複數個染色體區分析生物體之生物樣品之方法3600的流程圖。

圖37A展示對患者TBR36之腫瘤組織、未經亞硫酸氫鹽(BS)處理之血漿DNA及經亞硫酸氫鹽處理之血漿DNA (由內向外)的CNA分析。圖

37B為展示針對患者TBR36使用1 Mb區間之經亞硫酸氫鹽及未經亞硫酸氫鹽處理之血漿進行CNA偵測之z評分之關係的散點圖。

圖38A展示對患者TBR34之腫瘤組織、未經亞硫酸氫鹽(BS)處理之血漿DNA及經亞硫酸氫鹽處理之血漿DNA (由內向外)的CNA分析。圖38B為展示針對患者TBR34使用1 Mb區間之經亞硫酸氫鹽處理及未經亞硫酸氫鹽處理之血漿進行CNA偵測之z評分之關係的散點圖。

圖39A為展示對HCC患者TBR240之經亞硫酸氫鹽處理之血漿之CNA (內環)及甲基化分析(外環)的Circos圖。圖39B為展示對HCC患者TBR164之經亞硫酸氫鹽處理之血漿之CNA (內環)及甲基化分析(外環)的Circos圖。

圖40A展示對患者TBR36之處理前樣品及處理後樣品的CNA分析。圖40B展示對患者TBR36之處理前樣品及處理後樣品的甲基化分析。圖41A展示對患者TBR34之處理前樣品及處理後樣品的CNA分析。圖41B展示對患者TBR34之處理前樣品及處理後樣品的甲基化分析。

圖42展示利用不同數目之定序讀段進行全基因組低甲基化分析之診斷效能的圖式。

圖43為展示基於全基因組低甲基化分析在不同區間大小(50 kb、100 kb、200 kb及1 Mb)下偵測癌症之ROC曲線的圖式。

圖44A展示具有畸變之區間之累積概率(CP)及百分比之診斷效能。圖44B展示整體低甲基化、CpG島高甲基化及CNA之血漿分析之診斷效能。

圖45展示肝細胞癌患者中之整體低甲基化、CpG島高甲基化及CNA之結果的表。

圖46展示罹患除肝細胞癌以外之癌症的患者中之整體低甲基化、

CpG島高甲基化及CNA之結果的表。

圖47展示病例TBR34之血漿甲基化之系列分析。

圖48A展示顯示HCC患者TBR36之經亞硫酸氫鹽處理之血漿DNA中之CNA (內環)及甲基化變化(外環)的Circos圖。圖48B為HCC患者TBR36之染色體得失區域及無複本數變化區域之甲基化z評分的圖。

圖49A展示顯示HCC患者TBR34之經亞硫酸氫鹽處理之血漿DNA中之CNA (內環)及甲基化變化(外環)的Circos圖。圖49B為HCC患者TBR34之染色體得失區域及無複本數變化區域之甲基化z評分的圖。

圖50A及50B展示SLE患者SLE04及SLE10之血漿低甲基化及CNA分析之結果。

圖51A及51B展示對兩個HCC患者(TBR34及TBR36)之血漿之有CNA區域及無CNA區域的 $Z_{meth}$ 分析。圖51C及51D展示對兩個SLE患者(SLE04及SLE10)之血漿之有CNA區域及無CNA區域的 $Z_{meth}$ 分析。

圖52A展示使用CNA、整體甲基化及CpG島甲基化之A組特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。圖52B展示使用CNA、整體甲基化及CpG島甲基化之B組特徵的階層式叢集。

圖53A展示使用A組CpG島甲基化特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。圖53B展示使用A組整體甲基化密度對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。

圖54A展示使用A組整體CNA對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。圖54B展示使用B組CpG島

甲基化密度對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。

圖55A展示使用B組整體甲基化密度對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。圖55B展示使用B組整體甲基化密度對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。

圖56展示32個健康個體中之1 Mb區間(紅色點)之平均甲基化密度。  
定義

「甲基化模式組」提供基因組中之複數個位點或基因座處DNA甲基化之量的量度。甲基化模式組可對應於所有基因組、相當大部分之基因組或相對較小部分之基因組。「胎兒甲基化模式組」對應於妊娠女性之胎兒之甲基化模式組。胎兒甲基化模式組可使用多種胎兒組織或胎兒DNA來源來測定，包括胎盤組織及母體血漿中之細胞游離胎兒DNA。「腫瘤甲基化模式組」對應於生物體(例如，人類)之腫瘤之甲基化模式組。腫瘤甲基化模式組可使用腫瘤組織或母體血漿中之細胞游離腫瘤DNA來測定。胎兒甲基化模式組及腫瘤甲基化模式組為相關甲基化模式組之實例。相關甲基化模式組之其他實例為可向體液(例如血漿、血清、汗液、唾液、尿液、生殖道分泌物、精液、大便液、腹瀉液、腦脊液、胃腸道分泌物、胰臟分泌物、小腸分泌物、痰液、淚液、來自乳房及甲狀腺之抽吸液等)中貢獻DNA之器官之甲基化模式組(例如，腦細胞、骨、肺、心臟、肌肉及腎等之甲基化模式組)。該等器官可為移植器官。

「血漿甲基化模式組」為自動物(例如人類)之血漿或血清中測定之甲基化模式組。血漿甲基化模式組為細胞游離甲基化模式組之一個實例，因

為血漿及血清包括細胞游離DNA。血漿甲基化模式組亦為混合甲基化模式組之一個實例，因為其為胎兒/母體甲基化模式組或腫瘤/患者甲基化模式組之混合物。「胎盤甲基化模式組」可由絨毛膜絨毛樣品(CVS)或胎盤組織樣品(例如，分娩後獲得)中測定。「細胞甲基化模式組」對應於自患者細胞(例如血細胞)測定之甲基化模式組。血細胞之甲基化模式組稱為血細胞甲基化模式組(或血液甲基化模式組)。

「位點」對應於單一位點，其可為單一鹼基位置或相關鹼基位置之群組，例如CpG位點。「基因座」可對應於包括多個位點之區域。基因座可僅包括一個位點，在該情形下，其將使得該基因座等效於位點。

各基因組位點(例如，CpG位點)之「甲基化指數」係指該位點處展示甲基化之序列讀段(read)相對於覆蓋該位點之讀段總數之比例。區域之「甲基化密度」為展示甲基化之區域內之位點處的讀段之數目除以覆蓋該區域中之該等位點之讀段總數。該等位點可具有特異性特徵，例如為CpG位點。因此，區域之「CpG甲基化密度」為展示CpG甲基化之讀段之數目除以覆蓋該區域中之CpG位點(例如，特定CpG位點、CpG島內之CpG位點或較大區域)之讀段總數。舉例而言，人類基因組中各100 kb區間之甲基化密度可由亞硫酸氫鹽處理(其對應於甲基化胞嘧啶)後CpG位點處未轉化之胞嘧啶總數與定位於100 kb區域之序列讀段所覆蓋之所有CpG位點之比例決定。亦可針對其他區間大小，例如50 kb或1 Mb等，來進行此分析。區域可為整個基因組或染色體或染色體之一部分(例如，染色體臂)。當一個區域僅包括CpG位點時，該CpG位點之甲基化指數與該區域之甲基化密度相同。「甲基化胞嘧啶之比例」係指顯示欲甲基化(例如在亞硫酸氫鹽轉化後未轉化)之胞嘧啶位點「C」之數目相對於該區域中所分析之胞

嘧啶殘基之總數，亦即包括除CpG情形之外的胞嘧啶。甲基化指數、甲基化密度及甲基化胞嘧啶之比例為「甲基化程度」之實例。

「甲基化概況」(亦稱為甲基化狀態)包括與區域之DNA甲基化相關之資訊。與DNA甲基化相關之資訊可包括(但不限於)CpG位點之甲基化指數、區域中CpG位點之甲基化密度、CpG位點在連續區域上之分佈、含有超過一個CpG位點之區域內之各個別CpG位點之甲基化模式或程度及非CpG甲基化。基因組之相當大部分之甲基化概況可被視為等效於甲基化模式組。哺乳動物基因組中之「DNA甲基化」通常係指甲基添加至CpG二核苷酸中之胞嘧啶殘基(亦即5-甲基胞嘧啶)之5'碳。在其他情形下，例如CHG及CHH，其中H為腺嘌呤、胞嘧啶或胸嘧啶，DNA甲基化可在胞嘧啶中發生。胞嘧啶甲基化亦可呈5-羥甲基胞嘧啶形式。亦已報導非胞嘧啶甲基化，諸如N6-甲基腺嘌呤。

「組織」對應於任何細胞。不同類型組織可對應於不同類型細胞(例如，肝臟、肺或血液)，但亦可對應於來自不同生物體(母體相對於胎兒)之組織或健康細胞相對於腫瘤細胞。「生物樣品」係指取自個體(例如，人類，諸如孕婦、患有癌症之人士或疑似患有癌症之人士、器官移植接受者或疑似存在涉及器官(例如，心肌梗塞中之心臟或中風中之腦)之疾病過程之個體)且含有一或多個相關核酸分子之任何樣品。生物樣品可為體液，諸如血液、血漿、血清、尿液、陰道液、子宮或陰道沖洗液、胸腔積液、腹水、腦脊液、唾液、汗液、淚液、痰液、支氣管肺泡灌洗液等。亦可使用大便樣品。

術語「癌症程度」可指是否存在癌症、癌症之階段、腫瘤之大小、是否存在轉移、身體之總腫瘤負荷及/或癌症嚴重程度之其他量度。癌症

程度可為數字或其他特徵。該程度可為零。癌症程度亦包括與突變或許多突變相關的惡化前期或癌症前期病狀(狀態)。癌症程度可以各種方式使用。舉例而言，篩選可檢查癌症是否存在於先前不知患有癌症之某人中。評定可研究經診斷患有癌症之某人以監測癌症隨時間之進展、研究療法之效用或確定預後。在一個實施例中，預後可表述為患者死於癌症之概率，或癌症在特定持續時間或時間後進展之概率，或癌症轉移之概率。偵測可意謂『篩選』，或可意謂檢查具有癌症之暗示性特徵(例如症狀或其他陽性測試)之某人是否患有癌症。

### 【實施方式】

相關申請案之交叉參考

本申請案為主張以下各案之優先權的PCT申請案：2013年6月3日申請且標題為「Tumor Detection In Plasma Using Methylation Status And Copy Number」之美國臨時專利申請案第61/830,571號；及2013年3月15日申請且標題為「Non-Invasive Determination Of Methylome Of Fetus Or Tumor From Plasma」之美國申請案第13/842,209號，該申請案為2012年9月20日申請且標題為「Method Of Determining The Whole Genome DNA Methylation Status Of The Placenta By Massively Parallel Sequencing Of Maternal Plasma」之美國臨時專利申請案第61/703,512號之非臨時申請案且主張其權益，各案係出於所有目的而以全文引用的方式併入本文中。

表觀遺傳機制在胚胎及胎兒發育中起重要作用。然而，人類胚胎及胎兒組織(包括胎盤組織)不容易獲得(美國專利6,927,028)。某些實施例已藉由分析具有母體循環中所存在之細胞游離胎兒DNA分子之樣品而解決



此問題。可用多種方式推斷胎兒甲基化模式組。舉例而言，可將母體血漿甲基化模式組與細胞甲基化模式組(來自母親之血細胞)相比較且顯示差異與胎兒甲基化模式組相關。作為另一實例，胎兒特異性等位基因可用於測定特異性基因座處之胎兒甲基化模式組之甲基化。另外，在顯示大小與甲基化百分比之間的相關性時，片段大小可用作甲基化百分比之指標。

在一個實施例中，使用全基因組亞硫酸氫鹽定序來分析單一核苷酸解析度下母體血漿DNA之甲基化概況(部分或整個甲基化模式組)。藉由利用母親與胎兒之間的多態性差異，可由母體血液樣品收集胎兒甲基化模式組。在另一實施例中，未使用多態性差異，但可使用血漿甲基化模式組與血細胞甲基化模式組之間的差異。

在另一實施例中，藉由利用腫瘤基因組與非腫瘤基因組之間的單一核苷酸變化及/或複本數畸變及來自血漿(或其他樣品)之定序資料，可在疑似或已知患有癌症之患者之樣品中進行腫瘤之甲基化概況分析。當與一個健康對照或一組健康對照之血漿甲基化程度相比較時，測試個體之血漿樣品中之甲基化程度差異可允許測試個體鑑別為患有癌症。另外，甲基化概況可充當顯示該人士已罹患之癌症之類型(例如，來自何種器官)及是否已發生轉移的標誌。

由於該方法之非侵入性，吾等能夠連續評定來自在前三個月、後三個月及分娩後收集之母體血液樣品之胎兒及母體血漿甲基化模式組。觀測到妊娠相關之變化。該方法亦可應用於在中三個月期間獲得之樣品。由妊娠期間之母體血漿推斷之胎兒甲基化模式組類似於胎盤甲基化模式組。由母體血漿資料鑑別印記基因及差異性甲基化區域。

因此，吾等已開發一種用於非侵入性、連續且全面研究胎兒甲基化

模式組之方法，因此提供鑑別生物標記或直接測試妊娠相關病變之可能性。實施例亦可用以非侵入性、連續且全面研究腫瘤甲基化模式組，以便篩選或偵測個體是否罹患癌症、監測癌症患者之惡性疾病及預測。實施例可應用於任何癌症類型，包括(但不限於)肺癌、乳癌、結腸直腸癌、前列腺癌、鼻咽癌、胃癌、睪丸癌、皮膚癌(例如黑色素瘤)、影響神經系統之癌症、骨癌、卵巢癌、肝癌(例如肝細胞癌)、血液科惡性疾病、胰臟癌、子宮內膜癌、腎癌、子宮頸癌、膀胱癌等。

首先論述如何測定甲基化模式組或甲基化概況之說明，且隨後描述不同的甲基化模式組(諸如胎兒甲基化模式組、腫瘤甲基化模式組、母親或患者之甲基化模式組及混合甲基化模式組，例如來自血漿)。隨後使用胎兒特異性標記物或藉由比較混合甲基化概況與細胞甲基化概況來描述胎兒甲基化概況之測定。藉由比較甲基化概況來測定胎兒甲基化標記物。論述大小與甲基化之間的關係。亦提供甲基化概況用於偵測癌症之用途。

## **I. 測定甲基化模式組**

已使用無數方法來研究胎盤甲基化模式組，但各方法均具有其限制。舉例而言，可將未甲基化胞嘧啶殘基修飾成尿嘧啶且留下甲基化胞嘧啶不變的化合物亞硫酸氫鈉將胞嘧啶甲基化之差異轉化成基因序列差異供進一步查詢。研究胞嘧啶甲基化之黃金標準法係基於用亞硫酸氫鈉處理組織DNA，隨後對經亞硫酸氫鹽轉化之DNA分子之個別純系直接定序。在分析多個DNA分子純系之後，可獲得每個CpG位點之胞嘧啶甲基化模式及定量概況。然而，選殖亞硫酸氫鹽定序為不能容易地應用於全基因組規模之低產量且勞動密集性程序。

通常消化未甲基化的DNA之甲基化敏感性限制酶提供一種用於研究

DNA甲基化之低成本方法。然而，由該等研究產生之資料受限於具有酶識別基元之基因座且結果並非定量的。由抗甲基化胞嘧啶抗體結合之DNA之免疫沈澱可用於調查基因組之較大區段，但趨向於由於結合於該等區域之抗體的較高強度而向具有緻密甲基化之基因座偏移。基於微陣列之方法視查詢探針之先驗設計及該等探針與靶DNA之間的雜交效率而定。

為全面查詢甲基化模式組，一些實施例使用大規模平行定序(MPS)基於每個核苷酸及每個等位基因提供全基因組資訊及甲基化程度之定量評定。近來，亞硫酸氫鹽轉化隨後進行全基因組MPS已具有可行性(R Lister等人2008 Cell; 133: 523-536)。

在應用全基因組亞硫酸氫鹽定序來研究人類甲基化模式組之少數已公開研究(R Lister等人2009 Nature; 462: 315-322 ; L Laurent等人2010 Genome Res; 20: 320-331 ; Y Li等人2010 PLoS Biol; 8: e1000533 ; 及 M Kulis等人2012 Nat Genet; 44: 1236-1242)中，兩項研究聚焦於胚胎幹細胞及胎兒纖維母細胞(R Lister等人2009 Nature; 462: 315-322 ; L Laurent等人2010 Genome Res; 20: 320-331)。兩項研究均分析來源於細胞株之DNA。

#### A. 全基因組亞硫酸氫鹽定序

某些實施例可克服上述挑戰並且使得能夠全面、非侵入性並且連續查詢胎兒甲基化模式組。在一個實施例中，使用全基因組亞硫酸氫鹽定序來分析在孕婦循環中發現之細胞游離胎兒DNA分子。儘管血漿DNA分子具有較低豐度及片段化性質，但吾等能夠由母體血漿收集高解析度胎兒甲基化模式組且連續觀測隨妊娠進展之變化。鑒於強烈關注非侵入性產前測

試(NIPT)，實施例可為胎兒生物標記發現提供有效新工具或充當用於達成對胎兒或妊娠相關疾病之NIPT的直接平台。現提供來自對各種樣品之全基因組亞硫酸氫鹽定序之資料，由其可獲得胎兒甲基化模式組。在一個實施例中，該技術可應用於併發子癩先兆或子宮內生長停滯或早產之妊娠中的甲基化概況分析。對於該等有併發症之妊娠，該技術由於其非侵入性而可連續使用，從而允許監測及/或預測及/或對治療之反應。

圖1A展示根據本發明實施例之母體血液、胎盤及母體血漿之定序結果之表100。在一個實施例中，對使用甲基化DNA文庫接附子(Illumina)製備(R Lister等人2008 Cell; 133: 523-536)之前三個月收集之血液樣品之血細胞、CVS、妊娠期收集之胎盤組織、在前三個月及後三個月及產後期過程中收集之母體血漿樣品之經亞硫酸氫鹽轉化之DNA文庫進行完整基因組定序。亦分析獲自一個成年男性及一個成年未妊娠女性之血細胞及血漿DNA樣品。在此研究中產生總計95億對原始序列讀段。各樣品之定序覆蓋率示於表100中。

對於前三個月、後三個月及分娩後母體血漿樣品，可唯一定位於人類參考基因組之序列讀段分別達到平均單倍體基因組覆蓋率之50倍、34倍及28倍。對於獲自妊娠之樣品，基因組中之CpG位點之覆蓋率在81%至92%範圍內。對於前三個月、後三個月及分娩後母體血漿樣品，跨越CpG位點之序列讀段分別總計達平均單倍體覆蓋率之每股33倍、每股23倍及每股19倍。所有樣品之亞硫酸氫鹽轉化效率均>99.9% (表100)。

在表100中，含糊率(標記為「a」)係指定位於參考人類基因組之沃森(Watson)及克立克(Crick)股兩者上之讀段的比例。 $\lambda$ 轉化率係指內部 $\lambda$ DNA對照中藉由亞硫酸氫鹽修飾轉化成「胸嘧啶」殘基之未甲基化胞嘧

嘧之比例。H一般等於A、C或T。「a」係指可定位於特定基因組基因座但不能指派至沃森或克立克股之讀段。「b」係指具有一致起點及終點座標之配對讀段。對於「c」，在亞硫酸氫鹽轉化之前將 $\lambda$  DNA摻入各樣品中。 $\lambda$ 轉化率係指在亞硫酸氫鹽轉化之後仍呈胞嘧啶形式之胞嘧啶核苷酸之比例且用作成功亞硫酸氫鹽轉化率之指示。「d」係指存在於參考人類基因組中且在亞硫酸氫鹽轉化後仍呈胞嘧啶序列形式的胞嘧啶核苷酸之數目。

在亞硫酸氫鹽修飾期間，將未甲基化胞嘧啶轉化成尿嘧啶且隨後在PCR擴增後轉化成胸嘧啶，而甲基化胞嘧啶將保持完整(M Frommer等人 1992 Proc Natl Acad Sci USA;89:1827-31)。在CpG情形下，在定序並比對後，可因此根據胞嘧啶殘基處之甲基化序列讀段「M」(甲基化)之計數及未甲基化序列讀段「U」(未甲基化)之計數來推斷個別CpG位點之甲基化狀態。使用亞硫酸氫鹽定序資料，構築母體血液、胎盤及母體血漿之整個甲基化模式組。可使用以下等式計算母體血漿中特定基因座之平均甲基化CpG密度(亦稱為甲基化密度MD)：

$$MD = \frac{M}{M + U}$$

其中M為基因座內CpG位點處之甲基化讀段之計數且U為未甲基化讀段之計數。若基因座內存在超過一個CpG位點，則M及U對應於所有位點之計數。

## B. 各種技術

如上文所述，可使用對經亞硫酸氫鹽轉化之血漿DNA之大規模平行定序(MPS)進行甲基化概況分析。經亞硫酸氫鹽轉化之血漿DNA之MPS可用隨機或鳥槍方式進行。定序深度可根據相關區域之大小而變化。

在另一實施例中，可首先使用基於溶液相或固體相雜交之製程捕捉經亞硫酸氫鹽轉化之血漿DNA中之相關區域，隨後進行MPS。大規模平行定序可使用以下進行：諸如Illumina之邊合成邊定序(sequencing-by-synthesis)平台、諸如得自Life Technologies之SOLiD平台之邊接合邊定序(sequencing-by-ligation)平台、諸如得自Life Technologies之Ion Torrent或Ion Proton平台之基於半導體之定序系統、或諸如Helicos系統或Pacific Biosciences系統或基於奈米孔之定序系統之單一分子定序系統。基於奈米孔之定序包括使用例如脂質雙層及蛋白質奈米孔構築之奈米孔及固態奈米孔(諸如基於石墨烯之奈米孔)。因為所選單一分子定序平台將允許在不進行亞硫酸氫鹽轉化的情況下直接闡明DNA分子(包括N6-甲基腺嘌呤、5-甲基胞嘧啶及5-羥甲基胞嘧啶)之甲基化狀態(BA Flusberg等人2010 Nat Methods; 7: 461-465；J Shim等人2013 Sci Rep; 3:1389. doi: 10.1038/srep01389)，故使用該等平台將允許分析未經亞硫酸氫鹽轉化之樣品DNA (例如血漿DNA)之甲基化狀態。

除定序以外，亦可使用其他技術。在一個實施例中，甲基化概況分析可以藉由以下方式進行：進行甲基化特異性PCR或甲基化敏感性限制酶消化，隨後進行PCR或連接酶鏈式反應，隨後進行PCR。在其他實施例中，PCR為單一分子或數位PCR形式(B Vogelstein等人1999 Proc Natl Acad Sci USA; 96: 9236-9241)。在其他實施例中，PCR可為即時PCR。在其他實施例中，PCR可為多路PCR。

## II. 分析甲基化模式組

一些實施例可使用整個基因組亞硫酸氫鹽定序來測定血漿DNA之甲基化概況。胎兒之甲基化概況可藉由對母體血漿DNA樣品進行定序來測

定，如下文所述。因此，在妊娠期間非侵入性地獲得胎兒DNA分子(及胎兒甲基化模式組)，且隨著妊娠進展連續監測變化。由於定序資料之全面性，故吾等能夠在單一核苷酸解析度下以全基因組規模研究母體血漿甲基化模式組。

因為定序讀段之基因組座標為已知的，故此等資料使得吾人能夠研究基因組中之甲基化模式組或任何相關區域之總體甲基化程度並且在不同基因元件之間進行比較。另外，多個序列讀段覆蓋各CpG位點或基因座。現提供用於量測甲基化模式組之一些量度的描述。

#### A. 血漿DNA分子之甲基化

DNA分子以較低濃度且呈長度通常類似於單核小體單元之片段化形式存在於人類血漿中(YMD Lo等人2010 *Sci Transl Med*; 2: 61ra91；及YW Zheng等人2012 *Clin Chem*; 58: 549-558)。儘管存在此等限制，但全基因組亞硫酸氫鹽定序渠道能夠分析血漿DNA分子之甲基化。在其他實施例中，由於所選單一分子定序平台將允許在不進行亞硫酸氫鹽轉化的情況下直接闡明DNA分子之甲基化狀態(BA Flusberg等人2010 *Nat Methods*; 7: 461-465；J Shim等人2013 *Sci Rep*; 3:1389. doi: 10.1038/srep01389)，故使用該等平台將允許使用未經亞硫酸氫鹽轉化之血漿DNA來測定血漿DNA之甲基化程度或測定血漿甲基化模式組。該等平台可偵測N6-甲基腺嘌呤、5-甲基胞嘧啶及5-羥甲基胞嘧啶，由此可提供與不同甲基化形式之不同生物功能相關之經改良結果(例如經改良敏感性或特異性)。當應用實施例來偵測或監測特定病症，例如子癩先兆或特定類型癌症時，該等經改良結果可適用。

亞硫酸氫鹽定序亦可辯別不同甲基化形式。在一個實施例中，可包

括可區分5-甲基胞嘧啶與5-羥甲基胞嘧啶之附加步驟。一種此類方法為氧化性亞硫酸氫鹽定序(oxBS-seq)，其可在單鹼基解析度下闡述5-甲基胞嘧啶及5-羥甲基胞嘧啶之位置(MJ Booth等人2012 *Science*; 336: 934-937；MJ Booth等人2013 *Nature Protocols*; 8: 1841-1851)。在亞硫酸氫鹽定序中，5-甲基胞嘧啶及5-羥甲基胞嘧啶兩者均讀作胞嘧啶且因此無法鑑別。另一方面，在oxBS-seq中，藉由用高鈳酸鉀(KRuO<sub>4</sub>)處理將5-羥甲基胞嘧啶特異性氧化成5-甲醯基胞嘧啶，隨後使用亞硫酸氫鹽轉化將新形成之5-甲醯基胞嘧啶轉化成尿嘧啶將允許區分5-羥甲基胞嘧啶與5-甲基胞嘧啶。因此，5-甲基胞嘧啶之讀出可獲自單次oxBS-seq運作，且藉由與亞硫酸氫鹽定序結果比較來推斷5-羥甲基胞嘧啶含量。在另一實施例中，可使用Tet輔助之亞硫酸氫鹽定序(TAB-seq)來區分5-甲基胞嘧啶與5-羥甲基胞嘧啶(M Yu等人2012 *Nat Protoc*; 7: 2159-2170)。TAB-seq可在單一鹼基解析度下鑑別5-羥甲基胞嘧啶以及測定其在各修飾位點處之豐度。此方法涉及β-葡糖基轉移酶介導之5-羥甲基胞嘧啶保護(葡糖基化)及重組小鼠Tet1(mTet1)介導之5-甲基胞嘧啶氧化成5-羧基胞嘧啶。在隨後進行亞硫酸氫鹽處理及PCR擴增後，將胞嘧啶及5-羧基胞嘧啶(衍生自5-甲基胞嘧啶)兩者轉化成胸嘧啶(T)，而5-羥甲基胞嘧啶將讀作C。

圖1B展示根據本發明實施例之經定序樣品之1 Mb窗口中之甲基化密度。圖150為描繪基因組上之1 Mb窗口中之母體血漿及基因組DNA之甲基化密度的Circos圖。自外向內：染色體表意文字可在順時針方向上呈pter-qter定向(著絲點以紅色顯示)，母體血液(紅色)、胎盤(黃色)、母體血漿(綠色)、母體血漿中之共有讀段(藍色)及母體血漿中之胎兒特異性讀段(紫色)。母體血細胞、胎盤及母體血漿之總體CpG甲基化程度(亦即，密度水



準)可見於表100中。在整個基因組上，母體血細胞之甲基化程度一般高於胎盤之甲基化程度。

#### B. 比較亞硫酸氫鹽定序與其他技術

吾等使用大規模平行亞硫酸氫鹽定序來研究胎盤甲基化模式組。另外，吾等使用覆蓋人類基因組中之約480,000個CpG位點之寡核苷酸陣列平台(Illumina)來研究胎盤甲基化模式組(M Kulis等人2012 Nat Genet; 44: 1236-1242；及C Clark等人2012 PLoS One; 7: e50233)。在使用基於微珠晶片之基因型分析及甲基化分析之一個實施例中，使用Illumina HumanOmni2.5-8基因型分析陣列根據製造商方案進行基因型分析。使用Genome Studio軟體(Illumina)之GenCall演算法來檢出基因型。檢出率(call rate)高於99%。對於基於微陣列之甲基化分析，使用Zymo EZ DNA甲基化套組(Zymo Research, Orange, CA, USA)根據製造商關於Illumina Infinium甲基化分析之推薦用亞硫酸氫鈉處理基因組DNA (500至800 ng)。

根據Infinium HD甲基化分析方案在50 ng/μl下對4 μl經亞硫酸氫鹽轉化之基因組DNA進行甲基化分析。在Illumina iScan儀器上掃描雜交微珠晶片。藉由GenomeStudio (v2011.1)甲基化模組(v1.9.0)軟體，藉由針對內部對照進行標準化並且減去背景來分析DNA甲基化資料。個別CpG位點之甲基化指數由β值(β)表示，該值係使用甲基化與未甲基化等位基因之間的螢光強度比來計算：

$$\beta = \frac{\text{甲基化等位基因之強度}}{\text{未甲基化等位基因之強度} + \text{甲基化等位基因之強度} + 100}$$

對於表示於陣列上且定序至至少10倍之覆蓋率的CpG位點，吾等比較由陣列獲得之β值與藉由對同一位點進行定序而測定之甲基化指數。β值

表示呈覆蓋同一CpG位點之甲基化與未甲基化的探針之組合強度之比例形式的甲基化探針強度。各CpG位點之甲基化指數係指甲基化讀段相對於覆蓋該CpG之讀段總數的比例。

圖2A至2C展示可藉由兩個平台查詢之藉由Illumina Infinium人類甲基化450K微珠晶片陣列測定之 $\beta$ 值相對於藉由對相應CpG位點之全基因組亞硫酸氫鹽定序測定之甲基化指數的圖：(A)母體血細胞；(B)絨毛膜絨毛樣品；(C)足月胎盤組織。來自兩個平台之資料高度一致且母體血細胞、CVS及足月胎盤組織之皮爾遜相關係數(Pearson correlation coefficient)分別為0.972、0.939及0.954，且 $R^2$ 值分別為0.945、0.882及0.910。

吾等進一步比較吾等之定序資料與使用覆蓋約27,000個CpG位點之寡核苷酸陣列研究12對CVS及母體血細胞DNA樣品之甲基化概況之Chu等人所報導之資料(T Chu等人2011 PLoS One; 6: e14723)。CVS及母體血細胞DNA及先前研究中之12對樣品中之每一者的定序結果之間的相關性資料得到母體血液之平均皮爾森係數(0.967)及 $R^2$  (0.935)以及CVS之平均皮爾森係數(0.943)及 $R^2$  (0.888)。在兩個陣列上所表示之CpG位點中，吾等之資料與公開資料高度相關。母體血細胞、CVS及胎盤組織之非CpG甲基化率<1% (表100)。此等結果與以下當前信念一致：相當大量之非CpG甲基化主要侷限於多能細胞(R Lister等人2009 Nature; 462: 315-322；L Laurent等人2010 Genome Res; 20: 320-331)。

### C. 比較未妊娠個體之血漿與血液甲基化模式組

圖3A及3B展示自成年男性及未妊娠成年女性收集之血漿及血細胞中之甲基化CpG位點之百分比的直方圖：(A)常染色體；(B)染色體X。該圖展示男性及未妊娠女性之血漿與血液甲基化模式組之間的相似性。男性及

未妊娠女性血漿樣品中甲基化之CpG位點之總體比例與相應血細胞DNA幾乎相同(表100及圖2A及2B)。

吾等接下來以基因座特異性方式研究血漿及血細胞樣品之甲基化概況之相關性。吾等藉由測定CpG位點處未轉化胞嘧啶總數與定位於100 kb區域之序列讀段所覆蓋之所有CpG位點之比例來測定人類基因組中各100 kb區間的甲基化密度。男性血漿樣品及相應血細胞DNA以及女性樣品之間的甲基化密度高度一致。

圖4A及4B展示血細胞DNA及血漿DNA中之相應基因座之甲基化密度的圖：(A)未妊娠成年女性；(B)成年男性。未妊娠女性樣品之皮爾遜相關係數及 $R^2$ 值分別為0.963及0.927，且男性樣品之皮爾遜相關係數及 $R^2$ 值分別為0.953及0.908。此等資料與基於同種異體造血幹細胞移植接受者之血漿DNA分子之基因型評定之先前發現一致，該先前發現顯示造血細胞為人類血漿中之DNA之主要來源(YW Zheng等人2012 Clin Chem; 58: 549-558)。

#### D. 整個甲基化模式組之甲基化程度

吾等接下來研究母體血漿DNA、母體血細胞及胎盤組織之DNA甲基化程度以測定甲基化程度。測定重複序列區域、非重複序列區域及總體之程度。

圖5A及5B展示自妊娠收集之樣品中之甲基化CpG位點之百分比的直方圖：(A)常染色體；(B)染色體X。對於前三個月及後三個月母體血漿樣品，甲基化CpG之總體比例分別為67.0%及68.2%。與獲自未妊娠個體之結果不同，此等比例低於前三個月母體血細胞樣品但高於CVS及足月胎盤組織樣品(表100)。應注意，分娩後母體血漿樣品中之甲基化CpG之百分

比為73.1%，該值類似於血細胞資料(表100)。在分佈於所有常染色體以及染色體X上且跨越人類基因組之非重複序列區域及多類重複序列元件之CpG中觀測到此等趨勢。

發現胎盤中之重複序列及非重複序列元件相對於母體血細胞經低甲基化。該等結果與文獻中之以下發現一致：胎盤相對於其他組織(包括周邊血液細胞)經低甲基化。

介於71%至72%之間的經定序CpG位點在來自孕婦、非孕婦及成年男性之血細胞DNA中經甲基化(圖1之表100)。此等資料與Y Li等人2010 PLoS Biol; 8: e1000533所報導之血液單核細胞之68.4% CpG位點之報導相當。與關於胎盤組織之低甲基化性質之先前報導一致，CVS及足月胎盤組織中分別55%及59% CpG位點經甲基化(表100)。

圖6展示母體血液、胎盤及母體血漿之人類基因組之不同重複序列類別之甲基化程度的直方圖。重複序列類別如UCSC基因組瀏覽器所定義。所顯示之資料係來自前三個月樣品。與表明主要在基因組之某些重複序列類別中觀測到胎盤組織之低甲基化性質的早期資料(B Novakovic等人2012 Placenta; 33: 959-970)不同，在此吾等展示參考血細胞，胎盤實際上在大部分基因組元件類別中經低甲基化。

#### E. 甲基化模式組之相似性

實施例可使用同一平台來測定胎盤組織、血細胞及血漿之甲基化模式組。因此，可直接比較該等生物樣品類型之甲基化模式組。男性及未妊娠女性之血細胞與血漿甲基化模式組之間以及母體血細胞與分娩後母體血漿樣品之間的高度類似性進一步證實造血細胞為人類血漿中之主要DNA來源(YW Zheng等人2012 Clin Chem; 58: 549-558)。

就甲基化CpG在基因組中之總體比例以及血細胞DNA及血漿DNA中之相應基因座之間的甲基化密度高度相關性兩者而言，類似性顯而易見。又，當與母體血細胞資料或分娩後母體血漿樣品相比較時，前三個月及後三個月母體血漿樣品中之甲基化CpG之總體比例降低。妊娠期間甲基化程度降低係由於母體血漿中所存在之胎兒DNA分子之低甲基化性質。

分娩後母體血漿樣品中之甲基化概況逆轉從而變得與母體血細胞更類似表明胎兒DNA分子已自母體循環中移除。基於胎兒之SNP標記物計算胎兒DNA濃度實際上顯示濃度自分娩前之33.9%變化至分娩後樣品中之僅4.5%。

#### F. 其他應用

實施例已經由血漿DNA之MPS分析而成功收集DNA甲基化模式組。能夠自母體血漿測定胎盤或胎兒甲基化模式組提供一種非侵入性方法以測定、偵測及監測與妊娠相關病狀(諸如子癩先兆、子宮內生長限制、早產及其他)相關之異常甲基化概況。舉例而言，偵測疾病特異性異常甲基化標誌允許篩選、診斷及監測該等妊娠相關病狀。量測母體血漿甲基化程度允許篩選、診斷及監測該等妊娠相關病狀。除直接應用於妊娠相關病狀之研究以外，該方法可應用於血漿DNA分析相關之其他藥物領域。舉例而言，可由癌症患者之血漿DNA測定癌症之甲基化模式組。如本文所述，由血漿分析癌症甲基化模式組可能為一種由血漿分析癌症基因組之協同技術(KCA Chan等人2013 Clin Chem; 59: 211-224；及RJ Leary等人2012 Sci Transl Med; 4:162ra154)。

舉例而言，測定血漿樣品之甲基化程度可用於篩選癌症。當血漿樣品之甲基化程度與健康對照相比顯示異常程度時，可疑似患有癌症。隨後

可藉由測定不同基因組基因座處之血漿甲基化概況或藉由血漿基因組分析以偵測腫瘤相關複本數畸變、染色體易位及單一核苷酸變異體來進一步確定及評估癌症類型或癌症之組織來源。實際上，在本發明之一個實施例中，可同時進行血漿癌症甲基化模式組及基因組概況分析。或者，基於血漿甲基化程度分析，放射性及成像研究(例如電腦斷層攝影術、磁共振成像、正電子發射斷層攝影術)或內視鏡檢法(例如上胃腸道內視鏡檢法或結腸鏡檢法)可用於進一步研究疑似患有癌症之個體。

對於癌症篩選或偵測，測定血漿(或其他生物學)樣品之甲基化程度可與諸如前列腺特异性抗原量測(例如針對前列腺癌)、癌胚抗原(例如針對結腸直腸癌、胃癌、胰臟癌、肺癌、乳癌、甲狀腺髓質癌)、甲胎蛋白(例如針對肝癌或生殖細胞腫瘤)、CA125 (例如針對卵巢癌及乳癌)及CA19-9 (例如針對胰臟癌)之其他癌症篩選或偵測用模態聯合使用。

另外，可對其他組織進行定序以獲得細胞甲基化模式組。舉例而言，可分析肝組織以測定對肝臟之甲基化模式特異性，該對肝臟之甲基化模式特異性可用於鑑別肝臟病變。亦可加以分析之其他組織包括腦細胞、骨、肺、心臟、肌肉及腎等。各種組織之甲基化概況可時常變化，例如，由於發育、老化、疾病過程(例如發炎或肝硬化或自體免疫過程(諸如在全身性紅斑狼瘡中))或治療(例如用諸如5-氮雜胞嘧啶核苷及5-氮雜脫氧胞嘧啶核苷之脫甲基劑治療)。DNA甲基化之動態性質使得此種分析可能對監測生理及病理過程非常有價值。舉例而言，若在個體血漿甲基化模式組中偵測到與當其健康時所獲得之基線值相比有所變化，則隨後可偵測器官中對血漿DNA有貢獻之疾病過程。

此外，可由器官移植接受者之血漿DNA測定移植器官之甲基化模式

組。如本發明中所述，由血漿分析移植物甲基化模式組可能為一種由血漿分析移植物基因組之協同技術(YW Zheng等人, 2012；YMD Lo等人1998 Lancet; 351: 1329-1330；及TM Snyder等人2011 Proc Natl Acad Sci USA; 108: 6229-6234)。由於血漿DNA一般被視為細胞死亡之標記物，故自移植器官釋放之DNA之血漿含量增加可用作該器官之細胞死亡增加的標記物，諸如排斥反應發作或涉及該器官之其他病理過程(例如感染或膿腫)。在成功實行抗排斥療法之情況下，預期由移植器官釋放之DNA之血漿含量將減少。

### **III.使用SNP測定胎兒或腫瘤甲基化模式組**

如上所述，血漿甲基化模式組對應於未妊娠正常人之血液甲基化模式組。然而，對於妊娠女性，甲基化模式組不同。在大多數母體DNA背景中，胎兒DNA分子在母體血漿中循環(YMD Lo等人1998 Am J Hum Genet; 62: 768-775)。因此，對於妊娠女性，血漿甲基化模式組在很大程度上為胎盤甲基化模式組與血液甲基化模式組之複合物。因此，可自血漿提取胎盤甲基化模式組。

在一個實施例中，使用母親與胎兒之間的單核苷酸多態性(SNP)差異來鑑別母體血漿中之胎兒DNA分子。一個目的在於鑑別母親為同型組合但胎兒為異型組合的SNP基因座；胎兒特異性等位基因可用於測定何種DNA片段來自胎兒。使用SNP基因型分析陣列Illumina HumanOmni2.5-8來分析來自母體血細胞之基因組DNA。另一方面，對於母親為異型組合且胎兒為同型組合的SNP基因座，則可使用母親特異性SNP等位基因來測定何種血漿DNA片段來自母親。該等DNA片段之甲基化程度將體現母親中相關基因組區域的甲基化程度。

#### A. 胎兒特異性讀段之甲基化與胎盤甲基化模式組之相關性

由生物樣品之定序結果鑑別具有兩個不同等位基因之基因座，其中一個等位基因(B)之量顯著少於另一等位基因(A)。覆蓋B等位基因之讀段被認為具胎兒特異性(胎兒特異性讀段)。確定母親為A同型組合且胎兒為A/B異型組合，且因此覆蓋A等位基因之讀段由母親及胎兒共有(共有讀段)。

在所分析之用於說明本發明中之若干概念的一種妊娠情況中，發現妊娠母親在常染色體上之1,945,516個基因座處為同型組合。檢驗覆蓋此等SNP之母體血漿DNA定序讀段。在107,750個基因座處偵測到攜帶非母體等位基因之讀段，且此等被視為資訊基因座。在各資訊SNP處，並非來自母親之等位基因稱為胎兒特異性等位基因，而另一等位基因稱為共有等位基因。

可測定母體血漿中之部分胎兒/腫瘤DNA濃度(亦稱為胎兒DNA百分比)。在一個實施例中，母體血漿中之部分胎兒DNA濃度f係由以下等式測定：

$$f = \frac{2p}{p+q}$$

其中p為具有胎兒特異性等位基因之定序讀段的數目且q為具有母親與胎兒之間的共有等位基因之定序讀段的數目(YMD Lo等人2010 Sci Transl Med; 2: 61ra91)。發現前三個月、後三個月及分娩後母體血漿樣品中之胎兒DNA比例分別為14.4%、33.9%及4.5%。亦使用與染色體Y比對之讀段之數目來計算胎兒DNA比例。基於染色體Y資料，在前三個月、後三個月及分娩後母體血漿樣品中，結果分別為14.2%、34.9%及3.7%。

藉由獨立地分析胎兒特異性或共有定序讀段，實施例顯示循環胎兒



DNA分子之低甲基化程度比背景DNA分子大得多。胎兒特異性母體血漿讀段中相應基因座之甲基化密度與前三個月及後三個月兩者之胎盤組織資料的比較顯示高度相關性。與基於來源於所選基因座之資訊的先前證據相比，此等資料提供胎盤為母體血漿中來源於胎兒之DNA分子之主要來源的基因組含量證據且向進了一大步。

吾等使用覆蓋與資訊SNP相鄰之CpG位點的胎兒特異性或共有讀段來測定基因組中之各1 Mb區域之甲基化密度。自母體血漿序列讀段收集之胎兒及非胎兒特異性甲基化模式組可顯示於例如Circos圖中(M Krzywinski等人2009 *Genome Res*; 19: 1639-1645)。亦測定母體血細胞及胎盤組織樣品之每1 Mb區間之甲基化密度。

圖7A展示前三個月樣品之Circos圖700。圖7B展示後三個月樣品之Circos圖750。圖700及750展示每1 Mb區間之甲基化密度。染色體表意文字(最外環)在順時針方向上呈pter-qter定向(著絲點以紅色顯示)。第二最外部軌道顯示相應1 Mb區域中之CpG位點數。紅色條之標度顯示為每1 Mb區間多達20,000個位點。基於中心所示之色彩方案，在其他軌道中展示相應1 Mb區域之甲基化密度。

對於前三個月樣品(圖7A)，由內向外，軌道為：絨毛膜絨毛樣品、母體血漿中之胎兒特異性讀段、母體血漿中之母體特異性讀段、母體血漿中之胎兒與非胎兒組合讀段及母體血細胞。對於後三個月樣品(圖7B)，軌道為：足月胎盤組織、母體血漿中之胎兒特異性讀段、母體血漿中之母體特異性讀段、母體血漿中之胎兒與非胎兒組合讀段、分娩後母體血漿及母體血細胞(來自前三個月血液樣品)。可瞭解對於前三個月及後三個月血漿樣品兩者，胎兒甲基化模式組之低甲基化程度大於非胎兒特異性甲基化模

式組。

胎兒甲基化模式組之總體甲基化概況更緊密類似CVS或胎盤組織樣品。相反，血漿中之共有讀段(主要為母體DNA)之DNA甲基化概況更緊密類似於母體血細胞。吾等隨後對母體血漿DNA讀段與母體或胎兒組織之甲基化密度進行系統性逐基因座比較。吾等測定作為資訊SNP存在於同一序列讀段上且覆蓋至少5個母體血漿DNA序列讀段之CpG位點之甲基化密度。

圖8A至8D展示比較資訊性單核苷酸多態性周圍之CpG位點之基因組組織DNA相對於母體血漿DNA之甲基化密度的圖。圖8A展示前三個月母體血漿樣品中胎兒特異性讀段之甲基化密度相對於CVS樣品中讀段之甲基化密度。如可見，胎兒特異性值良好對應於CVS值。

圖8B展示後三個月母體血漿樣品中胎兒特異性讀段之甲基化密度相對於足月胎盤組織中讀段之甲基化密度。密度組又良好對應，表明可藉由分析具有胎兒特異性等位基因之讀段來獲得胎兒甲基化概況。

圖8C展示前三個月母體血漿樣品中共有讀段之甲基化密度相對於母體血細胞中讀段之甲基化密度。鑒於大部分共有讀段來自母親，故兩組值良好對應。圖8D展示後三個月母體血漿樣品中共有讀段之甲基化密度相對於母體血細胞中讀段之甲基化密度。

對於母體血漿中之胎兒特異性讀段，前三個月母體血漿與CVS之間的斯皮爾曼相關係數(Spearman correlation coefficient)為0.705 ( $P < 2.2 \times 10^{-16}$ )；且後三個月母體血漿與足月胎盤組織之間的斯皮爾曼相關係數為0.796 ( $P < 2.2 \times 10^{-16}$ ) (圖8A及8B)。對母體血漿與母體血細胞資料中之共有讀段進行類似比較。前三個月血漿樣品之皮爾遜相關係數為

0.653 ( $P < 2.2 \times 10^{-16}$ ) 且後三個月血漿樣品之皮爾遜相關係數為0.638 ( $P < 2.2 \times 10^{-16}$ ) (圖8C及8D)。

### B. 胎兒甲基化模式組

在一個實施例中，為自母體血漿中收集胎兒甲基化模式組，吾等淘選跨越至少一個資訊性胎兒SNP位點且在同一讀段內含有至少一個CpG位點的序列讀段。顯示胎兒特異性等位基因之讀段包括在胎兒甲基化模式組之集合中。顯示共有等位基因，亦即非胎兒特異性等位基因之讀段包括在主要包含來源於母體之DNA分子的非胎兒特異性甲基化模式組之集合中。

對於前三個月母體血漿樣品，胎兒特異性讀段覆蓋常染色體上之218,010個CpG位點。後三個月及分娩後母體血漿樣品之相應個數分別為263,611及74,020。平均而言，共有讀段分別平均覆蓋33.3、21.7及26.3倍該等CpG位點。對於前三個月、後三個月及分娩後母體血漿樣品，胎兒特異性讀段分別覆蓋3.0、4.4及1.8倍該等CpG位點。

胎兒DNA代表母體血漿中之微量群體，且因此胎兒特異性讀段對該等CpG位點之覆蓋率與樣品之胎兒DNA百分比成比例。對於前三個月母體血漿樣品，甲基化CpG在胎兒讀段中之總體百分比為47.0%，而共有讀段為68.1%。對於後三個月母體血漿樣品，胎兒讀段之甲基化CpG之百分比為53.3%，而共有讀段為68.8%。此等資料顯示，母體血漿中之胎兒特異性讀段之低甲基化程度大於母體血漿中之共有讀段。

### C. 方法

上述技術亦可用於測定腫瘤甲基化概況。現描述用於測定胎兒及腫瘤甲基化概況之方法。

圖9為說明根據本發明實施例之用於測定生物體之生物樣品之第一甲基化概況之方法900的流程圖。方法900可根據母體血漿之甲基化概況構築胎兒之表觀遺傳圖譜。該生物樣品包括細胞游離DNA，該細胞游離DNA包含來源於第一組織及第二組織之細胞游離DNA之混合物。作為實例，第一組織可來自胎兒、腫瘤或移植器官。

在方塊910處，分析來自生物樣品之複數個DNA分子。分析DNA分子可包括測定DNA分子在生物體基因組中之位置、測定DNA分子之基因型及測定DNA分子是否在一或多個位點甲基化。

在一個實施例中，使用DNA分子之序列讀段來分析DNA分子，其中定序獲知甲基化。因此，序列讀段包括來自生物樣品之DNA分子之甲基化狀態。甲基化狀態可包括特定胞嘧啶殘基是5-甲基胞嘧啶或是5-羥甲基胞嘧啶。序列讀段可獲自各種定序技術、PCR技術、陣列及適用於鑑別片段之序列的其他技術。可如本文所述獲得序列讀段之位點之甲基化狀態。

在方塊920處，鑑別複數個第一基因座，其中第一組織之第一基因組對於各別第一等位基因及各別第二等位基因為異型組合且第二組織之第二基因組對於各別第一等位基因為同型組合。舉例而言，可在複數個第一基因座處鑑別胎兒特異性讀段。或者，可在複數個第一基因座處鑑別腫瘤特異性讀段。可由第二等位基因之序列讀段之百分比處於特定範圍(例如約3%至25%)內的定序讀段中鑑別組織特異性讀段，藉此指示來自該基因座處之異型組合基因組之DNA片段之少數群體及來自該基因座處之同型組合基因組之多數群體。

在方塊930處，分析位於各第一基因座之一或多個位點處的DNA分子。測定在某一位點處經甲基化且對應於基因座之各別第二等位基因的許

多DNA分子。每個基因座可存在超過一個位點。舉例而言，SNP可指示片段具胎兒特異性，且該片段可具有多個甲基化狀態已測定之位點。可測定各位點處經甲基化之讀段的數目，且可測定基因座之甲基化讀段之總數。

基因座可由特定數目之位點、一組特定位點或包含組織特異性等位基因之變化形式周圍之區域的特定大小定義。基因座可僅具有一個位點。該等位點可具有特異性，例如，作為CpG位點。測定許多未甲基化的讀段為等效的，且涵蓋在測定甲基化狀態內。

在方塊940處，對於各第一基因座，基於在基因座之一或多個位點處經甲基化且對應於基因座之各別第二等位基因之DNA分子之數目來計算甲基化密度。舉例而言，可測定對應於基因座之CpG位點的甲基化密度。

在方塊950處，由第一基因座之甲基化密度產生第一組織之第一甲基化概況。第一甲基化概況可對應於特定位點，例如CpG位點。甲基化概況可針對具有胎兒特異性等位基因之所有基因座或僅針對該等基因座中的一些。

#### IV. 使用血漿與血液甲基化模式組之差異

上文顯示來自血漿之胎兒特異性讀段與胎盤甲基化模式組相關。因為母體血漿甲基化模式組之母體組分主要由血細胞貢獻，故血漿甲基化模式組與血液甲基化模式組之間的差異可用於測定所有基因座之胎盤甲基化模式組，而不僅是胎兒特異性等位基因之位置。血漿甲基化模式組與血液甲基化模式組之間的差異亦可用以測定腫瘤之甲基化模式組。

##### A. 方法

圖10為說明根據本發明實施例之用於測定生物體之生物樣品之第一

甲基化概況之方法1000的流程圖。生物樣品(例如血漿)包括細胞游離DNA，該細胞游離DNA包含來源於第一組織及第二組織之細胞游離DNA之混合物。第一甲基化概況對應於第一組織(例如，胎兒組織或腫瘤組織)之甲基化概況。方法1200可由母體血漿推斷差異性甲基化區域。

在方塊1010處，接收生物樣品。在機器(例如，定序機器)處簡單地接收生物樣品。生物樣品可呈取自生物體之形式或可呈經處理形式，例如，樣品可為自血液樣品提取之血漿。

在方塊1020處，獲得對應於第二組織之DNA的第二甲基化概況。可自記憶體讀取第二甲基化概況，因為其可能先前已測定。可由第二組織測定第二甲基化概況，例如僅含有或主要含有第二組織之細胞的不同樣品。第二甲基化概況可對應於細胞甲基化概況且獲自細胞DNA。作為另一實例，可由在妊娠之前或在癌症發展之前收集之血漿樣品測定第二概況，因為未患癌症之非妊娠人士之血漿甲基化模式組極類似於血細胞之甲基化模式組。

第二甲基化概況可提供生物體基因組中之複數個基因座中之每一者處的甲基化密度。特定基因座處之甲基化密度對應於第二組織之甲基化DNA之比例。在一個實施例中，甲基化密度為CpG甲基化密度，其中與基因座相關之CpG位點用於測定甲基化密度。若基因座存在一個位點，則甲基化密度可等於甲基化指數。甲基化密度亦對應於未甲基化密度，因為該兩個值互補。

在一個實施例中，藉由對來自生物體樣品之細胞DNA進行可獲知甲基化之定序來獲得第二甲基化概況。可獲知甲基化之定序之一個實例包括用亞硫酸氫鈉處理DNA且隨後進行DNA定序。在另一實例中，可獲知甲

基化之定序可在不使用亞硫酸氫鈉的情況下藉由以下方式來進行：使用單一分子定序平台，該定序平台將在不進行亞硫酸氫鹽轉化的情況下直接闡明DNA分子(包括N6-甲基腺嘌呤、5-甲基胞嘧啶及5-羥甲基胞嘧啶)之甲基化狀態(AB Flusberg等人2010 Nat Methods; 7: 461-465；J Shim等人2013 Sci Rep; 3:1389. doi: 10.1038/srep01389)；或藉由甲基化胞嘧啶之免疫沈澱，例如藉由使用針對甲基胞嘧啶之抗體或藉由使用甲基化DNA結合蛋白或肽(LG Acevedo等人2011 Epigenomics; 3: 93-101)，隨後定序；或藉由使用甲基化敏感性限制酶，隨後定序。在另一實施例中，使用非定序技術，諸如陣列、數位PCR及質譜。

在另一實施例中，可預先自個體或其他個體之對照樣品獲得第二組織之第二甲基化密度。來自另一個體之甲基化密度可充當具有參考甲基化密度之參考甲基化概況。可自多個樣品測定參考甲基化密度，其中基因座處之不同甲基化密度之平均程度(或其他統計值)可用作該基因座處之參考甲基化密度。

在方塊1030處，自混合物之細胞游離DNA測定細胞游離甲基化概況。細胞游離甲基化概況提供複數個基因座中之每一者處的甲基化密度。可藉由接收來自細胞游離DNA之定序的序列讀段來測定細胞游離甲基化概況，其中甲基化資訊係利用序列讀段獲得。可用與細胞甲基化模式組相同之方式測定細胞游離甲基化概況。

在方塊1040處，測定來自生物樣品中之第一組織之細胞游離DNA之百分比。在一個實施例中，第一組織為胎兒組織，且相應DNA為胎兒DNA。在另一實施例中，第一組織為腫瘤組織，且相應DNA為腫瘤DNA。可用多種方式測定百分比，例如使用胎兒特異性等位基因或腫瘤

特異性等位基因。複本數亦可用於測定百分比，例如，如2013年3月13日申請之標題為「**Mutational Analysis Of Plasma DNA For Cancer Detection**」之美國專利申請案13/801,748中所述，該申請案係以引用的方式併入本文中。

在方塊1050處，鑑別複數個基因座以供測定第一甲基化模式組。此等基因座可對應於用於測定細胞游離甲基化概況及第二甲基化概況之基因座中每一者。因此，複數個基因座可對應。有可能可使用更多基因座來測定細胞游離甲基化概況及第二甲基化概況。

在一些實施例中，例如可使用母體血細胞鑑別第二甲基化概況中經高甲基化或經低甲基化之基因座。為鑑別母體血細胞中經高甲基化之基因座，可自染色體之一端掃描甲基化指數 $\geq X\%$  (例如80%)之CpG位點。隨後可在下游區域內(例如，在下游200 bp內)搜尋下一CpG位點。若下一下游CpG位點之甲基化指數亦 $\geq X\%$  (或其他規定量)，則可將第一及第二CpG位點分組。分組可持續，直至下一下游區域內無其他CpG位點；或下一下游CpG位點之甲基化指數 $< X\%$ 。若經分組CpG位點之區域含有至少五個緊密相鄰之高甲基化CpG位點，則該區域可報導為在母體血細胞中經高甲基化。可進行類似分析以在母體血細胞中經低甲基化之基因座中搜尋甲基化指數 $\leq 20\%$ 之CpG位點。可計算候選基因座之第二甲基化概況之甲基化密度且用於推斷相應基因座之第一甲基化概況(例如胎盤組織甲基化密度)，例如根據母體血漿亞硫酸氫鹽定序資料。

在方塊1060處，藉由計算差異性參數(包括複數個基因座中每一者之第二甲基化概況之甲基化密度與細胞游離甲基化概況之甲基化密度之間的差異)來測定第一組織之第一甲基化概況。該差異以百分比定標。



在一個實施例中，第一(例如胎盤)組織中之基因座之第一甲基化密度(D)係使用以下等式推斷：

$$D = mbc - \frac{(mbc - mp)}{f * CN} \quad (1)$$

其中 $mbc$ 表示基因座(例如候選基因座，如母體血細胞亞硫酸氫鹽定序資料中所測定)處之第二甲基化概況之甲基化密度； $mp$ 表示母體血漿亞硫酸氫鹽定序資料中相應基因座之甲基化密度； $f$ 表示來自第一組織之細胞游離DNA之百分比(例如，部分胎兒DNA濃度)，且 $CN$ 表示該基因座處之複本數(例如，相對於正常情況，較高擴增值或較低缺失數)。若第一組織中不存在擴增或缺失，則 $CN$ 可為一。對於三體性(或腫瘤或胎兒中之區域複製)， $CN$ 將為1.5 (因為自2個複本增至3個複本)且單體性將具有0.5。較高擴增可以0.5之增量增加。在此實例中， $D$ 可對應於差異性參數。

在方塊1070處，轉換第一甲基化密度以獲得第一組織之經修正第一甲基化密度。該轉換可解決第一組織之差異性參數與實際甲基化概況之間的固定差異。舉例而言，該等值可相差固定常數或某一斜率。該轉換可為線性或非線性的。

在一個實施例中，發現推斷值 $D$ 之分佈低於胎盤組織之實際甲基化程度。舉例而言，推斷值可使用來自CpG島之資料進行線性轉換，該等CpG島為過度呈現CpG位點之基因組區段。本研究中所使用之CpG島之基因組位置係獲自UCSC基因組瀏覽器資料庫(NCBI版本36/hg18) (PA Fujita等人2011 *Nucleic Acids Res*; 39: D876-882)。舉例而言，CpG島可定義為GC含量 $\geq 50\%$ 、基因組長度 $> 200$  bp且觀測/預期CpG數之比率 $> 0.6$ 的基因組區段(M Gardiner-Garden等人1987 *J Mol Biol*; 196: 261-282)。

在一個實施例中，為導出線性轉換等式，定序樣品中可包括具有至

少4個CpG位點且每個CpG位點之平均讀取深度 $\geq 5$ 的CpG島。在確定CVS或妊娠期胎盤中之CpG島之甲基化密度與推斷值D之間的線性關係之後，使用以下等式來測定預測值：

$$\text{前三個月預測值} = D \times 1.6 + 0.2$$

$$\text{後三個月預測值} = D \times 1.2 + 0.05$$

## B. 胎兒實例

如上文所提及，方法1000可用於由母體血漿推斷胎盤之甲基化型態。血漿中之循環DNA主要來源於造血細胞。由其他內臟貢獻之細胞游離DNA之比例仍為未知的。此外，來源於胎盤之細胞游離DNA佔母體血漿中之總DNA之約5%-40%，其平均值為約15%。因此，可假定母體血漿中之甲基化程度等效於現存背景甲基化加妊娠期間之胎盤貢獻，如上文所述。

母體血漿甲基化程度MP可使用以下等式測定：

$$MP = BKG \times (1 - f) + PLN \times f$$

其中BKG為來源於血細胞及內臟之血漿中的背景DNA甲基化程度，PLN為胎盤之甲基化程度，且f為母體血漿中之部分胎兒DNA濃度。

在一個實施例中，胎盤之甲基化程度理論上可由以下等式推斷：

$$PLN = \frac{MP - BKG \times (1 - f)}{f} \quad (2)$$

當CN等於一、D等於PLN且BKG等於mbc時，等式(1)及(2)等效。在另一實施例中，部分胎兒DNA濃度可假定或設定為規定值，例如，作為所存在之最小f之假定值的一部分。

獲取母體血液之甲基化程度來代表母體血漿之背景甲基化。除母體血細胞中經高甲基化或低甲基化之基因座以外，吾等進一步研究藉由聚焦

於具有臨床相關性之所定義區域(例如人類基因組中之CpG島)的推斷方法。

常染色體及chrX上之總計27,458個CpG島(NCBI版本36/hg18)之平均甲基化密度來源於母體血漿及胎盤之定序資料。在所有分析樣品中，包括胎盤、母體血液及母體血漿，僅選擇覆蓋 $\geq 10$ 個CpG位點且每個所覆蓋CpG位點之平均讀取深度 $\geq 5$ 的CpG島。因此，26,698個CpG島(97.2%)仍有效，且使用血漿甲基化資料及部分胎兒DNA濃度根據以上等式推斷其甲基化程度。

應注意，在胎盤組織中，推斷PLN值之分佈低於CpG島之實際甲基化程度。因此，在一個實施例中，推斷PLN值或簡單推斷值(D)用作任意單位以評估胎盤中之CpG島之甲基化程度。在轉換後，推斷值呈線性且其分佈更類似於實際資料集。經轉換推斷值稱為甲基化預測值(MPV)且隨後用於預測胎盤中之基因座之甲基化程度。

在此實例中，將CpG島基於其在胎盤中之甲基化密度分為3類：低( $\leq 0.4$ )、中間( $>0.4$ - $<0.8$ )及高( $\geq 0.8$ )。使用推斷等式，吾等計算同一組CpG島之MPV且隨後使用該值以相同截止值將其分為3類。藉由比較實際與推斷資料集，吾等發現根據其MPV，75.1%之候選CpG島可與組織資料中之相同類別正確匹配。約22%之CpG島被指派至具有1級差異(高相對於中間，或中間相對於低)之群組，且小於3%將為完全錯分類(高相對於低)(圖12A)。亦測定總體分級效能：正確地將胎盤中甲基化密度 $\leq 0.4$ 、 $>0.4$ - $<0.8$ 及 $\geq 0.8$ 之CpG島中之86.1%、31.4%及68.8%推斷為「低」、「中間」及「高」(圖12B)。

圖11A及11B展示根據本發明實施例之使用母體血漿資料及部分胚胎

DNA濃度之預測演算法之效能的圖式。圖11A為展示使用MPV正確分類(推斷類別準確匹配實際資料集)、1級差異(推斷類別與實際資料集相差1級)及錯分類(推斷類別與實際資料集相對)之CpG島分類準確度的圖式1100。圖11B為展示分類於各推斷類別中之CpG島之比例的圖式1150。

鑒於母體背景甲基化在各別基因組區域中較低，循環中存在來源於高甲基化胎盤之DNA將使總體血漿甲基化程度增加一定程度，視部分胎兒DNA濃度而定。當所釋放之胎兒DNA完全甲基化時，可觀測到顯著變化。相反，當母體背景甲基化較高時，若釋放低甲基化胎兒DNA，則血漿甲基化程度之變化程度將變得更顯著。因此，當推斷已知在母體背景與胎盤之間互異之基因座的甲基化程度時，推斷流程可能更實際，尤其是對於胎盤中之高甲基化及低甲基化標記物。

圖12A為展示根據本發明實施例之用於甲基化預測之15個所選基因組基因座之詳情的表1200。為確定技術，吾等選擇先前已研究之15個差異性甲基化基因組基因座。推斷所選區域之甲基化程度且與先前研究之15個差異性甲基化之基因座相比較(RWK Chiu等人2007 *Am J Pathol*; 170: 941-950; S.S.C. Chim等人2008 *Clin Chem*; 54: 500-511; SSC Chim等人2005 *Proc Natl Acad Sci U S A*; 102: 14753-14758; DWY Tsui等人2010 *PLoS One*; 5: e15069)。

圖12B為展示15個所選基因組基因座之推斷類別及其在胎盤中之相應甲基化程度的圖式1250。推斷甲基化類別為：低， $\leq 0.4$ ；中間， $> 0.4 - < 0.8$ ；高， $\geq 0.8$ 。表1200及圖式1300顯示可正確推斷其在胎盤中之甲基化程度，但以下若干個除外：RASSF1A、CGI009、CGI137及VAPA。在該4個標記物中，僅CGI009與實際資料集展示顯著差異。其他僅稍微錯分

類。

在表1200中，「1」係指藉由以下等式計算之推斷值(D)：

$$D = \frac{MP - BKG \times (1 - f)}{f}$$

，其中f為部分胎兒DNA濃度。標記「2」係指甲

基化預測值(MPV)指代使用以下等式獲得之經線性轉換之推斷值： $MPV = D \times 1.6 + 0.25$ 。標記「3」係指推斷值之分類截止值：低， $\leq 0.4$ ；中(中等)， $> 0.4 - < 0.8$ ；高， $\geq 0.8$ 。標記「4」係指實際胎盤資料集之分類截止值：低， $\leq 0.4$ ；中(中等)， $> 0.4 - < 0.8$ ；高， $\geq 0.8$ 。標記「5」表示胎盤狀態係指胎盤之甲基化狀態相對於母體血細胞之甲基化狀態。

### C. 計算胎兒DNA之部分濃度

在一個實施例中，來自第一組織之胎兒DNA之百分比可使用男性胎兒之Y染色體。母體血漿樣品中之染色體Y序列之比例(chrY%)為來源於男性胎兒之染色體Y讀段及錯指派至染色體Y之母體(女性)讀段之數目的組合(RWK Chiu等人2011 BMJ; 342: c7401)。因此，樣品中之chrY%與部分胎兒DNA濃度(f)之間的關係可由以下等式獲得：

$$\text{chrY}\% = \text{chrY}_{\text{男性}}\% \times f + \text{chrY}_{\text{女性}}\% \times (1 - f)$$

其中chrY<sub>男性</sub>%係指含有100%男性DNA之血漿樣品中被指派為染色體Y之讀段的比例；且chrY<sub>女性</sub>%係指含有100%女性DNA之血漿樣品中被指派為染色體Y之讀段的比例。

可由來自懷有男性胎兒之女性之樣品中被指派至染色體Y且無錯配之讀段(例如，其中該等讀段來自經亞硫酸氫鹽轉化之樣品)確定chrY%。chrY<sub>男性</sub>%值可獲自對兩個成年男性血漿樣品之亞硫酸氫鹽定序。chrY<sub>女性</sub>%值可獲自對兩個未妊娠成年女性血漿樣品之亞硫酸氫鹽定序。

在其他實施例中，胎兒DNA百分比可由常染色體上之胎兒特異性等

位基因確定。作為另一實例，表觀遺傳標記物可用於測定胎兒DNA百分比。亦可使用測定胎兒DNA百分比之其他方式。

#### D. 使用甲基化確定複本數之方法

胎盤基因組之低甲基化程度大於母體基因組。如上文所論述，孕婦血漿之甲基化視母體血漿中來源於胎盤之胎兒DNA之部分濃度而定。因此，藉由分析染色體區域之甲基化密度，可偵測胎兒組織對母體血漿之貢獻差異。舉例而言，在懷有三體性胎兒(例如罹患第21對染色體三體症或第18對染色體三體症或第13對染色體三體症)之孕婦中，當與二體性染色體相比時，胎兒將由三體性染色體向母體血漿貢獻額外量之DNA。在此情形下，三體性染色體(或具有擴增之任何染色體區域)之血漿甲基化密度將低於二體性染色體。可藉由數學計算藉由考慮血漿樣品中之部分胎兒DNA濃度來預測差異程度。血漿樣品中之部分胎兒DNA濃度愈高，三體性與二體性染色體之間的甲基化密度差異將愈大。對於具有缺失之區域，甲基化密度將較高。

當女性胎兒將僅具有一個染色體X複本時，缺失之一個實例為特納氏症候群(Turner syndrome)。在此情形下，對於懷有罹患特納氏症候群之胎兒的孕婦，其血漿DNA中之染色體X之甲基化密度將高於懷有具有正常數目染色體X之女性胎兒之相同孕婦的情形。在此策略之一個實施例中，可首先分析母體血漿中存在或不存在染色體Y序列(例如，使用基於MPS或PCR之技術)。若存在染色體Y序列，則胎兒可分類為男性且將不需要進行以下分析。另一方面，若母體血漿中不存在染色體Y序列，則胎兒可分類為女性。在此情形下，可隨後分析母體血漿中之染色體X之甲基化密度。染色體X甲基化密度高於正常將指示胎兒具有高特納氏症候群風險。

此方法亦可應用於其他性別染色體非整倍性。舉例而言，對於受XYY影響之胎兒，母體血漿中Y染色體之甲基化密度將低於母體血漿中具有類似含量之胎兒DNA之正常XY胎兒。作為另一實例，對於罹患克萊恩費特氏症候群(Klinefelter syndrome)之胎兒(XXY)，染色體Y序列存在於母體血漿中，但母體血漿中之染色體X之甲基化密度將低於母體血漿中具有類似含量之胎兒DNA之正常XY胎兒。

根據先前論述，二體性染色體之血漿甲基化密度( $MP_{\text{非非整倍體}}$ )可計算為： $MP_{\text{非非整倍體}} = BKG \times (1 - f) + PLN \times f$ ，其中BKG為來源於血細胞及內臟之血漿中的背景DNA甲基化程度，PLN為胎盤之甲基化程度，且f為母體血漿中之部分胎兒DNA濃度。

三體性染色體之血漿甲基化密度( $MP_{\text{非整倍體}}$ )可計算為： $MP_{\text{非整倍體}} = BKG(1 - f) + PLNf \times 1.5$ ，其中1.5對應於複本數CN且再增加一個染色體為50%增加。三體性與二體性染色體之間的差異( $MP_{\text{差異}}$ )將為：

$$MP_{\text{差異}} = PIN \times f \times 0.5。$$

在一個實施例中，比較潛在非整倍體染色體(或染色體區域)與一或多個其他假定非非整倍體染色體之甲基化密度或基因組之總體甲基化密度可用於有效標準化血漿樣品中之胎兒DNA濃度。該比較可經由計算兩個區域之甲基化密度之間的參數(例如，涉及比率或差異)以獲得標準化甲基化密度。該比較可移除對所得甲基化程度之依賴性(例如，確定為來自兩種甲基化密度之參數)。

若潛在非整倍體染色體之甲基化密度未針對一或多個其他染色體或體現部分胎兒DNA濃度之其他參數進行標準化，則部分濃度將為影響血漿中之甲基化密度的主要因素。舉例而言，懷有第21對染色體三體症胎兒

且部分胎兒DNA濃度為10%之孕婦之染色體21之血漿甲基化密度將與懷有整倍體胎兒且部分胎兒DNA濃度為15%之孕婦相同，而標準化甲基化密度將顯示差異。

在另一實施例中，潛在非整倍體染色體之甲基化密度可針對部分胎兒DNA濃度進行標準化。舉例而言，可應用以下等式以標準化甲基化密度： $MP_{標準化} = MP_{未標準化} + (BKG - PLN) \times f$ ，其中 $MP_{標準化}$ 為經血漿中之部分胎兒DNA濃度標準化之甲基化密度， $MP_{未標準化}$ 為量測之甲基化密度， $BKG$ 為來自母體血細胞或組織之背景甲基化密度， $PLN$ 為胎盤組織中之甲基化密度，且 $f$ 為部分胎兒DNA濃度。 $BKG$ 及 $PLN$ 之甲基化密度可基於先前由獲自健康妊娠之母體血細胞及胎盤組織確定之參考值。不同的遺傳及表觀遺傳方法可用於測定血漿樣品中之部分胎兒DNA濃度，例如藉由使用大規模平行定序或PCR對未經亞硫酸氫鹽轉化之DNA量測來自染色體Y之序列讀段之百分比。

在一個實施例中，潛在非整倍體染色體之標準化甲基化密度可與由懷有整倍體胎兒之孕婦組成之參考組相比。可測定參考組之標準化甲基化密度之平均值及SD。隨後，測試病例之標準化甲基化密度可表示為z評分，該z評分指示得自參考組之平均值的SD數值： $z_{評分} = \frac{MP_{標準化} - 平均值}{SD}$ ，其中 $MP_{標準化}$ 為測試病例之標準化甲基化密度，平均值為參考病例之標準化甲基化密度之平均值，且SD為參考病例之標準化甲基化密度之標準差。若染色體顯著低甲基化，則截止值，例如z評分<-3，可用於分類，且因此，用於確定樣品之非整倍性狀態。

在另一實施例中， $MP_{差異}$ 可用作標準化甲基化密度。在該實施例中，可推斷 $PLN$ ，例如使用方法1000。在一些實施例中，參考甲基化密度(其



可使用f標準化)可由非非整倍體區域之甲基化程度確定。舉例而言，平均值可由相同樣品之一或多個染色體區域確定。截止值可藉由f定標，或僅設定為足夠程度，只要存在最小濃度即可。

因此，區域之甲基化程度與截止值之比較可用各種方式實現。該比較可涉及標準化(例如，如上所述)，該標準化可等效地對甲基化程度或截止值進行，視如何界定該等值而定。因此，區域之所測定甲基化程度是否在統計上不同於參考程度(由相同樣品或其他樣品測定)可用多種方式確定。

以上分析可應用於分析染色體區域，該等染色體區域可包括完整染色體或部分染色體，包括染色體之連續或不連貫亞區。在一個實施例中，潛在非整倍體染色體可分成許多區間。該等區間可具有相同或不同的大小。各區間之甲基化密度可針對樣品之部分濃度或一或多個假定非非整倍體染色體之甲基化密度或基因組之總體甲基化密度進行標準化。各區間之標準化甲基化密度可隨後可與參考組相比較以確定其是否顯著低甲基化。隨後可測定顯著低甲基化之區間之百分比。顯著低甲基化之區間之截止值，例如大於5%、10%、15%、20%或30%，可用於對病例之非整倍性狀態進行分類。

當測試擴增或缺失時，可比較甲基化密度與可能對所測試之特定區域具有特異性的參考甲基化密度。各區域可具有不同的參考甲基化密度，因為甲基化可能因區域而各異，特定言之視區域之大小而定(例如，區域愈小，將展示愈多變化)。

如上文所提及，一或多個各自懷有整倍體胎兒之孕婦可用於定義相關區域之甲基化密度之正常範圍或兩個染色體區域之間的甲基化密度差

異。亦可測定PLN之正常範圍(例如，藉由直接量測或藉由方法1000推斷)。在其他實施例中，可使用兩種甲基化密度之間的比率，例如，潛在非整倍體染色體及非非整倍體染色體之甲基化密度比率可用於該分析來替代其差異。此甲基化分析方法可與序列讀段計數方法(RWK Chiu等人2008 Proc Natl Acad Sci USA;105:20458-20463)及涉及血漿DNA之大小分析之方法(美國專利2011/0276277)組合以測定或確定非整倍性。與甲基化分析組合使用之序列讀段計數方法可使用隨機定序(RWK Chiu等人2008 Proc Natl Acad Sci USA;105:20458-20463；DW Bianchi DW等人2012 Obstet Gynecol 119:890-901)或靶向定序(AB Sparks等人2012 Am J Obstet Gynecol 206:319.e1-9；B Zimmermann等人2012 Prenat Diagn 32:1233-1241；GJ Liao等人2012 PLoS One; 7:e38154)來進行。

使用BKG可解決樣品之間的背景的變化。舉例而言，一個女性可能具有與另一女性不同的BKG甲基化程度，但在該等情形下，在樣品之間可使用BKG與PLN之間的差異。不同染色體區域之截止值可不同，例如，當基因組之一個區域之甲基化密度相對於基因組之另一區域有所不同時。

此方法可經通用化以偵測胎兒基因組中之任何染色體畸變，包括缺失及擴增。另外，此分析之解析度可調節至所要程度，例如基因組可分成10 Mb、5 Mb、2 Mb、1 Mb、500 kb、100 kb區間。因此，此技術亦可用於偵測亞染色體複製或亞染色體缺失。此技術將因此允許非侵入性獲得產前胎兒分子核型。當以此方式使用時，此技術可與基於分子計數之非侵入性產前測試方法(A Srinivasan等人2013 Am J Hum Genet;92:167-176；SCY Yu等人2013 PLoS One 8: e60968)組合使用。在其他實施例

中，區間之大小不需要一致。舉例而言，區間之大小可經調節以使得各區間含有一致數目之CpG二核苷酸。在此情況下，區間之實際大小將不同。

等式可重寫如下以應用於不同類型之染色體畸變：

$MP_{\text{差異}} = (BKG - PLN) \times f \times 0.5 \times CN$ 。此處CN表示受影響區域處複本數變化之數目。CN在獲得1個染色體複本時等於1，在獲得2個染色體複本時等於2，且在損失兩個同源染色體之一時(例如，偵測其中女性胎兒損失一個X染色體，從而產生XO核型的胎兒特納氏症候群時)等於-1。當區間之大小變化時，此等式不需要變化。然而，當使用較小區間大小時，敏感性及特異性可能降低，因為較小數目之CpG二核苷酸(或在胎兒DNA與母體DNA之間顯示差異性甲基化的其他核苷酸組合)將存在於較小區間中，從而在量測甲基化密度時增加隨機變化。在一個實施例中，所需要之讀段之數目可藉由分析甲基化密度之變化係數及所要敏感性程度來確定。

為證明此方法之可行性，吾等已分析來自9個孕婦之血漿樣品。在五個孕婦中，各自懷有整倍體胎兒，且其他四人各自懷有第21對染色體三體症(T21)胎兒。隨機選擇五個整倍體妊娠中之三個以形成參考組。使用此方法分析其餘兩個整倍體妊娠病例(Eu1及Eu2)及四個T21病例(T21-1、T21-2、T21-3及T21-4)以測試潛在T21狀態。使用Illumina HiSeq2000平台對血漿DNA進行亞硫酸氫鹽轉化及定序。在一個實施例中，計算個別染色體之甲基化密度。隨後確定染色體21與其他21對常染色體之平均值之間的甲基化密度差異以獲得標準化甲基化密度(表1)。參考組之平均值及SD用於計算六個測試病例之z評分。

	Eu1	Eu2	T21-1	T21-2	T21-3	T21-4
染色體21與其他常染色體之間的 $MP_{\text{差異}}$ 之z評分	-1.48	1.09	-4.46	-5.30	-8.06	-5.69

**表1**：使用<-3之z評分截止值將樣品分類為T21，對所有整倍體及T21病例

之分類均正確。

在另一實施例中，將基因組分成1 Mb區間且測定各1 Mb區間之甲基化密度。潛在非整倍體染色體上之所有區間之甲基化密度可利用位於假定非非整倍體染色體上之所有區間之中值甲基化密度標準化。在一個實施例中，對於各區間，可計算與非非整倍體區間之中值相距之甲基化密度差異。可使用參考組之平均值及SD值來計算此等值之z評分。可確定顯示低甲基化之區間之百分比(表2)且與截止百分比相比較。

	Eu1	Eu2	T21-1	T21-2	T21-3	T21-4
染色體21上具有MP <sub>差異</sub> <-3之z評分之區間的百分比	0%	0%	33.3%	58.3%	19.4%	52.8%

**表2**：針對染色體21上顯著更低甲基化之區間使用5%作為截止值，對於T21狀態，所有病例均正確分類。

此種用於偵測胎兒染色體或亞染色體畸變之基於DNA甲基化之方法可與基於分子計數之方法聯合使用，諸如藉由DNA分子之定序(RWK Chiu等人2008 Proc Natl Acad Sci USA; 105: 20458-20463)或數位PCR(YMD Lo等人2007 Proc Natl Acad Sci USA; 104: 13116-13121)或大小分析(美國專利公開案2011/0276277)。該組合(例如DNA甲基化加分子計數，或DNA甲基化加大小分析，或DNA甲基化加分子計數加大小分析)將具有協同效應，該協同效應在臨床處置中將為有利的，例如改良敏感性及/或特異性。舉例而言，可減少將需要例如藉由定序加以分析之DNA分子數目而不會不利地影響診斷準確性。此特徵將允許該等測試更經濟地進行。作為另一實例，對於指定數目之所分析DNA分子，組合方法將允許在較低部分胎兒DNA濃度下偵測胎兒染色體或亞染色體畸變。

圖13為偵測來自生物體之生物樣品之染色體異常的方法1300的流程

圖。生物樣品包括細胞游離DNA，該細胞游離DNA包含來源於第一組織及第二組織之細胞游離DNA之混合物。第一組織可來自胎兒或腫瘤且第二組織可來自妊娠女性或患者。

在方塊1310處，分析來自生物樣品之複數個DNA分子。分析DNA分子可包括測定DNA分子在生物體基因組中之位置及測定DNA分子是否在一或多個位點處經甲基化。該分析可藉由接收來自可獲知甲基化之定序的序列讀段來進行，且因此可僅對先前獲自DNA之資料進行分析。在其他實施例中，該分析可包括實際定序或獲得該資料之其他有效步驟。

位置之測定可包括將DNA分子定位(例如經由定序讀段)於人類基因組之各別部分，例如定位於特定區域。在一個實施例中，若讀段未定位於相關區域，則該讀段可忽略。

在方塊1320處，對於複數個位點中之每一者，測定該位點處經甲基化之DNA分子之各別數目。在一個實施例中，位點為CpG位點，且可僅為某些CpG位點，如使用本文所提及之一或多個準則所選擇。甲基化DNA之數目等效於在使用特定位點處所分析之DNA分子總數(例如，定序讀段之總數)進行標準化後測定未甲基化數目。

在方塊1330處，基於第一染色體區域內之位點處之經甲基化DNA分子之各別數目來計算第一染色體區域之第一甲基化程度。第一染色體區域可具有任何大小，例如，上文所提及之大小。甲基化程度可說明指派至第一染色體區域之DNA分子總數，例如作為標準化程序之一部分。

第一染色體區域可具有任何大小(例如，整個染色體)且可由不連貫亞區構成，亦即，亞區彼此分開。可測定各亞區之甲基化程度且組合為例如平均值或中值以測定第一染色體區域之甲基化程度。

在方塊1340處，將第一甲基化程度與截止值相比較。截止值可為參考甲基化程度或係關於參考甲基化程度(例如，與正常程度相距規定距離)。截止值可由懷有第一染色體區域中無染色體異常之胎兒的其他女性妊娠個體、無癌症個體之樣品、或已知與非整倍性無關之生物體基因座(亦即，二體性區域)測定。

在一個實施例中，截止值可定義為與參考甲基化程度相比具有差異  $(BKG - PLN) \times f \times 0.5 \times CN$ ，其中BKG為女性背景(或來自其他個體之平均值或中值)，f為來源於第一組織之細胞游離DNA之部分濃度，且CN為所測試之複本數。CN為對應於一種異常類型(缺失或複製)之標度因子之實例。CN 1之截止值可用於初步測試所有擴增，且隨後可使用其他截止值來測定擴增程度。截止值可基於來源於第一組織之細胞游離DNA之部分濃度以測定基因座之預期甲基化程度，例如在不存在複本數畸變的情況下。

在方塊1350處，第一染色體區域之異常之分類係基於比較而確定。統計上顯著之程度差異可指示胎兒具有染色體異常之風險增加。在各種實施例中，染色體異常可為第21對染色體三體症、第18對染色體三體症、第13對染色體三體症、特納氏症候群或克萊恩費特氏症候群。其他實例為亞染色體缺失、亞染色體複製或狄喬治氏症候群(DiGeorge syndrome)。

## V. 測定標記物

如上所述，胎兒基因組之某些部分以不同於母體基因組之方式經甲基化。此等差異在整個妊娠中可為常見的。不同甲基化之區域可用於鑑別來自胎兒之DNA片段。

### A. 用於測定來自胎盤組織及母體組織之DMR的方法

胎盤具有組織特異性甲基化標誌。已基於在胎盤組織與母體血細胞

之間差異性甲基化之基因座而開發出胎兒特異性DNA甲基化標記物用於母體血漿偵測及非侵入性產前診斷應用(SSC Chim等人2008 Clin Chem; 54: 500-511 ; EA Papageorgiou等人2009 Am J Pathol; 174: 1609-1618 ; 及T Chu等人2011 PLoS One; 6: e14723)。提供基於全基因組搜尋該等差異性甲基化區域(DMR)之實施例。

圖14為根據本發明實施例之用於藉由比較胎盤甲基化概況與母體甲基化概況(例如，由血細胞測定)來鑑別甲基化標記物之方法1400的流程圖。方法1400亦可用於藉由比較腫瘤甲基化概況與對應於健康組織之甲基化概況來測定腫瘤標記物。

在方塊1410處，獲得胎盤甲基化模式組及血液甲基化模式組。可由胎盤樣品，例如CVS或妊娠期胎盤測定胎盤甲基化模式組。甲基化模式組應理解為可包括僅部分基因組之甲基化密度。

在方塊1420處，鑑別包括規定量之位點(例如，5個CpG位點)且已獲得足夠數目之讀段的區域。在一個實施例中，鑑別始於各染色體之一端以定位含有至少五個合格CpG位點之前500 bp區域。若CpG位點被至少五個序列讀段覆蓋，則可認為該位點合格。

在方塊1430處，計算各位點之胎盤甲基化指數及血液甲基化指數。舉例而言，對於各500 bp區域內之所有合格CpG位點，個別地計算甲基化指數。

在方塊1440處，比較母體血細胞與胎盤樣品之間的甲基化指數以確定指數集彼此之間是否不同。舉例而言，使用例如曼-惠特尼檢驗(Mann-Whitney test)比較母體血細胞與CVS或妊娠期胎盤之間的甲基化指數。例如P值 $\leq 0.01$ 被視為統計上顯著不同，但可使用其他值，其中較低數字將減

少假陽性區域。

在一個實施例中，若合格CpG位點數目小於五或曼-惠特尼檢驗為非顯著，則500 bp區域向下游位移100 bp。該區域繼續向下游位移直至曼-惠特尼檢驗對500 bp區域變得顯著。隨後將考慮下一500 bp區域。若曼惠特尼檢驗發現下一區域展現統計顯著性，則將其加入當前區域，只要組合連續區域不大於1,000 bp即可。

在方塊1450處，可合併統計上顯著不同(例如，藉由曼-惠特尼檢驗)之相鄰區域。注意：該差異在兩個樣品之甲基化指數之間。在一個實施例中，若相鄰區域彼此在規定距離(例如1,000 bp)內且若其顯示類似甲基化概況，則將其合併。在一個實施例中，相鄰區域之間的甲基化概況之相似性可使用以下任一者來定義：(1)胎盤組織中展現與母體血細胞相同之趨勢，例如胎盤組織中兩個區域之甲基化程度大於血細胞中；(2)胎盤組織中之相鄰區域之甲基化密度差異小於10%；及(3)母體血細胞中相鄰區域之甲基化密度差異小於10%。

在方塊1460處，計算該等區域處來自母體血細胞DNA之血液甲基化模式組及胎盤樣品(例如CVS或足月胎盤組織)之甲基化密度。可如本文所述測定甲基化密度。

在方塊1470處，確定該區域中所有位點處之總胎盤甲基化密度及總血液甲基化密度在統計上顯著不同之假定DMR。在一個實施例中，對合併區域內之所有合格CpG位點進行 $\chi^2$ 檢驗。 $\chi^2$ 檢驗評定所合併區域內所有合格CpG位點中之甲基化胞嘧啶數目(呈甲基化及未甲基化胞嘧啶之比例形式)在母體血細胞與胎盤組織之間是否在統計上顯著不同。在一個實施例中，對於 $\chi^2$ 檢驗，P值 $\leq 0.01$ 可視為統計上顯著不同。藉由 $\chi^2$ 檢驗顯示顯



著性之合併區段被視為假定DMR。

在方塊1480處，鑑別母體血細胞DNA之甲基化密度高於高截止值或低於低截止值之基因座。在一個實施例中，鑑別母體血細胞DNA之甲基化密度 $\leq 20\%$ 或 $\geq 80\%$ 之基因座。在其他實施例中，可使用除母體血液以外的體液，包括(但不限於)唾液、來自女性生殖道之子宮或子宮頸灌洗液、淚液、汗液、唾液及尿液。

成功開發在母體血漿中具胎兒特異性之DNA甲基化標記物之關鍵可為母體血細胞之甲基化狀態為儘可能高度甲基化或儘可能未甲基化。由此可降低(例如減至最小)具有母體DNA分子之概率，該等母體DNA分子干擾對顯示相對甲基化概況之來源於胎盤之胎兒DNA分子進行分析。因此，在一個實施例中，藉由進一步過濾來選擇候選DMR。候選低甲基化基因座為顯示在母體血細胞中之甲基化密度 $\leq 20\%$ 且在胎盤組織中之甲基化密度高出至少20%的低甲基化基因座。候選高甲基化基因座為顯示在母體血細胞中之甲基化密度 $\geq 80\%$ 且在胎盤組織中之甲基化密度低至少20%的高甲基化基因座。可使用其他百分比。

在方塊1490處，隨後藉由比較與臨限值之差異來鑑別胎盤甲基化密度顯著不同於血液甲基化密度之基因座子集中的DMR。在一個實施例中，該臨限值為20%，故甲基化密度與母體血細胞之甲基化密度相差至少20%。因此，可計算所鑑別之各基因座處胎盤甲基化密度與血液甲基化密度之間的差異。該差異可為簡單減法。在其他實施例中，定標因數及其他函數可用於確定該差異(例如，該差異可為應用於簡單減法之函數之結果)。

在一個實施例中，使用此方法，自前三個月胎盤樣品中鑑別11,729

個高甲基化及239,747個低甲基化基因座。前100個高甲基化基因座列於附錄之表S2A中。前100個低甲基化基因座列於附錄之表S2B中。表S2A及S2B列出染色體、起始及結束位置、區域大小、母體血液中之甲基化密度、胎盤樣品中之甲基化密度、P值(均極小)及甲基化差異。該等位置對應於參考基因組 hg18，該基因組可見於 [hgdownload.soe.ucsc.edu/goldenPath/hg18/chromosomes](http://hgdownload.soe.ucsc.edu/goldenPath/hg18/chromosomes)。

自後三個月胎盤樣品中鑑別11,920個高甲基化及204,768個低甲基化基因座。後三個月之前100個高甲基化基因座列於表S2C中，且前100個低甲基化基因座列於表S2D中。據先前報導在母體血細胞與前三個月胎盤組織之間經差異性甲基化之三十三個基因座用於驗證吾等之前三個月候選者之清單。藉由使用吾等之演算法，33個基因座中有79%已鑑別為DMR。

圖15A為展示使用前三個月資料參考33個先前報導之前三個月標記物之DMR鑑別演算法之效能的表1500。在該表中，「a」指示基因座1至15先前描述於(RWK Chiu等人2007 *Am J Pathol*; 170:941-950；及SSC Chim等人2008 *Clin Chem*; 54:500-511)中；基因座16至23先前描述於(KC Yuen, 2007年論文, The Chinese University of Hong Kong, Hong Kong)中；且基因座24至33先前描述於(EA Papageorgiou等人2009 *Am J Pathol*; 174:1609-1618)中。「b」指示此等資料來源於以上公開文獻。

「c」指示可由本研究中產生之定序資料中但基於原始研究所提供之基因組座標觀測出母體血細胞及絨毛膜絨毛樣品之甲基化密度及其差異。

「d」指示使用方法1400之實施例在不參考以上所引用之Chiu等人(2007)、Chim等人(2008)、Yuen (2007)及Papageorgiou等人(2009)之公開文獻的情況下在亞硫酸氫鹽定序資料上鑑別關於基因座之資料。基因座

之跨度包括先前報導之基因組區域，但一般跨越更大區域。「e」指示基於在母體血細胞及絨毛膜絨毛樣品中之DMR之對應基因組座標之甲基化密度之間觀測到 $>0.20$ 差異的必要條件將候選DMR分類為真陽性(TP)或假陰性(FN)。

圖15B為展示使用後三個月資料且與分娩時所獲得之胎盤樣品相比較之DMR鑑別演算法之效能的表1550。「a」指示使用與圖17A中所述相同之33個基因座之清單。「b」指示由於33個基因座係先前自早期妊娠樣品中鑑別，故其可能不適用於後三個月資料。因此，查閱本研究中基於原始研究所提供之基因組座標產生之關於足月胎盤組織之亞硫酸氫鹽定序資料。母體血細胞與足月胎盤組織之間的甲基化密度差異 $>0.20$ 用於確定基因座在後三個月中確實為真DMR。「c」指示使用方法1400在不參考先前所引用之Chiu等人(2007)、Chim等人(2008)、Yuen(2007)及Papageorgiou等人(2009)之公開文獻的情況下在亞硫酸氫鹽定序資料上鑑別關於基因座的資料。基因座之跨度包括先前報導之基因組區域，但一般跨越更大區域。「d」指示基於在母體血細胞及足月胎盤組織中之DMR之對應基因組座標之甲基化密度之間觀測到 $>0.20$ 差異的必要條件將含有由於在後三個月中在差異性甲基化時合格之基因座的候選DMR分類為真陽性(TP)或假陰性(FN)。對於在後三個月中在差異性甲基化時未合格之基因座，其不存在於DMR清單中或存在含有該基因座但顯示甲基化差異 $<0.20$ 之DMR被視為真陰性(TN) DMR。

#### B. 來自母體血漿定序資料之DMR

應能夠由母體血漿DNA亞硫酸氫鹽定序資料直接鑑別胎盤組織DMR，其限制條件為樣品之部分胎兒DNA濃度亦為已知的。可能因為胎

盤為母體血漿中之胎兒DNA之主要來源(SSC Chim等人2005 Proc Natl Acad Sci USA 102, 14753-14758)，且吾等在此研究中展示母體血漿中之胎兒特異性DNA之甲基化狀態與胎盤甲基化模式組相關。

因此，可實施方法1400之態樣以使用血漿甲基化模式組替代使用胎盤樣品來測定推斷胎盤甲基化模式組。因此，可組合方法1000及方法1400以測定DMR。方法1000可用於測定胎盤甲基化概況之預測值且將其用於方法1400。對於此分析，實例亦聚焦於母體血細胞中 $\leq 20\%$ 或 $\geq 80\%$ 甲基化之基因座。

在一個實施例中，為推斷胎盤組織相對於母體血細胞中經高甲基化之基因座，吾等在血細胞甲基化密度與預測值之間相差至少50%的情況下淘選母體血細胞中根據預測值顯示 $\leq 20\%$ 甲基化及 $\geq 60\%$ 甲基化之基因座。為推斷胎盤組織相對於母體血細胞中經低甲基化之基因座，吾等在血細胞甲基化密度與預測值之間相差至少50%的情況下淘選母體血細胞中根據預測值顯示 $\geq 80\%$ 甲基化及 $\leq 40\%$ 甲基化之基因座。

圖16為展示基於直接分析母體血漿亞硫酸氫鹽定序資料而預測為高甲基化或低甲基化之位點之數目的表1600。「N/A」意謂不適用。「a」指示搜尋高甲基化基因座自顯示母體血細胞中之甲基化密度 $< 20\%$ 之基因座之清單起始。「b」指示搜尋低甲基化基因座自顯示母體血細胞中之甲基化密度 $> 80\%$ 之基因座之清單起始。「c」指示來自絨毛膜絨毛樣品之亞硫酸氫鹽定序資料用於驗證前三個月母體血漿資料，且足月胎盤組織用於驗證後三個月母體血漿資料。

如表1600中所示，大多數非侵入性推斷基因座在組織中顯示預期甲基化模式且與得自組織資料且存在於先前部分中之DMR重疊。附錄列出

自血漿中鑑別之DMR。表S3A列出由前三個月母體血漿亞硫酸氫鹽定序資料推斷為高甲基化之前100個基因座。表S3B列出由前三個月母體血漿亞硫酸氫鹽定序資料推斷為低甲基化之前100個基因座。表S3C列出由後三個月母體血漿亞硫酸氫鹽定序資料推斷為高甲基化之前100個基因座。表S3D列出由後三個月母體血漿亞硫酸氫鹽定序資料推斷為低甲基化之前100個基因座。

### C. 胎盤及胎兒甲基化模式組之妊娠期變化

CVS中之甲基化CpG之總體比例為55%，而在妊娠期胎盤中為59% (圖1之表100)。與妊娠期胎盤相比，可自CVS中鑑別出更多低甲基化DMR，而兩種組織之高甲基化DMR數目相似。因此，顯然CVS之低甲基化程度大於妊娠期胎盤。此妊娠期趨勢在母體血漿資料中亦顯而易見。在前三個月母體血漿中，甲基化CpG相對於胎兒特異性讀段之比例為47.0%，但在後三個月母體血漿中為53.3%。在前(1,457個基因座)及後三個月(1,279個基因座)母體血漿樣品中，驗證高甲基化基因座數目相似，但與後三個月樣品(12,677個基因座)相比，前三個月樣品(21,812個基因座)中存在實質上更多的低甲基化基因座(圖16之表1600)。

### D. 標記物之用途

差異性甲基化標記物或DMR適用於若干態樣。母體血漿中存在該等標記物指示並且證實存在胎兒或胎盤DNA。此證實可用作非侵入性產前測試之品質控制。DMR可充當母體血漿中之通用胎兒DNA標記物，且與依賴於母親與胎兒之間的基因型差異之標記物相比具有諸多優勢，諸如基於多態性之標記物或基於染色體Y之標記物。DMR為適用於所有妊娠之通用胎兒標記物。基於多態性之標記物僅適用於其中胎兒已繼承來自其父親

之標記物且母親之基因組中不具有此標記物的妊娠子集。另外，可藉由定量來源於該等DMR之DNA分子來量測母體血漿樣品中之胎兒DNA濃度。藉由獲知正常妊娠所預期之DMR概況，妊娠相關併發症，尤其涉及胎盤組織變化者，可藉由觀測母體血漿DMR概況或甲基化概況與正常妊娠所預期之母體血漿DMR概況或甲基化概況的偏差來偵測。涉及胎盤組織變化之妊娠相關併發症包括(但不限於)胎兒染色體非整倍性。實例包括第21對染色體三體症、子癩先兆、子宮內生長停滯及早產。

#### E. 使用標記物之套組

實施例可提供用於實踐本文所述之方法及其他可應用方法之組合物及套組。套組可用於進行分析以分析母體血漿中之胎兒DNA，例如，細胞游離胎兒DNA。在一個實施例中，套組可包括至少一種適用於與本文所鑑別之一或多個基因座進行特異性雜交之寡核苷酸。套組亦可包括至少一種適用於與一或多個參考基因座進行特異性雜交之寡核苷酸。在一個實施例中，量測胎盤高甲基化標記物。測試基因座可為母體血漿中之甲基化DNA且參考基因座可為母體血漿中之甲基化DNA。可構成類似套組以分析血漿中之腫瘤DNA。

在一些情況下，套組可包括至少兩種寡核苷酸引子，該等引子可用於擴增標靶基因座(例如，附錄中之基因座)及參考基因座之至少一個區段。替代引子或除引子之外，套組可包括用於偵測對應於標靶基因座及參考基因座之DNA片段的經標記探針。在各種實施例中，套組之一或多個寡核苷酸對應於附錄之表中的基因座。通常，套組亦提供使用手冊以指導使用者分析測試樣品且評定測試個體中之生理學或病理學狀態。

在各種實施例中，提供用於分析含有胎兒DNA與來自懷有胎兒之女

性個體之DNA之混合物的生物樣品中之胎兒DNA之套組。套組可包含一或多個與表S2A、S2B、S2C、S2D、S3A、S3B、S3C及S3D中所列出之基因組區域之至少一個區段特異性雜交之寡核苷酸。因此，可使用來自該等表之許多寡核苷酸僅來自一個表的情況。該等寡核苷酸可充當引子，且可組織為成對引子，其中一對對應於來自該等表之特定區域。

## VI. 大小與甲基化密度之關係

已知血漿DNA分子以短分子形式存在於循環中，其中大多數分子之長度為約160 bp (YMD Lo等人2010 *Sci Transl Med*; 2: 61ra91; YW Zheng等人2012 *Clin Chem*; 58: 549-558)。令人感興趣的是，吾等之資料顯示血漿DNA分子之甲基化狀態與大小之間的關係。因此，血漿DNA片段長度與DNA甲基化程度相關聯。血漿DNA分子之特徵大小概況表明大多數與可能來源於細胞凋亡期間之酶促降解的單核小體相關。

循環DNA本質上經片段化。詳言之，循環胎兒DNA比母體血漿樣品中來源於母體之DNA短(KCA Chan等人2004 *Clin Chem*; 50: 88-92)。由於配對端對準使得能夠對經亞硫酸氫鹽處理之DNA進行大小分析，故可直接評定血漿DNA分子之大小與其各別甲基化程度之間是否存在任何相關性。吾等在母體血漿以及未妊娠成年女性對照血漿樣品中研究此相關性。

對各DNA分子之兩端進行配對端定序(其包括定序整個分子)可用於分析此研究中之各樣品。藉由比對各DNA分子之末端序列對與參考人類基因組並且標註定序讀段最末端之基因組座標，可確定定序DNA分子之長度。將血漿DNA分子自然地片段化成小分子且通常在無任何片段化步驟的情況下製備血漿DNA定序文庫。因此，藉由定序推斷之長度代表原

始血漿DNA分子之大小。

在先前研究中，吾等測定母體血漿中之胎兒及母體DNA分子之大小概況(YMD Lo等人2010 Sci Transl Med; 2: 61ra91)。吾等證明血漿DNA分子之大小類似於單核小體且胎兒DNA分子比母體的短。在此研究中，吾等已確定血漿DNA分子之甲基化狀態與其大小之關係。

#### A. 結果

圖17A為展示母體血漿、未妊娠女性對照血漿、胎盤及周邊血液DNA之大小分佈的圖1700。對於母體樣品及未妊娠女性對照血漿，兩種經亞硫酸氫鹽處理之血漿樣品顯示與先前所報導相同之特徵大小分佈(YMD Lo等人2010 Sci Transl Med; 2: 61ra91)，其中長度為166-167 bp之總序列及比143 bp短之DNA分子之10 bp區間性最豐富。

圖17B為母體血漿、成年女性對照血漿、胎盤組織及成年女性對照血液之大小分佈及甲基化概況之圖1750。對於相同大小且含有至少一個CpG位點之DNA分子，計算其平均甲基化密度。吾等隨後將DNA分子之大小與其甲基化密度之間的關係繪圖。特定言之，對於覆蓋至少1個CpG位點之定序讀段，針對50 bp至多達180 bp範圍內之各片段長度測定平均甲基化密度。令人感興趣的是，甲基化密度隨血漿DNA大小而增加且在約166-167 bp處達到峰值。然而，在使用超音波處理器系統片段化之胎盤及對照血液DNA樣品中未觀測到此模式。

圖18展示血漿DNA分子之甲基化密度及大小的圖。圖18A為前三個月母體血漿之圖1800。圖18B為後三個月母體血漿之圖1850。關於覆蓋至少一個CpG位點之所有定序讀段的資料由藍色曲線1805表示。關於亦含有胎兒特異性SNP等位基因之讀段的資料由紅色曲線1810表示。關於亦含有



母體特異性SNP等位基因之讀段的資料由綠色曲線1815表示。

含有胎兒特異性SNP等位基因之讀段被視為來源於胎兒DNA分子。含有母體特異性SNP等位基因之讀段被視為來源於母體DNA分子。一般而言，具有高甲基化密度之DNA分子的大小較長。此趨勢存在於前三個月與後三個月兩者中之胎兒與母體DNA分子兩者中。如先前所報導，胎兒DNA分子之總體大小比母體的短。

圖19A展示成年未妊娠女性之定序讀段之甲基化密度及大小的圖1900。來自成年未妊娠女性之血漿DNA樣品亦顯示相同的DNA分子之大小與甲基化狀態之間的關係。另一方面，在MPS分析之前藉由超音波處理步驟將基因組DNA樣品片段化。如圖1900中所示，來自血細胞及胎盤組織樣品之資料未顯示相同趨勢。因為細胞之片段化為人工的，故將預期大小與密度無關。因為血漿中之自然片段化之DNA分子確實顯示對大小之依賴性，故可假定較低甲基化密度使得分子更可能斷裂成較小片段。

圖19B為展示母體血漿中之胎兒特異性及母體特異性DNA分子之大小分佈及甲基化概況的圖1950。胎兒特異性及母體特異性血漿DNA分子亦展現相同的片段大小與甲基化程度之間的相關性。來源於胎盤之DNA及母體循環細胞游離DNA之片段長度均隨甲基化程度而增加。此外，其甲基化狀態之分佈彼此不重疊，表明該現象不論循環DNA分子之來源之原始片段長度如何均存在。

## B. 方法

因此，大小分佈可用於評估血漿樣品之總甲基化百分比。隨後可在妊娠期間、在癌症監測期間或在藉由連續量測血漿DNA之大小分佈進行治療期間根據圖18A及18B中所示之關係追蹤此甲基化量測。甲基化量測

亦可用於探尋DNA自相關器官或組織中釋放之增加或減少。舉例而言，可特異性探尋對特定器官(例如肝臟)具特異性之DNA甲基化標誌且量測此等標誌在血漿中之濃度。由於DNA在細胞死亡時釋放至血漿中，故含量增加可意謂細胞死亡增加或該特定器官或組織損傷。特定器官之含量降低可意謂用於對抗該器官中之損傷或病理過程之治療在控制之下。

圖20為根據本發明實施例之用於評估生物體之生物樣品中之DNA甲基化程度之方法2000的流程圖。可估計基因組之特定區域或整個基因組之甲基化程度。若需要特定區域，則可使用僅來自該特定區域之DNA片段。

在方塊2010處，量測對應於各種大小之DNA片段之量。對於複數個大小之各大小，可量測來自生物樣品之對應於該大小之複數個DNA片段之量。舉例而言，可量測長度為140個鹼基之DNA片段之數目。該等量可另存為直方圖。在一個實施例中，量測來自生物樣品之複數個核酸中之每一者的大小，該量測可基於個體(例如，藉由對整個分子或僅分子末端進行單一分子定序)或基於群組(例如，經由電泳)來進行。大小可對應於範圍。因此，量可針對大小在特定範圍內之DNA片段。當進行配對端定序時，定位(對準)於特定區域之DNA片段(如藉由配對序列讀段所確定)可用於測定該區域之甲基化程度。

在方塊2020處，基於多種大小DNA片段之量計算第一參數之第一值。在一個態樣中，第一參數提供生物樣品中之DNA片段之大小概況之統計度量(例如直方圖)。該參數可稱為大小參數，因為其係由複數個DNA片段之大小確定。

第一參數可具有各種形式。一個參數為特定大小或大小範圍之DNA

片段相對於所有DNA片段或相對於另一大小或範圍之DNA片段之百分比。該參數為特定大小之DNA片段數目除以片段總數，該片段總數可獲自直方圖(提供特定大小之片段的絕對或相對計數的任何資料結構)。作為另一實例，參數可為特定大小或在特定範圍內之片段之數目除以另一大小或範圍之片段的數目。除法可充當標準化以說明對不同樣品分析不同數目之DNA片段。標準化可藉由對各樣品分析相同數目之DNA片段來實現，由此在除以所分析片段之總數時有效提供相同結果。參數及關於大小分析之其他實例可見於美國專利申請案13/789,553中，該申請案係出於所有目的而以引用的方式併入。

在方塊2030處，比較第一大小值與參考大小值。可根據參考樣品之DNA片段計算參考大小值。為測定參考大小值，可計算並且定量參考樣品之甲基化概況以及第一大小參數值。因此，當比較第一大小值與參考大小值時，可測定甲基化程度。

在方塊2040處，基於該比較評估甲基化程度。在一個實施例中，可確定第一參數之第一值是高於或是低於參考大小值，且藉此確定本發明樣品之甲基化程度是高於或是低於參考大小值之甲基化程度。在另一實施例中，藉由將第一值輸入校準函數中來實現該比較。校準函數可藉由鑑別對應於第一值之曲線上的點來有效比較第一值與校準值(參考大小值集)。隨後提供所評估之甲基化程度作為校準函數之輸出值。

因此，可校準大小參數與甲基化程度。舉例而言，可量測甲基化程度且其與該樣品之特定大小參數相關。隨後可將來自各種樣品之資料點擬合於校準函數。在一個實施例中，不同的校準函數可用於不同的DNA子集。因此，可基於關於特定DNA子集之甲基化與大小之間的關係的先前

知識進行某種形式的校準。舉例而言，對胎兒及母體DNA之校準可不同。

如上所示，當與母體血液相比較時，胎盤之低甲基化程度更大，且因此胎兒DNA由於甲基化程度較低而較小。因此，樣品之片段之平均大小(或其他統計值)可用於評估甲基化密度。由於片段大小可使用配對端定序而非潛在技術上更複雜之獲知甲基化之定序來量測，故此方法在臨床使用時將可能為成本有效的。此方法可用於監測與妊娠進展相關或與妊娠相關病症(諸如子癩先兆)、早產及胎兒病症(諸如由染色體或基因異常或子宮內生長停滯導致之病症)相關之甲基化變化。

在另一實施例中，此方法可用於偵測並監測癌症。舉例而言，在成功治療癌症之情況下，如使用此基於大小之方法所量測之血漿或另一體液中之甲基化概況將向未患癌症之健康個體之血漿或另一體液中之甲基化概況變化。反之，在癌症正在進展之情況下，則血漿或另一體液中之甲基化概況將偏離未患癌症之健康個體之血漿或另一體液中之甲基化概況。

總之，在血漿中，低甲基化分子比高甲基化分子短。在胎兒及母體DNA分子兩者中觀測到相同趨勢。因為已知DNA甲基化影響核小體包裝，因此吾等之資料表明低甲基化DNA分子或許密度較低地與組蛋白一起包裝且因此更容易酶促降解。另一方面，圖18A及18B中所呈現之資料亦顯示儘管胎兒DNA之低甲基化程度比母體讀段低得多，但胎兒與母體DNA之大小分佈並未彼此完全分開。在圖19B中，可見即使對於相同大小類別，胎兒及母體特異性讀段之甲基化程度亦彼此不同。此觀測結果表明，胎兒DNA之低甲基化狀態並非造成其相對於母體DNA相對較短的唯一因素。

## VII. 基因座之壓印狀態

可偵測母體血漿中與母親共有相同基因型但具有不同表觀遺傳標誌之來源於胎兒之DNA分子(LLM Poon等人2002 Clin Chem; 48: 35-41)。為顯示定序方法在收集母體血漿中來源於胎兒之DNA分子時具敏感性，吾等應用相同策略來偵測母體血漿樣品中之印記胎兒等位基因。鑑別兩個基因組印記區域：H19 (chr11:1,977,419-1,977,821，NCBI版本36/hg18)及MEST (chr7:129,917,976-129,920,347，NCBI版本36/hg18)。兩個區域均含有資訊性SNP以用於區分母體與胎兒序列。對於H19，即母體表現之基因，對於該區域中之SNP rs2071094 (chr11:1,977,740)，母親為同型組合(A/A)且胎兒為異型組合(A/C)。母體A等位基因中之一者完全甲基化且另一者未甲基化的。然而，在胎盤中，A等位基因未甲基化的，而自父體繼承之C等位基因完全甲基化。吾等在母體血漿中偵測兩個對應於來源於胎盤之印記父體等位基因之具有C基因型之甲基化讀段。

MEST，亦稱為PEG1，為父體表現之基因。對於印記基因座內之SNP rs2301335 (chr7:129,920,062)，母親與胎兒均為(A/G)。在母體血液中，G等位基因經甲基化，而A等位基因未甲基化。該甲基化模式在胎盤中逆轉，其中母體A等位基因經甲基化且父體G等位基因未甲基化。母體血漿中可偵測來源於父體之三個未甲基化G等位基因。相比之下，VAV1，即染色體19上之未印記基因座(chr19:6,723,621-6,724,121)在組織以及血漿DNA樣品中未顯示任何等位基因甲基化模式。

因此，甲基化狀態可用於測定何種DNA片段來自胎兒。舉例而言，當母親為GA異型組合時，僅偵測母體血漿中之A等位基因不能用作胎兒標記物。但若區分血漿中A分子之甲基化狀態，則甲基化A分子為胎兒特

異性，而未甲基化A分子為母體特異性，或反之亦然。

吾等接下來聚焦於胎盤組織中之據報導顯示基因組印記之基因座。基於Woodfine等人(2011 *Epigenetics Chromatin*; 4: 1)所報導之基因座清單，吾等進一步淘選印記控制區域內含有SNP的基因座。四個基因座滿足該準則且其為H19、KCNQ10T1、MEST及NESP。

關於母體血細胞樣品之針對H19及KCNQ10T1之讀段，母體讀段針對SNP為同型組合，且存在近似相等比例之甲基化及未甲基化讀段。CVS及足月胎盤組織樣品顯示胎兒就兩個基因座而言為異型組合且各等位基因排他性地經甲基化或未甲基化，亦即顯示單等位基因甲基化。在母體血漿樣品中，針對兩個基因座偵測自父體繼承之胎兒DNA分子。對於H19，自父體繼承之分子由含有胎兒特異性等位基因之定序讀段表示且經甲基化。對於KCNQ10T1，自父體繼承之分子由含有胎兒特異性等位基因之定序讀段表示且未甲基化。

另一方面，對於MEST及NESP，母親為異型組合。對於MEST，就SNP而言，母親及胎兒均為GA雜合子。然而，如自母體血細胞及胎盤組織之華生股資料顯而易見，與SNP相鄰之CpG之甲基化狀態在母親與胎兒中相對。在母親之DNA中，A等位基因未甲基化，但在胎兒之DNA中經甲基化。對於MEST，母體等位基因經甲基化。因此，可查明該胎兒已自其母親(在CVS中甲基化)繼承A等位基因且該母親已自其父親(母體血細胞中未甲基化)繼承A等位基因。令人感興趣的是，在母體血漿樣品中，可容易地區分所有四組分子，包括兩種母親等位基因中之每一者及兩種胎兒等位基因中之每一者。因此，藉由組合印記基因座處之基因型資訊與甲基化狀態，吾等可容易地區分自母體繼承之胎兒DNA分子與背景母體DNA分

子(LLM Poon等人2002 Clin Chem; 48: 35-41)。

此方法可用於偵測單親二體性。舉例而言，若已知此胎兒之父親就G等位基因而言為同型組合，則無法偵測母體血漿中之未甲基化G等位基因表示缺乏父體等位基因之貢獻。另外，在該情況下，當在此妊娠之血漿中偵測到甲基化G等位基因及甲基化A等位基因兩者時，將表明胎兒具有來自母親之單親異二體，亦即自母親繼承兩個不同的等位基因，而未自父親繼承等位基因。或者，若在母體血漿中偵測到甲基化A等位基因(自母親繼承之胎兒等位基因)及未甲基化A等位基因(自母方祖父繼承之母體等位基因)兩者而無未甲基化G等位基因(應由胎兒繼承之父體等位基因)，則將表明胎兒具有來自母親之單親同二體，亦即自母親繼承兩個一致等位基因，而未自父親繼承等位基因。

對於NESP，母親在SNP處為GA異型組合子，而胎兒就G等位基因而言為同型組合。父體等位基因就NESP而言經甲基化。在母體血漿樣品中，可容易地區分經甲基化之自父體繼承之胎兒G等位基因與未甲基化之背景母體G等位基因。

### VIII. 癌症/供體

一些實施例可用於使用循環血漿/血清DNA之甲基化分析來偵測、篩選、監測(例如用於復發、減輕或對治療之反應(例如存在或不存在))、分期、分類(例如用於輔助選擇最適當之治療模態)及預測癌症。

已知癌症DNA顯示異常DNA甲基化(JG Herman等人2003 N Engl J Med; 349: 2042-2054)。舉例而言，當與非癌細胞相比較時，基因(例如腫瘤抑制基因)之CpG島啟動子經高甲基化，而基因體中之CpG位點經低甲基化。鑒於癌細胞之甲基化概況可使用本文所述之方法由來源於腫瘤之

血漿DNA分子之甲基化概況來體現，吾等預期當與未患癌症之健康個體相比較時或當與癌症已治癒之個體相比較時，血漿中之總體甲基化概況在患有癌症之個體之間將有所不同。甲基化概況之差異類型可在於基因組甲基化密度及/或基因組區段之甲基化密度之定量差異方面。舉例而言，由於來自癌症組織之DNA之一般低甲基化性質(Gama-Sosa MA等人1983 *Nucleic Acids Res*; 11: 6883-6894)，將在癌症患者之血漿中觀測到血漿甲基化模式組或基因組區段之甲基化密度降低。

血漿甲基化模式組資料亦應體現甲基化概況之定性變化。舉例而言，當與來源於相同基因但在健康對照之樣品中之血漿DNA分子相比較時，來源於僅在癌細胞中高甲基化基因之血漿DNA分子將顯示在癌症患者之血漿中高甲基化。因為大部分癌症中發生異常甲基化，因此本文所述之方法可應用於偵測具有異常甲基化之所有惡性疾病形式，例如(但不限於)肺、乳房、結腸直腸、前列腺、鼻咽、胃、睪丸、皮膚、神經系統、骨、卵巢、肝臟、血液組織、胰臟、子宮、腎臟、膀胱、淋巴組織等中之惡性疾病。惡性疾病可具有多種組織學亞型，例如癌瘤、腺癌、肉瘤、纖維腺癌、神經內分泌亞型及未分化亞型等。

另一方面，吾等預期可區分來源於腫瘤之DNA分子與來源於背景非腫瘤之DNA分子，因為來源於腫瘤之DNA之總體較短大小概況著重於來源於具有腫瘤相關異常低甲基化之基因座之DNA分子，該腫瘤相關異常低甲基化對DNA分子之大小具有額外作用。此外，可使用與腫瘤DNA相關之多種特徵性特徵來區分來源於腫瘤之血漿DNA分子與來源於背景非腫瘤之血漿DNA分子，包括(但不限於)單一核苷酸變異體、複本數得失、易位、倒置、異常高甲基化或低甲基化及大小概況分析。由於所有此等變



化均可獨立地發生，故組合使用此等特徵可對血漿中之癌症DNA之敏感性及特異性偵測提供額外優勢。

#### A. 大小及癌症

血漿中來源於腫瘤之DNA分子之大小亦類似於單核小體單元之大小且比共存於癌症患者之血漿中的來源於背景非腫瘤之DNA分子短。大小參數已顯示與癌症有關，如美國專利申請案13/789,553中所述，該申請案係出於所有目的而以引用的方式併入。

因為血漿中來源於胎兒及來源於母體之DNA均顯示分子大小與甲基化狀態之間的關係，因此預期來源於腫瘤之DNA分子展現相同趨勢。舉例而言，在癌症患者之血漿中或在篩選癌症之個體中，低甲基化分子將比高甲基化分子短。

#### B. 癌症患者中之不同組織之甲基化密度

在此實例中，吾等分析肝細胞癌(HCC)患者之血漿及組織樣品。可在腫瘤手術切除之前及之後1週自HCC患者收集血液樣品。在血液樣品離心後收集血漿及白血球層。收集切除之腫瘤及相鄰非腫瘤肝組織。使用大規模平行定序在進行或不進行先前亞硫酸氫鹽處理的情況下分析自血漿及組織樣品提取之DNA樣品。亦分析來自未患癌症之四個健康個體之血漿DNA作為對照。DNA樣品之亞硫酸氫鹽處理將未甲基化胞嘧啶殘基轉化成尿嘧啶。在下游聚合酶鏈式反應及定序中，此等尿嘧啶殘基將充當胸苷。另一方面，亞硫酸氫鹽處理不會將甲基化胞嘧啶殘基轉化成尿嘧啶。在大規模平行定序後，藉由Methy-Pipe (P Jiang等人在IEEE生物資訊學與生物醫學國際研討會(IEEE International Conference on Bioinformatics and Biomedicine Workshops) (香港, 2010年12月18至21日)上提出之論文

Methy-Pipe: An integrated bioinformatics data analysis pipeline for whole genome methylome analysis)分析定序讀段以確定所有CG二核苷酸位置(亦即CpG位點)處之胞嘧啶殘基之甲基化狀態。

圖21A為展示HCC患者之手術前血漿及組織樣品之甲基化密度的表2100。相關區域(例如CpG位點、啟動子或重複序列區域等)之CpG甲基化密度係指顯示CpG甲基化之讀段相對於覆蓋基因組CpG二核苷酸之讀段總數的比例。白血球層及非腫瘤肝組織之甲基化密度相似。基於所有常染色體之資料，腫瘤組織之總體甲基化密度比白血球層及非腫瘤肝組織低25%。低甲基化在各個別染色體中一致。血漿之甲基化密度介於非惡性組織之值與癌症組織之值之間。此觀測結果與以下事實一致：癌症與非癌症組織均將對癌症患者之循環DNA有貢獻。已顯示造血系統為無活性惡性病狀之個體中之循環DNA之主要來源(YYN Lui等人2002 Clin Chem; 48: 421-7)。吾等因此亦分析獲自四個健康對照之血漿樣品。每個樣品之序列讀段之數目及所達到之定序深度示於圖21B之表2150中。

圖22為展示健康對照之血漿樣品中之常染色體甲基化密度(在71.2%至72.5%範圍內)的表220。此等資料顯示獲自不具有腫瘤DNA來源之個體之血漿樣品中之預期DNA甲基化程度。在癌症患者中，腫瘤組織亦將向循環中釋放DNA (KCA Chan等人2013 Clin Chem; 59: 211-224)；RJ Leary等人2012 Sci Transl Med; 4: 162ra154)。由於HCC腫瘤之低甲基化性質，當與健康對照之血漿含量相比較時，患者之手術前血漿中存在來源於腫瘤及非腫瘤之DNA兩者導致甲基化密度降低。實際上，手術前血漿樣品之甲基化密度介於腫瘤組織與健康對照之血漿之甲基化密度之間。原因係因為癌症患者之血漿DNA之甲基化程度將受腫瘤組織之異常甲基化

(在此情況下為低甲基化)之程度及循環中來源於腫瘤之DNA之部分濃度影響。腫瘤組織之甲基化密度較低及循環中來源於腫瘤之DNA之部分濃度較高將導致癌症患者中之血漿DNA之甲基化密度較低。據報導大部分腫瘤顯示整體低甲基化(JG Herman等人2003 N Engl J Med; 349: 2042-2054; MA Gama-Sosa等人1983 Nucleic Acids Res; 11: 6883-6894)。因此，HCC樣品中所見之當前觀測結果亦應適用於其他腫瘤類型。

在一個實施例中，當已知腫瘤組織之甲基化程度時，血漿DNA之甲基化密度可用於確定血漿/血清樣品中來源於腫瘤之DNA之部分濃度。若可獲得腫瘤樣品或可利用腫瘤之活組織檢查，則可獲得腫瘤組織之甲基化程度(例如甲基化密度)。在另一實施例中，關於腫瘤組織之甲基化程度的資訊可獲自對相似類型之腫瘤群組中之甲基化程度之調查，且使用本發明中所述之技術將此資訊(例如平均程度或中值程度)應用於欲分析之患者。腫瘤組織之甲基化程度可藉由分析患者之腫瘤組織來確定，或由分析具有相同或相似癌症類型之其他患者之腫瘤組織來推斷。腫瘤組織之甲基化可使用一定範圍之可獲知甲基化之平台來測定，包括(但不限於)大規模平行定序、單一分子定序、微陣列(例如寡核苷酸陣列)或質譜(諸如Epityper, Sequenom, Inc.分析)。在一些實施例中，可在該等分析之前進行對DNA分子之甲基化狀態敏感之程序，包括(但不限於)胞嘧啶免疫沈澱及可獲知甲基化之限制酶消化。當已知腫瘤之甲基化程度時，可在血漿甲基化模式組分析後計算癌症患者之血漿中之腫瘤DNA之部分濃度。

血漿甲基化程度P與部分腫瘤DNA濃度f及腫瘤組織甲基化程度TUM之間的關係可描述為： $P=BKG \times (1-f) + TUM \times f$ ，其中BKG為來源於血細胞及其他內臟之血漿中的背景DNA甲基化程度。舉例而言，據顯示，在獲

自此HCC患者之腫瘤活組織檢查組織中，所有常染色體之總體甲基化密度為42.9%，亦即此情況下之TUM值。來自四個健康對照之血漿樣品之平均甲基化密度為71.6%，亦即此情況下之BKG值。手術前血漿之血漿甲基化密度為59.7%。使用此等值，f估計為41.5%。

在另一實施例中，當已知血漿樣品中來源於腫瘤之DNA之部分濃度時，可基於血漿甲基化模式組資料非侵入性地評估腫瘤組織之甲基化程度。可藉由其他基因分析來測定血漿樣品中來源於腫瘤之DNA之部分濃度，例如先前所述之全基因組等位基因損失分析(GAAL)及單一核苷酸突變分析(美國專利申請案13/308,473；KCA Chan等人2013 Clin Chem; 59: 211-24)。該計算係基於與上述相同之關係，但在此實施例中，f值為已知的且TUM值為未知的。類似於對由母體血漿資料測定胎盤組織甲基化程度之情形所觀測之資料，可對整個基因組或基因組之部分進行推斷。

在另一實施例中，可使用甲基化密度之區間間變化或概況來區分患有癌症之個體與未患癌症之個體。可藉由將基因組分成特定大小之區間(例如1 Mb)來進一步增加甲基化分析之解析度。在該實施例中，對於所收集之樣品，例如白血球層、切除之HCC組織、與腫瘤相鄰之非腫瘤肝組織及腫瘤切除術前後所收集之血漿，計算各1 Mb區間之甲基化密度。在另一實施例中，區間大小不需要保持恆定。在一個實施例中，各區間內之CpG位點數目保持恆定，而區間自身之大小可變化。

圖23A及23B展示HCC患者之白血球層、腫瘤組織、非腫瘤肝組織、手術前血漿及手術後血漿之甲基化密度。圖23A為染色體1之結果的圖2300。圖23B為染色體2之結果的圖2350。

對於大部分1 Mb窗口，白血球層及與腫瘤相鄰之非腫瘤肝組織之甲

基化密度相似，而腫瘤組織之甲基化密度較低。手術前血漿之甲基化密度介於腫瘤與非惡性組織之甲基化密度之間。可使用手術前血漿之甲基化資料及部分腫瘤DNA濃度來推斷腫瘤組織中之所查詢基因組區域之甲基化密度。該方法與使用所有常染色體之甲基化密度值之上述方法相同。所述推斷腫瘤甲基化亦可使用血漿DNA之此更高解析度甲基化資料來進行。亦可使用其他區間大小，諸如300 kb、500 kb、2 Mb、3 Mb、5 Mb或大於5 Mb。在一個實施例中，區間大小不需要保持恆定。在一個實施例中，各區間內之CpG位點數目保持恆定，而區間自身之大小可變化。

### C. 比較癌症患者與健康個體之間的血漿甲基化密度

如2100中所示，手術前血漿DNA之甲基化密度低於癌症患者之非惡性組織之甲基化密度。此現象可能由存在來自腫瘤組織之低甲基化DNA造成。此較低血漿DNA甲基化密度可潛在地用作偵測及監測癌症之生物標記。對於癌症監測，若癌症正在進展，則血漿中來源於癌症之DNA之量將隨時間增加。在此實例中，血漿中來源於癌症之循環DNA之量增加將在全基因組層面上進一步降低血漿DNA甲基化密度。

反之，若癌症響應於治療，則血漿中來源於癌症之DNA之量將隨時間降低。在此實例中，血漿中來源於癌症之DNA之量降低將導致血漿DNA甲基化密度增加。舉例而言，若已用靶向療法(例如酪胺酸激酶抑制)治療具有表皮生長因子受體突變之肺癌患者，則血漿DNA甲基化密度增加將表示反應。隨後，抗酪胺酸激酶抑制之腫瘤純系之出現將與血漿DNA甲基化密度降低相關，其將指示復發。

可連續進行血漿甲基化密度量測，且可計算該等量測之變化率並且用於預測臨床進展或減輕或預後，或與臨床進展或減輕或預後相關。對於

在癌症組織中經高甲基化且在正常組織中經低甲基化之所選基因組基因座，例如許多腫瘤抑制基因之啟動子區域，癌症進展與有利治療反應之間的關係將與上述模式相對。

為證明此方法之可行性，吾等比較在手術移除腫瘤前後自癌症患者收集之血漿樣品與獲自四個健康對照個體之血漿DNA之DNA甲基化密度。

表2200展示癌症患者之手術前及手術後血漿樣品及四個健康對照個體之各常染色體之DNA甲基化密度及所有常染色體之組合值。對於所有染色體，手術前血漿DNA樣品之甲基化密度低於手術後樣品及來自四個健康個體之血漿樣品。手術前及手術後樣品之間的血漿DNA甲基化密度差異提供手術前血漿樣品中之較低甲基化密度係由於存在來自HCC腫瘤之DNA的支持性證據。

手術後血漿樣品中之DNA甲基化密度逆轉至類似於健康對照之血漿樣品之程度表明大部分來源於腫瘤之DNA已由於手術移除來源(亦即腫瘤)而消失。此等資料表明，如使用可獲自較大基因組區域(諸如所有常染色體或個別染色體)之資料確定之手術前血漿之甲基化密度與健康對照相比具有較低甲基化程度，從而允許鑑別(亦即診斷或篩選)測試病例患有癌症。

手術前血漿之資料亦顯示比手術後血漿低得多的甲基化程度，表明血漿甲基化程度亦可用於監測腫瘤負荷，因此用於預測及監測患者之癌症進展。可由健康對照或有癌症風險但目前未患癌症之人士之血漿測定參考值。有HCC風險之人士包括患有慢性B型肝炎或C型肝炎感染者、患有血色素沈著症者及患有肝硬化者。

基於參考值，超過(例如低於)所定義之截止值的血漿甲基化密度值可用於評定未妊娠人士之血漿是否具有或不具有腫瘤DNA。為偵測低甲基化循環腫瘤DNA之存在，截止值可定義為低於對照群體之值的5%或1%，或基於比對照之平均甲基化密度值低一定數目的標準差，例如2或3個標準差(SD)，或基於測定多個中值(MoM)。對於高甲基化腫瘤DNA，截止值可定義為高於對照群體之值的95%或99%，或基於比對照之平均甲基化密度值高出一定數目的標準差，例如2或3個標準差(SD)，或基於測定多個中值(MoM)。在一個實施例中，對照群體在年齡方面與測試個體相匹配。年齡匹配不需要精確，且可在多個年齡段(例如30至40歲，針對35歲之測試個體)中進行。

吾等接下來比較癌症患者與四個對照個體之血漿樣品之間的1 Mb區間之甲基化密度。出於說明目的，展示染色體1之結果。

圖24A為展示得自HCC患者之手術前血漿之甲基化密度的圖2400。圖24B為展示得自HCC患者之手術後血漿之甲基化密度的圖2450。藍色點表示對照個體之結果，紅色點表示HCC患者之血漿樣品之結果。

如圖24A中所示，對於大部分區間而言，來自HCC患者之手術前血漿之甲基化密度低於對照個體。對於其他染色體，觀測到類似模式。如圖24B中所示，對於大部分區間而言，來自HCC患者之手術後血漿之甲基化密度與對照個體類似。對於其他染色體，觀測到類似模式。

為評定測試個體是否患有癌症，將測試個體之結果與參考組之值相比較。在一個實施例中，參考組可包含許多健康個體。在另一實施例中，參考組可包含患有非惡性病狀(例如慢性B型肝炎感染或肝硬化)之個體。隨後可定量測試個體與參考組之間的甲基化密度差異。

在一個實施例中，參考範圍可來源於對照組之值。隨後測試個體之結果距參考組之上限或下限之偏差可用於確定個體是否具有腫瘤。此量將受血漿中來源於腫瘤之DNA之部分濃度及惡性與非惡性組織之間的甲基化程度差異影響。血漿中來源於腫瘤之DNA之部分濃度較高將導致測試血漿樣品與對照之間的甲基化密度差異較大。惡性及非惡性組織之甲基化程度之差異程度較大亦與測試血漿樣品與對照之間的較大甲基化密度差異相關。在另一實施例中，對於不同年齡範圍之測試個體，選擇不同的參考組。

在另一實施例中，針對各1 Mb區間計算四個對照個體之甲基化密度之平均值及SD。隨後對於對應區間，計算HCC患者之甲基化密度與對照個體之平均值之間的差異。在一個實施例中，此差異隨後除以對應區間之SD以確定z評分。換言之，z評分代表測試與對照血漿樣品之間的甲基化密度差異，其表示為相對於對照個體之平均值之SD之數目。區間之z評分 $>3$ 指示在該區間中HCC患者之血漿DNA之高甲基化程度比對照個體大3個SD，而區間中之z評分 $<-3$ 指示在該區間中HCC患者之血漿DNA之低甲基化程度比對照個體大3個SD。

圖25A及25B展示使用四個健康對照個體之血漿甲基化模式組資料作為參考針對染色體1對HCC患者之手術前(圖2500)及手術後(圖2550)血漿樣品之血漿DNA甲基化密度的z評分。各點表示一個1 Mb區間之結果。黑色點表示z評分在-3與3之間的區間。紅色點表示z評分 $<-3$ 的區間。

圖26A為展示手術前及手術後血漿之z評分之資料的表2600。在手術前血漿樣品中，染色體1上之大部分區間(80.9%)之z評分 $<-3$ ，表明HCC患者之手術前血漿DNA之低甲基化程度顯著大於對照個體。相反，在手



術後血漿樣品中紅色點之數目實質上降低(染色體1上之區間中之8.3%)，表明大部分腫瘤DNA已由於手術切除循環腫瘤DNA之來源而自循環中移除。

圖26B為展示使用四個健康對照個體作為參考針對由所有常染色體分析之1 Mb區間對HCC患者之手術前及手術後血漿樣品之血漿DNA甲基化密度之z評分的Circos圖2620。最外環顯示人類常染色體之表意文字。中間環顯示手術前血漿樣品之資料。最內環顯示手術後血漿樣品之資料。各點表示一個1 Mb區間之結果。黑色點表示z評分在-3與3之間的區間。紅色點表示z評分<-3的區間。綠色點表示z評分>3之區間。

圖26C為展示HCC患者之手術前及手術後血漿樣品兩者中之整個基因組之1 Mb區間之z評分之分佈的表2640。結果指示在整個基因組中，對於大多數區域(1 Mb區間中之85.2%)，HCC患者之手術前血漿DNA之低甲基化程度大於對照。相反，手術後血漿樣品中之大多數區域(1 Mb區間中之93.5%)顯示與對照相比無顯著高甲基化或低甲基化。此等資料指示此HCC之大部分腫瘤DNA (本質上主要經低甲基化)不再存在於手術後血漿樣品中。

在一個實施例中，z評分<-3之區間之數目、百分比或比例可用於指示是否存在癌症。舉例而言，如表2640中所示，在手術前血漿中，2734個所分析區間中之2330個(85.2%)顯示z評分<-3，而在手術後血漿中，2734個所分析區間中僅171個(6.3%)顯示z評分<-3。資料指示手術前血漿中之腫瘤DNA負載比手術後血漿中高得多。

可使用統計方法確定區間數目之截止值。舉例而言，基於正態分佈，將預期約0.15%區間之z評分<-3。因此，區間之截止數目可為所分析

之區間總數之0.15%。換言之，若來自未妊娠個體之血漿樣品顯示多於0.15%之區間之z評分 $<-3$ ，則血漿中存在低甲基化DNA來源，亦即癌症。舉例而言，吾等在此實例中已分析之2734個1 Mb區間中之0.15%為約4個區間。使用此值作為截止值，手術前及手術後血漿樣品均含有來源於腫瘤之低甲基化DNA，但該量在手術前血漿樣品中比手術後血漿樣品中高得多。對於四個健康對照個體，無一個區間顯示顯著高甲基化或低甲基化。可使用其他截止值(例如1.1%)，且其可視所使用之分析要求而變化。作為其他實例，截止百分比可基於統計分佈以及所要敏感性及可接受之特異性而變化。

在另一實施例中，截止數目可藉由接受者工作特徵(receiver operator characteristic, ROC)曲線分析藉由分析許多癌症患者及未患癌症個體來確定。為進一步驗證此方法之特異性，分析來自尋求非惡性病狀之醫學諮詢之患者的血漿樣品(C06)。1.1%區間之z評分 $<-3$ 。在一個實施例中，可使用不同的臨限值對不同程度之疾病狀態進行分類。可使用較低百分比臨限值來區分健康狀態與良性病狀，且較高百分比臨限值用於區分良性病狀與惡性疾病。

使用大規模平行定序之血漿低甲基化分析之診斷效能似乎優於使用基於聚合酶鏈式反應(PCR)之擴增特定類別重複序列元件所獲得的效能，例如長散佈核元件1 (LINE-1) (P Tangkijvanich 等人 2007 Clin Chim Acta; 379:127-133)。此觀測結果之一種可能解釋為儘管低甲基化在腫瘤基因組中非常普遍，但其在一個基因組區域與下一個之間具有一定程度的異質性。

實際上，吾等觀測到參考個體之平均血漿甲基化密度在整個基因組

上變化(圖56)。圖56中之各紅色點顯示32個健康個體中之一個1 Mb區間之平均甲基化密度。該圖展示整個基因組上所分析之所有1 Mb區間。各方框內之數字表示染色體數目。吾等觀測到平均甲基化密度因區間而各異。

基於PCR之簡單分析將不能夠在其診斷演算法中考慮到該區域-區域異質性。該異質性將拓寬健康個體中所觀測到之甲基化密度之範圍。被視為顯示低甲基化之樣品則需要甲基化密度之較大降低幅度。由此將導致測試敏感性降低。

相比之下，基於大規模平行定序之方法將基因組分成1 Mb區間(或其他大小之區間)且個別地量測該等區間之甲基化密度。此方法減小該等變化在不同基因組區域之基線甲基化密度中之影響，因為在測試樣品與對照之間比較各區域。實際上，在同一區間內，32個健康對照之個體間變化相對較小。在32個健康對照中，95%區間之變異係數(CV) $\leq$ 1.8%。又，為進一步增強對癌症相關低甲基化之偵測敏感性，該比較可在多個基因組區域之間進行。敏感性將藉由測試多個基因組區域來增強，因為此舉將在僅測試一個區域時在癌症樣品之特定區域碰巧不顯示低甲基化時避免生物學變化之效應。

比較對照與測試樣品之間的等效基因組區域之甲基化密度(例如獨立地測試各基因組區域，且隨後可能梳理該等結果)且對多個基因組區域進行此比較的方法在偵測與癌症相關之低甲基化時具有較高信雜比。以說明之方式展示此大規模平行定序方法。可測定多個基因組區域之甲基化密度且允許比較對照與測試樣品之間的對應區域之甲基化密度的其他方法學據預測將達成類似效應。舉例而言，可設計出可靶向來源於特定基因組區域

之血漿DNA分子以及測定該區域之甲基化程度的雜交探針或分子倒置探針以達成所要效應。

在另一實施例中，所有區間之z評分之總和可用於確定癌症是否存在或用於監測血漿DNA甲基化程度之連續變化。由於腫瘤DNA之總體低甲基化性質，故自患有癌症之個體收集之血漿中的z評分之總和將低於健康對照。HCC患者之手術前及手術後血漿樣品之z評分之總和分別為-49843.8及-3132.13。

在其他實施例中，其他方法可用於調查血漿DNA之甲基化程度。舉例而言，可使用質譜(ML Chen等人2013 Clin Chem; 59: 824-832)或大規模平行定序來測定甲基化胞嘧啶殘基相對於胞嘧啶殘基總含量之比例。然而，作為大部分胞嘧啶殘基不在CpG二核苷酸中之情形，當與CpG二核苷酸情形下所估計之甲基化程度相比較時，甲基化胞嘧啶相對於總胞嘧啶殘基之比例將相對較小。吾等測定獲自HCC患者之組織及血漿樣品以及四個獲自健康對照之血漿樣品的甲基化程度。在CpG情形下、任何胞嘧啶、在CHG及CHH情形下使用全基因組大規模平行定序資料量測甲基化程度。H係指腺嘌呤、胸嘧啶或胞嘧啶殘基。

圖26D為展示當使用CHH及CHG情形時與一些對照血漿樣品重疊之腫瘤組織及手術前血漿樣品之甲基化程度的表2660。當與白血球層、非腫瘤肝組織、手術後血漿樣品及健康對照血漿樣品相比較時，在CpG及未規定胞嘧啶兩者中，腫瘤組織及手術前血漿樣品之甲基化程度一貫較低。然而，基於甲基化CpG之資料，亦即甲基化密度，與基於甲基化胞嘧啶之資料相比顯示較寬動態範圍。

在其他實施例中，血漿DNA之甲基化狀態可藉由使用針對甲基化胞

嘧啶之抗體的方法，例如甲基化DNA免疫沈澱(MeDIP)來測定。然而，預期此等方法之精確度因為抗體結合之變異性而次於基於定序之方法。在另一實施例中，可測定血漿DNA中之5-羥甲基胞嘧啶之含量。就此而言，已發現5-羥甲基胞嘧啶之含量降低為某些癌症(例如黑色素瘤)之表觀遺傳特徵(CG Lian等人2012 Cell; 150: 1135-1146)。

除HCC以外，吾等亦研究此方法是否可應用於其他類型癌症。吾等分析來自以下患者之血漿樣品：2個肺腺癌患者(CL1及CL2)、2個鼻咽癌患者(NPC1及NPC2)、2個結腸直腸癌患者(CRC1及CRC2)、1個轉移性神經內分泌腫瘤患者(NE1)及1個轉移性平滑肌肉瘤患者(SMS1)。此等個體之血漿DNA經亞硫酸氫鹽轉化且使用Illumina HiSeq2000平台在一端針對50 bp進行定序。以上所提及之四個健康對照個體用作參考組以分析此8個患者。使用一端之50 bp序列讀段。將整個基因組分成1 Mb區間。使用來自參考組之資料計算各區間之甲基化密度之平均值及SD。隨後將8個癌症患者之結果表示為z評分，該等z評分表示距參考組之平均值之SD數目。正值指示測試病例之甲基化密度低於參考組之平均值，且負值指示測試病例之甲基化密度高於參考組之平均值。每個樣品之序列讀段之數目及所達到之定序深度示於圖27I之表2780中。

圖27A至27H展示根據本發明實施例之8個癌症患者之甲基化密度的Circos圖。各點表示1 Mb區間之結果。黑色點表示z評分在-3與3之間的區間。紅色點表示z評分<-3的區間。綠色點表示z評分>3之區間。兩條連續線之間的時間隔表示z評分差異為20。

在大部分類型癌症(包括肺癌、鼻咽癌、結腸直腸癌及轉移性神經內分泌腫瘤)患者之基因組中的多個區域中觀測到顯著低甲基化。令人感興

趣的是，除低甲基化以外，在患有轉移性平滑肌肉瘤之病例中之基因組中的多個區域中觀測到顯著高甲基化。平滑肌肉瘤之胚胎來源為中胚層，而其餘7個患者中之其他類型癌症之胚胎來源為外胚層。因此，有可能肉瘤之DNA甲基化模式可不同於癌瘤。

如由此情況可見，血漿DNA之甲基化模式亦可用於區分不同類型癌症，在此實例中為區分癌瘤及肉瘤。此等資料亦表明該方法可用於偵測與惡性疾病相關之異常高甲基化。對於所有此8種情況，僅利用血漿樣品且未分析腫瘤組織。此表明即使不存在腫瘤組織之先前甲基化概況或甲基化程度，亦可使用所述方法容易地偵測血漿中來源於腫瘤之DNA。

圖27J為表2790，該表為展示患有不同惡性疾病之患者血漿中之整個基因組之1 Mb區間之z評分之分佈的表。展示各病例下z評分<-3、-3至3及>3之區間的百分比。在所有病例下，超過5%區間之z評分<-3。因此，若吾等使用5%區間顯著低甲基化之截止值將樣品分類為對癌症呈陽性，則所有此等病例將分類為對癌症呈陽性。吾等之結果顯示低甲基化可能為不同類型癌症之一般現象，且血漿甲基化模式組分析將適用於偵測不同類型癌症。

#### D. 方法

圖28為根據本發明實施例分析生物體之生物樣品以確定癌症程度分類之方法2800的流程圖。生物樣品包括來源於正常細胞之DNA且可潛在包括來自與癌症相關之細胞之DNA。至少一些DNA在生物樣品中可為細胞游離DNA。

在方塊2810處，分析來自生物樣品之複數個DNA分子。分析DNA分子可包括測定DNA分子在生物體基因組中之位置及測定DNA分子是否在

一或多個位點處經甲基化。該分析可藉由接收來自可獲知甲基化之定序的序列讀段來進行，且因此可僅對先前獲自DNA之資料進行分析。在其他實施例中，該分析可包括實際定序或獲得該資料之其他有效步驟。

在方塊2820處，對於複數個位點中之每一者，測定該位點處經甲基化之DNA分子之各別數目。在一個實施例中，該等位點為CpG位點，且可僅為某些CpG位點，如使用本文所提及之一或多個準則所選擇。甲基化DNA分子之數目等效於在使用特定位點處所分析之DNA分子之總數，例如序列讀段之總數進行標準化後確定未甲基化之數目。舉例而言，區域之CpG甲基化密度增加等效於同一區域之未甲基化CpG之密度降低。

在方塊2830處，基於複數個位點處經甲基化之DNA分子之各別數目來計算第一甲基化程度。第一甲基化程度可對應於基於對應於該複數個位點之DNA分子數目而確定的甲基化密度。該等位點可對應於複數個基因座或僅一個基因座。

在方塊2840處，將第一甲基化程度與第一截止值相比較。該第一截止值可為參考甲基化程度或係關於參考甲基化程度(例如，與正常程度相距規定距離)。參考甲基化程度可由來自未患癌症之個體或已知與生物體之癌症無關聯之基因座或生物體之樣品來測定。該第一截止值可由該生物體之在測試該生物樣品前所獲得之先前生物樣品測定之參考甲基化程度確定。

在一個實施例中，該第一截止值為距由獲自健康生物體之生物樣品確定之參考甲基化程度之規定距離(例如，規定數目之標準差)。該比較可藉由確定第一甲基化程度與參考甲基化程度之間的差異，且隨後比較該差異與對應於該第一截止值之臨限值(例如，以確定甲基化程度是否在統計

上不同於參考甲基化程度)來進行。

在方塊2850處，基於該比較來確定癌症程度之分類。癌症程度之實例包括個體是否患有癌症或惡化前病狀，或發生癌症之可能性增加。在一個實施例中，該第一截止值可由來自個體之先前獲得之樣品確定(例如，可由該先前樣品確定參考甲基化程度)。

在一些實施例中，該第一甲基化程度可對應於甲基化程度超過臨限值之許多區域。舉例而言，可鑑別生物體之基因組之複數個區域。可使用本文所提及之例如某些長度或某些位點數之準則來鑑別該等區域。可在各區域內鑑別一或多個位點(例如CpG位點)。可計算各區域之區域甲基化程度。第一甲基化程度係針對第一區域。將各區域甲基化程度與各別區域截止值相比較，該各別區域截止值相對於多個區域可能相同或有所變化。第一區域之區域截止值為第一截止值。各別區域截止值可為距參考甲基化程度之規定量(例如0.5)，藉此僅計數與參考具有顯著差異之區域，該參考可由非癌症個體測定。

可確定區域甲基化程度超過各別區域截止值之區域之第一數目，且與臨限值相比較以確定分類。在一個實施例中，該臨限值為百分比。比較第一數目與臨限值可包括在與臨限值相比較之前將區域之第一數目除以區域之第二數目(例如，所有區域)，例如，作為標準化過程之一部分。

如上所述，生物樣品中之腫瘤DNA之部分濃度可用於計算第一截止值。該部分濃度可簡單評估為大於最小值，而部分濃度低於最小值之樣品可標記為例如不適於分析。該最小值可基於腫瘤之甲基化程度相對於參考甲基化程度之預期差異而確定。舉例而言，若差異為0.5 (例如，用作截止值)，則將需要某一腫瘤濃度足夠高以發現此差異。



來自方法1300之特異性技術可應用於方法2800。在方法1300中，可測定腫瘤之複本數變化(例如，其中可測試腫瘤之第一染色體區域相對於腫瘤之第二染色體區域具有複本數變化)。因此，方法1300可假定存在腫瘤。在方法2800中，可測試樣品究竟是否存在任何腫瘤之指示，而不考慮任何複本數特徵。該兩種方法之一些技術可相似。然而，與相對於癌症DNA與一些區域可能具有複本數變化之非癌症DNA之混合物之參考甲基化程度之差異相對，對於非癌症DNA，方法2800之截止值及甲基化參數(例如標準化甲基化程度)可偵測相對於參考甲基化程度之統計差異。因此，方法2800之參考值可由未患癌症之樣品測定，諸如未患癌症之生物體或同一患者之非癌症組織(例如，得到已知不具有癌症之樣品之前或同時獲取之血漿，該癌症可由細胞DNA確定)。

#### *E. 使用血漿DNA甲基化分析預測欲偵測之腫瘤DNA之最小部分濃度*

用於使用血漿DNA之甲基化程度量測該方法偵測癌症之敏感性的一種方式係關於來源於腫瘤之DNA之最小部分濃度，其要求當與對照比較時顯示血漿DNA甲基化程度之變化。測試敏感性亦視腫瘤組織與健康對照之基線血漿DNA甲基化程度或血細胞DNA之間的DNA甲基化之差異程度而定。血細胞為健康個體之血漿中之主要DNA來源。該差異愈大，愈可容易辨別癌症患者與非癌症個體，且將體現為血漿中來源於腫瘤之偵測下限較低且偵測癌症患者時之臨床敏感性較高。另外，健康個體或不同年齡之個體中之血漿DNA甲基化之變化(G Hannum等人2013 Mol Cell; 49: 359-367)亦將影響偵測與存在癌症相關之甲基化變化之敏感性。健康個體中之血漿DNA甲基化之變化較小將使得更容易偵測由存在少量來源於癌症之DNA所致之變化。

圖29A為展示參考個體中之甲基化密度分佈的圖2900，假定此分佈遵循正態分佈。此分析係基於各血漿樣品僅提供一個甲基化密度值，例如，所有常染色體或特定染色體之甲基化密度。其說明分析之特異性將如何受影響。在一個實施例中，使用比參考個體之平均DNA甲基化密度低3個SD之截止值來確定測試樣品之低甲基化程度是否顯著大於來自參考個體之樣品。當使用此截止值時，預期約0.15%非癌症個體將具有被分類為患有癌症之假陽性結果，從而獲得99.85%之特異性。

圖29B為展示參考個體及癌症患者中之甲基化密度分佈的圖2950。該截止值比參考個體之甲基化密度平均值低3個SD。若癌症患者之甲基化密度平均值比該截止值低2個SD（亦即，比參考個體之平均值低5個SD），則將預期97.5%癌症個體具有低於該截止值之甲基化密度。換言之，若針對各個體提供一個甲基化密度值，例如當分析整個基因組、所有常染色體或特定染色體之總甲基化密度時，預期敏感性將為97.5%。兩個群體之平均甲基化密度之間的差異受兩個因素影響，亦即癌症與非癌症組織之間的甲基化程度之差異程度及血漿樣品中來源於腫瘤之DNA之部分濃度。此兩個參數之值愈大，此兩個群體之甲基化密度值之差異愈大。另外，兩個群體之甲基化密度分佈之SD愈低，兩個群體之甲基化密度分佈之重疊愈少。

此處吾等使用假想實例來說明此概念。假定腫瘤組織之甲基化密度為約0.45且健康個體之血漿DNA之甲基化密度為約0.7。此等假定值類似於獲自吾等之HCC患者之值，其中對於來自健康對照之血漿樣品，常染色體之總體甲基化密度為42.9%，且常染色體之平均甲基化密度為71.6%。假定量測整個基因組之血漿DNA甲基化密度之CV為1%，該截止值將為

$0.7 \times (100\% - 3 \times 1\%) = 0.679$ 。為達成97.5%之敏感性，癌症患者血漿DNA之平均甲基化密度需為約 $0.679 - 0.7 \times (2 \times 1\%) = 0.665$ 。以f表示血漿樣品中來源於腫瘤之DNA之部分濃度。則f可計算為 $(0.7 - 0.665) / 0.05 = 0.7 - 0.665$ 。因此，f為約14%。根據此計算，據估計，若使用整個基因組之總甲基化密度作為診斷參數，則為了達成97.5%之診斷敏感性，血漿中可偵測到之最小部分濃度為14%。

接下來吾等對獲自HCC患者之資料進行此分析。關於此說明，基於由所有常染色體評估之值，僅對各樣品進行一次甲基化密度量測。在獲自健康個體之血漿樣品中，平均甲基化密度為71.6%。此四個樣品之甲基化密度之SD為0.631%。因此，血漿甲基化密度之截止值將需要為 $71.6\% - 3 \times 0.631\% = 69.7\%$ 以達到z評分 $< -3$ 且特異性為99.85%。為達成97.5%之敏感性，癌症患者之平均血漿甲基化密度將需要比截止值低2個SD，亦即68.4%。因為腫瘤組織之甲基化密度為42.9%且使用下式： $P = BKG \times (1 - f) + TUM \times f$ ，故f將需要為至少11.1%。

在另一實施例中，可獨立地分析不同基因組區域之甲基化密度，例如，如圖25A或26B中所示。換言之，多次量測各樣品之甲基化程度。如下文所示，可在低得多的部分腫瘤DNA濃度下在血漿中偵測到顯著低甲基化，且因此血漿DNA甲基化分析用於癌症偵測之診斷效能將得到增強。可對相對於參考群體顯示顯著甲基化密度偏差之基因組區域之數目進行計數。隨後可將基因組區域之數目與截止值相比較以確定所調查之基因組區域群體(例如，整個基因組之1 Mb區間)中是否存在血漿DNA之總體顯著低甲基化。可藉由分析未患癌症之參考個體群組確定截止值，或以數學方式，例如根據正態分佈函數獲得截止值。

圖30為展示健康個體及癌症患者之血漿DNA之甲基化密度分佈的圖3000。比較各1 Mb區間之甲基化密度與參考組之相應值。確定顯示顯著低甲基化(比參考組之平均值低3個SD)之區間之百分比。10%顯著低甲基化之截止值用於確定血漿樣品中是否存在來源於腫瘤之DNA。根據測試之所要敏感性及特異性，亦可使用其他截止值，諸如5%、15%、20%、25%、30%、35%、40%、45%、50%、60%、70%、80%或90%。

舉例而言，為將樣品分類為含有來源於腫瘤之DNA，吾等可使用10% 1 Mb區間顯示顯著低甲基化(z評分<-3)作為截止值。若超過10%區間之低甲基化程度顯著大於參考組，則將該樣品分類為對癌症測試呈陽性。對於各1 Mb區間，比參考組之平均甲基化密度低3個SD之截止值用於將樣品定義為低甲基化程度顯著更大。對於各1 Mb區間，若癌症患者之平均血漿DNA甲基化密度比參考個體之平均血漿DNA甲基化密度低1.72個SD，則癌症患者之任何特定區間之甲基化密度值將低於截止值(亦即，z評分<-3)且獲得陽性結果的概率為10%。隨後，若吾等觀察整個基因組之所有1 Mb區間，則將預期約10%區間顯示具有顯著較低甲基化密度之陽性結果(亦即z評分<-3)。假定健康個體之血漿DNA之總體甲基化密度為約0.7且量測各1 Mb區間之血漿DNA甲基化密度之變異係數(CV)為1%，則癌症患者之血漿DNA之平均甲基化密度將需要為 $0.7 \times (100\% - 1.72 \times 1\%) = 0.68796$ 。以f為血漿中來源於腫瘤之DNA之部分濃度，以便達成此平均血漿DNA甲基化密度。假定腫瘤組織之甲基化密度為0.45，則可使用以下等式計算f：

$$(\bar{M}_{P_{\text{參考}}} - M_{\text{腫瘤}}) \times f = \bar{M}_{P_{\text{參考}}} - \bar{M}_{P_{\text{癌症}}},$$

其中 $\bar{M}_{P_{\text{參考}}}$ 表示參考個體中之血漿DNA之平均甲基化密度； $M_{\text{腫瘤}}$ 表示

癌症患者中之腫瘤組織之甲基化密度；且  $\bar{M}_{p_{癌症}}$  表示癌症患者中之血漿DNA之平均甲基化密度。

使用此等式， $(0.7-0.45) \times f = 0.7 - 0.68796$ 。因此，最小部分濃度可使用此方法偵測，將推斷為4.8%。藉由降低低甲基化程度顯著較大之區間之截止百分比，例如自10%至5%，可進一步增強敏感性。

如以上實例中所示，此方法之敏感性係由癌症與非癌症組織(例如血細胞)之間的甲基化程度差異之程度決定。在一個實施例中，僅選擇顯示非癌症個體之血漿DNA與腫瘤組織之間的甲基化密度差異較大的染色體區域。在一個實施例中，僅選擇甲基化密度差異 $>0.5$ 之區域。在其他實施例中，0.4、0.6、0.7、0.8或0.9之差異可用於選擇適合區域。在另一實施例中，基因組區域之物理大小不固定。作為替代，例如基於固定讀取深度或固定CpG位點數目來定義基因組區域。對於各樣品，評定多個此等基因組區域之甲基化程度。

圖31為展示健康個體之血漿DNA平均值與HCC患者之腫瘤組織之間的甲基化密度差異之分佈的圖3100。正值表示甲基化密度高於健康個體之血漿DNA，且負值表示甲基化密度高於腫瘤組織。

在一個實施例中，可選擇癌症與非癌症組織之間的甲基化密度差異最大，例如差異 $>0.5$ 的區間，而不考慮此等區間之腫瘤是低甲基化或是高甲基化。鑒於血漿中來源於腫瘤之DNA之部分濃度相同，因為癌症與非癌症個體之間的血漿DNA甲基化程度分佈之間的差異較大，故可藉由聚焦於此等區間而降低血漿中來源於腫瘤之DNA之部分濃度的偵測極限。舉例而言，若僅使用差異 $>0.5$ 之區間且採用10%區間之低甲基化程度顯著較大之截止值用於確定測試個體是否患有癌症，則所偵測之來源於腫瘤之

DNA 之 最 小 部 分 濃 度 (f) 可 使 用 以 下 等 式 計 算：  
 $(\bar{M}_{P_{\text{參考}}} - M_{\text{腫瘤}}) \times f = \bar{M}_{P_{\text{參考}}} - \bar{M}_{P_{\text{癌症}}}$ ，其中  $\bar{M}_{P_{\text{參考}}}$  表示參考個體之血漿DNA之平均甲基化密度； $M_{\text{腫瘤}}$  表示癌症患者之腫瘤組織之甲基化密度；且  $\bar{M}_{P_{\text{癌症}}}$  表示癌症患者之血漿DNA之平均甲基化密度。

儘管參考個體之血漿與腫瘤組織之間的甲基化密度差異為至少0.5。隨後吾等得到 $0.5 \times f = 0.7 - 0.68796$ 及 $f = 2.4\%$ 。因此，藉由聚焦於癌症與非癌症組織之間具有較高甲基化密度差異之區間，部分來源於腫瘤之DNA之下限可自4.8%降至2.4%。關於何種區間將在癌症與非癌症組織(例如血細胞)之間顯示較大程度之甲基化差異的資訊可由獲自其他個體之同一器官或同一組織學類型之腫瘤組織測定。

在另一實施例中，參數可來源於所有區間之血漿DNA甲基化密度且考慮癌症與非癌症組織之間的甲基化密度差異。具有較大差異之區間可給予較大權重。在一個實施例中，若特定區間用於計算最終參數，則各區間之癌症與非癌症組織之間的甲基化密度差異可直接用作權重。

在另一實施例中，不同類型之癌症可在腫瘤組織中具有不同的甲基化模式。癌症特異性權重概況可來源於特定類型癌症之甲基化程度。

在另一實施例中，可在患有及未患癌症之個體中測定甲基化密度之區間間關係。在圖8中，吾等觀測到在較少數目之區間中，腫瘤組織之甲基化程度大於參考個體之血漿DNA。因此，可選擇具有最大差異極值，例如差異 $>0.5$ 及差異 $<0$ 之區間。此等區間之甲基化密度之比率隨後可用於指示測試個體是否患有癌症。在其他實施例中，不同區間之甲基化密度之差及商可用作指示區間間關係之參數。

吾等進一步評定如藉由獲自HCC患者之資料所說明，使用多個基因

組區域之甲基化密度來偵測或評定腫瘤之方法的偵測敏感性。首先，吾等混合來自手術前血漿之讀段與獲自健康對照之血漿樣品之讀段以模擬含有在20%至0.5%範圍內之部分腫瘤DNA濃度之血漿樣品。吾等隨後用等效於z評分<-3之甲基化密度對1 Mb區間(自整個基因組中之2,734個區間中)之百分比進行評分。當血漿中之部分腫瘤DNA濃度為20%時，80.0%區間顯示顯著低甲基化。關於血漿中部分腫瘤DNA濃度10%、5%、2%、1%及0.5%之相應資料分別為67.6%、49.7%、18.9%、3.8%及0.77%區間顯示低甲基化。因為對照樣品中顯示z評分<-3之區間數目之理論極限為0.15%，故吾等之資料顯示更多區間(0.77%)超過理論截止極限，即使當腫瘤部分濃度僅為0.5%時。

圖32A為展示當血漿樣品含有5%或2%腫瘤DNA時降低定序深度之效應的表3200。當平均定序深度僅為單倍體基因組之0.022倍時，仍可偵測到較高比例之區間(>0.15%)顯示顯著低甲基化。

圖32B為展示四個健康對照個體之血漿、HCC患者之白血球層、正常肝組織、腫瘤組織、手術前血漿及手術後血漿樣品中之重複序列元件及非重複序列區域之甲基化密度的圖式3250。可觀測到重複序列元件之甲基化程度(甲基化密度)大於(高於)癌症及非癌症組織中之非重複序列區域。然而，當與腫瘤組織相比較時，重複序列元件與非重複序列區域之間的甲基化差異大於非癌症組織及健康個體之血漿DNA。

因此，癌症患者之血漿DNA在重複序列元件之甲基化密度方面與非重複序列區域相比具有較大降低。對於重複序列元件及非重複序列區域，四個健康對照之平均值與HCC患者之間的血漿DNA甲基化密度差異分別為0.163及0.088。關於手術前及手術後血漿樣品之資料亦顯示甲基化密度

變化之動態範圍在重複序列中大於非重複序列區域。在一個實施例中，重複序列元件之血漿DNA甲基化密度可用於確定患者是否受癌症影響或用於監測疾病進展。

如上文所論述，參考個體之血漿中之甲基化密度變化亦將影響區分癌症患者與非癌症個體之準確性。甲基化密度分佈愈緊密(亦即，標準差愈小)，將能愈準確地區分癌症與非癌症個體。在另一實施例中，在參考組中，1 Mb區間之甲基化密度之變異係數(CV)可用作選擇具有較低血漿DNA甲基化密度變化性之區間的準則。舉例而言，僅選擇CV<1%之區間。其他值，例如0.5%、0.75%、1.25%及1.5%亦可用作選擇具有較低甲基化密度變化性之區間的準則。在另一實施例中，選擇準則可包括區間之CV及癌症與非癌症組織之間的甲基化密度差異兩者。

當已知腫瘤組織之甲基化密度時，甲基化密度亦可用於評估血漿樣品中來源於腫瘤之DNA之部分濃度。此資訊可藉由分析患者之腫瘤或調查許多患有同一類型癌症之患者之腫瘤而獲得。如上文所論述，血漿甲基化密度(P)可使用以下等式表示： $P = BKG \times (1 - f) + TUM \times f$ ，其中BKG為血細胞及其他器官之背景甲基化密度，TUM為腫瘤組織之甲基化密度，且f為血漿樣品中來源於腫瘤之DNA之部分濃度。此等式可重寫為：

$$f = \frac{BKG - P}{BKG - TUM}。$$

BKG之值可藉由在不存在癌症之時間點分析患者之血漿樣品或調查未患癌症之個體之參考組來測定。因此，在量測血漿甲基化密度之後，可確定f。

#### F. 與其他方法組合

本文所述之甲基化分析方法可與基於血漿中來源於腫瘤之DNA之基



因變化的其他方法組合使用。該等方法之實例包括分析血漿中之癌症相關染色體畸變(KCA Chan等人2013 Clin Chem; 59:211-224 ; RJ Leary等人2012 Sci Transl Med; 4:162ra154)及癌症相關單一核苷酸變化(KCA Chan等人2013 Clin Chem; 59:211-224)。與該等基因方法相比，甲基化分析方法存在優勢。

如圖21A中所示，腫瘤DNA之低甲基化為涉及跨越幾乎整個基因組分佈之區域的整體現象。因此，來自所有染色體區域之DNA片段關於來源於腫瘤之低甲基化DNA對患者血漿/血清DNA之潛在貢獻將具資訊性。相比之下，染色體畸變(染色體區域之擴增或缺失)僅存在於一些染色體區域中，且腫瘤組織中來自無染色體畸變之區域的DNA片段在分析中將不具資訊性(KCA Chan等人2013 Clin Chem; 59: 211-224)。同樣，在各癌症基因組中僅觀測到數千單一核苷酸變化(KCA Chan等人2013 Clin Chem; 59: 211-224)。不與此等單一核苷酸變化重疊之DNA片段在確定血漿中是否存在來源於腫瘤之DNA時將不具資訊性。因此，與用於偵測循環中之癌症相關變化之該等基因方法相比，此甲基化分析方法可能更具成本有效性。

在一個實施例中，可藉由增濃最具資訊性之區域(例如癌症與非癌症組織之間的差異性甲基化差異最高的區域)之DNA片段來進一步增強血漿DNA甲基化分析之成本有效性。增濃此等區域之方法的實例包括使用雜交探針(例如 Nimblegen SeqCap系統及 Agilent SureSelect 標靶增濃系統)、PCR擴增及固相雜交。

#### G. 組織特異性分析/供體

來源於腫瘤之細胞侵入並轉移至相鄰或遠端器官。經侵入組織或轉

移性病灶由於細胞死亡而向血漿中貢獻DNA。藉由分析癌症患者之血漿中之DNA甲基化概況及偵測組織特異性甲基化標誌之存在，可偵測涉及疾病過程之組織的類型。此方法提供涉及癌症過程之組織之非侵入性解剖掃描以輔助鑑別作為原發性及轉移性位點之所涉及器官。監測血漿中之所涉及器官之甲基化標誌的相對濃度亦將允許評定該等器官之腫瘤負荷且確定該器官中之癌症過程是劣化或是改良或已治癒。舉例而言，若在肝臟中基因X經特異性甲基化。則預期癌症(例如結腸直腸癌)對肝臟之轉移性涉及將增加血漿中來自基因X之甲基化序列之濃度。亦將存在與基因X具有類似甲基化特徵之另一序列或序列群組。隨後可組合來自該等序列之結果。類似考慮適用於其他組織，例如腦、骨、肺及腎臟等。

另一方面，已知來自不同器官之DNA展現組織特異性甲基化標誌(BW Futscher等人2002 Nat Genet; 31:175-179；SSC Chim等人2008 Clin Chem; 54: 500-511)。因此，血漿中之甲基化概況可用於闡明來自各種器官之組織對血漿之貢獻。闡明該貢獻可用於評定器官損傷，因為據信當細胞死亡時釋放血漿DNA。舉例而言，諸如肝炎(例如由病毒、自體免疫過程等所致)或由藥物所致之肝毒性(例如藥物過度劑量(諸如由撲熱息痛(paracetamol)引起)或毒素(諸如由酒精引起))之肝臟病理學與肝細胞損傷相關，且將預期與血漿中來源於肝臟之DNA之含量增加相關。舉例而言，若在肝臟中基因X經特異性甲基化。則將預期肝臟病理學增加血漿中來自基因X之甲基化序列之濃度。反之，若在肝臟中基因Y經特異性低甲基化。則將預期肝臟病理學降低血漿中來自基因Y之甲基化序列之濃度。在其他實施例中，基因X或Y可經可能不為基因且在體內不同組織中展現差異性甲基化之任何基因組序列置換。

本文所述之技術亦可應用於評定器官移植接受者之血漿中來源於供體之DNA (YMD Lo等人1998 Lancet; 351:1329-1330)。供體與接受者之間的多態性差異已用於區分血漿中來源於供體之DNA與來源於接受者之DNA (YW Zheng等人2012 Clin Chem; 58: 549-558)。吾等提議移植器官之組織特異性甲基化標誌亦可用作偵測接收者血漿中之供體DNA的方法。

藉由監測供體DNA之濃度，可非侵入性地評定移植器官之狀態。舉例而言，當與當患者處於穩定病狀時之時間相比較時或當與其他穩定移植接受者或未進行移植之健康對照相比較時，移植物排斥反應與較高細胞死亡速率相關且因此接收者血漿(或血清)中之供體DNA濃度(如移植器官之甲基化標誌所反映)將增加。類似於關於癌症所描述，可藉由偵測所有或一些特徵性特徵，包括多態性差異、針對移植固體器官之大小較短之DNA (YW Zheng等人2012 Clin Chem; 58: 549-558)及組織特異性甲基化概況來鑑別移植接受者之血漿中來源於供體之DNA。

#### H. 基於大小標準化甲基化

如上文及 Lun 等人 (FMF Lun 等人 Clin. Chem. 2013; doi:10.1373/clinchem.2013.212274)所述，甲基化密度(例如血漿DNA之甲基化密度)與DNA片段之大小相關。較短血漿DNA片段之甲基化密度分佈顯著低於較長片段。吾等提議一些具有異常血漿DNA片段化模式之非癌症病狀(例如，全身性紅斑狼瘡(SLE))可由於存在更豐富之較少甲基化之短血漿DNA片段而展現明顯血漿DNA低甲基化。換言之，血漿DNA之大小分佈可為血漿DNA甲基化密度之混淆因素。

圖34A展示SLE患者SLE04之血漿DNA之大小分佈。九個健康對照個

體之大小分佈顯示為灰色虛線，且SLE04之大小分佈顯示為黑色實線。與九個健康對照個體中相比，SLE04中之短血漿DNA片段更豐富。由於較短DNA片段一般較少甲基化，故此大小分佈模式可混淆對血漿DNA之甲基化分析且引起更明顯之低甲基化。

在一些實施例中，所量測之甲基化程度可經標準化以減少大小分佈對血漿DNA甲基化分析之混淆效應。舉例而言，可量測複數個位點處之DNA分子之大小。在各種實施例中，該量測可提供DNA分子之特定大小(例如長度)或簡單確定該大小處於特定範圍內，該範圍亦可對應於大小。隨後可比較標準化甲基化程度與截止值。存在若干種用於進行標準化之方式以減少大小分佈對血漿DNA甲基化分析之混淆效應。

在一個實施例中，可進行DNA(例如血漿DNA)之大小分級。大小分級可確保用與截止值一致的方式使用類似大小之DNA片段來確定甲基化程度。作為大小分級之一部分，可選擇具有第一大小(例如，第一長度範圍)之DNA片段，其中該第一截止值對應於該第一大小。可藉由僅使用所選DNA片段計算甲基化程度來達成標準化。

可用各種方式達成大小分級，例如藉由物理分離不同大小之DNA分子(例如藉由基於電泳或微流體之技術，或基於離心之技術)或藉由電子雜交分析達成。對於電子雜交分析，在一個實施例中，可進行血漿DNA分子之配對末端大規模平行定序。隨後可藉由將血漿DNA分子之兩個末端中之每一者的位置與參考人類基因組相比較來推斷定序分子之大小。隨後，可藉由選擇匹配一或多個大小選擇準則(例如，大小在規定範圍內之準則)的定序DNA分子來進行後續分析。因此，在一個實施例中，可分析具有類似大小(例如在規定範圍內)之片段的甲基化密度。可基於相同大小

範圍內之片段來確定截止值(例如，在方法2800之方塊2840中)。舉例而言，可由已知患有癌症或未患癌症之樣品確定甲基化程度，且可由此等甲基化程度確定截止值。

在另一實施例中，可確定循環DNA之甲基化密度與大小之間的函數關係。該函數關係可由函數之資料點或係數來定義。該函數關係可提供對應於各別大小之標度值(例如，較短大小可具有甲基化之相應增加)。在各種實施例中，該標度值可介於0與1之間或大於1。

該標準化可基於平均大小來進行。舉例而言，可計算對應於用於計算第一甲基化程度之DNA分子之平均大小，且該第一甲基化程度可乘以相應標度值(亦即，對應於平均大小)。作為另一實例，各DNA分子之甲基化密度可根據DNA分子之大小及DNA大小與甲基化之間的關係而加以標準化。

在另一實施例中，標準化可基於每個分子來進行。舉例而言，可獲得特定位點處之DNA分子之各別量(例如，如上所述)，且可由該函數關係鑑別對應於各別大小之標度值。對於非標準化計算，將在測定該位點處之甲基化指數時同樣地對各分子進行計數。對於標準化計算，分子對甲基化指數之貢獻可藉由對應於分子大小之定標因數加權。

圖34B及34C展示對來自SLE患者SLE04 (圖34B)及HCC患者TBR36 (圖34C)之血漿DNA之甲基化分析。外部圓顯示未進行電子雜交大小分級時血漿DNA之 $Z_{\text{甲基化}}$ 結果。內部圓顯示130 bp或更長血漿DNA之 $Z_{\text{甲基化}}$ 結果。對於SLE患者SLE04，在未進行電子雜交大小分級的情況下，84%區間顯示低甲基化。當僅分析130 bp或更長片段時，顯示低甲基化之區間之百分比降至15%。對於HCC患者TBR36，在進行及未進行電子雜交大小分

級的情況下，分別98.5%及98.6%區間顯示血漿DNA低甲基化。此等結果表明電子雜交大小分級可有效減少與血漿DNA片段化增加相關之假陽性低甲基化結果，例如在患有SLE之患者中或在其他發炎性病狀中。

在一個實施例中，可比較進行及未進行大小分級之分析的結果以指示大小是否對甲基化結果存在任何混淆效應。因此，另外或替代標準化，計算特定大小之甲基化程度可用於確定當進行或不進行大小分級的情況下高於截止值之區間百分比有所不同時是否存在假陽性之可能性，或是否僅特定甲基化程度有所不同。舉例而言，進行及未進行大小分級之樣品的結果之間存在顯著差異可用於指示由於異常片段化模式所致之假陽性結果之可能性。用於確定差異是否顯著之臨限值可經由分析一群癌症患者及一群非癌症對照個體來確定。

#### I. 分析血漿中之全基因組CpG島高甲基化

除一般低甲基化以外，癌症中通常亦觀測到CpG島高甲基化(SB Baylin等人2011 Nat Rev Cancer; 11: 726-734；PA Jones等人2007, Cell; 128: 683-692；M Esteller等人2007 Nat Rev Genet 2007; 8: 286-298；M Ehrlich等人2002 Oncogene 2002; 21: 5400-5413)。在此部分中，吾等描述使用CpG島高甲基化之全基因組分析來偵測並監測癌症。

圖35為根據本發明實施例之基於CpG島之高甲基化確定癌症程度分類之方法3500的流程圖。方法2800之複數個位點可包括CpG位點，其中該等CpG位點被組織成複數個CpG島，各CpG島包括一或多個CpG位點。各CpG島之甲基化程度可用於確定癌症程度之分類。

在方塊3510處，鑑別欲分析之CpG島。在此分析中，作為一個實例，吾等首先定義一組欲分析之CpG島，該等CpG島係用健康參考個體之

血漿中之相對較低甲基化密度加以表徵。在一個態樣中，參考組中之甲基化密度之變化可相對較小，以便允許更容易地偵測癌症相關之高甲基化。在一個實施例中，該等CpG島之平均甲基化密度小於參考組之第一百分比，且該參考組中之甲基化密度之變異係數小於第二百分比。

作為一個實例，出於說明目的，使用以下準則來鑑別適用CpG島：

- i. 參考組(例如健康個體)中之CpG島之平均甲基化密度 $<5\%$
- ii. 用於分析參考組(例如健康個體)之血漿中之甲基化密度的變異係數 $<30\%$ 。

此等參數可經調節以用於特定應用。根據吾等之資料集，基因組中之454個CpG島滿足此等準則。

在方塊3520處，計算各CpG島之甲基化密度。甲基化密度可如本文所述來測定。

在方塊3530處，確定各CpG島是否經高甲基化。舉例而言，為分析測試病例之CpG島高甲基化，可比較各CpG島之甲基化密度與參考組之相應資料。可將甲基化密度(甲基化程度之一個實例)與一或多個截止值相比較以確定特定島是否經高甲基化。

在一個實施例中，第一截止值可對應於參考組之甲基化密度平均值加規定百分比。另一截止值可對應於參考組之甲基化密度平均值加規定數目之標準差。在一個實施例中，計算z評分( $Z_{\text{甲基化}}$ )且與截止值相比較。舉例而言，若測試個體(例如針對癌症篩選之個體)中之CpG島滿足以下準則，則其被認為顯著高甲基化：

- i. 其甲基化密度比參考組之平均值高出 $2\%$ ；及
- ii.  $Z_{\text{甲基化}} > 3$

此等參數亦可經調節以用於特定應用。

在方塊3540處，高甲基化CpG島之甲基化密度(例如z評分)用於確定累積評分。舉例而言，在鑑別所有顯著高甲基化之CpG島之後，可計算涉及所有高甲基化CpG島之z評分之總和或z評分之函數的評分。評分之一個實例為累積機率(CP)評分，如另一部分中所述。累積機率評分使用 $Z_{\text{甲基化}}$ 來確定根據概率分佈(例如，具有3個自由度之史都登氏t概率分佈(Student's t probability distribution))偶然獲得該觀測結果的機率。

在方塊3550處，將累積評分與累積臨限值相比較以確定癌症程度之分類。舉例而言，若所鑑別之CpG島中之總高甲基化足夠大，則該生物體可鑑別為患有癌症。在一個實施例中，該累積臨限值對應於該參考組之最高累積評分。

## IX. 甲基化及CNA

如上文所提及，本文所述之甲基化分析方法可與基於血漿中來源於腫瘤之DNA之基因變化的其他方法組合使用。該等方法之實例包括分析癌症相關之染色體畸變(KCA Chan等人2013 Clin Chem; 59: 211-224; RJ Leary等人2012 Sci Transl Med; 4: 162ra154)。複本數畸變(CNA)之態樣描述於美國專利申請案第13/308,473號中。

### A. CNA

複本數畸變可藉由對與基因組之特定部分比對之DNA片段進行計數、標準化該計數且比較該計數與截止值來偵測。在各種實施例中，可藉由與該基因組之同一部分之另一單倍體比對之DNA片段之計數(相對單倍體劑量(RHDO))或與該基因組之另一部分比對之DNA片段之計數來進行標準化。



RHDO方法依賴於使用異型組合基因座。此部分中所述之實施例亦可藉由比較兩個區域而非同一區域之兩個單倍體而用於同型組合基因座，且因此為非單倍體特異性的。在相對染色體區域劑量方法中，將來自一個染色體區域之片段之數目(例如，如藉由對與該區域比對之序列讀段進行計數所測定)與預期值(其可來自參考染色體區域或來自已知健康之另一樣品中之相同區域)相比較。以此方式，將對片段之染色體區域進行計數，而不考慮定序標籤來自何種單倍體。因此，仍可使用不含異型組合基因座之序列讀段。為進行比較，一個實施例可在比較之前標準化該標籤計數。各區域係由至少兩個基因座(彼此分開)定義，且此等基因座處之片段可用於獲得關於該區域之集體值。

可藉由將與該區域比對之定序讀段之數目除以可與整個基因組比對之定序讀段之總數來計算特定區域之定序讀段(標籤)之標準化值。此標準化標籤計數允許比較來自一個樣品之結果與另一樣品之結果。舉例而言，標準化值可為預期來自特定區域之定序讀段之比例(例如，百分比或分數)，如上文所述。在其他實施例中，其他標準化方法為可能的。舉例而言，可藉由將一個區域之計數之數目除以參考區域之計數之數目來進行標準化(在以上情況下，參考區域剛好為整個基因組)。隨後可比較此標準化標籤計數與可由一或多個未展現癌症之參考樣品測定之臨限值。

隨後將測試病例之標準化標籤計數與一或多個參考個體(例如未患癌症者)之標準化標籤計數相比較。在一個實施例中，該比較係藉由計算該病例之特定染色體區域之z評分來進行。z評分可使用以下等式計算： $z$  評分=(該病例之標準化標籤計數-平均值)/SD，其中「平均值」為與參考樣品之特定染色體區域比對之平均標準化標籤計數；且SD為與參考樣品之

特定區域比對之標準化標籤計數之數目的標準差。因此， $z$ 評分為測試病例之染色體區域之標準化標籤計數與一或多個參考個體之相同染色體區域之平均標準化標籤計數相距的標準差數目。

在當測試生物體患有癌症時的情形下，腫瘤組織中經擴增之染色體區域將在血漿DNA中過度呈現。由此將產生 $z$ 評分之正值。另一方面，腫瘤組織中缺失之染色體區域將在血漿DNA中呈現不足。由此將產生 $z$ 評分之負值。 $z$ 評分之量值係由若干因素決定。

一個因素為生物樣品(例如血漿)中來源於腫瘤之DNA之部分濃度。樣品(例如血漿)中來源於腫瘤之DNA之部分濃度愈高，測試病例與參考病例之標準化標籤計數之間的差異將愈大。因此，將產生較大量值之 $z$ 評分。

另一因素為一或多個參考病例中之標準化標籤計數之變化。在測試病例之生物樣品(例如血漿)中之染色體區域具有相同程度之過度呈現時，參考組中之標準化標籤計數之變化愈小(亦即，標準差愈小)， $z$ 評分將愈高。同樣，在測試病例之生物樣品(例如血漿)中之染色體區域具有相同程度之呈現不足時，參考組中之標準化標籤計數之標準差愈小， $z$ 評分之負值愈大。

另一因素為腫瘤組織中之染色體畸變之量值。染色體畸變之量值係指特定染色體區域之複本數目變化(得/失)。腫瘤組織中之複本數目變化愈大，血漿DNA中特定染色體區域之過度呈現或呈現不足程度愈大。舉例而言，與損失染色體之兩個複本之一相比，染色體之兩個複本均損失將引起血漿DNA中之染色體區域之呈現不足程度更大，且因此產生負值更大的 $z$ 評分。通常，癌症中存在多個染色體畸變。各癌症中之染色體畸變可

進一步改變其性質(亦即擴增或缺失)、其程度(單一或多個複本得失)及其程度(畸變相對於染色體長度之大小)。

量測標準化標籤計數之精確度受所分析之分子之數目影響。吾等預期當部分濃度分別為約12.5%、6.3%及3.2%時，將需要分析15,000、60,000及240,000個分子以偵測在一個複本變化(得/失)下的染色體畸變。不同染色體區域之用於偵測癌症之標籤計數之其他細節描述於Lo等人之標題為「**Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing**」之美國專利公開案第2009/0029377號中，該公開案之全文係出於所有目的以引用的方式併入本文中。

實施例亦可使用大小分析作為標籤計數法之替代方案。亦可使用大小分析作為標準化標籤計數之替代方案。大小分析可使用如本文及美國專利申請案第12/940,992號中所提及之各種參數。舉例而言，可使用來自上文之Q或F值。該等大小值不需要藉由來自其他區域之計數進行標準化，因為此等值未用讀段之數目定標。單倍體特異性方法(諸如上文及美國專利申請案第13/308,473號中更詳細描述之RHDO方法)之技術亦可用於非特異性方法。舉例而言，可使用涉及區域之深度及改進之技術。在一些實施例中，當比較兩個區域時可考慮特定區域之GC偏移。因為RHDO方法使用同一區域，故不需要進行此修正。

雖然某些癌症通常可與特定染色體區域中之畸變一起存在，但該等癌症不會始終與該等區域中之畸變一起排他性地存在。舉例而言，其他染色體區域可顯示畸變，且該等其他區域之位置可能為未知的。此外，當篩選患者以鑑別早期癌症時，可能希望鑑別可顯示整個基因組中存在畸變之較寬範圍之癌症。為處理此等情形，實施例可以系統方式分析複數個區域

以確定何種區域顯示畸變。畸變數目及其位置(例如其是否連續)可用於例如基於展現畸變之各種區域之數目及位置確定畸變、確定癌症階段、提供癌症診斷(例如若該數目大於臨限值)及提供預後。

因此，實施例可基於顯示畸變之區域之數目來鑑別生物體是否患有癌症。因此，可測試複數個區域(例如3,000個)以鑑別許多展現畸變之區域。該等區域可覆蓋整個基因組或僅基因組之數個部分，例如非重複序列區域。

圖36為根據本發明實施例之使用複數個染色體區分析生物體之生物樣品之方法3600的流程圖。生物樣品包括核酸分子(亦稱為片段)。

在方塊3610處，鑑別生物體基因組之複數個區域(例如不重疊區域)。各染色體區域包括複數個基因座。區域可為1 Mb大小，或一些其他相等大小。對於區域為1 Mb大小之情形，整個基因組則可包括約3,000個區域，各區域具有預定大小及位置。該等預定區域可變化以適應欲使用之特定染色體之長度或規定數目之區域及本文所提及之任何其他準則。若區域具有不同長度，則該等長度可用於標準化結果，例如，如本文所述。可基於特定生物體之某些準則及/或基於所測試癌症之知識特異性地選擇該等區域。亦可任意地選擇該等區域。

在方塊3620處，對於複數個核酸分子中之每一者，鑑別生物體之參考基因組中之核酸分子之位置。可用本文所提及之任何方式測定位置，例如藉由對片段進行定序以獲得定序標籤且比對定序標籤與參考基因組。對於單倍體特異性方法，亦可測定分子之特定單倍體。

對該等染色體區域中之每一者進行方塊3630至3650。在方塊3630處，基於所鑑別之位置將核酸分子之各別群組鑑別為來自該染色體區域。

各別群組可包括至少一個位於該染色體區域之複數個基因座中之每一者處的核酸分子。在一個實施例中，該群組可為與該染色體區域之特定單倍體比對(例如，如以上RHDO方法中)之片段。在另一實施例中，該群組可具有與該染色體區域比對之任何片段。

在方塊3640處，電腦系統計算核酸分子之各別群組之各別值。各別值定義各別群組之核酸分子之性質。各別值可為本文所提及之值中的任何一者。舉例而言，該值可為該群組中之片段數目或該群組中之片段之大小分佈的統計值。各別值亦可為標準化值，例如，區域之標籤計數除以樣品標籤計數之總數或參考區域之標籤計數之數目。該各別值亦可為相對於另一值(例如，在RHDO中)之差異或比率，藉此提供該區域之差異性質。

在方塊3650處，將各別值與參考值比較以確定第一染色體區域是否展現缺失或擴增之分類。此參考值可為本文所述之任何臨限值或參考值。舉例而言，該參考值可為對正常樣品測定之臨限值。對於RHDO，該各別值可為兩個單倍體之標籤計數之差異或比率，且該參考值可為用於確定存在統計上顯著之偏差的臨限值。作為另一實例，該參考值可為另一單倍體或區域之標籤計數或大小值，且該比較可包括獲取差異或比率(或其函數)且隨後確定該差異或比率是否大於臨限值。

該參考值可基於其他區域之結果而變化。舉例而言，若鄰近區域亦顯示偏差(但與一個臨限值，例如，z評分3相比較小)，則可使用較低臨限值。舉例而言，若三個連續區域均大於第一臨限值，則較可能為癌症。因此，此第一臨限值可低於自非連續區域鑑別癌症所需之另一臨限值。具有三個即使具有較小偏差之區域(或多於三個)亦可具有足夠低的可保持敏感性及特異性之概率效應的機率。

在方塊3660處，測定分類為展現缺失或擴增之基因組區域之量。所計數之染色體區域可具有限制。舉例而言，僅可計數與至少一個其他區域連續之區域(或可需要連續區域具有某一大小，例如4個或4個以上區域)。對於區域不相等之實施例，數目亦可說明各別長度(例如，數目可為異常區域之總長度)。

在方塊3670處，將該量與量臨限值比較以確定樣品之分類。作為實例，該分類可為生物體是否患有癌症、癌症之階段及癌症之預後。在一個實施例中，對所有異常區域進行計數且使用單一臨限值，而不考慮區域出現在何處。在另一實施例中，臨限值可基於所計數之區域之位置及大小而變化。舉例而言，可將特定染色體或染色體臂上之區域之量與特定染色體(或臂)之臨限值相比較。可使用多個臨限值。舉例而言，特定染色體(或臂)上之異常區域之量必須大於第一臨限值，且基因組中之異常區域之總量必須大於第二臨限值。該臨限值可為被確定為展現缺失或擴增之區域的百分比。

區域之量的此臨限值亦可視所計數之區域之不平衡的強烈程度而定。舉例而言，用作確定癌症分類之臨限值的區域之量可視用於偵測各區域中之畸變的特異性及敏感性(異常臨限值)而定。舉例而言，若異常臨限值較低(例如z評分2)，則量臨限值可選擇為較高(例如150)。但，若異常臨限值較高(例如z評分3)，則量臨限值可能較低(例如50)。顯示畸變之區域之量亦可為加權值，例如一個顯示較高不平衡之區域與僅顯示極小不平衡之區域(亦即，對於畸變存在多於僅陽性及陰性之分類)相比可具有較高權重。作為一個實例，可使用z評分之總和，藉此使用加權值。

因此，顯示標準化標籤計數(或群組之性質的其他各別值)之顯著過度

呈現或呈現不足之染色體區域之量(其可包括數目及/或大小)可用於體現疾病之嚴重程度。具有異常標準化標籤計數之染色體區域之量可由兩個因素決定，即腫瘤組織中之染色體畸變之數目(或大小)及生物樣品(例如血漿)中來源於腫瘤之DNA之部分濃度。較晚期癌症趨向於展現較多(及較大)染色體畸變。因此，樣品(例如血漿)中將可能偵測到較多癌症相關之染色體畸變。在患有較晚期癌症之患者中，較高腫瘤負荷將引起血漿中來源於腫瘤之DNA之部分濃度較高。因此，血漿樣品中將更容易偵測到腫瘤相關之染色體畸變。

一種在不犧牲特異性的情況下改良敏感性之可能之方法將考慮相鄰染色體區段之結果。在一個實施例中，z評分之截止值仍為 $>2$ 及 $<-2$ 。然而，僅當兩個連續區段將顯示相同類型畸變，例如兩個區段之z評分 $>2$ 時，將染色體區域分類為可能異常。在其他實施例中，可使用較高截止值將鄰近區段之z評分加和在一起。舉例而言，可對三個連續區段之z評分進行求和且可使用截止值5。此概念可延伸至三個以上連續區段。

量與異常臨限值之組合亦可視分析目的及生物體之任何先前知識(或其缺乏)而定。舉例而言，若針對癌症篩選正常健康群體，則通常將可能在區域之量(亦即，區域數目之高臨限值)與當區域鑑別為具有畸變時之異常臨限值兩方面使用高特異性。但在具有較高風險之患者(例如，申訴腫瘤或家族史之患者、吸菸者、慢性人類乳頭狀瘤病毒(HPV)攜帶者、肝炎病毒攜帶者或其他病毒攜帶者)中，則臨限值可能較低以便具有較高敏感性(較少假陰性)。

在一個實施例中，若在偵測染色體畸變時使用1 Mb解析度及來源於腫瘤之DNA之6.3%較低偵測極限，則各1 Mb區段中之分子數目將需要為

60,000。對於整個基因組，此需要將轉變為約18000萬(60,000個讀段/Mb×3,000 Mb)可比對讀段。

區段大小愈小，用於偵測較小染色體畸變之解析度將愈高。然而，由此將增加對所分析之分子總數之需求。較大區段大小將以解析度為代價減少分析所需之分子數目。因此，僅可偵測較大畸變。在一個實施例中，可使用較大區域，可將顯示畸變之區段再劃分，且分析此等亞區以獲得較佳解析度(例如，如上文所述)。若獲得欲偵測之缺失或擴增之大小(或欲偵測之最小濃度)之估計值，則可確定欲分析之分子數目。

#### B. 基於經亞硫酸氫鹽處理之血漿DNA之定序的CNA

在腫瘤組織中往往可觀測到全基因組低甲基化及CNA。此處，吾等證明可由血漿DNA之亞硫酸氫鹽定序同時獲得CNA及癌症相關之甲基化變化之資訊。因為可對同一資料集進行兩種類型分析，故實際上CNA分析無額外成本。其他實施例可使用不同程序來獲得甲基化資訊及基因資訊。在其他實施例中，可結合CNA分析對癌症相關之高甲基化進行類似分析。

圖37A展示對患者TBR36之腫瘤組織、未經亞硫酸氫鹽(BS)處理之血漿DNA及經亞硫酸氫鹽處理之血漿DNA (由內至外)的CNA分析。圖37A展示對患者TBR36之腫瘤組織、未經亞硫酸氫鹽(BS)處理之血漿DNA及經亞硫酸氫鹽處理之血漿DNA (由內至外)的CNA分析。最外環展示染色體表意文字。各點表示1 Mb區域之結果。綠色、紅色及灰色點分別表示具有複本數獲得、複本數損失及無複本數變化之區域。對於血漿分析，展示z評分。兩個同軸線之間存在差異5。對於腫瘤組織分析，展示複本數。兩個同軸線之間存在一個複本差異。圖38A展示對患者TBR34之腫



瘤組織、未經亞硫酸氫鹽(BS)處理之血漿DNA及經亞硫酸氫鹽處理之血漿DNA (由內至外)的CNA分析。在經亞硫酸氫鹽及未經亞硫酸氫鹽處理之血漿樣品中所偵測到之CNA模式一致。

在腫瘤組織、未經亞硫酸氫鹽處理之血漿及經亞硫酸氫鹽處理之血漿中所偵測到之CNA模式一致。為進一步評估經亞硫酸氫鹽及未經亞硫酸氫鹽處理之血漿之結果之間的一致性，建構散點圖。圖37B為展示針對患者TBR36使用1 Mb區間之經亞硫酸氫鹽及未經亞硫酸氫鹽處理之血漿進行CNA偵測之z評分之關係的散點圖。觀測到兩種分析之z評分之正相關性( $r=0.89$ ， $p<0.001$ ，皮爾森相關性(Pearson correlation))。圖38B為展示針對患者TBR34使用1 Mb區間之經亞硫酸氫鹽處理及未經亞硫酸氫鹽處理之血漿進行CNA偵測之z評分之關係的散點圖。觀測到兩種分析之z評分之正相關性( $r=0.81$ ， $p<0.001$ ，皮爾森相關性)。

### C. 癌症相關之CNA及甲基化變化的協同分析

如上所述，CNA分析可涉及對各1 Mb區域中之序列讀段之數目進行計數，而甲基化密度分析可涉及偵測經甲基化之CpG二核苷酸處之胞嘧啶殘基之比例。此兩種分析之組合可獲得關於癌症偵測之協同資訊。舉例而言，甲基化分類及CNA分類可用於確定癌症程度之第三分類。

在一個實施例中，存在癌症相關之CNA或甲基化變化可用於指示可能存在癌症。在該實施例中，當測試個體之血漿中存在CNA或甲基化變化時，可增加偵測癌症之敏感性。在另一實施例中，存在兩種變化可用於指示存在癌症。在該實施例中，可改良測試特異性，因為一些非癌症個體中可能偵測到兩種變化類型中之任一者。因此，僅當該第一分類及該第二分類均指示癌症時，該第三分類可對癌症呈陽性。

徵募26個HCC患者及22個健康個體。自各個體收集血液樣品且在亞硫酸氫鹽處理後對血漿DNA進行定序。對於HCC患者，在診斷時收集血液樣品。舉例而言，存在大量CNA定義為有>5%區間顯示z評分<-3或>3。存在大量癌症相關之低甲基化定義為有>3%區間顯示z評分<-3。作為實例，區域(區間)之量可表示為原始區間計數、百分比及區間長度。

表3展示對經亞硫酸氫鹽處理之血漿DNA使用大規模平行定序在26個HCC患者之血漿中偵測到大量CNA及甲基化變化。

		CNA	
		存在	不存在
甲基化變化	存在	12	6
	不存在	1	7

表3

癌症相關之甲基化變化及CNA之偵測率分別為69%及50%。若任一準則之存在用於指示可能存在癌症，則檢測率(亦即，診斷敏感性)改良至73%。

據顯示，兩個患者之結果顯示存在CNA (圖39A)或甲基化變化(圖39B)。圖39A為展示對HCC患者TBR240之經亞硫酸氫鹽處理之血漿之CNA (內環)及甲基化分析(外環)的Circos圖。對於CNA分析，綠色、紅色及灰色點分別表示具有染色體獲得、損失及無複本數變化之區域。對於甲基化分析，綠色、紅色及灰色點分別表示具有高甲基化、低甲基化及正常甲基化之區域。在此患者中，在血漿中偵測到癌症相關之CNA，而甲基化分析未顯示大量癌症相關之低甲基化。圖39B為展示對HCC患者TBR164之經亞硫酸氫鹽處理之血漿之CNA (內環)及甲基化分析(外環)的Circos圖。在此患者中，在血漿中偵測癌症相關之低甲基化。然而，無法觀測到大量CNA。顯示存在CNA及甲基化變化兩者之兩個患者之結果示

於圖48A (TBR36)及49A (TBR34)中。

表4展示對經亞硫酸氫鹽處理之血漿DNA使用大規模平行定序在22個對照個體之血漿中偵測到大量CNA及甲基化變化。使用自展法(亦即留一法)來評估各對照個體。因此，當評估特定個體時，使用其他21個個體來計算對照組之平均值及SD。

		CNA	
		存在	不存在
甲基化變化	存在	1	2
	不存在	1	18

表4

偵測大量甲基化變化及CNA之特異性分別為86%及91%。若需要存在兩種準則以指示可能存在癌症，則特異性改良至95%。

在一個實施例中，CNA及/或低甲基化陽性樣品被視為對癌症呈陽性，且當兩者均不可偵測時樣品被視為陰性。使用「或」邏輯提供較高敏感性。在另一實施例中，僅CNA及低甲基化陽性樣品被視為對癌症呈陽性，藉此提供較高特異性。在另一實施例中，可使用三級分類。將個體分類為i. 均正常；ii. 一項異常；iii. 均異常。

不同後續策略可用於此三個分類。舉例而言，可對(iii)之個體進行最密集之後續方案，例如涉及整體成像；可對(ii)之個體進行較不密集之後續方案，例如在若干週之相對較短時間間隔後重複血漿DNA定序；且可對(i)之個體進行最不密集之後續方案，諸如在許多年後再測試。在其他實施例中，甲基化及CNA量測可與其他臨床參數(例如成像結果或血清生物化學)聯合用於進一步改進該分類。

#### D. 出於治療性目的之治療後血漿DNA分析之預測值

血漿中存在癌症相關之CNA及/或甲基化變化將指示癌症患者之循環

中存在來源於腫瘤之DNA。將預期在治療(例如，手術)後減少或清除此等癌症相關之變化。另一方面，治療後血漿中之此等變化持續存在可能指示所有腫瘤細胞自體內不完全移除，且可為疾病復發之適用預測物。

在出於治癒性目的手術切除腫瘤後一週，自兩個HCC患者TBR34及TBR36收集血液樣品。對經亞硫酸氫鹽處理之治療後血漿樣品進行CNA及甲基化分析。

圖40A展示對在HCC患者TBR36手術切除腫瘤之前(內環)及之後(外環)收集之經亞硫酸氫鹽處理之血漿DNA之CNA分析。各點表示1 Mb區域之結果。綠色、紅色及灰色點分別表示具有複本數獲得、複本數損失及無複本數變化之區域。在治療前觀測到之CNA中的大部分在腫瘤切除術消失。顯示z評分 $<-3$ 或 $>3$ 之區間的比例自25%降至6.6%。

圖40B展示對HCC患者TBR36手術切除腫瘤之前(內環)及之後(外環)收集之經亞硫酸氫鹽處理之血漿DNA之甲基化分析。綠色、紅色及灰色點分別表示具有高甲基化、低甲基化及正常甲基化之區域。顯示顯著低甲基化之區間之比例自90%顯著降至7.9%，且低甲基化程度亦顯示顯著降低。此患者在腫瘤切除術之後22個月具有完全臨床減輕。

圖41A展示對HCC患者TBR34手術切除腫瘤之前(內環)及之後(外環)收集之經亞硫酸氫鹽處理之血漿DNA之CNA分析。雖然在手術切除腫瘤之後顯示CNA之區間之數目及受影響區間中之CNA之量值均有所減少，但在手術後血漿樣品中可觀測到殘餘CNA。紅色圓突出顯示殘餘CNA最明顯之區域。顯示z評分 $<-3$ 或 $>3$ 之區間的比例自57%降至12%。

圖41B.展示對HCC患者TBR34手術切除腫瘤之前(內環)及之後(外環)收集之經亞硫酸氫鹽處理之血漿DNA之甲基化分析。腫瘤切除術後低甲

基化量值降低，其中低甲基化區間之平均 $z$ 評分自 $-7.9$ 降至 $-4.0$ 。然而， $z$ 評分 $<-3$ 之區間的比例顯示相對變化，自 $41\%$ 增至 $85\%$ 。此觀測結果可能指示治療後存在殘餘癌細胞。臨床上，在腫瘤切除術後3個月在其餘未切除肝臟中偵測到多個腫瘤節結病灶。在手術後4個月觀測到肺轉移。患者在手術後8個月死於局部復發及轉移性疾病。

此兩個患者(TBR34及TBR36)中之觀測結果表明存在CNA及低甲基化之殘餘癌症相關變化可用於在出於治癒目的之治療後監測及預測癌症患者。資料亦顯示所偵測之血漿CNA量之變化程度可與評定血漿DNA低甲基化程度之變化程度一起協同使用以預測及監測治療效力。

因此，在一些實施例中，在治療之前獲得一個生物樣品且在治療(例如手術)之後獲得第二生物樣品。針對第一樣品獲得第一值，諸如區域之 $z$ 評分(例如，區域甲基化程度及CNA之標準化值)及顯示低甲基化及CNA(例如擴增或缺失)之區域之數目。可針對第二樣品獲得第二值。在另一實施例中，可在治療後獲得第三或甚至其他樣品。顯示低甲基化及CNA(例如擴增或缺失)之區域之數目可獲自第三或甚至其他樣品。

如關於圖40A及41A所描述，可將第一樣品中顯示低甲基化之區域之第一數目與第二樣品中顯示低甲基化之區域之第二量相比較。如關於圖40B及41B所描述，可將第一樣品中顯示低甲基化之區域之第一量與第二樣品中顯示低甲基化之區域之第二量相比較。比較第一量與第二量及第一數目與第二數目可用於確定治療之預後。在各種實施例中，僅一個比較對預後可為決定性的，或可使用兩個比較。在獲得第三或甚至其他樣品之實施例中，此等樣品中之一或多者可獨立地或與第二樣品聯合用於確定治療之預後。

在一個實施例中，當第一量與第二量之間的第一差異低於第一差異臨限值時預測預後較糟。在另一實施例中，當第一數目與第二數目之間的第二差異低於第二差異臨限值時預測預後較糟。臨限值可相同或不同。在一個實施例中，第一差異臨限值及第二差異臨限值為零。因此，對於以上實例，甲基化值之間的差異將指示患者TBR34之預後較糟。

若第一差異及/或第二差異高於同一臨限值或各別臨限值，則預後可較佳。預後分類可視差異高於或低於臨限值的程度而定。可使用多個臨限值來提供各種分類。較大差異可預測較佳結果，且較小差異(及甚至負值)可預測較糟結果。

在一些實施例中，亦注意獲取各種樣品之時間點。藉由該等時間參數，可確定量之動力學或變化率。在一個實施例中，血漿中腫瘤相關之低甲基化快速降低及/或血漿中腫瘤相關之CNA快速降低將預示良好預後。反之，血漿中腫瘤相關之低甲基化靜止或快速增加及/或腫瘤相關之CNA靜止或快速增加將預示不良預後。甲基化及CNA量測可與其他臨床參數(例如成像結果或血清生物化學或蛋白質標記物)聯合用於預測臨床結果。

實施例可使用除血漿以外的其他樣品。舉例而言，可由癌症患者血液中循環之腫瘤細胞、細胞游離DNA或尿液、大便、唾液、痰液、膽液、胰液、子宮頸拭棉、生殖道(例如陰道)分泌物、腹水、胸膜液、精液、汗液及淚液中之腫瘤細胞量測腫瘤相關之甲基化畸變(例如低甲基化)及/或腫瘤相關之CNA。

在各種實施例中，可由患有乳癌、肺癌、結腸直腸癌、胰臟癌、卵巢癌、鼻咽癌、子宮頸癌、黑色素瘤、腦腫瘤等之患者的血液或血漿偵測腫瘤相關之甲基化畸變(例如低甲基化)及/或腫瘤相關之CNA。實際上，

因為甲基化及基因變化(諸如CNA)在癌症中為普遍現象，故所描述之方法可用於所有癌症類型。甲基化及CNA量測可與其他臨床參數(例如成像結果)聯合用於預測臨床結果。實施例亦可用於篩選並監測患有瘤前期病變(例如腺瘤)之患者。

因此，在一個實施例中，在治療前獲取生物樣品，且在治療後重複CNA及甲基化量測。量測可產生被確定為展現缺失或擴增之區域的隨後第一量，且可產生被確定為區域甲基化程度超過各別區域截止值之區域的隨後第二量。可比較該第一量與該隨後第一量，且可比較該第二量與該隨後第二量，以確定該生物體之預後。

用於確定生物體之預後的比較可包括確定第一量與隨後第一量之間的第一差異，且可比較該第一差異與一或多個第一差異臨限值以確定預後。用於確定生物體之預後的比較亦可包括確定第二量與隨後第二量之間的第二差異，且可比較該第二差異與一或多個第二差異臨限值。該等臨限值可為零或另一數值。

與當第一差異高於第一差異臨限值時相比，當第一差異低於第一差異臨限值時可預測預後較糟。與當第二差異高於第二差異臨限值時相比，當第二差異低於第二差異臨限值時可預測預後較糟。治療之實例包括免疫療法、手術、放射線療法、化學療法、基於抗體之療法、基因療法、表觀遺傳療法或靶向療法。

## E. 效能

現針對CNA及甲基化分析描述不同數目之序列讀段及區間大小之診斷效能。

### 1. 序列讀段之數目

根據一個實施例，吾等分析32個健康對照個體、26個罹患肝細胞癌之患者及20個罹患其他類型癌症(包括鼻咽癌、乳癌、肺癌、神經內分泌癌及平滑肌肉瘤)之患者的血漿DNA。隨機選擇32個健康個體中的二十二個作為參考組。此22個參考個體之平均值及標準差(SD)用於測定甲基化密度及基因組呈現之正常範圍。使用Illumina配對末端定序套組，使用自各個體之血漿樣品提取之DNA進行定序文庫建構。隨後對定序文庫進行亞硫酸氫鹽處理，由此將未甲基化之胞嘧啶殘基轉化成尿嘧啶。使用Illumina HiSeq2000定序器之一個泳道對各血漿樣品之經亞硫酸氫鹽轉化之定序文庫進行定序。

在鹼基檢出之後，移除片段末端上之接附序列及低品質鹼基(亦即品質評分<5)。隨後藉由稱為Methy-Pipe之甲基化資料分析渠道處理呈FASTQ格式之經修切讀段(P Jiang等人2010, IEEE生物資訊學與生物醫學國際會議, doi:10.1109/BIBMW.2010.5703866)。為比對經亞硫酸氫鹽轉化之定序讀段，吾等首先使用參考人類基因組(NCBI版本36/hg19)將沃森股及克立克股上之所有胞嘧啶殘基獨立地電子雜交轉化成胸嘧啶。隨後，吾等將所有經處理讀段中之各胞嘧啶電子雜交轉化成胸嘧啶，且保存各經轉化殘基之位置資訊。使用SOAP2來比對經轉化之讀段與兩個轉化前參考人類基因組(R Li等人2009 *Bioinformatics* 25:1966-1967)，其中各比對讀段最多允許兩個錯配。僅可定位於獨特基因組位置之讀段用於下游分析。含糊讀段定位於沃森股及克立克股兩者，且移除複製之(純系)讀段。CpG二核苷酸情形下之胞嘧啶殘基用於下游甲基化分析。在比對後，基於電子雜交轉化期間保存之位置資訊恢復最初存在於定序讀段上之胞嘧啶。將CpG二核苷酸中之經恢復胞嘧啶評分為甲基化。將CpG二核苷酸中之胸



嘧啶評分為未甲基化。

對於甲基化分析，將基因組分成相等大小之區間。所測試區間之大小包括50 kb、100 kb、200 kb及1 Mb。各區間之甲基化密度計算為CpG二核苷酸情形下之甲基化胞嘧啶之數目除以CpG位置處之胞嘧啶之總數。在其他實施例中，區間大小在基因組上可為不相等的。在一個實施例中，在多個個體之間比較該等具有不相等大小之區間中之各區間。

為確定測試病例之血漿甲基化密度是否正常，將甲基化密度與參考組之結果相比較。隨機選擇32個健康個體中之二十二個作為參考組用於計算甲基化z評分( $Z_{\text{甲基化}}$ )。

$$Z_{\text{甲基化}} = \frac{MD_{\text{測試}} - \overline{MD}_{\text{參考}}}{MD_{\text{SD}}}$$

其中  $MD_{\text{測試}}$  為測試病例之特定1 Mb區間之甲基化密度； $\overline{MD}_{\text{參考}}$  為參考組之對應區間之平均甲基化密度；且  $MD_{\text{SD}}$  為參考組之對應區間之甲基化密度之SD。

對於CNA分析，確定定位於各1 Mb區間之定序讀段之數目(KCA Chan等人2013 Clin Chem 59:211-24)。在使用如先前所述(EZ Chen等人2011 PLoS One 6: e21791)之局部加權散點圖平滑回歸(Locally Weighted Scatter Plot Smoothing regression)修正GC偏移後測定各區間之定序讀段密度。對於血漿分析，將測試病例之定序讀段密度與參考組相比較以計算CNA z評分( $Z_{\text{CNA}}$ )：

$$Z_{\text{CNA}} = \frac{RD_{\text{測試}} - \overline{RD}_{\text{參考}}}{RD_{\text{SD}}}$$

其中  $RD_{\text{測試}}$  為測試病例之特定1 Mb區間之定序讀段密度； $\overline{RD}_{\text{參考}}$  為參考組之對應區間之平均定序讀段密度；且  $RD_{\text{SD}}$  為參考組之對應區間之定序讀段密度之SD。若區間之 $Z_{\text{CNA}} < -3$ 或 $> 3$ ，則該區間定義為展現CNA。

每個病例平均獲得9300萬個比對讀段(範圍：3900萬至14200萬)。為評估定序讀段之數目減少對診斷效能之效應，吾等自各病例隨機選擇1000萬個比對讀段。對於定序讀段有所減少之資料集，使用同一參考個體集來確定各1 Mb區間之參考範圍。對於各病例，測定顯示顯著低甲基化，亦即 $Z_{\text{甲基化}} < -3$ 之區間的百分比及具有CNA，亦即 $Z_{\text{CNA}} < -3$ 或 $> 3$ 之區間的百分比。使用接受者工作特徵(ROC)曲線來說明具有來自1個泳道之所有定序讀段及1000萬個讀段/病例之資料集之全基因組低甲基化及CNA分析之診斷效能。在ROC分析中，所有32個健康個體均用於分析。

圖42展示利用不同數目之定序讀段進行全基因組低甲基化分析之診斷效能的圖式。對於低甲基化分析，分析來自一個泳道之所有定序讀段及1000萬個讀段/病例之兩個資料集之間的ROC曲線之曲線下面積無顯著不同( $P=0.761$ )。對於CNA分析，診斷效能劣化，其中當定序讀段之數目自使用一個泳道之資料減至1000萬時曲線下面積顯著減小( $P < 0.001$ )。

## 2. 使用不同區間大小之效應

除將基因組分成1 Mb區間以外，吾等亦研究是否可使用較小區間大小。理論上，使用較小區間可能降低區間內之甲基化密度之變化性。此係因為不同基因組區域之間的甲基化密度可廣泛變化。當區間較大時，包括具有不同甲基化密度之區域的概率將增加，且因此，將導致區間之甲基化密度之變化性總體增加。

雖然使用較小區間大小可能降低與區域間差異相關之甲基化密度之變化性，但另一方面由此將減少定位於特定區間之定序讀段之數目。定位於個別區間之讀段減少將增加由於取樣變化所致之變化性。對於特定診斷應用之需求，例如每個樣品之定序讀段總數及所使用之DNA定序器類

型，可產生甲基化密度之最低總體變化性之最佳區間大小可以實驗方式測定。

圖43為展示基於全基因組低甲基化分析在不同區間大小(50 kb、100 kb、200 kb及1 Mb)下偵測癌症之ROC曲線的圖式。所示P值係針對在1 Mb區間大小下進行的曲線下面積比較。當區間大小自1 Mb降至200 kb時，可見改良趨勢。

#### F. 累積機率評分

甲基化及CNA之區域之量可為各種值。以上實例描述許多區域超過截止值或該等顯示顯著低甲基化或CNA之區域之百分比作為對樣品是否與癌症相關進行分類的參數。該等方法不考慮個別區間之畸變量值。舉例而言， $Z_{\text{甲基化}}$ 為-3.5之區間將與 $Z_{\text{甲基化}}$ 為-30之區間相同，因為兩者均將歸類為具有顯著低甲基化。然而，血漿中之低甲基化變化程度，亦即 $Z_{\text{甲基化}}$ 值之量值，受樣品中與癌症相關之DNA之量影響，且因此，可補充顯示畸變之區間之百分比之資訊以體現腫瘤負荷。血漿樣品中之腫瘤DNA之部分濃度較高將引起較低甲基化密度且此結果將轉變為較低 $Z_{\text{甲基化}}$ 值。

##### 1. 作為診斷參數之累積機率評分

為使用來自畸變量值之資訊，吾等開發出一種稱為累積機率(CP)評分之方法。基於正態分佈概率函數，將各 $Z_{\text{甲基化}}$ 值轉變為偶然具有該觀測結果之機率。

對於 $Z_{\text{甲基化}} < -3$ 之區間(i)，CP評分計算為：

$$\text{CP 評分} = \sum -\log(\text{Prob}_i)$$

其中 $\text{Prob}_i$ 為根據具有3個自由度之史都登氏t分佈區間(i)之 $Z_{\text{甲基化}}$ 的機率，且 $\log$ 為自然對數函數。在另一實施例中，可使用以10 (或其他數字)

為基數之對數。在其他實施例中，可應用其他分佈，例如但不限於正態分佈及 $\gamma$ 分佈，將 $z$ 評分轉換成CP。

較大CP評分指示正態群體中偶然具有該偏離甲基化密度之機率較低。因此，高CP評分將指示樣品中具有異常低甲基化之DNA，例如存在癌症相關之DNA的概率較高。

與顯示畸變之區間的百分比相比，CP評分量測具有較高動態範圍。雖然不同患者之間的腫瘤負荷可廣泛變化，但較大CP值範圍將適用於體現具有相對較高及相對較低腫瘤負荷之患者的腫瘤負荷。另外，使用CP評分可能對偵測血漿中之腫瘤相關之DNA之濃度變化更敏感。此有利於監測治療反應及預測。因此，治療期間CP評分降低指示對治療之良好反應。治療期間缺乏CP評分降低或甚至增加將指示不良反應或缺乏反應。對於預測，較高CP評分指示較高腫瘤負荷且表明不良預後(例如死亡或腫瘤進展之概率較高)。

圖44A展示具有畸變之區間之累積機率(CP)及百分比之診斷效能。該兩種類型診斷演算法之曲線下面積之間無顯著差異( $P=0.791$ )。

圖44B展示整體低甲基化、CpG島高甲基化及CNA之血漿分析之診斷效能。在每個樣品一個定序泳道(對於低甲基化分析為200 kb區間大小，對於CNA為1 Mb區間大小，且CpG島係根據加利福尼亞州大學聖克魯斯分校(The University of California, Santa Cruz, UCSC)主持之資料庫定義)的情況下，所有三種類型分析之曲線下面積均大於0.90。

在後續分析中，對於三種類型分析中之每一種，將對照個體中之最高CP評分用作截止值。此等截止值之選擇得到100%診斷特異性。一般低甲基化、CpG島高甲基化及CNA分析之診斷敏感性分別為78%、89%及

52%。在46個癌症患者中的43個中，偵測三種類型畸變中之至少一種，因此產生93.4%敏感性及100%特異性。吾等之結果指示該三種類型分析可協同用於偵測癌症。

圖45展示肝細胞癌患者中之整體低甲基化、CpG島高甲基化及CNA之結果的表。該三種類型分析之CP評分截止值分別為960、2.9及211。正CP評分結果呈粗體且加下劃線。

圖46展示罹患除肝細胞癌以外之癌症的患者中之整體低甲基化、CpG島高甲基化及CNA之結果的表。該三種類型分析之CP評分截止值分別為960、2.9及211。正CP評分結果呈粗體且加下劃線。

## 2. 應用CP評分進行癌症監測

在治療前後自HCC患者TBR34收集系統樣品。分析樣品之整體低甲基化。

圖47展示病例TBR34之血漿甲基化之系列分析。最內環展示白血球層(黑色)及腫瘤組織(紫色)之甲基化密度。對於血漿樣品，展示各1 Mb區間之 $Z_{\text{甲基化}}$ 。兩條線之間的差異表示 $Z_{\text{甲基化}}$ 差異為5。紅色及灰色點表示低甲基化及甲基化密度與參考組相比無變化之區間。自第2個內環向外分別為治療前、腫瘤切除術後3天及2個月時獲取之血漿樣品。治療前，在血漿中可觀測到高度低甲基化且超過18.5%區間具有 $<-10$ 之 $Z_{\text{甲基化}}$ 。在腫瘤切除術後3天時，可觀測到血漿中之低甲基化程度降低，其中無區間具有 $<-10$ 之 $Z_{\text{甲基化}}$ 。

病例編號	時間點	甲基化分析		
		顯示顯著低甲基化之區間的百分比	累積機率(CP)評分	累積z評分
TBR34	OT之前	62.6%	37,573	14,285
	OT之後3天	80.5%	17,777	9,195
	OT之後2個月	40.1%	15,087	5,201

表5

表5顯示雖然手術切除腫瘤後3天時低甲基化變化之量值降低，但展現畸變之區間的百分比顯示反常增加。另一方面，CP評分更準確地顯示血漿中之低甲基化程度之降低，且可進一步體現腫瘤負荷之變化。

在OT後2個月時，仍存在顯著百分比的顯示低甲基化變化之區間。CP評分亦在約15,000處保持靜止。隨後將此患者診斷為在手術後3個月時在仍未切除之肝臟中具有多病灶腫瘤沈積(先前手術時未知)且指示在手術後4個月時具有多個肺轉移。該患者在手術後8個月時死於轉移性疾病。此等結果表明對於體現腫瘤負荷，CP評分可能比具有畸變之區間之百分比更有效。

總體而言，CP可適用於需要量測血漿中之腫瘤DNA之量的應用。該等應用之實例包括：預測及監測癌症患者(例如觀測對治療之反應，或觀測腫瘤進展)。

累加z評分為z評分之直接和，亦即，未轉化成機率。在此實例中，累加z評分與CP評分顯示相同特性。在其他情況中，對於監測殘餘疾病而言，由於CP評分之動態範圍較大，故CP可比累加z評分更敏感。

## **X. CNA對甲基化之影響**

使用CNA及甲基化確定癌症程度之各別分類描述於上文，其中將該等分類組合以提供第三分類。除該組合以外，CNA可用於改變甲基化分析之截止值及藉由比較具有不同CNA特徵之區域群組的甲基化程度來鑑別假陽性。舉例而言，可將過度豐度之甲基化程度(例如， $Z_{CNA} > 3$ )與正常豐度之甲基化程度(例如， $-3 < Z_{CNA} < 3$ )相比較。首先，描述CNA對甲基化程度之影響。

### **A. 改變具有染色體得失之區域處的甲基化密度**

由於腫瘤組織一般顯示總體低甲基化，故癌症患者之血漿中存在來源於腫瘤之DNA將引起甲基化密度當與非癌症個體比較時有所降低。癌症患者血漿中之低甲基化程度理論上與血漿樣品中來源於腫瘤之DNA之部分濃度成比例。

對於腫瘤組織中顯示染色體獲得之區域，將有額外劑量之腫瘤DNA自經擴增之DNA區段釋放至血漿中。腫瘤DNA對血漿之此種增加貢獻理論上將引起受影響區域之血漿DNA之低甲基化程度較高。額外因素為將預期顯示擴增之基因組區域可賦予腫瘤細胞以生長優勢，且因此將預期得以表現。該等區域一般經低甲基化。

相比之下，對於腫瘤組織中顯示染色體損失的區域，腫瘤DNA對血漿之貢獻減少將引起與無複本數變化之區域相比較低程度之低甲基化。另一因素為腫瘤細胞中缺失之基因組區域可含有腫瘤抑制基因且可有利於腫瘤細胞使該等區域沉默。因此，預期該等區域經高甲基化之概率較高。

在此，吾等使用兩個HCC患者(TBR34及TBR36)之結果來說明此效應。圖48A (TBR36)及49A (TBR34)具有突出顯示具有染色體得失之區域及相應甲基化分析之圖。圖48B及49B分別展示患者TBR36及TBR34之損失、正常及獲得之甲基化z評分的圖。

圖48A展示顯示HCC患者TBR36之經亞硫酸氫鹽處理之血漿DNA中之CNA (內環)及甲基化變化(外環)的Circos圖。紅色圓突出顯示具有染色體得失之區域。展示染色體獲得之區域之低甲基化程度大於無複本數變化之區域。展示染色體損失之區域之低甲基化程度小於無複本數變化之區域。圖48B為HCC患者TBR36之具有染色體得失之區域及無複本數變化之區域的甲基化z評分的圖。與無複本變化之區域相比較，具有染色體獲得

之區域具有負值較大之 $z$ 評分(低甲基化程度較大)，且具有染色體損失之區域具有負值較小之 $z$ 評分(低甲基化程度較小)。

圖49A展示顯示HCC患者TBR34之經亞硫酸氫鹽處理之血漿DNA中之CNA (內環)及甲基化變化(外環)的Circos圖。圖49B為HCC患者TBR34之具有染色體得失之區域及無複本數變化之區域的甲基化 $z$ 評分的圖。患者TBR36中具有染色體得失之區域之間的甲基化密度差異大於患者TBR34中，因為前一患者中來源於腫瘤之DNA之部分濃度較高。

在此實例中，用於測定CNA之區域與用於測定甲基化之區域相同。在一個實施例中，各別區域截止值視各別區域是否展現缺失或擴增而定。在一個實施例中，當各別區域展現擴增時，與當未展現擴增時相比，各別區域截止值(例如，用於測定低甲基化之 $z$ 評分截止值)具有較大量值(例如，該量值可大於3，且可使用小於-3之截止值)。因此，為測試低甲基化，當各別區域展現擴增時，與當未展現擴增時相比，各別區域截止值可具有較大負值。預期該實施例可改良用於偵測癌症之測試的特異性。

在另一實施例中，當各別區域展現缺失時，與當未展現缺失時相比，各別區域截止值具有較少量值(例如小於3)。因此，為測試低甲基化，當各別區域展現缺失時，與當未展現缺失時相比，各別區域截止值可具有較小負值。預期該實施例可改良用於偵測癌症之測試的敏感性。以上實施例中之截止值之調節可視特定診斷情形之所要敏感性及特異性而變化。在其他實施例中，甲基化及CNA量測可與其他臨床參數(例如成像結果或血清生物化學)聯合用於預測癌症。

## B. 使用CNA來選擇區域

如上所述，吾等已展示腫瘤組織中具有複本數畸變之區域中的血漿



甲基化密度將改變。在腫瘤組織中具有複本數獲得之區域處，低甲基化腫瘤DNA對血漿之貢獻增加將引起與無複本數畸變之區域相比血漿DNA之低甲基化程度較大。反之，在腫瘤組織中具有複本數損失之區域處，來源於癌症之低甲基化DNA對血漿之貢獻減少將引起血漿DNA之低甲基化程度較小。血漿DNA之甲基化密度與相對呈現之間的此種關係可能用於區分與存在癌症相關之DNA相關之低甲基化結果與血漿DNA低甲基化之其他非癌症病因(例如SLE)。

為說明此方法，吾等分析兩個肝細胞癌(HCC)患者及兩個無癌症SLE患者之血漿樣品。此兩個SLE患者(SLE04及SLE10)顯示血漿中明顯存在低甲基化及CNA。對於患者SLE04，84%區間顯示低甲基化且11.2%區間顯示CNA。對於患者SLE10，10.3%區間顯示低甲基化且5.7%區間顯示CNA。

圖50A及50B展示SLE患者SLE04及SLE10之血漿低甲基化及CNA分析之結果。外部圓展示1 Mb解析度下之甲基化z評分( $Z_{\text{甲基化}}$ )。甲基化 $Z_{\text{甲基化}} < -3$ 之區間呈紅色，且 $Z_{\text{甲基化}} > -3$ 之區間呈灰色。內部圓展示CNA z評分( $Z_{\text{CNA}}$ )。綠色、紅色及灰色點分別表示 $Z_{\text{CNA}} > 3$ 、 $< 3$ 及介於-3至3之間的區間。在此兩個SLE患者中，在血漿中觀測到低甲基化及CNA變化。

為確定甲基化及CNA之變化是否與血漿中存在來源於癌症之DNA一致，吾等比較 $Z_{\text{CNA}} > 3$ 、 $< -3$ 及介於-3至3之間的區域的 $Z_{\text{甲基化}}$ 。對於由血漿中來源於癌症之DNA所貢獻之甲基化變化及CNA，將預期 $Z_{\text{CNA}} < -3$ 之區域具有較小甲基化程度及負值較小之 $Z_{\text{甲基化}}$ 。相比之下，將預期 $Z_{\text{CNA}} > 3$ 之區域具有較大低甲基化程度及負值較大之 $Z_{\text{甲基化}}$ 。出於說明目的，吾等應用

單側秩和檢驗來比較具有CNA之區域(亦即 $Z_{CNA} < -3$ 或 $> 3$ 的區域)與無CNA之區域(亦即 $Z_{CNA}$ 介於 $-3$ 與 $3$ 之間的區域)的 $Z_{\text{甲基化}}$ 。在其他實施例中，可使用其他統計檢驗，例如但不限於史都登氏t檢驗、方差分析(ANOVA)檢驗及克魯斯凱-沃利斯檢驗(Kruskal-Wallis test)。

圖51A及51B展示對兩個HCC患者(TBR34及TBR36)之血漿之有CNA區域及無CNA區域的 $Z_{\text{甲基化}}$ 分析。 $Z_{CNA} < -3$ 及 $> 3$ 之區域分別表示血漿中呈現不足及過度呈現之區域。在TBR34及TBR36兩者中，與血漿中正常呈現之區域(亦即， $Z_{CNA}$ 介於 $-3$ 與 $3$ 之間的區域)相比，血漿中呈現不足之區域(亦即 $Z_{CNA} < -3$ 之區域)具有顯著較高 $Z_{\text{甲基化}}$ ( $P$ 值 $< 10^{-5}$ ，單側秩和檢驗)。正常呈現對應於預期整倍體基因組。對於血漿中過度呈現之區域(亦即 $Z_{CNA} > 3$ 之區域)，與血漿中正常呈現之區域相比，其具有顯著較低 $Z_{\text{甲基化}}$ ( $P$ 值 $< 10^{-5}$ ，單側秩和檢驗)。所有此等變化與血漿樣品中存在低甲基化腫瘤DNA一致。

圖51C及51D展示對兩個SLE患者(SLE04及SLE10)之血漿之有CNA區域及無CNA區域的 $Z_{\text{甲基化}}$ 分析。 $Z_{CNA} < -3$ 及 $> 3$ 之區域分別表示血漿中呈現不足及過度呈現之區域。對於SLE04，與血漿中正常呈現之區域(亦即 $Z_{CNA}$ 介於 $-3$ 與 $3$ 之間的區域)相比，血漿中呈現不足之區域(亦即 $Z_{CNA} < -3$ 之區域)不具有顯著較高之 $Z_{\text{甲基化}}$ ( $P$ 值 $= 0.99$ ，單側秩和檢驗)，且與血漿中正常呈現之區域相比，血漿中過度呈現之區域(亦即 $Z_{CNA} > 3$ 之區域)不具有顯著較低之 $Z_{\text{甲基化}}$ ( $P$ 值 $= 0.68$ ，單側秩和檢驗)。此等結果由於血漿中存在來源於腫瘤之低甲基化DNA而不同於預期變化。同樣，對於SLE10，與 $Z_{CNA}$ 介於 $-3$ 與 $3$ 之間的區域相比， $Z_{CNA} < -3$ 之區域不具有顯著較高之 $Z_{\text{甲基化}}$ ( $P$ 值 $= 0.99$ ，單側秩和檢驗)。

SLE患者中之 $Z_{\text{甲基化}}$ 與 $Z_{\text{CNA}}$ 之間不具有典型癌症相關模式之原因在於在SLE患者中，CNA不存在於亦展現低甲基化之特定細胞類型中。作為替代，觀測到明顯存在CNA及低甲基化係由於SLE患者中之循環DNA之大小分佈有所改變。大小分佈改變可能改變不同基因組區域之定序讀段密度，從而引起明顯CNA，因為參考來源於健康個體。如先前部分中所述，循環DNA片段之大小與其甲基化密度之間存在相關性。因此，大小分佈改變亦可引起異常甲基化。

雖然 $Z_{\text{CNA}} > 3$ 之區域之甲基化程度稍低於 $Z_{\text{CNA}}$ 介於-3與3之間的區域，但該比較之p值遠高於在兩個癌症患者中所觀測到的。在一個實施例中，p值可用作用以確定測試病例患有癌症之機率的參數。在另一實施例中，具有正常及異常呈現之區域之間的 $Z_{\text{甲基化}}$ 差異可用作用於指示存在癌症之機率的參數。在一個實施例中，可使用一組癌症患者來確定 $Z_{\text{甲基化}}$ 與 $Z_{\text{CNA}}$ 之間的相關性並且確定不同參數之臨限值，以便指示該等變化與測試血漿樣品中存在來源於癌症之低甲基化DNA一致。

因此，在一個實施例中，可進行CNA分析以確定均展現以下中之一者的區域的第一集：缺失、擴增或正常呈現。舉例而言，第一區域集可均展現缺失，或均展現擴增，或均展現正常呈現(例如具有正常第一量之區域，諸如正常 $Z_{\text{甲基化}}$ )。可測定此第一區域集之甲基化程度(例如，方法2800之第一甲基化程度可對應於第一區域集)。

CNA分析可測定第二區域集，該等區域均展現以下中之另一者：缺失、擴增或正常呈現。第二區域集與第一集相比將以不同方式展現。舉例而言，若第一區域集正常，則第二區域集可展現缺失或擴增。可基於該第二區域集中之位點處之甲基化DNA分子之各別數目來計算第二甲基化程

度。

隨後可計算第一甲基化程度與第二甲基化之間的參數。舉例而言，可計算差異或比率並且與截止值比較。該差異或比率亦可經歷機率分佈(例如，作為統計檢驗之一部分)以確定獲得該值之機率，且可將此機率與截止值比較以基於甲基化程度確定癌症程度。可選擇該截止值以區分患有癌症之樣品與未患癌症之樣品(例如SLE)。

在一個實施例中，可測定第一區域集或區域混合物(亦即，顯示擴增、缺失及正常之區域的混合物)之甲基化程度。隨後可將此甲基化程度與第一截止值比較作為第一分析階段之一部分。若超過截止值，藉此指示癌症可能性，則可進行以上分析以確定該指示是否為假陽性。癌症程度之最終分類因此可包括比較兩種甲基化程度之參數與第二截止值。

第一甲基化程度可為針對第一區域集之各區域計算的區域甲基化程度之統計值(例如平均值或中值)。第二甲基化程度亦可為針對第二區域集之各區域計算的區域甲基化程度的統計值。作為實例，可使用單側秩和檢驗、史都登氏t檢驗、方差分析(ANOVA)檢驗或克魯斯凱-沃利斯檢驗來確定該等統計值。

## **XI. 癌症類型分類**

除確定生物體是否患有癌症以外，實施例可鑑別與樣品相關之癌症類型。此鑑別癌症類型可使用整體低甲基化、CpG島高甲基化及/或CNA之模式。該等模式可涉及使用所量測之區域甲基化程度、區域之各別CNA值及CpG島之甲基化程度對具有已知診斷之患者進行叢集分析。以下結果顯示具有類似癌症類型之生物體具有類似區域及CpG島值，以及非癌症患者具有類似值。在叢集分析中，區域或島之各值可為叢集分析過程

中之獨立維度。

已知相同癌症類型將共有相似基因及表觀遺傳變化(E Gebhart等人 2004 Cytogenet Genome Res; 104: 352-358 ; PA Jones等人 2007 Cell; 128: 683-692)。以下，吾等描述血漿中所偵測之CNA及甲基化變化模式如何適用於推斷癌症之來源或類型。使用例如階層式叢集分析對來自HCC患者、非HCC患者及健康對照個體之血漿DNA樣品進行分類。該分析係使用例如R指令碼套件中之heatmap.2功能([cran.r-project.org/web/packages/gplots/gplots.pdf](http://cran.r-project.org/web/packages/gplots/gplots.pdf))進行。

為說明此方法之潛力，吾等使用兩組準則(A組及B組)作為實例來鑑別適用於對血漿樣品進行分類之特徵(參見表6)。在其他實施例中，可使用其他準則來鑑別該等特徵。所使用之特徵包括1 Mb解析度下之整體CNA、1 Mb解析度下之整體甲基化密度及CpG島甲基化。

	1 Mb解析度下之整體甲基化	
	A組準則	B組準則
準則	>20個癌症病例具有 >3或<-3之z評分	>20個癌症病例具有 >2.5或<-2.5之z評分
所鑑別之特徵數目	584	1,911

	CNA特徵	
	A組準則	B組準則
準則	>10個癌症病例具有 >3或<-3之z評分	>10個癌症病例具有 >2.5或<-2.5之z評分
所鑑別之特徵數目	355	759

	CpG島甲基化	
	A組準則	B組準則
準則	>5個癌症病例在特 定CpG島處之甲基化 密度與參考之平均 值相差2%	>1個癌症病例在特 定CpG島處之甲基 化密度與參考之平 均值相差2%
所鑑別之特徵數目	110	191

表6

在前兩個實例中，吾等使用所有CNA、1 Mb解析度下之整體甲基化及CpG島甲基化特徵進行分類。在其他實施例中，可使用其他準則，例如但不限於量測參考組之血漿中之特徵的精確度。

圖52A展示使用所有1,130個A組特徵(包括355個CNA、1 Mb解析度下之584個整體甲基化特徵及110個CpG島之甲基化狀態)對來自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品進行階層式叢集分析。上側色彩條表示樣品群組：綠色、藍色及紅色分別表示健康個體、HCC及非HCC癌症患者。一般而言，三組個體趨向於叢集在一起。縱軸表示分類特徵。將不同個體中具有類似模式之特徵叢集在一起。此等結果表明CpG島甲基化變化、1 Mb解析度下之全基因組甲基化變化及血漿中之CNA之模式可能用於確定具有未知原發性之患者中的癌症來源。

圖52B展示使用所有2,780個B組特徵(包括759個CNA、1 Mb解析度下之1,911個整體甲基化及191個CpG島之甲基化狀態)對來自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品進行階層式叢集分析。上側色彩條表示樣品群組：綠色、藍色及紅色分別表示健康個體、HCC及非HCC癌症患者。一般而言，三組個體趨向於叢集在一起。縱軸表示分類特徵。將不同個體中具有類似模式之特徵叢集在一起。此等結果表明不同組CpG島甲基化變化、1 Mb解析度下之全基因組甲基化變化及血漿中之CNA之模式可用於確定具有未知原發性之患者中的癌症來源。分類特徵之選擇可針對特定應用加以調節。另外，可根據個體對不同癌症類型之先前機率對癌症類型預測加權。舉例而言，患有慢性病毒性肝炎之患者易於發生肝細胞癌且長期吸菸者易於發生肺癌。因此，可使用例如但不限於邏輯、多重或叢集回歸來計算癌症類型之加權機率。

在其他實施例中，單一類型特徵可用於分類分析。舉例而言，在以下實例中，僅1 Mb解析度下之整體甲基化、CpG島高甲基化或1 Mb解析度下之CNA用於階層式叢集分析。當使用不同特徵時，差異力可能不同。進一步改進分類特徵可能改良分類準確性。

圖53A展示使用A組CpG島甲基化特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。一般而言，將癌症患者叢集在一起且非癌症個體處於另一叢集中。然而，與使用所有三種類型特徵相比較，HCC及非HCC患者相隔較小。

圖53B展示使用A組1 Mb解析度下之整體甲基化密度作為分類特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。觀察到HCC及非HCC患者優先叢集。

圖54A展示使用A組1 Mb解析度下之整體CNA作為分類特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。可見HCC及非HCC患者優先叢集。

圖54B展示使用B組CpG島甲基化密度作為分類特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。可觀察到HCC及非HCC癌症患者優先叢集。

圖55A展示使用B組1 Mb解析度下之整體甲基化密度作為分類特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。可觀察到HCC及非HCC癌症患者優先叢集。

圖55B展示使用B組1 Mb解析度下之整體CNA作為分類特徵對得自HCC患者、非HCC癌症患者及健康對照個體之血漿樣品之階層式叢集分析。可觀察到HCC及非HCC癌症患者優先叢集。

血漿樣品之此等階層式叢集結果表明不同特徵之組合可能用於鑑別原發性癌症類型。進一步改進選擇準則可能進一步改良分類準確性。

因此，在一個實施例中，當甲基化分類指示生物體存在癌症時，可藉由比較甲基化程度(例如，來自方法2800之第一甲基化或任何區域甲基化程度)與自其他生物體(亦即相同類型之其他生物體，諸如人類)測定之相應值來鑑別與該生物體相關之癌症類型。相應值可針對計算甲基化程度之相同區域或位點集。其他生物體中之至少兩個鑑別為患有不同類型之癌症。舉例而言，相應值可組織成叢集，其中兩個叢集與不同的癌症相關。

此外，當CNA及甲基化一起用於獲得癌症程度之第三分類時，可將CNA及甲基化特徵與來自其他生物體之相應值比較。舉例而言，可將展現缺失或擴增之區域之第一量(例如，來自圖36)與由其他生物體測定之相應值比較以鑑別與生物體相關之癌症類型。

在一些實施例中，甲基化特徵為基因組之複數個區域之區域甲基化程度。可使用據測定具有超過各別區域截止值之區域甲基化程度的區域，例如，可將生物體之區域甲基化程度與其他生物體之基因組之相同區域之區域甲基化程度比較。該比較可允許區分癌症類型，或僅提供額外過濾以確定癌症(例如鑑別假陽性)。因此，可基於該比較確定生物體是患有第一類型癌症、不存在癌症或是第二類型癌症。

可使用區域甲基化程度對其他生物體(以及測試生物體)進行叢集分析。因此，比較區域甲基化程度可用於確定生物體屬於何種叢集。叢集分析亦可使用被確定為展現缺失或擴增之區域之CNA標準化值，如上文所述。且叢集分析可使用高甲基化CpG島之各別甲基化密度。

為說明此方法之原理，吾等展示對兩個未知樣品之分類使用邏輯回



歸之實例。此分類之目的在於確定此兩個樣品是為HCC或是為非HCC癌症。彙集包括自HCC患者收集之23個血漿樣品及來自罹患非HCC之癌症之患者的18個樣品的訓練樣品組。因此，訓練組中存在總計41個病例。在此實例中，選擇13個特徵，包括五個關於CpG島甲基化之特徵(X1至X5)、六個關於1 Mb區域甲基化之特徵(X6至X11)及2個關於1 Mb區域CNA之特徵(X12至X13)。基於訓練組中至少15個病例具有 $>3$ 或 $<-3$ 之z評分的準則來選擇CpG甲基化特徵。基於訓練組中至少39個病例具有 $>3$ 或 $<-3$ 之z評分的準則來選擇1 Mb甲基化特徵。基於至少20個病例具有 $>3$ 或 $<-3$ 之z評分的準則來選擇CNA特徵。對此訓練組之樣品進行邏輯回歸以測定各特徵(X1至X13)之回歸係數。回歸係數具有較大量值(不論其是正或負意義)之特徵可較佳辨別HCC與非HCC樣品。對於各別特徵，各病例之z評分用作獨立變量之輸入值。隨後針對13個特徵分析兩個血漿樣品，一個來自HCC患者(TBR36)，且一個來自罹患肺癌之患者(TBR177)。

在此癌症類型分類分析中，假定此兩個樣品收集自罹患未知來源之癌症的患者。對於各樣品，將各別特徵之z評分輸入邏輯回歸方程以確定比值比之自然對數( $\ln(\text{比值比})$ )，其中該比值比表示患有HCC之機率與未患有HCC之機率的比率(HCC/非HCC)。

表7展示邏輯回歸方程之13個特徵之回歸係數。亦展示兩個測試病例(TBR36及TBR177)之各別特徵之z評分。TBR36及TBR177之HCC之 $\ln(\text{比值比})$ 分別為37.03及-4.37。由此等比值比，自HCC患者收集之血漿樣品之機率計算分別為 $>99.9\%$ 及 $1\%$ 。簡短言之，TBR36為來自HCC患者之樣品的可能性較高，而TBR177為來自HCC患者之樣品的可能性較低。

特徵	回歸係數	各別特徵之z評分	
		TBR36	TBR177
X1	-2.9575	14.8	0
X2	2.2534	21.3	0
X3	-1.5099	6.1	0
X4	-0.236	34.0	0
X5	0.7426	17.3	0
X6	-0.6682	-26.3	-1.5
X7	-0.2828	-13.9	-2.6
X8	-0.7281	-9.4	-4.4
X9	1.0581	-7.8	-3.7
X10	0.3877	-20.8	-4.3
X11	0.3534	-15.5	-3.1
X12	-1.1826	4.8	3.3
X13	-0.3805	-11.7	-1.4
	ln(比值比)	37.03	-4.37463

表7

在其他實施例中，階層式叢集回歸、分類樹分析及其他回歸模型可用於確定癌症之可能原發性來源。

## XII. 材料及方法

### A. 製備經亞硫酸氫鹽處理之DNA文庫並定序

藉由Covaris S220系統(Covaris)將添加有0.5% (w/w)未甲基化的 $\lambda$  DNA (Promega)之基因組DNA (5  $\mu$ g)片段化至長度為約200 bp。使用配對末端定序樣品製備套組(Illumina)根據製造商之說明書製備DNA文庫，但改將甲基化接附子(Illumina)接合至DNA片段。在使用AMPure XP磁性珠粒(Beckman Coulter)進行兩輪純化之後，接合產物分成2個部分，用EpiTect亞硫酸氫鹽套組(Qiagen)對其中一個進行2輪亞硫酸氫鹽修飾。插入物中之CpG位點處之未甲基化胞嘧啶轉化成尿嘧啶，而甲基化胞嘧啶保持不變。使用以下配方藉由10個PCR循環增濃經接附子接合之DNA分子(經亞硫酸氫鈉處理或未經亞硫酸氫鈉處理)：含2.5 U PfuTurboCx hotstart DNA聚合酶(Agilent Technologies)、1×PfuTurboCx反應緩衝

液、25  $\mu$ M dNTPs、1  $\mu$ l PCR引子PE 1.0及1  $\mu$ l PCR引子PE 2.0 (Illumina)之50  $\mu$ l反應物。熱循環概況為：95°C/2分鐘，98°C/30秒，隨後10循環之98°C/15秒，60°C/30秒，及72°C/4分鐘，最終步驟為72°C/10分鐘(R Lister等人2009 Nature; 462: 315-322)。使用AMPure XP磁性珠粒純化PCR產物。

向自3.2至4 ml母體血漿樣品中提取之血漿DNA中摻入片段化 $\lambda$  DNA (每毫升血漿25 pg)且如上所述進行文庫建構(RWK Chiu等人2011 BMJ; 342: c7401)。在接合於甲基化接附子之後，將接合產物分成2半，且對一部分進行2輪亞硫酸氫鹽修飾。隨後藉由如上所述進行10個PCR循環來增濃經亞硫酸氫鹽處理或未經處理之接合產物。

在HiSeq2000儀器(Illumina)上用配對末端格式針對75 bp對經亞硫酸氫鹽處理或未經處理之DNA文庫進行定序。在cBot儀器(Illumina)上用配對末端叢集產生套組v3產生DNA叢集。使用HiSeq控制軟體(HCS) v1.4及即時分析(RTA)軟體v1.13 (Illumina)進行即時影像分析及鹼基檢出，其中自動化基質及相計算係基於用DNA文庫定序之摻入PhiX對照v3。

#### B. 序列比對及鑑別甲基化胞嘧啶

在鹼基檢出之後，移除片段末端上之接附序列及低品質鹼基(亦即品質評分<20)。隨後藉由稱為Methy-Pipe之甲基化資料分析渠道處理呈FASTQ格式之經修切讀段(P Jiang等人在IEEE生物資訊學與生物醫學國際研討會(香港, 2010年12月18至21日)上提出之論文Methy-Pipe: An integrated bioinformatics data analysis pipeline for whole genome methylome analysis)。為比對經亞硫酸氫鹽轉化之定序讀段，吾等首先使用參考人類基因組(NCBI版本36/hg18)將沃森股及克立克股上之所有胞嘧

啶殘基獨立地電子雜交轉化成胸嘧啶。吾等隨後將所有經處理讀段中之各胞嘧啶電子雜交轉化成胸嘧啶，且保存各經轉化殘基之位置資訊。使用 SOAP2 (R Li等人2009 *Bioinformatics*; 25: 1966-1967)來比對經轉化之讀段與兩個轉化前參考人類基因組，其中各比對讀段最多允許兩個錯配。僅選擇可定位於獨特基因組位置之讀段。移除定位於沃森股及克立克股兩者之含糊讀段及具有相同起始及結束基因組位置之複製(純系)讀段。保留插入物大小 $\leq 600$  bp之定序讀段用於甲基化及大小分析。

CpG二核苷酸情形下之胞嘧啶殘基為下游DNA甲基化研究之主要標靶。在比對後，基於電子雜交轉化期間保存之位置資訊恢復最初存在於定序讀段上之胞嘧啶。將CpG二核苷酸中之經恢復胞嘧啶評分為甲基化。將CpG二核苷酸中之胸嘧啶評分為未甲基化。所包括之未甲基化的 $\lambda$  DNA在文庫製備期間充當評估亞硫酸氫鈉修飾效率之內部對照。若亞硫酸氫鹽轉化效率為100%，則 $\lambda$  DNA上之所有胞嘧啶均應已轉化成胸嘧啶。

### **XIII. 概述**

利用本文所述之實施例，可使用例如個體之血漿來非侵入性地篩選、偵測、監測或預測癌症。亦可藉由自母體血漿推斷胎兒DNA之甲基化概況來對胎兒進行產前篩選、診斷、研究或監測。為說明該方法之效率，吾等展示可自母體血漿直接評定習知經由胎盤組織研究獲得之資訊。舉例而言，基因座之印記狀態、鑑別在胎兒與母體DNA之間具有差異性甲基化之基因座及基因座之甲基化概況之妊娠期變化係藉由直接分析母體血漿DNA來達成。吾等之方法的主要優勢在於可在妊娠期間全面評定胎兒甲基化模式組而不破壞妊娠或不需要對胎兒組織進行侵入性取樣。鑒於DNA甲基化狀態改變與許多妊娠相關病狀之間的已知關聯性，本研究中

所述之方法可充當研究該等病狀之病理生理學及鑑別該等病狀之生物標記的重要工具。藉由聚焦於印記基因座，吾等證明可由母體血漿評定父體傳遞以及母體傳遞之胎兒甲基化概況。此方法可能適用於研究印記疾病。實施例亦可直接應用於胎兒或妊娠相關疾病之產前評估。

吾等已證明全基因組亞硫酸氫鹽定序可應用於研究胎盤組織之DNA甲基化概況。人類基因組中存在約28M CpG位點(C Clark等人2012 PLoS One; 7: e50233)。吾等之CVS及足月胎盤組織樣品之亞硫酸氫鹽定序資料覆蓋超過80% CpG。此代表實質上比使用其他高產量平台可達成之覆蓋率更廣泛之覆蓋率。舉例而言，用於先前胎盤組織研究之Illumina Infinium人類甲基化27K微珠晶片陣列(T Chu等人2011 PLoS One; 6: e14723)僅覆蓋基因組中之0.1% CpG。近年來可利用之Illumina Infinium人類甲基化450K微珠晶片陣列僅覆蓋1.7% CpG (C Clark等人2012 PLoS One; 7: e50233)。因為MPS方法不存在與探針設計、雜交效率或抗體捕捉強度相關之限制，故可評定CpG島內外及大部分序列情形下之CpG。

#### **XIV. 電腦系統**

本文中所提及之任何電腦系統皆可利用任何適合數目之子系統。該等子系統之實例示於圖33中之電腦設備3300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，電腦系統可包括多個具有內部組件之電腦設備，各電腦設備為子系統。

圖33中所示之子系統經由系統匯流排3375互連。顯示其他子系統，諸如印表機3374、鍵盤3378、儲存裝置3379、耦合於顯示器配接器3382之監視器3376及其他。耦合於I/O控制器3371之周邊裝置及輸入/輸出

(I/O)裝置可藉由此項技術中已知的許多手段連接於電腦系統，諸如串聯埠 3377。舉例而言，串聯埠 3377 或外部介面 3381 (例如乙太網路 (Ethernet)、Wi-Fi等)可用於將電腦系統3300連接於廣域網路，諸如網際網路、滑鼠輸入設備或掃描器。經由系統匯流排3375互連允許中央處理器3373與各子系統通信並且控制來自系統記憶體3372或儲存裝置3379 (例如固接磁碟)之指令的執行以及子系統之間的資訊交換。系統記憶體3372及/或儲存裝置3379可具體化為電腦可讀媒體。本文中所提及之任何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由外部介面3381或由內部界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、子系統或設備可經網路進行通信。在該等情況下，一個電腦可被視為用戶端且另一電腦可被視為伺服器，其中各自可為同一電腦系統之一部分。用戶端及伺服器可各包括多個系統、子系統或組件。

應瞭解本發明實施例中之任一者可使用硬體(例如特殊應用積體電路或場可程式化閘陣列)以控制邏輯形式實施及/或使用具有電腦軟體利用一般可程式化處理器以模組或積體方式實施。如本文中所使用，處理器包括處於同一積體晶片上之多核心處理器或處於單一電路板上或形成網路之多個處理單元。基於本發明及本文中所提供之教示，一般熟習此項技術者將已知且瞭解其他使用硬體及硬體與軟體之組合實施本發明之實施例的方式及/或方法。

使用任何適合電腦語言(諸如Java、C++或使用例如習知或物件導向技術之Perl)，本申請案中所述之任何軟體組件或函數皆可作為欲由處理器執行之軟體代碼來實施。軟體代碼可在用於儲存及/或傳輸之電腦可讀

媒體上儲存為一系列指令或命令，適合媒體包括隨機存取記憶體(RAM)、唯讀記憶體(ROM)、諸如硬碟機或軟碟之磁性媒體或諸如壓縮光碟(CD)或DVD(數位化通用光碟)之光學媒體、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經由符合包括網際網路在內之多種協定的有線、光學及/或無線網路傳輸之載波信號來編碼及傳輸。因此，根據本發明之一實施例的電腦可讀媒體可使用以該等程式編碼之資料信號產生。以程式代碼編碼之電腦可讀媒體可與相容裝置一起封裝或由其他裝置(例如經由網際網路下載)單獨提供。任何該電腦可讀媒體皆可駐留於單一電腦程式產品(例如硬碟機、CD或整個電腦系統)之上或之內，且可存在於系統或網路內之不同電腦程式產品之上或之內。電腦系統可包括監測器、印表機或向使用者提供本文中所提及之任何結果的其他適合顯示器。

本文所述之方法中之任一者均可完全或部分用包括一或多個處理器之電腦系統執行，該一或多個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述之任何方法之步驟的電腦系統，可能用不同組件執行各別步驟或各別步驟組。儘管本文方法之步驟呈編號步驟形式呈現，但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。亦可視情況選用步驟之全部或部分。另外，任何方法之任何步驟皆可用執行此等步驟之模組、電路或其他構件來執行。

在不背離本發明實施例之精神及範疇的情況下，特定實施例之特定細節可以任何適合方式組合。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例示性實施例之以上描述。該描述不欲為詳盡的或將本發明限制於所描述之精確形式，且根據以上教示，許多修改及變化為可能的。選擇並描述該等實施例，以便最佳地說明本發明之原理及其實際應用，以藉此使其他熟習此項技術者能夠在各種實施例中且在進行適合於預期特定用途之各種修改的情況下最佳地利用本發明。

除非相反地特別指示，否則「一(a/an)」或「該(the)」之敘述欲意謂「一或多」。

本文所提及之所有專利、專利申請案、公開案及描述係處於所有目的而以全文引用的方式併入。不承認任一者為先前技術。

#### 【符號說明】

3371	I/O控制器
3372	系統記憶體
3373	中央處理器
3374	印表機
3376	監視器
3377	串聯埠
3378	鍵盤
3379	儲存裝置
3381	外部界面
3382	顯示器配接器



表S2A. 自前三個月絨毛膜絨毛樣品及母體血細胞中鑑別之100個最高甲基化區域之清單

染色體	起始	結束	大小(bp)	母體 血細胞	CVS	P值	甲基化 差異
chr13	113063600	113064100	500	0.009	0.9	3.67E-15	0.891
chr6	36279700	36280200	500	0.0068	0.8957	2.39E-22	0.8889
chr16	66876000	66876500	500	0.0327	0.9211	3.82E-21	0.8884
chr10	163500	164000	500	0.0195	0.9034	3.60E-35	0.8839
chr9	3518300	3518800	500	0.0263	0.9045	1.32E-26	0.8782
chr12	31877100	31877600	500	0.007	0.8784	3.08E-22	0.8714
chr22	37477400	37478400	1000	0.0152	0.8848	0.00E+00	0.8696
chr4	148940500	148941000	500	0.0055	0.8717	4.40E-29	0.8662
chr5	131836300	131836800	500	0.075	0.9403	1.54E-10	0.8653
chr17	26661700	26663600	1900	0.0187	0.875	2.95E-38	0.8563
chr2	105758600	105759600	1000	0.0305	0.8828	1.19E-53	0.8523
chr22	39188800	39189800	1000	0	0.8514	2.05E-46	0.8514
chr3	153443900	153444900	1000	0.0436	0.8945	5.43E-34	0.8509
chr6	25149600	25150600	1000	0.0135	0.8632	0.00E+00	0.8497
chr5	98296800	98297300	500	0.0432	0.8925	4.97E-23	0.8493
chr7	150679900	150680400	500	0.0496	0.8944	6.50E-17	0.8448
chr7	107563100	107563600	500	0.0495	0.8895	9.58E-26	0.84
chr7	37348300	37349300	1000	0.0012	0.8409	0.00E+00	0.8397
chr14	58837800	58838300	500	0.0097	0.848	3.35E-16	0.8383
chr6	119238100	119238600	500	0.0899	0.928	2.38E-19	0.8381
chr15	93669900	93670400	500	0.0753	0.913	2.19E-10	0.8377
chr17	26669200	26670200	1000	0.0221	0.859	1.44E-29	0.8369
chr2	88108100	88108600	500	0.075	0.9109	3.55E-17	0.8359
chr13	98363800	98364300	500	0.11	0.9457	1.28E-11	0.8357
chr16	66948000	66948500	500	0.0331	0.8685	0.00E+00	0.8354
chr6	42098000	42098500	500	0.0484	0.8835	3.73E-16	0.8351
chr3	129876000	129876500	500	0.0565	0.8897	8.81E-17	0.8332
chr3	142700300	142700800	500	0.0063	0.8393	2.59E-22	0.833
chr8	145883800	145884300	500	0.0392	0.872	0.00E+00	0.8328
chr10	8320700	8321200	500	0.0566	0.8871	9.40E-09	0.8305
chr3	120438100	120438600	500	0.102	0.9292	7.09E-16	0.8272
chr3	173792600	173793100	500	0.0182	0.8453	2.84E-39	0.8271
chr17	40320700	40321200	500	0.0539	0.8788	6.50E-30	0.8249
chr15	72076200	72076700	500	0.0299	0.8525	6.48E-10	0.8226
chr16	29663900	29665400	1500	0.0081	0.8305	0.00E+00	0.8224
chr11	66961100	66962100	1000	0.0489	0.8712	0.00E+00	0.8223
chr9	27083100	27084100	1000	0.097	0.9177	2.37E-55	0.8207
chr9	111249600	111250100	500	0.0613	0.8795	1.99E-20	0.8182
chr14	101412400	101412900	500	0	0.8167	8.26E-32	0.8167
chr1	242549200	242549700	500	0	0.8155	3.50E-21	0.8155
chr8	38642800	38643300	500	0.0191	0.8346	3.22E-41	0.8155
chr4	85893600	85894100	500	0.0394	0.8533	1.45E-15	0.8139
chr5	142368600	142369100	500	0.0385	0.8523	1.18E-18	0.8138
chr8	130969500	130970000	500	0.069	0.8824	2.42E-24	0.8134
chr2	196783900	196784400	500	0	0.8123	3.63E-40	0.8123
chr16	49258100	49258600	500	0.0733	0.8851	4.29E-18	0.8118

chr1	232601200	232601700	500	0.0594	0.8707	1.73E-13	0.8113
chr1	109039500	109040000	500	0.0366	0.8471	1.07E-11	0.8105
chr17	59491300	59491800	500	0.0662	0.8758	2.15E-17	0.8096
chr21	42194100	42194600	500	0.11	0.9182	1.61E-12	0.8082
chr9	116174500	116175500	1000	0.0062	0.8132	1.98E-60	0.807
chr15	73429200	73429700	500	0	0.8066	9.81E-33	0.8066
chr6	157462800	157463300	500	0.0758	0.8819	7.94E-16	0.8061
chr3	16858500	16859500	1000	0.0021	0.8068	4.76E-68	0.8047
chr9	96662800	96663300	500	0.0614	0.8651	6.79E-28	0.8037
chr9	88143000	88143500	500	0.1538	0.9559	7.43E-09	0.8021
chr19	16090000	16091000	1000	0.0899	0.8904	1.60E-53	0.8005
chr15	29436300	29437300	1000	0.0553	0.8556	1.18E-80	0.8003
chr11	77816100	77816600	500	0.1069	0.9068	2.31E-17	0.7999
chr10	30346800	30347300	500	0.1212	0.9211	7.48E-07	0.7999
chr1	89510300	89511300	1000	0.0203	0.8191	3.53E-77	0.7988
chr3	125986100	125986600	500	0.1686	0.9674	5.24E-22	0.7988
chr19	60162800	60163300	500	0.0127	0.8113	9.99E-19	0.7986
chr16	73655900	73656900	1000	0.082	0.8806	4.48E-41	0.7986
chr16	30104300	30105800	1500	0.0298	0.8282	0.00E+00	0.7984
chr10	118642400	118642900	500	0.0588	0.8571	8.63E-11	0.7983
chr16	4495000	4496000	1000	0.0632	0.8615	2.27E-44	0.7983
chr1	2048300	2048800	500	0.0309	0.8289	1.19E-80	0.798
chr2	136481800	136482800	1000	0.0554	0.8533	8.50E-48	0.7979
chr10	29959200	29959700	500	0.1429	0.94	2.60E-08	0.7971
chr6	139642400	139642900	500	0.0618	0.8585	2.16E-29	0.7967
chr14	69825300	69825800	500	0.0654	0.8615	6.85E-14	0.7961
chr8	49739700	49740200	500	0.0324	0.828	2.88E-30	0.7956
chr17	42205700	42206200	500	0.057	0.852	2.11E-30	0.795
chr4	77445300	77445800	500	0.0442	0.8377	1.79E-35	0.7935
chr17	53762700	53766300	3600	0.0003	0.7926	0.00E+00	0.7923
chr17	44269900	44270400	500	0.026	0.8182	3.49E-21	0.7922
chr6	42462700	42463200	500	0.0761	0.8678	4.74E-22	0.7917
chr2	23396200	23396700	500	0.0333	0.8235	1.25E-14	0.7902
chr9	100921100	100921600	500	0.0244	0.814	3.32E-21	0.7896
chr7	74016100	74016600	500	0.1442	0.9333	6.74E-10	0.7891
chr6	157879000	157879500	500	0.133	0.9219	6.36E-17	0.7889
chr3	3189400	3190400	1000	0.0693	0.8571	1.38E-24	0.7878
chr16	29581500	29584500	3000	0.0081	0.7956	0.00E+00	0.7875
chr17	42201800	42202800	1000	0.0884	0.8751	0.00E+00	0.7867
chr11	94257000	94257500	500	0.1122	0.8986	4.29E-10	0.7864
chr10	14741600	14742100	500	0.0139	0.8	1.73E-20	0.7861
chr21	33826900	33827400	500	0.0879	0.8739	2.81E-11	0.786
chr4	130057200	130057700	500	0.0893	0.875	1.76E-13	0.7857
chr21	35343400	35343900	500	0	0.7853	7.43E-18	0.7853
chr12	105372800	105373300	500	0.0923	0.8767	8.67E-22	0.7844
chr5	10799800	10800300	500	0.1429	0.9263	8.21E-17	0.7834
chr5	16753100	16753600	500	0.041	0.8241	1.40E-15	0.7831
chr3	135746000	135746500	500	0.1429	0.9259	2.86E-09	0.783
chr6	53708300	53708800	500	0.0412	0.8235	2.74E-31	0.7823
chr2	128122900	128123400	500	0.0634	0.8455	4.82E-21	0.7821
chr5	150574200	150574700	500	0.0876	0.8696	1.56E-20	0.782

chr16	84326000	84327000	1000	0.1071	0.8891	3.58E-61	0.782
chr1	26744500	26745500	1000	0.0336	0.8152	0.00E+00	0.7816
chr2	234882000	234882500	500	0.0392	0.819	7.63E-14	0.7798

表S2B. 自前三個月絨毛膜絨毛樣品及母體血細胞中鑑別之100個最低甲基化區域之清單

染色體	起始	結束	大小(bp)	母體 血細胞	CVS	P值	甲基化 差異
chr18	12217500	12218500	1000	0.9873	0	3.05E-25	0.9873
chr17	22885400	22885900	500	0.9714	0.0161	8.92E-12	0.9553
chr3	184827100	184827600	500	0.9875	0.033	4.79E-16	0.9545
chr5	148968300	148968800	500	0.98	0.0426	6.70E-09	0.9374
chr10	104794500	104795000	500	0.973	0.0385	9.33E-10	0.9345
chr4	84977900	84978400	500	0.9643	0.0417	2.98E-08	0.9226
chr3	180395300	180395800	500	0.9877	0.0667	6.72E-08	0.921
chr2	138908300	138908800	500	0.939	0.0208	1.10E-16	0.9182
chr6	139873100	139873600	500	0.9667	0.0526	1.29E-07	0.914
chr8	59604700	59605200	500	0.9468	0.033	2.88E-14	0.9138
chr6	167622300	167622800	500	0.9452	0.0316	3.86E-14	0.9136
chr3	175701300	175701800	500	0.9846	0.0735	7.43E-10	0.9111
chr13	59246400	59246900	500	0.9402	0.0313	2.31E-11	0.9089
chr12	71263600	71264100	500	0.9296	0.0213	1.08E-08	0.9083
chr5	39459400	39459900	500	0.9219	0.014	5.01E-22	0.9079
chr17	24904700	24905200	500	0.9161	0.0092	5.04E-35	0.9069
chr12	31889900	31890400	500	0.9524	0.0465	6.78E-13	0.9059
chr3	152897800	152898300	500	0.9402	0.0345	1.70E-17	0.9057
chr1	40378700	40379200	500	0.9565	0.0526	3.31E-09	0.9039
chr12	43979300	43979800	500	0.952	0.05	6.68E-13	0.902
chr18	1395900	1397400	1500	0.9308	0.0293	0.00E+00	0.9015
chr1	223482900	223483400	500	0.9579	0.0575	3.36E-24	0.9004
chr9	130357000	130357500	500	0.9282	0.0286	9.19E-13	0.8996
chr3	72878300	72878800	500	0.9612	0.0625	8.20E-14	0.8987
chr7	84347200	84348700	1500	0.9401	0.0418	0.00E+00	0.8983
chr15	37317500	37318000	500	0.9358	0.0385	7.58E-14	0.8973
chr8	42528600	42529100	500	0.9302	0.0337	1.73E-14	0.8965
chr6	134914000	134914500	500	0.9037	0.0076	4.84E-21	0.8961
chr13	56207100	56208100	1000	0.9184	0.0245	0.00E+00	0.894
chr2	209074000	209074500	500	0.9309	0.037	6.13E-27	0.8938
chr12	74021100	74022100	1000	0.9513	0.058	0.00E+00	0.8933
chr4	118939300	118939800	500	0.9192	0.0276	6.58E-27	0.8916
chr5	12626600	12628600	2000	0.9266	0.0355	0.00E+00	0.8911
chr5	105517300	105518300	1000	0.927	0.0359	0.00E+00	0.891
chr12	70056300	70057300	1000	0.9488	0.0609	0.00E+00	0.888
chr6	153238200	153239200	1000	0.9123	0.0244	0.00E+00	0.8879
chr17	60374800	60375300	500	0.9655	0.0777	3.64E-14	0.8878
chr14	68272700	68273200	500	0.9389	0.0523	1.23E-22	0.8866
chr19	54533800	54534800	1000	0.9117	0.0262	0.00E+00	0.8855
chr12	15392200	15393200	1000	0.9307	0.0457	0.00E+00	0.885
chr1	212517400	212517900	500	0.9266	0.0417	9.81E-12	0.8849
chr10	49344400	49345400	1000	0.9422	0.0579	0.00E+00	0.8844
chr3	47410400	47410900	500	0.9213	0.0381	5.59E-16	0.8832
chr3	879500	880000	500	0.9455	0.0625	8.06E-06	0.883
chr2	31572400	31573400	1000	0.9176	0.0357	0.00E+00	0.8819

chr1	89131200	89131700	500	0.9314	0.0498	5.15E-70	0.8816
chr8	94832000	94832500	500	0.9156	0.0351	2.16E-65	0.8805
chr7	14008300	14009800	1500	0.9349	0.0545	0.00E+00	0.8804
chr12	12971300	12972300	1000	0.9361	0.0559	0.00E+00	0.8802
chr5	43114700	43115200	500	0.9638	0.0842	1.79E-13	0.8796
chr11	107872400	107872900	500	0.9472	0.0677	2.31E-32	0.8794
chr8	49757600	49758100	500	0.9048	0.0269	3.15E-52	0.8779
chr13	33106400	33106900	500	0.9384	0.0606	9.54E-15	0.8778
chr3	190658800	190659300	500	0.9388	0.0617	2.71E-22	0.8777
chr1	181508000	181508500	500	0.9259	0.0495	3.78E-15	0.8764
chr1	180436900	180437400	500	0.9412	0.0652	2.36E-13	0.876
chr6	122642800	122643800	1000	0.9218	0.0458	0.00E+00	0.8759
chr5	166429300	166429800	500	0.9551	0.08	5.26E-05	0.8751
chr12	14972900	14973400	500	0.9483	0.0733	2.10E-18	0.8749
chr5	123933900	123934400	500	0.943	0.0683	1.12E-39	0.8746
chr2	15969400	15970400	1000	0.8939	0.0196	9.43E-46	0.8743
chr3	167635200	167636200	1000	0.9363	0.0625	9.03E-41	0.8738
chr5	159442700	159443200	500	0.9174	0.044	6.27E-14	0.8734
chr4	48027200	48027700	500	0.9839	0.1111	8.89E-06	0.8728
chr6	140071500	140072000	500	0.9234	0.0506	4.39E-33	0.8728
chr10	22356300	22356800	500	0.9548	0.0822	1.04E-18	0.8726
chr8	61007300	61007800	500	0.9197	0.0476	1.24E-15	0.8721
chr11	95463500	95464000	500	0.9348	0.0629	1.16E-20	0.8718
chr2	216399800	216400300	500	0.938	0.0667	5.98E-06	0.8713
chr18	57359700	57360200	500	0.9293	0.0584	1.77E-19	0.871
chr3	102734400	102734900	500	0.8917	0.0207	7.94E-22	0.871
chr1	173605700	173606200	500	0.96	0.0891	5.46E-13	0.8709
chr2	86993700	86995700	2000	0.8965	0.0261	0.00E+00	0.8704
chr3	162621100	162621600	500	0.9226	0.0526	7.89E-38	0.8699
chr12	10144800	10145300	500	0.929	0.0598	3.45E-17	0.8691
chr3	113855100	113855600	500	0.9667	0.0982	3.97E-14	0.8685
chr2	156958200	156959200	1000	0.9252	0.0571	8.89E-50	0.8681
chr2	55775000	55776000	1000	0.9159	0.0483	0.00E+00	0.8676
chr6	124898400	124898900	500	0.8987	0.0313	1.91E-15	0.8675
chr5	42003700	42004700	1000	0.9262	0.0588	0.00E+00	0.8674
chr3	24162200	24162700	500	0.883	0.0161	1.75E-27	0.8668
chr6	35394000	35395000	1000	0.9204	0.0539	0.00E+00	0.8665
chr17	8451800	8453300	1500	0.9376	0.0714	0.00E+00	0.8662
chr14	53487700	53488700	1000	0.9013	0.0353	0.00E+00	0.866
chr7	98572800	98573300	500	0.9651	0.0995	8.37E-26	0.8656
chr6	52298700	52299200	500	0.9427	0.0772	1.31E-28	0.8655
chr6	159047900	159048400	500	0.908	0.0426	3.34E-08	0.8655
chr14	22152600	22153100	500	0.9085	0.0435	4.43E-17	0.865
chr12	103285000	103285500	500	0.9321	0.0674	0.00E+00	0.8647
chr7	43302200	43302700	500	0.968	0.1037	6.40E-16	0.8643
chr14	22247400	22247900	500	0.9804	0.1163	3.83E-06	0.8641
chr2	66780900	66781400	500	0.9355	0.0714	8.37E-09	0.8641
chr12	97393000	97393500	500	0.9045	0.0408	3.46E-21	0.8637
chr5	162797900	162798900	1000	0.9271	0.0635	1.75E-57	0.8636
chr2	83598400	83599400	1000	0.9354	0.0719	0.00E+00	0.8635
chr11	111358800	111359300	500	0.9156	0.0523	4.15E-24	0.8632

chr11	104891100	104892600	1500	0.9164	0.0533	0.00E+00	0.863
chr1	184583600	184584100	500	0.9647	0.1026	3.02E-13	0.8621
chr5	132350500	132351500	1000	0.9042	0.0426	1.86E-36	0.8616
chr5	53268300	53268800	500	0.972	0.1111	4.76E-16	0.8609

表S2C. 自後三個月胎盤組織及母體血細胞中鑑別之100個最高甲基化區域之清單

染色體	起始	結束	大小(bp)	母體 血細胞	妊娠期 胎盤	P值	甲基化 差異
chr4	78129700	78130200	500	0.0488	0.9747	3.97E-33	0.926
chr5	131467400	131467900	500	0.0213	0.9275	7.10E-27	0.9063
chr17	26661700	26663600	1900	0.0187	0.9226	1.79E-41	0.9039
chr4	148940500	148941000	500	0.0055	0.9079	1.82E-29	0.9024
chr9	100921100	100921600	500	0.0244	0.9242	1.38E-25	0.8998
chr6	137114200	137114700	500	0	0.8934	8.87E-14	0.8934
chr3	173792600	173793100	500	0.0182	0.9091	1.70E-42	0.8908
chr5	98296800	98297300	500	0.0432	0.9333	2.58E-23	0.8901
chr12	44898000	44898500	500	0	0.8889	4.47E-11	0.8889
chr3	197328900	197329400	500	0.0169	0.9048	5.55E-10	0.8878
chr8	49739700	49740200	500	0.0324	0.9194	5.71E-34	0.887
chr12	122279300	122279800	500	0.0135	0.8969	3.46E-21	0.8834
chr17	43092200	43092700	500	0	0.8824	4.34E-10	0.8824
chr7	107563100	107563600	500	0.0495	0.931	1.05E-28	0.8815
chr11	72543200	72543700	500	0.0377	0.9167	2.94E-09	0.8789
chr14	58837800	58838300	500	0.0097	0.886	9.16E-18	0.8763
chr3	153443900	153444900	1000	0.0436	0.9197	6.24E-39	0.876
chr3	16953200	16953700	500	0.0896	0.9655	6.78E-09	0.876
chr17	42205700	42206200	500	0.057	0.933	1.13E-31	0.8759
chr6	53217600	53218100	500	0.0818	0.9571	1.54E-19	0.8754
chr3	112749000	112749500	500	0.0403	0.9154	4.11E-22	0.8752
chr8	22453700	22454200	500	0.003	0.8765	1.64E-50	0.8735
chr1	162860900	162861400	500	0.023	0.8932	8.37E-14	0.8702
chr6	36279700	36280200	500	0.0068	0.8762	1.14E-21	0.8694
chr5	80962500	80963000	500	0	0.8679	2.08E-15	0.8679
chr16	11312500	11313000	500	0	0.8679	2.14E-10	0.8679
chr16	29663900	29665400	1500	0.0081	0.8759	0.00E+00	0.8678
chr3	120438100	120438600	500	0.102	0.9639	2.98E-15	0.8618
chr8	134157000	134157500	500	0.0625	0.9219	6.10E-20	0.8594
chr6	42620900	42621400	500	0	0.8571	5.68E-08	0.8571
chr5	131836300	131836800	500	0.075	0.9315	1.26E-10	0.8565
chr14	60290000	60290500	500	0	0.8544	5.63E-14	0.8544
chr6	42850300	42851300	1000	0.0676	0.9211	2.38E-24	0.8534
chr8	28974100	28974600	500	0.0394	0.8927	2.03E-51	0.8533
chr22	22368500	22369000	500	0.0248	0.8778	1.18E-70	0.8529
chr14	69825300	69825800	500	0.0654	0.9174	2.73E-14	0.852
chr3	142700300	142700800	500	0.0063	0.8582	2.56E-23	0.8519
chr17	59491300	59491800	500	0.0662	0.9175	1.81E-16	0.8513
chr15	30881700	30882200	500	0.0493	0.8995	2.38E-26	0.8502
chr15	91496300	91496800	500	0	0.85	3.13E-17	0.85
chr17	18745300	18745800	500	0.0294	0.8775	3.47E-51	0.848
chr15	29436500	29437000	500	0.0336	0.8811	2.62E-66	0.8476
chr2	217795300	217795800	500	0	0.8472	1.78E-22	0.8472
chr11	16328100	16328600	500	0.0278	0.875	3.43E-11	0.8472
chr13	113063500	113064000	500	0.0102	0.8571	1.82E-15	0.8469

chr5	40472400	40472900	500	0.0197	0.8664	7.54E-35	0.8467
chr1	242549200	242549700	500	0	0.8462	8.53E-23	0.8462
chr11	58099100	58099600	500	0.0162	0.8612	4.45E-35	0.845
chr9	16020400	16020900	500	0.0132	0.8555	8.05E-23	0.8423
chr8	37550700	37551200	500	0.0093	0.8512	1.11E-16	0.8419
chr5	75722400	75722900	500	0.1215	0.9627	5.97E-23	0.8411
chr19	60454700	60455200	500	0.0316	0.8722	2.44E-62	0.8405
chr4	99587100	99587600	500	0.0128	0.8526	1.49E-12	0.8398
chr6	25149600	25150600	1000	0.0135	0.8514	0.00E+00	0.8379
chr1	32065200	32065700	500	0	0.8371	1.09E-44	0.8371
chr7	5337200	5337700	500	0.0727	0.9098	2.18E-14	0.8371
chr17	44269900	44270400	500	0.026	0.8621	3.94E-22	0.8361
chr1	36180800	36181300	500	0.0714	0.9067	1.23E-09	0.8352
chr18	10472700	10473700	1000	0.0713	0.9064	9.35E-70	0.8351
chr5	350000	350500	500	0.0297	0.8643	1.49E-16	0.8346
chr2	136481800	136482800	1000	0.0554	0.8887	1.87E-52	0.8332
chr4	89241100	89241600	500	0.1091	0.9423	1.05E-12	0.8332
chr17	40320700	40321200	500	0.0539	0.8859	6.94E-31	0.832
chr7	133897200	133897700	500	0.0769	0.9077	1.64E-24	0.8308
chr8	98060600	98061100	500	0.0741	0.9048	3.11E-07	0.8307
chr8	134141500	134142000	500	0	0.829	2.77E-58	0.829
chr14	80250600	80251100	500	0.0839	0.9122	2.05E-18	0.8283
chr2	100730900	100731400	500	0.0787	0.9067	4.85E-11	0.828
chr2	88108100	88108600	500	0.075	0.901	2.10E-16	0.826
chr19	16338500	16339500	1000	0.0011	0.8259	0.00E+00	0.8247
chr5	141791900	141792900	1000	0.0225	0.8467	0.00E+00	0.8243
chr11	116227400	116227900	500	0	0.8242	1.01E-17	0.8242
chr22	48705500	48706000	500	0.0649	0.8891	4.00E-76	0.8242
chr9	3518300	3518800	500	0.0263	0.8493	6.66E-25	0.823
chr11	16791000	16791500	500	0.1095	0.9322	1.05E-22	0.8228
chr3	135746000	135746500	500	0.1429	0.9651	7.63E-10	0.8223
chr1	19323400	19323900	500	0.0411	0.8624	2.36E-20	0.8213
chr9	96662800	96663300	500	0.0614	0.8826	1.07E-28	0.8212
chr7	37348300	37349300	1000	0.0012	0.821	0.00E+00	0.8198
chr2	234882000	234882500	500	0.0392	0.8591	5.36E-15	0.8198
chr6	44694000	44694500	500	0.1024	0.9222	5.68E-19	0.8198
chr17	18320500	18321000	500	0	0.8197	2.78E-39	0.8197
chr22	28992000	28994000	2000	0.0012	0.8195	0.00E+00	0.8183
chr17	53762700	53766300	3600	0.0003	0.8179	0.00E+00	0.8176
chr1	114215500	114216000	500	0	0.8169	2.24E-20	0.8169
chr6	13381700	13382700	1000	0.0037	0.8206	1.03E-40	0.8169
chr5	17045400	17045900	500	0.0235	0.84	3.24E-13	0.8165
chr12	110924300	110924800	500	0.0855	0.9016	1.01E-18	0.816
chr1	200499800	200500300	500	0.011	0.8269	9.73E-24	0.8159
chr4	8311000	8311500	500	0.053	0.8687	7.82E-18	0.8157
chr8	6535300	6535800	500	0.0667	0.8824	5.25E-09	0.8157
chr6	42462700	42463200	500	0.0761	0.8919	5.78E-23	0.8157
chr1	91969900	91970400	500	0.0172	0.8325	5.23E-18	0.8152
chr2	105758600	105759600	1000	0.0305	0.8455	5.15E-52	0.815
chr21	37538500	37539000	500	0.1595	0.9745	8.81E-16	0.815
chr9	92953000	92953500	500	0.0189	0.8333	1.88E-15	0.8145



chr16	30104400	30105900	1500	0.0505	0.8636	0.00E+00	0.8131
chr1	234184400	234185400	1000	0.0346	0.8477	9.66E-31	0.813
chr8	19116400	19116900	500	0	0.8125	9.33E-11	0.8125
chr4	141194300	141195300	1000	0.0865	0.899	5.46E-30	0.8125

表S2D. 自後三個月胎盤組織及母體血細胞中鑑別之100個最低甲基化區域之清單

染色體	起始	結束	大小(bp)	母體血 細胞	妊娠期 胎盤	P值	甲基化 差異
chr9	40380300	40380800	500	0.9667	0	1.13E-06	0.9667
chr1	31769200	31769700	500	0.9548	0.0256	5.57E-25	0.9291
chr18	12217600	12218100	500	0.9873	0.0602	1.63E-19	0.9271
chr20	19704400	19704900	500	0.9426	0.018	4.34E-18	0.9246
chr15	37317500	37318000	500	0.9358	0.0132	1.90E-25	0.9226
chrX	83368400	83368900	500	0.913	0	3.15E-07	0.913
chr11	27549100	27549600	500	0.9224	0.0123	3.92E-24	0.9101
chr18	58141500	58142000	500	0.9737	0.0645	1.07E-09	0.9092
chr1	159897000	159897500	500	0.9067	0	2.53E-16	0.9067
chr7	84347200	84348700	1500	0.9401	0.0407	0.00E+00	0.8994
chr2	216916100	216916600	500	0.9695	0.0714	2.13E-16	0.8981
chr7	144200000	144200500	500	0.9294	0.0317	1.24E-10	0.8977
chr1	241331600	241332100	500	0.9198	0.0227	0.00E+00	0.8971
chr7	123190000	123191000	1000	0.9341	0.0384	0.00E+00	0.8957
chr5	12626600	12628600	2000	0.9266	0.0321	0.00E+00	0.8944
chr12	12971300	12972300	1000	0.9361	0.0438	0.00E+00	0.8923
chr22	20936500	20937000	500	0.9528	0.0606	1.87E-06	0.8922
chr13	31321900	31322400	500	0.9231	0.0313	1.43E-06	0.8918
chr22	21701500	21702000	500	0.9579	0.0667	1.30E-09	0.8912
chr10	104794400	104794900	500	1	0.1111	6.10E-09	0.8889
chr7	21835800	21836300	500	0.9156	0.0267	3.85E-13	0.8889
chr10	16134800	16135300	500	0.95	0.0635	4.79E-10	0.8865
chr3	47410400	47410900	500	0.9213	0.0357	6.63E-17	0.8855
chr10	49344400	49345400	1000	0.9422	0.0571	0.00E+00	0.8851
chr2	209073900	209074400	500	0.9196	0.0353	1.63E-22	0.8843
chr1	89131200	89131700	500	0.9314	0.0472	1.05E-75	0.8842
chr3	167118500	167119500	1000	0.9365	0.0527	0.00E+00	0.8838
chr18	1395900	1397400	1500	0.9308	0.0472	0.00E+00	0.8836
chr2	59670300	59670800	500	0.9433	0.0599	5.09E-23	0.8834
chr14	28368900	28369400	500	0.9446	0.0619	5.03E-64	0.8827
chr3	126028800	126029300	500	0.9379	0.0556	1.83E-20	0.8823
chr9	69378900	69379900	1000	0.8816	0	6.02E-51	0.8816
chr5	105517300	105518300	1000	0.927	0.0461	0.00E+00	0.8808
chr2	31572400	31573400	1000	0.9176	0.037	0.00E+00	0.8806
chr5	42003700	42004700	1000	0.9262	0.0462	0.00E+00	0.88
chr14	94718300	94718800	500	0.9548	0.0764	6.67E-19	0.8784
chr19	56417800	56418300	500	0.9615	0.0833	1.71E-06	0.8782
chr2	70183000	70183500	500	0.9694	0.0914	9.49E-39	0.878
chr4	118939300	118939800	500	0.9192	0.0412	2.20E-34	0.878
chr13	59246400	59246900	500	0.9402	0.0633	5.40E-12	0.8769
chr12	74021100	74022100	1000	0.9513	0.0752	0.00E+00	0.8761
chr2	173432500	173433000	500	0.9529	0.0778	5.39E-12	0.8752
chr16	24004400	24004900	500	0.9239	0.0488	3.25E-23	0.8751
chr13	27596300	27597300	1000	0.9538	0.0795	0.00E+00	0.8743
chr15	88904300	88904800	500	0.9212	0.0481	7.69E-27	0.8731

chr18	12720200	12721200	1000	0.9346	0.0618	0.00E+00	0.8728
chr15	60975900	60976900	1000	0.9311	0.0587	0.00E+00	0.8724
chr21	39630100	39631100	1000	0.9423	0.07	4.68E-43	0.8723
chr5	123933900	123934400	500	0.943	0.0707	2.60E-38	0.8722
chr8	77382600	77383600	1000	0.9117	0.0395	0.00E+00	0.8722
chr21	32238800	32239300	500	0.9396	0.0677	1.28E-18	0.8719
chr5	175019600	175020100	500	0.9542	0.0828	4.34E-20	0.8714
chr8	134437400	134438400	1000	0.9083	0.037	4.79E-29	0.8713
chr5	69668800	69669300	500	0.9194	0.0492	2.88E-09	0.8702
chr1	60877900	60878900	1000	0.9378	0.068	0.00E+00	0.8698
chr16	80650400	80650900	500	0.9309	0.0611	2.49E-32	0.8698
chr18	59388800	59389300	500	0.9706	0.1008	4.84E-15	0.8697
chr2	15969400	15970400	1000	0.8939	0.0244	3.07E-52	0.8695
chr13	56207100	56208100	1000	0.9184	0.0505	0.00E+00	0.868
chr3	180395300	180395800	500	0.9877	0.12	1.73E-09	0.8677
chr6	153238200	153239200	1000	0.9123	0.0452	0.00E+00	0.8671
chr18	61635100	61635600	500	0.9268	0.06	2.77E-13	0.8668
chr3	177562200	177563200	1000	0.9121	0.0455	0.00E+00	0.8666
chr4	160368300	160370200	1900	0.9272	0.0606	0.00E+00	0.8665
chr6	144626900	144627400	500	0.9114	0.046	2.10E-12	0.8654
chr16	59885500	59886500	1000	0.9407	0.0757	1.12E-62	0.865
chr1	55667100	55667600	500	0.9095	0.0446	8.62E-39	0.8649
chr2	83598300	83599300	1000	0.9366	0.0718	0.00E+00	0.8648
chr4	105135200	105136200	1000	0.913	0.0486	0.00E+00	0.8644
chr14	32048400	32048900	500	0.9142	0.0499	5.43E-53	0.8643
chr1	223482700	223483700	1000	0.9636	0.0997	2.69E-34	0.864
chr14	47487700	47488200	500	0.915	0.0514	5.45E-33	0.8636
chr3	104515000	104515500	500	1	0.1373	1.08E-06	0.8627
chr7	14008300	14009800	1500	0.9349	0.0725	0.00E+00	0.8624
chr1	243134000	243135500	1500	0.9208	0.0588	0.00E+00	0.8619
chr10	14156400	14156900	500	0.9105	0.0489	0.00E+00	0.8616
chr2	118616200	118617200	1000	0.9178	0.0565	0.00E+00	0.8613
chr17	8455500	8456000	500	0.8941	0.0331	1.94E-18	0.8611
chr12	15392200	15393200	1000	0.9307	0.0697	0.00E+00	0.861
chr8	81275900	81276900	1000	0.9291	0.0684	0.00E+00	0.8606
chr1	234269300	234269800	500	0.9471	0.087	2.25E-25	0.8602
chr1	181970300	181970800	500	0.9167	0.0566	2.15E-08	0.8601
chr2	55775000	55776000	1000	0.9159	0.0559	0.00E+00	0.8599
chr3	88338000	88339000	1000	0.8909	0.0311	0.00E+00	0.8598
chr5	140078700	140079200	500	0.8852	0.0253	0.00E+00	0.8598
chr21	16720900	16721400	500	0.9317	0.0721	1.38E-15	0.8596
chr11	104891100	104892600	1500	0.9164	0.0569	0.00E+00	0.8595
chr1	184204700	184205200	500	0.9194	0.0603	8.16E-16	0.859
chr6	160732500	160733000	500	0.9191	0.0606	1.31E-10	0.8585
chr8	37134300	37134800	500	0.9151	0.0567	1.21E-26	0.8584
chr18	5869800	5870300	500	0.913	0.0548	1.21E-09	0.8582
chr1	98448100	98448600	500	0.9574	0.1	2.08E-05	0.8574
chr3	152897800	152898300	500	0.9402	0.0828	3.28E-18	0.8574
chr1	110304000	110304500	500	0.9783	0.121	2.60E-18	0.8572
chr2	86993600	86995600	2000	0.8965	0.0395	0.00E+00	0.857
chr19	15428100	15430600	2500	0.9424	0.0862	0.00E+00	0.8563

chr13	75176800	75177800	1000	0.9258	0.0697	3.09E-47	0.8561
chr13	24126700	24127200	500	0.9498	0.0938	1.57E-17	0.856
chr16	28238500	28240500	2000	0.9427	0.0868	0.00E+00	0.8559
chr2	158079500	158080500	1000	0.9199	0.0642	0.00E+00	0.8557

表S3A. 根據前三個月母體血漿亞硫酸氫鹽定序資料推斷經高甲基化之前100個基因座之清單

染色體	起始	結束	母體血 細胞	CVS	甲基化 差異
chr22	39189067	39189863	0	0.8444	0.8444
chr17	53763065	53764027	0	0.7922	0.7922
chr7	41887694	41888212	0	0.7614	0.7614
chr2	1.14E+08	1.14E+08	0	0.751	0.751
chr12	25096242	25097206	0	0.7098	0.7098
chr1	66574104	66574793	0	0.7025	0.7025
chr6	11489985	11490755	0	0.7004	0.7004
chr6	1.07E+08	1.07E+08	0	0.6978	0.6978
chr10	30858286	30858871	0	0.6693	0.6693
chr17	21131574	21132167	0	0.6496	0.6496
chr18	13454740	13455292	0	0.5468	0.5468
chr16	11298755	11299326	0	0.5373	0.5373
chr2	1.75E+08	1.75E+08	0	0.5196	0.5196
chr19	44060511	44061036	0	0.5128	0.5128
chr6	1.08E+08	1.08E+08	0	0.5	0.5
chr3	71261611	71262501	0	0.4587	0.4587
chr9	36247847	36248885	0	0.447	0.447
chr19	17819240	17820082	0	0.4279	0.4279
chr17	53769900	53770731	0	0.4102	0.4102
chr1	1.12E+08	1.12E+08	0.0002	0.6167	0.6166
chr7	1.34E+08	1.34E+08	0.0003	0.4351	0.4348
chr3	11658550	11659929	0.0004	0.4299	0.4295
chr17	53764417	53765963	0.0005	0.7967	0.7961
chr10	11246762	11249052	0.0005	0.4002	0.3997
chr22	28992647	28993434	0.0006	0.8092	0.8087
chr15	62460278	62461007	0.0006	0.4334	0.4328
chr1	31002038	31003474	0.0007	0.5926	0.5919
chr19	3129246	3132159	0.0008	0.7725	0.7717
chr12	1.21E+08	1.21E+08	0.0008	0.7303	0.7295
chr19	12304446	12305741	0.0009	0.6986	0.6978
chr3	67788734	67789395	0.001	0.9131	0.9121
chr9	1.32E+08	1.32E+08	0.001	0.7047	0.7037
chr19	6723370	6724479	0.001	0.689	0.688
chr3	1.84E+08	1.84E+08	0.001	0.4384	0.4374
chr2	53848089	53849214	0.001	0.4368	0.4358
chr17	59450886	59452113	0.0012	0.469	0.4678
chr5	1.72E+08	1.72E+08	0.0014	0.578	0.5766
chr21	35342527	35343373	0.0014	0.5392	0.5378
chr21	45164804	45165437	0.0015	0.4251	0.4236
chrX	3742417	3744601	0.0016	0.4486	0.447
chr21	45158293	45159003	0.0017	0.7799	0.7782
chr7	39839340	39839876	0.0017	0.4074	0.4057
chr2	1.75E+08	1.75E+08	0.0018	0.4816	0.4797
chr12	1.24E+08	1.24E+08	0.0019	0.6306	0.6287

chr3	50352688	50353823	0.002	0.624	0.622
chr9	97264382	97265523	0.0021	0.5008	0.4987
chr7	64178628	64179354	0.0021	0.4088	0.4066
chr9	94767202	94767802	0.0023	0.7568	0.7544
chr5	42986308	42988304	0.0023	0.4882	0.4859
chr17	63854127	63854693	0.0024	0.8266	0.8242
chr12	1.22E+08	1.22E+08	0.0024	0.4869	0.4844
chr17	16260170	16260909	0.0026	0.6404	0.6378
chr4	39874787	39875456	0.0027	0.7233	0.7206
chr12	6441080	6441608	0.0027	0.6228	0.6201
chr19	45015653	45016886	0.0028	0.5444	0.5416
chr6	30757752	30758823	0.0028	0.4783	0.4755
chr6	41636176	41637112	0.0028	0.4254	0.4226
chr12	6315199	6315765	0.0029	0.4613	0.4584
chr14	76576283	76577070	0.0029	0.4365	0.4336
chr16	48857790	48858300	0.0031	0.5625	0.5594
chr5	1.7E+08	1.7E+08	0.0031	0.4752	0.4721
chr13	26897813	26898557	0.0032	0.4354	0.4322
chr14	52753948	52754571	0.0032	0.4221	0.4189
chr1	1.66E+08	1.66E+08	0.0033	0.5579	0.5545
chr12	56157424	56158348	0.0033	0.47	0.4667
chr22	16079971	16080532	0.0034	0.6226	0.6193
chr7	1946410	1946975	0.0036	0.6826	0.6789
chr11	258799	259749	0.0036	0.5072	0.5037
chr6	13381944	13382477	0.0037	0.5945	0.5908
chr7	1.27E+08	1.27E+08	0.0037	0.5096	0.5058
chr13	23743886	23744467	0.0037	0.4534	0.4497
chr2	1.21E+08	1.21E+08	0.0038	0.7175	0.7137
chr21	25855853	25857105	0.0039	0.4661	0.4622
chr2	43211724	43212565	0.0039	0.4345	0.4306
chr12	1.08E+08	1.08E+08	0.0041	0.6024	0.5983
chr15	92928924	92929575	0.0041	0.4074	0.4033
chr19	10731043	10731636	0.0042	0.5868	0.5826
chr6	1.45E+08	1.45E+08	0.0043	0.5783	0.574
chr1	52875323	52875907	0.0044	0.4145	0.4101
chr14	75058186	75058956	0.0045	0.602	0.5975
chr12	1.21E+08	1.21E+08	0.0045	0.4821	0.4776
chr17	76873737	76874417	0.0046	0.6012	0.5966
chr2	2.38E+08	2.38E+08	0.0049	0.7654	0.7604
chr2	1.98E+08	1.98E+08	0.0049	0.7228	0.7179
chr6	1.47E+08	1.47E+08	0.0049	0.4967	0.4918
chr9	1.36E+08	1.36E+08	0.0049	0.4584	0.4535
chr1	67545402	67546771	0.005	0.4971	0.4921
chr6	1.58E+08	1.58E+08	0.0052	0.6145	0.6093
chr3	1.7E+08	1.7E+08	0.0052	0.5845	0.5794
chr1	2.34E+08	2.34E+08	0.0053	0.7033	0.6979
chr10	80715722	80716751	0.0053	0.6515	0.6462
chr4	48602901	48603736	0.0053	0.6315	0.6262
chr19	13957965	13958580	0.0053	0.599	0.5937
chr1	90081114	90082367	0.0053	0.4574	0.4521
chr2	1.06E+08	1.06E+08	0.0054	0.8858	0.8804

chr16	29664213	29665369	0.0054	0.8339	0.8285
chr1	1.59E+08	1.59E+08	0.0054	0.7663	0.7608
chr13	97926489	97927025	0.0054	0.6229	0.6175
chr1	41604452	41605277	0.0054	0.6011	0.5956
chr9	1.28E+08	1.28E+08	0.0054	0.5871	0.5818

表S3B. 根據前三個月母體血漿亞硫酸氫鹽定序資料推斷經低甲基化之前100個基因座之清單

染色體	起始	結束	母體血細胞	CVS	甲基化差異
chr1	235771917	235772426	0.9868	0.549	0.4377
chr1	97357972	97358622	0.9835	0.4805	0.503
chr1	4490516	4491074	0.9826	0.4793	0.5032
chr4	181124168	181124671	0.9825	0.4725	0.5099
chr16	71908694	71909213	0.982	0.5581	0.4239
chr3	182727915	182728477	0.981	0.3577	0.6233
chr5	115339535	115340038	0.9802	0.5455	0.4347
chr3	195855575	195856122	0.9801	0.3793	0.6008
chr6	155437621	155438161	0.9799	0.5991	0.3808
chr9	20468093	20468904	0.9798	0.4271	0.5527
chr10	90702298	90702987	0.9787	0.3324	0.6463
chr1	170581654	170582162	0.9785	0.4817	0.4968
chr3	108816849	108817794	0.9783	0.4793	0.4989
chr20	36912749	36913319	0.9783	0.5	0.4783
chr13	72517281	72517839	0.9782	0.4855	0.4927
chr12	103553001	103553677	0.9774	0.492	0.4854
chr22	27638905	27639408	0.9766	0.5385	0.4382
chr7	17290850	17291462	0.9763	0.59	0.3863
chr6	17227866	17228510	0.976	0.4058	0.5703
chr15	56998547	56999107	0.9754	0.3766	0.5988
chr7	70965945	70966842	0.9753	0.5893	0.386
chr3	32159338	32160065	0.9752	0.5379	0.4372
chr16	17043258	17043854	0.9752	0.5521	0.4231
chr16	22776223	22776850	0.9752	0.5735	0.4017
chr5	169344029	169344869	0.9751	0.4211	0.5541
chr11	34324955	34325722	0.975	0.5561	0.4189
chr8	58554745	58555376	0.9747	0.5784	0.3964
chr1	153933389	153934121	0.9746	0.463	0.5116
chr14	88003983	88004485	0.9745	0.5379	0.4366
chr3	151738501	151739120	0.9741	0.4901	0.484
chr14	105618699	105619606	0.974	0.3457	0.6283
chr16	24060085	24060702	0.9738	0.3991	0.5747
chr8	68941792	68942711	0.9738	0.5449	0.429
chr12	53208707	53209304	0.9737	0.4847	0.489
chr7	76892564	76893249	0.9736	0.5664	0.4072
chr3	69464294	69464971	0.9736	0.5893	0.3843
chr19	61401137	61401745	0.9732	0.4933	0.4799
chr11	124569867	124570490	0.9732	0.5136	0.4595
chr18	42618440	42619096	0.9732	0.5942	0.379
chr5	169398896	169399637	0.9731	0.498	0.4751
chr5	169328124	169328983	0.9731	0.572	0.401
chr20	34679880	34680448	0.9731	0.5922	0.3809
chr16	9042198	9042702	0.973	0.4286	0.5444
chr10	90205044	90205701	0.973	0.4407	0.5323
chr13	33236454	33236997	0.973	0.5906	0.3824



chr16	73284579	73285087	0.9729	0.5602	0.4127
chr8	29100691	29101428	0.9728	0.505	0.4678
chr2	202383851	202384447	0.9727	0.5461	0.4267
chr3	179501620	179502300	0.9722	0.5766	0.3956
chr6	107674976	107675906	0.9719	0.4434	0.5285
chr6	107880632	107881161	0.9718	0.5623	0.4095
chr12	56350283	56350933	0.9718	0.5909	0.3809
chr19	40636458	40637339	0.9717	0.4941	0.4776
chr2	223472599	223473287	0.9714	0.1824	0.7891
chr22	20709067	20709787	0.9714	0.5149	0.4565
chr19	46095583	46096190	0.9713	0.5385	0.4328
chr6	90258338	90259318	0.9712	0.3415	0.6297
chr2	54598347	54598933	0.9712	0.5894	0.3819
chr3	114810453	114811493	0.9711	0.5166	0.4545
chr19	15851125	15851654	0.9711	0.5236	0.4476
chr8	42889138	42890084	0.9711	0.5652	0.4059
chr18	52354390	52355064	0.971	0.598	0.373
chr15	38206236	38207010	0.9709	0.4186	0.5523
chr7	99700554	99701110	0.9708	0.305	0.6658
chr12	19487336	19487855	0.9708	0.4105	0.5603
chr7	87996908	87997437	0.9708	0.5462	0.4246
chr6	63628653	63629378	0.9707	0.529	0.4417
chr15	38209108	38209618	0.9706	0.5882	0.3821
chr19	6623769	6624450	0.9704	0.5179	0.4526
chr2	10794513	10795242	0.9704	0.5976	0.3728
chr2	118472785	118474454	0.9704	0.5992	0.3712
chr5	57820209	57820801	0.9701	0.5815	0.3886
chr10	100183380	100184702	0.9701	0.5826	0.3875
chr2	8151989	8152646	0.97	0.4701	0.4999
chr10	3938374	3938914	0.9699	0.1741	0.7958
chr9	123724524	123725439	0.9697	0.57	0.3997
chr14	89085469	89086097	0.9696	0.3278	0.6418
chr16	14129437	14130133	0.9695	0.5304	0.4392
chr5	60746367	60747191	0.9695	0.5571	0.4124
chr1	92002953	92003729	0.9694	0.52	0.4494
chr6	31264677	31265413	0.9693	0.5135	0.4558
chr7	99317013	99318281	0.9692	0.5117	0.4574
chr8	8808867	8809422	0.9692	0.5691	0.4002
chr19	20052165	20052720	0.969	0.2792	0.6898
chr8	129139026	129139573	0.969	0.3458	0.6232
chr11	122314929	122315458	0.969	0.4232	0.5458
chr13	98377663	98378165	0.9688	0.3319	0.6369
chr9	107606194	107606872	0.9688	0.449	0.5198
chr8	56096904	56097736	0.9688	0.5267	0.4422
chr7	128093836	128094339	0.9688	0.5929	0.3758
chr2	103109370	103109916	0.9686	0.3333	0.6352
chr3	101803534	101804063	0.9686	0.5027	0.4659
chr10	69505720	69506278	0.9684	0.2515	0.7169
chr13	26608225	26608754	0.9683	0.3614	0.6069
chr1	90993315	90993828	0.9683	0.5519	0.4164
chr6	11361243	11361801	0.9681	0.2578	0.7103

chr21	36529300	36529981	0.968	0.1944	0.7736
chr21	37813953	37814521	0.9679	0.2175	0.7505
chr2	15226273	15227211	0.9679	0.5134	0.4545
chr19	4102809	4103443	0.9679	0.5646	0.4034

表S3C. 根據後三個月母體血漿亞硫酸氫鹽定序資料推斷經高甲基化之前100個基因座之清單

染色體	起始	結束	母體血細胞	妊娠期胎盤	甲基化差異
chr17	53763065	53764027	0.0000	0.8680	0.8680
chr22	39189067	39189863	0.0000	0.8233	0.8233
chr10	30858286	30858871	0.0000	0.7713	0.7713
chr7	41887694	41888212	0.0000	0.7578	0.7578
chr2	1.14E+08	1.14E+08	0.0000	0.7500	0.7500
chr12	25096242	25097206	0.0000	0.7332	0.7332
chr6	1.07E+08	1.07E+08	0.0000	0.7229	0.7229
chr1	66574104	66574793	0.0000	0.7136	0.7136
chr16	11298755	11299326	0.0000	0.7005	0.7005
chr6	11489985	11490755	0.0000	0.6935	0.6935
chr18	13454740	13455292	0.0000	0.6594	0.6594
chr6	1.08E+08	1.08E+08	0.0000	0.6231	0.6231
chrX	3627885	3628549	0.0000	0.6133	0.6133
chr12	7979754	7980413	0.0000	0.6118	0.6118
chr3	71261611	71262501	0.0000	0.5938	0.5938
chr17	53769900	53770731	0.0000	0.5586	0.5586
chr11	1.18E+08	1.18E+08	0.0000	0.5558	0.5558
chr19	44060511	44061036	0.0000	0.5464	0.5464
chr2	2.38E+08	2.38E+08	0.0000	0.5330	0.5330
chr1	1.91E+08	1.91E+08	0.0000	0.5294	0.5294
chr1	1.44E+08	1.44E+08	0.0000	0.4857	0.4857
chr2	1.75E+08	1.75E+08	0.0000	0.4785	0.4785
chr4	15366889	15367646	0.0000	0.4729	0.4729
chr2	19537237	19537737	0.0000	0.4599	0.4599
chr1	1.15E+08	1.15E+08	0.0000	0.4351	0.4351
chr1	1.54E+08	1.54E+08	0.0000	0.4299	0.4299
chr14	51383387	51384149	0.0000	0.4186	0.4186
chr1	1.12E+08	1.12E+08	0.0002	0.5350	0.5348
chr3	11658550	11659929	0.0004	0.5579	0.5575
chr17	53764417	53765963	0.0005	0.7894	0.7889
chr22	28992647	28993434	0.0006	0.8053	0.8047
chr6	27214981	27215823	0.0006	0.4593	0.4587
chr1	31002038	31003474	0.0007	0.6309	0.6302
chr12	1.21E+08	1.21E+08	0.0008	0.7360	0.7352
chr19	3129246	3132159	0.0008	0.7257	0.7249
chr19	12304446	12305741	0.0009	0.6397	0.6388
chr6	28723918	28724965	0.0009	0.4344	0.4335
chr19	6723370	6724479	0.0010	0.7280	0.7270
chr2	53848089	53849214	0.0010	0.4060	0.4050
chr9	1.32E+08	1.32E+08	0.0010	0.7558	0.7548
chr3	67788734	67789395	0.0010	0.9219	0.9209
chr19	18276368	18277132	0.0011	0.5136	0.5125
chr17	59450886	59452113	0.0012	0.5196	0.5184
chr17	74243852	74244670	0.0012	0.4117	0.4105

chr3	1.85E+08	1.85E+08	0.0014	0.4961	0.4948
chr21	35342527	35343373	0.0014	0.5126	0.5112
chr5	1.72E+08	1.72E+08	0.0014	0.6531	0.6516
chr21	45164804	45165437	0.0015	0.4364	0.4349
chrX	3742417	3744601	0.0016	0.7517	0.7501
chr21	45158293	45159003	0.0017	0.8180	0.8163
chr2	1.75E+08	1.75E+08	0.0018	0.6214	0.6196
chr12	1.24E+08	1.24E+08	0.0019	0.5906	0.5887
chr3	50352688	50353823	0.0020	0.6082	0.6062
chr9	94767202	94767802	0.0023	0.8327	0.8304
chr17	63854127	63854693	0.0024	0.7886	0.7862
chr12	1.22E+08	1.22E+08	0.0024	0.5021	0.4997
chr17	16260170	16260909	0.0026	0.5780	0.5754
chr12	6441080	6441608	0.0027	0.7471	0.7444
chr4	39874787	39875456	0.0027	0.7962	0.7935
chr18	50536032	50536649	0.0027	0.4920	0.4893
chr6	30757752	30758823	0.0028	0.4029	0.4001
chr19	2571763	2572292	0.0031	0.4200	0.4169
chr5	1.7E+08	1.7E+08	0.0031	0.4218	0.4187
chr13	26897813	26898557	0.0032	0.6485	0.6453
chr12	56157424	56158348	0.0033	0.5541	0.5508
chr1	1.66E+08	1.66E+08	0.0033	0.5147	0.5113
chr22	16079971	16080532	0.0034	0.6265	0.6231
chr6	16820551	16821134	0.0035	0.4800	0.4765
chr11	258799	259749	0.0036	0.5475	0.5439
chr7	1946410	1946975	0.0036	0.8251	0.8215
chr6	13381944	13382477	0.0037	0.8221	0.8183
chr7	1.27E+08	1.27E+08	0.0037	0.4767	0.4730
chr2	1.21E+08	1.21E+08	0.0038	0.6734	0.6697
chr2	43211724	43212565	0.0039	0.4256	0.4217
chr15	92928924	92929575	0.0041	0.5605	0.5564
chr12	1.08E+08	1.08E+08	0.0041	0.7313	0.7271
chr19	10731043	10731636	0.0042	0.5668	0.5626
chr6	1.45E+08	1.45E+08	0.0043	0.5910	0.5867
chr1	52875323	52875907	0.0044	0.6115	0.6071
chr12	1.21E+08	1.21E+08	0.0045	0.5884	0.5839
chr14	75058186	75058956	0.0045	0.6534	0.6489
chr17	76873737	76874417	0.0046	0.5658	0.5612
chr6	1.47E+08	1.47E+08	0.0049	0.5826	0.5777
chr2	1.98E+08	1.98E+08	0.0049	0.7944	0.7895
chr2	2.38E+08	2.38E+08	0.0049	0.7328	0.7278
chr8	1.42E+08	1.42E+08	0.0050	0.7728	0.7679
chr3	1.7E+08	1.7E+08	0.0052	0.7227	0.7176
chr6	1.58E+08	1.58E+08	0.0052	0.6389	0.6337
chr2	2.38E+08	2.38E+08	0.0052	0.4238	0.4185
chr11	56948854	56949496	0.0053	0.4484	0.4431
chr4	48602901	48603736	0.0053	0.5920	0.5867
chr5	1.31E+08	1.31E+08	0.0053	0.4858	0.4805
chr10	80715722	80716751	0.0053	0.5249	0.5196
chr19	13957965	13958580	0.0053	0.4379	0.4326
chr1	2.34E+08	2.34E+08	0.0053	0.8440	0.8387

chr13	97926489	97927025	0.0054	0.7233	0.7179
chr9	1.28E+08	1.28E+08	0.0054	0.7312	0.7258
chr2	1.06E+08	1.06E+08	0.0054	0.8513	0.8459
chr2	96556705	96557637	0.0054	0.4095	0.4041
chr16	29664213	29665369	0.0054	0.8837	0.8783

表S3D. 由後三個月母體血漿亞硫酸氫鹽定序資料推斷經低甲基化之前 100 個基因座之清單

染色體	起始	結束	母體血 細胞	妊娠期 胎盤	甲基化 差異
chr10	7548948	7549483	0.9866	0.5685	0.4181
chr1	4490516	4491074	0.9826	0.5015	0.4810
chr4	1.81E+08	1.81E+08	0.9825	0.5981	0.3843
chr3	1.83E+08	1.83E+08	0.9810	0.2925	0.6886
chr3	1.96E+08	1.96E+08	0.9801	0.4643	0.5158
chr6	1.55E+08	1.55E+08	0.9799	0.4610	0.5189
chr1	1.71E+08	1.71E+08	0.9785	0.5122	0.4662
chr20	36912749	36913319	0.9783	0.4513	0.5269
chr22	38583100	38583616	0.9783	0.5428	0.4355
chr1	19391314	19392207	0.9778	0.5273	0.4505
chr5	1.74E+08	1.74E+08	0.9770	0.5852	0.3918
chr19	13678906	13679531	0.9760	0.5812	0.3949
chr14	83650790	83651395	0.9760	0.5378	0.4382
chr15	56998547	56999107	0.9754	0.4691	0.5063
chr16	22776223	22776850	0.9752	0.5114	0.4638
chr5	1.69E+08	1.69E+08	0.9751	0.4809	0.4943
chr8	58554745	58555376	0.9747	0.5977	0.3770
chr14	1.06E+08	1.06E+08	0.9740	0.2069	0.7671
chr8	68941792	68942711	0.9738	0.5872	0.3866
chr16	24060085	24060702	0.9738	0.3470	0.6268
chr12	53208707	53209304	0.9737	0.5278	0.4459
chr5	1.69E+08	1.69E+08	0.9731	0.5057	0.4673
chr16	9042198	9042702	0.9730	0.1860	0.7869
chr10	90205044	90205701	0.9730	0.5922	0.3808
chr3	1.89E+08	1.89E+08	0.9720	0.4949	0.4771
chr6	1.08E+08	1.08E+08	0.9719	0.5825	0.3894
chr2	2.23E+08	2.23E+08	0.9714	0.3333	0.6381
chr19	46095583	46096190	0.9713	0.5065	0.4648
chr8	1.41E+08	1.41E+08	0.9713	0.5753	0.3959
chr6	90258338	90259318	0.9712	0.4357	0.5355
chr13	51403556	51404069	0.9710	0.3980	0.5731
chr18	66875048	66875726	0.9710	0.5259	0.4451
chr7	99700554	99701110	0.9708	0.3757	0.5951
chr7	87996908	87997437	0.9708	0.5720	0.3988
chr19	6623769	6624450	0.9704	0.4774	0.4930
chr1	97639047	97639749	0.9701	0.4148	0.5553
chr16	23892096	23892772	0.9701	0.5000	0.4701
chr10	3938374	3938914	0.9699	0.1148	0.8551
chr14	89085469	89086097	0.9696	0.2964	0.6732
chr8	1.29E+08	1.29E+08	0.9690	0.3565	0.6126
chr13	98377663	98378165	0.9688	0.3123	0.6566
chr8	56096904	56097736	0.9688	0.4562	0.5126
chr2	1.03E+08	1.03E+08	0.9686	0.3459	0.6227
chr13	26608225	26608754	0.9683	0.4562	0.5121
chr2	22738157	22738760	0.9682	0.5122	0.4560
chr6	11361243	11361801	0.9681	0.2646	0.7035
chr21	36529300	36529981	0.9680	0.1829	0.7852

chr21	37813953	37814521	0.9679	0.3061	0.6619
chr2	2.43E+08	2.43E+08	0.9679	0.5750	0.3929
chr4	12413543	12414103	0.9679	0.5944	0.3735
chr3	1.27E+08	1.27E+08	0.9677	0.4030	0.5648
chr7	33509047	33509556	0.9676	0.4627	0.5048
chr14	59284846	59285553	0.9674	0.5254	0.4420
chr17	42623453	42624024	0.9673	0.4318	0.5355
chr19	6778363	6779377	0.9671	0.4416	0.5255
chr4	41798250	41798788	0.9670	0.5000	0.4670
chr5	88054080	88054588	0.9669	0.2238	0.7431
chr16	24109379	24110289	0.9669	0.5062	0.4607
chr10	13847159	13847895	0.9667	0.3188	0.6479
chr10	1.27E+08	1.27E+08	0.9667	0.5423	0.4244
chr12	1.12E+08	1.12E+08	0.9663	0.3722	0.5941
chr10	17220886	17221845	0.9662	0.4455	0.5207
chr8	5947355	5947862	0.9662	0.5171	0.4491
chr3	73740840	73741439	0.9659	0.3657	0.6002
chr14	57945953	57946875	0.9658	0.5357	0.4301
chr14	50905777	50906333	0.9658	0.3008	0.6650
chr15	90275374	90276000	0.9657	0.5409	0.4248
chr22	24717299	24718197	0.9657	0.5160	0.4497
chr7	36530128	36530987	0.9656	0.5194	0.4462
chr2	1.31E+08	1.31E+08	0.9655	0.4384	0.5271
chr4	42116988	42117788	0.9654	0.5195	0.4459
chr12	1.16E+08	1.16E+08	0.9653	0.5594	0.4059
chr2	7491785	7492736	0.9652	0.4556	0.5097
chr19	6599638	6600187	0.9652	0.5488	0.4163
chr6	25326803	25327398	0.9651	0.3974	0.5677
chr4	1.7E+08	1.7E+08	0.9651	0.4933	0.4718
chr7	99875338	99876155	0.9650	0.2696	0.6953
chr14	97144328	97145208	0.9649	0.5377	0.4272
chr3	11718596	11719163	0.9649	0.5521	0.4128
chr14	1E+08	1E+08	0.9649	0.3794	0.5855
chr7	1.5E+08	1.5E+08	0.9648	0.3327	0.6322
chr12	56357827	56358328	0.9648	0.4217	0.5430
chr10	8275750	8276276	0.9647	0.3100	0.6547
chr11	16999685	17000209	0.9647	0.2765	0.6882
chr22	34419356	34419861	0.9646	0.4245	0.5401
chr18	72453151	72453725	0.9646	0.4700	0.4946
chr5	49919879	49920699	0.9645	0.3169	0.6476
chr1	24580891	24581805	0.9643	0.3565	0.6078
chr22	18233774	18234492	0.9641	0.5205	0.4436
chr14	45356178	45356903	0.9640	0.3934	0.5706
chr3	53007193	53008661	0.9638	0.4902	0.4737
chr4	55027912	55028539	0.9637	0.5254	0.4384
chr5	1.37E+08	1.37E+08	0.9637	0.5290	0.4347
chr1	2.23E+08	2.23E+08	0.9636	0.0997	0.8640
chr7	1.35E+08	1.35E+08	0.9636	0.2959	0.6677
chr5	80350438	80351169	0.9636	0.4969	0.4667
chr12	31889600	31890343	0.9636	0.1745	0.7891
chr12	8365395	8366096	0.9636	0.5721	0.3914

chr19	15424819	15425355	0.9635	0.2836	0.6799
chr10	10985469	10986409	0.9635	0.4877	0.4759



## 【發明摘要】

### 【中文發明名稱】

來自血漿之胚胎或腫瘤甲基化模式組(METHYLOME)之非侵入性測定

### 【英文發明名稱】

NON-INVASIVE DETERMINATION OF METHYLOME OF FETUS OR TUMOR FROM PLASMA

### 【中文】

本發明係關於可測定並使用各種組織及樣品之甲基化概況的系統、方法及設備。提供實例。可基於血漿甲基化(或具有細胞游離DNA之其他樣品)與母體/患者之甲基化概況之比較來推斷胎兒/腫瘤組織之甲基化概況。當樣品具有DNA混合物時，可使用組織特異性等位基因測定胎兒/腫瘤組織之甲基化概況以鑑別來自胎兒/腫瘤之DNA。甲基化概況可用於確定胎兒/腫瘤之基因組中的複本數目變化。已經由各種技術鑑別胎兒之甲基化標記物。可藉由測定DNA片段之大小分佈之大小參數來測定甲基化概況，其中大小參數之參考值可用於確定甲基化程度。另外，甲基化程度可用於測定癌症程度。

### 【英文】

Systems, methods, and apparatuses can determine and use methylation profiles of various tissues and samples. Examples are provided. A methylation profile can be deduced for fetal/tumor tissue based on a comparison of plasma methylation (or other sample with cell-free DNA) to a methylation profile of the mother/patient. A methylation

profile can be determined for fetal/tumor tissue using tissue-specific alleles to identify DNA from the fetus/tumor when the sample has a mixture of DNA. A methylation profile can be used to determine copy number variations in genome of a fetus/tumor. Methylation markers for a fetus have been identified via various techniques. The methylation profile can be determined by determining a size parameter of a size distribution of DNA fragments, where reference values for the size parameter can be used to determine methylation levels. Additionally, a methylation level can be used to determine a level of cancer.

**【指定代表圖】**

圖1A、1B

**【代表圖之符號簡單說明】**

無

## 【發明申請專利範圍】

### 【第1項】

一種分析生物體之生物樣品的方法，該生物樣品包括來源於正常細胞及可能來源於與癌症相關之細胞的核酸分子，其中該等核酸分子中至少一些在該生物樣品中為細胞游離核酸分子，該方法包含：

分析來自該生物樣品之複數個DNA分子，其中分析DNA分子包括：

測定該DNA分子在該生物體之基因組中之位置；

測定該DNA分子是否在一或多個位點處經甲基化；

對於複數個位點中之每一者：

測定在該位點處經甲基化之DNA分子之各別數目；

基於在該複數個位點處經甲基化之DNA分子之該等各別數目來計算第一甲基化程度；

比較該第一甲基化程度與第一截止值；

基於該比較測定癌症程度之第一分類。

### 【第2項】

如請求項1之方法，其中測定該DNA分子是否在一或多個位點處經甲基化包括：

進行可獲知甲基化之定序。

### 【第3項】

如請求項2之方法，其中進行可獲知甲基化之定序包括：

用亞硫酸氫鈉處理該等DNA分子；及

對該等經處理之DNA分子進行定序。

**【第4項】**

如請求項3之方法，其中使用亞硫酸氫鈉對該等細胞游離DNA分子之該處理為用於偵測5-羥甲基胞嘧啶之Tet輔助亞硫酸氫鹽轉化或氧化性亞硫酸氫鹽定序之一部分。

**【第5項】**

如請求項2之方法，其中各區域之DNA分子之該等各別數目係藉由比對獲自該可獲知甲基化之定序之序列讀段來測定。

**【第6項】**

如請求項1之方法，其中測定該DNA分子是否在一或多個位點處經甲基化包括在不進行亞硫酸氫鈉處理的情況下使用甲基化敏感性限制酶消化、甲基化特異性PCR、甲基化依賴性DNA沈澱、甲基化DNA結合蛋白/肽或單一分子定序。

**【第7項】**

如請求項1之方法，其進一步包含：

對於該基因組之第一複數個區域中之每一者：

測定來自該區域之DNA分子之各別數目；

由該各別數目計算各別標準化值；

比較該各別標準化值與參考值以確定該各別區域是否展現缺失或擴增；

測定被確定為展現缺失或擴增之區域的第一量；

比較該第一量與第一臨限值以確定癌症程度之第二分類；及

使用該第一分類及該第二分類以確定癌症程度之第三分類。

**【第8項】**

如請求項7之方法，其中該第一臨限值為被確定為展現缺失或擴增之該第一複數個區域之百分比。

**【第9項】**

如請求項7之方法，其中僅當該第一分類及該第二分類均指示癌症時，該第三分類為對癌症呈陽性。

**【第10項】**

如請求項7之方法，其中當該第一分類或該第二分類指示癌症時，該第三分類為對癌症呈陽性。

**【第11項】**

如請求項1或7之方法，其中該第一分類指示該生物體中存在癌症，該方法進一步包含：

藉由比較該第一甲基化程度與由其他生物體測定之相應值來鑑別與該生物體相關之癌症類型，其中該等其他生物體中之至少兩者被鑑別為患有不同類型癌症。

**【第12項】**

如請求項11之方法，其中該第三分類指示該生物體中存在癌症，該方法進一步包含：

藉由比較區域之該第一量與由該等其他生物體測定之相應值來鑑別與該生物體相關之癌症類型。

**【第13項】**

如請求項1或7之方法，其中計算該第一甲基化程度包括：

鑑別該基因組之第二複數個區域；

鑑別該等區域中之每一者內的一或多個位點；

計算各區域之區域甲基化程度，其中該第一甲基化程度係針對第一區域，

該方法進一步包含：

比較該等區域甲基化程度中之每一者與各別區域截止值，包括比較該第一甲基化程度與該第一截止值；

測定被確定為區域甲基化程度超過該各別區域截止值之區域的第二量；及

比較區域之該第二量與第二臨限值以確定該第一分類。

#### **【第14項】**

如請求項13之方法，其中被確定為區域甲基化程度超過該各別區域截止值之該等區域對應於第一區域集，該方法進一步包含：

比較該第一區域集之區域甲基化程度與其他生物體之該第一區域集之相應區域甲基化程度，該等其他生物體具有以下中之至少兩者：第一類型癌症、不存在癌症及第二類型癌症；及

基於該比較確定該生物體是否患有該第一類型癌症、不存在癌症或患有該第二類型癌症。

#### **【第15項】**

如請求項14之方法，其進一步包含：

基於該第一區域集之該等區域甲基化程度對該等其他生物體進行叢集分析，其中該等叢集中之兩者對應於以下中之任何兩者：該第一類型癌症、不存在癌症及該第二類型癌症，

其中該等區域甲基化程度之該比較確定該生物體屬於何種叢集。

#### **【第16項】**

如請求項15之方法，其中該等其他生物體之該叢集分析使用該生物體之該等區域甲基化程度。

**【第17項】**

如請求項15之方法，其中該等叢集包括對應於該第一類型癌症之第一叢集、對應於該第二癌症類型之第二叢集及對應於不存在癌症之第三叢集。

**【第18項】**

如請求項15之方法，其中該等其他生物體之該叢集分析進一步基於該等其他生物體之第二區域集之各別標準化值，該第二區域集對應於被確定為展現缺失或擴增之區域，其中區域之該各別標準化值係由來自該區域之DNA分子之各別數目確定，該方法進一步包含：

對於該第二區域集中之每一者：

測定來自該區域之DNA分子之各別數目；

由該各別數目計算各別標準化值；及

比較該生物體之該第二區域集之該等各別標準化值與該等其他生物體之該等各別標準化值作為確定該生物體屬於何種叢集之一部分。

**【第19項】**

如請求項18之方法，其中該等其他生物體之該叢集分析進一步基於高甲基化CpG島之各別甲基化密度，該方法進一步包含：

對於該等高甲基化CpG島中之每一者：

測定各別甲基化密度；

比較該生物體之該等高甲基化CpG島之該等各別甲基化密度與該

等其他生物體之該等甲基化密度作為確定該生物體屬於何種叢集之一部分。

**【第20項】**

如請求項13之方法，其進一步包含：

對於該第二複數個區域中之每一者：

計算該區域甲基化程度與該各別區域截止值之間的各別差異；

及

計算對應於該各別差異之各別機率；

其中測定區域之該第二量包括：

計算包括該等各別機率之累積評分。

**【第21項】**

如請求項20之方法，其中計算該累積評分包括：

獲取該各別機率之對數以獲得各別對數結果；及

計算包括該等各別對數結果之總和。

**【第22項】**

如請求項21之方法，其中該累積評分為該等各別對數結果之該總和之負數。

**【第23項】**

如請求項20之方法，其中各差異係用與該各別區域截止值相關之標準差標準化。

**【第24項】**

如請求項20之方法，其中根據統計分佈，該各別機率對應於該各別差異之機率。



**【第25項】**

如請求項20之方法，其中該第二臨限值對應於參考組之最高累積評分。

**【第26項】**

如請求項13之方法，其進一步包含：

對於該第一複數個區域中之每一者：

計算該各別標準化值與該參考值之間的各別差異；及

計算對應於該各別差異之各別機率；

其中測定區域之該第一量包括：

計算包括該等各別機率之第一總和。

**【第27項】**

如請求項13之方法，其中該等各別區域截止值為與參考甲基化程度相差之規定量。

**【第28項】**

如請求項13之方法，其中該第二臨限值為百分比，且其中比較區域之該第二量與第二臨限值包括：

在與該第二臨限值相比較之前，將區域之該第二量除以該第二複數個區域之第二數目。

**【第29項】**

如請求項28之方法，其中該第二數目對應於所有該第二複數個區域。

**【第30項】**

如請求項13之方法，其中該第一複數個區域與該第二複數個區域相

同，且其中該等各別區域截止值視該各別區域是否展現缺失或擴增而定。

**【第31項】**

如請求項30之方法，其中當該各別區域展現擴增時，與當未展現擴增時相比，各別區域截止值具有較大量值，且其中當該各別區域展現缺失時，與當未展現缺失時相比，各別區域截止值具有較小量值。

**【第32項】**

如請求項31之方法，其中各別區域截止值測試該第二複數個區域之低甲基化，其中當該各別區域展現擴增時，與當未展現擴增時相比，各別區域截止值具有較大負值，且其中當該各別區域展現缺失時，與當未展現缺失時相比，各別區域截止值具有較小負值。

**【第33項】**

如請求項13之方法，其中該生物樣品係在治療前獲取，該方法進一步包含：

對治療後獲取之另一生物樣品重複如請求項13之方法以獲得：

被確定為展現缺失或擴增之區域的隨後第一量；及

被確定為區域甲基化程度超過該各別區域截止值之區域的隨後第二量；

比較該第一量與該隨後第一量以及該第二量與該隨後第二量以確定該生物體之預後。

**【第34項】**

如請求項33之方法，其中比較該第一量與該隨後第一量以及該第二量與該隨後第二量以確定該生物體之預後包括：

確定該第一量與該隨後第一量之間的第一差異；

比較該第一差異與一或多個第一差異臨限值；  
確定該第二量與該隨後第二量之間的第二差異；及  
比較該第二差異與一或多個第二差異臨限值。

**【第35項】**

如請求項34之方法，其中當該第一差異低於一個第一差異臨限值時與當該第一差異高於該一個第一差異臨限值時相比該預後可預測為較糟，且其中當該第二差異低於一個第二差異臨限值時與當該第二差異高於該一個第二差異臨限值時相比該預後可預測為較糟。

**【第36項】**

如請求項35之方法，其中該一個第一差異臨限值及該一個第二差異臨限值為零。

**【第37項】**

如請求項33之方法，其中該治療為免疫療法、手術、放射線療法、化學療法、基於抗體之療法、表觀遺傳療法或靶向療法。

**【第38項】**

如請求項1之方法，其中該第一截止值為距由獲自健康生物體之生物樣品確定之參考甲基化程度之規定距離。

**【第39項】**

如請求項38之方法，其中該規定距離為距該參考甲基化程度之規定數目之標準差。

**【第40項】**

如請求項1之方法，其中該第一截止值係由該生物體之在測試該生物樣品前所獲得之先前生物樣品測定之參考甲基化程度確定。

**【第41項】**

如請求項1之方法，其中比較該第一甲基化程度與該第一截止值包括：

確定在該第一甲基化程度與參考甲基化程度之間的差異；及  
比較該差異與對應於該第一截止值之臨限值。

**【第42項】**

如請求項1之方法，其進一步包含：

測定該生物樣品中之腫瘤DNA之部分濃度；  
基於該部分濃度計算該第一截止值。

**【第43項】**

如請求項1之方法，其進一步包含：

確定該生物樣品中之腫瘤DNA之部分濃度是否大於最小值；及  
若該部分濃度不大於該最小值，則標記該生物樣品。

**【第44項】**

如請求項43之方法，其中該最小值係基於腫瘤之甲基化程度相對於參考甲基化程度之預期差異而確定。

**【第45項】**

如請求項1之方法，其進一步包含：

在該複數個位點處量測DNA分子之大小；及  
在比較該第一甲基化程度與該第一截止值之前，基於該等DNA分子之該所量測大小標準化該第一甲基化程度。

**【第46項】**

如請求項45之方法，其中基於該所量測大小標準化該第一甲基化程

度包括：

選擇具有第一大小之DNA分子；

使用該等所選DNA分子來計算該第一甲基化程度、對應於該第一大小之該第一截止值。

**【第47項】**

如請求項46之方法，其中該第一大小為一定範圍之長度。

**【第48項】**

如請求項46之方法，其中基於視大小而定之物理分離而選擇該等DNA分子。

**【第49項】**

如請求項46之方法，其中選擇具有第一大小之DNA分子包括：

對該複數個DNA分子進行配對末端大規模平行定序以獲得該等DNA分子中之每一者之序列對；

藉由比較該序列對與參考基因組來確定DNA分子之大小；及

選擇具有該第一大小之DNA分子。

**【第50項】**

如請求項45之方法，其中基於該所量測大小標準化該第一甲基化程度包括：

獲得大小與甲基化程度之間的函數關係；及

使用該函數關係來標準化該第一甲基化程度。

**【第51項】**

如請求項50之方法，其中該函數關係提供對應於各別大小之標度值。

**【第52項】**

如請求項51之方法，其進一步包含：

計算對應於用於計算該第一甲基化程度之DNA分子之平均大小；

及

將該第一甲基化程度乘以該相應標度值。

**【第53項】**

如請求項51之方法，其進一步包含：

對於該複數個位點中之每一者：

對於位於該位點處之該等DNA分子中之每一者：

獲得該位點處之該DNA分子之各別大小；及

使用對應於該各別大小之該標度值來標準化該DNA分子對該位點處經甲基化之DNA分子之該各別數目的貢獻。

**【第54項】**

如請求項1之方法，其中該複數個位點包括CpG位點，其中該等CpG位點組織成複數個CpG島，各CpG島包括一或多個CpG位點，其中該第一甲基化程度對應於第一CpG島。

**【第55項】**

如請求項54之方法，其中該等CpG島之平均甲基化密度小於參考組之第一百分比，且其中該參考組中之甲基化密度之變異係數小於第二百分比。

**【第56項】**

如請求項54之方法，其進一步包含：

對於該等CpG島中之每一者：

藉由比較該CpG島之甲基化程度與各別截止值來確定該CpG島相對於其他生物體之樣品之參考組是否經高甲基化；

對於該等高甲基化CpG島中之每一者：

測定各別甲基化密度；

由該等各別甲基化密度計算累積評分；及

比較該累積評分與累積截止值以確定該第一分類。

#### 【第57項】

如請求項56之方法，其中由該等各別甲基化密度計算該累積評分包括：

對於該等高甲基化CpG島中之每一者：

計算該各別甲基化密度與參考密度之間的各別差異；及

計算對應於該各別差異之各別機率；及

使用該等各別機率來確定該累積評分。

#### 【第58項】

如請求項56之方法，其中該累積評分係藉由以下方式測定：

獲取該各別機率之對數以獲得各別對數結果；及

計算包括該等各別對數結果之總和，該累積評分為該總和之負數。

#### 【第59項】

如請求項56之方法，其中各各別差異係用與該參考密度相關之標準差標準化。

#### 【第60項】

如請求項56之方法，其中該累積截止值對應於該參考組之最高累積

評分。

**【第61項】**

如請求項56之方法，其中確定該第一CpG島是否經高甲基化包括：

比較該第一甲基化程度與該第一截止值及第三截止值，

其中該第一截止值對應於該參考組之甲基化密度平均值加規定百分比，且其中該第三截止值對應於規定數目之標準差加該參考組之甲基化密度之平均值。

**【第62項】**

如請求項61之方法，其中該規定百分比為2%。

**【第63項】**

如請求項61之方法，其中該標準差之規定數目為三。

**【第64項】**

如請求項1之方法，其進一步包含：

對於該基因組之第一複數個區域中之每一者：

測定來自該區域之DNA分子之各別數目；

由該各別數目計算各別標準化值；

比較該各別標準化值與參考值以確定該各別區域是否展現缺失或擴增；

測定第一區域集，該等區域被確定為均展現以下中之一者：缺失、擴增或正常呈現，其中該第一甲基化程度對應於該第一區域集；

測定第二區域集，該等區域被確定為均展現以下中之另一者：缺失、擴增或正常呈現；



基於該第二區域集中之位點處甲基化之DNA分子之該等各別數目來計算第二甲基化程度，

其中比較該第一甲基化程度與該第一截止值包括：

計算該第一甲基化程度與該第二甲基化之間的參數；及  
比較該參數與該第一截止值。

**【第65項】**

如請求項64之方法，其中該第一甲基化程度為針對該第一區域集之各區域計算之區域甲基化程度之統計值，且其中該第二甲基化程度為針對該第二區域集之各區域計算之區域甲基化程度之統計值。

**【第66項】**

如請求項65之方法，其中使用史都登氏t檢驗(Student's t-test)、方差分析(ANOVA)檢驗或克魯斯凱-沃利斯檢驗(Kruskal-Wallis test)來確定該等統計值。

**【第67項】**

如請求項64之方法，其中該參數包括比率或差異。

**【第68項】**

如請求項67之方法，其中計算該參數包括將機率分佈應用於該比率或該差異。

**【第69項】**

一種由生物體之生物樣品測定第一甲基化概況之方法，該生物樣品包括包含來源於第一組織及第二組織之細胞游離DNA分子之混合物的細胞游離DNA，該方法包含：

獲得對應於該第二組織之DNA分子之第二甲基化概況，該第二甲

基化概況提供該生物體之基因組中之複數個基因座中之每一者處的甲基化密度，特定基因座處之該甲基化密度對應於該第二組織之經甲基化之DNA分子之比例；

由該混合物之該等細胞游離DNA分子測定細胞游離甲基化概況，該細胞游離甲基化概況提供該複數個基因座中之每一者處的甲基化密度；

測定該混合物中來自該第一組織之該等細胞游離DNA分子之百分比；及

藉由以下方式測定該第一組織之該第一甲基化概況：

對於該複數個基因座中之每一者：

計算差異性參數，該參數包括該第二甲基化概況之該甲基化密度與該細胞游離甲基化概況之該甲基化密度之間的差異，該差異係由百分比定標。

#### 【第70項】

如請求項69之方法，其進一步包含：

由該第一甲基化概況確定第一甲基化程度；及

比較該第一甲基化程度與第一截止值以基於該比較確定癌症程度之分類。

#### 【第71項】

如請求項69之方法，其進一步包含：

轉換該第一甲基化概況以獲得經修正之第一甲基化概況。

#### 【第72項】

如請求項71之方法，其中該轉換為線性轉換。

**【第73項】**

如請求項69之方法，其中基因座之該差異性參數D定義為 $D = mbc - \frac{(mbc-mp)}{f * CN}$ ，其中mbc表示該基因座處之該第二甲基化概況之該甲基化密度，mp表示該基因座處之該細胞游離甲基化概況之該甲基化密度，f為該生物樣品中來自該第一組織之細胞游離DNA分子之百分比，且CN表示該基因座處之複本數。

**【第74項】**

如請求項73之方法，其中CN為該基因座處為二體性之該第一組織之複本數。

**【第75項】**

如請求項73之方法，其進一步包含：  
鑑別D超過臨限值之區域。

**【第76項】**

如請求項69之方法，其進一步包含：

藉由選擇具有以下準則中之任何一或多者的基因座來鑑別該複數個基因座：

GC含量大於50%；

該第二甲基化概況之甲基化密度小於第一臨限值或大於第二臨限值；及

最少五個CpG位點在由該基因座所定義之區域中。

**【第77項】**

如請求項69之方法，其中該生物樣品係來自懷有胎兒之女性個體，且其中該第一組織係來自該胎兒或胎盤，且該第二組織來自該女性個體。

**【第78項】**

如請求項77之方法，其進一步包含：

由該第一甲基化概況確定第一甲基化程度；及

比較該第一甲基化程度與第一截止值以確定醫學病狀之分類。

**【第79項】**

如請求項78之方法，其中該醫學病狀為子癩先兆或該第一組織中之染色體異常。

**【第80項】**

如請求項69之方法，其中該第一組織係來自該生物體內之腫瘤且該第二組織係來自該生物體之非惡性組織。

**【第81項】**

如請求項69之方法，其中該生物樣品亦包括來自該第二組織之DNA分子，且其中獲得該第二甲基化概況包括：

分析來自該第二組織之該等DNA分子以確定該第二甲基化概況。

**【第82項】**

如請求項69之方法，其中該所獲得之第二甲基化概況對應於獲自對照樣品組之平均甲基化概況。

**【第83項】**

如請求項69之方法，其中確定該複數個基因座之第一基因座之該細胞游離甲基化概況之甲基化密度包括：

確定來自該第一基因座之該等細胞游離DNA分子之第一數目；

確定該第一數目之各DNA分子在一或多個位點處是否甲基化以獲得甲基化DNA分子之第二數目；及

由該第一數目及該第二數目計算該甲基化密度。

**【第84項】**

如請求項83之方法，其中該一或多個位點為CpG位點。

**【第85項】**

如請求項69之方法，其中確定該細胞游離甲基化概況包括：

用亞硫酸氫鈉處理該等細胞游離DNA分子；及

對該等經處理之DNA分子進行定序。

**【第86項】**

如請求項85之方法，其中對該等細胞游離DNA分子之該處理為Tet輔助亞硫酸氫鹽轉化之一部分，或包括用高鈳酸鉀(KRuO<sub>4</sub>)處理該等細胞游離DNA分子。

**【第87項】**

如請求項69之方法，其中確定該細胞游離甲基化概況包括：

對該DNA進行單一分子定序，其中該單一分子定序包括確定DNA分子之至少一個位點是否經甲基化。

**【第88項】**

一種電腦產品，其包含非暫時式電腦可讀媒體，該非暫時式電腦可讀媒體儲存複數個當被執行時控制電腦系統執行如請求項1之方法的指令。

**【第89項】**

一種由生物體之生物樣品測定第一甲基化概況之方法，該生物樣品包括包含來源於第一組織及第二組織之細胞游離DNA之混合物的細胞游離DNA，該方法包含：

分析來自該生物樣品之複數個DNA分子，其中分析DNA分子包括：

測定該DNA分子在該生物體之基因組中之位置；

測定該DNA分子之基因型；

測定該DNA分子是否在一或多個位點處經甲基化；

鑑別複數個第一基因座，在該複數個第一基因座處該第一組織之第一基因組對於各別第一等位基因及各別第二等位基因為異型組合且該第二組織之第二基因組對於該各別第一等位基因為同型組合；

對於該等第一基因座中之每一者：

對於與該基因座相關之一或多個位點中之每一者：

測定在該位點處經甲基化且對應於該基因座之該各別第二等位基因之DNA分子之數目；

基於該基因座之該一或多個位點處經甲基化且對應於該基因座之該各別第二等位基因之DNA分子之數目計算甲基化密度；

由該第一基因座之該等甲基化密度產生該第一組織之該第一甲基化概況。

#### 【第90項】

如請求項89之方法，其中該生物樣品係來自懷有胎兒之女性個體，且其中該第一組織係來自該胎兒或胎盤，且該第二組織對應於該女性個體。

#### 【第91項】

如請求項89之方法，其中該等位點對應於CpG位點。

#### 【第92項】

如請求項89之方法，其中各基因座包括至少一個CpG位點。

**【第93項】**

如請求項89之方法，其中各基因座與至少一個CpG位點相鄰。

**【第94項】**

一種用於分析含有胎兒DNA與來自懷有胎兒之女性個體之DNA之混合物的生物樣品中之胎兒DNA的套組，該套組包含：

一或多個與表S2A、S2B、S2C、S2D、S3A、S3B、S3C及S3D中所列出之基因組區域之至少一個區段特異性雜交的寡核苷酸。

**【第95項】**

如請求項94之套組，其中該等寡核苷酸中之至少一者為對應於該等基因組區域中之一者之至少一個區段的引子。

**【第96項】**

如請求項94之套組，其中該等寡核苷酸中之至少一者為對應於該等基因組區域中之一者之至少一個區段的探針。

**【第97項】**

如請求項94之套組，其進一步包含：

一或多種甲基化敏感性限制酶。

**【第98項】**

如請求項94之套組，其中該一或多種甲基化敏感性限制酶中之至少一者切割低甲基化DNA分子。

**【第99項】**

如請求項94之套組，其中該一或多種甲基化敏感性限制酶中之至少一者切割高甲基化DNA分子。

**【第100項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自表S2A。

**【第101項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自表S2B。

**【第102項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自表S2C。

**【第103項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自表S2D。

**【第104項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自表S3A。

**【第105項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自表S3B。

**【第106項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自表S3C。

**【第107項】**

如請求項94之套組，其中該一或多種寡核苷酸中之至少一者係選自



表S3D。

**【第108項】**

一種評估生物體之生物樣品中之DNA之甲基化程度的方法，該方法包含：

對於複數個大小中之每一大小：

量測該生物樣品中對應於該大小之複數個DNA片段之量；

用電腦系統基於多個大小之DNA片段之量來計算第一參數之第一值，該第一參數提供該生物樣品中之DNA片段之大小概況的統計度量；

比較該第一大小值與參考大小值，該參考大小值係由參考樣品之DNA片段計算；及

基於該比較評估該甲基化程度。

**【第109項】**

如請求項108之方法，其中量測對應於該等大小之DNA片段之量包括：

對於來自該生物樣品之複數個DNA片段中之每一者：

量測該DNA片段之大小，

其中量測DNA片段之大小包括：

進行該DNA片段之配對末端定序以獲得配對序列讀段；

比對該等配對序列讀段與參考基因組；及

使用該等經比對之位置來確定該DNA片段之大小。

**【第110項】**

如請求項109之方法，其中該甲基化程度藉由測定僅來自與特定區域

比對之DNA片段之該第一大小值而對應於該特定區域。

**【第111項】**

如請求項108之方法，其中量測對應於該等大小之DNA片段之量包括使用電泳。

**【第112項】**

如請求項108之方法，其中該參考樣品係在比該生物樣品早之時間獲自該生物體，其中該所評估之甲基化程度為對應於該第一大小值相對於該參考大小值之增加或降低量的相對程度。

**【第113項】**

如請求項108之方法，其中該參考大小值為校準曲線之一部分。

**【第114項】**

如請求項108之方法，其中該甲基化程度為甲基化密度。

**【第115項】**

如請求項114之方法，其中該甲基化密度係針對CpG位點。

**【第116項】**

一種偵測生物體之生物樣品之染色體異常的方法，該生物樣品包括包含來源於第一組織及第二組織之細胞游離DNA之混合物的細胞游離DNA，該方法包含：

分析來自該生物樣品之複數個DNA分子，其中分析DNA分子包括：

測定該DNA分子在參考基因組中之位置；

測定該DNA分子是否在一或多個位點處經甲基化；

對於複數個位點中之每一者：

測定在該位點處經甲基化之DNA分子之各別數目；

基於第一染色體區域內之位點處之經甲基化DNA分子之該等各別數目來計算該第一染色體區域之第一甲基化程度；

比較該第一甲基化程度與截止值；

基於該比較確定該第一組織中該第一染色體區域之異常分類。

**【第117項】**

如請求項116之方法，其中比較該第一甲基化程度與截止值包括：

標準化該第一甲基化程度及比較該標準化第一甲基化程度與該截止值。

**【第118項】**

如請求項117之方法，其中該標準化使用第二染色體區域之第二甲基化程度。

**【第119項】**

如請求項117之方法，其中該標準化使用來自該第一組織之細胞游離DNA之部分濃度。

**【第120項】**

如請求項116之方法，其中計算第一甲基化程度包括：

鑑別該生物體之基因組之複數個區域；

鑑別該等區域中之每一者內的一或多個位點；

計算各區域之區域甲基化程度，其中該第一甲基化程度係針對第一區域，及

其中比較該第一甲基化程度與第一截止值包括：

比較該等區域甲基化程度中之每一者與各別區域截止值；

確定區域甲基化程度超過該各別區域截止值之區域的第一數目；及

比較該第一數目與臨限值以確定該分類。

**【第121項】**

如請求項120之方法，其中該臨限值為百分比，且其中比較該第一數目與臨限值包括：

在與該臨限值比較之前將區域之該第一數目除以區域之第二數目。

**【第122項】**

如請求項121之方法，其中區域之該第二數目為所鑑別之該複數個區域的全部。

**【第123項】**

如請求項120之方法，其中該等各別區域截止值為與參考甲基化程度相差之規定量。

**【第124項】**

如請求項116之方法，其中該生物樣品係來自懷有胎兒之女性個體，其中該第一組織係來自該胎兒或胎盤，且該第二組織來自該女性個體，且其中該染色體異常為胎兒染色體異常。

**【第125項】**

如請求項116之方法，其中該截止值係基於與來自該女性個體之該第二組織之DNA相關之背景甲基化程度。

**【第126項】**

如請求項124之方法，其中該截止值係由懷有該第一染色體區域中無

染色體異常之胎兒之其他女性妊娠個體測定。

**【第127項】**

如請求項124之方法，其中該染色體異常為第21對染色體三體症或第18對染色體三體症或第13對染色體三體症或特納氏症候群(Turner syndrome)或克萊恩費特氏症候群(Klinefelter syndrome)。

**【第128項】**

如請求項116之方法，其中該第一組織係來自該生物體內之腫瘤且該第二組織係來自該生物體之非惡性組織。

**【第129項】**

如請求項116之方法，其中該截止值係由未患癌症之其他生物體測定。

**【第130項】**

如請求項116之方法，其中截止值係基於該生物樣品中來源於該第一組織之細胞游離DNA之濃度。

**【第131項】**

如請求項130之方法，其中該截止值係基於對應於異常類型之標度因子，其中該異常類型為缺失或複製。

**【第132項】**

如請求項116之方法，其中確定該DNA分子在參考基因組中之位置涉及確定該位置是否在該第一染色體區域內。

**【第133項】**

如請求項132之方法，其中確定該DNA分子在該參考基因組中之位置係藉由確定該DNA分子是否定位於該第一染色體區域來實現。

**【第134項】**

如請求項116之方法，其中該染色體異常為亞染色體缺失或亞染色體複製或狄喬治氏症候群(DiGeorge syndrome)。



























































































































































## 【發明申請專利範圍】

### 【第1項】

一種藉由分析來自生物體之生物樣品之無細胞(cell-free)DNA分子以確定癌症類型的方法，該方法包含：

(a) 從生物樣品之無細胞DNA分子獲得序列讀段，其中該序列讀段包含該無細胞DNA分子之甲基化狀態；

(b) 分析該序列讀段以確定，在單一核苷酸解析度下，位於該無細胞DNA分子之複數個位點之甲基化狀態；

(c) 確定來自該複數個位點之甲基化狀態的甲基化概況；及

(d) 至少部分地基於該甲基化概況確定癌症類型。

### 【第2項】

如請求項1之方法，其中該生物樣品係選自由血液、血漿及血清組成之群。

### 【第3項】

如請求項1之方法，其中該生物樣品係選自由尿液、陰道液、子宮或陰道沖洗液、胸腔積液、腹水、腦脊液、唾液、汗液、淚液、痰液、支氣管肺泡灌洗液及大便組成之群。

### 【第4項】

如請求項1至3中任一項之方法，其中該癌症類型包含已罹患癌症之器官。

### 【第5項】

如請求項1至3中任一項之方法，其中該癌症類型包含癌症來源之組織。

**【第6項】**

如請求項5之方法，其中該來源之組織係選自由腦、骨、肺、心臟、腎臟、肝臟、乳房、結腸、前列腺、鼻咽、胃、睪丸、皮膚、卵巢、胰臟、子宮及淋巴組成之群。

**【第7項】**

如請求項1至3中任一項之方法，其中該癌症類型係選自由肺癌、乳癌、結腸直腸癌、前列腺癌、鼻咽癌、胃癌、睪丸癌、皮膚癌、影響神經系統之癌症、骨癌、卵巢癌、肝癌、血液科惡性疾病、胰臟癌、子宮內膜癌及腎癌組成之群。

**【第8項】**

如請求項1至3中任一項之方法，其進一步包含定序該無細胞DNA分子以獲得該序列讀段。

**【第9項】**

如請求項8之方法，其中該定序包含可獲知甲基化之定序(methylation-aware sequencing)。

**【第10項】**

如請求項9之方法，其中該可獲知甲基化之定序包含亞硫酸氫鹽定序。

**【第11項】**

如請求項10之方法，其中該亞硫酸氫鹽定序包含全基因組亞硫酸氫鹽定序。

**【第12項】**

如請求項9之方法，其中該可獲知甲基化之定序包含配對末端大規模

平行定序或完整基因組定序。

**【第13項】**

如請求項8之方法，其進一步包含在定序前增濃(enriching)該無細胞DNA分子。

**【第14項】**

如請求項13之方法，其中該增濃包含使用雜交探針、聚合酶鏈式反應或固體相雜交。

**【第15項】**

如請求項1至3中任一項之方法，其中該複數個位點包含一或多個CpG位點。

**【第16項】**

如請求項1至3中任一項之方法，其中該複數個位點包含一或多個CpG位點，其被組織成一或多個CpG島。

**【第17項】**

如請求項16之方法，其中該確定複數個位點之甲基化概況包含使用該甲基化狀態計算該一或多個CpG島之甲基化密度，該甲基化狀態由位於該一或多個CpG島之序列讀段所確定。

**【第18項】**

如請求項1至3中任一項之方法，其中該確定複數個位點之甲基化概況包含使用該甲基化狀態計算該複數個位點之個別位點之甲基化密度，該甲基化狀態由位於該位點之序列讀段所確定。

**【第19項】**

如請求項1至3中任一項之方法，其中該分析進一步包含確定該無細

胞DNA分子在基因組中之位置。

**【第20項】**

如請求項19之方法，其中該確定甲基化概況包含使用該甲基化狀態確定在一或多個基因組區域之一或多個甲基化程度，該甲基化狀態由位於該一或多個基因組區域之該無細胞DNA分子之序列讀段所確定。

**【第21項】**

如請求項20之方法，其中該一或多個基因組區域包含複數個規定於基因組中之基因組區間(bins)。

**【第22項】**

如請求項21之方法，其中該複數個基因組區間具有50 kb至1 Mb之大小。

**【第23項】**

如請求項21之方法，其中該複數個基因組區間具有不同大小。

**【第24項】**

如請求項21之方法，其進一步包含藉由比較該基因組區間之每一者之甲基化程度與截止值，選擇基因組區間群組。

**【第25項】**

如請求項24之方法，其進一步包含至少部分地基於該選擇之基因組區間群組之甲基化程度確定癌症類型。

**【第26項】**

如請求項25之方法，其中該截止值由已知患有癌症之個體所確定。

**【第27項】**

如請求項25之方法，其中該截止值由健康個體所確定。

**【第28項】**

如請求項1至3中任一項之方法，其中該確定癌症類型包含比較該甲基化概況與參考甲基化概況。

**【第29項】**

如請求項28之方法，其中該參考甲基化概況對應於健康個體。

**【第30項】**

如請求項28之方法，其中該參考甲基化概況係由已知患有癌症之個體所獲得之至少一個樣品所獲得。

**【第31項】**

如請求項1至3中任一項之方法，其進一步包含基於該甲基化概況確定存在或不存在複本數畸變(copy number aberration)。

**【第32項】**

如請求項1至3中任一項之方法，其進一步包含確定無細胞DNA分子中位於一或多個基因組位點之一或多個甲基化程度。

**【第33項】**

如請求項1至3中任一項之方法，其中確定該甲基化概況包含確定一或多個CpG島之一或多個甲基化程度。

**【第34項】**

一種分析來自個體之生物樣品的方法，該方法包含：

(a) 從生物樣品之無細胞DNA分子獲得序列讀段，其中該序列讀段包含該無細胞DNA分子之甲基化狀態；

(b) 基於複數個位點之甲基化狀態，分析該序列讀段以確定該複數個位點之甲基化概況；及

(c) 至少部分地基於該甲基化概況，確定來自該生物樣品之至少一部分之無細胞DNA分子之組織來源。

**【第35項】**

如請求項34之方法，其中至少一部分之無細胞DNA分子為源於癌症之分子，該方法進一步包含至少部分地基於該甲基化概況，確定該個體之癌症類型。

**【第36項】**

如請求項35之方法，其中該癌症類型係選自由肺癌、乳癌、結腸直腸癌、前列腺癌、鼻咽癌、胃癌、睪丸癌、皮膚癌、影響神經系統之癌症、骨癌、卵巢癌、肝癌、血液科惡性疾病、胰臟癌、子宮內膜癌及腎癌組成之群。

**【第37項】**

如請求項34之方法，其進一步包含定序該無細胞DNA分子以獲得該序列讀段。

**【第38項】**

如請求項37之方法，其中該定序包含可獲知甲基化之定序。

**【第39項】**

如請求項38之方法，其中該可獲知甲基化之定序包含亞硫酸氫鹽定序。

**【第40項】**

如請求項39之方法，其中該複數個位點包含一或多個CpG位點。

**【第41項】**

一種電腦產品，其包含電腦可讀媒體，該電腦可讀媒體儲存複數個



當被執行時控制電腦系統以執行如請求項1至40中任一項之方法的指令，該指令包含如請求項1至40中任一項之步驟。

**【第42項】**

一種電腦系統，其包含一或多個處理器，其經程式化以實現如請求項1至40中任一項之方法。