

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号

特開2024-105502

(P2024-105502A)

(43)公開日 令和6年8月6日(2024.8.6)

(51)国際特許分類

F I

C 1 2 Q 1/6869(2018.01)

C 1 2 Q 1/6869

Z Z N A

C 1 2 Q 1/6837(2018.01)

C 1 2 Q 1/6837

Z

C 1 2 N 15/12 (2006.01)

C 1 2 Q 1/6869

Z

C 1 2 N 15/12

審査請求 有 請求項の数 3 O L (全80頁)

(21)出願番号 特願2024-77519(P2024-77519)
 (22)出願日 令和6年5月10日(2024.5.10)
 (62)分割の表示 特願2020-567152(P2020-567152)
)の分割
 原出願日 令和1年7月9日(2019.7.9)
 (31)優先権主張番号 62/695,429
 (32)優先日 平成30年7月9日(2018.7.9)
 (33)優先権主張国・地域又は機関 米国(US)

(71)出願人 520465839
 エイチケージー エピセラピューティクス
 リミテッド
 HKG EPITHERAPEUTICS
 LIMITED
 中国, 香港, 九龍, 尖沙咀, 広東道 3 0
 號, 塔 1, 新港中心 8 1 2
 8 1 2 Silvercord, Tow
 er 1, 30 Canton Road
 Tsimshatsui, Kowloo
 n, Hong Kong, China
 (74)代理人 110001999
 弁理士法人はなぶさ特許商標事務所
 (72)発明者 チェイシビリ, デビッド
 カナダ国, H 9 K 1 N 2, ケベック, ピ
 最終頁に続く

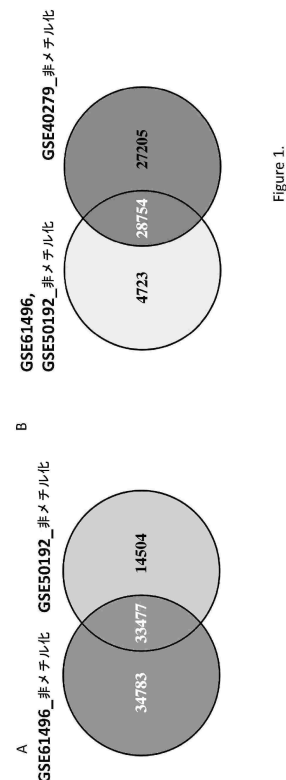
(54)【発明の名称】 がんの非侵襲的検出のためのDNAメチル化マーカーとその使用

(57)【要約】 (修正有)

【課題】がんの非侵襲的検出のためのDNAメチル化マーカーとその使用を提供する。

【解決手段】患者由来の生体物質中のDNAにあるがんを検出し、他の組織の無細胞DNAおよび血球DNAと区別するために、ヒトゲノム(CGID)内の少数の精巧なDNAメチル化位置の組み合わせを見つけるための「バイナリーカテゴリー区別」法である。腫瘍DNAの起源の組織を検出する別の方法は、ヒトゲノム(CGID)の固有のDNAメチル化位置の組み合わせを使用する。腫瘍DNAに由来するCGIDの組み合わせは、がんを正確に検出するために、特定のCGIDの組み合わせのDNAメチル化を測定し、「メチル化スコア」を導出することにより、開示されている。少量の血漿から、多重化された次世代シーケンスメチル化アッセイ、パイロシーケンスアッセイそしてメチル化特異的PCRを使用して、CGIDを使用してがんを予測するためのキットである。

【選択図】図1



【特許請求の範囲】

【請求項 1】

「バイナリーカテゴリー区別 (BCD)」法を使用してゲノムワイドDNAメチル化マップからバイナリーカテゴリーDNAメチル化マーカーを導き出すステップを含む、がんの検出の補助手段として、がんのバイナリーカテゴリーDNAメチル化マーカーを使用する方法。

【請求項 2】

前記ゲノムワイドDNAメチル化マップが、がん細胞、正常な組織および血液DNAの1つ以上である、請求項1に記載の方法。

【請求項 3】

前記「バイナリーカテゴリー区別 (BCD)」法が、ゲノムワイドDNAメチル化アレイ、DNAマイクロアレイ、ゲノムワイドバイサルファイトシーケンス、メチル化DNA免疫沈降 (MeDIP) シーケンス、およびオリゴヌクレオチドアレイとのハイブリダイゼーションの1つ以上の使用を含む、請求項1に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ヒトDNA特に分子診断の分野におけるDNAメチル化シグネチャに関する。

【背景技術】

【0002】

がんは人間の主要な死因になっている。がんの早期発見は、治癒率を大幅に改善し、患者とその家族および医療システムへの甚大な個人的および経済的コストを削減できる。例えば、肝細胞がん (HCC) は、世界で5番目に多いがんである (El-Serag、2011年)。これは特にアジアで蔓延しており、その発生率はB型肝炎が蔓延している地域で最も高く、因果関係の可能性を示している (Flores & Marrero、2014年)。慢性肝炎患者などの高リスク集団のフォローアップおよび慢性肝炎からHCCへの移行の早期診断は、治癒率を改善するだろう。肝細胞がんの生存率は、ほとんどの場合末期に診断されるため、現在非常に低い。早期に診断されれば、肝臓がんは80%を超える治癒率で効果的に治療できる。イメージングの進歩により、HCCの非侵襲的検出が改善された (Tan、Low、& Thng、2011; Valente et al.、2014)。しかし、 α -フェトプロテインなどの単一タンパク質を用いたイメージングや免疫アッセイを含む現在の診断方法では、HCCを早期に診断できないことがよくある (Flores & Marrero、2014年)。これらの課題はHCCに限定されず、他のがんにも共通している。例えば、乳がんと大腸がんの早期発見は、罹患率と死亡率、および公衆衛生システムと保険会社のコストを劇的に削減する可能性がある。さらに、膵臓がんなどの特定のがんは、ほぼ常に遅れて検出され、その結果、事実上一定の死亡率を生じる。イメージングの進歩によりがんの早期発見が改善されているが、MRIなどの高解像度イメージングは高価であり、高度な訓練を受けた人材が必要であり、多くの場所では利用できない。それはまだ多数の集団をスクリーニングする方法に進化していない。がんによる罹患率と死亡率の低減に影響を与えるには、集団の定期的なスクリーニングのために広い地理的領域で使用できる、非侵襲的で強力にもかかわらず低コストの方法を開発する必要がある。主な課題は、固形腫瘍が臓器に隠れ、臨床症状を示さずと前に進化することである。しかしながら、腫瘍材料を非侵襲的に得ることは可能である。

【0003】

腫瘍DNAがシステムに放出され、血漿で見つけられ得ること (Warton & Samimi、2015)、そして尿や唾液などのその他の分泌された体液や糞中に見つかる可能性があることは、今では広く確立されている。腫瘍DNAの分子特性を測定することにより、体液に見つかったDNAが腫瘍に由来することを確認できる (Zhai et al.、2012)。腫瘍細胞は腫瘍DNAと正常細胞DNAを区別できる突然変異を発生

10

20

30

40

50

させるが、起こり得る突然変異の数は膨大であり、すべての腫瘍で一般的な突然変異が起こるわけではない (Dominguez - Vigil, Moreno - Martinez, Wang, Roehrl, & Barrera - Saldana, 2018)。

【0004】

DNAの共有結合修飾であるDNAメチル化は、ゲノム機能のエピジェネティック制御の主要なメカニズムであり、腫瘍内で遍在的に変化している (Aguirre - Ghiso, 2007; Baylin et al., 2001; Ehrlich, 2002; Issa et al., 1993)。腫瘍のDNAメチル化プロファイルは、腫瘍の分類、予後、および化学療法に対する反応の予測のための強力なツールとなる可能性がある (Stefansson et al., 2014)。早期診断で腫瘍DNAメチル化を使用することの主な欠点は、疑わしい腫瘍の侵襲的手順と解剖学的視覚化が必要なことである。循環腫瘍細胞は非侵襲的な腫瘍DNAの供給源であり、腫瘍抑制遺伝子のDNAメチル化を測定するために使用される (Radpour et al., 2011)。HCC

10

DNAの低メチル化は患者の血液で検出可能であり (Ramzy, Omran, Hamad, Shaker, & Abboud, 2011)、最近ゲノムワイドバイサルファイトシーケンス (bisulfite sequencing) が、HCC患者の血漿中の低メチル化DNAを検出するために適用された (Chan et al., 2013)。ただし、特にがんの初期段階では、この情報源は限られており、DNAメチル化プロファイルは宿主 (host) のDNAメチル化プロファイルによって混同されている。ゲノムワイドバイサルファイトシーケンスは比較的費用のかかる手順であり、重要なバイオインフォマティクス分析を必要とするため、スクリーニングツールとしては実行不可能である。したがって、課題は、腫瘍DNAと非腫瘍DNAを確実に区別できる少数のCGを描き、広範で多様な地理的領域の幅広い集団のスクリーニングを可能にする低コストハイスループットアッセイを開発することである。最近、いくつかのグループが、がんと正常なDNAおよび血中DNAのゲノムワイドなDNAメチル化マップの比較分析を実施した (Zhai et al., 2012)。ただし、これらのアプローチの主な課題は、先験的に予期しないさまざまなレベルで血液中に見られる他の組織からの無細胞DNAを考慮に入れていないことである。がん組織と同様のメチル化プロファイルを持つ別の組織からのDNAの混入は、偽陽性を引き起こす可能性がある。さらに、過去のアプローチでは、正常組織とがん組織におけるDNAメチル化を定量的に比較している。この定量的な違いは、腫瘍DNAが他の非形質転換組織からの異なる未知の量のDNAと混合されると希釈され、偽陰性を引き起こす可能性がある。現在の方法におけるこれらの欠陥は、本発明の主題において開示される異なるアプローチを必要とする。

20

30

【0005】

がんを検出するためのシステムと方法の使用に関連するさらなる出版物は次のとおりである: Grigg G, Clark S. ゲノムDNAの5 - メチルシトシン残基のシーケンス. *Bioessays*. 1994年6月; 16(6): 431 - 6, 431; Zeschneigk M, Schmitz B, Ditttrich B, Buiting K, Horsthemke B, Doerfler W. ヒトのゲノムにインプリントされたセグメント: ゲノムシーケンス法で決定されたプラダー・ウィリー/アンジェルマン症候群領域の異なるDNAメチル化パターン. *Hum Mol Genet*. 1997年3月; 6(3): 387 - 95; Feil R, Charlton J, Bird A P, Walter J, Reik W. 個々の染色体のメチル化分析: バイサルファイトゲノムシーケンスのための改良されたプロトコル. *Nucleic Acids Res*.

40

1994年2月25日; 22(4): 695 - 6; Martin V, Ribieras S, Song - Wang X, Rio M C, Dante R. ゲノムシーケンスは、p52遺伝子の5'領域のDNA低メチル化とヒト乳がん細胞株におけるその発現との相関を示す. *Gene*. 1995年5月19日; 157(1 - 2): 261 - 4; WO 97 46705、WO 95 15373およびWO 45560。

Aguirre - Ghiso, J. A. (2007)。がん休眠のモデル、メカニズム

50

- 、および臨床的証拠。Nat Rev Cancer、7(11)、834 - 846。doi:10.1038/nrc2256
- Baylin, S. B., Esteller, M., Rountree, M. R., Bachman, K. E., Schuebel, K., & Herman, J. G. (2001)。がんにおけるDNAメチル化、クロマチン形成および遺伝子発現の異常なパターン。Hum Mol Genet、10(7)、687 - 692。
- Breitbach, S., Tug, S., Helmig, S., Zahn, D., Kubiak, T., Michal, M., . . . Simon, P. (2014)。未精製血漿からの無細胞循環DNAの直接定量。PLOS One、9(3)、e87838。doi:10.1371/journal.pone.0087838 10
- Chan, K. C., Jiang, P., Chan, C. W., Sun, K., Wong, J., Hui, E. P., . . . Lo, Y. M. (2013)。血漿DNAパイサルファイトシーケンスによるがん関連ゲノムワイドの低メチル化とコピー数異常の非侵襲的検出。Proc Natl Acad Sci U S A、110(47)、18761 - 18768。doi:10.1073/pnas.1313995110
- Dominguez-Vigil, I. G., Moreno-Martinez, A. K., Wang, J. Y., Roehrl, M. H. A., & Barrera-Saldana, H. A. (2018)。がんとの闘いにおけるリキッドバイオプシーの夜明け。Oncotarget、9(2)、2912 - 2922。doi:10.18632/oncotarget.23131 20
- Ehrlich, M. (2002)。がんにおけるDNAメチル化：多すぎるが少なすぎる。Oncogene、21(35)、5400 - 5413。
- El-Serag, H. B. (2011)。肝細胞がん。N Engl J Med、365(12)、1118 - 1127。doi:10.1056/NEJMra1001683
- Flores, A., & Marrero, J. A. (2014)。肝細胞がんの最新動向：診断と治療に焦点を当てる。Clin Med Insights Oncol、8、71 - 76。doi:10.4137/CMO.S9926
- Issa, J. P., Vertino, P. M., Wu, J., Sazawal, S., Celano, P., Nelkin, B. D., . . . Baylin, S. B. (1993)。結腸がんの進行中のシトシンDNA - メチルトランスフェラーゼ活性の増加。J Natl Cancer Inst、85(15)、1235 - 1240。 30
- Luczak, M. W., & Jagodzinski, P. P. (2006)。がん進行におけるDNAメチル化の役割。Folia Histochem Cytobiol、44(3)、143 - 154
- Radpour, R., Barekati, Z., Kohler, C., Lv, Q., Burki, N., Diesch, C., . . . Zhong, X. Y. (2011)。乳がんの血液検査を開発するための重要な調節経路に關与する腫瘍抑制遺伝子の高メチル化。PLOS One、6(1)、e16080。doi:10.1371/journal.pone.0016080 40
- Ramzy, I I, Omran, D. A., Hamad, O., Shaker, O., & Abboud, A. (2011)。肝細胞がんの予後マーカーとしての血清LINE - 1低メチル化の評価。Arab J Gastroenterol、12(3)、139 - 142。doi:10.1016/j.ajg.2011.07.002
- Stefansson, O. A., Moran, S., Gomez, A., Sayols, S., Arribas-Jorba, C., Sandoval, J., . . . Esteller, M. (2014)。生物学的に区別される乳がんのサブタイプのDNAメチル化に基づく定義。Mol Oncol。doi:10.1016/j.molonc.2014.10.012
- Tan, C. H., Low, S. C., & Thng, C. H. (2011)。肝細胞が 50

んの画像診断に関するAPASLおよびAASLDコンセンサスガイドライン：レビュー。*Int J Hepatol*、2011、519783。doi:10.4061/2011/519783

Valente, S., Liu, Y., Schnekenburger, M., Zwerger, C., Cosconati, S., Gros, C., . . . Mai, A. (2014)。がん幹細胞を含むがんで活性化するヒトDNAメチルトランスフェラーゼの選択的非ヌクレオシド阻害剤。*J Med Chem*、57(3)、701-713。doi:10.1021/jm4012627

Warton, K., & Samimi, G. (2015)。がんの診断における無細胞循環DNAのメチル化。*Front Mol Biosci*、2、13。doi:10.3389/fmolb.2015.00013 10

Xu, R. H., Wei, W., Krawczyk, M., Wang, W., Luo, H., Flagg, K., . . . Zhang, K. (2017)。肝細胞がんの診断と予後のための循環腫瘍DNAメチル化マーカー。*Nat Mater*、16(11)、1155-1161。doi:10.1038/nmat4997

Zhai, R., Zhao, Y., Su, L., Cassidy, L., Liu, G., & Christiani, D. C. (2012)。食道腺がんおよびパレット食道における無細胞血清DNAのゲノムワイドなDNAメチル化プロファイリング。*Neoplasia*、14(1)、29-33。

【発明の概要】

【0006】

請求される主題の実施形態は、がんが、任意の正常組織および血球のDNAメチル化プロファイルとは異なる、一連の「カテゴリー的に」区別されるDNAメチル化シグネチャに関連付けられることを示す。これらのサイト(site)は、がんと他の組織との間に二元的な分化を作り、これにより、これらのサイトはがんでのみメチル化され、他のがんでは完全にメチル化されない。したがって、ディープ次世代シーケンスを使用して、メチル化の正常な細胞DNAプロファイルのバックグラウンドにあるがん細胞の数個の分子さえも検出することが可能である。本発明の主題の実施形態は、他の組織からの無細胞DNAの高いバックグラウンド上でさえ、無細胞腫瘍DNAの検出を可能にし、したがって、例えば、唾液、血漿、尿、糞便など体液から抽出された無細胞(CF)DNAを使用するがんの早期検出に特に適している。実施形態はまた、パップテストなどの組織スミアならびにバイオプシーおよび針バイオプシーにおけるがんの早期検出を可能にする。従来技術での以前の分析では、同じ組織と血液の正常細胞とがん細胞、およびDNAメチル化レベルが量的に異なる派生サイト(site)のみが比較されていた(Xu et al., 2017)。ただし、このような先行技術の分析で発見されたサイトは、他の組織のCF DNAと混合した場合、CF腫瘍DNAを検出できない(中山大学がん病院のHCCのctDNAマーカーについては、図2を参照する)。本発明請求の主題の一実施形態は、すべての組織ではメチル化されていないが、特定のがんではメチル化されているユニークなサイトのセットを明らかにしている。別の実施形態は、次世代シーケンシング、MeDIPアレイ、MeDIPシーケンスなどによって得られたゲノムワイドなDNAメチル化データのさまざまなソースを使用して、「バイナリーカテゴリー区別(BCD)法」(binary-categorical differentiation method)と呼ばれるがん、他の組織および他の疾患におけるカテゴリー的に区別されるメチル化サイトを発見する方法を明らかにする。一実施形態は、ゲノムワイドデータの発見セットにおけるa. 肝細胞がん(HCC)、b. 肺がん、c. 前立腺がん、d. 乳がん、e. 大腸がん、f. 頭頸部扁平上皮がん(HNSC)、g. 膵臓がん、h. 脳がん(膠芽腫)、i. 胃がん、j. 卵巣がん、k. 子宮頸がん、l. 食道がん、m. 膀胱がん、n. 腎臓がん、o. 精巣がん、p. 一般的な固形腫瘍、q. 血液がんのプロファイルの検出のための「カテゴリー」DNAメチル化サイトの組み合わせを明らかにする。別の実施形態はまた、起源の組織によって腫瘍を区別する「カテゴリー」DNAメチル化サイトの組み合わせ 50

を明らかにする。この実施形態は、組織特異性が低いメチル化 C F DNA を検出するための先行技術の方法とアッセイを区別する。実施形態は高い感度および特異性で、何百人もの患者からの DNA メチル化データならびに腫瘍の起源の組織におけるがんの検出のためのポリジーン DNA メチル化アッセイを検証する。本発明は、ターゲット特異的プライマーとそれに続くバーコードプライマーによる逐次増幅、および単一の次世代 M i s e q シーケンス反応におけるマルチプレックスシーケンス、血漿、唾液、尿などの少量の体液からのデータ抽出とメチル化の定量化により、何百人もの人々の C G I D のポリジーンセットにおける DNA メチル化を同時に正確に測定する方法を開示する。本発明の主題の別の実施形態はまた、パイロシーケンスアッセイまたはメチル化特異的 P C R を使用する、前記 DNA メチル化 C G I D のメチル化の測定を開示する。別の実施形態は、がんを有する人を健康な人から区別する、「カテゴリー」またはポリジーン加重メチル化スコアのいずれかの計算を開示する。別の実施形態は、血漿、尿、糞便、組織バイオプシーまたは組織スワブからがんの他の臨床的証拠がない人のがんの予測に至る新規プロセスを開示する。別の実施形態は、がんならびに、細胞死およびアルツハイマー病およびニューロンの他の神経変性疾患、心筋細胞の心臓病などの C F DNA の放出を含む他の疾患を検出するために当業者によって使用され得る。実施形態に記載されている DNA メチル化マーカー (C G I D) は、下記に利用される。 a . 定期的な「健康診断」による、すなわち「健康な」人のがんの非侵襲的早期発見。 b . H C C のリスクが高い慢性肝炎患者や肺がんのリスクが高い喫煙者などの、「リスクの高い」人の監視。 c . がん治療を受けている患者の治療への反応を監視し、再発または転移を検出。

10

20

【 0 0 0 7 】

実施形態は、本明細書に開示される DNA メチル化測定方法に基づいて、ポリジーンまたはカテゴリースコアを使用して未知のサンプルのがんを検出する有用性を実証する。開示された実施形態は、体液、糞便、尿、および任意のがんの組織または罹患組織におけるがんを検出するために、例えば次世代バイサルファイトシーケンス (b i s u l f i t e s e q u e n c i n g)、I l l u m i n a E p i c マイクロアレイ、キャプチャーシーケンス、メチル化 DNA 免疫沈降 (M e D I P) メチル化特異的 P C R、および利用可能になるメチル化測定法など、当業者が利用できるメチル化分析の方法を使用することで、当業者によって使用され得る。

【 0 0 0 8 】

実施形態はまた、次世代バイサルファイトシーケンス、M e D i p シーケンス、イオントレントシーケンス、E p i c マイクロアレイなどに続いて、疾患の非侵襲的検出に使用される特定の高感度マーカーを発見するためのバイナリーカテゴリー区別 (B C D) 法の分析など、ゲノムワイドシーケンスのために当業者が利用可能な任意の方法を使用して、他のがんおよび疾患の新しい「ポリジーン」カテゴリー DNA メチル化マーカーの発見の可能性を開示する。

30

実施形態はまた、ゲノムワイド DNA メチル化マップから「バイナリーカテゴリー区別 (B C D) 」法を使用して「バイナリー」DNA メチル化マーカーを導出するステップを含む、がんの「バイナリーカテゴリー」DNA メチル化マーカーを使用して無細胞 DNA 中の腫瘍由来の DNA からがんを検出することで、正常 DNA の量にかかわらず、がんとすべての正常組織との間でバイナリー的にかつカテゴリー的に区別する領域を発見する方法を開示する。

40

前記ゲノムワイド DNA メチル化マップが、がん細胞、正常な組織および血液 DNA の 1 つ以上である。

前記「バイナリーカテゴリー区別 (B C D) 」法が、ゲノムワイド DNA メチル化アレイ、DNA マイクロアレイ、ゲノムワイドバイサルファイトシーケンス、メチル化 DNA 免疫沈降 (M e D I P) シーケンス、およびオリゴヌクレオチドアレイとのハイブリダイゼーションの 1 つ以上の使用を含む。

【 0 0 0 9 】

本発明の主題の実施形態は、以下を含む。

50

第1の態様では、実施形態は、がんの早期検出のために、血漿などの体液中の無細胞DNAにおけるがんのポリジーンDNAメチル化マーカーを提供し、前記ポリジーンDNAメチル化マーカーセットは、本明細書に開示されている「バイナリーカテゴリー区別（BCD）分析」を使用して、Illumina 450KやEPICアレイ、ゲノムワイドバイサルファイトシーケンス、メチル化DNA免疫沈降（MeDIP）シーケンス、またはオリゴヌクレオチドアレイとのハイブリダイゼーションなどのマッピング方法によって得られたゲノム全体のDNAメチル化から得られる。

【0010】

他の実施形態では、血漿CF DNAまたは他の体液CF DNAを使用した肝臓がん肝細胞がん（HCC）の早期検出のために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせ（または表1の下に例として示される本リストの短いサブセット）である。

10

【0011】

【表1】

肝臓がん_検出 (Detect)					
cg00370303	cg10900437	cg02012576	cg16460359	cg04035559	cg17419241
cg00931619	cg11223367	cg03768777	cg16977570	cg04085025	cg18607529
cg05040544	cg19289599	cg06233293	cg24804544	cg26523670	cg09992116
cg05739190	cg24599205				

20

【0012】

検出するためのサブセット：

30

cg02012576, cg03768777, cg24804544, cg05739190

【0013】

他の実施形態では、血漿CF DNAまたは他の体液DNAを使用してがんの起源をHCCとして特定し、そして他の10種の一般的な固形腫瘍がんと区別するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせまたは本リストの短いサブセット（表2の下に例として示される）である。

【0014】

40

50

【表 2】

肝臓がん_スペック (Spec)					
cg12137206	cg14126493	cg06105778	cg22076972	cg13341720	cg03705926
cg09363194	cg07036412	cg02702614	cg10181419	cg05876864	cg11068343
cg17167468	cg15375239	cg00026222	cg17283781	cg16147221	cg26386472
cg14570307	cg06207432	cg07610192	cg03422204	cg11684022	cg23693289
cg21107197	cg04920951	cg20385508	cg25296314	cg20707679	cg26703661
cg00456086	cg05009389	cg19388016	cg08460435	cg04739306	cg04221886
cg26797073	cg04109768	cg05337743	cg00483503	cg18668780	cg10604002
cg27650175	cg05684891	cg26026416	cg00177496	cg14221460	cg16551483
cg13438961	cg24432073	cg21059834	cg23305567	cg04809136	cg21105227

10

【0015】

特異性 (s p e c i f i c i t y 、 スペック) のためサブセット :

c g 1 4 1 2 6 4 9 3

【0016】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用した肺がんの早期検出のために、ポリジーン DNA メチル化マーカーは、以下のリストの C G ID の組み合わせまたは本リストの短いサブセット (表 3 の下に例として示される) である。

20

【0017】

【表 3】

肺がん_検出					
cg04223424	cg06530490	cg08944430	cg09463882	cg11017065	cg12405785
cg16405026	cg21410080	cg23141355	cg25024074		

30

【0018】

検出するためのサブセット :

c g 0 4 2 2 3 4 2 4 , c g 2 3 1 4 1 3 5 5

【0019】

他の実施形態では、血漿 C F DNA または他の体液 DNA を使用してがんの起源を肺がんとして特定し、そして他の 10 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G ID の組み合わせまたは本リストの短いサブセット (表 4 の下に例として示される) である。

40

【0020】

50

【表 4】

肺がん_スベック					
cg05917732	cg02682457	cg23407396	cg23601095	cg23141355	cg15386964
cg02578368	cg24631970	cg27487839	cg16405026		
cg23141355	cg06530490	cg04223424	cg25470077	cg07138603	cg23460835
cg20678442	cg15436096				

10

【0021】

スベックのためのサブセット：

cg05917732, cg25470077

【0022】

他の実施形態では、血漿CF DNAまたは他の体液CF DNAを使用して、前立腺がんを早期に検出し、そしてがんの起源を前立腺がんとして特定し、ならびに他の16種の一般的な固形腫瘍がんと区別するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせ（または表5の下に例として示される本リストの短いサブセット）である。

20

【0023】

【表 5】

前立腺がん_検出_スベック			
cg14283569	cg18085998	cg13303553	cg17929627

30

【0024】

検出_スベックのためのサブセット：

cg14283569

[上記の表に示された4つの組み合わせのサブセットである]

【0025】

他の実施形態では、血漿CF DNAまたは他の体液CF DNAを使用して、乳がんを早期に検出するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせまたは本リストの短いサブセット（表6の下に示される例など）である。

【0026】

【表 6】

乳がん_検出					
cg13031251	cg19595750	cg16842053	cg18565473	cg09734791	cg25694349
cg09695735	cg03637878	cg26261793	cg18800085	cg16132520	cg05617413
cg17228900	cg06945936	cg08406370	cg24427504	cg26937500	cg11297107
cg02215070	cg14140647	cg05377226	cg07070305	cg24899571	cg07844931

40

【0027】

50

検出するためのサブセット：

c g 1 3 0 3 1 2 5 1 , c g 0 9 7 3 4 7 9 1 , c g 0 9 6 9 5 7 3 5 , c g 0 3 6 3 7 8 7 8

【 0 0 2 8 】

他の実施形態では、血漿 C F DNA または他の体液 DNA を使用してがんの起源を乳がんとして特定し、そして他の 1 0 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G ID の組み合わせまたは本リストの短いサブセット（表 7 の下に例として示される）である。

【 0 0 2 9 】

【 表 7 】

乳がん_スペック				
cg00467244	cg04194674	cg06879394	cg10720997	cg19966212
cg00722320	cg04751811	cg06998282	cg11498607	cg24525457
cg01308827	cg06282270	cg08066035	cg14862207	cg26228266
cg03113878	cg06405186	cg08296680	cg17945153	cg20180843

10

20

【 0 0 3 0 】

スペックのためのサブセット：

c g 0 3 1 1 3 8 7 8 , c g 2 0 1 8 0 8 4 3

【 0 0 3 1 】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、大腸がん（CRC）を早期に検出し、そしてがんの起源を大腸がんとして特定し、ならびに他の 1 6 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G ID の組み合わせまたは本リストの短いサブセット（表 8 の下に例として示される）である。

30

【 0 0 3 2 】

【 表 8 】

CRC_検出_スペック					
cg08808128	cg03788131	cg09854653	cg21627760	cg08169901	cg07494047
cg01566242	cg13788592	cg24102266	cg17716617	cg16733654	

40

【 0 0 3 3 】

検出 - スペックのためのサブセット：

c g 0 9 8 5 4 6 5 3 , c g 0 1 5 6 6 2 4 2

【 0 0 3 4 】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、膵臓がんを早期に検出するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G

50

I Dの組み合わせまたは本リストの短いサブセット（表9の下に例として示される）である。

【0035】

【表9】

膵臓がん_検出					
cg11017065	cg23833588	cg25024074	cg21277995	cg22286978	cg06241792
cg11591516	cg10096177	cg12035092	cg03611007	cg17996329	cg13807970
cg16678602	cg15386964	cg01423964	cg07900968	cg19118812	cg06728579
cg16232979	cg08406370				

10

【0036】

検出するためのサブセット：

cg25024074, cg15386964, cg16232979

【0037】

他の実施形態では、血漿CF DNAまたは他の体液DNAを使用してがんの起源を膵臓がんとして特定し、そして他の10種の一般的な固形腫瘍がんと区別するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせまたは本リストの短いサブセット（表10の下に例として示される）である。

20

【0038】

【表10】

膵臓がん _スペック					
cg01237565	cg08182975	cg15323936	cg19102272	cg01311909	cg26927232
cg22290704	cg15832577	cg20983577	cg26466027	cg09796911	cg15850155
cg15510118	cg25591377	cg16165258	cg25595541	cg18151519	cg19749445
cg14870128	cg13441142				

30

【0039】

スペックのためのサブセット：

cg01237565, cg08182975, cg20983577, cg25591377

40

【0040】

他の実施形態では、血漿CF DNAまたは他の体液CF DNAを使用して、脳がん（膠芽腫）を早期に検出し、そしてがんの起源を脳がん（膠芽腫）として特定し、ならびに他の10種の一般的な固形腫瘍がんと区別するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせまたは本リストの短いサブセット（表11の下に例として示される）である。

【0041】

50

【表 1 1】

脳がん_検出-スペック					
cg06887260	cg19116006	cg10938374	cg03663746	cg25568243	cg27449131
cg24917627	cg04134305	cg09797645	cg02892595	cg13231951	cg26269703
cg09183941	cg16842053	cg26551897	cg26988692	cg07849581	cg25026703
cg06798642	cg22963915	cg04245373	cg25533556	cg24917627	cg27659841
cg04692993	cg06045408	cg06887260	cg11345323	cg17167076	cg17526812
cg19929355	cg22513169	cg22865720			

10

【 0 0 4 2 】

検出 - スペックのためのサブセット :

c g 1 9 9 2 9 3 5 5

【 0 0 4 3 】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、胃 (s t o m a c h) [胃 (g a s t r i c)] がんを早期に検出するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット (表 1 2 の下に例として示される) である。

20

【 0 0 4 4 】

【表 1 2】

胃がん_検出					
cg04125371	cg05377226	cg05611779	cg06241792	cg07900968	cg09734791
cg11017065	cg12510981	cg13807970	cg15760257	cg18323466	cg19118812
cg19419279	cg19769760	cg20334243	cg26261793		

【 0 0 4 5 】

検出するためのサブセット :

c g 0 5 6 1 1 7 7 9 , c g 0 9 7 3 4 7 9 1 , c g 1 5 7 6 0 2 5 7

30

【 0 0 4 6 】

他の実施形態では、血漿 C F DNA または他の体液 DNA を使用してがんの起源を胃がんとして特定し、そして他の 1 0 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット (表 1 3 の下に例として示される) である。

【 0 0 4 7 】

【表 1 3】

胃がん_スペック					
cg17187167	cg07350470	cg05000488	cg25612145	cg09851120	cg00904726
cg04911739	cg18192294	cg11861709	cg19452853	cg02706110	cg03768513
cg05611779	cg06981182	cg06118999	cg04812509	cg10131095	cg05339066
cg19235339					

40

【 0 0 4 8 】

スペックのためのサブセット :

50

c g 0 5 1 1 7 7 9 , c g 1 9 2 3 5 3 3 9

【 0 0 4 9 】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、卵巣がんを早期に検出するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット（表 1 4 の下に例として示される）である。

【 0 0 5 0 】

【 表 1 4 】

卵巣がん_検出					
cg24339193	cg04008429	cg06849719	cg22694153	cg11252337	cg15804105
cg12479674	cg06961071	cg21210985	cg01556502	cg02537149	cg23983315
cg03597143	cg27209395				

10

【 0 0 5 1 】

検出するためのサブセット：

c g 2 4 3 3 9 1 9 3 , c g 2 2 6 9 4 1 5 3 , c g 1 1 2 5 2 3 3 7 , c g 2 1 2 1 0 9 8 5

20

【 0 0 5 2 】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用してがんの起源を卵巣がんとして特定し、そして他の 1 0 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット（表 1 5 の下に例として示される）である。

【 0 0 5 3 】

【 表 1 5 】

卵巣がん_スペック					
cg00895834	cg01159194	cg01961086	cg02649698	cg03345116	cg03456771
cg09392827	cg10178270	cg10581012	cg13459217	cg15701612	cg17130982
cg05901462	cg07068768	cg08389588	cg09173621	cg19276014	cg19846609
cg18476766	cg19129687				

30

【 0 0 5 4 】

スペックのためのサブセット：

c g 0 7 0 6 8 7 6 8 , c g 1 9 8 4 6 6 0 9

【 0 0 5 5 】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、子宮頸がんを早期に検出するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット（表 1 6 の下に例として示される）である。

40

【 0 0 5 6 】

50

【表 1 6】

子宮頸がん_検出					
cg00578154	cg13644629	cg01423964	cg15745619	cg04522671	cg21621906
cg00757182	cg14962363	cg01601746	cg17228900	cg07126167	cg22806837
cg08134106	cg22922289	cg09260640	cg24625128	cg11259628	cg27420520
cg08535260	cg23141355	cg09734791	cg25024074		

10

【 0 0 5 7 】

検出するためのサブセット：

c g 0 0 7 5 7 1 8 2 , c g 0 1 6 0 1 7 4 6

【 0 0 5 8 】

他の実施形態では、血漿 C F DNA または他の体液 DNA を使用してがんの起源を子宮頸がんとして特定し、そして他の 1 0 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G ID の組み合わせまたは本リストの短いサブセット（表 1 7 の下に例として示される）である。

20

【 0 0 5 9 】

【表 1 7】

子宮頸がん_スペック					
cg00829990	cg02401399	cg04996219	cg07066594	cg07195011	cg07576142
cg09260640	cg12961842	cg13668618	cg18543270		

30

【 0 0 6 0 】

スペックのためのサブセット：

c g 0 7 0 6 6 5 9 4 , c g 0 9 2 6 0 6 4 0 , c g 1 2 9 6 1 8 4 2

【 0 0 6 1 】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、頭頸部扁平上皮がん（HNSC）を早期に検出するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G ID の組み合わせまたは本リストの短いサブセット（表 1 8 の下に例として示される）である。

40

【 0 0 6 2 】

【表 1 8】

HNSC_検出					
cg01613638	cg02776314	cg03280624	cg04524120	cg05151803	cg07746323
cg07900968	cg08406370	cg11108676	cg12083965	cg15397448	cg17428324
cg18403606	cg20334243	cg26770917	cg27009208	cg27420520	

50

【 0 0 6 3 】

検出するためのサブセット：

c g 0 7 9 0 0 9 6 8 , c g 2 0 3 3 4 2 4 3 , c g 2 7 4 2 0 5 2 0

【 0 0 6 4 】

他の実施形態では、血漿 C F DNA または他の体液 DNA を使用してがんの起源を頭頸部扁平上皮がん (H N S C) として特定し、そして他の 1 0 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット (表 1 9 の下に例として示される) である。

【 0 0 6 5 】

【 表 1 9 】

10

HNSC_スペック					
cg01217080	cg03522799	cg06672120	cg09136346	cg10155875	cg18006328
cg18443253	cg19287220				

【 0 0 6 6 】

スペックのためのサブセット：

c g 1 8 0 0 6 3 2 8 , c g 1 9 2 8 7 2 2 0

【 0 0 6 7 】

20

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、食道がんを早期に検出するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット (表 2 0 の下に例として示される) である。

【 0 0 6 8 】

【 表 2 0 】

30

食道がん_検出					
cg03280624	cg03735888	cg06530490	cg06963053	cg08944430	cg09734791
cg11017065	cg12035092	cg18344922	cg19118812	cg20334243	cg22128431
cg23141355	cg24740531	cg27009208	cg27420520		

【 0 0 6 9 】

検出するためのサブセット：

c g 0 3 2 8 0 6 2 4 , c g 0 3 7 3 5 8 8 8 , c g 0 9 7 3 4 7 9 1 , c g 2 7 4 2 0 5 2 0

【 0 0 7 0 】

40

一実施形態では、血漿 C F DNA または他の体液 DNA を使用してがんの起源を食道がんとして特定し、そして他の 1 0 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット (表 2 1 の下に例として示される) である。

【 0 0 7 1 】

50

【表 2 1】

食道がん_スベック					
cg00532449	cg02763101	cg04743654	cg08055087	cg08932440	cg09556952
cg10608333	cg12473285	cg12966367	cg17579667	cg17949440	cg18723937
cg21554552	cg22647407	cg23286646	cg23730575	cg27569446	

10

【0072】

スベックのためのサブセット：

C g 0 9 5 5 6 9 5 2 , c g 1 2 4 7 3 2 8 5

【0073】

他の実施形態では、血漿 C F DNA または他の体液 C F DNA を使用して、膀胱がんを早期に検出するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット（表 2 2 の下に例として示される）である。

20

【0074】

【表 2 2】

膀胱がん_検出					
cg01423964	cg04223424	cg01556502	cg23141355	cg10723962	cg09734791
cg25024074	cg21039778	cg19113641	cg04342821	cg09008417	cg05873285

【0075】

検出するためのサブセット：

c g 0 4 2 2 3 4 2 4 , c g 1 0 7 2 3 9 6 2 , c g 2 5 0 2 4 0 7 4

30

【0076】

他の実施形態では、血漿 C F DNA または他の体液 DNA を使用してがんの起源を膀胱がんとして特定し、そして他の 10 種の一般的な固形腫瘍がんと区別するために、ポリジーン DNA メチル化マーカーは、以下のリストの C G I D の組み合わせまたは本リストの短いサブセット（表 2 3 の下に例として示される）である。

【0077】

【表 2 3】

膀胱がん_スベック					
cg14773260	cg08827322	cg05085809	cg13544006	cg02293936	cg10153335
cg05039004	cg14718568	cg07790170	cg12995090	cg16495265	cg01508045
cg04933208	cg03231163	cg06708937	cg00675569	cg06312813	cg14126688
cg12984729	cg02384661	cg08307030	cg26884027	cg20540209	cg04355159
cg08857479	cg23395715	cg22006640	cg26279336	cg19898108	cg17446010
cg12911122					

40

【0078】

50

スペックのためのサブセット：

cg13544006

【0079】

他の実施形態では、血漿CF DNAまたは他の体液DNAを使用して、腎臓（renal）[腎臓（kidney）]がんを早期に検出し、がんの起源を腎臓がんとして特定し、そして他の10種の一般的な固形腫瘍がんと区別するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせまたは本リストの短いサブセット（表24の下に例として示される）である。

【0080】

【表24】

10

腎臓がん_検出_スペック					
cg08884571	cg00011225	cg14535274	cg05507490	cg10367244	cg07190763
cg02971546	cg19990785	cg03084949	cg08765317	cg02820958	cg23946709
cg26642667					

【0081】

検出 スペックのためのサブセット：

cg08884571, cg00011225, cg23946709

20

【0082】

他の実施形態では、血漿CF DNAまたは他の体液CF DNAを使用して、精巣がんを早期に検出し、がんの起源を精巣がんとして特定し、そして他の10種の一般的な固形腫瘍がんと区別するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせまたは本リストの短いサブセット（表25の下に例として示される）である。

【0083】

【表25】

30

精巣がん_検出_スペック					
cg14531093	cg11777290	cg13283877	cg03966099	cg26939375	cg26297530
cg16157016	cg14002772	cg09890687	cg05134918	cg09719850	cg06456125
cg17978367	cg20100671	cg26627956	cg19789755	cg14393609	cg21634331
cg02039634	cg25159927	cg01895439			

【0084】

検出とスペックのためのサブセット：

cg14531093, cg25159927

40

【0085】

他の実施形態では、血漿CF DNAまたは他の体液CF DNAを使用して、13種の最も一般的な固形腫瘍の1つを早期に検出するために、ポリジーンDNAメチル化マーカーは、以下のリストのCG IDの組み合わせまたは本リストの短いサブセット（表26の下に例として示される）である。

【0086】

50

【表 2 6】

汎がん_検出					
cg01423964	cg04223424	cg06530490	cg06543087	cg07900968	cg08406370
cg08848774	cg09734791	cg10723962	cg11017065	cg15759056	cg15760257
cg16405026	cg17228900	cg21277995	cg22286978	cg23141355	cg24427504
cg25024074	cg27420520				

10

【0087】

検出するためのサブセット：

cg10723962, cg15759056, cg24427504, cg25024074

【0088】

他の実施形態では、白血球、血漿CF DNAまたは他の体液CF DNAを使用して、AML、CLLなどの血液がんを早期に検出するために、ポリジーンDNAメチル化マーカーは、表27に示されるようなゲノムワイドなDNAメチル化データに対してのBCD法によって描かれたCG IDの組み合わせ（または表27の下に示される本組み合わせの短いサブセット）である。

20

【0089】

【表 2 7】

AML_検出- スペック					
cg00594866	cg05608159	cg18658397	cg18780412	cg20439288	cg22828045
cg25375340					

30

【0090】

検出 - スペックのためのサブセット：

cg18658397, cg18780412, cg20439288, cg22828045, cg25375340

【0091】

他の実施形態では、血漿CF DNAまたは他の体液CF DNAを使用して、黒色腫を早期に検出し、がんの起源を黒色腫として特定し、そして他の16種一般的な固形腫瘍がんと区別するために、ポリジーンDNAメチル化マーカーは、以下のリストに示されるCG IDの組み合わせ（または表28の下に例として示される本リストの短いサブセット）である。

40

【0092】

50

【表 2 8】

黒色腫 検出-スペック					
cg00325866	cg01228636	cg04616691	cg04824711	cg05569109	cg06019853
cg08611163	cg10830649	cg12593303	cg15307891	cg18866529	cg19530886
cg19634213	cg20146030	cg21755725	cg22280705	cg24217704	cg24678095
cg26127345	cg27084903				

10

【0093】

検出 - スペックのためのサブセット：

cg15307891, cg18866529, cg27084903

【0094】

本発明の主題の別の態様では、ポリジーンDNAメチル化マーカーのDNAメチル化測定値を検出するための手段および試薬を含む、がんを検出するためのキットおよびプロセスが提供される。

【0095】

一実施形態では、表1および表2のCG IDのDNAメチル化測定値のための手段および試薬を含む、肝細胞がんを検出するためのキットが提供される。

20

【0096】

別の実施形態では、表3および表4のCG IDのDNAメチル化測定値のための手段および試薬を含む、肺がんを検出するためのキットが提供される。

【0097】

別の実施形態では、表5のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、前立腺がんを検出するためのキットが提供される。

【0098】

別の実施形態では、表6および表7のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、乳がんを検出するためのキットが提供される。

30

【0099】

別の実施形態では、表8のCG IDのDNAメチル化測定値のための手段および試薬を含む、大腸がんを検出するためのキットが提供される。

【0100】

別の実施形態では、表9および表10のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、膵臓がんを検出するためのキットが提供される。

【0101】

さらに別の実施形態では、表11のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、脳がんを検出するためのキットが提供される。

【0102】

別の実施形態では、表12および表13のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、胃がんを検出するためのキットが提供される。

40

【0103】

別の実施形態では、表14および表15のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、卵巣がんを検出するためのキットが提供される。

【0104】

別の実施形態では、表16および表17のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、子宮頸がんを検出するためのキットが提供される。

【0105】

別の実施形態では、表18および表19のCG IDのDNAメチル化測定値を検出す

50

るための手段および試薬を含む、頭頸部扁平上皮がん（HNSC）を検出するためのキットが提供される。

【0106】

別の実施形態では、表20および表21のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、食道がんを検出するためのキットが提供される。

【0107】

別の実施形態では、表22および表23のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、膀胱がんを検出するためのキットが提供される。

【0108】

別の実施形態では、表24のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、腎臓がんを検出するためのキットが提供される。 10

【0109】

別の実施形態では、表25のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、精巣がんを検出するためのキットが提供される。

【0110】

他の実施形態では、表26のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、13種の一般的ながん（膀胱、脳、乳房、子宮頸部、大腸、食道、HNSC、HCC（肝臓）、肺、卵巣、膵臓、前立腺、胃）のうちの1つを検出するためのキットが提供される。

【0111】

別の実施形態では、表27の血液がんの異なるサブタイプに特異的であるBCD法によって検出されたCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、AMLおよびCLLなどの血液がんを検出するためのキットが提供される。 20

【0112】

別の実施形態では、表28のCG IDのDNAメチル化測定値を検出するための手段および試薬を含む、黒色腫を検出するためのキットが提供される。

【0113】

別の実施形態では、DNAパイロシーケンスメチル化アッセイは、上記にリストされるCG IDを使用することにより、例えば以下に開示されるプライマーおよびパイロシーケンス反応の標準条件を使用することにより、血漿CF DNAなどの体液中のHCCを予測するために使用される。 30

cg 02012576

フォワード:GGTAGTTAGGAAGTTTLAGAGGTTGTAGTA

リバーズ (ピオチン標識された): ACCACTACCCCAACCCAACCCCTA

配列:GGTTTTAGGATGTTTTG

cg 03768777 (VASH2)

フォワード:AGAATAATATTAGAGAATGGGATATGGAA

リバーズ (ピオチン標識された): ACAACTCCAAAATCCTACCT

配列:GAATGGGATATGGAATGA

cg 05739190 (CCNJ)

フォワード:GTTTAGGAGTTGGGTTTTAGTTGAG

リバーズ (ピオチン標識された): ACCCCACCCTAACTCCCTTACC

配列:TGGGTTTTAGTTGAGG

cg 24804544 (GRID2IP)

フォワード (ピオチン標識された): GGGTAGGGGAGGGTTTTGAAATA

リバーズ:TAACCCCCCTCCAACCTCATTC

配列:CACCCAACCTTCTCAAT

がんの起源の組織の特異性は、以下のCG ID cg 02012576 (HPX)のDNAメチル化を測定することによって決定される。

フォワード (ピオチン標識された): ATTTTTATGGGTATTAGTTTTAGGGAGAA

リバース (ピオチン標識された): CCAAAACTATCCTATAACCTCTACAACCTCA

配列: ACCATTACCACCCCT

【0114】

別の実施形態では、ポリジーン多重増幅バイサルファイトシーケンスDNAメチル化アッセイ (polygenic multiplexed amplicon bisulfite sequencing DNA methylation assay) が、上記のCG IDを使用することにより、血漿CF DNAなどの体液中のがんを予測するために使用される。例えば、以下に開示されているプライマーと標準条件を使用して前立腺がんを予測する。該標準条件は、バイサルファイト変換、ターゲット特異的プライマー (PCR1) とそれに続くバーコードプライマー (PCR2) による逐次増幅、および単一の次世代Miseqシーケンサー (Illumina) でのマルチプレックスシーケンス、Illuminaソフトウェアを使用した逆多重化 (demultiplexing)、Methylkitなどのメチル化分析の標準的な方法を使用したデータ抽出とメチル化の定量化、その後の加重DNAメチル化 (weighted DNA methylation) スコアの計算と、血漿、唾液または尿などの少量の体液からのがんの予測を含む。

10

【0115】

第一のPCRで前立腺がんを検出するステップは次のとおりである：

CG ID cg02879662の場合

フォワードプライマー：

5' ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGGTAGGAGTTTTGGGAATTGG3'

20

リバースプライマー：

5' gTgACTggAgTTCAGACgTgTgCTCTTCCgATCTCCACCCCTACAATCCCTAA3' CG ID cg16232979の場合

フォワードプライマー：

5' ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNYGGTTTTYGGGTTTTYGTATT3'

リバースプライマー：

5' gTgACTggAgTTCAGACgTgTgCTCTTCCgATCTACRCAAAAATATAAATCRACRATC3'

30

【0116】

がんが前立腺に特異的に発生していることをテストするために、第一のPCRは次のように実行される：

CG ID : cg14041701およびcg14498227の場合

フォワードプライマー：

5' ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGTTTTGYGTTTTYGGATTTGGGTT3'

リバースプライマー：

5' gTgACTggAgTTCAGACgTgTgCTCTTCCgATCTCATAAACACACCTTTAAATAAACACTAAA3'

40

【0117】

サンプルをバーコード化するには、次のプライマーを用いて第二のPCR反応を使用する。

フォワードプライマー：

5' AATgATACggCgACCACCgAgATCTACACTCTTTCCCTACACgAC3'

バーコードプライマー (リバース)：

5' CAAGCAGAAgACggCATACgAgATAGTCATCGgTgACTggAgTTCAGACgTg3'

(赤い塩基はインデックスであり、このインデックスの200のバリエーションが使用さ

50

れる)

【 0 1 1 8 】

他の実施形態では、受信者動作特性 (R O C) アッセイは、C G I D の加重 D N A メチル化測定値を使用して、がんと正常との間のしきい値を定義することによってがんを検出するために使用される。しきい値を超える / 下回るサンプルは、がんとして分類される。例えば、H C C を検出するための上記の C G I D がある :

cg02012576	cg03768777	cg05739190	cg24804544
------------	------------	------------	------------

10

【 0 1 1 9 】

別の実施形態では、上記にリストされる C G I D のメチル化の測定値を使用することによってがんを予測するために、階層的クラスタリング分析アッセイが使用される。

【 0 1 2 0 】

本発明の主題の別の態様では、がんおよび他の疾患を検出するための D N A メチル化マーカーを同定する方法は、臨床サンプルから得られた D N A メチル化測定値に関して以前に開示された「バイナリーカテゴリー区別 (B C D)」法で統計分析を行うステップを含む。

【 0 1 2 1 】

別の実施形態では、この方法は、サンプルから得られた D N A メチル化測定値に対して統計分析および「バイナリーカテゴリー区別 (B C D)」法を実行することを含み、I l l u m i n a B e a d c h i p 4 5 0 K または少なくとも 1 つのサンプルから抽出された D N A の E P I C アレイを実行することによって得られた D N A メチル化測定値を伴う。

20

【 0 1 2 2 】

別の実施形態において、D N A メチル化測定値は、サンプルから抽出された D N A の D N A パイロシーケンス、続いて質量分析ベース (E p i t y p e r ^{T M})、P C R ベースのメチル化アッセイ、およびバイサルファイト変換された D N A に続いて、増幅の第二のセットにおけるバーコード化、そして I l l u m i n a 次世代シーケンサーでのインデックス付きマルチプレックスシーケンスからの、本明細書に開示される、標的 C G I D にまたがる領域の標的増幅の実行により得られる。

30

【 0 1 2 3 】

他の実施形態では、統計分析は、受信者動作特性 (R O C) アッセイを含む。

【 0 1 2 4 】

他の実施形態では、統計分析は、階層的クラスタリング分析アッセイを含む。

【 0 1 2 5 】

定義

本明細書で使用されるように、「C G」という用語は、シトシンおよびグアノシン塩基を含む D N A 中のジヌクレオチド配列を指す。これらのジヌクレオチドシーケンスは、人間や他の動物の D N A でメチル化される可能性がある。C G I D は、I l l u m i n a 4 5 0 K マニフェストで定義されているように、ヒトゲノムにおけるその位置を明らかにする (ここにリストされている C G の注釈は、<https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylation450k.db.html> で公開され、R パッケージ I l l u m i n a H u m a n M e t h y l a t i o n 4 5 0 k . d b I l l u m i n a H u m a n M e t h y l a t i o n 4 5 0 k . d b : I l l u m i n a H u m a n M e t h y l a t i o n 4 5 0 k 注釈データとしてインストールされている。R パッケージバージョン 2 . 0 . 9 .) 。

40

【 0 1 2 6 】

50

本明細書で使用されるように、診断機器または機器という用語は、1つ以上の試薬と共に使用して、例えば請求される主題の実施形態によるDNAメチル化スコアのDNAメチル化測定値を導出するために使用される診断テストを実行できる、当業者に知られている任意の機器である。

【0127】

本明細書で使用されるように、「ベータ値」という用語は、式 $\text{ベータ値} = \frac{\text{メチル化C強度}}{\text{メチル化C強度} + \text{非メチル化C強度}}$ を使用するメチル化プローブと非メチル化プローブの強度比を使用して、Illumina 450Kアレイの正規化および定量化によって導き出されたCG ID位置での推定のメチル化レベルを指す。ベータ値は0と1の間で、0は完全にメチル化されておらず、1は完全にメチル化されている。

10

【0128】

本明細書で使用されるように、「ペナルティ付き回帰」という用語は、例えば、R統計パッケージにおける、「ペナルティ付き」が、Goeman J. J.、Cox比例ハザードモデルでのL1ペナルティ付き推定 *Biometrical Journal* 52(1), 70-84で説明されているように、実施されるバイオマーカーのより大きなリストから結果を予測するために必要な予測因子の最小数を特定することを目的とした統計的方法を指す。

【0129】

本明細書で使用されるように、「クラスタリング」という用語は、同じグループ(クラスターと呼ばれる)内の対象が、他のグループ(クラスター)より(ある意味では)互いに類似するように、対象セットをグループ化することを指す。

20

【0130】

本明細書で使用するように、「階層的クラスタリング」という用語は、例えば、Kaufman, L.; Rousseeuw, P. J. (1990) データ内のグループの検索: クラスタ分析の概要(1版) New York: John Wiley. ISBN 0-471-87876-6で説明されているように、互いにクラスターがどの程度類似(近い)または類似していない(遠い)かに基づいて「クラスター」の階層を構築する統計的方法を指す。

【0131】

本明細書で使用されるように、「受信者動作特性(ROC)アッセイ」という用語は、予測因子の性能を例示するグラフィカルプロットを作成する統計的方法を指す。例えば、Hanley, James A., McNeil, Barbara J. (1982) 「受信者操作特性(ROC)曲線下の面積の意味と使用」 *Radiology* 143(1): 29-36で説明されているように、予測の真の陽性率は、予測因子のさまざまなしきい値設定(すなわち、メチル化の異なる%)での偽陽性率に対してプロットされる。

30

【0132】

本明細書で使用される「多変量またはポリジーン線形回帰」という用語は、CG IDのメチル化の割合などの複数の「独立変数」または「予測値」と、がんなどの「従属変数」の間の関係を推定する統計的方法を指す。この方法は、CG IDなどのいくつかの「独立変数」がモデルに含まれている場合の「結果」(がんなどの従属変数)を予測する際に、各CG IDの「重み(weight)」または係数(coefficient)を決定する。

40

【図面の簡単な説明】

【0133】

【図1】図1は、何百人もの血液サンプルと正常組織にまたがる完全にメチル化されていないサイトの候補リストを示す。図Aは、Illumina 450Kのゲノムワイドなメチル化アレイのすべての個体(<0.1)(GSE50192)でメチル化されていない17組織にわたるCG IDが、312の個体からの血液サンプル(GSE61496)のゲノムワイドなDNAメチル化アレイの非メチル化CG IDと重複して33477個のCG IDのリストを生成したことを示している。Bは、最も強力な非メチル化CG

50

I Dの候補リストを示しており、Aの33477個のCG I Dのリストは、19歳から101歳までの656人（女性および男性）の血液サンプル（GSE40279）のDNAメチル化アレイの非メチル化CG I Dと重複していた。結合された部分について、これらの分析により、すべての年齢層の多くの個人の組織と血液サンプルにわたってメチル化されていない信頼度の高い28754個のCG I Dのリストが生成された。これらの28754個の位置は、本発明の主題によって開示される「バイナリーカテゴリー区別（BCD）」法を使用して、がんにおいてカテゴリー的にメチル化されているが他の組織ではメチル化されていないサイトの発見に使用された。

【図2】図2は、HCCに対する現在の循環DNAマーカーの組織特異性の欠如を示す図である。図示されたヒートマップは、他の正常な組織におけるこれらのサイトのHCCとメチル化レベルのバイオマーカーとしてXu等（Xu等、2017）に候補としてリストされた10個のCG I Dを示す。HCCの特定のバイオマーカーとして提案されているCG I Dのいくつかは、他の組織でもメチル化されており、そして血中DNAのメチル化レベルがさまざまであることを示す。（青は0メチル化、暗赤色は100%メチル化）

10

【図2A】図2Aは、本発明の主題の実施形態による図2の一部の分解図である。

【図2B】図2Bは、本発明の主題の実施形態による図2の一部の分解図である。

【図2C】図2Cは、本発明の主題の実施形態による図2の一部の分解図である。

【図3】図3は、がんDNAに対してBCD法を使用して発見されたHCC DNAメチル化マーカーの特異性を示す図である。図示されたヒートマップは、ここで説明するBCD法によってHCC DNAメチル化マーカーとして選択された4つのCG I Dを示す。メチル化レベルは、がん（HCC）と正常な組織および血液との間でカテゴリー的に異なり、これにより、血液および他の組織のすべての個体で該サイトがメチル化されず、HCCで測定可能な程度にメチル化される。

20

【図3A】図3Aは、本発明の主題の実施形態による図3の一部の分解図である。

【図3B】図3Bは、本発明の主題の実施形態による図3の一部の分解図である。

【図3C】図3Cは、本発明の主題の実施形態による図3の一部の分解図である。

【図4】図4は、大腸がんに対する現在のDNAメチル化マーカーのがん組織起源特異性の欠如、および本発明の主題の実施形態による「検出-スベック」法との比較を示す図である。図Aは、大腸がんのCF DNAメチル化マーカー「Epi-大腸がん」（エピゲノミクス社より販売）に含まれるSept9遺伝子のCGサイトを示しており、マーカーはがんDNAメチル化データのTCGAコレクションからのメチル化データを利用して他の多くのがんを検出するために使用できるため、大腸がんに対する特異性に欠けている（HKG-大腸がん（HKG-epiCRC）、青）。BCD法（HKG-大腸がんオレンジ）（表8）を用いて発見された大腸がんの検出のために本発明の主題で開示されるマーカー（表8）は、他の一般的な固形腫瘍がんに対して試験した場合、大腸がん非常に特異的である。図Bと図Cは、HKG-大腸がん（HKG-epiCRC）（B）またはEpi-大腸がん（C）のいずれかのDNAメチル化マーカーを使用した、さまざまながんの異なる個人からの腫瘍DNAのDNAメチル化値の散布図である。注目すべきは、HKG-epiCRCマーカー（B）対Epi-大腸がんマーカー（C）の散在する異質なプロファイルを使用した、大腸がんと他のがんの間のDNAメチル化の厳格なカテゴリー的な違いである。

30

40

【図4A】図4Aは、本発明の主題の実施形態による図4の一部の分解図である。

【図4B】図4Bは、本発明の主題の実施形態による図4の一部の分解図である。

【図5】図5は、肝臓がん（HCC）の早期発見のためのポリジーンDNAメチル化マーカーの発見を示す図である。図Aは、ソースと患者数をリストした表を示し、これらの患者のメチル化データは、BCD法（表1）および起源特定のがんを決定するためのCG I D（表2）を使用する実施形態による、HCCの検出のための4つの一連のCG I Dの発見に使用された。図5の左下のパネルの図B（検出）は、1~145（正常79名およびHCC66名）からリストされた各テスト対象者のこれらのCG I D（表1）のメチル化スコアの合計を示す。該ポリジーンスコアは、HCCを有する人と正常な肝組織を有

50

する人をカテゴリー的に区別する。右下のパネルの図Cは、8種の異なる腫瘍（表2）のデータを使用して特定の腫瘍の起源を検出する1つのCGID（表2）のメチル化スコアを示す。マーカーは、他の起源のがんとHCCをカテゴリー的に区別する。

【図6】図6は、GSE76269（ $n = 227$ ）からのDNAメチル化データを使用した、HCC（スペック）のポリジーンDNAメチル化マーカーの検証の図である。図Aは、肝臓がん患者227名のDNAメチル化データと正常10名を使用したHCC DNAメチル化マーカーの曲線下の領域を示すROCプロットである。図6の図Bは、HCC検出の感度、特異性、および精度を示す。図Cは、検証データセットにおけるHCCの検出の予測率を示す。

【図7】図7は、TCGAメチル化データ（ $n = 4166$ ）における、HCC対他のがん 10
に対するポリジーンHKG - epi肝臓がん - 検出およびスペックマーカーの精度および
特異性の検証の図である。図7の図Aは、異なるがんを有する患者のHKG - 肝臓検出/
スペックマーカーDNAメチル化データの検出率を示す。HCCのほぼ完全な特異性に注
目すべきである。図Bは、TCGAにおける4166人の患者のDNAメチル化データに
おける、HCCのHKG - 肝臓 - 検出マーカーの特異性と感度のROCプロットである。
図Cは、HCC対他の起源のがんに対する感度と特異性である。

【図8】図8は、肺がんのポリジーンDNAメチル化マーカーの発見の図である。図8の
図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BC
D法（表3）および起源特定のがん組織を決定するためのCGID（表4）を使用する
実施形態に開示される、肺がんの検出のための一連のCGIDの発見に使用された。図8 20
の左下のパネルの図B（検出）は、1～20名（正常10名および肺がん10名）からリ
ストされた各テスト対象者のこれらのCGID（表3）のメチル化スコアの合計を示す
。該ポリジーンスコアは、肺がんを有する人と正常な組織を有する人をカテゴリー的に区
別する。図8の右下のパネルの図Cは、8種の異なる腫瘍（ $n = 80$ ）のある人からのデ
ータを使用して特定の腫瘍の起源を検出するCGID（表4）のメチル化スコアを示す。
これらの実施形態では、マーカーは他の起源のがんと肺がんをカテゴリー的に区別する。

【図9】図9は、TCGAメチル化データ（ $n = 4166$ ）における、肺がん対他のがん 30
に対するポリジーンHKG - epi肺がん - 検出およびスペックマーカーの精度および特
異性の検証の図である。図9の図Aは、異なるがんを有する患者のDNAメチル化デー
タを使用したHKG - epi肺がん検出/スペックマーカーの検出率を示す。肺がんの特異
性に注目すべきである。図9の図Bは、TCGAからの4166名の患者のDNAメチル
化データにおける、肺がんのHKG - 肺がん - 検出マーカーの特異性と感度のROCプロ
ットを示す。図9の図Cは、肺がん対他の起源のがんに対する感度と特異性を示す。

【図10】図10は、前立腺がんのポリジーンDNAメチル化マーカーの発見の図である
。図10の図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化デー
タは、BCD法（表5）および起源特定のがん組織を決定するためのCGID（表5）
を使用する実施形態に開示される、前立腺がんの検出のための一連のCGIDの発見に使
用された。図10の左下のパネルの図B（検出）は、1～15（正常5名および前立腺が
ん10名）からリストされた各テスト対象者のこれらのCGID（表5）のメチル化ス
コアの合計を示す。該ポリジーンスコアは、前立腺がんを有する人と正常な人 40
をカテゴリー的に区別する。図10の右下のパネルの図Cは、8種の異なる腫瘍（ $n = 80$ ）のある
人からのデータを使用して起源特定の前立腺組織を検出するCGID（表5）のメチル化スコア
を示す。これらの実施形態では、マーカーは他の起源のがんと前立腺がんをカテゴリー的
に区別する。

【図11】図11は、TCGAメチル化データ（ $n = 4166$ ）における、前立腺がん対
他のがんに対するポリジーンHKG - epi前立腺がん - 検出およびスペックマーカーの
精度および特異性の検証の図である。図11の図Aは、異なるがんを有する患者のDNA
メチル化データを使用したHKG - 前立腺がん検出/スペックマーカーの検出率を示す。
前立腺がんの特異性に注目すべきである。図11の図Bは、TCGAにある4166名の
患者からのDNAメチル化データを使用する、前立腺がんのHKG - 前立腺がん - 検出マ 50

ーカーの特異性と感度のROCプロットである。図11の図Cは、前立腺対他の起源のがんに対する感度と特異性を示す。

【図12】図12は、乳がんのポリジーンDNAメチル化マーカーの発見の図である。図12の図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表6)および起源特定のがんを決定するためのCGID(表7)を使用する実施形態に開示される、乳がんの検出のための一連のCGの発見に使用された。図12の左下のパネルの図B(検出)は、1~27(正常17名および乳がん10名)からリストされた各テスト対象者のこれらのCGID(表6)のメチル化スコアの合計を示す。該ポリジーンスコアは、乳がんを有する人と正常な乳組織を有する人をカテゴリー的に区別する。図12の右下のパネルの図Cは、8種の異なる腫瘍(n=80)のある人からのデータを

10

図13は、検証コホートGSE60185(n=285)において、HKG-epi乳がん-検出ポリジーンDNAメチル化マーカーが非浸潤性および浸潤性乳がんを検出する図である。図13の図Aは、239名の乳がん患者のDNAメチル化データ、17名の乳がんではない乳房形成術患者および29名の隣接組織を使用した乳がんポリジーンDNAメチル化マーカーの曲線下の領域を示すROCプロットである。すべての乳がんの感度、特異性および精度をBに示し、DCIS(非浸潤性乳管がん)、浸潤性乳がんおよび混合乳がんのサンプルの予測率を図13の図Cに示す。注目すべきは、乳がんマーカーが非常に早期の乳がん(DCIS)を検出することである。

20

【図14】図14は、TCGAメチル化データ(n=4166)における、乳がん対他のがんに対するポリジーンHKG-epi乳がん-検出およびスペックマーカーの精度および特異性の検証の図である。図14の図Aは、異なるがんを有する患者からのDNAメチル化データにおいて、HKG-epi乳がん検出/スペックマーカーの検出率を示す。乳がんの特異性に注目すべきである。図14の図Bは、TCGAにおける4166人患者のDNAメチル化データを使用して乳がんを検出するための、HKG-乳がん-検出マーカーの特異性と感度のROCプロットである。図14の図Cは、乳がん対他の起源のがんに対する感度と特異性を示す。

【図15】図15は、大腸がん(CRC)のポリジーンDNAメチル化マーカーの発見の図である。図15の図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表8)および起源特定のがんを決定するためのCGID(表8)を使用する実施形態に開示される、大腸がんの検出のための一連のCGIDの発見に使用された。図15の左下のパネルの図B(検出)は、1~75(正常25名および大腸がん50名)からリストされた各テスト対象者のこれらのCGIDのメチル化スコアの合計を示す。該ポリジーンスコアは、がんを有する人と正常な組織を有する人をカテゴリー的に区別する。図15の右下のパネルの図Cは、8種の異なる腫瘍(n=80)のある人からのDNAメチル化データを使用して腫瘍の特定の起源を検出するCGIDのメチル化スコアを示す。これらの実施形態では、マーカーは他の起源のがんと大腸がんをカテゴリー的に区別する。

30

40

【図16】図16は、TCGA DNAメチル化データセット(n=4166)を使用した、大腸がん対他のがんに対するポリジーンHKG-epiCRC-検出およびスペックマーカーの精度および特異性の検証の図である。図16の図Aは、異なるがんを有する患者のDNAメチル化データを使用するHKG-epiCRC検出/スペックマーカーの検出率を示す。大腸がんの特異性に注目すべきである。図16の図Bは、TCGAにおける4166名の患者からのDNAメチル化データを使用する、大腸がんのHKG-epi大腸がん-検出マーカーの特異性と感度のROCプロットである。図16の図Cは、大腸がん対他の起源のがんに対する感度と特異性を示す。

【図17】図17は、膵臓がんのポリジーンDNAメチル化マーカーの発見の図である。図17の図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データ

50

は、BCD法(表9)および起源特定のがんを決定するためのCGID(表10)を使用する本発明に開示される膵臓がんの検出のための一連のCGIDの発見に使用された。図17の左下のパネルの図B(検出)は、1~32(正常12名および膵臓がん20名)からリストされた各テスト対象者のこれらのCGID(表9)のメチル化スコアの合計を示す。該ポリジーンスコアは、膵臓がんを有する人と正常な組織を有する人をカテゴリー的に区別する。図17の右下のパネルの図Cは、10種の異なる腫瘍(n=100)のある人からのデータを使用して腫瘍の特定の起源を検出するCGID(表10)のメチル化スコアを示す。これらの実施形態では、マーカーは他の起源のがんと膵臓がんをカテゴリー的に区別する。

【図18】図18は、TCGAメチル化データ(n=4854)における、膵臓がん対他のがんに対するポリジーンHKG-epi膵臓がん-検出およびスペックマーカーの精度および特異性の検証の図である。図18の図Aは、異なるがんを有する患者のDNAメチル化データを使用する、HKG-epi膵臓がん検出/スペックマーカーの検出率である。膵臓がんの特異性に注目すべきである。図18の図Bは、TCGAにおける4854名の患者からのDNAメチル化データを使用する、膵臓がんのHKG-epi膵臓がん-検出マーカーの特異性と感度のROCプロットである。図Cは、膵臓がん対他の起源のがんに対する感度と特異性を示す。

【図19】図19は、脳がん(膠芽腫)のポリジーンDNAメチル化マーカーの発見の図である。図19の図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表11)およびがんの特定起源を決定するためのCGID(表11)を使用する、本発明に開示される、脳がんの検出のための一連のCGIDの発見に使用された。左下のパネルの図B(検出/スペック)は、1~16(正常6名および脳がん10名)からリストされた各テスト対象者のこれらのCGID(表11)のメチル化スコアの合計を示す。該ポリジーンスコアは、脳がんを有する人、110名の別のがんを有する人と正常な組織を有する人をカテゴリー的に区別する。

【図20】図20は、TCGAメチル化データ(n=4854)における、脳がん対他のがんに対するポリジーンHKG-epi脳がん-検出およびスペックマーカーの精度および特異性の検証の図である。図Aは、異なるがんを有する患者のDNAメチル化データを使用する、HKG-epi脳がん検出/スペックマーカーの検出率である。脳がんの特異性に注目すべきである。図Bは、TCGAにおける4854名の患者からのDNAメチル化データを使用する、脳がんのHKG-epi脳がん-検出マーカーの特異性と感度のROCプロットである。図Cは、TCGAデータセット(n=695)における脳がんに対する感度と特異性を示す。

【図21】図21は、胃(gastric)[胃(stomach)]がんのポリジーンDNAメチル化マーカーの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表12)およびがんの特定起源を決定するためのCGID(表13)を使用する、本発明に開示される胃がんの検出のための一連のCGIDの発見に使用された。図21の左下のパネルの図B(検出)は、1~28(正常14名および胃がん20名)からリストされた各テスト対象者のこれらのCGID(表12)のメチル化スコアの合計を示す。該ポリジーンスコアは、胃がんを有する人と正常な組織を有する人をカテゴリー的に区別する。図21の右下のパネルの図C(スペック)は、10種の異なる腫瘍(n=100)のある人のポリジーンメチル化スコアを示す。これらの実施形態では、マーカーは他の起源のがんと胃がんをカテゴリー的に区別する。

【図22】図22は、TCGAメチル化データ(n=4817)における、胃がん対他のがんに対するポリジーンHKG-胃-検出およびスペックマーカーの精度および特異性の検証の図である。図Aは、異なるがんを有する患者のDNAメチル化データを使用する、HKG-epi胃がん検出/スペックマーカーの検出率である。胃がんの特異性に注目すべきである。図Bは、TCGAにおける4420名の患者からのDNAメチル化データを使用する、胃(胃がん)のHKG-epi胃がん-検出 スペック 1 マーカーの特異性

10

20

30

40

50

と感度のROCプロットである。図Cは、TCGAにおける4854名の患者からのDNAメチル化データを使用する、胃がんのHKG-epi胃がん-スベック1マーカーの特異性と感度のROCプロットである。注目すべきは、大腸がんおよび食道がんとの有意な交差反応性があり、それが共通の起源であることを証明していることである。

【図23】図23は、卵巣がんのポリジーンDNAメチル化マーカーの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表14)および起源特定のがんを決定するためのCGID(表15)を使用する、本発明に開示される卵巣がんの検出のための一連のCGIDの発見に使用された。図23の左下のパネルの図B(検出)は、1~15(正常5名および卵巣がん10名)からリストされた各テスト対象者のこれらのCGIDのメチル化スコアの合計を示す。該ポリジーンスコアは、卵巣がんを有する人と正常な組織を有する人をカテゴリー的に区別する。図23の右下のパネルの図Cは、11種の異なる腫瘍(n=110)のある人からのデータを使用して特定の腫瘍起源を検出するCGIDのメチル化スコアを示す。これらの実施形態では、マーカーは他の起源のがんと卵巣がんをカテゴリー的に区別する。

【図24】図24は、TCGAメチル化データ(n=6522)における、卵巣がん対他のがんに対するポリジーンHKG-epi卵巣がん-検出およびスベックマーカーの精度および特異性の検証の図である。図Aは、異なるがんを有する患者のDNAメチル化データを使用する、HKG-epi卵巣がん検出/スベックマーカーの検出率である。卵巣がんの特異性に注目すべきである。図Bは、TCGAにおける4723名の患者からのDNAメチル化データにおける、卵巣がんのHKG-epi卵巣がん-検出およびスベックマーカーの特異性と感度のROCプロットである。図Cは、卵巣がんに対する感度と特異性を示す。

【図25】図25は、子宮頸がんのポリジーンDNAメチル化マーカーの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表16)および起源特定のがんを決定するためのCG(表17)を使用する、本発明に開示される子宮頸がんの検出のための一連のCGIDの発見に使用された。図25の左下のパネルの図B(検出)は、1~30(正常20名および子宮頸がん10名)からリストされた各テスト対象者のこれらのCGID(表16)のメチル化スコアの合計を示す。該ポリジーンスコアは、子宮頸がんを有する人と正常な組織を有する人をカテゴリー的に区別する。図25の右下のパネルの図Cは、8種の異なる腫瘍(n=80)のある人からのデータを使用して腫瘍の特定起源を検出するCGID(表17)のメチル化スコアを示す。これらの実施形態では、マーカーは他の起源のがんと子宮頸がんをカテゴリー的に区別するが、大腸がんのいくつかの測定可能な検出に留意されたい。

【図26】図26は、TCGAメチル化データ(n=6522)における、子宮頸がん対他のがんに対するポリジーンHKG-子宮頸がん-検出およびスベックマーカーの精度および特異性の検証の図である。図Aは、異なるがんを有する患者のDNAメチル化データを使用する、HKG-子宮頸がん検出/スベックマーカーの検出率を示す。子宮頸がんの特異性に注目すべきである。図Bは、TCGAにおける4420名の患者からのDNAメチル化データを使用する、子宮頸がんのHKG-子宮頸がん-検出スベックマーカーの特異性と感度のROCプロットである。図Cは、子宮頸がんに対する感度と特異性を示す。

【図27】図27は、頭頸部扁平上皮がん(HNSC)のポリジーンDNAメチル化マーカーの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表18)および起源特定のがんを決定するためのCG(表19)を使用する、本発明に開示されるHNSCの検出のための一連のCGIDの発見に使用された。図27の左下のパネルの図Bは、1~140(がん10名、正常10名および他のがん120名)からリストされた各テスト対象者のこれらのCGID(表18)のメチル化スコアの合計を示す。図Cは、実施形態においてHNSCと正常組織サンプルをカテゴリー的に区別するとともに、他の起源からのがんとHNSCをカテゴリー的に区別するポリジーンスコアを示す。

10

20

30

40

50

【図28】図28は、TCGAメチル化データ(n=4166)における、HNSC対他のがんに対するポリジーンHKG-epiHNSC-検出/スベックマーカの精度および特異性の検証の図である。図Aは、異なるがんを有する患者のDNAメチル化データを使用する、HKG-epiHNSC 検出/スベックマーカの検出率である。HNSCの特異性に注目すべきである。図Bは、TCGAにおける4166名の患者からのDNAメチル化データにおけるHNSCのHKG-epiHNSC-検出マーカの特異性と感度のROCプロットである。図Cは、HNSC対他の起源のがんに対する感度と特異性を示す。

【図29】図29は、食道がんのポリジーンDNAメチル化マーカの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表20)およびがんの特定起源を決定するためのCGID(表21)を使用する、実施形態に開示される、食道がんの検出のための一連のCGIDの発見に使用された。図29の左下のパネルの図Bは、1~22(正常8名、がん14名)からリストされた各テスト対象者のこれらのCGID(表20)のメチル化スコアの合計を示す。図Cは、実施形態において食道がんと正常組織をカテゴリー的に区別するとともに、他の起源からのがんと1~220(がん20名、他のがん190名および正常血液10名)からリストされた食道がんをカテゴリー的に区別するポリジーンスコアを示す。

【図30】図30は、TCGAメチル化データ(n=7102)における、食道がん対他のがんに対するポリジーンHKG-epi食道がん-検出/スベックマーカの精度および特異性の検証の図である。図Aは、異なるがんを有する患者のDNAメチル化データを使用する、HKG-epi食道がん検出/スベックマーカの検出率を示す。食道がんの特異性に注目すべきである。図Bは、TCGAにおける4166名の患者のDNAメチル化データにおける食道がんのHKG-epi食道がん-検出マーカの特異性と感度のROCプロットである。図Cは、食道がん対他の起源のがんに対する感度と特異性を示す。

【図31】図31は、膀胱がんのポリジーンDNAメチル化マーカの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法(表22)およびがんの特定起源を決定するためのCGID(表23)を使用する、実施形態に開示される膀胱がんの検出のための一連のCGIDの発見に使用された。図31の左下のパネルの図B(検出)は、1~15(正常5名、膀胱がん10名)からリストされた各テスト対象者のこれらのCGID(表22)のメチル化スコアの合計を示す。図31の右下のパネルの図Cは、13種の異なる腫瘍(n=130)を有する人からのデータを使用して腫瘍の特定の起源を検出するCGID(表23)のメチル化スコアを示す。これらの実施形態では、マーカは、他の起源からのがんと膀胱がんとを区別する。これらのマーカによる大腸がんの測定可能な検出も注目すべきである。

【図32】図32は、TCGA(n=4723)における膀胱がん対他のがんに対するポリジーンHKG-epiBladder-検出およびスベックマーカの精度および特異性の検証の図である。図Aは、異なるがん(A)および膀胱がん(B)を有する患者のDNAメチル化データにおいて、HKG-epi膀胱がんスベックマーカ(A)および検出マーカ(B)の検出率を示す。図Cは、TCGAにおける4420名の患者からのDNAメチル化データを使用した、膀胱がんのHKG-epi膀胱がんスベックマーカの特異性と感度のROCプロットである。図Dは、膀胱がんのHKG-epi膀胱がん検出マーカの特異性と感度のROCプロットである(n=440)。

【図33】図33は、腎臓がんのポリジーンDNAメチル化マーカの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD(hypo)法を使用する、実施形態で開示された腎臓がんの検出およびがんの特定の起源の決定(表24)のための一連のCGIDの発見に使用された。図33の左下のパネルの図B(検出/スベック)は、1~226(その他のがん180名、健康な血液10名、正常な腎臓6名、腎がん30名)からリストされた各テスト対象者のこれらのCGID(表24)のメチル化スコアの合計を示す。これらの実施形態では、該ポリジーンスコアは腎臓がん、別のがんと正常な血液をカテゴリー的に区別する。

10

20

30

40

50

【図34】図34は、TCGA DNAメチル化データ (n = 7102) を使用する、腎臓がん対他のがんおよび正常組織に対するポリジーンHKG-epi腎臓がん-検出およびスペックマーカの精度および特異性の検証の図である。図Aは、異なるがんからのDNAメチル化データを使用する、HKG-epi腎臓がん検出/スペックマーカの検出率である。腎臓がんの特異性に注目すべきである。図Bは、TCGAにおける6367のがんのDNAメチル化データを使用する、腎臓がんのHKG-epi腎臓がん-検出スペックマーカの特異性と感度のROCプロットである。図Cは、腎臓 (renal) [腎臓 (kidney)] がんに対する感度と特異性を示す。さらに注目すべきは、脳、HCCおよび精巣がんとクロスオーバーである。

【図35】図35は、精巣がんのポリジーンDNAメチル化マーカの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD (hypo) 法を使用する実施形態で開示された精巣がんの検出および起源特定のがんの決定 (表25) のための一連のCGIDの発見に使用された。図35の左下のパネルの図B (検出/スペック) は、1~190 (精巣がん10名、その他のがんおよび血液180名) からリストされた各テスト対象者のこれらのCGID (表25) のメチル化スコアの合計を示す。これらの実施形態では、該ポリジーンスコアは精巣がん、正常な血液と別のがんをカテゴリー的に区別する。

【図36】図36は、TCGAメチル化データ (n = 7102) における、精巣がん対他の正常な組織およびがんに対するポリジーンHKG-epi精巣がん-検出およびスペックマーカの精度および特異性の検証の図である。図Aは、異なるがんを有する患者からのDNAメチル化データを使用する、HKG-epi精巣がん検出/スペックマーカの検出率を示す。精巣がんの特異性に注目すべきである。図Bは、TCGAにおける6367名の患者のDNAメチル化データを使用する、精巣がんのHKG-epi精巣がん-検出スペックマーカの特異性と感度のROCプロットである。図Cは、精巣がんに対する感度と特異性を示す。

【図37】図37は、13種の一般的ながんの汎がんポリジーンDNAメチル化マーカの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法 (表26) を使用する実施形態で開示された13種の一般的ながん (表26) (膀胱がん、脳がん、乳がん、子宮頸がん、大腸がんCRC、食道がん、肝臓がん、肺がん、卵巣がん、膵臓がん、前立腺がん、胃がん) の検出のための一連のCGIDの発見に使用された。図Bは、1~180 (がん170名、健康な血液10名) からリストされた各テスト対象者のこれらのCGIDのメチル化スコアの合計を示す。これらの実施形態では、該ポリジーンスコアはがんと正常な組織を区別する。

【図38】図38は、TCGAメチル化データ (n = 7102) における、ポリジーンHKG-epi汎がんマーカの精度および特異性の検証の図である。図Aは、TCGAデータを使用し、13人の異なるがん患者のepi汎がんポリジーンDNAメチル化マーカを使用して計算されたメチル化スコアを示す。図Bは、TCGAにおける4878名の患者からのすべてのがんのDNAメチル化データを使用する、HKG-汎がん検出およびスペックマーカの特異性と感度のROCプロットである。図Cは、がんを検出するための汎がんマーカの全体的な感度と特異性を示す。これらの実施形態では、1つ以上の色、例えば、オレンジ (加重メチル化スコア) および青 (サンプルあたり1つのBCDマーカの検出が陽性がんとしてスコア付けされる) が使用される。

【図39】図39は、黒色腫のポリジーンDNAメチル化マーカの発見の図である。図Aは、ソースと患者数をリストした表であり、これらの患者のメチル化データは、BCD法 (表45) を使用する実施形態で開示された黒色腫の検出 (表28) のための一連のCGIDの発見に使用された。図Bは、1~220 (その他のがんおよび健康な血液) および10名の黒色腫を有する患者からリストされた各テスト対象者のこれらのCGIDのメチル化スコアの合計である。これらの実施形態では、該ポリジーンスコアは黒色腫、別のがんそして正常な組織を区別する。

【図40】図40は、TCGAメチル化データ (n = 7102) における、黒色腫対他の

10

20

30

40

50

正常な組織およびがんに対するポリジーン H K G - e p i 黒色腫 - 検出およびスペックマーカの精度および特異性の検証の図である。図 A は、異なるがんを有する患者からの D N A メチル化データを使用する、H K G - e p i 黒色腫 検出 / スペックマーカの検出率を示す。黒色腫（肝臓がん、脳がんおよび前立腺がんとの重複検出）の特異性に注目すべきである。図 B は、T C G A にある 6 3 6 7 名の患者の D N A メチル化データを使用する、黒色腫の H K G - 黒色腫 - 検出 スペックマーカの特異性と感度の R O C プロットである。図 C は、黒色腫に対する感度と特異性を示す。

【図 4 1】図 4 1 は、血液がん（急性骨髄性白血病（A M L））のポリジーン D N A メチル化マーカの発見の図である。図 A は、ソースと患者数をリストした表であり、これらの患者のメチル化データは、B C D 法（表 2 7）を使用する実施形態で開示された血液がん A M L の検出（表 2 7）のための一連の C G I D の発見に使用された。図 B は、1 ~ 1 0（健康な血液）および 1 0 名の A M L を有する患者からリストされた各テスト対象者のこれらの C G I D のメチル化スコアの合計である。これらの実施形態では、該ポリジーンスコアは A M L と正常な血液を区別する。

【図 4 2】図 4 2 は、G S E 8 6 4 0 9（n = 7 9）および T C G A（n = 1 4 0）における A M L 対 G S E 4 0 2 7 9 および G S E 6 1 4 9 6（n = 9 6 8）における正常な血液に対するポリジーン H K G - e p i A M L - 検出およびスペックマーカの精度および特異性の検証の図である。図 A は、A M L を有する患者及び健康な血液を有する者からの D N A メチル化データを使用する、H K G - e p i A M L 検出 / スペックマーカの検出率を示す。図 B は、G S E 8 6 4 0 9（n = 7 9）、T C G A（n = 1 4 0）、G S E 4 0 2 7 9 および G S E 6 1 4 9 6（n = 9 6 8）からの D N A メチル化データを使用する、A M L の H K G - A M L - 検出 スペックマーカの特異性と感度の R O C プロットである。図 C は、A M L に対する感度と特異性を示す。

【図 4 3】図 4 3 は、正常な人に由来する血漿において B C D 特性 ~ 0 のメチル化を示す異なるがんを検出するために選択されたプライマーの検証の図である（各サンプルは正常な患者からの血漿の混合物である）。特定の C G をターゲットとする第一の P C R 1 反応は、シーケンスターゲットプライマーを使用して実行された。第二の P C R の後、増幅されたフラグメントを精製し、次世代シーケンスを行った。D N A メチル化は、示された各々の C G I D 位置で定量化された。

【図 4 4】図 4 4 は、正常な人に由来する血漿において B C D 特性 ~ 0 のメチル化を示す異なるがんを検出するために選択された、示されたプライマーの検証の図である（各サンプルは正常な患者からの血漿の混合物である）。

【図 4 5】図 4 5 は、多重増幅およびシーケンスのためのプライマー設計の図である。第一の P C R 反応は対象となる特定の関心領域をターゲットにするが、P C R 1 プライマーは第二の P C R 2 プライマーに相補的な配列を持っていることに注意されたい。プライマーの第二セットは、各患者のインデックスと、リバーズおよびフォワードシーケンスプライマーを導入する。

【図 4 6】図 4 6 は、前立腺がんを検出するための P C R 条件の最適化の図である。右側のパネルには、前立腺がんの 3 つのマーカ H I F 3 A 2 3 2 b p、T P M 4 2 1 3 b p、および C T T N 1 9 9 b p について、示されている D N A のようなさまざまなプライマー濃度を使用したマルチプレックス P C R 1 反応が示されている。

【図 4 7】図 4 7 は、D N A メチル化レベルを決定するためのバイオフィォマティクスワークフローの図である。P C R 2 の産物を組み合わせ、定量化および精製し、M i s e q I l l u m i n a シーケンサーで次世代シーケンスを行う。シーケンスは逆多重化（d e m u l t i p l e x e d）され、F A S T Q ファイルは患者ごとに生成され、スキームに示されているワークフローで分析される。D N A メチル化スコアは患者ごとに計算される。

【発明を実施するための形態】

【0 1 3 4】

詳細な説明

10

20

30

40

50

図面のすべての説明は、選択された実施形態を説明するためのものであり、請求される主題の範囲を限定することを意図するものではない。

【0135】

実施形態1：正常組織および血液DNAにおける数百人の個体にわたるカテゴリー的にメチル化されていないCGIDの発見

【0136】

腫瘍に由来する無細胞DNAは、血漿、尿などの体液や糞中に見られることが知られている。CF腫瘍DNAのDNAメチル化プロファイルが腫瘍DNAに類似していることも確立されている(Dominguez-Vigil et al., 2018)。膨大な量のデータにより、腫瘍DNAは正常組織と比較して区別的にメチル化されていることが確立されている(Luczak & Jagodzinski, 2006)。したがって、多くのグループは、がん性の組織とその正常な起源の組織、例えば肝臓がんと隣接する肝臓組織との間で区別的にメチル化されているDNAのCGID位置(Illumina 450K マニフェストのCG ID)をロジスティック回帰によって描写しようとした。ただし、これらの方法は、カテゴリー的な質的な違いではなく、がんと非形質転換組織の間の量的な違いを測定するため、腫瘍と正常組織の間の量的な違いが正常組織からのCF DNAによって希釈および消去され、偽陰性と感度低下につながる。さらに、分析に含まれなかった他の組織は、腫瘍DNAと同様のDNAメチル化プロファイルを持っている可能性があり、そしてほとんどの研究は腫瘍DNAを、他の組織ではなくその非形質転換対応物とのみ比較するため、これは偽陽性につながる可能性がある。異なる組織からの変動する予測不可能な量のDNAがCF DNAで検出されている(Breitbach et al., 2014)ため、測定されたDNAメチル化は、異なるソースからの組織DNAと腫瘍DNAの未知の予測不可能な混合物の複合を反映する。数千の腫瘍サンプルがIllumina 450Kアレイを使用したゲノムワイドDNAメチル化分析にかけられており、パブリックドメイン(TCGA)で発見されている。多くの正常組織およびがん組織のメチル化のプロファイルを調べると、発明者らは、すべての正常組織で完全にメチル化されていないが、腫瘍のDNAでメチル化されているゲノムにCGの重要なグループがあることを発見した。これらのサイトのサブセットは、DNAメチル化がパブリックドメインでプロファイルされた多数の個人にわたってメチル化されていない。本発明者らはまた、多くのがんにおいて、これらの強力に非メチル化されたサイトががんにおいてメチル化されることに気づいた。したがって、腫瘍DNAと血液中找到る可能性のある他のすべてのDNAとの間に質的な「カテゴリー的な違い」を生み出す。深い次世代シーケンスを使用すると、完全にメチル化されていないコピーのバックグラウンドで、わずかなメチル化分子でも簡単に識別できる。

【0137】

データベース； Illumina 450K DNAメチル化データ

遺伝子発現オムニバス(Gene Expression Omnibus)(GEO) <https://www.ncbi.nlm.nih.gov/geo/>またはがんゲノムアトラス(The Cancer Genome Atlas) TCGA <https://cancergenome.nih.gov/> 公開データベースのいずれかに委託された多数の個人からのヒトゲノム全体の~450,000CGのメチル化の正規化されたベータ値の公開されているデータベースを使用した。次のデータベースを使用して、多くの正常組織および血液DNA：GSE61496、GSE40279にある強力な非メチル化CG IDのリストを取得した。

【0138】

白血球からのDNAは、血漿中のCF DNAの主要なソースの1つである。発明者らはまず、GSE50192のIllumina 450KデータとExcelの論理式COUNTIFおよびIF関数を使用して、17の異なる体細胞ヒト組織のすべての個体でメチル化されていない47981個のCGIDのリストを生成した：

$NmCGID_x = COUNTIF(betaCGID_{x n_1:n_i}, "0.1")$

10

20

30

40

50

$umCGID_x = IF(NmCGID_x = 0, TRUE, FALSE)$

$NmCGID_x$ = メチル化された $CGID_x$ を有する正常な被験者の数。

$umCGID_x$ = すべての被験者における非メチル化 $CGID_x$

$betaCGID_x$ = 与えられた $CGID_x$ のメチル化値

$x = Illumina\ 450k$ アレイ上の任意の $CGID$

n_1 = アレイにおける最初のサブジェクト

n_i = アレイにおける最後のサブジェクト。

【0139】

次に、発明者らは、同じ基準を使用して、312個体からの血液DNA中の68260個の非メチル化 $CGID$ ($UMCGID$) のリストを生成した。次に、発明者らは、47981と68260の $CGID$ のリストを重ね、すべての個人の血液組織と体細胞組織の両方でメチル化されていない33477の $CGID$ のリストを取得した(図1A)。非メチル化 $CGID$ のこのリストの頑健性を高めるために、発明者らは、19歳から101歳までの656人の男性と女性($GSE40279$)の全血DNAの $Illumina\ 450K$ アレイにおける非メチル化 $CGID$ の60,379の $CGID$ のリストを描いた。これらは、数百人の性別や年齢に依存しない血液中における強力にメチル化されていない部位である。この60,379の $CGID$ のリストは、体組織と血液の両方でメチル化されていない33,477の $CGID$ のリストと重複して、がんのカテゴリー的なメチル化マーカーの発見に使用された28,754の $CGID$ の最終リストを生成した。このリストには、組織および個人間で強力にメチル化されていない $CGID$ の位置が含まれる。

【0140】

がんと正常組織との間でカテゴリー的に異なるDNAメチル化位置を特定するために、発明者らはこれらの28754の $CGID$ のいずれかが異なるがんでメチル化されているかどうかを調べた。発明者らは、腫瘍DNAメチル化データの調査に続いて、これらの28754の $CGID$ のサブセットのメチル化が、個々の患者からの腫瘍DNAにおいて一般的であることに気付いた。ただし、すべての個人が同じ位置でメチル化されているわけではない。そのため、高い特異性でがんを検出するには、 $CGID$ の組み合わせが必要である。したがって、本発明者らは、がんの検出のための $CGID$ のポリジーンの組み合わせを発見した。

【0141】

発明者らは、TCGAまたはGEOのパブリックドメインの10~50個のDNAメチル化プロファイルを「発見セット」として使用し、メチル化状態が腫瘍と正常組織の間で「カテゴリー的に」異なり、最高の感度と特異性でがんを検出できる $CGID$ のポリジーンセットを発見した。次に、これらの $CGID$ を「検証セット」として数百のTCGAおよびGEO腫瘍DNAメチル化アレイデータでテストし、実施形態2に開示されているように、がんを検出するためのポリジーンDNAメチル化マーカーの感度と特異性を検証した。

【0142】

実施形態2：無細胞DNA中のがんを検出するためのバイナリーカテゴリー区別(BCD)法。

ヒトゲノム全体の~450,000の CG ($CGID$) のメチル化の正規化されたベータ値の公開されている以下のデータベースを使用して、がん特有のDNAメチル化マーカーのリストを導き出した。

【0143】

【表 29】

表 29 肝臓がん			
病状	ソース	コーホート N	検出/スペック
正常な肝臓	GSE61258	発見	79 検出
HCC	TCGA	発見	50 検出
HCC	TCGA	発見	10 スペック
非HCCがん	TCGA	発見	80 スペック
HCC	GSE75041	検証	66 検出
HCC	GSE76269	検証	227 検出
HCC	TCGA	検証	430 検出
膀胱がん	TCGA	検証	439 スペック
脳がん	TCGA	検証	689 スペック
乳がん	TCGA	検証	891 スペック
子宮頸がん	TCGA	検証	312 スペック
CRCがん	TCGA	検証	459 スペック
ESCAがん	TCGA	検証	202 スペック
腎臓がん	TCGA	検証	871 スペック
肺がん	TCGA	検証	919 スペック
卵巣がん	TCGA	検証	10 スペック
PANCがん	TCGA	検証	195 スペック
PRADがん	TCGA	検証	553 スペック
胃がん	TCGA	検証	397 スペック
精巣がん	TCGA	検証	156 スペック
正常な肝臓	GSE76269	検証	10 スペック
正常な肝臓	GSE69852	検証	6 スペック
正常な肝臓	GSE75041	検証	10 スペック
正常な血液	GSE40279	検証	656 スペック
正常な血液	GSE61496	検証	312 スペック

10

20

30

40

50

【表 3 0】

表 30 肺がん				
病状	ソース	コーホート	N	検出/スペック
肺がん	GSE63704	発見	10	検出
正常な肺	GSE66836	発見	10	検出
肺がん	TCGA	発見	10	スペック
非肺がん	TCGA	発見	80	スペック
肺がん	GSE66836	検証	164	検出
肺がん	GSE63704	検証	17	検出
肺がん	GSE76269	検証	56	検出
肺がん	TCGA	検証	919	スペック
膀胱がん	TCGA	検証	439	検出
脳がん	TCGA	検証	689	スペック
乳がん	TCGA	検証	891	スペック
子宮頸がん	TCGA	検証	312	スペック
CRC	TCGA	検証	459	スペック
ESCA	TCGA	検証	202	スペック
HCC	TCGA	検証	430	スペック
腎臓がん	TCGA	検証	871	スペック
卵巣がん	TCGA	検証	10	スペック
PANC	TCGA	検証	195	スペック
PRAD	TCGA	検証	553	スペック
胃がん	TCGA	検証	396	スペック
精巣がん	TCGA	検証	156	スペック
正常な肺	GSE63704	検証	112	検出

10

20

30

40

50

【表 3 1】

表 31 前立腺がん				
病状	ソース	コーホート	N	検出/スペック
コントロール	GSE52955	発見		5 検出
PRAD	TCGA	発見		10 検出
PRAD	TCGA	発見		10 スペック
非PRAD	TCGA	発見		80 スペック
PRAD	GSE73549	検証		77 検出
PRAD	GSE52955	検証		25 検出
PRAD	TCGA	検証		553 スペック
膀胱	TCGA	検証		439 スペック
脳	TCGA	検証		689 スペック
乳	TCGA	検証		891 スペック
子宮頸	TCGA	検証		312 スペック
CRC	TCGA	検証		459 スペック
ESCA	TCGA	検証		202 スペック
HCC	TCGA	検証		430 スペック
腎臓	TCGA	検証		871 スペック
肺	TCGA	検証		919 スペック
卵巣	TCGA	検証		10 スペック
PANC	TCGA	検証		195 スペック
胃	TCGA	検証		397 スペック
精巣	TCGA	検証		156 スペック
正常な前立腺	GSE73549	検証		15 検出
正常な前立腺	GSE52955	検証		5 検出
正常な血液	GSE40279	検証		656 スペック
正常な血液	GSE61496	検証		312 スペック

10

20

30

40

50

【表 3 2】

表 32 乳がん				
病状	ソース	コーホート	N	検出/スペック
健康	GSE60185	発見	17	検出
乳がん	TCGA	発見	10	検出
乳がん	TCGA	発見	10	スペック
非乳がん	TCGA	発見	80	スペック
乳がん	GSE60185	検証	239	検出
乳がん	GSE75067	検証	188	検出
乳がん	TCGA	検証	891	検出
膀胱がん	TCGA	検証	439	スペック
脳がん	TCGA	検証	689	スペック
子宮頸がん	TCGA	検証	312	スペック
CRC	TCGA	検証	459	スペック
ESCA	TCGA	検証	202	スペック
HCC	TCGA	検証	430	スペック
腎臓がん	TCGA	検証	871	スペック
肺がん	TCGA	検証	919	スペック
卵巣がん	TCGA	検証	10	スペック
PANC	TCGA	検証	195	スペック
PRAD	TCGA	検証	553	スペック
胃がん	TCGA	検証	396	スペック
精巣がん	TCGA	検証	156	スペック
健康な乳房組織	GSE101961	検証	121	検出
健康な乳房組織	GSE60185	検証	17	検出
正常な隣接組織	GSE60185	検証	29	検出
正常な血液	GSE40279	検証	656	スペック
正常な血液	GSE61496	検証	312	スペック

10

20

30

40

50

【表 3 3】

表 33 大腸がんCRC			
病状	ソース	コーホート N	検出/スペック
コントロール	GSE6550	発見	25 検出
CRC	TCGA	発見	50 検出
CRC	TCGA	発見	10 スペック
非HCC	TCGA	発見	80 スペック
CRC	TCGA	検証	260 検出
HCC	TCGA	検証	459 検出
膀胱	TCGA	検証	439 スペック
脳	TCGA	検証	689 スペック
乳	TCGA	検証	891 スペック
子宮頸	TCGA	検証	312 スペック
ESCA	TCGA	検証	202 スペック
HCC	TCGA	検証	430 スペック
腎臓	TCGA	検証	871 スペック
肺	TCGA	検証	919 スペック
卵巣	TCGA	検証	10 スペック
PANC	TCGA	検証	195 スペック
PRAD	TCGA	検証	553 スペック
胃	TCGA	検証	397 スペック
精巣	TCGA	検証	156 スペック
正常な大腸	GSE6550	検証	8 検出
正常な血液	GSE40279	検証	656 スペック
正常な血液	GSE61496	検証	312 スペック

10

20

30

40

50

【表 3 4】

表 34 膵臓がん			
病状	ソース	コーホート N	検出/スペック
健康	GSE53051	発見	12 検出
膵臓がん	TCGA	発見	20 検出
膵臓がん	TCGA	発見	20 スペック
非膵臓がん	TCGA	発見	100 スペック
膵臓がん	E-MTAB-5571	検証	24 検出
膵臓がん	TCGA	検証	195 検出
膀胱がん	TCGA	検証	439 スペック
脳がん	TCGA	検証	689 スペック
乳がん	TCGA	検証	891 スペック
子宮頸がん	TCGA	検証	312 スペック
CRC	TCGA	検証	459 スペック
ESCA	TCGA	検証	202 スペック
HCC	TCGA	検証	430 スペック
腎臓がん	TCGA	検証	871 スペック
肺がん	TCGA	検証	919 スペック
卵巣がん	TCGA	検証	10 スペック
PRAD	TCGA	検証	553 スペック
胃がん	TCGA	検証	396 スペック
精巣がん	TCGA	検証	156 スペック
正常な血液	GSE40179	検証	656 スペック
正常な血液	GSE61496	検証	312 スペック

10

20

30

40

50

【表 3 5】

表 35 脳がん				
病状	ソース	コーホート	N	検出/スペック
脳がん	TCGA	発見	10	スペック-検出
非脳がん	TCGA	発見	158	スペック-検出
脳がん	GSE36278	検証	136	スペック-検出
脳がん	GSE58298	検証	40	スペック-検出
脳がん	GSE58218	検証	228	スペック-検出
脳がん	TCGA	検証	689	スペック-検出
膀胱がん	TCGA	検証	439	スペック-検出
乳がん	TCGA	検証	891	スペック-検出
子宮頸がん	TCGA	検証	312	スペック-検出
CRC	TCGA	検証	459	スペック-検出
ESCA	TCGA	検証	202	スペック-検出
HCC	TCGA	検証	430	スペック-検出
腎臓がん	TCGA	検証	871	スペック-検出
肺がん	TCGA	検証	919	スペック-検出
卵巣がん	TCGA	検証	10	スペック-検出
PANC	TCGA	検証	195	スペック-検出
PRAD	TCGA	検証	553	スペック-検出
胃がん	TCGA	検証	396	スペック-検出
精巣がん	TCGA	検証	156	スペック-検出
脳がん	GSE36278	検証	6	スペック-検出
正常な血液	GSE40279	検証	656	スペック-検出
正常な血液	GSE61496	検証	312	スペック-検出

10

20

30

40

50

【表 3 6】

表 36 胃がん				
病状	ソース	コーホート	N	検出/スペック
健康	GSE99553	発見	14	検出
胃がん	TCGA	発見	20	検出
胃がん	TCGA	発見	7	スペック
非胃がん	TCGA	発見	100	スペック
胃がん	TCGA	検証	396	検出
膀胱がん	TCGA	検証	439	スペック
脳がん	TCGA	検証	689	スペック
乳がん	TCGA	検証	891	スペック
子宮頸がん	TCGA	検証	312	スペック
CRC	TCGA	検証	459	スペック
ESCA	TCGA	検証	202	スペック
HCC	TCGA	検証	430	スペック
腎臓がん	TCGA	検証	871	スペック
肺がん	TCGA	検証	919	スペック
卵巣がん	TCGA	検証	10	スペック
PANC	TCGA	検証	195	スペック
PRAD	TCGA	検証	553	スペック
精巣がん	TCGA	検証	156	スペック
正常な組織	GSE99553	検証	42	検出
正常な血液	GSE40179	検証	656	スペック
正常な血液	GSE61496	検証	312	スペック

10

20

30

40

50

【表 37】

表 37 卵巣がん				
病状	ソース	コーホート	N	検出/スペック
健康	GSE65821	発見		6 検出
卵巣がん	TCGA	発見		10 検出
卵巣がん	TCGA	発見		10 スペック
非卵巣がん	TCGA	発見		110 スペック
卵巣がん	GSE65821	検証		113 検出
膀胱がん	TCGA	検証		439 スペック
脳がん	TCGA	検証		689 スペック
乳がん	TCGA	検証		891 検出
子宮頸がん	TCGA	検証		312 スペック
CRC	TCGA	検証		459 スペック
ESCA	TCGA	検証		202 スペック
HCC	TCGA	検証		430 スペック
腎臓がん	TCGA	検証		871 スペック
肺がん	TCGA	検証		919 スペック
PANC	TCGA	検証		195 スペック
PRAD	TCGA	検証		553 スペック
胃がん	TCGA	検証		396 スペック
精巣がん	TCGA	検証		156 スペック
健康な卵巣組織	GSE87621	検証		9 検出
健康な卵巣組織	GSE74845	検証		216 検出
健康な卵巣組織	GSE81228	検証		10 検出
正常な血液	GSE40279	検証		656 スペック

10

20

30

40

50

【表 3 8】

表 38 子宮頸がん				
病状	ソース	コーホート	N	検出/スペック
健康	GSE46306	発見	20	検出
子宮頸がん	TCGA	発見	10	検出
子宮頸がん	TCGA	発見	10	スペック
非子宮頸がん	TCGA	発見	110	スペック
子宮頸がん	GSE68339	検証	270	検出
子宮頸がん	TCGA	検証	312	検出
膀胱がん	TCGA	検証	439	スペック
脳がん	TCGA	検証	689	スペック
乳がん	TCGA	検証	891	スペック
CRC	TCGA	検証	459	スペック
ESCA	TCGA	検証	202	スペック
HCC	TCGA	検証	430	スペック
腎臓がん	TCGA	検証	871	スペック
肺がん	TCGA	検証	919	スペック
卵巣がん	TCGA	検証	10	スペック
PANC	TCGA	検証	195	スペック
PRAD	TCGA	検証	553	スペック
胃がん	TCGA	検証	396	スペック
精巣がん	TCGA	検証	156	スペック
正常な血液	GSE40279	検証	656	スペック
正常な血液	GSE61496	検証	312	スペック

10

20

30

40

50

【表 3 9】

表 39 HNSC				
病状	ソース	コーホート	N	検出/スペック
HNSC	TCGA	発見	20	検出
コントロール	GSE75537	発見	10	検出
HNSC	TCGA	発見	20	Spec
非HNSC	TCGA	発見	190	Spec
健康	GSE40279	発見	10	Spec
HNSC	TCGA	検証	580	検出
コントロール	GSE52068	検証	24	検出
HNSC	GSE75537	検証	54	検出
HNSC	GSE79556	検証	83	検出
コントロール	GSE75537	検証	54	検出
HNSC	GSE52068	検証	24	検出
膀胱	TCGA	検証	439	スペック
脳	TCGA	検証	689	スペック
乳	TCGA	検証	891	スペック
子宮頸	TCGA	検証	312	スペック
HCC	TCGA	検証	459	スペック
ESCA	TCGA	検証	202	スペック
HCC	TCGA	検証	430	スペック
腎臓	TCGA	検証	871	スペック
肺	TCGA	検証	919	スペック
卵巣	TCGA	検証	10	スペック
PANC	TCGA	検証	195	スペック
PRAD	TCGA	検証	553	スペック
胃	TCGA	検証	396	スペック
精巣	TCGA	検証	156	スペック
正常な血液	GSE40279	検証	656	スペック
正常な血液	GSE52068	検証	312	スペック

10

20

30

40

50

【表 4 0】

表 40 食道がん				
病状	ソース	コーホート	N	検出/スペック
食道がん	GSE52826	発見	6	検出
ESCA	TCGA	発見	10	検出
非食道がん	TCGA	発見	190	スペック
健康な血液	GSE40279	発見	10	スペック
ESCA	TCGA	発見	20	スペック
食道がん	TCGA	検証	202	検出
膀胱がん	TCGA	検証	439	スペック
脳がん	TCGA	検証	689	スペック
乳がん	TCGA	検証	891	スペック
子宮頸がん	TCGA	検証	312	スペック
CRCがん	TCGA	検証	459	スペック
腎臓がん	TCGA	検証	871	スペック
HNSC	TCGA	検証	580	スペック
HCC	TCGA	検証	430	スペック
肺がん	TCGA	検証	919	スペック
卵巣がん	TCGA	検証	10	スペック
膵臓がん	TCGA	検証	195	スペック
前立腺がん	TCGA	検証	553	スペック
胃がん	TCGA	検証	397	スペック
精巣	TCGA	検証	156	スペック
正常な血液	GSE40279	検証	656	スペック
正常な血液	GSE61496	検証	312	スペック
正常な組織	Mix	検証	115	スペック
正常な組織	GSE85566	検証	189	スペック

10

20

30

40

50

【表 4 1】

表 41 膀胱がん				
病状	ソース	コーホート	N	検出/スペック
膀胱がん	TCGA	発見	10	検出
正常な膀胱	GSE52955	発見	5	検出
膀胱がん	TCGA	発見	10	スペック
非膀胱がん	TCGA	発見	180	スペック
膀胱がん	TCGA	検証	439	検出
脳がん	TCGA	検証	689	スペック
乳がん	TCGA	検証	891	スペック
子宮頸がん	TCGA	検証	312	スペック
CRC	TCGA	検証	459	スペック
ESCA	TCGA	検証	202	スペック
HCC	TCGA	検証	430	スペック
腎臓がん	TCGA	検証	871	スペック
肺がん	TCGA	検証	919	スペック
卵巣がん	TCGA	検証	10	スペック
PANC	TCGA	検証	195	スペック
PRAD	TCGA	検証	553	スペック
胃がん	TCGA	検証	396	スペック
精巣がん	TCGA	検証	156	スペック
正常な組織	GSE50192	検証	70	スペック
正常な血液	GSE40279	検証	656	スペック
正常な血液	GSE61496	検証	312	スペック

10

20

30

40

50

【表 4 2】

表 42 腎臓がん				
病状	ソース	コーホート	N	検出/スペック
コントロール	GSE52955	発見	6	スペック-検出
腎臓がん	TCGA	発見	30	スペック-検出
健康な血液	GSE40279	発見	10	スペック-検出
非腎臓がん	TCGA	発見	180	スペック-検出
腎臓	TCGA	検証	871	スペック-検出
膀胱	TCGA	検証	439	スペック-検出
脳	TCGA	検証	689	スペック-検出
乳	TCGA	検証	891	スペック-検出
子宮頸	TCGA	検証	312	スペック-検出
CRC	TCGA	検証	459	スペック-検出
ESCA	TCGA	検証	202	スペック-検出
HNSC	TCGA	検証	580	スペック-検出
HCC	TCGA	検証	430	スペック-検出
肺	TCGA	検証	919	スペック-検出
卵巣	TCGA	検証	10	スペック-検出
PANC	TCGA	検証	195	スペック-検出
PRAD	TCGA	検証	553	スペック-検出
胃	TCGA	検証	397	スペック-検出
精巣	TCGA	検証	156	スペック-検出
正常な血液	GSE40279	検証	656	スペック-検出
正常な血液	GSE61496	検証	312	スペック-検出
正常な組織	Mix	検証	189	スペック-検出
正常な組織	GSE85566	検証	115	スペック-検出

10

20

30

40

50

【表 4 3】

表 43 精巣がん				
病状	ソース	コーホート	N	検出/スペック
精巣がん	TCGA	発見	10	スペック-検出
健康な血液	GSE40279	発見	10	スペック-検出
非精巣がん	TCGA	発見	170	スペック-検出
精巣がん	TCGA	検証	156	スペック-検出
膀胱がん	TCGA	検証	439	スペック-検出
脳がん	TCGA	検証	689	スペック-検出
乳がん	TCGA	検証	891	スペック-検出
子宮頸がん	TCGA	検証	312	スペック-検出
CRC	TCGA	検証	459	スペック-検出
ESCA	TCGA	検証	202	スペック-検出
HCC	TCGA	検証	430	スペック-検出
腎臓がん	TCGA	検証	871	スペック-検出
肺がん	TCGA	検証	919	スペック-検出
卵巣がん	TCGA	検証	10	スペック-検出
PANC	TCGA	検証	195	スペック-検出
PRAD	TCGA	検証	553	スペック-検出
胃がん	TCGA	検証	396	スペック-検出
正常な組織	GSE50192	検証	70	スペック-検出
正常な組織	Mix	検証	119	スペック-検出
正常な血液	GSE40279	検証	656	スペック-検出
正常な血液	GSE61496	検証	312	スペック-検出

10

20

30

40

50

【表 4 4】

表 44 汎がん				
病状	ソース	コーホート	N	検出/スペック
健康な血液	GSE40279	発見	10	検出
TCGAがん	TCGA	発見	170	検出
膀胱がん	TCGA	検証	439	検出
脳がん	TCGA	検証	689	検出
乳がん	TCGA	検証	891	検出
子宮頸がん	TCGA	検証	312	検出
CRCがん	TCGA	検証	459	検出
食道がん	TCGA	検証	202	検出
HNSCがん	TCGA	検証	580	検出
HCCがん	TCGA	検証	430	検出
肺がん	TCGA	検証	919	検出
卵巣がん	TCGA	検証	10	検出
膵臓がん	TCGA	検証	195	検出
前立腺がん	TCGA	検証	553	検出
胃がん	TCGA	検証	397	検出
正常な血液	GSE40279	検証	656	検出
正常な血液	GSE61496	検証	312	検出

10

20

【表 4 5】

表 45 黒色腫				
病状	ソース	コーホート	N	検出/スペック
膀胱がん	TCGA	検証	439	検出/スペック
脳がん	TCGA	検証	689	検出/スペック
乳がん	TCGA	検証	891	検出/スペック
子宮頸がん	TCGA	検証	312	検出/スペック
CRCがん	TCGA	検証	459	検出/スペック
食道がん	TCGA	検証	202	検出/スペック
HNSCがん	TCGA	検証	580	検出/スペック
HCCがん	TCGA	検証	430	検出/スペック

30

40

50

【表 4 6】

表 46 AML				
病状	ソース	コーホート	N	検出/スペック
正常な血液	GSE40279	検証	656	検出/スペック
正常な血液	GSE61496	検証	312	検出/スペック
AML	GSE86409	検証	79	検出/スペック
AML	TCGA	検証	140	検出/スペック

10

【0144】

BCD法

以下は、異なるがんの早期予測のためのポリジーンDNAメチル化マーカーを発見するために実施形態で使用されるバイナリーカテゴリー区別法（BCD）のステップである。

【0145】

正常な組織で強力にメチル化されていない28,754個のCGIDをフィルター処理した。

発見コーホートについては、Microsoft ExcelのCOUNTIFおよびIF関数を使用して、正常な組織で強力にメチル化されていない28,754個のCGIDのリスト内に描いた。該CGIDは特定のがんではカテゴリー的にメチル化され、影響を受けていない組織および正常な組織ではメチル化されない。

20

$NmcCGIDx = COUNTIF(\text{beta}CGIDxCancer\ n1:ni, "0.2")$

$NmnCGIDx = COUNTIF(\text{beta}CGIDxNormal\ n1:ni, "0.1")$

$DMCGIDx = IF((AND(NmcCGIDx = 0, NmnCGIDx = 0)), "TRUE", "FALSE")$

DM CGIDxは最も高い番号から最も低い番号へとソートされた。

上位20までのTRUE DM CGIDxの位置が選択された。

$NmcCGIDx$ = メチル化されたCGIDxを有するがん患者の数

Nmn = メチル化されたCGIDxを持つ正常な隣接または類似の組織サンプルの数

$\text{beta}CGIDx$ = CGIDxのメチル化のレベル

30

$n = 1$ から i までの患者

DM = 区別的にメチル化されたCGIDx

【0146】

本発明者らは、精巣がんおよび腎臓がんがすべての組織で高度にメチル化されているCGIDでメチル化の広範な欠如を示すことに気付いた。したがって、がんではカテゴリー的にメチル化されておらず正常な組織ではメチル化されることに対して、「BCD hypo」と呼ぶ改良したBCD法を使用して、精巣がんおよび腎臓(kidney) [腎臓(renal)]がんのカテゴリー的に区別的にメチル化されたCGIDの位置を発見した。以下のステップは、精巣がんと腎臓がんにおいて区別的に低(hypo)メチル化されたCGIDの位置を発見するために使用された。

40

【0147】

発見コーホートでは、ExcelのCOUNTIFおよびIF関数を使用して、正常な組織で完全にメチル化されている精巣または腎臓における低メチル化CGIDを描いた。

$NucCGIDx = COUNTIF(\text{beta}CGIDxCancer\ n1:ni, "0.2")$

$NunCGIDx = COUNTIF(\text{beta}CGIDxNormal\ n1:ni, "0.9")$

$DHMcGIDx = IF((AND(NucCGIDx = 0, NunCGIDx = 0)), "TRUE", "FALSE")$

DHM CGIDの位置が最も高い番号から最も低い番号へとソートされた。

上位20のTRUE DHMのサイトが選択され、罰則付き回帰分析が行われた。

$NucCGIDx$ = 非メチル化CGID Xのがん患者の数

$NunCGIDx$ = 非メチル化CGID Xの正常な組織サンプルの数

50

$n = 1$ から i までの患者

DHM = 区別的に低メチル化された CGID

【0148】

次に、発明者らは、上位20のDM（またはDHM）CGID \times で、Rで罰則するパッケージを使用して罰則付き回帰を実行し、最高の感度と特異性でがんを予測するCGID \times の最小の組み合わせを描いた。ポリジーンの組み合わせとがんにおけるこれらのCGIDのメチル化レベル間の回帰係数を決定するために、CGID \times のポリジーンの組み合わせが多変量線形回帰方程式でさらにテストされた。このモデルを使用して、典型的ながんの各患者のメチル化スコアを計算した。

$$Ms = \alpha + \sum_{i=1}^n \beta_i CG_i$$

10

Ms = メチル化スコア、 α = 切片、 β_i = CG ID $_i$ の係数、 CG_i = CGあたりのメチル化レベルの組み合わせ。1から i = 組み合わせたCGの数。

【0149】

実施形態3：肝臓がん（HCC）のポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE61258（正常な肝臓）から、そしてHCC DNAメチル化データのTCGA HCCコレクションからランダムに選択された66個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」
コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGID \times を候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性でHCCを検出する（図5B、表1）（検出）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、8種の異なる腫瘍タイプを表すTCGAからランダムに選択された80個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、HCCと他の腫瘍との間で区別的にメチル化された
CGIDのポリジーンセットを発見した（図5C、表2）（スペック）。実施形態2で説明したように、加重DNAメチル化スコアがCGIDのために開発された。

20

30

【0150】

実施形態4：HCCを検出するためのHCCポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、加重HCC DNAメチル化スコアが、表1のCGIDの227人のHCC患者のGSE76269からの正規化されたIllumina 450K DNAメチル化ベータ値を含む「検証コーホート」でHCCを検出したことを実証した。この方法を使用して、95%のHCCサンプルがHCCとして検出された（図6C）。図6Aに示されるROC曲線は、がんを検出するためのこのメチル化スコアの特異性（1）および感度（0.96）を明らかにしている。次に、発明者らは、HCCと他の8種類のがんのGSE75041とTCGAからのメチル化データと共に「検証コーホート」を使用して、HCCの検出およびHCCと他のがんを区別するための、スペックおよび検出のDNAメチル化スコアを組み合わせる有用性を実証した。図7Bに示されるROC曲線は、HCCを他の正常な組織および他のがんと区別するためのこのメチル化スコアの特異性（0.97）および感度（0.95）を明らかにしている。これらのDNAメチル化マーカーと計算されたメチル化スコアは、組織、糞便、唾液、血漿および尿などの人とは異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんのスクリーニングと早期発見に使用できる。

40

【0151】

50

実施形態 5：肺がんのポリジーン DNA メチル化マーカーの発見。

本発明者らは、GSE61258（正常な肺）からの 10 人の、そして肺がん DNA メチル化データの TCGA 肺がんコレクションからランダムに選択された 10 個のサンプルからの正規化された Illumina 450K DNA メチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態 1 で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット 28754 CGID を候補リストに挙げた。次に、発明者らは、実施形態 2 に記載の BCD 法を使用して、トレーニングコーホートにおいて高い感度および特異性で肺がん（腺がんおよび扁平上皮がんの両方を含むサンプル）を検出する（図 8 B、表 3）（検出）、バイナリーカテゴリー区分的にメチル化された CGID のポリジーンセットを発見した。実施形態 2 で説明したように、加重 DNA メチル化スコアおよびがんの閾値が CGID のために開発された。次に、発明者らは、8 種の異なる腫瘍タイプを表す TCGA からランダムに選択された 80 個の DNA メチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、肺がんと他の腫瘍との間で区分的にメチル化された CGID のポリジーンセットを発見した（図 8 C、表 4）（スペック）。実施形態 2 で説明したように、加重 DNA メチル化スコアが CGID のために開発された。

10

【0152】

実施形態 6：肺がんを検出するための肺がんポリジーン DNA メチル化マーカーの有用性。

20

次に、発明者らは、実施形態 5（検出）で開発された加重肺がん DNA メチル化スコアおよび閾値が、GSE66836、GSE63704、GSE76269 からの、および TCGA の 919 人の肺がん患者からの正規化された Illumina 450K DNA メチル化ベータ値が含まれる「検証コーホート」で肺がんを検出することを実証した。この方法を使用して、96% の肺がんサンプルが肺がんとして検出された（図 9 A）。次に、発明者らは、肺がんと他の 8 種類のがんの GSE と TCGA からのメチル化データと共に「検証コーホート」を使用して、肺がんと他のがんを区別するための、スペックおよび検出の DNA メチル化スコアを組み合わせた有用性を実証した（図 9 A）。図 9 B に示される ROC 曲線は、肺がんを他の正常な組織および他のがんから検出するための、このメチル化スコアの特異性（0.96）および感度（0.84）を明らかにしている（図 9 C）。これらの DNA メチル化マーカーと計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

30

【0153】

実施形態 7：前立腺がんのポリジーン DNA メチル化マーカーの発見。

本発明者らは、GSE52955（正常な前立腺）からの 5 人の、そして前立腺がん DNA メチル化データの TCGA 前立腺がんコレクションからランダムに選択された 10 個のサンプルからの正規化された Illumina 450K DNA メチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態 1 で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット 28754 CGID を候補リストに挙げた。次に、発明者らは、実施形態 2 に記載の BCD 法を使用して、トレーニングコーホートにおいて高い感度および特異性で前立腺がんを検出する（図 10 B、表 5）（検出）、バイナリーカテゴリー区分的にメチル化された CGID のポリジーンセットを発見した。実施形態 2 で説明したように、加重 DNA メチル化スコアおよびがんの閾値が CGID のために開発された。次に、発明者らは、8 種の異なる腫瘍タイプを表す TCGA からランダムに選択された 80 個の DNA メチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、前立腺がんと他の腫瘍との間で区分的にメチル化された CGID のポリジーンセットを発見した（図 10 C、表 5）（スペック）。実施形態 2 で説明したように、加重 DNA メチル化スコアが CGID のために

40

50

開発された。

【0154】

実施形態8：前立腺がんを検出するための前立腺がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態7（検出）で開発された加重前立腺がんDNAメチル化スコアおよび閾値が、GSE73549、GSE2955からの、およびTCGAの430人の前立腺がん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で前立腺がんを検出することを実証した。この方法を使用して、99%の前立腺がんサンプルが前立腺がんとして検出された（図11A）。次に、発明者らは、前立腺がんと他の8種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、前立腺がんと他のがんを区別するための、スペックおよび検出のDNAメチル化スコアを組み合わせた有用性を実証した（図11A）。図11Bに示されるROC曲線は、前立腺がんを他の正常な組織および他のがんから検出するための、このメチル化スコアの特異性（0.99）および感度（0.98）を明らかにしている（図11C）。これらのDNAメチル化マーカーと計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

10

【0155】

実施形態9：乳がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE60185（正常な乳房）からの17人の、そして乳がんDNAメチル化データのTCGA乳がんコレクションからランダムに選択された10個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で乳がんを検出する（図12B、表6）（検出）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、8種の異なる腫瘍タイプを表すTCGAからランダムに選択された80個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、乳がんと他の腫瘍との間で区別的にメチル化されたCGIDのポリジーンセットを発見した（図12C、表7）（スペック）。実施形態2で説明したように、加重DNAメチル化スコアがCGIDのために開発された。

20

30

【0156】

実施形態10：乳がんを検出するための乳がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態9（検出）で開発された加重乳がんDNAメチル化スコアおよび閾値が、GSE60185、GSE75067からの、およびTCGAからの正規化されたIllumina 450K DNAメチル化ベータ値を使用する891人の乳がん患者が含まれる「検証コーホート」で乳がんを検出することを実証した。この方法を使用して、91%の乳がんサンプルが乳がんとして検出され（図13A）、DCISと浸潤性がんの両方が検出された。次に、発明者らは、乳がんと他の8種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、乳がんと他のがんを区別するための、スペックおよび検出のDNAメチル化スコアを組み合わせた有用性を実証した（図14A）。図14Bに示されるROC曲線は、乳がんを他の正常な組織および他のがんと区別するためのこのメチル化スコアの特異性（0.89）および感度（0.87）を明らかにしている（図14C）。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある女性や一般的な健康な人における乳がんの早期

40

50

発見に使用できる。

【0157】

実施形態11：大腸がん（CRC）のポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE(32146)（正常）からの25人の、そして大腸がんDNAメチル化データのTCGA大腸がんコレクションからランダムに選択された50個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で大腸がんを検出する（図15B、表8）（検出）、バイナリーカテゴリー区分的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、8種の異なる腫瘍タイプを表すTCGAからランダムに選択された80個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、大腸がんと他の腫瘍との間で区分的にメチル化されたCGIDのポリジーンセットを発見した（図15C、表8）（スペック）。実施形態2で説明したように、加重DNAメチル化スコアがCGIDのために開発された。

10

【0158】

実施形態12：大腸がんを検出するための大腸がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態11（検出）で開発された加重大腸がんDNAメチル化スコアおよび閾値が、GSE69550からの、およびTCGAの459人の大腸がん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で大腸がんを検出することを実証した。この方法を使用して、98%の大腸がんサンプルが大腸がんとして検出された（図16A）。次に、発明者らは、大腸がんと他の8種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、大腸がんと他のがんを区別するための、スペックおよび検出のDNAメチル化スコアを組み合わせた有用性を実証した（図16A）。図16Bに示されるROC曲線は、大腸がんを他の正常な組織および他のがんを区別するためのこのメチル化スコアの特異性（0.96）および感度（0.98）を明らかにしている（図16C）。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、CRCのリスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

30

【0159】

実施形態13：膵臓がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE53051（正常）からの12人の、そして膵臓がんDNAメチル化データのTCGAコレクションからランダムに選択された20個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で膵臓がんを検出する（図17B、表9）（検出）、バイナリーカテゴリー区分的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、10種の異なる腫瘍タイプを表すTCGAからランダムに選択された100個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、膵臓がんと他の腫瘍との間で区分的にメチル化さ

40

50

れたCGIDのポリジーンセットを発見した(図17C、表10)(スペック)。実施形態2で説明したように、加重DNAメチル化スコアがCGIDのために開発された。

【0160】

実施形態14：膵臓がんを検出するための膵臓がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態13(検出)で開発された加重膵臓がんDNAメチル化スコアおよび閾値が、TCGAからの891人の膵臓がん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で膵臓がんを検出することを実証した。この方法を使用して、86%の膵臓がんサンプルが膵臓がんとして検出された(図18A)。次に、発明者らは、膵臓がんと他の9種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、膵臓がんと他のがんを区別するための、スペックおよび検出のDNAメチル化スコアを組み合わせた有用性を実証した(図18A)。図18Bに示されるROC曲線は、膵臓がんの検出および膵臓がんを他の正常な組織および他のがんと区別するためのこのメチル化スコアの特異性(0.93)および感度(0.86)を明らかにしている(図18C)。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

10

【0161】

実施形態15：脳がんのポリジーンDNAメチル化マーカーの発見。

20

本発明者らは、GSE65820(正常)からの10人の、そして脳がんDNAメチル化データのTCGAコレクションからランダムに選択された10個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で脳がんを検出する(図19B、表11)(検出)、バイナリーカテゴリー区別的にメチル化されたCGIDのセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、11

30

【0162】

実施形態16：脳がんを検出するための脳がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態15(検出)で開発された加重脳がんDNAメチル化スコアおよび閾値が、TCGAからの689人の脳がん患者、GSE58298からの40人の患者およびGSE36278からの136人の患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で脳がんを検出することを実証した。この方法を使用して、91%~97%の脳がんサンプルが脳がんとして検出された(図20A)。次に、発明者らは、脳がんと他の9種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、脳がんと他のがんを区別するための、同様のCGIDの有用性を実証した(図20A)。図20Bに示されるROC曲線は、脳がんを他の正常な組織および他のがんから検出するためのこのメチル化スコアの特異性(1)および感度(0.97)を明らかにしている(図20C)。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一

40

50

一般的な健康な人におけるがんの早期発見に使用できる。

【0163】

実施形態17：胃がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE99553（正常）からの18人の、そして胃がんDNAメチル化データのTCGAコレクションからランダムに選択された10個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で胃がんを検出する（図21B、表12）（検出）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、11種の異なる腫瘍タイプを表すTCGAからランダムに選択された100個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、胃がんと他の腫瘍との間で区別的にメチル化されたCGIDのポリジーンセットを発見した（図21C、表13）（スペック）。実施形態2で説明したように、加重DNAメチル化スコアがCGIDのために開発された。

10

【0164】

実施形態18：胃がんを検出するための胃がんポリジーンDNAメチル化マーカーの有用性。

20

次に、発明者らは、実施形態17（検出）で開発された加重胃がんDNAメチル化スコアおよび閾値が、TCGAからの397人の胃がん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で胃がんを検出することを実証した。この方法を使用して、88%の胃がんサンプルが胃がんとして検出された（図23A）。次に、発明者らは、胃がんと他の10種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、胃がんと他のがんを区別するための、スペックおよび検出のDNAメチル化スコアを組み合わせた有用性を実証した（図23A）。図22Bに示されるROC曲線は、胃がんを他の正常な組織および他のがんから検出するためのこのメチル化スコアの特異性（0.9）および感度（0.9）を明らかにしている（図22C）。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんのスクリーニングと早期発見に使用できる。

30

【0165】

実施形態19：卵巣がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE65820（正常）からの5人の、そして卵巣がんDNAメチル化データのTCGAコレクションからランダムに選択された10個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で卵巣がんを検出する（図23B、表14）（検出）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、10種の異なる腫瘍タイプおよび血液を表すTCGAからランダムに選択された100個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、卵巣がんと他の腫瘍との間で区別的にメチル化されたCGIDのポリジーンセットを発見した（図23C、表15）（スペック

40

50

）。実施形態 2 で説明したように、加重 DNA メチル化スコアが CGI D のために開発された。

【 0 1 6 6 】

実施形態 2 0 : 卵巣がんを検出するための卵巣がんポリジーン DNA メチル化マーカーの有用性。

次に、発明者らは、実施形態 1 9 (検出) で開発された加重卵巣がん DNA メチル化スコアおよび閾値が、TCGA からの 1 1 4 人の卵巣がん患者からの正規化された Illumina 4 5 0 K DNA メチル化ベータ値が含まれる「検証コーホート」で卵巣がんを検出することを実証した。この方法を使用して、8 6 % の卵巣がんサンプルが卵巣がんとして検出された (図 2 4 A) 。次に、発明者らは、卵巣がんと他の 9 種類のがんの GSE と TCGA からのメチル化データと共に「検証コーホート」を使用して、卵巣がんと他のがんを区別するための、スベック DNA メチル化スコアの有用性を実証した (図 2 4 A) 。図 2 4 B に示される ROC 曲線は、卵巣がんを他の正常な組織および他のがんから区別するためのこのメチル化スコアの特異性 (0 . 9 9) および感度 (1) を明らかにしている (図 2 4 C) 。これらの DNA メチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

【 0 1 6 7 】

実施形態 2 1 : 子宮頸がんのポリジーン DNA メチル化マーカーの発見。

本発明者らは、GSE 4 6 3 0 6 (正常) からの 2 0 人の、そして子宮頸がん DNA メチル化データの TCGA コレクションからランダムに選択された 1 0 個のサンプルからの正規化された Illumina 4 5 0 K DNA メチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態 1 で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット 2 8 7 5 4 CGI D を候補リストに挙げた。次に、発明者らは、実施形態 2 に記載の BCD 法を使用して、トレーニングコーホートにおいて高い感度および特異性で子宮頸がんを検出する (図 2 5 B 、表 1 6) (検出) 、バイナリーカテゴリー区別的にメチル化された CGI D のポリジーンセットを発見した。実施形態 2 で説明したように、加重 DNA メチル化スコアおよびがんの閾値が CGI D のために開発された。次に、発明者らは、8 種の異なる腫瘍タイプおよび血液を表す TCGA からランダムに選択された 8 0 個の DNA メチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、子宮頸がんと他の腫瘍との間で区別的にメチル化された CGI D のポリジーンセットを発見した (図 2 5 C 、表 1 7) (スベック) 。実施形態 2 で説明したように、加重 DNA メチル化スコアが CGI D のために開発された。

【 0 1 6 8 】

実施形態 2 2 : 子宮頸がんを検出するための子宮頸がんポリジーン DNA メチル化マーカーの有用性。

次に、発明者らは、実施形態 2 1 (検出) で開発された加重子宮頸がん DNA メチル化スコアおよび閾値が、TCGA からの 3 1 3 人の子宮頸がん患者からの正規化された Illumina 4 5 0 K DNA メチル化ベータ値が含まれる「検証コーホート」で子宮頸がんを検出することを実証した。この方法を使用して、9 1 % の子宮頸がんサンプルが子宮頸がんとして検出された (図 2 6 A) 。次に、発明者らは、子宮頸がんと他の 9 種類のがんの GSE と TCGA からのメチル化データと共に「検証コーホート」を使用して、子宮頸がんと他のがんを区別するための、スベック DNA メチル化スコアの有用性を実証した (図 2 6 A) 。図 2 6 B に示される ROC 曲線は、子宮頸がんの検出、および子宮頸がんを他の正常な組織および他のがんから区別するためのこのメチル化スコアの特異性 (0 . 9) および感度 (0 . 9) を明らかにしている (図 2 6 C) 。これらの DNA メチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人

におけるがんの早期発見に使用できる。

【0169】

実施形態23：頭頸部扁平上皮がん（HNSC）のポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE（52068）（正常）からの10人の、そしてHNSC DNAメチル化データのTCGAがんコレクションからランダムに選択された10個のサンプルから正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性でHNSCを検出する（図27B、表18）（検出）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、12種の異なる腫瘍タイプを表すTCGAからランダムに選択された80個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、HNSCと他の腫瘍との間で区別的にメチル化されたCGIDのポリジーンセットを発見した（図27C、表19）（スペック）。

10

【0170】

実施形態24：HNSCを検出するための頭頸部扁平上皮がん（HNSC）ポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態23（検出）で開発された加重HNSC DNAメチル化スコアおよび閾値が、GSE52068、GSE75537およびGSE79556からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」でHNSCを検出することを実証した。この方法を使用して、88%~96%のHNSCサンプルが検出された（図28A）。次に、発明者らは、HNSCと他の12種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、HNSCと他のがんを区別するための、DNAメチル化検出スコアの有用性を実証した（図28A）。図28Bに示されるROC曲線は、HNSCを他の正常な組織および他のがんから区別するためのこのメチル化スコアの特異性（0.86）および感度（0.88）を明らかにしている（図28C）。マーカーは、他のいくつかのがんも検出する（比較的高感度であるため、これらのがんに対する特異性は限られている）。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

20

30

【0171】

実施形態25：食道がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE（52068）（正常）からの10人の、そして食道がんDNAメチル化データのTCGAがんコレクションからランダムに選択された10個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で食道がんを検出する（図29B、表20）（検出）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、12種の異なる腫瘍タイプを表すTCGAからランダムに選択された80個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者

40

50

らは、このトレーニングコーホートを使用して、食道がんと他の腫瘍との間で区別的にメチル化されたCGIDのポリジーンセットを発見した(図29C、表21)(スペック)。

【0172】

実施形態26：食道がんを検出するための食道がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態25(検出)で開発された加重食道がんDNAメチル化スコアおよび閾値が、GSE52068、GSE75537およびGSE79556からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で食道がんを検出することを実証した。この方法を使用して、88%~96%の食道がんサンプルが検出された(図30A)。次に、発明者らは、食道がんと他の12種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、食道がんと他のがんを区別するための、検出DNAメチル化スコアの有用性を実証した(図30A)。図30Bに示されるROC曲線は、食道がんを他の正常な組織および他のがんから区別するためのこのメチル化スコアの特異性(0.86)および感度(0.88)を明らかにしている(図30C)。マーカーは、他のいくつかのがんも検出する(比較的高感度であるため、これらのがんに対する特異性は限られている)。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

【0173】

実施形態27：膀胱がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE52955(正常)からの5人の、そして膀胱がんDNAメチル化データのTCGAがんコレクションからランダムに選択された10個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCD法を使用して、トレーニングコーホートにおいて高い感度および特異性で膀胱がんを検出する(図31B、表22)(検出)、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。次に、発明者らは、13種の異なる腫瘍タイプおよび正常な血液を表すTCGAからランダムに選択された80個のDNAメチル化サンプルから「トレーニングコーホート」を生成した。本発明者らは、このトレーニングコーホートを使用して、膀胱がんと他の腫瘍との間で区別的にメチル化されたCGIDのポリジーンセットを発見した(図31C、表23)(スペック)。実施形態2で説明したように、加重DNAメチル化スコアがCGIDのために開発された。

【0174】

実施形態28：膀胱がんを検出するための膀胱がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態27(検出)で開発された加重膀胱がんDNAメチル化スコアおよび閾値が、TCGAからの439人の膀胱がん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で膀胱がんを検出することを実証した。この方法を使用して、96%の膀胱がんサンプルが膀胱がんとして検出された(図32B)。次に、発明者らは、膀胱がんと他の13種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、膀胱がんと他のがんを区別するための、スペックDNAメチル化スコアの有用性を実証した(図32B)。図32Cに示されるROC曲線は、膀胱がんを他の正常な組織および他のがんから区別するためのこのメチル化スコアの特異性(0.86)および感度(0.88)を明ら

かにしている（図32C）。しかし、かなり高い割合で胃がん、膵臓がん、食道がん、および大腸がんの交差検出がある。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

【0175】

実施形態29：腎臓がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE52955（正常）からの10人の、そしてTCGAデータセットにある13種のがんよりがんごとにランダムに選択された10個のサンプルからの、腎臓がんのための正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用し、そして正常な組織および血液（GSE40279、GSE52955）を使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCDhyo法を使用して、トレーニングコーホートにおいて高い感度および特異性で腎臓がんを検出し、そして他のがんに対して腎臓がんの特異的である「検出 - スペック」（図33B、表24）（検出 - スペック）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。

10

20

【0176】

実施形態30：腎臓がんを検出するための腎臓がんポリジーンDNAメチル化マーカーの有用性。

発明者らは、実施形態29（検出 - スペック）で開発された加重腎臓がんDNAメチル化スコアおよび閾値が、TCGAからの871人の腎臓がん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で腎臓がんを検出し、腎臓がんを他のがんから区別することを実証した。この方法を使用して、90%の腎臓がんのサンプルが腎臓がんとして検出された（図34A）。次に、発明者らは、腎臓がんと他の13種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、腎臓がんと他のがんを区別するための、「検出 - スペック」DNAメチル化スコアの有用性を実証した（図34A）。図34Bに示されるROC曲線は、腎臓がんを他の正常な組織および他のがんから検出するためのこのメチル化スコアの特異性（0.87）および感度（0.91）を明らかにしている（図34C）（HCC、脳がん、精巣がんとの高い交差）。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人における腎臓がんの早期発見に使用できる。

30

【0177】

実施形態31：精巣がんのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GSE46306（正常）からの13人の、そしてTCGAデータセットにある13種のがんよりがんごとにランダムに選択された10個のサンプルからの、正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用し、そして正常な組織および血液（GSE40279、GSE61496）を使用した。本発明者らは最初に、実施形態1で正常な組織および血液サンプル全体で強力にメチル化されていないサイトとして発見された、「トレーニングコーホート」データセット28754 CGIDを候補リストに挙げた。次に、発明者らは、実施形態2に記載のBCDhyo法を使用して、トレーニングコーホートにおいて高い感度および特異性で精巣がんを検出し、そして他のがんに対して精巣がんの特異的である「検出 - スペック」（図35B、表25）（検出 - スペック）、バイナリーカテゴリー区別的にメチル化されたCGIDのポリジーンセットを発見した。実施形態2で説明したように、加重D

40

50

NAメチル化スコアおよびがんの閾値がCGIDのために開発された。

【0178】

実施形態32：精巣がんを検出するための精巣がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態31（検出 - スペック）で開発された加重精巣がんDNAメチル化スコアおよび閾値が、TCGAからの156人の精巣がん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で精巣がんを検出し、精巣がんを他のがんから区別することを実証した。この方法を使用して、96%の精巣がんサンプルが精巣がんとして検出された（図36A）。次に、発明者らは、精巣がんと他の13種類のがんのGSEとTCGAからのメチル化データと共に「検証コーホート」を使用して、精巣がんと他のがんを区別するための、「検出 - スペック」DNAメチル化スコアの有用性を実証した（図36A）。図36Bに示されるROC曲線は、精巣がんを他の正常な組織および他のがんから検出するためのこのメチル化スコアの特異性（0.97）および感度（0.96）を明らかにしている（図36C）これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

10

【0179】

実施形態33：13種の一般的な固形腫瘍のポリジーン汎がんDNAメチル化マーカーの発見。

本発明者らは、TCGAデータセットにある13種類のがん（膀胱がん、脳がん、乳がん、子宮頸がん、大腸がん、食道がん、HNSC、肝臓がん、肺がん、卵巣がん、膵臓がん、前立腺がん、胃がん）よりがんごとにランダムに選択された10個のサンプルからの、正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用し、そしてTCGAとGEOからの正常な組織および血液を使用した。次に、発明者らは、表x-yにリストされている10種類のがんを検出するCGIDの組み合わせリストおよび10種の一般的ながんのいずれかを高い感度および特異性で検出する候補リストに挙げられたCGIDに対して罰則付き回帰を実行した（図37B、表26）（検出）。実施形態2で説明したように、加重DNAメチル化スコアおよびがんの閾値がCGIDのために開発された。

20

30

【0180】

実施形態34：がんを検出するための汎がんポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態33（「検出」）で開発された加重がんDNAメチル化スコアおよび閾値が、他の正常な組織によるTCGAからの3644人のがん患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で13種の一般的なのがん（膀胱がん、脳がん、乳がん、子宮頸がん、大腸がん、食道がん、HNSC、肝臓がん、肺がん、卵巣がん、膵臓がん、前立腺がん、胃がん）を検出することを実証した。この方法を使用して、90%~95%のがんサンプルが検出された（図38A）。図38Bに示されるROC曲線は、13種類のがんを他の正常な組織から検出するためのこのメチル化スコアの特異性（0.99）および感度（0.95）を明らかにしている（図38C）。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人におけるがんの早期発見に使用できる。

40

【0181】

実施形態35：黒色腫を検出するためのポリジーンDNAメチル化マーカーの発見。

本発明者らは、TCGAおよびGEOデータセット内のランダムに選択された10個の黒色腫サンプルと他のがん（膀胱がん、脳がん、乳がん、子宮頸がん、大腸がん、食道がん、HNSC、肝臓がん、肺がん、卵巣がん、膵臓がん、前立腺がん、胃がん）および正

50

常血液からの220個のサンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。次に、発明者らは、黒色腫の検出のためのCGIDの組み合わせリスト、および黒色腫を高い感度および特異性で検出する候補リストに挙げられたCGIDに対して罰則付き回帰を実行した(図39、表28)(検出-スペック)。実施形態2で説明したように、加重DNAメチル化スコアおよび黒色腫の閾値がCGIDのために開発された。

【0182】

実施形態36：黒色腫を検出するための黒色腫ポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態35(「検出-スペック」)で開発された加重黒色腫DNAメチル化スコアおよび閾値が、他のがんおよび正常な組織によるTCGAからの475人の黒色腫患者からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」で黒色腫を検出することを実証した。この方法を使用して、98%の黒色腫サンプルが検出された(図40A)。図40Bに示されるROC曲線は、黒色腫を他の正常な組織および他のがんから検出するためのこのメチル化スコアの特異性(0.98)および感度(0.95)を明らかにしている(図40C)。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、患者からの組織、糞便、唾液、血漿および尿からの異なる生体材料を使用して、リスクのある人や一般的な健康な人における黒色腫の早期発見に使用できる。

【0183】

実施形態37：急性骨髄性白血病(AML)を検出するためのポリジーンDNAメチル化マーカーの発見。

本発明者らは、GEOデータセット内のランダムに選択された10個のAMLサンプルと10個の正常な血液サンプルからの正規化されたIllumina 450K DNAメチル化データを「トレーニング」コーホートとして使用した。次に、発明者らは、AMLの検出のためのCGIDの組み合わせリスト、およびAMLを高い感度および特異性で検出する候補リストに挙げられたCGIDに対して罰則付き回帰を実行した(図41、表27)(検出-スペック)。実施形態2で説明したように、加重DNAメチル化スコアおよびAMLの閾値がCGIDのために開発された。

【0184】

実施形態38：血液DNA中のAMLを検出するための急性骨髄性白血病(AML)ポリジーンDNAメチル化マーカーの有用性。

次に、発明者らは、実施形態37(「検出-スペック」)で開発された加重AML DNAメチル化スコアおよび閾値が、GEOからの79人のAML患者およびTCGAからの140人の患者、そして正常な血液からの正規化されたIllumina 450K DNAメチル化ベータ値が含まれる「検証コーホート」でAMLを検出することを実証した。この方法を使用して、100%のAMLサンプルが検出された(図42A)。図42Bに示されるROC曲線は、血液からのAMLを検出するためのこのメチル化スコアの特異性(1)および感度(1)を明らかにしている(図42C)。これらのDNAメチル化マーカーとメチル化値からの計算されたメチル化スコアは、血液DNAを使用して、リスクのある人や一般的な健康な人におけるAMLの早期発見に使用できる。

【0185】

実施形態39：前立腺がんを予測するためのバイサルファイト変換、多重増幅および次世代シーケンスおよびメチル化スコアの計算。

血液は、K3-EDTAを含む9mlチューブに収集され、1時間以内に処理された。新鮮な血液サンプルを4で1000gで10分間遠心分離した。細胞層を乱すことなく上澄みをファルコンチューブに注意深く移し、残りの細胞を完全に除去するために再度10分間遠心分離し、-80で凍結した。血漿サンプルを解凍し、血漿DNA用のQiaagenキットやEZ DNA直接抽出法など、血漿DNA抽出用のいくつかの利用可能な方法と市販のキットでDNAを抽出する。AMPure XP磁気ビーズなどの市販の方

10

20

30

40

50

法を使用してDNAを精製し、精製したDNAを、例えばEZ DNAバイサルファイト処理キットを使用して亜硫酸水素ナトリウムで処理する。ターゲット配列のライブラリーは、2段階のPCR反応によって生成される(図40)。第一のPCR反応は、表5および表6の特定のCGIDをターゲットにする。PCR1プライマーは、第二のPCR2プライマーと相補的な配列を持っていることに注意する(図45)。本発明者らは、HEK293細胞からのヒトバイサルファイト変換ゲノムDNAを使用して、標準的なTaqポリメラーゼ反応で以下のプライマーを使用したマルチプレックスPCR反応において、HIF3A(232塩基対領域)、TPM4(213塩基対領域)、およびCTTN(199塩基対領域)から前立腺がんを検出するCGIDを含むDNAの3つの配列を同時に増幅した: CGID cg02879662の場合、フォワードプライマー:

10

5' ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGGTAGGAGTTTTGGG AATTGG3' およびリバースプライマー:

5' gTgACTggAgTTCAgACgTgTgCTCTTCCgATCTCCACCCCTACAATCCCTAA3' CGID cg16232979の場合、フォワードプライマー:

5' ACACTCTTTCCCTACACgACgCTCTTCCgATCT NNNNNYGGTTTTYGGGTTTTYGT ATT3'

およびリバースプライマー:

5' gTgACTggAgTTCAgACgTgTgCTCTTCCgATCTACRCAAAAATATAAATCRACR ATC3'

CGID : cg14041701およびcg14498227の場合、フォワードプライマー:

20

5' ACACTCTTTCCCTACACgACgCTCTTCCgATCTNNNNNGTTTTGYGTTTTYGGA TTTGGGTT3'

およびリバースプライマー:

5' gTgACTggAgTTCAgToACgTgTgCTCTTCCgATCTCATAAACAAACACCTTTAAATAAACACTAAA3'。増幅した断片をアガロースゲルで分画した。

【0186】

サンプルをバーコード化するために、次のプライマーを使用した第二のPCR反応を使用する:

フォワードプライマー:

30

5' AATgATACggCgACCACCgAgATCTACACTCTTTCCCTACACgAC3'

バーコードプライマー(リバース):

5' CAAGCAgAAgACggCATACgAgATAGTCATCGgTgACTggAgTTCAgACgTg3'

(太字の塩基はインデックスである。このインデックスの200のバリエーションが使用される。プライマーの第二のセットは、各患者のインデックスと、リバースおよびフォワードシーケンスプライマーを導入する。前立腺がんHIF3A 232 bp、TPM4 213 bp、およびCTTN 199 bpの3つのマーカーに対するマルチプレックスPCR1反応は、図46に示すように、さまざまなプライマー濃度を使用する右側のパネルに表示される。

【0187】

40

実施形態40: バイサルファイト変換、マルチプレックス増幅および次世代シーケンスの方法の有用性、およびがんを予測するためのメチル化スコアの計算。

本発明者らは、実施形態39が、何百人もの患者からの血漿サンプルを使用して、前立腺がんおよび他のがんを同時にハイスループット(high throughput)予測することに使用され得ることを実証する。高度に予測可能なCGIDのインデックス付き増幅と、がんを示すメチル化スコアを計算するための合理化された方法は、前立腺がんと他のがんの早期発見に使用できる。

【0188】

実施形態41: 選択されたバイオマーカーが真のBCD特性を示すことの実証は、健康な人々の血漿では完全に低メチル化されている。

50

血漿DNAは、40人の健康な個人から調製された血漿から抽出され、次のがんのがん特異的プライマーによる標的増幅に供された：肝臓がん、前立腺がん、肺がん（図43）および胃がん、汎がんおよびCRC（図44）、その後、実施形態39および実施形態40に記載されているように、第2の増幅セット（PCR2）および次世代シーケンスを使用したバーコード化を行った。すべてのCGは、健康な人々からの血漿において非常に低レベルのメチル化を示した（図43および図44）。

【0189】

実施形態42：DNAメチル化レベルを決定するためのバイオインフォマティクスワークフロー。

PCR2の産物は、定量化と精製を組み合わせ、Miseq Illuminaシーケンサーで次世代シーケンスにかけられる。シーケンスは、インデックス付きシーケンス用のIlluminaのソフトウェアを使用して逆多重化され（demultiplexed）、各患者に対してFASTQファイルが生成される。Perlテキスト編集スクリプト<https://www.activestate.com/activeperl/downloads>を使用して、患者ごとのCG IDごとのFASTQファイル内のTとCをカウントし、C/C+Tの数を除算して患者の一つのCG ID内のメチル化したCの割合を定量化する。（図42のスキームを参照）。出力CSVファイルは、方程式を使用して各患者のメチル化スコア（Ms）を計算するために使用される。方程式：

10

$$Ms = \alpha + \sum_{i=1}^n \beta_i CG_i$$

20

、 α = 切片、 β_i = CG ID_iの係数、CG = 1からnまでのCG IDの組み合わせでのCGあたりのメチル化レベル。n = 組み合わせたCGの数。Ms = メチル化スコア。

【0190】

本発明の主題の用途

本発明の主題の用途は、一般に、分子診断およびがんの早期予測の分野にある。当業者は、本発明の主題を使用して、神経疾患、糖尿病、肝硬変および心血管疾患における心臓組織の損傷などの心臓病など、細胞死および無細胞DNAのシステムへの脱落を伴う他のがんおよび他の疾患の早期予測のための同様の非侵襲性バイオマーカーを導き出すことができる。本発明の主題は、BCDおよびBCDhyp法を使用して、特定の細胞のタイプおよび組織の精巧なメチル化マーカーを見つける方法を提供する。がんを早期に検出し、生存率を劇的に高め、がんから治療するために当業者が使用することができる広範囲のがんの早期予測のための方法およびバイオマーカーも開示される。本発明により開示された方法は、健康な人々に定期的な毎年のスクリーニングを行い、がんを発症し始めている人々を特定し、直ちに治療し、がんの死亡率と罹患率の悲惨な個人的社会的および経済的影響を防止するため、また、「リスクの高い」人々を監視し、再発または転移を検出するために治療を受けている患者の治療に対する反応を監視するために、当業者により使用され得る。ここに記載された本発明の、医療提供者および健康診断施設による日常の医療管理のための採用は、がんの負担および医療費の削減に大きな影響を与えるであろう。

30

40

【0191】

本発明の主題が多数の異なる従属請求項を含むという事実は、がんを予測するためにこれらの請求項の組み合わせを使用できないことを意味するものではない。がんを測定し、統計的に分析および予測するために本明細書に開示されている実施形態は、限定的であると考えられるべきではない。Illumina EPICアレイ、キャプチャーアレイシーケンス、次世代シーケンス、メチル化特異的PCR、エピタイパー、制限酵素ベースの分析、そしてパブリックドメインにあるその他の方法など、がん患者のDNAメチル化を測定するためのさまざまな他の変更が当業者には明らかである。同様に、患者サンプル中のがんの予測のために本発明の主題を使用するために、ここに挙げられているものに加えて、パブリックドメインには多数の統計方法がある。

50

【 0 1 9 2 】

本発明の主題は、1つ以上の好ましい実施形態を含むその実施形態に関連して説明されてきたが、請求される主題の精神および範囲から逸脱することなく、他の多くの可能な修正および変更を行うことができることを理解されたい。

【 図 面 】

【 図 1 】

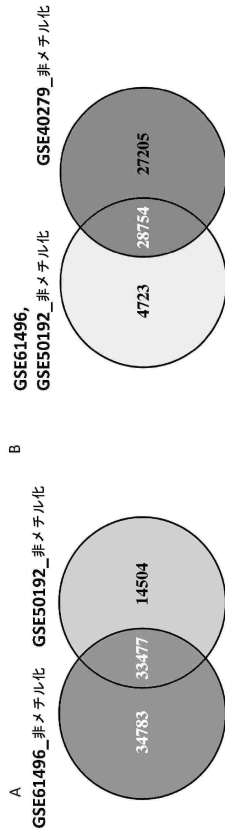


Figure 1.

【 図 2 】

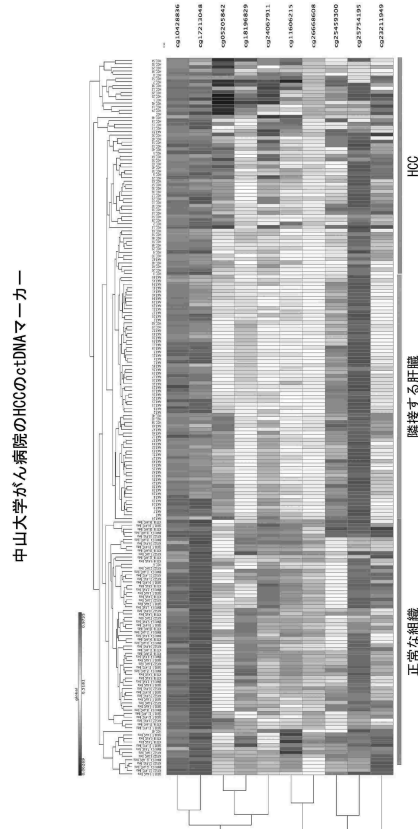


Figure 2.

10

20

30

【 図 2 A 】

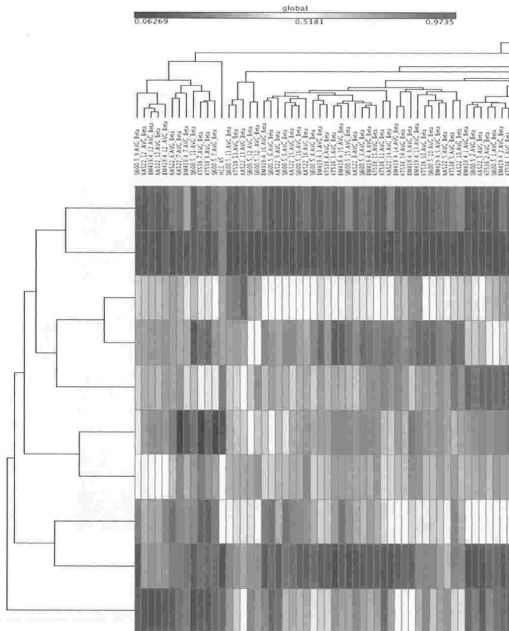


FIG. 2A

【 図 2 B 】

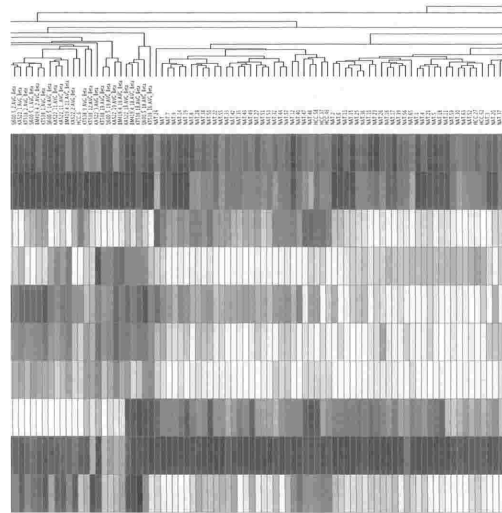


FIG. 2B

40

50

【 図 3 C 】

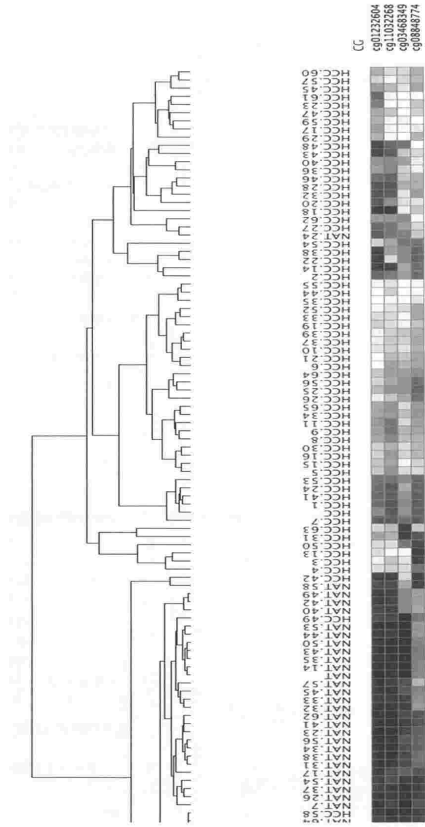


FIG. 3C

【 図 4 】

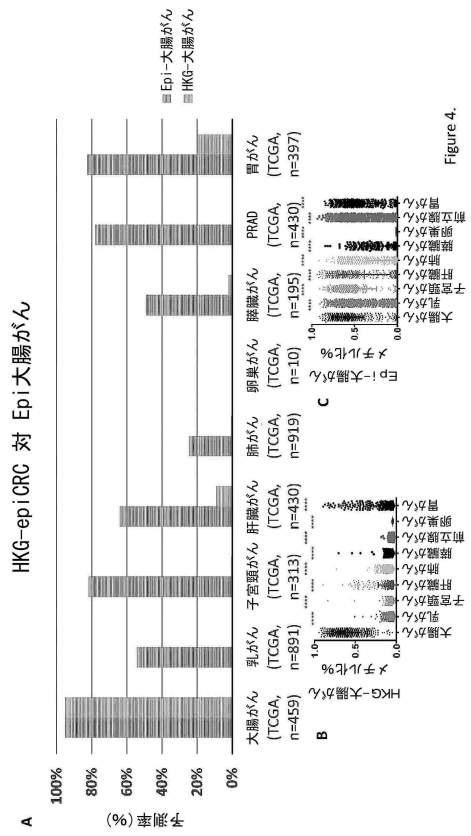


Figure 4.

10

20

【 図 4 A 】

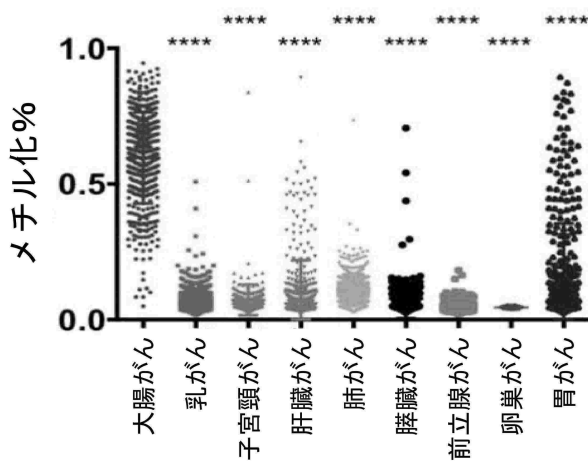


FIG. 4A

【 図 4 B 】

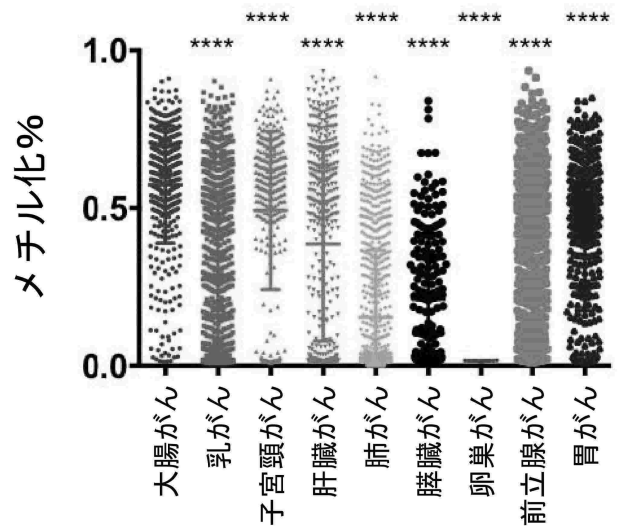


FIG. 4B

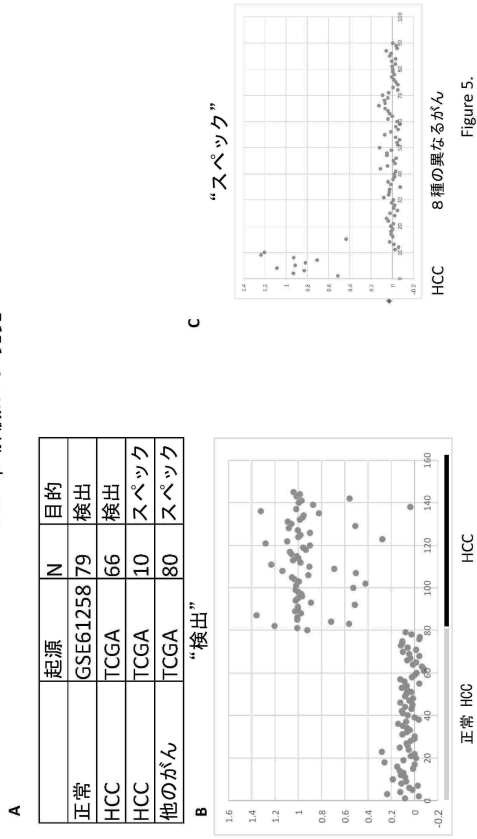
30

40

50

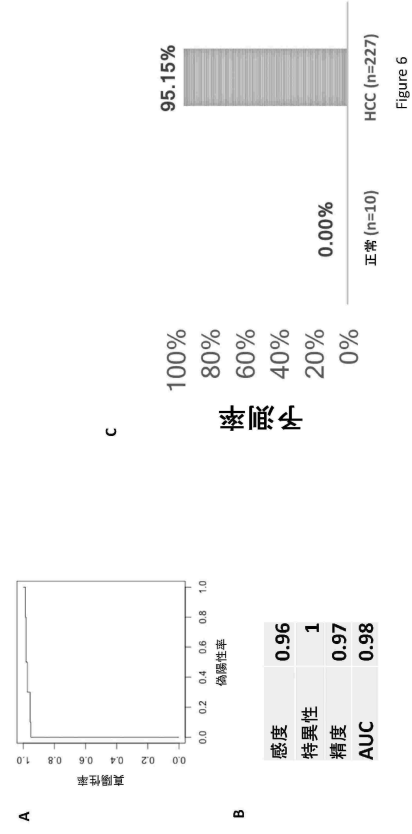
【 図 5 】

HKG-epi 肝臓がん：発見



【 図 6 】

HKG-epi 肝臓がん-検出・検証GSE76269

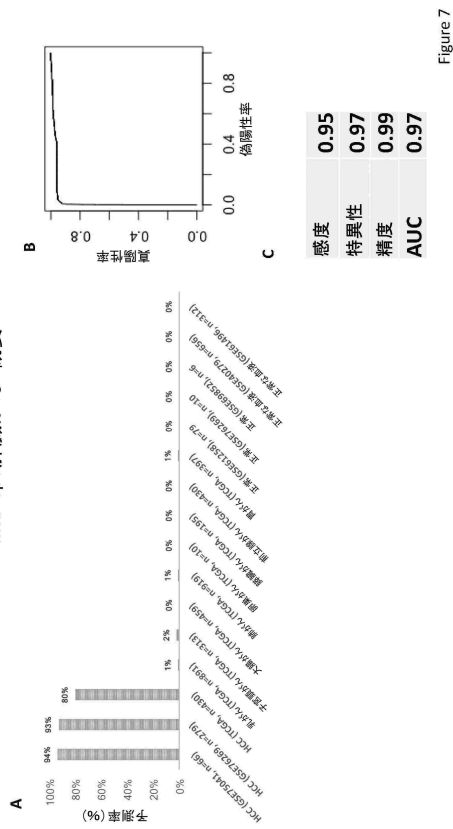


10

20

【 図 7 】

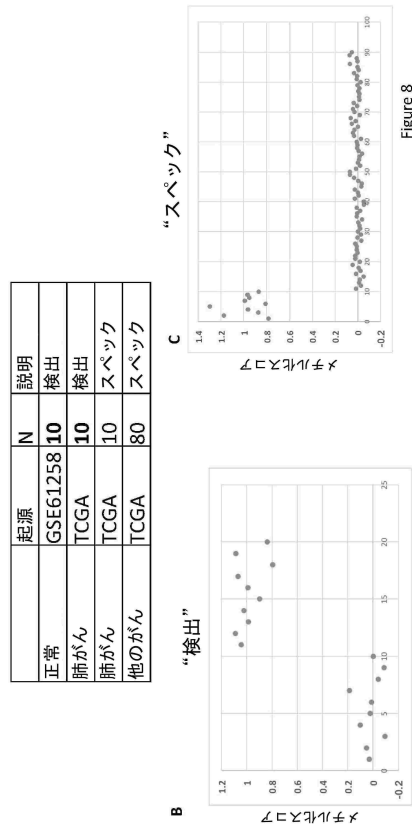
HKG-epi 肝臓がん：概要



30

【 図 8 】

HKG-epi 肺がん：発見



40

50

【 図 9 】

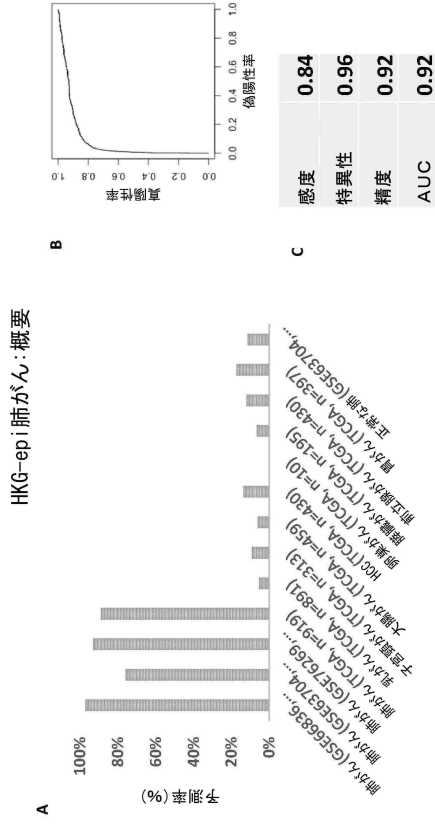


Figure 9

(69)

【 図 10 】

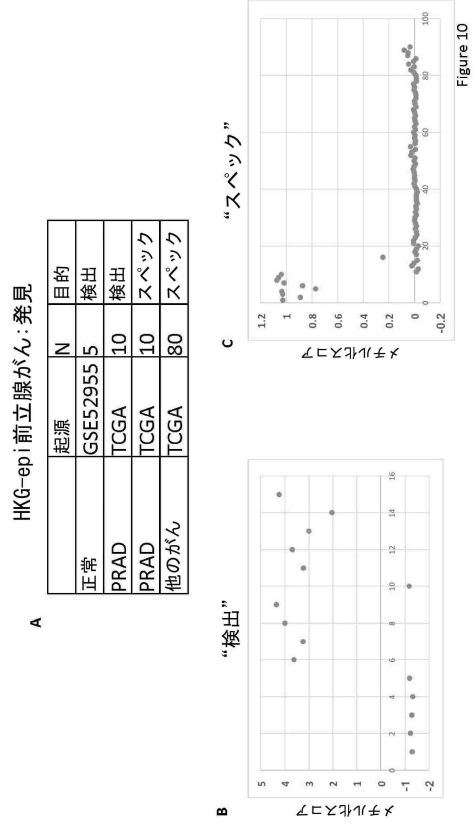


Figure 10

10

20

【 図 11 】

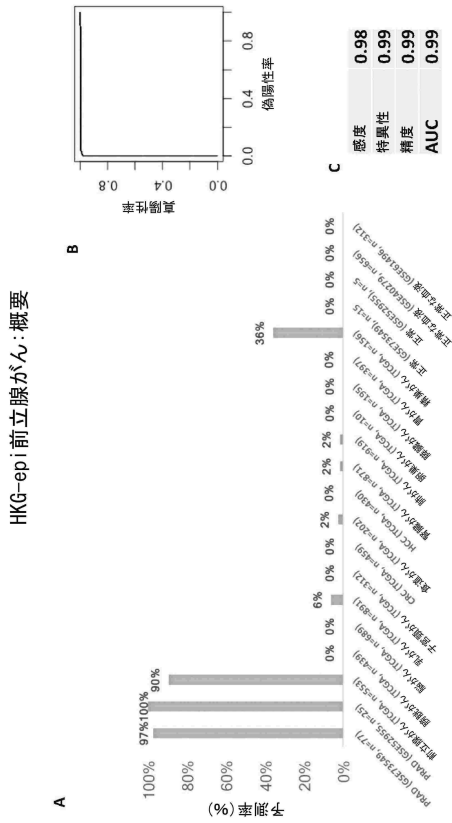


Figure 11

(70)

【 図 12 】

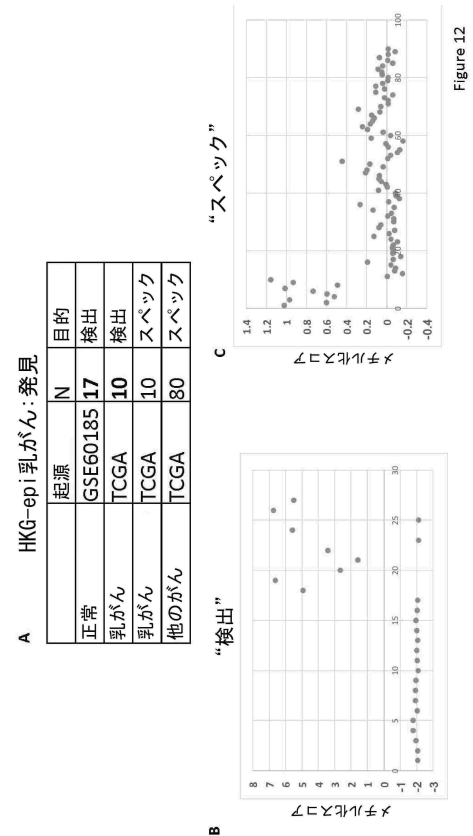


Figure 12

30

40

50

【 図 1 3 】

HKG-epi乳がん-検出: 検証GSE60185

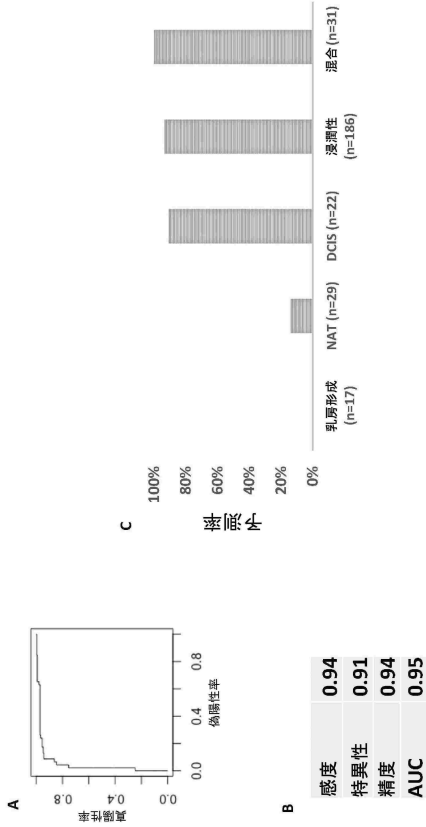


Figure 13

【 図 1 4 】

HKG-epi乳がん: 概要

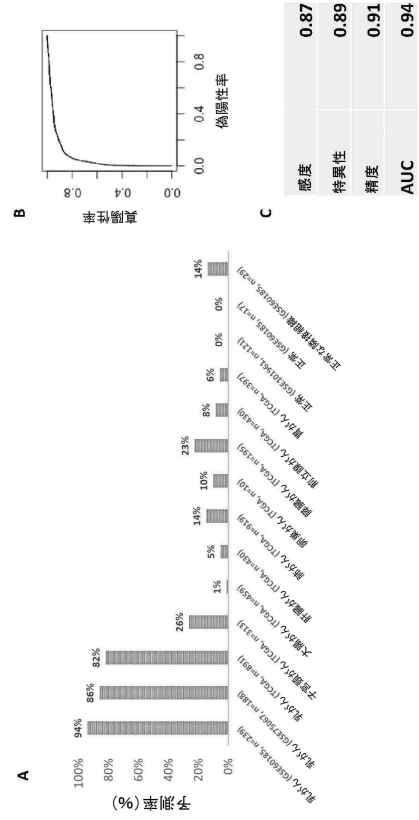


Figure 14

10

20

【 図 1 5 】

HKG-epiCRC: 発見

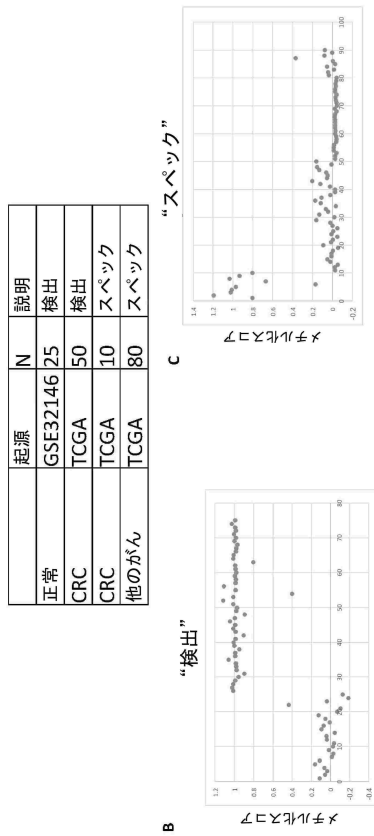


Figure 15

【 図 1 6 】

HKG-epiCRC: 概要

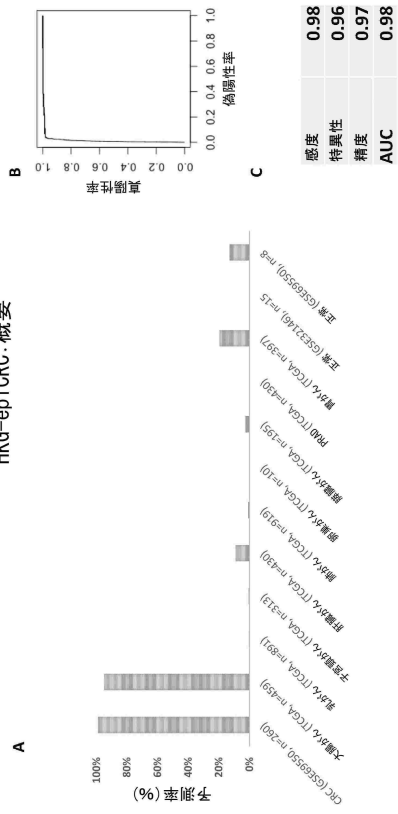


Figure 16

30

40

50

【 図 1 7 】

HKG-ep i 膵臓がん: 発見

A	起源	N	説明
健康	GSE53051	12	検出
PANC	TCGA	20	検出
PANC	TCGA	20	スペースック
他のがん	TCGA	100	スペースック

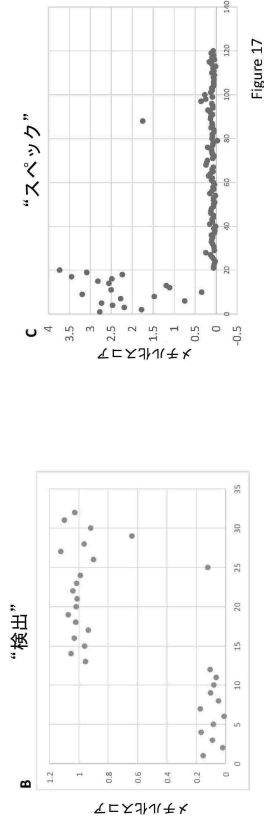


Figure 17

【 図 1 8 】

HKG-ep 膵臓がん: 概要

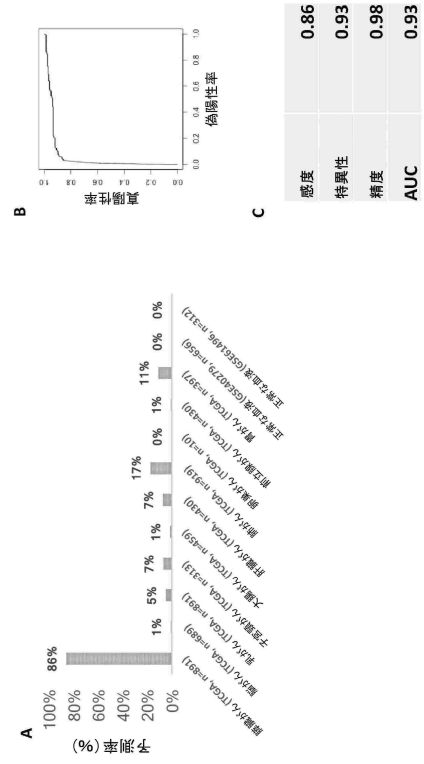


Figure 18

10

20

【 図 1 9 】

HKG-ep i 脳がん (膠芽腫): 発見

A	起源	N	説明
健康	GSE6820	6	検出
脳がん	TCGA	10	検出
脳がん	TCGA	10	スペースック
他のがん	TCGA	110	スペースック

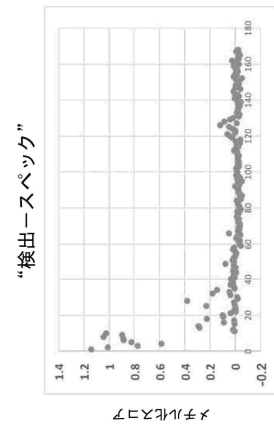


Figure 19

【 図 2 0 】

HKG-ep i 脳がん (膠芽腫): 概要



Figure 20

30

40

50

【 図 2 1 】

HKG-epi 胃がん-検出: 発見

起原	N	目的
胃がん	7	スペースック
他のがん	100	スペースック
正常	GSE99553 14	検出
胃がん	TCGA 20	検出

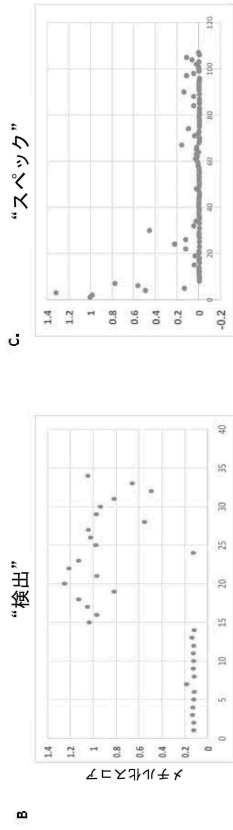


Figure 21

【 図 2 2 】

HKG-epi 胃がん: 概要

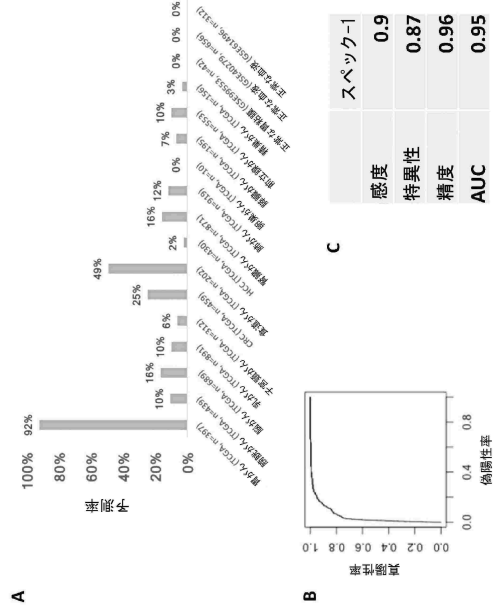


Figure 22

10

20

【 図 2 3 】

HKG-epi 卵巣がん: 発見

起原	N	説明
健康	GSE65820 5	検出
卵巣がん	TCGA 10	検出
卵巣がん	TCGA 10	スペースック
他のがん	TCGA 110	スペースック



Figure 23

【 図 2 4 】

HKG-epi 卵巣がん: 概要

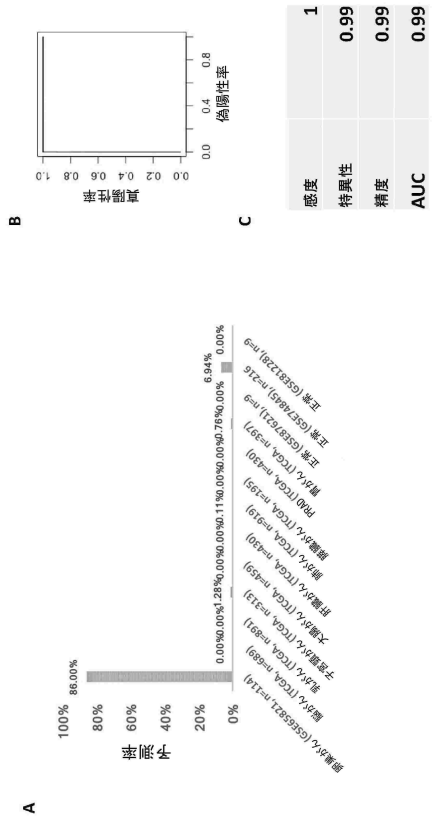


Figure 24

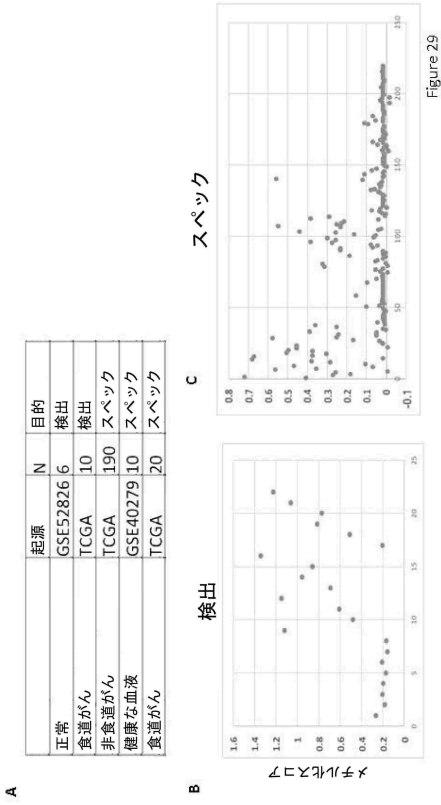
30

40

50

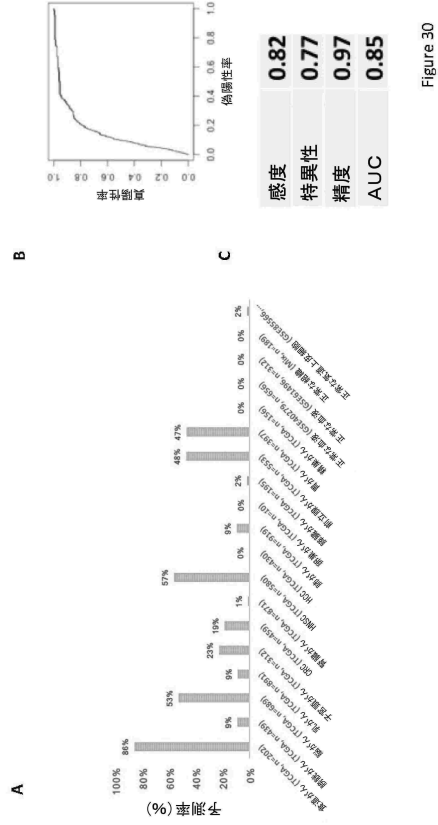
【 図 2 9 】

HKG-epi 食道がん：発見



【 図 3 0 】

HKG-epi 食道がん：概要

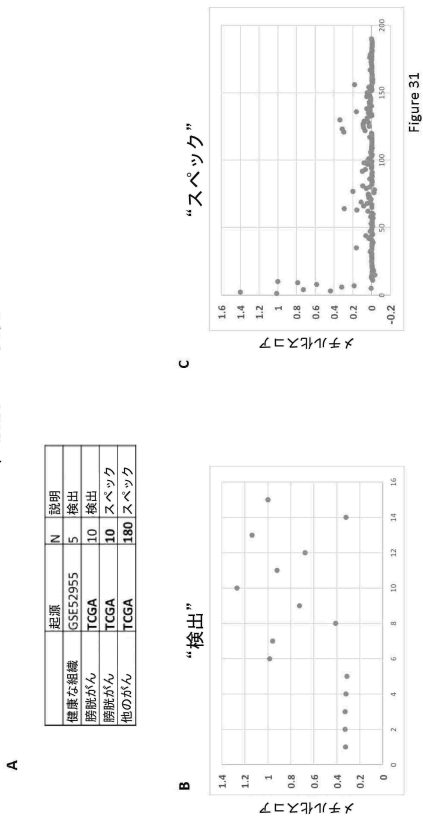


10

20

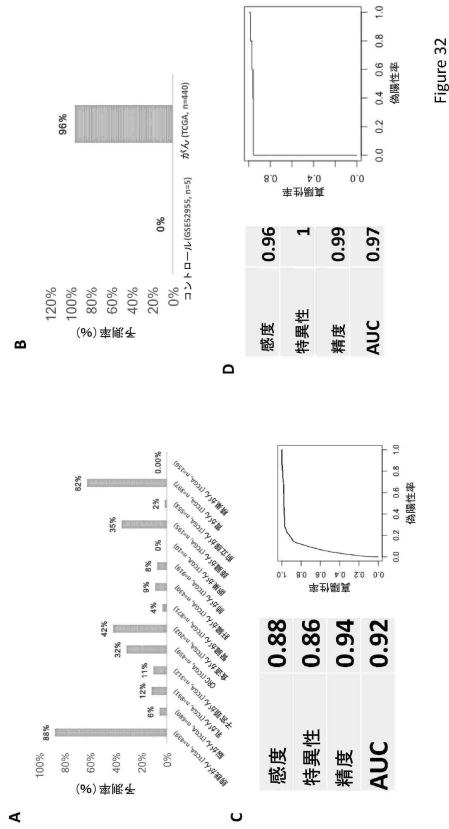
【 図 3 1 】

HKG-epi 膀胱がん：発見



【 図 3 2 】

HKG-epi 膀胱がん：概要



30

40

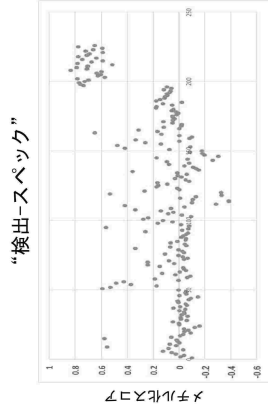
50

【 図 3 3 】

HKG-epi 腎臓がん: 発見

	起源	N	目的
他のがん	TCGA	180	検出-スベック
健康な血液	GSE40279	10	検出-スベック
正常な腎臓	GSE52955	6	検出-スベック
腎臓がん	TCGA	30	検出-スベック

A



B

Figure 33

【 図 3 4 】

HKG-epi 腎臓がん: 概要

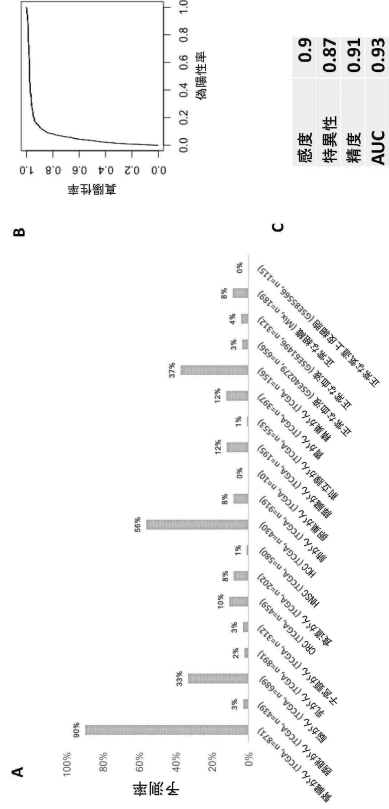


Figure 34

10

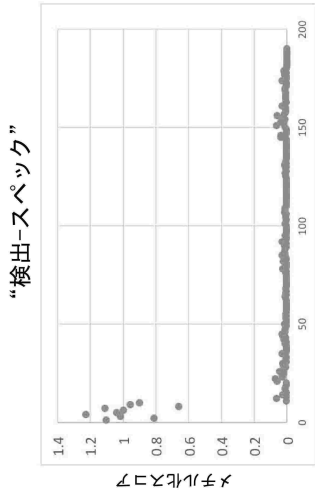
20

【 図 3 5 】

HKG-epi 精巣がん: 発見

	起源	N	目的
他のがんおよび血液	GSE46306	180	検出-スベック
精巣がん	TCGA	10	検出-スベック

A



B

Figure 35

【 図 3 6 】

HKG-epi 精巣がん: 概要

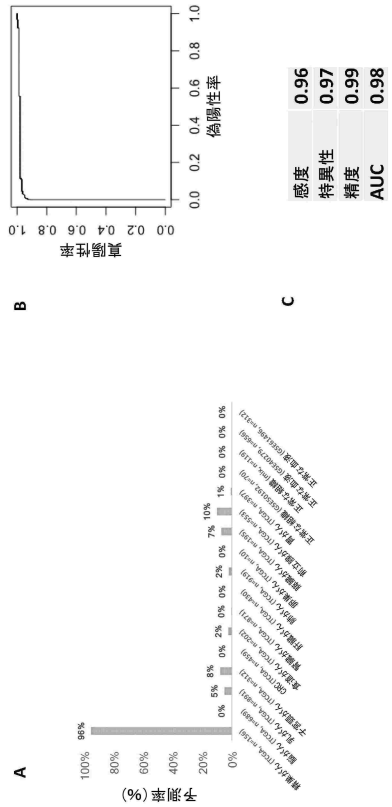


Figure 36

30

40

50

【 図 3 7 】

HKG-epi 汎がん 発見

A

健康な血液	起源	N	目的
TCGA_がん	GSE40279	10	検出
	TCGA	170	検出

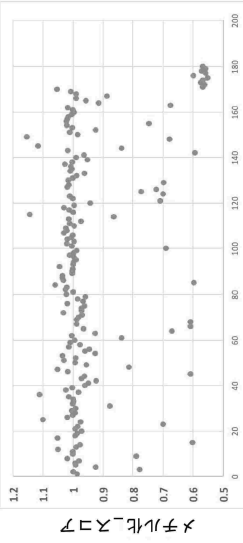


Figure 37

【 図 3 8 】

HKG-epi 汎がん 概要

A

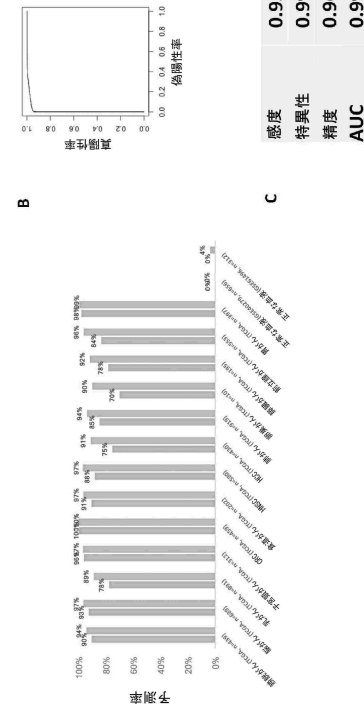


Figure 38

【 図 3 9 】

HKG-epi 黒色腫: 発見

A

黒色腫 (SKCM)	起源	N	目的
他のがんおよび血液	TCGA	10	検出-スベック
	TCGA	220	検出-スベック

B

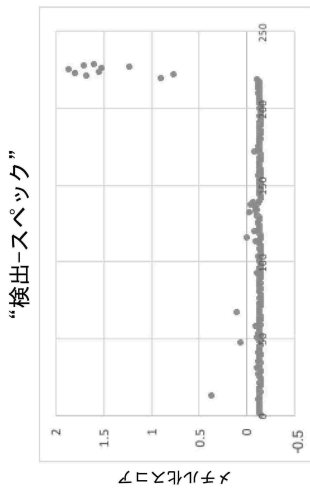


Figure 39

【 図 4 0 】

HKG-epi 黒色腫 概要

A

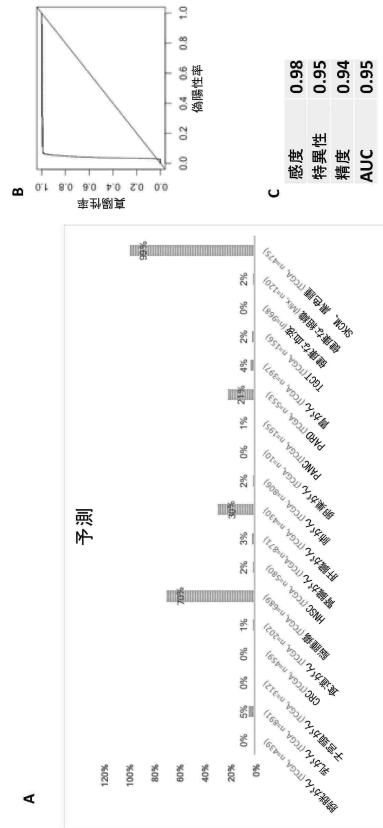


Figure 40

汎がんマーカー : 13の所在、コントロール : 968の健康な血液コントロール

【 図 4 1 】

HKG-epi AML : 発見

AML	起源	N	目的
AML	GSE86409	10	検出-スベック
血液	GSE40279	10	検出-スベック

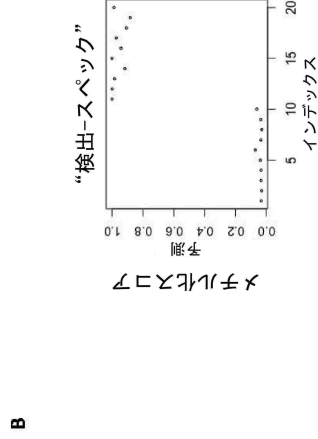


Figure 41

【 図 4 2 】

HKG-epi AML : 概要

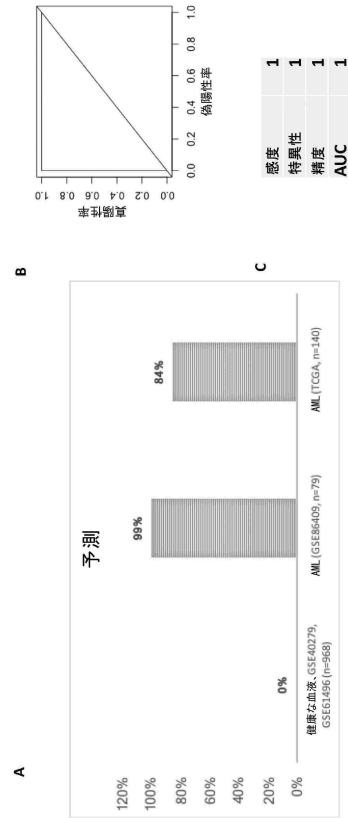


Figure 42

10

20

【 図 4 3 】

健康な血漿のBCD検証

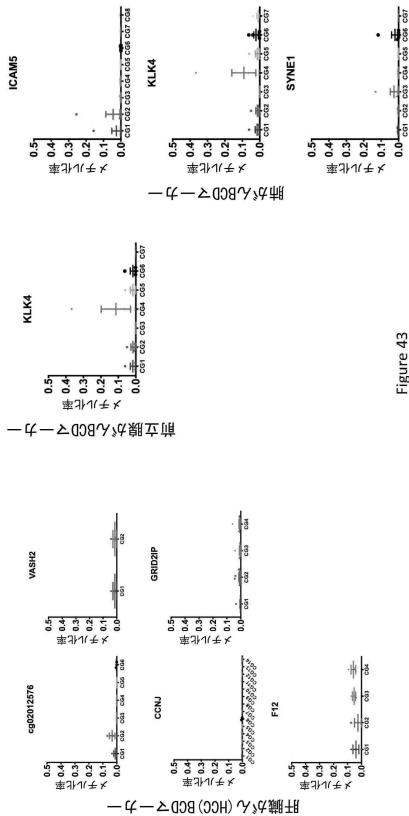


Figure 43

【 図 4 4 】

健康な血漿のBCD検証

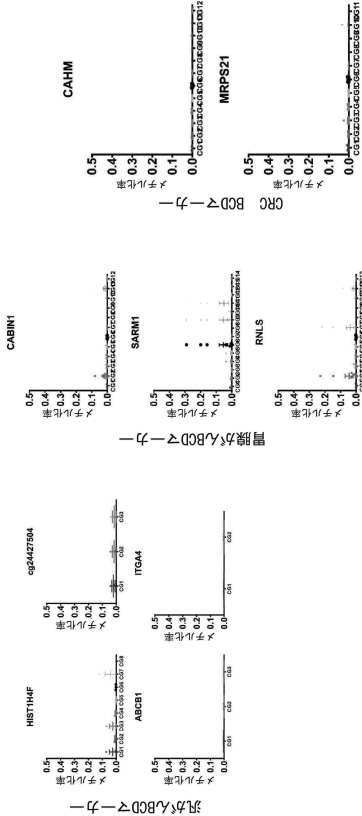


Figure 44

30

40

50

【配列表】

2024105502000001.xml

10

20

30

40

50

フロントページの続き

- エールフォン, セシル4 5 6 7
(72)発明者 リー, ファイ
中国, 香港, 香港サイエンスパーク, 11サイエンスパークウェストアベニュー, バイオテックセンター2、1/F, ユニット613
- (72)発明者 ウォン, チーフアット
中国, 香港, N.T., 青衣, チュンファットエスト, チュンファットハウス1009号室