



등록특허 10-2307424



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년09월29일
(11) 등록번호 10-2307424
(24) 등록일자 2021년09월24일

- (51) 국제특허분류(Int. Cl.)
C12Q 1/6886 (2018.01) C12Q 1/6881 (2018.01)
- (52) CPC특허분류
C12Q 1/6886 (2018.05)
C12Q 1/6881 (2018.05)
- (21) 출원번호 10-2020-7019527(분할)
- (22) 출원일자(국제) 2013년09월20일
심사청구일자 2020년07월24일
- (85) 번역문제출일자 2020년07월06일
- (65) 공개번호 10-2020-0085928
- (43) 공개일자 2020년07월15일
- (62) 원출원 특허 10-2015-7010184
원출원일자(국제) 2013년09월20일
심사청구일자 2018년09월19일
- (86) 국제출원번호 PCT/AU2013/001088
- (87) 국제공개번호 WO 2014/043763
국제공개일자 2014년03월27일
- (30) 우선권주장
61/703,512 2012년09월20일 미국(US)
(뒷면에 계속)
- (56) 선행기술조사문헌
WO2011141711 A1
WO2007132166 A2
- (73) 특허권자
더 차이니즈 유니버시티 오브 홍콩
중국 홍콩 엔. 티. 새턴 더 차이니즈 유니버시티 오브 홍콩
- (72) 발명자
치우 로사 와이 켄
중국 홍콩 에스에이알 뉴 테리토리즈 새턴 마 록 패스 52 더블 헤이븐 하우스 31
- (74) 대리인
김진희, 김태홍

전체 청구항 수 : 총 43 항

심사관 : 이준혁

(54) 발명의 명칭 혈장으로부터 태아 또는 중앙 메틸롬의 비침습적 결정

(57) 요약

본 발명은, 다양한 조직 및 표본의 메틸화 프로필을 결정하고 이용할 수 있는 시스템, 방법, 및 장치에 관한 것이다. 본 발명에는 그 예가 제공되어 있다. 혈장 메틸화(또는 무세포 DNA를 갖는 다른 표본)와 모체/환자의 메틸화 프로필의 비교에 기반하여 태아/중앙 조직에 대해 임의의 메틸화 프로필이 추정될 수 있다. 표본이 DNA 혼합

(뒷면에 계속)

대표도

100

	모체 혈장 (표1 식별)		CSF		만기 태반		모체 혈장 (표1 식별)		모체 혈장 (표3 식별)		모체 혈장 (표3 식별)	
	원소	크리	원소	크리	원소	크리	원소	크리	원소	크리	원소	크리
미가량 측정지 카본트	54.7522112	54.7522112	42.2873853	63.7531710	63.7531710	45.67228854	45.67228854	12.88331674	12.88331674	12.77615263	12.77615263	
중 합성 가능성(%)	45.09	44.39	44.22	43.83	46.70	44.40	29.82	35.81	35.39	37.40	37.40	
중 합성 가능 측정지	24.640482	24.640482	19.8805988	10.3403028	29.7735873	23.3050215	13.61938564	45.1659366	45.0248036	44.3697863	47.770078	
모체 비율(%)	6.72	6.83	5.64	5.89	5.25	5.64	6.35	6.43	5.86	5.93	5.90	6.85
중합 비율(%)	6.60	5.88	9.52	9.27	32.94	30.39	64.66	64.41	25.70	25.18	41.01	40.55
모체 및 태아 간 변이 비율의 재가 후	11.29	11.19	8.36	8.31	9.43	9.25	25.40	25.29	19.55	19.05	14.98	14.00
전체 재가 없이(%)	99.00	99.58	99.93	99.84	99.95	99.96	99.97	99.96	99.96	99.96	99.96	99.95
개별 C 사열 N 카본트	20846956	205652085	196003145	108440370	220378306	223722965	640307054	530295175	456260420	453648655	397072165	395158659
개별 C 사열 U 카본트	448494136	444964957	423238295	420135075	4098445023	400956234	1.1099E+10	1.1641E+10	8576357110	8538105885	689711473	866535640
개별 C 카본트	68.45	66.39	62.66	62.55	67.49	67.40	80.48	81.46	69.84	69.76	67.68	67.61
개별 C 사열 길이	6.14	6.07	7.57	7.51	8.89	8.78	21.12	21.01	15.46	15.39	12.48	12.43
개별 C 메틸화 정도(%)	5.65	5.63	4.31	4.29	4.36	4.36	5.19	5.17	5.85	5.04	5.45	5.44
CpG 사열 N 카본트	205233963	202943447	182423116	181245307	2137571505	211325313	627666979	522753476	443220176	440725023	388913638	387475582
CpG 사열 U 카본트	104163057	103975011	145623017	140490340	147248261	146617792	310300075	306133584	206357482	206477045	142859625	142271154
CpG 카본트	85.17	85.12	81.02	80.92	84.23	84.15	92.02	92.02	85.13	85.06	84.93	84.85
CpG 사열 길이	13.12	13.01	11.68	11.71	12.82	12.71	33.90	33.85	23.06	22.94	18.91	18.81
CpG 메틸화 정도(%)	71.80	71.63	55.72	54.97	59.21	59.04	66.96	66.90	68.23	68.20	73.18	73.14
HiCG 사열 N 카본트	3275043	3181847	7290715	7162205	12802148	12801418	12810914	12532520	12897169	12843942	7858475	7740641
HiCG 사열 U 카본트	430228351	423959404	3579820287	3571155466	4569807406	453754784	1.1356E+10	1.1308E+10	8367572695	836055296	6749182596	6719112482
HiCG 카본트	0.13	0.13	0.12	0.12	0.13	0.13	0.16	0.16	0.14	0.14	0.13	0.13
HiCG 사열 길이	1.52	1.52	1.41	1.41	1.62	1.61	4.02	4.00	2.96	2.95	2.39	2.38
HiCG 메틸화 정도(%)	0.88	0.87	0.18	0.18	0.27	0.27	0.11	0.11	0.16	0.15	0.12	0.12
CHG 사열 N 카본트	1025734	1001267	7108977	7071015	1222342	1216751	3766662	3691180	3626463	3606177	2147382	2142523
CHG 사열 U 카본트	1345458602	1341383986	1217862955	1216820371	1361862765	1357566940	341655503	3404725988	2579517702	2569459447	2036318055	2030929603
CHG 카본트	77.90	77.62	73.26	73.13	77.50	77.38	88.16	88.12	79.68	79.59	77.84	77.75
CHG 사열 길이	10.69	10.65	9.86	9.85	11.64	11.60	27.65	27.55	23.40	23.32	16.48	16.42
CHG 메틸화 정도(%)	0.88	0.87	0.17	0.17	0.24	0.23	0.11	0.11	0.14	0.14	0.12	0.11
CHH 사열 N 카본트	2249309	2145760	5165738	5091211	8079803	8124657	9052012	8841730	9353607	9251225	5512083	5428318
CHH 사열 U 카본트	2666642429	2643700465	2781814332	2745457867	3198114541	3180185844	7939454828	7903454828	5848914821	5811006511	4712645421	4691636798
CHH 카본트	61.95	61.50	58.43	58.32	63.64	63.45	77.52	77.52	65.96	65.89	63.65	63.60
CHH 사열 길이	6.84	6.81	6.40	6.38	7.42	7.36	18.38	18.36	13.55	13.48	10.91	10.85
CHH 메틸화 정도(%)	0.88	0.87	0.19	0.19	0.29	0.28	0.11	0.11	0.16	0.16	0.12	0.12

물을 갖는 경우 태아/종양 유래 DNA를 확인하기 위해 조직-특이적 대립유전자를 이용하여 태아/종양 조직에 대해 임의의 메틸화 프로필이 결정될 수 있다. 태아/종양의 계통에서 복제수 변동을 결정하는데 임의의 메틸화 프로필이 이용될 수 있다. 태아에 대한 메틸화 마커가 다양한 기법을 통해 확인되고 있다. 그 메틸화 프로필은 DNA 단편 크기 분포의 크기 파라미터를 결정함으로써 결정될 수 있고, 여기서 메틸화 수준을 결정하는데 크기 파라미터에 대한 기준값이 이용될 수 있다. 추가로, 압 수준을 결정하는데 임의의 메틸화 수준이 이용될 수 있다.

- | | |
|--------------------------------|-------------------------------|
| (52) CPC특허분류 | (30) 우선권주장 |
| <i>C12Q 2600/154</i> (2013.01) | 13/842,209 2013년03월15일 미국(US) |
| (72) 발명자 | 61/830,571 2013년06월03일 미국(US) |

로 유크 밍 테니스

중국 홍콩 (에스에이알) 호만틴 코울룬 킹 탁 스트리트 7 포스 플로어

룬 미우 판

중국 홍콩 (에스에이알) 에스테이트콰이 충 라이 킹 앳 킹 하우스 룸 1116

지앙 페이용

중국 홍콩 에스에이알 뉴 테러토리즈 새턴 시우 렉 유엔 로드 코트 넘버 62-66 (넘버 62) 랭 램 블록 비 퍼스트 플로어 플랫 7

찬 와이 만

중국 홍콩 (에스에이알) 알드리치 가든-샤우케이완 블록 6 플랫 에이 31/에프

명세서

청구범위

청구항 1

유기체의 생물학적 표본으로부터의 무세포 DNA 분자를 분석하는 것에 의해 암의 유형을 결정하는 방법으로서, 상기 방법은

- (a) 생물학적 표본으로부터의 무세포 DNA 분자로부터 서열 리드를 얻는 단계로서, 상기 서열 리드는 각각의 무세포 DNA 분자의 하나 이상의 부위에서의 메틸화 상태를 포함하는 것인 단계;
- (b) 상기 서열 리드를 분석하여 복수의 부위 각각에서 메틸화된 무세포 DNA 분자의 수를 결정하고, 상기 수를 이용하여 단일 뉴클레오타이드 해상에서 무세포 DNA 분자 내 복수의 부위에 대한 메틸화 상태를 얻는 단계;
- (c) 상기 복수의 부위에 대한 메틸화 상태로부터 메틸화 프로필을 결정하는 단계; 및
- (d) 상기 메틸화 프로필에 적어도 부분적으로 기초하여 암의 유형을 결정하는 단계로서, 상기 암의 유형을 결정하는 단계는 상기 메틸화 프로필을 하나 이상의 기준 메틸화 프로필과 비교하는 단계를 포함하고, 상기 하나 이상의 기준 메틸화 프로필의 적어도 하나는 암을 갖는 것으로 알려진 대상자로부터 얻어진 하나 이상의 표본으로부터 얻어진 것인 단계

를 포함하는 것인 방법.

청구항 2

제1항에 있어서, 상기 생물학적 표본은 혈액, 혈장 및 혈청으로 이루어진 군으로부터 선택되는 것인 방법.

청구항 3

제1항에 있어서, 상기 생물학적 표본은 소변, 질액, 자궁 또는 질 세척액, 복수, 양수, 뇌척수액, 침, 땀, 눈물, 가래, 기관지 폐포 세척액, 체액 및 대변으로 이루어진 군으로부터 선택되는 것인 방법.

청구항 4

제1항에 있어서, 상기 암의 유형은 암이 발생한 기관에 기초하는 것인 방법.

청구항 5

제1항에 있어서, 상기 암의 유형은 암의 기원의 조직에 기초하는 것인 방법.

청구항 6

제5항에 있어서, 기원의 조직은 뇌, 뼈, 폐, 심장, 신장, 간, 유방, 결장, 전립선, 코인두, 위, 정소, 피부, 난소, 췌장, 자궁 및 림프절로 이루어진 군에서 선택되는 것인 방법.

청구항 7

제1항에 있어서, 상기 암의 유형은 폐암, 유방암, 결직장암, 전립선암, 코인두암, 위암, 정소암, 피부암, 신경계에 영향을 주는 암, 뼈암, 난소암, 간암, 혈액 악성물, 췌장암, 자궁내막암종, 및 신장암으로 이루어진 군으로부터 선택되는 것인 방법.

청구항 8

제1항에 있어서, 무세포 DNA 분자를 서열분석하여 서열 리드를 얻는 단계를 추가로 포함하는 것인 방법.

청구항 9

제8항에 있어서, 서열분석은 메틸화-인식 서열분석(methylation-aware sequencing)을 포함하는 것인 방법.

청구항 10

제9항에 있어서, 상기 메틸화-인식 서열분석은 바이셀파이트 서열분석을 포함하는 것인 방법.

청구항 11

제10항에 있어서, 상기 바이셀파이트 서열분석은 게놈에 걸친 바이셀파이트 서열분석을 포함하는 것인 방법.

청구항 12

제9항에 있어서, 상기 메틸화-인식 서열분석은 페어-엔드(paired-end) 대량 병렬 서열분석 또는 전체 게놈 서열 분석을 포함하는 것인 방법.

청구항 13

제8항에 있어서, 상기 서열분석 전에 무세포 DNA 분자를 농축하는 단계를 추가로 포함하는 것인 방법.

청구항 14

제13항에 있어서, 상기 농축 단계는 혼성화 프로브, 폴리머라아제 연쇄 반응(PCR) 증폭 또는 고상 혼성화의 사용을 포함하는 것인 방법.

청구항 15

제1항에 있어서, 상기 복수의 부위는 하나 이상의 CpG 부위를 포함하는 것인 방법.

청구항 16

제1항에 있어서, 상기 복수의 부위는 하나 이상의 CpG 섬으로 조직되는 하나 이상의 CpG 부위를 포함하는 것인 방법.

청구항 17

제16항에 있어서, 상기 복수의 부위에 대한 메틸화 프로필을 결정하는 단계는 상기 하나 이상의 CpG 섬에서의 서열 리드로부터 결정된 메틸화 상태를 사용하여 상기 하나 이상의 CpG 섬에 대한 메틸화 밀도를 계산하는 것을 포함하는 것인 방법.

청구항 18

제1항에 있어서, 상기 복수의 부위에 대한 메틸화 프로필을 결정하는 단계는 상기 복수의 부위의 각각의 부위에서의 서열 리드로부터 결정된 메틸화 상태를 사용하여 상기 복수의 부위의 각각의 부위에 대한 메틸화 밀도를 계산하는 것을 포함하는 것인 방법.

청구항 19

제1항에 있어서, 상기 분석은 게놈 내 무세포 DNA 분자의 위치를 결정하는 것을 추가로 포함하는 것인 방법.

청구항 20

제19항에 있어서, 상기 메틸화 프로필을 결정하는 단계는 하나 이상의 게놈 영역 내 위치한 무세포 DNA 분자의 서열 리드로부터 결정된 메틸화 상태를 사용하여 하나 이상의 게놈 영역 내 하나 이상의 메틸화 수준을 결정하는 것을 포함하는 것인 방법.

청구항 21

제20항에 있어서, 상기 하나 이상의 게놈 영역은 게놈 내에 특정된 복수의 게놈 bin을 포함하는 것인 방법.

청구항 22

제21항에 있어서, 상기 복수의 게놈 bin은 50 kb 내지 1 Mb의 크기를 갖는 것인 방법.

청구항 23

제21항에 있어서, 상기 복수의 계놈 bin은 상이한 크기를 갖는 것인 방법.

청구항 24

제21항에 있어서, 각각의 계놈 bin의 메틸화 수준을 계놈 bin이 저메틸화 또는 고메틸화되었음을 나타내는 컷오프 값과 비교함으로써 계놈 bin의 그룹을 선택하는 단계를 추가로 포함하는 것인 방법.

청구항 25

제24항에 있어서, 상기 선택된 계놈 bin의 그룹의 메틸화 수준에 적어도 부분적으로 기초하여 암의 유형을 결정하는 단계를 추가로 포함하는 것인 방법.

청구항 26

제25항에 있어서, 상기 하나 이상의 기준 메틸화 프로파일 중 다른 하나는 다른 유형의 암을 갖는 대상자로부터 얻어진 것인 방법.

청구항 27

제25항에 있어서, 상기 하나 이상의 기준 메틸화 프로파일 중 다른 하나는 건강한 대상자로부터 얻어진 것인 방법.

청구항 28

제1항에 있어서, 상기 하나 이상의 기준 메틸화 프로파일은 무세포 DNA 분자를 포함하는 하나 이상의 기준 표본으로부터 얻어지는 것인 방법.

청구항 29

제28항에 있어서, 상기 하나 이상의 기준 메틸화 프로파일 중 다른 하나가 건강한 대상자에 대응하는 것인 방법.

청구항 30

제28항에 있어서, 상기 하나 이상의 기준 표본은 각각의 대상자에 대응하는 복수의 기준 표본을 포함하고, 상기 메틸화 프로파일을 하나 이상의 기준 메틸화 프로파일과 비교하는 단계는

상기 복수의 기준 표본으로부터 결정된 복수의 기준 메틸화 프로파일을 클러스터링하여 복수의 기준 메틸화 프로파일의 하나 이상의 클러스터를 얻는 단계; 및

상기 메틸화 프로파일을 하나 이상의 클러스터와 비교하여 상기 메틸화 프로파일과 연관된 암의 유형을 결정하는 단계를 포함하는 것인 방법.

청구항 31

제1항에 있어서, 상기 메틸화 프로파일에 기초하여 카피 수 이상의 존재 또는 부재를 결정하는 단계를 추가로 포함하는 것인 방법.

청구항 32

제1항에 있어서, 상기 무세포 DNA 분자 내 하나 이상의 계놈 부위에서 하나 이상의 메틸화 수준을 결정하는 단계를 추가로 포함하는 것인 방법.

청구항 33

제1항에 있어서, 상기 메틸화 프로파일을 결정하는 단계는 하나 이상의 CpG 섬에 대한 하나 이상의 메틸화 수준을 결정하는 것을 포함하는 것인 방법.

청구항 34

대상자로부터의 생물학적 표본을 분석하는 방법으로서, 상기 방법은

- (a) 상기 대상자의 생물학적 표본으로부터 무세포 DNA 분자에 대한 서열 리드를 얻는 단계로서, 상기 서열 리드는 단일 뉴클레오타이드 해상에서 무세포 DNA 분자에 대한 복수의 부위에서의 메틸화 상태를 함유하는 것인 단계;
- (b) 상기 서열 리드를 분석하여 복수의 부위 각각에서 메틸화된 무세포 DNA 분자의 수를 결정하고, 상기 수를 이용하여 상기 복수의 부위에 대한 메틸화 상태에 기초하여 상기 복수의 부위에 대한 메틸화 프로필을 결정하는 단계; 및
- (c) 상기 메틸화 프로필에 적어도 부분적으로 기초하여 상기 생물학적 표본으로부터 무세포 DNA 분자의 적어도 일부에 대한 조직 원천을 결정하는 단계로서, 상기 조직 원천을 결정하는 단계는 상기 메틸화 프로필을 하나 이상의 기준 메틸화 프로필과 비교하는 것을 포함하고, 상기 하나 이상의 기준 메틸화 프로필의 적어도 하나는 암을 갖는 것으로 알려진 대상자로부터 얻어진 하나 이상의 표본으로부터 얻어진 것인 단계를 포함하는 것인 방법.

청구항 35

제34항에 있어서, 상기 무세포 DNA 분자의 적어도 일부는 암-유래 분자이며, 상기 방법은 메틸화 프로필에 적어도 부분적으로 기초하여 상기 대상자의 암의 유형을 결정하는 단계를 추가로 포함하는 것인 방법.

청구항 36

제35항에 있어서, 상기 암의 유형은 폐암, 유방암, 결직장암, 전립선암, 코인두암, 위암, 정소암, 피부암, 신경계에 영향을 주는 암, 뼈암, 난소암, 간암, 혈액 악성물, 췌장암, 자궁내막암종, 및 신장암으로 이루어진 군으로부터 선택되는 것인 방법.

청구항 37

제34항에 있어서, 상기 무세포 DNA 분자를 서열분석하여 서열 리드를 얻는 단계를 추가로 포함하는 것인 방법.

청구항 38

제37항에 있어서, 상기 서열분석은 메틸화-인식 서열분석을 포함하는 것인 방법.

청구항 39

제38항에 있어서, 상기 메틸화-인식 서열분석은 바이셀파이트 서열분석을 포함하는 것인 방법.

청구항 40

제39항에 있어서, 상기 복수의 부위는 하나 이상의 CpG 부위를 포함하는 것인 방법.

청구항 41

컴퓨터가 제1항 내지 제7항, 제15항 내지 제36항 및 제40항 중 어느 한 항의 방법을 실행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 매체.

청구항 42

하나 이상의 프로세서 및 제41항의 컴퓨터 판독 가능한 매체를 포함하는 시스템.

청구항 43

제1항 내지 제7항, 제15항 내지 제36항 및 제40항 중 어느 한 항의 방법을 구현하도록 프로그램된 하나 이상의 프로세서를 함유하는 컴퓨터 시스템.

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

청구항 101

삭제

청구항 102

삭제

청구항 103

삭제

청구항 104

삭제

청구항 105

삭제

청구항 106

삭제

청구항 107

삭제

청구항 108

삭제

청구항 109

삭제

청구항 110

삭제

청구항 111

삭제

청구항 112

삭제

청구항 113

삭제

청구항 114

삭제

청구항 115

삭제

청구항 116

삭제

청구항 117

삭제

청구항 118

삭제

청구항 119

삭제

청구항 120

삭제

청구항 121

삭제

청구항 122

삭제

청구항 123

삭제

청구항 124

삭제

청구항 125

삭제

청구항 126

삭제

청구항 127

삭제

청구항 128

삭제

청구항 129

삭제

청구항 130

삭제

청구항 131

삭제

청구항 132

삭제

청구항 133

삭제

청구항 134

삭제

발명의 설명

기술 분야

[0001] 관련 출원에 대한 교차 참조

[0002] 본 출원은 2013. 6. 3.에 출원되고 "Tumor Detection In Plasma Using Methylation Status And Copy Number"를 표제로 하는 U.S. 특허 가출원 번호 61/830,571; 및 2012. 9. 20.에 출원되고 "Method Of Determining The Whole Genome DNA Methylation Status Of The Placenta By Massively Parallel Sequencing Of Maternal Plasma"를 표제로 하는 U.S. 특허 가출원 번호 61/703,512의 정규 출원이고 그 이익을 향유하는 2013. 3. 15.에 출원되고 "Non-Invasive Determination Of Methylome Of Fetus Or Tumor From Plasma"를 표제로 하는 U.S. 출원 번호 13/842,209를 우선권으로 청구하는 PCT 출원이며, 본원에 이들의 전문이 모든 목적에 대해 참조로 도입된다.

[0003] 분야

[0004] 본 개시는 일반적으로 DNA의 메틸화 패턴(메틸롬)의 결정, 보다 구체적으로 상이한 게놈 유래(예로, 태아 및 모체 유래, 또는 종양 및 정상 세포 유래) DNA 혼합물을 포함하는 생물학적 표본(예로, 혈장)을 분석하여 소수 게놈의 메틸화 패턴(메틸롬)을 결정하는 것에 관한 것이다. 결정된 메틸롬의 용도도 기재된다.

배경 기술

[0005] 배아 및 태아 발생은 복잡한 절차이며 일련의 고도로 조직화된 유전적 및 후생적 사건이 수반된다. 암 발생도 전형적으로 여러 유전적 및 후생적 단계가 수반되는 복잡한 절차이다. 발생 절차의 후생적 제어 이상이 불임,

자연 유산, 자궁내 성장 이상 및 출생 후 결과에 수반된다. DNA 메틸화는 가장 빈번하게 연구되는 후생적 기전 중 하나이다. DNA의 메틸화는 대부분 CpG 디뉴클레오타이드 중 시토신 잔기의 5' 탄소에 대한 메틸기 부가의 맥락에서 일어난다. 시토신 메틸화는 유전자 전사 및 DNA 기능에 대한 제어층을 부가한다. 예를 들어, CpG 섬으로 불리는 CpG 디뉴클레오타이드가 풍부한 유전자 프로모터의 고메틸화는 전형적으로 유전자 기능의 억제에 연관된다.

[0006] 발생 절차의 매개에서 후생적 기전의 중요한 역할에도 불구하고, 인간 배아 및 태아 조직은 분석을 위해 쉽게 접근할 수 없다(종양도 유사하게 접근 불가능할 수 있다). 인간의 산전 기간 동안 건강 및 질환에서의 이러한 후생적 절차의 동적 변화 연구는 실질적으로 불가능하다. 산전 진단 절차의 일환으로 또는 출생 후 획득될 수 있는 배아 외 조직, 특히 태반이 이러한 연구를 위한 주요 방법 중 하나를 제공해왔다. 그러나 이러한 조직은 침습적 절차를 필요로 한다.

[0007] 인간 태반의 DNA 메틸화 프로파일은 수십 년 간 연구자들의 흥미를 끌었다. 인간 태반은 DNA 메틸화가 수반되는 독특한 생리학적 특성의 과다를 나타낸다. 전반적 수준에서, 태반 조직은 대부분의 신체 조직과 비교될 때 저메틸화된다. 유전자 수준에서, 선택된 게놈 유전자좌의 메틸화 상태는 태반 조직의 특이적 성질이다. 전반적 및 유전자좌-특이적 메틸화 프로파일은 모두 임신 나이 의존적 변화를 나타낸다. 각인 유전자, 즉 그 발현이 대립유전자의 부모 기원에 의존하는 유전자가 태반에서 주요 기능을 수행한다. 태반은 위-악성으로 설명되었으며, 몇몇 종양 억제인자 유전자의 고메틸화가 관찰되었다.

[0008] 태반 조직의 DNA 메틸화 프로파일 연구는 임신 연관 또는 발생 관련 질환, 예컨대 자간전증 및 자궁내 성장 제한의 병태생리에 대한 식견을 제공하였다. 게놈 각인 장애는 발생 장애, 예컨대 프라더-윌리 증후군 및 앵겔만 증후군에 연관된다. 태반 및 태아 조직에서의 게놈 각인 및 전반적 DNA 메틸화의 변경된 프로파일이 보조 생식 기법으로 일어난 임신에서 관찰되었다(H Hiura 등, 2012 Hum Reprod; 27: 2541-2548). 여러 환경적 요인, 예컨대 모체 흡연(KE Haworth 등, 2013 Epigenomes; 5: 37-49), 모체 식이 요인(X Jiang 등, 2012 FASEB J; 26: 3563-3574) 및 모체 대사 상태, 예컨대 당뇨병(N Hajj 등, Diabetes. doi: 10.2337/db12-0289)은 후손의 후생적 이상에 연관되었다.

[0009] 수십 년 간의 노력에도 불구하고, 임신 전반에 걸쳐 또는 질환 절차, 예컨대 악성화 동안 태아 또는 종양 메틸롬을 연구하고 동적 변화를 모니터링할 수 있는 어떠한 실제적 수단이 없었다. 따라서, 태아 메틸롬 및 종양 메틸롬의 전체 또는 일부를 비침습적으로 분석하기 위한 방법을 제공하는 것이 바람직하다.

발명의 내용

[0010] 요약

[0011] 구현예는 다양한 조직 및 표본의 메틸화 프로파일을 결정하고 이용하기 위한 시스템, 방법, 및 장치를 제공한다. 본원에는 그 예가 제공되어 있다. 혈장 메틸화(또는 무세포 DNA를 갖는 다른 표본, 예로 소변, 침, 생식기 세척액)와 모체/환자의 메틸화 프로파일의 비교에 기반하여 태아/종양 조직에 대해 임의의 메틸화 프로파일 추정될 수 있다. 표본이 DNA 혼합물을 갖는 경우 태아/종양 유래 DNA를 확인하기 위해 조직-특이적 대립유전자를 이용하여 태아/종양 조직에 대해 임의의 메틸화 프로파일 결정될 수 있다. 태아/종양의 게놈에서 복제수 변동을 결정하는데 임의의 메틸화 프로파일 이용될 수 있다. 태아에 대한 메틸화 마커가 다양한 기법을 통해 확인되고 있다. 그 메틸화 프로파일은 DNA 단편 크기 분포의 크기 파라미터를 결정함으로써 결정될 수 있고, 여기서 메틸화 수준을 결정하는데 크기 파라미터에 대한 기준값이 이용될 수 있다.

[0012] 추가로, 암 수준을 결정하는데 임의의 메틸화 수준이 이용될 수 있다. 암의 맥락에서, 혈장 중 메틸롬(methylomic) 변화의 측정으로, 모니터링을 위해(예를 들면, 항암 치료 후 반응을 검출하기 위해; 그리고 암 재발을 검출하기 위해), 그리고 예진을 위해(예를 들어, 체내 암 세포 부하를 측정하기 위해, 또는 단계분석 목적을 위해, 또는 질환 또는 질환 진행 또는 전이 과정으로부터의 사망 가능성을 평가하기 위해) 암을 검출할 수 있다(예를 들면, 스크리닝 목적을 위해).

[0013] 본 발명의 구현예의 성질 및 장점의 더 우수한 이해는 하기 상세한 설명 및 첨부 도면을 참조하여 얻어질 수 있다.

도면의 간단한 설명

[0014] 도 1a는 본 발명의 구현예에 따른 모체 혈액, 태반, 및 모체 혈장에 대한 서열분석 결과의 표 100을 나타낸다.

도 1b는 본 발명의 구현예에 따른 서열분석 표본의 1Mb 윈도우에서의 메틸화 밀도를 나타낸다.

도 2a-2c는 메틸화 지수에 대한 베타-값의 그래프를 나타낸다: (a) 모체 혈액 세포, (b) 융모막 융모 표본, (c) 만기 태반 조직.

도 3a 및 3b는 성인 남성 및 비임신 성인 여성에서 수집된 혈장 및 혈액 세포에서 메틸화된 CpG 부위 백분율의 막대 그래프를 나타낸다: (a) 상염색체, (b) 염색체 X.

도 4a 및 4b는 혈액 세포 DNA 및 혈장 DNA에서 대응 유전자좌의 메틸화 밀도 그래프를 나타낸다: (a) 비임신 성인 여성, (b) 성인 남성.

도 5a 및 5b는 임신부에서 수집된 표본 중 메틸화 CpG 부위 백분율의 막대 그래프를 나타낸다: (A) 상염색체, (B) 염색체 X.

도 6은 모체 혈액, 태반 및 모체 혈장에 대한 인간 게놈의 상이한 반복 클래스의 메틸화 수준의 막대 그래프를 나타낸다.

도 7a는 제1 석달 표본에 대한 Circos 그래프 700을 나타낸다. 도 7b는 제3 석달 표본에 대한 Circos 그래프 750을 나타낸다.

도 8a-8d는 정보를 제공하는 단일 뉴클레오타이드 다형성 주변의 CpG 부위에 대한 모체 혈장 DNA 대비 게놈 조직 DNA의 메틸화 밀도의 비교 그래프를 나타낸다.

도 9는 본 발명의 구현예에 따른 유기체의 생물학적 표본으로부터 제1 메틸화 프로필을 결정하기 위한 방법 900을 예시하는 순서도이다.

도 10은 본 발명의 구현예에 따른 유기체의 생물학적 표본으로부터 제1 메틸화 프로필을 결정하기 위한 방법 1000을 예시하는 순서도이다.

도 11a 및 11b는 본 발명의 구현예에 따른 모체 혈장 데이터 및 분획 태아 DNA 농도를 이용한 예측 알고리즘의 수행 그래프를 나타낸다

도 12a는 본 발명의 구현예에 따른 메틸화 예측을 위한 15개의 선택된 게놈 유전자좌의 상세내용을 나타내는 표 1200이다. 도 12b는 태반에서 15개의 선택된 게놈 유전자좌 및 이들의 대응 메틸화 수준의 추정 범주를 나타내는 그래프 1250이다.

도 13은 적어도 하나의 태아를 임신한 여성 대상자의 생물학적 표본에서 태아 염색체 이상을 검출하기 위한 방법 1300의 순서도이다.

도 14는 본 발명의 구현예에 따라 태반 메틸화 프로필과 모체 메틸화 프로필을 비교하여 메틸화 마커를 확인하기 위한 방법 1400의 순서도이다.

도 15a는 33개의 이전에 보고된 제1 석달 마커를 참조하여 제1 석달 데이터를 이용한 DMR 확인 알고리즘의 수행을 나타내는 표 1500이다. 도 15b는 출산 시 수득한 태반 표본과 비교되는 제3 석달 데이터를 이용한 DMR 확인 알고리즘의 성능을 나타내는 표 1550이다.

도 16은 모체 혈장 바이셀라이트-서열분석 데이터의 직접 분석에 기반하여 고메틸화 또는 저메틸화된 것으로 예측되는 유전자좌의 수를 나타내는 표 1600이다.

도 17a는 모체 혈장, 비임신 여성 대조군 혈장, 태반 및 말초 혈액 DNA의 크기 분포를 나타내는 그래프 1700이다. 도 17b는 모체 혈장, 성인 여성 대조군 혈장, 태반 조직 및 성인 여성 대조군 혈액의 크기 분포 및 메틸화 프로필의 그래프 1750이다.

도 18a 및 18b는 본 발명의 구현예에 따른 혈장 DNA 분자의 메틸화 밀도 및 크기 그래프이다.

도 19a는 성인 비임신 여성에 대한 서열분석 측정치의 크기 및 메틸화 밀도의 그래프 1900을 나타낸다. 도 19b는 모체 혈장 중 태아-특이적 및 모체-특이적 DNA 분자의 크기 분포 및 메틸화 프로필을 나타내는 그래프 1950이다.

도 20은 본 발명의 구현예에 따른 유기체의 생물학적 표본 중 DNA의 메틸화 수준을 산정하기 위한 방법 2000의 순서도이다.

도 21a는 간세포암종(HCC) 환자의 수술 전 혈장 및 조직 표본의 메틸화 밀도를 나타내는 표 2100이다. 도 21b는 표본 별로 달성된 서열 리드의 수 및 서열분석 깊이를 나타내는 표 2150이다.

도 22는 건강한 대조군의 혈장 표본 중 71.2% 내지 72.5% 범위의, 상염색체에서의 메틸화 밀도를 나타내는 표 220이다.

도 23a 및 23b는 HCC 환자의 버피 코트, 종양 조직, 비종양 간 조직, 수술 전 혈장 및 수술 후 혈장의 메틸화 밀도를 나타낸다.

도 24a는 HCC 환자로부터의 수술 전 혈장의 메틸화 밀도를 나타내는 그래프 2400이다. 도 24b는 HCC 환자로부터의 수술 후 혈장의 메틸화 밀도를 나타내는 그래프 2450이다.

도 25a 및 25b는 염색체 1에 대한 기준으로 4명의 건강한 대조군 대상자의 혈장 메틸롬 데이터를 이용하여 HCC 환자의 수술 전(그래프 2500) 및 수술 후(그래프 2550) 혈장 표본에 대한 혈장 DNA 메틸화 밀도의 z-스코어를 나타낸다.

도 26a는 수술 전 및 수술 후 혈장에 대한 z-스코어 데이터를 나타내는 표 2600이다. 도 26b는 전체 상염색체에서 분석된 1Mb bin에 대한 기준으로 4명의 건강한 대조군 대상자를 이용하여 HCC 환자의 수술 전 및 수술 후 혈장 표본에 대한 혈장 DNA 메틸화 밀도의 z-스코어를 나타내는 Circos 그래프 2620이다. 도 26c는 HCC 환자의 수술 전 및 수술 후 혈장 표본 모두에서 전체 게놈에 대한 1Mb bin의 z-스코어 분포를 나타내는 표 2640이다. 도 26d는 CHH 및 CHG 맥락을 이용하는 경우 일부 대조군 혈장 표본과 중복되는 종양 조직 및 수술 전 혈장 표본의 메틸화 수준을 나타내는 표 2660이다.

도 27a-h는 본 발명의 구현예에 따른 8명의 암 환자의 메틸화 밀도의 Circos 그래프를 나타낸다. 도 27i는 표본 별로 달성한 서열 리드의 수 및 서열분석 깊이를 나타내는 표 2780이다. 도 27j는 상이한 악성물을 갖는 환자의 혈장에서 전체 게놈에 대해 1Mb bin의 z-스코어 분포를 나타내는 표 2790이다. CL=폐의 선암종; NPC=코인두암종; CRC=결장직장암종; NE=신경내분비암종; SMS=평활근육종.

도 28은 본 발명의 구현예에 따른 암 수준의 분류를 결정하기 위한 유기체의 생물학적 표본을 분석하는 방법 2800의 순서도이다.

도 29a는 상기 분포가 정규 분포를 따른다고 가정하는 기준 대상자에서 메틸화 밀도의 분포를 나타내는 그래프 2900이다. 도 29b는 상기 분포가 정규 분포를 따른다고 가정하는 암 대상자에서 메틸화 밀도의 분포를 나타내는 그래프 2950이며, 평균 메틸화 수준은 컷 오프의 2 표준 편차 미만이다.

도 30은 건강한 대상자 및 암 환자의 혈장 DNA의 메틸화 밀도 분포를 나타내는 그래프 3000이다.

도 31은 HCC 환자의 종양 조직 및 건강한 대상자의 혈장 DNA의 평균 간 메틸화 밀도 차이의 분포를 나타내는 그래프 3100이다.

도 32a는 혈장 표본이 5% 또는 2% 종양 DNA를 함유한 경우 서열분석 깊이의 감소 효과를 나타내는 표 3200이다.

도 32b는 4명의 건강한 대조군 대상자의 혈장, 버피 코트, 정상 간 조직, HCC 환자의 종양 조직, 수술 전 혈장 및 수술 후 혈장 표본 중 반복 요소 및 비반복 영역의 메틸화 밀도를 나타내는 그래프 3250이다.

도 33은 본 발명의 구현예에 따른 시스템 및 방법에 이용 가능한 예시적 컴퓨터 시스템 3300의 블록 구성도를 나타낸다.

도 34a는 전신 홍반 루푸스(SLE) 환자 SLE04에서 혈장 DNA의 크기 분포를 나타낸다. 도 34b 및 34c는 SLE 환자 SLE04(도 34b) 및 HCC 환자 TBR36(도 34c)으로부터의 혈장 DNA에 대한 메틸화 분석을 나타낸다.

도 35는 본 발명의 구현예에 따른 CpG 섬의 고메틸화에 기반한 암 수준의 분류를 결정하는 방법 3500의 순서도이다.

도 36은 본 발명의 구현예에 따른 복수의 염색체 영역을 이용한 유기체의 생물학적 표본을 분석하는 방법 3600의 순서도이다.

도 37a는 환자 TBR36에 대한 종양 조직, 바이셀파이트(BS)-미처리 혈장 DNA 및 바이셀파이트-처리 혈장 DNA(내부에서 외부로)에 대한 CNA 분석을 나타낸다. 도 37b는 환자 TBR36에 대한 1Mb bin의 바이셀파이트-처리 및 바이셀파이트-미처리 혈장을 이용한 CNA의 검출을 위한 z-스코어 간 관계를 나타내는 산포도이다.

도 38a는 환자 TBR34에 대한 종양 조직, 바이셀과이트(BS)-미처리 혈장 DNA 및 바이셀과이트-처리 혈장 DNA(내부에서 외부로)에 대한 CNA 분석을 나타낸다. 도 38b는 환자 TBR34에 대한 1Mb bin의 바이셀과이트-처리 및 바이셀과이트-미처리 혈장을 이용한 CNA의 검출을 위한 z -스코어 간 관계를 나타내는 산포도이다.

도 39a는 HCC 환자 TBR240의 바이셀과이트-처리 혈장에 대한 CNA(내부 고리) 및 메틸화 분석(외부 고리)을 나타내는 Circos 그래프이다. 도 39b는 HCC 환자 TBR164의 바이셀과이트-처리 혈장에 대한 CNA(내부 고리) 및 메틸화 분석(외부 고리)을 나타내는 Circos 그래프이다.

도 40a는 처리-전 표본 및 처리-후 표본에 대한 환자 TBR36의 CNA 분석을 나타낸다. 도 40b는 처리-전 표본 및 처리-후 표본에 대한 환자 TBR36의 메틸화 분석을 나타낸다. 도 41a는 처리-전 표본 및 처리-후 표본에 대한 환자 TBR34의 CNA 분석을 나타낸다. 도 41b는 처리-전 표본 및 처리-후 표본에 대한 환자 TBR34의 메틸화 분석을 나타낸다.

도 42는 상이한 수의 서열분석 측정치를 갖는 게놈에 걸친 저메틸화 분석의 진단 수행 도형을 나타낸다.

도 43은 상이한 bin 크기(50kb, 100kb, 200kb 및 1Mb)를 갖는 게놈에 걸친 저메틸화 분석에 기반한 암 검출을 위한 ROC 곡선을 나타내는 도형이다.

도 44a는 이상이 있는 bin의 누적 확률(CP) 및 백분율에 대한 진단 성능을 나타낸다. 도 44b는 전반적 저메틸화, CpG 섬 고메틸화 및 CNA에 대한 혈장 분석을 위한 진단 성능을 나타낸다.

도 45는 간세포암종 환자에서 전반적 저메틸화, CpG 섬 고메틸화 및 CNA에 대한 결과표를 나타낸다.

도 46은 간세포암종 이외의 암 환자에서의 전반적 저메틸화, CpG 섬 고메틸화 및 CNA에 대한 결과표를 나타낸다.

도 47은 케이스 TBR34의 혈장 메틸화에 대한 연속 분석을 나타낸다.

도 48a는 HCC 환자 TBR36에 대한 바이셀과이트-처리 혈장 DNA 중 CNA(내부 고리) 및 메틸화 변화(외부 고리)를 나타내는 Circos 그래프를 나타낸다. 도 48b는 HCC 환자 TBR36에 대해 염색체 획득 및 손실을 갖는 영역 그리고 복제수 변화가 없는 영역에 대한 메틸화 z -스코어의 그래프이다.

도 49a는 HCC 환자 TBR34에 대한 바이셀과이트-처리 혈장 DNA 중 CNA(내부 고리) 및 메틸화 변화(외부 고리)를 나타내는 Circos 그래프를 나타낸다. 도 49b는 HCC 환자 TBR34에 대해 염색체 획득 및 손실을 갖는 영역, 및 복제수 변화가 없는 영역에 대한 메틸화 z -스코어의 그래프이다.

도 50a 및 50b는 SLE 환자 SLE04 및 SLE10에 대한 혈장 저메틸화 및 CNA 분석 결과를 나타낸다.

도 51a 및 51b는 두 HCC 환자(TBR34 및 TBR36)의 혈장에 대해 CNA를 포함하는 및 포함하지 않는 영역에 대한 Z_{meth} 분석을 나타낸다. 도 51c 및 51d는 두 SLE 환자(SLE04 및 SLE10)의 혈장에 대해 CNA를 포함하는 및 포함하지 않는 영역에 대한 Z_{meth} 분석을 나타낸다.

도 52a는 CNA, 전반적 메틸화, 및 CpG 섬 메틸화에 대한 A군 특성을 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. 도 52b는 CNA, 전반적 메틸화, 및 CpG 섬 메틸화에 대한 B군 특성을 이용한 계층 클러스터링을 나타낸다.

도 53a는 A군 CpG 섬 메틸화 특성을 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. 도 53b는 A군의 전반적 메틸화 밀도를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다.

도 54a는 A군의 전반적 CNA를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. 도 54b는 B군의 CpG 섬 메틸화 밀도를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다.

도 55a는 B군의 전반적 메틸화 밀도를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. 도 55b는 B군의 전반적 메틸화 밀도를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다.

도 56은 32명의 건강한 대상자 간 1Mb bin(빨간 점)의 평균 메틸화 밀도를 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0015] 정의
- [0016] "메틸롬"은 게놈의 복수의 부위 또는 유전자좌에서 DNA 메틸화량의 측정치를 제공한다. 메틸롬은 게놈의 전체, 게놈의 상당 부분 또는 게놈의 상대적으로 작은 부분(들)에 대응할 수 있다. "태아 메틸롬"은 임신 여성의 태아의 메틸롬에 대응한다. 태아 메틸롬은 모체 혈장 중의 무세포 태아 DNA 및 태반 조직을 포함하는 다양한 태아 조직 또는 태아 DNA의 원천을 이용해서 결정될 수 있다. "종양 메틸롬"은 유기체(예로 인간)의 종양의 메틸롬에 대응한다. 종양 메틸롬은 모체 혈장 중 무세포 종양 DNA 또는 종양 조직을 이용하여 결정될 수 있다. 태아 메틸롬 및 종양 메틸롬이 관심 메틸롬의 예이다. 관심 메틸롬의 다른 예는 체액(예로 혈장, 혈청, 땀, 침, 소변, 생식기 분비물, 정액, 대변액, 설사액, 뇌척수액, 위장관 분비물, 체장 분비물, 장 분비물, 가래, 눈물, 유방 및 갑상샘으로부터의 흡인액 등) 내로의 DNA에 기여할 수 있는 기관의 메틸롬(예로 뇌 세포, 뼈, 폐, 심장, 근육 및 신장 등의 메틸롬)이다. 기관은 이식된 기관일 수 있다.
- [0017] "혈장 메틸롬"은 동물(예로 인간)의 혈장 또는 혈청에서 결정된 메틸롬이다. 혈장 메틸롬은 혈장 및 혈청에 무세포 DNA가 포함되므로, 무세포 메틸롬의 예이다. 혈장 메틸롬은 태아/모체 메틸롬 또는 종양/환자 메틸롬의 혼합물이므로, 또한 혼합 메틸롬의 예이다. "태반 메틸롬"은 융모막 융모 표본(CVS) 또는 태반 조직 표본(예로, 출산 후 수득된 것)에서 결정될 수 있다. "세포 메틸롬"은 환자의 세포(예로 혈액 세포)에서 결정되는 메틸롬에 대응한다. 혈액 세포의 메틸롬은 혈액 세포 메틸롬(또는 혈액 메틸롬)으로 불린다.
- [0018] "부위"는 단일 부위에 대응하며, 이는 단일 염기 위치 또는 관련된 염기 위치의 그룹, 예로 CpG 부위일 수 있다. "유전자좌"는 다중 부위를 포함하는 영역에 대응할 수 있다. 유전자좌에는 단 하나의 부위가 포함될 수 있고, 이는 그 맥락에서 유전자좌를 부위와 동등하게 만들 것이다.
- [0019] 각각의 게놈 부위(예로, CpG 부위)에 대한 "메틸화 지수"는 그 부위를 커버하는 측정치의 총 수에 걸쳐 부위에서 메틸화를 나타내는 서열 리드 비율을 나타낸다. 영역의 "메틸화 밀도"는 영역에서 부위를 커버하는 측정치의 총 수로 나눈 메틸화를 나타내는 영역 내 부위에서의 측정치의 수이다. 부위는 특이적 특징을 가질 수 있다, 예로 CpG 부위일 수 있다. 따라서, 영역의 "CpG 메틸화 밀도"는 영역에서 CpG 부위(예로, 특정 CpG 부위, CpG 섬 내에서의 CpG 부위, 또는 더 큰 영역)를 커버하는 측정치의 총 수로 나눈 CpG 메틸화를 나타내는 측정치의 수이다. 예를 들어, 인간 게놈에서 각각의 100kb bin에 대한 메틸화 밀도는 CpG 부위에서 바이셀파이트 처리 후 전환되지 않은 시토신(메틸화 시토신에 대응함)의 총 수에서 100-kb 영역에 맵핑된 서열 리드로 커버되는 모든 CpG 부위의 비율로 결정될 수 있다. 상기 분석은 다른 bin 크기, 예로 50-kb 또는 1-Mb 등에 대해서도 수행될 수 있다. 영역은 전체 게놈 또는 염색체 또는 염색체의 일부(예로 염색체 팔)일 수 있다. CpG 부위의 메틸화 지수는 영역에 CpG 부위만 포함되는 경우의 영역에 대한 메틸화 밀도와 동일하다. "메틸화 시토신의 비율"은 분석된 시토신 잔기의 총 수에 걸쳐, 즉 영역에서 CpG 맥락 외부의 시토신을 포함하여 메틸화된 것으로 나타나는(예를 들어 바이셀파이트 전환 후 미전환된) 시토신 부위 "C"의 수를 나타낸다. 메틸화 지수, 메틸화 밀도 및 메틸화 시토신의 비율은 "메틸화 수준"의 예이다.
- [0020] "메틸화 프로파일"(또한 메틸화 상태로도 불림)에는 영역에 대한 DNA 메틸화에 관련된 정보가 포함된다. DNA 메틸화에 관련된 정보에는 비제한적으로 CpG 부위의 메틸화 지수, 영역에서 CpG 부위의 메틸화 밀도, 인접 영역에 걸친 CpG 부위의 분포, 둘 이상의 CpG 부위를 함유하는 영역 내의 각각의 개별 CpG 부위에 대한 메틸화 패턴 또는 수준 및 비-CpG 메틸화가 포함될 수 있다. 게놈의 상당 부분의 메틸화 프로파일은 메틸롬과 동등한 것으로 간주될 수 있다. 포유류 게놈에서의 "DNA 메틸화"는 전형적으로 CpG 디뉴클레오타이드 간 시토신 잔기의 5' 탄소에 대한 메틸기 첨가(즉 5-메틸 시토신)를 나타낸다. DNA 메틸화는 다른 맥락에서, 예를 들어 CHG 및 CHH 중 시토신에서 일어날 수 있고, 여기서 H는 아데닌, 시토신 또는 티민이다. 시토신 메틸화는 또한 5-히드록시메틸시토신의 형태일 수 있다. 비-시토신 메틸화, 예컨대 N6-메틸아데닌도 보고되었다.
- [0021] "조직"은 임의 세포에 대응한다. 상이한 유형의 조직은 상이한 유형의 세포(예로, 간, 폐, 또는 혈액)에 대응할 수 있지만, 또한 상이한 유기체(모체 vs. 태아)로부터의 조직 또는 건강한 세포 vs. 종양 세포에 대응할 수도 있다. "생물학적 표본"은 대상자(예로 인간, 예컨대 임신 여성, 암이 있는 사람 또는 암이 있는 것으로 추정되는 사람, 기관 이식 수령자 또는 기관(예로 심근경색에서의 심장 또는 뇌졸중에서의 뇌)를 갖는 것으로 추정되는 대상자)이 수반된 질환 절차에서 얻어지고, 하나 이상의 관심 핵산 분자(들)를 함유하는 임의 표본을 나타낸다. 생물학적 표본은 체액, 예컨대 혈액, 혈장, 혈청, 소변, 질액, 자궁 또는 질 세척액, 복수, 양수, 뇌척수액, 침, 땀, 눈물, 가래, 기관지폐포 세척액 등일 수 있다. 대변 표본도 이용될 수 있다.

- [0022] 용어 "암의 수준"은 암이 존재하는지의 여부, 암의 단계, 종양의 크기, 전이가 존재하는지의 여부, 신체의 총 종양 부담, 및/또는 암의 다른 중증도 측정치를 나타낼 수 있다. 암의 수준은 수 또는 다른 특징일 수 있다. 수준은 0일 수 있다. 암의 수준에는 또한 돌연변이 또는 여러 돌연변이와 연관된 전악성 또는 전암성 상태(단계)가 포함된다. 암의 수준은 다양한 방식으로 이용될 수 있다. 예를 들어, 스크리닝은 암이 이전에 암을 가진 것으로 알려지지 않은 사람에서 존재하는지를 확인할 수 있다. 평가는 암으로 진단받은 사람을 조사하여 경시적으로 암의 진행을 모니터링하거나, 치료법의 유효성을 연구하거나 또는 예후를 결정할 수 있다. 하나의 구현예에서, 예후는 환자가 암으로 죽을 가능성 또는 특정 기간 또는 시간 후 암이 진행할 가능성 또는 암의 전이 가능성으로 표현될 수 있다. 검출은 암을 제시하는 특성(예로 증상 또는 다른 양성 평가)이 있는 사람이 암을 갖는지의 '스크리닝'을 의미할 수도 있고 확인을 의미할 수도 있다.
- [0023] 상세한 설명
- [0024] 후생적 기전은 배아 및 태아 발생에서 중요한 역할을 수행한다. 그러나, 인간 배아 및 태아 조직(태반 조직 포함)에는 쉽게 접근할 수 없다(U.S. 특허 6,927,028). 특정 구현예에서는 모체 순환 중에 존재하는 무세포 태아 DNA 분자를 갖는 표본을 분석하여 상기 문제를 해결하였다. 태아 메틸롬은 다양한 방식으로 추정될 수 있다. 예를 들어, 모체 혈장 메틸롬이 세포 메틸롬(모체의 혈액 세포 유래)과 비교될 수 있고, 차이는 태아 메틸롬과 관련된 것으로 나타난다. 다른 예로서, 태아-특이적 대립유전자를 이용하여 특정 유전자좌에서 태아 메틸롬의 메틸화를 결정할 수 있다. 또한, 크기 및 메틸화 백분율 간 상관관계가 나타나므로, 단편 크기가 메틸화 백분율의 지표로 이용될 수 있다.
- [0025] 하나의 구현예에서, 게놈에 걸친 바이셀과이트 서열분석을 이용하여 단일 뉴클레오타이드 해상에서 모체 혈장 DNA의 메틸화 프로파일(메틸롬의 일부 또는 전체)을 분석한다. 모체 및 태아 간 다형성 차이를 탐색하여, 태아 메틸롬이 모체 혈액 표본에서 조립될 수 있다. 다른 구현예에서는, 다형성 차이가 이용되지 않았지만 혈장 메틸롬 및 혈액 세포 메틸롬 간 차이가 이용될 수 있다.
- [0026] 다른 구현예에서, 종양 게놈 및 비종양 게놈 간 단일 뉴클레오타이드 변동 및/또는 복제수 이상, 및 혈장(또는 다른 표본) 서열분석 데이터를 탐색함으로써, 종양의 메틸화 프로파일 확인이 암을 갖는 것으로 추정되거나 알려진 환자의 표본에서 수행될 수 있다. 건강한 대조군 또는 건강한 대조군 그룹의 혈장 메틸화 수준과 비교되는 경우, 평가 개인의 혈장 표본 중 메틸화 수준의 차이는 암을 보유한 평가 개인을 확인할 수 있게 한다. 또한, 메틸화 프로파일은, 예를 들어 개인에서 발생한 암의 유형, 예를 들어 기관의 유래 및 전이가 일어났는지를 드러내는 특징으로 작용할 수 있다.
- [0027] 상기 접근의 비침습적 성질로 인해, 제1 석달, 제3 석달 중 및 출산 후 수집된 모체 혈액 표본에서 태아 및 모체 혈장 메틸롬을 연속 평가할 수 있었다. 임신 관련 변화가 관찰되었다. 접근은 제2 석달 동안 수득된 표본에도 적용될 수 있다. 임신 중 모체 혈장에서 추정된 태아 메틸롬은 태반 메틸롬과 유사하였다. 각인 유전자 및 상이하게 메틸화된 영역이 모체 혈장 데이터에서 확인되었다.
- [0028] 따라서 태아 메틸롬을 비침습적으로, 연속적으로 그리고 통합적으로 연구하여 임신 관련 병리의 바이오마커 확인 또는 직접적 평가를 위한 가능성을 제공하는 접근을 개발하였다. 구현예는 대상자에 암이 있는지를 스크리닝하거나 검출하기 위해, 암 환자에서 악성 질환을 모니터링하기 위해 그리고 예진을 위해, 종양 메틸롬을 비침습적으로, 연속적으로 그리고 통합적으로 연구하기 위해서도 이용될 수 있다. 구현예는 비제한적으로 폐암, 유방암, 결직장암, 전립선암, 코인두암, 위암, 정소암, 피부암(예로 흑색종), 신경계에 영향을 주는 암, 뼈암, 난소암, 간암(예로 간세포암종), 혈액 악성물, 췌장암, 자궁내막암종, 신장암, 자궁경부암, 방광암 등을 포함하는 임의의 암 유형에 적용될 수 있다.
- [0029] 메틸롬 또는 메틸화 프로파일을 결정하는 방법에 대한 설명이 먼저 논의된 후, 상이한 메틸롬(예컨대 태아 메틸롬, 종양 메틸롬, 모체 또는 환자의 메틸롬, 및 혼합 메틸롬, 예로 혈장 유래)이 기재된다. 이어서 태아 메틸화 프로파일의 결정이 태아-특이적 마커를 이용해서 또는 혼합 메틸화 프로파일을 세포 메틸화 프로파일과 비교하여 기재된다. 태아 메틸화 마커는 메틸화 프로파일을 비교하여 결정된다. 크기 및 메틸렌 간 관계가 논의된다. 암을 검출하기 위한 메틸화 프로파일의 용도도 제공된다.
- [0030] I. 메틸롬의 결정
- [0031] 여러 접근을 이용하여 태반 메틸롬을 조사하였으나, 각각의 접근은 그 한계를 갖는다. 예를 들어, 비메틸화 시토신 잔기를 우라실로 개질하고 메틸화된 시토신은 변하지 않은 채 놔두는 화학물질인 나트륨 바이셀과이트는 시토신 메틸화의 차이를 추가 의문을 위한 유전적 서열 차이로 전환한다. 시토신 메틸화를 연구하는 최고의 방

법은 조직 DNA를 나트륨 바이설파이트로 처리한 뒤 바이설파이트 전환 DNA분자의 개별 클론의 직접적 서열분석에 기반한다. 여러 DNA 분자 클론의 분석 후, CpG 부위 별 정량적 프로파일 및 시토신 메틸화 패턴이 수득될 수 있다. 그러나 클로닝된 바이설파이트 서열분석은 계놈에 걸친 규모 상에 쉽게 적용될 수 없는 저처리량의 노동 집약적 절차이다.

[0032] 전형적으로 비메틸화 DNA를 소화시키는 메틸화-감수성 제한 효소는 DNA 메틸화를 연구하기 위한 저비용 접근을 제공한다. 그러나 이러한 연구에서 생성된 데이터는 효소 인식 모티프를 갖는 유전자좌에 제한되며 결과가 정량적이지 않다. 항-메틸화 시토신 항체에 결합된 DNA의 면역침전을 이용하여 큰 계놈 절편을 조사할 수 있지만, 이러한 영역에 대한 더 강한 항체 결합 강도로 인해 조밀한 메틸화를 갖는 유전자좌에 대해 편향되는 경향이 있다. 마이크로어레이 기반 접근은 프로브 및 표적 DNA 간 연구 프로브의 선택적 설계 및 혼성화 효율에 의존한다.

[0033] 메틸롬을 통합적으로 조사하기 위해, 일부 구현에는 대량 병렬 서열분석(MPS)을 이용하여 뉴클레오타이드 별로 그리고 대립유전자 기준 별로 메틸화 수준의 정량적 평가 및 계놈에 걸친 정보를 제공한다. 최근, 바이설파이트 전환에 이은 계놈에 걸친 MPS가 타당해졌다(R Lister 등, 2008 Cell; 133: 523-536).

[0034] 인간 메틸롬의 조사를 위해 계놈에 걸친 바이설파이트 서열분석을 적용한 소수의 공개 연구(R Lister 등, 2009 Nature; 462: 315-322; L Laurent 등, 2010 Genome Res; 20: 320-331; Y Li 등, 2010 PLoS Biol; 8: e1000533; 및 M Kulis 등, 2012 Nat Genet; 44: 1236-1242) 중에서, 두 연구가 배아 줄기 세포 및 태아 섬유아 세포에 초점을 맞췄다(R Lister 등, 2009 Nature; 462: 315-322; L Laurent 등, 2010 Genome Res; 20: 320-331). 두 연구 모두 세포주 유래 DNA를 분석하였다.

[0035] A. 계놈에 걸친 바이설파이트 서열분석

[0036] 특정 구현에는 상기 난제를 극복하고 태아 메틸롬을 통합적이고 비침습적으로 그리고 연속적으로 조사할 수 있게 한다. 하나의 구현예에서, 계놈에 걸친 바이설파이트 서열분석을 이용하여 임신 여성의 순환에서 확인되는 무세포 태아 DNA 분자를 분석하였다. 혈장 DNA 분자의 낮은 풍부도 및 단편화된 성질에도 불구하고, 모체 혈장에서 고해상 태아 메틸롬을 조립하고 임신 진행에 따른 변화를 연속 관찰할 수 있었다. 비침습적 산전 평가(NIPT)에 대한 큰 관심에 기반하여, 구현예는 태아 바이오마커 발견을 위한 강력한 신규 도구를 제공하거나 태아 또는 임신 연관 질환의 NIPT를 달성하기 위한 직접적 플랫폼으로 작용할 수 있다. 이제 태아 메틸롬이 유도될 수 있는 다양한 표본의 계놈에 걸친 바이설파이트 서열분석으로부터의 데이터가 제공된다. 하나의 구현예에서, 상기 기술이 자간전증, 또는 자궁내 성장 지연을 합병증으로 갖는 임신 또는 조산에서의 메틸화 프로파일 확인을 위해 적용될 수 있다. 이러한 합병증을 갖는 임신에서, 상기 기술은 그 비침습적 성질로 인해 연속 이용하여 모니터링 및/또는 예진 및/또는 치료에 대한 반응을 허용할 수 있다.

[0037] 도 1a는 본 발명의 구현예에 따른 모체 혈액, 태반, 및 모체 혈장에 대한 서열분석 결과표 100을 나타낸다. 하나의 구현예에서, 전체 계놈 서열분석이 제1 석달에 수집된 혈액 표본의 혈액 세포, CVS, 출산 시 수집된 태반 조직, 제1 및 제3 석달 동안 및 산후기에 수집된 모체 혈장 표본의 메틸화 DNA 라이브러리 어댑터(Illumina)를 이용해서 제조된 바이설파이트-전환 DNA 라이브러리 상에서 수행되었다(R Lister 등, 2008 Cell; 133: 523-536). 1명의 성인 남성 및 1명의 비임신 성인 여성에서 수득한 혈액 세포 및 혈장 DNA 표본도 분석되었다. 총 95억쌍의 미가공 서열 리드가 상기 연구에서 생성되었다. 각 표본의 서열분석 커버 범위를 표 100에 나타낸다.

[0038] 인간 기준 계놈에 독특하게 맵핑될 수 있던 서열 리드는 제1 석달, 제3 석달 및 산후 모체 혈장 표본에 대해 각각 50배, 34배 및 28배의 평균 홀배수체 계놈 커버 범위에 도달하였다. 계놈에서 CpG 부위의 커버 범위는 임신 부에서 수득한 표본에 대해 81% 내지 92% 범위였다. CpG 부위에 걸친 서열 리드는 제1 석달, 제3 석달 및 산후 모체 혈장 표본에 대해 각각 가닥 당 33배, 가닥 당 23배 및 가닥 당 19배의 평균 홀배수체 커버 범위에 달하였다. 모든 표본에 대한 바이설파이트 전환 효율은 >99.9%였다(표 100).

[0039] 표 100에서, 모호 비율("a"로 표시)은 기준 인간 계놈의 왓슨 및 크릭 가닥 모두에 맵핑된 측정치의 비율을 나타낸다. 람다 전환 비율은 바이설파이트 개질에 의해 "티민" 잔기로 전환된 내부 람다 DNA 대조군 중 비메틸화 시토신의 비율을 나타낸다. H는 일반적으로 A, C, 또는 T와 같다. "a"는 특정 계놈 유전자좌에 맵핑될 수 있지만 왓슨 또는 크릭 가닥에 지정될 수 없는 측정치를 나타낸다. "b"는 동일한 시작 및 끝 좌표를 갖는 페어를 이룬 측정치를 나타낸다. "c"에 있어서, 람다 DNA는 바이설파이트 전환 전에 각 표본 내로 스파이크 처리되었다. 람다 전환 비율은 바이설파이트 전환 후 시토신으로 유지되는 시토신 뉴클레오타이드의 비율을 나타내며, 성공적인 바이설파이트 전환 비율의 지표로 이용된다. "d"는 기준 인간 계놈에 존재하고 바이설파이트 전환 후 시토신 서열로 유지되는 시토신 뉴클레오타이드의 수를 나타낸다.

[0040] 바이선폴아이트 개질 동안, 비메틸화 시토신은 PCR 증폭 후 우라실, 이어서 티민으로 전환되는 반면, 메틸화 시토신은 온전히 유지될 것이다(M Frommer 등, 1992 Proc Natl Acad Sci USA;89:1827-31). 따라서 서열분석 및 정렬 후, 개별 CpG 부위의 메틸화 상태는 CpG 맥락 내 시토신 잔기에서 메틸화 서열 리드 "M"(메틸화) 카운트 및 비메틸화 서열 리드 "U"(비메틸화) 카운트에서 도출될 수 있다. 바이선폴아이트 서열분석 데이터를 이용하여, 모체 혈액, 태반 및 모체 혈장의 전체 메틸롬이 정립되었다. 모체 혈장 중 특정 유전자좌의 평균 메틸화 CpG 밀도 (또한 메틸화 밀도 MD로 불림)가 다음 식을 이용해서 계산될 수 있다:

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

[0041]

[0042] 식 중, M은 유전적 유전자좌 내 CpG 부위에서의 메틸화 측정치의 카운트이고, U는 비메틸화 측정치의 카운트이다. 유전자좌 내에 둘 이상의 CpG 부위가 있는 경우, M 및 U는 부위에 걸친 카운트에 대응한다.

[0043] B. 다양한 기법

[0044] 전술된 바와 같이, 바이선폴아이트 전환 혈장 DNA의 대량 병렬 서열분석(MPS)을 이용하여 메틸화 프로파일 확인을 수행할 수 있다. 바이선폴아이트 전환 혈장 DNA의 MPS는 무작위로 또는 샷건 방식으로 수행할 수 있다. 서열분석의 깊이는 관심 영역의 크기에 따라 변할 수 있다.

[0045] 다른 구현예에서, 바이선폴아이트 전환 혈장 DNA 중의 관심 영역(들)은 먼저 용액상 또는 고상 혼성화 기반 절차를 이용해서 포획된 후 MPS가 뒤따를 수 있다. 합성 시 서열분석 플랫폼, 예컨대 Illumina, 결찰 시 서열분석 플랫폼, 예컨대 Life Technologies의 SOLiD 플랫폼, 반도체 기반 서열분석 시스템, 예컨대 Ion Torrent 또는 Life Technologies의 Ion Proton 시스템, 또는 단일 분자 서열분석 시스템, 예컨대 Helicos 시스템 또는 Pacific Biosciences 시스템 또는 나노포어 기반 서열분석 시스템을 이용하여 대량 병렬 서열분석을 수행할 수 있다. 나노포어-기반 서열분석은, 예를 들어 지질 이중층 및 단백질 나노포어, 그리고 고상 나노포어(예컨대 그 래핀 기반 나노포어)를 이용하여 정립된 나노포어를 포함한다. 선택된 단일 분자 서열분석 플랫폼은 DNA 분자 (N6-메틸아데닌, 5-메틸시토신 및 5-히드록시메틸시토신 포함)의 메틸화 상태가 바이선폴아이트 전환 없이 직접 규명될 수 있도록 할 것이며(BA Flusberg 등, 2010 Nat Methods; 7: 461-465; J Shim 등, 2013 Sci Rep; 3:1389. doi: 10.1038/srep01389), 이러한 플랫폼의 이용은 비-바이선폴아이트 전환 표본 DNA(예로 혈장 DNA)의 메틸화 상태를 분석할 수 있도록 할 것이다.

[0046] 서열분석 이외에도 다른 기법을 이용할 수 있다. 하나의 구현예에서, 메틸화 프로파일 확인은 메틸화-특이적 PCR 또는 메틸화-감수성 제한 효소 분해에 이어 PCR 또는 결찰효소 연쇄 반응에 이어 PCR에 의해 수행될 수 있다. 다른 구현예에서, PCR은 단일 분자 또는 디지털 PCR의 형태이다(B Vogelstein 등, 1999 Proc Natl Acad Sci USA; 96: 9236-9241). 추가 구현예에서, PCR은 실시간 PCR일 수 있다. 다른 구현예에서, PCR은 멀티플렉스 PCR 일 수 있다.

[0047] II. 메틸롬의 분석

[0048] 일부 구현예는 전체 게놈 바이선폴아이트 서열분석을 이용하여 혈장 DNA의 메틸화 프로파일을 결정할 수 있다. 태아의 메틸화 프로파일은 후술되는 바와 같이 모체 혈장 DNA 표본의 서열분석에 의해 결정될 수 있다. 따라서 태아 DNA 분자(및 태아 메틸롬)가 임신 동안 비침습적으로 평가되고, 임신이 진행됨에 따라 변화가 연속적으로 모니터링되었다. 서열분석 데이터의 통합성으로 인해, 모체 혈장 메틸롬을 게놈에 걸친 규모 상에서 단일 뉴클레오타이드 해상도로 연구할 수 있었다.

[0049] 서열분석 측정치의 게놈 좌표가 알려져 있었으므로, 이들 데이터는 메틸롬의 전체적 메틸화 수준 또는 게놈 중 임의의 관심 영역을 연구하고, 상이한 유전 요소 간 비교를 수행할 수 있도록 하였다. 또한, 다중 서열 리드가 각각의 CpG 부위 또는 유전자좌를 커버하였다. 이제 메틸롬을 측정하기 위해 이용된 일부 측정법의 설명이 제공된다.

[0050] A. 혈장 DNA 분자의 메틸화

[0051] DNA 분자는 저농도로 그리고 단편화된 형태로, 전형적으로 단일 뉴클레오타이드 단위와 유사한 길이로 인간 혈장에 존재한다(YMD Lo 등, 2010 Sci Transl Med; 2: 61ra91; 및 YW Zheng 등, 2012 Clin Chem; 58: 549-558). 이러한 제한에도 불구하고, 게놈에 걸친 바이선폴아이트-서열분석 파이프라인으로 혈장 DNA 분자의 메틸화를 분석할

수 있었다. 다른 구현예에서, 선택된 단일 분자 서열분석 플랫폼이 바이설파이트 전환 없이 DNA 분자의 메틸화 상태가 직접 규명될 수 있도록 할 것이므로(BA Flusberg 등, 2010 Nat Methods; 7: 461-465; J Shim 등, 2013 Sci Rep; 3:1389. doi: 10.1038/srep01389), 이러한 플랫폼의 이용은 비-바이설파이트 전환 혈장 DNA가 혈장 DNA의 메틸화 수준을 결정하기 위해 또는 혈장 메틸롬을 결정하기 위해 이용될 수 있도록 할 것이다. 이러한 플랫폼은 N6-메틸아데닌, 5-메틸시토신, 및 5-히드록시메틸시토신을 검출할 수 있고, 이는 상이한 메틸화 형태의 상이한 생물학적 기능에 관련된 개선된 결과(예로 개선된 민감성 또는 특이성)를 제공할 수 있다. 이러한 개선된 결과는 검출을 위해 구현예를 적용하거나 특정 장애, 예로 자간전증 또는 특정 유형의 암을 모니터링할 때 유용할 수 있다.

[0052] 바이설파이트 서열분석은 또한 상이한 메틸화 형태 간을 구별할 수 있다. 하나의 구현예에서, 5-메틸시토신을 5-히드록시메틸시토신과 구별할 수 있는 추가 단계를 포함시킬 수 있다. 이러한 접근의 하나는 산화적 바이설파이트 서열분석(oxBS-seq)으로, 단일 염기 해상에서 5-메틸시토신 및 5-히드록시메틸시토신의 위치를 규명할 수 있다(MJ Booth 등, 2012 Science; 336: 934-937; MJ Booth 등, 2013 Nature Protocols; 8: 1841-1851). 바이설파이트 서열분석에서, 5-메틸시토신과 5-히드록시메틸시토신은 모두 시토신으로 측정되므로, 구별될 수 없다. 반면 oxBS-seq에서는, 칼륨 퍼루테네이트(KRuO₄)를 이용한 처리에 의해 5-히드록시메틸시토신이 5-포르밀시토신으로 특이적 산화된 후 바이설파이트 전환을 이용해서 새로 형성된 5-포르밀시토신이 우라실로 전환되어 5-히드록시메틸시토신을 5-메틸시토신과 구별할 수 있도록 한다. 따라서 5-메틸시토신 관독이 단일 oxBS-seq 수행에서 수득될 수 있고, 5-히드록시메틸시토신 수준이 바이설파이트 서열분석 결과와의 비교에 의해 추정된다. 다른 구현예에서, 5-메틸시토신은 Tet-보조 바이설파이트 서열분석(TAB-seq)을 이용하여 5-히드록시메틸시토신과 구별될 수 있다(M Yu 등, 2012 Nat Protoc; 7: 2159-2170). TAB-seq는 단일 염기 해상에서 5-히드록시메틸시토신을 확인할 수 있을 뿐만 아니라 각 개질 부위에서 그 풍부도를 결정할 수 있다. 상기 방법에는 5-히드록시메틸시토신의 β-글루코실트랜스퍼라아제 매개 보호(글루코실화) 및 5-메틸시토신의 5-카르복실시토신으로의 재조합 마우스 Tet1(mTet1)-매개 산화가 수반된다. 후속 바이설파이트 처리 및 PCR 증폭 후, 시토신 및 5-카르복실시토신(5-메틸시토신 유래)은 모두 티민(T)으로 전환되는 반면, 5-히드록시메틸시토신은 C로 관독될 것이다.

[0053] 도 1b는 본 발명의 구현예에 따른 서열분석된 표본의 1Mb 윈도우에서의 메틸화 밀도를 나타낸다. 그래프 150은 게놈에 걸친 1Mb 윈도우에서 모체 혈장 및 게놈 DNA에서의 메틸화 밀도를 나타내는 Circos 그래프이다. 외부에서 내부로: 염색체 핵형도는 시계 방향으로(중심절은 빨간색으로 나타냄), 모체 혈액(빨간색), 태반(노란색), 모체 혈장(녹색), 모체 혈장 중 공유 측정치(파란색), 및 모체 혈장 중 태아-특이적 측정치(보라색)로 배향될 수 있다. 모체 혈액 세포, 태반 및 모체 혈장의 전체적 CpG 메틸화 수준(즉, 밀도 수준)은 표 100에서 확인할 수 있다. 모체 혈액 세포 중 메틸화 수준은 일반적으로 전체 게놈에 걸친 태반의 메틸화 수준보다 높다.

[0054] B. 다른 기법에 대한 바이설파이트 서열분석의 비교

[0055] 대량 병렬 바이설파이트 서열분석을 이용하여 태반 메틸롬을 연구하였다. 또한 인간 게놈에서 약 480,000 CpG 부위를 커버하는 올리고뉴클레오타이드 어레이 플랫폼(Illumina)을 이용하여 태반 메틸롬을 연구하였다(M Kulis 등, 2012 Nat Genet; 44: 1236-1242; 및 C Clark 등, 2012 PLoS One; 7: e50233). 비드칩 기반 유전형 분석 및 메틸화 분석을 이용하는 하나의 구현예에서, 유전형 분석은 제조업체의 프로토콜에 따라 Illumina HumanOmni2.5-8 유전형 분석 어레이를 이용하여 수행되었다. 유전형은 Genome Studio 소프트웨어(Illumina)의 GenCall 알고리즘을 이용하여 호출되었다. 호출 비율은 99%를 초과하였다. 마이크로어레이 기반 메틸화 분석에 있어서, 게놈 DNA(500-800ng)는 Illumina Infinium 메틸화 분석에 대한 제조업체의 권장사항에 따라 Zymo EZ DNA 메틸화 키트(Zymo Research, Orange, CA, USA)를 이용하여 나트륨 바이설파이트로 처리하였다.

[0056] 메틸화 분석은 Infinium HD 메틸화 분석 프로토콜에 따라 50ng/μl로 4μl 바이설파이트-전환 게놈 DNA 상에서 수행하였다. 혼성화된 비드칩을 Illumina iScan 기구 상에서 스캐닝하였다. DNA 메틸화 데이터를 내부 대조군 및 배경 감산에 대해 정규화하여 GenomeStudio(v2011.1) 메틸화 모듈(v1.9.0) 소프트웨어로 분석하였다. 개별 CpG 부위에 대한 메틸화 지수를 베타값(β)으로 나타내었고, 이는 메틸화 및 비메틸화 대립유전자 간 형광 강도의 비를 이용하여 계산하였다:

[0057]
$$\beta = (\text{메틸화 대립유전자의 강도}) / (\text{비메틸화 대립유전자의 강도} + \text{메틸화 대립유전자의 강도} + 100)$$

[0058] 어레이 상에 나타내고 적어도 10배의 커버 범위에 대해 서열분석된 CpG 부위에 있어서, 어레이에 의해 수득된 베타값을 동일 부위의 서열분석에 의해 결정된 메틸화 지수와 비교하였다. 베타값은 동일한 CpG 부위를 커버하는 메틸화 및 비메틸화 프로브의 조합 강도의 비율로 메틸화 프로브의 강도를 나타내었다. 각각의 CpG 부위에 대한 메틸화 지수는 CpG를 커버하는 측정치의 총 수에 걸친 메틸화 측정치의 비율을 나타낸다.

- [0059] 도 2a-2c는 두 플랫폼에 의해 연구된 대응 CpG 부위의 게놈에 걸친 바이실라이트 서열분석에 의해 결정된 메틸화 지수에 대한 Illumina Infinium 인간메틸화 450K 비드칩 어레이에 의해 결정된 베타값의 그래프를 나타낸다: (A) 모체 혈액 세포, (B) 용모막 용모 표본, (C) 만기 태반 조직. 두 플랫폼의 데이터는 유의미하게 합치하였고, 모체 혈액 세포, CVS 및 만기 태반 조직 각각에 대해 Pearson 상관관계 계수는 0.972, 0.939 및 0.954였고, R^2 값은 0.945, 0.882 및 0.910이었다.
- [0060] 본 발명자들은 자체 서열분석 데이터를 약 27,000 CpG 부위를 커버하는 올리고뉴클레오타이드 어레이를 이용해서 12쌍의 CVS 및 모체 혈액 세포 DNA 표본의 메틸화 프로필을 연구한 Chu 등이 보고한 것과 추가 비교하였다(T Chu 등, 2011 PLoS One; 6: e14723). CVS 및 모체 혈액 세포 DNA의 서열분석 결과 및 이전 연구에서 각각의 12쌍의 표본 간 상관관계 데이터는 모체 혈액에 대해 평균 Pearson 계수(0.967) 및 R^2 (0.935)를 제공하였고, CVS에 대해 평균 Pearson 계수(0.943) 및 R^2 (0.888)를 제공하였다. 두 어레이에 제시된 CpG 부위 가운데, 본 발명자들의 데이터가 공개된 데이터와 유의미하게 연관되었다. 비-CpG 메틸화 비율은 모체 혈액 세포, CVS 및 태반 조직에 대해 <1%이었다(표 100). 이들 결과는 상당량의 비-CpG 메틸화가 주로 다능 세포로 제한된다는 현재의 믿음과도 일치하였다(R Lister 등, 2009 Nature; 462: 315-322; L Laurent 등, 2010 Genome Res; 20: 320-331).
- [0061] C. 비임신 대상자에 대한 혈장 및 혈액 메틸롬의 비교
- [0062] 도 3a 및 3b는 성인 남성 및 비임신 성인 여성에서 수집된 혈장 및 혈액 세포 중 메틸화 CpG 부위의 백분율의 막대 그래프를 나타낸다: (A) 상염색체, (B) 염색체 X. 그래프는 남성 및 비임신 여성의 혈장 및 혈액 메틸롬 간 유사성을 나타낸다. 남성 및 비임신 여성 혈장 표본에서 메틸화된 CpG 부위의 전체적 비율은 대응 혈액 세포 DNA와 거의 동일하였다(표 100 및 도 2a 및 2b).
- [0063] 다음으로 유전자좌-특이적 방식으로 혈장 및 혈액 세포 표본의 메틸화 프로필의 상관관계를 연구하였다. CpG 부위에서 미전환 시토신의 총 수를 100kb 영역에 맵핑된 서열 리드로 커버되는 모든 CpG 부위의 비율로 결정하여, 인간 게놈에서 각각의 100kb bin의 메틸화 밀도를 결정하였다. 메틸화 밀도는 남성뿐만 아니라 여성 표본의 혈장 표본 및 대응 혈액 세포 DNA 간에 상당히 일치하였다.
- [0064] 도 4a 및 4b는 혈액 세포 DNA 및 혈장 DNA에서 대응 유전자좌의 메틸화 밀도 그래프를 나타낸다: (A) 비임신 성인 여성, (B) 성인 남성. 비임신 여성 표본에 대한 Pearson 상관관계 계수 및 R^2 값은 각각 0.963 및 0.927이었으며, 남성 표본에 대한 것은 각각 0.953 및 0.908이었다. 이들 데이터는 조혈 세포가 인간 혈장 중 DNA의 주요한 원천임을 나타낸 동종이계 조혈 줄기 세포 이식 수령자의 혈장 DNA 분자의 유전형 평가에 기반한 이전 발견과 일치한다(YW Zheng 등, 2012 Clin Chem; 58: 549-558).
- [0065] D. 메틸롬에 걸친 메틸화 수준
- [0066] 다음으로 모체 혈장 DNA, 모체 혈액 세포, 및 태반 조직의 DNA 메틸화 수준을 연구하여 메틸화 수준을 결정하였다. 수준을 반복 영역, 비반복 영역, 및 전체에 대해 결정하였다.
- [0067] 도 5a 및 5b는 임신부에서 수집된 표본 간 메틸화 CpG 부위의 백분율의 막대 그래프를 나타낸다: (A) 상염색체, (B) 염색체 X. 메틸화 CpG의 전체적 비율은 각각 제1 및 제3 석달 모체 혈장 표본에 대해 67.0% 및 68.2%였다. 비임신 개인에서 취득된 결과와는 달리, 이들 비율은 제1 석달 모체 혈액 세포 표본에서보다 낮았지만 CVS 및 만기 태반 조직 표본에서보다 높았다(표 100). 주목할 것은, 산후 모체 혈장 표본에 대한 메틸화 CpG의 백분율이 혈액 세포 데이터와 유사한 73.1%였다(표 100). 이들 경향은 모든 상염색체뿐만 아니라 염색체 X에 걸쳐 분포된 CpG에서 관찰되었으며, 인간 게놈의 반복 요소의 다중 클래스 및 비반복 영역 모두에 걸쳐 존재하였다.
- [0068] 태반에서의 반복 및 비-반복 요소는 모두 모체 혈액 세포에 비해 저메틸화된 것으로 나타났다. 이 결과는 태반이 말초 혈액 세포를 포함하여 다른 조직에 비해 저메틸화된다는 문헌에서의 발견과 일치하였다.
- [0069] 서열분석된 CpG 부위의 71% 내지 72%가 임신 여성, 비임신 여성 및 성인 남성의 혈액 세포 DNA에서 메틸화되었다(도 1의 표 100) 이들 데이터는 Y Li 등(2010 PLoS Biol; 8: e1000533)에서 보고된 혈액 단핵구의 CpG 부위의 68.4% 보고와 필적한다. 태반 조직의 저메틸화 성질에 대한 이전 보고와 일치하게, CpG 부위의 55% 및 59%가 각각 CVS 및 만기 태반 조직에서 메틸화되었다(표 100).
- [0070] 도 6은 모체 혈액, 태반 및 모체 혈장에 대한 인간 게놈의 상이한 반복 클래스의 메틸화 수준의 막대 그래프를 나타낸다. 반복 클래스는 UCSC 게놈 브라우저에 의해 정의된 바와 같다. 나타난 데이터는 제1 석달 표본의 것이

다. 태반 조직의 저메틸화 성질이 게놈의 특정 반복 클래스에서 주로 관찰됨을 제시한 이전 데이터와는 달리(B Novakovic 등, 2012 Placenta; 33: 959-970), 여기서는 태반이 실제로 혈액 세포에 비해 게놈 요소의 대부분의 클래스에서 저메틸화되었음을 나타낸다.

[0071] E. 메틸롬의 유사성

[0072] 구현예는 동일한 플랫폼을 이용하여 태반 조직, 혈액 세포 및 혈장의 메틸롬을 결정할 수 있다. 따라서 이들 생물학적 표본 유형의 직접 비교가 가능하였다. 남성 및 비임신 여성의 혈액 세포 및 혈장 메틸롬 간 뿐만 아니라 모체 혈액 세포 및 산후 모체 혈장 표본 간 고수준의 유사성은 조혈 세포가 인간 혈장에서 DNA의 주요 원천임을 추가로 확인시켰다(YW Zheng 등, 2012 Clin Chem; 58: 549-558).

[0073] 유사성은 게놈에서의 메틸화 CpG의 전체적 비율뿐만 아니라 혈액 세포 DNA 및 혈장 DNA의 대응 유전자좌 간 메틸화 밀도의 높은 연관 관점에서 모두 자명하다. 그러나 제1 석달 및 제3 석달 모체 혈장 표본 중 메틸화 CpG의 전체적 비율은 모체 혈액 세포 데이터 또는 산후 모체 혈장 표본과 비교 시 감소되었다. 임신 동안 감소된 메틸화 수준은 모체 혈장에 존재하는 태아 DNA 분자의 저메틸화 성질 때문이었다.

[0074] 산후 모체 혈장 표본의 메틸화 프로파일은 모체 혈액 세포의 메틸화 프로파일과 더 유사해지는 역전은 태아 DNA 분자가 모체 순환에서 제거되었음을 제시한다. 태아의 SNP 마커에 기반한 태아 DNA 농도의 계산은 실제로 농도가 산전 33.9%에서 산후 표본에서의 불과 4.5%로 변화하였음을 나타내었다.

[0075] F. 다른 용도

[0076] 구현예는 혈장 DNA의 MPS 분석을 통해 DNA 메틸롬을 성공적으로 조립하였다. 모체 혈장에서 태반 또는 태아 메틸롬을 결정하는 능력은 임신 연관 상태, 예컨대 자간전증, 자궁내 성장 제한, 조산 등에 연관된 이상 메틸화 프로파일을 결정하고, 검출하고 모니터링하기 위한 비침습적 방법을 제공한다. 예를 들어, 질환-특이적 이상 메틸화 특징의 검출은 이러한 임신 연관 상태의 스크리닝, 진단 및 모니터링을 허용한다. 모체 혈장 메틸화 수준의 측정은 이러한 임신 연관 상태의 스크리닝, 진단 및 모니터링을 허용한다. 임신 연관 상태의 조사에 대한 직접 적용 이외에도, 이러한 접근은 혈장 DNA 분석이 관심의 대상인 다른 의약 분야에 적용될 수 있다. 예를 들어, 암의 메틸롬이 암 환자의 혈장 DNA에서 결정될 수 있다. 본원에 기재된 바와 같이, 혈장으로부터의 암 메틸롬 분석은 혈장으로부터의 암 게놈 분석에 대해 잠재적으로 상승 효과를 갖는 기술이다(KCA Chan 등, 2013 Clin Chem; 59: 211-224 및 RJ Leary 등, 2012 Sci Transl Med; 4:162ra154).

[0077] 예를 들어, 혈장 표본의 메틸화 수준 결정을 암을 스크리닝하기 위해 이용할 수 있다. 혈장 표본의 메틸화 수준이 건강한 대조군에 비해 이상 수준을 나타내는 경우, 암을 추정할 수 있다. 이어서 암의 유형 또는 암의 조직 기원의 추가 확인 및 평가가 상이한 게놈 유전자좌에서 메틸화의 혈장 프로파일을 결정하여 또는 종양 연관 복제 수 이상, 염색체 전위 및 단일 뉴클레오타이드 변이를 검출하기 위한 혈장 게놈 분석에 의해 수행될 수 있다. 실제로 본 발명의 하나의 구현예에서, 혈장 암 메틸 및 게놈 프로파일 확인이 동시에 수행될 수 있다. 대안적으로, 방사선 및 조영 조사(예로 전산화 단층촬영, 자기 공명 조영, 양성자 방출 단층촬영) 또는 내시경(예로 상부 위장관 시험 내시경 또는 대장내시경 검사)이 혈장 메틸화 수준 분석에 기반하여 암을 갖는 것으로 추정되는 개인을 추가 조사하기 위해 이용될 수 있다.

[0078] 암 스크리닝 또는 검출을 위해, 혈장(또는 다른 생물학적) 표본의 메틸화 수준의 결정은 암 스크리닝 또는 검출을 위한 다른 양식, 예컨대 전립선 특이적 항원 측정(예로 전립선암을 위한), 암배아 항원(예로 결직장암종, 위암종, 췌장암종, 폐암종, 유방암종, 수질성 갑상샘암종을 위한), 알파 태아단백질(예로 간암 또는 생식 세포 종양을 위한), CA125(예로 난소암 및 유방암을 위한) 및 CA19-9(예로 췌장암종을 위한)와 함께 이용될 수 있다.

[0079] 또한, 다른 조직을 서열분석하여 세포 메틸롬을 수득할 수 있다. 예를 들어, 간 조직을 분석하여 간에 특이적인 메틸화 패턴을 결정하고, 이를 간 병리를 확인하기 위해 이용할 수 있다. 또한 분석할 수 있는 다른 조직에는 뇌 세포, 뼈, 폐, 심장, 근육 및 신장 등이 포함된다. 다양한 조직의 메틸화 프로파일은, 예로 발생, 노화, 질환 절차(예로 염증 또는 경화 또는 자가면역 절차(예컨대 진신 홍반 루푸스에서)) 또는 치료(예로 탈메틸화제, 예컨대 5-아자시티딘 및 5-아자데옥시시티딘을 이용한 치료)의 결과 때때로 변할 수 있다. DNA 메틸화의 동적 성질은 이러한 분석을 생리학적 및 병리학적 절차의 모니터링을 위해 잠재적으로 매우 중요하게 만든다. 예를 들어, 건강할 때 수득한 베이스라인값에 비해 개인의 혈장 메틸롬 변화를 검출하는 경우, 혈장 DNA에 기여하는 기관에서 질환 절차를 검출할 수 있다.

[0080] 또한, 이식 기관의 메틸롬은 기관 이식 수령자의 혈장 DNA에서 결정될 수 있다. 본 발명에 기재된 바와 같이, 혈장으로부터의 이식 메틸롬 분석은 혈장으로부터의 이식 게놈 분석에 대한 잠재적으로 상승효과를 갖는 기술이

다(YW Zheng 등, 2012; YMD Lo 등, 1998 Lancet; 351: 1329-1330; 및 TM Snyder 등, 2011 Proc Natl Acad Sci USA; 108: 6229-6234). 혈장 DNA는 일반적으로 세포사의 마커로 간주되므로, 이식 기관에서 방출된 DNA의 혈장 수준 증가는 그 기관으로부터의 증가된 세포사, 예컨대 거부 증례 또는 그 기관이 수반되는 다른 병리적 절차(예로 감염 또는 농양)에 대한 마커로 이용될 수 있다. 항-거부 치료법이 성공적으로 시행되는 경우, 이식 기관에서 방출되는 DNA의 혈장 수준이 감소할 것으로 예측될 것이다.

[0081] III. SNP를 이용한 태아 또는 종양 메틸롬의 결정

[0082] 전술된 바와 같이, 혈장 메틸롬은 비임신 정상 개인에 대한 혈액 메틸롬에 대응한다. 그러나 임신 여성에 있어서는 메틸롬이 상이하다. 태아 DNA 분자는 모체 DNA의 대부분의 배경 가운데 모체 혈장에서 순환한다(YMD Lo 등, 1998 Am J Hum Genet; 62: 768-775). 따라서 임신 여성에 있어서, 혈장 메틸롬은 크게는 태반 메틸롬 및 혈액 메틸롬의 복합물이다. 따라서 혈장에서 태반 메틸롬을 추출할 수 있다.

[0083] 하나의 구현예에서, 모체 및 태아 간 단일 뉴클레오티드 다형성(SNP) 차이가 모체 혈장 중에서 태아 DNA 분자를 확인하기 위해 이용된다. 목적은 모체가 동종접합성이지만 태아가 이종접합성인 SNP 유전자좌를 확인하는 것이었다; 태아-특이적 대립유전자를 이용하여 어느 DNA 단편이 태아에서 유래되는지를 결정할 수 있다. 모체 혈액 세포로부터의 게놈 DNA는 SNP 유전형 분석 어레이, Illumina HumanOmni2.5-8을 이용하여 분석될 수 있다. 반면, 모체가 이종접합성이고 태아가 동종접합성인 SNP 유전자좌에 있어서, 모체에 대해 특이적인 SNP 대립유전자를 이용하여 어느 혈장 DNA 단편이 모체에서 유래되는지를 결정할 수 있다. 이러한 DNA 단편의 메틸화 수준은 모체에서 관련 게놈 영역에 대한 메틸화 수준을 반영할 것이다.

[0084] A. 태아-특이적 측정치의 메틸화 및 태반 메틸롬의 상관관계

[0085] 하나의 대립유전자(B)의 양이 다른 대립유전자(A)보다 훨씬 더 적은 두 상이한 대립유전자를 갖는 유전자좌가 생물학적 표본의 서열분석 결과에서 확인되었다. B 대립유전자를 커버하는 측정치는 태아-특이적으로(태아-특이적 측정치로) 간주되었다. 모체는 A에 대해 동종접합성인 것으로 그리고 태아는 A/B에 대해 이종접합성인 것으로 결정되며, 이에 따라 A 대립유전자를 커버하는 측정치는 모체 및 태아에 의해 공유된다(공유 측정치).

[0086] 본 발명의 몇몇 개념을 예시하기 위해 이용된 하나의 분석된 임신 케이스에서, 임신 모체는 상염색체 상 1,945,516 유전자좌에서 동종접합성으로 나타났다. 이들 SNP를 커버하는 모체 혈장 DNA 서열분석 측정치가 조사되었다. 비-모체 대립유전자를 수반하는 측정치가 107,750 유전자좌에서 검출되었고, 이들은 정보를 제공하는 유전자좌로 간주되었다. 각각의 정보를 제공하는 SNP에서, 모체에서 유래되지 않은 대립유전자가 태아-특이적 대립유전자로 명명된 반면, 다른 것은 공유 대립유전자로 명명되었다.

[0087] 모체 혈장 중 분획 태아/종양 DNA 농도(또한 태아 DNA 백분율로 불림)가 결정될 수 있다. 하나의 구현예에서, 모체 혈장 중 분획 태아 DNA 농도, f 는 다음 공식으로 결정된다:

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

[0088]

[0089] 식 중, p 는 태아-특이적 대립유전자의 서열분석 측정치의 수이며, q 는 모체 및 태아 간 공유 대립유전자의 서열분석 측정치의 수이다(YMD Lo 등, 2010 Sci Transl Med; 2: 61ra91). 제1 석달, 제3 석달 및 산후 모체 혈장 표본 중 태아 DNA 비율은 각각 14.4%, 33.9% 및 4.5%로 나타났다. 태아 DNA 비율은 또한 염색체 Y에 지정된 측정치의 수를 이용해서 계산되었다. 염색체 Y 데이터에 기반하여, 제1 석달, 제3 석달 및 산후 모체 혈장 표본에서의 결과는 각각 14.2%, 34.9% 및 3.7%였다.

[0090] 태아-특이적 또는 공유 서열 리드를 별도로 분석함으로써, 구현예는 순환 태아 DNA 분자가 배경 DNA 분자보다 훨씬 더 저메틸화되었음을 나타낸다. 제1 및 제3 석달 모두에 대한 태반 조직 데이터 및 태아-특이적 모체 혈장 측정치에서 대응 유전자좌의 메틸화 밀도 비교는 높은 수준의 상관관계를 드러내었다. 이들 데이터는 태반이 모체 혈장에서 태아 유래 DNA 분자의 주요 원천이며 선택된 유전자좌에서 유래된 정보에 기반한 이전 증거에 비해 주요 발전 단계를 나타내는 게놈 수준 증거를 제공하였다.

[0091] 정보를 제공하는 SNP에 인접한 CpG 부위를 커버하는 태아-특이적 또는 공유 측정치를 이용하여 게놈에서 각각의 1Mb 영역의 메틸화 밀도를 결정하였다. 모체 혈장 서열 리드에서 조립된 태아 및 비-태아-특이적 메틸롬은, 예를 들어 Circos 그래프에 디스플레이될 수 있다(M Krzywinski 등, 2009 Genome Res; 19: 1639-1645). 1Mb bin

별 메틸화 밀도도 모체 혈액 세포 및 태반 조직 표본에 대해 결정되었다.

- [0092] 도 7a는 제1 석달 표본에 대한 Circos 그래프 700을 나타낸다. 도 7b는 제3 석달 표본에 대한 Circos 그래프 750을 나타낸다. 그래프 700 및 750은 1Mb bin 별 메틸화 밀도를 나타낸다. 염색체 핵형도(최외부 고리)는 시계 방향으로 배향된다(중심절은 빨간색으로 나타냄). 두 번째 최외부 궤도는 대응 1Mb 영역에서 CpG 부위의 수를 나타낸다. 나타난 빨간색 막대 스케일은 1Mb bin 별로 최대 20,000 부위이다. 대응 1Mb 영역의 메틸화 밀도를 중심에 나타난 컬러 방식에 기반한 다른 궤도에 나타낸다.
- [0093] 제1 석달 표본(도 7a)에 대해, 내부에서 외부로 궤도는 다음과 같다: 융모막 융모 표본, 모체 혈장 중 태아-특이적 측정치, 모체 혈장 중 모체-특이적 측정치, 모체 혈장 중 조합된 태아 및 비-태아 측정치, 및 모체 혈액 세포. 제3 석달 표본(도 7b)에 대해, 궤도는 다음과 같다: 만기 태반 조직, 모체 혈장 중 태아-특이적 측정치, 모체 혈장 중 모체-특이적 측정치, 모체 혈장 중 조합된 태아 및 비-태아 측정치, 산후 모체 혈장 및 모체 혈액 세포(제1 석달 혈액 표본 유래). 제1 및 제3 석달 혈장 표본 모두에 대해, 태아 메틸롬이 비-태아-특이적 메틸롬에 비해 더 저메틸화되었음이 이해될 수 있다.
- [0094] 태아 메틸롬의 전체적 메틸화 프로파일은 CVS 또는 태반 조직 표본과 보다 더 유사하였다. 반대로, 주로 모체 DNA 였던 혈장 중 공유 측정치의 DNA 메틸화 프로파일은 모체 혈액 세포와 보다 더 유사하였다. 이어서 모체 혈장 DNA 측정치 및 모체 또는 태아 조직의 메틸화 밀도의 체계적 유전자좌별 비교를 수행하였다. 적어도 5 모체 혈장 DNA 서열 리드로 커버되며 정보를 제공하는 SNP와 동일한 서열 리드 상에 존재하는 CpG 부위의 메틸화 밀도를 결정하였다.
- [0095] 도 8a-8d는 정보를 제공하는 단일 뉴클레오타이드 다형성 주위의 CpG 부위에 대한 모체 혈장 DNA 대비 게놈 조직 DNA의 메틸화 밀도의 비교 그래프를 나타낸다. 도 8a는 CVS 표본에서의 측정치에 대한 메틸화 밀도 대비 제1 석달 모체 혈장 표본에서의 태아-특이적 측정치에 대한 메틸화 밀도를 나타낸다. 알 수 있듯이, 태아-특이적 값은 CVS값에 잘 대응한다.
- [0096] 도 8b는 만기 태반 조직에서의 측정치에 대한 메틸화 밀도 대비 제3 석달 모체 혈장 표본에서의 태아-특이적 측정치에 대한 메틸화 밀도를 나타낸다. 또 다시, 밀도 세트가 잘 대응하여, 태아 메틸화 프로파일이 태아-특이적 대립유전자의 측정치를 분석하여 수득될 수 있음을 시사한다.
- [0097] 도 8c는 모체 혈액 세포에서의 측정치에 대한 메틸화 밀도 대비 제1 석달 모체 혈장 표본에서의 공유 측정치에 대한 메틸화 밀도를 나타낸다. 대부분의 공유 측정치가 모체에서 유래되었다는 점에서, 두 세트값은 잘 대응한다. 도 8d는 모체 혈액 세포에서의 측정치에 대한 메틸화 밀도 대비 제3 석달 모체 혈장 표본에서의 공유 측정치에 대한 메틸화 밀도를 나타낸다.
- [0098] 모체 혈장에서의 태아-특이적 측정치에 있어서, 상기 제1 석달 모체 혈장 및 CVS 간 Spearman 상관관계 계수는 0.705였고($P < 2.2 \times 10^{-16}$); 제3 석달 모체 혈장 및 만기 태반 조직 간에는 0.796이었다($P < 2.2 \times 10^{-16}$)(도 8a 및 8b). 유사한 비교를 모체 혈액 세포 데이터로 모체 혈장에서의 공유 측정치에 대해 수행하였다. Pearson 상관관계 계수는 제1 석달 혈장 표본에 대해 0.653이었고($P < 2.2 \times 10^{-16}$), 제3 석달 혈장 표본에 대해서는 0.638이었다($P < 2.2 \times 10^{-16}$)(도 8c 및 8d).
- [0099] **B. 태아 메틸롬**
- [0100] 하나의 구현예에서, 모체 혈장으로부터 태아 메틸롬을 조립하기 위해, 적어도 하나의 정보를 제공하는 태아 SNP 부위에 걸치고 동일 측정치 내에서 적어도 하나의 CpG 부위를 함유한 서열 리드에 대해 정렬하였다. 태아-특이적 대립유전자를 나타낸 측정치는 태아 메틸롬의 조립에 포함되었다. 공유 대립유전자, 즉 비-태아-특이적 대립유전자를 나타낸 측정치는 주로 모체 유래 DNA 분자로 이루어진 비-태아-특이적 메틸롬의 조립에 포함되었다.
- [0101] 태아-특이적 측정치는 제1 석달 모체 혈장 표본에 대해 상염색체 상의 218,010 CpG 부위를 커버하였다. 제3 석달 및 산후 모체 혈장 표본에 대한 대응 수치는 각각 263,611 및 74,020이었다. 평균적으로, 공유 측정치는 각각 평균 33.3, 21.7 및 26.3배의 이들 CpG 부위를 커버하였다. 태아-특이적 측정치는 제1 석달, 제3 석달 및 산후 모체 혈장 표본에 대해 각각 3.0, 4.4 및 1.8배로 이들 CpG 부위를 커버하였다.
- [0102] 태아 DNA는 모체 혈장에서 소량 모집단을 나타내며, 이에 따라 태아-특이적 측정치에 의한 이들 CpG 부위의 커버 범위는 표본의 태아 DNA 백분율에 비례하였다. 제1 석달 모체 혈장 표본에 있어서, 태아 측정치 중 메틸화 CpG의 전체적 백분율은 47.0%였던 반면, 공유 측정치에 대한 백분율은 68.1%였다. 제3 석달 모체 혈장 표본에 있어서, 태아 측정치의 메틸화 CpG의 백분율은 53.3%였던 반면, 공유 측정치에 대한 백분율은 68.8%였다. 이들

데이터는 모체 혈장 중 태아-특이적 측정치가 모체 혈장 중 공유 측정치에 비해 더 저메틸화됨을 나타내었다.

[0103] C. 방법

[0104] 상술된 기법은 또한 종양 메틸화 프로필을 결정하기 위해 이용될 수 있다. 이제 태아 및 종양 메틸화 프로필을 결정하기 위한 방법을 설명한다.

[0105] 도 9는 본 발명의 구현예에 따른 유기체의 생물학적 표본에서 제1 메틸화 프로필을 결정하기 위한 방법 900을 예시하는 순서도이다. 방법 900은 모체 혈장의 메틸화 프로필에서 태아의 후생적 맵을 정립할 수 있다. 생물학적 표본에는 제1 조직 및 제2 조직에서 유래되는 무세포 DNA의 혼합물을 포함하는 무세포 DNA가 포함된다. 예로서, 제1 조직은 태아, 종양, 또는 이식 기관 유래일 수 있다.

[0106] 블록 910에서, 복수의 DNA 분자가 생물학적 표본에서 분석된다. DNA 분자의 분석에는 유기체의 게놈에서 상기 DNA 분자의 위치를 결정하고, 상기 DNA 분자의 유전형형을 결정하고, 상기 DNA 분자가 하나 이상의 부위에서 메틸화되어 있는지의 여부를 결정하는 것이 포함될 수 있다.

[0107] 하나의 구현예에서, DNA 분자는 DNA 분자의 서열 리드를 이용하여 분석되며, 여기서 서열분석은 메틸화를 인식한다. 따라서 서열 리드에는 생물학적 표본 유래 DNA 분자의 메틸화 상태가 포함된다. 메틸화 상태에는 특정 시토신 잔기가 5- 메틸시토신 또는 5-히드록시메틸시토신인지가 포함될 수 있다. 서열 리드는 다양한 서열분석 기법, PCR-기법, 어레이, 및 단편의 서열을 확인하기 위한 다른 적합한 기법에서 획득될 수 있다. 서열 리드의 부위의 메틸화 상태는 본원에 기재된 바와 같이 획득될 수 있다.

[0108] 블록 920에서, 제1 조직의 제1 게놈이 각각의 제1 대립유전자 및 각각의 제2 대립유전자에 대해 이중접합성이고, 제2 조직의 제2 게놈이 각각의 제1 대립유전자에 대해 동종접합성인 복수의 제1 유전자좌가 확인된다. 예를 들어, 태아-특이적 측정치가 복수의 제1 유전자좌에서 확인될 수 있다. 또는, 종양-특이적 측정치가 복수의 제1 유전자좌에서 확인될 수 있다. 조직-특이적 측정치는 제2 대립유전자의 서열 리드 백분율이 특정 범위, 예로 약 3%-25% 내에 속하여 유전자좌에서 이중접합성 게놈 유래의 소량 모집단의 DNA 단편 및 유전자좌에서 동종접합성 게놈 유래의 다량 모집단을 시사하는 서열분석 측정치에서 확인될 수 있다.

[0109] 블록 930에서, 각각의 제1 유전자좌의 하나 이상의 부위에 위치한 DNA 분자가 분석된다. 부위에서 메틸화되고 유전자좌의 각각의 제2 대립유전자에 대응하는 여러 DNA 분자가 결정된다. 유전자좌 당 둘 이상의 부위가 존재할 수 있다. 예를 들어, SNP는 단편이 태아-특이적임을, 그리고 단편이 메틸화 상태가 결정된 여러 부위를 가질 수 있음을 시사할 수 있다. 메틸화된 각 부위에서 측정치의 수가 결정될 수 있고, 유전자좌에 대한 메틸화 측정치의 총 수가 결정될 수 있다.

[0110] 유전자좌는 조직-특이적 대립유전자를 포함하는 변동 주위 영역에 대한 특정 수의 부위, 특정 세트의 부위 또는 특정 크기로 정의될 수 있다. 유전자좌는 단 하나의 부위를 가질 수 있다. 부위는 특이적 특성을 가질 수 있다, 예로 CpG일 수 있다. 비메틸화된 여러 측정치의 결정은 동등하며, 메틸화 상태의 결정 내에 포함된다.

[0111] 블록 940에서, 각각의 제1 유전자좌에 대해 메틸화 밀도는 유전자좌의 하나 이상의 부위에서 및 유전자좌의 각각의 제2 대립유전자에 대응하여 메틸화된 DNA 분자의 수에 기반하여 계산된다. 예를 들어, 메틸화 밀도는 유전자좌에 대응하는 CpG 부위에 대해 결정될 수 있다.

[0112] 블록 950에서, 제1 조직의 제1 메틸화 프로필은 제1 유전자좌에 대한 메틸화 밀도에서 생성된다. 제1 메틸화 프로필은 특정 부위, 예로 CpG 부위에 대응할 수 있다. 메틸화 프로필은 태아-특이적 대립유전자를 갖는 모든 유전자좌 또는 이들 유전자좌의 단지 일부에 대한 것일 수 있다.

[0113] IV. 혈장 및 혈액 메틸롬의 차이 이용

[0114] 상기에서, 혈장의 태아-특이적 측정치가 태반 메틸롬과 연관됨이 나타났다. 모체 혈장 메틸롬의 모체 성분은 주로 혈액 세포에 의해 기여되므로, 혈장 메틸롬 및 혈액 메틸롬 간 차이가 태아-특이적 대립유전자의 위치뿐만 아니라 모든 유전자좌에 대한 태반 메틸롬을 결정하기 위해 이용될 수 있다. 혈장 메틸롬 및 혈액 메틸롬 간 차이는 또한 종양 메틸롬을 결정하기 위해 이용될 수 있다.

[0115] A. 방법

[0116] 도 10은 본 발명의 구현예에 따른 유기체의 생물학적 표본으로부터 제1 메틸화 프로필을 결정하는 방법 1000을 예시하는 순서도이다. 생물학적 표본(예로, 혈장)에는 제1 조직에서 그리고 상기 제2 조직에서 기원하는 무세포 DNA 혼합물을 포함하는 무세포 DNA가 포함된다. 제1 메틸화 프로필은 제1 조직(예로, 태아 조직 또는 종양

조직)의 메틸화 프로필에 대응한다. 방법 1200은 모체 혈장으로부터 상이하게 메틸화된 영역의 추정을 제공할 수 있다.

- [0117] 블록 1010에서, 상기 생물학적 표본이 수신된다. 생물학적 표본은 기계(예로, 서열분석기)에서 단순히 수신될 수 있다. 생물학적 표본은 유기체에서 얻은 형태일 수도 있고, 또는 가공된 형태일 수 있다, 예로 표본은 혈액 표본에서 추출된 혈장일 수 있다.
- [0118] 블록 1020에서, 상기 제2 조직의 DNA에 대응하는 제2 메틸화 프로필이 획득된다. 제2 메틸화 프로필은 이전에 결정되었을 수 있으므로, 메모리에서 읽혀질 수 있다. 제2 메틸화 프로필은 제2 조직, 예로 제2 조직의 세포만을 또는 이를 주로 함유하는 상이한 표본에서 결정될 수 있다. 제2 메틸화 프로필은 세포 메틸화 프로필에 대응할 수 있고, 세포 DNA에서 획득될 수 있다. 다른 예로서, 암이 없는 비임신 개인의 혈장 메틸롬이 혈액 세포의 메틸롬과 매우 유사하므로, 제2 프로필은 임신 전에 또는 암 발생 전에 수집된 혈장 표본에서 결정될 수 있다.
- [0119] 제2 메틸화 프로필은 유기체의 게놈에서 각각의 복수의 유전자좌에서 메틸화 밀도를 제공할 수 있다. 특정 유전자좌에서의 메틸화 밀도는 메틸화된 상기 제2 조직의 DNA의 비율에 대응한다. 하나의 구현예에서, 메틸화 밀도는 CpG 메틸화 밀도이며, 여기서 유전자좌와 연관된 CpG 부위는 메틸화 밀도를 결정하기 위해 이용된다. 유전자좌에 대해 하나의 부위가 존재하는 경우, 메틸화 밀도는 메틸화 지수와 같을 수 있다. 메틸화 밀도는 두 값이 상보적이므로, 비메틸화 밀도에도 대응한다.
- [0120] 하나의 구현예에서, 제2 메틸화 프로필은 유기체의 표본에서 세포 DNA의 메틸화 인식 서열분석(methylation-aware sequencing)을 수행하여 획득된다. 메틸화-인식 서열분석의 한 예에는 나트륨 바이설파이트를 이용한 DNA 처리 및 DNA 서열분석의 수행이 포함된다. 다른 예에서, 메틸화-인식 서열분석은 나트륨 바이설파이트를 이용하지 않고, 상기 DNA 분자(N6-메틸아데닌, 5-메틸시토신 및 5-히드록시메틸시토신 포함)의 메틸화 상태가 바이설파이트 전환 없이 직접 규명될 수 있도록 하는 단일 분자 서열분석 플랫폼을 이용하여(AB Flusberg 등, 2010 Nat Methods; 7: 461-465; J Shim 등, 2013 Sci Rep; 3:1389. doi: 10.1038/srep01389); 또는 메틸화 시토신의 면역침전을 통해(예로 메틸시토신에 대한 항체를 이용하거나 메틸화 DNA 결합 단백질 또는 펩티드를 이용한 후(LG Acevedo 등, 2011 Epigenomes; 3: 93-101) 서열분석을 통해); 또는 메틸화-감수성 제한 효소의 이용 후 서열분석을 통해 수행될 수 있다. 다른 구현예에서, 비-서열분석 기법, 예컨대 어레이, 디지털 PCR 및 질량 분광측정이 이용된다.
- [0121] 다른 구현예에서, 제2 조직의 제2 메틸화 밀도는 대상자의 대조군 표본에서 또는 다른 대상자에서 이전에 획득될 수 있다. 다른 대상자로부터의 메틸화 밀도는 기준 메틸화 밀도를 갖는 기준 메틸화 프로필로 작용할 수 있다. 기준 메틸화 밀도는 여러 표본에서 결정될 수 있고, 여기서 유전자좌에서 상이한 메틸화 밀도의 평균 수준(또는 다른 통계값)이 유전자좌에서 기준 메틸화 밀도로 이용될 수 있다.
- [0122] 블록 1030에서, 무세포 메틸화 프로필이 혼합물의 무세포 DNA에서 결정된다. 무세포 메틸화 프로필은 각각의 복수의 유전자좌에서 메틸화 밀도를 제공한다. 무세포 메틸화 프로필은 무세포 DNA의 서열분석에서 서열 리드를 수신하여 결정될 수 있고, 여기서 메틸화 정보가 서열 리드와 함께 획득된다. 무세포 메틸화 프로필은 세포 메틸롬과 동일한 방식으로 결정될 수 있다.
- [0123] 블록 1040에서, 상기 생물학적 표본 중 제1 조직으로부터의 무세포 DNA의 백분율이 결정된다. 하나의 구현예에서, 제1 조직은 태아 조직이고, 대응 DNA는 태아 DNA이다. 다른 구현예에서, 제1 조직은 종양 조직이고, 대응 DNA는 종양 DNA이다. 백분율은 다양한 방식, 예로 태아-특이적 대립유전자 또는 종양-특이적 대립유전자를 이용하여 결정될 수 있다. 복제수는 또한 참조로 도입된, 2013. 3. 13.에 출원된 "Mutational Analysis Of Plasma DNA For Cancer Detection"을 표제로 하는 U.S. 특허 출원 13/801,748에 기재된 바와 같이 백분율을 결정하기 위해 이용될 수 있다.
- [0124] 블록 1050에서, 제1 메틸롬을 결정하기 위한 복수의 유전자좌가 확인된다. 이들 유전자좌는 무세포 메틸화 프로필 및 제2 메틸화 프로필을 결정하기 위해 이용되는 각각의 유전자좌에 대응할 수 있다. 따라서 복수의 유전자좌가 대응할 수 있다. 더 많은 유전자좌가 무세포 메틸화 프로필 및 제2 메틸화 프로필을 결정하기 위해 이용될 수 있다.
- [0125] 일부 구현예에서, 제2 메틸화 프로필에서 고메틸화 또는 저메틸화된 유전자좌가, 예로 모체 혈액 세포를 이용해서 확인될 수 있다. 모체 혈액 세포에서 고메틸화된 유전자좌를 확인하기 위해, 메틸화 지수 $\geq X\%$ (예로, 80%)인 CpG 부위에 대해 염색체의 한 말단부터 스캐닝할 수 있다. 이어서 하류 영역 내에서(예로, 200-bp 하류 내에서) 다음 CpG 부위에 대해 검색할 수 있다. 바로 하류 CpG 부위도 메틸화 지수 $\geq X\%$ (또는 다른 특정된 양)를 가진

경우, 제1 및 제2 CpG 부위가 그룹핑될 수 있다. 그룹핑은 다음 하류 영역 내에 다른 CpG 부위가 없을 때까지; 또는 바로 하류 CpG 부위가 메틸화 지수 <X%를 가질 때까지 계속될 수 있다. 그룹핑된 CpG 부위의 영역은 그 영역이 적어도 5개의 바로 인접한 고메틸화 CpG 부위를 함유하는 경우, 모체 혈액 세포에서 고메틸화된 것으로 보고될 수 있다. 메틸화 지수 ≤20%인 CpG 부위에 대해 모체 혈액 세포 중에서 저메틸화된 유전자좌에 대해 검색하기 위해 유사한 분석이 수행될 수 있다. 제2 메틸화 프로필에 대한 메틸화 밀도는 짧게 기재된 유전자좌에 대해 계산되고, 예로 모체 혈장 바이셀파이트-서열분석 데이터로부터 대응 유전자좌의 제1 메틸화 프로필(예로, 태반 조직 메틸화 밀도)를 추정하기 위해 이용될 수 있다.

[0126] 블록 1060에서, 제1 조직의 제1 메틸화 프로필은 상기 제2 메틸화 프로필의 메틸화 밀도 및 각각의 복수의 유전자좌에 대한 무세포 메틸화 프로필의 메틸화 밀도 간 차이를 포함하는 차등적 파라미터를 계산하여 결정된다. 차이는 백분율로 등급화된다.

[0127] 하나의 구현예에서, 제1(예로, 태반) 조직에서 유전자좌의 제1 메틸화 밀도(D)는 다음 공식을 이용하여 추정되었다:

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

[0128]

[0129] 식 중, mbc 는 유전자좌(예로, 모체 혈액 세포 바이셀파이트-서열분석 데이터에서 결정된 짧게 기재된 유전자좌)에서 상기 제2 메틸화 프로필의 메틸화 밀도를 나타내며; mp 는 모체 혈장 바이셀파이트-서열분석 데이터에서 대응 유전자좌의 메틸화 밀도를 나타내고; f 는 제1 조직(예로, 분획 태아 DNA 농도)으로부터 무세포 DNA의 백분율을 나타내고, CN은 유전자좌에서의 복제수(예로, 정상 대비 증폭에 대해 더 높은 값 또는 결실에 대해 더 낮은 값)를 나타낸다. 제1 조직에 증폭 또는 결실이 없는 경우, CN은 1일 수 있다. 삼염색체증(또는 종양 또는 태아에서 영역의 중복)에 있어서, CN은 1.5일 것이고(증가가 2개 카피 내지 3개 카피이므로), 일염색체증은 0.5를 가질 것이다. 더 높은 증폭은 0.5 증분씩 증가할 수 있다. 상기 예에서, D 는 차등적 파라미터에 대응할 수 있다.

[0130] 블록 1070에서, 제1 메틸화 밀도는 제1 조직의 보정된 제1 메틸화 밀도를 획득하기 위해 변환된다. 변환은 제1 조직의 차등적 파라미터 및 실제 메틸화 프로필 간의 고정 차이를 설명할 수 있다. 예를 들어, 값은 고정 상수만큼 또는 경사도만큼 다를 수 있다. 변환은 선형 또는 비선형일 수 있다.

[0131] 하나의 구현예에서, 추정값의 분포, D 는 태반 조직의 실제 메틸화 수준보다 낮은 것으로 나타났다. 예를 들어, 추정값은 CpG 부위의 상향 제시를 가진 계층 절편인 CpG 섬의 데이터를 이용해서 선형으로 변환될 수 있다. 상기 연구에서 이용된 CpG 섬의 계층 위치는 UCSC 계층 브라우저 데이터베이스(NCBI build 36/hg18)에서 획득되었다(PA Fujita 등, 2011 Nucleic Acids Res; 39: D876-882). 예를 들어, CpG 섬은 GC 함량 ≥50%, 계층 길이 >200bp 및 관찰/예측 CpG수의 비 >0.6인 계층 절편으로 정의될 수 있다(M Gardiner-Garden 등, 1987 J Mol Biol; 196: 261-282).

[0132] 하나의 시행에서, 선형 변환식을 유도하기 위해 서열분석된 표본에서 적어도 4 CpG 부위 및 CpG 부위 별 평균 측정 값이 ≥5를 갖는 CpG 섬이 포함될 수 있다. CVS 또는 만기 태반에서 CpG 섬의 메틸화 밀도 및 추정값, D 간의 선형 관계를 결정한 후, 하기 공식이 예측값을 결정하기 위해 이용될 수 있다:

[0133] 제1 식달 예측값 = $D \times 1.6 + 0.2$

[0134] 제3 식달 예측값 = $D \times 1.2 + 0.05$

[0135] B. 태아 예

[0136] 상술된 바와 같이, 방법 1000은 모체 혈장으로부터 태반의 메틸화 상황을 추정하기 위해 이용될 수 있다. 혈장 중 순환 DNA는 주로 조혈 세포에서 유래된다. 여전히 여기에는 다른 내부 기관에서 기여되는 미지 비율의 무세포 DNA가 존재한다. 또한, 태반-유래 무세포 DNA가 모체 혈장 중 전체 DNA의 대략 5-40%를 나타내며, 평균 대략 15%이다. 따라서 모체 혈장 중 메틸화 수준이 전술된 바와 같이 임신 동안 기존 배경 메틸화 + 태반 분포와 동일하다고 가정할 수 있다.

[0137] 모체 혈장 메틸화 수준, MP 는 하기 식을 이용해서 결정될 수 있다:

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

[0138]

[0139] 식 중, BKG 는 혈액 세포 및 내부 기관에서 유래된 혈장 중의 배경 DNA 메틸화 수준이며, PLN 은 태반의 메틸화 수준이고, f 는 모체 혈장 중 분획 태아 DNA 농도이다.

[0140] 하나의 구현예에서, 태반의 메틸화 수준은 이론적으로 다음과 같이 추정될 수 있다:

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

[0141]

[0142] 식 (1) 및 (2)는 CN 이 1이고, D 가 PLN 이고, BKG 가 mbc 인 경우 동일하다. 다른 구현예에서, 분획 태아 DNA 농도는, 예로 최소 f 가 존재한다는 가정의 일부로서, 특정값으로 가정되거나 또는 설정될 수 있다.

[0143] 모체 혈액의 메틸화 수준은 모체 혈장의 배경 메틸화를 나타내기 위해 얻을 수 있다. 모체 혈액 세포에서 고메틸화 또는 저메틸화된 유전자좌에 부가하여, 임상적 관련성을 갖는 정의된 영역, 예를 들어 인간 게놈에서의 CpG 섬에 초점을 맞춰 추정 접근을 추가 탐색하였다.

[0144] 상염색체 및 chrX에 대한 총 27,458개의 CpG 섬(NCBI Build36/hg18)의 평균 메틸화 밀도는 모체 혈장 및 태반의 서열분석 데이터에서 유래되었다. ≥ 10 CpG 부위를 가지고 태반, 모체 혈액 및 모체 혈장을 포함하는 모든 분석 표본에서 커버되는 CpG 부위 별 평균 측정 값이 ≥ 5 인 것들만 선택되었다. 그 결과, 26,698개의 CpG 섬(97.2%)이 유효하게 남았고, 이들의 메틸화 수준이 상기 식에 따라 혈장 메틸화 데이터 및 분획 태아 DNA 농도를 이용하여 추정되었다.

[0145] 추정된 PLN 값의 분포는 태반 조직에서 CpG 섬의 실제 메틸화 수준보다 낮음이 주지되었다. 따라서 하나의 구현예에서, 추정된 PLN 값, 또는 단순 추정값(D)이 태반에서 CpG 섬의 메틸화 수준을 추산하기 위한 임의의 단위로 이용되었다. 변환 후, 추정값은 선형으로 그리고 이들의 분포는 실제 데이터세트와 더 유사해졌다. 변환된 추정값은 메틸화 예측값(MPV)으로 명명되었고, 이어서 태반에서 유전적 유전자좌의 메틸화 수준을 추정하기 위해 이용되었다.

[0146] 상기 예에서, CpG 섬은 태반에서 이들의 메틸화 밀도에 기반하여 3 범주로 분류되었다: 저(≤ 0.4), 중간(>0.4 < 0.8) 및 고(≥ 0.8). 추정 공식을 이용하여, 동일 세트의 CpG 섬의 MPV를 계산한 뒤 값들을 이용해서 동일한 컷오프를 갖는 3 범주로 이들을 분류하였다. 실제 및 추정 데이터세트를 비교하여, 짧게 기재된 CpG 섬의 75.1%가 이들의 MPV에 따른 조직 데이터에서 동일한 범주와 정확히 매치될 수 있음을 확인하였다. CpG 섬의 약 22%는 1-수준 차이를 갖는 그룹(고 대 중간 또는 중간 대 저)으로 지정되며, 3% 미만만 완벽하게 오분류될 것이다(고 대 저) (도 12a). 전체적 분류 성능도 결정되었다: 태반에서 메틸화 밀도 ≤ 0.4 , >0.4 < 0.8 및 ≥ 0.8 을 갖는 CpG 섬의 86.1%, 31.4% 및 68.8%가 정확히 "저", "중간" 및 "고"로 추정되었다(도 12b).

[0147] 도 11a 및 11b는 본 발명의 구현예에 따른 모체 혈장 데이터 및 분획 태아 DNA 농도를 이용한 예측 알고리즘의 성능 그래프를 나타낸다. 도 11a는 MPV 수정 분류를 이용한 CpG 섬 분류의 정확성(추정 카테고리가 실제 데이터 세트와 정확히 매치된다); 1-수준 차이(추정 카테고리가 실제 데이터 세트와 1-수준 상이하다); 및 오분류(추정 분류가 실제 데이터 세트와 반대이다)를 나타내는 그래프 1100이다. 도 11b는 각각의 추정 카테고리에 분류된 CpG 섬의 비율을 나타내는 그래프 1150이다.

[0148] 모체 배경 메틸화가 각각의 게놈 영역에서 낮다면, 순환 중 고메틸화 태반-유래 DNA의 존재는 분획 태아 DNA 농도에 따른 정도로 전체적 혈장 메틸화 수준을 증가시킬 것이다. 배출된 태아 DNA가 완전 메틸화된 경우, 현저한 변화가 관찰될 수 있다. 반대로 모체 배경 메틸화가 높은 경우, 혈장 메틸화 수준의 변화 정도는 저메틸화 태아 DNA가 배출된 경우 더 커질 것이다. 따라서, 추정 방식은 메틸화 수준이 모체 배경 및 태반 간에 구별되는 것이 알려져 있는 유전적 유전자좌에 대해, 특히 태반에서 고메틸화 및 저메틸화 마커에 대해 추정되는 경우 더 실용적이 될 수 있다.

[0149] 도 12a는 본 발명의 구현예에 따른 메틸화 예측을 위한 15개의 선택된 게놈 유전자좌의 상세내용을 나타내는 표 1200이다. 기법을 확인하기 위해, 이전에 연구된 15개의 차등적으로 메틸화된 게놈 유전자좌를 선택하였다. 선

택된 영역의 메틸화 수준을 추정하고 이전에 연구된 15개의 차등적으로 메틸화된 유전적 유전자좌와 비교하였다 (RWK Chiu 등, 2007 Am J Pathol; 170: 941-950; S.S.C. Chim 등, 2008 Clin Chem; 54: 500-511; SSC Chim 등, 2005 Proc Natl Acad Sci U S A; 102: 14753-14758; DWY Tsui 등, 2010 PLoS One; 5: e15069).

[0150] 도 12b는 태반에서 15개의 선택된 계놈 유전자좌 및 이들의 대응 메틸화 수준의 추정 범주를 나타내는 그래프 1250이다. 추정 메틸화 범주는 다음과 같다: 저, ≤ 0.4 ; 중간, $>0.4 < 0.8$; 고, ≥ 0.8 . 표 1200 및 그래프 1300은 태반에서의 이들의 메틸화 수준이 몇몇 예외 RASSF1A, CGI009, CGI137 및 VAPA를 가지며 정확히 추정될 수 있었음을 나타낸다. 이들 4개 마커 중 CGI009만이 실제 데이터세트와 현저한 불일치를 나타내었다. 나머지는 약간만 오분류되었다.

[0151] 표 1200에서, "1"은 하기 식에 의해 계산된 추정값(D)을 나타낸다:

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

[0152]

[0153] 식 중, f 는 분획 태아 DNA 농도이다. 표시 "2"는 하기 식을 이용해서 선형으로 변환된 추정값에 대한 메틸화 예측값(MPV)을 나타낸다:

[0154] $MPV = D \times 1.6 + 0.25$. 표시 "3"은 추정값에 대한 분류 컷오프를 나타낸다: 저, ≤ 0.4 ; 중간, $>0.4 < 0.8$; 고, ≥ 0.8 . 표시 "4"는 실제 태반 데이터세트에 대한 분류 컷오프를 나타낸다: 저, ≤ 0.4 ; 중간, $>0.4 < 0.8$; 고, ≥ 0.8 . 표시 "5"는 태반 상태가 모체 혈액 세포에 대비한 태반의 메틸화 상태를 나타냄을 표시한다.

[0155] C. 태아 DNA의 분획 농도 계산

[0156] 하나의 구현예에서, 제1 조직의 태아 DNA의 백분율은 남성 태아에 대해 Y 염색체를 이용할 수 있다. 모체 혈장 표본 중 염색체 Y 서열의 비율(%chrY)은 남성 태아에서 유래된 염색체 Y 측정치 및 염색체 Y에 대해 잘못 정렬된 모체(여성) 측정치의 수의 합성 함수였다(RWK Chiu 등, 2011 BMJ; 342: e7401). 따라서 표본에서 %chrY 및 분획 태아 DNA 농도(f) 간 관계는 다음과 같이 주어질 수 있다:

$$\%chrY = \%chrY_{male} \times f + \%chrY_{female} \times (1 - f)$$

[0157]

[0158] 식 중, %chrY_{male}은 100% 남성 DNA를 함유하는 혈장 표본 중 염색체 Y에 정렬된 측정치의 비율을 나타내며; %chrY_{female}은 100% 여성 DNA를 함유하는 혈장 표본 중 염색체 Y에 정렬된 측정치의 비율을 나타낸다.

[0159] %chrY는 남성 태아를 갖는 여성 임산부의 표본에 대해 미스매치가 없는 염색체 Y에 정렬된 측정치로부터, 예로 측정치가 바이셀과이트-전환 표본에서 유래된 경우 결정될 수 있다. %chrY_{male}값은 두 성인 남성 혈장 표본의 바이셀과이트-서열분석에서 수득될 수 있다. %chrY_{female}값은 두 비임신 성인 여성 혈장 표본의 바이셀과이트-서열분석에서 수득될 수 있다.

[0160] 다른 구현예에서, 태아 DNA 백분율은 삼염색체 상의 태아-특이적 대립유전자에서 결정될 수 있다. 다른 예로서, 후생적 마커가 태아 DNA 백분율을 결정하기 위해 이용될 수 있다. 태아 DNA 백분율의 다른 결정 방식도 이용될 수 있다.

[0161] D. 복제수를 결정하기 위한 메틸화 이용 방법

[0162] 태반 계놈은 모체 계놈에 비해 더 저메틸화된다. 위에서 논의된 바와 같이, 임신 여성의 혈장 메틸화는 모체 혈장 중 태반 유래 태아 DNA의 분획 농도에 의존한다. 따라서 염색체 영역의 메틸화 밀도 분석을 통해, 모체 혈장에 대한 태아 조직의 기여에서의 차이를 검출할 수 있다. 예를 들어, 삼염색체 태아(예로 삼염색체중 21 또는 삼염색체중 18 또는 삼염색체중 13을 갖는)를 수반하는 임신 여성에 있어서, 태아는 이염색체와 비교 시 모체 혈장에 대해 삼염색체로부터 추가량의 DNA에 기여할 것이다. 상기 상황에서, 삼염색체(또는 증폭을 갖는 임의의 염색체 영역)에 대한 혈장 메틸화 밀도는 이염색체에 대한 것보다 낮을 것이다. 차이의 정도는 혈장 표본 중 분획 태아 DNA 농도를 고려하여 수학적 계산에 의해 예측될 수 있다. 혈장 표본 중 분획 태아 DNA 농도가 높을수록, 삼염색체 및 이염색체 간 메틸화 밀도 차이는 더 커질 것이다. 결실을 갖는 영역에 대해서는 메틸화 밀도가 더 높을 것이다.

[0163] 결실의 한 예는 여성 태아가 염색체 X의 한 개 사본만을 가지게 되는 터너 증후군이다. 상기 상황에서, 터너 증후군을 겪는 태아를 수반하는 임신 여성에 있어서, 그녀의 혈장 DNA 중 염색체 X의 메틸화 밀도는 정상 수의 염색체 X를 갖는 여성 태아를 수반하는 동일 임신 여성의 상황에서보다 높을 것이다. 상기 전략의 하나의 구현예에서, 염색체 Y 서열의 존재 또는 부재에 대해 먼저 모체 혈장을 분석할 수 있다(예로 MPS 또는 PCR-기반 기법을 이용해서). 염색체 Y 서열이 존재하는 경우, 태아는 남성으로 분류될 수 있고, 후속 분석이 필요하지 않을 것이다. 반면, 염색체 Y 서열이 모체 혈장 중에 부재하는 경우, 태아는 여성으로 분류될 수 있다. 상기 상황에서는, 모체 혈장 중 염색체 X의 메틸화 밀도를 분석할 수 있다. 정상보다 높은 염색체 X 메틸화 밀도는 태아가 터너 증후군을 가질 높은 위험이 있음을 시사할 것이다. 상기 접근은 다른 성 염색체 홀배수체에도 적용될 수 있다. 예를 들어, XYY를 겪는 태아에 있어서, 모체 혈장 중 Y 염색체에 대한 메틸화 밀도는 모체 혈장 중 유사한 수준의 태아 DNA를 갖는 정상 XY 태아에 비해 더 낮을 것이다. 다른 예로서, 클라인펠터 증후군(XXY)을 겪는 태아에 있어서, 염색체 Y 서열은 모체 혈장 중에 존재하지만, 모체 혈장 중 염색체 X의 메틸화 밀도는 모체 혈장 중 유사한 수준의 태아 DNA를 갖는 정상 XY 태아에 비해 더 낮을 것이다.

[0164] 이전 논의로부터, 이염색체에 대한 혈장 메틸화 밀도($MP_{Non-aneu}$)는 다음과 같이 계산될 수 있다:

$$MP_{Non-aneu} = BKG \times (1 - f) + PLN \times f$$

[0165]

[0166] 식 중, BKG 는 혈액 세포 및 내부 기관에서 유래된 혈장 중의 배경 DNA 메틸화 수준이며, PLN 은 태반의 메틸화 수준이고, f 는 모체 혈장 중 분획 태아 DNA 농도이다.

[0167] 삼염색체에 대한 혈장 메틸화 밀도(MP_{Aneu})는 다음과 같이 계산될 수 있다:

$$MP_{Aneu} = BKG \times (1 - f) + PLN \times f \times 1.5$$

[0168]

[0169] 식 중, 1.5는 복제수 CN에 대응하며, 한 개 염색체의 추가는 50% 증가이다. 삼염색체 및 이염색체 간 차이(MP_{Diff})는 다음과 같을 것이다:

$$MP_{Diff} = PLN \times f \times 0.5.$$

[0170]

[0171] 하나의 구현예에서, 잠재적 홀배수체 염색체(또는 염색체 영역)의 메틸화 밀도와 하나 이상의 다른 추정 비-홀배수체 염색체(들) 또는 계놈의 전체적 메틸화 밀도의 비교는 혈장 표본 중 태아 DNA 농도를 효과적으로 정규화하기 위해 이용될 수 있다. 비교는 정규화된 메틸화 밀도를 획득하기 위해 두 영역의 메틸화 밀도 간 파라미터(예로 비 또는 차이가 수반되는)의 계산을 통할 수 있다. 비교는 생성 메틸화 수준의 의존성을 제거할 수 있다(예로 두 메틸화 밀도에서 파라미터로 결정됨).

[0172] 잠재적 홀배수체 염색체의 메틸화 밀도가 하나 이상의 다른 염색체(들)의 메틸화 밀도 또는 태아 DNA의 분획 농도를 반영하는 다른 파라미터에 대해 정규화되지 않은 경우, 분획 농도는 혈장 중 메틸화 밀도에 영향을 미치는 주요 요인일 것이다. 예를 들어, 분획 태아 DNA 농도 10%로 삼염색체중 21 태아를 수반하는 임신 여성의 염색체 21의 혈장 메틸화 밀도는 정배수체 태아를 수반하며 분획 태아 DNA 농도가 15%인 임신 여성과 동일할 것인 반면, 정규화된 메틸화 밀도는 차이를 나타낼 것이다.

[0173] 다른 구현예에서, 잠재적 홀배수체 염색체의 메틸화 밀도는 분획 태아 DNA 농도에 대해 정규화될 수 있다. 예를 들어, 하기 식이 메틸화 밀도를 정규화하기 위해 적용될 수 있다:

$$MP_{Normalized} = MP_{non-normalized} + (BKG - PLN) \times f$$

[0174]

[0175] 식 중, $MP_{Normalized}$ 는 혈장 중 분획 태아 DNA 농도로 정규화된 메틸화 밀도이며, $MP_{non-normalized}$ 는 측정된 메틸화 밀도이고, BKG 는 모체 혈액 세포 또는 조직으로부터의 배경 메틸화 밀도이고, PLN 은 태반 조직 중 메틸화 밀도이고, f 는 분획 태아 DNA 농도이다. BKG 및 PLN 의 메틸화 밀도는 건강한 임신부에서 획득된 모체 혈액 세포 및 태반 조직에서 이전에 정립된 기준값에 기반할 수 있다. 상이한 유전적 및 후생적 방법이 혈장 표본 중 분획 태아 DNA 농도의 결정을 위해, 예를 들어 대량 병렬 서열분석 또는 비-바이선폴라이트-전환 DNA 상의 PCR을 이용하여

염색체 Y로부터 서열 리드 백분율의 측정에 의해 이용될 수 있다.

- [0176] 하나의 시행에서, 잠재 홀배수체 염색체에 대한 정규화된 메틸화 밀도는 정배수체 태아를 수반하는 임신 여성으로 구성된 기준 군과 비교될 수 있다. 기준 군의 정규화된 메틸화 밀도의 평균 및 SD가 결정될 수 있다. 이어서 평가 케이스의 정규화된 메틸화 밀도가 하기 식에 의해 기준 군의 평균으로부터 SD의 수를 시사하는 z-스코어로 표현될 수 있다:

$$z\text{-스코어} = \frac{MP_{\text{Normalized}} - \text{Mean}}{SD}$$

- [0178] 식 중, $MP_{\text{Normalized}}$ 는 평가 케이스에 대해 정규화된 메틸화 밀도이고, Mean 은 기준 케이스의 정규화된 메틸화 밀도의 평균이고 SD는 기준 케이스의 정규화된 메틸화 밀도의 표준 편차이다. 염색체가 크게 저메틸화된 경우 분류하기 위해, 그리고 이에 따라 표본의 홀배수체 상태를 결정하기 위해, 예를 들어 z-스코어 <-3의 컷오프가 이용될 수 있다.

- [0179] 다른 구현예에서, MP_{Diff} 가 정규화된 메틸화 밀도로 이용될 수 있다. 이러한 구현예에서, PLN은, 예로 방법 1000을 이용해서 추정될 수 있다. 일부 시행에서, 기준 메틸화 밀도(f를 이용해서 정규화될 수 있음)가 비-홀배수체 영역의 메틸화 수준에서 결정될 수 있다. 예를 들어, Mean은 동일한 표본의 하나 이상의 염색체 영역에서 결정될 수 있다. 컷오프는 f에 의해 스케일링되거나 최소 농도가 존재하는 한 충분한 수준으로 설정될 수 있다.

- [0180] 따라서, 영역의 메틸화 수준의 컷오프에 대한 비교는 다양한 방식으로 수행될 수 있다. 비교에는 정규화가 수반될 수 있고(예로, 전술된 바와 같이), 이는 값들이 정의되는 방식에 따라 메틸화 수준 또는 컷오프값에서 동등하게 수행될 수 있다. 따라서 영역의 결정된 메틸화 수준이 기준 수준과 통계적으로 상이한지 여부는(동일한 표본 또는 다른 표본에서 결정됨) 다양한 방식으로 결정될 수 있다.

- [0181] 상기 분석은 염색체 영역의 분석에 적용될 수 있고, 여기에는 염색체의 인접 또는 비인접 서브영역을 포함하는 염색체의 일부 또는 염색체의 전부가 포함될 수 있다. 하나의 구현예에서, 잠재적인 홀배수체 염색체는 여러 bin으로 구분될 수 있다. bin은 동일한 크기 또는 상이한 크기일 수 있다. 각각의 bin의 메틸화 밀도는 표본의 분해 농도 또는 하나 이상의 추정 비-홀배수체 염색체(들)의 메틸화 밀도 또는 계놈의 전체적 메틸화 밀도에 대해 정규화될 수 있다. 이어서 각각의 bin의 정규화된 메틸화 밀도는 이것이 유의미하게 저메틸화되었는지를 결정하기 위해 기준 군과 비교될 수 있다. 이어서 유의미하게 저메틸화된 bin의 백분율이 결정될 수 있다. 예를 들어 5%, 10%, 15%, 20% 또는 30% 초과와 bin이 유의미하게 저메틸화된 컷오프가 케이스의 홀배수체 상태를 분류하기 위해 이용될 수 있다.

- [0182] 증폭 또는 결실을 평가하는 경우, 메틸화 밀도를 평가되는 특정 영역에 특이적일 수 있는 기준 메틸화 밀도에 대해 비교할 수 있다. 각각의 영역은 메틸화가 특히 영역의 크기에 따라(예로 더 작은 영역이 더 변동을 보일 것임) 영역 별로 변할 수 있으므로, 상이한 기준 메틸화 밀도를 가질 수 있다.

- [0183] 상술된 바와 같이, 각각 정배수체 태아를 수반하는 하나 이상의 임신 여성이 두 염색체 영역 간 메틸화 밀도 차이 또는 관심 영역에 대한 메틸화 밀도의 정상 범위를 정의하기 위해 이용될 수 있다. 정상 범위는 또한 PLN에 대해(예로 직접 측정에 의해 또는 방법 1000으로 추정되는 바와 같이) 결정될 수 있다. 다른 구현예에서, 예로 잠재적으로 홀배수체 염색체의 두 메틸화 밀도 간의 비가 이용될 수 있고, 비-홀배수체 염색체가 이들의 차이 대신 분석에 이용될 수 있다. 상기 메틸화 분석 접근은 홀배수체를 결정하거나 확인하기 위한 서열 리드 카운트 접근(RWK Chiu 등, 2008 Proc Natl Acad Sci USA;105:20458-20463) 및 혈장 DNA의 크기 분석이 수반되는 접근(US 특허 2011/0276277)과 조합될 수 있다. 메틸화 분석과 조합 이용되는 서열 리드 카운트 접근은 무작위 서열 분석(RWK Chiu 등, 2008 Proc Natl Acad Sci USA;105:20458-20463; DW Bianchi DW 등, 2012 Obstet Gynecol 119:890-901) 또는 표적화된 서열분석(AB Sparks 등, 2012 Am J Obstet Gynecol 206:319.e1-9; B Zimmermann 등, 2012 Prenat Diagn 32:1233-1241; GJ Liao 등, 2012 PLoS One; 7:e38154)을 이용해서 수행될 수 있다.

- [0184] BKG의 이용은 표본 간 배경 변동을 설명할 수 있다. 예를 들어, 한 여성은 다른 여성과 상이한 BKG 메틸화 수준을 가질 수 있지만, BKG 및 PLN 간 차이가 이러한 상황에서 표본에 걸쳐 이용될 수 있다. 상이한 염색체 영역에 대한 컷오프는, 예로 계놈의 한 영역의 메틸화 밀도가 계놈의 다른 영역에 비해 상이한 경우, 상이할 수 있다.

- [0185] 상기 접근은 태아 계놈에서의 결실 및 증폭을 포함하는 임의의 염색체 이상을 검출하기 위해 일반화될 수 있다. 또한, 상기 분석의 해상도는 원하는 수준으로 조정될 수 있다, 예를 들어 계놈이 10Mb, 5Mb, 2Mb, 1Mb, 500kb,

100kb bin으로 구분될 수 있다. 따라서 상기 기술은 아염색체 중복 또는 아염색체 결실을 검출하기 위해서도 이용될 수 있다. 따라서 상기 기술은 산전 태아 분자 핵형이 비침습적으로 획득될 수 있도록 할 것이다. 상기 방식으로 이용되는 경우, 상기 기술은 분자 카운팅에 기반한 비침습적 산전 평가 방법과 조합 이용될 수 있다(A Srinivasan 등, 2013 Am J Hum Genet;92:167-176; SCY Yu 등, 2013 PLoS One 8: e60968). 다른 구현예에서, bin의 크기가 동일할 필요는 없다. 예를 들어, bin의 크기는 각각의 bin이 동일한 수의 CpG 디뉴클레오타이드를 함유하도록 조정될 수 있다. 이러한 경우, bin의 물리적 크기가 상이할 것이다.

[0186] 공식은 상이한 유형의 염색체 이상에 적용하기 위해 다음과 같이 다시 쓰일 수 있다:

$$MP_{Diff} = (BKG - PLN) \times f \times 0.5 \times CN.$$

식 중, CN은 관련 영역에서 복제수 변화의 수를 나타낸다. CN은 염색체 1개 사본의 획득에 대해 1이고, 염색체 2개 사본의 획득에 대해 2이며, 두 개의 상동성 염색체 중 하나의 손실에 대해 -1이다(예로 여성 태아가 X 염색체 중 하나를 소실하여 XO 핵형으로 이어지는 태아 터너 증후군의 검출에 있어서). bin의 크기가 변하는 경우, 상기 식이 변할 필요는 없다. 그러나 더 작은 크기의 bin이 이용되는 경우, 더 적은 수의 CpG 디뉴클레오타이드(또는 태아 DNA 및 모체 DNA 간 차등적 메틸화를 나타내는 다른 뉴클레오타이드 조합)가 더 적은 bin에 존재하여 메틸화 밀도 측정의 증가된 확률적 변동으로 이어질 것이므로 민감성 및 특이성이 감소할 수 있다. 하나의 구현예에서, 필요한 측정치의 수는 메틸화 밀도의 변동 계수 및 원하는 민감성 수준을 분석하여 결정될 수 있다.

[0187] 상기 접근의 타당성을 나타내기 위해, 9명의 임신 여성으로부터 혈장 표본을 분석하였다. 5명의 임신 여성에서, 각각은 정배수체 태아를 수반하였으며, 다른 4명은 삼염색체중 21(T21) 태아를 수반하였다. 5명의 정배수체 임신부 중 3명이 무작위 선택되어 기준 군을 형성하였다. 나머지 2명의 정배수체 임신 케이스(Eu1 및 Eu2) 및 4명의 T21 케이스(T21-1, T21-2, T21-3 및 T21-4)는 잠재적인 T21 상태에 대해 평가하기 위해 상기 접근을 이용해서 분석되었다. 혈장 DNA는 바이셀파이트-전환되고 Illumina HiSeq2000 플랫폼을 이용해서 서열분석되었다. 하나의 구현예에서, 개별 염색체의 메틸화 밀도가 계산되었다. 이어서 염색체 21 및 다른 21개 상염색체의 평균 간 메틸화 밀도 차이가 결정되어 정규화된 메틸화 밀도가 획득되었다(표 1). 기준 군의 평균 및 SD가 6개 평가 케이스의 z-스코어 계산을 위해 이용되었다.

표 1

[0188] 표본을 T21로 분류하기 위한 z-스코어 <-3의 컷오프 이용으로, 모든 정배수체 및 T21 케이스의 분류가 정확하였음.

	Eu1	Eu2	T21-1	T21-2	T21-3	T21-4
chr 21 및 다른 상염색체 간 MP_{Diff} 에 대한 z-스코어	-1.48	1.09	-4.46	-5.30	-8.06	-5.69

[0189] 다른 구현예에서, 계층을 1Mb bin으로 나누고 각각의 1Mb bin에 대한 메틸화 밀도를 결정하였다. 잠재적인 홀배수체 염색체에 대한 모든 bin의 메틸화 밀도는 추정되는 비-홀배수체 염색체 상에 존재하는 모든 bin의 중앙값 메틸화 밀도로 정규화될 수 있다. 하나의 시행에서, 각각의 bin에 대해 비-홀배수체 bin의 중앙값으로부터의 메틸화 밀도 차이가 계산될 수 있다. z-스코어는 기준 군의 평균 및 SD값을 이용하여 이들 값에 대해 계산될 수 있다. 저메틸화를 나타내는 bin의 백분율이 결정되고 컷오프 백분율과 비교할 수 있다(표 2).

표 2

[0190] 염색체 21 상에 유의미하게 더 저메틸화된 bin에 대해 컷오프 5% 이용으로, 모든 케이스가 T21 상태에 대해 정확히 분류되었음.

	Eu1	Eu2	T21-1	T21-2	T21-3	T21-4
MP_{Diff} <-3의 z-스코어를 갖는 chr 21 상 bin의 백분율	0%	0%	33.3%	58.3%	19.4%	52.8%

[0191] 태아 염색체 또는 아염색체 이상을 검출하기 위한 상기 DNA 메틸화-기반 접근은 분자 카운팅 기반 접근, 예컨대 서열분석(RWK Chiu 등, 2008 Proc Natl Acad Sci USA; 105: 20458-20463) 또는 디지털 PCR(YMD Lo 등, 2007

Proc Natl Acad Sci USA; 104: 13116-13121), 또는 DNA 분자의 크기별 배열(US 특허 공개 2011/0276277)과 함께 이용될 수 있다. 이러한 조합(예로 DNA 메틸화 + 분자 카운팅, 또는 DNA 메틸화 + 크기별 배열, 또는 DNA 메틸화 + 분자 카운팅 + 크기별 배열)은 임상 환경에서 유리할, 예로 민감성 및/또는 특이성을 개선할 상승 효과를 가질 것이다. 예를 들어, 예로 서열분석으로 분석되어야 할 DNA 분자의 수가 진단 정확성에 부정적 영향을 미치지 않고 감소될 수 있다. 상기 특성은 이러한 평가가 보다 경제적으로 수행될 수 있도록 할 것이다. 다른 예로서, 주어진 수의 분석되는 DNA 분자에 대해, 조합 접근은 더 낮은 분획 농도의 태아 DNA로 태아 염색체 또는 아염색체 이상이 검출될 수 있도록 할 것이다. 도 13은 유기체의 생물학적 표본에서 염색체 이상을 검출하기 위한 방법 1300의 순서도이다. 생물학적 표본에는 제1 조직에서 그리고 상기 제2 조직에서 기원하는 무세포 DNA의 혼합물을 포함하는 무세포 DNA가 포함된다. 제1 조직은 태아 또는 종양 유래일 수 있고 상기 제2 조직은 임신 여성 또는 환자 유래일 수 있다.

[0192] 블록 1310에서, 상기 생물학적 표본으로부터의 복수의 DNA 분자가 분석된다. DNA 분자의 분석에는 유기체의 게놈에서 상기 DNA 분자의 위치 결정 및 DNA 분자가 하나 이상의 부위에서 메틸화되어 있는지의 여부를 결정하는 것이 포함될 수 있다. 분석은 메틸화-인식 서열분석에서 서열 리드를 수신하여 수행될 수 있고, 이에 따라 분석은 DNA에서 이전에 획득된 데이터에 대해서만 수행될 수 있다. 다른 구현예에서, 분석에는 실제 서열분석 또는 데이터 수득의 다른 유효 단계가 포함될 수 있다.

[0193] 위치 결정에는 인간 게놈의 각 부분, 예로 특정 영역으로의 DNA 분자의 맵핑(예로 서열 리드를 통해)이 포함될 수 있다. 하나의 시행에서, 측정치가 관심 영역에 맵핑되지 않는 경우, 측정치는 무시될 수 있다.

[0194] 블록 1320에서, 부위에서 메틸화되어 있는 DNA 분자의 각각의 수가 각각의 복수의 부위에 대해 결정된다. 하나의 구현예에서, 부위는 CpG 부위이며, 본원에 언급된 하나 이상의 기준을 이용하여 선택된 바와 같은, 특정 CpG 부위만일 수도 있다. 메틸화되는 DNA의 수는 특정 부위에서 분석되는 DNA 분자의 총 수, 예로 서열 리드의 총 수를 이용하여 정규화가 수행되면 비메틸화되는 수를 결정하는 것과 동등하다.

[0195] 블록 1330에서, 제1 염색체 영역의 제1 메틸화 수준은 제1 염색체 영역 내 부위에서 메틸화되어 있는 DNA 분자의 각각의 수에 기반하여 계산된다. 제1 염색체 영역은 임의 크기, 예로 상기 언급된 크기일 수 있다. 메틸화 수준은, 예로 정규화 절차의 일부로서 제1 염색체 영역에 지정된 DNA 분자의 총 수를 설명할 수 있다.

[0196] 제1 염색체 영역은 임의 크기(예로 전체 염색체)일 수 있고, 분리된 서브영역으로 이루어질 수도 있다, 즉 서브영역이 서로 분리된다. 각각의 서브영역의 메틸화 수준은, 예로 평균 또는 중앙값으로 결정되고 조합되어 제1 염색체 영역에 대한 메틸화 수준을 결정할 수 있다.

[0197] 블록 1340에서, 제1 메틸화 수준이 컷오프값과 비교된다. 컷오프값은 기준 메틸화 수준이거나 기준 메틸화 수준에 관련될 수 있다(예로 정규화 수준에서 특정된 거리에 있음). 컷오프값은 제1 염색체 영역에 대해 염색체 이상이 없는 태아를 수반하는 다른 임신 대상자, 암이 없는 개인의 표본, 또는 홀배수체와 연관되지 않은 것으로 알려진 유기체의 유전자좌(즉 이염색체인 영역)에서 결정될 수 있다.

[0198] 하나의 구현예에서, 컷오프값은 다음의 기준 메틸화 수준으로부터의 차이를 갖는 것으로 정의될 수 있다:

$(BKG - PLN) \times f \times 0.5 \times CN$, 식 중 BKG는 여성의 배경(또는 다른 대상자로부터의 평균 또는 중앙값)이고, f 는 제1 조직에서 기원하는 무세포 DNA의 분획 농도이고, CN은 평가되는 복제수이다. CN은 한 유형의 이상(결실 또는 중복)에 대응하는 스케일 인자의 예이다. CN에 대해 컷오프 1이 초기에 모든 증폭을 평가하기 위해 이용될 수 있고, 이어서 추가 컷오프가 증폭 정도를 결정하기 위해 이용될 수 있다. 컷오프값은, 예로 복제수 이상이 존재하지 않는 경우 유전자좌에 대한 예측 메틸화 수준을 결정하기 위해, 제1 조직에서 기원하는 무세포 DNA의 분획 농도에 기반할 수 있다.

[0199] 블록 1350에서, 제1 염색체 영역에 대한 이상의 분류는 비교에 기반하여 결정된다. 통계적으로 유의미한 수준 차이는 태아가 염색체 이상을 가질 증가된 위험을 시사할 수 있다. 다양한 구현예에서, 염색체 이상은 삼염색체 증 21, 삼염색체 증 18, 삼염색체 증 13, 터너 증후군, 또는 클라인펠터 증후군일 수 있다. 다른 예는 아염색체 결실, 아염색체 중복, 또는 디조지 증후군이다.

[0200] V. 마커 결정

[0201] 위에서 주지된 바와 같이, 태아 게놈의 특정 부분은 모체 게놈과 상이하게 메틸화된다. 이들 차이는 임신부에 걸쳐 일반적일 수 있다. 상이한 메틸화 영역이 태아에서 유래된 DNA 단편을 확인하기 위해 이용될 수 있다.

- [0202] A. 태반 조직 및 모체 조직에서 DMR을 결정하기 위한 방법
- [0203] 태반은 조직-특이적 메틸화 특징을 갖는다. 태아-특이적 DNA 메틸화 마커가 태반 조직 및 모체 혈액 세포 간에 차등적으로 메틸화된 유전자좌에 기반하여 모체 혈장 검출을 위해 그리고 비침습적 산전 진단 적용을 위해 개발되었다(SSC Chim 등, 2008 Clin Chem; 54: 500-511; EA Papageorgiou 등, 2009 Am J Pathol; 174: 1609-1618; 및 T Chu 등, 2011 PLoS One; 6: e14723). 게놈에 걸친 기준으로 이러한 차등적으로 메틸화된 영역(DMR)을 마 이닝하기 위한 구현예가 제공된다.
- [0204] 도 14는 본 발명의 구현예에 따라 태반 메틸화 프로필을 모체 메틸화 프로필(예로, 혈액 세포에서 결정됨)과 비 교하여 메틸화 마커를 확인하기 위한 방법 1400의 순서도이다. 방법 1400은 또한 종양 메틸화 프로필을 건강한 조직에 대응하는 메틸화 프로필과 비교하여 종양에 대한 마커를 결정하기 위해 이용될 수 있다.
- [0205] 블록 1410에서, 태반 메틸롬 및 혈액 메틸롬이 수득된다. 태반 메틸롬은 태반 표본, 예로 CVS 또는 만기 태반에 서 결정될 수 있다. 메틸롬은 게놈 일부만의 메틸화 밀도를 포함할 수 있는 것임이 이해되어야 한다.
- [0206] 블록 1420에서, 특정한 수의 부위(예로, 5개 CpG 부위)를 포함하고 충분한 수의 측정치가 수득된 영역이 확인된 다. 하나의 구현예에서, 적어도 5개의 적격 CpG 부위를 포함하는 첫 번째 500-bp 영역을 배치하기 위해 각 염색 체의 한 말단부터 확인이 시작되었다. CpG 부위는 그 부위가 적어도 5개의 서열 리드로 커버되는 경우, 적격으 로 간주될 수 있다.
- [0207] 블록 1430에서, 태반 메틸화 지수 및 혈액 메틸화 지수가 각 부위에 대해 계산된다. 예를 들어, 메틸화 지수는 각각의 500bp 영역 내에서 모든 적격 CpG 부위에 대해 개별적으로 계산되었다.
- [0208] 블록 1440에서, 메틸화 지수는 지수 세트가 서로 상이한지를 결정하기 위해 모체 혈액 세포 및 태반 표본 간에 비교되었다. 예를 들어, 메틸화 지수는, 예를 들어 Mann-Whitney 평가를 이용하여 모체 혈액 세포 및 CVS 또는 만기 태반 간에 비교되었다. 더 적은 수가 위양성 영역을 감소시킬 경우 다른 값이 이용될 수 있지만, 예를 들 어 ≤ 0.01 의 P -값이 통계적으로 유의미하게 상이한 것으로 간주되었다.
- [0209] 하나의 구현예에서, 적격 CpG 부위의 수가 5 미만이거나 Mann-Whitney 평가가 유의미하지 않은 경우, 500-bp 영 역이 100bp 하류로 이동하였다. 영역은 Mann-Whitney 평가가 500bp 영역에 대해 유의미해질 때까지 하류로 계속 이동하였다. 이어서 다음 500-bp 영역이 고려될 것이다. 다음 영역이 Mann-Whitney 평가에 의해 통계적 유의성 을 나타내는 것으로 확인되면, 조합된 인접 영역이 1,000bp 이하인 한, 현재 영역에 추가될 것이다.
- [0210] 블록 1450에서, 통계적으로 유의미하게 상이한(예로, Mann-Whitney 평가에 의해) 인접 영역은 병합될 수 있다. 차이는 두 표본에 대한 메틸화 지수 간에 대한 것임을 주지하라. 하나의 구현예에서, 인접 영역이 서로 특정된 거리(예로, 1,000bp) 내에 있고 이들이 유사한 메틸화 프로필을 나타낸 경우, 이들은 병합될 수 있다. 하나의 시행에서, 인접 영역 간 메틸화 프로필의 유사성은 하기 중 임의의 이용하여 정의될 수 있다: (1) 모체 혈액 세 포에 비해 태반 조직에서 동일한 경향성을 나타냄, 예로 두 영역이 혈액 세포에 비해 태반 조직에서 더 많이 메틸화됨; (2) 태반 조직에서 인접 영역에 대해 10% 미만의 메틸화 밀도 차이를 가짐; 및 (3) 모체 혈액 세포에서 인접 영역에 대해 10% 미만의 메틸화 밀도 차이를 가짐.
- [0211] 블록 1460에서, 영역에서 모체 혈액 세포 DNA 및 태반 표본(예로, CVS 또는 만기 태반 조직)으로부터의 혈액 메틸롬의 메틸화 밀도가 계산되었다. 메틸화 밀도는 본원에 기재된 바와 같이 결정될 수 있다.
- [0212] 블록 1470에서, 영역에서 모든 부위에 대한 총 태반 메틸화 밀도 및 총 혈액 메틸화 밀도가 통계적으로 유의미 하게 상이한 추정 DMR이 결정된다. 하나의 구현예에서, 병합 영역 내의 모든 적격 CpG 부위가 χ^2 평가를 거친 다. 병합 영역 내의 모든 CpG 부위에 걸친 메틸화 및 비메틸화 시토신의 비율로서의 메틸화 시토신의 수가 모체 혈액 세포 및 태반 조직 간에 통계적으로 유의미하게 상이한 경우 χ^2 평가를 평가하였다. 하나의 시행에서, χ^2 평가에 대해 ≤ 0.01 의 P -값이 통계적으로 유의미하게 상이한 것으로 간주될 수 있다. χ^2 평가에 의해 유의성을 나타낸 병합 절편이 추정 DMR로 간주되었다.
- [0213] 블록 1480에서, 모체 혈액 세포 DNA의 메틸화 밀도가 상한 컷오프 초과 또는 하한 컷오프 미만인 유전자좌가 확 인되었다. 하나의 구현예에서, 모체 혈액 세포 DNA의 메틸화 밀도가 $\leq 20\%$ 또는 $\geq 80\%$ 인 유전자좌가 확인되었다. 다른 구현예에서, 침, 소변 또는 여성 생식기의 자궁 경부 세척액, 눈물, 땀, 침 및 소변을 비제한 적으로 포함하는 모체 혈액 이외의 체액이 이용될 수 있다.
- [0214] 모체 혈장 중 태아-특이적인 DNA 메틸화 마커의 성공적 개발을 위한 핵심은 모체 혈액 세포의 메틸화 상태가 가

능한 한 고메틸화되거나 비메틸화되는 것일 수 있다. 이는 반대의 메틸화 프로필을 나타내는 태반-유래 태아 DNA 분자의 분석을 방해하는 모체 DNA 분자를 가질 가능성을 감소(예로 최소화)시킬 수 있다. 따라서 하나의 구현예에서, 후보 DMR이 추가 필터링에 의해 선택되었다. 후보 저메틸화 유전자좌는 모체 혈액 세포 중 메틸화 밀도 $\leq 20\%$ 및 태반 조직 중 적어도 20% 더 높은 메틸화 밀도를 나타낸 것이었다. 후보 고메틸화 유전자좌는 모체 혈액 세포 중 메틸화 밀도 $\geq 80\%$ 및 태반 조직 중 적어도 20% 더 낮은 메틸화 밀도를 나타낸 것이었다. 다른 백분율이 이용될 수 있다.

[0215] 블록 1490에서, 차이를 역치와 비교하여 태반 메틸화 밀도가 혈액 메틸화 밀도와 유의미하게 상이한 유전자좌 서브세트 중에서 DMR이 확인되었다. 하나의 구현예에서, 역치는 20%로, 메틸화 밀도는 모체 혈액 세포의 메틸화 밀도와 적어도 20% 상이하였다. 따라서 각각의 확인된 유전자좌에서의 태반 메틸화 밀도 및 혈액 메틸화 밀도 간 차이가 계산될 수 있다. 차이는 단순 차감일 수 있다. 다른 구현예에서, 차이를 결정하기 위해 스케일링 인자 및 다른 함수가 이용될 수 있다(예로, 차이는 단순 차감에 대해 적용되는 함수의 결과일 수 있다).

[0216] 하나의 시행에서, 상기 방법을 이용하여 11,729개의 고메틸화 및 239,747개의 저메틸화 유전자좌가 제1 석달 태반 표본에서 확인되었다. 상위 100개의 고메틸화 유전자좌가 첨부 표 S2A에 기재된다. 상위 100개의 저메틸화 유전자좌가 첨부 표 S2B에 기재된다. 표 S2A 및 S2B는 염색체, 시작 및 끝 위치, 영역의 크기, 모체 혈액 중 메틸화 밀도, 태반 표본 중 메틸화 밀도, P-값(모두 매우 작음), 및 메틸화 차이를 기재한다. 위치는 기준 게놈 hg18에 대응하며, 이는 hgdownload.soe.ucsc.edu/goldenPath/hg18/chromosomes에서 확인할 수 있다.

[0217] 11,920개의 고메틸화 및 204,768개의 저메틸화 유전자좌가 제3 석달 태반 표본에서 확인되었다. 제3 석달에 대한 상위 100개의 고메틸화 유전자좌가 표 S2C에 기재되며, 상위 100개의 저메틸화 유전자좌가 표 S2D에 기재된다. 이전에 모체 혈액 세포 및 제1 석달 태반 조직 간에 상이하게 메틸화된 것으로 이전에 보고된 33 유전자좌가 제1 석달 후보 목록을 검증하기 위해 이용되었다. 33 유전자좌의 79%는 본 발명자들의 알고리즘을 이용하여 DMR로 확인되었다.

[0218] 도 15a는 이전에 보고된 33의 제1 석달 마커에 대해 제1 석달 데이터를 이용한 DMR 확인 알고리즘의 성능을 나타내는 표 1500이다. 표에서, "a"는 유전자좌 1 내지 15가 이전에 (RWK Chiu 등, 2007 Am J Pathol; 170:941-950 및 SSC Chim 등, 2008 Clin Chem; 54:500-511)에 보고되었으며; 유전자좌 16 내지 23이 (KC Yuen, thesis 2007, The Chinese University of Hong Kong, Hong Kong)에 기재되었고; 유전자좌 24 내지 33이 (EA Papageorgiou 등, 2009 Am J Pathol; 174:1609-1618)에 기재되었음을 시사한다. "b"는 이들 데이터가 상기 문헌에서 유도되었음을 시사한다. "c"는 모체 혈액 세포 및 용모막 용모 표본의 메틸화 밀도 및 이들의 차이가 본 연구에서 생성된 서열분석 데이터에서 관찰되었지만 원래 연구에서 제공되는 게놈 좌표에 기반함을 시사한다. "d"는 유전자좌 상의 데이터가 Chiu 등(2007), Chim 등(2008), Yuen(2007) 및 Papageorgiou 등(2009)에 의해 위에 언급된 문헌을 참조하지 않고 바이설파이트 서열분석 데이터 상에서 방법 1400의 구현예를 이용하여 확인되었음을 시사한다. 유전자좌 범위에는 이전에 보고된 게놈 영역이 포함되지만 일반적으로 더 넓은 영역에 걸쳐 있었다. "e"는 후보 DMR이 모체 혈액 세포 및 용모막 용모 표본 중 DMR의 대응 게놈 좌표의 메틸화 밀도 간에 >0.20 차이를 관찰하는 요건에 기반하여 진양성(TP) 또는 위음성(FN)으로 분류되었음을 시사한다.

[0219] 도 15b는 출산 시 수득된 태반 표본에 비해 제3 석달 데이터를 이용한 DMR 확인 알고리즘의 성능을 나타내는 표 1550이다. "a"는 도 17a에 기재된 것과 동일한 33 유전자좌 목록이 이용되었음을 시사한다. "b"는 33 유전자좌가 초기 임신 표본에서 이전에 확인되었으므로, 이들이 제3 석달 데이터에 적용 가능하지 않을 수 있음을 시사한다. 따라서 원래 연구에서 제공되는 게놈 좌표에 기반한 만기 태반 조직 상에서 본 연구에서 생성된 바이설파이트 서열분석 데이터가 검토되었다. 유전자좌가 제3 석달에 실제로 진DMR인지를 결정하기 위해 모체 혈액 세포 및 만기 태반 조직 간 메틸화 밀도에서 >0.20 의 차이가 이용되었다. "c"는 유전자좌 상의 데이터가 Chiu 등(2007), Chim 등(2008), Yuen(2007) 및 Papageorgiou 등(2009)에 의해 전에 언급된 문헌을 참조하지 않고 바이설파이트 서열분석 데이터 상에서 방법 1400을 이용하여 확인되었음을 시사한다. 유전자좌 범위에는 이전에 보고된 게놈 영역이 포함되지만 일반적으로 더 넓은 영역에 걸쳐 있었다. "d"는 제3 석달에 상이하게 메틸화된 것이 적절한 유전자좌를 포함한 후보 DMR이 모체 혈액 세포 및 만기 태반 조직 중 DMR의 대응 게놈 좌표의 메틸화 밀도 간에 >0.20 차이를 관찰하는 요건에 기반하여 진양성(TP) 또는 위음성(FN)으로 분류되었음을 시사한다. 제3 석달에 상이하게 메틸화된 것에 적격이 아닌 유전자좌에 있어서, DMR 목록에서의 이들의 부재 또는 유전자좌를 함유하지만 메틸화 차이 <0.20 을 나타내는 DMR의 존재는 진음성(TN) DMR로 간주되었다.

[0220] B. 모체 혈장 서열분석 데이터로부터의 DMR

[0221] 표본의 분획 태아 DNA 농도가 또한 알려진 경우, 모체 혈장 DNA 바이설파이트-서열분석 데이터에서 직접 태반

조직 DMR을 확인할 수 있어야 한다. 태반은 모체 혈장 중 태아 DNA의 주요 원천이므로 이것이 가능하며(SSC Chim 등, 2005 Proc Natl Acad Sci USA 102, 14753-14758) 본 연구에서 모체 혈장 중 태아 특이적 DNA의 메틸화 상태가 태반 메틸롬과 연관되었음을 나타내었다.

[0222] 따라서 방법 1400의 측면은 태반 표본을 이용하는 대신 추정된 태반 메틸롬을 결정하기 위해 혈장 메틸롬을 이용하여 시행할 수 있다. 이에 따라 방법 1000 및 방법 1400이 DMR을 결정하기 위해 조합될 수 있다. 방법 1000은 태반 메틸화 프로필에 대한 예측값을 결정하고 이들을 방법 1400에서 이용하기 위해 이용될 수 있다. 상기 분석을 위해, 실시예는 또한 모체 혈액 세포 중 $\leq 20\%$ 또는 $\geq 80\%$ 메틸화된 유전자좌에 초점을 맞춘다.

[0223] 하나의 시행에서, 모체 혈액 세포에 비해 태반 조직 중에서 고메틸화된 유전자좌를 추정하기 위해, 모체 혈액 세포 중 $\leq 20\%$ 메틸화 및 혈액 세포 메틸화 밀도와 추정값 간 적어도 50% 차이를 가지는 추정값에 따른 $\geq 60\%$ 메틸화를 나타낸 유전자좌에 대해 정렬하였다. 모체 혈액 세포에 비해 태반 조직 중에서 저메틸화된 유전자좌를 추정하기 위해, 모체 혈액 세포 중 $\geq 80\%$ 메틸화 및 혈액 세포 메틸화 밀도 및 추정값 간 적어도 50% 차이를 가지는 추정값에 따른 $\leq 40\%$ 메틸화를 나타낸 유전자좌에 대해 정렬하였다.

[0224] 도 16은 모체 혈장 바이설파이트-서열분석 데이터의 직접 분석에 기반하여 고메틸화 또는 저메틸화된 것으로 예측되는 유전자좌의 수를 나타내는 표 1600이다. "N/A"는 해당되지 않음을 의미한다. "a"는 고메틸화 유전자좌에 대한 검색이 모체 혈액 세포 중 $<20\%$ 메틸화 밀도를 나타내는 유전자좌 목록에서 시작하였음을 시사한다. "b"는 저메틸화 유전자좌에 대한 검색이 모체 혈액 세포 중 $>80\%$ 메틸화 밀도를 나타내는 유전자좌 목록에서 시작하였음을 시사한다. "c"는 용모막 용모 표본으로부터의 바이설파이트-서열분석 데이터가 제1 석달 모체 혈장 데이터의 확인을 위해 이용되었고 만기 태반 조직이 제3 석달 모체 혈장 데이터의 확인을 위해 이용되었음을 시사한다.

[0225] 표 1600에 나타낸 바와 같이, 대부분의 비침습적으로 추정된 유전자좌는 조직에서 예측되는 메틸화 패턴을 나타내었고, 조직 데이터에서 마이닝되고 이전 섹션에 제시된 DMR과 중복되었다. 첨부에서는 혈장에서 확인된 DMR을 기재한다. 표 S3A는 제1 석달 모체 혈장 바이설파이트-서열분석 데이터에서 고메틸화된 것으로 추정된 상위 100개 유전자좌를 기재한다. 표 S3B는 제1 석달 모체 혈장 바이설파이트-서열분석 데이터에서 저메틸화된 것으로 추정된 상위 100개 유전자좌를 기재한다. 표 S3C는 제3 석달 모체 혈장 바이설파이트-서열분석 데이터에서 고메틸화된 것으로 추정된 상위 100개 유전자좌를 기재한다. 표 S3D는 제3 석달 모체 혈장 바이설파이트-서열분석 데이터에서 저메틸화된 것으로 추정된 상위 100개 유전자좌를 기재한다.

[0226] *C. 태반 및 태아 메틸롬의 임신 중 변동*

[0227] CVS 에서 메틸화 CpG의 전체적 비율은 55%였던 반면 만기 태반에 대해서는 59%였다(도 1의 표 100). 더 많은 저메틸화 DMR이 만기 태반에서보다 CVS에서 확인될 수 있던 반면, 고메틸화 DMR의 수는 두 조직에 대해 유사하였다. 따라서, CVS가 만기 태반에 비해 더 저메틸화됨이 명백하였다. 이러한 임신 경향은 모체 혈장 데이터에서도 명백하였다. 태아-특이적 측정치 중 메틸화 CpG의 비율은 제1 석달 모체 혈장에서 47.0%였지만 제3 석달 모체 혈장에서는 53.3%였다. 검증된 고메틸화 유전자좌의 수는 제1(1,457 유전자좌) 및 제3 석달(1,279 유전자좌) 모체 혈장 표본에서 유사하였지만 제3 석달(12,677 유전자좌) 표본에 비해 제1 석달(21,812 유전자좌) 표본에서 실질적으로 더 많은 저메틸화 유전자좌가 존재하였다(도 16의 표 1600).

[0228] *D. 마커의 이용*

[0229] 상이하게 메틸화된 마커, 또는 DMR은 여러 측면에서 유용하다. 모체 혈장 중 이러한 마커의 존재는 태아 또는 태반 DNA의 존재를 시사하고 확인시킨다. 상기 확인은 비침습적 산전 평가를 위한 품질 대조군으로 이용될 수 있다. DMR은 모체 혈장 중 일반적인 태아 DNA 마커로 작용할 수 있고, 모체 및 태아 간 유전형 차이에 의존하는 마커, 예컨대 다형성 기반 마커 또는 염색체 Y에 기반하는 마커에 비해 장점을 가질 수 있다. DMR은 모든 임신에 대해 유용한 일반적 태아 마커이다. 다형성 기반 마커는 태아가 마커를 그 아버지에게 유전받았고 모체가 그 녀의 계놈에 상기 마커를 보유하지 않는 임신의 하위세트에만 적용 가능하다. 또한, 이들 DMR에서 유래되는 DNA 분자의 정량에 의해 모체 혈장 표본에서 태아 DNA 농도를 측정할 수 있다. 정상 임신에 대해 예측되는 DMR 프로필을 인지함으로써, 임신 연관 합병증, 특히 태반 조직 변화가 수반되는 것들이 정상 임신에서 예측되는 것으로부터 모체 혈장 DMR 프로필 또는 메틸화 프로필의 일탈을 관찰하여 검출될 수 있다. 태반 조직 변화가 수반되는 임신 연관 합병증에는 비제한적으로 태아 염색체 흡배수체가 포함된다. 예에는 삼염색체증 21, 자간전증, 자궁 내 성장 지연 및 조산이 포함된다.

[0230] *E. 마커를 이용하는 키트*

- [0231] 구현예에서는 본원에 기재된 방법 및 다른 적용 가능한 방법을 실시하기 위한 조성물 및 키트를 제공할 수 있다. 키트는 태아 DNA, 예로 모체 혈장 중 무세포 태아 DNA를 분석하기 위해 분석을 수행하는데 이용될 수 있다. 하나의 구현예에서, 키트에는 본원에서 확인된 하나 이상의 유전자좌와의 특이적 혼성화에 유용한 적어도 하나의 올리고뉴클레오티드가 포함될 수 있다. 키트에는 또한 하나 이상의 기준 유전자좌와의 특이적 혼성화에 유용한 적어도 하나의 올리고뉴클레오티드가 포함될 수 있다. 하나의 구현예에서, 태반 고메틸화 마커가 측정된다. 평가 유전자좌는 모체 혈장 중의 메틸화 DNA일 수 있고, 기준 유전자좌는 모체 혈장 중의 메틸화 DNA일 수 있다. 유사한 키트가 혈장 중 종양 DNA의 분석을 위해 구성될 수 있다.
- [0232] 일부 경우, 키트에는 표적 유전자좌(예로, 첨부된 유전자좌) 및 기준 유전자좌의 적어도 한 구획의 증폭에 이용될 수 있는 적어도 두 올리고뉴클레오티드 프라이머가 포함될 수 있다. 프라이머 대신 또는 이에 부가하여, 키트에는 표적 유전자좌 및 기준 유전자좌에 대응하는 DNA 단편을 검출하기 위한 표지된 프로브가 포함될 수 있다. 다양한 구현예에서, 키트의 하나 이상의 올리고뉴클레오티드는 첨부된 표에서의 유전자좌에 대응한다. 전형적으로, 키트는 또한 사용자가 평가 표본을 분석하고 평가 대상자에서의 생리 또는 병리 상태를 평가하기 위한 명령을 제공하는 명령 매뉴얼을 제공한다.
- [0233] 다양한 구현예에서, 태아 DNA 및 태아를 임신한 여성 대상자 유래 DNA의 혼합물을 함유하는 생물학적 표본 중 태아 DNA를 분석하기 위한 키트가 제공된다. 키트는 표 S2A, S2B, S2C, S2D, S3A, S3B, S3C, 및 S3D에 기재된 게놈 영역의 적어도 한 구획에 특이적으로 혼성화하기 위한 하나 이상의 올리고뉴클레오티드를 포함할 수 있다. 따라서, 표에 걸쳐 또는 단 하나의 표로부터의 임의의 수의 올리고뉴클레오티드가 이용될 수 있다. 올리고뉴클레오티드는 프라이머로 작용할 수 있고, 프라이머쌍으로 조직화될 수 있으며, 여기서 한 쌍이 표로부터의 특정 영역에 대응한다.
- [0234] VI. 크기 및 메틸화 밀도의 관계
- [0235] 혈장 DNA 분자는 짧은 분자의 형태로 순환에 존재하는 것으로 알려져 있고, 대부분의 분자가 약 160bp 길이이다(YMD Lo 등, 2010 Sci Transl Med; 2: 61ra91, YW Zheng 등, 2012 Clin Chem; 58: 549-558). 흥미롭게도 본 발명자들의 데이터는 혈장 DNA 분자의 메틸화 상태 및 크기 간 관계를 드러내었다. 즉, 혈장 DNA 단편 길이는 DNA 메틸화 수준에 연관된다. 혈장 DNA 분자의 특징적인 크기 프로파일은 대부분이 아마도 아포토시스 동안의 효소 분해에서 유도될 수 있는 단일 뉴클레오솜과 연관됨을 제시한다.
- [0236] 순환 DNA는 성질 상 단편화되어 있다. 특히 순환 태아 DNA는 모체 혈장 표본 중의 모체 유래 DNA보다 짧다(KCA Chan 등, 2004 Clin Chem; 50: 88-92). 페어-엔드(paired-end) 정렬이 바이셀파이트-처리 DNA의 크기 분석을 가능케 하므로, 혈장 DNA 분자의 크기 및 이들 각각의 메틸화 수준 간에 존재하는 경우 임의 상관관계가 직접 평가할 수 있다. 이를 모체 혈장뿐만 아니라 비임신 성인 여성 대조군의 혈장 표본 중에서 탐색하였다.
- [0237] 각각의 DNA 분자의 양 말단에 대한 페어-엔드 서열분석(전체 분자의 서열분석이 포함됨)이 본 연구에서 각 표본을 분석하기 위해 이용되었다. 각 DNA 분자의 말단 서열 페어를 기준 인간 게놈에 대해 정렬하고 서열분석 측정치의 극 말단의 게놈 좌표를 주지하여, 서열분석된 DNA 분자의 길이를 결정할 수 있다. 혈장 DNA 분자는 자연적으로 소분자로 단편화되며, 혈장 DNA에 대한 서열분석 라이브러리는 전형적으로 임의의 단편화 단계 없이 제조된다. 따라서 서열분석으로 추정된 길이는 원래 혈장 DNA 분자의 크기를 나타내었다.
- [0238] 이전 연구에서, 모체 혈장 중 태아 및 모체 DNA 분자의 크기 프로파일을 결정하였다(YMD Lo 등, 2010 Sci Transl Med; 2: 61ra91). 혈장 DNA 분자가 단일 뉴클레오솜과 유사한 크기를 가졌으며 태아 DNA 분자가 모체 분자에 비해 더 짧았음을 나타내었다. 본 연구에서는 혈장 DNA 분자의 메틸화 상태와 이들의 크기에 대한 관계를 결정하였다.
- [0239] A. 결과
- [0240] 도 17a는 모체 혈장, 비임신 여성 대조군 혈장, 태반 및 말초 혈액 DNA의 크기 분포를 나타내는 그래프 1700이다. 모체 표본 및 비임신 여성 대조군 혈장에 있어서, 두 바이셀파이트-처리 혈장 표본은 이전에 보고된 것과 동일한 특징적 크기 분포를 나타내었고(YMD Lo 등, 2010 Sci Transl Med; 2: 61ra91) 가장 풍부한 전체 서열은 166-167bp 길이이고 143bp보다 짧은 DNA 분자가 10-bp 주기로 나타났다.
- [0241] 도 17b는 모체 혈장, 성인 여성 대조군 혈장, 태반 조직 및 성인 여성 대조군 혈액 크기 분포 및 메틸화 프로파일 그래프 1750이다. 동일한 크기이고 적어도 하나의 CpG 부위를 함유하는 DNA 분자에 있어서, 이들의 평균 메틸화 밀도가 계산되었다. 이어서 DNA 분자의 크기 및 이들의 메틸화 밀도 간 관계를 그래프화하였다. 구체적으로 평

균 메틸화 밀도가 적어도 1 CpG 부위를 커버하는 서열분석 측정치에 대해 50bp 내지 최대 180bp 범위의 각각의 단편 길이에 대해 결정되었다. 흥미롭게도, 메틸화 밀도는 혈장 DNA 크기와 함께 증가하였고, 166-167bp 근처에서 피크를 나타냈다. 그러나 상기 패턴은 초음파 분쇄 시스템을 이용하여 단편화된 태반 및 대조군 혈액 DNA 표본에서는 관찰되지 않았다.

[0242] 도 18은 혈장 DNA 분자의 메틸화 밀도 및 크기 그래프를 나타낸다. 도 18a는 제1 석달 모체 혈장에 대한 그래프 1800이다. 도 18b는 제3 석달 모체 혈장에 대한 그래프 1850이다. 적어도 하나의 CpG 부위를 커버하는 모든 서열분석 측정치에 대한 데이터는 파란색 곡선 1805로 나타낸다. 태아-특이적 SNP 대립유전자를 또한 함유하는 측정치에 대한 데이터는 빨간색 곡선 1810으로 나타낸다. 모체-특이적 SNP 대립유전자를 또한 함유하는 측정치에 대한 데이터는 녹색 곡선 1815로 나타낸다.

[0243] 태아-특이적 SNP 대립유전자를 함유한 측정치는 태아 DNA 분자에서 유래된 것으로 간주된다. 모체-특이적 SNP 대립유전자를 함유한 측정치는 모체 DNA 분자에서 유래된 것으로 간주된다. 일반적으로, 높은 메틸화 밀도를 갖는 DNA 분자의 크기가 더 길었다. 이러한 경향은 제1 및 제3 석달 모두의 태아 및 모체 DNA 분자 둘 다에 존재하였다. 태아 DNA 분자의 전체적 크기는 이전에 보고된 바와 같이 모체보다 짧았다.

[0244] 도 19a는 성인 비임신 여성에 대한 서열분석 측정치의 크기 및 메틸화 밀도의 그래프 1900을 나타낸다. 성인 비임신 여성의 혈장 DNA 표본은 또한 DNA 분자의 크기 및 메틸화 상태 간에 동일한 관계를 나타내었다. 반면, 게놈 DNA 표본은 MPS 분석 전에 초음파 분쇄 단계로 단편화되었다. 그래프 1900에 나타난 바와 같이, 혈액 세포 및 태반 조직 표본으로부터의 데이터는 동일한 경향을 드러내지 않았다. 세포의 단편화가 인공적이므로, 크기 및 밀도 간에 관계를 갖지 않을 것으로 예상하게 된다. 혈장 중 천연 단편화된 DNA 분자는 크기에 대한 의존성을 나타내며, 메틸화 밀도가 낮을수록 분자가 더 작은 단편으로 분쇄되기 쉬워진다고 추정될 수 있다.

[0245] 도 19b는 모체 혈장 중 태아-특이적 및 모체-특이적 DNA 분자의 크기 분포 및 메틸화 프로필을 나타내는 그래프 1950을 나타낸다. 태아-특이적 및 모체-특이적 혈장 DNA 분자는 또한 단편 크기 및 메틸화 수준 간에 동일한 상관관계를 나타내었다. 태반 유래 및 모체 순환 무세포 DNA의 단편 길이는 모두 메틸화 수준과 함께 증가하였다. 또한, 이들의 메틸화 상태 분포가 서로 중복되지 않아, 그러한 현상이 순환 DNA 분자 원천의 원래 단편 길이와 무관하게 존재함을 제시하였다.

[0246] B. 방법

[0247] 따라서, 혈장 표본의 총 메틸화 백분율을 산정하기 위해 크기 분포가 이용될 수 있다. 이어서 이러한 메틸화 측정이 임신 동안, 암 모니터링 동안 또는 치료 동안 도 18a 및 18b에 나타난 관계에 따라 혈장 DNA 크기 분포의 연속 측정에 의해 추적될 수 있다. 메틸화 측정은 또한 관심 기관 또는 조직 유래 DNA의 증가되거나 감소된 배출을 살피기 위해 이용될 수 있다. 예를 들어, 특정 기관(예로 간)에 특이적인 DNA 메틸화 특징을 특이적으로 살피고 혈장 중 이들의 특징의 농도를 측정할 수 있다. DNA는 세포가 사멸할 때 혈장 중으로 배출되므로, 수준 증가는 그 특정 기관 또는 조직에서의 세포사 또는 손상 증가를 의미할 수 있다. 특정 기관으로부터의 수준 감소는 그 기관에서의 손상 또는 병리적 절차에 대응하기 위한 치료가 조절되고 있음을 의미할 수 있다.

[0248] 도 20은 본 발명의 구현예에 따른 유기체의 생물학적 표본에서 DNA의 메틸화 수준을 산정하기 위한 방법 2000의 순서도이다. 메틸화 수준은 게놈의 특정 영역 또는 전체 게놈에 대해 산정될 수 있다. 특정 영역을 원하는 경우, 그 특정 영역에서만 유래된 DNA 단편이 이용될 수 있다.

[0249] 블록 2010에서, 다양한 크기에 대응하는 DNA 단편의 양이 측정된다. 복수의 크기의 각각의 크기에 대해, 그 크기에 대응하는 생물학적 표본으로부터의 복수의 DNA 단편의 양이 측정될 수 있다. 예를 들어, 140염기 길이를 갖는 DNA 단편의 수가 측정될 수 있다. 그 양은 히스토그램으로 저장될 수 있다. 하나의 구현예에서, 상기 생물학적 표본으로부터의 복수의 핵산의 각각의 크기가 측정되고, 이는 개별 기준으로(예로 전체 분자 또는 분자 말단만의 단일 분자 서열분석에 의해) 또는 그룹 기준으로(예로 전기영동을 통해) 수행될 수 있다. 크기는 범위에 대응할 수 있다. 따라서 양은 특정 범위 내의 크기를 갖는 DNA 단편에 대한 것일 수 있다. 페어-엔드 서열분석이 수행되는 경우, 특정 영역에 대한 DNA 단편(페어 형성 서열 리드로 결정됨) 맵핑(정렬)이 영역의 메틸화 수준을 결정하기 위해 이용될 수 있다.

[0250] 블록 2020에서, 제1 파라미터의 제1값은 여러 크기의 DNA 단편의 양에 기반하여 계산된다. 하나의 측면에서, 제1 파라미터는 생물학적 표본 중 DNA 단편의 크기 프로필의 통계적 측정치(예로, 히스토그램)를 제공한다. 파라미터는 이것이 복수의 DNA 단편의 크기로부터 결정되므로, 크기 파라미터로 불릴 수 있다.

[0251] 제1 파라미터는 다양한 형태일 수 있다. 하나의 파라미터는 전체 DNA 단편에 대한 또는 다른 크기 또는 범위의

DNA 단편에 대한 특정 크기 또는 크기 범위의 DNA 단편의 백분율이다. 이러한 파라미터는 단편의 총 수로 나누어지는 특정 크기의 DNA 단편의 수이며, 이는 히스토그램(특정 크기 단편의 절대 또는 상대 카운트를 제공하는 임의의 데이터 구조)에서 수득될 수 있다. 다른 예로서, 파라미터는 다른 크기 또는 범위의 단편의 수로 나누어지는 특정 크기의 또는 특정 범위 내의 단편의 수일 수 있다. 구분은 상이한 수의 DNA 단편이 상이한 표본에 대해 분석되는 것을 설명하기 위한 정규화로 작용할 수 있다. 정규화는 각각의 표본에 대해 동일한 수의 DNA 단편으로 분석하여 달성될 수 있고, 이는 분석되는 단편의 총 수로 나누는 것과 동일한 결과를 효과적으로 제공한다. 파라미터 및 크기 분석에 대한 추가 예는 모든 목적에 대해 참조로 도입된 U.S. 특허 출원 13/789,553에서 확인할 수 있다.

[0252] 블록 2030에서, 제1 크기값은 기준 크기값과 비교된다. 기준 크기값은 기준 표본의 DNA 단편에서 계산될 수 있다. 기준 크기값을 결정하기 위해, 메틸화 프로필이 기준 표본뿐만 아니라 제1 크기 파라미터값에 대해 계산되고 정량될 수 있다. 따라서 제1 크기값이 기준 크기값과 비교되는 경우, 메틸화 수준이 결정될 수 있다.

[0253] 블록 2040에서, 메틸화 수준이 비교에 기반하여 산정된다. 하나의 구현예에서, 제1 파라미터의 제1값이 기준 크기값 초과 또는 미만인지를 결정하고, 이에 따라 대응 표본의 메틸화 수준이 기준 크기값에 대한 메틸화 수준의 초과 또는 미만인지를 결정할 수 있다. 다른 구현예에서, 비교는 제1값을 검정 함수로 입력하여 달성된다. 검정 함수는 제1값에 대응하는 곡선 상의 지점을 확인하여 제1값을 검정값(기준 크기값 세트)과 효과적으로 비교할 수 있다. 이어서 산정된 메틸화 수준이 검정 함수의 출력값으로 제공된다.

[0254] 따라서 크기 파라미터를 메틸화 수준에 대해 검정할 수 있다. 예를 들어, 메틸화 수준이 측정되고 그 표본에 대한 특정 크기의 파라미터와 연관될 수 있다. 이어서 다양한 표본으로부터의 데이터 지점이 검정 함수에 맞추어질 수 있다. 하나의 시행에서, 상이한 검정 함수가 상이한 DNA 서브세트에 대해 이용될 수 있다. 따라서 DNA의 특정 서브세트에 대한 메틸화 및 크기 간 관계에 대한 이전 지식에 기반한 일부 검정 형태가 존재할 수 있다. 예를 들어, 태아 및 모체 DNA에 대한 검정이 상이할 수 있다.

[0255] 위에 나타난 바와 같이, 태반은 모체 혈액에 비해 더 저메틸화되어 있고, 이에 따라 태아 DNA는 더 낮은 메틸화로 인해 더 작다. 따라서 표본 단편의 평균 크기(또는 다른 통계값)를 이용하여 메틸화 밀도를 산정할 수 있다. 단편 크기는 잠재적으로는 기술적으로 더 복잡한 메틸화-인식 서열분석보다는 페어-엔드 서열분석을 통해 측정될 수 있으므로, 상기 접근은 임상적으로 이용되는 경우 잠재적으로 비용 효과적일 것이다. 상기 접근은 임신 진행 또는 임신 연관 장애, 예컨대 자간전증, 조산 및 태아 장애(예컨대 염색체 또는 유전적 비정상성 또는 자궁내 성장 지연으로 야기되는 것들)에 연관된 메틸화 변화의 모니터링을 위해 이용될 수 있다.

[0256] 다른 구현예에서, 상기 접근은 암의 검출 및 모니터링을 위해 이용될 수 있다. 예를 들어, 성공적인 암 치료를 이용하면, 상기 크기 기반 접근을 이용하여 측정되는 혈장 또는 다른 체액 중 메틸화 프로필이 암이 없는 건강한 개인의 프로필로 변할 것이다. 반대로, 암이 진행 중인 경우, 혈장 또는 다른 체액 중의 메틸화 프로필은 암이 없는 건강한 개인의 프로필로부터 이탈할 것이다.

[0257] 요약하면, 저메틸화 분자가 혈장 중의 고메틸화 분자에 비해 더 짧았다. 동일 경향이 태아 및 모체 DNA 분자 모두에서 관찰되었다. DNA 메틸화는 뉴클레오타이드 충전에 영향을 미치는 것으로 알려져 있으므로, 본 발명자들의 데이터는 아마도 저메틸화 DNA 분자가 히스톤으로 덜 밀집되게 충전되며 이에 따라 효소 분해에 더 민감하였음을 제시한다. 반면, 도 18a 및 18b에 제시된 데이터는 또한 태아 DNA가 모체 측정치에 비해 훨씬 더 저메틸화됨에도 불구하고, 태아 및 모체 DNA의 크기 분포는 서로 완벽히 구분되지 않음을 나타내었다. 도 19b에서, 동일한 크기 분류에 대해서도 태아- 및 모체-특이적 측정치의 메틸화 수준이 서로 상이함을 알 수 있다. 이러한 관찰은 태아 DNA의 저메틸화 상태가 모체 DNA에 대비한 그 상대적 단축성을 설명하는 유일한 요인이 아님을 제시한다.

[0258] VII. 유전자 유전자좌의 각인 상태

[0259] 동일한 유전형을 그러나 모체 혈장 중 모체와 상이한 후생적 특징을 공유하는 태아-유래 DNA 분자가 검출될 수 있다(LLM Poon 등, 2002 Clin Chem; 48: 35-41). 서열분석 접근이 모체 혈장 중 태아 유래 DNA 분자의 선별에 민감함을 나타내기 위해, 모체 혈장 표본 중 각인된 태아 대립유전자를 검출하기 위해 동일한 전략을 적용하였다. 두 계층 각인 영역이 확인되었다: H19(chr11:1,977,419-1,977,821, NCBI Build36/hg18) 및 MEST(chr7:129,917,976-129,920,347, NCBI Build36/hg18). 둘 다 모체 및 태아 서열 간 차별화를 위해 정보를 제공하는 SNP를 함유한다. H19, 모체 발현 유전자의 경우, 모체는 영역 중의 SNP rs2071094(chr11:1,977,740)에 대해 동종접합성(A/A)이고 태아는 이종접합성(A/C)이었다. 모체 A 대립유전자 중 하나는 완전 메틸화되고 다른 것은 비메틸화된다. 그러나 태반에서는, A 대립유전자가 비메틸화된 반면, 부계에서 유전된 C 대립유전자는

완전 메틸화되었다. 본 발명자들은 모체 혈장 중, 태반에서 유래된 각인된 부계 대립유전자에 대응하는 C 유전형질을 갖는 두 메틸화 측정치를 검출하였다.

[0260] 또한 PEG1로도 알려져 있는 MEST는 부계 발현 유전자이다. 모체 및 태아는 모두 각인 유전자와 내에서 SNP rs2301335(chr7:129,920,062)에 대해 이종접합성(A/G)이었다. 모체 혈액 중 G 대립유전자는 메틸화된 반면 A 대립유전자는 비메틸화되었다. 태반에서는 메틸화 패턴이 역전되어 모체 A 대립유전자가 메틸화되고 부계 G 대립유전자가 비메틸화되었다. 부계 유래된 세 비메틸화 G 대립유전자가 모체 혈장 중에 검출될 수 있었다. 대조적으로, 염색체 19 상의 비각인 유전자 유전자좌(chr19:6,723,621-6,724,121)인 VAV1은 조직에서만 아니라 혈장 DNA 표본에서 어떠한 대립유전자 메틸화 패턴도 나타내지 않았다.

[0261] 따라서 메틸화 상태를 이용하여 어느 DNA 단편이 태아에서 유래되는지를 결정할 수 있다. 예를 들어, 모체 혈장에서 A 대립유전자만을 검출하는 것은 모체가 GA 이종접합성인 경우 태아 마커로 이용될 수 없다. 그러나 혈장에서 A 분자의 메틸화 상태를 구별하면, 메틸화 A 분자는 태아-특이적인 반면 비메틸화 A-분자는 모체-특이적이거나, 또는 반대이다.

[0262] 다음으로 태반 조직에서 게놈 각인을 나타내는 것으로 보고된 유전자좌에 초점을 맞추었다. Woodfine 등(2011 Epigenetics Chromatin; 4: 1)이 보고한 유전자좌 목록에 기반하여, 각인 조절 영역 내에 SNP를 함유한 것들에 대해 추가 정렬하였다. 네 유전자좌가 요건을 충족하였고, 이들은 H19, KCNQ10T1, MEST 및 NESP였다.

[0263] H19 및 KCNQ10T1에 대한 모체 혈액 세포 표본의 측정치에 있어서, 모체 측정치는 SNP에 대해 동종접합성이었고, 이들은 메틸화 및 비메틸화 측정치의 비율이 대략 같았다. CVS 및 만기 태반 조직 표본은 태아가 두 유전자좌에 대해 이종접합성이었으며 각각의 대립 유전자가 배타적으로 메틸화 또는 비메틸화되었음을, 즉 단일대립유전자 메틸화되었음을 드러내었다. 모체 혈장 표본에서, 부계 유전된 태아 DNA 분자가 두 유전자좌 모두에 대해 검출되었다. H19에 있어서, 부계 유전된 분자는 태아-특이적 대립유전자를 함유하고 메틸화된 서열분석 측정치로 나타났다. KCNQ10T1에 있어서, 부계 유전된 분자는 태아-특이적 대립유전자를 함유하고 비메틸화된 서열분석 측정치로 나타났다.

[0264] 반면, 모체는 MEST 및 NESP 모두에 대해 이종접합성이었다. MEST에 있어서, 모체 및 태아는 모두 SNP에 대해 GA 이종접합체였다. 그러나 모체 혈액 세포 및 태반 조직에서 왓슨 가닥에 대한 데이터에서 명백하듯이, SNP에 인접한 CpG에 대한 메틸화 상태는 모체 및 태아에서 반대였다. A-대립유전자는 모체의 DNA에서 비메틸화된 반면 태아의 DNA에서는 메틸화되었다. MEST에 있어서, 모체 대립유전자는 메틸화되었다. 따라서 태아가 그 모체로부터 A-대립유전자를 유전받았고(CVS 중 메틸화됨), 모체는 그 아버지로부터 A-대립유전자를 유전받았음(모체 혈액 세포 중 비메틸화됨)을 찾아낼 수 있었다. 흥미롭게도 모체 혈장 표본에서, 모체의 각각의 두 대립유전자 및 태아의 각각의 두 대립유전자를 포함하는 모든 네 분자 그룹이 쉽게 구별될 수 있었다. 따라서 각인 유전자좌에서 유전형 정보와 메틸화 상태를 조합하여, 배경 모체 DNA 분자로부터 모체에서 유전된 태아 DNA 분자를 쉽게 구별할 수 있었다(LLM Poon 등, 2002 Clin Chem; 48: 35-41).

[0265] 상기 접근을 이용하여 단부모 이염색체를 검출할 수 있었다. 예를 들어, 상기 태아의 아버지가 G-대립유전자에 대해 동종접합성으로 알려진 경우, 모체 혈장 중 비메틸화 G-대립유전자의 검출 실패는 부계 대립유전자의 기여 부재를 나타낸다. 또한 이러한 상황 하에서, 메틸화 G-대립유전자 및 메틸화 A-대립유전자가 모두 상기 임신부의 혈장에서 검출되었다면, 이는 태아가 모체로부터 이중이염색체를 가짐을, 즉 아버지로부터의 간섭 없이 모체로부터 두 상이한 대립유전자가 유전되었음을 제시할 것이다. 대안적으로 메틸화 A-대립유전자(모체에서 유전된 태아 대립유전자) 및 비메틸화 A-대립유전자(모체 할아버지에서 유전된 모체 대립유전자)가 모두 비메틸화 G-대립유전자(태아가 유전받았어야 하는 부계 대립유전자) 없이 모체 혈장에서 검출되는 경우, 이는 태아가 모체로부터 동종이염색체를 가짐을, 즉 아버지로부터의 간섭 없이 모체로부터 두 동일한 대립유전자가 유전되었음을 제시할 것이다.

[0266] NESP에 있어서, 모체는 SNP에서 GA 이종접합체인 반면 태아는 G-대립유전자에 대해 동종접합성이었다. 부계 대립유전자는 NESP에 대해 메틸화되었다. 모체 혈장 표본에서, 메틸화된, 부계 유전된 태아 G-대립유전자는 비메틸화된 배경 모체 G-대립유전자로부터 쉽게 구별될 수 있었다.

[0267] VIII. 암/공여자

[0268] 일부 구현예가 순환 혈청/혈장 DNA의 메틸화 분석을 이용하여 암의 검출, 스크리닝, 모니터링(예로 치료에 대한 재발, 완화, 또는 반응(예로 존재 또는 부재)), 단계 설정, 분류(예로 가장 적절한 치료 양식의 선택을 보조하기 위해) 및 예진을 위해 이용될 수 있다.

- [0269] 암 DNA는 이상 DNA 메틸화를 나타내는 것으로 알려져 있다(JG Herman 등, 2003 N Engl J Med; 349: 2042-2054). 예를 들어, 유전자, 예로 종양 억제인자 유전자의 CpG 섬 프로모터는 고메틸화된 반면, 유전자 본체에서의 CpG 부위는 비암세포에 비해 저메틸화되어 있다. 암 세포의 메틸화 프로파일은 본원에 기재된 방법을 이용하여 종양 유래 혈장 DNA의 분자의 메틸화 프로파일로 반영될 수 있는 한, 혈장 중의 전체적 메틸화 프로파일은 암이 있는 개인을 암이 없는 건강한 개인과 비교 시 또는 암이 치유된 개인과 비교 시 상이할 것으로 예측된다. 메틸화 프로파일의 차이 유형은 계놈의 메틸화 밀도 및/또는 계놈 절편의 메틸화 밀도에서의 정량적 차이의 관점일 수 있다. 예를 들어 암 조직 DNA의 일반적인 저메틸화 성질로 인해(Gama-Sosa MA 등, 1983 Nucleic Acids Res; 11: 6883-6894), 혈장 메틸롬 또는 계놈 절편에서의 메틸화 밀도 감소가 암 환자의 혈장에서 관찰될 것이다.
- [0270] 메틸화 프로파일의 정량적 변화는 또한 혈장 메틸롬 데이터에 걸쳐 반영될 수 있다. 예를 들어, 암 세포에서만 고메틸화된 유전자에서 유래되는 혈장 DNA 분자는 동일한 유전자에서 유래되지만 건강한 대조군 표본 중의 혈장 DNA 분자에 비해 암 환자의 혈장에서 고메틸화를 나타낼 것이다. 이상 메틸화는 대부분의 암에서 일어나므로, 본원에 기재된 방법은 이상 메틸화를 갖는 모든 형태의 악성물, 예를 들어 폐, 유방, 결직장, 전립선, 코인두, 위, 정소, 피부, 신경계, 뼈, 난소, 간, 혈액 조직, 췌장, 자궁, 신장, 방광, 림프 조직 등에서 비제한적으로 포함하는 악성물의 검출에 적용될 수 있다. 악성물은 다양한 조직 하위유형, 예를 들어 암종, 선암종, 육종, 섬유선암종, 신경내분비 및 미분화된 것 등일 수 있다.
- [0271] 반면, 본 발명자들은 DNA 분자의 크기에 추가 효과를 가질 종양 연관된 이상 저메틸화를 갖는 유전자좌 유래의 DNA 분자에서 종양 유래 DNA 분자의 전체적으로 짧은 크기 프로파일은 강조되므로, 종양 유래 DNA 분자가 배경 비-종양 유래 DNA 분자와 구별될 수 있는 것으로 예상한다. 또한, 종양 유래 혈장 DNA 분자는 단일 뉴클레오타이드 변이체, 복제수 획득 및 손실, 전위, 역전, 이상 고- 또는 저-메틸화 및 크기 프로파일 확인을 비제한적으로 포함하는 종양 DNA와 연관된 여러 특징적 특성을 이용하여 배경 비-종양 유래 혈장 DNA 분자와 구별될 수 있다. 이들 변화가 모두 독립적으로 일어날 수 있으므로, 이들 특성의 조합 이용은 혈장 중 암 DNA의 민감하고 특이적인 검출을 위해 추가적인 장점을 제공할 수 있다.
- [0272] A. 크기 및 암
- [0273] 혈장 중 종양-유래 DNA 분자의 크기는 또한 단일 뉴클레오타이드 단위체의 크기와 유사하고, 암 환자의 혈장 중에 공존하는 배경 비-종양-유래 DNA 분자보다 짧다. 크기 파라미터는 모든 목적을 위해 참조로 도입된, U.S. 특허 출원 13/789,553에 기재된 바와 같이 암에 연관된 것으로 나타났다.
- [0274] 혈장 중 태아-유래 및 모체-유래 DNA는 모두 분자의 크기 및 메틸화 상태 간의 관계를 나타내었으므로, 종양-유래 DNA 분자는 동일한 경향을 나타낼 것으로 예상된다. 예를 들어, 저메틸화 분자는 암에 대해 스크리닝되는 대상자 또는 암 환자의 혈장 중에서 고메틸화 분자에 비해 더 짧을 것이다.
- [0275] B. 암 환자에서 상이한 조직의 메틸화 밀도
- [0276] 본 예에서, 간세포암종(HCC) 환자의 혈장 및 조직 표본을 분석하였다. 혈액 표본을 종양의 수술적 절제 이전 및 이후 1주에 HCC 환자에서 수집하였다. 혈액 표본의 원심분리 후 혈장 및 버피 코트를 수확하였다. 절제된 종양 및 인접한 비-종양 간 조직을 수집하였다. 혈장 및 조직 표본에서 추출된 DNA 표본을 사전 바이설파이트 처리와 함께 또는 이것이 없이 대량 병렬 서열분석을 이용하여 분석하였다. 암이 없는 4명의 건강한 개인의 혈장 DNA도 대조군으로 분석하였다. DNA 표본의 바이설파이트 처리는 비메틸화 시토신 잔기를 우라실로 전환할 것이다. 하류 폴리머라아제 연쇄 반응 및 서열분석에서, 이들 우라실 잔기는 티미딘으로 작용할 것이다. 반면, 바이설파이트 처리는 메틸화 시토신 잔기를 우라실로 전환하지 않을 것이다. 대량 병렬 서열분석 후, 서열분석 측정치를 Methy-Pipe(P Jiang, 등, Methy-Pipe: 전체 계놈 메틸롬 분석을 위한 통합 바이오인포매틱스 데이터 분석 파이프라인, 논문은 2010. 12. 18-21, IEEE International Conference on Bioinformatics and Biomedicine Workshops, Hong Kong에서 발표됨)로 분석하여 모든 CG 디뉴클레오타이드 위치, 즉 CpG 부위에서 시토신 잔기의 메틸화 상태를 결정하였다.
- [0277] 도 21a는 HCC 환자의 수술 전 혈장 및 조직 표본의 메틸화 밀도를 나타내는 표 2100이다. 관심 영역(예로 CpG 부위, 프로모터, 또는 반복 영역 등)에 대한 CpG 메틸화 밀도는 계놈 CpG 디뉴클레오타이드를 커버하는 측정치의 총 수에 걸쳐 CpG 메틸화를 나타내는 측정치의 비율을 나타낸다. 버피 코트 및 비종양 간 조직의 메틸화 밀도는 유사하다. 모든 상염색체에서의 데이터에 기반한 종양 조직의 전체적 메틸화 밀도는 버피 코트 및 비종양 간 조직에 대한 것보다 25% 더 낮았다. 저메틸화는 각각의 개별 염색체에 걸쳐 일치하였다. 혈장의 메틸화 밀도는 비악성 조직 및 암 조직의 값 사이에 있었다. 상기 관찰은 암 및 비암 조직이 모두 암 환자의 순환 DNA에 기여할

것이라는 점과 일치한다. 조혈계는 활성 악성 상태가 없는 개인에서 순환 DNA의 주요 원천임이 나타났다(YYN Lui, 등, 2002 Clin Chem; 48: 421-7). 따라서 4명의 건강한 대조군에서 수득한 혈장 표본을 또한 분석하였다. 표본 별로 획득한 서열 리드의 수 및 서열분석 깊이를 도 21b의 표 2150에 나타낸다.

[0278] 도 22는 건강한 대조군의 혈장 표본 중 71.2% 내지 72.5% 범위의 상염색체에서의 메틸화 밀도를 나타내는 표 220이다. 이들 데이터는 종양 DNA 원천이 없는 개인에서 수득한 혈장 표본 중 DNA 메틸화의 예측 수준을 나타내었다. 암 환자에서, 종양-조직은 또한 DNA를 순환 내로 배출할 것이다(KCA Chan 등, 2013 Clin Chem; 59: 211-224); RJ Leary 등, 2012 Sci Transl Med; 4: 162ra154). HCC 종양의 저메틸화 성질로 인해, 환자의 수술 전 혈장 중 종양- 및 비-종양-유래 DNA 모두의 존재가 건강한 대조군의 혈장 수준과 비교 시 메틸화 밀도 감소를 일으켰다. 사실 상, 수술 전 혈장 표본의 메틸화 밀도는 종양 조직 및 건강한 대조군의 혈장의 메틸화 밀도 사이에 있었다. 그 이유는 암 환자의 혈장 DNA의 메틸화 수준이 순환에서 종양-유래 DNA의 분획 농도 및 종양 조직의 이상 메틸화, 이 경우에는 저메틸화 정도에 영향을 받을 것이기 때문이다. 종양 조직의 더 낮은 메틸화 밀도 및 순환 중 종양-유래 DNA의 더 높은 분획 농도는 암 환자에서 혈장 DNA의 더 낮은 메틸화 밀도로 이어질 것이다. 대부분의 종양은 전반적 저메틸화를 나타내는 것으로 보고되었다(JG Herman 등, 2003 N Engl J Med; 349: 2042-2054; MA Gama-Sosa 등, 1983 Nucleic Acids Res; 11: 6883-6894). 따라서 HCC 표본에서 나타난 본 관찰도 다른 유형의 종양에도 적용 가능할 것이다.

[0279] 하나의 구현예에서, 종양 조직의 메틸화 수준을 아는 경우, 혈장 DNA의 메틸화 밀도를 이용하여 혈장/혈청 표본 중 종양-유래 DNA의 분획 농도를 결정할 수 있다. 종양 조직의 메틸화 수준, 예로 메틸화 밀도는 종양 표본을 이용할 수 있거나 종양 생검을 이용할 수 있는 경우 수득될 수 있다. 다른 구현예에서, 종양 조직의 메틸화 수준에 대한 정보는 유사한 유형의 종양 그룹에서 메틸화 수준 조사로 수득될 수 있고, 상기 정보(예로 평균 수준 또는 중앙값 수준)는 본 발명에 기재된 기술을 이용하여 분석될 환자에게 적용된다. 종양 조직의 메틸화 수준은 환자의 종양 조직 분석에 의해 결정되거나, 동일하거나 유사한 암 유형을 갖는 다른 환자의 종양 조직 분석에서 도출될 수 있다. 종양 조직의 메틸화는 비제한적으로 대량 병렬 서열분석, 단일 분자 서열분석, 마이크로어레이(예로 올리고뉴클레오타이드 어레이), 또는 질량 분광측정(예컨대 Epityper, Sequenom, Inc., 분석)을 포함하는 다양한 메틸화 인식 플랫폼을 이용하여 결정될 수 있다. 일부 구현예에서, 이러한 분석에는 비제한적으로 시토신 변형침전 및 메틸화 인식 제한효소 분해를 포함하는 DNA 분자의 메틸화 상태에 민감한 절차가 선행될 수 있다. 종양의 메틸화 수준을 아는 경우, 암 환자의 혈장 중 종양 DNA의 분획 농도가 혈장 메틸롬 분석 후 계산될 수 있다.

[0280] 혈장 메틸화 수준, P와 분획 종양 DNA 농도, f , 및 종양 조직 메틸화 수준, TUM 간 관계는 다음과 같이 기재될

$$P=BKG \times (1-f)+TUM \times f,$$

수 있다. 식 중, BKG는 혈액 세포 유래 혈장 및 다른 내부 기관 중 배경 DNA 메틸화 수준이다. 예를 들어, 상기 HCC 환자에서 수득한 종양 생검 조직 중 모든 상염색체의 전체적 메틸화 밀도, 즉 이 경우 TUM값은 42.9%로 나타났다. 4명의 건강한 대조군으로부터의 혈장 표본의 평균 메틸화 밀도, 즉 이 경우 BKG값은 71.6%였다. 수술 전 혈장에 대한 혈장 메틸화 밀도는 59.7%였다. 이들 값을 이용해서, f 는 41.5%로 산정된다.

[0281] 다른 구현예에서, 종양 조직의 메틸화 수준은 혈장 표본에서의 종양-유래 DNA의 분획 농도를 아는 경우, 혈장 메틸롬 데이터에 기반하여 비침습적으로 산정될 수 있다. 혈장 표본 중 종양-유래 DNA의 분획 농도는 다른 유전적 분석, 예를 들어 전술된 바와 같이 대립유전자 상실에 대한 게놈에 걸친 분석(GAAL) 및 단일 뉴클레오타이드 돌연변이 분석(US 특허 출원 13/308,473; KCA Chan 등, 2013 Clin Chem; 59: 211-24)에 의해 결정될 수 있다. 계산은 본 구현예에서 f 값이 알려져 있고 TUM값을 알 수 없다는 것을 제외하고는 상술된 바와 동일한 관계에 기반한다. 추정치는 모체 혈장 데이터에서 태반 조직 메틸화 수준 결정 맥락에 대해 관찰되는 데이터와 유사하게, 전체 게놈에 대해 또는 게놈의 일부에 대해 수행될 수 있다.

[0282] 다른 구현예에서, 메틸화 밀도에서의 bin간 변동 또는 프로필을 이용하여 암이 있는 및 암이 없는 대상자를 구별할 수 있다. 메틸화 분석의 해상도는 게놈을 특정 크기, 예로 1Mb의 bin으로 나누어 추가 증가될 수 있다. 이러한 구현예에서, 각각의 1Mb bin의 메틸화 밀도가 수집된 표본, 예로 버피 코트, 절제 HCC 조직, 종양에 인접한 비종양 간 조직 및 종양 절제 이전 및 이후에 수집된 혈장에 대해 계산되었다. 다른 구현예에서, bin 크기가 일정하게 유지될 필요는 없다. 하나의 시행에서, CpG 부위의 수는 각각의 bin 내에서 일정하게 유지되는 반면 bin 자체는 크기가 변할 수 있다.

[0283] 도 23a 및 23b는 HCC 환자의 버피 코트, 종양 조직, 비종양 간 조직, 수술 전 혈장 및 수술 후 혈장의 메틸화

밀도를 나타낸다. 도 23a는 염색체 1에 대한 결과 그래프 2300이다. 도 23b는 염색체 2에 대한 결과 그래프 2350이다.

- [0284] 대부분의 1Mb 윈도우에 있어서, 버피 코트 및 종양에 인접한 비종양 간 조직에 대한 메틸화 밀도는 유사한 반면 종양 조직의 밀도는 더 낮았다. 수술 전 혈장의 메틸화 밀도는 종양 및 비악성 조직의 사이에 놓인다. 종양 조직에서 조사 게놈 영역의 메틸화 밀도는 수술 전 혈장 및 분획 종양 DNA 농도의 메틸화 데이터를 이용해서 추정될 수 있다. 방법은 모든 상염색체의 메틸화 밀도값을 이용하여 전술된 바와 같이 동일하다. 기재된 종양 메틸화의 추정은 혈장 DNA의 이러한 더 높은 해상의 메틸화 데이터를 이용해서 수행될 수 있다. 다른 bin 크기, 예컨대 300kb, 500kb, 2Mb, 3Mb, 5Mb 또는 5Mb 초과도 이용될 수 있다. 하나의 구현예에서, bin 크기가 일정하게 유지될 필요는 없다. 하나의 시행에서, CpG 부위의 수는 각각의 bin 내에서 일정하게 유지되는 반면 bin 자체는 크기가 변할 수 있다.
- [0285] C. 암 환자 및 건강한 개인 간 혈장 메틸화 밀도의 비교
- [0286] 2100에 나타난 바와 같이, 수술 전 혈장 DNA의 메틸화 밀도는 암 환자에서 비악성 조직의 밀도보다 더 낮았다. 이는 저메틸화된 종양 조직 유래 DNA의 존재에 기인할 수 있다. 이러한 더 낮은 혈장 DNA 메틸화 밀도는 잠재적으로 암 검출 및 모니터링을 위한 바이오마커로 이용될 수 있다. 암 모니터링을 위해, 암이 진행되고 있는 경우, 혈장 중 암-유래 DNA량의 경시적 증가가 있을 것이다. 이러한 예에서, 혈장 중 순환 암-유래 DNA의 증가된 양은 게놈에 걸친 수준에서 혈장 DNA 메틸화 밀도의 추가 감소로 이어질 것이다.
- [0287] 반대로 암이 치료에 반응하는 경우, 혈장 중 암-유래 DNA의 양은 경시적으로 감소할 것이다. 이러한 예에서, 혈장 중, 암-유래 DNA량의 감소는 혈장 DNA 메틸화 밀도의 증가로 이어질 것이다. 예를 들어, 표피 성장 인자 수용체 돌연변이를 갖는 폐암 환자가 표적화된 치료법, 예로 티로신 키나아제 저해로 치료받은 경우, 혈장 DNA 메틸화 밀도 증가는 반응을 나타낼 것이다. 이어서 티로신 키나아제 저해에 대해 내성이 있는 종양 클론의 출현은 재발을 시사할 혈장 DNA 메틸화 밀도의 감소와 연관될 것이다.
- [0288] 혈장 메틸화 밀도 측정은 연속 수행될 수 있고, 이러한 측정의 변화율이 계산되고 임상적 진행 또는 완화 또는 예후를 예측하거나 연관시키기 위해 이용될 수 있다. 암 조직에서 고메틸화되지만 정상 조직에서 저메틸화된 선택된 게놈 유전자좌, 예로 여러 종양 억제인자 유전자좌의 프로모터 영역에 있어서, 암 진행 및 치료에 대해 바람직한 반응 간의 관계는 전술된 패턴에 반대일 것이다.
- [0289] 상기 접근의 타당성을 나타내기 위해, 종양의 수술적 제거 이전 및 이후 암 환자에서 수집된 혈장 표본의 DNA 메틸화 밀도를 4명의 건강한 대조군 대상자에서 수득한 혈장 DNA와 비교하였다.
- [0290] 표 2200은 각각의 상염색체의 DNA 메틸화 밀도 및 암 환자의 수술 전 및 수술 후 혈장 표본의 모든 상염색체 및 4명의 건강한 대조군 대상자에서의 조합값을 나타낸다. 모든 염색체에 있어서, 수술 전 혈장 DNA 표본의 메틸화 밀도는 수술 후 표본 및 4명의 건강한 대상자로부터의 혈장 표본에서보다 낮았다. 수술 전 및 수술 후 표본 간 혈장 DNA 메틸화 밀도의 차이는 수술 전 혈장 표본에서의 더 낮은 메틸화 밀도가 HCC 종양 유래 DNA 존재에 기인하였음을 지지하는 증거를 제공하였다.
- [0291] 수술 후 혈장 표본 중 DNA 메틸화 밀도의 건강한 대조군의 혈장 표본과 유사한 수준으로의 역전은 다수의 종양-유래 DNA가 원천, 즉 종양의 수술적 제거로 인해 사라졌음을 제시하였다. 이들 데이터는 큰 게놈 영역, 예컨대 모든 상염색체 또는 개별 염색체에서 이용 가능한 데이터를 이용해서 결정되는 수술 전 혈장의 메틸화 밀도가 건강한 대조군에서보다 낮은 메틸화 수준을 가져서 암을 갖는 평가 케이스의 확인, 즉 진단 또는 스크리닝을 허용함을 제시한다.
- [0292] 수술 전 혈장의 데이터는 또한 수술 후 혈장에서보다 훨씬 더 낮은 메틸화 수준을 나타내어 혈장 메틸화 수준이 또한 종양 부하의 모니터링에 이용될 수 있고, 이에 따라 환자에서 암의 진행을 예진하고 모니터링할 수 있음을 제시한다. 기준값은 건강한 대조군 또는 암에 대한 위험이 있지만 현재 암이 없는 개인의 혈장에서 결정될 수 있다. HCC에 대한 위험이 있는 개인에는 만성 B형 간염 또는 만성 C형 간염이 감염된 환자, 혈색소증 환자 및 간 경화 환자가 포함된다.
- [0293] 기준값에 기반한 정의된 컷오프를 넘어서는, 예를 들어 그보다 낮은 혈장 메틸화 밀도값을 이용해서 비임신 개인의 혈장이 종양 DNA를 갖는지의 여부를 결정할 수 있다. 저메틸화 순환 종양 DNA의 존재를 검출하기 위해, 컷오프는 대조군 모집단값의 제5 또는 제1 퍼센타일보다 낮은 것으로, 또는 표준 편차의 수, 예를 들어 대조군의 평균 메틸화 밀도값의 2 또는 3 표준 편차(SD) 미만에 기반하여, 또는 다수의 중앙값(MoM) 결정에 기반하여 정의될 수 있다. 고메틸화 종양 DNA에 있어서, 컷오프는 대조군 모집단값의 제95 또는 제99 퍼센타일보다 높은 것

으로, 또는 표준 편차의 수, 예를 들어 대조군의 평균 메틸화 밀도값의 2 또는 3 표준 편차(SD) 초과에 기반하여, 또는 다수의 중앙값(MoM) 결정에 기반하여 정의될 수 있다. 하나의 구현예에서, 대조군 모집단은 평가 대상자에 대해 연령 매치된다. 연령 매치가 정확할 필요는 없고, 연령 구간(예로 35세 평가 대상자에 대해 30 내지 40세)로 수행될 수 있다.

[0294] 다음으로 암 환자 및 4명의 대조군 대상자의 혈장 표본 간 1Mb bin의 메틸화 밀도를 비교하였다. 예시적 목적으로 염색체 1의 결과를 나타낸다.

[0295] 도 24a는 HCC 환자의 수술 전 혈장의 메틸화 밀도를 나타내는 그래프 2400이다. 도 24b는 HCC 환자의 수술 후 혈장의 메틸화 밀도를 나타내는 그래프 2450이다. 파란색 점은 대조군 대상자의 결과를 나타내고, 빨간색 점은 HCC 환자의 혈장 표본의 결과를 나타낸다.

[0296] 도 24a에 나타난 바와 같이, HCC 환자의 수술 전 혈장의 메틸화 밀도는 대부분의 bin에 대해 대조군 대상자에서 보다 낮았다. 유사한 패턴이 다른 염색체에 대해 관찰되었다. 도 24b에 나타난 바와 같이, HCC 환자의 수술 후 혈장의 메틸화 밀도는 대부분의 bin에 대해 대조군 대상자와 유사하였다. 유사한 패턴이 다른 염색체에 대해 관찰되었다.

[0297] 평가 대상자가 암을 갖는지를 결정하기 위해, 평가 대상자의 결과가 기준 군의 값과 비교될 것이다. 하나의 구현예에서, 기준 군은 여러 건강한 대상자를 포함할 수 있다. 다른 구현예에서, 기준 군은 비-악성 상태, 예를 들어, 만성 B형 간염 감염 또는 경화가 있는 대상자를 포함할 수 있다. 이어서 평가 대상자 및 기준 군 간 메틸화 밀도 차이가 정량될 수 있다.

[0298] 하나의 구현예에서, 기준 범위는 대조군의 값에서 유도될 수 있다. 이어서 기준 군의 상한 또는 하한으로부터 평가 대상자 결과의 일탈을 이용하여 대상자가 종양을 갖는지를 결정할 수 있다. 상기 양은 혈장 중 종양-유래 DNA의 분획 농도 및 악성 및 비-악성 조직 간 메틸화 수준 차이에 의해 영향을 받을 것이다. 혈장 중 더 높은 분획 농도의 종양-유래 DNA는 평가 혈장 표본 및 대조군 간에 더 큰 메틸화 밀도 차이로 이어질 것이다. 악성 및 비-악성 조직의 더 큰 정도의 메틸화 수준 차이도 평가 혈장 표본 및 대조군 간의 더 큰 메틸화 밀도 차이와 연관된다. 다른 구현예에서, 상이한 기준 군이 상이한 연령 범위의 평가 대상자에 대해 선택된다.

[0299] 다른 구현예에서, 4명의 대조군 대상자의 메틸화 밀도의 평균 및 SD가 각각의 1Mb bin에 대해 계산되었다. 이어서 대응 bin에 대해, HCC 환자의 메틸화 밀도 및 대조군 대상자의 평균값 간 차이가 계산되었다. 하나의 구현예에서, 상기 차이가 대응 bin의 SD로 나뉘어 z-스코어를 결정하였다. 다시 말하면, z-스코어는 대조군 대상자의 평균으로부터 SD의 수로 표현되는 평가 및 대조군 혈장 표본 간의 메틸화 밀도 차이를 나타낸다. 3을 초과하는 z-스코어의 bin은 HCC 환자의 혈장 DNA가 대조군 대상자에 비해 그 bin에서 3SD를 초과해서 더 고메틸화됨을 시사하는 반면, -3 미만의 z-스코어의 bin은 HCC 환자의 혈장 DNA가 대조군 대상자에 비해 그 bin에서 3SD를 초과해서 더 저메틸화됨을 시사한다.

[0300] 도 25a 및 25b는 염색체 1에 대한 기준으로 4명의 건강한 대조군 대상자의 혈장 메틸롬 데이터를 이용해서 HCC 환자의 수술 전(그래프 2500) 및 수술 후(그래프 2550) 혈장 표본에 대한 혈장 DNA 메틸화 밀도의 z-스코어를 나타낸다. 각각의 점은 하나의 1Mb bin의 결과를 나타낸다. 검은 점은 -3 내지 3의 z-스코어를 갖는 bin을 나타낸다. 빨간색 점은 -3 미만의 z-스코어를 갖는 bin을 나타낸다.

[0301] 도 26a는 수술 전 및 수술 후 혈장에 대한 z-스코어에 대한 데이터를 나타내는 표 2600이다. 수술 전 혈장 표본에서 염색체 1의 대부분의 bin(80.9%)은 -3 미만의 z-스코어를 가져서 HCC 환자의 수술 전 혈장 DNA가 대조군 대상자에 비해 유의미하게 더 저메틸화됨을 시사하였다. 반대로 빨간색 점의 수는 수술 후 혈장 표본에서 실질적으로 감소되어(염색체 1 상의 bin의 8.3%) 대부분의 종양 DNA가 순환 종양 DNA 원천의 수술적 절제로 인해 순환에서 제거되었음을 제시하였다.

[0302] 도 26b는 모든 상염색체에서 분석된 1Mb bin에 대한 기준으로 4명의 건강한 대조군 대상자를 이용해서 HCC 환자의 수술 전 및 수술 후 혈장 표본에 대한 혈장 DNA 메틸화 밀도의 z-스코어를 나타내는 Circos 그래프 2620이다. 최외부 고리는 인간 상염색체의 핵형도를 나타낸다. 가운데 고리는 수술 전 혈장 표본에 대한 데이터를 나타낸다. 최내부 고리는 수술 후 혈장 표본에 대한 데이터를 나타낸다. 각각의 점은 하나의 1Mb bin의 결과를 나타낸다. 검은 점은 -3 내지 3의 z-스코어를 갖는 bin을 나타낸다. 빨간색 점은 -3 미만의 z-스코어를 갖는 bin을 나타낸다. 녹색 점은 3 초과 z-스코어를 갖는 bin을 나타낸다.

[0303] 도 26c는 HCC 환자의 수술 전 및 수술 후 혈장 표본 모두에서 전체 게놈에 대한 1Mb bin의 z-스코어 분포를 나타내는 표 2640이다. 결과는 HCC 환자의 수술 전 혈장 DNA가 전체 게놈에서 대부분의 영역(1Mb bin의 85.2%)에

대해 대조군에 비해 더 저메틸화되었음을 시사한다. 반대로, 수술 후 혈장 표본에서 대부분의 영역(1Mb bin의 93.5%)은 대조군에 비해 유의미한 고메틸화 또는 저메틸화를 나타내지 않았다. 이들 데이터는 상기 HCC에 대해서 성질 상 주로 저메틸화된 다수의 종양 DNA가 더 이상 수술 후 혈장 표본에 존재하지 않았음을 시사한다.

[0304] 하나의 구현예에서, -3 미만의 z-스코어를 갖는 bin의 수, 백분율 또는 비율을 이용하여 암이 존재하는지를 나타낼 수 있다. 예를 들어 표 2640에 나타낸 바와 같이, 분석된 2734개의 bin 중 2330개(85.2%)가 수술 전 혈장에서 -3 미만의 z-스코어를 나타낸 반면, 분석된 2734개의 bins 중 171개(6.3%)만이 수술 후 혈장에서 -3 미만의 z-스코어를 나타냈다. 데이터는 수술 전 혈장에서의 종양 DNA 부하가 수술 후 혈장에서보다 훨씬 더 높음을 시사하였다.

[0305] bin 수의 컷오프값은 통계적 방법을 이용해서 결정될 수 있다. 예를 들어, 대략 0.15%의 bin이 정규 분포에 기반하여 -3 미만의 z-스코어를 가질 것으로 예상될 것이다. 따라서 bin의 컷오프 수는 분석되는 bin의 총 수의 0.15%일 수 있다. 다시 말하면, 비임신 개인의 혈장 표본이 -3 미만의 z-스코어를 갖는 0.15% 초과 bin을 나타내는 경우, 혈장 중 저메틸화 DNA의 원천, 즉 암이 존재한다. 예를 들어, 상기 예에서 분석한 2734개의 1Mb bin의 0.15%는 약 4개 bin이다. 상기 값을 컷오프로 이용하면, 그 양이 수술 후 혈장 표본에 비해 수술 전 혈장 표본에서 훨씬 더 크지만, 수술 전 및 수술 후 혈장 표본은 모두 저메틸화 종양-유래 DNA를 함유하였다. 4명의 건강한 대조군 대상자에 있어서, 어느 bin도 유의미한 고메틸화 또는 저메틸화를 나타내지 않았다. 다른 컷오프 값(예로, 1.1 %)이 이용될 수 있고, 이용되는 분석의 요건에 따라 변할 수 있다. 다른 예로서, 컷오프 백분율은 통계적 분포뿐만 아니라 요구되는 민감성 및 허용 가능한 특이성에 기반하여 변할 수 있다.

[0306] 다른 구현예에서, 컷오프 수는 암 환자 및 암이 없는 개인의 수를 분석하여 수신자 조작 특징(ROC) 곡선 분석으로 결정될 수 있다. 상기 접근의 특이성을 추가 검증하기 위해, 비-악성 상태에 대한 의학적 자문을 구하는 환자의 혈장 표본(C06)이 분석되었다. bin의 1.1%는 -3 미만의 z-스코어를 가졌다. 하나의 구현예에서, 상이한 역치를 이용하여 상이한 수준의 질환 상태를 분류할 수 있다. 더 낮은 백분율 역치를 이용해서 건강한 상태와 양성 상태를 구별하고, 더 높은 백분율 역치로 양성 상태와 악성을 구별할 수 있다.

[0307] 대량 병렬 서열분석을 이용한 혈장 저메틸화 분석을 위한 진단 수행은 특정 클래스의 반복 요소, 예로 장산재핵 요소-1(LINE-1)의 폴리머라아제 연쇄 반응(PCR) 기반 증폭을 이용하여 획득된 것보다 더 우수한 것으로 나타난다(P Tangkijvanich 등, 2007 Clin Chim Acta; 379:127-133). 상기 관찰에 대한 하나의 가능한 설명은 저메틸화가 종양 게놈에서 전반적인 반면, 하나의 게놈 영역에서 다른 게놈 영역 간 어느 정도의 이종성을 갖는다는 것이다.

[0308] 사실 상, 기준 대상자의 평균 혈장 메틸화 밀도가 게놈에 걸쳐 변하는 것을 관찰하였다(도 56). 도 56에서 각각의 빨간색 점은 32명의 건강한 대상자 중 하나의 1Mb bin의 평균 메틸화 밀도를 나타낸다. 그래프는 게놈에 걸쳐 분석되는 모든 1Mb bin을 나타낸다. 각각의 상자 내의 수는 염색체 번호를 나타낸다. 평균 메틸화 밀도가 bin에 따라 변화함을 관찰하였다.

[0309] 단순 PCR-기반 분석은 이러한 영역 별 이종성을 그 진단 알고리즘 내에서 고려할 수 없을 것이다. 이러한 이종성은 건강한 개인 간에 관찰되는 메틸화 밀도 범위를 확장시킬 것이다. 이어서 표본이 저메틸화를 나타내는 것으로 간주되기 위해 더 큰 크기의 메틸화 밀도 감소가 필요할 것이다. 이는 평가 민감성 감소를 일으킬 것이다.

[0310] 반대로, 대량 병렬 서열분석-기반 접근은 게놈을 1Mb bin(또는 다른 크기의 bin)으로 나누며 이러한 bin에 대한 메틸화 밀도를 개별적으로 측정한다. 각각의 영역이 평가 표본 및 대조군 간에 비교되므로 상기 접근은 상이한 게놈 영역에 걸쳐 베이스라인 메틸화 밀도의 변동 영향을 감소시킨다. 실제로 동일한 bin 내에서, 32명의 건강한 대조군에 걸친 개인 간 변동은 상대적으로 작았다. 95%의 bin은 32명의 건강한 대조군에 걸쳐 1.8% 이하의 변동 계수(CV)를 가졌다. 그러나 암 연관 저메틸화 검출을 위한 민감성을 추가 증강시키기 위해, 비교가 여러 게놈 영역에 걸쳐 수행될 수 있다. 하나의 영역만 평가될 때에는 암 표본이 특정 영역에 대해 저메틸화를 나타내지 않는 경우 생물학적 변동 효과에 대해 보장할 수 있으므로, 여러 게놈 영역을 평가하여 민감성이 증강될 것이다.

[0311] 대조군 및 평가 표본 간 대등한 게놈 영역의 메틸화 밀도를 비교(예로, 각각의 게놈 영역을 별도로 평가한 뒤 가능하면 이러한 결과를 조합)하고 상기 비교를 다수의 게놈 영역에 대해 수행하는 접근은 암에 연관된 저메틸화의 검출에 대해 더 높은 신호 대 노이즈비를 갖는다. 이러한 대량 병렬 서열분석 접근을 예로서 나타낸다. 다수의 게놈 영역의 메틸화 밀도를 결정하고 대조군 및 평가 표본 간 대응 영역의 메틸화 밀도 비교를 허용할 수 있는 다른 방법론은 유사한 효과를 달성할 것으로 예측될 것이다. 예를 들어, 특정 게놈 영역에서 유래된 혈장

DNA 분자를 표적화할 수 있을 뿐만 아니라 그 영역의 메틸화 수준을 결정할 수 있는 혼성화 프로브 또는 분자 역전 프로브를 설계하여 원하는 효과를 달성할 수 있다.

[0312] 다른 구현예에서, 모든 bin에 대한 z -스코어의 합을 이용하여 암이 존재하는지를 결정하거나 혈장 DNA 메틸화 수준의 일련의 변화를 모니터링하기 위해 이용할 수 있다. 종양 DNA의 전체적인 저메틸화 성질로 인해, z -스코어의 합은 건강한 대조군에 비해 암을 갖는 개인에서 수집된 혈장에서 더 낮을 것이다. HCC 환자의 수술 전 및 후 혈장 표본에 대한 z -스코어의 합은 각각 -49843.8 및 -3132.13이었다.

[0313] 다른 구현예에서, 다른 방법을 이용하여 혈장 DNA의 메틸화 수준을 조사할 수 있다. 예를 들어, 시토신 잔기의 총 함량에 대한 메틸화 시토신 잔기의 비율은 질량 분광측정(ML Chen 등, 2013 Clin Chem; 59: 824-832) 또는 대량 병렬 서열분석을 이용하여 결정될 수 있다. 그러나 대부분의 시토신 잔기가 CpG 디뉴클레오타이드 맥락에 있지 않으므로, 총 시토신 잔기 중 메틸화 시토신의 비율이 CpG 디뉴클레오타이드 맥락에서 산정된 메틸화 수준에 비해 상대적으로 작을 것이다. HCC 환자에서 획득된 조직 및 혈장 표본뿐만 아니라 건강한 대조군에서 획득한 4 개의 혈장 표본의 메틸화 수준을 결정하였다. 메틸화 수준은 게놈에 걸친 대량 병렬 서열분석 데이터를 이용하여 CpG 맥락에서, 임의의 시토신, CHG 및 CHH 맥락에서 측정되었다. H는 아데닌, 티민 또는 시토신 잔기를 나타낸다.

[0314] 도 26d는 CHH 및 CHG 맥락을 이용하는 경우 일부 대조군 혈장 표본과 중복되는 종양 조직 및 수술 전 혈장 표본의 메틸화 수준을 나타내는 표 2660이다. 종양 조직 및 수술 전 혈장 표본의 메틸화 수준은 CpG 및 특정되지 않은 시토신 모두에서 버피 코트, 비-종양 간 조직, 수술 후 혈장 표본 및 건강한 대조군 혈장 표본과 비교 시 일관되게 더 낮았다. 그러나 메틸화 CpG, 즉 메틸화 밀도에 기반한 데이터는 메틸화 시토신에 기반한 데이터에 비해 더 넓은 동적 범위를 나타내었다.

[0315] 다른 구현예에서, 혈장 DNA의 메틸화 상태는 메틸화 시토신에 대한 항체를 이용하는 방법, 예를 들어 메틸화 DNA 면역침전(MeDIP)에 의해 결정될 수 있다. 그러나 이들 방법의 정밀도는 항체 결합의 변동성으로 인해 서열 분석 기반 방법에 비해 열등한 것으로 예상된다. 다른 구현예에서, 혈장 DNA 중 5-히드록시메틸시토신의 수준이 결정될 수 있다. 이와 관련하여, 5-히드록시메틸시토신의 수준 감소는 특정 암, 예로 흑색종의 후생적 특성인 것으로 나타났다(CG Lian, 등, 2012 Cell; 150: 1135-1146).

[0316] HCC에 부가하여, 상기 접근이 다른 유형의 암에도 적용될 수 있는지를 조사하였다. 2명의 폐 선암종 환자(CL1 및 CL2), 2명의 코인두암종 환자(NPC1 및 NPC2), 2명의 결직장암 환자(CRC1 및 CRC2), 1명의 전이성 신경내분비 종양 환자(NE1) 및 1명의 전이성 평활근육종 환자(SMS1)로부터의 혈장 표본을 분석하였다. 이들 대상자의 혈장 DNA가 바이셀과이트-전환되고 한 말단에서 50bp 대에 Illumina HiSeq2000 플랫폼을 이용하여 서열분석되었다. 상기 언급된 4명의 건강한 대조군 대상자가 이들 8명의 환자 분석을 위한 기준 군으로 이용되었다. 한 말단에서의 서열 리드 50bp가 이용되었다. 전체 게놈은 1Mb bin으로 나뉘었다. 메틸화 밀도의 평균 및 SD가 기준 군의 데이터를 이용하여 각각의 bin에서 계산되었다. 이어서 8명의 암 환자의 결과가 기준 군의 평균으로부터의 SD 수로 나타내는 z -스코어로 표현되었다. 양의 값은 평가 케이스의 메틸화 밀도가 기준 군의 평균보다 낮다는 것을 시사하며, 그리고 음의 값은 그 반대를 시사한다. 표본 별로 달성되는 서열 리드의 수 및 서열분석 깊이를 도 27i의 표 2780에 나타낸다.

[0317] 도 27a-h는 본 발명의 구현예에 따른 8명의 암 환자의 메틸화 밀도의 Circos 그래프를 나타낸다. 각각의 점은 1Mb bin의 결과를 나타낸다. 검은색 점은 -3 내지 3의 z -스코어를 갖는 bin을 나타낸다. 빨간색 점은 -3 미만의 z -스코어를 갖는 bin을 나타낸다. 녹색 점은 3 초과인 z -스코어를 갖는 bin을 나타낸다. 두 연속선 사이의 간격은 20의 z -스코어 차이를 나타낸다.

[0318] 폐암, 코인두암종, 결직장암 및 전이성 신경내분비 종양을 포함하는 대부분의 암 유형의 환자에 대해 게놈에 걸쳐 다수의 영역에서 유의미한 저메틸화가 관찰되었다. 흥미롭게도, 저메틸화에 부가하여 전이성 평활근육종 케이스에서 게놈에 걸쳐 다수의 영역에서 유의미한 고메틸화가 관찰되었다. 평활근육종의 배아 기원은 중배엽인 반면, 나머지 7명 환자에서 다른 유형의 암의 배아 기원은 외배엽이다. 따라서 육종의 DNA 메틸화 패턴이 암종의 패턴과 상이할 수 있음이 가능하다.

[0319] 이 경우에 나타난 바와 같이, 혈장 DNA의 메틸화 패턴은 또한 상이한 암 유형의 구별, 이 예에서는 암종 및 육종의 구별을 위해 유용할 수 있다. 이들 데이터는 또한 접근을 이용하여 악성에 연관된 이상 고메틸화를 검출할 수 있음을 제시한다. 이들 전체 8 케이스에 있어서, 혈장 표본만 이용 가능하였고 종양 조직은 분석되지 않았다. 이는 이전의 메틸화 프로파일 또는 종양 조직의 메틸화 수준 없이도 종양-유래 DNA가 기재된 방법을 이용

하여 혈장에서 쉽게 검출될 수 있음을 나타내었다.

- [0320] 도 27j는 상이한 악성 환자의 혈장에서 전체 게놈에 대해 1Mb bin의 z-스코어 분포를 나타내는 표인 표 2790이다. 각 케이스에 대해 -3 미만, -3 내지 3 및 3 초과 z-스코어를 갖는 bin의 백분율을 나타낸다. 모든 케이스에 있어서, 5% 초과 bin이 -3 미만의 z-스코어를 가졌다. 따라서 표본을 암에 대해 양성으로 분류하기 위해 bin의 5% 컷오프를 유의미하게 저메틸화된 것으로 이용하는 경우, 이들 전체 케이스가 암에 대해 양성으로 분류될 것이다. 본 발명자들의 결과는 저메틸화가 상이한 유형의 암에 대해 일반적인 현상일 수 있으며, 혈장 메틸롬 분석이 상이한 유형의 암 검출을 위해 유용할 것임을 나타낸다.
- [0321] D. 방법
- [0322] 도 28은 본 발명의 구현예에 따른 암 수준의 분류를 결정하기 위해 유기체의 생물학적 표본을 분석하는 방법 2800의 순서도이다. 생물학적 표본에는 정상 세포에서 유래된 DNA가 포함되며, 잠재적으로 암에 연관된 세포 유래 DNA가 포함될 수 있다. 적어도 일부 DNA는 생물학적 표본에서 무세포일 수 있다.
- [0323] 블록 2810에서, 상기 생물학적 표본으로부터 복수의 DNA 분자가 분석된다. DNA 분자의 분석에는 유기체의 게놈에서의 DNA 분자 위치 결정 및 DNA 분자가 하나 이상의 부위에서 메틸화되어 있는지의 여부를 결정하는 것이 포함될 수 있다. 분석은 메틸화-인식 서열분석으로부터 서열 리드를 수신하여 수행될 수 있고, 이에 따라 분석은 DNA에서 이전에 획득된 데이터 상에서만 수행될 수 있다. 다른 구현예에서, 분석에는 실제 서열분석 또는 데이터를 획득하는 다른 유효 단계가 포함될 수 있다.
- [0324] 블록 2820에서, 부위에서 메틸화되어 있는 DNA 분자의 각각의 수가 각각의 복수의 부위에 대해 결정된다. 하나의 구현예에서, 부위는 CpG 부위이고 본원에 언급된 하나 이상의 기준을 이용하여 선택된 특정 CpG 부위만일 수 있다. 메틸화된 DNA 분자의 수는 특정 부위에서 분석되는 DNA 분자의 총 수, 예로 서열 리드의 총 수를 이용하여 정규화가 수행되는 경우 비메틸화되는 수의 결정과 대등하다. 예를 들어, 영역의 CpG 메틸화 밀도 증가는 동일한 영역의 비메틸화 CpG 밀도의 감소와 대등하다.
- [0325] 블록 2830에서, 제1 메틸화 수준은 상기 복수의 부위에서 상기 메틸화된 DNA 분자의 각각의 수에 기반하여 계산된다. 제1 메틸화 수준은 복수의 부위에 대응하는 DNA 분자의 수에 기반하여 결정되는 메틸화 밀도에 대응할 수 있다. 부위는 복수의 유전자좌 또는 단 하나의 유전자좌에 대응할 수 있다.
- [0326] 블록 2840에서, 제1 메틸화 수준이 제1 컷오프값과 비교된다. 제1 컷오프값은 기준 메틸화 수준일 수 있거나 기준 메틸화 수준에 관련될 수 있다(예로, 정상 수준으로부터 특정된 거리). 기준 메틸화 수준은 암이 없는 개인의 표본에서 또는 유기체의 암과 연관되지 않은 것으로 알려진 유전자좌 또는 유기체에서 결정될 수 있다. 제1 컷오프값은 평가되는 생물학적 표본에 대해 이전에 획득된 유기체의 이전 생물학적 표본에서 결정된 기준 메틸화 수준에서 정립될 수 있다.
- [0327] 하나의 구현예에서, 제1 컷오프값은 건강한 유기체에서 획득된 생물학적 표본에서 정립된 기준 메틸화 수준으로부터의 특정된 거리(예로 표준 편차의 특정된 수)이다. 비교는 제1 메틸화 수준 및 기준 메틸화 수준 간 차이를 결정한 뒤 그 차이를 제1 컷오프값에 대응하는 역치와 비교하여 수행될 수 있다(예로 메틸화 수준이 기준 메틸화 수준과 통계적으로 상이한지를 결정하기 위해).
- [0328] 블록 2850에서, 암의 수준 분류는 비교에 기반하여 결정된다. 암의 수준의 예에는 대상자가 암 또는 전악성 상태, 또는 증가된 암 발생 가능성을 갖는지의 여부가 포함된다. 하나의 구현예에서, 제1 컷오프값은 대상자에서 이전에 획득된 표본에서 결정될 수 있다(예로, 기준 메틸화 수준이 이전 표본에서 결정될 수 있다).
- [0329] 일부 구현예에서, 제1 메틸화 수준은 그 메틸화 수준이 역치값을 초과하는 여러 영역에 대응할 수 있다. 예를 들어, 유기체 게놈의 복수 영역이 확인될 수 있다. 영역은 본원에 언급된 기준, 예로 특정 길이 또는 특정 수의 부위를 이용해서 확인될 수 있다. 하나 이상의 부위(예로, CpG 부위)가 각각의 영역 내에서 확인될 수 있다. 영역의 메틸화 수준이 각 영역에 대해 계산될 수 있다. 제1 메틸화 수준은 제1 영역에 대한 것이다. 각각의 영역 메틸화 수준은 각각의 영역 컷오프값과 비교되며, 이는 영역 간에 동일하거나 변할 수 있다. 제1 영역에 대한 영역 컷오프값이 제1 컷오프값이다. 각각의 영역 컷오프값은 기준 메틸화 수준에서 특정된 양(예로 0.5)이고, 이에 따라 기준으로부터 유의미한 차이를 갖는 영역만을 카운트할 수 있으며, 이는 비-암 대상자에서 결정될 수 있다.
- [0330] 그 영역 메틸화 수준이 각각의 영역 컷오프값을 초과하는 영역의 제1 수가 결정되고 역치값과 비교되어 분류를 결정한다. 하나의 시행에서, 역치값은 백분율이다. 제1 수와 역치값의 비교에는, 예로 정규화 절차의 일환으로,

역치값과의 비교 전에 영역의 제1 수를 영역의 제2 수(예로, 전체 영역)으로 나누는 것이 포함될 수 있다.

[0331] 전술된 바와 같이, 상기 생물학적 표본 중 종양 DNA의 분획 농도를 이용하여 제1 컷오프값을 계산할 수 있다. 분획 농도는 단순히 최소값보다 크게 산정될 수 있는 반면, 최소값 미만의 분획 농도를 갖는 표본은, 예로 분석을 위해 적합하지 않은 것으로 표시될 수 있다. 최소값은 기준 메틸화 수준 대비 종양에 대한 메틸화 수준의 예상 차이에 기반하여 결정될 수 있다. 예를 들어, 차이가 0.5(예로, 컷오프값으로 사용되는 것)인 경우, 특정 종양 농도는 이러한 차이를 보기에 충분히 높도록 요구될 것이다.

[0332] 방법 1300으로부터의 특정 기법이 방법 2800을 위해 적용될 수 있다. 방법 1300에서, 복제수 변동이 종양에 대해 결정될 수 있다(예로, 종양의 제1 염색체 영역이 종양의 제2 염색체 영역에 비해 복제수 변화를 갖는지가 평가될 수 있는 경우). 즉, 방법 1300은 종양이 존재함을 추정할 수 있다. 방법 2800에서, 표본은 임의의 복제수 특징과 무관하게 결국 임의의 종양의 존재가 시사되는지에 대해 평가될 수 있다. 두 방법의 일부 기법은 유사할 수 있다. 그러나 방법 2800에 대한 컷오프값 및 메틸화 파라미터(예로, 정규화된 메틸화 수준)는 복제수 변동을 가질 수 있는 일부 영역을 포함하는 암 DNA 및 비-암 DNA의 혼합물에 대한 기준 메틸화 수준으로부터의 차이와 대조되어 비-암 DNA에 대한 기준 메틸화 수준으로부터의 통계적 차이를 검출할 수 있다. 따라서 방법 2800에 대한 기준값은, 암이 없는 표본, 예컨대 암이 없는 유기체 또는 동일한 환자의 비-암 조직(예로, 세포 DNA에서 결정될 수 있는, 이전에 채취된 혈장 또는 암을 갖지 않는 것으로 알려진 동시에 획득된 표본)에서 결정될 수 있다.

[0333] E. 혈장 DNA 메틸화 분석을 이용해서 검출되는 종양-DNA의 최소 분획 농도의 예측

[0334] 혈장 DNA의 메틸화 수준을 이용하여 암을 검출하기 위한 접근의 민감성을 측정하는 하나의 방식은 대조군과 비교되는 경우 혈장 DNA 메틸화 수준 변화를 드러내는데 필요한 최소 분획 종양-유래 DNA 농도에 관련된다. 평가 민감성은 또한 건강한 대조군 또는 혈액 세포 DNA에서의 베이스라인 혈장 DNA 메틸화 수준 및 종양 조직 간의 DNA 메틸화 차이 정도에 의존한다. 혈액 세포는 건강한 개인의 혈장에서 주요한 DNA 원천이다. 차이가 클수록, 암 환자를 비-암 개인과 구별하기 쉬울 수 있고, 더 높은 임상적 민감성 및 혈장 중 더 낮은 종양-유래 검출 한도로 암 환자 검출에 반영될 것이다. 또한, 상이한 연령의 대상자 또는 건강한 대상자에서 혈장 DNA 메틸화의 변동(G Hannum 등, 2013 Mol Cell; 49: 359-367)도 암의 존재에 연관된 메틸화 변화 검출의 민감성에 영향을 미칠 것이다. 건강한 대상자에서 혈장 DNA 메틸화의 변동이 적으면 소량의 암 유래 DNA 존재로 야기되는 변화의 검출이 더 쉬워질 것이다.

[0335] 도 29a는 상기 분포가 정규 분포를 따른다고 가정하는 기준 대상자에서 메틸화 밀도 분포를 나타내는 그래프 2900이다. 상기 분석은 하나의 메틸화 밀도값, 예를 들어 모든 상염색체 또는 특정 염색체의 메틸화 밀도만을 제공하는 각각의 혈장 표본에 기반을 둔다. 이는 분석의 특이성이 어떻게 영향을 받을지를 나타낸다. 하나의 구현예에서, 기준 대상자의 평균 DNA 메틸화 밀도의 3SD 미만의 컷오프를 이용하여 평가 표본이 기준 대상자의 표본보다 유의미하게 더 저메틸화되어 있는지가 결정된다. 상기 컷오프가 이용되는 경우, 비-암 대상자의 대략 0.15%가 암을 갖는 것으로 분류되는 위양성 결과를 가져서 99.85%의 특이성을 야기할 것으로 예상된다.

[0336] 도 29b는 기준 대상자 및 암 환자에서 메틸화 밀도의 분포를 나타내는 그래프 2950이다. 컷오프값은 기준 대상자의 메틸화 밀도 평균의 3SD 미만이다. 암 환자의 메틸화 밀도 평균이 컷오프값의 2SD 미만(즉 기준 대상자 평균의 5SD 미만)인 경우, 97.5%의 암 대상자는 컷오프값 미만의 메틸화 밀도를 가질 것으로 예상될 것이다. 다시 말하면, 하나의 메틸화 밀도값이 각각의 대상자에 대해 제공되는 경우, 예를 들어 전체 게놈, 모든 상염색체 또는 특정 염색체의 총 메틸화 밀도가 분석되는 경우, 예상되는 민감성은 97.5%일 것이다. 두 모집단의 평균 메틸화 밀도 간 차이는 두 요인, 즉 암 및 비-암 조직 간 메틸화 수준의 차이 정도 및 혈장 표본 중 종양-유래 DNA의 분획 농도에 의해 영향을 받는다. 이들 두 파라미터값이 높을수록, 이들 두 모집단의 메틸화 밀도값의 차이가 더 클 것이다. 또한, 두 모집단의 메틸화 밀도 분포의 SD가 낮을수록, 두 모집단의 메틸화 밀도 분포의 중복이 더 적다.

[0337] 여기서는 상기 개념을 나타내기 위해 가상의 예를 이용한다. 종양 조직의 메틸화 밀도가 약 0.45이고, 건강한 대상자의 혈장 DNA가 약 0.7이라고 가정해 보자. 이러한 가정값은 상염색체의 전체적 메틸화 밀도가 42.9%이고 건강한 대조군의 혈장 표본에 대한 상염색체의 평균 메틸화 밀도가 71.6%인 HCC 환자에서 수득된 것과 유사하다. 전체 게놈에 대해 혈장 DNA 메틸화 밀도 측정 CV가 1%라고 가정하면, 컷오프값은 $0.7 \times (100\% - 3 \times 1\%) = 0.679$ 일 것이다. 97.5%의 민감성을 확보하기 위해, 암 환자에 대한 혈장 DNA의 평균 메틸화 밀도는 대략 $0.679 - 0.7 \times (2 \times 1\%) = 0.665$ 여야 한다. f 가 혈장 표본 중 종양-유래 DNA의 분획 농도를 나타낸다고 하자. 그러면 f 는 $(0.7 - 0.45) \times f = 0.7 - 0.665$ 로 계산될 수 있다. 따라서 f 는 대략 14%이다. 상기 계산으로부터, 전체 게놈의 총

메틸화 농도가 진단 파라미터로 이용되는 경우, 97.5%의 진단 민감성을 달성하기 위해 혈장에서 검출될 수 있는 최소 분획 농도는 14%인 것으로 산정된다.

[0338] 다음으로 HCC 환자에서 수득한 데이터 상에서 상기 분석을 수행하였다. 이 예시를 위해, 각 표본에 대해 모든 상염색체에서 산정된 값에 기반하여 단 하나의 메틸화 밀도만 측정하였다. 건강한 대상자에서 수득된 혈장 표본 간 평균 메틸화 밀도는 71.6%였다. 이들 네 표본의 메틸화 밀도의 SD는 0.631%였다. 따라서 혈장 메틸화 밀도의 컷오프값은 -3 미만의 z-스코어 및 99.85%의 특이성에 도달하기 위해 $71.6\% - 3 \times 0.631\% = 69.7\%$ 여야 할 것이다. 97.5%의 민감성에 도달하기 위해, 암 환자의 평균 혈장 메틸화 밀도는 컷오프의 2SD 미만, 즉 68.4%여야 할 것이다. 종양 조직의 메틸화 밀도가 42.9%이므로, 식 $P = BKG \times (1-f) + TUM \times f$ 를 이용하면 f 는 적어도 11.1%여야 할 것이다.

[0339] 다른 구현예에서, 상이한 게놈 영역의 메틸화 밀도는, 예로 도 25a 또는 26b에 나타난 바와 같이 별도로 분석될 수 있다. 다시 말하면, 각 표본에 대해 메틸화 수준의 복수 측정이 수행될 수 있다. 아래 나타난 바와 같이, 유의미한 저메틸화는 혈장 중 훨씬 더 낮은 분획 종양 DNA 농도를 검출할 수 있고, 이에 따라 암 검출을 위한 혈장 DNA 메틸화 분석의 진단 성능이 증강될 것이다. 기준 모집단에서 메틸화 밀도의 유의미한 일탈을 나타내는 게놈 영역의 수가 카운트될 수 있다. 이어서 게놈 영역의 수를 컷오프값과 비교하여 조사되는 게놈 영역 모집단, 예를 들어 전체 게놈의 1Mb bin에 걸쳐 혈장 DNA의 전체적으로 유의미한 저메틸화가 존재하는지를 결정할 수 있다. 컷오프값은 암이 없는 기준 대상자군의 분석에 의해 정립되거나 또는 예를 들어 정규 분포 함수에 따라 수학적으로 유도될 수 있다.

[0340] 도 30은 건강한 대상자 및 암 환자의 혈장 DNA의 메틸화 밀도 분포를 나타내는 그래프 3000이다. 각각의 1Mb bin의 메틸화 밀도가 기준 군의 대응값과 비교된다. 유의미한 저메틸화(기준 군 평균의 3SD 미만)를 나타내는 bin의 백분율이 결정되었다. 유의미한 저메틸화인 10% 컷오프를 이용하여 종양-유래 DNA가 혈장 표본 중에 존재하는지를 결정하였다. 다른 컷오프값, 예컨대 5%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80% 또는 90%가 또한 원하는 평가 민감성 및 특이성에 따라 이용될 수 있다.

[0341] 예를 들어, 표본을 종양-유래 DNA를 함유하는 것으로 분류하기 위해, 유의미한 저메틸화(z-스코어 < -3)를 나타내는 1Mb bin의 10%를 컷오프로 이용할 수 있다. 기준 군보다 유의미하게 더 저메틸화된 bin이 10% 초과로 존재하는 경우, 표본은 암 평가에 대해 양성으로 분류된다. 각각의 1Mb bin에 대해, 기준 군의 평균 메틸화 밀도의 3SD 미만의 컷오프를 이용하여 표본이 유의미하게 더 저메틸화된 것으로 정의한다. 각각의 1Mb bin에 대해, 암 환자의 평균 혈장 DNA 메틸화 밀도가 기준 대상자의 평균 혈장 DNA 메틸화 밀도보다 1.72SD 미만인 경우, 암 환자의 임의의 특정 bin의 메틸화 밀도값이 컷오프(즉, z-스코어 < -3)보다 낮고 양성 결과를 나타낼 확률은 10%이다. 이어서 전체 게놈에 대해 모든 1Mb bin을 살펴보면, 대략 10%의 bin이 유의미하게 더 낮은 메틸화 밀도(즉 z-스코어 < -3)를 갖는 양성 결과를 나타낼 것으로 예상될 것이다. 건강한 대상자의 혈장 DNA의 전체적 메틸화 밀도가 대략 0.7이고 각각의 1Mb bin에 대한 혈장 DNA 메틸화 밀도 측정의 변동 계수(CV)가 1%라고 가정하면, 암 환자의 혈장 DNA의 평균 메틸화 밀도는 $0.7 \times (100\% - 1.72 \times 1\%) = 0.68796$ 이어야 할 것이다. 상기 평균 혈장 DNA 메틸화 밀도를 수득하기 위해, 혈장 중 종양-유래 DNA의 분획 농도를 f 라고 하자. 종양 조직의 메틸화 밀도가 0.45라고 가정하면, f 는 다음 식을 이용해서 계산될 수 있다:

$$(\overline{M}_{P_{ref}} - M_{umor}) \times f = \overline{M}_{P_{ref}} - \overline{M}_{P_{cancer}}$$

[0342] 식 중, $\overline{M}_{P_{ref}}$ 는 기준 개인에서 혈장 DNA의 평균 메틸화 밀도를 나타내며, M_{umor} 는 암 환자에서 종양 조직의 메틸화 밀도를 나타내고; $\overline{M}_{P_{cancer}}$ 는 암 환자에서 혈장 DNA의 평균 메틸화 밀도를 나타낸다.

[0344] 상기 식을 이용하면, $(0.7 - 0.45) \times f = 0.7 - 0.68796$ 이다. 따라서 상기 접근을 이용해서 검출될 수 있는 최소 분획 농도는 4.8%로 추정될 것이다. 민감성은 bin의 컷오프 백분율을 유의미하게 더 저메틸화되도록, 예를 들어 10%에서 5%로 감소시켜 추가 증강될 수 있다.

[0345] 상기 예에 나타난 바와 같이, 상기 방법의 민감성은 암 및 비-암 조직, 예를 들어 혈액 세포 간 메틸화 수준의 차이 정도로 결정된다. 하나의 구현예에서, 비-암 대상자의 혈장 DNA 및 종양 조직 간에 큰 메틸화 밀도 차이를 나타내는 염색체 영역만 선택된다. 하나의 구현예에서, 0.5 초과와 메틸화 밀도 차이를 갖는 영역만 선택된다. 다른 구현예에서, 0.4, 0.6, 0.7, 0.8 또는 0.9의 차이가 적합한 영역을 선택하기 위해 이용될 수 있다. 다른

구현예에서, 게놈 영역의 물리적 크기는 고정되지 않는다. 대신 게놈 영역은, 예를 들어 고정된 측정치 깊이 또는 고정된 수의 CpG 부위에 기반하여 정의된다. 복수의 이들 게놈 영역에서의 메틸화 수준이 각 표본에 대해 평가된다.

[0346] 도 31은 건강한 대상자의 혈장 DNA 및 HCC 환자의 종양 조직의 평균 간 메틸화 밀도 차이의 분포를 나타내는 그래프 3100이다. 양의 값은 메틸화 밀도가 건강한 대상자의 혈장 DNA에서 더 높음을 표시하고, 음의 값은 메틸화 밀도가 종양 조직에서 더 높음을 표시한다.

[0347] 하나의 구현예에서, 암 및 비-암 조직의 메틸화 밀도 간에 가장 큰 차이를 갖는 bin이, 예를 들어 종양이 이들 bin에 대해 저메틸화 또는 고메틸화되어 있는지와 무관하게 0.5 초과 차이를 갖는 것들이 선택될 수 있다. 혈장 중 종양-유래 DNA 분획 농도의 검출 한계는 혈장 중 종양-유래 DNA의 동일한 분획 농도가 주어진 암 및 비-암 대상자 간 혈장 DNA 메틸화 수준의 분포 간의 더 큰 차이로 인해 이들 bin에 초점을 맞춰 낮춰질 수 있다. 예를 들어, 평가된 개인이 암을 갖는지를 결정하기 위해 0.5 초과 차이를 갖는 bin만 이용되고 10%의 bin이 유의미하게 더 저메틸화된 컷오프가 채택되는 경우, 검출된 종양 유래 DNA의 최소 분획 농도(f)는 다음 공식을 이

용해서 계산될 수 있다: $(\overline{M}_{P_{ref}} - M_{tumor}) \times f = \overline{M}_{P_{ref}} - \overline{M}_{P_{cancer}}$, 식 중, $\overline{M}_{P_{ref}}$ 는 기준 개인에서 혈장 DNA의 평균 메틸화 밀도를 나타내며, M_{tumor} 는 암 환자에서 종양 조직의 메틸화 밀도를 나타내고; $\overline{M}_{P_{cancer}}$ 는 암 환자에서 혈장 DNA의 평균 메틸화 밀도를 나타낸다.

[0348] 기준 대상자의 혈장 및 종양 조직 간 메틸화 밀도 차이는 적어도 0.5이다. 그러면 $0.5 \times f = 0.7 - 0.68796$ 이고 $f = 2.4\%$ 이다. 따라서 암 및 비-암 조직 간 메틸화 밀도 차이가 더 높은 bin에 초점을 맞추므로써, 분획 종양-유래 DNA의 하한이 4.8%에서 2.4%로 낮춰질 수 있다. 어느 bin이 암 및 비-암 조직, 예를 들어 혈액 세포 간에 더 큰 정도의 메틸화 차이를 나타내는지에 대한 정보는 동일한 기관의 종양 조직 또는 다른 개인에서 수득된 동일한 조직 유형에서 결정될 수 있다.

[0349] 다른 구현예에서, 파라미터는 암 및 비-암 조직 간 메틸화 밀도 차이를 고려하여 모든 bin의 혈장 DNA의 메틸화 밀도에서 유도될 수 있다. 더 큰 차이를 갖는 bin에 더 큰 가중치가 제공될 수 있다. 하나의 구현예에서, 각각의 bin의 암 및 비-암 조직 간 메틸화 밀도 차이는 특정 bin이 최종 파라미터 계산에 있는 경우 가중치로 직접 이용될 수 있다.

[0350] 다른 구현예에서, 상이한 유형의 암은 종양 조직에서 상이한 메틸화 패턴을 가질 수 있다. 암-특이적 가중치 프로파일은 특정 유형의 암의 메틸화 정도에서 유도될 수 있다.

[0351] 다른 구현예에서, 메틸화 밀도의 bin-간 관계는 암이 있는 및 암이 없는 대상자에서 결정될 수 있다. 도 8에서, 작은 수의 bin에서 종양 조직이 기준 대상자의 혈장 DNA에 비해 더 메틸화되었음을 관찰할 수 있다. 따라서 가장 심한 차이값, 예로 0.5 초과 차이 및 0 미만의 차이를 갖는 bin이 선택될 수 있다. 이어서 이들 bin의 메틸화 밀도비를 이용하여 평가된 개인이 암을 갖는지를 나타낼 수 있다. 다른 구현예에서, 상이한 bin의 메틸화 밀도 차이 및 할당이 bin간 관계를 나타내기 위한 파라미터로 이용될 수 있다.

[0352] HCC 환자에서 수득된 데이터로 나타내는 바와 같이 다수 게놈 영역의 메틸화 밀도를 이용하여 종양을 검출하거나 평가하기 위한 접근의 검출 민감성을 추가로 평가하였다. 먼저, 수술 전 혈장으로부터의 측정치를 건강한 대조군의 혈장 표본에서 수득된 것과 혼합하여 20% 내지 0.5% 범위인 분획 종양 DNA 농도를 함유하는 혈장 표본을 시뮬레이션하였다. 이어서 z-스코어 <-3과 대등한 메틸화 밀도를 갖는 1Mb bin의 백분율을(전체 게놈에서의 2,734개 bin 중에) 스코어링하였다. 혈장 중 분획 종양 DNA 농도가 20%였다면, bin의 80.0%가 유의미한 저메틸화를 나타내었다. 10%, 5%, 2%, 1% 및 0.5%의 혈장 중 분획 종양 DNA 농도에 대한 대응 데이터는 각각 저메틸화를 나타내는 bin 67.6%, 49.7%, 18.9%, 3.8% 및 0.77%였다. 대조군 표본에서 z-스코어 <-3을 나타내는 bin 수의 이론적 한계가 0.15%이므로, 본 발명자들의 데이터는 종양 분획 농도가 단지 0.5%였던 경우에도 이론적 컷오프 한계를 넘어서는 훨씬 더 많은 bin(0.77%)이 존재하였음을 나타낸다.

[0353] 도 32a는 혈장 표본이 5% 또는 2%의 종양 DNA를 함유하는 경우, 서열분석 깊이를 감소시키는 효과를 나타내는 표 3200이다. 평균 서열분석 깊이가 홀배수체 게놈의 단지 0.022배였던 경우에도 유의미한 저메틸화를 나타내는 높은 비율(>0.15%)의 bin이 여전히 검출될 수 있었다.

[0354] 도 32b는 4명의 건강한 대조군 대상자의 혈장, 버피 코트, 정상 간 조직, 종양 조직, HCC 환자의 수술 전 혈장 및 수술 후 혈장 표본에서 반복 요소 및 비반복 영역의 메틸화 밀도를 나타내는 그래프 3250이다. 반복 요소가

암 및 비-암 조직 모두에서 비반복 영역에 비해 더 메틸화되었음을(더 높은 메틸화 밀도) 관찰할 수 있다. 그러나 반복 요소 및 비반복 영역 간 메틸화 밀도의 차이는 종양 조직과 비교 시 건강한 대상자의 혈장 DNA 및 비-암 조직에서 더 컸다.

[0355] 그 결과, 암 환자의 혈장 DNA는 비반복 영역에서보다 반복 영역에서 더 큰 메틸화 밀도 감소를 가졌다. 4명의 건강한 대조군 평균 및 HCC 환자 간 혈장 DNA 메틸화 밀도의 차이는 반복 요소 및 비반복 영역에 대해 각각 0.163 및 0.088이었다. 수술 전 및 수술 후 혈장 표본에 대한 데이터도 메틸화 밀도 변화에서의 동적 범위가 비반복 영역에 비해 반복 영역에서 더 큼을 나타내었다. 하나의 구현예에서, 반복 요소의 혈장 DNA 메틸화 밀도는 환자가 암에 의해 영향받는지를 결정하기 위해 또는 질환 진행을 모니터링하기 위해 이용될 수 있다.

[0356] 상기 논의된 바와 같이, 기준 대상자의 혈장 중 메틸화 밀도의 변동은 또한 암 환자와 비-암 개인의 구별의 정확성에 영향을 미칠 것이다. 메틸화 밀도의 분포가 조밀할수록(즉 표준 편차가 작을수록), 암 및 비-암 대상자를 더 정확히 구별하게 될 것이다. 다른 구현예에서, 1Mb bin의 메틸화 밀도의 변동 계수(CV)는 기준 군에서 혈장 DNA 메틸화 밀도의 낮은 변동성을 갖는 bin을 선택하기 위한 기준으로 이용될 수 있다. 예를 들어, 1% 미만의 CV를 갖는 bin만 선택된다. 다른 값, 예를 들어 0.5%, 0.75%, 1.25% 및 1.5%도 메틸화 밀도에서 낮은 변동성을 갖는 bin의 선택을 위한 기준으로 이용될 수 있다. 다른 구현예에서, 선택 기준에는 bin의 CV 및 암 및 비-암 조직 간 메틸화 밀도 차이 모두가 포함될 수 있다.

[0357] 종양 조직의 메틸화 밀도를 아는 경우, 메틸화 밀도를 또한 이용하여 혈장 표본 중 종양-유래 DNA의 분획 농도를 산정할 수 있다. 상기 정보는 동일한 유형의 암을 갖는 여러 환자의 종양 조사로부터 또는 환자 종양의 분석에 의해 수득될 수 있다. 상기 논의된 바와 같이, 혈장 메틸화 밀도(P)는 다음 공식을 이용하여 표현될 수

있다:
$$P = BKG \times (1 - f) + TUM \times f$$
, 식 중 BKG는 혈액 세포 및 다른 기관의 배경 메틸화 밀도이고, TUM은 종양 조직에서의 메틸화 밀도이고, f는 혈장 표본 중 종양-유래 DNA의 분획 농도이다. 이는 다음과

$$f = \frac{BKG - P}{BKG - TUM}.$$

같이 다시 쓰여질 수 있다:

[0358] BKG값은 암이 없는 개인 기준 군 조사에서 또는 암이 존재하지 않는 시점에 환자의 혈장 표본을 분석하여 결정될 수 있다. 따라서 혈장 메틸화 밀도의 측정 후, f가 결정될 수 있다.

[0359] F. 다른 방법과의 조합

[0360] 본원에 기재된 메틸화 분석 접근은 혈장 중 종양-유래 DNA의 유전적 변화에 기반한 다른 방법과 조합되어 이용될 수 있다. 이러한 방법의 예에는 암 연관 염색체 이상에 대한 분석(KCA Chan 등, 2013 Clin Chem; 59:211-224; RJ Leary 등, 2012 Sci Transl Med; 4:162ra154) 및 혈장 중의 암-연관 단일 뉴클레오티드 변동(KCA Chan 등, 2013 Clin Chem; 59:211-224)이 포함된다. 이러한 유전적 접근에 비해 메틸화 분석 접근의 장점이 존재한다.

[0361] 도 21a에 나타낸 바와 같이, 종양 DNA의 저메틸화는 거의 전체 게놈에 걸쳐 분포된 영역이 수반되는 전반적 현상이다. 따라서 모든 염색체 영역 유래 DNA 단편은 환자에서 종양-유래 저메틸화 DNA의 혈장/혈청 DNA에 대한 잠재적 분포에 관해 정보를 제공할 것이다. 대조적으로, 염색체 이상(염색체 영역의 증폭 또는 결실)은 일부 염색체 영역에만 존재하며, 종양 조직 중 염색체 이상이 없는 영역 유래 DNA 단편은 분석에서 정보를 제공하지 않을 것이다(KCA Chan 등, 2013 Clin Chem; 59: 211-224). 유사하게 각각의 암 게놈에서 수 천개의 단일 뉴클레오티드 변형만이 관찰된다(KCA Chan 등, 2013 Clin Chem; 59: 211-224). 이들 단일 뉴클레오티드 변화와 중복되지 않는 DNA 단편은 종양-유래 DNA가 혈장 중에 존재하는지를 결정하는데 있어 정보를 제공하지 않을 것이다. 따라서 상기 메틸화 분석 접근은 순환 중 암 연관 변화를 검출하기 위한 이러한 유전적 접근에 비해 잠재적으로 더 비용 효과적이다.

[0362] 하나의 구현예에서, 혈장 DNA 메틸화 분석의 비용 효율성은 가장 많은 정보를 제공하는 영역, 예를 들어 암 및 비-암 조직 간에 가장 높은 차등적 메틸화 차이를 갖는 영역으로부터 DNA 단편을 농축하여 추가 증강될 수 있다. 이러한 영역에 대한 농축 방법의 예에는 혼성화 프로브의 이용(예로 Nimblegen SeqCap 시스템 및 Agilent SureSelect 표적 농축 시스템), PCR 증폭 및 고상 혼성화가 포함된다.

[0363] G. 조직-특이적 분석/ 공여자

- [0364] 종양-유래 세포는 인접한 또는 먼 기관으로 침입하고 전이한다. 침입된 조직 또는 전이성 병터는 세포사 결과 혈장 내로의 DNA에 기여한다. 암 환자의 혈장 중 DNA의 메틸화 프로필을 분석하고 조직-특이적 메틸화 특징의 존재를 검출함으로써, 질환 절차에 수반된 조직 유형을 검출할 수 있다. 상기 접근은 일차 및 전이성 부위로 수반된 기관의 확인을 보조하기 위해 암성 절차에 수반된 조직의 비침습적 해부학적 스캔을 제공한다. 혈장 중 수반 기관의 메틸화 특징의 상대 농도 모니터링도 그 기관에서의 암 절차가 악화 중 또는 개선 중인지 또는 치유되었는지를 결정하고 그 기관의 종양 부하를 평가할 수 있도록 할 것이다. 예를 들어, 유전자 X가 간에서 특이적으로 메틸화된 경우, 암(예로 결직장암)에 의한 간의 전이성 수반은 혈장 중 유전자 X의 메틸화 서열 농도를 증가시킬 것으로 예상될 것이다. 또한 유전자 X와 유사한 메틸화 특징을 갖는 다른 서열 또는 서열 그룹이 있을 것이다. 이어서 이러한 서열로부터의 결과를 조합할 수 있다. 유사한 고려사항이 다른 조직, 예로 뇌, 뼈, 폐 및 신장 등에 적용 가능하다.
- [0365] 반면, 상이한 기관 유래 DNA는 조직-특이적 메틸화 특징을 나타내는 것으로 알려져 있다(BW Futscher 등, 2002 Nat Genet; 31:175-179; SSC Chim 등, 2008 Clin Chem; 54: 500-511). 따라서 혈장 중의 메틸화 프로필 확인이 다양한 기관으로부터 혈장 내로의 조직의 기여를 규명하기 위해 이용될 수 있다. 세포 사멸 시 혈장 DNA가 배출되는 것으로 여겨지므로, 이러한 기여의 규명은 기관 손상을 평가하기 위해 이용될 수 있다. 예를 들어, 간 병리, 예컨대 간염(예로 바이러스, 자가면역 절차 등에 의한) 또는 약물에 의해 야기되는 간 독성(예로 약물 과용량(예컨대 파라세타몰에 의한) 또는 독소(예컨대 알코올))은 간 세포 손상과 연관되며, 혈장 중 간-유래 DNA의 증가된 수준과 연관된 것으로 예상될 것이다. 예를 들어 유전자 X가 간에서 특이적으로 메틸화된 경우, 간 병리는 혈장 중 유전자 X의 메틸화 서열 농도를 증가시킬 것으로 예상될 것이다. 반대로 유전자 Y가 간에서 특이적으로 저메틸화된 경우, 간 병리는 혈장 중 유전자 Y의 메틸화 서열 농도를 감소시킬 것으로 예상될 것이다. 다른 구현예에서, 유전자 X 또는 Y는 유전자가 아닐 수 있고 체내에서 상이한 조직 중 차등적 메틸화를 나타내는 임의의 게놈 서열로 대체될 수 있다.
- [0366] 본원에 기재된 기법은 또한 기관 이식 수령자의 혈장 중 공여자 유래 DNA의 평가에 적용될 수 있다(YMD Lo 등, 1998 Lancet; 351:1329-1330). 공여자 및 수령자 간 다형성 차이를 이용하여 혈장 중 공여자 유래 DNA를 수령자 유래 DNA와 구별하였다(YW Zheng 등, 2012 Clin Chem; 58: 549-558). 본 발명자들은 이식 기관의 조직-특이적 메틸화 특징이 수령자의 혈장 중 공여자의 DNA를 검출하기 위한 방법으로도 이용될 수 있음을 제안한다.
- [0367] 공여자의 DNA 농도를 모니터링함으로써, 이식된 기관의 상태를 비침습적으로 평가할 수 있다. 예를 들어, 이식 거부는 더 높은 비의 세포사와 연관되며, 따라서 환자가 안정한 상태에 있는 때와 비교 시 또는 이식 없이 건강한 대조군 또는 다른 안정한 이식 수령자와 비교 시 이식 기관의 메틸화 특징으로 반영되는 수령자 혈장(또는 혈청) 중 공여자의 DNA 농도가 증가될 것이다. 암에 대해 기재된 바와 유사하게, 공여자-유래 DNA는 다형성 차이, 이식 고형 기관에 대해 더 짧은 크기의 DNA(YW Zheng 등, 2012 Clin Chem; 58: 549-558) 및 조직-특이적 메틸화 프로필을 포함하는 전체 또는 일부 특징적 특성에 대한 검출에 의해 이식 수령자의 혈장 중에서 확인될 수 있다.
- [0368] *H. 크기에 기반한 메틸화의 정규화*
- [0369] 전술되고 Lun 등(FMF Lun 등, Clin. Chem. 2013; doi:10.1373/clinchem.2013.212274)에 의해 기재된 바와 같이, 메틸화 밀도(예로, 혈장 DNA의 메틸화 밀도)는 DNA 단편의 크기와 연관된다. 더 짧은 혈장 DNA 단편에 대한 메틸화 밀도의 분포는 더 긴 단편에 비해 유의미하게 더 낮았다. 혈장 DNA의 비정상적 단편화 패턴을 갖는 일부 비-암 상태(예로, 전신 홍반 루푸스(SLE))는 덜 메틸화된 보다 풍부한 짧은 혈장 DNA 단편의 존재로 인해 혈장 DNA의 현저한 저메틸화를 나타낼 수 있음을 제안한다. 다시 말하면, 혈장 DNA의 크기 분포는 혈장 DNA에 있어서 메틸화 밀도에 대한 교란 교인일 수 있다.
- [0370] 도 34a는 SLE 환자 SLE04에서 혈장 DNA의 크기 분포를 나타낸다. 9명의 건강한 대조군 대상자의 크기 분포는 회색 점선으로 나타내며, SLE04에 대한 것은 검은색 실선으로 나타낸다. 짧은 혈장 DNA 단편이 9명의 건강한 대조군 대상자에 비해 SLE04에서 더 풍부하였다. 더 짧은 DNA 단편은 일반적으로 덜 메틸화되므로, 상기 크기 분포 패턴은 혈장 DNA에 대한 메틸화 분석을 교란시킬 수 있고, 보다 명백한 저메틸화로 이어질 수 있다.
- [0371] 일부 구현예에서, 측정된 메틸화 수준을 정규화하여 혈장 DNA 메틸화 분석에 대한 크기 분포의 교란 효과를 감소시킬 수 있다. 예를 들어, 복수의 부위에서의 DNA 분자의 크기가 측정될 수 있다. 다양한 시행에서, 측정은 DNA 분자에 특정 크기(예로 길이)를 제공하거나 또는 단순히 크기가 특정 범위 내에 속하는지를 결정할 수 있고, 이것도 크기에 대응할 수 있다. 이어서 정규화된 메틸화 수준이 컷오프값과 비교될 수 있다. DNA 메틸화

분석에 있어서 크기 분포의 교란 효과를 감소시키기 위해 정규화를 수행하기 위한 몇몇 방식이 존재한다.

- [0372] 하나의 구현예에서, DNA(예로, 혈장 DNA)의 크기 분획화가 수행될 수 있다. 크기 분획화는 유사한 크기의 DNA 단편을 이용하여 컷오프값과 일치하는 방식으로 메틸화 수준을 결정하도록 할 수 있다. 크기 분획화의 일환으로, 제1 크기(예로, 제1 길이 범위)를 갖는 DNA 단편이 선택될 수 있고, 상기 제1 컷오프값은 제1 크기에 대응한다. 정규화는 선택된 DNA 단편만을 이용하여 메틸화 수준을 계산함으로써 달성될 수 있다.
- [0373] 크기 분획화는 다양한 방식, 예로 상이한 크기의 DNA 분자의 물리적 분리에 의해(예로 전기영동 또는 마이크로 유체 기반 기술 또는 원심분리 기반 기술에 의해) 또는 컴퓨터내 분석에 의해 달성될 수 있다. 컴퓨터내 분석에 있어서, 하나의 구현예에서 혈장 DNA 분자의 페어-엔드 대량 병렬 서열분석을 수행할 수 있다. 이어서 혈장 DNA 분자의 각각의 두 말단의 위치를 기준 인간 게놈과 비교하여 서열분석된 분자의 크기를 추정할 수 있다. 그 뒤 하나 이상의 크기 선택 기준(예로 특정된 범위 내인 크기 기준)에 매치되는 서열분석된 DNA 분자의 선택에 의해 후속 분석을 수행할 수 있다. 따라서 하나의 구현예에서, 메틸화 밀도가 유사한 크기를 갖는(예로 특정된 범위 내인) 단편에 대해 분석될 수 있다. 컷오프값(예로, 방법 2800의 블록 2840에서)은 동일한 크기 범위 내의 단편에 기반하여 결정될 수 있다. 예를 들어, 메틸화 수준은 암을 갖는 또는 암을 갖지 않는 것으로 알려진 표본에서 결정될 수 있고, 컷오프값은 이들 메틸화 수준으로부터 결정될 수 있다.
- [0374] 다른 구현예에서, 순환 DNA의 메틸화 밀도 및 크기 간 함수 관계가 결정될 수 있다. 함수 관계는 함수의 계수 또는 데이터 지점에 의해 정의될 수 있다. 함수 관계는 상기 각각의 크기에 대응하는 스케일링값을 제공할 수 있다(예로 더 짧은 크기는 메틸화에 대해 대응하는 증가를 가질 수 있다). 다양한 시행에서, 스케일링값은 0 내지 1 또는 1 초과일 수 있다.
- [0375] 평균 크기에 기반하여 정규화가 이루어질 수 있다. 예를 들어, 제1 메틸화 수준을 계산하기 위해 이용되는 DNA 분자에 대응하는 평균 크기가 컴퓨터 연산되고, 상기 제1 메틸화 수준이 대응 스케일링값으로 곱해질 수 있다(즉 평균 크기에 대응됨). 다른 예로서, 각각의 DNA 분자의 메틸화 밀도는 DNA 분자의 크기와 DNA 크기 및 메틸화 간 관계에 따라 정규화될 수 있다.
- [0376] 다른 시행에서, 정규화는 분자 기준 별로 수행될 수 있다. 예를 들어, 특정 부위에서 DNA 분자의 각각의 크기가 수득될 수 있고(예로, 전술된 바와 같이), 각각의 크기에 대응하는 스케일링값이 함수 관계로부터 확인될 수 있다. 비-정규화된 계산에 있어서, 상기 각각의 분자는 부위에서의 메틸화 지수 결정에 있어서 동일하게 카운트될 것이다. 정규화된 계산에 있어서, 메틸화 지수에 대한 분자의 기여는 분자 크기에 대응하는 스케일링 인자에 의해 가중치를 얻을 것이다.
- [0377] 도 34b 및 34c는 SLE 환자 SLE04(도 34b) 및 HCC 환자 TBR36(도 34c)로부터의 혈장 DNA에 대한 메틸화 분석을 나타낸다. 외부 원은 컴퓨터내 크기 분획화 없이 혈장 DNA에 대한 Z_{meth} 결과를 나타낸다. 내부 원은 130bp 이상의 혈장 DNA에 대한 Z_{meth} 결과를 나타낸다. SLE 환자 SLE04에 있어서, 84%의 bin이 컴퓨터내 크기 분획화 없이 저메틸화를 나타내었다. 130bp 이상의 단편만 분석된 경우 저메틸화를 나타내는 bin의 백분율이 15%로 감소되었다. HCC 환자 TBR36에 있어서, 98.5% 및 98.6%의 bin이 각각 컴퓨터내 크기 분획화로 또는 이것이 없이 혈장 DNA에 대해 저메틸화를 나타내었다. 이들 결과는 컴퓨터내 크기 분획화가, 예로 SLE 또는 다른 염증 상태 환자에서, 혈장 DNA의 증가된 단편화에 관련된 위양성 저메틸화 결과를 효과적으로 감소시킬 수 있음을 제시한다.
- [0378] 하나의 구현예에서, 크기 분획화를 포함하는 및 이를 포함하지 않는 분석에 대한 결과를 비교하여 메틸화 결과 상에 크기의 임의 교란 효과가 있는지를 나타낼 수 있다. 따라서 정규화에 부가하여 또는 대신에, 특정 크기에서 메틸화 수준의 계산을 이용하여 컷오프값을 초과하는 bin의 백분율이 크기 분획화를 포함하는 및 포함하지 않는 것과 더불어 상이한 경우 위양성 가능성이 있는지 또는 단지 특정 메틸화 수준이 상이한 것인지를 결정할 수 있다. 예를 들어, 크기 분획화를 포함하는 및 포함하지 않는 표본에 대한 결과 간의 유의미한 차이의 존재를 이용하여 비정상적인 단편화 패턴으로 인한 위양성 결과의 가능성을 나타낼 수 있다. 차이가 유의미한지를 결정하기 위한 역치는 암 환자 코호트 및 비-암 대조군 대상자의 코호트 분석을 통해 정립될 수 있다.
- [0379] I. 혈장 중 게놈에 걸친 CpG 섬 고메틸화에 대한 분석
- [0380] 일반적인 저메틸화에 부가하여, CpG 섬의 고메틸화도 암에서 일반적으로 관찰된다(SB Baylin 등, 2011 Nat Rev Cancer; 11: 726-734; PA Jones 등, 2007, Cell; 128: 683-692; M Esteller 등, 2007 Nat Rev Genet 2007; 8: 286-298; M Ehrlich 등, 2002 Oncogene 2002; 21: 5400-5413). 이 섹션에서는 암의 검출 및 모니터링을 위한 CpG 섬 고메틸화에 대한 게놈에 걸친 분석의 이용을 설명한다.

- [0381] 도 35는 본 발명의 구현예에 따른 CpG 섬의 고메틸화에 기반한 암 수준의 분류 결정 방법 3500의 순서도이다. 방법 2800의 복수의 부위에는 CpG 부위가 포함될 수 있고, 여기서 CpG 부위는 복수의 CpG 섬으로 조직되며, 각각의 CpG 섬은 하나 이상의 CpG 부위를 포함한다. 각각의 CpG 섬에 대한 메틸화 수준을 이용하여 암의 수준 분류를 결정할 수 있다.
- [0382] 블록 3510에서, 분석될 CpG 섬이 확인된다. 상기 분석에서, 예로서 먼저 분석될 CpG 섬 세트를 정의하였고, 이는 건강한 기준 대상자의 혈장 중 상대적으로 낮은 메틸화 밀도를 특징으로 한다. 하나의 측면에서, 기준 군에서의 메틸화 밀도 변동은 암-연관 고메틸화의 검출을 보다 쉽게 허용하기 위해 상대적으로 작을 수 있다. 하나의 구현예에서, CpG 섬은 기준 군에서 제1 백분율보다 작은 평균 메틸화 밀도를 가지며, 기준 군에서 메틸화 밀도에 대한 변동 계수는 제2 백분율보다 작다.
- [0383] 예로서, 예시적 목적으로 다음 기준이 유용한 CpG 섬의 확인을 위해 이용된다:
- [0384] i. 기준 군(예로 건강한 대상자)에서 CpG 섬에 대한 평균 메틸화 밀도 <5%
- [0385] ii. 기준 군(예로 건강한 대상자)에서 혈장 중 메틸화 밀도의 분석을 위한 변동 계수 <30%.
- [0386] 이들 파라미터는 특정 용도를 위해 조정될 수 있다. 본 발명자들의 데이터세트로부터, 게놈 중 454개 CpG 섬이 상기 기준을 충족하였다.
- [0387] 블록 3520에서, 각각의 CpG 섬에 대해 메틸화 밀도가 계산된다. 메틸화 밀도는 본원에 기재된 바와 같이 결정될 수 있다.
- [0388] 블록 3530에서, 각각의 CpG 섬이 고메틸화되어 있는지의 여부가 결정된다. 예를 들어 평가 케이스의 CpG 섬 고메틸화에 대한 분석을 위해, 각각의 CpG 섬의 메틸화 밀도가 기준 군의 대응 데이터와 비교되었다. 메틸화 밀도(메틸화 수준의 예)를 하나 이상의 컷오프값과 비교하여 특정 섬이 고메틸화되어 있는지의 여부를 결정할 수 있다.
- [0389] 하나의 구현예에서, 제1 컷오프값은 기준 군에 대한 메틸화 밀도의 평균 + 특정된 백분율에 대응할 수 있다. 다른 컷오프값은 기준 군에 대한 메틸화 밀도의 평균 + 표준 편차의 특정된 수에 대응할 수 있다. 하나의 시행에서, z -스코어(Z_{meth})가 계산되고 컷오프값과 비교되었다. 예로서, 평가 대상자(예로 암에 대해 스크리닝되는 대상자)에서의 CpG 섬은 이것이 다음 기준을 충족하는 경우, 유의미하게 고메틸화된 것으로 간주되었다:
- [0390] i. 그 메틸화 밀도가 기준 군의 평균보다 2% 이상 더 높았음, 및
- [0391] ii. $Z_{\text{meth}} > 3$.
- [0392] 이들 파라미터도 특정 용도를 위해 조정될 수 있다.
- [0393] 블록 3540에서, 고메틸화 CpG 섬의 메틸화 밀도(예로, z -스코어로)를 이용하여 누적 스코어를 결정한다. 예를 들어, 모든 유의미하게 고메틸화된 CpG 섬의 확인 후, z -스코어의 합 또는 모든 고메틸화 CpG 섬의 z -스코어의 함수가 수반되는 스코어가 계산될 수 있다. 스코어의 한 예는 다른 섹션에 기재되는 누적 확률(CP) 스코어이다. 누적 확률 스코어는 Z_{meth} 를 이용하여 확률 분포에 따라 우연한 관찰을 가질 확률을 결정한다(예로, 3의 자유도를 갖는 Student t 확률 분포).
- [0394] 블록 3550에서, 누적 스코어를 누적 역치와 비교하여 암의 수준 분류를 결정한다. 예를 들어, 확인된 CpG 섬에서의 총 고메틸화가 충분히 크면, 유기체가 암을 갖는 것으로 확인될 수 있다. 하나의 구현예에서, 누적 역치는 기준 군으로부터의 최고 누적 스코어에 대응한다.
- [0395] IX. 메틸화 및 CNA
- [0396] 상술된 바와 같이, 본원에 기재된 메틸화 분석 접근은 혈장 중 종양-유래 DNA의 유전적 변화에 기반한 다른 방법과 조합되어 사용될 수 있다. 이러한 방법의 예에는 암-연관 염색체 이상 분석이 포함된다(KCA Chan 등, 2013 Clin Chem; 59: 211-224; RJ Leary 등, 2012 Sci Transl Med; 4: 162ra154). 복제수 이상(CNA)의 측면은 U.S. 특허 출원 번호 13/308,473에 기재된다.
- [0397] A. CNA
- [0398] 복제수 이상은 게놈의 특정 부분에 정렬하는 DNA 단편을 카운트하고, 카운트를 정규화하고, 카운트를 컷오프값과 비교하여 검출될 수 있다. 다양한 구현예에서, 정규화는 게놈의 동일 부분의 다른 일배체형에 정렬된 DNA 단

편(상대 일배체형 용량(RHDO))의 카운트에 의해 또는 게놈의 다른 부분에 정렬된 DNA 단편의 카운트에 의해 수행될 수 있다.

- [0399] RHDO 방법은 이중접합성 유전자좌의 이용에 의존한다. 본 섹션에 기재된 구현에는 또한 동일한 영역의 두 일배체형이 아지며, 따라서 비-일배체형 특이적인 두 영역을 비교하여 동중접합성인 유전자좌에 대해 이용될 수 있다. 상대 염색체 영역 용량 방법에서, 한 염색체 영역으로부터의 단편의 수(예로, 그 영역에 정렬된 서열 리드의 카운팅에 의해 결정됨)가 예상값(건강한 것으로 알려진 다른 표본에서의 동일한 영역에서 또는 기준 염색체 영역에서 유래될 수 있음)과 비교된다. 이러한 방식으로, 단편은 서열분석된 태그가 유래되는 일배체형과 무관하게 염색체 영역에 대해 카운트될 것이다. 따라서 이중접합성 유전자좌를 함유하지 않는 서열 리드가 여전히 이용될 수 있다. 비교를 수행하기 위해, 하나의 구현에는 비교 전에 태그 카운트를 정규화할 수 있다. 각각의 영역은 적어도 두 유전자좌(서로 분리됨)에 의해 정의되며, 이들 유전자좌에서의 단편을 이용하여 영역에 대한 종합값을 수득할 수 있다.
- [0400] 특정 영역에 대한 서열분석 측정치(태그)에 대한 정규화값은 그 영역에 정렬하는 서열분석 측정치의 수를 전체 게놈에 정렬 가능한 서열분석 측정치의 총 수로 나누어 계산될 수 있다. 상기 정규화된 태그 카운트는 하나의 표본으로부터의 결과가 다른 표본의 결과와 비교될 수 있도록 한다. 예를 들어, 정규화값은 위에 언급된 바와 같이, 특정 영역 유래로 예상되는 서열분석 측정치의 비율(예로, 백분율 또는 분획)일 수 있다. 다른 구현에 있어서, 정규화를 위해 다른 방법이 가능하다. 예를 들어, 한 영역에 대한 카운트 수를 기준 영역에 대한 카운트 수로 나누어 정규화할 수 있다(상기 경우, 기준 영역은 단지 전체 게놈이다). 이어서 이러한 정규화된 태그 카운트가 역치값에 대해 비교될 수 있고, 이는 암을 나타내지 않는 하나 이상의 기준 표본에서 결정될 수 있다.
- [0401] 그 뒤, 평가된 케이스의 정규화된 태그 카운트는 하나 이상의 기준 대상자, 예로 암이 없는 대상자의 정규화된 태그 카운트와 비교될 것이다. 하나의 구현예에서, 특정 염색체 영역에 대한 케이스의 z -스코어를 계산하여 비교가 수행된다. z -스코어는 다음 식을 이용해서 계산될 수 있다: $z\text{-스코어} = (\text{케이스의 정규화된 태그 카운트} - \text{평균}) / \text{SD}$,
- [0402] 식 중, "평균"은 기준 표본에 대해 특정 염색체 영역에 정렬하는 평균 정규화 태그 카운트이고; SD는 기준 표본에 있어서 특정 영역에 정렬하는 정규화 태그 카운트 수의 표준 편차이다. 따라서 z -스코어는 평가 케이스에 대한 염색체 영역의 정규화 태그 카운트가 하나 이상의 기준 대상자의 동일한 염색체 영역에 대한 평균 정규화 태그 카운트에서 벗어나는 표준 편차의 수이다.
- [0403] 평가 유기체가 암을 갖는 상황에서, 종양 조직에서 증폭된 염색체 영역은 혈장 DNA 중에 상향 제시될 것이다. 이는 양의 값의 z -스코어를 생성할 것이다. 반면, 종양 조직에서 결실된 염색체 영역은 혈장 DNA에서 하향 제시될 것이다. 이는 음의 값의 z -스코어를 생성할 것이다. z -스코어의 크기는 몇몇 요인에 의해 결정된다.
- [0404] 하나의 요인은 생물학적 표본(예로 혈장) 중 종양-유래 DNA의 분획 농도이다. 표본(예로 혈장) 중 종양-유래 DNA의 분획 농도가 높을수록, 평가 케이스 및 기준 케이스의 정규화 태그 카운트 간 차이가 클 것이다. 따라서 더 큰 크기의 z -스코어를 생성할 것이다.
- [0405] 다른 요인은 하나 이상의 기준 케이스에서의 정규화 태그 카운트의 변동이다. 평가 케이스의 생물학적 표본(예로 혈장) 중 염색체 영역의 동일한 정도의 상향 제시에 있어서, 기준 군에서 정규화 태그 카운트의 더 작은 변동(즉, 더 작은 표준 편차)는 더 높은 z -스코어를 생성할 것이다. 유사하게, 평가 케이스의 생물학적 표본(예로 혈장) 중 염색체 영역의 동일한 정도의 하향 제시에 있어서, 기준 군에서 정규화 태그 카운트의 더 작은 표준 편차는 더 음의 z -스코어를 생성할 것이다.
- [0406] 또 다른 요인은 종양 조직에서의 염색체 이상의 크기이다. 염색체 이상의 크기는 특정 염색체 영역에 대한 복제수 변화(획득 또는 손실)를 나타낸다. 종양 조직에서의 복제수 변화가 높을수록, 혈장 DNA에서 특정 염색체 영역의 상향 또는 하향 제시의 정도가 더 높다. 예를 들어, 염색체의 두 사본 모두의 손실은 염색체의 두 사본 중 하나의 손실에 비해 혈장 DNA 중 염색체 영역의 더 큰 하향 제시를 일으킬 것이므로, 이에 따라 더 음의 z -스코어를 생성하였다. 전형적으로, 암에는 여러 염색체 이상이 존재한다. 각각의 암에서의 염색체 이상은 그 성질(즉 증폭 또는 결실), 그 정도(단일 또는 복수의 복제 획득 또는 손실) 및 그 범위(염색체 길이의 관점에서 이상의 크기)에 의해 추가로 변할 수 있다.
- [0407] 정규화 태그 카운트의 측정 정밀도는 분석되는 분자의 수에 의해 영향을 받는다. 분획 농도가 각각 대략 12.5%, 6.3% 및 3.2%인 경우, 하나의 복제 변화(획득 또는 손실)를 갖는 염색체 이상을 검출하기 위해 15,000, 60,000 및 240,000 분자가 분석되어야 할 것으로 예상된다. 상이한 염색체 영역에 대해 암 검출을 위한 태그 카운팅의

추가 상세내용은 Lo 등의 "Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing"을 표제로 하는 U.S. 특허 공개 번호 2009/0029377에 기재되며, 그 전문은 본원에 모든 목적에 대해 참조로 도입된다.

- [0408] 구현예는 또한 태그 카운팅 방법 대신에 크기 분석을 이용할 수 있다. 크기 분석은 또한 정규화 태그 카운트 대신에 이용될 수 있다. 크기 분석은 본원 및 U.S. 특허 출원 번호 12/940,992에서 언급된 다양한 파라미터를 이용할 수 있다. 예를 들어, 상기로부터의 Q 또는 F값이 이용될 수 있다. 이러한 크기값은 이들 값이 측정치의 수와 함께 스케일이 변하지 않으므로 다른 영역으로부터의 카운트에 의해 정규화될 필요가 없다. 일배체형-특이적 방법의 기법, 예컨대 상술되고 U.S. 특허 출원 번호 13/308,473에 보다 상세히 기재된 RHDO 방법이 비특이적 방법에 대해서도 이용될 수 있다. 예를 들어, 영역의 깊이 및 정련이 수반되는 기법이 이용될 수 있다. 일부 구현예에서, 두 영역의 비교 시 특정 영역에 대한 GC 바이어스가 고려될 수 있다. RHDO 방법은 동일한 영역을 이용하므로, 이러한 수정이 필요하지 않다.
- [0409] 특정 암은 전형적으로 특정 염색체 영역에서의 이상을 제시하지만, 이러한 암이 항상 이러한 영역에서의 이상을 배타적으로 제시하지는 않는다. 예를 들어, 추가적인 염색체 영역이 이상을 나타낼 수 있고, 이러한 추가적 영역의 위치가 알려지지 않을 수 있다. 또한 암의 초기 단계를 확인하기 위한 환자 스크리닝 시, 다양한 암을 확인하고 싶을 수 있고, 이는 게놈에 걸쳐 존재하는 이상을 나타낼 수 있다. 이러한 상황을 해결하기 위해, 구현예는 어느 영역이 이상을 나타내는지를 결정하기 위한 체계적 방식으로 복수 영역을 분석할 수 있다. 이상의 수 및 이들의 위치(예로 이들이 인접한지 여부)를 이용하여, 예를 들어 이상을 확인하고, 암의 단계를 결정하고, 암의 진단을 제공하고(그 수가 역치값보다 큰 경우), 이상을 나타내는 다양한 영역의 수 및 위치에 기반하여 예후를 제공할 수 있다.
- [0410] 따라서 구현예는 이상을 나타내는 영역의 수에 기반하여 유기체가 암을 갖는지를 확인할 수 있다. 따라서 복수 영역(예로, 3,000)을 평가하여 이상을 나타내는 영역의 수를 확인할 수 있다. 영역은 전체 게놈 또는 게놈의 단지 일부, 예로 비반복 영역을 커버할 수도 있다.
- [0411] 도 36은 본 발명의 구현예에 따른 복수의 염색체 영역을 이용한 유기체의 생물학적 표본의 분석 방법 3600의 순서도이다. 생물학적 표본에는 핵산 분자(단편으로도 불림)가 포함된다.
- [0412] 블록 3610에서, 유기체 게놈의 복수 영역(예로, 비-중복 영역)이 확인된다. 각각의 염색체 영역에는 복수의 유전자좌가 포함된다. 영역은 1Mb 크기일 수도, 또는 일부 다른 동일-크기일 수도 있다. 영역이 1Mb 크기인 상황에 있어서, 전체 게놈은 각각 소정 크기 및 위치의 약 3,000개 영역을 포함할 수 있다. 이러한 소정 영역은 특정 염색체의 길이 이용될 영역의 특정된 수 및 본원에서 언급된 임의의 다른 기준에 맞춰 변할 수 있다. 영역이 상이한 길이를 갖는 경우, 이러한 길이를 이용하여 결과를, 예로 본원에 기재된 바와 같이 정규화할 수 있다. 영역은 평가되는 암의 지식에 기반하여 및/또는 특정 유기체의 특정 요건에 기반하여 구체적으로 선택될 수 있다. 영역은 또한 임의로 선택될 수 있다.
- [0413] 블록 3620에서, 유기체의 기준 게놈 중 핵산 분자의 위치가 각각의 복수의 핵산 분자에 대해 확인된다. 위치는 본원에서 언급된 임의의 방식, 예로 분석된 태그를 수득하기 위해 단편을 서열분석하고, 서열분석된 태그를 기준 게놈에 대해 정렬하여 결정될 수 있다. 분자의 특정 일배체형도 일배체형-특이적 방법에 대해 결정될 수 있다.
- [0414] 블록 3630-3650이 각각의 염색체 영역에 대해 수행된다. 블록 3630에서, 각 그룹의 핵산 분자는 확인된 위치에 기반하여 염색체 영역에서 유래된 것으로 확인된다. 각 그룹에는 염색체 영역의 각각의 복수 유전자좌에 위치한 적어도 하나의 핵산 분자가 포함될 수 있다. 하나의 구현예에서, 그룹은, 예로 상기 RHDO 방법에서와 같이 염색체 영역의 특정 일배체형에 정렬하는 단편일 수 있다. 다른 구현예에서, 그룹은 염색체 영역에 정렬하는 임의의 단편일 수 있다.
- [0415] 블록 3640에서, 컴퓨터 시스템은 각 그룹의 핵산 분자의 각각의 값을 계산한다. 각각의 값은 각 그룹의 핵산 분자의 특성을 정의한다. 각각의 값은 본원에 언급된 임의의 값일 수 있다. 예를 들어, 값은 그룹 내 단편의 수 또는 그룹 내 단편의 크기 분포의 통계값일 수 있다. 각각의 값은 또한 정규화값, 예로 기준 영역에 대한 태그 카운트의 수 또는 표본에 대한 태그 카운트의 총 수로 나눈 영역의 태그 카운트일 수 있다. 각각의 값은 또한 다른 값으로부터의 차이 또는 비(예로 RHCO 중)일 수 있고, 이에 따라 영역에 대한 차이 특성을 제공할 수 있다.
- [0416] 블록 3650에서, 각각의 값을 기준값과 비교하여 제1 염색체 영역이 결실 또는 증폭을 나타내는지의 분류를 결정한다. 상기 기준값은 본원에 기재된 임의의 역치 또는 기준값일 수 있다. 예를 들어, 기준값은 정상 표본에 대

해 결정된 역치값일 수 있다. RHDO에 있어서, 상기 각각의 값은 두 일배체형에 대한 태그 카운트의 차이 또는 비일 수 있고, 기준값은 통계적으로 유의미한 일탈이 존재하는지를 결정하기 위한 역치일 수 있다. 다른 예로서, 기준값은 다른 일배체형 또는 영역에 대한 태그 카운트 또는 크기값일 수 있고, 비교에는 차이 또는 비(또는 그 함수)를 얻은 뒤 차이 또는 비가 역치값보다 큰지를 결정하는 것이 포함될 수 있다.

[0417] 기준값은 다른 영역의 결과에 기반하여 변할 수 있다. 예를 들어, 주변 영역도 일탈을 보이는 경우(하나의 역치에 비해 적더라도, 예로 z -스코어 3이더라도), 더 낮은 역치가 이용될 수 있다. 예를 들어, 세 연속 영역이 모두 제1 역치보다 위에 있는 경우, 암의 가능성이 더 높을 수 있다. 따라서 상기 제1 역치는 비연속 영역에서 암을 확인하기 위해 필요한 다른 역치보다 낮을 수 있다. 더 작은 일탈을 갖는 세(또는 넷 이상의) 영역을 갖는 것은 민감성 및 특이성이 보존될 수 있는 충분히 낮은 기회 효과 확률을 가질 수 있다.

[0418] 블록 3660에서, 결실 또는 증폭을 나타내는 것으로 분류되는 게놈 영역의 양이 결정된다. 카운트되는 염색체 영역은 제한을 가질 수 있다. 예를 들어, 적어도 하나의 다른 영역에 인접한 영역만 카운트될 수 있다(또는 인접 영역이 특정 크기, 예로 4개 이상의 영역이어야 할 수 있다). 영역이 동일하지 않은 구현예에 있어서, 상기 각각의 길이에 대한 수도 고려될 수 있다(예로, 그 수는 이상 영역의 총 길이일 수 있다).

[0419] 블록 3670에서, 양을 양의 역치값과 비교하여 표본의 분류를 결정한다. 예로서, 분류는 유기체에 암이 있는지의 여부, 암의 단계, 및 암의 예후일 수 있다. 하나의 구현예에서, 모든 이상 영역이 카운트되고 영역이 나타나는 위치와 무관하게 단일 역치값이 이용된다. 다른 구현예에서, 역치값은 카운트되는 영역의 위치 및 크기에 따라 변할 수 있다. 예를 들어, 특정 염색체 또는 염색체 팔 상의 영역의 양은 그 특정 염색체(또는 팔)에 대한 역치와 비교될 수 있다. 다중 역치가 이용될 수 있다. 예를 들어, 특정 염색체(또는 팔)에서 이상 영역의 양은 제1 역치값보다 커야 하며, 게놈에서 이상 영역의 총량은 제2 역치값보다 커야 한다. 역치값은 결실 또는 증폭을 나타내는 것으로 결정되는 영역의 백분율일 수 있다.

[0420] 영역의 양에 대한 상기 역치값은 또한 카운트되는 영역에 대한 불균형이 얼마나 강력한지에 의존할 수 있다. 예를 들어, 암의 분류 결정을 위한 역치로 이용되는 영역의 양은 각 영역에서 이상을 검출하기 위해 이용되는 특이성 및 민감성(이상 역치)에 의존할 수 있다. 예를 들어, 이상 역치가 낮은 경우(예로 z -스코어 2), 양의 역치는 높게 선택될 수 있다(예로 150). 그러나 이상 역치가 높은 경우(예로, z -스코어 3), 양의 역치는 낮아질 수 있다(예로, 50). 이상을 나타내는 영역의 양은 또한 가중치가 부여된 값일 수 있다, 예로 높은 불균형을 나타내는 하나의 영역이 거의 불균형을 보이지 않는 영역에 비해 더 가중치가 높을 수 있다(즉, 이상에 대해 단지 양 및 음의 분류보다 더 많은 분류가 있다). 예로서, z -스코어의 합이 이용되어 가중치가 부여된 값을 이용할 수 있다.

[0421] 따라서 정규화 태그 카운트(또는 그룹 특성에 대한 다른 각각의 값)의 유의미한 상향- 또는 하향-제시를 나타내는 염색체 영역의 양(수 및/또는 크기를 포함할 수 있음)이 질환의 중증도를 반영하기 위해 이용될 수 있다. 이상 정규화 태그 카운트를 갖는 염색체 영역의 양은 두 요인, 즉 종양 조직 중 염색체 이상의 수(또는 크기) 및 생물학적 표본(예로 혈장) 중 종양-유래 DNA의 분획 농도에 의해 결정될 수 있다. 보다 진행된 암은 더 많은(그리고 더 큰) 염색체 이상을 보이는 경향이 있다. 따라서 더 많은 암 연관 염색체 이상이 잠재적으로 표본(예로 혈장) 중에서 검출 가능할 것이다. 더 진행된 암 환자에서, 더 높은 종양 부하는 혈장 중 종양-유래 DNA의 더 높은 분획 농도로 이어질 것이다. 그 결과, 종양 연관 염색체 이상이 혈장 표본 중에서 보다 쉽게 검출될 것이다.

[0422] 특이성을 손상시키지 않고 민감성을 개선하기 위한 하나의 가능한 접근은 인접한 염색체 절편의 결과를 고려하는 것이다. 하나의 구현예에서, z -스코어에 대한 컷오프는 >2 및 <-2 로 유지된다. 그러나 염색체 영역은 두 연속 절편이 동일한 유형의 이상을 나타내는 경우에만, 예로 두 절편이 모두 2 초과와 z -스코어를 갖는 경우에만 잠재적으로 이상한 것으로 분류될 것이다. 다른 구현예에서, 주변 절편의 z -스코어가 더 높은 컷오프값을 이용하여 함께 추가될 수 있다. 예를 들어, 세 연속 절편의 z -스코어가 합산될 수 있고 컷오프값 5가 이용될 수 있다. 상기 개념이 넷 이상의 연속 절편으로 확장될 수 있다.

[0423] 양 및 이상 역치의 조합은 또한 분석의 목적 및 유기체의 임의의 사전 지식(또는 그 부재)에 의존할 수 있다. 예를 들어, 암에 대해 건강한 정상 모집단을 스크리닝하는 경우, 영역이 이상을 갖는 것으로 확인된 경우에 대해 잠재적으로 영역의 양(즉 영역의 수에 대한 높은 역치) 및 이상 역치 모두에서 전형적으로 높은 특이성을 이용할 것이다. 그러나 더 높은 위험을 갖는 환자(예로 종괴 또는 가족력이 있는 환자, 흡연자, 만성 인간 유두종 바이러스(HPV) 보균자, 간염 바이러스 보균자 또는 다른 바이러스 보균자)에서, 역치는 더 큰 민감성(더 적은 위음성)을 갖기 위해 낮아질 수 있다.

- [0424] 하나의 구현예에서, 염색체 이상을 검출하기 위해 1-Mb 해상 및 종양-유래 DNA의 더 낮은 검출 하한 6.3%를 이용하는 경우, 각각의 1Mb 절편에서의 분자 수는 60,000이 되어야 할 것이다. 이는 전체 게놈에 대해 대략 1억 8천만 개(60,000측정치/Mb×3,000Mb)의 정렬 가능한 측정치로 환산될 것이다.
- [0425] 더 작은 절편 크기는 더 작은 염색체 이상을 검출하기 위해 더 높은 해상도를 제공할 것이다. 그러나 이는 전체적으로 분석될 분자 수의 요건을 증가시킬 것이다. 더 큰 절편 크기는 해상도를 소모하며 분석을 위해 필요한 분자 수를 감소시킬 것이다. 따라서 더 큰 이상만 검출될 수 있다. 하나의 시행에서, 더 큰 영역이 이용될 수 있고, 이상을 나타내는 절편이 하위 구분되고 이들 하위 영역이 분석되어 더 우수한 해상도를 얻을 수 있다(예로 전술된 바와 같이). 검출될 결실 또는 증폭의 크기(또는 검출할 최소 농도)에 대한 산정치가 있는 경우, 분석할 분자 수가 결정될 수 있다.
- [0426] *B. 바이설파이트-처리 혈장 DNA의 서열분석에 기반한 CNA*
- [0427] 게놈에 걸친 저메틸화 및 CNA는 종양 조직에서 빈번히 관찰될 수 있다. 여기서, CNA 및 암 연관 메틸화 변화 정보가 혈장 DNA의 바이설파이트 서열분석에서 동시에 획득될 수 있음을 나타낸다. 두 유형의 분석이 동일한 데이터 세트 상에서 수행될 수 있으므로, 실질적으로 CNA 분석에 대한 추가 비용이 없다. 다른 구현예는 상이한 절차를 이용해서 메틸화 정보 및 유전 정보를 획득할 수 있다. 다른 구현예에서, CNA 분석과 함께 암 연관 고메틸화에 대한 유사 분석을 수행할 수 있다.
- [0428] 도 37a는 환자 TBR36에 대한 종양 조직, 바이설파이트(BS)-비처리 혈장 DNA 및 바이설파이트-처리 혈장 DNA에 대한 CNA 분석(내부에서 외부로)을 나타낸다. 도 37a는 환자 TBR36에 대한 종양 조직, 바이설파이트(BS)-비처리 혈장 DNA 및 바이설파이트-처리 혈장 DNA에 대한 CNA 분석(내부에서 외부로)을 나타낸다. 최외부 고리는 염색체 핵형도를 나타낸다. 각각의 점은 1-Mb 영역의 결과를 나타낸다. 녹색, 빨간색 및 회색 점은 각각 복제수 획득, 복제수 손실을 갖는 영역 및 복제수 변화가 없는 영역을 나타낸다. 혈장 분석을 위해, z-스코어를 나타낸다. 두 중심 선 사이에는 5의 차이가 존재한다. 종양 조직 분석을 위해, 복제수를 나타낸다. 두 중심 선 사이에 하나의 사본 차이가 존재한다. 도 38a는 환자 TBR34에 대한 종양 조직, 바이설파이트(BS)-비처리 혈장 DNA 및 바이설파이트-처리 혈장 DNA에 대한 CNA 분석(내부에서 외부로)을 나타낸다. 바이설파이트-처리 및 바이설파이트-비처리 혈장 표본에서 검출된 CNA 패턴은 일치하였다.
- [0429] 종양 조직, 바이설파이트-비처리 혈장 및 바이설파이트-처리 혈장에서 검출된 CNA 패턴은 일치하였다. 바이설파이트-처리 및 바이설파이트-비처리 혈장 결과 간 일치를 추가 평가하기 위해, 산포 그래프가 생성된다. 도 37b는 환자 TBR36에 대한 1Mb bin의 바이설파이트-처리 및 바이설파이트-비처리 혈장을 이용한 CNA 검출을 위한 z-스코어 간 관계를 나타내는 산포 그래프이다. 두 분석의 z-스코어 간 양의 상관관계가 관찰되었다($r = 0.89$, $p < 0.001$, Pearson 상관관계). 도 38b는 환자 TBR34에 대한 1Mb bin의 바이설파이트-처리 및 바이설파이트-비처리 혈장을 이용한 CNA 검출을 위한 z-스코어 간 관계를 나타내는 산포 그래프이다. 두 분석의 z-스코어 간 양의 상관관계가 관찰되었다($r = 0.81$, $p < 0.001$, Pearson 상관관계).
- [0430] *C. 암 연관 CNA 및 메틸화 변화의 상승적 분석*
- [0431] 전술된 바와 같이, CNA에 대한 분석에는 각각의 1Mb 영역에서 서열 리드의 수의 카운팅이 수반될 수 있는 반면, 메틸화 밀도에 대한 분석에는 CpG 디뉴클레오타이드에서 시토신 잔기가 메틸화된 비율의 검출이 수반될 수 있다. 이들 두 분석의 조합은 암 검출을 위해 상승적 정보를 제공할 수 있다. 예를 들어, 메틸화 분류 및 CNA 분류를 이용하여 암 수준의 제3 분류를 결정할 수 있다.
- [0432] 하나의 구현예에서, 암 연관 CNA 또는 메틸화 변화의 존재를 이용하여 암의 잠재적 존재를 나타낼 수 있다. 이러한 구현예에서, 암 검출의 민감성은 CNA 또는 메틸화 변화가 평가된 대상자의 혈장 중에 존재하는 경우 증가될 수 있다. 다른 구현예에서, 두 변화의 존재를 이용하여 암의 존재를 나타낼 수 있다. 이러한 구현예에서, 두 유형의 변화 중 하나가 일부 비-암 대상자에서 잠재적으로 검출될 수 있으므로, 평가의 특이성이 개선될 수 있다. 따라서 제3 분류는 상기 제1 분류 및 제2 분류가 모두 암을 나타내는 경우에만 암에 대해 양성일 수 있다.
- [0433] 26명의 HCC 환자 및 22명의 건강한 대상자가 모집되었다. 혈액 표본이 각 대상자에서 수집되었고, 혈장 DNA가 바이설파이트 처리 후 서열분석되었다. HCC 환자에 있어서, 혈액 표본이 진단 시 수집되었다. 유의미한 양의 CNA의 존재는, 예를 들어 <-3 또는 >3 의 z-스코어를 나타내는 bin을 5% 초과로 갖는 것으로 정의되었다. 유의미한 양의 암 연관 저메틸화의 존재는 <-3 의 z-스코어를 나타내는 bin을 3% 초과로 갖는 것으로 정의되었다. 예로서, 영역(bin)의 양은 bin의 미가공 카운트, 백분율 및 bin의 길이로 표현될 수 있다.

[0434] 표 3은 바이셀과이트-처리 혈장 DNA 상에서 대량 병렬 서열분석을 이용한 26명의 HCC 환자 혈장 중 유의미한 양의 CNA 및 메틸화 변화의 검출을 나타낸다.

표 3

[0435]

		CNA	
		존재	부재
메틸화 변화	존재	12	6
	부재	1	7

[0436] 암 연관 메틸화 변화 및 CNA의 검출 비율은 각각 69% 및 50%였다. 어느 한 기준의 존재를 이용하여 암의 잠재적 존재를 나타낸 경우, 검출 비율(즉 진단 민감성)은 73%로 개선되었다.

[0437] CNA(도 39a) 또는 메틸화 변화(도 39b)의 존재를 나타내는 두 환자의 결과를 나타낸다. 도 39a는 HCC 환자 TBR240에 있어서 바이셀과이트-처리 혈장에 대한 CNA(내부 고리) 및 메틸화 분석(외부 고리)을 나타내는 Circos 그래프이다. CNA 분석에 있어서, 녹색, 빨간색 및 회색 점은 각각 염색체 획득, 손실을 갖는 영역 및 복제수 변화가 없는 영역을 나타낸다. 메틸화 분석에 있어서, 녹색, 빨간색 및 회색 점은 각각 고메틸화, 저메틸화 및 정상 메틸화를 갖는 영역을 나타낸다. 상기 환자에서, 암 연관 CNA가 혈장 중에 검출된 반면, 메틸화 분석은 유의미한 양의 암 연관 저메틸화를 드러내지 않았다. 도 39b는 HCC 환자 TBR164에 있어서 바이셀과이트-처리 혈장에 대한 CNA(내부 고리) 및 메틸화 분석(외부 고리)을 나타내는 Circos 그래프이다. 상기 환자에서, 암 연관 저메틸화가 혈장에서 검출되었다. 그러나, 유의미한 양의 CNA는 관찰할 수 없었다. CNA 및 메틸화 변화 모두의 존재를 나타내는 두 환자의 결과를 도 48a(TBR36) 및 49a(TBR34)에 나타낸다.

[0438] 표 4는 바이셀과이트-처리 혈장 DNA 상에서 대량 병렬 서열분석을 이용한 22명의 대조군 대상자 혈장 중 유의미한 양의 CNA 및 메틸화 변화의 검출을 나타낸다. 부트스트래핑(즉 LOO(하나를 남기고 없애는) 접근)이 각각의 대조군 대상자의 평가를 위해 이용되었다. 따라서 특정 대상자가 평가되는 경우, 다른 21명의 대상자가 대조군 그룹의 평균 및 SD 계산을 위해 이용되었다.

표 4

[0439]

		CNA	
		존재	부재
메틸화 변화	존재	1	2
	부재	1	18

[0440] 유의미한 양의 메틸화 변화 및 CNA의 검출 특이성은 각각 86% 및 91%였다. 두 기준 모두의 존재가 암의 잠재적 존재를 나타내기 위해 필요한 경우, 특이성은 95%로 개선되었다.

[0441] 하나의 구현예에서, CNA 및/또는 저메틸화에 대해 양성인 표본은 암에 대해 양성으로 간주되며, 둘 다 검출 불가능한 경우 표본은 음성으로 간주된다. "또는" 논리의 이용이 더 높은 민감성을 제공한다. 다른 구현예에서, CNA 및 저메틸화 모두에 대해 양성인 표본만 암에 대해 양성으로 간주되어, 더 높은 특이성을 제공한다. 다른 구현예에서, 3중 분류가 이용될 수 있다. 대상자는 i. 둘 다 정상; ii. 하나가 비정상; iii. 둘 다 비정상으로 분류된다.

[0442] 상이한 추적 전략이 이들 세 분류에 대해 이용될 수 있다. 예를 들어, (iii)에 대한 대상자는, 예로 전신 조영이 수반되는 가장 집중적인 추적 프로토콜을 거칠 수 있고; (ii)에 대한 대상자는, 예로 상대적으로 짧은 수주의 간격 후 혈장 DNA 서열분석을 반복하는, 덜 집중적인 추적 프로토콜을 거칠 수 있고; (i)에 대한 대상자는, 수 년 후 재평가와 같이 가장 덜 집중적인 추적 프로토콜을 거칠 수 있다. 다른 구현예에서, 메틸화 및 CNA 측정은 분류의 추가 정련을 위해 다른 임상적 파라미터(예로 조영 결과 또는 혈청 생화학)와 함께 이용될 수 있다.

[0443] D. 근처 목적 치료 후 혈장 DNA 분석의 예후값

[0444] 혈장 중 암 연관 CNA 및/또는 메틸화 변화의 존재는 암 환자의 순환 중 종양-유래 DNA의 존재를 시사할 것이다. 이들 암 연관 변화의 감소 또는 제거가 치료(예로 수술) 후 예상될 것이다. 반면, 치료 후 혈장 중 이들 변화의 지속은 신체로부터 모든 종양 세포의 불완전한 제거를 시사할 수 있고, 질환 재발에 대해 유용한 예후 지표일

수 있다.

- [0445] 혈액 표본이 종양의 근치 목적 수술적 절제 1주 후 2명의 HCC 환자 TBR34 및 TBR36에서 수집되었다. CNA 및 메틸화 분석이 바이셀파이트-처리 치료-후 혈장 표본에서 수행되었다.
- [0446] 도 40a는 HCC 환자 TBR36에 대한 종양의 수술적 절제 이전(내부 고리) 및 이후(외부 고리)에 수집된 바이셀파이트-처리 혈장 DNA 상의 CNA 분석을 나타낸다. 각각의 점은 1-Mb 영역의 결과를 나타낸다. 녹색, 빨간색 및 회색 점은 각각 복제수 획득, 복제수 손실을 갖는 영역 및 복제수 변화가 없는 영역을 나타낸다. 치료 전에 관찰된 대부분의 CNA는 종양 절제 후 사라졌다. <-3 또는 >3 의 z-스코어를 나타내는 bin의 비율은 25%에서 6.6%로 감소하였다.
- [0447] 도 40b는 HCC 환자 TBR36에 대한 종양의 수술적 절제 이전(내부 고리) 및 이후(외부 고리)에 수집된 바이셀파이트-처리 혈장 DNA 상의 메틸화 분석을 나타낸다. 녹색, 빨간색 및 회색 점은 각각 고메틸화, 저메틸화 및 정상 메틸화를 갖는 영역을 나타낸다. 유의미한 저메틸화를 나타내는 bin의 비율에서 90%에서 7.9%로의 현저한 감소가 있었으며 저메틸화 정도도 현저한 감소를 나타내었다. 상기 환자는 종양 절제 후 22개월에 완전히 임상적으로 완화되었다.
- [0448] 도 41a는 HCC 환자 TBR34에 대한 종양의 수술적 절제 이전(내부 고리) 및 이후(외부 고리)에 수집된 바이셀파이트-처리 혈장 DNA 상의 CNA 분석을 나타낸다. CNA를 나타내는 bin의 수 및 종양의 수술적 절제 후 영향받은 bin에서 CNA의 크기 모두에 감소가 존재하지만, 잔여 CNA가 수술 후 혈장 표본에서 관찰될 수 있었다. 빨간색 원은 잔여 CNA가 가장 뚜렷한 영역을 강조한다. <-3 또는 >3 의 z-스코어를 나타내는 bin의 비율은 57%에서 12%로 감소하였다.
- [0449] 도 41b는 HCC 환자 TBR34에 대한 종양의 수술적 절제 이전(내부 고리) 및 이후(외부 고리)에 수집된 바이셀파이트-처리 혈장 DNA 상의 메틸화 분석을 나타낸다. 저메틸화의 크기는 종양 절제 후 감소하여 저메틸화 bin에 대한 평균 z-스코어는 -7.9에서 -4.0으로 감소하였다. 그러나 <-3 의 z-스코어를 갖는 bin의 비율은 반대 변화를 나타내어 41%에서 85%로 증가하였다. 상기 관찰은 치료 후 잔여 암 세포의 존재를 잠재적으로 시사한다. 임상적으로, 종양 절제 후 3개월에 종양 결절의 다중 병터가 잔여 비-절제 간에서 검출되었다. 수술 후 4개월째부터 폐 전이가 관찰되었다. 환자는 수술 8개월 후 국소 재발 및 전이성 질환으로 사망하였다.
- [0450] 이들 두 환자(TBR34 및 TBR36)에서의 관찰은 CNA 및 저메틸화의 잔여 암 연관 변화의 존재가 근치 목적 치료 후 암 환자의 모니터링 및 예진을 위해 이용될 수 있음을 제시한다. 데이터는 또한 검출된 혈장 CNA의 양에서의 변화 정도가 치료 유효성의 예진 및 모니터링을 위한 혈장 DNA 저메틸화의 정도의 변화 정도 평가와 함께 상승 효과를 가지고 이용될 수 있음을 나타내었다.
- [0451] 따라서 일부 구현예에서, 하나의 생물학적 표본은 치료 전에 획득되고 제2의 생물학적 표본은 치료(예로, 수술) 후에 획득된다. 제1값, 예컨대 영역의 z-스코어(예로, 영역 메틸화 수준 및 CNA에 대한 정규화값) 및 저메틸화 및 CNA(예로, 증폭 또는 결실)를 나타내는 영역의 수가 제1 표본에 대해 획득된다. 제2값이 제2 표본에 대해 획득될 수 있다. 다른 구현예에서, 제3, 또는 더 추가적인 표본이 치료 후 획득될 수 있다. 저메틸화 및 CNA(예로, 증폭 또는 결실)를 나타내는 영역의 수가 제3, 또는 더 추가적인 표본에서 획득될 수 있다.
- [0452] 도 40a 및 41a에 대해 전술된 바와 같이, 제1 표본에 대해 저메틸화를 나타내는 영역의 제1 수는 제2 표본에 대한 저메틸화를 나타내는 영역의 제2 양과 비교될 수 있다. 도 40b 및 41b에 대해 전술된 바와 같이, 제1 표본에 대해 저메틸화를 나타내는 영역의 제1 양은 제2 표본에 대해 저메틸화를 나타내는 영역의 제2 양과 비교될 수 있다. 제1 양과 제2 양, 및 제1 수와 제2 수의 비교를 이용하여 치료의 예후를 결정할 수 있다. 다양한 구현예에서, 하나의 비교로만 예후가 결정될 수 있거나 두 비교가 모두 이용될 수 있다. 제3 또는 더 추가적인 표본이 획득되는 구현예에서, 하나 이상의 이들 표본을 이용하여 자체적으로 또는 제2 표본과 함께 치료의 예후를 결정할 수 있다.
- [0453] 하나의 시행에서, 제1 양 및 제2 양 간의 제1 차이가 제1 차이 역치 미만인 경우, 예후가 악화될 것으로 예측된다. 다른 시행에서, 제1 수 및 제2 수 간의 제2 차이가 제2 차이 역치 미만인 경우, 예후가 악화될 것으로 예측된다. 역치는 동일하거나 상이할 수 있다. 하나의 구현예에서, 제1 차이 역치 및 제2 차이 역치는 0이다. 따라서 상기 예에 있어서, 메틸화 값 간의 차이는 환자 TBR34에 대해 더 나쁜 예후를 시사할 것이다.
- [0454] 예후는 제1 차이 및/또는 제2 차이가 동일한 역치 또는 각각의 역치 초과인 경우 더 우수할 수 있다. 예후에 대한 분류는 차이가 역치보다 얼마나 위에 또는 아래에 있는지에 의존할 수 있다. 다중 역치를 이용하여 다양한 분류를 제공할 수 있다. 차이가 클수록 더 우수한 결과를 예측할 수 있고, 차이가 작을수록(심지어 음의 값일 수

록) 더 나쁜 결과를 예측할 수 있다.

- [0455] 일부 구현예에서, 다양한 표본을 얻는 시점도 주지된다. 이러한 시간적 파라미터로, 양 변화의 역학 또는 비율을 결정할 수 있다. 하나의 구현예에서, 혈장 중 종양 연관 저메틸화의 빠른 감소 및/또는 혈장 중 종양-연관 CNA의 빠른 감소는 우수한 예후를 예측할 것이다. 반대로 혈장 중 종양 연관 저메틸화의 정적이거나 빠른 증가 및/또는 종양-연관 CNA의 정적이거나 빠른 증가는 나쁜 예후를 예측할 것이다. 메틸화 및 CNA 측정은 임상 결과의 예측을 위한 다른 임상적 파라미터(예로 조영 결과 또는 혈청 생화학 또는 단백질 마커)와 함께 이용될 수 있다.
- [0456] 구현예는 혈장 이외에 다른 표본을 이용할 수 있다. 예를 들어, 종양-연관 메틸화 이상(예로 저메틸화) 및/또는 종양-연관 CNA는 암 환자의 혈액 중에 순환하는 종양 세포로부터, 소변, 대변, 침, 가래, 담액, 체장액, 경부면봉채취물, 생식관으로부터의(예로 질로부터의) 분비물, 복수, 흉수, 정액, 땀 및 눈물 중 무세포 DNA 또는 종양 세포로부터 측정될 수 있다.
- [0457] 다양한 구현예에서, 종양-연관 메틸화 이상(예로 저메틸화) 및/또는 종양-연관 CNA는 유방암, 폐암, 결직장암, 췌장암, 난소암, 코인두암종, 자궁경부암, 흑색종, 뇌 종양 등의 환자 혈액 또는 혈장에서 검출될 수 있다. 실제로 메틸화 및 유전적 변형, 예컨대 CNA가 암에서 일반적인 현상이므로, 기재된 접근은 모든 암 유형에 대해 이용될 수 있다. 메틸화 및 CNA 측정은 임상 결과의 예측을 위한 다른 임상적 파라미터(예로 조영 결과)와 함께 이용될 수 있다. 구현예는 또한 전-암성 병소, 예로 선암종 환자의 스크리닝 및 모니터링을 위해 이용될 수 있다.
- [0458] 따라서 하나의 구현예에서, 상기 생물학적 표본이 치료 전에 얻어지고, CNA 및 메틸화 측정이 치료 후 반복된다. 측정은 결실 또는 증폭을 나타내는 것으로 결정된 영역의 후속 제1 양을 산출할 수 있고, 각각의 영역 컷오프값을 초과하는 영역 메틸화 값을 갖는 것으로 결정된 영역의 후속 제2 양을 산출할 수 있다. 제1 양이 후속 제1 양과 비교되고, 상기 제2 양이 후속 제2 양과 비교되어 유기체의 예후를 결정할 수 있다.
- [0459] 유기체의 예후를 결정하기 위한 비교에는 제1 양 및 후속 제1 양 간의 제1 차이의 결정이 포함될 수 있고, 상기 제1 차이를 하나 이상의 제1 차이 역치와 비교하여 예후를 결정할 수 있다. 유기체의 예후를 결정하기 위한 비교에는 또한 상기 제2 양 및 후속 제2 양 간의 제2 차이의 결정이 포함될 수 있고, 상기 제2 차이를 하나 이상의 제2 차가 역치와 비교할 수 있다. 역치는 0이거나 다른 수일 수 있다.
- [0460] 예후는 제1 차이가 제1 차이 역치 초과일 때보다 제1 차이가 제1 차이 역치 미만일 때 악화될 것으로 예측될 수 있다. 예후는 제2 차이가 제2 차이 역치 초과일 때보다 제2 차이가 제2 차이 역치 미만일 때 악화될 것으로 예측될 수 있다. 치료의 예에는 면역치료법, 수술, 방사선치료법, 화학치료법, 항체 기반 치료법, 유전자 치료법, 후생적 치료법 또는 표적화 치료법이 포함된다.
- [0461] E. 성능
- [0462] 이제 상이한 수의 서열 리드 및 bin 크기에 대한 진단 성능이 CNA 및 메틸화 분석에 대해 설명된다.
- [0463] 1. 서열 리드의 수
- [0464] 하나의 구현예에 따르면, 32명의 건강한 대조군 대상자, 26명의 간세포암종 환자 및 코인두암종, 유방암, 폐암, 신경내분비암 및 평활근육종을 포함하는 20명의 다른 유형의 암 환자의 혈장 DNA를 분석하였다. 32명의 건강한 대상자 중 22명이 기준 군으로 무작위 선택되었다. 이들 22명의 기준 개인의 평균 및 표준 편차(SD)가 메틸화 밀도 및 게놈 표시의 정상 범위를 결정하기 위해 이용되었다. 각 개인의 혈장 표본에서 추출된 DNA가 Illumina 페어-엔드 서열분석 키트를 이용한 서열분석 라이브러리 정립을 위해 이용되었다. 이어서 서열분석 라이브러리는 비메틸화 시토신 잔기를 우라실로 전환하는 바이설파이트 처리를 거쳤다. 각 혈장 표본에 대한 바이설파이트 전환 서열분석 라이브러리가 Illumina HiSeq2000 서열분석기 중 하나의 레인을 이용하여 서열분석되었다.
- [0465] 염기 호출 후, 단편 말단 상에서 어댑터 서열 및 저품질 염기(즉 품질 스코어 <5)가 제거되었다. FASTQ 포맷의 가공 측정치를 Methy-Pipe(P Jiang 등, 2010, IEEE International Conference on Bioinformatics and Biomedicine, doi:10.1109/BIBMW.2010.5703866)로 불리는 메틸화 데이터 분석 파이프라인으로 처리하였다. 바이설파이트 전환 서열분석 측정치를 정렬하기 위해, 먼저 기준 인간 게놈(NCBI build 36/hg19)을 이용하여 왓슨 및 크릭 가닥 상에서 각각 모든 시토신 잔기의 티민으로의 컴퓨터내 전환을 수행하였다. 이어서 모든 처리된 측정치에서 각각의 시토신의 티민으로의 컴퓨터내 전환을 수행하고, 각각의 전환 잔기에 대한 위치 정보를 유지하였다. SOAP2를 이용해서 전환 측정치를 2개의 사전 전환된 기준 인간 게놈에 대해 정렬시켰고(R Li 등, 2009

Bioinformatics 25:1966-1967), 각각의 정렬된 측정치에 대해 최대 2개의 미스매치가 허용되었다. 고유한 게놈 위치에 맵핑 가능한 측정치만 하류 분석을 위해 이용되었다. 왓슨 및 크릭 가닥 모두에 맵핑되는 모호한 측정치 및 중복(클론) 측정치가 제거되었다. CpG 디뉴클레오타이드 맥락에서의 시토신 잔기가 하류 메틸화 분석을 위해 이용되었다. 정렬 후, 서열분석 측정치 상에 원래 존재하는 시토신이 컴퓨터내 전환 동안 유지된 위치 정보에 기반하여 회수되었다. CpG 디뉴클레오타이드 중에서 회수된 시토신은 메틸화된 것으로 스코어링되었다. CpG 디뉴클레오타이드 중의 티민은 비메틸화된 것을 스코어링되었다.

[0466] 메틸화 분석을 위해, 게놈이 동일한 크기의 bin으로 구분되었다. 평가된 bin의 크기에는 50kb, 100kb, 200kb 및 1Mb가 포함된다. 각각의 bin에 대한 메틸화 밀도는 CpG 디뉴클레오타이드 맥락에서의 메틸화 시토신 수를 CpG 위치에서 시토신의 총 수로 나누어 계산되었다. 다른 구현예에서, bin 크기는 게놈에 걸쳐 동일하지 않을 수 있다. 하나의 구현예에서, 동일하지 않은 크기의 이러한 bin 중의 각각의 bin이 다중 대상자에 걸쳐 비교된다.

[0467] 평가 케이스의 혈장 메틸화 밀도가 정상인지를 결정하기 위해, 메틸화 밀도가 기준 군의 결과와 비교되었다. 32명의 건강한 대상자 중 22명이 메틸화 z -스코어(Z_{meth})의 계산을 위한 기준 군으로 무작위 선택되었다.

$$Z_{meth} = \frac{MD_{test} - \overline{MD}_{ref}}{MD_{SD}}$$

[0468]

[0469] 식 중, MD_{test} 는 특정 1Mb bin에 대한 평가 케이스의 메틸화 밀도였고; $\overline{MD}_{ref}$ 는 대응 bin에 대한 기준 군의 평균 메틸화 밀도였고; MD_{SD} 는 대응 bin에 대한 기준 군의 메틸화 밀도의 SD였다.

[0470] CNA 분석을 위해, 각각의 1-Mb bin에 맵핑된 서열분석 측정치의 수가 결정되었다(KCA Chan 등, 2013 Clin Chem 59:211-24). 서열분석 측정치 밀도는 전술된 바와 같은 국소 가중치 부여 산포 그래프 평활화 회귀를 이용하여 GC 바이어스에 대한 수정 후 각각의 bin에 대해 결정되었다(EZ Chen 등, 2011 PLoS One 6: e21791). 혈장 분석을 위해, 평가 케이스의 서열분석 측정치 밀도가 기준 군과 비교되어 CNA z -스코어(Z_{CNA})가 계산되었다:

$$Z_{CNA} = \frac{RD_{test} - \overline{RD}_{ref}}{RD_{SD}}$$

[0471]

[0472] 식 중, RD_{test} 는 특정 1-Mb bin에 대한 평가 케이스의 서열분석 측정치 밀도였고; $\overline{RD}_{ref}$ 는 대응 bin에 대한 기준 군의 평균 서열분석 측정치 밀도였고; RD_{SD} 는 대응 bin에 대한 기준 군의 서열분석 측정치 밀도의 SD였다. bin의 Z_{CNA} 가 <-3 또는 >3인 경우, bin은 CNA를 나타내는 것으로 정의되었다.

[0473] 케이스 별로 평균 9300만 개(범위: 3900만 내지 1억 4200만 개)의 정렬된 측정치의 평균이 획득되었다. 진단 성능에 대한 서열분석 측정치 수의 감소 효과를 평가하기 위해, 각 케이스로부터 1000만 개의 정렬된 측정치를 무작위 선택하였다. 동일한 세트의 기준 개인이 감소된 서열분석 측정치를 갖는 데이터세트에 대해 각각의 1Mb bin의 기준 범위를 정립하기 위해 이용되었다. 유의미한 저메틸화를 나타내는 bin의 백분율, 즉 $Z_{meth} < -3$ 및 CNA를 갖는 bin의 백분율, 즉 $Z_{CNA} < -3$ 또는 >3이 각각의 케이스에 대해 결정되었다. 수신자 조작 특징(ROC) 곡선을 이용하여 케이스 당 1000만 개 측정치 및 1개 레인으로부터의 모든 서열분석 측정치의 데이터세트에 대해 게놈에 걸친 저메틸화 및 CNA 분석의 진단 성능을 나타내었다. ROC 분석에서, 모든 32명의 건강한 대상자가 분석을 위해 이용되었다.

[0474] 도 42는 상이한 수의 서열분석 측정치를 갖는 게놈에 걸친 저메틸화 분석의 진단 성능의 다이어그램을 나타낸다. 저메틸화 분석에 있어서, ROC 곡선에 대한 곡선 하 면적은 케이스 당 1000만 개 측정치 및 1개 레인으로부터의 모든 서열분석 측정치를 분석한 두 데이터세트 간에 유의미하게 상이하지 않았다($P=0.761$). CNA 분석에 있어서, 1000만 개에 대한 하나의 레인의 데이터를 이용하여 서열분석 측정치의 수가 감소된 경우, 곡선 하 면적의 유의미한 감소를 가지며 진단 성능이 악화되었다($P < 0.001$).

[0475] 2. 상이한 bin 크기의 이용 효과

- [0476] 계놈을 1-Mb bin으로 구분하는 것에 부가하여, 더 작은 bin 크기가 이용될 수 있는지도 탐색하였다. 이론적으로 더 작은 bin의 이용은 bin 내의 메틸화 밀도 변동을 잠재적으로 감소시킬 수 있다. 이는 상이한 계놈 영역 간 메틸화 밀도가 크게 변할 수 있기 때문이다. bin이 더 크면, 상이한 메틸화 밀도를 갖는 영역을 포함할 가능성이 증가할 것이고, 이에 따라 bin의 메틸화 밀도 변동의 전체적 증가로 이어질 것이다.
- [0477] 더 작은 bin 크기의 이용이 영역 간 차이에 관련된 메틸화 밀도의 변동을 잠재적으로 감소시킬 수 있지만, 반면에 이는 특정 bin에 맵핑되는 서열분석 측정치의 수를 감소시킬 것이다. 개별 bin에 맵핑되는 측정치의 감소는 샘플링 변동으로 인해 변동성을 증가시킬 것이다. 메틸화 밀도에서 가장 적은 전체적 변동을 줄 수 있는 최적의 bin 크기는 특정 진단 용도의 요건, 예로 이용되는 DNA 서열분석기의 유형 및 표본 별 서열분석 측정치의 총 수에 대해 실험적으로 결정될 수 있다.
- [0478] 도 43은 상이한 bin 크기(50kb, 100kb, 200kb 및 1Mb)를 갖는 계놈에 걸친 저메틸화 분석에 기반하여 암의 검출을 위한 ROC 곡선을 나타내는 다이어그램이다. 나타난 P값은 1Mb의 bin 크기를 갖는 곡선 하 면적 비교를 위한 것이다. bin크기가 1Mb에서 200kb로 감소된 경우 개선 경향이 나타날 수 있다.
- [0479] *F. 누적 확률 스코어*
- [0480] 메틸화 및 CNA에 대한 영역의 양은 다양한 값일 수 있다. 상술된 예는 표본이 암에 연관된 경우 분류를 위한 파라미터로서 유의미한 저메틸화 또는 CNA를 나타낸 이러한 영역의 백분율 또는 컷오프값을 초과하는 영역의 수를 기재하였다. 이러한 접근은 개별 bin에 대한 이상의 크기를 고려하지 않는다. 예를 들어, $Z_{meth} < -3.5$ 의 bin과 $Z_{meth} < -30$ 의 bin은 이들이 둘 다 유의미한 저메틸화를 갖는 것으로 분류될 것이므로 동일할 것이다. 그러나 혈장 중 저메틸화의 변화 정도, 즉 Z_{meth} 값의 크기는 표본 중 암 연관 DNA의 양에 의해 영향을 받으며, 이에 따라 종양 부하를 반영하기 위해 이상을 나타내는 bin의 백분율 정보를 보강할 수 있다. 혈장 표본 중 종양 DNA의 분획 농도가 높을수록 더 낮은 메틸화 밀도로 이어질 것이며, 이는 더 낮은 Z_{meth} 값으로 환산될 것이다.
- [0481] 1. 진단 파라미터로서의 누적 확률 스코어
- [0482] 이상의 크기에서 정보를 이용하기 위해, 누적 확률(CP) 스코어로 불리는 접근을 개발한다. 정규 분포 확률 함수에 기반하여, 각각의 Z_{meth} 값이 이러한 관찰을 할 확률로 환산되었다.
- [0483] CP 스코어는 다음과 같이 계산되었다:
- [0484] $Z_{meth} < -3$ 인 bin (i)에 대해, CP 스코어 = $-\sum \log(\text{Prob}_i)$
- [0485] 식 중, Prob_i 는 3의 자유도를 갖는 스튜던트 t 분포에 따른 bin (i)의 Z_{meth} 에 대한 확률이며, $\log$ 는 자연 로그 함수이다. 다른 구현예에서, 기수가 10(또는 다른 수)인 로그가 이용될 수 있다. 다른 구현예에서, 다른 분포, 예를 들어 비제한적으로 정규 분포 및 감마 분포를 적용하여 z-스코어를 CP로 변환할 수 있다.
- [0486] 더 큰 CP 스코어는 이러한 일탈된 메틸화 밀도를 정상 모집단에서 가질 확률이 더 낮음을 나타낸다. 따라서 높은 CP 스코어는 표본 중 비정상적으로 저메틸화된 DNA, 예로 암 연관 DNA의 존재를 가질 가능성이 더 높음을 나타낼 것이다.
- [0487] 이상을 보이는 bin의 백분율에 비해, CP 스코어 측정은 더 높은 동적 범위를 갖는다. 상이한 환자 간 종양 부하가 크게 변할 수 있으므로, 더 큰 범위의 CP값이 상대적으로 높은 및 상대적으로 낮은 종양 부하를 갖는 환자의 종양 부하를 반영하기 위해 유용할 것이다. 또한, CP 스코어의 이용은 혈장 중 종양-연관 DNA 농도의 변화를 검출하기 위해 잠재적으로 더 민감할 수 있다. 이는 치료 반응의 모니터링 및 예진을 위해 유리하다. 따라서 치료 동안 CP 스코어의 감소는 치료에 대한 우수한 반응을 시사한다. 치료 동안 CP 스코어의 감소 부재 또는 심지어 증가하는 불량한 반응 또는 반응의 부재를 시사할 것이다. 예진에 있어서, 높은 CP 스코어는 높은 종양 부하를 시사하며 불량한 예후(예로 더 높은 사망 또는 종양 진행 가능성)를 제시한다.
- [0488] 도 44a는 이상을 갖는 bin의 누적 확률(CP) 및 백분율에 대한 진단 성능을 나타낸다. 두 유형의 진단 알고리즘에 대한 곡선 하 면적 간에는 유의미한 차이가 없었다($P=0.791$).
- [0489] 도 44b는 전반적 저메틸화, CpG 섬 고메틸화 및 CNA에 대한 혈장 분석을 위한 진단 성능을 나타낸다. 표본 당 하나의 서열분석 레인에서(The University of California, Santa Cruz(UCSC)에서 호스팅하는 데이터베이스에

따라 정의된 저메틸화 분석에 대한 200kb bin 크기 및 CNA와 CpG 섬에 대한 1Mb bin 크기), 모든 세 유형의 분석에 대한 곡선 하 면적은 0.90 초과였다.

[0490] 후속 분석에서, 대조군 대상자에서 가장 높은 CP 스코어는 각각의 세 유형의 분석에 대한 컷오프로 이용되었다. 이들 컷오프의 선택은 100%의 진단 특이성을 제공하였다. 일반적 저메틸화, CpG 섬 고메틸화 및 CNA 분석에 대한 진단 민감성은 각각 78%, 89% 및 52%였다. 46명의 암 환자 중 43명에서, 세 유형의 이상 중 적어도 하나가 검출되었고, 이에 따라 93.4%의 민감성 및 100%의 특이성을 제공하였다. 본 발명자들의 결과는 세 유형의 분석이 암 검출을 위해 상승 효과를 가지고 이용될 수 있음을 시사한다.

[0491] 도 45는 간세포암종 환자에서 전반적 저메틸화, CpG 섬 고메틸화 및 CNA에 대한 결과표를 나타낸다. 세 유형의 분석에 대한 CP 스코어 컷오프값은 각각 960, 2.9 및 211이었다. 양의 CP 스코어 결과에는 진한 글씨로 밑줄을 쳤다.

[0492] 도 46은 간세포암종 이외의 암 환자에서 전반적 저메틸화, CpG 섬 고메틸화 및 CNA에 대한 결과표를 나타낸다. 세 유형의 분석에 대한 CP 스코어 컷오프값은 각각 960, 2.9 및 211이었다. 양의 CP 스코어 결과에는 진한 글씨로 밑줄을 쳤다.

[0493] 2. 암 모니터링을 위한 CP 스코어의 적용

[0494] 치료 이전 및 이후 HCC 환자 TBR34에서 연속 표본이 수집되었다. 표본은 전반적 저메틸화에 대해 분석되었다.

[0495] 도 47은 케이스 TBR34에 있어서 혈장 메틸화에 대한 연속 분석을 나타낸다. 최내부 고리는 버피 코트(검은색) 및 종양 조직(보라색)의 메틸화 밀도를 나타낸다. 혈장 표본에 있어서, Z_{meth} 를 각각의 1Mb bin에 대해 나타낸다. 두 선 간의 차이는 5의 Z_{meth} 차이를 나타낸다. 빨간색 및 회색 점은 기준 군에 비해 저메틸화된 및 메틸화 밀도 변화를 갖지 않는 bin을 나타낸다. 2번째 내부 고리부터 바깥쪽으로는 각각 치료 이전, 종양 절제 후 3일 및 2개월에 얻은 혈장 표본이다. 치료 전에는 높은 정도의 저메틸화가 혈장에서 관찰될 수 있고, 18.5%를 초과하는 bin이 <-10 의 Z_{meth} 를 가졌다. 종양 절제 3일 후에는 저메틸화 정도가 혈장에서 감소된 것이 관찰될 수 있고, 어느 bin도 <-10 의 Z_{meth} 를 갖지 않았다.

표 5

[0496]

케이스 번호	시점	메틸화 분석		
		유의미한 저메틸화를 나타내는 bin의 백분율	누적 확률(CP) 스코어	합산 z-스코어
TBR34	OT 전	62.6%	37,573	14,285
	OT 후 3일	80.5%	17,777	9,195
	OT 후 2개월	40.1%	15,087	5,201

[0497] 표 5는 저메틸화 변화 크기가 종양의 수술적 절제 3일 후에 감소됨에도 불구하고, 이상을 나타내는 bin의 백분율은 역설적 증가를 나타내었음을 나타낸다. 반면, CP 스코어는 혈장 중 저메틸화 정도의 감소를 더 정확히 드러내었고, 종양 부하의 변화를 더 반영할 수 있다.

[0498] OT 2개월 후에는 유의미한 백분율의 bin이 여전히 저메틸화 변화를 나타내었다. CP 스코어도 대략 15,000으로 안정되게 유지되었다. 상기 환자는 이후 3개월째에 잔여 비절제 간에서 다병태성 종양 침적물(이전 수술 시에는 알지 못했음)을 갖는 것으로 진단되었고, 수술 4개월 후에 다중 폐 전이를 갖는 것이 주지되었다. 환자는 수술 8개월 후 전이성 질환으로 사망하였다. 이들 결과는 CP 스코어가 종양 부하의 반영을 위해 이상을 갖는 bin의 백분율보다 더 강력할 수 있음을 제시하였다.

[0499] 전체적으로, CP는 혈장 중 종양 DNA의 양 측정을 필요로 하는 용도를 위해 유용할 수 있다. 이러한 용도의 예에는 암 환자의 예진 및 모니터링(예로 치료에 대한 반응을 관찰하기 위한, 또는 종양 진행을 관찰하기 위한)이 포함된다.

[0500] 합산 z-스코어는 z-스코어의 직접적인, 즉 확률으로의 전환이 없는 합이다. 상기 예에서, 합산 z-스코어는 CP 스코어와 동일한 거동을 나타낸다. 다른 예에서, CP는 CP 스코어에 대한 더 큰 동적 범위로 인해 잔여 질환의 모니터링을 위해 합산 z-스코어보다 더 민감할 수 있다.

- [0501] X. 메틸화에 대한 CNA 영향
- [0502] 분류가 조합되어 제3 분류를 제공하는, 암 수준에 대해 각각의 분류를 결정하기 위한 CNA 및 메틸화의 이용이 전술되었다. 이러한 조합에 부가하여, CNA를 이용하여 메틸화 분석을 위한 컷오프값을 변경하고 상이한 CNA 특징을 갖는 영역 그룹에 대해 메틸화 수준을 비교하여 위양성을 확인할 수 있다. 예를 들어, 초과-존재에 대한 메틸화 수준(예로, $Z_{CNA} > 3$)이 정상 존재에 대한 메틸화 수준(예로, $-3 < Z_{CNA} < 3$)과 비교될 수 있다. 먼저, 메틸화 수준에 대한 CNA의 영향이 설명된다.
- [0503] A. 염색체 획득 및 손실을 갖는 영역에서의 메틸화 밀도 변경
- [0504] 종양 조직은 일반적으로 전체적인 저메틸화를 나타내므로, 암 환자의 혈장 중 종양-유래 DNA의 존재는 비-암 대상자와 비교되는 경우, 메틸화 밀도 감소로 이어질 것이다. 암 환자의 혈장 중 저메틸화 정도는 이론적으로 혈장 표본 중 종양-유래 DNA의 분획 농도와 비례한다.
- [0505] 종양 조직 중 염색체 획득을 나타내는 영역에 있어서, 종양 DNA의 추가 용량이 증폭된 DNA 절편에서 혈장 내로 배출될 것이다. 종양 DNA의 혈장에 대한 상기 증가된 기여는 이론적으로 영향을 받는 영역에 대해 혈장 DNA 중 더 높은 정도의 저메틸화로 이어질 것이다. 추가 요인은 증폭을 나타내는 게놈 영역이 종양 세포에 성장 이점을 부여하는 것이 예상될 것이며, 이에 따라 발현될 것으로 예상된다는 것이다. 이러한 영역은 일반적으로 저메틸화된다.
- [0506] 대조적으로, 종양 조직 중 염색체 손실을 보이는 영역에 있어서, 종양 DNA의 혈장에 대한 감소된 기여는 복제수 변화가 없는 영역에 비해 더 낮은 정도의 저메틸화로 이어질 것이다. 추가 요인은 종양 세포에서 결실된 게놈 영역이 종양 억제인자 유전자를 함유할 수 있고, 이러한 영역이 침묵화된 것이 종양 세포에 유리할 수 있다는 것이다. 따라서 이러한 영역은 더 높은 고메틸화 가능성을 가질 것으로 예상된다.
- [0507] 여기서, 상기 효과를 나타내기 위해 두 HCC 환자(TBR34 및 TBR36)의 결과를 이용한다. 도 48a(TBR36) 및 49a(TBR34)는 염색체 획득 또는 손실을 갖는 영역을 강조하는 원 및 대응하는 메틸화 분석을 나타낸다. 도 48b 및 49b는 환자 TBR36 및 TBR34에 대해 각각 손실, 정상 및 획득에 대한 메틸화 z-스코어 그래프를 나타낸다.
- [0508] 도 48a는 HCC 환자 TBR36에 대해 바이셀파이트-처리 혈장 DNA 중 CNA(내부 고리) 및 메틸화 변화(외부 고리)를 나타내는 Circos 그래프를 나타낸다. 빨간색 원은 염색체 획득 또는 손실을 갖는 영역을 강조한다. 염색체 획득을 나타내는 영역은 복제수 변화가 없는 영역에 비해 더 저메틸화되었다. 염색체 손실을 나타내는 영역은 복제수 변화가 없는 영역에 비해 덜 저메틸화되었다. 도 48b는 HCC 환자 TBR36에 대해 염색체 획득 및 손실을 갖는 영역, 및 복제수 변화가 없는 영역에 대한 메틸화 z-스코어 그래프이다. 복제수 변화가 없는 영역에 비해, 염색체 획득을 갖는 영역은 더 큰 음의 z-스코어를 가졌고(더 저메틸화됨) 염색체 손실을 갖는 영역은 더 작은 음의 z-스코어를 가졌다(덜 저메틸화됨).
- [0509] 도 49a는 HCC 환자 TBR34에 대해 바이셀파이트-처리 혈장 DNA 중 CNA(내부 고리) 및 메틸화 변화(외부 고리)를 나타내는 Circos 그래프를 나타낸다. 도 49b는 HCC 환자 TBR34에 대해 염색체 획득 및 손실을 갖는 영역, 및 복제수 변화가 없는 영역에 대한 메틸화 z-스코어 그래프이다. 환자 TBR36에서의 종양-유래 DNA의 분획 농도가 더 높았으므로, 염색체 획득 및 손실을 갖는 영역 간 메틸화 밀도 차이는 환자 TBR36가 환자 TBR34에서보다 더 컸다.
- [0510] 상기 예에서, CNA를 결정하기 위해 이용된 영역은 메틸화를 결정하기 위해 이용된 영역과 동일하다. 하나의 구현예에서, 각각의 영역 컷오프값은 각각의 영역이 결실 또는 증폭을 나타내는지에 의존한다. 하나의 시행에서, 각각의 영역 컷오프값(예로, 저메틸화를 결정하기 위해 이용된 z-스코어 컷오프)은 증폭이 나타나지 않는 경우에 비해 각 영역이 증폭을 나타내는 경우 더 큰 크기를 가졌다(예로, 크기는 3 초과일 수 있고, -3 미만의 컷오프가 이용될 수 있다). 따라서 저메틸화를 평가하기 위해, 증폭이 나타나지 않는 경우에 비해 각 영역이 증폭을 나타내는 경우 각각의 영역 컷오프값이 더 큰 음의 값을 가질 수 있다. 이러한 시행은 암을 검출하기 위한 평가의 특이성을 개선할 것으로 예상된다.
- [0511] 다른 시행에서, 결실이 나타나지 않는 경우에 비해 각 영역이 결실을 나타내는 경우 각각의 영역 컷오프값은 더 작은 크기(예로, 3 미만)를 갖는다. 따라서 저메틸화를 평가하기 위해, 결실이 나타나지 않는 경우에 비해 각 영역이 결실을 나타내는 경우 각각의 영역 컷오프값은 더 작은 음의 값을 가질 수 있다. 이러한 시행은 암을 검출하기 위한 평가의 민감성을 개선할 것으로 예상된다. 상기 시행에서 컷오프값의 조정은 특히 진단 시나리오를 위해 원하는 민감성 및 특이성에 따라 변할 수 있다. 다른 구현예에서, 메틸화 및 CNA 측정은 암을 예측하기 위

한 다른 임상적 파라미터(예로 조영 결과 또는 혈청 생화학)와 함께 이용될 수 있다.

[0512] B. 영역을 선택하기 위한 CNA의 이용

[0513] 전술된 바와 같이, 혈장 메틸화 밀도가 종양 조직 중 복제수 이상을 갖는 영역에서 변경될 것임을 나타내었다. 종양 조직 중 복제수 획득을 갖는 영역에서, 저메틸화된 종양 DNA의 혈장에 대한 증가된 기여는 복제수 이상이 없는 영역에 비해 혈장 DNA의 더 큰 정도의 저메틸화로 이어질 것이다. 반대로 종양 조직 중 복제수 손실을 갖는 영역에서, 저메틸화된 종양 DNA의 혈장에 대한 감소된 기여는 혈장 DNA의 더 작은 정도의 저메틸화로 이어질 것이다. 혈장 DNA의 메틸화 밀도 및 상대적 제시 간의 이러한 관계는 혈장 DNA 중 암 연관 DNA 및 다른 비-암성 요인(예로, SLE)의 저메틸화의 존재와 연관된 저메틸화 결과를 구별하기 위해 잠재적으로 이용될 수 있다.

[0514] 상기 접근을 예시하기 위해, 2명의 간세포암종(HCC) 환자 및 2명의 암이 없는 SLE 환자의 혈장 표본을 분석하였다. 이들 두 SLE 환자(SLE04 및 SLE10)는 혈장 중 뚜렷한 저메틸화 및 CNA의 존재를 나타내었다. 환자 SLE04에 있어서, 84%의 bin이 저메틸화를 나타내었고 11.2%의 bin이 CNA를 나타내었다. 환자 SLE10에 있어서, 10.3%의 bin이 저메틸화를 나타내었고 5.7%의 bin이 CNA를 나타내었다.

[0515] 도 50a 및 50b는 SLE 환자 SLE04 및 SLE10에 대한 혈장 저메틸화 및 CNA 분석 결과를 나타낸다. 외부의 원은 1Mb 해상에서의 메틸화 z-스코어(Z_{meth})를 나타낸다. 메틸화 $Z_{meth} < -3$ 인 bin은 빨간색이었고 $Z_{meth} > 3$ 인 bin은 회색이었다. 내부의 원은 CNA z-스코어(Z_{CNA})를 나타낸다. 녹색, 빨간색 및 회색 점은 각각 $Z_{CNA} > 3$, < 3 및 -3 내지 3인 bin을 나타낸다. 이들 두 SLE 환자에서, 저메틸화 및 CNA 변화가 혈장 중에서 관찰되었다.

[0516] 메틸화의 변화 및 CNA가 혈장 중 암-유래 DNA의 존재와 일치하는지를 결정하기 위해, $Z_{CNA} > 3$, < -3 및 -3 내지 3인 영역에 대해 Z_{meth} 를 비교하였다. 혈장 중 암-유래 DNA가 기여하는 메틸화 변화 및 CNA에 있어서, $Z_{CNA} < -3$ 인 영역은 덜 저메틸화되고 더 작은 음의 Z_{meth} 를 가질 것으로 예상될 것이다. 대조적으로 $Z_{CNA} > 3$ 인 영역은 더 메틸화되고 더 큰 음의 Z_{meth} 를 가질 것으로 예상될 것이다. 예시적 목적으로, 1측 순위 합산 평가를 적용하여 CNA를 갖는 영역(즉 $Z_{CNA} < -3$ 또는 > 3 인 영역)에 대한 Z_{meth} 를 CNA를 갖지 않는 영역(즉 $Z_{CNA} -3$ 내지 3인 영역)에 대한 Z_{meth} 와 비교하였다. 다른 구현예에서, 다른 통계학적 평가, 예를 들어 비제한적으로 스튜던트 t-테스트, 분산 분석(ANOVA) 테스트 및 크루스칼-왈리스(Kruskal-Wallis) 테스트가 이용될 수 있다.

[0517] 도 51a 및 51b는 두 HCC 환자(TBR34 및 TBR36)의 혈장에 대해 CNA를 갖는 및 갖지 않는 영역에 대한 Z_{meth} 분석을 나타낸다. $Z_{CNA} < -3$ 및 > 3 인 영역은 각각 혈장 중 하향- 또는 상향-제시를 갖는 영역을 나타낸다. TBR34 및 TBR36 모두에서, 혈장 중 하향 제시된 영역(즉 $Z_{CNA} < -3$ 인 영역)은 혈장 중 정상 제시(normal representation)되는 영역(즉 $Z_{CNA} -3$ 내지 3인 영역)에 비해 유의미하게 더 높은 Z_{meth} 를 가졌다(P -값 $< 10^{-5}$, 1측 순위 합산 평가). 정상 제시는 정배수체 게놈에 대해 예상되는 것에 대응한다. 혈장 중 상향 제시되는 영역(즉 $Z_{CNA} > 3$ 인 영역)에 있어서, 이들은 혈장 중 정상 제시되는 영역에 비해 유의미하게 더 낮은 Z_{meth} 를 가졌다(P -값 $< 10^{-5}$, 1측 순위 합산 평가). 이들 모든 변화는 혈장 표본 중 저메틸화 종양 DNA의 존재와 일치하였다.

[0518] 도 51c 및 51d는 두 SLE 환자(SLE04 및 SLE10)의 혈장에 대해 CNA를 갖는 및 갖지 않는 영역에 대한 Z_{meth} 분석을 나타낸다. $Z_{CNA} < -3$ 및 > 3 인 영역은 각각 혈장 중 하향- 또는 상향-제시를 갖는 영역을 나타낸다. SLE04에 있어서, 혈장 중 하향 제시된 영역(즉 $Z_{CNA} < -3$ 인 영역)은 혈장 중 정상 제시되는 영역(즉 $Z_{CNA} -3$ 내지 3인 영역)에 비해 유의미하게 더 높은 Z_{meth} 를 갖지 않았고(P -값=0.99, 1측 순위 합산 평가), 혈장 중 상향 제시된 영역(즉 $Z_{CNA} > 3$ 인 영역)은 혈장 중 정상 제시되는 영역에 비해 유의미하게 더 높은 Z_{meth} 를 갖지 않았다(P -값=0.68, 1측 순위 합산 평가). 이들 결과는 혈장 중 종양-유래 저메틸화 DNA의 존재로 인해 예상되는 변화와 상이하였다. 유사하게 SLE10에 있어서, $Z_{CNA} < -3$ 인 영역은 $Z_{CNA} -3$ 내지 3인 영역에 비해 유의미하게 더 높은 Z_{meth} 를 갖지 않았다(P -값=0.99, 1측 순위 합산 평가).

[0519] SLE 환자에서 Z_{meth} 및 Z_{CNA} 간 전형적인 암 연관 패턴을 갖지 않는 이유는, SLE 환자에서 CNA가 저메틸화를 또한 나타내는 특정 세포 유형을 제시하지 않는 데 있다. 그 대신, 관찰되는 CNA 및 저메틸화의 뚜렷한 존재는 SLE 환자에서 순환 DNA의 변형된 크기 분포에 기인한다. 기준이 건강한 대상자에서 유래되므로, 변형된 크기 분포는

상이한 게놈 영역에 대한 서열분석 측정치 밀도를 잠재적으로 변형하여 뚜렷한 CNA로 이어질 수 있다. 이전 섹션에서 기재된 바와 같이, 순환 DNA 단편의 크기 및 그 메틸화 밀도 간에는 상관관계가 존재한다. 따라서 변형된 크기 분포도 이상 메틸화로 이어질 수 있다.

[0520] $Z_{CNA} > 3$ 인 영역이 $Z_{CNA} - 3$ 내지 3인 영역보다 약간 더 낮은 메틸화 수준을 가졌지만, 비교를 위한 p-값은 두 명의 암 환자에서 관찰된 것에 비해 훨씬 더 높았다. 하나의 구현예에서, p-값은 평가 케이스가 암을 가질 가능성을 결정하기 위한 파라미터로 이용될 수 있다. 다른 구현예에서, 정상 및 이상 제시를 갖는 영역 간 Z_{meth} 차이는 암의 존재 가능성을 시사하기 위한 파라미터로 이용될 수 있다. 하나의 구현예에서, 암 환자 그룹을 이용하여 평가되는 혈장 표본 중 암-유래 저메틸화 DNA의 존재와 변화가 일치함을 나타내기 위해 Z_{meth} 및 Z_{CNA} 간 상관관계를 정립하고 차등적 파라미터에 대한 역치를 결정할 수 있다.

[0521] 따라서 하나의 구현예에서, CNA 분석을 수행하여 모두 결실, 증폭, 또는 정상 제시 중 하나를 나타내는 제1 영역 세트를 결정할 수 있다. 예를 들어, 제1 영역 세트는 모두 결실을 나타내거나, 또는 모두 증폭을 나타내거나, 또는 모두 정상 제시를 나타낼 수 있다(예로, 정상인 영역의 제1 양, 예컨대 정상 Z_{meth} 를 가짐). 메틸화 수준이 상기 제1 영역 세트에 대해 결정될 수 있다(예로, 방법 2800의 제1 메틸화 수준이 제1 영역 세트에 대응할 수 있다).

[0522] CNA 분석은 모두 두 번째로 결실, 증폭, 또는 정상 제시를 나타내는 제2 영역 세트를 결정할 수 있다. 제2 영역 세트는 제1 세트와는 상이하게 나타날 것이다. 예를 들어, 제1 영역 세트가 정상인 경우, 제2 영역 세트는 결실 또는 증폭을 나타낼 수 있다. 제2 메틸화 수준은 상기 제2 영역 세트의 부위에서 메틸화되어 있는 DNA 분자의 각각의 수에 기반하여 계산될 수 있다.

[0523] 이어서 파라미터는 제1 메틸화 수준 및 제2 메틸화 간에 컴퓨터로 연산될 수 있다. 예를 들어, 차이 또는 비가 컴퓨터로 연산되고 컷오프값과 비교될 수 있다. 차이 또는 비는 또한 값을 수득할 가능성을 결정하기 위해 확률 분포(예로 통계 평가의 일환으로)를 거칠 수 있고, 상기 확률을 컷오프값과 비교하여 메틸화 수준에 기반하여 암의 수준을 결정할 수 있다. 이러한 컷오프는 암을 갖는 표본 및 암을 갖지 않는 표본(예로 SLE)을 구별하기 위해 선택될 수 있다.

[0524] 하나의 구현예에서, 메틸화 수준은 제1 영역 세트 또는 영역 혼합(즉 증폭, 결실, 및 정상을 나타내는 영역의 혼합)에 대해 결정될 수 있다. 이어서 상기 메틸화 수준이 분석의 제1 단계의 일환으로 제1 컷오프와 비교될 수 있다. 컷오프를 초과하고, 이에 따라 암의 가능성을 시사하는 경우, 위의 분석을 수행하여 그 시사가 위양성인지 여부를 결정할 수 있다. 따라서 암의 수준에 대한 최종 분류에는 두 메틸화 수준에 대한 파라미터와 제2 컷오프의 비교가 포함될 수 있다.

[0525] 제1 메틸화 수준은 제1 영역 세트의 각 영역에 대해 계산된 영역 메틸화 수준의 통계값(예로, 평균 또는 중앙값)일 수 있다. 제2 메틸화 수준은 또한 제2 영역 세트의 각 영역에 대해 계산된 영역 메틸화 수준의 통계값일 수 있다. 예로서, 통계값은 1측 순위 합산 평가, 스튜던트 t-테스트, 분산 분석(ANOVA) 테스트, 또는 크루스칼-왈리스(Kruskal-Wallis) 테스트를 이용해서 결정될 수 있다.

[0526] XI. 암 유형 분류

[0527] 유기체가 암을 갖는지의 여부를 결정하는 것에 부가하여, 구현예는 표본과 연관된 암의 유형을 확인할 수 있다. 상기 암 유형의 확인은 전반적 저메틸화, CpG 섬 고메틸화, 및/또는 CNA의 패턴을 이용할 수 있다. 패턴에는 측정된 영역 메틸화 수준, 영역에 대한 각각의 CNA값, 및 CpG 섬에 대한 메틸화 수준을 이용하여 알려진 공지 진단으로 환자를 클러스터링하는 것이 수반될 수 있다. 아래 결과는 유사한 유형의 암을 갖는 유기체가 영역 및 CpG 섬에 대해 유사한 값을 가질 뿐만 아니라 비-암 환자도 유사한 값을 가짐을 나타낸다. 클러스터링에서, 영역 또는 섬에 대한 각각의 값은 클러스터링 절차에서 별개의 차원일 수 있다.

[0528] 동일한 유형의 암이 유사한 유전적 및 후생적 변화를 공유할 것임이 알려졌다(E Gebhart 등, 2004 CytoGenet Genome Res; 104: 352-358; PA Jones 등, 2007 Cell; 128: 683-692). 아래에서는, 혈장 중에 검출된 CNA 및 메틸화 변화의 패턴이 암의 기원 또는 유형의 도출을 위해 어떻게 유용한지를 설명한다. HCC 환자, 비-HCC 환자 및 건강한 대조군 대상자로부터의 혈장 DNA 표본이, 예를 들어 계층 클러스터링 분석을 이용하여 분류되었다. 분석은, 예를 들어 R script 패키지(cran.r-project.org/web/packages/gplots/gplots.pdf)의 heatmap.2 기능을 이용하여 수행되었다.

[0529] 상기 접근의 가능성을 예시하기 위해, 혈장 표본의 분류를 위한 유용한 특성을 확인하기 위한 예로 2개 세트의

기준(A군 및 B군)을 이용하였다(표 6 참고). 다른 구현예에서, 다른 기준이 특성 확인을 위해 이용될 수 있다. 이용된 특성에는 1Mb 해상에서의 전반적 CNA, 1Mb 해상에서의 전반적 메틸화 밀도 및 CpG 섬 메틸화가 포함되었다.

표 6

1Mb 해상에서의 전반적 메틸화		
	A군 기준	B군 기준
기준	z-스코어 >3 또는 <-3인 20 초과 의 암 케이스	z-스코어 >2.5 또는 <-2.5인 20 초과의 암 케이스
확인된 특성의 수	584	1,911
CNA 특성		
	A군 기준	B군 기준
기준	z-스코어 >3 또는 <-3인 10 초과 의 암 케이스	z-스코어 >2.5 <-2.5인 10 초과 의 암 케이스
확인된 특성의 수	355	759
CpG 섬 메틸화		
	A군 기준	B군 기준
기준	특정 CpG 섬에서 기준 평균과 2% 상 이한 메틸화 밀도 를 갖는 5 초과 의 암 케이스	특정 CpG 섬에서 기준 평균과 2% 상 이한 메틸화 밀도 를 갖는 1 초과 의 암 케이스
확인된 특성의 수	110	191

[0530]

[0531]

처음 두 예에서, 분류를 위해 CNA, 1Mb 해상에서의 전반적 메틸화 및 CpG 섬 메틸화 특성을 모두 이용하였다. 다른 구현예에서, 다른 기준, 예를 들어 비제한적으로 기준 군의 혈장 중 특성 측정의 정밀도가 이용될 수 있다.

[0532]

도 52a는 355 CNA, 1Mb 해상에서의 584 전반적 메틸화 특성 및 110 CpG 섬의 메틸화 상태를 포함하는 1,130의 모든 A군 특성을 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. 상부측 컬러 막대는 표본군을 나타낸다: 녹색, 파란색 및 빨간색은 각각 건강한 대상자, HCC 및 비-HCC 암 환자를 나타낸다. 일반적으로 세 군의 대상자는 함께 클러스터되는 경향을 보였다. 세로축은 분류 특성을 나타낸다. 상이한 대상자에 걸쳐 유사한 패턴을 갖는 특성이 함께 클러스터링되었다. 이들 결과는 CpG 섬 메틸화 변화, 1Mb 해상에서의 게놈에 걸친 메틸화 변화 및 혈장 중 CNA의 패턴이 일차 원인을 알지 못하는 환자에서 암의 기원을 결정하기 위해 잠재적으로 이용될 수 있음을 제시한다.

[0533]

도 52b는 759 CNA, 1Mb 해상에서의 1,911 전반적 메틸화 특성 및 191 CpG 섬의 메틸화 상태를 포함하는 2,780의 모든 B군 특성을 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. 상부측 컬러 막대는 표본군을 나타낸다: 녹색, 파란색 및 빨간색은 각각 건강한 대상자, HCC 및 비-HCC 암 환자를 나타낸다. 일반적으로 세 군의 대상자는 함께 클러스터되는 경향을 보였다. 세로축은 분류 특성을 나타낸다. 상이한 대상자에 걸쳐 유사한 패턴을 갖는 특성이 함께 클러스터링되었다. 이들 결과는 CpG 섬 메틸화 변화, 1Mb 해상에서의 게놈에 걸친 메틸화 변화 및 혈장 중 CNA의 패턴이 일차 원인을 알지 못하는 환자에서 암의 기원을 결정하기 위해 이용될 수 있음을 제시한다. 분류 특성의 선택은 구체적인 용도를 위해 조정될 수 있다. 또한, 상이한 유형의 암에 대한 대상자의 이전 확률에 따라 암 유형 예측에 가중치가 주어질 수 있다. 예를 들어, 만성 바이러스성 간염 환자는 간세포암종이 발생하기 쉽고, 만성 흡연자는 폐암이 발생하기 쉽다. 따라서 가중치가 부여된 암 유형의 확률이, 예를 들어 비제한적으로 논리적, 다중 또는 클러스터링 회귀를 이용하여 계산될 수 있다.

[0534]

다른 구현예에서, 단일 유형의 특성이 분류 분석을 위해 이용될 수 있다. 예를 들어 하기 예에서, 1Mb 해상에서의 전반적 메틸화, CpG 섬 고메틸화 또는 1Mb 해상에서의 CNA만 계층 클러스터링 분석을 위해 이용되었다. 상이한 특성이 이용되는 경우 구별력이 상이할 수 있다. 분류 특성의 추가 정련이 분류 정확성을 잠재적으로 개선할 수 있다.

- [0535] 도 53a는 A군의 CpG 섬 메틸화 특성을 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. 일반적으로, 암 환자가 함께 클러스터링되었고 비-암 대상자는 다른 클러스터에 있었다. 그러나 HCC 및 비-HCC 환자는 모든 세 유형의 특성에 비해 덜 분리되었다.
- [0536] 도 53b는 분류 특성으로 1Mb 해상에서의 A군의 전반적 메틸화 밀도를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. HCC 및 비-HCC 환자의 우선적 클러스터링이 관찰되었다.
- [0537] 도 54a는 분류 특성으로 1Mb 해상에서의 A군의 전반적 CNA를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. HCC 및 비-HCC 환자의 우선적 클러스터링이 나타났다.
- [0538] 도 54b는 분류 특성으로 B군의 CpG 섬 메틸화 밀도를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. HCC 및 비-HCC 암 환자의 우선적 클러스터링이 관찰될 수 있었다.
- [0539] 도 55a는 분류 특성으로 1Mb 해상에서의 B군의 전반적 메틸화 밀도를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. HCC 및 비-HCC 암 환자의 우선적 클러스터링이 관찰될 수 있었다.
- [0540] 도 55b는 분류 특성으로 1Mb 해상에서의 B군의 전반적 CNA를 이용한 HCC 환자, 비-HCC 암 환자 및 건강한 대조군 대상자로부터의 혈장 표본에 대한 계층 클러스터링 분석을 나타낸다. HCC 및 비-HCC 암 환자의 우선적 클러스터링이 관찰될 수 있었다.
- [0541] 혈장 표본에 대한 이들 계층 클러스터링 결과는 상이한 특성의 조합이 일차 암 유형의 확인을 위해 잠재적으로 이용될 수 있음을 제시한다. 선택 기준의 추가 정련은 분류의 정확성을 잠재적으로 추가 개선할 수 있다.
- [0542] 따라서 하나의 구현예에서, 메틸화 분류가 상기 유기체에 있어 암이 존재함을 시사하는 경우, 유기체에 연관된 암의 유형은 메틸화 수준(예로 방법 2800로부터의 제1 메틸화 또는 임의 영역의 메틸화 수준)을 다른 유기체(즉 동일한 유형, 예컨대 인간의 다른 유기체)로부터 결정된 대응값과 비교하여 확인될 수 있다. 대응값은 메틸화 수준이 계산되는 동일한 영역 또는 부위 세트에 대한 것일 수 있다. 적어도 두 명의 다른 유기체가 상이한 유형의 암을 갖는 것으로 확인된다. 예를 들어, 대응값은 클러스터로 조직될 수 있고, 여기서 클러스터 중 2 개는 상이한 암에 연관된다.
- [0543] 또한 CNA 및 메틸화가 암 수준의 제3 분류를 획득하기 위해 함께 이용되는 경우, CNA 및 메틸화 특성은 다른 유기체로부터의 대응값과 비교될 수 있다. 예를 들어, 결실 또는 증폭을 나타내는 영역의 제1 양(예로, 도 36의)을 다른 유기체로부터 결정된 대응값과 비교하여 상기 유기체와 연관된 암 유형을 확인할 수 있다.
- [0544] 일부 구현예에서, 메틸화 특성은 게놈의 복수 영역의 영역 메틸화 수준이다. 각각의 영역 컷오프값을 초과하는 영역 메틸화 수준을 갖는 것으로 결정된 영역이 이용될 수 있다. 예로 유기체의 영역 메틸화 수준이 게놈의 동일한 영역에 대한 다른 유기체의 영역 메틸화 수준과 비교될 수 있다. 비교는 암 유형을 구별할 수 있도록 하거나 단지 암을 확인하기 위한(예로 위양성을 확인하기 위한) 추가 필터를 제공할 수 있다. 따라서 상기 비교에 기반하여 유기체가 제1 유형의 암, 암의 부재 또는 제2 유형의 암을 갖는지가 결정될 수 있다.
- [0545] 다른 유기체(평가되는 것과 함께)가 영역 메틸화 수준을 이용해서 클러스터링될 수 있다. 따라서 영역 메틸화 수준의 비교를 이용하여 유기체가 어느 클러스터에 속하는지를 결정할 수 있다. 클러스터링은 또한 상술된 바와 같이 결실 또는 증폭을 나타내는 것으로 결정된 영역에 대한 CNA 정규화값을 이용할 수 있다. 또한 클러스터링은 고메틸화 CpG 섬의 각각의 메틸화 밀도를 이용할 수 있다.
- [0546] 상기 방법의 원리를 예시하기 위해, 두 미지의 표본 분류를 위해 논리 회귀를 이용하는 예를 나타낸다. 이 분류의 목적은 이들 두 표본이 HCC 또는 비-HCC 암인지를 결정하기 위한 것이었다. HCC 환자에서 수집된 23개 혈장 표본 및 HCC 이외 암 환자의 18개 표본을 포함하는 표본 교육 세트가 컴파일링되었다. 따라서 교육 세트에는 총 41 케이스가 있었다. 상기 예에서, CpG 섬(X1-X5)의 메틸화에 대한 5 특성(X1-X5), 1Mb 영역의 메틸화에 대한 6 특성(X6-X11) 및 1Mb 영역의 CNA에 대한 2 특성(X12-X13)을 포함하는 13 특성이 선택되었다. CpG 메틸화 특성은 >3 또는 <-3 의 z -스코어를 갖는 교육 세트에서 적어도 15 케이스의 기준에 기반하여 선택되었다. 1-Mb 메틸화 특성은 >3 또는 <-3 의 z -스코어를 갖는 교육 세트에서 적어도 39 케이스의 기준에 기반하여 선택되었다. CNA 특성은 >3 또는 <-3 의 z -스코어를 갖는 적어도 20 케이스의 기준에 기반하여 선택되었다. 각 특성(X1-X13)에 대한

회귀 계수를 결정하기 위해, 상기 교육 세트의 표본 상에서 논리 회귀가 수행되었다. 더 큰 크기의 회귀 계수를 갖는 특성(이것이 양의 관점인지 또는 음의 관점인지와 무관하게)이 HCC 및 비-HCC 표본 간에 더 우수한 구별을 제공하였다. 각각의 특성에 대한 각 케이스의 z -스코어가 독립 변수의 입력 값으로 이용되었다. 이어서 HCC 환자로부터 하나(TBR36) 및 폐암 환자로부터 하나(TBR177)의 두 혈장 표본이 13 특성에 대해 분석되었다.

[0547] 상기 암 유형 분류 분석에서, 이들 두 표본은 미지 기원의 암 환자에서 수집되는 것으로 가정되었다. 각 표본에 있어서, 상기 각각의 특성에 대한 z -스코어가 교차비의 자연 로그($\ln(\text{교차비})$)를 결정하기 위한 논리 회귀 공식에 대입되었으며, 여기서 교차비는 HCC를 갖는 및 HCC를 갖지 않는 확률의 비(HCC/비-HCC)를 나타내었다.

[0548] 표 7은 논리 회귀 공식의 13 특성에 대한 회귀 계수를 나타낸다. 두 평가 케이스(TBR36 및 TBR177)의 각각의 특성에 대한 z -스코어도 나타낸다. TBR36 및 TBR177에 대한 HCC의 $\ln(\text{교차비})$ 는 각각 37.03 및 -4.37이었다. 이들 교차비로부터, 혈장 표본이 HCC 환자에서 수집될 확률은 각각 >99.9% 및 1%로 계산되었다. 요약하면, TBR36은 HCC 환자로부터의 표본일 가능성이 높았던 반면, TBR177은 HCC 환자로부터의 표본일 가능성이 낮았다.

표 7

특성	회귀 계수	각 특성의 z -스코어	
		TBR36	TBR177
X1	-2.9575	14.8	0
X2	2.2534	21.3	0
X3	-1.5099	6.1	0
X4	-0.236	34.0	0
X5	0.7426	17.3	0
X6	-0.6682	-26.3	-1.5
X7	-0.2828	-13.9	-2.6
X8	-0.7281	-9.4	-4.4
X9	1.0581	-7.8	-3.7
X10	0.3877	-20.8	-4.3
X11	0.3534	-15.5	-3.1
X12	-1.1826	4.8	3.3
X13	-0.3805	-11.7	-1.4
	$\ln(\text{교차비})$	37.03	-4.37463

[0550] 다른 구현예에서, 계층 클러스터링 회귀, 분류 트리 분석 및 다른 회귀 모델이 암의 가능한 일차 기원을 결정하기 위해 이용될 수 있다.

[0551] XII. 재료 및 방법

[0552] A. 바이설파이트-처리 DNA 라이브러리의 제조 및 서열분석

[0553] 0.5%(w/w) 비메틸화 람다 DNA(Promega)가 첨가된 게놈 DNA(5 μ g)를 Covaris S220 시스템(Covaris)에 의해 대략 200bp 길이로 단편화하였다. 메틸화 어댑터(Illumina)를 DNA 단편에 결합한 것을 제외하고는 제조업체의 지시에 따라 페어-엔드 서열분석 표본 제조 키트(Illumina)를 이용하여 DNA 라이브러리를 제조하였다. AMPure XP 자기 비드(Beckman Coulter)를 이용한 2회 정제 후, 결합 산물을 2개 부분으로 분할하여 그 중 하나를 EpiTect 바이설파이트 키트(Qiagen)로 2회의 바이설파이트 개질을 거쳤다. 삽입물 중 CpG 부위에서의 비메틸화 시토신은 우라실로 전환된 반면, 메틸화 시토신은 변하지 않고 유지되었다. 나트륨 바이설파이트로 처리한 또는 처리하지 않은 어댑터가 결합된 DNA 분자를 다음 조성을 이용하여 10사이클의 PCR로 농축하였다: 50 μ l-반응에서 2.5U PfuTurboCx 핫스타트 DNA 폴리머라아제(Agilent Technologies), 1X PfuTurboCx 반응 완충액, 25 μ M dNTPs, 1 μ l PCR 프라이머 PE 1.0 및 1 μ l PCR 프라이머 PE 2.0(Illumina). 열 사이클링 프로파일은 다음과 같았다: 95 $^{\circ}$ C에서 2분, 98 $^{\circ}$ C에서 30초, 이어서 98 $^{\circ}$ C에서 15초, 60 $^{\circ}$ C에서 30초 및 72 $^{\circ}$ C에서 4분을 10사이클, 마지막 단계로 72 $^{\circ}$ C에서 10분(R Lister, 등, 2009 Nature; 462: 315-322). PCR 산물을 AMPure XP 자기 비드를 이용하여 정제하였다.

[0554] 3.2-4ml의 모체 혈장 표본에서 추출한 혈장 DNA를 단편화된 람다 DNA(25pg/혈장ml)과 함께 스파이크 처리하고 전술된 바와 같이 라이브러리 정립을 거쳤다(RWK Chiu 등, 2011 BMJ; 342: c7401). 메틸화 어댑터로의 결합 후, 결합 산물을 2개로 절반씩 분할하고, 1개 부분으로 2회의 바이설파이트 개질을 거쳤다. 이어서 바이설파이트-처

리 또는 미처리 결찰 산물을 전술된 바와 같이 10사이클의 PCR로 농축하였다.

[0555] 바이설파이트-처리 또는 미처리 DNA 라이브러리를 HiSeq2000 장비(Illumina) 상에서 페어-엔드 포맷으로 75bp에 대해 서열분석하였다. DNA 클러스터를 cBot 장비(Illumina) 상에서 페어-엔드 클러스터 생성 키트 v3으로 생성하였다. 실시간 영상 분석 및 염기 호출을 HiSeq 제어 소프트웨어(HCS) v1.4 및 실시간 분석(RTA) 소프트웨어 v1.13(Illumina)을 이용해서 수행하였고, 자동화 매트릭스 및 위상 계산은 DNA 라이브러리로 서열분석된 스파이크-인 PhiX 제어 v3에 기반하였다.

[0556] B. 메틸화 시토신의 서열 정렬 및 확인

[0557] 염기 호출 후, 단편 말단 상의 어댑터 서열 및 저품질 염기(즉 품질 스코어 <20)를 제거하였다. 이어서 FASTQ 포맷의 가공 측정치를 Methy-Pipe(P Jiang, 등, Methy-Pipe: 전체 게놈 메틸롬 분석을 위한 통합 바이오인포매틱스 데이터 분석 파이프라인, 논문은 2010. 12. 18~21, IEEE International Conference on Bioinformatics and Biomedicine Workshops, Hong Kong에서 발표됨)로 불리는 메틸화 데이터 분석 파이프라인으로 처리하였다. 바이설파이트 전환 서열분석 측정치를 정렬하기 위해, 먼저 기준 인간 게놈(NCBI build 36/hg18)을 이용하여 각각 왓슨 및 크릭 가닥 상에서 모든 시토신 잔기의 티민으로의 컴퓨터내 전환을 수행하였다. 이어서 모든 처리된 측정치에서 각각의 시토신의 티민으로의 컴퓨터내 전환을 수행하고, 각각의 전환된 잔기의 위치 정보를 유지하였다. SOAP2(R Li, 등, 2009 Bioinformatics; 25: 1966-1967)를 이용하여 전환된 측정치를 2 사전 전환된 기준 인간 게놈에 대해 정렬하고, 각각의 정렬된 측정치에 대해 최대 2 미스매치를 허용하였다. 고유한 게놈 위치에 맵핑 가능한 측정치만 선택하였다. 왓슨 및 크릭 가닥 모두에 맵핑되는 모호한 측정치 및 동일한 시작 및 끝 게놈 위치를 가진 중복된(클론) 측정치를 제거하였다. 삽입물 크기가 ≤600bp인 서열분석 측정치를 메틸화 및 크기 분석을 위해 보유하였다.

[0558] CpG 디뉴클레오타이드 맥락의 시토신 잔기가 하류 DNA 메틸화 연구를 위한 주요 표적이었다. 정렬 후, 원래 서열 분석 측정치에 존재하는 시토신을 컴퓨터내 전환 동안 유지된 위치 정보에 기반하여 회수하였다. CpG 디뉴클레오타이드 중에서 회수된 시토신을 메틸화된 것으로 스코어링하였다. CpG 디뉴클레오타이드 중의 티민은 비메틸화된 것으로 스코어링하였다. 라이브러리 제조 동안 포함된 비메틸화 람다 DNA는 나트륨 바이설파이트 개질의 효율 산정을 위한 내부 대조군으로 작용하였다. 바이설파이트 전환 효율이 100%인 경우, 람다 DNA 상의 모든 시토신이 티민으로 전환되어야 한다.

[0559] XIII. 요약

[0560] 본원에 기재된 구현예를 이용하여, 예를 들어 대상자의 혈장을 이용하여 비침습적으로 암을 스크리닝, 검출, 모니터링 또는 예진할 수 있었다. 또한 모체 혈장으로부터 태아 DNA의 메틸화 프로필을 추정하여 태아의 산전 스크리닝, 진단, 연구 또는 모니터링을 수행할 수 있었다. 접근의 파워를 예시하기 위해, 태반 조직의 연구를 통해 통상적으로 수득된 정보가 모체 혈장에서 직접 평가될 수 있음을 나타내었다. 예를 들어, 유전자 유전자좌의 각인 상태, 태아 및 모체 DNA 간 차등적 메틸화를 갖는 유전자좌의 확인 및 유전자 유전자좌의 메틸화 프로필의 임신 중 변동이 모체 혈장 DNA의 직접적 분석을 통해 달성되었다. 본 발명자들의 접근의 주요 장점은 태아 메틸롬이 임신에 대한 손상 또는 태아 조직의 침습적 표본 채집이 필요 없이 임신 동안 통합적으로 평가될 수 있다는 것이다. 변경된 DNA 메틸화 상태 및 여러 임신 연관 상태 간의 기지의 상관관계에 있어서, 본 연구에 기재된 접근은 이들 상태에 대한 병태생리의 조사 및 바이오마커의 확인을 위한 중요한 도구로 작용할 수 있다. 각인된 유전자좌에 초점을 맞추으로써, 모계 전달된 태아 메틸화 프로필뿐만 아니라 부계 전달된 것이 모두 모체 혈장에서 평가될 수 있음을 나타내었다. 상기 접근은 각인 질환의 조사를 위해 잠재적으로 유용할 수 있다. 구현예는 또한 태아의 산전 평가 또는 임신 연관 질환에 대해 직접 적용될 수 있다.

[0561] 게놈에 걸친 바이설파이트 서열분석이 태반 조직의 DNA 메틸화 프로필을 조사하기 위해 적용될 수 있음을 나타내었다. 인간 게놈에는 대략 2800만 개의 CpG 부위가 존재한다(C Clark 등, 2012 PLoS One; 7: e50233). CVS 및 만기 태반 조직 표본에 대한 본 발명자들의 바이설파이트 서열분석 데이터는 CpG를 80% 초과 커버하였다. 이는 다른 고처리량 플랫폼을 이용하여 달성 가능한 것보다 실질적으로 더 넓은 커버 범위를 나타낸다. 예를 들어, 태반 조직에 대한 이전 연구(T Chu 등, 2011 PLoS One; 6: e14723)에서 이용된 Illumina Infinium HumanMethylation 27K 비드칩 어레이는 게놈에서 CpG의 0.1%만을 커버하였다. 보다 최근에 이용 가능해진 Illumina Infinium HumanMethylation 450K 비드칩 어레이는 CpG의 1.7%만을 커버하였다(C Clark 등, 2012 PLoS One; 7: e50233). MPS 접근에는 프로브 설계, 혼성화 효율 또는 항체 포획 강도에 관련된 제한이 없으므로, CpG 섬 내 또는 그를 넘어선 그리고 대부분의 서열 맥락에서의 CpG가 평가될 수 있다.

- [0562] XIV. 컴퓨터 시스템
- [0563] 본원에서 언급된 임의의 컴퓨터 시스템은 임의의 적합한 수의 서브시스템을 이용할 수 있다. 이러한 서브시스템의 예를 도 33에서 컴퓨터 장치 3300에 나타낸다. 일부 구현예에서, 컴퓨터 시스템에는 단일 컴퓨터 장치가 포함되며, 여기서 서브시스템은 컴퓨터 장치의 컴포넌트일 수 있다. 다른 구현예에서, 컴퓨터 시스템에는 다중 컴퓨터 장치가 포함될 수 있고, 각각이 내부 컴포넌트를 갖는 서브시스템이다.
- [0564] 도 33에 나타난 서브시스템은 시스템 버스 3375를 통해 상호연결된다. 추가적인 서브시스템, 예컨대 프린터 3374, 키보드 3378, 저장 장치(들) 3379, 어댑터 3382를 디스플레이하도록 커플링된 모니터 3376 등을 나타낸다. I/O 컨트롤러 3371에 커플링된 주변 및 입력/출력(I/O) 장치는 당분야에 공지된 임의의 수의 수단, 예컨대 시리얼 포트 3377에 의해 컴퓨터 시스템에 연결될 수 있다. 예를 들어, 시리얼 포트 3377 또는 외부 인터페이스 3381(예로 이더넷, 와이파이 등)을 이용하여 컴퓨터 시스템 3300을 광역 네트워크, 예컨대 인터넷, 마우스 입력 장치 또는 스캐너에 연결할 수 있다. 시스템 버스 3375를 통한 상호연결은 중앙 프로세서 3373이 각각의 서브시스템과 통신하고 시스템 메모리 3372 또는 저장 장치(들) 3379(예로 고정 디스크)로부터의 지시 수행뿐만 아니라 서브시스템 간 정보 교환을 제어할 수 있도록 한다. 시스템 메모리 3372 및/또는 저장 장치(들) 3379는 컴퓨터에서 읽을 수 있는 매체를 구현할 수 있다. 본원에 언급된 임의의 값은 하나의 컴포넌트로부터 다른 컴포넌트로 출력될 수 있고 사용자에게 출력될 수 있다.
- [0565] 컴퓨터 시스템에는, 예로 외부 인터페이스 3381에 의해 또는 내부 인터페이스에 의해 함께 연결된, 복수의 동일한 컴포넌트 또는 서브시스템이 포함될 수 있다. 일부 구현예에서, 컴퓨터 시스템, 서브시스템, 또는 장치는 네트워크를 통해 통신할 수 있다. 이러한 상황에서, 하나의 컴퓨터가 클라이언트로 그리고 다른 컴퓨터가 서버로 간주될 수 있고, 여기서 각각은 동일한 컴퓨터 시스템의 일부일 수 있다. 클라이언트 및 서버에는 각각 다중 시스템, 서브시스템, 또는 컴포넌트가 포함될 수 있다.
- [0566] 본 발명의 임의의 구현예가 모듈 또는 집적 방식으로 하드웨어(예로 어플리케이션 특이적 집적 회로 또는 현장에서 프로그래밍 가능한 게이트 어레이)를 이용해서 및/또는 일반적으로 프로그래밍 가능한 프로세서를 갖는 컴퓨터 소프트웨어를 이용해서 제어 논리의 형태로 시행될 수 있음이 이해되어야 한다. 본원에서 이용된 바와 같이, 프로세서에는 동일한 집적 칩 상의 멀티 코어 프로세서 또는 단일 회로판 상의 또는 네트워크화된 다중 프로세싱 유닛이 포함된다. 본원에 제공된 개시 및 교시에 기반하여, 당업자는 하드웨어 및 하드웨어와 소프트웨어의 조합을 이용하여 본 발명의 구현예를 시행하기 위한 다른 방식 및/또는 방법을 알고 이해할 것이다.
- [0567] 본 출원에 기재된 임의의 소프트웨어 컴포넌트 또는 기능은, 예를 들어 통상적인 또는 오브젝트형 기법을 이용하여 임의의 적합한 컴퓨터 언어, 예를 들어 Java, C++ 또는 Perl을 이용하여 프로세서에 의해 시행될 소프트웨어 코드로 시행될 수 있다. 소프트웨어 코드는 저장 및/또는 전송을 위해 컴퓨터에서 읽을 수 있는 매체 상에 일련의 명령 또는 명령으로 저장될 수 있고, 적합한 매체에는 램(random access memory, RAM), 롬(read only memory, ROM), 자기 매체, 예컨대 하드 드라이브 또는 플로피 디스크, 또는 광 매체, 예컨대 콤팩트 디스크(CD) 또는 DVD(digital versatile disk), 플래시 메모리 등이 포함된다. 컴퓨터 컴퓨터로 판독 가능한 매체는 이러한 저장 또는 전송 장치의 임의의 조합일 수 있다.
- [0568] 이러한 프로그램은 또한 인터넷을 포함하는 다양한 프로토콜에 적합한 유선, 광, 및/또는 무선 네트워크를 통한 전송을 위해 채택된 캐리어 신호를 이용하여 인코딩되고 전송될 수 있다. 이와 같이, 본 발명의 구현예에 따른 컴퓨터로 판독 가능한 매체는 이러한 프로그램으로 인코딩되는 데이터 신호를 이용하여 생성될 수 있다. 프로그램 코드로 인코딩되는 컴퓨터로 판독 가능한 매체는 호환 장치와 패키지화되거나 다른 장치에서 별도로 제공될 수 있다(예로 인터넷 다운로드를 통해). 임의의 이러한 컴퓨터로 판독 가능한 매체는 단일 컴퓨터 프로그램 제품(예로 하드 드라이브, CD 또는 전체 컴퓨터 시스템) 상에 또는 내에 존재할 수도 있고, 시스템 또는 네트워크 내에서 상이한 컴퓨터 프로그램 제품 상에 또는 내에 존재할 수도 있다. 컴퓨터 시스템에는 본원에서 언급된 임의의 결과를 사용자에게 제공하기 위한 모니터, 프린터, 또는 다른 적합한 디스플레이가 포함될 수 있다.
- [0569] 본원에 기재된 임의의 방법은 하나 이상의 프로세서를 포함하는 컴퓨터 시스템으로 완전히 또는 부분적으로 수행될 수 있고, 이는 단계들을 수행하기 위해 구성될 수 있다. 따라서, 구현예는 잠재적으로 각각의 단계 또는 각각의 단계군을 수행하는 상이한 컴포넌트를 갖는, 본원에 기재된 임의의 방법 단계를 수행하기 위해 구성된 컴퓨터 시스템에 관한 것일 수 있다. 번호가 지정된 단계로 제시되었으나, 본원의 방법 단계는 동시에 또는 상이한 순서로 수행될 수 있다. 또한 이들 단계의 일부가 다른 방법으로부터의 다른 단계의 일부와 함께 이용될 수도 있다. 또한 단계의 전부 또는 일부가 선택적일 수 있다. 또한 임의의 방법의 임의의 단계가 모듈, 회로 또는 이들

단계를 수행하기 위한 다른 수단으로 수행될 수 있다.

[0570] 특정 구현예의 구체적 상세내용은 본 발명의 구현예의 요지 및 범위에서 벗어나지 않고 임의의 적합한 방식으로 조합될 수 있다. 그러나 본 발명의 다른 구현예는 각각의 개별 측면에 관한 특정 구현예 또는 이들 개별 측면의 특정 조합에 관한 것일 수 있다.

[0571] 본 발명의 예시적 구현예의 상기 설명은 예시 및 설명의 목적을 위해 제시되었다. 이는 본 발명을 기재된 정확한 형태로 제한하거나 배타적이기 위한 것이 아니며, 여러 개질 및 변형이 상기 교시의 관점에서 가능하다. 구현예는 본 발명의 이론 및 그 실제 적용을 가장 잘 설명하여 당분야의 다른 숙련자가 본 발명을 예시된 특정 용도에 적합하게 다양한 구현예 및 다양한 개질로 가장 잘 이용할 수 있도록 하기 위해 선택되고 기재되었다.

[0572] 단수의 언급은 구체적으로 반대로 나타내지 않는 이상 "하나 이상"을 의미하는 것이다.

[0573] 본원에서 언급된 모든 특허, 특허 출원, 공보 및 기재는 모든 목적을 위해 이들의 전문이 참조로 도입된다. 어느 것도 선행 기술로 인정되지는 않는다.

[0574] [표 S2A]

[0575] 제1 식달 용모막 용모 표본 및 모체 혈액 세포에서 확인된 상위 100개의 고메틸화 영역의 목록

염색체	시작	끝	크기 (bp)	모체 혈액 세포	CVS	P-값	메틸화 차이
chr13	113063600	113064100	500	0.009	0.9	3.67E-15	0.891
chr6	36279700	36280200	500	0.0068	0.8957	2.39E-22	0.8889
chr16	66876000	66876500	500	0.0327	0.9211	3.82E-21	0.8884
chr10	163500	164000	500	0.0195	0.9034	3.60E-35	0.8839
chr9	3518300	3518800	500	0.0263	0.9045	1.32E-26	0.8782
chr12	31877100	31877600	500	0.007	0.8784	3.08E-22	0.8714
chr22	37477400	37478400	1000	0.0152	0.8848	0.00E+00	0.8696
chr4	148940500	148941000	500	0.0055	0.8717	4.40E-29	0.8662
chr5	131836300	131836800	500	0.075	0.9403	1.54E-10	0.8653
chr17	26661700	26663600	1900	0.0187	0.875	2.95E-38	0.8563
chr2	105758600	105759600	1000	0.0305	0.8828	1.19E-53	0.8523
chr22	39188800	39189800	1000	0	0.8514	2.05E-46	0.8514
chr3	153443900	153444900	1000	0.0436	0.8945	5.43E-34	0.8509
chr6	25149600	25150600	1000	0.0135	0.8632	0.00E+00	0.8497
chr5	98296800	98297300	500	0.0432	0.8925	4.97E-23	0.8493
chr7	150679900	150680400	500	0.0496	0.8944	6.50E-17	0.8448
chr7	107563100	107563600	500	0.0495	0.8895	9.58E-26	0.84
chr7	37348300	37349300	1000	0.0012	0.8409	0.00E+00	0.8397
chr14	58837800	58838300	500	0.0097	0.848	3.35E-16	0.8383
chr6	119238100	119238600	500	0.0899	0.928	2.38E-19	0.8381
chr15	93669900	93670400	500	0.0753	0.913	2.19E-10	0.8377
chr17	26669200	26670200	1000	0.0221	0.859	1.44E-29	0.8369
chr2	88108100	88108600	500	0.075	0.9109	3.55E-17	0.8359
chr13	98363800	98364300	500	0.11	0.9457	1.28E-11	0.8357
chr16	66948000	66948500	500	0.0331	0.8685	0.00E+00	0.8354
chr6	42098000	42098500	500	0.0484	0.8835	3.73E-16	0.8351
chr3	129876000	129876500	500	0.0565	0.8897	8.81E-17	0.8332
chr3	142700300	142700800	500	0.0063	0.8393	2.59E-22	0.833
chr8	145883800	145884300	500	0.0392	0.872	0.00E+00	0.8328
chr10	8320700	8321200	500	0.0566	0.8871	9.40E-09	0.8305
chr3	120438100	120438600	500	0.102	0.9292	7.09E-16	0.8272
chr3	173792600	173793100	500	0.0182	0.8453	2.84E-39	0.8271
chr17	40320700	40321200	500	0.0539	0.8788	6.50E-30	0.8249
chr15	72076200	72076700	500	0.0299	0.8525	6.48E-10	0.8226
chr16	29663900	29665400	1500	0.0081	0.8305	0.00E+00	0.8224
chr11	66961100	66962100	1000	0.0489	0.8712	0.00E+00	0.8223
chr9	27083100	27084100	1000	0.097	0.9177	2.37E-55	0.8207
chr9	111249600	111250100	500	0.0613	0.8795	1.99E-20	0.8182
chr14	101412400	101412900	500	0	0.8167	8.26E-32	0.8167
chr1	242549200	242549700	500	0	0.8155	3.50E-21	0.8155
chr8	38642800	38643300	500	0.0191	0.8346	3.22E-41	0.8155
chr4	85893600	85894100	500	0.0394	0.8533	1.45E-15	0.8139
chr5	142368600	142369100	500	0.0385	0.8523	1.18E-18	0.8138
chr8	130969500	130970000	500	0.069	0.8824	2.42E-24	0.8134
chr2	196783900	196784400	500	0	0.8123	3.63E-40	0.8123
chr16	40258100	40258600	500	0.0733	0.8851	4.29E-18	0.8118

[0576]

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

[0577]

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

[0578]

[0580] [표 S2B]

[0581] 제1 식달 응모막 응모 표본 및 모체 혈액 세포에서 확인된 상위 100개의 저메틸화 영역의 목록

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

[0582]

chr1	89131200	89131700	500	0.9314	0.0498	5.15E-70	0.8816
chr8	94832000	94832500	500	0.9156	0.0351	2.16E-65	0.8805
chr7	14008300	14009800	1500	0.9349	0.0545	0.00E+00	0.8804
chr12	12971300	12972300	1000	0.9361	0.0559	0.00E+00	0.8802
chr5	43114700	43115200	500	0.9638	0.0842	1.79E-13	0.8796
chr11	107872400	107872900	500	0.9472	0.0677	2.31E-32	0.8794
chr8	49757600	49758100	500	0.9048	0.0269	3.15E-52	0.8779
chr13	33106400	33106900	500	0.9384	0.0606	9.54E-15	0.8778
chr3	190658800	190659300	500	0.9388	0.0617	2.71E-22	0.877
chr1	181508000	181508500	500	0.9259	0.0495	3.78E-15	0.8764
chr1	180436900	180437400	500	0.9412	0.0652	2.36E-13	0.876
chr6	122642800	122643800	1000	0.9218	0.0458	0.00E+00	0.8759
chr5	166429300	166429800	500	0.9551	0.08	5.26E-05	0.8751
chr12	14972900	14973400	500	0.9483	0.0733	2.10E-18	0.8749
chr5	123933900	123934400	500	0.943	0.0683	1.12E-39	0.8746
chr2	15969400	15970400	1000	0.8939	0.0196	9.43E-46	0.8743
chr3	167635200	167636200	1000	0.9363	0.0625	9.03E-41	0.8738
chr5	159442700	159443200	500	0.9174	0.044	6.27E-14	0.8734
chr4	48027200	48027700	500	0.9839	0.1111	8.89E-06	0.8728
chr6	140071500	140072000	500	0.9234	0.0506	4.39E-33	0.8728
chr10	22356300	22356800	500	0.9548	0.0822	1.04E-18	0.8726
chr8	61007300	61007800	500	0.9197	0.0476	1.24E-15	0.8721
chr11	95463500	95464000	500	0.9348	0.0629	1.16E-20	0.8718
chr2	216399800	216400300	500	0.938	0.0667	5.98E-06	0.8713
chr18	57359700	57360200	500	0.9293	0.0584	1.77E-19	0.871
chr3	102734400	102734900	500	0.8917	0.0207	7.94E-22	0.871
chr1	173605700	173606200	500	0.96	0.0891	5.46E-13	0.8709
chr2	86993700	86995700	2000	0.8965	0.0261	0.00E+00	0.8704
chr3	162621100	162621600	500	0.9226	0.0526	7.89E-38	0.8699
chr12	10144800	10145300	500	0.929	0.0598	3.45E-17	0.8691
chr3	113855100	113855600	500	0.9667	0.0982	3.97E-14	0.8685
chr2	156958200	156959200	1000	0.9252	0.0571	8.89E-50	0.8681
chr2	55775000	55776000	1000	0.9159	0.0483	0.00E+00	0.8676
chr6	124898400	124898900	500	0.8987	0.0313	1.91E-15	0.8675
chr5	42003700	42004700	1000	0.9262	0.0588	0.00E+00	0.8674
chr3	24162200	24162700	500	0.883	0.0161	1.75E-27	0.8668
chr6	35394000	35395000	1000	0.9204	0.0539	0.00E+00	0.8665
chr17	8451800	8453300	1500	0.9376	0.0714	0.00E+00	0.8662
chr14	53487700	53488700	1000	0.9013	0.0353	0.00E+00	0.866
chr7	98572800	98573300	500	0.9651	0.0995	8.37E-26	0.8656
chr6	52298700	52299200	500	0.9427	0.0772	1.31E-28	0.8655
chr6	159047900	159048400	500	0.908	0.0426	3.34E-08	0.8655
chr14	22152600	22153100	500	0.9085	0.0435	4.43E-17	0.865
chr12	103285000	103285500	500	0.9321	0.0674	0.00E+00	0.8647
chr7	43302200	43302700	500	0.968	0.1037	6.40E-16	0.8643
chr14	22247400	22247900	500	0.9804	0.1163	3.83E-06	0.8641
chr2	66780900	66781400	500	0.9355	0.0714	8.37E-09	0.8641
chr12	97393000	97393500	500	0.9045	0.0408	3.46E-21	0.8637
chr5	162797900	162798900	1000	0.9271	0.0635	1.75E-57	0.8636
chr2	83598400	83599400	1000	0.9354	0.0719	0.00E+00	0.8635
chr11	111358800	111359300	500	0.9156	0.0523	4.15E-24	0.8632

[0583]

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

[0584]

[0586] [표 S2C]

[0587] 제3 석달 태반 조직 및 모체 혈액 세포에서 확인된 상위 100개의 고메틸화 영역의 목록

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

[0588]

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

[0589]

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

[0590]

[0592] [표 S2D]

[0593] 제3 석달 태반 조직 및 모체 혈액 세포에서 확인된 상위 100개의 저메틸화 영역의 목록

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

[0594]

chr18	12720200	12721200	1000	0.9346	0.0618	0.00E+00	0.8728
chr15	60975900	60976900	1000	0.9311	0.0587	0.00E+00	0.8724
chr21	39630100	39631100	1000	0.9423	0.07	4.68E-43	0.8723
chr5	123933900	123934400	500	0.943	0.0707	2.60E-38	0.8722
chr8	77382600	77383600	1000	0.9117	0.0395	0.00E+00	0.8722
chr21	32238800	32239300	500	0.9396	0.0677	1.28E-18	0.8719
chr5	175019600	175020100	500	0.9542	0.0828	4.34E-20	0.8714
chr8	134437400	134438400	1000	0.9083	0.037	4.79E-29	0.8713
chr5	69668800	69669300	500	0.9194	0.0492	2.88E-09	0.8702
chr1	60877900	60878900	1000	0.9378	0.068	0.00E+00	0.8698
chr16	80650400	80650900	500	0.9309	0.0611	2.49E-32	0.8698
chr18	59388800	59389300	500	0.9706	0.1008	4.84E-15	0.8697
chr2	15969400	15970400	1000	0.8939	0.0244	3.07E-52	0.8695
chr13	56207100	56208100	1000	0.9184	0.0505	0.00E+00	0.868
chr3	180395300	180395800	500	0.9877	0.12	1.73E-09	0.8677
chr6	153238200	153239200	1000	0.9123	0.0452	0.00E+00	0.8671
chr18	61635100	61635600	500	0.9268	0.06	2.77E-13	0.8668
chr3	177562200	177563200	1000	0.9121	0.0455	0.00E+00	0.8666
chr4	160368300	160370200	1900	0.9272	0.0606	0.00E+00	0.8665
chr6	144626900	144627400	500	0.9114	0.046	2.10E-12	0.8654
chr16	59885500	59886500	1000	0.9407	0.0757	1.12E-62	0.865
chr1	55667100	55667600	500	0.9095	0.0446	8.62E-39	0.8649
chr2	83598300	83599300	1000	0.9366	0.0718	0.00E+00	0.8648
chr4	105135200	105136200	1000	0.913	0.0486	0.00E+00	0.8644
chr14	32048400	32048900	500	0.9142	0.0499	5.43E-53	0.8643
chr1	223482700	223483700	1000	0.9636	0.0997	2.69E-34	0.864
chr14	47487700	47488200	500	0.915	0.0514	5.45E-33	0.8636
chr3	104515000	104515500	500	1	0.1373	1.08E-06	0.8627
chr7	14008300	14009800	1500	0.9349	0.0725	0.00E+00	0.8624
chr1	243134000	243135500	1500	0.9208	0.0588	0.00E+00	0.8619
chr10	14156400	14156900	500	0.9105	0.0489	0.00E+00	0.8616
chr2	118616200	118617200	1000	0.9178	0.0565	0.00E+00	0.8613
chr17	84555500	8456000	500	0.8941	0.0331	1.94E-18	0.8611
chr12	15392200	15393200	1000	0.9307	0.0697	0.00E+00	0.861
chr8	81275900	81276900	1000	0.9291	0.0684	0.00E+00	0.8606
chr1	234269300	234269800	500	0.9471	0.087	2.25E-25	0.8602
chr1	181970300	181970800	500	0.9167	0.0566	2.15E-08	0.8601
chr2	55775000	55776000	1000	0.9159	0.0559	0.00E+00	0.8599
chr3	88338000	88339000	1000	0.8909	0.0311	0.00E+00	0.8598
chr5	140078700	140079200	500	0.8852	0.0253	0.00E+00	0.8598
chr21	16720900	16721400	500	0.9317	0.0721	1.38E-15	0.8596
chr11	104891100	104892600	1500	0.9164	0.0569	0.00E+00	0.8595
chr1	184204700	184205200	500	0.9194	0.0603	8.16E-16	0.859
chr6	160732500	160733000	500	0.9191	0.0606	1.31E-10	0.8585
chr8	37134300	37134800	500	0.9151	0.0567	1.21E-26	0.8584
chr18	5869800	5870300	500	0.913	0.0548	1.21E-09	0.8582
chr1	98448100	98448600	500	0.9574	0.1	2.08E-05	0.8574
chr3	152897800	152898300	500	0.9402	0.0828	3.28E-18	0.8574
chr1	110304000	110304500	500	0.9783	0.121	2.60E-18	0.8572
chr2	86993600	86995600	2000	0.8965	0.0395	0.00E+00	0.857
chr19	15428100	15430600	2500	0.9424	0.0862	0.00E+00	0.8563

[0595]

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

[0596]

[0597] [표 S3A]

[0598] 제1 석달 모체 혈장 바이셀파이트-서열분석 데이터로부터 고메틸화된 것으로 추정된 상위 100개 유전자좌의 목록

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

[0599]

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

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

[표 S3B]

제1 석달 모체 혈장 바이설파이트-서열분석 데이터로부터 저메틸화된 것으로 추정된 상위 100개 유전자좌의 목록

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

[0605]

chr16	73284579	73285087	0.9729	0.5602	0.4127
chr8	29100691	29101428	0.9728	0.505	0.4678
chr2	202383851	202384447	0.9727	0.5461	0.4267
chr3	179501620	179502300	0.9722	0.5766	0.3956
chr6	107674976	107675906	0.9719	0.4434	0.5285
chr6	107880632	107881161	0.9718	0.5623	0.4095
chr12	56350283	56350933	0.9718	0.5909	0.3809
chr19	40636458	40637339	0.9717	0.4941	0.4776
chr2	223472599	223473287	0.9714	0.1824	0.7891
chr22	20709067	20709787	0.9714	0.5149	0.4565
chr19	46095583	46096190	0.9713	0.5385	0.4328
chr6	90258338	90259318	0.9712	0.3415	0.6297
chr2	54598347	54598933	0.9712	0.5894	0.3819
chr3	114810453	114811493	0.9711	0.5166	0.4545
chr19	15851125	15851654	0.9711	0.5236	0.4476
chr8	42889138	42890084	0.9711	0.5652	0.4059
chr18	52354390	52355064	0.971	0.598	0.373
chr15	38206236	38207010	0.9709	0.4186	0.5523
chr7	99700554	99701110	0.9708	0.305	0.6658
chr12	19487336	19487855	0.9708	0.4105	0.5603
chr7	87996908	87997437	0.9708	0.5462	0.4246
chr6	63628653	63629378	0.9707	0.529	0.4417
chr15	38209108	38209618	0.9706	0.5882	0.3824
chr19	6623769	6624450	0.9704	0.5179	0.4526
chr2	10794513	10795242	0.9704	0.5976	0.3728
chr2	118472785	118474454	0.9704	0.5992	0.3712
chr5	57820209	57820801	0.9701	0.5815	0.3886
chr10	100183380	100184702	0.9701	0.5826	0.3875
chr2	8151989	8152646	0.97	0.4701	0.4999
chr10	3938374	3938914	0.9699	0.1741	0.7958
chr9	123724524	123725439	0.9697	0.57	0.3997
chr14	89085469	89086097	0.9696	0.3278	0.6418
chr16	14129437	14130133	0.9695	0.5304	0.4392
chr5	60746367	60747191	0.9695	0.5571	0.4124
chr1	92002953	92003729	0.9694	0.52	0.4494
chr6	31264677	31265413	0.9693	0.5135	0.4558
chr7	99317013	99318281	0.9692	0.5117	0.4574
chr8	8808867	8809422	0.9692	0.5691	0.4002
chr19	20052165	20052720	0.969	0.2792	0.6898
chr8	129139026	129139573	0.969	0.3458	0.6232
chr11	122314929	122315458	0.969	0.4232	0.5458
chr13	98377663	98378165	0.9688	0.3319	0.6369
chr9	107606194	107606872	0.9688	0.449	0.5198
chr8	56096904	56097736	0.9688	0.5267	0.4422
chr7	128093836	128094339	0.9688	0.5929	0.3758
chr2	103109370	103109916	0.9686	0.3333	0.6352
chr3	101803534	101804063	0.9686	0.5027	0.4659
chr10	69505720	69506278	0.9684	0.2515	0.7169
chr13	26608225	26608754	0.9683	0.3614	0.6069
chr1	90993315	90993828	0.9683	0.5519	0.4164
chr6	11361243	11361801	0.9681	0.2578	0.7103

[0606]

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

[0607]

[0609] [표 S3C]

[0610] 제3 석달 모체 혈장 바이셀파이트-서열분석 데이터로부터 고메틸화된 것으로 추정된 상위 100개 유전자좌의 목록

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

[0611]

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

[0612]

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

[0613]

[0614] [표 S3D]

[0615] 제3 식달 모체 혈장 바이설파이트-서열분석 데이터로부터 저메틸화된 것으로 추정된 상위 100개 유전자좌의 목록

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

[0616]

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

[0617]

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

[0618]

도면

도면1aa

100

이기관 측정치 카운트	54752712	54752712	422873853	422873853	63753170	63753170	456722854	456722854	1289337674	1289337674	1277615263	1277615263
총 배위 가능성(%)	45.08	44.39	44.22	43.83	46.70	46.70	29.82	29.82	35.59	35.59	37.40	37.40
총 배위 가능 측정치	246000492	243023006	180808398	185346026	297735873	297735873	1361959504	1361959504	456222254	456222254	40697003	40697003
모호 비율(%)	6.72	6.03	5.04	5.04	5.25	5.25	64.66	64.66	25.70	25.70	41.01	40.55
중복 비율(%)	6.60	5.86	9.52	9.27	32.94	32.94	64.66	64.66	25.70	25.70	41.01	40.55
모호 및 랑다 전환 비율의 제거 후	11.29	11.19	8.36	8.31	9.43	9.43	25.40	25.40	16.95	16.95	14.08	14.00
개별 C 사열 카운트	99.98	99.98	99.93	99.94	99.95	99.95	99.96	99.96	99.96	99.96	99.96	99.96
개별 C 사열 M 카운트	258458956	265652985	190603545	189460370	226376366	226376366	640597054	640597054	453206420	453206420	397672185	395158659
개별 C 사열 U 카운트	4464594136	4486493557	4223202985	4201359735	4598446023	4598446023	16399110	16399110	8576357118	8576357118	6895711473	6863336640
개별 C 사열 길이	66.45	66.39	62.86	62.86	67.49	67.49	80.48	80.48	69.76	69.76	67.88	67.61
개별 C 사열 길이	8.14	8.07	7.57	7.51	8.89	8.89	21.12	21.12	15.46	15.46	12.48	12.43
개별 C 매칭화 정도(%)	5.65	5.63	4.31	4.29	4.36	4.36	5.19	5.19	5.06	5.06	5.45	5.44
CpG 사열 M 카운트	255203953	262504347	183242316	181245307	213757505	213757505	627966879	627966879	443226176	443226176	38810638	387417592
CpG 사열 U 카운트	104163657	103975011	145623817	149496340	147240201	146617702	310000675	308133594	20637402	205477045	142659625	142271154
CpG 카버(%)	85.17	85.12	81.02	80.92	84.23	84.15	92.05	92.05	86.13	86.06	84.93	84.85
CpG 사열 길이	13.12	13.01	11.68	11.71	12.82	12.71	33.30	33.05	23.06	22.94	18.91	18.81
CpG 매칭화 정도(%)	71.80	71.63	55.72	54.97	59.21	59.04	66.95	66.95	68.20	68.20	73.18	73.14
HiCpG 사열 M 카운트	3279043	3147047	7286715	7162225	12301418	12301418	12818574	12532920	12979169	12979169	7859475	7740641
HiCpG 사열 U 카운트	4302333851	4285094041	3579826287	3971182568	4550807406	4537834784	11358110	11308110	8330532598	8330532598	6749182596	6719112682
HiCpG 카버(%)	0.13	0.13	0.12	0.12	0.13	0.13	0.16	0.16	0.14	0.14	0.13	0.13
HiCpG 사열 길이	1.52	1.52	1.41	1.41	1.62	1.62	4.02	4.02	2.96	2.96	2.39	2.38
HiCpG 매칭화 정도(%)	0.08	0.07	0.18	0.18	0.27	0.27	0.11	0.11	0.15	0.15	0.12	0.12
CHG 사열 M 카운트	1029734	1000287	2100977	2071015	3223342	3176751	3785562	3691190	3625552	3592817	2347392	2314323
CHG 카버(%)	1345439602	1341393096	1271962855	1276609371	1352882785	1357566840	341653203	3404225988	2519511702	2509439947	2035336055	2028059603
CHG 사열 길이	77.90	77.82	73.26	73.13	77.50	77.38	88.16	88.12	79.68	79.59	77.84	77.75
CHG 매칭화 정도(%)	10.89	10.85	9.86	9.95	11.04	11.00	27.65	27.55	20.40	20.32	16.48	16.42
CHH 사열 M 카운트	2249309	2146760	5185738	5091211	9278803	9124657	9052012	8841730	9353607	9251325	5512083	5426318
CHH 카버(%)	796694249	2943700945	2761843432	2754573597	3181726541	3180265544	799564596	7903545496	5848054967	5821092551	4712844541	4691053079
CHH 사열 길이	61.95	61.90	58.43	58.32	63.54	63.45	77.53	77.52	65.96	65.89	63.65	63.58
CHH 매칭화 정도(%)	6.84	6.81	6.40	6.38	7.42	7.38	18.36	18.30	13.55	13.48	10.91	10.86
CHH 카버(%)	0.06	0.07	0.19	0.18	0.29	0.29	0.11	0.11	0.16	0.16	0.12	0.12

도면lab

미가공 측정치 카운트

총 맵핑 가능성(%)

총 맵핑 가능 측정치

모호 비율(%)^a중복 비율(%)^b모호 및 일단 전함 비율의 제거
후 전체 계층 깊이(%)^c계층 C 서열 M 카운트^d

계층 C 서열 U 카운트

계층 C 커버(%)

계층 C 서열 깊이

계층 C 메틸화 정도(%)

CpG 서열 M 카운트

CpG 서열 U 카운트

CpG 커버(%)

CpG 서열 깊이

CpG 메틸화 정도(%)

비CpG 서열 M 카운트

비CpG 서열 U 카운트

비CpG 커버(%)

비CpG 서열 깊이

비CpG 메틸화 정도(%)

CHG 서열 M 카운트

CHG 서열 U 카운트

CHG 커버(%)

CHG 서열 깊이

CHG 메틸화 정도(%)

CHH 서열 M 카운트

CHH 서열 U 카운트

CHH 커버(%)

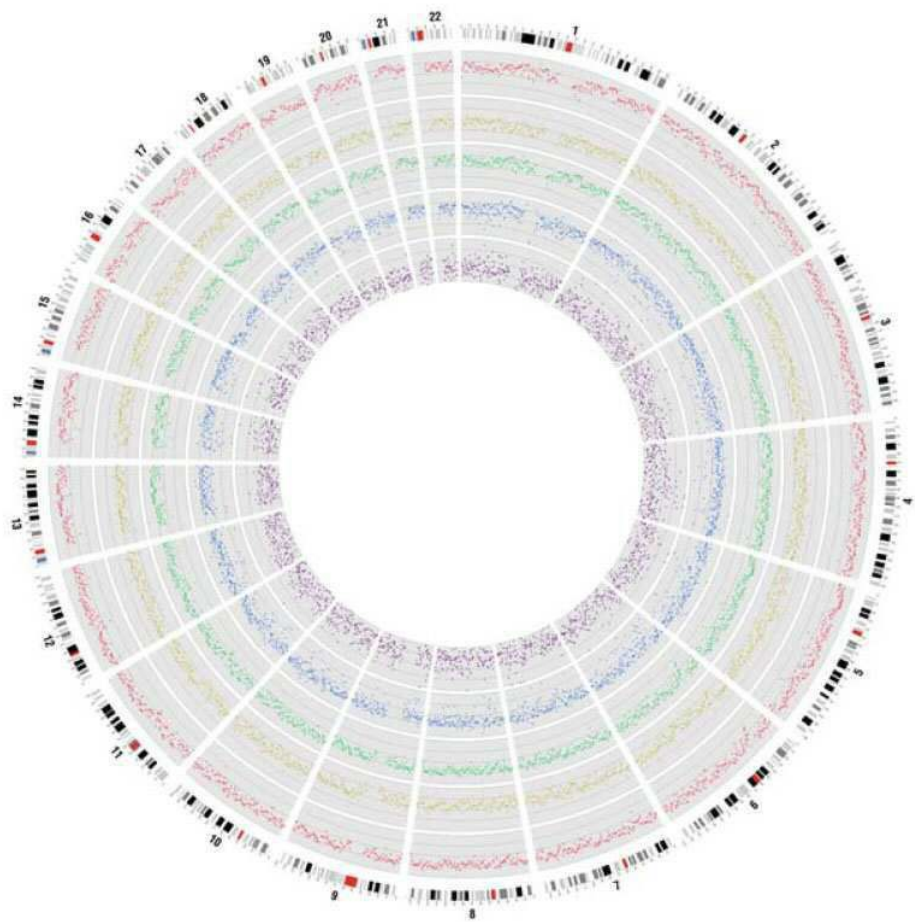
CHH 서열 깊이

CHH 메틸화 정도(%)

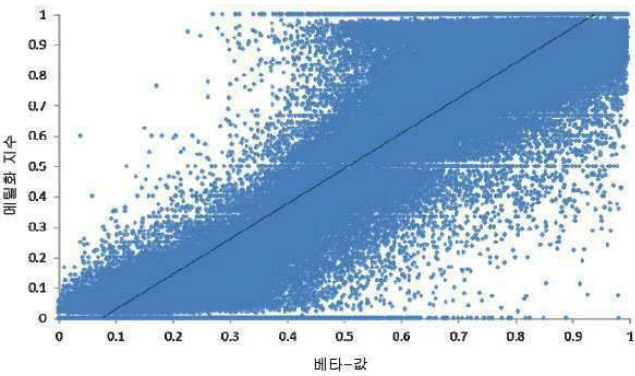
	성인 남성 혈액		성인 여성 혈액		성인 남성 혈장		성인 여성 혈장	
	앞손	크리	앞손	크리	앞손	크리	앞손	크리
196644792	196544792	197387755	197387755	197387755	179688600	179688600	141199439	141199439
43.52	42.97	44.94	44.35	44.35	22.47	22.29	18.98	18.79
85526774	84462496	88699242	87542953	87542953	40380938	40048964	26795699	26537762
6.05	6.13	6.29	6.37	6.37	6.09	6.14	6.29	6.35
1.99	1.73	1.88	1.53	1.53	7.89	7.64	11.71	11.47
4.13	4.09	4.28	4.24	4.24	1.83	1.82	1.16	1.15
99.96	99.97	99.97	99.97	99.97	99.96	99.97	99.96	99.97
127164246	125487793	124983462	123238608	123238608	56049489	54630306	37403328	37048480
2047971831	2026731827	2058153525	2039985644	2039985644	938186429	933377753	598496591	594974253
61.66	61.65	66.17	66.06	66.06	43.76	43.66	38.39	38.28
3.72	3.68	3.74	3.70	3.70	1.70	1.69	1.09	1.08
5.85	5.83	5.72	5.70	5.70	5.54	5.53	5.88	5.86
125161042	123578850	123020928	121374309	121374309	54155037	53756413	36788819	36453865
49564325	49936003	47562383	47812716	47812716	21930864	21830555	13971713	13857179
73.37	73.25	76.24	76.12	76.12	63.93	63.80	57.87	57.77
6.20	6.16	6.06	6.01	6.01	2.70	2.68	1.80	1.79
71.53	71.22	72.12	71.74	71.74	71.18	71.12	72.48	72.44
1979410	1899750	1951160	1866642	1866642	894139	873671	614025	594289
1960312392	1950307595	1976031265	1966232257	1966232257	915375235	910910490	583173346	580050464
0.10	0.10	0.11	0.11	0.11	0.08	0.08	0.07	0.07
0.69	0.69	0.70	0.70	0.70	0.32	0.32	0.21	0.21
0.10	0.10	0.10	0.09	0.09	0.10	0.10	0.11	0.10
619974	598717	594336	573203	573203	279945	272311	191120	186135
612387336	610126996	607312325	605033927	605033927	283916057	282626050	180374803	179612404
63.24	63.11	66.75	66.62	66.62	54.69	54.56	48.57	48.45
4.96	4.94	4.91	4.90	4.90	2.30	2.29	1.46	1.45
0.10	0.10	0.10	0.09	0.09	0.10	0.10	0.11	0.10
1360436	1301033	1356824	1283439	1283439	614194	601360	422905	408154
1347925056	1340180599	1368718940	1361138330	1361138330	631459168	628084440	402798543	400438060
46.93	46.83	50.48	50.38	50.38	39.32	39.22	34.20	34.11
3.12	3.10	3.17	3.15	3.15	1.46	1.45	0.93	0.93
0.10	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.10

도면1b

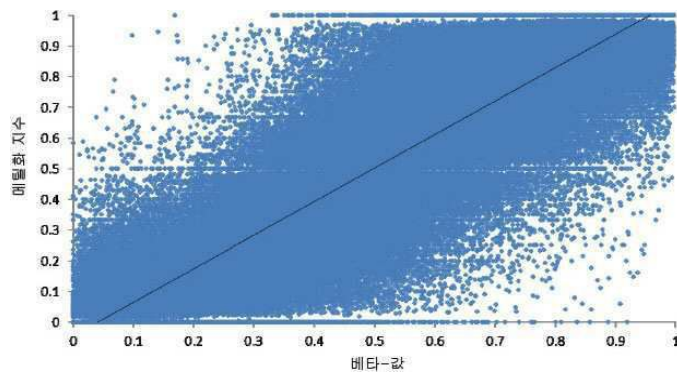
150



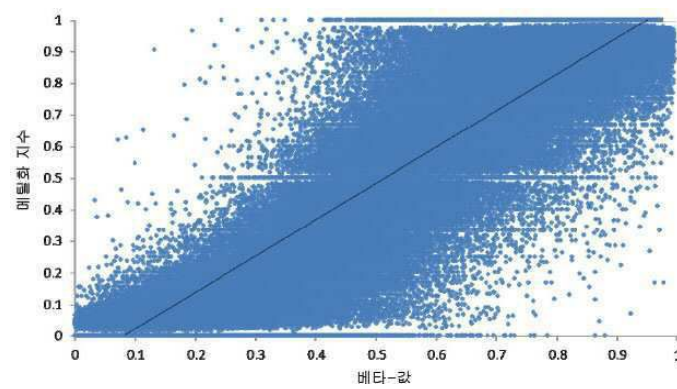
도면2a



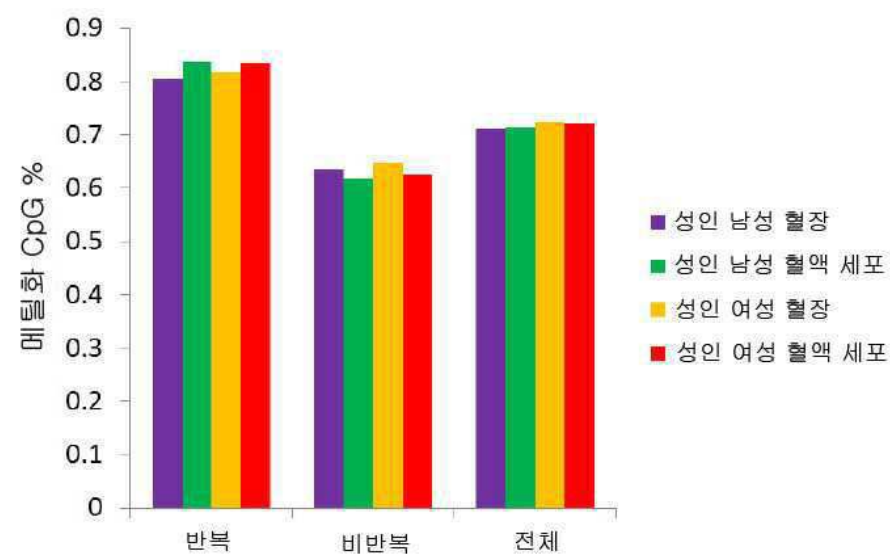
도면2b



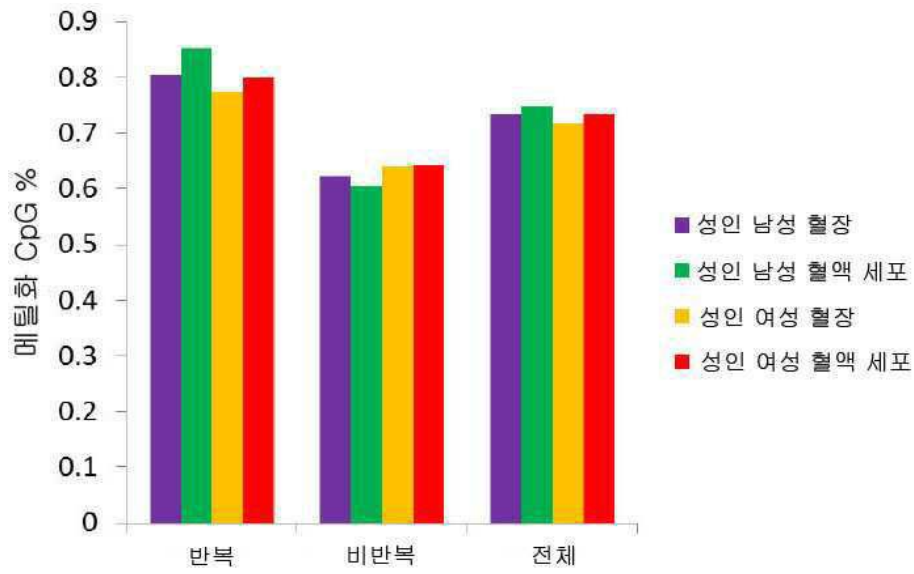
도면2c



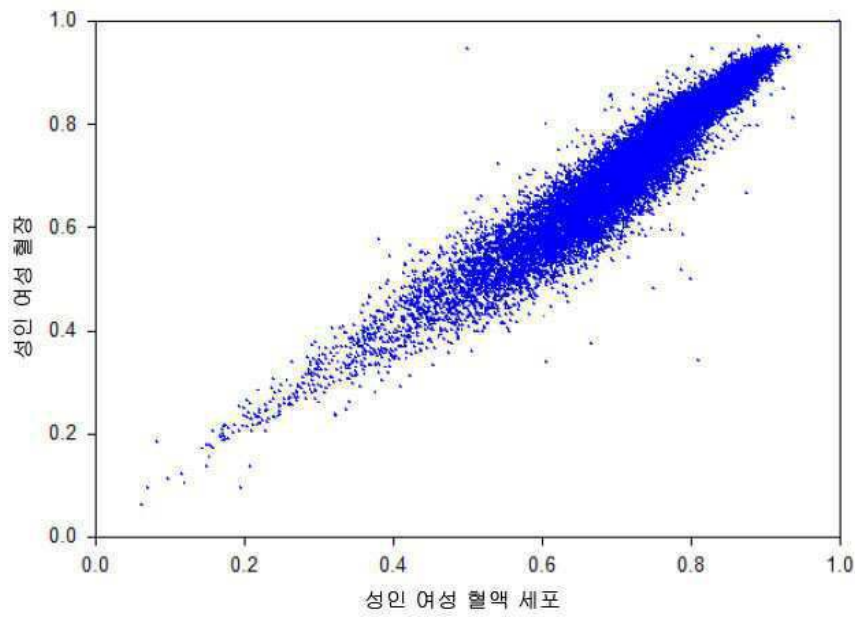
도면3a



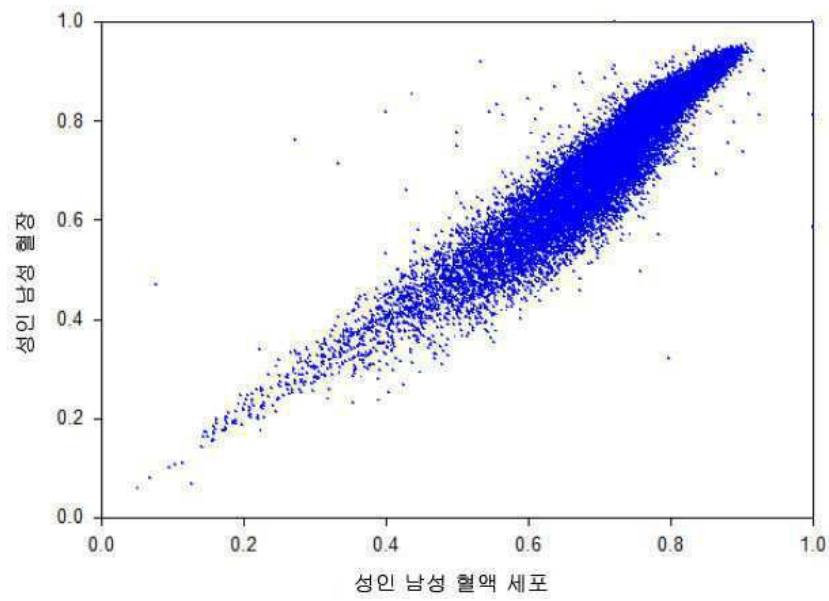
도면3b



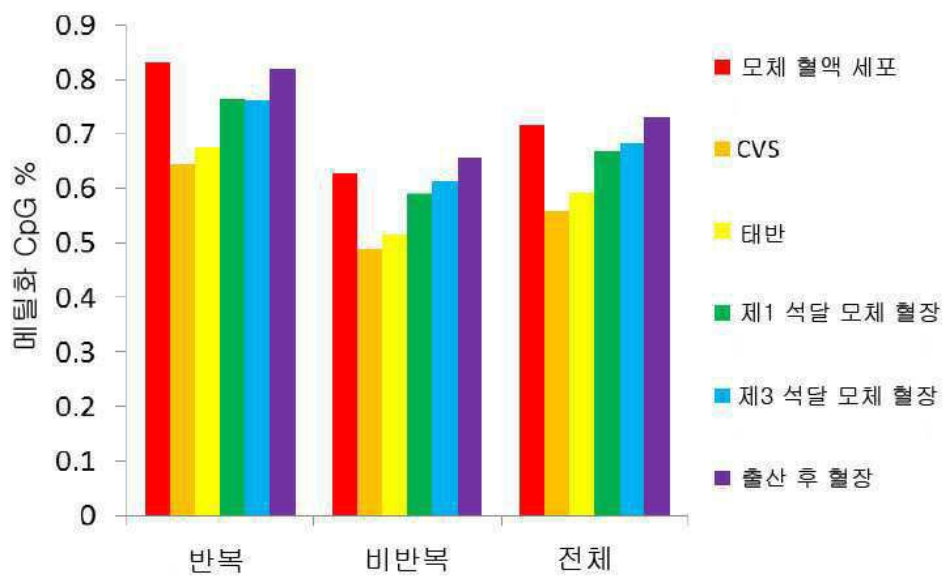
도면4a



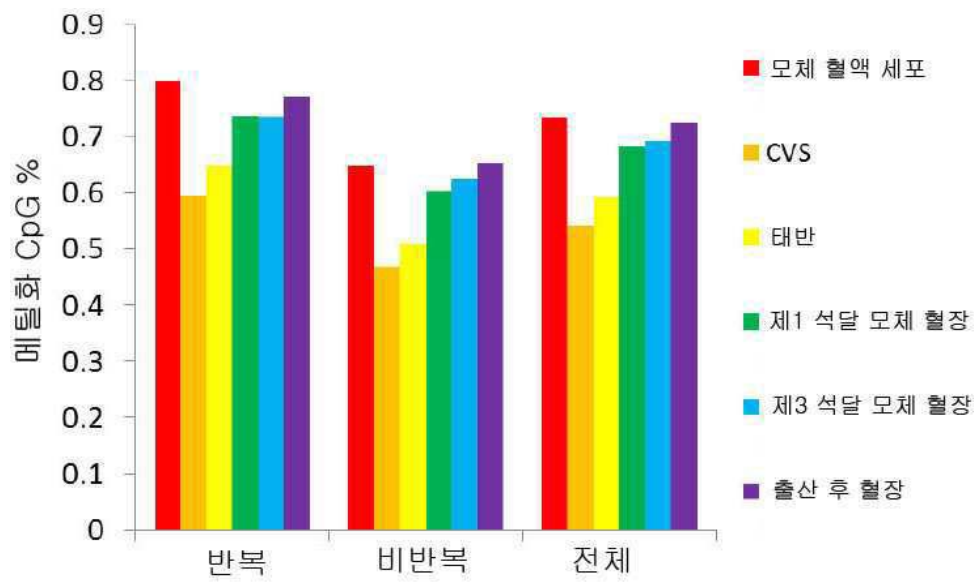
도면4b



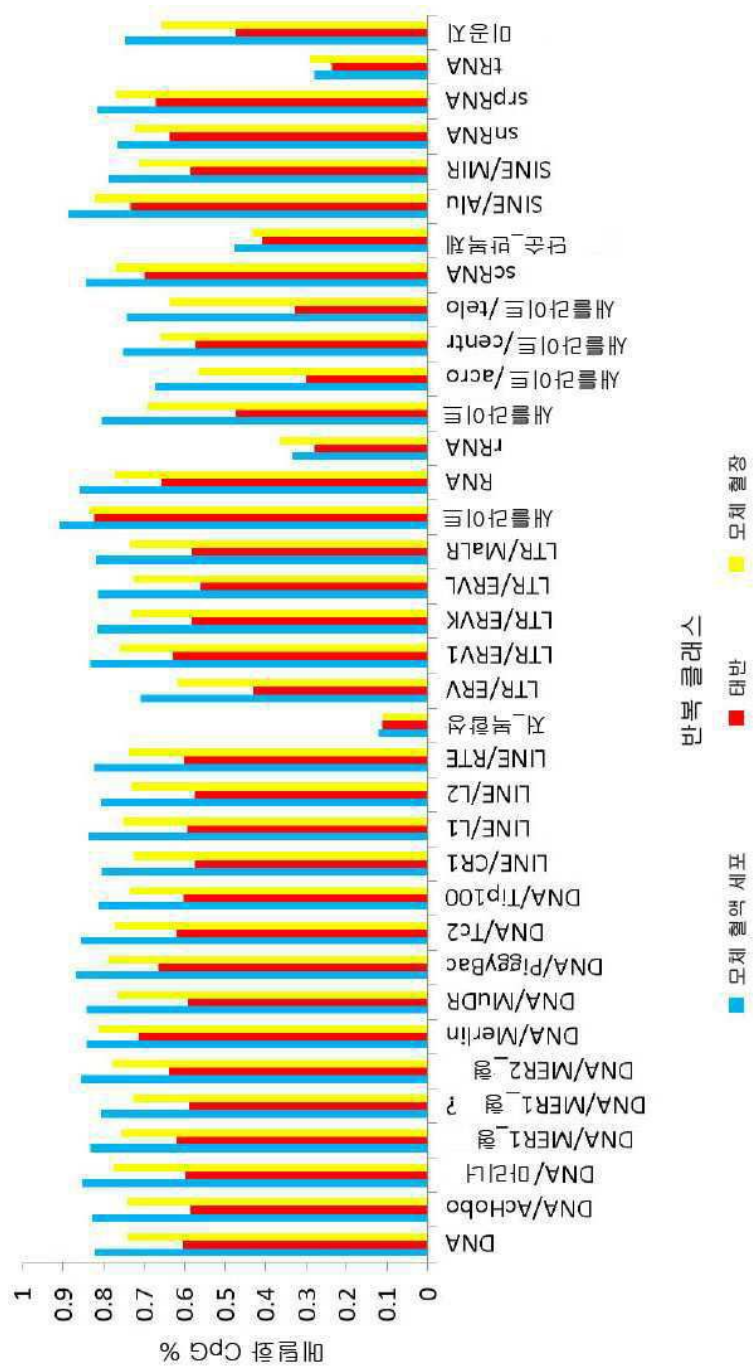
도면5a



도면5b



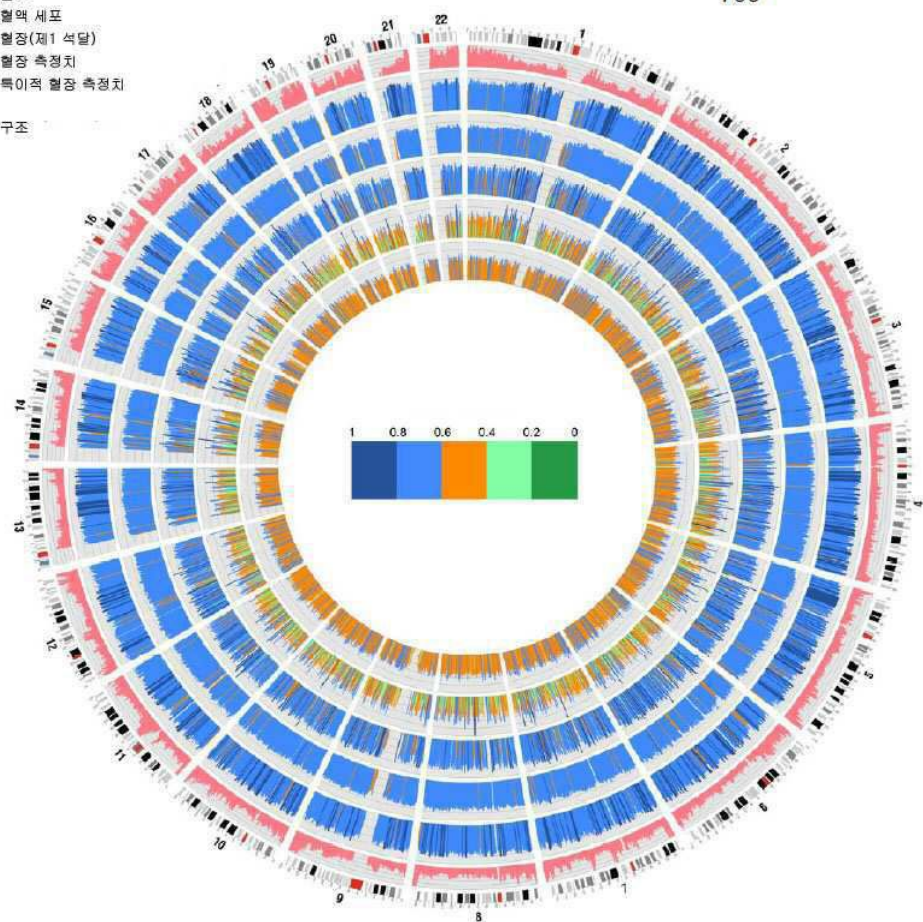
도면6



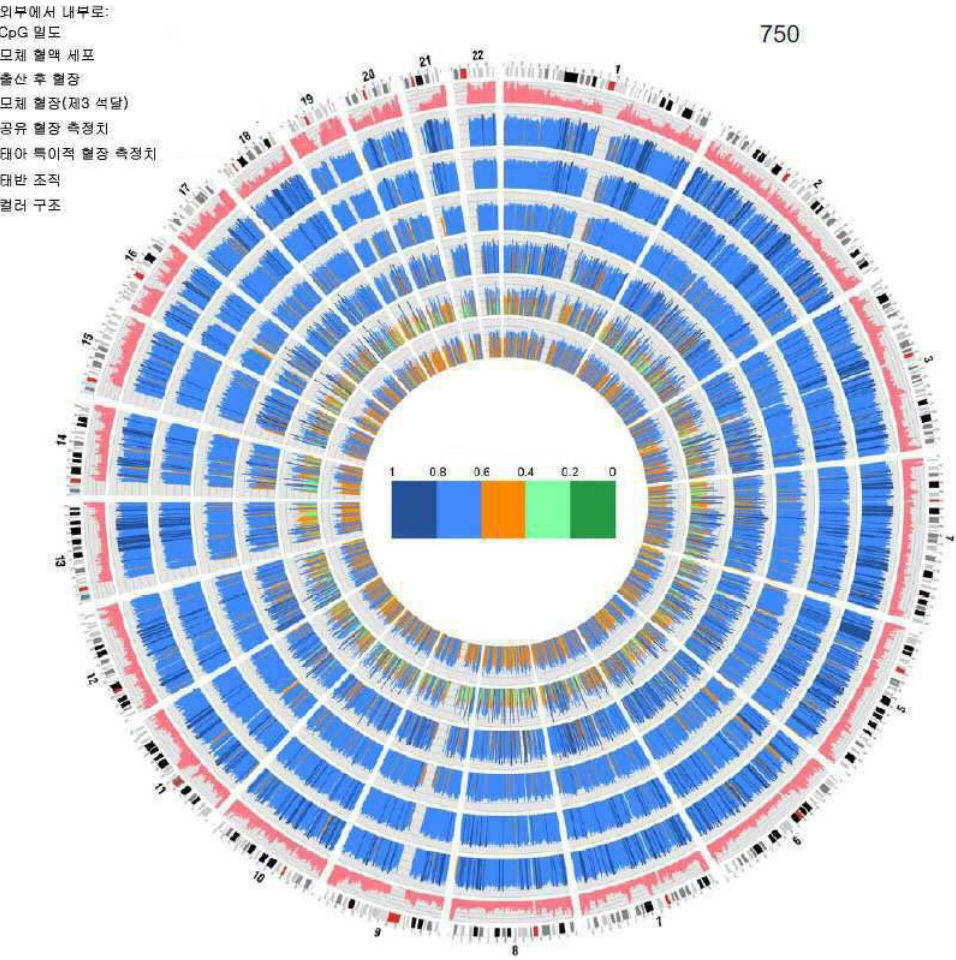
도면7a

외부에서 내부로:
CpG 밀도
모체 혈액 세포
모체 혈장(제1 식달)
공유 혈장 측정치
태아 특이적 혈장 측정치
CVS
컬러 구조

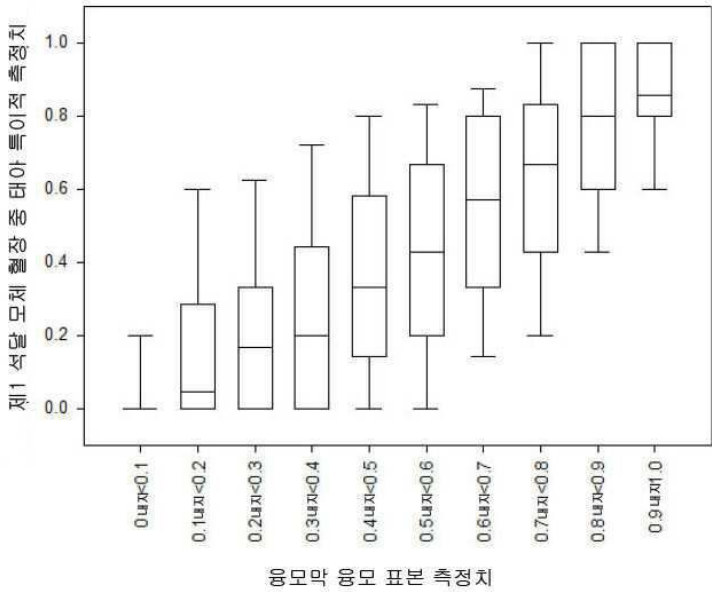
700



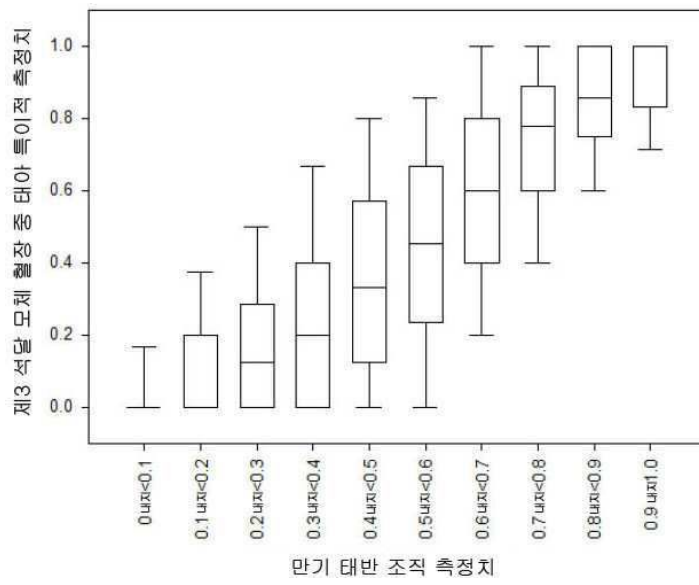
도면7b



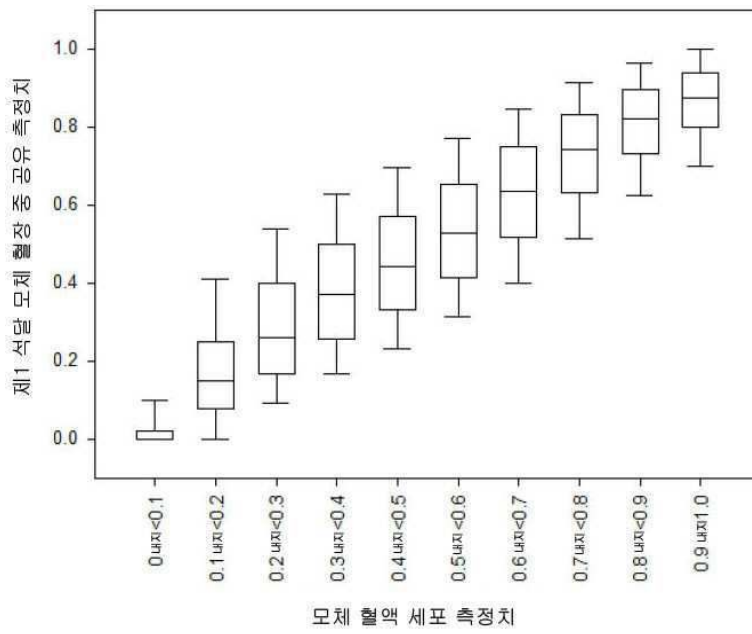
도면8a



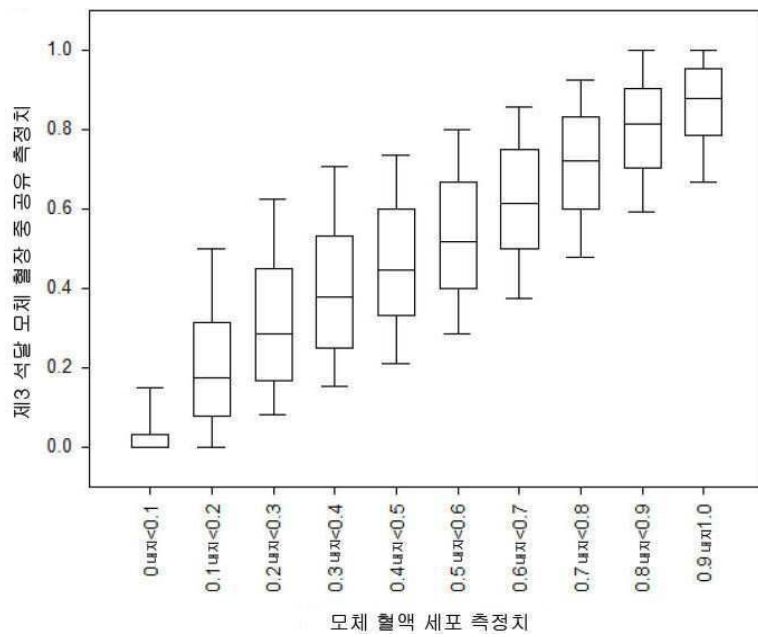
도면8b



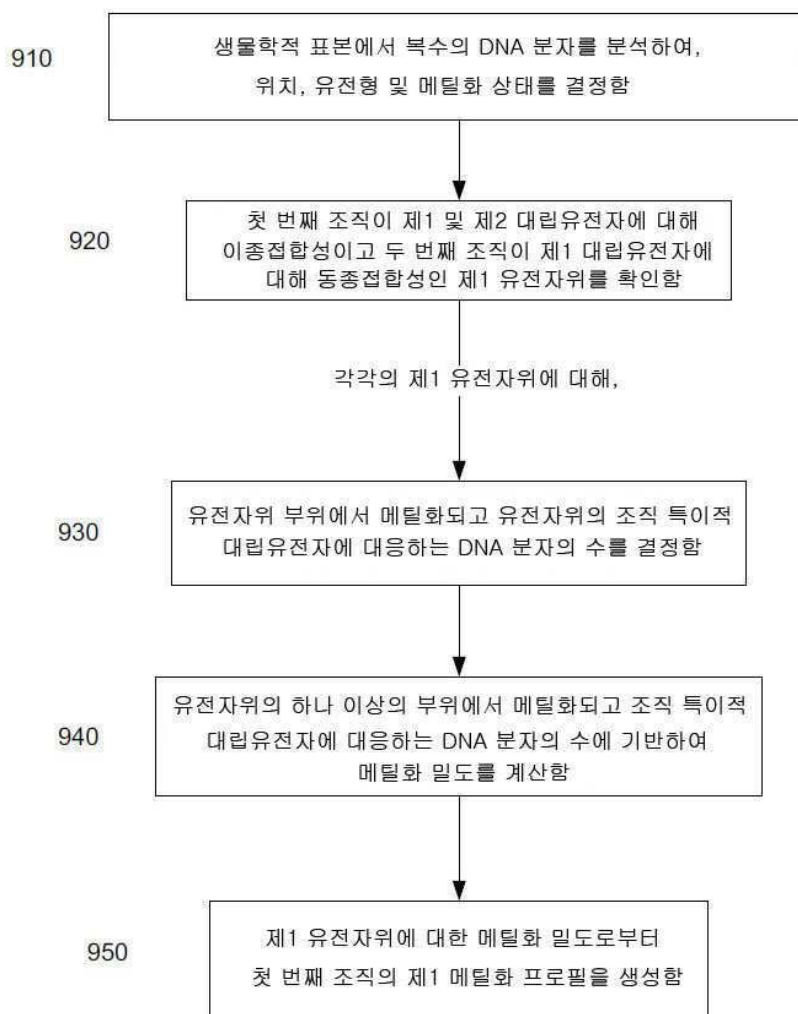
도면8c



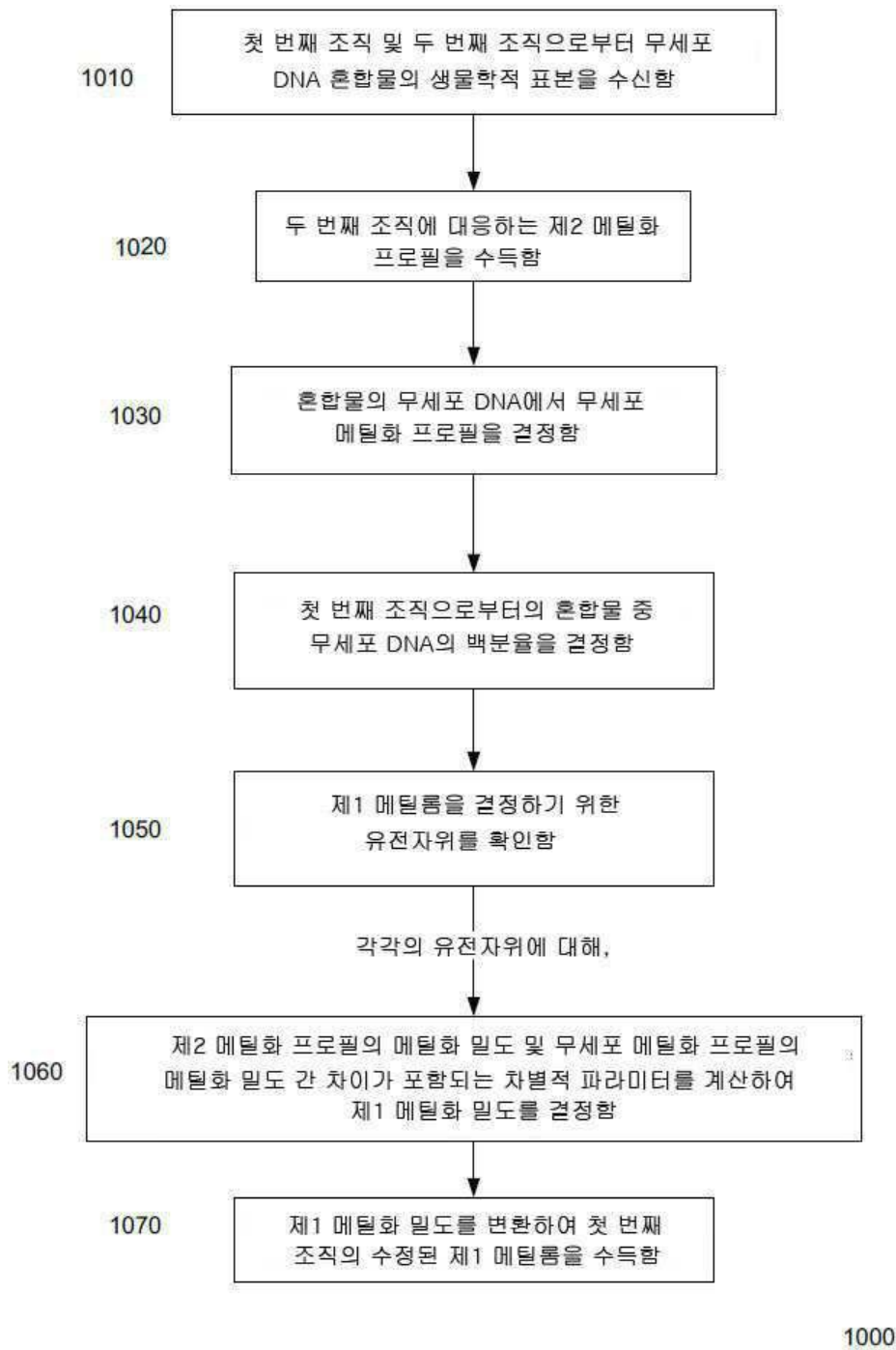
도면8d



도면9



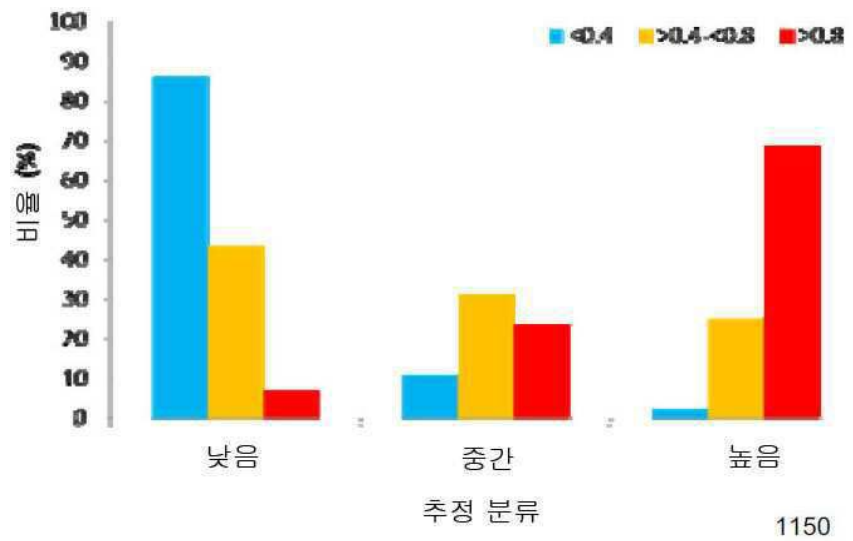
도면10



도면11a



도면11b

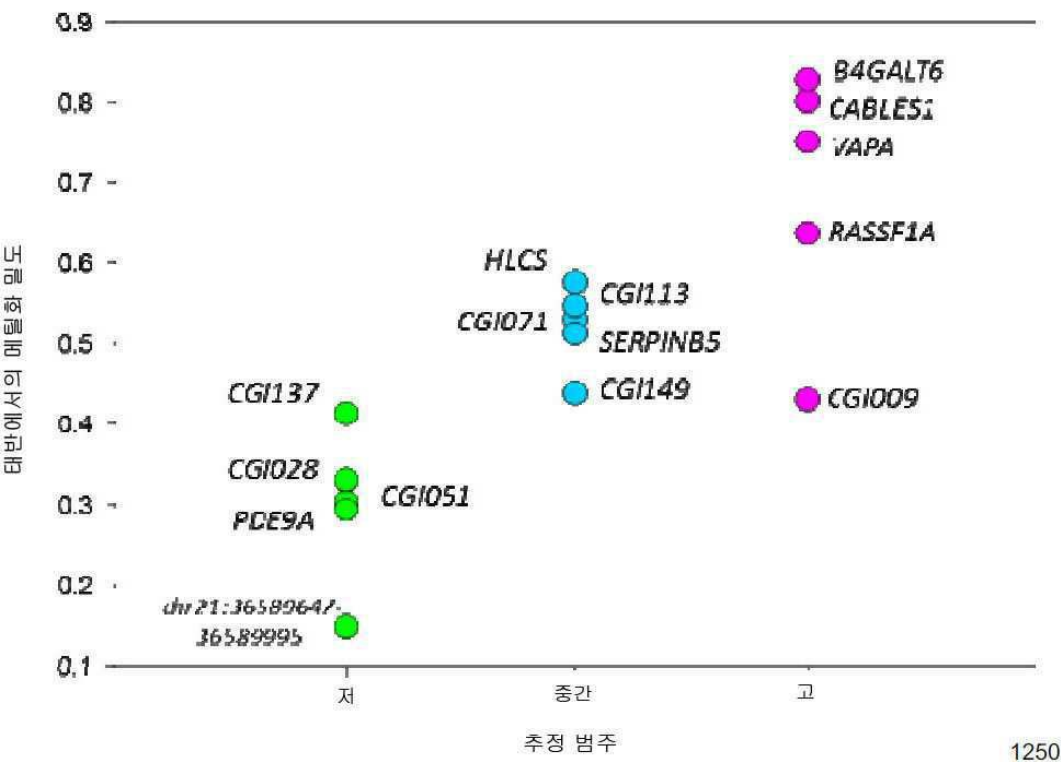


도면12a

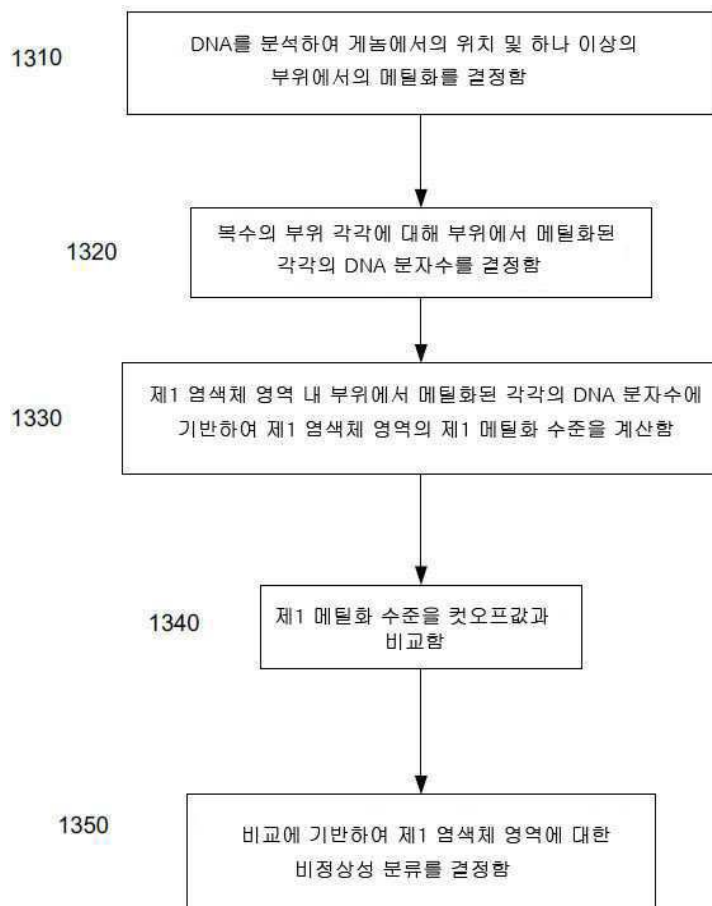
메틸화 및도(서열분석 데이터)										추정		문헌에서의 메틸화 상태	
번호	유전자위	게놈 위치	모체 혈액 (BKG)	태반	모체 혈장 (MP)	추정 값 (D) ¹	메틸화 예측 스코어 (MPS) ²	추정 변주 ³	태반 메틸화 수준 ⁴	태반 상태 ⁵	참고 문헌		
1	RASSF1A	chr3:50352938-50353234	0.000	0.637	0.067	0.506	1.060	고	중간	고메틸화	(1)		
2	CGI009	chr21:25856489-25856756	0.001	0.432	0.050	0.374	0.849	고	중간	고메틸화	(2)		
3	CGI028	chr21:33321942-33322182	0.035	0.331	0.040	0.073	0.367	저	저	고메틸화	(2)		
4	CGI051	chr21:36995232-36995509	0.043	0.303	0.046	0.072	0.365	저	저	고메틸화	(2)		
5	CGI071	chr21:41140466-41140898	0.031	0.531	0.046	0.146	0.483	중간	중간	고메틸화	(2)		
6	CGI113	chr21:44953820-44953972	0.013	0.547	0.051	0.306	0.739	중간	중간	고메틸화	(2)		
7	CGI137	chr21:46249872-46250196	0.940	0.413	0.820	0.030	0.299	저	중간	저메틸화	(2)		
8	CGI149	chr21:46911575-46912533	0.004	0.440	0.030	0.202	0.573	중간	중간	고메틸화	(2)		
9	HLCS	chr21:37274945-37275031	0.038	0.576	0.059	0.203	0.575	중간	중간	고메틸화	(2)		
10	PDE9A	chr21:42979270-42979525	0.950	0.303	0.823	-0.018	0.222	저	저	저메틸화	(2)		
11	-	chr21:36589647-36589995	0.952	0.149	0.827	0.007	0.262	저	저	저메틸화	미공개		
12	SERPINB5	chr18:59294739-59295038	0.936	0.530	0.856	0.329	0.777	중간	중간	저메틸화	(3)		
13	VAPA	chr18:10022563-10023186	0.036	0.752	0.082	0.382	0.861	고	중간	고메틸화	(4)		
14	CABLES1	chr18:19035954-19036069	0.335	0.801	0.491	1.519	2.680	고	고	고메틸화	(4)		
15	B4GALT6	chr18:27486040-27486932	0.113	0.827	0.331	1.764	3.072	고	고	고메틸화	(4)		

1200

도면12b

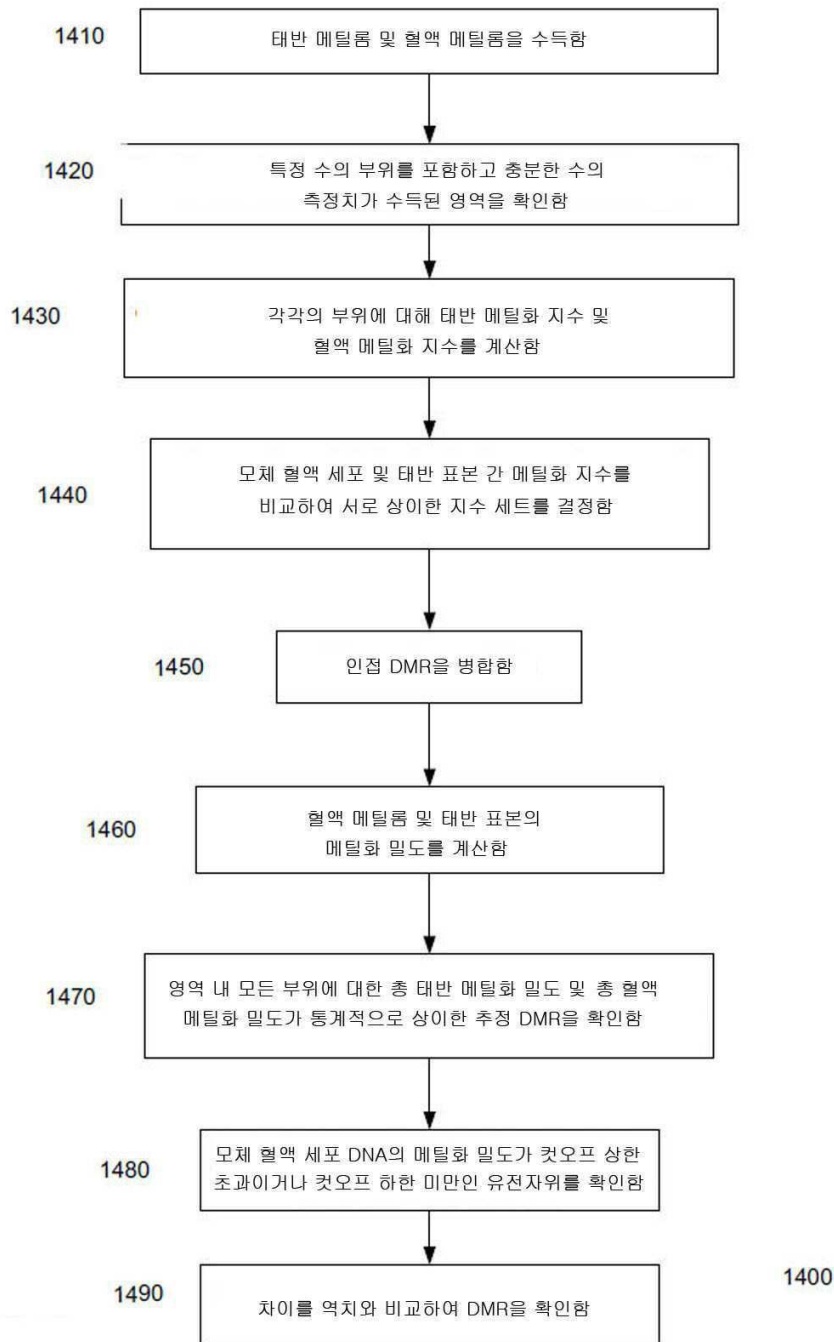


도면13



1300

도면14



도면15a

					BS-Seq ^a			DMR ^d							
번호 ^a	유전자위	개능 위치 ^b	크기 ^c	대만 상태 ^b	모체 혈액 세포	CVS	Meth diff	Chr	시작	끝	크기	모체 혈액 세포	CVS	p 값	Meth 클래스
1	RASSF1A	chr3:50352938-50353234	297	고메틸화	0.00	63.73	63.73	chr3	50351000	50354000	3000	6.26	66.68	0.00E+00	60.42 TP
2	CGI009	chr21:25855489-25856758	268	고메틸화	0.08	43.20	43.13	chr21	25855500	25857500	2000	4.97	48.70	0.00E+00	43.73 TP
3	CGI028	chr21:33321942-33322182	241	고메틸화	3.51	33.10	29.59	chr21	33316400	33323900	7500	4.52	23.85	0.00E+00	19.33 FN
4	CGI051	chr21:36995232-36995509	278	고메틸화	4.26	30.33	26.07	chr21	36992800	37003900	11100	3.12	21.91	0.00E+00	18.79 FN
5	CGI071	chr21:41140466-41140898	433	고메틸화	3.09	53.08	49.97	chr21	41140000	41141000	1000	3.40	52.00	0.00E+00	48.80 TP
6	CGI113	chr21:44953820-44953972	153	고메틸화	1.27	54.69	53.42	chr21	44948500	44954700	6200	7.18	57.23	0.00E+00	50.05 TP
7	CGI137	chr21:46249872-46250196	325	저메틸화	94.00	41.34	52.65	chr21	46227800	46250400	22600	89.74	52.67	0.00E+00	37.07 TP
8	CGI149	chr21:46911575-46912533	959	고메틸화	0.40	43.95	43.56	chr21	46911200	46912700	1500	5.31	46.17	0.00E+00	40.86 TP
9	HLC5	chr21:37274945-37275031	87	고메틸화	3.77	57.61	53.85	chr21	37274200	37275200	1000	7.33	54.24	0.00E+00	46.91 TP
10	PDE9A	chr21:42979270-42979525	256	저메틸화	95.01	30.33	64.68	chr21	42978100	43004800	26700	89.97	51.44	0.00E+00	38.53 TP
11	-	chr21:36589647-36589995	349	저메틸화	95.17	14.89	80.28	chr21	36589200	36590200	1000	94.49	19.27	0.00E+00	75.22 TP
12	SERPINB5	chr18:59294739-59295038	300	저메틸화	93.64	53.00	40.64	chr18	59294800	59296300	1500	88.47	40.76	3.33E-27	47.71 TP
13	VAPA	chr18:10022563-10023187	625	고메틸화	3.63	75.17	71.54	chr18	10021800	10023300	1500	24.08	80.33	0.00E+00	56.25 TP
14	CABLES1	chr18:19035854-19036070	117	고메틸화	33.47	80.09	46.62	-	-	-	-	-	-	-	FN
15	B4GALT6	chr18:27486040-27486933	894	고메틸화	11.34	82.75	71.41	chr18	27486100	27487100	1000	11.34	82.75	9.6408E-88	71.41 TP
16	CASP8	chr2:201830342-201831885	1544	고메틸화	0.52	47.54	47.03	chr2	201830200	201832200	2000	3.77	49.13	4.22E-42	45.36 TP
17	RARB	chr3:25444840-25445090	251	고메틸화	2.28	26.93	24.65	chr3	25443900	25445400	1500	3.44	22.97	9.65E-50	19.53 FN
18	APC	chr5:112100999-112101943	945	고메틸화	0.71	45.78	45.06	chr5	112100300	112101800	1500	3.59	48.09	0.00E+00	44.50 TP
19	DAB2IP_ori	chr9:121541221-121541480	240	저메틸화	84.87	52.63	32.24	chr9	121540400	121541500	1100	85.21	36.86	5.2605E-10	48.35 TP
20	DAB2IP_ex	chr9:123128261-123129077	817	저메틸화	95.09	32.99	62.10	chr9	123127700	123129200	1500	90.72	50.61	1.4706E-19	40.11 TP
21	THY1	chr11:118798952-118799382	431	고메틸화	0.88	39.32	38.44	chr11	118797600	118799600	2000	2.25	27.28	0.00E+00	25.03 TP
22	PTPN6	chr12:6930211-6933772	3562	고메틸화	1.52	79.80	78.29	chr12	6930000	6934000	4000	4.26	80.13	0.00E+00	75.87 TP
23	PYCARD_ex	chr16:31121808-31122146	339	고메틸화	4.90	46.51	41.61	-	-	-	-	-	-	-	FN
24	chr18A	chr18:55090284-55090606	323	고메틸화	6.14	31.79	25.65	chr18	55089700	55094200	4500	4.78	22.91	0.00E+00	18.13 FN
25	chr21A	chr21:39279723-39280004	282	고메틸화	12.80	64.02	51.21	chr21	39279300	39279800	500	4.74	62.07	7.17E-62	57.33 TP
26	chr21B	chr21:44161027-44161371	345	고메틸화	0.86	83.90	83.04	chr21	44160700	44161700	1000	11.06	77.32	0.00E+00	66.26 TP
27	chr21C	chr21:33320544-33320829	286	고메틸화	5.49	37.92	32.43	chr21	33316400	33323900	7500	4.52	23.85	0.00E+00	19.33 FN
28	chr21D	chr21:42189223-42189683	461	고메틸화	13.02	85.84	72.82	chr21	42188700	42189700	1000	7.78	81.67	3.62E-77	73.89 TP
29	chr21E	chr21:42355712-42355908	197	고메틸화	0.00	57.85	57.85	chr21	42355000	42357500	2500	5.21	62.23	0.00E+00	57.02 TP
30	chr21EII	chr21:42357215-42357341	127	고메틸화	3.57	48.25	44.68	chr21	42355000	42357500	2500	5.21	62.23	0.00E+00	57.02 TP
31	chr21H	chr21:32268803-32268943	141	고메틸화	0.49	78.84	78.35	chr21	32268300	32269300	1000	20.49	85.25	4.9563E-34	64.76 TP
32	chr21I	chr21:44079235-44079535	301	고메틸화	1.23	67.44	66.21	chr21	44079400	44080400	1000	15.19	68.06	4.9E-59	52.87 TP
33	SERPINB5_P	chr18:59294811-	465	저메틸화	91.96	39.71	52.25	chr18	59294800	59296300	1500	88.47	40.76	3.3294E-	47.71 TP

1500

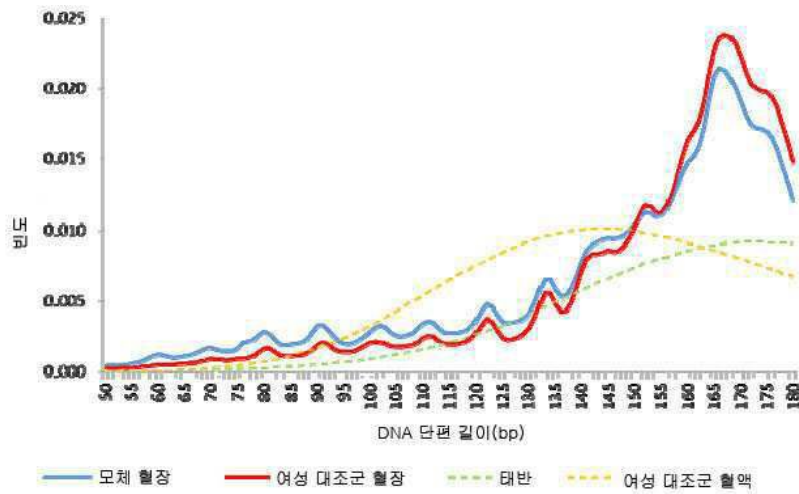
도면15b

번호 ^a	유전자위	BS-Seq			DMR ^c								클래스 ^d
		모체 혈액 세포	태반	Meth diff ^b	Chr	시작	끝	크기	모체 혈액 세포	태반	p 값	Meth diff	
1	RASSF1A	0.00	60.84	60.84	chr3	50352100	50354100	2000	2.16	62.21	0.00E+00	60.05	TP
2	CGI009	0.08	20.50	20.43	chr21	25855600	25857100	1500	2.04	24.37	0.00E+00	22.33	TP
3	CGI028	3.51	17.38	13.86	-	-	-	-	-	-	-	-	TN
4	CGI051	4.26	13.93	9.67	-	-	-	-	-	-	-	-	TN
5	CGI071	3.09	24.85	21.76	chr21	41139900	41141400	1500	6.83	29.28	0.00E+00	22.45	TP
6	CGI113	1.27	40.24	38.97	chr21	44948500	44954700	6200	7.18	45.00	0.00E+00	37.82	TP
7	CGI137	94.00	37.10	56.90	chr21	46248400	46250400	2000	92.28	43.85	1.66E-76	48.43	TP
8	CGI149	0.40	22.19	21.79	chr21	46911200	46912700	1500	5.31	25.73	0.00E+00	20.42	TP
9	HLCS	3.77	58.49	54.72	chr21	37274200	37275200	1000	7.33	56.83	0.00E+00	49.50	TP
10	PDE9A	95.01	32.41	62.60	chr21	42978900	42990100	11200	90.89	44.38	0.00E+00	46.51	TP
11	-	95.17	19.92	75.24	chr21	36589200	36590200	1000	94.49	22.66	0.00E+00	71.83	TP
12	SERPINB5	93.64	28.06	65.57	chr18	59294500	59295500	1000	92.11	36.53	3.85E-31	55.58	TP
13	VAPA	3.63	73.88	70.25	chr18	10022100	10023100	1000	11.67	75.76	0.00E+00	64.09	TP
14	CABLES1	33.47	81.40	47.93	-	-	-	-	-	-	-	-	FN
15	B4GALT6	11.34	88.57	77.23	chr18	27485800	27486800	1000	28.97	90.00	4.3779E-59	61.03	TP
16	CASP8	0.52	56.27	55.75	chr2	201830200	201832200	2000	3.77	57.60	1.33E-52	53.83	TP
17	RARB	2.28	38.18	35.90	chr3	25444000	25445000	1000	2.36	24.42	8.00E-54	22.06	TP
18	APC	0.71	34.89	34.18	chr5	112100700	112101700	1000	0.38	32.97	0.00E+00	32.59	TP
19	DAB2IP_ori	84.87	40.28	44.60	chr9	121540400	121541400	1000	84.79	30.63	1.7338E-14	54.16	TP
20	DAB2IP_ex	95.09	16.14	78.95	chr9	123128600	123129100	500	95.14	17.42	9.6247E-47	77.72	TP
21	THY1	0.88	22.28	21.40	chr11	118797800	118799300	1500	1.22	20.50	0.00E+00	19.28	FN
22	PTPN6	1.52	81.33	79.82	chr12	6930000	6934000	4000	4.26	81.58	0.00E+00	77.32	TP
23	PYCARD_ex	4.90	48.73	43.83	chr16	31121700	31122200	500	4.85	32.87	1.1158E-34	28.02	TP
24	chr18A	6.14	16.25	10.10	-	-	-	-	-	-	-	-	TN
25	chr21A	12.80	74.65	61.85	chr21	39278300	39280300	2000	9.96	75.07	0.00E+00	65.11	TP
26	chr21B	0.86	81.56	80.70	chr21	44160700	44161700	1000	11.06	78.6	0.00E+00	67.54	TP
27	chr21C	5.49	14.91	9.42	-	-	-	-	-	-	-	-	TN
28	chr21D	13.02	88.71	75.70	chr21	42188700	42189700	1000	7.78	84.42	3.5806E-77	76.64	TP
29	chr21EI	0.00	73.21	73.21	chr21	42355300	42357800	2500	4.66	79.19	0.00E+00	74.53	TP
30	chr21EII	3.57	84.00	80.43	chr21	42355300	42357800	2500	4.66	79.19	0.00E+00	74.53	TP
31	chr21H	0.49	77.49	77.00	chr21	32268400	32268900	500	18.4	83.85	2.1972E-30	65.45	TP
32	chr21I	1.23	58.62	57.39	chr21	44079300	44080300	1000	12.16	72.55	2.3559E-60	60.39	TP
33	SERPINB5_P	91.96	28.61	63.35	chr18	59294500	59295500	1000	92.11	36.53	3.85E-31	55.58	TP

도면16

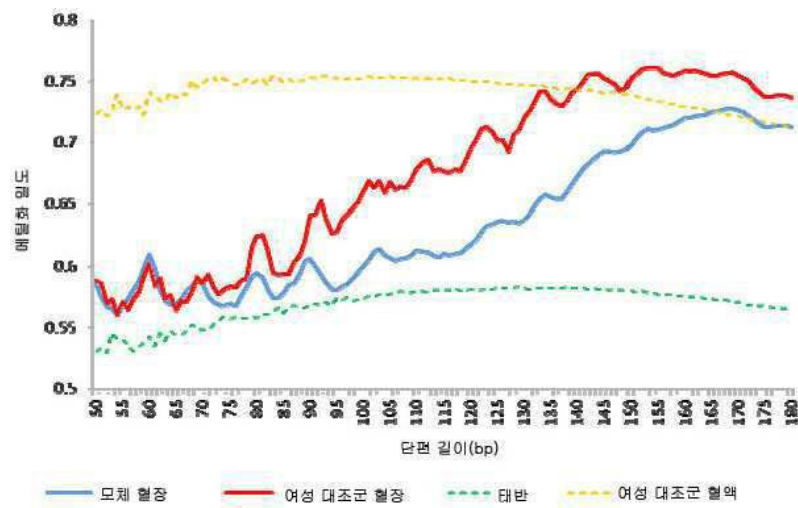
	제1 석달 모체 혈장		제3 석달 모체 혈장	
	고메틸화 ^a	저메틸화 ^b	고메틸화 ^a	저메틸화 ^b
예측된 유전자위의 수	3081	44455	1746	14930
태반 조직 ^c 데이터 에서 메틸화 밀도 > 40%인 유전자위의 수	1,678	N/A	1,525	N/A
태반 조직 ^c 데이터 에서 메틸화 밀도 < 60%인 유전자위의 수	N/A	23468	N/A	13475
태반 조직 ^c 및 모체 혈액 세포 데이터에서 마이닝된 DMR과 중복된 유전자위의 수	1457	21812	1279	12677

도면17a



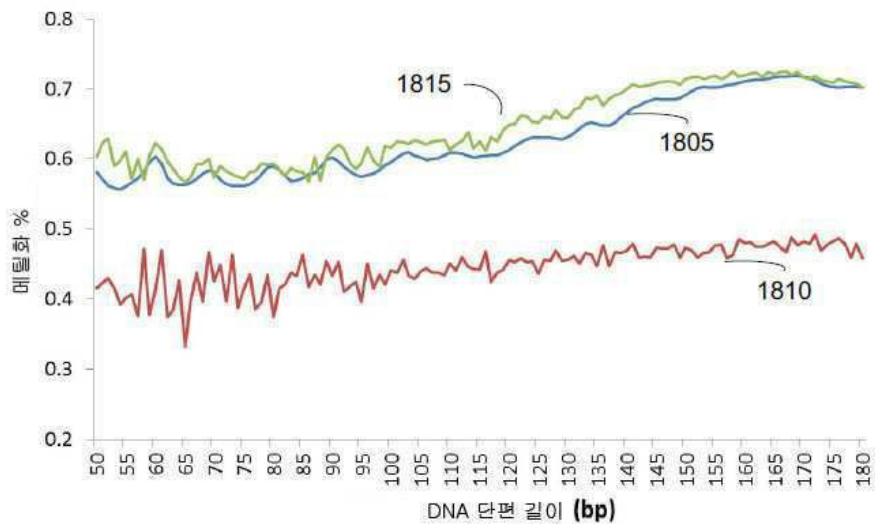
1700

도면17b



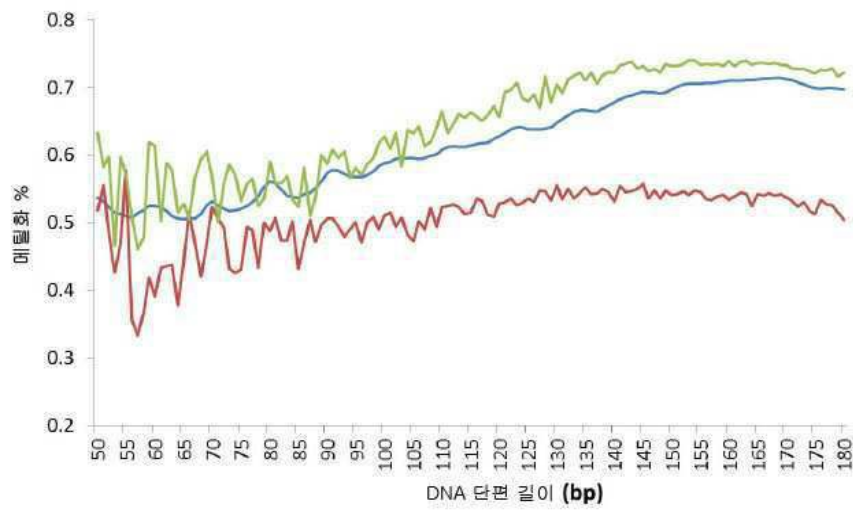
1750

도면18a



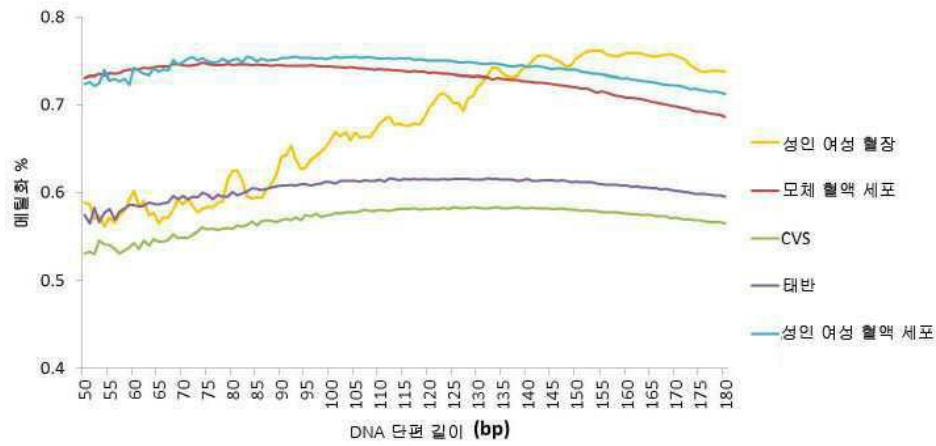
1800

도면18b



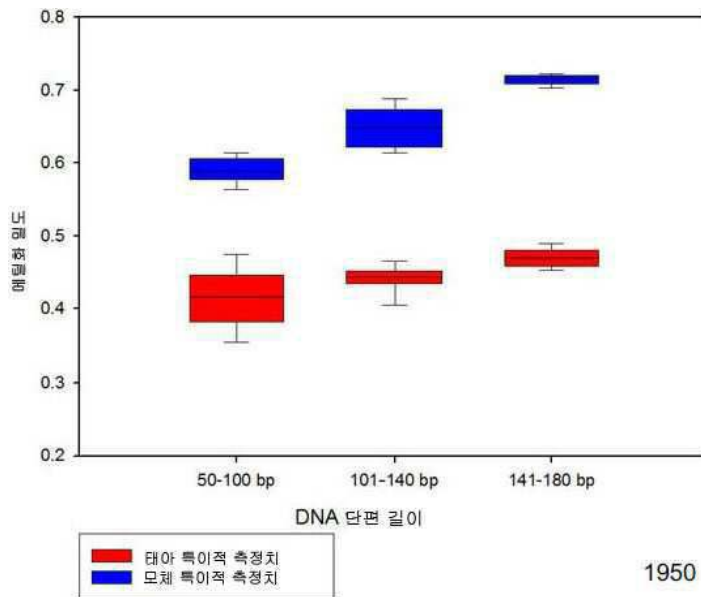
1850

도면19a



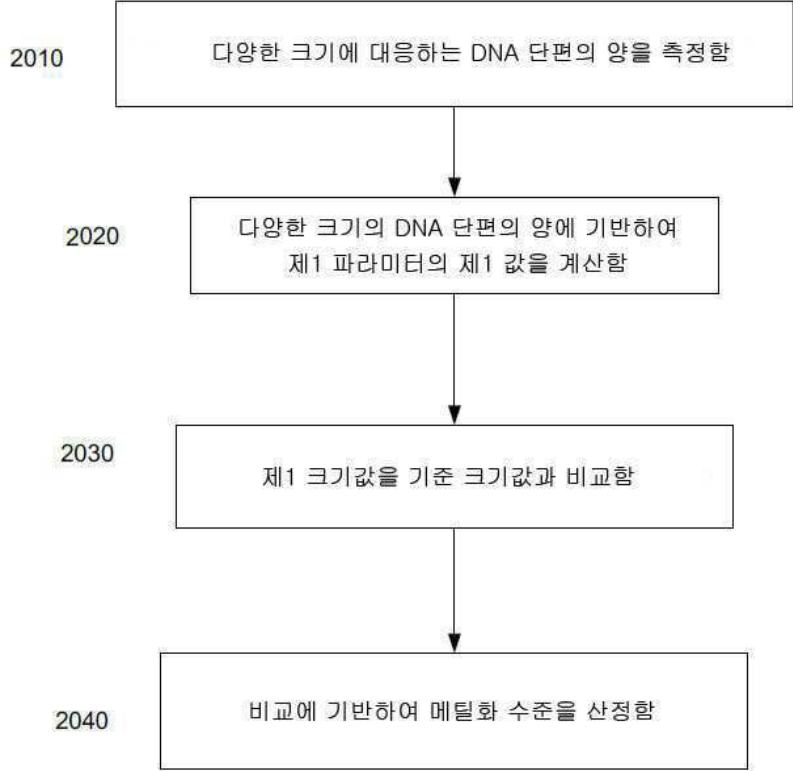
1900

도면19b



1950

도면20



2000

도면21a

염색체	HCC 환자			
	버피 코트	비종양 간 조직	종양 조직	수술 전 혈장
1	65.4%	65.7%	44.1%	60.2%
2	68.3%	68.2%	42.0%	58.6%
3	68.2%	68.5%	44.6%	60.9%
4	68.9%	67.7%	37.8%	56.9%
5	67.4%	66.7%	37.8%	56.9%
6	67.4%	67.7%	43.2%	59.9%
7	69.1%	68.8%	43.0%	60.0%
8	69.1%	68.0%	40.4%	60.0%
9	66.4%	66.2%	44.9%	60.6%
10	68.9%	68.2%	40.8%	58.5%
11	65.5%	65.6%	41.6%	58.5%
12	67.6%	67.4%	45.9%	62.0%
13	69.8%	68.8%	40.6%	60.2%
14	66.4%	66.2%	41.4%	59.2%
15	66.1%	66.8%	44.6%	61.0%
16	68.9%	67.6%	45.6%	60.7%
17	64.0%	64.9%	45.8%	60.2%
18	69.9%	69.1%	42.9%	60.0%
19	62.3%	62.7%	47.5%	61.0%
20	66.0%	65.3%	39.1%	57.8%
21	61.8%	61.1%	39.9%	57.7%
22	67.9%	67.6%	49.1%	63.8%
모든 상염색체	67.1%	66.9%	42.9%	59.7%

2100

도면21b

	미가공 측정치 카운트	맵핑된 측정치 카운트	비중복 단위 (상염색체만)	이용 가능한 측정치의 비율 %	가닥 별 깊이	CpG 커버%	가닥 별 커버되는 측정치를 갖는 CpG 부위 상 깊이
대조군 1	159554637	56323206	41653660	26%	1.12	54.86%	3.56
대조군 2	118056735	49232079	45741385	39%	1.23	60.26%	3.24
대조군 3	181913746	75753907	68081010	37%	1.82	65.74%	4.37
대조군 4	142782870	50195168	42370350	30%	1.13	59.15%	3.14
C06	262763161	197344994	88266088	33%	2.11	64.31%	3.65
버피 코트	579329822	439663895	375814248	65%	10.07	78.49%	21.67
비암 간 조직	549773334	417736311	355154964	65%	9.51	84.89%	18.21
종양 조직	568004640	487494291	176985129	31%	4.74	83.48%	8.32
수술 전 혈장	796183998	637735722	474751039	60%	12.72	87.18%	18.69
수술 후 혈장	921833848	624851729	550104801	60%	14.73	89.94%	24.81

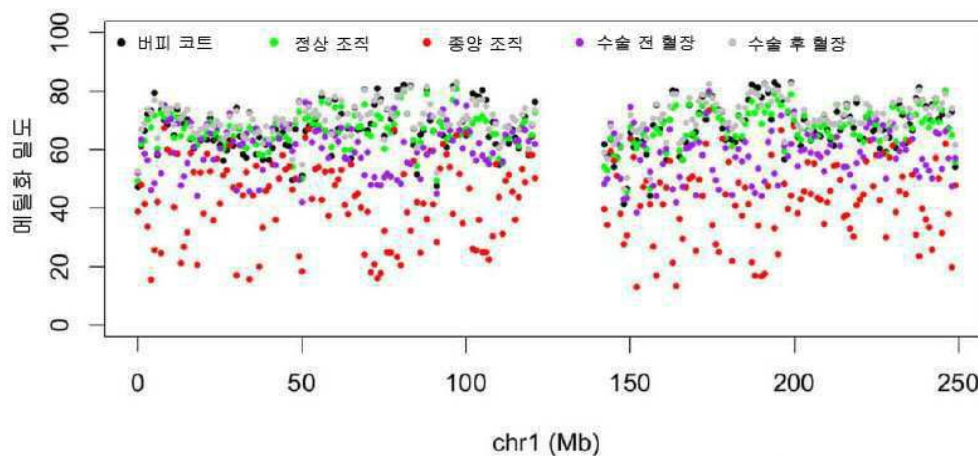
2150

도면22

염색체	혈장 DNA 메틸화 밀도					
	HCC 환자		대조군 1	대조군 2	대조군 3	대조군 4
	수술 전	수술 후				
1	60.2%	69.6%	70.0%	70.3%	70.1%	71.3%
2	58.6%	71.9%	72.2%	72.4%	72.3%	73.5%
3	60.9%	72.3%	72.7%	73.1%	72.9%	73.9%
4	56.9%	71.8%	71.9%	72.2%	72.2%	73.4%
5	56.9%	70.9%	71.0%	71.3%	71.3%	72.6%
6	59.9%	71.2%	71.6%	71.9%	71.7%	72.8%
7	60.0%	72.2%	72.8%	72.8%	72.5%	73.9%
8	60.0%	72.0%	72.6%	72.6%	72.4%	73.7%
9	60.6%	70.4%	71.0%	71.1%	71.0%	72.2%
10	58.5%	71.7%	72.0%	72.1%	72.0%	73.2%
11	58.5%	69.6%	70.2%	70.4%	70.2%	71.6%
12	62.0%	71.0%	71.6%	71.7%	71.5%	72.7%
13	60.2%	72.8%	73.2%	73.4%	73.2%	74.3%
14	59.2%	70.2%	70.7%	70.9%	70.8%	72.0%
15	61.0%	70.5%	70.6%	71.0%	70.9%	72.0%
16	60.7%	71.3%	72.4%	71.8%	71.6%	73.3%
17	60.2%	68.6%	69.3%	69.4%	69.3%	70.8%
18	60.0%	72.7%	72.9%	73.1%	73.0%	74.2%
19	61.0%	66.9%	68.1%	68.1%	67.6%	69.3%
20	57.8%	69.6%	70.3%	70.3%	70.3%	71.6%
21	57.7%	67.0%	64.0%	64.5%	65.7%	67.4%
22	63.8%	71.4%	72.5%	72.0%	71.9%	73.3%
모든 상염색체	59.7%	70.8%	71.2%	71.3%	71.2%	72.5%

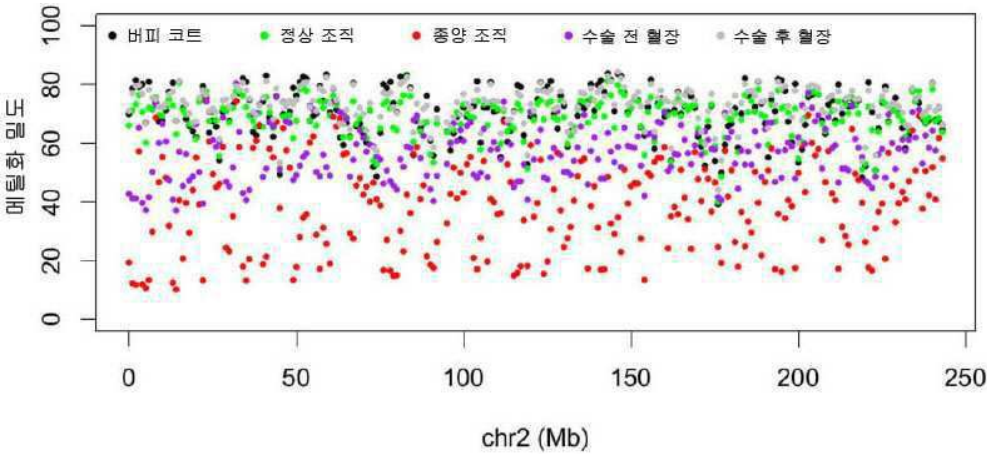
2200

도면23a



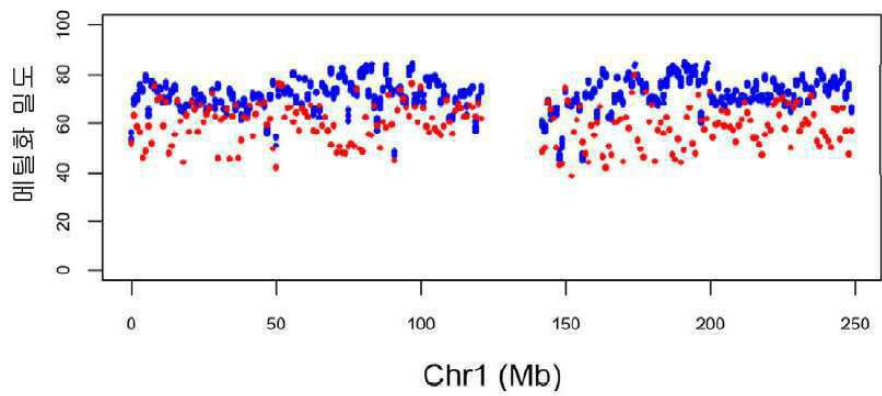
2300

도면23b



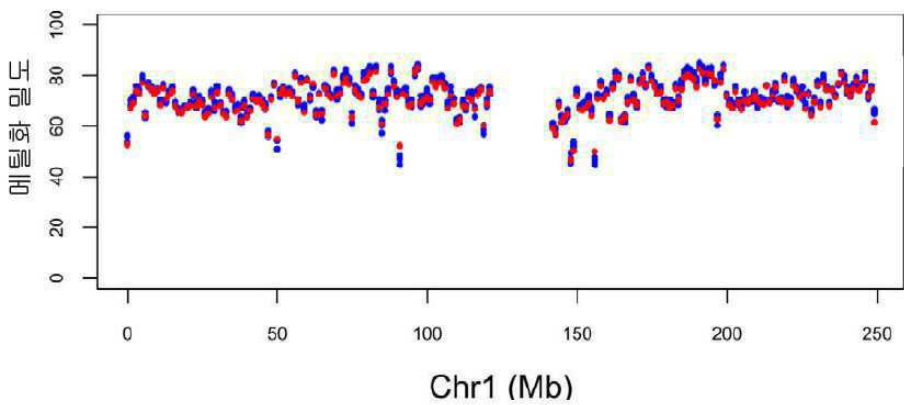
2350

도면24a



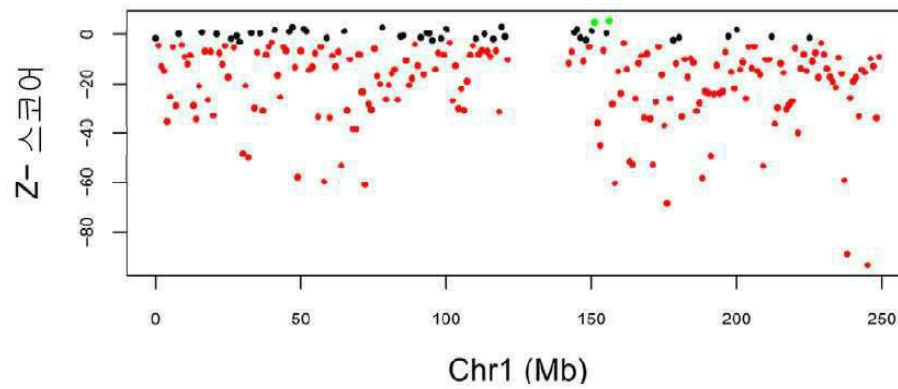
2400

도면24b



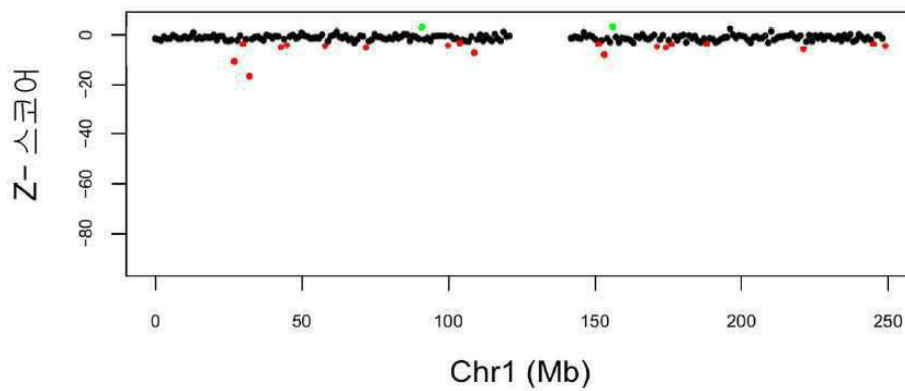
2450

도면25a



2500

도면25b



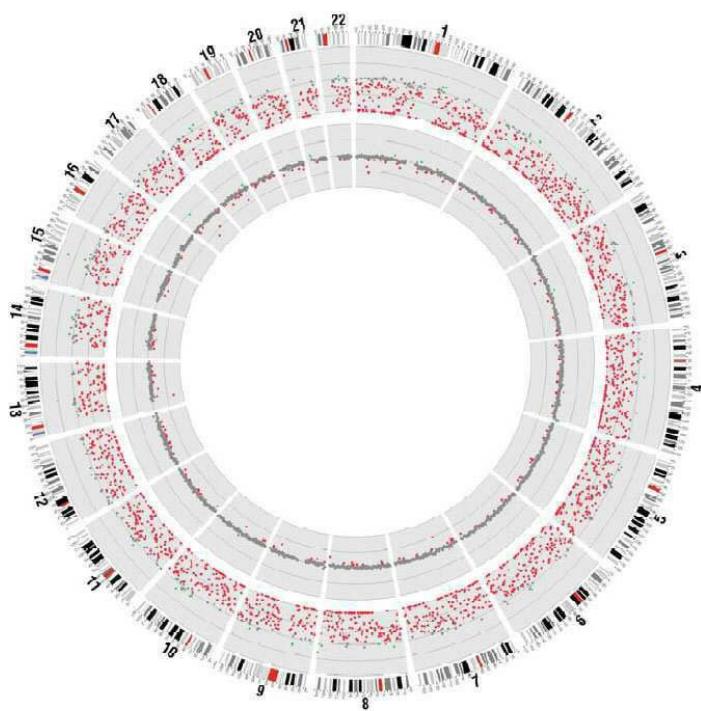
2550

도면26a

z-스코어	수술 전 혈장 (bin의 수)	수술 후 혈장 (bin의 수)
<-3	186 (80.9%)	19 (8.3%)
-3 내지 3	42 (18.3%)	209 (90.3%)
>3	2 (0.9%)	2 (0.9%)
	230 (100%)	230 (100%)

2600

도면26b



2620

도면26c

z-스코어	수술 전 혈장 (bin의 수)	수술 후 혈장 (bin의 수)	C06	대조군 1 (bin의 수)	대조군 2 (bin의 수)	대조군 3 (bin의 수)	대조군 4 (bin의 수)
<-3	2330 (85.2%)	171 (6.3%)	29 (1.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
-3 to 3	362 (13.3%)	2556 (93.5%)	2688 (98.3%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)
>3	42 (1.5%)	7 (0.2%)	17 (0.62%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
총	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)

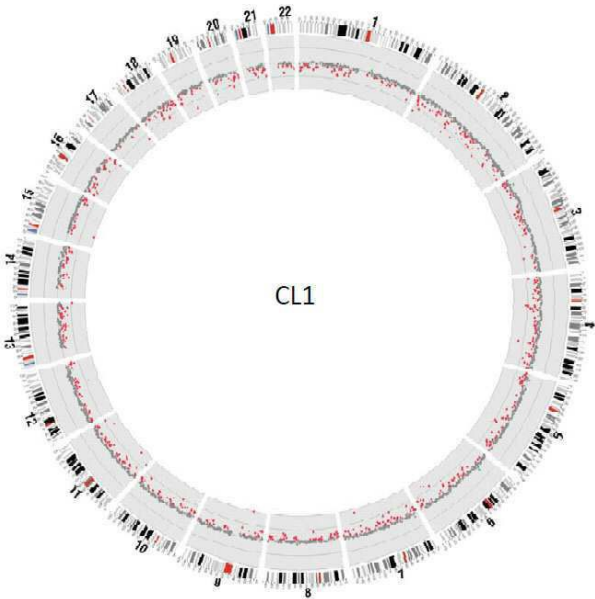
2640

도면26d

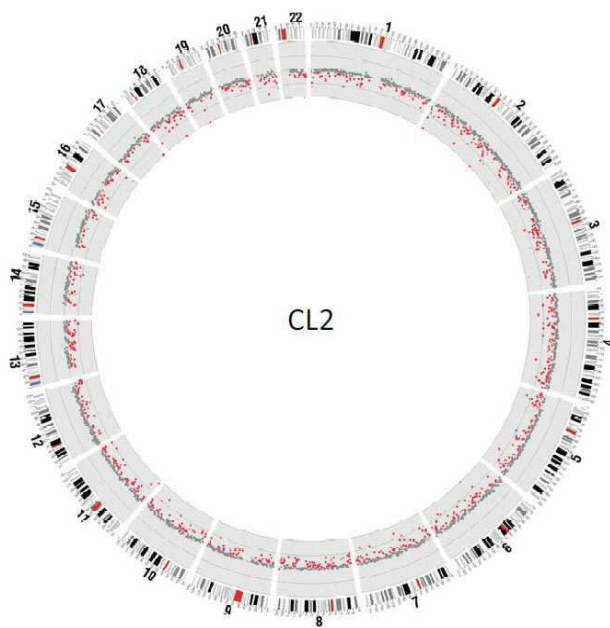
	메틸화 CHH의 비율	메틸화 CHG의 비율	메틸화 시토신의 비율	메틸화 CpG의 비율 (메틸화 밀도)
HCC 환자의 버피 코트	0.13	0.14	5.74	67.08
비종양 간 조직	0.16	0.15	5.45	66.85
종양 조직	0.09	0.09	3.29	42.88
수술 전 혈장	0.09	0.10	3.99	59.67
수술 후 혈장	1.25	0.96	6.52	70.75
대조군 혈장 1	0.12	0.12	6.38	71.17
대조군 혈장 2	0.10	0.10	5.62	71.33
대조군 혈장 3	0.09	0.09	5.57	71.20
대조군 혈장 4	0.10	0.10	5.93	72.48

2660

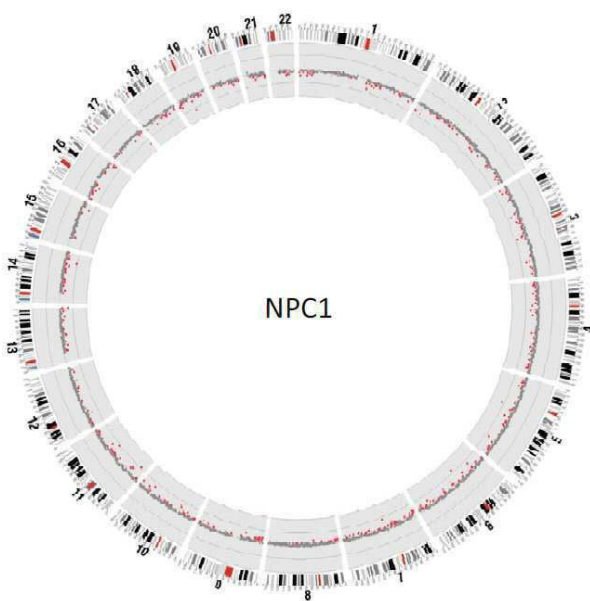
도면27a



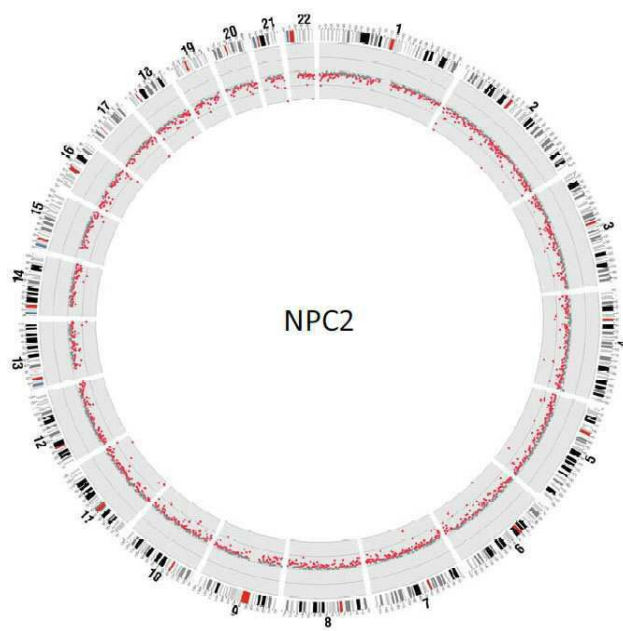
도면27b



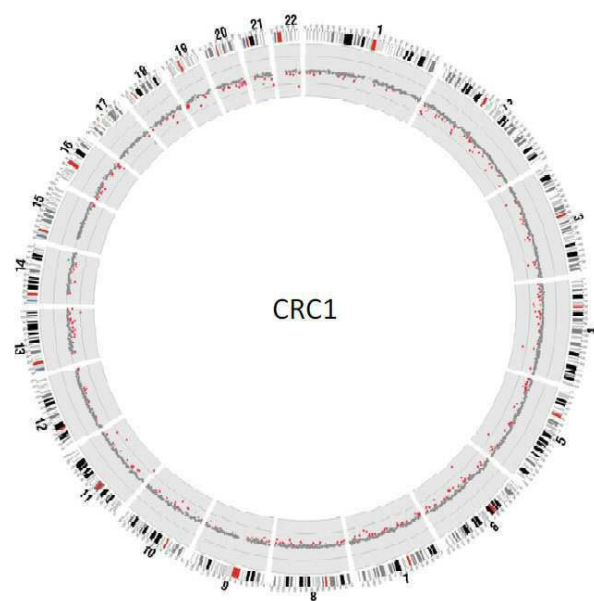
도면27c



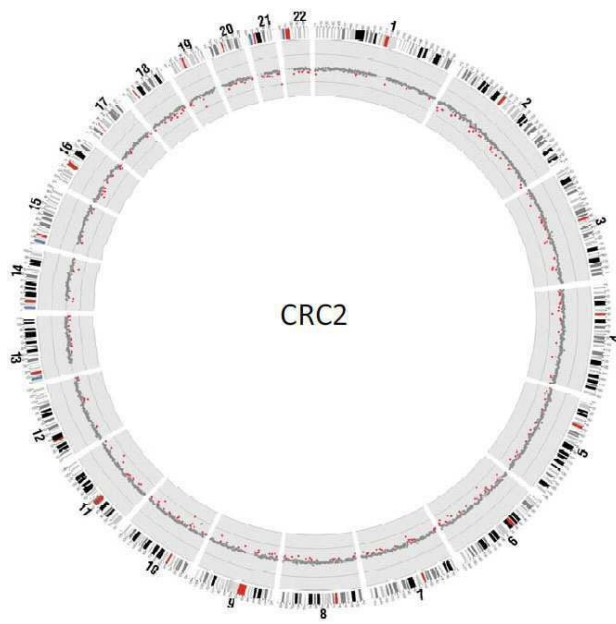
도면27d



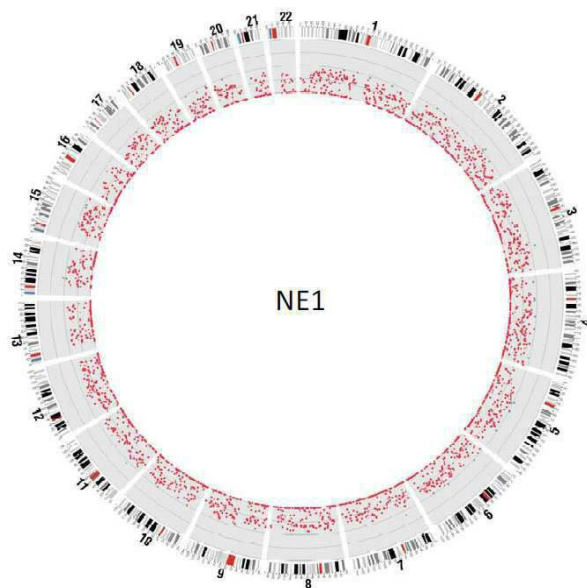
도면27e



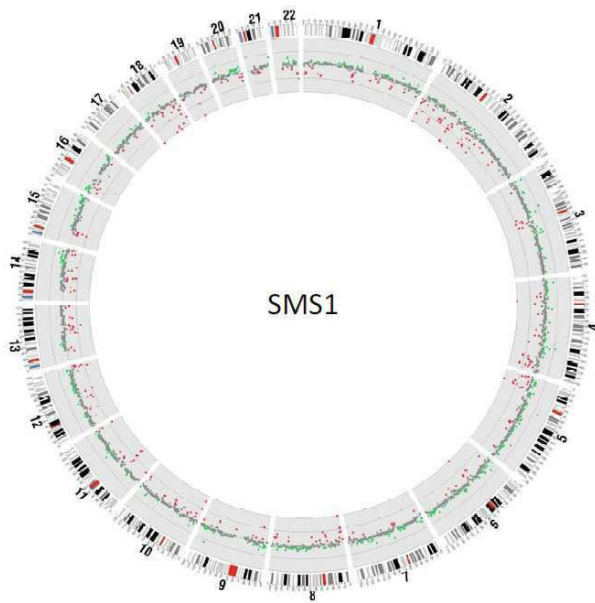
도면27f



도면27g



도면27h



도면27i

	미가공 측정치 카운트	맵핑된 측정치 카운트	비중복 단위	이용 가능한 측정치의 비율 %	가닥 별 길이	CpG 커버 %	가닥 별 커버되는 측정치를 갖는 CpG 부위 상 길이
CL1	120807588	91750018	76890830	64%	0.69	33.32%	2.87
CL2	101344850	76473792	66246435	65%	0.59	30.26%	2.79
NPC1	101883643	75281330	57236188	56%	0.51	27.87%	2.72
NPC2	102745073	74211781	56841611	55%	0.51	28.74%	2.70
CRC1	87732645	66456695	56112543	64%	0.50	27.38%	2.63
CRC2	109955730	82594007	67253101	61%	0.60	29.91%	2.90
NE1	105514000	79923333	64534726	61%	0.58	27.86%	2.58
SMS1	94727279	71698718	64426529	68%	0.58	29.23%	2.79

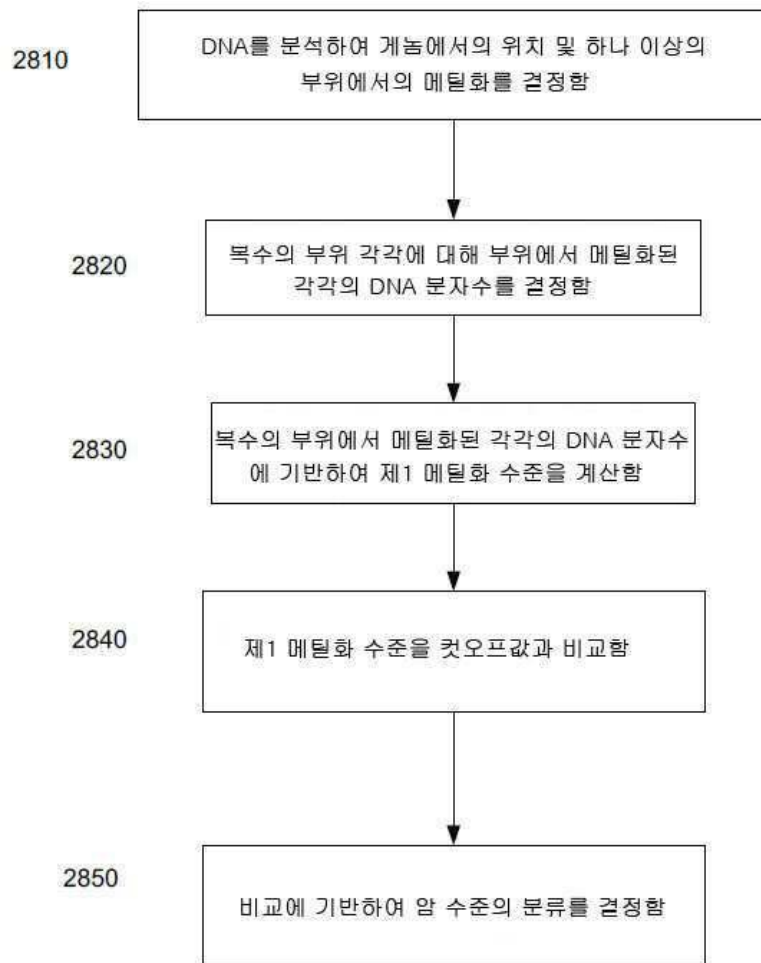
2780

도면27j

z-스코어	CL1	CL2	NPC1	NPC2	CRC1	CRC2	NE1	SMS1
< -3	34.9%	33.9%	17.6%	54.3%	13.5%	9.8%	98.1%	11.2%
-3 내지 3	65.0%	65.9%	82.4%	45.6%	86.5%	90.2%	1.7%	70.8%
>3	0.07%	0.15%	0.04%	0%	0.07%	0%	0.15%	18.0%

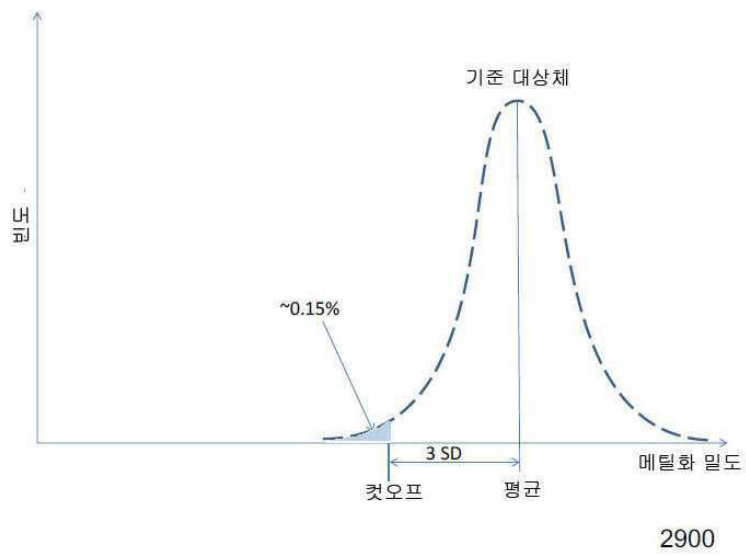
2790

도면28

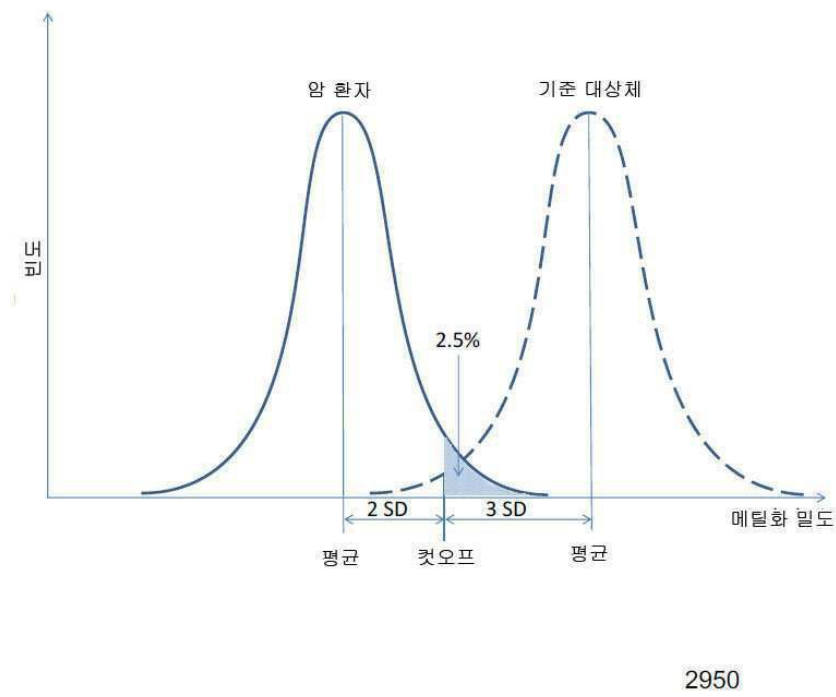


2800

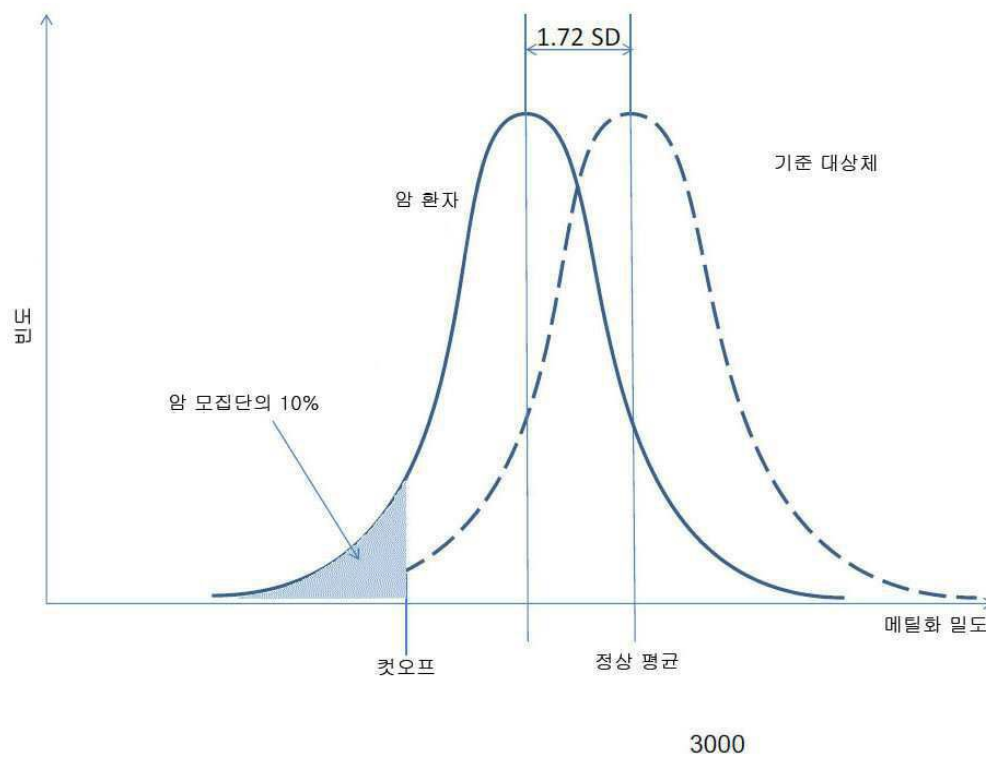
도면29a



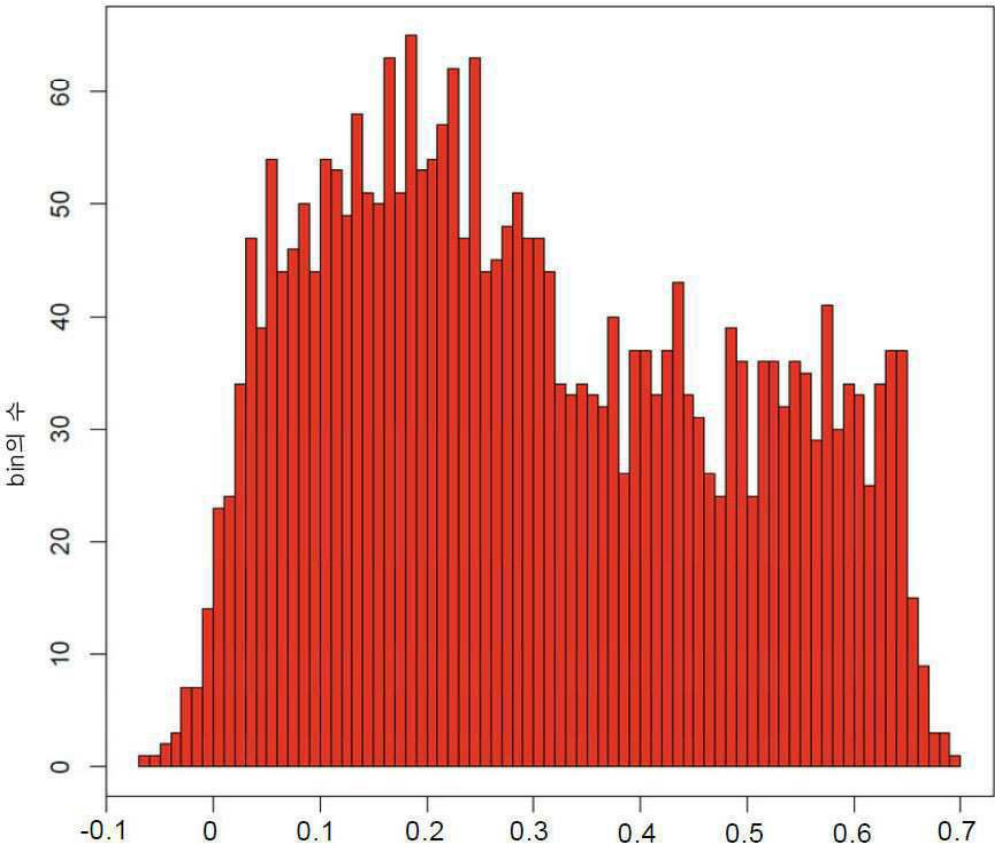
도면29b



도면30



도면31



건강한 대상체의 혈장 DNA 및 암 환자 종양 조직의 평균 간 메틸화 밀도 차이

3100

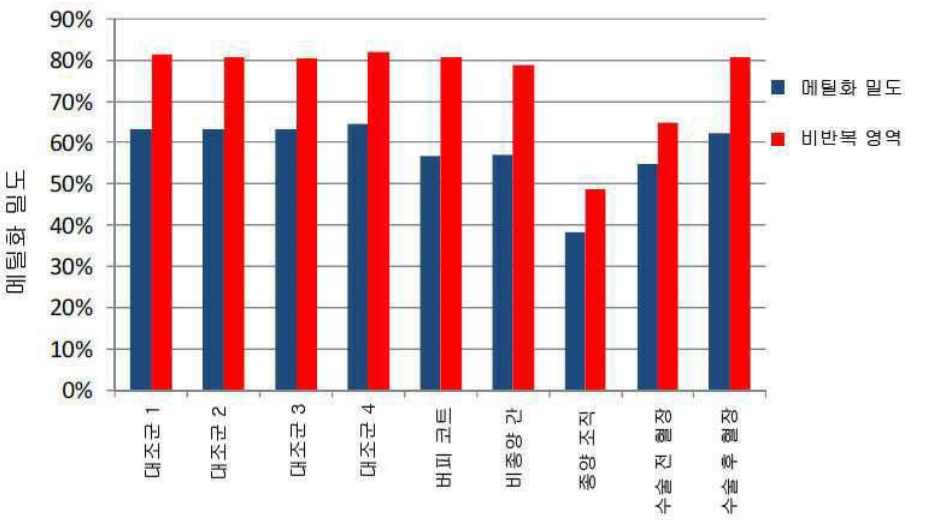
도면32a

5%에서의 깊이 회석 인자	서열분석 깊이 (일배체형 게놈의 배)	수술 전 혈장 중 저메틸화 bin(z- 스코어 <-3)의 수	수술 전 혈장 중 저메틸화 bin(z- 스코어 <-3)의 수
0.8	1.76	1314	48.06
0.6	1.32	1323	48.39
0.4	0.88	1220	44.62
0.2	0.44	1094	40.01
0.1	0.22	900	32.92
0.05	0.11	645	23.59
0.01	0.022	207	7.57

2%에서의 깊이 회석 인자	서열분석 깊이 (일배체형 게놈의 배)	수술 전 혈장 중 저메틸화 bin(z- 스코어 <-3)의 %	수술 전 혈장 중 저메틸화 bin(z- 스코어 <-3)의 %
0.8	1.76	491	17.96
0.6	1.32	493	18.03
0.4	0.88	460	16.83
0.2	0.44	439	16.06
0.1	0.22	356	13.02
0.05	0.11	297	10.86
0.01	0.022	231	8.45

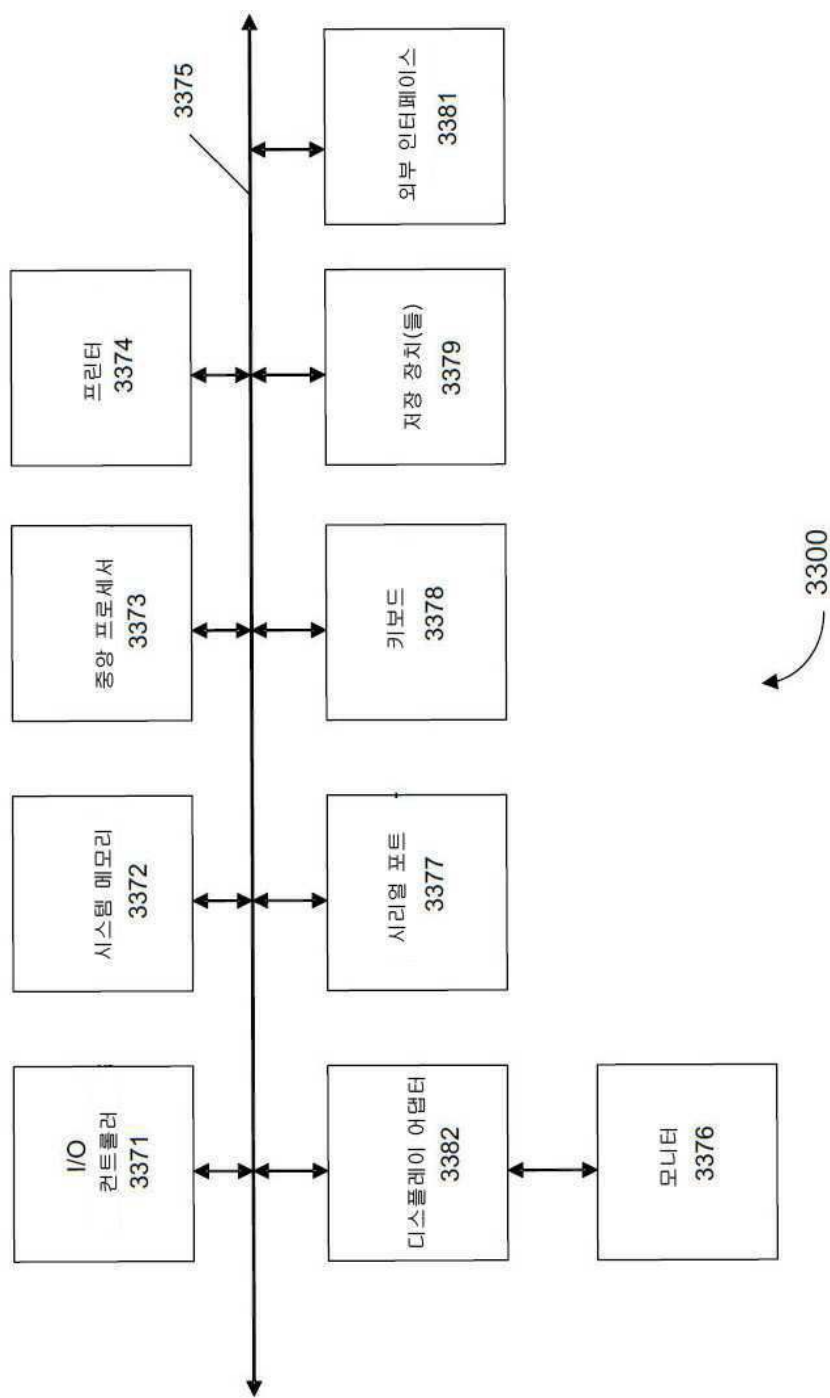
3200

도면32b

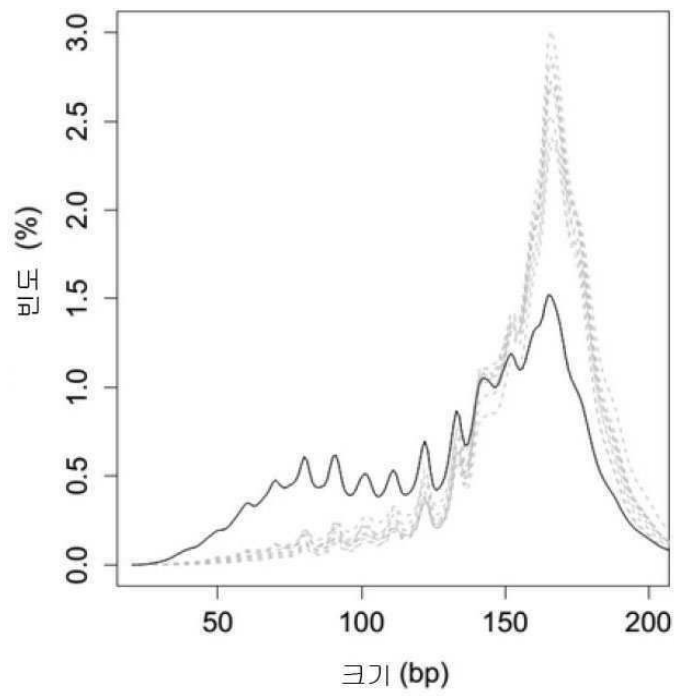


3250

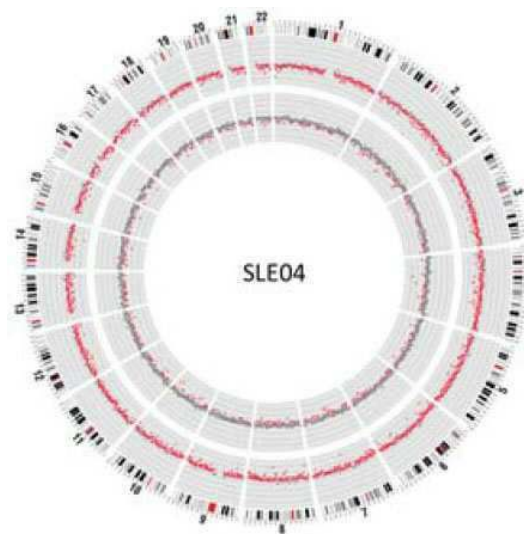
도면33



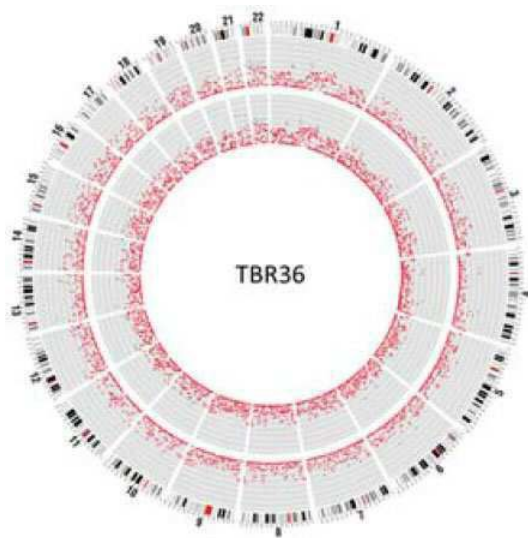
도면34a



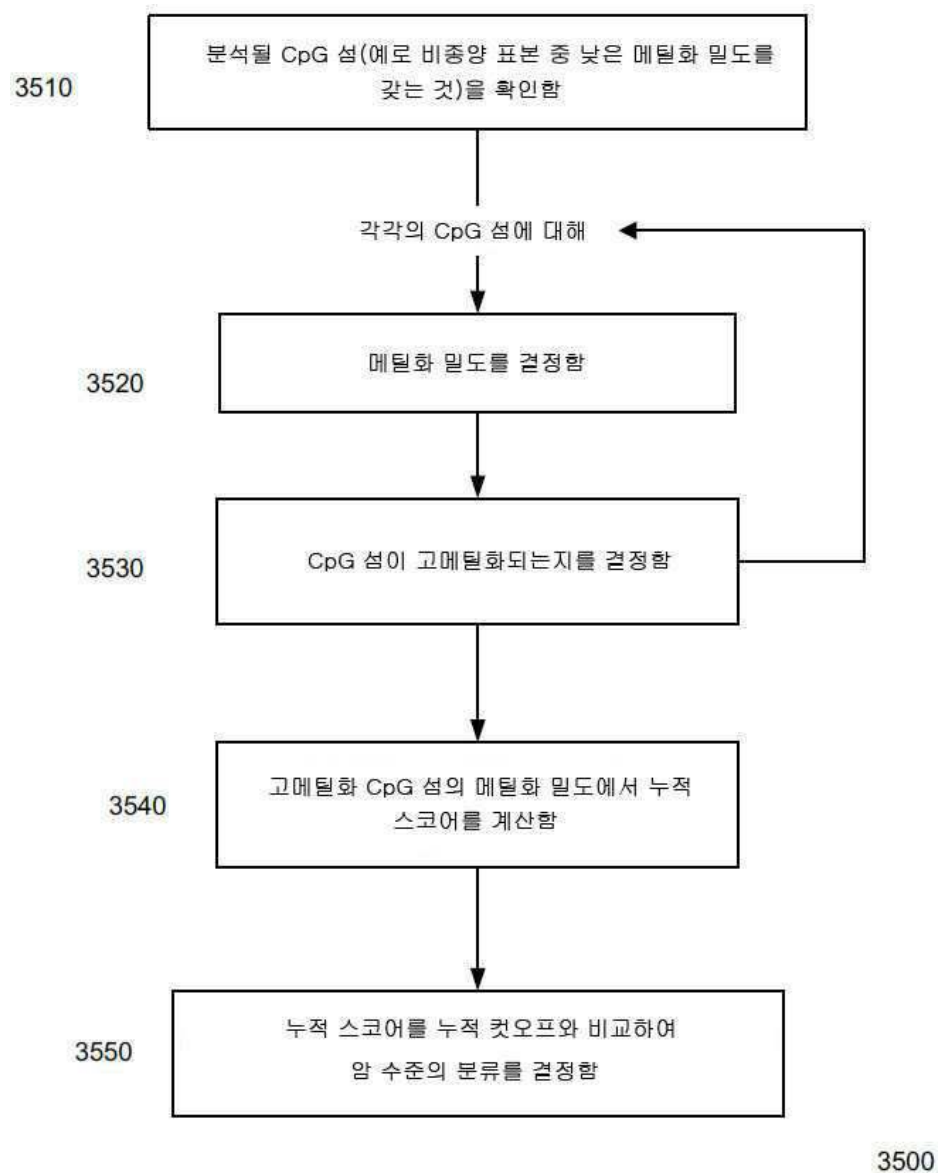
도면34b



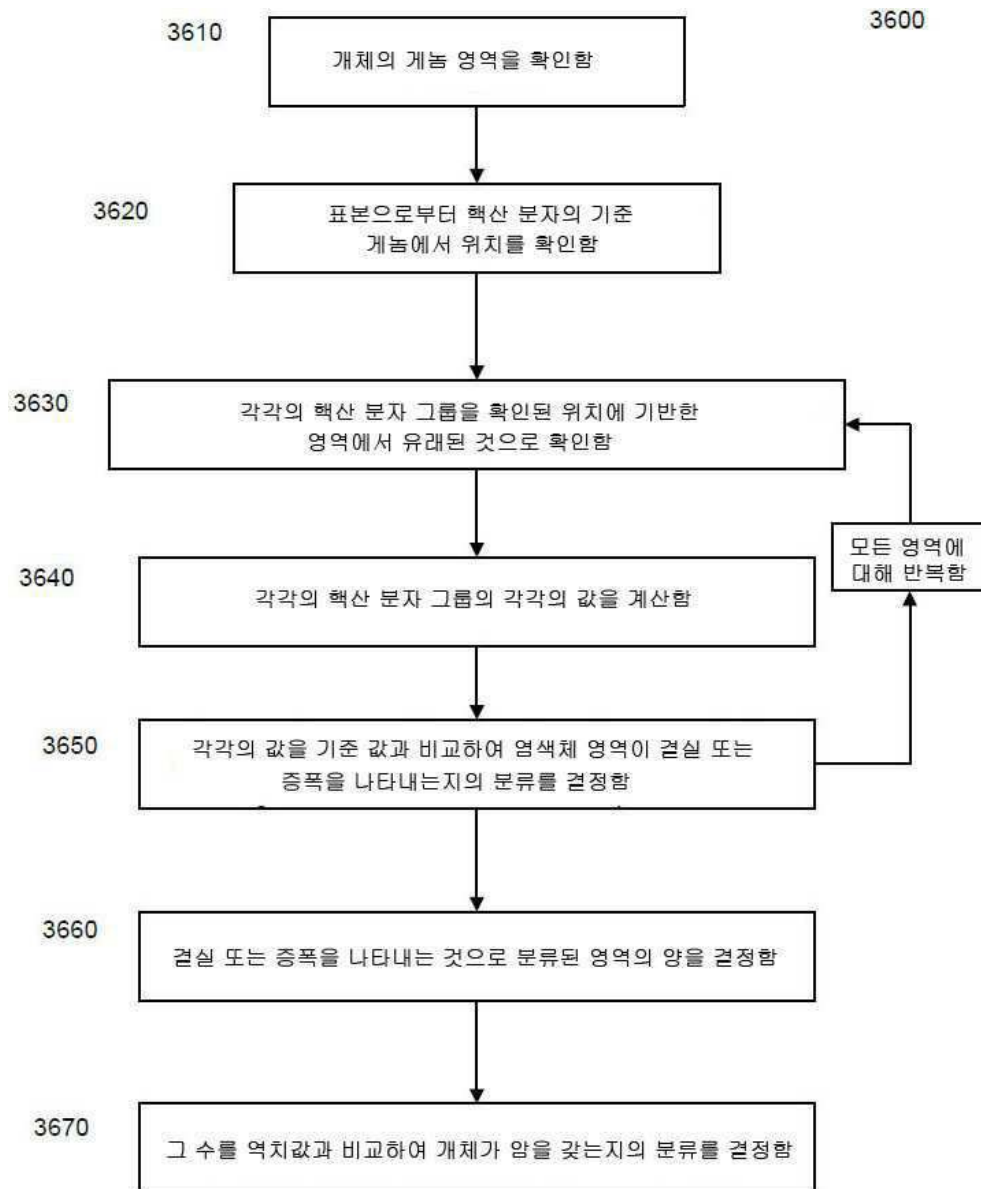
도면34c



도면35



도면36

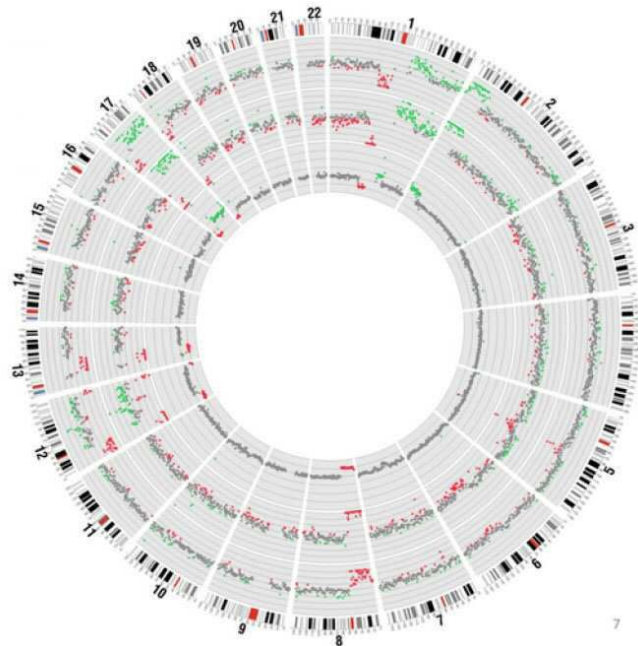


도면37a

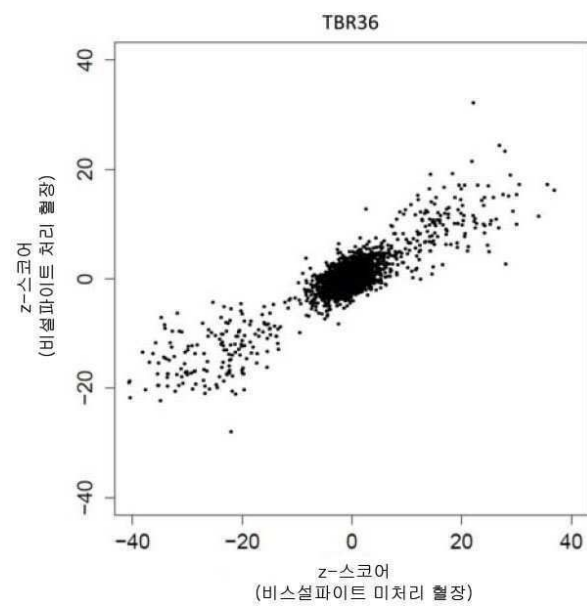
표본: TBR36
종양 DNA 분획: 40%

내부에서 외부로:

- 1) 종양 CNA
- 2) 혈장 CNA(비-BS)
- 3) 혈장 CNA(BS)



도면37b



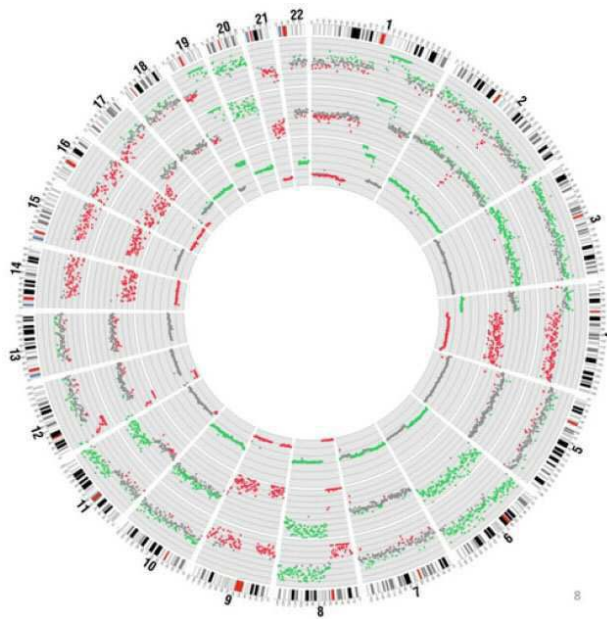
도면38a

표본: TBR34

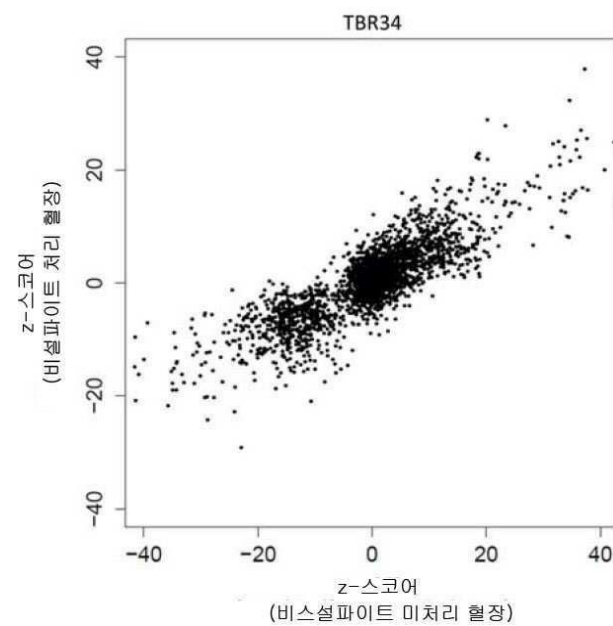
종양 DNA 분석: 30%

내부에서 외부로:

- 1) 종양 CNA
- 2) 혈장 CNA(비-BS)
- 3) 혈장 CNA(BS)



도면38b

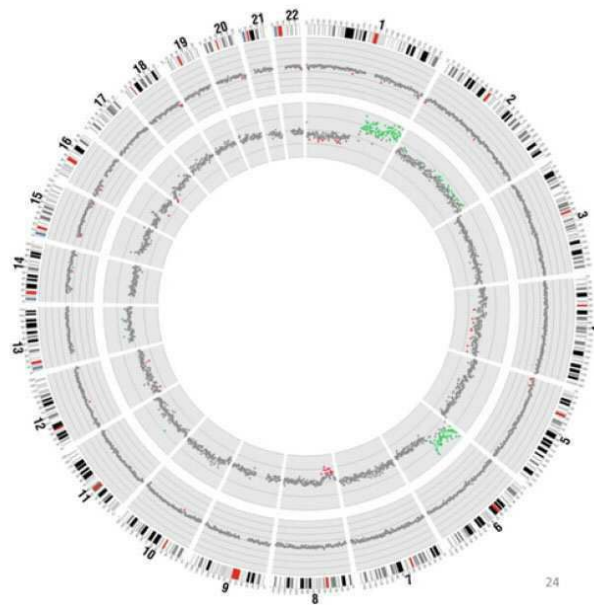


도면39a

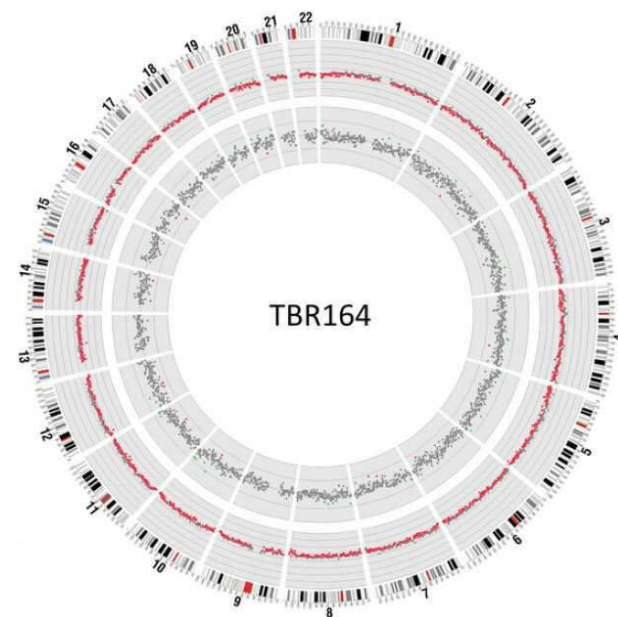
표본: TBR240
종양 크기: 2.8cm x 2.3cm

내부에서 외부로:

- 1) 혈장 CNA z-스코어
- 2) 혈장 메틸화 z-스코어



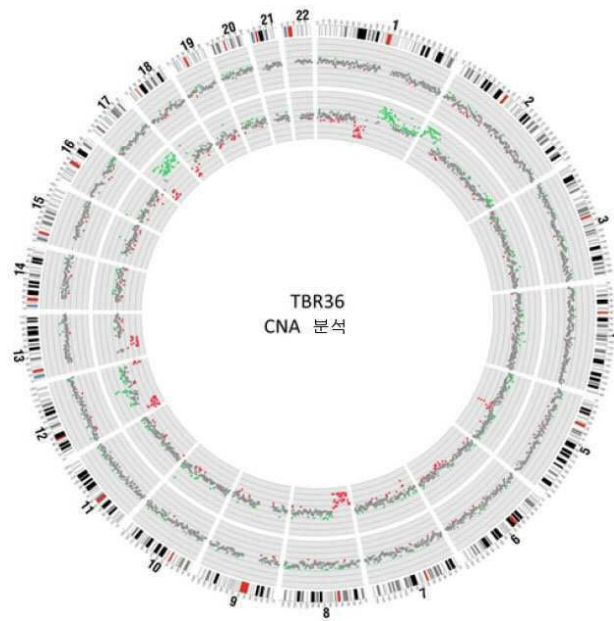
도면39b



도면40a

내부 고리 :
처리 전 혈장

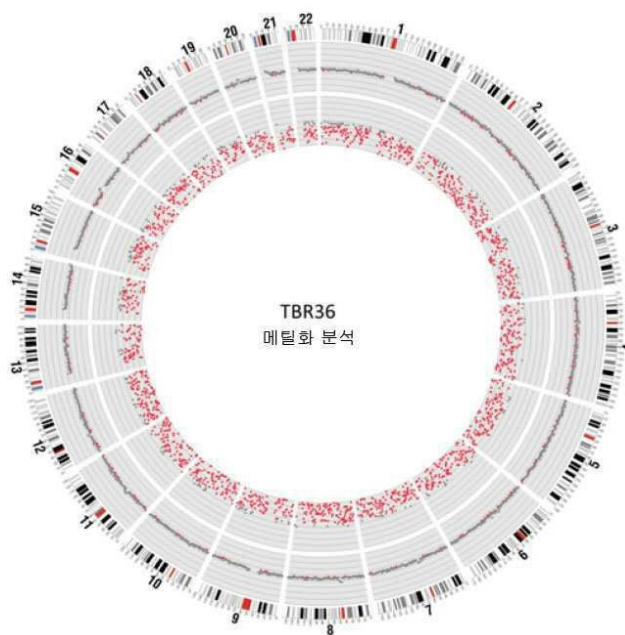
외부 고리 :
처리 후 혈장



도면40b

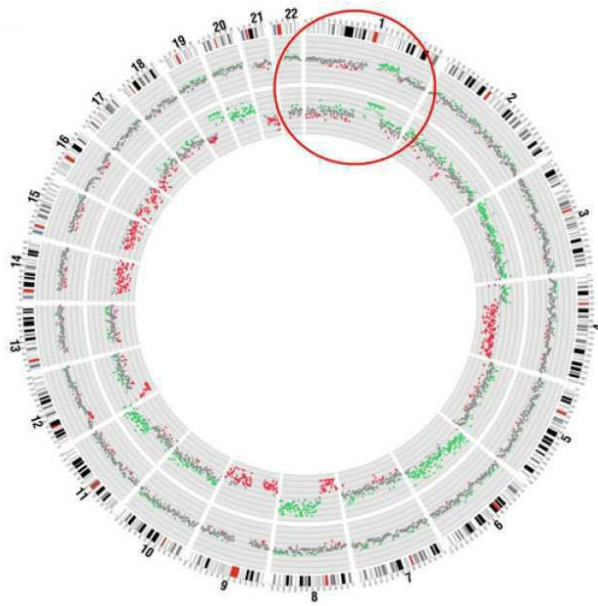
내부 고리 :
처리 전 혈장

외부 고리 :
처리 후 혈장



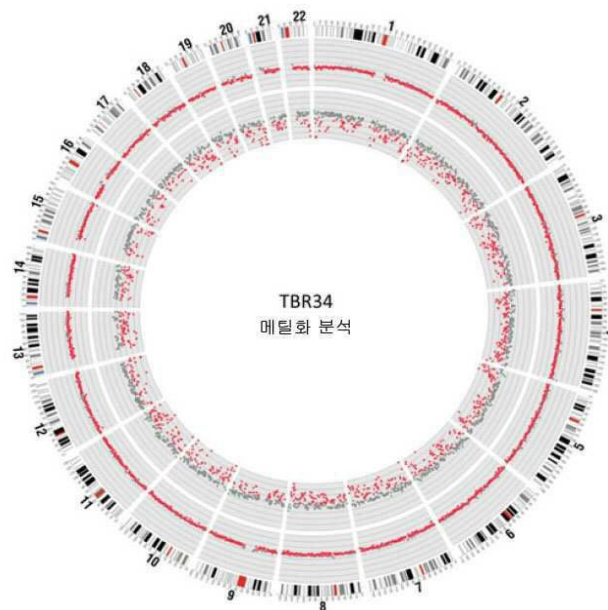
도면41a

표본: TBR34
 종양 DNA 분획: 30%
 외부에서 내부로:
 1) 수술 후 CNV
 2) 수술 전 CNV

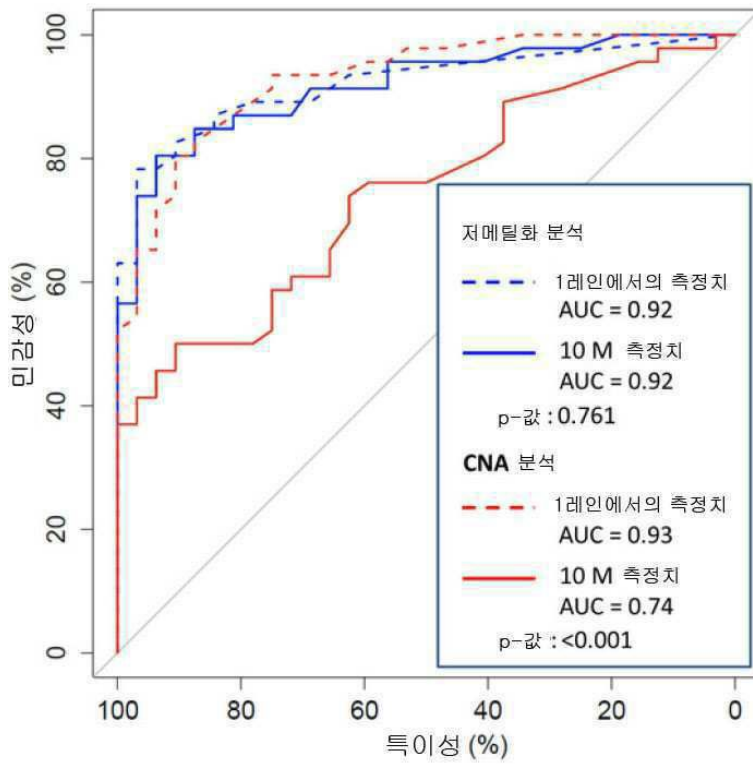


도면41b

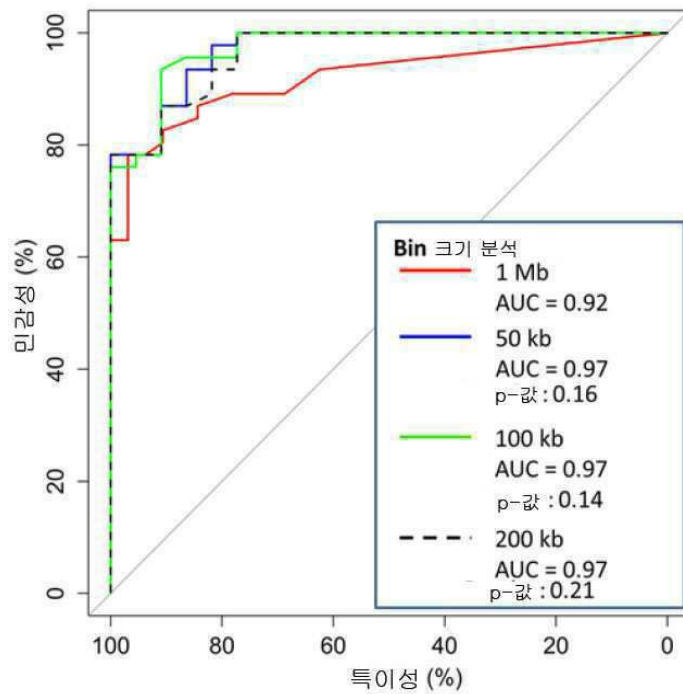
내부 고리 :
 처리 전 혈장
 외부 고리 :
 처리 후 혈장



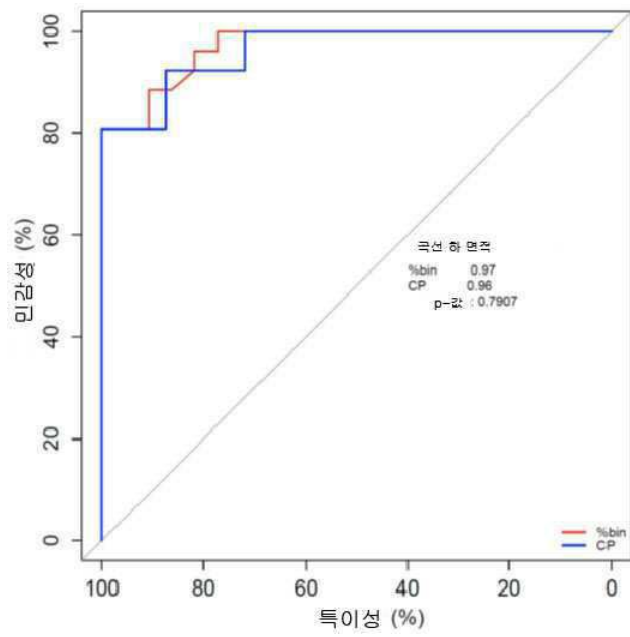
도면42



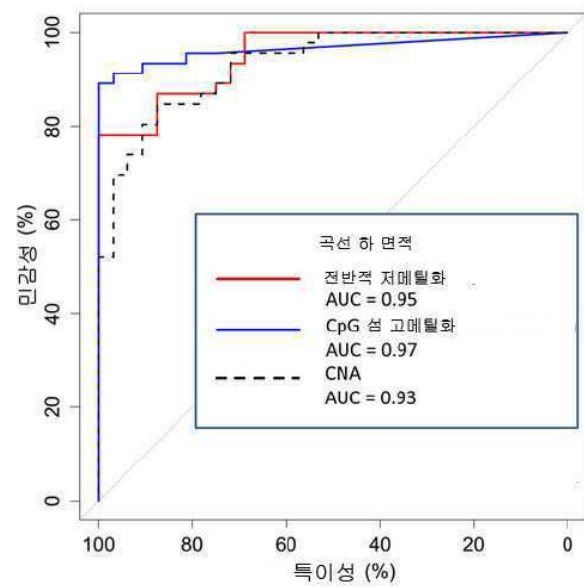
도면43



도면44a



도면44b



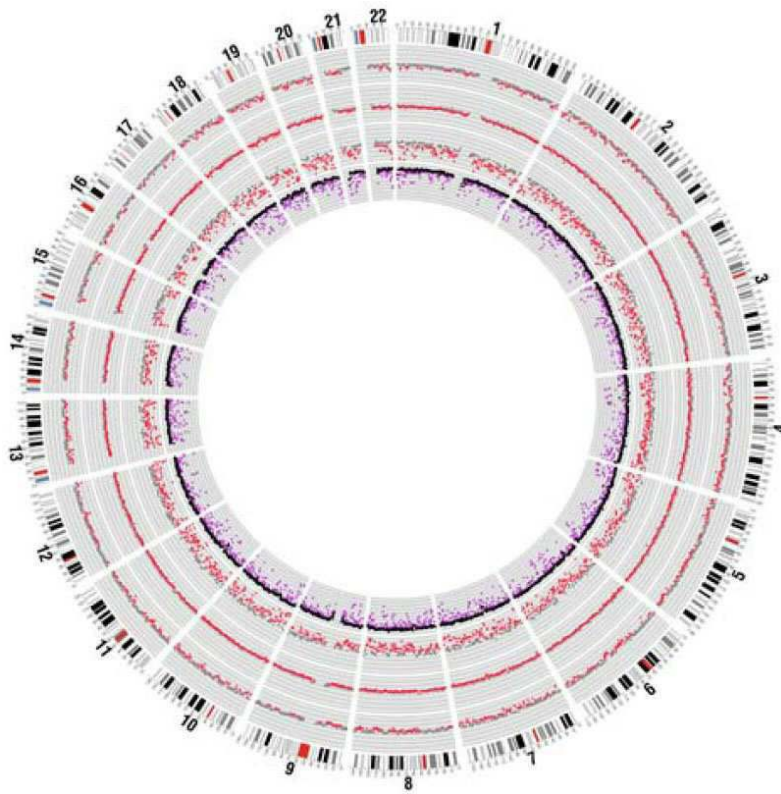
도면45

케이스 번호	BCLC 단계	종양(들) 최대 치수 (cm)	CP 스코어			임의 유형의 변화의 존재
			일반적 저메틸화	CpG 섬의 고메틸화	CNA	
TBR36	A1	16	<u>88585</u>	<u>80.9</u>	<u>3777</u>	양성
TBR34	A1	13	<u>37573</u>	<u>221.7</u>	<u>8650</u>	양성
HOT198	A1	12	<u>93132</u>	<u>52.5</u>	<u>6803</u>	양성
HOT205	A1	6	<u>1146</u>	1.7	175	양성
HOT192	A1	5.5	<u>25914</u>	<u>119.5</u>	<u>1668</u>	양성
HOT227	A1	5.5	<u>36333</u>	<u>155.4</u>	<u>5008</u>	양성
HOT238	A1	5.5	<u>12187</u>	<u>105.3</u>	<u>1978</u>	양성
HOT197	B	5.3	<u>1714</u>	<u>9.5</u>	28	양성
HOT156	A1	5	<u>16529</u>	<u>115.9</u>	<u>1947</u>	양성
HOT236	A1	3.7	<u>7993</u>	<u>5.9</u>	85	양성
HOT170	A1	3.5	<u>22271</u>	<u>81.9</u>	<u>1542</u>	양성
HOT229	A1	3.1	383	<u>5.4</u>	71	양성
HOT233	A1	3	189	<u>7.5</u>	<u>322</u>	양성
HOT240	A1	2.8	<u>1277</u>	<u>14.0</u>	<u>972</u>	양성
HOT222	A1	2.6	<u>18832</u>	<u>10.5</u>	172	양성
HOT162	A1	2.5	<u>26112</u>	<u>41.0</u>	<u>3048</u>	양성
HOT172	A1	2.5	<u>12909</u>	<u>98.1</u>	<u>1305</u>	양성
HOT164	A1	2.3	<u>12604</u>	<u>4.8</u>	210	양성
HOT215	A1	2.3	<u>25824</u>	<u>54.8</u>	<u>714</u>	양성
HOT207	A1	2.1	<u>3377</u>	<u>12.0</u>	101	양성
HOT224	A1	2	367	0.0	13	음성
HOT159	A1	1.5	<u>1367</u>	1.4	26	양성
HOT204	A1	1.5	102	<u>4.5</u>	198	양성
HOT151	A1	1.5	153	2.8	9	음성
HOT208	A1	1.2	<u>5867</u>	<u>4.7</u>	<u>214</u>	양성
HOT167	A1	1	<u>21544</u>	<u>4.5</u>	54	양성

도면46

종양 유형	케이스 번호	단계/종양 최대 치수	CP 스코어			임의 유형의 변화의 존재
			일반적 저메틸화	CpG 섬의 고메틸화	CNA	
유방암	TBR150	전이	<u>14581</u>	<u>146.7</u>	<u>5123</u>	양성
	TBR126	T2N1M0 4.5 cm	<u>10415</u>	<u>34.9</u>	<u>1951</u>	양성
	TBR127	T2N1M0 2.5 cm	<u>2598</u>	<u>4.3</u>	54	양성
	TBR128	T2N0M0 2 cm	157	<u>7.7</u>	<u>2250</u>	양성
	TBR111	T1N0M0 0.9 cm	298	0.0	128	음성
평활근 육종	TBR051	전이	<u>4787</u>	<u>79.8</u>	<u>7986</u>	양성
신경내분비 종양	TBR052	전이	<u>112518</u>	<u>119.8</u>	<u>14681</u>	양성
폐암	TBR164	전이	<u>51818</u>	<u>47.3</u>	<u>4714</u>	양성
	TBR012	전이	<u>4351</u>	<u>7.7</u>	<u>2213</u>	양성
	TBR014	전이	<u>2356</u>	<u>12.2</u>	<u>825</u>	양성
	TBR177	T2N1M0 3.9 cm	<u>1072</u>	<u>10.6</u>	25	양성
크린두암	TBR031	전이	<u>1028</u>	<u>4.4</u>	<u>1860</u>	양성
	TBR125	전이	<u>20949</u>	<u>162.5</u>	<u>6708</u>	양성
	TBR124	T4N1	270	<u>4.9</u>	37	양성
	TBR123	T1N2M0	<u>6871</u>	<u>11.8</u>	80	양성
	TBR062	T3N1M0	111	<u>30.9</u>	41	양성
	TBR108	T3N1M0	93	<u>7.7</u>	26	양성
	TBR107	T3N0M0	<u>2439</u>	<u>41.2</u>	205	양성
	TBR106	T1N1M0	<u>2281</u>	<u>5.9</u>	97	양성
	TBR099	T1N0M0	<u>6187</u>	<u>10.2</u>	26	양성

도면47

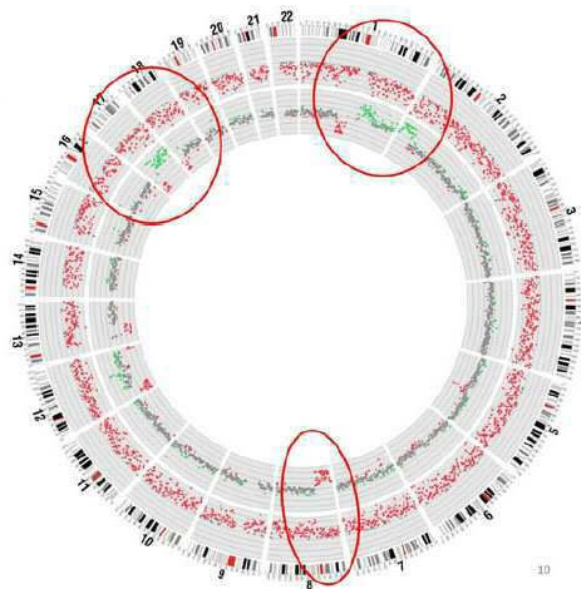


도면48a

표본: TBR36

내부에서 외부로:

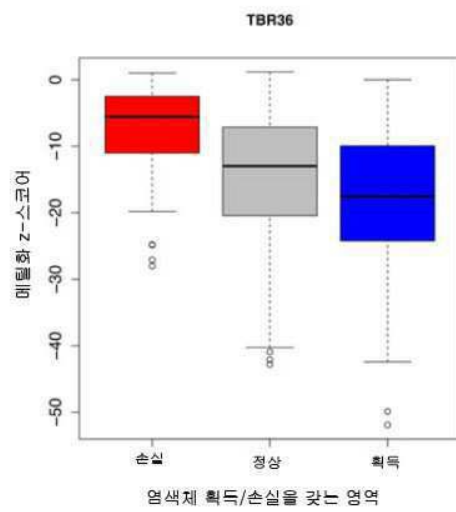
- 1) CNA z-스코어
- 2) 메틸화 z-스코어



도면48b

표본 : TBR36

중량 DNA 분석: 40%



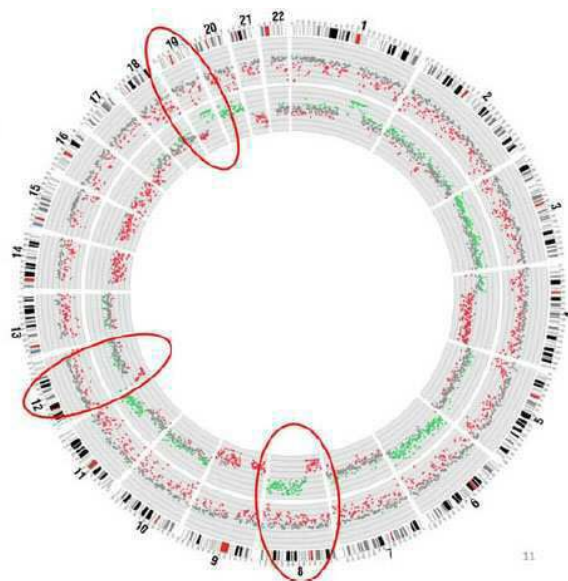
도면49a

표본 : TBR34

중량 DNA 분석: 30%

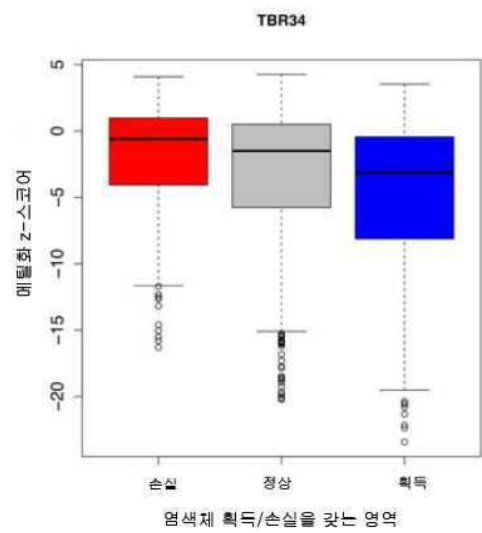
내부에서 외부로:

- 1) CNA z-스코어
- 2) 메틸화 z-스코어

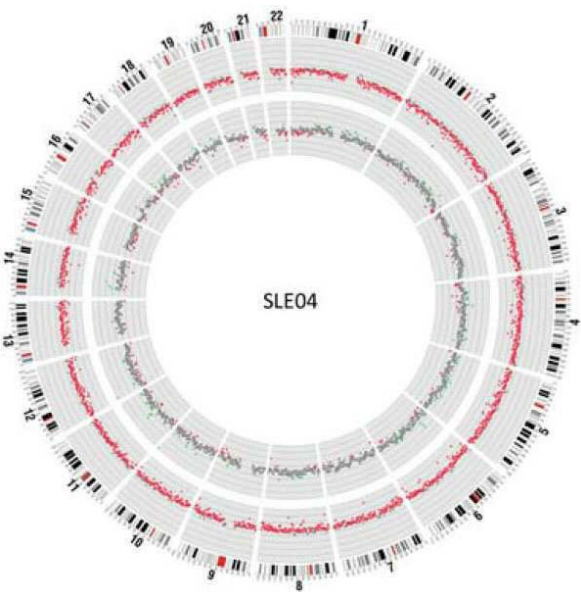


도면49b

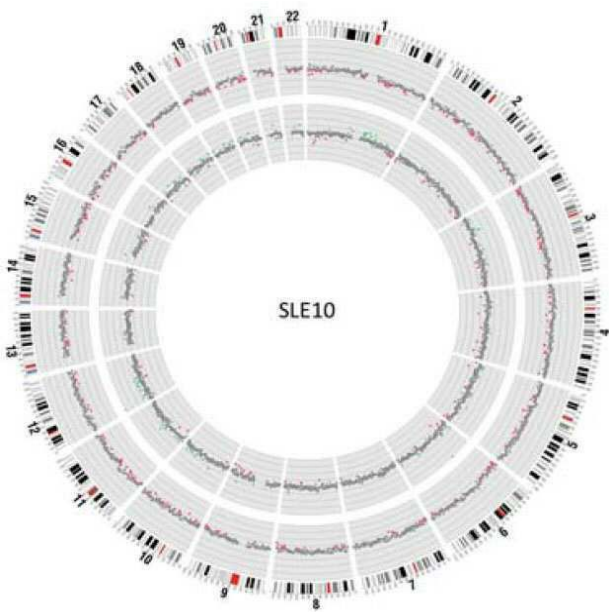
표본 : TBR34
중양 DNA 분획: 30%



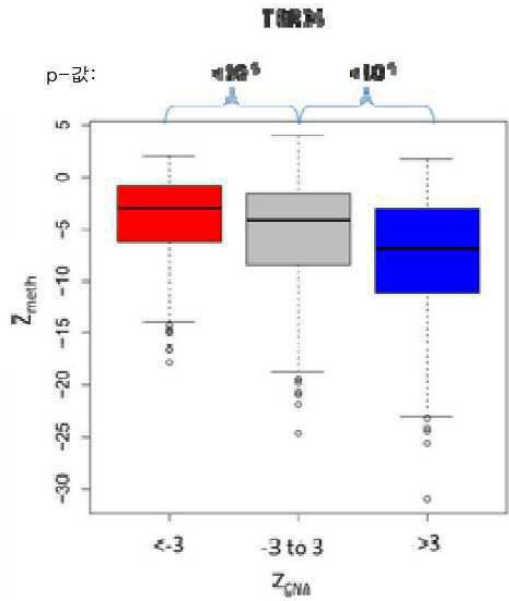
도면50a



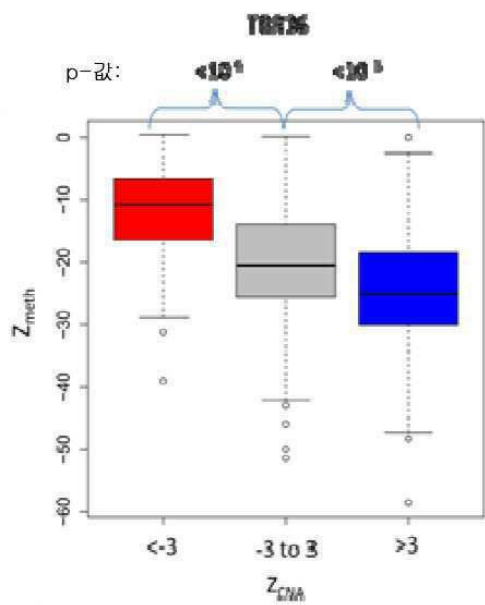
도면50b



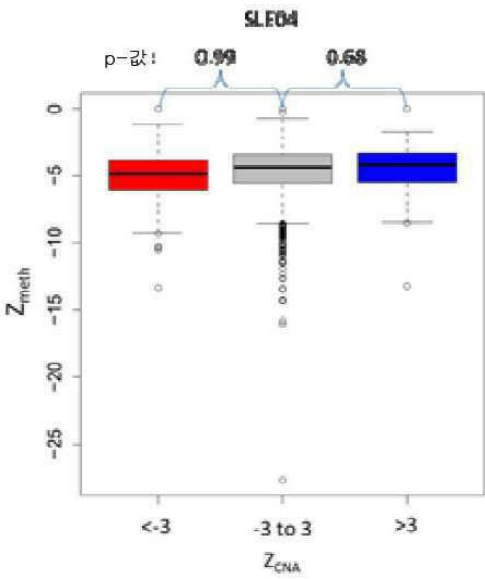
도면51a



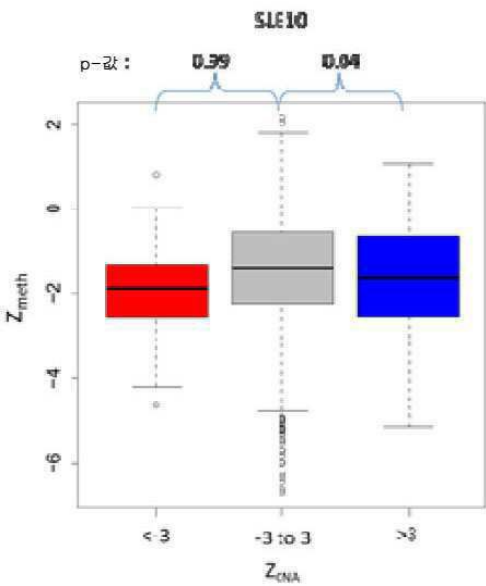
도면51b



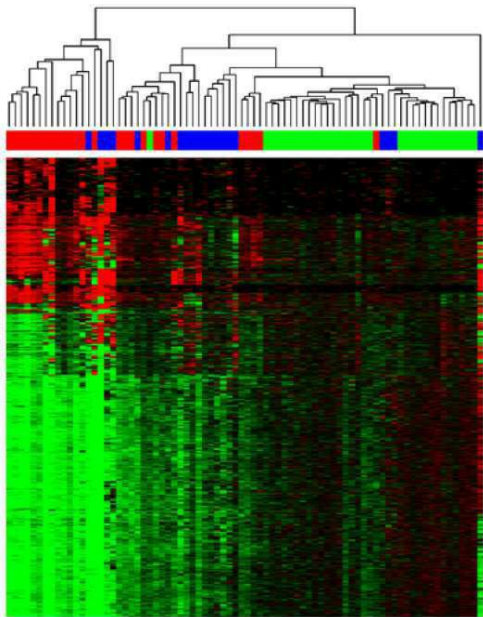
도면51c



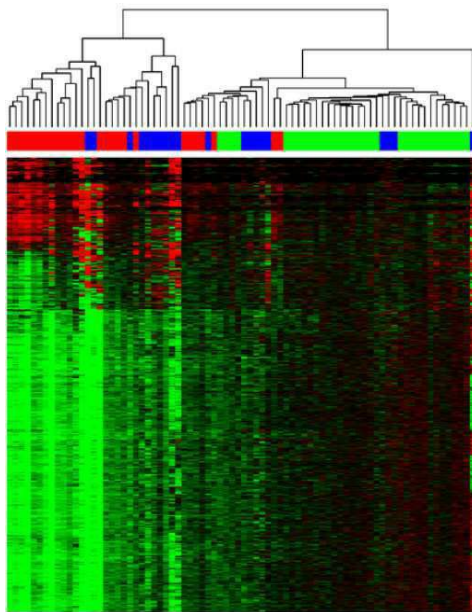
도면51d



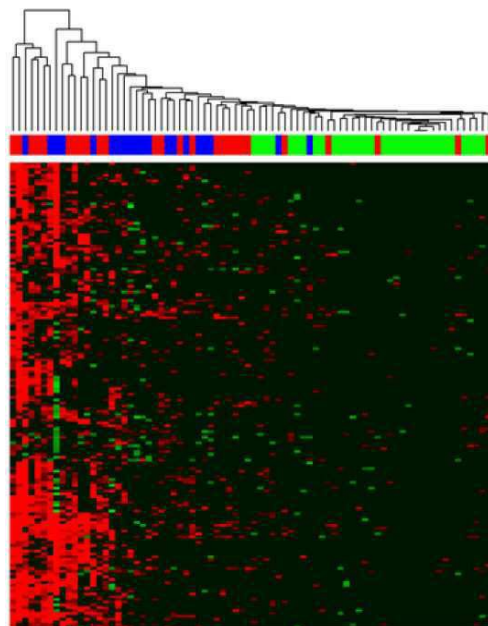
도면52a



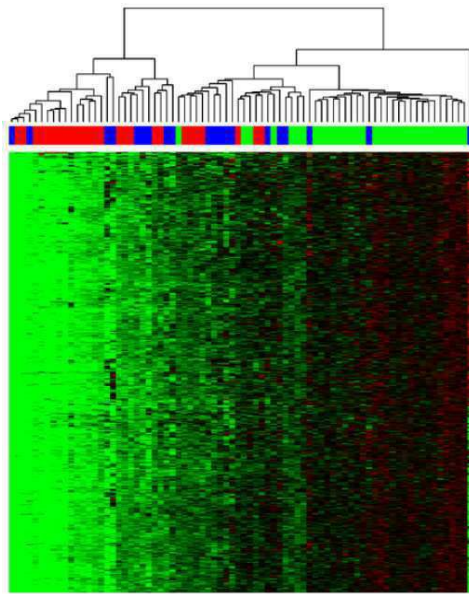
도면52b



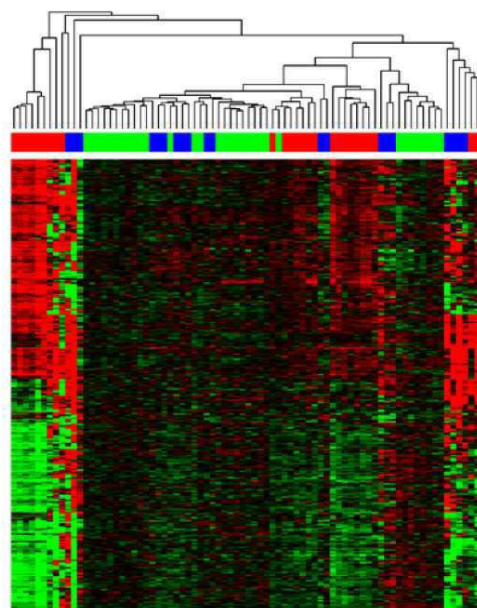
도면53a



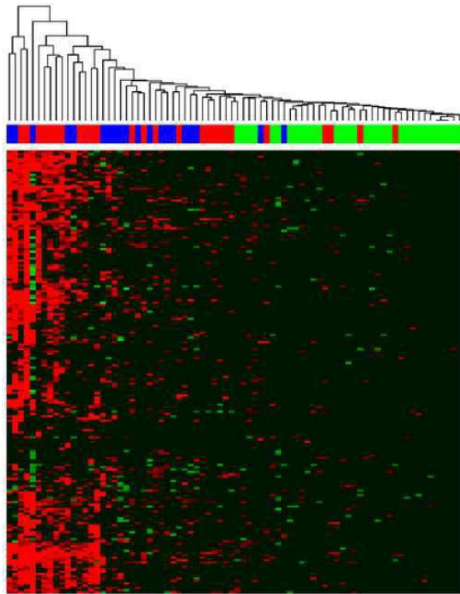
도면53b



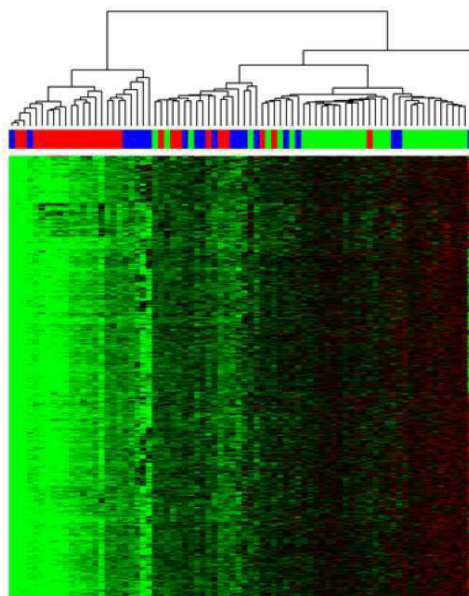
도면54a



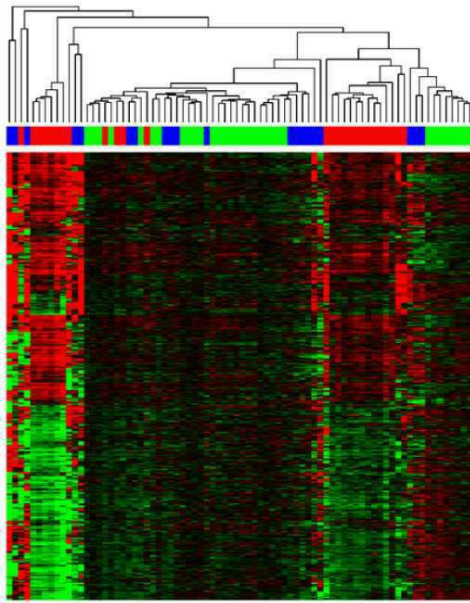
도면54b



도면55a



도면55b



도면56

