

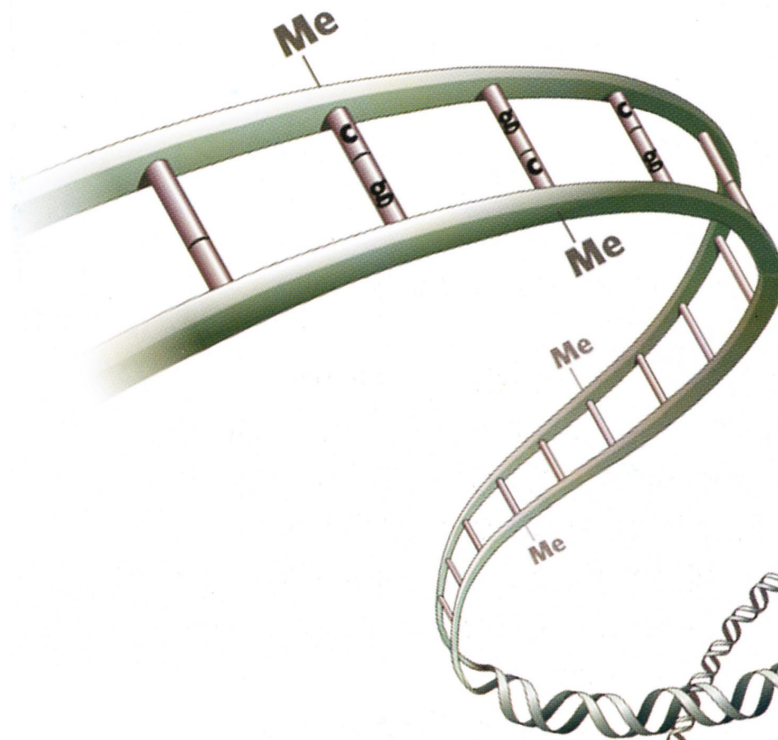
illumina Webinar 2012.7.12

エピゲノム解析シリーズ

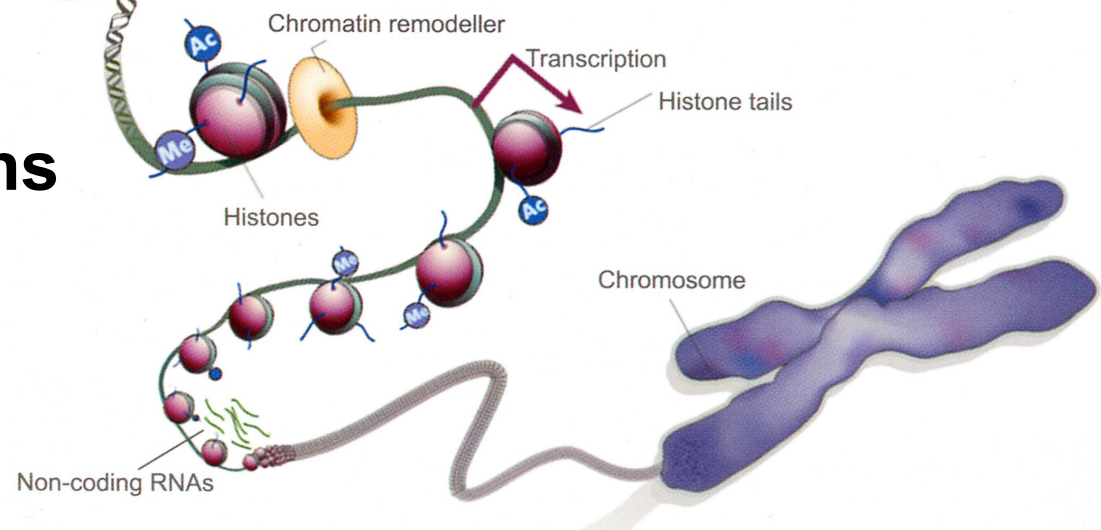
「DNAメチル化・基礎からの網羅的解析」

国立がん研究センター研究所 エピゲノム解析分野
山下 聡

Epigenetics



1. **DNA methylation**
2. **Histone modifications**
3. **Chromatin structure**
4. **Imprinting**
5. **Non-coding RNAs**

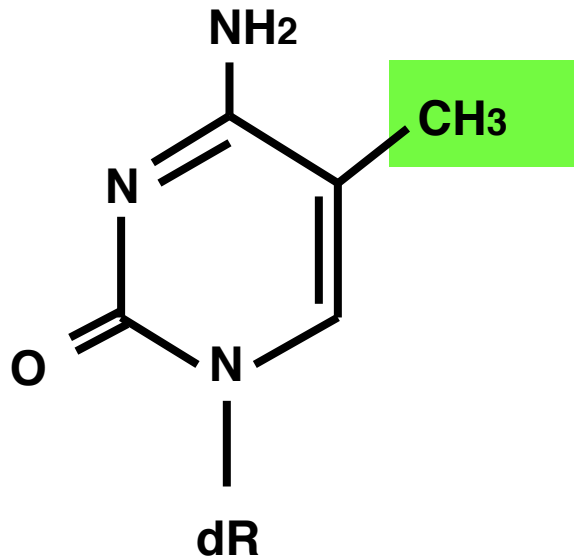


Topics

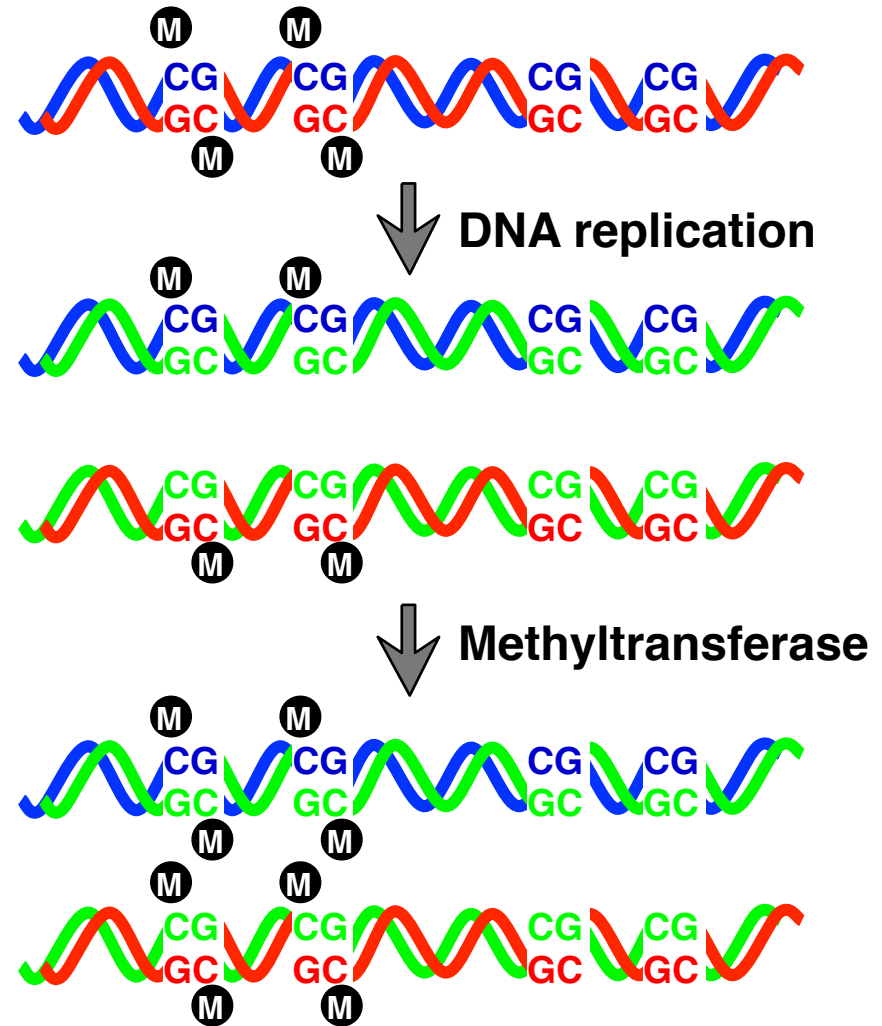
1. DNAメチル化解析の意義
2. 網羅的DNAメチル化解析の概要
3. 網羅的メチル化解析例とその簡単なデータ処理
4. 領域個別のメチル化解析

酵素のDNAメチル化

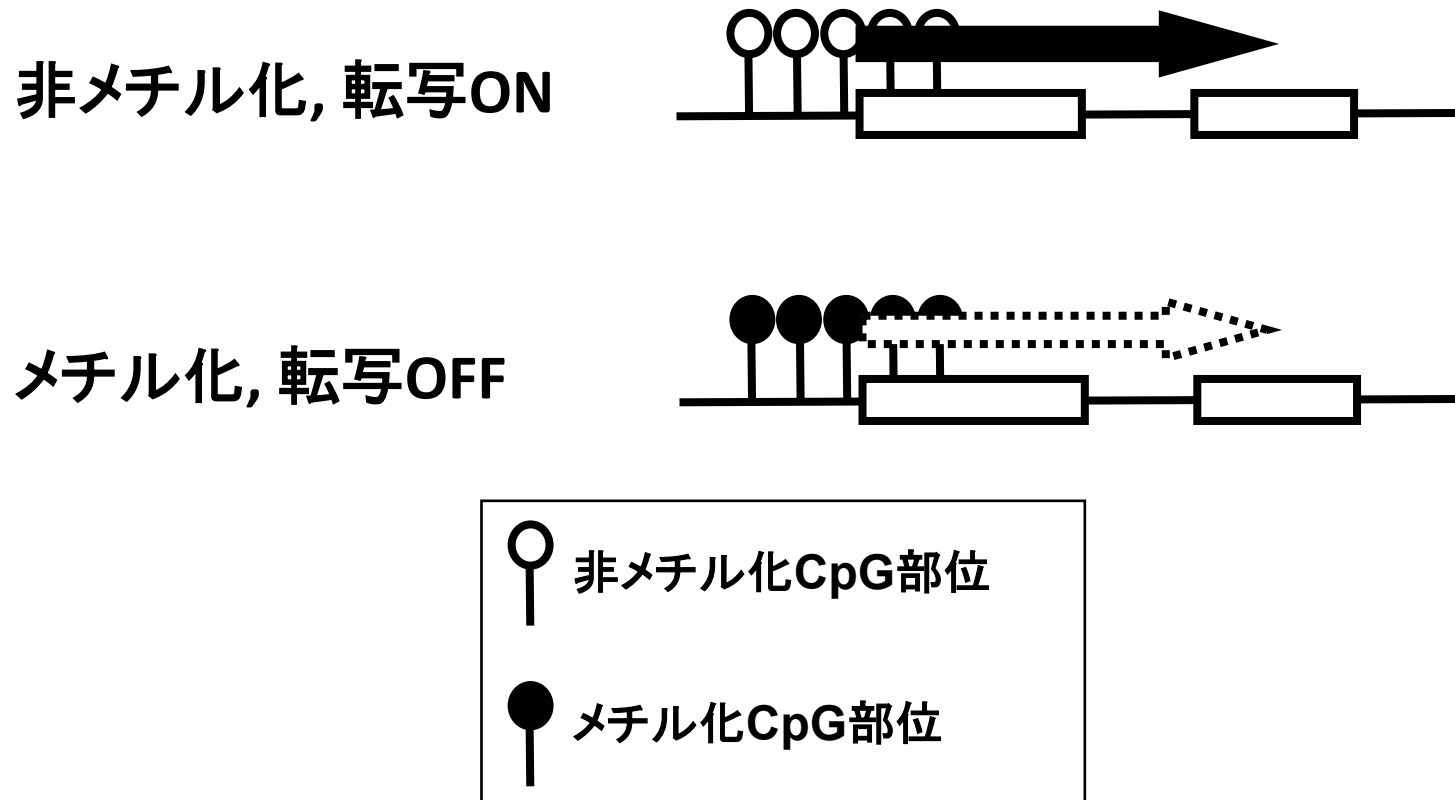
5-Methylcytosine



Inheritance upon replication



遺伝子プロモーター領域CpGアイランドのメチル化は 遺伝子サイレンシングをもたらす



癌抑制遺伝子の不活化機構として重要

なぜDNAメチル化解析を行うのか？

- 後天的な（環境による）変化
- 遺伝子発現（表現型）が異なっている原因かもしれない
- 病態マーカーとして有用かもしれない

DNAメチル化異常の特徴

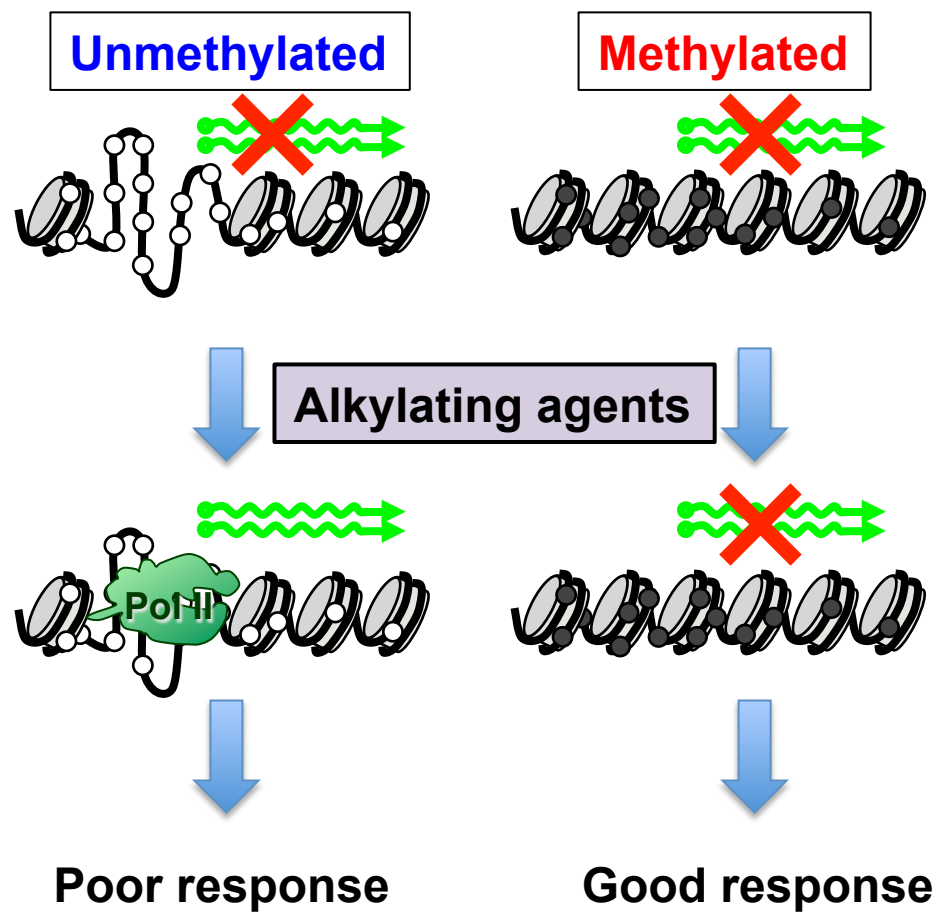
トランスクリプトーム・プロテオームとの違い

1. プロモーター領域のメチル化は、その時たまたま発現していないことではなく、どんな時でも「発現し得ない」ことを示す。
2. 一部細胞での高発現により、全体像が影響されることがない。
3. 細胞の状態や環境により簡単には変わらない。
4. RNAや蛋白質と異なり、DNAは化学的に安定である。サンプル入手で有利。

ゲノム異常との違い

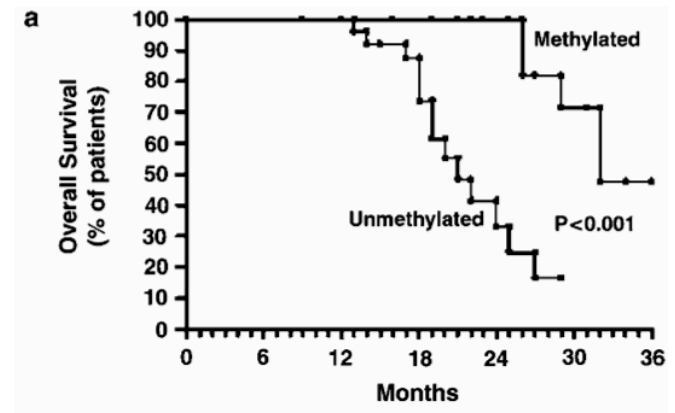
1. 非がん部に非常にたくさん蓄積しうる。
2. 異常誘発には特定の慢性炎症が重要である。
3. 誘発要因に応じて、特徴的な遺伝子に誘発される。
4. がん部での異常が非常に多く、その分、随伴変化も多い。
5. 脱メチル化剤で解除できうる。

アルキル化剤反応性予測マーカー: *MGMT* のメチル化



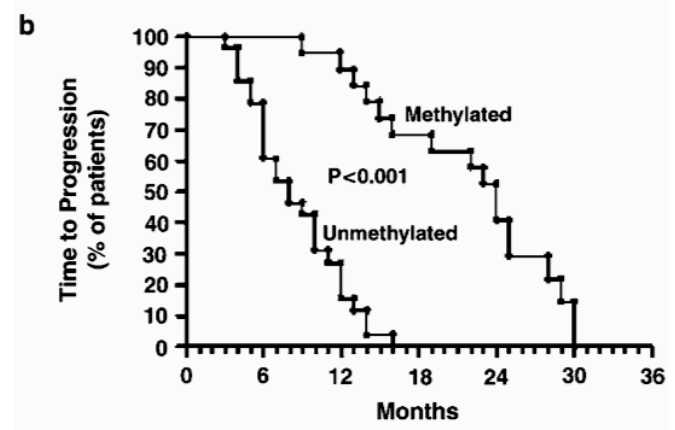
(From Esteller and Herman, *Oncogene*, 2004)

40 glioma cases



NO. OF SUBJECTS

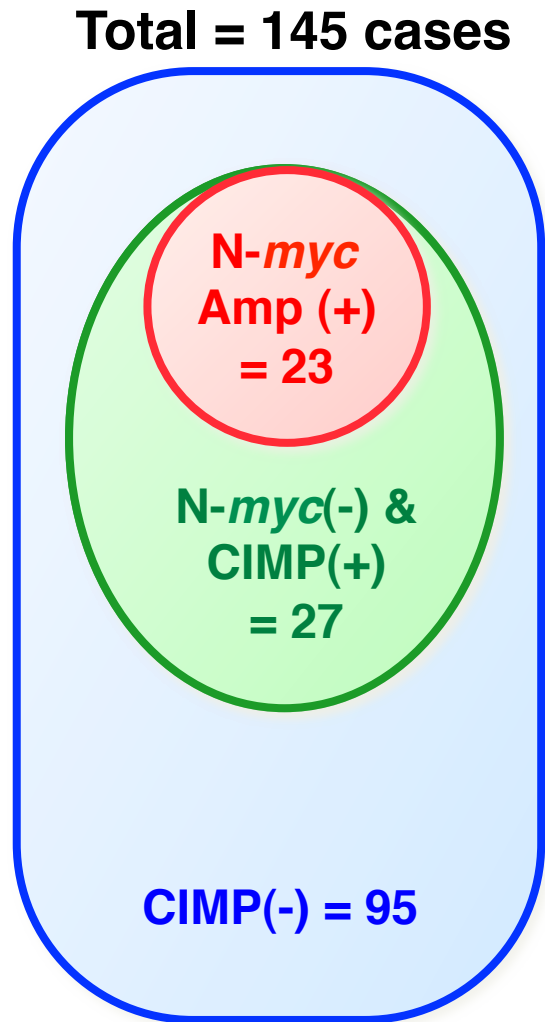
Methylated	19	19	19	18	13
Unmethylated	28	28	26	19	5



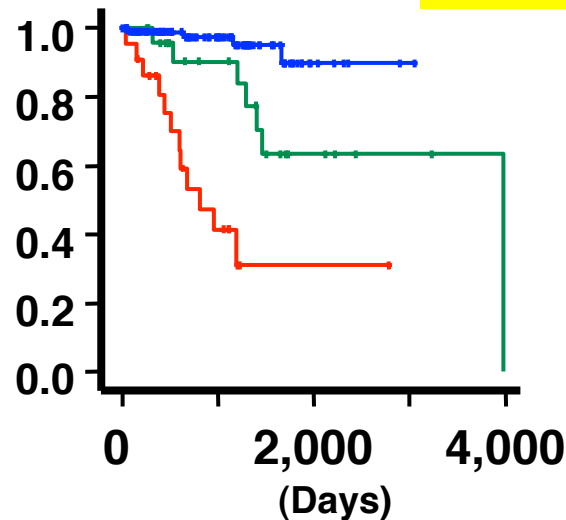
NO. OF SUBJECTS

Methylated	19	19	18	14	9	3
Unmethylated	28	22	7			

DNAメチル化(CIMP)の神経芽細胞腫予後マーカーとしての有用性

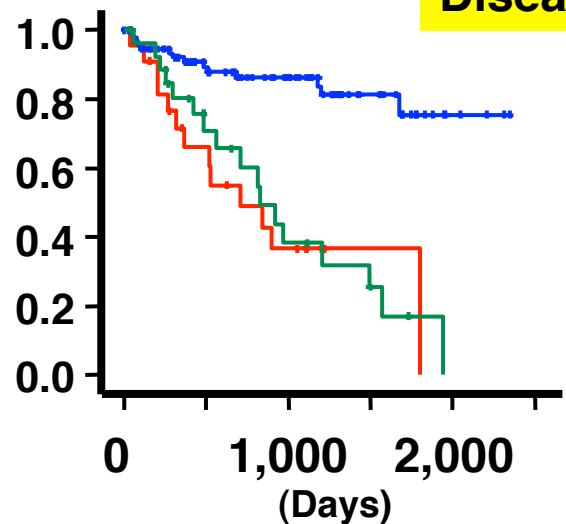


Overall survival



	HR	P
CIMP	9.5 (3.2-28.1)	<0.0001
N- <i>myc</i>	12 (4.9-29)	<0.0001
B to A	4.5 (1.3-16)	0.020
C to B	4.8 (1.7-14)	0.003
C to A	22 (6.8-69)	<0.0001

Disease-free survival



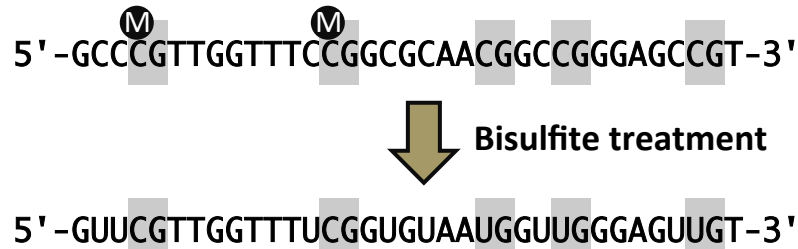
	HR	P
CIMP	5.4 (2.9-10)	<0.0001
N- <i>myc</i>	3.1 (1.6-6.0)	0.0007
B to A	5.2 (2.6-11)	<0.0001
C to B	1.1 (0.53-2.3)	0.821
C to A	5.7 (2.6-12)	<0.0001

Topics

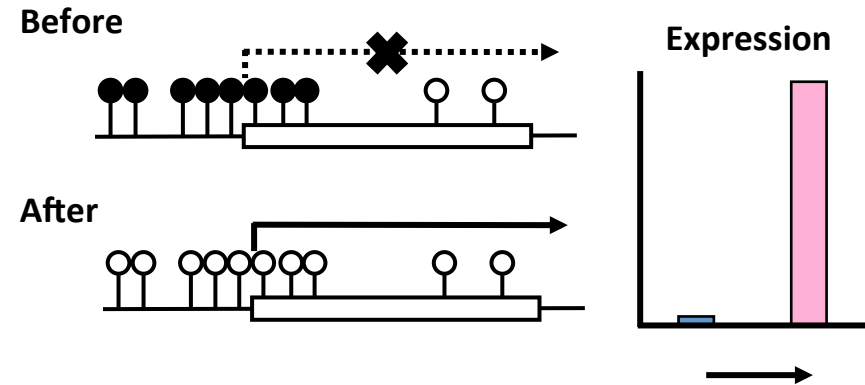
1. DNAメチル化解析の意義
2. 網羅的DNAメチル化解析の概要
3. 網羅的メチル化解析例とその簡単なデータ処理
4. 領域個別のメチル化解析

DNAメチル化解析の原理

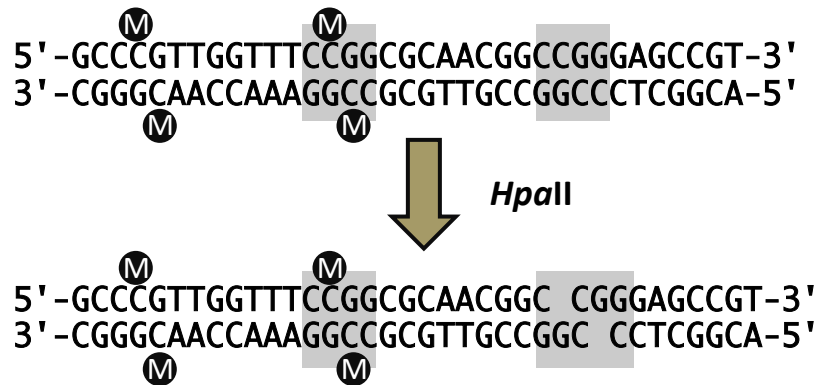
Bisulfite treatment (最重要)



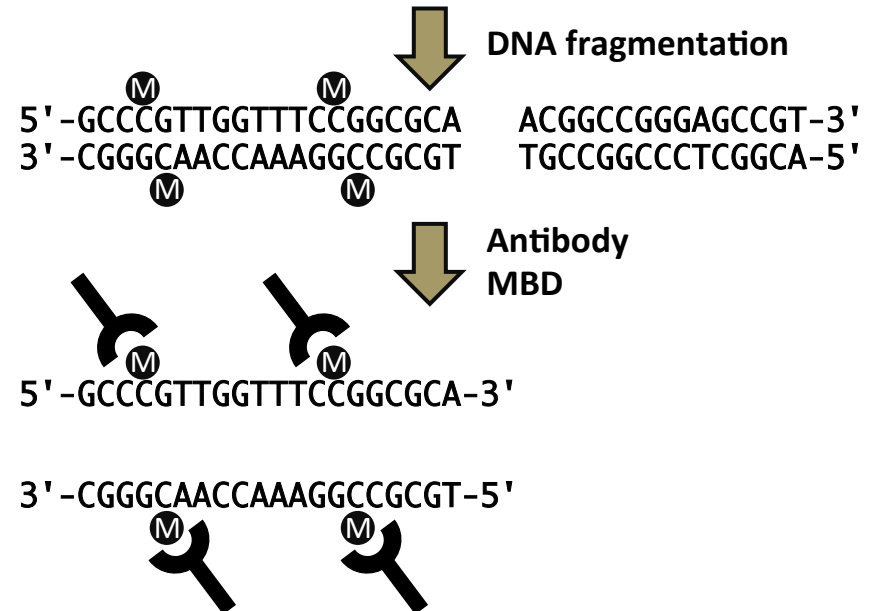
Demethylating agent



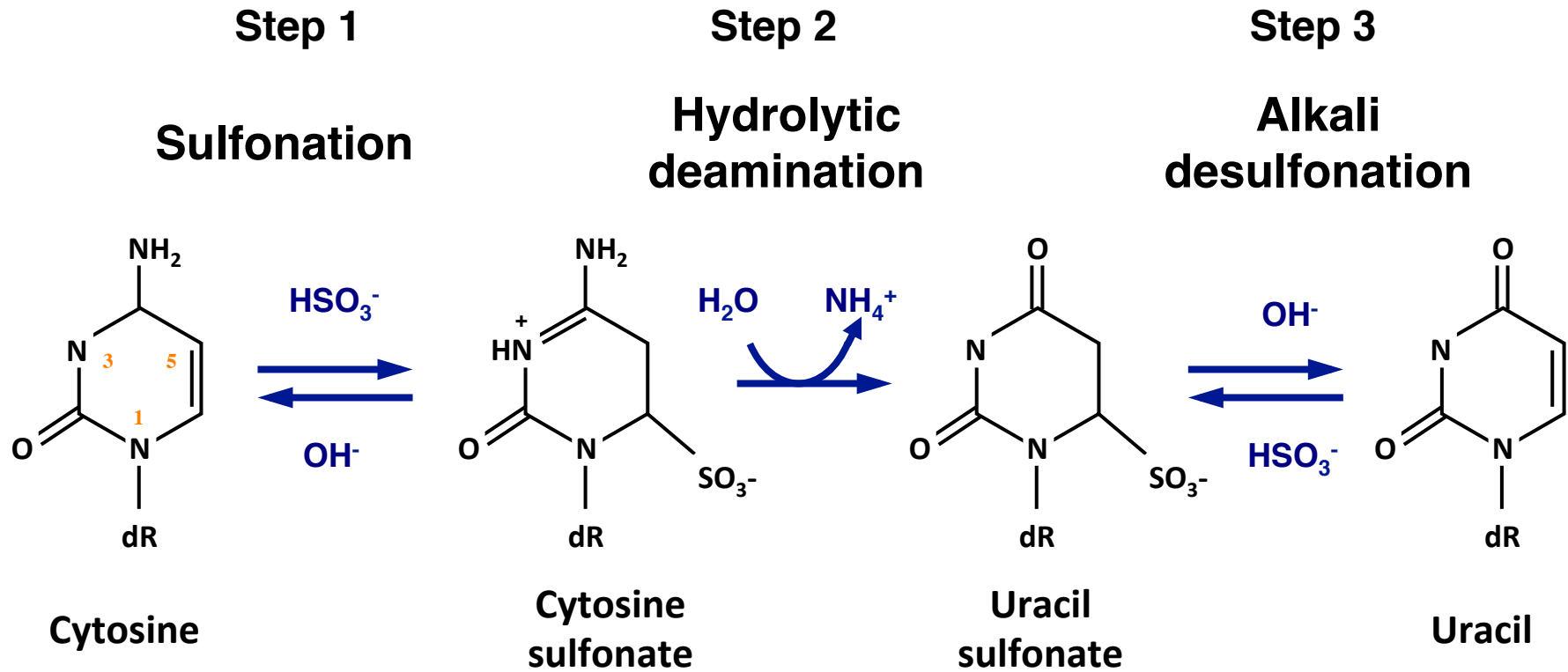
Methylation-sensitive restriction enzyme



Affinity



Bisulfite 反応によるシトシンの変換



Bisulfite:
NaHSO₃

Bisulfite conversion		PCR	
M C	→	M C	→ C
C	→	U	→ T

網羅的DNAメチル化解析の選択肢

	MethylC-seq	MeDIP-seq, MethylCap-seq, MBD-seq	Reduced representation bisulfite sequencing	MeDIP-CGI microarray	Infinium -27K	Infinium -450K
Genomic DNA	5 µg	0.3-5 µg	0.05 µg	5 µg	0.5-1 µg	
Assay	Bisulfite conversion	Capture with antibody or MeCP2 or MBD	Bisulfite conversion	Capture with antibody	Bisulfite conversion	
Readout	Sequence			Array		
Resolution	1 bp	100-1,000 bp	1 bp	100-1,000 bp	1 bp	
Theoretical coverage	~100%		~10% (CpG)	~70% (CGI)	~0.1% (CpG)	~2% (CpG)
Cost	~\$20 K	~\$2 K		~\$0.5 K		
Species	All			Customization available	Human	
Ease of use	Hard	Medium		Easy		
Conclusion	Good standard	Good all-rounder	Good for CGIs		Good for promoters	Good all-rounder

(Beck et al. 2010を元にYamashitaが大幅改訂)

Illumina HumanMethylation450 BeadChip

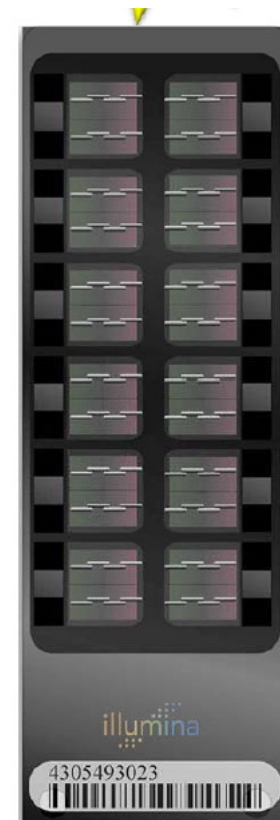
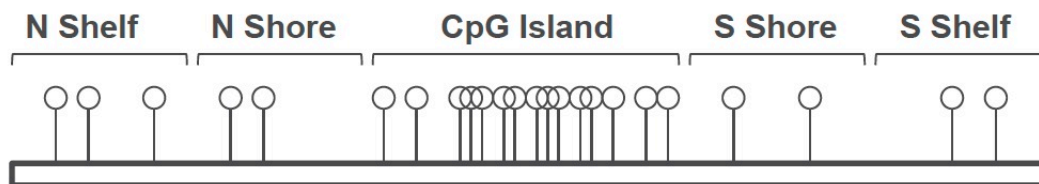
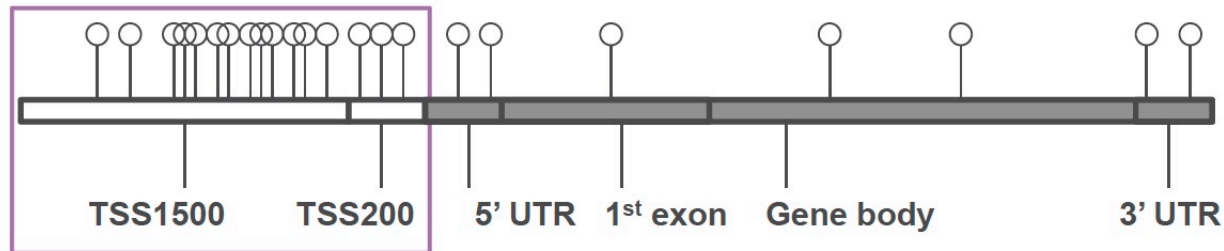
450K (450,000) probes

95% CpG islands

12 samples / slide

41,300 yen / sample

human



[From Illumina]

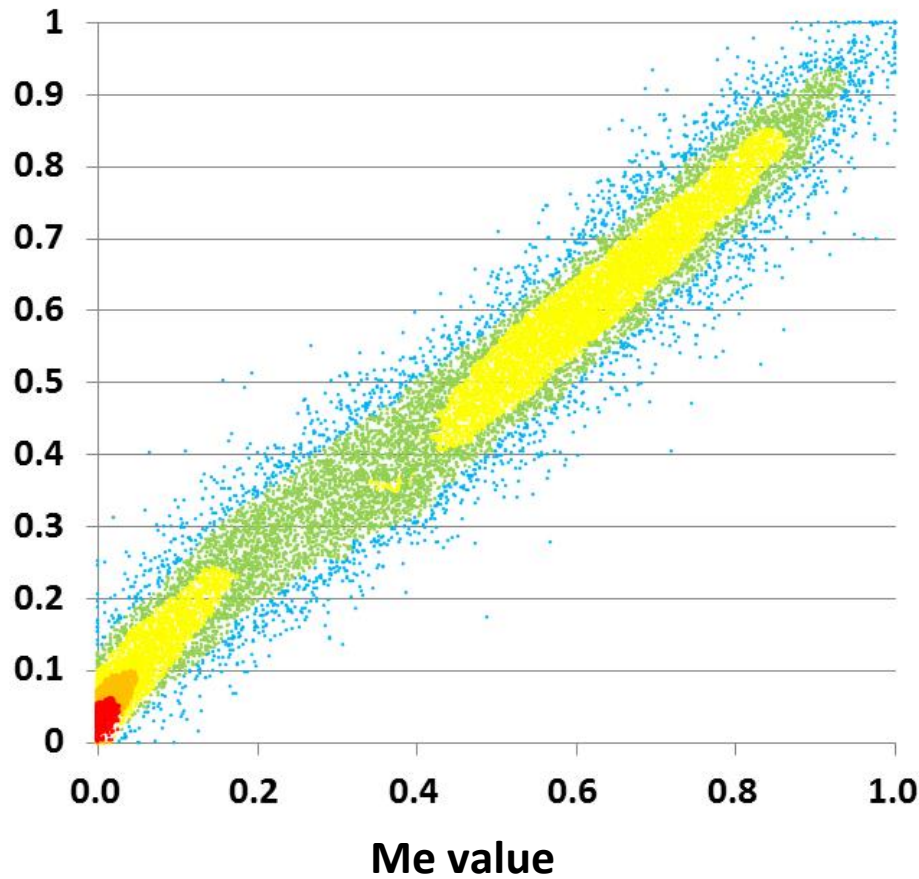
HumanMethylation450 BeadChipの特徴

1. 網羅性 - プロモーター領域以外、CpG island以外もカバー
2. 定量性, 再現性 - 10%のメチル化の違いも検出し得ると期待できる
3. 簡便性 - (機器さえ揃っていれば)プロトコールが整備されていて扱いやすい。得られるデータがシンプル。
4. その他 - FFPEサンプルにも対応、アノテーションが充実 (転写開始点からの距離以外)

ゲノム網羅的DNAメチル化解析の再現性

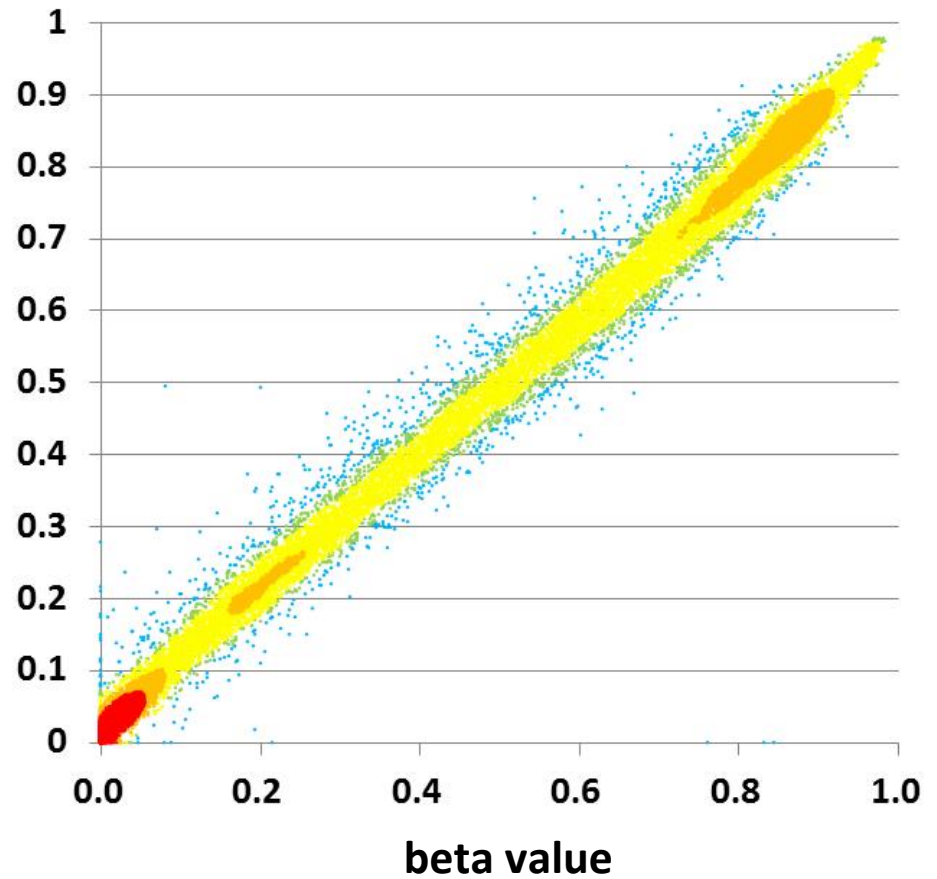
MeDIP-CGI microarray

$R^2 = 0.962$



HumanMethylation450

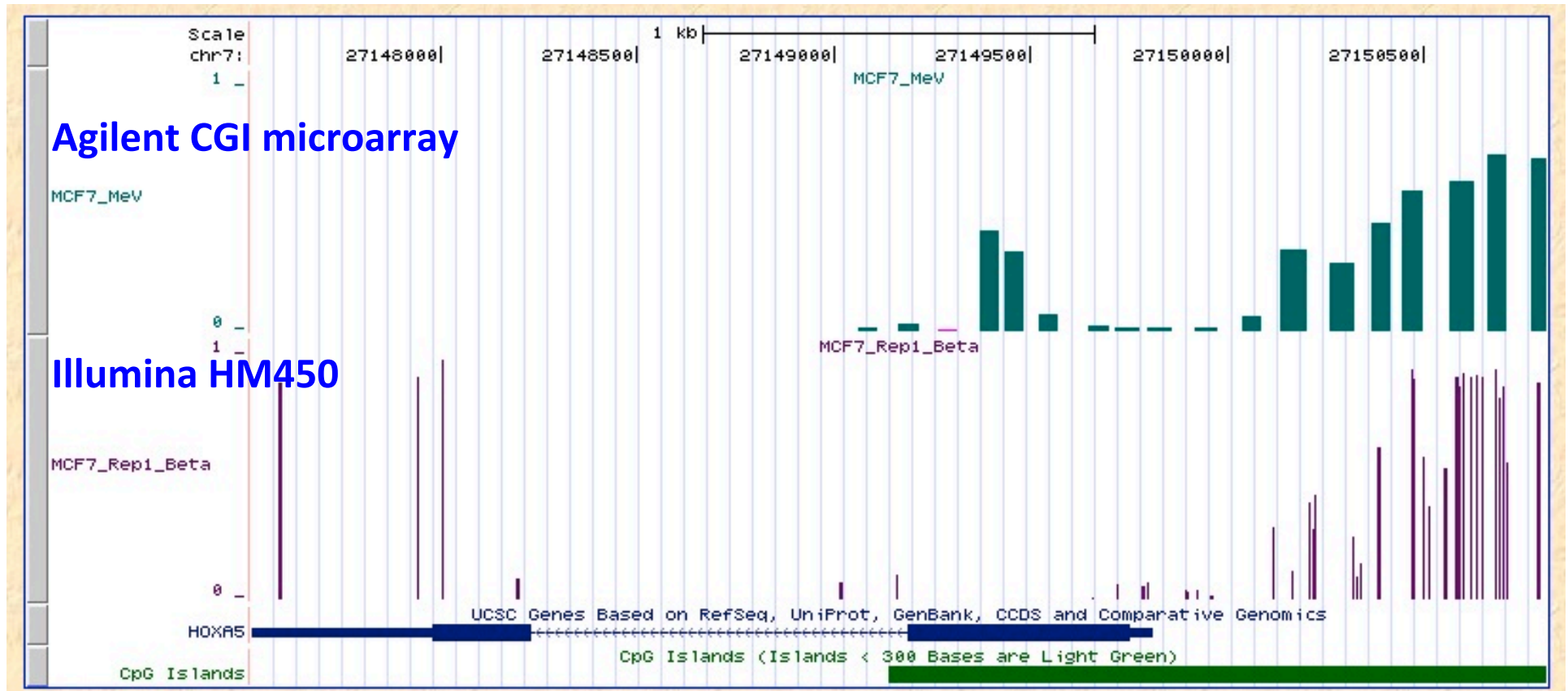
$R^2 = 0.993$



[Yamashita *et al.* unpublished]

異なるプラットフォームでは異なる領域が計測可能

HOXA5 promoter



UCSC Genome Browser

Topics

1. DNAメチル化解析の意義
2. 網羅的DNAメチル化解析の概要
3. 網羅的メチル化解析例とその簡単なデータ処理
4. 領域個別のメチル化解析

まずは図で様子を見る

- Illumina GenomeStudioもかなりよいが
- UCSC Genome Browser (<http://genome.ucsc.edu/>)で他のデータと比べるのがおすすめ
- Importに必要なBedファイルはExcelで作製可能

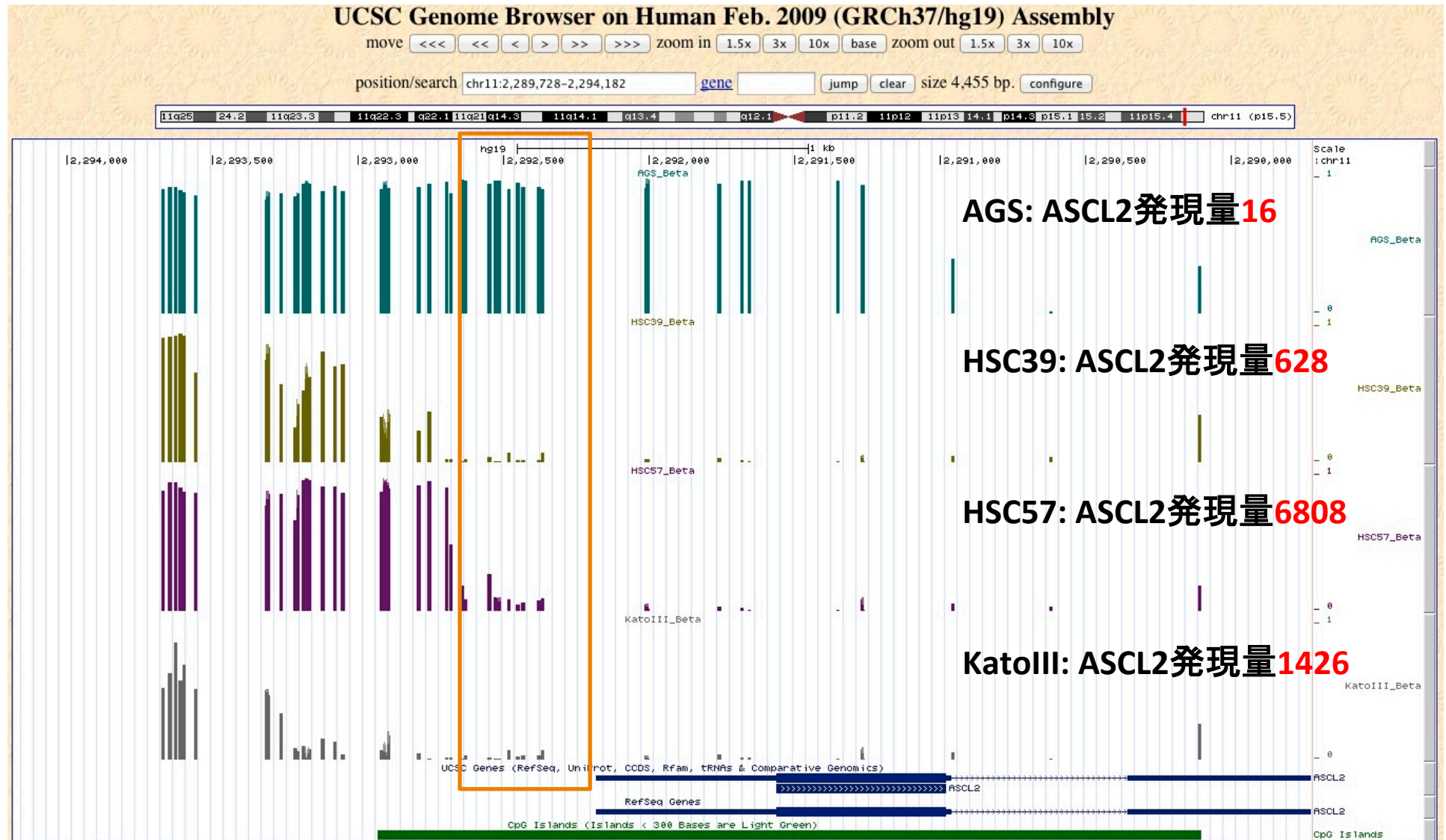
	A	B	C	D	E	F	G	H	I
1	track type=bedGraph name="AGS_Beta" description="AGS_Beta" color=102,0,102								
2	chr16	53468112	53468121	0.3608217					
3	chr3	37459206	37459215	0.9382988					
4	chr3	171916037	171916046	0.8856235					
5	chr1	91194674	91194683	0.8744467					
6	chr8	42263294	42263303	0.8944484					
7	chr14	69341139	69341148	0.694135					
8	chr16	28890100	28890109	0.912966					
9	chr8	41167802	41167811	0.7912668					
10	chr1	230560793	230560802	0.8947818					
11	chr15	23034447	23034456	0.00394956					
12	chr9	139997924	139997933	0.8280725					
13	chr19	54695678	54695687	0.3309445					
14	chr6	25282779	25282788	0.9378259					
15	chr3	128902377	128902386	0.0685074					
16	chr12	124086477	124086486	0.03259799					
17	chr2	23913414	23913423	0.9129239					

読み込みエラーを避けるために

- Textファイルにコンバートした時、1行目の書式を確認
- 時々あるbeta値(空欄)はゼロで埋める

染色体 位置 位置+10 Beta値

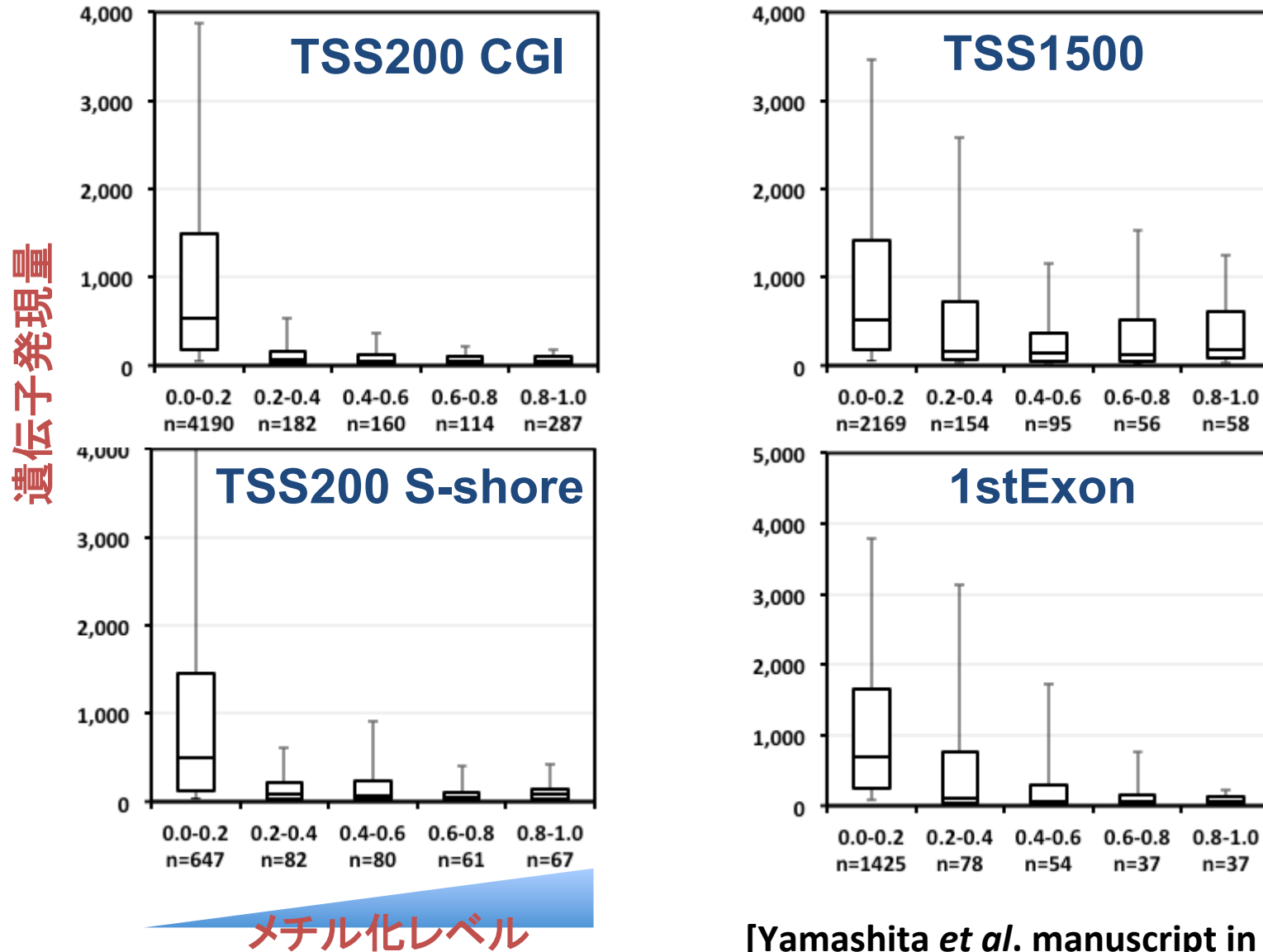
発現量とメチル化が関係のあるプロモーター領域は限られている



[Yamashita *et al.* manuscript in preparation]

CpGアイランドやプロモーターでなくともメチル化サイレンシングはあり得る

HSC39 cells



[Yamashita *et al.* manuscript in preparation]

網羅的DNAメチル化解析によるマーカー同定の例 (1)

- **食道扁平上皮癌(ESCC)**
- 手術前に転移の有無について情報を得るためのマーカー同定
- 転移有りと転移なし原発巣との比較(合計3組の比較)
- MeDIP + Agilent Human CpG island array (24万プローブ)
- メチル化の差が0.4 (Me value, 最小0-最大1) 以上のプローブを抽出

数百プローブ

- **3個以上のプローブで連続して差があるものに限定**

52領域

ExcelではIF関数で可能



個別領域のメチル化解析によるアレイ結果の検証

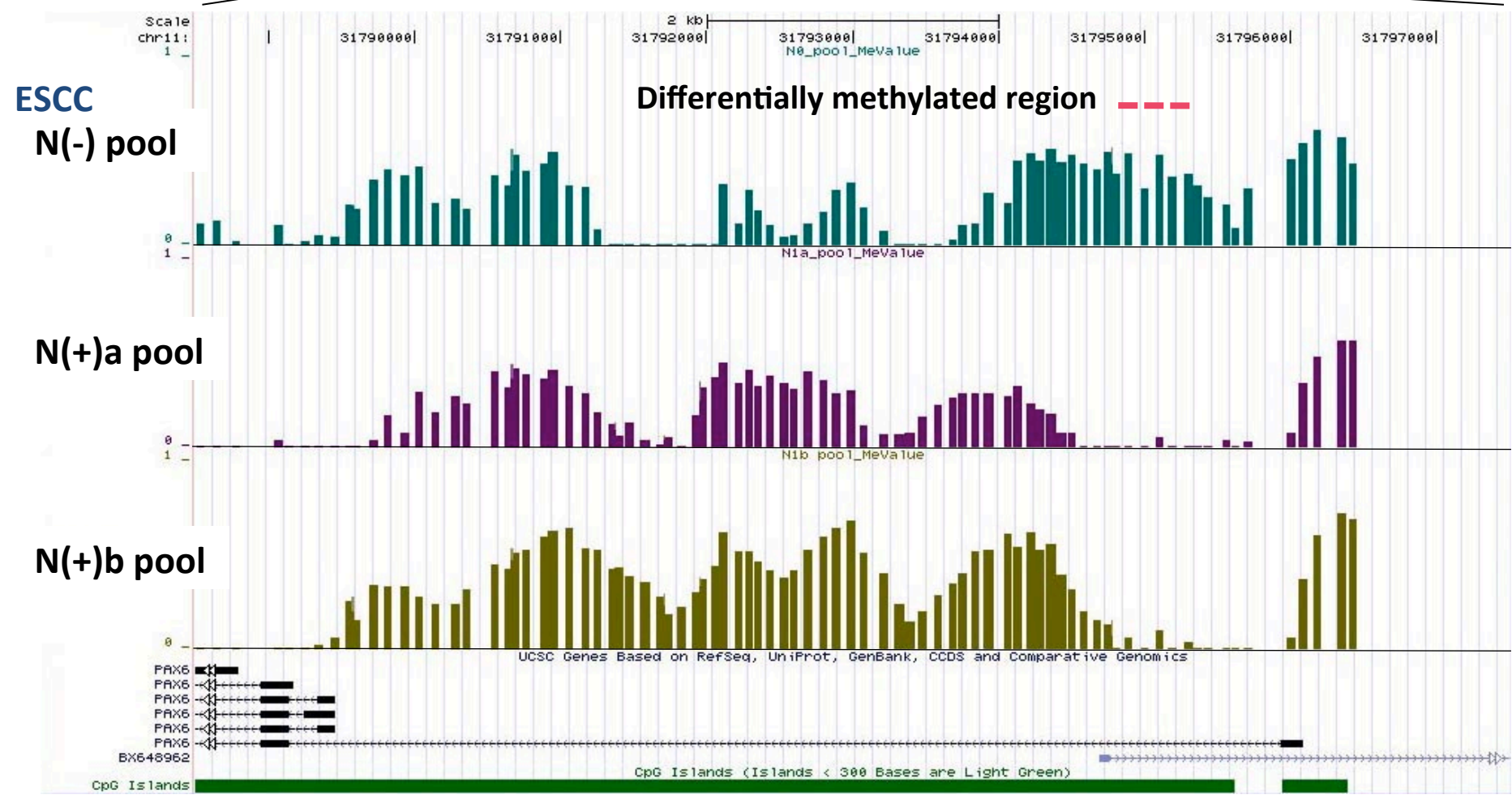
25領域



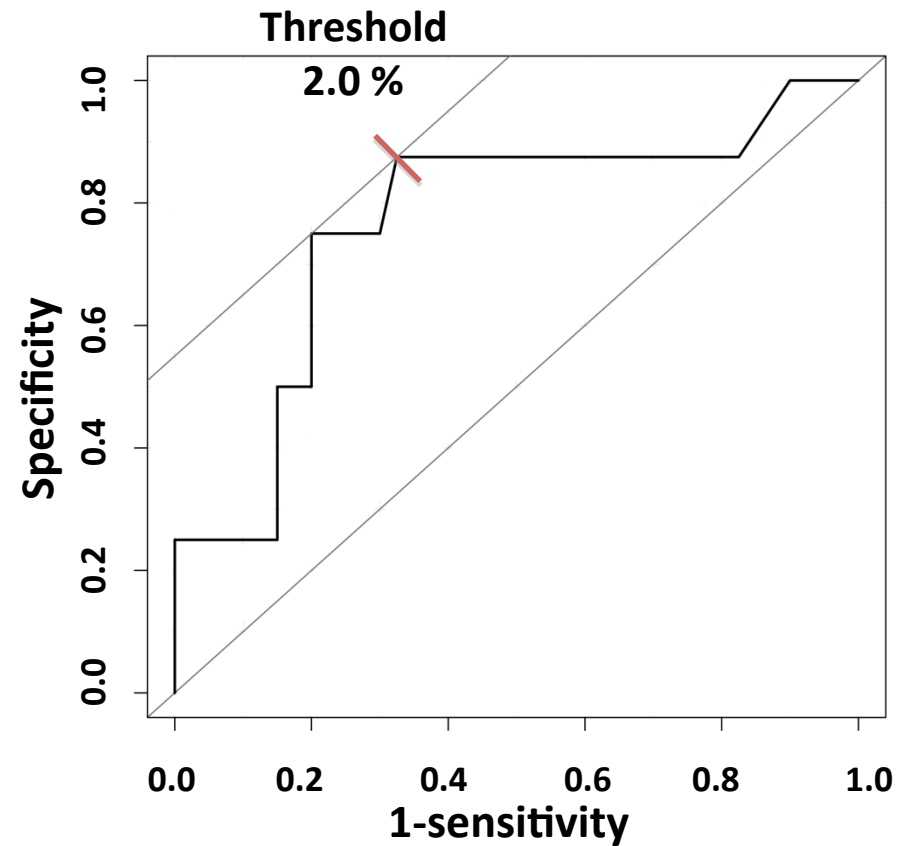
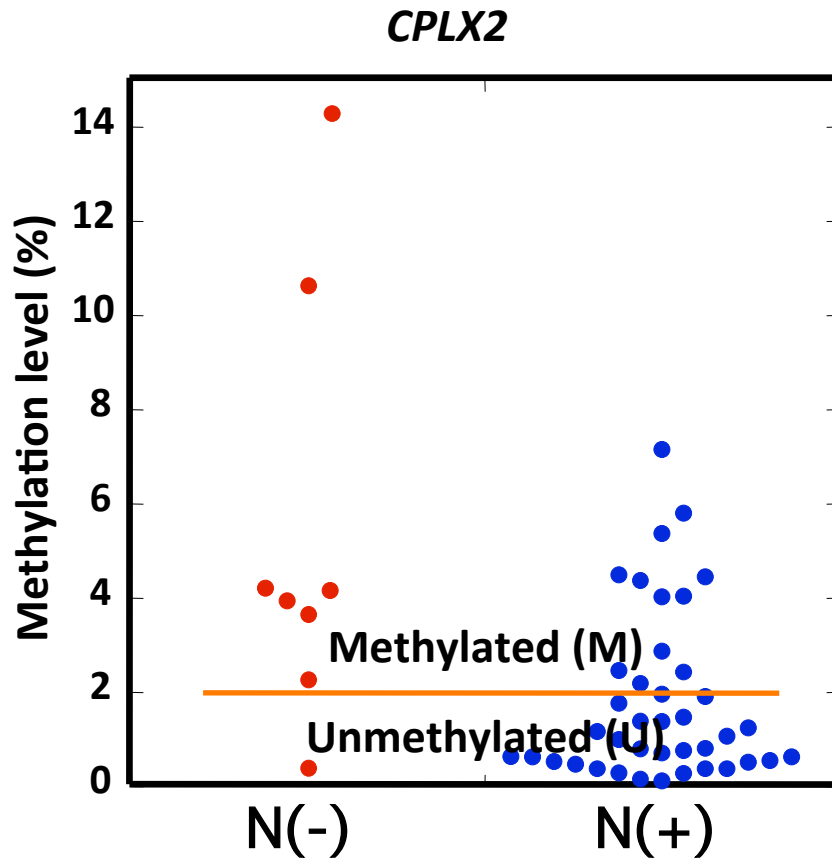
マーカーのscreeningとvalidationへ

Data of MeDIP-CGI microarray analysis in genomic regions around *PAX6* CGIs

Chr 11 (p13) 



Quantitative Methylation Analysis of Candidate CGIs in Screening Set and Construction of ROC Curve



	N(-)	N(+)
M	7	13
U	1	27

$p=0.0061$ (Fisher's exact test)

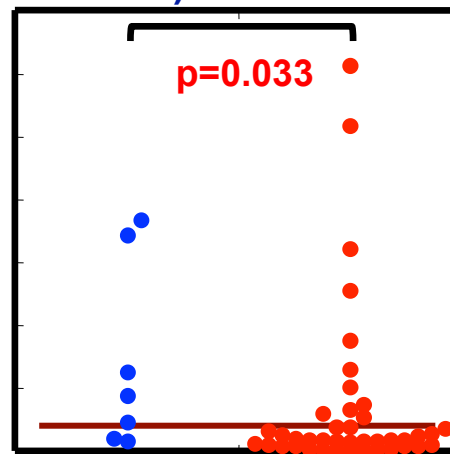
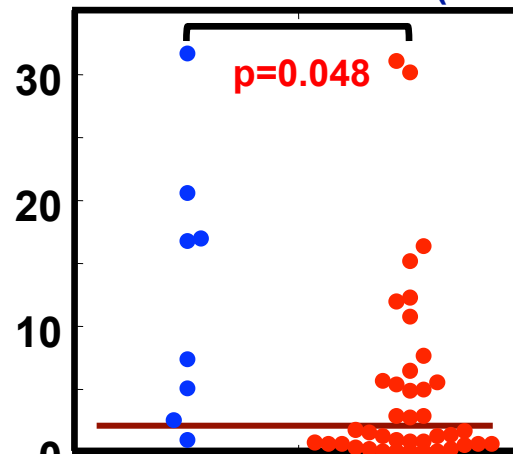
25 CGIs → 7 CGIs ($p < 0.05$)

食道がんのリンパ節転移予測マーカー

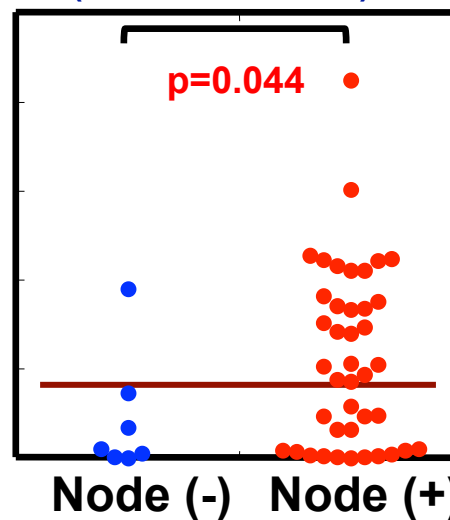
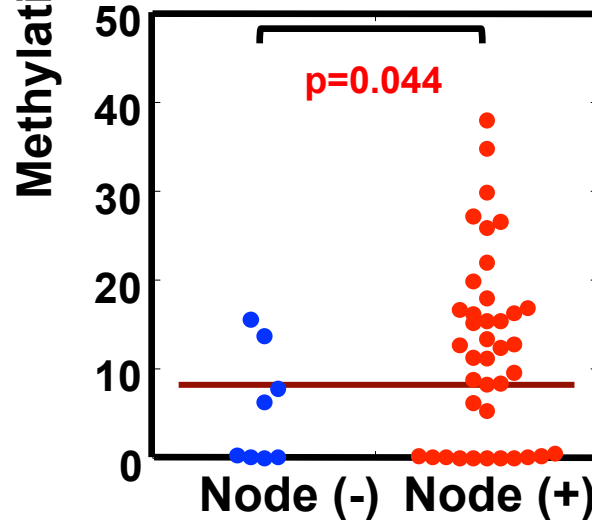
Screening set
(N=48)

Validation set
(N=48)

PAX6 (Threshold 2.0 %)



ENST00000363328 (Threshold 8.3%)



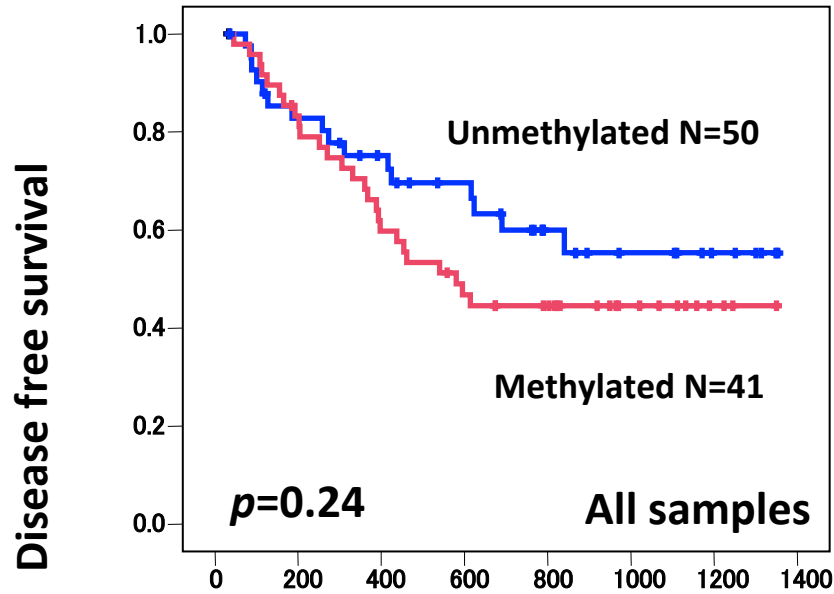
Combination marker

PAX6 M and U
ENST U or M
C. marker (-) (+)

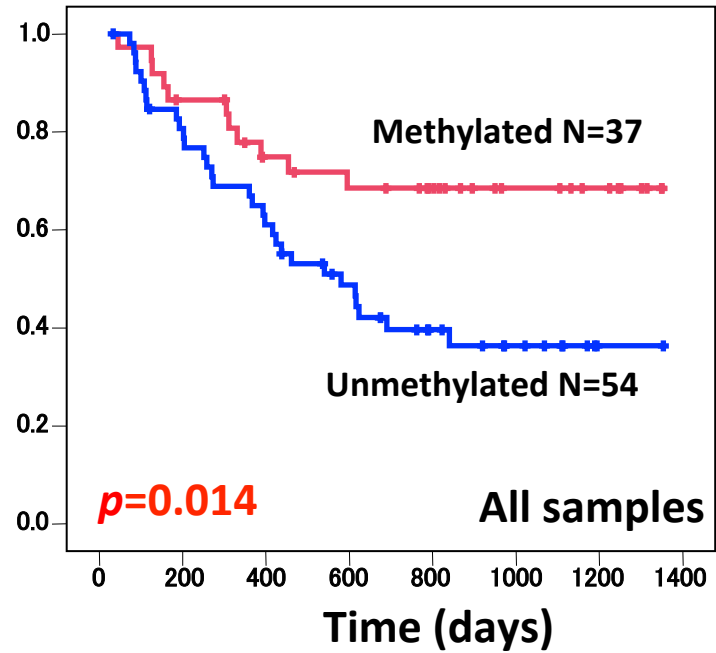
	C. marker	
	(-)	(+)
Node meta-stasis (-)	10	5
Node meta-stasis (+)	7	74

Sensitivity 91%
Specificity 67%
Accuracy 88%

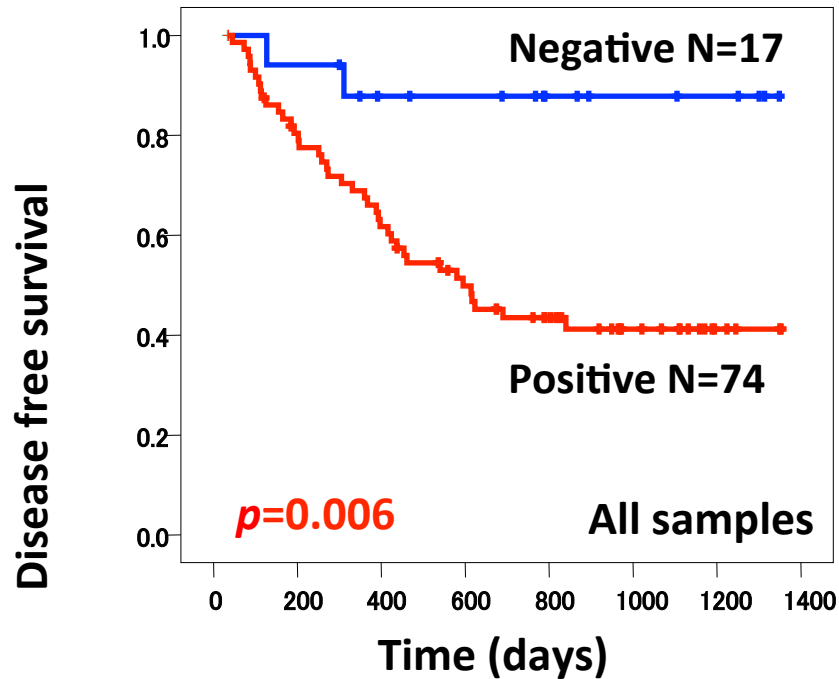
ENST00000363328



PAX6



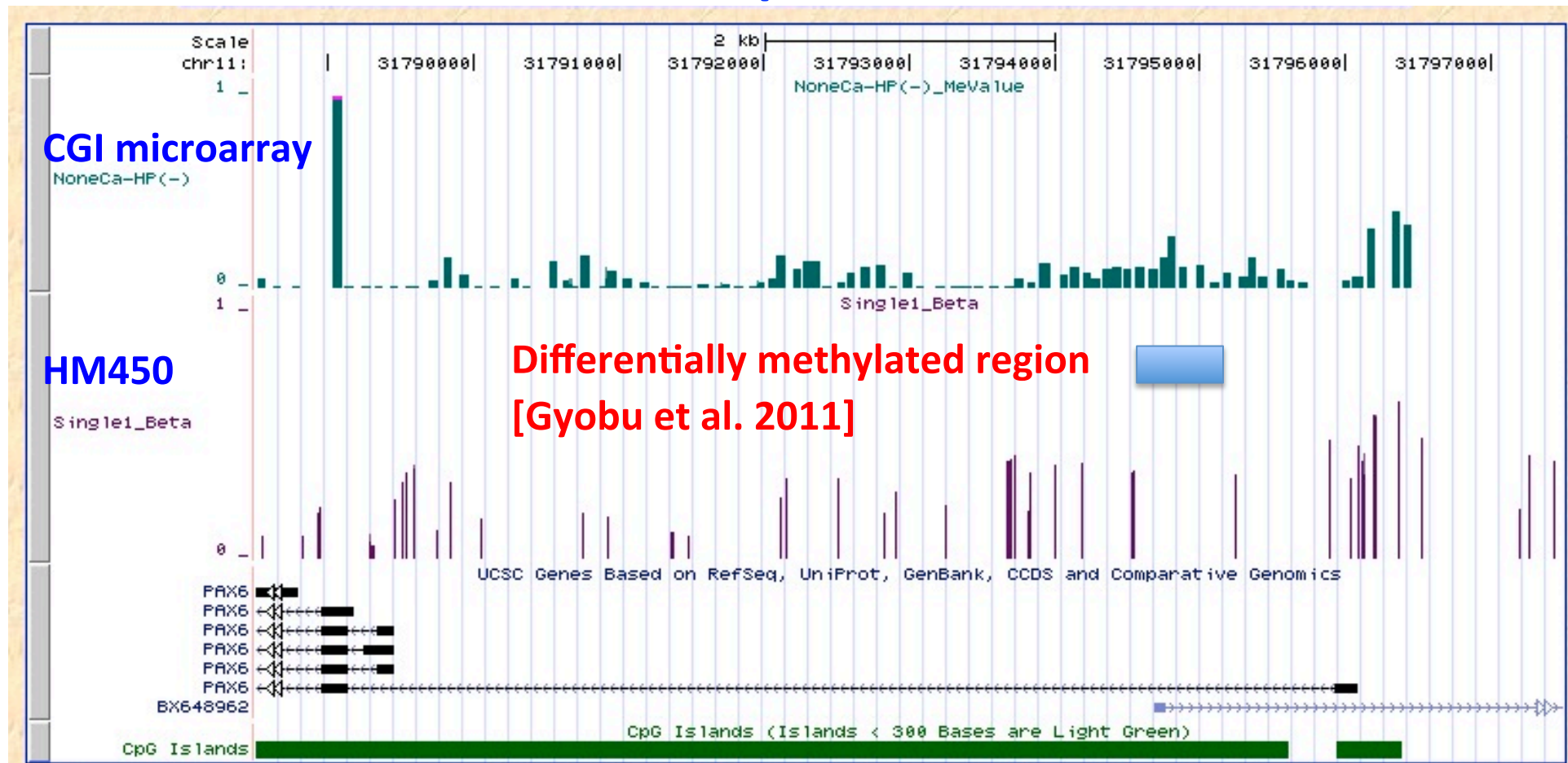
Combination markers



Analysis of Association with Methylation Status and Prognosis

実はこのマーカー領域はHM450ではプローブが存在しなかった

PAX6 promoter



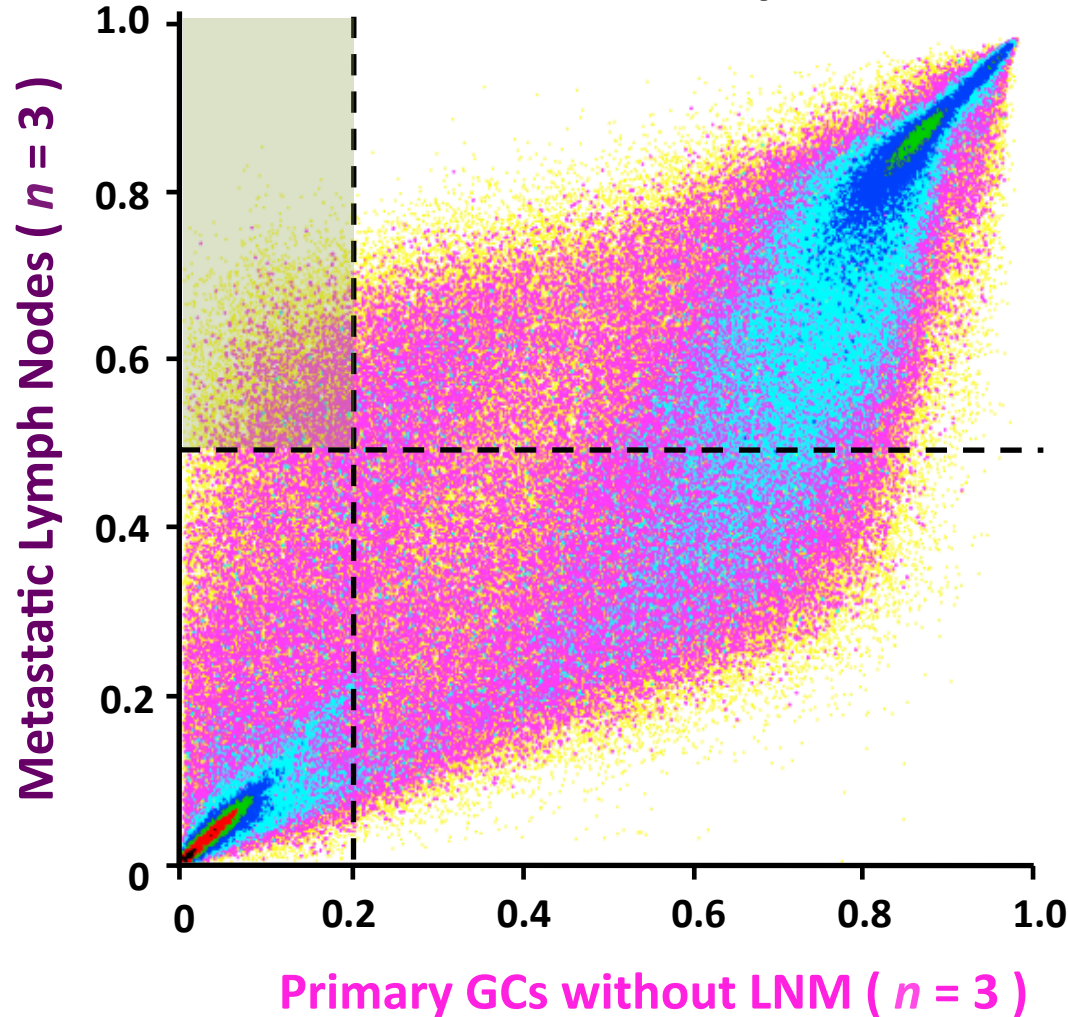
網羅的DNAメチル化解析によるマーカー同定の例 (2)

- **胃癌(GC)**
- 手術前に転移の有無について情報を得るためのマーカー同定
- 転移巣と転移なし原発巣との比較
- **Illumina HumanMethylation450**
- **差ではなく、転移なし原発巣でメチル化レベルが低く、転移巣でメチル化レベルが高い領域を抽出**

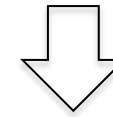
MLN vs Primary GC without LNM

Mean methylation
Level

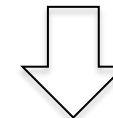
> 480,000 CpG sites



762 candidates
CpG sites



- Methylation level < 0.2 in all 3 GCs without LNM
- Methylation level > 0.5 in all 3 Metastatic lymph nodes

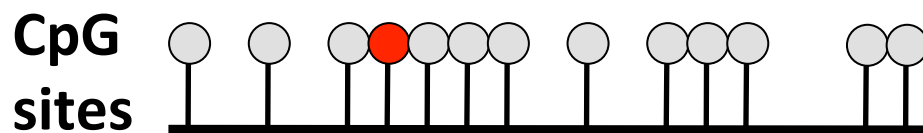


31 CpG sites

Mean methylation
Level

Methods for analysis of candidates

31 CpG sites



MSP primers

10 CpG sites

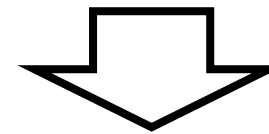


qMSP



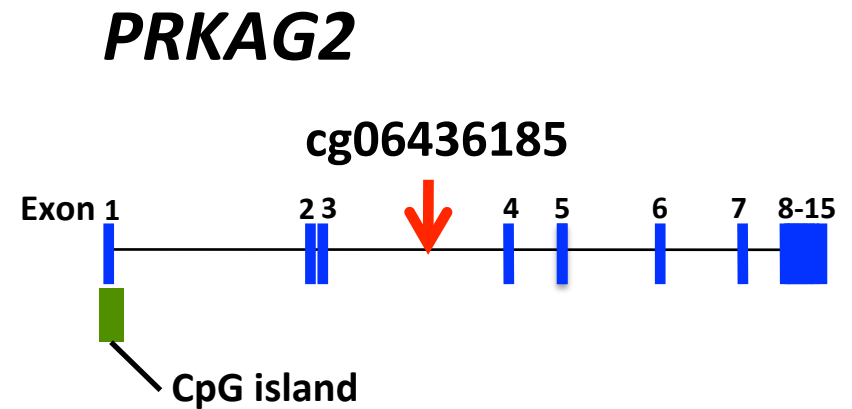
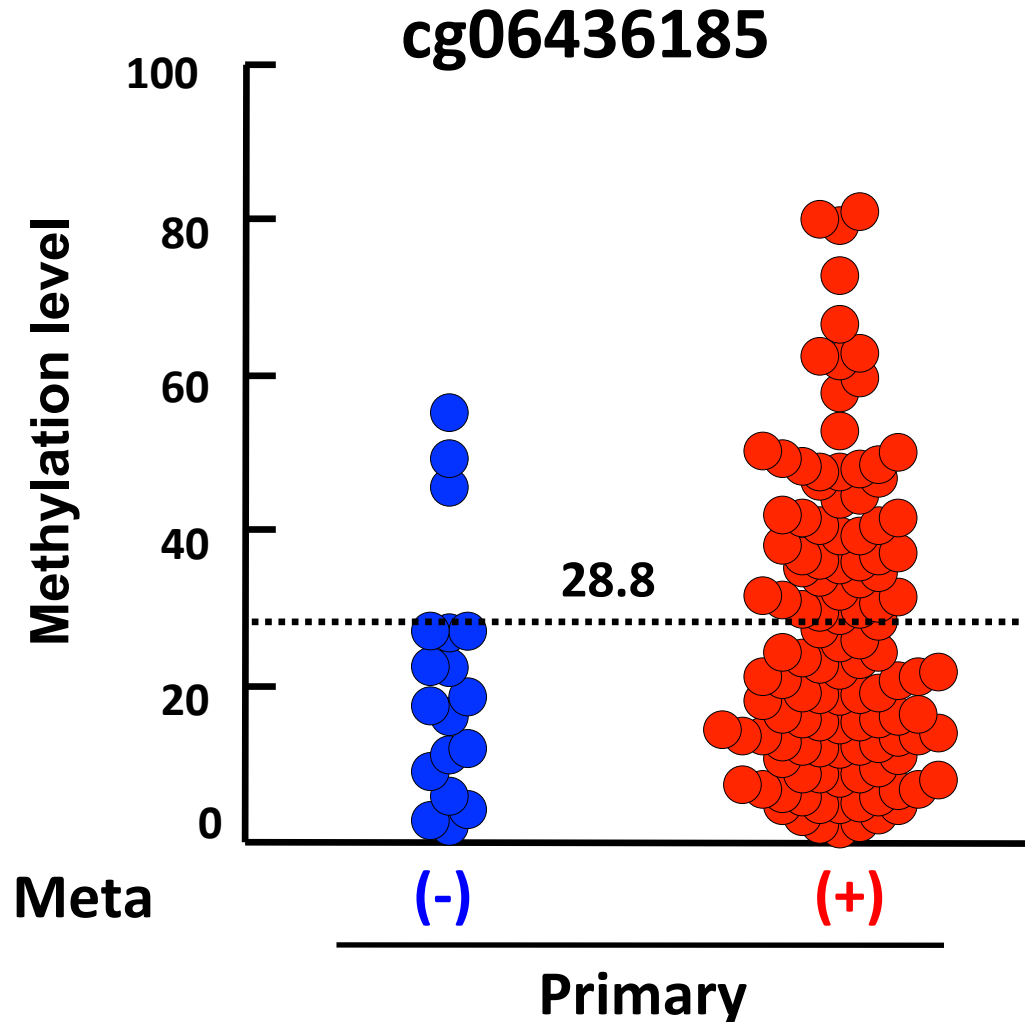
PCR primers

21 CpG sites



qPCR-MR

Validated one CpG site (cg06436185)



Sensitivity 43%

Specificity 85%

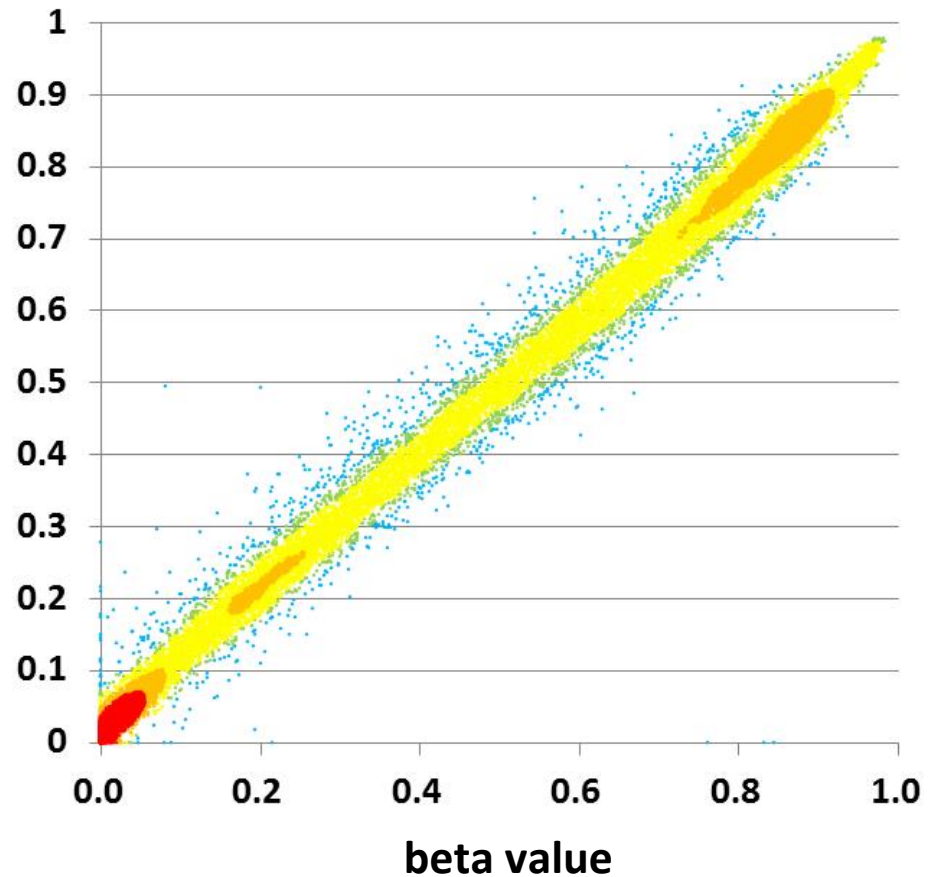
[Shigematsu, Oncol Lett, 2012]

HM450の結果をあらかじめより確かなものにできないか?

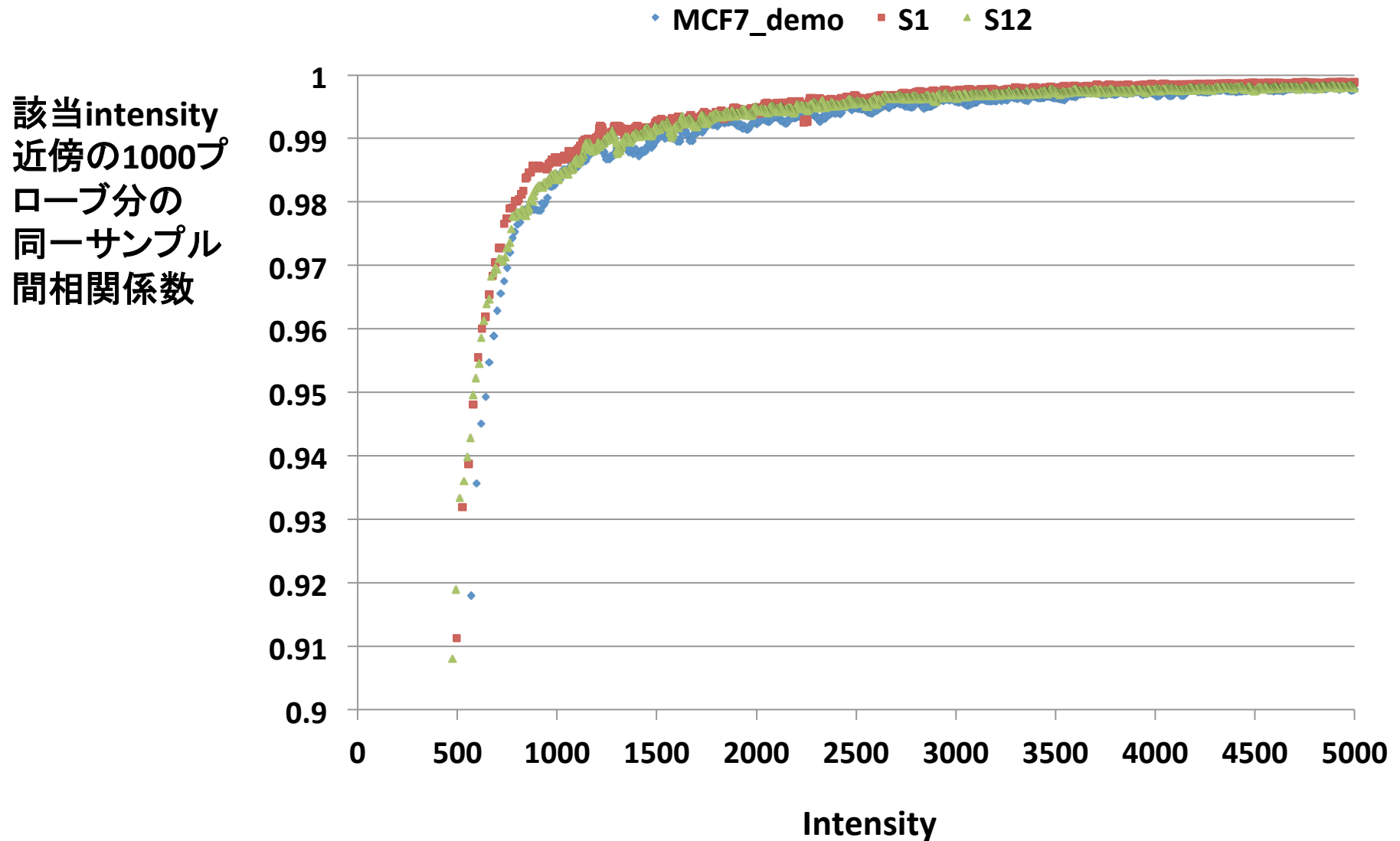
1. 複数回実験
2. 塩基多型データの活用
(Illumina)
3. Detection P-valueの活用
(Illumina)
4. エラー傾向プローブの
除去

HumanMethylation450

$R^2 = 0.993$



HumanMethylation450の再現性はintensityに依存する

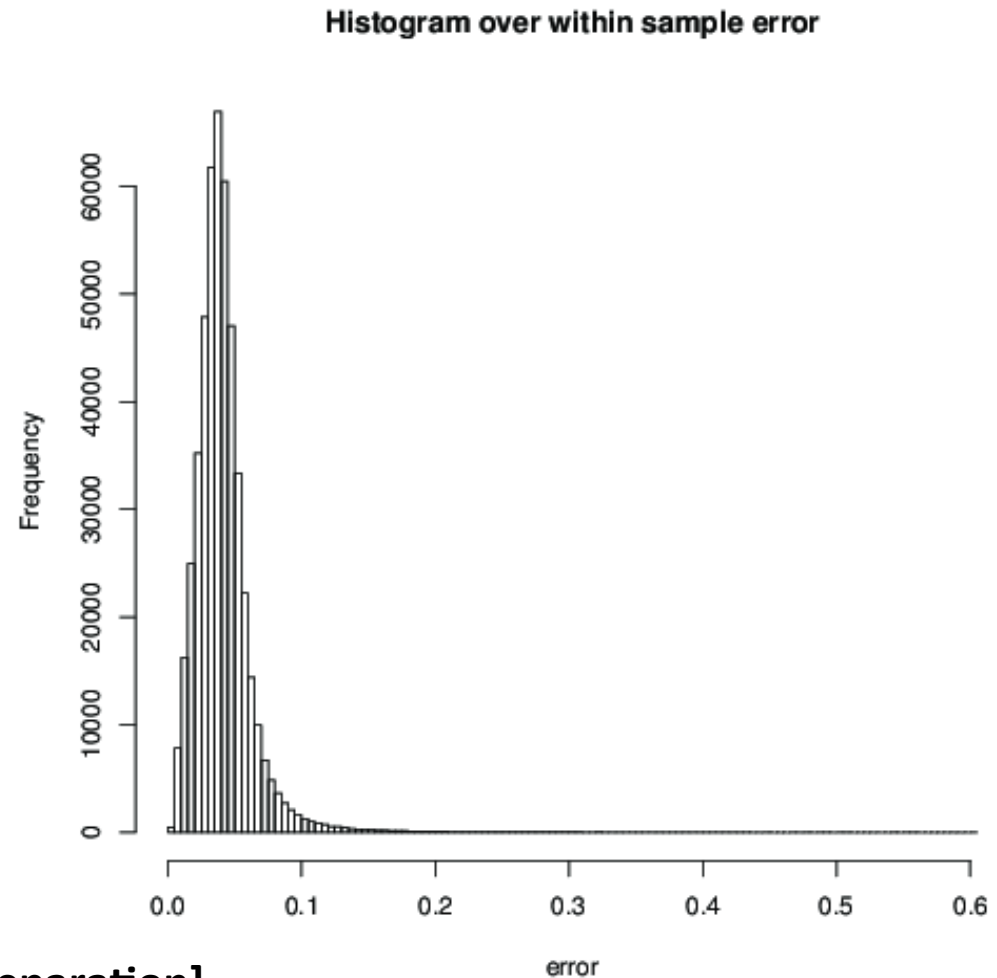


Intensity500以下を示したプローブの結果は無かったことにする

Estimation of the probe errors and their histogram

- Repeated measurements are available for 10 individuals.
- The within-sample variation is estimated by use of "analysis of variance" (ANOVA).

Quantile	Error
95%	0.074
96%	0.079
97%	0.086
98%	0.096
99%	0.117
99.5%	0.144
99.9%	0.222



[Rehnberg *et al.* manuscript in preparation]

薬剤投与によるメチル化変化結果(上位20プローブ)

AGS+5-aza-dC

	error	diff
cg14341177	0.044	0.894
cg01618928	0.064	0.709
cg03281139	0.071	0.698
cg19095568	0.060	0.683
cg09220326	0.049	0.678
cg16402006	0.050	0.677
cg04016485	0.046	0.657
cg20455854	0.044	0.649
cg09373676	0.041	0.643
cg20695611	0.052	0.642
cg18764836	NA	0.640
cg05255994	0.036	0.635
cg08635395	0.042	0.627
cg13321688	0.072	0.626
cg09312135	0.021	0.624
cg14251216	0.249	0.603
cg11481582	0.040	0.600
cg02927821	0.031	0.594
cg20519665	0.051	0.587
cg22950831	0.035	0.580

SH-SY5Y+CDDP

	error	diff
cg06647360	0.033	0.884
cg26446133	0.031	0.868
cg16066505	0.370	0.790
cg25720930	0.455	0.679
cg00040566	0.045	0.628
cg07296841	0.290	0.622
cg18443450	NA	0.562
cg23476209	0.230	0.500
cg25376032	0.232	0.487
cg19871348	0.178	0.471
cg02076747	0.134	0.466
cg11822772	NA	0.458
cg05393828	0.243	0.446
cg18825597	0.231	0.433
cg13769609	0.072	0.433
cg06813250	0.159	0.432
cg04054303	0.226	0.427
cg07565150	0.193	0.423
cg06255004	0.106	0.415
cg03562350	0.291	0.410

メチル化変化の
大半はエラー

[Rehnberg *et al.* manuscript in preparation]

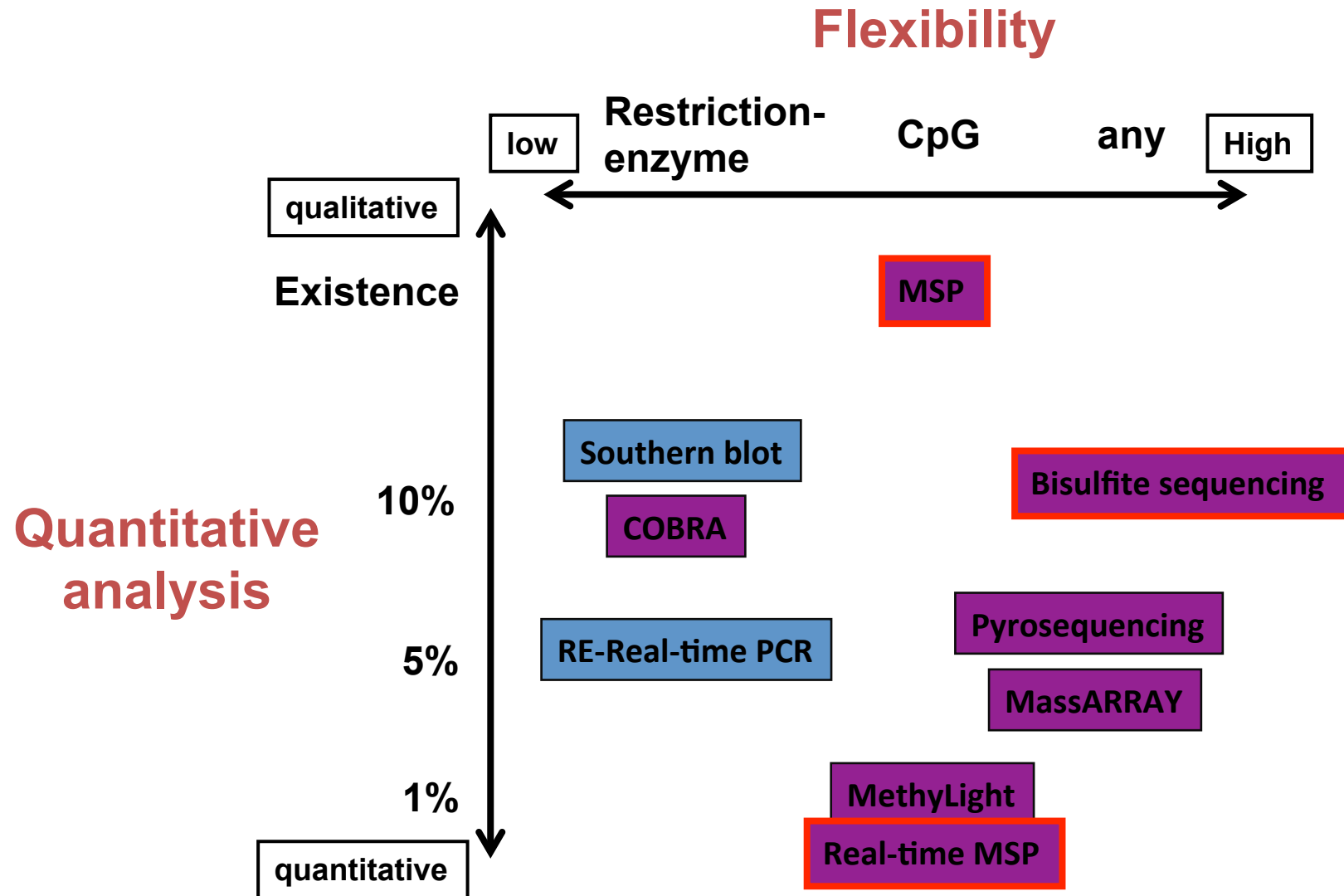
Biologicallyに意味があるDNAメチル化変化の解析

- TSS200/CpG island領域
 - 発現との強い関係が既知
- 狭い領域内での連続性
- 発現解析結果との統合
 - プラットホームが違えば煩雑
 - Excelで行う場合は遺伝子名で揃える。VLOOKUP関数を使えば何とかなる
- GeneOntology解析
 - 何はともあれDAVID (<http://david.abcc.ncifcrf.gov/>)

Topics

1. DNAメチル化解析の意義
2. 網羅的DNAメチル化解析の概要
3. 網羅的メチル化解析例とその簡単なデータ処理
4. 領域個別のメチル化解析

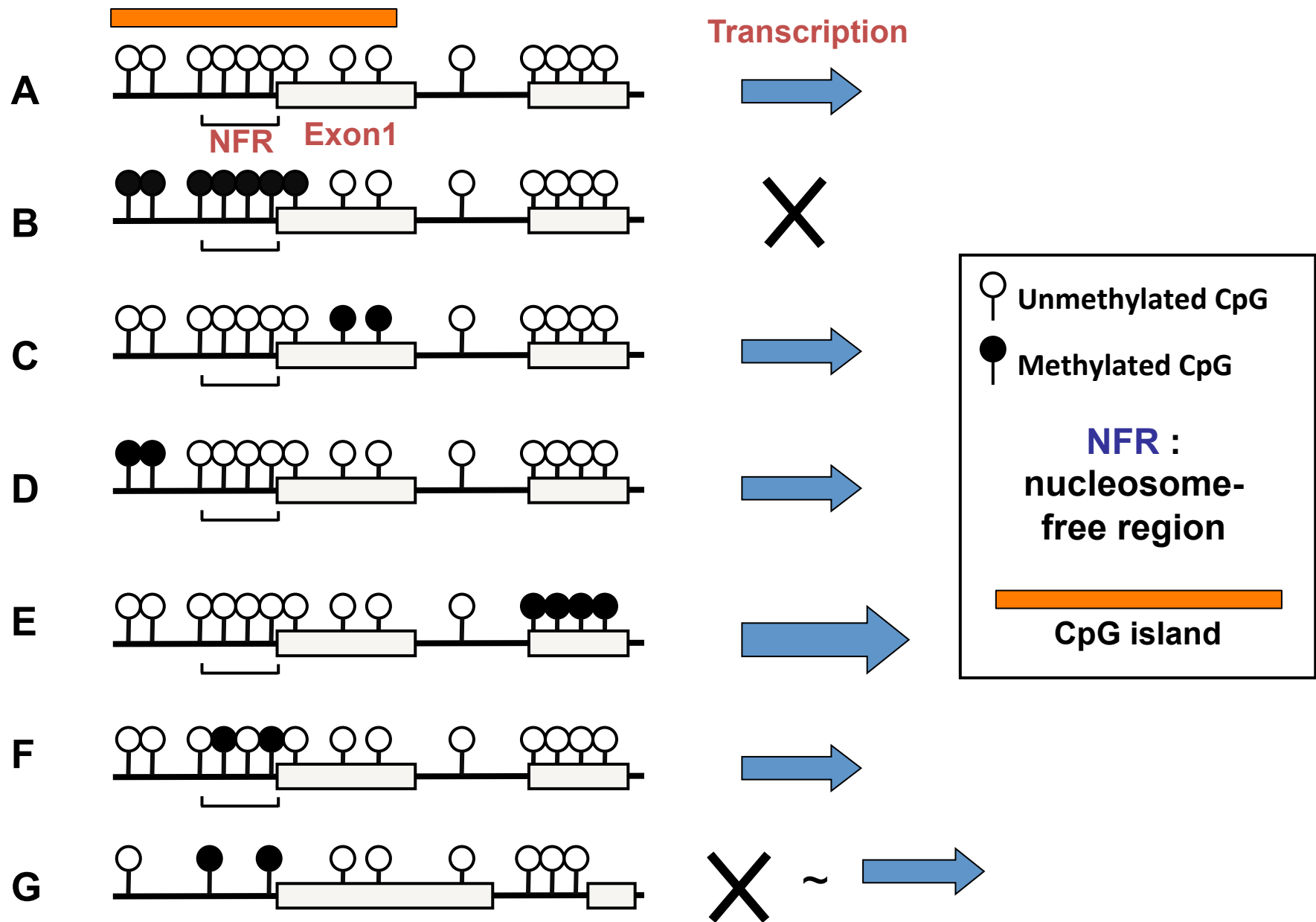
DNAメチル化 (個別領域) 解析方法の選択



Check Point of MSP

- **Primer location**
- **Quantity and quality of template**
- **PCR conditions**

メチル化CpGの位置と発現量の関係 (原則)



CpGアイランドの位置を調べる

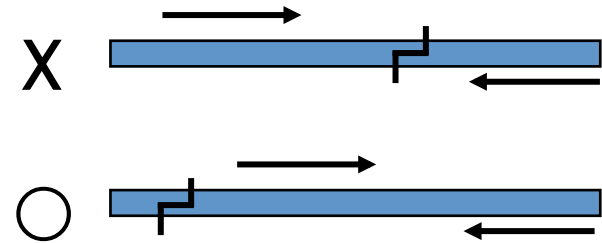
The screenshot displays the NCBI Map Viewer interface for the GATA2 gene region on chromosome 21. The main track shows CpG islands as blue vertical bars. Other tracks include the human genome overview (Build 37.1 and 36.3), UniGene (Hs.709024, Hs.367725, Hs.593717, Hs.287582), RefSeq (GATA2), and Genes_seq (NM_001141, NM_032636). A vertical scale on the right indicates genomic coordinates from 128197K to 128213K. A search bar at the top contains the query '2624[gene_id]'. Navigation and search options like 'Find in This View' and 'Advanced Search' are visible. A sidebar on the left provides map viewer controls and an ideogram of chromosome 21.

NCBI Map ViewerでCpG
アイランドを表示させる

Bisulfite 处理

✓ Fragmentation of genomic DNA

- *Bam*HI or Sonication
- Adequate **fragmentation**

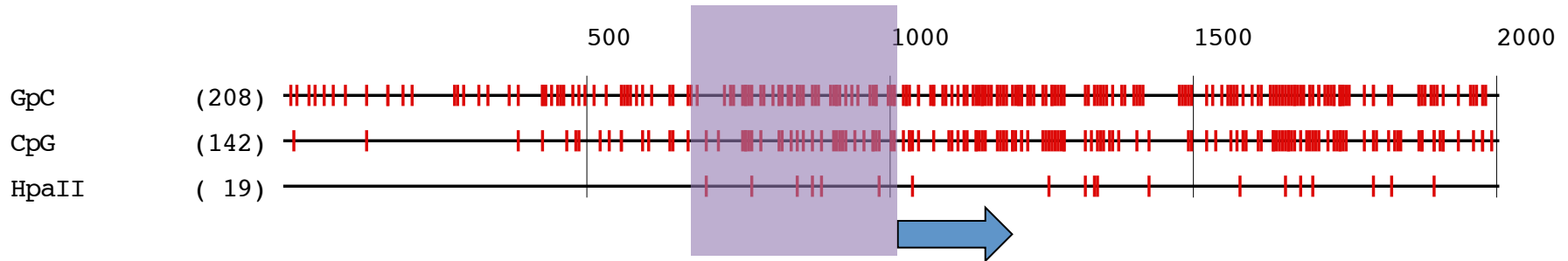


✓ Our protocol

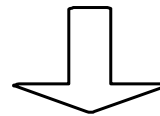
1. 3.1M sodium bisulfite, 6mM hydroquinone, pH 5.0
2. [95°C, 30 sec ⇌ 50°C, 15 min] 15 cycles (x 50°C 16 hours)
3. Desulfonation and purification using Spin column (**Zymo Research**)
 - **pH**
 - High concentration bisulfite, high temperature
 - Commercial kits (Invitrogen or Zymo)

遺伝子転写開始点近傍のMSPプライマー

ヒトGATA2遺伝子



Original sequences (Top strand) 5' - TTACGTGAACAAATAGCTGAGGGGCGGCCGGGCCAGAACG-3'



Bisulfite treatment

M primer (forward)

ATAAATAGTTGAGGGGCGGTC

Methylated DNA 5' -ttaCGtgaauaaatagutgaggggCGgtCGgguuagaaCG -3'

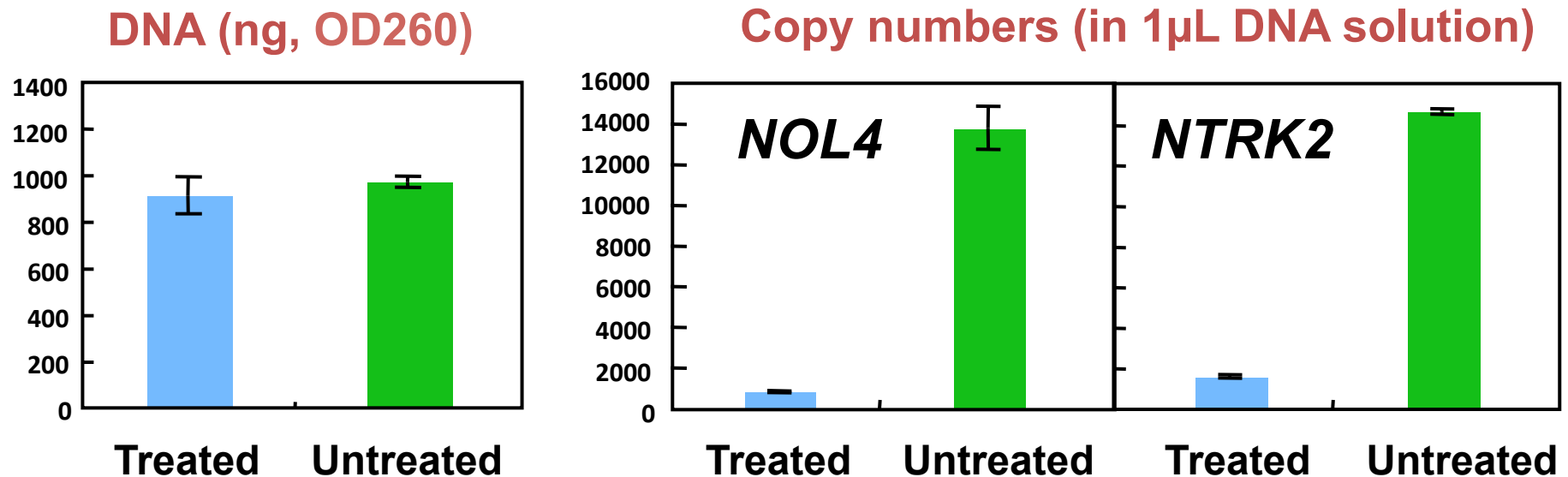
U primer (forward)

GAATAAATAGTTGAGGGGTGGTT

Unmethylated DNA 5' -ttaUGtgaauaaatagutgaggggUGgtUGgguuagaaUG -3'

- CG site on primer 3' end =1, others = 1 to 4

PCR Template DNAはBisulfite処理により 大半が失われる



- 14,000 copies (50 ng) → 700-1,700 copies (5-12%)

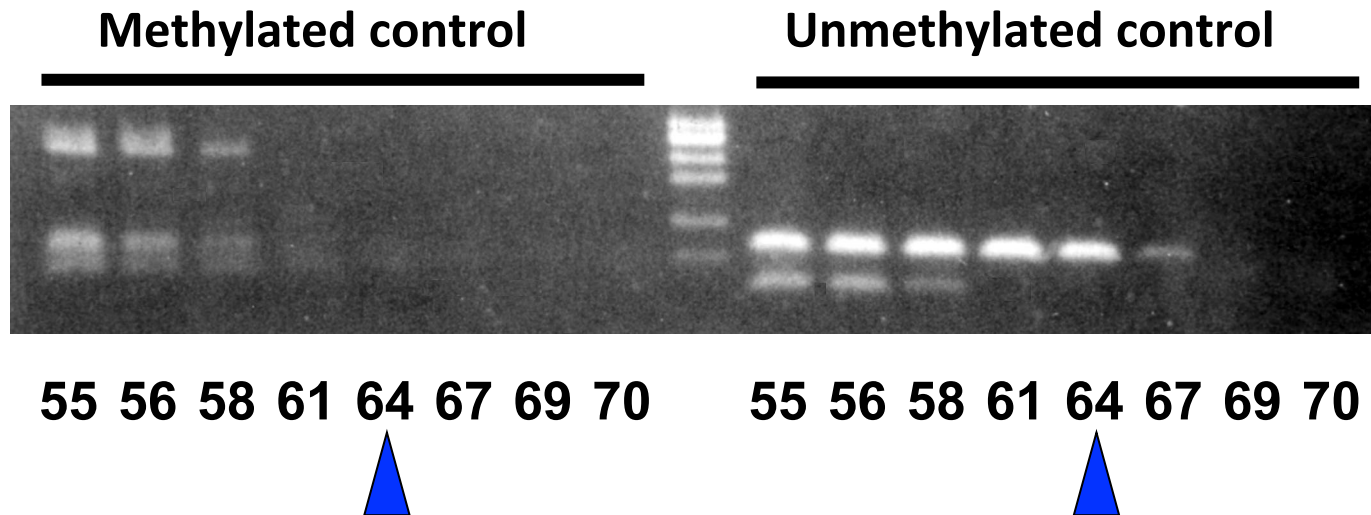
- Inevitable (Any kit!)

- 10 copies for DNA detection / 1 MSP reaction
= 200 copies (100 cells) before bisulfite treatment
= 0.6 ng (human, mouse) / 1 MSP reaction

Our protocol: DNA 1 μg (40-80 PCRs) = 12-25 ng / 1 MSP

Annealing Temperature

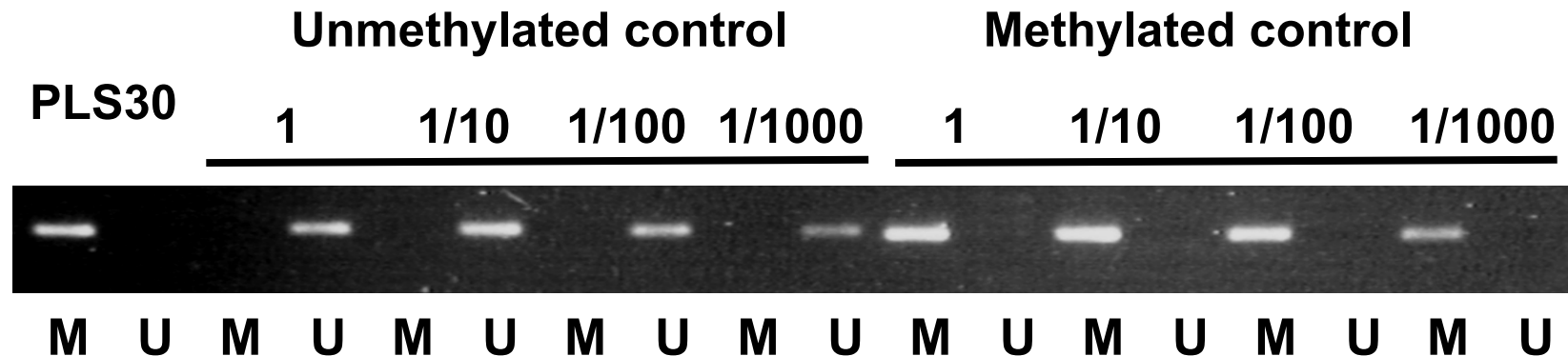
Primer for unmethylated DNA



- Gradient PCR system is useful.

Excessive Number of PCR Cycles

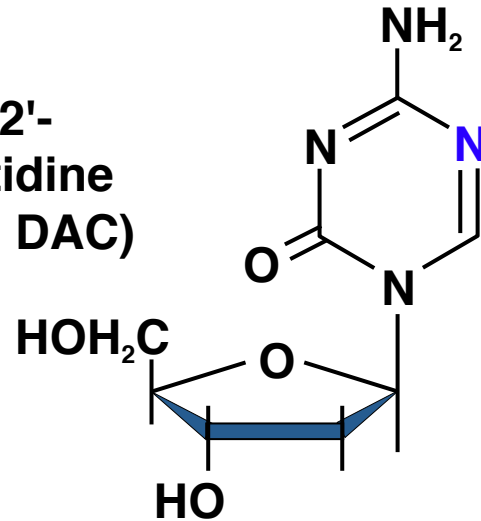
PCR 40 cycles



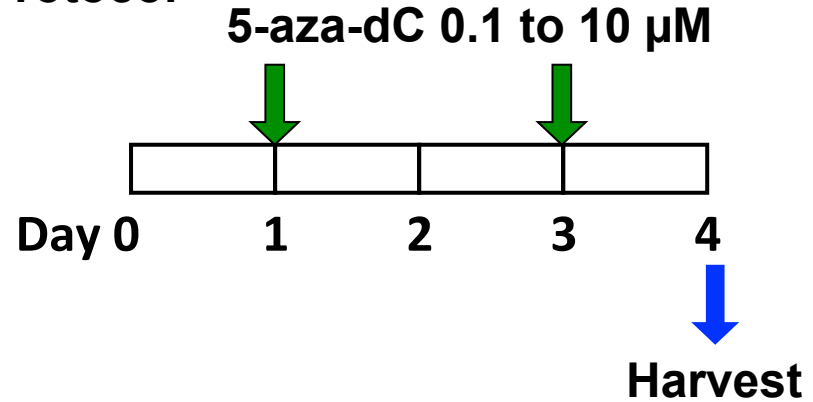
- 0.1 % of methylated DNA can be amplified.
0.1% methylated in cancer??
- **Minimum cycles** to obtain **visible bands** with positive control samples
- 4 cycles were added for test samples (total of 30-36 cycles).

5-Aza-dC 処理による脱メチル化

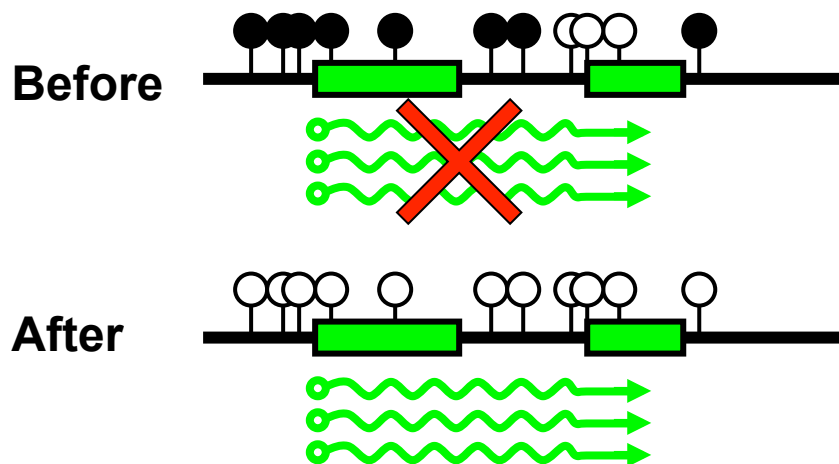
5-Aza-2'-
deoxycytidine
(5-aza-dC, DAC)



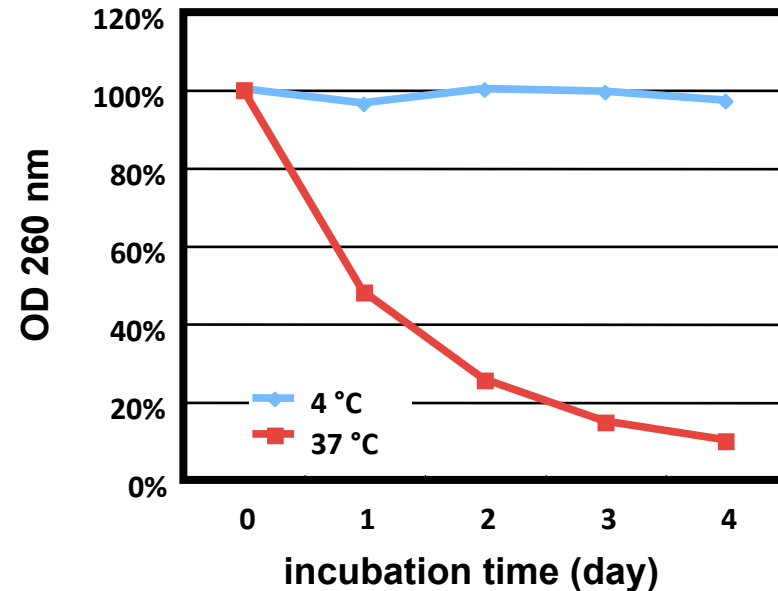
Protocol



Re-expression by demethylation



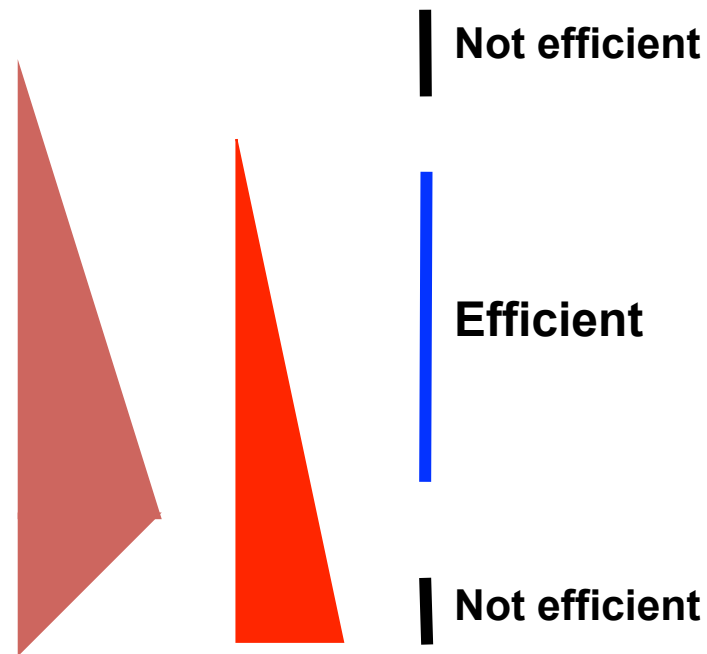
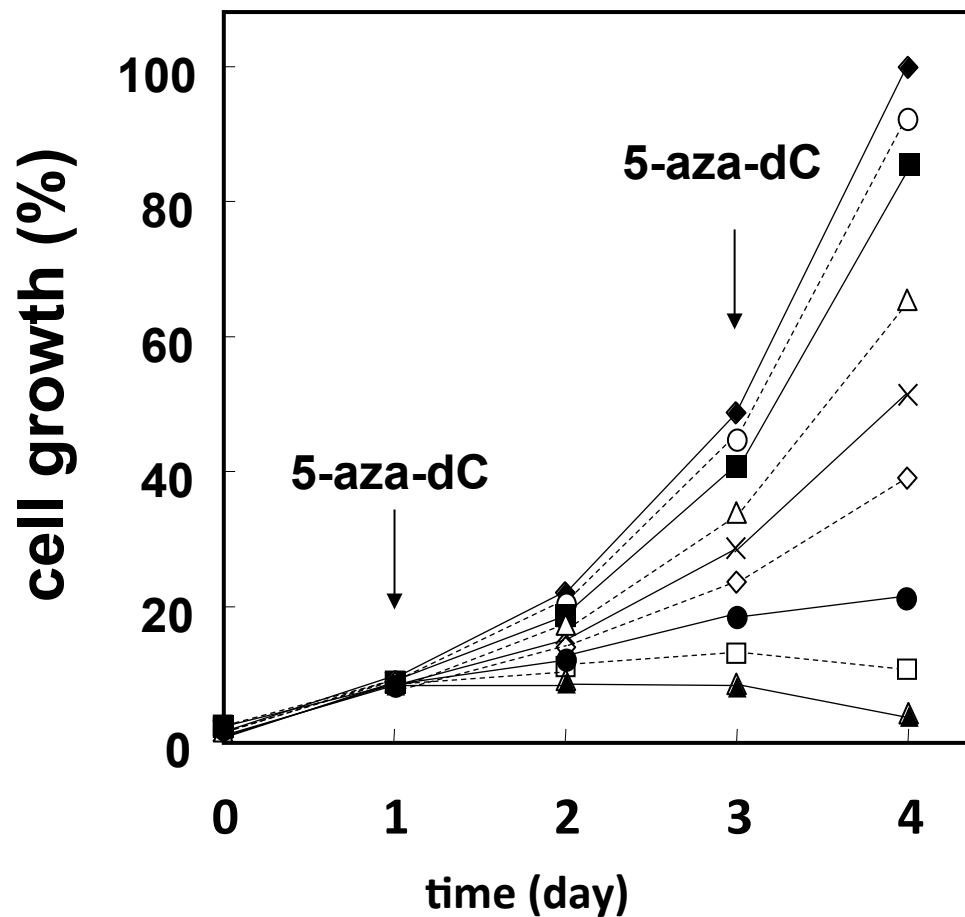
Stability of 5-aza-dC in PBS



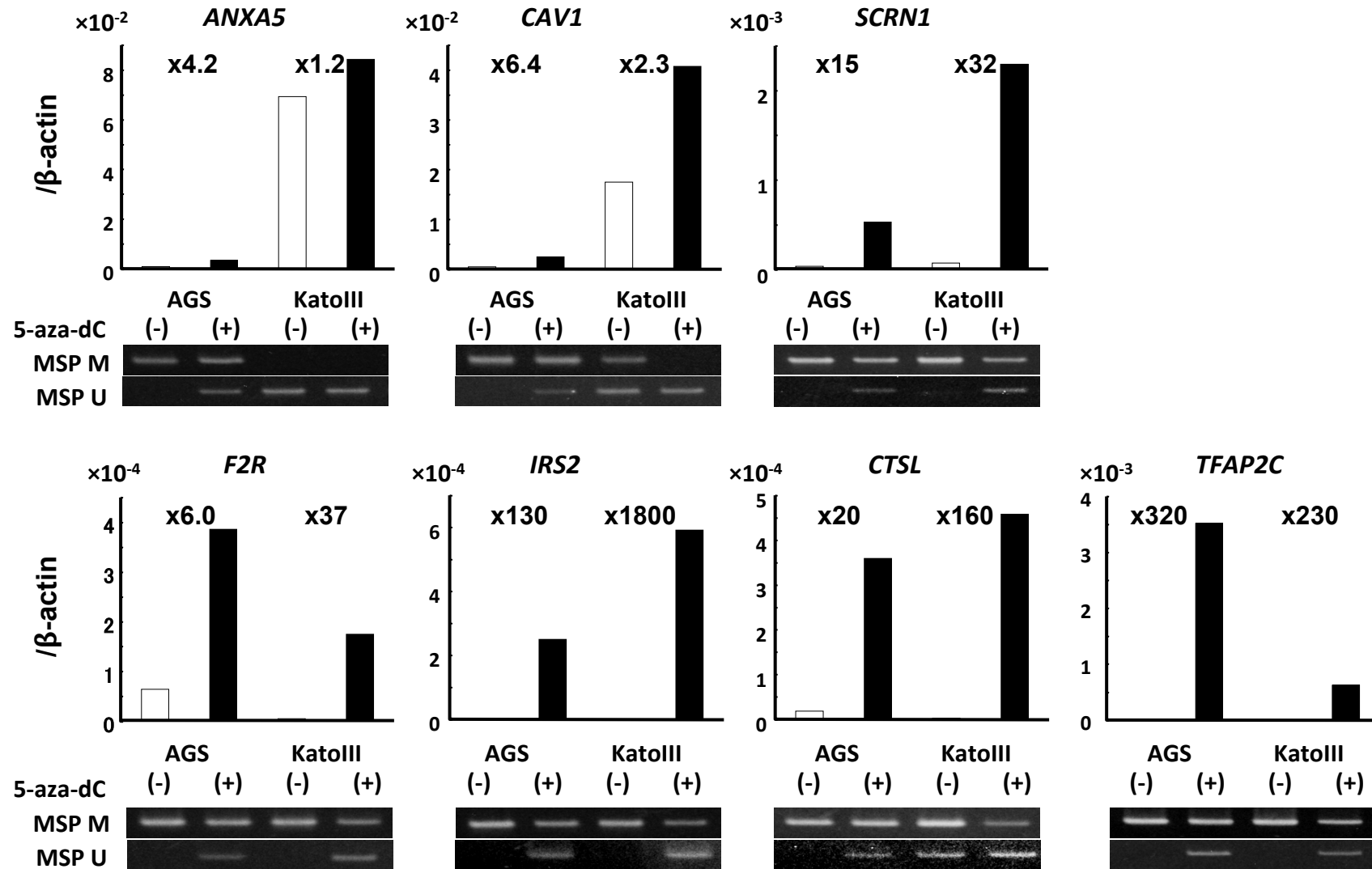
5-Aza-dC の処理濃度

Up-regulation
by demethylation

Up-regulation
by something else



メチル化サイレンシングされていた遺伝子は5-Aza-dC 処理により**大幅に**遺伝子発現が増大する



[Nakajima et al. IJC 2009]

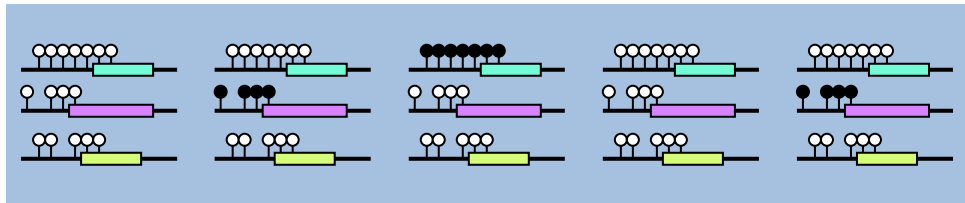
Quantification of DNA Methylation

- **Reliability**
- **Methylation level as marker (Cancer diagnosis, risk)**

	Number of molecules	Low level methylation	Precision	Flexibility	Ease of use	Cost
Bisulfite sequencing	×	△	△	◎	○	×
Combined bisulfite and restriction analysis (COBRA)	×	○	△	△	◎	◎
Pyrosequencing	×	○	○	○	○	△
MethyLight	◎	◎	◎	○	○	△
Real-time methylation-specific PCR (SYBR Green I)	◎	◎	◎	○	○	○
MSP	×	×	×	○	◎	◎

Principle of Real-time MSP

Bisulfite treatment



Real-time PCR
(SYBR Green)

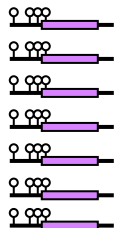


Primer for
unmethylated
(universal)

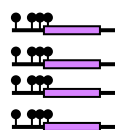
Primer for
methylated

of DNA molecules

Unmethylated

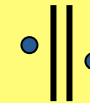


Methylated

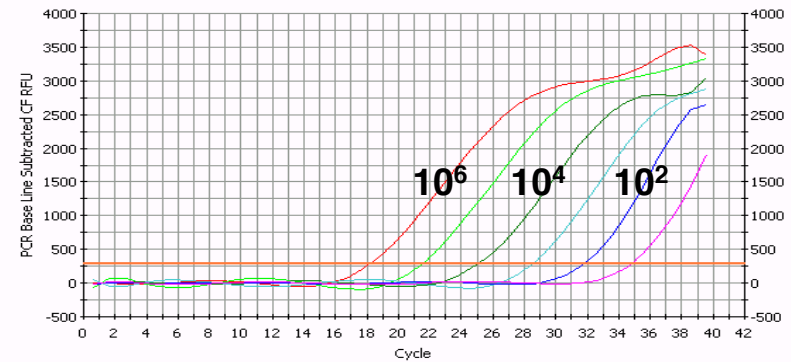


Methylation level (%)

$$= \frac{\# \text{ of M}}{\# \text{ of U} + \# \text{ of M}} \times 100$$

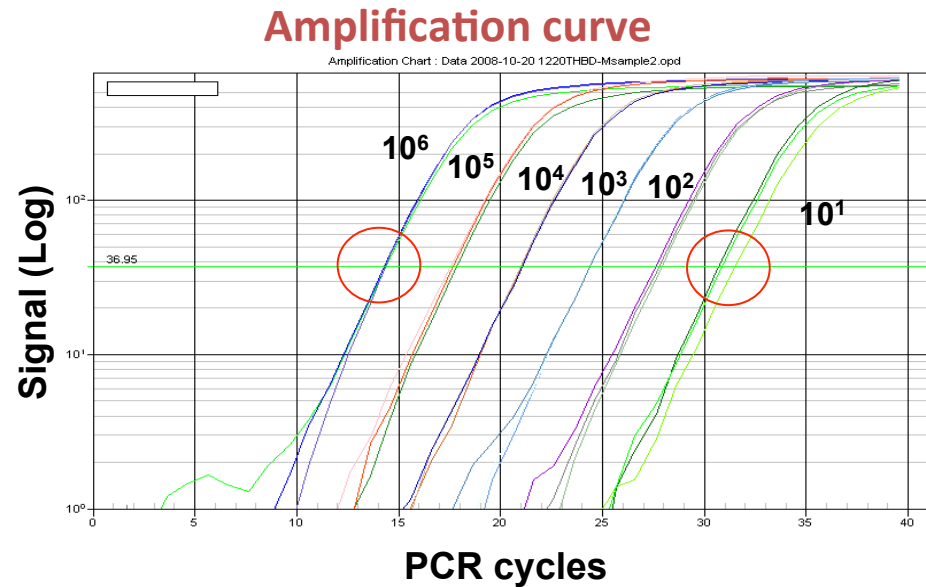


Fraction of cells with methylation



Real-time MSP needs DNA

[Quantification of high copy number of DNA is stable.]

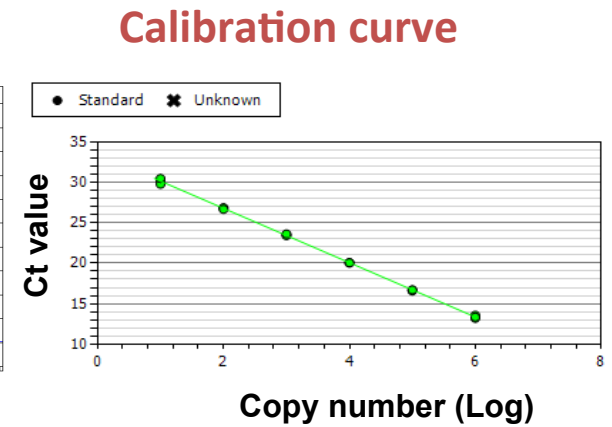
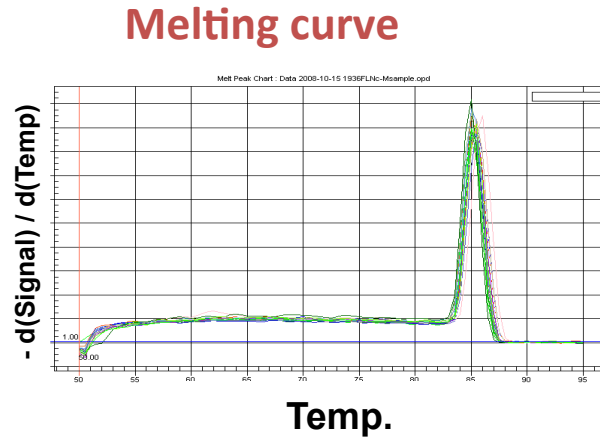
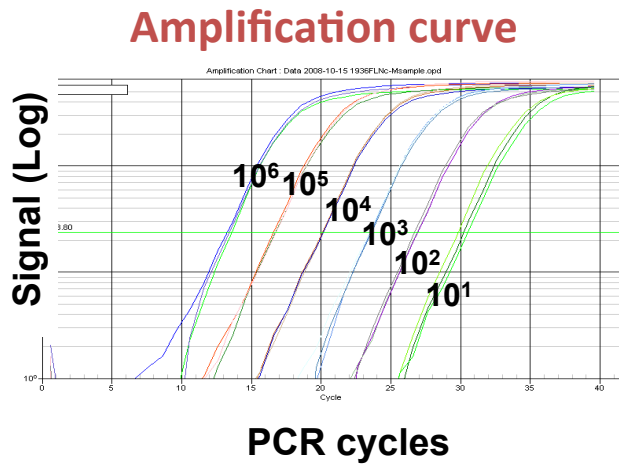
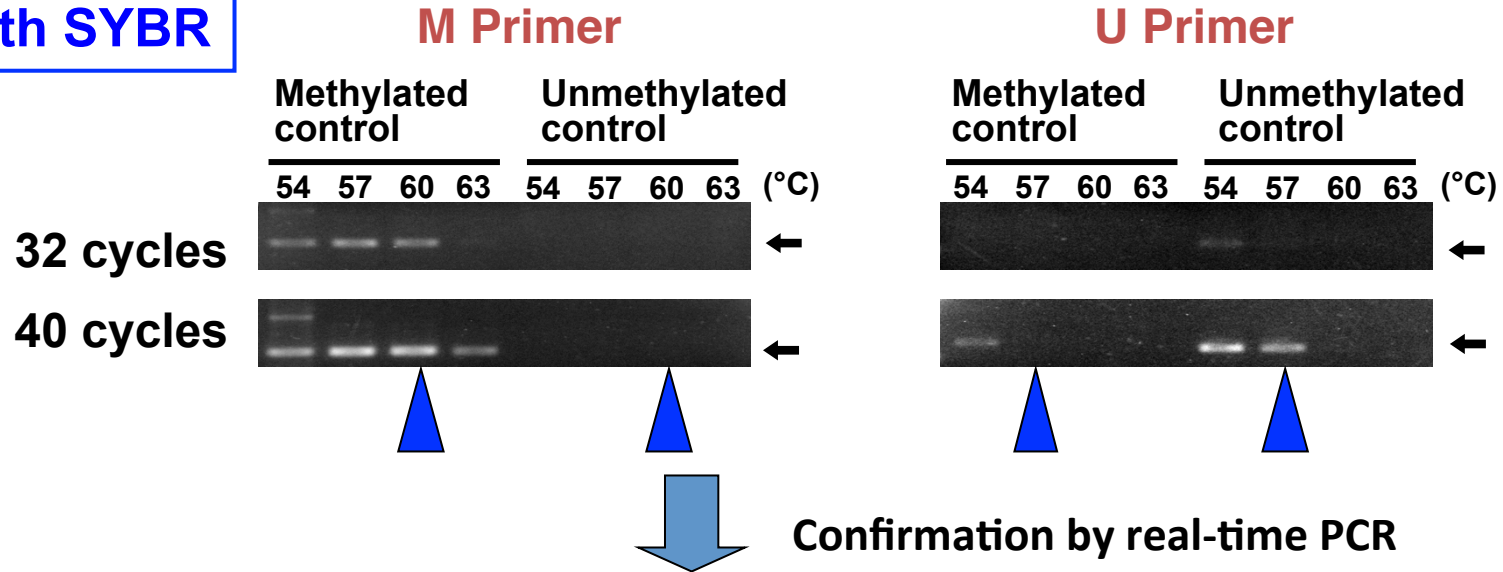


[PCR Template DNA is lost by bisulfite treatment.]

	DNA needed	DNA before bisulfite treatment
1% precision	100 copies	more than 2,000 copies (6 ng)
0.1% precision	1,000 copies	more than 20,000 copies (60 ng)

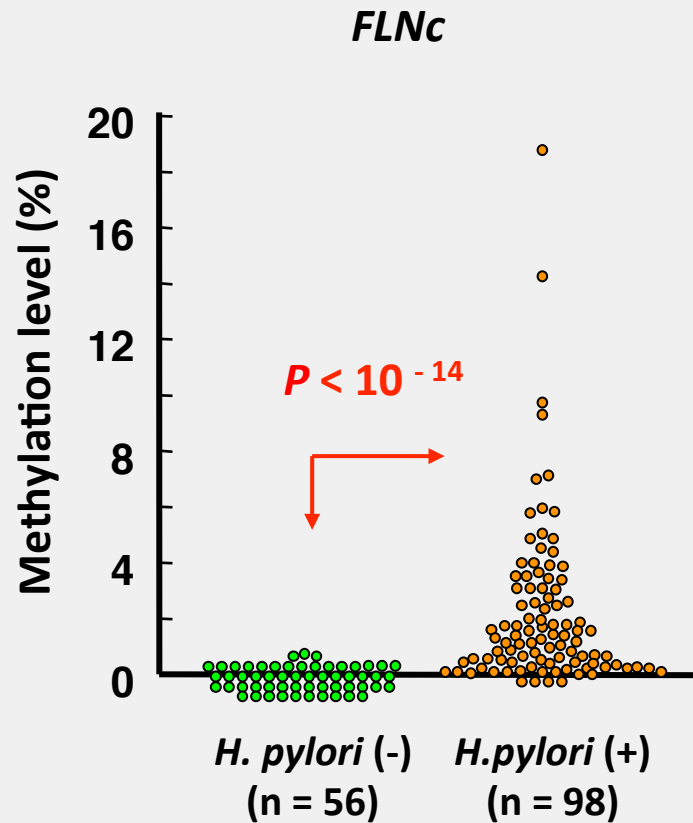
Condition of Real-time MSP

PCR with SYBR

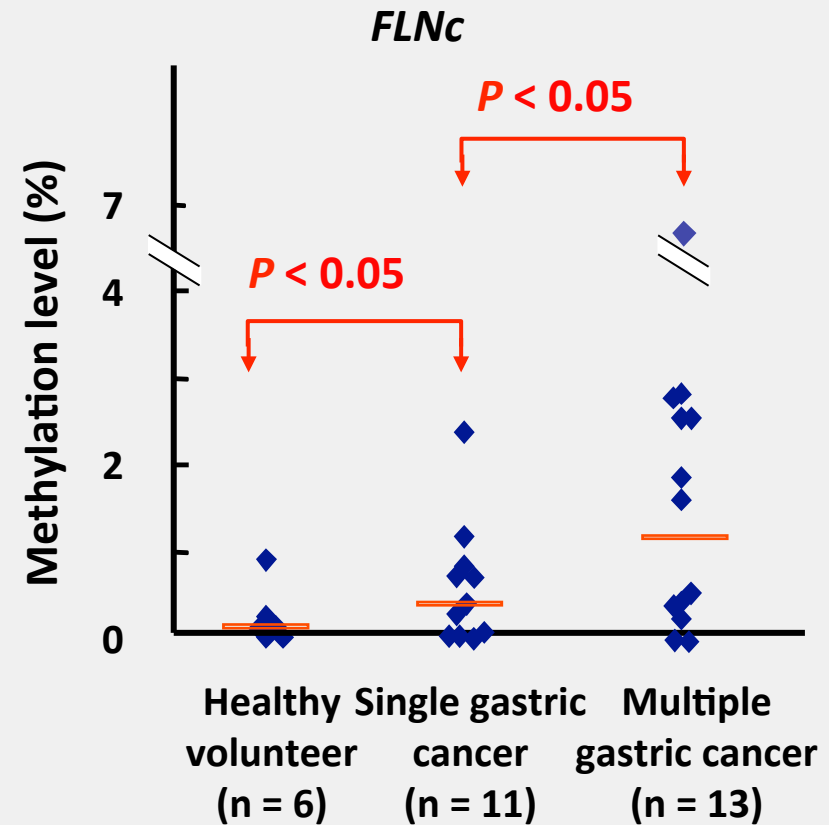


Correlation coefficient : 0.999
PCR efficiency : 98.3%

Quantitative Methylation Analysis Revealed Epigenetic Field for Cancerization



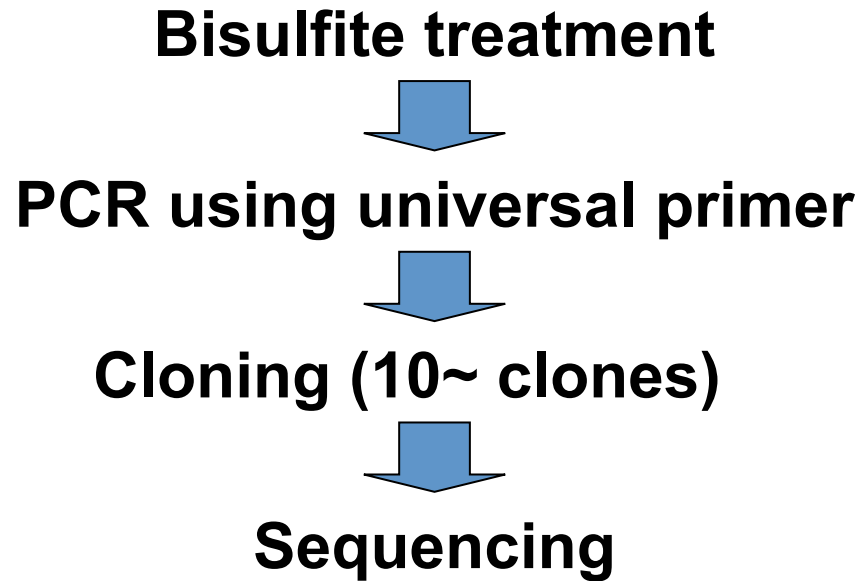
[Maekita, *Clin Cancer Res*, 2006]



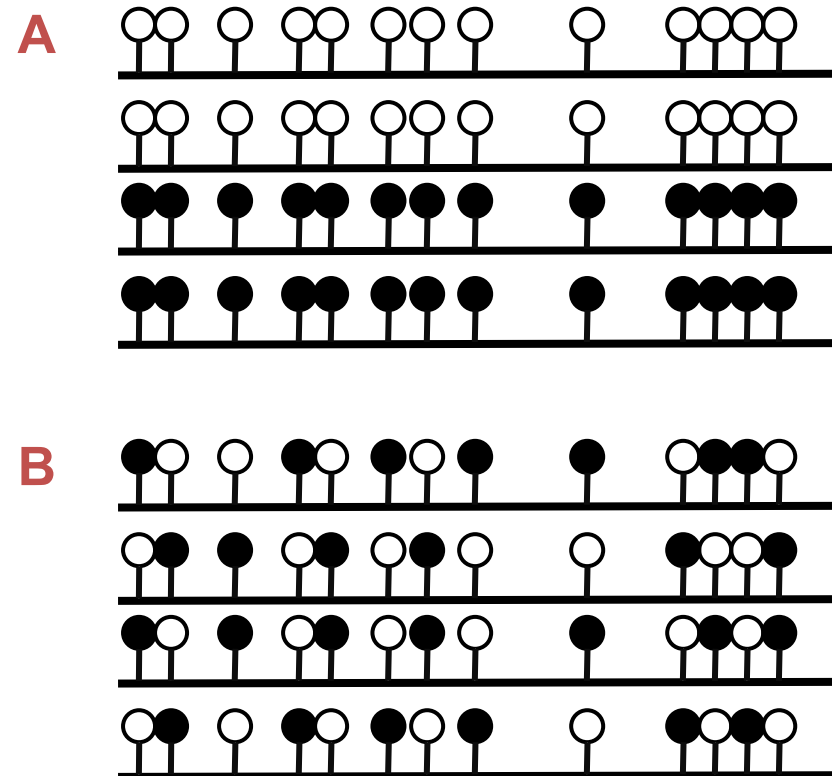
[Nakajima, *CEBP*, 2006]

- ✓ These findings showed the presence of an epigenetic field for cancerization in gastric mucosa.

Bisulfite Sequencing

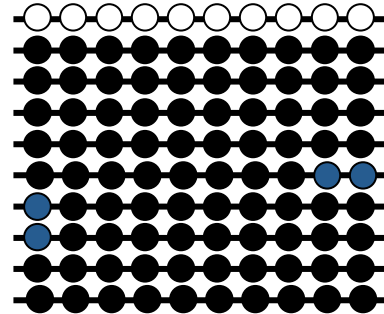
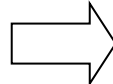
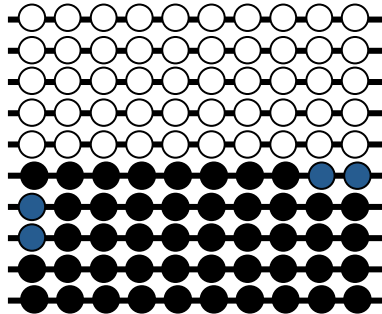
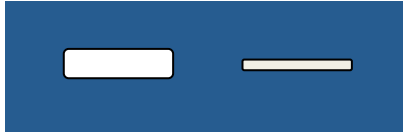
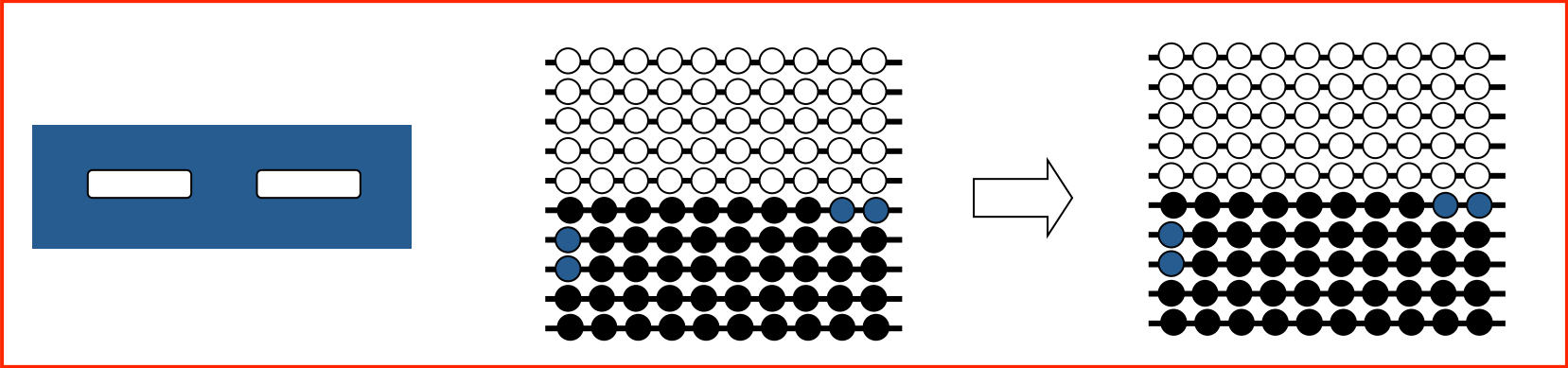
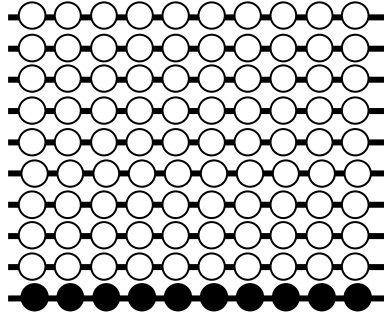
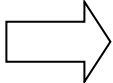
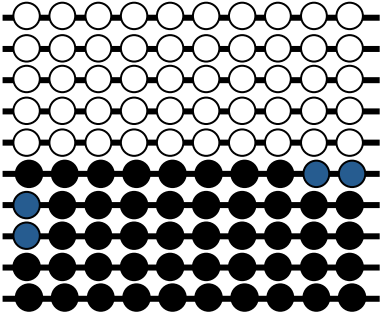
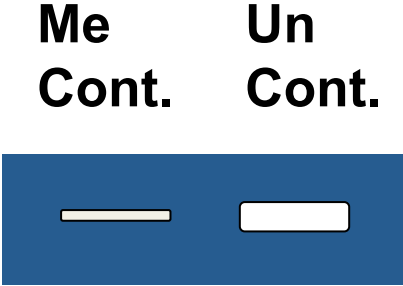


- **Gold standard**
- **Methylation pattern**

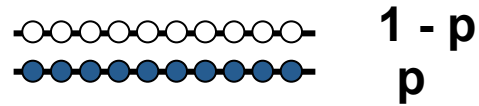


PCR Bias

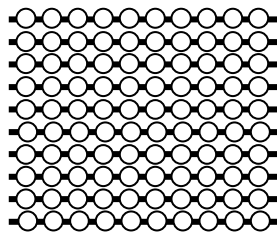
Bisulfite Sequencing



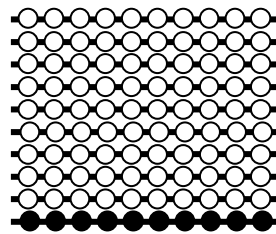
Number of Analyzed Clones in Bisulfite Sequencing



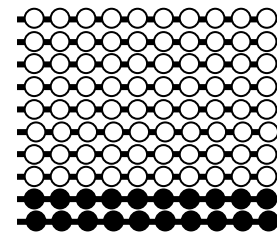
Binomial distribution ($p = 0.5, n = 10$)



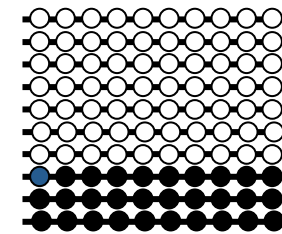
0.1%



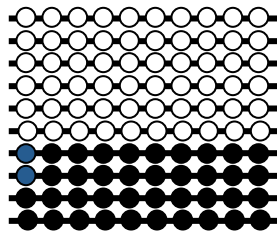
1.0%



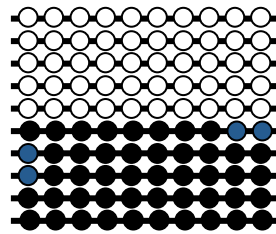
4.4%



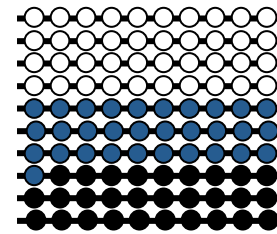
11.7%



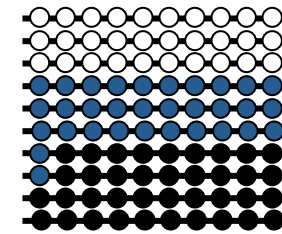
20.5%



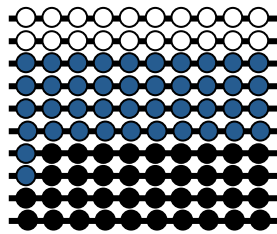
24.6%



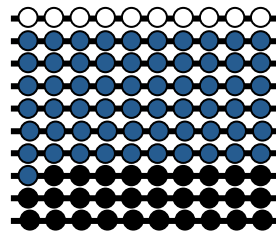
20.5%



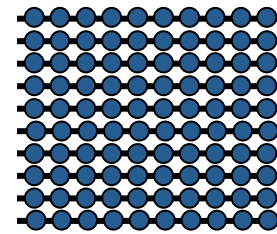
11.7%



4.4%



1.0%



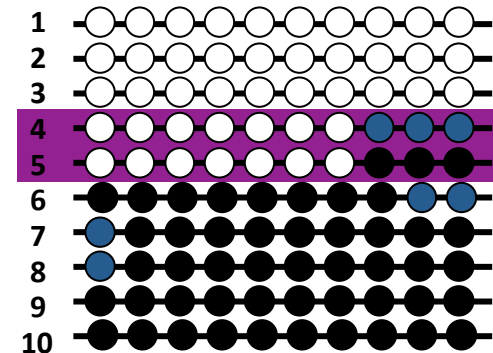
0.1%

(From Sasaki et. al)

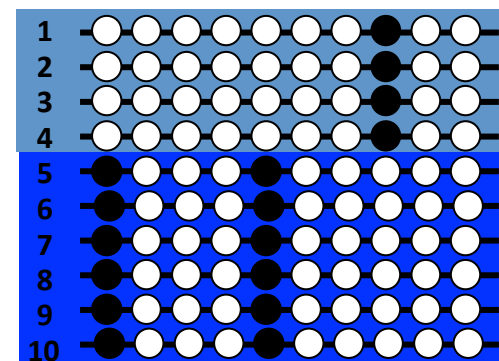
Number of DNA Molecules in Bisulfite Sequencing

- 100 copies for methylation pattern / 1 reaction
= **2,000 copies** (1000 cells) before bisulfite treatment
= **6 ng** (human, mouse) / 1 reaction

- **Nonspecific products of PCR**



- **Duplication of clones**



参考書

羊土社 実験医学別冊

エピジェネティクス実験プロトコール
DNAメチル化網羅的解析については古い

マイクロアレイデータ統計解析プロトコール

Acknowledgment

National Cancer Center Research Institute

Toshikazu Ushijima

Tohru Niwa

Hideyuki Takeshima

Yasuyuki Shigematsu

Ken Gyobu

Yasunori Matsuda

Kosuke Hosoya

Monali Lahoti

Satomi Takahashi