UC San Diego
UC San Diego Previously Published Works

Title

Quantitative comparison of DNA methylation assays for biomarker development and clinical applications

Permalink

https://escholarship.org/uc/item/7tf2342w

Journal

Nature Biotechnology, 34(7)

ISSN

1087-0156

Authors

Bock, Christoph
Halbritter, Florian
Carmona, Francisco J
et al.

Publication Date

2016-07-01

DOI

10.1038/nbt.3605

Peer reviewed

ANALYSIS

Quantitative comparison of DNA methylation assays for biomarker development and clinical applications

Christoph Bock[1–3], Florian Halbritter[1], Francisco J Carmona[4], Sascha Tierling[5], Paul Datlinger[1], Yassen Assenov[6], María Berdasco[4], Anke K Bergmann[7,8], Keith Booher[9], Florence Busato[10], Mihaela Campan[11], Christina Dahl[12], Christina M Dahmcke[12], Dinh Diep[13], Agustín F Fernández[14,15,16], Clarissa Gerhauser[6], Andrea Haake[7], Katharina Heilmann[6], Thomas Holcomb[17], Dianna Hussmann[18], Mitsuteru Ito[19], Ruth Kläver[20], Martin Kreutz[20], Marta Kulis[21], Virginia Lopez[14,15,16], Shalima S Nair[22,23], Dirk S Paul[24], Nongluk Plongthongkum[13], Wenjia Qu[22], Ana C Queirós[21], Frank Reinicke[20], Guido Sauter[25], Thorsten Schlomm[25], Aaron Statham[22], Clare Stirzaker[22,23], Ruslan Strogantsev[19], Rocío G Urdinguio[14,15,16], Kimberly Walter[17], Dieter Weichenhan[6], Daniel J Weisenberger[11], Stephan Beck[24], Susan J Clark[22,23], Manel Esteller[4,26,27], Anne C Ferguson-Smith[19], Mario F Fraga[14–16], Per Guldberg[12], Lise Lotte Hansen[18], Peter W Laird[11,28], José I Martín-Subero[21], Anders O H Nygren[29], Ralf Peist[20], Christoph Plass[6], David S Shames[17], Reiner Siebert[7,30], Xueguang Sun[9], Jörg Tost[10], Jörn Walter[5] & Kun Zhang[13] for the BLUEPRINT consortium

DNA methylation patterns are altered in numerous diseases and often correlate with clinically relevant information such as disease subtypes, prognosis and drug response. With suitable assays and after validation in large cohorts, such associations can be exploited for clinical diagnostics and personalized treatment decisions. Here we describe the results of a community-wide benchmarking study comparing the performance of all widely used methods for DNA methylation analysis that are compatible with routine clinical use. We shipped 32 reference samples to 18 laboratories in seven different countries. Researchers in those laboratories collectively contributed 21 locus-specific assays for an average of 27 predefined genomic regions, as well as six global assays. We evaluated assay sensitivity on low-input samples and assessed the assays' ability to discriminate between cell types. Good agreement was observed across all tested methods, with amplicon bisulfite sequencing and bisulfite pyrosequencing showing the best all-round performance. Our technology comparison can inform the selection, optimization and use of DNA methylation assays in large-scale validation studies, biomarker development and clinical diagnostics.

DNA methylation is an epigenetic mark widely studied for its association with diseases such as cancer[1] and autoimmune disorders[2], with environmental exposures[3] and with other biological phenomena[4,5]. Strong associations between DNA methylation patterns and clinical phenotypes can be used as biomarkers for diagnosing diseases and guiding treatment[6,7]. For example, DNA methylation biomarkers have been shown to support clinical decisions in various cancers[8–14] and

are also used for noninvasive prenatal testing[15], for quality control of cultured cells[16] and for forensic applications[17,18].

DNA methylation biomarkers have several advantages that qualify them for broad use as *in vitro* diagnostics: (i) DNA methylation is cell-type-specific but robust toward transient perturbations, thus complementing static DNA-sequence-based biomarkers and volatile RNA-expression-based biomarkers. (ii) DNA methylation is a binary mark (i.e., for a single cell and allele, each CpG is either methylated or unmethylated), which facilitates reliable measurements on heterogeneous and degraded samples. (iii) The infrastructure for assaying DNA methylation biomarkers is already present in many clinical diagnostics laboratories, as the assays are similar to those used for DNA-sequence-based biomarkers. (iv) DNA methylation biomarkers are straightforward to integrate into routine clinical workflows because DNA is more stable than RNA and does not require any special handling. (v) DNA methylation patterns are faithfully retained during long-term storage as fresh-frozen or formalin-fixed, paraffin-embedded (FFPE) samples.

Genome-wide mapping and analysis of DNA methylation has become feasible for patient cohorts with thousands of samples[19,20], and epigenome-wide association studies have been conducted for numerous biomedically relevant phenotypes[21,22]. To translate relevant epigenome associations into clinically useful biomarkers, it is necessary to select a manageable set of highly informative genomic regions, to target these loci with DNA methylation assays that are sufficiently fast, cheap, robust and widely available to be useful for routine clinical diagnostics[23–25], and to confirm their predictive value in large validation cohorts.

Here we systematically compared and evaluated the most promising assays for measuring DNA methylation in large cohorts, clinical diagnostics and biomarker development. This multicenter study included research groups from seven countries across three continents, organized

---

A full list of affiliations appears at the end of the paper.

by the BLUEPRINT project[26] in the context of the International Human Epigenome Consortium[27] and as a follow-up to a previous comparison of genome-wide DNA methylation assays[28–30]. Overall, our results show that most assays provide high accuracy and robustness, although we observed some differences between assay types and laboratories. We provide detailed documentation of all contributed assays (**Supplementary Data 1**), such that this study can be used not only to guide assay selection but also as a resource of validated DNA methylation protocols.

RESULTS

Study design and assay selection

We selected assays based on comprehensive literature review, and for each promising assay we selected at least one research group that had extensive prior experience using that particular assay (**Fig. 1a**). In total, we invited 25 research groups, of which 19 agreed to participate. All participants received DNA aliquots for 32 reference samples, together with a list of 48 preselected genomic regions to be targeted. They designed the assays independently, analyzed the reference samples with their assays of choice and submitted the final results together with a detailed assay report for centralized benchmarking analysis by the study coordinator. Ultimately, 18 of the 19 participating research groups submitted complete analysis reports for a total of 27 assays (**Table 1** and **Supplementary Table 1**). All contributed assays can be classified into one of three categories, which we summarize below.

First, absolute DNA methylation assays provide a quantitative measure of DNA methylation levels at single-CpG resolution. We included 16 absolute assays based on four technologies: (i) amplicon bisulfite sequencing (AmpliconBS) uses next-generation sequencing (NGS) of pooled PCR amplicons derived from bisulfite-converted DNA[31–33]. (ii) Enrichment bisulfite sequencing (EnrichmentBS) is similar to AmpliconBS in its use of bisulfite conversion and NGS, but it uses highly scalable techniques such as padlock probes or microdroplet-based amplification to enrich many genomic regions in parallel rather than relying on separate PCRs for each individual region[34,35]. (iii) Mass spectrometric analysis of DNA methylation (EpiTyper) combines bisulfite conversion, *in vitro* transcription and uracil-specific cleavage with mass-spectrometry-based quantification of fragment lengths[36–38]. (iv) Bisulfite pyrosequencing (Pyroseq) applies sequencing by synthesis[39] to single PCR amplicons obtained from bisulfite-converted DNA[40,41].

Second, relative DNA methylation assays measure DNA methylation by comparing samples to a suitable reference. This approach is mainly used for detecting methylated DNA fragments in an excess of unmethylated fragments, but it also provides rough estimates of absolute DNA methylation levels. We included five relative DNA methylation assays based on three alternative technologies: (v) MethyLight uses PCR amplification of bisulfite-converted DNA in combination with fluorescently labeled probes that hybridize specifically to a predefined DNA methylation pattern, typically that of fully methylated DNA[42,43]. (vi) Methylation-specific melting assays, including methylation-sensitive high-resolution melting (MS-HRM) and methylation-specific melting curve analysis (MS-MCA), apply melting curve analysis to amplicons obtained from bisulfite-converted DNA, which provides a semiquantitative measure of cytosines that have been converted to thymines[44–46]. (vii) Quantitative methylation-specific PCR (qMSP) uses DNA-methylation-specific primers in combination with real-time PCR to compare the prevalence of a specific DNA methylation pattern with that of a suitable reference[47,48].

Third, global DNA methylation assays measure a sample's total DNA methylation content, which can be useful for measuring hypomethylation in cancer[49] and the response to drugs that inhibit DNA methylation[50]. We included five global assays based on three alternative technologies: (viii) High-performance liquid chromatography followed by mass spectrometry (HPLC-MS) quantifies the amount of 5-methylcytosine based on its mass difference compared to unmethylated cytosine[51]. (ix) Immunoquantification of global DNA methylation (Immunoquant) uses a modified enzyme-linked immunosorbent assay (ELISA) with an antibody against 5-methylcytosine to quantify the total amount of methylated DNA in a given DNA sample[52]. (x) Bisulfite pyrosequencing of repetitive DNA elements (Pyroseq AluYb8/D4Z4/LINE/NBL2) applies pyrosequencing to amplicons obtained from bisulfite-converted DNA using primers that amplify multiple instances of the selected type of repeat[53–56], which assumes that averaged local DNA methylation levels across specific repetitive regions correlate with global DNA methylation levels.

Given the study's focus on clinically applicable assays, we did not include emerging technologies that have not yet been shown to be practically useful in large-scale studies, e.g., nanopores, nanowire transistors, quantum dots, single-molecule real-time sequencing and atomic force spectroscopy[57]. We also did not include genome-wide assays such as whole-genome bisulfite sequencing, reduced-representation bisulfite sequencing, methylated DNA immunoprecipitation sequencing or methyl-CpG binding domain enriched sequencing, given that these assays have been benchmarked previously[28–30], and are currently too cumbersome and expensive for routine clinical diagnostics. However, we did include the Infinium 450k assay[58], which we also used to select the target regions, and we performed a limited amount of clonal bisulfite sequencing[59,60], given that this assay was until recently considered the gold standard but has been largely superseded by less labor-intensive assays. Our benchmarking did not explicitly address non-CpG methylation nor DNA methylation variants (5hmC, 5fC and 5caC), but most of the included assays can be used to measure non-CpG methylation as well as CpG methylation, and they can also be adapted to distinguish between DNA methylation variants[61–63]. Finally, we note that all contributed locus-specific assays were bisulfite-based, although we had invited four research groups that had expertise in alternative technologies.

Reference samples and target regions

We prepared 32 reference samples that mimic typical applications of DNA methylation assays in biomedical research and clinical diagnostics (**Supplementary Table 2**). This sample set included DNA extracted from six pairs of primary colon tumor and adjacent normal colon tissue samples ('tumor/normal), DNA from two cell lines before and after treatment with a demethylation-inducing drug ('drug/control'), a titration series with partially methylated DNA spiked into unmethylated DNA ('titration 1'), another titration series with DNA from a cancer cell line spiked into whole blood DNA ('titration 2'), and DNA from two matched pairs of fresh-frozen and FFPE xenograft tumors ('frozen/FFPE').

To establish suitable targets for the locus-specific assays, we performed genome-scale DNA methylation analysis with the Infinium 450k assay and selected 48 differentially methylated CpGs that cover a broad range of technical challenges encountered in biomarker development (**Supplementary Table 3**). For example, we included genomic regions with high and low CpG density, GC content and repetitive DNA overlap. As an additional challenge, we included a single-nucleotide polymorphism (SNP) that replaces a potentially methylated CpG by an always unmethylated TpG dinucleotide in some of the reference samples.

To each contributing laboratory we sent aliquots of ~1 µg DNA for each of the 32 reference samples. In addition, we provided a
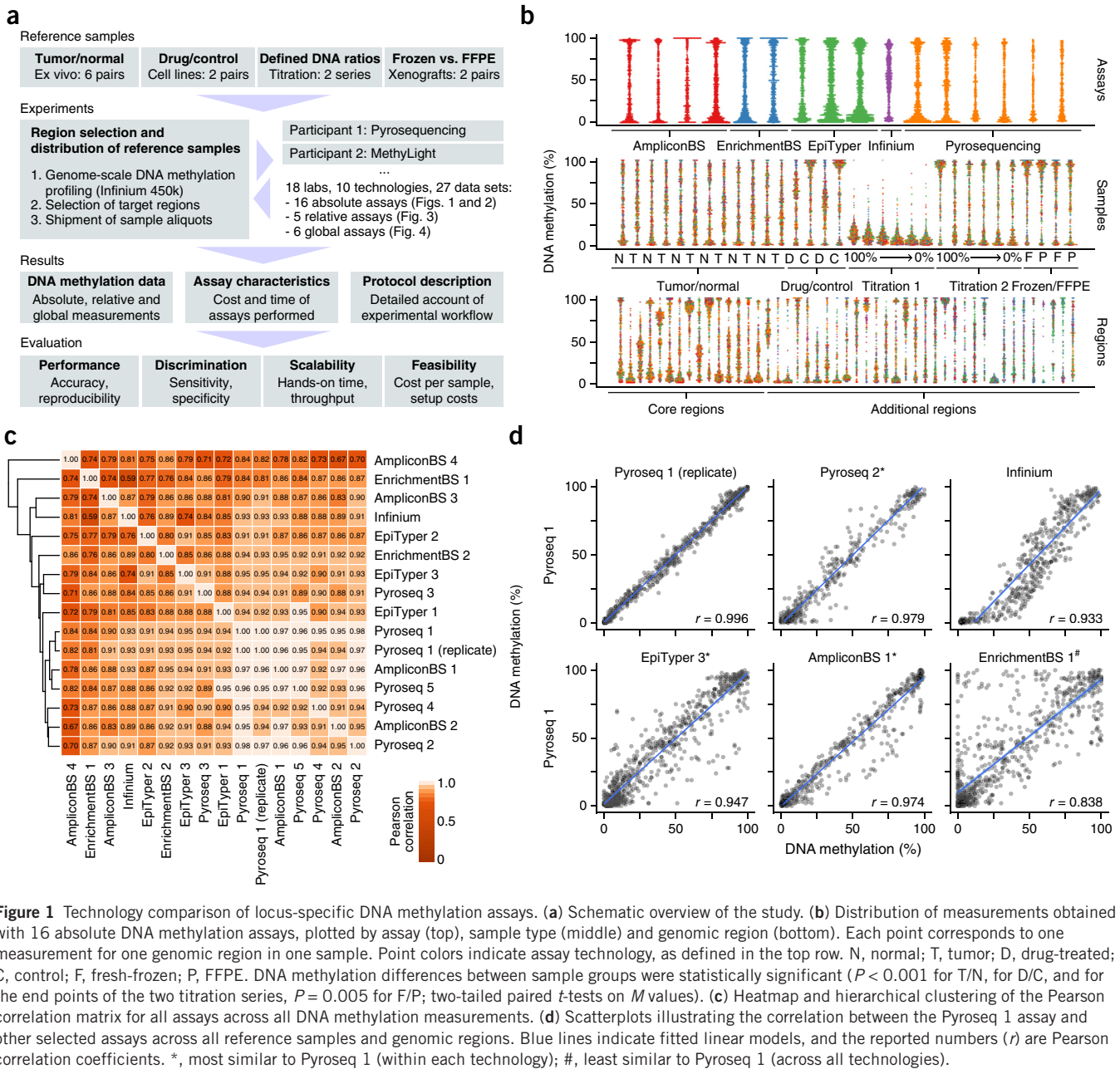
a

Reference samples

Tumor/normal	Drug/control	Defined DNA ratios	Frozen vs. FFPE
Ex vivo: 6 pairs	Cell lines: 2 pairs	Titration: 2 series	Xenografts: 2 pairs

Experiments

Region selection and distribution of reference samples

1. Genome-scale DNA methylation profiling (Infinium 450k)
2. Selection of target regions
3. Shipment of sample aliquots

Participant 1: Pyrosequencing

Participant 2: MethyLight

...

18 labs, 10 technologies, 27 data sets:
- 16 absolute assays (Figs. 1 and 2)
- 5 relative assays (Fig. 3)
- 6 global assays (Fig. 4)

Results

DNA methylation data	Assay characteristics	Protocol description
Absolute, relative and global measurements	Cost and time of assays performed	Detailed account of experimental workflow

Evaluation

Performance	Discrimination	Scalability	Feasibility
Accuracy, reproducibility	Sensitivity, specificity	Hands-on time, throughput	Cost per sample, setup costs

b

c

d

Figure 1 Technology comparison of locus-specific DNA methylation assays. (**a**) Schematic overview of the study. (**b**) Distribution of measurements obtained with 16 absolute DNA methylation assays, plotted by assay (top), sample type (middle) and genomic region (bottom). Each point corresponds to one measurement for one genomic region in one sample. Point colors indicate assay technology, as defined in the top row. N, normal; T, tumor; D, drug-treated; C, control; F, fresh-frozen; P, FFPE. DNA methylation differences between sample groups were statistically significant ($P < 0.001$ for T/N, for D/C, and for the end points of the two titration series, $P = 0.005$ for F/P; two-tailed paired t-tests on M values). (**c**) Heatmap and hierarchical clustering of the Pearson correlation matrix for all assays across all DNA methylation measurements. (**d**) Scatterplots illustrating the correlation between the Pyroseq 1 assay and other selected assays across all reference samples and genomic regions. Blue lines indicate fitted linear models, and the reported numbers (r) are Pearson correlation coefficients. *, most similar to Pyroseq 1 (within each technology); #, least similar to Pyroseq 1 (across all technologies).

standardized information package comprising general instructions, documentation templates and the list of the 48 target genomic regions (**Supplementary Data 2**). Each region had one designated target CpG for which the DNA methylation level was to be measured, and we asked the contributing research groups to return DNA methylation measurements for each of the reference samples. We gave no further instructions on how to design the assays or how to derive the DNA methylation measurements for the target CpG from the raw data. Moreover, we asked research groups not to exchange any information among each other, and they did not have access to the Infinium 450k data used for region selection.

We designated 16 of the 48 target regions as mandatory ('core regions') and let scientists in each contributing research group for themselves decide how many of the remaining 32 regions they would cover in addition to the core regions. On average, assay design was

attempted for 30 genomic regions and was successful in 95% of cases (**Supplementary Table 1**). The known SNP at one of the target CpGs was detected and reported by 9 of the 17 research groups who contributed locus-specific assays. We removed the SNP-containing region from further analysis to avoid bias, but we emphasize the importance of double-checking for known SNPs during assay design. In total, scientists from 18 laboratories contributed 16 absolute, 5 relative and 6 global assays (**Supplementary Fig. 1**), giving rise to a benchmarking data set with 16,435 locus-specific and 192 global DNA methylation measurements (**Supplementary Data 3**).

Performance of absolute DNA methylation assays

All absolute assays detected the expected bimodal pattern of DNA methylation, with most regions being either highly or lowly methylated (**Fig. 1b**). NGS-based assays (i.e., AmpliconBS and EnrichmentBS)

Table 1 Overview of the evaluated DNA methylation assays

Assay name	Assay description	Assay type	Regions included[a]	Design success[b]	Assay success[c]	Total measurements[d]
AmpliconBS 1	Amplicon bisulfite sequencing with manually prepared PCR amplicons for each region and multiplex sequencing on the medium-throughput MiSeq platform	Absolute	25	100.0%	100.0%	800
AmpliconBS 2	Same as previous, but with sequencing on the high-throughput HiSeq platform	Absolute	15	100.0%	100.0%	480
AmpliconBS 3	Amplicon bisulfite sequencing with automatically prepared PCR amplicons on the AccessArray platform and multiplex sequencing on MiSeq	Absolute	37	86.5%	87.2%	586
AmpliconBS 4	Amplicon bisulfite sequencing using the bisulfite profiling (Bi-PROF) protocol with manually prepared PCR amplicons for each region and multiplex sequencing on MiSeq	Absolute	47	93.6%	99.0%	1,394
EnrichmentBS 1	Enrichment bisulfite sequencing using the RainDrop bisulfite sequencing protocol with massively parallel enrichment on the ThunderStorm platform, followed by sequencing on MiSeq	Absolute	47	93.6%	95.2%	1,340
EnrichmentBS 2	Enrichment bisulfite sequencing using a library-free protocol with bisulfite padlock probes for massively parallel enrichment, followed by sequencing on MiSeq	Absolute	47	91.5%	91.9%	1,265
EpiTyper 1	MassARRAY EpiTYPER protocol with bisulfite conversion, PCR for manually designed amplicons, in vitro transcription, cleavage and mass spectrometric quantification	Absolute	25	100.0%	97.6%	781
EpiTyper 2	Same as previous	Absolute	47	93.6%	97.2%	1,369
EpiTyper 3	Same as previous	Absolute	47	93.6%	99.7%	1,404
Infinium	Genome-scale analysis of DNA methylation using the Infinium 450k microarray	Absolute	47	100.0%	100.0%	846
Pyroseq 1	Bisulfite pyrosequencing of manually prepared PCR amplicons for each region	Absolute	31	100.0%	100.0%	992
Pyroseq 1 (replicate)	Same as previous	Absolute	31	100.0%	100.0%	806
Pyroseq 2	Same as previous	Absolute	15	100.0%	100.0%	480
Pyroseq 3	Same as previous	Absolute	20	95.0%	93.8%	570
Pyroseq 4	Same as previous	Absolute	16	81.2%	97.1%	404
Pyroseq 5	Same as previous	Absolute	34	64.7%	100.0%	572
MethyLight	DNA-methylation-specific amplification by qPCR with a methylation-specific probe	Relative	15	100.0%	100.0%	480
MS-HRM	Non-discriminative qPCR amplification of bisulfite-treated templates followed by DNA-methylation-sensitive high-resolution melting analysis	Relative	15	93.3%	99.8%	447
MS-MCA	Non-discriminative qPCR amplification of bisulfite-treated templates followed by DNA-methylation-sensitive melting curve analysis	Relative	15	100.0%	100.0%	390
qMSP (preamp)	DNA-methylation-specific amplification by qPCR with preamplification for increased sensitivity to low concentrations of methylated DNA	Relative	15	86.7%	100.0%	416
qMSP (standard)	DNA-methylation-specific amplification by qPCR	Relative	15	86.7%	100.0%	416
HPLC-MS	Quantification of total 5mC content using high-performance liquid chromatography combined with electrospray ionization mass spectrometry	Global	n/a	n/a	100.0%	32
Immunoquant	Quantification of total 5mC content using a modified enzyme-linked immunosorbent assay (ELISA) with an antibody against 5-methylcytosine	Global	n/a	n/a	100.0%	32
Pyroseq AluYb8	Bisulfite pyrosequencing using primers that amplify AluYb8 repetitive DNA	Global	n/a	n/a	100.0%	32
Pyroseq D4Z4	Bisulfite pyrosequencing using primers that amplify D4Z4 repetitive DNA	Global	n/a	n/a	100.0%	32
Pyroseq LINE1	Bisulfite pyrosequencing using primers that amplify LINE1 repetitive DNA	Global	n/a	n/a	100.0%	32
Pyroseq NBL2	Bisulfite pyrosequencing using primers that amplify NBL2 repetitive DNA	Global	n/a	n/a	100.0%	32
ClonalBS	Clonal bisulfite sequencing using Sanger technology with a target of 10–20 sequenced clones	Validation	15 + 2[e]	100.0%	100.0%	197

[a]The number of genomic regions for which assay design was attempted. This information does not apply to the global DNA methylation assays, which are not locus-specific and result in a single measurement for each sample. Further details are provided in **Supplementary Figure 1** and **Supplementary Table 1**. [b]The percentage of target genomic regions for which valid measurements were obtained, relative to the number of regions for which assay design was attempted. [c]The percentage of DNA methylation measurements for a given sample and genomic region that were successful and passed all quality control steps. [d]The total number of DNA methylation measurements across samples and genomic regions that were successful and passed all quality control steps. [e]For clonal bisulfite sequencing, two regions were analyzed independently in two laboratories giving rise to two replicates that were compared to each other (**Supplementary Fig. 8**). n/a, not applicable.
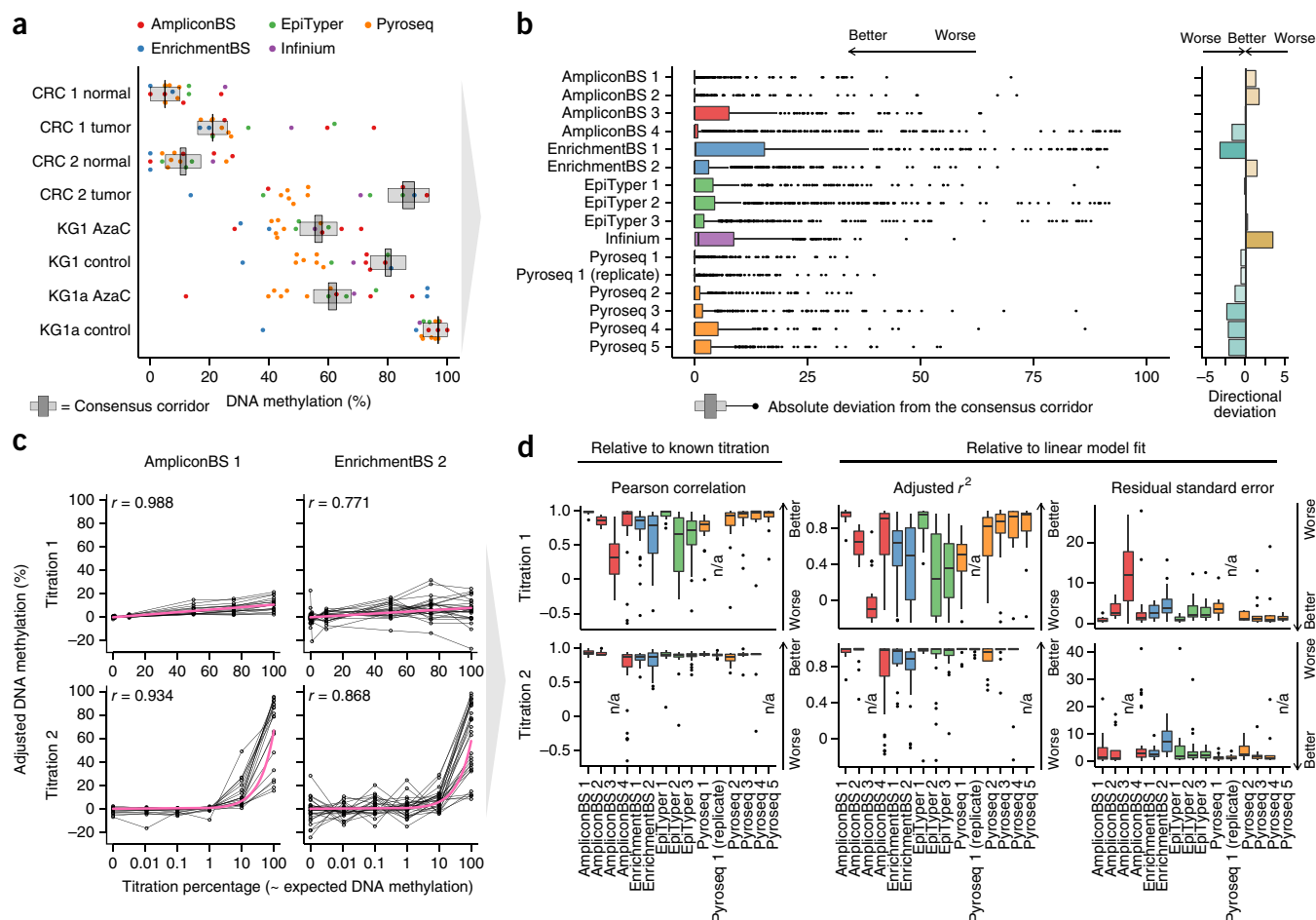
Figure 2 Evaluation of absolute DNA methylation assays. (**a**) Illustration of the consensus corridor as a reference for quantifying assay performance. The true DNA methylation levels of each region in each sample are estimated as the smallest interval that comprises DNA methylation measurements obtained with at least three different assay technologies (inner gray boxes), extended by an uncertainty margin of five percentage points on either side (outer gray boxes). (**b**) Boxplots summarizing the distribution of absolute deviation (left) and directional deviation (right) across samples and regions, comparing each assay's DNA methylation measurements to the closest boundary of the corresponding consensus corridor. Measurements falling inside the corridor were assigned a deviation of zero. (**c**) DNA methylation measurements for the best-performing assay (left) and worst-performing assay (right) in two titration series (top, partially methylated DNA vs. unmethylated DNA; bottom: cancer DNA vs. blood DNA). The measurements (y axis) were adjusted for different basal levels of DNA methylation (**Supplementary Fig. 7**) and plotted against the corresponding titration ratios (x axis). Pink lines represent the slope of the linear model fitted to the data. (**d**) Boxplots summarizing assay performance on the titration series. Shown are Pearson correlation coefficients relative to expected DNA methylation levels (left), and adjusted r^2 values (center) as well as residual standard errors (right) of a fitted linear model that tests for the expected linearity of DNA methylation measurements in the titration series. n/a, not analyzed.

reported extreme values of 0% and 100% more frequently than the other assays, which can be explained by their digital counting of methylated and unmethylated cytosines. The distribution plots confirmed the expected differences among the 32 reference samples (**Fig. 1b**), with higher DNA methylation levels for colon tumors than in matched normal tissue in the target regions, lower DNA methylation in the drug-treated leukemia cell lines, decreasing DNA methylation with decreasing concentrations of *in vitro* methylated DNA (titration 1) and cancer cell line DNA (titration 2), and similar DNA methylation levels for DNA extracted from fresh-frozen vs. FFPE xenografts. These plots also illustrate the broad range of different DNA methylation distributions among the selected target regions (**Fig. 1b**).

To assess global similarity among the absolute DNA methylation assays, we calculated Pearson correlation coefficients across all measurements for each pair of assays (**Supplementary Fig. 2**) and performed hierarchical clustering (**Fig. 1c**). 85% of between-assay comparisons resulted in correlations above 0.8, and 46% even

exceeded 0.9, indicating an excellent overall agreement between many of the tested assays. Correlations were high for technical replicates in the same laboratory (Pyroseq 1 vs. Pyroseq 1 (replicate): $r = 0.996$), the same technology between laboratories (e.g., Pyroseq 1 vs. Pyroseq 2: $r = 0.98$) and between assays of different types (e.g., Pyroseq 1 vs. EpiTyper 3: $r = 0.95$; Pyroseq 1 vs. AmpliconBS 1: $r = 0.97$). However, not all assays agreed equally well (**Fig. 1d** and **Supplementary Fig. 2**). For instance, the Infinium assay reported higher DNA methylation levels for CpGs that other assays identified as lowly methylated, while reporting slightly reduced DNA methylation levels for highly methylated CpGs; and the EnrichmentBS 1 assay gave rise to a substantial number of outliers when compared to any of the other assays.

To quantify the accuracy of individual assays, a reference is needed against which to evaluate the measurements. Synthesized DNA with predefined DNA methylation patterns would be one option, but this is currently feasible only for fully methylated DNA spiked into fully

unmethylated DNA, thus ignoring the challenges posed by heterogeneous DNA methylation patterns[64]. For this reason, we chose two alternative approaches for quantifying assay performance in the presence of epigenetic heterogeneity.

First, we combined data from several assays into high-confidence consensus estimates to establish target DNA methylation levels for the reference samples (**Fig. 2a** and **Supplementary Fig. 3a**). For each sample and genomic region, we identified the smallest interval comprising measurements by at least three of the five technologies (AmpliconBS, EnrichmentBS, EpiTyper, Infinium and Pyroseq), which minimizes the impact of outliers and technology-specific artifacts. Moreover, we extended these intervals with flanking windows of five percentage points on either side to account for small deviations (**Fig. 2a**). We used the resulting 'consensus corridor' as a surrogate for the true DNA methylation level (which is unknown) of each target CpG in each reference sample. All assays contributed to the consensus corridor (**Supplementary Fig. 3b,c**), and sensitivity analysis confirmed that the ranking of assay performance was robust to the exact definition of the consensus corridor (**Supplementary Note** and **Supplementary Fig. 4**).

Evaluating each assay against this corridor (**Fig. 2b** and **Supplementary Table 1**), we observed the lowest mean absolute deviation for Pyroseq 1 and its replicate (1.1 and 1.2, respectively), closely followed by AmpliconBS 1 (1.6), Pyroseq 2 (2.5), AmpliconBS 2 (2.8) and Pyroseq 5 (3.3). The highest mean absolute deviation was observed for EpiTyper 2 and EnrichmentBS 1 (6.8 and 11.4, respectively). We also assessed the bias of each assay, which we defined as the directional (rather than absolute) deviation from the consensus corridor. For the Infinium assay we observed an overall tendency to over-estimate DNA methylation levels, whereas all Pyroseq assays tended to underestimate DNA methylation levels. For AmpliconBS EnrichmentBS, and EpiTyper, the average direction of the deviation depended on the laboratory.

Furthermore, to understand which properties make genomic regions difficult to measure, we fitted a linear model that predicts the deviation from the consensus based on each region's estimated DNA methylation level, GC content, CpG observed vs. expected ratio and content of repetitive DNA (**Supplementary Fig. 5**). Four assays (AmpliconBS 4, EnrichmentBS 1, Pyroseq 4 and Pyroseq 5) showed significantly increased deviation in highly methylated regions, whereas the Infinium assay was comparably more accurate in highly methylated regions. GC content, CpG density and repetitive DNA also affected the deviation in some cases, but the best-performing assays did not show any significant biases (**Supplementary Fig. 5**). Finally, we compared assay performance between matched fresh-frozen and FFPE samples (**Supplementary Fig. 6**). We obtained highly similar results, showing that all tested assays are compatible with DNA from FFPE material.

Second, as a complementary approach to consensus corridors, we assessed the performance of each assay using two titration series with known ratios. The titration samples included heterogeneous DNA methylation patterns, which make them more challenging than titrations of fully methylated DNA used for assay calibration. For the first titration series, we created partially and heterogeneously methylated DNA by incomplete *in vitro* methylation (less than 20% methylated cytosines) and combined it with unmethylated DNA at ratios of 100%, 75%, 50%, 10%, 1% and 0%. The second titration series mimics the diagnostic task of detecting hypermethylated cancer DNA against a background of blood-derived DNA. To that end, we spiked DNA from a colon cancer cell line (HCT-116) at ratios of 100%, 10%, 1%, 0.1%, 0.01% and 0% into DNA extracted from whole blood. We then fitted linear models to the DNA methylation measurements (**Fig. 2c**

and **Supplementary Fig. 7**) to assess consistency with the known titration ratios. The agreement was high for most assays and for three alternative metrics (**Fig. 2d** and **Supplementary Table 1**). Best results were achieved by AmpliconBS 1 with median Pearson correlation coefficients of 0.99 and 0.93 in the two titration series (**Fig. 2c**). By contrast, EnrichmentBS 2 is an example of an assay that showed more variability with correlation of coefficients 0.77 and 0.87 (**Fig. 2c**). The results for the titration series were in good agreement with the assay performance in the consensus-based validation (**Fig. 2b**), with AmpliconBS 1, AmpliconBS 2, Pyroseq 1 and Pyroseq 3 being among the best in both analyses.

Finally, we assessed how clonal bisulfite sequencing[59,60] would fare in our benchmarking, given that it was previously considered the gold standard for locus-specific DNA methylation mapping. At a target coverage of 10–20 Sanger sequencing clones, fully unmethylated and fully methylated CpGs gave rise to consistent measurements between replicates, but regions with intermediate DNA methylation levels agreed less well (**Supplementary Fig. 8a**). Diverging measurements appeared to be caused by random noise resulting from sequencing few clones, and both replicates clustered similarly well with other assays (Pearson correlation above 0.9 for all but one assay; **Supplementary Fig. 8b**). We did not observe any directional deviation from the consensus corridor (**Supplementary Fig. 8c**), and Pearson correlation coefficients in comparison to other assays were in the range of 0.7 to 0.9 in comparison to other assays (**Supplementary Fig. 8d**). Overall, clonal bisulfite sequencing performed reasonably well in our analysis but did not reach the accuracy and reproducibility of the top-ranking assays.

Performance of relative DNA methylation assays

Relative DNA methylation assays detect DNA molecules with a predefined DNA methylation pattern, e.g., identifying fully methylated, tumor-derived DNA fragments in an excess of blood DNA. This approach is less suited for measuring quantitative DNA methylation levels at single-CpG resolution, which prompted two of the research groups contributing relative assays to report their measurements as ranges (e.g., 0.1–1%, 1–10%, or 10–100%). Indeed, the observed correlations were generally lower for relative assays than for absolute assays, ranging from 0.5 to 0.7 in most comparisons (**Supplementary Fig. 9**).

To benchmark the relative assays in a way that accounts for their strengths and characteristics, we assessed their ability to detect differences in DNA methylation between pairs of samples. For each assay and each pairwise comparison we discretized the measurements into three categories ('+', higher DNA methylation in first sample; '−', lower DNA methylation in first sample and '=', no detectable difference) and calculated the agreement between the different assays (**Fig. 3a**). Percent concordance values were high for closely related assays (68% for MS-HRM vs. MS-MCA and 75% for qMSP preamp vs. qMSP standard), but substantially lower between technologies (**Fig. 3b**). Furthermore, when we compared these results with the concordance values observed for absolute assays using suitable thresholds (Online Methods), the relative assays were as similar to some of the absolute assays as they were to each other, but they did not reach the high concordance (90% and above) observed among the best-performing absolute assays (**Fig. 3b**).

We also assessed the discriminatory power of the relative assays for DNA methylation differences identified by the consensus corridor, and for the known ratios in the two titration series (**Fig. 3c**). In these analyses, all relative assays accurately detected DNA methylation differences that exceeded 25%, whereas the performance for smaller differences varied between assays. MethyLight and qMSP
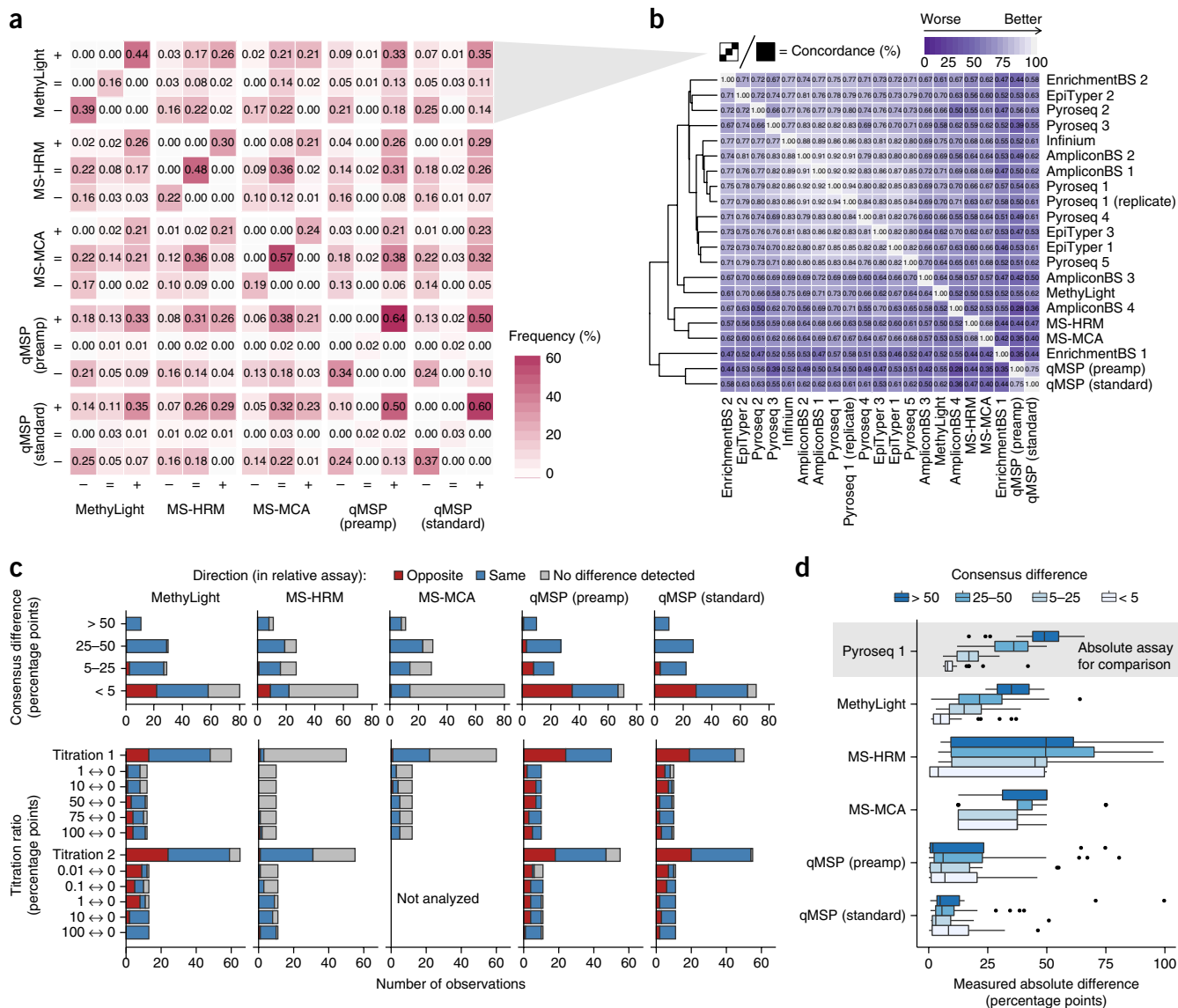
Figure 3 Evaluation of relative DNA methylation assays. (**a**) Pairwise contingency tables for DNA methylation differences detected by the relative DNA methylation assays ('+', higher; '−' lower; '=',: similar). High values in the diagonal of any given 3 × 3 matrix are indicative of strong agreement between the corresponding pair of assays. (**b**) Clustered heatmap of the concordance (i.e., the sum of the diagonal divided by the sum of the 3 × 3 matrix) between all pairs of relative and absolute assays. Lighter colors indicate higher concordance. (**c**) Frequency of detecting the expected direction of DNA methylation difference for the relative assays evaluated against consensus corridor estimates (top) and the titration series (bottom). (**d**) Boxplots showing the quantitative differences reported by each relative assay (and by Pyroseq 1 as an example of an absolute assay), grouped by the consensus difference and focusing only on those genomic regions for which the direction was correctly identified by the relative assay (corresponding to blue bars in **c**).

detected many small differences with the expected direction, but reported the opposite direction for a considerable number of genomic regions. By contrast, MS-MCA and MS-HRM were more conservative and yielded very few mistakes regarding direction, but at the cost of detecting fewer differences.
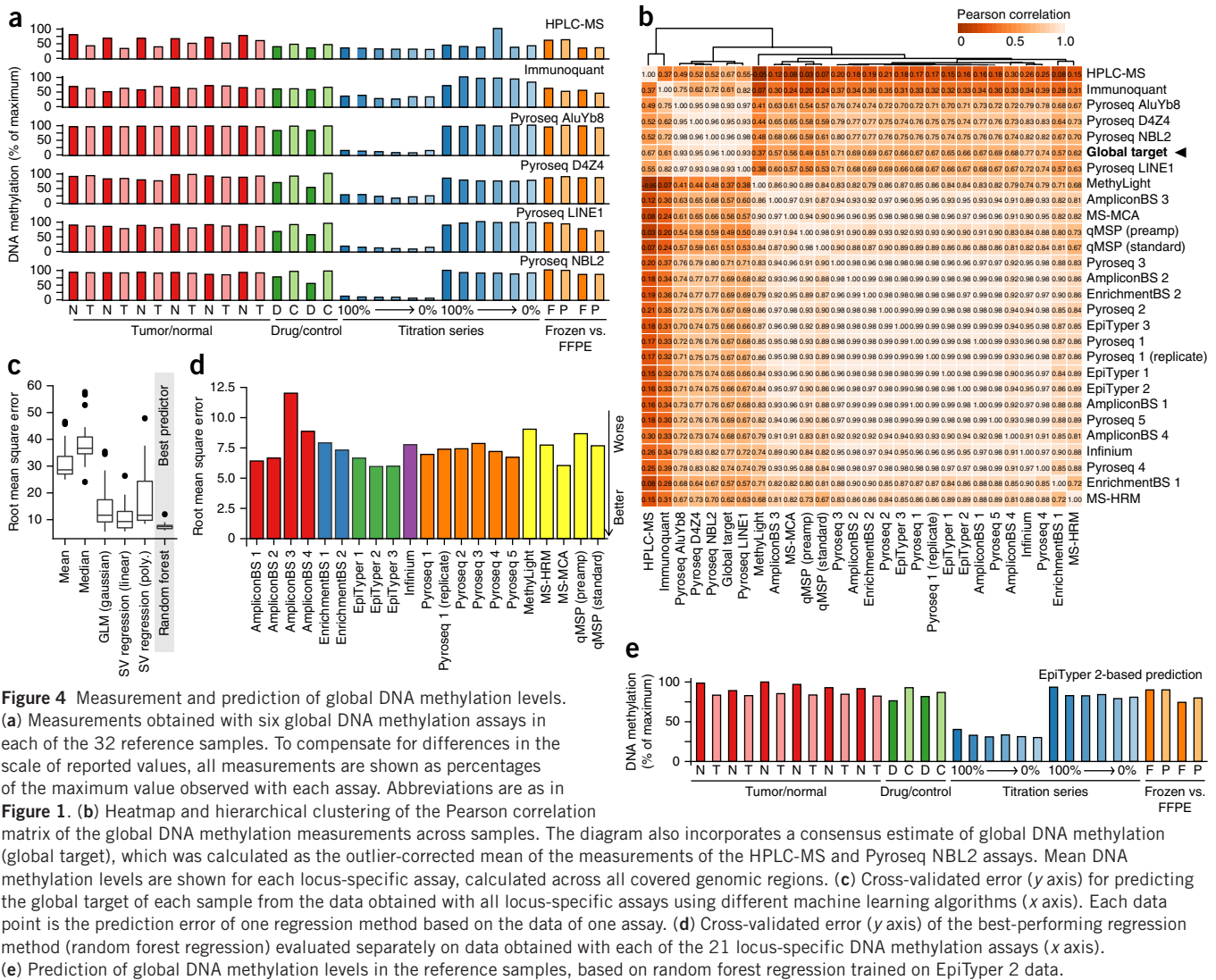
Finally, we asked how well the relative DNA methylation assays captured quantitative differences in DNA methylation between samples. To that end, we took the quantitative differences reported by the relative assays for regions that were correctly classified and plotted them against the difference in consensus corridor estimates (**Fig. 3d**). The differences in the consensus corridor were most accurately recapitulated by the MethyLight assay. By contrast, the measurements of the other relative assays did not correlate well with the difference obtained from the consensus corridor, supporting the notion that MS-MCA,

MS-HRM and qMSP should only be used for the type of qualitative comparisons that they were originally developed for.

Finally, given that sample material is scarce in some applications of DNA methylation biomarkers (e.g., needle biopsies and blood-plasma-derived DNA), we also analyzed the effect of input DNA amounts on assay performance. This analysis confirmed that DNA amounts were not limiting the assay performance in the main part of our comparison, but only the AmpliconBS and Pyroseq technologies were able to cope with severely reduced amounts and/or high fragmentation of input DNA (**Supplementary Note** and **Supplementary Figs. 10–12**).

Performance of global DNA methylation assays

Global DNA methylation assays report a single measurement value for each sample, indicative of its total DNA methylation content (**Fig. 4a**).

Figure 4 Measurement and prediction of global DNA methylation levels. (**a**) Measurements obtained with six global DNA methylation assays in each of the 32 reference samples. To compensate for differences in the scale of reported values, all measurements are shown as percentages of the maximum value observed with each assay. Abbreviations are as in **Figure 1**. (**b**) Heatmap and hierarchical clustering of the Pearson correlation matrix of the global DNA methylation measurements across samples. The diagram also incorporates a consensus estimate of global DNA methylation (global target), which was calculated as the outlier-corrected mean of the measurements of the HPLC-MS and Pyroseq NBL2 assays. Mean DNA methylation levels are shown for each locus-specific assay, calculated across all covered genomic regions. (**c**) Cross-validated error (y axis) for predicting the global target of each sample from the data obtained with all locus-specific assays using different machine learning algorithms (x axis). Each data point is the prediction error of one regression method based on the data of one assay. (**d**) Cross-validated error (y axis) of the best-performing regression method (random forest regression) evaluated separately on data obtained with each of the 21 locus-specific DNA methylation assays (x axis). (**e**) Prediction of global DNA methylation levels in the reference samples, based on random forest regression trained on EpiTyper 2 data.

For HPLC-MS, the results were generally consistent with expectations, showing global hypomethylation for the tumor samples (as opposed to locus-specific hypermethylation in the target regions of the absolute and relative assays) and for the drug-treated cell lines (**Fig. 4a**), similar values for fresh-frozen and FFPE samples from the same xenograft, and gradually decreasing DNA methylation from left to right in the two titration series (with relatively small differences and one strong outlier). The data for the Immunoquant assay did not agree well with the characteristics of the reference samples. Finally, the pyrosequencing assays, which focused on different types of repetitive elements (AluYb8, D4Z4, LINE1 and NBL2), detected the drug-induced demethylation and, to a lesser extent, also the tumor hypomethylation (LINE1 and NBL2) and the differences in the titration series (NBL2).

With correlations of 0.37 to 0.82 between the three technologies (**Fig. 4b**), there was less agreement among the global DNA methylation assays than we had observed for the locus-specific DNA methylation assays. This result prompted us to explore whether global DNA methylation levels could be inferred from locus-specific data, as a potential alternative to measuring them with global assays. We defined the 'global target' as the outlier-corrected mean of the two best-performing global assays (HPLC-MS and Pyroseq NBL2), and

we tested several approaches for predicting the sample-specific global target values from the locus-specific data. Averaging across locus-specific measurements did not provide an accurate prediction (correlations of 0.37 to 0.77, **Fig. 4b**), likely because the target regions were enriched for regulatory elements with different DNA methylation dynamics compared to the bulk of the genome. By contrast, machine learning methods such as the generalized linear model, support vector regression and random forest regression compensated for these differences and predicted the global target values much more accurately (**Fig. 4c,d**). These results suggest that locus-specific assays in combination with statistical methods can be used to detect sample-specific differences in global DNA methylation (**Fig. 4e**).

Accuracy and robustness as epigenetic biomarkers

Epigenetic biomarker development is an important application of DNA methylation assays, requiring robust discrimination between cell types or disease states. We observed good separation between the different cell types using unsupervised methods (**Supplementary Fig. 13**), and we sought to quantify the assays' discriminatory power by supervised analysis focusing on the colon tumor and adjacent normal samples (**Fig. 5**). To that end, we trained support vector machines to distinguish between tumor and normal samples based on the data
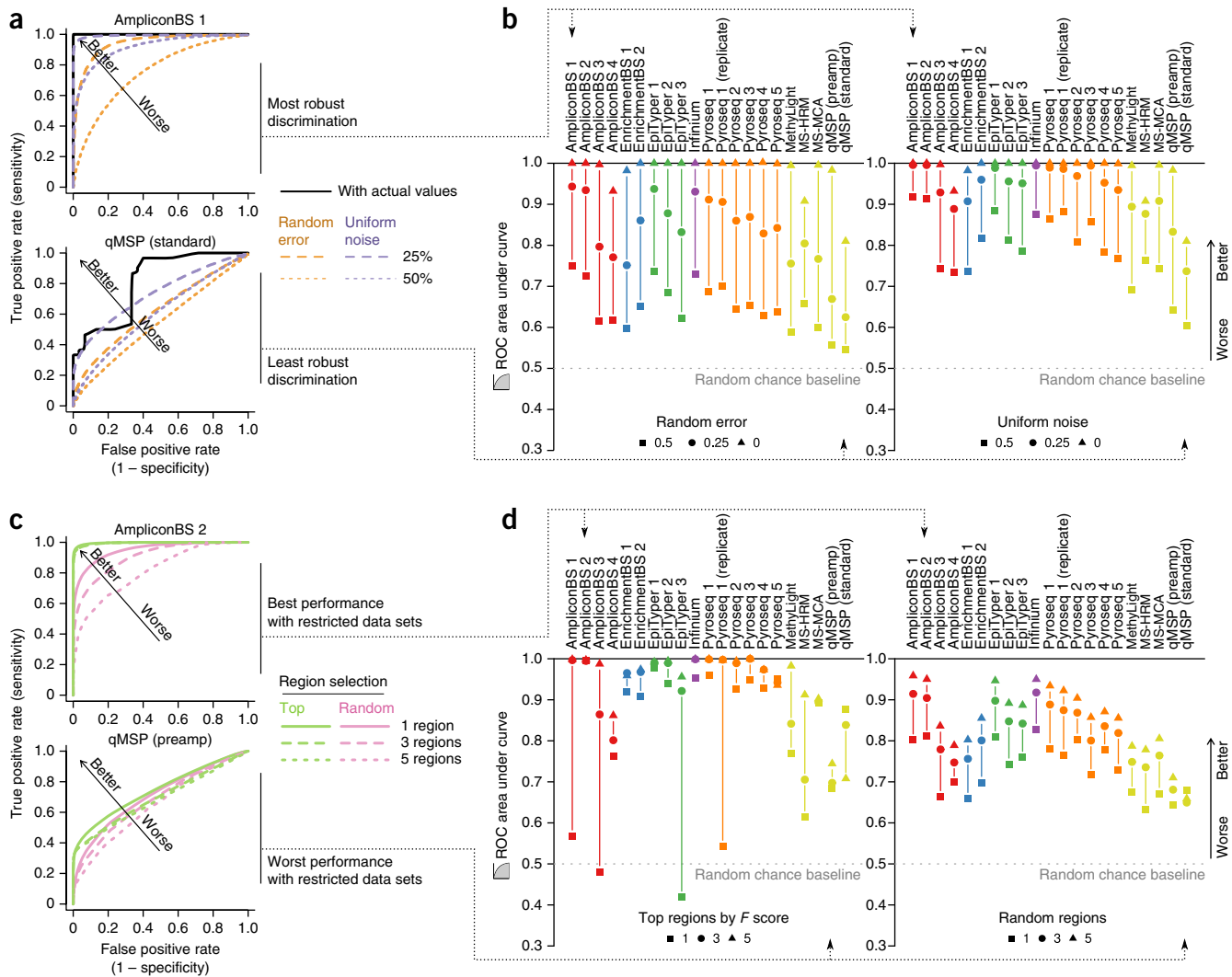
Figure 5 Discriminatory power of locus-specific DNA methylation assays. (**a**) Cross-validated performance of support vector machines discriminating between tumor and adjacent normal colon samples, based on DNA methylation values obtained with the best-performing assay (top) and worst-performing assay (bottom). Shown are receiver operating characteristic (ROC) curves for classifiers trained on the actual data and on artificially noisy data (random error: a defined percentage of measurements was replaced by random values; uniform noise: uniformly distributed random values between zero and a defined percentage were added or subtracted). (**b**) Cross-validated prediction performance (measured as the ROC area under curve) for support vector machines trained and evaluated separately based on data from each assay. The same error models were used as in **a**, and the performance was calculated separately for random error (left) and uniform noise (right). ROC curves are shown in **Supplementary Figure 14a**. (**c**) Cross-validated prediction performance after restricting the data set to the most informative regions (top) or to randomly selected regions (random). ROC curves are shown for the best-performing assay (top) and worst-performing assay (bottom) at a constant level of 25% uniform noise. (**d**) Cross-validated ROC area under curve for each assay after restricting the data set to the most informative (left) or to randomly selected (right) regions. ROC curves are shown in **Supplementary Figure 14b**.

of each assay. We used four tumor-normal pairs for training, and evaluated the prediction performance on test sets consisting of the two remaining pairs, constituting a threefold cross-validation. Receiver operating characteristic (ROC) curves show excellent prediction performance for most assays (**Fig. 5a** and **Supplementary Fig. 14a**), which is not unexpected because DNA methylation patterns are known to be different between colon tumor and adjacent normal tissue, and because we selected several target regions based on their differential DNA methylation in colon cancer.

To simulate the complications of working with clinical samples of varying quality, we added noise to the data and assessed how the prediction performance was affected. Two types of noise were introduced (Online Methods): erroneous measurements were simulated

by randomly replacing a fraction of DNA methylation measurements with other measurements (random error), and inaccurate measurements were simulated by adding random noise to each measurement (uniform noise) (**Fig. 5a,b**). Using the ROC area under curve to assess performance, we found that low noise levels were well tolerated by all assays, but at higher noise levels some of the absolute assays (AmpliconBS 1, AmpliconBS 2, EpiTyper 1, EpiTyper 2 and Pyroseq 1) outperformed the other assays.

We also assessed the effect of reducing the number of genomic regions contributing to the analysis (**Supplementary Fig. 14b**). When we trained and evaluated each classifier on the one, three or five most discriminatory genomic regions at a constant level of 25% uniform noise (**Fig. 5c,d**), the prediction accuracy remained high for most

assays (in some cases it even increased because the removal of less informative regions reduced noise in the data set). However, when we repeated the analysis with randomly sampled regions (**Fig. 5c,d**), we observed major differences: for many of those assays that performed well at high noise levels (**Fig. 5b**), the drop in performance was low, and satisfactory results were achieved using just three randomly selected regions. By contrast, other assays (in particular the relative DNA methylation assays) performed well only with combined data for many different genomic regions.

Finally, we selected two assays (EpiTyper 3 and Infinium) and tested their predictive performance on a much larger cohort comprising 160 prostate tumor and 8 normal prostate samples. The concordance between the two assays was high for nine of ten tested CpGs, with Pearson correlations of 0.75 to 0.93, whereas one CpG showed little biological variation within the cohort and a low correlation coefficient (**Supplementary Fig. 15a**). When we trained and evaluated support vector machines for distinguishing between tumor and normal samples, we observed higher accuracy using the EpiTyper data than for the Infinium data, indicating that the locus-specific assays outperforms the Infinium assay in terms of accuracy and discriminatory power (**Supplementary Fig. 15b**).

DISCUSSION

We conducted a multicenter benchmarking study evaluating all DNA methylation assays that are strong candidates for clinical use. Most assays proved to be accurate and reproducible. The results also agreed well between laboratories and between technologies, which is notable because assay design (e.g., selection of primer sites and protocol parameters), execution (e.g., bisulfite conversion and sequencing) and data processing (e.g., normalization and quality control) were conducted independently by members of each contributing laboratory. Overall, our study demonstrates that locus-specific DNA methylation assays can be considered a mature technology ready for widespread use in biomarker development and clinical applications.

Despite generally consistent results, we observed characteristic strengths and weaknesses of the tested assays. The relative assays were generally less accurate and less concordant with each other than the absolute assays. This observation is not unexpected given that relative assays work best for detecting fully methylated regions, whereas many of the selected target regions were heterogeneously methylated. Despite their lower quantitative accuracy, the relative assays distinguished robustly between methylated and unmethylated regions, and they discriminated well between tumor and normal samples. Among the global assays, the HPLC-MS measurements most accurately reflected the expected differences in global DNA methylation levels, whereas the Immunoquant assay did not provide reliable results. Bisulfite pyrosequencing of repetitive DNA gave rise to highly reproducible results, but these repetitive DNA methylation levels did not correlate well with the expected differences in global DNA methylation. By contrast, good results were obtained when predicting global DNA methylation from locus-specific measurements, which may become a viable alternative to measuring global DNA methylation directly.

To capture not only the quantitative performance but also other relevant aspects of each assay, members of the contributing laboratories wrote detailed reports (**Supplementary Data 1**). These reports include protocol descriptions, comments on the practical strengths and limitations of each assay, and detailed time and cost calculations for running the assays in the respective laboratories. Drawing upon the cumulative experiences of our study, we arrive at the following conclusions and recommendations.

(i) Absolute DNA methylation assays are the method of choice when validating DNA methylation differences in large cohorts, and they are also an excellent technology for developing epigenetic biomarkers. (ii) Relative DNA methylation assays are not a good replacement for absolute assays. However, experiences of scientists in the contributing laboratories suggest that carefully selected, designed and validated relative assays can cost-effectively detect minimal traces of methylated DNA against an excess of unmethylated DNA. (iii) Global DNA methylation assays suffer from noisy data and divergent results between technologies. Locus-specific assays (possibly combined with prediction) provide a more robust alternative. (iv) Among the absolute DNA methylation assays, AmpliconBS and Pyroseq showed the best all-round performance, closely followed by EpiTyper. AmpliconBS is the best choice for assaying dozens of genomics regions in parallel, EpiTyper provides the highest sample throughput, and Pyroseq can work well even on minute amounts of highly fragmented DNA. (v) EnrichmentBS and Infinium can measure many more CpGs simultaneously than the other tested assays, but this comes at the cost of lower accuracy and higher cost per sample. (vi) Clonal bisulfite sequencing suffers from a high level of technical noise when sequencing 10–20 clones per sample. Given its high labor intensity and the availability of alternate assays with equal or better performance (as demonstrated in this study), clonal bisulfite sequencing is not recommended for large-scale validation and biomarker development.

We conclude that the accuracy and robustness, discriminatory power, cost structure and practical feasibility of current DNA methylation assays are sufficient for large-scale validation studies and epigenetic biomarker development. We expect that DNA methylation assays will become widely useful for clinical diagnostics and personalized therapies, as companion diagnostics of targeted drugs, in forensic testing of tissue types and in many other applications. Our study may serve as a starting point for broader standardization efforts involving academic and clinical laboratories as well as the commercial sector and regulatory agencies, to fully embrace the potential of DNA methylation biomarkers for precision medicine.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. GEO: GSE77965.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHORS' CONTRIBUTIONS
C.B. designed and coordinated the study. F.J.C., S.T. and P.D. prepared and characterized the reference samples. Y.A., M.B., A.K.B., K.B., F.B., M.C., C.D., C.M.D., D.D., A.F.F., C.G., A.H., K.H., T.H., D.H., M.I., R.K., M.Kr., M.Ku., V.L., S.S.N., D.S.P., N.P., W.Q., A.C.Q., F.R., A.S., C.S., R.St., S.T., R.G.U., K.W., D.W. and D.J.W. performed the experiments and prepared assay-specific analysis reports. S.B., S.J.C., M.E., A.C.F.-S., M.F.F., P.G., L.L.H., P.W.L., J.I.M.-S., A.O.H.N., R.P.,

C.P., D.S.S., R.Si., X.S., J.T., J.W. and K.Z. contributed to the analysis reports and supervised the work. G.S. and T.S. provided the prostate cancer samples. C.B. and F.H. performed integrative data analysis, and drafted the manuscript with contributions from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Baylin, S.B. & Jones, P.A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).
2. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068 (2010).
3. Feil, R. & Fraga, M.F. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* **13**, 97–109 (2011).
4. Benayoun, B.A., Pollina, E.A. & Brunet, A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell Biol.* **16**, 593–610 (2015).
5. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
6. Laird, P.W. The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* **3**, 253–266 (2003).
7. Bock, C. Epigenetic biomarker development. *Epigenomics* **1**, 99–110 (2009).
8. Claus, R. *et al.* Validation of ZAP-70 methylation and its relative significance in predicting outcome in chronic lymphocytic leukemia. *Blood* **124**, 42–48 (2014).
9. deVos, T. *et al.* Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clin. Chem.* **55**, 1337–1346 (2009).
10. Ebert, M.P. *et al.* TFAP2E-DKK4 and chemoresistance in colorectal cancer. *N. Engl. J. Med.* **366**, 44–53 (2012).
11. Fernandez, A.F. *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res.* **22**, 407–419 (2012).
12. Hegi, M.E. *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **352**, 997–1003 (2005).
13. Mahon, K.L. *et al.* & PRIMe consortium. Methylated Glutathione S-transferase 1 (mGSTP1) is a potential plasma free DNA epigenetic marker of prognosis and response to chemotherapy in castrate-resistant prostate cancer. *Br. J. Cancer* **111**, 1802–1809 (2014).
14. Queirós, A.C. *et al.* A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* **29**, 598–605 (2015).
15. Papageorgiou, E.A. *et al.* Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21. *Nat. Med.* **17**, 510–513 (2011).
16. Bock, C. *et al.* Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).
17. Park, J.L. *et al.* Identification of body fluid-specific DNA methylation markers for use in forensic science. *Forensic Sci. Int. Genet.* **13**, 147–153 (2014).
18. Yi, S.H., Xu, L.C., Mei, K., Yang, R.Z. & Huang, D.X. Isolation and identification of age-related DNA methylation markers for forensic age-prediction. *Forensic Sci. Int. Genet.* **11**, 117–125 (2014).
19. Bock, C. Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* **13**, 705–719 (2012).
20. Laird, P.W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **11**, 191–203 (2010).
21. Rakyan, V.K., Down, T.A., Balding, D.J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12**, 529–541 (2011).
22. Michels, K.B. *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat. Methods* **10**, 949–955 (2013).
23. Kristensen, L.S. & Hansen, L.L. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin. Chem.* **55**, 1471–1483 (2009).
24. Mikeska, T., Bock, C., Do, H. & Dobrovic, A. DNA methylation biomarkers in cancer: progress towards clinical implementation. *Expert Rev. Mol. Diagn.* **12**, 473–487 (2012).
25. Umer, M. & Herceg, Z. Deciphering the epigenetic code: an overview of DNA methylation analysis methods. *Antioxid. Redox Signal.* **18**, 1972–1986 (2013).
26. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
27. Satterlee, J.S., Schübeler, D. & Ng, H.H. Tackling the epigenome: challenges and opportunities for collaboration. *Nat. Biotechnol.* **28**, 1039–1044 (2010).
28. Beck, S. Taking the measure of the methylome. *Nat. Biotechnol.* **28**, 1026–1028 (2010).
29. Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.* **28**, 1106–1114 (2010).
30. Harris, R.A. *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.* **28**, 1097–1105 (2010).
31. Gries, J. *et al.* Bi-PROF: bisulfite profiling of target regions using 454 GS FLX Titanium technology. *Epigenetics* **8**, 765–771 (2013).
32. Masser, D.R., Berg, A.S. & Freeman, W.M. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics Chromatin* **6**, 33 (2013).
33. Varley, K.E. & Mitra, R.D. Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. *Genome Res.* **20**, 1279–1287 (2010).
34. Diep, D. *et al.* Library-free methylation sequencing with bisulfite padlock probes. *Nat. Methods* **9**, 270–272 (2012).
35. Paul, D.S. *et al.* Assessment of RainDrop BS-seq as a method for large-scale, targeted bisulfite sequencing. *Epigenetics* **9**, 678–684 (2014).
36. Coolen, M.W., Statham, A.L., Gardiner-Garden, M. & Clark, S.J. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res.* **35**, e119 (2007).
37. Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **102**, 15785–15790 (2005).
38. Ehrich, M. *et al.* Cytosine methylation profiling of cancer cell lines. *Proc. Natl. Acad. Sci. USA* **105**, 4844–4849 (2008).
39. Ronaghi, M., Uhlén, M. & Nyrén, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
40. Tost, J., Dunker, J. & Gut, I.G. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* **35**, 152–156 (2003).
41. Tost, J. & Gut, I.G. DNA methylation analysis by pyrosequencing. *Nat. Protoc.* **2**, 2265–2275 (2007).
42. Campan, M., Weisenberger, D.J., Trinh, B. & Laird, P.W. MethyLight. *Methods Mol. Biol.* **507**, 325–337 (2009).
43. Eads, C.A. *et al.* MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* **28**, E32 (2000).
44. Guldberg, P., Worm, J. & Grønbaek, K. Profiling DNA methylation by melting analysis. *Methods* **27**, 121–127 (2002).
45. Wojdacz, T.K., Dobrovic, A. & Hansen, L.L. Methylation-sensitive high-resolution melting. *Nat. Protoc.* **3**, 1903–1908 (2008).
46. Worm, J., Aggerholm, A. & Guldberg, P. In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin. Chem.* **47**, 1183–1189 (2001).
47. Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D. & Baylin, S.B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. USA* **93**, 9821–9826 (1996).
48. Walter, K. *et al.* DNA methylation profiling defines clinically relevant biological subsets of non-small cell lung cancer. *Clin. Cancer Res.* **18**, 2360–2373 (2012).
49. Ehrich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239–259 (2009).
50. Aparicio, A. *et al.* LINE-1 methylation in plasma DNA as a biomarker of activity of DNA methylation inhibitors in patients with solid tumors. *Epigenetics* **4**, 176–184 (2009).
51. Berdasco, M., Fraga, M.F. & Esteller, M. Quantification of global DNA methylation by capillary electrophoresis and mass spectrometry. *Methods Mol. Biol.* **507**, 23–34 (2009).
52. Kremer, D., Metzger, S., Kolb-Bachofen, V. & Kremer, D. Quantitative measurement of genome-wide DNA methylation by a reliable and cost-efficient enzyme-linked immunosorbent assay technique. *Anal. Biochem.* **422**, 74–78 (2012).
53. Bollati, V. *et al.* Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res.* **67**, 876–880 (2007).
54. Choi, S.H. *et al.* Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *Int. J. Cancer* **125**, 723–729 (2009).
55. Martínez, J.G. *et al.* Hypomethylation of LINE-1, and not centromeric SAT-α, is associated with centromeric instability in head and neck squamous cell carcinoma. *Cell Oncol. (Dordr.)* **35**, 259–267 (2012).
56. Yang, A.S. *et al.* A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res.* **32**, e38 (2004).
57. Shanmuganathan, R. *et al.* Conventional and nanotechniques for DNA methylation profiling. *J. Mol. Diagn.* **15**, 17–26 (2013).
58. Bibikova, M. *et al.* Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* **1**, 177–200 (2009).
59. Clark, S.J., Harrison, J., Paul, C.L. & Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**, 2990–2997 (1994).
60. Clark, S.J., Statham, A., Stirzaker, C., Molloy, P.L. & Frommer, M. DNA methylation: bisulphite modification and analysis. *Nat. Protoc.* **1**, 2353–2364 (2006).
61. Booth, M.J. *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* **8**, 1841–1851 (2013).
62. Lu, X. *et al.* Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J. Am. Chem. Soc.* **135**, 9315–9317 (2013).
63. Song, C.X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
64. Mikeska, T., Candiloro, I.L. & Dobrovic, A. The implications of heterogeneous DNA methylation for the accurate quantification of methylation. *Epigenomics* **2**, 561–573 (2010).

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. [2]Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria. [3]Max Planck Institute for Informatics, Saarbrücken, Germany. [4]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain. [5]Department of Genetics/Epigenetics, FR8.3 Life Sciences, Saarland University, Saarbrücken, Germany. [6]Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany. [7]Institute of Human Genetics, Christian-Albrechts-University Kiel & University Hospital Schleswig-Holstein, Kiel, Germany. [8]Department of General Pediatrics, University Hospital Schleswig-Holstein, Kiel, Germany. [9]Zymo Research Corporation, Irvine, California, USA. [10]Laboratory for Epigenetics and Environment, Centre National de Génotypage, CEA-Institut de Génomique, Evry, France. [11]USC Epigenome Center, University of Southern California, Los Angeles, California, USA. [12]Danish Cancer Society Research Center, Copenhagen, Denmark. [13]Department of Bioengineering, Institute for Genomic Medicine and Institute of Engineering in Medicine, University of California at San Diego, La Jolla, California, USA. [14]Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA), HUCA, Universidad de Oviedo, Oviedo, Spain. [15]Department of Immunology and Oncology, National Center for Biotechnology, CNB-CSIC, Cantoblanco, Madrid, Spain. [16]Nanomaterials and Nanotechnology Research Center (CINN-CSIC), El Entrego, Spain. [17]Genentech Inc., Oncology Biomarker Development, South San Francisco, California, USA. [18]Department of Biomedicine, Aarhus University, Aarhus C, Denmark, Copenhagen. [19]Department of Genetics, University of Cambridge, Downing Site, Cambridge, United Kingdom. [20]QIAGEN GmbH, Sample and Assay Technologies, Hilden, Germany. [21]Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Department of Anatomic Pathology, Pharmacology and Microbiology, University of Barcelona, Barcelona, Spain. [22]Epigenetics Laboratory, Genomics and Epigenetics Division, Garvan Institute of Medical Research, Darlinghurst, New South Wales, Australia. [23]St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales, New South Wales, Australia. [24]UCL Cancer Institute, University College London, London, United Kingdom. [25]Martini-Clinic, Prostate Cancer Center & Department of Urology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [26]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain. [27]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain. [28]Van Andel Research Institute, Grand Rapids, Michigan, USA. [29]Agena Bioscience, San Diego, California, USA. [30]Institute of Human Genetics, University Hospital of Ulm, Ulm, Germany. Correspondence should be addressed to C.B. (cbock@cemm.oeaw.ac.at).

ONLINE METHODS

Preparation of reference DNA samples. Six pairs of fresh-frozen colon tumor and adjacent normal colon tissue samples were obtained from the IDIBELL Tissue Biobank following approval by the corresponding ethics committee. For DNA extraction, frozen tumors were incubated overnight at 37 °C with DNA lysis buffer (10 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA pH 8.0, 10% SDS) and proteinase K, and phenol-chloroform extraction was carried out.

KG1 (ref. 65) and KG1a[66] leukemia cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). The cells were cultured in RPMI medium (Sigma-Aldrich) supplemented with 15% FCS (PAA) and 1% penicillin/streptomycin (Sigma-Aldrich) at 37 °C and 5% CO_2. DNA demethylation was induced by treatment with 5-aza-2′-deoxycytidine (Sigma-Aldrich) at the half-maximal inhibitory concentration (IC50) determined by MTT viability tests for each cell line (80 nM for KG1 and 300 nM for KG1a). Stocks were prepared fresh for each experiment in 0.1% phosphate-buffer-saline and kept at −20 °C until use. KG1 and KG1a cell lines where seeded in 75 cm^2 tissue culture flasks (Sarstedt) at a concentration of 2.5 × 10^5 cells/ml 24 h before treatment. Cells were treated for three days at 24-h intervals. Every 24 h, medium change was performed by centrifugation of cells at 900 r.p.m. for 5 min and resuspension of cell pellet in fresh warm medium. Cells were allowed to recover for 1 h before addition of the drug. Cells where harvested by centrifugation 96 h after seeding, and the cell pellet was washed twice with 1% PBS.

For the first titration series, the REPLI-g Whole Genome Amplification kit (Qiagen) was used to produce unmethylated DNA following the manufacturer's protocol. Part of the whole genome amplified DNA was incompletely methylated *in vitro* using M.SssI (New England BioLabs) such that ~20% of all cytosines became methylated. For the second titration series, DNA was extracted from HCT-116 cells grown *in vitro* (as described below) and from whole blood of a healthy donor using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich) according to the manufacturer's protocol.

For the comparison of fresh-frozen and FFPE material, xenografts were prepared as follows. Two colon cancer cell lines (HCT-15, derived from a Dukes' type C colorectal adenocarcinoma, and HCT-116, derived from a colorectal carcinoma, both obtained from ATCC) were grown in DMEM with 10% FBS supplemented with antibiotic (37 °C, 5% CO_2). Cells were harvested, filtered and aliquoted in PBS, and two SCID mice were subcutaneously injected in the flank with 3 × 10^6 tumor cells per animal and maintained over four weeks. Tumors were extracted, and each tumor was split in two. One part was formalin-fixed and embedded in paraffin, whereas the other part was frozen at −80 °C. For the FFPE samples, DNA was extracted using the QIAamp DNA FFPE Tissue kit (Sigma-Aldrich) following the manufacturer's protocol. DNA was resuspended in TE buffer pH 7.5 and treated with RNase (Sigma-Aldrich) for 45min at 37 °C. DNA from the fresh-frozen samples was extracted in the same way as for the primary colon tumors.

All reference DNA samples were quantified using Qubit 2.0 (Invitrogen) and quality-checked by gel electrophoresis. Homogeneous aliquots of equal volume corresponding to a target DNA amount of 1 μg were prepared for all reference samples and shipped on dry ice to the contributing laboratories. Scientists in each laboratory confirmed DNA quality by gel electrophoresis and DNA amounts by one of four alternative technologies (Qubit, Quant-iT, NanoDrop, TapeStation), and the results in the analysis reports.

Selection of target genomic regions. To select informative and challenging target regions, we initially ran the Infinium 450k assay on all reference samples excluding the titration series and the FFPE samples. The resulting data were preprocessed and analyzed with RnBeads[67], and we selected 1,072 genomic regions (corresponding to 16 core regions, 32 additional regions, and 1024 further regions for the EnrichmentBS assays) from the list of Infinium probes that passed quality control. Each of the regions was 122 base pairs wide, with the designated target CpG at the center and a window of 60 base pairs in both directions. 25% of these regions were selected based on differential DNA methylation for colon tumor vs. normal, 25% based on differential DNA methylation for drug treatment vs. control, 25% based on differential DNA methylated between the colon cancer cell lines and whole blood, and 25% were randomly selected. Within each block of differentially methylated CpGs we selected: (i) a subset of the most strongly differentially methylated

CpGs (ranks 1, 2, 3, 5, 7, 10, 13, 17, 21 and 25); (ii) the 1% highest and lowest regions in terms of repetitive DNA content, GC content, CpG content, CpA content and CpG observed vs. expected ratio; and (iii) the 1% CpGs with DNA methylation values in the first group that were closest to 0%, 25%, 50%, 75% and 100%, respectively. We also included CpGs in genomic regions that had previously been described as epigenetic biomarkers for colon cancer[68]. From this list of 1,072 genomic regions, we manually selected 16 core regions and 32 additional regions such that the selected regions capture as much as possible of the technical challenges in assay design and analysis with locus-specific DNA methylation assays. One of the core regions (region 11) was chosen such that the target CpG overlapped with a common SNP, in order to provide an additional challenge during assay design. Furthermore, some of the remaining 1,024 genomic regions were covered by the two assays that easily scale to hundreds of genome regions (EnrichmentBS 1 and EnrichmentBS 2). These measurements are available from **Supplementary Data 3**, but for reasons of comparability across assays they were not included in the benchmarking.

Documentation of DNA methylation assays, results and analysis workflows. The details for all contributed DNA methylation assays are available in **Supplementary Data 1**. These reports include a short assay summary, quality control data for the received reference DNA samples, and detailed descriptions of the design and execution of each contributed assay. They follow the standardized reporting template from the information package that was sent to all contributing laboratories (**Supplementary Data 2**). DNA methylation measurements for each assay, genomic region and reference sample are available in **Supplementary Data 3**. Illumina 450k microarray data are available at the NCBI Gene Expression Omnibus under the accession number GSE77965. Finally, the source code (written in R) underlying the bioinformatic analysis is available in a public repository (http://biomarker-benchmark.computational-epigenetics.org/), to foster transparency and reuse in the spirit of open science and reproducible research[69].

Bioinformatic analysis of absolute DNA methylation assays. To quantify assay performance without a priori knowledge of the true DNA methylation values in the reference samples, we defined target DNA methylation values by consensus. The consensus corridor was calculated as the narrowest interval containing measurements from three different technologies, extended by an additional flanking region of five percentage points in both directions. We chose this corridor (rather than, e.g., the arithmetic or geometric mean between all measurements) to minimize bias toward overrepresented assays. Based on the consensus corridor we calculated the absolute difference $|d_{a,r,s}|$ between the DNA methylation values $v_{a,r,s}$ (as measured by an assay a for a region r in a sample s) and the closest boundary of the corresponding consensus corridor $c_{r,s}$: $|d_{a,r,s}| = v_{a,r,s} - c_{r,s}$. According to this metric, assays with smaller absolute differences are in better agreement with the consensus. To assess whether certain assays tend to systematically overestimate or underestimate DNA methylation values, we also calculated the directional deviation b_a as the mean of all differences:

$$b_a = \frac{\Sigma_{r \in R} \Sigma_{s \in S} d_{a,r,s}}{|R| \times |S|} .$$

Finally, we performed a sensitivity analysis for the consensus corridor (**Supplementary Note** and **Supplementary Fig. 4**), which confirmed that our choice of parameters (i.e., using three different technologies and flanking regions of five percentage points to constitute the consensus corridor) was appropriate for robustly ranking the assays by their performance.

We also quantified the absolute assay performance in an alternative way, which does not rely on any consensus values but makes use of the two titration series. The DNA methylation values in both titration series are expected to be proportional to the titration ratios, which are known. In contrast, the DNA methylation values at the two extreme points of the titration series are different between regions and a priori unknown. Therefore, as outlined in **Supplementary Figure 7**, we first calculated the difference between the median of the consensus corridors for each titration series and each region at the 0% and 100% titration ratios. We then removed all regions that did not change by at least five percentage points to focus the analysis on regions with a clear-cut

change in DNA methylation over the titration series. Next, regions with a negative change between the 0% and 100% consensus values were inverted by subtracting their measured DNA methylation value from the maximum corresponding to complete DNA methylation. This procedure reversed directionality for the particular region and therefore standardized the direction across all regions. Finally, we adjusted for different offsets of DNA methylation levels by fitting a linear model to the values of each region and then subtracting the linear model offset (intersect) from these values. Using the adjusted DNA methylation values we then evaluated the Pearson correlation of the measured values to the titration ratios, which is the titration-based estimate of the correct value. To evaluate how well the assays captured the linearity of the DNA methylation values along the titration series, we also fitted a second intercept-free linear model to the adjusted DNA methylation values across all regions and samples, and we recorded the adjusted r^2 and residual standard error of the fitted model. Assays with higher adjusted r^2 values and lower residual standard error were considered in better agreement with the expectation that was based on the known titration ratios.

Bioinformatic analysis of relative DNA methylation assays. We compared the relative assays among each other by calculating pairwise 3-by-3 contingency tables for the differences between each pair of samples recorded by each assay. Measurements that agreed on the direction of change in both assays appear on the diagonal of the contingency table, and the higher the percentage of measurements on the diagonal, the more concordant both assays are. We formalize the agreement between assays as a numeric value, the percent concordance:

$$A = \frac{\Sigma_i \Sigma_j c_{i,j} d_{i,j}}{\Sigma_i \Sigma_j c_{i,j}},$$

where $c_{i,j}$ is the value in the ith row and jth column of the contingency table and

$$d_{i,j} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}.$$

This approach readily generalizes to the absolute assays, where we considered samples with an absolute difference of less than five percentage points as concordant.

In a separate and complementary analysis, we evaluated the ability of the relative assays to detect the correct direction of change between any two samples by using the median of the three DNA methylation values spanning the previously defined consensus corridor as reference. For each pair of samples, we determined the target direction and magnitude of change as the difference between the two median values, and we checked for each relative assay whether the difference between the corresponding measurements had the same or opposite direction of change. If no difference was detected in the relative assays, this was also recorded. The differences in the medians were divided into four bins: marginal change (absolute difference below five percentage points), small change (5–25 percentage points) medium change (25–50 percentage points), and strong change (above 50 percentage points). Finally, we also evaluated the relative assays based on the titration series, including only those regions with a difference above five percentage points between the two extreme points according to the consensus corridor. Results were regarded as consistent with the titration series if the direction of change observed for the relative assay was the same as the direction of the change in the titration ratio, taking into account the two extreme points according to the consensus corridor.

Correlation between input DNA amounts and assay performance. Two alternative approaches were used to assess the effect of DNA amounts on assay performance (**Supplementary Note**). First, owing to normal variation in the extracted DNA quality/quantity and in the initial quantification, the DNA amounts varied slightly between reference samples, e.g., ranging from 875 ng to 1,843 ng in the primary tumor/normal samples (**Supplementary Fig. 10a**).

Each laboratory was provided with the exact same volume of homogeneous aliquots for these samples, such that these differences between samples did not result in differences between laboratories. To correlate input DNA amounts with assay performance, we fitted a linear model predicting the deviation from the consensus corridor for each sample and assay using two alternative measures of input DNA amounts: the first value based on the median of concentration measurements across all laboratories multiplied by the volume of DNA used for a given assay, and the second value based on the DNA amounts that each research group reported to have used according to their own concentration measurements. For each assay and each of the two measurements of DNA amount, P values were calculated with linear models and adjusted for multiple testing using the Benjamini-Hochberg method. We used an adjusted P-value threshold of 0.05 to call assays significantly influenced by DNA amount, but no associations were significant at this level.

Second, to assess the impact of DNA amounts in a much lower range (0.3 ng to 100 ng), we established a titration series with target DNA amounts of 100 ng, 30 ng, 10 ng, 3 ng, 1 ng and 0.3 ng from one of the tumor reference samples (CRC 2 tumor). To exclude differences owing to variation among different bisulfite conversion protocols and reagents[70], these samples were bisulfite-converted centrally, using the EZ DNA Methylation-Direct Kit (Zymo Research, D5020) with the following deviations from the manufacturer's protocol: Conversion reagent was applied at 0.9× concentration, reactions incubated for 20 cycles of 1 min at 95 °C and 10 min at 60 °C, and the desulphonation time was extended to 30 min. The converted DNA was shipped on dry ice to nine laboratories that repeated their assays on these samples. We also analyzed the impact of reductions in DNA quality by fragmenting DNA from one of the tumor reference samples (CRC 1 tumor) to an average fragment length of 200 base pairs. To that end, batches of 600 ng DNA were digested with NEBNext dsDNA Fragmentase (New England BioLabs, M0348L) for exactly 60 min at 37 °C, stopping the fragmentation reactions by addition of 5μl 0.5M EDTA stop solution. The fragmented batches were combined, titrated to the same amounts as above, bisulfite-converted and shipped to the contributing laboratories.

Prediction of global DNA methylation levels. The global DNA methylation assays give rise to one single value per sample, which made it possible to plot all data points into one diagram (**Fig. 4a**) and to assess the overall consistency of the results by visual inspection. In addition, we explored whether we could predict global DNA methylation values from the results of the locus-specific DNA methylation assays, either by using the mean or median of the DNA methylation levels or by more complex machine learning methods such as generalized linear models, support vector regression (linear and polynomial kernels) and random forest regression. To compensate for the fact that not all assays were run on all samples, we first imputed missing values by filling in the values of the most closely related other assay based on Pearson correlation. We trained the regression models using leave-one-out cross-validation to make optimal use of the limited data set. For each method and each analysis, we recorded the root mean square error (RMSE) between the prediction and the target value. As no single global assay gave fully consistent results, we chose as global target the mean of the two best-performing assays (HPLC-MS and Pyroseq NBL2), and we replaced the four mean values that were inconsistent with the known change in concentration in the titration series by imputed values that were calculated as the mean of the two neighboring values in the titration series. The e1071 R package was used for support vector regression, randomForest for random forest regression and DMwR for cross-validation.

Analysis of discriminatory power. We trained linear support vector machines using patient-stratified cross-validation, such that each prediction used four tumor/patient pairs for training and left two pairs out for test-set validation. The e1071 R package was used to train the classifiers and the ROCR package[71] to calculate the ROC area under curve as the main performance metric. We further examined the robustness of the classifiers in presence of two different error models: (i) random error and (ii) uniform noise.

Random error. We simulated faulty measurements by replacing a defined fraction of measurements by random numbers drawn from the pool of all measurements of that assay. In this manner, we ensured that the simulated

erroneous measurements were drawn from the same distribution as the correct measurements without making assumptions about the statistical distribution of the data.

Uniform noise. We simulated inaccurate measurements by adding a random number to each measurement. At any given noise level n, this random number was sampled uniformly from the interval $[-n \times r; n \times r]$, where r is the range spanned by all DNA methylation values for the same assay. To assess the prediction performance, we tested each classifier in a stratified three-fold cross-validation: for each error model, noise/error level, assay, and selection of training and test set, we performed 1,000 repetitions of the analysis with randomized noise/error. To assess the robustness toward fewer measurements, we repeated the analysis with 25% uniform noise after removing the majority of regions from the training and test sets. The choice of regions retained (either 1, 3 or 5) was either entirely random or guided by the information content of each region for the prediction. We calculated the information content separately for each assay and region as the F score[72]. As before, we performed patient-stratified cross-validation with random repetitions. Finally, we analyzed a much larger cohort with 160 primary prostate tumor samples and 8 nonmatched normal prostate samples, comparing the EpiTyper 3 and Infinium assays with each other in terms of their correlation and discriminatory power.

66. Koeffler, H.P. & Golde, D.W. Acute myelogenous leukemia: a human cell line responsive to colony-stimulating activity. *Science* **200**, 1153–1154 (1978).
67. Koeffler, H.P., Billing, R., Lusis, A.J., Sparkes, R. & Golde, D.W. An undifferentiated variant derived from the human acute myelogenous leukemia cell line (KG-1). *Blood* **56**, 265–273 (1980).
68. Assenov, Y. *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* **11**, 1138–1140 (2014).
69. Heyn, H. & Esteller, M. DNA methylation profiling in the clinic: applications and challenges. *Nat. Rev. Genet.* **13**, 679–692 (2012).
70. Gentleman, R. & Temple Lang, D. Statistical analyses and reproducible research. Bioconductor project working papers. Working paper 2. http://biostats.bepress.com/bioconductor/paper2##(2004).
71. Holmes, E.E. *et al.* Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One* **9**, e93933 (2014).
72. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
73. Chang, Y.-W. & Lin, C.-J. Feature ranking using linear SVM. *JMLR Workshop Conf. Proc.* **3**, 53–64 (2008).