Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

A comparison of existing global DNA methylation assays to low-coverage whole-genome bisulfite sequencing for epidemiological studies

Florence K. Crary-Dooley[a], Mitchell E. Tam[a], Keith W. Dunaway[a], Irva Hertz-Picciotto[b,c], Rebecca J. Schmidt[b,c], and Janine M. LaSalle[a,c]

[a]Department of Medical Microbiology and Immunology, Genome Center, University of California, Davis, CA, USA; [b]Department of Public Health Sciences, University of California, Davis, CA, USA; [c]MIND Institute, University of California, Davis, CA, USA

ABSTRACT

DNA methylation is an epigenetic mark at the interface of genetic and environmental factors relevant to human disease. Quantitative assessments of global DNA methylation levels have therefore become important tools in epidemiology research, particularly for understanding effects of environmental exposures in complex diseases. Among the available methods of quantitative DNA methylation measurements, bisulfite sequencing is considered the gold standard, but whole-genome bisulfite sequencing (WGBS) has previously been considered too costly for epidemiology studies with high sample numbers. Pyrosequencing of repetitive sequences within bisulfite-treated DNA has been routinely used as a surrogate for global DNA methylation, but a comparison of pyrosequencing to WGBS for accuracy and reproducibility of methylation levels has not been performed. This study compared the global methylation levels measured from uniquely mappable (non-repetitive) WGBS sequences to pyrosequencing assays of several repeat sequences and repeat assay-matched WGBS data and determined uniquely mappable WGBS data to be the most reproducible and accurate measurement of global DNA methylation levels. We determined sources of variation in repetitive pyrosequencing assays to be PCR amplification bias, PCR primer selection bias in methylation levels of targeted sequences, and inherent variability in methylation levels of repeat sequences. Low-coverage, uniquely mappable WGBS showed the strongest correlation between replicates of all assays. By using multiplexing by indexed bar codes, the cost of WGBS can be lowered significantly to improve the accuracy of global DNA methylation assessments for human studies.

Introduction

Within the field of molecular and genetic epidemiology, an increasing number of studies have sought to associate specific environmental exposures with changes to global DNA methylation levels and disease outcome measures.[1-6] The epigenetic marker 5-methylcytosine (5mC) is a stable covalent modification that can be measured from DNA isolated of any tissue type, including easily obtainable peripheral blood. However, determining which assay of global DNA methylation levels is the most accurate and cost effective has remained a challenge.

There are a variety of different methods to assess genome-wide DNA methylation, including array-based, antibody-based, and sequencing-based approaches. In a recent community-based benchmarking study comparing DNA methylation assays across laboratories, the sequencing-based assays of amplicon bisulfite sequencing and pyrosequencing were all-around better performers than array-based or antibody-based approaches.[7] In this same benchmarking study focused on cancer relevance, assessments of global DNA methylation showed poor correlations between the 3 different assays of repeat-based pyrosequencing, HPLC-MS, or an enzyme-linked immunsorbent assay (ELISA) assay. Whole-genome bisulfite sequencing (WGBS) was not included in this

benchmarking study, but could be an affordable alternative to improving the accuracy of global DNA methylation. Suggested coverage for differentially methylated region identification is 5x to 15x CpG coverage;[8] however, lower sequencing depth may be sufficient for assaying global DNA methylation levels. Even when performed at low-coverage through the use of indexed bar codes in sequencing library preparation, WGBS has the potential advantage of providing random sampling of CpGs over the entire genome compared with repeat-based pyrosequencing.

Repetitive elements make up almost two-thirds of the human genome,[9] thereby contributing greatly to global DNA methylation levels. Reactivation of transposable repeats, such as LINE-1 and Alu, through hypomethylation can increase genome instability,[10] reactivate lowly expressed genes,[11] or disrupt gene function,[12] thereby potentially contributing to disease risk. Since LINE-1 and Alu repeats are represented at a high frequency in the human genome,[13] methylation assays designed to recognize these repeats are considered the best available estimates of "global" DNA methylation levels in the sample.[14]

Pyrosequencing of LINE-1 or Alu repeats is commonly used in epidemiological studies and favored for its low price and

CONTACT Janine M. LaSalle, Ph.D. ✉ jmlasalle@ucdavis.edu 🖃 Department of Medical Microbiology and Immunology, One Shields Ave., Davis, CA, USA.

⊕ Supplemental data for this article can be accessed on the publisher's website.

short processing time, despite several recognized limitations. First, for the synthesis-coupled light emission to be detected by the pyrosequencer, target sequences need to be heavily amplified. PCR amplification bias is likely to occur due to the 45 cycles of amplification required and multiple templates targeted.[15,16] This type of amplification bias presents a challenge to replicating results since a different random subset of amplicons is selected in each PCR reaction. Second, bias also can be introduced by primer affinity differences, so that a subset of targeted repeat sequences can be preferentially amplified based on their sequence or methylation state. Last, repetitive sequence methylation levels may not correlate with genome-wide assessments. Specifically, LINE-1 global pyrosequencing assay did not correlate with Illumina Infinium 27K array data in one study[17] or the luminometric methylation assay (LUMA) of $C^m CGG$ recognition sites genome-wide in another.[18] However, methylation values from pyrosequencing LINE-1 or Alu repeat assays have not been previously compared for the same samples to the same target sequences within WGBS data or to uniquely mappable (non-repetitive) WGBS average methylation levels.

In this study, global methylation was measured from 12 human whole-blood samples using both low-coverage multiplexed WGBS and repeat-based pyrosequencing techniques. Cost and data quality are considered to determine if low-coverage WGBS could be an improved method for global methylation measurement in studies involving high sample numbers.

Results

The effect of coverage on methylation and sensitivity in WGBS

For whole-genome bisulfite sequencing to be useful for global DNA methylation assessments in epidemiology studies, it must be comparable with pyrosequencing in price and data quality. One concern about the existing LINE-1 pyrosequencing assay (PyroMark Q24 CpG LINE-1) was the large 18% range of variability we observed using an identical DNA sample assayed 38 different times over a period of 20 months (Fig. 1). In addition, the mean percent methylation estimate of all the pyrosequencing replicates (74.5%) was substantially lower than the same sample previously assessed by WGBS (82.4%).[19]

To test the hypothesis that low-coverage WGBS using indexed libraries could improve quality and reproducibility of global DNA methylation assessments compared with LINE-1 pyrosequencing, 12 human whole-blood samples from pregnant mothers were used for comparison benchmarking by multiple methods. Pyrosequencing was performed in triplicate using 4 different assays targeting LINE-1 and ALU repetitive elements: commercial L1 (Qiagen), novel NL1 primers designed from LINE-1 repetitive sequences extracted from hg38, as well as USCL1 and ALUYB8 previously published.[20,21] For low-coverage WGBS, the 12 samples were run with 12 individual barcodes in a single lane of whole-genome bisulfite sequencing (WGBS1), as well as in duplicate using 24 barcodes in a second lane of sequencing (WGBS2 & WGBS3). For all 3 WGBS data sets, we analyzed both total average percent methylation after mapping to human genome (uniquely mappable WGBS), as well as specific repetitive sequence categories. We
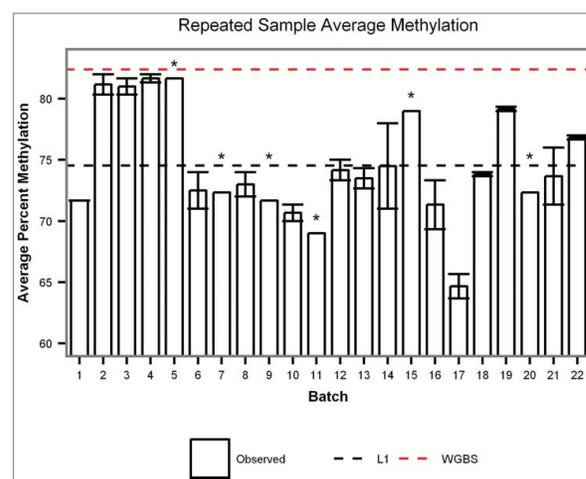

Figure 1. L1 pyrosequencing variability of the same sample across time. A single male human cord blood sample was pyrosequenced in duplicate wells per run (once if marked by asterisk) for a total of 38 assays over 22 runs using the PyroMark Q24 CpG LINE-1 assay over a period of 20 months. Average percent methylation for each batch is shown with standard error. Average methylation of all 38 measurements is marked by the dashed black line [mean = 74.54, standard deviation (sd) = 4.6, range = 18.33]. Uniquely mappable WGBS percent methylation level of the same sample is marked by the red dashed line. (WGBS % me = 82.4).

define "uniquely mappable" WGBS reads as those that align to a single location in the genome. In contrast, the specific repeat sequences were identified from sequence reads before genome alignment.

Since multiplexing reduces the coverage per sample, it is important to first determine if the coverage of sequencing reads from the uniquely mappable WGBS data will affect the measurement of global percent methylation. Sequencing reads from each of the 12 samples in the 12 per lane sequencing run were "down-sampled" until coverage for all samples was less than 10^{-5} (Fig. 2a). No difference in average percent methylation was found among the different sequence subsets within the range of 0.1x to 0.77x coverage, and only minor differences were observed below 0.1x coverage. These results demonstrate that depth of sequencing coverage does not detectably affect the accuracy of global percent methylation by WGBS. In addition, no chromosomal bias in the read coverage was observed in either 12 per lane coverage, 24 per lane coverage, or simulated reduced coverage (Fig. S1).

Previously published WGBS data on natural killer (NK) cells and kidney[22] were randomly down-sampled as in Fig. 2a to create simulated NK and kidney WGBS data sets of similar coverage (Fig. 2b). Mixed data sets at ratios 25:75, 50:50, and 75:25 were created by combining randomly sampled portions of the tissue specific data. A methylation difference of 2.6% between tissue types remains detectable even as coverage decreased below 0.1x coverage. Mixed data sets fell as expected between the non-mixed data sets, suggesting that tissue-specific differences lower than 2.6% can be detectable at coverage lower than 0.1x.

To specifically compare the sensitivity of WGBS and pyrosequencing global methylation assays, sample number and standard deviation can be used to calculate minimum difference needed to produce a power of 0.95 and significant P-value of 0.05. Minimum percent methylation difference, or delta, was calculated for each pyrosequencing assay and WGBS at
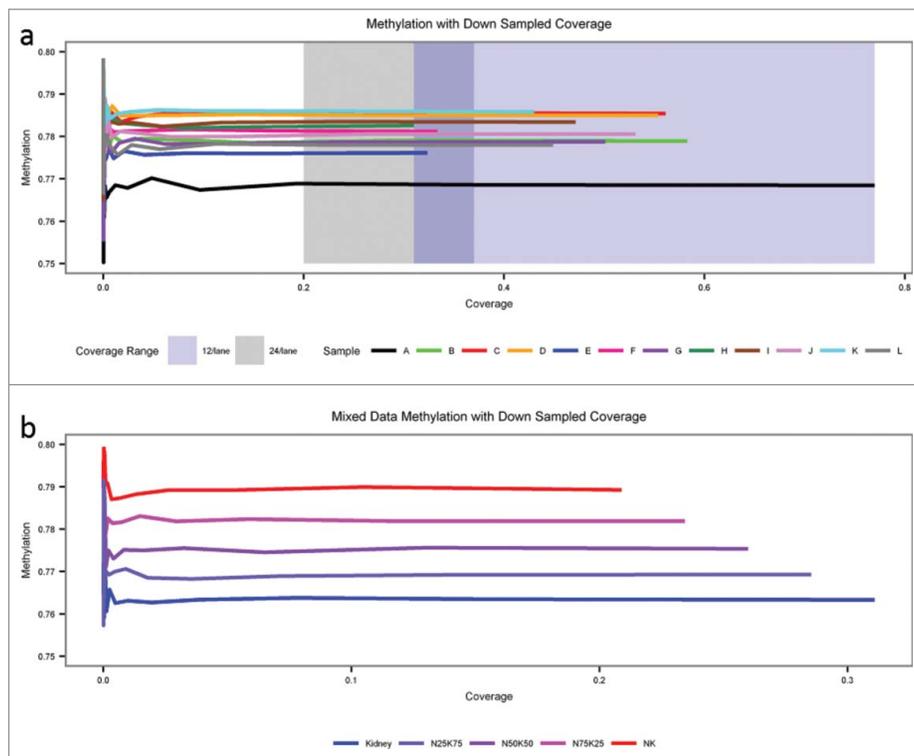
Figure 2. WGBS global methylation levels are not detectably affected by sequence coverage. (a) Uniquely mappable WGBS reads were down-sampled from the 12 sample per lane sequencing run. An increasingly small subset of all mappable reads was randomly selected until coverage for all samples was less than 10^{-5}. The areas shaded blue and gray represent the range of coverage for sequencing at 12 per lane and 24 per lane respectively. (b) Previously published WGBS data from human NK cells and kidney tissue was down-sampled as in (a). Reads from NK and kidney were mixed at different proportions; 50/50, 75/25, 25/75. The red line represents NK cell methylation while the blue line represents kidney tissue methylation. The lines in between NK and kidney represent different mixtures of sequencing data.

different coverage levels. WGBS had a delta of 0.807% at 0.15x coverage and delta remained under 1 until 5e-04x coverage (Table S1). Delta was 2.3–4.6 times lower for WGBS at 0.15x coverage compared with pyrosequencing assays, confirming that WGBS is more sensitive than pyrosequencing at detecting small methylation changes. Furthermore, in a simulated grouping of the 12 samples into either "high methylation" or "low methylation" groups, a significant difference between groups was still detectable in simulated WGBS data sets as low as 0.002x coverage (Fig. S2).

Comparing variability of repetitive pyrosequencing assays directly to WGBS data

To determine if the methylation variability seen in pyrosequencing replicates was affected by the specific repetitive sequences targeted in the repeat-based global pyrosequencing method, we directly compared variability of pyrosequencing to assay-matched WGBS. The pyrosequencing assay target sequences were extracted from raw WGBS data before mapping to produce corresponding WGBS methylation data of the same CpG sites, termed "assay-matched WGBS methylation" for each assay. Absolute deviation from the mean was calculated for each pyrosequencing and assay-matched WGBS, and compared with variability in uniquely mappable WGBS data (Fig. 3, Table S2). A high absolute deviation indicates high run-to-run variability. Average absolute deviation of uniquely mappable WGBS (gray) was lower than absolute deviations of
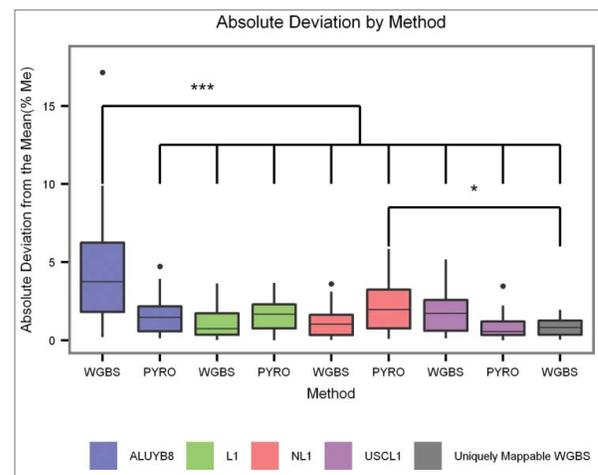

Figure 3. Comparison of absolute deviations between different global pyrosequencing assays and assay-matched WGBS. Twelve human whole-blood samples were pyrosequenced in triplicate using 4 separate assays targeting repetitive elements. Methylation at CpG sites corresponding to each pyrosequencing assay was extracted from unmapped reads as assay-matched WGBS methylation from 3 replicates of WGBS data for each sample. Three pyrosequencing replicates from each repeat-type assay for each of the 12 samples are averaged together to produce an assay average. The absolute deviation from the mean is calculated by taking the absolute value of the difference between the assay average and an average methylation measurement. Direct pairwise comparisons of global pyrosequencing and WGBS are indicated by matching colors, where absolute deviation from the mean for each pyrosequencing assay is shown on the right and absolute deviation from the mean for the corresponding CpG sites by WGBS on the left. For comparison, uniquely mappable WGBS from the same 12 samples is shown on the left. (*** $P < 0.0005$, * $P < 0.05$).
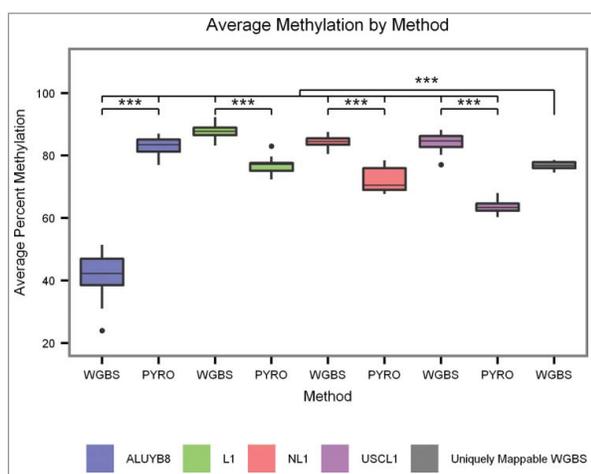
Figure 4. Average methylation from 4 pyrosequencing assays compared with assay-matched WGBS CpG sites, as well as 3 replicates of uniquely mappable WGBS. The absolute levels of methylation are significantly different between pyrosequencing and assay-matched WGBS assays performed on the same 12 samples. Uniquely mappable WGBS is significantly different from all other measurements except for L1 pyrosequencing. ($^{***}P<0.0005$).

Figure 5. Pooling of PCR reactions improves variability of L1 pyrosequencing assay. Either 2 or 6 PCR reactions were pooled, mixed, and re-aliquoted before pyrosequencing using L1. Absolute deviation from the mean is shown grouped by the number of PCR reactions that were pooled and split: single (n = 72), 2 (n = 36) and 6 (n = 216). Significance was not affected when n was normalized. Mean was calculated as the average across all pyrosequencing reactions from the same pool. Pooled samples, regardless of the number of reactions pooled, had lower absolute deviation than single reactions. The difference remains significant when an equal number of reactions are sampled from each group. ($^{***}P<0.0005$).

pyrosequencing measurements across all assays except for USCL1 (purple). In assay-matched CpG sites, however, absolute deviation was lower in WGBS than pyrosequencing for only 2 assays: L1 (green) and NL1 (red). Variability of WGBS was greater than pyrosequencing for assay-matched CpG sites in the other 2 assays, ALUYB8 and USCL1.

In addition to differences in variability, absolute levels of DNA methylation varied widely between pyrosequencing and assay-matched WGBS for the same samples and CpG sites (Fig. 4, Table S2). For all 3 of the LINE-1 pyrosequencing assays (L1, NL1, USCL1), absolute methylation levels were significantly 10–20% lower than the same repeat type assayed by WGBS, suggesting either a bias toward lower methylated regions in the pyrosequencing assays or a bias toward higher methylated regions in the WGBS library preparation. The ALUYB8 assayed showed an assay difference of over 40% from WGBS methylation levels in the opposite direction, with pyrosequencing showing higher percent methylation levels compared with assay-matched WGBS (Fig. 4, blue), suggesting that the direction of the bias is dependent on the repeat type.

PCR amplification bias introduces variability to global DNA methylation assays

To test the hypothesis that the observed variability in L1 pyrosequencing over time using the same sample (Fig. 1) is due to PCR amplification differences, replicates were homogenized through pooling of separate reactions before pyrosequencing. Replicates of either 2 or 6 PCR reactions were pooled after amplification, thoroughly mixed, and separated back into 2 or 6 aliquots. Pooled replicates were then pyrosequenced using PyroMark L1 and absolute deviation was compared with single reaction measurements (Fig. 5). Both pools of 2 replicates and pools of 6 replicates showed significantly lower absolute deviation than single PCR reaction pyrosequencing measurements. There was no difference in absolute deviation between pools of 2 replicates and pools of 6 replicates. These results support the
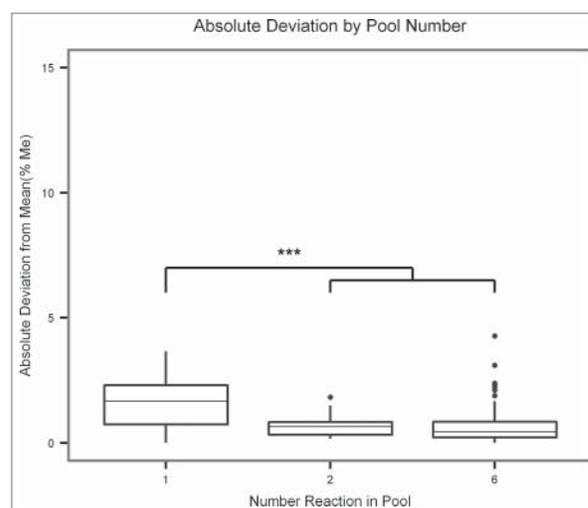
hypothesis that PCR amplification differences between reaction tubes is a significant source of variability. The multi-template nature of repetitive element PCR for pyrosequencing may introduce inevitable variability into global methylation pyrosequencing assays.

Repetitive element methylation CpG sites assayed by pyrosequencing are PCR-biased and more variable compared with uniquely mappable WGBS analysis of global DNA methylation levels

Repetitive elements in the human genome often contain degenerate sequences and therefore may not be well suited for global methylation assessment. As pyrosequencing primer sets are designed using consensus sequences, templates with higher complementarity to the assay primers are more likely to be amplified. To test the assay-matched WGBS reads for methylation-dependent sequence bias, the regions flanking the L1 and USCL1 target sequences were extracted from the WGBS and represented as degenerate sequence frequencies ordered according to target sequence percent methylation (Fig. 6, Fig. S3). For L1 targets, where primer sequences are not available, several methylation dependent sequence frequency variations were observed. Furthermore, USCL1 target region methylation with exact sequencing primer matches showed lower methylation of target regions than those without exact matches, a result that is consistent with the lower methylation observed with the USCL1 pyrosequencing assay compared with the WGBS assay-matched methylation levels. Further support for the hypothesis that the USCL1 pyrosequencing primers are selecting for a subset of the target sequences with lower methylation is observed as the change in primer match frequency with increased methylation of target sequences (Table 1).
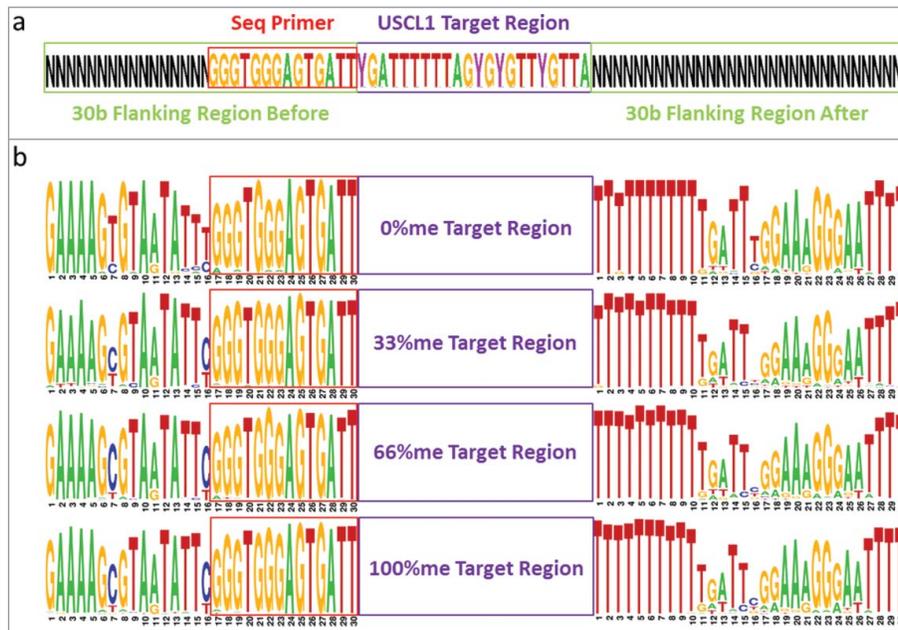
Figure 6. USCL1 flanking sequences are associated with different methylation levels. (a) Flanking 30 bases before and after the target sequence from the USCL1 assay were extracted from the WGBS data. The red-boxed region indicates the sequencing primer. The purple box indicates the target region where the letter Y signifies a site of potential methylation. Due to limits in read length, extending the flanking region to include the forward and reverse primers was not possible. (b) Sequence logos were created from the extracted flanking regions. Multiple letters in a stack represents a degenerate site in the sequence and the height of the letter reflects abundance of that sequence within the reads. Different average percent methylation is grouped and displayed in each row. Sequence degeneracies are found in the flanking regions at specific target sequence methylation levels.

We then determined if estimates of repetitive element methylation levels correlated with global percent methylation from uniquely mappable WGBS data (Fig. 7, Fig. S4). Both pyrosequencing and WGBS assay-matched determinations of repetitive methylation levels correlated poorly with actual global methylation of the same samples determined from uniquely mappable WGBS data. Furthermore, in a correlation analysis of the 3 WGBS replicates of the 12 different individuals, a better correlation was observed in percent methylation for the uniquely mappable WGBS (black) compared with the repetitive assay-matched WGBS estimates (blue, purple, green, red) (Fig. 8, Table S3). Taken together, the lack of correlation and comparatively high variation observed over methylation analyses of repeats suggest that repetitive element methylation is not an accurate representation of whole-genome methylation.

Table 1. Sequence degeneracy at primer template skews methylation. USCL1 target region methylation with exact sequencing primer matches in the flanking region is lower than the methylation of target regions with inexact matches. The ratio of inexact match (Fail) to exact match (Pass) increases with increasing percent methylation of target.

Average methylation of all extracted target regions:			81.172%
Average methylation of exact primer match target regions:			80.508%
Average methylation of inexact match target regions			82.786%
% methylation of target	Exact primer match	# of matches	Ratio fail/pass
0%	Fail	7	0.206
	Pass	34	
33.33%	Fail	23	0.383
	Pass	60	
66.67%	Fail	106	0.417
	Pass	254	
100%	Fail	199	0.427
	Pass	466	

Discussion

In this study, we have performed direct comparisons of WGBS vs. pyrosequencing on the same 12 blood DNA samples to assess accuracy and reproducibility, and investigated possible sources of bias or inaccuracies with the methods. From these comparisons, we demonstrate that uniquely mappable CpG percent methylation from low-coverage WGBS is the most sensitive and least variable method for estimating global DNA
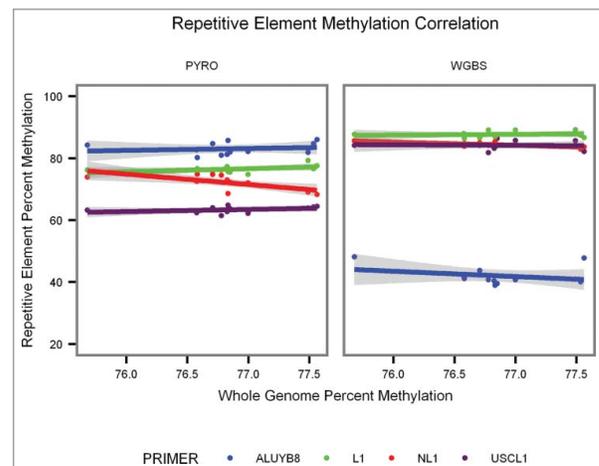

Figure 7. Poor correlation between repeat-based assays and uniquely mappable WGBS methylation levels. A direct comparison between uniquely mappable WGBS vs. repetitive element methylation for averaged methylation values of technical replicates is assessed by either pyrosequencing (left) or WGBS (right) methods on the same 12 samples. Correlation is poor between repetitive elements and whole-genome methylation regardless of assay or method (ALUYB8: $R^2_{pyro} = 0.025$, $R^2_{WGBS} = 0.086$; L1: $R^2_{pyro} = 0.188$, $R^2_{WGBS} = 0.013$; NL1: $R^2_{pyro} = 0.494$, $R^2_{WGBS} = 0.471$; USCL1: $R^2_{pyro} = 0.126$, $R^2_{WGBS} = 0.005$).
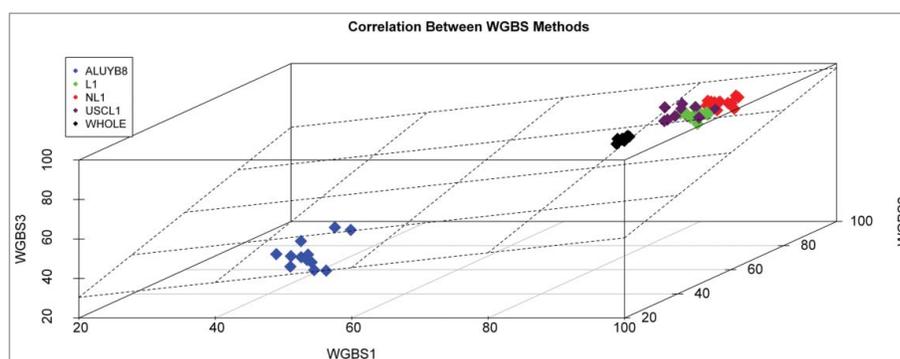
Figure 8. Inter-replicate variability of WGBS analysis of uniquely mappable WGBS (black) vs. repetitive elements (colors). Twelve human whole-blood samples were whole-genome bisulfite sequenced in 2 different libraries and 2 separate lanes of sequencing resulting in 3 replicates of sequencing for each sample. One lane contained the 12 samples prepared and multiplexed using the EpiGnome Kit for a total of 12 indexed libraries in a single lane (WGBS1). The other lane contained the 12 samples prepared and multiplexed using the NextFlex kit in duplicate for a total of 24 indexed libraries in a single lane (WGBS2 & WGBS3). Assay-matched target regions (defined from pyrosequencing assays) were extracted from each sample WGBS replicate. Each sample is shown in a 3D scatterplot where each axis is a replicate. All LINE-1 assay-average methylation cluster together (purple, green, red). ALUYB8 assay matched and whole-genome methylation also cluster independently. Unselected whole-genome methylation samples (black) cluster the most closely with the lowest standard deviation (ALUYB8: mean = 42.05, sd = 6.41; L1: mean = 87.73, sd = 1.8; NL1: mean = 84.37, sd = 1.65; USCL1: mean = 84.23, sd = 2.49; WHOLE: mean = 76.87, sd = 1.15).

methylation in human blood samples. Rather than sampling a single repetitive element, WGBS provides a random sampling of CpGs over the entire mappable genome. We also demonstrate that repetitive targets are poor surrogates for global DNA methylation levels regardless of the assay type, although of the available pyrosequencing assays, the USCL1 primers showed the lowest variability. Lastly, we show that LINE-1 pyrosequencing assays could be improved by increasing the number of replicates and pooling PCR reactions to reduce amplification bias.

In a comparison of cost, time, and quality of the global DNA methylation assays (Table 2), pyrosequencing clearly has advantages of low price and short analysis time that explain the method's popularity with high sample number studies, especially in the field of epidemiology. Upon closer examination, however, pyrosequencing of repetitive elements may not provide accurate or sensitive enough estimates of global methylation, especially when performed without replicates. A lack of correlation between repetitive element methylation and global methylation from uniquely mappable WGBS, in conjunction with inherent individual variability in repetitive element methylation suggests that repetitive element methylation may not be representative of global methylation. There could be times when assaying methylation of repetitive elements would be useful for understanding the repeats themselves (summarized in Table 2), or perhaps reducing problems with cell type heterogeneity.

Our results are also consistent with prior comparisons of PyroMark LINE-1 analysis to other methylation assays that showed poor correlation between methods.[7,17,18] The high number of PCR amplification cycles required for signal detection in pyrosequencing and the multiple templates that increase the probability of PCR bias are 2 likely sources for the wide variability between measurements on the same sample. This lack of precision in pyrosequencing may obscure small differences in global methylation levels, thereby outweighing the relevance of its advantages. A potential limitation to our study was that the impact of PCR amplification in introducing variability was not separated from instrumentation differences and other

methodological differences when comparing WGBS to pyrosequencing. Another limitation of our study was that the replication analyses were based on results within a single laboratory, although the methylation levels observed were within a comparable range to those reported in other human blood studies.[1-6] Future studies could perform consensus benchmarking studies of the same blood samples across different laboratories, as was performed between cancer laboratories.[7]

Our results also suggest that measuring methylation levels of repeats in the human genome is inherently more variable and less accurate that those of the uniquely mappable genome. While measuring methylation levels of specific subsets of repetitive sequences may be of interest in some studies, our results and those of others[7,17,18] caution against the simple interpretation that methylation measurements of repeat categories are an accurate surrogate for global DNA methylation levels in the non-repetitive portion of the genome.

Whole-genome bisulfite sequencing is currently the most expensive method of global methylation assay, but it provides many advantages while overcoming challenges present in pyrosequencing. Although PCR amplification is included in the preparation of libraries for sequencing, the number of cycles is less than a third of those with pyrosequencing and PCR duplicates are excluded at the data processing level.[23] An added benefit of WGBS not available with pyrosequencing is the ability to trace methylation back to a genomic region. While the 0.2x coverage WGBS performed here is not sufficient for accuracy at single CpG sites, percent methylation of individual chromosomes or large genomic regions, such as chromosome bands or partially methylated domains can be assessed.[24,25] Because of the reduced variability between replicates with uniquely mappable WGBS compared with other methods, the need for replicates may be eliminated when sample numbers are high. Despite lower coverage in multiplexed sequencing, a greater diversity of CpGs sites in the genome is analyzed and the selection of sites is randomly determined. The price of a single sample is thus reduced from $2,000-$3,000 to only $127, only 2 times more than running triplicates using pyrosequencing (Table 2). With current multiplexing technology, only 24

Table 2. A comparison of all methods used. Price, DNA required, time required, PCR cycle number, and overall assay quality is considered. WGBS assay time is based on the current estimate of sample turnover at the Vincent J. Coates Genomics Sequencing Laboratory (GSL) at the University of California, Berkeley. Run is defined as one pyrosequencing plate, one ELISA plate, or one Hi-Seq 4000 flow cell. * Calculated from 12/lane and 24/lane WGBS runs combined.

| Method | ng bsDNA needed | Amp Cycle | Rxn/run | Prep time/run (h) | Assay time/run (h) | Analysis time/run(h) | | Total time (h) | Time/rxn | Price/rxn (USD$) | Variability | Sensitivity (delta) | Recommended Use |
						User Input	Computer processing						
12/lane indexed WGBS	100	14	96	52	84	0.5	384	520.5	5.42	204.79	0.85 ± 0.56*	0.8% me difference*	Global methylation differences
24/lane indexed WGBS	100	18	192	64	84	0.5	384	532.5	2.77	126.74	0.85 ± 0.56*	0.8% me difference*	Global methylation differences
LINE-1 Pyrosequencing	50	45	24	6	0.6	0.5	0	7.1	0.3	20.64	1.66 ± 1.04	2% me difference	LINE-1 specific
NL1–1 Pyrosequencing	50	45	24	6	1	0.5	0	7.5	0.31	14.5	2.14 ± 1.60	3.7% me difference	Not recommended
ALUYB8 Pyrosequencing	50	50	24	6	0.6	0.5	0	7.1	0.3	14.5	1.58 ± 1.18	2.8% me difference	ALUYB8 specific
USCL1 Pyrosequencing	50	45	24	6	0.5	0.5	0	7	0.29	14.5	0.84 ± 0.73	1.8% me difference	LINE-1 specific
2 Rxn Pooling LINE-1	100	45	12	6.5	0.6	0.5	0	7.6	0.63	41.28	0.67 ± 0.43	2.63% me difference	LINE-1 specific

samples can be combined into a single lane of WGBS sequencing. Based on our WGBS "down-sampling" analyses, however, if 48 or 96 bar codes were available for WGBS, the cost could be significantly reduced without compromising accurate global methylation levels.

Although price can be a highly limiting factor in studies with high sample numbers, a low price per sample can become an irrelevant advantage if the disadvantages of the assay outweigh the benefits. Since most DNA methylation differences observed in epidemiology studies are small effect sizes, pyrosequencing may not be the best method available in the future for global methylation measurements. In the future, we expect that both improved multiplexing, specifically library sequencing kits with >24 bar codes, and increased sequencing yields will produce higher coverage and higher resolution for WGBS approaches.

Materials and methods

DNA isolation and bisulfite conversion

DNA was isolated from 0.2 to 1 mL of human peripheral whole-blood from pregnant mothers in the MARBLES study (Markers of Autism in Babies: Learning Early Signs) using the Gentra Puregene Blood Kit (Qiagen cat. 158445). The same protocol was used to isolate DNA from one male human cord blood sample from the MARBLES study, which was assayed 38 times using pyrosequencing. Purified DNA (500 ng) was bisulfite converted using EZ DNA Methylation-Lightning[TM] Kit (Zymo cat. D5030). Completely methylated and unmethylated DNA standards were obtained from EpiTect PCR Control DNA Set (Qiagen cat. 59695).

Pyrosequencing

Bisulfite converted DNA was eluted into 20 μL of elution buffer. For a single run, 2 μL ($\sim$50 ng) of converted DNA was PCR amplified using the PyroMark PCR Kit (Qiagen cat. 978703). Four different primer sets were used; 3 assays targeting different regions of LINE-1 repetitive elements and one assay targeting ALUYB8 repetitive elements. PCR cycling conditions for LINE-1 targeting primer sets were 95°C for 15 min initial denaturation, 95°C for 30 sec, annealing temperature for 30 sec, and 72°C for 30 sec for 45 cycles. Proprietary primers from PyroMark Q24 CpG LINE-1 (Qiagen cat. 970042) are called "L1," target sequence: TTYGTGGTGYG-TYGTTTTTTAAGTYGGTTT, annealing temperature: 50°C. Previously published LINE-1 primers are called "USCL1" — forward primer: TTGAGTTAGGTGTGGGATATA, reverse primer: bio-AAACATTACCTCACCTAAAAAA, sequencing primer: GGG TGGGAGTGATT, target sequence: YGATTTTTTAGYGYGT-TYGTTA, annealing temperature: 60°C.[20,21] Novel LINE-1 primers based on bioinformatics search are called "NL1" — forward primer: ATAGAGAAGTGTTTAAAGGAGTTGATG, reverse primer: bio-CATTCATTTCATCTTCCATTACTAATACC, sequencing primer: GGAGTTGAAAATTAAGGT, target sequence: TYGAGAATTAYGTGAAGAATGTAGAAGTTTTAGGAGTY-GATGYGAATT, annealing temperature: 54°C. PCR cycling conditions for ALU targeting primer set was 95°C for 15 min initial denaturation, 96°C for 50 sec, 53°C for 30 sec, and 72°C for 30 sec

for 50 cycles. Previously published ALUYB8 — forward primer: bio-AGATTATTTTGGTTAATAAG, reverse primer: AACTA-CYAACTACAATAAC, sequencing primer: AATAACTAAAAT-TACAAA, target sequence: RCCCRCCACCRCRCCCRACTA.[20] The PCR products were then assayed on the PyroMark Q24 with PyroMark Gold Q24 Reagents (Qiagen cat. 970802) and analyzed with accompanying software. A single methylation measurement is the average of all CpG sites in the assay except for position 4 in assay L1 and position 2 in assay USCL1 because of the lack of correlation with other CpG sites in the assay.

For pooled pyrosequencing, several PCR amplification reactions were prepared for a single biologic sample. The replicate PCR reactions were pooled after amplification and then separated again into 25 μL aliquots and used for separate pyrosequencing reactions. The resulting number of pyrosequencing replicates equal the initial number of PCR replicates; however, each 25 μL going into each pyrosequencing replicate is a homogenous mixture of all the initial PCR replicate products. For each of the 12 samples, 3 pools of 2 reactions (6 pyrosequencing measurements) and 3 pools of 6 reactions (18 pyrosequencing measurements) were run on the pyrosequencer using the L1 assay. Absolute deviation was calculated by taking the absolute value of the difference between an individual pyrosequencing measurement and the average methylation across all measurements in the same pool.

Whole-genome bisulfite sequencing

Libraries were prepared according to included protocols using the TrueSeq DNA Methylation Kit (Illumina cat. EGMK81312) with 12 TruSeq DNA Methylation Index PCR Primers (Illumina cat. EGIDX81312) or the NEXTflex Bisulfite Sequencing Kit (Bioo cat. 5119–02) with 24 NEXTflex Bisulfite Sequencing Barcodes (Bioo cat. 511913). The bisulfite conversion steps were prepared using EZ DNA Methylation-Lightning[TM] Kit as mentioned above. Libraries were sequenced on the HiSeq 4000 System (Illumina). Whole-genome methylation was determined from uniquely mappable reads aligned to hg38 using BS-Seeker2.[23] Target sequences from the pyrosequencing assays were used to search whole-genome bisulfite sequencing reads to determine methylation at assay matched CpG sites. Flanking regions to L1 and USCL1 target sequences were extracted from WGBS reads. Sequence logos were created using WebLogo.[26]

Average deviation from the mean

Absolute deviation from the mean was calculated by first averaging methylation across a group or across all replicates, then taking the absolute value of the difference between an individual measurement and the group average.

Down-sampling mixing of WGBS data

Simulated mixed sequencing data was created by concatenating sampled portions of previously published natural killer (NK) cell blood isolate and kidney tissue WGBS data using the fastq-sample function from fastq-tools.[27] Decreasing coverage was simulated by randomly sampling $1/2^x$ of all mappable reads, where x is every integer from 0 to 18. Additionally, sampling by

a calculated number of reads from total mappable reads was performed to achieve specific coverage levels. Average percent methylation was then calculated from all sampled reads.

References

1. Shigaki H, Baba Y, Watanabe M, Iwagami S, Miyake K, Ishimoto T, Iwatsuki M, Baba H. LINE-1 hypomethylation in noncancerous esophageal mucosae is associated with smoking history. Ann Surg Oncol 2012; 19:4238-43; PMID:22766991; http://dx.doi.org/10.1245/s10434-012-2488-y

2. El-Maarri O, Walier M, Behne F, van Uum J, Singer H, Diaz-Lacava A, Nusgen N, Niemann B, Watzka M, Reinsberg J, et al. Methylation at global LINE-1 repeats in human blood are affected by gender but not by age or natural hormone cycles. PLoS One 2011; 6:e16252; PMID:21311577; http://dx.doi.org/10.1371/journal.pone.0016252

3. Guerrero-Preston R, Goldman LR, Brebi-Mieville P, Ili-Gangas C, Lebron C, Witter FR, Apelberg BJ, Hernandez-Roystacher M, Jaffe A, Halden RU, et al. Global DNA hypomethylation is associated with in utero exposure to cotinine and perfluorinated alkyl compounds. Epigenetics 2010; 5:539-46; PMID:20523118

4. Tarantini L, Bonzini M, Apostoli P, Pegoraro V, Bollati V, Marinelli B, Cantone L, Rizzo G, Hou L, Schwartz J, et al. Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation. Environ Health Perspect 2009; 117:217-22; PMID:19270791; http://dx.doi.org/10.1289/ehp.11898

5. Subbalekha K, Pimkhaokham A, Pavasant P, Chindavijak S, Phokaew C, Shuangshoti S, Matangkasombut O, Mutirangura A. Detection of LINE-1s hypomethylation in oral rinses of oral squamous cell carcinoma patients. Oral Oncol 2009; 45:184-91; PMID:18715815; http://dx.doi.org/10.1016/j.oraloncology.2008.05.002

6. Rusiecki JA, Baccarelli A, Bollati V, Tarantini L, Moore LE, Bonefeld-Jorgensen EC. Global DNA hypomethylation is associated with high serum-persistent organic pollutants in Greenlandic Inuit. Environ Health Perspect 2008; 116:1547-52; PMID:19057709; http://dx.doi.org/10.1289/ehp.11338

7. consortium B. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol 2016; 34:726-37; PMID:27347756; http://dx.doi.org/10.1038/nbt.3605

8. Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods 2015; 12:230-2, 1 p following 2; PMID:25362363; http://dx.doi.org/10.1038/nmeth.3152

9. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 2011; 7:e1002384; PMID:22144907; http://dx.doi.org/10.1371/journal.pgen.1002384

10. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. Science 2003; 300:455; PMID:12702868; http://dx.doi.org/10.1126/science.1083557

11. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet 2010; 6:e1000917; PMID:20421991; http://dx.doi.org/10.1371/journal.pgen.1000917

12. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 1992; 52:643-5; PMID:1310068

13. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM, Moran JV. LINE-1 retrotransposition activity in human genomes. Cell 2010; 141:1159-70; PMID:20602998; http://dx.doi.org/10.1016/j.cell.2010.05.021

14. Lisanti S, Omar WA, Tomaszewski B, De Prins S, Jacobs G, Koppen G, Mathers JC, Langie SA. Comparison of methods for quantification of global DNA methylation in human cells and tissues. PLoS One 2013; 8:e79044; PMID:24260150; http://dx.doi.org/10.1371/journal.pone.0079044

15. Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng 2003; 96:317-23; PMID:16233530; http://dx.doi.org/10.1016/S1389-1723(03)90130-7

16. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol 1998; 64:3724-30; PMID:9758791

17. Price EM, Cotton AM, Penaherrera MS, McFadden DE, Kobor MS, Robinson W. Different measures of "genome-wide" DNA methylation exhibit unique properties in placental and somatic tissues. Epigenetics 2012; 7:652-63; PMID:22531475; http://dx.doi.org/10.4161/epi.20221

18. Knothe C, Shiratori H, Resch E, Ultsch A, Geisslinger G, Doehring A, Lotsch J. Disagreement between two common biomarkers of global DNA methylation. Clin Epigenetics 2016; 8:60; PMID:27222668; http://dx.doi.org/10.1186/s13148-016-0227-0

19. Schroeder DI, Jayashankar K, Douglas KC, Thirkill TL, York D, Dickinson PJ, Williams LE, Samollow PB, Ross PJ, Bannasch DL, et al. Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. PLoS Genet 2015; 11:e1005442; PMID:26241857; http://dx.doi.org/10.1371/journal.pgen.1005442

20. Breton CV, Byun HM, Wenten M, Pan F, Yang A, Gilliland FD. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. Am J Respir Crit Care Med 2009; 180:462-7; PMID:19498054; http://dx.doi.org/10.1164/rccm.200901-0135OC

21. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, Byun HM, Jiang J, Marinelli B, Pesatori AC, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res 2007; 67:876-80; PMID:17283117; http://dx.doi.org/10.1158/0008-5472.CAN-06-2995

22. Schroeder DI, Blair JD, Lott P, Yu HO, Hong D, Crary F, Ashwood P, Walker C, Korf I, Robinson WP, et al. The human placenta methylome. Proc Natl Acad Sci U S A 2013; 110:6037-42; PMID:23530188; http://dx.doi.org/10.1073/pnas.1215145110

23. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Gen 2013; 14:774; PMID:24206606; http://dx.doi.org/10.1186/1471-2164-14-774

24. Schroeder DI, Schmidt RJ, Crary FK, Walker CK, Ozonoff S, Tancredi DJ, Hertz-Picciotto I, LaSalle JM. Placental methylome analysis from a prospective autism study. Mol Autism 2016; 7:51; PMID:28018572; http://dx.doi.org/10.1186/s13229-016-0114-8

25. Schmidt R, Schroeder D, Crary-Dooley F, Barkoski J, Tancredi DJ, Walker C, Ozonoff S, Hertz-Picciotto I, LaSalle J. Self-reported pregnancy exposures and placental DNA methylation in the MARBLES propsective autism sibling study. Environmental Epigenetics in press; http://dx.doi.org/10.1093/eep/dvw024

26. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004; 14:1188-90; PMID:15173120; http://dx.doi.org/10.1101/gr.849004

27. Jones DC. Seattle, Washington: University of Washington: 2012 [accessed 2016 November 17]. http://homes.cs.washington.edu/~dcjones/fastq-tools/