

BRIF-Seq: Bisulfite-Converted Randomly Integrated Fragments Sequencing at the Single-Cell Level

Xiang Li^{1,2}, Lu Chen^{1,2}, Qinghua Zhang¹, Yonghao Sun¹, Qing Li^{1,*} and Jianbing Yan^{1,*}

¹National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

²These authors contributed equally to this article.

*Correspondence: Qing Li (qingli@mail.hzau.edu.cn), Jianbing Yan (yjianbing@mail.hzau.edu.cn)

<https://doi.org/10.1016/j.molp.2019.01.004>

ABSTRACT

Single-cell bisulfite sequencing (scBS-seq) was developed to assess DNA methylation heterogeneity in human and mouse. However, the reads are under-represented in regions with high DNA methylation, because these regions are usually fragmented into long segments and are seldom sequenced on the Illumina platform. To reduce the read distribution bias and maximize the use of these long segments, we developed bisulfite-converted randomly integrated fragments sequencing (BRIF-seq), a method with high rates of read mapping and genome coverage. Single microspore of maize, which has a highly methylated and repetitive genome, was used to perform BRIF-seq. High coverage of the haploid genome was obtained to evaluate the methylation states of CG, CHG, and CHH (H = A, C, or T). Compared with scBS-seq, BRIF-seq produced reads that were distributed more evenly across the genome, including regions with high DNA methylation. Surprisingly, the methylation rates among the four microspores within one tetrad were similar, but differed significantly among tetrads, suggesting that non-simultaneous methylation reprogramming could occur among tetrads. Similar levels of heterogeneity, which often occur in low-copy regions, were detected in different genetic backgrounds. These results suggest that BRIF-seq can be applied for single-cell methylome analysis of any species with diverse genetic backgrounds.

Key words: single-cell methylome, whole-genome amplification, reprogramming

Li X., Chen L., Zhang Q., Sun Y., Li Q., and Yan J. (2019). BRIF-Seq: Bisulfite-Converted Randomly Integrated Fragments Sequencing at the Single-Cell Level. *Mol. Plant*. ■ ■, 1–9.

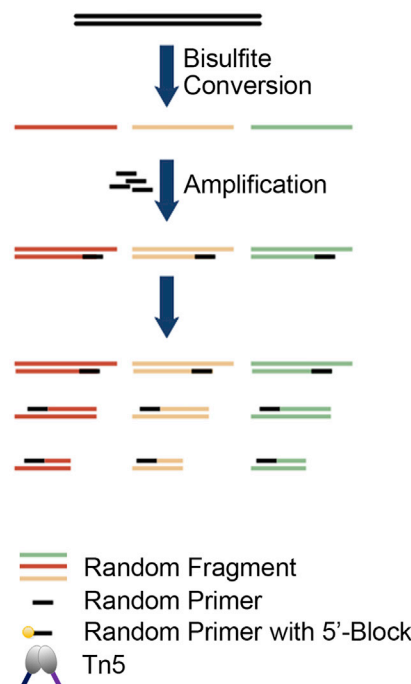
INTRODUCTION

5-Methylcytosine (5mC) is a major class of DNA modifications, and is associated with regulation of gene expression, transposon silencing, chromatin stability, and transgenerational reprogramming (Jones, 2012; Umer and Herceg, 2013; Walker et al., 2018). While CG is the primary site for DNA methylation in mammals, CHG and CHH (H = A, C, or T) methylation is also common in plants (Law and Jacobsen, 2010). DNA methylation levels are usually assessed in samples consisting of different cell types, potentially masking functionally important heterogeneity (Luo et al., 2017; Zhu et al., 2018). Single-cell reduced-representation bisulfite sequencing (scRRBS) was developed to analyze DNA methylation at the single-cell level (Guo et al., 2013). However, the sequencing reads were not evenly distributed and were enriched in regions with high GC content, partly due to biased digestion by restriction endonuclease *MspI* and preferential PCR amplification during library preparation. In addition, bisulfite conversion was performed after adapter ligation, which could degrade

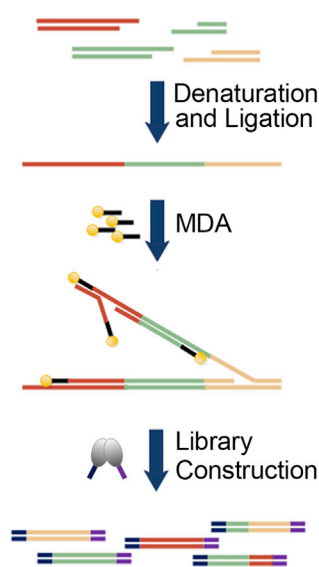
library fragments and reduce coverage. The sci-MET method (Mulqueen et al., 2018) was recently developed and was reported to have a high mapping rate. However, bisulfite conversion was performed after transposase tagmentation, resulting in a fragmented library and leading to low coverage. Therefore, increasing genome coverage remains a challenge for single-cell sequencing. Another method, single-cell bisulfite sequencing (scBC-seq) (Smallwood et al., 2014), is currently the mainstream approach. Recently, post-bisulfite adapter ligation (PBAL) (Hui et al., 2018) was developed. Both scBC-seq and PBAL incorporate bidirectional library amplification following bisulfite conversion, thus improving read coverage. scBS-seq has been applied to assess the heterogeneity of single human reproductive cells (Guo et al., 2015) and embryonic cells during embryogenesis (Zhu et al., 2018). However, hypermethylated genomes cannot be covered efficiently by scBS-seq reads,

Molecular Plant

Step 1: Fragment Amplification



Step 2: Ligated Library Construction



BRIF-Seq for Assessing Heterogeneity

Step 3: Reads Alignment

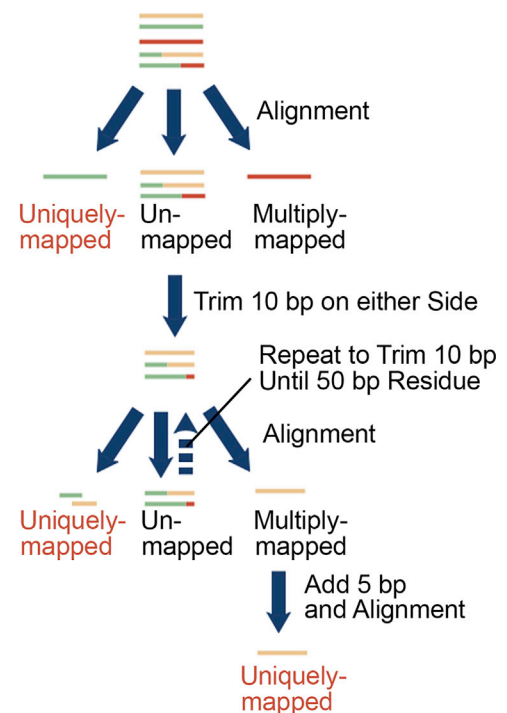


Figure 1. Schematic Representation of BRIF-Seq.

Both experimental and data analysis steps are included.

resulting in a potential bias in methylation calling (Supplemental Table 1).

It is known that bisulfite conversion changes unmethylated cytosine to uracil. During this treatment, a higher frequency of depyrimidination occurs at unmethylated rather than methylated cytosine sites, producing abasic sites and resulting in DNA fragmentation (Tanaka and Okamoto, 2007). Therefore, regions with a much lower (or higher) number of unmethylated cytosine sites could possibly be fragmented into longer (or shorter) segments during bisulfite conversion. Many genomes also contain a high percentage of transposons (occupying 85% of the maize genome), which are usually organized into nested structures and are silenced by high DNA methylation (Schnable et al., 2009; Jiao et al., 2017). These genomes usually contain successive methylated cytosine sites over a long distance, and will be fragmented into long segments during bisulfite treatment. During the random primer-based amplification step of scBS-seq (Smallwood et al., 2014), these long converted segments are amplified into a large number of long sequences and a small number of short sequences (Supplemental Figure 1). Thus, most of the fragments in these libraries cannot be sequenced on the HiSeq platform, leading to reduced genome coverage, especially in regions with high DNA methylation. It was also observed that the insertion size of an scBS-seq library covering hypermethylated regions was larger than that covering hypomethylated regions (Supplemental Figure 2). To maximize the sequencing of converted segments with different sizes and to reduce the read distribution bias, we designed a new method for single-cell DNA methylome

sequencing that is independent of DNA methylation state, called bisulfite-converted randomly integrated fragments sequencing (BRIF-seq). This method ensures the production of library fragments with uniform size and improves genome coverage. Single-cell BRIF-seq (scBRIF-seq) was tested on maize microspores with high DNA methylation levels. High rates of read mapping and genome coverage allowed us to identify heterogeneous sites, suggesting that BRIF-seq is an effective method for studying fundamental biological processes.

RESULTS AND DISCUSSION

Characterization of the BRIF-Seq Method

BRIF-seq consists of three steps (Figure 1). In step 1, bisulfite-converted DNA fragments were subjected to five rounds of amplification with untagged primers containing nine random nucleotides, resulting in a moderate number of sequences for the next step. We do not recommend more rounds of amplification since this may result in the overrepresentation of short fragments, complicating subsequent analysis. Step 1 is similar to the first part of scBS-seq, except for the use of untagged primers. In step 2, after removing the single-strand primers, double-stranded fragments were denatured and ligated together as single long strands. 5'-blocked primers containing nine random nucleotides were used in the subsequent multiple displacement amplification (MDA), avoiding ligation of primers to DNA templates by the residual ligase. The library was then constructed using Tn5 transposase. In step 2, the use of converted segments with different sizes was maximized, and

BRIF-Seq for Assessing Heterogeneity

Molecular Plant

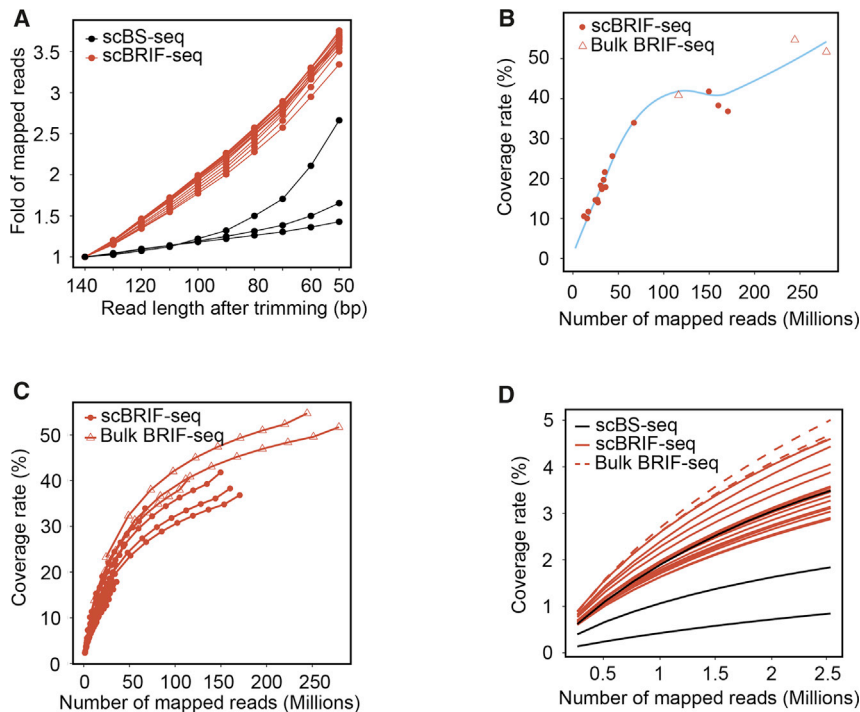


Figure 2. BRIF-Seq Improves Read Mapping and Genome Coverage.

(A) Read trimming increases the mapping rate of libraries constructed with scBRIF-seq. The fold increase in the number of mapped reads after trimming (130–50 bp) compared with the number of mapped full-length (140 bp) reads is shown. ScBS-seq libraries were used as a negative control since they were prepared without the ligation step and thus, nucleotide trimming was not expected to improve the mapping rate. Each line represents one library prepared using either scBS-seq or scBRIF-seq.

(B) Genome coverage rates for BRIF-seq libraries. Each dot represents one library. The x axis represents the total number of mapped reads for each sample.

(C) Random sampling of each library to show the correlation between genome coverage and number of mapped reads. For each BRIF-seq sample, a certain proportion of mapped reads (0%–100% with a step size of 10%) was randomly extracted and was used to estimate genome coverage.

(D) Comparison of genome coverage for libraries prepared with either BRIF-seq or scBS-seq. A maximum of 2.5 million mapped reads for each sample was extracted to estimate genome coverage. Each line represents one sample.

size-appropriate libraries that are suitable for sequencing on the Illumina platform were generated. In step 3, sequencing reads were aligned to the reference genome and divided into three classes: uniquely mapped, unmapped, and multiply mapped. Only the uniquely mapped reads were used for SNP and 5mC calling. Some reads contained multiple fragments from different chromosomal locations as a result of the ligation step during library construction and could not be mapped. To maximize the utilization of these ligated reads, we trimmed the unmapped reads by 10 bp on either side and remapped them, and this process was repeated multiple times until they could be uniquely mapped. Reads that were smaller than 50 bp after trimming were discarded. For reads mapped to multiple positions after trimming, 5 bp was added back to the trimmed reads, which were then remapped. These reads were kept if they could now be mapped uniquely and were discarded if not. We did not trim the reads further if they could not be mapped when trimmed down to 50 bp, since further trimming would increase coverage by just 0.45% (Supplemental Figure 3) and these reads were potentially prone to mapping error. All the uniquely mapped reads were used for subsequent analysis.

Using BRIF-seq, the genome-wide DNA methylation levels of 16 single microspores from four tetrads of an F_2 individual derived from a cross of Zheng58 and CF3 were assessed. Two bulk samples containing two tetrads from each of the two parents, and another bulk sample containing a dozen leaf protoplasts from Zheng58, were separately analyzed with BRIF-seq under the same conditions (Supplemental Table 2). As controls, three single microspores were subjected to scBS-seq. Through high-throughput sequencing, a total of 928 Gb of raw data were obtained (Supplemental Table 2). A higher fraction of trimmed reads from scBRIF-seq than from scBS-seq could be realigned to the reference genome (B73, v3.26) (Figure 2A). For example,

trimming to 50 bp led to an ~2.6-fold increase in mapped reads for scBRIF, which is significantly higher than that for scBS-seq libraries where fragment ligation was not used (Figure 2A). This confirmed that fragment ligation indeed occurs during BRIF-seq library construction. It is interesting to note that mapping success increased at a similar rate for different scBRIF-seq libraries, indicating the high reproducibility and reliability of scBRIF-seq.

Because increasing the read number improves genome coverage until library saturation, samples were sequenced at different depths (on average 632 million reads for tetrad 1; 147 million for tetrads 2–4; 503 million for bulk samples) to identify an optimal sequencing depth for BRIF-seq. The read mapping rate ranged from 7.0% to 45.0%, which is higher than that for scBS-seq (3.3%–4.1%; Supplemental Table 2). For each sample, the number of mapped reads ranged from 12 to 280 million (Supplemental Table 2). The genome coverage increased exponentially from 10.6% to 34.0% as the mapped read number increased from 12 to ~70 million. It then slowly climbed to 54.7% (for the bulk sample containing eight microspores) when the read number reached 244 million (Figure 2B). The highest coverage of a single-cell genome reached 41.8%. Thus 70 million mapped reads could be used to economically and efficiently reach a genome coverage of ~35%. Random sampling of the mapped reads from 0% to 100% with steps of 10% showed that 70 million mapped reads is an inflection point for increased genome coverage (Figure 2C). Considering an average mapping success of ~24%, the appropriate number of BRIF-seq reads for maize would be 290 million (70 million divided by 24%).

The coverage and resolution obtained by scBS-seq in mammals cannot be compared with those in maize, as the levels of genome methylation and repetitive sequence differ. Thus in this study,

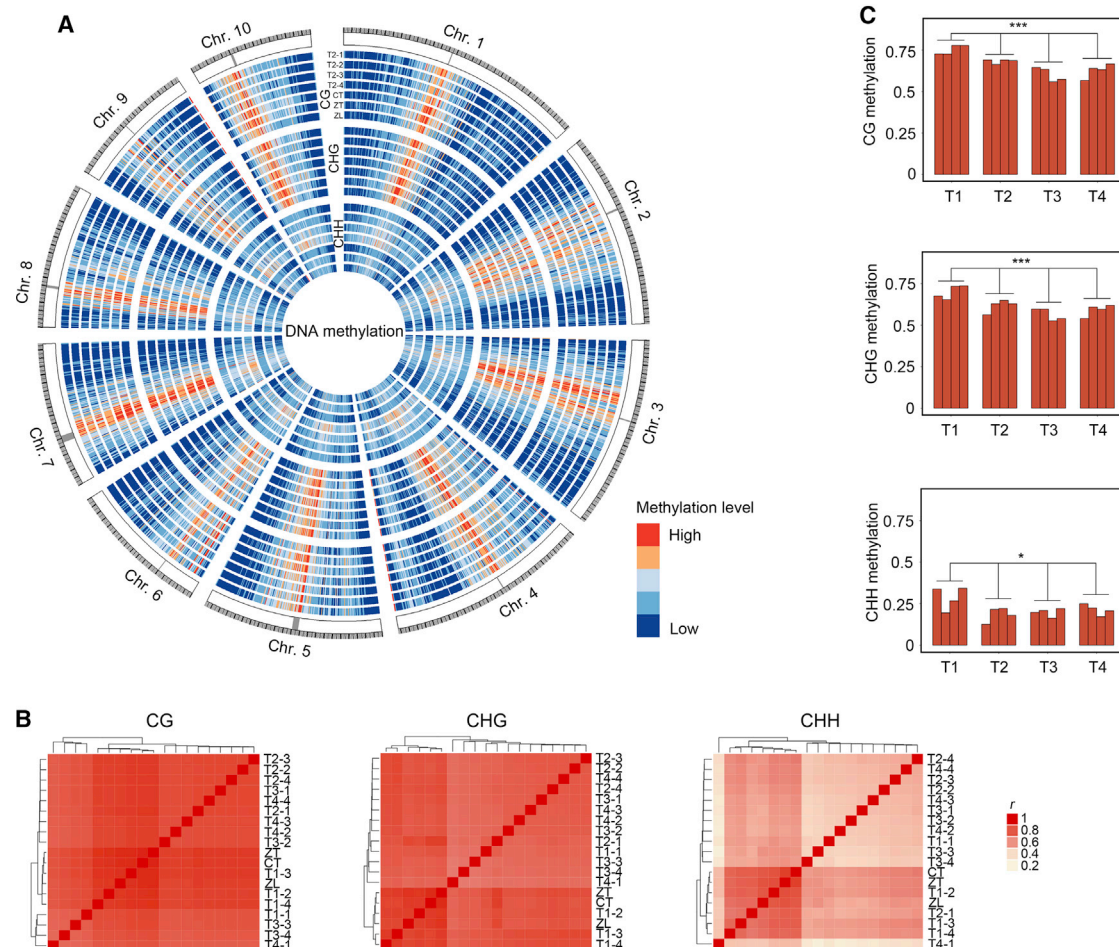


Figure 3. Overall DNA Methylation in Maize Microspore Determined by scBRIF-Seq.

(A) Genome-wide landscape of CG, CHG, and CHH methylation within 1-Mb windows.

(B) Heatmaps showing the Pearson correlation coefficients between samples for CG, CHG, and CHH methylation.

(C) The average levels of CG, CHG, and CHH methylation for each microspore. ****P* value of ANOVA between tetrads is less than 10⁻³; **P* value is less than 0.05. T1 to T4 are short for tetrads 1 to 4. T1-1, T1-2, T1-3, and T1-4 are the abbreviations for tetrad 1 microspores 1 to 4. CT, ZT, and ZL are the abbreviations for bulk CF3 tetrads, bulk Zheng58 tetrads, and bulk Zheng58 leaf cells, respectively.

both scBS-seq and scBRIF-seq were performed on maize microspores for a direct comparison. Owing to the limited number of scBS-seq reads, a comparable number of BRIF-seq reads were extracted randomly for comparison. The curves for different scBRIF-seq libraries were close to each other, and the coverage rates of the scBRIF-seq libraries were slightly lower than those of the bulk BRIF libraries, but were significantly higher than those of the scBS-seq libraries at all mapping depths (Figure 2D), implying that scBRIF-seq is a more efficient and stable method than scBS-seq for sequencing hypermethylated repetitive genomes, resulting in a higher mapping rate and genome coverage.

DNA Methylation Patterns in Maize Microspores

Using the BRIF-seq method, genome-wide CG, CHG, and CHH methylation levels were estimated in maize microspores (Figure 3A and Supplemental Figure 4). In theory, each cytosine site in a single microspore with a haploid genome must be completely methylated or unmethylated. Indeed, on average 91.9% of CG sites, 91.3% of CHG sites, and 93.4% of

CHH sites were identified as completely methylated or unmethylated in single-cell samples (Supplemental Figure 5A). The percentages of cytosine sites that were completely methylated or unmethylated were also calculated for different sequencing depths (Supplemental Figure 5B). At high depths (≥ 2), the ratio of completely methylated and unmethylated cytosine sites in single-cell samples was higher than that in bulk samples, especially for CG and CHG sites, suggesting that methylation heterogeneity exists between cells in the bulk samples.

The coverage rates of CG, CHG, and CHH loci reached 38.8%, 39.9%, and 37.0% (Supplemental Table 2), respectively, which is lower than the whole-genome coverage rate (54.7%). This may be partly attributed to the complexity of the maize genome and the fact that a number of unmethylated sequences were converted to multiply mapped reads. In the microspore samples, the methylation rates of CG and CHG were high, with an average of 66.7% (55.9%–77.9%) and 61.8% (52.6%–73.7%), respectively. CHH methylation rates were low, ranging

BRIF-Seq for Assessing Heterogeneity

from 12.6% to 34.3% with an average of 22.1%. CG and CHG methylation occur more frequently in peri-centromeric regions, while CHH methylation is relatively uniformly distributed along the entire chromosome (Figure 3A; for details see Supplemental Figure 4). In transposable elements (TEs), CG (average 82%, ranging from 73% to 87%) and CHG (average 73%, ranging from 60% to 80%) methylation is higher than CHH methylation (average 14%, ranging from 6% to 21%), which indicates that TEs may be silenced by high levels of CG and CHG methylation. These results are consistent with previous reports (Regulski et al., 2013; Walker et al., 2018). Furthermore, the high correlations of genome-wide CG (average Pearson's $r = 0.84$) and CHG (average Pearson's $r = 0.78$) methylation rates between each sample pair indicates the reliability of scBRIF-seq (Figure 3B). The methylation rates around genes and splicing sites were profiled. Consistent with previous observations (Regulski et al., 2013), the rate of CG, CHG, and CHH methylation dropped at transcriptional start sites and transcriptional termination sites, and the rate of CG methylation increased slightly within the gene body but was still lower than that in intergenic regions (Supplemental Figure 6). The level of CG methylation in exons was higher than that in introns, while CHG and CHH methylation remained low, with no difference between exons and introns (Supplemental Figure 7). The overall trends are consistent with results of a previous report (Regulski et al., 2013), confirming the high reliability of scBRIF-seq.

Surprisingly, methylation rates among the four microspores within one tetrad are similar, but CG (ANOVA $P = 2.2 \times 10^{-4}$), CHG (ANOVA $P = 1.8 \times 10^{-3}$), and CHH methylation (ANOVA $P = 0.040$) rates differ significantly between tetrads (Figure 3C). This suggests that methylation reprogramming could occur around the division into four microspores (Walker et al., 2018), with the reprogramming dynamics differing among tetrads. Additionally, the correlation among maize microspores observed for CG (average Pearson's $r = 0.84$), CHG (average Pearson's $r = 0.78$), and CHH (average Pearson's $r = 0.55$) methylation is still lower than that observed for CG methylation between mouse metaphase II oocytes (average Pearson's $r = 0.92$) (Smallwood et al., 2014). This implies the presence of epigenetic reprogramming in maize around the tetrad stage (Walker et al., 2018), as discussed in detail below.

Analysis of Heterogeneity in Maize Sexual Reproduction

DNA methylation reprogramming during sexual reproduction reduces the inheritance of epi-phenotypes. To assess the heterogeneity of haploid microspores, we subjected 177 035 641 cytosine sites (18.38% in total) detected by scBRIF-seq to strict filtering (see Methods). This led to the identification of 99 348 413 cytosine sites, with 16 163 929 in CG, 12 722 559 in CHG, and 70 461 925 in CHH contexts. These microspores were derived from one hybrid F_2 individual with 25% of genomic regions predicted to be homozygous for the CF3 background, 25% homozygous for the Zheng58 background, and 50% heterozygous for the CF3 and Zheng58 backgrounds. Genetic backgrounds were profiled using an average of 69 498 SNPs for each microspore (Supplemental Figure 8). As expected, 50.2% of the genome exhibited Mendelian segregation in a tetrad, with 39.70% of the regions being

homozygous for either the CF3 or Zheng58 background. This suggests that BRIF-seq is highly sensitive to genetic background and appropriate for the analysis of samples with diverse backgrounds. Heterogeneity in each microspore was defined as changes in methylation state, including methylation loss and gain, compared with the parents in either segregating or non-segregating regions (Figure 4A). In theory, methyltransferase- and demethylase-induced changes should occur in clusters. Thus the adjacent heterogeneous sites that could be assigned with high confidence were retained during analysis. In total, a median of 24 749 CG sites (1.2%), 21 805 (1.3%) CHG sites, and 87 273 (1.5%) CHH sites (Figure 4B) were found to be heterogeneous. These numbers are significantly higher than the random expectations for each microspore (Supplemental Figure 9). Moreover, high proportions (8.3%–10.8% for CG, 7.8%–8.5% for CHG, and 5.8%–6.0% for CHH) of adjacent heterogeneous sites coexisted in at least two microspores, and these proportions are significantly higher than the expectation (less than 1%) (Supplemental Figure 10). The sum of adjacent heterogeneous sites across all microspores represents 6.9% (435 853) of CG, 7.4% (387 128) of CHG, and 7.0% (1 667 084) of CHH sites (Figure 4C). Together these data suggest that methylation reprogramming does occur during male gametogenesis.

A median of 88.6%, 93.3%, and 98.9% of heterogeneous CG, CHG, and CHH sites, respectively, represented methylation gain (Figure 4D), while a high proportion of methylation loss occurred in two microspores, tetrad 1-2 and tetrad 2-1, in which 53.7% and 60.2% of CG sites and 39.7% and 62.2% of CHG sites lost methylation, respectively. These results suggest that microspores undergo extensive methylation changes during the non-synchronous reprogramming process, as has been reported to occur in plants during reproduction (Walker et al., 2018). As mentioned above, the microspores were from an F_2 individual, so we had the opportunity to assess whether heterogeneity is affected by genetic background. Among segregating and background-specific non-segregating regions, the rates for adjacent heterogeneous sites for each microspore were similar (Figure 4B), as were the overall rates for adjacent heterogeneous sites and the proportions of methylation gain and loss (Figure 4C and 4D). This suggests that the level of heterogeneity during reprogramming is seldom affected by genetic background, although methylation state was reported to be heavily dependent on genetic background. To estimate the bias in the identification of reprogrammed sites in different genomic regions, we profiled the distribution of adjacent heterogeneous sites using all the detected cytosine sites as a control (Figure 5). Most of the adjacent heterogeneous sites were located within/near genic regions except CHH. Interestingly, adjacent heterogeneously methylated cytosine sites were less likely to occur in transposons (Figure 5, 9.3% versus 17.1% for CG, 7.9% versus 13.8% for CHG, and 32.5% versus 44.4% for CHH) (chi-square tests, all $P = 0$). These results suggest that heterogeneity due to reprogramming is uncommon in transposon sequences. The biased distribution of heterogeneous sites between genomic elements was also observed in segregating and background-specific non-segregating regions (Supplemental Figure 11). This provides further support that heterogeneity is triggered by reprogramming and is seldom affected by genetic background.

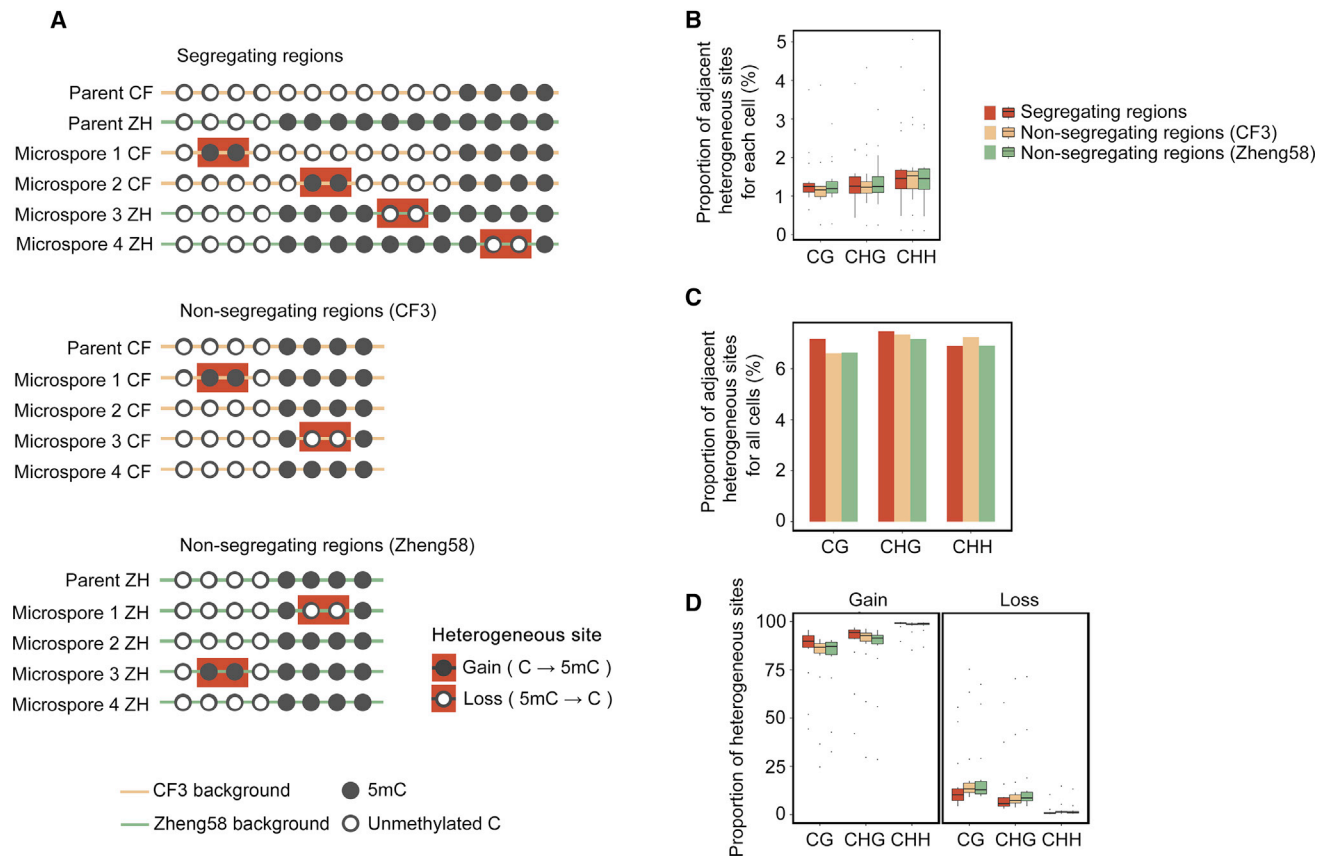


Figure 4. Methylation Heterogeneity during Maize Male Gametogenesis.

(A) Definition of heterogeneity.

(B) Box plot illustrating the proportion of adjacent heterogeneous sites in each microspore.

(C) The proportion of non-redundant adjacent heterogeneous sites from all microspores.

(D) The ratio of methylation gain and loss is shown for heterogeneous CG, CHG, and CHH sites.

In (B) to (D), the proportions of segregating (in red) and non-segregating regions with CF3 (in yellow) or Zheng58 (in green) backgrounds were calculated. In (B and D), the horizontal line represents the median, and vertical lines mark the range from the 5th and 95th percentile of the total data.

In summary, this work introduces BRIF-seq for single-cell methylome analysis. scBRIF-seq reads have higher coverage than scBS-seq in regions with high CG and CHG methylation (Figure 6A), while scBRIF-seq and scBS-seq reads share similar coverage in hypomethylated regions (Supplemental Figure 12). In addition, scBRIF-seq can cover regions with a

wide range of GC content with no obvious bias (Figure 6B). These results suggest that BRIF-seq reads could uniformly cover the genome independently of methylation state. We also showed that dynamic heterogeneity occurs during maize sexual reproduction, and scBRIF-seq provides an opportunity for further dissection of this process.

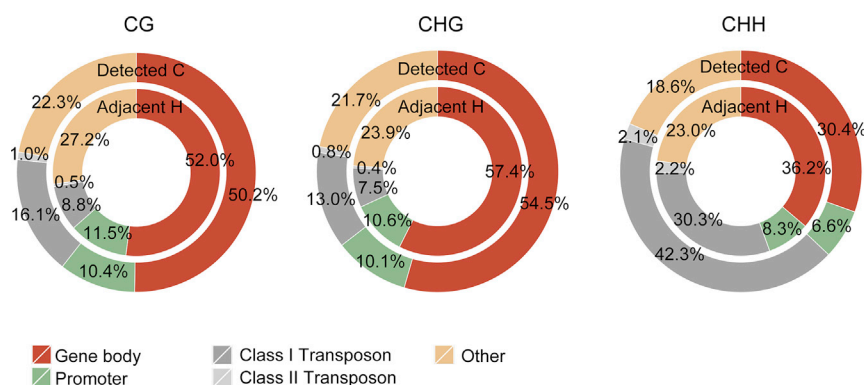


Figure 5. Distribution of Methylation Heterogeneity in Different Genome Features.

The inner rings represent the proportions for adjacent heterogeneous sites (Adjacent H), while the outer rings represent all detected cytosine sites (Detected C), which serve as a control.

BRIF-Seq for Assessing Heterogeneity

Molecular Plant

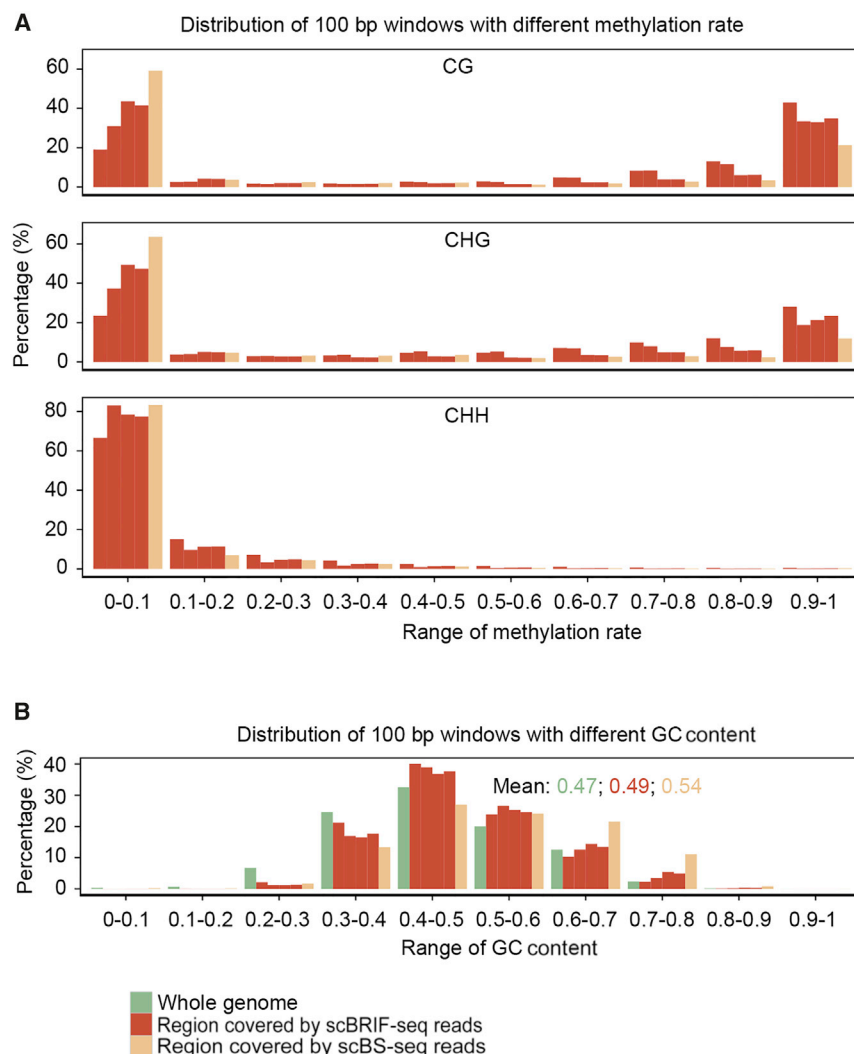


Figure 6. Even Distribution of scBRIF-Seq Reads.

The 100-bp windows covered by at least two reads (depth ≥ 2) were defined for each sample. Methylation rate and GC content for each window were calculated.

(A) The distribution of 100-bp windows with different levels of DNA methylation.

(B) Distribution of 100-bp windows with different GC contents. The GC content of windows covered by scBRIF-seq reads (0.49) approximates the whole-genome level (0.47), but the GC content for scBS-seq is higher (0.54) than the whole-genome level. Each red bar represents the average of four microspore sets from one tetrad sequenced by scBRIF-seq, and each yellow bar represents the average of three single microspore sets sequenced by scBS-seq.

The mixture was incubated at 37°C for 1 h for cell lysis. Bisulfite conversion was performed using the Imprint DNA Modification Kit (Sigma-Aldrich Merck, Darmstadt, Germany) as follows. Balance Solution (1 μ l) was added, followed by incubation at 37°C for 15 min. DNA modification solution was prepared by adding 1.1 ml of DNA modification solution and 40 μ l of Balance Solution to the bottle of DNA modification powder. Then 62.5 μ l of the DNA modification agent was added to the DNA samples, and bisulfite conversion was performed by incubating the samples successively at 65°C for 1.5 h, 95°C for 3 min, 65°C for 20 min, and 4°C for 1 min. The converted DNA fragments were purified using a PureLink PCR Micro Kit (Thermo Fisher, Waltham, MA, USA). The purified DNA (20 μ l) was added into the primary amplification mix (2.5 μ l of 10 \times NEBuffer 2, 1 μ l of 10 mM deoxynucleotide triphosphate [dNTP] mix, 1 μ l of 10 μ M oligo

NNNNNNNNN), incubated at 65°C for 3 min, and put on ice. Then 1 μ l of Klenow fragment (high concentration, M0212M; NEB, Ipswich, MA, USA) was added. The following steps were performed four times: incubating at 4°C for 5 min, increasing temperature by 1°C per 15 s to 37°C, incubating at 37°C for 30 min, 4°C for 4 min, and 95°C for 45 s, keeping on ice for 1 min, and adding the second amplification mix (0.25 μ l of 10 \times NEBuffer 2, 0.1 μ l of 10 mM dNTP, 1 μ l of 10 μ M oligo NNNNNNNNN, 0.5 μ l of Klenow fragment, 0.65 μ l of water). After the above steps were repeated four times, the sample was incubated at 4°C for 5 min, the temperature was increased by 1°C per 15 s to 37°C, then the sample was incubated at 37°C for 90 min and 4°C for 10 min. The resulting amplified product was purified using 0.9X AmpureXP beads (Beckman Coulter, Brea, CA, USA). Then 10 μ l of purified DNA was incubated at 95°C for 45 s and transferred onto ice for 1 min. The ligation mix (2.35 μ l of water, 1 μ l of T4 RNA ligase 1, 1.5 μ l of T4 10 \times buffer, 0.15 μ l of 100 mM ATP) was added, and the sample was thoroughly mixed and incubated at 37°C for 30 min. Thereafter, the MDA mix (5 μ l of 10 \times Phi29 buffer, 10 μ l of 100 μ M oligo 5'-biotin-NNNNNNNNN, 14.5 μ l of water, 2.5 μ l of 10 mM dNTP, 1 μ l of 100 \times BSA, 2 μ l of Phi29 polymerase) was added, and the sample was incubated at 30°C for 2.5 h. The polymerase was inactivated by incubating at 65°C for 10 min. The MDA product was purified using 0.9X AmpureXP beads (Beckman Coulter) to produce a 12- μ l sample. One microliter of sample was used to measure DNA concentration with a Qubit (Thermo Fisher). The sequencing library was made using the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme Biotech, Nanjing, China). Water was added to 5 ng of purified DNA up to a volume of 11 μ l.

METHODS

Cell Collection

The F₂ individuals derived from a cross between inbred lines Zheng58 and CF3 were planted in the field. Immature anthers were collected and placed in a drop of isolation buffer (27% sorbitol) on a slide. Immature pollens were released and those at the tetrad stage were selected. For the single-cell samples, a single tetrad was transferred into a new drop of isolation buffer using a micropipette. Four microspores were separated by repeated aspiration with the micropipette. A single microspore was then transferred to 4 μ l of water in a PCR well. For the bulk tetrad samples, two tetrads from each parent were transferred to 4 μ l of water in a PCR well. For the bulk protoplast sample, young leaves from Zheng58 were cut to pieces, incubated with 10 ml of enzyme mixture (12.5 g/l cellulase R10, 3 g/l macerozyme R10, 0.4 M mannitol, 20 μ M KCl, 20 μ M 2-(N-morpholino)ethanesulfonic acid, 10 mM CaCl₂, 0.0342% mercaptoethanol, and 1 g/l BSA) at room temperature for 4 h. Separated protoplasts were observed under a microscope, and a dozen protoplasts were transferred to 4 μ l of water in a PCR well. These samples were either put on ice for library construction or stored at -80°C.

BRIF Library Construction

Eight microliters of lysis buffer (15 mM Tris-Cl [pH 7.4], 0.9% SDS, and 0.5 μ l of proteinase K) was added to 4 μ l of sample prepared as described above.

Molecular Plant

Four microliters of 5× TTBL [TruePrep Tagment Buffer L] and 5 µl of TTE [TruePrep Tagment Enzyme] Mix V5 was added, and the sample was mixed thoroughly, incubated at 55°C for 10 min, and held at 10°C. Five microliters of 5× TS [Terminate Solution] was added, the sample was mixed thoroughly to stop the reaction, and the PCR mix was added (4 µl of water, 10 µl of 5× TAB [TruePrep Amplify Buffer], 5 µl of Index N5, 5 µl of Index N7, and 1 µl of TAE [TruePrep Amplify Enzyme]). The mixture was incubated at 72°C for 3 min and 98°C for 30 s. Three steps of incubation at 98°C for 15 s, 60°C for 30 s, and 72°C for 3 min were repeated for 10 cycles. The reaction was then incubated at 72°C for 5 min and held at 4°C for 1 min. Following library amplification, a 0.6-fold volume of AmpureXP beads was added into the PCR product, mixed thoroughly, and incubated at room temperature for 5 min. After placing the sample onto a magnet for 5 min, the supernatant was transferred to a new well with a 0.15-fold volume of AmpureXP beads, mixed thoroughly, and incubated at room temperature for 5 min. After placing the sample onto a magnet for 5 min, the supernatant was removed. The beads were washed twice with 80% ethanol, dried, and resuspended in elution buffer. After removing the beads, the library with a median size of 330 bp was selected and pair-end sequenced on the Illumina HiSeq 3000 platform.

Alignment of Reads Containing Ligated Sequences

Raw reads were trimmed to remove the low-quality data with Trim_Galore (v0.4.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The trimmed reads were aligned to the B73 reference genome v3 under the single-end mode using Bismark (v0.14.5; -bowtie2 -non_directional -score_min L,0,-0.3 -ambiguous -un -N 1) (Krueger and Andrews, 2011). Duplicate reads were removed, and all uniquely mapped reads were used for downstream analysis.

Methylation and Genetic Marker Calling

The methylation state and C/T read number for each cytosine site were calculated using bismark_methylation_extractor, a tool in the Bismark software. The scBRIF-seq conversion rate was estimated based on the rate of unmethylated cytosines in the chloroplast genome. SAMtools (v0.1.19) (Li et al., 2009; Li, 2011) was used to call SNPs for each sample. Because the G/A and C/T polymorphisms (reference/variant) in bisulfite-converted DNA could not be distinguished between true SNPs and bisulfite-converted SNPs, these sites were not used for SNP calling. To obtain high-quality SNPs, we filtered the identified SNPs based on mapping quality (MapQ ≥ 30.0), minimum coverage (DP ≥ 2) and maximum coverage (DP ≤ 200) thresholds.

Identification of Genetic Background

The maize genome was divided into 1-Mb non-overlapping bins, and the genetic background of each bin for each single microspore was determined by calculating the percentage of SNPs derived from a particular parent. We defined the ratio of CF3-derived SNPs in window i for each microspore j as $r_{i,j} = C_{i,j}/(C_{i,j} + Z_{i,j})$, where $C_{i,j}$ is the sum of SNPs with CF3 alleles and $Z_{i,j}$ is the sum of SNPs with Zheng58 alleles. We then calculated the difference in $r_{i,j}$ between the four microspores from a single tetrad k as $d_{i,j,k} = \max(r_{i,j,k}) - \min(r_{i,j,k})$. If $d_{i,j,k}$ was higher than 0.4, the genetic backgrounds of the two microspores with higher $r_{i,j}$ were defined as CF3-derived and the other two were defined as Zheng58-derived. On the other hand, if $d_{i,j,k}$ was smaller than 0.3, all microspores in tetrad k were defined as sharing the same genetic background. The backgrounds were identified as CF3-derived if the number of microspores exhibiting $r_{i,j} > 0.5$ was two or more; otherwise they were identified as Zheng58-derived.

Enrichment of DNA Methylation Heterogeneity

To assess heterogeneity, we filtered the completely methylated and unmethylated cytosine sites based on the following criteria: (1) at least two reads available for each sample; (2) more than 90% of the reads have the same methylation states in each sample; (3) detected in at least two microspores and have differences in methylation among microspores;

BRIF-Seq for Assessing Heterogeneity

(4) adjacent cytosine sites have the same methylation status in parental lines. The identified heterogeneous sites were assigned to the following four groups: within gene body, gene promoter, transposons, and other. Gene annotation was based on the B73 reference genome V3.29. The 2000-bp region upstream of transcriptional start site for each gene was defined as the promoter. Transposon information was extracted from the B73 V3.29 annotations using the sources from Repeatmasker. Cytosine sites that could not be assigned to gene, promoter, or transposons were classified as “other”

Data Availability

Raw reads from this study have been deposited in the NCBI Sequence Read Archive under accession number SRP139279.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This research was supported by the National Natural Science Foundation of China (31730064, 31801125), National Key Research and Development Program of China (2016YFD0101003), National Postdoctoral Program for Innovative Talents (BX201700091), China Postdoctoral Science Foundation (2017M620325), the Hubei Provincial Natural Science Foundation of China (2015CFA008), and the Huazhong Agricultural University Scientific & Technological Self-innovation Foundation.

AUTHOR CONTRIBUTIONS

X.L. designed the BRIF-seq method. J.Y. and Q.L. supervised this study. X.L., Q.Z., and Y.S. performed the experiments. L.C. analyzed sequencing data. X.L., L.C., Q.L., and J.Y. prepared the manuscript.

ACKNOWLEDGMENTS

The authors declare that they have no competing interests.

Received: October 8, 2018

Revised: December 12, 2018

Accepted: January 6, 2019

Published: January 10, 2019

REFERENCES

- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161**:1437–1452.
- Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **23**:2126–2135.
- Hui, T., Cao, Q., Wegrzyn-Woltosz, J., O'Neill, K., Hammond, C.A., Knapp, D.J.H.F., Laks, E., Moksa, M., Aparicio, S., Eaves, C.J., et al. (2018). High-resolution single-cell DNA methylation measurements reveal epigenetically distinct hematopoietic stem cell subpopulations. *Stem Cell Rep.* **11**:578–592.
- Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B., Campbell, M.S., Stein, J.C., Wei, X., Chin, C.S., et al. (2017). Improved maize reference genome with single-molecule technologies. *Nature* **546**:524–527.
- Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**:484–492.
- Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**:1571–1572.
- Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**:204–220.

BRIF-Seq for Assessing Heterogeneity

Molecular Plant

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**:2078–2079.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993.
- Luo, C., Keown, C.L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J.R., Sandoval, J.P., et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**:600–604.
- Mulqueen, R.M., Pokholok, D., Norberg, S.J., Torkenczy, K.A., Fields, A.J., Sun, D., Sinnamon, J.R., Shendure, J., Trapnell, C., O’Roak, B.J., et al. (2018). Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* **36**:428–431.
- Regulski, M., Lu, Z., Kendall, J., Donoghue, M.T., Reinders, J., Llaca, V., Deschamps, S., Smith, A., Levy, D., McCombie, W.R., et al. (2013). The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Res.* **23**:1651–1662.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**:1112–1115.
- Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**:817–820.
- Tanaka, K., and Okamoto, A. (2007). Degradation of DNA by bisulfite treatment. *Bioorg. Med. Chem. Lett.* **17**:1912–1915.
- Umer, M., and Herceg, Z. (2013). Deciphering the epigenetic code: an overview of DNA methylation analysis methods. *Antioxid. Redox Signal.* **18**:1972–1986.
- Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D., and Feng, X. (2018). Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis*. *Nat. Genet.* **50**:130–137.
- Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., et al. (2018). Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet.* **50**:12–19.