nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-021-00630-1

Check for updates

Genome-wide quantification of transcription factor binding at single-DNA-molecule resolution using methyl-transferase footprinting

Rozemarijn W. D. Kleinendorst[1,4], Guido Barzaghi[1,2,4], Mike L. Smith[1], Judith B. Zaugg[3] and Arnaud R. Krebs[1]✉

Precise control of gene expression requires the coordinated action of multiple factors at *cis*-regulatory elements. We recently developed single-molecule footprinting to simultaneously resolve the occupancy of multiple proteins including transcription factors, RNA polymerase II and nucleosomes on single DNA molecules genome-wide. The technique combines the use of cytosine methyltransferases to footprint the genome with bisulfite sequencing to resolve transcription factor binding patterns at *cis*-regulatory elements. DNA footprinting is performed by incubating permeabilized nuclei with recombinant methyltransferases. Upon DNA extraction, whole-genome or targeted bisulfite libraries are prepared and loaded on Illumina sequencers. The protocol can be completed in 4–5 d in any laboratory with access to high-throughput sequencing. Analysis can be performed in 2 d using a dedicated R package and requires access to a high-performance computing system. Our method can be used to analyze how transcription factors cooperate and antagonize to regulate transcription.

Introduction

Transcription factors (TFs) modulate transcription through the recruitment of coactivators and RNA polymerase II (Pol II) at the promoters of genes. There are several technologies available to directly measure the binding of transcriptional regulators (i.e., ChIP-seq[1], CUT&RUN[2]) or to indirectly infer protein occupancy through footprints in chromatin accessibility (DNase I hypersensitive sites sequencing (DNase-seq[3]), Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq[4])). These methods have led to extensive insights into the identity of TFs involved in *cis*-regulatory element (CRE) activation in various cell types and tissues. Most TFs are unable to activate their target CRE alone. Cooperativity between TFs has been shown to be an essential mechanism used by TFs to bind and activate CREs[5–8]. Most genomics methods used to measure TF binding are bulk assays that typically average binding information from millions of cells. Most of these assays enrich for a single feature of interest (e.g., TF, chromatin mark), disregarding the potential co-occurrence of other binding events and ignoring potential heterogeneity of occupancy at CREs.

Recently, several approaches have been developed that employ exogenous DNA methyltransferases to footprint protein–DNA contacts in the genome[9–11]. These approaches were shown to accurately quantify DNA occupancy by nucleosomes[9], TFs[8], general TFs and Pol II[10]. Coupling methylation footprinting with various sequencing technologies has made it possible to resolve protein–DNA contacts continuously over several hundreds[8,10] to several thousand[12–15] base pairs on individual DNA molecules. This unprecedented resolution has enabled new insights into transcription initiation dynamics[10], TF cooperativity[8], transcriptional coordination[13] and chromatin fiber organization[12,14].

Here we describe single-molecule footprinting (SMF), which we recently developed to resolve the occupancy of multiple TFs and Pol II simultaneously on single DNA molecules[8,10]. The technique combines the use of cytosine methyltransferases to footprint the genome with bisulfite sequencing to resolve the molecular binding patterns of TFs at CREs. Footprinting is performed on permeabilized

[1]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. [2]Faculty of Biosciences, Collaboration for Joint PhD Degree between EMBL and Heidelberg University, Heidelberg, Germany. [3]European Molecular Biology Laboratory (EMBL), Structural and Computational Biology Unit, Heidelberg, Germany. [4]These authors contributed equally: Rozemarijn W. D. Kleinendorst, Guido Barzaghi.
✉e-mail: arnaud.krebs@embl.de

nuclei using commercially available recombinant methyltransferases. Bisulfite libraries (whole-genome or prepared using an optional DNA capture step to enrich sequences of interest) can be generated using commercial protocols and are sequenced on Illumina MiSeq or NextSeq sequencers.

Continuity in footprinting information allows the study of whether binding events occur simultaneously at CREs with molecular resolution. Specifically, it enables quantification of the degree of co-occupancy of TFs on the same DNA molecules and linking their binding in a way that is impossible with bulk data[16] (discussed in 'Advantages'). We have successfully used this strategy to study co-occupancy of TFs, to identify dependencies between TFs and to reveal cooperativity mechanisms underlying their action at CREs[8].

Overview of the protocol

SMF requires the extraction and permeabilization of nuclei from cell lines or tissues. Purified nuclei are sequentially incubated with recombinant methyltransferases that methylate GpCs (M.CviPI) and possibly CpGs (M.SssI) (Fig. 1a). To obtain reproducible methylation footprints, it is essential to carefully quantify the number of cells used to maintain a constant enzyme/DNA substrate ratio. The number of cells to be used has to be adjusted per species according to its genome size (i.e., 2.5×10^6 for *Drosophila* and 0.25×10^6 for mouse or human). DNA is extracted, sheared into large molecules (300–500 bp) by sonication and bisulfite converted for whole-genome DNA methylation profiling. When whole-genome profiling is not suitable, several targeted SMF strategies can be implemented (see 'Sequencing strategy and coverage requirements'). For studying TF binding in the mouse genome, we added a DNA capture step using a library of RNA baits tiling 297,000 CREs. This step enriches libraries for regions of interest prior to bisulfite conversion and reduces the sequencing effort to 2% of the genome. With this strategy, a molecular coverage compatible with single-molecule analysis (>40×) can be reached at a large majority of TF binding sites with a reasonable sequencing effort (200×10^6 reads) (Fig. 1b,c). An alternative strategy consists of designing primers against regions of interest to prepare amplicon libraries that lead to very high molecular coverage (>1,000×) at defined regions with limited sequencing effort (1×10^6 reads) (Fig. 1d,e).

Advantages

Most genomics methods used to measure protein–DNA interactions such as ChIP-seq, CUT&RUN, DNase-seq or ATAC-seq are bulk assays that average binding signals over millions of cells. These assays are based on the selective sequencing of protein-bound DNA fragments following their fragmentation and enrichment. This enrichment step implies that only the protein-bound DNA molecules are quantified; however, potential heterogeneity such as the competition between nucleosomes and TFs would be ignored. Moreover, all these protocols disrupt the chromatin template, thus precluding the measurement of multiple factors interacting with DNA. In SMF, molecules are sequenced regardless of their accessibility status. Thus, at any given locus, the competitive occupancy by TFs and nucleosomes can be quantified simultaneously, providing valuable information on the frequency of CRE usage in cellular populations[8]. Moreover, deposition of methylation in SMF preserves the integrity of DNA, allowing quantification of the footprints created by multiple proteins over a stretch of 300–500 bp of a DNA molecule. This information can be used to infer dependencies between binding events in the genome and has, for instance, allowed us to resolve the mechanism of TF cooperativity in vivo[8].

Recently, single-cell protocols have been developed for most genomics assays[17], resolving the heterogeneity of CRE usage in individual cells. The generated data have sufficient resolution to precisely infer the cell type composition of heterogeneous populations[18,19]. However, information per single cell is sparse, and a given CRE rarely has more than a couple of informative reads per cell which is insufficient to dissect the logic of protein binding events at CREs[1617]. SMF provides complementary information to single-cell approaches as it resolves details of the molecular occupancy patterns at CREs[17]. In turn, SMF comes as a method of choice when dissecting molecular mechanisms regulating transcription.

Limitations

In SMF, protein–DNA contacts are detected as DNA regions that are protected from the exogenous methylation signal. As with any footprinting method, SMF is agnostic to the identity of the protein creating the footprints. Thus, interpretation of the SMF signal requires the integration of other sources of information. For instance, we have demonstrated that combining TF recognition motifs

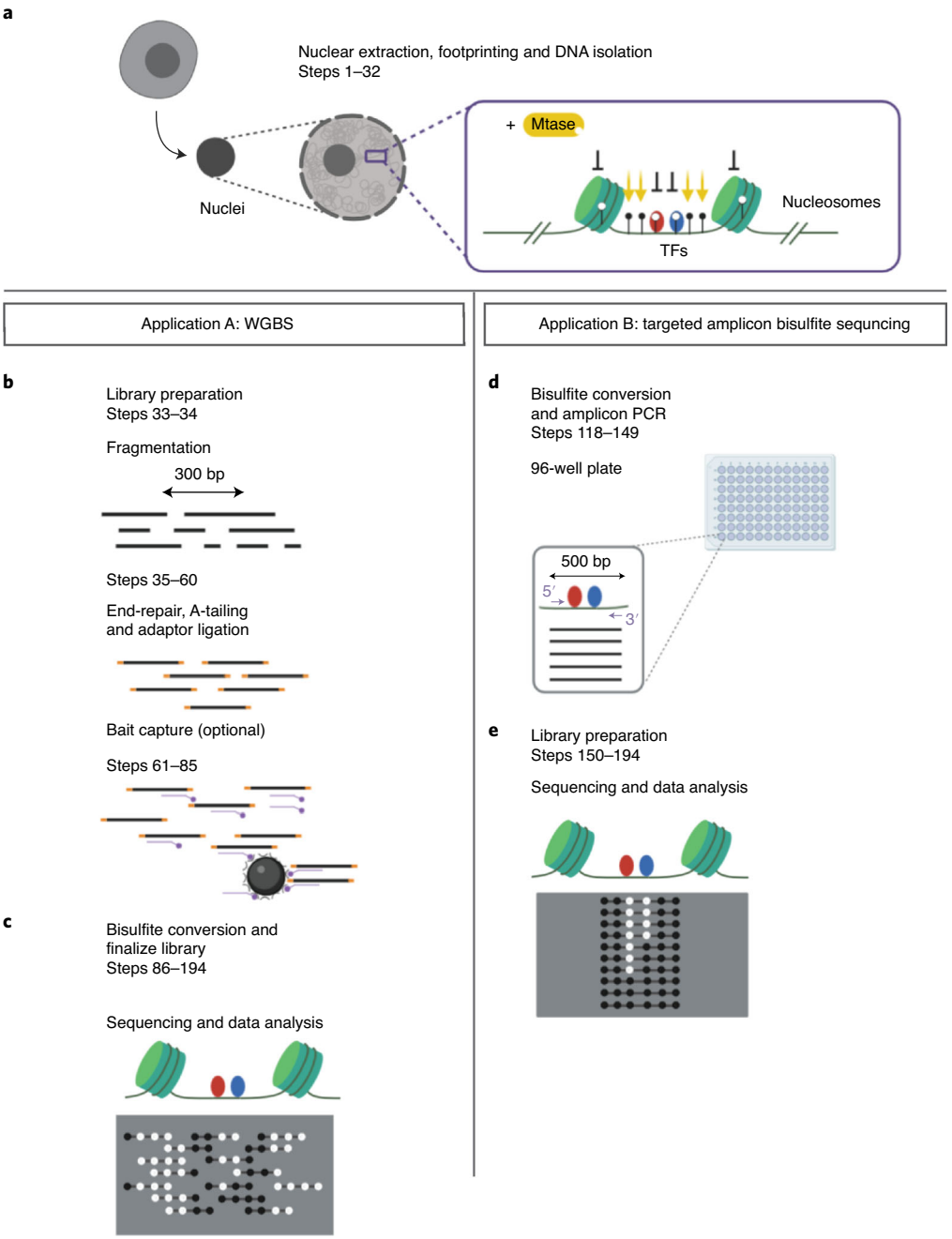

Fig. 1 | Overview of the experimental workflow. a, Nuclei are extracted using a hypotonic buffer. Methylation footprinting is performed by incubating the nuclei with a GpC (M.CviPI) and, optionally, CpG (M.SssI) methyltransferase (Mtase). Regions accessible to the enzymes are methylated, while regions bound by proteins (TFs, nucleosomes) are protected, creating footprints of various sizes. DNA is extracted and used for whole-genome (left) or targeted amplicon (right) analysis. **b**, For whole-genome analysis, DNA is fragmented to a target size range of 300–500 bp. DNA is end-repaired, and sequencing adapters are ligated. An optional capture step can be performed to enrich the library for regions of interest such as CREs and reduce the sequencing depth required for single-molecule analysis. **c**, DNA is bisulfite converted and the library is amplified before sequencing on Illumina MiSeq and NextSeq platforms. **d**, Alternatively to the whole-genome approach, primers can be designed to target 96 loci using amplicon bisulfite PCR. Amplicons are typically designed to cover 300–500 bp of the CRE. **e**, Amplicons are pooled, and the library is prepared. Up to 12 libraries can be multiplexed and sequenced on a MiSeq instrument. The read ends in amplicon data are identical for every molecule, creating focused high-coverage views of the targeted loci.

and ChIP-seq data can be used to accurately identify the TFs creating footprints detectable by SMF[8]. Similarly, we have shown that footprints created by general TFs and Pol II at core promoters can be identified by their relative position to transcriptional start sites as defined by CAGE data[10]. It is

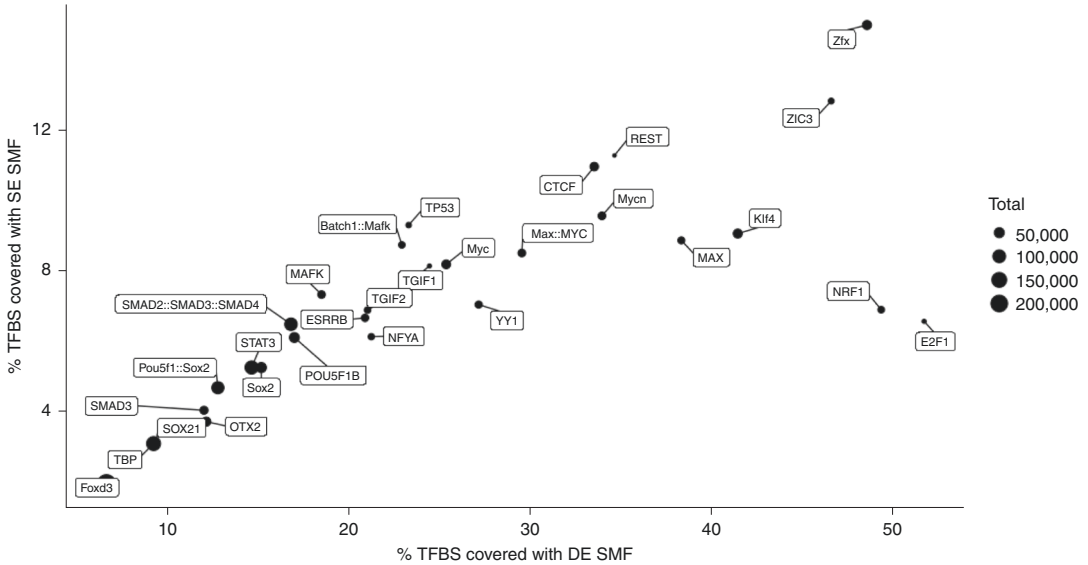

Fig. 2 | Number of TFBSs that can be studied by SMF. Classification of the single molecules at a TFBS requires the presence of informative cytosines in each of the classification bins. The scatterplot shows the percentages of TFBSs that can be analyzed when performing SMF with the GpC methyltransferase M.CviPI (single enzyme, SE; *y* axis) or in combination with the CpG methyltransferase M.SssI (dual enzyme, DE; *x* axis). The percentages are calculated with respect to the total number of TFBSs (dot size) mapped to the mouse genome using JASPAR[26] positional weight matrices and confirmed via publicly available ChIP-seq evidence (the datasets used are detailed in Table S1 of Sönmezer et al.[8]). For TFs such as NRF1, E2F1 and Klf4, there is quite a clear advantage in performing DE SMF as compared with SE SMF.

therefore recommended that SMF be applied to understand the dependencies between binding events for which the identity of the factors and their binding location is documented by orthogonal methods. Moreover, confirming the identity of the factor through downregulation is advisable to unambiguously identify the factors creating the footprints[8,10]. As SMF measures protein–DNA contacts, it is intrinsically unable to resolve footprints from factors that regulate CREs but do not directly contact DNA.

SMF uses M.CviPI and/or M.SssI that methylate cytosines in GpC and CpG context, respectively. This implies that molecular accessibility can only be resolved at regions containing sufficient density of these dinucleotides. The use of methylation in the CpG context is restricted to biological systems where endogenous DNA methylation is absent at these sites. This is the case for flies and embryonic stem cells that can proliferate in absence of endogenous DNA methylation. We empirically defined that the footprints created by TFs or Pol II span 15–20 bp[8,10]. This is compatible with the resolution of SMF performed using either the GpC methyltransferase only (~14 bp) or both enzymes (~7 bp). However, dinucleotide distribution is not even across the genome. For a given genome, only a fraction of the binding sites for every TF (i.e., ~20% for repressor element 1-silencing transcription factor (REST)[8]) or Pol II pausing sites (~40% in flies[10]) will be analyzable by SMF. It is therefore important to analyze the dinucleotide compositions of the regions of interest prior to SMF profiling (see 'Number of TF binding sites analyzed' under 'Anticipated results', Fig. 2).

DNAse-seq and ATAC-seq are based on the selective sequencing of accessible regions of the genome, while SMF sequences all DNA molecules regardless of their accessibility status. Performing SMF is substantially more expensive than other DNA footprinting methods. Costs are in large part attributed to the requirement of high coverage (>40×) for sound statistical analysis of binding frequencies. This problem is enhanced for mammalian genomes that are 20 times bigger than fly genomes. We have developed several strategies for targeted SMF that enable cost-efficient high coverage SMF on hundreds (PCR-based[8,10]) to tens of thousands of loci (bait capture on mouse CREs[8]), thereby focusing the sequencing efforts to the regulatory regions that represent only 5–10% of mammalian genomes.

Applications

SMF has been applied to several *Drosophila* and mouse cell lines successfully. Moreover, various protocols for methylation footprinting have been developed and used in yeast[20,21] and humans[9,11]. In

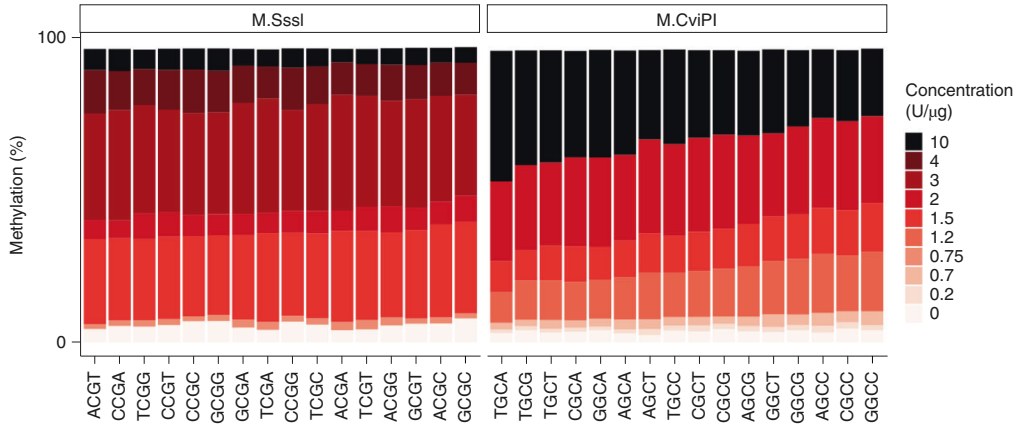

Fig. 3 | Methylation efficiency of M.SssI and M.CviPI is not affected by the sequence context when saturating conditions are used. In vitro methylation of naked lambda DNA using various concentrations of M.SssI (left) or M. CviPI (right) shows moderate sequence preferences at nonsaturating enzyme concentrations (up to 2 units/μg of DNA). Importantly, these differences become negligible under saturating conditions (>10 units/μg of DNA), such as the ones used during SMF experiments (200 units/μg of DNA).

principle, SMF can be adapted to any cell type or tissue for which nuclei can be purified and permeabilized. Efficient footprinting depends on a homogeneous nuclear extract as the persistence of cytoplasmic membrane will prevent the penetration of methyl-transferases; therefore, the protocol may have to be adapted according to the cell type. SMF is performed on purified nuclei under native conditions. An alternative approach consists of performing the methylation footprinting on cross-linked chromatin. We and others have successfully implemented such a protocol[8,20,21]. This strategy globally generates results comparable to native SMF[8]. However, the ability to add accessibility information on stable protein–DNA complexes has the added advantage of enabling the coupling of SMF with other approaches such as ChIP or Hi-C. In turn, such technology could resolve the genomic and epigenomic context in which TF binding occurs at the molecular level[16].

Recently, several studies have demonstrated the possibility to couple methylation footprinting with long read sequencing[12–15]. The advancements made by these studies enable haplotype-resolved maps of accessibility over several kilobases. This continuous accessibility information can reveal co-regulatory patterns and dependencies between distant CREs. Future improvements of these sequencing methods in terms of throughput and accuracy of methylation calls could enable measuring the degree of TF co-occupancy at distant regulatory regions.

Experimental design

Footprinting efficiency

Preparation of the biological material for footprinting is key to successful SMF experiments. The number of cells to use has to be adjusted based on the genome size. The presented conditions allow efficient footprinting of ~1 μg of DNA, which corresponds to 0.25×10^6 mammalian cells or 2.5×10^6 Drosophila cells. This material is sufficient to prepare targeted (96 bisulfite PCR reactions) or whole-genome bisulfite libraries. Homogeneity in nuclear extraction and permeabilization is important as the cytoplasmic membrane would prevent the penetration of the methyltransferases. This would lead to artifactual heterogeneity in the footprinting patterns (fully inaccessible molecules). It is therefore important to use a nuclear extraction protocol adapted to the cell type or tissue used. The current protocol is robust and has been used successfully for various fly cell lines (Schneider S2, Ovarian Somatic Cells) and mammalian cell types (mouse embryonic stem cell (mESC), Neuronal Progenitors, mouse erythroleukemia cells, C2C12, HeLa). It is, however, advisable to routinely check the homogeneity of the nuclear preparations using trypan blue before performing SMF.

Unbiased quantification of protein–DNA contacts in the genome requires uniform ectopic methylation of CpGs or GpCs in all possible sequence contexts. To evaluate the sequence preferences of M.SssI and M.CviPI, we have performed in vitro methylation at various nonsaturating enzymes concentrations and evaluated the methylation levels of cytosines in all possible 4mer contexts (Fig. 3). We observed a modest preference of M.CviPI for certain sequence contexts under low enzyme-to-substrate ratio and nearly no preference for M.SssI (Fig. 3). Importantly, under saturating conditions

(>10 units/μg) these preferences become negligible, in agreement with the fact that these differences between sequence contexts cannot be observed in SMF data. We thus recommend keeping saturating levels of methyl-transferases (>200 units/μg) when performing SMF in order to ensure that every GpC and CpG can be analyzed unbiasedly.

Enzyme selection

SMF can be performed using the GpC methyltransferase M.CviPI alone or in combination with the CpG methyltransferase M.SssI. The tandem treatment increases the spatial resolution of the assay from one observation every 10 bp to one every 7 bp (median)[10]. However, tandem methylation footprinting can only be performed in cell types or tissues that do not have endogenous methylation signals in the CpG context. We have used this strategy successfully in fly cell lines[10] as well as in mESCs where endogenous methyltransferases are genetically depleted[8,10]. This is, however, not applicable to somatic cell types that do not survive depletion of endogenous methylation. Using only GpC methylation reduces the number of analyzable binding sites by a factor of about two[8]. This nevertheless leaves several thousands of binding events representing each TF and is still useful to derive general rules about their function (see 'Anticipated results', Fig. 2).

Sequencing strategy and coverage requirements

Sequencing a sufficient number of DNA molecules to cover the loci of interest is essential for accurate SMF analysis. The typical coverage reached in genome-wide experiments is ~40 molecules per locus. The coverage requirement depends on the binding frequency of the studied protein. For TFs, we typically observe binding frequencies between 1% and 40%. While frequencies >20% would be accurately quantified with coverage of 40× (8/40 molecules), lower binding frequencies would require higher sequencing depth. This consideration is even more critical when aiming to jointly analyze multiple binding events to allow accurate quantification of all the combinations of binding states. Calculations of the theoretical coverage should be conducted to decide on the sequencing strategy applied to footprinted DNA. For instance, performing a whole-genome SMF sequencing experiment on a NextSeq 550 lane with 150 paired-end reads leads to ~350 × 10^6 clusters of 300 bp (cost ~4,500 EUR). Accounting for the lower mapping rates of bisulfite libraries (~60%), this achieves a theoretical coverage of 252× for the fly genome and of 20× for the mouse genome. For the mouse genome, this is insufficient for single-molecule analysis and targeted sequencing approaches should be considered.

Primer design

We recommend the use of Primer 3[22] with an in silico bisulfite-converted genome to identify suitable primers for targeted SMF experiments. Primers should be designed such that the region to amplify is centered around the feature of interest (i.e., TF binding sites), and should not exceed 500 bp in width. Primers that will be used for the same experiments should have a uniform melting temperature (Tm difference of <4 °C, i.e., 55 °C < Tm < 58 °C) to enable their parallel amplification in 96- or 384-well plates. Primers should not overlap cytosines in the CpG or GpC contexts to avoid amplification biases towards certain methylation states. This makes the design of regions enriched with these dinucleotides more challenging (e.g., CpG islands). Since bisulfite conversion differentially alters the sequence of the plus and minus DNA strands, we recommend designing primers for both strands to increase the chance to identify efficient primer pairs. Default Primer 3 design parameters typically allow designing primers for 70–80% of the regions of interest, which leads to an amplicon for >85% of the targets. Releasing Primer 3 stringency will increase the success rate of the primer design, but also reduce the success rate of amplification.

Quality controls

Several controls can be implemented to ensure the quality of SMF libraries. Measuring accessibility over long DNA molecules (>200 bp) is critical for the interpretation of SMF data at the single-molecule level. Bisulfite conversion leads to substantial DNA fragmentation. It is therefore important to experimentally determine the DNA fragment size distribution using Agilent Bioanalyzer (Fig. 4a). Before investing in deep sequencing of the sample, it is also advised to generate low coverage data (<1 × 10^6 reads) to verify the basic features of the libraries. These include the efficiency of bisulfite conversion, mapping rates, complexity of the library (duplication rates), and fragment size distribution (Fig. 5b). Bisulfite conversion is estimated by calculating the average conversion of cytosines that are neither in the CpG nor in the GpC context. As these cytosines are not methylated in vivo, thymine frequency is expected to exceed 95% in this context. Additionally, capture efficiency can be

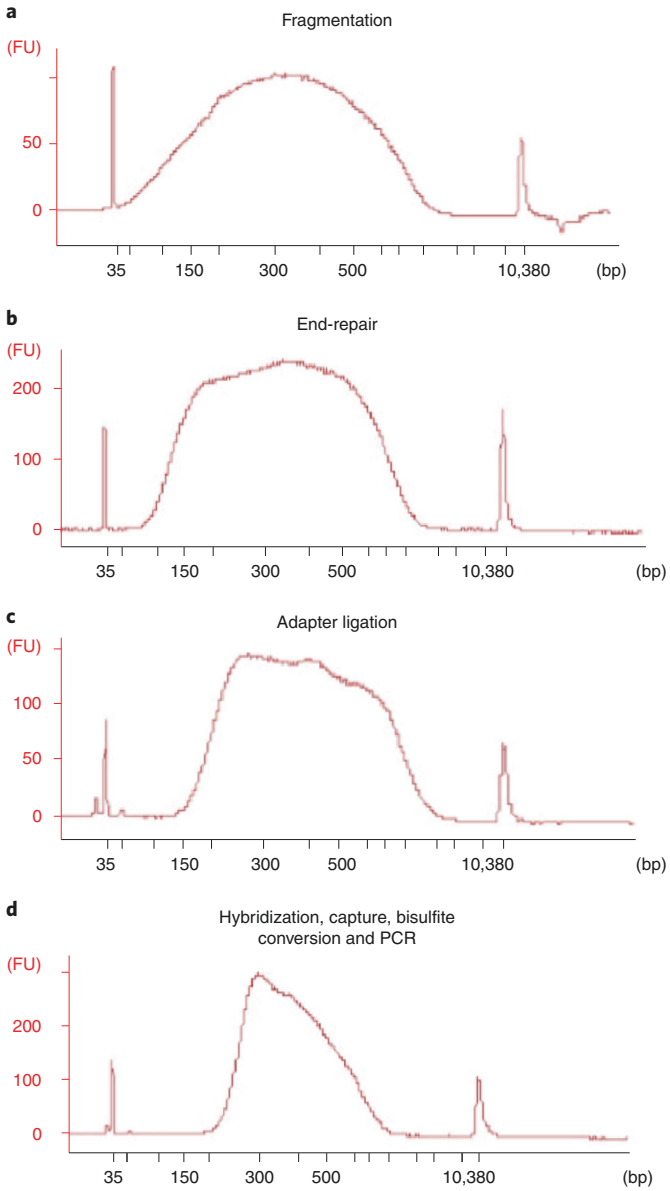

Fig. 4 | QCs during the preparation of bait-captured SMF samples. Bioanalyzer traces after various steps of the protocol. **a**,**b**, Footprinted DNA is fragmented with Covaris (300–500 bp) (Step 34) (**a**) and subjected to end-repair and A-tailing (Step 49) (**b**). **c**, A 50–100 bp shift in size distribution is detected at the adapter ligation step (Step 60). The library is then subjected to bait capture and bisulfite conversion. **d**, The size distribution is further shifted upon library amplification to a final library size of 300–600 bp representing DNA fragments of ~150–500 bp (Step 115).

estimated by calculating the fraction of mapped reads falling within the bait regions in the case of targeted enrichment in mammalian genomes. Finally, footprinting efficiency can be evaluated by comparing the observed methylation with reference high-coverage datasets (Fig. 6) (see 'Quality controls' under 'Bioinformatics analysis').

Bioinformatics analysis

SMF data can be interpreted in bulk and at the single-molecule level. The bulk level analysis is performed by calculating average methylation using all sequencing reads covering a locus. The generated profiles typically show large footprints at nucleosome occupied sites (~150 bp) and shorter footprints at TF-bound regions (<25 bp) (Fig. 7). The binding frequency of TFs or nucleosomes in the cell population can be further quantified by analyzing the data at single-molecule resolution.

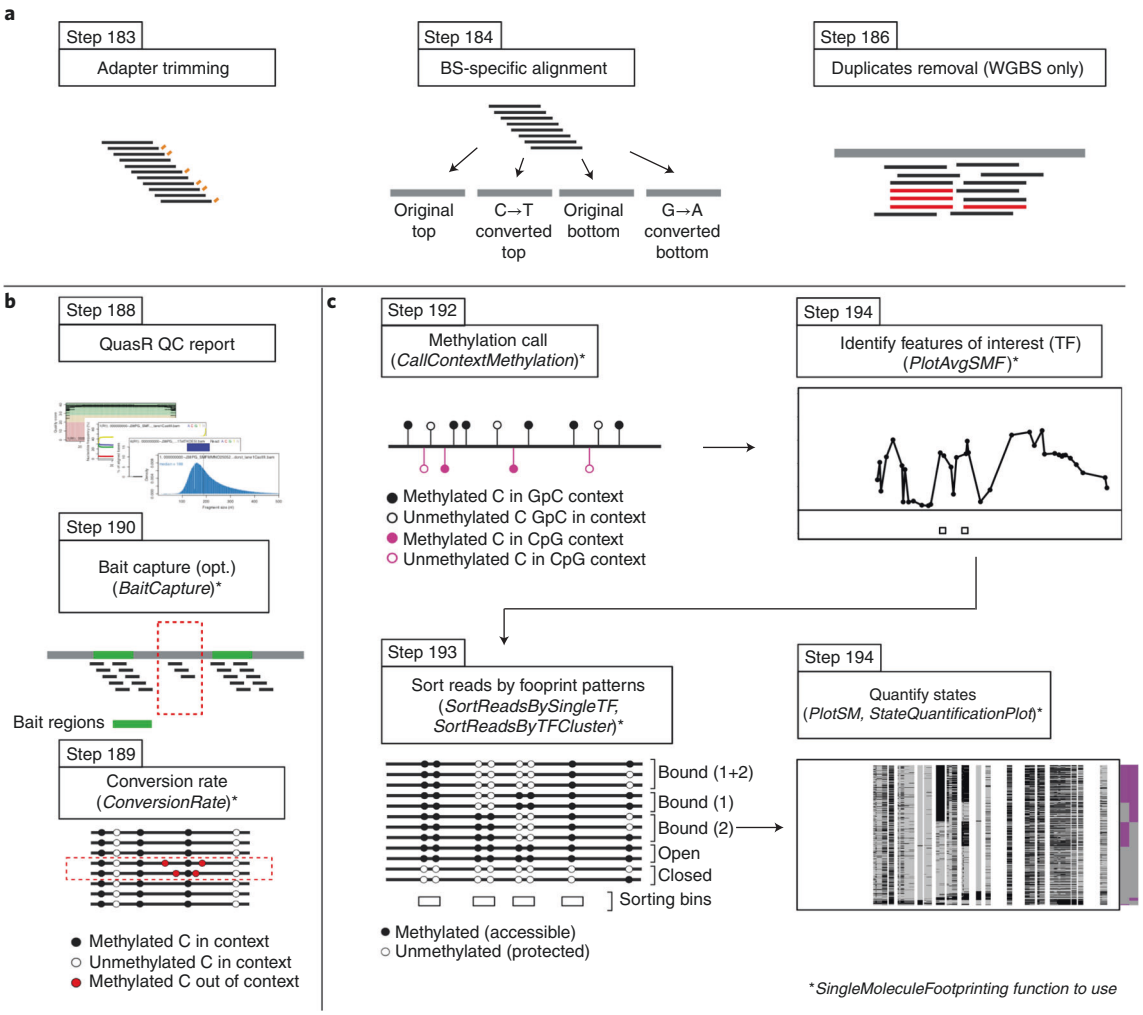

Fig. 5 | Overview of the computational workflow. a, The sequencing reads are preprocessed. Illumina adapters are removed, and low-quality bases are trimmed. The reads undergo bisulfite (BS)-specific alignment against a bisulfite-converted genome. PCR duplicates are removed only for WGBS. **b**, The quality of the library is assessed by performing several generic QCs including estimating the mapping rate, duplication rates and fragment length distribution. In addition, SMF-specific control steps, such as estimating bait capture efficiency and the conversion rate, are performed. **c**, A series of functions have been implemented in the *SingleMoleculeFootpring* R package to facilitate data interpretation. These include functions to call average methylation in the relevant genomic contexts (GpC and CpG); sort the reads according to their footprint patterns; and plot average and single-molecule footprints at individual loci.

This allows one to determine the proportion of sequenced molecules that show a footprint for a given factor at individual binding sites (Fig. 7). These proportions reflect the binding frequencies of either TFs or nucleosomes at a given locus and quantitatively describe the heterogeneity of CRE usage in a cell population. Additionally, the degree of co-occurrence between neighboring binding events can be quantified.

Data preprocessing

Base calling and barcode demultiplexing of raw Illumina data is performed using manufacturer instructions and software. The resulting fastq files are used as an input for trimming adapters and low-quality bases. Reads are aligned using the Bioconductor package QuasR[23], which performs bisulfite alignments using Bowtie 1[24]. If a tool other than QuasR is required, users should restrict their choice to an aligner based on Bowtie 1.x.x rather than later versions (Bowtie 2.x.x). This is critical to ensure compatibility with our *SingleMoleculeFootprinting*[25] R package, which has QuasR and its functions at its core. The expected mapping rate for a typical mouse SMF experiment is ~60%. Technical replicates are pooled and PCR duplicates are removed using Picard Tools v2.15.0 (http://broadinstitute.github.io/picard/). Duplicates removal should not be performed in the case of amplicon sequencing experiments.

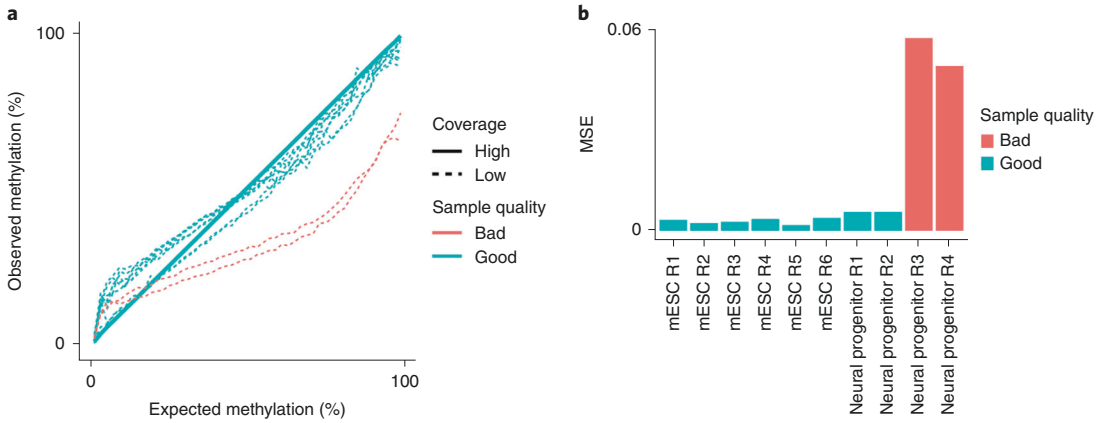

Fig. 6 | Controlling footprinting efficiency with low-coverage sequencing data. The efficiency of footprinting can be controlled using low-coverage samples (<1 × 10^6 reads) and comparing them with existing reference datasets. The comparison is made under the assumption that most of the SMF signal is invariable between conditions since it mostly represents nucleosome occupancy across the genome. **a**, Comparison of expected versus observed methylation rate values for several low-coverage samples, two of which were identified to be undermethylated (red lines). The high-coverage reference sample is used to group cytosines based on their reference methylation. The methylation of each group of cytosines is calculated using all reads covering cytosines of a given group that have similar accessibility profiles. **b**, The deviation of each sample from the reference dataset where the observed values perfectly equal the expected values is quantified as the mean squared error (MSE), successfully identifying undermethylated samples. This procedure allows control for the efficiency of footprinting before investing in deep sequencing of SMF samples.

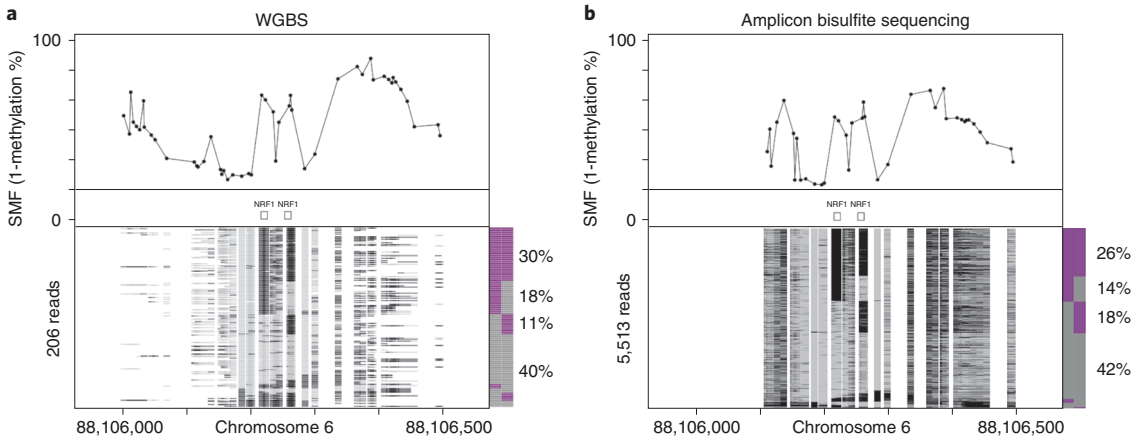

Fig. 7 | SMF data visualization. a,**b**, Single-molecule analysis of a *Mus musculus* genomic locus harboring two NRF1 binding sites using WGBS (**a**) or amplicon bisulfite sequencing data (**b**). The upper panels show the average SMF signal (1-methylation). The lower panels show stacks of single DNA molecules sorted according to the occupancy pattern of the two NRF1 binding sites. The frequency of the states is displayed in the barplot next to the single-molecule stacks. In this particular case, both NRF1 binding sites are co-occupied in 30% and 26% of the reads in the WGBS and amplicon sequencing experiment, respectively. Binding at individual NRF1 sites is observed at between 11% and 18% of the reads, and the region is accessible in ~40% of the molecules. Signal amplification in the amplicon experiment increases coverage to 5,513 reads versus the 206 of the genome-wide experiment.

Quality controls

Before performing a whole-genome bisulfite sequencing (WGBS) run at high depth, we advise assessing the quality of the sequencing libraries by producing shallow sequencing data and running the following quality controls (QCs). The *qQCreport* function from the Bioconductor package *QuasR*[23] can be used to produce a QC report providing an assessment of the quality of the SMF libraries. A typical library has a mapping rate of >50%, read duplication rate of <20% (does not apply to amplicon sequencing experiments) and median fragment size distribution of >200 bp. In addition, the function *ConversionRateEstimate* from our *SingleMoleculeFootprinting*[25] package estimates the bisulfite conversion rate by measuring the conversion of cytosines outside of methylated contexts. Conversion rates should exceed 95%, while the majority of methylation rate values for footprinted contexts are expected to fall between 15% and 60%. The efficiency of the optional capture step for large genomes can be calculated using the *BaitCapture* function from *SingleMoleculeFootprinting*. For a typical captured SMF library, >70% of the reads are expected to fall within the region targeted by baits. Finally, the efficiency of footprinting can be assessed by calculating the methylation distribution in the sample using the function *LowCoverageMethRateDistribution* from the

SingleMoleculeFootprinting package. The function uses an existing high-coverage reference dataset to group cytosines based on their methylation scores. This grouping enables aggregation of reads for multiple cytosines that have similar expected footprinting levels to reach the coverage needed to confidently measure them at low sequencing depth ($<1 \times 10^6$ reads). The generated methylation profiles can discriminate samples based on the efficiency of the footprinting step before investing in deep sequencing experiments (Fig. 6). A curve lying above the reference line indicates over-methylation of the sample, possibly indicating the presence of naked DNA in the nuclear preparation (i.e., broken nuclei). Curves lying under the reference line indicate under-methylation of the samples (Fig. 6). This is possibly due to incomplete footprinting that can arise if nuclear extraction is not complete or if enzyme activity is too low. For more details, see the Troubleshooting section.

Quantification of bulk protein occupancy levels

Average methylation is computed at all genomic cytosines and reduced to the relevant contexts using the *SingleMoleculeFootprinting* function *CallContextMethylation*, which employs at its core the *QuasR* function *qMeth*. GpC and CpG contexts can be analyzed together when performing dual-enzyme footprinting (e.g., in *Drosophila*). However, they have to be interpreted separately when performing GpC-only footprinting in mammalian cell lines that have endogenous CpG methylation. In this case, accessibility footprints should only be analyzed at DGCH contexts, where, in International Union of Pure and Applied Chemistry (IUPAC) code, D stands for any nucleotide except C and H for any nucleotide except G. This is aimed at excluding ambiguous contexts such as GCG that are also targeted by endogenous methyltransferases.

Assigning the identity of SMF footprints

As with any other footprinting method, SMF is agnostic to the identity of the protein creating the footprints. Thus, SMF data interpretation requires the association of the observed footprint with protein–DNA binding data to identify the protein that creates it. Scanning the genome with known positional weight matrices[26] can be used to annotate the footprints and identify putative TF binding events. This process is, however, very noisy, typically leading to multiple overlapping motifs, most of which are not bound. Therefore, we advise to subset the list of putative transcription factor binding sites (TFBSs) for evidence of in vivo binding as measured by orthogonal methods such as ChIP-seq. For instance, UniBind[27] offers a reference map of putative TFBSs predicted from ChIP-seq data.

Quantification of protein occupancy at the single-molecule level

We developed several strategies to sort molecules according to their occupancy states and to calculate the frequency of those states at individual loci (Fig. 8). In the case of TFs, we distinguish molecules that are bound by one or multiple TFs from molecules that are fully accessible and molecules that are occupied by nucleosomes. Given a set of n TFBS coordinates as input, we draw $n + 2$ bins: n that are 15–30 bp in width for the TFBSs, plus one upstream and one downstream both 10 bp in width (Fig. 8). Methylation values for each read are averaged and rounded within each bin such that each read becomes described by a string of $n + 2$ binary digits. There are at this point 2^{n+2} possible methylation patterns that can be biologically interpreted in terms of molecular occupancy. The functions *SortReadsBySingleTF* and *SortReadsByTFCluster* from our *SingleMoleculeFootprinting* package can be used to sort reads based on the footprint left by one or multiple TFs, respectively.

Single-locus visualization

The bulk footprinting signal can be displayed for a single locus using the function *PlotAvgSMF*. Accessibility information for single molecules can be visualized using the *PlotSM* function, while the proportions of reads found in each state can be obtained using the *StateQuantificationPlot* function. Finally, the function *PlotSingleSiteSMF* offers a convenience wrapper to generate the three plots at once (Fig. 7). The vignette for the SingleMoleculeFootprinting R package can be found at https://www.bioconductor.org/packages/release/bioc/vignettes/SingleMoleculeFootprinting/inst/doc/SingleMoleculeFootprinting.html.

Expertise needed to implement the protocol

The protocol described here requires knowledge of standard molecular biology techniques. Production of SMF data requires access to a dedicated sequencing facility. To analyze SMF data, the user should have access to a high-performance computing system with a Linux distribution installed in order to perform some of the data preprocessing steps. The user should be comfortable with R scripting and minimal command line usage.

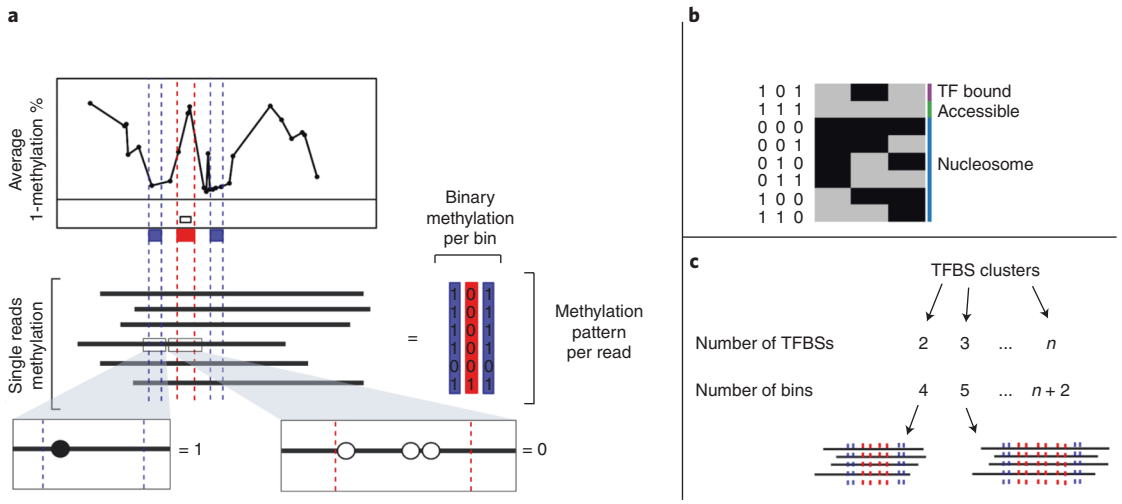

Fig. 8 | Single-molecule sorting. a, Single reads can be sorted according to the occupancy pattern over a genomic feature of interest. Here, a TFBS is depicted as a box in the lower part of the average SMF plot. Three collection bins are drawn: one centered on the TFBS (red box), one upstream and one downstream of it (green boxes). For each read, the methylation information is averaged and rounded within the bins (as shown in the callout windows). The result is that each read is now reduced to three binary values. **b**, There are 2^3 possible methylation patterns. One of those is '101', which represents the cases where the TFBS bin is found occupied (unmethylated) and the two surrounding bins are found accessible (methylated). When the methylation pattern of a read corresponds to '101', it is interpreted as in the 'TF bound' state. Alternatively, the sequence '111' would correspond to the 'accessible' state. The remaining combinations are interpreted as 'nucleosome occupied' states. **c**, Single reads can also be sorted according to the occupancy pattern over multiple genomic features, such as TFBS clusters. In this case, the number of bins that are drawn is $n + 2$, where n equals the number of TFBS in the cluster. Notably, the number of possible states, and therefore the complexity of the biological interpretation, increases with the number of TFBSs. Figure adapted from ref. [8].

Materials

Reagents

Biological materials

- Cell suspension. mESCs; 159[28] (https://scicrunch.org/resolver/CVCL_IT51) and 159 DNMT TKO[29], a knockout cell line of the three DNA methyltransferases DNMT1, 3a, and 3b in the 159 cell line **! CAUTION** The cell lines used should be regularly checked to ensure they are authentic and are not infected with mycoplasma.

Common reagents

- Nuclease-free water (not diethylpyrocarbonate-treated) (Ambion, cat. no. AM9937)
- Qubit dsDNA HS Assay Kit (Life Technologies, cat. no. Q32851)

Cell culture

- Dulbecco's Modified Eagle Medium, high glucose (Gibco, cat. no. 41965039)
- FBS Embryomax (Millipore, cat. no. ES-009-B)
- Gelatin, from porcine skin (Sigma, cat. no. G-1890)
- L-glutamine (Gibco, cat. no. A2916801)
- Leukemia inhibitory factor (prepared in house, 10 mg/ml in PBS)[30]
- MEM Non-Essential Amino Acids Solution (100×) (Gibco, cat. no. 11140050)
- 2-Mercaptoethanol (Merck, cat. no. M6250) **! CAUTION** Toxic and irritant; avoid inhalation, and wear personal protective equipment (PPE).
- Sodium pyruvate (Gibco, cat. no. 11360070)
- PBS (prepared in house)
- Trypan blue solution, 0.4% (wt/vol) (Gibco, cat. no. 15250061)
- Trypsin-EDTA (0.25%) (Gibco, cat. no. 25200056)

SMF treatment

- CpG methyltransferase (M.SssI) (NEB, cat. no. M0226L)
- GpC methyltransferase (M.CviPI) (NEB, cat. no. M0227L)

- IGEPAL CA-630 (Sigma, cat. no. I8896) **! CAUTION** Eye irritant.
- Magnesium chloride ($MgCl_2$) (Sigma, cat. no. M8266)
- *S*-adenosyl-methionine (SAM), 32 mM (NEB, cat. no. B9003S)
- Sodium chloride (NaCl) (Sigma-Aldrich, cat. no. S7653)
- Sodium dodecyl sulfate solution (10%) (Sigma-Aldrich, cat. no. 71736)
- Sucrose Ultrapure MB grade (Affymetrix, cat. no. 21938)
- Titriplex III (EDTA disodium salt dihydrate) (Sigma, cat. no. 1.08421)
- Trizma base (Sigma-Aldrich, cat. no. T1503)

DNA extraction

- Chloroform (Sigma, cat. no. 366919) **! CAUTION** Harmful and irritant; avoid inhalation, and wear PPE.
- Glycogen from *Mytilus edulis* (blue mussel) (Sigma-Aldrich, cat. no. G1767)
- Phenol equilibrated, stabilized:chloroform:isoamyl alcohol 25: 24: 1 (PanReac AppliChem, cat. no. A0889) **! CAUTION** Phenol is corrosive and toxic; chloroform is harmful and an irritant. Avoid inhalation, and wear PPE.
- 2-propanol (Sigma, cat. no. I9516) **! CAUTION** Flammable.
- Proteinase K (Sigma-Aldrich, cat. no. 124568)
- RNase A, DNase- and protease-free (Sigma, cat. no. R6513)

Capture library

- EZ DNA Methylation-Gold Kit (Zymo Research, cat. no. D5005)
- Sodium hydroxide solution, 10 M (Sigma-Aldrich, cat. no. 72068) **! CAUTION** Corrosive; wear PPE.
- SureSelectXT Methyl-Seq Reagent Kit (Agilent, cat. no. G9651A)
- SureSelectXT Mouse Methyl-Seq Capture Library (Agilent, cat. no. 931052)

Amplicon library

- Agarose (Sigma, cat. No. A9539)
- Epitect bisulfite conversion kit (Qiagen cat. no. 59104)
- Ethidium bromide solution 1% (Roth, cat. no. 2218.1)
- GeneRuler 1 kb DNA ladder, ready-to-use (Thermo, cat. no. SM0313)
- GeneRuler 100 bp DNA ladder, ready-to-use (Thermo, cat. no. SM0244)
- KAPA HiFi HotStart Uracil+ ReadyMix (2×) (Roche, cat. no. KK2802 07959079001)
- NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, cat. no. E7645L)
- NEBNext Multiplex Oligos for Illumina (Index Primers 1–12) (NEB, cat. no. E7335L)
- Primers, bisulfite specific, resulting in amplicons ranging from 300 to 500 bp (Sigma)

Equipment

Common

- Bioanalyzer 2100 instrument (Agilent, cat. no. G2939BA)
- Bioanalyzer DNA 1000 Kit (Agilent, cat. no. 5067-1504)
- Bioanalyzer High Sensitivity DNA Kit (Agilent, cat. no. 5067-4627)
- Centrifuge, refrigerated, with fixed-angle rotor (Eppendorf, model no. 5427R)
- Centrifuge with fixed-angle rotor (Eppendorf, model no. 5425)
- Centrifuge with swinging bucket (Eppendorf, model no. 5810R)
- Heater block with wells for 1.5 ml tubes (e.g., Thermo), set to 37 and 56 °C
- Magnetic rack for PCR tubes
- Magnetic rack for 1.5 ml tubes; Dynamag (Thermo, cat. no. 12321D)
- Microcentrifuge (e.g., Roth)
- 1.5 ml microcentrifuge tubes (Eppendorf, cat. no. 22-282)
- 1.5 ml microcentrifuge safe-lock tubes (Eppendorf, cat. no. 30120086)
- 1.5 ml microcentrifuge DNA LoBind tubes (Eppendorf, cat. no. 30108051)
- 0.2 ml PCR tubes (Eppendorf, cat. no. 30124359)
- Thermal cycler (Biorad, C1000 touch, cat. no. 1851148/1851196)
- Vortex mixer (e.g., Vortex Genie; VWR)
- Water baths (e.g., VWR), set to 37 °C

Capture library
- MicroTUBE holder (Covaris, cat. no. 500114)
- S-series focused ultrasonicator (Covaris, S2 model)
- Snap-Cap microTUBEs (Covaris, cat. no. 520045)
- Vacuum concentrator (e.g., Eppendorf)

Amplicon library
- Agarose gel chamber and power supply (Bio-rad, cat. no. 1640301)
- PCR adhesive film (Eppendorf, cat. no. 0030127781)
- Twin.tec PCR plates (Eppendorf, cat. no. 0030133366)

Hardware and software
- High-performance computting system running on a Linux distribution (e.g., CentOS)
- Trimmomatic (v0.36 or higher) installation (http://www.usadellab.org/cms/?page=trimmomatic)
- Picard (v2.15.0 or higher) installation (https://broadinstitute.github.io/picard/)
- R-4.1.0, or higher (https://www.r-project.org)
- Rstudio (optional) (https://www.rstudio.com)
- Bioconductor v3.13 (https://bioconductor.org)
- Primer 3 (https://primer3.ut.ee/)
- Bioconductor packages
  - QuasR v1.32.0 (or higher) (https://bioconductor.org/packages/QuasR/)
  - SingleMoleculeFootprinting v1.0.0 (https://bioconductor.org/packages/SingleMoleculeFootprinting/)
  - SingleMoleculeFootprintingData v1.0.0 (https://bioconductor.org/packages/SingleMoleculeFootprintingData/)

Reagent setup

mESC culture medium
Supplement Dulbecco's Modified Eagle Medium with 15% (vol/vol) FBS Embryomax, 2mM L-glutamine, 1% MEM Non-Essential Amino Acids Solution (vol/vol), 1 mM sodium pyruvate, 0.001% (vol/vol) 2-mercaptoethanol and 20 ng/ml leukemia inhibitory factor. Store at 4 °C for up to 1 month.

0.2% Gelatin
Prepare stock solution in water. Sterilize by autoclaving. Can be kept at room temperature (RT, 19–22 °C) for 1 year.

20% (vol/vol) IGEPAL CA-630
Prepare stock solution in water. Store at RT for at least 1 year.

1 M $MgCl_2$
Prepare stock solution in water, and store at RT for 1 year.

5 M NaCl
Prepare stock solution in water, and store at RT for at least 1 year.

1 M Sucrose
Prepare stock solution in water, and store at 4 °C for up to 1 month.

0.5 M EDTA
Prepare stock solution in water, and store at RT for at least 1 year.

1 M Tris-HCl (pH 7.4–7.6 and pH 7.9)
Prepare stock solution 1 M in water. Adjust pH accordingly with HCl to pH 7.4–7.6 and pH 7.9, and store at RT for at least 1 year.

0.1× TE
Prepare a stock solution of 1× Tris–EDTA (TE) with 10 mM Tris and 0.1 mM EDTA. Dilute to 0.1× for a working solution. Store both solutions at RT for at least 1 year.

Glycogen
Stock solution at 20 mg/ml, and store aliquots at −20 °C for 1 year.

Proteinase K
Prepare stock solution at 20 mg/ml in water, aliquot and store at −20 °C for 1 year.

RNase A
Prepare stock solution at 10 mg/ml in water, aliquot and store at −20 °C for 1 year.

10.1 M NaOH
Prepare fresh 0.1 M NaOH by diluting 10 M NaOH stock solution. Per sample prepare 20 µl plus excess.

Lysis buffer
10 mM Tris (pH 7.4-7.6), 10 mM NaCl, 3 mM MgCl2, 0.1 mM EDTA and 0.5% (vol/vol) IGEPAL CA-630. Buffer is stable for up to 1 month at 4 °C.

Wash buffer
10 mM Tris (pH 7.4–7.6), 10 mM NaCl, 3 mM $MgCl_2$ and 0.1 mM EDTA. Buffer is stable for up to 1 month at 4 °C.

GpC methyltransferase mix
1× M.GpC buffer, 300 mM sucrose and 64 µM SAM. Add SAM shortly before using the mix. To be made fresh every time.

Stop solution
20 mM Tris-HCl (pH 7.9), 600 mM NaCl, 1% (wt/vol) sodium dodecyl sulfate solution and 10 mM EDTA. Needs to be heated up to 37 °C prior to usage to eliminate precipitates. Buffer is stable for up to 1 month at RT.

Primers for amplicon bisulfite sequencing library
The amplicons size should range from 300 to 500 bp in size with the majority of amplicons being >450 bp. Primers are ordered in 96-mixed well format resuspended in 100 µM of water (by the manufacturer). A working dilution is obtained by diluting the forward and reverse primers to a 2 µM mix with RNase-free water in a 96-well format, in a total volume of 200 µl. Working plates can be stored at 4 °C, while stock plates are kept at −20 °C.

Procedure

Nuclei extraction ● Timing 50 min

▲ **CRITICAL** For downstream application of genome wide footprinting, generally 3 µg of DNA is needed; therefore, it is recommended to perform three reactions in parallel and pool them at the DNA capture step (Step 61). 0.25×10^6 mESCs cells are needed per reaction.

▲ **CRITICAL** We do not recommend processing more than eight samples at a time as enzymes and cofactors have to be replenished individually during the treatment.

1 Trypsinize actively growing cells, spin down the cells for 5 min at 314*g* and wash the pelleted cells once with cold (4 °C) PBS.

▲ **CRITICAL STEP** The following steps are all done at 4 °C. The lysis and wash buffers are also kept at 4 °C. Prewarm the stop solution at 37 °C.

▲ **CRITICAL STEP** It is essential to ensure that a single-cell suspension has been obtained in this step, since a mistake can only be read out upon data analysis. Therefore, confirm single-cell distribution under a microscope or with a cell counter.

2 Resuspend 0.25×10^6 cells in 1 ml of ice-cold lysis buffer and incubate on ice for 5–10 min, inverting the tubes occasionally.

3 Centrifuge for 5 min at 1,000*g* at 4 °C. Discard the supernatant.

4 Resuspend nuclei in 250 µl ice-cold wash buffer.

5 Centrifuge for 5 min at 1,000*g* at 4 °C and discard the supernatant.

6 Resuspend the nuclei in 94.5 µl 1× M.GpC buffer and keep on ice until methyltransferase treatment.

GpC methyltransferase treatment ● Timing 20 min

7 To each sample containing 94.5 µl of nuclei, add 150 µl of GpC methyltransferase mix.
▲ CRITICAL STEP The enzymatic treatment starts here. It is important to minimize pipetting time in order to keep incubation time consistent between samples.
8 Add 50 µl M.CviPI and mix by pipetting with a P200 pipette. Do not vortex.
9 Incubate at 37 °C for 7.5 min.
▲ CRITICAL STEP Keep samples at RT for the following additions.
10 Add 25 µl M.CviPI, then 4 µl SAM and mix by pipetting with P200. Do not vortex.
11 Incubate at 37 °C for 7.5 min. Proceed directly to Step 14 if not performing CpG treatment.

CpG methyltransferase treatment ● Timing 10 min

▲ CRITICAL CpG methyltransferase treatment is optional and is only applicable to cells without endogenous methylation, such as DNMT TKO.
▲ CRITICAL Keep samples at RT for the following additions.
12 Add in this order: 3.5 µl $MgCl_2$ (1 M), 15 µl M.SssI and 4 µl SAM. Mix by pipetting with a P200 pipette. Do not vortex.
▲ CRITICAL STEP SAM is an unstable substrate and degrades at elevated temperatures. Therefore any leftover SAM should be discarded and not be saved for later use.
13 Incubate at 37 °C for 7.5 min.

Finalize treatment ● Timing 5 min

14 Add 300 µl of prewarmed stop solution and 6 µl proteinase K, and mix briefly by vortex.
15 Incubate overnight at 55 °C.
■ PAUSE POINT Samples can be stored at −20 °C for several weeks.

DNA extraction ● Timing 2 h

! CAUTION Phenol and chloroform are hazardous chemicals. Perform Steps 16–21 in the chemical hood, avoid inhalation and wear PPE.
16 Extract DNA by adding 600 µl phenol:chloroform to each sample.
17 Shake hard 15 times and centrifuge for 5 min at RT at maximum speed.
18 Transfer the aqueous phase to a new 1.5 ml safe-lock tube.
19 Add 600 µl chloroform, shake hard 15 times and centrifuge for 5 min at RT at maximum speed.
20 Transfer the aqueous phase to a new 1.5 ml safe-lock tube.
21 Precipitate the DNA by adding 600 µl isopropanol and 1 µl glycogen.
22 Incubate at RT shaking continuously at 300 rpm in a thermomixer for 10 min, alternatively mixing the tube by occasional inversion every 2 min.
23 Centrifuge for 20 min at maximum speed.
24 Remove the supernatant and discard.
25 Wash the pellet with 1 ml 70% (vol/vol) ethanol (EtOH) kept on ice.
26 Centrifuge for 15 min at 4 °C at maximum speed.
27 Remove the supernatant thoroughly without disturbing the pellet and discard.
28 Air-dry the pellet by laying the tubes on the bench. This can take 5–30 min.
▲ CRITICAL STEP Try not to over-dry the pellet as resuspension may become difficult, since the pellet is quite small and therefore may be lost.
29 Resuspend pellet in 20 µl H_2O for targeted amplicon enrichment (application B) and in 30 µl H_2O for whole-genome enrichment (application A).
30 Add 1 µl RNAse A and incubate at 37 °C for 30 min.
31 Let the pellet dissolve fully for 2 h at 37 °C or overnight at 4 °C.
■ PAUSE POINT Samples can be stored at −20 °C for several weeks.
32 Quantify DNA concentration by Qubit 1X DNA HS measurement and DNA quality by Nanodrop.
▲ CRITICAL STEP Expect ~1 µg DNA per reaction.
▲ CRITICAL STEP To continue with whole-genome bisulfite library preparation, proceed with application A. For amplicon bisulfite library, proceed with application B.

Application A: WGBS library with targeted enrichment

▲ **CRITICAL** Libraries are prepared based on the SureSelect XT Mouse Methyl-Seq Kit Enrichment System for Illumina Multiplexed Sequencing Library protocol (Agilent Technologies, version E0, April 2018).

▲ **CRITICAL** WGBS libraries can also be prepared without targeted enrichment. In this case, the hybridization and capture of the library steps (from Steps 61–85) are omitted. Following ligation (Step 60), proceed directly to Step 86.

▲ **CRITICAL** Samples are fragmented with a Covaris model S2. Consult the S-series setup and instruction manual for start-up procedures (Covaris, Chapter 4.1, Rev F October 2020).

Library preparation ● **Timing 3 h**

▲ **CRITICAL** Prepare the Covaris device ahead of time as it will take time to cool down.

33 Prepare a dilution of one footprinted reaction (from Step 31), ranging from 1 to 2.5 µg DNA, in 60 µl H_2O. Shortly before fragmentation, transfer the diluted DNA to a Covaris microtube.
▲ **CRITICAL STEP** Do not keep the DNA in the Covaris microtube for an extended time.

34 Fragment DNA using a Covaris device. Aim to obtain 300 bp fragments via sonication. For Covaris S2, this will be duty factor 10%, intensity 4 and 200 cycles per burst for 100 s. Check fragmentation quality with bioanalyzer by running 1 µl sample diluted 1:5. A successful example is shown in Fig. 4a.
? TROUBLESHOOTING

35 Prepare the end repair master mix as follows and keep on ice.

Reagent	Volume (for one reaction)
10× End Repair Buffer (clear cap)	10 µl
dNTP mix (green cap)	1.6 µl
Klenow DNA polymerase (yellow cap)	2 µl
T4 polynucleotide kinase (orange cap)	2.2 µl
T4 DNA polymerase (purple cap)	1 µl
Total	16.8 µl

36 Add water to the fragmented DNA to a final volume of 83.2 µl.

37 Add the end repair master mix to the DNA. Mix briefly by vortex and spin for a few seconds on a table top centrifuge. Incubate the sample in a thermal cycler with the following program.

Step	Temperature	Time
1	20 °C	30 min
2	4 °C	Hold

38 Add 180 µl of AMPure XP beads to 100 µl of end repaired sample (1.8× ratio) and mix by pipetting approximately ten times.
▲ **CRITICAL STEP** Prior to usage, put AMPure XP beads at RT. To ensure a correct ratio is maintained, mix the beads well by vortexing shortly before adding them to the sample.

39 Incubate sample for 5 min at RT.

40 Place the sample on a magnetic stand until the solution is clear. This will take ~5 min.

41 While keeping the tubes on the magnetic stand, remove the supernatant and discard.

42 Wash the beads with 200 µl freshly prepared 80% (vol/vol) EtOH, while still keeping the sample on the magnetic stand.
▲ **CRITICAL STEP** Freshly prepared EtOH should not be older than 48 h as this will reduce the concentration of the washing solution and, in effect, the DNA yield.

43 Repeat Step 42 once more.

44 After the second wash, remove the supernatant completely and lay the tubes on their side to air-dry the beads. This will take ~1–3 min.
▲ **CRITICAL STEP** Air-drying the beads properly is critical. Avoid over-drying the beads, as this would result in a substantial loss of material when rehydrating the beads.

45 Once the beads are no longer glossy, add 44 µl water and resuspend the dried pellet by tapping the tube.
46 Incubate for ~2 min.
47 Place the sample on a magnetic stand until the solution is clear. This will take ~2 min.
48 Transfer 42 µl of eluate to a new tube.
49 Prepare an aliquot (1 µl, 4× diluted) to check the quality with a bioanalyzer DNA HS chip later. ▲CRITICAL STEP This will be run later together with the ligated sample obtained at Step 60. A successful example of end-repaired DNA is shown in Fig. 4b.
50 Prepare the A-tailing master mix as follows and keep on ice.

Reagent	Volume (for one reaction)
10× Klenow Polymerase Buffer (blue cap)	5 µl
dATP (green cap)	1 µl
Exo(-) Klenow (red cap)	3 µl
Total	9 µl

51 Add 9 µl prepared A-tailing mix to 41 µl purified end-repaired DNA (from Step 48), mix briefly by vortex and spin down for a few seconds in a table top centrifuge. Incubate the sample in a thermal cycler with the following program.

Step	Temperature	Time
1	37 °C	20 min
2	4 °C	Hold

52 Clean up the A-tailed sample by adding 90 µl of AMPure XP beads to 50 µl of the sample (1.8× ratio).
53 Follow Steps 39–44 to carry out the cleanup.
54 Elute the sample by adding 35 µl water (follow Steps 46–47) and transfer 33.5 µl of eluate to a new tube.
55 Prepare the ligation master mix as follows and keep on ice.

Reagent	Volume (for one reaction)
SureSelect Methyl-Seq Methylated Adapter (green cap)	5 µl
5× T4 DNA ligase buffer (green cap)	10 µl
T4 DNA ligase (red cap)	1.5 µl
Total	16.5 µl

56 Add 16.5 µl ligation master mix to 33.5 µl purified A-tailed DNA, mix briefly by vortex and spin down. Incubate the sample in a thermal cycler with the following program.

Step	Temperature	Time
1	20 °C	15 min
2	4 °C	Hold

57 Clean up the ligated sample by adding 32.5 µl of AMPure XP beads to 50 µl of the sample (0.65× ratio).
58 Follow Steps 39–44 to carry out the cleanup.
59 Elute the sample by adding 24 µl water (follow Steps 46–47) and transfer 22 µl of eluate to a new tube.
60 Check sample quality with a bioanalyzer DNA HS chip (1 µl, 4× diluted) and run the aliquot from Step 49 in parallel. An example of a successful library is shown in Fig. 4c. In addition, quantify the ligated sample with Qubit 1X DNA HS.

▲ CRITICAL STEP If <350 ng of adapter ligated DNA is recovered, repeat the library preparation to obtain more material. Expect 500–800 ng adapter ligated material when starting with 3 µg footprinted DNA.

■ PAUSE POINT Samples can be stored at −20 °C for several weeks.

Hybridization and capture of the library ● Timing 4 h + 16 h incubation

▲ CRITICAL At this point, it is good to keep in mind that the hybridization step takes 16 h. Remember when planning this part of the experiment that there is no pause point for the remainder of the protocol.

61 Using a vacuum concentrator, reduce the adapter ligated DNA down to 3.4 µl. Prepare a test tube with 3.4 µl water as a reference. In case the volume accidentally reduces to <3.4 µl, adjust the water level up to 3.4 µl.

▲ CRITICAL STEP It is important not to mix the sample with a pipette, since this can result in loss of material.

▲ CRITICAL STEP It will take 12–15 min at 45 °C to reduce the adapter ligated DNA from a volume of 20 µl to 3.4 µl.

62 Prepare the blocking mix as outlined below and add 5.6 µl to the 3.4 µl concentrated adapter ligated library.

Reagent	Volume (for one reaction)
Indexing Block 1 (green cap)	2.5 µl
Block 2 (blue cap)	2.5 µl
Methyl-Seq Block 3 (brown cap)	0.6 µl
Total	5.6 µl

Mix the reaction mixture gently with a pipette and incubate the sample in a thermal cycler with the following program.

Step	Temperature	Time
1	20 °C	5 min
2	65 °C	2 min
3	65 °C	Hold

63 While the samples are at 65 °C in the thermal cycler, prepare the Methyl-Seq Capture Library Hybridization Mix. Firstly, prepare the RNAse blocking solution as follows and keep on ice.

Reagents for RNase blocking solution	Volume (for one reaction)
RNase Block (purple cap)	0.5 µl
Nuclease-free water	1.5 µl
Total	2 µl

Prepare the hybridization buffer by mixing the following and keep at RT:

Reagents for hybridization buffer	Volume (for one reaction)
Hyb 1 (orange cap)	6.63 µl
Hyb 2 (red cap)	0.27 µl
Hyb 3 (yellow cap)	2.65 µl
Hyb 4 (black cap)	3.45 µl
Total	13 µl

Finally, prepare the hybridization mix at RT:

Reagents for hybridization mix	Volume (for one reaction)
Hybridization buffer	13 µl
RNase blocking solution	2 µl
Mouse Methyl-Seq Capture Library	5 µl
Total	20 µl

▲ **CRITICAL STEP** For this part of the protocol, use the SureSelectXT Mouse Methyl-Seq Capture Library part of the SureSelectXT Mouse Methyl-Seq Capture system.

▲ **CRITICAL STEP** It is important to note that the hybridization mix is prepared at RT but can only be kept at RT for a short amount of time owing to stability of the other components.

64 Keep the PCR tube containing the DNA library with the blocking mix at 65 °C in the thermal cycler while adding 20 µl of the Capture Library Hybridization Mix. Gently mix the reactions by pipetting.

65 Incubate the hybridization mixture for 16 h at 65 °C with a heated lid set to 105 °C.

66 Resuspend the MyOne Streptavidin T1 Dynabeads on a vortex mixer.

▲ **CRITICAL STEP** 50 µl of the magnetic bead suspension is needed for one hybridization sample. In case of multiple samples, pool the beads in a 1.5 ml safe-lock tube to prepare the beads (Steps 67–70) for the capture.

67 Wash 50 µl streptavidin beads with 200 µl SureSelect Binding Buffer by mixing the beads by pipetting up and down ten times.

68 Place the sample on a magnetic stand until the solution is clear, then remove and discard the supernatant.

69 Repeat Steps 67 and 68 two more times.

70 Resuspend the washed beads in 200 µl of SureSelect Binding Buffer.

71 Keep the PCR tube with the hybridization reaction at 65 °C while transferring the entire volume of the hybridization mixture to the PCR tube containing the 200 µl of washed streptavidin beads. Slowly pipette up and down until the beads are fully resuspended.

72 Cap the tube and seal with parafilm. Then incubate the capture reaction by putting the PCR tube on a vortex mixing continuously at full speed for 30 min at RT. Make sure the sample is mixing properly in the tube.

73 During the 30 min incubation for capture, prewarm Wash Buffer 2 at 65 °C by placing 200 µl aliquots of Wash Buffer 2 in PCR tubes. Aliquot three tubes of buffer for each DNA capture sample.

74 Place the aliquots with Wash Buffer 2 in the thermal cycler, with the heated lid on, held at 65 °C.

75 After the 30 min incubation period, briefly spin the capture reaction tube in a centrifuge.

76 Place the sample on a magnetic stand until the solution is clear, then remove and discard the supernatant.

77 Resuspend the beads in 200 µl of SureSelect Wash Buffer 1 and mix by pipetting until the beads are fully resuspended.

78 Incubate the sample for 15 min at RT. Afterwards, briefly spin in a centrifuge.

▲ **CRITICAL STEP** During the 15 min incubation, prepare fresh 0.1 M NaOH (see 'Reagent setup') to elute the captured library from the beads.

79 Place the sample on a magnetic stand until the solution is clear, then remove and discard the supernatant.

80 Wash the beads with 200 µl of 65 °C prewarmed Wash Buffer 2. Pipette up and down until beads are fully resuspended.

81 Cap the wells, then incubate the sample for 10 min at 65 °C on the thermal cycler.

82 Place the sample on a magnetic stand until the solution is clear, then remove and discard the supernatant.

83 Repeat washing steps Steps 80–82 twice more. Make sure all of the wash buffer has been removed during the final wash.

84 To elute the captured DNA, add 20 µl of the freshly prepared 0.1 M NaOH solution to the bead-bound sample and mix on a vortex mixer for 5 s to resuspend the beads. Then incubate the sample for 20 min at RT.

▲ **CRITICAL STEP** During the 20 min incubation, prepare the EZ DNA Methylation-Gold Kit CT conversion reagent.

85 Collect the beads from the elution mixture by placing the sample on a magnetic stand for ~2 min. Transfer the eluate, containing the captured DNA, to a new PCR tube.

Bisulfite conversion ● Timing 2 h 15 min + 2 h 30 min incubation

▲ **CRITICAL** Captured libraries are converted with the ZYMO EZ DNA Methylation-Gold Kit according to the manufacturer's protocol.

86 Prepare the CT conversion reagent mix by reconstituting one vial of solid CT conversion reagent with 900 µl of nuclease-free water, 300 µl of M-Dilution Buffer and 50 µl of M-Dissolving Buffer.
▲ **CRITICAL STEP** Prepare the appropriate number of vials for the number of samples in the run. One vial is sufficient for ten samples.

87 Mix by continuous vortexing for 10 min at RT.

88 Add 130 µl of the prepared CT conversion reagent to the 20 µl of captured library sample (from Step 85). Mix by brief vortexing, then briefly spin in a centrifuge.

89 Divide the bisulfite conversion reaction over two PCR tubes. Place the tubes in a thermal cycler and run the following program.

Step	Temperature	Time
1	64 °C	2.5 h
2	4 °C	Hold

▲ **CRITICAL STEP** When hybridization and capture are omitted (Steps 61–85), an additional step at 98 °C for 10 min is required prior to step 1 in the program of Step 89.

90 Combine the two 75 µl bisulfite conversion reactions to get a total volume of 150 µl for each DNA library.
▲ **CRITICAL STEP** Before starting the desulfonation procedure, make sure that the EtOH has been added to the M-Wash buffer provided with the EZ DNA Methylation-Gold Kit, according to the kit instructions.

91 First, add 600 µl of M-Binding Buffer to a Zymo-Spin IC column and place the column in a collection tube. Then, load the 150 µl bisulfite-converted DNA sample onto the column.

92 Cap the column and mix well by inverting the column five times. Spin down at RT for 1 min at 17,000*g*. Discard the flow-through, then place the column back in the same collection tube.

93 Wash the column by adding 100 µl of prepared M-Wash Buffer. Spin down at RT for 1 min at 17,000*g*. Discard the flow-through, then place the column back in the same collection tube.

94 Add 200 µl of M-Desulphonation Buffer to the column. Incubate at RT for 20 min.

95 Spin down at RT for 1 min at 17,000*g*. Discard the flow-through, then place the column back in the same collection tube.

96 Add 200 µl of prepared M-Wash Buffer to the column. Spin down at RT for 1 min at 17,000*g*. Discard the flow-through, then place the column back in the same collection tube.

97 Repeat Step 96 once more.

98 Spin down at RT once more for 1 min at 17,000*g*.

99 Place the column in a fresh 1.5 ml tube. Allow the column to sit at RT for 2 min.

100 Add 10 µl of M-Elution Buffer to the column and incubate at RT for 3 min.

101 Spin down at RT for 1 min at 17,000*g*.

102 Keep the flow-through in the collection tube and add an additional 10 µl of M-Elution Buffer to the column. Incubate at RT for 3 min.

103 Spin down at RT for 1 min at 17,000*g* and continue with the combined eluate for further processing.

104 To amplify the bisulfite-converted library, prepare the following PCR master mix and keep on ice.

Reagent	Volume (for one reaction)
Nuclease-free water	30 µl
SureSelect Methyl-Seq PCR Master Mix	50 µl
Methyl-Seq PCR1 Primer F	1 µl
Methyl-Seq PCR1 Primer R	1 µl
Total	82 µl

105 Add 82 µl PCR master mix to 18 µl bisulfite-converted library (from Step 103), mix briefly by vortex and spin down. Place the sample in a thermal cycler and run the following program.

Step	No. of cycles	Temperature	Time
1	1	95 °C	2 min
2	8	95 °C	30 s
		60 °C	30 s
		72 °C	30 s
3	1	72 °C	7 min
4	1	4 °C	Hold

106 Clean up the amplified bisulfite-converted library by adding 180 µl of AMPure XP beads to 100 µl of the sample (1.8× ratio).
107 Follow Steps 39–44 to carry out the cleanup.
108 Elute sample by adding 22 µl water (follow Steps 46–47) and transfer 19.5 µl of eluate to a new tube.

Library indexing ● Timing 1 h 10 min

109 Prepare the indexing PCR master mix as follows and keep on ice.

Reagent	Volume (for one reaction)
SureSelect Methyl-Seq PCR Master Mix	25 µl
SureSelect Methyl-Seq Indexing Primer Common	0.5 µl
Total	25.5 µl

▲ **CRITICAL STEP** Assign the indexing barcodes in such a way that optimal diversity is guaranteed. Consult the indexing list in the manufacturer's protocol.

110 Add 25.5 µl indexing PCR master mix to 19.5 µl amplified bisulfite-converted library (from Step 108).
111 Finally, add 5 µl of the selected indexing primer, mix briefly by vortex and spin down. Place the sample in a thermal cycler and run the following program.

Step	No. of cycles	Temperature	Time
1	1	95 °C	2 min
2	6	95 °C	30 s
		60 °C	30 s
		72 °C	30 s
3	1	72 °C	7 min
4	1	4 °C	Hold

112 Clean up the final library by adding 90 µl of AMPure XP beads to 50 µl of the sample (1.8× ratio).
113 Follow Steps 39–44 to carry out the cleanup.
114 Elute the sample by adding 26 µl water (follow Steps 46–47) and transfer 24 µl of eluate to a new tube.
115 Check the quality with a bioanalyzer DNA HS chip and the quantity with Qubit 1X DNA HS. A successful example is shown in Fig. 4d.
? TROUBLESHOOTING
116 Run the sample on an Illumina sequencing platform. A MiSeq 150 bp paired-end run gives an indication to the quality of the library. A good library can then be run on a Nextseq High 150 bp in paired-end mode.
117 Proceed to Step 179 for the computational analysis of the sequencing data.

Application B: amplicon bisulfite sequencing library

Bisulfite conversion of footprinted DNA ● Timing 1 h + 5 h 30 min incubation

▲ **CRITICAL** Footprinted DNA is converted with the Qiagen Epitect bisulfite kit based on the manufacturer's protocol with some modifications.

▲ **CRITICAL** Before starting the desulfonation procedure, make sure that the EtOH has been added to the buffer BD and BW provided with the Epitect bisulfite kit, according to the kit instructions.

118 Dissolve the required number of aliquots of bisulfite mix by adding 800 µl RNase-free water to the aliquot. One vial is sufficient for eight samples. Vortex until the bisulfite mix is completely dissolved. This can take up to 5 min.

▲ **CRITICAL STEP** If necessary, heat the bisulfite mix–RNase-free water solution to 60 °C and vortex again.

▲ **CRITICAL STEP** Do not place dissolved bisulfite mix on ice.

119 Prepare the bisulfite reaction by adding 85 µl bisulfite mix and 35 µl DNA protect buffer to 20 µl footprinted DNA (from Step 31) in a PCR tube.

120 Place the sample in a thermal cycler and run the following program.

Step	Time	Temperature
Denaturation	5 min	95 °C
Incubation	25 min	60 °C
Denaturation	5 min	95 °C
Incubation	85 min (1 h 25 min)	60 °C
Denaturation	5 min	95 °C
Incubation	175 min (2 h 55 min)	60 °C
Hold	Indefinite	20 °C

121 Next, briefly centrifuge the PCR tubes containing the bisulfite reactions and transfer to new 1.5 ml tubes.

▲ **CRITICAL STEP** Transfer of precipitates in the solution will not affect the performance or yield of the reaction.

122 Add 560 µl buffer BL to the sample. Mix the solution by vortexing and then centrifuge briefly.

123 Place the necessary number of EpiTect spin columns and collection tubes in a suitable rack. Transfer the mixture from Step 122 into the corresponding EpiTect spin column.

124 Centrifuge the spin columns at RT on maximum speed for 1 min. Discard the flow-through and place the spin columns back into the collection tubes.

125 Add 500 µl buffer BW to the spin column and centrifuge at maximum speed for 1 min. Discard the flow-through and place the spin columns back into the collection tubes.

126 Add 500 µl buffer BD to the spin column and incubate for 15 min at RT.

▲ **CRITICAL STEP** If there are precipitates in buffer BD, avoid transferring them to the spin columns. The bottle containing buffer BD should be closed immediately after use to avoid acidification from carbon dioxide in the air. It is important to close the lids of the spin columns before incubation.

127 Centrifuge the spin columns at maximum speed for 1 min. Discard the flow-through and place the spin columns back into the collection tubes.

128 Add 500 µl buffer BW to the spin column and centrifuge at maximum speed for 1 min. Discard the flow-through and place the spin columns back into the collection tubes.

129 Repeat Step 128 once more.

130 Place the spin columns into new 2 ml collection tubes and centrifuge the spin columns at maximum speed for 1 min to remove any residual liquid.

131 Place the spin columns with open lids into new 1.5 ml tubes and incubate for 5 min at 56 °C in a heating block.

132 Place the spin columns into new 1.5 ml tubes. Dispense 200 µl buffer EB onto the center of the membrane.

133 Incubate for 5 min at RT. Elute the purified DNA by centrifugation for 1 min at ~15,000*g*.

■ **PAUSE POINT** Samples can be stored at −20 °C for several weeks.

Plate-based PCR for amplicon generation ● Timing **2 h 40 min**

▲ **CRITICAL** Pipetting in this section must be performed in a PCR hood to minimize the risk of cross contamination.

134 Prepare a PCR plate by aliquoting 5 µl of each primer from the working plate of PCR primers for amplicon bisulfite sequencing library (see 'Reagent setup') to a new 96-well plate.

135 Prepare the following PCR master mix and add 11 µl per well to the PCR plate with the aliquoted primer mix.

Reagents for PCR master mix	Volume for one plate
PCR-grade water	130 µl
2× KAPA HiFi HotStart Uracil+ ReadyMix	880 µl
Bisulfite-converted DNA (Step 133)	200 µl

136 Place the sample plate in a thermal cycler and run the following program:

Step	No. of cycles	Temperature	Time
1	1	95 °C	3 min
2	35	98 °C	20 s
		56 °C	30 s
		72 °C	1 min
3	1	72 °C	5 min
4	1	4 °C	Hold

■ **PAUSE POINT** Samples should be frozen at −20 °C if you are not proceeding directly with DNA purification.

137 To verify product quality and quantity, run 5 µl of each sample on a 2% (wt/vol) TBE gel. Figure 9a shows an example where the desired amplicons range from 300 to 500 bp.

▲ **CRITICAL STEP** For bisulfite specific primers, generally an 80% success rate is expected after conversion. When first using a new set of amplicon primers, it is advisable to run aliquots from each well of the PCR plate on a gel. For subsequent experiments, it would be sufficient to check only a few amplicons.

138 Upon confirmation by gel, pool 10 µl of each reaction together.

▲ **CRITICAL STEP** Use a multichannel pipette and PCR strip tubes for convenience.

139 Take 800 µl of the pooled sample, add 640 µl of AMPure XP beads (0.8× ratio) and mix by pipetting about ten times.

▲ **CRITICAL STEP** Prior to usage, put AMPure XP beads at RT. To ensure a correct ratio is maintained, mix the beads well on a vortexer shortly before adding them to the sample.

140 Incubate the sample for 5 min at RT.

141 Place the sample on a magnetic stand until the solution is clear. This will take ~5 min.

142 While keeping the tubes on the magnetic stand, remove and discard the supernatant.

143 Wash the beads with 200 µl freshly prepared 80% (vol/vol) EtOH, while still keeping the sample on a magnetic stand.

▲ **CRITICAL STEP** Freshly prepared EtOH should not be older than 48 h as this will reduce the concentration of the washing solution and, in effect, the output of the sample.

144 Repeat Step 143 once more.

145 After the second wash, remove the supernatant completely and lay the tubes on their side to air-dry the beads. This will take ~1–3 min.

▲ **CRITICAL STEP** Air-drying the beads properly is critical. Avoid over-drying the beads, as this would result in a substantial loss of material when rehydrating them.

146 Once the beads are no longer glossy, add 52 µl water and resuspend the dried pellet by tapping the tube.

147 Incubate for 2 min at RT.

148 Place the sample on a magnetic stand until the solution is clear. This will take ~2 min.

149 Transfer 50 µl of eluate to a new PCR tube.

■ **PAUSE POINT** Samples can be stored at −20 °C for several weeks.

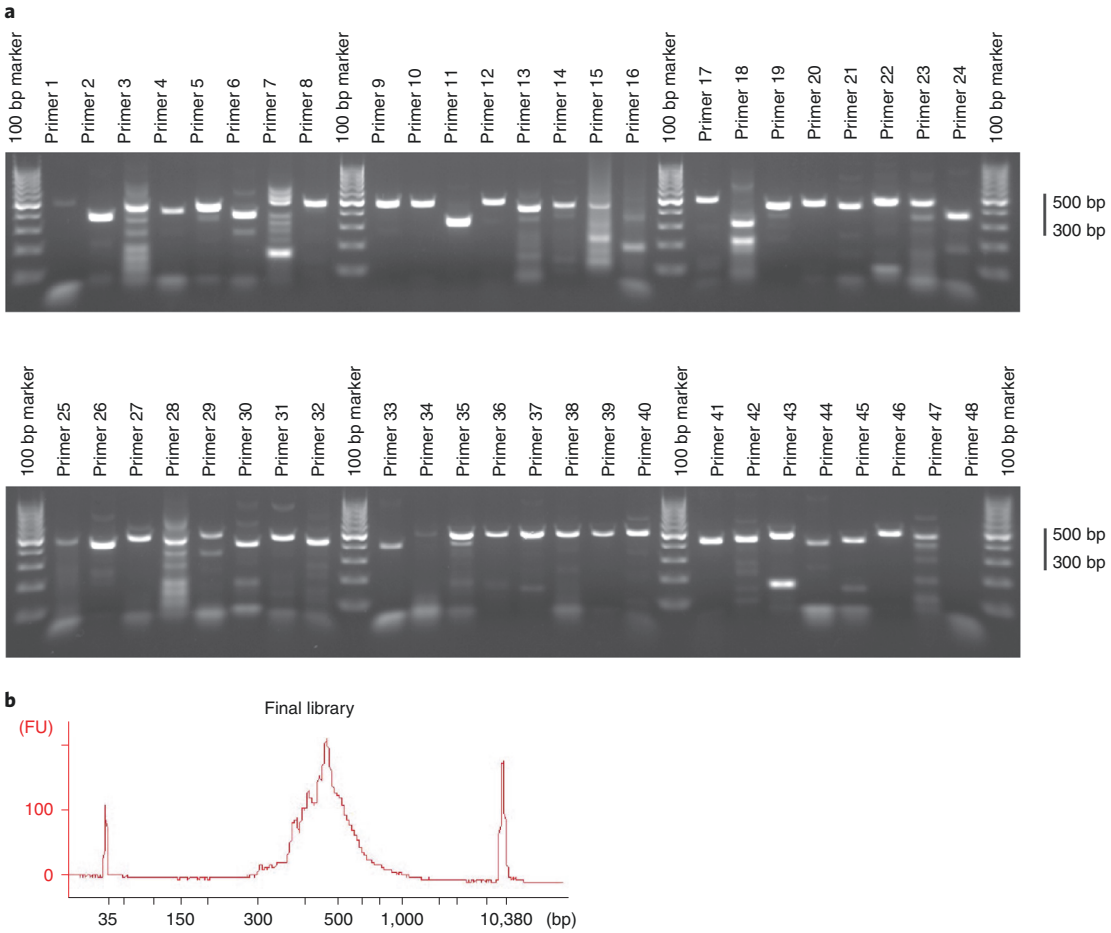

Fig. 9 | QCs during the preparation of amplicon SMF samples. 1–2 µg of footprinted DNA is bisulfite converted and used as an input for 96 parallel PCR reactions. **a**, PCR efficiency is checked by loading an aliquot on a 2% agarose gel. With standard bisulfite primer design parameters, 80–90% of the reactions lead to a detectable product and amplicon size ranges between 300 and 500 bp (Step 137). An aliquot of each PCR product is pooled and used as an input for sequencing library preparation. **b**, The size distribution of the final library is verified on an Agilent Bioanalyzer, with an expected size of 430–630 bp (Step 176).

Library preparation ● **Timing 3 h 40 min**

▲ **CRITICAL** Libraries are prepared based on NEBNext DNA Ultra II library preparation protocol.

▲ **CRITICAL** Up to 12 amplicon bisulfite samples can be multiplexed for library preparation with the NEBNext DNA Ultra II library preparation kit.

150 Quantify the amplicon pool with Qubit 1X DNA HS. Up to 1 µg in 50 µl can be used as input for library preparation.

151 Prepare the end repair master mix as follows and keep on ice.

Reagent	Volume (for one reaction)
End Prep Enzyme Mix (green cap)	3 µl
End Prep Reaction Buffer (green cap)	7 µl
Total	10 µl

152 Add 10 µl end repair master mix to 50 µl purified amplicon DNA (from Step 149), mix briefly by pipetting and spin down. Place the sample in a thermal cycler with a heated lid and run the following program.

Step	Temperature	Time
1	20 °C	30 min
2	65 °C	30 min
3	4 °C	Hold

153 Add the following ligation reagents to the end prepped sample.

Reagents	Volume
Ligation master mix (red cap)	30 µl
Ligation Enhancer (red cap)	1 µl
Adaptor (red cap)	2.5 µl

154 Mix by pipetting, quickly spin down and incubate 15 min at 20 °C without a heated lid.

155 Add 3 µl USER, mix by pipetting, quickly spin down and incubate 15 min at 37 °C with a heated lid.
▲ **CRITICAL STEP** The conditions for size selection of the adapter ligated library depend on the amplicon pool. In this case, the amplicon range is from 300 bp up to 500 bp in size. Consult the Ultra II protocol for the size selection criteria suitable to your conditions.

156 Add 17.5 µl of AMPure XP beads to 96.5 µl of adapter ligated library and mix by pipetting about ten times.
▲ **CRITICAL STEP** Prior to usage, put AMPure XP beads at RT. To ensure a correct ratio is maintained, mix the beads well on a vortex shortly before adding them to the sample.

157 Incubate the sample for 5 min at RT.

158 Place the sample on a magnetic stand until the solution is clear. This will take ~5 min.

159 While keeping the tube on the magnetic stand, transfer the supernatant to a new PCR tube.

160 Add another 17.5 µl of AMPure XP beads to the supernatant and mix by pipetting about ten times.

161 Incubate the sample for 5 min at RT.

162 Place the sample on a magnetic stand until the solution is clear. This will take ~5 min.

163 While keeping the tube on the magnetic stand, remove and discard the supernatant.

164 Wash the beads with 200 µl freshly prepared 80% (vol/vol) EtOH, while keeping the sample on a magnetic stand.
▲ **CRITICAL STEP** Freshly prepared EtOH should not be older than 48 h as this will reduce the concentration of the washing solution and, in effect, the DNA yield.

165 Repeat washing step Step 164 once more.

166 After the second wash, remove the supernatant completely and lay the tubes on their side to air-dry the beads. This will take ~1–3 min.
▲ **CRITICAL STEP** Air-drying the beads is very critical. Avoid over-drying the beads, as this would mean a substantial loss of material when rehydrating the beads.

167 Once the beads are no longer glossy, add 17 µl 0.1× TE and resuspend the dried pellet by tapping the tube.

168 Incubate for 2 min at RT.

169 Place the sample on a magnetic stand until the solution is clear. This will take ~2 min.

170 Transfer 15 µl of eluate to a new tube.

171 Add the following components to 15 µl of purified adapter-ligated library, up to a final volume of 50 µl.

Reagents for final PCR master mix	Volume
Q5 Master Mix (blue cap)	25 µl
Universal primer (blue cap)	5 µl
Indexing primer (blue cap)	5 µl

▲ **CRITICAL STEP** Assign the indexing barcodes in such a way that optimal diversity is guaranteed. Consult the indexing list in the manufacturer's protocol.

172 Mix by pipetting, spin down, place the sample in a thermal cycler with a heated lid and run the following program.

Step	No. of cycles	Time	Temperature
Initial denaturation	1	30 s	98 °C
Denaturation	3	10 s	98 °C
Annealing/extension		75 s	65 °C
Final extension	1	5 min	65 °C
Hold		Indefinite	4 °C

173 Clean up the final library by adding 45 µl of AMPure XP beads to 50 µl of the sample (0.9× ratio).

174 Follow Steps 140–145 to carry out the cleanup.

▲ **CRITICAL STEP** Air-drying the beads properly is critical. Avoid over-drying the beads; once they are no longer glossy, after ~1 min, elute the final library.

175 Elute the final library by adding 33 µl 0.1× TE (follow Steps 147–148) and transfer 30 µl of eluate to a new tube.

176 Check the quality of the final library with a bioanalyzer DNA 1000 chip and the quantity with Qubit 1X DNA HS. A successful example is shown in Fig. 9b.

177 Run the sample on an Illumina sequencing platform. Typically, a MiSeq 250 bp paired-end run gives sufficient results.

178 Proceed to Step 179 for the computational analysis of the sequencing data.

Computational analysis

Software installation ● Timing 30 min

▲ **CRITICAL** All the steps related to the computational analysis of SMF data depend on external software. Steps 179, 180, 183, 185 and 186 require bash command line usage. Steps 181, 182 and 184 involve code in R.

179 To install Trimmomatic, navigate to http://www.usadellab.org/cms/?page=trimmomatic and download the binary for the version of your choice. We recommend using version 0.36. Uncompress Trimmomatic-0.36.zip and move the resulting folder to a location of your choice. No further action is required for installation.

180 To install Picard, navigate to https://broadinstitute.github.io/picard/, uncompress the.zip file and move the.jar file to a location of your choice. No further action is required for installation. We recommend using version 2.15.0.

181 Read alignment and downstream analysis use the Biostrings-based genome data package for the species of interest. This can be installed as indicated in the relevant Bioconductor webpage and is exemplified below for *Mus musculus*:

```
if (!requireNamespace("BiocManager", quietly = TRUE))

install.packages("BiocManager")
BiocManager::install("BSgenome.Mmusculus.UCSC.mm10")
```

182 Analysis of SMF data can be performed using the SingleMoleculeFootprinting[25] R package, which can be installed as follows:

```
if (!requireNamespace("BiocManager", quietly = TRUE))

install.packages("BiocManager")
BiocManager::install("SingleMoleculeFootprinting")
```

Raw data processing ● Timing 1 d

183 With increasing read length (>100 bp), the quality of the base calling tends to decrease. To avoid errors in the methylation calls, we recommend trimming the low-quality 3′ end of the reads using the Trailing function of Trimmomatic.

Read trimming can be performed using Trimmomatic as exemplified by the following code:

```
java -jar /installation_path/Trimmomatic-0.36/trimmomatic-0.36.jar \
PE \
-threads 5 \
Sample_Name_R1.fq.gz Sample_Name_R2.fq.gz \
Sample_Name_forward_paired.fq.gz \
Sample_Name_forward_unpaired.fq.gz \
Sample_Name_reverse_paired.fq.gz \
Sample_Name_reverse_unpaired.fq.gz \
ILLUMINACLIP:/installation_path/Trimmomatic-0.36/adapters/TruSeq3-
PE.fa:2:30:10 \
LEADING:3 \
TRAILING:3 \
SLIDINGWINDOW:4:18
```

184 Align reads using the following QuasR function:

```
library(QuasR)
library(BSgenome.Mmusculus.UCSC.mm10)
proj=qAlign(sampleFile=sampleSheet,
genome="BSgenome.Mmusculus.UCSC.mm10",
aligner="Rbowtie",
projectName="ProjectName",
paired="fr",
bisulfite="undir",
alignmentsDir="./",
alignmentParameter="-e 70 -X 1000 -k 2 --best -strata",
cacheDir=tempdir())
```

▲ **CRITICAL STEP** The *sampleSheet* path to pass as *sampleFile* argument should point to a tab delimited file describing the location of the input fastq files. For more details, refer to the *qAlign* documentation https://www.rdocumentation.org/packages/QuasR/versions/1.12.0/topics/qAlign.
▲ **CRITICAL STEP** Depending on the reference genome and sample coverage, this step can be quite demanding in terms of computational resources and execution time. This is particularly relevant in case the reference genome has not yet been bisulfite converted and indexed, in which case QuasR will perform these steps internally.

185 If applicable, merge technical replicates using the following Picard utility function:

```
java -Xmx4g -jar /installation_path/picard.jar MergeSamFiles \
I=Replicate1.bam \
I=Replicate2.bam \
O=Merged.bam
```

186 Deduplicate reads by identifying and removing sequencing reads with identical start and end coordinates, as well as mapping orientation, using the following code. This step is not recommended for amplicon experiments.

```
java -Xmx4g -jar /installation_path/picard.jar MarkDuplicates \
INPUT=Merged.bam \
OUTPUT=Merged_deduplicated.bam \
METRICS_FILE=Deduplication_stats.txt \
VALIDATION_STRINGENCY=LENIENT \
REMOVE_DUPLICATES=true \
TMP_DIR=TmpDir
```

▲ **CRITICAL** Steps 187–194 contain code to be executed in R. All the necessary tools, including QuasR, are imported with *SingleMoleculeFootprinting* except for the Biostrings-based data package

containing the genome for the species of interest. The steps below use the mouse genome as an example of such a genome data package.

? TROUBLESHOOTING

Library QCs ● Timing 3 h

187 Load libraries:

```
library(SingleMoleculeFootprinting)
library(BSgenome.Mmusculus.UCSC.mm10)
```

188 To assess the quality of the sequencing libraries, we recommend inspecting canonical QC metrics that can be produced using the QuasR QC report as follows:

```
qQCReport(Qinput, pdfFilename=NULL, chunkSize=1e6L, useSampleNa-
mes=FALSE, clObj=NULL)
```

For more details, refer to the qQCreport documentation at https://www.rdocumentation.org/packages/QuasR/versions/1.12.0/topics/qQCReport.

189 The conversion rate measures the conversion of cytosines outside of methylated contexts and is expected to exceed 95%. It is generally sufficient to calculate it for a single chromosome as follows:

```
ConversionRate(sampleSheet, genome, chr)
```

? TROUBLESHOOTING

190 If applicable, the efficiency of the bait capture process, which is expected to exceed 70%, can be verified as follows:

```
BaitCapture(sampleSheet, genome, baits)
```

This function computes the ratio between the number of reads aligned to the genomic regions passed as *baits* argument over the total number of aligned reads.

For the mouse genome, we provide the coordinates of the genomic regions that are expected to be enriched due to the bait capture step as part of the *SingleMoleculeFootprintingData* package, a Bioconductor ExperimentData package allowing access to a few convenience data objects. Users can access these enrichment regions as follows:

```
SingleMoleculeFootprintingData::EnrichmentRegions_mm10.rds()
```

? TROUBLESHOOTING

191 The efficiency of footprinting can be evaluated from a shallow sequencing experiment by using the *LowCoverageMethRateDistribution* function distributed in the dev version of the *SingleMoleculeFootprinting* package, which can be installed as follows.

```
remotes::install_github(
repo = "https://github.com/Krebslabrep/SingleMoleculeFootprinting.git",
ref = "dev", build_vignettes = FALSE)
```

The LowCoverageMethRateDistribution function can be used in the following form:

```
LowCoverageMethRateDistribution(LowCoverage,
 LowCoverage_samples,
 HighCoverage,
 HighCoverage_samples,
 returnDF = FALSE,
 returnPlot = TRUE,
 MSE = TRUE,
 return_MSE_DF = FALSE,
```

`return_MSE_plot = TRUE)` where the arguments LowCoverage and HighCoverage are the GRanges objects returned by calling the function CallContextMethylation (see next step for details) on the low coverage sample and the high coverage reference one, respectively. For details on the other arguments, consult the manual page for this function as follows: **`?LowCoverageMethRateDistribution`**

▲ **CRITICAL STEP** The *coverage* argument to the *CallContextMethylation* function should be set to 1 when used on the low coverage sample, in order to unbiasedly estimate methylation rates.

▲ **CRITICAL STEP** Ideally, *LowCoverageMethRateDistribution* should be run using data for a whole chromosome (e.g., chr19 for mouse). However, the memory required to call methylation at the single-molecule level would be quite elevated, and this level of detail is not required here. Therefore, we suggest employing the *CallContextMethylation* function from the *dev* version of the package, where the argument *returnSM* can be set to FALSE, thus returning the bulk methylation only. For additional details, we recommend consulting the manual page of the function.

? TROUBLESHOOTING

SMF data analysis ● Timing 5 h

192 Extract methylation calls at the single-molecule level for a region of interest as follows:

```
CallContextMethylation(sampleSheet, sample, genome, range, coverage,
ConvRate.thr)
```

This function calls methylation events for the cytosines in the genomic context relevant for the experiment (single enzyme, double enzyme, etc.), filters reads based on conversion rate, collapses strands and filters cytosines for coverage.

The output consists of a list of two objects. The first is a GRanges object reporting the average methylation values for the covered cytosines in the region of interest. The second is a matrix reporting the binary methylation information at the single-molecule level. This object can be used directly for plotting or for single-read sorting as explained in Steps 193 and 194.

The argument sample should be a string for one of the sample names as it appears in the SampleName field of the sampleSheet file.

The argument range should be a GRanges object that can be defined as follows:

```
GRanges(seqnames = "chr6",
 ranges = IRanges(start = 88106000, end = 88106500),
 strand = "*")
```

▲ **CRITICAL STEP** Note that the resulting matrix can require a large amount of memory. We, therefore, recommend keeping the region of interest, passed as the *range* parameter, short, which, depending on the user's system, can mean 10^5–10^7 bp.

▲ **CRITICAL STEP** For the quantification of occupancy of a single TF, we recommend setting the *coverage* argument to 20.

▲ **CRITICAL STEP** The argument *ConvRate.thr* should be used with caution: while it marginally improves final results by filtering few reads with atypical conversion rates, it can lead to skewed results towards reads with too few 'out of context' cytosines. For example, the desired behavior when setting the value of *ConvRate.thr* to 0.8 is to discard reads that have at least 20% of their 'out of context' cytosines unconverted. The length of Illumina reads can easily cause the number of 'out of context' cytosines to be insufficient (e.g., three cytosines) for accurate estimation of the true conversion rate. We therefore suggest setting a threshold only when suspecting the conversion rate to be the cause of artifacts.

193 To sort single reads based on their occupancy patterns, the user can employ either of the following functions depending on whether the sorting is to be carried over a single TF or multiple TFs.

```
SortReadsBySingleTF(MethSM, TFBS)
SortReadsByTFCluster(MethSM, TFBSs)
```

SortReadsBySingleTF will sort reads considering a single TF (Fig. 8a,b) whose genomic location should be indicated through the TFBS argument in the form of a GRanges object of length 1. In this case, the default sorting bins are defined by the coordinates [−35;−25], [−15;15] and [25;35] relative to the center of the TFBS.

SortReadsByTFCluster will sort reads considering multiple TFs bound at one locus (Fig. 8c). The argument TFBSs should be a GRanges of length ≥1. The function will design sorting bins flanking the TFBS cluster with coordinates [−35;−25] and [25;35], and at each of the binding sites composing the cluster [−7;7].

In case the user wishes to customize the design of the sorting bins, the *SortReads* function can be used with the argument *SortByCluster*=FALSE and the argument *BinsCoord* passed as a list of vectors, each containing the relative coordinates of the desired bins, as follows:

```
SortReads(MethSM, TFBS, BinsCoord=list(c(.,.),…, c(.,.)),
SortByCluster=FALSE)
```

▲ **CRITICAL STEP** As discussed in 'Limitations', the outcome of sorting fundamentally depends on the genomic location of the TFBS(s) of interest. Therefore, we strongly advise the user to employ TFBSs identified with high confidence (e.g., high positional weight matrix match score and/or ChIP-seq evidence).

▲ **CRITICAL STEP** For a single molecule to undergo sorting, each sorting bin has to be covered by at least one cytosine for which the methylation state has been observed. If any of the bins lack information (e.g., nucleotides trimmed due to poor quality), the read will not be included in the analysis.

194 Three types of information can be visualized for a single locus: the average SMF signal, the single-molecule SMF signal and a bar plot showing the fractions of molecules classified in each state. These can be obtained either individually or simultaneously using the following functions:

```
PlotAvgSMF(MethGR, range, TFBSs)
PlotSM(MethSM, range, SortedReads)
StateQuantificationPlot(SortedReads)
PlotSingleSiteSMF(ContextMethylation, sample, range, SortedReads,
TFBSs, saveAs)
```

The argument *TFBSs* should be a GRanges object of TF binding site coordinates to plot. It is expected to be already filtered for the sites that fall inside the genomic region passed in range.

The argument *SortedReads* for the last three functions can be set either to NULL to visualize unsorted reads, to 'HC' to perform hierarchical clustering (useful to spot PCR duplicates) or to the output of either of the *SortReads* functions from Step 193.

▲ **CRITICAL STEP** The argument *range* can be the same as the one passed to the function *CallContextMethylation* as long as the range width is less than a few hundred base pairs. More than that will impair the visual readability of the results since the SMF signal is most informative when examined within a width compatible with footprints left by TFs and nucleosomes.

▲ **CRITICAL STEP** For additional documentation and details on the functionalities of the *SingleMoleculeFootprinting* package, we recommend consulting the package vignette and the manual pages for the individual functions, which can be accessed by typing ?<function>.

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
34	Fragment size shifted towards higher molecular weight	DNA shearing is insufficient	Increase sonication by 20–40 s or repeat sonication. In case the issue remains, repeat the sonication with a new DNA extraction
115	Low concentration of the final library	Failed capture	Repeat the hybridization, using fresh beads for capture
			Check the expiration date of the baits. Check the thermal cycler; ensure that the lid is set to 65 °C to avoid sample evaporation

Table continued

Table 1 (continued)

Step	Problem	Possible reason	Solution
125	High PCR duplication rate (>20%), low-complexity library	Too low input	Increase input material for the hybridization (>1 μg), pooling several ligated reactions if needed
189	Low conversion rate	Conversion kit either too old or ineffective	Check expiration date and replace conversion kit
			When using an already prepared conversion mix, do not keep it for >1 week at −20 °C
190	Bait capture efficiency <70%	No enrichment of library	Repeat the hybridization, using fresh beads for capture
			Check the expiration date of the baits
			Check the thermal cycler; ensure that the lid is set to 65 °C to avoid sample evaporation
191	Footprinting is incomplete, methylation not saturated	Cell suspension not homogeneous	Ensure proper cell dissociation. Check the quality of the nuclear preparation and carefully verify the number of cells used
		Nuclear extraction incomplete	Repeat nuclear extraction, checking with microscopy. Specific cell types may require optimization
		Methyltransferase activity insufficient	Replace enzyme
			Check activity by in vitro methylation and digestion of genomic DNA with methyl-sensitive restriction enzymes such as HpaII for CpG methylation and BbvI for GpC methylation
		Insufficient amounts of SAM in the reaction	Use fresh SAM aliquots
		Strong variation of footprinting between samples	Keep incubation times similar between samples. Also do not process more than eight samples per round

Timing

Day 1
Steps 1–15, footprinting the DNA: 2 h 30 min

Day 2
Steps 16–32, isolating the DNA: 2 h
Steps 118–120 (B), bisulfite conversion: 10 min + 5 h 30 min incubation

Day 3
Steps 35–60 (A), library preparation up to hybridization: 6 h + 16 h incubation
Steps 121–149 (B), bisulfite conversion, amplicon PCR and gel check: 3 h 30 min

Day 4
Steps 61–117 (A), capture and final library preparation and sequencing: 8 h 30 min
Steps 150–178 (B), library preparation and sequencing: 3 h 40 min

Day 5
Steps 179–192, software installation and raw data processing: 1 d

Day 6
Steps 192–194, library QC and SMF analysis: 1 d

Anticipated results

Preparation of the libraries
A typical SMF reaction yields 0.7–1 μg of DNA. A minimum of 1 μg of DNA is required to prepare a genome-wide bisulfite library (with optional bait capture). We recommend performing three SMF

reactions in parallel yielding in total 2–3 μg of input DNA for the library preparation (corresponding to 0.75×10^6 mouse or human cells and 7.5×10^6 *Drosophila* cells). A successful library preparation yields 60–120 ng of DNA. Approximately 1 μg of DNA is recommended to perform an amplicon bisulfite library over 96 targets. The amount of DNA can be scaled down when targeting fewer loci. However, reducing the input DNA in individual reactions can lead to a reduction of library complexity and an increase in the proportion of identical molecules in the resulting datasets. Decreased complexity can be detected by an increase of the duplication rates above 20% in genome-wide experiments. This can be circumvented by increasing the DNA used as an input to the reaction.

Sequencing depth requirements

Accurate quantification of TF binding at single-molecule resolution requires generating sufficient sequencing reads spanning all the bins surrounding individual TFBS (Fig. 8a). Minimum coverage of 40 reads is recommended. This is sufficient to reproducibly quantify binding frequencies above 20%, where 8 reads out of 40 reads would have a TF footprint. Lower binding frequencies will have lower counts of bound molecules, and accurate quantification of the binding frequency requires increasing the total number of molecules sequenced at the locus. Moreover, when quantifying the binding of multiple TFs in a cluster, the classification algorithm will only consider the reads spanning all the studied binding sites (Fig. 8b). Therefore, it is important to adjust coverage as a function of the number of loci considered, the binding frequency of the TFs analyzed and the distance between the TFBS in a cluster. We recommend multiplexing a maximum of two SMF samples on a NextSeq 500, producing ~200×10^6 read pairs of length 150 bp, sequencing a total of 6×10^{10} nucleotides. Out of these, ~40% will not map and ~20% will be removed as PCR duplicates, leaving ~2.88×10^{10} analyzable nucleotides. This represents a theoretical coverage of ~144 times the *Drosophila* genome and ~360 times the 80 Mb of the mouse genome captured by the baits. However, reads are not evenly distributed in the genome, particularly when a capture step is applied, as pulldown efficiency varies between baits. Moreover, only a fraction of the reads covering a locus will span all the bins used to analyze the binding of one or more TFs (Fig. 8). In practice, this amount of sequencing leads to 50–100 usable reads to quantify TF binding.

Number of TF binding sites analyzed

SMF allows single-molecule quantification of the binding of a TF at a fraction of its binding sites (Fig. 2). The ability to analyze a given TFBS is defined by the sequence composition at its flanks because the classification of reads in various states requires measuring methylation in each of the bins used for sorting (Fig. 8). Each bin should contain informative cytosines that are in the GpC context when performing the experiment with M.CviPI, and in combination with CpGs when performing the treatment with M.SssI. Combining the two methyltransferases substantially extends the number of TFBS that can be analyzed (Fig. 2). The extent of the increase varies depending on the identity of the TF. For instance, there is only a moderate increase in the number of REST sites quantified in a double enzyme experiment (from ~200 to ~600 sites) while between seven and eight times more NRF1 binding sites can be analyzed (from ~900 to ~6,700) (Fig. 2). The number of analyzable binding sites is in the range of thousands for most TFs. This is sufficient to determine their global binding properties, and visualize their behavior at example loci. However, it is important to check the dinucleotide composition of target loci before performing an SMF experiment.

Data availability

The data used to produce Fig. 6 and Fig. 7 were produced within the scope of ref. [8] and are available at ArrayExpress: E-MTAB-9123 and E-MTAB-9033. The data used to produce Fig. 3 are available at ArrayExpress: E-MTAB-10815.

Code availability

The SingleMoleculeFootprinting[25] R package has been released and is available through Bioconductor. The code used to produce the figures for this paper is available at https://github.com/KrebsLab/Kleinendorst_et_al[31].

References

1. Raha, D., Hong, M. & Snyder, M. ChIP-Seq: a method for global identification of regulatory elements in the genome. *Curr. Protoc. Mol. Biol.* https://doi.org/10.1002/0471142727.mb2119s91 (2010).

2. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
3. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protoc.* https://doi.org/10.1101/pdb.prot5384 (2010).
4. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
5. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* **43**, 73–81 (2017).
6. Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* **47**, 1–8 (2017).
7. Ibarra, I. L. et al. Mechanistic insights into transcription factor cooperativity and its impact on protein–phenotype interactions. *Nat. Commun.* **11**, 124 (2020).
8. Sönmezer, C. et al. Molecular co-occupancy identifies transcription factor binding cooperativity in vivo. *Mol. Cell* **81**, 255–267.e6 (2021).
9. Kelly, T. K. et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).
10. Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* **67**, 411–422.e4 (2017).
11. Nabilsi, N. H. et al. Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res.* **24**, 329–339 (2014).
12. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
13. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
14. Abdulhay, N. J. et al. Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).
15. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
16. Krebs, A. R. Studying transcription factor function in the genome at molecular resolution. *Trends Genet.* https://doi.org/10.1016/j.tig.2021.03.008 (2021)
17. Minnoye, L. et al. Chromatin accessibility profiling methods. *Nat. Rev. Methods Prim.* **1**, 10 (2021).
18. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
19. Cusanovich, D. A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
20. Levo, M. et al. Systematic investigation of transcription factor activity in the context of chromatin using massively parallel binding and expression assays. *Mol. Cell* **65**, 604–617.e6 (2017).
21. Oberbeckmann, E. et al. Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* **29**, 1996–2009 (2019).
22. Untergasser, A. et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115–e115 (2012).
23. Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130–1132 (2015).
24. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
25. Barzaghi G, K. A. *SingleMoleculeFootprinting. Analysis tools for Single Molecule Footprinting (SMF) data. R package version 1.0.0*. (Bioconductor, 2021). https://doi.org/10.18129/B9.bioc.SingleMoleculeFootprinting
26. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkz1001 (2019)
27. Puig, R. R. et al. UniBind: maps of high-confidence direct TF-DNA interactions across nine species. *BMC Genomics* https://doi.org/10.1186/s12864-021-07760-6 (2021).
28. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
29. Domcke, S. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015).
30. Nicola, N. A. & Babon, J. J. Leukemia inhibitory factor (LIF). *Cytokine Growth Factor Rev.* **26**, 533–544 (2015).
31. GuidoBarzaghi, S. M. & KrebsLab. Krebslabrep/SingleMoleculeFootprinting: SingleMoleculeFootprinting. *Zenodo* https://doi.org/10.5281/ZENODO.4767134 (2021).

Acknowledgements

The authors are grateful to the members of the Krebs laboratory for helpful discussions, comments on the manuscript and feedback during the development of the R package. The authors would like to acknowledge C. Sönmezer for sharing data and E. Kreibich for sharing amplicon QC gels. The authors are thankful to L. Villacorta, V. Benes and the members of the Genomics Core facility for sequencing the libraries and technical assistance. The salary of G.B. is supported by the Deutsche Forschungsgemeinschaft (KR 5247/1-1). The authors thank W. Huber for supporting the development of the R package. The authors would like to thank C. Girardot and the

Genome Biology Computational Support. Research in the laboratory of A.R.K is supported by core funding of the European Molecular Biology Laboratory, Deutsche Forschungsgemeinschaft (KR 5247/1-1 and KR 5247/2-1). M.L.S. is funded by The German Network for Bioinformatics Infrastructure (de.NBI) Förderkennzeichen Nr. 031A537B.

Author contributions

A.R.K designed the study. R.W.D.K., G.B. and A.R.K wrote the manuscript. R.W.D.K performed the experiments. G.B. developed the package for data analysis with support from M.L.S. A.R.K supervised the conduction of the experiments and the data analysis. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Arnaud R. Krebs.

Peer review information *Nature Protocols* thanks Simon Bourdareau, Julia Zeitlinger and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 February 2021; Accepted: 6 September 2021;
Published online: 12 November 2021

Related links

Key references using this protocol

Sönmezer, C. et al. *Mol. Cell* **81**, 255–267 (2021): https://doi.org/10.1016/j.molcel.2020.11.015

Krebs, A. et al. *Mol. Cell* **67**, 411–422 (2017): https://doi.org/10.1016/j.molcel.2017.06.027