GwEEP - A comprehensive approach for genome-wide efficiency profiling of DNA modifying enzymes

Charalampos Kyriakopoulos[1,2], Karl Nordström[3,4], Paula Linh Kramer[1], Judith Gottfreund[3], Abdulrahman Salhab[3], Julia Arand[5], Fabian Müller[6], Ferdinand von Meyenn[7], Gabriella Ficz[8], Wolf Reik[9], Verena Wolf[1], Jörn Walter[3], and Pascal Giehr [*3,7]

[1]Computer Science Department, Saarland University, Campus E1.3, 66123 Saarbrücken, Germany
[2]Bristol Myers Squibb, Center for Innovation and Translational Research Europe, Calle Isaac Newton 4, 41092 Sevilla, Spain
[3]Department of Genetics and Epigenetics, Saarland University, Campus A2.4, 66123 Saarbrücken, Germany
[4]AstraZeneca, Pepparedsleden 1, 431 50 Mölndal, Schweden
[5]Division of Cell and Developmental Biology, Medical University of Vienna, 1090 Vienna, Austria
[6]Department of Integrative Cellular Biology and Bioinformatics, Campus A2.4, 66123 Saarbrücken, Germany
[7]Department of Health Sciences and Technology, ETH Zürich, Schorenstrasse 16, 8603 Schwerzenbach, Zürich, Switzerland
[8]Haemato-Oncology, Queen Mary University of London, EC1M 6BQ London, UK
[9]Epigenetics Department, Babraham Institute, CB22 3AT Cambridge, UK

Abstract

A precise understanding of DNA methylation dynamics on a genome wide scale is of great importance for the comprehensive investigation of a variety of biological processes such as reprogramming of somatic cells to iPSCs, cell differentiation and also cancer development. To date, a complex integration of multiple and distinct genome wide data sets is required to derive the global activity of DNA modifying enzymes. We present GwEEP - *Genome-wide Epigenetic Efficiency Profiling* as a versatile approach to infer dynamic efficiency changes of DNA modifying enzymes at base pair resolution on a genome wide scale. GwEEP relies on genome wide *oxidative Hairpin Bisulfite sequencing* (HPoxBS) data sets, which are translated by a sophisticated hidden Markov model into quantitative enzyme efficiencies with reported confidence around the estimates. GwEEP in its present form predicts *de novo* and maintenance methylation efficiencies of Dnmts, as well as the hydroxylation efficiency of Tets but its purposefully flexible design allows to capture further oxidation processes such as formylation and carboxylation given available data in the future. Applied to a well characterized ES cell model, GwEEP precisely predicts the complex epigenetic changes following a Serum-to-2i shift i.e., (i) instant reduction in maintenance efficiency (ii) gradually decreasing de novo methylation efficiency and (iii) increasing Tet efficiencies. In addition, a complementary analysis of Tet triple knock-out ES cells confirms the previous hypothesized mutual interference of Dnmts and Tets. GwEEP is applicable to a wide range of biological samples including cell lines, but also tissues and primary cell types.

Keywords: Epigenetics, Dnmts, Tets, DNA methylation, hairpin sequencing, hidden Markov model

[*]corresponding author (pgiehr@ethz.ch)

MOTIVATION

Dynamic changes of DNA methylation patterns are a common phenomenon in epigenetics. Although a stable DNA methylation profile is essential for cell identity, developmental processes require the rearrangement of 5-methylcytosine in the genome. Stable methylation patterns are the result of balanced Dnmts and Tets activities, while methylome transformation results from a coordinated change in Dnmt and Tet efficiencies. Such transformations occur on a global scale, for example during the reprogramming of maternal and paternal methylation patterns and the establishment of novel cell type specific methylomes during embryonic development *in vivo*, but also *in vitro* during (re)programming of induced pluripotent stem cells, as well as somatic cells. In addition, local (de)methylation events are key for gene regulation during cell differentiation. A detailed understanding of Dnmt and Tet cooperation is essential for understanding natural epigenetic adaptation as well as optimization of *in vitro* (re)programming protocols. For this purpose, we developed a pipeline for quantitative and precise estimation of Dnmt and Tet activity. Using only double strand methylation information, GwEEP infers accurate maintenance and *de novo* methylation efficiency of Dnmts, as well as hydroxylation efficiency of Tets at single base resolution. Thus, we believe GwEEP provides a powerful tool for the investigation of methylome rearrangements in various systems.

INTRODUCTION

Genetic information encoded in the DNA is regulated by epigenetic mechanisms, such as DNA methylation [1, 2, 3, 4]. In mammals, methylation of DNA is restricted to cytosine and is almost exclusively found in a palindromic CpG dinucleotide context [5, 6, 7]. Generation of 5-methylcytosine (5mC) is catalyzed by the DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. These enzymes catalyze the transfer of a methyl group from s-adenosyl methionine to the fifth carbon atom of cytosine. Dnmt1 is responsible for maintaining existing methylation patterns after replication. Via interaction with Uhrf1 and PCNA, Dnmt1 is tightly associated with the replication machinery [8, 9]. Furthermore, the cooperation with Uhrf1 modulates Dnmt1 to be receptive for hemimethylated DNA generated after replication [10, 11]. Thus, the protein complex post-replicatively copies the methylation pattern from the inherited to the newly synthesized DNA strand [12, 13]. Dnmt3a and Dnmt3b methylate DNA independently of its methylation status (hemimethylated or unmethylated) and are mainly responsible for the establishment of new methylation patterns during development [14, 15]. However, several studies indicate that the strict separation of Dnmt1 and Dnmt3a/b activity is not coherent and that under certain conditions, these enzymes exhibit overlapping functions [16, 17, 18]. Once established, 5mC can be further processed by a family of di-oxigenases, the ten-eleven translocation enzymes Tet1, Tet2 and Tet3 [19, 20, 21]. These Fe(II) and oxoglutarate-dependent enzymes consecutively oxidize 5mC to 5-hydroxymethyl cytosine (5hmC), 5-formyl cytosine (5fC) and ultimately to 5-carboxy cytosine (5caC) [22, 23]. 5hmC is the most abundant oxidative variant and can be found in numerous cell types [24, 25, 26]. Each oxidation step changes the chemical properties of the base and with it its biological function [27, 28, 29]. Several mechanisms have been proposed in which oxidative cytosine derivatives (oxC) serve as an intermediate during the course of active or passive demethylation [30, 31, 32, 33, 34]. Such removal of 5mC occurs locally during cell differentiation, but also on a genome-wide scale in the zygote, as well as during the maturation of primordial germ cells (PGCs) [35, 36, 37]. Global loss of 5mC has been observed in cultivated mouse embryonic stem cells (ES cells) during their transition from Serum to 2i medium. Under classical Serum/LIF conditions, ES cells exhibit DNA hypermethylation, whereas upon transition to GSK3 and Erk1/2 inhibitors (2i) containing medium, the cells experience a gradual genome-wide loss of 5mC [38, 39, 40]. Even though several studies examined the influence of Tets and oxCs [41, 42, 43, 44] within the genome, the precise contribution of Tets and oxCs towards maintaining or change existing cell type specific methylomes remains elusive. A contextual understanding of the local and spatial connections between Dnmts and Tets during processes of development, cell division and differentiation is of great importance, as they build the basis for a rational development of novel epigenetic cancer therapies and controlled reprogramming approaches in regenerative stem cell medicine.

To address the complex interplay between Dnmts and Tets we developed a comprehensive experimental and computational pipeline - GwEEP (**G**enome **w**ide **E**nzyme **E**fficiency **P**rofiling), which determines the efficiencies of DNA modifying enzymes across the genome. GwEEP comprises genome wide hairpin (oxidative) bisulfite sequencing - HP(ox)BS, a pipeline for extracting the double strand methylation pattern from HPBS, as well as a sophisticated hidden markov model to infer Dnmt's *de novo* and maintenance methylation, as well as Tet's hydroxylation efficiencies. Here we provide a detailed protocol for the generation of hairpin (oxidative) bisulfite libraries and furthermore, a detailed description data processing and HMM application. Lastly, we provide an application example where we applied GwEEP to a widely used model system, following the adaptation of ES cells during their transition from Serum/LIF to 2i based media.

MATERIAL

Oligo ID	**Short ID**	**Sequence [5'-to-3']**	**Modifications**	**Purification**
Hairpin linker	HP	GGGCCTADDD1DDDTAGGCCCT	1 = dT-biotin; 5'-phosphate	HPLC
Sequencing adapter top	SeqAdptTop	TCTTTCCCTACACGACGCTCTTCCGATCT	2 = 5mC	HPLC

Sequencing adapter bottom	SeqAdptBot	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC	2 = 5mC; 5'-phosphate	HPLC
Enrichment primer forward	EnPrimerFor	CAAGCAGAAGACGGCATACGAGAT[i7]CGTGATG-TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	none	HPLC
Enrichment primer reverse	EnPrimerRev	AATGATACGGCGACCACCGAGATCTACAC[i5]TC-TTTCCCTACACGACGCTCTTCCGATCT	none	HPLC

Enzyme	**Manufacturer**	**Catalog #**	**Application**
R.AluI	New England Biolabs (NEB)	R0137	Digestion
R.HaeIII	NEB	R0108	Digestion
R.HpyCH4V	NEB	R0619	Digestion
Klenow exo-	NEB	M0212	A-tailing
T4 DNA Ligase	NEB	M0202	Ligation
HotStarTaq DNA Polymerase	Qiagen	203207	Enrichment PCR

Chemical	**Manufacturer**	**Catalog #**	**Application**
Acetonitile	Fisher Chemical	A/0627/17	Clean-up
Ethanol	Fisher Chemical	E/0650DF/15	Clean-up
NaCl	ZChl UdS	701313	HP enrichment
NaOH	ZChl UdS	701340	HP enrichment
TrisHCl	Roth	9090.3	TE buffer
EDTA	Grüssing	103051000	TE buffer
SDS	BioChemica	A2572	HP enrichment
ATP	NEB	P0756S	HP ligation
dATP	Solis BioDyne	02-21-00100	A-Tailing
dNTPs	Solis BioDyne	02-21-00100	Enrichment PCR
CutSmart Buffer	NEB	B7204	Digestion
PCR Buffer	Qiagen	1005479	Enrichment PCR
$MgCl_2$	Qiagen	1005482	Enrichment PCR
Dynabeads™ M-280 Streptavidin	ThermoFisher	11205D	HP enrichment
AMPureXP® beads	Beckman Coulter	A63880	Clean-up
TrueMethyl® oxBS-Seq Module	TECAN	0414-32	BS and oxBS
Qubit™dsDNA HS Kit	Invitrogen	Q32851	Library QC
High Sensitivity DNA Kit	Agilent	5067-4626	Library QC
High Sensitivity DNA Reagents	Agilent	5067-4627	Library QC

Scripts and Pipelines	**Function**	**Resource**
HPpipeline	Processing of FASTQ files i.e. sequencing data	https://github.com/PascalGiehr/HPup
HMM (Python version)	Estimation of 5mC and 5hmC states, as well as enzyme efficiencies	https://github.com/PaulaKramer/GwEEP-tool

METHODS

Our pipeline consists of three major parts: (i) the construction of an oxidative hairpin bisulfite library (RRHPoxBS), (ii) a computational pipeline (HPup) which extracts the double strand DNA methylation values from Illumina sequencing data and (iii) a hidden Markov model (GwEEP-tool) that calculates conversion error corrected 5mC and 5hmC distributions and furthermore infers the enyzmatic efficiencies of Dnmts (maintenance and *de novo* methylation efficiency), as well as Tets (hydroxylation efficiency)

RRHPoxBS - Reduced Representation Hairpin oxidative Bisulfite Sequencing

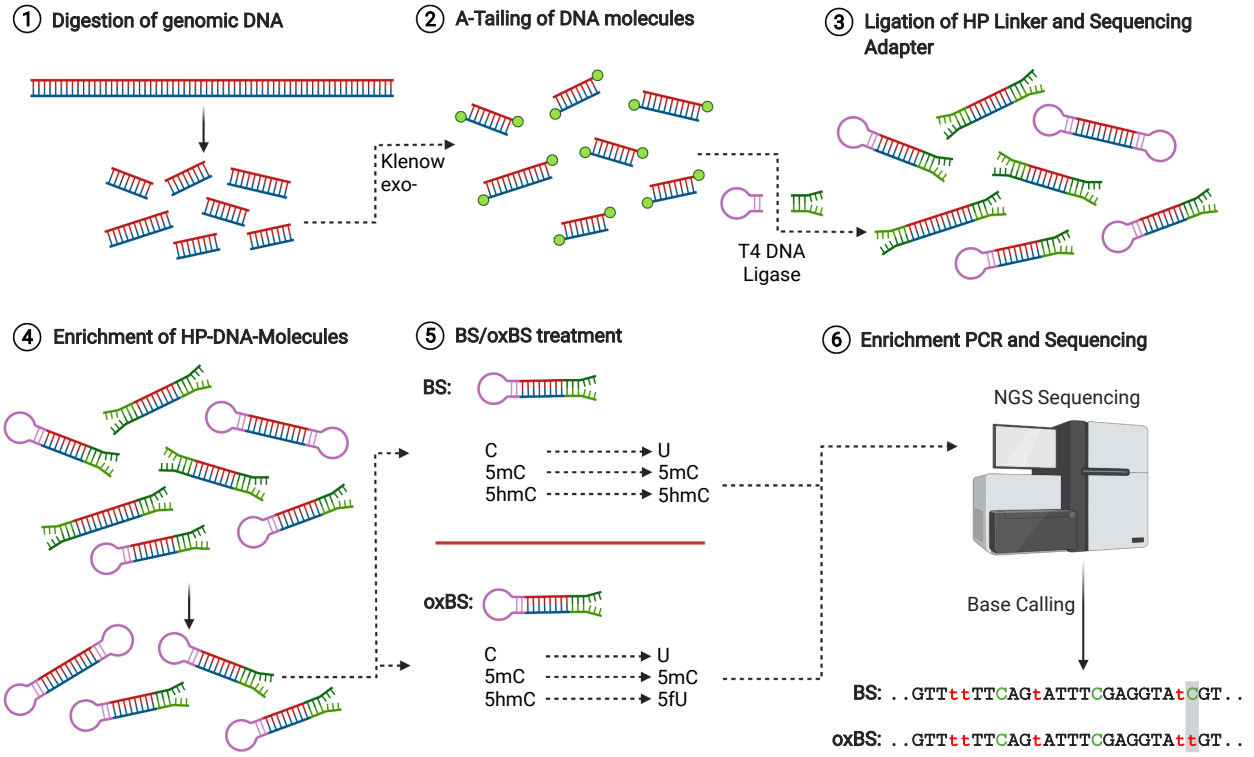


Figure 1: Laboratory Pipeline Overview - (1) genomic DNA is digested by endo nucleases, followed by (2) Klenow exo- catalyzed A-tailing. (3) A-Tailed DNA molecules are subjected to sequencing adapter and hairpin linker ligation and (4) subsequent enrichment of hairpin-ligated molecules. (5) half of the library is used for Bs, the other half for oxBs treatment. (6) After amplification and indexing using PCR, the libraries are sequenced on an Illumina platform with minimum 100 bp in a paired-end mode. Created with BioRender.com

Digestion of genomic DNA

For GwEEP we use endo nucleases (RE) in order to enrich for selected genomic regions. This increases the sensitivity of the assay (as little as 18 ng of DNA can be used - SI Sec 14) and, at the same time, reduces the sequencing costs considerably. The RE must not be sensitive against 5mC and 5hmC, in order to avoid a bias of the methylation analysis by blocked restriction. Ideally, the recognition/restriction site of the RE should not contain any CpG dyad. In the present protocol, we used a combination of three different REs i.e., R.AluI, R.HaeIII and R.HpyCH4V to obtain reads from regulatory (HaeIII), as well as inter and intra genomic regions (AluI and HpyCH4V). For the customization of GwEEP we recommend cuRRBS, which determines the best suited RE(s) based on a list of *regions of interest* [45].

Unless the amount of DNA is limited, we strongly advice a proper quality control using *Agarose-gel-elec-trophoresis* and *Qubit Fluorometer* based quantification. The best results are obtained by using 300 - 400 ng high quality DNA per used RE. However, as little as 2 ng per enzyme (6 ng in total) can be used. For each RE, prepare the reaction outlined in Table 2 and incubate for at least 3h at the temperature optimum of the RE (here 37°C), followed by a 20 min heat inactivation at 65°C or 80°C.

NOTE: Incubation over night might increase the sensitivity of the assay, especially when using low DNA amounts or dealing with sample impurities. However, this is only recommended for RE that do not exhibit any star activity.

After inactivation the reactions are combined and subjected to a purification step using magnetic beads i.e., SPRI® or AmpureXP® beads. Use a sample:bead ratio of 1:2 (60μl:120μl) for clean-up, wash twice with 200 μl freshly prepared 80% ethanol (EtOH) and elute the DNA in 17.5 μl of 1x CutSmart buffer.

At this stage we strongly recommend to add a suitable spike-in in order to calculate conversion rates of bisulfite and oxidative bisulfite reactions independently from the biological sample. We apply 4 pg of Sequencing Spike-in Control (CEGX) per sample (SI Sec. 6).

Table 2: Digest of genomic DNA

Component	**Concentration**	**Volume [μl]**
Genomic DNA	2 - 400 ng	X
Endo Nuclease (R.AluI, R.HaeII or R.HpyCH4V)	10 U	Y
CutSmart Buffer	10x	2.0
ddH_2O	–	ad 20.0
Total Volume		20.0

Table 3: Addition of Spike-in Control

Component	**Concentration [ng/μl]**	**Volume [μl]**
Digested DNA	-	17.5
Sequencing Spike-in Control	0.008	0.5
Total Volume		18.0

End-repair and A-tailing

In the present study, we applied REs which generate blunt-end DNA fragments and therefore require a subsequent A-tailing. For this, we rely on the polymerase I large (Klenow) fragment, which lacks both $5' - > 3'$, as well as $3' - > 5'$ exonuclease activity (Klenow exo-). The reaction given in Table 4 is incubated for 30 min at 37°C and inactivated for 20 min at 75°C:

Table 4: A-tailing reaction

Component	**Concentration**	**Volume [μl]**
Digested DNA	–	18.0
Klenow exo-	5 U/μl	1.0
dATP in 2x CutSmart	1 mM	1.0
Total Volume		20.0

NOTE: In case REs producing sticky-ends are used, the above reaction is not sufficient for library preparation. Instead, please rely on an end-repair and A-tailing workflow.

The A-tailing prevents self-ligation of DNA fragments and facilitates the ligation of hairpin linker (HP) and sequencing adapter (SA). Purification after heat-inactivation of Klenow exo- is not required. Instead, proceed immediately to section *Ligation of hairpin-linker and sequencing adapter*.

Ligation of hairpin-linker and sequencing adapter

Prior to the ligation reaction, SA and HP have to be transferred into a double stranded state. First, solve each oligo nucleotide in 1xTE for a final concentration of 100μM. Next, join 10 μl SeqAdptTop and 10 μl SeqAdptBot in a 0.2 ml reaction tube and fill 20μl of HP in a second 0.2 ml reaction tube. Place both tubes into a thermocylcer and incubate for 5 min at 95°C followed by a slow cool-down of about 0.3°C/sec. This will facilitate the annealing of SeqAdptTop and SeqAdptBot, as well as the proper folding of the HP.

Figure 2 displays the sequence and structure of SA and HP before and after annealing/folding. The HP comprises four distinct features, (i) a 3' T overhang complementary to the A-overhang of the DNA molecules, (ii) unmethylated cytosine within the sequence, which permits the calculation of C-to-T conversion rate after sequencing, (iii) a unique molecular identifier (UMI), which allows to identify individual original ligation events and the remove of clonal PCR amplificates and (iv) a biotinylated T, for enrichment of HP containing DNA molecules after ligation. The ligation of SA and HP are catalysed by the T4 DNA ligase following the reaction outlined in Table 5.

The reaction is incubated at 16°C for 16 h and inactivated at 65°C for 10 min. The ligation of SA and HP is a stochastic event, meaning that three distinct types of DNA molecules will be generated: (i) molecules with SA on both ends, (ii) molecules with HP on both ends and (iii) molecules with SA on one side and HP on the other side.

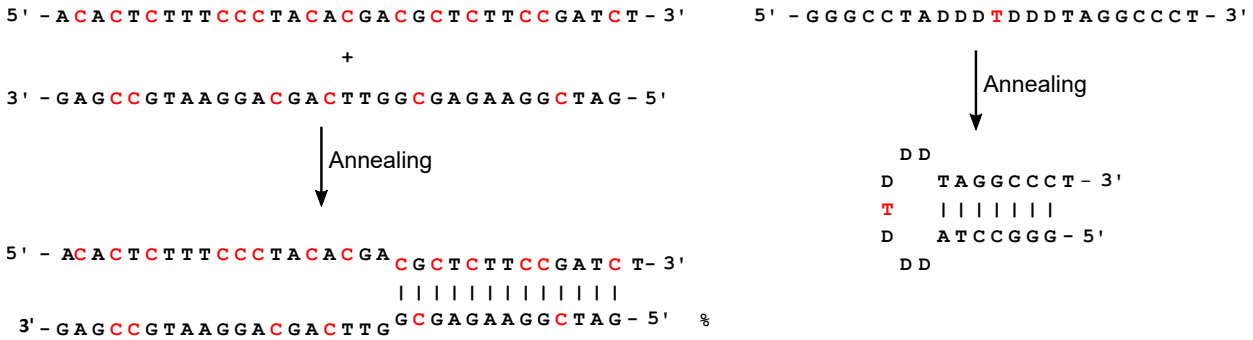


Figure 2: Annealing of Sequencing Adapter and Hairpin Linker - **Left:** Annealing of SeqAdptTop and SeqAdptBot, red Cs indicate methylated Cs; **Right:** Folding of the Hairpin linker, red T indicates biotinylated dT.

Table 5: Ligation reaction

Component	**Concentration**	**Volume [μl]**
A-tailed DNA	–	20.0
rCutSmart Buffer	10x	0.5
ATP	10 mM	2.5
Sequencing adapter (SA)	100 μM	0.5
Hairpin linker (HP)	100 μM	0.5
T4 DNA Ligase	2000 U/μl	1.0
Total Volume		25.0

This is a safe stopping point. After ligation the DNA can be stored for 24 h at 4°C or long term (days to weeks) at -20°C or -80°C. Please avoid repeated thawing and freezing, as this can lead to strand breaks and thus significantly reduces the number of usable DNA molecules.

NOTE: Including other modification such as 5mC, 5hmC, 5fC and 5caC, in the HP linker sequence, will allow to determine the conversion of non-canonical cytosine forms during BS and oxBS treatment.

Enrichment of Hairpin-DNA-adapter molecules

In order to deplete the unwanted non-HP molecules (SA on both sides) the library is subjected to a purification step using streptavidin coated beads. Start with an AMPure beads clean-up to remove proteins and excessive HPs that would likely saturate the streptavidin beads. Use a sample:bead ratio of 1:2 (25μl:50μl), wash twice with freshly prepared 80% EtOH and elute in 50μl ddH_2O.

1. Per library i.e., sample, transfer 10μl of Dynabeads^{TM} M-280 Streptavidin into a 1.5 ml reaction tube. Place the tube onto a magnetic stand and carefully remove the supernatant without disturbing the beads.

2. Add 1 ml 1xBW buffer (5 mM TrisHCl, 0.5 mM EDTA, 1 M NaCl), vortex thoroughly, place the mixture back onto the magnetic stand and wait until the solution is clear. Carefully remove and discard the supernatant.

3. Repeat step (2) for a total of 2 wash steps. At the end of wash step three, solve the beads in 50μl 2xBW buffer per library.

4. Transfer 50 μl of beads to each library, mix briefly by flicking the tube and incubate for 20 min at room-temperature while rotating.

5. Collect all liquid at the bottom of the tube though brief centrifugation and place the tube onto the magnetic stand. Wait until the solution is clear and carefully remove and discard the supernatant.

6. Add 200 μl 0.1N NaOH and mix by vortexing. Spin down all liquid, place the tube back onto the magnetic stand and wait until the solution is clear. Remove and discard the supernatant.

7. Repeat step (6) for a total of 2 wash steps.

8. Add 200 μl 0.5xTE buffer to the beads, mix by vortexing and place the tube back onto the magnetic stand. Wait until the solution is clear. Remove and discard the supernatant.

9. Repeat step (8) for a total of 2 wash steps. At the end of the second wash step, remove TE buffer completely, but do not let the beads dry fully.

10. Resuspend the beads in 50 μl 1xTE with 1% SDS.

11. Lock the tube and incubate the mixture for 5 min at 100°C in order to dissolve the the biotin-streptavidin-interaction. Place on ice for 2 min.

12. Briefly spin down and place the tube onto the magnetic stand. Wait until the solution is clear and transfer the supernatant containing the DNA into a new 1.5 ml DNA low-bind reaction tube.

The use of 1.5 ml DNA-low-bind reactions tubes might increase the final yield of the library. This is especially relevant when working with low input samples.

This is a safe stopping point. The DNA can be stored for 24 h at 4°C or long term (days to weeks) at -20°C or -80°C. Please avoid repeated thawing and freezing, as this can lead to strand breaks and thus significantly reduces the number of usable DNA molecules.

BS and oxBS treatment

For bisulfite and oxidative bisulfite treatment, we rely on the TrueMethyl®oxBS Module from TECAN. Here, we will provide a short summary of the individual steps of the protocol.

DNA Purification

Bring the Magnetic Bead Solution 1 (TrueMethyl Kit) and the Binding Buffer 1 (TrueMethyl Kit) to room temperature. Mix bead and buffer according to manufacturer instruction. To 50 μl DNA (in 1x TE and 1% SDS), add 100 μl bead-buffer solution and mix well by pipetting. Incubate at room temperature for 20 min. Transfer the solution onto the magnet and wait until the solution is clear (5 min). Carefully remove and discard the supernatant and wash 3x with 200 μl freshly prepared 80% acetonitrile. Let the beads dry for 5 min until all acetonitrile is evaporated. Remove the reaction tube from the magnet and solve the beads in 18 μl denaturation solution (TruMethyl Kit). Elute and denature the DNA at 37°C for 5 min. Put the tube back onto the magnet and wait until the solution is clear.

DNA oxidation

Transfer 9 μl into two 0.2 ml reaction tubes, respectively. To the tube intended for BS, add 1 μl ultrapure water (TrueMethyl Kit) in the other, meant for oxBS, add 1 μl oxidant solution. Mix both reactions by vortexing and immediately incubate at 40°C for 10 min. Centrifuge at 14000 x g for 10 min. A black precipitate will form at the bottom of the tube of the oxBS reactions. The supernatant should remain orange. Any other color (yellow, brown, transparent) indicates impurity of the sample and in our experiments indicates failing of the 5hmC oxidation. Transfer the supernatant from oxBS samples into a new 0.2 ml reaction tube.

Bisulfite treatment

Add 700 μl of the Bisulfite Dilutent (TrueMethyl Kit) to one aliquot of Bisulfite Reagent (TrueMethyl Kit). Incubate for 15 min at 60°C while shaking. Spin down briefly and add 30 μl of the bisulfite solution to BS and oxBS sample, respectively. Incubate the reaction according to the temperature profile outlined in table 6:

Prepare Magnetic Bead Solution 2 by mixing 2.4 μl Magnetic Beads Solution (TrueMethyl Kit) with 200 μl Binding Buffer 2 (TrueMEthyl Kit). Mix well by vortexing. Add 160 μl bead solution to 40 μl BS and 40 μl oxBS reaction, respectively. Mix well by pipetting and incubate for 5 min. Place the reaction onto an magnetic stand, wait until the solution is clear and discard the supernatant. Wash the beads with 200 μl freshly prepared 70% ethanol and resuspend the beads in 200 μl Desulfonation Buffer

Table 6: Bisulfite Conversion Temperature Profile

Step Number	Incubation Step	Temperature	Time
1	Denaturation	95°C	05:00 min
2	Sulfonation	60°C	20:00 min
3	Denaturation	95°C	05:00 min
4	Sulfonation	60°C	40:00 min
5	Denaturation	95°C	05:00 min
6	Sulfonation	95°C	45:00 min
7	Hold	20°C	≤ 16:00:00 h

(TrueMethyl Kit). Incubate for 5 min, remove the supernatant and wash the beads twice with 200 μl 70% ethanol. Let the beads dry completely for 15 min and elute the DNA in 12.5 μl Elution Buffer (TrueMethyl Kit).

Enrichment-PCR and sequencing

The enrichment PCR amplfies the library molecules and at the same time introduces the remaining part of the sequencing adapter i.e., indices (i5 and i7), as well as the sequences required for flow cell binding (P5 and P7). Table 7 and Table 8 provide the reaction conditions and the temperature profile of the enrichment PCR, respectively.

Table 7: Enrichment PCR

Component	Concentration	Volume [μl]
BS or oxBS DNA	–	10.0
HotStarTaq Buffer	10x	5.0
$MgCl_2$	25 mM	2.0
dNTPs	10 mM	4.0
EnPrimerFor	10 μM	0.8
EnPrimerRev	10 μM	0.8
HoStarTaq DNA Polymerase	5 U/μl	0.7
ddH_2O	–	26.7
Total Volume		50.0

Table 8: Enrichment PCR Temperature Profile

PCR Step	Temperature	Time	Number of Cycles
Initial Denaturation	95°C	5:00 min	1
Denaturation	94°C	0:30 min	
Annealing	58°C	1:00 min	7-18
Elongation	72°C	1:00 min	
Final Elongation	72°C	7:00 min	1
Hold	4°C	∞	1

After amplification, libraries are purified using AMPureXP®beads with a sample:bead ratio of 1:1 (50μl:50μl), wash twice with freshly prepared 80% EtOH and elute in 10μl ddH_2O or 0.1xTE. QC is performed by quantification of 2 μl of the library using Qubit HS Fluorometer and fragment size distribution is determined by loading 1 μl using the Agilent Bioanalyzer HS assay. In our case, the average library size lies by about 400bp. The library size might vary depending on the used restriction enzymes.

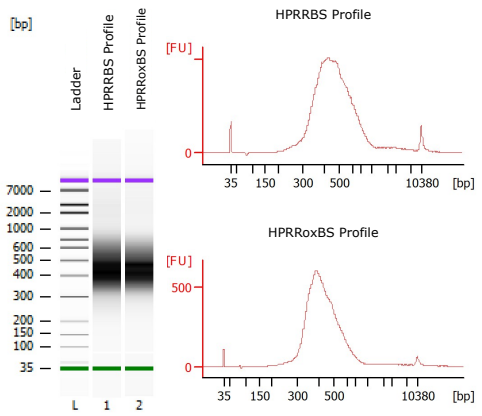


Figure 3: Bio-Analyzer Profiles - Typical Bio Analyzer profile of HPRRBS an HPRRoxBS libraries after enrichment PCR.

This is a safe stopping point. The DNA can be stored for 24 h at 4°C or long term (days to weeks) at -20°C or -80°C. Please avoid repeated thawing and freezing, as this can lead to strand breaks and thus significantly reduces the number of usable DNA molecules.

Computational Analysis

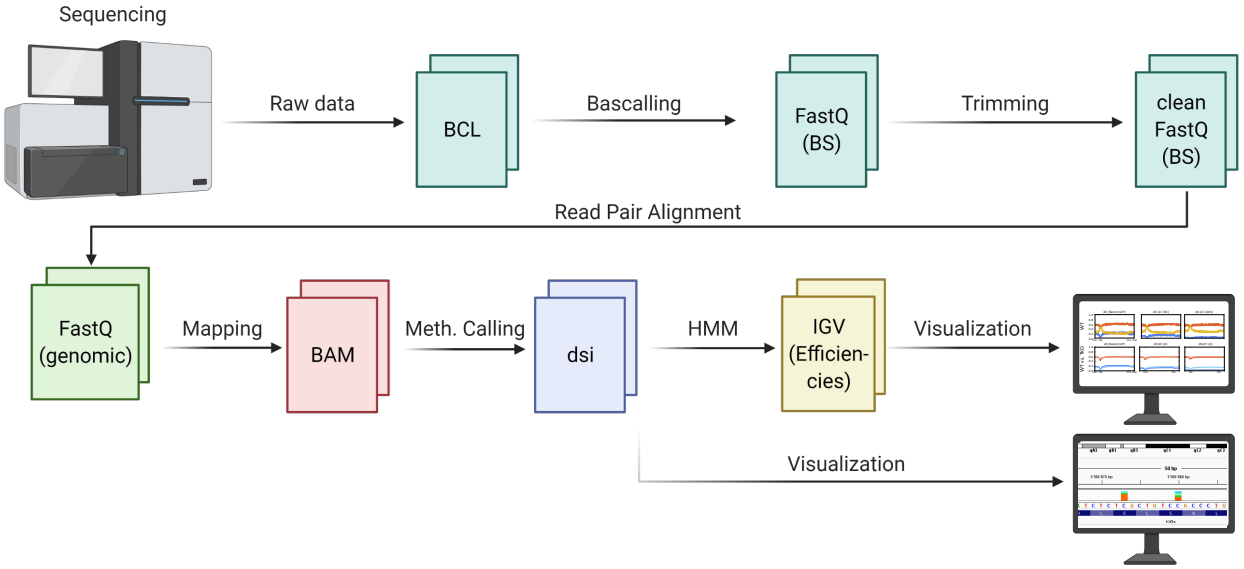


Figure 4: Data Processing and Analysis Pipeline - Illumina raw data is processed into base calls (FASTQ) and trimmed for adapter and hairpin linker sequences. Bisulfite reads from the same molecules are paired to restore the genomic sequence for efficient mapping. Subsequently, the doublestrand information is annotated and stored in - Created with BioRender.com

HPup - Processing Sequencing Data

In a first instance, the sequenced reads are subjected to a hairpin sequencing pipeline (HPup) that will restore the double strand information and output the methylation counts (uncorrected) for CpGs, GpCs and CpHs dyads. For this, the sequencing data was processed similar to what has been previously described by Porter *et al.* [46]. We used the Ruffus pipeline framework [47] for implementation of our workflow. In an initial step, sequencing adapter and low quality base calls (Q<20) were trimmed

from the ends of each read (FastQ files) using TrimGalore! [48]. Next, all reads were screened and purged from the hairpin linker sequence using cutadapt [49]. The hairpin linker sequence is stored in an additional file and used to calculate the C-to-T conversion rate during (oxidative) bisulfite treatment and furthermore, can be used to identify redundant reads generated by PCR.

Trimmed read pairs (read 1 and read 2 from paired-end-sequencing) were locally aligned with the Smith-Waterman algorithm. We allow for G-to-A and C-to-T mismatches to cope with bisulfite treatment. After aligning the bisulfite sequence of the read pairs, we can reconstitute the genomic sequence, which allows for a faster, more efficient and precise mapping of the reads to the reference genome. In addition, the pipeline determines the methylation state of each cytosine and classifies symmetric dyads, i.e., CpGs and GpCs into fully-methylated (both DNA strands are methylated), hemi-methylated (only one strand is methylated) at the plus strand, hemi-methylated on the minus strand and unmethylated (both strands are unmethylated). The resulting sequences were aligned to the mouse genome (mm10) with GEM-mapper (beta build 1.376) [50](54), after which the methylation information was reintroduced with a custom pileup function based on HTSJDK [51] and ratios for the four methylation states were calculated for each cytosine.

The hairpin pipeline comes with a configuration file in which all required parameters can be defined by the user. This includes the path to the input- and output-directory, as well as the adapter and hairpin linker sequences for trimming. The library generates three output files per sample: (i) a log file, containing the information about the individual processing steps (parameter, success and duration of individual steps). (ii) a statistic file which provides the read count after each processing step (raw reads, trimmed, paired, aligned), as well as the conversion rates of C, 5mC, 5hmC and 5fC, calculated based on hairpin linker and spike-in sequence. Lastly, (iii) a dsi-file (double strand information), which stores the methylation information of each cytosine is generated.

The DSI file is a header-less, tab-separated text-file and contains in total 11 columns. Columns 1 to 7 contain general information: (1) the chromosome number, (2) the genomic location of the analyzed cytosine, (3) the strand of the analyzed cytosine, (4) total number of reads, (5) number of methylated reads, (6) number of unmethylated reads and (7) sequence context of the cytosine (CG, GC or non-CpG). Columns 8 to 11 contain double strand specific methylation information: (8) counts of fully methylated dyads (CpG or GpC; NA = non-CpG), (9) counts of plus-strand methylated dyads, (10) counts of minus strand methylated dyads and (11) count of unmethylated dyads. DSI files can be visualised in the Integrative Genomics Viewer (IGV). Each sample will be displayed as bar diagram-track, in which each cytosine is represented by a stacked bar summarizing the frequency of the methylation states (fully = orange, hemi-plus = dark-green, hemi-minus = light-green and unmethylated = blue). Hovering the mouse cursor above the bar will show the entire 11-column information of the DSI file for the given position.

GwEEP-tool - Hidden Markov modelling of single CpG methylation

We provide the DSI files containing the BS and oxBS counts for each sequenced CpG as input to an HMM analysis (GwEEP-tool). We also incorporate the (oxidative-) bisulfite conversion rates (Figure 5) in our model to account for possible conversion errorsConversion rates of BS and oxBS are derived from a set of short (60bp) synthetic DNA sequences containing C, 5mC, 5hmC and 5fC, respectively (SI Sec. 6). The computational core of our model is the HMM previously described in [52, 53].We use Bayesian inference to estimate the distribution of methylation stages given the BS and oxBS counts. We also determine posterior distributions of the corresponding methylation efficiencies. In the HMM, we combine cell division i.e. DNA replication, methlyation (*de novo* and maintenance) as well as hydroxylmethylation efficiency. Figure 6 illustrates the possible transitions with the efficiencies that are to be estimated. The advantage of our Bayesian inference approach, compared to alternative approaches such as Maximum Likelihood Estimation, is that it yields useful estimates also in the case of relatively small counts. We apply the Metropolis-Hastings algorithm to a prior distribution and sample from the posterior distribution of the Dnmts, Tets efficiencies and the methylation levels of the CpGs at three available time points. Intuitively, the estimated efficiencies reflect the values that best explain the transition from the initial to the final DNA methylation state of the ES cells by considering all time points in between. Finally, our model also provides an estimate for maintenance of 5hmC during replication i.e. the maintenance probability p, which determines the likelihood that a given 5hmC is recognized by Dnmt1 as "methylated" after replication. Our model is now available as an efficient and user-friendly python script: `https://github.com/PaulaKramer/GwEEP-tool`.

RESULTS

We applied GwEEP on a well established ES cell system to precisely map 5mC and 5hmC across the genome in a time series experiment to study the enzymatic contribution of Tets and Dnmts for the progressive genome wide DNA (de)methylation. For this we generated a high resolution data set based on the above described genome wide hairpin sequencing approach: Reduced Representation Hairpin oxidative Bisulfite Sequencing (RRHPoxBS). The design of the RRHPoxBS (Combination of R.AluI, R.HaeIII and R.HpyCH4V) approach covered up to 4 million CpG dyads for which we could infer the precise distributions of 5mC and 5hmC. Following a strict read and conversion quality control, we filtered for sufficient sequencing depth and ended up with about 2 million CpGs per sample for subsequent comparative modeling. To follow the dynamics of the enzymes over time we generated six data sets for WT ES cells, i.e., BS and oxBS libraries for three different time points, starting with Serum/Lif (d0), followed by 72h 2i (d3) and 144h 2i (d6). For a comparison we also generated four datasets for Tet TKO cells starting with Serum/Lif (d0) followed by 48h in 2i (d2), 96h in 2i (d4) and 168h in 2i (d7).

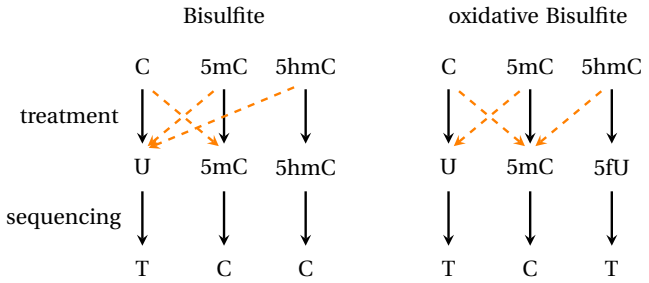
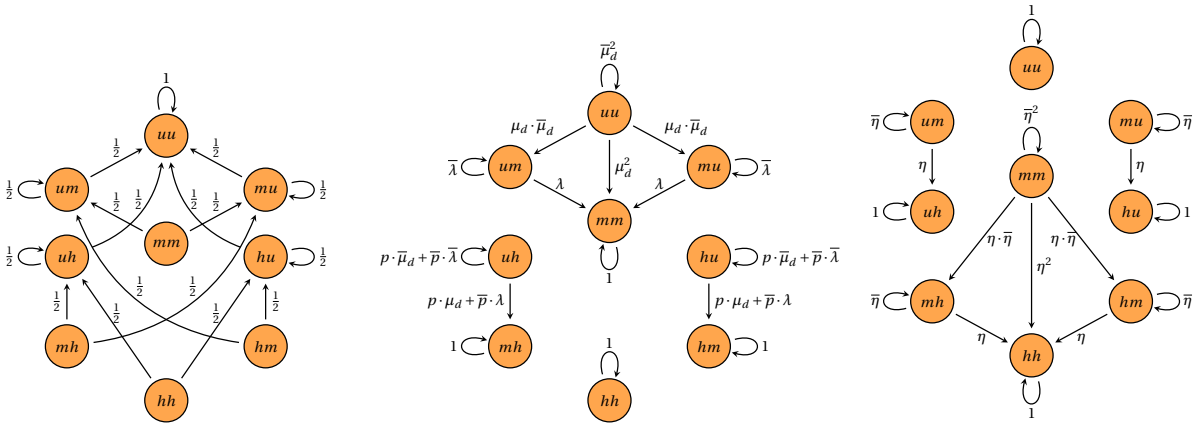


Figure 5: Bisulfite conversions (left) and oxidative bisulfite conversions (right). Possible conversion errors (orange arrows) can occur during the bisulfite or oxidative bisulfite sequencing approach and are considered in the modeling approach.



(a) Cell division (b) Methylation (c) Hydroxylation

Figure 6: Transitions between methylation states: u indicates unmethlyated, m represents methylated and h are hydroxylated stages. μ_d describes the efficiency of *de novo* methylation, μ_m the efficiency of maintenance methylation, η hydroxylation efficiency. λ, the overall methylation efficiency (*de novo* + maintenance), is defined as: $\lambda = \mu_m + \mu_d - \mu_m \cdot \mu_d$. The parameter p describes the probability that maintenance methylation does not consider hemi-hydroxylated sites.

Impaired loss of 5mC in Tet TKO ES cells

Under primed conditions (Serum/LIF) WT ES cells show a level of 78% methylation. About 56% of CpGs are fully methylated while 22% (Fig.: 7.A) are found in a hemimethylated state. Among cultivation in 2i medium the DNA becomes progressively demethylated, such that after six days in 2i only 30% of CpGs retain a methylated state (fully or hemimethylated). For all time points, we find that hemimethylation is equally distributed among both DNA strands (Fig.: 7.A and SI Sec. 4.3 Figure 23). Overall, our results agree nicely with previously published whole genome methylation profiles [38]. Note, that the data from Ficz *et. al* originate from the very same ES cell sample.

In addition, we observe that oxBS samples always display lower methylation levels than BS samples (Fig.: 7.A). This difference corresponds to the amount of 5hmC of each sample. We detect the main difference in the hemimethylated proportion, indicating that a considerable amount of 5hmC might exist in a hemi(hydroxy)- methylated (5hmC/C or C/5hmC) state (Fig.: 7.A). Initially, the amount of 5hmC is quite low but we observe a notable increase at d3, while later (d6) the amount of 5hmC decreases again (Fig.: 7.A).

ES cells lacking Tet enzymes (Tet TKO) show a marginal increase of methylated CpG dyads (82% fully or hemimethylated, Fig.: 7.B) in comparison to WT under primed serum conditions. However, in relation to WT the TKO cells show a higher frequency of fully methylated CpGs (72%) and a reduced proportion of hemimethylated CpGs (hemiCpGs; 10%) (Fig.: 7.B). We concluded that in WT ES cells, the enhanced presence of hemiCpGs is directly coupled to 5mC oxidation by Tets.

Moreover, in contrast to previous data sets, [38, 39, 54] the RRHoxBS data allow us distinguish hemi- from unmethylated CpGs and therefore to precisely estimate the demethylation kinetics, revealing that in WT ES cells the generation of unmethylated cytosine is 8% per day, while in Tet TKO cells it drops to 4.2% (SI Sec. 4.4 Figure 24). This indicates that the presence of Tets has a considerable influence on DNA demethylation kinetics.

RRHPoxBS sequencing also allowed us to accurately determine the amount, location and distribution of non-CpG methylation in (WT and Tet TKO) ES cells. In both, WT and TKO, we find CpA to be the most frequent methylated non-CpG motif SI

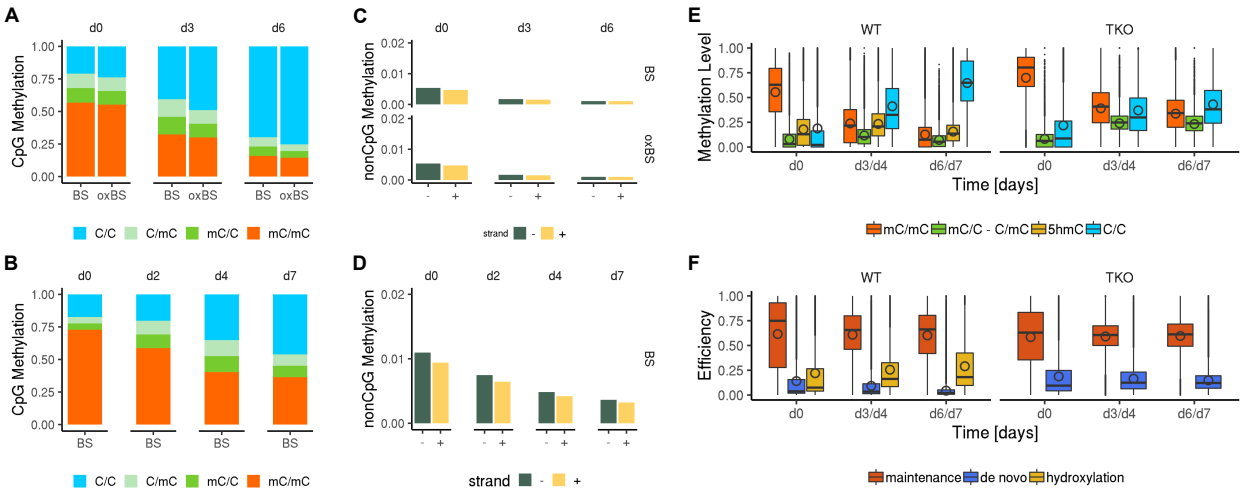


Figure 7: Hairpin Sequencing Data and HMM Results - **A** Average CpG methylation level based on uncorrected hairpin sequencing counts for WT ES cells; **B** Average CpG methylation level based on uncorrected hairpin sequencing counts for Tet TKO ES cells; **C** Average nonCpG methylation level based on uncorrected hairpin sequencing counts for WT ES cells; **D** Average CpG methylation level based on uncorrected hairpin sequencing counts for Tet TKO ES cells; **E** calculated and conversion error corrected 5mC and 5hmC distribution after HM modelling; **F** HMM derived maintenance methylation, *de novo* methylation and hydroxylation efficiencies.

Sec. 4.7 Figure 32 and 33). Over time, nonCpG methylation becomes gradually reduced upon cultivation in 2i. In WT ES cells the number of methylated nonCpGs was identical in BS and oxBS libraries, indicating that nonCpGs are not a substrate for Tet oxidation (Fig.: 7.C). In Tet TKO cells, the number of methylated nonCpGs is approximately doubled as compared to WT ES cells. Since nonCpG methylation is strictly dependent on the presence of *de novo* methylation activities by Dnmt3a/b [18], the higher nonCpG methylation in TKO cells, both under primed (=2%) and naive (=0.6% after 168h 2i) conditions (Fig.: 7.D), points clearly towards an increased *de novo* methylation activity by Dnmt3a/b in the absence of Tet enzymes.

5mC oxidation affects maintenance and *de novo* methylation

The estimated methylation levels for WT and Tet TKO ES cells fit nicely to the hairpin methylation data (SI Sec. 3, Fig.: 3), indicating a high accuracy of our model. Consequently, we observe a constant decline of fully methylated CpGs in WT ES cells over time (Fig.: 7.E).

Moreover, in WT ES cells, the HMM estimates a notable amount of 5hmC at all time points. Note that the displayed amount of 5hmC (yellow) refers to the sum of all possible 5hmC states, i.e., 5hmC/5hmC, 5hmC/5mC, 5mC/5hmC, 5hmC/C, C/5hmC. The highest amount of 5hmC is observed at d3, meaning that WT ES cells display a transient increase of 5hmC after cultivation in 2i. We observe a similar behaviour for hemiCpGs in WT ES cells. The parameter estimation by our model illustrates a mean maintenance methylation efficiency of about 61.4% at d0, which remains almost constant over time (60.1% at d6) (7.F). In contrast, *de novo* methylation efficiency shows a strong decrease (from 14.1% to 4.5% at d6) and the hydroxylation efficiency an increase (from 22.2% at d0 to 29.1% at d6) over time. This observation is in agreement with previous observations which demonstrated a reduction in RNA and protein levels of Dnmt3a/b in 2i, but an increased expression of Tet1/2 on a genome wide level [38, 54].

In Tet TKO cells maintenance efficiency lies by 58.8% at d0, which represents a marginal reduction compared to WT ES cells. Similarly to WT, maintenance efficiency remains stable over time (58.6% at d7) also in Tet TKO cells.

The most pronounced difference between WT and Tet TKO cells we see in *de novo* methylation efficiency. More specifically in Tet TKO *de novo* begins from (20.3%) at day 0 and exhibits only a slight decrease over time (16.2% at d7).

Finally, the model output also confirms the reduced demethylation rate in Tet TKO cells, previously observed in the hairpin sequencing data and suggests a substantial contribution of 5hmC and Tet enzyme on DNA demethylation. In fact, the model favors a scenario in which 5hmC is less well recognized (probability of non recognition equals p=0.66, SI Sec. 1.1) by the maintenance machinery after replication, promoting a faster demethylation process.

Tets prevent the spreading of DNA methylation

We next related the model estimates to genomic, enzymatic and epigenetic features first focusing on Dnmt and Tet enzyme efficiencies across large genome segments with distinct methylation states. We used MethylSeekR [55], to partition the genome

into four states: highly methylated regions (HMRs), partially methylated domains (PMDs), low methylated regions (LMRs) and unmethylated regions (UMRs). The segmentation was performed on whole genome bisulfite sequencing (WGBS) data from WT ES cells cultivated under Serum/Lif conditions (Ficz et al.) [38] on the identical cell batch used for our study. The estimated methylation levels (sum of 5mC and 5hmC) for WT ES cells by our model fully agreed with those derived from WGBS (SI Sec. 4.8, Fig.: 35.C and 35.E). This not only confirmed the accuracy of our model output but also denoted that we can use the precise WGBS segmentation for further analysis. We found that the majority of the WT ES cell genome (86%) consists of large HMRs (SI Sec. 4.8, Fig.: 35.A and Fig.: 35.B) followed by shorter (13%) PMDs. The residual 2% of the genome are found to be LMRs (0.4%) or UMRs (1.5%)(SI Sec. 4.8, Fig.: 35.A and Fig.: 35.B).

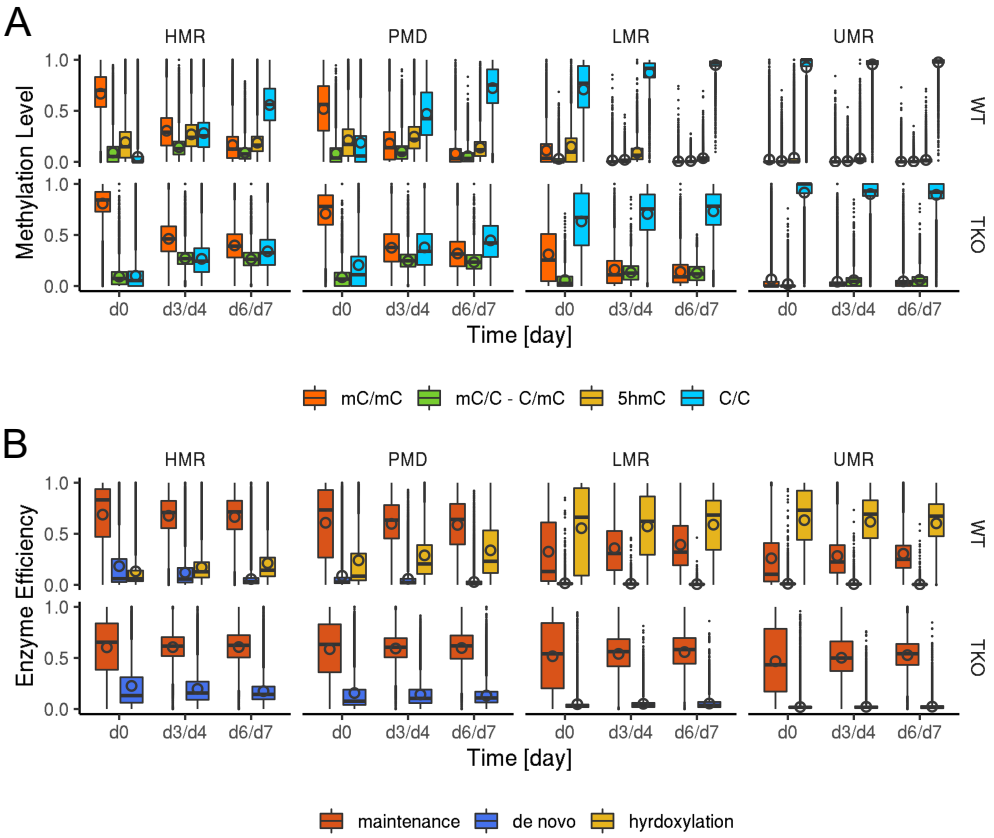


Figure 8: Methylation and Efficiency Profiles of Methylome Segments - (A) Estimated methylation distibution of HMRs, PMDs, LMRs and UMRs; orange = fully methylated CpGs, green = hemimethylated CpGs, yellow = CpG containing 5hmC and blue = umethylated CpGs. (B) Estimated enzyme efficiencies in HMRs, PMDs, LMRs and UMRs; red = maintenance methylation efficiency by Dnmts, blue = *de novo* methylation efficiency by Dnmts, yellow = hydroxylation efficiency by Tets

We assigned the 5mC and 5hmC modifications levels, their distribution at CpG dyads and the corresponding Dnmt/Tet enzyme efficiencies determined by our model to CpGs in the individual segments (Fig.: 8.A and 8.B). In WT ES cells, all segments show a continuous reduction in DNA methylation over time. This is particularly evident in segments with initial high 5mC levels, i.e., HMRs and PMDs. HMRs and PMDs also exhibit the highest amount of 5hmC and hemiCpGs, which transiently increases at d3 and d6 (Fig.: 8.A). LMRs and UMRs show different kinetics as both the amount of 5hmC and hemiCpGs constantly decline over time. The increase of hemiCpGs in HMRs and PMDs is a clear sign of impaired maintenance methylation in naive ES cells linked to the reported temporal increase in Tet expression and loss of Dnmt1 activity [38, 54].

Based on the methylation data, our model predicts high maintenance methylation efficiency in HMRs (69%) and PMDs (61%), but low maintenance efficiency in LMRs (32%) and UMRs (26%)(Fig.: 8.B). Additionally, we observe a relatively high *de novo* methylation efficiency at HMRs (18%) in primed ES cells (Fig.: 8.B). Overall, *de novo* methylation efficiency strongly decreases upon cultivation in 2i, which corresponds well with the previous described loss of Dnmt3a/b under these conditions. In contrast, hydroxylation efficiency is high in UMRs (63%) and LMRs (55%), but low in HMRs (13%) and PMDs (24%)(Fig.: 8.B). Together, our results indicate regional differences and an antagonistic behaviour of Dnmts and Tets. This antagonism has been validated by estimating a robust spatial correlation measure between the efficiencies of Dnmts and Tets across the whole genome (SI Sec. 3.3, Fig.: 12).

Both WT and TKO ES cells show overall a decline of DNA methylation across all segments over time (Fig.: 8.A). However,

in TKO cells, all segments retain a substantial higher frequency of fully methylated CpG dyads across all time points. This observation indicates a reduced demethylation rate in all segments under the absence of Tets. Surprisingly, under primed conditions (d0) Tet TKO cells show a higher number of unmethylated CpGs in HMRs as compared to WT ES cells.

Most importantly, comparing the Tet TKO with WT ES cells, we observe a strong change in Dnmt efficiencies. Maintenance methylation efficiency shows a reduction in HMRs and PMDs of TKO cells, while it clearly increases in LMRs and UMRs (Fig.: 8.B), resulting in almost equal maintenance activity across all segments. In the case of *de novo* methylation efficiency, we observe a more stable and slightly increased activity in all segments.

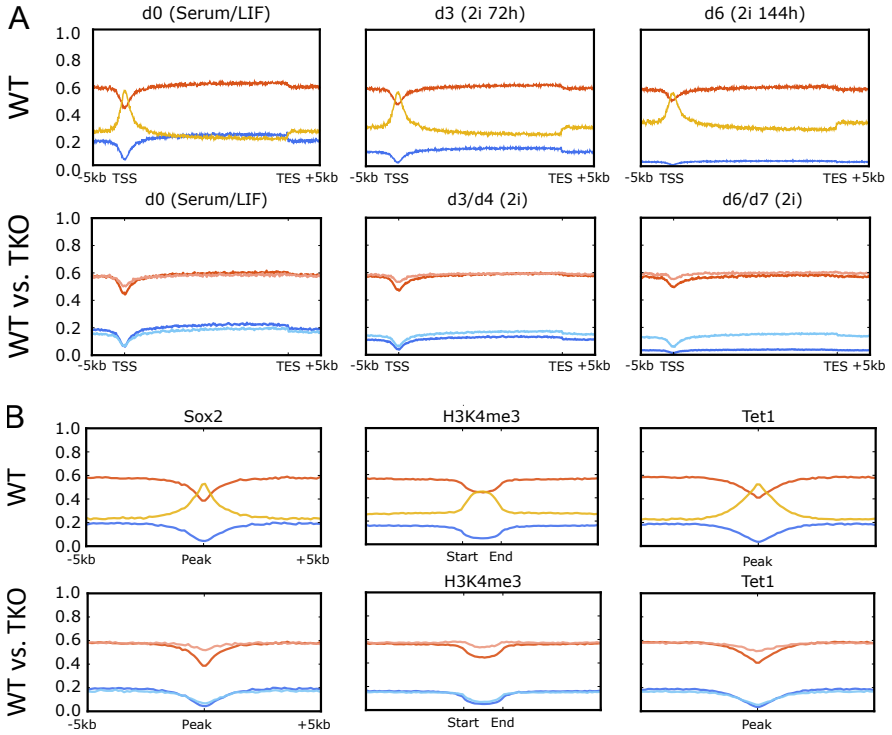


Figure 9: Average efficiency profiles across genes and protein binding sites. (A) Average maintenance, *de novo* and hydroxylation efficiency of WT and Tet TKO cells across genes; (B) Average maintenance, de novo, as well as hydroxylation efficiency across selected chromatin marks and protein binding sites. red = maintenance methylation efficiency, blue = *de novo* methylation efficiency, yellow = hydroxylation efficiency. Dark colors indicate the efficiencies in WT ES cells, while light colors refer to the efficiencies of Tet TKO ES cells.

Tets regulate Dnmts in TSS and TFBS

The genome wide antagonistic effects of Dnmts' and Tets' activity across segments and clusters prompted us to plot the enzymatic efficiencies of CpGs across genes, histone marks or ChIP profiled TFBS using DeepTools [56] (Fig.: 9) in order to investigate regularities and general local dependencies. In WT cells the enzymes' efficiencies across genes and TFBS show once more an opposing behavior: At transcription start sites (TSS) and TFBS, high hydroxylation efficiency is coupled to reduced methylation (both maintenance and *de novo*) efficiency. This inverse behavior at TSS remains upon 2i cultivation. *De novo* methylation almost disappears across the entire gene including the gene body. Under primed conditions *de novo* methylation is absent in TSS but has a strong presence in the gene body and it almost disappears from the entire gene over time after the transition to 2i. The observed efficiency profiles for maintenance methylation, de novo methylation and hydroxylation, correspond nicely to Uhrf1, Dnmt3a/b, as well as Tet1 ChIP profiles, respectively (SI Sec. 4.6 Fig. 30 and 31).

In Tet TKO ES cells the TSS associated drop in maintenance methylation is much less pronounced and almost absent at d6/d7. In addition, *de novo* methylation is only mildly reduced upon cultivation in 2i and clearly maintained across the gene body (Fig.: 8.A). Regulatory regions marked by Sox2, H3K4me3 and Tet1 enrichment show a strong hydroxylation activity in WT cells which is inversely linked to an impaired maintenance and de novo methylation activity. Interestingly, the lack of Tet activity in TKO cells does not change de novo methylation but maintenance activity across regulatory regions (Fig.: 8.B).

DISCUSSION

In this study, we provide a comprehensive approach for measuring genome wide DNA-methylation and efficiency profiling. GwEEP allowed us to infer how the activity of Dnmts and Tets contribute to modify CpGs and nonCpGs across the genome in a functional context. It is important to note that technically our RRHPoxBS data resemble and corroborate the overall methylation dynamics observed by classical RRBS and WGBS ([38, 54]). However, RRHPoxBS data provide three important novel features: (i) a genome-wide representation of up to 4 million CpGs uniformly distributed across the genome, (ii) a precise determination of 5mC and 5hmC levels at a single CpG dyad and (iii) a precise mapping of hemimethylated states and positions of nonCpG methylation. The combination with our HMM allows us to calculate accurate 5mC and 5hmC level by considering the conversion errors during BS and oxBS. Furthermore, compared to previous models, we simultaneously infer the genome wide efficiencies of maintenance methylation, *de novo* methylation and hydroxylation efficiencies.

The overall evaluation of our RRHPoxBS data showed that in mouse ES cells (as previously described for somatic cells [18]), hemiCpGs are almost equally distributed on both DNA strands, following the behavior of symmetric CpG methylation. This suggests that hemimethylation is most likely the result of (strand-) undirected *de novo* methylation or active and passive demethylation events, respectively. Furthermore, we detect more hemimethylation in WT compared to Tet TKO cells, which indicates that Tet enzymes enhance the passive loss of 5mC. Indeed, our model predicts that 5hmC is probably less well recognized by Dnmt1 after replication, such that hydroxylation enhances passive demethylation, which is in agreement with recent in vitro studies [30, 31]. In contrast to equally distributed hemimethylation we observe a slight increase in the minus strand presence of nonCpG methylation. We cannot find a simple biological (sequence context) or technical (calling/mapping) explanation for this bias. Non-CpG methylation is always occurring in close vicinity to CpG methylation (18) but in contrast to CpGs we find that nonCpGs are not a substrate for Tet enzymes, i.e., we do not find any indication of 5hmC in the non-CpG context. The amount of nonCpG methylation however is strongly enhanced in the absence of Tet enzymes, suggesting an increase of Dnmt3a and 3b efficiency in the absence of Tets. Our model provides strong evidence that Dnmts and Tets do not act independently at a given CpG, but clearly in an opposed manner. Generally, we observe a high maintenance and de novo efficiency at the majority of the genome, i.e., HMRs and PMDs (or inter-/intragenic regions), while the activity of Tet enzymes is highest at UMRs and LMRs, such as promoters, TFBS (Sox2, Pou5f1) and TSS. Recent studies based on chromatin immunoprecipitation support our findings, revealing binding of Dnmt3a/b at the gene body and HMRs, whereas Tet1 binding was observed across methylation valleys (LMRs and UMRs) [41, 57].

The impairment of maintenance methylation has been identified so far as the main driver of 2i induced DNA demethylation [54] and a role for Tet or oxidative cytosine forms, on the other hand, has only been recognized for selected loci [38, 54]. The comparison of WT and Tet TKO ES cells in the present study, however, discloses a notable reduction within the demethylation rate of Tet TKO, compared to WT ES cells. On average, we detect a reduction in the demethylation rate of almost 50% from around 8% to 4% loss per day. This indicates that Tets and their oxidized cytosine products are essential for an effective demethylation during the Serum-to-2i shift and probably other biological demethylation processes with similar enzymatic compositions.

The loss of Tet enzymes is naturally expected to result in an impaired removal of 5mC and it does at least for CpGs located in LMRs and UMRs, where we observe a notable increase in their methylation level. Nevertheless, under primed conditions and within HMRs, we paradoxically observe more unmethylated CpGs (hypomethylation) in Tet TKO ES cells compared to WT ES cells. Recently, L´opez-Moyado *et al.* conducted a systematic investigation of genome wide methylation profiles from various cell types carrying distinct Tet KO genotypes [42]. Similar to our observations they detected a pronounced loss of DNA methylation in heterochromatic compartments (i.e., HMRs and PMDs) of Tet TKO mouse ES cells.

L´opez-Moyado *et al.* propose a mutual exclusive localization of Dnmts and Tets in WT ES cells, while in Tet KO cells Dnmts invade domains which were previously occupied by Tets. Indeed, in the absence of Tets, our model predicts a clear misregulation in both maintenance and de novo methylation efficiency. In Tet TKO ES cells, we see an increase in maintenance methylation efficiency, but at the same time a reduction in HMRs and PMDs. Moreover, we observe an increase in de novo methylation efficiency at PMDs. Together, this indicates a displacement of Dnmt1, as well as Dnmt3a/b, which fits to the hypothesized model by L´opez-Moyado *et al.*. In addition, Tet TKO cells exhibit a more stable, almost persistent de novo methylation under naive conditions. The increased non-CpG methylation of Tet TKO cells detected by RRHPoxBS further supports this finding. This shows that in the absence of Tets, ES cells also fail to effectively down-regulate de novo methylation efficiency in 2i.

Taken together, we summarize that Tet enzymes work against methylation in three ways. (i) They guarantee an efficient conversion of 5mC at accessible regions and act against its establishment during a cell replication either via passive or active demethylation, (ii) They inhibit the effectiveness of the maintenance machinery over regions that should remain unmethylated. (iii) Finally, they ensure an efficient down-regulation of the de novo enzymes, which can not be observed in their absence.

CONCLUSION

We describe a novel combination of experimental and computational approach - GwEEP to investigate the contributions of Tets and Dnmts to the establishment of distinctive DNA methylation patterns across the genome. In GwEEP, we generate strand specific (hydroxy)methylation data and using a sophisticated HMM we infer the distriution of 5mC and 5hmC at individual CpGs across the genome and furthermore, derive accurate efficiency profiles of Dnmts (*de novo* and maintenance) and Tets (hydroxylation). GwEEP also works with low amounts of DNA and is there fore suitable for demanding samples and rare cell types. Moreover, by combining our hairpin protocol with 5fC or 5caC detecting chemistry, GwEEP is easily expandable for the

estimation of 5fC/5caC distribution and the inference of formylation and carboxylation efficiencies of Tets. Our analysis of WT and Tet TKO mES cells shows that Dnmts and Tets exhibit clear antagonistic efficiencies at individual CpGs. The comparison of WT and Tet TKO ES cells demonstrates that Tet enzymes contribute notably to the loss of DNA methylation in the present model system. Moreover, Tet enzymes seem to protect unmethylated regions against both *de novo* and maintenance methylation efficiency and to restrict the activity of Dnmts in highly methylated regions, guaranteeing the formation and maintenance of cell type specific methylation patterns.

DATA AVAILABILITY

All raw and processed hairpin sequencing data sets, as well as the HMM output are deposited at NCBI's Gene Expression Omnibus (GEO) repository under accession number GSE169070.

ACKNOWLEDGEMENTS

We thank Jasmin Kirch and Gilles Gasparoni for performing sequencing using the Illumina HiSeq2500 system and Andreas Firczynski for the optimization of the processing pipeline for the hairpin sequencing data. Particular thanks to Jim Robinson and the entire IGV team for implementing the dsi-file format into the IGV genome browser.

FUNDING

The project was funded by the DFG during the course of the SFB1309–325871075 and the German Epigenome Programme (DEEP) of the Federal Ministry of Education and Research in Germany (BMBF) [01KU1216 to J.W.]. AS was supported by the German Federal Ministry of Research and Education grant for de.NBI (031L0101D).

AUTHOR'S CONTRIBUTIONS

Conceived and designed the experiments: CK PG WR VW JW. Wrote the manuscript: CK PG JW VW. Performed the experiments: PG JA FvM GF. Processed the raw data: KN. Designed/implemented the computational methods: CK. Analyzed the data: CK. Performed meta analysis: CK PG KN AS FM.

SOFTWARE AVAILABILITY

All Matlab scripts written for implementing the single CpG stochastic model, run it in a parallel fashion on a multi-core environment and perform the subsequent computational analysis presented in the manuscript and the appendix are shared via GitHub (https://github.com/kyriakopou/hydroxyGit/tree/master/genomeWide)

CONFLICT of INTEREST STATEMENT

W.R. is a consultant and shareholder of Cambridge Epigenetix. All other authors declare no competing interests.

References

[1] R. Holliday, J. E. Pugh, Dna modification mechanisms and gene activity during development, Science 187 (4173) (1975) 226–232.

[2] A. D. Riggs, X inactivation, differentiation, and dna methylation, Cytogenetic and Genome Research 14 (1) (1975) 9–25.

[3] D. Bourc'his, T. H. Bestor, Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking dnmt3l, Nature 431 (7004) (2004) 96–99.

[4] E. Li, T. H. Bestor, R. Jaenisch, Targeted mutation of the dna methyltransferase gene results in embryonic lethality, Cell 69 (6) (1992) 915–926.

[5] B. H. Ramsahoye, D. Biniszkiewicz, F. Lyko, V. Clark, A. P. Bird, R. Jaenisch, Non-cpg methylation is prevalent in embryonic stem cells and may be mediated by dna methyltransferase 3a, Proceedings of the National Academy of Sciences 97 (10) (2000) 5237–5242.

[6] M. J. Ziller, F. Müller, J. Liao, Y. Zhang, H. Gu, C. Bock, P. Boyle, C. B. Epstein, B. E. Bernstein, T. Lengauer, et al., Genomic distribution and inter-sample variation of non-cpg methylation across human cell types, PLoS genetics 7 (12) (2011) e1002389.

[7] R. Lister, M. Pelizzola, R. Dowen, R. Hawkins, G. Hon, J. Tonti-Filippini, J. Nery, L. Lee, Z. Ye, Q. Ngo, et al., ntosiewicz-bourget, J., Stewart, R., Ruotti, V., Millar, AH, Thomson, JA, Ren, B., and Ecker, JR (2009) 315–322.

[8] H. Leonhardt, A. W. Page, H.-U. Weier, T. H. Bestor, A targeting sequence directs dna methyltransferase to sites of dna replication in mammalian nuclei, Cell 71 (5) (1992) 865–873.

[9] L. S.-H. Chuang, H.-I. Ian, T.-W. Koh, H.-H. Ng, G. Xu, B. F. Li, Human dna-(cytosine-5) methyltransferase-pcna complex as a target for p21waf1, Science 277 (5334) (1997) 1996–2000.

[10] M. Bostick, J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan, S. E. Jacobsen, Uhrf1 plays a role in maintaining dna methylation in mammalian cells, Science 317 (5845) (2007) 1760–1764.

[11] J. Sharif, M. Muto, S.-i. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, et al., The sra protein np95 mediates epigenetic inheritance by recruiting dnmt1 to methylated dna, Nature 450 (7171) (2007) 908–912.

[12] K. Arita, M. Ariyoshi, H. Tochio, Y. Nakamura, M. Shirakawa, Recognition of hemi-methylated dna by the sra protein uhrf1 by a base-flipping mechanism, Nature 455 (7214) (2008) 818–821.

[13] A. Hermann, R. Goyal, A. Jeltsch, The dnmt1 dna-(cytosine-c5)-methyltransferase methylates dna processively with high preference for hemimethylated target sites, Journal of Biological Chemistry 279 (46) (2004) 48350–48359.

[14] M. Okano, S. Xie, E. Li, Cloning and characterization of a family of novel mammalian dna (cytosine-5) methyltransferases, Nature genetics 19 (3) (1998) 219–220.

[15] M. Okano, D. W. Bell, D. A. Haber, E. Li, Dna methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development, Cell 99 (3) (1999) 247–257.

[16] D. Meilinger, K. Fellinger, S. Bultmann, U. Rothbauer, I. M. Bonapace, W. E. Klinkert, F. Spada, H. Leonhardt, Np95 interacts with de novo dna methyltransferases, dnmt3a and dnmt3b, and mediates epigenetic silencing of the viral cmv promoter in embryonic stem cells, EMBO reports 10 (11) (2009) 1259–1264.

[17] G. Liang, M. F. Chan, Y. Tomigahara, Y. C. Tsai, F. A. Gonzales, E. Li, P. W. Laird, P. A. Jones, Cooperativity between dna methyltransferases in the maintenance methylation of repetitive elements, Molecular and cellular biology 22 (2) (2002) 480–491.

[18] J. Arand, D. Spieler, T. Karius, M. R. Branco, D. Meilinger, A. Meissner, T. Jenuwein, G. Xu, H. Leonhardt, V. Wolf, et al., In vivo control of cpg and non-cpg dna methylation by dna methyltransferases, PLoS genetics 8 (6) (2012) e1002750.

[19] R. Ono, T. Taki, T. Taketani, M. Taniwaki, H. Kobayashi, Y. Hayashi, Lcx, leukemia-associated protein with a cxxc domain, is fused to mll in acute myeloid leukemia with trilineage dysplasia having t (10; 11)(q22; q23), Cancer research 62 (14) (2002) 4075–4080.

[20] R. Lorsbach, J. Moore, S. Mathew, S. Raimondi, S. Mukatira, J. Downing, Tet1, a member of a novel protein family, is fused to mll in acute myeloid leukemia containing the t (10; 11)(q22; q23), Leukemia 17 (3) (2003) 637–641.

[21] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids, Cell cycle 8 (11) (2009) 1698–1710.

[22] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian dna by mll partner tet1, Science 324 (5929) (2009) 930–935.

[23] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science 333 (6047) (2011) 1300–1303.

[24] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates, PloS one 5 (12) (2010) e15367.

[25] S. Kriaucionis, N. Heintz, The nuclear dna base 5-hydroxymethylcytosine is present in purkinje neurons and the brain, Science 324 (5929) (2009) 929–930.

[26] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic dna, Nucleic acids research 38 (19) (2010) e181–e181.

[27] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, 5-hydroxymethylcytosine is a predominantly stable dna modification, Nature chemistry 6 (12) (2014) 1049–1055.

[28] E.-A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, 5-formylcytosine alters the structure of the dna double helix, Nature structural & molecular biology 22 (1) (2015) 44–49.

[29] M. W. Kellinger, C.-X. Song, J. Chong, X.-Y. Lu, C. He, D. Wang, 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of rna polymerase ii transcription, Nature structural & molecular biology 19 (8) (2012) 831–833.

[30] H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, X. Cheng, Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, Nucleic acids research 40 (11) (2012) 4841–4849.

[31] V. Valinluck, L. C. Sowers, Endogenous cytosine damage products alter the site selectivity of human dna maintenance methyltransferase dnmt1, Cancer research 67 (3) (2007) 946–950.

[32] D. Ji, C. You, P. Wang, Y. Wang, Effects of tet-induced oxidation products of 5-methylcytosine on dna replication in mammalian cells, Chemical research in toxicology 27 (7) (2014) 1304–1309.

[33] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, et al., Tet-mediated formation of 5-carboxylcytosine and its excision by tdg in mammalian dna, Science 333 (6047) (2011) 1303–1307.

[34] A. Maiti, A. C. Drohat, Thymine dna glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of cpg sites, Journal of Biological Chemistry 286 (41) (2011) 35334–35338.

[35] Z. D. Smith, M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, A. Meissner, A unique regulatory phase of dna methylation in the early mammalian embryo, Nature 484 (7394) (2012) 339–344.

[36] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, Active demethylation of the paternal genome in the mouse zygote, Current Biology 10 (8) (2000) 475–478.

[37] P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway, Science 329 (5987) (2010) 78–82. doi:10.1126/science.1187945. URL https://www.science.org/doi/abs/10.1126/science.1187945

[38] G. Ficz, T. A. Hore, F. Santos, H. J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley, Y.-L. Paul, J. Walter, et al., Fgf signaling inhibition in escs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency, Cell stem cell 13 (3) (2013) 351–359.

[39] E. Habibi, A. B. Brinkman, J. Arand, L. I. Kroeze, H. H. Kerstens, F. Matarese, K. Lepikhov, M. Gut, I. Brun-Heath, N. C. Hubner, et al., Whole-genome bisulfite sequencing of two distinct interconvertible dna methylomes of mouse embryonic stem cells, Cell stem cell 13 (3) (2013) 360–369.

[40] M. Walter, A. Teissandier, R. Pérez-Palacios, D. Bourc'his, An epigenetic switch ensures transposon repression upon dynamic loss of dna methylation in embryonic stem cells, Elife 5.

[41] T. Gu, X. Lin, S. M. Cullen, M. Luo, M. Jeong, M. Estecio, J. Shen, S. Hardikar, D. Sun, J. Su, et al., Dnmt3a and tet1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells, Genome biology 19 (1) (2018) 1–15.

[42] I. F. López-Moyado, A. Tsagaratou, H. Yuita, H. Seo, B. Delatte, S. Heinz, C. Benner, A. Rao, Paradoxical association of tet loss of function with genome-wide dna hypomethylation, Proceedings of the National Academy of Sciences 116 (34) (2019) 16933–16942.

[43] P. A. Ginno, D. Gaidatzis, A. Feldmann, L. Hoerner, D. Imanci, L. Burger, F. Zilbermann, A. H. Peters, F. Edenhofer, S. A. Smallwood, et al., A genome-scale map of dna methylation turnover identifies site-specific dependencies of dnmt and tet activity, Nature communications 11 (1) (2020) 1–16.

[44] J. Charlton, E. J. Jung, A. L. Mattei, N. Bailly, J. Liao, E. J. Martin, P. Giesselmann, B. Brändl, E. K. Stamenova, F.-J. Müller, et al., Tets compete with dnmt3 activity in pluripotent cells at thousands of methylated somatic enhancers, Nature genetics 52 (8) (2020) 819–827.

[45] D. E. Martin-Herranz, A. J. Ribeiro, F. Krueger, J. M. Thornton, W. Reik, T. M. Stubbs, currbs: simple and robust evaluation of enzyme combinations for reduced representation approaches, Nucleic acids research 45 (20) (2017) 11559–11569.

[46] J. Porter, M.-a. Sun, H. Xie, L. Zhang, Investigating bisulfite short-read mapping failure with hairpin bisulfite sequencing data, BMC genomics 16 (11) (2015) 1–9.

[47] L. Goodstadt, Ruffus: a lightweight python library for computational pipelines, Bioinformatics 26 (21) (2010) 2778–2779.

[48] F. Krueger, Babraham bioinformatics: Trim galore!(2012) (2019).

[49] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet. journal 17 (1) (2011) 10–12.

[50] S. Marco-Sola, M. Sammeth, R. Guigó, P. Ribeca, The gem mapper: fast, accurate and versatile alignment by filtration, Nature methods 9 (12) (2012) 1185–1188.

[51] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup. 2009. the sequence alignment/map format and samtools, Bioinformatics 25 (16) (2009) 2078–2079.

[52] P. Giehr, C. Kyriakopoulos, G. Ficz, V. Wolf, J. Walter, The influence of hydroxylation on maintaining cpg methylation patterns: a hidden markov model approach, PLoS computational biology 12 (5) (2016) e1004905.

[53] C. Kyriakopoulos, P. Giehr, V. Wolf, H(O)TA: estimation of DNA methylation and hydroxylation levels and efficiencies from time course data, Bioinformatics 33 (11) (2017) 1733–1734. arXiv:https://academic.oup.com/bioinformatics/article-pdf/33/11/1733/25153659/btx042.pdf, doi:10.1093/bioinformatics/btx042. URL https://doi.org/10.1093/bioinformatics/btx042

[54] F. von Meyenn, M. Iurlaro, E. Habibi, N. Q. Liu, A. Salehzadeh-Yazdi, F. Santos, E. Petrini, I. Milagre, M. Yu, Z. Xie, et al., Impairment of dna methylation maintenance is the main cause of global demethylation in naive embryonic stem cells, Molecular cell 62 (6) (2016) 848–861.

[55] L. Burger, D. Gaidatzis, D. Schübeler, M. B. Stadler, Identification of active regulatory regions from dna methylation data, Nucleic acids research 41 (16) (2013) e155–e155.

[56] F. Ramırez, F. Dündar, S. Diehl, B. Grüning, T. Manke, deeptools: a flexible platform for exploring deep-sequencing data, Nucleic Acids Res 42 (W1) (2014) 187–91.

[57] T. Baubec, D. F. Colombo, C. Wirbelauer, J. Schmidt, L. Burger, A. R. Krebs, A. Akalin, D. Schübeler, Genomic profiling of dna methyltransferases reveals a role for dnmt3b in genic methylation, Nature 520 (7546) (2015) 243–247.