BiQ Analyzer HiMod: an interactive software tool for high-throughput locus-specific analysis of 5-methylcytosine and its oxidized derivatives

Journal Information
Journal ID (nlm-ta): Nucleic Acids Res
Journal ID (iso-abbrev): Nucleic Acids Res
Journal ID (hwp): nar
Journal ID (publisher-id): nar
Title: Nucleic Acids Research
ISSN (print): 0305-1048
ISSN (electronic): 1362-4962
Publisher: Oxford University Press

Article/Issue Information
Self URI: gku457.pdf
Date accepted: 08 May 2014
Date revision received: 2 May 2014
Date received: 15 March 2014
Publication date (print): 01 July 2014
Publication date (electronic): 29 May 2014
Publication date (pmc-release): 29 May 2014
Volume: 42
Issue: Web Server issue
Pages: W501-W507
Article Id (accession): PMC4086109
Article Id (pmcid): PMC4086109
Article Id (pmc-uid): 4086109
PubMed ID: 24875479
DOI: 10.1093/nar/gku457

Categories

Subject: Article

Counts

Pages: 7

Custom metadata

cover-date: 1 July 2014

BiQ Analyzer HiMod: an interactive software tool for high-throughput locus-specific analysis of 5-methylcytosine and its oxidized derivatives

Daniel Becker
Pavlo Lutsik
Peter Ebert
Christoph Bock
Thomas Lengauer
Jörn Walter

1 1Department of Genetics, Saarland University, Saarbrücken 66123, Germany

2 2Department of Computational Biology and Applied Algorithms, Max-Planck Institute for Informatics, Saarbrücken 66123, Germany

3 3CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna 1090, Austria

4 4Department of Laboratory Medicine, Medical University of Vienna, Vienna 1090, Austria

To whom correspondence should be addressed. Tel: +49 681 302 4367; Fax: +49 681 302 2703; Email: j.walter@mx.uni-saarland.de
Correspondence may also be addressed to Pavlo Lutsik. Tel: +49 681 302 2881; Fax: +49 681 302 2703; Email: p.lutsik@mx.uni-saarland.de

Abstract

Recent data suggest important biological roles for oxidative modifications of methylated cytosines, specifically hydroxymethylation, formylation and carboxylation. Several assays are now available for profiling these DNA modifications genome-wide as well as in targeted, locus-specific settings. Here we present BiQ Analyzer HiMod, a user-friendly software tool for sequence alignment, quality control and initial analysis of locus-specific DNA modification data. The software supports four different assay types, and it leads the user from raw sequence reads to DNA modification statistics and publication-quality plots. BiQ Analyzer HiMod combines well-established graphical user interface of its predecessor tool, BiQ Analyzer HT, with new and extended analysis modes. BiQ Analyzer HiMod also includes updates of the analysis workspace, an intuitive interface, a custom vector graphics engine and support of additional input and output data formats. The tool is freely available as a stand-alone installation package from http://biq-analyzer-himod.bioinf.mpi-inf.mpg.de/.

INTRODUCTION

DNA methylation is widely recognized as a key epigenetic mechanism playing a crucial role in development and disease (1,2). Sequencing of bisulfite-treated deoxyribonucleic acid (DNA) is the gold standard for base resolution mapping of DNA-methylation (3–5). Recently, oxidized derivatives of 5-methylcytosine (5mC) were identified that are assumed to have important biological function (6). Like 5mC, the oxidized derivatives occur predominantly at CpG-dinucleotide cytosines (7). However, conventional bisulfite

sequencing does not discriminate between different oxidized forms. Sodium bisulfite treatment converts unmodified cytosine, 5-carboxy (5caC) and 5-formylcytosyne (5fC) into deaminated forms, which are subsequently replaced by thymines during polymerase chainreaction (PCR) amplification, while methylated and hydroxymethylated (5hmC) cytosines remain unchanged (8). Several new methods have been suggested recently to overcome these limitations and transform the specific oxidative modifications into sequence-based signals. Oxidative bisulfite sequencing (oxBS-seq) involves an oxidation reaction to convert 5hmC to 5fC (9). A subsequent bisulfite reaction followed by the amplification step converts 5fC to thymine whereas the original 5mC is replaced by cytosines in PCR amplicons (Figure 1a). TET-assisted bisulfite sequencing (TAB-seq) utilizes the opposite scheme and applies an enzymatic oxidation by the TET protein to convert all 5mC to 5hmC and further to 5fC and 5caC, while the initially present 5hmC is protected by glycosylation prior to the bisulfite reaction (Supplementary Figure S1). The subsequent bisulfite treatment and PCR converts the initially present 5mC (as well as 5fC and 5caC) to thymine while the 5hmC is replaced with cytosines (10). Similar methods were developed for mapping 5fC and 5caC, such as formyl-chemically-assisted bisulfite sequencing (fCAB-seq) (11) (Figure 1b) and chemical-modification-assisted bisulfite sequencing (CAB-seq) (12) (Supplementary Figure S1). Following the modified bisulfite treatments of either oxBS-seq, TAB-seq, CAB-seq or fCAB-seq the DNA is amplified and sequenced. The abundance of any modification under inspection is estimated by comparing such sequencing results to data obtained by conventional bisulfite sequencing. Note that all modified bisulfite methods are only partially modification-specific and mostly produce cumulative estimates of several different modifications.

Figure 1.

Principal scheme of oxBS-seq and fCAB-seq methods. (**a**) oxBS-seq. Oxidative bisulfite treatment followed by PCR leads to conversion of the oxidative modifications to thymines while only 5mC appears as cytosine in the sequencing reads. The cumulative level of 5hmC can be established for each CpG site by comparing to ordinary bisulfite sequencing. (**b**) In fCAB-seq 5fC is protected from the conversion and together with 5mC and 5hmC appears as cytosine after PCR. The bulk 5fC abundance is calculated by subtracting the cumulative levels of 5mC + 5hmC pair obtained from ordinary bisulfite sequencing.

While several tools have been developed for the analysis of conventional locus-specific bisulfite sequencing data of various scale (13–18), no software package currently supports an integrated analysis of data from multiplexed locus-specific high-throughput DNA modification profiling. For the genome-scale data the MLML tool has been designed to produce accurate DNA methylation and hydroxymethylation levels from preprocessed oxBS-seq, TAB-seq and other similar data (19).

To fill this gap we developed BiQ Analyzer HiMod (or shortly BiQ HiMod), a specialized interactive software package for the preprocessing, quality control and analysis of various DNA modifications in high-throughput targeted sequencing experiments. BiQ HiMod continues our line of computational tools for locus-specific DNA methylation analysis (14,15,18) and is an extension of the previously published BiQ Analyzer HT (14). It provides a user-friendly interface, comprehensive and customizable streamlined analysis pipeline, rich graphical and tabular output. The tool currently supports four experimental assays and can be easily extended to include the upcoming methods.

BiQ Analyzer HiMod

Overview

BiQ Analyzer HiMod is a cross-platform interactive Java application, which can run on any system with properly installed and configured Java Runtime Environment. The package installation is performed via a simple ‘click-through’ installer. BiQ HiMod facilitates a fully interactive primary processing and analysis of sequencing data from locus-specific DNA modification profiling experiments, and currently supports oxBS-seq, TAB-seq, fCAB-seq and CAB-seq assays. The comprehensive and ergonomic graphical user interface provides means for multi-level overview of the complete multiplexed sequencing project, in which a number of genomic loci of interest are sequenced in several biological samples. The underlying software architecture uses the fact that most of the basic data processing steps are invariant to the type of chemical treatment and thus the reads from ordinary and modified bisulfite procedure can be preprocessed independently. During the quality control cycle, the user refines the quality thresholds of the processing pipeline until a satisfactory quality level is reached for each read batch. After low-level processing, cumulative levels of each DNA modification are summarized based on the results of the individual preprocessing pipeline runs. Final results can be exported as tables, figures, alignments and genome browser tracks (see Supplementary text for a detailed list). Importantly, BiQ HiMod maintains the full functionality of the previous versions by supporting the analysis of conventional bisulfite sequencing data in an independent analysis mode.

Data preparation and project setup

BiQ HiMod is designed to import multiple data sets of individual sequencing reactions and loci. Today's data sets are often generated by NGS-based multiplexed locus-specific sequencing (16,23,24). In NGS approaches sequences are usually indexed by short sequence tags (using multiplexing library kits), pooled and submitted to a high-throughput sequencing platform, e.g. Illumina MiSeq or Ion Torrent. In such a setting BiQ HiMod requires the direct sequencing output to be demultiplexed using available third-party tools. We recommend the barcode splitter from Galaxy (20–22) as an adequate solution which suits researchers with minimal bioinformatics experience. Reads ready to be loaded must be in FASTA or FASTQ format with one file per each sample-locus-treatment combination.

Besides demultiplexing, data preparation usually includes several additional standard steps, e.g. trimming of low quality parts of the sequence reads (often at the 3#-end) and—in case of paired-end sequencing—joining of overlapping mate reads. In the FAQ section of the BiQ HiMod web site (http://biq-analyzer-himod.bioinf.mpi-inf.mpg.de/FAQ.php) we describe an example of such a data preparation workflow in a step-by-step way using third-party tools integrated into the Galaxy framework (20–22).

Following the data preparation one needs to set up an analysis project and import the data. For a relatively small project this can be done directly using the GUI controls. For larger data sets one should prepare a spreadsheet-like text file, (see our example on the program web site). Samples are defined by text labels, while the genomic loci exhibit one-to-one mapping to their genomic reference sequences, supplied as single- or multi-sequence FASTA files. We recommend extracting sequence information via the Galaxy ‘Extract Sequence’ tool starting from a BED file with genomic coordinates of each amplicon, such that the generated FASTA records contain genomic location information. This information will be used further on, in order to bring the analysis into the genomic context and generate genome browser tracks.

Finally, sequencing reads can be imported directly from the genome-wide sequencing experiment (e.g. whole-genome or reduced-representation bisulfite sequencing), if the data are available as aligned sequencing reads (BAM files). BiQ HiMod facilitates loading the

reads of a corresponding region based on coordinate information and analyzing them in the context of the corresponding genomic reference. This feature of BiQ HiMod can be useful for quickly scanning locus-specific read distributions in regions of interest based on large genome-wide WGBS or RRBS datasets provided they are accessible as BAM files.

Primary processing pipeline

The software backend was subjected to several modifications compared to the predecessor version of the tool. The FASTQ files parsing is now supported using the data import library of the BioJava project (23). Loaded sequence reads are independently aligned to the corresponding reference sequence using the bisulfite-modified semi-global implementation of Needleman–Wunsch algorithm, which was further tweaked to improve robustness of the modification site alignment (see Supplementary Text and (14) for more details). The set of filtering criteria was extended to include sequencing quality scores, and default thresholds are set to ensure that only high quality bases are considered. The full list of quality metrics and DNA modification statistics are given in the Supplementary Text. All processing and analysis options remain fully customizable and can be refined after the initial analysis.

Quantifying the modification levels

A quantitative assessment of the relative abundances of different modifications depends on the comparison of two independent sequencing results. This comparative mode is a core feature implemented in BiQ Analyzer HiMod. However, such comparisons have two main limitations: (i) except for one modification type one can obtain estimates of modification differences at single CpG positions only as an average across all sequences and not per sequence (ii) in most comparisons combinations of several modification types are not separable from each other. For illustration, we consider the results of the oxBS-seq method. Here, the reads from the standard bisulfite treatment contain unconverted cytosines which were methylated or hydroxymethylated in the original DNA molecule. The read cytosines from the oxidative bisulfite treatment can only be mapped to the 5mC of the original DNA. All in all, oxBS-seq yields a single-molecule resolution map of 5mC, bulk per-CpG average levels of 5hmC and a combination of 5fC, 5caC and unmodified cytosines. Table 1 summarizes the information which can be extracted from each of the data types. For every type of analysis BiQ HiMod takes into account all potential modifications and summarizes the information accordingly.

Table 1.

Information that can be extracted from each DNA modification profiling method supported by BiQ Analyzer HiMod

SM—single-molecule resolution, B—bulk per-CpG measurement.

The estimation of bulk modification levels in BiQ Analyzer HiMod is performed by a simple subtraction of converted cytosine frequencies. In case of low abundance of the target modification, in particular 5fC and 5caC, the true biological changes of modification levels may overlap with the experimental and the technical error range, sometimes leading to negative DNA modification levels. To overcome such difficulties, on the experimental side we recommend gathering data from several independent experiments using very high sequencing coverage. Moreover, the statistical significance of the observed variation should be tested. Summarized data can be exported in bedGraph format and loaded into downstream statistical software, e.g. MLML or RnBeads (http://rnbeads.mpi-inf.mpg.de), for correction and proper statistical evaluation.

Visualization and data export

The frontend of BiQ HiMod provides capabilities for the top-down level-by-level exploration of a sequencing experiment. On the top level the global overview of the analysis project is given by the samples versus loci heat maps visualizing average modification abundances for each sample-locus pair (Figure 2a). More detailed profiles of each studied locus are provided by the locus-wide bar plots visualizing the modification levels at each CpG position with realistic relative genomic distances (Figure 2b). On the bottom level, sequence pileup, read-level cytosine pattern map and data tables are available for the most in-depth exploration and quality control similar to the analysis mode of BiQ Analyzer HT. The newly introduced diagnostic plots simplify the selection of the quality control thresholds by visualizing read batch-wide distributions of major analysis metrics, e.g. score and sequence identity of the read-reference alignment, bisulfite conversion rate, number of missing or mutated modification sites etc. (Figure 2c).

Figure 2.

BiQ Analyzer HiMod visualization features. (**a**). Global heat map of average bulk (5mC + 5hmC) modification levels at sequenced loci across the samples. The default color code is given in the color legend (see Supplementary Figure S2 for more details). (**b**) Bar chart displaying stacked levels of two modifications at each CpG position in a single locus (see also Supplementary Figure S3). (**c**) Diagnostic histogram visualizing the distribution of a quality control metric—sequence identity of each read-reference alignment—simplifying the sequence identity cutoff selection. (**d**) Integration with the IGV genome browser showing genomic tracks with 5mC and 5hmC levels of three samples (the value range is much smaller for the 5hmC tracks).

Read-level and summarized modification data can be exported as tab-delimited files for downstream statistical analysis and visualization in favorite statistical software packages, e.g. R/Bioconductor, SPSS and Excel, or custom bioinformatic pipelines. Furthermore, BiQ HiMod feeds into large-scale experimental projects by exporting its results as genome browser tracks. Methylation values for each analyzed CpG-site can be saved in bedGraph format importable into genome viewers and browsers, such as IGV (24) or UCSC Genome Browser (25) (http://genome.ucsc.edu/index.html) (see example in Figure 2d). This allows for a convenient display of amplicon validation data in the context of genome-wide epigenetic data tracks, such as DNA methylation, histone modification, chromatin accessibility maps etc. Finally, the summarized data in bedGraph format can be loaded into our in-house genome-wide DNA methylation data analysis package RnBeads (http://rnbeads.mpi-inf.mpg.de/) for statistically rigorous evaluation and comparison to corresponding genome-scale profiles in case such are available.

Software architecture, GUI improvements and the new graphics engine

BiQ HiMod builds upon the modular and robust design of its predecessor tool, BiQ Analyzer HT, with a multistep data processing pipeline, as its backend, combined with an interactive graphical user interface at the frontend. The backend takes raw sequencing reads and reference sequence information as input, performs a series of preprocessing steps—data loading, alignment, quality filtering, cytosine conversion calling—and generates the resulting tables and graphics for each sample-locus combination. The backend output is written directly to the hard drive and can be found in the project directory. This results in efficient use of the operating memory and guards against information loss. The interactive frontend, implemented in Java standard AWT/Swing framework, enables the setup of a large-scale multiplexed next-generation sequencing (NGS) analysis project for one of the

supported profiling methods, coordination of the pipeline runs, and summarization of the results at the sample, locus and project levels as well as export of the results.

BiQ HiMod expands the basic form of the GUI design of the predecessor tool, BiQ Analyzer HT, by introducing several major improvements. First, the project summary view has been turned into an ultimate control panel, affording an exhaustive overview without the need of frequent panel switching. It includes a project summary table, presenting the most crucial information about each read batch, the project-level heat maps visualizing average modification levels and locus-wide bar charts dynamically generated for each selected read batch. Second, the detailed views of individual read batches were updated and now include the summary pages presenting the levels of each modification or group of modifications as tables and plots.

BiQ HiMod features a completely new custom graphical engine generating high-quality vector images in scalable vector graphics (SVG) format. Most of the plotting parameters, including the flexible color scheme, are fully user-adjustable. The exported SVG files can be further edited via one of the popular image packages without any loss of quality or directly incorporated into manuscripts and reports.

VALIDATION ON ARTIFICIAL AND REAL BIOLOGICAL DATA AND PERFORMANCE ASSESSMENT

In order to test BiQ HiMod and assess its performance we generated artificial read batches for each supported type of experiment by simulation (see Supplementary Text 1 for details). We estimated the accuracy and robustness of BiQ HiMod results under different conditions.

We then applied BiQ HiMod to reprocess the earlier published oxBS-seq data set obtained by sequencing of repetitive elements in two types of mouse ES cells (26). In brief, serum cultured murine E14 ES cells and 2i-medium cultured ES cells from three different time points—1, 3 and 7 days of cultivation—were sampled, two replicates of each. After oxidative and conventional bisulfite treatment, hairpin amplicons of regions within two repeat types—intracisternal A particle, also known as IAP, and long interspersed elements, also known as LINE-1 or L1—were constructed as described in an earlier study (27). The amplicons were sequenced on the MiSeq platform from Illumina using paired end sequencing (2 × 150). The BiQ HiMod results document an accumulation of 5hmC at LINE1 but not at IAP elements during the first 72 h of cultivation in 2i-medium (Supplementary Figure S7).

We also benchmarked the runtime of BiQ Analyzer HiMod by analyzing batches of 1000 to 50 000 artificially generated reads. The tests were carried out on a machine with a 2.4 GHz Intel Core i5 Haswell dual core processor and 2 GB of RAM. The results of the benchmark, given in Table 2, show that moderate-sized analyzes can be performed in a reasonable amount of time on an average commodity laptop. For data from large-scale experiments amounting to several millions or more sequence reads, we recommend running BiQ HiMod on more powerful workstation computers with several multicore CPUs and a few dozen GB of RAM, where the program can benefit of its parallel computation capabilities.

Table 2.

Benchmarking of BiQ Analyzer HiMod on artificially generated read batches of variable size

Number. of reads in each batch[a]	1000	2000	5000	10 000	20 000	50 000
Running time, min.sec Normal	0.15	0.32	1.52	5.20	16.48	155.15
w/o pattern maps[b]	0.15	0.26	1.06	2.13	4.25	25.14

[a]Given are numbers of reads in each of two read batches (oxBS and conventional bisulfite) processed in every benchmarking step. Therefore, the actual number of processed reads was twice the given number in each step.

[b]The high computation time for large read sets is based on the size of the produced single read-resolution pattern maps. The user can decide which graphics to generate and by choosing to disable pattern maps the computation time can be significantly reduced.

CONCLUSIONS AND OUTLOOK

BiQ Analyzer HiMod is the first interactive software package for preprocessing, quality control and initial analysis of various DNA modifications from four different experimental methods. Developed on the basis of BiQ Analyzer and BiQ Analyzer HT (14,15), it features a thoroughly reworked GUI with a large number of analysis plots as well as a new customizable vector graphics engine, additional data import and export formats. The tool enables a comprehensive analysis workflow starting from raw sequence reads up to publication-quality plots and genome browser tracks. BiQ HiMod exploits modular and flexible software design allowing for painless further extensions to support upcoming experimental assays, such as MAB-seq (28). The standardized, comprehensive and user-friendly software paradigm implemented by BiQ HiMod will help the tool find numerous applications in basic biology and biomedical research.

AVAILABILITY

The BiQ Analyzer HiMod installation package is freely available from http://biq-analyzer-himod.bioinf.mpi-inf.mpg.de/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online including [1,2].

Supplementary Data

Acknowledgements

The authors would like to thank Julia Arand and Pascal Giehr for providing the oxBS-seq data for the biological example, Georg Friedrich for technical support with the BiQ HiMod web site and Karl Nordström for valuable advice on preprocessing of sequencing data.

FUNDING

European Union's Seventh Framework Programme (FP7/2007-2013) [267038 (NOTOX) to J.W. and P.L., HEALTH-F5-2011-282510 (BLUEPRINT)]; German Science Ministry [01KU1216A (DEEP Project) to P.E.]. Funding for open access charge: European Union's Seventh Framework Programme (FP7/2007-2013) [267038 (NOTOX), HEALTH-F5-2011-282510 (BLUEPRINT)].

Conflict of interest statement. None declared.

REFERENCES

1. Smith Z.D. Meissner A. DNA methylation: roles in mammalian development Nat. Rev. Genet. 2013 14 204 220 23400093

2. Bergman Y. Cedar H. DNA methylation dynamics in health and disease (vol 20, pg 274, 2013) Nat. Struct. Mol. Biol. 2013 20 1236 1236

3. Harris R.A. Wang T. Coarfa C. Nagarajan R.P. Hong C. Downey S.L. Johnson B.E. Fouse S.D. Delaney A. Zhao Y. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications Nat. Biotechnol. 2010 28 U1097–U1194

4. Bock C. Tomazou E.M. Brinkman A.B. Mueller F. Simmer F. Gu H. Jaeger N. Gnirke A. Stunnenberg H.G. Meissner A. Quantitative comparison of genome-wide DNA methylation mapping technologies Nat. Biotechnol. 2010 28 U1106–U1196

5. Bock C. Analysing and interpreting DNA methylation data Nat. Rev. Genet. 2012 13 705 719 22986265

6. Branco M.R. Ficz G. Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome Nat. Rev. Genet. 2012 13 7 13 22083101

7. Kohli R.M. Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation Nature 2013 502 472 479 24153300

8. Huang Y. Pastor W.A. Shen Y. Tahiliani M. Liu D.R. Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing PLoS One 2010 5 e8888 20126651

9. Booth M.J. Branco M.R. Ficz G. Oxley D. Krueger F. Reik W. Balasubramanian S. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution Science 2012 336 934 937 22539555

10. Schuler P. Miller A.K. Sequencing the sixth base (5-hydroxymethylcytosine): selective DNA oxidation enables base-pair resolution Angew. Chem., Int. Ed. Engl. 2012 51 10704 10707 23011896

11. Song C.X. Szulwach K.E. Dai Q. Fu Y. Mao S.Q. Lin L. Street C. Li Y. Poidevin M. Wu H. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming Cell 2013 153 678 691 23602153

12. Lu X. Song C.X. Szulwach K. Wang Z. Weidenbacher P. Jin P. He C. Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA J. Am. Chem. Soc. 2013 135 9315 9317 23758547

13. Kumaki Y. Oda M. Okano M. QUMA: quantification tool for methylation analysis Nucleic Acids Res. 2008 36 W170 W175 18487274

14. Lutsik P. Feuerbach L. Arand J. Lengauer T. Walter J. Bock C. BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing Nucleic Acids Res. 2011 39 W551 W556 21565797

15. Bock C. Reither S. Mikeska T. Paulsen M. Walter J. Lengauer T. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing Bioinformatics 2005 21 4067 4068 16141249

16. Rohde C. Zhang Y. Reinhardt R. Jeltsch A. BISMA–fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences BMC Bioinformatics 2010 11 e230

17. Hetzl J. Foerster A.M. Raidl G. Mittelsten Scheid O. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing Plant J. 2007 51 526 536 17559516

18. Schuffler P. Mikeska T. Waha A. Lengauer T. Bock C. MethMarker: user-friendly design and optimization of gene-specific DNA methylation assays Genome Biol 2009 10 eR105

19. Qu J. Zhou M. Song Q. Hong E.E. Smith A.D. MLML: consistent simultaneous estimates of DNA methylation and hydroxymethylation Bioinformatics 2013 29 2645 2646 23969133

20. Blankenberg D. Von Kuster G. Coraor N. Ananda G. Lazarus R. Mangan M. Nekrutenko A. Taylor J. Galaxy: a web-based genome analysis tool for experimentalists Curr. Protoc. Mol. Biol. 2010 Chapter 19 11 21 Unit 19 10

21. Giardine B. Riemer C. Hardison R.C. Burhans R. Elnitski L. Shah P. Zhang Y. Blankenberg D. Albert I. Taylor J. Galaxy: a platform for interactive large-scale genome analysis Genome Res. 2005 15 1451 1455 16169926

22. Goecks J. Nekrutenko A. Taylor J. Galaxy T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences Genome Biol. 2010 11 eR86

23. Holland R.C. Down T.A. Pocock M. Prlic A. Huen D. James K. Foisy S. Drager A. Yates A. Heuer M. BioJava: an open-source framework for bioinformatics Bioinformatics 2008 24 2096 2097 18689808

24. Robinson J.T. Thorvaldsdottir H. Winckler W. Guttman M. Lander E.S. Getz G. Mesirov J.P. Integrative genomics viewer Nat. Biotechnol. 2011 29 24 26 21221095

25. Kent W.J. Sugnet C.W. Furey T.S. Roskin K.M. Pringle T.H. Zahler A.M. Haussler D. The human genome browser at UCSC Genome Res. 2002 12 996 1006 12045153

26. Ficz G. Hore T.A. Santos F. Lee H.J. Dean W. Arand J. Krueger F. Oxley D. Paul Y.-L. Walter J. FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency Cell Stem Cell 2013 13 351 359 23850245

27. Arand J. Spieler D. Karius T. Branco M.R. Meilinger D. Meissner A. Jenuwein T. Xu G. Leonhardt H. Wolf V. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases PLoS Genet. 2012 8 e1002750 22761581

28. Hu X. Zhang L. Mao S.Q. Li Z. Chen J. Zhang R.R. Wu H.P. Gao J. Guo F. Liu W. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming Cell Stem Cell 2014 14 512 522 24529596