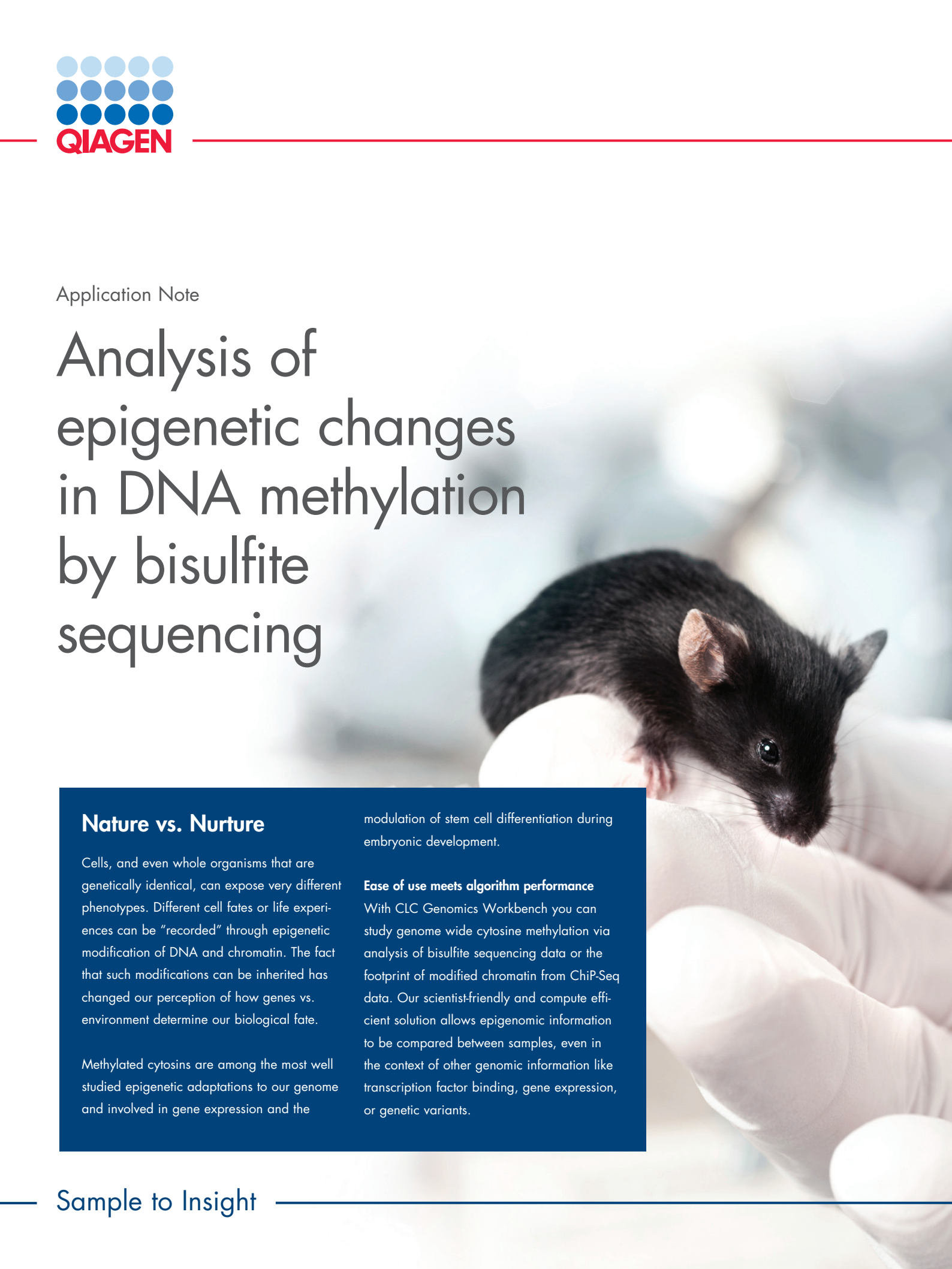
Application Note

Analysis of epigenetic changes in DNA methylation by bisulfite sequencing



Nature vs. Nurture

Cells, and even whole organisms that are genetically identical, can expose very different phenotypes. Different cell fates or life experiences can be "recorded" through epigenetic modification of DNA and chromatin. The fact that such modifications can be inherited has changed our perception of how genes vs. environment determine our biological fate.

Methylated cytosins are among the most well studied epigenetic adaptations to our genome and involved in gene expression and the modulation of stem cell differentiation during embryonic development.

Ease of use meets algorithm performance
With CLC Genomics Workbench you can study genome wide cytosine methylation via analysis of bisulfite sequencing data or the footprint of modified chromatin from ChiP-Seq data. Our scientist-friendly and compute efficient solution allows epigenomic information to be compared between samples, even in the context of other genomic information like transcription factor binding, gene expression, or genetic variants.

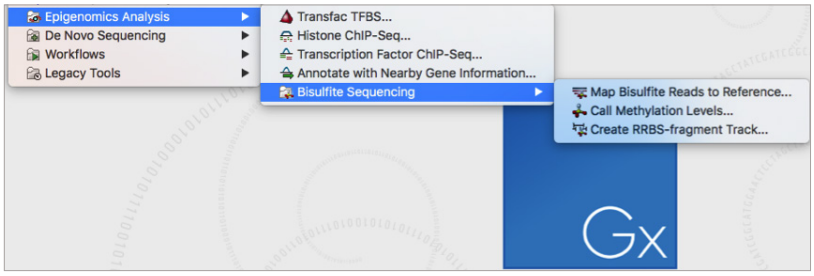


Figure 1. Bisulfite sequencing tools in the 'Epigenomics' toolbox upon plugin installation.

Here we demonstrate and benchmark the analysis of bisulfite sequencing data for human b-lymphocites differentiating from progenitor stem cells, using publicly available datasets.

Bisulfite sequencing tools and workflow

In CLC Genomics Workbench 8.5 and above the bisulfite sequencing functionality is available as a free downloadable plugin. Upon installation it creates three new tools in the 'Epigenomics' toolbox.

The first tool is for mapping in a special three-letter genetic code mode necessary for directional bisulfite sequncing reads. The second performs base level methylation calling on bisulfite read mapping(s), and statistical comparisons when more than one mapping is provided. The third tool is a utility to create a restriciton fragment track from a reference genome to limit the analysis with the first two tools in reduced-representation bisulfite sequencing.

The tools can be combined in workflows, and here we show an example based on our **tutorial***:

Here, case and control samples are mapped, methylation of individual cytosines called, and levels compared using Fisher's exact test in consecutive windows of fixed length.

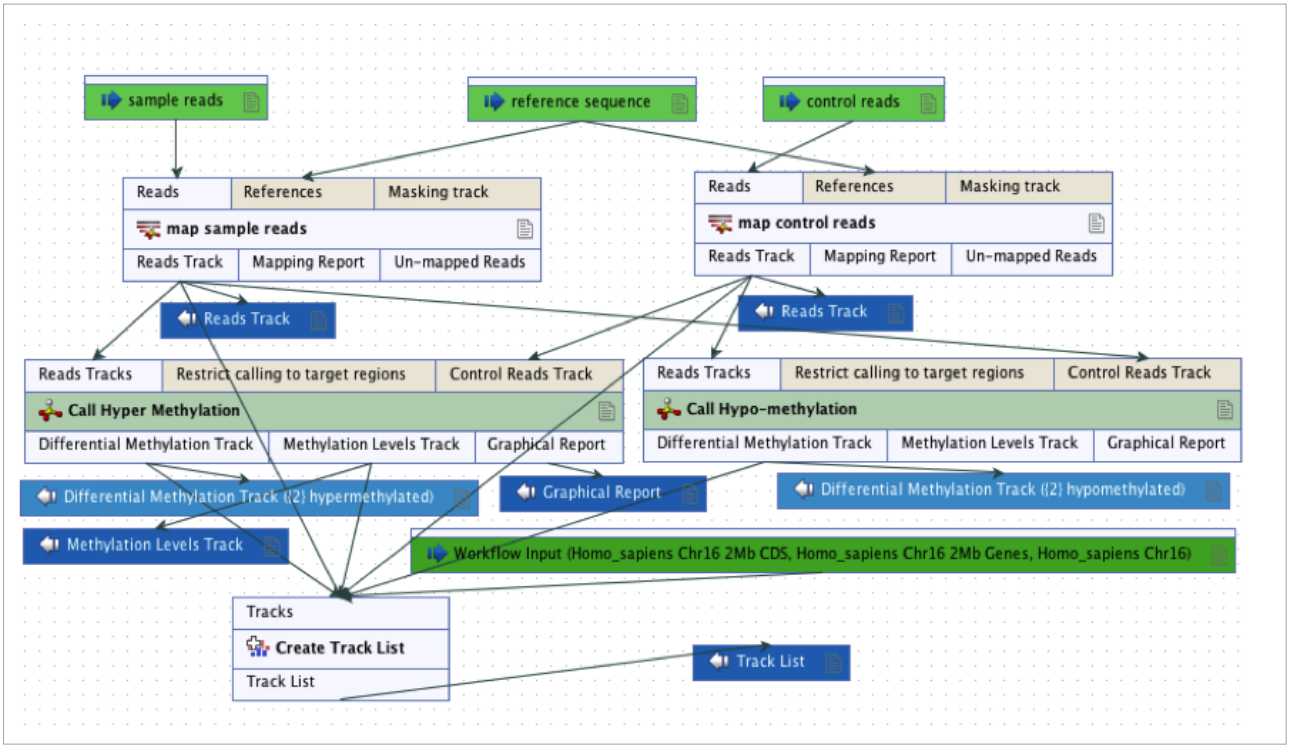


Figure 2. Example workflow to detect differential cytosine methylation and create numeric and visual outputs.

Hyper- and hypo-methylated regions are detected and can be investigated using a genome browser-like output (track list). It also produces a number of reports useful to interpret and trouble-shoot the results.

Figure 3. Visual output in the form of a Track List, showing hyper- and hypo-methylated sites in the context of genome annotations and mapped reads. Detected 5'-proximal hypomethylation in the CD19 gene is circled in red.

This example workflow was run on a subset of reads imported from NCBI Short Read Archive entries published in the Hodges, et al., 2011 paper. Running the example workflow on two sets of bisulfite sequencing reads, from human haemopoetic progenitor stem cells (hspc), and from differentiated b-lymphocytes (b-cells), reveals differential methylation in the CD19 gene. Hypomethylation in the 5'-end of the CD19 gene in b-cells (Figure 3) is easily detected, and is presumed to be responsible for the differential expression of the gene (Hodges et al., 2011). The observed methylation pattern is presumably responsible for the b-cell specific expression of this classical marker of the lymphocyte development.

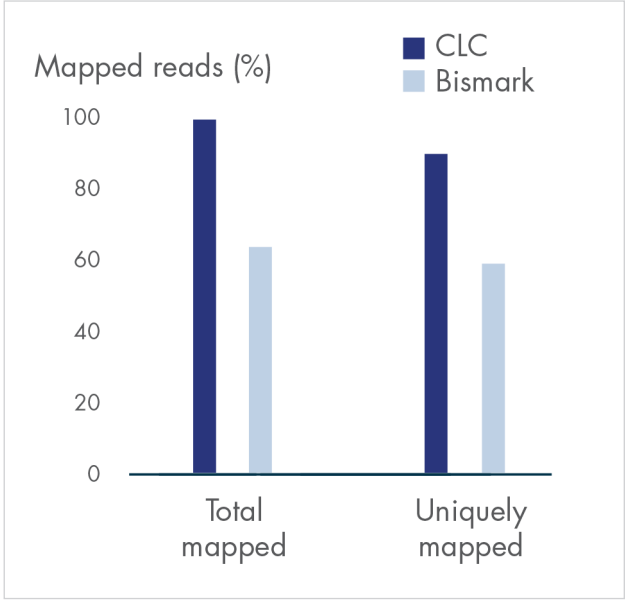


Figure 4. Mapping efficiency was consistently higher when comparing the CLC Bisulfite Seq Analysis tool with the open source standard Bismark.

Benchmarking

The data from Hodges et al. was used for comparison of various mappers available for the bisufite sequencing reads. In benchmarks published by Tran et al., 2014 Bismark emerged as the best mapper with respect to quality, comprehensiveness, and efficiency.

We therefore benchmarked the CLC bisufite mapper using the same datasets as Tran et al., 2014 against Bismark. We generally reproduced the results of Tran et al., 2014 with regard to Bismark's performance, which maps about 60% of reads in the test datasets, the highest fraction among the

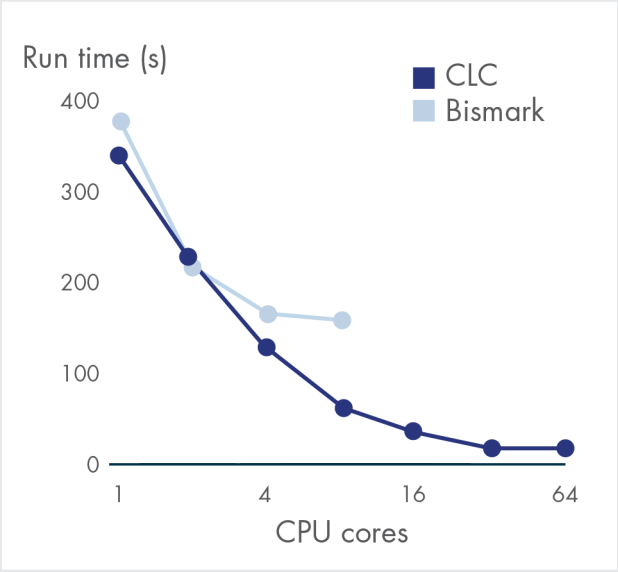


Figure 5. Superior run time and compute resource efficiency compared to leading alternative.

evaluated ones. However, the bisulfite mapper (CLC) is able to correctly place up to 30% more reads from the same datasets, achieving close to 100% efficiency of mapping in some cases. Close evaluation of mapping results confirmed that the additional reads were indeed mapped accurately.

In addition to accurate mapping, we also confirmed accurate detection and quantification of methylation levels by testing the workflow presented here using commercial methylation standard datasets for which methylation levels are known across 13 CG sites (Masser et al. 2013). Strong correlation (r^2=0.9845) between standard and CLC results indicates accurate determination of methylation levels.

Unlike Bismark, the CLC mapper is optimised for parallelized execution on high-performance architecture. This optimised multi-threaded execution drastically reduces the time you have to wait for your results. While Bismark's execution time actually degrades if more than 4 cores are allocated to it, the CLC mapper continues to benefit from up to 64 cores (highest investigated).

In conclusion, CLC Genomics Workbench tools for bisulfite sequencing data analysis extend the functionality of our scientist-friendly software. The tools deliver superior mapping performance without sacrificing accuracy. Optimised algorithms save run time and expensive compute resources and the ability to automate workflows adds to the ease of use.

*http://www.clcbio.com/files/tutorials/Bisulfite_Sequencing.pdf

References

Hodges, E., Molaro, A., Dos Santos, C. O., Thekkat, P., Song, Q., Uren, P. J., … Hannon, G. J. (2011). Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Molecular Cell, 44(1), 17–28. http://doi.org/10.1016/j.molcel.2011.08.026

Masser D. R., Berg A. S, Freeman W. M (2013). Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. Epigenetics & Chromatin 6:33, http://www.epigeneticsandchromatin.com/content/6/1/33

Tran, H., Porter, J., Sun, M. A., Xie, H., & Zhang, L. (2014). Objective and comprehensive evaluation of bisulfite short read mapping tools. Advances in Bioinformatics, 2014. http://doi.org/10.1155/2014/472045