

# Base-pair resolution DNA methylome of the EBV-positive Endemic Burkitt lymphoma cell line DAUDI determined by SOLiD bisulfite-sequencing

*Leukemia* (2013) **27**, 1751–1753; doi:10.1038/leu.2013.4

The Burkitt translocation t(8;14), first identified in the 1970s in biopsies and cell lines from Burkitt lymphoma (BL),<sup>1,2</sup> and its variants juxtapose the *MYC* oncogene to one of the immunoglobulin (*IG*) loci.<sup>3</sup> Nowadays, it is assumed that (nearly) all BL carry an *IG-MYC* translocation, rendering this somatic mutation a diagnostic marker for all three subtypes of BL (endemic, sporadic and immunodeficiency-related BL).

In contrast to many other lymphomas, BL show a quite simple karyotype, that is, with few if any secondary chromosomal changes in addition to the *IG-MYC* translocation.<sup>4</sup> Although there is evidence for some few recurrent secondary genetic changes the number of epigenetic alterations in BL as compared with normal B-cell subsets seems to outnumber the genetic changes by far.<sup>5–7</sup> Indeed, along with others, we have identified several hundred genes showing *de novo* DNA methylation in aggressive B-cell lymphoma, including BL as compared with normal B-cell subsets.<sup>8–10</sup> Nevertheless, the mentioned DNA methylation studies focused on a maximum of probably 10% (by HELP assays) of the CpGs of the genome, and were biased toward promoter regions and CpG islands and did not systematically analyze non-CpG methylation.<sup>8–11</sup> Therefore, we here aimed at generating a complete DNA methylome of a BL, allowing for unbiased analyses of all cytosines in the genome.

To this end, we chose the archetypal DAUDI cell line, established from an endemic BL (eBL) that was derived from a 16-year-old African male patient in 1967.<sup>12,13</sup> We selected this cell line as it has been pivotal for the identification of t(8;14), still carries a simple karyotype despite being many years in culture and because it shows the prototypic features of eBL. Moreover, considering the strong association of eBL with Epstein–Barr virus (EBV) infection, the EBV-positive DAUDI cell line offers the opportunity for a direct comparison of its lymphoma and EBV methylomes.

To obtain a base-pair resolution DNA methylome of a prototypic eBL, we subjected DNA of the DAUDI cell line to full bisulfite-sequencing (BS-seq) using the SOLiD two-base encoding (color-space) approach (for details, see Supplementary Information). Two bisulfite-converted SOLiD fragment libraries were constructed. Briefly, 15 µg of genomic DNA were sheared to approximately 125 bp. After end-repair of the DNA fragments, methyl-P1 and -P2 adaptors were ligated. The DNA was then size selected and nick translated with a modified deoxyribonucleoside triphosphate (dNTP) mix containing methyl-deoxycytidinetriphosphate (dCTP) instead of regular dCTPs. Bisulfite conversion was carried out in solution and recovered DNA fragments were PCR amplified using eight cycles. The bisulfite-converted fragment library was clonally amplified on SOLiD P1 beads using emulsion PCR. Templated (P2 positive) beads were then enriched and deposited on a slide for sequencing. Technical details on the genomic characterization of the cell line including karyotyping, single-nucleotide polymorphism (SNP) array analysis and exome sequencing as well as on control DNA methylation analyses using Illumina 450K BeadArray analysis (Illumina, San Diego, CA, USA), limited BS-seq using Illumina HiSeq2000 technology as well as Luminometric

methylation assay (LUMA) are provided in the Supplementary Information.

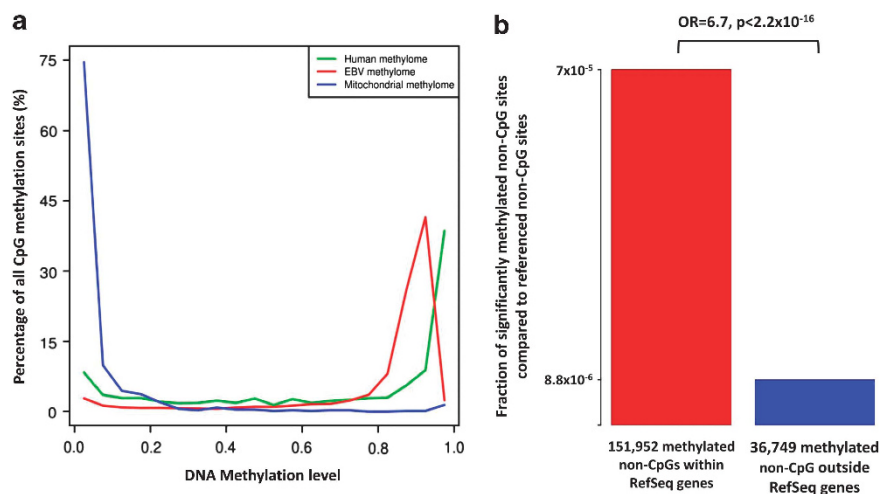
By karyotyping and SNP-array analysis, we confirmed that the cells under study show the typical features of BL, including the t(8;14) plus a few secondary chromosomal changes. Exome sequencing followed by filtering for known SNPs revealed a total of 2313 non-synonymous mutations (Supplementary Table 1a). Owing to the lack of a germline control from the patient from which the DAUDI cell line has been established, it is not possible to reliably differentiate somatic (lymphoma-associated) mutations from germline variants. Despite this limitation, exome sequencing in line with recent reports,<sup>5–7</sup> identified inactivating mutations in the *ID3* gene, which have been shown to co-operate with *MYC* activation in the pathogenesis of BL.<sup>5–7,14</sup> In addition, Sanger sequencing confirmed sequence variants in the genes *B2M*, *TET2* and *KIT* (Supplementary Table 1b).

We aligned 79.9 Gb of BS-seq of the SOLiD platform.<sup>15</sup> These were compared with 7.8 Gb aligned BS-seq of the HiSeq 2000 platform and the results to the DNA methylation levels determined by HumanMethylation450 BeadChip analysis (Supplementary Figure S1). We observed high correlation of the SOLiD data with both the sequence-based HiSeq 2000 (Pearson  $r = 0.86$ ; Supplementary Figure S2) and the array-based (Pearson  $r = 0.96$ ; Supplementary Figure S3) methylation levels. This led us to focus our further analyses on the most extensive data set derived from SOLiD BS-seq.

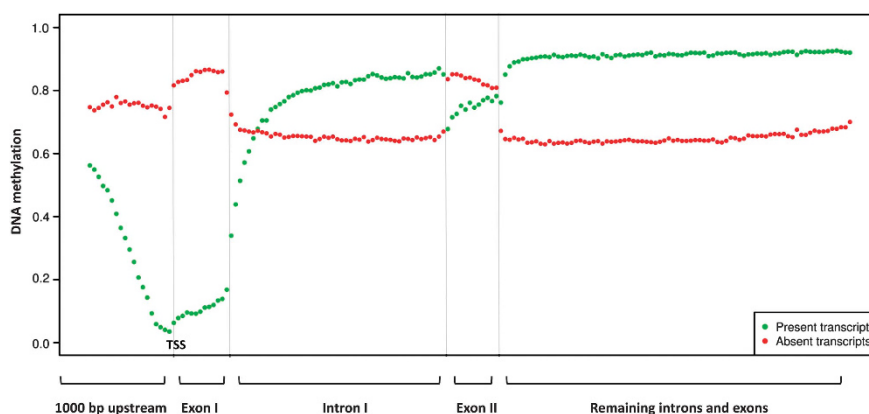
In total, 91.1% of all CpG sites and 90.2% of all non-CpG sites of the genome were covered by at least five SOLiD reads (Supplementary Figure S4). On the genome-wide level, 68.99% cytosines in CpG dinucleotides were methylated, which is in line with previous pyrosequencing-based determinations using LUMA. In contrast, the 450K BeadArray shows a mean methylation level of 59.24%, which is mostly due to the selection bias of the array loci, which are predominantly located within regions upstream of genes. We observed a mean CpG methylation level of 65.79% in LINEs and 78.84% in SINEs (Supplementary Figure S5). BS-seq shows the DNA methylation patterns on the forward and reverse strand to be comparably established (Pearson  $r = 0.90$ ).

Considering the recent description of non-CpG methylation in embryonic stem cells (ESCs), and the fact that the *MYC* oncogene deregulated in BL is also one of the four factors used to induce a stem cell-like phenotype in differentiated cells,<sup>16</sup> we analyzed the level of non-CpG methylation. The genome-wide fraction of methylated cytosines in a non-CpG context does not exceed the respective threshold of 0.003 given by the unmethylated lambda control DNA that was tested in parallel.<sup>11</sup> Moreover, we confirmed absence of non-CpG methylation at hallmark sites described in ESC<sup>11</sup> by bisulfite pyrosequencing (Supplementary Figure S6). Despite this overall low frequency of non-CpG methylation, we could identify a remarkable 6.7-fold enrichment of methylated non-CpG sites within genes ( $P < 2.2 \times 10^{-16}$ ; Figure 1b). Such non-CpG methylation might be linked to transcriptional activity (Supplementary Figure S7).

We next determined the sequence-based methylation status of 969 genes recently shown by us to exhibit *de novo* promoter hypermethylation in mature aggressive B-cell lymphoma (including BL) as compared with normal B cells.<sup>8</sup>



**Figure 1.** Distribution of DNA methylation in DAUDI cells. (a) The graphs show genome-wide distributions of CpG methylation of the human nuclear, EBV and mitochondrial genomes. The y axis indicates DNA methylation levels assessed by SOLiD BS-seq. Green: human nuclear CpG methylation; red: EBV CpG methylation; blue: mitochondrial CpG methylation. (b) Significantly methylated non-CpG sites of DAUDI within RefSeq genes (comprising 424 969 306 non-CpGs) are 6.7-fold enriched compared with those outside of RefSeq genes (comprising 689 750 660 non-CpGs). Red: fraction of significantly methylated non-CpGs of DAUDI within RefSeq genes; blue: fraction of significantly methylated non-CpGs of DAUDI outside of RefSeq genes. OR: odds ratio.



**Figure 2.** Correlation of DNA methylation levels and transcriptional states. CpG methylation levels were averaged for annotated RefSeq gene regions and transcripts are clustered by their expression level in present ( $n = 7662$  transcripts) and absent ( $n = 5429$  transcripts) calls. A strong dependency of the location of CpGs related to their distance to the TSS and the transcript expression level can be observed. Green: average methylation pattern for present transcripts; red: average methylation pattern for absent transcripts.

We could confirm that in DAUDI cells 91.21% of these genes have a DNA methylation level  $\geq 60\%$  in their promoter region and lack transcription. As compared with all other RefSeq genes, the mean CpG methylation level within promoter regions of the 969 genes was significantly higher (84% vs 41%; Supplementary Tables S2).

Gene expression analyses confirmed that DAUDI cells show the typical signature of molecular BL.<sup>17</sup> Correlating methylation and expression patterns in our data revealed that significant presence of transcripts is associated with absence of DNA methylation particular at and closely around the transcription start site (TSS). In contrast, DNA methylation exactly at the TSS correlates with lack of transcription (Figure 2). Although the group of non-expressed genes showed an overall high mean DNA methylation level across the whole gene with highest methylation levels in exons, genomic regions comprising expressed genes were characterized by particular high methylation levels in the first intron. Moreover, the patterns of both expressed and non-expressed genes were characterized by sharp transitions of methylation levels at exon-intron borders (Figure 2). Overall, when compared with transcriptional activity the patterns of DNA methylation across different

parts of genes were similar to those recently determined in non-neoplastic tissues including blood.<sup>18</sup>

Finally, we studied the DNA methylation of the mitochondrial and EBV genomes of DAUDI cells.<sup>19–21</sup> We estimated 80 EBV and 370 mitochondrial copies per DAUDI cell based on coverage analyses, which is in accordance with previous studies.<sup>22</sup> Although mitochondrial DNA is mostly unmethylated (mean methylation 6.43%; Figure 1a), CpG methylation in the human and EBV genome is comparably distributed, although the EBV genome exhibits hardly any fully methylated sites (Figure 1a). Overall, EBV shows a high level of DNA methylation (mean methylation 80.18%), as it was previously shown for BL cell lines.<sup>23</sup> Nevertheless, DNA methylation within the EBV genome correlates with expression only at high transcript levels (fragments per kilobase of exon model per million mapped (FPKM)  $\geq 15$ ) (Supplementary Figure S8).

In summary, we have characterized the nuclear DNA methylome of an eBL along with its mitochondrial and EBV methylome using colorspace BS-seq. We unravel significant differences between the different sub-methylomes and moreover show that gene transcription is associated with complex patterns of methylation, extending beyond simple promoter and CpG

methylation. As the DAUDI cell line has been used over decades in many laboratories in the world, the obtained methylome data might serve as a 'reference epigenome' for future studies.

Data availability: methylome data are available at [sftp://134.245.63.215/export/home/daudi](http://sftp://134.245.63.215/export/home/daudi) (login: daudi; password: daudismethylome2012).

## CONFLICT OF INTEREST

OA and RS received research support by Illumina on BeadChip Analyses. CABG is and VLB used to be employed by Life Technologies. The remaining authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This study was financed through a grant from the 'Medizinausschuss Schleswig-Holstein'. The project received support through the BMBF ICGC MMML-Seq project (FKZ 01KU1002A and 01KU1002E) and the EU Seventh Framework Programme (FP7/2007-2013, grant number 262055, ESGI), and infrastructure through the DFG Cluster of Excellence 'Inflammation at Interfaces'. Bisulfite pyrosequencing was supported by KinderKrebsInitiative Buchholz/Holm-Seppensen and gene expression data were obtained in the framework of the Deutsche Krebshilfe-Network 'Molecular Mechanisms in Malignant Lymphomas (MMML)'. This article is dedicated to Professor Lore Zech for her groundbreaking research on chromosomal aberrations in hematologic neoplasms in general and BL in particular.

## AUTHOR CONTRIBUTIONS

BK, JR, OA, PR, AF and RS designed the research, interpreted data and wrote the paper. PR, AF, OA, SSc and RS provided infrastructure and obtained grant support. IV, EMMP, HGD and RAFM performed cell line characterization. MH performed array-based gene expression profiling. BK, MB and FK analyzed BS-seq data. BK, JR and OA analyzed pyrosequencing data and array-based DNA methylation data. BK, DE and RH analyzed RNA-sequencing data. BSP analyzed exome sequencing data. CABG and VLB prepared the SOLiD BS-seq library. SSe and SSm prepared the Hi-Seq 2000 BS-seq libraries; and all authors critically reviewed the paper.

B Kreck<sup>1</sup>, J Richter<sup>2</sup>, O Ammerpohl<sup>2</sup>, M Barann<sup>1</sup>, D Esser<sup>1</sup>, BS Petersen<sup>1</sup>, I Vater<sup>2</sup>, EM Murga Penas<sup>2</sup>, CA Bormann Chung<sup>3</sup>, S Seisenberger<sup>4</sup>, V Lee Boyd<sup>3</sup>, S Smallwood<sup>4</sup>, HG Drexler<sup>5</sup>, RAF MacLeod<sup>5</sup>, M Hummel<sup>6</sup>, F Krueger<sup>4</sup>, R Häsler<sup>1</sup>, S Schreiber<sup>1,7</sup>, P Rosenstiel<sup>1</sup>, A Franke<sup>1,8</sup> and R Siebert<sup>2,8</sup>  
<sup>1</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany;  
<sup>2</sup>Institute of Human Genetics, Christian-Albrechts-University of Kiel and University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany;  
<sup>3</sup>Life Technologies, Foster City, CA, USA;  
<sup>4</sup>The Babraham Institute, Cambridge, UK;  
<sup>5</sup>Department of Human and Animal Cell Cultures, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany;  
<sup>6</sup>Institute of Pathology, Charité—University Medicine Berlin, Berlin, Germany and  
<sup>7</sup>Department of General Internal Medicine, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany  
 E-mail: [a.franke@mucosa.de](mailto:a.franke@mucosa.de) or [rsiebert@medgen.uni-kiel.de](mailto:rsiebert@medgen.uni-kiel.de)  
<sup>8</sup>These authors share senior authorship.

## REFERENCES

- Manolov G, Manolova Y. Marker band in one chromosome 14 from Burkitt lymphomas. *Nature* 1972; **237**: 33–34.

- Zech L, Haglund U, Nilsson K, Klein G. Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. *Int J Cancer* 1976; **17**: 47–56.
- Salaverria I, Siebert R. The gray zone between Burkitt's lymphoma and diffuse large B-cell lymphoma from a genetics perspective. *J Clin Oncol* 2011; **29**: 1835–1843.
- Boerma EG, Siebert R, Kluijn PM, Baudis M. Translocations involving 8q24 in Burkitt lymphoma and other malignant lymphomas: a historical review of cytogenetics in the light of today's knowledge. *Leukemia* 2009; **23**: 225–234.
- Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B *et al*. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat Genet* 2012 2012; **44**: 1316–1320.
- Love C, Sun Z, Jima D, Li G, Zhang J, Miles R *et al*. The genetic landscape of mutations in Burkitt lymphoma. *Nat Genet* 2012 2012; **44**: 1321–1325.
- Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M *et al*. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* 2012 2012; **490**: 116–20.
- Martin-Subero JI, Kreuz M, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E *et al*. New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood* 2009; **113**: 2488–2497.
- Shaknovich R, Geng H, Johnson NA, Tsikitas L, Cerchiatti L, Grealley JM *et al*. DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood* 2010; **116**: e81–e89.
- Shaknovich R, Cerchiatti L, Tsikitas L, Kormaksson M, De S, Figueroa ME *et al*. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* 2011; **118**: 3559–3569.
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J *et al*. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009; **462**: 315–322.
- Klein E, Klein G, Nadkarni JS, Nadkarni JJ, Wigzell H, Clifford P. Surface IgM-kappa specificity on a Burkitt lymphoma cell in vivo and in derived culture lines. *Cancer Res* 1968; **28**: 1300–1310.
- Nadkarni JS, Nadkarni JJ, Clifford P, Manolov G, Fenyö EM, Klein E. Characteristics of new cell lines derived from Burkitt lymphomas. *Cancer* 1969; **23**: 64–79.
- Sander S, Calado DP, Srinivasan L, Köcher K, Zhang B, Rosolowski M *et al*. Synergy between PI3K signaling and MYC in Burkitt lymphomagenesis. *Cancer Cell* 2012; **22**: 167–179.
- Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing data. *Nat Methods* 2012; **9**: 145–151.
- Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; **126**: 663–676.
- Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF *et al*. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med* 2006; **354**: 2419–2430.
- Li Y, Zhu J, Tian G, Li N, Li Q, Ye M *et al*. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol* 2010; **8**: e1000533.
- Lin Z, Xu G, Deng N, Taylor C, Zhu D, Flemington EK. Quantitative and qualitative RNA-Seq-based evaluation of Epstein-Barr virus transcription in type I latency Burkitt's lymphoma cells. *J Virol* 2010; **84**: 13053–13058.
- Flower K, Thomas D, Heather J, Ramasubramanian S, Jones S, Sinclair AJ. Epigenetic control of viral life-cycle by a DNA-methylation dependent transcription factor. *PLoS ONE* 2011; **6**: e25922.
- Dresang LR, Teuton JR, Feng H, Jacobs JM, Camp 2nd DG, Purvine SO *et al*. Coupled transcriptome and proteome analysis of human lymphotropic tumor viruses: insights on the detection and discovery of viral genes. *BMC Genomics* 2011; **12**: 625.
- Leenman EE, Panzer-Grümayer RE, Fischer S, Leitch HA, Horsman DE, Lion T *et al*. Rapid determination of Epstein-Barr virus latent or lytic infection in single human cells using in situ hybridization. *Mod Pathol* 2004; **17**: 1564–1572.
- Kim DN, Song YJ, Lee SK. The role of promoter methylation in Epstein-Barr virus (EBV) microRNA expression in EBV-infected B cell lines. *Exp Mol Med* 2011; **43**: 401–410.



This work is licensed under a Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>

Supplementary Information accompanies the paper on the Leukemia website (<http://www.nature.com/leu>)