Trisomy 8 Acute Myeloid Leukemia Analysis Reveals New Insights of DNA Methylome with Identification of HHEX as Potential Diagnostic Marker

Marwa H. Saied1,3, Jacek Marzec1,2, Sabah Khalid1, Paul Smith1, Gael Molloy1 and Bryan D. Young1

1Centre for Haematology-Oncology, Barts Cancer Institute, Barts and the London School of Medicine, Queen Mary University of London, London, United Kingdom. 2Centre for Molecular Oncology, Barts Cancer Institute, Barts and the London School of Medicine, Queen Mary University of London, London, United Kingdom. 3Clinical Pathology Department, Faculty of Medicine, Alexandria University, Alexandria, Egypt.

ABSTRACT: Trisomy 8 acute myeloid leukemia (AML) is the commonest numerical aberration in AML. Here we present a global analysis of trisomy 8 AML using methylated DNA immunoprecipitation-sequencing (MeDIP-seq). The study is based on three diagnostic trisomy 8 AML and their parallel relapse status in addition to nine non-trisomic AML and four normal bone marrows (NBMs). In contrast to non-trisomic DNA samples, trisomy 8 AML showed a characteristic DNA methylation distribution pattern because an increase in the frequency of the hypermethylation signals in chromosome 8 was associated with an increase in the hypomethylation signals in the rest of the chromosomes. Chromosome 8 hypermethylation signals were found mainly in the CpG island (CGI) shores and interspersed repeats. Validating the most significant differentially methylated CGI (P = 7.88 × 10−11) identified in trisomy 8 AML demonstrated a specific core region within the gene body of HHEX, which was significantly correlated with HHEX expression in both diagnostic and relapse trisomy 8 AML. Overall, the existence of extra chromosome 8 was associated with a global impact on the DNA methylation distribution with identification of HHEX gene methylation as a potential diagnostic marker for trisomy 8 AML.

KEYWORDS: trisomy 8, MeDIP-seq, CGI shores, HHEX

Introduction

Trisomy 8 is the most common numerical aberration in acute myeloid leukemia (AML).1 It occurs as a sole abnormality in 6% of AML and coexists with other numerical aberrations in 16% of AML.2 The extra chromosome 8 not only has been shown to affect the expression of genes located on chromosome 83 but also has an impact on the global gene expression.2,4 Using methylated DNA immunoprecipitation-sequencing (MeDIP-seq) followed by next-generation sequencing, we previously clustered trisomy 8 AML according to the DNA methylation of different genomic features, eg, promoters, gene bodies, CpG islands (CGIs), and CGI shores.5 Also, clusters of hypomethylated interspersed repeats, short interspersed nuclear elements.
t(8;21) AML patients (n = 3), t(15;17) AML patients (n = 3), and normal karyotype (NK) AML patients (n = 3), and four NBMs from healthy donors.\textsuperscript{5} The ethical approval was obtained from East London and the City Research Ethics Committee (ref. 10/H0704/65). Written informed consent was obtained from patients to store the excess diagnostic tissue for research purposes according to the Declaration of Helsinki. When patient consent was not obtained because of death, patient anonymity was preserved. The MeDIP-seq protocol\textsuperscript{6} was applied to 5 μg of DNA extracted from either peripheral blood or bone marrow. In brief, 5 μg genomic DNA undergoes sonication, correction to the ends of each fragment by enzymatic reactions in order to add the adaptors to both ends and for further treatment of the adaptors-ligated DNA fragments with a monoclonal antibody to 5-methylcytosine.\textsuperscript{6} The MeDIP-seq libraries were sequenced by Illumina Genome Analyzer II, followed by alignment of the reads to a human reference genome (NCBI36/hg18) by Mac\textsuperscript{7} and Bowtie\textsuperscript{8} algorithms. The methylation signals were further quantified by the Batman algorithm\textsuperscript{9} (a Bayesian tool for methylation analysis), which infers the absolute methylation state for 100-bp windows by estimating local sequencing read enrichment for methylation taking into account the varying densities of CpGs across the genome. Batman score \( \geq 0.6 \) implies a high methylated region, a score \( \leq 0.4 \) implies a low methylated region, and a score between 0.4 and 0.6 implies intermediate methylated region. For the detection of differentially methylated regions (DMRs), we applied the previously described MeDIP-seq statistical workflow.\textsuperscript{5} For detection of frequent genes in the DMRs, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatic tool.\textsuperscript{10} HHEX gene expression was measured relative to NBM derived from a healthy donor by relative quantitative real-time PCR (RT-PCR). TaqMan probe and primers are designed for each gene by Applied Biosystems. Primer details are listed in Supplementary Table 5. All samples are run in duplicates. A reaction mix contains 5 μL TaqMan Universal Master Mix II (catalog number 4440040, Applied Biosystems), 0.5 assay mix, 3.5 μL water, and 1 μL cDNA placed in each well of a 96-well reaction plate. RT-PCR reactions are run on the Applied Biosystems 7900HT Fast RT-PCR System using the standard thermal cycler protocol with an initial step at 95°C for 10 minutes, followed by 40 cycles at 95°C for 15 seconds and annealing and extension at 60°C for 1 minute. Ct (cycle threshold) for each gene is measured, which equals the mean of Ct duplicates for each sample. Ct is assessed relative to the Ct of a reference (18S). For calculating the expression fold change, these steps are performed:

\[
\Delta C_t = (C_{t_{\text{sample}}} - C_{t_{\text{reference}}})
\]

\[
\Delta \Delta C_t = (\Delta C_{t_{\text{sample}}} - \Delta C_{t_{\text{calibrator}}})
\]

Expression fold change = \( 2^{-\Delta C_t} \)

**Results and Discussion**

**Differences in DNA methylation distribution pattern between trisomic and non-trisomic DNA samples.** In order to determine the potential epigenomic impact of extra chromosome 8, DNA methylation distribution was first compared between chromosome 8 and the rest of the chromosomes in a single trisomy 8 AML patient. Chromosome 8 showed a higher frequency of Batman scores, \( >0.8 (>80\% \text{ methylation}) \), and a lower frequency of Batman scores, \( <0.4 (<40\% \text{ methylation}) \), than the rest of the chromosomes (Fig. 1A). To explore whether this pattern of DNA methylation distribution is specific to trisomy 8 AML, we repeated the comparison using the methylation scores of all trisomy 8 AML samples (three diagnostic and three relapse), the average of the methylation scores of nine non-trisomic AML patients, and the average of the methylation scores of four NBMs. Unlike trisomy 8 AML (Fig. 1B), there was no difference in the DNA methylation distribution between chromosome 8 and the rest of the chromosomes in either non-trisomic AML (Fig. 1C) or NBM (Fig. 1D). In order to localize the site of chromosome 8 extra methylation signals, DNA methylation of trisomy 8 AML (three diagnostic and three relapse) was compared to DNA methylation of four NBMs. Different chromosomes were frequently enriched in the DMRs between trisomy 8 AML and NBM with a notable representation of chromosome 8 in the intragenic differentially hypermethylated CGI shores (within promoters or inside gene bodies) in trisomy 8 AML (Fig. 2A and Supplementary Table 2). Additionally, the chromosomal DNA methylation distribution was compared between trisomic and non-trisomic DNA samples. Instead of investigating all the chromosomes, we chose three random chromosomes, eg, 1, 6, and 11, in addition to chromosome 8. Chromosomes 1, 6, and 11 showed a higher frequency of Batman scores, \( <0.4 \) in trisomy 8 AML than non-trisomic DNA samples (Supplementary Fig. 1). This was in contrast to chromosome 8 that showed a higher frequency of Batman scores, \( >0.8 \) in trisomy 8 AML than non-trisomic DNA samples (Supplementary Fig. 1). Collectively, it could be suggested that the presence of extra chromosome 8 led to an increase in the chromosome 8 methylation signals. Recently, an introduction of chromosome 8 into a normal human cell revealed an increase in the average level of gene expression on chromosome 8 compared to a diploid normal human cell.\textsuperscript{4}
Also, in the same study, the average level of gene expression of all non-trisomic chromosomes in the artificial trisomy 8 cells decreased. Similarly, we presented that there was no difference in the DNA methylation distribution between chromosome 8 and the rest of the chromosomes in either non-trisomic AML patients or NBMs. By contrast, trisomy 8 AML showed an increase in the methylated regions of chromosome 8 with an increase in the hypomethylated regions in the remaining non-trisomic chromosomes. Considering the previous evidence of the global impact of trisomy 8 on gene expression, it could be suggested that the presence of extra chromosome 8 not only affects chromosome 8 methylation but also might have had a global DNA methylation effect.

Global DNA methylation differences between diagnostic and relapse trisomy 8 AMLs. We investigated the DNA methylation differences between diagnostic and relapse trisomy 8 AMLs with a predicted increase in the CGIs DNA methylation in relapse. CGI shores showed the highest number of DMRs between diagnostic and relapse trisomy 8 AMLs among promoters, gene bodies, and CGIs (Supplementary Table 3). Most of those differentiating CGI shores were hypermethylated in diagnostic versus relapse status. Moreover, different chromosomes, not including chromosome 8, were frequently represented in the hypermethylated DMRs in relapse (Supplementary Table 3). Although the two-dimensional cluster analysis (Fig. 2A) showed clear discrimination between
Figure 2. Hierarchical clustering analysis of trisomy 8 AML versus NBM in four genomic features. In each cluster analysis, each column represents AML patient/NBM and each row represents a DMR. Color represents methylation level (red = high methylation and green = low methylation). In the four genomic features, promoters (A), gene bodies (B), CGIs (C), and CGI shores (D), diagnostic and relapse trisomy 8 AMLs were clustered together away from NBM with a high number of hypermethylated CGIs, and CGI shores were observed in trisomy 8 AML. HHEX gene methylation and expression in trisomy 8 AML (B). X-axis represents AML patients and NBMs; Y-axis represents Batman scores for each sample. The significant difference in HHEX gene methylation between trisomy 8 AML and the rest of the samples was observed in a CGI located within the gene body of HHEX, NK, NK AML, and NBM. (C) The significant methylated CGI (green-colored block) is located between exon 2 and exon 3 (blue-colored blocks) of HHEX gene. The blue arrow indicates the direction of the gene (the figure is adapted from UCSC Genome Browser). (D and E) Both CGI sequences (1 and 2) showed significant methylation differences among the groups investigated (Kruskal–Wallis test; \( P = 0.0004 \) and \( P = 0.0009 \), respectively). (D) Trisomy 8 AML methylation was significantly higher than the DNA methylation of t(8;21) AML, inversion 16 AML, and NBMs, while in (E), trisomy 8 AML methylation was significantly higher than the DNA methylation of only NBMs (Dunn’s multiple comparison tests; \( P < 0.05 \)). (F, G) There was a significant expression difference among the groups investigated in HHEX1 and HHEX2 (Kruskal–Wallis test; \( P = 0.02 \) and \( P = 0.04 \), respectively). HHEX gene expression was significantly less in trisomy 8 AML than t(8;21) AML in both (D) and (E) (Dunn’s multiple comparison tests; \( P < 0.05 \)). In each figure, each dot represents HHEX methylation/expression; the horizontal line represents the mean of these dots. N is the number of samples investigated, and PB is the peripheral blood from four healthy donors. HHEX methylation was the average methylation percentage of all CpG sites sequenced as calculated by pyrosequencing. HHEX gene expression was measured relative to NBM. RT-PCR primers and pyrosequencing primers are provided in the supplementary information (Supplementary Tables 4 and 5). Trisomy 8 AML D, diagnostic; trisomy 8 AML R, relapse.
diagnostic and relapse in the four genomic features, eg, promoters, gene bodies, CGIs, and CGIs shores, few DMRs passed the >25% absolute methylation difference. One of the interesting hypermethylated genes identified in relapse versus diagnostic was RGS2. RGS2 was repressed by AML-typical Fli3 mutations. RGS2 overexpression inhibited Fli3 internal tandem duplication (ITD)-induced transformation of myeloid cells and negatively influenced some Fli3-ITD-induced signaling events, such as phosphorylation of Akt or GSK3-β.3,13 Furthermore, SINE-, LINE-, and LTR-associated DMRs were more hypomethylated in relapse. Again, few DMRs associated with repeats showed absolute DNA methylation difference >25. Also, there were different frequently representative chromosomes, not including chromosome 8, in the differentiating repeats in relapse. Overall, this analysis revealed little significant methylation difference between diagnostic and relapse trisomy 8 AMLs, with most of the DMRs more hypomethylated in relapse. The more hypomethylation of interspersed repeats could be linked to the therapeutic resistance in relapse status because of loss of genome integrity.14,15

**HHEX gene methylation and expression in trisomy 8 AML.** We previously identified intragenic CGI-associated DMR within the gene body of HHEX that was significantly hypermethylated in trisomy 8 AML versus the other groups of AML and NBMs (P = 9 × 10^-4).9 This significant difference in methylation increased when more trisomy 8 AML patients (diagnostic plus relapse) were included (P = 7.88 × 10^-11). The CGI is located in chromosome 10 (chromosome 10: 94442180–94442408). Notably, from the MeDIP-seq results, we checked HHEX gene promoter and gene body methylation among all 15 AML patients (9 non-trisomic AML, and 3 diagnostic and 3 relapse trisomy 8 AML) and 4 NBMs. There was no significant methylation difference detected between trisomy 8 AML patients and the rest of the samples in either HHEX promoter or gene body (Fig. 2B). The CpG density of this island is around 9.2% (21 CpG sites/228 bp) and located between exon 2 and exon 3 within the gene body of HHEX (Fig. 2C). We validated CGI methylation by bisulfite pyrosequencing from CpG sites 8–16 and CpG sites 17–21 in five diagnostic samples and three relapse samples. For sequence 1 (CpGs 8–16), trisomy 8 AML showed significant methylation differences with t(8;21) AML, inversion 16 AML, and NBMs (P < 0.05) (Fig. 2D). However, in sequence 2 (CpGs 17–21), trisomy 8 AML showed a significant methylation difference only with NBMs (P < 0.05) (Fig. 2E). Investigating the HHEX expression of these patients by relative quantitative RT-PCR revealed that trisomy 8 AML expression was significantly less than the expression of t(8;21) AML in HHEX-sequence 1 (across exon boundaries 1 and 2) and HHEX-sequence 2 (across exon boundaries 3 and 4) (P < 0.05) (Fig. 2F and G). Correlating the gene expression with DNA methylation revealed that CGI methylation (CpGs 8–16) showed significant inverse correlations with HHEX-sequence 1 and HHEX-sequence 2 (Spearman \( r = -0.5, P = 0.006; \) Spearman \( r = -0.43, P = 0.02, \) respectively) (Supplementary Fig. 2). However, CGI methylation (CpGs 17–21) did not show such correlation with either HHEX-sequence 1 or HHEX-sequence 2 (P = 0.22 and P = 0.24, respectively). A previous study by Topisirovic et al reported that HHEX gene expression was altered in AML depending on the French-American-British (FAB) criteria.16 Topisirovic et al observed that HHEX gene was downregulated in M4/M5 AML (11 AML patients, one of them was +8), while the gene expression showed a similar level to NBM in M1/M2 AML (5 AML patients; 4 of them were t(8;21) AML). In the present study, trisomy 8 AML patients, who showed a significant decrease in HHEX expression, had different phenotypic criteria; four of them were M4/M5 AML and four of them were M1/M2 AML. Additionally, t(8;21) AML patients, who showed significant increase in HHEX gene expression, were all M2 AML. Therefore, the significant decrease in HHEX gene expression in trisomy 8 AML patients did not depend only on the phenotypic criteria of those patients. Together, the persistence of HHEX gene methylation in both diagnostic and relapse trisomy 8 AMLs and subsequent gene repression suggest that HHEX gene methylation may have an important role in the pathogenesis of trisomy 8 AML. HHEX gene becomes a transcriptional activator on fusion with proteins, eg, NUP9817 or V16.16 HHEX plays important roles in the development of hemangioblast from mesodern during embryogenesis19 and in the maturation of early hematopoietic progenitors.20 However, HHEX is repressed in terminal myeloid differentiation and T-cell lineage.21 Moreover, HHEX is involved in the pathogenesis of AML because HHEX interacts with the promyelocytic leukemia (PML) protein. This interaction is important for PML protein’s role in growth control. In t(15;17) AML, the fusion protein PML–RARα disrupts HHEX–PML interaction, thus, contributing to the pathogenesis of AML.22 Also, fusion of HHEX gene with NUP98 gene resulted in HHEX–NUP98 protein, which produced a gene expression profile similar to HOXA9–NUP98 fusion protein, which is highly leukemogenic.23 In addition, overexpression of HHEX in K562 cell line resulted in repression of vascular endothelial growth factor (VEGF) and VEGF receptor genes.24

In conclusion, the MeDIP-seq analysis of trisomy 8 AML revealed characteristic DNA methylation distribution pattern with identification of HHEX gene as a methylated and repressed gene in both diagnostic and relapse trisomy 8 AMLs.

**Author Contributions**

Designed research: MHS, JM, SK, and BDY. Performed research: MHS, PS, and GM. Analyzed data: MHS, JM, SK, and BDY. Wrote the manuscript: MHS. The present study has been seen and approved by all authors.

The following link will allow review of our MeDIP-seq relapse data submitted to GEO: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=tdtonuqwyweady&acc=GSE38483, in addition to the publicly available data of MeDIP-seq at GEO (record GSE28314).
**Supplementary Data**

**Supplementary table 1A.** Criteria of trisomy 8 AML samples.

**Supplementary table 1B.** Sequencing reads of trisomy 8 AML samples.

**Supplementary table 2.** DMRs detected in trisomy 8 AML with the most frequent representative chromosomes.

**Supplementary table 3.** List of the most significant DMRs between trisomy 8 AML diagnostic versus relapse and their location. AMD; absolute methylation difference, red is hypermethylated DMR in relapse, green is hypomethylated DMR in relapse.

**Supplementary table 4.** Pyrosequencing primers.

**Supplementary table 5.** RT-PCR primers.

**Supplementary figure 1.** Differences in DNA methylation distribution of chromosomes 1, 6, 11 and 8 between AML (non-trisomic), NBM and trisomy 8 AML. Chromosomes 1, 6, 8 and 11 showed similar DNA methylation distribution between AML (non-trisomic) (green line) and NBM (red line) with a lower frequency of Bamber scores >0.8 in AML than NBM. By contrast, trisomy 8 AML (blue line) showed a higher frequency of Bamber scores >0.8 in chromosome 8 and a higher frequency of Bamber scores <0.4 in chromosomes 1, 6, 11 than NBM and AML (non-trisomic).

**Supplementary figure 2.** Correlation between the intragenic CGI methylation and HHEX expression. (A-B) X-axis represents the percentage CpG methylation as calculated by pyrosequencing; Y-axis represents the relative expression of HHEX gene. There was a moderate but significant inverse correlation between CGI methylation (CpG 0-16) and HHEX gene expression; HHEX1 (across exon 1 and exon 2) and HHEX2 (across exon 3 and exon 4) (Spearman $r = -0.5$, $P = 0.006$, Spearman $r = -0.43$, $P = 0.02$ respectively). The red dots represent trisomy 8 AML patients with HHEX gene methylation >60%.

**REFERENCES**

1. Huret, JL. +8 or trisomy 8 Atlas Genet Cytogenet Oncol Haematol. 2009;13(3):75–77.

2. Paulsson K, Johansson B. Trisomy 8 as the sole chromosomal aberration in acute myeloid leukemia and myelodysplastic syndrome. Pathol Biol (Paris). 2007;55(1):37–48.

3. Virtanen K, Wright FA, Tanner SM, et al. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci U S A. 2001;98(3):1124–1129.

4. Nawata H, Kashino G, Tano K, et al. Dysregulation of gene expression in the artificial human trisomy cells of chromosome 8 associated with transformed cell phenotypes. PLoS One. 2011;6(9):e25519.

5. Saied MH, Marzec J, Khalid S, et al. Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. PLoS One. 2012;7(3):e33213.

6. Weber M, Davies JJ, Wirtz D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005;37(8):853–862.

7. Holt KE, Teo YY, Li H, et al. Detecting SNPs and estimating allele frequencies in clonal bacterial populations by sequencing pooled DNA. Bioinformatics. 2009;25(16):2074–2075.

8. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

9. Down TA, Rakyta VK, Turner DJ, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methlylation analysis. Nat Biotechnol. 2008;26(7):779–785.

10. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

11. Kroeger H, Jelinek J, Estocio MR, et al. Aberrant CpG island methylation in acute myeloid leukemia is accentuated at relapse. Blood. 2008;112(4):1366–1373.

12. Wilop S, Fernandez AF, Jor E, et al. Array-based DNA methylation profiling in acute myeloid leukemia. Br J Haematol. 2011;155(1):65–72.

13. Schwable J, Choudhury C, Thiede C, et al. RGS9 is an important target gene of Flt3-ITD mutations in AML and functions in myeloid differentiation and leukemic transformation. Blood. 2005;105(5):2107–2114.

14. Romar-M Gomez J, Jimenez-Velasco A, Agire X, et al. Promoter hypomethylation and global hypomethylation are independent epigenetic events in lymphoid leukemogenesis with opposing effects on clinical outcome. Leukemia. 2006;20(8):1445–1448.

15. Verma D, Katarzian H, Faderl S, et al. Late relapses in acute myeloid leukemia: analysis of characteristics and outcome. Leuk Lymphoma. 2010;51(5):779–782.

16. Topisirovic I, Guzman ML, McConnell MJ, et al. Aberrant eukaryotic transcription initiation factor 4E-dependent mRNA transport impedes hematopoietic differentiation and contributes to leukemogenesis. Mol Cell Biol. 2003;23(24):9992–9002.

17. Jankovic D, Gorello P, Liu T, et al. Leukemogenic mechanisms and targets of a NUP98/HHEX fusion in acute myeloid leukemia. Blood. 2008;111(12):5672–5682.

18. Brickman JM, Jones CM, Clements M, Smith JC, Beddington RS. Hex is a transcriptional repressor that contributes to anterior identity and suppresses Speckleless organizer function. Development. 2000;127(11):2301–2315.

19. Paz H, Lynch MR, Rogue CW, Gasson JC. The homeobox gene Hhex regulates the earliest stages of definitive hematopoiesis. Blood. 2010;116(8):1254–1262.

20. Bedford FK, Ashworth A, Enver T, Wiedemann LM. HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human. Nucleic Acids Res. 1993;21(5):1249–1249.

21. Mack DL, Leibowitz DS, Cooper S, Ramsey H, Broxmeyer HE, Hromas R. Down-regulation of the myeloid homeobox protein Hex is essential for normal T-cell development. Immunology. 2002;107(4):444–451.

22. Topica Z, Mack DL, Hromas RA, Borden KL. The pronucleolar leukemia protein PML interacts with the proline-rich homedomain protein PRH: a RING may link hematopoiesis and growth control. Oncogene. 1999;18(50):7091–7100.

23. Noy P, Williams H, Sawasdichai A, Gaston K, Jayaraman PS. PRH/Hhex controls cell survival through coordinate transcriptional regulation of vascular endothelial growth factor signaling. Mol Cell Biol. 2010;30(9):2120–2134.

24. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80.