

# New Possible Targetable Genes for Future Treatment of Mixed Lineage Leukemia

Senol Dogan\*

International Burch University, Sarajevo, Bosnia and Herzegovina

## Abstract

**Aim of study:** Leukemia has different subtypes, which present unique clinical and molecular characteristics. MLL (Mixed Lineage Leukemia) is one of the new different subtypes than AML and ALL.

**Materials and Methods:** Genomic characterization is the main key understanding the differences of MLL by analysis of differential gene expression, methylation patterns and mutational spectra that were compared and analyzed between MLL and AML types (n=197).

**Results:** According to the genomic characterization of MLL, differentially expressed 114 genes were selected and 37 of them targeted genes having more than 2 fold expression change, including *HOXA9*, *CFH*, *DDX4*, *MSH4*, *MSMB*, *TWIST1*, *ZSWIM2*, *POU6F2*. To measure the aberrant methylation is the second genomic characterization of this research because the rearrangements of MLL gene leading to aberrant methylation. The methylation data were compared between cancer and control, so high methylated genes have been detected between MLL and AML types. The methylation loci were categorized into two groups:  $\geq 10$  fold difference and  $\geq 5$  and  $\leq 10$  fold difference. Some of the genes high methylated more than one location such as; *RAET1E*, *HSD17B2*, *RNASE11*, *DGK1*, *POU6F2*, *NAGS*, *PIK3C2G*, *GADL1*, and *KRT13*. In addition to that, analysis of somatic mutation gives us that CFH has the highest point mutation 9,92%.

**Conclusion:** Overall, the MLL genomic characterization shows that it is different than AML and exhibits a unique molecular and biological phenotype and point to new possible targetable genes for future treatment of MLL leukemia are two important values.

**Keywords:** MLL type leukemia; Genomic analyzes; Gene expression; Methylation, Mutation

## Introduction

AML (Acute myeloid leukemia) and ALL (Lymphoid myeloid leukemia) subtypes are categorized and classified by different prognosis, diagnosis, treatment, survival rate and even different types of blood cancer [1,2]. MLL type is one of the subtypes of leukemia, which demonstrates aggressive characteristics. The rearrangement of chromosomes, including locus (11q23), results in MLL, can be found in AML and ALL that has shown some genomic abnormalities such as; chromosomal rearrangements, abnormal gene expression patterns, methylation status, mutational spectrum and microRNA studies [3-7]. The MLL gene is also properties of cancer and responsible for epigenetic regulation of histone H3 lys4 protein which has a significant impact on embryonic development and hematopoiesis [8]. The gene is also accompanied by MLL partner proteins which form a protein complex which can alter epigenetic profiles [9] and activate preleukemic target genes which are key for leukemogenesis, including *HOXA9*, *MEIS1* [10]. Although pediatric and adult AML patients show common properties, the differences have not yet been identified between MLL and AML [11]. Leukemia genomic abnormalities arise cytogenetic problems, which need different therapy and methods. Therefore, understanding genomic activities of MLL will let us make correct and precise therapy. For example, gene expression profile demonstrates a novel profile that is different than AML and ALL patients [12]. Novel methylation in the promoter region and histone protein change the epigenetics of MLL type leukemia and involve with abnormal gene expression signature [13,14]. MLL gene with partner fusion proteins binds DOT1L, which is a key gene to understand the methylation status of leukemia, methylates H3 at lysine 79 and initiates the chromatin modification during transcriptional elongation [15,16]. It is necessary to show the genomic difference and responsible genes of MLL than

AML. Genomic characterization of MLL type leukemia will help us to distinguish the differences between the cancer types by analyzing gene expression, methylation and mutational data. Therefore, the following research identifies possible target genes for future leukemia therapies.

## Materials

### Genomic and bioinformatics data materials

The genomic data were analyzed for gene expression, methylation and mutational perspectives present the differences of MLL and find new therapeutic target genes. Although TCGA (The Cancer Genome Atlas) was the main database, other databases such as; Cosmic, Genecards and Cancer Cell Lines, also were used to select target genes by analyzing different computer programs. According to the TCGA gene expression and clinical data, the ID of each patient was used to compare genomic data, which consist of gene expression, methylation, mutation and the clinical data simultaneously (Figure 1).

TCGA portal is the main source of data for the research which compare the AML and MLL type leukemia to find the differences. The steps of each data process are explained one by one in Figure 1A. For each step different computer programs have been used. The gene expression, methylation, mutational and clinical data are investigated to understand

\*Corresponding author: Senol Dogan, International Burch University, Sarajevo, Bosnia and Herzegovina, Tel: +387 33 944 400; E-mail: [senol1dogan3@gmail.com](mailto:senol1dogan3@gmail.com)

Received January 17, 2017; Accepted May 27, 2017; Published May 31, 2017

Citation: Dogan S (2017) New Possible Targetable Genes for Future Treatment of Mixed Lineage Leukemia. J Biom Biostat 8: 349. doi: [10.4172/2155-6180.1000349](https://doi.org/10.4172/2155-6180.1000349)

Copyright: © 2017 Dogan S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

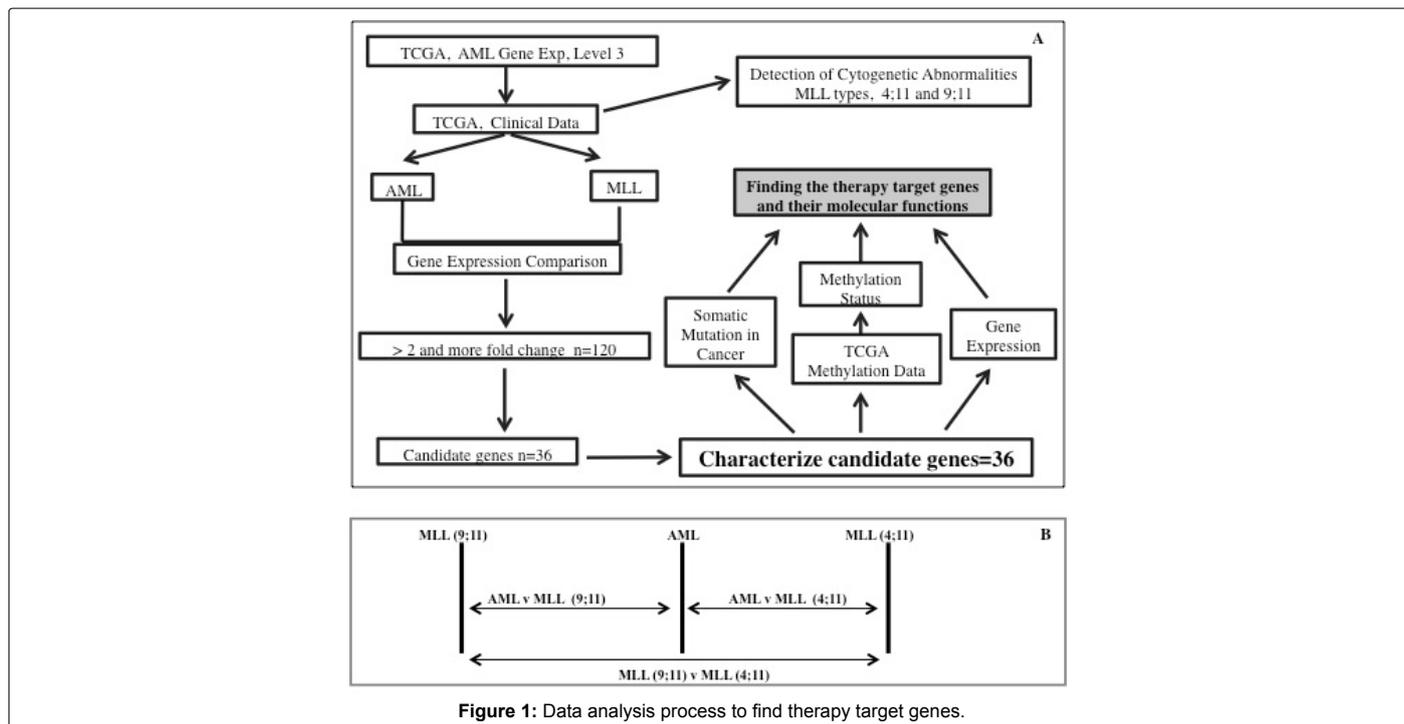


Figure 1: Data analysis process to find therapy target genes.

the genomic differences between AML and MLL. Although 120 genes are extracted depending on  $\geq 2$  fold gene expression abnormalities only 36 tumorigenesis related genes are selected as a candidate. To find the genomic characterization of MLL the candidate genes are elaborated from a different perspective, methylation, mutation and other cell lines gene expression. According to the data portal, two types MLL (4;11), (9;11) and AML, were compared for the research in Figure 1B.

### Gene expression data

The gene expression value is presented by Affymetrix U133+2 arrays platform, normalized log2 and downloaded as Level 3 type. 19800 different genes of 197 patients expression were analyzed depending on their TCGA ID's, which have been used to select the AML and MLL patients. Analysis of gene expression algorithms are helpful to select the target genes [14].

### DNA methylation data

The TCGA Illumina Human Methylation 450K platform was used to measure the AML methylation profile, which provides 194 patients and 199 control methylation data. The data consist of gene names, chromosome number, location and  $\beta$  value, by which methylation value was calculated and presented. If the value is closer to 1 it means hyper-methylated location, but if it is closer to 0 it means hypo-methylated location. Thus the hyper and hypo value allows us to see changes in the methylation profile between AML patients and control.

### Mutational data

Cosmic (The Catalogue of Somatic Mutation in Cancer) was the main mutation database in which any gene can be applied to find its mutational rate in different tissues (<http://cancer.sanger.ac.uk/cosmic>). The result of a gene mutation comes out as Histogram, Mutations, Fusions, Tissue, Distribution and CNV/Expr. details. In addition to that, the mutational spectrum shows the types of mutation, such as Missense, Non-sense, Frameshift, etc.

## Methods

### Gene expression comparison

The TCGA clinical data firstly was analyzed to select AML and MLL patients by their gene expression data using the same ID. The gene expression data were separated into 3 groups MLL (4;11), MLL (9;11) and AML (Figure 1). R (Statistical program) selected 197 patients and compared the different gene expression profiles of 19800 genes.  $\geq 2$  fold differentially expressed genes were detected between MLL and AML. Although there were 114 genes abnormally expressed, only 37 of them were selected as target genes because of their relation with tumorigenesis.

### Methylation data comparison

The methylation changes 194 AML and MLL patients were compared with 199 control. The  $\beta$  value of the comparison gives us the methylation fold change, which compares the high and the low methylation status of AML, MLL and control.

### Mutational data

The 37 target genes were applied to the COSMIC database to search the point mutation and copy number variation (Gain) and (Loss) of the genes. Each gene mutational rate is found in haematopoietic and lymphoid tissue level.

### Computer programs and bioinformatics tools

R program (<https://www.r-project.org>) was used to categorize the leukemia types, MLL (4;11), MLL (9;11) and AML to compare the gene expression and select the target genes depending on clinical and genomic data. The program selects abnormally expressed genes, which show  $\geq 2$  fold changes by observing 19800 genes. The target genes and their expression values have been applied to the HCE 3.5 tool (<http://www.cs.umd.edu/hcil/hce/>) to cluster and find the correlation of the genes.

## Result

### Selecting target genes

After the gene expression comparison has been completed 114 abnormally expressed genes were selected. 37 of the genes were selected as target genes because of their involvement with tumorigenesis among MLL (9;11), MLL (4;11) and AML (Figure 2).

The gene expression profile has been analyzed in two ways, common fold and high-low expressed genes number. The genes have been categorized depending on their fold expressed and separated as common fold genes from highest expressed to the lowest expressed genes in an order Figure 2A. The common fold genes as named as red the highest expressed, orange-yellow slightly high expressed and blue low expressed genes. According to the fold intensity, the number of genes and their intersection are shown in Figure 2A from highest fold to the lowest fold genes. The fold genes have been compared between AML and MLL types (4;11) and (9;11). The red color shows the highest expressed gene number and their fold interval depending on leukemia type comparison. Orange and yellow are slightly high expressed gene numbers. The black shows the genes number which expression does not change much, almost 7000. Turquoise and blue colors show the lowest expressed gene number Figure 2B. The highest gene expression fold and gene number belong to MLL (4;11) v MLL (9;11) comparison, 1033 genes expression changes 224-5.0 interval and 1213 genes 3-5 fold interval Figure 2B. Interestingly, the lowest expressed genes number shows almost similar changing around 5000 genes 0,005-0,7 interval fold. If we look at the sum of the gene number, the lowest expressed genes number are bigger than lowest expressed genes number.

### Finding high and low expressed genes

The gene expression profile has been analyzed in two ways, common fold and high-low expressed genes number. The genes have been categorized depending on their fold expressed and separated as common fold genes from highest expressed to the lowest expressed genes in an order. The common fold genes as named as red the highest expressed, orange-yellow slightly high expressed and blue low expressed genes. While red color stands for the highest (increase) fold

change, blue color represents the lowest (decrease) fold change. The first 4 rows, red, orange, yellow and black, present the number of high expressed genes in decreasing order, but the last two rows, turquoise and blue, show the number of low expressed genes (Figure 2A). The expression of 176 genes changes from 5 to 23.1 fold in AML v MLL (4;11), 1033 genes 5 to 224 fold in MLL (4;11) v MLL (9;11) and 82 genes 5 to 42.9 fold in AML v MLL (9;11). On the other hand, the expression of about 5000 different genes has changed from 0.0001 to 0.7 in all comparison. Turquoise and blue colors show the lowest expressed gene number (Figure 2B). In addition to that, some of the gene expressions were slightly changed and shown in yellow, black and turquoise colors. The black shows the genes number which expression does not change much, almost 7000.

### Common and different genes activity in MLL (4;11), MLL (9;11) and AML

The comparison between MLL (4;11), MLL (9;11) and AML helps us to find common and distinct expressed genes names and numbers, too. Therefore, the Venn diagrams present the highly expressed genes in red, slightly over expressed genes in orange and yellow, and the low expressed genes in blue colors (Figure 2A). The intersection numbers show unique and common expressed genes among the comparison. While the highest expressed genes are more unique, the low expressed genes are more common than others. The intersection gene numbers increase from A to D, 149, 197, 500 and 9369 (Figure 2A). The result gives that common gene of each group increase from highest to lowest expressed genes.

### Selection of the 2 and more fold changed genes

The selected target genes were clustered and their correlation was found by the HCE 3.5 program, which correlates genes as 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> group (Figure 3A). The genes' expression was calculated by Pearson's method depending on their highest correlation. The 1<sup>st</sup> group represents the high and low expressed together, the 2<sup>nd</sup> group represents low expressed genes and the 3<sup>rd</sup> group represents the highest expressed strongly correlated genes. Among the target genes, *MSMB*, *TWIST1*, *ZSWIM2* show the highest correlation with high gene expression.

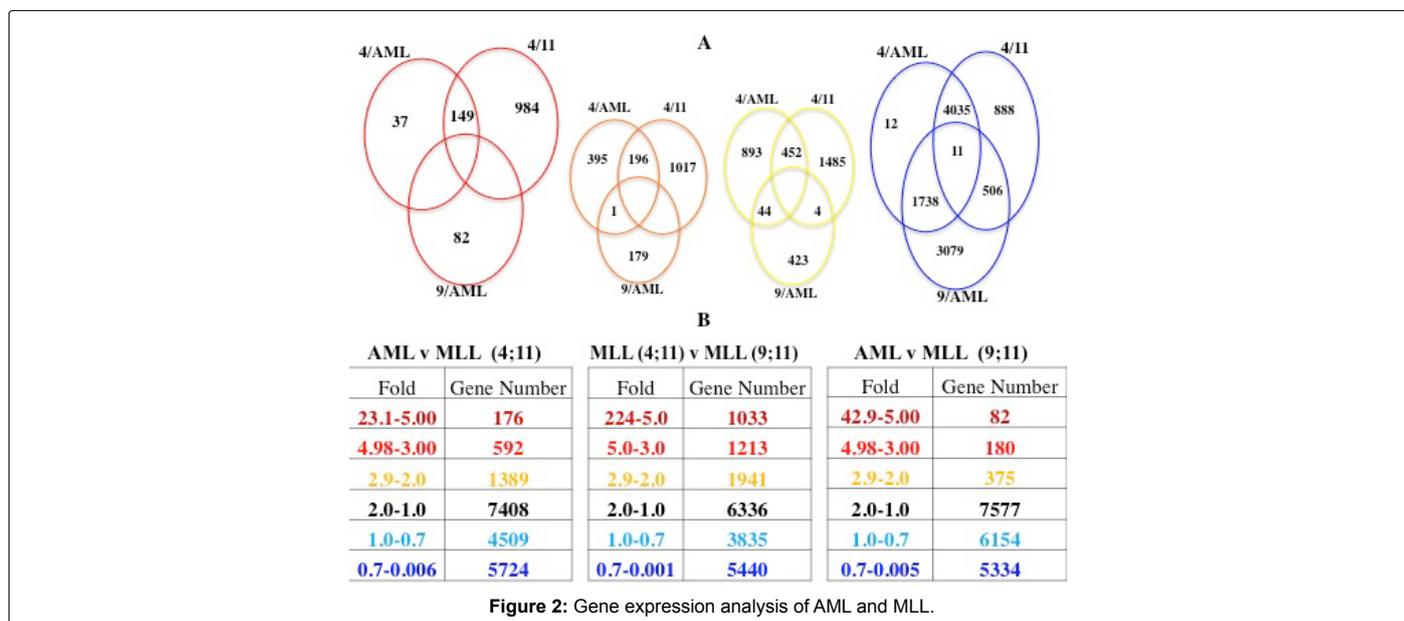


Figure 2: Gene expression analysis of AML and MLL.

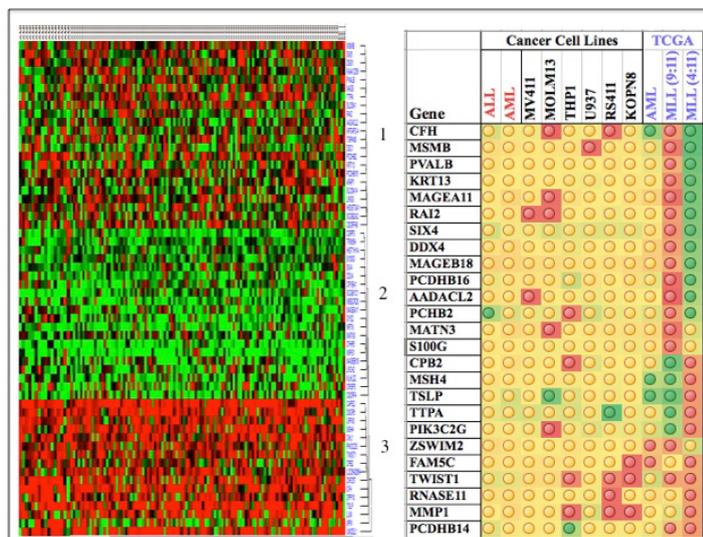


Figure 3: Analysis of gene expression.

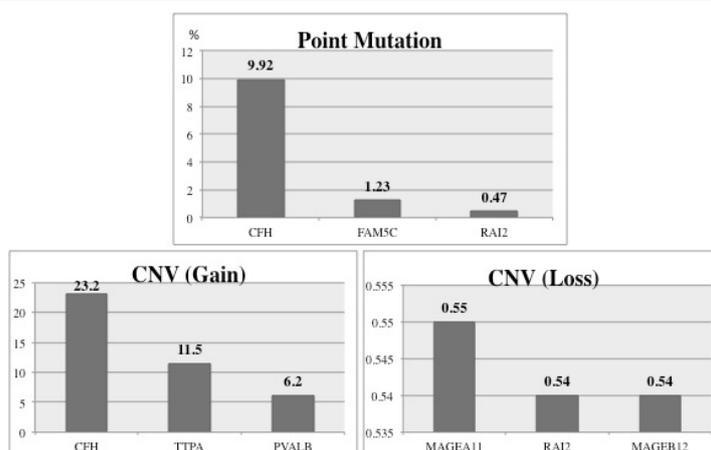


Figure 4: Mutational profile of the target genes.

### Other cancer cell lines and TCGA expression comparison

The target gene expression was compared with AML, ALL, and other cell lines, which include *MV41*, *MOLM13*, *THP1*, *U937*, *RS411* and *KOPN8* (Figure 3B). It is clearly seen that the target genes expression of MLL (4;11) and (9;11) are different in TCGA, comparison of other cell lines. Therefore, this is enough to say MLL type gene expressions are different than AML and ALL. After applying the genes into the other cell lines, *CFH*, *MSMB*, *PVALB*, *KRT13*, *MAGEA11*, *RAI2*, *SIX4*, *DDX4*, *MAGEB18*, *PCDHB16*, *AADAACL2*, *PCHB2*, *MATN3*, *S100G*, *C8orf4*, *CPB2*, *MSH4*, *TSLP*, *TTPA*, *PIK3C2G*, *ZSWIM2* genes appear much differently expressed in TCGA data than the other cell lines.

### Metabolic functions of the target genes

To understand the target gene function in a cell they all were applied to the web tool Panther, <http://pantherdb.org>, to discover their molecular functions, biological process, pathway, protein class and cellular component (Figure 4). The responsibility of the genes was detected in that order, 30% binding, 24.2% catalytic activity, 19.7% receptor activity, 7.6% transcription factor and transporter activity,

40.9% metabolic process, 39.4% developmental, 16.7% immune system process, 15.2% biological adhesion. Almost all pathways are affected by the same percentage, except WNT and Cadherin signaling pathway.

### Somatic mutation of target genes

After completing the whole gene expression analysis, the somatic mutation analysis of the target genes was detected. The genes were applied to COSMIC, <http://cancer.sanger.ac.uk/cosmic>, which is a somatic mutation database. The point mutation, copy number variation (Gain), and copy number variation (Loss) have been detected respectively for each target gene. As a result of somatic mutation rate *CFH* shows the maximum point mutation percentage, 9.92%. The gene is responsible for the regulation of complement activation and related to the defense mechanism to microbial infections, and is very closely related to blood functions and may have a potential risk of leukemia. *FAM5C* and *RAI2A* are other genes having high point mutation genes, 1.23% and 0.47%, respectively.

### Copy number variation (gain) and (loss)

Another mutational search is copy number variation (Gain) and

(Loss), which presents the profile of chromosomal activity of the target genes. *CFH*, *TTPA*, *PVALB* and *MMP1* are the highest CNVG genes 23.20%, 1.50%, 6.20% and 4.70% respectively. The highest CNVL genes are *RAI2*, *MAGEB12*, *MAGEB18* and *GAGE1* with almost same percentage 0.55%.

### Higher methylation of target genes and their location

The last genomic characterization of the target genes is methylation profile analysis. Acute myeloid leukemia methylation and control methylation (Level 3) data were compared to find the methylation profile of target genes by separating into two groups;  $\geq 10$  and  $\leq 10$  fold change (Figure 5). The highest methylation genes are *RAET1E*, *HSD17B2*, *RNASE11*, *DGKI*, *POU6F2*, *NAGS*, *PIK3C2G*, *GADL1* and *KRT13* more than  $\geq 10$  fold change (Figure 5A). The maximum methylation fold was 22.9 fold and was observed at 150211211 location of *RAET1E* in the C nucleotide. The gene is related to the tumor-immune system by stimulating the expansion of anti-tumor cytotoxic lymphocytes. The other highest methylated target genes are *KRT13* and *GADL1* 18.4, 16.4 in more than one location, respectively. *RAET1E* has two locations with high methylation with 10.2 and 22.9 fold. *DGKI* and *POUF62* have 3 high methylated locations with 10.9, 11.2, 13.9 fold and 11.3, 11.5, 15.7 fold, respectively. The second methylated target genes range between 5 to 10 fold. 19 locations of 14 different genes demonstrate higher than 5 lower than the 10 fold and five genes of this group *PVALB*, *PCDHB16*, *HSD17B2*, *DGKI*, *TSLP* were highly methylated in more than one location (Figure 5B).

### The target gene methylation level and fold change

The most highly methylated target genes have been selected, comparing MLL to control. While some of the target genes have only one location with high fold methylation, some of them present more than one location with high fold methylation, such as; 2 or 3 different locations (Table 1). *NAGS* has the highest methylation with 28.59 on the 42084953 location.

### Methylation and gene expression

Due to the relation between gene expression and methylation, highly methylated or slightly highly methylated genes may be found in different locations, but they still directly affect gene expression, not just in leukemia also in other cancers, such as colon [17]. The target genes' expression and their heavy methylation data were compared to understand the relation between AML and MLL (Table 2). The comparison of the 9 genes with  $\geq 10$  fold high methylation reveals that the high methylation of the genes has resulted from abnormal gene expression.

### Discussion

MLL leukemia has unique clinical and molecular characteristics. Using the genomic profiling of patient samples with acute myeloid leukemia (AML), we analyzed the differential gene expression, methylation patterns, and mutational spectra between MLL and other AML types (n=197). The type of cancer is combination of many genomic abnormalities, such as MLL fusion proteins expression and all genomic data [17]. We found that 114 genes were differentially expressed, where 37 genes had a more than 2 fold expression difference including *HOXA9*, *CFH*, *DDX4*, *MSH4*, *MSMB*, *TWIST1*, *ZSWIM2*, *POU6F2* and others. Since we analyzed expression from patient samples, we further compared the candidate genes to common MLL cell lines. Our results show that most genes did not have the same expression pattern in patients and in corresponding MLL cell lines. The rearrangements of the MLL gene result in aberrant methylation, so the differential methylation patterns between MLL and other AML types were identified as methylation hotspots. The methylation loci were categorized into two groups: group one ( $>5$  and  $<10$  fold difference) and group two ( $\geq 10$  fold difference). In the second group, 14 locations for 9 genes were found including *RAET1E*, *HSD17B2*, *RNASE11*, *DGKI*, *POU6F2*, *NAGS*, *PIK3C2G*, *GADL1*, and *KRT13*. MLL leukemia exhibits a unique clinical and biological phenotype. We conducted the analysis of MLL genomic data according to different

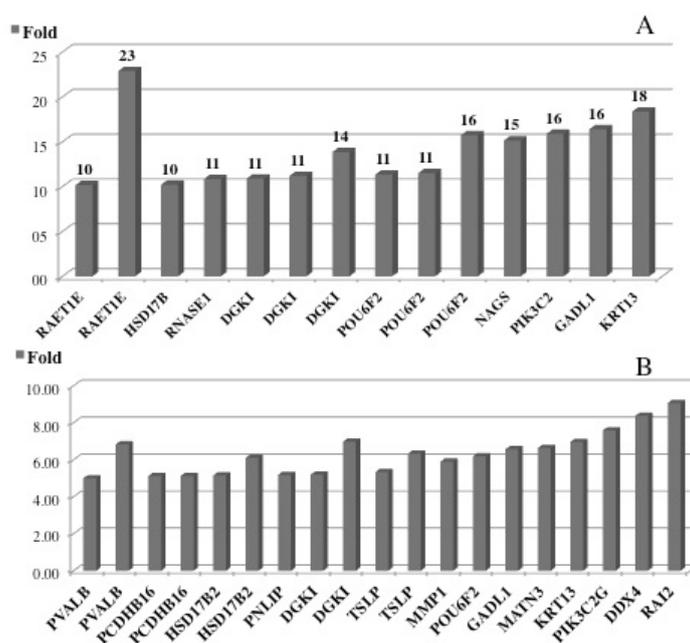


Figure 5: The highest methylated target genes.

Chromosome	Location	Gene	MLL/Control
17	42084953	NAGS	28.59
6	150209640	RAET1E	21.85
3	30812040	GADL1	16.21
17	39661855	KRT13	15.37
7	39350675	POU6F2	15.26
12	18584312	PIK3C2G	14.88
7	39332836	POU6F2	14.85
7	39044924	POU6F2	13.46
7	137367341	DGKI	12.26
7	137085181	DGKI	12.17
7	137410625	DGKI	11.42
7	39269105	POU6F2	11.38
5	140562007	PCDHB16	11.16
14	21053613	RNASE11	10.57

Table 1: The most methylated target genes and their location.

Gene Name	Fold Methylation		Gene Expression	
	MLL/Control	AML/Control	AML/Control	MLL/Control
DDX4	8.09	8.33	4,81	8,21
DGKI	12.22	10.88	6,31	3,97
FAM5C	2.69	2.67	4,26	5,59
GADL1	16,21	16,50	5,58	3,40
GAGE1	3,33	3,34	5,10	3,03
KRT13	6,89	6,96	5,69	8,83
MMP1	2,39	2,39	5,59	3,56
MSH4	3,94	3,95	7,13	3,54
PCDHB14	8,55	4,58	6,06	3,52
PCDHB16	11,16	4,96	6,18	2,31
PCDHB2	1,79	3,35	6,39	3,65
PIK3C2G	14,88	15,84	6,69	3,38
PNLIP	3,99	3,92	5,75	3,80
POU6F2	9,43	6,68	4,87	6,34
RAET1E	9,81	10,16	6,23	3,95
RNASE11	10,57	10,87	5,14	3,66
SIX4	3,81	3,32	5,26	8,58
SYDE2	2,52	1,79	6,03	3,70
TWIST1	0,17	0,56	6,85	3,72

Table 2: Target genes methylation and expression relation.

aspects, such as; gene expression, methylation, and mutation, in order to obtain a global look at the disease. As it is presented in the conclusion with tables and figures, some of the genes have different genomic activity simultaneously. The genes should be observed more than other target genes because their genomic profiles include more significant changes. If a gene has an abnormal expression, methylation, and mutational spectrum at the same time or at least two of them, it means it is a kind of therapeutic target gene. In the future, personalized medicine will be produced according to the patient gene expression and mutation. Our methodology and genomic data analyses may be useful for both classical and personalized medicine development. If we want to understand the genomic profile of MLL type leukemia, we should globally look at the genomic data and their relation. After our data analyses of MLL leukemia, it is clearly observed that the target genes present abnormal genomic profiles, which makes MLL the most aggressive example of AML subtypes. Therefore, our results point to new possible targetable genes for future treatment of MLL leukemia.

#### Acknowledgements

The paper has been awarded by EACR (European Association of Cancer) in NCRI, Liverpool, UK. We are thankful the all commission to select the work for

award. In addition to that, we are thankful for International Burch University and Bosna Sema Education Institutions.

#### References

- Armstrong (2002) MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 30: 41-47.
- Armstrong SA (2005) Molecular Genetics of Acute Lymphoblastic Leukemia. *J Clin Oncol* 23: 6306-6315.
- Slany RK (2009) The molecular biology of mixed lineage leukemia. *Haematologica* 94: 984-993.
- McCabe N, Van Der Poel SZ, Gill H, Espinosa R (1991) Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias.
- Streubel B, Valent P, Jäger U, Edelhäuser M, Wandt H, et al. (2000) Amplification of the MLL gene on double minutes, a homogeneously staining region, and ring chromosomes in five patients with acute myeloid leukemia or myelodysplastic syndrome. *Genes Chromosomes Cancer* 27: 380-386.
- Dogan S, Kurtovic-Kozaric A (2015) Changes of Molecular, Cellular and Biological Activities According to microRNA-mRNA Interactions in Ovarian Cancer. *Computational Molecular Biology*.
- Hess JL, Benjamin DY, Li B, Hanson R, Korsmeyer SJ. (1997) Defects in yolk sac hematopoiesis in Mll-null embryos. *Blood* 90: 1799-1806.
- Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ (2001) MLL and CREB Bind Cooperatively to the Nuclear Coactivator CREB-Binding Protein. *Mol Cell Biol* 21: 2249-2258.
- Muntean AG (2013) Mechanisms of mixed-lineage leukemia. *Int J Hematol Oncol* 2: 207-217.
- Creutzig U (2008) Significance of age in acute myeloid leukemia patients younger than 30 years: A common analysis of the pediatric trials AML-BFM 93/98 and the adult trials AMLCG 92/99 and AMLSG HD93/98A. *Cancer* 112: 562-571.
- Wakui M (2008) Diagnosis of acute myeloid leukemia according to the WHO classification in the Japan Adult Leukemia Study Group AML-97 protocol. *Int J Hematol* 87: 144-151.
- Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. *Carcinogenesis* 31: 27-36.
- Dogan S, Kurtovic-Kozaric A, Karli G (2015) The Detection of Extremely High and Low Expressed Genes by EGEF Algorithm in Invasive Breast Cancer. *J Biom Biostat* 7: 312.
- Guenther MG, Lawton LN, Rozovskaia T, Frampton GM, Levine SS, et al. (2008) Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev* 22: 3403-3408.
- Krivtsov AV, Feng Z, Lemieux ME, Faber J, Vempati S, et al. (2008) H3K79 Methylation Profiles Define Murine and Human MLL-AF4 Leukemias. *Cancer Cell* 14: 355-368.
- Dogan S, Kurtovic-Kozaric A, Hajrovic A, Lisic M, Gokgoz E (2015) Comparison of MLL Fusion Genes Expression among the Cytogenetics Abnormalities of Acute Myeloid Leukemia and Their Clinical Effects. *J Biom Biostat* 7: 312.
- Dogan S, Cilic A, Kurtovic-Kozaric A, Ozturk F (2015) Detection of G-type density in promoter sequence of colon cancer oncogenes and tumor suppressor genes. *Bioinformatics* 11: 290-295.