

## Association of Leukemia Target Genes *Tet2*, *Bcl2*, and *Slc23a2* in Vitamin C Pathways

JING LI<sup>1\*</sup>, WENQI ZHANG<sup>1,2\*</sup>, WEIDONG LIU<sup>1,3</sup>, JARRETT RONG<sup>1,4</sup>,  
YUHAN CHEN<sup>5</sup>, WEIKUAN GU<sup>1,6</sup> and WEI ZHANG<sup>1,7</sup>

<sup>1</sup>Department of Orthopedic Surgery and BME-Campbell Clinic,  
University of Tennessee Health Science Center, Memphis, TN, U.S.A.;

<sup>2</sup>College of Basic Medicine, Hebei Medical University, Shijiazhuang, P.R. China;

<sup>3</sup>The First Hospital of Qiqihar City, Heilongjiang, P.R. China;

<sup>4</sup>BSA Biology, Health Science Scholars Honors Program, the University of Texas at Austin, Austin, TX, U.S.A.;

<sup>5</sup>Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, P.R. China;

<sup>6</sup>Research Service, Memphis VA Medical Center, Memphis, TN, U.S.A.;

<sup>7</sup>Department of Pharmacology, School of Basic Medicine,  
Hebei University of Chinese Medicine, Shijiazhuang, P.R. China

**Abstract.** *Background:* Vitamin C has been used in combination with several target genes in the treatment of leukemia. *Tet* methylcytosine dioxygenase (*Tet2*), *B*-cell lymphoma 2 (*Bcl2*), and solute carrier family 23 member 2 (*Slc23a2*) are the major target genes in the treatment of leukemia and are relevant to vitamin C. *Materials and Methods:* Using whole-genome expression profiles from mouse livers, the expression quantitative trait locus (*eQTL*), correlation matrix, and gene network graph were constructed with probes from each of these three genes and with their relative genes. The function of key genes was examined by their pathways and reported information. The results indicated that although direct correlations among their expression levels were not strong, alternative connecting pathways were discovered. By comparing the expression levels of one probe with known sequences from each of the three genes, we identified several key genes, induced myeloid leukemia cell differentiation protein (*Mcl1*), far upstream element-binding protein 1 (*Fubp1*), and

tumor protein D52-like 2 (*Tpd52l2*), which play important roles in acute lymphocytic leukemia and acute myelocytic leukemia. In conclusion, Alternative pathways and key genes that connect *Tet2*, *Bcl2*, and *Slc23a2* for their therapeutic applications with vitamin C were identified.

Vitamin C, combined with other treatments, has positive effects on various diseases. It was recently reported that risk of several types of cancer has been linked with vitamin C deficiency when the plasma concentration is lower than 11.4  $\mu$ M (1), such as pancreatic cancer, ovarian cancer, prostate cancer, colon cancer, mesothelioma, breast cancer, and neuroblastoma (2). On the contrary, if intravenous ascorbate administration goes up to pharmacological concentrations (0.3-20 mM), it selectively kills cancer cells while not harming other normal cells (3, 4). Three genes, tet methylcytosine dioxygenase 2 (*TET2*), *B*-cell lymphoma 2 (*BCL2*), and solute carrier family 23 member 2 (*SLC23A2*) have been recently reported as being important in the vitamin C treatment of acute myeloid leukemia (AML) (3).

Several research studies have demonstrated that the cofactors of vitamin C in carcinogenesis, which include *TET2* (3) and *BCL2* (3, 5) are involved in several cancer types and functionally activate multiple enzymes. In this study, we focused on *TET2*, *BCL2*, and *SLC23A2*, which are located at different genetic loci and play different roles in jointly regulating the function of vitamin C (2).

*BCL2* inhibits the activation of *BCL2*-associated X (*BAX*) and *BCL2* homologous antagonist/killer *BAK*, the transcription factors of which are up-regulated by antioxidant vitamins (5). Recent studies have reported that *BCL2* plays roles in causing cell death by way of hindering the release of mitochondrial cytochrome *c*, impeding caspase activity,

This article is freely accessible online.

\*These Authors contributed equally to this study.

*Correspondence to:* Weikuan Gu, A310A Coleman Building, 956 Court Avenue, Memphis, TN 38163, U.S.A. Tel: +1 9014482259, Fax: +1 9014486062, e-mail: wgu@uthsc.edu; Wei Zhang, 3 Xingyuan Road, Luquan area, Shijiazhuang, Hebei 050200, PR China. Tel: +86 31189926226, Fax: +86 31189252651, e-mail: zhangwei666cn@foxmail.com

*Key Words:* Vitamin C, leukemia, *eQTL* mapping, key genes, alternative pathways.

and controlling mitochondrial membrane permeability (5-7). Antiapoptotic proteins including *BCL2* are inhibited by *miR-15a/16*, and their deletion can induce tumor B-cell survival *in vitro* and can also lead to chronic lymphocytic leukemia (CLL) development in mouse models (4).

Several types of cancer are related to genetic variation and mutations of *SLC23A1* and *SLC23A2* encoding sodium-ascorbate co-transporters 1 and 2 (SVCT1 and SVCT2), respectively (1). *SLC23A2* is closely associated with gastric cancer, lymphoma, and head and neck squamous cell carcinomas. As a key protein, SVCT2 is more interactive with vitamin C uptake than SVCT1 in the liver (2). *SLC23A2* should be expressed more highly in hematopoietic stem cells than in more differentiated cells (6). It has been reported that vitamin C influx into cells *via* SVCT2 increases levels of intracellular reactive oxygen species, which subsequently induces DNA damage and ATP depletion in hepatocellular carcinoma, leading to cell death partially *via* cell-cycle arrest and caspase-dependent apoptosis (2).

*TET* (ten-eleven translocation) has three types of enzymes: TET1, TET2 and TET3. *TET2* is the most expressed *TET* gene in hematopoietic tissue, especially in hematopoietic stem cells (7, 8). These proteins can promote post-transcriptional modifications of histones by recruiting the *O*-linked  $\beta$ -D-N-acetylglucosamine transfer enzyme to chromatin. The conversion of 5-methyl-cytosine (5-mC) to 5-hydroxymethyl-cytosine (5-hmC) is catalyzed by conserved dioxygenases of *TET2* (3). *TET2* is essential for preventing DNA demethylation from being activated in human monocytes. According to our research, single-cell-derived colony genotyping identified *TET2* mutation as one of the earliest events in the accumulation of genetic alternations that lead to leukemic clone expansion (6).

The function of these genes is highly correlated with vitamin C in the treatment of leukemia (7-9). However, the molecular pathway connections amongst them are still not completely understood. In this study, we took advantage of whole-genome expression profiles in mouse liver of recombinant inbred strains and analyzed the association of expression levels of these genes and key genes in their potential function connections.

## Materials and Methods

**Data source.** The study was carried out using multiple strategies and several steps. The target genes were identified from the whole-genome expression data of mouse liver located on the GeneNetwork website (<http://www.genenetwork.org/webqtl/main.py>). This dataset was produced by using liver mRNA from BXD recombinant inbred strains (9). Microarray analysis was carried out with Affy M430 2.0 (Santa Clara, CA, USA). The data were validated and normalized using standard procedures of Bennett *et al.* (9). These are the most optimal data from a well-developed recombinant inbred strain population. These mice were fed a chow diet for 8 weeks followed by a western diet for 16 weeks (9).

**Quantitative trait locus (QTL) mapping.** Transcriptomic expression QTL (eQTL), which regulates the expression level of each of the three genes, *Tet2*, *Bcl2*, and *Slc23a2*, were conducted using interval mapping platform at the GeneNetwork website (Memphis, TN, USA). The procedure follows that of our previous publication (10). Briefly, it includes three major steps. Firstly, all possible probes of each gene from the liver database were identified. Secondly, interval mapping was used to build transcriptome maps for each probe. Permutations of 1,000 tests were used to assess the strength and consistency of the linkages. Outlier data were excluded from the mapping. Finally, genomic regions and locations on chromosomes of the eQTL between and among the three genes were compared.

**Matrix comparison and pathway analysis among probes.** A comparison matrix was built using this function of GeneNetwork (11). The expression data of probes of *Tet2*, *Bcl2*, and *Slc23a2* were obtained by searching key words of gene names. All the data were then collected into a "BXD Trait collection" interactive sheet. The interaction matrix of all the collected probes was then produced by clicking the command "Matrix" for the selected probes. The matrix contains the correlation "r" of each probe to all the other probes. The data were converted into an Excel spreadsheet and Excel statistical functions (Microsoft, Redmond, WA, USA) were used to produce the data of correlations among these three sets of genes. Correlations then were analyzed to categorize the positive and negative associations among these genes and identify the most interesting genes from them.

The network graph was produced using the same set of data using the command "Graph". The network graph showed the most likely association, either positive or negative, among the probes for these three genes.

**Expanded matrix comparison and alternative indirect pathway analysis.** Additional known relative genes based on the search results from the Online Mendelian Inheritance in Man (OMIM) information at NCBI (<https://www.ncbi.nlm.nih.gov/>) were incorporated into the matrix for the construction of an expanded matrix. The three genes were sought by name using the OMIM database. The names of the reference genes and the gene used for the search were obtained. These genes then were selected as the list for further study. These genes were then put in the liver mRNA expression database for searching. The probes of these genes and the three genes themselves were next used for matrix and pathway analysis, performed in the same manner as for the initial analysis on the three genes. The matrices and network graph were examined for genes that significantly correlated with any of the probes.

**Exploration of additional pathways using probes with known sequences.** This comparison was carried out by using the expression levels of the top 100 correlated genes using one probe of known sequence from *Tet2*, *Bcl2*, and *Slc23a2*. For each gene, the top 100 genes whose expression levels most highly correlated with that of the expression level of each gene were identified. These probes were identified from the NCBI gene database (<https://www.ncbi.nlm.nih.gov/gene>) as: *Bcl2*: Probel1457687|PRIMERF Forward polymerase chain reaction (PCR) primer (outermost) (41b) TGTA AAAACGACGCCAGTGCTGAAGGGTATGTGATTGGCAG> Probel1457687|PRIMERR Reverse PCR primer (outermost) (42b) CAGGAAACAGCTATGACCCCCTATCACTTGAGCCAAGCACT; *Tet2*: Probel1455300|PRIMERF Forward PCR primer (outermost) (38b)

TGTAACGACGGCCAGTTGGGCATGTCGTAGTTGC  
 CT>Probel1455300|PRIMERR Reverse PCR primer (outermost) (42b)  
 CAGGAAACAGCTATGACCGTAATTGCTCATTTCAGTTGGCCTT;  
*Slc23a2*: Probel1442983|PRIMERF Forward PCR primer (outermost)  
 (38b) TGTAACGACGGCCAGTAAAGCGCTTCAGCACGCA  
 AC>Probel1442983|PRIMERR Reverse PCR primer (outermost) (38b)  
 CAGGAAACAGCTATGACCGGGCCCGACTTCAACTCCAA

For each of the top 100 genes that correlated with each probe, their data were used to build a correlative matrix with all other probes using Gene Network.

*Methods of comparison and statistical analysis.* Standard criteria were used to categorize the strength of the correlation as we reported previously (11). Thus, R values of  $-0.7$  or  $0.7$  or more were regarded as reflecting a strong positive or negative association; R values between  $0.35$  and  $0.69$  or between  $-0.35$  and  $-0.69$  were regarded as indicating present but not strong correlation; and R values between  $0$  and  $0.35$  or between  $0$  and  $-0.35$  were treated as not showing correlation (10). According to our study, the most interesting genes were selected according to their R values for correlations to all other relevant genes.

Several additional parameters were used in our study (11). The sum of R values or  $\Sigma R$  of the target genes represents the total of all R values.  $\Sigma R/n$  represents the average of R values for the selected gene (Re).  $\Sigma |R|$  represents the sum of all absolute values for R, while Rn is the net sum of every positive or negative R value for the probes of *Tet2*, *Bcl2*, and *Slc23a2*. The relative association level (Ra) between a target probe and the selected gene was calculated as  $\Sigma |R|/n$ , where  $|R|$  is the absolute value of R for a probe and the selected gene, and the variable  $n$  is the number of selected genes, which can range from 1 to 50.

## Results

*QTL mapping of probes of three genes.* From the liver database, the expression data from eight probes of *Bcl2*, four probes of *Slc23a2*, and one of *Tet2* were extracted. They were used for the eQTL mapping using the interval map method. From eight probes of *Bcl2*, seven produced loci that were at or above the suggestive likelihood threshold, which suggests that a potential eQTL exists. These seven probes produced a total of 13 eQTL loci on nine chromosomes, namely chromosomes 1, 2, 6, 7, 8, 10, 12, 15 and 16 (Figure 1). Among them, the eQTL on chromosome 1 was mapped by three probes, with one near the significant threshold. Three out of four probes of *Slc23a2* produced an eQTL above the suggestive threshold. A total of five eQTLs were mapped on five chromosomes, namely chromosomes 1, 2, 5, 8 and 18. The probe of *Tet2* produced two eQTLs at the suggestive level on chromosomes 2 and X. There was no overlapping eQTL among the probes of the three genes. Although probes from different genes mapped their eQTL on the same chromosomes, these were not on the same genomic regions. Three probes from *Bcl2* and one probe from *Slc23a2* mapped their eQTL on chromosome 1. The eQTL location of all three probes of *Bcl2* was located approximately between 100 and 130 Mb, while the eQTL of probe 1442983

of *Slc23a2* was located approximately between 150 and 165 Mb. One probe from *Bcl2* (probe 1427818) and *Slc23a2* (probe 1450404) and the probe of *Tet2* mapped their eQTL on chromosome 2. However, the eQTL of *Tet2* was mapped to the genomic region approximately between 130 Mb and 145 Mb. *Slc23a2* was mapped by eQTL to approximately between 5 Mb and 15 Mb and *Bcl2* to between 25 Mb and 45 Mb.

### *Matrix comparison and pathway analysis among probes.*

Matrix comparison was constructed using all of the probes from the three genes (Table I) (Figure 2A). The comparison shows that probes from *Tet2* and one of the probes from *Slc23a2* (1442983) were positively correlated, with an R value of  $0.576$  (Figure 2B). *Tet2* did not show any significant correlation with probes of *Bcl2*. The relation between *Bcl2* and *Slc23a2* is complex. While most of the probes of *Slc23a2* showed negative correlation (144383 vs. 1421921 in Figure 2C as example), one probe of *Slc23a2* (1417330) was negatively correlated to one *Bcl2* probe (1437122) with  $R=-0.570$  (Figure 2D), but it was positively correlated to the other two probes, with probe 1426054 with  $R=0.522$  (Figure 2E) and with 1422938  $R=0.514$  (Figure 2F). These correlations were at suggestive but not significant levels. However, these analyses explored all direct connections among these three genes. Alternatively, there is a possibility that *Tet2*, *Bcl2*, and *Slc23a2* are indirectly connected with each other.

### *Expanded matrix comparison and alternative indirect pathway analysis.*

In order to explore the alternative pathways among these three genes, the second cycle of matrix comparison was conducted with additional known relative genes based on the OMIM information at NCBI. These relevant genes include BCL2-associated athanogene 1 (*Bag1*), *Bag2*, *Bag3*, *Bag4*, *Bag5*, BCL2-modifying factor (*Bmf*), Bcl2-like 1 (*Bcl2l1*), *Bcl2l2*, *Bcl2l1l1*, *Bcl2l2*, *Bcl2l1l3*, *Bcl2l1l4*, myeloid cell leukemia sequence 1 (*Mcl1*), BCL2-associated X protein (*Bax*), BCL2/adenovirus E1B-interacting protein 2 (*Bclaf1*), beclin-1 (*Becn1*), BCL2-interacting killer-like (*Biklk*), CXXC finger 4 (*Cxxc4*), estrogen related receptor, beta (*Esrrb*), myelodysplasia syndrome 1 homolog (human) (*Mds1*), *Tet1*, *Tet3*, O-linked N-acetylglucosamine (*GlcNAc*), transferase (*Ogt*), Nanog homeobox (*Nanog*), and gulonolactone (L-) oxidase (*Gulo*).

The network construction produced two indirect pathways through which the expression level of *Bcl2* were negatively associated to that of *Slc23a2* (Figure 3A). The strength of correlation is represented by lines of different color that connect two genes in Figures 3 and 4. R Values between the two genes that are higher than  $0.7$  are displayed in red lines; those between  $0.5$  and  $0.7$ , in orange lines; between  $-0.5$  and  $-0.7$ , in green lines, and less than  $-0.7$  (between  $-0.7$  and  $-1.0$ ) in blue lines.

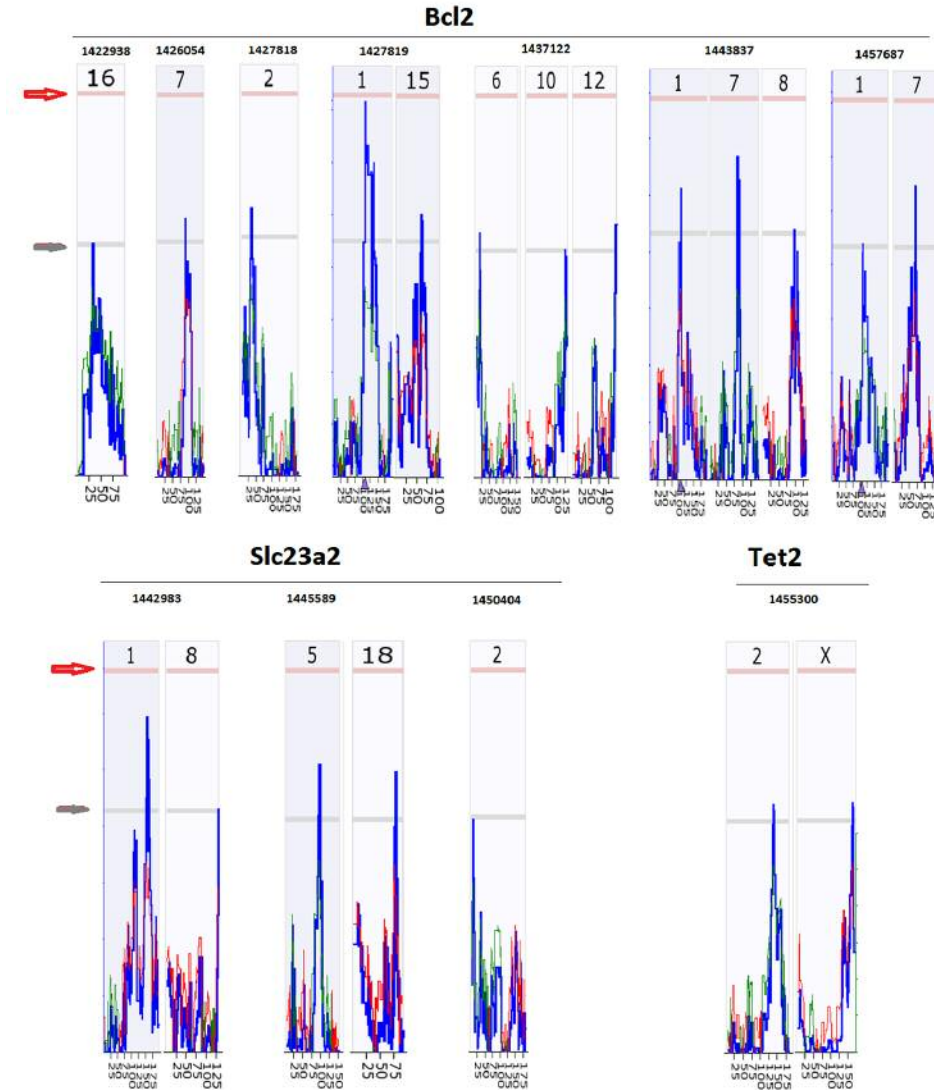


Figure 1. The eQTL locations of probes of the three genes *Tet* methylcytosine dioxygenase (*Tet2*), *B*-cell lymphoma 2 (*Bcl2*), and solute carrier family 23 member 2 (*Slc23a2*) on different chromosomes. The number identifier of the chromosomes is on the top of each individual graph, while the Mb of the physical distance is indicated at the bottom. The pink lines (indicated by the red arrow) indicate the threshold of significance level, and the gray lines (indicated by the gray arrow) indicate the threshold of the level suggestive of expression quantitative trait locus (eQTL). Blue traces indicate LSR scores by chromosome, while red and green traces (on each individual chromosome) indicate regions where B6 or D2 alleles tended to increase the expression level of the gene, respectively. The top line is all eQTL from probes (probe numbers are on top of each graph) of *Bcl2*. The bottom left is eQTL from probes of *Slc23a2*. The bottom right is the eQTL from *Tet2*.

In Figure 3, One pathway goes through five steps, which involves four genes (marked as P1 in Figure 3A, light blue), *Bag4*, *Bcl2l13*, *Mcl1* and *Becn1*, before reaching *Slc23a2*. The effect starting with *Bcl2* was strongly negatively associated with *Bag4* (Figure 3B), and *Bag4* was strongly positively associated with *Bcl2l13* (Figure 3C). *Bcl2l13* was negatively correlated with *Mcl1* (Figure 3D), *Mcl1* was positively correlated with *Becn1* (Figure 3E), and *Becn1* finally negatively correlated with *Slc23a2* (Figure 3F). The other pathway goes through the *Mcl1* and *Becn1* nodes (marked as P2 in Figure

3A, red). Unlike the direct association between *Bcl2* and *Slc23a2*, the genes in the indirect association pathways were all strongly associated with each other. In addition, *Mcl1* also appears to serve as an alternative connection between *Bcl2* and *Tet2*. As indicated by the arrows in Figure 3G, the expression level of *Bcl2* was strongly positively associated with that of *Mcl1*. The expression level of *Mcl1* is then positively associated with that of *Tet2* (Figure 3H). Thus, *Mcl1* is important in both pathways, between *Slc23a2* and *Bcl2*, and between *TET2* and *Bcl2*.

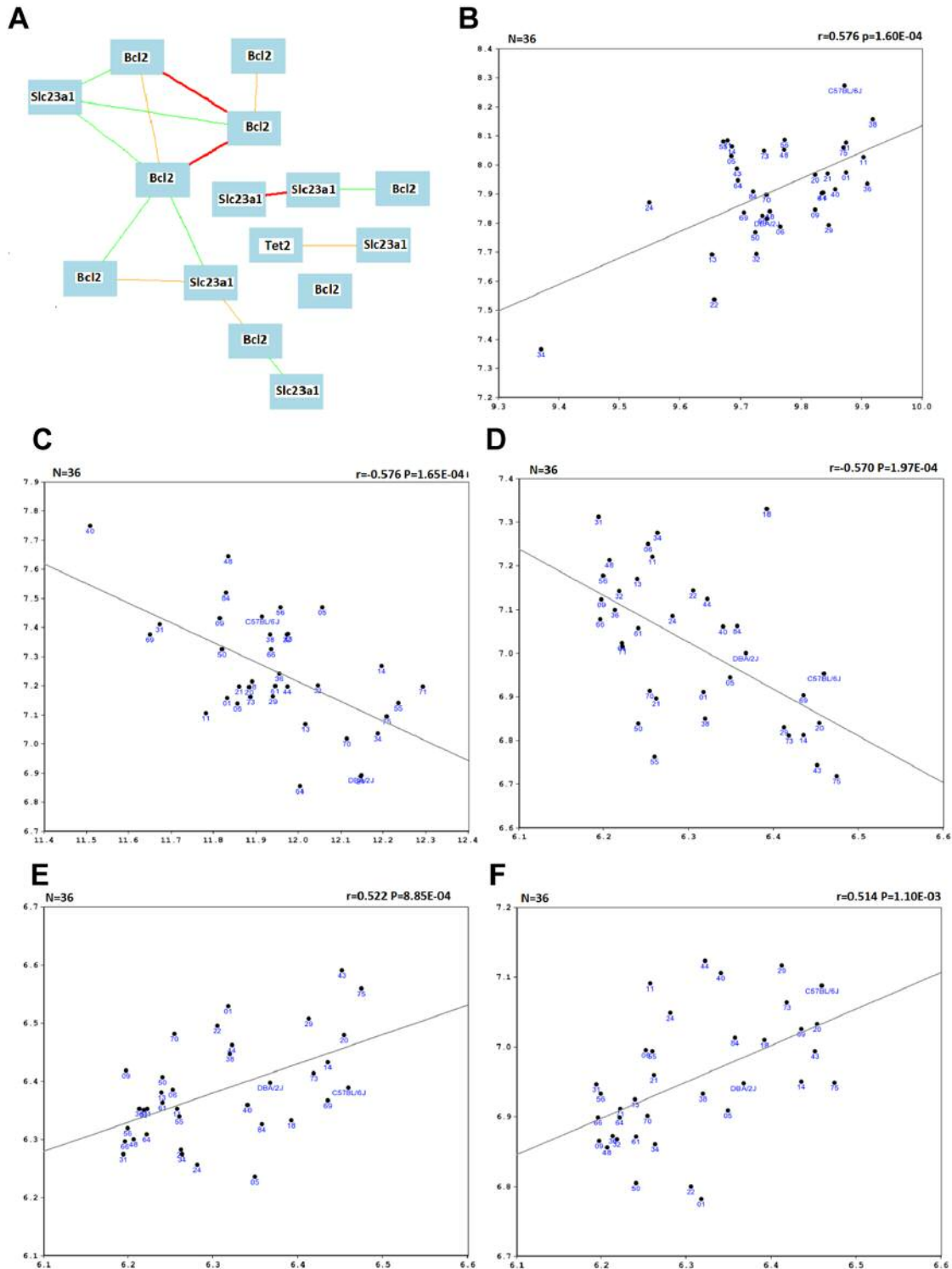


Figure 2. Gene network and pathway composed of one probe of Tet methylcytosine dioxygenase (*Tet2*), eight probes of B-cell lymphoma 2 (*Bcl2*), and six probes of solute carrier family 23 member 2 (*Slc23a2*). A: The network graphs. The 16 nodes in the graph show the selected traits. The 15 curved edges between the nodes were filtered from the total of 15 edges. Only Pearson correlation coefficients greater than 0.43 or less than  $-0.43$  are displayed. B: The positive correlation between the expression level of a probe of *Tet2* (probe 1455300) and that of a probe of *Slc23a2* (probe 1445983). C: The negative association between a probe of *Bcl2* (1443874) and a probe of *Slc23a2* (probe 1421912). D: A probe of *Slc23a2* (1417330) shows negative correlation with one *Bcl2* probe (1437122). E: A probe of *Slc23a2* (1417330) was positively correlated with a probe (1426054) of *Bcl2*. F: A probe of *Slc23a2* (1417330) was positively correlated with a probe (1422938) of *Bcl2*.

Table I. Correlation R values for Tet2, Bcl2 and Slc23a2\*.

	Spearman rank correlation (rho), Trait														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pearson															
Trait 1: UCLA_BXD_Liv_Jan16::1457687_at Bcl2 on Chr 1 @ 106.538248 Mb extreme distal 3' UTR		0.750	0.154	-0.200	0.600	0.142	0.853	0.046	-0.270	-0.249	-0.454	-0.129	-0.304	-0.274	0.109
Trait 2: UCLA_BXD_Liv_Jan16::1437122_at Bcl2 on Chr 1 @ 106.542085 Mb mid distal 3' UTR	0.701		-0.184	-0.501	0.336	0.344	0.674	0.088	-0.243	-0.203	-0.486	-0.565	-0.190	-0.326	0.008
Trait 3: UCLA_BXD_Liv_Jan16::1422938_at Bcl2 on Chr 1 @ 106.543246 Mb exons 1 and 2	0.201	-0.116		0.096	0.268	-0.103	0.024	-0.057	-0.182	-0.151	0.089	0.576	-0.455	0.213	0.191
Trait 4: UCLA_BXD_Liv_Jan16::1426054_at Bcl2 on Chr 1 @ 106.624135 Mb intron 1	-0.166	-0.537	0.042		-0.013	-0.428	-0.099	-0.112	0.001	0.014	0.211	0.482	0.070	-0.036	0.244
Trait 5: UCLA_BXD_Liv_Jan16::1427819_at Bcl2 on Chr 1 @ 106.69806 Mb intron 1 (from BC027249)	0.632	0.390	0.325	-0.043		0.115	0.496	-0.356	-0.038	0.062	-0.191	0.060	-0.377	-0.333	0.007
Trait 6: UCLA_BXD_Liv_Jan16::1427818_at Bcl2 on Chr 1 @ 106.698992 Mb antisense in intron of Bcl2 (from EST AK047377)	0.133	0.372	-0.153	-0.410	0.106		-0.024	-0.377	0.339	0.419	0.009	-0.305	0.057	-0.064	-0.080
Trait 7: UCLA_BXD_Liv_Jan16::1440770_at Bcl2 on Chr 1 @ 106.71037 Mb exon 1 or segment of intron 1	0.849	0.643	0.104	-0.043	0.494	-0.002		0.148	-0.327	-0.377	-0.510	-0.256	-0.186	-0.273	0.247
Trait 8: UCLA_BXD_Liv_Jan16::1443837_x_at Bcl2 on Chr 1 @ 106.711825 Mb 3' UTR of isoform 2	0.072	0.078	0.016	-0.074	-0.263	-0.423	0.129		-0.382	-0.464	-0.033	-0.045	0.013	0.282	0.180
Trait 9: UCLA_BXD_Liv_Jan16::1450404_at Slc23a1 on Chr 18 @ 35.614355 Mb	-0.300	-0.243	-0.172	0.092	-0.037	0.281	-0.331	-0.497		0.924	0.099	0.088	-0.017	0.088	-0.483
Trait 10: UCLA_BXD_Liv_Jan16::1421912_at Slc23a1 on Chr 18 @ 35.615432 Mb	-0.342	-0.236	-0.208	0.065	-0.011	0.358	-0.455	-0.576	0.900		0.143	0.082	-0.008	0.034	-0.514
Trait 11: UCLA_BXD_Liv_Jan16::1417329_at Slc23a2 on Chr 2 @ 132.052618 Mb distal 3' UTR	-0.526	-0.505	0.074	0.151	-0.252	0.010	-0.513	0.039	0.110	0.205		0.306	0.326	0.444	0.206

Table I. Continued

Table I. Continued

	Spearman rank correlation (rho), Trait														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Trait 12: UCLA_BXD_ Liv_Jan16::1417330_at <i>Slc23a2</i> on Chr 2 @ 132.056296 Mb last exon and proximal 3' UTR	-0.145	-0.570	0.514	0.522	0.020	-0.359	-0.215	0.037	0.070	0.012	0.311		-0.278	0.127	0.006
Trait 13: UCLA_BXD_ Liv_Jan16::1445589_at <i>Slc23a2</i> on Chr 2 @ 132.092737 Mb intron 3	-0.330	-0.166	-0.520	0.043	-0.416	0.088	-0.240	0.065	0.003	0.054	0.387	-0.268		0.268	0.209
Trait 14: UCLA_BXD_ Liv_Jan16::1442983_at <i>Slc23a2</i> on Chr 2 @ 132.124984 Mb first intron	-0.266	-0.380	0.305	0.038	-0.237	-0.256	-0.279	0.260	-0.049	-0.058	0.434	0.210	0.247		0.351
Trait 15: UCLA_BXD_ Liv_Jan16::1455300_at <i>Tet2</i> on Chr 3 @ 133.463827 Mb	0.028	-0.136	0.207	0.305	0.020	-0.307	0.185	0.291	-0.444	-0.488	0.167	0.082	0.172	0.576	

\*Lower left cells list Pearson product-moment correlations; upper right cells list Spearman rank order correlations. Each cell included 36 counts. *Bcl2*: B-Cell leukemia/lymphoma 2; *Slc23a1*: solute carrier family 23 (nucleobase transporters), member 1; *Slc23a2*: solute carrier family 23 (nucleobase transporters), member 2; *Tet2*: *TET* oncogene family member 2. R Values higher than 0.7 are displayed in red; those between 0.5 and 0.7, in orange; between -0.5 and -0.7, in green; Values between -0.499 and 0.499 are displayed in blue.

*Association analysis at the whole-genome level for indirect connections.* In order to further explore novel genes that may participate in the pathways that connect these three genes, analyses at the whole-genome level were conducted to examine the relationships between these three genes using one probe identified from the NCBI database from each of the three genes.

*Slc23a2.* The top 100 genes found to be closely correlated with the *Slc23a2* probe were identified by the function "Trait Correlation" on the GeneNetwork website (<http://www.genenetwork.org/webqtl/main.py?sid=c53bac41bbf57ff26fdd53e4aa139c2cc7cfe59b>). These genes almost show the same level of correlation with *Bcl2* and *Tet2*, so we selected 25 genes from both *Bcl2* and *Tet2* whose R values exceeded  $\pm 0.5$  for further analysis. The R values for these 25 selected genes with *Tet2* and *Bcl2* were 0.02344 and 0.15336, respectively. The corresponding R values were 0.39736 and 0.3604. As shown in Figure 5A, genes ribosomal protein S7 (*Rps7*), tumor protein D52-like 2 (*Tpd52l2*), Small ubiquitin-related modifier 2 (*Sumo2*), palmitoyl-protein thioesterase 1 (*Ppt1*), nucleic acid binding protein 1 (*5830411E10Rik*), vacuolar protein sorting-associated protein 29 (*Vps29*), IMP U3 small nucleolar ribonucleoprotein 4 (*Imp4*), vacuolar protein-sorting-associated protein 24 (*Vps24*), *Ube2l3* and UTP18 small subunit processome component (*Utp18*) were

strongly positively correlated with *Bcl2*, with R values ranging from 0.641 to 0.502. Genes CREB regulated transcription coactivator 2 (*Crtc2*), transmembrane protein 259 (*Tmem259*), and activin A receptor type 2B (*Acvr2b*) were negatively correlated with *Bcl2*, with R values ranging from -0.61 to -0.5. However, for these two groups of genes, there were weaker negative correlations with *Tet2*, with R values from -0.138 to -0.37. The other genes we found to have stronger correlations with *Tet2*. The genes mixed lineage leukemia 4 (*Mll4*), CTD small phosphatase 2 (*Ctdsp2*, *Centd2*), O-linked N-acetylglucosamine (*Ogt*), Jumonji and AT-rich interaction domain containing 2 (*Jarid2*), far upstream element binding protein 1 (*Fubp1*), and meningioma-expressed antigen 5 (*Mgea5*) had positive correlations, with R values from 0.524 to 0.643, and genes ubiquitin-like autophagy protein (*Apg12*), *riken cDNA 1600012H06 (1600012H06Rik)*, *Cse11*, and forty-two-three domain containing 1 (*Fyttl1*) were negatively correlated with *Tet2*, with R values ranging from -0.505 to -0.578.

Gene ontology (GO) analysis suggested that the majority of these 25 genes with significant correlation with *Slc23a2* are involved in biological processes, with only a few genes in cellular components, and no genes involved in molecular functions (Figure 4B). As shown in Figure 4C, from the aspect of cellular components, two genes were found to be

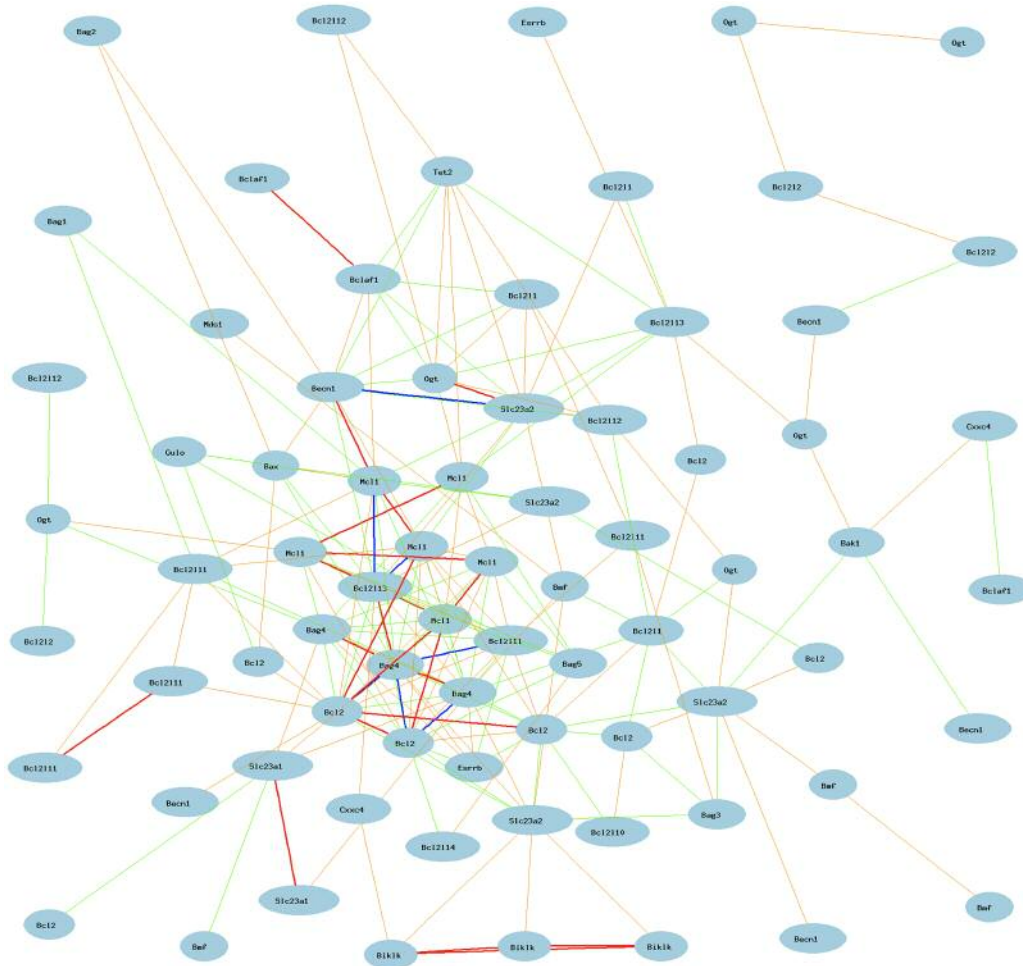


Figure 3. Key genes and the connection among genes of the full list. The 92 nodes show the selected traits. Only nodes with edges are displayed. The 216 edges between the nodes, filtered from the 4,186 total edges and drawn as lines, show Pearson correlation coefficients greater than 0.5 or less than  $-0.5$ . Different color of lines represents different R values between the two genes as explained in the Materials and Methods section.

involved in pore complex. Nineteen out of these 25 genes are functional in intercellular parts, such as the intercellular membrane-bounded organelle, also related to the nucleus. For example, nucleic acid binding protein 1 (*Nabp1*) can regulate nucleic acid binding protein 1. As to biological processes, the correlations with these functions are complex. Ten genes are involved in cellular component organization and biogenesis, such as *Ppt1* and *Atg12* for vacuole organization. For the function of localization, there are seven genes involved in macromolecule localization, such as protein localization. There are also eight genes involved in cellular localization, for example, *Chmp3* and *Vps29* can regulate vacuole localization. Three genes take part in signal release. Eleven genes are involved in nucleobase-containing compound metabolic processes.

*Tet2*. The 100 top genes that are closely correlated with the expression of probes of *Tet2* were identified by using the

function “Trait Correlation” at GeneNetwork (<http://www.genenetwork.org/webqtl/main.py?sid=dabd84131eaa36908ab381e9616642cb4a39803>). Compared to *Bcl2*, the probe *Slc23a2* showed much stronger correlation with *Tet2*. We selected genes between *Tet2* and *Slc23a2* whose R values were up to 0.6. These genes were *Fubp1*, *Centd2*, *Mll4*, *BC013481*, *PsmA4*, synaptobrevin homolog *YKT6* (*Ykt6*), splicing factor 3B, 14 kDa subunit (*Sf3b14*) and BH3 interacting domain death agonist (*Bid*) (Figure 5D).  $\Sigma R$  for *Slc23a2* was 0.099, while  $R_e$  was 0.01237. The  $R_n$  of *Slc23a2* was 5.363, while  $R_a$  was 0.6703. From Figure 4D, *Fubp1*, *Centd2*, *Mll4* and *BC013481* can be seen to be strongly positively correlated with *Slc23a2*, with R values close to 0.7. In contrast, *PsmA4*, *Ykt6*, *Sf3b14*, and *Bid* were strongly negatively correlated with *Slc23a2*. However, these genes had typical standard correlation with *Bcl2*, which probably means there is no interaction between *Tet2* and *Bcl2*.



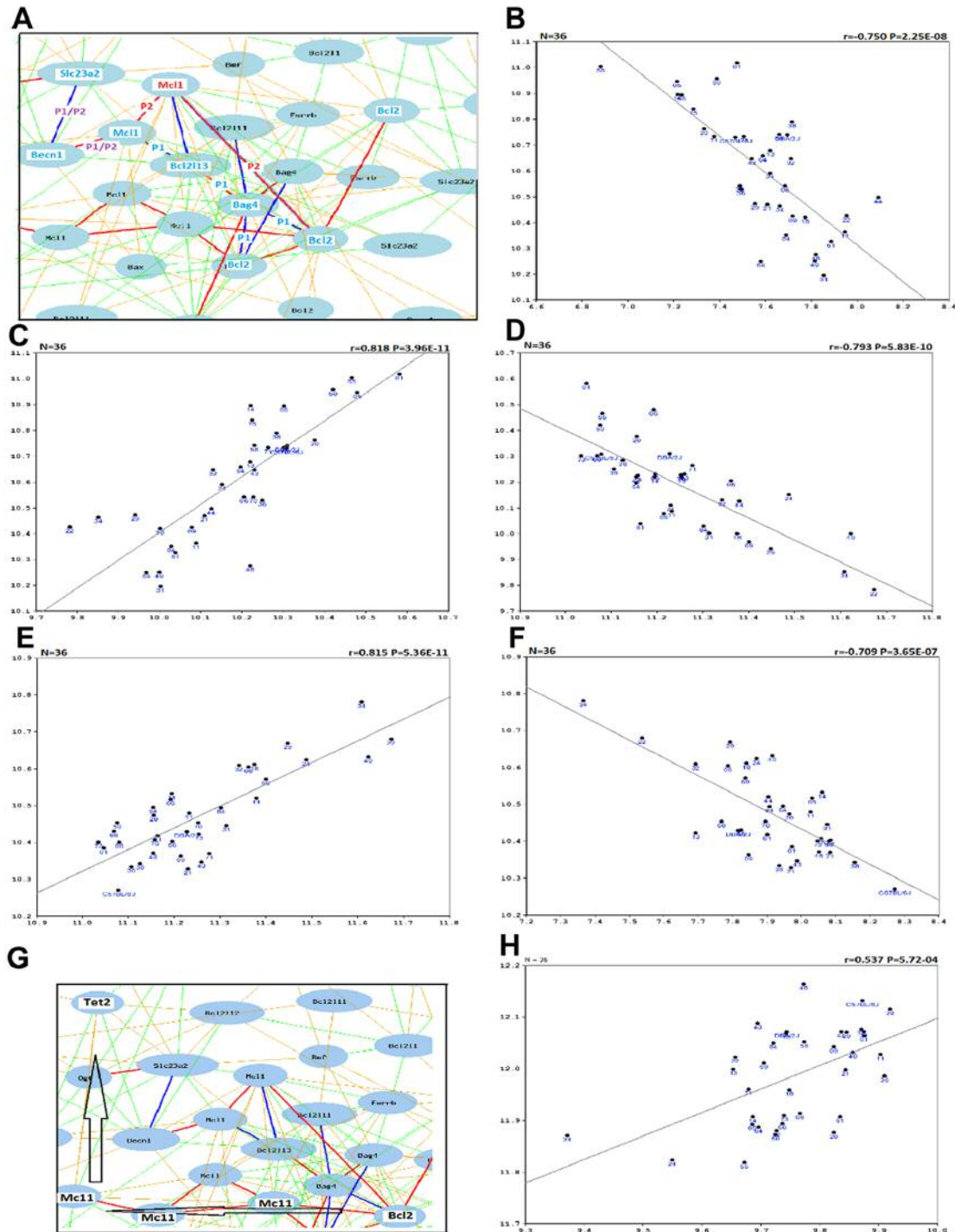


Figure 4. Alternative gene networks and pathways composed of additional relevant genes. In A and G, different color of lines represents different R values between the two genes as explained in the Materials and Methods section of the manuscript. In B, C, D, E, F and H, mouse strain names are written in blue color. A: The two alternative pathways between B-cell lymphoma 2 (Bcl2), and six probes of solute carrier family 23 member 2 (Slc23a2) (marked as P1, light blue, and P2, red). B: The expression level of a probe of Bcl2 (probe 1440770) was negatively correlated with that of a probe of BCL2-associated athanogene 4 (Bag4) (probe 1439024). C: A probe of Bag4 (1439024) was positively correlated with a probe of BCL2-like 13 (1429539). D: A probe of BCL2-like 13 (Bcl2l13) (1429539) was negatively correlated with a probe of myeloid leukemia cell differentiation protein (Mcl1) (1416881). E: A probe of Mcl1 (1416881) was positively correlated with a probe of beclin 1 (Becn1) (1460320). F: A probe of Becn1 (1460320) was negatively correlated to a probe of Slc23a2 (1442983). G: The alternative pathway between Bcl2 and Tet2 (indicated by arrows). H: The expression level of Mcl1 (probe 1456381) was positively correlated to that of Tet2 (probe 1455300).

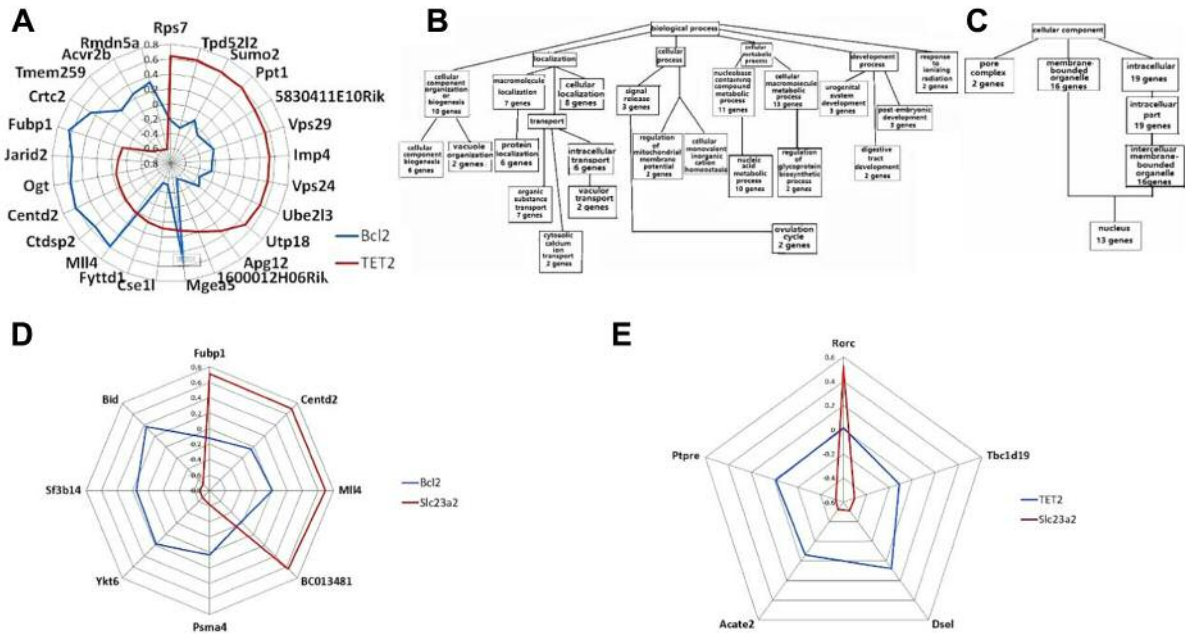


Figure 5. Association between the most correlated genes from the top 100 identified from *Tet* methylcytosine dioxygenase (*Tet2*), eight probes of *B*-cell lymphoma 2 (*Bcl2*), and six probes of solute carrier family 23 member 2 (*Slc23a2*) and functional analysis of the 25 selected genes. A: Genes according to different functional groups identified by GO analysis from 25 selected genes out of the top 100 and *Bcl2* and *Tet2*. These genes only have functional roles in two functional categories: biological processes and cellular components. B: Genes identified for biological processes. Some further functions in the database are not listed. C: Genes identified for correlation with cellular components. D: Association between *Bcl2*, *Slc23a2*, and the eight most highly correlated genes from the top 100 genes identified from *Tet2*. The expression levels of *Slc23a2* were highly correlated to these eight genes. E: Association between *Tet2*, *Slc23a2* and the five most highly correlated genes from the top 100 genes identified from *Bcl2*. Numbers listed vertically on the circular lines are the R values at different positions. Circular lines indicate the positions of different R values.

*Bcl2*. The top 100 genes that are closely correlated with the *Bcl2* probe (<http://www.genenetwork.org/webqtl/main.py?sid=fd3396eb5ffebcd0d5ad360d1edb3e3e43595205>) were identified in the same way as the analysis for *Tet2* and *Slc23a2*. Unlike the 100 genes for *Slc23a2* and *Tet2*, the top correlated genes for *Bcl2* had weak correlations with *Tet2* and *Slc23a2*, especially with *Tet2*, whose highest R value was  $-0.116$  (Figure 4E). For probe *Tet2*,  $\Sigma R$  was  $-0.1$ ,  $Re$  was  $-0.02$ ,  $Rn$  was  $0.284$ , while  $Ra$  ( $\Sigma IRI/n$ ) was  $0.0568$ . For probe *Slc23a2*,  $\Sigma R$  was  $-1.555$ ,  $Re$  was  $-0.311$ ,  $Rn$  was  $2.607$ , while  $Ra$  was  $0.5214$ . These data suggest that the interaction of *Tet2* with these 100 genes associated with *Bcl2* has no or very weak correlation and may even operate by a completely different mechanism of action. However, the R value for association between *Bcl2* and *Slc23a2* was around  $0.5$ , which is still at a typical intermediate level. We selected correlations with *Slc23a2* which gave R values over  $0.5$ , leading to selectin of genes RAR-related orphan receptor C (*Rorc*), TBC1 domain family member 19 (*Tbc1d19*), dermatan sulfate epimerase like (*Dsel*), *Acate2* and protein tyrosine phosphatase receptor type E (*Ptpre*). As shown in Figure 4E, the majority of these five genes negatively correlated with *Bcl2* at an intermediate level.

## Discussion

Our analysis indicated that the expression levels of *Tet2*, *Bcl2* and *Slc23a2* are associated differently with each other through a variety of pathways. The associations among them can be direct or indirect. The indirect associations could be mediated through multiple ways connected by a considerable number of related genes. Overall, the association between the expression levels of *Bcl2* and *Slc23a2* were stronger than those with *Tet2*. Although both *Bcl2* and *Slc23a2* are involved in vitamin C pathways, (3, 4) as far as we are aware, this is the first study that reveals such a close relationship between *Bcl2* and *Slc23a2*. Furthermore, deep analysis of related genes and at the whole-genome level revealed several novel genes that play important roles in the association of expression levels among these three genes. These novel genes may serve as new targets for vitamin C-related therapeutic applications.

By using direct association analysis among probes of the three genes, we realized the association among these three genes is complicated (Figure 2). By matrix and gene network graph, we obtained the complex associations between *Bcl2* and *Slc23a2*, with positive and negative connections among different probes of these two genes. *Tet2* was positively correlated to one probe

of *Slc23a2*. However, none of the associations were strong. Thus, these correlations may reflect association only, not the pathway relationship. Such speculation is also supported by the result from eQTL mapping, in which none of these probes mapped on the same loci, or to the location of other genes, among probes of these three genes (Figure 1). Therefore, we presumed that additional analysis with known relevant genes or with whole genome data may discover the key genes among their pathways or prove no strong correlations among them.

Our first expanded analysis using known relevant genes revealed new key genes in their associations and pathways. Firstly, we discovered two potential pathways in the association between the expression level of *Bcl2* and *Slc23a2*. One goes through five steps, which involve four genes, *Bag4*, *Bcl2l13*, *Mcl1*, and *Becn1*, before reaching *Slc23a2*. The other is through *Mcl1* and *Becn1*. Both eventually end up with negative associations. Our analysis also revealed an alternative correlation between *Bcl2* and *TET2* through *Mcl1*. A key discovery is that *Mcl1* is involved in both pathways – between *Bcl2* and *Slc23a2*, and between *Bcl2* and *TET2* (12-14). Thus, *Mcl1* may function as a key switch between these two pathways. The expression level of *Mcl1* will have positive correlation with *TET2*, while it negatively correlates with *Slc23a2*.

Our second whole-genome analysis using each probe with known sequence from each of these three genes discovered an additional set of important genes that potentially are the pathways connecting these three genes. Our data indicated that the probe *Fubp1* from the 100 top genes of *Slc23a2* (1442983\_at) can connect *Tet2* and *Slc23a2* (1442983\_at), which had the highest expression level. Surprisingly, in the group that compared *Slc23a2* with the 100 top genes from *Tet2*, probe *Fubp1* also had the highest correlation level. This indicates that gene *Fubp1* can play a crucial role in connecting *Slc23a2* with *Tet2*. *Fubp1* is a transcriptional regulator and plays an important role in precisely controlling the transcription of *c-Myc*. The c-Myc protein regulates the transcription of many different target genes for proliferation, cell cycle progression, differentiation, apoptosis, and cell metabolism (15, 16). Therefore, *Fubp1* gene also controls the same functions as c-Myc protein, which was verified by different experimental approaches. As a potential oncogene, *Fubp1* is overexpressed in different cancer types, such as in hepatocellular carcinoma (15). *Fubp1* down-regulates gene expression of the cell-cycle inhibitors p21 and p15, which causes cell -cycle acceleration (16). As *Fubp1* fulfills an important function in hepatocellular carcinoma, it may consequently have an impact on the production and metabolism of vitamin C. This might be the most significant result obtained from our analysis.

The GO analysis indicated that 25 genes selected from *Slc23a2*'s top 100 genes correlated with biological processes. If the expression of probe *Slc23a2* is inhibited, it will have an impact on many metabolic processes, including cellular macromolecule metabolic processes.

Analysis between *Bcl2* (probe 1457687) and *Slc23a2* (probe 1442983) found several genes highly correlated between them, such as *Rps7*, *Tpd52l2*, *Sumo2* and *Rmdn5a*. In particular, the expression of gene *Tpd52l2* may be a marker for acute lymphoblastic leukemia. Based on the current information, *Tpd52* and *Tpd52l2* are often co-expressed in childhood leukemia (17). In real-time PCR analyses, D54 expression was revealed in leukemic bone marrow samples (17, 18). Among patients with acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), D54 transcripts are also frequently detected in bone marrow and its expressions is highly positively correlated with disease. Several other studies have indicated that the expression of different TPD52-like genes varies in different leukocytes and hematological malignancies (18-20). Microarray analyses have also demonstrated that D54 levels fluctuate in human B-cells and neutrophils (21).

As in all genomic analyses, the data from our analysis need to be confirmed at the protein level and with other studies, especially for *Mcl1*, *Tpd52l2* and *Fubp1*. It is critical to determine whether they do have an impact on the function of vitamin C, or if their inhibition influences leukemia. We realize that our study is based on the liver genome, and that results obtained may not necessarily be the same as those obtained using the blood genome. Furthermore, the mouse model is different from humans in many ways. Future studies will be necessary to determine whether these pathways or key genes act similarly in humans. Nevertheless, our analysis reveals potential functions of these genes in the connections among the three key genes *Mcl1*, *Fubp1*, and *Tpd52l2* in the vitamin C pathways in relation to leukemia. These data set the foundation for further studies and potential future therapeutic applications.

## Authors' Contributions

Conceived and designed the experiments: WG, WZ. Performed data searching and collection: JL, W-QZ, WL, JR, YH, WG. Analyzed the data: JL, W-QZ, WL, YC, WG. Contributed analysis tools: WZ, WG. Wrote the article: JL, WL, W-QZ, JR, YC, WG.

## Conflicts of Interest

The Authors declare no competing financial interests regarding this study.

## Acknowledgements

The Authors thank Dr. David Armbruster for editing this article. This work was partially supported by funding from merit grant I01 BX000671 to WG from the Department of Veterans Affairs and the Veterans Administration Medical Center in Memphis, TN, USA. The key basic applied project of Hebei provincial Department of Science & Technology (grant #: 15967730D) was awarded to Wei Zhang.

References

- 1 Gillberg L, Ørskov AD, Liu M, Harsløf LBS, Jones PA and Grønbæk K: Vitamin C-A player in regulation of the cancer epigenome. *Semin Cancer Biol* 51: 59-67, 2018. PMID: 29102482. DOI: 10.1016/j.semcancer.2017.11.001
- 2 Lv H, Wang C, Fang T, Li T, Lv G, Han Q, Yang W and Wang H: Vitamin C preferentially kills cancer stem cells in hepatocellular carcinoma *via* SVCT-2. *NPJ Precis Oncol* 2(1): 1, 2018. PMID: 29872720. DOI: 10.1038/s41698-017-0044-8
- 3 Zhao H, Zhu H, Huang J, Zhu Y, Hong M, Zhu H, Zhang J, Li S, Yang L, Lian Y, Wang S, Mao J, Chen Y, Li J and Qian S: The synergy of Vitamin C with decitabine activates TET2 in leukemic cells and significantly improves overall survival in elderly patients with acute myeloid leukemia. *Leuk Res* 66: 1-7, 2018. PMID: 29331774. DOI: 10.1016/j.leukres.2017.12.009
- 4 Vineetha RC, Binu P, Arathi P and Nair RH: L-Ascorbic acid and  $\alpha$ -tocopherol attenuate arsenic trioxide-induced toxicity in H9c2 cardiomyocytes by the activation of Nrf2 and Bcl2 transcription factors. *Toxicol Mech Methods* 28(5): 353-360, 2018. PMID: 29297235. DOI: 10.1080/15376516.2017.1422578
- 5 Chauhan D, Bartok E, Gaidt MM, Bock FJ, Herrmann J, Seeger JM, Broz P, Beckmann R, Kashkar H, Tait SWG, Müller R and Hornung V: BAX/BAK-induced apoptosis results in caspase-8-dependent IL-1 $\beta$  maturation in macrophages. *Cell Rep* 25(9): 2354-2368.e5, 2018. PMID: 30485805. DOI: 10.1016/j.celrep.2018.10.087
- 6 Miller PG and Ebert BL: Vitamin C regulates stem cells and cancer. *Nature Sep* 549(7673): 462-464, 2017. PMID: 28869971. DOI: 10.1038/nature23548
- 7 Solary E, Bernard OA, Tefferi A, Fuks F and Vainchenker W: The ten-eleven translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia* 28(3): 485-496, 2014. PMID: 24220273. DOI: 10.1038/leu.2013.337
- 8 Pan W, Zhu S, Qu K, Meeth K, Cheng J, He K, Ma H, Liao Y, Wen X, Roden C, Tobiasova Z, Wei Z, Zhao J, Liu J, Zheng J, Guo B, Khan SA, Bosenberg M, Flavell RA and Lu J: The DNA methylcytosine dioxygenase TET2 sustains immunosuppressive function of tumor-infiltrating myeloid cells to promote melanoma progression. *Immunity* 47(2): 284-297.e5, 2017. PMID: 28813659. DOI: 10.1016/j.immuni.2017.07.020
- 9 Bennett BJ, Davis RC, Civelek M, Orozco L, Wu J, Qi H, Pan C, Packard RR, Eskin E, Yan M, Kirchgessner T, Wang Z, Li X, Gregory JC, Hazen SL, Gargalovic PS and Lusis AJ: Genetic architecture of atherosclerosis in mice: a systems genetics analysis of common inbred strains. *PLoS Genet* 11(12): e1005711, 2015. PMID: 26694027. DOI: 10.1371/journal.pgen.1005711
- 10 Jiao Y, Chen H, Yan J, Wang L, Huang Y, Liu X, Williams RW, Lu L, Wang Y and Gu W: Genome-wide gene expression profiles in antioxidant pathways and their potential sex differences and connections to vitamin C in mice. *Int J Mol Sci* (5): 10042-10062, 2013. PMID: 23665904. DOI: 10.3390/ijms 140510042
- 11 Cui J, Liu L, Lu H, Wei D, Jiao Y, Jablonski MM, Williams RW, Gu W and Chen H: Potential effect on molecular pathways in different targeted genes in the VEGF family in retina - From the genomic point of view. *Exp Eye Res* 176: 78-87, 2018. PMID: 29944851. DOI: 10.1016/j.exer.2018.06.024
- 12 Kontos CK, Avgeris M, Vassilacopoulou D, Ardavanis A and Scorilas A: Molecular effects of treatment of human colorectal cancer cells with natural and classical chemotherapeutic drugs: alterations in the expression of apoptosis-related BCL2 family members, including BCL2L12. *Curr Pharm Biotechnol* 19(13): 1064-1075, 2018. PMID: 30417787. DOI: 10.2174/1389201019666181112101410
- 13 Urtishak KA, Wang LS, Culjkovic-Kraljacic B, Davenport JW, Porazzi P, Vincent TL, Teachey DT, Tasian SK, Moore JS, Seif AE, Jin S, Barrett JS, Robinson BW, Chen IL, Harvey RC, Carroll MP, Carroll AJ, Heerema NA, Devidas M, Dreyer ZE, Hilden JM, Hunger SP, Willman CL, Borden KLB and Felix CA: Targeting EIF4E signaling with ribavirin in infant acute lymphoblastic leukemia. *Oncogene* 38(13): 2241-2262, 2019. PMID: 30478448. DOI: 10.1038/s41388-018-0567-7
- 14 Chen YJ, Huang CH, Shi YJ, Lee YC, Wang LJ and Chang LS: The suppressive effect of arsenic trioxide on TET2-FOXp3-LYN-AKT axis-modulated MCL1 expression induces apoptosis in human leukemia cells. *Toxicol Appl Pharmacol* 358: 43-55, 2018. PMID: 30213730. DOI: 10.1016/j.taap.2018.09.008
- 15 Gerlach K and Zörnig M: FUBP1 (far upstream element (FUSE) binding protein 1). *Atlas Genet Cytogenet Oncol Haematol* 17(12): 816-820, 2013.
- 16 Rabenhorst U, Thalheimer FB, Gerlach K, Kijonka M, Böhm S, Krause DS, Vauti F, Arnold HH, Schroeder T, Schnütgen F, von Melchner H, Rieger MA and Zörnig M: Single-stranded DNA-binding transcriptional regulator FUBP1 is essential for fetal and adult hematopoietic stem cell self-renewal. *Cell Rep* 11(12): 1847-1855, 2015. PMID: 26095368. DOI: 10.1016/j.celrep.2015.05.038
- 17 Barbaric D, Byth K, Dalla-Pozza L and Byrne JA: Expression of tumor protein D52-like genes in childhood leukemia at diagnosis: Clinical and sample considerations. *Leuk Res* 30(11): 1355-1363, 2006. PMID: 16620967. DOI: 10.1016/j.leukres.2006.03.009
- 18 Hofmann W-K, deVos S, Komor M, Hoelzer D, Wachsman W and Koeffler HP: Characterization of gene expression of CD34+ cells from normal and myelodysplastic bone marrow. *Blood* 100(10): 3553-3560, 2002. PMID: 12411319. DOI: 10.1182/blood.V100.10.3553
- 19 Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK, Liu HC, Mahfouz R, Raimondi SC, Lenny N, Patel A and Downing JR: Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* 102(8): 2951-2959, 2003. PMID: 12730115. DOI: 10.1182/blood-2003-01-0338
- 20 Mircean C, Tabus I, Kobayashi T, Yamaguchi M, Shiku H, Shmulevich I and Zhang W: Pathway analysis of informative genes from microarray data reveals that metabolism and signal transduction genes distinguish different subtypes of lymphomas. *Int J Oncol* 24(3): 497-504, 2004. PMID: 14767533.
- 21 Subrahmanyam YV, Yamaga S, Prashar Y, Lee HH, Hoe NP, Kluger Y, Gerstein M, Goguen JD, Newburger PE and Weissman SM: RNA expression patterns change dramatically in human neutrophils exposed to bacteria. *Blood* 97(8): 2457-2468, 2001. PMID: 11290611.

Received March 26, 2019  
 Revised June 15, 2019  
 Accepted June 20, 2019