Gene set enrichment and topological analyses based on interaction networks in pediatric acute lymphoblastic leukemia

SHUXIANG SUI, XIN WANG, HUA ZHENG, HUA GUO, TONG CHEN and DONG-MEI JI

Department of Pediatrics, Shandong Dongying People's Hospital, Dongying, Shandong 257091, P.R. China

Abstract. Pediatric acute lymphoblastic leukemia (ALL) accounts for over one-quarter of all pediatric cancers. Interacting genes and proteins within the larger human gene interaction network of the human genome are rarely investigated by studies investigating pediatric ALL. In the present study, interaction networks were constructed using the empirical Bayesian approach and the Search Tool for the Retrieval of Interacting Genes/proteins database, based on the differentially-expressed (DE) genes in pediatric ALL, which were identified using the RankProd package. Enrichment analysis of the interaction network was performed using the network-based methods EnrichNet and PathExpand, which were compared with the traditional expression analysis systematic explored (EASE) method. In total, 398 DE genes were identified in pediatric ALL, and LIF was the most significantly DE gene. The co-expression network consisted of 272 nodes, which indicated genes and proteins, and 602 edges, which indicated the number of interactions adjacent to the node. Comparison between EASE and PathExpand revealed that PathExpand detected more pathways or processes that were closely associated with pediatric ALL compared with the EASE method. There were 294 nodes and 1,588 edges in the protein-protein interaction network, with the processes of hematopoietic cell lineage and porphyrin metabolism demonstrating a close association with pediatric ALL. Network enrichment analysis based on the PathExpand algorithm was revealed to be more powerful for the analysis of interaction networks in pediatric ALL compared with the EASE method. LIF and MLLT11 were identified as the most significantly DE genes in pediatric ALL. The process of hematopoietic cell lineage was the pathway most significantly associated with pediatric ALL.

Correspondence to: Dr Shuxiang Sui, Department of Pediatrics, Shandong Dongying People's Hospital, 317 Nanyi Road, Dongying, Shandong 257091, P.R. China
E-mail: shuxiangsui@yeah.net

Introduction

Acute lymphoblastic leukemia (ALL) is the most common malignant tumors in children, and accounts for >80% of all cases of acute leukemia (1). It is estimated that ~4% of children <15 years of age develop ALL, which comprises over one-quarter of all pediatric cancers (2). Therefore, pediatric ALL has attracted increasing attention in studies.

Rapid progress in molecular biology has aided in the selection of patients with specific poor prognostic genetic factors, and sensitive monitoring of minimal residual disease (MRD) has also contributed significantly to the management of pediatric ALL (3). Previous studies have indicated that monitoring MRD constitutes an essential marker, and that detection of MRD, particularly at the end of induction and subsequent to treatment completion, was significantly predictive for the outcome of patients (4,5). In addition, the application of emerging DNA and RNA techniques, DNA methylation arrays and gene association studies has led to the identification of additional genetic and epigenetic alterations in pediatric ALL (6,7). However, studies have focused on the effects of an individual gene and have not considered that genes not only encode as individual genes or proteins, but also as sub-networks of interacting proteins within the larger human gene interaction network of the human genome (8). As a result, the exact mechanism of pediatric ALL remains to be elucidated.

Due to the difficulties of uncovering the pathogenesis of pediatric ALL via individual genes, the availability of large protein networks provides a method to partially address these challenges. Since large protein networks have become available for humans, a number of approaches have been used to extract relevant functional pathways based on the relevant databases (9). Subsequent to the measurement of sufficient protein interaction data, a large number of distinct functional pathways may be identified, which enables novel opportunities for elucidating pathways involved in major diseases and pathologies (10). In addition, numerous studies have hypothesized more effective methods to combine gene expression measurements over groups of genes that fall within the pathways (11-13).

The aim of the present study is to investigate the genes and pathways associated with pediatric ALL using a combination of analysis methods. The current study pre-processed downloaded datasets and identified differentially expressed

(DE) genes using the Bioconductor RankProd package (14). A co-expression network was constructed using the empirical Bayesian approach based on DE genes in ALL. Protein-protein networks were constructed and analyzed by enrichment and topological methods. In the enrichment analysis, the traditional expression analysis systematic explored (EASE) method of pathway analysis was compared with the network-based algorithm PathExpand.

Materials and methods

Identification of gene expression datasets. The microarray expression profiles of E-GEOD-26713 (15), E-GEOD-34670 and E-GEOD-42221 were obtained from the Array Express database (European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK), and were selected for identifying DE genes in pediatric ALL. E-GEOD-26713, which was included in the Human Genome U133 Plus 2.0 Array Platform (Affymetrix, Inc., Santa Clara, CA, USA), consisted of 7 normal controls and 117 pediatric ALL patients. E-GEOD-34670, which was included in the Human Genome U133A Array Platform (Affymetrix, Inc.), consisted of 12 normal controls and 25 pediatric ALL patients. E-GEOD-42221, which was also included in the Affymetrix Human Genome U133A Array Platform, consisted of 4 normal controls and 7 pediatric ALL patients.

Pre-processing of datasets. The expresso function of the Affy-Batch data structure was selected to control the quality of the gene microarray probe-level data (16). The rma method was used to perform background correction (17), and the observed perfect match (PM) probes were modeled as the sum of a normal noise component (N), with a mean indicated by μ and a variance indicated by σ^2, and an exponential signal component (S), with a mean indicated by α. To avoid the occurrence of

$$E\left(s|O=o\right) = a + b \, \frac{\phi\left(\frac{a}{b}\right) - \phi\left(\frac{o-a}{b}\right)}{\Phi\left(\frac{a}{b}\right) + \left(\frac{o-a}{b}\right) - 1}$$

negatives, the normal was truncated at zero. Since the observed intensity (O) had already been obtained, an adjustment was performed, as follows:
where a = s - μ - $\sigma^2\alpha$ and b = σ. ϕ and Φ were the standard normal distribution density and distribution functions, respectively. Mismatch (MM) probe intensities were not corrected by the rma method.

The quantile method, which was introduced by Bolstad *et al*, was used as the normalization method (18). The aim of the method is to provide each microarray chip with the same empirical distribution. The quantile normalization method uses the transformation $x'_i = F^{-1}\left(G\left(x_i\right)\right)$, where G is estimated by the empirical distribution of each array and F is estimated using the empirical distribution of the averaged sample quantiles.

The mas method was used to perform PM/MM correction (16). The ideal MM is subtracted from PM in this method. The ideal MM is always less than the corresponding PM, and therefore the MM may be subtracted from the PM without the risk of negative values.

The medianpolish summarization method was also used in the present study (14). A multichip linear model was fitted to

$$\log_2(PM_{ij}^{(k)}) = \alpha_i^{(k)} + \beta_j^{(k)} + \varepsilon_{ij}^{(k)}$$

the data from each probe set. For the probe set k with $i=1,\dots,I_k$ probes and data from $j=1,\dots,J$ arrays, the following model is used:
where α_i was the probe effect and β_j was the $\log_2$ expression value.

In total, 20,109, 12,493 and 12,493 genes were identified subsequent to pre-processing using the E-GEOD-26713, E-GEOD-34670 and E-GEOD-42221 datasets, respectively. The intersect function of the probe package was used to remove the genes identified by all three datasets, termed the common genes, in order to identify DE genes.

Analysis of DE genes. The RankProd package provides a novel and intuitive tool for detecting DE genes under two experimental conditions (14). The package modifies and extends the rank product method proposed by Breitling *et al* to integrate multiple microarray studies from various platforms (19). The significance of the detection was assessed using a non-parametric permutation test, and the associated P-value and false discovery rate (FDR) or percentage of false-positive (pfp) were included in the output, in addition to the genes that were detected by user-defined criteria. The RPadvance function was employed subsequent to pre-processing to identify the DE genes associated with pediatric ALL in the datasets. pfp≤0.01 was considered to indicate a significantly DE gene. In addition, a $\log_2$ fold change >2 in genes was considered to indicate a DE gene that required additional investigation.

Co-expression network construction. The co-expression network was constructed using the empirical Bayesian (EB) approach (20), which provided a FDR-controlled list of significant differential co-expression (DC) gene pairs, without sacrificing power (21). An m by n matrix of expression values was produced, where m was the number of genes or probes under consideration and n was the total number of microarrays over all conditions. These values were normalized to obtained X. For a conditions array with length n, all members of this array should take values in 1-K, where K was the total number of conditions. The ebPatterns function was used to define equal co-expression/DC classes.

From X and the conditions array, intra-group correlations were calculated for all $p=m^*(m\text{-}1)/2$ gene pairs. This step was accomplished through the makeMyD function and the resulting D matrix of correlations was p-by-K. The initializeHP function initialized the hyper parameters to identify the component normal mixture model that best fit the correlations of D subsequent to transformation. The zero-step version did not perform any expectation maximization (EM) calculations and instead used the initial estimates of the hyper parameters to generate posterior probabilities of DC (22). After the selected EM function had finished its computations, the fit of the model chosen by the EM with the data was assessed visually using the prior diagnostic function. The crit.fun function of meta analysis was used to provide a soft threshold (FDR≤0.05) and simulations indicated that the soft threshold was desirable.

Construction of the protein-protein interaction (PPI) network. To obtain the protein interaction data, a human PPI dataset obtained from the Search Tool for the Retrieval of Interacting Genes/proteins (23) was used. In addition, the PPI network was visualized using Cytoscape (24), a free software package for visualizing, modeling and analyzing the integration of bimolecular interaction networks with high-throughput expression data and other molecular states.

Enrichment analysis of PPI network and DE genes Method. Two enrichment analysis methods, PathExpand and EASE, were used to determine the interaction networks and DE genes in the present study.

Enrichment analysis based on PathExpand. Java enrichment of pathways extended to topology (JEPETTO) is a Cytoscape 3.1.0 plugin that performs integrative human gene set analysis (25). JEPETTO identifies functional associations between genes and known cellular pathways, and processes these associations using protein interaction networks and topological analysis.

EnrichNet and PathExpand were used to perform the enrichment analysis. EnrichNet mapped the input gene set onto a molecular interaction network and, using a random walk (26), scored the distances between the genes and pathways or processes in a reference database (26). This network-based association score (XD-score) was relative to the average distance to all pathways and represents a deviation, either positive or negative, from the average distance. PathExpand mapped the input pathway/process onto the human PPI network and extended the network with proteins that were strongly associated with the pathway nodes and increased the pathway compactness by connecting the disconnected members (27).

Enrichment analysis based on EASE. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a knowledge base for the systematic analysis of gene functions, linking genomic information with higher order functional information (28). In the present study, KEGG pathway enrichment analysis for DE genes was performed using the online Database for Annotation, Visualization and Integrated Discovery (DAVID) (29).

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}}$$

KEGG pathways with a P-value <0.01 were selected based on the EASE test implemented using DAVID (30). The principle of EASE was as follows:

where n is the number of background genes, a is the gene number of one gene set in the gene lists, $a + b$ is the number of genes in the gene list, including at least one gene set, $a + c$ is the gene number of one gene list in the background genes. a may be replaced by $a = a - 1$.

Topological analysis of the PPI network. Topological analysis was performed using TopoGSA (31). TopoGSA mapped the input gene set on an interaction network, computed the topological signature and compared the signature with the signatures of the pathways and processes in a reference database. The degree of a node, which represented a gene or protein, was the average number of edges or interactions adjacent to this node. The degree quantified the local topology of each gene, by calculating the sum of the number of genes adjacent to the target gene. The shortest path analysis was the most basic function of the network analysis, and the Dijkstra algorithm was chosen to calculate the shortest path length (32).

Results

Identification of DE genes. In total, 12,493 common genes were identified using the E-GEOD-26713, E-GEOD-34670 and E-GEOD-42221 datasets. The RankProd package was then used to analyze these genes. The DE genes were identified by assimilating a set of genes with the threshold of pfp≤0.01 and $\log_2$ fold change >2. In total, 398 DE genes in pediatric ALL were identified, and the top 100 DE genes were revealed in Table I. The most significantly DE genes were LIF, DEFA4, CEACAM8, S100A12 and ELANE.

Co-expression network construction. The simplest method for identifying DC gene pairs was conducting pair-specific tests for selected pairs within one condition, identifying those pairs that were strongly or significantly co-expressed and defining the DC pairs as those co-expressed. The EB approach provided a much-required method for identifying DC pairs while controlling a specified FDR. In the present study, 627 pairs of associated genes and the expression values were obtained using the EB approach, and Cytoscape was used to construct a co-expression network, which is shown in Fig. 1. The co-expression network revealed the complicated regulation pattern of DE genes and consisted of 272 nodes and 602 edges. MLLT11 demonstrated the highest degree with 38, followed by ADA, with a degree of 35. SP110 and CD177 possessed the same degree of 22.

PPI network construction and analysis. The PPI network was constructed based on the DE genes of pediatric ALL. In the network, nodes indicated a DE gene and the edges between the nodes indicated the interaction of genes. The PPI network was comprised of 294 nodes and 1,588 edges, which consisted of 398 DE genes (Fig. 2). Among the nodes, TSPO exhibited the highest degree of 62, followed by TYROBP (degree, 62), ITGB2 (degree, 60), TLR2 (degree, 55) and CTSS (degree, 54).

In order to further study the PPI network, the JEPETTO plugin was selected to perform enrichment and topological analysis based on the interaction networks, to exhibit the target gene set within the interaction environment and to identify possible gene cofactors and topologically-associated pathways and processes.

Enrichment analysis. The enrichment analysis found pathways significantly associated with the DE genes in terms of significance of interaction network distance between input gene set and pathways (XD-score), reported in Table II. Porphyrin metabolism (XD-score=1.077), starch and sucrose metabolism (XD-score, 0.788), nicotinate and nicotinamide metabolism (XD-score, 0.563), galactose metabolism (XD-score, 0.541) and hematopoietic cell lineage (XD-score, 0.536) were the top five pathways of pediatric ALL.

Table I. Top 100 differentially expressed genes in pediatric acute lymphoblastic leukemia.

No.	Gene
1	LTF
2	DEFA4
3	CEACAM8
4	S100A12
5	ELANE
6	PPBP
7	LCN2
8	TSPAN7
9	CRISP3
10	RCBTB2
11	MLLT11
12	S100P
13	LPAR6
14	CAMP
15	ANXA3
16	PLBD1
17	CSTA
18	MNDA
19	MS4A3
20	TOX
21	ALOX5
22	CLC
23	MZB1
24	BASP1
25	LYN
26	BPI
27	C5AR1
28	FCN1
29	AZU1
30	OLFM4
31	FPR1
32	SOX4
33	MMP8
34	PTGS2
35	MS4A1
36	GCA
37	ARG1
38	NCF2
39	ADA
40	PRTN3
41	ITGAM
42	TCN1
43	CLEC7A
44	PTX3
45	VCAN
46	CHI3L1
47	MARCKS
48	FAT1
49	CSF2RB
50	MMP9

Table I. Continued.

| No.	Gene
51	RAB31
52	MPO
53	GPR56
54	RNASE2
55	SCAI
56	ACSL1
57	CD1E
58	ALDH1A2
59	NCF4
60	ELOVL4
61	STOM
62	S100A9
63	IL1RN
64	CYBB
65	PGLYRP1
66	CTSS
67	CDA
68	RNASE3
69	TRAT1
70	HCK
71	PADI4
72	ORM1
73	LILRB2
74	IGLC1
75	CD14
76	LY86
77	RETN
78	CD1B
79	SKAP2
80	CFP
81	CTSG
82	CST7
83	SLC15A3
84	GPR65
85	B4GALT5
86	CD24
87	VNN2
88	GSAP
89	MAP1A
90	HCAR3
91	CD79B
92	ANPEP
93	CBFA2T3
94	ATHL1
95	MYO1F
96	BCL2A1
97	PF4
98	FGR
99	FCGR3B
100	ARHGEF10
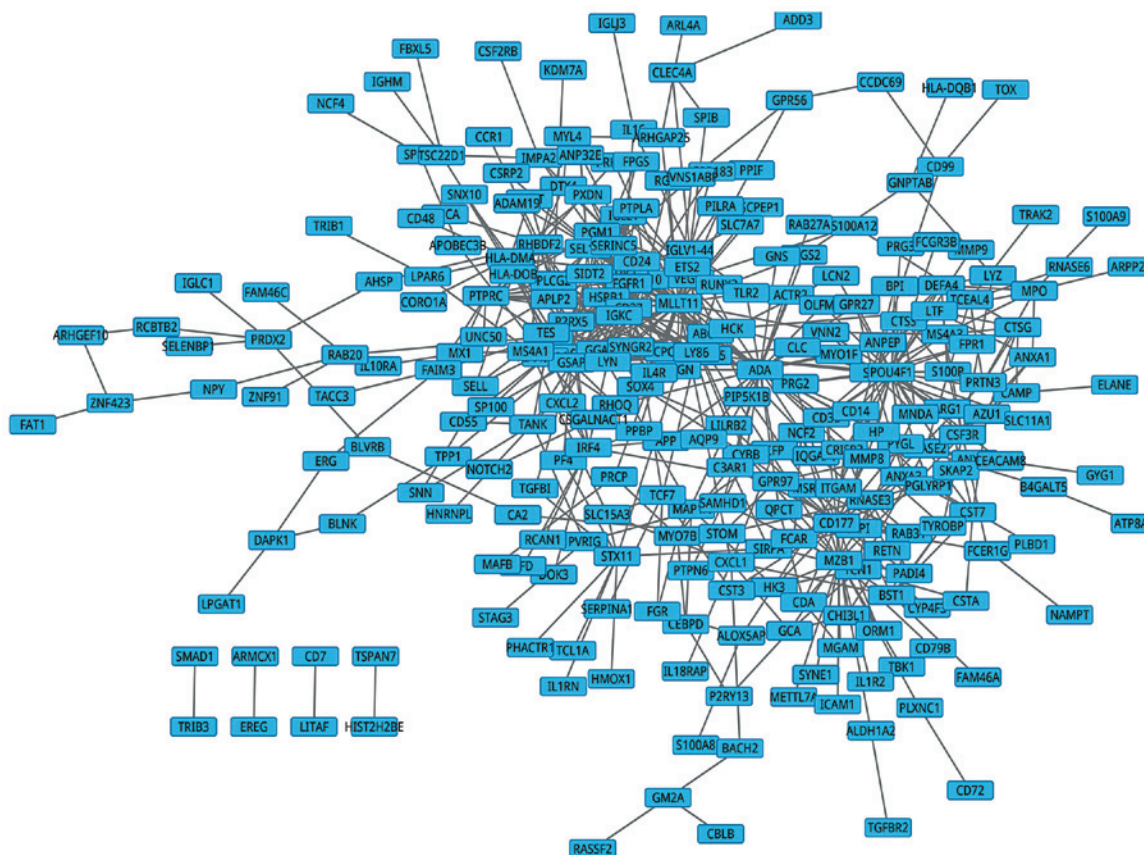
Figure 1. Co-expression network based on the DE genes in pediatric acute lymphoblastic leukemia. The network contained 272 nodes and 602 edges. Nodes referred to the DE genes and the edges between nodes indicated the interaction of genes. MLLT11 exhibited the highest degree of 38. DE, differentially-expressed.

Table II. Top 20 enrichment pathways of the protein-protein interaction network.

Pathway or process	XD-score
Porphyrin metabolism	1.07738
Starch and sucrose metabolism	0.78793
Nicotinate and nicotinamide metabolism	0.56282
Galactose metabolism	0.54143
Hematopoietic cell lineage	0.53555
Arachidonic acid metabolism	0.4894
Amino sugar and nucleotide sugar metabolism	0.47555
Asthma	0.46713
Leishmaniasis	0.45056
Primary immunodeficiency	0.44834
Phagosome	0.36488
Renin-angiotensin system	0.34249
Amoebiasis	0.33596
Drug metabolism-other enzymes	0.33555
Fc gamma R-mediated phagocytosis	0.29335
Salivary secretion	0.29058
Proximal tubule bicarbonate reclamation	0.28617
Natural killer cell mediated cytotoxicity	0.28415
B-cell receptor signaling pathway	0.27853
Lysosome	0.27703

XD-score, network-based association score.

Topological analysis. Topological analysis identified pathways with a pattern of interactions that were most similar to those in the input gene set and visually compared the topological properties of these pathways (Fig. 3). As shown in Fig. 3, the most notable environmental information processing pathway was the calcium signaling pathway and ECM receptor interaction. In addition, these pathways are mainly involved in cellular processes (cell adhesion molecules, cell communication, and hematopoietic cell lineage), cell metabolism (pyrimidine metabolism, inositol phosphate metabolism and porphyrin metabolism) and genetic information processing (RNA polymerases and DNA polymerases).

Comparison of enrichment analysis between EASE and PathExpand. Comparison of the pathways in PathExpand of JEPETTO and EASE of DAVID clarified that JEPETTO was a more powerful and efficient tool compared with traditional methods. The KEGG pathway database is a collection of manually drawn pathway maps for metabolism, genetic information processing, environmental information processing, including signal transduction, and various other cellular processes and human diseases. Table III reports the 12 enriched terms of the DE genes, and reveals that the most significantly DE pathway was hematopoietic cell lineage (P<0.001).

The enrichment analysis result of the PPI network is partly reported in Table II. The total number of terms was 177, which was markedly higher compared with the results of KEGG. Thus, enrichment analysis of the PPI network identified novel
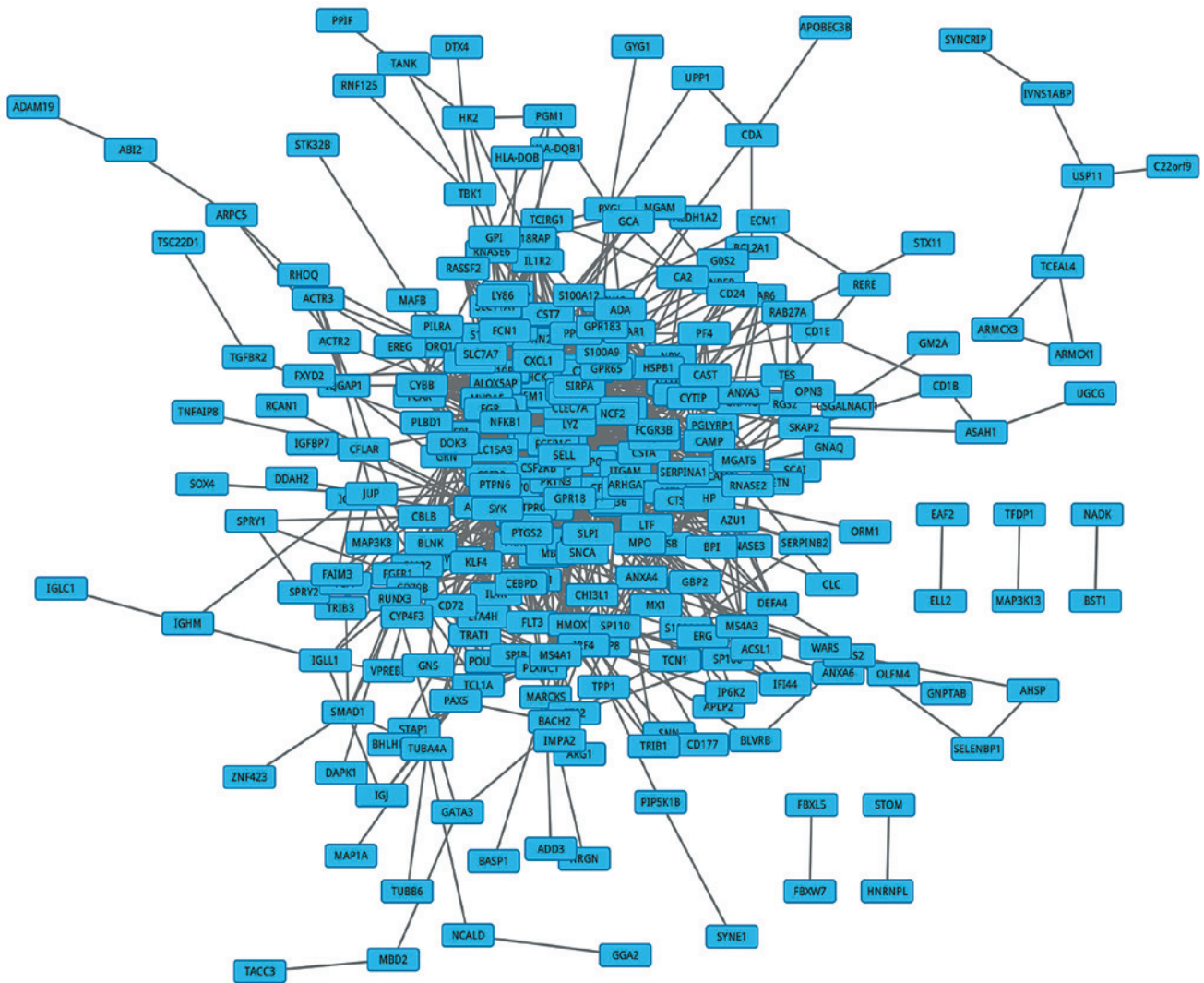
Figure 2. PPI network of pediatric acute lymphoblastic leukemia DE genes. In total, 294 nodes and 1,588 edges were identified in the PPI network, which included 398 DE genes. The nodes refer to the DE genes and the edges between nodes indicated the interaction of genes. Out of the nodes, TSPO exhibited the highest degree (degree, 62), followed by TYROBP (degree, 62), ITGB2 (degree, 60), TLR2 (degree, 55) and CTSS (degree, 54). PPI, protein-protein interaction; DE, differentially-expressed.

pathways that KEGG did not. The terms that were present in the DE gene and PPI networks were selected for additional analysis, as listed in Table III. Genes involved in the lysosome and asthma exerted a notable effect on the expression of ALL through KEGG analysis, but not through the JEPETTO analysis.

Discussion

In the present study, 398 DE genes were identified in pediatric ALL using the Bioconductor RankProd package, with the most significantly DE gene being LIF. The co-expression network was constructed using the EB method, based on the identified DE genes, and contained 272 nodes and 602 edges. The PPI network was constructed and the network was analyzed using the JEPETTO plugin. Gene enrichment and topological analyses revealed that hematopoietic cell lineage and porphyrin metabolism were the pathways most significantly associated with pediatric ALL. Comparison between the results of PathExpand and EASE revealed that PathExpand was able to

identify more processes closely associated with pediatric ALL than EASE.

LIF is involved in the induction of hematopoietic differentiation in normal and myeloid leukemia cells (33). LIF signaling regulates cellular processes to maintain the self-renewal and pluripotency of cells (34). The protein encoded by this gene is a pleiotropic cytokine that performs roles in various systems, particularly in lymphoblastic leukemia (35). It has previously been reported that an increased level of LIF may be a factor in the development of Cushing's syndrome (CS) in patients with ALL, and central nervous system infiltration should be considered in leukemic patients that have developed CS (36). Therefore, the LIF gene was concluded to be closely associated with pediatric ALL.

The characterization, description and extraction of information from networks have received attention in previous studies (37,38). The co-expression network constructed using the EB method in the present study consisted of 272 nodes and 602 edges. The most significantly co-expressed gene was MLLT11. MLLT11, which has been termed ALL1, HRX or

Table III. Comparison of pathways between expression analysis systematic explored and PathExpand.

| Pathway or process	EASE, P-value	PathExpand, q-value
Hematopoietic cell lineage	4.2345800×10^{-7}	0.00021
Lysosome	3.2443600×10^{-4}	0.26732
Asthma	2.0248050×10^{-3}	0.32768
B-cell receptor signaling	9.7318830×10^{-3}	0.18788
Natural killer cell mediated cytotoxicity	9.8639250×10^{-3}	0.06947
Starch and sucrose metabolism	1.0380592×10^{-2}	0.03252
Fc gamma R-mediated phagocytosis	1.0701949×10^{-2}	0.17438
Cytokine-cytokine receptor interaction	2.0008775×10^{-2}	0.17438
Primary immunodeficiency	2.4587264×10^{-2}	0.14603
Cell adhesion molecules	2.5103397×10^{-2}	0.39694
Leukocyte transendothelial migration	3.4812818×10^{-2}	0.17438
Galactose metabolism	4.9005430×10^{-2}	0.27336
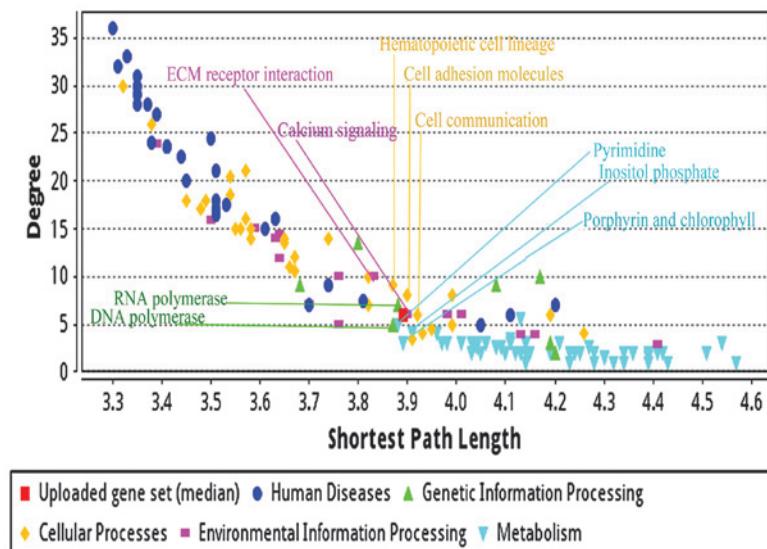
Figure 3. Comparative analysis of the topological properties of pathways in acute lymphoblastic leukemia. The red square next to the calcium signaling pathway represented the target network. ECM, extracellular matrix.

MLL, is located on chromosome 1 band q21 and has been demonstrated to be fused with a number of translocation partners in patients with leukemia (39). MLLT11 is a poor prognostic biomarker for pediatric acute myeloid leukemia, and MIR29B directly regulates MLLT11 expression (40). Complex duplication or translocation events of the MLLT11 locus have since been reported in hematological malignancies and high MLLT11 mRNA levels have been reported as markers of poor prognosis in leukemia and myelodysplastic syndromes (41,42). Therefore, it was concluded in the present study that MLLT11 exerts an important effect on pediatric lymphoblastic leukemia.

The JEPETTO plugin integrates three network-centric human gene set analysis methods under the single interface of the Cytoscape 3.1.0 environment. This was then used to perform enrichment and topological analyses based on the interaction networks, and exhibit the target gene set within the interaction environment. Possible gene cofactors and topologically associated pathways and processes that were unlikely to be detected using traditional term-based analysis were also identified (25). Comparison of the pathways identified by PathExpand and EASE confirmed that JEPETTO was a more powerful and efficient tool compared with traditional methods. As aforementioned, hematopoietic cell lineage and porphyrin metabolism were the pathways most significantly associated with pediatric ALL. Bonnet and Dick previously indicated that human ALL is organized as a hierarchy that originates from a primitive hematopoietic cell (43). It is possible that a hematopoietic cell lineage-specific proteoglycan is involved in lymphoid cell adherence and activation (44), which is in accordance with the finding that cell adhesion molecules comprised a significant cellular process of pediatric ALL in the present study. However, bone marrow hematopoietic stem cells are likely to be distinguished by a distinct B-cell lineage commitment (45). It has been reported that macrophages and dendrite cells are key components of cellular immunity and

are considered to originate and renew from hematopoietic stem cells (46). Therefore, the results of the gene enrichment and topological analyses are valid and demonstrated they are associated with genes and pathways associated with pediatric ALL.

In conclusion, 398 genes were identified as DE genes in pediatric ALL, and the co-expression and PPI networks were constructed based on these genes. Gene enrichment and topological analyses for the interaction networks identified the most significantly DE genes, such as MLLT11, and pathways, such as the hematopoietic cell lineage, associated with pediatric ALL. The present study may provide further insight of the diagnosis and treatment of pediatric ALL at the molecular level.

References

1. Carroll WL, Bhojwani D, Min DJ, Raetz E, Relling M, Davies S, Downing JR, Willman CL and Reed JC: Pediatric acute lymphoblastic leukemia. Hematology Am Soc Hematol Educ Program: 102-131, 2003.
2. Abboud MR, Ghanem K and Muwakkit S: Acute lymphoblastic leukemia in low and middle-income countries: disease characteristics and treatment results. Curr Opin Oncol 26: 650-655, 2014.
3. Chang TT and Lin PC: Treatment of pediatric acute lymphoblastic leukemia and recent advances. In: Novel Aspects in Acute Lymphoblastic Leukemia. Faderl S (ed). InTech, East Providence, pp101-116, 2011.
4. Stow P, Key L, Chen X, Pan Q, Neale GA, Coustan-Smith E, Mullighan CG, Zhou Y, Pui CH and Campana D: Clinical significance of low levels of minimal residual disease at the end of remission induction therapy in childhood acute lymphoblastic leukemia. Blood 115: 4657-4663, 2010.
5. Katsibardi K, Moschovi MA, Braoudaki M, Papadhimitriou SI, Papathanasiou C and Tzortzatou-Stathopoulou F: Sequential monitoring of minimal residual disease in acute lymphoblastic leukemia: 7-year experience in a pediatric hematology/oncology unit. Leuk Lymphoma 51: 846-852, 2010.
6. Schotte D, Pieters R and Den Boer ML: MicroRNAs in acute leukemia: From biological players to clinical contributors. Leukemia 26: 1-12, 2012.
7. Yu L, Slovak ML, Mannoor K, Chen C, Hunger SP, Carroll AJ, Schultz RA, Shaffer LG, Ballif BC and Ning Y: Microarray detection of multiple recurring submicroscopic chromosomal aberrations in pediatric T-cell acute lymphoblastic leukemia. Leukemia 25: 1042-1046, 2011.
8. Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, Neumüller RA, Mohr SE and Perrimon N: Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. Nat Methods 11: 94-99, 2014.
9. Cui D, Ou S and Patel S: Protein-spanning water networks and implications for prediction of protein-protein interactions mediated through hydrophobic effects. Proteins 82: 3312-3326, 2014.
10. Yang Z, Zheng R, Gao Y and Zhang Q: Gene expression profiles on predicting protein interaction network and exploring of new treatments for lung cancer. Mol Biol Rep 41: 8203-8210, 2014.
11. Pavlidis P, Qin J, Arango V, Mann JJ and Sibille E: Using the gene ontology for microarray data mining: A comparison of methods and application to age effects in human prefrontal cortex. Neurochem Res 29:1213-1222, 2004.
12. Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC and Conklin BR: MAPPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. Genome Biol 4: R7, 2003.
13. Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS and Park PJ: Discovering statistically significant pathways in expression profiling studies. Proc Natl Acad Sci USA 102: 13544-13549, 2005.
14. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL and Chory J: RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 22: 2825-2827, 2006.
15. Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, Vuerhard M, Buijs-Gladdines J, Kooi C, Klous P, *et al*: Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. Cancer Cell 19: 484-97, 2011.
16. Bolstad B: affy: Built-in Processing Methods. 2013. www.bioconductor.org/packages/release/bioc/vignettes/affy/inst/doc/builtinMethods.pdf. Accessed July 11, 2013.
17. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B and Speed TP: Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31: e15, 2003.
18. Bolstad BM, Irizarry RA, Astrand M and Speed TP: A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19: 185-193, 2003.
19. Breitling R, Armengaud P, Amtmann A and Herzyk P: Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 573: 83-92, 2004.
20. Dawson JA, Ye S and Kendziorski C: R/EBcoexpress: an empirical Bayesian framework for discovering differential co-expression. Bioinformatics 28: 1939-1940, 2012.
21. Dawson JA and Kendziorski C: An empirical Bayesian approach for identifying differential coexpression in high-throughput experiments. Biometrics 68: 455-465, 2012.
22. Moon TK: The expectation-maximization algorithm. IEEE Signal Proc Mag 13: 47-60, 1996.
23. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, *et al*: STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43: D447-452, 2015.
24. Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27: 431-432, 2011.
25. Winterhalter C, Widera P and Krasnogor N: JEPETTO: A cytoscape plugin for gene set enrichment and topological analysis based on interaction networks. Bioinformatics 30: 1029-1030, 2014.
26. Glaab E, Baudot A, Krasnogor N, Schneider R and Valencia A: EnrichNet: Network-based gene set enrichment analysis. Bioinformatics 28: i451-i457, 2012.
27. Glaab E, Baudot A, Krasnogor N and Valencia A: Extending pathways and processes using molecular interaction networks to analyse cancer genome data. BMC Bioinformatics 11: 597, 2010.
28. Kanehisa M and Goto S: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30, 2000.
29. Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57, 2009.
30. Wang X and Simon R: Microarray-based cancer prediction using Single genes. BMC Bioinformatics 12: 391, 2011.
31. Glaab E, Baudot A, Krasnogor N and Valencia A: TopoGSA: network topological gene set analysis. Bioinformatics 26: 1271-1272, 2010.
32. Qu H, Yi Z and Yang SX: Efficient shortest-path-tree computation in network routing based on pulse-coupled neural networks. IEEE Trans Cybern 43: 995-1010, 2013.
33. Waring PM: Leukemia inhibitory factor. In: Colony-Stimulating Factors: Molecular and Cellular Biology. Garland JM, Quesenberry PJ and Hilton DJ (eds). 2nd edition. Marcel Dekker, New York, NY: pp467-513, 1997.
34. Kuphal S, Wallner S and Bosserhoff AK: Impact of LIF (leukemia inhibitory factor) expression in malignant melanoma. Exp Mol Pathol 95: 156-165, 2013.
35. Telugu BP, Ezashi T, Sinha S, Alexenko AP, Spate L, Prather RS and Roberts RM: Leukemia inhibitory factor (LIF)-dependent, pluripotent stem cells established from inner cell mass of porcine embryos. J Biol Chem 286: 28948-28953, 2011.
36. Sahin NM, Avci Z, Malbora B, Abaci A, Kinik ST and Ozbek NY: Cushing syndrome related to leukemic infiltration of the central nervous system: a case report and a possible role of LIF. J Pediatr Endocrinol Metab 26: 967-970, 2013.
37. Bhalla US and Iyengar R: Emergent properties of networks of biological signaling pathways. Science 283: 381-387, 1999.
38. Strogatz SH: Exploring complex networks. Nature 410: 268-276, 2001.
39. Tse W, Zhu W, Chen H and Cohen A: A novel gene, AF1q, fused to MLL in t (1; 11) (q21; q23), is specifically expressed in leukemic and immature hematopoietic cells. Blood 85: 650-656, 1995.

40. Xiong Y, Li Z, Ji M, Tan AC, Bemis J, Tse JV, Huang G, Park J, Ji C, Chen J, et al: MIR29B regulates expression of MLLT11 (AF1Q), an MLL fusion partner and low MIR29B expression associates with adverse cytogenetics and poor overall survival in AML. Br J Haematol 153: 753-757, 2011.
41. Ussowicz M, Jaśkowiec A, Meyer C, Marschalek R, Chybicka A, Szczepański T and Haus O: A three-way translocation of MLL, MLLT11 and the novel reciprocal partner gene MYO18A in a child with acute myeloid leukemia. Cancer Genet 205: 261-265, 2012.
42. Parcelier A, Maharzi N, Delord M, Robledo-Sarmiento M, Nelson E, Belakhdar-Mekid H, Pla M, Kuranda K, Parietti V, Goodhardt M, et al: AF1q/MLLT11 regulates the emergence of human prothymocytes through cooperative interaction with the Notch signaling pathway. Blood 118: 1784-1796, 2011.
43. Bonnet D and Dick JE: Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 3: 730-737, 1997.
44. Toyama-Sorimachi N, Sorimachi H, Tobita Y, Kitamura F, Yagita H, Suzuki K and Miyasaka M: A novel ligand for CD44 is serglycin, a hematopoietic cell lineage-specific proteoglycan possible involvement in lymphoid cell adherence and activation. J Biol Chem 270: 7437-7444, 1995.
45. Ghosn EE, Yamamoto R, Hamanaka S, Yang Y, Herzenberg LA, Nakauchi H and Herzenberg LA: Distinct B-cell lineage commitment distinguishes adult bone marrow hematopoietic stem cells. Proc Natl Acad Sci U S A 109: 5394-5398, 2012.
46. Schulz C, Gomez Perdiguero E, Chorro L, Szabo-Rogers H, Cagnard N, Kierdorf K, Prinz M, Wu B, Jacobsen SE, Pollard JW, et al: A lineage of myeloid cells independent of Myb and hematopoietic stem cells. Science 336: 86-90, 2012.