

RESEARCH

Open Access



Identification of novel drug targets for diamond-blackfan anemia based on *RPS19* gene mutation using protein-protein interaction network

Abbas Khan¹, Arif Ali¹, Muhammad Junaid¹, Chang Liu¹, Aman Chandra Kaushik¹, William C. S. Cho² and Dong-Qing Wei^{1*}

From The 11th International Conference on Systems Biology (ISB 2017) Shenzhen, China. 18-21 August 2017

Abstract

Background: Diamond-Blackfan anemia (DBA) is a congenital erythroid aplasia that usually presents in infancy. In order to explore the molecular mechanisms of wild and mutated samples from DBA patients were exposed to bioinformatics investigation. Biological network of differentially expressed genes was constructed. This study aimed to identify novel therapeutic signatures in DBA and uncovered their mechanisms. The gene expression dataset of GSE14335 was used, which consists of 6 normal and 4 diseased cases. The gene ontology (GO), as well as Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed, and then protein-protein interaction (PPI) network of the identified differentially expressed genes (DEGs) was constructed by Cytoscape software.

Results: A total of 607 DEGs were identified in DBA, including 433 upregulated genes and 174 downregulated genes. GO analysis results showed that upregulated DEGs were significantly enriched in biological processes, negative regulation of transcription from RNA polymerase II promoter, chemotaxis, inflammatory response, immune response, positive regulation of cell proliferation, negative regulation of cell proliferation, response to mechanical stimulus, positive regulation of cell migration, response to lipopolysaccharide, and defence response. KEGG pathway analysis revealed the TNF signalling pathway, Osteoclast differentiation, Chemokine signalling pathway, Cytokine-cytokine receptor interaction, Rheumatoid arthritis, Biosynthesis of amino acids, Biosynthesis of antibiotics and Glycine, serine and threonine metabolism. The top 10 hub genes, AKT1, IL6, NFKB1, STAT3, STAT1, RAC1, EGR1, IL8, RELA, RAC3, mTOR and CCR2 were identified from the PPI network and sub-networks.

Conclusion: The present study flagged that the identified DEGs and hub genes enrich our understanding of the molecular mechanisms underlying the development of DBA, and might shine some lights on identifying molecular targets and diagnostic biomarkers for DBA.

Keywords: Diamond-Blackfan anemia (DBA), Microarray, Differentially expressed gene, KEGG pathways, Gene expression omnibus

* Correspondence: dqwei@sjtu.edu.cn

¹Department of Bioinformatics and Biostatistics, College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
Full list of author information is available at the end of the article



Background

Diamond-Blackfan anemia (DBA) is a rare disease mostly affecting child during infancy. This disease is characterized by defect in red blood cells which result in the reduction or absence of erythroid precursors in bone marrow. The cephalic area is affected in most of the cases but other parts such as heart, limb and tract were also reported. This disease is a complicated disorder, however, its evolution remains erratic [1, 2]. Diamond-Blackfan anemia (DBA) is categorized as a rare genetic diseases characterized by cancer predisposition, bone marrow failure, pro-apoptotic haematopoiesis and congenital anomalies. It is also known as the inherited bone marrow failure syndromes (IBMFS) [3–5]. A wide range of mutations have been identified in DBA patients, from missense to nonsense mutations and from partial to complete deletion of one allele. DBA is the only disease known to be caused by defect in ribosomal protein [1, 2, 6, 7]. To date, DBA was reported as sporadic but latest cases approximately 45% of cases are familial [8]. DBA can be caused by mutations in the RPL5, RPL11, RPL35A, RPS7, RPS10, RPS17, RPS19, RPS24, and RPS26 genes [9]. However, RPS19 mutation account for 25% of the cases. RPS19 is one of the important proteins which interact with 18S rRNA and form the ribosomal small subunit 40S and perform its normal functioning of translation [1, 6]. Other than protein synthesis, RPS19 is also involved in many functions such as attraction of monocytes. Loss of non-ribosomal function could be the possible cause of DBA [10]. There is a case control study reported that patient with DBA has reduced RPS gene expression and so ribosomal synthesis defect is the underlying cause of DBA [11]. Some missense mutations in RPS19 was also reported to be involved in the reduction in the intracellular transport and stability of RPS19. With such mutations in cells are also reported to reduce differentiation and proliferation of erythroleukemic cell lines or CD34+ [12–14]. Furthermore, Chiocchetti et al. reported that RPS19 interact with PIM-1 oncoprotein in DBA. RPS19 with FGF-2 and with some other proteins of unknown function named S19-BP are also reported to interact with each others [15–17]. To date no definitive treatment is available for this complicated disease. Steroid therapy is widely used for the treatment which is responsive in 60% of the patients but sometimes patients suffered from severe complications such as iron overdose. DBA patients with steroid resistance required blood transfusion for a life time. Some cases successfully treated with interleukin-3 supplement [18], bone marrow transplantation [19] but not easy and affordable. It is necessary to find a feasible and cost-effective way of treatment.

Network based gene expression profiling is a proposed methodology to discover therapeutic signatures by

integrating multiple factors including disease genes, gene expression intensities and proteins network [20–22]. Systems biology is one among such approaches, which rely on a global approach by analyzing the whole interacting network rather than a single protein, gene or enzyme to be analyzed. Systems biology reported that cellular protein does not function alone but these genes/proteins are clustered together to form an interconnected molecular networks to perform a specific function. Information regarding the individual gene/protein and function of a living system can be obtained through traditional approaches but systems biology alternatively access the mechanism at a systemic level. For this reason, systems biology employ biological relationships to construct a network consist of nodes (protein, enzyme, genes) interacting with another partner [23–26].

Proteomics and transcriptomic modeling of molecular networks from microarray data to discover potential biomarkers in DBA has not yet been resolved. Therefore, we used gene expression data by using computational systems biology approach based on microarray dataset analysis to identify the possible therapeutic gene/protein signatures for the treatment of DBA. Initially statistical approaches were applied to identify differentially expressed genes (DEGs) based on nominal *p*-value and false discovery rate (FDR). Furthermore, the subnetwork modules were constructed and the obtained DEGs were analyzed for the biological processes, molecular components, KEGG pathways and cellular component analysis. Finally, we mapped out hub genes from the differentially expressed genes network that could act as possible drugs targets.

Methods

Microarray data

Differential analysis of one channel microarray data (Accession No:GSE14335) [27] based on the GPL570 Affymetrix Human Genome U133A plus 2.0 Array platform was retrieved from NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>) [28]. The dataset submitted by Avondo et al. 2009 consists of ten samples including four DBA and six control samples. Pre-processing and differential analysis of the dataset was carried out to conclude the final results. In the pre-processing, a collectively phenotypic and the samples information stored in .cel files were combined to precisely approach the differential expression step. The expression intensity values were subjected to quantile normalization to remove the noise from the microarray data [29] using an integrated GCRMA [30] package of Rstudio v 3.0.2 [31]. Quantile normalization is perhaps the most widely implemented method for considering microarray data produced by Affymetrix GeneChip platform. Following the normalization, an

Table 1 The dataset divided into two groups based on mutation

No	Group No	Number of Samples	Phenotype
1	Group I	4	RPS19 Mutated
2	Group II	6	Normal

expression set was built which contain information regarding the assay probes, features, phenotype and the experimental setup. Based on the presence or absence of RPS19 gene mutation, the data was divided into two groups including the control and diseased (Table 1). The methodological flow of the work as defined by (ref [32]) is given in the Fig. 1.

DEGs screening

The expression set contains the log expression values was used to design a model matrix to accurately adjust the total 54,675 probes according to their phenotypic information. This matrix includes a coefficient for the diseased vs normal, rather than applying another approach to

assign a separate coefficient to each condition and then using a contrast to extract the difference. A pertinent statistical test such as Student’s *t*-test [33] is applied to these *normalized* gene expression levels. The resulting *p*-values are adjusted by a multiple testing procedure (MTP) in order to control certain quantity of per-family Type I error, such as family-wise error rate (FWER) [34–36] and false discovery rate (FDR) [37]. Differentially expressed genes are identified based on a pre-specified threshold of adjusted *p*-values. More detailed introduction of statistical methods for detecting differentially expressed genes can be found in [32, 38, 39]. Using Benjamin-Hochberg [40] multiple testing method the $FDR < 0.05$ with fold-change ≥ 1 and the adjusted *p*-value < 0.05 was selected as the threshold for DEGs identification.

Gene ontology and pathway enrichment analysis of DEGs

Gene ontology (GO) is a useful tool to annotate genes and its products. Attributes of high throughput genomics and transcriptomics data could be obtained through GO

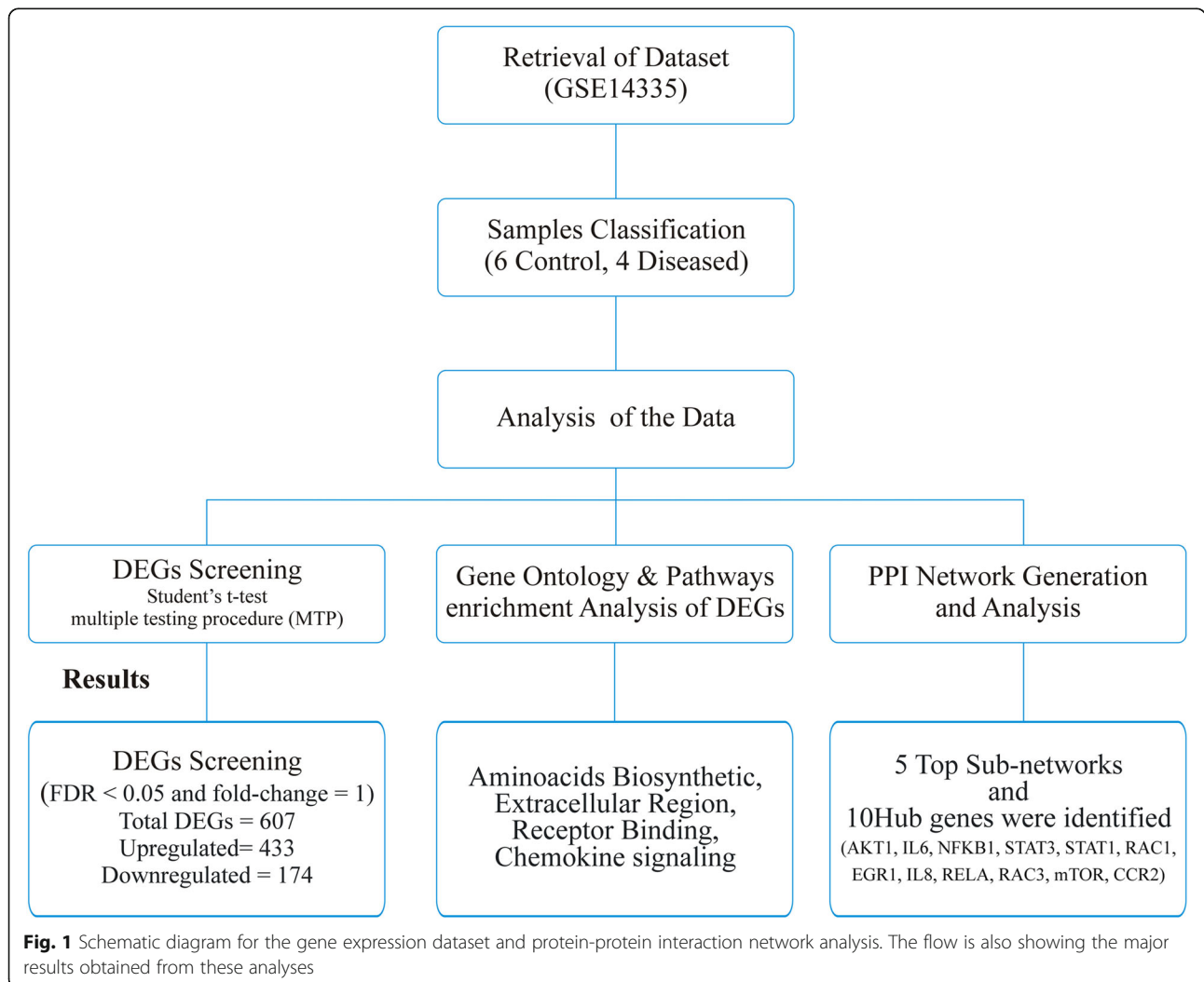


Fig. 1 Schematic diagram for the gene expression dataset and protein-protein interaction network analysis. The flow is also showing the major results obtained from these analyses

analysis [41, 42]. KEGG (<http://www.genome.jp/kegg/pathway.html>) provides a detail information about gene function and pathways and also link the genomics data with the high order functional information [43]. For the functional analysis DAVID (The Database for Annotation, Visualization and Integrated Discovery) (<https://david.ncifcrf.gov/>) is an essential online server which can functionally annotate genes with high success [44]. Here we used DAVID to label the mapped DEGs to their functional class, pathways and GO enrichment processes. A p -value < 0.05 was defined as significant threshold for the essential annotations.

Protein-protein interaction (PPI) network generation

PPI network of the total DEGs was retrieved from STRING database [45] and GeneMANIA [46]. Current STRING database contains 9,643,763 proteins and 2031 total organisms while GeneMANIA owned 747 data sets. The interactions were loaded into the Cytoscape v 3.4 [47] and were analysed using different integrated functions.

Network topological parameters

Numbers of topological parameters are available to analyse and compare the network. Cytoscape is a freely available software which provides an integrated function “NetworkAnalyzer” to analyse the gene/protein network. Here we also used “NetworkAnalyzer” to calculate the parameters for all the constructed networks. The primary parameters which were analysed includes power law of node distribution, distribution of node degree, clustering coefficient, network centralization and density to distinguish the three constructed networks [48].

Hub genes identification

Cytohubba is a well-known integrated plugin in Cytoscape which analyse the network features and rank the nodes in the network accordingly [49]. Cytohubba uses 11 different methods to analyse the network including the identification of hub genes/nodes in a network. We used Cytohubba to find out the hub genes in our constructed network of total DEGs which could be the possible new drug targets for the treatment of DBA.

Molecular complex detection analysis (MCODE)

MCODE is an automated algorithm which can be used as an integrated plugin in Cytoscape provides a way to identify highly connected dense subnetwork in a PPI/gene networks [50]. To cluster the subnetworks in the total DEGs we also used MCODE. The interconnected nodes in the subgraphs were identified and selected for further analysis

based on number of node. We used $n > 10$ as a parameter for selecting highly interconnected sub-networks.

Results

Identification of differentially expressed genes

Analysing the network calculated $R^2 = 0.26$ which shows that the network is scale-free. The complex biological system is composed of thousands of genes and its products. These genes and products interact randomly and form a complicated network.. Microarray expression analysis may pose many regular variations. To overcome these variations Normalization process is usually carried out. These variations may reveal different expression level then actual which is a major problem in gene expression analysis. Statistical models are proposed to perform normalization. Here, we also used “ArrayQualityMetrics” and “GCRMA” to perform normalization. The density plots and boxplot of the data before and after normalization is shown in the Fig. 2. Based on the defined criteria (FDR < 0.05 and fold-change ≥ 1) we compared the two types of samples in the dataset that result a total of 607 differentially expressed genes (Additional file 1: Table S1) using R Bioconductor. Among the total DEGs 433 were upregulated while the rest 174 were downregulated. Using this criteria we identified some novel DEGs with important functions.

GO term enrichment analysis of the up and downregulated DEGs

Functional exploration of functional genomics data can be obtained by utilizing GO analysis process. Here, we also utilized a well-known functional annotation database DAVID to understand the GO processes of identified DEGs. For enrichment analysis of DEGs we selected GO biological processes, GO Molecular function, cellular component and KEGG pathways analysis. In biological processes negative regulation of transcription from RNA polymerase II promoter, chemotaxis, inflammatory response, immune response, positive regulation of cell proliferation, negative regulation of cell proliferation, response to mechanical stimulus, positive regulation of cell migration, response to lipopolysaccharide, response to cytokine, regulation of cell proliferation, positive regulation of tyrosine phosphorylation of Stat3 protein, positive regulation of MAPK cascade, positive regulation of transcription from RNA polymerase II promoter, positive regulation of smooth muscle cell proliferation, chemokine-mediated signaling pathway, cellular response to lipopolysaccharide, cellular response to tumor necrosis factor, cellular response to hypoxia from upregulated while downregulated genes were found to be involved in L-serine biosynthetic process and cellular amino acid biosynthetic process. In molecular function protein binding,

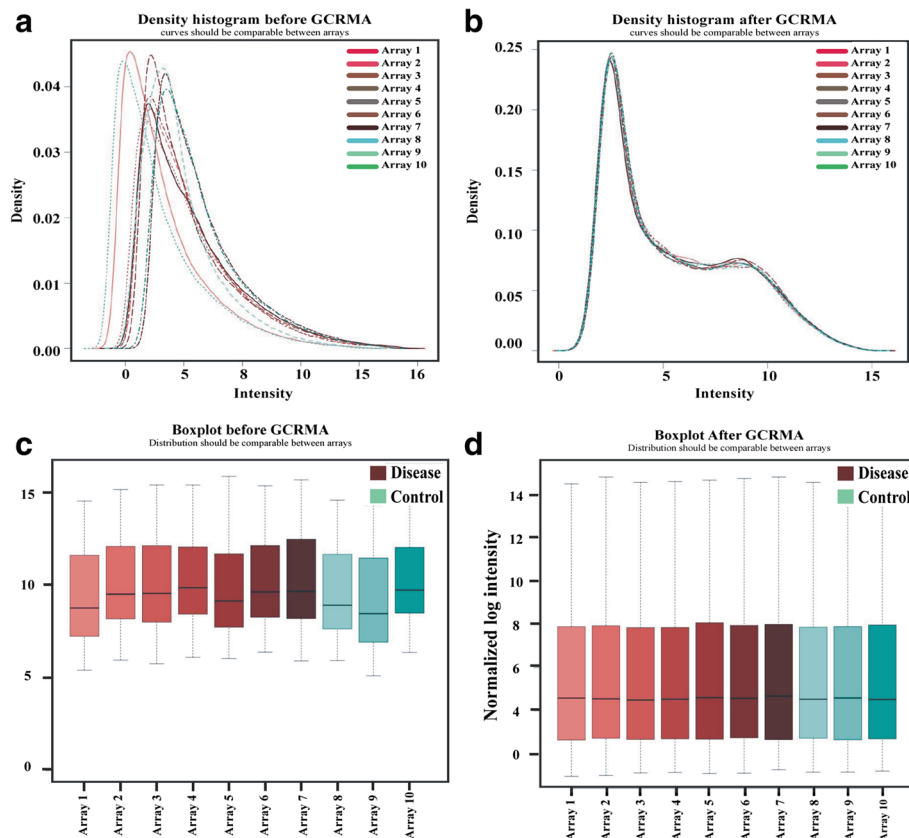


Fig. 2 The figure shows the quality of the data before and after normalization. **a** and **b** are the density plot. **a** is before the normalization in which the data is not good enough to be compared while the **(b)** graph is showing the overlapped density plot. **c** and **d** are the normalization plot which is also revealing the quality of the data is comparable

receptor binding and chemokine activity were only observed in upregulated genes. Upregulated genes were enriched in TNF signaling pathway, osteoclast differentiation, chemokine signaling pathway, cytokine-cytokine receptor interaction, rheumatoid arthritis, while the only three KEGG pathways found in the downregulated genes were biosynthesis of amino acids, biosynthesis of antibiotics and glycine, serine and threonine metabolism. A detail of these processes including the *p*-value and FDR (0.05) are given in the Table 2 and Fig. 3.

PPI network analysis

PPI network is an important approach towards the understanding of biological problems and its elucidation. Nodes and edges are the important combinations to construct a network. The mapped DEGs identified by comparing the data from control and diseased samples were visualized in Cytoscape and hub genes were identified using Cytohubba. Identification of hub genes shown in the Fig. 4 was followed by the identification of highly interconnected dense sub-networks. The *k*-score was set 2.0, the node score cut off was set 0.2 while the

maximum depth for seed node was set 100 for efficiency. MCODE was parameterized by using node degree > 10 which identified a total of 12 modules. Among the 12 modules only 5 were found to be under the defined parameters (nodes > 10 & node score > 2.0). A total of 157 proteins and 731 interactions were observed in these sub-network modules. Initial confirmation of the hub genes was carried out by comparing genes from the sub-networks with the identified hub genes, which revealed consistency in the results and confirm the existence of almost all the hub genes in the subnetworks. Final functional annotation of sub-networks and hub genes was carried out to confirm the reliability of our initial results.

The five different subnetworks generated were as shown in the Fig. 5 and their properties including score, number of proteins and interaction are given in the Table 3. The enrichment analysis of these sub-networks also confirm the validity and they were found to be involved in immune response, inflammatory responses, response to cytokines, chemotaxis, positive and negative regulation of RNA polymerase II transcription activities.

Table 2 Gene Ontology/functional enrichment analysis of up and downregulated DEGs associated with DBA patients. The FDR and *p*-value cut-off criteria was set > 0.05. The table is enriched with GO biological processes, Molecular Functions and KEGG pathways

Category	Term	Count	<i>p</i> -value	False discovery rate
GO-BP:0000122	negative regulation of transcription from RNA polymerase II promoter	44	6.19E-11	1.10E-07
GO-BP:0006935	chemo taxis	14	7.72E-07	0.001365
GO-BP:0006954	inflammatory response	30	3.82E-10	6.77E-07
GO-BP:0006955	immune response	28	6.64E-08	1.17E-04
GO-BP:0008284	positive regulation of cell proliferation	28	5.10E-07	9.02E-04
GO-BP:0008285	negative regulation of cell proliferation	25	1.01E-06	0.001791
GO-BP:0009612	response to mechanical stimulus	10	1.97E-06	0.003493
GO-BP:0030335	positive regulation of cell migration	16	3.39E-06	0.005999
GO-BP:0032496	response to lipopolysaccharide	18	2.31E-08	4.10E-05
GO-BP:0034097	response to cytokine	10	6.52E-07	0.001153
GO-BP:0042127	regulation of cell proliferation	15	1.68E-05	0.029732
GO-BP:0042517	positive regulation of tyrosine phosphorylation of Stat3 protein	9	5.71E-07	0.001010
GO-BP:0043410	positive regulation of MAPK cascade	11	3.88E-06	0.006861
GO-BP:0045944	positive regulation of transcription from RNA polymerase II promoter	45	2.14E-07	3.78E-04
GO-BP:0048661	positive regulation of smooth muscle cell proliferation	12	1.95E-08	3.45E-05
GO-BP:0070098	chemokine-mediated signaling pathway	10	9.53E-06	0.016859
GO-BP:0071222	cellular response to lipopolysaccharide	13	2.14E-06	0.003778
GO-BP:0071356	cellular response to tumor necrosis factor	14	2.30E-07	4.07E-04
GO-BP:0071456	cellular response to hypoxia	11	1.80E-05	0.031815
GO-BP:0006564	L-Serine biosynthetic Process	4	1.35E-06	0.002131
GO-BP:0008652	Cellular amino acid biosynthetic process	5	3.07E-05	0.048645
GO-CC:0005576	extracellular region	60	1.27E-07	1.71E-04
GO-CC:0005615	extracellular space	61	6.12E-11	8.24E-08
GO-CC:0048471	perinuclear region of cytoplasm	30	3.33E-06	0.004483
GO-MF:0005515	protein binding	209	1.90E-07	2.80E-04
GO-MF:0005102	receptor binding	24	2.07E-07	3.05E-04
GO-MF:0008009	chemokine activity	9	3.05E-06	0.004492
KEGG_Pathway	TNF signaling pathway	22	7.91E-13	1.01E-09
KEGG_Pathway	Osteoclast differentiation	16	3.16E-06	0.004022
KEGG_Pathway	Chemokine signaling pathway	18	1.52E-05	0.019366
KEGG_Pathway	Cytokine-cytokine receptor interaction	20	2.02E-05	0.025631
KEGG_Pathway	Rheumatoid arthritis	12	2.97E-05	0.037766
KEGG_Pathway	Biosynthesis of amino acids	9	3.22E-07	0.000387
KEGG_Pathway	Biosynthesis of antibiotics	12	3.11E-06	0.003751
KEGG_Pathway	Glycine, serine and threonine metabolism	6	2.62E-05	0.031616

Discussion

Many different computational approaches such as Single Nucleotide polymorphism (SNPs) analysis, Genome wide association studies (GWAS), disease, pathway biomarkers, network module biomarkers and especially gene expression microarray analysis are available to

analyze different genomics data and access significant information regarding the disease condition, ranging from diagnosis to treatment [51]. All these approaches can be used to access the available diverse data from different levels including genomics, proteomics, transcriptomics, metagenomics, epigenomics, and metabolomics to

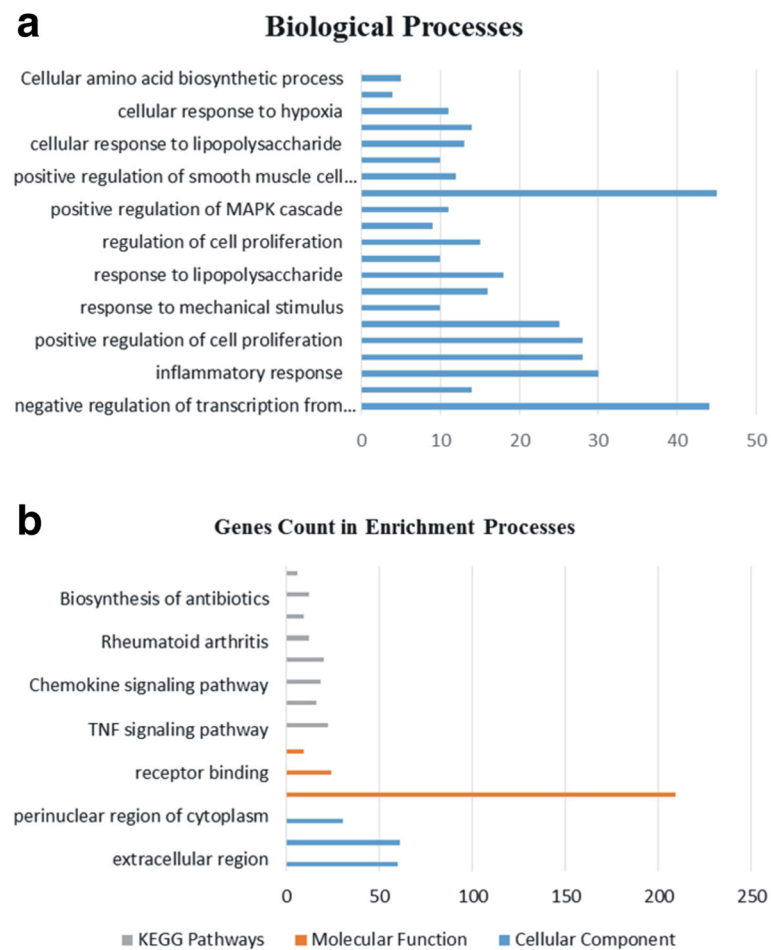


Fig. 3 The figure shows the number of genes enriched in different biological processes, cellular components, molecular functions and the KEGG pathways they are involved in. **a** is showing the biological processes only while **(b)** is showing the cellular components, molecular functions and KEGG pathways along with the number of genes involved in it

frequently aid the prediction and development of both predictive and prognostic biomarkers. PPI network analysis has been widely utilized to support the process of understanding the mechanism of different diseases, identifying drug targets and metabolic processes. The systemic interactions of different proteins in different biological process ranging from normal to disease phenotypes are playing significant role. Analysing microarray gene expression dataset and the identification of differentially expressed genes in a diseased condition compared to the normal, provides a way of targeting different nodes for the discovery of novel drug candidates. It was reported that biomarkers related to a disease through molecular network interactions are more accurate and vigorous. Recently a new method for the identification of miRNAs related to cancer without the prior knowledge of miRNA expression profiling. This method also explains how

such miRNA aid the development and progression of cancer [52]. Here, we used microarray gene expression dataset submitted to GEO under accession number GSE14335. Different statistical analysis were carried out such as student-*t* test, Pearson correlation test and Benjamin-Hochberg multiple testing method ($FDR < 0.05$ with fold-change ≥ 1) and the adjusted *p*-value (< 0.05) was selected as the threshold for DEGs identification, which result a total of 607 differentially expressed genes, of which 433 were upregulated while the rest 174 were downregulated. Among the downregulated genes COMP (cartilage oligomeric matrix protein) was found to be to most downregulated gene with the fold change value of -4.003 followed by COL15A1 (collagen type XV alpha 1 chain) and WFDC1(WAP four-disulfide core domain 1) with the fold change values of -3.188 and -3.050 respectively. Of the identified upregulated DEGs, EGFL6 (EGF like domain multiple 6) with fold change 4.27, TNFAIP3

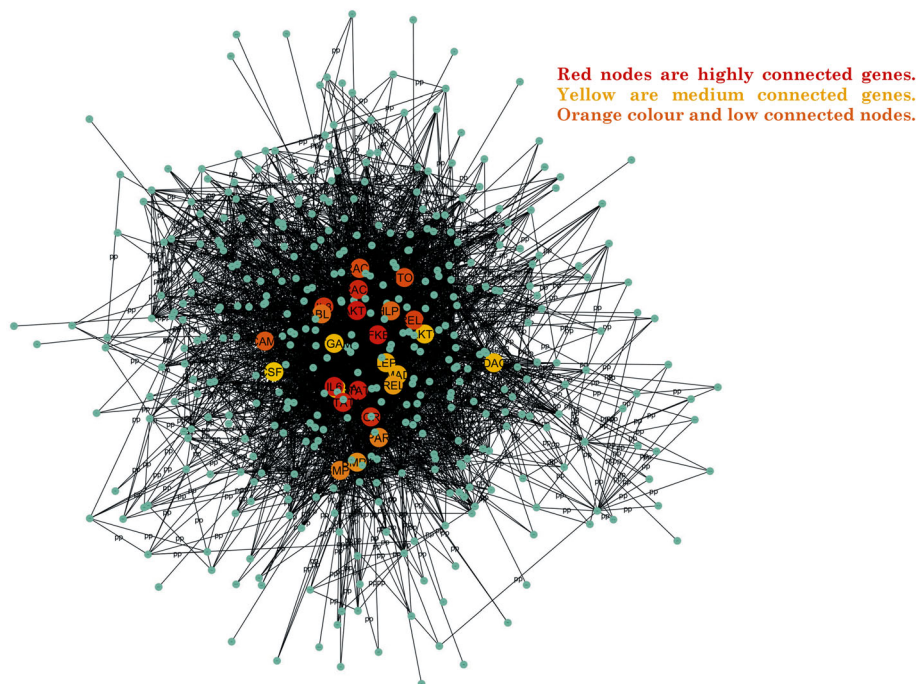
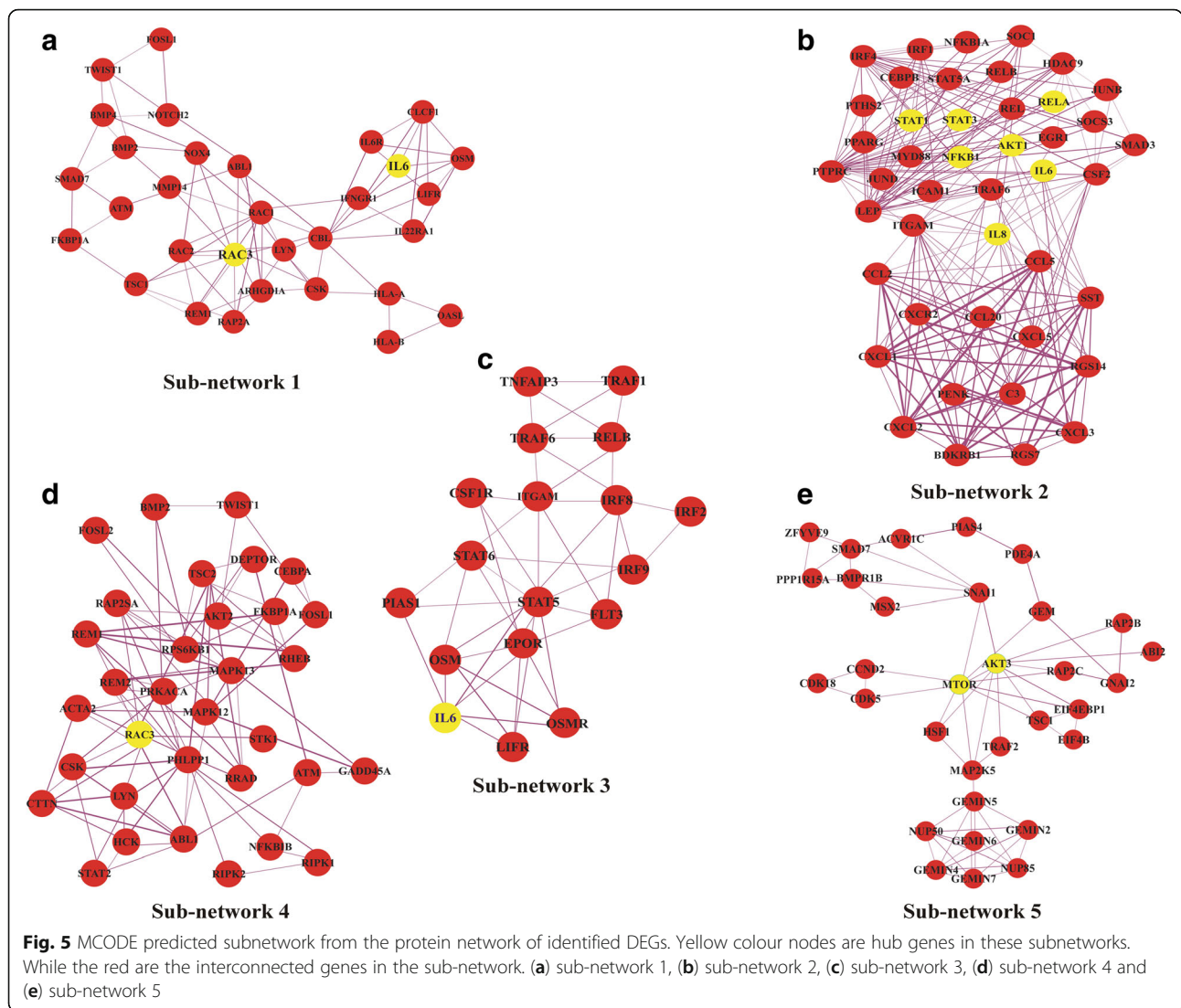


Fig. 4 The figure is illustrating the hub genes based on the degree of nodes. Red nodes are highly connected genes, yellow and orange colour are medium and low connected nodes. The Red nodes act more to be hub genes

(TNF alpha induced protein 3) and SERPINB2 (serpin family B member 2) with 4.49 and 5.33 fold change respectively. Upon subjection to enrichment analysis these DEGs revealed that they are involved in diverse array of processes. Among them the most enriched GO biological processes were immune response, positive and negative regulation of cell proliferation, positive regulation of transcription from RNA polymerase II promoter, extracellular region, extracellular space and protein binding. However, Positive regulation of cell proliferation, negative regulation of cell proliferation, L-Serine biosynthetic Process, Cellular amino acid biosynthetic process, Biosynthesis of amino acids, Glycine, serine and threonine metabolism reported to be associated with DBA [27]. Mapping of DEGs on Cytoscape top 10 hub genes were identified using Cytohubba. For hub genes degree in Cytohubba was set as a parameter. Among the top hub genes AKT1 with highest degree 171, IL6 and NFKB1 with 101, STAT3 with 99 and MTOR with 75 degree was found in the total DEGs network. These hub genes were also found in sub-networks shown in the Table 3 which is a way of validating out results.

The enrichment of hub genes given in the Table 4 reports the same category of processes and thus validate our results. RPS family is considered to be mostly involved in the causing of DBA. Many different mutations including missense and non-sense mutation in the

RPS19 is reported to be the most common. This mutant condition is characterized by haploinsufficiency [53]. Among the hub genes, AKT1 was reported to be the best target for the treatment of DBA. Hypothesis developed from different studies by *Gazda & Sieff* suggested that restriction of recruitment of polysome to other complexes for translation is regulated by AKT pathway. Not only AKT but also mTOR was reported to be involved in the restriction of forming such complexes which result haploinsufficiency [54]. *Gazda et al., 2006* in reported that Ribosomal Protein Gene Expression Alters Oncogenic Pathways in DBA [55]. Functional enrichment analysis also reported the AKT1 as the negative regulator of transcription from RNA polymerase II promoter Another study conducted by *Payne et al., 2012* and *Stipanuk 2007*, reported that treatment of RPS19 and RPS14 defected cells using Zebrafish model, skeletal muscle of rats and human CD34⁺ cells reported increased production of proteins. This condition was justified by as that L-leucine activate mTOR pathway and thus improves anemia in the DBA patients by promoting mRNA translation. On the other hand Isoquinoline-5-sulfonamides, Azepane derivatives, Aminofurazans, heterocyclic 6-5 fused rings compounds, Phenylpyrazole derivatives and Thiophenecarboxamides and derivatives are reported inhibitors reverse the negative role AKT1 in diseases. Clinical trials of L-leucine mediated



mTOR pathway activation and treatment of DBA is under process [56, 57]. IL-6 with degree value of 101 in the hub genes could be another possible target for DBA treatment. It has been confirmed while culturing of DBA patient cells in a liquid containing IL-6 media

reported decrease proliferation [58]. Furthermore, the analysis reported by previous disseminate interfaces interaction between RPS19 and the proto-oncogenic protein PIM-1. This study explained that the interaction RPS19 and the proto-oncogenic protein is mediated by the cytokines in the intracellular environment. Likewise many cytokines such as IL-6 and IL-8 are among the hub genes reported in this study could be the biomarkers to be probed for the robust treatment against the DBA infirmity [16]. This bioinformatics pipeline proposed a list of potential cellular proteins which are identified as hub genes and could be the targets for the treatment of DBA.

Table 3 Statistics for top 5 sub-networks identified in the PPI network constructed from DEGs by MCODE algorithm using Cytoscape

Sub-Network	Score	Proteins	Interaction
1.	8.18	33	131
2.	21.67	44	466
3.	5.67	18	48
4.	3.4	30	60
5.	4.45	32	69

Conclusion

In this study computational systems biology approach was utilized to effectively find DEGs that might lead to DBA condition. A total of 607 DEGs were identified, of

Table 4 Significant GO Biological terms and pathway analysis for top 10 hub genes identified from the whole PPI network of DEGs based degree of node

Gene	Degree	GO Term
AKT1	171	cellular response to lipopolysaccharide, response to cytokine, inflammatory response, negative regulation of transcription from RNA polymerase II promoter, immune response
IL6	101	cellular response to tumor necrosis factor, positive regulation of smooth muscle cell proliferation, negative regulation of cell proliferation
NFKB1	101	positive regulation of macromolecule biosynthetic process, cellular response to tumor necrosis factor, immune response
STAT3	99	positive regulation of cell proliferation, response to cytokine, inflammatory response
STAT1	91	cellular response to tumor necrosis factor, cellular response to lipopolysaccharide, response to cytokine, immune response
RAC1	88	immune response
EGR1	79	cellular response to hypoxia
IL8	77	cellular response to lipopolysaccharide, response to cytokine, inflammatory response, immune response, regulation of cell communication
RELA	76	positive regulation of cell proliferation, response to cytokine, inflammatory response
RAC3	75	positive regulation of cell proliferation
mTOR	75	positive regulation of macromolecule biosynthetic process, inflammatory response, immune response

which many new upregulated were identified in the DBA samples. Overall, the current findings speculate a potential association of these DEGs with DBA using in silico gene expression patterns and network topological analysis. Among the total identified DEGs EGFL6, TNFAIP3 and SERPINB2 were top upregulated which might implicate in the pathogenesis of DBA through interacting with one another. These findings may provide new insights on the DBA pathogenesis, and lay the foundation for the targeted therapy of the disease. We envision our results from bioinformatics analysis will be validated by experimental work.

Additional file

Additional file 1: Table S1. The identified differently expressed genes are listed in the Table S1. Both upregulated and downregulated genes are given based on the defined criteria. (XLSX 52 kb)

Abbreviations

CCR2: Chemokine receptor type 2; EGR1: Early Growth Response 1; GEO: Gene Expression Omnibus; GWAS: Genome Wide Association Studies; IL-6: Interleukin-6; IL8 : Interleukin 8; mTOR: Mammalian target of rapamycin;

NCBI: National Centre for Biotechnology Information; NFKB1: Nuclear factor NF-kappa-B1; RAC1: Ras-related C3 botulinum toxin substrate 1; RAC3: Ras-related C3 botulinum toxin substrate 3; SNP: Single Nucleotide Polymorphism; STAT1 : Signal transducer and activator of transcription 1; STAT3 : Signal transducer and activator of transcription 3

Acknowledgements

We acknowledge HPC Center, Shanghai Jiao Tong University for the computational framework setup.

Funding

Publication costs were funded by the Key Research Area Grant 2016YFA0501703 from the Ministry of Science and Technology of China.

Availability of data and materials

No data is applicable.

About this supplement

This article has been published as part of *BMC Systems Biology* Volume 12 Supplement 4, 2018: Selected papers from the 11th International Conference on Systems Biology (ISB 2017). The full contents of the supplement are available online at <https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-12-supplement-4>.

Authors' contributions

AK is the first author and has major contribution in the design, analysis and interpretation of the data as well as drafting the manuscript. MJ and AA were actively involved in the study design. CL and ACK helped in the final manuscript preparation and revision. DW and WCSC collectively supervised, checked and refined the study. The manuscript is approved by all the authors.

Ethics approval and consent to participate

N/A.

Consent for publication

N/A.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Bioinformatics and Biostatistics, College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.

²Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong.

Published: 24 April 2018

References

- Campagnoli MF, Garelli E, Quarello P, Carando A, Varotto S, Nobili B, Longoni D, Pecile V, Zecca M, Dufour C. Molecular basis of diamond-Blackfan anemia: new findings from the Italian registry and a review of the literature. *Haematologica*. 2004;89(4):480–9.
- Da Costa L, Willig T-N, Fixler J, Mohandas N, Tchernia G. Diamond-Blackfan anemia. *Curr Opin Pediatr*. 2001;13(1):10–5.
- Gripp KW, McDonald-McGinn DM, La Rossa D, McGain D, Federman N, Vlachos A, Glader BE, McKenzie SE, Lipton JM, Zackai EH. Bilateral microtia and cleft palate in cousins with diamond-Blackfan anemia. *Am J Med Genet*. 2001;101(3):268–74.
- Lipton JM, Federman N, Khabbaze Y, Schwartz CL, Hilliard LM, Clark JJ, Vlachos A. Osteogenic sarcoma associated with diamond-Blackfan anemia: a report from the diamond-Blackfan Anemia registry. *J Pediatr Hematol Oncol*. 2001;23(1):39–44.
- Young NS, Scheinberg P, Calado RT. Aplastic anemia. *Curr Opin Hematol*. 2008;15(3):162.
- Draptchinskaia N, Gustavsson P, Andersson B, Pettersson M, Dianzani I, Ball S, Tchernia G, Klar J, Matsson H, Tentler D. The gene encoding ribosomal

- protein S19 is mutated in diamond-Blackfan anaemia. *Nat Genet.* 1999;21(2):169–75.
7. Liu JM, Ellis SR. Ribosomes and marrow failure: coincidental association or molecular paradigm? *Blood.* 2006;107(12):4583–8.
 8. Ball S, McGuckin C, Jenkins G, Gordon-Smith E. Diamond-Blackfan anaemia in the UK: analysis of 80 cases from a 20-year birth cohort. *Br J Haematol.* 1996;94(4):645–53.
 9. Boria I, Garelli E, Gazda HT, Aspesi A, Quarello P, Pavesi E, Ferrante D, Meerpohl JJ, Kartal M, Da Costa L. The ribosomal basis of diamond-blackfan anemia: mutation and database update. *Hum Mutat.* 2010;31(12):1269–79.
 10. Revollo J, Nishiura H, Shibuya Y, Oda Y, Nishino N, Yamamoto T. Agonist and antagonist dual effect of the cross-linked S19 ribosomal protein dimer in the C5a receptor-mediated respiratory burst reaction of phagocytic leukocytes. *Inflamm Res.* 2005;54(2):82–90.
 11. Koga Y, Ohga S, Nomura A, Takada H, Hara T. Reduced gene expression of clustered ribosomal proteins in diamond-Blackfan anemia patients without RPS19 gene mutations. *J Pediatr Hematol Oncol.* 2006;28(6):355–61.
 12. Flygare J, Kiefer T, Miyake K, Utsugisawa T, Hamaguchi I, Da Costa L, Richter J, Davey EJ, Matsson H, Dahl N. Deficiency of ribosomal protein S19 in CD34+ cells generated by siRNA blocks erythroid development and mimics defects seen in diamond-Blackfan anemia. *Blood.* 2005;105(12):4627–34.
 13. Miyake K, Flygare J, Kiefer T, Utsugisawa T, Richter J, Ma Z, Wiznerowicz M, Trono D, Karlsson S. Development of cellular models for ribosomal protein S19 (RPS19)-deficient diamond-Blackfan anemia using inducible expression of siRNA against RPS19. *Mol Ther.* 2005;11(4):627–37.
 14. Ebert BL, Lee MM, Pretz JL, Subramanian A, Mak R, Golub TR, Sieff CA. An RNA interference model of RPS19 deficiency in diamond-Blackfan anemia recapitulates defective hematopoiesis and rescue by dexamethasone: identification of dexamethasone-responsive genes by microarray. *Blood.* 2005;105(12):4620–6.
 15. Soulet F, Al Saati T, Roga S, Amalric F, Bouche G. Fibroblast growth factor-2 interacts with free ribosomal protein S19. *Biochem Biophys Res Commun.* 2001;289(2):591–6.
 16. Chiochetti A, Gibello L, Carando A, Aspesi A, Secco P, Garelli E, Loreni F, Angelini M, Biava A, Dahl N. Interactions between RPS19, mutated in diamond-Blackfan anemia, and the PIM-1 oncoprotein. *Haematologica.* 2005;90(11):1453–62.
 17. Maeda N, Toki S, Kenmochi N, Tanaka T. A novel nucleolar protein interacts with ribosomal protein S19. *Biochem Biophys Res Commun.* 2006;339(1):41–6.
 18. El-Beshlawy A, Ibrahim IY, Rizk S, Eid K. Study of 22 Egyptian patients with diamond-Blackfan anemia, corticosteroids, and cyclosporin therapy results. *Pediatrics.* 2002;110(4):e44.
 19. Dunbar CE, Smith DA, Kimball J, Garrison L, Nienhuis AW, Young NS. Treatment of diamond-Blackfan anaemia with haematopoietic growth factors, granulocyte-macrophage colony stimulating factor and interleukin 3: sustained remissions following IL-3. *Br J Haematol.* 1991;79(2):316–21.
 20. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol.* 2007;3(1):140.
 21. Pujana MA, Han J-DJ, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet.* 2007;39(11):1338–49.
 22. Sadhu A, Bhattacharyya B. Common subcluster Mining in Microarray Data for molecular biomarker discovery. *Interdiscip Sci Comput Life Sci.* 2017;9:1–12.
 23. Wheelock CE, Wheelock AM, Kawashima S, Diez D, Kanehisa M, van Erk M, Kleemann R, Haeggström JZ, Goto S. Systems biology approaches and pathway tools for investigating cardiovascular disease. *Mol BioSyst.* 2009;5(6):588–602.
 24. Arrell D, Terzic A. Network systems biology for drug discovery. *Clin Pharmacol Ther.* 2010;88(1):120–5.
 25. Barabasi A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;5(2):101–13.
 26. Zhao M, Wei D-Q. Rare Diseases: Drug Discovery and Informatics Resource. *Interdiscip Sci Comput Life Sci.* 2017;9:1–10.
 27. Avondo F, Roncaglia P, Crescenzo N, Krmac H, Garelli E, Armiraglio M, Castagnoli C, Campagnoli MF, Ramenghi U, Gustincich S. Fibroblasts from patients with diamond-Blackfan anaemia show abnormal expression of genes involved in protein synthesis, amino acid metabolism and cancer. *BMC Genomics.* 2009;10(1):442.
 28. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207–10.
 29. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003;4(2):249–64.
 30. Wu J, Irizarry R, MacDonald J, Gentry J. Gcrma: background adjustment using sequence information. R package version. 2012;2200
 31. Team R. RStudio: integrated development for R. Boston: RStudio Inc.; 2015. In
 32. Dudoit S, Yang YH, Callow MJ, Speed TP. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat Sin.* 2002;12(1):111–39.
 33. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci.* 2001;98(9):5116–21.
 34. Šidák Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc.* 1967;62(318):626–33.
 35. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat.* 1979;65–70.
 36. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika.* 1986;73(3):751–4.
 37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 1995;57(1):289–300.
 38. Lee M-LT. Analysis of microarray gene expression data. Springer Sci Bus Media; 2007. p. 143–55.
 39. Bremer M, Himelblau E, Madlung A. Introduction to the statistical analysis of two-color microarray data. *Stat Methods Mol Biol.* 2010;620:287–313.
 40. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med.* 1990;9(7):811–8.
 41. Consortium GO. The gene ontology (GO) project in 2006. *Nucleic Acids Res.* 2006;34(suppl 1):D322–6.
 42. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25–9.
 43. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
 44. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 2003;4(9):R60.
 45. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2014;43:D447. gku1003
 46. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38(suppl 2):W214–20.
 47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
 48. Assenov Y, Ramirez F, Schelhorn S-E, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics.* 2008;24(2):282–4.
 49. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8(4):S11.
 50. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4(1):2.
 51. Qin G, Zhao X-M. A survey on computational approaches to identifying disease biomarkers based on molecular networks. *J Theor Biol.* 2014;362:9–16.
 52. Zhao X-M, Liu K-Q, Zhu G, He F, Duval B, Richer J-M, Huang D-S, Jiang C-J, Hao J-K, Chen L. Identifying cancer-related microRNAs based on gene expression data. *Bioinformatics.* 2014;31(8):1226–34.
 53. Gustavsson P, Garelli E, Draptchinskaia N, Ball S, Willig T-N, Tentler D, Dianzani U, Punnett HH, Shafer FE, Cario H. Identification of microdeletions spanning the diamond-Blackfan anemia locus on 19q13 and evidence for genetic heterogeneity. *Am J Hum Genet.* 1998;63(5):1388–95.
 54. Gazda HT, Sieff CA. Recent insights into the pathogenesis of diamond-Blackfan anaemia. *Br J Haematol.* 2006;135(2):149–57.
 55. Gazda HT, Kho AT, Sanoudou D, Zaucha JM, Kohane IS, Sieff CA, Beggs AH. Defective ribosomal protein gene expression alters transcription, translation, apoptosis, and oncogenic pathways in diamond-Blackfan Anemia. *Stem Cells.* 2006;24(9):2034–44.
 56. Payne EM, Virgilio M, Narla A, Sun H, Levine M, Paw BH, Berliner N, Look AT, Ebert BL, Khanna-Gupta A. L-leucine improves the anemia and

developmental defects associated with diamond-Blackfan anemia and del (5q) MDS by activating the mTOR pathway. *Blood*. 2012;120(11):2214–24.

57. Stipanuk MH. Leucine and protein synthesis: mTOR and beyond. *Nutr Rev*. 2007;65(3):122–9.
58. Hamaguchi I, Flygare J, Nishiura H, Brun AC, Ooka A, Kiefer T, Ma Z, Dahl N, Richter J, Karlsson S. Proliferation deficiency of multipotent hematopoietic progenitors in ribosomal protein S19 (RPS19)-deficient diamond-Blackfan anemia improves following RPS19 gene transfer. *Mol Ther*. 2003;7(5):613–22.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

