Construction of polycythemia vera protein interaction network and prediction of related biological functions

L.-J. Liu[1], X.-J. Cao[2], C. Zhou[1], Y. Sun[3], Q.-L. Lv[3], F.-B. Feng[1], Y.-Y. Zhang[4] and C.-G. Sun[1]

[1]Cancer Center, Weifang Traditional Chinese Medicine Hospital,
Kuiwen District, Weifang, Shandong Province, China
[2]The First Clinical College, Shandong University of Chinese Medicine,
Changqing District, Jinan, Shandong Province, China
[3]Clinical Medical School, Weifang Medical University, Kuiwen District,
Weifang, Shandong Province, China
[4]China Academy of Traditional Chinese Medicine, Dongcheng District,
Beijing, China

Corresponding author: C.-G. Sun
E-mail: hxdhxd19852003@163.com

ABSTRACT. Here, polycythemia vera (PV)-related genes were screened by the Online Mendelian Inheritance in Man (OMIM), and literature pertaining to the identified genes was extracted and a protein-protein interaction network was constructed using various Cytoscape plugins. Various molecular complexes were detected using the Clustervize plugin and a gene ontology-enrichment analysis of the biological pathways, molecular functions, and cellular components of the selected molecular complexes were identified using the BiNGo plugin. Fifty-four PV-related genes were identified in OMIM. The protein-protein interaction network contains 5 molecular complexes with correlation integral values >4. These complexes regulated various biological processes (peptide tyrosinase acidification, cell metabolism, and

macromolecular biosynthesis), molecular functions (kinase activity, receptor binding, and cytokine activity), and the cellular components were mainly concentrated in the nucleus, intracellular membrane-bounded organelles, and extracellular region. These complexes were associated with the JAK-STAT signal transduction pathway, neurotrophic factor signaling pathway, and Wnt signaling pathway, which were correlated with chronic myeloid leukemia and acute myeloid leukemia.

Key words: Protein-protein interaction network; Polycythemia vera; Online Mendelian Inheritance in Man; Molecular complexes; Biological pathways

INTRODUCTION

Polycythemia vera (PV) is a chronic and cloned myeloproliferative disease with predominant erythrocytosis and abnormal erythrocyte, granulocyte, and megakaryocyte proliferation. The clinical features of PV include erythrocytosis, neutrophilic leukocytosis, thrombocytosis, and splenomegaly; however, until the latter stages of disease, PV patients exhibit progressive myelofibrosis, anemia, and up-regulation of splenomegaly, symptoms that are also observed in idiopathic myelofibrosis syndrome.

A number of researchers have attempted to elucidate the pathogenesis of PV, with variable degrees of success, since 2005. Mutations in the *JAK2* gene have been specifically identified in PV patients by multiple research groups, and is currently considered the standard for PV pathogenesis (Levine and Gilliland, 2008). Additionally, approximately 40% of the PV patients are believed to express an abnormal chromosomal karyotype. Kralovics et al. (2002) reported that *9pLOH* was the most common abnormal chromosome in PV patients, accounting for approximately 33% of the total. However, these chromosomal abnormalities are also known to affect chronic myeloid leukemia; therefore, these cannot be classified as PV-specific abnormalities (Kralovics et al., 2002). The discovery of mutations in the *JAK2* gene is believed to have opened new avenues for research into the pathogenesis of PV; however, as this mutation is also observed in other myeloproliferative disorders, and is not universal to all PV patients, research into PV pathogenesis must not be limited to the *JAK2* mutation. Recent advances in the field of bioinformatics have led to the development of new tools (e.g., for the construction of a protein-protein interaction network) to elucidate the genesis and development of PV.

Here, previously confirmed PV-related genes were identified by Online Mendelian Inheritance in Man (OMIM) and a literature search was performed using the Cytoscape and Agilent Literature Search plugin; the obtained information was used to build a protein-protein interaction network. The MCOMD algorithm, obtained from the Clusterviz plugin, was used to identify possible molecular complexes in the network, and the Bingo plugin was used for gene ontology (GO)-enrichment analysis of the biological processes, molecular functions, and cellular components of the complex. The DAVID program was used to enrich the biological pathways of molecular complexes, which in turn provides a basis for further research into the genesis and development of PV.

MATERIAL AND METHODS

Materials

PV-related genes were extracted from the OMIM database, using the search term "Polycythemia vera" in the "Gene" section (Amberger et al., 2009; Burkard et al., 2010).

Methods

Construction of gene/protein interaction network

The identified PV-related genes were searched using the Agilent Literature Search 2.7.7 plugin (Cytoscape 2.8.2), which performed a comprehensive literature search of the related literature in PubMed (Vailaya et al., 2005); the retrieved results were read and all false-positive interaction information was excluded. The identified (included) gene/protein interaction relationships were visualized in the form of a network in Cytoscape 2.8.2 (Shannon et al., 2003).

Network analysis

The MCOMD algorithm in the Clusterviz 1.2 plugin of Cytoscape 2.8.2 was used to perform a regional correlation analysis of the biological network. This analysis provided the integral correlation values and helped identify possible protein groups forming molecular complexes in the entire network. Molecular complexes were annotated as areas with correlation integral values greater than 4.

The BiNGO 2.44 network analysis plugin of Cytoscape 2.8.2 was used to annotate and enrich the function of each gene/protein in the molecular complexes related to the biological process, molecular function, and cellular components. The "Hypergenometric test" was performed, with P < 0.05 as the significant threshold (http://apps.cytoscape.org/apps/bingo).

The names of genes/proteins included in the molecular complexes were submitted to "The Database for Annotation, Visualization, and Integrated Discovery" (Dennis Jr. et al., 2003; Huang et al., 2009); a relative search of the KEGG, BBID, PANTHER, PATHWAY, BIOCARTA, EC_NUMBER, and REACTOME_PATHWAY databases helped identify the biological pathways regulated by PV-related genes. The P values of relative target genes predicted using the software were calculated by the incidental Fisher exact test. P < 0.05 indicated the role of these genes in transduction pathways with statistical significance.

Major outcome measures

The nodes (proteins) and sides (interaction) of a protein-protein interaction network comprised of PV-related genes in the molecular complexes, and the correlation integral values were determined. The biological processes regulated by molecular functions, cellular components, and biological pathways of these molecular complexes were identified.

RESULTS

PV-related genes in OMIM

Fifty-six PV-related genes were identified using the search criterion (up to 2015-02-09); fifty-four genes (up to 2015-02-09) were retained after screening and removal of duplicate genes: *CTNNB1*, *ACSL6*, *ASXL1*, *BCL2L1*, *BCR*, *CALR*, *CD177*, *CRP*, *CSF2*, *CTNNB1*, *DKK3*, *DNMT3A*, *EIF2S1*, *ERCC2*, *ETV6*, *F2*, *F5*, *FIP1L1*, *HAMP*, *GSTT1*, *HDAC11*, *HDAC9*, *HGF*, *HMGA2*, *IDH1*, *IDH2*, *IFNA1*, *IL11*, *IL23A*, *JAK2*, *KIT*, *L3MBTL1*, *LCN2*, *MEFV*, *MPL*, *MTHFR*, *NFE2*, *NFIX*, *NR3C1*, *PDGFRA*, *PLVAP*, *PRM2*, *PTX3*, *SH2B3*, *SOCS2*, *SOCS3*, *SPRY1*, *SPRY2*, *SPRY4*, *STAT3*, *STAT5A*, *TET2*, *TNFRSF11B*, and *TNF*.

Gene/protein interaction network

A gene/protein interaction network, consisting of 541 nodes (protein) and 1771 sides (interactions), was constructed after a comprehensive literature search of the 54-related genes using the Agilent Literature Search engine (Figure 1). The yellow color in Figure 1 indicates genetic disease-related proteins in OMIM, while the red color indicates the proteins identified by a literature search.

Figure 1. Protein-protein interaction network of polycythemia vera (the whole and the part).

Molecular complex

The MCOMD algorithm was used to analyze the interaction network; five molecular complexes with correlation integral values higher than 4 (Figure 2) were identified. The yellow color in Figure 2 indicates genetic disease-related proteins in OMIM, while the red color indicates the proteins identified by a literature search.
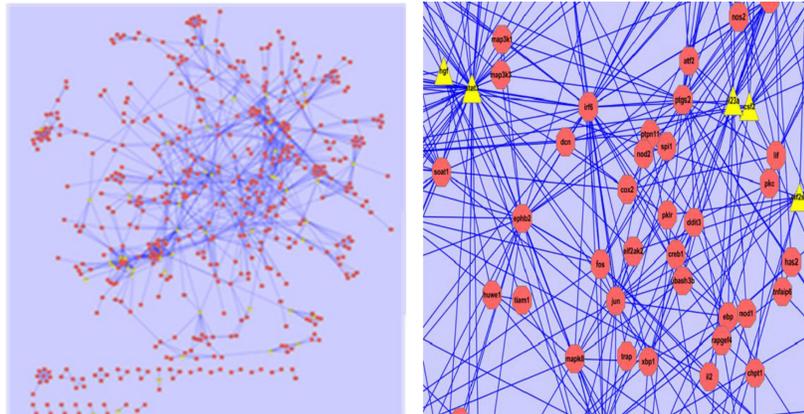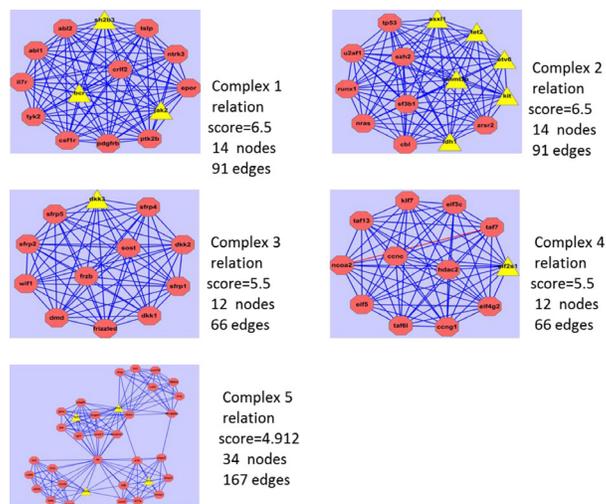

Figure 2. Molecular complexes obtained by MCOMD algorithm analysis.

GO analysis of gene/protein in the molecular complexes

The GO database was used for gene annotation and function-enrichment analysis of the 5 molecular complexes. The PV-related target genes were separately enriched in various biological processes, including peptide tyrosinase acidification, cell metabolism, macromolecular biosynthesis, gene expression, Wnt signal pathway regulation, and immune regulation. The major molecular functions of these genes included kinase activity, receptor binding, and cytokine activity, while the cell components were mainly concentrated in the nucleus, intracellular membrane-bound organelles, and the extracellular region. The five lowest enriched genes and protein in each molecular complex were chosen; these were concentrated in the nucleo-cytoplasm, cell membrane, extracellular space, cell surface membrane raft, and cell components, and regulated molecular functions such as nucleic acid binding, cytokine activity, protein activity, enzyme activity, and growth factor activity regulation (Tables 1-3). The biological process, molecular function, and cellular components of all molecular complexes are shown in Figures 3-5. The node size represented the number of genes, and the depth of color indicated the P value.

Table 1. Gene ontology (GO) analysis of biological processes related to the identified molecular complexes.

GO-ID	Description	P value	Genes	Total frequency[#]
Complex 1				
48169	Regulation of long-term neuronal synaptic plasticity	1.6730 E-4	NRAS, KIT	20/14305, 0.1%
8542	Visual learning	2.4249 E-4	NRAS, KIT	24/14305, 0.1%
7632	Visual behavior	3.0786 E-4	NRAS, KIT	27/14305, 0. 1%
9416	Response to light stimulus	3.3523 E-4	NRAS, TP53, KIT	144/14305, 1.0%
8306	Associative learning	4.6156 E-4	NRAS, KIT	33/14305, 0.2%
Complex 2				
18108	Peptidyl-tyrosine phosphorylation	2.5047 E-13	TYK2, PTK2B, PDGFRB, JAK2, ABL1, ABL2	39/14306, 0.2%
18212	Peptidyl-tyrosine modification	3.4494 E-13	TYK2, PTK2B, PDGFRB, JAK2, ABL1, ABL2	41/14306, 0.2%
6468	Protein amino acid phosphorylation	1.6742 E-10	TYK2, NTRK3, BCR, PTK2B, PDGFRB, JAK2, ABL1, ABL2, CSF1R	657/14306, 4.5%
16310	Phosphorylation	8.2677 E-10	TYK2, NTRK3, BCR, PTK2B, PDGFRB, JAK2, ABL1, ABL2, CSF1R	786/14306, 5.4%
18193	Peptidyl-amino acid modification	1.4434 E-9	TYK2, PTK2B, PDGFRB, JAK2, ABL1, ABL2	158/14306, 1.1%
Complex 3				
31323	Regulation of cellular metabolic process	3.8111 E-7	EIF4G2, TAF13, KLF7, NCOA2, HDAC2, EIF2S1, EIF5, TAF7, CCNC, TAF6L, CCNG1	3736/14306, 26.1%
19222	Regulation of metabolic process	6.4535 E-7	EIF4G2, TAF13, KLF7, NCOA2, HDAC2, EIF2S1, EIF5, TAF7, CCNC, TAF6L, CCNG1	3919/14306, 27.3%
10556	Regulation of macromolecule biosynthetic process	9.6108 E-7	EIF4G2, TAF13, KLF7, NCOA2, HDAC2, EIF2S1, EIF5, TAF7, CCNC, TAF6L	2877/14306, 20.1%
10468	Regulation of gene expression	1.1415 E-6	EIF4G2, TAF13, KLF7, NCOA2, HDAC2, EIF2S1, EIF5, TAF7, CCNC, TAF6L	2928/14306, 20.4%
31326	Regulation of cellular biosynthetic process	1.5499 E-6	EIF4G2, TAF13, KLF7, NCOA2, HDAC2, EIF2S1, EIF5, TAF7, CCNC, TAF6L	3021/14306, 21.1%
Complex 4				
30178	Negative regulation of Wnt receptor signaling pathway	9.8887 E-13	DKK2, DKK3, DKK1, SFRP1, SFRP2, FRZB	54/14306, 0.3%
30111	Regulation of Wnt receptor signaling pathway	1.9144 E-11	DKK2, DKK3, DKK1, SFRP1, SFRP2, FRZB	87/14306, 0.6%
35467	Negative regulation of signaling pathway	1.9780 E-10	DKK2, DKK3, DKK1, SOST, SFRP1, SFRP2, FRZB	262/14306, 1.8%
16055	Wnt receptor signaling pathway	2.2316 E-10	SFRP5, SOST, SFRP1, SFRP2, SFRP4, WIF1	130/14306, 0.9%
10648	Negative regulation of cell communication	9.5324 E-10	DKK2, DKK3, DKK1, SOST, SFRP1, SFRP2, FRZB	328/14306, 2.2%
Complex 5				
9617	Response to bacterium	4.4647 E-23	TNF, PTGS2, TLR2, NFKBIA, TLR3, IL10, CXCL10, PCSK1, CD86, TNFA, IFNB1, SERPINE1, IFNG, IL1B, NOS2, CD14, IL1A	232/14305, 1.6%
42127	Regulation of cell proliferation	9.6585 E-21	CSF3, TNF, IL8, PTGS2, PDGFB, IGF1, NFKBIA, IL10, TIMP1, IL11, CXCL10, CD86, TNFA, IFNB1, IRF6, TEK, SERPINE1, IFNG, IL1B, ANGPT1, NOS2, IL1A	849/14305, 5.9%
48518	Positive regulation of biological process	4.6294 E-20	CSF3, TNF, PDGFB, PTGS2, TLR2, TLR3, NFKBIA, IL10, CXCL10, IL11, TIMP1, PCSK1, IL17A, TNFA, SERPINE1, IFNG, TEK, IL1B, ANGPT1, NOS2, IL1A, ICAM1, IL8, IGF1, CD86, IFNB1, IRF6, CD14	2209/14305, 15.4%
51707	Response to other organism	7.8380 E-20	TNF, PTGS2, TLR2, NFKBIA, TLR3, IL10, CXCL10, PCSK1, CD86, TNFA, IFNB1, SERPINE1, IFNG, IL1B, NOS2, CD14, IL1A	358/14305, 2.5%
2376	Immune system process	1.0806 E-19	CSF3, ICAM1, TNF, IL8, PDGFB, CXCL2, TLR2, NFKBIA, TLR3, IL10, TIMP1, IL11, CXCL10, CD86, IL17A, TNFA, IFNB1, IFNG, IL1B, NOS2, CD14, IL1A	949/14305, 6.6%

Only the first 5 molecular function categories of enrichment analysis listed. [#]Total percentage of number of target genes in function categories with respect to the functional annotation of target gene.

Molecular complex pathway enrichment

The proteins included in the 5 molecular complexes were submitted online in order to identify the relevant pathways; these proteins were seen to mainly regulate the JAK-STAT signal transduction pathway, neurotrophic factor signaling pathway, Wnt signaling pathway, Toll-like receptor signal transduction pathway, and RIG-I-like receptor signal transduction pathway, as well as chronic myeloid leukemia, cancer, and acute myeloid leukemia (Table 4).

Table 2. Gene ontology (GO) analysis of molecular functions of the identified molecular complexes.

GO-ID	Description	P value	Genes	Total frequency#
Complex 1				
3682	Chromatin binding	4.5153 E-4	SF3B1, DNMT3A, TP53	172/15443, 1.1%
3676	Nucleic acid binding	5.6971 E-4	SF3B1, DNMT3A, CBL, EZH2, U2AF1, TP53, ZRSR2, ETV6, RUNX1	3252/15443, 21.0%
5020	Stem cell factor receptor activity	9.0656 E-4	KIT	1/15443, 0.0%
51718	DNA (cytosine-5-)-methyltransferase activity, acting on CpG substrates	1.8124 E-3	DNMT3A	2/15443, 0.0%
35033	Histone deacetylase regulator activity	1.8124 E-3	TP53	2/15443, 0.0%
Complex 2				
4713	Protein tyrosine kinase activity	1.0082 E-13	TYK2, NTRK3, PTK2B, PDGFRB, JAK2, ABL1, ABL2, CSF1R	139/1544, 4 0.9%
4715	Non-membrane spanning protein tyrosine kinase activity	1.7700 E-10	TYK2, PTK2B, JAK2, ABL1, ABL2	40/1544, 0.2%
4672	Protein kinase activity	1.9642 E-10	TYK2, NTRK3, BCR, PTK2B, PDGFRB, JAK2, ABL1, ABL2, CSF1R	568/1544, 4 3.6%
16773	Phosphotransferase activity, alcohol group as acceptor	9.8214 E-10	TYK2, NTRK3, BCR, PTK2B, PDGFRB, JAK2, ABL1, ABL2, CSF1R	681/1544, 4 4.4%
16301	Kinase activity	2.0710 E-9	TYK2, NTRK3, BCR, PTK2B, PDGFRB, JAK2, ABL1, ABL2, CSF1R	741/1544, 4 4.7%
Complex 3				
3743	Translation initiation factor activity	2.6179 E-6	EIF4G2, EIF2S1, EIF5	40/15444, 0.2%
16563	Transcription activator activity	6.5016 E-6	KLF7, NCOA2, HDAC2, TAF7, TAF6L	429/1544, 4 2.7%
3713	Transcription coactivator activity	1.1164 E-5	KLF7, NCOA2, TAF7, TAF6L	215/1544, 4 1.3%
8135	Translation factor activity, nucleic acid binding	1.6270 E-5	EIF4G2, EIF2S1, EIF5	73/15444, 0.4%
8134	Transcription factor binding	1.6886 E-5	KLF7, NCOA2, HDAC2, TAF7, TAF6L	522/1544, 4 3.3%
Complex 4				
17147	Wnt-protein binding	1.2056 E-10	SFRP1, SFRP2, SFRP4, FRZB	15/15444, 0.0%
2162	Dystroglycan binding	1.9414 E-3	DMD	3/15444, 0.0%
0998	Nitric-oxide synthase binding	3.2337 E-3	DMD	5/15444, 0.0%
5109	Frizzled binding	3.8793 E-3	SFRP1	6/15444, 0.0%
Complex 5				
5125	Cytokine activity	2.8668 E-17	CSF3, IL17A, TNFA, TNF, IL8, IFNB1, IFNG, CXCL2, IL1B, IL1A, IL10, CXCL10, IL11	198/15443, 1.2%
5126	Cytokine receptor binding	8.0428 E-16	CSF3, TNFA, TNF, IL8, IFNB1, IFNG, CXCL2, IL1B, IL1A, IL10, CXCL10, IL11	187/15443, 1.2%
5102	Receptor binding	1.6455 E-13	CSF3, ICAM1, TNF, IL8, PDGFB, CXCL2, IGF1, IL10, IL11, CXCL10, IL17A, TNFA, IFNB1, IFNG, IL1B, ANGPT1, IL1A	923/15443, 5.9%
70851	Growth factor receptor binding	1.2325 E-8	CSF3, PDGFB, IL1B, IL1A, IL10, IL11	82/15443, 0.5%
8083	Growth factor activity	6.6913 E-7	CSF3, PDGFB, IGF1, IL1B, IL10, IL11	160/15443, 1.0%

List only the first 5 molecular function categories of enrichment analysis results. #Total percentage of the number of target gene in the current functional categories to the functional annotation of target gene.

Table 3. Gene ontology (GO) analysis of cellular components of the identified molecular complexes.

GO-ID	Description	P value	Genes	Total frequency#
Complex 1				
44428	Nucleus	1.2587 E-5	SF3B1, DNMT3A, ASXL1, EZH2, U2AF1, TP53, ZRSR2, ETV6	1740/16377, 10.6%
31519	PcG protein complex	7.8578 E-5	ASXL1, EZH2	17/16377, 0.1%
5681	Spliceosomal complex	1.4427 E-4	SF3B1, U2AF1, ZRSR2	134/16377, 0.8%
43231	Intracellular membrane-bounded organelle	1.5615 E-4	DNMT3A, EZH2, ASXL1, CBL, TP53, KIT, SF3B1, NRAS, U2AF1, IDH1, ZRSR2, ETV6, RUNX1	8348/16377, 50.9%
43227	Membrane-bounded organelle	1.5786 E-4	DNMT3A, EZH2, ASXL1, CBL, TP53, KIT, SF3B1, NRAS, U2AF1, IDH1, ZRSR2, ETV6, RUNX1	8355/16377, 51.0%
Complex 2				
44459	Plasma membrane part	3.5034 E-4	NTRK3, PTK2B, PDGFRB, EPOR, JAK2, IL7R, CSF1R	1999/16378, 12.2%
Complex 3				
5669	Transcription factor TFIID complex	6.9693 E-5	TAF13, TAF7	2/11, 18.1%
70461	SA GA-type complex	1.0296 E-4	TAF7, TAF6L	2/11, 18.1%
118	Histone deacetylase complex	2.5517 E-4	HDAC2, TAF6L	2/11, 18.1%
44451	Nucleoplasm part	4.3225 E-4	TAF13, HDAC2, TAF7, TAF6L	4/11, 36.3%
123	Histone acetyltransferase complex	4.9362 E-4	TAF7, TAF6L	2/11, 18.1%
Complex 4				
5576	Extracellular region	8.0772 E-9	SFRP5, DKK2, DKK3, DKK1, SOST, SFRP1, SFRP2, SFRP4, WIF1, FRZB	2027/16378, 12.3%
5615	Extracellular space	1.1225 E-7	SFRP5, DKK2, DKK3, SOST, SFRP1, SFRP4, FRZB	747/16378, 4.5%
44421	Extracellular region part	7.4280 E-7	SFRP5, DKK2, DKK3, SOST, SFRP1, SFRP4, FRZB	985/16378, 6.0%
Complex 5				
44421	Extracellular region part	4.7822 E-19	CSF3, ICAM1, TNF, IL8, CXCL2, IGF1, MMP1, IL10, TIMP1, IL11, CXCL10, PCSK1, IL17A, TNFA, IFNB1, IFNG, SERPINE1, IL1B, ANGPT1, CD14, IL1A	986/16378, 6.0%
5615	Extracellular space	2.3005 E-18	CSF3, ICAM1, TNF, IL8, CXCL2, IGF1, IL10, IL11, CXCL10, PCSK1, IL17A, TNFA, IFNB1, SERPINE1, IFNG, IL1B, ANGPT1, CD14, IL1A	748/16378, 4.5%
5576	Extracellular region	6.5374 E-14	CSF3, ICAM1, TNF, IL8, PDGFB, CXCL2, IGF1, MMP1, IL10, TIMP1, IL11, CXCL10, PCSK1, IL17A, TNFA, IFNB1, IFNG, SERPINE1, IL1B, ANGPT1, CD14, IL1A	2028/16378, 12.3%
9986	Cell surface	1.6899 E-7	ICAM1, CD86, TNFA, TNF, PDGFB, TEK, TLR2, IL1A	341/16378, 2.0%
45121	Membrane raft	3.1278 E-7	TNFA, TNF, PTGS2, TEK, ANGPT1, CD14	149/16378, 0.9%

List only the first 5 molecular function categories of enrichment analysis results. #Total percentage of the number of target gene in the current functional categories to the functional annotation of target gene.

Figure 3. Biological processes of complex 5 (the whole and the part).

Figure 4. Molecular function of complex 5.

Figure 5. Cell components of complex 5.

Table 4. Pathways related to the identified molecular complexes.

Category	Term	P value	Gene	FDR
Complex 1				
KEGG_PATHWAY	Chronic myeloid leukemia	2.4293 E-4	NRAS, CBL, TP53, RUNX1	0.2327
KEGG_PATHWAY	Pathways in cancer	1.6521 E-3	NRAS, CBL, TP53, KIT, RUNX1	1.5728
KEGG_PATHWAY	Acute myeloid leukemia	4.3727 E-3	NRAS, KIT, RUNX1	4.1147
PANTHER_PATHWAY	mRNA splicing	2.2561 E-3	U2AF1, ZRSR2	16.3551
Complex 2				
EC_NUMBE	2.7.10.2	5.2684 E-7	TYK2, PTK2B, JAK2, ABL1, ABL2	1.6585E-4
KEGG_PATHWAY	Jak-STAT signaling pathway	2.6059 E-5	TYK2, TSLP, CRLF2, EPOR, JAK2, IL7R	0.02159
KEGG_PATHWAY	Cytokine-cytokine receptor interaction	3.1971 E-4	TSLP, CRLF2, PDGFRB, EPOR, IL7R, CSF1R	0.2646
EC_NUMBER	2.7.10.1	5.2021 E-3	NTRK3, PDGFRB, CSF1R	1.6284
KEGG_PATHWAY	Hematopoietic cell lineage	1.9540 E-2	EPOR, IL7R, CSF1R	15.0824
KEGG_PATHWAY	Neurotrophin signaling pathway	3.8602 E-2	NTRK3, SH2B3, ABL1	27.8300
KEGG_PATHWAY	Pathways in cancer	4.6901 E-2	BCR, PDGFRB, ABL1, CSF1R	32.8320
Complex 3				
KEGG_PATHWAY	Basal transcription factors	2.9417 E-4	TAF13, TAF7, TAF6L	0.0926
Complex 4				
KEGG_PATHWAY	Wnt signaling pathway	5.0840 E-10	SFRP5, DKK2, DKK1, SFRP1, SFRP2, SFRP4, WIF1	1.1477E-7
PANTHER_PATHWAY	Wnt signaling pathway	5.6881 E-5	SFRP5, SFRP1, SFRP2, SFRP4, FRZB	0.0057
PANTHER_PATHWAY	Angiogenesis	1.5156 E-3	SFRP5, SFRP1, SFRP2, FRZB	0.1516
Complex 5				
KEGG_PATHWAY	Cytokine-cytokine receptor interaction	1.3021 E-8	CSF3,IL17A, IL8, PDGFB, IFNB1, CXCL2, IFNG, IL10, IL1A, IL11, CXCL10	1.2857E-5
BIOCARTA	Cytokines and inflammatory response	2.0616 E-7	CSF3, IL8, IFNB1, IFNG, IL10, IL1A, IL11	2.1209E-4
KEGG_PATHWAY	Toll-like receptor signaling pathway	1.6368 E-6	IL8, IFNB1, TLR2, NFKBIA, TLR3, CD14, CXCL10	0.0016
BIOCARTA	Cytokine network	1.7027 E-6	IL17A, IL8, IFNB1, IFNG, IL10, IL1A	0.0018
BIOCARTA	Adhesion and diapedesis of granulocytes	2.2430 E-5	CSF3, ICAM1, IL8, IFNG, IL1A	0.0231
PANTHER_PATHWAY	Toll receptor signaling pathway	3.2975 E-5	PTGS2, IFNB1, TLR2, NFKBIA, TLR3, CD14	0.0247
BBID	Cytokine-chemokine_CNS	3.1438 E-4	IL8, IFNB1, IFNG, IL10, IL1A	0.2873
BIOCARTA	Erythrocyte differentiation pathway	3.6613 E-4	CSF3, IGF1, IL1A, IL11	0.3760
REACTOME_PATHWAY	Signaling in immune system	5.6719 E-4	ICAM1, TEK, TLR2, NFKBIA, LR3, MMP1, CD14	0.3260
BBID	Immnosurveillance	1.0734 E-3	ICAM1, IL8, IFNG, IL10, IL1A	0.9777
KEGG_PATHWAY	Pathways in cancer	1.2300 E-3	PTGS2, IL8, PDGFB, NFKBIA, IGF1, NOS2, MMP1	1.2080
BIOCARTA	Dendritic cells in regulating TH1 and TH2 development	1.6054 E-3	IFNB1, IFNG, TLR2, IL10	1.6393
KEGG_PATHWAY	RIG-I-like receptor signaling pathway	2.5081 E-3	IL8, IFNB1, NFKBIA, CXCL10	2.4492
KEGG_PATHWAY	Jak-STAT signaling pathway	2.7476 E-3	CSF3, IFNB1, IFNG, IL10, IL11	2.6802
BBID	Alternatively_Activated_APC	4.1034 E-3	IFNG, IL10, CD14, IL1A	13.6920
KEGG_PATHWAY	Hematopoietic cell lineage	4.3251 E-3	CSF3, CD14, IL1A, IL11	4.1897
BIOCARTA	Toll-Like receptor pathway	8.3371 E-3	TLR2, NFKBIA, TLR3, CD14	8.2524

DISCUSSION

The major purpose of a bioinformatic analysis of molecular regulation networks is to elucidate the temporo-spatial mechanism of molecular regulation using a model. Identification of the precise molecular mechanism and complicated regulation of gene expression could be useful in establishing various network models to describe the behavior of biological systems. Gene expression is dependent on its interaction with other genes and regulation. Elucidation of the relationship between diseases, genes, and drugs using a network plays a significant role in the design of drugs.

A molecular regulation network construct is based on the foundation of the formation and development of the internet for the study of a network dynamic model. The most common network evolution models of this day include the rule network model, random network model, small-world network model, and scale-free network model. In general, complicated molecular regulation networks display 4 properties: degree distribution, the shortest path, aggregation, and robustness. The study of the systemic network properties of living organisms is at a preliminary stage; however, organismal networks are based on some general universal rules. Ravasz et al. (2002) split an algorithm-based metabolic network into smaller modular sub-networks, and discovered that the majority of the reactions within the same module belonged to the same metabolic pathways. This indicated that each sub-module has relatively independent biological functions; therefore, these were denoted as biological function modules. Based on these results, a compilation of genes recovered from OMIM were combined with the results of literature mining to build a network of protein-protein interactions in PV in this study. Although the extracted experimental data were related to genes, the identified literature proved its function and interaction with other molecules at a protein level. Therefore, the network set up via literature mining could illuminate protein-protein interactions in molecular complexes.

A PV protein-protein interaction network was built using the 54 genes identified from OMIM; this network included 541 nodes (proteins) and 1771 sides (mutual effects). Because of the enormity of the internet, the MCOMD algorithm was used to estimate the network area, through a correlation integral. Correlation integrals illustrate the correlation of proteins at a specific area. Proteins in these complexes that regulated known biological processes and pathways, and with known molecular functions and cellular components were excluded to allow for the identification of the function of unknown genes or novel molecular functions (Bader and Hogue, 2003; Zhou et al., 2014). Five molecular compounds with correlation integral >4 were identified in this study.

Gene ontology is a new perspective in the study of functional genomics. GO develops effective bioinformatic methods to realize the targets of genetic network analysis and the functional prediction of unknown genes using previously existing biological architecture and knowledge. GO analyzes for three stand-alone systems: cellular component, biological process, and molecular function. Based on these, GO analysis identified 5 compounds that were considerably enriched, and the characteristics of classification were determined by regional GO. DAVID enriches the biological information of single genes in addition to annotating large genes from different species.

Compound 2 was found to regulate several biological processes, including the phosphorylation of the peptide tyrosine and the amino acids in proteins; the molecular function of compound 2 included the regulation of tyrosine kinase and phosphatase activity, while its cellular component was concentrated in the plasma membrane. Studies conducted since 2005 revealed that the morbidity of PV was closely related to the *JAK* gene mutation. JAK, a tyrosine phosphatase

protein family expressed in mammals, is composed of four members: JAK1, JAK2, JAK3, and TYK2. Tyrosine phosphatase, which suppresses cytokine signaling and inhibits activated STATs, negatively regulates the JAK/STAT pathway (Rawlings et al., 2004; Wang et al., 2009).

Phosphorylation of the tyrosyl residue is the key regulatory mechanism in the eukaryotic signaling pathway, and is regulated by PTKs and PTPs. PTP de-phosphorylation and PTK phosphorylation are in dynamic equilibrium, and are closely related to cellular proliferation, differentiation, signal transduction, and regulation of cell cycle. Some studies have reported that this balance can be disrupted by up-regulating PTK expression; this could lead to the accumulation of proteins related to tyrosine phosphorylation in the cell, which in turn causes abnormal cell proliferation and differentiation, as well as disease. This imbalance could also be caused by the decline in PTP activity as a result of gene mutation, leading to an increase in GAP-associated proteins (affecting tyrosine phosphorylation) in the cell. The increase in intracellular protein phosphorylation could be a key to unlocking the cell proliferation and differentiation mechanism (Weerkamp et al., 2006).

The gene/protein interaction in molecule 4 is responsible for various biological processes (regulation of Wnt signal pathway) and molecular functions (combining Wnt protein with nitric oxide synthase), while its cellular component is located in the extracellular space and region. The classical Wnt pathway has been extensively researched in malignant diseases of the blood. Weerkamp et al. (2006) and Khan et al. (2007) reported that ALL patients expressed abnormalities in the Wnt pathway; additionally, many genes and receptors belonging to the Wnt family have been identified in precursor B-cell lymphoblastic leukemic cells. Wnt3 stimulation leads to β-catenin accumulation, rapid cell proliferation, and a discernible change in cell cycle and apoptosis-related gene function. Therefore, abnormalities in the classical Wnt pathway is a potential pathogenesis of malignant disease (Valencia et al., 2009). A previous study on AML reported that the genes encoding the Wnt pathway antagonist proteins DKK1 and DKK3 undergo methylation *in vitro* (cell line and specimen) (Wu et al., 2009). B-CLL patients highly express Wnt proteins and their homologous receptor genes. The study of B-CLL and its development in an animal (mouse) model revealed that an increase in impaired Fz-6 and β-catenin expression led to a decrease in disease progression (Ashihara et al., 2009). β-catenin is highly expressed in myeloma cell lines; cell proliferation is disrupted upon deletion of the β-catenin small interfering RNA (Riley et al., 2008).

The biological process of compound 1 is similar to that of 3 and 5, and includes the regulation of cellular metabolism, macromolecular biosynthesis, regulation of gene expression, cell proliferation, and immune regulation. The molecular function of compound 1 includes nucleic acid clustering, cytokine activity, enzymatic activity, and transcription factor clustering. Compound 1 is mainly located in the nucleus, intracellular membrane-bound organelle, nucleoplasm, and the extracellular space. *TP53* (coding for p53) is an important tumor suppressor gene, which retards the cell cycle, restrains cell proliferation, induces cell differentiation, activates cell aging and apoptosis, participates in DNA damage and repair (maintain genomic balance), regulates the energy metabolism, and restrains tumor angiogenesis, via transcriptional regulation of many key genes such as *p21*, *Bax*, and *Puma* when the cell is stimulated by carcinogenic factors, thereby preventing the genesis and development of tumor (Teodoro et al., 2006; Riley et al., 2008; Feng and Levine, 2010; Feng et al., 2011; Kumar et al., 2011). More than 50% tumors are estimated to express mutated *TP53*, which could lead to an abnormal p53 signal transduction pathway (Xiong et al., 2008; Thol et al., 2010; Kumar et al., 2011). Interferon α (IFNα) up-regulates the activity of the *NK* gene via the STATE1 and STAT3 signal pathways. Isocitrate dehydrogenase catalyzes the

conversion of isocitric acid to α-oxoglutarates, such as IDHA, IDH2, and IDH3. IDH1/2 mutations can undergo one of two mechanisms resulting in tumor formation: 1) decline in the ability of isocitric acid to introduce α-KG, leading to the accumulation of hypoxia-inducible factor 1a (HIF1a), and influencing the formation, transference, and resistance of tumor vessels (Rakheja et al., 2011); 2) development of new enzymatic ability by the mutated IDH1/2, leading to the accumulation of D-2-HG and subsequent competitive inhibition of α-KG-dependent dioxygenase, which leads to the abnormal expression of a series of signal conductors, thereby influencing transcription regulation, angiogenesis, differentiation, proliferation, anti-apoptosis activity, transference, and gene invasion (Ward et al., 2010; Frezza et al., 2010; Andersson et al., 2011; Westman et al., 2013).

The biological pathways regulated by the various molecular complexes were separately enriched in our study: compound 1 regulated the Jak-STAT signaling pathway, neurotrophin signaling pathway, Wnt signaling pathway, Toll-like receptor signaling pathway, Toll receptor signaling pathway, RIG-I-like receptor signaling pathway, and erythrocyte differentiation pathway; compound 2 was associated with cancer pathways that were associated with different types of cancer, including chronic myeloid leukemia and acute myeloid leukemia; compound 3 regulated complex biological behaviors, including that of hematopoietic cell lineage, immune function, signal transcription, angiogenesis, cytokine activation and function, and inflammatory response. Previous studies have indicated that myeloproliferative diseases are highly sensitive to cytokines, and the genes initiating and regulating the related signal pathways could offer a means for the molecular treatment and cure of PV. The identified compounds are representative of the complicated biological behavior of PV, and its extensive correlation with other tumors; additionally, these compounds describe PV as a network system composed of various signal pathways and genes. Therefore, PV can be considered as a complicated disease requiring coordinated regulation of multiple genes and signal pathways (and not caused by any individual gene or signal pathway). However, this network could contain a specific "target spot" to facilitate its treatment. The genes and signal pathways identified in this experiment should offer a reliable direction for further studies into, and treatment of, PV. However, our results must be further verified in the lab.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Amberger J, Bocchini CA, Scott AF and Hamosh A (2009). McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.* 37: D793-D796. http://dx.doi.org/10.1093/nar/gkn665

Andersson AK, Miller DW, Lynch JA, Lemoff AS, et al. (2011). IDH1 and IDH2 mutations in pediatric acute leukemia. *Leukemia* 25: 1570-1577. http://dx.doi.org/10.1038/leu.2011.133

Ashihara E, Kawata E, Nakagawa Y, Shimazaski C, et al. (2009). beta-catenin small interfering RNA successfully suppressed progression of multiple myeloma in a mouse model. *Clin. Cancer Res.* 15: 2731-2738. http://dx.doi.org/10.1158/1078-0432.CCR-08-1350

Bader GD and Hogue CW (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2. http://dx.doi.org/10.1186/1471-2105-4-2

Burkard TR, Rix U, Breitwieser FP, Superti-Furga G, et al. (2010). A computational approach to analyze the mechanism of action of the kinase inhibitor bafetinib. *PLOS Comput. Biol.* 6: e1001001. http://dx.doi.org/10.1371/journal.pcbi.1001001

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4: 3. http://dx.doi.org/10.1186/gb-2003-4-5-p3

Feng Z and Levine AJ (2010). The regulation of energy metabolism and the IGF-1/mTOR pathways by the p53 protein. *Trends*

Cell Biol. 20: 427-434. http://dx.doi.org/10.1016/j.tcb.2010.03.004

Feng Z, Zhang C, Wu R and Hu W (2011). Tumor suppressor p53 meets microRNAs. *J. Mol. Cell Biol.* 3: 44-50. http://dx.doi.org/10.1093/jmcb/mjq040

Frezza C, Tennant DA and Gottlieb E (2010). IDH1 mutations in gliomas: when an enzyme loses its grip. *Cancer Cell* 17: 7-9 http://apps.cytoscape.org/apps/bingo. http://dx.doi.org/10.1016/j.ccr.2009.12.031

Huang W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57. http://dx.doi.org/10.1038/nprot.2008.211

Khan NI, Bradstock KF and Bendall LJ (2007). Activation of Wnt/beta-catenin pathway mediates growth and survival in B-cell progenitor acute lymphoblastic leukaemia. *Br. J. Haematol.* 138: 338-348. http://dx.doi.org/10.1111/j.1365-2141.2007.06667.x

Kralovics R, Guan Y and Prchal JT (2002). Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. *Exp. Hematol.* 30: 229-236. http://dx.doi.org/10.1016/S0301-472X(01)00789-5

Kumar M, Lu Z, Takwi AA, Chen W, et al. (2011). Negative regulation of the tumor suppressor p53 gene by microRNAs. *Oncogene* 30: 843-853. http://dx.doi.org/10.1038/onc.2010.457

Levine RL and Gilliland DG (2008). Myeloproliferative disorders. *Blood* 112: 2190-2198. http://dx.doi.org/10.1182/blood-2008-03-077966

Rakheja D, Konoplev S, Su M, Wheeler D, et al. (2011). High incidence of IDH mutations in acute myeloid leukaemia with cuplike nuclei. *Br. J. Haematol.* 155: 125-128. http://dx.doi.org/10.1111/j.1365-2141.2011.08646.x

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, et al. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-1555. http://dx.doi.org/10.1126/science.1073374

Rawlings JS, Rosler KM and Harrison DA (2004). The JAK/STAT signaling pathway. *J. Cell Sci.* 117: 1281-1283. http://dx.doi.org/10.1242/jcs.00963

Riley T, Sontag E, Chen P and Levine A (2008). Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.* 9: 402-412. http://dx.doi.org/10.1038/nrm2395

Shannon P, Markiel A, Ozier O, Baliga NS, et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498-2504. http://dx.doi.org/10.1101/gr.1239303

Teodoro JG, Parker AE, Zhu X and Green MR (2006). p53-mediated inhibition of angiogenesis through up-regulation of a collagen prolyl hydroxylase. *Science* 313: 968-971. http://dx.doi.org/10.1126/science.1126391

Thol F, Weissinger EM, Krauter J, Wagner K, et al. (2010). IDH1 mutations in patients with myelodysplastic syndromes are associated with an unfavorable prognosis. *Haematologica* 95: 1668-1674. http://dx.doi.org/10.3324/haematol.2010.025494

Vailaya A, Bluvas P, Kincaid R, Kuchinsky A, et al. (2005). An architecture for biological information extraction and representation. *Bioinformatics* 21: 430-438. http://dx.doi.org/10.1093/bioinformatics/bti187

Valencia A, Román-Gómez J, Cervera J, Such E, et al. (2009). Wnt signaling pathway is epigenetically regulated by methylation of Wnt antagonists in acute myeloid leukemia. *Leukemia* 23: 1658-1666. http://dx.doi.org/10.1038/leu.2009.86

Wang Z, Jin H, Xu R, Mei Q, et al. (2009). Triptolide downregulates Rac1 and the JAK/STAT3 pathway and inhibits colitis-related colon cancer progression. *Exp. Mol. Med.* 41: 717-727. http://dx.doi.org/10.3858/emm.2009.41.10.078

Ward PS, Patel J, Wise DR, Abdel-Wahab O, et al. (2010). The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* 17: 225-234. http://dx.doi.org/10.1016/j.ccr.2010.01.020

Weerkamp F, van Dongen JJ and Staal FJ (2006). Notch and Wnt signaling in T-lymphocyte development and acute lymphoblastic leukemia. *Leukemia* 20: 1197-1205. http://dx.doi.org/10.1038/sj.leu.2404255

Westman MK, Pedersen-Bjergaard J, Andersen MT and Andersen MK (2013). IDH1 and IDH2 mutations in therapy-related myelodysplastic syndrome and acute myeloid leukemia are associated with a normal karyotype and with der(1;7)(q10;p10). *Leukemia* 27: 957-959. http://dx.doi.org/10.1038/leu.2012.347

Wu QL, Zierold C and Ranheim EA (2009). Dysregulation of Frizzled 6 is a critical component of B-cell leukemogenesis in a mouse model of chronic lymphocytic leukemia. *Blood* 113: 3031-3039. http://dx.doi.org/10.1182/blood-2008-06-163303

Xiong W, Wu X, Starnes S, Johnson SK, et al. (2008). An analysis of the clinical and biologic significance of TP53 loss and the identification of potential novel transcriptional targets of TP53 in multiple myeloma. *Blood* 112: 4235-4246. http://dx.doi.org/10.1182/blood-2007-10-119123

Zhou C, Teng WJ, Yang J, Hu ZB, et al. (2014). Construction of a protein-protein interaction network for chronic myelocytic leukemia and pathway prediction of molecular complexes. *Asian Pac. J. Cancer Prev.* 15: 5325-5330. http://dx.doi.org/10.7314/APJCP.2014.15.13.5325