A network biology approach evaluating the anticancer effects of bortezomib identifies SPARC as a therapeutic target in adult T-cell leukemia cells

Junko H Ohyashiki[1]
Ryoko Hamamura[2]
Chiaki Kobayashi[2]
Yu Zhang[2]
Kazuma Ohyashiki[2]

[1]Intractable Immune System Disease Research Center, Tokyo Medical University, Tokyo, Japan; [2]First Department of Internal Medicine, Tokyo Medical University, Tokyo, Japan

Correspondence: Junko H Ohyashiki
Intractable Immune System Diseases Research Center, Tokyo Medical University, 6-7-1, Nishi-shinjuku, Shinjuku-ku, Tokyo, 160-0023, Japan
Tel +81 3 3342 6111 ext. 5999
Fax +81 3 5381 6651
Email junko@hh.iij4u.or.jp

Abstract: There is a need to identify the regulatory gene interaction of anticancer drugs on target cancer cells. Whole genome expression profiling offers promise in this regard, but can be complicated by the challenge of identifying the genes affected by hundreds to thousands of genes that induce changes in expression. A proteasome inhibitor, bortezomib, could be a potential therapeutic agent in treating adult T-cell leukemia (ATL) patients, however, the underlying mechanism by which bortezomib induces cell death in ATL cells via gene regulatory network has not been fully elucidated. Here we show that a Bayesian statistical framework by VoyaGene® identified a secreted protein acidic and rich in cysteine (SPARC) gene, a tumor-invasiveness related gene, as a possible modulator of bortezomib-induced cell death in ATL cells. Functional analysis using RNAi experiments revealed that inhibition of the expression SPARC by siRNA enhanced the apoptotic effect of bortezomib on ATL cells in accordance with an increase of cleaved caspase 3. Targeting SPARC may help to treat ATL patients in combination with bortezomib. This work shows that a network biology approach can be used advantageously to identify the genetic interaction related to anticancer effects.

Keywords: network biology, adult T cell leukemia, bortezomib, SPARC

Introduction

Gene expression microarrays yield quantitative and semiquantitative data on the cell status in a specific condition and time. The aim is to infer, or 'reverse-engineer', from gene expression data, the regulatory interactions among genes using computational algorithms. The interaction between two genes in a gene network does not necessarily imply a physical interaction, but can also refer to an indirect regulation via proteins, metabolites and noncoding RNA that have not been measured directly. In practical terms, however, the meaning of interactions is not well defined and depends on mathematical formations used to model networks (Bansal et al 2007). In this study, we show that a network biology approach can be used for this. In particular, we show that reverse-engineered gene networks can be combined with expression profiles to identify the genes which may play an important role in genetic networks. We used a Bayesian network using a computational tool, VoyaGene® (Mitsui Knowledge Industry, Tokyo, Japan), which has been validated as a means to identify the genes related to viral infection (Takaku et al 2005).

Adult T-cell leukemia/lymphoma (ATL), an aggressive malignancy of CD4+ lymphocytes, is caused by the retrovirus HTLV-1 (Matsuoka and Jeang 2005). ATL is resistant to chemotherapy and has a very poor prognosis, which underlines the need for new and effective therapeutic approaches. The proteasome inhibitor,

bortezomib (also known as Velcade or PS-341), represents a new class of anticancer drugs which has been shown to inhibit the growth and/or progression of human cancers, including multiple myeloma (Mitsiades et al 2002; Cavo 2006; Rajkumar et al 2006). Using DNA microarray analysis, we have recently shown that bortezomib-induced cell death via multiple pathways in ATL cell lines and induction of heme oxygenase-1 by cobalt protoporphyrin enhances the antitumour effect of bortezomib in adult T-cell leukemia cells (Hamamura et al 2007). Despite the fact that bortezomib affects several pathways critical for the survival of HTLV-1-positive and negative malignant T cell (Satou et al 2004; Nasr et al 2006), the gene regulatory network by which bortezomib inhibits ATL cell growth has not been fully elucidated. In addition, threshold-dominated analysis of DNA microarray analysis hides critical genes which may interfere with large numbers of genes.

The above points led us to hypothesize that a network biology approach might be able to indicate regulatory interactions among genes associated with the anticancer effect of bortezomib on ATL cells. We constructed a pathway-focused oligonucleotide array which contained a limited number of genes, and explored a gene network-based analysis, seeking to evaluate the molecular pathways which might be involved in the bortezomib-induced antiproliferative effect on ATL cells.

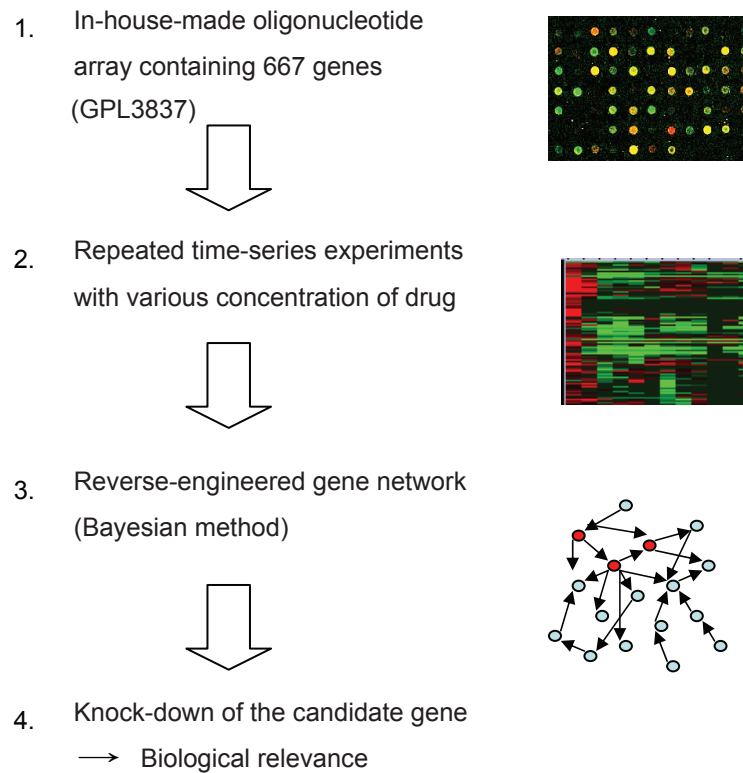## Materials and methods
Reagents and cells
Time-course experiments were done as follows; An ATL cell line, designated, TaY (Zhou 1998) was treated with 0.5 to 100 nM of bortezomib for 8 h, 12 h, 24 h, 48 h, and 72 h. The antiproliferative effect of bortezomib on TaY (IC_{50}; 5 nM) has been reported elsewhere (Hamamura et al 2007). Two interleukin-2 (IL-2) independent ATL cell lines (MT-2, and MT-4) and, after obtaining written informed consent, peripheral blood cells isolated from a patient with ATL were also used in a part of our study (Hamamura et al 2007). Bortezomib was kindly provided by Millennium Pharmaceutials Inc., (Cambridge, MA, USA). The inhibitory effect of bortezomib on cell growth was assessed by a cell counting kit (Wako Chemicals, Tokyo, Japan). For detection of apoptosis, the Annexin V-binding capacities of the treated cells were examined by Annexin V-biotin apoptosis detection kit (Calbiochem, La Jolla, CA, USA) using an Agilent 2100 Bioanalyzer (Agilent, MA, USA), according to the manufacturer's instructions.

DNA microarray analysis, data validation, and statistics
A schematic diagram of our strategy to identify the gene network which may be involved with bortezomib-induced cell death in ATL cells is shown in Figure 1. We designed an in-house pathway-focused low-density oligonucleotide microarray (GPL 3837: Novusgene Inc., Tokyo, Japan) in order to perform repeated time course experiments and construct a gene network by the following Bayesian method. For dual color analysis, cDNA obtained from untreated cells was labeled with Cy3 as a reference, and cDNA obtained from the bortezomib-treated cells was labeled with Cy5. Hybridization was carried out automatically using GeneTAC Hybstation (Genomic Solutions, AnnArbor, MI, USA), as reported previously (Zhang et al 2006). The hybridization signals were scanned by GenePix 4000B (Axon Instruments, Union City, CA, USA). For statistical analysis of gene expression, we utilized a GeneSifter® (VizXLabs, Seattle, WA, USA). Analysis of variance (ANOVA), and Student's t-test, were done using GeneSifter®. P-values of less than 0.05 were considered to indicate a statistically significant difference and the Benjamini-Hochberg algorithm was used for estimation of false discovery rates, as reported previously (Zhang et al 2006).

Genetic network analysis using a VoyaGene®
We next performed a gene network analysis employing a Bayesian network using the VoyaGene® to construct an artificial genetic network from the microarray gene expression data. The Bayesian network is a graph-based model that statistically investigates the characteristics of dependence and conditional independence between data sets of variables (Maki et al 2004). We exploited plausible network structures from the causal relationships with the statistically highest possibility between the variables. Since it is difficult to investigate all network structure comprehensively for large number of genes by conventional Bayesian networks, we established an original evaluation formula we named "score value" focusing on the binary relations included in the network structure. When we consider graphs over three nodes X1, X2 and X3, the total number of possible network structures is 25 (Figure 2). The detailed formula for the score value of each binary relations included in network structure, Sk ($k = 1, 2,...$, the number of binary relations) is shown in Supplementary File 1. The closer the value of Sk approaches 1.0, the greater the possibility that the binary relation k exists. In contrast, the closer the value of Sk approached

1. In-house-made oligonucleotide array containing 667 genes (GPL3837)

2. Repeated time-series experiments with various concentration of drug

3. Reverse-engineered gene network (Bayesian method)

4. Knock-down of the candidate gene
 $\longrightarrow$ Biological relevance

Figure 1 A schematic diagram of gene expression analysis using a Bayesian method as applied to identify genetic interaction for anticancer effect of bortezomib in ATL cells.

zero, the less likely the existence of the binary relation k. In the current study, we selected relations with score value was greater than 0.2 to validate the correlation of each gene.

Small interfering RNA (siRNA)

To confirm the results obtained from the network-based analysis, RNAi experiments were done. siRNA oligonucleotides for the secreted protein acidic and rich in cysteine (SPARC) and β-actin (control) were purchased from Ambion (Austin, TX, USA) and resuspended in RNase-free-H_2O as 20 μM solutions according to the manufacturer's instructions. TaY cells were transfected with SPARC or control siRNA in the presence or absence of 5n M of bortezomib. For cell transfection, approximately 5×10^5 cells were plated in six-well plates
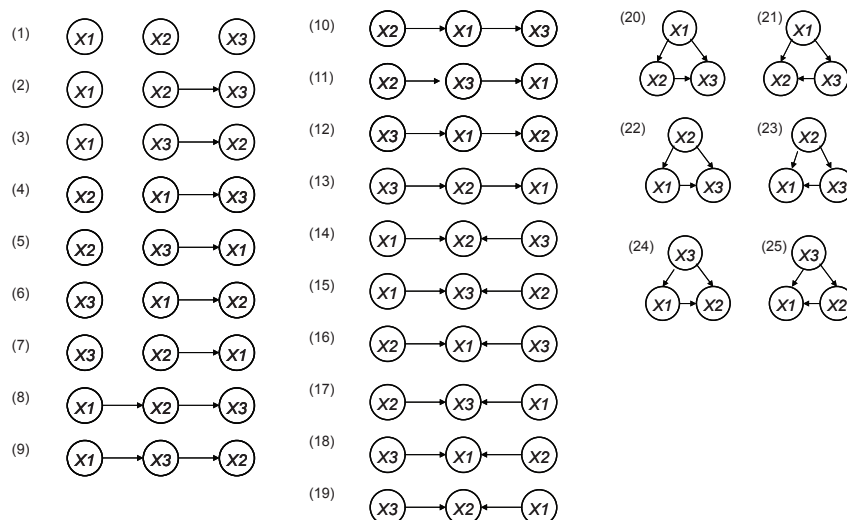
Figure 2 Estimated binary relations among 3 genes. When we consider permutations of three nodes X1, X2 and X3, the total number of the possible network structures is 25.

to give 50% confluency. The cells were transfected with a final siRNA concentration of 30 nM using siPORT™ NeoFX™ Transfection Agent (Ambion). After 48 h, cell viability and apoptosis assays were done. Efficacy of transfection was evaluated by Western blotting as well as real-time RT-PCR as reported elsewhere (Hamamura et al 2007). Primary antibodies used in the current study are anti-SPARC, anti-vascular endothelial growth factor-B (anti-VEGF-B), fibroblast growth factor-1 (FGF-1) (Santa Cruz Biotechnology, Santa Cruz, CA, USA), antiactin (Chemicon), and anti-GAPDH (Abcam Plc., Cambridge, UK).

Results

Gene network analysis by Bayesian statistical framework in bortezomib-treated TaY cells

Repeated time-series experiments were done using in-house oligonucleotide DNA microarray which contains a limited number of genes (NCBI Gene expression omnibus, GSE5794, released), since the Bayesian method requires abundant experimental data sets to obtain statistically significant results. To seek how each gene interfered with each other in bortezomib-induced cell death, we analyzed the relationship of 667 genes by the Bayesian statistical network analysis using VoyaGene®. We extracted several relations the score value of which was greater than 0.2, and the list of selected genes was attached as Supplementary File 2. Based on our evaluation system (Supplementary File 1), we extracted an artificial gene network involving 3 genes: SPARC, NLR family, pyrin domain containing 12 (NAPL12), ankyrin 1 (ANK1). Among them, we in particular focused a gene network involving SPARC (Figure 3). Artificial gene network involving NALP12 and ANK1 are shown in Supplementary File 3.

Down-regulation of SPARC in bortezomib-treated ATL cells

To validate whether SPARC indeed reflect molecular pathways in bortezomib-treated ATL cells, we analyzed the gene expression levels of SPARC in bortezomib-treated ATL cell lines as well as ATL cells obtained from a patient.
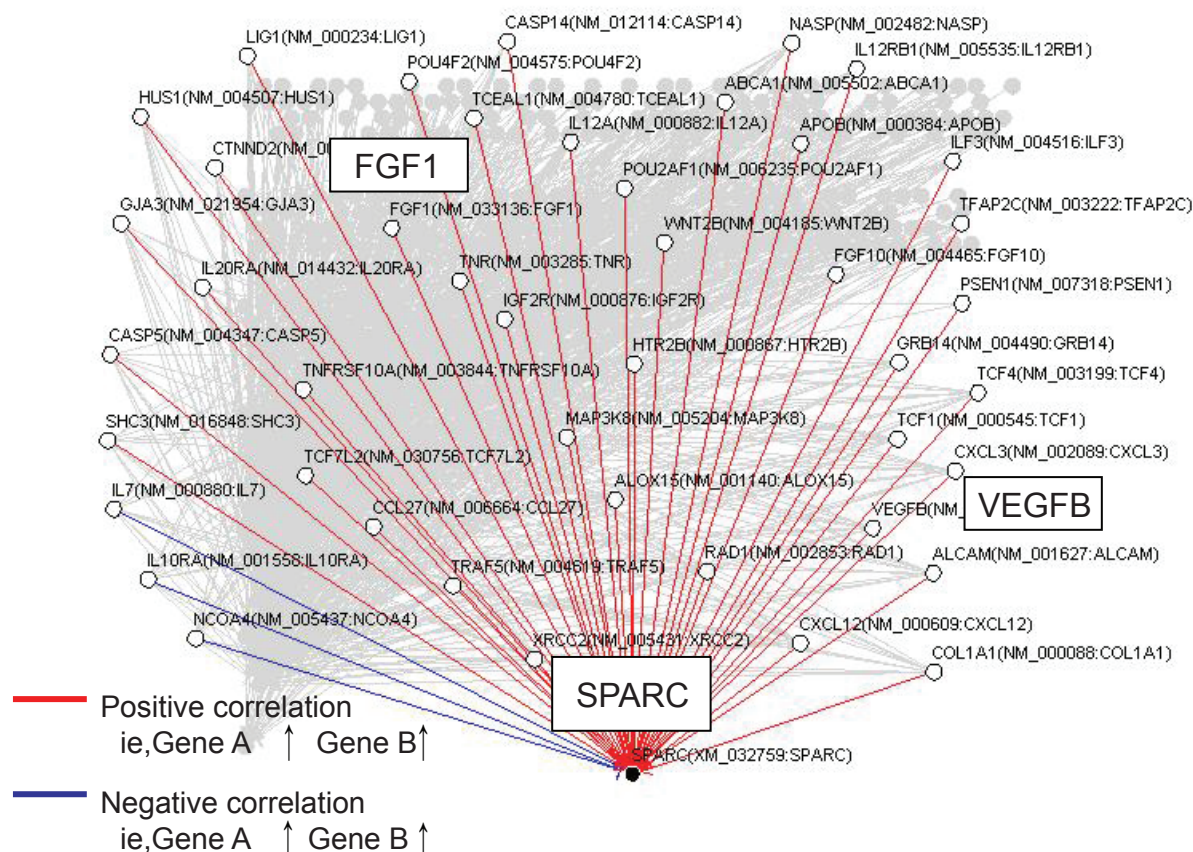
Figure 3 Artificial gene regulatory network integrated to SPARC conduced by VoyaGene®. Genes are analyzed by the Bayesian statistical network analysis. Nodes indicate genes and edges represent regulations between genes; red line (when gene A is up-regulated, gene B is up-regulated), blue line (when gene A is up-regulated, gene B is down-regulated).

The SPARC expression level was elevated in bortezomib-treated fresh ATL cells as well as an ATL cell lines. This indicates that the molecular pathway involving SPARC may play some roles in *de novo* ATL cells treated with bortezomib (Figure 4).

Knock-down of SPARC enhance the bortezomib-induced cell death in TaY cells

To test the functional importance of SPARC in bortezomib-induced cell death, we performed the RNAi experiment in the presence or absence of pharmacological dose of bortezomib. TaY cells were transfected with control siRNA (β-actin) or SPRC siRNA. The SPARC siRNA specifically reduced SPARC expression, while the β-actin siRNA did not affect the SPARC expression level (Figure 5A). The reduction of SPARC expression was more evident in cells treated with bortezomib and SPARC siRNA. The control siRNA showed modest reduction of β-actin expression at the protein level, however, real-time PCR revealed the reduction of β-actin 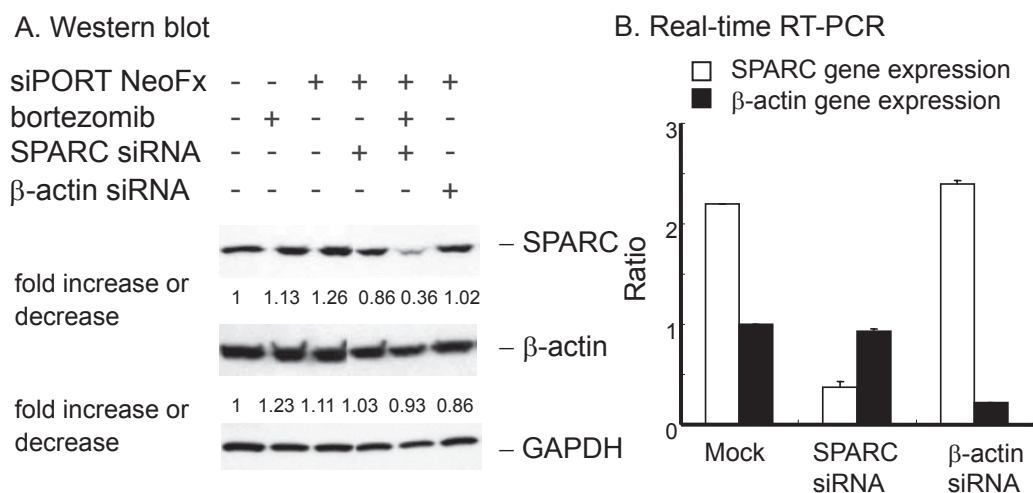mRNA (Figure 5B). We then collected cells at 48 h after transfection, to measure cell viability and apoptosis. Inhibition of SPARC gene expression by SPARC siRNA alone did not affect the cell viability, however, it enhanced the antiproliferative activity of bortezomib (p = 0.042) (Figure 5). In contrast, inhibition of control gene expression did not affect antiproliferative activity of bortezomib. The synergic effect of bortezomib and SPARC siRNA with respect to cell viability was relatively small possibly due to cytotoxicity of bortezomib, increasing apoptotic effect of SPARC siRNA was more evident in the presence of 5 nM of bortezomib (p = 0.011) (Figure 7).

Effect of SPARC RNAi on VEGF-B and FGF1 expression

To elucidate the possible correlation between genes involved in the artificial network, we next analyzed the effect of SPARC RNAi on VEGF-B and FGF-1. In the presence of bortezomib, inhibition of SPARC gene expression remarkably suppressed either VEGF-B or FGF-1 expression (Figure 8). This indicates that positive correlations figured out by the Bayesian method make perfect sense in bortezomib-treated TaY cells.
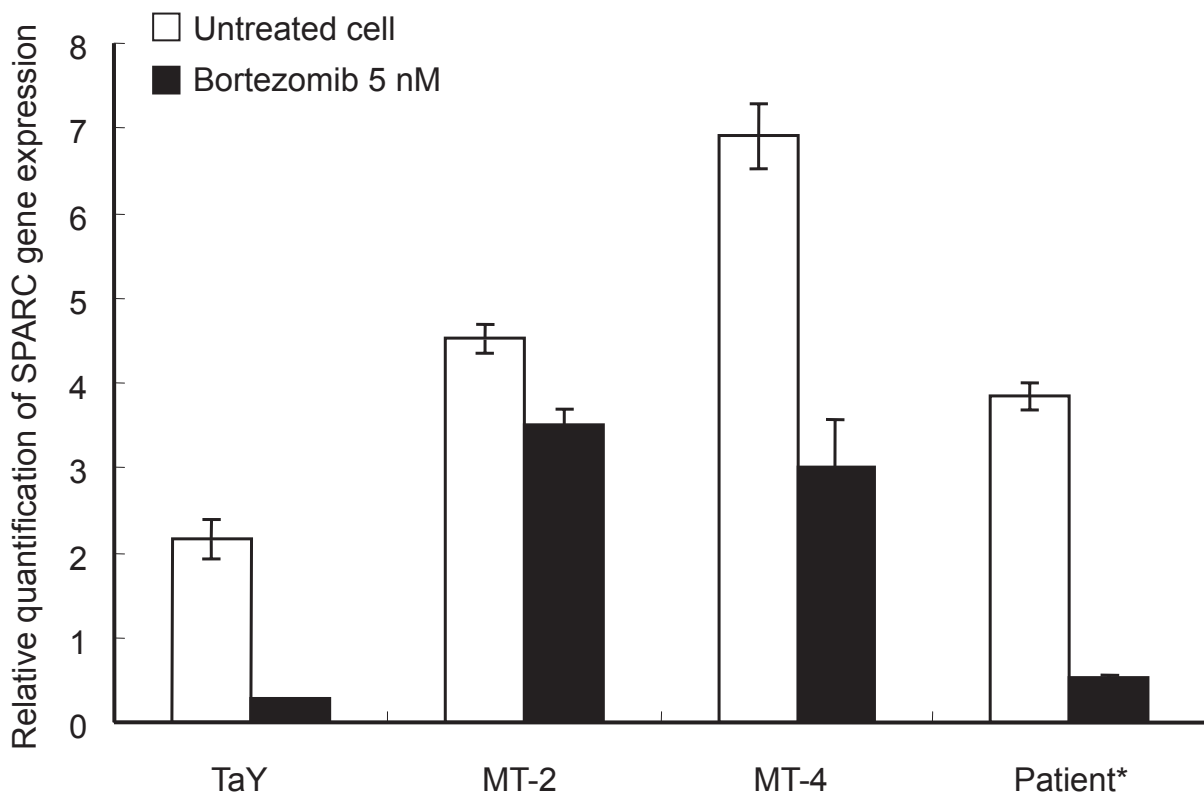
Figure 4 Comparison of gene expression levels between cell lines and a patient sample. Percent of ATL cells in the peripheral blood specimens is 76% respectively. Relative gene expression levels are expressed as ratios (copy numbers of target gene / copy numbers of GAPDH). In the presence of bortezomib, and down-regulation of SPARC is evident in both cell lines and a patient sample.

A. Western blot

siPORT NeoFx	-	-	+	+	+	+
bortezomib	-	+	-	-	+	-
SPARC siRNA	-	-	-	+	+	-
β-actin siRNA	-	-	-	-	-	+

– SPARC

fold increase or decrease

1   1.13 1.26 0.86 0.36 1.02

– β-actin

fold increase or decrease

1   1.23 1.11 1.03 0.93 0.86

– GAPDH

B. Real-time RT-PCR

☐ SPARC gene expression
■ β-actin gene expression

Figure 5 Effect of SPRC siRNA on TaY cells. Twenty four hours after transfection with SPARC or control siRNA, bortezomib is added to the culture. A: Western blot reveals that SPARC siRNA induce modest reduction of SPARC protein levels in TaY cells, while SPARC siRNA does not affect β-actin protein levels. SPARC expression is remarkably reduced when cells were treated with SPARC siRNA followed by bortezomib. B: Reduction of target gene by RNAi: relative gene expression levels are expressed as ratios (copy numbers of target gene/copy numbers of GAPDH).

In addition, SPRC RNAi directly affects the increase of cleaved caspase 3. This suggests that SPARC RNAi enhances apoptosis in the presence of bortezomib, accompanied with VEGF-B and FGF-1 inhibition.

Discussion

Construction of gene networks from microarray gene expression data is becoming an important area in the post-genome era (Zhou et al 2003; Hubner et al 2005). Gene network interference algorithms are becoming accurate enough to be of practical value. Recently, Ergun and colleagues (2007) have reported a network biology approach to prostate cancer utilizing the microarray data sets deposited in a public database. In this study, we performed wet-experiments using an *in vitro* model, and performed computational analysis using Bayesian algorithms from the aspects of a network biology approach. Although we

Neofx	-	+	+	+	+	-	+	+	+
SPARC siRNA	-	-	+	-	-	-	-	+	-
β-actin siRNA	-	-	-	+	-	-	-	-	+
bortezomib 5 nM	-	-	-	-	-	+	+	+	+
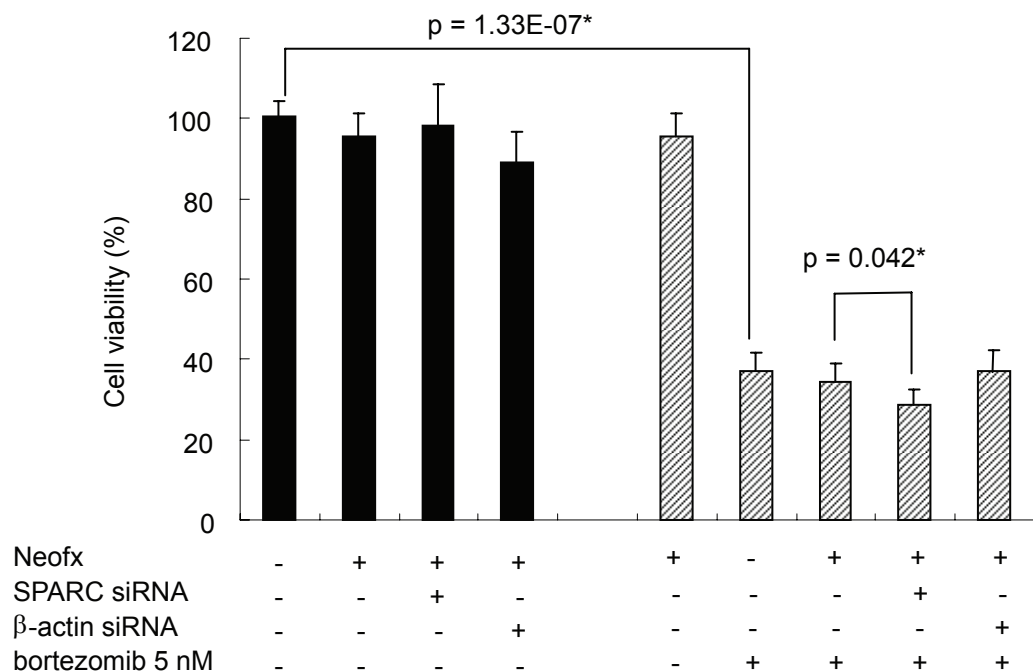
Figure 6 SPARC siRNA enhances bortezomib-induced cell death on TaY cells. Bortezomib induces antiproliferative effect in TaY cells (p = 1.33E-07). Inhibition of SPARC gene expression by SPARC siRNA also enhances the antiproliferative activity of bortezomib (p = 0.039).
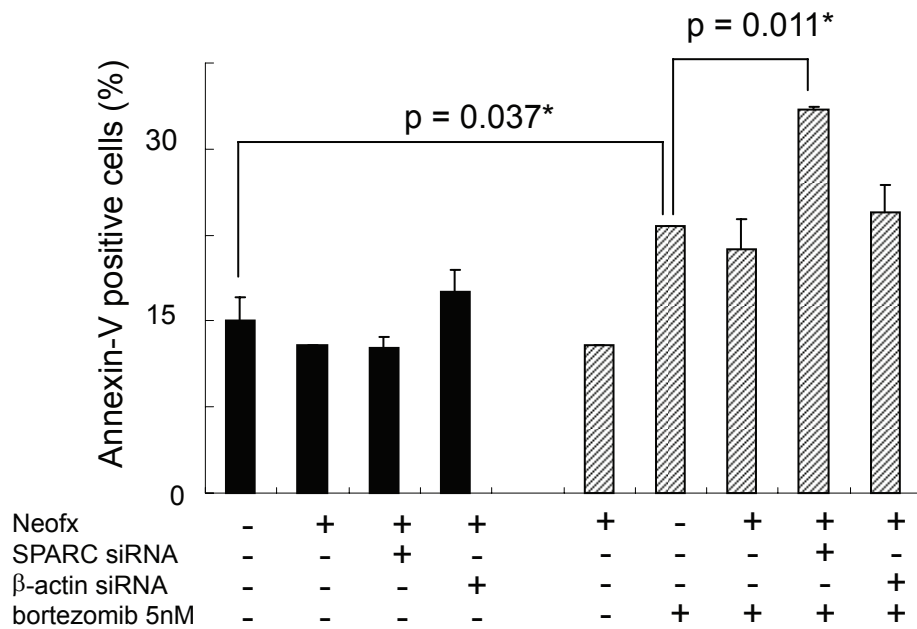
Figure 7 SPARC siRNA enhances bortezomib-induced apoptosis in TaY cells. Apoptosis is expressed as a percent of Annexin V positive cells in bortezomib and or siRNA treated cells. SPARC siRNA induces apoptosis in the presence of 5 nM of bortezomib (p = 0.011).

could not perform network analysis using high-density DNA microarrays such as Affimetrix's GeneChip, our time series data allowed us to investigate the dynamics of activation (inhibition) of genes in response to bortezomib treatment in ATL cells.

Our final goal was to assimilate cellular biological research (ie, wet-experiments) and bioinformatics, and we found that a reverse-engineered gene network can be used to identify the molecular target related to anticancer effect.
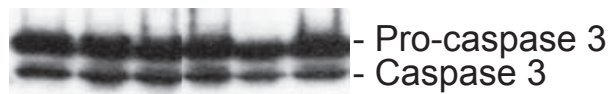

Figure 8 Inhibition of VEGF-B and FGF-1 expression in bortezomib and siRNA treated TaY cells. In the presence of bortezomib, inhibition of SPARC gene expression remarkably suppresses either VEGF-B or FGF-1 expression. In contrast, SPARC siRNA enhanced the cleavage of caspase 3 in the presence of bortezomib, indicating that SPARC siRNA enhanced bortezomib-induced apoptosis in TaY cells.

SPARC/osteonectin/BM-40 is a matrix-cellular protein that is thought to play some roles in tumor progression (Hadvat et al 2003; Kato et al 2005; Rich et al 2005; Chlenski et al 2006; Infante et al 2007; Lien et al 2007). We investigated the possibility of inferring the local network of regulatory interaction surrounding SPARC even when there is no *a priori* knowledge of genes belonging to the network. Recent microarray and immunohistochemical analysis of human cancer indicate a relatively high level of SPARC (Kato et al 2005; Lien et al 2007). These reports have suggested a correlation between the level of SPARC expression and clinical outcome (Rich et al 2005; Infante et al 2007). Conversely, a high level of SPARC was correlated with a better prognosis in neuroblastoma, possibly due to impaired angiogenesis (Chlenski et al 2006). Although the mechanism underlying the capacity of SPARC to promote or inhibit tumor progression in different tumor microenvironments remains under investigation, it may depend on the types of cancer, the source of SPARC, and the site of tumor growth.

Recent studies have demonstrated that SPARC is related to the VEGF functions, tumor angiogenesis, and extravasation of tumors mediated by the increased permeability of the endothelial barrier (Goldblum et al 1994; Kato et al 2001). A recent study demonstrated that ATL- and HTLV-I-associated myelopathy patients exhibit high plasma levels of functional VEGF and basic FGF (Bazabachi et al 2004). The interaction of HTLV-I-transformed cells with endothelial cells induces the gelatinase activity of matrix metalloproteinase (MMP)-2 and MMP-9 in endothelial cells and down-regulates the tissue inhibitor of MMP (Bazabachi et al 2004). This leads to subendothelial basement membrane degradation followed by endothelial cell retraction, allowing neoplastic lymphocyte extravasation. In the current study, we also found a positive correlation between SPARC and FGF-1. It is likely that SPARC acts at the crossroads of cell-matrix communication (Brekken and Sage 2001). Extra-nodal infiltration of ATL cell such as skin and spleen is frequently seen and this invasive characteristic of ATL cells seems to be one factor affecting the poor prognosis in ATL (Yamada et al 1997; Yagi et al 2003). Our results suggest that bortezomib has the potential to prevent ATL cell invasion by inhibiting SPARC. It should obviously be clarified in large numbers of patient specimen.

Conclusion

Our study clearly shows that the reverse-engineered analysis worked adequately and that the extracted gene, SPARC, may be involved in bortezomib-induced cell death in ATL cells. The integration of bioinformatics, system biology, and carefully designed models will provide more insight into the biological role of SPARC. Targeting SPARC may also be effective in treating ATL patients exhibiting extra-nodal lesions such as skin invasion. Applying network-based analysis in addition to gene expression profiling may be a new option to obtain novel insights into the proteasome inhibitor, bortezomib, in not only ATL, but also various types of human cancer.

Acknowledgment

References

Bansal M, Belcastro V, Ambesi-Impiombato A, et al. 2007. How to infer gene networks from expression profiles. *Mol Syst Biol*, 3:7–8.

Bansal M, Gatta GD, di Bernardo D. 2006. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*, 22:815–22.

Bazarbachi A, Abou Merhi R, Gessain A, et al. 2004. Human T-cell lymphotropic virus type I-infected cells extravasate through the endothelial barrier by a local angiogenesis-like mechanism. *Cancer Res*, 264:2039–46.

Brekken RA, Sage EH. 2001. SPARC, a matricellular protein: at the crossroads of cell-matrix communication. *Matrix Biol*, 19:816–27.

Cavo M. 2006. Proteasome inhibitor bortezomib for the treatment of multiple myeloma. *Leukemia*, 20:1341–52.

Chlenski A, Liu S, Guerrero LJ, et al. 2006. SPARC expression is associated with impaired tumor growth, inhibited angiogenesis and changes in the extracellular matrix. *Int J Cancer*, 118:310–16.

Ergun A, Lawrence CA, Kohanski MA, et al. 2007. A network biology approach to prostate cancer. *Mol Syst Biol*, 3:1–8.

Goldblum SE, Ding X, Funk SE, et al. 1994. SPARC (secreted protein acidic and rich in cysteine) regulates endothelial cell shape and barrier function. *Proc Natl Acad Sci U S A*, 91:3448–55.

Hamamura RS, Ohyashiki JH, Kurashina R, et al. 2007. Induction of heme oxygenase-1 by cobalt protoporphyrin enhances the antitumour effect of bortezomib in adult T-cell leukaemia cells. *Br J Cancer*, 97:1099–105.

Hedvat CV, Comenzo RL, Teruya-Feldstein J, et al. 2003. Insights into extramedullary tumour cell growth revealed by expression profiling of human plasmacytomas and multiple myeloma. *Br J Haematol*, 122:728–44.

Hubner N, Wallace CA, Zimdahl H, et al. 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*, 37:243–53.

Infante JR, Matsubayashi H, Sato N, et al. 2007. Peritumoral fibroblast SPARC expression and patient outcome with resectable pancreatic adenocarcinoma. *J Clin Oncol*, 25:319–25.

Kato Y, Lewalle JM, Baba Y, et al. 2001. Expression of SPARC by VEGF in human vascular endothelial cells. *Biochem Biophys Res Commun*, 287:422–6.

Kato Y, Nagashima Y, Baba Y, et al. 2005. Expression of SPARC in tongue carcinoma of stage II is associated with poor prognosis: an immunohistochemical study of 86 cases. *Int J Mol Med*, 16:263–8.

Lien HC, Hsiao YH, Lin YS, et al. 2007. Molecular signatures of metaplastic carcinoma of the breast by large-scale transcriptional profiling: identification of genes potentially related to epithelial-mesenchymal transition. *Oncogene*, 26:7859–71.

Maki Y, Takahashi Y, Arikawa Y, et al. 2004. An integrated comprehensive workbench for inferring genetic networks: voyagene. *J Bioinform Comput Biol*, 2:533–50.

Matsuoka M, Jeang KT. 2005. Human T-cell leukemia virus type I at age 25: a progress report. *Cancer Res*, 65:4467–70.

Mitsiades N, Mitsiades CS, Poulaki V, et al. 2002. Molecular sequelae of proteasome inhibition in human multiple myeloma cells. *Proc Natl Acad Sci U S A*, 99:14374–9.

Nasr R, El-Sabban ME, Karam JA, et al. 2005. Efficacy and mechanism of action of the proteasome inhibitor PS-341 in T-cell lymphomas and HTLV-I associated adult T-cell leukemia/lymphoma. *Oncogene*, 24:419–30.

Rajkumar SV, Richardson PG, Hideshima T, et al. 2006. Proteasome inhibition as a novel therapeutic target in human cancer. *J Clin Oncol*, 23:630–9.

Rich JN, Hans C, Jones B, et al. 2005. Gene expression profiling and genetic markers in glioblastoma survival. *Cancer Res*, 65:4051–8.

Satou Y, Nosaka K, Koya Y, et al. 2004. Proteasome inhibitor, bortezomib, potently inhibits the growth of adult T-cell leukemia cells both in vivo and in vitro. *Leukemia*, 18:1357–63.

Takaku T, Ohyashiki JH, Yu Zhang, et al. 2005. Estimating immunoregulatory gene networks in human herpesvirus type 6 (HHV-6)-infected T cells. *Biophys Res Commun*, 336:469–77.

Yagi H, Takigawa M, Hashizume H. 2003. Cutaneous type of adult T cell leukemia/lymphoma: a new entity among cutaneous lymphomas. *J Dermatol*, 30:641–43.

Yamada Y, Kamihira S, Murata K, et al. 1997. Frequent hepatic involvement in adult T cell leukemia: comparison with non-Hodgkin's lymphoma. *Leuk Lymphoma*, 26:327–35.

Zhang Y, Ohyashiki JH, Takaku T, et al. 2006. Transcriptional profiling of Epstein-Barr virus (EBV) genes and host cellular genes in nasal NK/T-cell lymphoma and chronic active EBV infection. *Br J Cancer*, 94:599–608.

Zhou CF, Abe K, Wang P, et al. 1998. Susceptibility of the adult T cell leukemia (ATL) cell lines to HHV-6B. *Leukemia*, 12:1001.

Zhou XJ, Kao MC, Huang H, et al. 2003. Functional annotation and network reconstruction through cross-platform integration of microarray data. *Nat Biotechnol*, 23:238–43.

Supplementary File 1
Formula for score value

We arbitrarily define the possibility of each network structure to be $P(B_i)(i = 1, 2, \ldots, 25)$. In the cases of three-gene networks, six kinds of binary relations can be considered; *X1->X2, X1->X3, X2->X1, X2->X3, X3->X1, X3->X2*. The score value of each binary relations included in network structure, S_k ($k = 1, 2, \ldots$; the number of binary relations), can be defined as follows:

$$S_k = \frac{\Sigma_i \{(-1)^r P(B_i)\}}{\Sigma_i \{P(B_i)\}}$$

($i = 1, 2, \ldots$; the number of possible network structures) (for cases when the binary relation k is not included in the ith network structure, B_i, $r = 1$; for the case when

binary relation k is included in the network structure, B_i, $r = 2$).

The score value of binary relation $X1->X2$ depicted in Figure, defined by S1, can be calculated as follows;

$S1 = \{P(B6) + P(B8) + P(B12) + P(B14) + P(B19) + P(B20) + P(B21) + P(B24)$
$- P(B1) - P(B2) - P(B3) - P(B4) - P(B5) - P(B7) - P(B9)$
$- P(B10) - P(B11) - P(B13)$
$- P(B15) - P(B16) - P(B17) - P(B18) - P(B22) - P(B23)$
$- P(B25)\}/$
$\qquad \Sigma_i \{P(B_i)\}$ $(i = 1, 2, \ldots, 25)$

The closer the value of S_k approaches 1.0, the greater the possibility that the binary relation k exists. In contrast, the closer the value of S_k approached zero, the less likely the existence of the binary relation k.
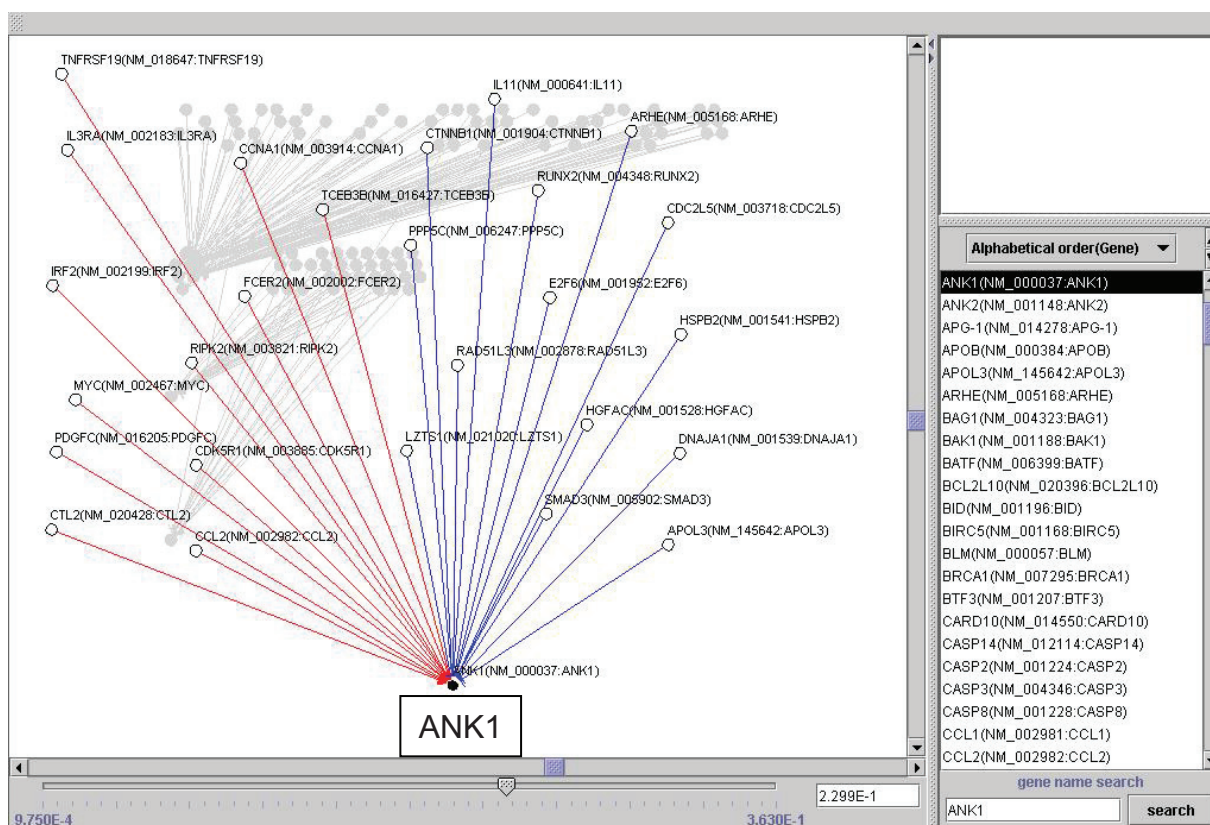
Supplementary File 3a Artificial gene regulatory network integrated to NALP12 conducted by VoyaGene®.
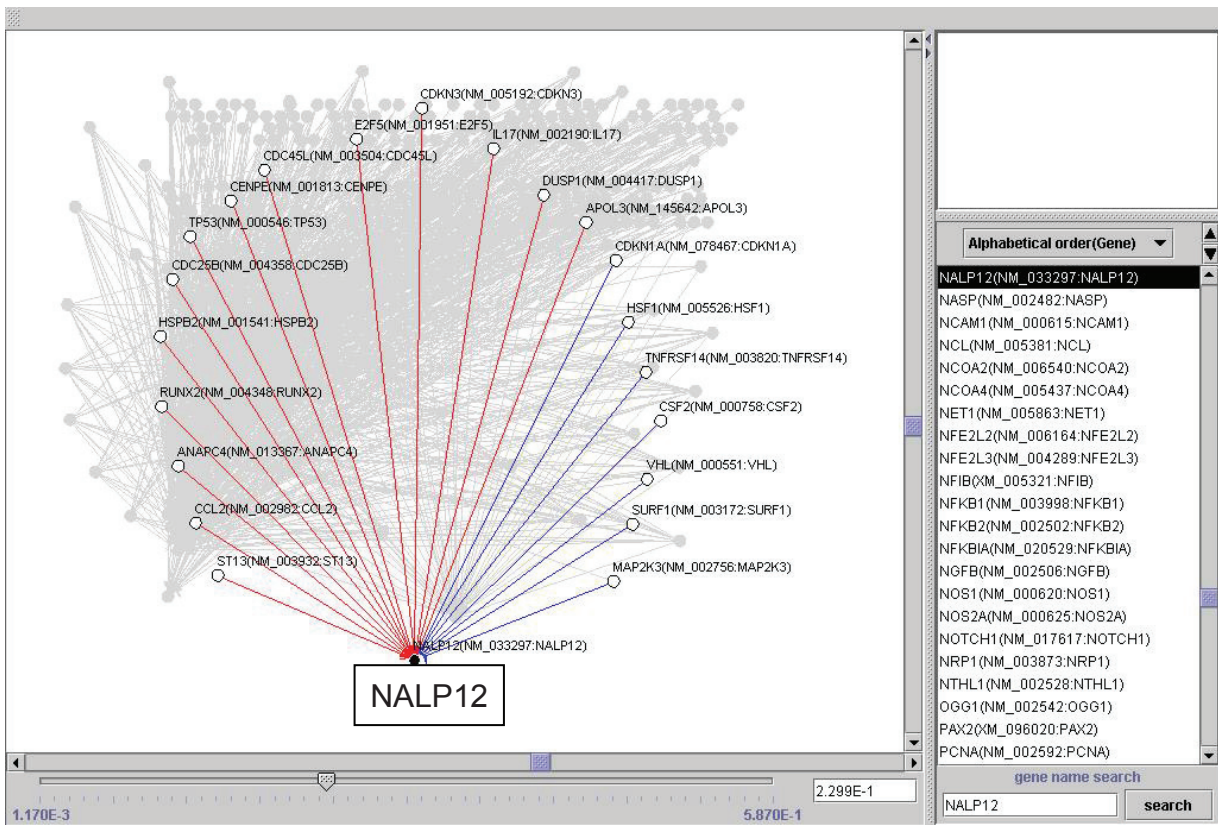
Supplementary File 3b Artificial gene regulatory network integrated to ANK1 conducted by VoyaGene®.

Supplementary File 2
List of genes with scores greater than 0.2 selected by Voyagene® (Raw data)

Three networks (A to C) with scores greater than 0.2 was shown below.

Relation 1: positive correlation

Relation 1: negative correlation

A: an artificial gene network terminating and integrating into SPARC

B: an artificial gene network terminating and integrating into NALP12

C: an artificial gene network terminating and integrating into ANK1

Genes	Relation	Network	Score value
TCF7L2(NM_030756:TCF7L2)	1	A	0.2299
CCL27(NM_006664:CCL27)	1	A	0.2299
APOB(NM_000384:APOB)	1	A	0.2299
FGF1(NM_033136:FGF1)	1	A	0.2299
TCF1(NM_000545:TCF1)	1	A	0.2299
CXCL3(NM_002089:CXCL3)	1	A	0.2299
RAD1(NM_002853:RAD1)	1	A	0.2299
POU2AF1(NM_006235:POU2AF1)	1	A	0.2299
NCOA4(NM_005437:NCOA4)	−1	A	0.2299
IL12A(NM_000882:IL12A)	1	A	0.2299
TNR(NM_003285:TNR)	1	A	0.2299
WNT2B(NM_004185:WNT2B)	1	A	0.2299
TCEAL1(NM_004780:TCEAL1)	1	A	0.2299
XRCC2(NM_005431:XRCC2)	1	A	0.2299
COL1A1(NM_000088:COL1A1)	1	A	0.2299
CASP5(NM_004347:CASP5)	1	A	0.2299
TFAP2C(NM_003222:TFAP2C)	1	A	0.2299
HTR2B(NM_000867:HTR2B)	1	A	0.2299
ABCA1(NM_005502:ABCA1)	1	A	0.2299
IL7(NM_000880:IL7)	−1	A	0.2299
CXCL12(NM_000609:CXCL12)	1	A	0.2299
IL12RB1(NM_005535:IL12RB1)	1	A	0.2299
TRAF5(NM_004619:TRAF5)	1	A	0.2299
VEGFB(NM_003377:VEGFB)	1	A	0.2299
IGF2R(NM_000876:IGF2R)	1	A	0.2299
IL10RA(NM_001558:IL10RA)	−1	A	0.2299
FGF10(NM_004465:FGF10)	1	A	0.2299
TCF4(NM_003199:TCF4)	1	A	0.2299
ILF3(NM_004516:ILF3)	1	A	0.2299
NASP(NM_002482:NASP)	1	A	0.2299
TNFRSF10A(NM_003844:TNFRSF10A)	1	A	0.2299
ALOX15(NM_001140:ALOX15)	1	A	0.2299
POU4F2(NM_004575:POU4F2)	1	A	0.2299
CTNND2(NM_001332:CTNND2)	1	A	0.2299
IL20RA(NM_014432:IL20RA)	1	A	0.2299
GJA3(NM_021954:GJA3)	1	A	0.2299
HUS1(NM_004507:HUS1)	1	A	0.2299

(*Continued*)

Supplementary file 2 (*Continued*)

Genes	Relation	Network	Score value
CASP14(NM_012114:CASP14)	I	A	0.2299
PSEN1(NM_007318:PSEN1)	I	A	0.2299
GRB14(NM_004490:GRB14)	I	A	0.2299
LIG1(NM_000234:LIG1)	I	A	0.2299
ALCAM(NM_001627:ALCAM)	I	A	0.2299
MAP3K8(NM_005204:MAP3K8)	I	A	0.2299
SHC3(NM_016848:SHC3)	I	A	0.2299
SURF1(NM_003172:SURF1)	−I	B	0.2299
ANAPC4(NM_013367:ANAPC4)	I	B	0.2299
CENPE(NM_001813:CENPE)	I	B	0.2299
VHL(NM_000551:VHL)	−I	B	0.2299
MAP2K3(NM_002756:MAP2K3)	−I	B	0.2299
ST13(NM_003932:ST13)	I	B	0.2299
CDC25B(NM_004358:CDC25B)	I	B	0.2299
CDC45L(NM_003504:CDC45L)	I	B	0.2299
TP53(NM_000546:TP53)	I	B	0.2299
CCL2(NM_002982:CCL2)	I	B	0.2299
HSPB2(NM_001541:HSPB2)	I	B	0.2299
RUNX2(NM_004348:RUNX2)	I	B	0.2299
CSF2(NM_000758:CSF2)	−I	B	0.2299
DUSP1(NM_004417:DUSP1)	I	B	0.2299
CDKN1A(NM_078467:CDKN1A)	−I	B	0.2299
APOL3(NM_145642:APOL3)	I	B	0.2299
CDKN3(NM_005192:CDKN3)	I	B	0.2299
HSF1(NM_005526:HSF1)	−I	B	0.2299
TNFRSF14(NM_003820:TNFRSF14)	−I	B	0.2299
E2F5(NM_001951:E2F5)	I	B	0.2299
IL17(NM_002190:IL17)	I	B	0.2299
CCL2(NM_002982:CCL2)	I	C	0.2299
RUNX2(NM_004348:RUNX2)	−I	C	0.2299
FCER2(NM_002002:FCER2)	I	C	0.2299
RIPK2(NM_003821:RIPK2)	I	C	0.2299
CDK5R1(NM_003885:CDK5R1)	I	C	0.2299
DNAJA1(NM_001539:DNAJA1)	−I	C	0.2299
SMAD3(NM_005902:SMAD3)	−I	C	0.2299
CTNNB1(NM_001904:CTNNB1)	−I	C	0.2299
CCNA1(NM_003914:CCNA1)	I	C	0.2299
PPP5C(NM_006247:PPP5C)	−I	C	0.2299
CDC2L5(NM_003718:CDC2L5)	−I	C	0.2299
RAD51L3(NM_002878:RAD51L3)	−I	C	0.2299
IRF2(NM_002199:IRF2)	I	C	0.2299
HSPB2(NM_001541:HSPB2)	−I	C	0.2299
APOL3(NM_145642:APOL3)	−I	C	0.2299
IL3RA(NM_002183:IL3RA)	I	C	0.2299
ARHE(NM_005168:ARHE)	−I	C	0.2299
LZTS1(NM_021020:LZTS1)	−I	C	0.2299

Supplementary file 2 (*Continued*)

Genes	Relation	Network	Score value
PDGFC(NM_016205:PDGFC)	I	C	0.2299
MYC(NM_002467:MYC)	I	C	0.2299
CTL2(NM_020428:CTL2)	I	C	0.2299
TCEB3B(NM_016427:TCEB3B)	I	C	0.2299
E2F6(NM_001952:E2F6)	−I	C	0.2299
TNFRSF19(NM_018647:TNFRSF19)	I	C	0.2299
IL11(NM_000641:IL11)	−I	C	0.2299
HGFAC(NM_001528:HGFAC)	−I	C	0.2299