

## RESEARCH ARTICLE

# Integrative Meta-Analysis of Multiple Gene Expression Profiles in Acquired Gemcitabine-Resistant Cancer Cell Lines to Identify Novel Therapeutic Biomarkers

Young Seok Lee<sup>1</sup>, Jin Ki Kim<sup>1</sup>, Seoung Won Ryu<sup>1</sup>, Se Jong Bae<sup>1</sup>, Kang Kwon<sup>2</sup>, Yun Hee Noh<sup>1</sup>, Sung Young Kim<sup>1\*</sup>

## Abstract

In molecular-targeted cancer therapy, acquired resistance to gemcitabine is a major clinical problem that reduces its effectiveness, resulting in recurrence and metastasis of cancers. In spite of great efforts to reveal the overall mechanism of acquired gemcitabine resistance, no definitive genetic factors have been identified that are absolutely responsible for the resistance process. Therefore, we performed a cross-platform meta-analysis of three publically available microarray datasets for cancer cell lines with acquired gemcitabine resistance, using the R-based RankProd algorithm, and were able to identify a total of 158 differentially expressed genes (DEGs; 76 up- and 82 down-regulated) that are potentially involved in acquired resistance to gemcitabine. Indeed, the top 20 up- and down-regulated DEGs are largely associated with a common process of carcinogenesis in many cells. For the top 50 up- and down-regulated DEGs, we conducted integrated analyses of a gene regulatory network, a gene co-expression network, and a protein-protein interaction network. The identified DEGs were functionally enriched via Gene Ontology hierarchy and Kyoto Encyclopedia of Genes and Genomes pathway analyses. By systemic combinational analysis of the three molecular networks, we could condense the total number of DEGs to final seven genes. Notably, *GJA1*, *LEF1*, and *CCND2* were contained within the lists of the top 20 up- or down-regulated DEGs. Our study represents a comprehensive overview of the gene expression patterns associated with acquired gemcitabine resistance and theoretical support for further clinical therapeutic studies.

**Keywords:** meta-analysis - microarray - DEG - acquired drug resistance - gemcitabine

*Asian Pac J Cancer Prev*, 16 (7), 2793-2800

## Introduction

Although there have been innumerable clinical trials over the last few decades related to the prevention, detection, and treatment of cancer, acquired drug resistance remains an inevitable obstacle to successful chemotherapy in various types of cancer, causing treatment failure in over 90% of patients with metastatic cancer (Szakacs et al., 2006; Saraswathy and Gong, 2013). Acquired drug resistance may be the result of a multifactorial etiology that emerges due to a variety of reasons, including host environmental factors, as well as genetic or epigenetic alterations in cancer cells (Longley and Johnston, 2005; Chen and Sikic, 2012; Foo and Michor, 2014). According to the most well known reported studies, acquired drug resistance in cancer cells could be caused by three major mechanisms, as follows: (i) decreased uptake of water-soluble drugs which require membrane transporters to enter cancer cells, (ii) various changes in cancer cells that affect the capacity of cytotoxic drugs to kill them, including alterations in the cell cycle, increased repair of

DNA damage, reduced apoptosis, and altered metabolism of drugs, and (iii) increased energy-dependent efflux of hydrophobic drugs.

Gemcitabine (2', 2'-difluorodeoxycytidine) is a deoxycytidine analog with a broad spectrum that was established as the first-line chemotherapeutic treatment for locally advanced and metastatic pancreatic cancer in the late 1990s; it has generally been used to treat a variety of solid tumors, including breast, ovarian, pancreatic and non-small cell lung carcinoma, especially in combination with the platinum-based drugs, cisplatin and carboplatin (Toschi et al., 2005; Mini et al., 2006; Toschi and Cappuzzo, 2009; Zhang et al., 2013; de Sousa Cavalcante and Monteiro, 2014). Gemcitabine is first transported into the cancer cell by nucleoside transporters, which include the concentrative and equilibrative transporters. When phosphorylated by deoxycytidine kinase (DCK) to generate its active forms, the diphosphate and triphosphate, gemcitabine interferes with DNA replication and inhibits cancer cell growth, by modulating dNTP pools via the inhibition of ribonucleotide reductase

<sup>1</sup>Department of Biochemistry, School of Medicine, Konkuk University, Seoul, <sup>2</sup>School of Korean Medicine, Pusan National University, Pusan, Republic of Korea \*For correspondence: [palelamp@kku.ac.kr](mailto:palelamp@kku.ac.kr)

(composed of RRM1 and RRM2 polypeptides). Several mechanisms of acquired gemcitabine resistance (AGR) in cancers have been reported; these are anatomical (e.g., desmoplasia, epithelial-mesenchymal transition (EMT), and inherent cancer stem cell resistance), pathophysiological (e.g., abnormal tumor growth, tumor angiogenesis, altered cancer cell survival, and anti-apoptosis), or pharmacological (e.g., the necessity of phosphorylation for prodrug activation) (Nakano et al., 2007; Toschi and Cappuzzo, 2009; Tufman and Huber, 2010; Hung et al., 2012; de Sousa Cavalcante and Monteiro, 2014). However, none of these mechanisms was confirmed as the etiology of AGR, and the fundamental reason for-and exact process of-AGR is still being studied, in many different ways. Despite gemcitabine's relatively broad and frequent use, the acquired resistance to it during cancer treatment is a common phenomenon that is caused by the establishment of complex crosstalk between many different cellular pathways. High-throughput microarray technologies, based on genome-wide probe selection and regression analysis, are widely used to elucidate global gene expression in complex diseases such as cancer. By applying these technologies, researchers have improved understanding of the cellular and molecular changes occurring during the development of acquired drug resistance in cancers. In three previous studies that used microarrays, various differentially expressed genes (DEGs) were identified as candidate factors that may influence AGR (Tooker et al., 2007; Saiki et al., 2012). However, published lists of identified genes show large inconsistencies because of the small sample size, low sample quality, and different laboratory protocol and platform in each individual study. In order to overcome these limitations, we identified DEGs that consistently appeared in all pairwise samples by meta-analysis of multiple microarray datasets, and performed integrative analysis of systemic molecular networks at the gene and/or protein level, in order to establish a theoretical framework for prospective molecular biological and clinical experiments. To our knowledge, we are the first to perform a cross-platform meta-analysis of the gene expression profiles associated with AGR in different cancer cell lines.

## Materials and Methods

### *Selection of microarray datasets eligible for meta-analysis*

We conducted a narrow search of microarray datasets for meta-analysis, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines published in 2009. We collated data from microarray gene expression studies related to acquired-gemcitabine-resistant cancer cell lines from the PubMed, National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), and European Bioinformatics Institute ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) databases. For objective assessment, two independent reviewers extracted data from the original studies; any discrepancies that arose between these reviewers were resolved either by

consensus or by consultation with a third reviewer. The keywords "gemcitabine," "gemcitabine cancer," "acquired gemcitabine resistance (resistant)," "acquired drug resistance (resistant)," and "gene and/or expression and/or profile" were used in the search for studies. We included a study in the analysis if it contained the following: (i) gene expression profiling of gemcitabine-resistant cancer cell lines or gemcitabine-resistant derivatives of cancer cell lines that had been generated by stepwise selection, and (ii) sufficient data and the correct platform to facilitate meta-analysis. Studies that reported non-human data or used intrinsically drug-resistant cells were excluded from the meta-analysis.

### *Meta-analysis of multiple microarray datasets*

We carried out a cross-platform meta-analysis of gene expression profiles in the selected microarray datasets using the rank product method (RankProd package in the R software, <http://www.r-project.org/>). Before meta-analysis of the three datasets, all genes or probe IDs from each dataset were annotated as Entrez database IDs, for data consistency, and the expression values for corresponding genes in the samples were  $\log_2$ -transformed and quantile-normalized so that they had zero mean and unit variance. According to the RankProd algorithm non-parametric permutation test, which considers all possible pair-wise comparisons, the DEGs that appeared consistently in whole datasets were assigned to a higher rank, depending on the percentage of false-positives predicted in a given number of replicates, multiplied across the different microarray datasets.

### *GO hierarchy and KEGG pathway enrichment analysis*

In order to discern biological attributes of the identified DEGs in the acquired-gemcitabine-resistant cancer cell lines, functional enrichment via Gene Ontology (GO) hierarchy and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed using Database for Annotation, Visualization, and Integrated Discovery (DAVID) bioinformatics resources (<http://david.abcc.ncifcrf.gov/>), with a significance threshold of  $p < 0.05$ .

### *Gene regulatory network analysis*

In order to determine the gene regulatory network of the identified DEGs, we carried out enrichment analysis for potential transcription factors and microRNAs, based on a comparison of upstream DNA sequences with an assembly of previously discovered gene sets retrieved from the Molecular Signatures database (MSigDB, <http://www.broadinstitute.org/gsea/msigdb/index.jsp>), with a significance threshold of  $p < 0.05$  (Shi et al., 2014). The hypergeometric algorithm and Benjamini-Hochberg adjustment were used for statistical processing and multiple-test revision of the network analysis, respectively.

### *Gene co-expression network analysis*

In order to predict the biological activity of the identified DEGs at the gene level, we constructed a gene co-expression network of the top 50 up- and down-regulated DEGs using the GeneMANIA web server (<http://www.genemania.org/>) (Warde-Farley et al., 2010;

Molina-Navarro et al., 2014). The relationships between genes in the network were determined by GO term (biological process)-based weighting, and filtered by including only gene co-expression relationships with a significance threshold for weight value of  $p > 0.05$ . Within the network, distinct modules were identified based on the fast-greedy HEN (G) algorithm, using the Community Clusters GLayer plug-in (<http://cytoscape.wodaklab.org/wiki/CommunityStructureLayout>) in the Cytoscape software (<http://www.cytoscape.org/>) (Gupta et al., 2012; Firoz et al., 2014). Overrepresented biological functions within each module were examined using the functional enrichment analyses in the DAVID and g:Profiler (<http://biit.cs.ut.ee/gprofiler/>) programs.

#### Protein-protein interaction network analysis

In order to predict the biological activity of the identified DEGs at the protein level, the top 50 (20 up- and down-regulated DEGs) were imported into a protein-protein interaction (PPI) network downloaded from the Biological General Repository for Interaction Datasets (BioGRID, <http://thebiogrid.org/>, (Pan, 2012). The PPI network was screened on a genome-wide scale using the Cytoscape software. Within the PPI network, we identified the hub proteins of distinct protein clusters using the Cytoscape plug-in, ClusterONE (<http://apps.cytoscape.org/apps/clusterone>) (Nepusz et al., 2012).

Overrepresented biological functions within each protein cluster surrounding a hub protein were examined by the functional enrichment analyses in the DAVID and g:Profiler online programs.

## Results

#### Selection of microarray datasets related to AGR for meta-analysis

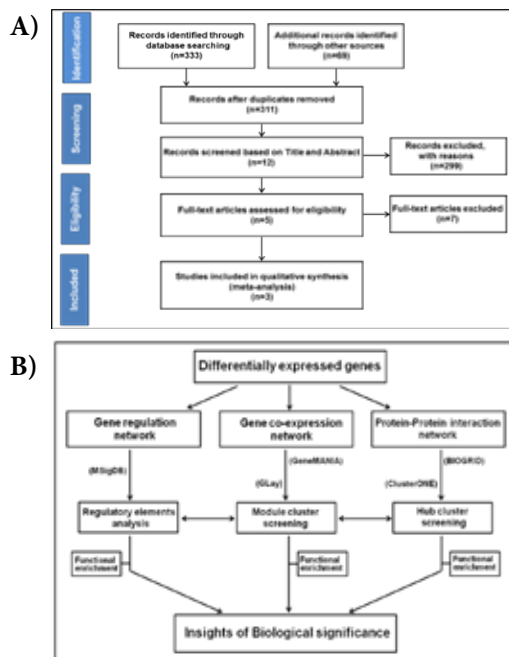
From microarray datasets in the NCBI GEO database, we extracted 22 GEO samples within three GEO series (GSEs) related to AGR that met our criteria (see Materials and Methods; Figure 1A). All three GSEs were solely derived from cancer cell lines that had acquired drug resistance by stepwise treatment with gemcitabine, such as lung adenocarcinoma, pancreatic cancer, and epidermoid carcinoma (Table 1). With regard to the microarray platforms used, GSE 35141 was obtained using Illumina BeadChip (Illumina, San Diego, CA) and the other two datasets (GSE 6914 and GSE 3344) were obtained using Affymetrix GeneChip (Affymetrix, Santa Clara, CA).

#### Identification of up- or downregulated DEGs in the meta-analysis

From cross-platform microarray meta-analysis based on the RankProd algorithm, we identified a total of 158 DEGs, including 76 up- and 82 down-regulated genes with a significance threshold of  $p < 0.05$ . The top 20 up- and down-regulated genes among the total DEGs are listed in order of significance (by p value) in Table 2. The up-regulated DEGs with the lowest value of  $p < 1.0 \times 10^{-10}$  were *CALB1*, *ADAM28*, *TRIM22*, *MSMB*, *TLE4*, *INHBB*, *ADH1C*, *IL1R2*, and *TRIM21*. The down-regulated DEGs with the lowest observed value of  $p = 0.00250$  were *ARHGAP29* and *PTX3*. The up- and down-regulated DEGs with the largest mean  $\log_2$  fold change were *CALB1* (calbindin 1, 28 kDa) and *ARHGAP29* (Rho GTPase activating protein 29), respectively. In order to interpret the biological significance of the identified DEGs in different cancer cell lines with AGR, we attempted the systemic approach of using various *in silico* analyses that might identify gene regulation, gene co-expression, and PPI networks, accompanied by functional enrichment analysis (Figure 1B).

#### Functional and pathway enrichment analysis of all the identified DEGs

A total of 158 DEGs identified by the meta-analysis were classified according to GO hierarchy functional category (biological process, molecular function, and cellular component) and KEGG pathway, with a significance threshold of  $p < 0.05$  (Table 3). The most overrepresented GO terms under biological process were



**Figure 1. General Workflow Depicting the Systematic Review of this Study.** (A) Selection of microarray datasets for meta-analysis of the acquired gemcitabine resistant cancer cell lines, according to Prisma 2009 flow diagram; (B) The integrated in-silico analysis of DEGs identified by meta-analysis

**Table 1. Characteristics of Individual Studies Selected from GEO of NCBI for Meta-analysis**

Dataset	Sample		Drug	Cancer cell line	Platform
	PC	AGR			
GSE 6914	4	4	Gemcitabine	lung adenocarcinoma (Calu3)	Affymetrix Human Genome U133A
GSE 35141	6	6	Gemcitabine	pancreatic cancer (PK)	Agilent-014850 Whole Human Genome Microarray 4x44K
GSE 3344	1	1	Gemcitabine	epidermoid carcinoma (KB)	Affymetrix Human Genome U95 Version 2

\*GEO gene expression omnibus, PC parental control, AGR acquired gemcitabine resistance

**Table 2. The top 20 Most Strongly Up- or Down-regulated Genes in the DEGs Identified by Meta-analysis**

Enterz ID	Gene symbol	Log2 FC	p value	Gene name
Up-regulated genes				
793	CALB1	-2.20630		calbindin 1, 28kDa
10863	ADAM28	-1.99019		metallopeptidase domain 28
10346	TRIM22	-1.78943		tripartite motif containing 22
4477	MSMB	-1.76001		microseminoprotein, beta-
7091	TLE4	-1.73700	<1.0E-5	transducin-like enhancer of split 4
3625	INHBB	-1.40772		inhibin, beta B
126	ADH1C	-1.33476		alcohol dehydrogenase 1C (class I), gamma polypeptide
7850	IL1R2	-1.31214		interleukin 1 receptor, type II
6737	TRIM21	-1.11247		tripartite motif containing 21
25849	PARM1	-1.66358	0.00250	prostate androgen-regulated mucin-like protein 1
1004	CDH6	-1.63775	0.00272	cadherin 6, type 2, K-cadherin (fetal kidney)
51176	LEF1	-1.49318	0.00300	lymphoid enhancer-binding factor 1
2697	GJA1	-1.45898	0.00307	gap junction protein, alpha 1, 43kDa
4856	NOV	-1.70069	0.00357	nephroblastoma overexpressed
124	ADH1A	-1.34103	0.00437	alcohol dehydrogenase 1A (class I), alpha polypeptide
1002	CDH4	-1.48827	0.00466	cadherin 4, type 1, R-cadherin (retinal)
10231	RCAN2	-1.21286	0.00588	regulator of calcineurin 2
6304	SATB1	-0.38100	0.00611	SATB homeobox 1
6240	RRM1	-0.93649	0.00631	ribonucleotide reductase M1
6649	SOD3	-0.88427	0.00636	superoxide dismutase 3, extracellular
Down-regulated genes				
9411	ARHGAP29	1.94459	0.00250	Rho GTPase activating protein 29
5806	PTX3	1.54702	0.00250	pentraxin 3, long
50810	HDFGRP3	1.60953	0.00285	hepatoma-derived growth factor, related protein 3r
72	ACTG2	1.28665	0.00300	actin, gamma 2, smooth muscle, enteric
8187	ZNF239	1.37927	0.00333	zinc finger protein 239
6696	SPP1	1.34688	0.00333	secreted phosphoprotein 1
26150	RIBC2	1.32767	0.00333	RIB43A domain with coiled-coils 2
6414	SEPP1	2.21272	0.00400	selenoprotein P, plasma, 1
7292	TNFSF4	1.50729	0.00461	tumor necrosis factor (ligand) superfamily, member 4
7431	VIM	1.61811	0.00470	vimentin
1382	CRABP2	1.24545	0.00473	cellular retinoic acid binding protein 2
1633	DCK	2.13356	0.00500	deoxycytidine kinase
2313	FLI1	1.39167	0.00500	Fli-1 proto-oncogene, ETS transcription factor
4633	MYL2	1.28676	0.00500	myosin, light chain 2, regulatory, cardiac, slow
5272	SERPINB9	1.02120	0.00500	serpin peptidase inhibitor, clade B (ovalbumin), member 9
147	ADRA1B	0.82309	0.00500	adrenoceptor alpha 1B
2947	GSTM3	0.80997	0.00521	glutathione S-transferase mu 3 (brain)
10669	CGREF1	1.78892	0.00533	cell growth regulator with EF-hand domain 1
6515	SLC2A3	1.68493	0.00545	solute carrier family 2 (facilitated glucose transporter), member 3
894	CCND2	1.13272	0.00545	cyclin D2

\*FC fold change

enriched in the following descending order: “Defense response” (GO 0006952), “Response to extracellular stimulus” (GO 0009991), and “Response to drug” (GO 0042493). The most enriched GO terms under molecular function and cellular component were “Tumor necrosis factor receptor superfamily binding” (GO 0032813) and “Extracellular region (GO 0005576). The most enriched KEGG pathway terms were (descending order): “Cytokine-cytokine receptor interaction” (Hsa 04060), “Metabolism of xenobiotics by cytochrome P450” (Hsa 00980), and “Regulation of actin cytoskeleton” (Hsa 04810).

#### Gene regulation network analysis of the top 50 up- and down-regulated DEGs

In order to identify the network regulating gene expression of the top 50 up- and down-regulated DEGs, which might directly influence AGR, we analyzed potential regulatory elements that target the DEGs

depending on their upstream DNA sequence (Table 4). The target sites of the following transcription factors were significantly enriched in the DEGs: JUN, LEF1, NFAT, MAZ, MLLT4, and TCF1. The target sites of the following microRNAs were also significantly enriched in the DEGs: miR-200B/200C/429, miR-19A/19B, miR-520G/520H, miR-524, miR-23A/23B, miR-153, miR-409, miR145, miR-9, and miR-129.

#### Gene co-expression network analysis of the top 50 up- and down-regulated DEGs

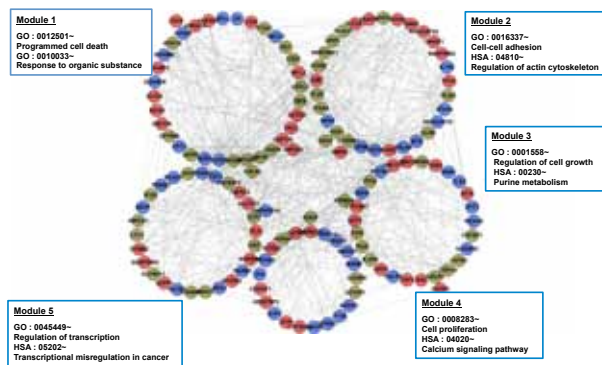
We constructed a co-expression network for the top 50 up- and down-regulated DEGs, with gene-correlation interactions consisting of 143 nodes and 764 edges, by mapping the DEGs onto a massive database of functional-interaction datasets in the GeneMANIA web server (Figure 2). The network was further subdivided into five functional modules that were closely connected by >20 nodes, using the fast-greedy HEN (G) algorithm

**Table 3. The Top 15enrichment of GO Hierarchy and KEGG Pathway for the Total DEGs**

GO ID	GO Term	Genes	p value
GO_CC:0005576	Extracellular region	45	9.92E-07
GO_MF:0032813	Tumor necrosis factor receptor superfamily binding	5	3.91E-04
GO_BP:0006952	Defense response	15	3.17E-03
GO_BP:0009991	Response to extracellular stimulus	8	6.28E-03
GO_BP:0042493	Response to drug	7	2.05E-02
GO_BP:0008284	Positive regulation of cell proliferation	10	2.19E-02
GO_BP:0043067	Regulation of programmed cell death	15	3.02E-02
GO_MF:0008083	Growth factor activity	6	3.61E-02
GO_BP:0051046	Regulation of secretion	6	5.04E-02
GO_BP:0007267	Cell-cell signaling	11	7.44E-02
GO_CC:0031988	Membrane-bounded vesicle	11	8.21E-02
GO_BP:0006355	Regulation of transcription, DNA-dependent	20	8.99E-02
GO_CC:0031090	Organelle membrane	13	1.03E-01
GO_BP:0043933	Macromolecular complex subunit organization	8	1.74E-01
GO_MF:0030528	Transcription regulator activity	16	1.95E-01

KEGG ID	KEGG Pathway	Genes	p value
Hsa:04060	Cytokine-cytokine receptor interaction	8	2.82E-02
Hsa:00980	Metabolism of xenobiotics by cytochrome P450	4	3.04E-02
Hsa:04810	Regulation of actin cytoskeleton	7	3.45E-02
Hsa:04270	Vascular smooth muscle contraction	5	3.81E-02
Hsa:04020	Calcium signaling pathway	5	4.39E-02
Hsa:00982	Drug metabolism	3	4.57E-02
Hsa:04510	Focal adhesion	5	4.95E-02
Hsa:05215	Prostate cancer	3	5.79E-02
Hsa:04540	Gap junction	3	6.98E-02
Hsa:00240	Pyrimidine metabolism	3	7.36E-02
Hsa:04010	MAPK signaling pathway	5	1.53E-01
Hsa:04530	Tight junction	3	2.46E-01
Hsa:04310	Wnt signaling pathway	3	3.18E-01
Hsa:04144	Endocytosis	3	4.62E-01
Hsa:05200	Pathways in cancer	4	5.30E-01



**Figure 2. Functional Modules Clusters in Gene Co-expression Network of the Identified DEGs.** From gene co-expression network of the top 50 up- and down-regulated DEGs, five functional module clusters were identified. The node and edge of each module cluster stand for genes with the DEGs and interaction of the genes, respectively. The color of node signifies as follows: Light blue-up-regulated DEGs, Light red-down-regulated DEGs, and Light brown-additional genes in GeneMANIA

of the Cytoscape GLayer plug-in, followed by functional enrichment analysis according to GO hierarchy and KEGG pathway. For example, “Module 1,” of maximum size in the network, was significantly enriched using biological terms of the GO hierarchy such as “Programmed cell death” (GO 0012501) and “Response to organic substance” (GO 0010033).

The smallest module, “Module 4,” was significantly

**Table 4. The potential regulatory elements of the top 50 up- and down-regulated DEGs**

Transcription factor	Target sequence	Genes	p value
JUN	TGANTCA	20	1.93E-09
LEF1	CTTTGT	26	3.56E-09
NFAT	TGGAAA	25	7.38E-09
MAZ	GGGAGRR	27	1.54E-08
MLLT4	TTGTTT	25	3.76E-08
TCF1	WRGTAAATNA	9	2.51E-07
	TTAACNNN		

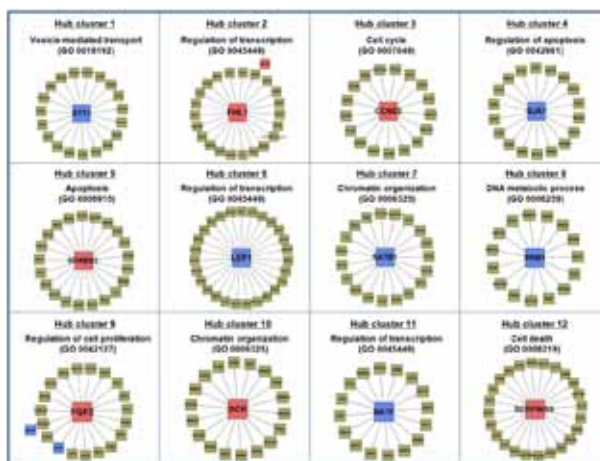
microRNA	Target sequences	Genes	p value
MIR-200B/200C/429	CAGTATT	14	1.02E-09
MIR-19A/19B	TTTGAC	11	1.84E-06
MIR-520G/520H	CACTTTG	7	1.88E-05
MIR-524	CTTTGTA	9	2.00E-05
MIR-23A/23B	AATGTGA	6	1.05E-04
MIR-153	CTATGCA	8	1.05E-04
MIR-409-3P	AACATTC	5	1.34E-04
MIR-145	AACGGGA	6	1.63E-04
MIR-9	TAGCTTT	6	1.70E-04
MIR-129	GCAAAA	5	4.33E-04

enriched using biological terms such as “Cell proliferation” (GO 0008283) and “Calcium signaling pathway” (Hsa 04020), with regard to GO hierarchy and KEGG pathway, respectively.

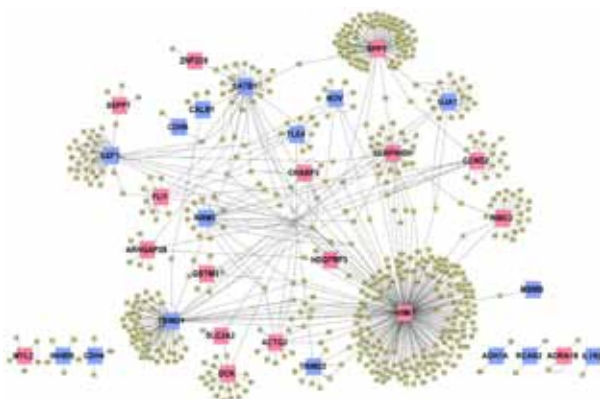
*Protein-protein interaction network analysis of the top 50 up- and down-regulated DEGs*

We constructed a PPI network for the top 50 up- and





**Figure 3. Functional Hub Clusters in PPI Network of the Identified DEGs.** From PPI network of proteins encoded by the top 50 up- and down-regulated DEGs, twelve functional hub clusters were identified. The node and edge of each hub cluster stand for protein encoded by genes with the identified DEGs and interaction of the proteins, respectively. The color of node signifies proteins that are encoded by the following DEGs: Light blue-up-regulated DEGs, Light red-down-regulated DEGs, and Light brown-additional genes in BioGRID



**Figure 4. The Constructed PPI Network of the Top 20 Up- and Down-regulated DEGs.** The PPI network of the top 20 up- and down-regulated DEGs was constructed by mapping them into massive database of BIOGRID program. The node and edge of each hub cluster stand for protein encoded by genes with the identified DEGs and interaction of the proteins, respectively. The color of node signifies proteins that are encoded by the following DEGs: Light blue-up-regulated DEGs, Light red-down-regulated DEGs, and Light brown-additional genes in BioGRID

down-regulated DEGs, composed of 1004 nodes and 1186 edges, by mapping them onto a very large database of PPI datasets downloaded from the BioGRID web server (Figure 3). From fifteen distinct protein clusters surrounding hub proteins, with >15 nodes, twelve functional hub cluster proteins were specifically identified, based on their p value and node density derived using the Cytoscape ClusterONE plug-in followed by functional enrichment analysis by GO hierarchy, as follows: SYT1, FHL1, CCND2, GJA1, SORBS2, LEF1, SATB1, RRM1, FGF2, DCK, BATF, and SERPINB9. While six of these hub proteins (SYT1, GJA1, LEF1, SATB1, RRM1, and BATF) represented up-regulated DEGs, the other six (FHL1, CCND2, SORBS2, FGF2, DCK, and SERPINB9)

represented down-regulated DEGs. By constructing a PPI network for the top20 up- and down-regulated DEGs, we also observed the whole schematic diagram that the two kinds of DEGs interact directly or indirectly in the network (Figure 4).

## Discussion

Gemcitabine is a molecular-targeted cancer drug for the standard chemotherapeutic treatment of patients with various solid tumors, but its clinical impact is limited by the high degree of inherent or acquired drug resistance; no definitive genetic factors have been reported to be solely responsible for the AGR of cancers. Investigating alterations in gene expression, which characterize the response of cancer cells to gemcitabine treatment during the process of AGR, would help us to understand the underlying mechanisms of drug resistance and improve the efficacy of therapeutic strategies for this deadly disease. In this context, we performed a cross-platform meta-analysis of three independent microarray datasets and attempted an integrative analysis of three systemic molecular networks (gene regulation network, gene co-expression network, and PPI network) for the identified DEGs. In the case of the top 20 up- and down-regulated DEGs, most of the genes have been reported to be involved in carcinogenesis of many tumor and cancer types, suggesting that they may be potential key factors in the AGR process. Functional enrichment analysis of all the identified DEGs revealed that they were mainly classified as related to biological functions such as the cell cycle, homeostasis, immune response, apoptosis, replication, and signal transduction that are associated with the general process of carcinogenesis. In particular, the following GO and KEGG enrichment terms had direct relevance to the mechanisms by which cancers acquire their gemcitabine-resistant property: the GO terms, “Response to drug,” “Regulation of programmed cell death,” “Regulation of secretion,” and “Regulation of transcription, DNA-dependent,” and the KEGG terms, “Metabolism of xenobiotics by cytochrome P450,” “Drug metabolism,” “Pyrimidine metabolism,” “MAPK and Wnt signaling pathway,” “Endocytosis,” and “Pathways in cancer” (Nakano et al., 2007; Toschi and Cappuzzo, 2009; Tufman and Huber, 2010; Hung et al., 2012; de Sousa Cavalcante and Monteiro, 2014). In evaluating the biological significance of the identified DEGs in the complex process of AGR, comprehensive information on the topological positions of the DEGs within a network at the transcriptome level is no less valuable than the fold-change and p values of individual DEGs. Systemic analysis of the gene regulation network of the identified DEGs and their potential regulatory elements (targets of transcription factors and microRNAs) may facilitate a macroscopic view of AGR, by looking into regulatory mechanisms governing the expression and the function of (many different) genes and cellular processes, respectively, in cancers that acquire gemcitabine resistance. In practice, most of regulatory elements enriched by the top 50 up- and down-regulated DEGs were reported to be involved as oncogenes or tumor suppressors in a variety of human cancers including colon,

gastric, prostate, lung, pancreatic, and breast cancer. For example, among transcription factors, it was reported that activation of the JUN-JNK complex was required for development of AGR in lung cancer H1299 cells, and increased expression of NFAT was correlated with tumor cell survival against apoptosis in drug-resistant pancreatic cancer (Teraishi et al., 2005; Griesmann et al., 2013). LEF1 (lymphoid enhancer-binding factor 1) was identified as a mediator of the Wnt/ $\beta$ -catenin signaling pathway during metastasis of lung adenocarcinoma (Bleckmann et al.). In the case of microRNAs, it was reported that the miR-200 family could serve as regulators of EMT in metastasis of ovarian, breast, and pancreatic cancer, and the attenuated expression of miR-200b/200c was found in gemcitabine-resistant pancreatic cancer cells (Ali et al., 2010). The microRNA, miR-145, is known to be a novel regulator of *MUC13* that is highly involved in the progression of pancreatic cancer, and miR-19a was discovered as a prognostic factor for poor outcome in patients with non-small cell lung cancer (Lin et al., 2013; Khan et al., 2014). In another systemic approach for identifying expression patterns of DEGs that may be related to AGR in cancers, we evaluated the functional enrichment of DEGs into distinct modules, where they are co-located and form functional interactions with each other, within the network that was constructed by mapping the DEGs onto the massive gene co-expression database of the GeneMANIA online resource. Five modules, composed of the top 50 up- and down-regulated DEGs, were identified; genes within these modules and other genes already known to be in the network were largely involved in processes that are representative of AGR, including indefinite cell proliferation, for example, via abnormal apoptosis (modules 1, 3, and 4), intercellular membrane transport of small molecules (modules 2 and 4), deregulated transcription (modules 3 and 4), and arrested replication escape by reactivation of DNA synthesis (module 3). In parallel with the two above-mentioned network analyses, clustering and enrichment of functional hub clusters were analyzed at the protein level within the PPI network in order to identify hub proteins with a high degree of interaction. In many studies, hub nodes have been found to be necessary factors for the specific function that is executed by their corresponding network in an organic system, and play important functions in maintaining that network within the system. The twelve functional hub DEGs identified in the PPI network were enriched using GO terms for biological processes with a close relationship to the AGR process, such as abnormal apoptosis (hub clusters 3, 4, 5, 9, and 12), membrane transport of small molecules (hub cluster 1), deregulated transcription (hub clusters 2, 6, 8, and 11), and arrested replication escape by reactivation of DNA synthesis (hub clusters 7, 8, and 10). By comparing four lists of DEGs, for the gene regulation network, gene co-expression network, PPI network, and the top 50 up- and down-regulated DEGs, seven DEGs were shortlisted as AGR candidate genes from the total of 158 DEGs identified by meta-analysis; these included four up-regulated DEGs (*SYT1*, *GJAI*, *LEF1*, and *SATB1*) and three down-regulated

DEGs (*FHL1*, *CCND2*, and *SORBS2*) that appeared in all four lists. In particular, *GJAI*, *LEF1*, and *CCND2* were affiliated to the lists of the top 20 up- and down-regulated DEGs more likely to be crucial for the etiology of AGR.

LEF1 belongs to a family that shares conserved amino acid sequence homology with high-mobility group protein 1 (Nguyen et al., 2009; Bleckmann et al., 2013). Many previous studies showed that a LEF1/TCF4 complex is closely associated with poor survival in patients with different cancers (colon, gastric, breast, lung, and pancreatic), via regulation of cell proliferation, migration, invasion, and metastasis, as a transcription factor that mediates the Wnt/ $\beta$ -catenin signaling pathway in the nucleus. Another upregulated DEG, *GJAI* (encoding gap junction protein, alpha 1, 43 kDa), is a member of the connexin gene family that encode components of gap junctions, which act as intercellular channels for the diffusion of low molecular weight materials from cell to cell (McLachlan et al., 2006; Li et al., 2007). *GJAI*, better known as connexin 43 (Cx43), was reported to function as a tumor suppressor that inhibits tumor growth, via regulation of EMT and angiogenesis, in breast and prostate cancer. Gene expression analysis of cisplatin-sensitive and -resistant ovarian cancer cells showed that *GJAI* expression was highly elevated in cisplatin-resistant ovarian cancer cells, by whole-genome oligonucleotide microarray analysis, and by immunoblotting and immunofluorescence analyses using a Cx43-specific antibody. *CCND2* (cyclin D2), identified by the down-regulated DEG, is known to function as a regulatory subunit of cyclin-dependent kinases that are involved in the G1/S transition in the mitotic cell cycle (Koyama-Nasu et al., 2013). Alteration of *CCND2* gene expression promoted phosphorylation and subsequent inactivation of the retinoblastoma tumor suppressor protein, RB1, which causes dysregulation of the G1/S transition as a common event in the tumorigenesis of many cancers. There was no choice but to exclude DCK (deoxycytidine kinase) and RRM1, already known as molecular targets in cytotoxic mechanisms of gemcitabine, from the final gene lists; although they were contained in module 3 of the gene co-expression network and hub clusters 8 and 13 of the PPI network (highly related to AGR), we could not analyze their potential regulatory elements in the gene regulation network, owing to computational limitation using databases containing only experimentally discovered relationships.

In conclusion, by performing a cross-platform meta-analysis of three microarray datasets for different cancer cell lines with AGR, we have identified a total of 158 candidate DEGs that have a high probability of being involved in the molecular mechanism of AGR. We have also provided a comprehensive overview of the gene expression pattern of the AGR-related DEGs by attempting integrated *in silico* analysis of three molecular networks. This topological approach of integrative network analysis could help to provide new insights into the complex nature of AGR and may be useful for studying prospective chemotherapeutic strategies.

## Acknowledgements

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government(NRF-2013R1A1A1075999).

## References

- Ali S, Ahmad A, Banerjee S, et al (2010). Gemcitabine sensitivity can be induced in pancreatic cancer cells through modulation of miR-200 and miR-21 expression by curcumin or its analogue CDF. *Cancer Res*, **70**, 3606-17.
- Bleckmann A, Siam L, Klemm F, et al (2013). Nuclear LEF1/TCF4 correlate with poor prognosis but not with nuclear beta-catenin in cerebral metastasis of lung adenocarcinomas. *Clin Exp Metastasis*, **30**, 471-82.
- Chen KG, Sikic BI (2012). Molecular pathways: regulation and therapeutic implications of multidrug resistance. *Clin Cancer Res*, **18**, 1863-9.
- de Sousa Cavalcante L, Monteiro G (2014). Gemcitabine: metabolism and molecular mechanisms of action, sensitivity and chemoresistance in pancreatic cancer. *Eur J Pharmacol*, **741**, 8-16.
- Firoz A, Malik A, Singh SK, et al (2014). Comparative analysis of glycogene expression in different mouse tissues using RNA-Seq data. *Int J Genomics*, **2014**, 837365.
- Foo J, Michor F (2014). Evolution of acquired resistance to anti-cancer therapy. *J Theor Biol*, **355**, 10-20.
- Griesmann H, Ripka S, Pralle M, et al (2013). WNT5A-NFAT signaling mediates resistance to apoptosis in pancreatic cancer. *Neoplasia*, **15**, 11-22.
- Gupta A, Schulze TG, Nagarajan V, et al (2012). Interaction networks of lithium and valproate molecular targets reveal a striking enrichment of apoptosis functional clusters and neurotrophin signaling. *Pharmacogenomics J*, **12**, 328-41.
- Hung SW, Mody HR, Govindarajan R (2012). Overcoming nucleoside analog chemoresistance of pancreatic cancer: a therapeutic challenge. *Cancer Lett*, **320**, 138-49.
- Khan S, Ebeling MC, Zaman MS, et al (2014). MicroRNA-145 targets MUC13 and suppresses growth and invasion of pancreatic cancer. *Oncotarget*, **5**, 7599-609.
- Koyama-Nasu R, Nasu-Nishimura Y, Todo T, et al (2013). The critical role of cyclin D2 in cell cycle progression and tumorigenicity of glioblastoma stem cells. *Oncogene*, **32**, 3840-5.
- Li J, Wood WH, 3rd, Becker KG, et al (2007). Gene expression response to cisplatin treatment in drug-sensitive and drug-resistant ovarian cancer cells. *Oncogene*, **26**, 2860-72.
- Lin Q, Chen T, Lin Q, et al (2013). Serum miR-19a expression correlates with worse prognosis of patients with non-small cell lung cancer. *J Surg Oncol*, **107**, 767-71.
- Longley DB, Johnston PG (2005). Molecular mechanisms of drug resistance. *J Pathol*, **205**, 275-92.
- McLachlan E, Shao Q, Wang HL, et al (2006). Connexins act as tumor suppressors in three-dimensional mammary cell organoids by regulating differentiation and angiogenesis. *Cancer Res*, **66**, 9886-94.
- Mini E, Nobili S, Caciagli B, et al (2006). Cellular pharmacology of gemcitabine. *Ann Oncol*, **17**, 7-12.
- Molina-Navarro MM, Trivino JC, Martinez-Dolz L, et al (2014). Functional networks of nucleocytoplasmic transport-related genes differentiate ischemic and dilated cardiomyopathies. A new therapeutic opportunity. *PLoS One*, **9**, 104709.
- Nakano Y, Tanno S, Koizumi K, et al (2007). Gemcitabine chemoresistance and molecular markers associated with gemcitabine transport and metabolism in human pancreatic cancer cells. *Br J Cancer*, **96**, 457-63.
- Nepusz T, Yu H, Paccanaro A (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods*, **9**, 471-2.
- Nguyen DX, Chiang AC, Zhang XH, et al (2009). WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. *Cell*, **138**, 51-62.
- Pan XH (2012). Pathway crosstalk analysis based on protein-protein network analysis in ovarian cancer. *Asian Pac J Cancer Prev*, **13**, 3905-9.
- Saiki Y, Yoshino Y, Fujimura H, et al (2012). DCK is frequently inactivated in acquired gemcitabine-resistant human cancer cells. *Biochem Biophys Res Commun*, **421**, 98-104.
- Saraswathy M, Gong S (2013). Different strategies to overcome multidrug resistance in cancer. *Biotechnol Adv*, **31**, 1397-407.
- Shi WY, Liu KD, Xu SG, et al (2014). Gene expression analysis of lung cancer. *Eur Rev Med Pharmacol Sci*, **18**, 217-28.
- Szakacs G, Paterson JK, Ludwig JA, et al (2006). Targeting multidrug resistance in cancer. *Nat Rev Drug Discov*, **5**, 219-34.
- Teraishi F, Zhang L, Guo W, et al (2005). Activation of c-Jun NH2-terminal kinase is required for gemcitabine's cytotoxic effect in human lung cancer H1299 cells. *FEBS Lett*, **579**, 6681-7.
- Tooker P, Yen WC, Ng SC, et al (2007). Bexarotene (LGD1069, Targretin), a selective retinoid X receptor agonist, prevents and reverses gemcitabine resistance in NSCLC cells by modulating gene amplification. *Cancer Res*, **67**, 4425-33.
- Toschi L, Cappuzzo F (2009). Gemcitabine for the treatment of advanced nonsmall cell lung cancer. *Oncotargets Ther*, **2**, 209-17.
- Toschi L, Finocchiaro G, Bartolini S, et al (2005). Role of gemcitabine in cancer therapy. *Future Oncol*, **1**, 7-17.
- Tufman A, Huber RM (2010). Biological markers in lung cancer: A clinician's perspective. *Cancer Biomark*, **6**, 123-35.
- Warde-Farley D, Donaldson SL, Comes O, et al (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*, **38**, W214-20.
- Zhang X, Jin FS, Zhang LG, et al (2013). Predictive and prognostic roles of ribonucleotide reductase M1 in patients with pancreatic cancer treated with gemcitabine: a meta-analysis. *Asian Pac J Cancer Prev*, **14**, 4261-5.